

UC Berkeley

UC Berkeley Electronic Theses and Dissertations

Title

Constraining NOx and CO Emission Factors from Urban Transportation and Residential Heating Using Dense Sensor Network Observations

Permalink

<https://escholarship.org/uc/item/3v78669c>

ISBN

9798189270826

Author

Zhu, Yishu

Publication Date

2026-05-01

Peer reviewed|Thesis/dissertation

Constraining NO_x and CO Emission Factors from Urban Transportation and Residential
Heating Using Dense Sensor Network Observations

by

Yishu Zhu

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Earth and Planetary Science

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Ronald C. Cohen, Chair

Professor Joshua S. Apte

Professor Inez Fung

Spring 2026

Constraining NO_x and CO Emission Factors from Urban Transportation and Residential
Heating Using Dense Sensor Network Observations

Copyright 2026
by
Yishu Zhu

Abstract

Constraining NO_x and CO Emission Factors from Urban Transportation and Residential Heating Using Dense Sensor Network Observations

by

Yishu Zhu

Doctor of Philosophy in Earth and Planetary Science

University of California, Berkeley

Professor Ronald C. Cohen, Chair

Cities are dominant sources of air pollutants and greenhouse gases, with transportation and residential combustion representing two major sectors. Yet urban emissions are rarely observed at the spatial resolution needed to resolve source heterogeneity or evaluate the efficacy of local emission reduction efforts. As a result, persistent discrepancies remain between activity-based emission inventories and estimates derived from atmospheric observations, with unknown uncertainties across sectors. Seeking to address that gap, I developed and applied an analytical and calibration framework that leverages dense networks of miniaturized sensors to infer sectoral emission factors in cities and diagnose biases in current emission inventories.

Dense urban sensor networks, such as the Berkeley Environmental Air Quality and CO_2 Network (BEA CO_2N) used in this study, have unveiled fine-scale spatial variability in atmospheric composition that was previously unresolved due to the limited spatial density of traditional observing systems. This new observational capability also raises a new challenge: how to translate dense concentration measurements into quantitative constraints on emission sources. In the first part of this work, I developed an analytical approach to inferring transportation and residential heating emission factors in the San Francisco (SF) Bay Area. The framework models their decay away from major highways using a simplified transport scheme and interprets spatial patterns in enhancement ratios (ERs), where enhancements are defined as the difference between measured concentrations and a network-derived background (i.e., $\Delta\text{NO}_x/\Delta\text{CO}_2$ and $\Delta\text{CO}/\Delta\text{CO}_2$). Comparing the inferred emission factors with those assumed in existing inventories revealed systematic differences, particularly for residential NO_x emissions.

Ensuring accurate and consistent sensor measurements across diverse urban settings presents a second challenge, as co-located reference instruments are often sparse or absent depending on deployment location. To address this limitation, the second part of the dissertation de-

veloped a physically grounded calibration strategy designed for dense sensor networks. The method combines multi-variable linear regression with constraints from atmospheric chemistry, specifically the NO-O₃ titration events that frequently occur in urban environments. This calibration framework enables robust measurement of multiple pollutants while minimizing dependence on co-located reference instruments, thereby providing a scalable pathway for maintaining measurement accuracy in large urban sensor networks at manageable cost.

Equipped with the analytical and calibration frameworks developed above, I applied them in the third part of this dissertation to multiple U.S. cities where BEACO₂N observations were established—the SF Bay Area, CA, Los Angeles, CA, and Providence, RI—to track and assess how urban emissions vary across space and time under differing infrastructure, energy use patterns, and regulatory implementations, an area that still requires further attention from the research community. Transportation emissions were broadly consistent with inventory expectations, while residential heating emissions showed larger discrepancies. Inventories overestimated residential NO_x emission factors by a factor of 2–3, and residential CO emission factors exhibited inter-city variability. These results identify distributed residential combustion as a major remaining uncertainty in urban emission inventories and highlight the value of dense sensor networks for diagnosing sector-specific emissions across diverse cities.

Together, this dissertation demonstrates that dense urban sensor networks, combined with scalable and physically grounded calibration, provide the observational resolution needed to refine emission inventories and inform city air quality and climate actions.

To my parents,
to everyone who has made this life meaningful,
to everything that has brought me here.

Contents

Contents	ii
List of Figures	iv
List of Tables	xiii
1 Introduction	1
1.1 Urban Emissions, Emission Factors, and the Central Question	1
1.2 Observational Approaches to Constraining Urban Emissions	2
1.3 Calibration as a Scientific Prerequisite	3
1.4 From Bay Area Inference to Cross-City Evaluation	4
1.5 Aims and Chapter Overview	5
2 Observational Inferences of Emission Factors	7
2.1 Introduction	7
2.2 Materials and Methods	9
2.2.1 BEACO ₂ N Measurements	9
2.2.2 Enhancement Ratio Calculation	9
2.2.3 Ancillary Datasets	11
2.3 Results and Discussion	12
2.3.1 Idealized Mixing of Transportation and Residential Emissions	12
2.3.2 Observed Emission Ratios versus Distance to Highway	14
2.3.3 BEACO ₂ N-Derived Emission Factors Compared to CARB and GRA ² PES Emission Inventories	18
2.4 Conclusions	19
3 Physically Constrained NO₂ Sensor Calibration	20
3.1 Introduction	20
3.2 Methods and Materials	23
3.2.1 Sensor network and reference measurements	24
3.2.2 Calibration framework overview	25
3.2.3 Validation design	25
3.2.4 Stage 1: Reference-based baseline calibration	27

3.2.5	Stage 2: Titration-based gain correction	29
3.2.6	Stage 3: Rolling offset correction	32
3.3	Results and Discussion	34
3.3.1	Validation against withheld colocated truth	34
3.3.2	Titration ratio diagnostics	36
3.3.3	Remote calibration error structure	38
3.3.4	Gain-curve convergence across nodes and cities	42
3.3.5	Recovery of site-level NO ₂ variability	44
3.3.6	Limitations	45
3.4	Conclusions	47
4	Cross-City Emission Factor Comparison	49
4.1	Introduction	49
4.2	Measurements and Methods	51
4.2.1	Networks and observations	51
4.2.2	Enhancement ratio computation and distance-decay fitting	51
4.2.3	Comparison with GRA ² PES inventory	53
4.3	Results and Discussion	53
4.3.1	Transportation emission factors from near-highway gradients	54
4.3.2	Residential emission factors from the far-field plateau	57
4.3.3	Limitations and outlook	60
4.4	Conclusions	61
5	Concluding Remarks	63
	Bibliography	65
A	Supplementary to Chapter 2	75
B	Supplementary to Chapter 3	85
C	Supplementary to Chapter 4	118

List of Figures

2.1	The location of BEACO ₂ N sites (left panel) in the SF Bay Area and time series (right panel) of (a) NO, (b) NO ₂ , (c) CO, and (d) CO ₂ measurements from September 2021 to April 2023. All available observations in the BEACO ₂ N network are shown as gray dots, while median values of the network for 1-hour time bins are shown by the solid lines. Background map credits: © OpenStreetMap (https://www.openstreetmap.org/copyright) contributors; © CARTO (https://carto.com).	10
2.2	The simulated mixing of area source (residential heating) and line source (transportation) within a 10 km × 10 km modeling domain. The spatial gradient profiles of enhancement ratios under different relative emission source strengths (area source: line source) are labeled by different colors.	14
2.3	BEACO ₂ N observed $\Delta\text{NO}_x/\Delta\text{CO}_2$ and $\Delta\text{CO}/\Delta\text{CO}_2$ in 2022 versus distance to highways. Panels (a) and (c) show the summer values (June–September), and panels (b) and (d) display winter values (January, February, and December). Morning (6:00–9:00 AM local time) data and distance-decay fitting lines are represented in green for $\text{ER}_{\text{NO}_x,\text{integrated}}(d)$ and pink for $\text{ER}_{\text{CO},\text{integrated}}(d)$. Midnight (0:00–3:00 AM local time) data and distance-decay fitting lines are shown in blue for $\text{ER}_{\text{NO}_x,\text{integrated}}(d)$ and purple for $\text{ER}_{\text{CO},\text{integrated}}(d)$. Only data corresponding to wind directions between 120° and 260° within each time window were included. Uncertainties of the ER estimates are represented as error bars. Parameters of the exponential fitting for each line are provided in Table 1.	16
3.1	Study-region map of BEACO ₂ N nodes and reference monitors	23
3.2	NO ₂ calibration pipeline overview	26
3.3	Calibrated NO ₂ versus withheld colocated truth at site 43 for remote WLS (left column) and remote WLS+GC (right column), evaluated over three windows: summer (JAS; top), winter (DJF; middle), and full year (bottom). Colour indicates hexbin point density on a log scale. Dashed line: 1:1. Metrics are computed from a single representative remote reference assignment; Table 3.5 reports the median across all four assignments.	37
3.4	Titration ratio R vs. environmental variables, node 261	39
3.5	Residual diagnostics for remote calibration	41
3.6	Cross-city gain curve comparison	43

3.7	Network-wide calibration diagnostics, Providence	45
4.1	Sensor node locations for the three BEACO ₂ N deployments: (a) San Francisco Bay Area, (b) Los Angeles, (c) Providence. Major highways (Interstate and state routes) are shown as grey lines. Filled blue circles indicate where the sensor nodes are installed.	52
4.2	Enhancement ratios as a function of distance to the nearest highway for the Bay Area summer morning rush hour and Providence winter morning rush hour: (a) $ER_{NO_x} = \Delta NO_x / \Delta CO_2$ and (b) $ER_{CO} = \Delta CO / \Delta CO_2$. Points show per-node ER values from bivariate least-squares regression of ΔX versus ΔCO_2 hourly pairs within the respective seasonal, wind-direction, and time-of-day windows. Dashed curves are least-squares fits to Equation 4.1. Stars mark $EF_{transportation} = a + c$ obtained by extrapolating the fit to $d = 0$. Per-node error bars are bivariate regression uncertainties. Fit parameters are listed in Table 4.1. Per-city analysis settings are listed in Supplementary Table S1; full ER-versus-distance panels for all profiles are shown in Supplementary Figures S1–S3.	55
4.3	Winter midnight enhancement ratios as a function of distance to the nearest highway for the three BEACO ₂ N cities: (a) $ER_{NO_x} = \Delta NO_x / \Delta CO_2$ and (b) $ER_{CO} = \Delta CO / \Delta CO_2$. Each point represents one sensor node from its city’s winter midnight pool. Dashed horizontal lines indicate the fitted far-field plateau c for each city, with shaded bands showing the fit uncertainty. The vertical dotted line at 2 km marks the threshold beyond which near-highway contribution is negligible given the k values recovered in Figure 4.2. The plateau c is identified with $EF_{residential}$ following Zhu et al. [107].	58
A.1	Map of BEACO ₂ N and MesoWest site locations in the study region. Background map credits: © OpenStreetMap contributors © CARTO.	76
A.2	Wind speed distributions from MesoWest measurements and HRRR reanalysis for June–September and December–February.	76
A.3	Modeled ER profile comparisons for $WS = 0.5, 1, 1.5, 3$, and 5 m/s.	79
A.4	Comparison of simulated ER profiles using NO_x lifetimes of $\infty, 10, 5, 1$ and 0.2 hours.	80
A.5	Monthly vehicle miles traveled (VMT) and monthly residential natural gas consumption in the SF Bay Area in 2021–2023.	80
A.6	Map of BEACO ₂ N site locations labeled with site ID. Background map credits: © OpenStreetMap contributors © CARTO.	84
S1	Progressive comparison results at site 42 (EBMUD truth), in the same format as main-text Table 3.5: four rows $\times$ three evaluation windows (full year, JAS, DJF), metrics = RMSE (ppb), bias (ppb), OLS slope, R^2 . Results are consistent with site 43 (main text), confirming that the progressive comparison findings are not specific to a single site or remote reference assignment.	91

S2	Seasonal RMSE (ppb; top) and OLS slope (bottom) for all five calibration methods (Table 3.2) at both validation sites (43 and 42), presented as box plots across all validation configurations. Rows correspond to the three evaluation windows (full year, JAS, DJF); columns correspond to the two metrics. Each box spans the interquartile range across jobs for a given method and site; the horizontal line shows the median. JAS consistently shows the largest improvement from the gain correction (WLS+GC vs WLS remote), consistent with warm temperatures and elevated O_3 producing the strongest gain error in summer. DJF results confirm that the correction remains effective in cold conditions. Colocated methods are shown as horizontal reference lines (single values, not varying across jobs). . . .	92
S3	Per-pair OLS slope heatmap for the WLS+GC remote pipeline (Row 5), showing slope against the withheld colocated truth for all combinations of test site (rows) and remote reference station (columns). Each cell reports the mean slope across month strategies. Colour scale: green = slope near 1; red = slope well below 1. The heatmap confirms that the gain correction produces consistent slope recovery across different remote reference assignments.	93
S4	Kernel density estimates (KDE) of the titration ratio R before (grey fill) and after (red line) per-node gain and offset correction for all 22 Providence nodes, arranged in a grid ordered by node ID. The dashed vertical line in each panel marks $R = 1$. Median R before and after is annotated in each panel. All panels share an identical x -axis range ($R \in [0, 2]$) for direct comparison. The anomalous node ([ID]; red border) retains a post-correction median R that does not converge toward unity, consistent with its positive dg/dT slope.	94
S5	Per-node KDE of R before (grey fill) and after (red line) correction for all nodes in the San Francisco Bay Area, in the same format as Supplementary Figure S4.	95
S6	Per-node KDE of R before (grey fill) and after (red line) correction for all nodes in Los Angeles, in the same format as Supplementary Figure S4. A residual post-correction negative skew is visible at most nodes and is most pronounced at eastern basin nodes where O_3 is highest.	96
S7	Calibration performance spread across remote reference assignments for both Bay Area validation sites (43 top, 42 bottom). Panels show RMSE (ppb), OLS slope, and R^2 evaluated against the withheld colocated truth. Each point represents one remote reference station; box plots show the interquartile range across references. Three remote calibration methods are compared: MLR (CO_2 index), WLS (CO_2 index), and WLS+GC (full pipeline). WLS+GC consistently achieves the highest slope and tightest cross-reference spread, confirming that the gain correction reduces dependence on which remote reference station is used for calibration. . .	96

S8	Temperature dependence of the NO cross-interference term ($\text{term_no} = \beta_{\text{NO}} \cdot V_{\text{NO}}$, ppb) pooled across all Providence nodes, shown as a hexbin density plot on a logarithmic colour scale. Blue points show the bin median. The dashed line marks zero. The bin medians remain near zero across the full temperature range, with no systematic trend. The median Spearman ρ across nodes is 0.530, but reflecting scatter rather than a coherent directional bias. This near-zero median term_no justifies excluding it from the gain-scaled sum S_{gain} (Equation B.1; S-Methods-1), and confirms that the gain correction need not be applied to the NO cross-interference term.	97
S9	Fraction of minute-level NO_2 data for which $T > T_{\text{CAP}}$ and the WLS output is therefore preserved unchanged (gain correction bypassed) in each city. All cities use $T_{\text{CAP}} = 35^\circ\text{C}$. The fraction of data above the cap is small in all cities, indicating that the T_{CAP} guard does not substantially limit the temporal coverage of the correction.	98
S10	Mean NO_2 zero-level estimate Z_{NO_2} (ppb) as a function of ambient temperature for all nodes in each city. Three panels: (a) Providence, (b) San Francisco Bay Area, (c) Los Angeles. Thin grey lines = individual nodes; bold coloured line = network median. The floor is positive in all cities, reflecting the O_3 cross-sensitivity contribution to S_{gain} under near-zero NO_2 conditions. A residual temperature dependence of Z_{NO_2} is visible — the floor is systematically lower at warm temperatures in all three cities.	98
S11	Z_{NO_2} network distribution	99
S12	Z_{NO_2} parameter distributions	100
S13	Z_{NO_2} predictor model comparison	101
S14	WLS coefficient time series (Providence)	102
S15	WLS coefficient time series (Bay Area)	103
S16	WLS coefficient time series (Los Angeles)	104
S17	Pairwise mean NO_2 bias (ppb), RMSE, and correlation among EPA reference stations in the San Francisco Bay Area, computed over the study period as a function of inter-station distance. Pairwise biases represent the irreducible spatial heterogeneity in NO_2 across the Bay Area regulatory network. This range is used in the main text as the spatial representativeness benchmark against which post-correction bias is judged.	105

- S18 Titration event frequency as a function of calendar month (x-axis) and local hour of day (y-axis) for the three BEACO₂N cities: (a) Providence, (b) San Francisco Bay Area, (c) Los Angeles. Colour indicates the number of events detected per node-month in each month-hour bin, averaged across all nodes. Events are most frequent in Providence during winter evenings (November–March, 17:00–23:00), where long nights and sustained heating emissions produce frequent fresh-NO pulses into an O₃-rich residual layer. Bay Area events are similarly concentrated but show suppression in summer months due to marine-layer O₃ depletion. Los Angeles events are more evenly distributed year-round, consistent with narrower seasonal variation in nocturnal O₃ and NO in the basin. 105
- S19 Rolling monthly midday offset $b(m)$ (ppb) applied in Stage 3, shown as a time series for all nodes in each city. Three panels: (a) Providence, (b) San Francisco Bay Area, (c) Los Angeles. Thin grey lines = individual nodes; bold coloured line = representative node. Providence shows larger positive offsets in winter, consistent with a mild residual sensor low-bias at cold temperatures not fully resolved by the gain correction and with the seasonal spatial gradient between nodes and reference stations. Los Angeles shows a persistent small positive offset in summer, consistent with residual O₃ cross-sensitivity under high-ozone conditions. The offset magnitude is generally smaller than the gain correction at any individual temperature but provides a meaningful improvement in long-term mean bias. 106
- S20 OAT event detection sensitivity analysis, part 1: baseline stability (**max_base_span**), end condition (**end_k**), and end mode (**dual_end**). For each parameter, the figure shows: **left** — median per-node $g(T)$ with IQR shading for each tested value, compared to the default (bold black); **right** — post-correction KDE of R (network-pooled) for each value. Parameter values tested are listed in Supplementary Table S3. The gain curve changes by less than 0.04 units at any temperature bin across all tested configurations, confirming robustness of the gain correction to reasonable variation in these detection parameters. 107
- S21 OAT event detection sensitivity analysis, part 2: **(a)** night window definition (19:00–05:00 default, 21:00–04:00, 22:00–04:00); **(b)** R_CLIP_FIT lower bound; **(c)** temperature bin width (1°C, 2°C default, 5°C). Each row shows $g(T)$ (left), post-correction R distribution, calibration metrics, event count, and related diagnostics. The 2°C default provides the best balance of resolution and stability. 108

- S22 Sensitivity of calibration diagnostics to five pipeline parameters. Each row shows $g(T)$, post-correction R distribution, and calibration metrics for a swept parameter. **(a)** WLS regression window length (3, 5, 7, 9, 12 months). **(b)** Stage 3 offset window length (3, 5, 7 months). **(c)** Stage 3 offset percentile threshold (5th, 10th, 15th, 20th, 25th). **(d)** CO₂ enhancement index cap (10%–30%). **(e)** Z_{NO_2} predictor model (V_{O_x} alone, $+T$, $+RH$, $+T+RH$, $T+RH$). All models that include V_{O_x} produce nearly identical gain curves; the $T + RH$ model without the electrode term performs substantially worse. Final R diagnostics are comparable across all tested configurations. 109
- S23 Comparison of four bin-weighting schemes for the gain curve construction. **Left:** median per-node $g(T)$ with IQR shading for each scheme. **Right:** KDE of the network-pooled R distribution after correction under each scheme, with the dashed vertical line at $R = 1$. Schemes: $w_{\Delta\text{NO}_2}$ (chosen; solid black), w_{base} (blue dashed), w_{unw} (unweighted median; orange dashed), $w_{\Delta\text{O}_3}$ ($|\Delta\text{O}_3|$ -weighted; green dashed). All four schemes produce statistically indistinguishable post-correction R distributions (Supplementary Table S4), confirming that the main results are not sensitive to the choice of weighting scheme. 110
- S24 Sensitivity of the per-node gain curve $g(T)$ to the choice of O₃ reference station used in the gain correction, evaluated on the Providence network. **Left:** median per-node $g(T)$ and inter-node IQR (shading) computed using each node’s assigned O₃ reference (default; black) versus alternative EPA O₃ stations. **Right:** distribution of the per-node absolute difference in $g(T)$ at each temperature bin between the default and each alternative O₃ reference. The curves differ by less than 0.03 units at any temperature bin, indicating that the circular dependency between O₃ and NO₂ calibrations does not materially bias the gain curve. 111
- S25 Joint sensitivity of the gain curve and calibration performance to the Z_{NO_2} regression window length and NO₂ background percentile threshold, evaluated on the Providence network. **(a)** Window length OAT: regression window varied across 1, 2, 3 (default), 4, and 6 months with the percentile held at 10th. Shorter windows (≤ 2 months) produce noisier α and β estimates; longer windows (≥ 4 months) smooth over within-season variation. **(b)** Percentile OAT: NO₂ percentile varied across 2nd, 5th, 10th (default), 15th, and 25th with the window held at 3 months. Higher percentiles include hours with non-negligible NO₂, biasing the regression. **(c)** Joint 2D heatmap of calibration metrics across all window $\times$ percentile combinations. The adopted configuration (3 months, 10th percentile) lies in the optimal region. The interaction between the two parameters is weak. . . . 112

- S26 Sensitivity of calibration performance to the Z_{NO_2} scaling factor λ and offset formulation. **(a)** Calibration metrics as a function of the scaling factor $\lambda \in \{0, 0.3, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 1.25\}$. $\lambda = 0$ corresponds to no zero-level subtraction; $\lambda = 0.7$ is the adopted default. Performance degrades at both extremes, confirming that the zero-level estimate is well-calibrated near $\lambda = 0.7$. **(b)** Effect of scaling factor on the gain curve $g(T)$. **(c)** Comparison of the temperature-dependent monthly offset $b(T, m)$ (default) versus a scalar monthly offset $b(m)$ and no offset ($b = 0$). The temperature-dependent form produces lower RMSE and higher slope than the scalar, consistent with its ability to address residual temperature structure. 113
- S27 Investigation of the 15–20°C gain curve feature observed at a Bay Area site. Three diagnostic angles: **(a)** Seasonal event pools: gain curves fitted using only events from restricted calendar months to determine whether the bump is seasonal. The bump persists in all seasonal subsets. **(b)** Gain curve smoothing: Savitzky–Golay and rolling-mean smoothers applied to the raw bin medians. Smoothing attenuates but does not eliminate the feature. **(c)** Upper R clip: R_{hi} varied. The bump is reduced when R_{hi} is lowered, suggesting that a small population of high- R events in the 15–20°C range contributes to the feature. 114
- S28 Sensitivity of the gain curve and calibration performance to an OLS R^2 quality threshold applied to individual titration events. **(a)** Median per-node $g(T)$ with IQR shading and post-correction R KDE for R^2 thresholds of 0.0 (no filter; default), 0.3, 0.5, 0.7, and 0.8. The gain curve is stable across thresholds up to 0.5; above 0.7, warm-temperature bins become under-sampled. **(b)** Event count per temperature bin across thresholds, showing that warm-temperature bins are disproportionately depleted by stricter thresholds. **(c)** Calibration RMSE and slope against withheld truth at the Bay Area validation site for each threshold. **(d)** Seasonal breakdown (full year, JAS, DJF) confirming that JAS performance is most sensitive to event depletion. 115
- S29 Sensitivity of calibration performance to the temperature cap T_{CAP} above which the gain correction is bypassed. T_{CAP} values of 25, 28, 30, 33, 35 (default), 40, and 45°C are tested at the Bay Area validation site. **(a)** RMSE and slope against withheld truth for each cap value. Lower caps bypass the correction at progressively more hours; higher caps extend the correction into temperature bins where events are sparse. **(b)** Effect of each cap value on the gain curve $g(T)$. **(c)** Fraction of data bypassed at each cap value per city. The adopted $T_{\text{CAP}} = 35^\circ\text{C}$ bypasses a small fraction of data while avoiding extrapolation into unsupported temperature ranges. 116
- S30 Validation results across all site–reference pairs in the Bay Area, showing the full matrix of calibration performance under the WLS+GC remote pipeline. Each panel corresponds to a different combination of validation site and remote reference station, confirming that the gain correction produces consistent improvement regardless of which remote reference is used. 117

S1	Bay Area enhancement ratios as a function of distance to the nearest highway: NO_x/CO_2 and CO/CO_2 for summer and winter, with morning rush (AM) and midnight (MN) overlaid. The summer AM NO_x and CO profiles are the cases used for the Bay Area transportation EF in main-text Table 4.2. Per-profile settings are listed in Table S1.	124
S2	Providence enhancement ratios as a function of distance to the nearest highway, same layout as Figure S1. The winter AM NO_x and CO profiles are the cases used for the Providence transportation EF in main-text Table 4.2. Per-profile settings are listed in Table S1.	125
S3	Los Angeles enhancement ratios as a function of distance to the nearest highway, same layout as Figure S1. No exponential decay is resolvable because all nodes lie beyond the near-highway k scale recovered in the Bay Area and Providence. Fitted EFs show very large uncertainties or non-convergent fits (Table S1), confirming that the network geometry does not support transportation EF recovery.	126
S4	Residential heating fuel shares for the three study regions, from the 2023 American Community Survey 5-year estimates. The Bay Area and Los Angeles are dominated by utility gas ($\sim 55\text{--}60\%$), whereas Providence has a substantial fuel oil and kerosene fraction ($\sim 15\%$) alongside gas ($\sim 68\%$). This contrast in residential fuel mix is the basis for interpreting the cross-city variation in observed residential emission factors (Section 4.3.2). Error bars indicate the survey margin of error.	127
S5	Monthly composition and annual distribution of CO_2 emissions from GRA ² PES v1.0, 2021. Stacked areas (left axis) show the fractional contribution of each sector to the monthly total, renormalised so each month sums to unity. The overlaid dashed line (right axis) shows each month's share of the annual total, so the 12 monthly values integrate to 100%.	128
S6	Monthly composition and annual distribution of NO_x emissions from GRA ² PES v1.0, 2021, same layout as Figure S5. Residential combustion contributes a larger fraction of winter NO_x in Providence than in the California cities, consistent with the stronger observed residential EF_{NO_x} (Section 4.3.2).	129
S7	Monthly composition and annual distribution of CO emissions from GRA ² PES v1.0, 2021, same layout as Figure S5. Off-road sources (red) account for roughly 30–50% of total CO emissions in all three cities, a substantially larger share than for CO_2 or NO_x . This structural difference in the CO emission budget is discussed in Section 4.3.2.	130
S8	Seasonal cycle of observed far-pool enhancement ratios (distance > 2 km) at midnight hours for (a) Bay Area, and (b) Providence. The elevated winter ERs in Providence are consistent with seasonal heating activity.	131

S9	Sectoral NO _x fractions during morning rush (06–09 local) and midnight (00–03 local), January 2021. Transportation dominates at rush hour in all three cities. At midnight, residential and commercial combustion account for larger shares, confirming that the midnight observational window isolates non-transportation sources.	132
S10	Sectoral CO fractions during rush and midnight, January 2021. The residential fraction grows substantially from rush to midnight in all three cities.	133
S11	Sectoral CO ₂ fractions during rush and midnight, January 2021. CO ₂ shifts from transportation-dominated at rush to residential-dominated at midnight in the Bay Area and Providence.	134
S12	Sectoral NO _x fractions during rush and midnight, July 2021. Summer rush NO _x is more transportation-dominated than in winter.	135
S13	Sectoral CO fractions during rush and midnight, July 2021. Summer rush CO is gasoline-vehicle-dominated. Midnight summer CO shows large off-road fractions in the Bay Area and Los Angeles.	136
S14	Sectoral CO ₂ fractions during rush and midnight, July 2021.	137
S15	Bay Area per-pixel NO _x /CO ₂ , January 2021. Gasoline and diesel panels are single-valued; the residential panel shows mild texture (CoV $\approx$ 0.06).	138
S16	Bay Area per-pixel NO _x /CO ₂ , July 2021.	138
S17	Los Angeles per-pixel NO _x /CO ₂ , January 2021.	139
S18	Los Angeles per-pixel NO _x /CO ₂ , July 2021.	139
S19	Providence per-pixel NO _x /CO ₂ , January 2021.	140
S20	Providence per-pixel NO _x /CO ₂ , July 2021.	140
S21	Bay Area per-pixel CO/CO ₂ , January 2021. The residential panel shows substantial sub-grid texture (CoV $\approx$ 0.71, p_{10} – p_{90} = 2.3–17.3 ppbv/ppmv), absent in the gasoline and diesel panels.	141
S22	Bay Area per-pixel CO/CO ₂ , July 2021. Summer residential CO shows similar texture (CoV $\approx$ 0.68), confirming a structural inventory feature rather than a seasonal artefact.	141
S23	Los Angeles per-pixel CO/CO ₂ , January 2021. The residential panel is more uniform than the Bay Area (CoV $\approx$ 0.44).	142
S24	Los Angeles per-pixel CO/CO ₂ , July 2021.	142
S25	Providence per-pixel CO/CO ₂ , January 2021 (CoV $\approx$ 0.52).	143
S26	Providence per-pixel CO/CO ₂ , July 2021.	143

List of Tables

2.1	Summary of e-folding distances (k), mixing length scales ($k/0.36$), and integrated emission ratios for NO_x and CO at peak and low traffic volumes, and during periods of peak and low natural gas usage.	17
2.2	BEACO ₂ N-derived emission factors and sectoral emission factors in current emission inventories.	19
3.1	Deployment summary	24
3.2	Progressive comparison design	27
3.3	Event detection parameters	30
3.4	Calibration models for all gas channels	33
3.5	Progressive comparison results at site 43, evaluated against the withheld colocated reference on hourly data. Colocated methods use on-site reference hours and yield single values; remote methods report the median ± 1 s.d. across four remote reference assignments. Slope is from OLS regression of calibrated values against truth (ideal = 1.0). Site 42 results are in SI Table S3.	35
3.6	Cross-city R diagnostics	44
4.1	Fitted parameters of Equation 4.1 for the near-highway decay cases shown in Figure 4.2. a : near-highway amplitude; k : e-folding mixing length; c : far-field plateau; $\text{EF}_{\text{transportation}}$ ($= a + c$): transportation emission factor at $d = 0$. All emission-related quantities are in ppbv/ppmv; k is in meters.	55
4.2	Comparison of observed transportation emission factors with GRA ² PES on-road sub-sector values. Observed EFs are the extrapolated near-highway value $a + c$: Bay Area from summer morning rush hour (JJA, 06–09 local), Providence from winter morning rush hour (DJF, 05–09 local). Los Angeles is omitted because the network geometry does not resolve a transportation EF. GRA ² PES columns show on-road gasoline, on-road diesel, and the combined on-road total, extracted over each network footprint for the corresponding season and time-of-day window. Full sub-sectoral breakdown in Supplementary Table S2. All values in ppbv/ppmv.	56

4.3	Comparison of observed residential emission factors with GRA ² PES residential-sector values. Observed values are the far-field plateau c from the winter midnight ER profiles (Figure 4.3). GRA ² PES values are the pixel-median EF within each network footprint (winter DJF, 00–03 local window); the median is less sensitive to the heavy-tailed spatial distribution of residential CO than the pixel-mean aggregate (see Supplementary Table S3 for per-pixel statistics). All values in ppbv/ppmv.	59
A.1	Description and values of parameters used in the 3-D advection-diffusion model.	79
A.2	Summary of BEACO ₂ N site locations and inclusion status for ER analysis. Symbols indicate inclusion status: ✓ = passed all inclusion criteria and used in ER distance-decay fitting; • = passed data completeness but not the R^2 threshold; ○ = no nearby upwind highway; blank = did not pass data completeness criteria.	81
S1	AirNow reference station metadata for all three BEACO ₂ N deployments. Data capture rate is the annual percentage of valid hourly records passing quality screening. Distance to nearest node is the minimum node-to-station distance within each city. N assigned is the number of sensor nodes whose Stage 1 and 3 calibration is anchored to that station.	89
S2	Sensitivity analysis: parameter ranges and outcome summary. For each parameter varied, the table reports the default value, range tested, and physical motivation. Tests on event detection parameters (Group A) were conducted on the Providence network; pipeline parameters (Groups B–I) use the Bay Area validation site unless noted.	90
S3	Progressive comparison results at site 42 (EBMUD truth), evaluated against the withheld colocated reference on hourly data. Colocated methods use on-site reference hours and yield single values; remote methods report the median across four remote reference assignments. Slope is from OLS regression of calibrated values against truth (ideal = 1.0). Compare with main-text Table 3.5 (site 43). .	91
S1	Per-profile analysis settings and fitted transportation emission factors for all 24 city / species / season / time-of-day combinations. Months: calendar months in each seasonal pool. h : starting local hour (AM = morning rush, MN = late night at 00:00). WD: wind-direction window (degrees). $R_{\min}^2$: minimum bivariate regression R^2 for node inclusion. $\text{EF}_{\text{transportation}} = a + c$: fitted transportation emission factor (ppbv/ppmv) with ± 1 SE. N : nodes passing all filters. Profiles where the covariance matrix could not be estimated are marked $\pm\infty$	120
S2	Sub-sectoral GRA ² PES emission factors across season and time-of-day slices. Extends main-text Tables 4.2 and 4.3. All values in ppbv/ppmv.	121

S3	Per-pixel spatial variability of GRA ² PES emission ratios across each network footprint. For each combination, the ratio $X_p/\text{CO}_{2,p}$ is computed at each masked grid pixel. CoV = SD/mean. On-road rows are abbreviated because all pixels carry the same value within each city. All values in ppbv/ppmv except CoV (dimensionless). Abbreviations: BA = Bay Area, LA = Los Angeles, PVD = Providence.	122
S4	Residential far-field plateau c ($= \text{EF}_{\text{residential}}$) for each city and species, winter midnight pool. ± 1 SE from the nonlinear least-squares fit. N : number of nodes beyond 2 km contributing to each fit.	123

Acknowledgments

A dissertation, like a long voyage, is rarely undertaken alone. The pages that follow bear my name, but they carry the patience, generosity, and kindness of many who walked part of this road with me.

I am deeply grateful to my advisor, Ron Cohen, for the freedom to explore and the guidance to find my way back. He taught me how to ask the right questions, and to keep asking when answers did not come easily. His trust and support shaped not only this work, but also the kind of scientist and mentor I hope to become.

My thanks also go to Inez Fung, my second advisor, for her rigor, wisdom, and warmth. I will always remember her curiosity about the world and her creative way of thinking, which continually encouraged me to see problems from new perspectives.

I am thankful for the guidance of my committee members, Josh Apte and Nick Swanson-Hysell, whose thoughtful questions and sharp insights pushed this work further while giving me room to grow into it. I am also grateful to Allen Goldstein and Michael Manga at Berkeley, and to the mentors at Peking University and beyond, who have been generous with their time and their advice at every turn.

Sharing an office for nearly five years with Anna and Milan was one of the great fortunes of this journey. What is an office, if not a small world built on shared milk tea, frustration, and joy? We grew together in that space, and I would not have made it through without them.

The Cohen Lab made this experience both intellectually rewarding and personally enjoyable. Naomi, Pietro, Hikari, Sam, Danika, Natalie, Serena, Raffy, Toireasa — thank you for making the long days lighter and the difficult ones manageable. You were collaborators, sounding boards, and friends. Lab alumni and visitors, especially Alex Turner and Josh Laughner, offered mentorship that arrived exactly when it was needed.

Time spent with the Goldstein, Apte, and Ivey labs was always a pleasure. Beyond the lab, friends across Chemistry, EPS, and CEE — Wendy, Becca, Anika, Erin, Michael, Noella, Yiming, Ziman, Yuancong, Langlang, Xiaodi — brought balance to these years. The dinners, conversations, and quiet solidarity made graduate school feel far less solitary.

My roommates, Yaqin, Sen, and Yi, made Berkeley feel like home. Their kindness, shared meals, and late-night conversations turned an apartment into a place of rest and belonging.

Across continents, Meng and Taiyi remained constant. Distance never made them feel far. Their steady presence across years and time zones kept me grounded.

Finally, my deepest gratitude goes to my family in China. My mother, Beibei Liu, and my father, Zhenghua Zhu, have given me more than I can express. They raised me to be curious, to be brave, and to know I always have a way home. My grandmothers, Jingxian Guo and Jinlan Yuan, have been the quiet foundation beneath everything I am. And my uncle, aunt, and cousins, who crossed oceans to visit me, reminded me that distance is never truly distance when the heart knows where home is.

Chapter 1

Introduction

1.1 Urban Emissions, Emission Factors, and the Central Question

Fossil fuel combustion in cities produces both air pollutants—nitrogen oxides (NO_x), carbon monoxide (CO), and particulate matter—and the greenhouse gas carbon dioxide (CO_2). Because more than half of the global population now lives in urban areas, and because cities concentrate transportation, building energy use, and industrial activity within relatively small footprints, urban combustion sources exert an outsized influence on both regional air quality and global climate forcing [1, 2, 3, 4]. Among these sources, on-road transportation and residential heating are especially important: they are spatially distributed throughout the urban fabric, difficult to measure directly, and central to policies aimed at improving air quality and reducing greenhouse gas emissions [5, 6, 7]. Accurate and current quantification of emissions from these sectors is therefore essential not only for tracking progress toward climate goals, but also for evaluating whether the anticipated co-benefits for public health are being achieved.

Urban emissions are commonly represented through inventories that have historically developed along two largely separate paths. Air pollutant inventories, such as those produced by the U.S. Environmental Protection Agency National Emission Inventory (NEI) and the California Air Resources Board (CARB), are designed to support air quality management and are typically updated on multi-year cycles [8]. Greenhouse gas inventories, including research products such as Vulcan, Hestia, ACES, and the more recent GRA^2PES , provide CO_2 estimates at spatial resolutions of 1–10 km and have emerged primarily from the research community [9, 10, 11, 12, 13]. Despite their different policy contexts, both traditions rely on the same general accounting structure: emissions are estimated by multiplying activity data—such as vehicle miles traveled or therms of natural gas consumed—by emission factors that describe the amount of pollutant released per unit of activity [14]. While emissions from large point sources can often be tied to direct facility-level measurements and remain reasonably current, distributed urban sources such as vehicles and residential heating appliances are

more difficult to constrain. Their representation in inventories depends heavily on assumed emission factors and source allocations that may not reflect the conditions, technologies, or fuel mixes actually present in a given city at a given time [15, 16, 17, 18, 19].

Emission factors are not merely accounting parameters; they are also physical descriptors of combustion processes. The ratio of a co-emitted pollutant to CO_2 —expressed, for example, in ppbv of NO_x per ppmv of CO_2 —encodes information about combustion temperature, fuel type, installed technology, and the overall source mixture contributing to an observation [20, 21]. These quantities vary across vehicle classes, engine operating conditions, appliance types, fuel compositions, and regulatory environments [22]. A single emission factor assigned to a broad sector in an inventory may therefore fail to represent the actual source mixture in a particular city or neighborhood. When those values are inaccurate or not locally representative, the resulting inventories inherit the bias—not only in the total emissions they report, but also in how they distribute emissions across sectors and across space [23].

This dissertation asks whether dense urban atmospheric observations can provide robust, observation-based constraints on transportation and residential emission factors for NO_x and CO , and whether those constraints can be extended consistently across cities with different climates, vehicle fleets, and residential fuel mixes. It addresses that question by combining multispecies enhancement-ratio analysis with dense sensor network observations, while also confronting a critical prerequisite for quantitative NO_x inference: preserving the fidelity of network NO_2 measurements through physically constrained calibration. The sections that follow provide the observational, analytical, and methodological background needed to motivate and contextualize this work.

1.2 Observational Approaches to Constraining Urban Emissions

Over the past decade, several observational approaches have emerged to improve fine-scale mapping of urban air pollution and greenhouse gases beyond what conventional regulatory monitoring provides [24, 25, 26, 27]. Satellite remote sensing now offers broad spatial coverage of column-integrated NO_2 and CO_2 , but remains limited in temporal resolution, in sensitivity to near-surface sources, and in its ability to resolve intra-urban gradients at the neighborhood scale [28, 29, 30, 7]. Mobile monitoring platforms—including dedicated laboratory vehicles and fleet-based sensor deployments—can achieve very high spatial resolution along driven routes, but their temporal sampling is sparse and irregular, making it difficult to capture the diurnal, weekly, and seasonal variability needed to distinguish source sectors [31, 32, 33]. Dense stationary sensor networks take a different approach: they deploy tens to hundreds of fixed, low-cost measurement nodes at roughly 1–3 km spacing across a city, providing continuous multispecies observations that resolve both persistent spatial gradients and rapid temporal fluctuations [34, 35, 36, 37]. Each platform addresses a different part of the observational gap, but dense stationary networks are especially well suited to the analy-

ses in this dissertation because they provide the combination of neighborhood-scale spatial resolution and continuous temporal coverage needed to track distance-dependent emission gradients, seasonal changes in source composition, and contrasts between transportation and residential sectors over extended periods.

The analytical power of dense networks becomes especially clear when their observations are interpreted through enhancement-ratio analysis. Pollutants such as NO_x and CO are co-emitted with CO_2 during fossil fuel combustion, and the ratio of their atmospheric enhancements— $\Delta X/\Delta\text{CO}_2$, where X is NO_x or CO—is approximately conserved during atmospheric transport and dilution because background subtraction and mixing affect the numerator and denominator similarly [38, 39, 40, 41]. This ratio therefore serves as an observational proxy for the effective emission factor of the local source mixture, linking ambient concentration measurements directly to emission characteristics without requiring a full atmospheric transport inversion. Previous studies have used enhancement ratios at individual regulatory monitoring stations to track long-term trends in fleet-average on-road emission factors [39, 40], at near-road sites and along roadway transects to separate contributions from different vehicle classes [16, 17], and from eddy-covariance flux towers to evaluate city-scale inventory totals [42]. However, most of these applications sample a limited range of source types at one location and cannot resolve how enhancement ratios vary spatially across the roadway and residential sectors within an urban area. Combining enhancement-ratio analysis with a spatially distributed dense network makes it possible to distinguish sector-specific emission signatures—separating the near-highway transportation signal from the distributed residential background—rather than inferring only a single citywide average.

1.3 Calibration as a Scientific Prerequisite

The scientific value of dense sensor networks depends on calibration quality, and this dependence is especially acute for reactive trace gases measured by low-cost electrochemical sensors. These instruments are subject to cross-interference from co-occurring species, temperature-dependent sensitivity, and long-term drift [43]. Among the species considered in this dissertation, NO_2 presents the greatest calibration challenge: the Alphasense NO_2 -B43F electrochemical cell is sensitive to interference from O_3 and NO, exhibits nonlinear temperature-dependent gain, and can drift over deployment periods of months to years. At the same time, NO_2 is one of the most important species for this work, because the NO_x/CO_2 enhancement ratio is a central diagnostic of source composition and combustion conditions. If the calibrated NO_2 signal does not faithfully represent the true atmospheric variability, the enhancement-ratio analyses that underpin this dissertation are directly compromised.

Previous work has advanced electrochemical sensor calibration from colocated multivariate linear regression, in which each sensor is trained against a reference instrument at the same site, toward remote-transfer and network-coherence-based approaches that extend calibration to nodes distributed across a city [44, 45]. These methods have enabled long-term deployment of dense networks even when most nodes cannot be permanently colocated with

reference monitors. The Berkeley Environmental Air-quality and CO₂ Network (BEACO₂N) calibration framework, for instance, achieves $R^2 = 0.66$ and RMSE = 3.16 ppb for remotely calibrated NO₂ over a full year using network-coherence-based training-hour selection and rolling multivariate regression [45]. Multi-year evaluations further show that the Alphasense electrochemical sensors maintain stable performance over deployments of three years or more, indicating that the remaining calibration errors reflect limitations of the calibration model rather than rapid sensor degradation [46]. An important remaining limitation is that unresolved nonlinear gain behavior in the NO₂ sensor compresses the dynamic range of the calibrated signal: a single linear sensitivity coefficient fit across the full range of operating temperatures necessarily represents only an average response and systematically attenuates both positive and negative departures from the mean [43]. This is not simply a technical inconvenience. The analyses in this dissertation depend not only on recovering the correct mean NO₂ concentration, but on preserving the real spatial and temporal variability in NO_x needed for enhancement-ratio-based emission-factor inference.

The atmosphere itself can provide physically meaningful calibration information that statistical regression alone cannot recover. After sunset, the gas-phase reaction of NO with O₃ drives a near-1:1 stoichiometric exchange in which each molecule of O₃ consumed produces approximately one molecule of NO₂ [47]. If the calibrated NO₂ signal carries a multiplicative gain error, the observed ratio $\Delta\text{NO}_2/(-\Delta\text{O}_3)$ during these nocturnal titration events will depart systematically from unity, directly diagnosing the temperature-dependent gain bias that a linear model leaves unresolved. Earlier work showed that titration chemistry can provide useful calibration constraints for BEACO₂N sensors through zero-offset estimation [44]. This dissertation extends that principle from offset estimation to gain diagnosis and uses the titration-derived gain curve as the foundation of a physically constrained calibration framework developed in Chapter 3. That framework strengthens the quantitative basis for the emission-factor analyses in the surrounding chapters.

1.4 From Bay Area Inference to Cross-City Evaluation

The San Francisco Bay Area provides the initial development site for the observational emission-factor framework in this dissertation. The BEACO₂N network in the Bay Area is the most mature deployment, with approximately 50 sensor nodes operating since 2012 at roughly 2 km spacing [35, 48]. The region features strong spatial gradients between major highway corridors and adjacent residential neighborhoods, a residential heating sector dominated by natural gas, and a moderate maritime climate. Prior studies using this network have quantified urban CO₂ emissions through inverse modeling [49, 50], tracked multi-year emission trends, and detected CO₂ reductions associated with vehicle electrification [51, 52, 53]. These characteristics make the Bay Area an ideal setting in which to develop and demonstrate an enhancement-ratio-versus-distance framework for separating transportation

and residential emission signatures.

A single-city result, however well characterized, cannot establish whether inferred discrepancies with inventories are locally contingent or reflect more general features of how inventories parameterize distributed combustion sources. Extending the analysis to Los Angeles and Providence, Rhode Island, introduces experimental contrasts that test both the portability of the framework and the generality of its findings. Los Angeles shares the Bay Area’s natural-gas-dominated residential fuel mix but differs markedly in urban form, traffic patterns, and network geometry. Providence introduces a different residential fuel composition—with appreciable fractions of fuel oil and liquefied petroleum gas in addition to natural gas—together with a colder winter climate and stronger heating demand. These contrasts make it possible to distinguish results that are reproducible across gas-dominated cities from those that respond to fuel mix, climate, or deployment configuration. The physically constrained NO_2 calibration for non-colocated sites developed in Chapter 3 is what makes internally consistent NO_x comparisons across all three cities possible.

Comparison with gridded emission inventories is a unifying theme across the dissertation, but the goal is not simply to ask whether inventories are biased high or low in aggregate. A deeper question is whether they correctly represent the source partitioning and local source mixtures resolved by the dense network, as well as the processes that govern how emissions respond to changing societal and environmental conditions. The two species examined here provide complementary leverage in this evaluation: NO_x enhancement ratios are governed primarily by combustion temperature and fuel type, making them especially sensitive to whether the inventory represents local source composition correctly, whereas CO enhancement ratios depend more strongly on combustion completeness and on whether CO-producing sources are allocated to the same sector sampled by the observations. Together, these tracers help distinguish composition-related inventory mismatches from allocation-related ones, an interpretive framework developed fully in Chapter 4.

1.5 Aims and Chapter Overview

This study has three linked aims:

1. Develop and apply an observational framework based on multispecies enhancement ratios and spatial gradients to infer transportation and residential NO_x and CO emission factors from dense urban sensor network observations.
2. Develop a physically constrained NO_2 calibration approach for electrochemical sensors in dense urban networks that resolves nonlinear temperature-dependent gain errors and preserves the atmospheric variability required for quantitative emission-factor inference.

3. Test the generalizability of the observational framework across multiple cities and evaluate how inferred source-specific emission factors compare with current gridded emission inventories.

Chapter 2 develops the enhancement-ratio-versus-distance framework in the San Francisco Bay Area and uses it to infer sector-specific NO_x and CO emission factors for on-road transportation and residential heating. Comparison with the CARB and GRA²PES inventories reveals that transportation emission factors are broadly consistent with inventory values, whereas the residential NO_x emission factor is a factor of four lower than current inventory estimates. That result raises the question of whether the discrepancy is specific to the Bay Area or reflects a more general feature of residential-emission parameterization. Chapter 3 addresses the calibration prerequisite underlying reliable cross-city NO_x comparison by developing a physically constrained framework that preserves NO_2 dynamic range across deployments without requiring permanent colocation with reference instruments. Chapter 4 extends the emission-factor analysis to three cities—the Bay Area, Los Angeles, and Providence—using gain-corrected observations from Chapter 3. It tests whether the observational framework is portable across different urban settings and uses the dual NO_x/CO comparison to distinguish composition-related from allocation-related mismatches between observed and inventory-based emission factors.

Together, these chapters demonstrate how dense urban observing networks can move beyond qualitative pollution mapping toward quantitative, city-resolved evaluation of transportation and residential emissions, inventory structure, and policy-relevant changes in urban combustion sources.

Chapter 2

Observational Inferences of NO_x and CO Emission Factors for Vehicles and Homes in the San Francisco Bay Area

The chapter was adapted from: Y. Zhu et al. “Observational Inferences of NO_x and CO Emission Factors for Vehicles and Homes in the San Francisco Bay Area”. ACS ES&T Air 2.8 (Aug. 2025), pp. 1478–1487. issn: 2837-1402, 2837-1402. doi: 10.1021/acsestair.5c00004. url: <https://pubs.acs.org/doi/10.1021/acsestair.5c00004>

2.1 Introduction

The substantial urban share of air pollutants and greenhouse gases (GHGs) emissions originating from fossil fuel combustion is expected to undergo rapid change in the coming decades [1, 2, 3, 4]. Cities are leading efforts to mitigate air pollution and prepare for climate change by implementing policies that promote a transition to lower emissions of air pollutants and GHGs [5, 6, 7, 54]. Emissions from urban sectors such as transportation, building energy, and industrial processes vary drastically across space and time due to the complex distribution of sources and associated activity levels within cities. Accurate quantification of these emissions and tracking their changes in response to specific policies enable the evaluation of policy effectiveness. Such observational assessments are essential to verify that anticipated reductions in GHG emissions, along with co-benefits for public health and reduced disparities in air pollution exposure, are achieved as intended.

Urban emission inventories for air quality and GHGs in the U.S. have historically proceeded along separate paths, with few exceptions, such as the recent release of GRA²PES [13]. Air pollutant inventories are generally produced by national, state, or local governments to guide air quality improvements [8]. For instance, the U.S. Environmental Protection Agency National Emission Inventory (NEI) is provided at the county level with updates every three years since 1970. In contrast, city-scale GHG inventories are typically not supported by na-

tional agencies and are primarily developed by the research community. Inventories such as VULCAN (The Vulcan Project, North Arizona University), HESTIA (The Hestia Project, North Arizona University), ACES (Anthropogenic Carbon Emissions System, North American Carbon Program), and GRA²PES (The GReenhouse gas And Air Pollutant Emissions System, National Institute of Standards and Technology) have recently emerged, offering annual CO₂ estimates at 1–10 km resolution [9, 10, 11, 12]. Efforts to build emission inventory estimates are commonly divided into two major approaches: activity-based and observation-based [14]. Activity-based approaches estimate emissions by multiplying emission factors (EFs) by corresponding activity data. EFs, representing the amount of pollutant released per unit of activity (e.g., g CO₂ per mile traveled or per therm of gas consumed), are thus foundational to these accounting methods. Many existing inventories are activity-based, and one key limitation is that the traceability of emissions to observational data differs. While emissions from large point sources can often be tied to direct measurements and remain up to date, emission estimations on mobile and other distributed sources rely on EFs and activity proxies that may be changing over time and thus require extrapolation from the date of inventory preparation to estimate current emissions. In contrast, observation-based approaches use ambient atmospheric measurements to infer emissions directly.

Atmospheric concentration measurements provide an objective and consistent evaluation of inventories when linked to emissions through an appropriate model. Nevertheless, discrepancies among emission inventories persist [18, 19]. For example, there has been extensive discussion about the bias in inventory estimates for on-road traffic emissions [15, 16, 17, 23]. These discrepancies are particularly notable in cities, where emissions and concentration profiles vary significantly on fine spatial scales and where understanding the impacts of urban emissions on public health is especially important because of high population density [55, 56, 57, 58].

Over the past decade, several innovative techniques have emerged to improve fine-scale mapping of air pollution and GHGs. These include: (i) dense stationary networks, (ii) mobile monitoring, and (iii) satellite remote sensing [24, 25, 26, 27, 31, 32, 30, 7, 28, 29]. Among these new techniques, dense stationary networks offer the distinct advantage of providing multi-species continuous temporal information while simultaneously capturing spatial patterns at high resolution [34, 35, 36, 37, 33]. This capability is particularly valuable for tracking temporal variations across daily, seasonal, and multi-year timescales with neighborhood-level details. By combining measurements of multiple chemical species, observational constraints on sectoral emissions can be developed based on the characteristic EFs of different sources [59, 42]. For example, NO_x EFs (defined as ratios to CO₂) from heavy-duty trucks are much higher than NO_x EFs from passenger cars due to differences in combustion conditions and emission control technologies (see Table 1) [22].

Here, we use observations from the Berkeley Environmental Air Quality and CO₂ Network (BEACO₂N), a ~ 2 km-spaced multipollutant monitoring network, to map the distribution of observed enhancement ratios (ERs, i.e., $\Delta\text{NO}_x/\Delta\text{CO}_2$ and $\Delta\text{CO}/\Delta\text{CO}_2$) in the San Francisco (SF) Bay Area and examine the spatial and seasonal patterns of these ratios. We organize the observed ERs by distance from highways (as done previously to assess transportation

CO₂ emission rates [51, 52]), which allows us to derive observational estimates for NO_x and CO EFs from vehicles and residential homes in the region.

2.2 Materials and Methods

In Section 2.1, we discuss the BEACO₂N measurements of NO, NO₂, CO, and CO₂. In Section 2.2, we describe the calculation of ERs based on these measurements. In Section 2.3, we provide a brief overview of the ancillary datasets used in this analysis, including vehicle miles traveled (VMT), natural gas usage, wind data, and emission inventories.

2.2.1 BEACO₂N Measurements

Hourly-averaged BEACO₂N measurements from approximately 40 locations in the SF Bay Area that were active for most of 2022 were used in this analysis. NO, NO₂, and CO were measured by Alphasense electrochemical gas sensors: NO-B4, NO₂-B43F, and CO-B4. CO₂ was measured by a non-dispersive infrared CO₂ sensor: Vaisala CarboCap GMP 343. Figure 2.1 displays a map of BEACO₂N site locations and measurement time series. Approximately 75% of the BEACO₂N sensor nodes are mounted on one-story school rooftops, with inlet heights around 4 meters above ground level. The remaining sites are located on multi-story buildings, such as museums or municipal buildings, with inlet heights ranging from approximately 12 to 16 meters. The data are transmitted in real time through Wi-Fi or collected manually using a cellular hotspot and are publicly available (<https://beacon.berkeley.edu/>). Winter et al. provides details on the calibration and performance evaluation of Alphasense sensors in the field [45]. For NO, NO₂, and CO sensors, uniform periods in concentrations across the network were leveraged to train multivariate linear regression models for calibration at each site. The calibration models included terms for the target analyte, cross sensitivities to interfering species, and zero-offset corrections with respect to temperature and relative humidity. Measurement accuracy was estimated at 1.1 ppbv, 2.4 ppbv, 20.2 ppbv, and 1 ppmv for NO, NO₂, CO, and CO₂, respectively [45, 60]. Measurement precision was 1.6 ppbv for NO, 3.2 ppbv for NO₂, 33.3 ppbv for CO, and 1.6 ppmv for CO₂, based on a 1-hour averaging period. Further details of BEACO₂N instrumentation, descriptions of CO₂ and aerosol particle sensor calibration, and quality assurance (QA)/quality control (QC) methods are available in previous publications [35, 48, 61].

2.2.2 Enhancement Ratio Calculation

Enhancement ratios of co-emitted species are conserved at spatial and temporal scales in urban settings and are independent of atmospheric dilution into background air, which makes them robust quantities linking dense sensor network data to emission information [38, 39, 40, 41]. The slope of the best-fit line between ΔCO_2 and its co-emitted species ΔX represents

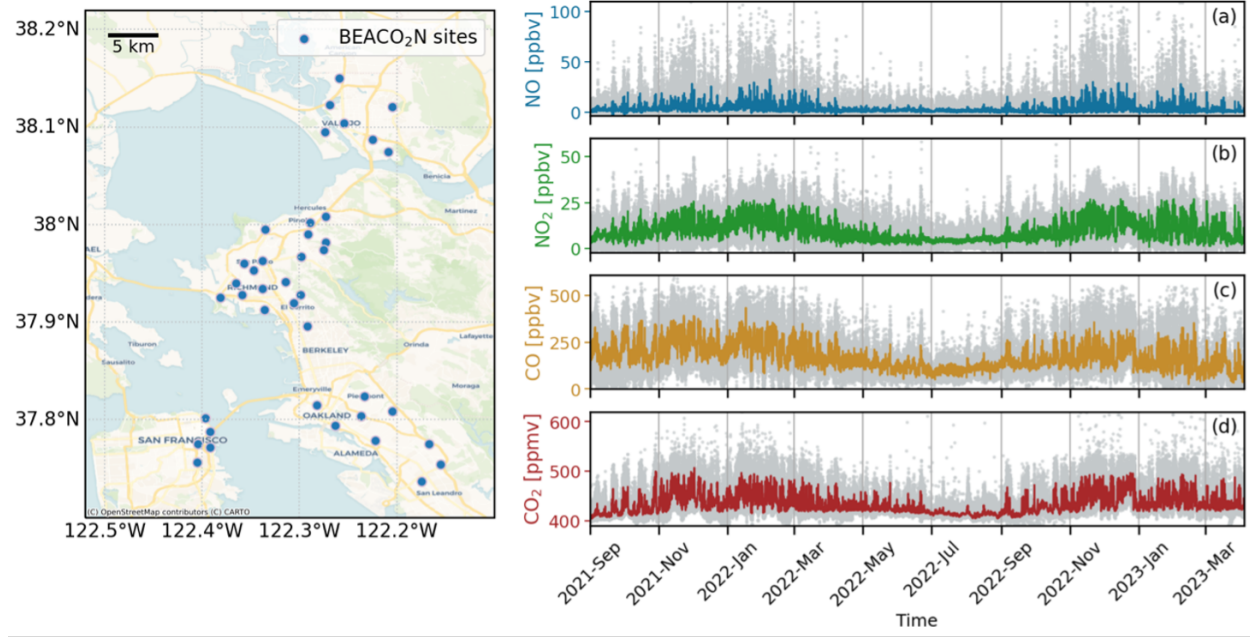


Figure 2.1: The location of BEACO₂N sites (left panel) in the SF Bay Area and time series (right panel) of (a) NO, (b) NO₂, (c) CO, and (d) CO₂ measurements from September 2021 to April 2023. All available observations in the BEACO₂N network are shown as gray dots, while median values of the network for 1-hour time bins are shown by the solid lines. Background map credits: © OpenStreetMap contributors (© CARTO (<https://www.openstreetmap.org/copyright>)) contributors; © CARTO (<https://carto.com>).

the enhancement ratio at each location. This observed ratio essentially reflects the ER of species X to CO₂, integrated across all sectors and as detected by the site (Equation 2.1), where X denotes either NO _{x} or CO.

$$\frac{\Delta X}{\Delta \text{CO}_2} = \frac{X_{\text{measured}} - X_{\text{background}}}{\text{CO}_{2,\text{measured}} - \text{CO}_{2,\text{background}}} = ER_{X,\text{integrated}} \quad (2.1)$$

The measurement uncertainties for CO₂ are comparable to those of CO and NO _{x} , necessitating the use of bivariate least-squares linear regression analysis, which considers errors in both variables, to determine the ERs of NO _{x} and CO with respect to CO₂ (i.e., $\Delta \text{NO}_x / \Delta \text{CO}_2$ and $\Delta \text{CO} / \Delta \text{CO}_2$) [62, 63, 64, 65]. Data for enhancement ratio calculations were chosen to be within selected ranges of wind direction and time of day. A modified interquartile range method was adopted to exclude time periods reflecting exceptional events (e.g., wildfires, cross-validated against CAL FIRE incident reports, <https://www.fire.ca.gov/incidents>, accessed 2024-10-06) [66], resulting in the removal of less than 0.5% of the data. Concentration enhancements are defined as the differences between measured concentrations and background concentrations. Background concentrations were estimated using the lowest 5th

percentile of the network measurements for each hour within a 4-week moving window.

The influence of background choice on ER fitting has been discussed widely in the literature. The lowest 5th percentile method used here has been shown to work well in urban areas in prior urban studies and is well-suited for dense networks like BEACO₂N, where a subset of sites typically samples upwind or cleaner air [41, 67, 68, 69, 42]. Enhancement ratio pairs with squared Pearson correlation coefficient (r^2) values below 0.5 were excluded (i.e., 5 out of 25 sites reporting NO_x ratios and 12 of 33 sites reporting CO ratios were not included). This threshold was consistent with previous urban emission studies and was set to ensure that the results reflect strong co-variation between CO₂ and NO_x or CO enhancements [41, 70]. Sites with lower r^2 values were typically influenced by non-synchronous emission sources that introduce independent variability and weaken the collinearity between CO₂ and NO_x or CO concentrations. These sources include industrial activities (e.g., refinery operations, power generation, product manufacturing), commercial activities (e.g., restaurant cooking), biospheric fluxes (e.g., biogenic CO₂ uptake), and off-road sources (e.g., construction equipment, lawn and garden equipment). Although a comprehensive understanding of all non-synchronous emission sources has not yet been achieved, four of the sites were excluded due to their proximity to the largest point source in the region, the Richmond Chevron Refinery, which accounts for nearly one third of the total CO₂ emissions in the study region and exhibits a distinct emission profile [71, 50]

2.2.3 Ancillary Datasets

The Caltrans Performance Measurement System (PeMS) monitors traffic on major highways throughout California, with approximately 580 traffic counting stations located within the study area. These stations count vehicle flow using magnetic loops embedded in roadways and estimate the truck proportion based on vehicle speed and length [72]. After filtering for data with $\geq 99\%$ observed and $\geq 95\%$ time coverage during the study period, 358 PeMS sites remained. For each traffic monitor, hourly data on vehicle flow, VMT, and average fleet speed from January to December 2022 were used.

The Pacific Gas and Electric Company (PG&E) supplies natural gas to the SF Bay Area and publishes gas usage data aggregated by ZIP code and by month, on a quarterly basis (available at https://pge-energydatarequest.com/public_datasets, accessed 2024-10-06). Monthly total natural gas usage trends for 2022 were obtained.

Wind data were sourced from MesoWest, a cooperative weather database managed by the University of Utah, which provides spatial coverage overlapping with BEACO₂N [73]. The distance to the nearest highway for each site, considering wind directions, was determined for the hourly measurements with the assumption that wind speed (WS) variability within the observed range does not affect ER-based EF analysis. A sensitivity test (detailed in Text S2–S3 and Figure S3) showed that varying WS from 0.5 to 5 m s^{−1} changed the emission mixing length but did not alter key features of the ER profile (e.g., values at $d = 0$ and $d > 3$ km), supporting the robustness of the EF inference. Wind data were obtained from 44 MesoWest stations across the study region, reported at 15-minute intervals and averaged to

hourly resolution. These stations overlap well spatially with BEACO₂N sensor locations (see Figure S1), allowing site-level wind matching. To assess representativeness, we compared MesoWest wind speed distributions with HRRR reanalysis winds (hourly, 3-km grid) in summer (June–September) and winter (December–February). After adjusting HRRR winds to 5 m height, we found consistent modal WS values between MesoWest and HRRR (1.8 m s⁻¹ in summer, 1.0 m s⁻¹ in winter; Figure S2).

Sectoral EFs from two inventories are discussed below: the criteria pollutant inventory (<https://ww2.arb.ca.gov/criteria-pollutant-emission-inventory-data>, accessed 2024-10-06) and the GHG emission inventory (<https://ww2.arb.ca.gov/ghg-inventory-data>, accessed 2024-10-06) compiled by the California Air Resources Board (CARB), and the U.S. GHG and Air Pollutant Emissions System (GRA²PES, <https://data.nist.gov/od/id/mds2-3520>, accessed 2024-10-25). Statewide annual emission estimates for the transportation and residential sectors in 2022 were retrieved to calculate EFs in the CARB inventories. The GRA²PES inventory provides hourly emission information, including on-road vehicle emissions based on 2021 taxable fuel sales, and scales the residential sector from the 2017 NEI at a 4-km resolution.

2.3 Results and Discussion

Our analysis begins with a model describing the mixing of emissions from a line source (e.g., a highway) into a uniform field of emissions (e.g., residential housing). Transportation CO₂ emission trends in the SF Bay Area have previously been assessed by BEACO₂N observations coupled with multiple linear regression, Gaussian plume modeling, or inverse modeling techniques [51, 52, 50]. The BEACO₂N inference of CO₂ emission reductions due to improved vehicle fuel economy during 2018–2022 aligned with estimates from the California state emission model and electric vehicle sales. In Section 3.1, we present a numerical advection–diffusion model to illustrate how ER profiles depend on the EFs of the two sectors, the mixing length scale, and the relative strength of the two emission sources. In Section 3.2, we describe the observations organized by distance from highways, and in Section 3.3 we compare the BEACO₂N-derived EFs to the CARB and GRA²PES inventories.

2.3.1 Idealized Mixing of Transportation and Residential Emissions

We developed a simplified three-dimensional numerical advection–diffusion model, describing a 10 km-long highway line source located at the edge of the domain as it mixes into an adjacent 10 km × 10 km residential neighborhood downwind. The wind direction was set to be perpendicular to the highway, with crosswind and vertical wind velocities set to be zero. Many BEACO₂N sites are located within 5 km downwind of a nearby highway, with wind directions selected for the corresponding time periods (see Figure 2.1). The forward Euler scheme (first order in time, second order in space) was used to discretize the advection–

diffusion equation [74]. The vertical dispersion parameter was adjusted for surface roughness using the Hosker Index for urban terrain [75, 76, 77]. Other wind and dispersion parameters were chosen to match the e-folding length scales for NO_x concentration decay from highways, approximately 300–500 meters in urban settings [78, 79]. Details on model configurations and sensitivity tests are provided in the Supplementary.

We used $\text{EF}_{X,\text{transportation}} = 1.2 \text{ ppbv/ppmv}$ and $\text{EF}_{X,\text{residential}} = 0.2 \text{ ppbv/ppmv}$ for the transportation and residential sectors to simulate the distribution of ER_X as a function of distance from the highway. These EF values were consistent with the observations shown in Section 3.2. Emission rates of the line and area sources were estimated based on the posterior CO_2 emissions from Asimow et al.: $0.15 \text{ tC hr}^{-1} \text{ km}^{-2}$ from building natural gas usage (mostly heating and hot water) and $0.82 \text{ tC hr}^{-1} \text{ km}^{-1}$ from highways in summer (June–September); $0.34 \text{ tC hr}^{-1} \text{ km}^{-2}$ from buildings and $0.68 \text{ tC hr}^{-1} \text{ km}^{-1}$ from highways in winter (December–February) [50]. The relative emission intensities in the $10 \text{ km} \times 10 \text{ km}$ model domain were 2:1 in summer and 6:1 in winter for building versus highway emissions.

We modeled the two scenarios representing summer (2:1 emission intensity) and winter (6:1 emission intensity) conditions and added two additional extreme cases: one with no building emissions (0:1 emission intensity) and another with very high building emissions (25:1 emission intensity), as illustrated in Figure 2.2. When there are no area source emissions, the modeled ER remains constant at the value of transportation $\text{EF}_{X,\text{transportation}} = 1.2 \text{ ppbv/ppmv}$ regardless of proximity to the highway. As the area source contribution increases, the deviation of $\text{ER}_{X,\text{integrated}}$ from $\text{EF}_{X,\text{transportation}}$ at the origin grows. If the chemical species X has a 5-hour lifetime, which is approximately the loss rate for NO_x at midday in summer, then a slow decrease reaching about 10% occurs over 8 km. A 5-hour NO_x lifetime was used in the model simulations, and we evaluated sensitivity to chemical reactivity by simulating lifetimes ranging from 10 hours to 0.2 hours, as shown in Figure S4.

The small difference ($\sim 10\%$) between the starting points of the 0:1 and 2:1 area-source-to-line-source scenarios, as well as the convergence of the ERs after 3 km for all three scenarios with area source contributions, suggested that transportation $\text{EF}_{X,\text{transportation}}$ and residential $\text{EF}_{X,\text{residential}}$ can be estimated from the integrated $\text{ER}_{X,\text{integrated}}$ at $d = 0 \text{ km}$ and $d > 3 \text{ km}$, respectively. The simulated profiles under summer and winter conditions both followed an exponential decay with distance from the highway, as described by

$$\text{ER}_{X,\text{integrated}}(d) = 0.92 \exp(-d/0.87 \text{ km}) + 0.19$$

for summer, and

$$\text{ER}_{X,\text{integrated}}(d) = 0.73 \exp(-d/0.56 \text{ km}) + 0.20$$

for winter, where d is the distance from highways, the offset at large distance reflects the EF of the residential sector and the y -intercept represents a lower bound to the highway EF. While this ER-based EF estimation method is applicable in both summer and winter, optimal separation of source signals can be achieved by using summer morning ERs near the highway ($d = 0 \text{ km}$) to infer transportation-related emissions. Far-field ($d > 3 \text{ km}$) residential EFs

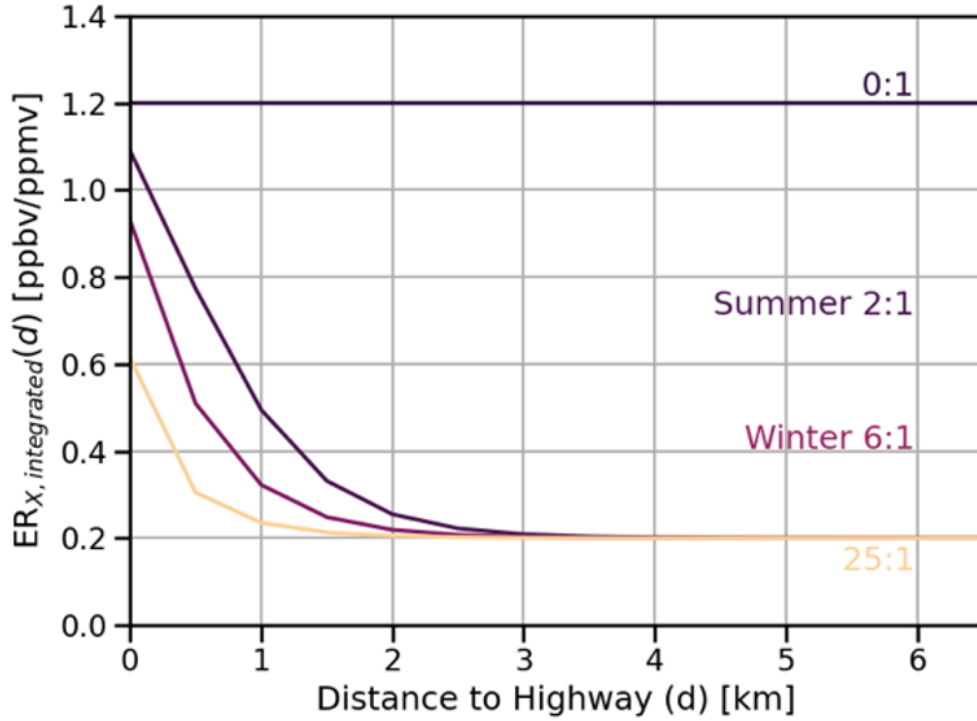


Figure 2.2: The simulated mixing of area source (residential heating) and line source (transportation) within a $10 \text{ km} \times 10 \text{ km}$ modeling domain. The spatial gradient profiles of enhancement ratios under different relative emission source strengths (area source: line source) are labeled by different colors.

are consistent across all times of day and both seasons studied. The e-folding length scale (k) of the winter scenario ($k = 0.56 \text{ km}$) was shorter than that of the summer scenario fitting ($k = 0.87 \text{ km}$), reflecting the greater influence of the area source in winter.

2.3.2 Observed Emission Ratios versus Distance to Highway

We generate observational analogs of Figure 2 by calculating the distance to the nearest highway along the wind vector for each hour and plotting the observed ERs as a function of these distances. Although changes in wind speed can affect the absolute concentration magnitudes measured at a site and the mixing length scale, the enhancement ratios derived from relative differences in concentrations remain little affected for the range of wind speeds in the study region (see Figure S3). Also, our analysis based on ERs between co-emitted species is largely insensitive to planetary boundary layer height (PBLH) variations since both species undergo the same mixing.

We compared observed ERs during peak and low traffic hours (6:00–9:00 AM and 0:00–

3:00 AM) as well as during peak and low natural gas usage months (December–February and June–September of 2022). Monthly VMT measured by Caltrans PeMS and natural gas consumption data reported by PG&E for 2021–2023 are shown in Figure S5. During peak morning traffic, vehicle flow was approximately 10 times higher than during late night hours, and natural gas usage in winter months was over three times higher than in summer months. These patterns are consistent with previous studies describing the diurnal and seasonal variation of transportation and natural gas use in California, and are similar to the CO₂ emission ratios for transportation and building energy reported by Asimow et al [50, 80, 81, 82].

Figure 2.3 displays the spatial gradients and distance-decay relationship of observed ERs during morning rush hours and late night, in both summer (June–September) and winter (December–February). Each point represents the ER calculated at a single BEACO₂N site for a given season, time window, and wind direction (120°–260°) to ensure comparability of downwind transport from major highways. Although the BEACO₂N network includes 40 active sites, only data passing three filters were used in the analysis: (1) data completeness > 80% in 2022 (25 NO_x sites, 33 CO sites); (2) regression $r^2 > 0.5$ (yielding 20 NO_x and 21 CO sites); and (3) location constraints—one site upwind and two sites > 7 km from highways were not shown in the figure but included in distance-decay fitting. Site-level inclusion in the distance-decay analysis was based on data completeness, regression quality, and proximity to highways; full site status is documented in Table S2.

During summer morning rush hours, marked by low natural gas usage and peak traffic volumes, elevated NO_x and CO ERs were observed at the three sites closest to highways (Figure 2.3a and 2.3c), along with near-constant lower values observed beyond 3 km from the highways. The e-folding length obtained from exponential decay fitting represents the distance at which the ER drops to $1/e$ ($\approx 36.7\%$) of its initial value. To estimate the characteristic mixing length, the distance required for the ER to decay fully to the far-field offset, we divide the e-folding length by 0.36. These mixing length scales, calculated as $k/0.36$, provide a physically interpretable measure of the spatial extent over which highway emissions blend well with building emissions. The e-folding distances k for the NO_x and CO ER fittings were 0.95 km and 0.66 km, which indicated that transportation NO_x and CO emissions mixed with other urban emissions over distances of 2.6 km and 1.8 km, respectively. At $d = 0$ km, where transportation emissions dominate, we estimated that

$$\text{EF}_{\text{NO}_x, \text{transportation}} \approx \text{ER}_{\text{NO}_x, \text{integrated}}(d = 0 \text{ km}) = 1.0 \text{ ppbv/ppmv}$$

and

$$\text{EF}_{\text{CO}, \text{transportation}} \approx \text{ER}_{\text{CO}, \text{integrated}}(d = 0 \text{ km}) = 7.8 \text{ ppbv/ppmv},$$

recognizing that the model shown above indicates these values are lower-bound limits. At long distances $d > 3$ km, where building energy use dominates, the well-mixed ERs were 0.21 ppbv/ppmv for NO_x and 4.1 ppbv/ppmv for CO. These findings align with expectations that $\text{EF}_{\text{NO}_x, \text{transportation}}$ and $\text{EF}_{\text{CO}, \text{transportation}}$ are higher than EFs for the building energy sector [20].

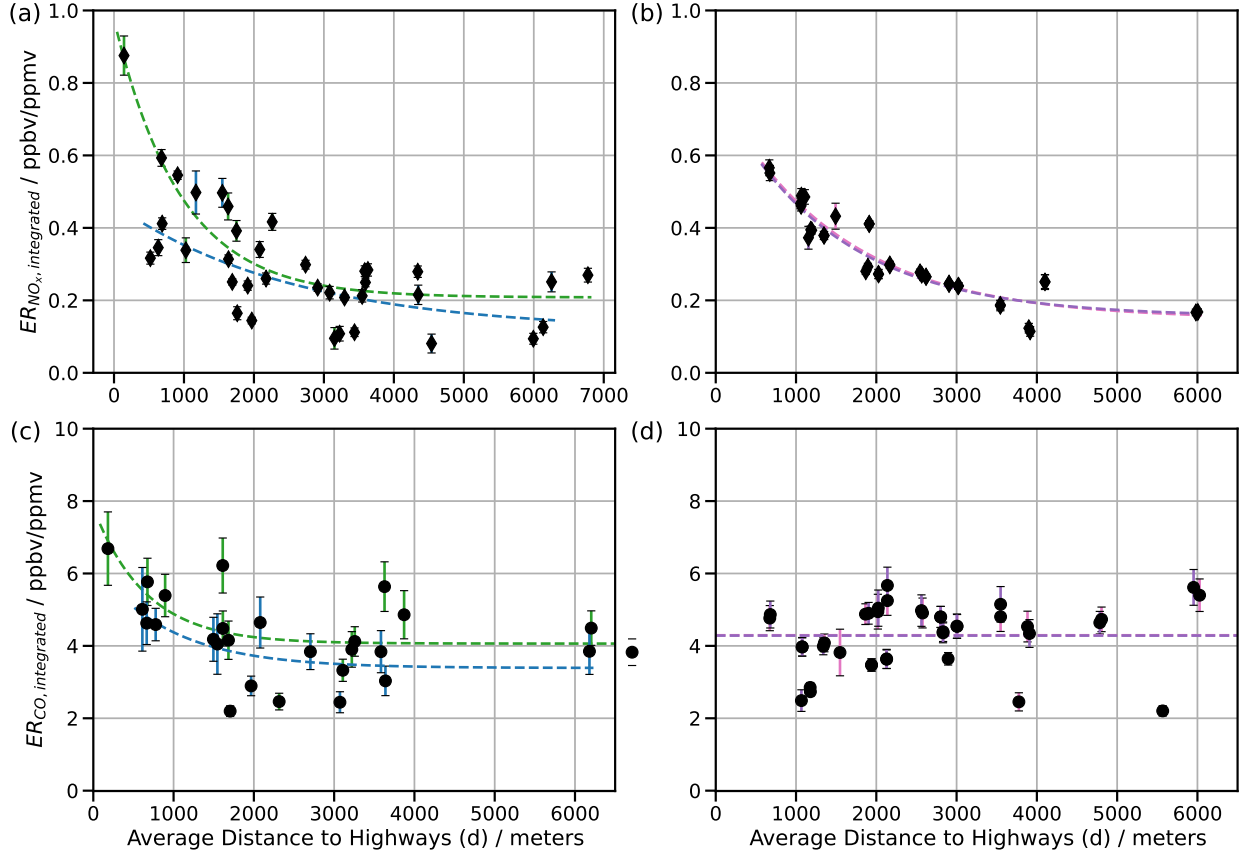


Figure 2.3: BEACO₂N observed $\Delta NO_x / \Delta CO_2$ and $\Delta CO / \Delta CO_2$ in 2022 versus distance to highways. Panels (a) and (c) show the summer values (June–September), and panels (b) and (d) display winter values (January, February, and December). Morning (6:00–9:00 AM local time) data and distance-decay fitting lines are represented in green for $ER_{NO_x, integrated}(d)$ and pink for $ER_{CO, integrated}(d)$. Midnight (0:00–3:00 AM local time) data and distance-decay fitting lines are shown in blue for $ER_{NO_x, integrated}(d)$ and purple for $ER_{CO, integrated}(d)$. Only data corresponding to wind directions between 120° and 260° within each time window were included. Uncertainties of the ER estimates are represented as error bars. Parameters of the exponential fitting for each line are provided in Table 1.

At night during summer, the $ER_{NO_x, \text{integrated}}$ at $d = 0$ km dropped by about 50%, and $ER_{CO, \text{integrated}}$ decreased by 25%, reflecting reduced transportation on the highways and consequently increased effects of building energy emissions on the observed ERs. The smaller fractional reduction in $ER_{CO, \text{integrated}}(d = 0 \text{ km})$ than in $ER_{NO_x, \text{integrated}}(d = 0 \text{ km})$ is consistent with the smaller difference between transportation and residential EFs for CO compared to NO_x . Day–night differences in decay length and offset for summer were not significant.

In winter, while elevated NO_x ERs were still observable, CO ERs became flat with respect to proximity to the highway (Figure 2.3b and 2.3d). This shift is consistent with the threefold increase in building energy use during the winter months. Seasonal differences in observed ERs, driven by changes in residential heating, were also identified in the city of Paris, France [67]. The day–night variations seen in summer data were absent in winter, indicating that residential heating emissions remained more stable and pronounced throughout the day, masking the influence of transportation emissions, as the approximately doubled CO_2 emissions in the denominator dampened the spatial variations [50]. These summer–winter differences highlight how additional residential emissions from heating during winter months influence the spatial gradients of integrated ERs. During winter midnight hours at $d = 6$ km, where residential emissions dominate, we estimated that

$$EF_{NO_x, \text{residential}} \approx ER_{NO_x, \text{integrated}}(d = 6 \text{ km}) = 0.15 \text{ ppbv/ppmv}$$

and

$$EF_{CO, \text{residential}} \approx ER_{CO, \text{integrated}}(d = 6 \text{ km}) = 4.3 \text{ ppbv/ppmv}.$$

The residential EFs estimated at long distances from highways were nearly identical for observations from summer and winter. This consistency supports interpreting the ERs beyond 3 km from the highway as representing the building energy EF. The e-folding length (k), mixing length scales ($k/0.36$), and integrated ERs for NO_x and CO at peak and low traffic volumes, as well as periods of peak and low natural gas usage, are summarized in Table 1.

Table 2.1: Summary of e-folding distances (k), mixing length scales ($k/0.36$), and integrated emission ratios for NO_x and CO at peak and low traffic volumes, and during periods of peak and low natural gas usage.

Time Period	k for ER_{NO_x} (km)	ER_{NO_x} mixing length (km)	k for ER_{CO} (km)	ER_{CO} mixing length (km)	ER_{NO_x} at $d = 0$ km (ppbv/ppmv)	ER_{NO_x} at $d = 6$ km (ppbv/ppmv)	ER_{CO} at $d = 0$ km (ppbv/ppmv)	ER_{CO} at $d = 6$ km (ppbv/ppmv)
Summer 6:00–9:00 am	1.0 ± 0.2	2.6 ± 0.6	0.66 ± 0.5	1.8 ± 1.4	1.0 ± 0.02	0.21 ± 0.02	7.8 ± 0.6	4.1 ± 0.6
Summer 0:00–3:00 am	2.5 ± 2.2	7.0 ± 7.6	0.93 ± 0.9	2.5 ± 2.4	0.51 ± 0.02	0.11 ± 0.02	6.3 ± 0.6	3.1 ± 0.6
Winter 6:00–9:00 am	1.5 ± 0.4	4.1 ± 1.1	n/a	n/a	0.75 ± 0.01	0.15 ± 0.01	4.3 ± 0.3	4.3 ± 0.3
Winter 0:00–3:00 am	1.4 ± 0.4	3.9 ± 1.1	n/a	n/a	0.75 ± 0.01	0.15 ± 0.01	4.3 ± 0.3	4.3 ± 0.3

Although the observed spatial patterns and seasonality of emission ratios in Figure 2.3 generally matched the simulated spatial gradients of $ER_{X, \text{integrated}}$ for area-source-to-line-

source ratios of 2:1 and 6:1 in Figure 2, there was one difference worth noting. The shorter mixing length modeled for winter data was not observed in measurements, though the modeled and observed mixing length scales were comparable (approximately 2–3 km; see Table 2.1). This discrepancy could be explained by faster turbulent dispersion during the measurement periods than the $10^3 \text{ m}^2 \text{ s}^{-1}$ used in the model [83].

2.3.3 BEACO₂N-Derived Emission Factors Compared to CARB and GRA²PES Emission Inventories

A comparison of EFs derived from the BEACO₂N observations to EFs in the CARB and GRA²PES emission inventories is shown in Table 2.2. Our goal in this study is to infer up-to-date EFs for the overall transportation and residential sectors in the SF Bay Area, rather than to resolve detailed EFs for individual subcategories. We acknowledge that transportation EFs can vary with vehicle type, engine load, and driving conditions, and that residential EFs depend on factors such as fuel type, installed equipment, and infrastructure design, which present valuable opportunities for future investigation.

Our transportation CO EF estimate of 7.8 ppbv/ppmv is higher than the inventories by 15% relative to GRA²PES and 18% relative to CARB. The BEACO₂N-derived residential CO EF of 4.3 ppbv/ppmv is close to the GRA²PES value for building energy and is about 33% lower than the CARB value. The transportation NO_x EF estimate of 1.0 ppbv/ppmv is within 8% of both the GRA²PES and CARB values. In contrast, the observed building energy NO_x EF is four times lower than the identical estimates from these inventories. It is worth noting that CARB and GRA²PES values represent statewide averages, which may not fully capture localized differences in installed equipment or regulatory enforcement.

We speculate that the effective flame temperatures assumed in inventories for natural-gas-fired equipment are overestimated. Given the strong temperature dependence of NO_x formation during combustion, even small errors in temperature assumptions can result in large errors in EF estimates [21]. While some emission measurements exist for commercial boilers, most research on natural-gas-fired appliances in residential settings has focused on their impact on indoor air quality [84, 85, 86]. Detailed characterization of residential appliance-level combustion temperatures remains rare. Although overestimated temperatures most likely contribute to this discrepancy, we also note that the Bay Area Air Quality Management District has implemented more stringent regulations on NO_x emissions from natural-gas-fired water heaters, boilers, and process heaters than other air districts in California since 2008 [87]. These local policies may have further contributed to the lower residential NO_x EF observed in our study, while the turnover time of building energy systems is typically expected to be on the scale of decades. Stricter regulations and lower combustion temperatures together could explain the observed discrepancy between BEACO₂N-derived and inventory-based residential NO_x EF.

2.4 Conclusions

The lower NO_x EF observed for the building energy sector, compared to current inventory estimates, underscores the importance of observational efforts in verifying emissions and emission factors. Beyond identifying mismatches between observed and inventory-based EFs, our framework offers a scalable and demonstrative pathway for emissions monitoring and verification. By leveraging spatial gradients in ERs across a dense urban sensor network, we generate emission factor estimates that are both spatially disaggregated and temporally current, filling a critical gap in city-level emissions data. This method is particularly valuable in regions where activity-based inventories lack the resolution or update frequency to track changes in urban infrastructure, policy, or technology. Looking ahead, this approach can be used to assess the real-world impacts of emerging technologies (e.g., electric vehicles, heat pumps) on emissions patterns, as cities accelerate efforts to decarbonize.

Furthermore, while this study primarily focused on network-wise ER profiles, future work examining temporal variations at diurnal, weekly, and seasonal scales at individual sites holds promise for detecting emission signatures from additional sources, such as industrial point sources and biogenic fluxes. Together, these extensions lay a foundation for tracking the spatial and temporal evolution of urban emissions in support of both air quality management and climate action goals.

Table 2.2: BEACO₂N-derived emission factors and sectoral emission factors in current emission inventories.

Emission Factor (ppbv/ppmv)	Inventory	Transportation	Residential
$\Delta\text{NO}_x/\Delta\text{CO}_2$	BEACO ₂ N	1.0 ± 0.02	0.15 ± 0.01
	CARB	1.0	0.60
	GRA ² PES	1.1	0.60
$\Delta\text{CO}/\Delta\text{CO}_2$	BEACO ₂ N	7.8 ± 0.6	4.3 ± 0.3
	CARB	6.6	6.2
	GRA ² PES	6.8	4.0

Chapter 3

Physically Constrained Non-Collocated Calibration of Electrochemical NO₂ Sensors in Dense Urban Networks

3.1 Introduction

Urban emissions of air pollutants and greenhouse gases exhibit strong variability at spatial scales that have historically been difficult to observe directly. Within cities, pollutants such as NO₂ and other reactive tracers can vary by factors of two to five over distances of only a few hundred metres, driven by proximity to traffic corridors, industrial point sources, and the channeling of emissions within the urban canopy [79, 78]. These fine-scale gradients are increasingly recognized as central to understanding air pollution inequities and the ways in which pollution burdens are distributed across communities [88, 89]. Observations that resolve this heterogeneity are therefore needed not only to refine urban emission inventories and quantify spatial and temporal variability in emissions, but also to evaluate disparities in exposure and to assess how local pollution patterns align with land use, infrastructure, and demographic patterns. Conventional observing platforms, including regulatory monitoring networks and satellite remote sensing, provide essential information at regional to national scales, but they are typically too sparse in space or too coarse in time to fully resolve intra-urban variability at the scales most relevant for neighbourhood exposure and source attribution [90, 91].

Dense networks of electrochemical and metal-oxide sensors, deployed at tens to hundreds of fixed locations within a single city, offer a path toward closing this observational gap at manageable capital and labor cost. Operating at high temporal resolution, such networks can capture both persistent spatial gradients and rapid fluctuations associated with traffic pulses, episodic emissions, and meteorological transitions. Compared with ground-based reg-

ulatory networks, they are better suited to characterizing the temporal evolution of pollution within and between neighbourhoods, especially where existing monitoring sites are sparse or unevenly distributed across communities. This capability creates new opportunities for exposure mapping, inventory evaluation, source attribution, and neighbourhood-scale analysis of urban chemistry.

Realizing those applications, however, depends on a scientific prerequisite: the measurements must be sufficiently well calibrated to preserve spatial and temporal variability. These calibration challenges are particularly acute for NO₂. Among the species measured by typical multi-pollutant sensor nodes, NO₂ is both the most susceptible to electrochemical artefacts—in particular, pronounced temperature-dependent gain and cross-interference from co-occurring O₃ and NO—and one of the most important for constraining urban emissions, diagnosing neighbourhood-scale chemistry, and assessing disparities in exposure. By comparison, the O₃, NO, and CO channels are generally more straightforward to calibrate [43].

In situ measurements of NO₂ are central to atmospheric chemistry, and several reference-grade techniques are well established. These methods are briefly summarized here to contrast their strengths and operational constraints with those of electrochemical NO₂ sensors. The most widely used method converts NO₂ to NO and detects the resulting NO through its chemiluminescent reaction with O₃ [92]. Optical absorption methods, including cavity ring-down and cavity-attenuated phase-shift spectroscopy [93, 94, 95], measure NO₂ directly with delicate optical path alignment. Both approaches are typically calibrated at least daily by measuring an instrument zero with NO₂-free air and an instrument response with a known amount of NO₂. For chemiluminescence instruments, biases from non-specific conversion of other reactive nitrogen species, such as organic nitrates (e.g., methyl nitrate) and HNO₃, have motivated the use of more selective photolytic converters in place of thermal converters [96, 97, 98]. These instruments are commercially available from multiple vendors, but require trained personnel, temperature-controlled enclosures, and dedicated sample inlets to reach their nominal performance: typical accuracy of 5–10% with precision around 0.1 ppb min⁻¹ [99]. Capital costs are typically \$10,000–\$100,000 per instrument. Where high time resolution is not required, absorbent cartridges followed by ion chromatography can also map NO₂ on day-to-week timescales [97].

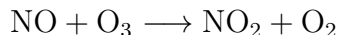
The electrochemical sensors used in dense networks operate on a fundamentally different principle and under a different set of operational constraints. They generate a current associated with redox reactions of the target gas at a working electrode, achieve few-second time resolution, and cost orders of magnitude less than reference instruments, making it possible to deploy many nodes without temperature-controlled enclosures or intensive maintenance. However, deployed electrochemical sensors are rarely accompanied by on-site gas standards, so their zero and analyte response cannot be directly assessed during routine operation. Bringing gas standards to every node would undermine the cost and deployment advantages that motivate their use. As a result, electrochemical sensors are often calibrated initially at a colocated reference site and then moved to remote locations under the assumption that the calibration is stable and transferable across time and local conditions. In practice, this assumption is problematic: zero offsets drift over time, and both baseline and sensitivity have

complex temperature and humidity dependencies that may not be captured during a short colocation and may themselves evolve over the deployment. This open problem—preserving the spatial coverage enabled by electrochemical sensors while maintaining sufficient accuracy and precision for quantitative network analysis—motivates the calibration framework developed in this chapter.

Currently, most of the framing of the calibration problem emphasizes reference distance: standard practice colocates each sensor with a reference instrument and trains a multivariate linear regression, which performs well locally but does not scale naturally to distributed deployments where most nodes lack direct reference supervision. Machine learning methods and other statistical approaches exploiting network-wide properties have been developed to extend calibration to remote nodes, but generalisability across locations remains limited. The BEACO₂N calibration achieved $R^2 = 0.66$ and $\text{RMSE} = 3.16$ ppb for remotely calibrated NO₂ over a full year using network-coherence index selection and rolling multivariate regression, representing the best published result for this network [45]. However, its training-hour selection relies on the relative standard deviation of NO₂ across multiple reference stations, limiting portability to other cities with fewer regulatory grade instruments. Winter et al. further showed that the Alphasense electrochemical sensors maintain consistent performance over three years [46], indicating that the remaining calibration errors are more likely to reflect model limitations than sensor degradation. The remaining question is therefore more specific: what error persists even after a well-designed remote linear calibration?

We show here that the dominant remaining obstacle is not reference distance alone, but unresolved nonlinear sensor response. The NO₂-B43F electrochemical cell exhibits non-linear temperature-dependent sensitivity [43], so a single linear sensitivity coefficient necessarily represents only an average response across the training period. The result is dynamic-range compression: positive and negative departures from the mean are both damped, even when the calibrated record appears reasonable in average level. This problem persists even under colocation, implying that the main limitation is model structure rather than distance from the reference.

This study shows that nocturnal NO–O₃ titration provides the diagnostic needed to measure and correct that gain error. Previous work showed that titration chemistry can provide useful calibration constraints for BEACO₂N sensors through zero-offset estimation [44]. Here we extend the same physical principle from offset estimation to gain diagnosis. After sunset, the reaction



drives a near-1:1 exchange in which each ppb of O₃ consumed produces approximately one ppb of NO₂ [47]. Systematic departures of the observed ratio $\Delta\text{NO}_2/(-\Delta\text{O}_3)$ from unity directly quantify the multiplicative gain error. By accumulating these events across the seasonal temperature cycle, a temperature-dependent gain curve can be estimated independently at each sensor node from its own atmospheric observations, without additional hardware, laboratory characterization, or a colocated reference monitor.

Building on the existing BEACO₂N calibration infrastructure [44, 45], we develop a three-

stage calibration framework. The framework is evaluated using a withheld-truth design at two Bay Area colocated sites and then applied at network scale in the Bay Area, Providence, and Los Angeles. The goals in this study are threefold: to determine whether unresolved temperature-dependent gain is the principal source of remaining NO₂ calibration error; to test whether node-specific nocturnal titration signals can improve non-colocated calibration beyond what remote linear transfer alone achieves; and to evaluate whether the resulting framework recovers site-level NO₂ variability robustly across nodes and deployment settings.

3.2 Methods and Materials

This section describes the observational platform, reference data, and calibration framework used to develop and evaluate a three-stage titration-based calibration method for electrochemical NO₂ sensors in BEACO₂N. We first describe the sensor network and regulatory reference measurements that define the observational constraints of the problem. We then introduce the three-stage calibration framework and the validation design used to test it. Finally, we describe the implementation of each calibration stage in sequence.

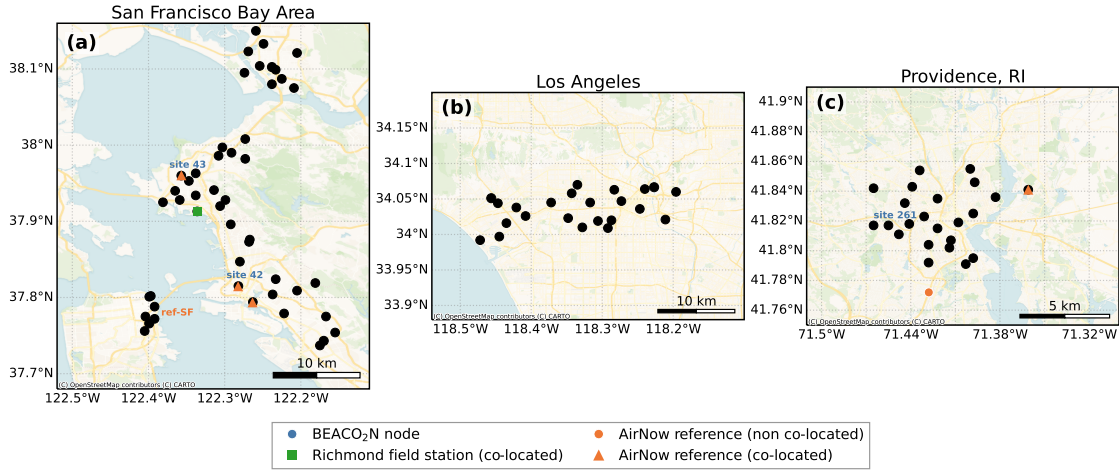


Figure 3.1: Locations of BEACO₂N nodes and AirNow reference monitors in the three study regions: (a) San Francisco Bay Area, (b) Los Angeles, and (c) Providence, Rhode Island. Blue circles denote BEACO₂N nodes, orange circles denote non-colocated AirNow reference monitors, orange triangles denote colocated AirNow reference monitors, and the green square marks the Richmond Field Station colocated site. Sites highlighted for later analyses are labeled in panels (a) and (c). Scale bars are 10 km in (a) and (b) and 5 km in (c).

3.2.1 Sensor network and reference measurements

This study uses observations from the Berkeley Environmental Air-quality and CO₂ Network (BEACO₂N), a dense urban sensor network deployed across three U.S. metropolitan regions (Figure 3.1; Table 3.1): the San Francisco Bay Area (~ 50 nodes), Los Angeles (~ 20 nodes), and Providence, Rhode Island (25 nodes, of which 22 have sufficient data coverage for April 2023–April 2024). Sensor nodes are typically spaced 1–3 km apart and measure NO₂, O₃, NO, and CO using Alphasense B-type electrochemical sensors (NO₂-B43F, O_x-B431, NO-B4, and CO-B4) mounted on individual sensor boards (ISBs). Signals are sampled every 8 s and averaged to one-minute intervals [44, 45]. Sensor-box temperature (T) and relative humidity (RH) are recorded concurrently with an Adafruit BME280. Winter et al. [46] show that these sensors maintain stable performance over multi-year deployments, indicating that the calibration errors addressed here arise primarily from model limitations rather than rapid sensor degradation. Because all three deployments use identical hardware and firmware, differences in calibration performance across cities should mainly reflect atmospheric conditions and calibration transfer challenges rather than instrument design.

Reference measurements are obtained from the U.S. EPA AirNow system [99], which provides quality-controlled hourly NO₂ (chemiluminescence), NO (chemiluminescence), and O₃ (UV photometry) measurements from Federal Reference or Equivalent Method analysers. Reference observations are aligned with sensor data at hourly resolution. Each sensor node is assigned to the AirNow station whose NO_x regime best matches its local environment. In Providence, 16 residential nodes are assigned to the East Providence (EP) station and 7 traffic-adjacent nodes to the South Providence (CCRI) station. Laney College is excluded from the NO₂ calibration index because it lacks an O₃ measurement. The Richmond Field Station (RFS) ozone measurement is used for ozone measurement error characterization. Station metadata are listed in Supplementary Table S1.

Table 3.1: Deployment summary. Colocated nodes are within 10 metre of a reference station.

City	N nodes	Period	Dist. range (km)	Median dist. (km)	N colocated
SF Bay Area	~ 50	2018–2025	0.01–21.4	3.6	3
Providence	~ 25	2022–2025	0.01–9.5	6.0	1
Los Angeles	~ 20	2022–2025	1.2–10.5	7.7	0

Together, these deployments provide a common instrumental framework across diverse atmospheric settings, allowing the calibration method to be evaluated under both colocated and non-colocated conditions. We next describe the three-stage calibration pipeline applied to these observations.

3.2.2 Calibration framework overview

The starting point for this work is the BEACO₂N multivariate linear regression framework [44, 45], in which each gas channel is calibrated against a reference monitor using electrode voltages and sensor temperature and humidity covariates. The present study retains that general structure for all species, but extends the NO₂ calibration with two additional stages that target errors a linear model cannot resolve. The resulting framework (Figure 3.2) consists of three stages: a reference-based baseline calibration, a physically constrained gain correction, and a residual offset adjustment.

Stage 1 removes linear cross-interferences through weighted least-squares (WLS) regression against an assigned reference, selecting training hours with a CO₂ enhancement-based coherence index. This index replaces the multi-reference NO₂ index used by Winter et al. [45], which requires three or more NO₂ reference stations, with a metric derived from the CO₂ measurements already available at every BEACO₂N node. This generalization is necessary because the multi-reference approach cannot be applied in cities with fewer than three NO₂ references; for example, Providence has only two NO₂ reference stations. Stage 2 applies a titration-based gain correction using only the sensor node’s own observations, with the gain curve applied up to a sensor-box temperature cap of 35°C — extending the corrected range by approximately 10% of the data relative to a more conservative 30°C threshold in previous BEACO₂N NO₂ calibration. Stage 3 re-anchors the absolute concentration level with a temperature-dependent rolling offset.

The same three-stage framework is used for both colocated and non-colocated nodes (Figure 3.2). In all cases, Stage 2 relies only on the sensor’s own observations, whereas Stages 1 and 3 use information from the assigned reference station. The O₃ data are calibrated using the Stage 1 regression only, and the calibrated outputs are then used as inputs to the NO₂ gain-correction framework.

3.2.3 Validation design

Before describing the implementation details of each stage, we briefly outline the evaluation design used throughout the paper. The framework is tested using a progressive comparison of four calibration methods at two Bay Area nodes colocated with regulatory reference monitors (sites 43 and 42; locations in Figure 3.1). In this validation design, the colocated reference is withheld entirely from the remote calibration steps. Stages 1 and 3 therefore use only remote AirNow stations, and Stage 2 uses only the sensor’s own atmospheric observations. The colocated reference is used solely as independent truth for evaluation. This design reproduces the information available at a truly non-colocated node while still retaining an end-to-end benchmark.

Table 3.2 lists the four calibration method choices. The comparison of the first two methods tests whether the gain correction improves calibration even when a colocated reference is available. This comparison separates structural limitations of the linear model from errors that arise only because the reference is remote. The latter two methods then evaluate

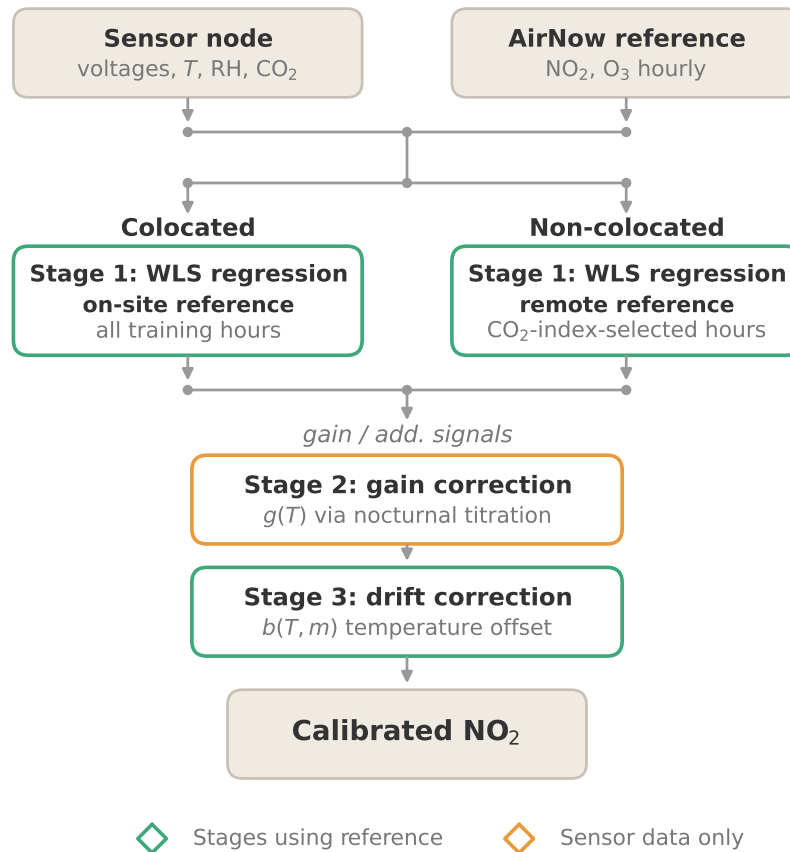


Figure 3.2: Calibration pipeline for the BEACO₂N NO₂ channel. Both colocated and non-collocated nodes pass through the same three stages. Stage 1 and Stage 3 use reference data, whereas Stage 2 uses only the sensor’s own nocturnal titration events. For colocated nodes, Stage 1 uses all training hours with an on-site reference; for non-collocated nodes, it uses CO₂-index-selected hours with a remote reference.

the remote pipeline with and without the gain correction, isolating its contribution under realistic dense-network conditions.

Table 3.2: Progressive comparison: four calibration methods evaluated against the withheld colocated EPA reference. The colocated reference is never used as a calibration input except in the colocated MLR baseline.

Method	Reference	Index	What it tests
MLR	Colocated	All hours	Baseline performance of a linear calibration with direct reference supervision
WLS+GC	Colocated	CO ₂ index	Effect of gain correction under colocated conditions
WLS	Remote	CO ₂ index	Remote pipeline without gain correction
WLS+GC	Remote	CO ₂ index	Full remote pipeline with gain correction

Performance is evaluated over the full deployment period at each site. In addition to full-year metrics, seasonal diagnostics are reported separately for summer (July–September, JAS) and winter (December–February, DJF). JAS isolates the period when temperature-dependent gain error is expected to be largest, owing to higher temperatures and elevated O₃, whereas DJF provides a complementary test under colder conditions, when titration events are most frequent but the gain departure from unity is expected to be smallest. This seasonal decomposition helps distinguish improvements that are concentrated in one regime from those that hold across the full annual cycle.

Results for site 43 under a representative remote reference assignment are presented in the main text; site 42 results and the full validation matrix across all remote reference assignments are provided in Supplementary Figures S1 and S30, respectively.

With the evaluation framework established, we now describe the three calibration stages in detail.

3.2.4 Stage 1: Reference-based baseline calibration

Stage 1 establishes the baseline calibration by removing the dominant linear cross-interferences and additive meteorological dependence through multivariate regression against the assigned reference station. This stage includes three components: pre-processing, training-hour selection, and weighted least-squares fitting.

Pre-processing

Raw sensor data are first cleaned by converting invalid values to NaN, computing working-minus-auxiliary electrode differentials, and applying manufacturer temperature and humidity corrections [43]. Outliers are then removed using an asymmetric interquartile range filter ($1.5 \times \text{IQR}$ below Q_1 ; $2.2 \times \text{IQR}$ above Q_3). This filtering removes approximately 0.6% of minute-level data, predominantly instrument transients and communication dropouts rather than atmospheric signals.

Training-hour selection

For non-colocated nodes, training hours are selected using a network-coherence index designed to identify periods when the sensor and its assigned remote reference are most likely sampling comparable air. The index is computed from the coefficient of variation (CoV) of CO₂ enhancement, defined as the difference between each node’s CO₂ measurement and a rolling network background, rather than from absolute CO₂. Computing the CoV on enhancement rather than on the raw CO₂ signal is important because the ~ 420 ppm common-mode background would otherwise compress all CoV values toward zero and reduce the index’s ability to distinguish spatially homogeneous from heterogeneous periods. For the NO₂ channel, this CoV is further scaled by the diurnal reference NO₂ profile in order to suppress photochemically active afternoon hours:

$$\text{CoV}_{\text{scaled}} = \frac{\text{CoV}_{\text{CO}_2}}{\widetilde{\text{NO}_{2,\text{ref}}(h)}/P_{95}} \quad (3.1)$$

Within rolling 3-month windows, hours falling in the lowest 25th percentile of this scaled CoV are selected for training. The O₃ calibration uses an analogous index that favours midday hours. For colocated nodes, all hours in the training window are used. Further details of the CO₂ enhancement calculation are provided in S-Methods-2.

Weighted least-squares fitting

The baseline NO₂ model is

$$\text{NO}_{2,\text{base}} = \beta_{\text{NO}_2} V_{\text{NO}_2} + \beta_{\text{NO}} V_{\text{NO}} + \beta_{\text{O}_x} V_{\text{O}_x} + \beta_T T + \beta_{\text{RH}} \text{RH} + \beta_0 \quad (3.2)$$

This model is fitted in rolling 9-month windows using the selected training hours. Observation weights are defined as

$$w_i = \frac{1}{\sigma_{\text{diff}}^2(y_i)} \quad (3.3)$$

where $\sigma_{\text{diff}}(y_i)$ is the standard deviation of the difference between colocated BEACO₂N measurements and the reference, evaluated at concentration y_i .

A key limitation of this linear model is that the coefficient β_{NO_2} represents only the average NO₂ sensitivity across the temperatures sampled during training. Because the B43F

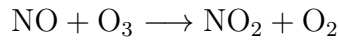
working-electrode sensitivity is temperature dependent [43], a single value of β_{NO_2} cannot fully represent the sensor response across the full temperature range. This remaining temperature dependence is non-linear, and therefore cannot be resolved by a linear model alone (i.e., the colocated multivariate linear regression). In Winter et al. [45], this issue was resolved by omitting data at temperatures above 30 °C where the issue is most significant. Stage 2 is therefore designed to diagnose and correct that temperature-dependent gain error using the sensor’s own atmospheric observations.

3.2.5 Stage 2: Titration-based gain correction

Stage 2 uses the sensor’s own nocturnal atmospheric observations to diagnose and correct the temperature-dependent multiplicative gain error that remains after Stage 1. The method exploits the stoichiometry of NO–O₃ titration and extends earlier uses of titration chemistry as a calibration constraint from zero-offset estimation [44] to gain-curve diagnosis.

Physical basis

After sunset, the gas-phase reaction



drives a 1:1 exchange in which each ppb of O₃ consumed produces one ppb of NO₂ [47]. If the electrochemical NO₂ channel carries a multiplicative gain error $g(T)$, the observed ΔNO_2 equals $g(T)$ times the true value. The O₃ channel is not assumed to be free of error, but it is generally more stable in this framework and O₃ is spatially more homogeneous than NO₂, so $-\Delta\text{O}_3$ provides the most useful available event-scale reference for the expected NO₂ response. We therefore define the titration ratio through a per-event regression,

$$R = m \quad \text{from} \quad \Delta\text{NO}_2 = m(-\Delta\text{O}_3) + a \quad (3.4)$$

where m is the event slope and a is an intercept that absorbs residual baseline mismatch between the two channels. Including a ensures that small baseline offsets do not bias the slope estimate; in practice, the fitted intercepts are small (median $|a| = 0.12$ ppb across all Providence events). Under ideal 1:1 titration and a correctly calibrated NO₂ response, R should be close to unity. Systematic departures of R from unity therefore diagnose the multiplicative error remaining in the NO₂ calibration.

Importantly, this diagnostic constrains only the part of the Stage 1 prediction that participates directly in the apparent NO₂ response during titration. In the Stage 1 model, that response is carried primarily by the V_{NO_2} term and by the V_{O_x} cross-sensitivity term. Both arise from electrochemical channels whose effective sensitivity varies with temperature, so both are expected to inherit the same multiplicative gain structure diagnosed by titration. By contrast, the NO cross-interference term and the explicit meteorological and intercept

terms are additive regression adjustments; they are not identified by the titration slope itself and should not be multiplicatively rescaled by a titration-derived gain factor. Stage 2 therefore applies the gain correction only to the NO₂ and O₃-cross-sensitivity terms, while leaving the other Stage 1 terms outside the gain scaling.

By accumulating R over the annual temperature cycle, a continuous gain curve can be estimated entirely from the sensor’s own observations.

Event detection

Nocturnal titration events are identified from the minute-level NO₂ and O₃ time series between 20:00 and 05:00 local time. The algorithm first applies a 3-minute rolling median to both channels to suppress short-timescale instrument noise. It then estimates the background noise level in each channel from successive minute-to-minute differences. An event onset is flagged when the rate of change in both channels exceeds 1.2 times this noise level. The event end is defined as the first point at which both channels have remained stable for at least five consecutive minutes.

Candidate events are then filtered to retain only those that are physically consistent with NO–O₃ titration. First, the NO₂ change during the event must be at least 1.0 ppb, ensuring sufficient signal above the instrument noise floor. Second, the pre-event baseline must be stable, with a concentration span no greater than 6.0 ppb, so that the event amplitude is referenced to a well-defined background level. Third, the fitted slope of ΔNO_2 against $-\Delta\text{O}_3$ must fall between 0.10 and 1.80, excluding both failed regressions and events influenced by fresh traffic plumes that can produce anomalously large slopes. Finally, the coefficient of determination (R^2) of the per-event regression must exceed 0.50, retaining only events whose NO₂ and O₃ changes are reasonably well described by the expected titration relationship.

All detection parameters (Table 3.3) are held fixed across nodes and cities. A one-at-a-time sensitivity analysis shows that the resulting minimal gain curve changes across all tested configurations (Figures S20, S21, and S28; Table S2).

Table 3.3: Event detection parameters, held constant across all nodes and cities.

Parameter	Value	Physical motivation
Night window	20:00–05:00 local	Captures titration without photolysis
NO ₂ change minimum	1.0 ppb	Signal-to-noise floor for reliable slope fitting
Baseline span	≤ 6.0 ppb	Pre-event stability requirement
End condition	5 stable minutes, both channels	Ensures physically complete events
Slope acceptance range	0.10–1.80	Excludes failed fits and traffic contamination
Per-event R^2 minimum	0.50	Stoichiometric consistency filter

Gain-curve construction

At each node, the retained per-event R values are binned by temperature at event onset using 2°C bins. Within each bin, the gain $g(T)$ is estimated as the weighted median

$$w_i = \frac{N_i \cdot \text{median}(|\Delta\text{NO}_{2,i}|)}{\text{IQR}(R_i)} \quad (3.5)$$

where N_i is the number of events in the bin, $|\Delta\text{NO}_2|$ is the event amplitude, and $\text{IQR}(R_i)$ is the within-bin interquartile range of R . This weighting favours bins that are well sampled, contain larger-amplitude events, and exhibit more internally consistent stoichiometry. Bins with fewer than 5 events are excluded.

The gain curve is fit within rolling 9-month centered windows rather than from all events pooled together, allowing $g(T)$ to evolve over the deployment. For each calendar month m , events within $[m-4, m+4]$ months are used; edge months inherit the nearest well-sampled curve. The bin medians are interpolated to a continuous function, and outside the populated temperature range $g(T)$ is held at the nearest valid value. Sensitivity to window length and weighting scheme is evaluated in Supplementary Figures S21 and S23.

Signal decomposition and gain application

The titration-derived gain correction is not applied to the full Stage 1 output. Instead, the baseline prediction is separated into the part whose amplitude is directly constrained by NO–O₃ titration and the part that is not:

$$S_{\text{gain}} = \beta_{\text{NO}_2} V_{\text{NO}_2} + \beta_{\text{O}_x} V_{\text{O}_x} \quad (3.6)$$

$$S_{\text{add}} = \beta_{\text{NO}} V_{\text{NO}} + \beta_T T + \beta_{\text{RH}} \text{RH} + \beta_0 \quad (3.7)$$

This grouping follows directly from what the titration diagnostic can and cannot constrain. During a nocturnal titration event, the fitted slope $R = \Delta\text{NO}_2/(-\Delta\text{O}_3)$ diagnoses the amplitude of the apparent NO₂ response relative to the paired O₃ loss. In the Stage 1 model, those event-scale changes are carried primarily by the V_{NO_2} term and by the V_{O_x} cross-sensitivity term. The gain correction is therefore applied only to S_{gain} .

By contrast, the remaining terms are not identified by the NO–O₃ titration slope. The $\beta_{\text{NO}} V_{\text{NO}}$ term represents cross-interference from the NO channel rather than the NO₂ response itself, and it shows little systematic temperature dependence (Supplementary Figure S8). The terms $\beta_T T$, $\beta_{\text{RH}} \text{RH}$, and β_0 are additive regression adjustments that account for baseline offset, not signal amplitudes diagnosed by titration stoichiometry.

Within S_{gain} , the correction should be applied only to the NO₂-related anomaly above the sensor’s effective zero, that is, the output level when ambient NO₂ is absent. Under zero-NO₂ conditions, the sensor still responds to O₃ through the $\beta_{\text{O}_x} V_{\text{O}_x}$ cross-sensitivity term, so S_{gain} remains non-zero even when there is no ambient NO₂. If S_{gain} were divided

directly by $g(T)$, this O₃-driven floor would be spuriously rescaled together with the true NO₂ signal.

The NO₂ zero level, denoted Z_{NO_2} , estimates this floor. It is obtained by regressing S_{gain} against V_{O_x} , T , and RH during hours when ambient NO₂ is near zero (remote reference below its rolling 10th percentile within a 3-month window). During these hours, S_{gain} is dominated by the O₃ cross-sensitivity term, and the regression captures how this floor varies with the O₃ electrode signal and meteorological conditions (Supplementary Section S-Methods-1).

Because V_{O_x} responds to both O₃ and common-mode noise shared with the NO₂ electrode, but does not perfectly reproduce the O₃-related contribution captured in the NO₂ channel, the raw regression estimate $\hat{Z}_{\text{NO}_2}$ is expected to overestimate the effective zero level. A scaling factor $\lambda = 0.7$ is therefore applied:

$$Z_{\text{NO}_2} = \lambda \cdot \hat{Z}_{\text{NO}_2} \quad (3.8)$$

The sensitivity of the calibration performance to λ is evaluated in Supplementary Figure S26. Calibration performance is optimised near $\lambda = 0.7$, so this value is adopted empirically in the present framework. This result indicates that the unscaled regression estimate tends to overestimate the effective zero level.

The gain-corrected concentration is then

$$\text{NO}_{2,\text{gc}} = \frac{S_{\text{gain}} - Z_{\text{NO}_2}}{g(T)} + Z_{\text{NO}_2} + S_{\text{add}} \quad (3.9)$$

The numerator isolates the part of the Stage 1 signal whose amplitude is diagnosed by titration. Dividing by $g(T)$ restores that amplitude, after which the effective zero level and the additive terms are restored without gain scaling. The correction is bypassed above a sensor-box temperature cap ($T_{\text{CAP}} = 35^\circ\text{C}$), where titration events are too sparse to constrain the gain curve reliably, and when $\text{NO}_{2,\text{base}} \leq -2.5$ ppb to avoid amplifying noise near the detection limit. The adopted cap of 35°C represents an increase from the 30°C value used in Winter et al. [45], extending the correction to approximately 10% more of the minute-level data. Sensitivity testing across caps of 25 – 45°C (Supplementary Figure S29) shows that full-year slope improves monotonically up to $T_{\text{CAP}} \approx 35^\circ\text{C}$, beyond which the gain curve relies increasingly on extrapolation into temperature bins with few events and performance levels off.

3.2.6 Stage 3: Rolling offset correction

After the gain correction, a slowly varying residual bias remains. This bias reflects both true sensor drift and the persistent spatial NO₂ gradient between the sensor node and its assigned reference. Stage 3 addresses this remaining error with a temperature-dependent rolling offset that is intended to correct low-frequency residual bias without altering the gain-corrected variability recovered in Stage 2.

Table 3.4: Calibration models for all gas channels. V denotes the working-minus-auxiliary electrode differential. O₃, NO, and CO are calibrated by Stage 1 regression only, and their outputs feed the NO₂ correction stages. β coefficients follow the notation of [45] with chemically descriptive subscripts.

Species	Model
NO ₂ (Stage 1)	$\beta_{\text{NO}_2} V_{\text{NO}_2} + \beta_{\text{NO}} V_{\text{NO}} + \beta_{\text{O}_x} V_{\text{O}_x} + \beta_T T + \beta_{\text{RH}} \text{RH} + \beta_0$
NO ₂ (Stage 3)	$\frac{\beta_{\text{NO}_2} V_{\text{NO}_2} + \beta_{\text{O}_x} V_{\text{O}_x} - Z_{\text{NO}_2}}{g(T)} + \beta_{\text{NO}} V_{\text{NO}} + \beta_{T,\text{new}} T + \beta_{\text{RH}} \text{RH} + \beta_{0,\text{new}}$
O ₃	$\beta_{\text{O}_3} V_{\text{O}_3} + \beta_{\text{NO}_2} V_{\text{NO}_2} + \beta_{\text{NO}} V_{\text{NO}} + \beta_T T + \beta_{\text{RH}} \text{RH} + \beta_0$
NO	$\beta_{\text{NO},T} V_{\text{NO},T} + \beta_{(\text{NO},T)^2} V_{\text{NO},T}^2 + \beta_{\text{NO}_2} V_{\text{NO}_2} + \beta_T T + \beta_{\text{RH}} \text{RH} + \beta_0$
CO	$\beta_{\text{CO}} V_{\text{CO}} + \beta_{\text{NO}_2} V_{\text{NO}_2} + \beta_T T + \beta_{\text{RH}} \text{RH} + \beta_0$

In the main derivation, $S_{\text{gain}} = \beta_{\text{NO}_2} V_{\text{NO}_2} + \beta_{\text{O}_x} V_{\text{O}_x}$ and $S_{\text{add}} = \beta_{\text{NO}} V_{\text{NO}} + \beta_T T + \beta_{\text{RH}} \text{RH} + \beta_0$. In the compact Stage 3 table entry, $\beta_{T,\text{new}} T \equiv \beta_T T + b(T, m)$ and $\beta_{0,\text{new}} \equiv \beta_0 + Z_{\text{NO}_2}$ are used only as shorthand to highlight where the offset terms are updated after the gain correction. The Stage 1 NO₂ model follows the multivariate linear regression form of Winter et al. [45], but is fit here using weighted least squares. When $g(T) = 1$ and $b = 0$, the Stages 3 model reduces to Stage 1.

Within each calendar month, centred in a rolling 5-month window, background hours are identified as those when the remote reference NO₂ falls below its 15th percentile. These are the conditions under which the reference is expected to be most spatially representative of the sensor location. Within each window, the background hours are binned by temperature (5°C bins), and the offset is defined as

$$b(T_k, m) = \overline{\text{NO}_{2,\text{ref}}}_{T_k, m} - \overline{\text{NO}_{2,\text{gc}}}_{T_k, m} \quad (3.10)$$

The per-bin offsets are linearly interpolated to form a continuous $b(T, m)$. When temperature coverage is insufficient, the offset falls back to a scalar $b(m)$. This final stage therefore updates only the additive part of the calibration, not the multiplicative gain correction established in Stage 2. In the compact notation of Table 3.4, this update is represented through the effective terms $\beta_{T,\text{new}} T$ and $\beta_{0,\text{new}}$, which absorb the temperature-dependent offset surface and the restored constant offset, respectively.

Sensitivity to the rolling-window length and percentile threshold is shown in Supplementary Figure S22. The final calibrated concentration is

$$\text{NO}_{2,\text{final}} = \text{NO}_{2,\text{gc}} + b(T, m) \quad (3.11)$$

For clarity and completeness, the calibration models used in this study are summarised in Table 3.4.

3.3 Results and Discussion

This section evaluates the titration-based gain correction from four complementary perspectives. We first test whether the correction improves agreement with independent truth using Bay Area sites with withheld colocated reference measurements (Section 3.3.1). We then examine the titration-ratio diagnostics that identify the underlying error source in Providence (Section 3.3.2), followed by residual diagnostics that show which systematic biases are removed and which remain (Section 3.3.3). We next ask whether the fitted gain curves behave as a reproducible sensor property across nodes and cities (Section 3.3.4). We then show the practical network-scale consequence of the correction, namely recovery of site-level NO₂ variability (Section 3.3.5). Finally, we discuss the main limits of the framework (Section 3.3.6).

3.3.1 Validation against withheld colocated truth

We begin with the most direct test of performance: agreement with independent truth. This validation is carried out in the Bay Area, the only BEACO₂N deployment with more than two colocated regulatory reference sites and therefore the only city where a withheld-truth design can be implemented while preserving realistic remote-reference options. We evaluate the four methods in Table 3.2 at two Bay Area sites: node 43, colocated with the San Pablo reference, and node 42, colocated with the EBMUD reference. Table 3.5 summarises full-year performance at site 43. For the colocated methods, the table reports single values because the on-site reference assignment is unique. For the remote methods, it reports the median and standard deviation across four remote reference assignments, thereby capturing both typical performance and sensitivity to reference choice. Site 42 results are similar and are provided in Supplementary Table S3.

The colocated and remote comparisons address two distinct questions. The colocated case tests whether the dominant calibration error persists even when the sensor is trained against an on-site reference. The remote case then tests whether the same correction remains effective under the conditions that apply to most of the network, where no colocated reference is available. This separation distinguishes structural limitations of the linear model from errors introduced by remote reference transfer.

In the colocated case, the baseline MLR already performs reasonably well by conventional metrics, with RMSE of 2.71 ppb and $R^2 = 0.74$, yet its slope against truth is only 0.72. Even under direct local reference supervision, a true 10 ppb change in ambient NO₂ is represented as only about 7 ppb. Applying the gain correction raises the slope to 1.00.

That slope recovery is accompanied by increases in RMSE (2.71 to 3.62 ppb) and MAE (1.80 to 2.49 ppb), with a modest decrease in R^2 (0.74 to 0.69). This pattern is physically consistent with the removal of dynamic-range compression. A compressed calibration pulls predictions toward the mean and suppresses apparent scatter around the reference, even when it is damping real atmospheric variability. Once the gain correction restores the full amplitude of the signal, scatter-based metrics increase accordingly. For applications that

Table 3.5: Progressive comparison results at site 43, evaluated against the withheld colocated reference on hourly data. Colocated methods use on-site reference hours and yield single values; remote methods report the median ± 1 s.d. across four remote reference assignments. Slope is from OLS regression of calibrated values against truth (ideal = 1.0). Site 42 results are in SI Table S3.

Method	RMSE (ppb)	NRMSE	MAE (ppb)	Bias (ppb)	Slope	R ²
<i>Colocated reference</i>						
MLR	2.71	0.44	1.80	-0.15	0.72	0.74
WLS+GC	3.62	0.59	2.49	-0.28	1.00	0.69
<i>Remote reference (median ± 1 s.d.)</i>						
WLS	3.48 $\pm$ 0.52	0.56 $\pm$ 0.09	2.51 $\pm$ 0.51	0.85 $\pm$ 1.26	0.77 $\pm$ 0.11	0.66 $\pm$ 0.05
WLS+GC	3.64 $\pm$ 0.34	0.59 $\pm$ 0.06	2.70 $\pm$ 0.28	-0.79 $\pm$ 0.71	0.92 $\pm$ 0.07	0.65 $\pm$ 0.02

depend on recovering neighbourhood-scale contrasts, this tradeoff is not a weakness of the correction but a consequence of evaluating a less compressed and therefore more realistic signal.

The remote case is more operationally relevant. Under remote WLS alone, the median slope is 0.77 ± 0.11 . Adding the gain correction raises the median slope to 0.92 ± 0.07 , showing that most of the missing amplitude can be recovered from the sensor node’s own atmospheric observations. The spread in slope across reference assignments also decreases, indicating that the correction makes performance less sensitive to which remote reference is used in Stages 1 and 3. The remaining departure from unity is therefore more plausibly attributed to spatial mismatch in the reference-based stages than to failure of Stage 2.

The bias shifts from $+0.85 \pm 1.26$ ppb under remote WLS to -0.79 ± 0.71 ppb under remote WLS+GC. Restoring dynamic range and reapplying the rolling offset changes not only the slope but also the mean level. More importantly, the standard deviation of the bias is reduced, so the final bias is more consistent across remote reference assignments. Its remaining magnitude lies within the range of pairwise biases among Bay Area EPA reference stations themselves (Supplementary Figure S17), suggesting that the residual error has been reduced to the level of spatial variability already present in the regulatory network.

These full-year metrics also place the new framework in context. Winter et al. [45] report $R^2 = 0.66$ and $\text{RMSE} = 3.16$ ppb for remotely calibrated NO₂ using a linear model alone. The remote WLS+GC pipeline achieves a comparable RMSE (3.64 ± 0.34 ppb) while improving slope from 0.77 ± 0.11 to 0.92 ± 0.07 . The main value of the new framework is not a reduction in headline RMSE, but a substantially more faithful recovery of atmospheric variability.

Seasonal diagnostics further clarify both the value and the current limits of the framework.

Figure 3.3 presents seasonal scatter plots for remote WLS and remote WLS+GC at site 43 under a representative reference assignment. Summer (JAS) is the most challenging regime. Under remote WLS, the JAS slope falls to 0.37 with $R^2 = 0.10$; in winter (DJF), the corresponding values are 0.64 and 0.53. The gain correction improves both seasons: in JAS, slope rises from 0.37 to 0.64 and R^2 from 0.10 to 0.28; in DJF, slope rises from 0.64 to 0.75 and R^2 from 0.53 to 0.61. The largest absolute improvement occurs in summer, where the temperature-dependent gain error is expected to be strongest.

At the same time, WLS+GC does not eliminate seasonal degradation. Two factors likely contribute to the remaining summer difficulty. First, ambient NO₂ concentrations are lower in summer, so the signal lies closer to the electrochemical noise floor and the effective signal-to-noise ratio is reduced. Second, summertime photochemistry may introduce additional cross-interferences from reactive species that are not represented explicitly in the current calibration model. The gain correction removes the dominant temperature-dependent bias, but it cannot compensate for these additional error sources.

A further practical advantage of the remote framework is that training-hour selection does not depend on a dense set of NO₂ reference sites. Instead, it uses the accompanying BEACO₂N CO₂ observations to identify spatially homogeneous periods through the CO₂-index method. This makes the approach portable to cities where a dense NO₂ reference network does not exist. The withheld-truth results therefore establish the central empirical result of the paper: relative to linear remote transfer alone, the titration-based gain correction recovers substantially more of the true NO₂ variability under the conditions that matter for dense-network deployment. The same qualitative pattern is observed at site 42 (Supplementary Table S3), confirming that the improvement is not specific to a single colocated site.

3.3.2 Titration ratio diagnostics

The withheld-truth analysis shows that the gain correction improves performance. We next turn to the physical diagnostic that explains why. A temperature-dependent multiplicative gain error in the NO₂ electrochemical cell predicts a simple and testable pattern: the titration ratio $R = \Delta\text{NO}_2/(-\Delta\text{O}_3)$ should deviate systematically from unity as temperature changes. Providence provides a particularly strong setting for this test because the network spans a broad seasonal temperature range and all nodes rely on remote-reference calibration. We focus here on node 261, calibrated against a remote reference approximately 9 km away. The event-detection algorithm identifies 1,114 nocturnal titration events over the 10-month deployment period. Figure 3.4 shows how R varies with five environmental quantities before and after the per-node gain and offset correction.

Before correction, the clearest structure is a pronounced negative correlation between R and temperature (Spearman $\rho = -0.454$, $p < 0.001$): cold-weather events cluster near or above unity, whereas warm-weather events fall systematically below 1. This is the expected signature of a multiplicative gain error whose magnitude increases with temperature. A weaker but still significant negative correlation with relative humidity is also present ($\rho =$

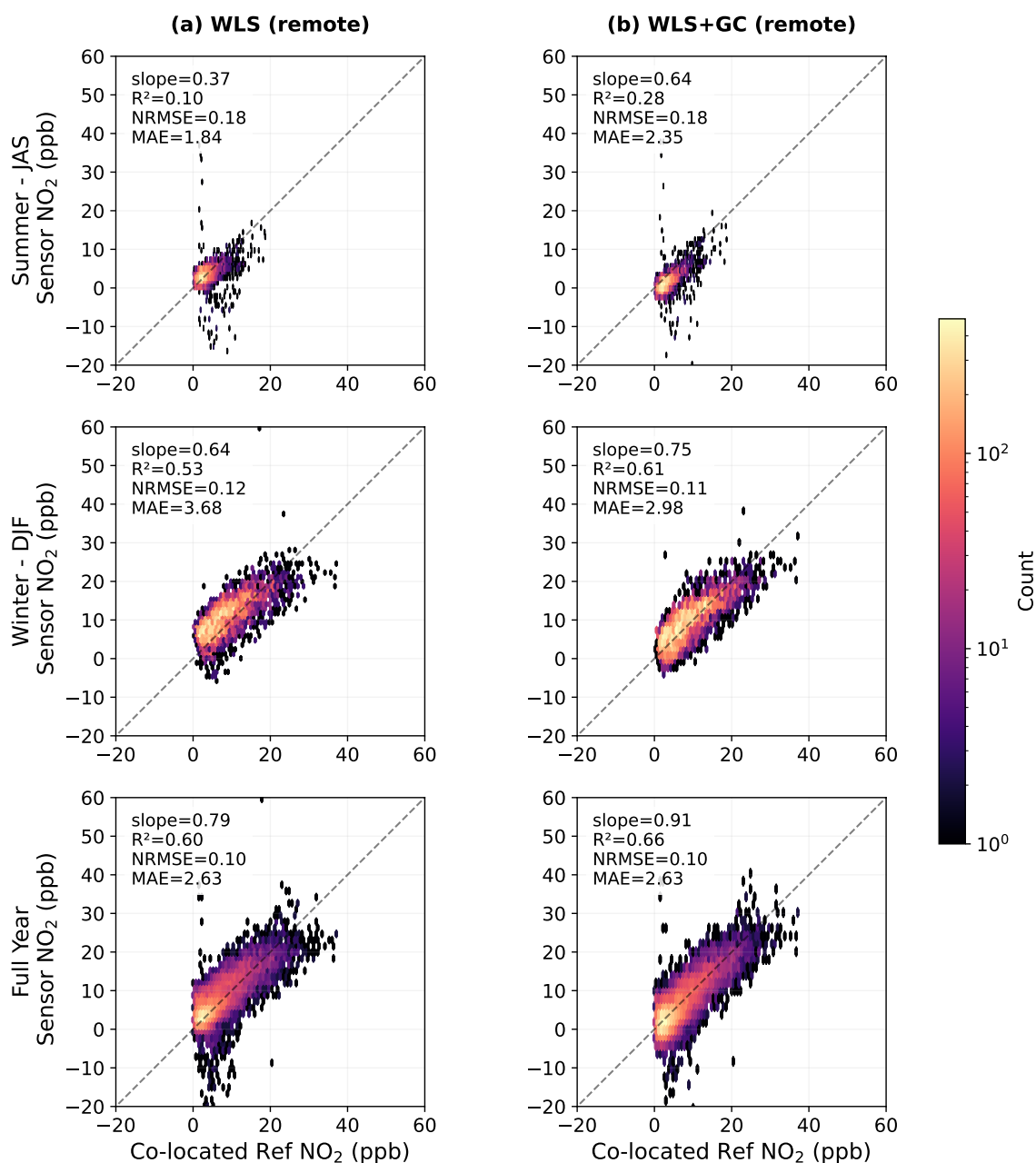


Figure 3.3: Calibrated NO₂ versus withheld colocated truth at site 43 for remote WLS (left column) and remote WLS+GC (right column), evaluated over three windows: summer (JAS; top), winter (DJF; middle), and full year (bottom). Colour indicates hexbin point density on a log scale. Dashed line: 1:1. Metrics are computed from a single representative remote reference assignment; Table 3.5 reports the median across all four assignments.

-0.140 , $p < 0.001$), together with a positive correlation with baseline NO₂ ($\rho = +0.326$, $p < 0.001$) and a weaker positive correlation with O₃ drawdown amplitude ($\rho = +0.164$, $p < 0.001$). In contrast, baseline O_x shows essentially no relationship with R before correction ($\rho = +0.004$, $p = 0.91$).

After applying the per-node gain curve and monthly offset, the structure of R changes in a physically interpretable way. The temperature dependence is effectively removed ($\rho: -0.454 \rightarrow -0.035$, $p = 0.24$), and the humidity correlation collapses at the same time ($\rho: -0.140 \rightarrow +0.026$, $p = 0.39$), even though the correction is parameterized only on temperature. That joint disappearance strongly suggests that the two relationships do not reflect independent environmental controls, but rather a shared underlying sensor process, most plausibly the effects of temperature on the electrochemical kinetics at the working electrode [100].

The remaining structure is weaker and more selective. The positive dependence of R on baseline NO₂ is reduced but not eliminated ($\rho: +0.326 \rightarrow +0.210$, $p < 0.001$), indicating that some concentration-dependent structure remains after the multiplicative correction. The correlation with O₃ drawdown amplitude weakens substantially and becomes only marginally significant ($\rho: +0.164 \rightarrow +0.058$, $p = 0.055$), showing that most of the amplitude dependence is removed together with the gain bias. Finally, baseline O_x, which showed no detectable relationship with R before correction, develops a weak positive correlation after correction ($\rho = +0.095$, $p = 0.0016$). Because this residual structure is small relative to the pre-correction temperature signal and appears only after the dominant bias is removed, it can be reasonably interpreted as a secondary atmospheric or concentration-dependent effect than as evidence against the correction.

Taken together, these diagnostics show that the correction is targeting the correct error source. The strongest pre-correction structure in R is the one expected from sensor physics, and it is the one that collapses after applying the node-specific gain curve. What remains is weaker: a reduced dependence on baseline NO₂, a marginal dependence on event amplitude, and a small positive dependence on baseline O_x. Those residuals are consistent with effects that a temperature-only multiplicative correction would not be expected to resolve fully. The same qualitative behaviour is observed across all 22 Providence nodes (Supplementary Figure S4) and in the Bay Area and Los Angeles networks (Supplementary Figures S5 and S6), indicating that the ideal performance of gain correction at node 261 is representative.

The logic connecting this section to the withheld-truth validation is straightforward. The Bay Area analysis showed that WLS+GC recovers slope against independent truth. The Providence analysis shows that this improvement is not a black-box statistical effect: it corresponds to removal of the specific temperature-structured departure from 1:1 NO–O₃ titration expected from a temperature-dependent gain error.

3.3.3 Remote calibration error structure

The progressive comparison shows that the gain correction primarily improves calibration by restoring slope. We next examine the residual structure in order to identify which systematic

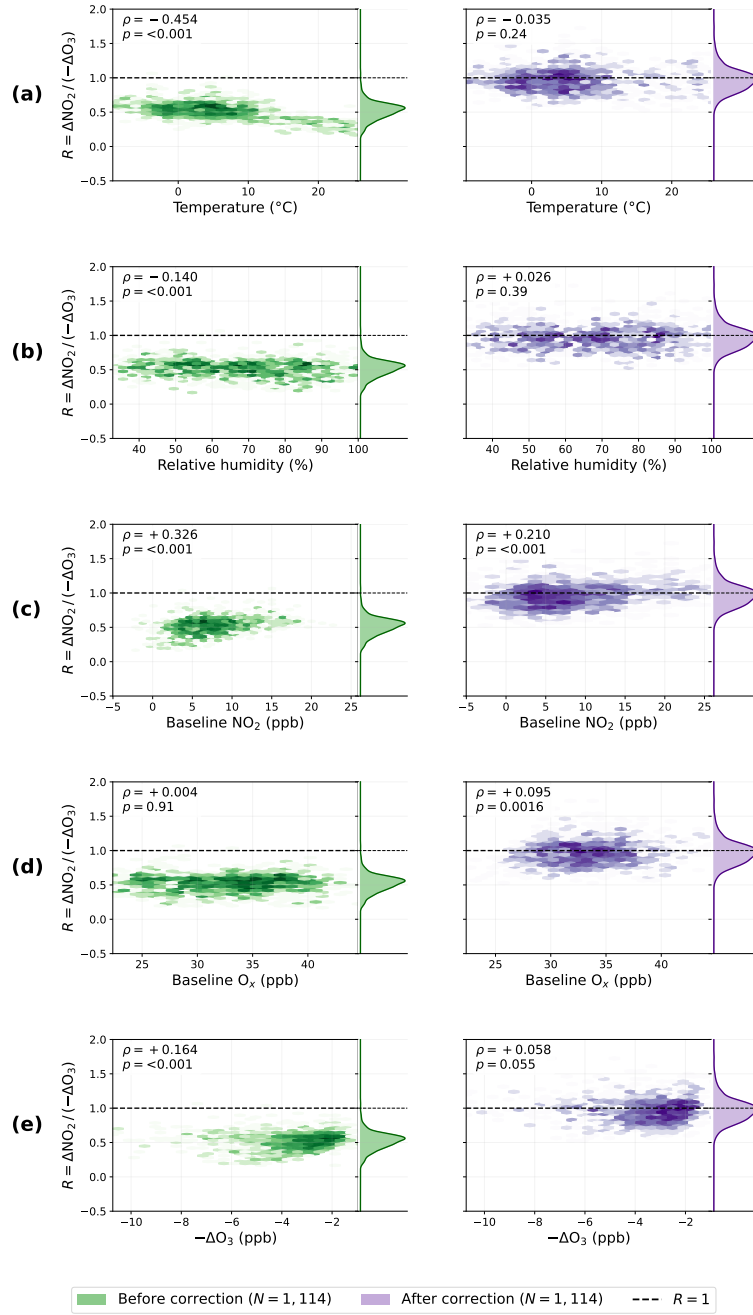


Figure 3.4: Titration ratio $R = \Delta\text{NO}_2 / (-\Delta\text{O}_3)$ versus five environmental variables before (left, green) and after (right, purple) per-node gain and offset correction at Providence node 261 ($N = 1,114$ events). Rows show (a) sensor-box temperature, (b) relative humidity, (c) baseline NO₂, (d) baseline O_x, and (e) O₃ drawdown amplitude. Spearman ρ and p -values are annotated in each panel; marginal distributions are shown at right. The dashed horizontal line marks $R = 1$. Network-wide results are in Supplementary Figure S4.

biases are removed by the correction and which persist. Figure 3.5 shows the calibration residual (calibrated NO₂ minus withheld truth) as a function of temperature, relative humidity, and reference NO₂ concentration for the two remote methods at site 43.

Under remote WLS alone, the most prominent feature is a clear temperature-dependent decline in the residual. Residuals are near or slightly above zero under cool conditions, but become increasingly negative as temperature rises, reaching their most negative values under the warmest conditions. This is the same bias diagnosed through the titration-ratio analysis in Providence, now seen directly against withheld truth. In other words, the baseline remote WLS calibration systematically underestimates NO₂ under warm conditions, which is exactly the behaviour expected from a linear model that uses a single temperature-averaged sensitivity coefficient.

A corresponding but weaker structure appears in the humidity panels. Residuals are most negative at the driest conditions, increase toward near-zero or slightly positive values at intermediate relative humidity, and then decrease again somewhat at the highest humidity. This is consistent with the fact that low humidity in the Bay Area often coincides with warm conditions where the gain bias is strongest. The humidity dependence therefore appears not as an independent source of error, but largely as a secondary reflection of the same temperature-dependent sensor response.

After applying the gain correction, the temperature-structured bias is substantially flattened. Residuals remain close to zero across most of the observed temperature range, and the strong warm-temperature underestimation seen under WLS alone is greatly weakened. A modest negative residual persists above approximately 35°C, where nocturnal titration events are sparse and the gain curve is therefore least well constrained. The humidity dependence is flattened at the same time, reinforcing the interpretation from Section 3.3.2 that the apparent humidity sensitivity is largely inherited from the underlying temperature-dependent gain bias rather than reflecting a separate correction target.

By contrast, the residual dependence on reference NO₂ concentration changes relatively little after the gain correction. In both methods, residuals remain near zero at low NO₂ and become increasingly negative at higher concentrations. That remaining structure is not unexpected, because the gain correction is designed to remove temperature-dependent multiplicative bias, not all residual error. Two effects likely contribute. First, remaining instrument noise is partly multiplicative and therefore grows with signal amplitude. Second, spatial mismatch between the sensor and its assigned remote reference becomes most important during high-NO₂ episodes, when local source influence is strongest and the two sites are least likely to sample the same air mass.

This residual analysis clarifies the division of labour among the three stages of the pipeline. Stage 2 removes the dominant temperature-structured component that cannot be represented by a linear model. What remains is concentrated in regimes where either the gain curve is weakly constrained, such as very warm conditions, or where the remote reference itself becomes a limited proxy for local truth, such as high-concentration episodes.

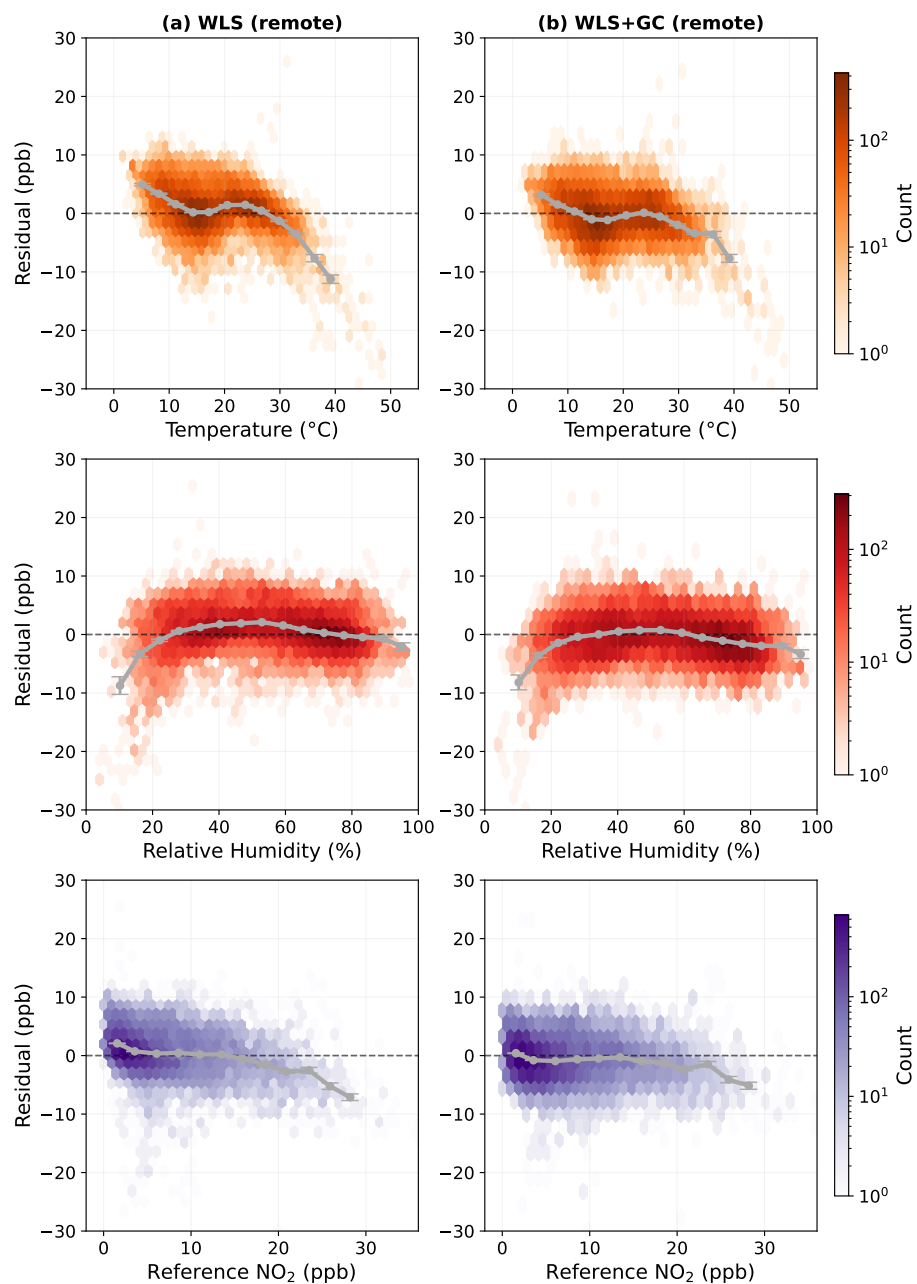


Figure 3.5: Calibration residual (calibrated NO₂ minus withheld truth, ppb) as a function of temperature (top), relative humidity (middle), and reference NO₂ concentration (bottom) at site 43, full-year evaluation. Left column: remote WLS; right column: remote WLS+GC. Colour indicates point density on a logarithmic scale; grey lines show bin means with ± 1 standard error bars. The dashed horizontal line marks zero residual.

3.3.4 Gain-curve convergence across nodes and cities

The preceding sections show that the gain correction improves agreement with withheld truth and removes the temperature-structured bias expected from sensor physics. The next question is whether the fitted gain curve reflects a reproducible property of the B43F sensor response or instead absorbs site-specific atmospheric structure. If the titration diagnostic is primarily capturing sensor behaviour, then independently estimated gain curves should converge across nodes exposed to different local environments and, more broadly, across cities with different climates.

Figure 3.6 presents the per-node gain curves for all three BEACO₂N cities. In each network, the independently fitted curves cluster around a common negative temperature dependence without any city-specific retuning of the method. The network-median slopes are $-0.004\text{ }^{\circ}\text{C}^{-1}$ in Providence (22 nodes; median per-node $dg/dT = -0.002\text{ }^{\circ}\text{C}^{-1}$ [$-0.004, -0.0005$]), $-0.001\text{ }^{\circ}\text{C}^{-1}$ in the Bay Area (41 nodes; $-0.0003\text{ }^{\circ}\text{C}^{-1}$ [$-0.002, +0.0001$]), and $-0.003\text{ }^{\circ}\text{C}^{-1}$ in Los Angeles (7 nodes; $-0.004\text{ }^{\circ}\text{C}^{-1}$ [$-0.006, -0.002$]). All three are negative, consistent in sign with the temperature-dependent sensitivity reported for the B43F cell [43]. At a given temperature, the interquartile spread across nodes within each city remains modest relative to the total change in the network median over the observed temperature range. This convergence across nodes exposed to distinct local environments argues that the titration diagnostic is capturing a reproducible sensor response rather than simply fitting local atmospheric variability.

The differences in slope magnitude across cities are also physically interpretable. Providence, which spans the widest annual temperature range, shows the steepest median slope; the Bay Area, with its more moderate maritime climate, shows the shallowest; and Los Angeles falls between the two. This ordering is consistent with the same underlying B43F temperature dependence being expressed under different ranges and covariances of temperature and humidity. The Bay Area curves show greater scatter near the temperature extremes, where nocturnal titration events are sparse and the fitted curves are therefore less well constrained. Los Angeles, with only 7 nodes, shows the widest interquartile range, reflecting both the smaller sample size and the more heterogeneous photochemical environment of the LA basin. Even so, the network medians in all three cities remain monotonically decreasing with temperature.

The network-level R diagnostics support the same interpretation (Table 3.6). In Providence and the Bay Area, the pre-correction median R is well below unity (0.65 and 0.54, respectively), indicating that the Stage 1 calibration systematically under-reports NO₂ variability during titration events. After correction, both networks shift toward nearly the same median value ($R \approx 0.9$) with similar interquartile ranges, and the fraction of events within the physically expected range $R \in [0.7, 1.3]$ roughly doubles (from 39% to 81% in Providence and from 24% to 79% in the Bay Area). Although the fitted gain curves differ in slope and shape between climates, the correction moves both networks toward the same physically expected 1:1 titration behaviour.

Los Angeles presents a qualitatively different pre-correction pattern: the median R is 1.10,

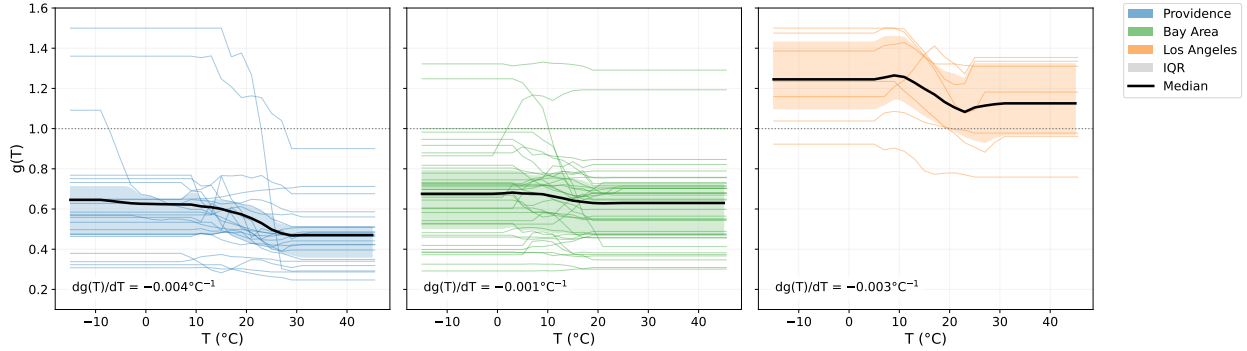


Figure 3.6: Per-node gain curves across three cities: (a) Providence, (b) Bay Area, (c) Los Angeles. Thin lines show individual nodes; bold black lines show the network median; shading indicates the interquartile range. The median slopes are annotated. The dotted horizontal line marks $g(T) = 1$.

above unity rather than below it, indicating that the Stage 1 calibration over-reports NO₂ variability during titration events. This most likely reflects the proximity of both available LA reference stations to major transportation sources. North Main Street station is located within 1 km of highways in all directions, and West Los Angeles station is approximately 220 m from the nearest highway. Both stations therefore read systematically higher in NO₂ than most sensor locations in the network. The Stage 1 WLS regression, trained against this elevated reference, fits an inflated NO₂ sensitivity coefficient that amplifies the calibrated ΔNO_2 during titration events at the sensor sites. Because O₃ is more spatially homogeneous, the calibrated $-\Delta\text{O}_3$ is less affected, and the ratio R exceeds unity. In Providence and the Bay Area, by contrast, nodes can be assigned to the reference station that best matches their local NO_x regime, reducing this reference-mismatch effect. Despite the opposite starting point in Los Angeles, the per-node gain correction brings the network to a post-correction median of 0.99 with 89% of events within $[0.7, 1.3]$, closer to the ideal than either of the other two cities. That the same framework corrects both under- and over-reporting of NO₂ variability without city-specific tuning provides additional evidence that the gain curve is diagnosing a genuine sensor property rather than a statistical artefact of a particular reference configuration.

The full per-node distributions of R before and after correction are shown in Supplementary Figure S4 (Providence) and Figures S5 and S6 (Bay Area and Los Angeles). In all three cities, the distributions shift toward more compact shapes near $R = 1$.

Together, these convergence tests strengthen the physical interpretation of the method. The gain curve is not merely a flexible empirical adjustment that improves validation statistics at an individual site. Rather, it behaves as a reproducible feature of the sensor response that can be estimated independently from each node’s own atmospheric observations and that corrects gain errors in both directions across climatically distinct deployment settings.

Table 3.6: Titration ratio R diagnostics before and after gain and offset correction. Med. R is the median across per-node median values; IQR is the interquartile range across per-node medians; $R \in [.7, 1.3]$ is the percentage of individual events falling within that range.

City	N_{evt}	Before correction			After correction		
		Med. R	IQR	$[0.7, 1.3]$	Med. R	IQR	$[0.7, 1.3]$
Providence (22 nodes)	39,653	0.65	[0.50, 0.83]	39%	0.91	[0.78, 1.04]	81%
Bay Area (41 nodes)	111,382	0.54	[0.42, 0.71]	24%	0.91	[0.77, 1.04]	79%
Los Angeles (7 nodes)	41,476	1.10	[0.89, 1.28]	69%	0.99	[0.89, 1.10]	89%

3.3.5 Recovery of site-level NO₂ variability

The preceding sections show that the gain correction improves agreement with withheld truth, removes the dominant temperature-structured bias, and behaves consistently across nodes and cities. We now turn to its practical consequence at network scale: recovery of the site-level dynamic range that remote linear calibration compresses.

A remote-reference linear calibration assigns a single effective sensitivity to the NO₂ electrode across all training temperatures, optimized to minimize the average residual against a distant reference. That objective can produce reasonable agreement in the mean while attenuating departures from the local site mean in both directions. Because the gain curve is estimated from each node’s own titration events, the per-site correction restores the node-specific dynamic range that the remote linear fit cannot represent.

Figure 3.7 summarizes this effect using three network-wide diagnostics for the Providence deployment, evaluated against each node’s assigned reference and interpreted alongside the Bay Area inter-reference diagnostics (Supplementary Figure S17), which provide a benchmark for the disagreement expected solely from spatial heterogeneity among regulatory-grade stations.

Mean bias (Panel a) is already near zero before correction at most nodes, as expected from the WLS objective function. Because Stage 1 is trained to minimize squared residuals against the assigned reference, small mean bias does not by itself imply accurate site-level calibration. After correction, biases shift modestly at some nodes as the gain correction changes the effective sensitivity and Stage 3 re-anchors the absolute level. The post-correction biases lie within the range of pairwise biases among Bay Area EPA stations (roughly 1–6 ppb at separations of 1–40 km; Supplementary Figure S17).

The variability ratio (Panel b) and OLS slope (Panel c) provide more direct measures of dynamic-range recovery. Before correction, both are suppressed below unity at most nodes, indicating that the baseline calibration compresses the temporal spread of the NO₂ signal. After correction, both shift upward toward unity across the network, with post-correction values falling within the range of inter-reference agreement among Bay Area EPA AirNow stations. This recovery of variance and slope is the clearest network-scale consequence of

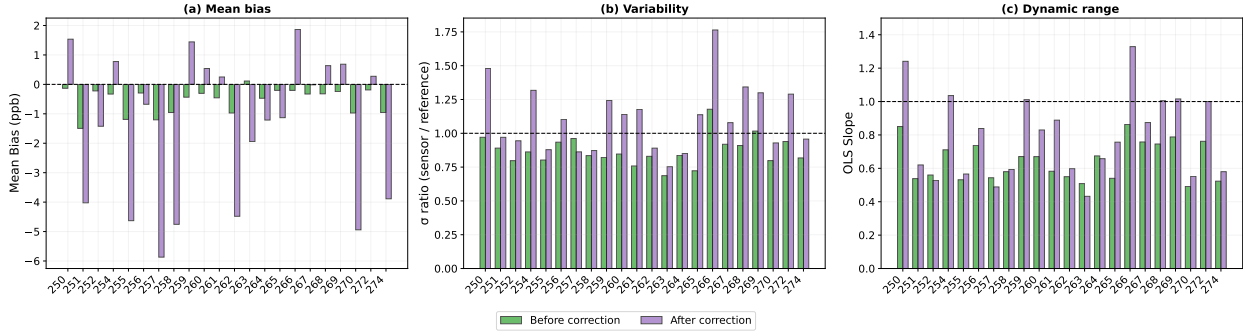


Figure 3.7: Network-wide calibration diagnostics for all 22 Providence nodes, evaluated against each node’s assigned reference. (a) Mean bias (ppb). (b) Variability ratio: standard deviation of the calibrated sensor divided by that of the reference. (c) OLS slope of calibrated sensor against reference. Green: before gain correction; purple: after gain and offset correction. Dashed lines mark zero bias, unit variability ratio, and unit slope.

applying a per-site multiplicative correction informed by each node’s own atmospheric observations.

The variability-ratio diagnostic is conceptually related to moment matching [101], a remote calibration approach that derives slope and offset corrections by matching the mean and standard deviation of the sensor distribution to those of a proxy reference over the same period. That approach is effective for correcting bulk distributional shifts, but it assumes that the true concentration variability at every sensor site matches that of the proxy reference. In a dense urban network, this assumption suppresses exactly the site-to-site differences in NO₂ variability that arise from local source proximity, street geometry, and ventilation. Those differences are central to the dense-network applications that motivate the observations in the first place, including exposure disparity mapping and neighbourhood-scale emission quantification. The titration-based gain correction avoids this limitation because it diagnoses multiplicative error from each node’s own atmospheric signal rather than rescaling the node to match a distant reference distribution. It therefore restores local dynamic range while preserving the heterogeneous variability that dense urban networks are designed to resolve.

3.3.6 Limitations

The analyses above show that the gain correction removes the non-linear temperature-dependent bias in the NO₂ calibration, but they also unveil the current limits of the framework. These limits are most evident under summer conditions, when ambient NO₂ is low and the signal approaches the electrochemical noise floor. More broadly, the remaining uncertainties arise from incomplete warm-temperature constraint on the gain curve, unresolved behaviour of the O₃ electrode and the NO₂ zero level, use of remote hourly references, and departures from ideal titration chemistry.

A first limitation is the sparse warm-temperature coverage of titration events. Nocturnal titration events occur predominantly on cooler nights, so the warm-temperature bins used to construct $g(T)$ are relatively poorly sampled even though the gain error is expected to be largest at high temperature. The hold extrapolation and T_{CAP} bypass reduce the risk of overcorrection in this regime, but they do not fully compensate for the lack of direct observational constraint. This limited warm-temperature sampling likely contributes to the residual summer degradation seen in the withheld-truth validation, together with the smaller summertime NO₂ signal itself.

A second limitation is that the O₃ electrode likely has its own temperature-dependent sensitivity ($g_{\text{ox}}(T) \neq 1$), which cannot be separated cleanly from the NO₂ response using a single NO–O₃ titration constraint. Degradation of the ozone filter on the B43F sensor further complicates this interpretation [43, 46]. As the filter depletes, the NO₂ electrode becomes increasingly sensitive to O₃, which can bias both the WLS decomposition and the inferred titration slope. Winter et al.[46] suggest that this degradation is slow enough to be a minor effect over the deployment periods considered here, but it may become more important over longer deployments.

Related to this, estimation of the NO₂ zero level Z_{NO_2} remains one of the more weakly constrained parts of the pipeline. This component is sensitive to tuning choices (Supplementary Figures S25 and S26). The regression used to estimate Z_{NO_2} relies on hours identified as near-zero NO₂ by the remote reference, and the empirical scaling factor λ is introduced to compensate for the imperfect correspondence between the O_x electrode signal and the actual O₃ contribution to the NO₂ channel. Better characterization of O₃ cross-sensitivity, or direct measurement of ozone-filter attenuation, would help constrain this part of the pipeline more physically.

Additional limitations arise from the use of remote hourly reference data. The rolling offset applied in Stage 3 necessarily conflates true sensor drift with the persistent spatial NO₂ gradient between the sensor node and its assigned remote reference. Without a colocated reference, these two contributions cannot be separated. In addition, because AirNow observations are available only at hourly resolution, the pipeline cannot correct sub-hourly bias structure. More broadly, any remaining interference from reactive species or chemical-state variables that are not measured across the network cannot be represented explicitly in the current model. This is likely part of the reason that summer calibration remains more difficult even after the gain correction.

The framework also assumes that the detected events approximate ideal 1:1 NO–O₃ titration. In practice, incomplete mixing within the sensor footprint, nearby fresh-NO plumes arriving during an event, and residual photochemistry near the edges of the night window can all perturb the observed $\Delta\text{NO}_2/(-\Delta\text{O}_3)$ ratio. The event-quality filters are designed to remove the most contaminated events, but they cannot eliminate all departures from ideal stoichiometry.

The main conclusions are not highly sensitive to reasonable implementation choices despite these limitations. The circular dependency between O₃ and NO₂ calibrations changes $g(T)$ by less than 0.2 units across alternative O₃ references (Supplementary Figure S24), and

one-at-a-time changes in event-detection parameters alter $g(T)$ by less than 0.1 units across all other tested configurations (Supplementary Table S2). These sensitivity tests indicate that, although the framework remains imperfect, its central findings do not depend on a narrow set of tuning choices. With the currently available observations, WLS+GC therefore represents the most complete implementable framework for network-wide NO₂ correction, even though further improvement will likely require additional measurements beyond those currently available at BEACO₂N nodes.

3.4 Conclusions

This study develops a physically constrained calibration framework for electrochemical NO₂ sensors in dense urban monitoring networks. The central result is that the dominant remaining error in non-colocated NO₂ calibration is not reference distance alone, but unresolved temperature-dependent multiplicative gain. Standard linear calibration assigns a single effective sensitivity across the full range of operating conditions and therefore compresses the sensor dynamic range. Because this limitation persists even under colocation, the main obstacle is not simply the lack of a nearby reference monitor, but the inability of a linear model to represent the underlying sensor response.

We address this problem by combining remote WLS regression with a titration-based gain correction that uses each sensor node’s own nocturnal atmospheric observations. During NO–O₃ titration events, the stoichiometric relationship between O₃ loss and NO₂ production provides a physically interpretable diagnostic of gain error. This allows a temperature-dependent gain curve to be estimated independently at each node, after which a rolling temperature-dependent offset is applied to correct the slower residual bias that remains.

Four main conclusions follow from the results. First, the gain correction is a dominant contributor to improved performance under realistic non-colocated conditions. In Bay Area withheld-truth tests, it recovers much of the slope lost under both colocated and remote linear calibration, when other calibration performance metrics (i.e., RMSE, NRMSE, MAE, R², bias) change only modestly. Second, the correction targets the physically relevant error source. The node-specific gain correction removes the strong temperature dependence of the titration ratio and simultaneously collapses its residuals with respect to temperature, humidity, and baseline NO₂, indicating that the dominant residual structure arises from a common underlying sensor process. Third, independently estimated gain curves converge across nodes within Providence and show qualitatively consistent behavior in the Bay Area and Los Angeles, supporting the interpretation that the titration diagnostic captures a reproducible sensor property rather than a purely local atmospheric artefact. Fourth, at network scale, the correction provides a practical scientific benefit beyond improved slope: it restores the site-level dynamic range of NO₂ variability that a remote linear fit cannot fully preserve.

These results show that atmospheric chemical constraints can provide calibration information that cannot be recovered from remote regression alone. In this framework, the sensor node’s own atmospheric observations are not only an evaluation target but also an essential

calibration input. That distinction is especially important for dense urban networks, where most nodes cannot be permanently colocated with reference instruments but where accurate recovery of local variability is necessary to interpret neighbourhood-scale gradients, chemical environments, and exposure contrasts.

The framework also has clear limitations. Warm-temperature conditions remain less well constrained because strong nocturnal titration events are less common on warm nights, even though gain departures are expected to be largest there. The method does not fully separate the temperature-dependent responses of the NO₂ and O₃ electrodes, estimation of the effective NO₂ zero level remains partly empirical, and the rolling offset necessarily combines true sensor drift with persistent node-to-reference spatial gradients. In addition, because the method relies on remote hourly reference data, it cannot resolve all sub-hourly or chemistry-dependent residual errors.

Despite these limitations, the main conclusions are robust to reasonable implementation choices, and the framework substantially improves the scientific usefulness of non-colocated NO₂ measurements. With the currently available observations, WLS+GC provides the most complete implementable approach for network-wide NO₂ correction in BEACO₂N-alike sensor networks. More broadly, this work shows that physically constrained calibration can extend accurate NO₂ measurements beyond permanently colocated sites and provides a scalable path for improving high-resolution urban observations in applications ranging from exposure assessment to emissions evaluation.

Chapter 4

Cross-City Comparison of Transportation and Residential NO_x and CO Emission Factors Inferred from Dense Urban Observations

4.1 Introduction

Urban emissions of air pollutants and greenhouse gases remain difficult to represent accurately. Activity-based inventories must assign emission factors to diverse combustion sources—vehicles, residential heating systems, and industrial activities—on grids of a few to tens of kilometers, often using values derived from laboratory tests, fleet models, or fuel-use statistics that may not reflect local conditions. Urban sources are also highly heterogeneous in space: highways, residential neighbourhoods, and industrial facilities can lie within hundred-meter scales of one another while contributing very different emission signatures. Residential heating is especially variable, spanning modern low- NO_x gas furnaces as well as older oil- and propane-fired systems, with the source mix differing sharply across cities and neighbourhoods. Capturing this fine-scale structure is essential for evaluating mitigation policies and understanding disparities in air pollution exposure. Yet the emission factors embedded in current inventories are rarely tested against observations at the scales where these differences matter most.

Dense urban sensor networks offer a way to verify these emission factors observationally. By measuring co-emitted tracers such as NO_x , CO, and CO_2 at 20 – 50 of fixed locations within a city, these networks capture both the spatial decay of roadway signals and the distributed background from area sources such as residential heating. Because NO_x and CO are predominantly co-emitted with CO_2 during fossil fuel combustion, the surface enhancement ratio $\Delta X/\Delta \text{CO}_2$ reflects the effective emission factor of the local source mix, with atmospheric dilution affecting numerator and denominator similarly. This approach has been

used at individual monitoring stations, near-road transects, and satellite pixels, but these applications usually sample only a limited range of source types at one location and cannot resolve how enhancement ratios vary across source sectors within an urban area [102, 39, 65]. Dense networks extend this framework by resolving both the near-roadway signal and the broader urban background across space.

The Berkeley Environmental Air-quality and CO₂ Network (BEACO₂N) has operated in the San Francisco Bay Area at roughly 2 km spacing since 2012 [35], with newer deployments in Los Angeles and Providence since 2022. Its combination of CO₂ and air-quality measurements at neighbourhood scale has supported a growing body of observational emissions research, including city-scale CO₂ trend detection and sectoral attribution [49, 52, 50, 103], point-source emissions quantification [104, 105], and characterisation of residential heating behavior across cities [106]. More recently, enhancement-ratio analyses showed that $\Delta\text{NO}_x/\Delta\text{CO}_2$ and $\Delta\text{CO}/\Delta\text{CO}_2$ decay approximately exponentially with distance from major highways, allowing sector-specific emission factors to be inferred from the fitted near-highway amplitude and far-field plateau [107]. In that framework, the near-highway limit represents transportation sources, whereas the far-field plateau represents distributed residential combustion after direct roadway influence has mixed away.

That framework has so far been demonstrated in only one city. In the San Francisco Bay Area, transportation $\Delta\text{NO}_x/\Delta\text{CO}_2$ and $\Delta\text{CO}/\Delta\text{CO}_2$ agree well with on-road inventory values, whereas the residential NO_x emission factor is substantially lower than the inventory [107]. This result suggested that inventories may overestimate the NO_x intensity of residential combustion, but it remained unclear whether the discrepancy was specific to the Bay Area or reflected a more general feature of residential-emission parameterisation. The portability of the ER-versus-distance framework to cities with different climates, fleet compositions, and fuel mixes has also not been tested. A recently developed physically constrained calibration framework for electrochemical NO₂ sensors resolves nonlinear temperature-dependent gain errors and enables consistent NO_x comparisons across multiple deployments (see Chapter 3), making such a cross-city test possible.

Here we extend the observational emission-factor framework to a three-city comparison spanning the San Francisco Bay Area, Los Angeles, and Providence, Rhode Island, using BEACO₂N deployments in all three cities. These cities provide a useful experimental contrast. The two California cities are both dominated by natural gas in the residential sector, but differ substantially in urban layout, traffic characteristics, and the spatial configuration of the sensor network. Providence, by contrast, has a colder winter climate, stronger heating demand, and a more diverse residential fuel mix that includes appreciable fractions of fuel oil and liquefied petroleum gas in addition to natural gas. Applying the same ER-versus-distance framework across all three cities makes it possible to distinguish patterns that are reproducible across gas-dominated cities from those that vary with fuel mix, climate, or deployment geometry.

This study evaluates the consistency of observationally inferred transportation and residential emission factors across the three cities and examines the factors that shape cross-city variation. It also compares the inferred emission factors with a gridded inventory—here

GRA²PES v1.0 [12]—to assess the character of observation–inventory disagreement. The two species provide complementary leverage in that comparison. $\Delta\text{NO}_x/\Delta\text{CO}_2$ is more closely linked to fuel type, combustion conditions, and source composition, making it especially informative about whether the inventory represents the local source mix correctly. By contrast, $\Delta\text{CO}/\Delta\text{CO}_2$ reflects a broader set of combustion conditions and source contributions, making it useful for evaluating how inventories partition emissions across sectors. Together, these two ratios provide a framework for distinguishing mismatches related mainly to source composition from those related more to sector allocation.

4.2 Measurements and Methods

4.2.1 Networks and observations

This analysis uses observations from the Berkeley Environmental Air-quality and CO₂ Network (BEACO₂N) in three U.S. metropolitan regions: the San Francisco Bay Area, Los Angeles, and Providence, Rhode Island. Each network uses Alphasense B-type electrochemical sensors for NO, NO₂, and CO, together with Vaisala CarboCap nondispersive infrared CO₂ sensors. Nodes are mounted primarily on one-story school rooftops at $\sim 1\text{--}3$ km spacing [35, 44, 45]. Hardware, firmware, and primary data-processing pipelines are common across all three deployments. NO_x is defined here as the sum of NO and NO₂, computed from the independently calibrated NO and NO₂ channels at each node.

NO₂ is calibrated using the physically constrained three-stage framework described in Chapter 3. NO, CO, and CO₂ are calibrated using the rolling multivariate linear regression framework described by Winter et al [45]. Wind data used for highway-distance calculations and wind-direction filtering are obtained from the Synoptic Data MesoWest cooperative mesonet [73], averaged to hourly resolution, and matched to the nearest BEACO₂N node.

Figure 4.1 shows sensor locations and nearby major highways for the three deployments. Approximately 25 of Bay Area nodes (2022), 8 of Los Angeles nodes (2023), and 20 of Providence nodes (2023) provided data passed the data-completeness and regression-quality filters described in Section 4.2.2.

4.2.2 Enhancement ratio computation and distance-decay fitting

Enhancement ratios and their distance-decay relationships are computed following Zhu et al. [107], extended here from the Bay Area to Los Angeles and Providence. ER_{NO_x} and ER_{CO} denote the measured enhancement-ratio profiles ($\Delta\text{NO}_x/\Delta\text{CO}_2$ and $\Delta\text{CO}/\Delta\text{CO}_2$ as functions of distance), whereas EF_{NO_x} and EF_{CO} denote the emission factors inferred from those profiles or extracted from GRA²PES.

For each sensor node, hourly enhancements ΔX and ΔCO_2 , where X is NO_x or CO, are defined relative to a rolling network background taken as the 5th percentile of network-median concentrations within a 4-week moving window. Hourly ΔX – ΔCO_2 pairs are then

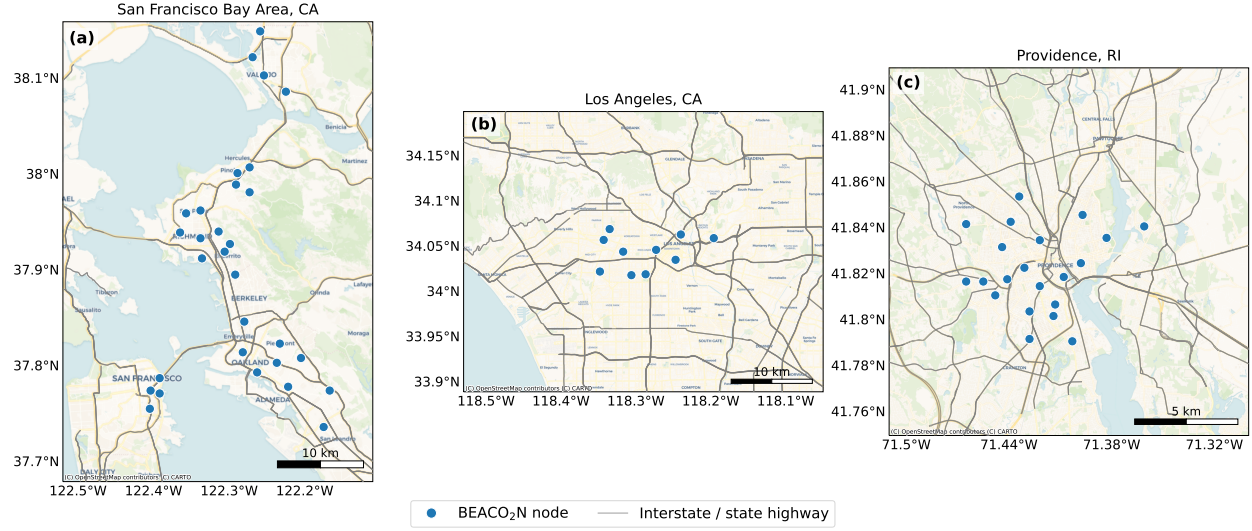


Figure 4.1: Sensor node locations for the three BEACO₂N deployments: (a) San Francisco Bay Area, (b) Los Angeles, (c) Providence. Major highways (Interstate and state routes) are shown as grey lines. Filled blue circles indicate where the sensor nodes are installed.

selected within a seasonal window, a time-of-day window, and a wind-direction window oriented toward the nearest major highway. The seasonal windows are summer (June, July, August, September, JJAS) and winter (December, January, February, DJF). The time-of-day windows are morning rush hour (06–09 local) and late night (00–03 local). City-specific wind-direction windows and other fitting settings are listed in Supplementary Table S1.

Within each pooled subset, the node-level enhancement ratio (ER) is defined as the bivariate least-squares slope of ΔX against ΔCO_2 in units of ppbv/ppmv [62]. Following Zhu et al. [107], nodes with regression $R^2 < 0.5$ are excluded from the subsequent distance-decay analysis.

The resulting node-level ERs are then arranged by each node’s distance to the nearest highway along the local hourly wind vector and fitted with

$$\text{ER}(d) = a \cdot \exp(-d/k) + c, \quad (4.1)$$

where a is the near-highway amplitude, k is the e-folding mixing length, and c is the far-field plateau. In this framework, the transportation emission factor is defined as the extrapolated near-highway value, $a + c$, while the residential emission factor is defined as the far-field plateau, c .

Uncertainties in the fitted parameters (a , k , and c) and in the derived emission factors are reported as ± 1 standard error from the nonlinear least-squares fit, based on the parameter covariance matrix. For the transportation emission factor, $a + c$, the propagated variance is

$$\text{Var}(a + c) = \text{Var}(a) + \text{Var}(c) + 2 \text{Cov}(a, c).$$

Per-city analysis settings, including seasonal month ranges, time-of-day windows, wind-direction windows, bivariate regression error scaling, and the set of major highways used in the distance calculation, are listed in Supplementary Table S1.

4.2.3 Comparison with GRA²PES inventory

Observed emission factors are compared against the U.S. Greenhouse Gas and Air Pollutant Emissions System (GRA²PES) v1.0 [12], a 4 km hourly gridded inventory with sector-resolved emissions and weekday–weekend and diurnal temporal profiles.

For each city, inventory emissions are extracted over the geographic footprint of the sensor network using a city-specific bounding box. A land mask based on the residential CO₂ emission field is applied to exclude unpopulated grid cells. For transportation, three on-road sector groupings are considered: on-road gasoline alone (ONROAD_GAS), on-road diesel alone (ONROAD_DSL), and the combined on-road total. Residential combustion is represented by the RES sector. For residential-sector comparisons, the lowest 5th percentile of CO₂ pixel-level values within each network footprint is excluded, removing about 3% of the data.

Inventory emissions are grouped by season, local time window, and sector using the same seasonal and time-of-day definitions applied to the observations. For each city, season, time-of-day window, and sector, NO_x, CO, and CO₂ emissions are summed across the masked network footprint and across the three months within that season before taking the ratio. The resulting seasonal emission ratio, computed in mol/mol, is converted to ppbv/ppmv by multiplication by 10³.

For comparison with the observationally inferred emission factors, the reported GRA²PES values are the pixel-median emission factors within each network footprint for the selected season and time-of-day window, reported separately for the transportation and residential sectors. Per-pixel spatial-variability statistics for the inventory fields are listed in Supplementary Table S3.

4.3 Results and Discussion

The results are organised in three parts. Section 4.3.1 examines whether near-highway gradients resolve transportation emission factors in each city and compares the inferred values with GRA²PES on-road sub-sectors. Section 4.3.2 turns to the far-field plateau under winter midnight conditions to isolate residential emission factors, estimates oil-heating emission factors from the cross-city contrast in fuel mix, and evaluates the inventory comparison for both species, including the role of off-road sources in explaining the different behaviour of NO_x and CO. Section 4.3.3 discusses the main uncertainties and limitations of the analysis.

4.3.1 Transportation emission factors from near-highway gradients

Figure 4.2 shows a clear decline in ER with distance from the nearest highway in two cases: the Bay Area during summer morning rush hour and Providence during winter morning rush hour. In both cities, the observations are consistent with the exponential form in Equation 4.1, indicating that the ER-versus-distance framework resolves a distinct near-highway enhancement against a broader urban background. The Bay Area profile drops rapidly and approaches a stable background within about 2 km, whereas the Providence profile decays more gradually over a longer distance range. The fitted parameters in Table 4.1 quantify this contrast. The Bay Area e-folding lengths are 186 ± 93 m for ER_{NO_x} and 200 ± 140 m for ER_{CO} , whereas Providence yields 444 ± 460 m and 524 ± 480 m, respectively. The Bay Area k values are reasonably well constrained, with fractional uncertainties of approximately 50–70%, whereas the Providence k values carry uncertainties comparable to the estimates themselves, reflecting the small number of near-highway nodes available to constrain the steep part of the decay. These k uncertainties do not enter the definition of $EF_{\text{transportation}}$ directly, because the transportation EF is defined as $a + c$. However, poor constraint on k can still broaden the uncertainty in the fitted values of a and c through parameter covariance, and therefore reduces the precision with which the decay structure can be compared across cities. Within each city, the e-folding lengths for NO_x and CO are similar (Bay Area: 186 m vs. 200 m; Providence: 444 m vs. 524 m). This similarity is expected because the decay is governed by physical turbulent mixing rather than chemical loss at these spatial scales. At typical wind speeds of 0.5–2 m/s, transport from the highway to 2 km downwind takes roughly 15–60 minutes, far shorter than the NO_x photochemical lifetime (approximately 5 hours at midday in summer); CO is effectively inert on these timescales. The modest difference that does exist— k is slightly shorter for NO_x in both cities—likely reflects differences in the spatial distribution of sources. Off-road equipment and other distributed combustion sources contribute a larger fraction of urban CO than of urban NO_x , which may smooth the highway-to-background transition over a slightly longer distance for CO.

Providence also has a higher NO_x far-field level ($c = 0.50$ versus 0.35 ppbv/ppmv), consistent with a stronger distributed winter background, likely reflecting residential heating. The shorter Bay Area decay length is consistent with stronger summer daytime mixing, whereas the longer Providence decay length is consistent with more stable winter conditions that allow roadway signals to persist over a broader horizontal footprint. Because the transportation EF is defined as $a + c$, the near-road signal is recovered relative to a city- and season-specific background rather than a common baseline.

The Bay Area transportation emission factors, $EF_{NO_x} = 1.21 \pm 0.23$ ppbv/ppmv and $EF_{CO} = 10.68 \pm 2.45$ ppbv/ppmv, are modestly higher than the 1.0 ± 0.02 and 7.8 ± 0.6 ppbv/ppmv reported by Zhu et al. [107] for the same network in 2022. This shift likely reflects a combination of methodological differences. First, Zhu et al. [107] reported uncertainty on a alone, whereas the present estimates propagate uncertainty through both a and c , including their

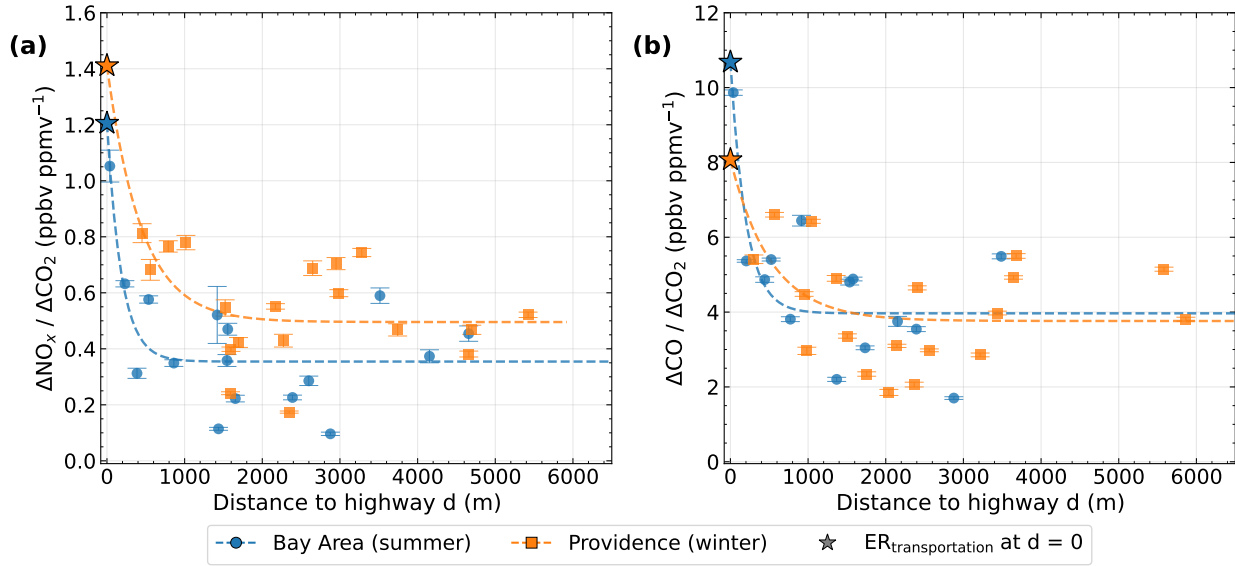


Figure 4.2: Enhancement ratios as a function of distance to the nearest highway for the Bay Area summer morning rush hour and Providence winter morning rush hour: (a) $\text{ER}_{\text{NO}_x} = \Delta\text{NO}_x / \Delta\text{CO}_2$ and (b) $\text{ER}_{\text{CO}} = \Delta\text{CO} / \Delta\text{CO}_2$. Points show per-node ER values from bivariate least-squares regression of ΔX versus ΔCO_2 hourly pairs within the respective seasonal, wind-direction, and time-of-day windows. Dashed curves are least-squares fits to Equation 4.1. Stars mark $\text{EF}_{\text{transportation}} = a + c$ obtained by extrapolating the fit to $d = 0$. Per-node error bars are bivariate regression uncertainties. Fit parameters are listed in Table 4.1. Per-city analysis settings are listed in Supplementary Table S1; full ER-versus-distance panels for all profiles are shown in Supplementary Figures S1–S3.

Table 4.1: Fitted parameters of Equation 4.1 for the near-highway decay cases shown in Figure 4.2. a : near-highway amplitude; k : e-folding mixing length; c : far-field plateau; $\text{EF}_{\text{transportation}}$ ($= a + c$): transportation emission factor at $d = 0$. All emission-related quantities are in ppbv/ppmv; k is in meters.

City	Species	a (ppbv/ppmv)	k (m)	c (ppbv/ppmv)	$\text{EF}_{\text{transportation}}$ (ppbv/ppmv)
SF Bay Area	ER_{NO_x}	0.85 ± 0.23	186 ± 93	0.35 ± 0.04	1.21 ± 0.23
	ER_{CO}	6.71 ± 2.40	200 ± 140	3.97 ± 0.46	10.68 ± 2.45
Providence	ER_{NO_x}	0.92 ± 1.10	444 ± 460	0.50 ± 0.05	1.41 ± 1.17
	ER_{CO}	4.31 ± 3.60	524 ± 480	3.76 ± 0.38	8.07 ± 3.72

covariance. Second, plausible alternatives in wind-direction and time-window selection can shift the inferred EF by up to $\sim 20\%$. Third, and most importantly for NO_x , the gain correction restores near-highway signal amplitude that was partly suppressed under the earlier calibration. Overall, agreement at the level of approximately 20% supports the robustness of the Bay Area estimate, while also indicating that transportation EFs inferred from the distance-decay framework should not be overinterpreted more precisely than that.

Los Angeles does not yield a resolvable transportation emission factor because the network lacks sites sufficiently close to major highways to constrain the near-highway amplitude. In that limit, Equation 4.1 becomes weakly constrained, and the network-median ER is better interpreted as a far-field quantity (Supplementary Figure S3 and Table S1). Rather than contradicting the framework, this outcome highlights a practical deployment requirement: recovering $\text{EF}_{\text{transportation}}$ observationally requires at least some sites within roughly 100 m of major roadways.

Table 4.2: Comparison of observed transportation emission factors with GRA^2PES on-road sub-sector values. Observed EFs are the extrapolated near-highway value $a + c$: Bay Area from summer morning rush hour (JJA, 06–09 local), Providence from winter morning rush hour (DJF, 05–09 local). Los Angeles is omitted because the network geometry does not resolve a transportation EF. GRA^2PES columns show on-road gasoline, on-road diesel, and the combined on-road total, extracted over each network footprint for the corresponding season and time-of-day window. Full sub-sectoral breakdown in Supplementary Table S2. All values in ppbv/ppmv.

City	Species	BEACO ₂ N	GRA ² PES		
		Observed	Gasoline	Diesel	Gasoline+Diesel
SF Bay Area	EF _{NO_x}	1.21 ± 0.23	0.64	2.77	1.10
	EF _{CO}	10.68 ± 2.45	8.14	1.83	6.78
Providence	EF _{NO_x}	1.41 ± 1.17	0.79	2.77	1.63
	EF _{CO}	8.07 ± 3.72	9.90	1.83	6.47

Table 4.2 compares the observational transportation EFs with GRA^2PES on-road values. For EF_{NO_x}, the observed values fall between the inventory gasoline and diesel sub-sector values in both cities. In the Bay Area, the observed value (1.21 ppbv/ppmv) is close to the combined on-road total (1.10 ppbv/ppmv). In Providence, the observed value (1.41 ppbv/ppmv) falls below the winter combined total (1.63 ppbv/ppmv) but remains within the range bounded by the gasoline and diesel values. This pattern is consistent with the Providence near-highway fleet being somewhat more gasoline-dominated than the footprint-average fleet composition in GRA^2PES , although the large uncertainty (± 1.17 ppbv/ppmv) means that this inference should be viewed as suggestive rather than definitive. Overall, the NO_x compar-

ison indicates that the inferred transportation signal is broadly consistent with the on-road inventory structure.

The CO comparison is less uniform. In the Bay Area, the estimate (10.68 ppbv/ppmv) exceeds both the gasoline sub-sector (8.14 ppbv/ppmv) and the combined on-road total (6.78 ppbv/ppmv), although the gasoline value still lies within the 1σ range of the observation. Several explanations are plausible. One is that sources assigned to other inventory sectors, particularly off-road emissions, contribute CO in near-highway environments and not in residential environments; the monthly sectoral composition in Supplementary Figure S7 shows that off-road sources represent a substantial fraction of summer CO emissions in all three cities. Cold-start and idling emissions may also contribute [108]. In Providence, by contrast, the observed value (8.07 ± 3.72 ppbv/ppmv) falls below the winter gasoline sub-sector (9.90 ppbv/ppmv) and within the gasoline–diesel range, remaining broadly consistent with the on-road inventory.

4.3.2 Residential emission factors from the far-field plateau

The transportation analysis above identifies the spatial scale over which direct roadway influence decays. Beyond about 2 km from the nearest highway, the resolved Bay Area and Providence profiles are already close to their background level, so the remaining ER can be interpreted primarily as a distributed urban signal. Following Zhu et al. [107], this stable far-field plateau is identified with the residential emission factor. Figure 4.3 shows winter midnight ER_{NO_x} and ER_{CO} for all three cities; per-city fit parameters are listed in Supplementary Table S4.

For both species, the Bay Area and Los Angeles yield similar values ($EF_{NO_x} \approx 0.27$ – 0.30 ppbv/ppmv and $EF_{CO} \approx 2.5$ – 2.9 ppbv/ppmv), whereas Providence is higher ($EF_{NO_x} = 0.45 \pm 0.06$ ppbv/ppmv and $EF_{CO} = 3.66 \pm 0.39$ ppbv/ppmv). The close agreement between the two California cities, despite their different urban layouts and network geometries, suggests that the residential signal is reasonably reproducible when the dominant combustion mix is similar.

A plausible explanation for the higher Providence values is fuel composition. The Bay Area and Los Angeles are overwhelmingly natural-gas-dominated, whereas Providence includes appreciable fractions of heating oil (15%) and liquefied petroleum (3%, LP) (Supplementary Figure S4). Oil- and LP-fired systems generally emit more NO_x or CO per unit CO_2 than modern gas appliances, and the same city ordering in $\Delta CO/\Delta CO_2$ is consistent with a larger contribution from less efficient combustion in Providence. The parallel behaviour across both species therefore supports a source-based interpretation, although it does not prove a unique mechanism. The GRA²PES monthly sectoral composition (Supplementary Figure S6) also shows residential combustion as a larger fraction of winter NO_x in Providence than in the two California cities, consistent with the stronger observed residential EF_{NO_x} there.

The cross-city contrast also permits a rough estimate of the emission factor associated with non-gas heating in Providence. Because electric heating does not produce combus-

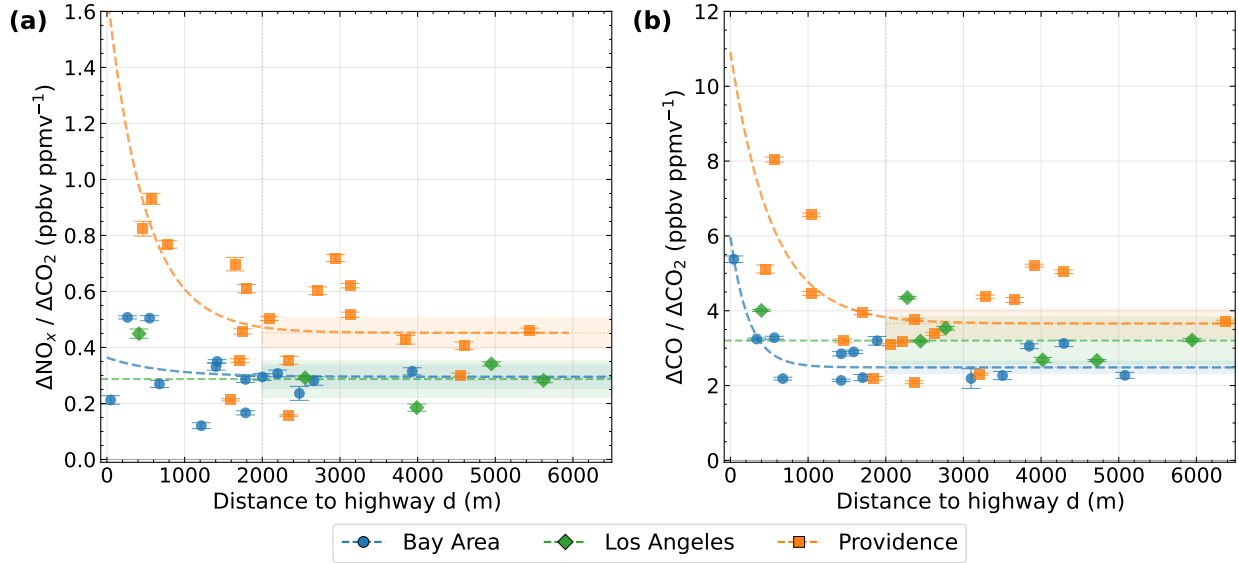


Figure 4.3: Winter midnight enhancement ratios as a function of distance to the nearest highway for the three BEACO₂N cities: (a) $\text{ER}_{\text{NO}_x} = \Delta\text{NO}_x / \Delta\text{CO}_2$ and (b) $\text{ER}_{\text{CO}} = \Delta\text{CO} / \Delta\text{CO}_2$. Each point represents one sensor node from its city’s winter midnight pool. Dashed horizontal lines indicate the fitted far-field plateau c for each city, with shaded bands showing the fit uncertainty. The vertical dotted line at 2 km marks the threshold beyond which near-highway contribution is negligible given the k values recovered in Figure 4.2. The plateau c is identified with $\text{EF}_{\text{residential}}$ following Zhu et al. [107].

tion emissions within the network domain, only fuel-burning households contribute to the observed $\Delta X / \Delta\text{CO}_2$. Excluding electric heating from the Providence fuel mix and rounding the census shares gives approximately 80% natural gas and 20% oil+LP (Supplementary Figure S4). Taking the California observations as an idealized natural-gas-only case ($\text{EF}_{\text{CO,gas}} \approx 2.5$ ppbv/ppmv and $\text{EF}_{\text{NO}_x,\text{gas}} \approx 0.28$ ppbv/ppmv), a simple two-component mixing model gives

$$\text{EF}_{X,\text{oil+LP}} = \frac{\text{EF}_{X,\text{observed}} - 0.80 \text{EF}_{X,\text{gas}}}{0.20}, \quad (4.2)$$

which yields $\text{EF}_{\text{CO,oil+LP}} \approx 8$ ppbv/ppmv and $\text{EF}_{\text{NO}_x,\text{oil+LP}} \approx 1.1$ ppbv/ppmv, roughly three to four times the corresponding gas values. These are order-of-magnitude estimates: housing-unit fractions are used as proxies for CO₂-weighted combustion shares, LP is grouped with oil, and the California observations are treated as a pure natural gas EF despite small non-gas fractions. Even so, the two species yield internally consistent results, and the implied values are physically plausible given that oil-fired systems typically operate at higher flame temperatures and with less emission control than modern gas appliances.

Table 4.3 compares the observed residential EFs with GRA²PES using the pixel-median inventory value within each network footprint. The median is used because residential CO

in GRA²PES is spatially heavy-tailed, so the footprint mean can be dominated by a small number of high-EF pixels.

Table 4.3: Comparison of observed residential emission factors with GRA²PES residential-sector values. Observed values are the far-field plateau c from the winter midnight ER profiles (Figure 4.3). GRA²PES values are the pixel-median EF within each network footprint (winter DJF, 00–03 local window); the median is less sensitive to the heavy-tailed spatial distribution of residential CO than the pixel-mean aggregate (see Supplementary Table S3 for per-pixel statistics). All values in ppbv/ppmv.

City	Species	BEACO ₂ N	GRA ² PES
		Observed	Residential
SF Bay Area	EF _{NO_x}	0.30 ± 0.04	0.76
	EF _{CO}	2.49 ± 0.16	6.37
Los Angeles	EF _{NO_x}	0.27 ± 0.05	0.60
	EF _{CO}	2.93 ± 0.82	2.50
Providence	EF _{NO_x}	0.45 ± 0.06	1.00
	EF _{CO}	3.66 ± 0.39	4.28

For EF_{NO_x}, the inventory preserves the same directional contrast seen in the observations — California cities lower, Providence higher — but every inventory value exceeds the corresponding observation by roughly a factor of two to three. Consistent with the previous study done for Bay Area [107], this pattern suggests that the inventory assigns residential-combustion NO_x emission factors that are systematically higher than those implied by the network observations. The agreement in cross-city ordering indicates that the inventory captures part of the underlying source contrast, while the magnitude mismatch points to a bias in how residential NO_x intensity is parameterized. Whether that bias reflects overestimated combustion temperatures, under-representation of modern low-NO_x appliances, or some combination of the two cannot be determined from the ER observations alone.

For EF_{CO}, the disagreement is not only in magnitude but also in ordering. The observed relationship among cities — Bay Area < Los Angeles < Providence — is not reproduced by the pixel-median inventory values (Los Angeles 2.50, Providence 4.28, Bay Area 6.37 ppbv/ppmv). In GRA²PES, residential CO/CO₂ is strongly heterogeneous within each city, with winter pixel CoV of 0.71 in the Bay Area, 0.52 in Providence, and 0.44 in Los Angeles (Supplementary Table S3), visible in the per-pixel maps (Supplementary Figures S21 and S22). No comparable heterogeneity exists for residential NO_x (CoV ≤ 0.06). The full set of per-pixel maps (Supplementary Figures S15–S26) illustrates this contrast: the on-road panels are spatially uniform, whereas the residential CO panels show pronounced sub-grid texture.

This difference between NO_x and CO can be understood in the context of the sectoral structure of each species' emission budget. In GRA²PES, CO_2 and NO_x are dominated by gasoline vehicles, diesel vehicles, and residential combustion (Supplementary Figures S5 and S6) — the same three sectors that the two-sector ER framework is designed to separate. CO has a different sectoral budget: off-road sources — construction equipment, lawn and garden machinery, generators, and other diffuse mobile sources — account for roughly 30–50% of total CO emissions in all three cities (Supplementary Figure S7). These sources are spatially diffuse, temporally variable, and not cleanly attributable to either the highway or the residential sector in the two-sector framework. Near highways, off-road CO that correlates spatially with traffic would inflate the apparent transportation EF_{CO} ; in the far field, it would contribute to the plateau alongside residential combustion. Because off-road activity patterns likely differ across cities, this unresolved third sector may also contribute to the cross-city CO ordering disagreement. This is a plausible interpretation consistent with the inventory sectoral budgets, but the ER observations alone cannot directly attribute the CO discrepancies to off-road sources. What the data do establish is that the two-sector decomposition is structurally better suited to the NO_x emission budget than to the CO budget, and that CO-based comparisons should be interpreted with that limitation in mind.

4.3.3 Limitations and outlook

Several caveats should apply. The fitted transportation EFs carry propagated uncertainties of roughly 20–50% (Table 4.1), and the Providence values in particular have standard errors comparable to the central estimates. The residential EFs are more precisely determined (± 10 –25%), but even for these the inferred values can shift by up to $\sim 20\%$ with plausible changes to the wind-direction window, seasonal pooling, or R^2 threshold (Supplementary Table S1). The observation–inventory comparisons should therefore be read as indicating the direction and approximate magnitude of agreement or disagreement, not as precise quantitative benchmarks.

The inventory comparison uses GRA²PES v1.0 for 2021, whereas the observational analysis spans later periods; this temporal mismatch is likely more important for transportation than for residential emissions because fleet composition is changing more rapidly than building stock. The Los Angeles deployment cannot constrain a transportation EF because it lacks near-highway sites, so the three-city comparison remains asymmetric across sectors. The interpretation of $\text{EF}_{\text{residential}}$ as the far-field plateau assumes that direct highway influence is negligible beyond 2 km; this is well supported by the resolved Bay Area and Providence decay lengths but cannot be verified independently in Los Angeles. Finally, the non-gas heating emission factors derived from the two-component mixing model (Equation 4.2) rest on several simplifying assumptions and should be treated as order-of-magnitude estimates.

These limitations also point to several directions for future work. A near-highway deployment in the sensor network would enable the transportation EF analysis to be extended to any other cities on equal footing. Additional information on building-level fuel use or appliance characteristics would help move the home heating EF estimate beyond the present

two-component approximation. More generally, the CO comparison suggests that future ER-based analyses should consider extending the current two-sector framework to include off-road or other diffuse mobile sources explicitly, particularly in cities where those sources make a substantial contribution to the urban CO budget.

4.4 Conclusions

This study extends the BEACO₂N enhancement-ratio framework from a single-city demonstration to a three-city comparison across the San Francisco Bay Area, Los Angeles, and Providence. By combining dense observations of NO_x, CO, and CO₂ with the distance-to-highway analysis, the framework separates a near-highway signal associated primarily with on-road transportation from a far-field signal associated primarily with distributed residential combustion. The cross-city design allows us to test both whether the inferred emission factors are reproducible across cities and whether observation–inventory disagreements are primarily local or structural.

Where the network includes sites within a few hundred metres of major roadways, the ER-versus-distance relationship resolves transportation emission factors with physically interpretable decay lengths. Both the Bay Area and Providence yield clear near-highway decay, and the inferred transportation EF_{NO_x} falls between the inventory gasoline and diesel values in both cities. Los Angeles does not resolve a transportation signal because the network lacks near-highway sites. This contrast highlights a practical design requirement: dense spatial coverage alone is not sufficient; the network must also sample the sharp gradients close to major roads.

The far-field winter midnight plateau provides the clearest cross-city constraint on residential combustion. The Bay Area and Los Angeles, both dominated by natural gas, yield similar residential EF_{NO_x} (~ 0.3 ppbv/ppmv) and EF_{CO} (~ 2.5 – 2.9 ppbv/ppmv), whereas Providence is higher in both species (0.45 and 3.66 ppbv/ppmv). This pattern is consistent with fuel composition as a primary driver: Providence has appreciable fractions of fuel oil and liquefied petroleum in its residential heating mix. A two-component mixing model using the California cities as a gas reference yields approximate oil-heating emission factors of ~ 8 ppbv/ppmv for CO and ~ 1.1 ppbv/ppmv for NO_x — roughly three to four times the gas values — providing initial observational constraints on a fuel type for which in-situ residential emission data remain scarce.

Comparison with GRA²PES shows that the strongest and most systematic observation–inventory mismatch lies in the residential sector. Residential EF_{NO_x} is lower than the inventory by a factor of two to three in all three cities, even though the inventory reproduces the broad California-low / Providence-high ordering. This indicates that the cross-city contrast in residential fuel mix is represented in the right direction, but that the NO_x intensity of residential combustion is systematically too high in the inventory, extending the earlier Bay Area finding of [107]. CO presents a more complex picture. The Bay Area transportation EF_{CO} exceeds the on-road inventory central estimate, and the observed cross-city pattern in

residential CO is not reproduced by the inventory. This difference likely reflects the fact that the CO emission budget is structurally less compatible with the present two-sector framework: off-road sources account for a large fraction of urban CO emissions but are not cleanly separated from either roadway or residential combustion in the ER analysis.

These results show that inventory uncertainty is larger for residential combustion than for transportation, providing useful guidance for future inventory improvement. In the near term, this points to the need for better characterization of building-sector emission factors, especially for NO_x , and for more explicit treatment of non-road CO sources that complicate interpretation of urban CO signals. In the longer term, these observational constraints will become increasingly important as urban energy systems change. Widespread transportation electrification is likely to reduce the relative importance of on-road combustion, while the building sector will increasingly reflect a mix of legacy combustion sources and electric technologies such as heat pumps.

Dense urban sensor networks are well suited to follow this transition. By providing observation-based constraints on how emission factors vary with fuel mix, technology adoption, and city structure, they can help inventories remain responsive to the changing composition of urban emissions rather than relying on static sectoral assumptions. The ER-versus-distance framework developed here can, in principle, be applied to multi-year records to track how sector-specific emission factors evolve over time. The calibration framework developed in Chapter 3 provides the consistency needed for such comparisons, and the BEACO₂N network is well positioned to support long-term observational evaluation as urban combustion sources continue to change. Because individual emission-factor estimates carry uncertainties of roughly 20% for the transportation sector and slightly less for the residential sector, the framework is unlikely to resolve small year-to-year changes. However, cumulative changes over 5–10 year timescales should be detectable if calibration, site selection, and background definitions remain consistent through time.

Chapter 5

Concluding Remarks

This dissertation asked whether dense urban sensor network observations can provide robust, observation-based constraints on transportation and residential emission factors for NO_x and CO, and whether those constraints generalize across cities with different climates, vehicle fleets, and residential fuel mixes. Across three chapters, the results show that dense urban observations can indeed recover sector-specific emission factors in a physically interpretable way, provided that the measurement and calibration framework is matched to the spatial and chemical structure of the problem. The enhancement-ratio-versus-distance framework yields reproducible constraints on transportation and residential combustion where network geometry is sufficient, the physically constrained calibration framework developed in Chapter 3 enables internally consistent cross-city NO_x comparison, and the resulting three-city analysis reveals systematic, species-dependent patterns of agreement and disagreement with current emission inventories.

This broader picture emerges when the results of the three chapters are considered together. The residential NO_x emission factor discrepancy between BEACO₂N observation and emission inventory first identified in the Bay Area appears unlikely to be only a local anomaly. A similar discrepancy is found in Los Angeles, and Providence shows the same tendency at a higher absolute level that tracks the city's larger oil- and LP-heating fractions. This cross-city consistency, made possible by the gain-corrected NO_2 calibration developed in Chapter 3, suggests that inventories systematically overestimate the NO_x intensity of residential combustion, especially in natural-gas-dominated cities. CO provides a useful complement, but a less uniform one. In the Bay Area, the near-highway CO signal exceeds the on-road gasoline inventory estimate, whereas in Providence the transportation CO result is broadly consistent with the gasoline–diesel range. Residential CO also shows cross-city ordering differences that NO_x does not. Together, these results suggest that inventories can capture broad sectoral contrasts while misrepresenting the source partitioning underneath.

The dissertation also makes two methodological contributions that extend beyond the specific cities studied. First, the enhancement-ratio-versus-distance framework provides a simple and physically grounded way to infer sector-resolved emission factors from routine network observations, without requiring a full atmospheric transport model run. Its main ob-

servational requirement is that the deployment span a clear highway-to-residential gradient. Second, the titration-based gain correction addresses a calibration limitation that appears to be structural to the Alphasense NO₂B43F electrochemical sensor. The gain curves converge across nodes and cities without city-specific retuning, and the method improves recovery of NO₂ variability under both under- and over-responsive reference configurations. Together, these contributions make dense sensor network observations more suitable for quantitative emission-factor inference.

Several limitations constrain the present results. The transportation framework depends on network geometry: without sites sufficiently close to major highways, as in Los Angeles, the near-highway amplitude cannot be resolved robustly. The warm-temperature end of the gain curve remains less well constrained because nocturnal titration events are sparse on warm nights, even though gain departures are expected to be largest there. The present two-sector decomposition into transportation and residential combustion also does not resolve industrial, commercial, or off-road sources explicitly. That limitation is likely especially important for CO, where off-road emissions appear to contribute substantially to the urban budget and to complicate comparison with the inventory. Finally, temporal mismatch between the GRA²PES v1.0 base year and the observation periods introduces additional uncertainty, particularly for transportation, where fleet composition is evolving more rapidly than building stock.

Looking ahead, the framework developed here points toward a broader role for dense urban observations in emissions science. As cities decarbonize through vehicle electrification, building electrification, and shifts in fuel use, the atmospheric signatures of transportation and residential combustion will also change, and those changes will not necessarily be captured promptly by activity-based inventories alone. Dense sensor networks provide a way to track these transitions in near real time, revealing not only whether total emissions are declining, but also how the underlying source mix is evolving across neighbourhoods and seasons. Future work that combines dense ground-based observations with improved sectoral representation, higher-resolution inventories, and complementary platforms such as mobile monitoring, towers, and satellites could extend this framework beyond transportation and residential combustion to a wider set of urban sources. In that broader context, the central contribution of this dissertation is to show that dense urban observing networks can do more than map pollution: when paired with physically grounded calibration and interpretable analytical frameworks, they can provide a scalable observational foundation for evaluating urban emissions, testing inventory assumptions, and supporting more responsive air quality and climate policy.

Bibliography

- [1] S. Angel et al. “The dimensions of global urban expansion: Estimates and projections for all countries, 2000–2050”. In: *Progress in Planning* 75.2 (2011), pp. 53–107.
- [2] N. B. Grimm et al. “Global Change and the Ecology of Cities”. In: *Science* 319.5864 (2008), pp. 756–760.
- [3] K. R. Gurney et al. “Greenhouse gas emissions from global cities under SSP/RCP scenarios, 1990 to 2100”. In: *Global Environmental Change* 73 (2022), p. 102478.
- [4] E. von Schneidemesser et al. “How will air quality effects on human health, crops and ecosystems change in the future?” In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 378.2183 (2020), p. 20190330.
- [5] H. Qian et al. “Air pollution reduction and climate co-benefits in China’s industries”. In: *Nature Sustainability* (2021).
- [6] R. Commane and L. D. Schiferl. “Climate mitigation policies for cities must consider air quality impacts”. In: *Chem* (2022).
- [7] Q. Yu et al. “California’s zero-emission vehicle adoption brings air quality benefits yet equity gaps persist”. In: *Nature Communications* 14.1 (2023), p. 7798.
- [8] S. J. Smith, E. E. McDuffie, and M. Charles. “Opinion: Coordinated development of emission inventories for climate forcers and air pollutants”. In: *Atmospheric Chemistry and Physics* 22.19 (2022), pp. 13201–13218.
- [9] C. Gately and L. R. Hutyra. *Anthropogenic Carbon Emission System, 2012–2017, Version 2*. Published: ORNL DAAC, Oak Ridge, Tennessee, USA. <https://doi.org/10.3334/ORNLDAAC/2022>.
- [10] K. R. Gurney et al. “The Vulcan version 3.0 high-resolution fossil fuel CO emissions for the United States”. In: *Journal of Geophysical Research: Atmospheres* 125.19 (2020), e2020JD032974.
- [11] K. R. Gurney et al. “The Hestia fossil fuel CO emissions data product for the Los Angeles megacity (Hestia-LA)”. In: *Earth System Science Data* 11.3 (2019), pp. 1309–1335.

- [12] Congmeng Lyu et al. “Development of the United States Greenhouse Gas and Air Pollutants Emissions System (GRA² PES)”. en. In: *Journal of Geophysical Research: Atmospheres* 130.20 (Oct. 2025), e2025JD043597. ISSN: 2169-897X, 2169-8996. DOI: 10.1029/2025JD043597. URL: <https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JD043597> (visited on 05/11/2026).
- [13] NOAA. *First-of-its-kind dataset connects greenhouse gases and air quality*. 2024. URL: <https://research.noaa.gov/2024/09/25/first-of-its-kind-dataset-connects-greenhouse-gases-and-air-quality/>.
- [14] National Research Council and Division on Earth and Life Studies and Board on Atmospheric Sciences and Committee on Methods for Estimating Greenhouse Gas Emissions. *Verifying greenhouse gas emissions: methods to support international climate agreements*. National Academies Press, 2010.
- [15] D. C. Carslaw et al. “Recent evidence concerning higher NO_x emissions from passenger cars and light duty vehicles”. In: *Atmospheric Environment* 45.39 (2011), pp. 7053–7063.
- [16] G. A. Bishop. “Does California’s EMFAC2017 vehicle emissions model underpredict California light-duty gasoline vehicle NO_x emissions?” In: *Journal of the Air & Waste Management Association* 71.5 (2021), pp. 597–606.
- [17] D. C. Anderson et al. “Measured and modeled CO and NO_y in DISCOVER-AQ: An evaluation of emissions and chemistry over the eastern US”. In: *Atmospheric Environment* 96 (2014), pp. 78–87.
- [18] Lucy R. Huttyra et al. “Urbanization and the carbon cycle: Current capabilities and research outlook from the natural sciences perspective”. en. In: *Earth’s Future* 2.10 (Oct. 2014), pp. 473–495. ISSN: 2328-4277, 2328-4277. DOI: 10.1002/2014EF000255. URL: <https://agupubs.onlinelibrary.wiley.com/doi/10.1002/2014EF000255> (visited on 05/11/2026).
- [19] Z. Jiang et al. “Decadal Variabilities in Tropospheric Nitrogen Oxides Over United States, Europe, and China”. In: *Journal of Geophysical Research: Atmospheres* 127.3 (2022).
- [20] M. Crippa et al. “Gridded emissions of air pollutants for the period 1970–2012 within EDGAR v4.3.2”. In: *Earth System Science Data* 10.4 (2018), pp. 1987–2013.
- [21] A. F. Sarofim and R. C. Flagan. “NO_x control for stationary combustion sources”. In: *Progress in Energy and Combustion Science* 2.1 (1976), pp. 1–25.
- [22] K. A. Yu, B. C. McDonald, and R. A. Harley. “Evaluation of Nitrogen Oxide Emission Inventories and Trends for On-Road Gasoline and Diesel Vehicles”. In: *Environmental Science & Technology* 55.10 (2021), pp. 6655–6664.
- [23] S. C. Anenberg et al. “Impacts and mitigation of excess diesel-related NO_x emissions in 11 major vehicle markets”. In: *Nature* 545.7655 (2017), pp. 467–471.

- [24] J. Bi et al. “Contribution of low-cost sensor measurements to the prediction of PM_{2.5} levels: A case study in Imperial County, California, USA”. In: *Environmental Research* 180 (2020), p. 108810.
- [25] R. Piedrahita et al. “The next generation of low-cost personal air quality sensors for quantitative exposure monitoring”. In: *Atmospheric Measurement Techniques* 7.10 (2014), pp. 3325–3336.
- [26] E. G. Snyder et al. “The Changing Paradigm of Air Pollution Monitoring”. In: *Environmental Science & Technology* 47.20 (2013), pp. 11369–11377.
- [27] R. S. Sokhi et al. “Advances in air quality research – current and emerging challenges”. In: *Atmospheric Chemistry and Physics* 22.7 (2022), pp. 4615–4703.
- [28] D. J. Varon et al. “Quantifying NO_x point sources with Landsat and Sentinel-2 satellite observations of NO₂ plumes”. In: *Proceedings of the National Academy of Sciences* 121.27 (2024), e2317077121.
- [29] D. H. Cusworth et al. “Synthesis of Methane Observations Across Scales: Strategies for Deploying a Multitiered Observing Network”. In: *Geophysical Research Letters* 47.7 (2020), e2020GL087869.
- [30] Q. Zhu et al. “Direct observations of NO_x emissions over the San Joaquin Valley using airborne flux measurements during RECAP-CA 2021 field campaign”. In: *Atmospheric Chemistry and Physics* 23.17 (2023), pp. 9669–9683.
- [31] S. E. Chambliss et al. “Comparison of Mobile and Fixed-Site Black Carbon Measurements for High-Resolution Urban Pollution Mapping”. In: *Environmental Science & Technology* 54.13 (2020), pp. 7848–7857.
- [32] F. Y. Majluf et al. “Mobile Near-Field Measurements of Biomass Burning Volatile Organic Compounds: Emission Ratios and Factor Analysis”. In: *Environmental Science & Technology Letters* 9.5 (2022), pp. 383–390.
- [33] B. Liu and N. Zimmerman. “Fleet-based vehicle emission factors using low-cost sensors: Case study in parking garages”. In: *Transportation Research Part D: Transport and Environment* 91 (2021), p. 102635.
- [34] J. S. Apte and C. Manchanda. “High-resolution urban air pollution mapping”. In: *Science* 385.6707 (2024), pp. 380–385.
- [35] A. A. Shusterman et al. “The BErkeley Atmospheric CO Observation Network: initial evaluation”. In: *Atmospheric Chemistry and Physics* 16.21 (2016), pp. 13449–13463.
- [36] M. Mueller, J. Meyer, and C. Hueglin. “Design of an ozone and nitrogen dioxide sensor unit and its long-term operation within a sensor network in the city of Zurich”. In: *Atmospheric Measurement Techniques* 10.10 (2017), pp. 3783–3799.

- [37] S. Jain, A. A. Presto, and N. Zimmerman. “Spatial Modeling of Daily PM_{2.5}, NO₂, and CO Concentrations Measured by a Low-Cost Sensor Network: Comparison of Linear, Machine Learning, and Hybrid Land Use Models”. In: *Environmental Science & Technology* 55.13 (2021), pp. 8631–8641.
- [38] D. D. Parrish. “Critical evaluation of US on-road vehicle emission inventories”. In: *Atmospheric Environment* 40.13 (2006), pp. 2288–2300.
- [39] B. Hassler et al. “Analysis of long-term observations of NO_x and CO in megacities and application to constraining emissions inventories”. In: *Geophysical Research Letters* 43.18 (2016), pp. 9920–9930.
- [40] I. B. Pollack et al. “Trends in ozone, its precursors, and related secondary oxidation products in Los Angeles, California: A synthesis of measurements from 1960 to 2010”. In: *Journal of Geophysical Research: Atmospheres* 118.11 (2013), pp. 5893–5911.
- [41] D. C. Anderson et al. “Urban Emissions of Nitrogen Oxides, Carbon Monoxide, and Methane Determined from Ground-Based Measurements in Philadelphia”. In: *Environmental Science & Technology* 55.8 (2021), pp. 4532–4541.
- [42] Jocelyn C. Turnbull et al. “Toward quantification and source sector identification of fossil fuel CO₂ emissions from an urban area: Results from the INFLUX experiment”. en. In: *Journal of Geophysical Research: Atmospheres* 120.1 (Jan. 2015), pp. 292–312. ISSN: 2169-897X, 2169-8996. DOI: 10.1002/2014JD022555. URL: <https://agupubs.onlinelibrary.wiley.com/doi/10.1002/2014JD022555> (visited on 05/11/2026).
- [43] Alphasense Ltd. *NO₂-B43F/NO₂-B43F+ Nitrogen Dioxide Sensor*. Technical specifications / datasheet Doc. Ref. NO₂-B43F/FEB24. Braintree, UK: Alphasense Ltd., Feb. 2024.
- [44] Jinsol Kim et al. “The BErkeley Atmospheric CO₂ Observation Network: field calibration and evaluation of low-cost air quality sensors”. en. In: *Atmospheric Measurement Techniques* 11.4 (Apr. 2018), pp. 1937–1946. ISSN: 1867-8548. DOI: 10.5194/amt-11-1937-2018. URL: <https://amt.copernicus.org/articles/11/1937/2018/> (visited on 04/09/2026).
- [45] Anna R. Winter et al. “A Scalable Calibration Method for Enhanced Accuracy in Dense Air Quality Monitoring Networks”. en. In: *Environmental Science & Technology* 59.5 (Feb. 2025), pp. 2599–2610. ISSN: 0013-936X, 1520-5851. DOI: 10.1021/acs.est.4c08855. URL: <https://pubs.acs.org/doi/10.1021/acs.est.4c08855> (visited on 05/11/2026).
- [46] Anna R. Winter et al. “Sustained Performance of Low-Cost Air Quality Sensors in Long-Term Deployments”. en. In: *ACS Sensors* 10.6 (June 2025), pp. 4329–4335. ISSN: 2379-3694, 2379-3694. DOI: 10.1021/acssensors.5c00566. URL: <https://pubs.acs.org/doi/10.1021/acssensors.5c00566> (visited on 04/09/2026).
- [47] Jacob, Daniel J. *Introduction to atmospheric chemistry*. Princeton university press, 1999.

- [48] E. R. Delaria et al. “The Berkeley Environmental Air-quality and CO Network: field calibrations of sensor temperature dependence and assessment of network scale CO accuracy”. In: *Atmospheric Measurement Techniques* 14.8 (2021), pp. 5487–5500.
- [49] Alexander J. Turner et al. “Observed Impacts of COVID-19 on Urban CO₂ Emissions”. en. In: *Geophysical Research Letters* 47.22 (Nov. 2020), e2020GL090037. ISSN: 0094-8276, 1944-8007. DOI: 10.1029/2020GL090037. URL: <https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2020GL090037> (visited on 04/21/2026).
- [50] N. G. Asimow, A. J. Turner, and R. C. Cohen. “Sustained Reductions of Bay Area CO Emissions 2018–2022”. In: *Environmental Science & Technology* 58.15 (2024), pp. 6586–6594.
- [51] A. A. Shusterman et al. “Observing local CO sources using low-cost, near-surface urban monitors”. In: *Atmospheric Chemistry and Physics* 18.18 (2018), pp. 13773–13785.
- [52] Jinsol Kim et al. “Observing Annual Trends in Vehicular CO₂ Emissions”. en. In: *Environmental Science & Technology* 56.7 (Apr. 2022), pp. 3925–3931. ISSN: 0013-936X, 1520-5851. DOI: 10.1021/acs.est.1c06828. URL: <https://pubs.acs.org/doi/10.1021/acs.est.1c06828> (visited on 04/21/2026).
- [53] Helen L. Fitzmaurice et al. “Assessing vehicle fuel efficiency using a dense network of CO₂ observations”. en. In: *Atmospheric Chemistry and Physics* 22.6 (Mar. 2022), pp. 3891–3900. ISSN: 1680-7324. DOI: 10.5194/acp-22-3891-2022. URL: <https://acp.copernicus.org/articles/22/3891/2022/> (visited on 04/21/2026).
- [54] T. Wang et al. “Health co-benefits of achieving sustainable net-zero greenhouse gas emissions in California”. In: *Nature Sustainability* 3.8 (2020), pp. 597–605.
- [55] M. Eeftens et al. “Development of land use regression models for PM_{2.5}, PM_{2.5} absorbance, PM₁₀ and PM_{coarse} in 20 European study areas; results of the ESCAPE project”. In: *Environmental Science & Technology* 46.20 (2012), pp. 11195–11205.
- [56] K. E. Kelly et al. “Community-Based Measurements Reveal Unseen Differences during Air Pollution Episodes”. In: *Environmental Science & Technology* 55.1 (2020), pp. 120–128.
- [57] H. Z. Li et al. “Application of mobile sampling to investigate spatial variation in fine particle composition”. In: *Atmospheric Environment* 142 (2016), pp. 71–82.
- [58] M. Wang et al. “Evaluation of Land Use Regression Models for NO₂ and Particulate Matter in 20 European Study Areas: The ESCAPE Project”. In: *Environmental Science & Technology* 47.9 (2013), pp. 4357–4364.
- [59] R. Bares et al. “The Wintertime Covariation of CO and Criteria Pollutants in an Urban Valley of the Western United States”. In: *Journal of Geophysical Research: Atmospheres* 123.5 (2018), pp. 2684–2703.

- [60] Vaisala. *GMP343 Carbon Dioxide Probe Datasheet*. URL: <https://docs.vaisala.com/v/u/B210688EN-H/en-US>.
- [61] M. Y. Patel et al. “Towards a hygroscopic growth calibration for low-cost PM_{2.5} sensors”. In: *Atmospheric Measurement Techniques* 17.3 (2024), pp. 1051–1060.
- [62] D. York et al. “Unified equations for the slope, intercept, and standard errors of the best straight line”. In: *American Journal of Physics* 72.3 (2004), pp. 367–375.
- [63] C. A. Cantrell. “Technical Note: Review of methods for linear least-squares fitting of data and application to atmospheric chemistry problems”. In: *Atmospheric Chemistry and Physics* 8.17 (2008), pp. 5477–5487.
- [64] C. Wu and J. Z. Yu. “Evaluation of linear regression techniques for atmospheric applications: the importance of appropriate weighting”. In: *Atmospheric Measurement Techniques* 11.2 (2018), pp. 1233–1250.
- [65] H. Simon et al. “Variability in Observation-Based Onroad Emission Constraints from a Near-Road Environment”. In: *Atmosphere* 11.11 (2020), p. 1243.
- [66] G. Barbato et al. “Features and performance of some outlier detection methods”. In: *Journal of Applied Statistics* 38.10 (2011), pp. 2133–2149.
- [67] L. Ammoura et al. “Exploiting stagnant conditions to derive robust emission ratio estimates for CO, CO and volatile organic compounds in Paris”. In: *Atmospheric Chemistry and Physics* 16.24 (2016), pp. 15653–15664.
- [68] N. Chandra et al. “Temporal variations of atmospheric CO and CO at Ahmedabad in western India”. In: *Atmospheric Chemistry and Physics* 16.10 (2016), pp. 6153–6173.
- [69] R. J. Yokelson, M. O. Andreae, and S. K. Akagi. “Pitfalls with the use of enhancement ratios or normalized excess mixing ratios measured in plumes to characterize pollution sources and aging”. In: *Atmospheric Measurement Techniques* 6.8 (2013), pp. 2155–2158.
- [70] H. S. Halliday et al. “Using Short-Term CO/CO Ratios to Assess Air Mass Differences Over the Korean Peninsula During KORUS-AQ”. In: *Journal of Geophysical Research: Atmospheres* 124.20 (2019), pp. 10951–10972.
- [71] T. S. Mangat et al. *Source inventory of Bay Area greenhouse gas emissions*. Tech. rep. San Francisco, CA: Bay Area Air Quality Management District, 2010.
- [72] J. Kwon, P. Varaiya, and A. Skabardonis. “Estimation of truck traffic volume from single loop detectors with lane-to-lane speed correlation”. In: *Transportation Research Record* 1856.1 (2003), pp. 106–117.
- [73] J. Horel et al. “MESOWEST: Cooperative Mesonets in the Western United States”. In: *Bulletin of the American Meteorological Society* 83.2 (2002), pp. 211–226.
- [74] G. P. Brasseur and D. J. Jacob. *Modeling of Atmospheric Chemistry*. Cambridge University Press, 2017.

- [75] F. Smith. “A scheme for estimating the vertical dispersion of a plume from a source near ground-level”. In: *Atmospheric Diffusion*. 1974.
- [76] R. Hosker. *Estimates of dry deposition and plume depletion over forests and grassland*. Tech. rep. Oak Ridge, TN: National Oceanic and Atmospheric Administration, 1974.
- [77] C. S. Borrego, M. S. Coutinho, and M. J. Costa. “Introduction of terrain roughness effects into a Gaussian dispersion model”. In: *Science of The Total Environment* 99.1 (1990), pp. 153–161.
- [78] J. S. Apte et al. “High-Resolution Air Pollution Mapping with Google Street View Cars: Exploiting Big Data”. In: *Environmental Science & Technology* 51.12 (2017), pp. 6999–7008.
- [79] A. A. Karner, D. S. Eisinger, and D. A. Niemeier. “Near-Roadway Air Quality: Synthesizing the Findings from Real-World Data”. In: *Environmental Science & Technology* 44.14 (2010), pp. 5334–5344.
- [80] B. C. McDonald et al. “High-resolution mapping of motor vehicle carbon dioxide emissions”. In: *Journal of Geophysical Research: Atmospheres* 119.9 (2014), pp. 5283–5298.
- [81] L. He et al. “Atmospheric Methane Emissions Correlate With Natural Gas Consumption From Residential and Commercial Sectors in Los Angeles”. In: *Geophysical Research Letters* 46.14 (2019), pp. 8563–8571.
- [82] R. A. Harley et al. “Changes in Motor Vehicle Emissions on Diurnal to Decadal Time Scales and Effects on Atmospheric Composition”. In: *Environmental Science & Technology* 39.14 (2005), pp. 5356–5362.
- [83] O. G. Sutton. “A theory of eddy diffusion in the atmosphere”. In: *Proceedings of the Royal Society of London. Series A* 135.826 (1932), pp. 143–165.
- [84] N. A. Mullen et al. “Results of the California Healthy Homes Indoor Air Quality Study of 2011–2013: impact of natural gas appliances on air pollutant concentrations”. In: *Indoor Air* 26.2 (2016), pp. 231–245.
- [85] B. C. Singer et al. “Pollutant concentrations and emission rates from natural gas cooking burners without and with range hood exhaust in nine California homes”. In: *Building and Environment* 122 (2017), pp. 215–229.
- [86] C. Arata et al. “Measurement of NO and NO in a Residential Kitchen”. In: *Environmental Science & Technology Letters* 5.10 (2018), pp. 595–599.
- [87] California Air Resources Board. *Current Air District Rules*. URL: <https://ww2.arb.ca.gov/capp/cst/tch/current-air-district-rules>.

- [88] Christopher W. Tessum et al. “Inequity in consumption of goods and services adds to racial–ethnic disparities in air pollution exposure”. en. In: *Proceedings of the National Academy of Sciences* 116.13 (Mar. 2019), pp. 6001–6006. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1818859116. URL: <https://pnas.org/doi/full/10.1073/pnas.1818859116> (visited on 04/10/2026).
- [89] Mary Angelique G. Demetillo et al. “Observing Nitrogen Dioxide Air Pollution Inequality Using High-Spatial-Resolution Remote Sensing Measurements in Houston, Texas”. en. In: *Environmental Science & Technology* 54.16 (Aug. 2020), pp. 9882–9895. ISSN: 0013-936X, 1520-5851. DOI: 10.1021/acs.est.0c01864. URL: <https://pubs.acs.org/doi/10.1021/acs.est.0c01864> (visited on 04/10/2026).
- [90] Julian D Marshall et al. “Advancing Methods and Models that Promote Equity in Ambient Air Quality”. en. In: ().
- [91] Yuzhou Wang, Julian D. Marshall, and Joshua S. Apte. “U.S. Ambient Air Monitoring Network Has Inadequate Coverage under New PM_{2.5} Standard”. en. In: *Environmental Science & Technology Letters* 11.11 (Nov. 2024), pp. 1220–1226. ISSN: 2328-8930, 2328-8930. DOI: 10.1021/acs.estlett.4c00605. URL: <https://pubs.acs.org/doi/10.1021/acs.estlett.4c00605> (visited on 04/10/2026).
- [92] Kley, DAMM and McFarland, M. “Chemiluminescence detector for NO and NO₂”. In: *Atmospheric Technology* 12 (1980).
- [93] Hendrik Fuchs et al. “A Sensitive and Versatile Detector for Atmospheric NO₂ and NO_x Based on Blue Diode Laser Cavity Ring-Down Spectroscopy”. en. In: *Environmental Science & Technology* 43.20 (Oct. 2009), pp. 7831–7836. ISSN: 0013-936X, 1520-5851. DOI: 10.1021/es902067h. URL: <https://pubs.acs.org/doi/10.1021/es902067h> (visited on 05/11/2026).
- [94] Paul L. Kebabian, Scott C. Herndon, and Andrew Freedman. “Detection of Nitrogen Dioxide by Cavity Attenuated Phase Shift Spectroscopy”. en. In: *Analytical Chemistry* 77.2 (Jan. 2005), pp. 724–728. ISSN: 0003-2700, 1520-6882. DOI: 10.1021/ac048715y. URL: <https://pubs.acs.org/doi/10.1021/ac048715y> (visited on 05/11/2026).
- [95] Paul L. Kebabian et al. “A Practical Alternative to Chemiluminescence-Based Detection of Nitrogen Dioxide: Cavity Attenuated Phase Shift Spectroscopy”. en. In: *Environmental Science & Technology* 42.16 (Aug. 2008), pp. 6040–6045. ISSN: 0013-936X, 1520-5851. DOI: 10.1021/es703204j. URL: <https://pubs.acs.org/doi/10.1021/es703204j> (visited on 05/11/2026).
- [96] T. B. Ryerson, E. J. Williams, and F. C. Fehsenfeld. “An efficient photolysis system for fast-response NO₂ measurements”. en. In: *Journal of Geophysical Research: Atmospheres* 105.D21 (Nov. 2000), pp. 26447–26461. ISSN: 0148-0227. DOI: 10.1029/2000JD900389. URL: <https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2000JD900389> (visited on 05/11/2026).

- [97] Ilana B. Pollack, Brian M. Lerner, and Thomas B. Ryerson. “Evaluation of ultraviolet light-emitting diodes for detection of atmospheric NO₂ by photolysis - chemiluminescence”. en. In: *Journal of Atmospheric Chemistry* 65.2-3 (Apr. 2010), pp. 111–125. ISSN: 0167-7764, 1573-0662. DOI: 10.1007/s10874-011-9184-3. URL: <http://link.springer.com/10.1007/s10874-011-9184-3> (visited on 05/11/2026).
- [98] Chris Reed et al. “Interferences in photolytic NO₂ measurements: explanation for an apparent missing oxidant?” en. In: *Atmospheric Chemistry and Physics* 16.7 (Apr. 2016), pp. 4707–4724. ISSN: 1680-7324. DOI: 10.5194/acp-16-4707-2016. URL: <https://acp.copernicus.org/articles/16/4707/2016/> (visited on 05/11/2026).
- [99] US Environmental Protection Agency. *Air Quality System Data Mart [internet database]* available via <https://www.epa.gov/outdoor-air-quality-data>. Accessed March 05, 2026.
- [100] Michael L. Hitchman and John R. Saffell. “Considerations of Thermodynamics and Kinetics for the Effects of Relative Humidity on the Electrolyte in Electrochemical Toxic Gas Sensors”. en. In: *ACS Sensors* 6.11 (Nov. 2021), pp. 3985–3993. ISSN: 2379-3694, 2379-3694. DOI: 10.1021/acssensors.1c01339. URL: <https://pubs.acs.org/doi/10.1021/acssensors.1c01339> (visited on 04/21/2026).
- [101] Georgia Miskell, Jennifer A. Salmond, and David E. Williams. “Solution to the Problem of Calibration of Low-Cost Air Quality Measurement Sensors in Networks”. en. In: *ACS Sensors* 3.4 (Apr. 2018), pp. 832–843. ISSN: 2379-3694, 2379-3694. DOI: 10.1021/acssensors.8b00074. URL: <https://pubs.acs.org/doi/10.1021/acssensors.8b00074> (visited on 04/21/2026).
- [102] Cameron G. MacDonald et al. “Estimating enhancement ratios of nitrogen dioxide, carbon monoxide and carbon dioxide using satellite observations”. en. In: *Atmospheric Chemistry and Physics* 23.6 (Mar. 2023), pp. 3493–3516. ISSN: 1680-7324. DOI: 10.5194/acp-23-3493-2023. URL: <https://acp.copernicus.org/articles/23/3493/2023/> (visited on 04/21/2026).
- [103] Jinsol Kim et al. “Observing Anthropogenic and Biogenic CO₂ Emissions in Los Angeles Using a Dense Sensor Network”. en. In: *Environmental Science & Technology* 59.7 (Feb. 2025), pp. 3508–3517. ISSN: 0013-936X, 1520-5851. DOI: 10.1021/acs.est.4c11392. URL: <https://pubs.acs.org/doi/10.1021/acs.est.4c11392> (visited on 04/21/2026).
- [104] Milan Y. Patel et al. “Plume Detection and Emissions Quantification Potential Using a Dense Sensor Network”. en. In: *ACS ES&T Air* 2.6 (June 2025), pp. 1099–1106. ISSN: 2837-1402, 2837-1402. DOI: 10.1021/acsestair.5c00069. URL: <https://pubs.acs.org/doi/10.1021/acsestair.5c00069> (visited on 04/21/2026).
- [105] Milan Y. Patel et al. “Using Network Observations to Constrain CO and CO₂ Emissions From an Oil Refinery in the San Francisco Bay Area”. en. In: *Journal of Geophysical Research: Atmospheres* 131.6 (Mar. 2026), e2025JD045035. ISSN: 2169-897X,

- 2169-8996. DOI: 10.1029/2025JD045035. URL: <https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JD045035> (visited on 04/21/2026).
- [106] Naomi G Asimow et al. *Differences in Regional Home Heating Behavior in Three U.S. Cities Revealed by Ground-Based Sensor Network*. en. Mar. 2025. DOI: 10.22541/essoar.174180753.30586978/v1. URL: <https://essopenarchive.org/doi/full/10.22541/essoar.174180753.30586978/v1> (visited on 04/21/2026).
- [107] Yishu Zhu et al. “Observational Inferences of NO_x and CO Emission Factors for Vehicles and Homes in the San Francisco Bay Area”. en. In: *ACS ES&T Air* 2.8 (Aug. 2025), pp. 1478–1487. ISSN: 2837-1402, 2837-1402. DOI: 10.1021/acsestair.5c00004. URL: <https://pubs.acs.org/doi/10.1021/acsestair.5c00004> (visited on 01/20/2026).
- [108] Matthew S. Reiter and Kara M. Kockelman. “The problem of cold starts: A closer look at mobile source emissions levels”. en. In: *Transportation Research Part D: Transport and Environment* 43 (Mar. 2016), pp. 123–132. ISSN: 13619209. DOI: 10.1016/j.trd.2015.12.012. URL: <https://linkinghub.elsevier.com/retrieve/pii/S1361920915002229> (visited on 04/21/2026).

Appendix A

Supplementary to Chapter 2

Text S1: MesoWest and HRRR Wind Speed Comparison

Wind speed data from 44 MesoWest stations were matched to BEACO₂N sites and compared to HRRR reanalysis fields at 10 m height. Using logarithmic wind profile scaling to approximate winds at 5 m, HRRR wind speeds showed reasonable agreement with MesoWest observations. Summer and winter WS distributions are shown in Figure A.2.

Text S2: Details on the Simplified 3-D Numerical Advection-Diffusion Model

Overview: A simplified three-dimensional numerical advection-diffusion model was developed to describe a 10 km-long highway line source located at the edge of the domain as it mixes into an adjacent 10 km $\times$ 10 km residential neighborhood downwind. The model simulates atmospheric concentrations based on height-averaged equations for advection and diffusion, assuming that the air above the ground surface is well-mixed. The influence of chemical reactions on the concentrations was tested separately. Enhancement ratios of the simulated concentrations for two species were analyzed to compare with the observed profiles. Emissions from both line source and area source are confined to the ground surface ($z = 0$ m) and occur at constant rates throughout the simulation. The area source emissions are distributed evenly over all the grid cells. The wind velocities in the y - and z - directions are set to zero, meaning that transport in these directions occurs solely through diffusion. The wind in the x -direction is perpendicular to the line source, located at $x = x_0$. The turbulent diffusion coefficient in the z -direction is adjusted for surface roughness using the Hosker Index (HI) for urban terrain.

The 3-D scalar equations for each species solved in the model are as follows:

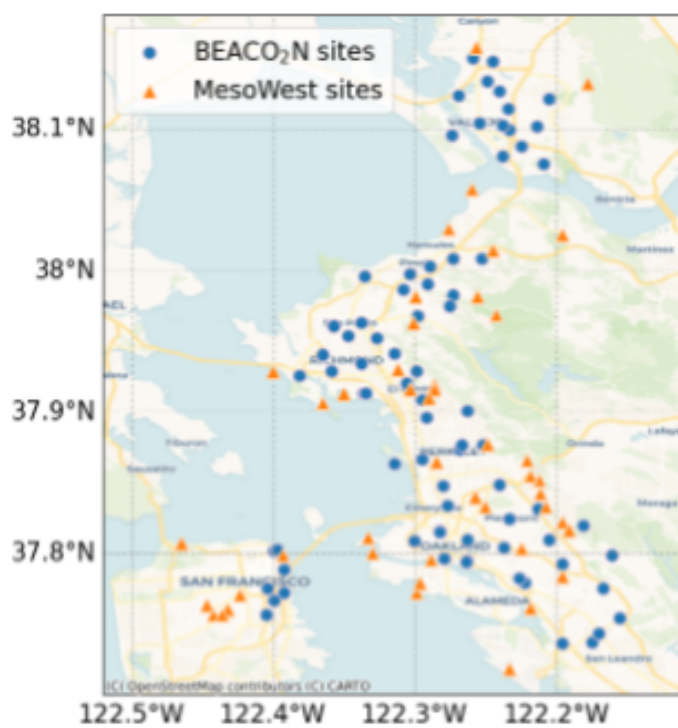


Figure A.1: Map of BEACO₂N and MesoWest site locations in the study region. Background map credits: © OpenStreetMap contributors © CARTO.

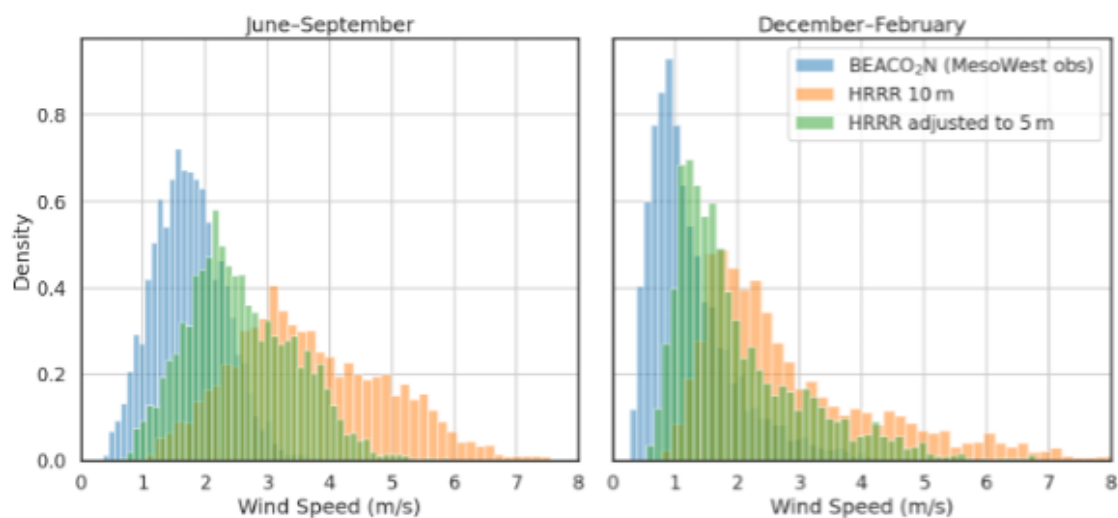


Figure A.2: Wind speed distributions from MesoWest measurements and HRRR reanalysis for June–September and December–February.

$$\frac{\partial A}{\partial t} + u \frac{\partial A}{\partial x} + v \frac{\partial A}{\partial y} + w \frac{\partial A}{\partial z} = D_x \frac{\partial^2 A}{\partial x^2} + D_y \frac{\partial^2 A}{\partial y^2} + D_z \frac{\partial^2 A}{\partial z^2} + \delta(x - x_0) \frac{Q_{A,\text{line}}}{L_y} + \frac{Q_{A,\text{area}}}{L_x L_y} \quad (\text{A.1})$$

$$\frac{\partial B}{\partial t} + u \frac{\partial B}{\partial x} + v \frac{\partial B}{\partial y} + w \frac{\partial B}{\partial z} = D_x \frac{\partial^2 B}{\partial x^2} + D_y \frac{\partial^2 B}{\partial y^2} + D_z \frac{\partial^2 B}{\partial z^2} + \delta(x - x_0) \frac{Q_{B,\text{line}}}{L_y} + \frac{Q_{B,\text{area}}}{L_x L_y} \quad (\text{A.2})$$

where:

- A represents the concentration of the co-emitted gas species (e.g., NO_x). Units: ppbv.
- B represents the concentration of CO_2 . Units: ppmv.
- $Q_{A,\text{line}}, Q_{B,\text{line}}$ are the total emission rates per unit length for the line source for species A and species B, respectively. Units: $\mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$.
- $Q_{A,\text{area}}, Q_{B,\text{area}}$ are the total emission rates per unit area for the area source for species A and species B, respectively. Units: $\mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$.
- L_x, L_y , and L_z are the lengths of the simulation domain in the x -, y - and z -directions. Units: m.
- u, v , and w are the velocities in the x -, y -, and z -directions respectively. Units: $\text{m}\cdot\text{s}^{-1}$.
- D_x, D_y , and D_z are the diffusion coefficients in the x -, y -, and z - directions respectively. Units: $\text{m}^2\cdot\text{s}^{-1}$.
- $\delta(x - x_0)$: Dirac delta function centered at $x = x_0$, representing the line source.
- t : Time. Units: s.
- x, y, z : Spatial coordinates in the x -, y -, and z -directions. Units: m.

The scalar equations are discretized using explicit first-order time stepping and second-order central differences in space for simplicity.

Initial and Boundary Conditions: At the start of the simulation, the concentrations of species are initialized to a background concentration, A_{bkg} or B_{bkg} , across the entire domain.

At the inflow sections, it is assumed that incoming air is at background concentrations, thus:

$$A(0, y, z, t) = A_{\text{bkg}}, \quad B(0, y, z, t) = B_{\text{bkg}} \quad (\text{A.3})$$

At the outflow section, air is allowed to leave the system by advection, and the diffusive flux is assumed negligible, i.e.:

$$\frac{\partial A}{\partial x} \Big|_{(L_x, y, z, t)} = 0, \quad \frac{\partial B}{\partial x} \Big|_{(L_x, y, z, t)} = 0 \quad (\text{A.4})$$

The flux at the upper boundary ($z = L_z$) is set to zero and the concentration at the upper boundary to the value of the second-to-last grid cell:

$$\frac{\partial A}{\partial z} \Big|_{(x,y,L_z,t)} = 0, \quad \frac{\partial B}{\partial z} \Big|_{(x,y,L_z,t)} = 0 \quad (\text{A.5})$$

Emission Rates for Line and Area Sources: Emission rates of CO₂ (i.e., $Q_{B,\text{line}}$, and $Q_{B,\text{area}}$) are estimated based on the posterior CO₂ results from Asimow et al. (see main text). Emission rates of co-emitted species (i.e., $Q_{A,\text{line}}$, and $Q_{A,\text{area}}$) for the line and area sources are determined by multiplying $Q_{B,\text{line}}$, and $Q_{B,\text{area}}$ by the corresponding emission factors. We used 1.2 ppbv/ppmv as the highway emission factor and 0.2 ppbv/ppmv as the area source emission factor to examine the distribution of enhancement ratios along with the distance from the highway. The emission rates for species A and B are as follows:

$$Q_{A,\text{line}} = 1.2 \times 10^{-3} \times Q_{B,\text{line}}, \quad Q_{A,\text{area}} = 0.2 \times 10^{-3} \times Q_{B,\text{area}} \quad (\text{A.6})$$

Domain setup and parameter configurations: The parameters used in the 3-D advection-diffusion model along with their values and units — including grid spacing, wind speeds, diffusion coefficients, and emission rates — are summarized in Table A.1.

Text S3. Wind Speed Sensitivity Analysis in the Idealized Advection-Diffusion Model

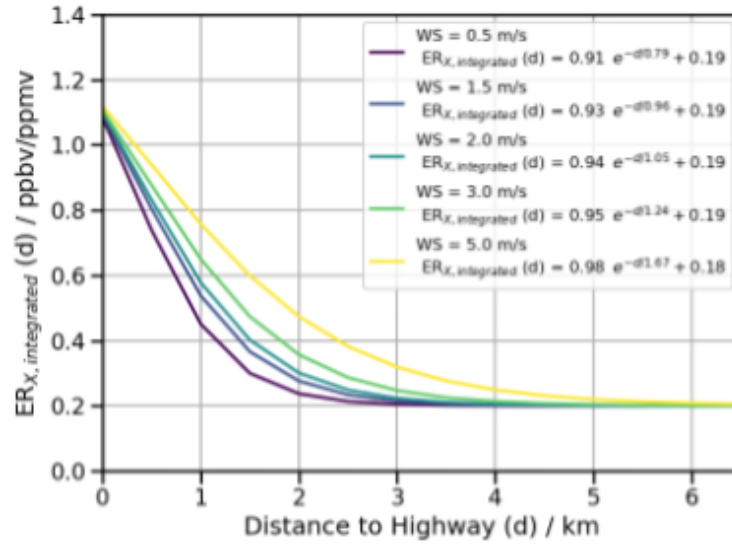
To examine how variations in WS influence the shape of ER profiles, we performed simulations using WS values corresponding to the lower bound (0.5 m/s) and upper bound (5 m/s) observed wind speeds. Results indicate that the mixing lengths for summer and winter ER profiles increase by 2.4 km over this range, but the ERs at $d = 0$ km and the offset at $d > 3$ km only show insignificant differences. These results suggest that the distance-decay patterns of ERs are relatively insensitive to WS variability within typical conditions in the SF Bay Area.

Text S4. Sensitivity of Modeled ER Decay to NO_x Lifetime

A sensitivity test was conducted varying the NO_x lifetimes in the advection-diffusion model ($\tau = \infty$, 10 hours, 5 hours, 1 hours, 0.2 hours). While the resulting ER profile when $\tau = 0.2$ hours showed a slightly faster decay, the derived EF values at the origin ($d = 0$ km) and at far downwind distances ($d > 3$ km) changed by less than 10%. This confirms that our conclusions are not sensitive to variations in NO_x lifetime.

Table A.1: Description and values of parameters used in the 3-D advection-diffusion model.

Symbol	Description	Value	Unit
dx, dy	Grid size in the x - and y -directions	500	meter
N_x, N_y	Number of grids in the x - and y -directions	20	–
L_x, L_y	Length of domain in the x - and y -directions	10000	meter
dz	Grid size in the z -direction	200	meter
N_z	Number of grids in the z -direction	3	–
L_z	Length of domain in the z -direction	600	meter
dt	Timesteps	2	second
N_t	Number of timesteps	25000	–
u	Horizontal wind velocity	1.5	m/s
v	Lateral wind velocity	0	m/s
w	Vertical wind velocity	0	m/s
D_x, D_y	Turbulent diffusivity in the x - and y - directions	1000	m^2/s
$D_{z,0}$	Initial turbulent diffusivity in the z -direction	150	m^2/s
D_z	$D_{z,0} \times \text{HI}$	$150 \times \text{HI}$	m^2/s
$Q_{A,\text{line}}$	Total emission rate per unit length for line source	$0.0012 \times Q_{B,\text{line}}$	$\mu\text{mol} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$
$Q_{B,\text{line}}$	Total emission rate per unit length for line source	See main text	$\mu\text{mol} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$
$Q_{A,\text{area}}$	Total emission rate per unit area for area source	$0.0002 \times Q_{B,\text{area}}$	$\mu\text{mol} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$
$Q_{B,\text{area}}$	Total emission rate per unit area for area source	See main text	$\mu\text{mol} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$
A_{bkg}	Background for species A (NO_x)	0.5	ppbv
B_{bkg}	Background for species B (CO_2)	400	ppmv

Figure A.3: Modeled ER profile comparisons for $\text{WS} = 0.5, 1, 1.5, 3, \text{ and } 5 \text{ m/s}$.

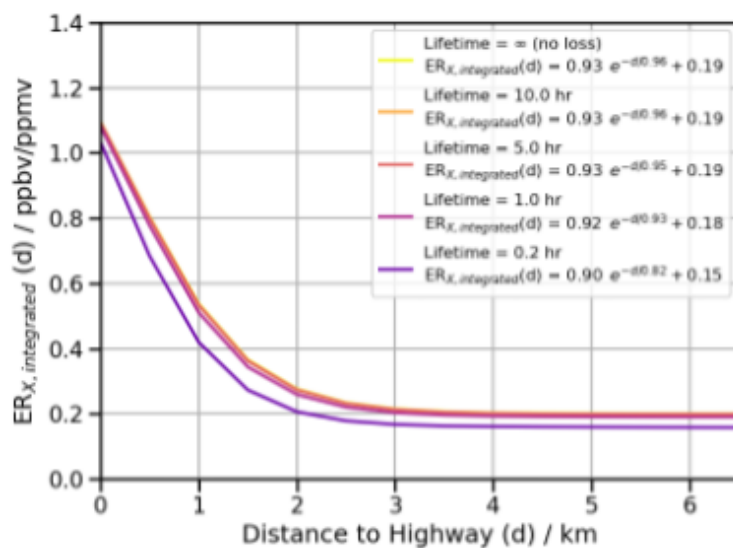


Figure A.4: Comparison of simulated ER profiles using NO_x lifetimes of ∞ , 10, 5, 1 and 0.2 hours.

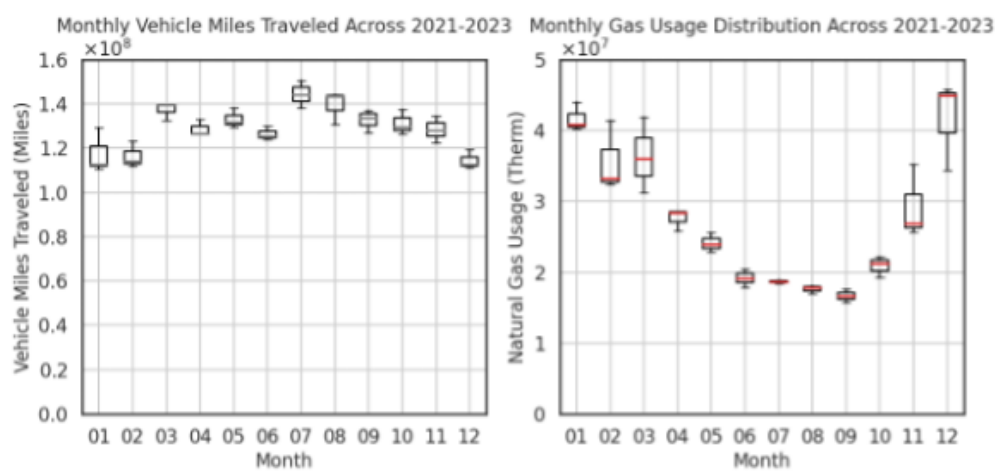


Figure A.5: Monthly vehicle miles traveled (VMT) and monthly residential natural gas consumption in the SF Bay Area in 2021–2023.

Text S5. Sectoral Activity Trends Supporting Emission Attribution

To define peak and low periods for traffic and building energy use, we compiled monthly vehicle miles traveled (VMT) and residential natural gas consumption data for 2021–2023. VMT data were obtained from the Caltrans Performance Measurement System (PeMS), and natural gas usage was sourced from Pacific Gas and Electric Company (PG&E) public datasets aggregated by ZIP code.

As shown in Figure A.5, natural gas usage consistently peaked during December–February and dropped to seasonal lows between June and September. These patterns support our classification of “peak” and “low” usage periods in the main analysis and provide a basis for interpreting the seasonal variation in observed ERs.

Text S6. BEACO₂N Site Information and Data Filtering

To clarify which BEACO₂N sites were included in the ER and EF analysis, we compiled a complete status table (Table A.2) and a corresponding site map (Figure A.6). Table A.2 includes each site’s coordinates, data completeness for NO_x and CO in 2022, regression R^2 value, and average distance to the nearest highway during the summer 6:00–9:00 AM time window with wind directions from 120°–260°.

Table A.2: Summary of BEACO₂N site locations and inclusion status for ER analysis. Symbols indicate inclusion status: ✓ = passed all inclusion criteria and used in ER distance-decay fitting; • = passed data completeness but not the R^2 threshold; ○ = no nearby upwind highway; blank = did not pass data completeness criteria.

Site ID	Name	Lat	Lon	Avg. Dist. to Hwy (km)	ER _{NO_x}	ER _{CO}	Note
3	Piedmont High School	37.82	–122.23	2.80		✓	<80% NO _x data coverage in 2022
11	Bay Area Metro Center	37.79	–122.39	1.75	✓		<80% CO data coverage in 2022
13	Richmond High School	37.95	–122.35	3.55	✓	✓	
14	Helms Middle School	37.96	–122.34				<80% NO _x or CO data coverage in 2022
15	Ohlone Elementary School	38.01	–122.27				<80% NO _x data coverage in 2022; R^2 for CO < 0.5
16	Ellerhorst Elementary School	37.98	–122.27		•	•	R^2 for NO _x and CO < 0.5
17	Stewart School	37.99	–122.29	3.22	✓	✓	
20	Nystrom School	37.93	–122.36		•	•	R^2 for NO _x and CO < 0.5

Continued on next page

Table A.2 continued from previous page

Site ID	Name	Lat	Lon	Avg. Dist. to Hwy (km)	ER _{NO_x}	ER _{CO}	Note
21	Washington Elementary School	37.93	-122.38		●		R^2 for NO _x and CO < 0.5
22	Dejean Middle School	37.93	-122.34	1.69	✓	●	R^2 for CO < 0.5
23	Mira Vista Elementary School	37.94	-122.31	2.26	✓	✓	
24	Madera Elementary School	37.93	-122.30				<80% NO _x or CO data coverage in 2022
25	Fred T. Korematsu Middle School	37.92	-122.31		●	●	R^2 for NO _x and CO < 0.5
26	Peres Elementary School	37.94	-122.37	1.63	✓	✓	
28	Collins Elementary School	38.00	-122.29		○	○	No upwind highway
29	Montalvin Elementary School	38.00	-122.34				<80% NO _x or CO data coverage in 2022
30	Sheldon Elementary School	37.97	-122.30			●	<80% NO _x data coverage in 2022; R^2 for CO < 0.5
31	Fred T. Korematsu Elementary School	37.74	-122.18	2.17	✓	✓	
33	Oakland High School	37.80	-122.24	1.94		✓	<80% NO _x data coverage in 2022
35	Bishop O'Dowd High School	37.75	-122.16				<80% NO _x or CO data coverage in 2022
38	St. Elizabeth High School	37.78	-122.22	2.89	✓	✓	
39	Hall of Justice	37.78	-122.40	0.90	✓	✓	
40	Head Royce School	37.81	-122.21	2.12	●	✓	R^2 for NO _x < 0.5
41	Richmond Field Station	37.91	-122.34	4.34		✓	<80% NO _x data coverage in 2022
42	EBMUD	37.82	-122.28	1.54	✓	✓	
43	BAAQMD sanpablo	37.96	-122.36	4.34	✓	●	R^2 for CO < 0.5
44	Albany High School	37.90	-122.29	3.58	✓	✓	
45	Laney College	37.79	-122.26	0.14	✓	✓	
46	SF Hospital Bldg 5	37.76	-122.41			●	<80% NO _x data coverage in 2022; R^2 for CO < 0.5
48	Exploratorium Bay	37.80	-122.40			●	<80% NO _x data coverage in 2022; R^2 for CO < 0.5
49	Burckhalter Elementary	37.78	-122.17	3.15	✓	●	R^2 for CO < 0.5
51	El Sobrante	37.97	-122.28	7.13	✓	✓	
60	Jesse Bethel High School	38.12	-122.21	3.88		✓	<80% NO _x data coverage in 2022
63	Dan Mini Elementary School	38.15	-122.26	1.17	✓	✓	
64	Federal Terrace Elementary School	38.12	-122.27			●	<80% NO _x data coverage in 2022; R^2 for CO < 0.5
65	Marc Island Health and Fitness Academy	38.10	-122.27	4.81	✓	✓	
78	Lincoln Elementary School	38.10	-122.25	2.83		✓	<80% NO _x data coverage in 2022

Continued on next page

Table A.2 continued from previous page

Site ID	Name	Lat	Lon	Avg. Dist. to Hwy (km)	ER _{NO_x}	ER _{CO}	Note
81	Beverly Hills Elementary School	38.09	-122.23	1.35	✓	✓	
82	Glen Cove Elementary School	38.08	-122.21				<80% NO _x or CO data coverage in 2022
125	SF Public Safety Building	37.77	-122.39	0.67	✓	✓	

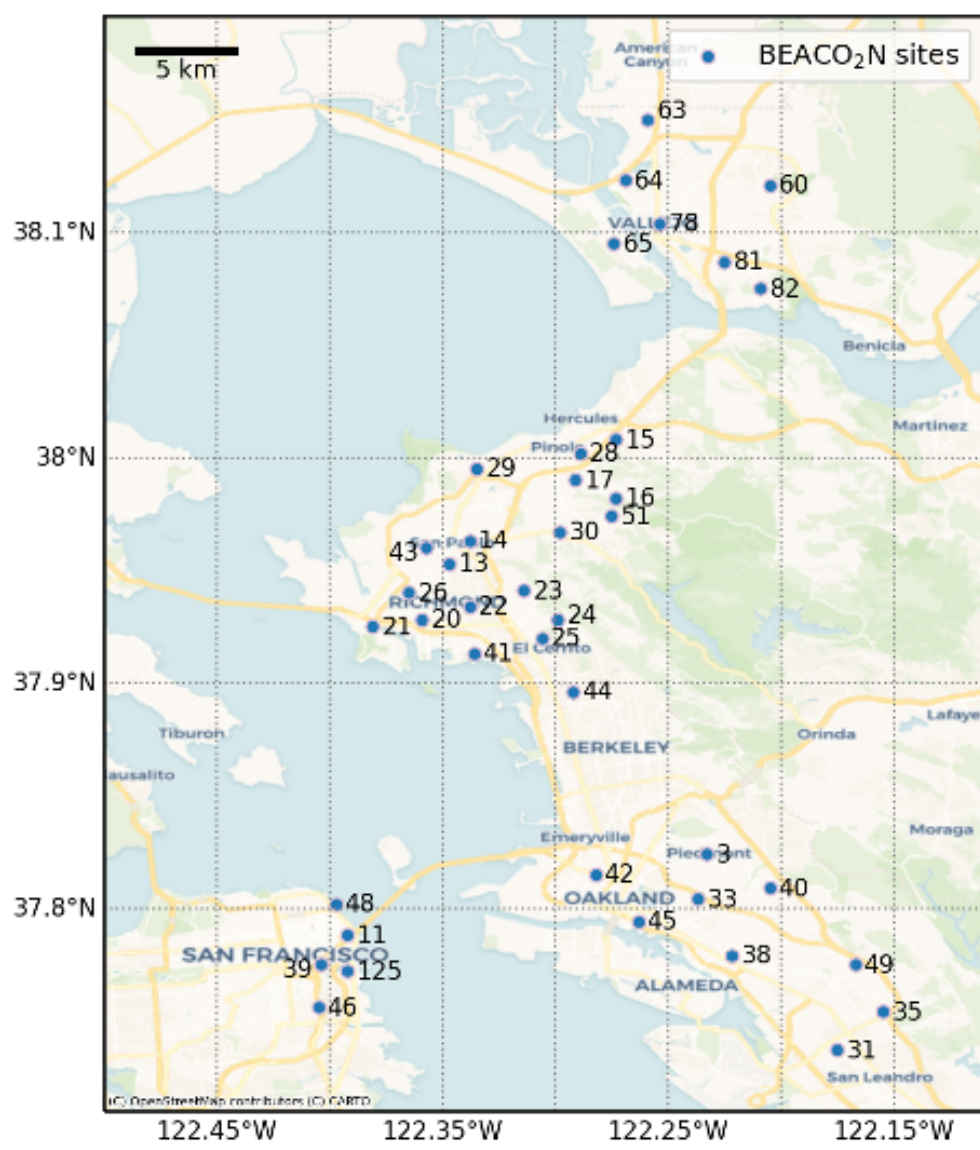


Figure A.6: Map of BEACO₂N site locations labeled with site ID. Background map credits: © OpenStreetMap contributors © CARTO.

Appendix B

Supplementary to Chapter 3

Supplementary Information

Contents

Supplementary Methods

- S-Methods-1: NO₂ zero-level estimation (Z_{NO_2})
- S-Methods-2: CO₂ enhancement index details
- S-Methods-3: Stage 3 offset design rationale

Supplementary Tables

- Table S1: Reference station metadata
- Table S2: Sensitivity parameter ranges and outcomes
- Table S3: Progressive comparison results, site 42 (EBMUD)

Group A — Bay Area progressive comparison

- Figure S1: Progressive comparison results, site 42 (robustness check)
- Figure S2: Seasonal breakdown across both sites and all validation jobs
- Figure S3: Per-pair slope heatmap, WLS+GC remote

Group B — Titration ratio diagnostics (per-city)

- Figure S4: Providence per-node KDE grid
- Figure S5: Bay Area per-node KDE grid

- Figure S6: Los Angeles per-node KDE grid

Group C — Bay Area validation robustness

- Figure S7: Reference-spread boxplot (RMSE / slope / R^2 across remote refs)

Group D — Pipeline component characterisation

- Figure S8: NO term vs ambient temperature (temperature dependence check)
- Figure S9: Fraction of data above T_{CAP} per city
- Figure S10: Z_{NO_2} vs temperature per city
- Figure S11–S13: Z_{NO_2} diagnostics: network distribution, parameter distributions, and predictor model comparison
- Figure S14–S16: WLS coefficient time series, all nodes per city (Providence, Bay Area, Los Angeles)
- Figure S17: Bay Area inter-reference bias matrix
- Figure S18: Titration event frequency heatmaps per city
- Figure S19: Monthly offset time series per city

Group E — Sensitivity analysis: event detection

- Figure S20: OAT event detection sensitivity, part 1 (baseline stability, end condition)
- Figure S21: OAT event detection sensitivity, part 2 (night window, R clipping, bin width)

Group F — Sensitivity analysis: WLS, index, offset, and Z_{NO_2} model parameters

- Figure S22: WLS window, offset window, percentile, index cap, and Z_{NO_2} model sensitivity
- Figure S23: Bin-weighting scheme comparison
- Figure S24: O_3 calibration coupling sensitivity

Group G — Sensitivity analysis: NO_2 zero level

- Figure S25: Z_{NO_2} window length $\times$ NO_2 percentile (joint sensitivity)
- Figure S26: Z_{NO_2} scaling factor and offset formulation

Group H — Sensitivity analysis: gain curve robustness

- Figure S27: 15–20°C $g(T)$ bump investigation
- Figure S28: Event R^2 quality threshold
- Figure S29: T_{CAP} sensitivity

Group I — Additional validation diagnostics

- Figure S30: Validation across all site–reference pairs

Supplementary Methods**S-Methods-1: NO₂ zero-level estimation (Z_{NO_2})**

The gain correction operates on the NO₂ signal *above* the sensor’s effective zero level Z_{NO_2} , which captures the O₃ cross-sensitivity contribution to the gain-scaled terms S_{gain} . This zero level must be subtracted before dividing by $g(T)$ and restored after; otherwise the O₃ cross-sensitivity — already correctly handled by the WLS coefficients — would be spuriously amplified. This section describes the estimation of Z_{NO_2} .

The O₃ cross-sensitivity term $\beta_{\text{O}_x} V_{\text{O}_x}$ requires careful treatment. V_{O_x} is the O₃ working electrode voltage, which responds to both ambient O₃ and to common-mode electromagnetic noise shared with the NO₂ electrode on the same sensor board. If S_{gain} were divided directly by $g(T)$, this O₃-driven floor would be amplified alongside the NO₂ signal — a spurious scaling of a component that the WLS coefficients already handle correctly.

The gain-scaled sum is defined as:

$$S_{\text{gain}}(t) = \beta_{\text{NO}_2} V_{\text{NO}_2}(t) + \beta_{\text{O}_x} V_{\text{O}_x}(t) \quad (\text{B.1})$$

The NO cross-interference term ($\beta_{\text{NO}} V_{\text{NO}}$) is excluded from S_{gain} because the NO electrode signal shows near-zero temperature dependence (Figure S8).

The NO₂ zero level is estimated by regressing S_{gain} against V_{O_x} , T , and RH during background hours:

$$\hat{Z}_{\text{NO}_2}(t) = \alpha \cdot V_{\text{O}_x}(t) + \beta \quad (\text{B.2})$$

where α and β are fitted by regression of S_{gain} against V_{O_x} during background hours. Background hours are defined as hours when the remote reference NO₂ falls below its rolling 10th percentile within a 3-month window. These are conditions when ambient NO₂ is near-zero and S_{gain} is dominated by the O₃ cross-sensitivity term.

Because V_{O_x} does not perfectly reproduce the O₃-related contribution captured in the NO₂ channel, the raw regression estimate $\hat{Z}_{\text{NO}_2}$ tends to overestimate the effective zero level. A scaling factor $\lambda = 0.7$ is therefore applied: $Z_{\text{NO}_2} = \lambda \cdot \hat{Z}_{\text{NO}_2}$ (see main text Equation 3.8; sensitivity in Figure S26).

The physical justification for using V_{O_x} as the predictor is that it is measured at the same sensor board at the same instant as V_{NO_2} , sharing the local O_3 environment without spatial lag and capturing common-mode electrode drift between the two channels. This is the best available estimate of the zero floor without a colocated O_3 reference.

Five candidate predictor sets for the Z_{NO_2} regression were compared: V_{O_x} alone, $V_{O_x} + T$, $V_{O_x} + RH$, $V_{O_x} + T + RH$, and $T + RH$ without any electrode term. Adding T or RH to the V_{O_x} -only model produced negligible improvement in cross-validated R^2 across both Providence and Bay Area nodes (Figure S13), confirming that the electrode voltage already captures the relevant variability. The V_{O_x} -only model is retained as the default for parsimony. Per-month regression diagnostics (α , β , R^2 , mean Z_{NO_2} , and background-hour temperature statistics) are stored for each node and presented in Figures S11–S13.

A systematic sensitivity analysis of the regression window length and NO_2 percentile threshold — the two parameters that define which hours enter the regression — is presented in Figure S25. The interaction between these parameters is weak, and the adopted configuration (3-month window, 10th percentile) lies in the optimal region of the joint parameter space. The effect of the scaling factor λ (testing whether the raw estimate over- or under-estimates the effective zero level) is shown in Figure S26.

S-Methods-2: CO_2 enhancement index details

The index framework was introduced by Kim et al. [44] and refined by Winter et al. [45], who select hours in the lowest 10th percentile of relative standard deviation across reference sites as training data for calibration. The modification adopted here computes the coefficient of variation on CO_2 enhancement rather than on absolute CO_2 concentration or reference-station relative standard deviation. Enhancement is defined as the difference between each node’s measured CO_2 and a rolling network background, computed as the 5th percentile of the network median CO_2 stratified by hour of day over a 3-week window. This isolates local source variability: the ~ 420 ppm common-mode background would otherwise compress all CoV values and reduce selectivity. The same enhancement-based index is used in all three cities. This CO_2 -based approach generalises to cities where only one reference station is available, unlike the reference-station relative standard deviation used by Winter et al. [45].

S-Methods-3: Stage 3 offset design rationale

Among the Stage 3 configurations evaluated, a temperature-dependent monthly offset $b(T, m)$ with $5^\circ C$ temperature bins was adopted over a scalar monthly offset $b(m)$. The T-dependent surface addresses residual temperature structure in the post-gain-correction signal that a single scalar cannot capture — particularly the tendency for background hours to be temperature-biased toward nighttime and winter conditions. At months or bins where temperature coverage is insufficient, the offset falls back to the scalar form automatically, ensuring robustness without sacrificing the additional correction where data support it. A direct comparison

of the scalar and temperature-dependent offset formulations is presented in Figure S26c. Sensitivity to window length and percentile threshold is shown in Figure S22.

Supplementary Tables

Table S1: AirNow reference station metadata for all three BEACO₂N deployments. Data capture rate is the annual percentage of valid hourly records passing quality screening. Distance to nearest node is the minimum node-to-station distance within each city. N assigned is the number of sensor nodes whose Stage 1 and 3 calibration is anchored to that station.

Station name	AQS ID	Species	Capture (%)	Dist. (km)	N assigned
<i>SF Bay Area</i>					
San Pablo	[ID]	NO, NO ₂ , O ₃	[%]	[km]	16
EBMUD	[ID]	NO, NO ₂ , O ₃	[%]	[km]	9
SF	[ID]	NO, NO ₂ , O ₃	[%]	[km]	7
Vallejo	[ID]	NO, NO ₂ , O ₃	[%]	[km]	7
Oakland East	[ID]	NO, NO ₂ , O ₃	[%]	[km]	3
Laney College	[ID]	NO, NO ₂	[%]	[km]	1
RFS	[ID]	NO, NO ₂ , O ₃	[%]	[km]	—
<i>Los Angeles</i>					
N. Main St.	[ID]	NO, NO ₂ , O ₃	[%]	[km]	9
West LA	[ID]	NO, NO ₂ , O ₃	[%]	[km]	5
<i>Providence</i>					
East Providence	[ID]	NO, NO ₂ , O ₃	[%]	[km]	16
South Providence	[ID]	NO, NO ₂ , O ₃	[%]	[km]	7

Laney College lacks O₃ and is excluded from the NO₂ calibration index (Section 3.2.4). RFS (Richmond Field Station) is used solely for derivation of the WLS error profile; it does not serve as a calibration reference for any sensor node. Colocated nodes: site 43 at San Pablo, site 42 at EBMUD, site 45 at Laney College.

Table S2: Sensitivity analysis: parameter ranges and outcome summary. For each parameter varied, the table reports the default value, range tested, and physical motivation. Tests on event detection parameters (Group A) were conducted on the Providence network; pipeline parameters (Groups B–I) use the Bay Area validation site unless noted.

Parameter	Default	Values tested	Physical motivation
<i>Event detection (Group A)</i>			
max_base_span (ppb)	6.0	3.0, 5.0, 6.0	Baseline stability
end_k	5	2, 5, 8	Stabilisation strictness
dual_end	True	True, False	Both vs one channel
Night window	20–05	19–05, 20–05, 21–04, 22–04	Capture vs contamination
R_CLIP_FIT	(0.15, 1.80)	(0.10–0.25, 1.80)	Accepted slope range
Temp. bin width (°C)	2	1, 2, 5	Resolution vs stability
<i>WLS and index (Group B)</i>			
WLS window (months)	9	3, 5, 7, 9, 12	Smoothing vs drift tracking
CO ₂ index cap (%)	25	5, 10, 15, 20, 25	Sample size vs quality
<i>Offset (Group B)</i>			
Offset window (months)	5	3, 5, 7	Stability vs responsiveness
Offset percentile	15th	5th, 10th, 15th, 20th, 25th	Background hour strictness
Offset formulation	$b(T, m)$	scalar $b(m)$, $b(T, m)$	T-dependent vs scalar
<i>NO₂ zero level (Group E)</i>			
Z _{NO₂} window (mo)	3	1, 2, 3, 4, 6	Zero-floor responsiveness
Z _{NO₂} NO ₂ pct1	10th	2nd, 5th, 10th, 15th, 25th	Background hour definition
Z _{NO₂} predictor	V _{O_x}	+T, +RH, +T+RH, T+RH	Complexity vs parsimony
Z _{NO₂} scale (λ)	0.7	0–1.25	Over-/under-estimation
<i>Gain curve robustness (Groups G–I)</i>			
Event R ² threshold	0.50	0.0, 0.3, 0.5, 0.7, 0.8	Stoichiometric quality
T _{CAP} (°C)	35	25, 28, 30, 33, 35, 40, 45	Sparse-event bypass

Table S3: Progressive comparison results at site 42 (EBMUD truth), evaluated against the withheld colocated reference on hourly data. Colocated methods use on-site reference hours and yield single values; remote methods report the median across four remote reference assignments. Slope is from OLS regression of calibrated values against truth (ideal = 1.0). Compare with main-text Table 3.5 (site 43).

Method	RMSE (ppb)	NRMSE	Slope	R ²	MAE (ppb)	Bias (ppb)
<i>Colocated reference</i>						
MLR	3.82	0.40	0.77	0.78	2.79	−0.24
WLS+GC	4.86	0.51	1.11	0.78	3.42	+0.58
<i>Remote reference (median over 4 pairs)</i>						
WLS (CO ₂ index)	5.50	0.58	0.53	0.72	3.91	−3.04
WLS+GC (O ₃ -scaled idx)	5.31	0.56	0.79	0.74	3.88	−3.03

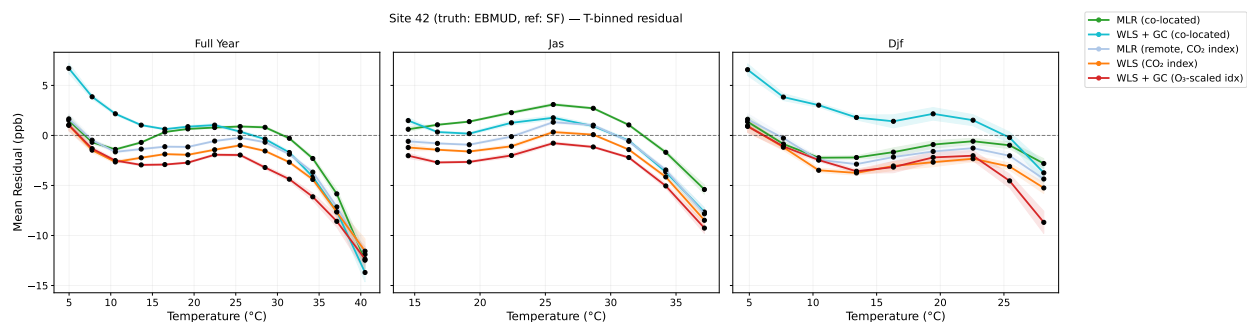


Figure S1: Progressive comparison results at site 42 (EBMUD truth), in the same format as main-text Table 3.5: four rows $\times$ three evaluation windows (full year, JAS, DJF), metrics = RMSE (ppb), bias (ppb), OLS slope, R². Results are consistent with site 43 (main text), confirming that the progressive comparison findings are not specific to a single site or remote reference assignment.

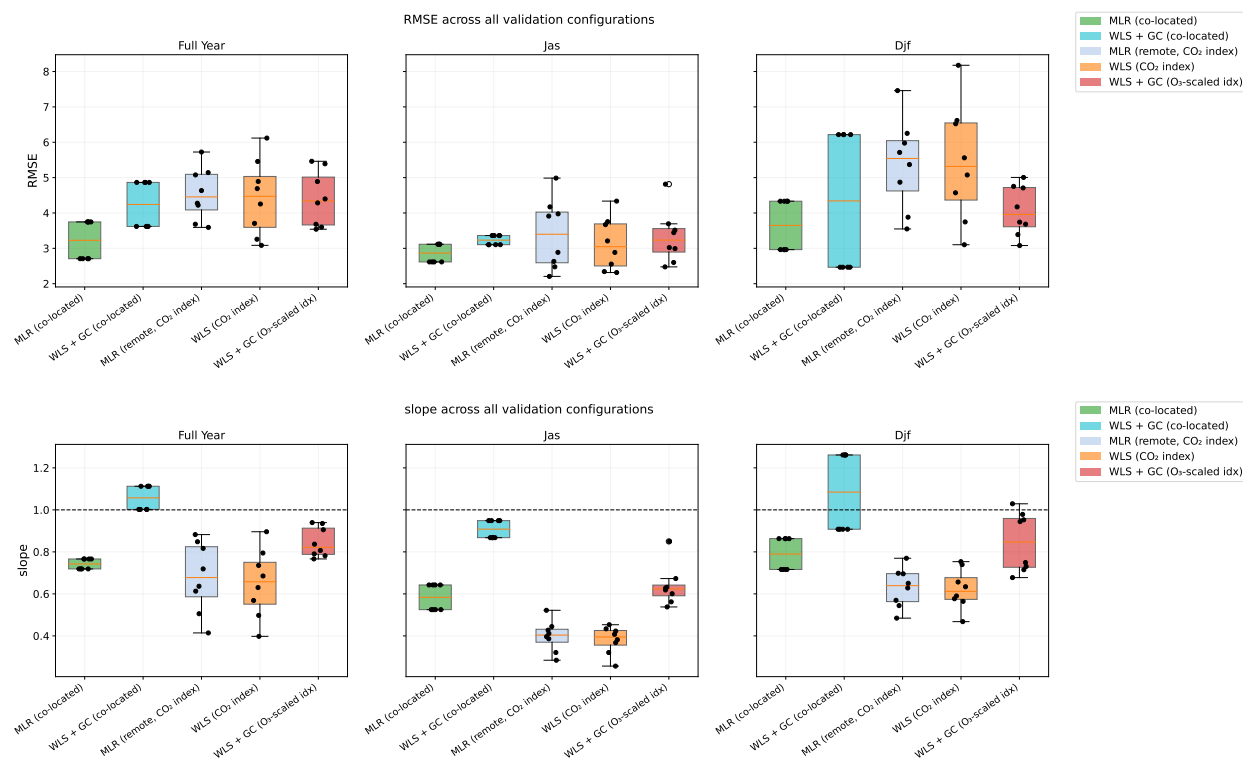


Figure S2: Seasonal RMSE (ppb; top) and OLS slope (bottom) for all five calibration methods (Table 3.2) at both validation sites (43 and 42), presented as box plots across all validation configurations. Rows correspond to the three evaluation windows (full year, JAS, DJF); columns correspond to the two metrics. Each box spans the interquartile range across jobs for a given method and site; the horizontal line shows the median. JAS consistently shows the largest improvement from the gain correction (WLS+GC vs WLS remote), consistent with warm temperatures and elevated O₃ producing the strongest gain error in summer. DJF results confirm that the correction remains effective in cold conditions. Colocated methods are shown as horizontal reference lines (single values, not varying across jobs).

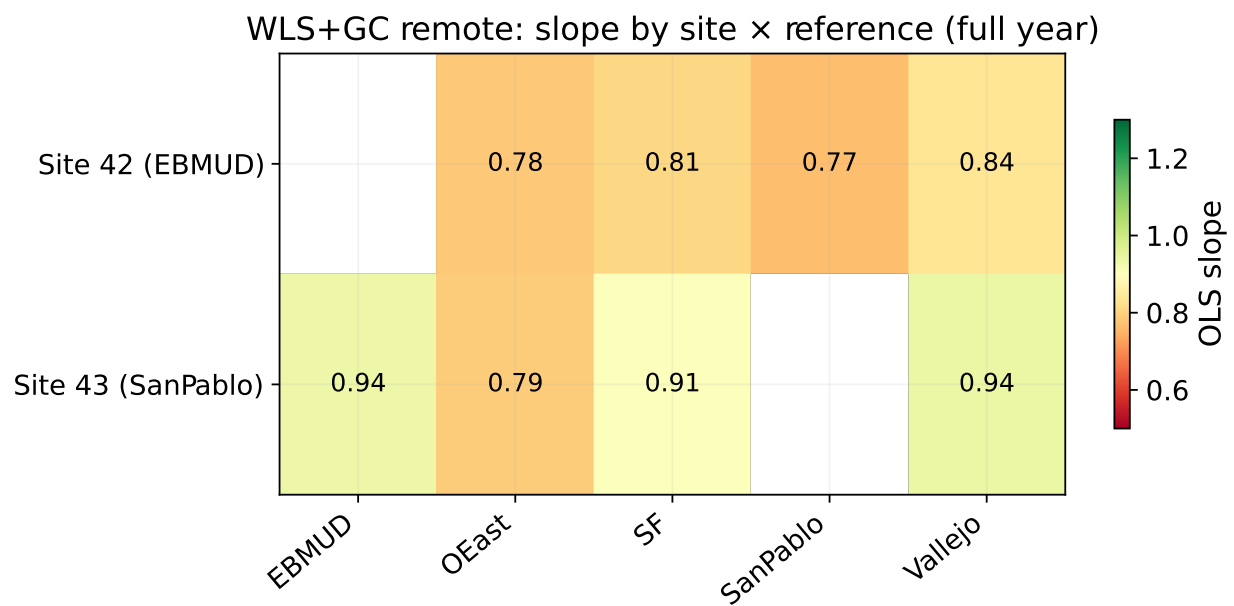


Figure S3: Per-pair OLS slope heatmap for the WLS+GC remote pipeline (Row 5), showing slope against the withheld colocated truth for all combinations of test site (rows) and remote reference station (columns). Each cell reports the mean slope across month strategies. Colour scale: green = slope near 1; red = slope well below 1. The heatmap confirms that the gain correction produces consistent slope recovery across different remote reference assignments.

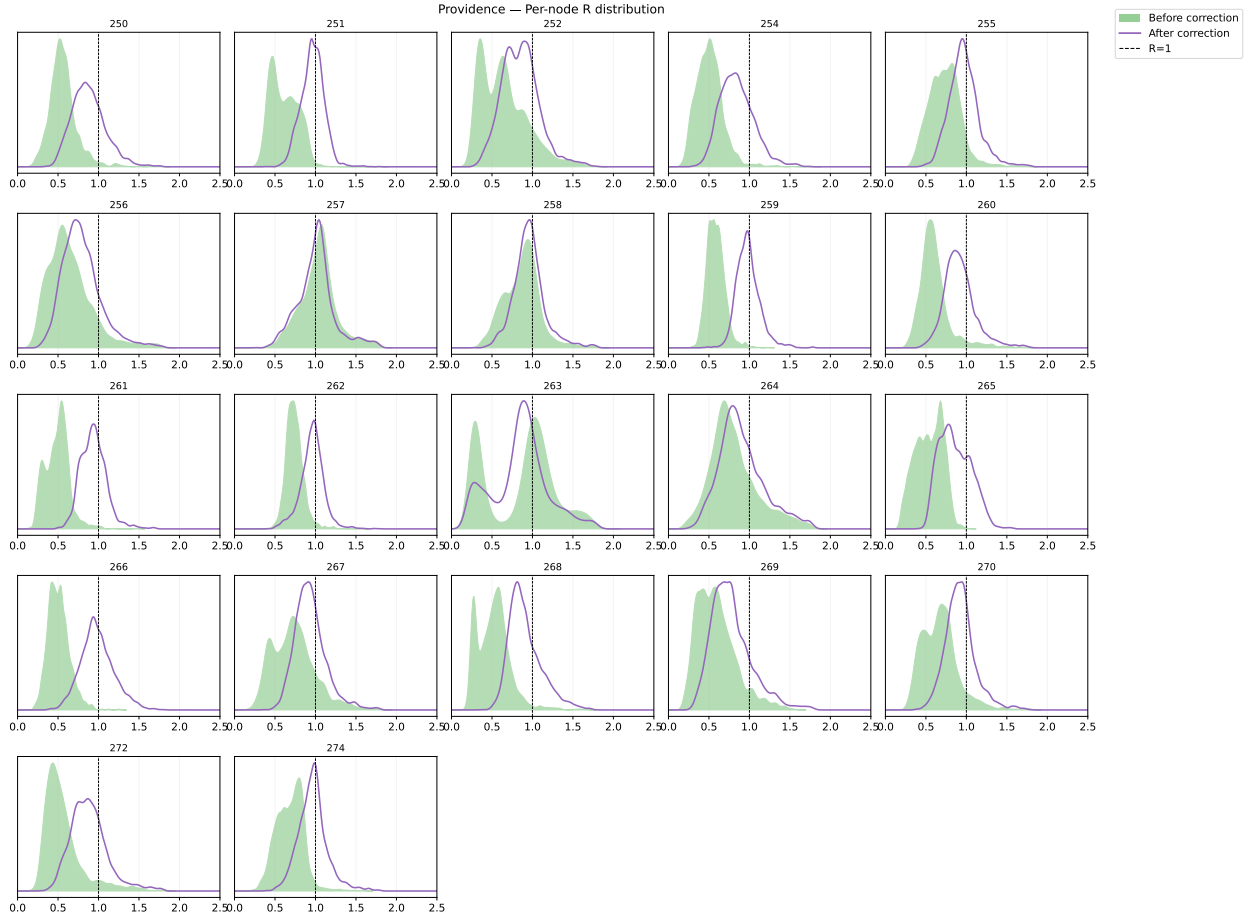


Figure S4: Kernel density estimates (KDE) of the titration ratio R before (grey fill) and after (red line) per-node gain and offset correction for all 22 Providence nodes, arranged in a grid ordered by node ID. The dashed vertical line in each panel marks $R = 1$. Median R before and after is annotated in each panel. All panels share an identical x -axis range ($R \in [0, 2]$) for direct comparison. The anomalous node ([ID]; red border) retains a post-correction median R that does not converge toward unity, consistent with its positive dg/dT slope.

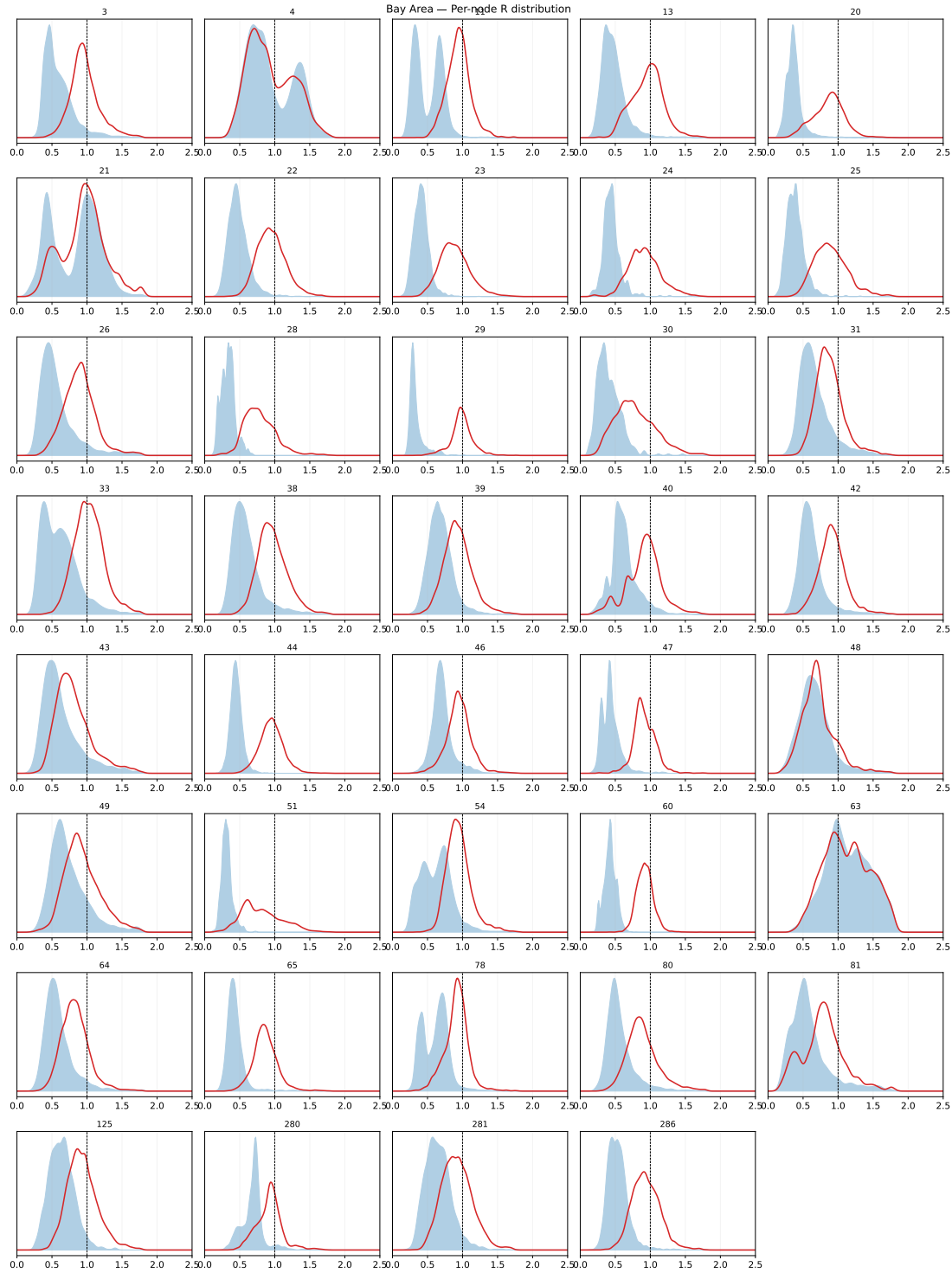


Figure S5: Per-node KDE of R before (grey fill) and after (red line) correction for all nodes in the San Francisco Bay Area, in the same format as Supplementary Figure S4.

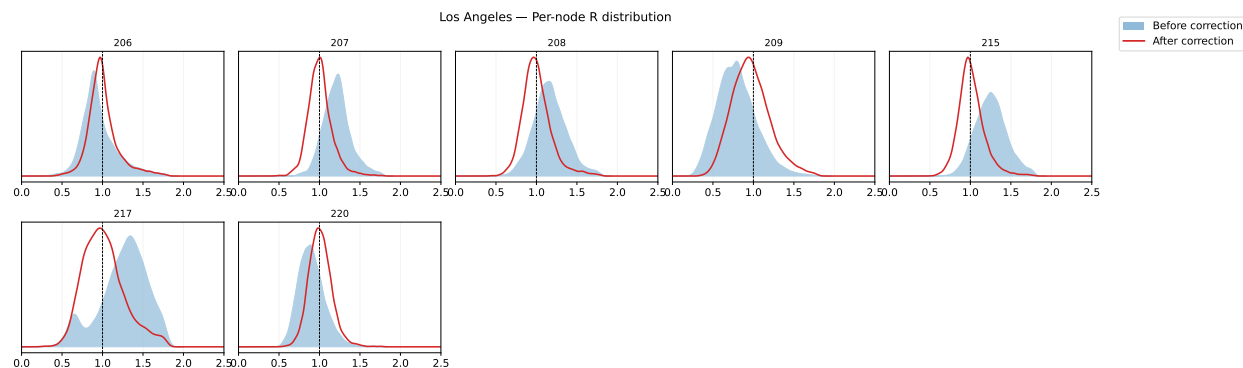


Figure S6: Per-node KDE of R before (grey fill) and after (red line) correction for all nodes in Los Angeles, in the same format as Supplementary Figure S4. A residual post-correction negative skew is visible at most nodes and is most pronounced at eastern basin nodes where O_3 is highest.

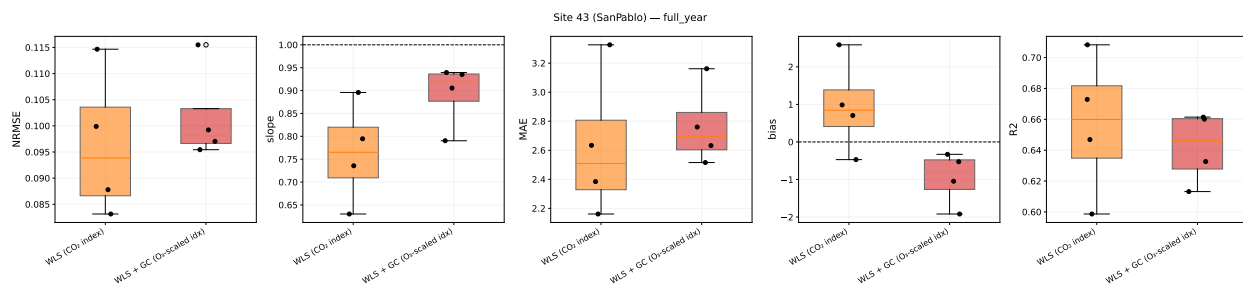


Figure S7: Calibration performance spread across remote reference assignments for both Bay Area validation sites (43 top, 42 bottom). Panels show RMSE (ppb), OLS slope, and R^2 evaluated against the withheld colocated truth. Each point represents one remote reference station; box plots show the interquartile range across references. Three remote calibration methods are compared: MLR (CO₂ index), WLS (CO₂ index), and WLS+GC (full pipeline). WLS+GC consistently achieves the highest slope and tightest cross-reference spread, confirming that the gain correction reduces dependence on which remote reference station is used for calibration.

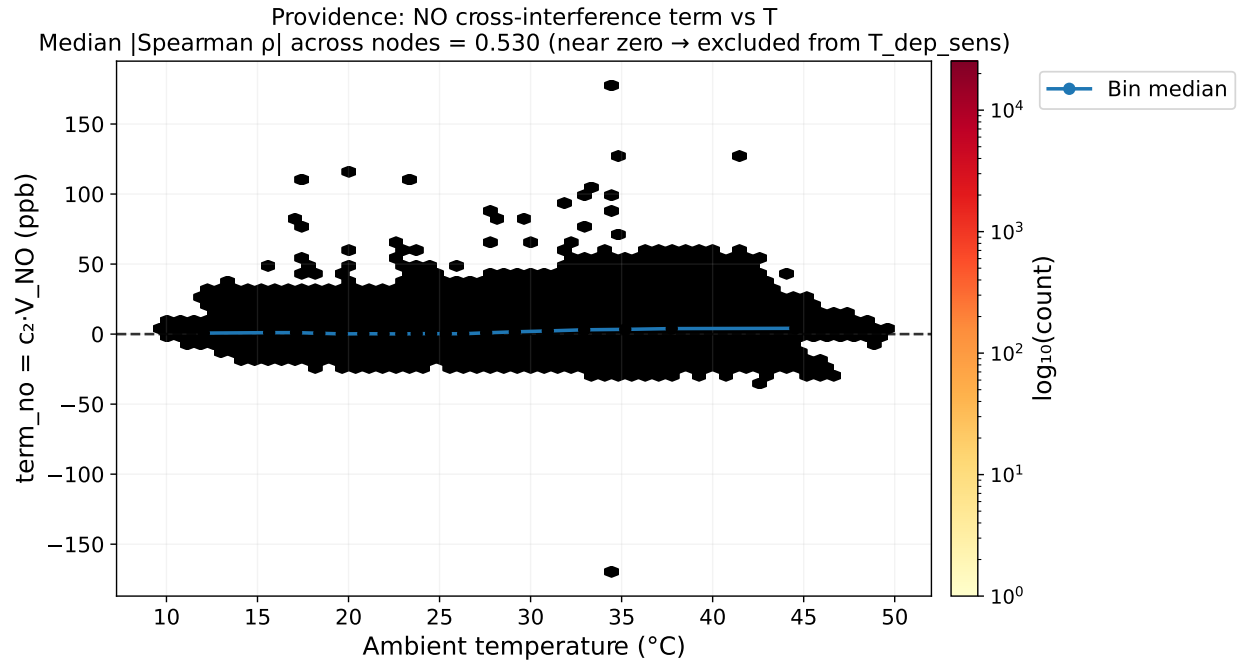


Figure S8: Temperature dependence of the NO cross-interference term ($\text{term_no} = \beta_{\text{NO}} \cdot V_{\text{NO}}$, ppb) pooled across all Providence nodes, shown as a hexbin density plot on a logarithmic colour scale. Blue points show the bin median. The dashed line marks zero. The bin medians remain near zero across the full temperature range, with no systematic trend. The median |Spearman ρ | across nodes is 0.530, but reflecting scatter rather than a coherent directional bias. This near-zero median term_no justifies excluding it from the gain-scaled sum S_{gain} (Equation B.1; S-Methods-1), and confirms that the gain correction need not be applied to the NO cross-interference term.

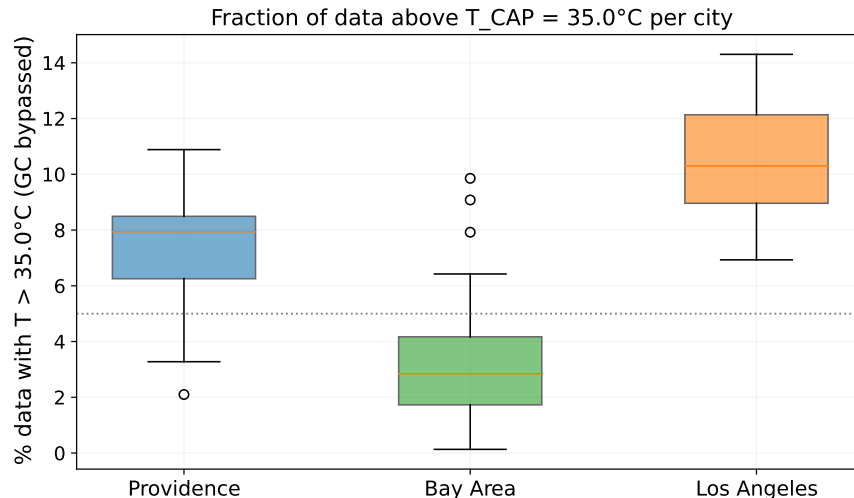


Figure S9: Fraction of minute-level NO_2 data for which $T > T_{CAP}$ and the WLS output is therefore preserved unchanged (gain correction bypassed) in each city. All cities use $T_{CAP} = 35^{\circ}\text{C}$. The fraction of data above the cap is small in all cities, indicating that the T_{CAP} guard does not substantially limit the temporal coverage of the correction.

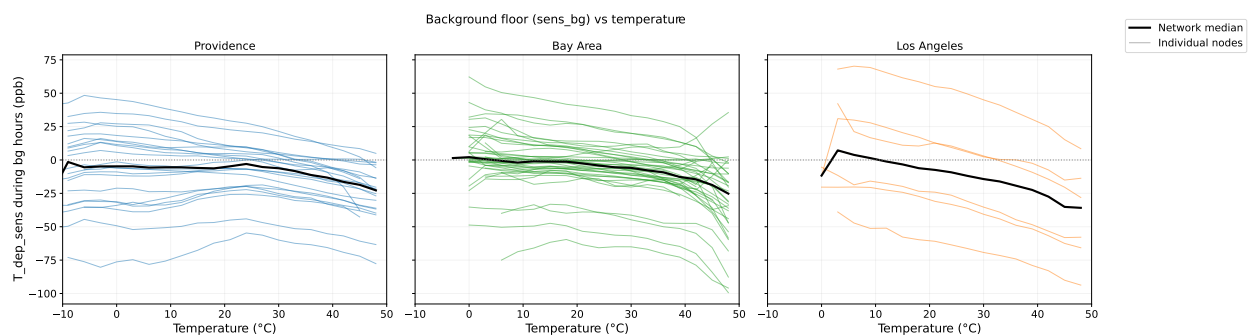


Figure S10: Mean NO_2 zero-level estimate Z_{NO_2} (ppb) as a function of ambient temperature for all nodes in each city. Three panels: (a) Providence, (b) San Francisco Bay Area, (c) Los Angeles. Thin grey lines = individual nodes; bold coloured line = network median. The floor is positive in all cities, reflecting the O_3 cross-sensitivity contribution to S_{gain} under near-zero NO_2 conditions. A residual temperature dependence of Z_{NO_2} is visible — the floor is systematically lower at warm temperatures in all three cities.

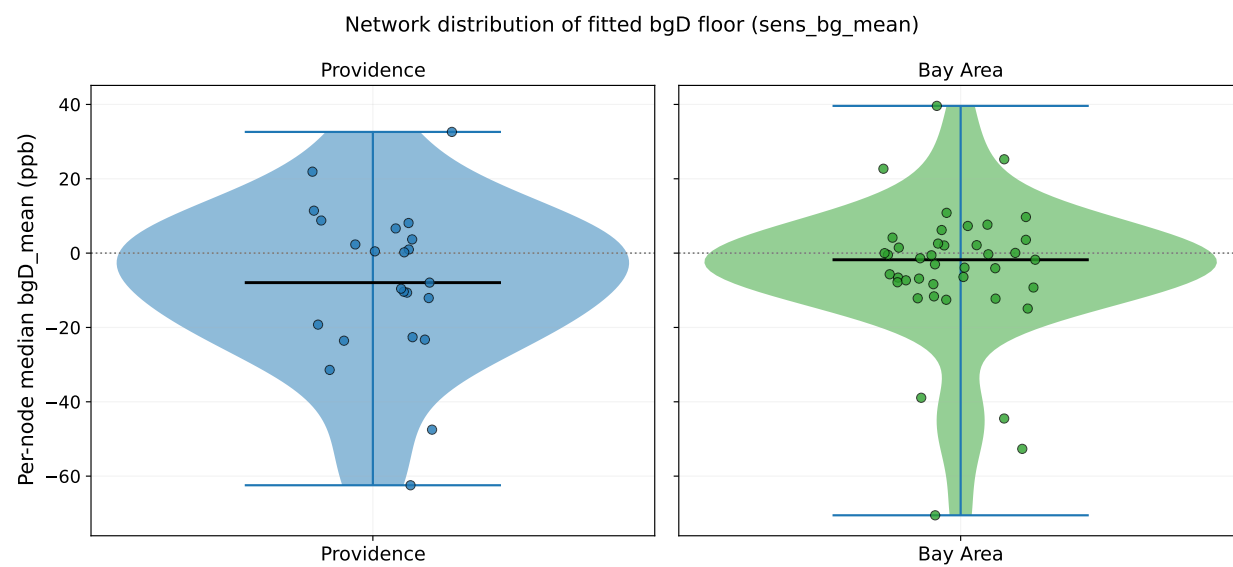


Figure S11: Z_{NO_2} regression diagnostics, part a: network distribution of the fitted zero level (mean Z_{NO_2}) across nodes and cities. Continued in Figures S10b–c.

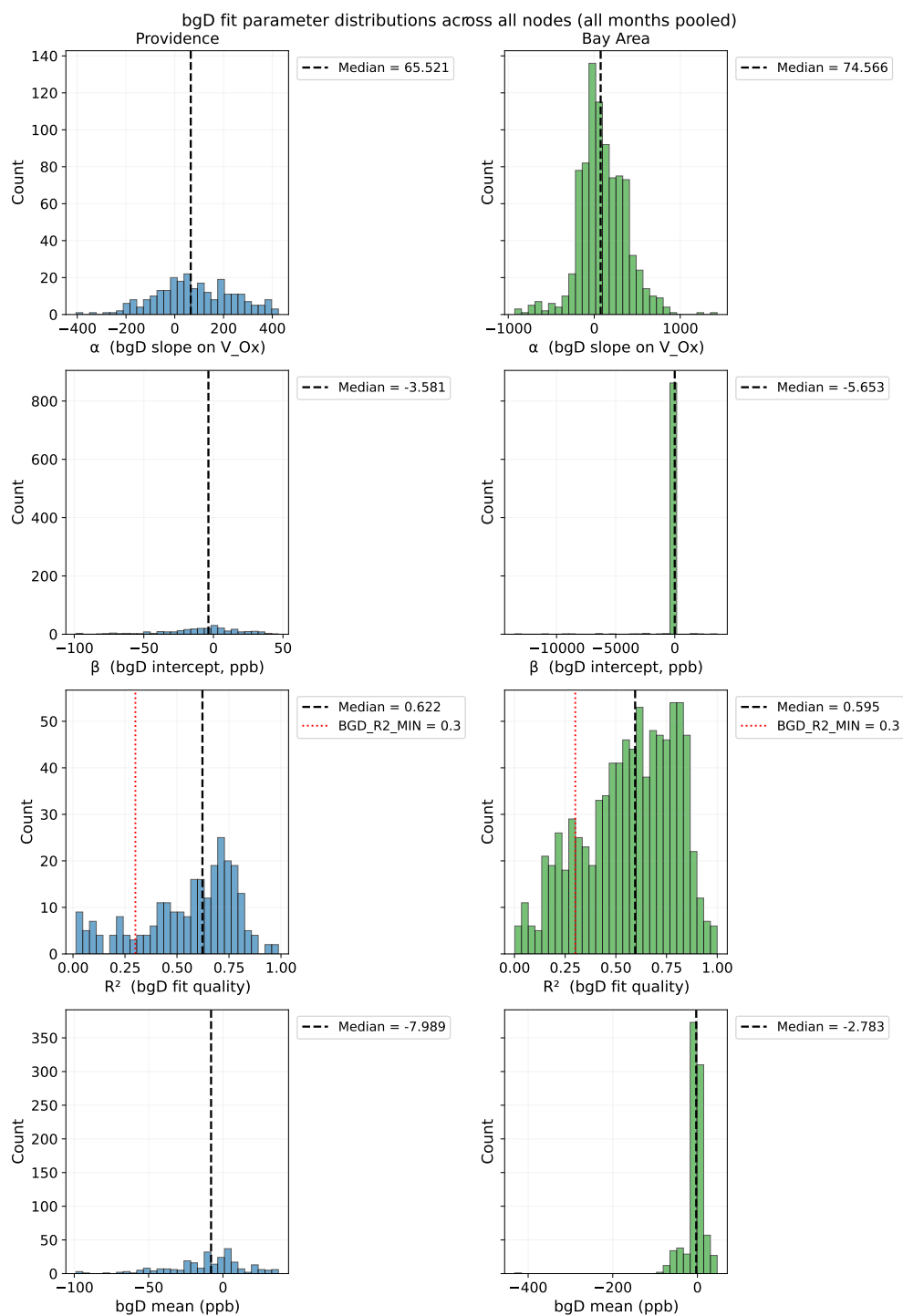


Figure S12: Z_{NO_2} regression diagnostics, part b: fit parameter distributions (α , β , R^2) across all nodes, all months pooled.

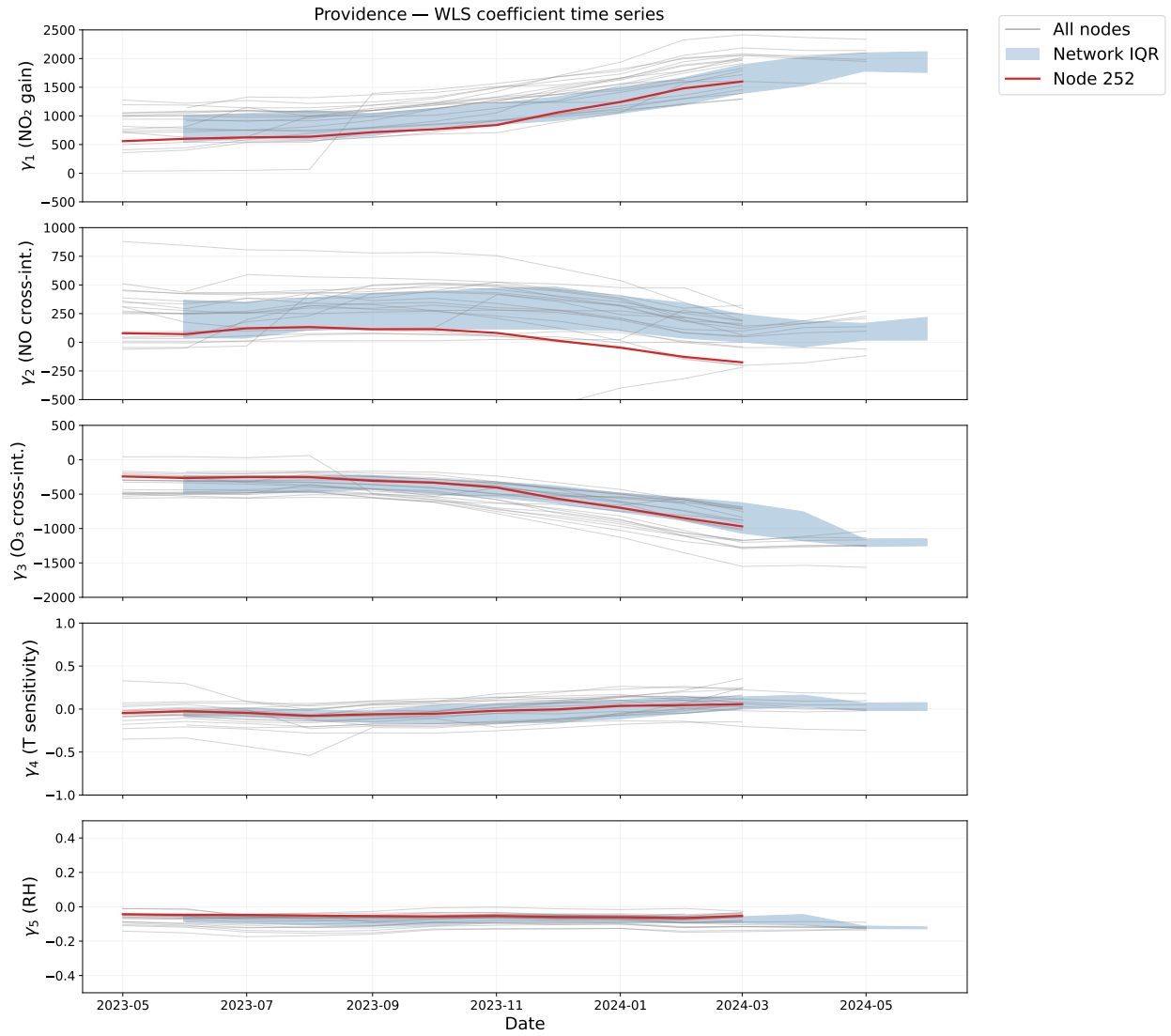


Figure S14: Rolling WLS coefficient time series, Providence. Subpanels show β_{NO_2} (NO₂ gain), β_{NO} (NO cross-interference), β_{O_x} (O₃ cross-interference), β_T (temperature sensitivity), and β_0 (intercept). Thin grey lines = all nodes; bold coloured line = representative node. Continued in Figures S11b–c.

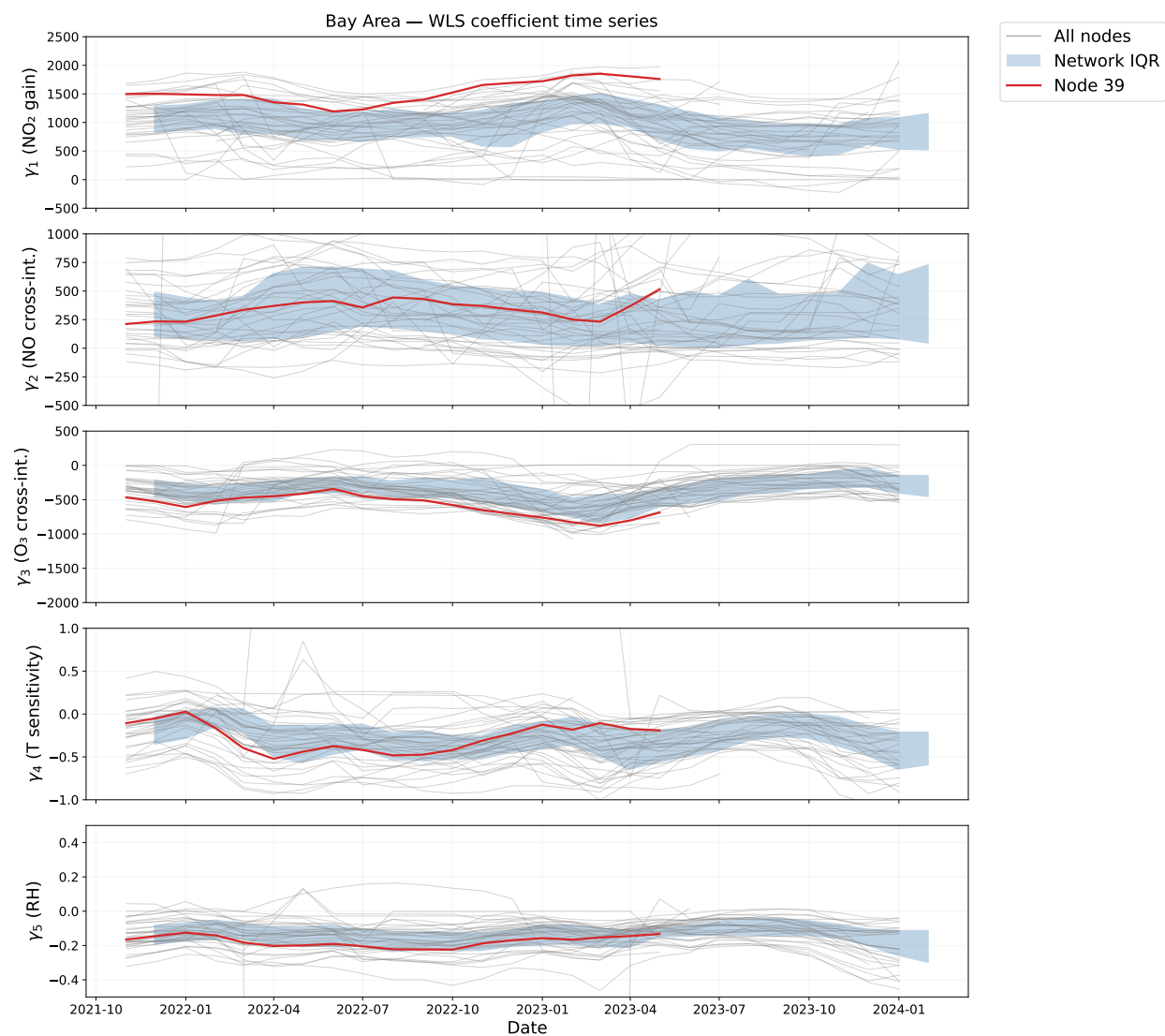


Figure S15: Rolling WLS coefficient time series, San Francisco Bay Area. Same format as Figure S11a.

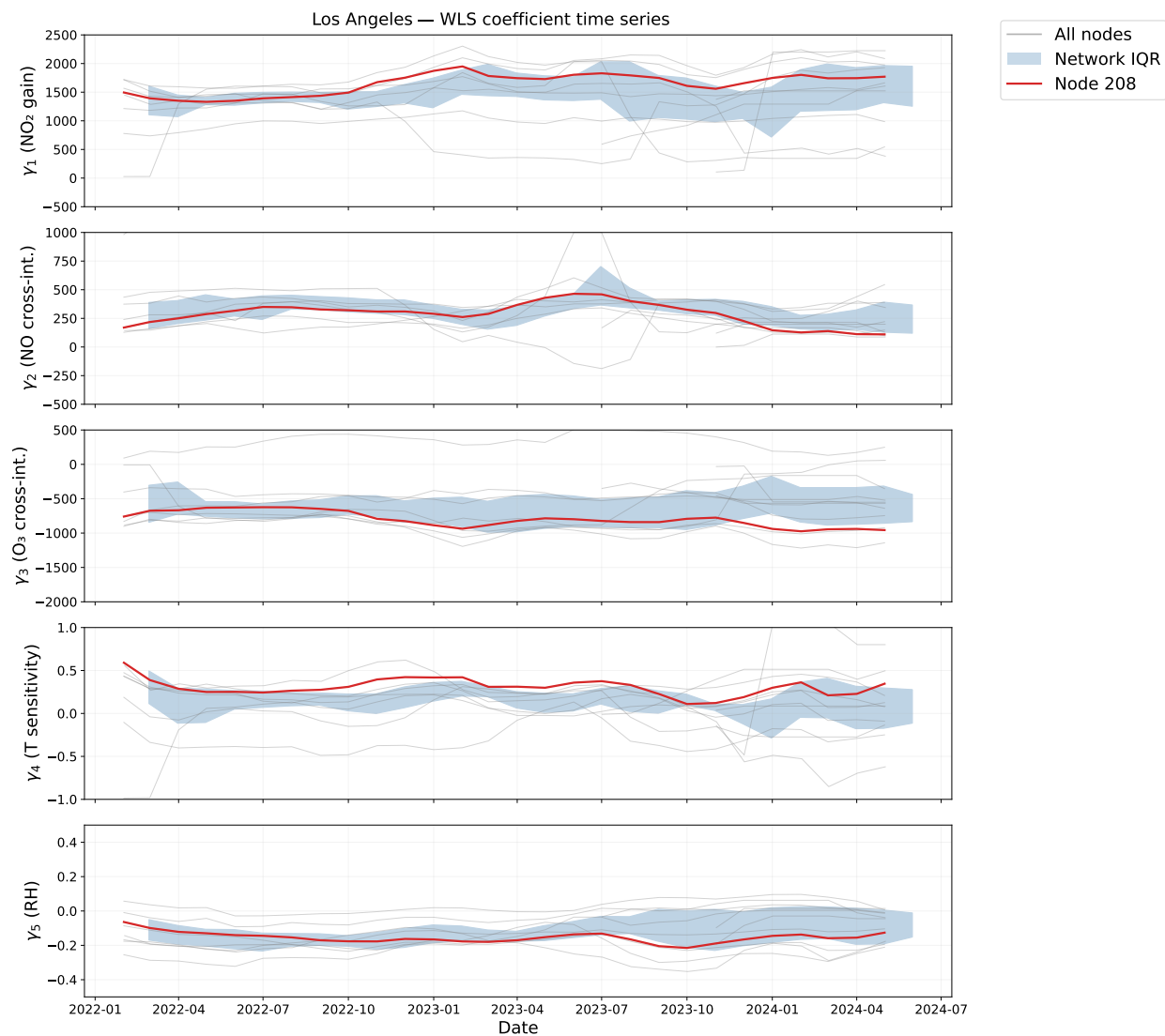


Figure S16: Rolling WLS coefficient time series, Los Angeles. Same format as Figure S11a. The sign and magnitude of each coefficient are consistent with known B43F cross-sensitivities and are stable over time across all three cities, confirming that Stage 1 is well-behaved.

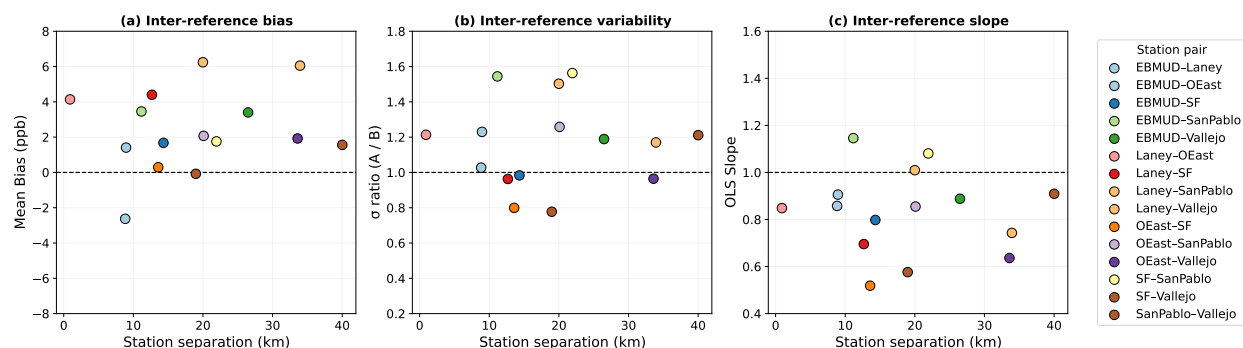


Figure S17: Pairwise mean NO_2 bias (ppb), RMSE, and correlation among EPA reference stations in the San Francisco Bay Area, computed over the study period as a function of inter-station distance. Pairwise biases represent the irreducible spatial heterogeneity in NO_2 across the Bay Area regulatory network. This range is used in the main text as the spatial representativeness benchmark against which post-correction bias is judged.

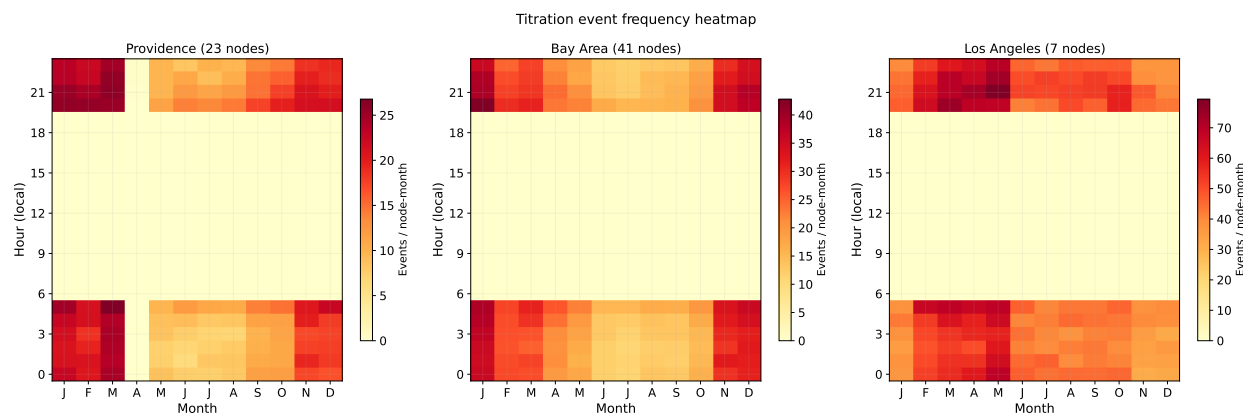


Figure S18: Titration event frequency as a function of calendar month (x-axis) and local hour of day (y-axis) for the three BEACO₂N cities: (a) Providence, (b) San Francisco Bay Area, (c) Los Angeles. Colour indicates the number of events detected per node-month in each month-hour bin, averaged across all nodes. Events are most frequent in Providence during winter evenings (November–March, 17:00–23:00), where long nights and sustained heating emissions produce frequent fresh- NO pulses into an O_3 -rich residual layer. Bay Area events are similarly concentrated but show suppression in summer months due to marine-layer O_3 depletion. Los Angeles events are more evenly distributed year-round, consistent with narrower seasonal variation in nocturnal O_3 and NO in the basin.

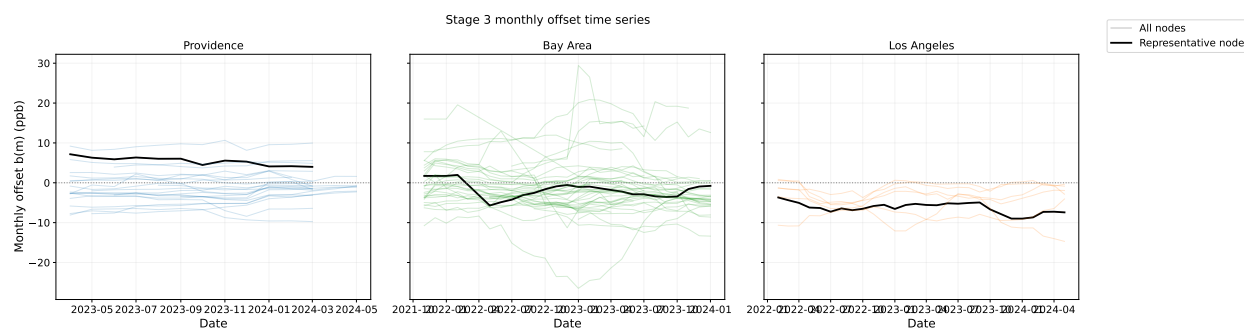


Figure S19: Rolling monthly midday offset $b(m)$ (ppb) applied in Stage 3, shown as a time series for all nodes in each city. Three panels: (a) Providence, (b) San Francisco Bay Area, (c) Los Angeles. Thin grey lines = individual nodes; bold coloured line = representative node. Providence shows larger positive offsets in winter, consistent with a mild residual sensor low-bias at cold temperatures not fully resolved by the gain correction and with the seasonal spatial gradient between nodes and reference stations. Los Angeles shows a persistent small positive offset in summer, consistent with residual O_3 cross-sensitivity under high-ozone conditions. The offset magnitude is generally smaller than the gain correction at any individual temperature but provides a meaningful improvement in long-term mean bias.

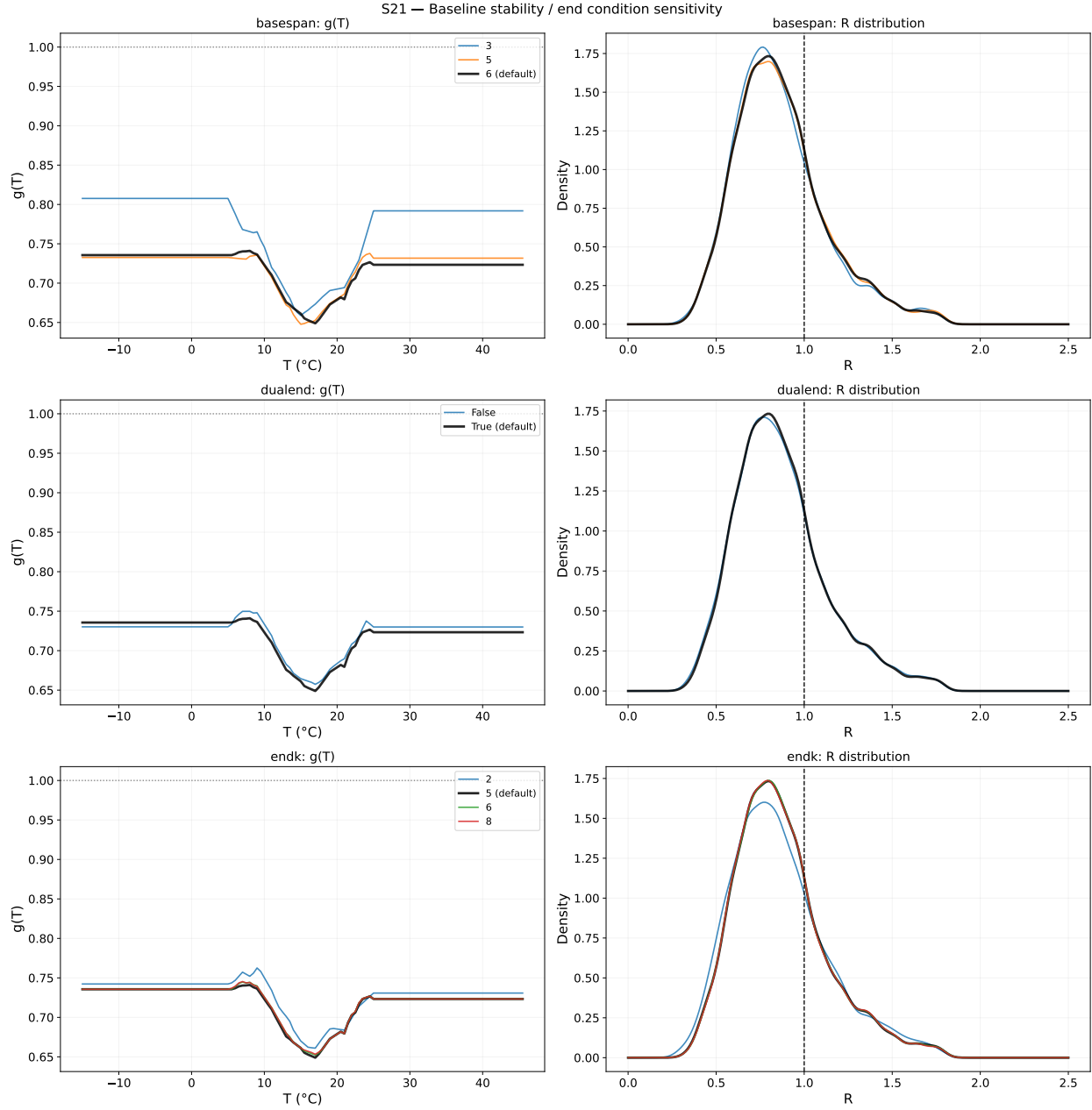


Figure S20: OAT event detection sensitivity analysis, part 1: baseline stability (`max_base_span`), end condition (`end_k`), and end mode (`dual_end`). For each parameter, the figure shows: **left** — median per-node $g(T)$ with IQR shading for each tested value, compared to the default (bold black); **right** — post-correction KDE of R (network-pooled) for each value. Parameter values tested are listed in Supplementary Table S3. The gain curve changes by less than 0.04 units at any temperature bin across all tested configurations, confirming robustness of the gain correction to reasonable variation in these detection parameters.

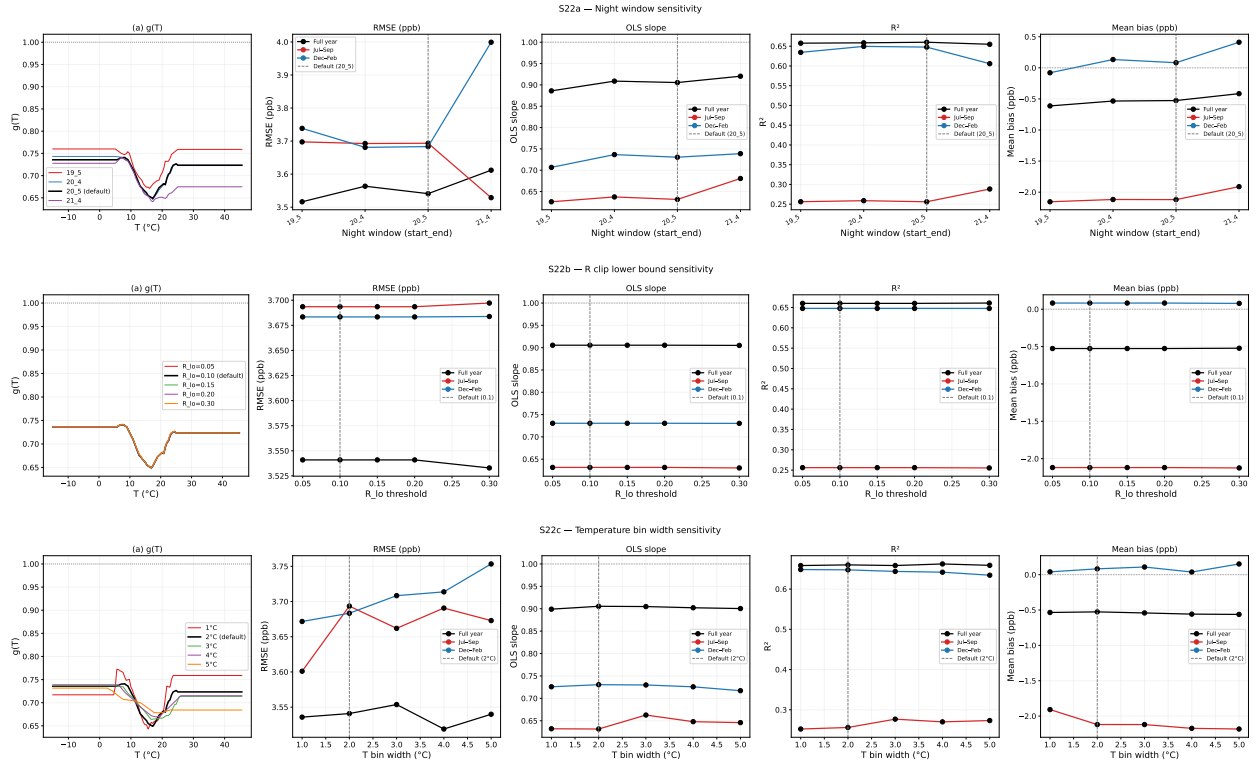


Figure S21: OAT event detection sensitivity analysis, part 2: **(a)** night window definition (19:00–05:00 default, 21:00–04:00, 22:00–04:00); **(b)** $R_{\text{CLIP_FIT}}$ lower bound; **(c)** temperature bin width (1°C, 2°C default, 5°C). Each row shows $g(T)$ (left), post-correction R distribution, calibration metrics, event count, and related diagnostics. The 2°C default provides the best balance of resolution and stability.

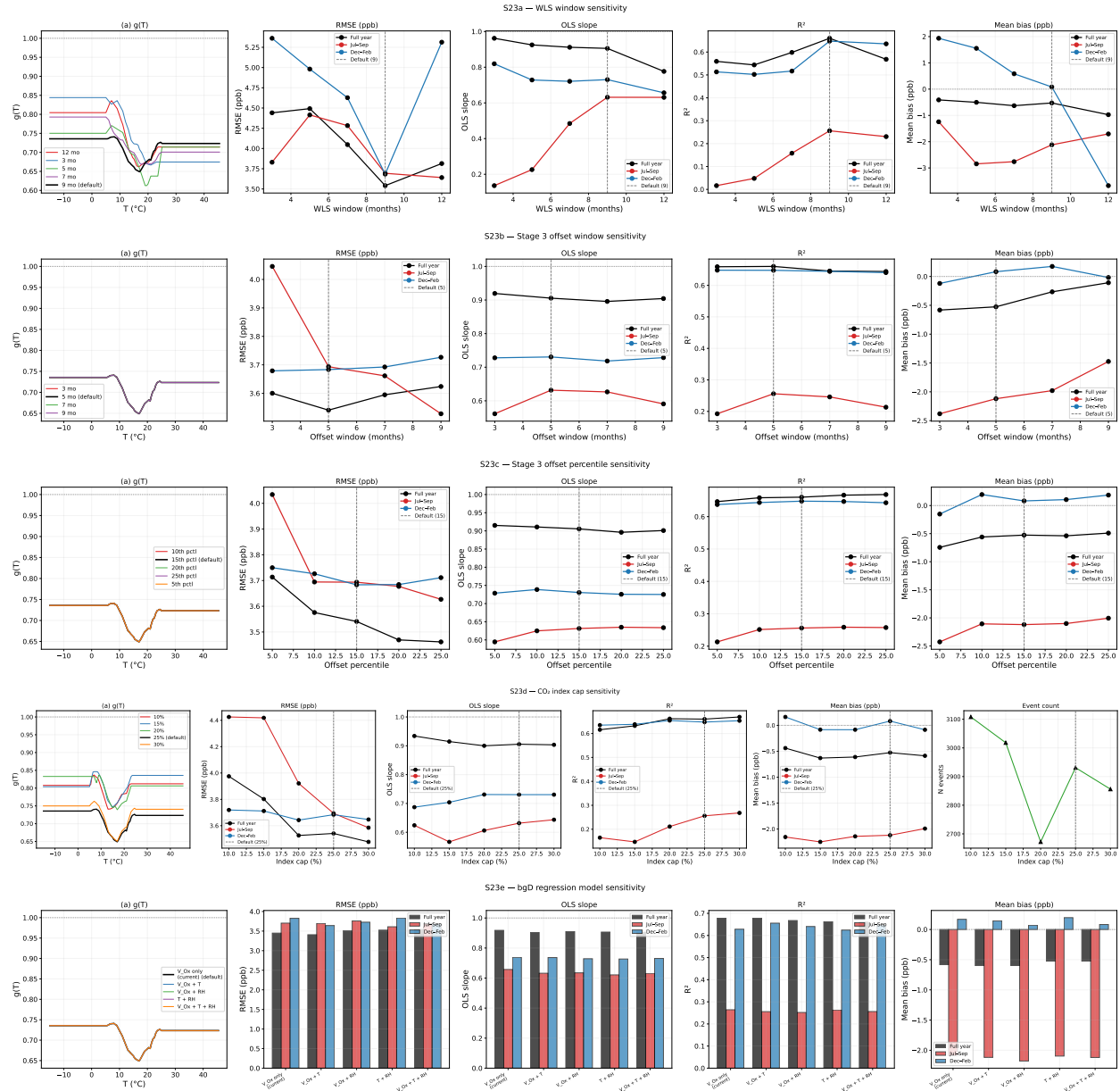


Figure S22: Sensitivity of calibration diagnostics to five pipeline parameters. Each row shows $g(T)$, post-correction R distribution, and calibration metrics for a swept parameter. (a) WLS regression window length (3, 5, 7, 9, 12 months). (b) Stage 3 offset window length (3, 5, 7 months). (c) Stage 3 offset percentile threshold (5th, 10th, 15th, 20th, 25th). (d) CO₂ enhancement index cap (10%–30%). (e) Z_{NO₂} predictor model (V_{O_x} alone, $+T$, $+RH$, $+T+RH$, $T+RH$). All models that include V_{O_x} produce nearly identical gain curves; the $T + RH$ model without the electrode term performs substantially worse. Final R diagnostics are comparable across all tested configurations.

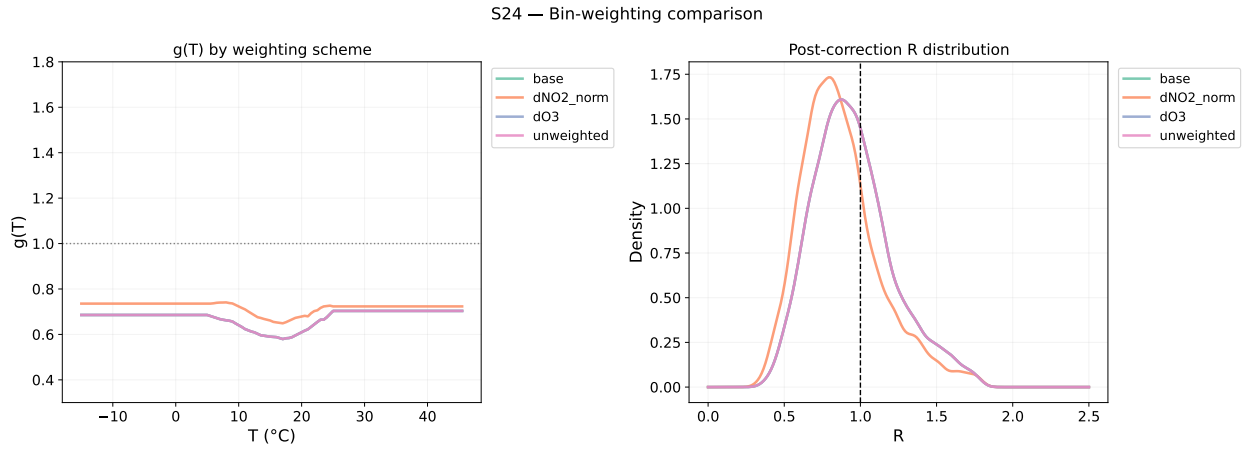


Figure S23: Comparison of four bin-weighting schemes for the gain curve construction. **Left:** median per-node $g(T)$ with IQR shading for each scheme. **Right:** KDE of the network-pooled R distribution after correction under each scheme, with the dashed vertical line at $R = 1$. Schemes: $w_{\Delta\text{NO}_2}$ (chosen; solid black), w_{base} (blue dashed), w_{unw} (unweighted median; orange dashed), $w_{\Delta\text{O}_3}$ ($|\Delta\text{O}_3|$ -weighted; green dashed). All four schemes produce statistically indistinguishable post-correction R distributions (Supplementary Table S4), confirming that the main results are not sensitive to the choice of weighting scheme.

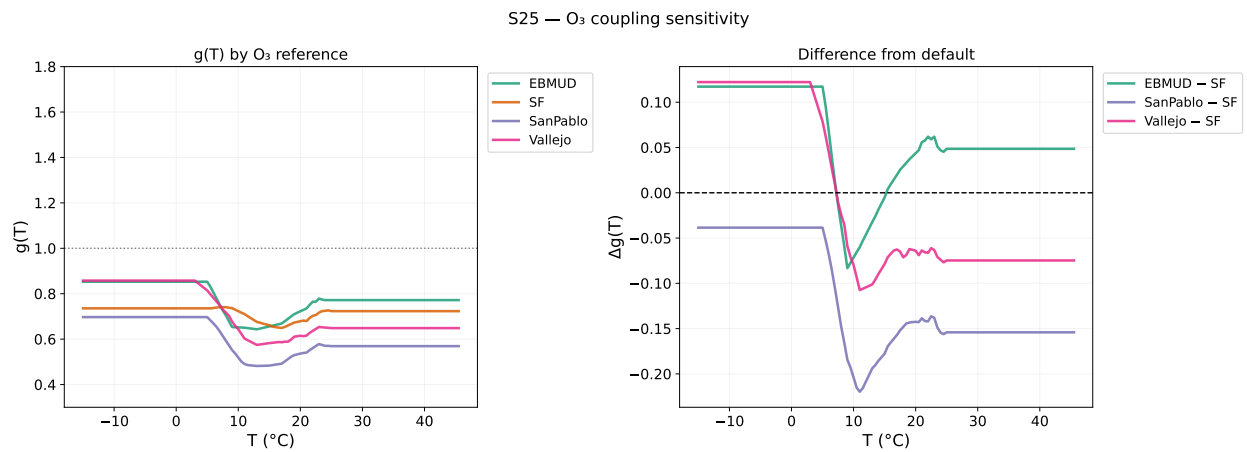


Figure S24: Sensitivity of the per-node gain curve $g(T)$ to the choice of O₃ reference station used in the gain correction, evaluated on the Providence network. **Left:** median per-node $g(T)$ and inter-node IQR (shading) computed using each node's assigned O₃ reference (default; black) versus alternative EPA O₃ stations. **Right:** distribution of the per-node absolute difference in $g(T)$ at each temperature bin between the default and each alternative O₃ reference. The curves differ by less than 0.03 units at any temperature bin, indicating that the circular dependency between O₃ and NO₂ calibrations does not materially bias the gain curve.

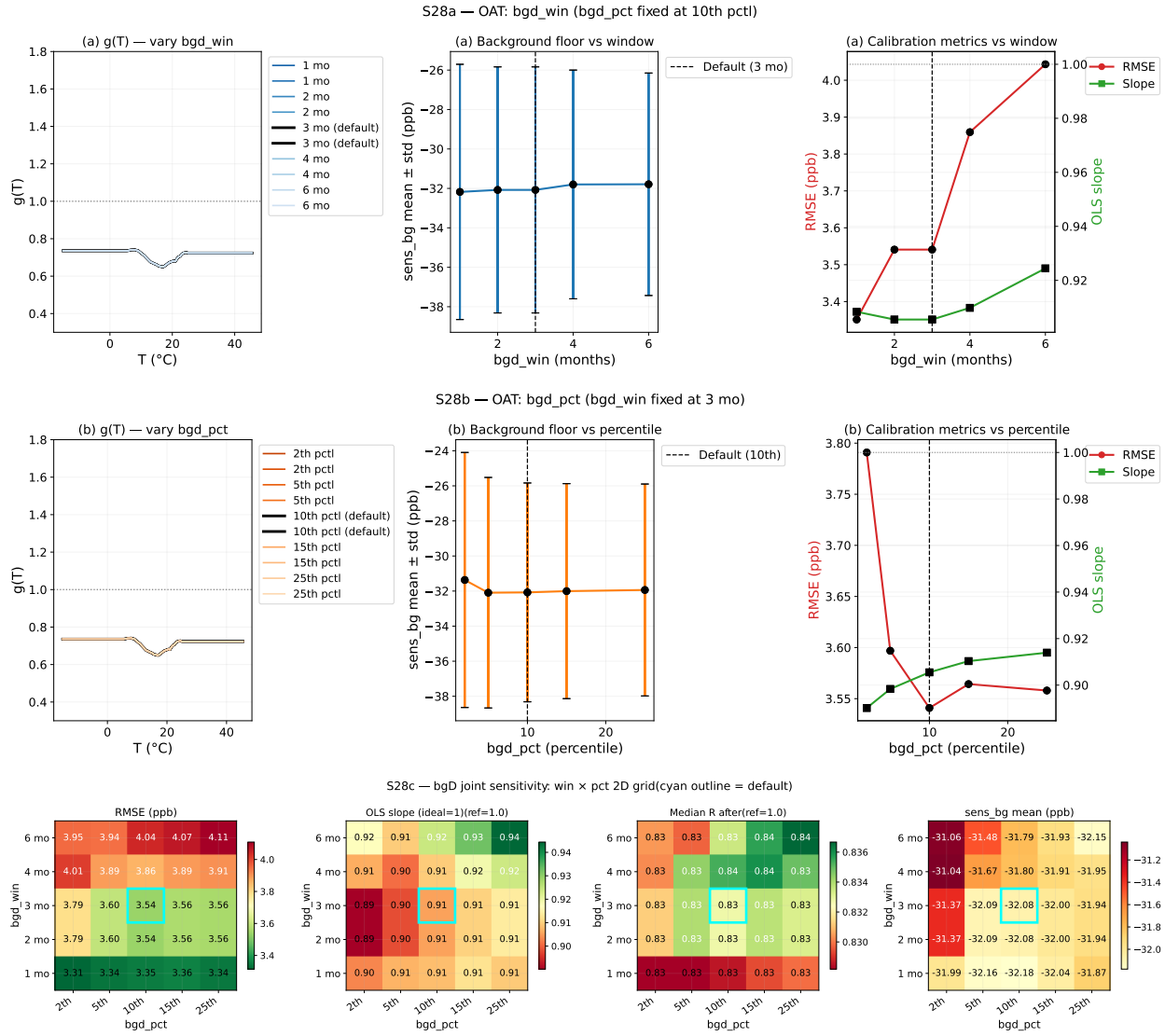


Figure S25: Joint sensitivity of the gain curve and calibration performance to the Z_{NO_2} regression window length and NO_2 background percentile threshold, evaluated on the Providence network. **(a)** Window length OAT: regression window varied across 1, 2, 3 (default), 4, and 6 months with the percentile held at 10th. Shorter windows (≤ 2 months) produce noisier α and β estimates; longer windows (≥ 4 months) smooth over within-season variation. **(b)** Percentile OAT: NO_2 percentile varied across 2nd, 5th, 10th (default), 15th, and 25th with the window held at 3 months. Higher percentiles include hours with non-negligible NO_2 , biasing the regression. **(c)** Joint 2D heatmap of calibration metrics across all window $\times$ percentile combinations. The adopted configuration (3 months, 10th percentile) lies in the optimal region. The interaction between the two parameters is weak.

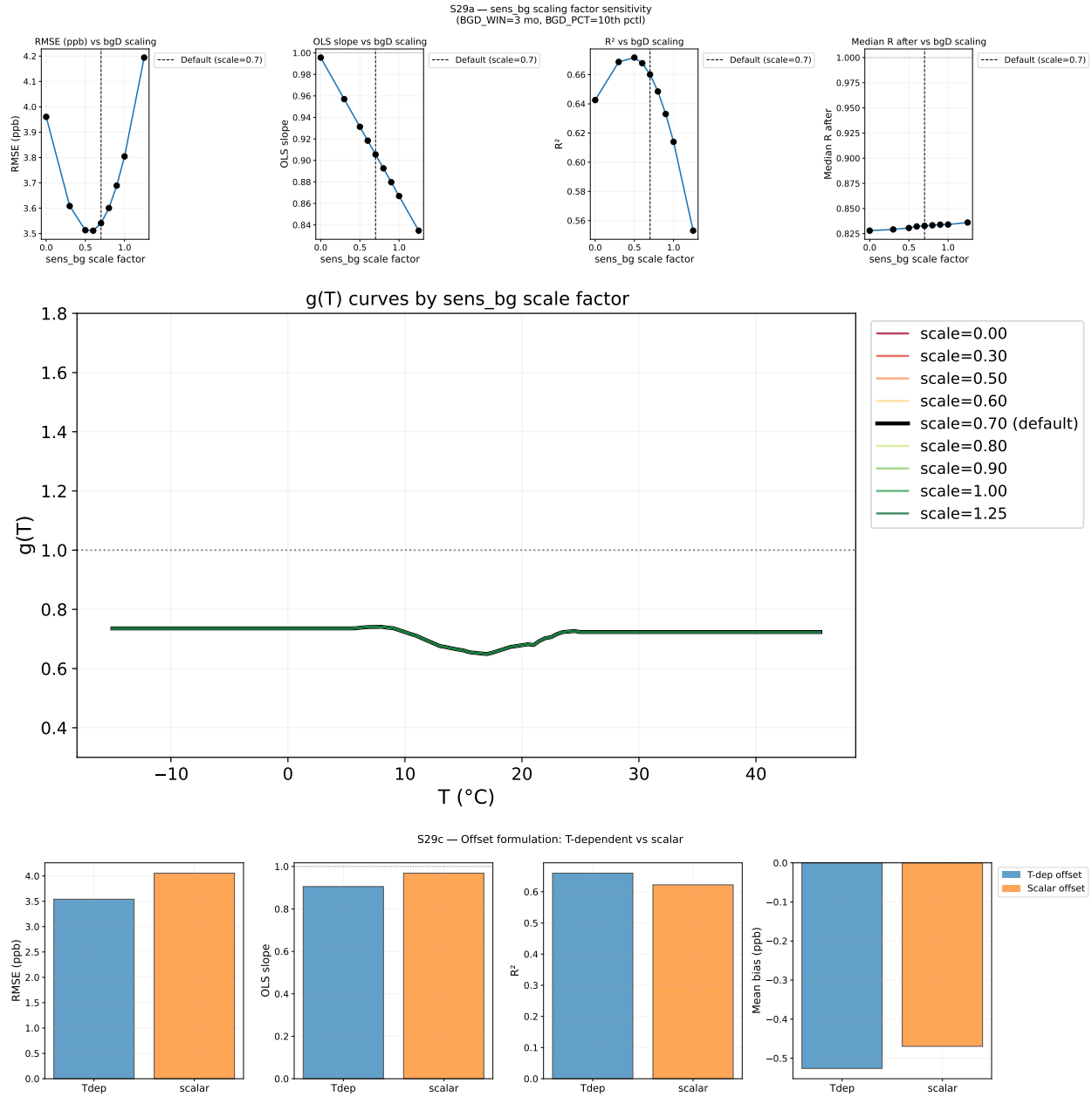


Figure S26: Sensitivity of calibration performance to the Z_{NO_2} scaling factor λ and offset formulation. **(a)** Calibration metrics as a function of the scaling factor $\lambda \in \{0, 0.3, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 1.25\}$. $\lambda = 0$ corresponds to no zero-level subtraction; $\lambda = 0.7$ is the adopted default. Performance degrades at both extremes, confirming that the zero-level estimate is well-calibrated near $\lambda = 0.7$. **(b)** Effect of scaling factor on the gain curve $g(T)$. **(c)** Comparison of the temperature-dependent monthly offset $b(T, m)$ (default) versus a scalar monthly offset $b(m)$ and no offset ($b = 0$). The temperature-dependent form produces lower RMSE and higher slope than the scalar, consistent with its ability to address residual temperature structure.

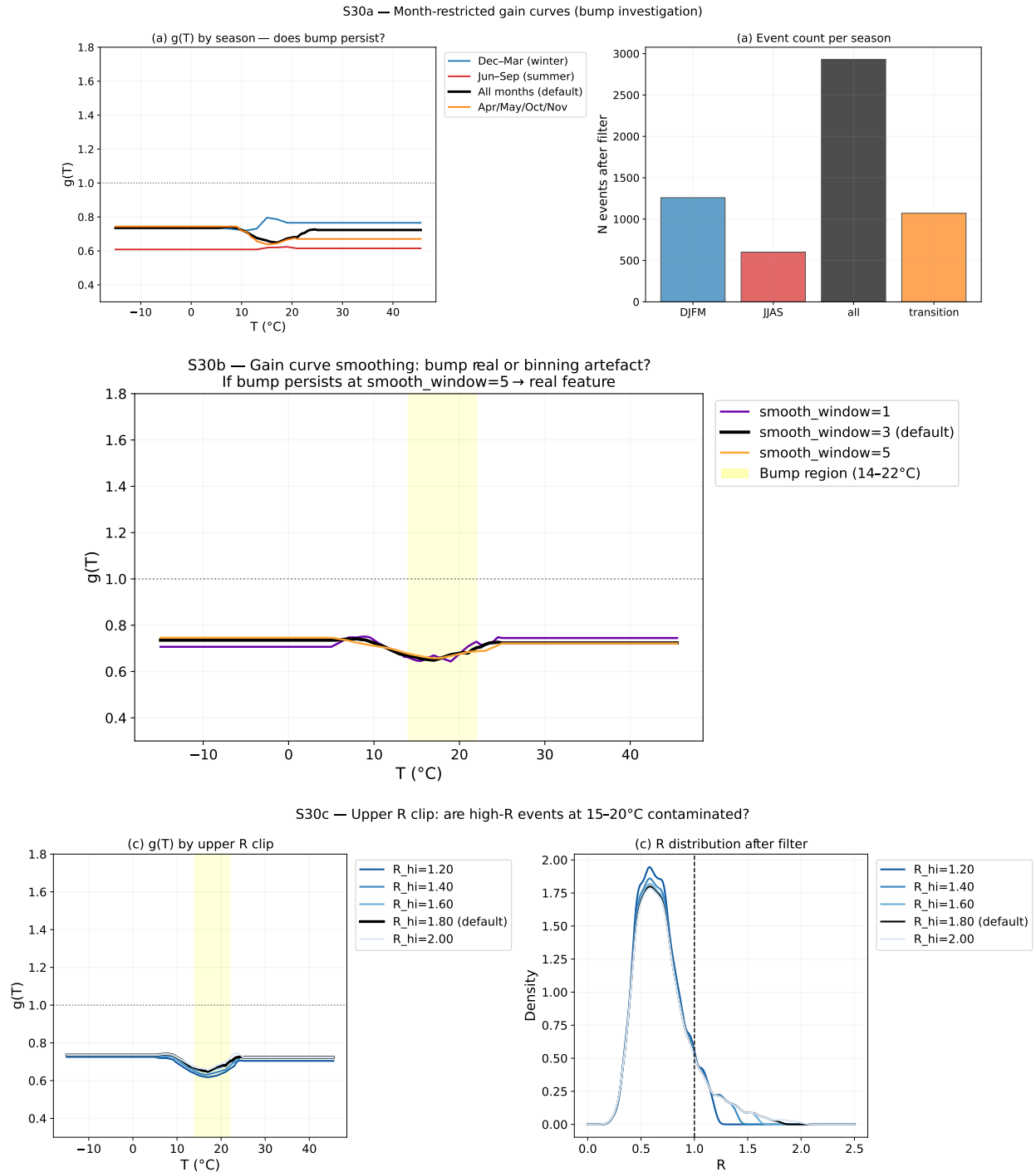


Figure S27: Investigation of the 15–20°C gain curve feature observed at a Bay Area site. Three diagnostic angles: **(a)** Seasonal event pools: gain curves fitted using only events from restricted calendar months to determine whether the bump is seasonal. The bump persists in all seasonal subsets. **(b)** Gain curve smoothing: Savitzky–Golay and rolling-mean smoothers applied to the raw bin medians. Smoothing attenuates but does not eliminate the feature. **(c)** Upper R clip: R_{hi} varied. The bump is reduced when R_{hi} is lowered, suggesting that a small population of high- R events in the 15–20°C range contributes to the feature.

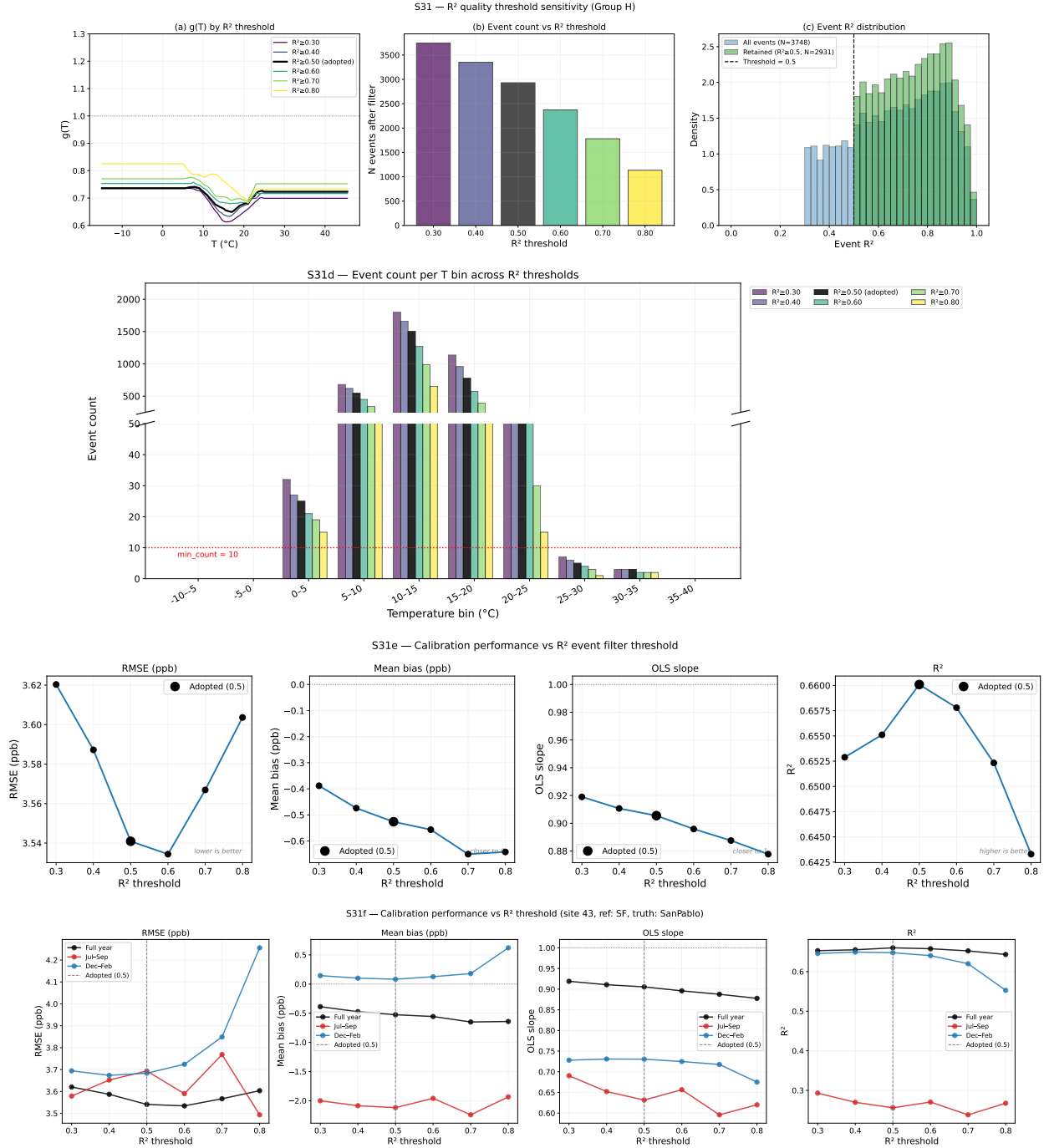


Figure S28: Sensitivity of the gain curve and calibration performance to an OLS R^2 quality threshold applied to individual titration events. **(a)** Median per-node $g(T)$ with IQR shading and post-correction R KDE for R^2 thresholds of 0.0 (no filter; default), 0.3, 0.5, 0.7, and 0.8. The gain curve is stable across thresholds up to 0.5; above 0.7, warm-temperature bins become under-sampled. **(b)** Event count per temperature bin across thresholds, showing that warm-temperature bins are disproportionately depleted by stricter thresholds. **(c)** Calibration RMSE and slope against withheld truth at the Bay Area validation site for each threshold. **(d)** Seasonal breakdown (full year, JAS, DJF) confirming that JAS performance is most sensitive to event depletion.

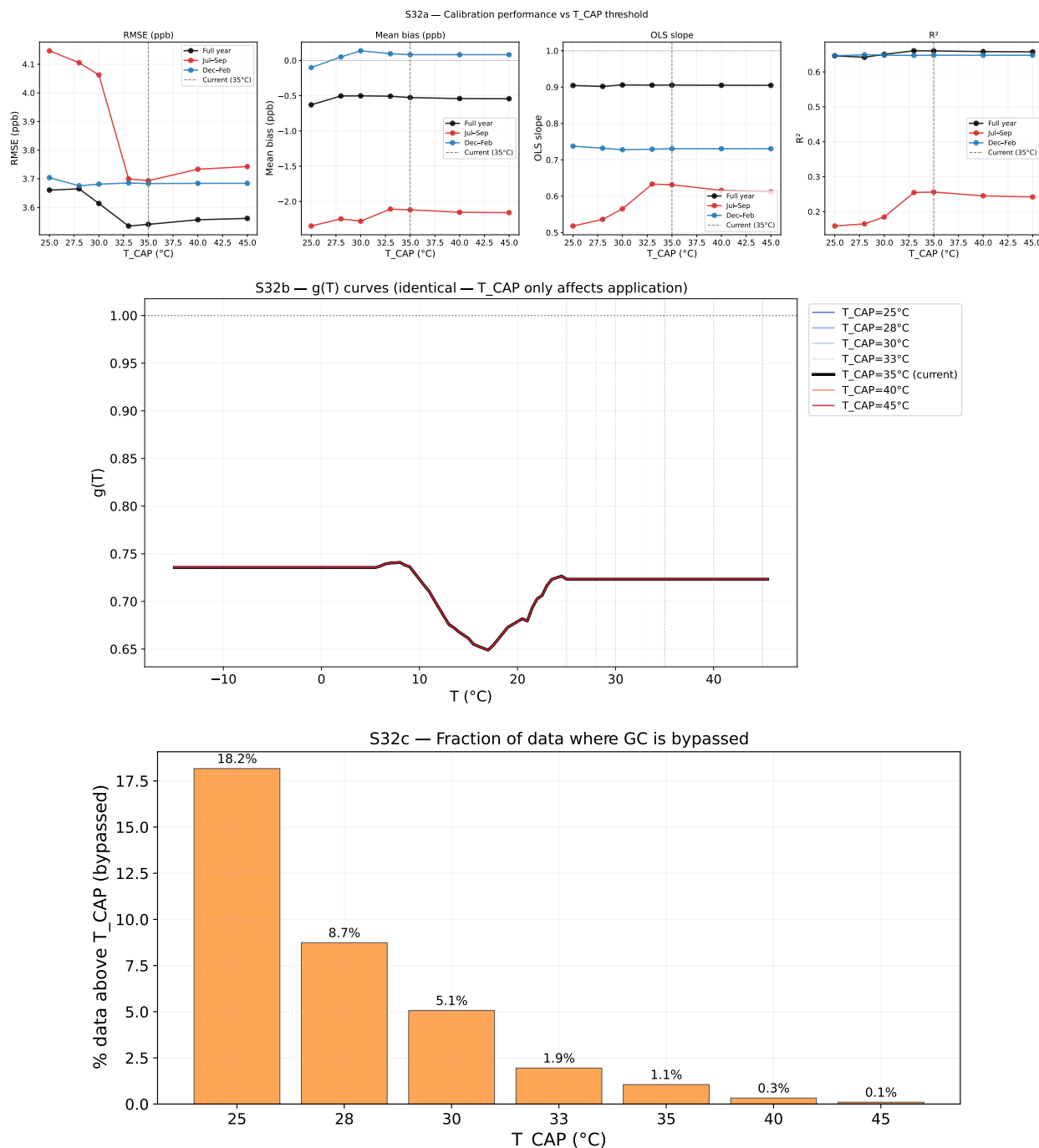


Figure S29: Sensitivity of calibration performance to the temperature cap T_{CAP} above which the gain correction is bypassed. T_{CAP} values of 25, 28, 30, 33, 35 (default), 40, and 45°C are tested at the Bay Area validation site. (a) RMSE and slope against withheld truth for each cap value. Lower caps bypass the correction at progressively more hours; higher caps extend the correction into temperature bins where events are sparse. (b) Effect of each cap value on the gain curve $g(T)$. (c) Fraction of data bypassed at each cap value per city. The adopted $T_{CAP} = 35^\circ\text{C}$ bypasses a small fraction of data while avoiding extrapolation into unsupported temperature ranges.

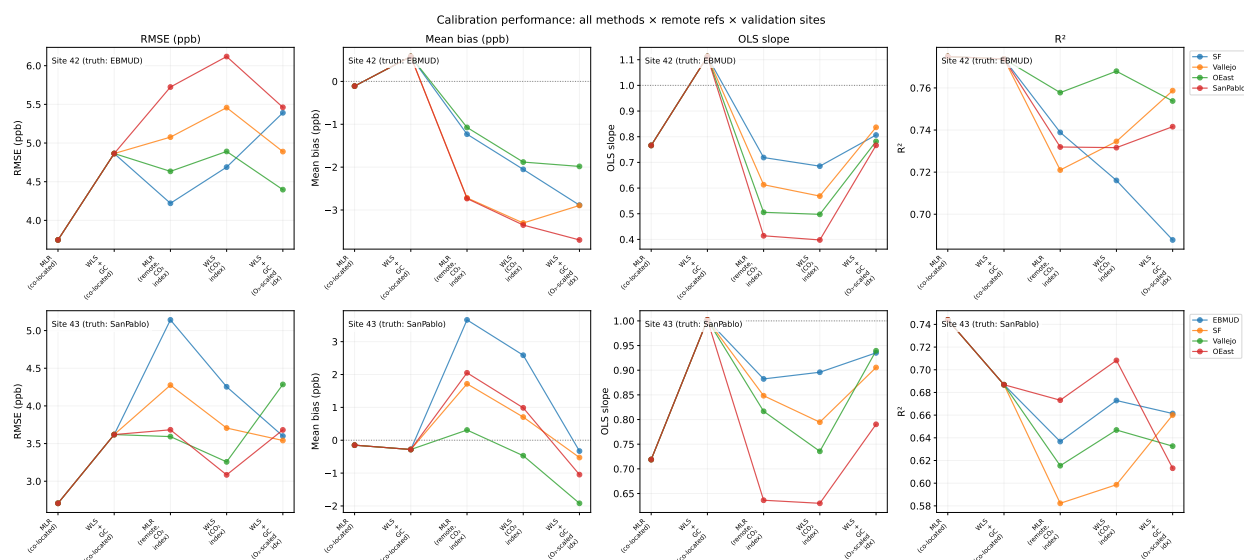


Figure S30: Validation results across all site-reference pairs in the Bay Area, showing the full matrix of calibration performance under the WLS+GC remote pipeline. Each panel corresponds to a different combination of validation site and remote reference station, confirming that the gain correction produces consistent improvement regardless of which remote reference is used.

Appendix C

Supplementary to Chapter 4

Supplementary Information

Contents

Supplementary Tables

- Table S1: Per-profile analysis settings and fitted EFs
- Table S2: Full sub-sectoral GRA²PES comparison
- Table S3: Per-pixel spatial variability summary
- Table S4: Residential fit parameters

Supplementary Figures

- Figures S1–S3: ER vs. distance panels (Bay Area, Providence, Los Angeles)
- Figure S4: Residential heating fuel shares by city
- Figures S5–S7: GRA²PES monthly sectoral composition and fraction of year (CO₂, NO_x, CO)
- Figure S8: Observational seasonal cycle of far-pool ERs
- Figures S9–S14: GRA²PES rush vs. midnight sectoral fractions
- Figures S15–S26: GRA²PES per-pixel EF maps

Table S1: Per-profile analysis settings and fitted transportation emission factors for all 24 city / species / season / time-of-day combinations. Months: calendar months in each seasonal pool. h : starting local hour (AM = morning rush, MN = late night at 00:00). WD: wind-direction window (degrees). $R_{\min}^2$: minimum bivariate regression R^2 for node inclusion. $\text{EF}_{\text{transportation}} = a + c$: fitted transportation emission factor (ppbv/ppmv) with ± 1 SE. N : nodes passing all filters. Profiles where the covariance matrix could not be estimated are marked $\pm \infty$.

City	Species	Season	TOD	Months	h	WD ($^\circ$)	$R_{\min}^2$	$\text{EF}_{\text{trans.}}$	N
<i>SF Bay Area</i>									
	NO _x	Summer	AM	7,8,9	06	140–280	0.5	1.21 ± 0.23	19
	NO _x	Summer	MN	7,8,9	00	140–280	0.5	0.53 ± 0.69	19
	CO	Summer	AM	6,7,8,9	06	100–300	0.5	10.68 ± 2.45	18
	CO	Summer	MN	6,7,8,9	00	100–300	0.5	7.67 ± 1.52	17
	NO _x	Winter	AM	11,12,1,2	07	140–300	0.5	0.68 ± 0.15	20
	NO _x	Winter	MN	11,12,1,2	00	140–300	0.5	0.36 ± 0.12	20
	CO	Winter	AM	12,1,2	07	140–280	0.5	6.41 ± 1.16	18
	CO	Winter	MN	12,1,2	00	140–280	0.5	5.97 ± 0.93	20
<i>Providence</i>									
	NO _x	Summer	AM	6,7,8	07	90–270	0.5	1.22 ± 12.41	11
	NO _x	Summer	MN	6,7,8	00	90–270	0.5	$0.82 \pm \infty$	3
	CO	Summer	AM	6,7,8,9	06	60–180	0.5	6.90 ± 3.82	14
	CO	Summer	MN	6,7,8,9	00	60–180	0.5	13.85 ± 12.50	6
	NO _x	Winter	AM	11,12,1,2	06	100–260	0.5	1.41 ± 1.17	19
	NO _x	Winter	MN	11,12,1,2	00	100–260	0.5	1.68 ± 1.15	19
	CO	Winter	AM	11,12,1,2	05	80–200	0.5	8.07 ± 3.72	21
	CO	Winter	MN	11,12,1,2	00	80–200	0.5	10.92 ± 5.93	20
<i>Los Angeles</i>									
	NO _x	Summer	AM	6,7,8,9	05	0–150	0.5	10.31 ± 44.83	7
	NO _x	Summer	MN	6,7,8,9	00	0–150	0.5	0.32 ± 0.63	7
	CO	Summer	AM	6,7,8,9	05	0–150	0.5	$2.85 \pm \infty$	8
	CO	Summer	MN	6,7,8,9	00	0–150	0.5	4.56 ± 3.22	8
	NO _x	Winter	AM	11,12,1,2	06	0–150	0.5	$4.04 \pm \infty$	6
	NO _x	Winter	MN	11,12,1,2	00	0–150	0.5	0.61 ± 1.01	5
	CO	Winter	AM	11,12,1,2	06	0–150	0.5	4.90 ± 559	7
	CO	Winter	MN	11,12,1,2	00	0–150	0.5	4.53 ± 0.94	7

Los Angeles profiles consistently show very large uncertainties or non-convergent fits, confirming that the network geometry does not support near-highway decay resolution (Section 4.3.1).

Table S2: Sub-sectoral GRA²PES emission factors across season and time-of-day slices. Extends main-text Tables 4.2 and 4.3. All values in ppbv/ppmv.

City	Season	TOD	Species	GRA ² PES sub-sector			
				Gasoline	Diesel	Gasoline+Diesel	Residential
<i>Summer AM</i>							
SF Bay Area	JJA	AM	NO _x	0.64	2.77	1.10	0.73
			CO	8.14	1.83	6.78	3.16
Los Angeles	JJA	AM	NO _x	0.64	2.77	1.08	0.59
			CO	8.14	1.83	6.84	1.68
Providence	JJA	AM	NO _x	0.74	2.77	1.23	0.96
			CO	7.14	1.83	5.86	2.46
<i>Summer MN</i>							
SF Bay Area	JJA	MN	NO _x				
			CO				
Los Angeles	JJA	MN	NO _x				
			CO				
Providence	JJA	MN	NO _x				
			CO				
<i>Winter AM</i>							
SF Bay Area	DJF	AM	NO _x				
			CO				
Los Angeles	DJF	AM	NO _x				
			CO				
Providence	DJF	AM	NO _x				
			CO				
<i>Winter MN</i>							
SF Bay Area	DJF	MN	NO _x	0.64	2.77	1.55	0.77
			CO	8.72	1.83	5.77	8.62
Los Angeles	DJF	MN	NO _x	0.64	2.77	1.51	0.61
			CO	8.72	1.83	5.92	3.71
Providence	DJF	MN	NO _x	0.79	2.77	1.63	1.00
			CO	9.90	1.83	6.47	6.00

Summer MN and Winter AM rows are blank pending extraction with those time-of-day windows.

Table S3: Per-pixel spatial variability of GRA²PES emission ratios across each network footprint. For each combination, the ratio $X_p/\text{CO}_{2,p}$ is computed at each masked grid pixel. $\text{CoV} = \text{SD}/\text{mean}$. On-road rows are abbreviated because all pixels carry the same value within each city. All values in ppbv/ppmv except CoV (dimensionless). Abbreviations: BA = Bay Area, LA = Los Angeles, PVD = Providence.

City	Sector	Species	Season	mean	p_{10}	p_{50}	p_{90}	CoV
<i>On-road: spatially uniform (CoV = 0).</i>								
BA	Gasoline	NO _x	all	0.64	0.64	0.64	0.64	0.00
BA	Diesel	NO _x	all	2.77	2.77	2.77	2.77	0.00
LA	Gasoline	NO _x	all	0.64	0.64	0.64	0.64	0.00
LA	Diesel	NO _x	all	2.77	2.77	2.77	2.77	0.00
PVD	Gasoline	NO _x	Winter	0.79	0.79	0.79	0.79	0.00
PVD	Diesel	NO _x	all	2.77	2.77	2.77	2.77	0.00
BA	Gasoline	CO	Summer	8.14	8.14	8.14	8.14	0.00
BA	Gasoline	CO	Winter	8.73	8.73	8.73	8.73	0.00
BA	Diesel	CO	all	1.83	1.83	1.83	1.83	0.00
PVD	Gasoline	CO	Winter	9.92	9.92	9.92	9.92	0.00
PVD	Diesel	CO	all	1.83	1.83	1.83	1.83	0.00
<i>Residential: sub-grid variability.</i>								
BA	Residential	NO _x	Summer	0.76	0.72	0.75	0.80	0.05
BA	Residential	NO _x	Winter	0.77	0.72	0.76	0.84	0.06
LA	Residential	NO _x	Summer	0.60	0.59	0.60	0.61	0.02
LA	Residential	NO _x	Winter	0.60	0.59	0.60	0.62	0.02
PVD	Residential	NO _x	Summer	0.99	0.94	0.98	1.04	0.04
PVD	Residential	NO _x	Winter	1.01	0.95	1.00	1.09	0.06
BA	Residential	CO	Summer	6.53	2.00	5.43	12.00	0.68
BA	Residential	CO	Winter	8.12	2.32	6.37	17.30	0.71
LA	Residential	CO	Summer	2.65	1.50	2.46	3.89	0.39
LA	Residential	CO	Winter	2.74	1.52	2.50	3.99	0.44
PVD	Residential	CO	Summer	4.50	2.37	4.02	7.79	0.51
PVD	Residential	CO	Winter	5.28	2.53	4.28	8.87	0.52

Residential CO has the highest CoV and widest p_{10} – p_{90} range, especially in the Bay Area (winter CoV = 0.71). Residential NO_x is nearly uniform (CoV $\leq$ 0.06). On-road sub-sectors are spatially constant. See Figures S15–S26 for the corresponding maps.

Table S4: Residential far-field plateau c ($= \text{EF}_{\text{residential}}$) for each city and species, winter midnight pool. ± 1 SE from the nonlinear least-squares fit. N : number of nodes beyond 2 km contributing to each fit.

City	Species	Season / TOD	c (ppbv/ppmv)	N
SF Bay Area	EF_{NO_x}	Winter MN	0.30 ± 0.04	[N]
	EF_{CO}	Winter MN	2.49 ± 0.16	[N]
Los Angeles	EF_{NO_x}	Winter MN	0.27 ± 0.05	[N]
	EF_{CO}	Winter MN	2.93 ± 0.82	[N]
Providence	EF_{NO_x}	Winter MN	0.45 ± 0.06	[N]
	EF_{CO}	Winter MN	3.66 ± 0.39	[N]

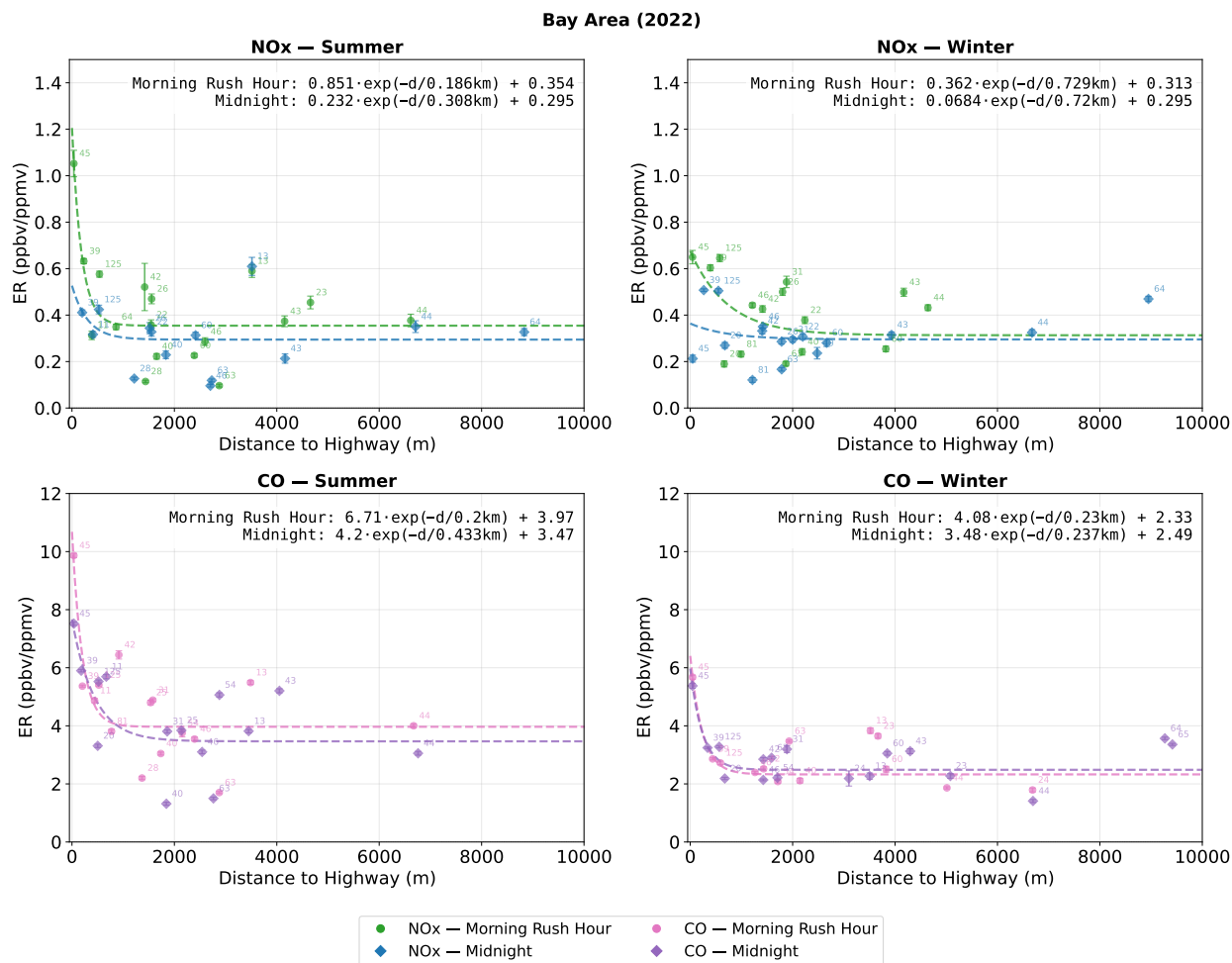


Figure S1: Bay Area enhancement ratios as a function of distance to the nearest highway: NO_x/CO_2 and CO/CO_2 for summer and winter, with morning rush (AM) and midnight (MN) overlaid. The summer AM NO_x and CO profiles are the cases used for the Bay Area transportation EF in main-text Table 4.2. Per-profile settings are listed in Table S1.

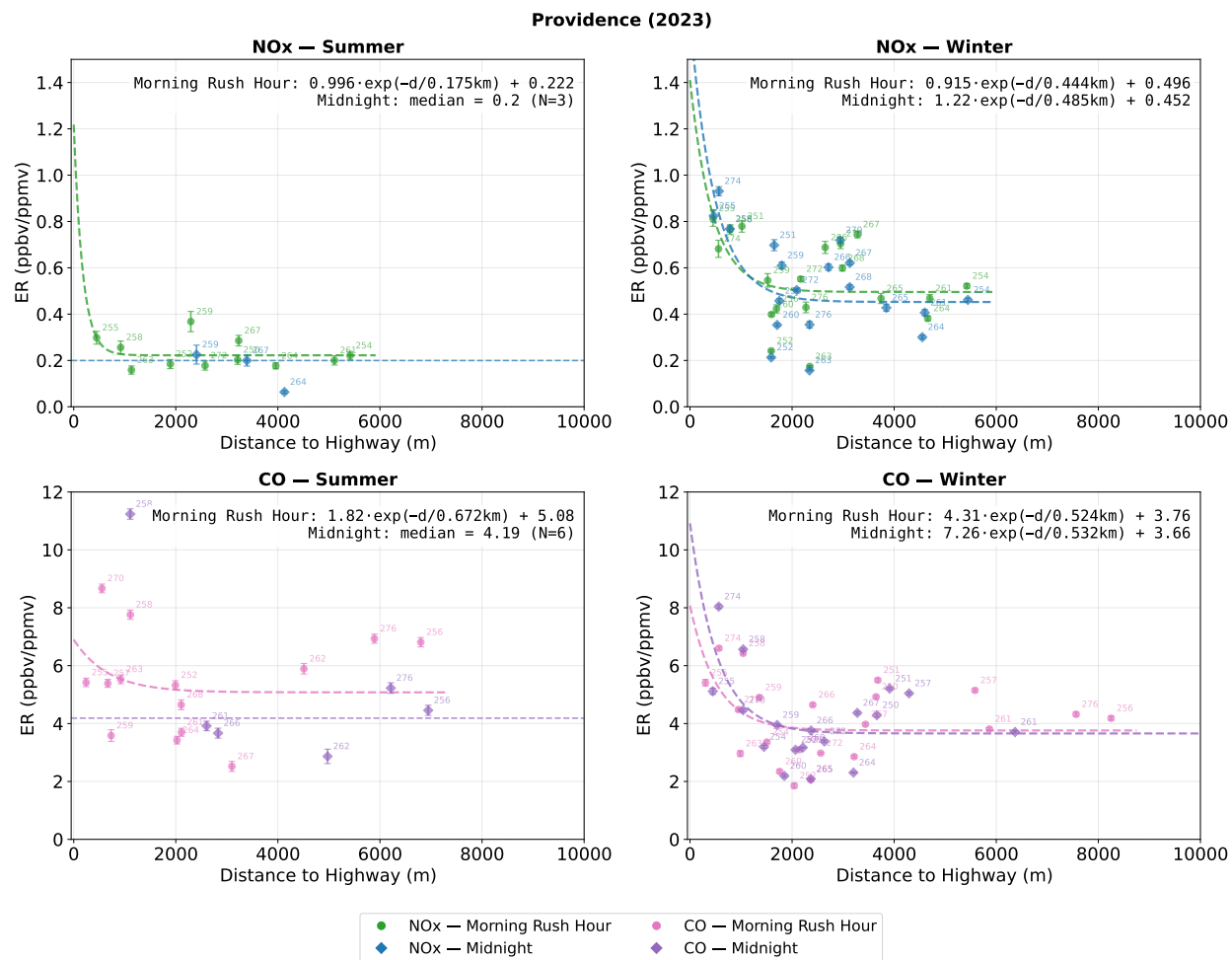


Figure S2: Providence enhancement ratios as a function of distance to the nearest highway, same layout as Figure S1. The winter AM NO_x and CO profiles are the cases used for the Providence transportation EF in main-text Table 4.2. Per-profile settings are listed in Table S1.

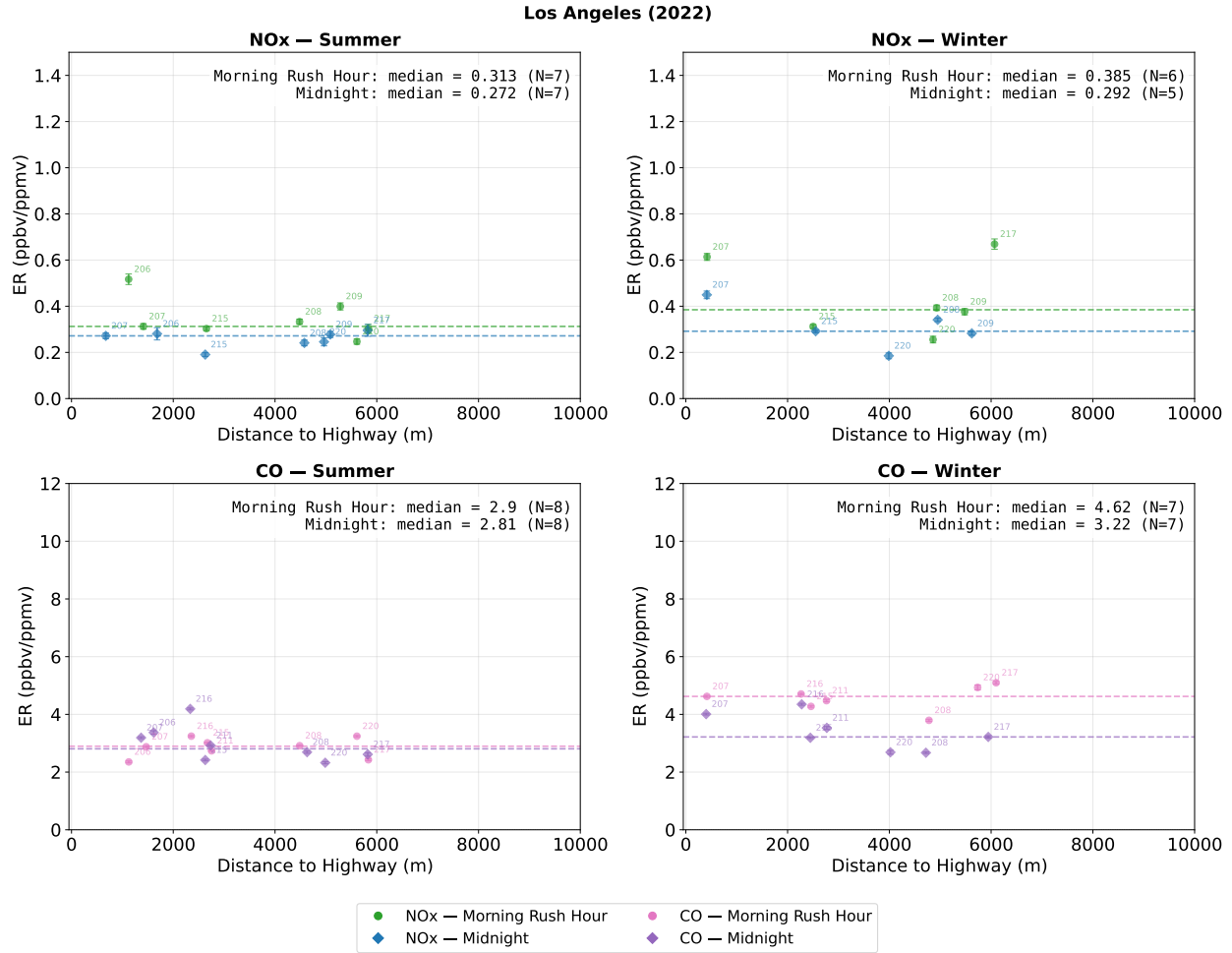


Figure S3: Los Angeles enhancement ratios as a function of distance to the nearest highway, same layout as Figure S1. No exponential decay is resolvable because all nodes lie beyond the near-highway k scale recovered in the Bay Area and Providence. Fitted EFs show very large uncertainties or non-convergent fits (Table S1), confirming that the network geometry does not support transportation EF recovery.

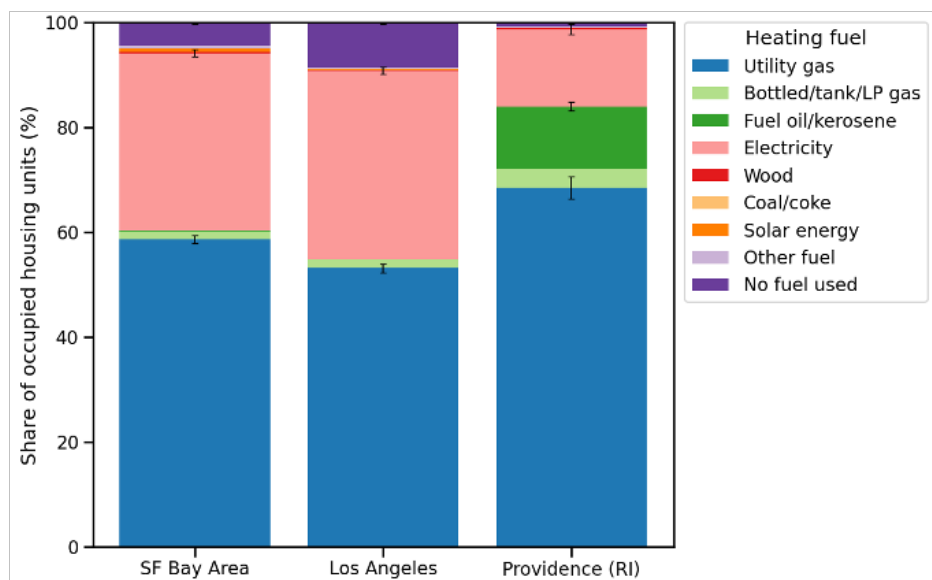


Figure S4: Residential heating fuel shares for the three study regions, from the 2023 American Community Survey 5-year estimates. The Bay Area and Los Angeles are dominated by utility gas ($\sim 55\text{--}60\%$), whereas Providence has a substantial fuel oil and kerosene fraction ($\sim 15\%$) alongside gas ($\sim 68\%$). This contrast in residential fuel mix is the basis for interpreting the cross-city variation in observed residential emission factors (Section 4.3.2). Error bars indicate the survey margin of error.

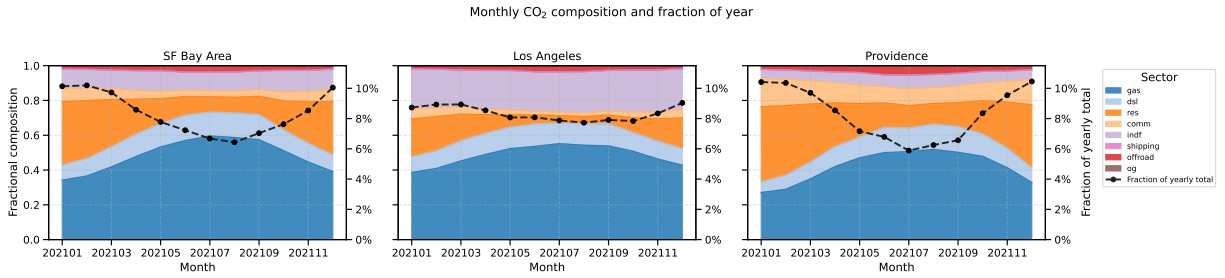


Figure S5: Monthly composition and annual distribution of CO₂ emissions from GRA²PES v1.0, 2021. Stacked areas (left axis) show the fractional contribution of each sector to the monthly total, renormalised so each month sums to unity. The overlaid dashed line (right axis) shows each month's share of the annual total, so the 12 monthly values integrate to 100%.

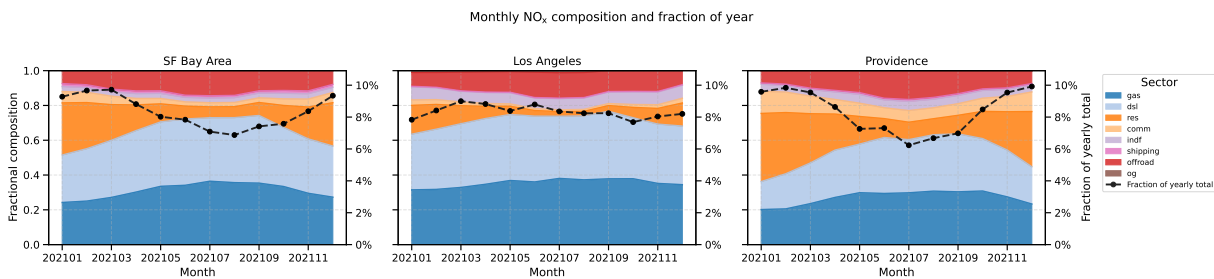


Figure S6: Monthly composition and annual distribution of NO_x emissions from GRA²PES v1.0, 2021, same layout as Figure S5. Residential combustion contributes a larger fraction of winter NO_x in Providence than in the California cities, consistent with the stronger observed residential EF_{NO_x} (Section 4.3.2).

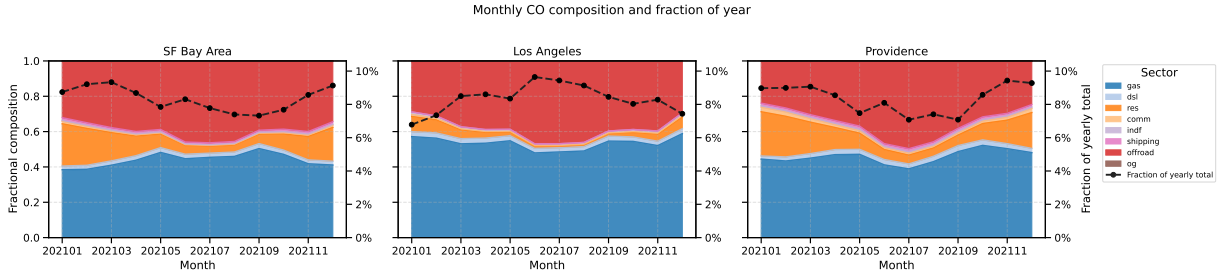


Figure S7: Monthly composition and annual distribution of CO emissions from GRA²PES v1.0, 2021, same layout as Figure S5. Off-road sources (red) account for roughly 30–50% of total CO emissions in all three cities, a substantially larger share than for CO₂ or NO_x. This structural difference in the CO emission budget is discussed in Section 4.3.2.

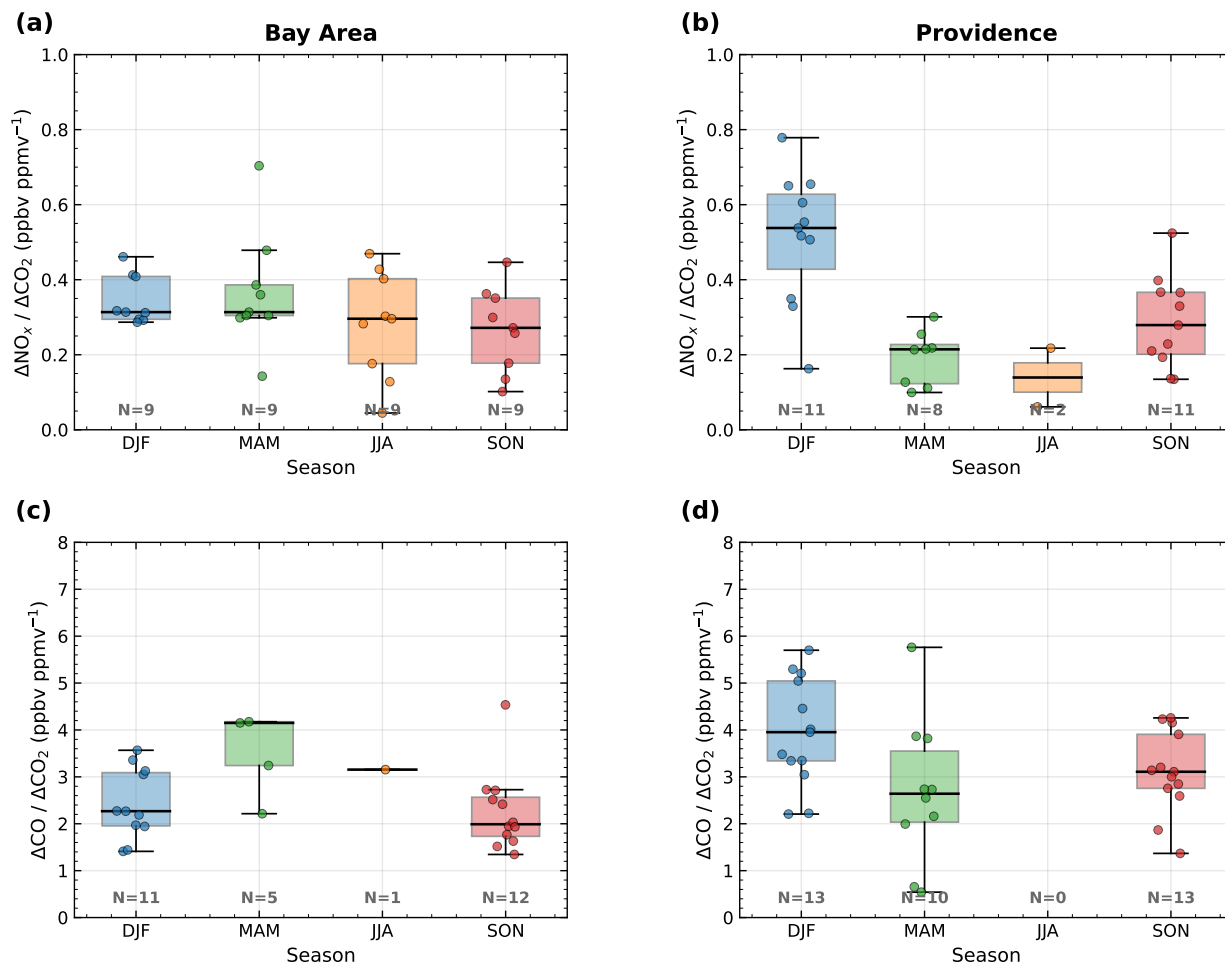


Figure S8: Seasonal cycle of observed far-pool enhancement ratios (distance > 2 km) at midnight hours for (a) Bay Area, and (b) Providence. The elevated winter ERs in Providence are consistent with seasonal heating activity.

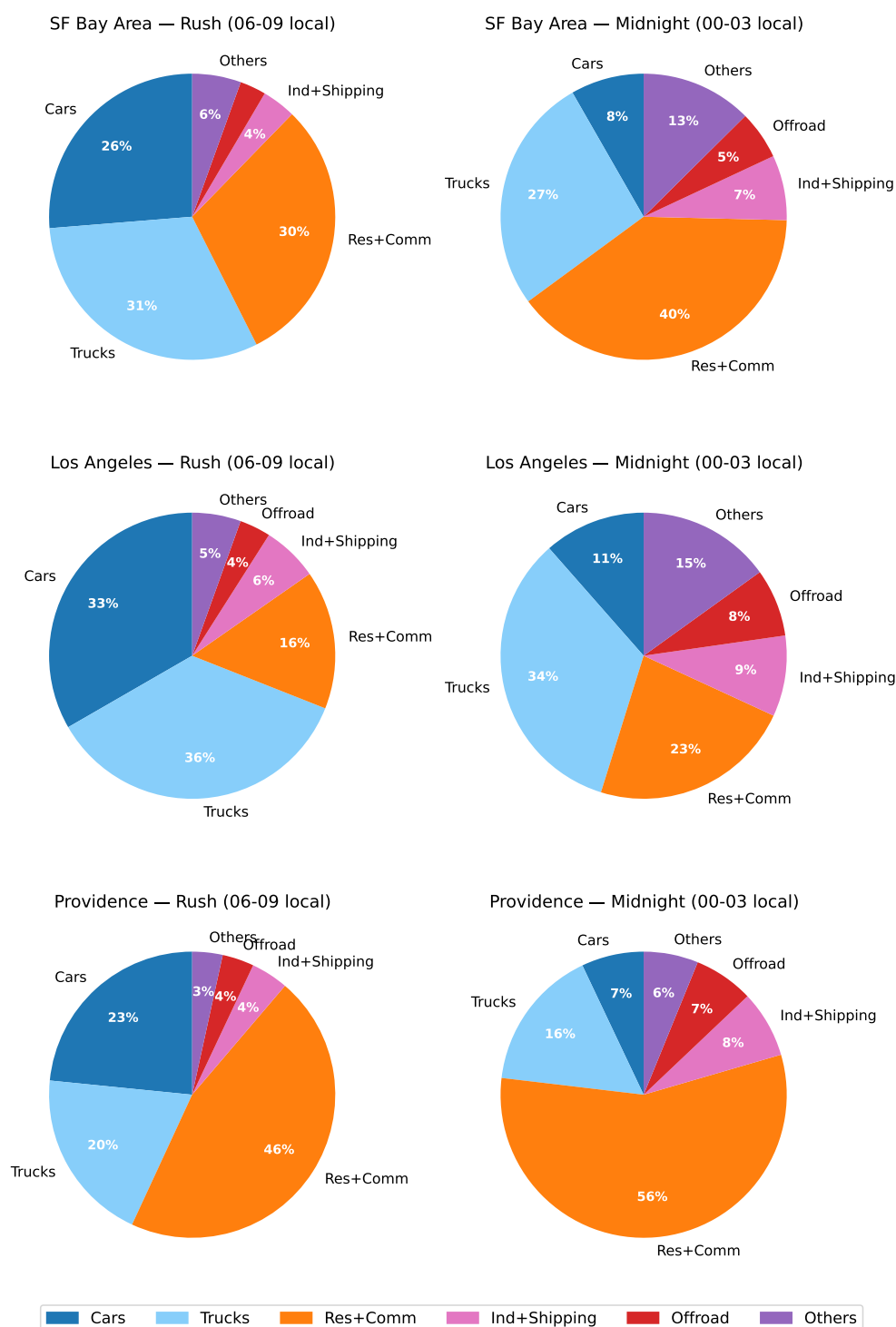
Sectoral NO_x fractions — 202101 (rush vs midnight)

Figure S9: Sectoral NO_x fractions during morning rush (06–09 local) and midnight (00–03 local), January 2021. Transportation dominates at rush hour in all three cities. At midnight, residential and commercial combustion account for larger shares, confirming that the midnight observational window isolates non-transportation sources.

Sectoral CO fractions — 202101 (rush vs midnight)

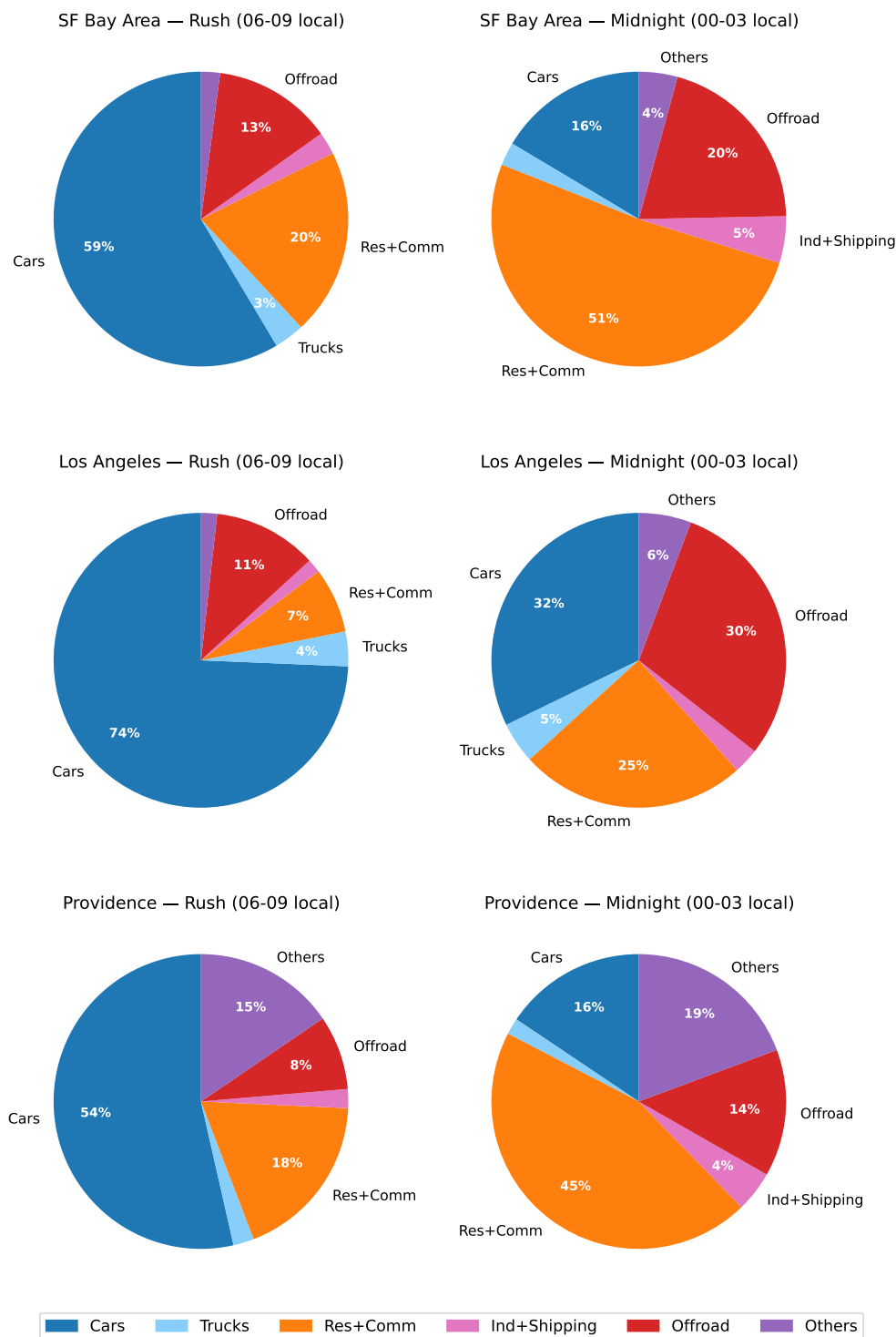


Figure S10: Sectoral CO fractions during rush and midnight, January 2021. The residential fraction grows substantially from rush to midnight in all three cities.

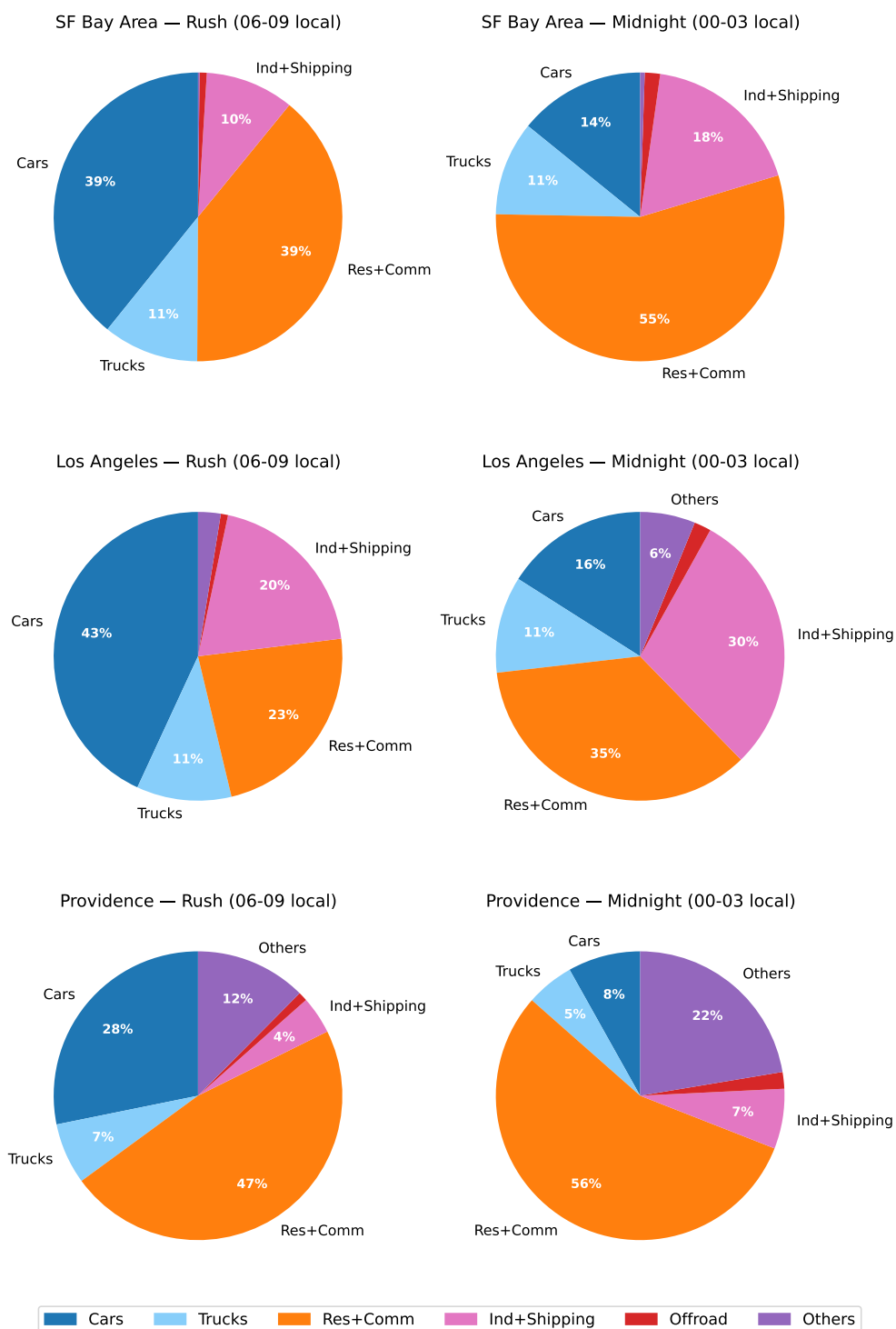
Sectoral CO₂ fractions — 202101 (rush vs midnight)

Figure S11: Sectoral CO₂ fractions during rush and midnight, January 2021. CO₂ shifts from transportation-dominated at rush to residential-dominated at midnight in the Bay Area and Providence.

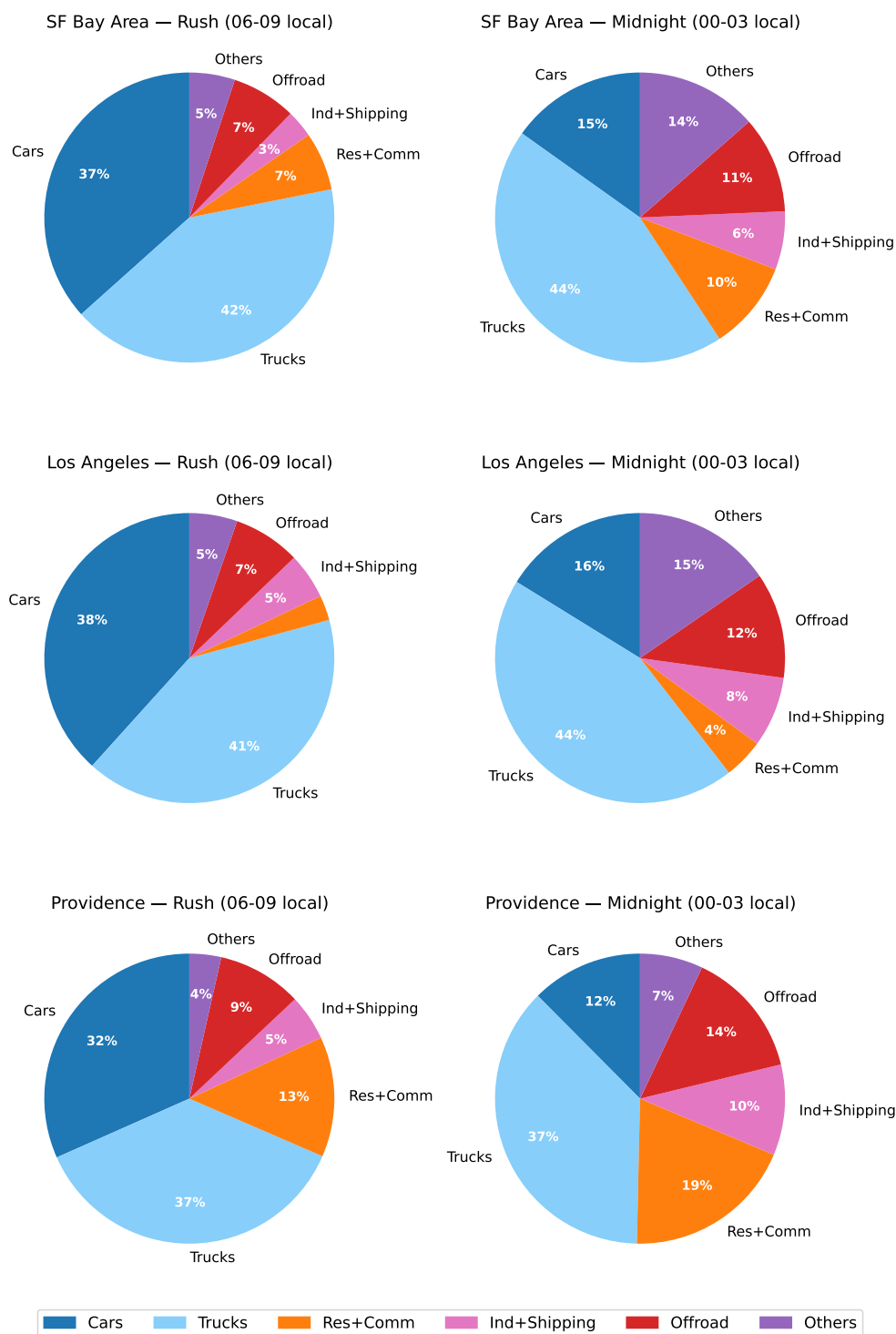
Sectoral NO_x fractions — 202107 (rush vs midnight)

Figure S12: Sectoral NO_x fractions during rush and midnight, July 2021. Summer rush NO_x is more transportation-dominated than in winter.

Sectoral CO fractions — 202107 (rush vs midnight)

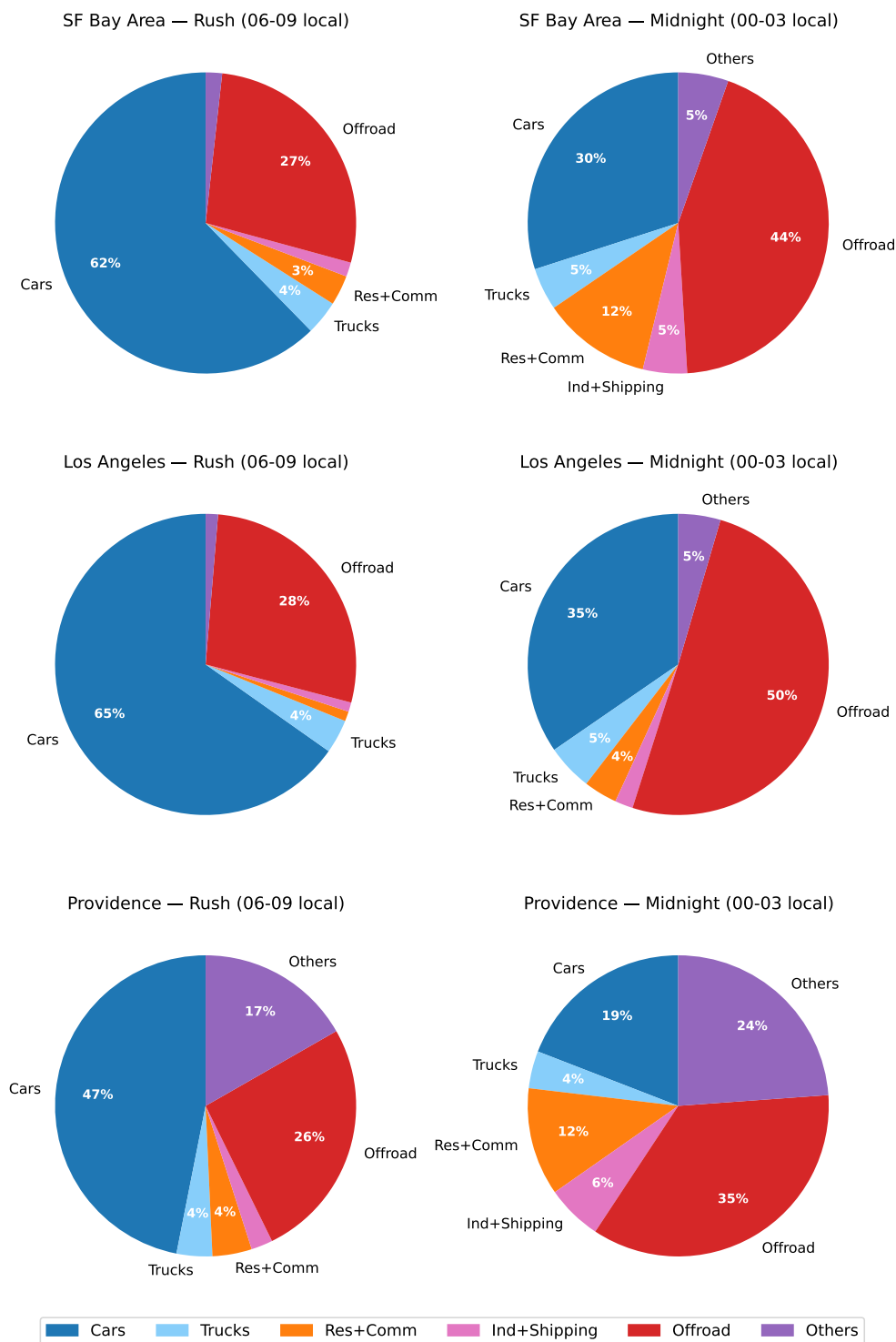
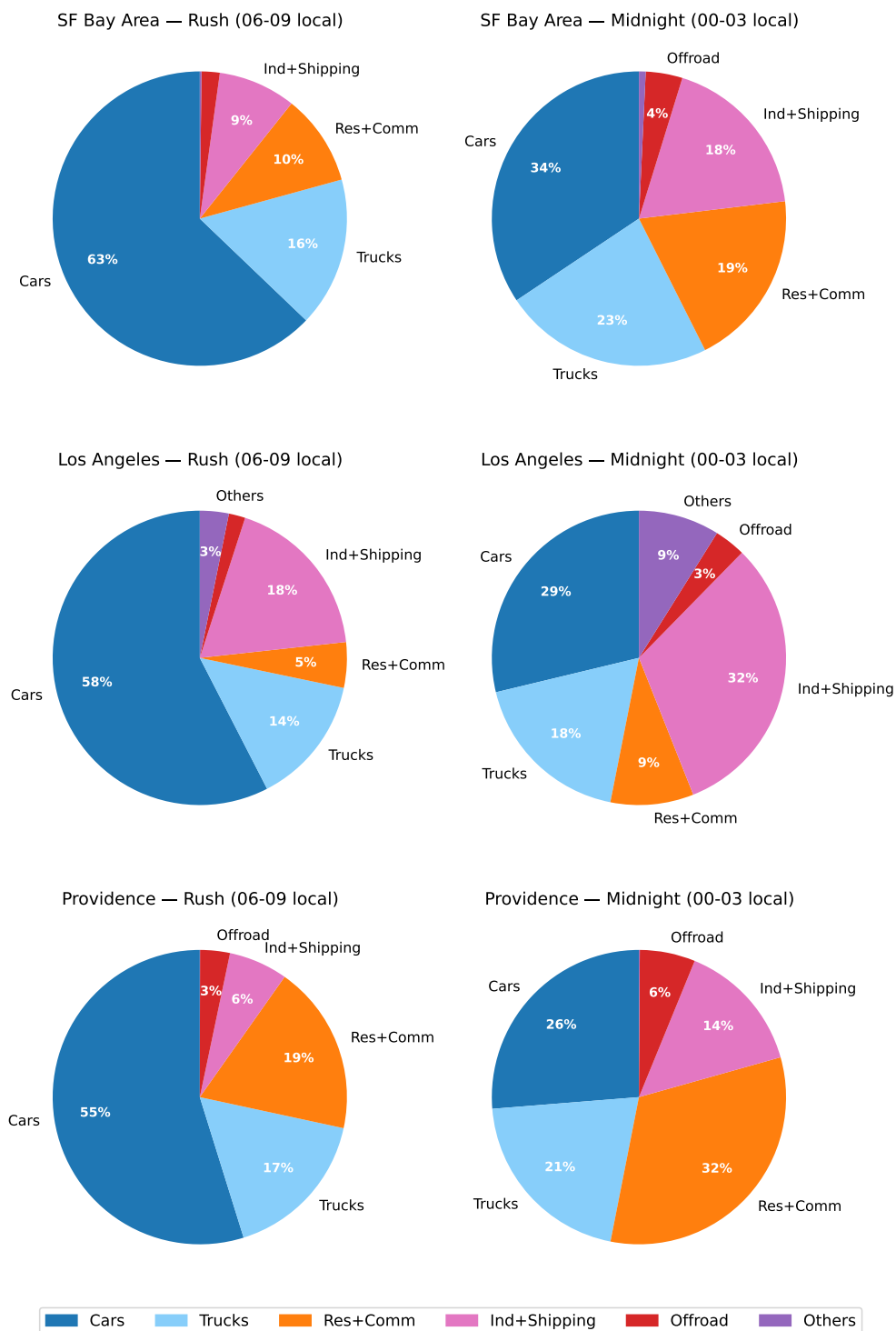


Figure S13: Sectoral CO fractions during rush and midnight, July 2021. Summer rush CO is gasoline-vehicle-dominated. Midnight summer CO shows large off-road fractions in the Bay Area and Los Angeles.

Sectoral CO₂ fractions — 202107 (rush vs midnight)Figure S14: Sectoral CO₂ fractions during rush and midnight, July 2021.

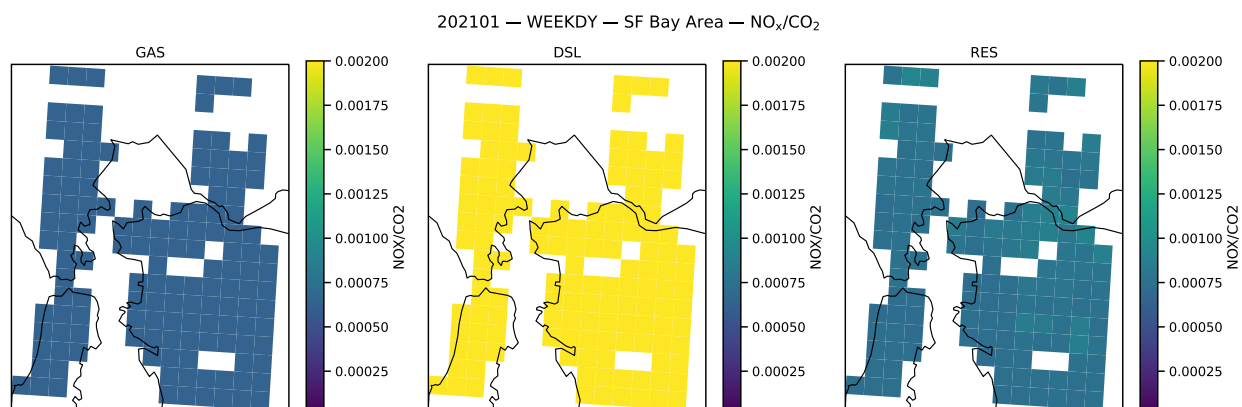


Figure S15: Bay Area per-pixel NO_x/CO_2 , January 2021. Gasoline and diesel panels are single-valued; the residential panel shows mild texture ($\text{CoV} \approx 0.06$).

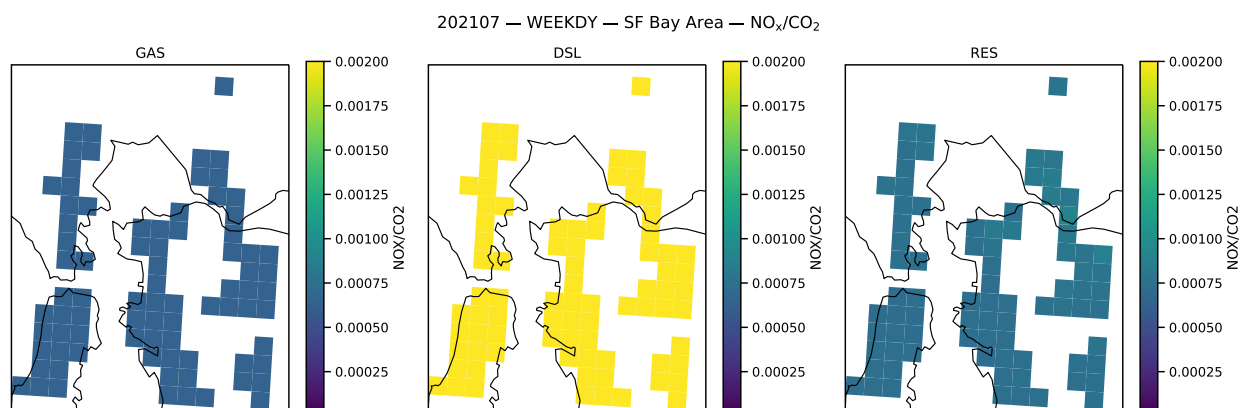
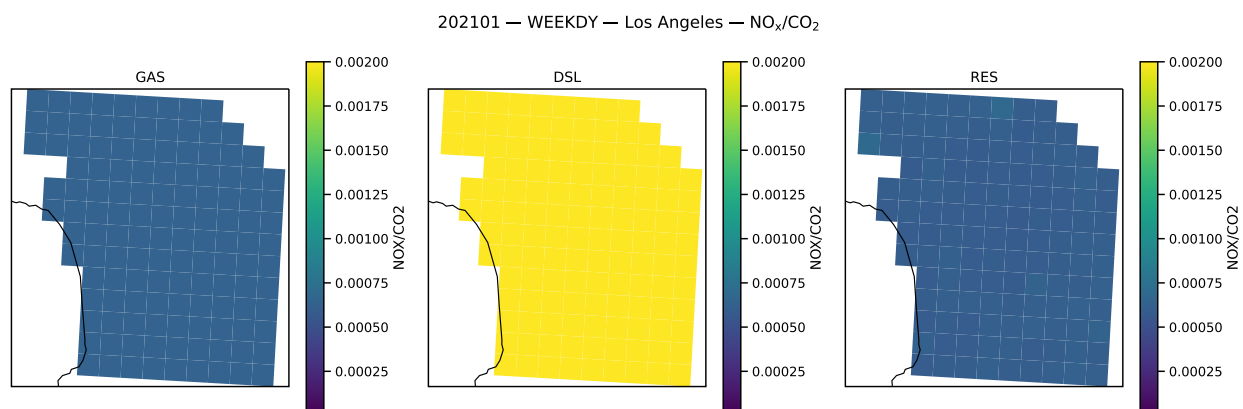
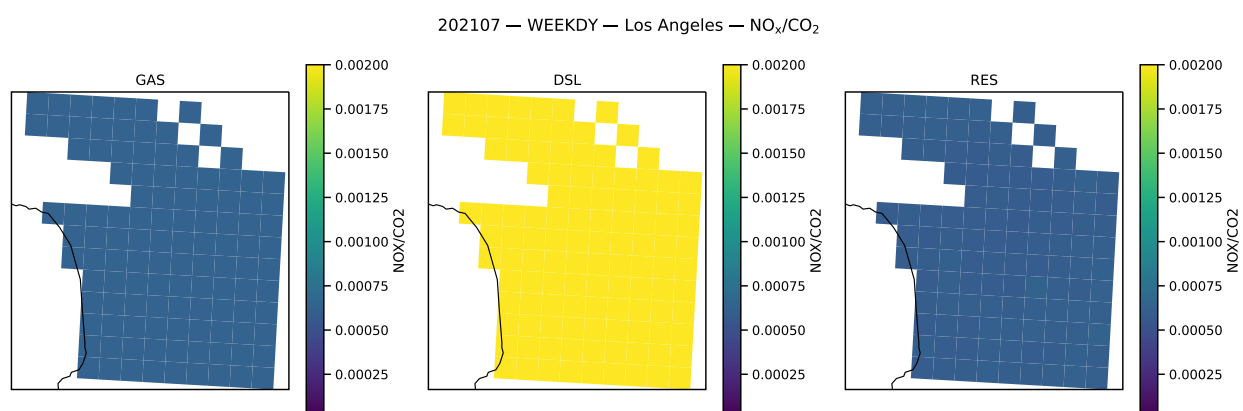
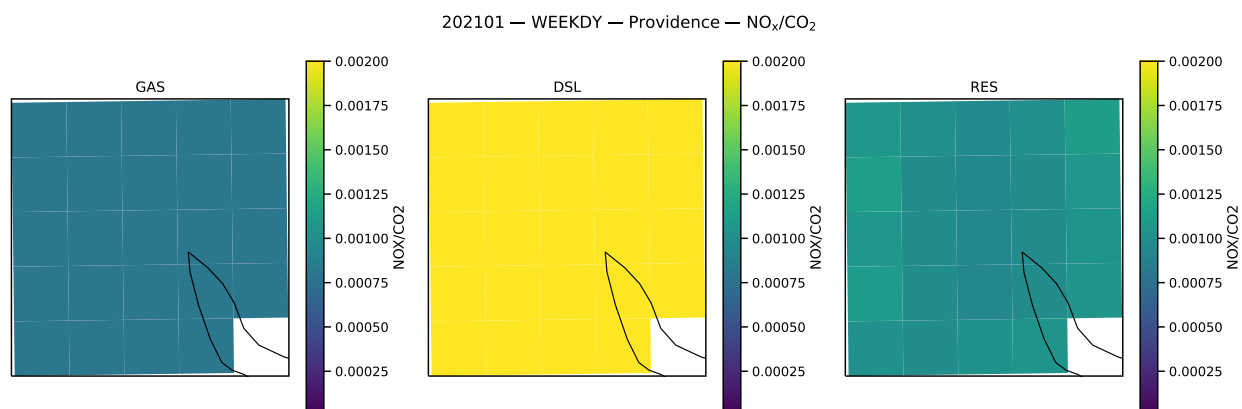
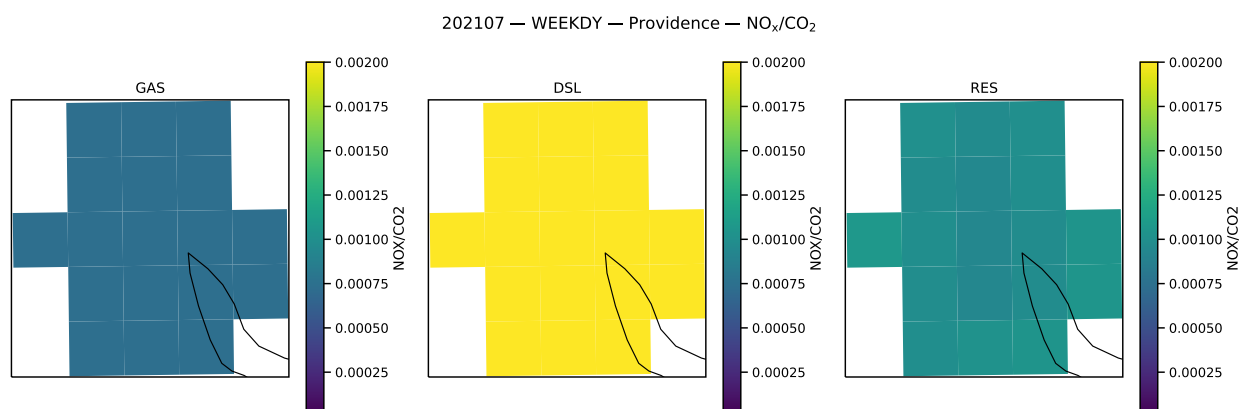


Figure S16: Bay Area per-pixel NO_x/CO_2 , July 2021.

Figure S17: Los Angeles per-pixel NO_x/CO_2 , January 2021.Figure S18: Los Angeles per-pixel NO_x/CO_2 , July 2021.

Figure S19: Providence per-pixel NO_x/CO_2 , January 2021.Figure S20: Providence per-pixel NO_x/CO_2 , July 2021.

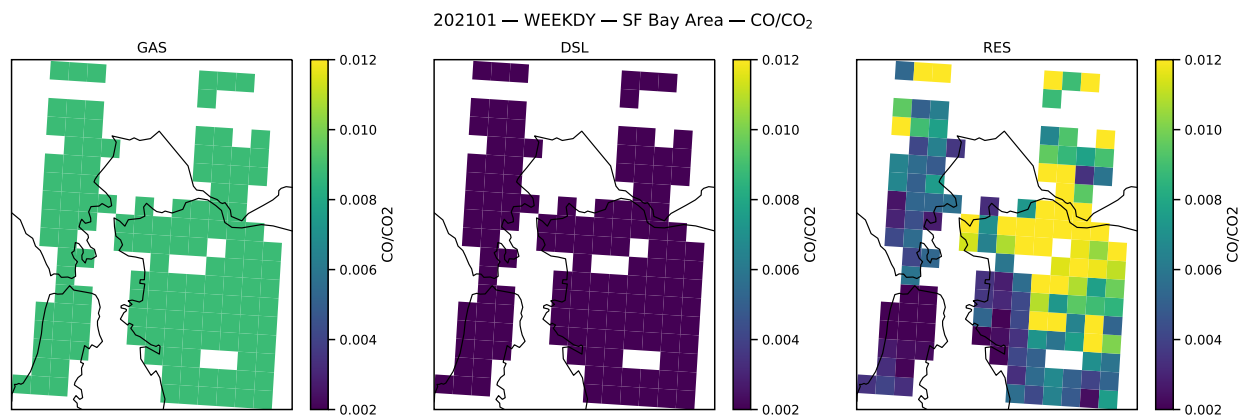


Figure S21: Bay Area per-pixel CO/CO₂, January 2021. The residential panel shows substantial sub-grid texture ($\text{CoV} \approx 0.71$, $p_{10}-p_{90} = 2.3-17.3$ ppbv/ppmv), absent in the gasoline and diesel panels.

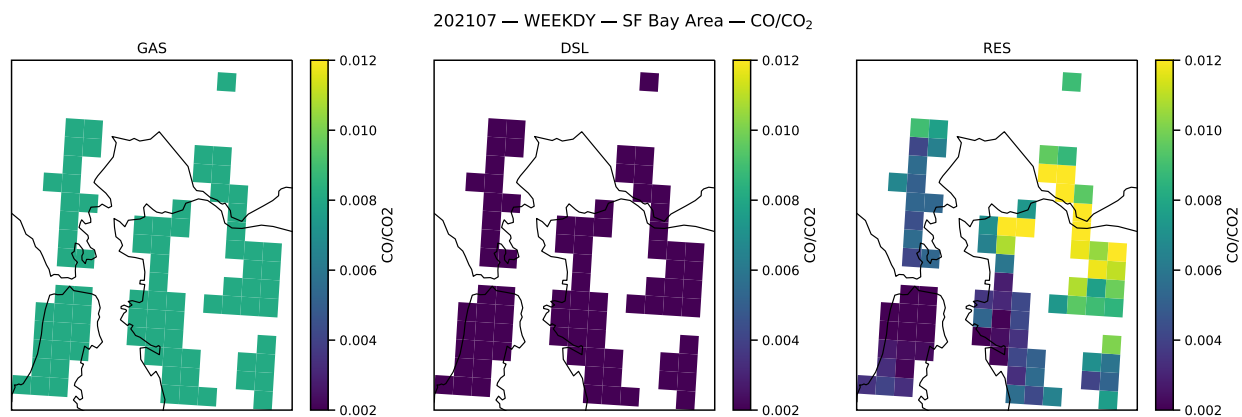


Figure S22: Bay Area per-pixel CO/CO₂, July 2021. Summer residential CO shows similar texture ($\text{CoV} \approx 0.68$), confirming a structural inventory feature rather than a seasonal artefact.

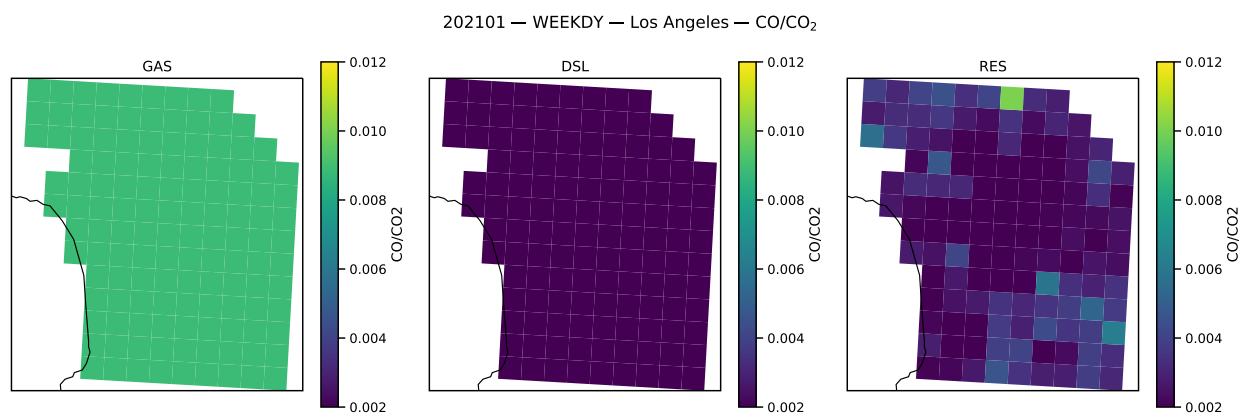


Figure S23: Los Angeles per-pixel CO/CO₂, January 2021. The residential panel is more uniform than the Bay Area (CoV ≈ 0.44).

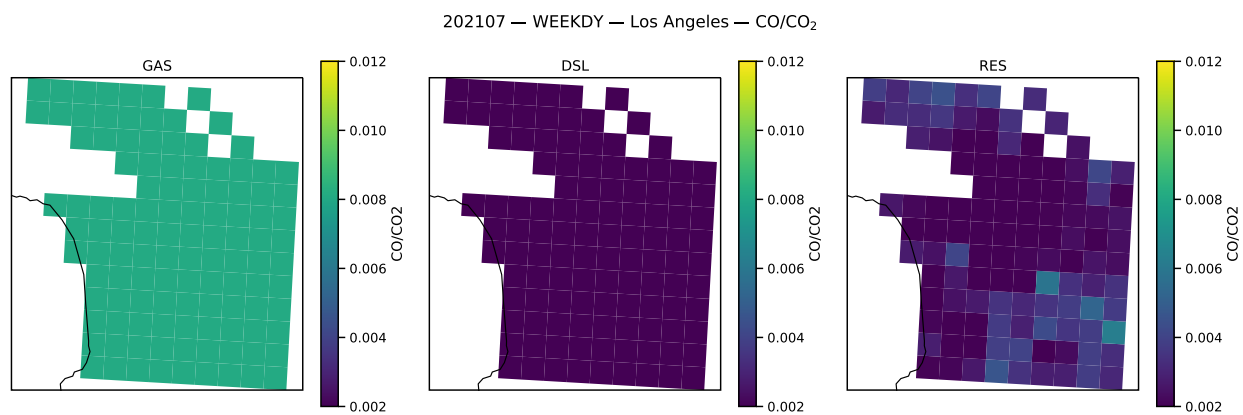
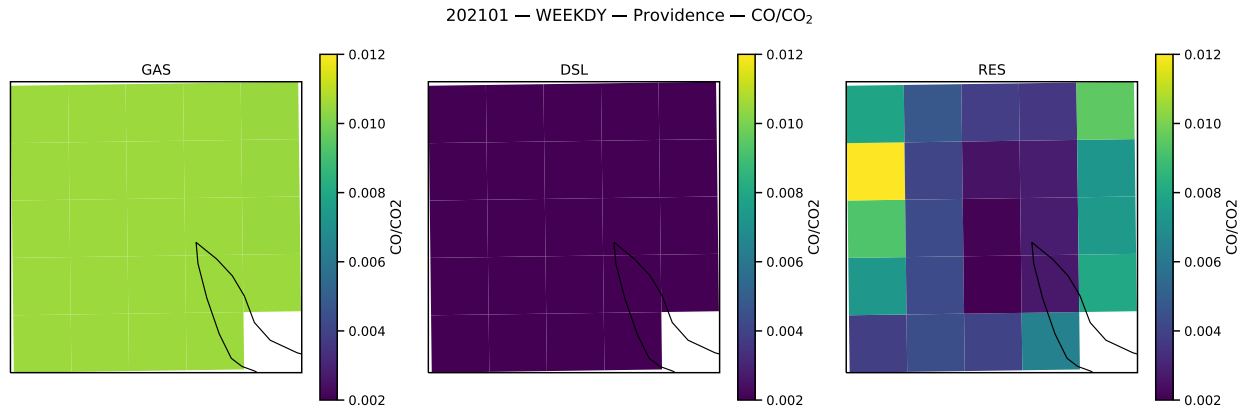
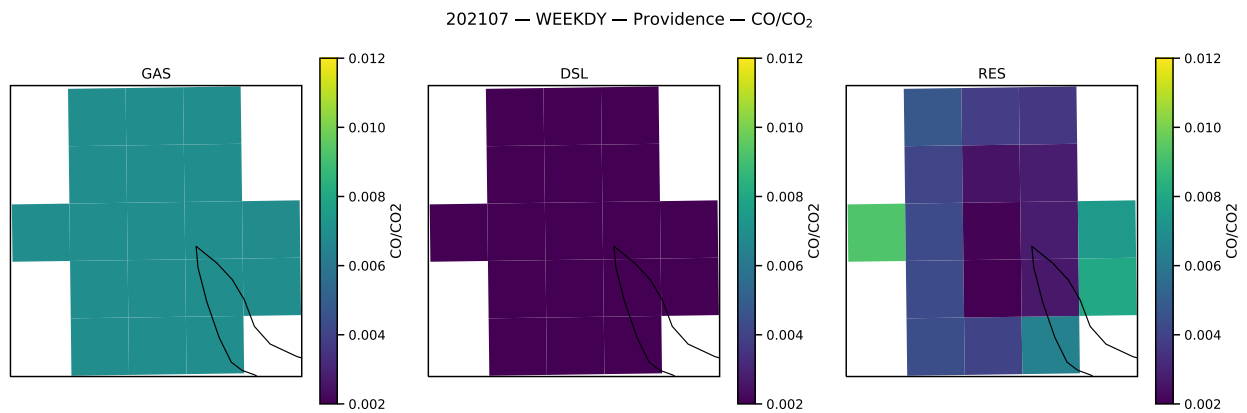


Figure S24: Los Angeles per-pixel CO/CO₂, July 2021.

Figure S25: Providence per-pixel CO/CO₂, January 2021 (CoV ≈ 0.52).Figure S26: Providence per-pixel CO/CO₂, July 2021.