

SIBaR: A New Method for Background Quantification and Removal from Mobile Air Pollution Measurements

Blake Actkinson¹, Katherine Ensor², Robert J. Griffin^{1,3}

¹Department of Civil and Environmental Engineering, Rice University, Houston, TX 77005, USA

5 ²Department of Statistics, Rice University, Houston, TX 77005, USA

³Department of Chemical and Biomolecular Engineering, Rice University, Houston, TX 77005, USA

Correspondence to: Robert Griffin (rob.griffin@rice.edu)

Abstract. Mobile monitoring is becoming increasingly popular for characterizing air pollution on fine spatial scales. In identifying local source contributions to measured pollutant concentrations, the detection and quantification of background are key steps in many mobile monitoring studies, but the methodology to do so requires further development to improve replicability. Here we discuss a new method for quantifying and removing background in mobile monitoring studies, State Informed Background Removal (SIBaR). The method employs Hidden Markov Models (HMMs), a popular modelling technique that detects regime changes in time series. We discuss the development of SIBaR and assess its performance on an external dataset. We find 83% agreement between the predictions made by SIBaR and the predetermined allocation of background and non-background data points. We then assess its application to a data set collected in Houston, by mapping the fraction of points designated as background and comparing source contributions to those derived using other published background detection and removal techniques. Presented results suggest that SIBaR modelled source contributions contain source influences left undetected by other techniques, but that it is prone to unrealistic source contribution estimates when it extrapolates. Results suggest that SIBaR could serve as a framework for improved background quantification and removal in future mobile monitoring studies while ensuring that cases of extrapolation are appropriately addressed.

1 Introduction

Understanding air pollution exposure is important, as it has been linked to various adverse health conditions (Caplin et al., 2019; Zhang et al., 2018). Mobile monitoring, a technique in which continuous air pollution measurements are collected using instrumentation on a mobile platform, is becoming increasingly important for characterizing exposure because air pollution varies on spatial scales finer than the typical distance between stationary monitors (Apte et al., 2017; Chambliss et al., 2020; Messier et al., 2018).

A key component of mobile monitoring analysis is identifying ambient background levels, defined here as measured air pollution concentrations independent of local source influences (Brantley et al., 2014). Background quantification is vital from both policy and exposure perspectives, as it is important to assess the contribution of local sources to pollution concentrations

accurately. Table 1 summarizes the wide variety of methods used to estimate background in studies incorporating mobile monitoring published within the past 5 years. The wide variance in the approaches used is problematic, as estimates of source contributions to measurements have been shown to be sensitive to the technique used (Brantley et al., 2014). To improve the replicability and power of mobile monitoring studies, a more consistent technique for background estimation is needed.

35

Designing a method to determine the background in mobile monitoring studies presents several challenges. Measurements in remote locations are often regarded as the most reliable representation of background concentrations; however, remote locations may be inaccessible for some mobile monitoring studies and are themselves subject to occasional source influences. These drawbacks make time series methods for determining background more desirable. However, many time series-based methods often rely on setting static time windows, which are usually determined by the expected duration of influence from source plumes within the mobile monitoring study (Bukowiecki et al., 2002). The underlying physical representation of these time series methods remains unclear for more extensive mobile monitoring campaigns, as the setting of static time windows does not often capture the entire variation in time scales that source impacts can have on mobile measurements.

40

Here we show the results of a newly developed method called **State-Informed Background Removal (SIBaR)** used to estimate background for several traffic related air pollutants, namely nitrogen oxides (NO_x) and carbon dioxide (CO_2). The method incorporates Hidden Markov Models (HMMs), a time series regime modelling technique used in a wide variety of contexts in signals processing, finance, and the social sciences and which has been used to model background in stationary monitors (Gómez-Losada et al., 2016, 2018, 2019; Visser and Speekenbrink, 2010). HMMs assume that observations within a time series are drawn from probability distributions governed by a hidden sequence of states. We propose decoding this hidden sequence of states as a way to determine whether measurements were taken during time periods representative of background versus time periods subject to local influences. We illustrate that a more physically meaningful representation of background is captured in this modelling context for mobile monitoring time series and show its application to a wide variety of traffic related air pollutant measurements. As a proof of concept, we run the method on a published external dataset already marked as background and non-background and assess its performance. As a first application and to provide further proof of concept, we map points binned as background by SIBaR to show their spatial distributions. As a proof of importance, we highlight differences in mapped source contributions derived from SIBaR background and background derived from other time series-based techniques. Results indicate that our consistent method for background identification and removal has noticeable impact on mapped mobile source contributions.

50

55

60

Study	Method Used to Determine Background Concentration
Apte et al., 2017	Applied 10-s moving average filter, then selected the smaller of the given data value or the 2-min 5 th percentile to derive baseline concentrations.
Brantley et al., 2019	Fitted quantile regression with cubic natural spline basis expansion of time with degrees of freedom equal to the number of hours in the time series.
Hankey and Marshall, 2015	Used pollutant-specific underwrite functions to estimate instantaneous background concentrations and subtracted these concentrations from the original time series, averaged reference monitor measurements, then added averaged measurements to underwrite adjusted time series.
Hankey et al., 2019	Used hourly averaged measurements in centrally located site for additive correction factor; used daily median fixed-site measurement for temporal correction factor.
Hudda et al., 2014	Applied rolling 30-s 5 th percentile of the original time series.
Larson et al., 2017	Applied 10-min rolling minimum.
Li et al., 2019	Applied 1-min moving median filter, then calculated 1-hr rolling 5 th percentile of smoothed data; additionally, used wavelet decomposition to isolate concentration changes across 8 hours at stationary monitors, then subtracted lowest decoupled concentration from mobile monitoring time series across 15-min time windows.
Patton et al., 2014	Used mobile measurements in designated urban background neighborhoods removed from highway.
Robinson et al., 2018	Linearly interpolated averaged data collected at designated background locations.
Shairsingh et al., 2018	Applied rolling 60-s mean, then applied spline of minimums technique (Brantley et al., 2014) across different time windows dependent on a desired background scale.
Tessum et al., 2018	Used daily 5 th percentile for all pollutants other than fine particle number concentration; used rolling 30-min 5 th percentile for fine particle number concentration.
Van den Bossche et al., 2015	Used averaged measurements from stationary monitor located in an urban green to apply additive correction factors to measurements greater than background then averaged site measurement and multiplicative correction factors to measurements lower than background.

Table 1. Summary of Previous Methodologies for Estimating Background Levels of Air Pollution in Mobile Monitoring Campaigns.

2 Methods

2.1 Mobile Campaign

65 Measurements were taken during the Houston Mobile Monitoring Google Street View (GSV) campaign and are described in detail elsewhere (Miller et al., 2020). Measurements were conducted over a 9 month period spanning July 2017 to March 2018. Sampling primarily took place between 7:00 and 16:00 local standard time (Miller et al., 2020) in a variety of census tracts across metropolitan Houston. Census tracts are included in the current analysis if they were sampled a minimum of 15 times during this 9-month period (Apte et al., 2017; Li et al., 2019). Details and names used to describe each census tract are given
70 in Table S1. The time of day and day of week for each census tract visit were predetermined to minimize temporal biases in sampling to the greatest extent possible. Instruments (Table S2) were loaded into 2 gasoline-powered GSV cars that sampled every drivable road in 22 different census tracts in the greater Houston area. Details and names used to describe each of the census tracts considered are given in Table S1. Individual observations are aggregated to 50-meter points in neighborhoods and 90-meter points on highways using a road network created from U.S. Census TigerLine roads (TIGER/Line Shapefile,
75 2018). More details on the road network creation and data quality control are provided elsewhere (Miller et al., 2020). Data quality and control measurements were implemented to ensure sound statistics were performed. Measurements were removed if they were taken during calibration periods, during periods of suspected instrument failure, and if they were outside of an instrument's reported operating range. Measurements were synchronized to GPS time stamps and adjusted for inlet residence time differences based on results from match strike tests. Measured pollutants include black carbon (BC), carbon dioxide
80 (CO_2), nitric oxide (NO), and nitrogen dioxide (NO_2) ($\text{NO}_x = \text{NO} + \text{NO}_2$),.

Bias, precision, and the minimum detection limit (MDL) for each instrument are provided in Table S2. Details concerning the calculation of each parameter for each instrument are given elsewhere (Miller et al., 2020). In brief, the bias for the T200 NO Analyzer and T500U NO_2 Analyzer were calculated from gas calibration checks performed every 2 weeks at the start of the study period and every month towards the end of the study period, because the checks routinely showed bias $< \pm 10\%$. The bias
85 for the Li-COR was determined from a gas phase calibration before the start of the study to match the manufacturer reported value. Precision values for the T200 and T500U were calculated as the standard deviation of zeroing periods taken throughout the entire campaign. Minimum detection limits for the T200 and T500U were determined as the mean of the time series zero $+ 3\sigma$. The minimum detection limit and precision of the Li-COR were not considered due to taking measurements at a consistently elevated global background and the latter manufacturer's reported value having a miniscule effect on the overall
90 uncertainty of the measurement. For the purposes of this work, we perform no MDL substitution, as MDL substitution would censor the underlying modelled background probability distribution.

2.2 Hidden Markov Model Categorization – The Background Partitioning Step

Time series observations are segregated by day and for each GSV car, and HMMs are fit to each day's worth of data. Before fitting the HMM to each day's time series realizations, we log transform them. HMMs attempt to maximize the log-likelihood, L_c , determined by the sum of the forward variables $\alpha_T(i)$:

$$L_c = \sum_i^N \alpha_T(i) \quad (1)$$

in which i designates state i (total states N) at the last realization of the time series T . The forward variables are derived recursively as:

$$\alpha_1(i) = \pi_i p(y_1 | \theta_i, z) \quad (2)$$

$$\alpha_{t+1}(j) = \sum_i^N (\alpha_t(i) a_{ij}) p(y_t | \theta_j, z) \quad (3)$$

in which π_i represents the initial probability for state i , a_{ij} represents the state transition probability from state i to state j , and $p(y_t | \theta_i, z)$ represents the conditional probability of observation y_t conditioned on the parameters θ_i governed by state i and any additional covariates z . For the purposes of our work, we assume that the probability distributions governing y_t are log normal and parametrize the mean of the response distribution as:

$$\mu_t = \widehat{\beta}_0 + \widehat{\beta}_1 t \quad (4)$$

where μ_t is the time-dependent mean of the response, $\widehat{\beta}_0$ and $\widehat{\beta}_1$ are estimated parameters, and t is time.

The log-likelihood of equation (1) is maximized using the expectation maximization algorithm (Dempster et al., 1977; Visser and Speekenbrink, 2010). Initial starting values of the transition probabilities are bootstrapped 150 times to produce 150 candidate models because convergence to a maximum likelihood can be affected by the starting values. The model with the greatest log-likelihood is then selected for decoding via the Viterbi algorithm (Forney, 1973). The Viterbi algorithm seeks to maximize the joint probability of both observations and state sequence $(q_1, \dots, q_T)$ given the parameters. We define a variable δ recursively as

$$\delta_{t+1}(j) = [\max_i \delta_t(i) a_{ij}] p(y_{t+1} | \theta_j, z) \quad (5)$$

with the initialization

$$\delta_1(i) = \pi_i p(y_1 | \theta_i, z) \quad (6)$$

To retrieve the state sequence, we create a matrix ψ such that

$$\psi_1(i) = 0 \quad 1 \leq i \leq N \quad (7)$$

$$\psi_t(j) = \operatorname{argmax}(\delta_{t-1}(i) a_{ij}) \quad 1 \leq j \leq N, 2 \leq t \leq T \quad (8)$$

We retrieve the state sequence by backtracking:

$$q_T = \operatorname{argmax}[\delta_T(i)] \quad 1 \leq i \leq N \quad (9)$$

$$q_t = \psi_{t+1}(q_{t+1}) \quad t = T - 1, T - 2, \dots, 1 \quad (10)$$

This state sequence is then used to designate observations as background or source. State assigned points with the lower median are designated background. An example of a decoded sequence is given in Figure 1 for NO_x (after retransformation).

HMM fits can be highly sensitive to time series outliers (Svensén and Bishop, 2005; Chatzis and Varvarigou, 2007; Chatzis et al., 2009). Additionally, while computationally cheap, the linearity assumption embedded in the time covariate could fail to capture more complex variations in background and produce flawed state categorizations. To capture misclassification instances, we recast the step as an unsupervised learning problem, design an empirical routine to evaluate the quality of created clusters, and incorporate it into SIBaR. The routine, coined the fitted line classifier, fits a line between averaged transition measurements and their corresponding transition times. The method then calculates the percentage of points above the line that are classified as background and the percentage of points below the line that are classified as source. If either percentage is greater than or equal to 50%, a predetermined percentage threshold, the method deems the series incorrectly classified. If a series is incorrectly classified, SIBaR breaks the series into two and performs the background partitioning step on each half chunk separately. 16 example time series, labelled as classified correctly or incorrectly, are depicted in Figures S1 and S2. After fitting HMMs to each separate chunk, SIBaR then uses the fitted line classifier on each chunk, repeating the process if any chunk's partitioning is labeled misclassified. The process continues recursively until all created partitions are deemed correctly classified. SIBaR then combines the state designations from all created chunks into one and returns those state designations as the corrected designations for the time series.

In running SIBaR on the campaign NO_x measurements, we note that the empirical classifier designates 96% of the original time series to be correctly classified for a 50% threshold. We run a sensitivity analysis on the percentage threshold and show the results in Figure S3. The figure illustrates that changing the percentage threshold causes changes in the percentage of correctly classified time series to range between 80-100%, dipping below 50% only for the most stringent requirement (5%). These results give us confidence in the partitioning step.

2.3 Natural Spline Fit

After HMMs have been fit to all time series data, natural splines are fit to the background points by day. As in the work published by Brantley et al. ("Brantley", Brantley et al., 2019), we select a natural spline basis with the degrees of freedom equal to the number of hours in the time series. However, we fit to the mean of our partitioned background time series, whereas in Brantley the focus is on a 10th quantile regression. An example of this spline fit is given in Figure 1.

Because SIBaR's partitioning step periodically generates background assigned points that differ from one another for the same time series, we perform a test to evaluate its robustness. We run SIBaR 25 times and evaluate the pairwise root-mean-square error (RMSE) between each set of generated background predictions for NO_x as defined below.

$$RMSE = \sqrt{\frac{\sum_t^T (n_{ta} - n_{tb})^2}{T}} \quad (11)$$

175

In which n_{ta} is the background realization at time t of signal a , n_{tb} is the background realization at time t of signal b , and T is the total number of realizations in the time series.

The pairwise RMSE values for the first 12 runs are given in Table S3. We calculate an average RMSE of 0.05 ± 0.02 ppb between each background signal and conclude that the fitting step is robust to small changes in background assigned points in the partitioning step.

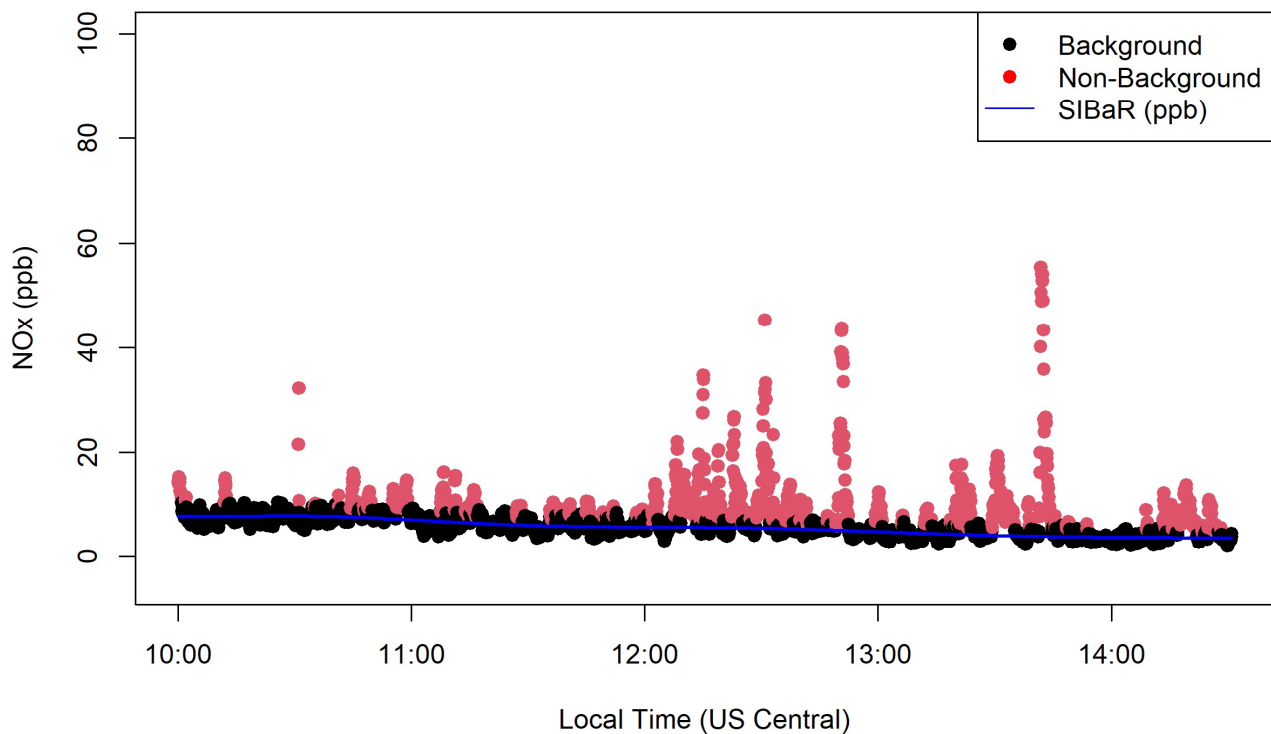


Figure 1. Example time series of SIBaR background signal (blue) being fit to background designated points (black) created in the SIBaR partitioning step. Points are colored red if designated as source in the partitioning step, black if designated background. Data presented in this time series were collected on 3/30/2018 starting at 10 AM.

2.4 Evaluating the Partitioning Step: Validation on an External Dataset

To test the validity of the partitioning step, we perform external validation using a mobile monitoring dataset published in Brantley et al. (2014). In that study, a van collecting mobile measurements of carbon monoxide (CO) systematically looped a route in which it drove through a predefined background location, on transects to a highway, and on the highway itself (Brantley et al., 2014). The measurements taken in the prescribed background location were marked as background, and all other measurements were marked as non-background. We run the partitioning step on these data to determine how well SIBaR captures the measurements taken in the background location of the study.

2.5 Generating Mapped Fractional Background Contribution and Source Contribution Maps

190 We explore the spatial extent of our HMM decoded categorizations from the partitioning step by creating mapped fractional background contribution maps. After aggregating time series observations (either CO₂ or NO_x, depending on the pollutant analysed) to road segment points created within our road segment network, we sum the number of observations designated as the background state and divide by the total number of observations assigned to that road segment point. We map the results and present them in section 3.2.

195

In section 3.3, we derive source contributions (source signal = original signal – background signal) using our background method and map them. To put these source contributions in context with previously published work, we repeat the same process using background derived from a moving 2 minute 5th percentile baseline (“Apte,” Apte et al., 2017) and the Brantley technique described previously in section 2.2 (“Brantley,” Brantley et al., 2019). To derive our source contributions, we make predictions
200 for the background for each time series realization collected using the derived background spline and then subtract those predictions from the original time series observations. We also derive source contributions using the Apte and Brantley techniques. We create the maps using the same methodology as Miller et al. (2020), described briefly here. Using our created road segment network, we take the mean of measurements collected as the GSV car drives past a road segment point in our network, coined the drive pass mean. We take the median of these drive pass means and map the result. Because we consider
205 drive pass means taken within 4 hours of one another to provide no new information about the air quality at that road segment, we take the median of drive pass means occurring within that 4 hour time window to generate a 4 hour median of drive pass means. Then, we take the median of all 4 hour medians of drive pass means at that road segment to derive its map reduced median. We perform this procedure for the source contributions derived using our method and the source contributions derived using the other published methods.

210 3 Results – Proof of Concept

3.1 Validating the Partitioning Step on an External Dataset

A comparison between SIBaR’s partitioning and the partitioning originally published by Brantley et al. (2014) is given in Figure 2. Initially, the HMM fitting step is performed and the resulting state sequence decoded. We run our classifier on the initially decoded time series and find it to be misclassified, which is apparent from panel (a) of Figure S4 that shows the

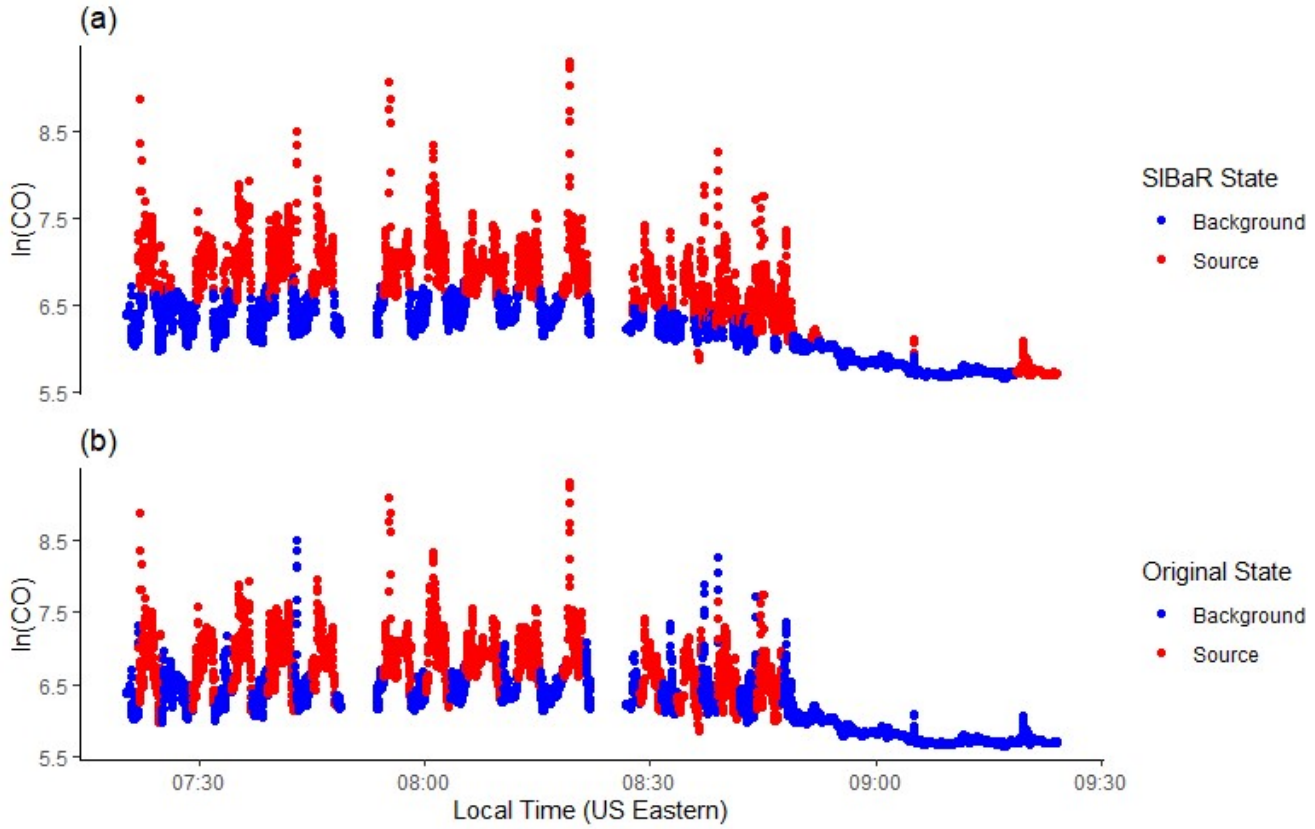


Figure 2. Comparison between SIBaR-predicted background and source states and originally published designations from Brantley et al. (2014) for log transformed CO. Background designated points are in blue, source designated points in red. (a) SIBaR decoded states for the mobile CO measurements. (b) Designations originally published by the authors of the study.

215 unsmoothed CO data before correction. The algorithm breaks the series into 2 chunks and refits the HMM to each part separately, resulting in the state designations in panel (a) of Figure 2. We then compute the percentage of matching background/non-background designations. The SIBaR partitioning step is able to match 83% of the originally published background/non-background designations. The mismatches could be attributed to the transition between the background/non-background portions of the route in the original study, which is observed in Figure 2 in the periods where background points

220 show larger values than source points near periods of the transition (for example., the last blue spike at approximately 8:45AM). Mismatches also could be a result of the effects of traffic on measurements in the background designated portion of the route. Finally, the mismatches could be attributed to the inability of the SIBaR linearity assumption to capture finer scale temporal variations within the background (see equations (2)-(4)).

225 In running this test, we note that the method is sensitive to a smoothing time window if one is used. Figure S4 illustrates SIBaR uncorrected decoded states for 3 different smoothing time windows on the same CO dataset and shows that the method

produces different state categorizations depending on the window used, even making correction unnecessary in the 30 s instance. We hypothesize that smoothing reduces the skewness of the data such that it better fits 2 switched lognormal Gaussian distributions.

230

3.2 Mapped Fractional Background State Contributions

For the Houston mobile campaign, maps detailing the fractional contribution of the background state to the overall mapped points are created for CO_2 and NO_x . Individual observations assigned to a road segment point have their decoded category designations assigned to the same point. The number of observations assigned the background category are then divided by the total number of observations assigned to the point to determine the fractional background state contribution. Figure 3 shows these census tract maps for NO_x . Figure S5 shows the maps for CO_2 . It is important to note that these maps represent the fraction of the measurements that are categorized as background or source for the given pollutant at a given location.

We note the following about the broad spatial patterns in mapped background state fraction presented in Figure 3. First, background state designated points dominate residential areas for both pollutants. This is encouraging, as it is expected that few point sources of these 2 pollutants would be found in residential neighborhoods except for those near industrial activity (Miller et al., 2020). Second, source state designated points dominate highways and busy arterials, which is expected given the large amounts of traffic on these roads. Finally, we note the appearance of source-dominated hotspots in front of point sources identified in our previous work (Miller et al., 2020) and denote their locations in Figure 3. This is encouraging given that we found these road segments to be elevated for NO and/or NO_2 compared to their surrounding neighbourhood domain.

We take the background state fractions depicted in Figure 3 and bin them by distance to highway. The results are presented in Figure 4. We do the same for CO_2 and present the results in Figures S5-S6. The exponential behaviour exhibited in Figure 4 mirrors published exponential decays in roadside source pollutant concentrations (Apte et al., 2017; Karner et al., 2010), while the sizeable interquartile ranges within each bin highlight the complexity and variability of source roadside gradients, which depend on emission rates, meteorology, geography, and other factors (Baldwin et al., 2015; Patton et al., 2014).

255

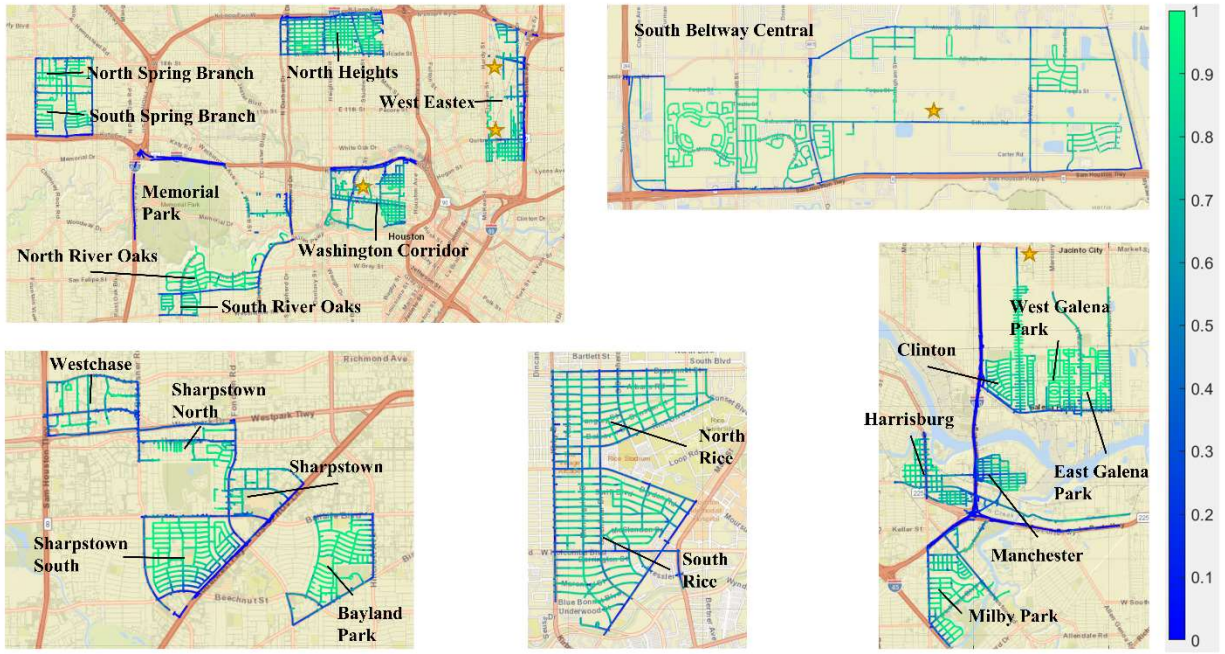


Figure 3. Fraction of points aggregated to road segment network designated as background in SIBaR decoded states for NO_x. Maps were generated following the methods outlined in Section 2.5. Points are mapped on a scale of 0 to 1; 1 implies all points aggregated to that road segment were designated as background, 0 implies all points were designated as non-background. Details of the census tracts are provided in Table S1. Gold stars indicate locations of elevated NO and/or NO₂ medians next to known industrial facilities published in Miller et al. (Miller et al., 2020). Basemap generated by Matlab geobasemap ‘streets’ and is hosted by ESRI (Sources: Esri, DeLorme, HERE, USGS, Intermap, iPC, NRCAN, Esri Japan, METI, Esri China (Hong Kong), Esri (Thailand), MapmyIndia, Tomtom).

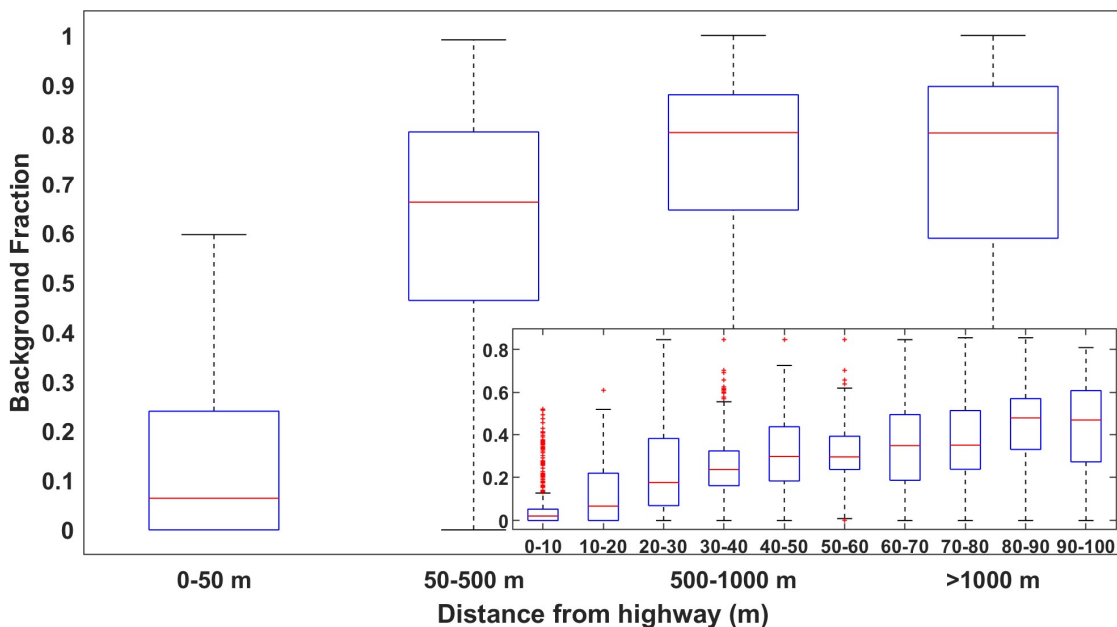


Figure 4. Boxplots of mapped background NO_x fractions, presented in Figure 3, binned by distance from highway. The red line represents the median, the top and bottom edges represent the 75th and 25th percentiles, respectively, and the whiskers extend to the most extreme data points not considered outliers.

3.3 Comparison of Source Contribution Maps Using Different Background Removal Techniques

To put SIBaR predicted source contributions in context, we compare the source contribution maps generated using SIBaR to the ones generated by the Apte and Brantley techniques. We zoom in on the Ship Channel domain for ease of comparison in Figure 5. We refer the reader to Figures S7-S15 to see maps for all other areas in the mobile monitoring campaign for both NO_x and CO_2 . The average NO_x background predicted by the Apte, Brantley, and SIBaR techniques are 15.25 ppb, 11.58 ppb, and 13.02 ppb respectively.

Figure 5 shows that the source contributions derived using the Apte technique are lower on highways compared to the source contributions derived using SIBaR and the Brantley techniques. Additionally, both Brantley and SIBaR techniques find higher source contributions on road segments with elevated NO and NO_2 concentrations found in Miller et al. (2020) compared to the Apte technique. We hypothesize this occurs due to the smaller time window utilized in the Apte technique. The GSV vehicles would often sit in traffic on highways for extended periods of time, making a 2 minute time window unsuitable for describing source durations during those time periods. While the 2 minute assumption would be better suited for situations in which the car was exposed to source durations within that time interval (which occurred in the Apte study), it would not be for source durations of a larger time interval, highlighting the challenges in assuming a static time window for extensive mobile monitoring campaigns with varying source durations.

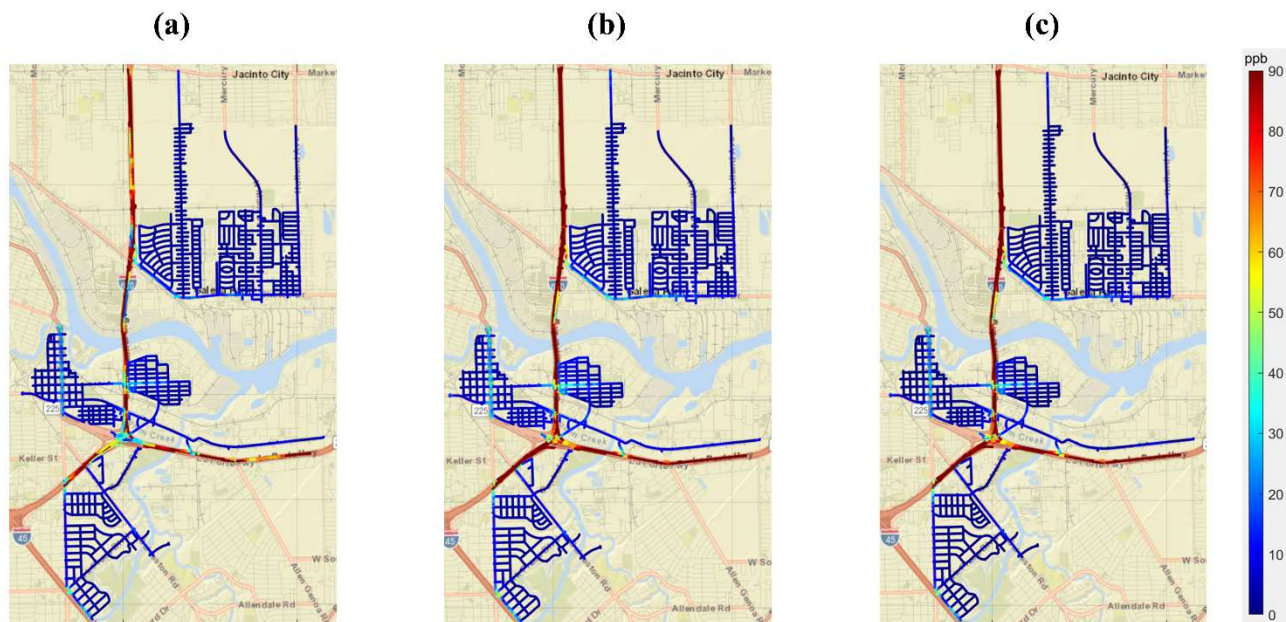


Figure 5. Comparison of source contributions derived using different techniques in the Ship Channel Domain. Source contributions were aggregated according to the methods described in Section 2.4. (a) Source contributions derived using the Apte technique. (b) Source contributions derived using the Brantley technique. (c) Source contributions derived using the SIBaR technique. Basemap generated by Matlab geobasemap ‘streets’ and is hosted by ESRI (Sources: Esri, DeLorme, HERE, USGS, Intermap, iPC, NRCAN, Esri Japan, METI, Esri China (Hong Kong), Esri (Thailand), MapmyIndia, Tomtom)

We plot road segment median source contributions derived by Apte and Brantley algorithms against the road segment median concentrations derived by SIBaR and present the results for NO_x in Figure 6. Additionally, we plot lines of best fit derived using ordinary least squares (OLS) regression. Panel (a) in Figure 6 illustrates that SIBaR derives higher source contributions medians than the Apte technique which is largely driven by differences in highway road segment medians. The slope determined using OLS regression suggests that, on average, SIBaR median source contributions are ~41% higher than Apte median source contributions. Panel (b) of Figure 6 comparing Brantley and SIBaR road segment medians indicates much closer agreement between the two techniques, with SIBaR estimating source contribution medians at an average offset of 2 ppb lower than Brantley source contribution medians. Data for CO_2 source contribution medians are shown in Figures S16 and S17.

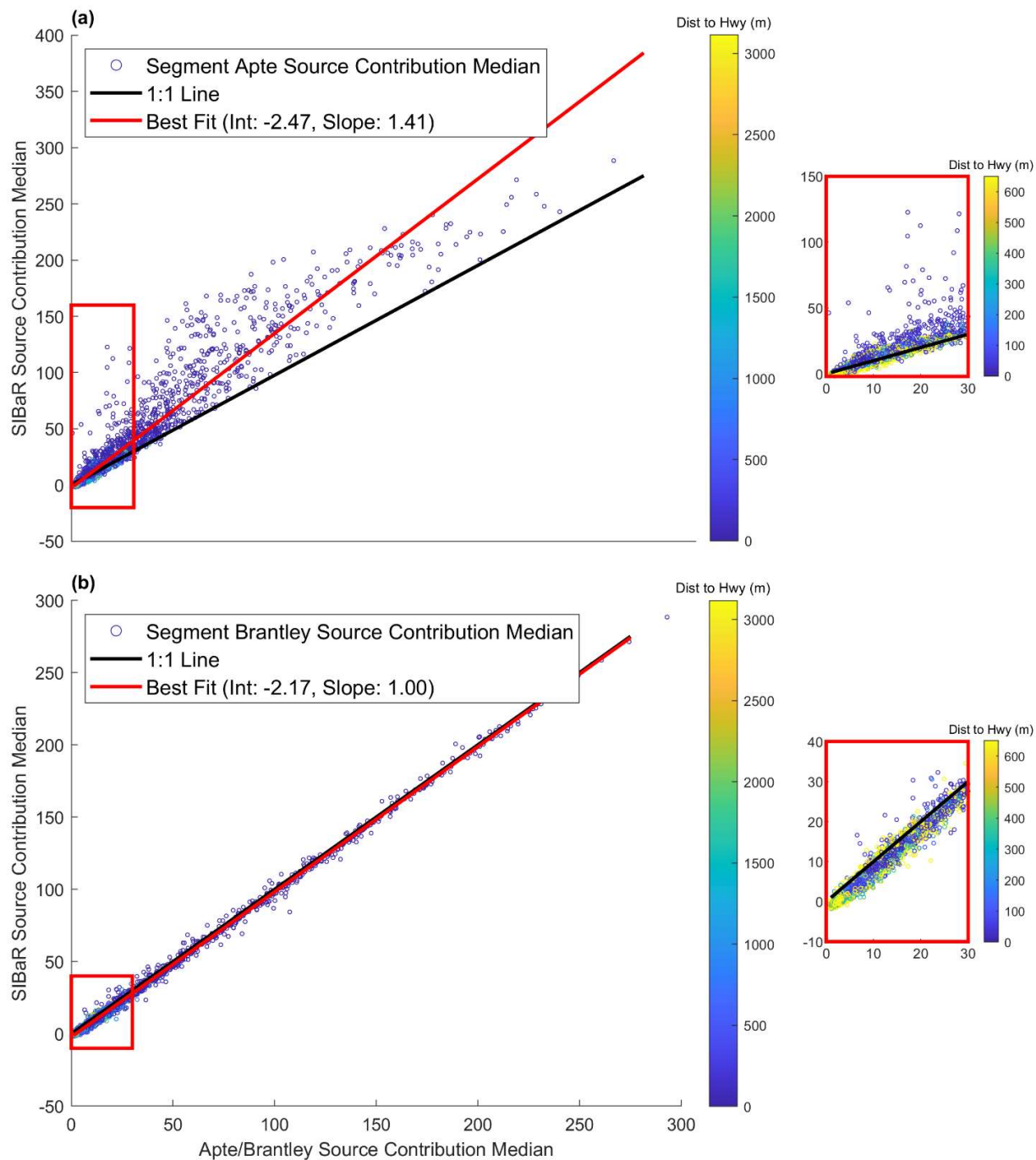


Figure 6. Scatterplots of road segment median source contributions predicted by two different techniques against their corresponding SIBaR median source contributions for NO_x . The line of best fit is derived using OLS regression and is depicted in red. The 1:1 line is depicted in black. Points are colored by their distance to the closest highway. (a) SIBaR source contribution medians plotted against Apte source contribution medians. (b) SIBaR source contribution medians plotted against Brantley source contribution medians. The plots in red rectangles designate a blown-up portion near the origin.

While the road segment median source contributions between the Brantley and SIBaR techniques exhibit strong agreement, we note that source contributions evaluated on a more granular level exhibit some disagreement. Figure 7 displays the inter quartile ranges (IQR) for source contributions assigned to each road segment plotted against each other for the SIBaR and Brantley techniques, again colored by distance to the closest highway. We display additional 1:1 plots of the IQR for different techniques and pollutants (NO_x and CO_2) in the supplement (Figures S18-S20). There are noticeable deviations from the 1:1 line in IQR between SIBaR and the Brantley technique for both NO_x and CO_2 , suggesting that the two techniques do disagree with one another on individual source contribution drive pass means. Figure S21 displays a histogram of differences in drive pass means between the 2 techniques. While SIBaR predicts lower source contributions compared to the Brantley technique on average, there are noticeable discrepancies captured in the tails of the distribution.

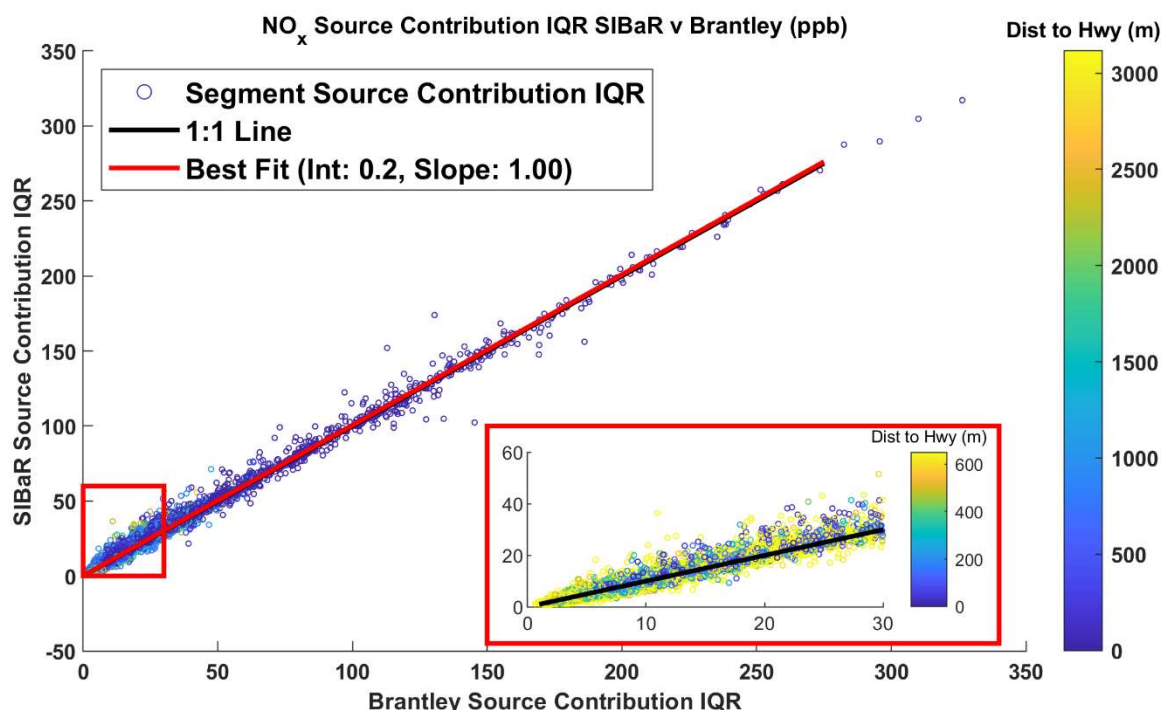


Figure 7. 1:1 scatterplot of the inter quartile range (IQR) of predicted NO_x source contributions at individual road segments for the SIBaR and Brantley techniques. The line of best fit is derived using OLS regression and is depicted in red. The 1:1 line is depicted in black. The inset, outlined by the red rectangle, shows the IQR at lower values of the Brantley source contribution IQR. Deviations from the 1:1 line suggest that SIBaR captures source influences the Brantley method fails to detect, despite predicting lower source contributions on average and the excellent agreement in median source contribution.

To provide further context for these results, we present 2 examples of daily time series of each background technique's predictions in Figure 8. It is apparent that the Apte technique overfits to the data in both cases. The top panel shows an example of SIBaR's predictions offering an advantage over Brantley's: since SIBaR is fit to a subset of the data, it avoids overfitting in the early morning hours of the time series that the Brantley time series incorporates. Panel (a) illustrates why the cases in the

right tail of the histogram in S21 exist. In contrast, the bottom panel showcases the potential faults in using SIBaR predictions: since there are no background designated points at the beginning of this time series example, the spline fit wildly extrapolates, resulting in unrealistic predictions that are captured in the left tail of the histogram in Figure S21. Both panels illustrate why the medians of Brantley and SIBaR agree so well with one another, yet display IQRs that deviate from their 1:1 line. Both
305 signals exhibit strong agreement with one another, but can capture different source influences periodically because of the assumptions inherent in each technique. It is also evident that the appropriate background fit would need to be investigated on a case-by-case basis, as one should avoid using the SIBaR technique in instances where extrapolation could occur.

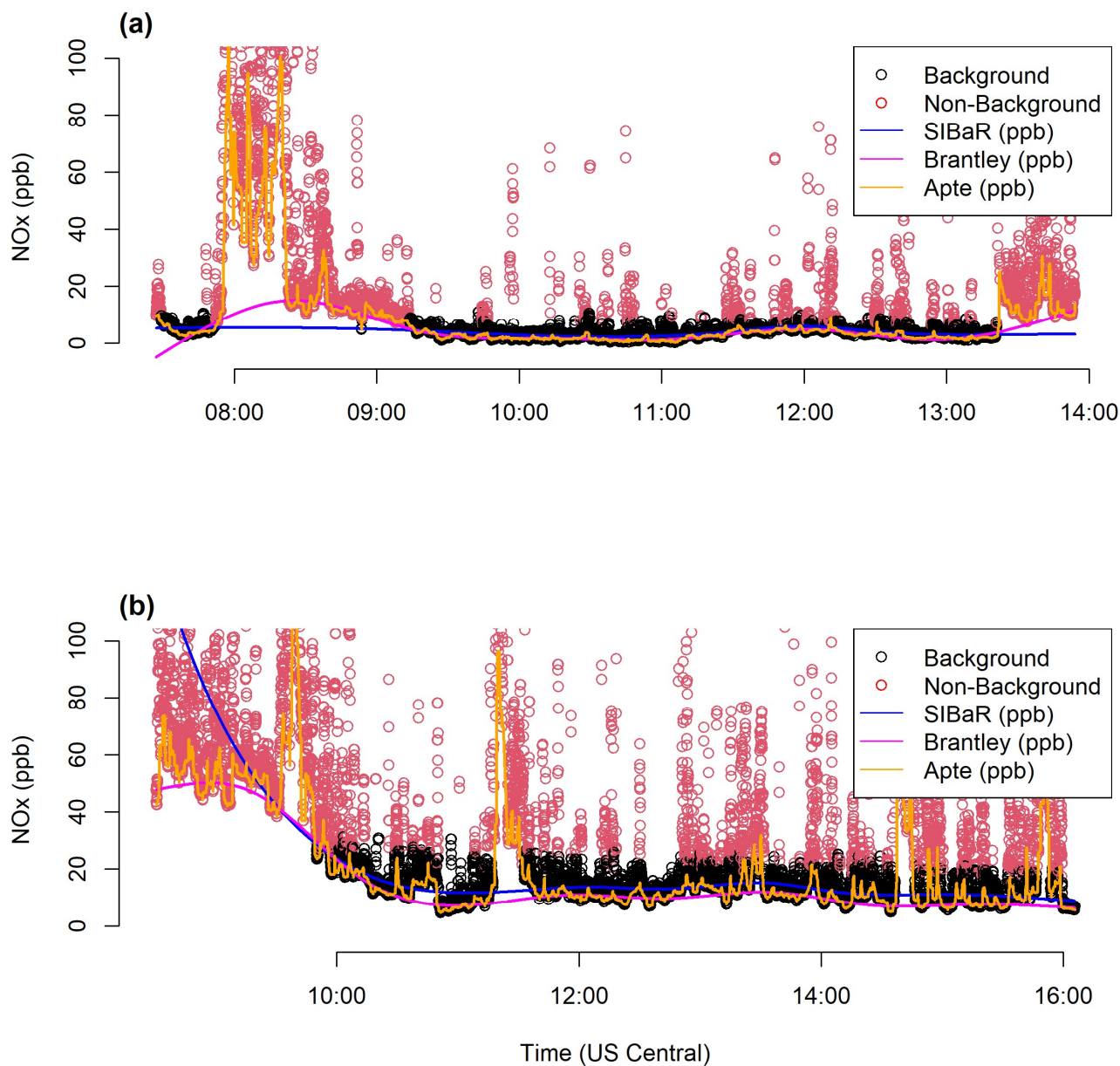


Figure 8. Time series plots which depict the original mobile campaign measurements, colored by their SIBaR decoded states (background and source), along with the background signals generated by the SIBaR, Brantley, and Apte techniques. (a) NO_x time series of mobile measurements taken on 10/3/2017 which displays the Apte and Brantley signals overfitting to data decoded as source by the SIBaR partitioning step. (b) NO_x time series of mobile measurements taken on 11/30/2017 which shows wildly extrapolated SIBaR predictions at the beginning of the time series due to the lack of background decoded states.

5 Concluding Remarks

We illustrate that SIBaR provides a defensible mechanism to quantify and remove background from air pollution monitoring data time series. The method's partitioning step is able to match 83% of a study's previously published background/non-background designations. Mapped distributions of the partitioning step's decoded states show high levels of background state assignment in residential areas, with notable exceptions in hotspots published in a previous study. Finally, we show the impact using SIBaR can have on deriving source contributions in comparing it to the background signals predicted by other techniques. Most notably, SIBaR does not rely on a static time window assumption to determine source impacts, and instead relies on fitting to a subset of the data generated with a time series regime change modelling technique. Setting a static time window can have significant impact on the derived source contributions, as exhibited by the discrepancies between the Apte and SIBaR methods shown in Section 3.3. While the SIBaR and Brantley techniques produce similar source contribution medians to one another in the context of this campaign's measurements, both capture different source influences based on the assumptions inherent in each respective technique.

Despite SIBaR's rigor and advancements relative to previously published methods, our approach needs careful consideration and improvement. The method is sensitive to how data in the time series are distributed, and transforming the measurements can provide different results. For example, Figure 9 exhibits a side-by-side comparison of SIBaR state predictions for transformed (a) and non-transformed (b) NO_x data. The transformation in this instance results in portions of the measurements in the early morning period being classified as background, whereas none are designated as background in the non-transformed case. While we think data are more appropriately described in the lognormal regime (Seinfeld and Pandis, 2016), careful consideration of transformation is necessary. Additionally, as discussed in Section 3.1 and exhibited in Figure S4, applying a smoothing time window can also affect the state categorizations.

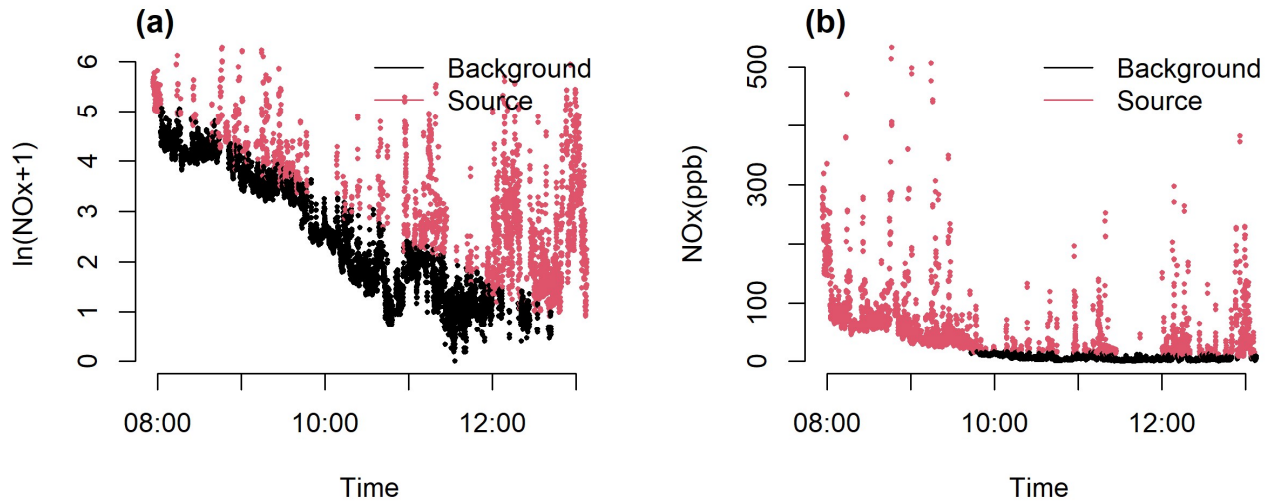


Figure 9. Comparison of SIBaR state designations for (a) log-transformed versus (b) non-transformed NO_x data on 10/30/2017 (Local Time, US Central) in the Houston mobile monitoring campaign. Transformation can affect state assignments, which in this case results in 38% of observations having a different categorization upon transformation.

While the linearity assumption in the time covariate is computationally cheap and easy to implement, it is limited. It is unrealistic to expect background air pollution to exhibit linear behavior, especially as time series duration extends (Luke et al., 2010). While the linearity assumption seems to be acceptable for time series of several hours of data, problems with that assumption arose in this work and will most likely arise on time series of data by day or when time series are impacted by abrupt meteorological changes. Future work should incorporate assumptions of non-linear behavior into analysis. Several studies have been published showing the applicability HMMs to covariates expressed as splines (Langrock et al., 2015, 2018). However, trade-offs between computational time and precision would need to be considered. In its current version, SIBaR takes ~6.5 hours to model background for millions of data points (performing the partitioning step, evaluating and/or correcting the fit, and fitting the spline for all time series). The Brantley technique, in contrast, takes several minutes.

Despite these shortcomings, SIBaR holds promise as a framework to quantify and remove background from air pollution monitoring time series. In its current state, it is inferior to the Brantley technique in regards to computation time. However, these problems with SIBaR are computational ones rather than problems with its underlying theory. The SIBaR partitioning step captures transient behavior between background and non-background quite well, as the diagnostic results of Section 3.1 and the maps in Section 3.2 indicate. In addition to addressing other issues highlighted here, future work should focus on methods to reduce its computational time to make its use more straightforward.

Code and Data Availability. Both the code and data are available on request. Additionally, time series comparisons for all 312 time series taken in the campaign, as well as a demo of the SIBaR partitioning step, are available here: <https://doi.org/10.5281/zenodo.5022590> (Actkinson, 2021). Data are also free to download on the Environmental Defense Fund's Air Quality Data commons (<https://aqdatacommons.org/>, Environmental Defense Fund, 2021).

Author Contributions. BA developed, wrote, and tested the method in R (R: The R Project for Statistical Computing, 2021) with critical input and scientific guidance from RG and KE. RG supervised the project and provided feedback on significance of method's results. BA wrote the manuscript. All authors contributed to the editing and review of this manuscript.

Competing interests. The authors declare that there is no conflict of interest.

Acknowledgements. The authors gratefully acknowledge the support of NIEHS (grant #R01ES028819-01). We thank Halley Brantley for the provision of data and comments concerning results in Section 3.1 of the manuscript. We also appreciate support from the Environmental Defense Fund for the collection and provision of mobile data.

References

Actkinson, B., Ensor, K., and Griffin, R.J.: Time Series Comparisons, Model Code, and a Demo Dataset for SIBaR: A New Method for Background Quantification and Removal from Mobile Air Pollution Measurements, Zenodo, doi:10.5281/zenodo.5022590, 2021

Apte, J. S., Messier, K. P., Gani, S., Brauer, M., Kirchstetter, T. W., Lunden, M. M., Marshall, J. D., Portier, C. J., Vermeulen, R. C. H., and Hamburg, S. P.: High-Resolution Air Pollution Mapping with Google Street View Cars: Exploiting Big Data, Environ. Sci. Technol., 51, 6999–7008, <https://doi.org/10.1021/acs.est.7b00891>, 2017.

Baldwin, N., Gilani, O., Raja, S., Batterman, S., Ganguly, R., Hopke, P., Berrocal, V., Robins, T., and Hoogterp, S.: Factors affecting pollutant concentrations in the near-road environment, Atmos. Environ., 115, 223–235, <https://doi.org/10.1016/j.atmosenv.2015.05.024>, 2015.

Brantley, H. L., Hagler, G. S. W., Kimbrough, E. S., Williams, R. W., Mukerjee, S., and Neas, L. M.: Mobile air monitoring data-processing strategies and effects on spatial air pollution trends, Atmospheric Meas. Tech., 7, 2169–2183, <https://doi.org/10.5194/amt-7-2169-2014>, 2014.

Brantley, H. L., Hagler, G. S. W., Herndon, S. C., Massoli, P., Bergin, M. H., and Russell, A. G.: Characterization of Spatial Air Pollution Patterns Near a Large Railyard Area in Atlanta, Georgia, Int. J. Environ. Res. Public. Health, 16, 535, <https://doi.org/10.3390/ijerph16040535>, 2019.

Bukowiecki, N., Dommen, J., Prévôt, A. S. H., Richter, R., Weingartner, E., and Baltensperger, U.: A mobile pollutant measurement laboratory—measuring gas phase and aerosol ambient concentrations with high spatial and temporal resolution, Atmos. Environ., 36, 5569–5579, [https://doi.org/10.1016/S1352-2310\(02\)00694-5](https://doi.org/10.1016/S1352-2310(02)00694-5), 2002.

- 385 Caplin, A., Ghandehari, M., Lim, C., Glimcher, P., and Thurston, G.: Advancing environmental exposure assessment science to benefit society, *Nat. Commun.*, 10, 1–11, <https://doi.org/10.1038/s41467-019-09155-4>, 2019.
- Chambliss, S. E., Preble, C. V., Caubel, J. J., Cados, T., Messier, K. P., Alvarez, R. A., LaFranchi, B., Lunden, M., Marshall, J. D., Szpiro, A. A., Kirchstetter, T. W., and Apte, J. S.: Comparison of Mobile and Fixed-Site Black Carbon Measurements for High-Resolution Urban Pollution Mapping, *Environ. Sci. Technol.*, 54, 7848–7857, <https://doi.org/10.1021/acs.est.0c01409>, 2020.
- 390 Chatzis, S. and Varvarigou, T.: A Robust to Outliers Hidden Markov Model with Application in Text-Dependent Speaker Identification, in: 2007 IEEE International Conference on Signal Processing and Communications, 2007 IEEE International Conference on Signal Processing and Communications, 804–807, <https://doi.org/10.1109/ICSPC.2007.4728441>, 2007.
- Chatzis, S. P., Kosmopoulos, D. I., and Varvarigou, T. A.: Robust Sequential Data Modeling Using an Outlier Tolerant Hidden Markov Model, *IEEE Trans. Pattern Anal. Mach. Intell.*, 31, 1657–1669, <https://doi.org/10.1109/TPAMI.2008.215>, 2009.
- 395 Dempster, A. P., Laird, N. M., and Rubin, D. B.: Maximum Likelihood from Incomplete Data Via the EM Algorithm, *J. R. Stat. Soc. Ser. B Methodol.*, 39, 1–22, <https://doi.org/10.1111/j.2517-6161.1977.tb01600.x>, 1977.
- Forney, G. D.: The viterbi algorithm, *Proc. IEEE*, 61, 268–278, <https://doi.org/10.1109/PROC.1973.9030>, 1973.
- Gómez-Losada, Á., Pires, J. C. M., and Pino-Mejías, R.: Characterization of background air pollution exposure in urban environments using a metric based on Hidden Markov Models, *Atmos. Environ.*, 127, 255–261, <https://doi.org/10.1016/j.atmosenv.2015.12.046>, 2016.
- 400 Gómez-Losada, Á., Pires, J. C. M., and Pino-Mejías, R.: Modelling background air pollution exposure in urban environments: Implications for epidemiological research, *Environ. Model. Softw.*, 106, 13–21, <https://doi.org/10.1016/j.envsoft.2018.02.011>, 2018.
- Gómez-Losada, Á., Santos, F. M., Gibert, K., and Pires, J. C. M.: A data science approach for spatiotemporal modelling of low and resident air pollution in Madrid (Spain): Implications for epidemiological studies, *Comput. Environ. Urban Syst.*, 75, 1–11, <https://doi.org/10.1016/j.compenvurbysys.2018.12.005>, 2019.
- 405 Hankey, S. and Marshall, J. D.: Land Use Regression Models of On-Road Particulate Air Pollution (Particle Number, Black Carbon, PM_{2.5}, Particle Size) Using Mobile Monitoring, *Environ. Sci. Technol.*, 49, 9194–9202, <https://doi.org/10.1021/acs.est.5b01209>, 2015.
- Hankey, S., Sforza, P., and Pierson, M.: Using Mobile Monitoring to Develop Hourly Empirical Models of Particulate Air Pollution in a Rural Appalachian Community, *Environ. Sci. Technol.*, 53, 4305–4315, <https://doi.org/10.1021/acs.est.8b05249>, 2019.
- Hudda, N., Gould, T., Hartin, K., Larson, T. V., and Fruin, S. A.: Emissions from an International Airport Increase Particle Number Concentrations 4-fold at 10 km Downwind, *Environ. Sci. Technol.*, 48, 6628–6635, <https://doi.org/10.1021/es5001566>, 2014.
- 415 Karner, A. A., Eisinger, D. S., and Niemeier, D. A.: Near-Roadway Air Quality: Synthesizing the Findings from Real-World Data, *Environ. Sci. Technol.*, 44, 5334–5344, <https://doi.org/10.1021/es100008x>, 2010.
- Langrock, R., Kneib, T., Sohn, A., and DeRuiter, S. L.: Nonparametric inference in hidden Markov models using P-splines, *Biometrics*, 71, 520–528, <https://doi.org/10.1111/biom.12282>, 2015.

- 420 Langrock, R., Adam, T., Leos-Barajas, V., Mews, S., Miller, D. L., and Papastamatiou, Y. P.: Spline-based nonparametric inference in general state-switching models, *Stat. Neerlandica*, 72, 179–200, <https://doi.org/10.1111/stan.12133>, 2018.
- Larson, T., Gould, T., Riley, E. A., Austin, E., Fintzi, J., Sheppard, L., Yost, M., and Simpson, C.: Ambient air quality measurements from a continuously moving mobile platform: Estimation of area-wide, fuel-based, mobile source emission factors using absolute principal component scores, *Atmos. Environ.*, 152, 201–211, <https://doi.org/10.1016/j.atmosenv.2016.12.037>, 2017.
- 425 Li, H. Z., Gu, P., Ye, Q., Zimmerman, N., Robinson, E. S., Subramanian, R., Apte, J. S., Robinson, A. L., and Presto, A. A.: Spatially dense air pollutant sampling: Implications of spatial variability on the representativeness of stationary air pollutant monitors, *Atmospheric Environ. X*, 2, 100012, <https://doi.org/10.1016/j.aeaoa.2019.100012>, 2019.
- Luke, W. T., Kelley, P., Lefer, B. L., Flynn, J., Rappenglück, B., Leuchner, M., Dibb, J. E., Ziemba, L. D., Anderson, C. H., and Buhr, M.: Measurements of primary trace gases and NO_y composition in Houston, Texas, *Atmos. Environ.*, 44, 4068–4080, <https://doi.org/10.1016/j.atmosenv.2009.08.014>, 2010.
- 430 Messier, K. P., Chambliss, S. E., Gani, S., Alvarez, R., Brauer, M., Choi, J. J., Hamburg, S. P., Kerckhoffs, J., LaFranchi, B., Lunden, M. M., Marshall, J. D., Portier, C. J., Roy, A., Szpiro, A. A., Vermeulen, R. C. H., and Apte, J. S.: Mapping Air Pollution with Google Street View Cars: Efficient Approaches with Mobile Monitoring and Land Use Regression, *Environ. Sci. Technol.*, 52, 12563–12572, <https://doi.org/10.1021/acs.est.8b03395>, 2018.
- 435 Miller, D. J., Actkinson, B., Padilla, L., Griffin, R. J., Moore, K., Lewis, P. G. T., Gardner-Frolick, R., Craft, E., Portier, C. J., Hamburg, S. P., and Alvarez, R. A.: Characterizing Elevated Urban Air Pollutant Spatial Patterns with Mobile Monitoring in Houston, Texas, *Environ. Sci. Technol.*, 54, 2133–2142, <https://doi.org/10.1021/acs.est.9b05523>, 2020.
- Patton, A. P., Perkins, J., Zamore, W., Levy, J. I., Brugge, D., and Durant, J. L.: Spatial and temporal differences in traffic-related air pollution in three urban neighborhoods near an interstate highway, *Atmos. Environ.*, 99, 309–321, <https://doi.org/10.1016/j.atmosenv.2014.09.072>, 2014.
- 440 R: The R Project for Statistical Computing: <https://www.r-project.org/>, last access: 22 June 2021.
- Robinson, E. S., Gu, P., Ye, Q., Li, H. Z., Shah, R. U., Apte, J. S., Robinson, A. L., and Presto, A. A.: Restaurant Impacts on Outdoor Air Quality: Elevated Organic Aerosol Mass from Restaurant Cooking with Neighborhood-Scale Plume Extents, *Environ. Sci. Technol.*, 52, 9285–9294, <https://doi.org/10.1021/acs.est.8b02654>, 2018.
- 445 Seinfeld, J. H. and Pandis, S. N.: *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change*, John Wiley & Sons, 1146 pp., 2016.
- Shairsingh, K. K., Jeong, C.-H., Wang, J. M., and Evans, G. J.: Characterizing the spatial variability of local and background concentration signals for air pollution at the neighbourhood scale, *Atmos. Environ.*, 183, 57–68, <https://doi.org/10.1016/j.atmosenv.2018.04.010>, 2018.
- 450 Svensén, M. and Bishop, C. M.: Robust Bayesian mixture modelling, *Neurocomputing*, 64, 235–252, <https://doi.org/10.1016/j.neucom.2004.11.018>, 2005.
- Tessum, M. W., Larson, T., Gould, T. R., Simpson, C. D., Yost, M. G., and Vedal, S.: Mobile and Fixed-Site Measurements To Identify Spatial Distributions of Traffic-Related Pollution Sources in Los Angeles, *Environ. Sci. Technol.*, <https://doi.org/10.1021/acs.est.7b04889>, 2018.
- 455

Van den Bossche, J., Peters, J., Verwaeren, J., Botteldooren, D., Theunis, J., and De Baets, B.: Mobile monitoring for mapping spatial variation in urban air quality: Development and validation of a methodology based on an extensive dataset, *Atmos. Environ.*, 105, 148–161, <https://doi.org/10.1016/j.atmosenv.2015.01.017>, 2015.

460 Visser, I. and Speekenbrink, M.: depmixS4: An R Package for Hidden Markov Models, *J. Stat. Softw.*, 36, 1–21, <https://doi.org/10.18637/jss.v036.i07>, 2010.

Zhang, X., Chen, X., and Zhang, X.: The impact of exposure to air pollution on cognitive performance, *Proc. Natl. Acad. Sci.*, 115, 9193–9197, <https://doi.org/10.1073/pnas.1809474115>, 2018.

465