# ADDITIONAL MATERIALS

## Research Report 228

## Optimizing Air Pollution Exposure Assessment with Application to Cognitive Function

**Lianne Sheppard et al.**

## Additional Materials: Chapters 3, 4, 6, 7, 8, 9, and 10

---

The Additional Materials were reviewed by the HEI Improved Exposure Assessment Studies Review Panel. They were not fully edited or formatted by HEI.

Correspondence may be addressed to Dr. Lianne Sheppard, Department of Environmental and Occupational Health Sciences, University of Washington, Box 351618, Seattle, WA 98195; email: *sheppard@uw.edu*.

Although this document was produced with partial funding by the United States Environmental Protection Agency under Assistance Award CR–83998101 to the Health Effects Institute, it has not been subjected to the Agency's peer and administrative review and may not necessarily reflect the views of the Agency; thus, no official endorsement by it should be inferred. It also has not been reviewed by private party institutions, including those that support the Health Effects Institute, and may not reflect the views or policies of these parties; thus, no endorsement by them should be inferred.

# Contents

# Chapter 3

## RATIONALE FOR NOT USING THE $PM_{2.5}$ DATA FROM THE MOBILE CAMPAIGN

For most of the analyses in this report (with the exception of some machine learning model performance comparisons in Chapter 8), we omitted the $PM_{2.5}$ data from the mobile campaign from our results. This is because we determined that the $PM_{2.5}$ predictions had limited variability, had poorer model performances, and were poorly correlated with predictions from the spatio-temporal model predictions from the same time period.

The range of $PM_{2.5}$ predictions from the mobile campaign at cohort locations is small, with a minimum of 3.2 and a maximum of 6.2 $\mu g/m^3$. Further, relative to the other pollutants measured in our campaign, much of the variation in $PM_{2.5}$ predictions in our mobile campaign was captured by the spatial smoothing (kriging) part of the model. We hypothesize that these features contribute to the poorer performance of $PM_{2.5}$ in the mobile campaign overall and with respect to our evaluation of the mobile monitoring study design. We included a nephelometer in our mobile campaign for completeness, recognizing that we don't advocate using mobile monitoring to characterize this pollutant. Mobile monitoring is a much more suitable approach for other traffic-related air pollutants, as we discuss in greater depth in our papers from this campaign (Blanco, Doubleday, et al., 2022; Blanco, Gassett, et al., 2022).

To document these conclusions further, we compared the 2019 annual average predictions at ACT-AP cohort locations for both $PM_{2.5}$ and $NO_2$ from our spatio-temporal models and our mobile campaigns. These predictions are constrained to the 11,612 residential locations that are within the mobile monitoring region shown in red in Figure S3.1.

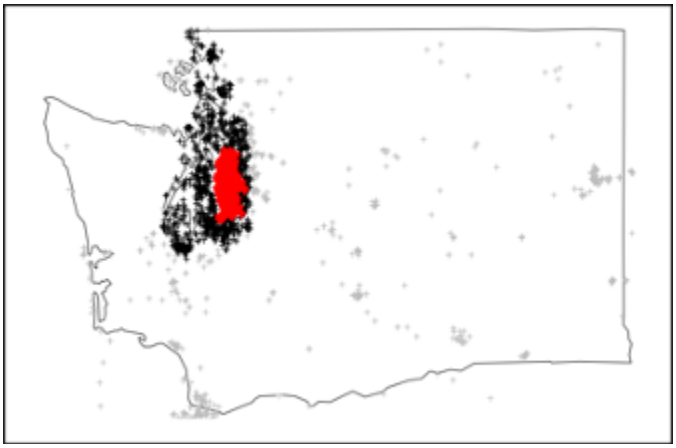


***Figure S3.1****: Prediction availability map. The mobile monitoring region is shown in red and contains 11,612 residential locations relevant to ACT participants' address histories. The spatio-temporal modeling region is shown in black. Locations without predictions are shown in gray.*

Table S3.1 indicates that marginally the two sets of predictions are fairly similar, though the mobile campaign has a lower mean, particularly for $PM_{2.5}$, while the maximum and minimum predictions are higher from the spatio-temporal model for both pollutants.

| Model | Min | 25th Percentile | Median | Mean | 75th Percentile | Max | SD |
|---|---|---|---|---|---|---|---|
| ST $PM_{2.5}$ ($\mu g/m^3$) | 3.5 | 5.1 | 5.4 | 5.4 | 5.8 | 7.6 | 0.5 |
| MM $PM_{2.5}$ ($\mu g/m^3$) | 3.2 | 4.5 | 4.8 | 4.8 | 5.2 | 6.2 | 0.5 |
| ST $NO_2$ (ppb) | 5.8 | 8.7 | 9.5 | 9.8 | 10.5 | 23.3 | 2.0 |
| MM $NO_2$ (ppb) | 3.8 | 8.4 | 9.3 | 9.7 | 10.6 | 20.6 | 2.2 |

***Table S3.1****: Summary statistics of predictions at 11,612 residential locations from ACT participants' address histories*

Observing from Figure S3.2, we note that the $PM_{2.5}$ predictions are poorly correlated between the two models, while the $NO_2$ predictions are much better correlated. These plots support our conclusion to avoid the use of $PM_{2.5}$ predictions from our mobile campaign in much of our work.

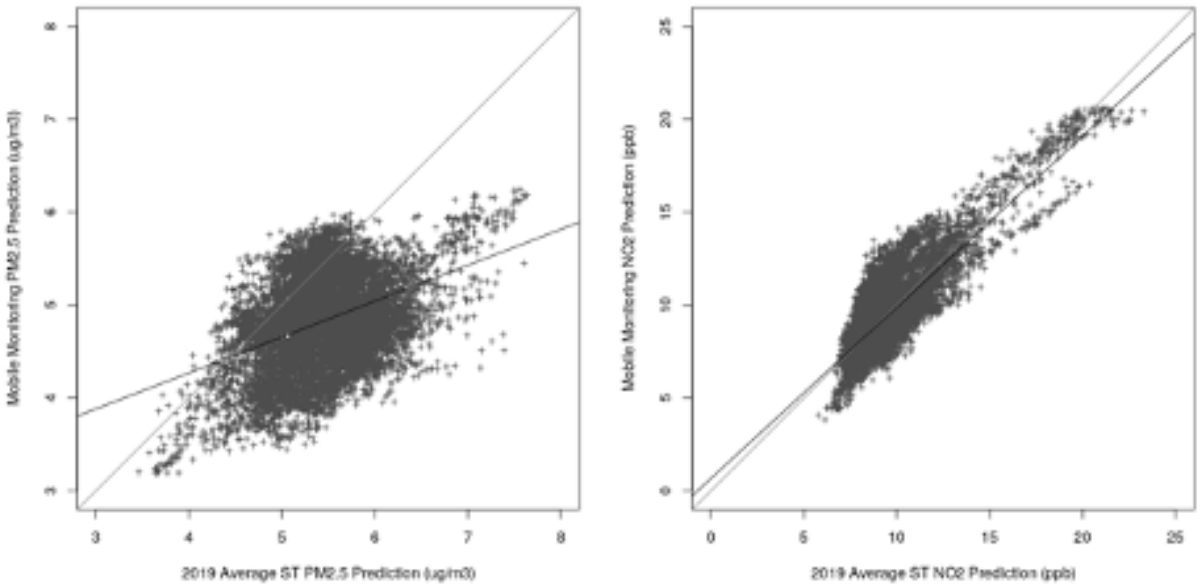



***Figure S3.2****: Scatterplots showing the association between different model predictions at 11,612 residential locations, with $PM_{2.5}$ ($\mu g/m^3$) on the left and $NO_2$ (ppb) on the right. The year 2019 annual average of 2-week spatio-temporal model predictions is presented on the x-axis in both plots. The y-axis presents the 2019 prediction based on a spatial-only model of mobile monitoring observations. The gray line in both plots is the 1-1 line, and the black line is the best-fit line.*

## Geographic Covariates

| Category | Measure | Variable Description | Source |
|---|---|---|---|
| Road | Distance to the nearest object<br><br>Sum within buffers of 0.3-15 km | A1 road, A2+A3 road, bus/truck route, intersection<br>A1 road, A2+A3 road, truck route, intersection | TeleAtlas |
| Population | Sum within buffers of 0.5-15 km | Population in block groups | U.S. Census Bureau |
| Land-Use | Percent within buffers of 0.05-15 km | Urban or built-up land<br><br>Agricultural land<br>Rangeland<br>Forest land<br>Wetland<br>Barren Land<br>Water | Multi-Resolution Land Characteristics (MRLC) National Landcover Dataset & USGS Historical Source |
| Facility & Others | Distance to the nearest object | Coastline<br><br>Commercial area<br>Railroad, railyard<br>Airport, major airport<br>Large port | TeleAtlas |
| Emission | Sum within buffers of 15-30 km | $PM_{2.5}$<br>$PM_{10}$<br>CO<br>$SO_2$<br>$NO_x$ | EPA National Emissions Inventory (NEI) |
| Vegetation | Mean within buffers of 0.25-10 km | Normalized Difference Vegetation Index (NDVI)<br>($25^{th}$, $50^{th}$, and $75^{th}$ percentiles, medians of January through March, April through September, and October through December) | Global Land Cover Facility (GLCF) at University of Maryland |
| Impervious Surfaces | Percent within buffers of 0.05-5 km | Impervious surface value | MRLC National Landcover Dataset |
| Elevation | Level or count within buffers of 1 or 5 km | Elevation above sea level<br><br>Count of points within 20/50 m of the same elevation<br>Count of points more than 20/50 m uphill or downhill | USGS National Elevation Dataset (NED) |

***Table S3.2****: List of selected informative geographic covariates used in the exposure prediction modeling*

**Cohort Inclusion**

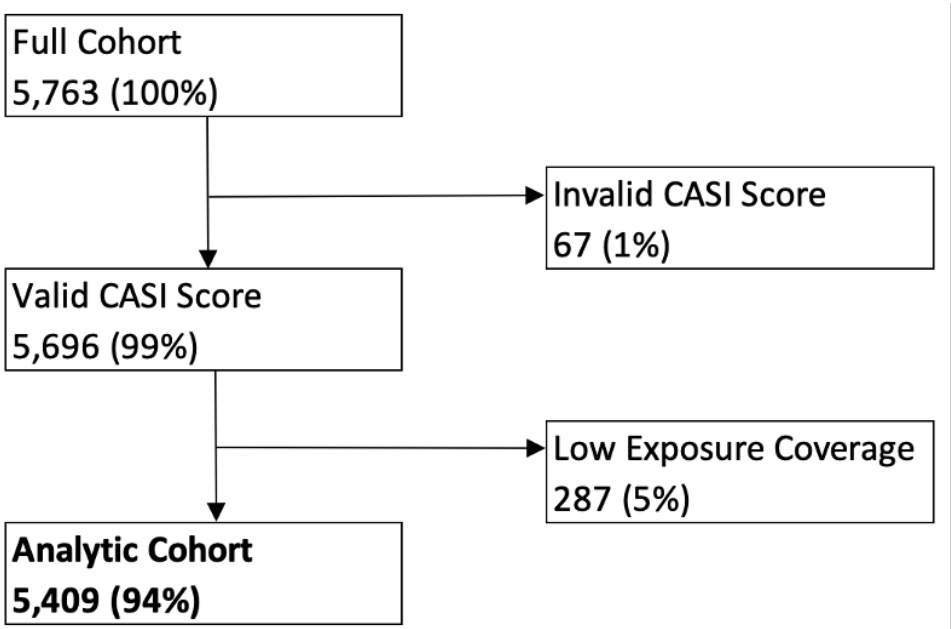



***Figure S3.3****: The Adult Changes in Thought (ACT) analytic cohort used in health inference analyses. Participants with low exposure coverage were those who did not reside in the mobile monitoring region for at least 95% of the previous five years.*

# Chapter 4

## METHODS

*Note S4.1. Temporal adjustment approach*

In summary, our temporal approach consisted of the following:

1. Simulate a long-term UFP monitoring at an urban background site (Beacon Hill; continuous measures were unavailable for the entire mobile monitoring study period) from periodic PNC measures, collocated $NO_2$ measures, and temporal indicators.
2. Generate adjustment factors, defined as the difference between the predicted hourly PNC and the long-term average PNC at Beacon Hill.
3. Apply these adjustment factors to the mobile monitoring data collected under business and rush hours designs.

More specifically, we simulated one based on available continuous $NO_2$ measurements and collocated PNC measurements collected periodically throughout the study period at an urban background site in the study area (Beacon Hill) since a long-term UFP monitoring site was unavailable during the original mobile monitoring study period. Appendix 4 Figure S4.1 depicts similar temporal trends between observed hourly PNC and $NO_2$ concentrations. Table S4.1 and Appendix 4 Figure S4.2 summarize the available PNC measures (available from 31 sampling days across four months and three seasons). Hourly $NO_2$ observations were from the US EPA regulatory air network (US EPA, 2023). We imputed missing hourly $NO_2$ values (3%) based on a regression model fitting existing log-transformed $NO_2$ observations against an indicator for the observation month and a cubic cyclic spline for each day of the week (Mon-Sun) based on the hour associated with that observation (See Equation S4.1 for a similar layout). While most days with missing observations had only one missing value, and we could have implemented simpler linear regression, we took this more flexible approach because 5 days had multiple (5+) missing values. We fit the following model to simulate a long-term PNC monitoring site:

$$\log(PNC_t) = \alpha + \beta * \log(NO_{2,t}) + I * DOW_t + s(hour_t) * DOW_t + \epsilon_t \quad \text{(S4.1)}$$

where the log-transformed hourly average PNC at time *t* was regressed against the hourly average log-transformed $NO_2$ concentration at the same time *t* and a cubic cyclic spline for each day of the week (DOW: Mon-Sun) for the hour (0-23) associated with *t* (i.e., there are seven cyclic splines). The resulting model residuals had a minimal temporal variation, suggesting that the model captured important PNC temporal trends (Figure S4.4); and it had an in-sample model $R^2$ of 0.52.

We used this model to predict hourly PNC concentrations during the study period and simulate a long-term PNC monitoring site. We winsorized (i.e., set) extreme predictions above the 95$^{th}$ (16,380 pt/cm$^3$) and below the 5$^{th}$ quantile (2,963 pt/cm$^3$) to those quantiles, respectively, to reduce extreme temporal adjustments in the next step. We calculated hourly adjustment factors by taking the difference between the long-term (i.e., annual) and each hourly average PNC site concentration:

$$\hat{\delta}_t = PNC_{LTA} - \widehat{PNC}_{t,winsorized} \quad \text{(S4.2)}$$

Finally, PNC samples collected under the business or rush hours designs were adjusted:

$$\widehat{PNC}_{t,adj} = \widehat{PNC}_t + \hat{\delta}_t \quad \text{(S4.3)}$$

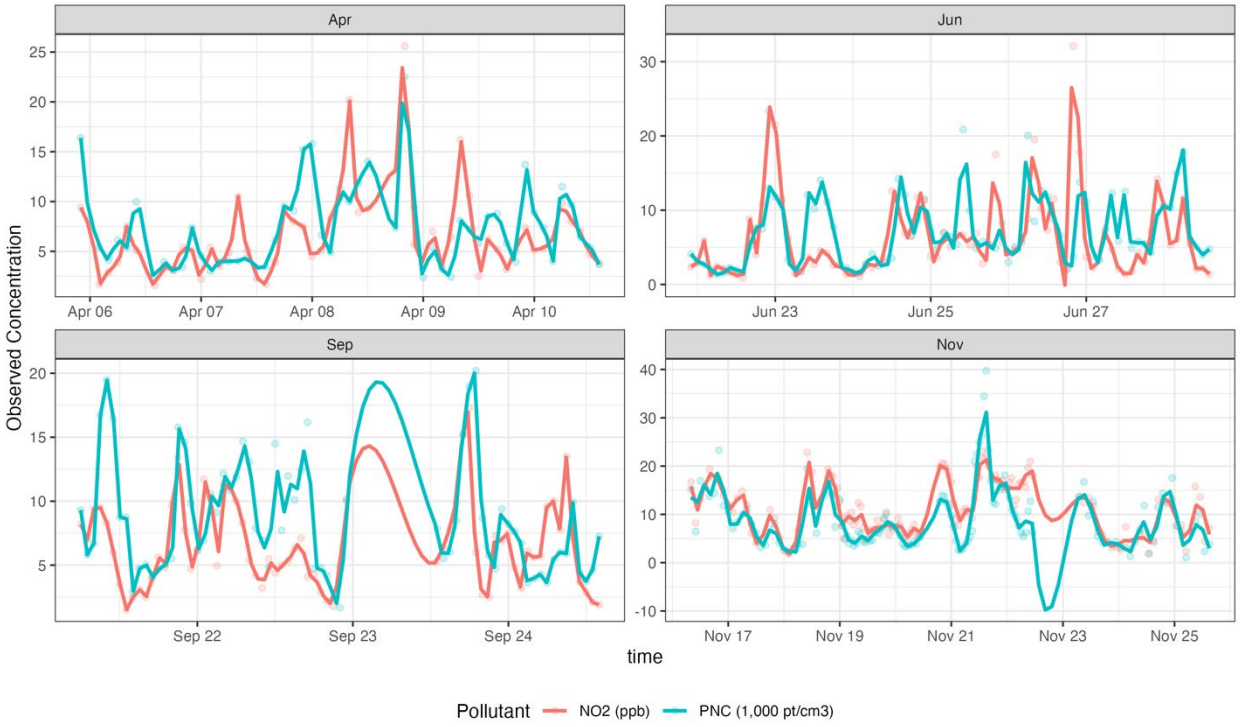



*Figure S4.1. Time series of hourly $NO_2$ (ppb) and UFP (1,000 pt/cm3) at Beacon Hill used to simulate a continuous, long-term PNC monitoring site, as described in Chapter 4's Methods. There are 632 paired hourly observations with a Pearson correlation (R) of 0.64. Smooth lines describe the general pollutant trends.*

*Table S4.1. Continuous overnight sampling PNC times at the Beacon Hill monitoring site. Data were used to simulate a continuous, long-term PNC monitoring site, as described in Chapter 4's Methods.*

| Month | Dates | Start | End | Days |
|---|---|---|---|---|
| Apr | 6 | 2019-04-05 | 2019-04-10 | Fri, Sat, Sun, Mon, Tue, Wed |
| Jun | 8 | 2019-06-21 | 2019-06-28 | Fri, Sat, Sun, Mon, Tue, Wed, Thu |
| Sep | 6 | 2019-09-19 | 2019-09-24 | Thu, Fri, Sat, Sun, Mon, Tue |
| Nov | 11 | 2019-11-15 | 2019-11-25 | Fri, Sat, Sun, Mon, Tue, Wed, Thu |

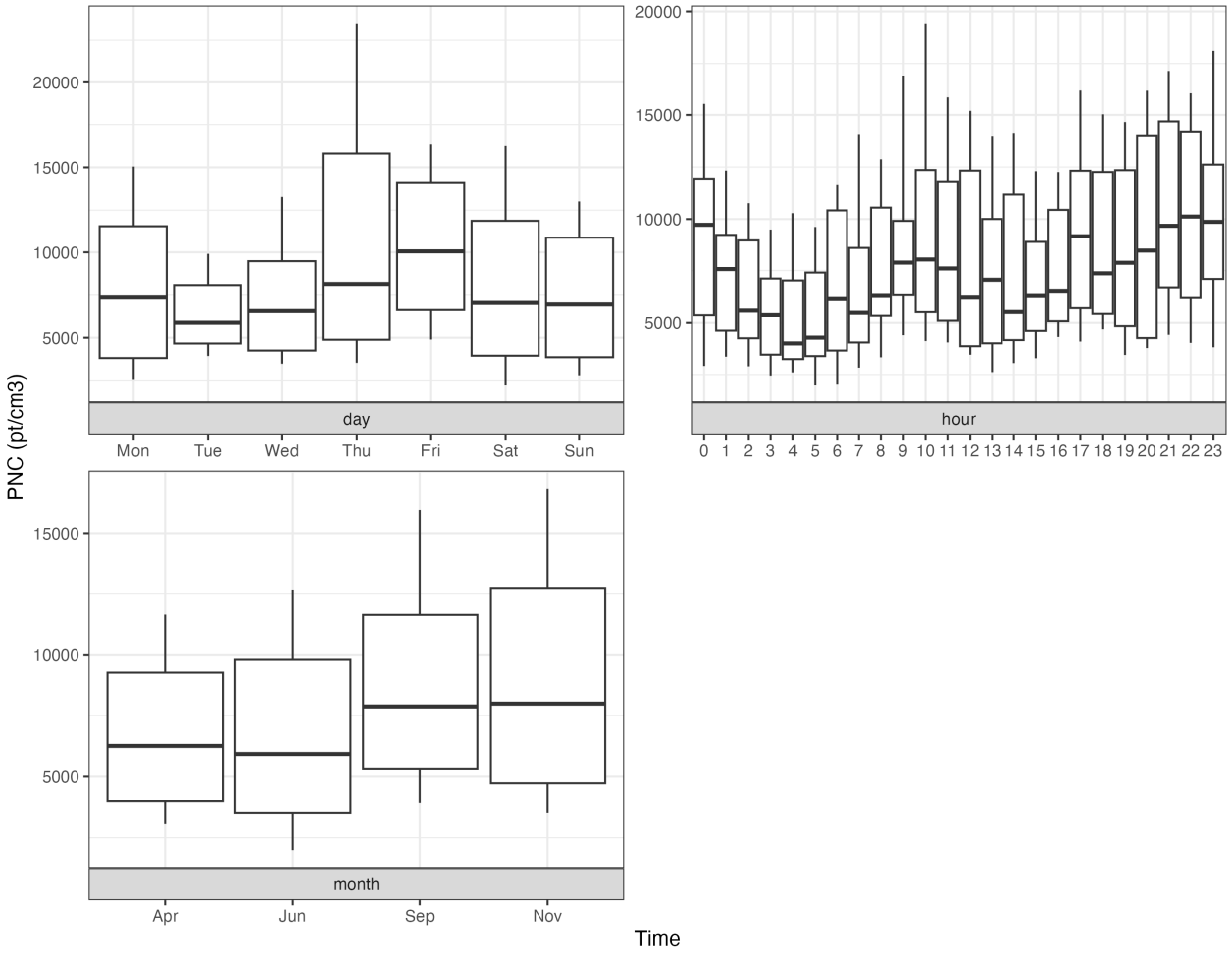



*Figure S4.2. Distribution of hourly PNC levels at Beacon Hill, stratified by day of the week, hour of the day, and sampling month.*

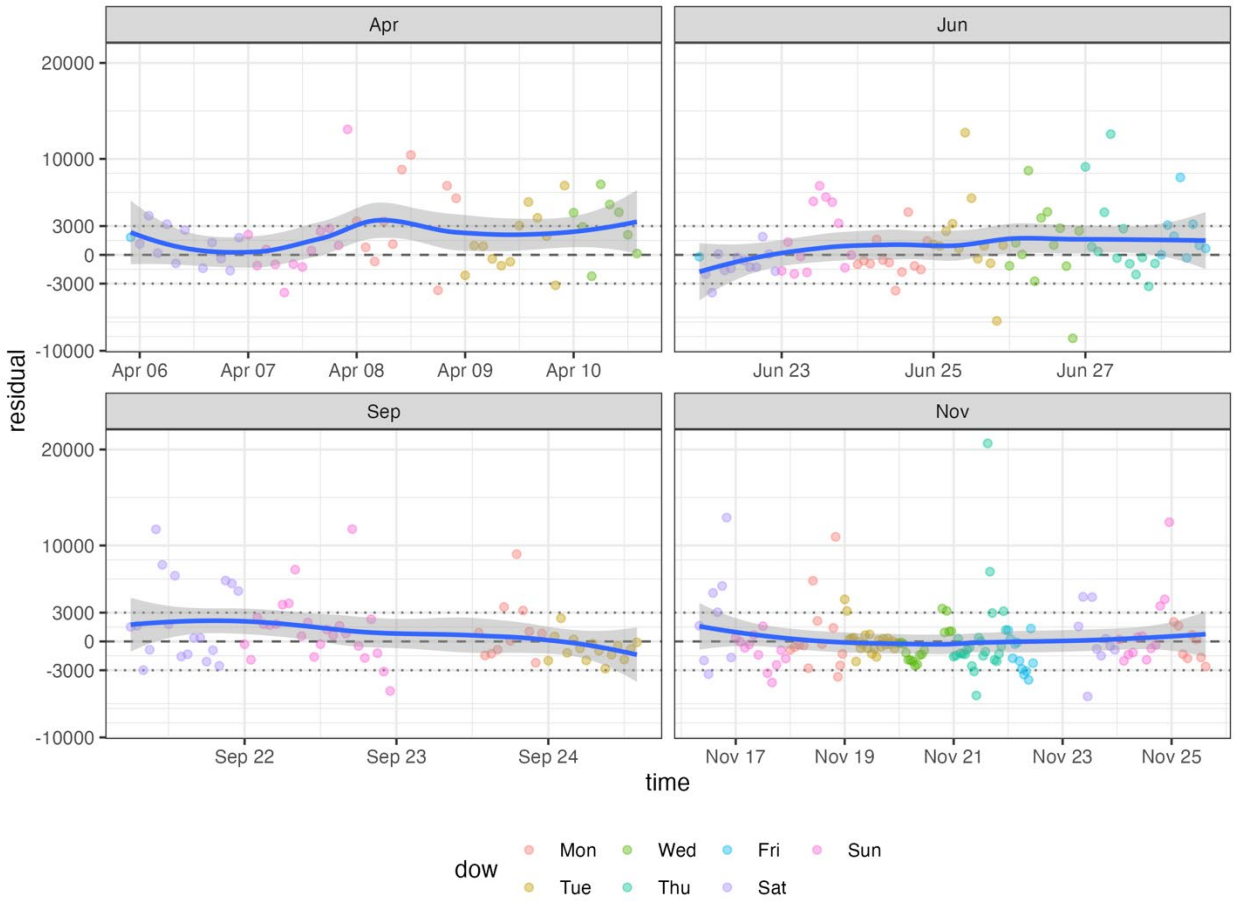



*Figure S4.3. Time series of simulated PNC model residuals using hourly Beacon Hill data. There is little remaining temporal trend in PNC.*

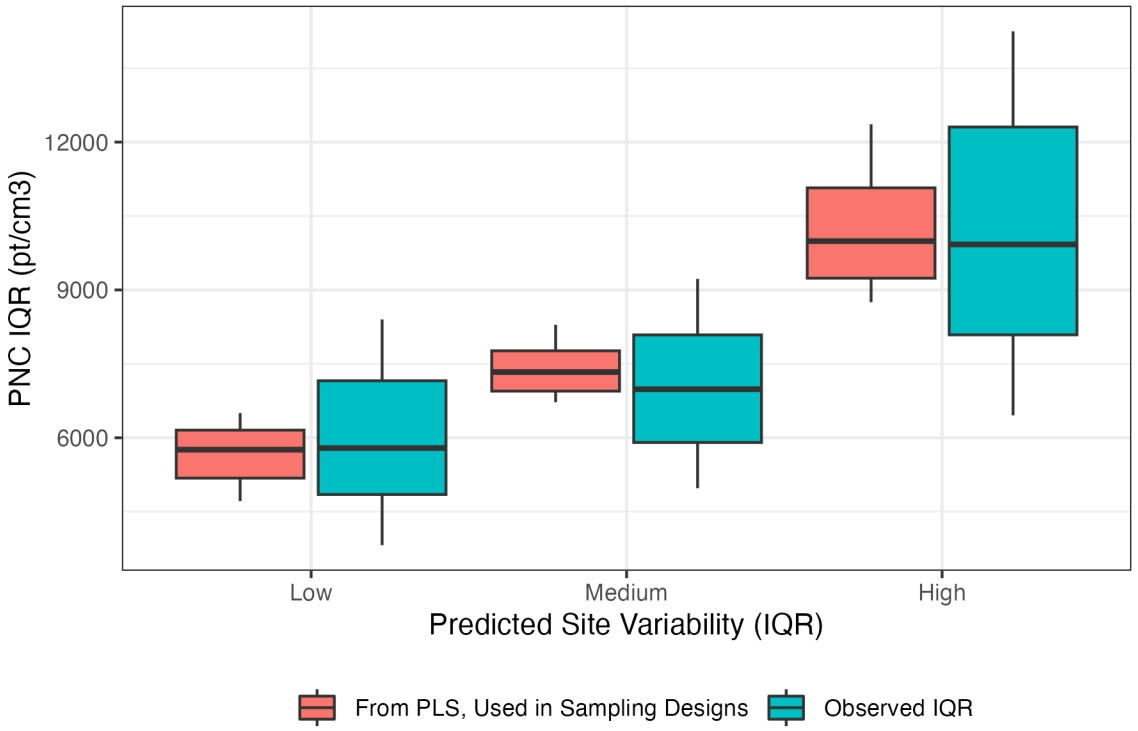



*Figure S4.4. Distribution of PNC IQR (pt/cm$^3$) for the unbalanced visits design, which categorizes sites as having low, medium, and high variability based on predicted IQR from PLS regression (see Chapter 4 Methods for modeling details). The observed IQR used to fit the model is also shown. Boxes show the median and IQR, whiskers show the 10$^{th}$ and 90$^{th}$ percentiles. Based on the stationary roadside NanoScan data from the Seattle mobile monitoring campaign.*

## RESULTS

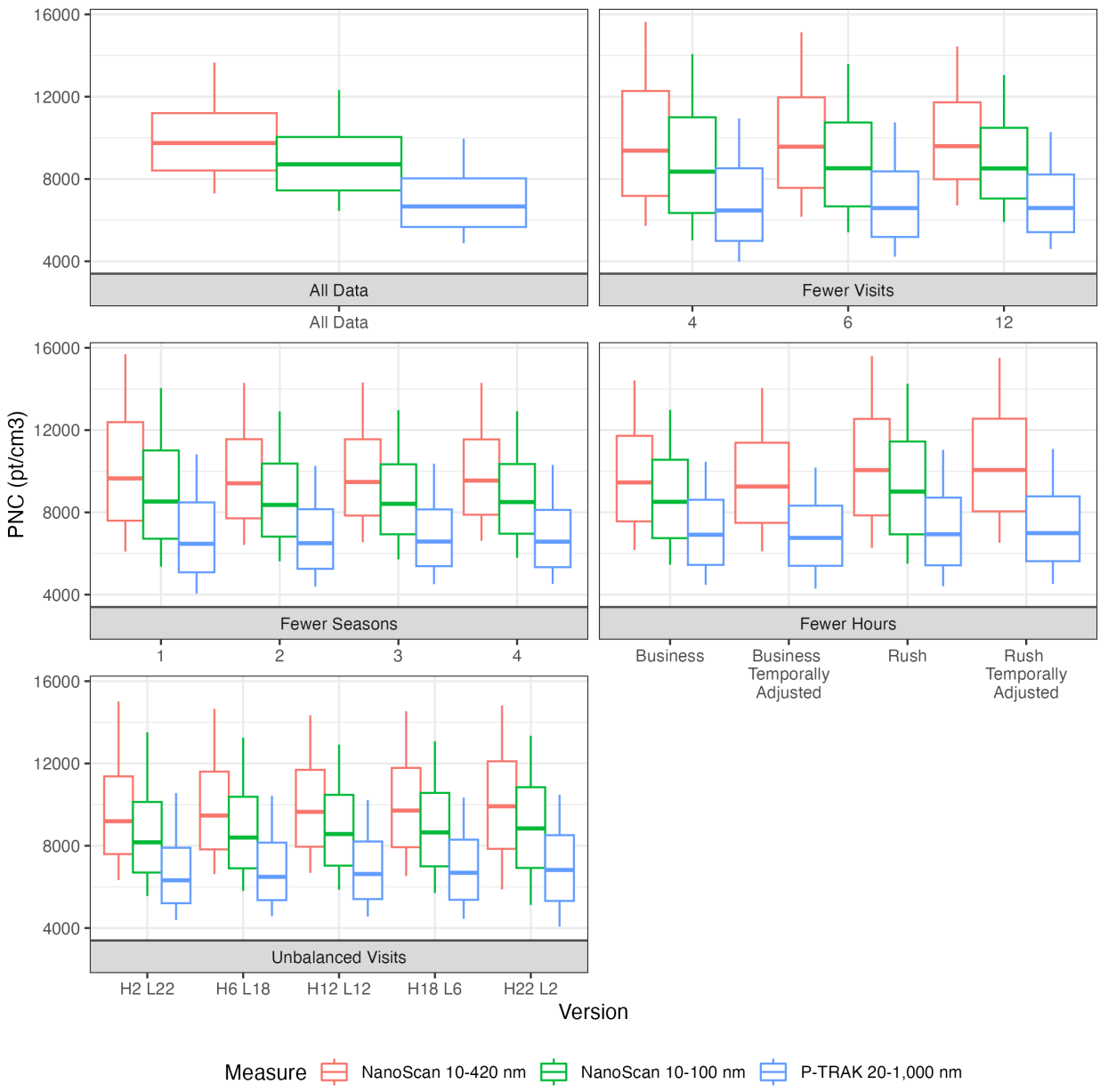



*Figure S4.5. Distribution of estimated annual average site concentrations for the all-data campaign (N=1 campaign x 309 sites) and each sampling design (N=30 campaigns per design x 309 sites each) for primary (NanoScan 10-420 nm) and sensitivity analyses using stationary roadside data from the Seattle mobile monitoring campaign.*

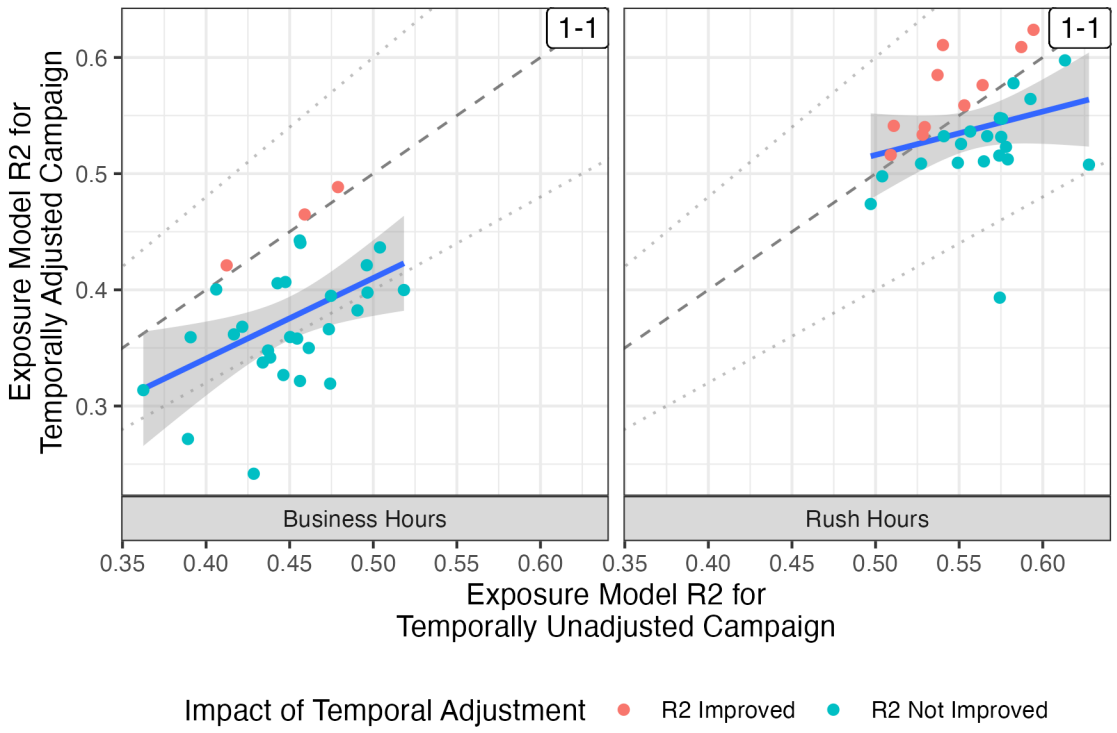



*Figure S4.6. Comparison of unadjusted and temporally adjusted business and rush hour campaign exposure models using stationary roadside NanoScan data from the Seattle mobile monitoring campaign. Temporal adjustment rarely improves business hour campaign models, and only sometimes improves rush hour campaign models. Dashed lines indicate the 1-1 and ±20% lines.*

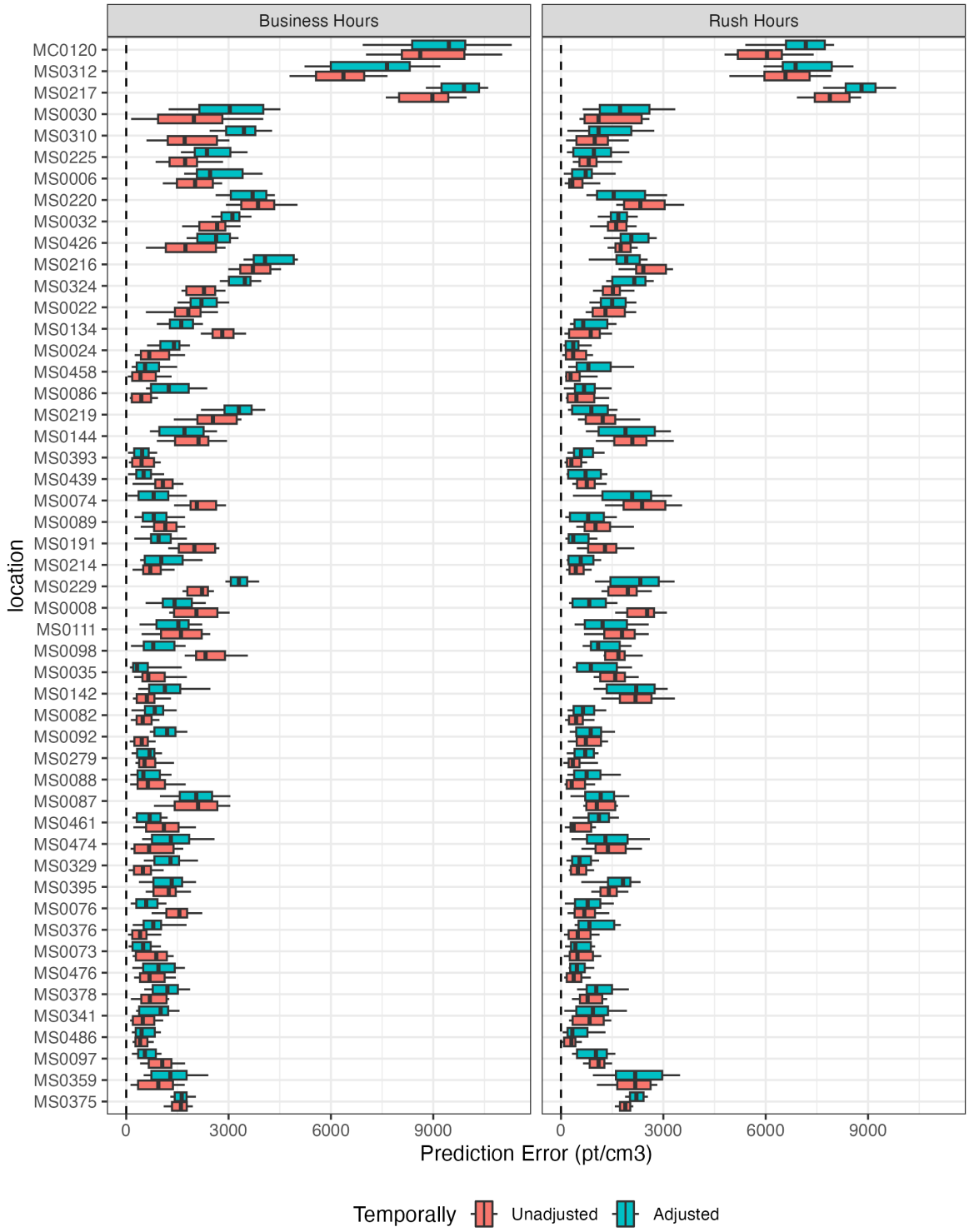



*Figure S4.7. Absolute UFP prediction errors from the business and rush hour designs for a sample of 50 sites (n=30 prediction errors per site and adjustment approach, i.e., per boxplot) using stationary roadside NanoScan data from the Seattle mobile monitoring campaign. Prediction errors are calculated by comparing the cross-validated site prediction from each campaign to the all-data site observation. Sites are arranged by their all-data annual average concentration, with higher concentration sites near the top.*

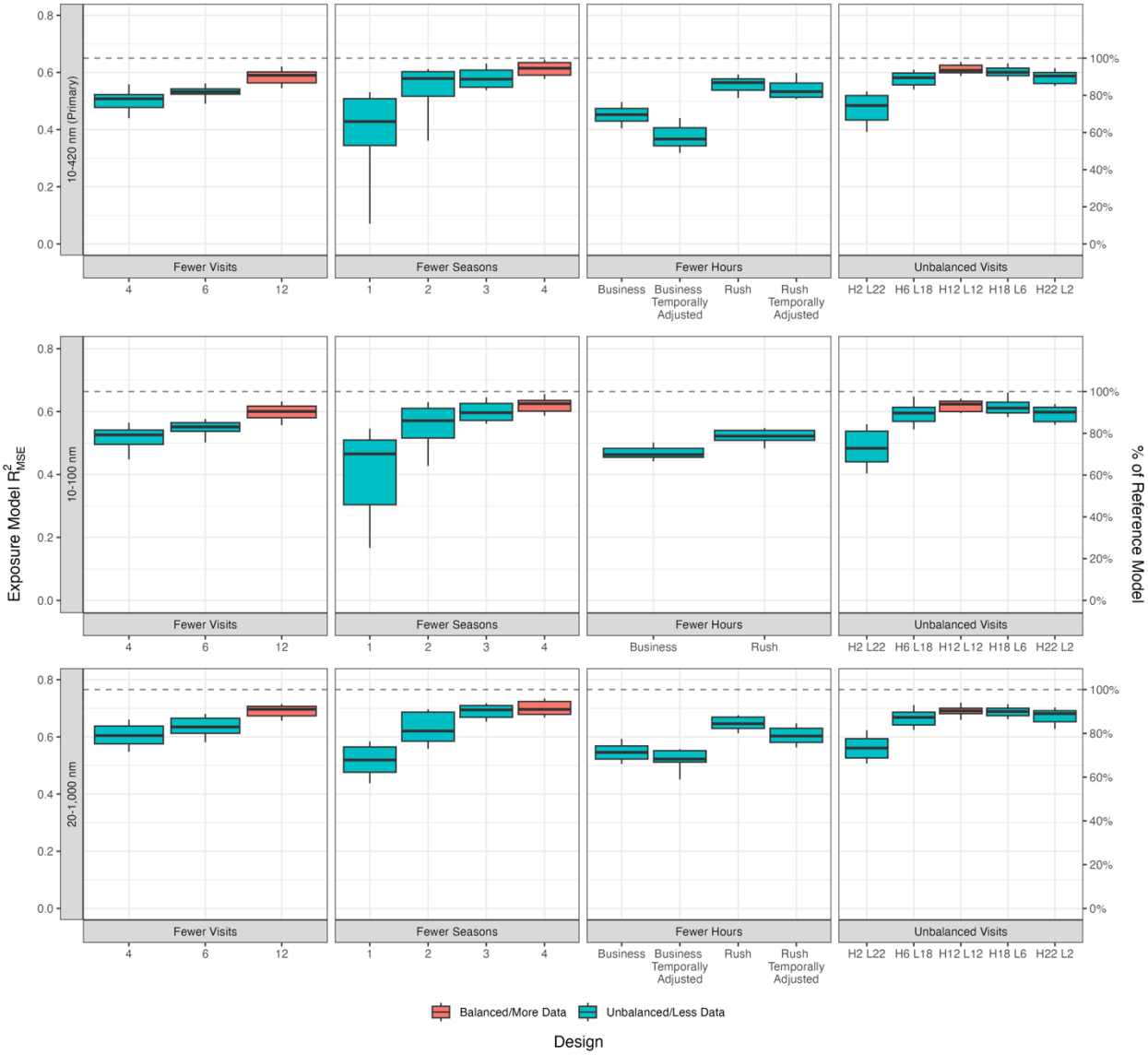
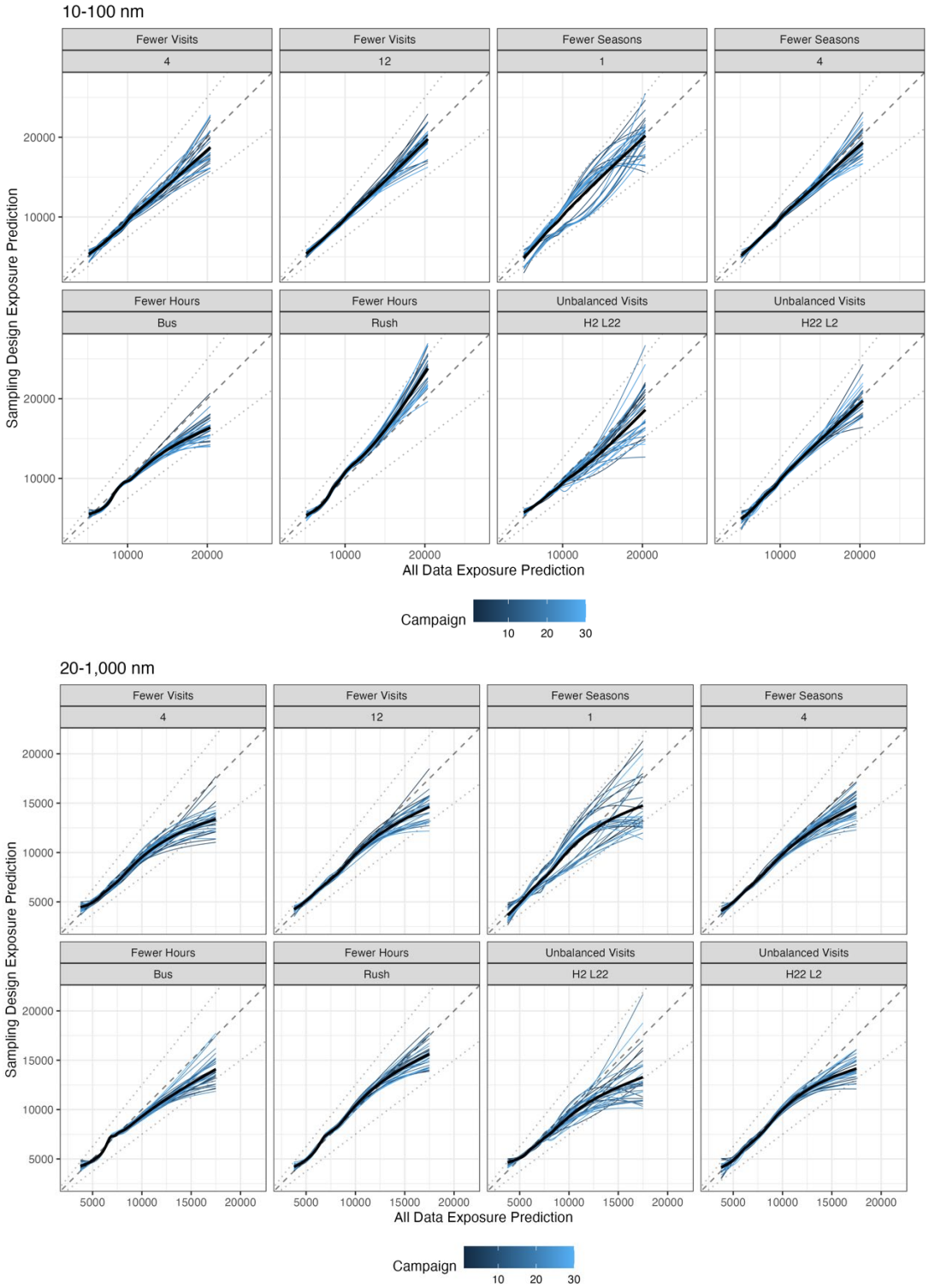



*Figure S4.8. Cross-validated model performances (N=30 campaigns per design) using stationary roadside data from the Seattle mobile monitoring campaign for primary and sensitivity exposure models. The dashed lines indicate the all-data campaign performance for primary and sensitivity analyses.*

(Figure S4.9. —see below for caption)



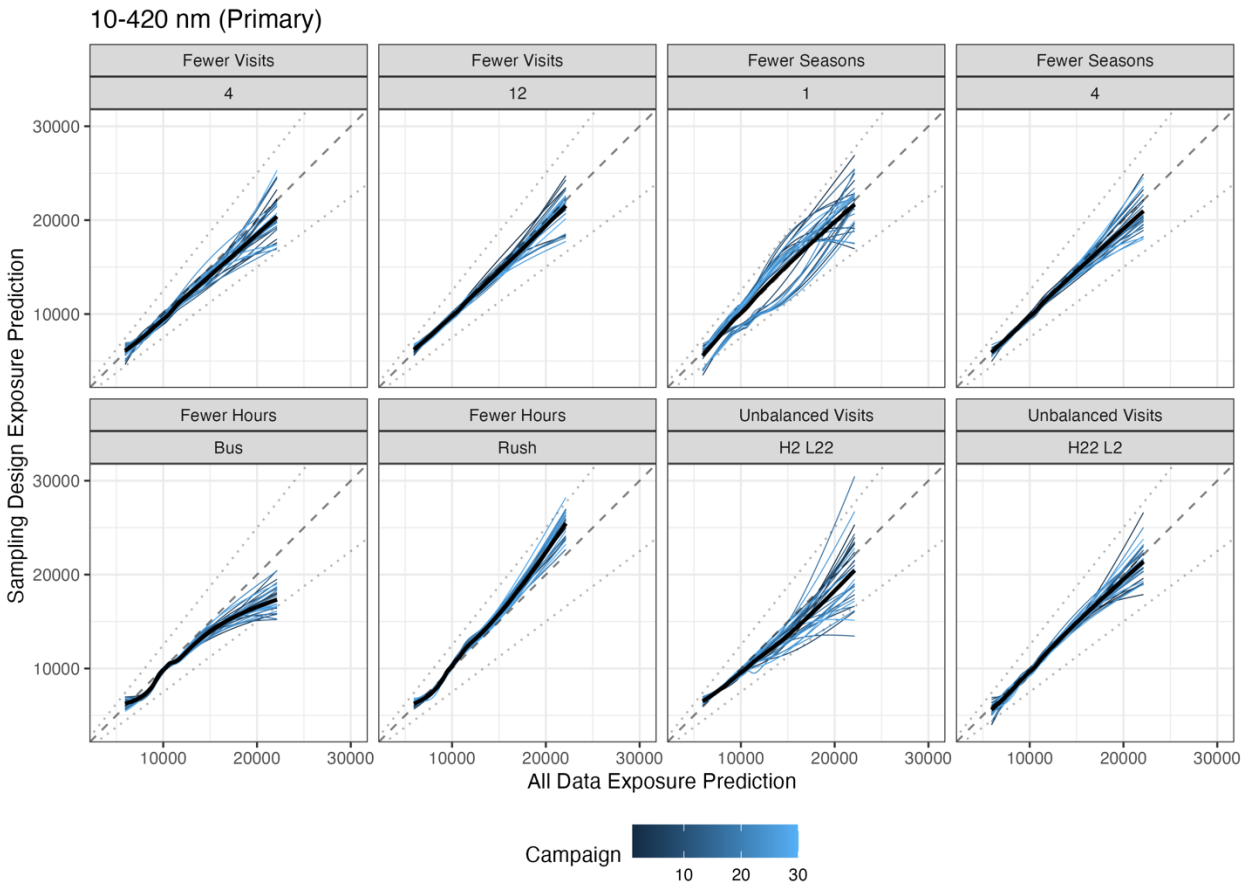



*Figure S4.9. Smooth lines comparing predicted five-year average participant PNC (pt/cm3) exposure from the all-data campaign (N=1 campaign) to exposure predictions from other sampling designs (N=30 campaigns) for primary and sensitivity exposure analyses using stationary roadside data from the Seattle mobile monitoring campaign. The black smooth line is the average trend for each design. The dashed lines indicate the 1-1 line as well as 25% above and below.*

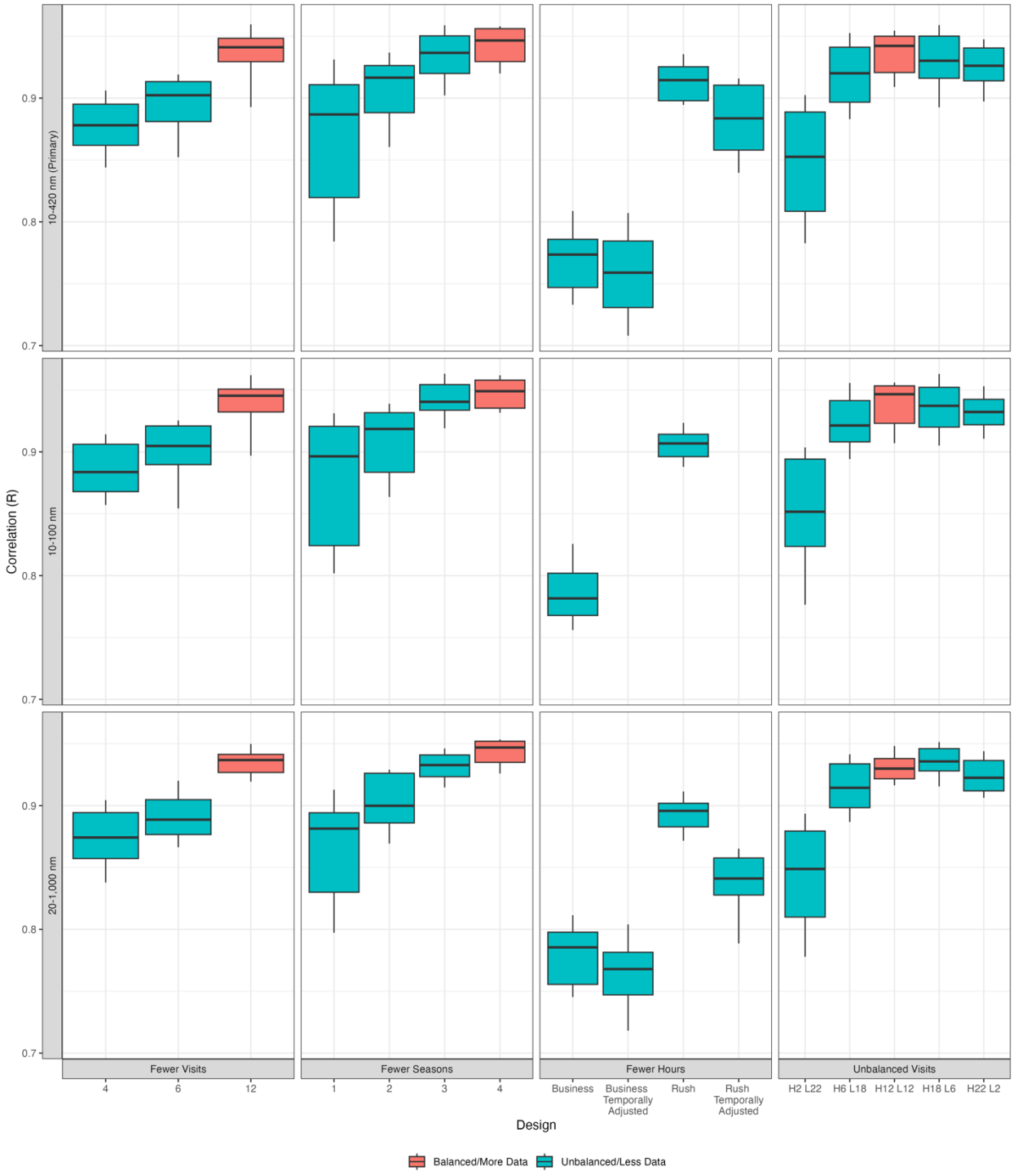



*Figure S4.10. Pearson correlations (R) for predicted five-year average air pollution participant exposure from the all-data campaign and each sampling design (N=30 campaign correlations per design) using stationary roadside data from the Seattle mobile monitoring campaign. Boxes show the median and IQR, whiskers show the 10th and 90th percentiles.*

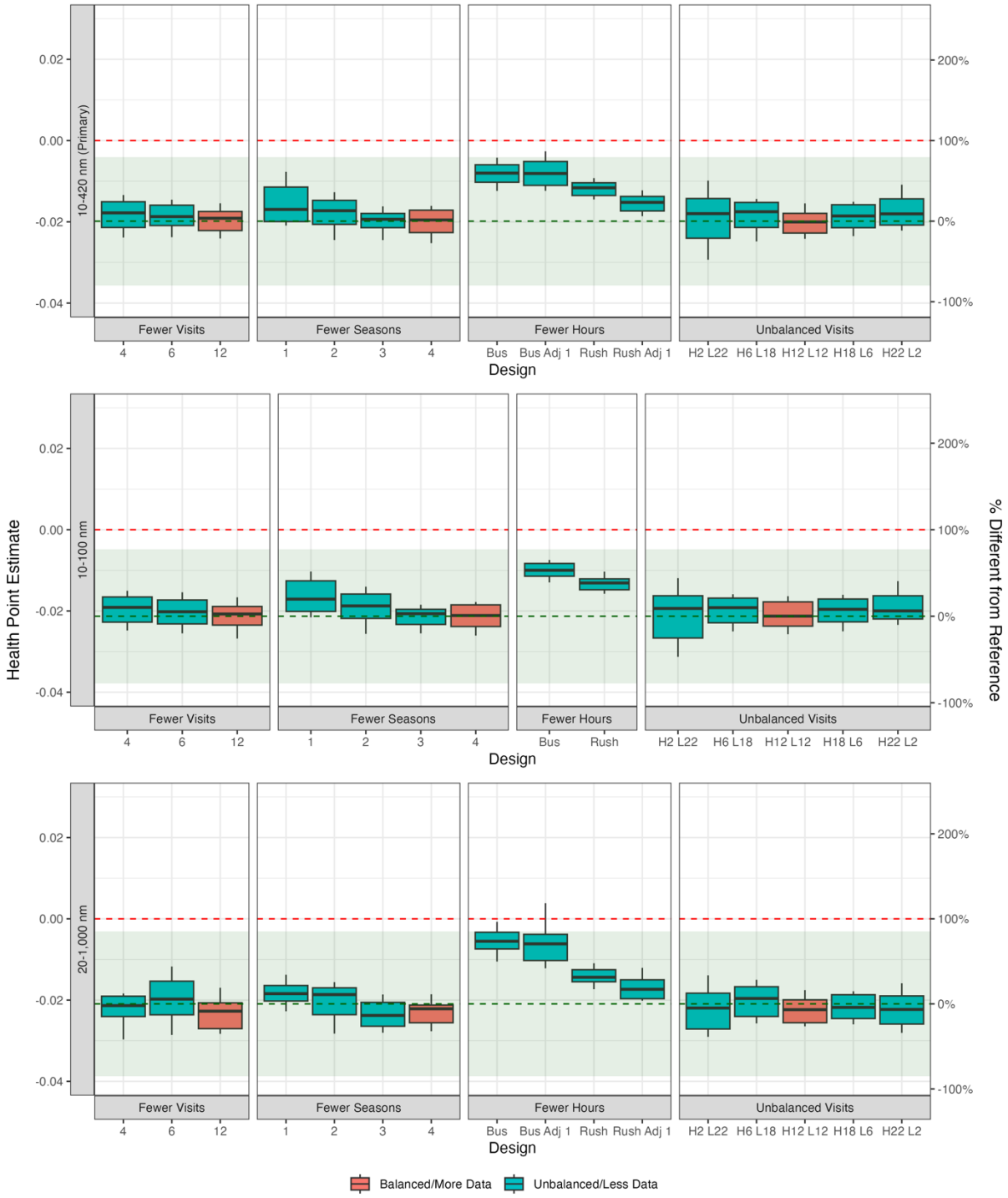
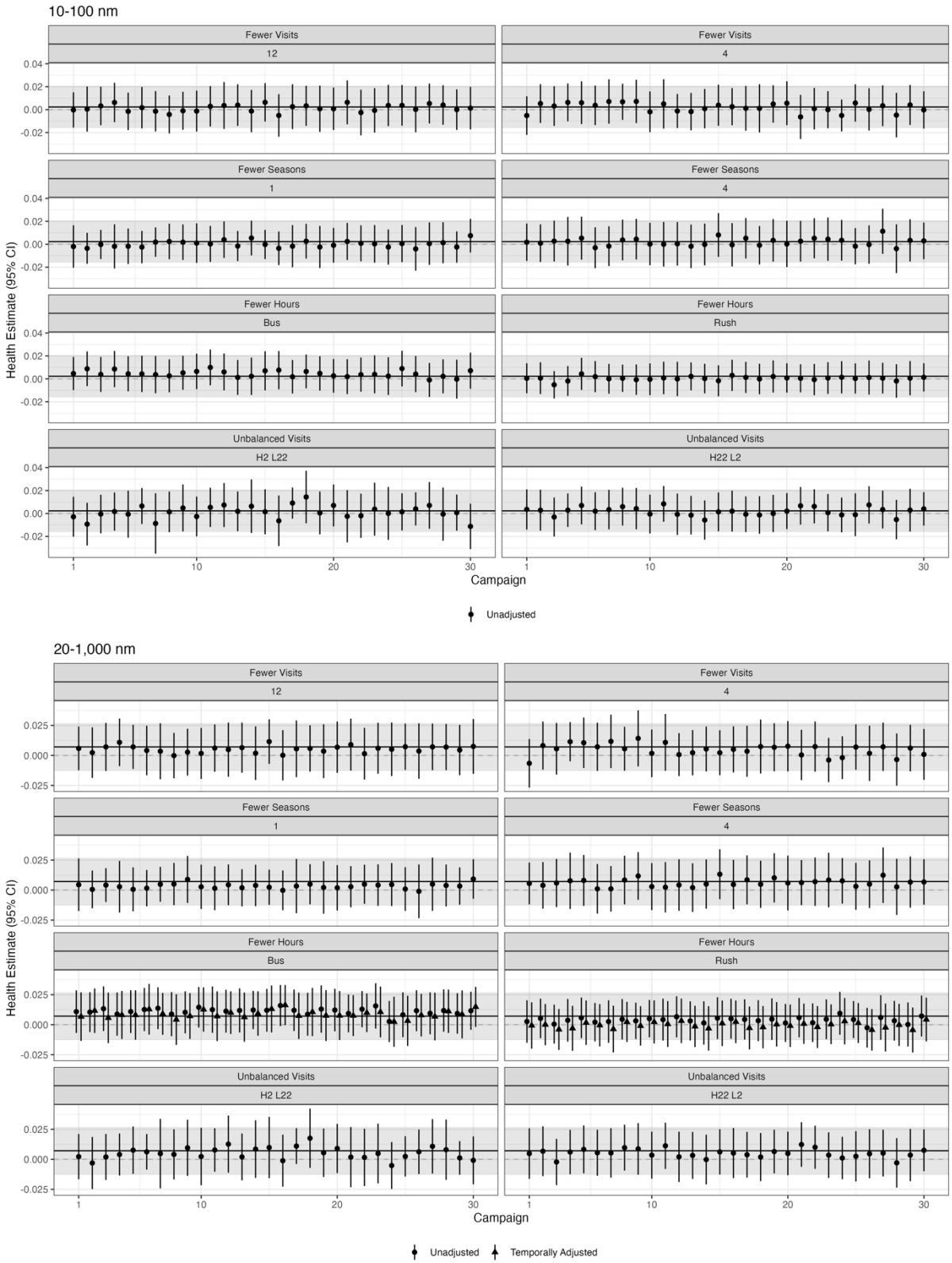



*Figure S4.11. Health estimates produced from different sampling designs for the health model between PNC (1,900 pt/cm3) and cognitive function (CASI-IRT) for primary (10-420 nm PNC) and sensitivity exposure analyses using stationary roadside data from the Seattle mobile monitoring campaign. Percentages on the y-axis are relative to the all-data exposure model health estimate. Boxes show the median and IQR, whiskers show the 10th and 90th percentiles.*

(Figure S4.12. — see below for caption)



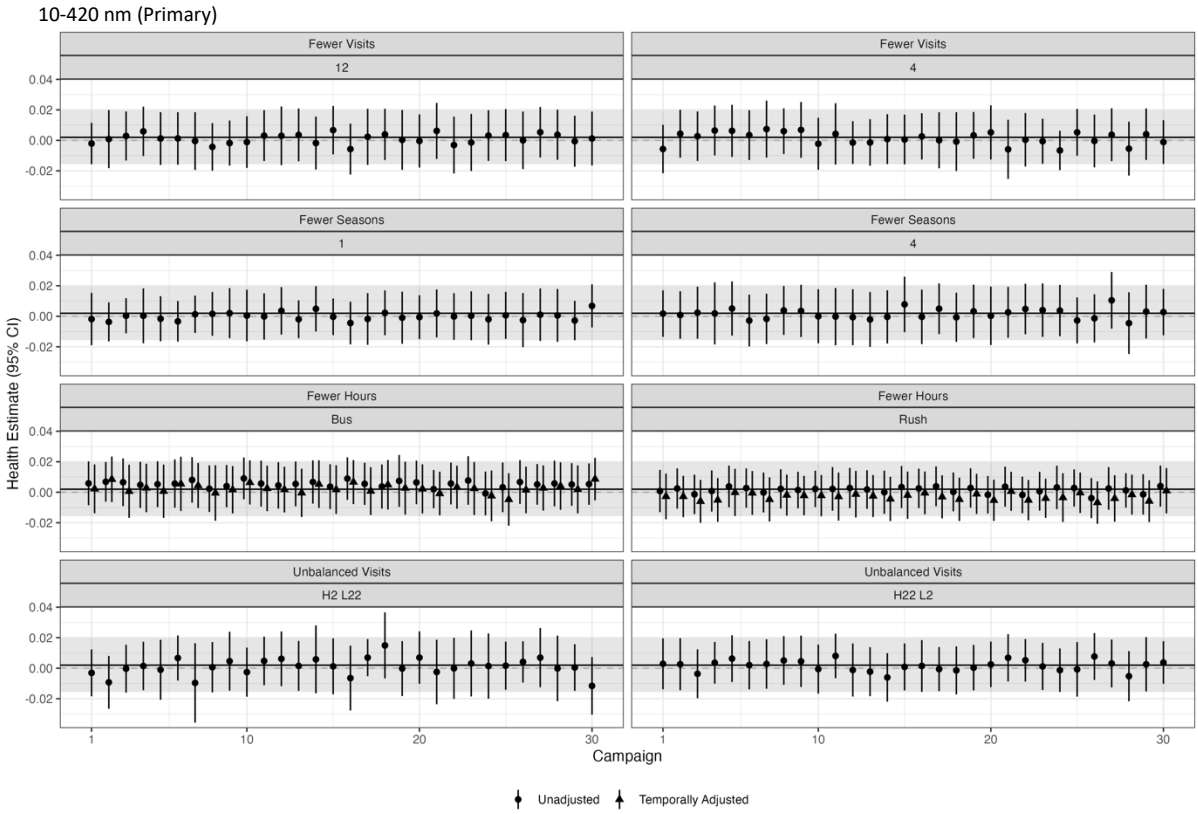



*Figure S4.12. Health estimates for the association between UFP (1,900 pt/cm$^3$) and CASI-IRT for primary (bottom set of panels) and sensitivity exposure analyses after adjusting for age, calendar year, sex, and education (main health model) using stationary roadside data from the Seattle mobile monitoring campaign. The horizontal black line and shaded gray area are the point and 95% confidence interval estimates from the all-data campaign, respectively. Each point range is the estimated health effect and 95% CI for a given sampling campaign.*

# Chapter 6

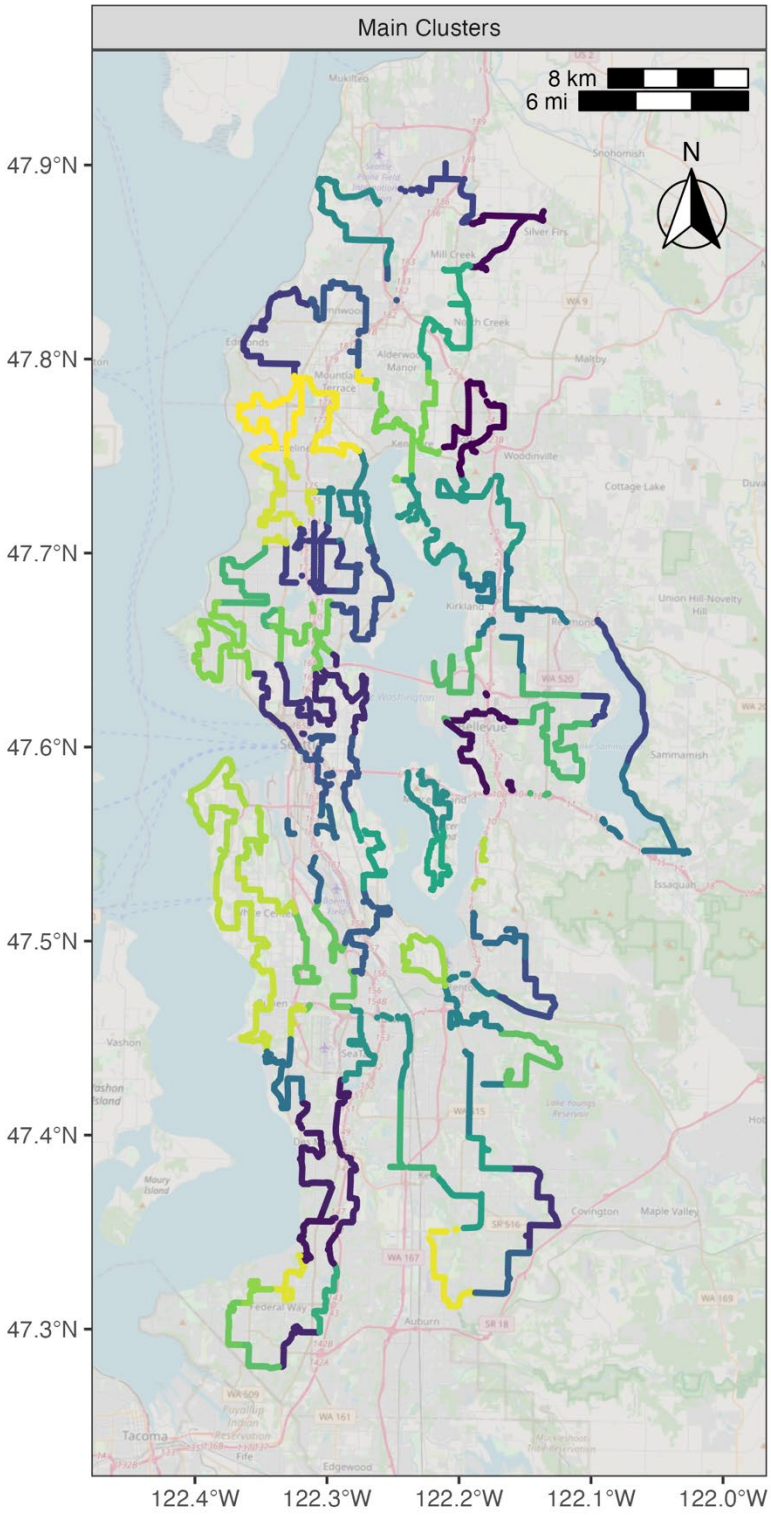

Main Clusters
8 km
6 mi
N
47.9°N
47.8°N
47.7°N
47.6°N
47.5°N
47.4°N
47.3°N
122.4°W
122.3°W
122.2°W
122.1°W
122.0°W

*Figure S6.1. Map of spatial clusters used to conduct the unbalanced sampling designs in the Seattle mobile monitoring campaign. There is an average of 93 100 m segments in each cluster, grouped based on their location along driving routes.*

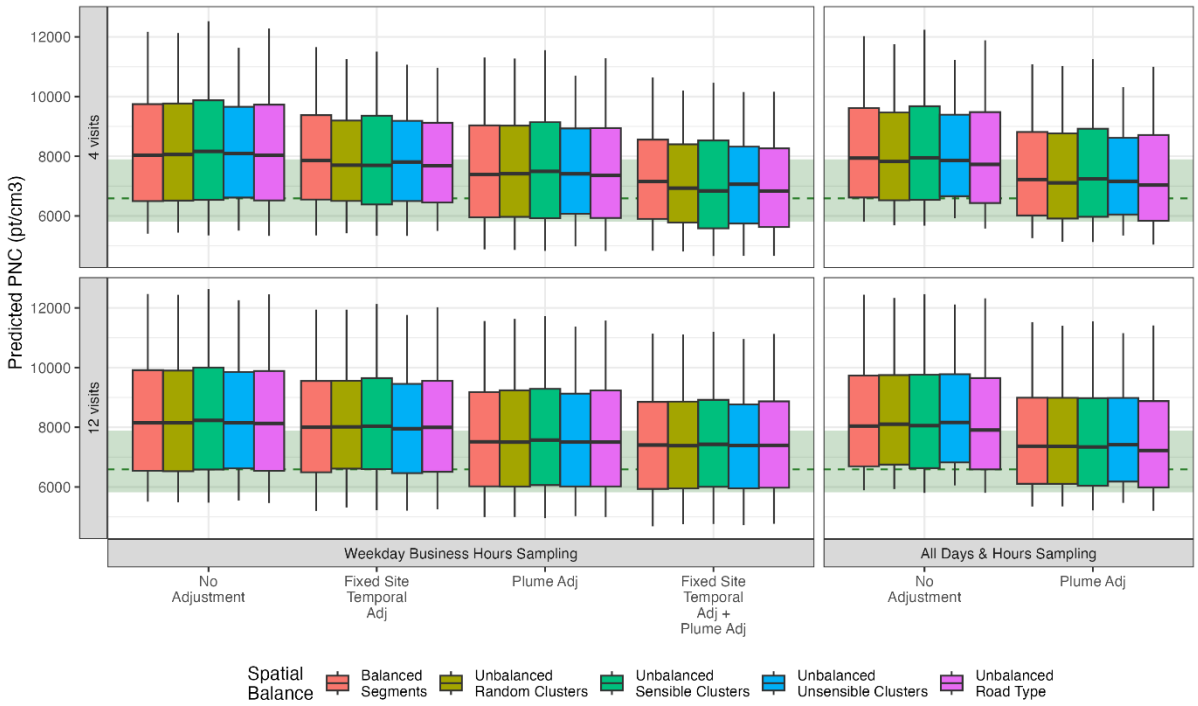



*Figure S6.2. Predicted annual average PNC at 309 stationary locations from on-road campaign exposure models. The dashed green line and area indicate the median (IQR) cross-validated PNC prediction from the reference all-data stationary exposure model. Boxes show the median and IQR, whiskers show the 10$^{th}$ and 90$^{th}$ percentiles.*

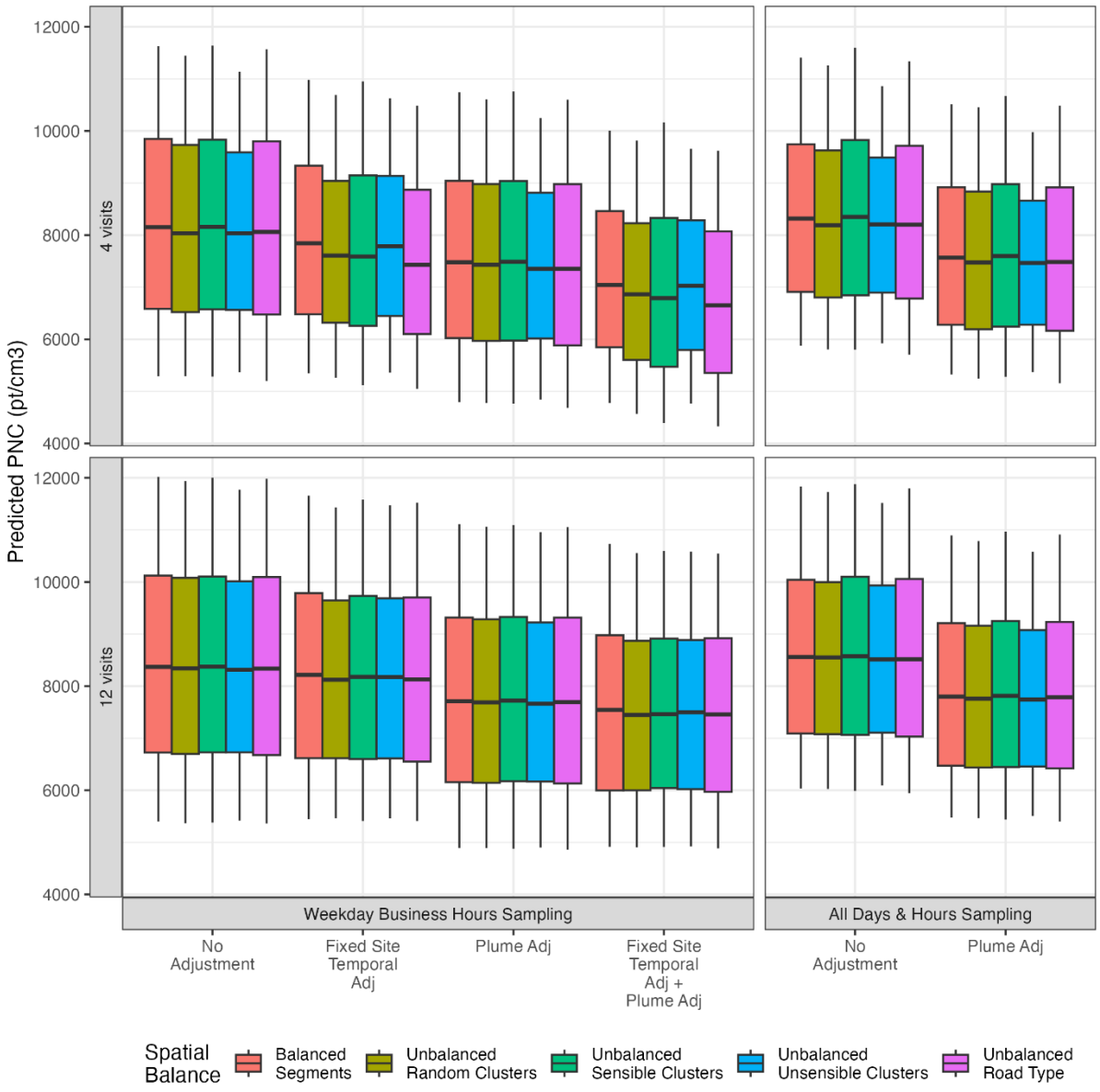



*Figure S6.3. Predicted PNC exposure for ACT participants across 30 sampling campaigns for each design (i.e., boxplot). Boxes show the median and IQR, whiskers show the 10$^{th}$ and 90$^{th}$ percentiles.*

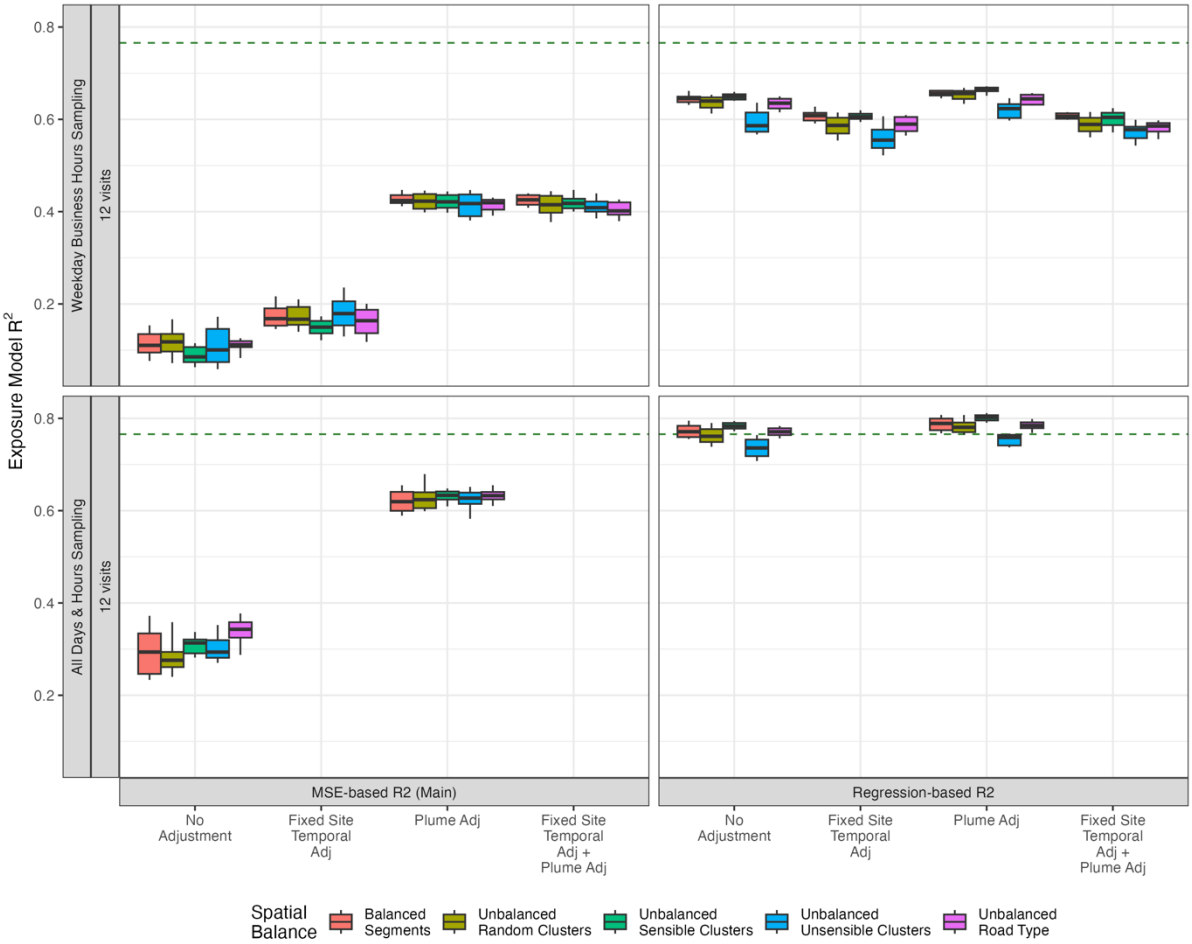



*Figure S6.4. Out-of-sample PNC (pt/cm$^3$) exposure model performances for on-road campaigns (N=30 campaigns per combination - i.e., boxplot). $R^2_{reg}$ is based on the comparison of the predicted PNC at 309 stationary locations and the annual average site estimate based on stationary measures. Models are for total PNC (20-1,000 nm) from the unscreened P-TRAK instrument. Boxes show the median and IQR, whiskers show the 10$^{th}$ and 90$^{th}$ percentiles. The dashed line indicates the $R^2$ from the reference all-data stationary model, which is 0.77, and the same for $R^2_{MSE}$ and $R^2_{reg}$.*

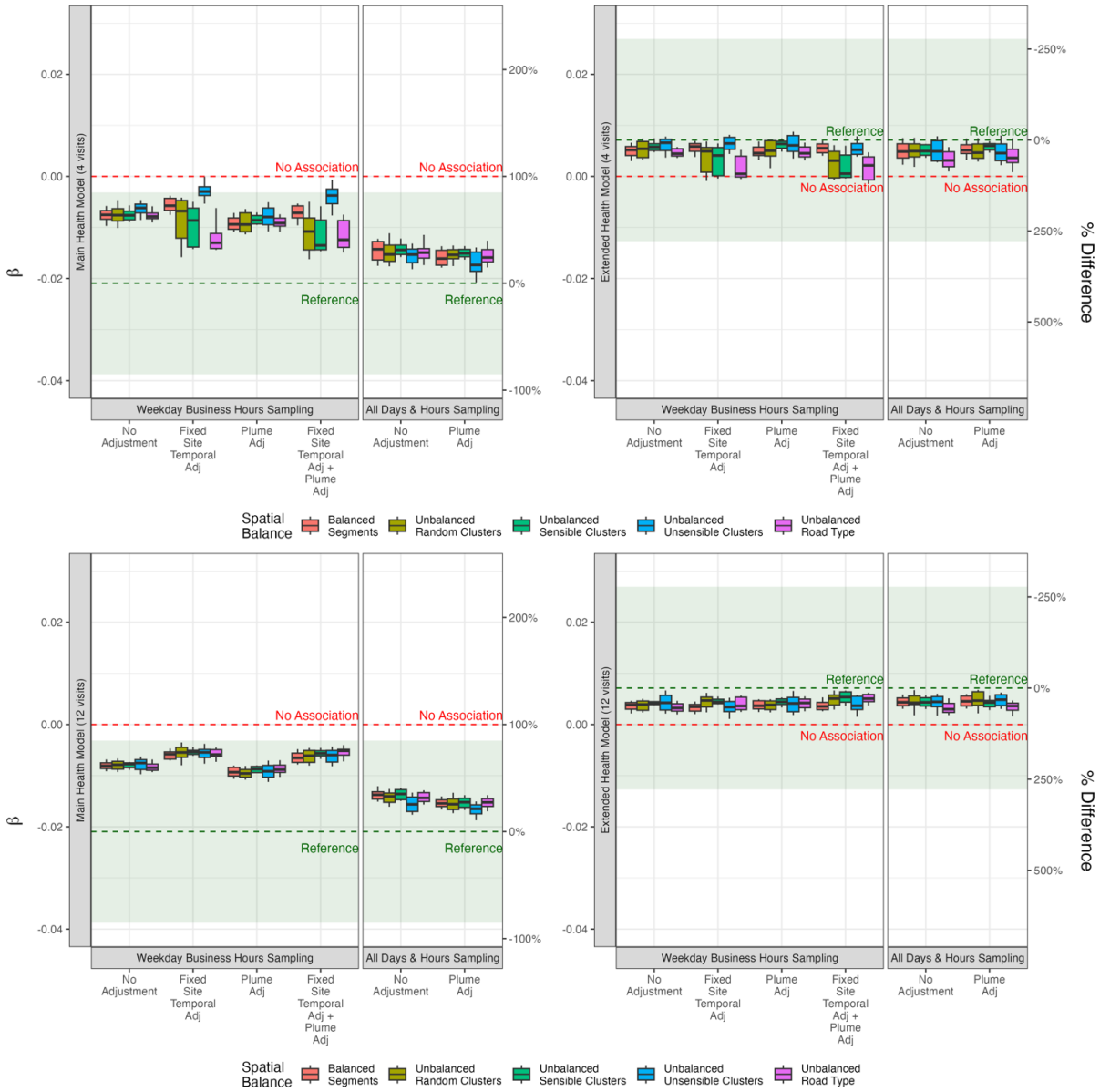



*Figure S6.5. Estimated association between cognitive function (CASI-IRT) and PNC (1,900 pt/cm$^3$) at baseline in the ACT cohort in the main and extended adjusted health models for 4-visit campaigns, with the 12-visit included for comparison. PNC exposures are predicted from on-road monitoring campaigns. The dashed green lines and shaded areas indicate the estimated point and 95% CIs from the all-data roadside exposure model, which are -0.021 (95% CI: -0.039, -0.003) in the main model and 0.007 (95% CI: -0.013, 0.027) in the extended model. The dashed red line indicates no association. Boxes show the median and IQR, whiskers show the 10$^{th}$ and 90$^{th}$ percentiles.*

# Chapter 7

## SUPPLEMENTARY MONITORING CAMPAIGN DESCRIPTIONS

Data from supplementary monitoring campaigns that were included in the spatiotemporal models included:

Diesel Exhaust Exposure in the Duwamish Study (DEEDS): measured $PM_{2.5}$, light-absorbing carbon, $NO_x$, $NO_2$, and $SO_2$ at sites in the South Park and Georgetown neighborhoods of Seattle. Ogawa Passive Samplers were deployed on 75 utility poles to measure $NO_2$, over two-week periods in a summer/non-heating season (August 18-30) and a winter/heating season (December 1-14, 2012).

Panel Study on Income Dynamics (PSID): air monitoring was conducted from October 1999 – May 2001 (in different seasons) at the homes of 108 individuals recruited from clinics, senior centers, retirement homes, and an allergy and asthma clinic. Ogawa Passive Samplers were used to measure $NO_2$ over 10-day periods.

Multi-Ethnic Study of Atherosclerosis and Air Pollution (MESA Air) Pilot (MAP): A pilot conducted to refine the field sampling protocol for MESA Air snapshot sampling. A grid-like pattern of Ogawa Passive Samplers was deployed on utility poles in residential neighborhoods across northern Seattle, including neighborhoods west of Highway 99, the University of Washington campus, between the University and Lake City Way; locations along I-5; locations along Highway 522/Lake City Way; and one site co-located with the AQS site at Beacon Hill. The study was conducted March 3-17, 2005.

Yesler Terrace study: The Seattle Housing Authority was planning a redevelopment of the Yesler Terrace neighborhood. Ogawa Passive Samplers were deployed at 28 locations in February 2010 and 86 locations in March 2010 to measure $NO_x$ and $NO_2$.

### $PM_{2.5}$ Data Description

| Category | Type | N of Monitors | N of Two-Week Measurements | Start Date | End Date | Mean $PM_{2.5}$ (SD) ($\mu g/m^3$) | Min, Max $PM_{2.5}$ ($\mu g/m^3$) |
|---|---|---|---|---|---|---|---|
| Reference-Grade Monitors | FRM | 19 | 3822 | 1/2/2010 | 9/26/2020 | 6.6 (5.5) | 0.4, 100.8 |
| | TEOM | 5 | 714 | 1/14/2012 | 9/26/2020 | 6.7 (6.1) | 1.2, 83.2 |
| | BAM | 2 | 184 | 2/20/2016 | 9/26/2020 | 7.0 (8.4) | 1.5, 81.7 |
| | NEPH | 27 | 1418 | 1/2/2010 | 9/26/2020 | 6.1 (5.0) | 1.3, 96.3 |
| | DEEDS | 25 | 48 | 8/25/2012 | 12/8/2012 | 5.3 (0.7) | 3.4, 7.4 |
| ACT-AP LCMs | Two-Period | 82 | 502 | 6/24/2017 | 5/25/2019 | 6.1 (2.2) | 2.6, 17.0 |
| | One-Period | 30 | 335 | 4/29/2017 | 9/26/2020 | 5.8 (2.1) | 2.8, 13.5 |

*Table S7.1: Summary statistics of the $PM_{2.5}$ monitors and measurements for the 2010 to 2020 time period. Summary statistics are calculated from two-week average concentrations. LCM: low-cost monitor; SD: standard deviation; FRM: Federal Reference Method; TEOM: tapered element oscillating microbalance; BAM: beta attenuation monitoring; NEPH: nephelometer; DEEDS: the Diesel Exhaust Exposure in the Duwamish Study; ACT-AP: the Adult Changes in Thought Air Pollution. Reprinted with permission from Bi et al. 2024 (Copyright 2024 Elsevier)*

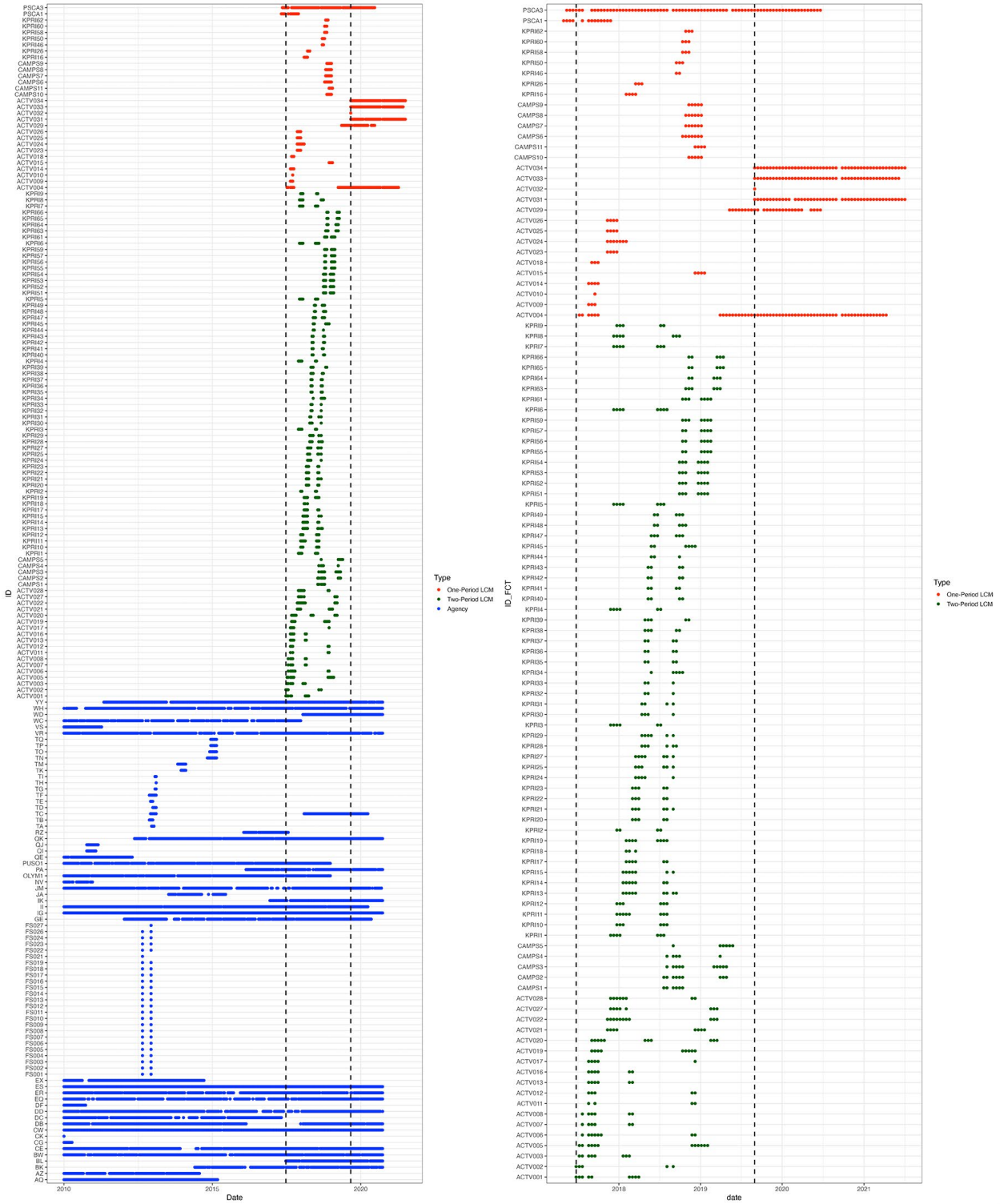



*Figure S7.1: **Left**: Time points of two-week measurements from different types of $PM_{2.5}$ monitors for 2010-2020, including the agency monitors (FRM, TEOM, BAM, NEPH, and DEEDS), as well as LCMs (the ACT-AP one- and two-period low-cost monitors). Note that the one-period monitors were not included in the models for 2010-2020. **Right**: The Adult Changes in Thought Air Pollution (ACT-AP) one- and two-period low-cost monitors. Reprinted with permission from Bi et al. 2024 (Copyright 2024 Elsevier)*

| Category | Type | N of Monitors | N of Two-Week Measurements | Start Date | End Date | Mean $PM_{2.5}$ (SD) ($\mu g/m^3$) | Min, Max $PM_{2.5}$ ($\mu g/m^3$) |
|---|---|---|---|---|---|---|---|
| Reference-Grade Monitors | FRM | 35 | 10,429 | 10/6/1979 | 9/26/2020 | 9.1 (7.1) | 0.4, 100.8 |
| | TEOM | 6 | 859 | 10/7/2000 | 9/26/2020 | 7.0 (5.7) | 1.2, 83.2 |
| | BAM | 2 | 184 | 2/20/2016 | 9/26/2020 | 7.0 (8.4) | 1.5, 81.7 |
| | NEPH | 48 | 3,491 | 2/25/1978 | 9/26/2020 | 9.8 (8.5) | 1.3, 98.9 |
| | DEEDS | 25 | 48 | 8/25/2012 | 12/8/2012 | 5.3 (0.7) | 3.4, 7.4 |
| ACT-AP LCMs | One- and Two-Period | 112 | 909 | 4/29/2017 | 7/3/2021 | 5.9 (2.1) | 2.6, 17.0 |

*Table S7.2: Summary statistics of the $PM_{2.5}$ monitors and measurements for the extended period 1978-2021. LCM: low-cost monitor; SD: standard deviation; FRM: Federal Reference Method; TEOM: tapered element oscillating microbalance; BAM: beta attenuation monitoring; NEPH: nephelometer; DEEDS: the Diesel Exhaust Exposure in the Duwamish Study; ACT-AP: the Adult Changes in Thought Air Pollution. Unlike the primary models for 2010-2020 without the one-period LCMs, because the extended-period models included all one-period LCMs, their modeling end year was 2021 rather than 2020. Note that all one-period LCM measurements after 9/26/2020 were excluded from the independent external validation of the primary models for 2010-2020. Reprinted with permission from Bi et al. 2024 (Copyright 2024 Elsevier)*

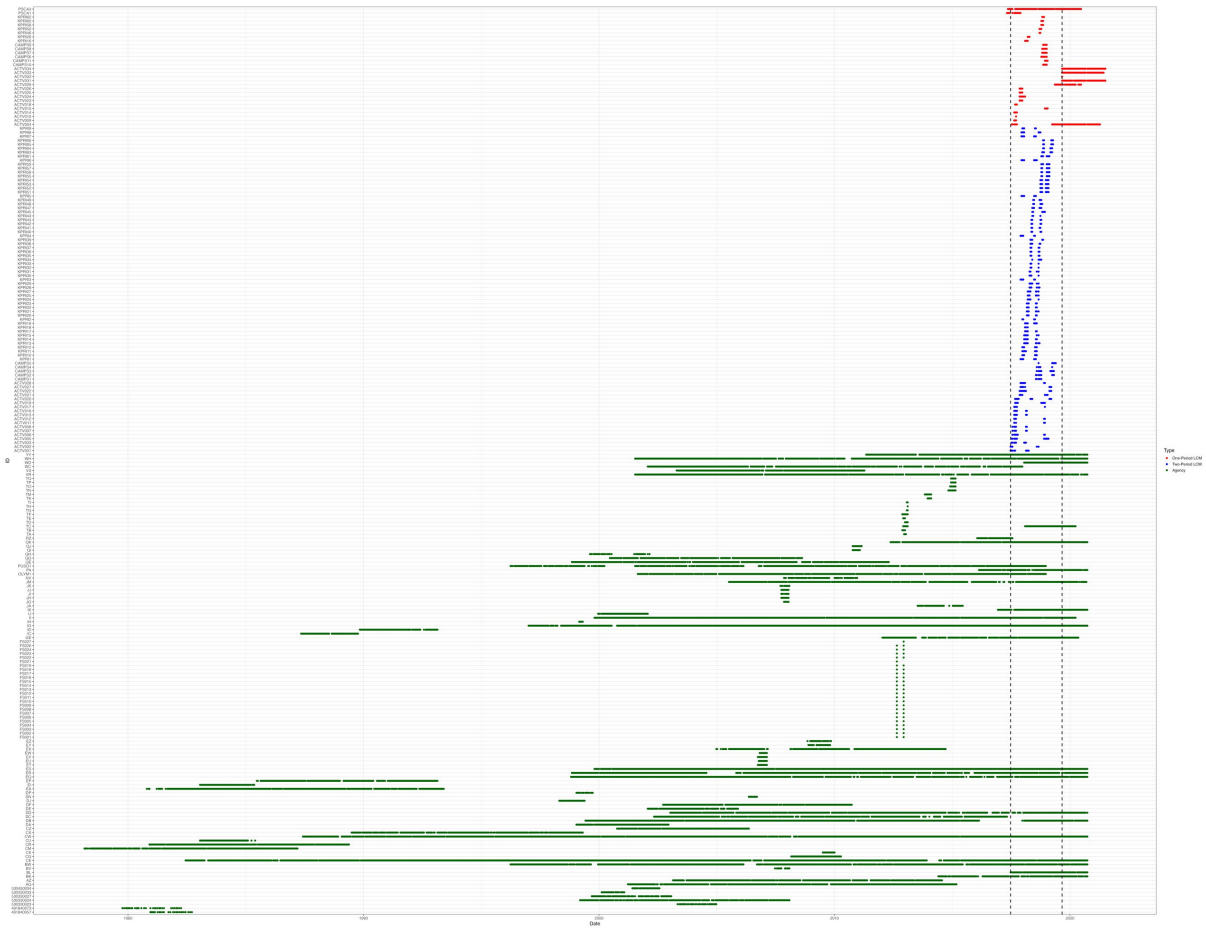


*Figure S7.2: Timepoints of two-week $PM_{2.5}$ measurements from different types of monitors for 1978-2021, including the agency monitors (Federal Reference Method [FRM] monitors, Federal Equivalent Method [FEM] monitors using the tapered element oscillating microbalance [TEOM] and beta attenuation monitoring [BAM] methods, nephelometers [NEPH], and the monitors from the Diesel Exhaust Exposure in the Duwamish Study [DEEDS]), as well as LCMs (the Adult Changes in Thought Air Pollution [ACT-AP] one- and two-period low-cost monitors). Reprinted with permission from Bi et al. 2024 (Copyright 2024 Elsevier)*

| Model | N of Reference-Grade Measurements (Monitors) for Modeling | N of Two-Period Low-Cost Measurements (Monitors) for Modeling | Mean (Standard Deviation) of Low-Cost Measurements ($\mu g/m^3$) | N of Combined Validation Measurements (Monitors) | Spatiotemporal Validation | | Spatial Validation | |
|---|---|---|---|---|---|---|---|---|
| | | | | | $R^2$ | RMSE ($\mu g/m^3$) | $R^2$ | RMSE ($\mu g/m^3$) |
| *Alldata* | 15418 (146) | 502 (82) | 1.7 (0.3) | 502 (82) | 0.84 | 0.9 | 0.55 | 0.6 |
| *Separate* | | 164 (82) | 1.7 (0.4) | | 0.81 | 1.0 | 0.46 | 0.6 |
| *Random Selection 25%* | | 131 (21) | 1.8 (0.3) | | 0.8 | 1.0 | 0.41 | 0.7 |
| *Short Distance to Road & High Population Density* | | 154 (26) | 1.8 (0.4) | | 0.8 | 1.0 | 0.45 | 0.7 |
| *No* | | 0 (0) | - | | 0.76 | 1.1 | 0.3 | 0.7 |

*Table S7.3: Combined validation performance of the extended 1978-2021 $PM_{2.5}$ model at the two-period low-cost monitor (LCM) locations for reduced LCM designs. $R^2$: coefficient of determination; RMSE: root-mean-square error. Reprinted with permission from Bi et al. 2024 (Copyright 2024 Elsevier)*

| Label | Description | Model Type | Model Time Period |
|---|---|---|---|
| adjacent | First 2 consecutive observations | SP, ST | 1978-2021 |
| first | First observation from each LCM | SP | 2010-2020 |
| full | Full or "Alldata" Model | SP, ST | 1978-2021 |
| last | Last observation from each LCM | SP | 2010-2020 |
| no | No LCMs | SP, ST | 1978-2021 |
| quadldhp | Long distance to road/high population density LCMs | SP | 2010-2020 |
| quadldlp | Long distance to road/low population density LCMs | SP | 2010-2020 |
| quadsdhp | Short distance to road/high population density LCMs | SP, ST | 1978-2021 |
| quadsdlp | Short distance to road/low population density LCMs | SP | 2010-2020 |
| random10 | Randomly select 10% of LCMs | SP | 2010-2020 |
| random25 | Randomly select 25% of LCMs | SP, ST | 1978-2021 |
| random50 | Randomly select 50% of LCMs | SP | 2010-2020 |
| random75 | Randomly select 75% of LCMs | SP | 2010-2020 |
| random90 | Randomly select 90% of LCMs | SP | 2010-2020 |
| region1 | LCMs from Region 1 | SP | 2010-2020 |
| region2 | LCMs from Region 2 | SP | 2010-2020 |
| region3 | LCMs from Region 3 | SP | 2010-2020 |
| region4 | LCMs from Region 4 | SP | 2010-2020 |
| region5 | LCMs from Region 5 | SP | 2010-2020 |
| region6 | LCMs from Region 6 | SP | 2010-2020 |
| region7 | LCMs from Region7 | SP | 2010-2020 |
| separate | First and last observation | SP, ST | 1978-2021 |

*Table S7.4: $PM_{2.5}$ Model descriptions for Figure 7.2*

## $NO_2$ Data Description

| Data Type | Locations (N) | Observations per location (N) | | Location $NO_2$ means (ppb) | | |
|---|---|---|---|---|---|---|
| | | Minimum | Maximum | Mean ± SD | Minimum | Maximum |
| Agency (in time trend) | 3 | 148 | 580 | 13.5 ± 7.3 | 5.6 | 20.1 |
| Agency (not in time trend) | 8 | 3 | 118 | 11.9 ± 7.1 | 4.9 | 21 |
| ACT-AP snapshot | 110 | 2 | 3 | 11.3 ± 3.8 | 6.4 | 26.2 |
| Low-cost sensors | 117 | 1 | 83 | 11.9 ± 3.1 | 7 | 28.6 |
| DEEDS | 100 | 2 | 2 | 17.1 ± 5.4 | 5.3 | 35.6 |
| MAP | 26 | 1 | 1 | 16.0 ± 2.4 | 11.6 | 21.4 |
| PSID | 42 | 1 | 12 | 15.4 ± 3.5 | 2.8 | 29.2 |
| Yesler Terrace | 86 | 1 | 2 | 26.0 ± 4.1 | 18.3 | 41.7 |

*Table S7.5: Summary statistics of the $NO_2$ monitors and measurements across 499 locations by monitoring type, which represent 479 unique locations. Summary statistics are calculated from two-week average concentrations over the 1996-2020 time period. Reprinted with permission from Zuidema et al. 2024 (Copyright 2024 Nature)*

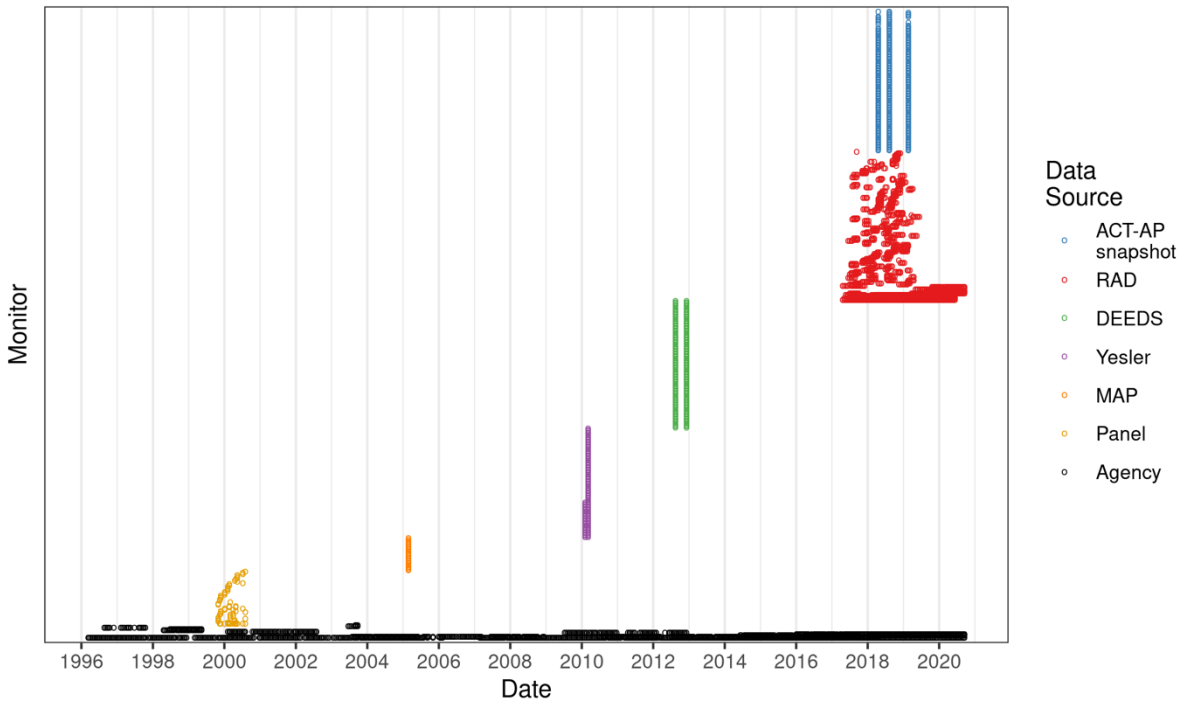



*Figure S7.3: Data availability for the $NO_2$ model. Each point represents a two-week average, with repeated observations from each location appearing in the same row. RAD refers to the low-cost sensor data, and Panel refers to PSID. Data reproduced with permission from Zuidema et al. 2024 (Copyright 2024 Nature)*

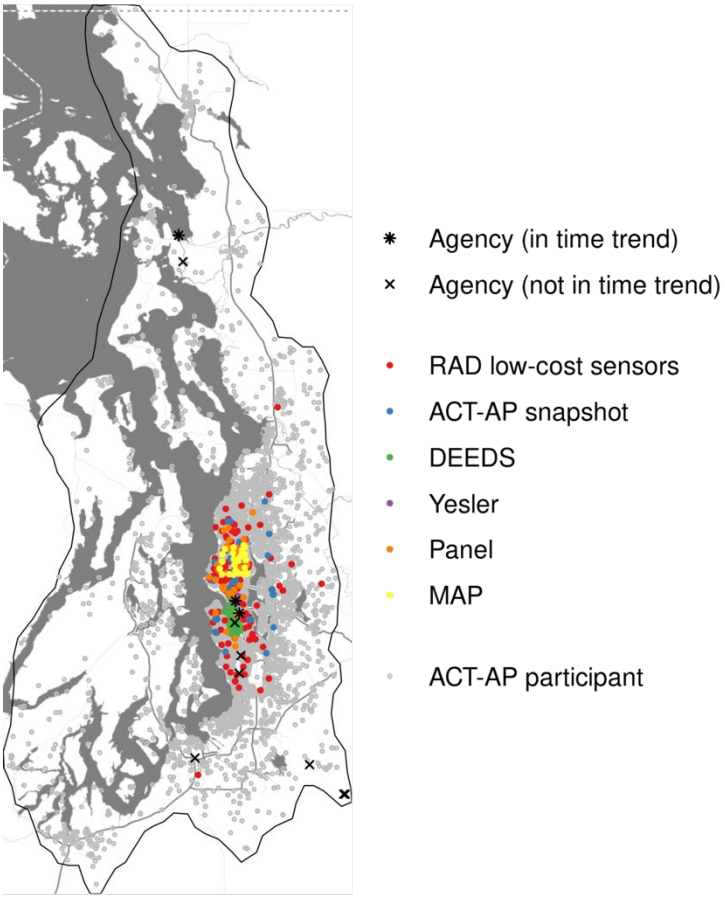



*Figure S7.4: Study region -- the Puget Sound region in Washington - with locations of $NO_2$ monitors; outline represents the modeling region. RAD refers to the low-cost sensor data, and Panel refers to PSID. Reprinted with permission from Zuidema et al. 2024 (Copyright 2024 Nature)*

# Chapter 8

## APPLICATION OF SPATIAL ENSEMBLE-LEARNING METHODS — INSIGHTS FROM VARIABLE IMPORTANCE METRICS

In this chapter supplement, we extend the presentation in Chapter 8 to discuss an example showing that the variable importance measure can provide insight into how the performance of various fitted models differs on newly observed data points. As depicted in Figure 8.2, we evaluated the UK-PLS and spatial RF models on an additional 2,815 grid locations that were not used to train the models. Figure S8.1 shows how the difference in predicted values between models varies with a selected set of predictors, namely those predictors where UK-PLS and spatial RF had the most different variable importance measures in the original training data. The analysis based on the original training data finds that spatial RF differs from UK-PLS by -4.11 units when looking at the contribution of NDVI (buffer size 5km), changing from its 25% and 50% quantiles, and differs by +41.02 units when looking at the change from 50% to 75% quantiles. Therefore, it is expected that at the higher end in the distribution of NDVI, we would observe predictions from a spatial RF model would grow faster, or decline slower, compared with UK-PLS, which is indeed reflected in the plot on the bottom right panel. Similar interpretations can be observed in the trend of evergreen forest land (buffer size 3km), highly developed land use (buffer size 5km), and truck route length (buffer size 10km, albeit less noticeable), among others. There are also signs of greater variability in the difference between predictions when the distance to A1/A3 intersections, A1/A1 intersections, or airports is large, which is consistent with the fact that these predictors were found to play different roles in the two models. Figure S8.2 presents a similar comparison, visualizing the differences in predictions at the residential locations of an epidemiological cohort throughout the greater Seattle area, as opposed to the grid locations in Figure S8.1. We observe that the difference between UK-PLS and spatial RF is less evident at the cohort locations, which are more spatially aligned with and better represented by the monitoring locations, in contrast to the gridded locations, which cover less populated areas as well. This supports the argument that models with similar behaviors on the training data (e.g., the monitoring sites) could have meaningful differences when extrapolated to new locations (e.g., the gridded locations); and with the aid of our variable importance measure, the latter can be anticipated and captured by a variable importance analysis on just the training data.

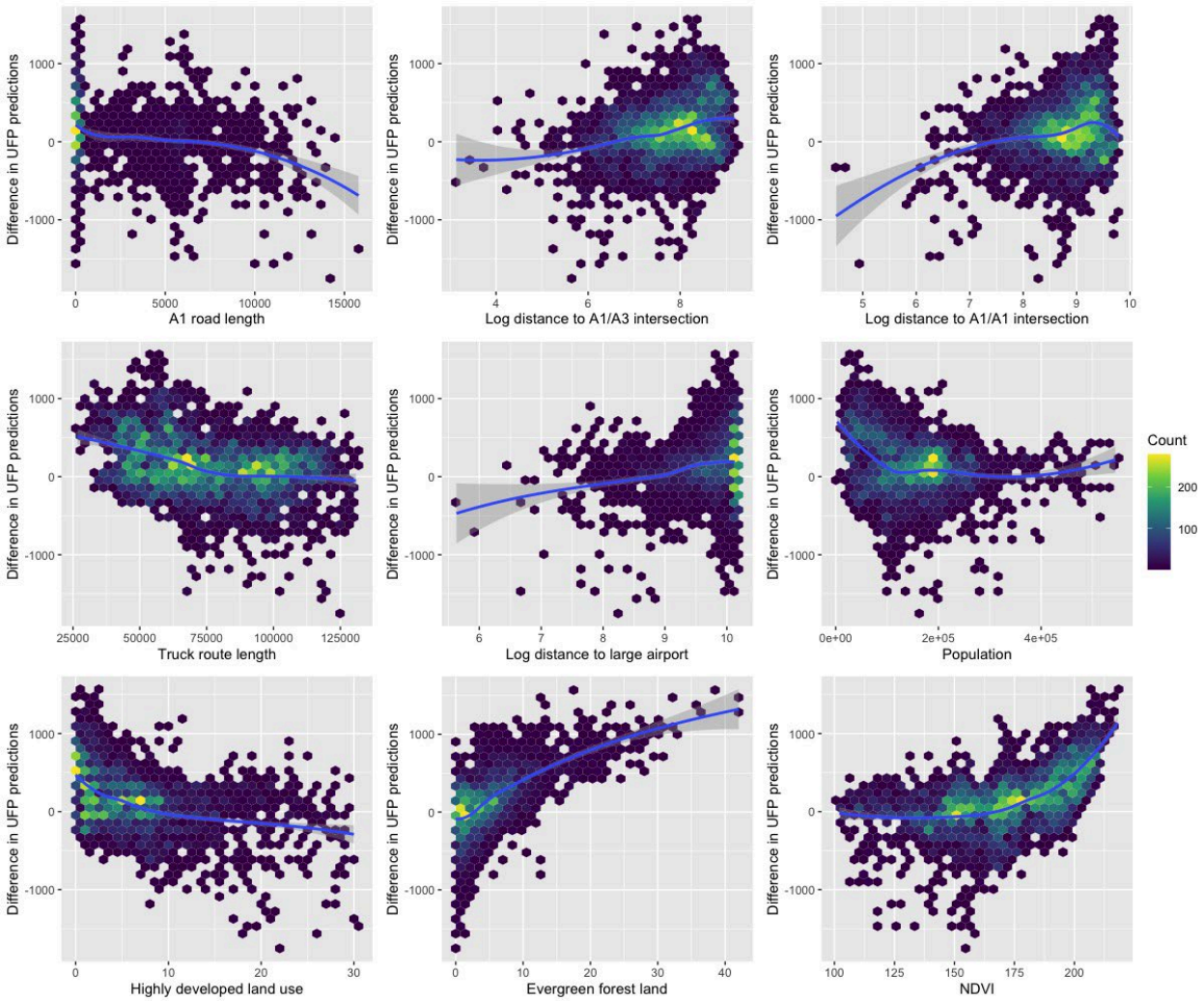



*Figure S8.1: Hexagonal bin plot showing the difference between spatial RF-PL and UK-PLS (the subtrahend) predictions of UFP concentrations at grid locations, versus the distribution of selected predictors with the greatest difference in variable importance between models. The color reflects the number of points falling into each small region of the plot. Locally weighted scatterplot smoothing (LOESS) curves have been added to show the overall trend.*

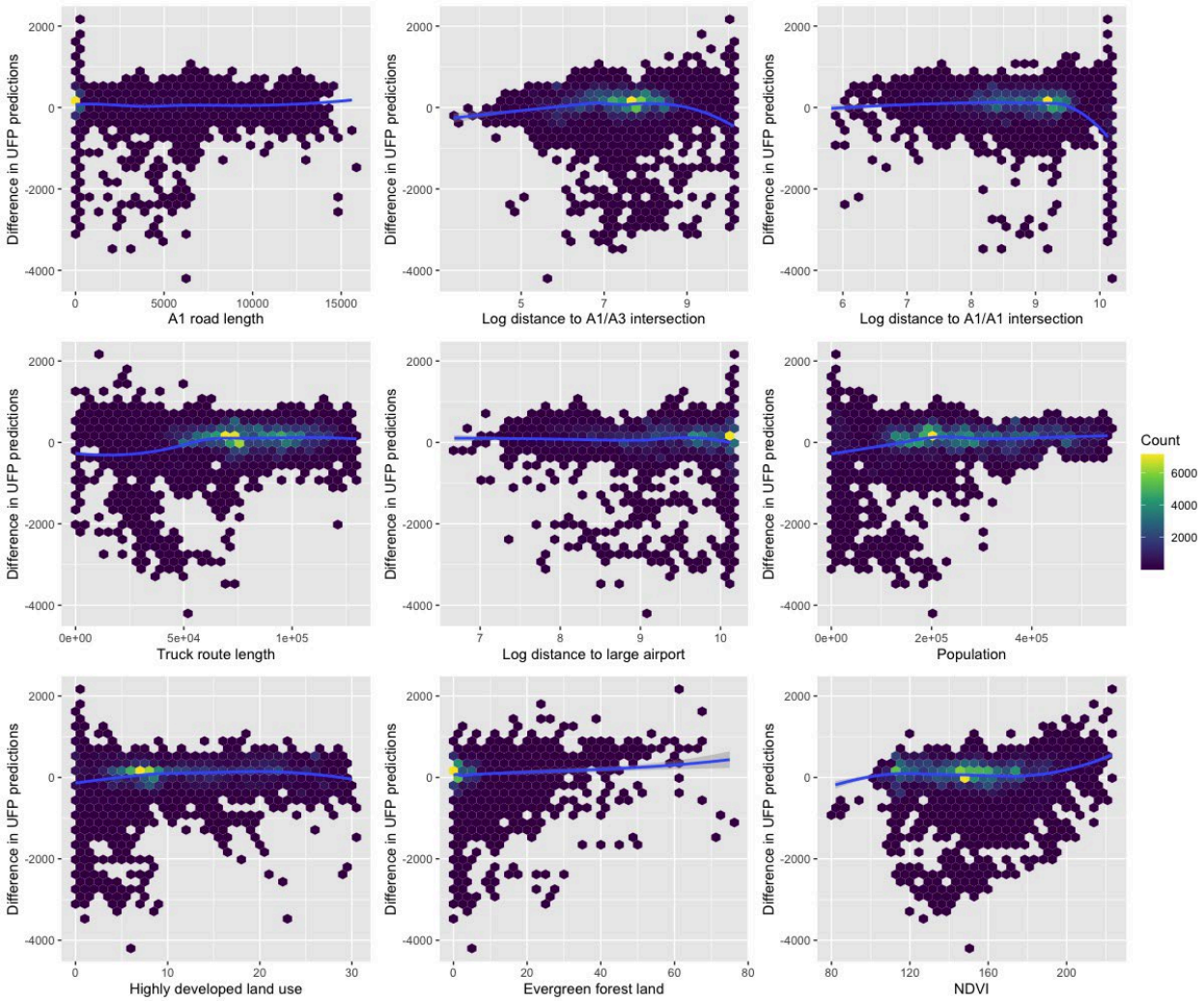



*Figure S8.2: Hexagonal bin plot showing the difference between spatial RF-PL and UK-PLS (the subtrahend predictions of UFP concentrations at ACT cohort residential locations, versus the distribution of selected predictors with the greatest difference in variable importance between models. The color reflects the number of points falling into each small region of the plot. Locally weighted scatterplot smoothing (LOESS) curves have been added to show the overall trend.*

# Chapter 9

## EXPOSURE MONITORING STUDY DESIGNS FOR EPIDEMIOLOGY: COST AND PERFORMANCE COMPARISONS

### Overview

This chapter supplement has two sections. The first section, *Supplement to: Mobile monitoring study designs for epidemiology*, provides the supplementary figures and tables referred to in that presentation. The second section, *Additional analyses: Cost and performance comparisons for low-cost monitors supplementing regulatory monitoring data*, describes our work on the value of information in supplementary low-cost monitoring campaigns. There is a summary of the write-up included in Chapter 9.

## Supplement to: Mobile Monitoring Study Designs for Epidemiology: Cost and Performance Comparisons

*Table S9.1. Characteristics and examples of two types of costs in mobile monitoring of air pollution*

| Type of cost | Item | Example | Staffing vs. material cost |
|---|---|---|---|
| Up-front | Purchase of instruments | • Purchase different instruments for UPF and gaseous pollutants | |
| | Fabrication or development | • Test/adjust the battery charging station<br>• Test/adjust air flow through the manifold<br>• Make adjustments to the manifold to get the right air flow | Relatively small materials cost, but large staffing cost |
| | Information technology or software development | • Design a data dashboard for data storage and quality control | Large cost for specialists (about 17,000 USD for developing a dashboard) |
| | Development of protocols | • Perform pilot drives and document the setup and take-down process of each instrument, including the support instruments such as GPS and hobo (temp/rh)<br>• Test and troubleshoot the airflow rate through each manifold/instrument<br>• Make sure there's enough battery power for all of the instruments for the duration of a long day of driving | Mostly staffing cost, including 2-3 person-months |
| | Maintenance | • Insurance<br>• Equipment service | |
| | Establishment of sampling locations | • Pilot sampling<br>• Protocol development and establishment of driving directions and routes, including taking photos for each location | Mostly staffing cost, including 20-30 minutes or more per location |
| Per-drive-day | Driving and quality control | • Regular and rescheduled driving<br>• Confirmation of data quality during daily sampling | Mostly staffing costs |
| | Sampling travel | • Vehicle use[a]<br>• Parking<br>• Gas | Larger material costs for a route with a lot of driving between stops |
| | Data management | • Data download from each equipment after a day of sampling<br>• Data cleaning (review error codes from instruments, remove data associated with instrument errors or severe deviations from sampling protocol, incorporate calibration data, review GPS data for positional problems) | Mostly staffing costs |
| | Others | • Equipment calibration with zero/filtered air or known gas concentration<br>• Perform additional co-location sampling of an instrument | Regular cost at a longer time scale than a day |

a. Because we were driving most days of the year, we were able to rent one of the University's vehicles at a long-term rate. This was $382/month plus about $0.30 per mile for gas and vehicle maintenance, plus we paid $120/month for parking in our building's garage. We averaged about 120 miles/day on data collection days, or about $36 of mileage-related expenses. For all but the most minimal designs, a long-term rental would be comparable or more cost-effective than renting a vehicle by the day, and significantly more convenient since a long-term rental permitted us to keep the instrument-stabilizing platform in the car at all times.

*Table S9.2. Model performance and estimated costs across different alternative sampling designs based on the data from the Adult Changes in Thought Traffic-Related Air Pollution (ACT-TRAP) mobile monitoring campaign and our assumption of a single UFP instrument (P-Trak) and a single study area*

| Design | | Design components | | | | | Cost | | | | | | | | Model performance |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | Work days for per-drive-day sampling | | | | Effort | Monetary value ($1,000)[a] | | | |
| | | Version of alternative | N of sites (per 100 km²) | N of visits per site[b] | Routes[c] | Seasons | Regular drive | Re-scheduled drive[d] | QC | Total | Staff effort[e] | Up-front[f] | Per-drive day[g] | Total | CV $R^2$ [h] |
| Complete | All[i] | | 309 (26) | 29 | 7 | 4 | 203 | 8 | 21 | 232 | 1.0 | $61 | $158 | $219 | 0.77 |
| Alternative | Fewer Sites | 250 | 250 (21) | 29 | 6 | 4 | 174 | 8 | 18 | 200 | 0.9 | $60 | $145 | $205 | 0.76 (0.74,0.78) |
| | | 200 | 200 (17) | 29 | 5 | 4 | 145 | 8 | 15 | 168 | 0.7 | $59 | $131 | $190 | 0.74 (0.70,0.76) |
| | | 150 | 150 (12) | 29 | 3 | 4 | 87 | 8 | 10 | 105 | 0.5 | $58 | $103 | $162 | 0.70 (0.67,0.75) |
| | | 100 | 100 (8) | 29 | 2 | 4 | 58 | 8 | 7 | 73 | 0.3 | $58 | $88 | $146 | 0.67 (0.59,0.73) |
| | Fewer Visits | 24 | 309 | 24 | 7 | 4 | 168 | 8 | 18 | 194 | 0.8 | $61 | $142 | $203 | 0.72(0.70,0.75) |
| | | 20 | 309 | 20 | 7 | 4 | 140 | 8 | 15 | 163 | 0.7 | $61 | $129 | $190 | 0.72 (0.67,0.74) |
| | | 16 | 309 | 16 | 7 | 4 | 112 | 8 | 12 | 132 | 0.6 | $61 | $116 | $177 | 0.71 (0.68,0.74) |
| | | 12 | 309 | 12 | 7 | 4 | 84 | 8 | 9 | 101 | 0.4 | $61 | $101 | $162 | 0.70 (0.65,0.72) |
| | | 6 | 309 | 6 | 7 | 4 | 42 | 8 | 5 | 55 | 0.2 | $61 | $71 | $132 | 0.63 (0.58,0.68) |
| | | 4 | 309 | 4 | 7 | 4 | 28 | 8 | 4 | 40 | 0.2 | $61 | $63 | $124 | 0.60 (0.52,0.67) |
| | Fewer Days[j] | Weekday | 309 | 12 | 7 | 4 | 84 | 8 | 9 | 101 | 0.4 | $61 | $101 | $162 | 0.65 (0.60,0.68) |
| | Fewer Hours[k] | Business hour | 309 | 12 | 7 | 4 | 84 | 8 | 9 | 101 | 0.4 | $61 | $101 | $162 | 0.55 (0.52,0.59) |
| | | Rush hour | 309 | 12 | 7 | 4 | 84 | 8 | 9 | 101 | 0.4 | $61 | $101 | $162 | 0.64 (0.60,0.68) |
| | Fewer Seasons[l] | 3 | 309 | 12 | 7 | 3 | 84 | 6 | 9 | 99 | 0.4 | $61 | $98 | $158 | 0.69 (0.64,0.72) |
| | | 2 | 309 | 12 | 7 | 2 | 84 | 4 | 9 | 97 | 0.4 | $61 | $94 | $155 | 0.67 (0.62,0.70) |

a. Based on the budget of the mobile monitoring campaign of the ACT-TRAP study in 2019-2000, multiplied by the inflation rate for 2020-2024
b. Maximum number of visits feasible given the restriction of reduced designs in a single mobile platform
c. A single route as a packet of driving covering an average of 35 sites, which provides the number of driving days
d. Make-up driving for at least 2 days per season
e. The number of technicians for the total required work days for drive sampling relative to 229 available working days
f. Up-front cost for Initial equipment, including P-TRAK, pilot-phase vehicle, pilot-phase route testing and documentation, and IT/data management investment
g. Staff cost for per-drive-day sampling plus vehicle cost
h. Median cross-validation (CV) $R^2$ and 5th and 95th percentiles across campaigns (N=30) for each design using predictions of UFP annual averages at all 309 sites; for fewer sites designs, predictions were obtained from CV for 100-250 sites and external validation for the rest of the sites not included in a given campaign.
i. 309 sites, 29 visits per site, all days, most hours for 5 a.m. – 11 p.m., and four seasons
j. We did not consider the designs of weekend-only and 1 season, which are not realistic.
k. Business hours: 9 a.m. – 5 p.m.; Rush Hours: 7-10 a.m. & 3-6 p.m.
l. Specific combination such as summer-winter and summer-spring

*Table S9.3. Estimated costs in monetary values in 1,000 US dollars across different alternative sampling designs based on the data from the Adult Changes in Thought Traffic-Related Air Pollution (ACT-TRAP) mobile monitoring campaign and monitoring scenarios by the number of instruments and the staff cost according to sampling day and time*

| Design | | Monetary values in 1000 US dollars [a] | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Scenario: | Single instrument [b] | | | Temporally-varying shift premium [c] | | | | | | | | | Multiple instruments [d] | | | |
| | | | | | Evening/weekend pay multiplier=1.3 | | | 1.5 | | | 2.0 | | | | | | |
| | | Version of alternative designs [e] | Up-front [f] | Per-drive day [g] | Total | Up-front | Per-drive day | Total | Up-front | Per-drive day | Total | Up-front | Per-drive day | Total | Up-front | Per-drive day | Total |
| Complete | All [h] | | $61 | $158 | $219 | $61 | $172 | $232 | $61 | $181 | $242 | $61 | $203 | $264 | $212 | $186 | $398 |
| Alternative | Fewer Sites | 250 | $60 | $145 | $205 | $60 | $156 | $216 | $60 | $164 | $224 | $60 | $183 | $244 | $211 | $159 | $370 |
| | | 200 | $59 | $131 | $190 | $59 | $141 | $200 | $59 | $147 | $207 | $59 | $164 | $223 | $211 | $145 | $355 |
| | | 150 | $58 | $103 | $162 | $58 | $109 | $168 | $58 | $113 | $172 | $58 | $124 | $182 | $209 | $118 | $327 |
| | | 100 | $58 | $88 | $146 | $58 | $93 | $150 | $58 | $95 | $153 | $58 | $103 | $160 | $209 | $103 | $312 |
| | Fewer Visits | 24 | $61 | $142 | $203 | $61 | $153 | $214 | $61 | $161 | $222 | $61 | $180 | $241 | $212 | $164 | $377 |
| | | 20 | $61 | $129 | $190 | $61 | $138 | $199 | $61 | $145 | $206 | $61 | $161 | $221 | $212 | $148 | $360 |
| | | 16 | $61 | $116 | $177 | $61 | $123 | $184 | $61 | $129 | $189 | $61 | $141 | $202 | $212 | $131 | $343 |
| | | 12 | $61 | $101 | $162 | $61 | $107 | $168 | $61 | $111 | $172 | $61 | $121 | $182 | $212 | $114 | $327 |
| | | 6 | $61 | $71 | $132 | $61 | $74 | $135 | $61 | $77 | $137 | $61 | $82 | $143 | $212 | $77 | $290 |
| | | 4 | $61 | $63 | $124 | $61 | $66 | $126 | $61 | $67 | $128 | $61 | $71 | $132 | $212 | $67 | $279 |
| | Fewer Days [i] | Weekday | $61 | $101 | $162 | $61 | $104 | $165 | $61 | $106 | $167 | $61 | $110 | $171 | $212 | $114 | $327 |
| | Fewer Hours [j] | Business hour | $61 | $101 | $162 | $61 | $105 | $166 | $61 | $107 | $168 | $61 | $112 | $173 | $212 | $114 | $327 |
| | | Rush hour | $61 | $101 | $162 | $61 | $105 | $166 | $61 | $107 | $168 | $61 | $112 | $173 | $212 | $114 | $327 |
| | Fewer Seasons [k] | 3 | $61 | $98 | $158 | $61 | $103 | $164 | $61 | $107 | $168 | $61 | $117 | $178 | $212 | $109 | $321 |
| | | 2 | $61 | $94 | $155 | $61 | $99 | $160 | $61 | $103 | $164 | $61 | $113 | $173 | $212 | $104 | $316 |

a. Based on the budget of the mobile monitoring campaign of the ACT-TRAP study in 2019-2000, multiplied by the average inflation rate of 1.2 for 2020-2024
b. P-Trak for PNC
c. 30, 50, and 100% higher staff cost rate for weekend and evening driving than that for weekdays and daytime driving
d. P-Trak and two additional instruments for PNC (NanoScan and DiSCmini) and four instruments for black carbon, nitrogen dioxide (NO2), carbon dioxide, and carbon monoxide (microAeth MA200, CAPS NO2, LI-850, and Langan, respectively); While additional PNC instruments are useful to verify PNC measurements, other pollutants can help adjust localized on-road plumes to assess representative residential exposure to UFP, as demonstrated in the ACT-TRAP study
e. All the other design components are identical to the complete design except for the application of 12 sites to fewer days, fewer hours, and fewer seasons designs
f. Up-front cost for Initial equipment, including P-TRAK, pilot-phase vehicle, pilot-phase route testing and documentation, and IT/data management investment
g. Staff cost for per-drive-day sampling plus vehicle cost
h. 309 sites, 29 visits per site, all days, most hours for 5 a.m. – 11 p.m., and four seasons
i. We did not consider the designs of weekend-only and 1 season, which are not realistic.
j. Business hours: 9 a.m. – 5 p.m.; Rush Hours: 7-10 a.m. & 3-6 p.m.
k. Specific combination such as summer-winter and summer-spring

*Table S9.4. Present and previous studies that investigated optimal mobile monitoring designs in vehicle platforms with the ultimate goal of epidemiological application*

| Study | Area | Period | Site selection | Temporal coverage | Pollutant | Prediction model | Main findings |
|---|---|---|---|---|---|---|---|
| Blanco 2023 [a] | Seattle, US (1200 $km^2$) | All days in March 2019-March 2020 | Residence | 5 a.m.–11 p.m./ weekdays and weekends/ 4 seasons | PNC, BC, $NO_2$, $PM_{2.5}$, $CO_2$, | Universal kriging | Model improvement for PNC with more sites and more repeat visits (4-8 visits, CV $R^2 > 0.6$) and less bias for all-day and most hours monitoring |
| Messier 2018 [b] | Oakland, US (30 $km^2$) | 327 days for May 2015-May 2017 | Land use | 9 a.m.–5 p.m./ weekdays/ 4 seasons | BC, NO | Universal kriging | Model improvement for NO with more repeat visits (4 visits, CV $R^2$=0.5) than larger road coverage |
| Kerckhoffs 2024 [c] | | | | | BC | Mixed model (random effect by road segment) | 5-15 repeat visits with CV $R^2$ of 0.8 and 0.9 |
| Hatzopoulou 2017 [d] | Montreal, Canada (470 $km^2$) | 34 days in 2009 | Land use | 9 a.m.–8 p.m./ 3 seasons | PNC, BC | Land use regression | Model improvement for PNC with more repeat visits (16 visits for CV $R^2$=0.70-0.74) than more road segments |

a. Monitoring design and data used in the present study
b. Investigation of the model performance using the updated data using the same monitoring design as Apte et al. 2017 which investigated optimal monitoring designs focusing on sampling frequency and spatial coverage based on the data only without applying prediction models
c. The same dataset as that of Messier et al. 2018, but with application of a new modeling approach
d. Investigation of designs using the data collected in a previous study (Levy et al. 2014) without new data collection

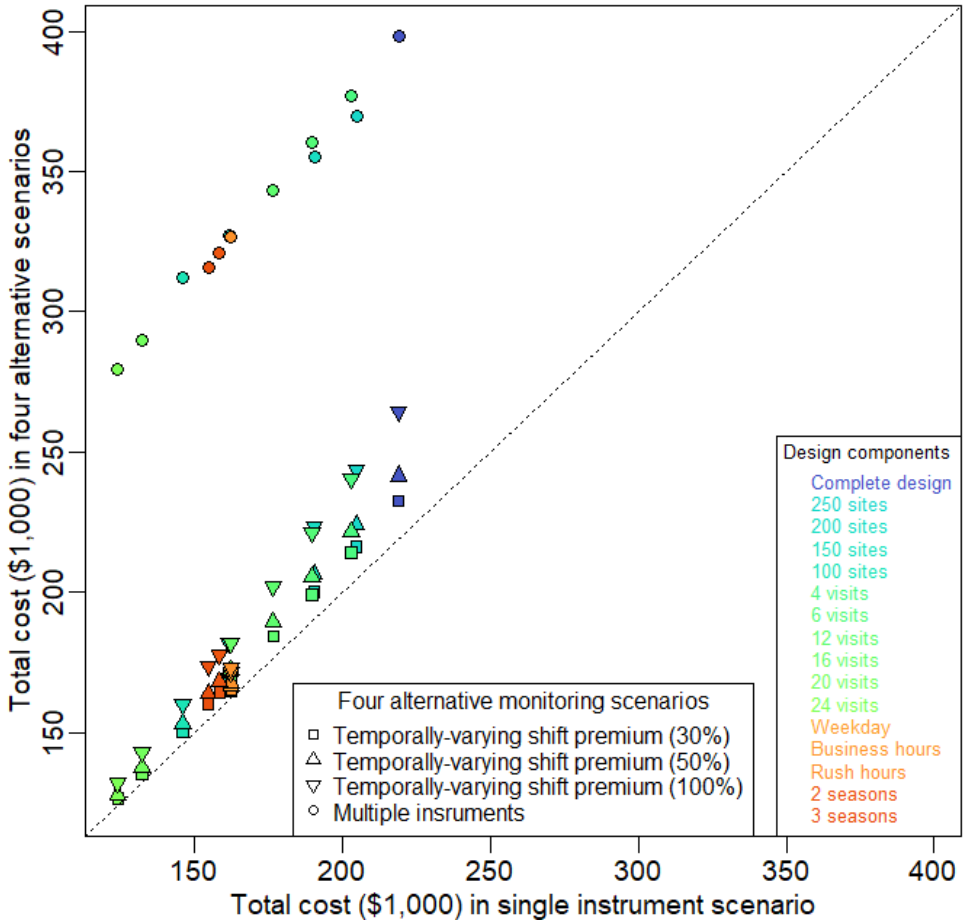



*Figure S9.1. Scatter plot of estimated total costs (up-front and per-drive-day costs) in 1,000 US dollars across the complete all-data and alternative designs by five design components (fewer sites, fewer visits, weekday only, business or rush hours only, and fewer seasons) between the single-instrument monitoring campaign scenario and alternative monitoring campaign scenarios that add 1) temporally-varying shift premium with 30, 50, and 100 % higher rate for evening and weekend driving or 2) multiple instruments*

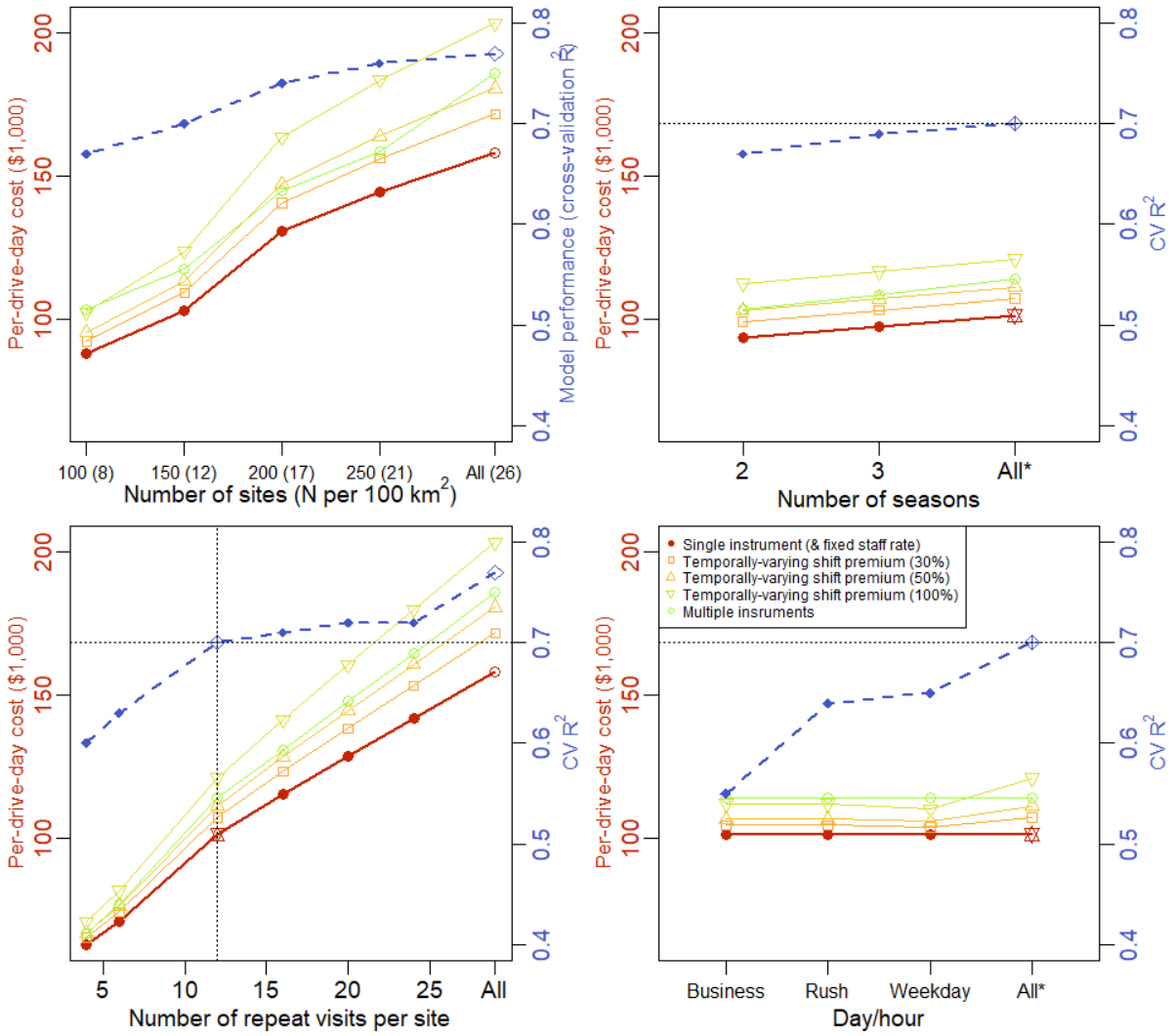



*Figure S9.2. Relationships between estimated per-drive-day costs in 1,000 US dollars and median cross-validated (CV) $R^2$s across complete and alternative mobile monitoring designs according to the number of sites, visits, seasons, days, and hours ("All" and "All*" indicating all sites, seasons, days or hours from the complete all-data and reduced (12 sites) reference designs, respectively) as well as different monitoring campaign scenarios according to the number of instruments or shift premium for evening and weekend driving*

## Additional Analyses: Cost and Performance Comparisons for Low-Cost Monitors Supplementing Regulatory Monitoring Data

Lead authors: Amanda Gassett, Jianzhao Bi, Magali Blanco, Lianne Sheppard

*Introduction*

Air pollution assessment campaigns are typically planned with certain hard constraints: a total budget, a deadline by which the data collection must be complete, and a geographic domain. Therefore, identifying the most efficient monitoring strategy - the strategy that produces the best predictions at the locations of interest while still being affordable – is critical for designing air monitoring studies. Thus, it is of interest to balance these logistical constraints with scientific principles that address key monitoring design considerations for estimating accurate long-term average exposures.

Criteria pollutants have a long history of regulation in the U.S., and Air Quality Systems (AQS) agencies are charged by the EPA to meet certain requirements for regular monitoring. Therefore, all urban communities and some rural communities in the U.S. have at least one local monitor observing concentrations for criteria pollutants such as $PM_{2.5}$ and $NO_2$. While a single monitor is a great resource for determining the trend in concentrations over time, it is not sufficient to characterize the spatial patterns that are necessary to study for informative inference in epidemiologic cohorts. Furthermore, even in areas with multiple regulatory monitors, those monitors typically are not sited to specifically represent the residential locations for a given study population. Cohort-informed exposure assessment campaigns are designed to provide the spatial coverage needed to improve air pollution predictions at the locations where participants live, and typically, a measurement method that is compatible with regulatory methods is chosen so that the temporal information from the AQS monitor(s) can be leveraged in exposure prediction models. Fortunately, a variety of affordable measurement methods exist for these pollutants, and some are suitable for simultaneous exposure measurement at numerous sites. In recent years, low-cost $PM_{2.5}$ sensors have become available to consumers and researchers. Further, $NO_2$ concentrations can be measured using the Ogawa method, which is appealing from a logistical standpoint because this method doesn't require a power source.

In terms of the geographic domain, epidemiological cohorts in the U.S. fall into one of three categories: single-center, multi-center, or national. For single and multi-center cohorts, it's important to characterize within-community spatial variability very well, since there is a lower amount of air pollution concentration variability compared to a national cohort. However, single-center and multi-center projects may have very different logistical constraints when it comes to collecting exposure data. Unless a multi-center study is particularly well-funded, exposure data collection is restricted to brief excursions by the field team, whereas a single-center study may be able to employ dedicated staff to collect data regularly throughout the monitoring period. For the purposes of this project, we will not consider national cohorts further. We will also focus exclusively on the greater Seattle area. However, in considering logistically feasible exposure assessment designs, our work will be informed by logistical constraints imposed by multi-center cohorts.

*Study Design and Methods*

*Brief Data Deployment Description*

Chapter 3 describes the monitoring data in the section "Low-cost monitoring and related regulatory monitoring data"; however, it is important to note that while the analysis presented in this chapter supplement is related to and informed by the analyses presented in Chapter 7, the comparative results presented here are based on different subsets of the data.

Briefly, LCMs were intended to be deployed at ACT participant and community volunteer homes for 8-week periods in two seasons. LCMs included sensors for both $NO_2$ and $PM_{2.5}$. Some participants chose not to host a monitor for a second round, and some participants chose to host a monitor for a single extended period of time. The calibration of these monitors has been described (Zuidema et al., 2021; Zusman et al., 2020).

Additionally, we collected $NO_2$ "snapshot" data by simultaneously deploying passive Ogawa samplers at 110 utility pole locations for 2 weeks in each of three seasons. Most of these locations were arranged as 17 clusters around major roads, most of which consisted of 6 monitors varying from < 50 m up to 350 m. There were also 9 isolated locations, sited on utility poles, monitored during previous research campaigns.

The datasets are summarized in Table S7.1 for $PM_{2.5}$ and Table S7.5 for $NO_2$. The $PM_{2.5}$ data in this analysis were restricted to the 2010-2020 time period, as was done for most of the $PM_{2.5}$ analyses reported in Chapter 7.

*DESIGNS CONSIDERED FOR $NO_2$ VALUE OF INFORMATION ANALYSES*

In the following subsections, we describe the designs we considered in the value of information (VOI) analyses that follow. The $NO_2$ designs are summarized in Table S9.5.

DESIGNS FOCUSED ON THE VALUE OF SNAPSHOT MONITORING

For these models, we excluded the LCM home monitoring and fit models based on varied subsets of snapshot monitoring. This included: All snapshot monitoring, 1 or 2 seasons at all locations, all seasons but reducing the number of locations in each cluster, and all seasons but reducing the number of locations at random.

The following describes the specific approach to sampling subsets of snapshot locations summarized in Table S9.5:

- 1/3 Snapshot locations; Stratified: The intention was to include one nearest-road and one farthest from road location per cluster. When limiting monitors to those with 3 $NO_2$ measurements, we end up with two groups of clusters. Those with 5+ qualifying monitors (14) and those that have less than 5 (3). For the clusters that have 5+ monitors, sample one monitor near the intersection and 1 monitor far from the intersection (2 total per cluster). For clusters that have fewer than 5 monitors, sample 1 monitor at random (1 total). For each model, this results in: 28 monitors (5+ monitor clusters) + 3 monitors (<5 monitor clusters) + 6 singletons (6 out of the 9 have 3 $NO_2$ measurements) + 3 AQS = 40 total monitors.
- 2/3 Snapshot locations; Stratified: The intention was to include at least one nearest-road and one farthest from road location per cluster, plus 2 at any distance. When limiting monitors to those with 3 $NO_2$ measurements, we end up with two groups of clusters. Those with 5+ qualifying monitors (14) and those that have less than 5 (3). For the clusters that have 5+ monitors, sample one monitor near the intersection, 1 monitor far from the intersection, and 2 of the remaining monitors at random (4 total per cluster). For clusters that have fewer than 5 monitors, sample 1 monitor at random (1 total). For each model, this results in: 56 monitors (5+ monitor clusters) + 3 monitors (<5 monitor clusters) + 6 singletons (6 out of the 9 have 3 $NO_2$ measurements) + 3 AQS = 68 total monitors.
- 1/3 Snapshot Locations; Random: The goal was to select one location per cluster plus the singletons. We chose at random from 97 snapshot locations, and matched the number of sites to 2/cluster. This results in: 28 monitors (5+ monitor clusters) + 3 monitors (<5 monitor clusters) + 6 singletons (6 out of the 9 have 3 $NO_2$ measurements) + 3 AQS = 40 total monitors.

### DESIGNS FOCUSED ON THE VALUE OF HOME MONITORING WITH LOW-COST MONITORS

For these models, we excluded the snapshot monitoring and fit models using varied subsets of the LCM home monitors. We sampled home monitoring observations to match the data support provided by the snapshot variation. Repeat samples at each home location were chosen to provide the maximum separation between observations. That is, for the 2-observation model, we selected the first and last observation. For the 3-observation model, we added the observation that was closest to the midpoint between the first and last observation.

*Table S9.5: Distribution of observations included in each $NO_2$ model.*

| Model | N Home Locations | N Snapshot Locations | N Seasons or Periods | N Models |
|---|---|---|---|---|
| Gold Standard – All data | 117 | 110 | All | 1 |
| 1 Snapshot | 0 | 97* | 1 | 3 |
| 2 Snapshot | 0 | 97* | 2 | 3 |
| All Snapshot | 0 | 97* | 3 | 1 |
| 1/3 Snapshot Locations; Stratified | 0 | 37 | 3 | 3 |
| 2/3 Snapshot Locations; Stratified | 0 | 65 | 3 | 3 |
| 1/3 Snapshot Locations; Random | 0 | 37 | 3 | 3 |
| 1 LCM Observation per Home | 97† | 0 | 1 | 3 |
| 2 LCM Observations per Home | 65 | 0 | 2 | 3 |
| 3 LCM Observations per Home | 37 | 0 | 3 | 1 |

* 13 samples were lost or damaged, so that some locations only had measurements for 2 seasons. We limited our analysis to the 97 locations with observations for all 3 seasons.

† To match the number of home locations to the number of snapshot locations, we limited our analysis to the 97 locations with the most available data.

### *DESIGNS CONSIDERED FOR $PM_{2.5}$ VALUE OF INFORMATION ANALYSES*

We did not have the equivalent of snapshot locations for $PM_{2.5}$ monitoring, so we investigated different kinds of monitoring designs, focusing on questions about spatial compatibility and temporal extent. This included randomly selecting monitoring subsets of 10% - 90% of the number used in the "all data" model, subsetting by land use (distance to road and population density), and restricting monitors to single monitoring regions. The temporally reduced models included models based either on one (first or last) observation or two (first and last or adjacent) observations. There was also a model that was based only on 78 reference-grade monitors and excluded all home monitoring data. These models were discussed in more depth in Chapter 7 and in Bi et al.(Bi et al., 2024).

### *DATA ANALYSIS – SPATIO-TEMPORAL MODELING AND PERFORMANCE STATISTICS*

All models were developed using the SpatioTemporal package.

For $NO_2$, our "gold standard" or "all-data" model for this analysis is the primary model developed by Zuidema et. al. (Zuidema et al., 2024) and generally described in the Chapter 3 section "Spatio-temporal Exposure Modeling Approach and Model Selection". $NO_2$ models were fit using the same number of time trends, the same number of PLS components, and the same residual structure as described in Chapter 7 and its source paper (Zuidema et al., 2024). Additional models were developed for this analysis primarily to disentangle the effect of the amount of monitoring from the type of monitoring. $NO_2$ models cover the time period 1996 – 2020.

$PM_{2.5}$ models were fit as described in Chapter 7 and its source paper (Bi et al., 2024). The models presented in this analysis covered the time period 2010 – 2020. No additional models were fit for the

analysis presented in this chapter supplement; rather, the results from Chapter 7 are presented and discussed here in the VOI context.

Tables S9.6 and S9.7 present the spatial or long-term average version of the MSE-based $R^2$ and RMSE (Keller et al., 2015), as also presented in Chapter 7. In other words, we focus on the whole modeling period (WMP or *spatial validation*) timescale in this chapter supplement. See Chapter 3 for additional description of the statistics and approach. The prediction and observation sets were each averaged prior to calculating long-term average metrics, to minimize the extent to which the metrics leverage temporal variability. The metrics reported for each specific model (i.e., the models listed in the rows of 6Tables S5 and S7) are based on a mix of cross-validation and external validation, which was referred to as "combined validation" in Chapter 7. 10-fold cross validation was performed for the $PM_{2.5}$ models, whereas 9-fold cross-validation was performed for $NO_2$ to divide the number of snapshot clusters evenly. Snapshot locations that were part of the same cluster were assigned to the same cross-validation group, to prevent kriging from monitors that are too nearby from causing overly optimistic predictions. Other monitors were also assigned to cross-validation groups so they would be evenly represented across the CV groups. Monitoring locations that did not contribute observations to the specific model under consideration are externally validated based on the corresponding prediction from the specific model that used all of the training data. We summarize these metrics separately for different types of monitors, focusing on the home locations (both $PM_{2.5}$ and $NO_2$) and snapshot sites ($NO_2$ only).

*Model Performance Results*

*$NO_2$ RESULTS*

Table S9.6 gives the cross-validation/external validation model performance results for the $NO_2$ models. The first section shows the results of sampling subsets of snapshot data, without any home sampling data. The second section shows subsets of home data without any snapshot data. Designs that included snapshot data and excluded home data performed poorly at homes by the MSE-based metrics. This was driven more by bias than poor correlation, as regression $R^2$ at the homes ranged from 0.43 - 0.51 for versions that included data from all 3 seasons. By this metric, these models had similar performance at home sites as models that used home site data.

*Table S9.6: Combined-validation results for $NO_2$ models. Ranges are provided for models with multiple iterations. Bolded results indicate reasonably good performance when compared to the gold standard. RMSE is in ppb. "Stratified samples" mean that both near-road and farthest-from-road locations within a cluster were included.*

| Model | MSE $R^2$: Homes | Reg $R^2$: Homes | MSE $R^2$: Snapshot Sites | RMSE (ppb): Homes | RMSE (ppb): Snapshot Sites |
|---|---|---|---|---|---|
| Gold Standard – All data | **0.51** | † | **0.61** | **2.80** | **2.70** |
| *Snapshot Focused Designs* | | | | | |
| 1 Snapshot – Spring | 0.00 | 0.31 | 0.49 | 4.22 | **2.70** |
| 1 Snapshot – Summer | 0.00 | 0.26 | 0.46 | 4.09 | **2.79** |
| 1 Snapshot – Winter | 0.00 | 0.09 | 0.36 | 3.42 | **3.04** |
| 2 Snapshot – Spring/Summer | 0.00 | 0.47 | **0.55** | 4.23 | **2.53** |
| 2 Snapshot – Spring/Winter | 0.00 | 0.20 | **0.54** | 3.19 | **2.56** |
| 2 Snapshot – Summer/Winter | 0.00 | 0.14 | 0.00* | 7.05* | 4.22** |
| 3 Snapshot | 0.03 | 0.50 | **0.56** | **3.03** | **2.52** |
| *1/3 Snapshot Locations; Stratified* | | | | | |
| Stratified Sample 1 | 0.04 | 0.50 | **0.56** | **3.01** | **2.52** |
| Stratified Sample 2 | 0.13 | 0.48 | **0.56** | **2.86** | **2.62** |
| Stratified Sample 3 | 0.18 | 0.51 | **0.52** | **2.79** | **2.69** |
| *2/3 Snapshot Locations* | | | | | |
| Stratified Sample 4 | 0.06 | 0.43 | **0.54** | **2.98** | **2.61** |
| Stratified Sample 5 | 0.00 | 0.46 | **0.57** | **3.18** | **2.49** |
| Stratified Sample 6 | 0.05 | 0.48 | **0.52** | **3.00** | **2.62** |
| 1/3 Snapshot Locations; Random | | | | | |
| Random Sample 1 | 0.00 | 0.47 | **0.53** | **3.26** | **2.65** |
| Random Sample 2 | 0.00 | 0.52 | 0.49 | **3.17** | **2.75** |
| Random Sample 3 | 0.05 | 0.42 | 0.41 | **2.99** | **2.92** |
| *Home Focused Designs* | | | | | |
| *1 Observation per Home* | | | | | |
| Random Observation Sample 1 | **0.54** | 0.56 | **0.66** | **2.08** | 3.23 |
| Random Observation Sample 2 | **0.50** | 0.51 | **0.64** | **2.18** | 2.90 |
| Random Observation Sample 3 | **0.52** | 0.54 | **0.53** | **2.13** | 3.31 |
| *2 Observations per Home* | | | | | |
| Random Observation Sample 4 | **0.43** | 0.45 | **0.59** | **2.31** | 3.49 |
| Random Observation Sample 5 | **0.40** | 0.44 | **0.51** | **2.37** | 3.63 |
| Random Observation Sample 6 | **0.41** | 0.48 | **0.56** | **2.37** | 3.69 |
| 3 Observations per Home | **0.42** | 0.49 | **0.55** | **2.33** | **2.55** |

* This result appears to be driven by a single highly influential point, with a predicted concentration of 70 ppb for an observed concentration of 17 ppb.
** This model also predicted about 37 ppb for a snapshot location with an observed concentration of about 7 ppb.
† Results not available.

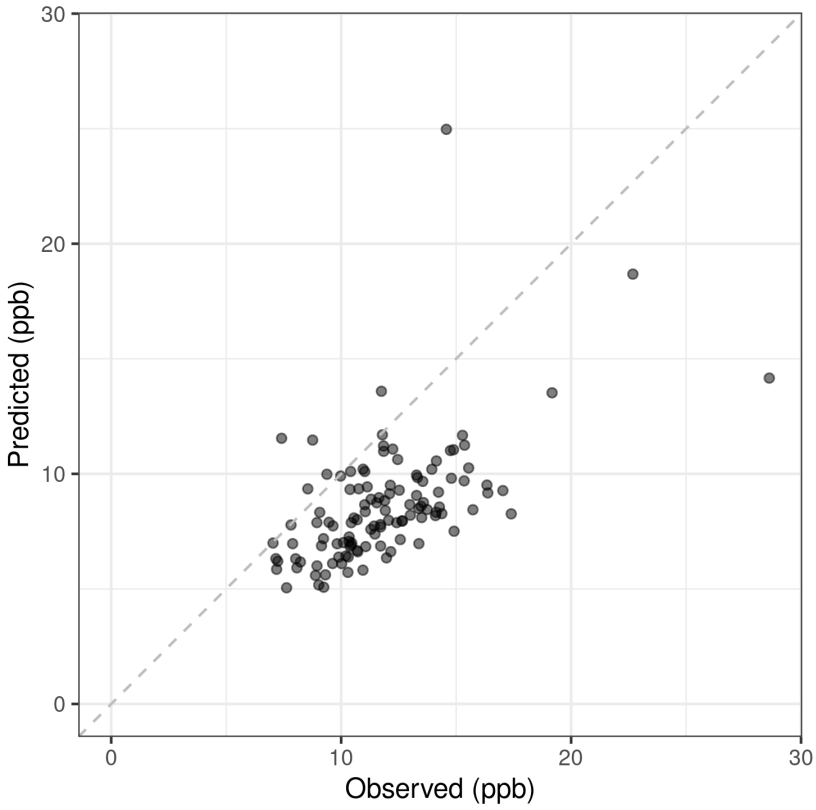



*Figure S9.3: Scatterplot of $NO_2$ predictions vs. observations at participant homes, using a model based on AQS data and a single snapshot. The dashed line is a 1-1 reference line.*

Figure S9.3 shows a scatterplot of the $NO_2$ predictions at participant homes for a model that uses all AQS data and a single snapshot campaign. Although the predictions and observations are somewhat correlated, the predictions are biased low.

*$PM_{2.5}$ RESULTS*

Model performance results from the $PM_{2.5}$ analysis are summarized in Table S9.7. Briefly, spatial MSE $R^2$'s ranged from 0.11 in the "no homes" model and one of the single-region models up to 0.57 in the "all data" model. Among the temporally reduced designs, spatial validation performance decreased moderately (spatial $R^2$ 0.37 – 0.48). Among the spatially reduced designs, the impact on the spatial $R^2$ was modest as long as at least 75% of the home locations were retained. In contrast, selecting home locations based on region or land use had a large to moderate impact moderately (spatial $R^2$ 0.11 – 0.40)

*Table S9.7. Selected combined-validation results from $PM_{2.5}$ analyses. Ranges are provided for rows summarizing multiple models; full results for each individual model are available in Table 7.2. Bolded results indicate reasonably good performance compared to the gold standard. RMSE is in $\mu g/m^3$.*

| Model | N Home Locations (N Observations) | MSE $R^2$: Homes | RMSE: Homes |
|---|---|---|---|
| All Data (Gold Standard) | 82 (502) | **0.57** | **0.6** |
| No Home Data | 0 (0) | 0.11 | 0.8 |
| *Temporally Reduced Designs* | | | |
| 1 Observation per Home | 82 (82) | 0.37 – 0.43 | 0.7 |
| 2 Observations per Home ("Separate" and "Adjacent" models) | 82 (164) | 0.41 – **0.48** | **0.6** – 0.7 |
| *Spatially Reduced Designs* | | | |
| 10% Locations | 9 (61) | 0.28 | 0.7 |
| 25% Locations | 21 (131) | 0.39 | 0.7 |
| 50% Locations | 41 (250) | 0.45 | 0.7 |
| 75% locations | 62 (382) | **0.5** | **0.6** |
| 90% Locations | 74 (458) | **0.57** | **0.6** |
| Single Region (7 regional models) | 8 – 14 (56 – 88) | 0.11 – 0.40 | 0.7 – 0.8 |
| Stratified by Land Use (4 models) | 15 – 26 (92 – 161) | 0.19 – 0.40 | 0.7 – 0.8 |

*Discussion and Staff Effort Estimates*

Results from both the $PM_{2.5}$ and $NO_2$ analyses clearly indicate that home monitoring adds value to models developed for epidemiological purposes. However, this type of monitoring is also very expensive. Home monitoring requires more staff time and potentially involves different staff positions, since participants need to be contacted for recruitment, to complete a consent process, and need to be present to provide access to their outdoor spaces for monitoring set-up and take-down. Home monitoring data is also subject to IRB oversight and cannot be publicly shared, since that would violate participants' privacy. In previous studies, we have found that a team of 2 field technicians could reasonably plan to visit about 5 homes over the course of a workday. In contrast, using a snapshot design and passive samplers, a 2-person team can typically visit 40 – 60 utility pole locations in a day, making it possible to collect data from more locations much more quickly. In Table S9.8, we show the approximate number of staff work days required just to collect each dataset. Datasets that included home monitoring required at least three times the amount of staff days as the most time-intensive snapshot. However, while the snapshot-based models have the lowest staff days needed, they are also more biased, as shown in Figure S9.4. In Table S9.6, we observed similar performance among the snapshot variations that included observations from all three seasons for each included location, though somewhat worse for the model with locations dropped at random rather than by cluster. Much larger reductions in performance were observed for designs with fewer repeated observations.

For $PM_{2.5}$, there are not many good alternatives to home monitoring. Because all monitoring methods for particles require electricity, the recruitment and coordination burden is relatively high, even if public locations could be used rather than homes. Given that public locations could have worse spatial compatibility than participant homes, we generally recommend conducting monitoring at participant homes (Bi et al., 2022; Szpiro & Paciorek, 2013).

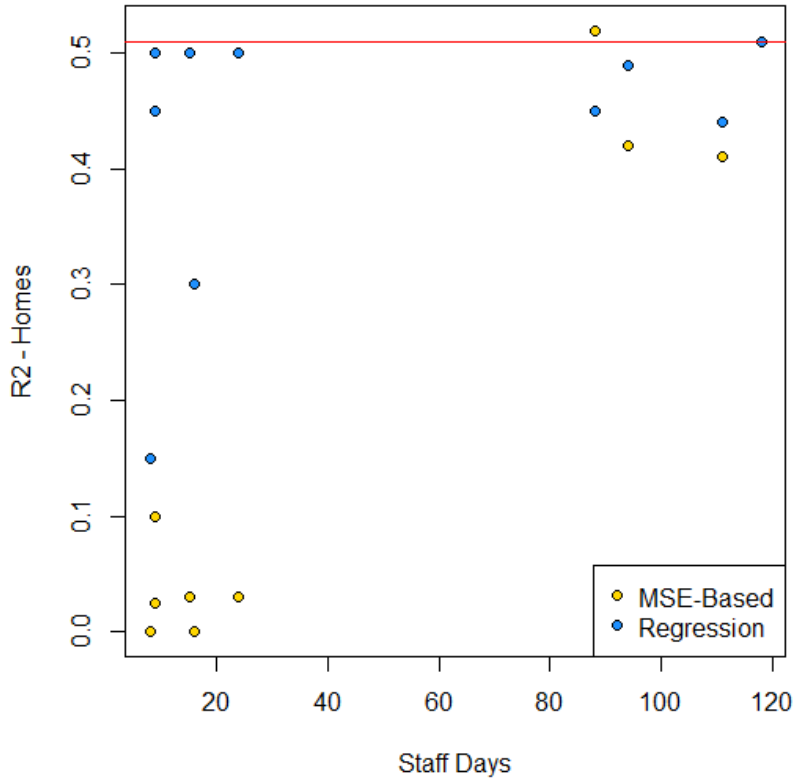



*Figure S9.4: Number of projected staff days versus model performance at participant homes for $NO_2$ models. Different performance metrics show that while most study designs produce appropriate spatial contrasts (high regression $R^2$), the snapshot-based models with the lowest staff days needed are more biased (low MSE-based $R^2$)*

Performance was more variable and more reduced among the spatially reduced $PM_{2.5}$ designs than among the temporally-reduced designs. However, it's important to note that because of the rolling nature of home monitoring, the temporally-reduced designs still cover the majority of the original temporal extent of the monitoring period. This can be contrasted with the snapshot designs used for the $NO_2$ model. While the snapshot locations cover a large spatial area, even the model that uses all three monitoring campaigns only covers six weeks of the 2-year period that the home monitors cover. Since we aren't able to assess performance at home monitors in a purely spatial way, it's difficult to assess the extent to which temporal incompatibility contributes to poor performance in the models. In an ideal world, we would assess this by using a contrasting snapshot monitoring design that observed a few locations at a time, on a rolling basis. This design would have the same spatial incompatibility with the homes as the design that was used, but would cover an equivalent temporal extent. That design would require considerable staff time and offer fewer advantages compared to home monitoring, but would answer the question about temporal versus spatial compatibility.

The biggest challenge with trying to interpret the results from the $NO_2$ analyses is that there were multiple differences between the snapshot-based and home-based supplementary monitoring designs. Snapshot locations used a less noisy, well-validated measurement method, but the locations were not residential (and thus had worse spatial compatibility), and the observations captured less temporal variability. Home monitoring used a noisy monitoring method, and observations were staggered over several years, covering far more temporal variability than the snapshot monitors. This means that the validation at participant homes is not purely spatial, even when averaging over multiple observation periods, whereas

the validation at snapshot locations is purely spatial. Another feature of this analysis is that the models presented here were not subject to the same level of scrutiny as the primary model was, since we used the parameters from the primary model and fit the models discussed here without additional tuning. We observed a fair amount of bias in the predictions at participant homes in the models based only on AQS and snapshot data, but we do not know whether that should be attributed to a method difference, to spatial incompatibilities between the two types of locations, or whether there were other temporal phenomena, such as wildfires or weather, that might be primarily responsible for the discrepancy. Therefore, we cannot be certain that simultaneous monitoring at a large number of participant homes would address the bias we observed.

*Table S9.8: Staff Time (number of days) Needed for Different Designs.*[a]

| Model | N Home Locations | N Snapshot Locations | N Seasons or Periods | Staff Time (Days) |
|---|---|---|---|---|
| **$NO_2$ Models** | | | | |
| Gold Standard | 117 | 110 | All | 204 |
| 1 Snapshot | 0 | 97 | 1 | 10 |
| 2 Snapshot | 0 | 97 | 2 | 20 |
| All Snapshot | 0 | 97 | 3 | 30 |
| 1/3 Snapshot Locations; Stratified | 0 | 37 | 3 | 11 |
| 2/3 Snapshot Locations; Stratified | 0 | 65 | 3 | 20 |
| 1/3 Snapshot Locations; Random | 0 | 37 | 3 | 11 |
| 1 LCS Observation per Home | 97 | 0 | 1 | 88 |
| 2 LCS Observations per Home | 65 | 0 | 2 | 111 |
| 3 LCS Observations per Home | 37 | 0 | 3 | 94 |
| **$PM_{2.5}$ Models** | | | | |
| Gold Standard | 117 | - | All | 171 |
| 1 Observation per Home | 82 | - | 1 | 74 |
| 2 Observations per Home | 82 | - | 2 | 140 |
| 10% Locations | 9 | - | 2 | 15 |
| 25% Locations | 21 | - | 2 | 36 |
| 50% Locations | 41 | - | 2 | 70 |
| 75% locations | 62 | - | 2 | 106 |
| 90% Locations | 74 | - | 2 | 125 |

a. A team of 2 field technicians visit each location for both set-up and take-down. For home locations, staff time is needed to recruit, consent, and schedule visits. Staff can visit approximately 55 snapshot locations or 5 home locations in 1 work day, and we budget 1 staff day per 10 locations for recruitment and consent. Ogawa snapshots require approximately 3 staff days per 110 locations for sampler assembly and disassembly.

*Conclusions*

In some settings, it may not be feasible to collect air pollution measurements at participant homes or on a rolling basis that allows measurement throughout the year. Particularly in multi-site designs where monitoring is conducted by excursions from the research center to the study sites, data collection designs are often logistically constrained. Snapshot designs are an appealing solution, but the snapshot designs in our study performed more poorly than anticipated. Because they were worse across the board, we cannot determine whether this is a spatial compatibility problem (near roads instead of at homes), a temporal compatibility problem (six weeks out of two years), or a method problem (Ogawa vs Alphasense

monitor). Based on the results presented in Table S9.6, we recommend reducing the number of $NO_2$ snapshot locations before reducing the number of monitoring seasons whenever possible. In contrast, the $PM_{2.5}$ results suggest that given a study design includes home monitoring that covers the full temporal extent of the possible monitoring period, it is preferable to reduce the number of repeat visits to the same location before reducing the number of locations, as long as the remaining measurements sufficiently cover the temporal extent of the monitoring period. It's also preferable to spread out those monitoring locations over space as much as is feasible.

*References for Cost and Performance Comparisons for Low-Cost Monitors Supplementing Regulatory Monitoring Data*


Bi, J., Burnham, D., Zuidema, C., Schumacher, C., Gassett, A. J., Szpiro, A. A., Kaufman, J. D., & Sheppard, L. (2024). Evaluating low-cost monitoring designs for PM2.5 exposure assessment with a spatiotemporal modeling approach. *Environmental Pollution*.

Bi, J., Carmona, N., Blanco, M. N., Gassett, A. J., Seto, E., Szpiro, A. A., V. Larson, T., Sampson, P. D., Kaufman, J. D., & Sheppard, L. (2022). Publicly available low-cost sensor measurements for PM2.5 exposure modeling: Guidance for monitor deployment and data selection. *Environment International*, *158*, 106897. https://doi.org/10.1016/j.envint.2021.106897

Blanco, M. N., Doubleday, A., Austin, E., Marshall, J. D., Seto, E., Larson, T. V., & Sheppard, L. (2022). Design and evaluation of short-term monitoring campaigns for long-term air pollution exposure assessment. *Journal of Exposure Science & Environmental Epidemiology*. https://doi.org/10.1038/s41370-022-00470-5

Blanco, M. N., Gassett, A., Gould, T., Doubleday, A., Slager, D. L., Austin, E., Seto, E., Larson, T. V., Marshall, J. D., & Sheppard, L. (2022). Characterization of Annual Average Traffic-Related Air Pollution Concentrations in the Greater Seattle Area from a Year-Long Mobile Monitoring Campaign. *Environmental Science & Technology*, *56*(16), 11460–11472. https://doi.org/10.1021/acs.est.2c01077

Keller, J. P., Olives, C., Kim, S.-Y., Sheppard, L., Sampson, P. D., Szpiro, A. A., Oron, A. P., Lindstrom, J., Vedal, S., & Kaufman, J. D. (2015). A Unified Spatiotemporal Modeling Approach for Predicting Concentrations of Multiple Air Pollutants in the Multi-Ethnic Study of Atherosclerosis and Air Pollution. *ENVIRONMENTAL HEALTH PERSPECTIVES*, *123*(4), 301–309. https://doi.org/10.1289/ehp.1408145

Szpiro, A. A., & Paciorek, C. J. (2013). Measurement error in two-stage analyses, with application to air pollution epidemiology. *Environmetrics*, *24*(8), 501–517. https://doi.org/10.1002/env.2233

US EPA. (2023). *Air Quality System Data Mart*. http://www.epa.gov/ttn/airs/aqsdatamart

Zuidema, C., Bi, J., Burnham, D., Carmona, N., Gassett, A. J., Slager, D. L., Schumacher, C., Austin, E., Seto, E., Szpiro, A. A., & Sheppard, L. (2024). Leveraging low-cost sensors to predict nitrogen dioxide for epidemiologic exposure assessment. *Journal of Exposure Science & Environmental Epidemiology*. https://doi.org/10.1038/s41370-024-00667-w

Zuidema, C., Schumacher, C. S., Austin, E., Carvlin, G., Larson, T. V., Spalt, E. W., Zusman, M., Gassett, A. J., Seto, E., Kaufman, J. D., & Sheppard, L. (2021). Deployment, Calibration, and Cross-Validation of Low-Cost Electrochemical Sensors for Carbon Monoxide, Nitrogen Oxides, and Ozone for an Epidemiological Study. *Sensors*, *21*(12), 4214. https://doi.org/10.3390/s21124214

Zusman, M., Schumacher, C. S., Gassett, A. J., Spalt, E. W., Austin, E., Larson, T. V., Carvlin, G., Seto, E., Kaufman, J. D., & Sheppard, L. (2020). Calibration of low-cost particulate matter sensors: Model development for a multi-city epidemiological study. *Environment International*, *134*, 105329.

# Chapter 10

To help fix ideas developed in Chapter 10's synthesis, several key figures are reproduced here.

## EXPOSURE MODEL RESULTS FROM THE SEATTLE MOBILE MONITORING CAMPAIGN

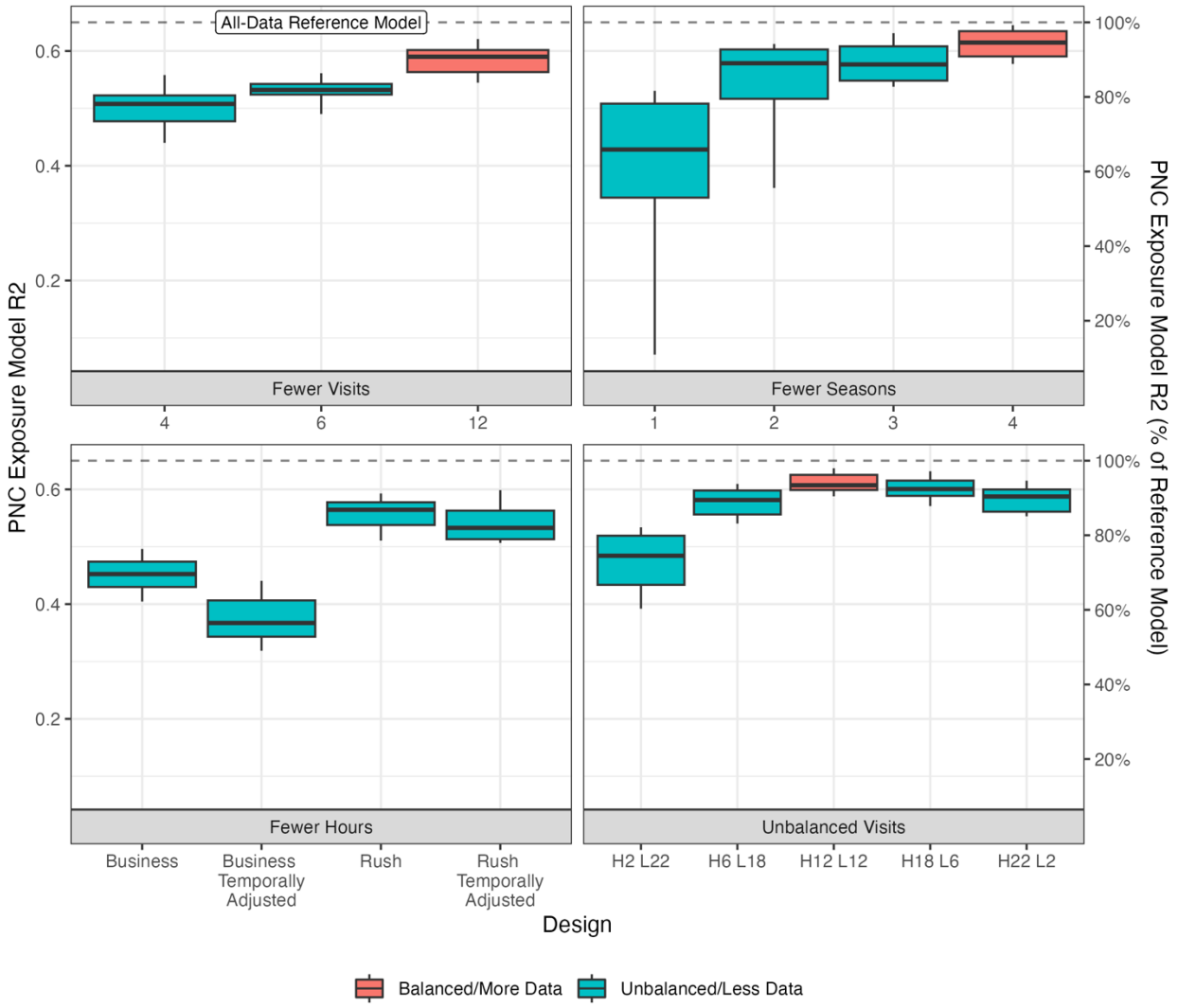



*Figure 4.4. Cross-validated UFP model performances (N=30 campaigns per design) using stationary roadside data from the Seattle mobile monitoring campaign. The dashed lines indicate the all-data campaign performance. Red design reference boxplots indicate the least restrictive or most balanced campaigns; any of these can serve as a reference for the business and rush hours designs. Business and rush hour designs produce annual average site estimates from unadjusted and temporally-adjusted visits. Models are for a total of 10-420 nm PNC (pt/cm3) from the NanoScan instrument. The dashed line indicates the $R^2$ from the reference all-data model, which is 0.65.*

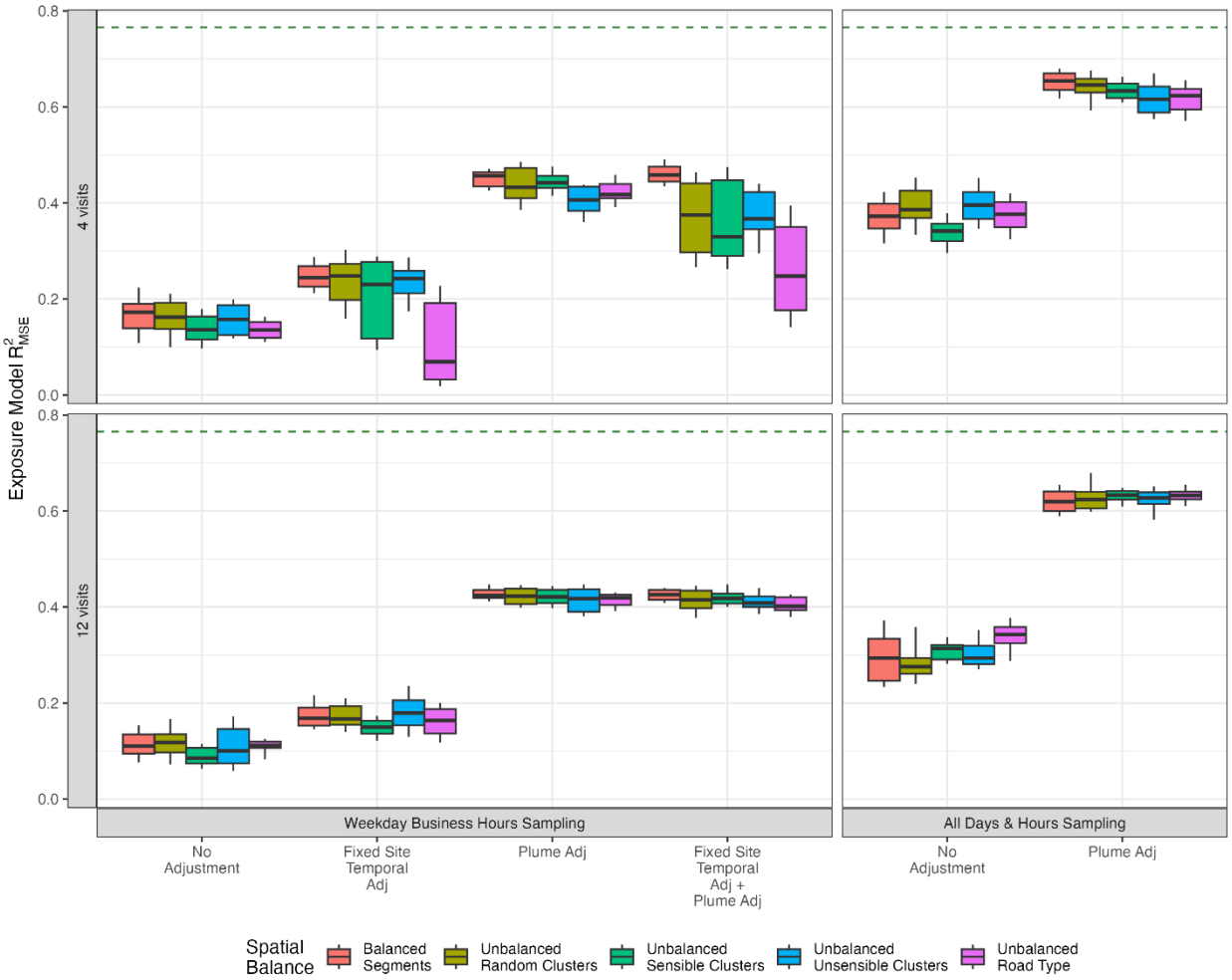



*Figure 6.2. Out-of-sample PNC (pt/cm$^3$) exposure model performances for on-road campaigns (N=30 campaigns per combination - i.e., boxplot). $R^2_{MSE}$ is based on the comparison of the predicted PNC at 309 stationary locations and the annual average site estimate from stationary roadside measures. Models are for total PNC (20-1,000 nm) from the unscreened P-TRAK instrument and the Seattle mobile monitoring campaign. Boxes show the median and IQR, whiskers show the 10$^{th}$ and 90$^{th}$ percentiles. The dashed line indicates the $R^2$ from the reference all-data stationary model, which is 0.77.*

## HEALTH MODEL RESULTS FROM THE SEATTLE MOBILE MONITORING CAMPAIGN

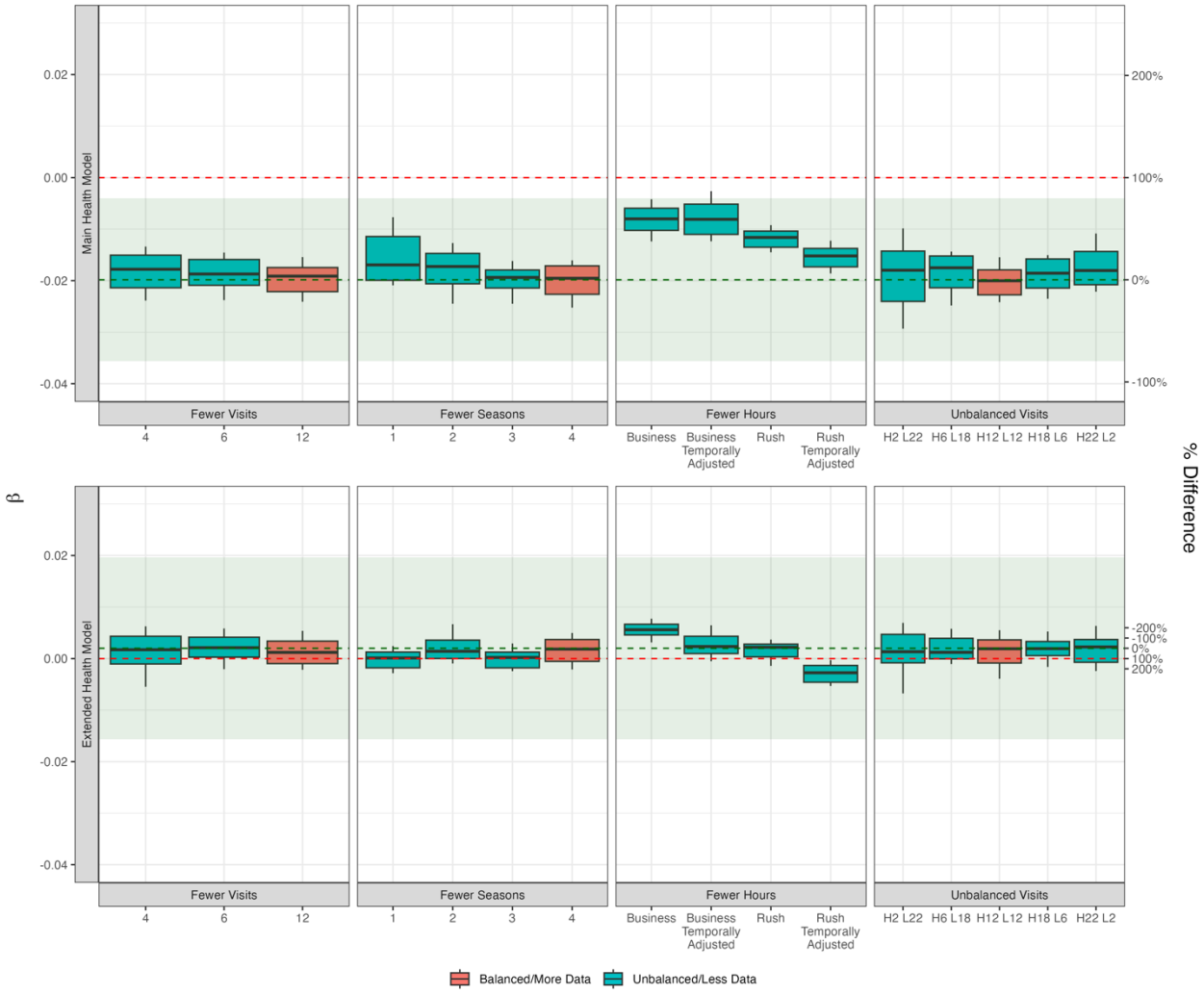



*Figure 4.5. Estimated adjusted health association between PNC (1,900 pt/cm$^3$) and cognitive function (CASI-IRT) adjusted for age, calendar year, sex, education (main health models, Model 1), and extended health models (Model 2) further adjusted for race and SES using stationary roadside data from the Seattle mobile monitoring campaign. The dashed green lines and shaded areas indicate the estimated point and 95% CIs from the all-data exposure model, which are -0.020 (95% CI: -0.036, -0.004) in the main model and 0.002 (95% CI: -0.016, 0.020) in the extended model. The dashed red line indicates no association. Boxplots show the results when using exposure estimates from reduced mobile monitoring sampling campaigns (N=30 estimates per boxplot). Boxes show the median and IQR, whiskers show the 10$^{th}$ and 90$^{th}$ percentiles. Percentages on the y-axis show the estimated association relative to using the all-data exposure model.*

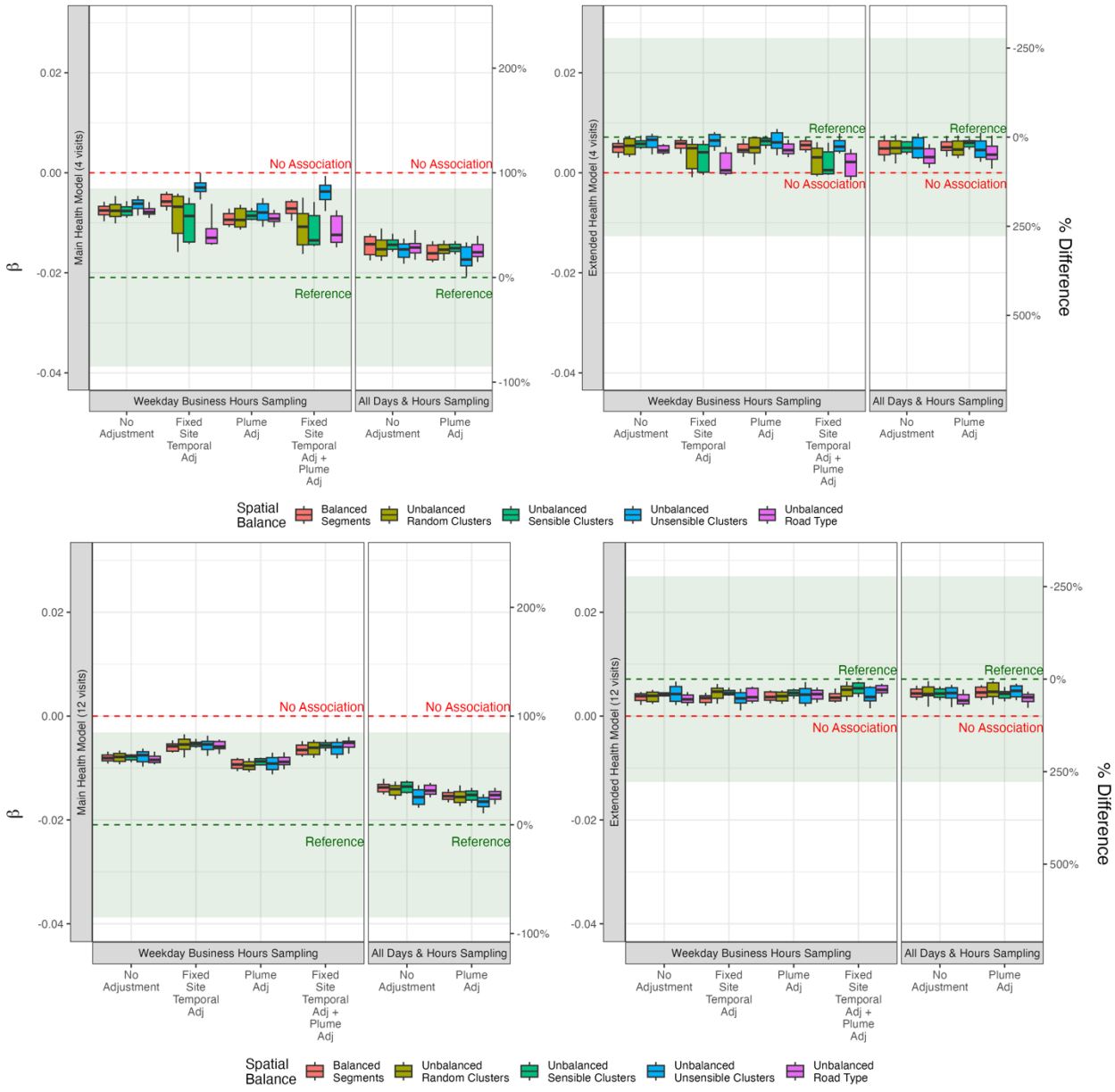



*Figure 6.3 and Figure S6.5. Estimated association between cognitive function (CASI-IRT) and PNC (1,900 pt/cm$^3$) at baseline in the ACT cohort in main (Model 1) and extended adjusted (Model 2) health models using 4-visit and 12-visit campaigns. PNC exposures are predicted from on-road monitoring campaigns. The dashed green lines and shaded areas indicate the estimated point and 95% CIs from the all-data roadside exposure model, which are -0.021 (95% CI: -0.039, -0.003) in the main model and 0.007 (95% CI: -0.013, 0.027) in the extended model. The dashed red line indicates no association. Boxes show the median and IQR, whiskers show the 10$^{th}$ and 90$^{th}$ percentiles.*