Leveraging Street View and Remote Sensing Imagery to Enhance Air Quality Modeling through Computer Vision and Machine Learning

Meng Qi

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
In
Planning, Governance, and Globalization

Steve Hankey, Chair
Theodore Lim
Linsey Marr
Wenwen Zhang

January 24, 2024
Blacksburg, Virginia

Keywords: air pollution, artificial intelligence, image sampling and processing, satellite, land use regression, exposure assessment, open-source, big geodata

Leveraging Street View and Remote Sensing Imagery to Enhance Air Quality Modeling through Computer Vision and Machine Learning

Meng Qi

ABSTRACT

Air pollution is associated with various adverse health impacts and is identified as one of the leading risk factors for global disease burden. Further, air pollution is one of the pathways through which climate change could negatively impact health. Field studies have shown that air pollution has high spatiotemporal variability and pollutant concentrations vary substantially within neighborhoods. Characterizing air pollution at a fine-grained level is essential for accurately estimating human exposure, assessing its impact to human health, and further aiding localized air pollution policy. Air quality models are developed to estimate air pollution at locations and time periods without monitors, and these estimates are commonly used for exposure and health effects studies. Traditional land use regression [LUR] models are one of the cost-effective empirical air quality models. LUR typically relies on fixed-site measurements, GIS-derived variables with limited spatial resolution, and captures linear relationships. In recent years, innovative open-source imagery datasets and their associated features (e.g., street view imagery, remote sensing imagery) have emerged and show potential to augment or replace traditional LUR predictors. Such imagery data sources embody abundant information of natural and built environment features. Advanced computer vision techniques enable feature extraction and quantification through these extensive imagery datasets. The overarching objective of this dissertation is to investigate the feasibility of leveraging open-source imagery datasets (i.e., Google Street View [GSV] imagery, Landsat imagery, etc.) and advanced machine learning algorithms to develop image-based empirical air quality models at both local and national scale. The first

study of this work established a pipeline of feature extraction through street view imagery sematic segmentation. The resulting street view features were used to predict street-level particulate air pollution for a single city. The results showed that solely using GSV-derived features can achieve comparable model fits as using traditional GIS-derived variables. Feature engineering improved model stability and interpretability through reducing spurious variables from potential misclassifications from computer vision algorithms. The second study further developed GSV-based models at national scale across multiple years. Random forest models were developed to capture the nonlinear relationship between air pollution and its impacting factors. The results showed that with sufficient street view images, GSV imagery alone may explain the variation of long-term national NO_2 concentrations. Adding satellite-derived aerosol estimates (i.e., OMI column density) can significantly boost model performance when GSV images are insufficient, but the addition narrows when more GSV images are available. Our systematic assessment of the impact of image availability on model performance suggested that a parsimonious image sampling strategy (i.e., one GSV image per 100m grid) may be sufficient and most cost-effective for model development and application. Our third study explored the feasibility of combining street view and remote sensing derived features for national NO_2 and $PM_{2.5}$ modeling and projection at high spatial resolution. We found that GSV-based models captured both the highest and lowest pollutant concentrations while remote sensing features tended to smooth the air pollution variations. The results suggested that GSV features may have the capability to better capture fine-scale air pollution variability. The resulting air pollution prediction product may serve a variety of applications, including providing new insights into environmental justice and epidemiological studies due to its high spatial resolution (i.e., street level).

Collectively, the result of this dissertation suggests that GSV imagery, processed with computer vision techniques, is a promising data source to develop empirical air quality models with high spatial resolution and consistent predictor variables processing protocol. Image-based features assisted with advanced ML approaches have the potential to greatly improve air quality modeling estimates, and successfully show comparable and even superior model performance than other modeling studies. Moreover, the ever-growing public imagery data sources are particularly promising for remote or less developed areas where traditional curated geodatabases are sparse or nonexistent.

Leveraging Street View and Remote Sensing Imagery to Enhance Air Quality Modeling through Computer Vision and Machine Learning

Meng Qi

GENERAL AUDIENCE ABSTRACT

Air pollution is detrimental to human health and well-being. Further, air pollutants concentrations can change rapidly within a short distance and temporal frame. Monitoring air pollution with high spatial-temporal resolution is important. Traditional air quality monitoring networks are expensive and sparsely distributed, leading to gaps in capturing the air pollution at small spatial scales. Air quality models are developed to estimate air pollution at locations and time periods without monitors. Empirical air quality models often use air measurements from stationary sites and GIS-derived features (e.g., traffic, population density, land use types, etc.) to develop regression models and use the regression formula to estimate air pollutant concentrations in unmonitored areas. However, GIS-derived features are often collected from curated GIS databases, which often have coarse resolution when available across large geography. Street view imagery and remote sensing imagery contains rich information of natural and built environments. Computer vision techniques can be applied to extract such information to replace or augment traditional GIS-derived features. Combined with advanced machine learning algorithms, features derived from open-access images are promising to develop air quality models with a consistent image collection and processing protocol. This dissertation examines the feasibility of using street view imagery (i.e., Google Street View [GSV] Imagery) and remote sensing imagery to develop air quality models at both local and national scales. Our results found that solely using GSV features to build local and national models can achieve good model performance, which is consistent or even better than other models using traditional GIS-

derived variables. For areas without sufficient GSV images, adding satellite observations for air pollution can significantly enhance model performance. Remote sensing features tend to smooth air pollution variation while GSV features tend to better capture fine-scale intra-urban air pollution variation. In conclusion, leveraging open-source imagery datasets with advanced machine learning methods are promising for estimating air pollution at high spatial resolution with good model fits.

ACKNOWLEDGEMENT

It feels like just yesterday when I arrived at the Blacksburg campus for the first time, ready to embark on my PhD. Along this journey, there are numerous moments of joy, excitement, and time of doubt and hesitation. Huge thanks to everyone I met during these years. Without your support, I would not have completed this incredible journey.

My deepest appreciation goes to my advisor Dr. Steve Hankey. Thank you for your openness and respect for our individual research interests. Your guidance led me into the field of image-based air quality modeling, granting me the freedom to explore research topics that I'm excited about, and help me grow into an independent researcher. Thank you for always encouraging me, supporting me, and guiding me for both in my academic pursuits and personal growth.

I would also like to express my gratitude to my committee members, Dr. Wenwen Zhang, Dr. Linsey Marr, and Dr. Theodore Lim. Thanks for all your support and insights into my research. I'm grateful to Dr. Wenwen Zhang for clarifying my confusions on data analytics and machine learning. Your insightful guidance always consistently directed me towards the right direction. I'm grateful to Dr. Linsey Marr for the support in my air quality studies. I learned a lot from your air quality courses and the AIR group you organized. My gratitude also goes to Dr. Theodore Lim for your guidance on heat datasets. It was a fulfilling experience working as your TA/RA for a semester.

I would also like to thank my collaborators Dr. Julian Marshall, Dr. Chunxue Xu, Dr. Matthias Demuzere, Dr. Perry Hystad, Dr. Tianjun Lu, Dr. Peter James, Dr. Benjamin Bechtel, Dr. Esra Suel, Dr. Debby Lin and Dr. Andrew Larkin for their valuable suggestions and comments on my research.

I’m grateful to be a member of the PGG program, a warm place where I met wonderful faculty, staff and fellow students. In particular, I’d like to thank Dr. Yang Zhang, Dr. Todd Schenk, Dr. Ralph Buehler, Kelly, Han, Matthew, Heather, Leslie, and Bradley. My heartful thanks also go out to Luke, Qiong, Huyen, Ayda, Kuldeep, Sajad, Kaidi, Ruixiang, Griffin, and other UAP members. Special thanks to my friends, Jin, Xinyue, Huan, Minzhe, Kaisen, Ran, Qishen, Chenyang, and many others. We had lots of joyful time, and you are my second family away from home.

Lastly, I would like to thank my very best friend and husband, Wei Wang. You always support every decision I make, applaud for every achievement I attain, and encourage me whenever I meet challenges. We’re blessed to have our precious little one, and I’m grateful for the boundless joy he brings. I could not thank my parents enough, who always selflessly support me and have faith in me. You are the wellspring of my courage for pursuing my dream and goals.

Table of Contents

ABSTRACT **2**

GENERAL AUDIENCE ABSTRACT **5**

ACKNOWLEDGEMENT **7**

Chapter 1 Introduction **12**

1.1 Background and Motivation 19

1.2 Objectives 23

1.3 Dissertation structure 23

1.4 References 25

Chapter 2 Using Street View Imagery to Predict Street-level Particulate Air Pollution **31**

2.1 Introduction 33

2.2 Materials and methods 35

2.2.1 Mobile monitoring data 35

2.2.2 Street-level predictor variables 36

2.2.3 Modeling approach 37

2.2.4 Feature selection 38

2.3 Results and discussion 40

2.3.1 Spatial distribution of street-level predictor features 40

2.3.2 Model results 42

2.3.3 Comparison among models 46

2.3.4 Results of spatial autocorrelation analysis and model uncertainty 48

2.4 Limitations and implications 52

2.5 Supporting Information 54

2.5.1 Description of data-driven feature selection 54

2.5.2 Reduced buffer-feature GSV-only models 55

2.6 References 82

Chapter 3 National Land Use Regression Model for NO_2 using Street View Imagery and Satellite Observations 91

3.1 Introduction 93

3.2 Materials and methods 95

3.2.1 Modeling variables 95

3.2.2 Modeling approach 97

3.2.3 Model evaluation 98

3.2.4 Sensitivity analysis for image sampling strategies 98

3.3 Results and discussion 99

3.3.1 Summary of model inputs 99

3.3.2 Model results of image-unlimited models 101

3.3.3 Sensitivity analysis using image-limited models 106

3.3.4 Model prediction and comparison 110

3.4 Limitations and implications for future studies 113

3.5 Supporting Information 115

3.5.1 Description of spatial and temporal cross validation 115

3.5.2 Results of spatial and temporal cross validation 116

3.6 References 138

Chapter 4 Assessment of Street View and Remote Sensing Imagery derived parameters in estimating national NO_2 and $PM_{2.5}$ pollution 147

4.1 Introduction 149

4.2 Materials and methods 151

4.2.1 Modeling variables 151

4.2.2 Modeling approach 153

4.2.3 Model evaluation 154

4.3 Results and discussion 154

4.3.1 Description of Model Inputs 154

4.3.2 Model Performance 156

4.3.3 Model prediction 161

4.4 Conclusion 166

4.5 Supporting Information 168

4.6 References 184

Chapter 5 Conclusion 190

5.1 Key findings 190

5.2 Limitations and Implications for future studies 193

List of Figures

Figure 2-1. Image sampling and feature extraction procedure. (A) Google Street View (GSV) images sampled around one air pollutant monitoring aggregation location. (B) GSV images retrieved for 4 directions at one GSV sampling location. (C) Feature extraction using PSPNet scene parsing. (D) Estimated feature percentage among the 150 segmented classes. 37

Figure 2-2. Comparison of model performance using theory-driven feature selection. Models developed with different categories of predictor variables are shown by color. Solid columns represent model adjusted R^2; hashed columns represent 10-fold cross validation R^2..... 44

Figure 2-3. Model performance using integrated theory- and data-driven feature selection. Each box represents 15 scenarios for different categories of predictor variables. Models developed with different categories of predictor variables are shown by color. Hollow boxes represent model adjusted R^2; hashed boxes represent 10-fold cross validation R^2. 46

Figure 2-4. Estimated BC and PN concentrations using 4 types of GSV-only models. Only grid centroids with GSV images within 50m are displayed. The scatter plots in the lower-left corner show model predicted air pollutant concentrations vs. mobile monitoring air pollutant concentrations. The dashed red line is the best fit line; the dashed black line is the 1:1 line. Due to the highly right-skewed distributions of model estimates (**Figure 2-S7**), we adjusted the color bar scale to visualize the spatial patterns of the models. Specifically, for each air pollutant, we binned the estimates among the 4 models and used the 95th percentile as the maximum for visualization..... 51

Figure 2-S1. Histogram of the mean of 1,350 predictor variables. The values were log-transformed to determine the magnitude of thresholds for data filtering…….. 71

Figure 2-S2. Histograms of the multiplicative background adjusted BC and PN concentrations. 72

Figure 2-S3. Spatial distribution of the sampled GSV images and examples of built and natural environment characteristics derived from the images. (a) GSV images sampled, (b) road distribution, and (c) tree cover distribution. 73

Figure 2-S4. Comparison of central-feature models with different distance ranges. The predictor variables for both models were "Theory-driven: Air pollution-related variables". 74

Figure 2-S5. Model performance of buffer-feature models among different scenarios using data-driven feature selection. The upper panel shows adjusted R^2 of the models and the bottom panel shows 10-fold cross validation R^2. Analogous plots for the central-feature models are in **Figure 2-S6**. 75

Figure 2-S6. Model performance of the central-feature models among different scenarios using the data-driven feature selection. The upper panel shows adjusted R^2 of the models and the bottom panel shows 10-fold cross validation R^2. 76

Figure 2-S7. Density plots of estimated BC and PN concentrations using 4 types of GSV-only models. All distributions were similar and highly positive-skewed..... 77

Figure 2-S8. Comparison among GSV-only model estimates and previous LUR model estimates. Figures from the left to right columns were estimates using reduced buffer-

feature models with the additive method, estimates using reduced models with the simplified method, and estimates using previous LUR models with traditional predictor variables. Due to highly right-skewed distributions of the predicted concentrations, the color bar was scaled from minimum to the 95th percentile to better visualize spatial patterns of the data. 78

Figure 2-S9. The spatial distribution of mobile monitoring measurements and the results of LISA test for GSV-only model residuals. We used the integrated buffer-feature models with air pollution-related variables to illustrate the results of the LISA analysis (500m threshold). High-High and Low-Low cluster represent statistically significant clusters of underestimates and overestimates of particulate air pollution – these clusters are the most important for the models presented in this paper. A High-Low (Low-High) outlier are where a high (low) residual value is surrounded primarily by low (high) residual values. 79

Figure 2-S10. Comparison of BC model uncertainty for the GSV-only model presented in this paper and the previously published LUR model with traditional GIS predictor variables. We used the integrated buffer-feature model with air pollution-related variables as an example. The upper panel shows model fits; the bottom panel shows standard error for model prediction. The model fits are transformed back to original values for illustration. 80

Figure 2-S11. Comparison of PN model uncertainty for the GSV-only model presented in this paper and the previously published LUR model with traditional GIS predictor variables. We used the integrated buffer-feature model with air pollution-related variables as an example. The upper panel shows model fits; the bottom panel shows standard error for model prediction. The model fits are transformed back to original values for illustration. 81

Figure 3-1. NO_2 measurement and GSV locations for the national NO_2 models. (A) Spatial distribution of NO_2 EPA monitors and GSV images within 500 m of monitors. (B) Example of GSV features extracted via image semantic segmentation. (C) Illustration of spatiotemporal variation in GSV image availability for one EPA monitor..... 98

Figure 3-2. Comparison of random 10-fold cross validation R^2 between GSV-only and GSV+OMI models. (A) shows the results of image-unlimited models while (B) shows the results of image-unlimited models with constrained training size. The training sizes of models in (A) and (B) are shown in the bar chart within (B). (A) also shows the feature importance score of the OMI variable in the GSV+OMI models among the 5 scenarios. Both (A) and (B) show results for 5 scenarios while subplots (C) to (F) only show results by scenario 5 from models in (A). (C) and (D) show hexagonal binned plots of model predictions versus observations. The color reflects the number of scatters within each hexagon. (E) and (F) show the rank of feature importance in the random forest models. Results of other scenarios are shown in **Figures 3-S5** and **3-S7**. 103

Figure 3-3. Results of sensitivity analyses among image sampling strategies. (A) and (B) show model results among all scenarios with varying maximum number of GSV images per 100 m × 100 m sampling grid. (C) and (D) show model results in scenario 5 with varying grid size. Each grid cell is allowed one GSV image for sampling. (E) and (F) show model predictions for Washington DC in 2019 using image-unlimited models and image-limited models with maximum one GSV image per 100 m × 100 m grid. 109

Figure 3-4. Comparison between GSV-based models and conventional national LUR models using prediction maps for Washington DC in 2015. (A) Model predictions for the GSV-only model, GSV+OMI model, and conventional national LUR model. For comparison, we masked GSV-based prediction maps based on where the conventional LUR predictions were available. (B) Predicted NO_2 concentrations among different models along the transect line. The solid lines represent actual model predictions; gaps in the conventional national LUR model (masked GSV-based models) were filled by interpolation (original GSV-based models) and represented by dotted lines. (C) Values of GSV feature groups along the transect line. 112

Figure 3-S1. Decreasing trend of annual NO_2 concentrations during 2007 to 2019 from the EPA regulatory network…….. 125

Figure 3-S2. Number of available NO_2-GSV pairs and GSV images in different years using scenario 1 and 5 for illustration. The left axis represents the number of GSV images found within 500 m of each NO_2 monitoring site in each year while the color of the stripes represents its frequency. Darker colors indicate more EPA sites with the same amount of GSV images. The right axis represents the number of NO_2-GSV pairs for model development in each year. 126

Figure 3-S3. Spatial distribution of NO_2-GSV pairs in different years using scenario 5 for illustration. The left column represents NO_2 monitors with sufficient GSV images for model development. The size and color of the stars represent the number of GSV images available within 500 m of NO_2 EPA monitors. The right column represents NO_2 monitors without sufficient GSV images that are excluded from model development. The size and color of the stars are fixed in the right column. 127

Figure 3-S4. Comparison of (A) root mean squared errors and (B) mean absolute errors between GSV-only and GSV+OMI image-unlimited models. 128

Figure 3-S5. Hexagonal binned plots of model predictions versus observations for both GSV and GSV+OMI models among different scenarios. The color reflects the number of scatters within each hexagon. 129

Figure 3-S6. Spatial distribution of NO_2 monitors with lowest and highest 5% errors (i.e., absolute error by GSV-only and GSV+OMI models using scenario 5). The left column represents the results of GSV-only model while the right column represents GSV+OMI model. The upper panel represents lowest 5% errors and the bottom panel represents the highest 5% errors. 130

Figure 3-S7. Rank of feature importance for GSV-only and GSV+OMI models among different scenarios. 131

Figure 3-S8. Distribution of the top 10 important features of GSV-only and GSV+OMI models among different scenarios. The numbers after the GSV variable names represent the buffer radii. 132

Figure 3-S9. Accumulated local effects (ALE) of several top important GSV features for the GSV-only model using scenario 5 for illustration. The suffix number of GSV features indicates the buffer level for tabulating the GSV features. The y axis shows how the features influence the prediction of GSV-only models on average. For instance, when the value of car_500 is ~0.02, the model prediction is higher by ~0.5. 133

Figure 3-S10. Comparison of three cross validation strategies (i.e., splitting training and testing set randomly, temporally, or spatially) for GSV-only and GSV+OMI models using scenario 5 for illustration. Bars with solid filling represent models using

250/500 m to tabulate GSV features. Bars with stripes represent models using 100/200/300/400/500 m buffer radii. 134

Figure 3-S11. Estimated NO_2 concentrations for Washington DC by GSV-only and GSV-OMI models using scenario 5. We display the prediction maps for 2007, 2011, 2015, and 2017 for illustration. 135

Figure 3-S12. Comparison between GSV-based models and conventional national LUR models using prediction maps for San Francisco in 2015. (A) Model predictions for the GSV-only model, GSV+OMI model, and conventional national LUR model. For comparison, we masked GSV-based prediction maps based on where the conventional LUR predictions were available. (B) Predicted NO_2 concentrations among different models along the transect line. The solid lines represent actual model predictions; gaps in the conventional national LUR model (masked GSV-based models) were filled by interpolation (original GSV-based models) and represented by dotted lines. (C) Values of GSV feature groups along the transect line. 136

Figure 3-S13. Spatial distribution of GSV feature groups for Washington DC in 2019. 137

Figure 4-1. Prediction versus observation plots and ranking of feature importance scores for the NO_2 and $PM_{2.5}$ full models. The left (right) panel shows results for NO_2 ($PM_{2.5}$) model. (A) Hexagonal binned plots of model predictions versus observations. The color reflects the number of scatters within each hexagon. (B) The ranking of feature importance scores based on aerosol estimates, aggregated GSV and Landsat feature groups (i.e., aggregation level 2). The inner plots show the ranking based on feature aggregation level 1. (C) Top 15 important features based on feature scores of aerosol estimates, individual GSV and Landsat features (aggregation level 3). The details of the aggregation levels are listed in **Table 4-S1**. Results of other models are shown in **Figures 4-S5 - 4-S7** 160

Figure 4-2. Predicted NO_2 and $PM_{2.5}$ air pollution by different image-based models at 100m spatial resolution using 2019 for illustration. The left panel shows results for NO_2 and the right panel shows results for $PM_{2.5}$. More details of sample cities are shown in **Figure 4-3**. 163

Figure 4-3. Spatial patterns of predicted annual NO_2 concentrations for four example cities (i.e., San Francisco-Oakland [SF-Oakland], Chicago, Washington, DC-Maryland-Virginia [DMV], and Houston) in 2019 by different image-based models. (A) to (F) represent GSV-only, GSV+OMI, Landsat-only, Landsat+OMI, and the full model, respectively. 164

Figure 4-4. Spatial patterns of predicted annual $PM_{2.5}$ concentrations for four example cities (i.e., SF-Oakland, Chicago, Washington, DMV, and Houston) in 2019 by different image-based models. (A) to (F) represent GSV-only, GSV+AOD, Landsat-only, Landsat+AOD, and the full model, respectively. 165

Figure 4-5. Statistical distribution of predicted NO_2 and $PM_{2.5}$ annual concentrations at urbanized area level by different image-based models using 2019 for illustration. The left panel shows results for NO_2, and the right panel shows results for $PM_{2.5}$. The first, second, and third rows represent the distribution of minimum, mean, and maximum annual concentrations, respectively. 166

Figure 4-S1. The number of $PM_{2.5}$ and NO_2 EPA monitors used or excluded for model development each year from 2007 to 2019. Bars with strips indicate monitors used for

modeling while solid bars indicate monitors excluded for modeling. The full bars are all available EPA monitors…… 175

Figure 4-S2. Spatial distribution of NO_2 and $PM_{2.5}$ EPA monitors during 2007-2019. Red stars represent EPA monitors used for model development while blue stars represent EPA monitors excluded for modeling. Monitors without GSV images within 250m were excluded for modeling due to an incomplete set of model independent features. 176

Figure 4-S3. Distribution of NO_2 and $PM_{2.5}$ EPA monitors by urban, suburban, and rural using 2019 for illustration. The urbanicity classification is based on the locale classification provided by the National Center for Education Statistics..... 177

Figure 4-S4. The decreasing trends of annual NO_2 and $PM_{2.5}$ concentrations during 2007-2019 from the EPA regulatory network. The blue line is the mean of annual concentrations from all modeled EPA monitors. The grey area represents the minimum and maximum annual concentrations for each year. 178

Figure 4-S5. Number of available NO_2-GSV pairs and GSV images in different years using scenario 1 and 5 for illustration. The left axis represents the number of GSV images found within 500m of each NO_2 monitoring site in each year while the color of the stripes represents its frequency. Darker colors indicate more EPA sites with the same amount of GSV images. The right axis represents the number of NO_2-GSV pairs for model development in each year. 179

Figure 4-S6. Ranking of feature importance scores for the NO_2 and $PM_{2.5}$ GSV-only models. (A) The ranking of feature importance scores based on aerosol estimates, aggregated GSV and Landsat feature groups (i.e., aggregation level 2). (C) Top 15 important features based on feature scores of aerosol estimates, individual GSV and Landsat features (aggregation level 3). The details of the aggregation levels are listed in **Table 4-S1**. 180

Figure 4-S7. Ranking of feature importance scores for the NO_2 GSV+OMI and $PM_{2.5}$ GSV+AOD models. (A) The ranking of feature importance scores based on aerosol estimates, aggregated GSV and Landsat feature groups (i.e., aggregation level 2). (C) Top 15 important features based on feature scores of aerosol estimates, individual GSV and Landsat features (aggregation level 3). The details of the aggregation levels are listed in **Table 4-S1**. 181

Figure 4-S8. Statistical distribution of predicted NO_2 and $PM_{2.5}$ annual concentrations at census tract level by different image-based models using 2019 for illustration. The left panel shows results for NO_2, and the right panel shows results for $PM_{2.5}$. The first, second, and third rows represent the distribution of minimum, mean, and maximum annual concentrations, respectively. 182

Figure 4-S9. Statistical distribution of predicted NO_2 and $PM_{2.5}$ annual concentrations at census block group level by different image-based models using 2019 for illustration. The left panel shows results for NO_2, and the right panel shows results for $PM_{2.5}$. The first, second, and third rows represent the distribution of minimum, mean, and maximum annual concentrations, respectively. 183

List of Tables

Table 2-1. Summary of model performance among all model types 50

Table 2-S1. Predictor variables in each category for the theory-driven feature selection……. 58

Table 2-S2. Descriptive statistics of features extracted from sampled GSV images within 2,000m of air quality monitoring locations 61

Table 2-S3. Model results for BC estimation using theory-driven feature selection .. 65

Table 2-S4. Model results for PN estimation using theory-driven feature selection .. 66

Table 2-S5. Number of input predictor variables offered to the buffer-feature models using different feature selection methods 67

Table 2-S6. Results of the best BC and PN buffer-feature models using data-driven feature selection 68

Table 2-S7. Summary of buffer-feature models and reduced buffer-feature models for air pollution estimation (all models used the integrated theory- and data-driven feature selection and the air pollution-related predictor variables)..... 69

Table 2-S8. Results of the Global Moran's I test for 4 types of GSV-only buffer-feature models 70

Table 3-1. Model performance of image-unlimited models……. 104

Table 3-S1. Predictor variables for model development……...118

Table 3-S2. Scenarios for GSV-NO_2 temporal match during model development. .119

Table 3-S3. Distribution of NO_2 EPA monitors by urban, suburban, rural using 2019 for illustration. The urbanicity classification is based on the locale classification provided by National Center for Education Statistics..... 120

Table 3-S4. Descriptive statistics of GSV features within 500 m of NO_2 monitors stratified by land use types. GSV features are unitless since they are extracted through image segmentation and represent pixel percentage by GSV variables. 121

Table 3-S5. Model performance of image-unlimited models using the same training size in 5 scenarios. 122

Table 3-S6. Model performance of image-limited models which used various maximum number of GSV images per 100 m × 100 m grid cell..... 123

Table 3-S7. Model performance of image-limited models using different grid sizes for sampling a maximum of one GSV image per grid cell. All models use scenario 5. .124

Table 4-1. Summary of model performance for models with different types of independent variables. Model evaluation metrics included random 10-fold R^2, mean absolute error (MAE), and root mean squared error (RMSE). 158

Table 4-S1. Feature aggregation levels for feature importance scores ranking…… 168

Table 4-S2. Descriptive statistics of GSV and Landsat-derived features within 250 m of NO_2 EPA monitors stratified by land use types. GSV features are applied with a scale of 10000. 170

Table 4-S3. Descriptive statistics of GSV and Landsat-derived features within 250 m of $PM_{2.5}$ EPA monitors stratified by land use types. GSV features are applied with a scale of 10000. 172

Table 4-S4. Summary of model performance for models with different types of independent variables using spatial hold-out cross validation. Model evaluation metrics included spatial hold-out 10-fold R^2, mean absolute error (MAE), and root mean squared error (RMSE). The difference of model fits by random and spatial hold-out CV was also shown in the table.....174

Chapter 1 Introduction

1.1 Background and Motivation

Exposure to air pollution has been associated with adverse health impacts by extensive epidemiological studies and is identified as one of the leading risk factors for global disease burden[1-6]. Further, air pollution is one of the pathways through which climate change could negatively impact health. According to the Global Burden of Disease Study, air pollution is estimated to cause 6.7 million premature deaths worldwide in 2019, of which ambient air pollution is responsible for 4.2 million[7]. In the United States, particulate matter exposure is the major risk factor among established air pollutants, responsible for 80% of deaths from air pollution causes and 3% of deaths from all causes[7]. Precise estimation of air pollution exposure and its impact on public health depends on the accuracy of monitored air pollutant concentrations and air pollution projection techniques. Regular air pollution monitoring has played an important role in air pollution regulation and exposure assessment. Although traditional fixed-site monitoring networks provide high-quality temporal-resolved data for air pollutant concentrations, they are insufficient to characterize air pollution at a fine spatial scale due to their sparse distribution. Furthermore, many field studies have shown that air pollution has high spatiotemporal variability and pollutant concentrations may vary rapidly within neighborhoods[8-10].

Air quality models are developed to estimate air pollution at locations and time periods without monitors, and these estimates are commonly used for exposure and health effects studies. Land Use Regression [LUR] is one of the methods that has been widely used as a cost-effective way to estimate fine-scale ambient air pollutant concentrations[11-16]. Compared to other air quality modeling approaches, LUR exhibits higher spatial resolution (neighborhood-scale) and thus is able to capture within-urban

variability[17]. LUR models are often applied to long-term air pollution studies (e.g., estimating annual mean concentrations), and the results have been widely used in exposure science and epidemiological studies[18-21]. Traditionally, LUR used air pollution data sampled from fixed sites and predictors obtained through geographic information systems [GIS]. Typical LUR predictors include various traffic representations, population density, land use, topography, morphography, and meteorology variables[15]. However, such predictors are often collected from curated GIS-derived databases, which may have been initially aimed to serve local administrative purposes[22]. The data availability and quality (e.g., spatial resolution, data format, etc.) may vary largely across jurisdictions. Generally, data which is available over large geographies often has less precision. For example, Larkin et al. developed a global NO_2 LUR model using satellite estimates and land use variables, of which the spatial resolution varies largely between 30 m to 10 km[23]. In recent years, innovative data sources and their associated features (e.g., street view imagery, remote sensing imagery, point of interest [POI]) have emerged to augment or replace traditional LUR predictors for empirical air quality modeling, and successfully shown comparable or superior model performance[24-27], although such research is in the nascent stage.

There are several general characteristics shared by the new data streams that are promising for air quality modeling. Firstly, information provided by traditional LUR predictors (e.g., land use, built environment) are also encoded in the new data sources. For example, advanced computer vision techniques can extract and quantify the encoded information in street view or remote sensing imagery for empirical modeling[28]. Secondly, these new variables are mostly open source and are promising for establishing a consistent data collection and processing protocol across large geography. This allows the easy scaling of the modeling approach for broader coverage.

Models can be well generalized across cities, regions, countries, or continents regardless of political boundaries. Thirdly, the new variables often have high spatial resolution and describe hyperlocal features of the built environment. For example, street view images can capture street-level attributes of the built and natural environment. Many large open-access street view imagery are available and growing rapidly, including Google Street View [GSV], Bing Maps Streetside, Apple Look Around, Tencent Maps, Daum Maps, CycloMedia, etc[29]. The application of satellite-based aerosol estimates (e.g., aerosol optical depth [AOD], Ozone Monitoring Instrument [OMI], TROPOspheric Monitoring Instrument [TROPOMI]) isn't new in air quality modeling[11,30-34], but more remote sensing products can be further leveraged with computer vision and advanced modeling algorithms. For example, Landsat, the longest continuous satellite program thus far, has a 30 m resolution which is considered medium resolution but is free of charge for use. Some commercial satellites may provide very high resolution (e.g., 30 cm) but the cost could be quite expensive (some commercial satellites may provide free use for education users).

The feasibility of new data sources and more advanced modeling approaches is largely facilitated by the ground-breaking advancements in artificial intelligence. Several studies have explored the application of the new data sources and the resulting variables in building empirical air quality models, most of which involve the application of computer vision or machine learning models[24-27,35-38]. For example, Zheng et al. first applied a convolutional neural network [CNN] to extract features from a high spatial-resolved micro-satellite imagery (3 m spatial resolution), and then applied a random forest [RF] model combining the satellite features with meteorological features to estimate ground-level $PM_{2.5}$ concentrations[37]. Lu et al. developed both traditional LUR models (e.g., stepwise linear regression) and machine learning models for six criteria

pollutants using traditional LUR variables versus microscale variables (e.g., Google POI, GSV Imagery)[25]. The study suggested that microscale variables are a suitable substitute for traditional GIS-derived variables. Lloyd et al. developed a hybrid approach which first trained CNN models on Google Maps satellite and street-level images to predict UFP and BC quintiles, and then combined the predictions with land use variables to further predict continuous particle measures. The results indicated redundant information when using both CNN predictions and land use variables, suggesting that satellite and street view images indeed capture some similar features of built environment[26].

In summary, there is a notable trend in recent air quality modeling studies towards the adoption of advanced machine learning approaches[39] to improve model performance, including random forest[40-43], gradient boosting[44,45], artificial neural network[30,46,47], deep learning[46,48,49], and hybrid algorithms. The emerging hyperlocal measures offer valuable information but are still underexplored. The model performance based on these novel variables indicate their great potential, either as additional variables to augment the traditional methods, or as an alternative to traditional predictors, which may be especially important for remote or less developed areas where curated GIS databases are sparse or nonexistent. The aim of this dissertation is to develop air quality modeling approaches leveraging street view and remote sensing imagery through computer vision and machine learning algorithms. The modeling framework is expected to be feasible for air pollution mapping with good model performance, consistent data collection and processing protocols across large geographies.

1.2 Objectives

The primary goal of this dissertation is to develop novel image-based predictors and new air quality empirical modeling approaches that are suitable for air pollution modeling and mapping. To achieve this goal, I apply computer vision techniques to extract built and natural environment features from street view and remote sensing imagery. Machine learning models are developed and tested for optimal model performance and predictive power. Specific objectives include:

(1) Using street view imagery to predict street-level particulate air pollution for a single city.

(2) Developing national LUR models for NO_2 using street view imagery and satellite-based aerosol estimates.

(3) Mapping national NO_2 and $PM_{2.5}$ air pollution at high resolution using street view and remote sensing imagery derived features.

1.3 Dissertation structure

The dissertation is structured as follows. The first chapter describes the background and motivation of dissertation research. Chapters 2-4 address the three objectives of the dissertation. The first study (chapter 2) is designed as a pilot study implemented in a single city (i.e., Blacksburg) to test the feasibility of solely using GSV images to build empirical air quality models and predict street-level air pollution. From this study, a pipeline to collect Google Street View imagery and extract a set of novel street-level predictor variables is established. Built upon the first study, the second study (chapter 3) extends the street view image-based models to the national level at multiple years. In addition, the impact of image availability on model performance is systematically assessed. A parsimonious image sampling strategy is proposed, and inclusion of

satellite-based aerosol estimates is suggested for modeling areas with insufficient street view images. To generate air pollution prediction surface in all regions, including areas with insufficient or no street view imagery, the third study (chapter 4) develops and compares models using different combinations of street view-derived and remote sensing-derived features. National annual NO_2 and $PM_{2.5}$ concentration maps are generated at high spatial resolution across 14 years using a cloud computing platform. Chapter 5 concludes the findings of this dissertation and implications for future studies.

1.4 References

1. Brauer, M. *et al.* Ambient air pollution exposure estimation for the Global Burden of Disease 2013. *Environ. Sci. Technol.* **50**, 79-88 (2016). https://doi.org:10.1021/acs.est.5b03709

2. Pope, C. A., Ezzati, M. & Dockery, D. W. Fine-Particulate Air Pollution and Life Expectancy in the United States. *N. Engl. J. Med.* **360**, 376-386 (2009). https://doi.org:10.1056/NEJMsa0805646

3. Lelieveld, J., Evans, J. S., Fnais, M., Giannadaki, D. & Pozzer, A. The contribution of outdoor air pollution sources to premature mortality on a global scale. *Nature* **525**, 367 (2015).

4. Kim, K. H., Kabir, E. & Kabir, S. A review on the human health impact of airborne particulate matter. *Environ. Int.* **74**, 136-143 (2015). https://doi.org:10.1016/j.envint.2014.10.005

5. Khomenko, S. *et al.* Premature mortality due to air pollution in European cities: a health impact assessment. *The Lancet Planetary Health* **5**, e121-e134 (2021).

6. de Bont, J. *et al.* Ambient air pollution and cardiovascular diseases: An umbrella review of systematic reviews and meta-analyses. *J. Intern. Med.* **291**, 779-800 (2022).

7. GBD Compare Data Visualization. http://vizhub.healthdata.org/gbd-compare (accessed March 20, 2023).

8. Alexeeff, S. E. *et al.* High-resolution mapping of traffic related air pollution with Google street view cars and incidence of cardiovascular events within neighborhoods in Oakland, CA. *Environ. Health* **17**, 13 (2018). https://doi.org:10.1186/s12940-018-0382-1

9. Apte, J. S. *et al.* High-resolution air pollution mapping with Google Street View cars: Exploiting big data. *Environ. Sci. Technol.* **51**, 6999-7008 (2017).

https://doi.org:10.1021/acs.est.7b00891

10. Miller, D. J. *et al.* Characterizing elevated urban air pollutant spatial patterns with mobile monitoring in Houston, Texas. *Environ. Sci. Technol.* **54**, 2133-2142 (2020). https://doi.org:10.1021/acs.est.9b05523

11. Kim, S. Y. *et al.* Concentrations of criteria pollutants in the contiguous US, 1979-2015: Role of prediction model parsimony in integrated empirical geographic regression. *PLoS One* **15**, e0228535 (2020). https://doi.org:10.1371/journal.pone.0228535

12. Schmitz, O. *et al.* Data Descriptor: High resolution annual average air pollution concentration maps for the Netherlands. *Sci. Data* **6**, 12 (2019). https://doi.org:10.1038/sdata.2019.35

13. Shi, Y., Lau, K. K. L. & Ng, E. Developing street-level $PM_{2.5}$ and PM_{10} land use regression models in high-density Hong Kong with urban morphological factors. *Environ. Sci. Technol.* **50**, 8178-8187 (2016). https://doi.org:10.1021/acs.est.6b01807

14. Hankey, S. & 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 (2015). https://doi.org:10.1021/acs.est.5b01209

15. Hoek, G. *et al.* A review of land-use regression models to assess spatial variation of outdoor air pollution. *Atmos. Environ.* **42**, 7561-7578 (2008). https://doi.org:10.1016/j.atmosenv.2008.05.057

16. Ryan, P. H. & LeMasters, G. K. A review of land-use regression models for characterizing intraurban air pollution exposure. *Inhal. Toxicol.* **19**, 127-133 (2007).

17. Marshall, J. D., Nethery, E. & Brauer, M. Within-urban variability in ambient air pollution: Comparison of estimation methods. *Atmos. Environ.* **42**, 1359-1369

(2008). https://doi.org:10.1016/j.atmosenv.2007.08.012

18. Pedersen, M. *et al.* Ambient air pollution and low birthweight: a European cohort study (ESCAPE). *Lancet Resp. Med.* **1**, 695-704 (2013). https://doi.org:10.1016/s2213-2600(13)70192-9

19. Xu, X. H., Ha, S. U. & Basnet, R. A review of epidemiological research on adverse neurological effects of exposure to ambient air pollution. *Front. Public Health* **4**, 31 (2016). https://doi.org:10.3389/fpubh.2016.00157

20. Hayes, R. B. *et al.* $PM_{2.5}$ air pollution and cause-specific cardiovascular disease mortality. *Int. J. Epidemiol.* **49**, 25-35 (2020). https://doi.org:10.1093/ije/dyz114

21. Eeftens, M. *et al.* Development of land use regression models for $PM_{2.5}$, $PM_{2.5}$ absorbance, PM_{10} and PM_{coarse} in 20 European study areas; Results of the ESCAPE project. *Environ. Sci. Technol.* **46**, 11195-11205 (2012). https://doi.org:10.1021/es301948k

22. Rundle, A. G., Bader, M. D. M., Richards, C. A., Neckerman, K. M. & Teitler, J. O. Using google street view to audit neighborhood environments. *Am. J. Prev. Med.* **40**, 94-100 (2011). https://doi.org:10.1016/j.amepre.2010.09.034

23. Larkin, A. *et al.* Global land use regression model for nitrogen dioxide air pollution. *Environ. Sci. Technol.* **51**, 6957-6964 (2017). https://doi.org:10.1021/acs.est.7b01148

24. Qi, M. & Hankey, S. Using street view imagery to predict street-level particulate air pollution. *Environ. Sci. Technol.* **55**, 2695-2704 (2021). https://doi.org:10.1021/acs.est.0c05572

25. Lu, T. *et al.* National empirical models of air pollution using microscale measures of the urban environment. *Environ. Sci. Technol.* **55**, 15519–15530 (2021).

26. Lloyd, M. *et al.* Predicting within-city spatial variations in outdoor ultrafine

particle and black carbon concentrations in Bucaramanga, Colombia: A hybrid approach using open-source geographic data and digital images. *Environ. Sci. Technol.* **55**, 12483-12492 (2021).

27. Hong, K. Y., Pinheiro, P. O. & Weichenthal, S. Predicting outdoor ultrafine particle number concentrations, particle size, and noise using street-level images and audio data. *Environ. Int.* **144**, 106044 (2020).

28. Weichenthal, S., Hatzopoulou, M. & Brauer, M. A picture tells a thousand...exposures: Opportunities and challenges of deep learning image analyses in exposure science and environmental epidemiology. *Environ. Int.* **122**, 3-10 (2019). https://doi.org:10.1016/j.envint.2018.11.042

29. Lefevre, S., Tuia, D., Wegner, J. D., Produit, T. & Nassar, A. S. Toward seamless multiview scene analysis from satellite to street level. *Proc. IEEE* **105**, 1884-1899 (2017). https://doi.org:10.1109/jproc.2017.2684300

30. Chan, K. L., Khorsandi, E., Liu, S., Baier, F. & Valks, P. Estimation of surface NO_2 concentrations over Germany from TROPOMI satellite observations using a machine learning method. *Remote Sens.* **13**, 969 (2021).

31. Demetillo, M. A. G. *et al.* Observing nitrogen dioxide air pollution inequality using high-spatial-resolution remote sensing measurements in Houston, Texas. *Environ. Sci. Technol.* **54**, 9882-9895 (2020). https://doi.org:10.1021/acs.est.0c01864

32. Ma, Z. W., Hu, X. F., Huang, L., Bi, J. & Liu, Y. Estimating ground-level $PM_{2.5}$ in China using satellite remote sensing. *Environ. Sci. Technol.* **48**, 7436-7444 (2014). https://doi.org:10.1021/es5009399

33. Di, Q. *et al.* Assessing $PM_{2.5}$ exposures with high spatiotemporal resolution across the continental United States. *Environ. Sci. Technol.* **50**, 4712-4721 (2016).

34. Lee, H. J. Benefits of high resolution $PM_{2.5}$ prediction using satellite MAIAC

AOD and land use regression for xxposure assessment: California examples. *Environ. Sci. Technol.* **53**, 12774-12783 (2019). https://doi.org:10.1021/acs.est.9b03799

35. Ganji, A., Minet, L., Weichenthal, S. & Hatzopoulou, M. Predicting traffic-related air pollution using feature extraction from built environment images. *Environ. Sci. Technol.* **54**, 10688–10699 (2020).

36. Hong, K. Y., Pinheiro, P. O. & Weichenthal, S. Predicting global variations in outdoor $PM_{2.5}$ concentrations using satellite images and deep convolutional neural networks. *arXiv preprint arXiv:1906.03975* (2019).

37. Zheng, T., Bergin, M. H., Hu, S., Miller, J. & Carlson, D. E. Estimating ground-level $PM_{2.5}$ using micro-satellite images by a convolutional neural network and random forest approach. *Atmos. Environ.* **230**, 117451 (2020).

38. Zhang, B. *et al.* Estimation of PM_x concentrations from landsat 8 oli images based on a multilayer perceptron neural network. *Remote Sens.* **11**, 646 (2019).

39. Li, Y., Sha, Z., Tang, A., Goulding, K. & Liu, X. The application of machine learning to air pollution research: A bibliometric analysis. *Ecotoxicol. Environ. Saf.* **257**, 114911 (2023).

40. Jain, S., Presto, A. A. & Zimmerman, N. Spatial modeling of daily $PM_{2.5}$, NO_2, and CO concentrations measured by a low-cost sensor network: Comparison of linear, machine learning, and hybrid land use models. *Environ. Sci. Technol.* **55**, 8631–8641 (2021).

41. Lim, C. C. *et al.* Mapping urban air quality using mobile sampling with low-cost sensors and machine learning in Seoul, South Korea. *Environ. Int.* **131**, 105022 (2019). https://doi.org:10.1016/j.envint.2019.105022

42. Vu, T. V. *et al.* Assessing the impact of clean air action on air quality trends in Beijing using a machine learning technique. *Atmos. Chem. Phys.* **19**, 11303-11314

(2019). https://doi.org:10.5194/acp-19-11303-2019

43. Li, Y. *et al.* Assessing the spatiotemporal characteristics, factor importance, and health impacts of air pollution in Seoul by integrating machine learning into land-use regression modeling at high spatiotemporal resolutions. *Environ. Sci. Technol.* **57**, 1225-1236 (2023).

44. He, Q. *et al.* Spatially and temporally coherent reconstruction of tropospheric NO_2 over China combining OMI and GOME-2B measurements. *Environ. Res. Lett.* **15**, 125011 (2020).

45. Kim, M., Brunner, D. & Kuhlmann, G. Importance of satellite observations for high-resolution mapping of near-surface NO_2 by machine learning. *Remote Sens. Environ.* **264**, 112573 (2021).

46. Wu, C.-l., Song, R.-f. & Peng, Z.-r. Prediction of air pollutants on roadside of the elevated roads with combination of pollutants periodicity and deep learning method. *Build. Environ.* **207**, 108436 (2022).

47. Kokkinos, K., Karayannis, V., Nathanail, E. & Moustakas, K. A comparative analysis of Statistical and Computational Intelligence methodologies for the prediction of traffic-induced fine particulate matter and NO_2. *J. Clean. Prod.* **328**, 129500 (2021).

48. Weichenthal, S., Dons, E., Hong, K. Y., Pinheiro, P. O. & Meysman, F. J. Combining citizen science and deep learning for large-scale estimation of outdoor nitrogen dioxide concentrations. *Environ. Res.* **196**, 110389 (2021).

49. Sun, H. *et al.* Spatial resolved surface ozone with urban and rural differentiation during 1990–2019: A space–time Bayesian neural network downscaler. *Environ. Sci. Technol.* **56**, 7337-7349 (2021).

Chapter 2 Using Street View Imagery to Predict Street-level Particulate Air Pollution

*Meng Qi, Steve Hankey**

School of Public and International Affairs, Virginia Tech, 140 Otey Street, Blacksburg, Virginia 24061, United States

* Corresponding author, School of Public and International Affairs, Virginia Tech

Email: hankey@vt.edu. Phone: 540.231.7508

The authors declare no competing interest.

ABSTRACT: Land-use regression (LUR) models are frequently applied to estimate spatial patterns of air pollution. Traditional LUR often relies on fixed-site measurements and GIS-derived variables with limited spatial resolution. We present an approach that leverages Google Street View (GSV) imagery to predict street-level particulate air pollution (i.e., black carbon [BC] and particle number [PN] concentrations). We developed empirical models based on mobile monitoring data and features extracted from ~52,500 GSV images using a deep learning model. We tested theory- and data-driven feature selection methods as well as models using images within varying buffer sizes (50-2,000m). Compared to LUR models with traditional variables, our models achieved similar model performance using the street-level predictors while also identifying additional potential hotspots. Adjusted R^2 (10-fold CV R^2) with integrated feature selection was 0.57-0.64 (0.50-0.57) and 0.65-0.73 (0.61-0.66) for BC and PN models, respectively. Models using only features near the measurement locations (i.e., GSV images within 250m) explained ~50% of air pollution variability, indicating PN and BC are strongly affected by the street-level built environment. Our results suggest that GSV imagery, processed with computer vision techniques, is a promising data source to develop LUR models with high spatial resolution and consistent predictor variables across administrative boundaries.

Keywords: integrated empirical model, exposure assessment, image processing, air quality, ultrafine

2.1 Introduction

The negative impacts of air pollution exposure on human health have been shown by many epidemiological studies[1-4]. Since air pollutant concentrations can vary rapidly over short distances[5, 6], accurately estimating the spatial patterns of air pollution is important for assessing human exposure and health outcomes[7, 8]. Numerous studies use measurements from fixed-site monitoring networks[9-12]. However, stationary monitoring is often cost-prohibitive, and thus sparsely distributed, which may lead to undetected hotspots and poor characterization of air quality at small spatial scales[13-16]. Compared to fixed-site monitoring, mobile monitoring offers an alternate approach to capture the spatial variation of street-level air pollution[5, 6]. For example, Apte et al.,[5] used Google Street View (GSV) cars to map air pollution at a spatial precision 4-5 orders of magnitude greater than that of routine monitoring stations. Consequently, empirical models based on dense mobile measurements may have the potential to better characterize street-level variation of air pollution to better describe health disparities at small spatial scales.

Land-use regression (LUR) is one type of empirical model which has been frequently applied to predict ambient air pollutant concentrations[17-20]. Typically, LUR models use measurements from fixed-site monitoring networks[17] and predictor variables derived from Geographic Information Systems (GIS) including land use types, traffic intensity, and population density[21]. While emerging studies use mobile monitoring data to increase the spatial density of measurements used in LUR models[22-25], other limitations arise when using traditional GIS-derived predictor variables that do not match the resolution of the mobile monitoring data. For example, GIS-derived variables are often jurisdiction specific since those data are usually collected to serve local administrative purposes[26, 27]. Similarly, variables available at larger geographic scales (e.g., national)

often provide less specific and spatially precise information. Since there are many mobile monitoring efforts being conducted across the world (which offer significant improvements in the spatial density of measurements), additional work is needed to develop predictor variables that (1) match the spatial scale of the measurements being collected and (2) are able to be derived in a consistent way across administrative and political boundaries. To address these limitations, we propose an image-derived empirical modeling approach that utilizes GSV imagery to extract street-level features as predictor variables.

Since its launch in 2007, GSV has collected a massive quantity of panoramic images across the world providing views of streetscapes at the eye-level[28]. GSV imagery is becoming a reliable data source[26, 28, 29] and a promising tool for research studies owing to high spatial resolution, extensive coverage, open access to the public, and data consistency – especially when combined with computer vision techniques[30-32]. Previous studies have used GSV images in various ways, for example, estimating the demographic makeup of neighborhoods[30], assessing urban vegetation and green space[33-35], auditing neighborhood built environments[26, 36], automating land-use classifications[37], or capturing changes in physical urban envirionments[38]. An emerging research area is using GSV imagery in air pollution research[39] to map hyperlocal air quality or to identify emission sources.

In this study, we investigate use of GSV imagery to predict street-level particulate air pollution. In previous work[24], we used a bicycle mobile monitoring platform to measure black carbon [BC] and particle number [PN] concentrations and developed LUR models based on traditional predictor variables. Based on the same air pollution measurements, we present here a series of empirical models with a promising new set of predictor variables: features extracted from GSV imagery by a deep learning scene

parsing technique[32]. The models are built solely with features from GSV images. Three different feature selection methods (i.e., theory-driven, data-driven, and integrated) are tested during model development to improve model robustness and interpretability. Our results demonstrate the combination of mobile monitoring and street scene images, assisted by computer vision techniques, constitutes a promising modeling approach to estimate air pollution at the street-level.

2.2 Materials and methods

2.2.1 Mobile monitoring data

Mobile monitoring data was collected in a previous study and full details of the monitoring campaign are described in Hankey et al[24]. Briefly, we conducted a bicycle mobile monitoring campaign to sample daytime street-level particulate air pollution in a small rural town (Blacksburg, VA) in 2016. We deployed microaethalometers (AE51; AethLabs) on bicycles to measure BC concentrations and Condensation Particle Counters (CPC 3007; TSI, Inc.) to measure PN concentrations. Two mobile monitoring routes, each ~20km in length, were repeatedly sampled and ~120h of mobile monitoring data was collected. We applied spatial aggregation at 100m intervals along the routes and the median concentrations of observations were tabulated for each aggregation location (n=422 aggregation locations). We tested different methods to adjust mobile monitoring data for background concentrations in Hankey et al[24]. In this study, we used concentrations adjusted by the multiplicative method, which aims to best approximate long-term air pollution, for model development. Briefly, the multiplicative background adjustment method calculated adjustment factors based on the ratio of the daily concentration to the hourly concentration at a central site used to measure background concentrations. Then, the adjustment factors were applied to all mobile monitoring

observations (based on the hour the mobile measurements were collected) to correct for variations in background concentrations.

2.2.2 Street-level predictor variables

Unlike traditional LUR models, the predictor variables in this study were features extracted from GSV imagery, which capture street-level characteristics of the built and natural environment. GSV images were sampled around each mobile monitoring aggregation location including (1) central sampling which randomly retrieved 6 locations of GSV imagery within 50m of the aggregation locations and (2) buffer sampling where we created a 100m ×100m grid within 2,000m of each aggregation location and then repeated the central sampling approach for the centroid of each grid cell (**Figure 2-1**). GSV images within the 2,000m buffer were further stratified into 8 circular buffers with radii ranging from 250m to 2,000m. Since GSV images are 360° photos, we retrieved 4 flat GSV images (640×640 pixels) at 90° intervals for each GSV sampling point. This procedure resulted in up to 24 images around each monitoring location or grid cell centroid depending on the availability of GSV images. To best match the mobile monitoring campaign, we preferentially collected GSV images photographed during the same year (2016). Images from the next closest year were included when year-2016 images were not available.

We used the Pyramid Scene Parsing Network (PSPNet), a deep learning model with state-of-the-art performance on various datasets, to process the GSV images[32]. Feature extraction was implemented at the pixel-wise level, i.e., each pixel in an image was classified into a specific feature. A total of 150 features were segmented, including natural features (e.g., trees, grass, plants), built environment features (e.g., roads, buildings, houses), and transportation features (e.g., cars, buses, vans). The percentage of each feature within an image was calculated (**Figure 2-1-D**). The mean percentage

of features among all images within a given buffer, representing the overall characteristics of that area, were used as street-level predictor variables for our empirical air quality models.

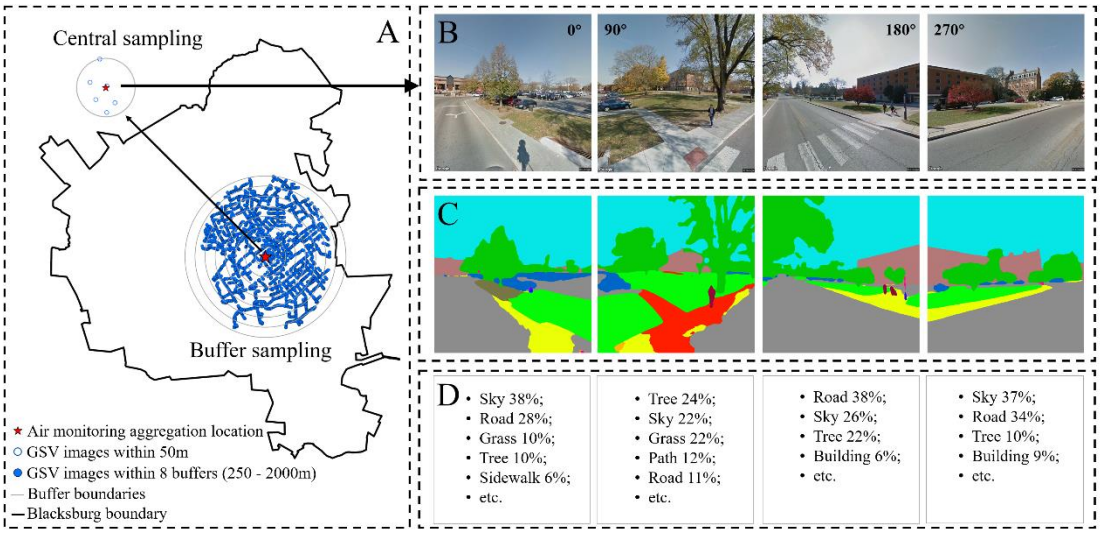

Figure 2-1. Image sampling and feature extraction procedure. (A) Google Street View (GSV) images sampled around one air pollutant monitoring aggregation location. (B) GSV images retrieved for 4 directions at one GSV sampling location. (C) Feature extraction using PSPNet scene parsing. (D) Estimated feature percentage among the 150 segmented classes.

2.2.3 Modeling approach

We developed GSV-only empirical models with predictor variables solely extracted from GSV images. To make direct comparisons with models using traditional variables, we used the same air pollution measurements and employed the same step-wise linear regression approach used by Hankey et al.[24]. Briefly, the model uses a forward selection procedure, adding the variable which has the highest correlation with the dependent variable and subsequently adding the variable which is most correlated with model residuals. The process was repeated until the added variable was either not statistically significant ($p < 0.05$) or the model had unacceptably high multicollinearity (Variance

Inflation Factor VIF > 5). All BC and PN concentrations were log transformed during model development and were transformed back to original values for model prediction.

To fully explore the role of GSV imagery in estimating street-level air pollution, we developed two types of GSV-only models: buffer-feature models and central-feature models. The buffer-feature models included both central features (i.e., the mean percentage of features for images within 50m of aggregation locations) and buffer features at various spatial scales (i.e., 8 circular buffers from 250m to 2,000m). During model development, each feature class was restricted to only be selected at a single buffer size. We also built central-feature models that only used features within 50m of measurement locations to assess the ability of using only images very near measurement sites to explain air pollution variability. A sensitivity analysis was developed to investigate how the choice of spatial scale of the central features (i.e., by varying the radius size that defines central features) impacts model performance. Specifically, we collected features from GSV images within 250, 500, 750 and 1000m of the mobile monitoring routes. Then, we compared central-feature models for each fixed radius.

Adjusted R^2 and root-mean-square error (RMSE) were used as metrics to compare model performance. To evaluate model robustness and generalizability, 10-fold cross validation was conducted for each model. Specifically, the data was randomly partitioned into 10 subsets. Then we held out each subset and used the remaining 9 subsets to build a model and predict the data for the hold out subset. R^2 was calculated after all subsets were predicted. We repeated the 10-fold cross validation 10 times. The mean R^2 was reported as the final 10-fold cross validation R^2.

2.2.4 Feature selection

We used a previously developed algorithm (PSPNet) to classify pixels in the GSV images[32]. Upon inspection of all 150 classes extracted from the images, we found that many classes are irrelevant to air pollution; furthermore, as expected, some pixels were misclassified during scene parsing. Including irrelevant and spurious features in empirical air quality models may cause overfitting and reduced accuracy when applying the models. It also makes it difficult to interpret the selected variables and their coefficients. One way to mitigate these problems is to conduct feature selection during the modeling process[40, 41]. In this study, we tested three feature selection methods to develop GSV-derived predictor datasets to offer to our empirical air quality models. To avoid confusion, we refer to feature selection in this paper as the procedure used to assemble a subset of the 150 features extracted from the GSV imagery to offer to our empirical air quality models; this process is independent of the variable selection process used within the step-wise regression. The three feature selection methods include: (1) theory-driven feature selection, (2) data-driven feature selection, and (3) integrated feature selection, which is the combination of the first two approaches. In general, theory-driven feature selection utilizes subjective background knowledge accumulated among many studies, while data-driven feature selection solely relies on statistical characteristics of the data itself.

We designed our theory-driven feature selection based on our prior knowledge of air pollution. To illustrate this approach, we developed three categories of predictor variables derived from the GSV images to offer to our air quality models: (1) "All variables" which included all 150 features, (2) "Theory-driven: Outdoor variables" which only included features that are commonly outdoors (n=57), and (3) "Theory-driven: Air pollution-related variables" which only included features that may influence air pollution concentrations (n=31). A full list of variables in each category as well as

the reason for each variable being included or excluded from each category is listed in **Table 2-S1**.

As a contrast to theory-driven feature selection, we designed our data-driven feature selection simply based on quantitative data filters without using any prior knowledge of air pollution. We used two data-filtering criteria and applied them to the "All variables" category. First, we filtered predictor variables based on the variable mean value averaged among all model building locations. The thresholds (i.e., 1×10^{-6}, 1×10^{-5} and 1×10^{-4}) were determined by the statistical distribution shown in **Figure 2-S1**. Second, we filtered predictor variables based on the nonzero percent of features among the aggregation locations. The thresholds were set to be equally distributed: 0%, 20%, 40%, 60% and 80%. A total of 15 scenarios were then created based on the combination of these criteria. Both criteria aim to remove uncommon features in the study area from the model building dataset. More details are described in SI.

Integrated theory- and data-driven feature selection was applied to leverage the strengths of both approaches. These models used the same data filtering criteria by the data-driven approach and applied them to the categories of predictor variables created by the theory-driven approach. To simplify model comparison, the best models among scenarios within the data-driven or integrated feature selection approaches were chosen based on the highest 10-fold CV R^2.

2.3 Results and discussion

2.3.1 Spatial distribution of street-level predictor features

The mean concentrations of BC and PN were 1.08 $\mu g/m^3$ and 6,059 pt/cm^3, respectively. The distribution of mobile monitoring data is shown in **Figure 2-S2**. **Figure 2-S3** shows the spatial distribution of GSV images sampled around the mobile monitoring routes. Approximately 52,500 GSV images were collected for model

development, including ~8,800 GSV images collected within 50m of the monitoring routes to capture central features and ~46,800 images within 2,000m to generate variables at various buffer radii. Corresponding to the two sources of the GSV imagery, Google (i.e., collecting images while GSV cars driving along roads) and Google users uploading photos, the GSV images were mostly distributed along the road network or locations with more human activities. Since the GSV images were predominantly on the road network, portions of the mobile monitoring routes (e.g., off-street trails far from major roads) did not have GSV images within 50m (11.6% of the mobile monitoring aggregation locations). Thus, we excluded these monitoring locations from our model building in this study.

A core advantage of GSV images is that they capture street-level built and natural environment features. Among the 150 features segmented by scene parsing, the top 10 most frequent features in Blacksburg were as follows: sky (average percent in images: 28.1%), road (25.8%), tree (24.3%), grass (10.8%), earth (2.7%), building (1.9%), plant (1.4%), sidewalk (1.3%), car (0.8%) and house (0.7%). To illustrate the spatial granularity of GSV features, road and tree (as indicators of traffic and green space) are shown in **Figure 2-S3**. The distribution of roads matched well with the road network in Blacksburg and had higher values along higher order roads (e.g., highways). The tree variable followed intuition and had higher values where there were large tracts of tree canopy or abundance of street trees. Tree cover was prone to systematic bias due to the way GSV images are collected. For example, large areas of tree cover exist in the NW corner of the study area, but they were not captured because this area has few roads and thus few GSV images. This bias is not the case for traffic related variables (e.g., roads, cars, sidewalks) that are only located on the transportation network. Detailed descriptive statistics of the extracted GSV features are listed in **Table 2-S2**.

2.3.2 Model results

2.3.2.1. Theory-driven models.

Overall, we found good model performance for the GSV-only empirical models. As shown in **Tables 2-S3** and **2-S4**, the adjusted R^2 (10-fold CV R^2) for the buffer-feature models using all variables was 0.77 (0.42) and 0.77 (0.57) for BC and PN, respectively. The decrease in R^2 after 10-fold cross validation was large, which can be explained by several factors. Stochastic effects could be significant as GSV images were sampled with very high spatial resolution and thus possibly reveal more landscape heterogeneity. More importantly, the scene parsing algorithm generates 150 features which is likely unnecessary as many of these features are irrelevant to air pollution. To reduce the amount of spurious variables, we conducted theory-driven feature selection and evaluated the change in model performance (**Figure 2-2**). As the selection criteria for including variables became more stringent, we generally observed a slight decrease in adjusted R^2, but importantly, a larger increase in 10-fold CV R^2 which reduced the gap between the two metrics. For example, the gap for the unconstrained BC (PN) models (i.e., using all variables) was 0.35 (0.20) as compared to 0.07 (0.07) when using "Theory-driven: Air pollution-related variables". This finding indicates that the most stringent theory-driven models reduced overfitting as well as improved model stability and variable interpretability while maintaining satisfying model fits. The adjusted R^2 (10-fold CV R^2) of the buffer-feature models with only air pollution-related variables were 0.63 (0.57) and 0.73 (0.65) for BC and PN, respectively. As expected, the model results show improved rationality in terms of the selected variables for the theory-driven feature selection as compared to the unconstrained models (**Tables 2-S3** and **2-S4**). We observed a slight increase in RMSE as theory-driven feature selection became stricter. For example, RMSE for the BC models using all variables, only outdoor

variables, and only air pollution-related variables were 0.20, 0.22 and 0.26 μg/m^3, respectively. This trend could be explained by the reduced overfitting for theory-driven models, which is consistent with the finding that theory-driven models have lower adjusted R^2 but higher 10-fold R^2 compared to unconstrained models.

To investigate the correlation of street-level features within the immediate vicinity of the mobile monitoring routes to particulate concentrations, central-feature models were developed. As shown in **Figure 2-2**, the adjusted R^2 (10-fold CV R^2) was 0.43 (0.16) for BC and 0.33 (0.13) for PN when using all variables. Similar trends to the buffer-feature models among different variable categories were also found for the central-feature models (**Figure 2-2**). However, the decrease in model fit before and after cross validation was more drastic for the central-feature models, possibly owing to the fact that only a limited number of images within 50m of mobile monitoring were used for generating central features. A sensitivity analysis was further developed to explore the influence of radius size used for defining central features. **Figure 2-S4** shows the performance of central-feature models developed using the air pollution-related variables at various spatial scales (i.e., from 250m to 1,000m). Both adjusted R^2 and 10-fold CV R^2 increased when the central features were extracted from larger areas. For example, when setting the central range to 250m, the central-feature models explained ~50% of the variance. Adjusted R^2 (10-fold CV R^2) increased to 0.51 (0.49) and 0.53 (0.48) for the BC model and the PN model, respectively. Prediction of PN concentrations was best using 250m buffers and decreased at larger distances; the BC models showed little variability between 250 and 750m. This finding suggests these pollutants were strongly affected by the immediate surrounding environment at the magnitude of several hundred meters.

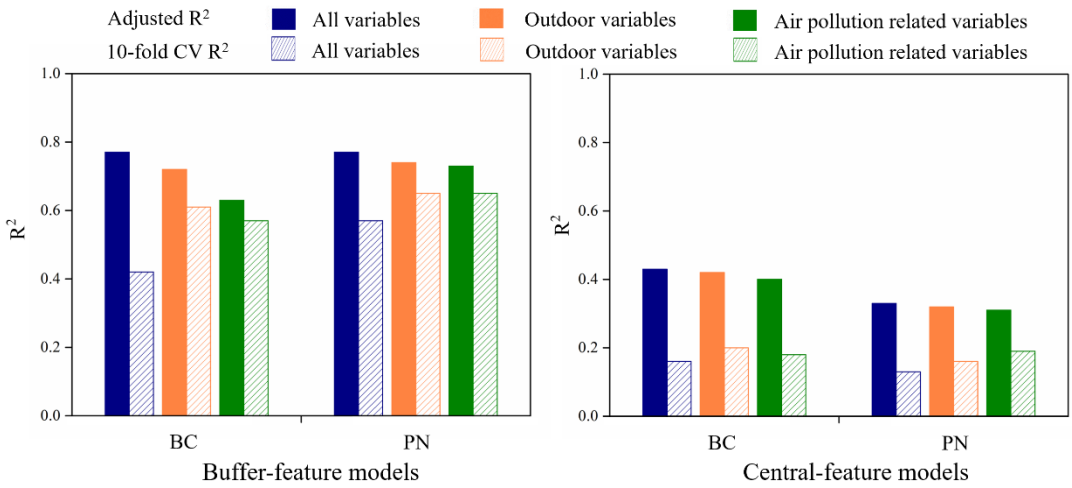

Figure 2-2. Comparison of model performance using theory-driven feature selection. Models developed with different categories of predictor variables are shown by color. Solid columns represent model adjusted R^2; hashed columns represent 10-fold cross validation R^2.

2.3.2.2 Data-driven models.

Table 2-S5 shows the number of predictor variables offered to our models using different feature selection methods. Although no prior knowledge of air pollution was applied, the number of predictor variables offered to the data-driven models among all scenarios (252-659 variables) aligned well with that of theory-driven models (279-513). Furthermore, we found a significant overlap of variables chosen by the two feature selection methods. For example, 67%-100% of "Theory driven: Air pollution-related variables" were also selected among the 15 data-driven scenarios. This demonstrates that even without background information on air quality, data-driven feature selection can achieve similar results to the theory-driven selection approach. Similar to the theory-driven models, the data-driven selection also showed the ability to enhance model robustness. Adjusted R^2 (10-fold CV R^2) of buffer-feature models through data-driven selection was 0.64-0.76 (0.35-0.60) for BC and 0.69-0.80 (0.63-0.70) for PN (**Figure 2-S5**). Compared to the unconstrained models, the data-driven models

improved 10-fold CV R^2 significantly while maintaining high adjusted R^2. In addition, the 10-fold CV R^2 was always the lowest in the least restrictive scenario, suggesting high multicollinearity when using too many variables. These trends were exacerbated in the central-feature models (**Figure 2-S6**). For example, the 10-fold CV R^2 of the BC central-feature model of the least and most restrictive scenarios were 0.07 and 0.33, respectively. Assuming the best model had the highest 10-fold CV R^2 among scenarios, adjusted R^2 (10-fold CV R^2) of the best BC and PN buffer-feature models were 0.72 (0.60) and 0.80 (0.70). RMSE of the best BC and PN buffer-feature models were 0.22 $\mu g/m^3$ and 807 pt/cm^3. Full model parameters are listed in **Table 2-S6**. Although the model fits were improved by data-driven variable selection, a few irrelevant variables were still selected in the models, leading to reduced model interpretability.

2.3.2.3. Integrated feature selection models.

As expected, similar trends of improved model robustness and interpretability when using the theory- and data-driven feature selection methods were also observed using the integrated method (**Figure 2-3**). The best integrated models always outperformed or matched the corresponding theory-driven models showing the utility of using both theory- and data-driven feature selection. We also found that the buffer-feature models generally had better model fits for PN while the central-feature models showed better performance for BC. This finding suggests that the effect of the street-level environment on BC was greater than on PN. This is reasonable since the dominant sources of BC are primary emissions related to combustion processes[42] while particles can be either emitted directly or generated from secondary sources (e.g., precursors of photochemical reactions[43, 44]).

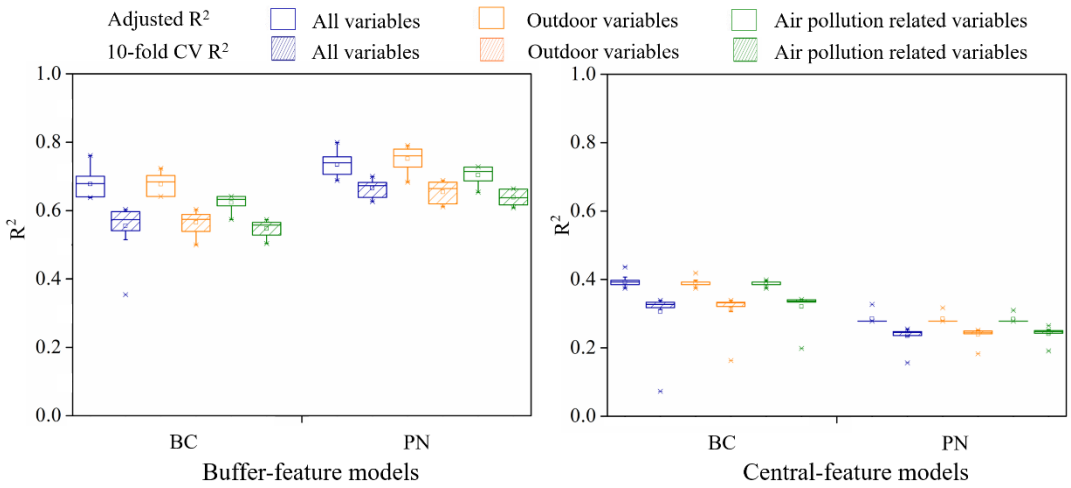

Figure 2-3. Model performance using integrated theory- and data-driven feature selection. Each box represents 15 scenarios for different categories of predictor variables. Models developed with different categories of predictor variables are shown by color. Hollow boxes represent model adjusted R^2; hashed boxes represent 10-fold cross validation R^2.

2.3.3 Comparison among models

Table 2-1 shows model performance for all GSV-only models as well as the previously published LUR models[24]. In general, GSV-only models had comparable model performance to LUR models using traditional GIS-derived variables. For buffer-feature models with only air pollution-related variables (best for interpretability), the 10-fold CV R^2 after integrated feature selection was as high as 0.57 for BC and 0.66 for PN. If pursuing the best goodness-of-fit and neglecting the theoretical basis of predictor variables, we can achieve better model fits. For example, the best data-driven models showed higher adjusted R^2 (10-fold CV R^2) and lower RMSE as compared to previous LUR models using traditional predictors. Our models suggest that features extracted from GSV images can be used to successfully develop predictive models in a variety of ways depending on the user's purpose, e.g., either pursuing best goodness-of-fit regardless of model interpretability or to be consistent with theory as much as possible.

We chose a subset of buffer-feature models to estimate BC and PN concentrations for the entire study area (~63,000 GSV images): (1) models based on all variables, (2) theory-driven models based on air pollution-related variables, (3) the best data-driven models, and (4) the best integrated theory- and data-driven models based on air pollution-related variables. Concentrations were mapped on a 100m×100m grid in **Figure 2-4**. Due to the limitation of GSV images mainly being captured along roadways, we were only able to make predictions for 39% of Blacksburg's land area. **Figure 2-4** also includes scatterplots of predicted vs. observed concentrations. In general, all models showed good agreement between model predictions and mobile monitoring data. The overall spatial patterns of estimated concentrations were similar among models, especially for areas where GSV images were more densely distributed. A few inconsistencies were apparent, for example, the SE corner of Blacksburg was estimated to have extremely low concentrations by the unconstrained PN model but was a hotspot using other models. Based on a review of these models, we found the low estimates resulted from some portions of GSV images being misclassified into the class 'bed' – a feature that is highly unlikely to appear frequently in outdoor images. Given that this area is an intersection of several major roads, it is expected to have elevated concentrations in this area. This example illustrates the importance of using feature selection to remove spurious predictor variables prior to building empirical models. Similar problems appeared in the NW corner of the map. In this case, the inconsistent predictions for BC concentrations mainly resulted from the sparsity of GSV images in that region along with a few unusual GSV features. Considering both the theoretical basis of input variables and model accuracy, we suggest using models with integrated variable selection for broader model applications.

Although the application of our models was limited by the availability of GSV images, alternative models could be constructed to mitigate this problem. For example, we developed reduced buffer-feature models using only larger buffer radii (full details are described in SI) and the models maintained good model performance (**Table 2-S7**). For example, adjusted R^2 (10-fold CV R^2) of the models with radii 500m or greater were 0.60 (0.53) for BC and 0.62 (0.56) for PN. Since 81% of the grid centroids for prediction had GSV images within 250m (94% when the radius was 500m), we applied the reduced buffer-feature models to a wider land area of Blacksburg to generate air pollution estimates (**Figure 2-S8**). These models showed reasonably consistent spatial patterns to the LUR models from our previous work[24]. In general, the hotspots identified by the LUR models with traditional GIS-derived variables were mainly distributed along major roads. While the GSV-only models also revealed these hotspots, our proposed models further identified other localized regions with high BC or PN concentrations that the GIS-derived models did not. For example, at the intersection of several major roads in the SE corner of Blacksburg, both GSV-only models and the previous LUR models estimated elevated PN concentrations in this region, while only the GSV-only models identified it as a BC hotspot as well. Future work may collect field observations in these potential hotspots to validate our new modeling approach.

2.3.4 Results of spatial autocorrelation analysis and model uncertainty

To compare among our models, we assessed spatial autocorrelation of model residuals and model uncertainty (full details are described in SI). Briefly, our model residuals demonstrate at least some spatial autocorrelation, partly due to the fine spatial resolution of the mobile monitoring data and GSV images (**Table 2-S8** and **Figure 2-S9**). Compared with the previous LUR models using traditional predictors, GSV-only models show slightly higher standard error (**Figures 2-S10** and **2-S11**). A useful

direction for future research is to use more advanced modeling approaches to account for spatial autocorrelation issues and reduce model uncertainty.

Table 2-1. Summary of model performance among all model types

Pollutants	Model type	Feature selection method	Predictor variables	Number of selected variables	Adjusted R^2	10-fold CV R^2	RMSE
BC ($\mu g/m^3$)	Buffer-feature models	None	All variables	20	0.77	0.42	0.20
		Theory-driven	Outdoor variables	19	0.72	0.61	0.22
		Theory-driven	Air pollution-related variables	10	0.63	0.57	0.26
		Data-driven	All variables	11-27	0.64-0.76	0.35-0.60	0.20-0.26
		Integrated	Outdoor variables	11-21	0.64-0.72	0.50-0.60	0.22-0.25
		Integrated	Air pollution-related variables	9-12	0.57-0.64	0.50-0.57	0.25-0.28
	Central-feature models	None	All variables	9	0.43	0.16	0.32
		Theory-driven	Outdoor variables	10	0.42	0.20	0.32
		Theory-driven	Air pollution-related variables	8	0.40	0.18	0.33
		Data-driven	All variables	5-13	0.37-0.44	0.07-0.34	0.32-0.34
		Integrated	Outdoor variables	5-10	0.37-0.42	0.16-0.34	0.32-0.34
		Integrated	Air pollution-related variables	5-8	0.37-0.40	0.20-0.34	0.33-0.34
	Previous LUR models[24]	None	Census data only; no GSV data	8	0.62	0.57	0.26*
PN (pt/cm^3)	Buffer-feature models	None	All variables	15	0.77	0.57	821
		Theory-driven	Outdoor variables	12	0.74	0.65	916
		Theory-driven	Air pollution-related variables	17	0.73	0.65	954
		Data-driven	All variables	7-22	0.69-0.80	0.63-0.70	807-1050
		Integrated	Outdoor variables	12-24	0.68-0.79	0.61-0.69	857-1005
		Integrated	Air pollution-related variables	11-17	0.65-0.73	0.61-0.66	941-1030
	Central-feature models	None	All variables	12	0.33	0.13	1379
		Theory-driven	Outdoor variables	8	0.32	0.16	1391
		Theory-driven	Air pollution-related variables	7	0.31	0.19	1408
		Data-driven	All variables	5-9	0.28-0.33	0.16-0.26	1379-1461
		Integrated	Outdoor variables	5-8	0.28-0.32	0.18-0.25	1391-1461
		Integrated	Air pollution-related variables	5-7	0.28-0.31	0.19-0.26	1408-1461
	Previous LUR models[24]	None	Census data only; no GSV data	16	0.78	0.70	906*

* RMSE for previous LUR models was calculated based on data from Hankey et al.,[24]

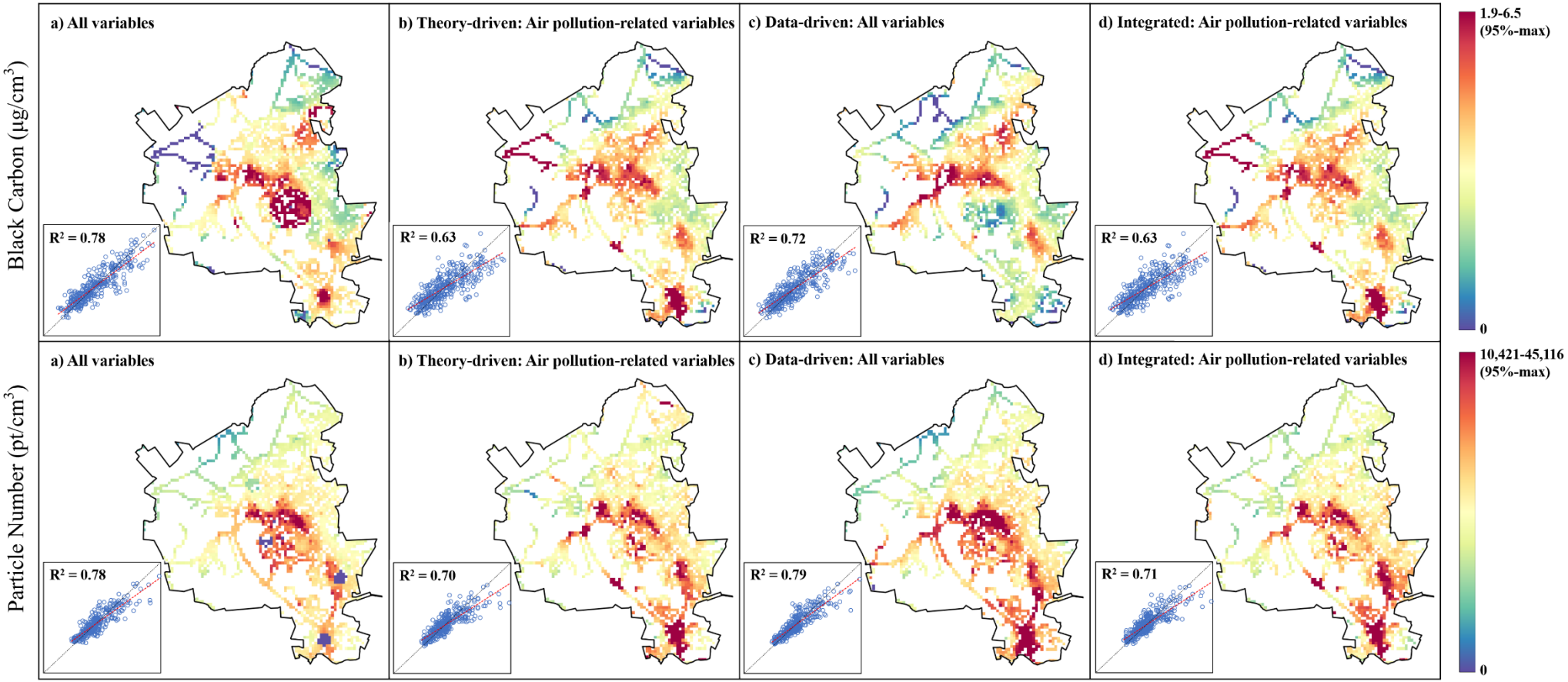

Figure 2-4. Estimated BC and PN concentrations using 4 types of GSV-only models. Only grid centroids with GSV images within 50m are displayed. The scatter plots in the lower-left corner show model predicted air pollutant concentrations vs. mobile monitoring air pollutant concentrations. The dashed red line is the best fit line; the dashed black line is the 1:1 line. Due to the highly right-skewed distributions of model estimates (**Figure 2-S7**), we adjusted the color bar scale to visualize the spatial patterns of the models. Specifically, for each air pollutant, we binned the estimates among the 4 models and used the 95th percentile as the maximum for visualization.

2.4 Limitations and implications

The usefulness of GSV-only models depends on the availability and quality of GSV images for any given region. In less developed or rural regions, this approach may miss large tracts of land where there is no road access and thus no GSV imagery. In addition, GSV images are static and in many areas are not updated at the same rate as urban development. Although we prioritized sampling GSV images from the same year of air quality monitoring, only 12% of GSV images retrieved were photographed in 2016. The impact of this temporal mismatch could be large or small depending on the rate of change of the built environment. Future work could apply our approach in urban areas where larger amounts of GSV images are available and updated more frequently. Given the continually growing repositories of public streetscape imagery[33], this limitation is expected to be mitigated in the future.

Introducing better temporal resolution is another important goal for accurate exposure assessment[24, 45, 46]. While finer spatial resolution can be achieved using GSV imagery, improving temporal resolution is a challenge in the short run. Multiple studies have used satellite imagery to map air pollution with high spatial-temporal resolution[47-49]. A possible solution is to use both GSV images and satellite imagery for model development. By combining remote sensing imagery (which can achieve daily updates) and ground-level imagery, both model accuracy and spatiotemporal precision may be leveraged as multi-view aspects of the built environment could be used[50, 51]. Another limitation is that the performance of the image processing technique (in this case PSPNet) for feature extraction also determines the performance of the image-based air

quality models. Our results show that feature selection can successfully reduce the number of spurious variables, which reduced issues caused by overcomplicated semantic segmentation or misclassification. Given the rapid development in computer vision and deep learning[52, 53], future work may also explore improved algorithms to reduce misclassification during scene parsing.

Our study suggests that GSV imagery alone, and thus likely any other large street scene imagery dataset with high resolution and broad coverage, may provide sufficient information to characterize air pollution patterns. Imagery based data sources offer potentially three major advantages: (1) finer spatial (i.e., street-level) resolution, (2) the ability to apply consistent data collection and processing protocols across large geographies and political boundaries, and (3) flexibility in method of data reduction for model development and application. The development of low-cost and portable sensors in conjunction with increased efforts to explore mobile monitoring provides an opportunity to characterize the spatial variability of air pollution at the street-level[54, 55]. Our proposed modeling approach leverages the strengths of mobile monitoring and GSV images in characterizing street-level gradients of air pollution. Our model results also highlight the impact of street-level factors (e.g., built environment, local emission sources) on air quality which suggests highly localized interventions may yield reductions in exposure. As low-cost air quality sensors and street scene images become more ubiquitous, the combination of mobile monitoring and automated street-level image analysis could become a scalable framework that could be applied nearly

anywhere in the world as a tool to track air pollution exposure and disparities at small spatial scales.

2.5 Supporting Information

2.5.1 Description of data-driven feature selection

For the data-driven feature selection, we filtered predictor variables according to the statistical distribution of the variables. We used two data-filtering criteria on predictor variables in the "All variables" category. First, we set up thresholds for the mean of the predictor variables to filter out uncommon features for the overall study area. There are 1,350 potential predictor variables (150 classes × 9 buffer sizes) in total. We calculated the mean for each predictor variable at all aggregation locations as this metric could serve as a proxy for how common each feature is in the entire study area. According to the statistical distribution shown in **Figure 2-S1**, we determined the thresholds of mean values to be 1×10^{-6}, 1×10^{-5} and 1×10^{-4}. The second filtering criteria was based on the nonzero percent of features among the aggregation locations to filter out uncommon features within measurement sites. Due to spatial discontinuity and heterogeneity, features along the monitoring routes varied as some features only appeared in a small number of places. For this criterion, the thresholds were set to be equally distributed: 0%, 20%, 40%, 60% and 80%. The aim of both criteria is to reduce uncommon features as these variables would harm the generalizability of our models. As a result, a total of 15 scenarios were created by combining these two filters for the data-driven feature selection.

2.5.2 Reduced buffer-feature GSV-only models

Since some grid centroids for air pollution prediction could not find GSV images nearby, we developed reduced buffer-feature models that avoid using central features or buffer features at low spatial radii. Specifically, we built models with the lowest buffer radius from 250m to 1,000m while the largest buffer size remained 2,000m. **Table 2-S7** lists the model parameters for the reduced buffer-feature GSV-only models as well as their model performance. All models used integrated feature selection with air pollution-related variables to prioritize model generalizability and interpretability. In addition, 81% of the grid centroids in the study area were able to find GSV images within 250m and the percentage was 94% when the radius extended to 500m. We explored two ways for applying the reduced buffer-feature models to make predictions for the 94% coverage of land area in Blacksburg. First, we developed an additive method which aimed to use models with the best possible spatial resolution for areas with limited GSV images. Specifically, for regions where GSV images were unavailable within 50 m but available within 250m, we used the 250m reduced buffer-feature models with buffer variables starting from 250m. For regions where GSV images were unavailable within 250 m but available within 500m, we used the 500m reduced buffer-feature models with buffer variables starting from 500m. Second, we used a simplified method, which simply used the reduced buffer-feature model with buffer variables starting from 500 m for all regions. **Figure 2-S8** shows the prediction maps using these methods as well as the estimates by the LUR models from our previous work[1] for comparison.

2.5.3 Results of spatial autocorrelation analysis and model uncertainty

Table 2-S8 shows the results of Global Moran's I tests for our model residuals using different weights matrices and distance thresholds. The smallest radius (300m) is the minimum distance where each aggregation location finds at least one neighbor. In general, the model residuals demonstrate at least some spatial autocorrelation. This may be partly due to the fine spatial resolution of the mobile monitoring data and GSV images. To identify clusters of spatial autocorrelation, we performed a LISA (local indicator of spatial association) analysis. As shown in **Figure 2-S9**, we chose the integrated buffer-feature models with air pollution-related variables for illustration. Generally, the high-high clusters (i.e., underprediction) are located where BC and PN concentrations are high while the low-low clusters (i.e., overprediction) are located where the observed concentrations are relatively low.

Figures 2-S10 and **2-S11** show model uncertainty by GSV-only models (integrated buffer-feature models with only air pollution-related variables for illustration) and previous LUR models[1]. Compared with the previous LUR models, GSV-only models show slightly higher standard error. The larger model uncertainty also tends to occur where BC and PN concentrations are at the higher or lower end of the observed distribution. As this study mainly aims to compare traditional GIS-derived predictors and our new GSV imagery-derived predictors, we chose to develop traditional step-wise linear regression models. A limitation of this choice is that our modeling framework does not account for spatial autocorrelation. A useful direction for future

research is to use more advanced modeling approaches to address this issue, especially to improve model performance for extremely high or low concentration regions.

Table 2-S1. Predictor variables in each category for the theory-driven feature selection

No.	Predictor variables	All variables	Theory-driven: Outdoor variables	Theory-driven: Air pollution-related variables	Reason to include/exclude variables from "Theory-driven: Air pollution related variables"
1	airplane	√	√		not found in GSV images in Blacksburg
2	animal	√	√		not found in GSV images in Blacksburg
3	apparel	√			indoor
4	arcade machine	√			indoor
5	armchair	√			indoor
6	ashcan	√			indoor
7	awning	√	√		no air pollutant emission
8	bag	√			indoor
9	ball	√			indoor
10	bannister	√			indoor
11	bar	√			indoor
12	barrel	√			indoor
13	base	√			indoor
14	basket	√			indoor
15	bathtub	√			indoor
16	bed	√			indoor
17	bench	√			indoor
18	bicycle	√	√	√	traffic
19	blanket	√			indoor
20	blind	√			indoor
21	boat	√	√		not found in GSV images in Blacksburg
22	book	√			indoor
23	bookcase	√			indoor
24	booth	√			indoor
25	bottle	√			indoor
26	box	√			indoor
27	bridge	√	√	√	traffic
28	buffet	√			indoor
29	building	√	√	√	pollutant dispersion
30	bulletin board	√			indoor
31	bus	√	√	√	traffic
32	cabinet	√			indoor
33	canopy	√	√		not found in GSV images in Blacksburg
34	car	√	√	√	traffic
35	case	√			indoor
36	ceiling	√			indoor
37	chair	√			indoor
38	chandelier	√			indoor
39	chest of drawers	√			indoor
40	clock	√			indoor
41	coffee table	√			indoor
42	column	√			indoor
43	computer	√			indoor
44	conveyer belt	√			indoor
45	counter	√			indoor
46	countertop	√			indoor
47	cradle	√			indoor
48	crt screen	√			indoor
49	curtain	√			indoor

No.	Predictor variables	All variables	Theory-driven: Outdoor variables	Theory-driven: Air pollution-related variables	Reason to include/exclude variables from "Theory-driven: Air pollution related variables"
50	cushion	√			indoor
51	desk	√			indoor
52	dirt track	√	√	√	dust
53	dishwasher	√			indoor
54	door	√			indoor
55	earth	√	√	√	dust/green space
56	escalator	√			indoor
57	fan	√			indoor
58	fence	√	√		no air pollutant emission
59	field	√	√	√	dust/green space
60	fireplace	√			indoor
61	flag	√	√		no air pollutant emission
62	floor	√			indoor
63	flower	√	√	√	green space
64	food	√			indoor
65	fountain	√	√		not found in GSV images in Blacksburg
66	glass	√			indoor
67	grandstand	√			indoor
68	grass	√	√	√	green space
69	hill	√	√	√	green space
70	hood	√			indoor
71	house	√	√	√	pollutant dispersion
72	hovel	√	√		not found in GSV images in Blacksburg
73	kitchen island	√			indoor
74	lake	√	√		not found in GSV images in Blacksburg
75	lamp	√			indoor
76	land	√	√	√	dust/green space
77	light	√			indoor
78	microwave	√			indoor
79	minibike	√	√	√	traffic
80	mirror	√			indoor
81	monitor	√			indoor
82	mountain	√	√	√	green space
83	ottoman	√			indoor
84	oven	√			indoor
85	painting	√			indoor
86	palm	√	√		not found in GSV images
87	path	√	√	√	traffic
88	person	√	√	√	possible emission source
89	pier	√	√		not found in GSV images in Blacksburg
90	pillow	√			indoor
91	plant	√	√	√	green space
92	plate	√			indoor
93	plaything	√			indoor
94	pole	√	√		no air pollutant emission
95	pool table	√			indoor
96	poster	√			indoor
97	pot	√			indoor
98	radiator	√			indoor
99	railing	√	√		no air pollutant emission
100	refrigerator	√			indoor
101	river	√	√		not found in GSV images in Blacksburg

No.	Predictor variables	All variables	Theory-driven: Outdoor variables	Theory-driven: Air pollution-related variables	Reason to include/exclude variables from "Theory-driven: Air pollution related variables"
102	road	√	√	√	traffic
103	rock	√	√	√	dust
104	rug	√			indoor
105	runway	√	√	√	traffic
106	sand	√	√	√	dust
107	sconce	√			indoor
108	screen	√			indoor
109	screen door	√			indoor
110	sculpture	√	√		no air pollutant emission
111	sea	√	√		not found in GSV images in Blacksburg
112	seat	√			indoor
113	shelf	√			indoor
114	ship	√	√		not found in GSV images in Blacksburg
115	shower	√			indoor
116	sidewalk	√	√	√	traffic
117	signboard	√	√	√	traffic
118	sink	√			indoor
119	sky	√	√	√	pollutant dispersion
120	skyscraper	√	√		not found in GSV images in Blacksburg
121	sofa	√			indoor
122	stage	√			indoor
123	stairs	√			indoor
124	stairway	√			indoor
125	step	√			indoor
126	stool	√			indoor
127	stove	√			indoor
128	streetlight	√	√	√	traffic
129	swimming pool	√	√		not found in GSV images in Blacksburg
130	swivel chair	√			indoor
131	table	√			indoor
132	tank	√			indoor
133	television receiver	√			indoor
134	tent	√	√		no air pollutant emission
135	toilet	√			indoor
136	towel	√			indoor
137	tower	√	√		no air pollutant emission
138	trade name	√	√		no air pollutant emission
139	traffic light	√	√	√	traffic
140	tray	√			indoor
141	tree	√	√	√	green space
142	truck	√	√	√	traffic
143	van	√	√	√	traffic
144	vase	√			indoor
145	wall	√	√	√	pollutant dispersion
146	wardrobe	√			indoor
147	washer	√			indoor
148	water	√	√		not found in GSV images in Blacksburg
149	waterfall	√	√		not found in GSV images in Blacksburg
150	windowpane	√	√		no air pollutant emission

Table 2-S2. Descriptive statistics of features extracted from sampled GSV images within 2,000m of air quality monitoring locations

No.	Feature	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Standard Deviation
1	airplane	0.00E+00	0.00E+00	0.00E+00	3.87E-06	0.00E+00	4.49E-03	9.23E-05
2	animal	0.00E+00	0.00E+00	0.00E+00	3.01E-06	0.00E+00	4.93E-03	7.17E-05
3	apparel	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
4	arcade machine	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
5	armchair	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
6	ashcan	0.00E+00	0.00E+00	0.00E+00	1.57E-04	0.00E+00	1.90E-02	8.46E-04
7	awning	0.00E+00	0.00E+00	0.00E+00	2.74E-05	0.00E+00	2.60E-02	4.94E-04
8	bag	0.00E+00	0.00E+00	0.00E+00	3.23E-06	0.00E+00	3.84E-03	6.61E-05
9	ball	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
10	bannister	0.00E+00	0.00E+00	0.00E+00	1.42E-05	0.00E+00	1.32E-02	2.17E-04
11	bar	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
12	barrel	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
13	base	0.00E+00	0.00E+00	0.00E+00	2.30E-05	0.00E+00	1.63E-02	3.54E-04
14	basket	0.00E+00	0.00E+00	0.00E+00	7.22E-07	0.00E+00	2.13E-03	3.08E-05
15	bathtub	0.00E+00	0.00E+00	0.00E+00	1.00E-06	0.00E+00	8.10E-03	8.03E-05
16	bed	0.00E+00	0.00E+00	0.00E+00	2.98E-06	0.00E+00	1.71E-02	1.94E-04
17	bench	0.00E+00	0.00E+00	0.00E+00	4.50E-05	0.00E+00	1.46E-02	3.49E-04
18	bicycle	0.00E+00	0.00E+00	0.00E+00	6.68E-06	0.00E+00	6.95E-03	1.35E-04
19	blanket	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
20	blind	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
21	boat	0.00E+00	0.00E+00	0.00E+00	5.81E-06	0.00E+00	8.19E-03	1.33E-04
22	book	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
23	bookcase	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
24	booth	0.00E+00	0.00E+00	0.00E+00	6.39E-08	0.00E+00	7.00E-04	6.69E-06
25	bottle	0.00E+00	0.00E+00	0.00E+00	1.61E-07	0.00E+00	1.19E-03	1.21E-05
26	box	0.00E+00	0.00E+00	0.00E+00	9.68E-05	0.00E+00	2.39E-02	7.09E-04
27	bridge	0.00E+00	0.00E+00	0.00E+00	3.87E-04	0.00E+00	1.47E-01	3.74E-03
28	buffet	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
29	building	0.00E+00	1.02E-03	6.16E-03	1.93E-02	1.87E-02	5.60E-01	3.91E-02
30	bulletin board	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
31	bus	0.00E+00	0.00E+00	0.00E+00	1.22E-04	0.00E+00	1.45E-01	2.82E-03
32	cabinet	0.00E+00	0.00E+00	0.00E+00	2.48E-06	0.00E+00	7.83E-03	1.18E-04
33	canopy	0.00E+00	0.00E+00	0.00E+00	1.85E-06	0.00E+00	8.43E-03	9.41E-05
34	car	0.00E+00	1.54E-04	1.46E-03	7.82E-03	6.27E-03	1.84E-01	1.68E-02
35	case	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
36	ceiling	0.00E+00	0.00E+00	0.00E+00	3.75E-04	0.00E+00	3.93E-01	7.78E-03
37	chair	0.00E+00	0.00E+00	0.00E+00	6.11E-06	0.00E+00	7.26E-03	1.23E-04
38	chandelier	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
39	chest of drawers	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
40	clock	0.00E+00	0.00E+00	0.00E+00	2.47E-07	0.00E+00	7.59E-04	1.20E-05
41	coffee table	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
42	column	0.00E+00	0.00E+00	0.00E+00	3.53E-05	0.00E+00	4.49E-02	7.36E-04
43	computer	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00

No.	Feature	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Standard Deviation
44	conveyer belt	0.00E+00	0.00E+00	0.00E+00	3.07E-07	0.00E+00	3.36E-03	3.21E-05
45	counter	0.00E+00	0.00E+00	0.00E+00	1.90E-06	0.00E+00	7.52E-03	9.73E-05
46	countertop	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
47	cradle	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
48	crt screen	0.00E+00	0.00E+00	0.00E+00	3.43E-08	0.00E+00	3.76E-04	3.59E-06
49	curtain	0.00E+00	0.00E+00	0.00E+00	4.89E-07	0.00E+00	4.95E-03	4.75E-05
50	cushion	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
51	desk	0.00E+00	0.00E+00	0.00E+00	2.31E-07	0.00E+00	1.34E-03	1.62E-05
52	dirt track	0.00E+00	0.00E+00	0.00E+00	1.10E-04	0.00E+00	6.09E-02	1.49E-03
53	dishwasher	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
54	door	0.00E+00	0.00E+00	0.00E+00	1.26E-05	0.00E+00	2.69E-02	3.60E-04
55	earth	0.00E+00	2.17E-04	5.88E-03	2.65E-02	3.24E-02	3.36E-01	4.52E-02
56	escalator	0.00E+00	0.00E+00	0.00E+00	7.37E-07	0.00E+00	3.90E-03	4.42E-05
57	fan	0.00E+00	0.00E+00	0.00E+00	6.17E-08	0.00E+00	6.64E-04	6.35E-06
58	fence	0.00E+00	0.00E+00	7.63E-05	3.76E-03	2.25E-03	1.98E-01	9.63E-03
59	field	0.00E+00	0.00E+00	0.00E+00	4.70E-03	3.30E-05	3.20E-01	1.99E-02
60	fireplace	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
61	flag	0.00E+00	0.00E+00	0.00E+00	9.14E-06	0.00E+00	2.03E-02	2.19E-04
62	floor	0.00E+00	0.00E+00	0.00E+00	1.46E-04	0.00E+00	1.34E-01	2.50E-03
63	flower	0.00E+00	0.00E+00	0.00E+00	3.90E-05	0.00E+00	2.47E-02	5.15E-04
64	food	0.00E+00	0.00E+00	0.00E+00	1.85E-06	0.00E+00	1.43E-02	1.41E-04
65	fountain	0.00E+00	0.00E+00	0.00E+00	2.20E-06	0.00E+00	5.87E-03	8.58E-05
66	glass	0.00E+00	0.00E+00	0.00E+00	1.84E-09	0.00E+00	2.01E-05	1.92E-07
67	grandstand	0.00E+00	0.00E+00	0.00E+00	8.38E-06	0.00E+00	3.83E-02	4.29E-04
68	grass	0.00E+00	5.63E-02	1.05E-01	1.08E-01	1.53E-01	3.51E-01	6.73E-02
69	hill	0.00E+00	0.00E+00	0.00E+00	2.27E-04	0.00E+00	6.62E-02	2.22E-03
70	hood	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
71	house	0.00E+00	0.00E+00	2.54E-04	7.09E-03	8.95E-03	1.21E-01	1.36E-02
72	hovel	0.00E+00	0.00E+00	0.00E+00	9.78E-06	0.00E+00	2.36E-02	3.69E-04
73	kitchen island	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
74	lake	0.00E+00	0.00E+00	0.00E+00	3.91E-06	0.00E+00	1.59E-02	1.97E-04
75	lamp	0.00E+00	0.00E+00	0.00E+00	2.20E-08	0.00E+00	1.02E-04	1.31E-06
76	land	0.00E+00	0.00E+00	0.00E+00	3.53E-06	0.00E+00	1.24E-02	1.27E-04
77	light	0.00E+00	0.00E+00	0.00E+00	3.18E-07	0.00E+00	1.15E-03	1.72E-05
78	microwave	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
79	minibike	0.00E+00	0.00E+00	0.00E+00	9.24E-06	0.00E+00	1.52E-02	2.48E-04
80	mirror	0.00E+00	0.00E+00	0.00E+00	9.31E-08	0.00E+00	1.02E-03	9.74E-06
81	monitor	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
82	mountain	0.00E+00	0.00E+00	0.00E+00	5.64E-04	2.32E-05	1.25E-01	4.07E-03
83	ottoman	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
84	oven	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
85	painting	0.00E+00	0.00E+00	0.00E+00	5.47E-07	0.00E+00	2.56E-03	2.96E-05
86	palm	0.00E+00	0.00E+00	0.00E+00	9.12E-06	0.00E+00	2.37E-02	3.35E-04
87	path	0.00E+00	0.00E+00	2.44E-06	4.81E-03	2.80E-03	2.13E-01	1.26E-02
88	person	0.00E+00	0.00E+00	0.00E+00	1.17E-04	3.05E-06	3.71E-02	7.35E-04
89	pier	0.00E+00	0.00E+00	0.00E+00	2.23E-08	0.00E+00	2.44E-04	2.33E-06

No.	Feature	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Standard Deviation
90	pillow	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
91	plant	0.00E+00	4.80E-04	4.87E-03	1.41E-02	1.72E-02	2.47E-01	2.31E-02
92	plate	0.00E+00	0.00E+00	0.00E+00	1.16E-07	0.00E+00	6.68E-04	7.14E-06
93	plaything	0.00E+00	0.00E+00	0.00E+00	3.14E-07	0.00E+00	2.18E-03	2.34E-05
94	pole	0.00E+00	0.00E+00	0.00E+00	3.07E-04	5.49E-05	2.80E-02	1.12E-03
95	pool table	0.00E+00	0.00E+00	0.00E+00	6.91E-09	0.00E+00	7.02E-05	6.73E-07
96	poster	0.00E+00	0.00E+00	0.00E+00	4.89E-07	0.00E+00	2.31E-03	2.72E-05
97	pot	0.00E+00	0.00E+00	0.00E+00	8.22E-06	0.00E+00	3.76E-03	8.72E-05
98	radiator	0.00E+00	0.00E+00	0.00E+00	3.27E-07	0.00E+00	3.58E-03	3.42E-05
99	railing	0.00E+00	0.00E+00	0.00E+00	7.83E-05	0.00E+00	4.96E-02	1.01E-03
100	refrigerator	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
101	river	0.00E+00	0.00E+00	0.00E+00	1.22E-04	0.00E+00	5.32E-02	1.50E-03
102	road	0.00E+00	2.08E-01	2.63E-01	2.58E-01	3.09E-01	4.61E-01	7.29E-02
103	rock	0.00E+00	0.00E+00	0.00E+00	2.40E-04	0.00E+00	3.53E-02	1.45E-03
104	rug	0.00E+00	0.00E+00	0.00E+00	5.91E-07	0.00E+00	4.30E-03	4.39E-05
105	runway	0.00E+00	0.00E+00	0.00E+00	4.60E-06	0.00E+00	9.24E-03	1.65E-04
106	sand	0.00E+00	0.00E+00	0.00E+00	8.68E-05	0.00E+00	5.84E-02	1.24E-03
107	sconce	0.00E+00	0.00E+00	0.00E+00	1.18E-07	0.00E+00	8.26E-04	8.52E-06
108	screen	0.00E+00	0.00E+00	0.00E+00	1.68E-07	0.00E+00	1.40E-03	1.40E-05
109	screen door	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
110	sculpture	0.00E+00	0.00E+00	0.00E+00	1.97E-06	0.00E+00	4.96E-03	6.49E-05
111	sea	0.00E+00	0.00E+00	0.00E+00	1.07E-06	0.00E+00	3.86E-03	4.27E-05
112	seat	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
113	shelf	0.00E+00	0.00E+00	0.00E+00	9.48E-08	0.00E+00	1.04E-03	9.92E-06
114	ship	0.00E+00	0.00E+00	0.00E+00	1.51E-06	0.00E+00	1.64E-02	1.56E-04
115	shower	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
116	sidewalk	0.00E+00	0.00E+00	8.39E-04	1.32E-02	1.53E-02	3.56E-01	2.53E-02
117	signboard	0.00E+00	1.47E-05	1.92E-04	6.30E-04	5.93E-04	4.69E-02	1.62E-03
118	sink	0.00E+00	0.00E+00	0.00E+00	4.34E-08	0.00E+00	3.03E-04	3.33E-06
119	sky	7.66E-03	1.98E-01	2.95E-01	2.81E-01	3.76E-01	4.85E-01	1.15E-01
120	skyscraper	0.00E+00	0.00E+00	0.00E+00	6.30E-09	0.00E+00	6.90E-05	6.59E-07
121	sofa	0.00E+00	0.00E+00	0.00E+00	3.24E-08	0.00E+00	3.55E-04	3.39E-06
122	stage	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
123	stairs	0.00E+00	0.00E+00	0.00E+00	9.92E-05	0.00E+00	1.98E-02	6.67E-04
124	stairway	0.00E+00	0.00E+00	0.00E+00	1.58E-05	0.00E+00	1.45E-02	2.40E-04
125	step	0.00E+00	0.00E+00	0.00E+00	2.87E-07	0.00E+00	2.14E-03	2.21E-05
126	stool	0.00E+00	0.00E+00	0.00E+00	2.29E-09	0.00E+00	2.44E-05	2.33E-07
127	stove	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
128	streetlight	0.00E+00	0.00E+00	0.00E+00	1.84E-04	1.08E-04	8.89E-03	4.77E-04
129	swimming pool	0.00E+00	0.00E+00	0.00E+00	1.20E-06	0.00E+00	1.23E-02	1.17E-04
130	swivel chair	0.00E+00	0.00E+00	0.00E+00	1.90E-09	0.00E+00	2.08E-05	1.98E-07
131	table	0.00E+00	0.00E+00	0.00E+00	5.41E-06	0.00E+00	9.82E-03	1.46E-04
132	tank	0.00E+00	0.00E+00	0.00E+00	9.96E-07	0.00E+00	5.30E-03	5.65E-05
133	television receiver	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
134	tent	0.00E+00	0.00E+00	0.00E+00	2.69E-05	0.00E+00	1.37E-01	1.34E-03
135	toilet	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00

No.	Feature	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Standard Deviation
136	towel	0.00E+00	0.00E+00	0.00E+00	8.09E-08	0.00E+00	8.86E-04	8.46E-06
137	tower	0.00E+00	0.00E+00	0.00E+00	3.84E-06	0.00E+00	8.84E-03	1.39E-04
138	trade name	0.00E+00	0.00E+00	0.00E+00	1.64E-05	0.00E+00	1.57E-02	3.40E-04
139	traffic light	0.00E+00	0.00E+00	0.00E+00	2.83E-05	0.00E+00	1.23E-02	3.80E-04
140	tray	0.00E+00	0.00E+00	0.00E+00	3.55E-08	0.00E+00	3.86E-04	3.69E-06
141	tree	2.23E-03	1.16E-01	2.17E-01	2.43E-01	3.48E-01	8.27E-01	1.54E-01
142	truck	0.00E+00	0.00E+00	0.00E+00	3.56E-04	0.00E+00	1.91E-01	4.02E-03
143	van	0.00E+00	0.00E+00	0.00E+00	1.64E-04	0.00E+00	6.99E-02	1.71E-03
144	vase	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
145	wall	0.00E+00	0.00E+00	1.20E-04	2.62E-03	1.52E-03	4.56E-01	1.07E-02
146	wardrobe	0.00E+00	0.00E+00	0.00E+00	3.09E-08	0.00E+00	3.39E-04	3.24E-06
147	washer	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
148	water	0.00E+00	0.00E+00	0.00E+00	4.62E-04	0.00E+00	9.80E-02	3.27E-03
149	waterfall	0.00E+00	0.00E+00	0.00E+00	9.88E-06	0.00E+00	1.64E-02	2.70E-04
150	windowpane	0.00E+00	0.00E+00	0.00E+00	2.12E-06	0.00E+00	5.39E-03	7.06E-05

Table 2-S3. Model results for BC estimation using theory-driven feature selection

No.	All variables		Theory-driven: Outdoor variables		Theory-driven: Air pollution-related variables	
1	variable	coefficient	variable	coefficient	variable	coefficient
2	building_250	1.48E+00	building_250	7.28E-01	building_250	6.38E-01
3	pool table_1750	1.87E+06	sea_1250	3.32E+04	traffic light_1000	1.01E+03
4	stool_1500	-1.16E+07	ship_1000	4.49E+03	plant_500	-4.85E+00
5	ashcan_1250	1.65E+02	road_250	1.37E+00	bridge_250	5.02E+01
6	hovel_2000	-6.51E+03	swimming pool_1250	6.21E+03	road_central	4.18E-01
7	bag_250	-2.10E+03	sand_1750	-1.54E+03	truck_2000	-1.88E+02
8	sand_500	5.17E+02	tent_250	8.37E+01	mountain_750	-1.81E+02
9	ship_250	5.20E+02	fence_1250	-3.27E+01	sand_500	4.34E+02
10	tent_250	5.73E+01	fountain_750	-4.62E+03	land_1000	1.34E+04
11	rock_250	-1.18E+02	awning_500	2.22E+02	streetlight_250	1.91E+02
12	cabinet_250	-1.50E+03	minibike_central	1.15E+02		
13	canopy_1000	4.11E+03	lake_750	2.28E+03		
14	tower_250	1.29E+03	hill_central	7.42E+01		
15	painting_central	1.74E+04	tower_250	1.03E+03		
16	sconce_1000	-7.73E+04	skyscraper_500	-3.03E+05		
17	wall_250	1.46E+01	hovel_250	-1.25E+03		
18	shelf_1500	1.45E+05	river_750	-1.19E+02		
19	waterfall_1750	-5.18E+03	runway_500	-9.76E+02		
20	flower_750	-3.08E+02	mountain_250	-3.75E+01		
21	poster_750	8.71E+03				
Intercept	1.10E+00		8.49E-01		1.03E+00	
adjusted R^2	0.77		0.72		0.63	
10-fold CV R^2	0.42		0.61		0.57	
RMSE ($\mu g/m^3$)	0.20		0.22		0.26	

Table 2-S4. Model results for PN estimation using theory-driven feature selection

No.	All variables		Theory-driven: Outdoor variables		Theory-driven: Air pollution-related variables	
1	variable	coefficient	variable	coefficient	variable	coefficient
2	river_2000	-2.29E+03	river_2000	-2.13E+03	streetlight_250	2.20E+02
3	signboard_250	6.22E+01	signboard_250	6.74E+01	truck_2000	9.36E+02
4	counter_1250	9.11E+03	grass_central	-8.95E-01	bridge_250	1.16E+02
5	awning_250	3.04E+02	ship_750	6.70E+03	traffic light_750	2.43E+03
6	bed_250	-5.99E+05	rock_2000	2.12E+03	grass_1000	4.84E+00
7	ship_750	5.27E+03	building_central	1.52E+00	path_central	-3.70E+00
8	grass_250	-2.11E+00	tent_250	2.42E+02	building_central	1.52E+00
9	food_750	-5.35E+03	sand_1500	-2.19E+03	sand_250	-5.50E+02
10	conveyer belt_1250	-5.83E+04	dirt track_1000	5.75E+02	plant_250	9.69E+00
11	building_central	1.19E+00	wall_250	2.23E+01	house_250	-3.31E+00
12	tent_250	1.93E+02	canopy_1000	5.84E+03	wall_250	2.38E+01
13	path_central	-3.41E+00	animal_250	9.06E+03	land_250	-3.93E+03
14	light_1500	9.72E+04			road_250	1.35E+00
15	desk_central	3.40E+04			car_250	4.73E+00
16	pot_central	2.58E+02			flower_1000	8.22E+02
17					mountain_250	6.32E+01
18					runway_1750	4.88E+03
Intercept	8.97E+00		8.44E+00		7.03E+00	
adjusted R^2	0.77		0.74		0.73	
10-fold CV R^2	0.57		0.65		0.65	
RMSE (pt/cm^3)	821		916		954	

Table 2-S5. Number of input predictor variables offered to the buffer-feature models using different feature selection methods

Feature selection method	Variable types	Number of variables offered
None	All variables	1350
Theory-driven	Outdoor variables	513
Theory-driven	Air pollution related variables	279
Data-driven	All variables	252-659
Integrated	Outdoor variables	216-478
Integrated	Air pollution related variables	188-278

Table 2-S6. Results of the best BC and PN buffer-feature models using data-driven feature selection

No.	BC		PN	
1	variable	coefficient	variable	coefficient
2	building_250	5.23E-01	river_2000	-2.50E+03
3	traffic light_1000	9.09E+02	signboard_250	5.49E+01
4	plant_500	-7.10E+00	counter_1250	1.19E+04
5	ashcan_1250	6.36E+02	awning_250	3.49E+02
6	road_central	1.94E-01	grandstand_1250	-3.98E+03
7	door_500	-9.14E+02	grass_250	-2.51E+00
8	water_1250	1.18E+02	field_500	1.81E+01
9	railing_1000	-5.17E+02	sand_250	-4.74E+02
10	signboard_500	1.39E+02	path_central	-3.08E+00
11	rock_250	-1.43E+02	house_central	-2.26E+00
12	streetlight_central	2.69E+01	canopy_1000	1.04E+04
13	grandstand_1750	3.49E+03	chair_2000	-2.08E+04
14	fence_750	-2.37E+01	building_central	8.54E-01
15	sand_500	3.39E+02	bathtub_1500	-2.69E+04
16	truck_1000	-1.10E+02	pot_central	2.40E+02
17	grass_250	-1.10E+00	door_500	-1.45E+03
18	hill_1500	-8.43E+01	wall_250	1.33E+01
19			tent_250	8.21E+01
20			van_central	-1.12E+01
21			land_1500	-8.59E+03
22			hovel_500	-9.56E+02
23			streetlight_central	4.52E+01
Intercept	1.12E+00		9.21E+00	
adjusted R^2	0.72		0.80	
10-fold CV R^2	0.60		0.70	
RMSE	0.22 ($\mu g/m^3$)		807 (pt/cm^3)	

Table 2-S7. Summary of buffer-feature models and reduced buffer-feature models for air pollution estimation (all models used the integrated theory- and data-driven feature selection and the air pollution-related predictor variables)

Pollutants	Model type	Predictor variables	Number of selected variables	Adjusted R^2	10-fold CV R^2	RMSE
BC ($\mu g/m^3$)	Buffer-feature model	central (50m) + buffer (250-2000m)	9~12	0.57~0.64	0.50~0.57	0.25-0.28
	Reduced buffer-feature models	buffer (250-2000m)	9~15	0.59~0.65	0.48~0.59	0.25-0.27
		buffer (500-2000m)	11~12	0.59~0.60	0.51~0.53	0.27-0.27
		buffer (750-2000m)	7~12	0.51~0.58	0.48~0.51	0.27-0.30
		buffer (1000-2000m)	4~12	0.36~0.56	0.34~0.50	0.28-0.34
PN (pt/cm^3)	Buffer-feature model	central (50m) + buffer (250-2000m)	11~17	0.65~0.73	0.61~0.66	941-1030
	Reduced buffer-feature models	buffer (250-2000m)	11~20	0.66~0.72	0.60~0.64	980-1076
		buffer (500-2000m)	10~14	0.55~0.62	0.48~0.56	1127-1213
		buffer (750-2000m)	6~9	0.48~0.53	0.42~0.49	1264-1329
		buffer (1000-2000m)	4~6	0.35~0.47	0.32~0.43	1324-1475

Table 2-S8. Results of the Global Moran's I test for 4 types of GSV-only buffer-feature models

Pollutants	Models	Inverse Distance				Inverse Distance Squared				Fixed Distance			
		300m		500m		300m		500m		300m		500m	
		Moran's I	p-value	Moran's I	p-value	Moran's I	p-value	Moran's I	p-value	Moran's I	p-value	Moran's I	p-value
BC ($\mu g/m^3$)	All variables	0.15	0.00	0.09	0.00	0.23	0.00	0.19	0.00	0.07	0.01	0.02	0.19
	Theory-driven: Air pollution related variables	0.32	0.00	0.22	0.00	0.40	0.00	0.34	0.00	0.25	0.00	0.12	0.00
	Data-driven: All variables	0.18	0.00	0.09	0.00	0.26	0.00	0.21	0.00	0.09	0.00	0.00	0.86
	Integrated: air pollution related variables	0.32	0.00	0.22	0.00	0.40	0.00	0.34	0.00	0.25	0.00	0.12	0.00
PN (pt/cm^3)	All variables	0.15	0.00	0.10	0.00	0.21	0.00	0.18	0.00	0.09	0.00	0.04	0.02
	Theory-driven: Air pollution related variables	0.21	0.00	0.15	0.00	0.25	0.00	0.22	0.00	0.16	0.00	0.09	0.00
	Data-driven: All variables	0.06	0.03	0.02	0.22	0.11	0.00	0.09	0.01	0.02	0.46	-0.03	0.20
	Integrated: air pollution related variables	0.22	0.00	0.16	0.00	0.26	0.00	0.22	0.00	0.17	0.00	0.10	0.00

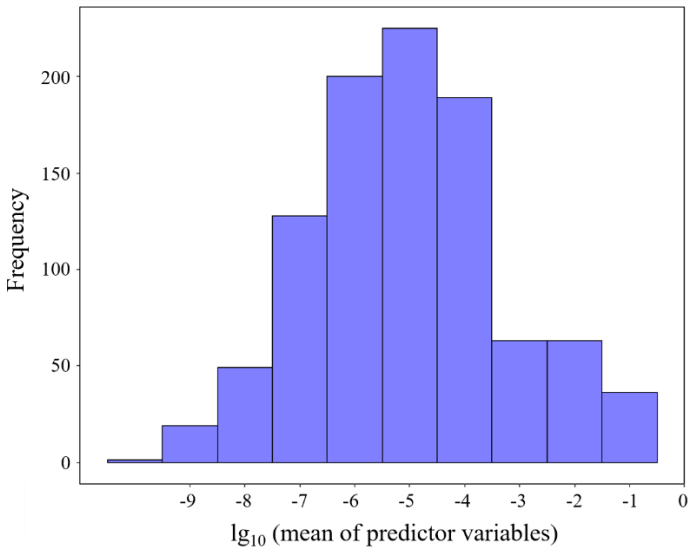

Figure 2-S1. Histogram of the mean of 1,350 predictor variables. The values were log-transformed to determine the magnitude of thresholds for data filtering.

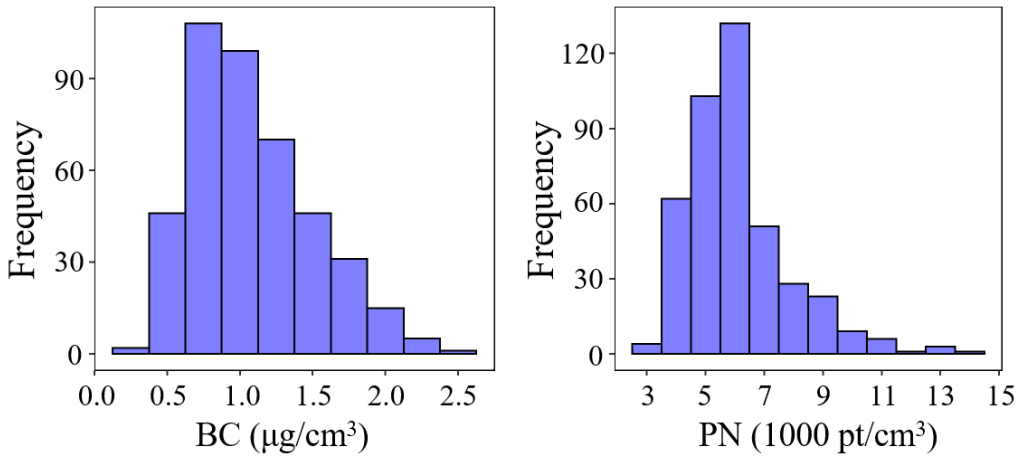

Figure 2-S2. Histograms of the multiplicative background adjusted BC and PN concentrations.

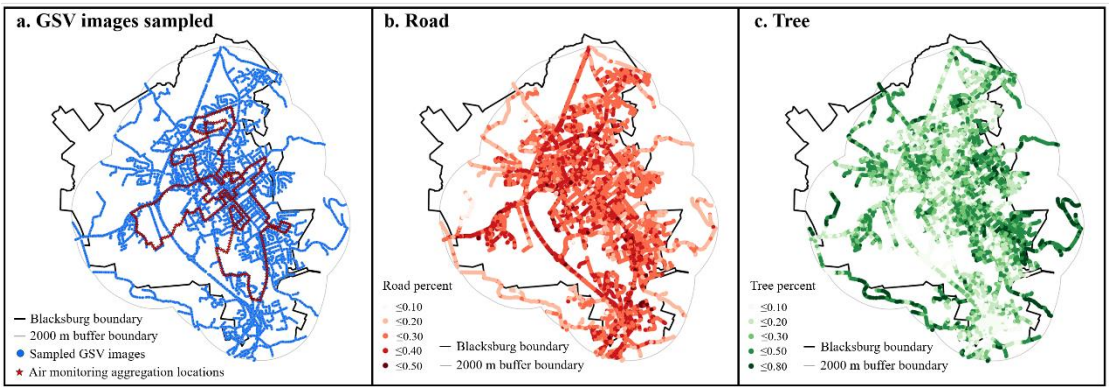

Figure 2-S3. Spatial distribution of the sampled GSV images and examples of built and natural environment characteristics derived from the images. (a) GSV images sampled, (b) road distribution, and (c) tree cover distribution.

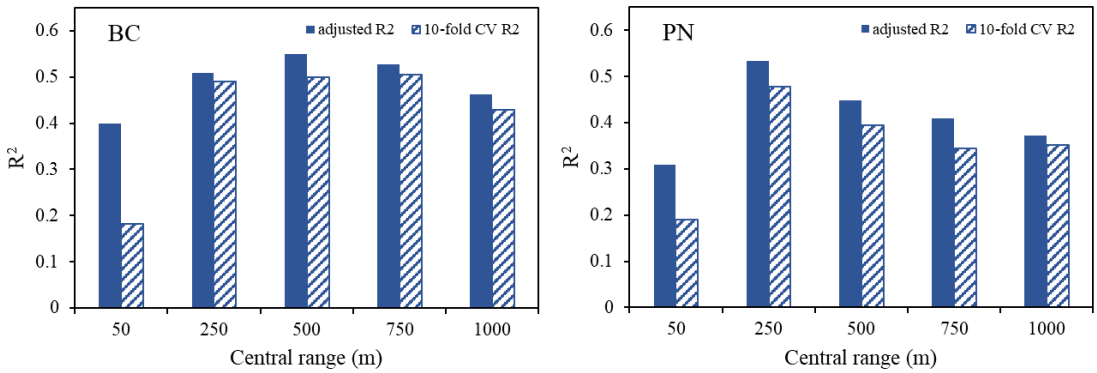

Figure 2-S4. Comparison of central-feature models with different distance ranges. The predictor variables for both models were "Theory-driven: Air pollution-related variables".

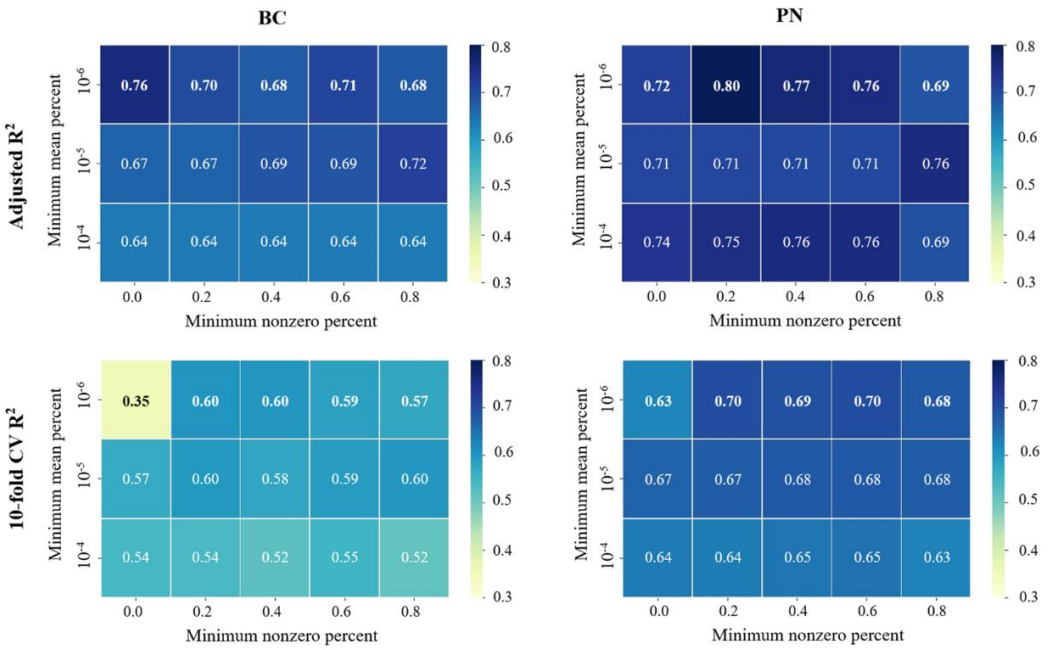

Figure 2-S5. Model performance of buffer-feature models among different scenarios using data-driven feature selection. The upper panel shows adjusted R^2 of the models and the bottom panel shows 10-fold cross validation R^2. Analogous plots for the central-feature models are in **Figure 2-S6**.

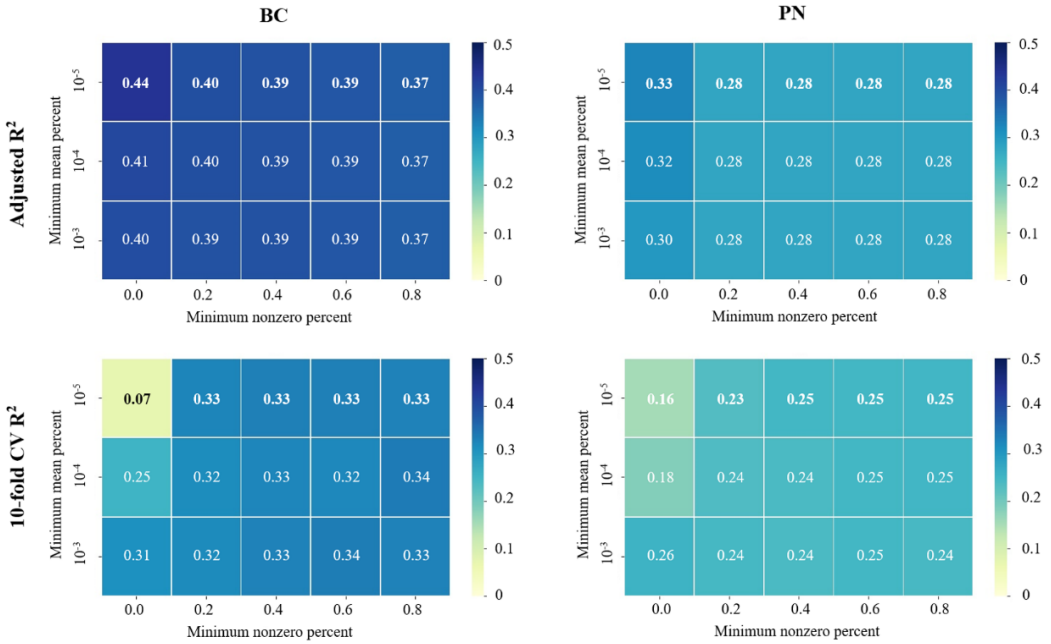

Figure 2-S6. Model performance of the central-feature models among different scenarios using the data-driven feature selection. The upper panel shows adjusted R^2 of the models and the bottom panel shows 10-fold cross validation R^2.

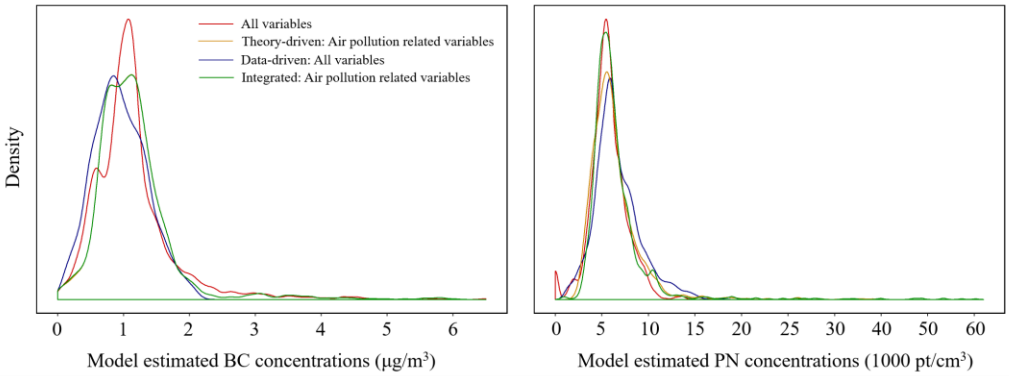

Figure 2-S7. Density plots of estimated BC and PN concentrations using 4 types of GSV-only models. All distributions were similar and highly positive-skewed.

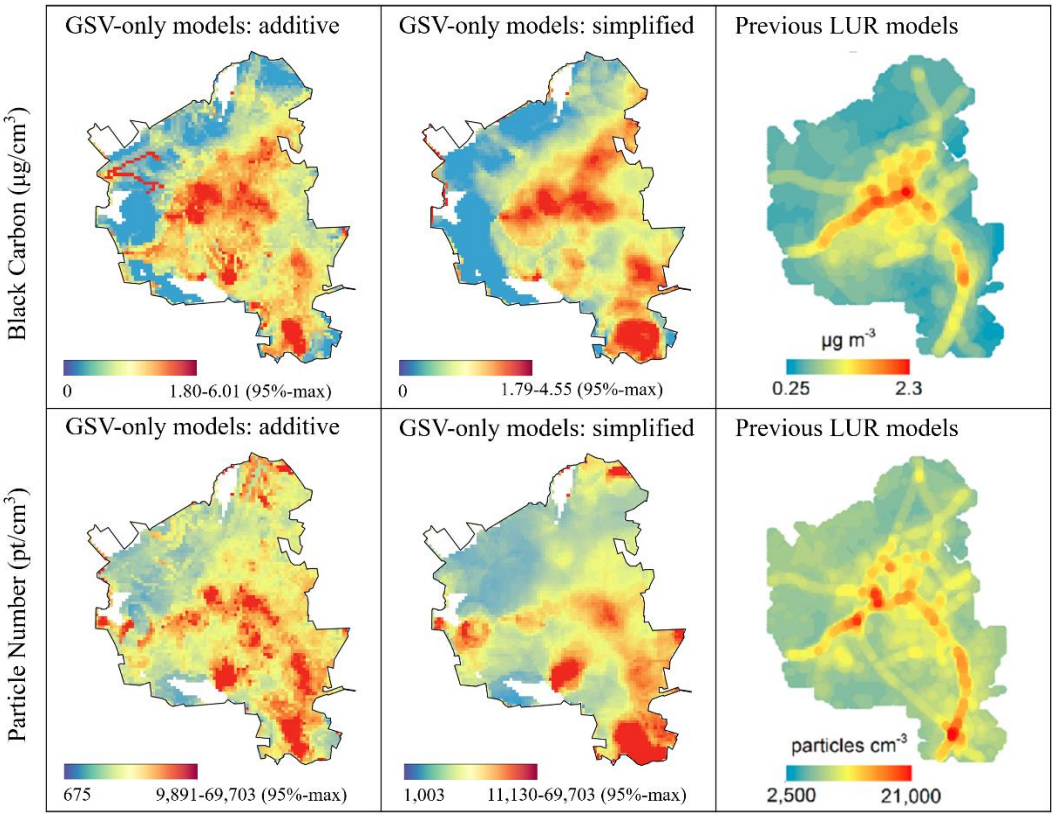

Figure 2-S8. Comparison among GSV-only model estimates and previous LUR model estimates. Figures from the left to right columns were estimates using reduced buffer-feature models with the additive method, estimates using reduced models with the simplified method, and estimates using previous LUR models with traditional predictor variables. Due to highly right-skewed distributions of the predicted concentrations, the color bar was scaled from minimum to the 95th percentile to better visualize spatial patterns of the data.

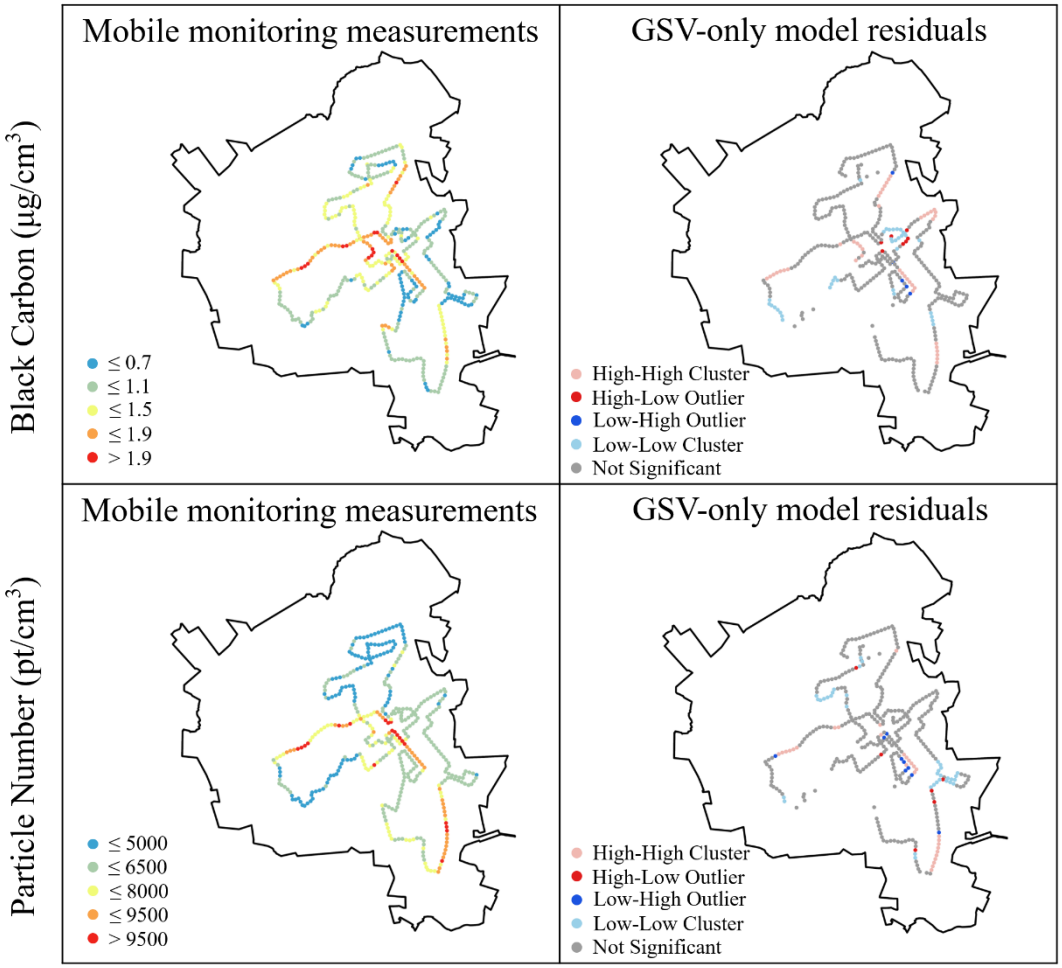

Figure 2-S9. The spatial distribution of mobile monitoring measurements and the results of LISA test for GSV-only model residuals. We used the integrated buffer-feature models with air pollution-related variables to illustrate the results of the LISA analysis (500m threshold). High-High and Low-Low cluster represent statistically significant clusters of underestimates and overestimates of particulate air pollution – these clusters are the most important for the models presented in this paper. A High-Low (Low-High) outlier are where a high (low) residual value is surrounded primarily by low (high) residual values.

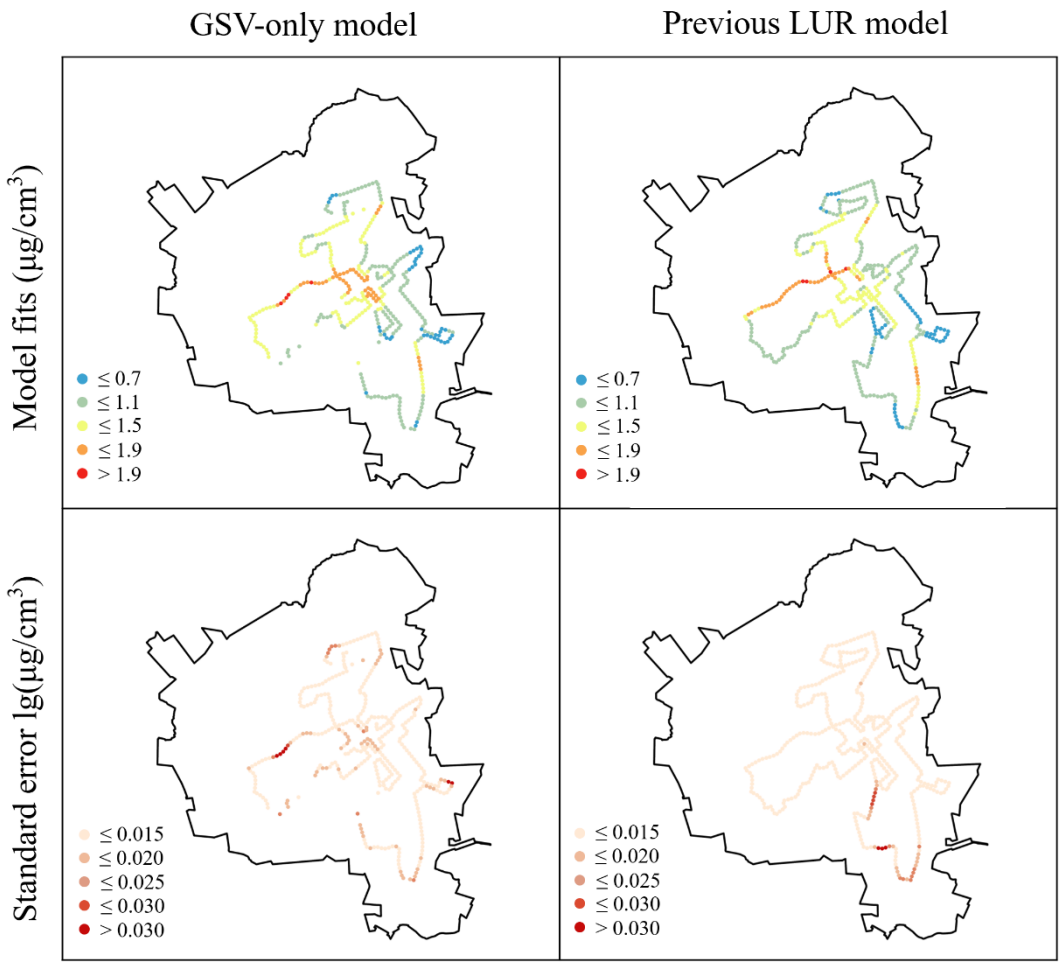

Figure 2-S10. Comparison of BC model uncertainty for the GSV-only model presented in this paper and the previously published LUR model with traditional GIS predictor variables. We used the integrated buffer-feature model with air pollution-related variables as an example. The upper panel shows model fits; the bottom panel shows standard error for model prediction. The model fits are transformed back to original values for illustration.

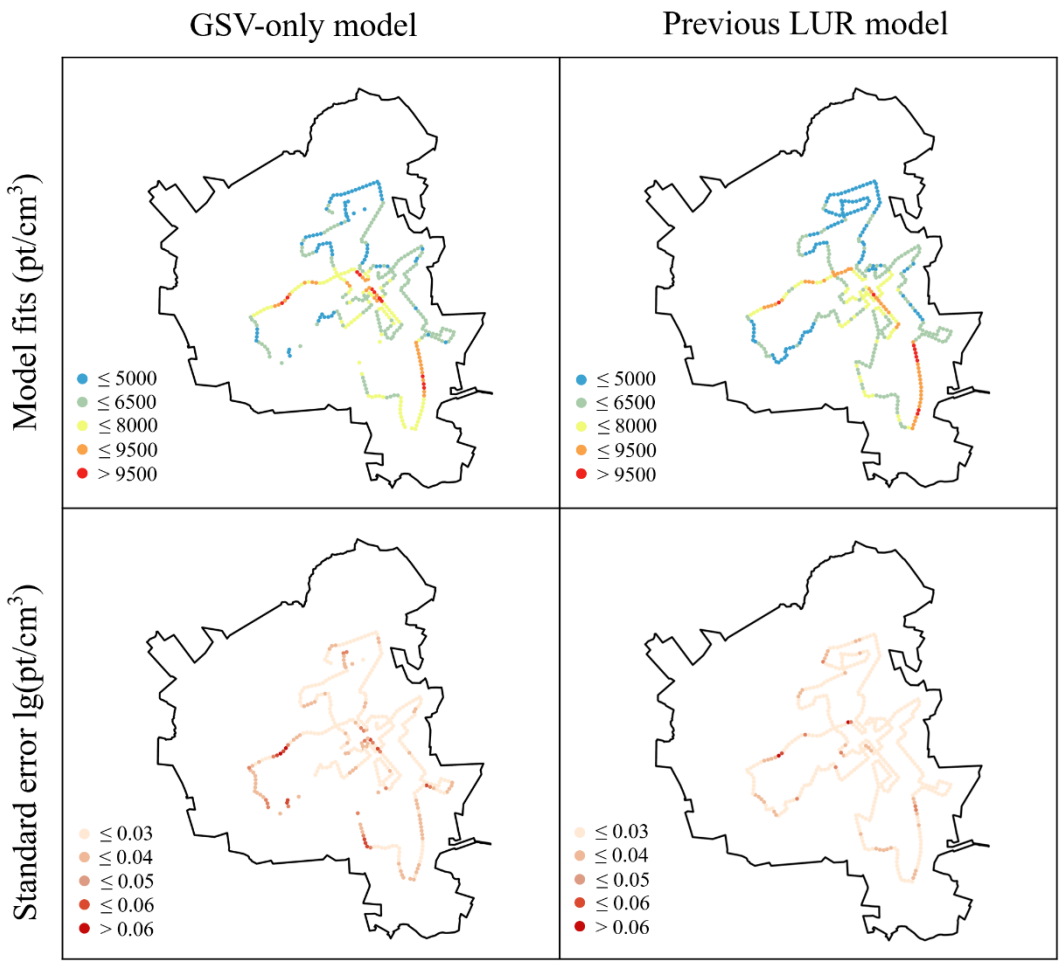

Figure 2-S11. Comparison of PN model uncertainty for the GSV-only model presented in this paper and the previously published LUR model with traditional GIS predictor variables. We used the integrated buffer-feature model with air pollution-related variables as an example. The upper panel shows model fits; the bottom panel shows standard error for model prediction. The model fits are transformed back to original values for illustration.

2.6 References

1. Pope, C. A.; Ezzati, M.; Dockery, D. W., Fine-Particulate Air Pollution and Life Expectancy in the United States. *N. Engl. J. Med.* **2009,** *360*, 376-386.

2. Hoek, G.; Krishnan, R. M.; Beelen, R.; Peters, A.; Ostro, B.; Brunekreef, B.; Kaufman, J. D., Long-term air pollution exposure and cardio- respiratory mortality: a review. *Environ. Health* **2013,** *12*, 43.

3. Kim, K. H.; Kabir, E.; Kabir, S., A review on the human health impact of airborne particulate matter. *Environ. Int.* **2015,** *74*, 136-143.

4. Shi, L. H.; Zanobetti, A.; Kloog, I.; Coull, B. A.; Koutrakis, P.; Melly, S. J.; Schwartz, J. D., Low-Concentration $PM_{2.5}$ and Mortality: Estimating Acute and Chronic Effects in a Population-Based Study. *Environ. Health Perspect.* **2016,** *124*, 46-52.

5. 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.; Hamburg, S. P., High-Resolution Air Pollution Mapping with Google Street View Cars: Exploiting Big Data. *Environ. Sci. Technol.* **2017,** *51*, 6999-7008.

6. 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.; Alvarez, R. A., Characterizing Elevated Urban Air Pollutant Spatial Patterns with Mobile Monitoring in Houston, Texas. *Environ. Sci. Technol.* **2020,** *54*, 2133-2142.

7. Alexeeff, S. E.; Roy, A.; Shan, J.; Liu, X.; Messier, K.; Apte, J. S.; Portier, C.; Sidney, S.; Van Den Eeden, S. K., High-resolution mapping of traffic related air pollution with Google street view cars and incidence of cardiovascular events within neighborhoods in Oakland, CA. *Environ. Health* **2018,** *17*, 38.

8. Lee, H. J., Benefits of High Resolution $PM_{2.5}$ Prediction using Satellite MAIAC

AOD and Land Use Regression for Exposure Assessment: California Examples. *Environ. Sci. Technol.* **2019,** *53*, 12774-12783.

9. Faridi, S.; Shamsipour, M.; Krzyzanowski, M.; Kunzli, N.; Amini, H.; Azimi, F.; Malkawi, M.; Momeniha, F.; Gholampour, A.; Hassanvand, M. S.; Naddafi, K., Long-term trends and health impact of $PM_{2.5}$ and O_3 in Tehran, Iran, 2006-2015. *Environ. Int.* **2018,** *114*, 37-49.

10. Vineis, P.; Hoek, G.; Krzyzanowski, M.; Vigna-Taglianti, F.; Veglia, F.; Airoldi, L.; Autrup, H.; Dunning, A.; Garte, S.; Hainaut, P.; Malaveille, C.; Matullo, G.; Overvad, K.; Raaschou-Nielsen, O.; Clavel-Chapelon, F.; Linseisen, J.; Boeing, H.; Trichopoulou, A.; Palli, D.; Peluso, M.; Krogh, V.; Tumino, R.; Panico, S.; Bueno-De-Mesquita, H. B.; Peeters, P. H.; Lund, E. E.; Gonzalez, C. A.; Martinez, C.; Dorronsoro, M.; Barricarte, A.; Cirera, L.; Quiros, J. R.; Berglund, G.; Forsberg, B.; Day, N. E.; Key, T. J.; Saracci, R.; Kaaks, R.; Riboli, E., Air pollution and risk of lung cancer in a prospective study in Europe. *Int. J. Cancer* **2006,** *119*, 169-174.

11. Ma, Z. W.; Hu, X. F.; Sayer, A. M.; Levy, R.; Zhang, Q.; Xue, Y. G.; Tong, S. L.; Bi, J.; Huang, L.; Liu, Y., Satellite-Based Spatiotemporal Trends in $PM_{2.5}$ Concentrations: China, 2004-2013. *Environ. Health Perspect.* **2016,** *124*, 184-192.

12. Sickles, J. E.; Shadwick, D. S., Air quality and atmospheric deposition in the eastern US: 20 years of change. *Atmos. Chem. Phys.* **2015,** *15*, 173-197.

13. Larson, T.; Henderson, S. B.; Brauer, M., Mobile Monitoring of Particle Light Absorption Coefficient in an Urban Area as a Basis for Land Use Regression. *Environ. Sci. Technol.* **2009,** *43*, 4672-4678.

14. Solomon, P. A.; Crumpler, D.; Flanagan, J. B.; Jayanty, R. K. M.; Rickman, E. E.; McDade, C. E., US National $PM_{2.5}$ Chemical Speciation Monitoring Networks-CSN and IMPROVE: Description of networks. *J. Air Waste Manag. Assoc.* **2014,** *64*, 1410-

1438.

15. Kumar, P.; Morawska, L.; Martani, C.; Biskos, G.; Neophytou, M.; Di Sabatino, S.; Bell, M.; Norford, L.; Britter, R., The rise of low-cost sensing for managing air pollution in cities. *Environ. Int.* **2015,** *75*, 199-205.

16. Minet, L.; Liu, R.; Valois, M. F.; Xu, J. S.; Weichenthal, S.; Hatzopoulou, M., Development and Comparison of Air Pollution Exposure Surfaces Derived from On-Road Mobile Monitoring and Short-Term Stationary Sidewalk Measurements. *Environ. Sci. Technol.* **2018,** *52*, 3512-3519.

17. Eeftens, M.; Beelen, R.; de Hoogh, K.; Bellander, T.; Cesaroni, G.; Cirach, M.; Declercq, C.; Dedele, A.; Dons, E.; de Nazelle, A.; Dimakopoulou, K.; Eriksen, K.; Falq, G.; Fischer, P.; Galassi, C.; Grazuleviciene, R.; Heinrich, J.; Hoffmann, B.; Jerrett, M.; Keidel, D.; Korek, M.; Lanki, T.; Lindley, S.; Madsen, C.; Molter, A.; Nador, G.; Nieuwenhuijsen, M.; Nonnemacher, M.; Pedeli, X.; Raaschou-Nielsen, O.; Patelarou, E.; Quass, U.; Ranzi, A.; Schindler, C.; Stempfelet, M.; Stephanou, E.; Sugiri, D.; Tsai, M. Y.; Yli-Tuomi, T.; Varro, M. J.; Vienneau, D.; von Klot, S.; Wolf, K.; Brunekreef, B.; Hoek, G., Development of Land Use Regression Models for $PM_{2.5}$, $PM_{2.5}$ Absorbance, PM_{10} and PMcoarse in 20 European Study Areas; Results of the ESCAPE Project. *Environ. Sci. Technol.* **2012,** *46*, 11195-11205.

18. Schmitz, O.; Beelen, R.; Strak, M.; Hoek, G.; Soenario, I.; Brunekreef, B.; Vaartjes, I.; Dijst, M. J.; Grobbee, D. E.; Karssenberg, D., Data Descriptor: High resolution annual average air pollution concentration maps for the Netherlands. *Sci. Data* **2019,** *6*, 190035.

19. Larkin, A.; Geddes, J. A.; Martin, R. V.; Xiao, Q. Y.; Liu, Y.; Marshall, J. D.; Brauer, M.; Hystad, P., Global Land Use Regression Model for Nitrogen Dioxide Air Pollution. *Environ. Sci. Technol.* **2017,** *51*, 6957-6964.

20. Shi, Y.; Lau, K. K. L.; Ng, E., Developing Street-Level $PM_{2.5}$ and PM_{10} Land Use Regression Models in High-Density Hong Kong with Urban Morphological Factors. *Environ. Sci. Technol.* **2016,** *50*, 8178-8187.

21. Hoek, G.; Beelen, R.; de Hoogh, K.; Vienneau, D.; Gulliver, J.; Fischer, P.; Briggs, D., A review of land-use regression models to assess spatial variation of outdoor air pollution. *Atmos. Environ.* **2008,** *42*, 7561-7578.

22. Hankey, S.; 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.* **2015,** *49*, 9194-9202.

23. 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.; Apte, J. S., Mapping Air Pollution with Google Street View Cars: Efficient Approaches with Mobile Monitoring and Land Use Regression. *Environ. Sci. Technol.* **2018,** *52*, 12563-12572.

24. Hankey, S.; Sforza, P.; Pierson, M., Using Mobile Monitoring to Develop Hourly Empirical Models of Particulate Air Pollution in a Rural Appalachian Community. *Environ. Sci. Technol.* **2019,** *53*, 4305-4315.

25. Hatzopoulou, M.; Valois, M. F.; Levy, I.; Mihele, C.; Lu, G.; Bagg, S.; Minet, L.; Brook, J., Robustness of Land-Use Regression Models Developed from Mobile Air Pollutant Measurements. *Environ. Sci. Technol.* **2017,** *51*, 3938-3947.

26. Rundle, A. G.; Bader, M. D. M.; Richards, C. A.; Neckerman, K. M.; Teitler, J. O., Using Google Street View to Audit Neighborhood Environments. *Am. J. Prev. Med.* **2011,** *40*, 94-100.

27. Hong, K. Y.; Pinheiro, P. O.; Minet, L.; Hatzopoulou, M.; Weichenthal, S., Extending the spatial scale of land use regression models for ambient ultrafine particles

using satellite images and deep convolutional neural networks. *Environ. Res.* **2019,** *176*, 108513.

28. Rzotkiewicz, A.; Pearson, A. L.; Dougherty, B. V.; Shortridge, A.; Wilson, N., Systematic review of the use of Google Street View in health research: Major themes, strengths, weaknesses and possibilities for future research. *Health Place* **2018,** *52*, 240-246.

29. Charreire, H.; Mackenbach, J. D.; Ouasti, M.; Lakerveld, J.; Compernolle, S.; Ben-Rebah, M.; McKee, M.; Brug, J.; Rutter, H.; Oppert, J. M., Using remote sensing to define environmental characteristics related to physical activity and dietary behaviours: A systematic review (the SPOTLIGHT project). *Health Place* **2014,** *25*, 1-9.

30. Gebru, T.; Krause, J.; Wang, Y. L.; Chen, D. Y.; Deng, J.; Aiden, E. L.; Li, F. F., Using deep learning and Google Street View to estimate the demographic makeup of neighborhoods across the United States. *Proc. Natl. Acad. Sci. U. S. A.* **2017,** *114*, 13108-13113.

31. Gong, F. Y.; Zeng, Z. C.; Zhang, F.; Li, X. J.; Ng, E.; Norford, L. K., Mapping sky, tree, and building view factors of street canyons in a high-density urban environment. *Build. Environ.* **2018,** *134*, 155-167.

32. Zhao, H. S.; Shi, J. P.; Qi, X. J.; Wang, X. G.; Jia, J. Y.; Ieee, Pyramid Scene Parsing Network. *Proc. IEEE Conf. Comput. Vis. Pattern Recog.* **2017,** pp 6230-6239.

33. Li, X. J.; Zhang, C. R.; Li, W. D.; Ricard, R.; Meng, Q. Y.; Zhang, W. X., Assessing street-level urban greenery using Google Street View and a modified green view index. *Urban For. Urban Green.* **2015,** *14*, 675-685.

34. Stubbings, P.; Peskett, J.; Rowe, F.; Arribas-Bel, D., A Hierarchical Urban Forest Index Using Street-Level Imagery and Deep Learning. *Remote Sens.* **2019,** *11*,

1395.

35. Seiferling, I.; Naik, N.; Ratti, C.; Proulx, R., Green streets - Quantifying and mapping urban trees with street-level imagery and computer vision. *Landsc. Urban Plan.* **2017,** *165*, 93-101.

36. Bader, M. D. M.; Mooney, S. J.; Lee, Y. J.; Sheehan, D.; Neckerman, K. M.; Rundle, A. G.; Teitler, J. O., Development and deployment of the Computer Assisted Neighborhood Visual Assessment System (CANVAS) to measure health-related neighborhood conditions. *Health Place* **2015,** *31*, 163-172.

37. Li, X. J.; Zhang, C. R.; Li, W. D., Building block level urban land-use information retrieval based on Google Street View images. *GISci. Remote Sens.* **2017,** *54*, 819-835.

38. Naik, N.; Kominers, S. D.; Raskar, R.; Glaeser, E. L.; Hidalgo, C. A., Computer vision uncovers predictors of physical urban change. *Proc. Natl. Acad. Sci. U. S. A.* **2017,** *114*, 7571-7576.

39. Ganji, A.; Minet, L.; Weichenthal, S.; Hatzopoulou, M., Predicting traffic-related air pollution using feature extraction from built environment images. *Environ. Sci. Technol.* **2020**. DOI: 10.1021/acs.est.0c00412.

40. Raghu, V. K.; Ge, X. Y.; Chrysanthis, P. K.; Benos, P. V.; Ieee, Integrated Theory- and Data-driven Feature Selection in Gene Expression Data Analysis. *2017 Ieee 33rd International Conference on Data Engineering*, **2017,** pp 1525-1532.

41. Karpatne, A.; Atluri, G.; Faghmous, J. H.; Steinbach, M.; Banerjee, A.; Ganguly, A.; Shekhar, S.; Samatova, N.; Kumar, V., Theory-Guided Data Science: A New Paradigm for Scientific Discovery from Data. *IEEE Trans. Knowl. Data Eng.* **2017,** *29*, 2318-2331.

42. Bond, T. C.; Doherty, S. J.; Fahey, D. W.; Forster, P. M.; Berntsen, T.; DeAngelo,

B. J.; Flanner, M. G.; Ghan, S.; Karcher, B.; Koch, D.; Kinne, S.; Kondo, Y.; Quinn, P. K.; Sarofim, M. C.; Schultz, M. G.; Schulz, M.; Venkataraman, C.; Zhang, H.; Zhang, S.; Bellouin, N.; Guttikunda, S. K.; Hopke, P. K.; Jacobson, M. Z.; Kaiser, J. W.; Klimont, Z.; Lohmann, U.; Schwarz, J. P.; Shindell, D.; Storelvmo, T.; Warren, S. G.; Zender, C. S., Bounding the role of black carbon in the climate system: A scientific assessment. *J. Geophys. Res.-Atmos.* **2013,** *118*, 5380-5552.

43. Klimont, Z.; Kupiainen, K.; Heyes, C.; Purohit, P.; Cofala, J.; Rafaj, P.; Borken-Kleefeld, J.; Schopp, W., Global anthropogenic emissions of particulate matter including black carbon. *Atmos. Chem. Phys.* **2017,** *17*, 8681-8723.

44. Hu, W.; Hu, M.; Hu, W. W.; Zheng, J.; Chen, C.; Wu, Y. S.; Guo, S., Seasonal variations in high time-resolved chemical compositions, sources, and evolution of atmospheric submicron aerosols in the megacity Beijing. *Atmos. Chem. Phys.* **2017,** *17*, 9979-10000.

45. Nyhan, M.; Grauwin, S.; Britter, R.; Misstear, B.; McNabola, A.; Laden, F.; Barrett, S. R. H.; Ratti, C., "Exposure Track" The Impact of Mobile-Device-Based Mobility Patterns on Quantifying Population Exposure to Air Pollution. *Environ. Sci. Technol.* **2016,** *50*, 9671-9681.

46. Stafoggia, M.; Schwartz, J.; Badaloni, C.; Bellander, T.; Alessandrini, E.; Cattani, G.; de' Donato, F.; Gaeta, A.; Leone, G.; Lyapustin, A.; Sorek-Hamer, M.; de Hoogh, K.; Di, Q.; Forastiere, F.; Kloog, I., Estimation of daily PM_{10} concentrations in Italy (2006-2012) using finely resolved satellite data, land use variables and meteorology. *Environ. Int.* **2017,** *99*, 234-244.

47. He, Q. Q.; Huang, B., Satellite-based mapping of daily high-resolution ground $PM_{2.5}$ in China via space-time regression modeling. *Remote Sens. Environ.* **2018,** *206*, 72-83.

48. Di, Q.; Kloog, I.; Koutrakis, P.; Lyapustin, A.; Wang, Y. J.; Schwartz, J., Assessing $PM_{2.5}$ Exposures with High Spatiotemporal Resolution across the Continental United States. *Environ. Sci. Technol.* **2016,** *50*, 4712-4721.

49. Lv, B. L.; Hu, Y. T.; Chang, H. H.; Russell, A. G.; Bai, Y. Q., Improving the Accuracy of Daily $PM_{2.5}$ Distributions Derived from the Fusion of Ground-Level Measurements with Aerosol Optical Depth Observations, a Case Study in North China. *Environ. Sci. Technol.* **2016,** *50*, 4752-4759.

50. Lefevre, S.; Tuia, D.; Wegner, J. D.; Produit, T.; Nassar, A. S., Toward Seamless Multiview Scene Analysis From Satellite to Street Level. *Proc. IEEE* **2017,** *105*, 1884-1899.

51. Weichenthal, S.; Hatzopoulou, M.; Brauer, M., A picture tells a thousand...exposures: Opportunities and challenges of deep learning image analyses in exposure science and environmental epidemiology. *Environ. Int.* **2019,** *122*, 3-10.

52. Guo, Y. M.; Liu, Y.; Oerlemans, A.; Lao, S. Y.; Wu, S.; Lew, M. S., Deep learning for visual understanding: A review. *Neurocomputing* **2016,** *187*, 27-48.

53. Ma, L.; Liu, Y.; Zhang, X. L.; Ye, Y. X.; Yin, G. F.; Johnson, B. A., Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS-J. Photogramm. Remote Sens.* **2019,** *152*, 166-177.

54. Lim, C. C.; Kim, H.; Vilcassim, M. J. R.; Thurston, G. D.; Gordon, T.; Chen, L. C.; Lee, K.; Heimbinder, M.; Kim, S. Y., Mapping urban air quality using mobile sampling with low-cost sensors and machine learning in Seoul, South Korea. *Environ. Int.* **2019,** *131*, 105022.

55. Morawska, L.; Thai, P. K.; Liu, X. T.; Asumadu-Sakyi, A.; Ayoko, G.; Bartonova, A.; Bedini, A.; Chai, F. H.; Christensen, B.; Dunbabin, M.; Gao, J.; Hagler, G. S. W.; Jayaratne, R.; Kumar, P.; Lau, A. K. H.; Louie, P. K. K.; Mazaheri, M.; Ning,

Z.; Motta, N.; Mullins, B.; Rahman, M. M.; Ristovski, Z.; Shafiei, M.; Tjondronegoro, D.; Westerdahl, D.; Williams, R., Applications of low-cost sensing technologies for air quality monitoring and exposure assessment: How far have they gone? *Environ. Int.* **2018,** *116*, 286-299.

Chapter 3 National Land Use Regression Model for NO_2 using Street View Imagery and Satellite Observations

*Meng Qi[1], Kuldeep Dixit[1], Julian D. Marshall[2], Wenwen Zhang[3], Steve Hankey[1]**

[1]School of Public and International Affairs, Virginia Tech, Blacksburg 24061 Virginia, United States

[2]Department of Civil & Environmental Engineering, University of Washington, Seattle 98195 Washington, United States

[3]Edward J. Bloustein School of Planning and Public Policy, Rutgers University, New Brunswick 08901 New Jersey, United States

* Corresponding author, School of Public and International Affairs, Virginia Tech

Email: hankey@vt.edu. Phone: 540.231.7508

The authors declare no competing interest.

ABSTRACT: Land Use Regression (LUR) models are widely applied to estimate intra-urban air pollution concentrations. National-scale LURs typically employ predictors from multiple curated geodatabases at neighborhood scales. In this study, we instead developed national NO_2 models relying on innovative street-level predictors extracted from Google Street View [GSV] imagery. Using machine learning (random forest), we developed two types of models: (1) GSV-only models, which use only GSV features, and (2) GSV+OMI models, which also include satellite observations of NO_2. Our results suggest that street view imagery alone may provide sufficient information to explain NO_2 variation. Satellite observations can improve model performance, but the contribution decreases as more images are available. Random 10-fold cross validation R^2 of our best models were 0.88 (GSV-only) and 0.91 (GSV+OMI) – performance that is comparable to traditional LUR approaches. Importantly, our models show that street-level features might have the potential to better capture intra-urban variation of NO_2 pollution than traditional LUR. Collectively, our findings indicate that street view image-based modeling has great potential for building large scale air quality models under a unified framework. Towards that goal, we describe a cost-effective image sampling strategy for future studies based on a systematic evaluation of image availability and model performance.

Keywords: empirical models, exposure assessment, air quality, image sampling and processing, computer vision, machine learning

3.1 Introduction

Nitrogen dioxide (NO_2) is an important traffic related air pollutant (TRAP) and a criteria pollutant. NO_2 has been associated with a wide range of adverse health effects, including premature mortality, asthma, lung cancer, and cardiovascular disease.[1-6] As a short-lived pollutant, NO_2 concentrations decay rapidly near emission sources, with published estimates of typical distance-decay gradients ranging from <500 m to <2000 m depending on the atmospheric dispersion conditions.[7-10] Capturing intra-urban variations of NO_2 concentrations is essential for accurately assessing human exposure to NO_2 and the corresponding health outcomes.[11,12] Among the many efforts to improve the spatial resolution of air quality models, land use regression (LUR) has been frequently applied as a cost-effective way to estimate fine-scale ambient air pollutant concentrations.[13-16] According to a review, LUR was the most popular method for NO_x studies in the last decade.[17]

National LUR models have been developed for, e.g., the United States[18-21], Australia[22,23], in Europe[24,25], and China[26,27]. Traditional LUR models rely on predictors collected from curated GIS-derived databases.[13,28] For example, Kim et al. collected geographic variables from 11 categories to build national air quality models, including traffic, population, land use, elevation, emissions, and satellite estimates.[18] Larkin et al. developed a global NO_2 LUR model using satellite estimates and land use variables, including normalized difference vegetation index, impervious surface, road length, and emission sources.[29] However, the availability of curated GIS-derived databases may vary across jurisdictions[30] and preprocessing geographic predictors originating from various data sources can be a labor intensive process. Additionally, there is evidence that traditional GIS variables may not capture street-level features that may be important determinants of air quality.[11,31]

In recent years, innovative image data sources (e.g., street view imagery, high resolution satellite imagery) have emerged as possible tools to capture hyperlocal characteristics of the natural and built environment. Image-based data may be promising for replacing or augmenting traditional LUR predictors[32,33] when enabled by imagery processing techniques (i.e., computer vision) and advanced modeling (e.g., machine learning).[34] For example, information (e.g., traffic, land use, built environment features) provided by traditional LUR predictors are also encoded in high-resolution digital images,[35] and can be extracted and quantified via advanced computer vision techniques. Compared to traditional curated geodatabases, image data sources allow consistent and uniform data collection and processing across jurisdictions.

As an emerging research topic, a small number of studies have explored using these new data sources to build empirical air quality models.[32,33,36-38] For example, Lu et al. compared national empirical models of six criteria pollutants using traditional LUR variables versus microscale variables (e.g., street view imagery, point of interest, local climate zones).[32] Similar model performance was achieved using the various combinations of variables, suggesting that microscale variables could be a suitable substitute for traditional predictors. Another study trained convolutional neural networks (CNN) on Google Maps satellite and street-level images, together with traditional LUR variables, to predict ultrafine particle (UFP) and black carbon (BC) concentrations.[33] Those results suggest that images may capture similar features as traditional GIS predictors while also allowing for inclusion of higher resolution (i.e., street-level) features that traditional GIS variables lack.[37] Another notable trend in recent air quality modeling studies is the adoption of advanced machine learning approaches to improve model performance, including random forest[39-41], gradient boosting[42,43], artificial neural network[44-46], and hybrid algorithms[47,48].

Our previous study successfully developed single-city LUR models solely using street view images.[36] In this study, we scale this approach to develop national NO_2 models for 2007-2019 using street view images and machine learning algorithms. Using the EPA fixed-site network as a basis (dependent variable), we employed the same image processing protocol developed in our previous study[36] to extract street-level features for modeling. We developed two types of random forest models, using: 1) Google Street View (GSV) imagery, and 2) GSV features plus satellite observations of NO_2 tropospheric vertical columns. To our knowledge, this is the first national NO_2 model which is based on street view imagery. We systematically assess the impact of GSV image availability on model performance, which may provide important evidence for optimal image sampling strategies for future image-based air quality modeling studies.

3.2 Materials and methods

3.2.1 Modeling variables

Figure 3-1 shows the sampling and processing procedure for preparing variables for modeling. We collected annual-average NO_2 concentrations at EPA monitors across the contiguous U.S. during 2007-2019 from the EPA AQS dataset. For predictor variables, we used street-level features extracted from GSV images and satellite observations of NO_2 tropospheric column density. In this paper, we use the term "GSV image" to refer to GSV panoramic images with unique image IDs. Each unique "GSV image" has 4 flat images corresponding to the 4 cardinal directions.

3.2.1.1 GSV image derived variables

We downloaded 336,544 GSV flat images within 500 m of EPA NO_2 monitors. For image collection and processing, we followed a similar procedure as detailed in our previous study.[36] Specifically, we created 100 m × 100 m grids within 500 m of each

EPA monitor and retrieved the metadata for all GSV images within the grid cells. To improve temporal consistency between the NO_2 observations and GSV features, we constrained the selection of GSV images such that only images taken during the same year or within the previous 4 years of the NO_2 observations were downloaded and processed. Further details about the choice of image inclusion by year are explained in the modeling approach section. After image retrieval, we used a deep learning model called Pyramid Scene Parsing Network (PSPNet),[49] to extract GSV features through image semantic segmentation (i.e., each pixel of the image is assigned to a feature category). A total of 150 GSV features were extracted through PSPNet processing. To improve the model interpretation, we removed features that are unrelated to air pollution (47 variables (**Table 3-S1**) remained for model development). We categorized GSV features into 7 subgroups: built environment (n=8), transport network (n=8), transport vehicles (n=9), natural (n=8), vegetation (n=6), water (n=6), and human (n=2). To tabulate GSV features for modeling, we stratified the GSV images into different buffer radii. Within each buffer radius, we calculated the mean percentage of GSV features to represent the surrounding natural and built environment characteristics. Unless specified otherwise, we used 250 m and 500 m buffer radii as the default for modeling.

3.2.1.2 Satellite variable

We also retrieved satellite observations of NO_2 from the Ozone Monitoring Instrument (OMI). OMI was launched in 2004 on the Aura platform and has a nadir footprint of approximately 13 km × 24 km.[50] Many studies have demonstrated the predictive power of satellite-based variables in NO_2 models.[19,21,43,51] Some studies found that directly adopting the satellite column density into the LUR model is sufficient for modeling, suggesting the extra step for estimating column-to-surface ratio may be

unnecessary.[20,22] In our study, we directly used the satellite observations (column density) at the EPA sites as an additional model predictor. We obtained OMI NO_2 tropospheric column density from the OMI L3 product (OMI/Aura NO_2 Tropospheric, Stratospheric & Total Columns MINDS Daily L3 Global Gridded 0.25° × 0.25°)[52] provided by the NASA Goddard Earth Sciences Data and Information Services Center (GESDISC). For data quality control, we used the cloud screened tropospheric vertical column density with effective cloud fraction (ECF) < 0.3, solar zenith angle (SZA) < 85 degree, and the primary summary quality flag indicating good data.

3.2.2 Modeling approach

Since GSV images are not updated at a uniform rate across regions, the density of available images in the vicinity of NO_2 monitors varies depending on the site location and year. **Figure 3-1-C** demonstrates the spatiotemporal variation of image availability using one EPA site for illustration; here (as with many locations), image availability varies by year and in extreme cases there may be no images for certain years. To systematically analyze the impact of GSV image availability on model performance, we designed 5 scenarios for temporal matching of images as model inputs (see **Table 3-S2**). The strictest scenario (i.e., scenario 1) uses only GSV images from the same year of the NO_2 observations. Since the built environment changes slowly in developed countries such as the U.S., we relaxed the temporal match criteria in other scenarios. The relaxed scenarios (i.e., scenario 2-5) further include GSV images from the previous 1-4 years before the NO_2 observations. As a result, the dataset for a stricter scenario (i.e., fewer years) is always a subset of the more relaxed scenario. Data summaries for each scenario are shown in **Table 3-S2**. Once all model variables were preprocessed, we developed two types of random forest models: 1) GSV-only models which solely relied on GSV features, and 2) GSV+OMI models which used both GSV features and

the OMI variable. Both models included the year of NO_2 observation as a dummy variable in the model.

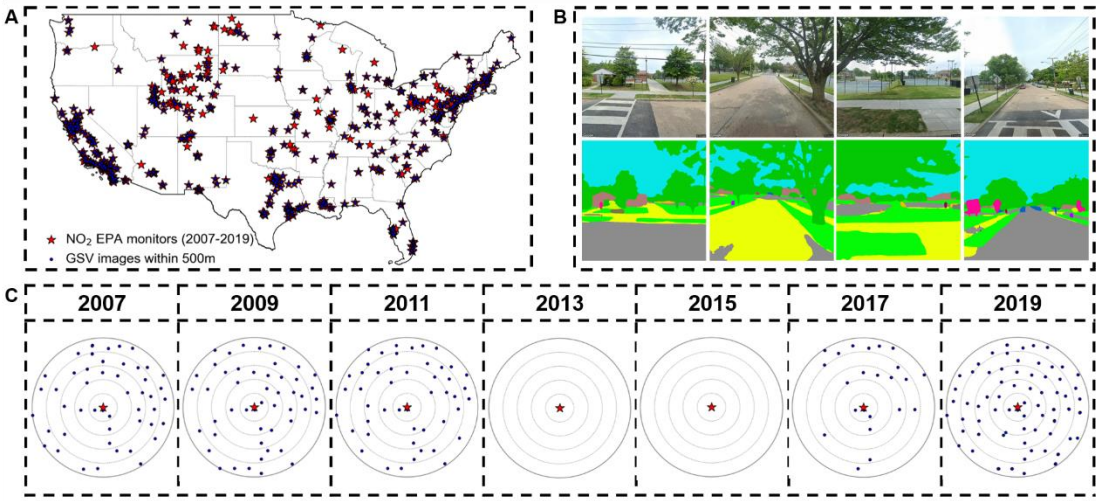

Figure 3-1. NO_2 measurement and GSV locations for the national NO_2 models. (A) Spatial distribution of NO_2 EPA monitors and GSV images within 500 m of monitors. (B) Example of GSV features extracted via image semantic segmentation. (C) Illustration of spatiotemporal variation in GSV image availability for one EPA monitor.

3.2.3 Model evaluation

To compare model performance among different scenarios, we report cross validation (CV) R^2, mean absolute error (MAE), and root mean squared error (RMSE). Our default cross validation approach is a random 10-fold cross validation for which all NO_2 records were randomly divided into 10 subsets. Each subset was sequentially held out as the test set, a random forest model was trained with the remaining 9 subsets, and model performance was evaluated on the unseen test set. We iterated this process for each subset and calculated the overall R^2 after all subsets were predicted. In addition to the conventional random 10-fold cross validation, we also report two additional cross validation strategies: spatial hold out and temporal hold out. More details are in the SI.

3.2.4 Sensitivity analysis for image sampling strategies

In addition to exploring the influence of temporal matching criteria for image-monitor pairs, we also examined the impact of image availability on model performance by using different image sampling strategies. Our base models downloaded all available images; in this sensitivity analysis, we limited the number of GSV images retrieved for each 100 m × 100 m grid within the 500 m buffer of NO_2 monitors. We chose thresholds of 1, 2, or 3 GSV panoramic images (i.e., 4, 8, or 12 GSV flat images) for each grid cell. In this paper, we refer to these models as image-limited models. We refer to models with no image inclusion threshold, i.e., including all GSV images available within the grid cell for modeling, as image-unlimited models. For image-limited models, when multiple images from different years were available, we chose images with the best NO_2-GSV temporal match. We also examined image sampling strategies among different sampling grid sizes. We compared 150 m × 150 m, 200 m × 200 m, and 250 m × 250 m grids. Direct comparison between image-unlimited and image-limited models was feasible within the same temporal matching scenario because the model training size remained the same among different sampling strategies.

3.3 Results and discussion

3.3.1 Summary of model inputs

We collected 4,867 unique NO_2 annual-average records for the contiguous U.S. during the 13-year study period. NO_2 concentrations declined over time (**Figure 3-S1**); the number of NO_2 EPA monitors slightly increased from 356 stations in 2007 to 407 in 2019 (average 374 stations/year during the study period).

GSV image availability varies across region and year (see **Figure 3-S2**). Hence, not every NO_2 site in every year has sufficient GSV images for model training. By including images from neighboring years, we were able to include a larger training dataset while striving to capture similar built environment characteristics as the target

year. In scenario 1, only 42% of NO_2 annual records had one or more images for the target year within 250 m of the monitor. Google started collecting street view images in 2007; image availability at EPA monitors varied by year. We found that 51% of NO_2 monitors had GSV images within 250 m in 2007, however, very few EPA monitors had images in 2008-2010 (e.g., only 5 locations had images in 2010). Relatively more images were collected in 2011 but the number of NO_2-GSV pairs dropped again in 2012-2014. In more recent years, the number of available GSV images generally increased. By relaxing the requirement for an exact temporal match (e.g., in scenario 5) far more images are available for modeling, leading to a more balanced training set across years. Overall, 72% of NO_2 annual records were successfully paired with GSV features in Scenario 5. Among these GSV-NO_2 pairs, the median (P10/P25/P75/P90) of GSV images within 500 m of one EPA monitor was 53 (12/26/104/164).

Figure 3-S3 shows the spatial distribution of NO_2-GSV pairs among years for scenario 5. GSV image availability near NO_2 monitors increased from early years to present. Variation in GSV image availability among monitor sites also increased, i.e., for a given year, some monitors include hundreds of images while others only have a few images. We also examined the spatial distribution of NO_2 monitors which were excluded from modeling due to insufficient images and found that monitors located around the Rocky Mountains have the fewest GSV images. In general, the EPA locates NO_2 monitors for a variety of purposes (e.g., near-road monitoring, area-wide monitoring, or to protect susceptible and vulnerable populations). **Table 3-S3** shows the distribution of NO_2 monitors by urbanicity (i.e., urban/suburban/rural) using 2019 for illustration: of the 407 monitors, 45% were located in urban areas, versus 28% (suburban) and 27% (rural). For context, the US population is 32% urban, 47% suburban, and 22% rural. Because of how GSV images are collected, more images were

available in urban than in other areas: the proportion of monitors having GSV images within 500 m in scenario 5 was 96% for urban areas, 84% for suburban areas, and 38% for rural areas. Detailed descriptive statistics of GSV features for NO_2 monitors by urbannicity are listed in **Table 3-S4**.

3.3.2 Model results of image-unlimited models

Using the image-unlimited models (i.e., including all images), we compared model performance among the different NO_2-GSV temporal match scenarios and investigated the ranking of feature importance among models.

3.3.2.1 Model performance

Figure 3-2-A shows the results of the random 10-fold cross validation for both GSV-only and GSV+OMI models. **Figure 3-S4** shows the corresponding results for MAE and RMSE. Full model performance is shown in **Table 3-1**. Both GSV-only and GSV+OMI models showed increasing performance from scenario 1 to scenario 5. With the addition of satellite observations, the GSV+OMI models outperformed GSV-only models in all scenarios. In the strictest scenario, the GSV-only model showed moderate performance (random 10-fold CV R^2: 0.50; MAE: 2.95 ppb; RMSE: 3.90 ppb) while the corresponding GSV+OMI model showed improved results (random 10-fold CV R^2: 0.69; MAE: 2.26 ppb; RMSE: 3.05 ppb). However, the advantage of adding satellite-based data gradually narrowed when more images were included in the model. From scenario 1 to scenario 2, the GSV-only model showed a dramatic jump in model fit with CV R^2 increasing 0.26. As more images were included in the model, the model performance increased but plateaued by scenario 5. Correspondingly, the difference in the random 10-fold CV R^2 between GSV+OMI models and GSV-only models decreased from 0.18 in scenario 1 to 0.02 in scenario 5. A similar decreasing trend was observed for the feature importance of the OMI variable in the GSV+OMI models. Both

the GSV-only model (10-fold CV R^2: 0.88; MAE: 1.25 ppb; RMSE: 1.88 ppb) and GSV+OMI model (10-fold CV R^2: 0.91; MAE: 1.13 ppb; RMSE: 1.69 ppb) achieved the best performance in scenario 5, where the advantage of the GSV+OMI model over the GSV-only model became marginal. Compared to other national NO_2 modeling studies (R^2: 0.72~0.91),[18,21,32,53] the performance of our best models was consistent with or better than previous efforts. Our results indicate that with sufficient images, GSV imagery alone may explain the variation of long-term national NO_2 concentrations.

Relaxing the temporal matching criteria for GSV-NO_2 pairs increased both the number of GSV images included in the model and the available GSV-NO_2 pairs for model training. To separate the effects of the two factors, we further constrained the training size of each scenario by choosing the same NO_2 records for model development. **Figure 3-2-B** illustrates that when limited to the same training data (i.e., the same NO_2 observations), adding more GSV images to the model still increases model performance significantly. For example, 10-fold CV R^2 for the GSV-only model in scenario 5 increased 0.25 compared to scenario 1. Comparatively, the increase for models without constraints in training size was 0.37. As expected, the model performance was worse when we limited the training data size and a larger gap between GSV-only and GSV+OMI models can be observed. The best model fits decreased to 0.76 for the GSV-only model and 0.82 for the GSV+OMI model. Full results are shown in **Table 3-S5**.

We examined model performance by comparing model predictions versus observations. **Figure 3-2-C** and **D** illustrate the results for scenario 5 while the full results are shown in **Figure 3-S5**. Overall, the trendlines for model predictions versus observations are closer to 1:1 line when including more images (i.e., moving from scenario 1 to scenario 5). In scenario 5, both GSV-only and GSV+OMI models show

good agreement between the predicted and observed NO_2 concentrations. No distinctive spatial pattern was observed in terms of monitors with lowest or highest errors (**Figure 3-S6**).

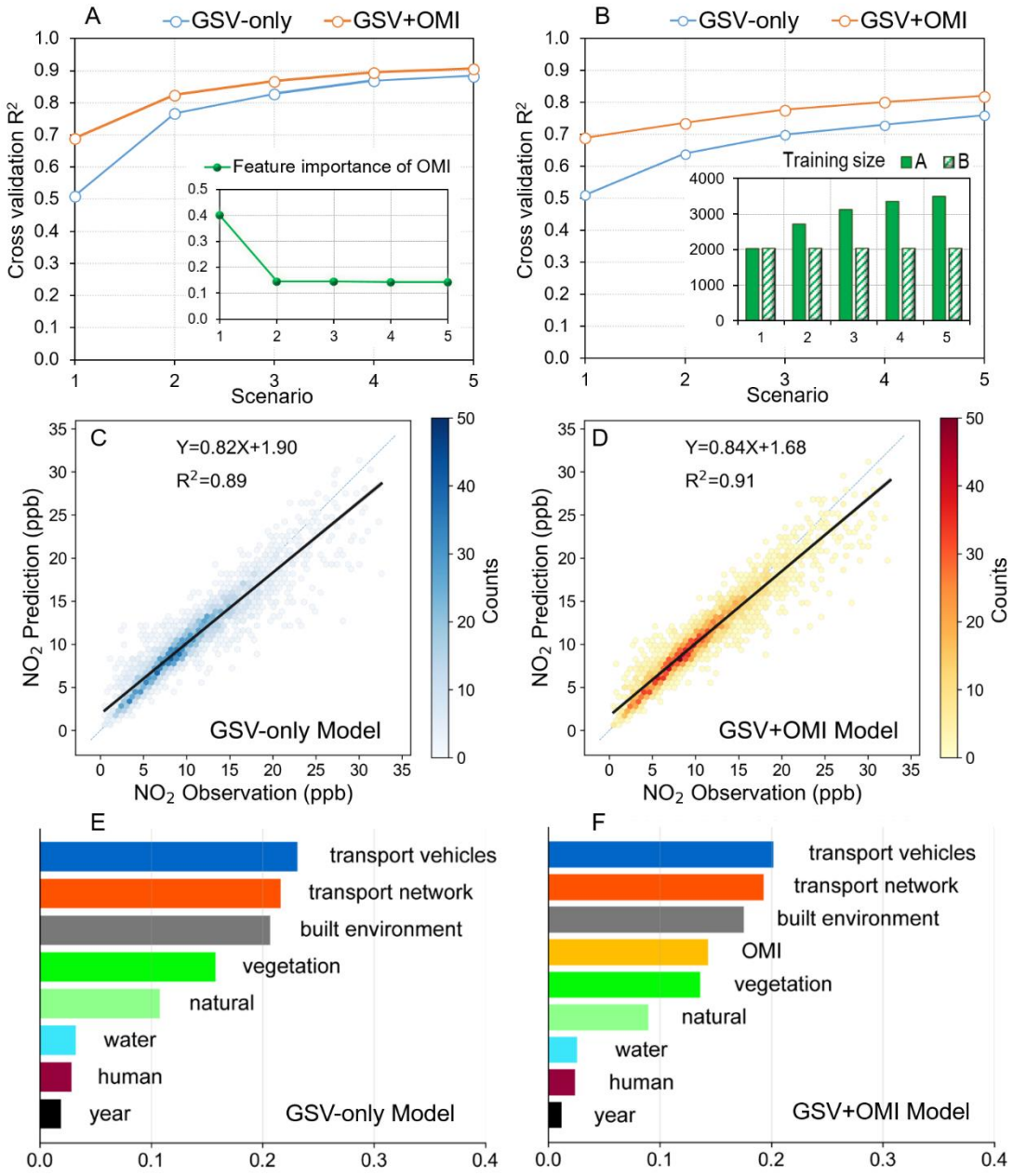

Figure 3-2. Comparison of random 10-fold cross validation R^2 between GSV-only and GSV+OMI models. (A) shows the results of image-unlimited models while (B) shows the results of image-unlimited models with constrained training size. The training sizes of models in (A) and (B) are shown in the bar chart within (B). (A) also shows the

feature importance score of the OMI variable in the GSV+OMI models among the 5 scenarios. Both (A) and (B) show results for 5 scenarios while subplots (C) to (F) only show results by scenario 5 from models in (A). (C) and (D) show hexagonal binned plots of model predictions versus observations. The color reflects the number of scatters within each hexagon. (E) and (F) show the rank of feature importance in the random forest models. Results of other scenarios are shown in **Figures 3-S5** and **3-S7**.

Table 3-1. Model performance of image-unlimited models.

Model type	Scenario	Random 10-fold CV R^2	Mean Absolute Error (MAE)	Root Mean Squared Error (RMSE)
GSV-only	1	0.51	2.89	3.83
	2	0.77	1.86	2.64
	3	0.83	1.55	2.27
	4	0.87	1.33	1.99
	5	0.88	1.25	1.88
GSV+OMI	1	0.69	2.26	3.05
	2	0.83	1.61	2.29
	3	0.87	1.37	2.00
	4	0.90	1.19	1.78
	5	0.91	1.13	1.69

3.3.2.2 Rank of feature importance

Feature importance scores (**Figure 3-2-E** and **F**) suggest that traffic-related features were the largest contributors to our NO_2 predictions. Generally, transport vehicle variables contributed most, followed by transport network, and built environment variables. This result is important as traditional GIS-derived variables are not able to resolve some of these differences (e.g., transport vehicles vs. transport network) that imagery was able to identify. This may be important as some parts of the urban environment are fixed (e.g., buildings, roads) while others are transient (e.g., vehicles); policy solutions may differ depending on the nature of these features.

The granularity of GSV features allowed further comparison among different components of the urban environment. **Figure 3-S8** shows disaggregated model predictors using the top 10 features for illustration. The OMI variable was always the strongest single predictor in our GSV+OMI models. In scenario 1, the feature importance score of the OMI variable was 0.40 while the second highest score was only 0.06 for car. However, when we added more images into the model, the contribution of satellite data decreased dramatically, and GSV features contributed more to the models overall. This finding may be partially explained by the different spatial resolutions of GSV features and the OMI variable. The satellite observations have coarser resolution (relative to GSV images) and thus contribute information on regional trends. Our preliminary analysis of an OMI-only model (i.e., using the OMI variable as the only predictor) showed that OMI alone could explain around half of the NO_2 variation (CV R^2: 0.49; MAE: 3.09 ppb; RMSE: 4.08 ppb). To further capture spatial trends of NO_2, predictors with much higher spatial resolution are needed.

Among GSV features, the most important features in both GSV-only and GSV+OMI models included transport vehicles (car, truck), transport network (sidewalk), built environment (wall, building), and vegetation (grass). These features have plausible explanations for inclusion in our models, e.g., traffic-related features reflect emission sources, built environment features may be associated with activity centers and pollution dispersion, and green spaces are typically low emission areas. We examined how GSV features influence our models through accumulated local effects plots using several important features and the GSV-only model for illustration. As shown in **Figure 3-S9**, car, truck, building, wall, and sidewalk positively contribute to NO_2 pollution while grass has a negative contribution.

3.3.2.3 Comparison of different cross validation strategies

Besides the conventional random 10-fold cross validation, we used two additional cross validation strategies, i.e., splitting the training and testing set spatially or temporally (Full details are described in the SI). Briefly, we found that temporal cross validation achieved similar performance as the random cross validation approach (**Figure 3-S10**). However, the spatial cross validation showed degraded performance, especially for the GSV-only models (CV R^2: 0.46; MAE: 3.05 ppb; RMSE: 4.04 ppb). The GSV+OMI model (CV R^2: 0.59; MAE: 2.62 ppb; RMSE: 3.52 ppb) showed slightly better performance, indicating improved model generalizability with the inclusion of regional information (i.e., OMI variable) in the model. One explanation may be that we were only able to use a small search area (i.e., 500 m) to retrieve street view images near NO_2 monitors. As a result, only hyperlocal information was included in the national model – a potential issue for model extrapolation in regions where there are few EPA monitors. We speculate that if a denser air pollution monitoring network were used (e.g., mobile monitoring) this issue could be mitigated.

3.3.3 Sensitivity analysis using image-limited models

We systematically controlled the image sampling strategy by varying the number of images sampled for each grid cell (as well as the sampling grid size) to investigate the impact on model performance. Full model results are in **Table 3-S6** and **3-S7**. As shown in **Figure 3-3-A** and **B**, image-limited models generally had decreased performance compared to image-unlimited models. However, the reduction in model fit was relatively small. For example, in scenario 5, when collecting one GSV image per 100 m × 100 m grid, the random 10-fold CV R^2 was 0.85 (0.88) for the GSV-only (GSV+OMI) model; comparatively, it was 0.88 (0.91) for the GSV-only (GSV+OMI) for the image-unlimited models. This result was due to the fact only a small portion of

grid cells had repetitive GSV images resulting in only 6.2% of GSV flat images being removed from the analysis.

Although only a small number of images were removed during model development, this finding is most important for the model application stage when the number of GSV images above a given threshold (e.g., 1, 2, or 3 per grid cell) could be huge. We used Washington DC as an example city to illustrate model prediction in 2019 at 100 m spatial resolution among different image sampling strategies (**Figure 3-3-E** and **F**). Interestingly, image-unlimited and image-limited models generated similar prediction surfaces for NO_2 while the former models used twice the number of images (158,276 vs 79,612 GSV flat images). This finding is important when making predictions at national or global scales. Our results suggest that a parsimonious image sampling strategy may be sufficient and most cost-effective for model development and application. A rationale for this approach is that images distributed geographically close are more likely to provide redundant information, resulting in little influence on the tabulated model predictors, i.e., the mean of GSV features within a certain buffer area. This issue could be further explored in future research, especially for differences in transient (e.g., vehicles, people) vs. fixed (e.g., roads, buildings) features of the urban environment.

We also explored the influence of various sampling grid sizes on model performance when sampling one GSV image per grid cell. Results for scenario 5 are in **Figure 3-3-C** and **D**. Increasing the size of grid cells led to a significant deterioration in model performance. When using a 150 m × 150 m (250 m × 250 m) grid and maximum of one GSV image per grid cell, the random 10-fold CV R^2 of the GSV-only model dropped to 0.74 (0.70). This result reflects the importance of spatial resolution of GSV images used in the models. For example, the 100 m × 100 m grid generated ~81 grid cells

within a 500 m buffer while the 150 m × 150 m (250 m × 250 m) grid generated ~29 (~7) grid cells. As expected, the GSV+OMI models were more robust to differences in grid size relative to the GSV-only models. Overall, our findings suggest that retrieving GSV images at 100 m resolution may be the most cost-effective way to establish national NO_2 models with high performance. Sparse resolution of image sampling may miss important hyperlocal features and lead to underfitting models. Overly dense sampling within a single grid cell may lead to an unnecessarily high workload – a cost-prohibitive (e.g., image downloading and processing) and time-consuming task for limited improvement in model fit.

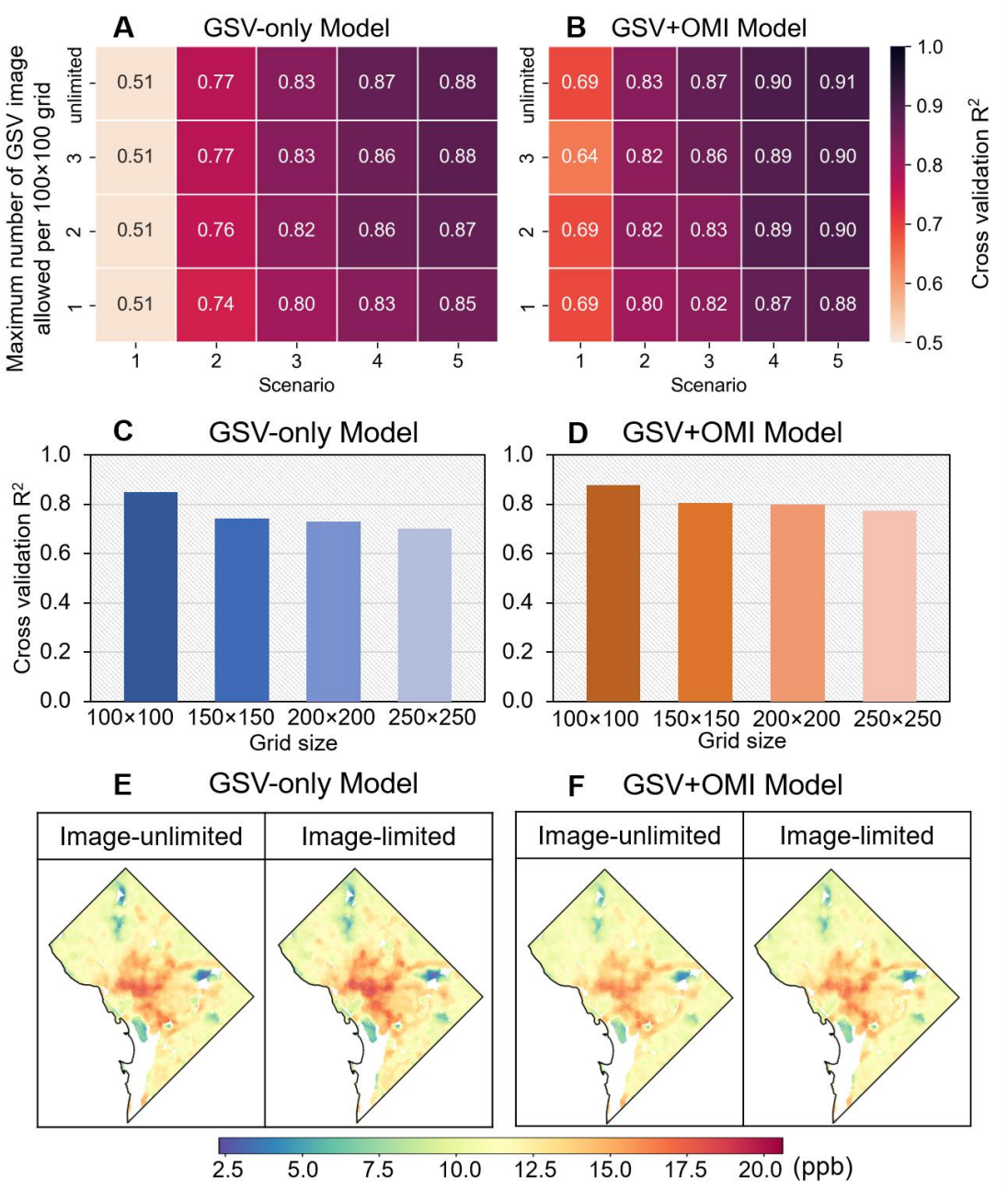

Figure 3-3. Results of sensitivity analyses among image sampling strategies. (A) and (B) show model results among all scenarios with varying maximum number of GSV images per 100 m × 100 m sampling grid. (C) and (D) show model results in scenario 5 with varying grid size. Each grid cell is allowed one GSV image for sampling. (E) and (F) show model predictions for Washington DC in 2019 using image-unlimited models and image-limited models with maximum one GSV image per 100 m × 100 m grid.

3.3.4 Model prediction and comparison

Following the same image collection and processing protocol as in the model development stage, we collected another 338,084 GSV flat images to predict NO_2 air pollution for Washington DC from 2007 to 2019 (**Figure 3-S11**). Both the GSV-only and GSV+OMI models captured fine scale gradients of NO_2 concentrations in Washington DC. Consistent with the decreasing national trend in EPA monitored NO_2 concentrations, our prediction maps also show reduced NO_2 concentrations over time. The most polluted area is located in the central city and the concentrations gradually decrease elsewhere. Major gaps in the surface are over rivers or large green spaces since most GSV images are collected by GSV cars driving along roads. Generally, GSV imagery has good coverage in this region and our GSV image-based models were able to predict 88.1%±1.6% of the city.

We compared our GSV models to another national NO_2 model developed using a conventional LUR approach (**Figure 3-4)**. We found that the GSV-only model predicted the highest NO_2 concentrations in the central city in Washington DC (the GSV+OMI model predicted slightly lower concentrations). The image-based models predicted significantly higher concentrations than the conventional LUR for the same region (**Figure 3-4-B**). Conversely, in the less polluted areas (e.g., city parks and rivers), the GSV-only model predicted the lowest NO_2 concentrations followed by GSV+OMI model; the conventional model predicted the highest concentrations. Previous studies show the tendency of LURs to underestimate in polluted regions and overestimate in clean regions,[29,54,55] which could be partly the result of uncaptured predictors in the model (e.g., street-level features). Our results suggest that leveraging street view images for air quality modeling may have the potential to mitigate this tendency and better capture intra-urban NO_2 variability than other LUR methods. We

found similar differences in another example city – San Francisco (**Figure 3-S12**). Future work could compare the results with mobile monitoring campaigns to investigate whether GSV image-based models capture more spatial variability of air pollution. Identifying the impact of different features on air pollution may provide important evidence for making local and regional air quality policies. We also plotted the distribution of GSV feature groups along a transect line (**Figure 3-4-C** and **Figure 3-S12-C)**. NO_2 predictions appear to show intuitive associations with the built environment - a surrogate for human activities. **Figure 3-S13** shows the distribution of GSV feature groups over the DC area in 2019 for illustration. The spatial patterns of different feature groups appear to be reasonably consistent with the land use characteristics of DC. For example, the transport vehicles and built environment features are highlighted in the central downtown area while transport network are more dispersed.

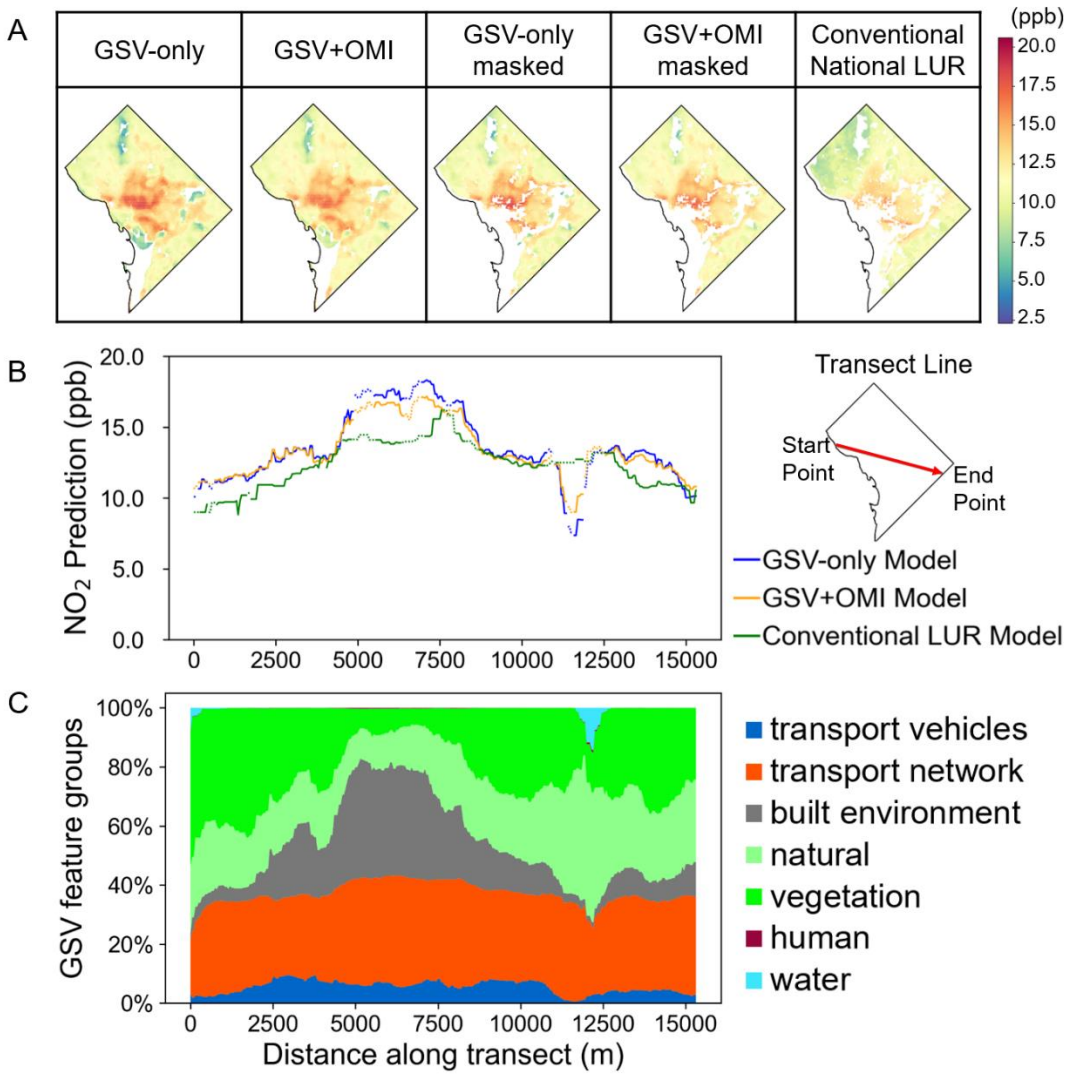

Figure 3-4. Comparison between GSV-based models and conventional national LUR models using prediction maps for Washington DC in 2015. (A) Model predictions for the GSV-only model, GSV+OMI model, and conventional national LUR model. For comparison, we masked GSV-based prediction maps based on where the conventional LUR predictions were available. (B) Predicted NO_2 concentrations among different models along the transect line. The solid lines represent actual model predictions; gaps in the conventional national LUR model (masked GSV-based models) were filled by interpolation (original GSV-based models) and represented by dotted lines. (C) Values of GSV feature groups along the transect line.

3.4 Limitations and implications for future studies

Our results demonstrate promising performance for GSV image-based air quality modeling, i.e., the potential to capture fine-scale NO_2 using a uniform modeling pipeline which can be easily scaled. The major limitation of our approach is the imbalance in image availability over time and space. This limitation is most prominent in early years and becomes less severe in more recent years as more images are available. Most GSV images are collected by Google GSV cars (with a smaller number shared by Google users). Depending on priorities set by the entity collecting images (in this case Google), the street-level imagery is updated at different frequencies over different regions and time periods. We are not able to fully utilize all monitoring sites from the EPA monitoring network for model development since some EPA sites (especially in the rural midwest and rocky mountain areas), do not have GSV images. Similarly, model application is affected by image availability and gaps in prediction maps may exist. However, since a priority is to collect images in areas where people live, the imagery does include the vast majority of the US population. Another limitation is our choice to relax the temporal matching criteria between air pollutant observations and GSV images. While we believe this approach is reasonable given slow changes in built environment in developed countries (and, the results indicate that the added images increase model performance), it is difficult to quantify the model uncertainty resulting from this choice. The lack of temporal resolution in the GSV imagery makes it challenging to use our approach to develop monthly or daily average models.

The satellite-based NO_2 variable has limitations similar to other air quality models that include satellite-based observations. For example, OMI has coarse spatial resolution, which is not capable of capturing local hotspots or intra-urban variation of

NO_2 concentrations. In addition, satellite-derived products reflect but do not offer exact attributes of the ground-level concentrations; retrieval of satellite-derived data involves many assumptions which may introduce uncertainties in the model.[56] Data availability and quality can be affected by multiple factors, e.g., errors in NO_2 column retrievals are large during cloudy days or over snow/ice surfaces.[52] All of these uncertainties may affect model performance.

Our models could be enhanced in several ways. We were only able to apply our models for two cities for illustration. As retrieval of large volumes of images becomes more common, our future work may map NO_2 concentrations at the national scale. Our models show reductions in model performance when using spatial cross validation. Although other studies also report decreases in performance when using spatial cross validation, the reductions were smaller (0.09-0.20).[21] Future work may mitigate this issue by extracting GSV features from a larger spatial buffer around each monitoring station, including interpolation components (e.g., kriging) in the modeling framework, or applying more advanced deep learning models to directly learn NO_2 pollution related features from the street view images[57].

Another promising direction is to utilize air pollution data from denser air quality monitoring networks that better match the granularity of the information in the street view images. For example, utilizing information from mobile monitoring studies that identify localized air quality patterns may unlock additional information in the street-level imagery that traditional, GIS-based predictors cannot. As mobile monitoring, low-cost sensing, and the frequency of image collection grows in future years, we expect this will be a significant area of research in the near future.

Our results suggest the promise of leveraging street view imagery and satellite data in building large scale empirical air quality models (e.g., national or global) under a

unified framework. Unlike traditional LUR predictors collected from multiple curated geodatabases, our approach requires few (e.g., single source for GSV-only models) data sources that achieves comparably high model performance to previous models. Adding satellite observations is a simple way to further boost model performance. Both data sources are publicly accessible and have a uniform data format, which make them excellent candidates in establishing a consistent data collection and processing protocol for model development over large geographies. This may be especially useful for less developed areas where municipal geodatabases are sparse or absent. Our study also presents a cost-efficient image sampling strategy (i.e., one image every 100 m) developed through a systematic sensitivity analysis – this finding may be helpful for future image-based modeling work. Our modeling approach is not restricted to GSV imagery. Many other large open-access street view image databases have emerged globally, e.g., Bing Maps Streetside, Apple Look Around, Tencent Maps (China), Daum Maps (South Korea), CycloMedia (Europe and U.S.), etc.[58] Our modeling framework can be applied to any street-level image. We expect that street view image-based air quality modeling may play a more important role in the near future given the fast growth of image resources.

3.5 Supporting Information

3.5.1 Description of spatial and temporal cross validation

In addition to the conventional random 10-fold cross validation, we also compared model fits with two additional cross validation strategies: spatial and temporal hold out. For the spatial 10-fold cross validation, we created 500 m × 500 m grids across the contiguous U.S. Each NO_2 record was mapped onto one grid and a unique geohash code was assigned based on the grid latitude and longitude. We then randomly divided the

NO_2 records into 10 groups based on the geohash code and conducted 10-fold cross validation. This method assured that data which was geographically close will not appear both in the training and testing set and may better reflect model extrapolation accuracy at unmonitored areas far from EPA sites. For the temporal cross validation, we treated each year as one hold-out subset and trained models with data from the other 12 years. This method estimates the model's interpolation ability for predicting future or past air quality given data from adjacent years.

3.5.2 Results of spatial and temporal cross validation

Besides the conventional random 10-fold cross validation, we also analyzed two additional cross validation strategies, i.e., splitting the training and testing set spatially or temporally. We chose to use scenario 5 for comparison based on the results discussed above. As shown in **Figure 3-S10**, the temporal cross validation achieved similar performance as the random cross validation approach. However, the spatial cross validation showed worse performance, especially for the GSV-only models (CV R^2: 0.46; MAE: 3.05 ppb; RMSE: 4.04 ppb). The GSV+OMI model (CV R^2: 0.59; MAE: 2.62 ppb; RMSE: 3.52 ppb) showed slightly better performance, indicating improved model generalizability with the inclusion of regional information (i.e., OMI variable) in the model. The vast decrease in GSV-only models might be explained by the small search area (i.e., 500m) for retrieving street view images around NO_2 monitors. As a result, only hyperlocal information was included in the national model, possibly exacerbating the difficulty in model extrapolation, especially in regions where EPA monitoring is lacking over a large area. Including additional images at larger buffers is out of the scope of this study due to computational constraints but would be a useful direction for future research.

Our hypothesis is that if our approach is applied to denser monitoring networks (e.g., large mobile monitoring campaigns) that this issue would be mitigated. This can be partially demonstrated by the comparison of models with different buffer radii as the choice of buffer size affects model training size. Additionally, we developed models using a different set of buffer radii to tabulate GSV features, i.e., 100/200/300/400/500 m. Such buffer radii offered finer features as the model input, but also led to 22.7% decrease in the number of GSV-NO_2 pairs for model development. While the results of temporal cross validation remained consistent, the spatial cross validation R^2 of the GSV-only model by 250/500 m buffers was significantly higher than that of using 100/200/300/400/500 m buffers, mainly because more NO_2-GSV pairs were available for the 250/500 m models. Hence, we speculate that if combined with a much denser air pollution monitoring network, the street view image-based air quality models may be enhanced remarkably.

Table 3-S1. Predictor variables for model development.

Predictor variables	Sub-group	Description
GSV features	built environment	building, canopy, house, hovel, pier, skyscraper, swimming pool*, wall,
	transport network	bridge, dirt track, path, road, runway, sidewalk, streetlight, traffic light
	transport vehicles	airplane, boat, bus, car, minibike, ship, signboard*, truck, van
	nature	animal, earth, hill, land, mountain, rock, sand, sky
	vegetation	field, flower, grass, palm, plant, tree
	water	fountain, lake, river, sea, water, waterfall
	human	person, bicycle
OMI	-	cloud screened tropospheric vertical column density
Year	-	year of NO_2 observation

GSV features marked with * represent that they may fall into multiple categories across different GSV images. For instance, signboard can be advertisement on trucks or walls. In our study, images with top signboard element mainly fitted into the first situation, hence we categorized it under transport vehicles. Swimming pool was categorized as built environment.

Table 3-S2. Scenarios for GSV-NO_2 temporal match during model development.

Scenario	GSV-NO_2 observation temporal match	No. of GSV images	No. of NO_2-GSV records
1	The same year as NO_2 observations	278,108	2,028
2	Same year and previous year of NO_2 observations	300,192	2,715
3	Same year and previous 2 years of NO_2 observations	312,272	3,129
4	Same year and previous 3 years of NO_2 observations	325,784	3,361
5	Same year and previous 4 years of NO_2 observations	336,544	3,503

Table 3-S3. Distribution of NO_2 EPA monitors by urban, suburban, rural using 2019 for illustration. The urbanicity classification is based on the locale classification provided by National Center for Education Statistics.

Land use type	No. of NO_2 monitors	No. of NO_2 monitors with GSV images within 500 m using scenario 5
Urban	182	174
Suburban	114	96
Rural	111	42

Table 3-S4. Descriptive statistics of GSV features within 500 m of NO_2 monitors stratified by land use types. GSV features are unitless since they are extracted through image segmentation and represent pixel percentage by GSV variables.

Land Use Type	Feature	Min.	25th percentile	Median (50th percentile)	Mean	75th percentile	Max.	Standard Deviation
Urban	transport vehicles	6.03E-04	1.54E-02	2.14E-02	2.60E-02	3.32E-02	9.67E-02	1.64E-02
	transport network	2.32E-01	3.19E-01	3.41E-01	3.38E-01	3.63E-01	4.07E-01	3.26E-02
	built environment	1.66E-03	3.74E-02	5.57E-02	6.91E-02	9.18E-02	2.77E-01	4.72E-02
	natural	2.07E-01	3.23E-01	3.68E-01	3.71E-01	4.21E-01	5.62E-01	7.04E-02
	vegetation	4.09E-02	1.14E-01	1.75E-01	1.81E-01	2.41E-01	4.08E-01	8.33E-02
	water	0.00E+00	5.02E-05	1.52E-04	4.87E-04	3.87E-04	1.41E-02	1.29E-03
	human	6.10E-07	5.92E-05	1.21E-04	1.79E-04	2.18E-04	1.57E-03	2.05E-04
Suburban	transport vehicles	1.18E-04	9.56E-03	1.66E-02	1.89E-02	2.39E-02	7.01E-02	1.37E-02
	transport network	1.72E-01	3.16E-01	3.41E-01	3.33E-01	3.65E-01	4.04E-01	4.42E-02
	built environment	1.02E-03	2.55E-02	4.38E-02	4.99E-02	6.24E-02	2.28E-01	3.74E-02
	natural	2.51E-01	3.52E-01	3.92E-01	3.99E-01	4.41E-01	6.85E-01	7.53E-02
	vegetation	2.18E-02	1.20E-01	1.70E-01	1.87E-01	2.37E-01	5.28E-01	9.22E-02
	water	0.00E+00	9.84E-06	8.58E-05	4.91E-04	3.31E-04	1.04E-02	1.32E-03
	human	1.77E-06	5.18E-05	8.92E-05	1.04E-04	1.42E-04	3.21E-04	7.67E-05
Rural	transport vehicles	7.68E-04	1.14E-02	1.73E-02	1.70E-02	2.29E-02	3.57E-02	9.98E-03
	transport network	2.79E-01	3.52E-01	3.59E-01	3.64E-01	3.99E-01	4.14E-01	4.22E-02
	built environment	2.52E-02	3.78E-02	4.52E-02	5.95E-02	5.95E-02	1.72E-01	4.48E-02
	natural	3.32E-01	3.85E-01	4.46E-01	4.25E-01	4.63E-01	4.69E-01	5.01E-02
	vegetation	5.05E-02	7.50E-02	1.01E-01	1.26E-01	1.54E-01	2.36E-01	6.36E-02
	water	0.00E+00	0.00E+00	5.89E-06	4.31E-05	2.48E-05	2.67E-04	8.72E-05
	human	3.26E-07	5.19E-05	9.63E-05	1.37E-04	1.34E-04	5.38E-04	1.64E-04

Table 3-S5. Model performance of image-unlimited models using the same training size in 5 scenarios.

Model type	Scenario	Random 10-fold CV R^2	Mean Absolute Error (MAE)	Root Mean Squared Error (RMSE)
GSV-only	1	0.51	2.89	3.83
	2	0.64	2.43	3.29
	3	0.70	2.17	3.00
	4	0.73	2.02	2.85
	5	0.76	1.88	2.69
GSV + OMI	1	0.69	2.26	3.05
	2	0.74	2.07	2.82
	3	0.78	1.87	2.59
	4	0.80	1.74	2.45
	5	0.82	1.64	2.32

Table 3-S6. Model performance of image-limited models which used various maximum number of GSV images per 100 m × 100 m grid cell.

Maximum number of GSV images per grid cell	Model type	Scenario	10-fold CV R^2	Mean Absolute Error (MAE)	Root Mean Squared Error (RMSE)
1	GSV-only	1	0.51	2.90	3.85
		2	0.74	1.96	2.79
		3	0.80	1.68	2.47
		4	0.83	1.50	2.23
		5	0.85	1.44	2.15
	GSV + OMI	1	0.69	2.25	3.06
		2	0.80	1.76	2.46
		3	0.82	1.66	2.34
		4	0.87	1.33	1.98
		5	0.88	1.28	1.91
2	GSV-only	1	0.51	2.88	3.83
		2	0.76	1.88	2.67
		3	0.82	1.58	2.31
		4	0.86	1.38	2.07
		5	0.87	1.32	1.97
	GSV + OMI	1	0.69	2.26	3.06
		2	0.82	1.63	2.33
		3	0.83	1.59	2.25
		4	0.89	1.23	1.85
		5	0.90	1.19	1.77
3	GSV-only	1	0.51	2.89	3.84
		2	0.77	1.86	2.65
		3	0.83	1.55	2.29
		4	0.86	1.35	2.02
		5	0.88	1.29	1.93
	GSV + OMI	1	0.64	2.45	3.29
		2	0.82	1.62	2.32
		3	0.86	1.38	2.02
		4	0.89	1.21	1.81
		5	0.90	1.16	1.73

Table 3-S7. Model performance of image-limited models using different grid sizes for sampling a maximum of one GSV image per grid cell. All models use scenario 5.

Model type	Grid size	10-fold CV R^2	Mean Absolute Error (MAE)	Root Mean Squared Error (RMSE)
GSV-only	100	0.85	1.44	2.15
	150	0.74	1.93	2.79
	200	0.73	1.96	2.85
	250	0.70	2.07	2.98
GSV + OMI	100	0.88	1.28	1.91
	150	0.81	1.66	2.42
	200	0.80	1.68	2.46
	250	0.77	1.77	2.59

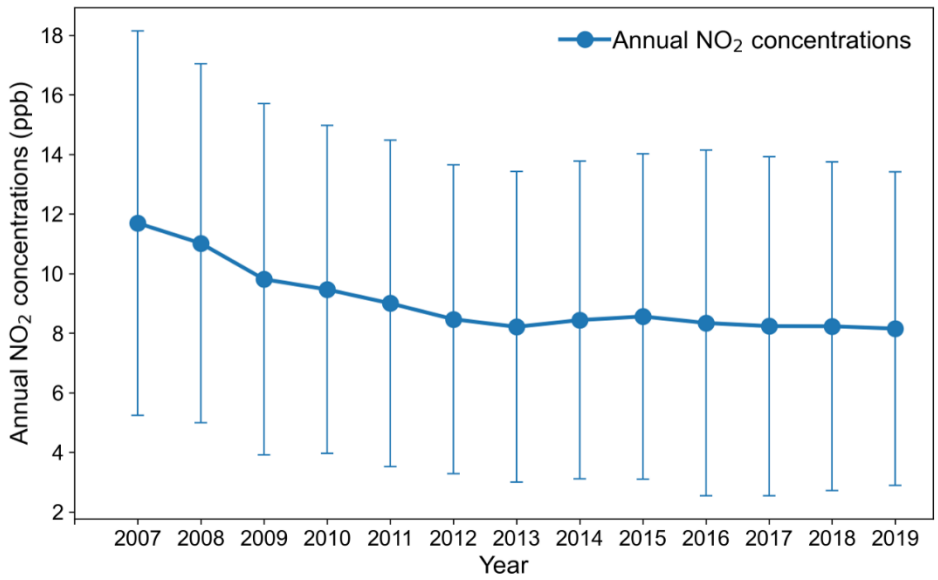

Figure 3-S1. Decreasing trend of annual NO_2 concentrations during 2007 to 2019 from the EPA regulatory network.

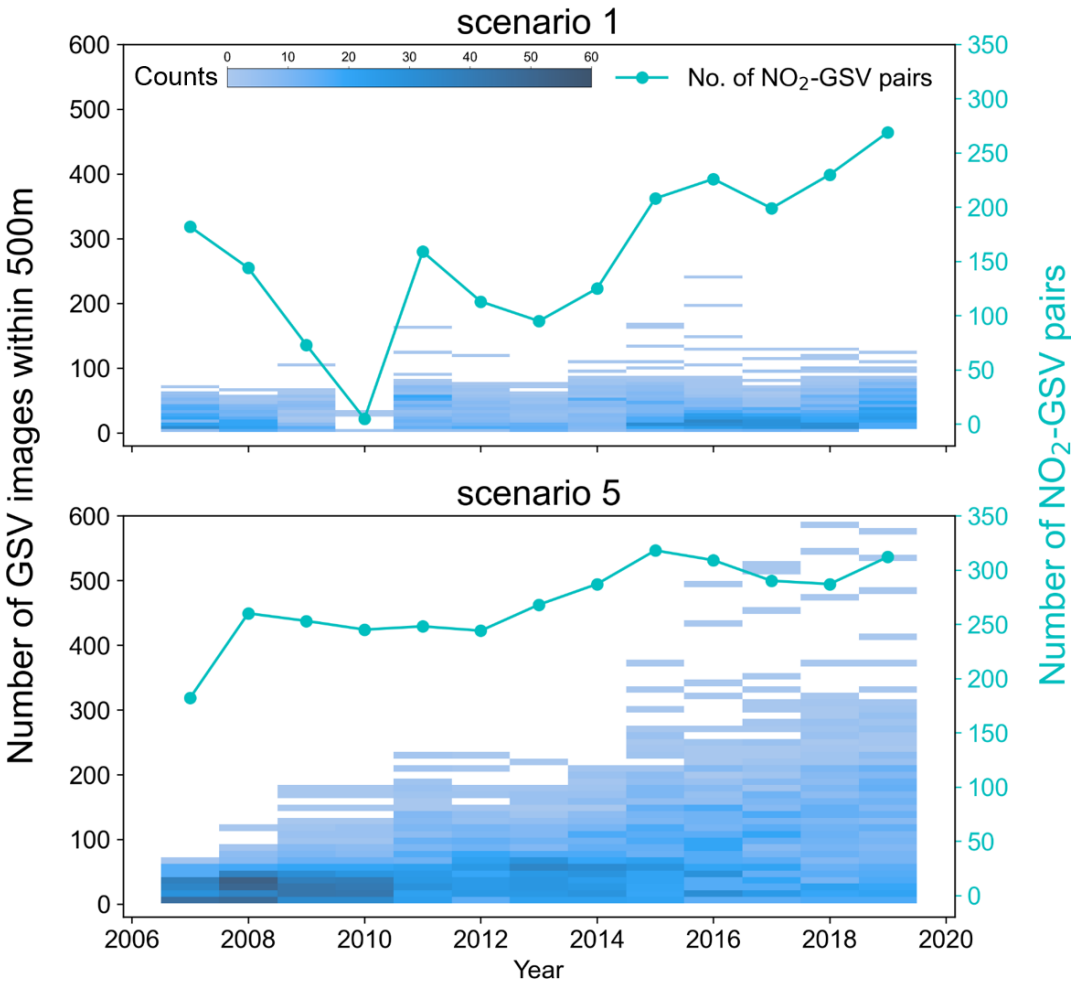

Figure 3-S2. Number of available NO_2-GSV pairs and GSV images in different years using scenario 1 and 5 for illustration. The left axis represents the number of GSV images found within 500 m of each NO_2 monitoring site in each year while the color of the stripes represents its frequency. Darker colors indicate more EPA sites with the same amount of GSV images. The right axis represents the number of NO_2-GSV pairs for model development in each year.

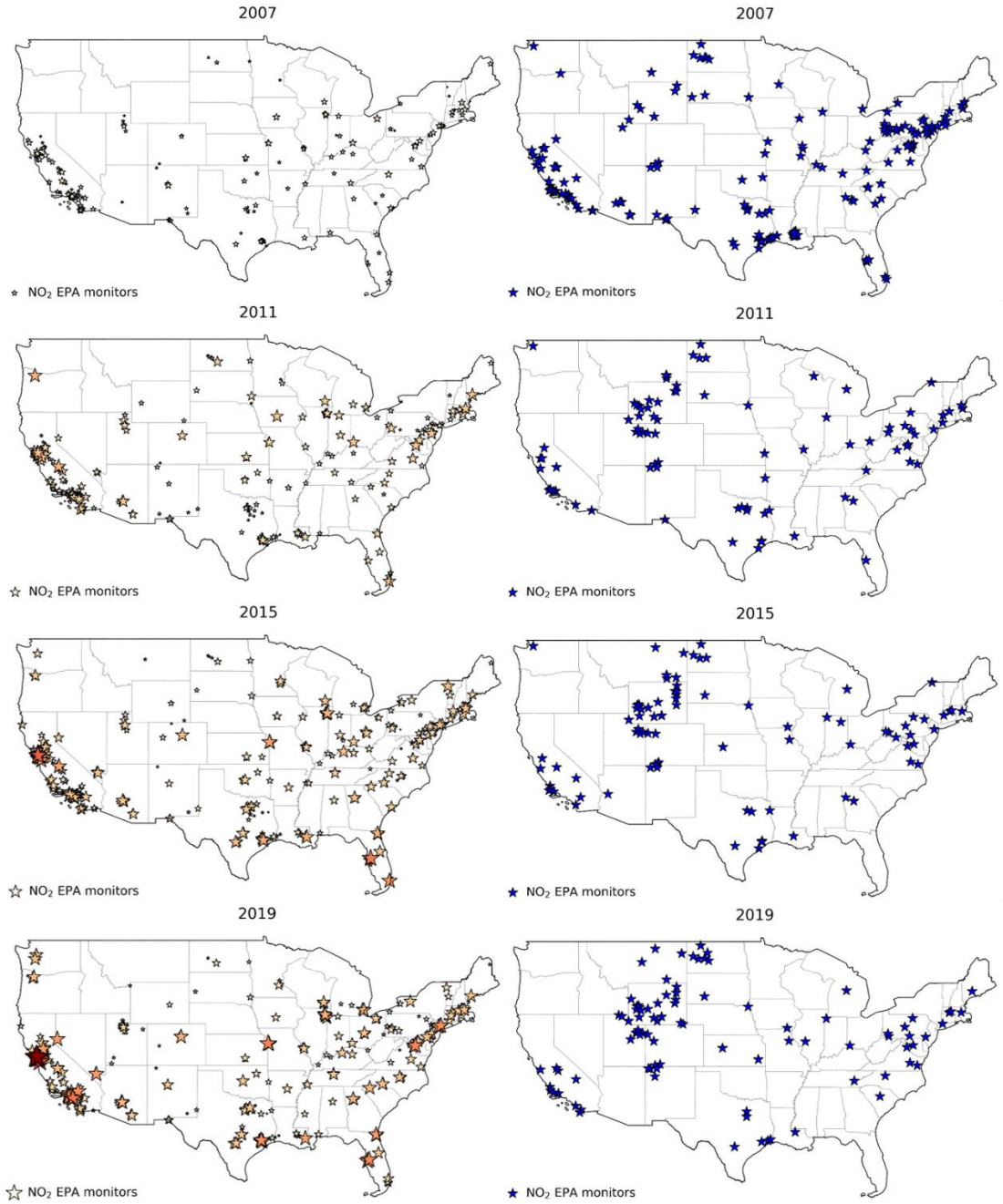

Figure 3-S3. Spatial distribution of NO_2-GSV pairs in different years using scenario 5 for illustration. The left column represents NO_2 monitors with sufficient GSV images for model development. The size and color of the stars represent the number of GSV images available within 500 m of NO_2 EPA monitors. The right column represents NO_2 monitors without sufficient GSV images that are excluded from model development. The size and color of the stars are fixed in the right column.

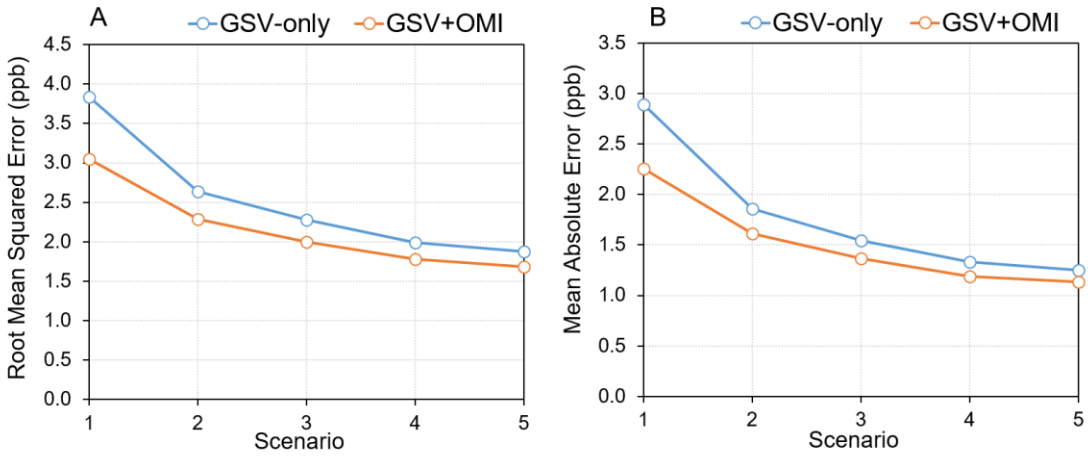

Figure 3-S4. Comparison of (A) root mean squared errors and (B) mean absolute errors between GSV-only and GSV+OMI image-unlimited models.

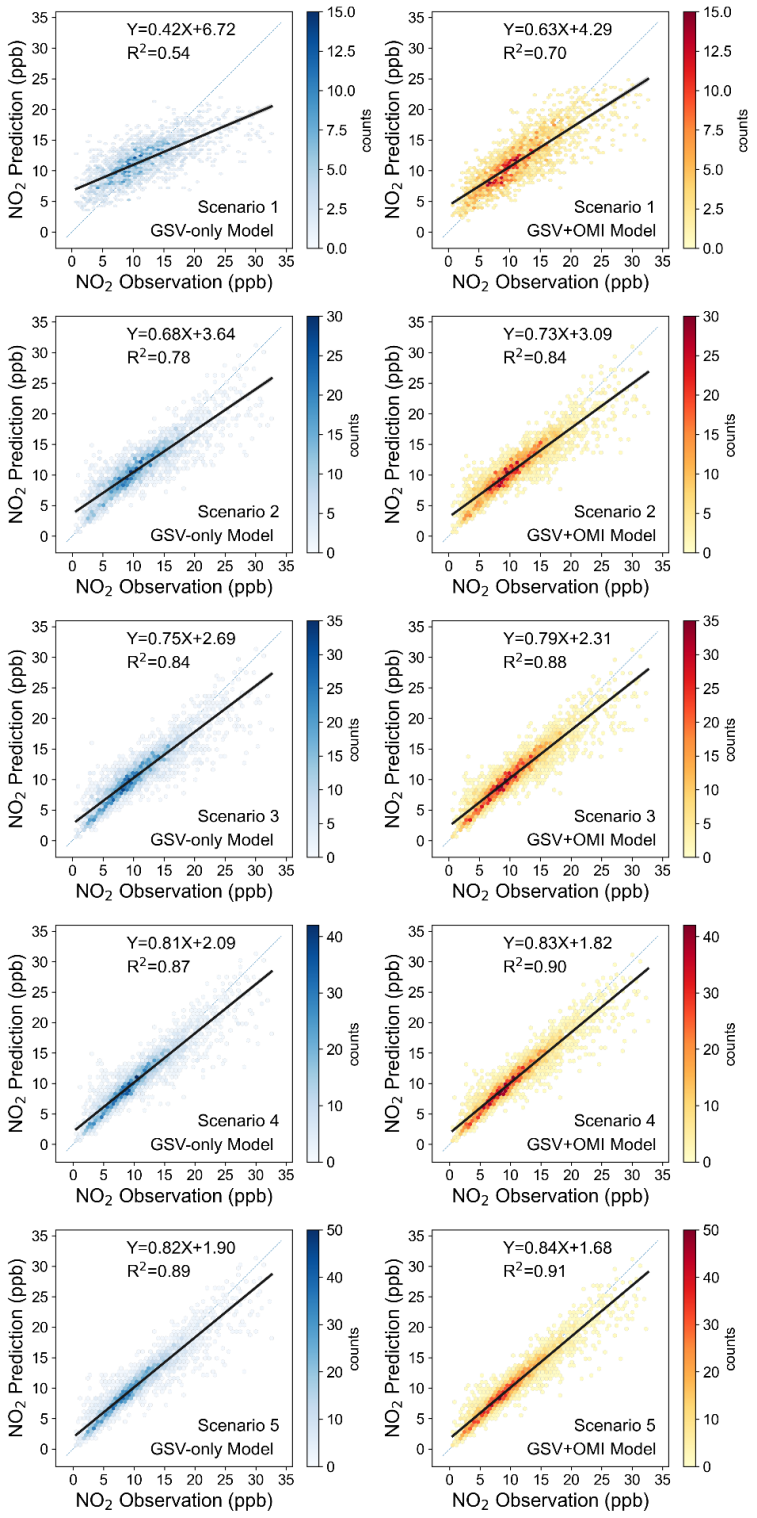

Figure 3-S5. Hexagonal binned plots of model predictions versus observations for both GSV and GSV+OMI models among different scenarios. The color reflects the number of scatters within each hexagon.

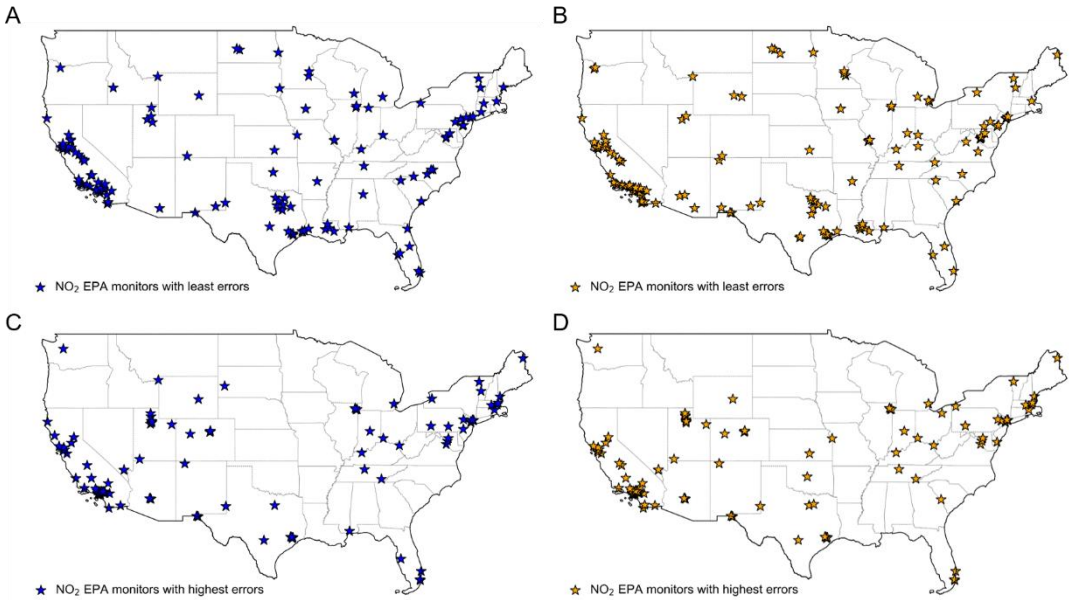

Figure 3-S6. Spatial distribution of NO_2 monitors with lowest and highest 5% errors (i.e., absolute error by GSV-only and GSV+OMI models using scenario 5). The left column represents the results of GSV-only model while the right column represents GSV+OMI model. The upper panel represents lowest 5% errors and the bottom panel represents the highest 5% errors.

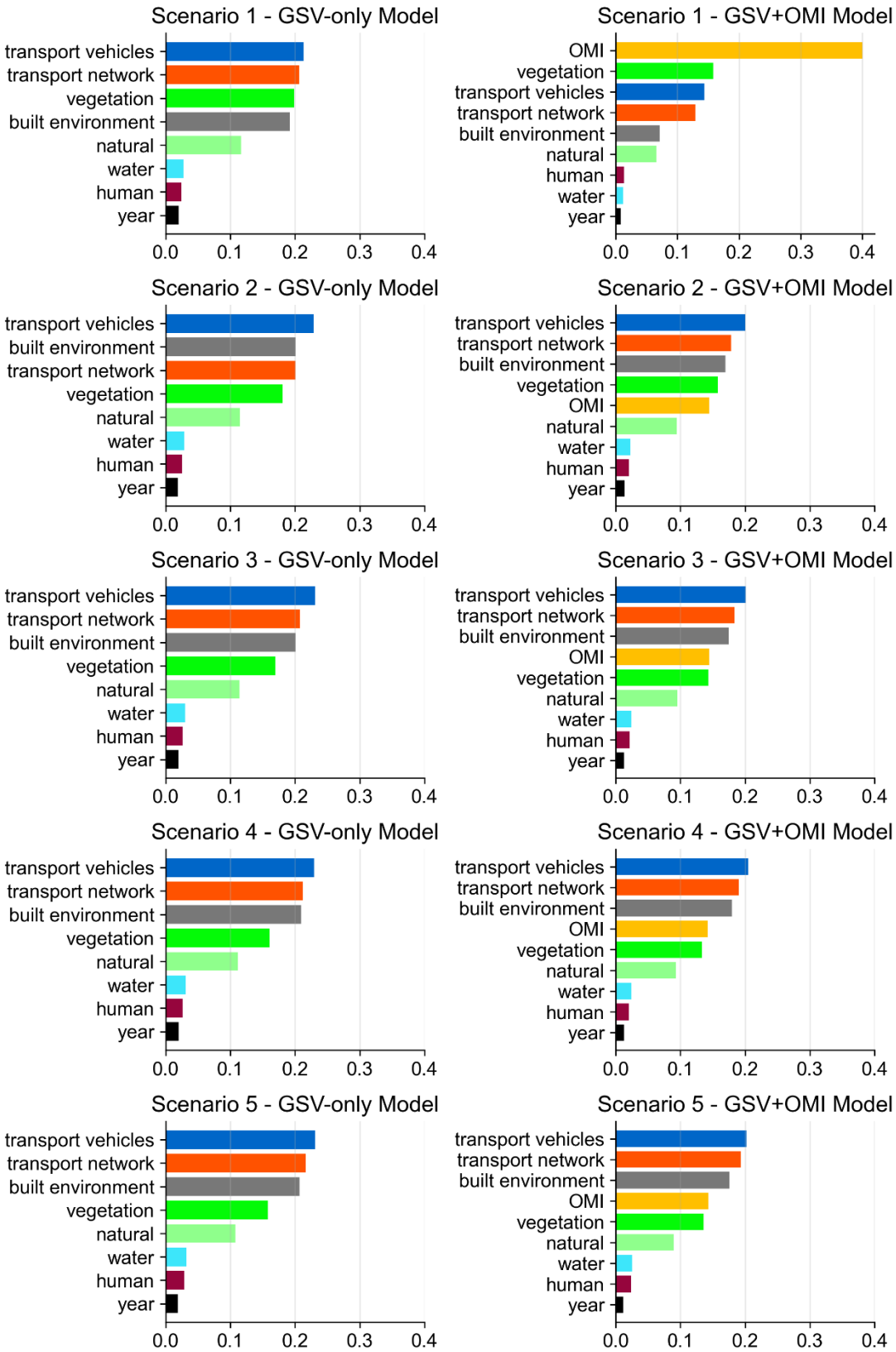

Figure 3-S7. Rank of feature importance for GSV-only and GSV+OMI models among different scenarios.

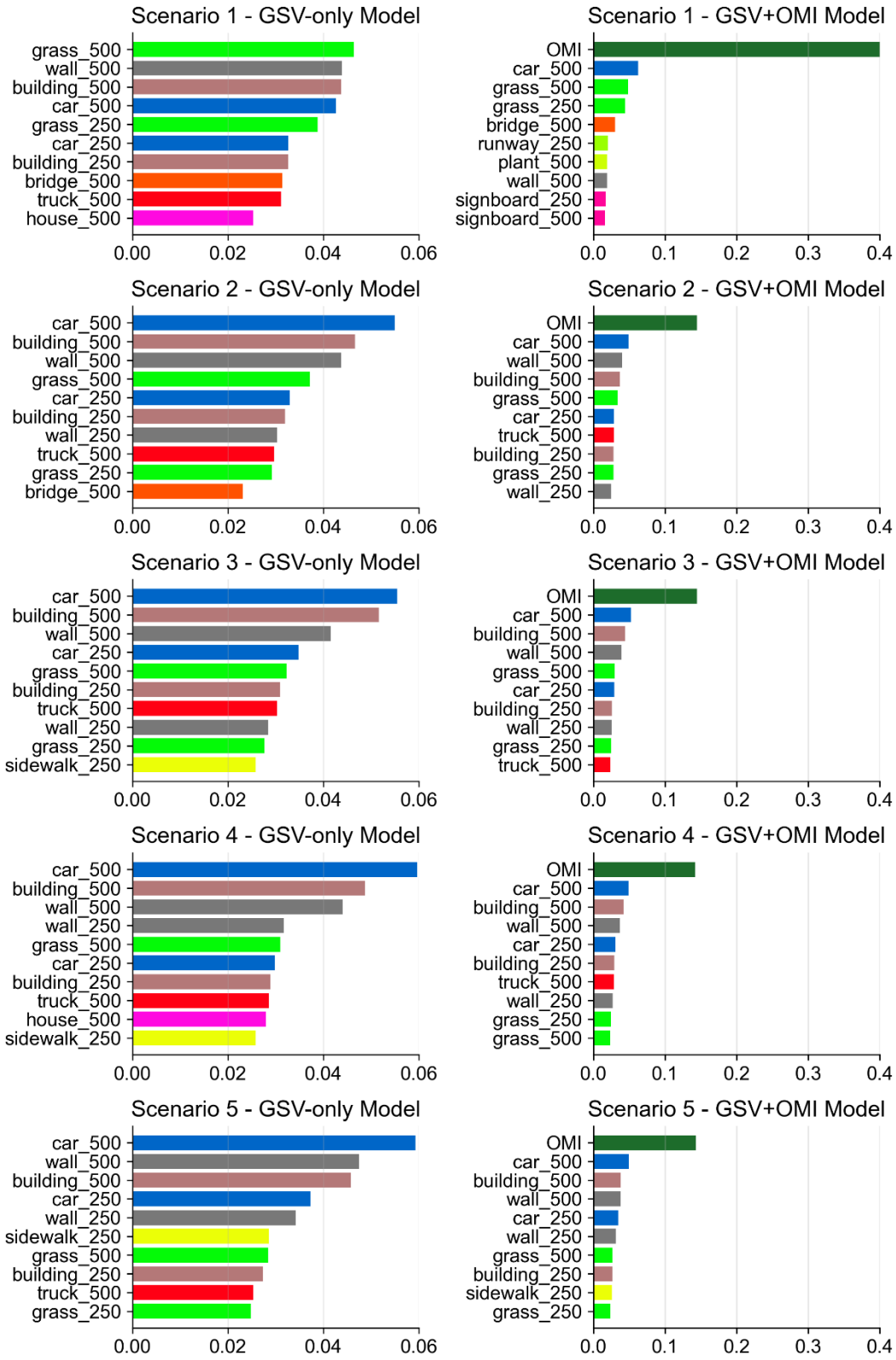

Figure 3-S8. Distribution of the top 10 important features of GSV-only and GSV+OMI models among different scenarios. The numbers after the GSV variable names represent the buffer radii.

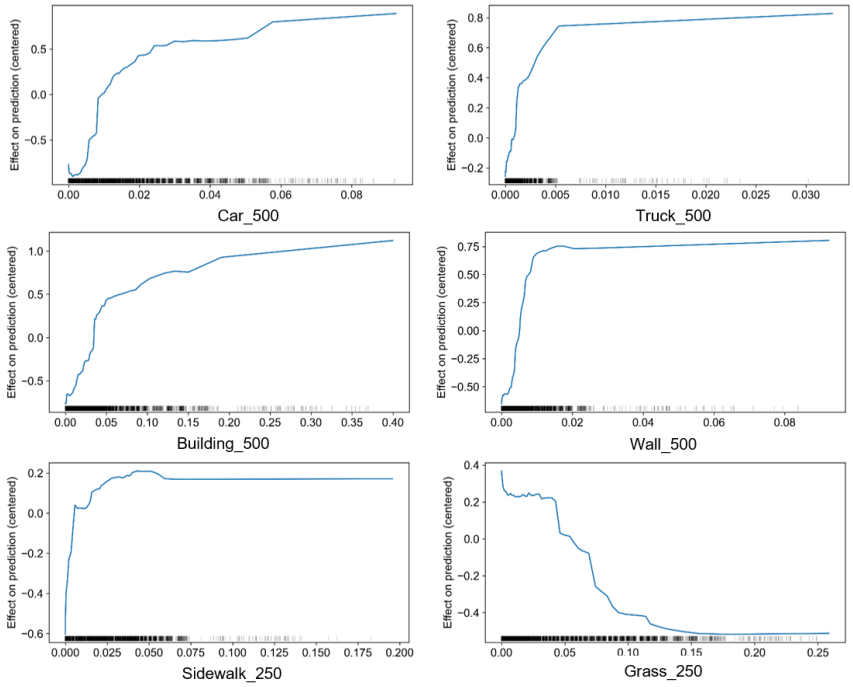

Figure 3-S9. Accumulated local effects (ALE) of several top important GSV features for the GSV-only model using scenario 5 for illustration. The suffix number of GSV features indicates the buffer level for tabulating the GSV features. The y axis shows how the features influence the prediction of GSV-only models on average. For instance, when the value of car_500 is ~0.02, the model prediction is higher by ~0.5.

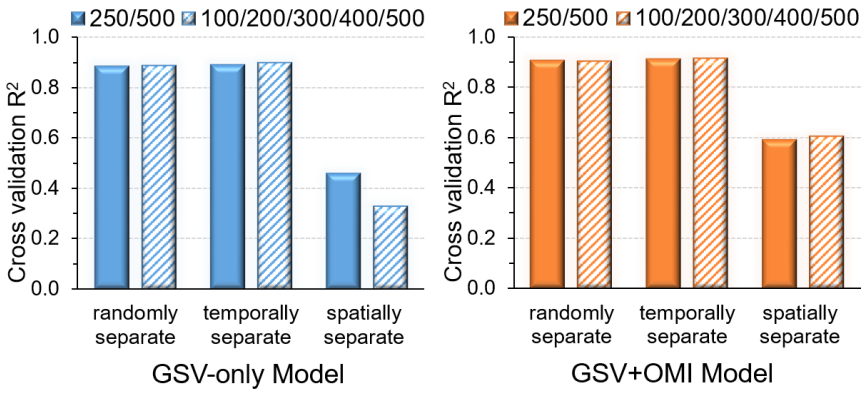

Figure 3-S10. Comparison of three cross validation strategies (i.e., splitting training and testing set randomly, temporally, or spatially) for GSV-only and GSV+OMI models using scenario 5 for illustration. Bars with solid filling represent models using 250/500 m to tabulate GSV features. Bars with stripes represent models using 100/200/300/400/500 m buffer radii.

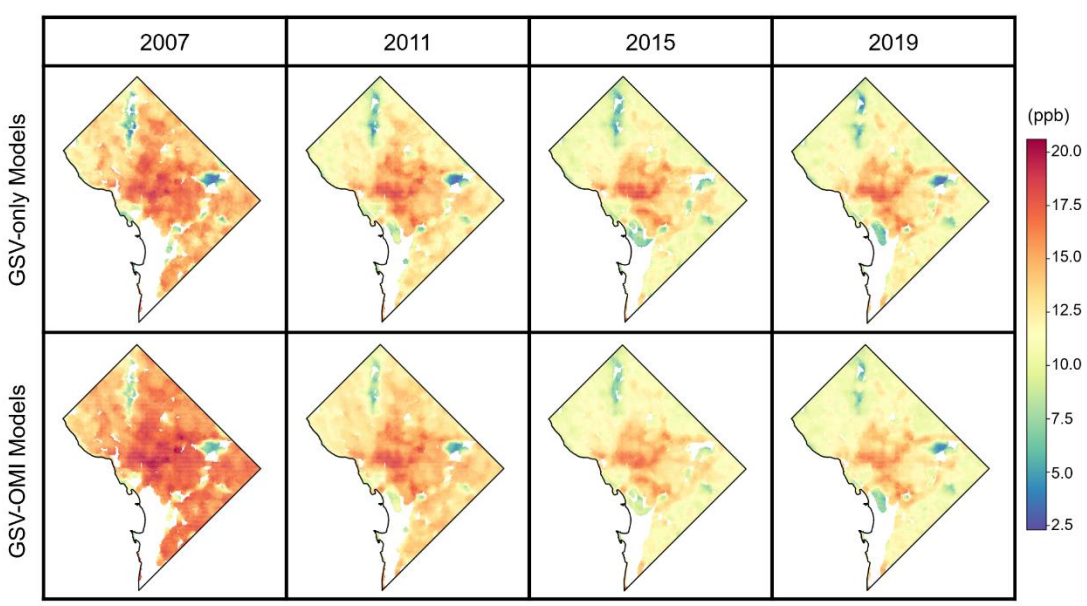

Figure 3-S11. Estimated NO_2 concentrations for Washington DC by GSV-only and GSV-OMI models using scenario 5. We display the prediction maps for 2007, 2011, 2015, and 2017 for illustration.

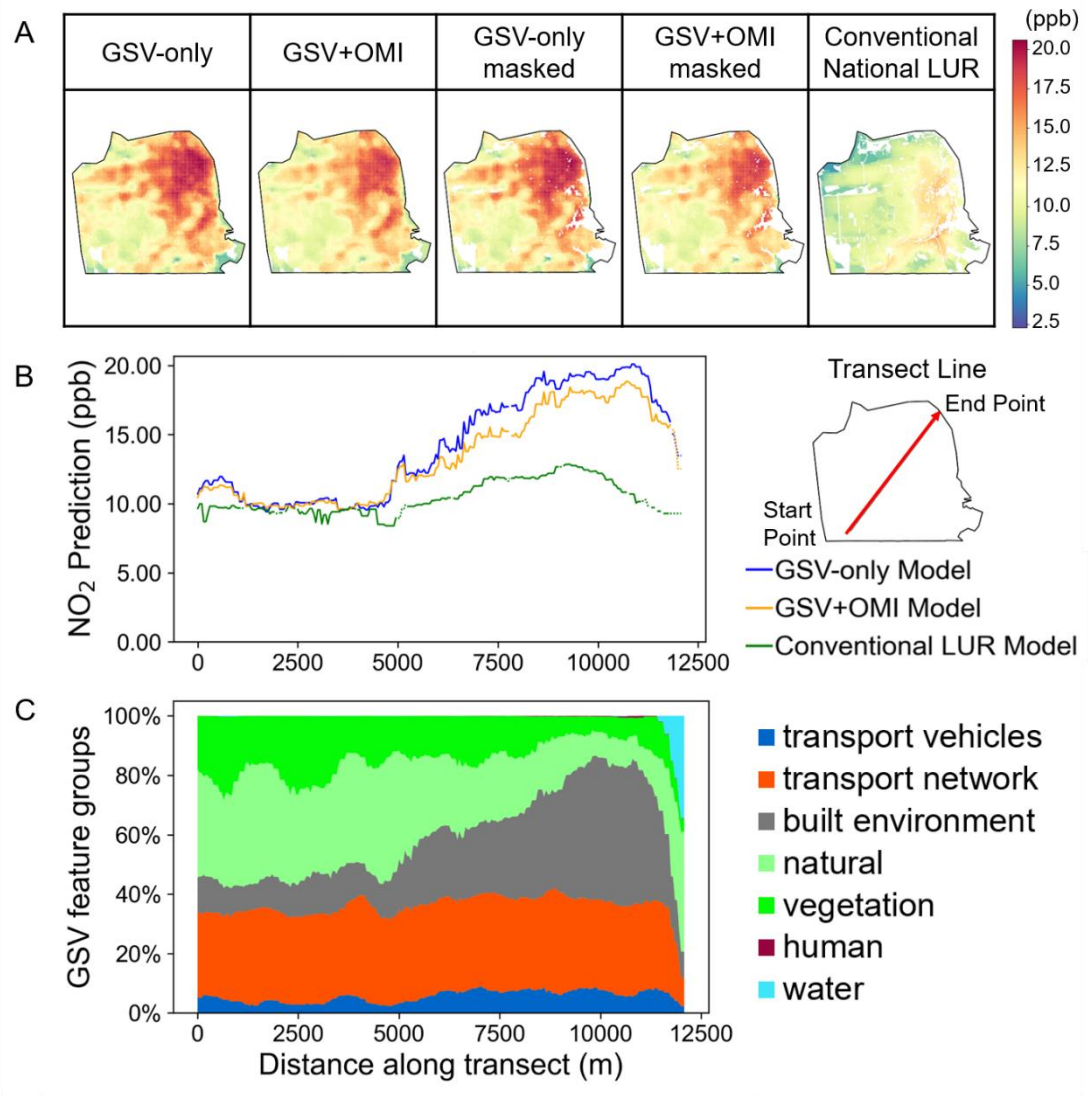

Figure 3-S12. Comparison between GSV-based models and conventional national LUR models using prediction maps for San Francisco in 2015. (A) Model predictions for the GSV-only model, GSV+OMI model, and conventional national LUR model. For comparison, we masked GSV-based prediction maps based on where the conventional LUR predictions were available. (B) Predicted NO_2 concentrations among different models along the transect line. The solid lines represent actual model predictions; gaps in the conventional national LUR model (masked GSV-based models) were filled by interpolation (original GSV-based models) and represented by dotted lines. (C) Values of GSV feature groups along the transect line.

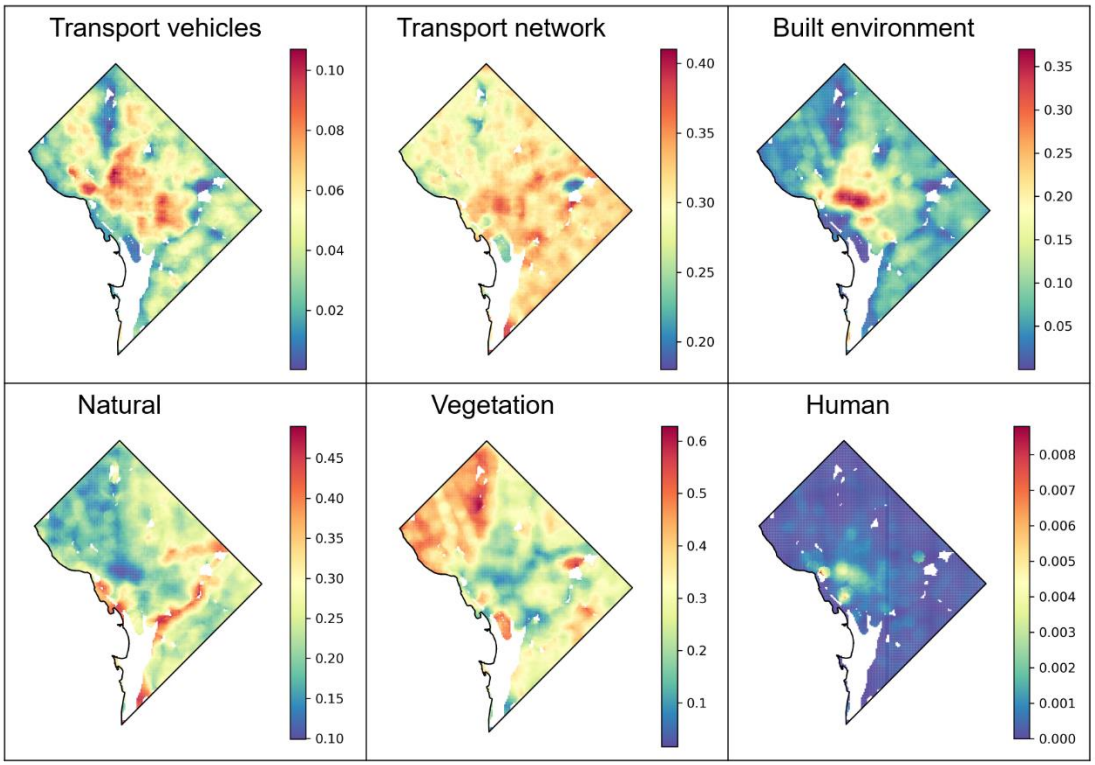

Figure 3-S13. Spatial distribution of GSV feature groups for Washington DC in 2019.

3.6 References

1. Atkinson, R. W.; Butland, B. K.; Anderson, H. R.; Maynard, R. L., Long-term concentrations of nitrogen dioxide and mortality: a meta-analysis of cohort studies. *Epidemiology* **2018,** *29*, 460-472.

2. Alotaibi, R.; Bechle, M.; Marshall, J. D.; Ramani, T.; Zietsman, J.; Nieuwenhuijsen, M. J.; Khreis, H., Traffic related air pollution and the burden of childhood asthma in the contiguous United States in 2000 and 2010. *Environ. Int.* **2019,** *127*, 858-867.

3. Collart, P.; Dubourg, D.; Levêque, A.; Sierra, N. B.; Coppieters, Y., Short-term effects of nitrogen dioxide on hospital admissions for cardiovascular disease in Wallonia, Belgium. *Int. J. Cardiol.* **2018,** *255*, 231-236.

4. He, M. Z.; Kinney, P. L.; Li, T. T.; Chen, C.; Sun, Q. H.; Ban, J.; Wang, J. N.; Liu, S. L.; Goldsmith, J.; Kioumourtzoglou, M. A., Short- and intermediate-term exposure to NO_2 and mortality: A multi-county analysis in China. *Environ. Pollut.* **2020,** *261*, 114165.

5. Achakulwisut, P.; Brauer, M.; Hystad, P.; Anenberg, S. C., Global, national, and urban burdens of paediatric asthma incidence attributable to ambient NO_2 pollution: estimates from global datasets. *Lancet Planet. Health* **2019,** *3*, E166-E178.

6. Crouse, D. L.; Peters, P. A.; Hystad, P.; Brook, J. R.; van Donkelaar, A.; Martin, R. V.; Villeneuve, P. J.; Jerrett, M.; Goldberg, M. S.; Pope, C. A.; Brauer, M.; Brook, R. D.; Robichaud, A.; Menard, R.; Burnett, R. T., Ambient $PM_{2.5}$, O_3, and NO_2 Exposures and Associations with Mortality over 16 Years of Follow-Up in the Canadian Census Health and Environment Cohort (CanCHEC). *Environ. Health Perspect.* **2015,** *123*, 1180-1186.

7. Demetillo, M. A. G.; Navarro, A.; Knowles, K. K.; Fields, K. P.; Geddes, J. A.;

Nowlan, C. R.; Janz, S. J.; Judd, L. M.; Al-Saadi, J.; Sun, K.; McDonald, B. C.; Diskin, G. S.; Pusede, S. E., Observing nitrogen dioxide air pollution inequality using high-spatial-resolution remote sensing measurements in Houston, Texas. *Environ. Sci. Technol.* **2020,** *54*, 9882-9895.

8. Karner, A. A.; Eisinger, D. S.; Niemeier, D. A., Near-roadway air quality: synthesizing the findings from real-world data. *Environ. Sci. Technol.* **2010,** *44*, 5334-5344.

9. Zhou, Y.; Levy, J. I., Factors influencing the spatial extent of mobile source air pollution impacts: a meta-analysis. *BMC Public Health* **2007,** *7*, 89.

10. 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.; Hamburg, S. P., High-resolution air pollution mapping with Google Street View cars: Exploiting big data. *Environ. Sci. Technol.* **2017,** *51*, 6999-7008.

11. Chambliss, S. E.; Pinon, C. P.; Messier, K. P.; LaFranchi, B.; Upperman, C. R.; Lunden, M. M.; Robinson, A. L.; Marshall, J. D.; Apte, J. S., Local-and regional-scale racial and ethnic disparities in air pollution determined by long-term mobile monitoring. *Proc. Natl. Acad. Sci. U.S.A.* **2021,** *118*, e2109249118.

12. Southerland, V. A.; Anenberg, S. C.; Harris, M.; Apte, J.; Hystad, P.; van Donkelaar, A.; Martin, R. V.; Beyers, M.; Roy, A., Assessing the distribution of air pollution health risks within cities: A neighborhood-scale analysis leveraging high-resolution data sets in the Bay Area, California. *Environ. Health Perspect.* **2021,** *129*, 037006.

13. Hoek, G.; Beelen, R.; de Hoogh, K.; Vienneau, D.; Gulliver, J.; Fischer, P.; Briggs, D., A review of land-use regression models to assess spatial variation of outdoor air pollution. *Atmos. Environ.* **2008,** *42*, 7561-7578.

14. Hankey, S.; 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.* **2015,** *49*, 9194-9202.

15. De Hoogh, K.; Chen, J.; Gulliver, J.; Hoffmann, B.; Hertel, O.; Ketzel, M.; Bauwelinck, M.; van Donkelaar, A.; Hvidtfeldt, U. A.; Katsouyanni, K., Spatial $PM_{2.5}$, NO_2, O_3 and BC models for Western Europe–Evaluation of spatiotemporal stability. *Environ. Int.* **2018,** *120*, 81-92.

16. 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.; Apte, J. S., Mapping air pollution with Google Street View cars: Efficient approaches with mobile monitoring and land use regression. *Environ. Sci. Technol.* **2018,** *52*, 12563-12572.

17. Liao, K.; Huang, X. H.; Dang, H. F.; Ren, Y.; Zuo, S. D.; Duan, C. S., Statistical approaches for forecasting primary air pollutants: A review. *Atmosphere* **2021,** *12*, 686.

18. Kim, S. Y.; Bechle, M.; Hankey, S.; Sheppard, L.; Szpiro, A. A.; Marshall, J. D., Concentrations of criteria pollutants in the contiguous US, 1979-2015: Role of prediction model parsimony in integrated empirical geographic regression. *PLoS One* **2020,** *15*, e0228535.

19. Novotny, E. V.; Bechle, M. J.; Millet, D. B.; Marshall, J. D., National satellite-based land-use regression: NO_2 in the United States. *Environ. Sci. Technol.* **2011,** *45*, 4407-4414.

20. Bechle, M. J.; Millet, D. B.; Marshall, J. D., National spatiotemporal exposure surface for NO_2: monthly scaling of a satellite-derived land-use regression, 2000–2010. *Environ. Sci. Technol.* **2015,** *49*, 12297-12305.

21. Young, M. T.; Bechle, M. J.; Sampson, P. D.; Szpiro, A. A.; Marshall, J. D.;

Sheppard, L.; Kaufman, J. D., Satellite-based NO_2 and model validation in a national prediction model based on universal kriging and land-use regression. *Environ. Sci. Technol.* **2016,** *50*, 3686-3694.

22. Knibbs, L. D.; Hewson, M. G.; Bechle, M. J.; Marshall, J. D.; Barnett, A. G., A national satellite-based land-use regression model for air pollution exposure assessment in Australia. *Environ. Res.* **2014,** *135*, 204-211.

23. Knibbs, L. D.; Van Donkelaar, A.; Martin, R. V.; Bechle, M. J.; Brauer, M.; Cohen, D. D.; Cowie, C. T.; Dirgawati, M.; Guo, Y.; Hanigan, I. C., Satellite-based land-use regression for continental-scale long-term ambient $PM_{2.5}$ exposure assessment in Australia. *Environ. Sci. Technol.* **2018,** *52*, 12445-12455.

24. Vienneau, D.; De Hoogh, K.; Bechle, M. J.; Beelen, R.; Van Donkelaar, A.; Martin, R. V.; Millet, D. B.; Hoek, G.; Marshall, J. D., Western European land use regression incorporating satellite-and ground-based measurements of NO_2 and PM_{10}. *Environ. Sci. Technol.* **2013,** *47*, 13555-13564.

25. De Hoogh, K.; Gulliver, J.; van Donkelaar, A.; Martin, R. V.; Marshall, J. D.; Bechle, M. J.; Cesaroni, G.; Pradas, M. C.; Dedele, A.; Eeftens, M., Development of West-European $PM_{2.5}$ and NO_2 land use regression models incorporating satellite-derived and chemical transport modelling data. *Environ. Res.* **2016,** *151*, 1-10.

26. Xu, H.; Bechle, M. J.; Wang, M.; Szpiro, A. A.; Vedal, S.; Bai, Y.; Marshall, J. D., National $PM_{2.5}$ and NO_2 exposure models for China based on land use regression, satellite measurements, and universal kriging. *Sci. Total Environ.* **2019,** *655*, 423-433.

27. Zhang, Z.; Wang, J.; Hart, J. E.; Laden, F.; Zhao, C.; Li, T.; Zheng, P.; Li, D.; Ye, Z.; Chen, K., National scale spatiotemporal land-use regression model for $PM_{2.5}$, PM_{10} and NO_2 concentration in China. *Atmos. Environ.* **2018,** *192*, 48-54.

28. Ryan, P. H.; LeMasters, G. K., A review of land-use regression models for

characterizing intraurban air pollution exposure. *Inhal. Toxicol.* **2007,** *19*, 127-133.

29. Larkin, A.; Geddes, J. A.; Martin, R. V.; Xiao, Q. Y.; Liu, Y.; Marshall, J. D.; Brauer, M.; Hystad, P., Global land use regression model for nitrogen dioxide air pollution. *Environ. Sci. Technol.* **2017,** *51*, 6957-6964.

30. Hong, K. Y.; Pinheiro, P. O.; Minet, L.; Hatzopoulou, M.; Weichenthal, S., Extending the spatial scale of land use regression models for ambient ultrafine particles using satellite images and deep convolutional neural networks. *Environ. Res.* **2019,** *176*, 108513.

31. Weissert, L.; Salmond, J.; Miskell, G.; Alavi-Shoshtari, M.; Williams, D., Development of a microscale land use regression model for predicting NO_2 concentrations at a heavy trafficked suburban area in Auckland, NZ. *Sci. Total Environ.* **2018,** *619*, 112-119.

32. Lu, T.; Marshall, J. D.; Zhang, W.; Hystad, P.; Kim, S.-Y.; Bechle, M. J.; Demuzere, M.; Hankey, S., National empirical models of air pollution using microscale measures of the urban environment. *Environ. Sci. Technol.* **2021,** *55*, 15519–15530.

33. Lloyd, M.; Carter, E.; Diaz, F. G.; Magara-Gomez, K. T.; Hong, K. Y.; Baumgartner, J.; Herrera G, V. c. M.; Weichenthal, S., Predicting within-city spatial variations in outdoor ultrafine particle and black carbon concentrations in Bucaramanga, Colombia: A hybrid approach using open-source geographic data and digital images. *Environ. Sci. Technol.* **2021,** *55*, 12483-12492.

34. Zhong, S.; Zhang, K.; Bagheri, M.; Burken, J. G.; Gu, A.; Li, B.; Ma, X.; Marrone, B. L.; Ren, Z. J.; Schrier, J., Machine learning: New ideas and tools in environmental science and engineering. *Environ. Sci. Technol.* **2021,** *55*, 12741-12754.

35. Weichenthal, S.; Hatzopoulou, M.; Brauer, M., A picture tells a thousand...exposures: Opportunities and challenges of deep learning image analyses in

exposure science and environmental epidemiology. *Environ. Int.* **2019,** *122*, 3-10.

36. Qi, M.; Hankey, S., Using street view imagery to predict street-level particulate air pollution. *Environ. Sci. Technol.* **2021,** *55*, 2695-2704.

37. Hong, K. Y.; Pinheiro, P. O.; Weichenthal, S., Predicting outdoor ultrafine particle number concentrations, particle size, and noise using street-level images and audio data. *Environ. Int.* **2020,** *144*, 106044.

38. Ganji, A.; Minet, L.; Weichenthal, S.; Hatzopoulou, M., Predicting traffic-related air pollution using feature extraction from built environment images. *Environ. Sci. Technol.* **2020,** *54*, 10688–10699.

39. Jain, S.; Presto, A. A.; Zimmerman, N., Spatial Modeling of Daily $PM_{2.5}$, NO_2, and CO Concentrations Measured by a Low-Cost Sensor Network: Comparison of Linear, Machine Learning, and Hybrid Land Use Models. *Environ. Sci. Technol.* **2021,** *55*, 8631–8641.

40. Lim, C. C.; Kim, H.; Vilcassim, M. J. R.; Thurston, G. D.; Gordon, T.; Chen, L. C.; Lee, K.; Heimbinder, M.; Kim, S. Y., Mapping urban air quality using mobile sampling with low-cost sensors and machine learning in Seoul, South Korea. *Environ. Int.* **2019,** *131*, 105022.

41. Vu, T. V.; Shi, Z. B.; Cheng, J.; Zhang, Q.; He, K. B.; Wang, S. X.; Harrison, R. M., Assessing the impact of clean air action on air quality trends in Beijing using a machine learning technique. *Atmos. Chem. Phys.* **2019,** *19*, 11303-11314.

42. Xiong, L.-H.; Huang, S.; Huang, Y.; Yin, F.; Yang, F.; Zhang, Q.; Cheng, J.; Zhang, R.; He, X., Ultrasensitive Visualization of Virus via Explosive Catalysis of an Enzyme Muster Triggering Gold Nano-aggregate Disassembly. *ACS Appl. Mater. Interfaces* **2020,** *12*, 12525-12532.

43. Kim, M.; Brunner, D.; Kuhlmann, G., Importance of satellite observations for

high-resolution mapping of near-surface NO_2 by machine learning. *Remote Sens. Environ.* **2021,** *264*, 112573.

44. Chan, K. L.; Khorsandi, E.; Liu, S.; Baier, F.; Valks, P., Estimation of surface NO_2 concentrations over Germany from TROPOMI satellite observations using a machine learning method. *Remote Sens.* **2021,** *13*, 969.

45. Wu, C.-l.; Song, R.-f.; Peng, Z.-r., Prediction of air pollutants on roadside of the elevated roads with combination of pollutants periodicity and deep learning method. *Build. Environ.* **2022,** *207*, 108436.

46. Kokkinos, K.; Karayannis, V.; Nathanail, E.; Moustakas, K., A comparative analysis of Statistical and Computational Intelligence methodologies for the prediction of traffic-induced fine particulate matter and NO_2. *J. Clean. Prod.* **2021,** *328*, 129500.

47. Di, Q.; Amini, H.; Shi, L.; Kloog, I.; Silvern, R.; Kelly, J.; Sabath, M. B.; Choirat, C.; Koutrakis, P.; Lyapustin, A., Assessing NO_2 concentration and model uncertainty with high spatiotemporal resolution across the contiguous United States using ensemble model averaging. *Environ. Sci. Technol.* **2019,** *54*, 1372-1384.

48. Wong, P.-Y.; Su, H.-J.; Lee, H.-Y.; Chen, Y.-C.; Hsiao, Y.-P.; Huang, J.-W.; Teo, T.-A.; Wu, C.-D.; Spengler, J. D., Using land-use machine learning models to estimate daily NO_2 concentration variations in Taiwan. *J. Clean. Prod.* **2021,** *317*, 128411.

49. Zhao, H. S.; Shi, J. P.; Qi, X. J.; Wang, X. G.; Jia, J. Y.; Ieee, Pyramid Scene Parsing Network. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* **2017**, 2881-2890.

50. Levelt, P. F.; Van Den Oord, G. H.; Dobber, M. R.; Malkki, A.; Visser, H.; De Vries, J.; Stammes, P.; Lundell, J. O.; Saari, H., The ozone monitoring instrument. *IEEE Trans. Geosci. Remote Sens.* **2006,** *44*, 1093-1101.

51. Hoek, G.; Eeftens, M.; Beelen, R.; Fischer, P.; Brunekreef, B.; Boersma, K. F.; Veefkind, P., Satellite NO_2 data improve national land use regression models for

ambient NO_2 in a small densely populated country. *Atmos. Environ.* **2015,** *105*, 173-180.

52. Lamsal, L. N.; Krotkov, N. A.; Marchenko, S. V.; Joiner, J.; Oman, L.; Vasilkov, A.; Fisher, B.; Qin, W.; Yang, E.-S.; Fasnacht, Z.; Choi, S.; Leonard, P.; Haffner, D., OMI/Aura NO_2 Tropospheric, Stratospheric & Total Columns MINDS Daily L3 Global Gridded 0.25 degree x 0.25 degree, NASA Goddard Space Flight Center, Goddard Earth Sciences Data and Information Services Center (GES DISC), Accessed: [Aug. 1, 2021], 10.5067/MEASURES/MINDS/DATA301. **2020**.

53. Hart, J. E.; Yanosky, J. D.; Puett, R. C.; Ryan, L.; Dockery, D. W.; Smith, T. J.; Garshick, E.; Laden, F., Spatial modeling of PM_{10} and NO_2 in the continental United States, 1985–2000. *Environ. Health Perspect.* **2009,** *117*, 1690-1696.

54. Gulliver, J.; de Hoogh, K.; Hansell, A.; Vienneau, D., Development and Back-Extrapolation of NO_2 Land Use Regression Models for Historic Exposure Assessment in Great Britain. *Environ. Sci. Technol.* **2013,** *47*, 7804-7811.

55. Cordioli, M.; Pironi, C.; De Munari, E.; Marmiroli, N.; Lauriola, P.; Ranzi, A., Combining land use regression models and fixed site monitoring to reconstruct spatiotemporal variability of NO_2 concentrations over a wide geographical area. *Sci. Total Environ.* **2017,** *574*, 1075-1084.

56. Zheng, T.; Bergin, M. H.; Hu, S.; Miller, J.; Carlson, D. E., Estimating ground-level $PM_{2.5}$ using micro-satellite images by a convolutional neural network and random forest approach. *Atmos. Environ.* **2020,** *230*, 117451.

57. Weichenthal, S.; Dons, E.; Hong, K. Y.; Pinheiro, P. O.; Meysman, F. J., Combining citizen science and deep learning for large-scale estimation of outdoor nitrogen dioxide concentrations. *Environ. Res.* **2021,** *196*, 110389.

58. Lefevre, S.; Tuia, D.; Wegner, J. D.; Produit, T.; Nassar, A. S., Toward seamless

multiview scene analysis from satellite to street level. *Proc. IEEE* **2017,** *105*, 1884-1899.

Chapter 4 Assessment of Street View and Remote Sensing Imagery derived parameters in estimating national NO_2 and $PM_{2.5}$ pollution

*Meng Qi, Steve Hankey**

School of Public and International Affairs, Virginia Tech, Blacksburg 24061 Virginia, United States

* Corresponding author, School of Public and International Affairs, Virginia Tech

Email: hankey@vt.edu. Phone: 540.231.7508

The authors declare no competing interest.

ABSTRACT: Land use regression (LUR) models have been widely used for estimating fine-scale ambient air pollutant concentrations. Street view and remote sensing imagery are promising in offering natural and built environment features which may be used to replace traditional LUR predictors. In this study, we investigated using different combinations of street view and remote sensing derived features to develop national NO_2 and $PM_{2.5}$ machine learning models. We collected annual NO_2 and $PM_{2.5}$ concentrations from the U.S. EPA monitoring network during 2007-2019. The modeling predictors included (1) Google Street View [GSV] derived features extracted from GSV imagery using a deep learning model, (2) satellite-based aerosol estimates (i.e., Ozone Monitoring Instrument [OMI] for NO_2 and aerosol optical depth [AOD] for $PM_{2.5}$), and (3) Landsat-derived features from with 12 bands and derived spectral indices. For both NO_2 and $PM_{2.5}$, the full models (i.e., using all image-derived features) achieved the best model performance. The random 10-fold cross validation R^2 was 0.91 for the NO_2 full model and 0.72 for the $PM_{2.5}$ full model. Models involving GSV features generally show enhanced model fits compared to models solely using remote sensing derived predictors. Despite that the difference of mode fits being relatively small among models, the spatial pattern of model prediction may vary significantly. Although the mean pollutant concentrations by different models were similar, an examination of model prediction at U.S. urbanized areas suggested that the GSV-based models captured both the highest and lowest pollutant concentrations while Landsat features tended to smooth the air pollution variations. The results indicated that GSV features may have a better capability to capture fine-scale air pollution.

Keywords: remote sensing, street view, model fits, spatial pattern, air pollution, machine learning

4.1 Introduction

The detrimental effects of air pollution have been extensively studied and well-documented[1-5]. Due to the sparsity of traditional ground monitoring network, air quality models and resulting air pollution projections are often used to serve epidemiology studies[6-8]. Since significant variation exists in the spatial-temporal distribution of air pollution and its emission sources[9, 10], high resolution estimates of air pollution which is capable of capturing local hot spots are a high priority. The precision and reliability of air pollution exposure and health effects estimates is based on the premise of high model performance and precision. Land use regression (LUR) is a popular empirical modeling approach for deriving high-resolution estimation of ambient air pollutant concentrations[11-13]. Liao et al. used bibliometrics to quantitatively review the statistical air pollution prediction models spanning 1990-2018 and found that LUR has been the most popular method for NOx prediction in the last decade.[14] LUR models are often applied to long-term air pollution studies (e.g., estimating annual mean concentrations), and the results have been widely used in exposure science and epidemiological studies.[15-18]

Many traditional LUR studies developed linear regression models based on the assumption of linear relationship between air pollutants and traditional LUR predictors (i.e., traffic, land use)[19]. However, the interactions between air pollution, its emission sources and other impacting factors can be complex and nonlinear. Thus, nonlinear models are needed to capture the complexity of air pollution dynamics[20, 21]. Moreover, traditional LUR predictors are typically collected from curated GIS databases, which are often limited either in spatial resolution or hard to collect over large geography. Benefited from the fast advance in the field of artificial intelligence [AI], an increasing number of studies have applied machine learning [ML] algorithms in air quality

modeling and achieved good performance[22-25]. Besides more advanced modeling algorithms, AI promotes the generation of new technologies and unlocked the potential of many data sources that were unavailable or too resource intensive in the past[23]. For example, numerous open-source imagery datasets are emerging each day, such as street view imagery, high-resolution remote sensing data, etc[26]. Some existing data sources (e.g., medium-resolution remote sensing imagery with long record) which may not have been feasible to utilize in the past now may be promising for multiple applications with the aid of advanced machine learning algorithms.

Our previous efforts have shown that open street view imagery is promising for replacing traditional LUR predictors[27, 28]. We successfully estimated street level air pollution with high model performance and resolution, but no large-scale pollution projection was generated partially due to computation constraints. Despite that the coverage and density of street view imagery increasing rapidly, it's challenging for remote areas to collect sufficient street view images for air quality modeling[27]. Various remote sensing products have wall-to-wall coverage and open access, thus are also promising for large scale air quality models with consistent data collection and data processing. Moreover, satellite imagery often has stable temporal resolution though its spatial resolution typically doesn't compete with street view imagery. The combination of street view imagery and remote sensing imagery may complement each other and thus maximize the benefits for air quality modeling.

In this study, we aim to use different combinations of image-derived features to develop national NO_2 and $PM_{2.5}$ models with the overarching goal of generating high-resolved time-series air pollution maps across the continental U.S. Three types of image-derived features were tested, including (1) street view derived variables, (2) remote sensing derived aerosol estimates, and (3) Landsat derived variables. Random

forest models were developed for capturing the complex nonlinear relationship between air pollutants and the predictor features. We investigate how different features contribute to the explanation of model variance and impact the model prediction surface. Our results indicate that street view derived features may have the capability to better capture fine-scale air pollution variability, which may further aid and provide insights for environmental justice, exposure assessment and health studies.

4.2 Materials and methods

4.2.1 Modeling variables

Our previous work successfully developed national NO_2 models using GSV-derived features and satellite derived aerosol estimates (i.e., OMI tropospheric column density)[28]. This study expanded on the effort from the previous work and investigated the inclusion of additional satellite-derived variables, i.e., Landsat-derived features. Moreover, national $PM_{2.5}$ models were established to explore the model reproducibility. Annual-average NO_2 and $PM_{2.5}$ concentrations at EPA monitors across the contiguous U.S. during 2007-2019 were collected as model dependent variables.

4.2.1.1 Street view derived variables.

We extracted GSV features from a national GSV raster database for both $PM_{2.5}$ and NO_2 models as the independent variables. The database collected ~350 million GSV images across the U.S. and semantically segmented the GSV images through a deep learning model called Pyramid Scene Parsing Network (PSPNet)[29]. By applying PSPNet, each pixel of the GSV image was assigned to a category representing different built and natural environment, e.g, tree, road, car, sidewalk, etc. The sum of the percentage for each category was aggregated for each GSV image. The GSV raster database was then aggregated at 100m spatial resolution. Each raster grid cell represents the mean percentage of GSV features among all GSV images located within the grid.

Following our previous work, a total of 47 GSV feature categories were used in the model development which were then divided into 7 subgroups: built environment, transport network, transport vehicles, natural, vegetation, water, and human (**Table 4-S1**). To improve temporal consistency, we constrained the selection of GSV features so that only images taken from the same year or within the previous 4 years of the air pollutant observations were matched to the corresponding dependent variables. More detailed can be found in Qi et al[28]. For model input, we extracted the GSV features at 8 buffers with radii ranging from 250m to 2000m with an interval of 250m.

4.2.1.2 Remote sensing derived aerosol estimates.

We collected OMI tropospheric column density for NO_2 and aerosol optical depth [AOD] for $PM_{2.5}$ as another type of model independent variables. The OMI NO_2 feature was obtained from the OMI/Aura NO_2 Tropospheric, Stratospheric & Total Columns MINDS Daily L3 Global Gridded 0.25° × 0.25° product provided by the NASA Goddard Earth Sciences Data and Information Services Center [GESDISC][30]. We used the cloud screened OMI tropospheric vertical column density as the satellite-derived NO_2 estimates. We chose effective cloud fraction (ECF) < 0.3, solar zenith angle (SZA) < 85 degree, and the primary summary quality flag indicating good data for data quality control. The AOD features were collected from the MODIS Terra and Aqua combined Multi-angle Implementation of Atmospheric Correction (MAIAC) Land Aerosol Optical Depth (AOD) gridded Level 2 1km product (https://lpdaac.usgs.gov/products/mcd19a2v061/) retrieved through Google Earth Engine platform. We used AOD over land retrieved in both MODIS blue band (0.47μm) and green band (0.55μm). For quality control, we only retrieved data with quality assessment [QA] band indicating the best quality. For both OMI and AOD variables,

we extracted the features directly at the EPA sites without a buffer due to their coarse spatial resolution.

4.2.1.3 Landsat-derived variables.

Besides satellite-derived aerosol estimates, we considered the use of other remote sensing products which were not originally designed for aerosol estimation. Specifically, we collected features derived from Landsat as the third type of model independent variables. Landsat is the longest continuous satellite program thus far (i.e., available as early as 1972). It has a 30 m resolution which is considered medium resolution and is free of charge for use. We generated Landsat yearly composite imagery using Landsat 7 and Landsat 8 imagery following the approach introduced in Demuzere et al[31]. A total of 12 bands were used, including 7 Landsat bands (i.e., blue, green, red, near infrared, shortwave infrared and thermal infrared) and 5 derived spectral indices (e.g., Normalized Difference Vegetation Index [NDVI], Biophysical Composition Index [BCI], Normalized Difference BAreness Index [NDBAI], and Normalized Difference Water Index [NDWI]). More details can be found in **Table 4-S1**. We collected the Landsat-derived features at the same multiple buffers as GSV features from 250m up to 2000m with an interval of 250m. Additionally, we included a buffer of 50m for the Landsat-derived features as an effort to capture local built information. We didn't include 50m for GSV-derived features mainly due to the limitation of GSV image availability. All satellite-derived feature collection was conducted through Google Earth Engine, a powerful cloud computing platform that provides various geospatial datasets at the petabyte-scale[32].

4.2.2 Modeling approach

We used different combinations of the street view and remote sensing derived features to develop and compare models, including (1) GSV-only, (2) GSV + aerosol estimates

(i.e., OMI or AOD), (3) Landsat-only, (4) Landsat + aerosol estimates, (5) GSV + Landsat + aerosol estimates (i.e., the full model). All models included the year of air pollutant measurements as the dummy variable. We developed random forest models, an ensemble machine learning model which is easy to train yet has demonstrated good model performance in many studies[25, 33-35]. We used a hybrid local-cloud modeling pipeline, i.e., local model fine-tuning and cloud model prediction, to accommodate the need for efficient model development and feasible large-scale model application. For main models, we use buffers up to 2000 m for both GSV features and Landsat-derived features. To explore the impact of buffers on model performance and application, we conducted a sensitivity analysis over the choice of Landsat-derived feature buffer sizes with potential buffer up to 10000 m.

4.2.3 Model evaluation

We assessed our models using both random 10-fold cross validation [CV] and spatial hold-out 10-fold cross validation. For random 10-fold CV, all air pollutant records were randomly divided into 10 subgroups. For spatial hold out CV, all air pollutant records were first geocoded based on 2000m spatial grids and then divided into 10 subgroups based on the geocodes. No records from the same spatial grids can be present in both training and testing sets to mitigate the impact of spatial autocorrelation. We reported 10-fold CV R^2, mean absolute error (MAE), and root mean squared error (RMSE) to compare model performance. We used the random 10-fold CV as our main model evaluation method. Unless specified as spatial hold-out CV, all 10-fold CV R^2 referred to the random 10-fold CV R^2 in default.

4.3 Results and discussion

4.3.1 Description of Model Inputs

We collected annual NO_2 and $PM_{2.5}$ concentrations from 483 and 1,069 unique EPA monitors across the continental U.S. during our study period. Due to the lack of GSV images in the vicinity of EPA monitors in certain areas or at certain times, monitors unable to assemble a complete set of modeling predictors were excluded for model development. **Figure 4-S1** shows the number of available and excluded EPA sites for modeling in each year. In average, 72.2% of $PM_{2.5}$ EPA monitors and 71.3% of NO_2 EPA monitors find GSV images within a 250m buffer (i.e., the minimum buffer for GSV features) during the same year or within the previous 4 years of the annual concentration measurement. **Figure 4-S2** shows the spatial distribution of the EPA monitors either used or excluded for model development. Excluded NO_2 monitors with insufficient GSV images were mainly located around the Rocky Mountains while excluded $PM_{2.5}$ monitors were more scattered across the country. **Figure 4-S3** showed the distribution of EPA monitors by urbanicity (i.e., urban/suburban/rural) for both air pollutants using 2019 for illustration. In general, the proportion of NO_2 sites in rural areas (27.3%) is relatively higher than that of $PM_{2.5}$ sites (21.7%). Due to the variation in GSV imagery availability, it's more challenging in collecting street view images for EPA sites in rural areas. For example, only 33.0% of $PM_{2.5}$ monitors in the rural areas successfully assembled a complete set for model training, while the percentage for monitors in urban and suburban areas are 94.5% and 80.2%, respectively. **Tables 4-S2** and **4-S3** show the descriptive summary of features within 250m of EPA sites by urbanicity. The trends of the annual concentrations from the modeled EPA stations showed that both pollutants in the U.S. continuously decreased during the study period in general (**Figure 4-S4**). The decreasing slope of $PM_{2.5}$ was slightly deeper than that of NO_2. The mean annual concentration dropped 37.3% for $PM_{2.5}$ and 31.5% for NO_2 from 2007 to 2019. Moreover, the maximum and minimum annual $PM_{2.5}$ concentrations

from individual sites also showed a generally continuous decreasing trend while the maximum and minimum annual NO_2 showed more fluctuations with slight surge after 2014.

4.3.2 Model Performance

Table 4-1 shows the results of random forest models using different combinations of street view derived and remote sensing derived features. As expected, the full model achieved the highest model performance for both NO_2 and $PM_{2.5}$. The best random 10-fold CV R^2 (MAE; RMSE) was 0.91 (1.13ppb; 2.58ppb) for NO_2 and 0.72 (0.94 $\mu g/m^3$; 1.62 $\mu g/m^3$) for $PM_{2.5}$. The results were comparable with or better than other reported national NO_2 models[24, 36-38], and consistent with or worse than other reported national $PM_{2.5}$ models[24, 36]. The model fits for $PM_{2.5}$ was generally lower than NO_2 for land use regression models, as $PM_{2.5}$ was both a primary and secondary air pollutant while NO_2 was a short-lived pollutant dominated by primary emission sources. For NO_2, all models achieved satisfying performance, especially models including GSV-derived features. It's noteworthy that using GSV features alone, the NO_2 models achieved almost the same model fits as the full model, indicating the ability of street view derived features in capturing the air pollutant variation. Models solely based on remote sensing derived features show relatively worse model performance. The worst NO_2 model was Landsat-only model, which achieved 0.82 for random 10-fold CV R^2, 1.69 ppb for MAE and 5.25 ppb for RMSE. Adding satellite-derived aerosol estimate helps with Landsat-based models and increased model fits by 0.04 from Landsat-only to Landsat+OMI model. The inclusion of OMI didn't increase random 10-fold CV R^2 for GSV-only NO_2 model but improved MAE and RMSE slightly. The potential explanation is that the street view features have captured sufficient variation of primary

air pollutant thus the inclusion of OMI aerosol estimate doesn't further enhance model performance.

Similar to NO_2 models, the GSV-based $PM_{2.5}$ models also outperformed the satellite-based $PM_{2.5}$ models. The GSV-only $PM_{2.5}$ model achieved 0.68 for random 10-fold CV R^2, 0.99 μg/m^3 for MAE and 1.84 μg/m^3 for RMSE. Adding AOD to the GSV-only model further boosted the model R^2 by 0.03. Unlike NO_2, the inclusion of satellite-derived aerosol estimates helped less for Landsat-based model but more for GSV-based model. In addition, the difference between GSV-based and satellite-based models was larger for $PM_{2.5}$ than that of NO_2. This finding indicated that GSV features are better at characterizing air pollution at local scale but relatively weak in providing regional information. The combination of GSV features and satellite-derived features have the potential to capture air pollution variation at both local and regional scale, thus achieved the best model performance. **Figure 4-1-A** shows the model predictions versus observations of the full models for both pollutants (see results for other models in **Figure 4-S5**). As expected, the trend lines from NO_2 models are closer to 1:1 line than the $PM_{2.5}$ models. For NO_2 models, the slope of the full model and GSV+OMI models are closest to 1 while for $PM_{2.5}$ models, the GSV-only and GSV+AOD models even achieved a higher slope than the full $PM_{2.5}$ model.

Spatial hold-out cross validation is a more stringent model evaluation strategy which assesses how model perform in regions far away from the training monitors, thus, it's expected that the model fits by spatial hold-out CV decrease compared against random 10-fold CV. **Table 4-S4** shows the results using the spatial hold-out cross validation and the difference between the two assessment strategies. The best spatial hold-out 10-fold CV R^2 (MAE; RMSE) was 0.57 (2.70ppb; 12.60ppb) for NO_2 and 0.41 (1.40 μg/m^3; 3.43 μg/m^3) for $PM_{2.5}$. The deterioration of model performance was most severe

for GSV-based models, especially the GSV-only models. Models using Landsat-derived features and aerosol estimates shown relatively lowest decrease, likely because the remote sensing derived features capture more regional information thus perform better in regions far away compared to street view features.

Table 4-1. Summary of model performance for models with different types of independent variables. Model evaluation metrics included random 10-fold R^2, mean absolute error (MAE), and root mean squared error (RMSE).

Air pollutant	Model name	Random 10-fold CV R^2	MAE	RMSE
NO_2	GSV-only	0.90	1.20	3.02
	Landsat-only	0.82	1.69	5.25
	GSV+OMI	0.90	1.16	2.79
	Landsat+OMI	0.86	1.49	4.07
	GSV+Landsat+OMI	0.91	1.13	2.58
$PM_{2.5}$	GSV-only	0.68	0.99	1.84
	Landsat-only	0.57	1.17	2.51
	GSV+AOD	0.71	0.96	1.70
	Landsat+AOD	0.59	1.16	2.40
	GSV+Landsat+AOD	0.72	0.94	1.62

We also investigated how different image-based features contributed to air pollution models. **Figure 4-1-B** and **C** show the model feature importance scores for the NO_2 and $PM_{2.5}$ full models (other two types of models were shown in **Figures 4-S6** and **4-S7** using GSV-only and GSV + aerosol estimates models for illustration). For both air pollutants, GSV-derived features contributed most to the full model, followed by the Landsat-derived features. GSV features contributed 70% for NO_2 and 67% for $PM_{2.5}$, respectively. The contribution of Landsat-derived features to NO_2 and $PM_{2.5}$ was 22% and 30%, respectively. The image-based features were further divided into different subgroups to identify major drivers of the model. The main contributors for the NO_2 model were transport and built environment related feature groups. For example, GSV-

transport vehicles and GSV-transport network showed the highest (i.e., 0.19) and the third highest (i.e., 0.14) feature importance scores for the NO_2 full model. GSV-built environment and Landsat-bci (an index representing impervious surface abundance) were also among the top 4 contributors. When further breaking down to more individualized features, the top 3 contributors for the NO_2 pollution variation were GSV-car, GSV-building, and Landsat-bci. It's noteworthy that traffic-related variables appeared in the majority of the top important individual features (e.g., car, sidewalk, truck, bridge, streetlight and traffic light) for the NO_2 model (**Figure 4-1-C**), which is consistent with the fact that NO_2 is a traffic-related air pollutant [TRAP]. The ranking of features by other types of models may differ slightly, however, the type of dominant feature groups remains unchanged.

While transport and built environment related feature groups were among the top important features to the $PM_{2.5}$ model, several natural environment features stood out and played a role in explaining the $PM_{2.5}$ variations. GSV-natural and GSV-vegetation are the second and third important feature group for $PM_{2.5}$ full model, slightly lower than the impact from GSV-transport network. Unlike NO_2 where the dummy variable showed little influence, the year variable in the $PM_{2.5}$ model is the most important single variable. This might be explained by the stronger decreasing trend of $PM_{2.5}$ pollution over years. It's also notable that the difference of feature importance scores among individual features or feature groups are smaller in the $PM_{2.5}$ model compared to the NO_2 model.

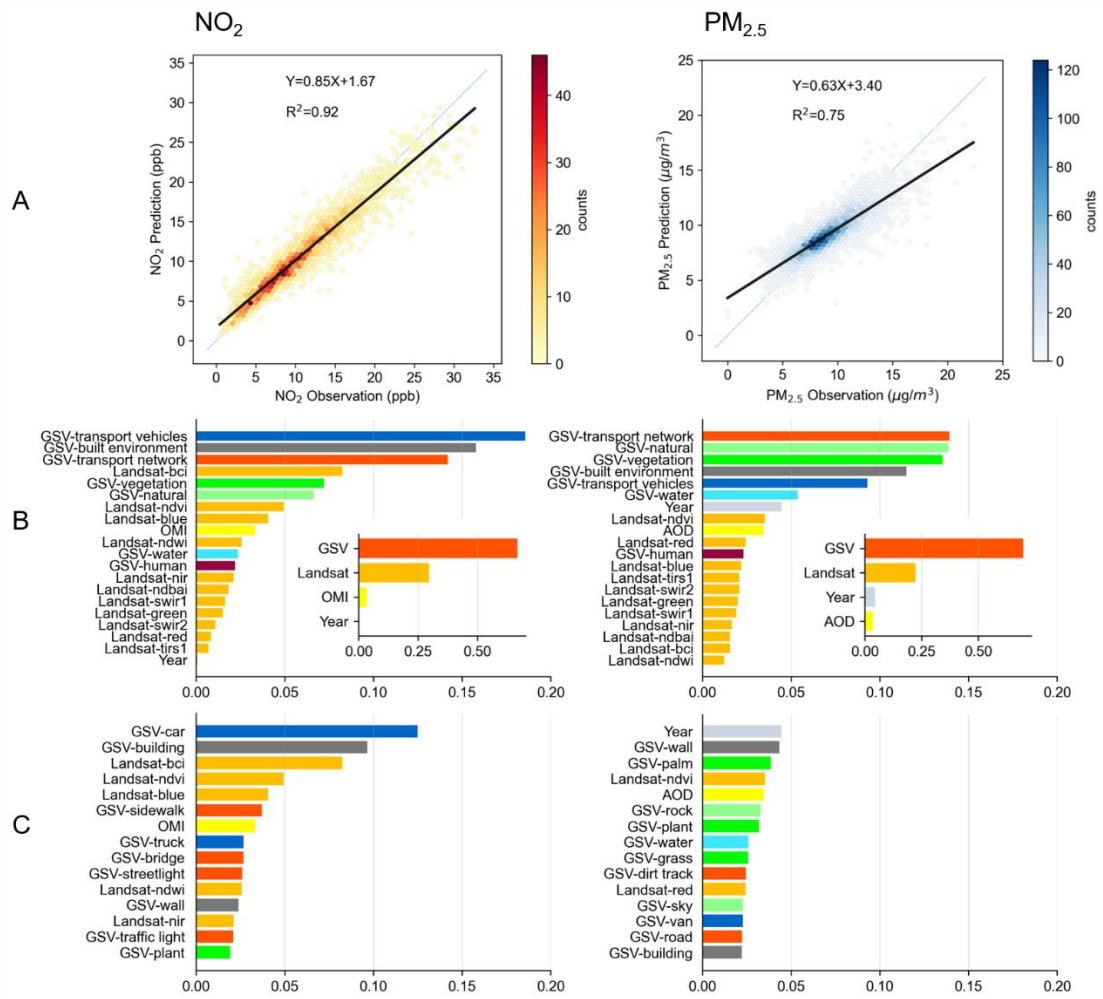

Figure 4-1. Prediction versus observation plots and ranking of feature importance scores for the NO_2 and $PM_{2.5}$ full models. The left (right) panel shows results for NO_2 ($PM_{2.5}$) model. (A) Hexagonal binned plots of model predictions versus observations. The color reflects the number of scatters within each hexagon. (B) The ranking of feature importance scores based on aerosol estimates, aggregated GSV and Landsat feature groups (i.e., aggregation level 2). The inner plots show the ranking based on feature aggregation level 1. (C) Top 15 important features based on feature scores of aerosol estimates, individual GSV and Landsat features (aggregation level 3). The details of the aggregation levels are listed in **Table 4-S1**. Results of other models are shown in **Figures 4-S5** to **4-S7.**

4.3.3 Model prediction

Figure 4-2 shows the prediction surface for both air pollutants at 100m spatial resolution using 2019 for illustration. Since GSV images are predominantly distributed in urban areas and largely limited in rural areas, models involving GSV features are unable to estimate air pollutant concentrations at locations without GSV images nearby (i.e., 500m due to the setting of our main models). This impact may be limited depending on the application. For example, 83% of the population in the U.S. lives in the urban area, highly overlapping with the spatial distribution of GSV imagery[39]. **Figures 4-3** and **4-4** show the predicted annual NO_2 and $PM_{2.5}$ concentrations for sample urbanized areas. For NO_2, the spatial patterns of the major pollutant areas are similar among models. Compared to Landsat-based models, models involving GSV-derived features show significant intra-urban variation in the prediction surface. GSV-only models not only identified more potential local hotspots but also achieved lower model estimates in cleaner areas. We aim to verify the authenticity of this pattern using low-cost sensor networks or mobile monitoring data in our future study. For $PM_{2.5}$, the Landsat-only models predicted higher concentrations in some regions than GSV-based models. The prediction surface by the Landsat-AOD model seem different from the Landsat-only model and hotspots seem to be random, requiring further examination.

To further explore how different types of image-based features impact the model prediction capability, three statistics (i.e., minimum, mean, maximum) of air pollutant estimates at various census unit levels by different models were summarized. **Figure 4-5** shows the results of statistical distribution at urbanized area level using 2019 for illustration while **Figures 4-S8** and **4-S9** show the results at census tract and block group level. It's noticeable that models involving GSV features estimated air pollution with a larger range. For example, GSV-based NO_2 models obtained both the lowest and

highest air pollutant estimates. The median of the maximum (minimum) NO_2 concentrations in an urbanized area in 2019 is 10.0 ppb (3.7 ppb), 9.6 ppb (3.5 ppb), and 10.6 ppb (5.1) by GSV-only, GSV+OMI, and the full model, respectively. Correspondingly, the value is 8.6 ppb (6.7 ppb) and 8.7 ppb (7.1 ppb) for Landsat-only and Landsat+OMI model. This result indicated that street view feature may have a better capability to capture fine-scale NO_2 air pollution variability while remote sensing based features seem to smooth out the variation. Landsat-based models generated relatively higher NO_2 estimates in general. The median of the mean NO_2 concentrations in an urbanized area is 6.9 ppb, 6.7 ppb, 7.3 ppb, 7.8 ppb and 8.0 ppb for GSV-only, GSV+OMI, Landsat-only, Landsat+OMI, and the full model, respectively. For $PM_{2.5}$, the difference among models is less distinctive and doesn't show a similar pattern as NO_2. This might be explained as street view features perform less good while remote sensing features perform relative better for pollutants with secondary emission sources from distant regions. For statistics at census tract and census block group level, the same distribution patterns were shown in **Figures 4-S8** and **4-S9**.

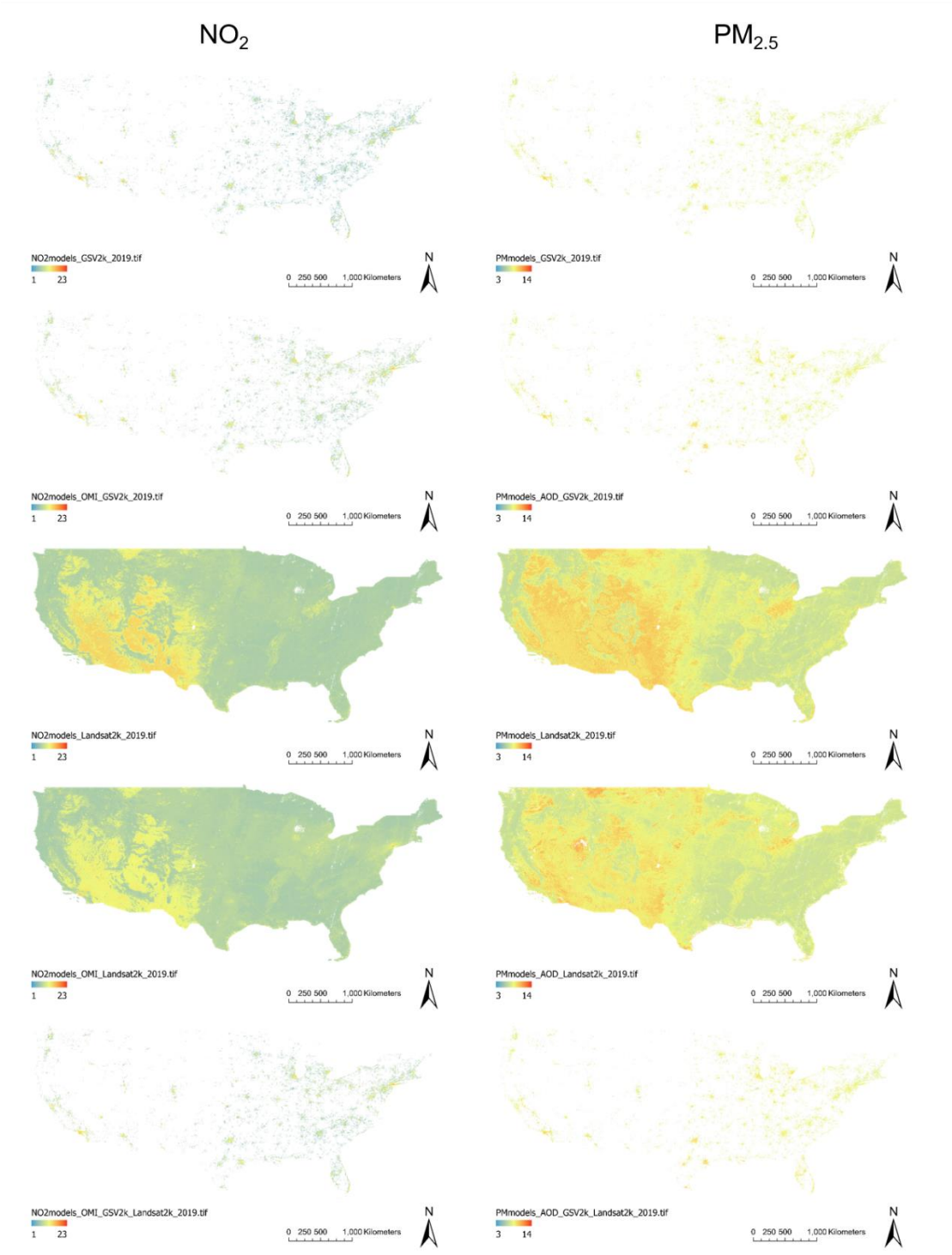

Figure 4-2. Predicted NO_2 and $PM_{2.5}$ air pollution by different image-based models at 100m spatial resolution using 2019 for illustration. The left panel shows results for NO_2 and the right panel shows results for $PM_{2.5}$.

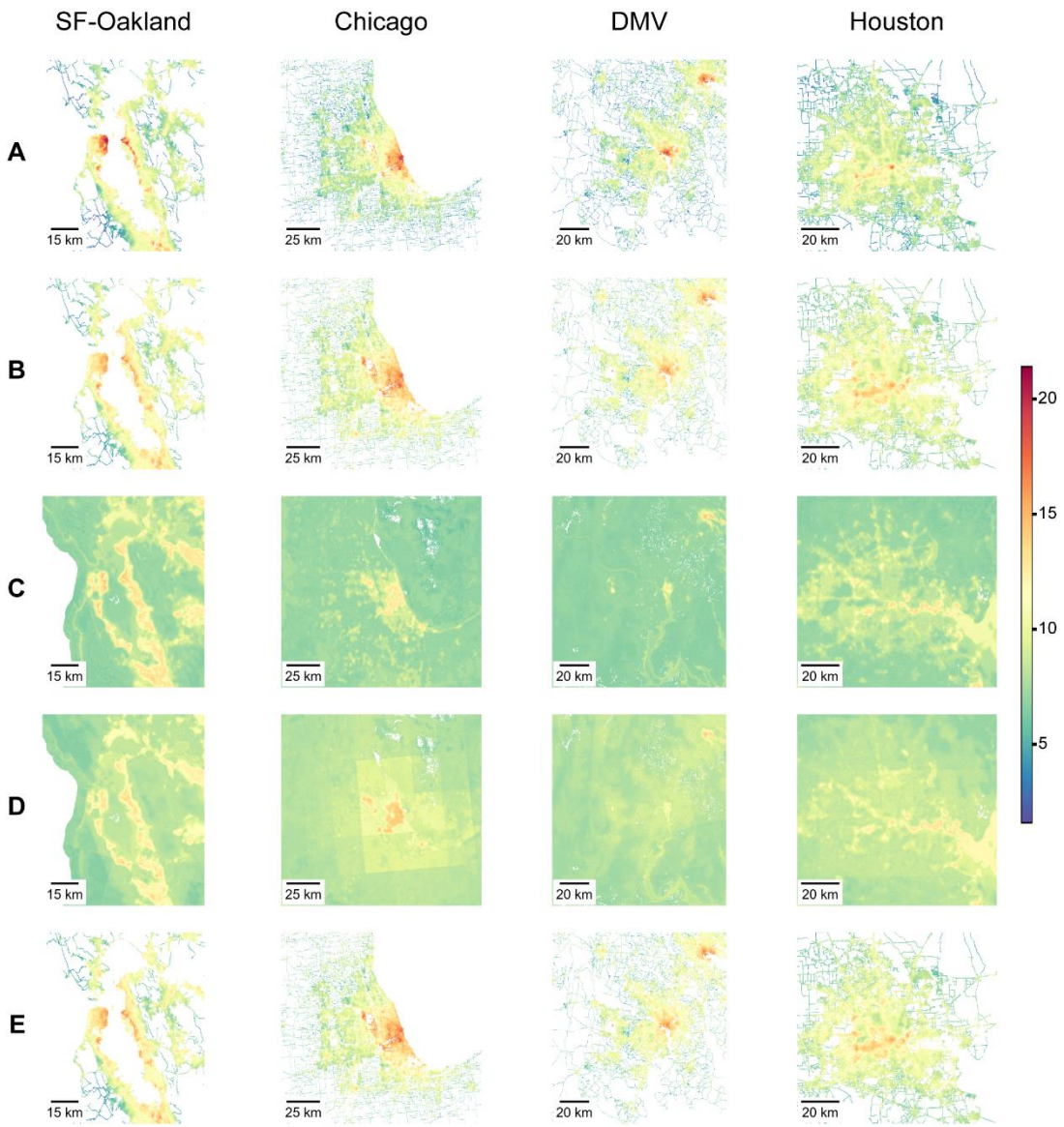

Figure 4-3. Spatial patterns of predicted annual NO_2 concentrations for four example cities (i.e., San Francisco-Oakland [SF-Oakland], Chicago, Washington, DC-Maryland-Virginia [DMV], and Houston) in 2019 by different image-based models. (A) to (F) represent GSV-only, GSV+OMI, Landsat-only, Landsat+OMI, and the full model, respectively.

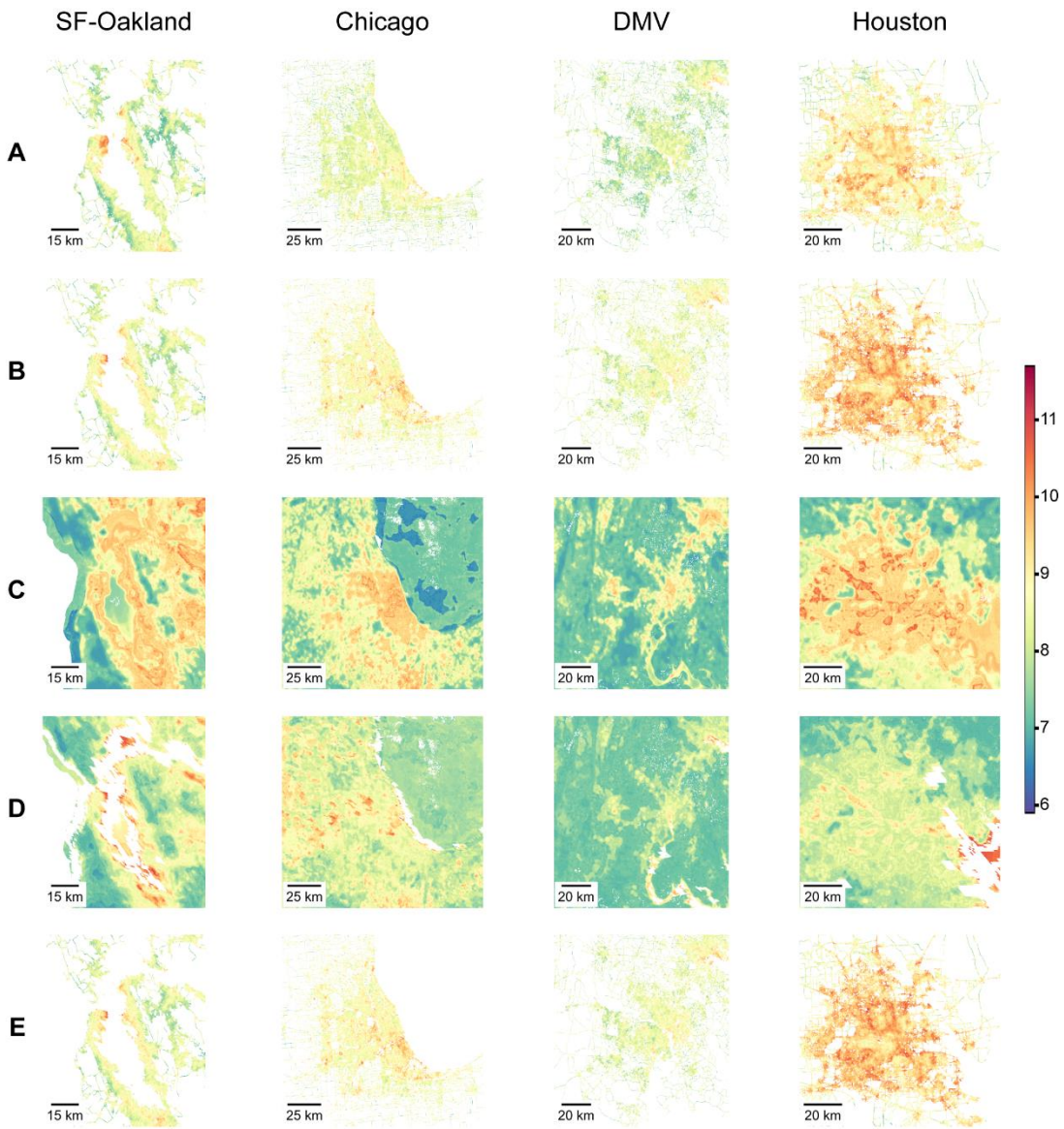

Figure 4-4. Spatial patterns of predicted annual $PM_{2.5}$ concentrations for four example cities (i.e., SF-Oakland, Chicago, Washington, DMV, and Houston) in 2019 by different image-based models. (A) to (F) represent GSV-only, GSV+AOD, Landsat-only, Landsat+AOD, and the full model, respectively.

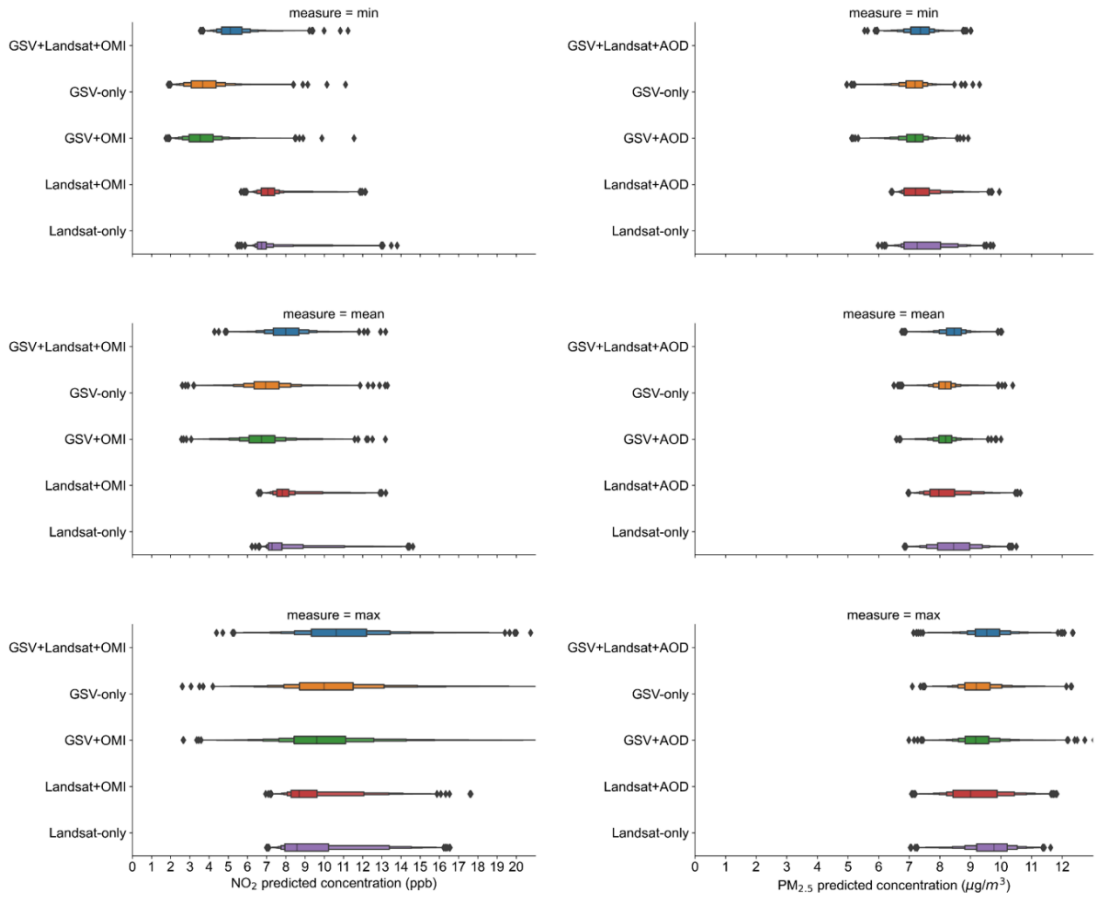

Figure 4-5. Statistical distribution of predicted NO_2 and $PM_{2.5}$ annual concentrations at urbanized area level by different image-based models using 2019 for illustration. The left panel shows results for NO_2, and the right panel shows results for $PM_{2.5}$. The first, second, and third rows represent the distribution of minimum, mean, and maximum annual concentrations, respectively.

4.4 Conclusion

Our work demonstrates the feasibility of utilizing street view imagery and remote sensing imagery to generate large-scale air pollution at high resolution. Our best models achieve good model performance. The random 10-fold cross validation R^2 was 0.91 and 0.72 for the NO_2 and $PM_{2.5}$ full model, respectively. Street view imagery is especially promising for air pollution modeling and projection as the street view derived features

alone achieved good model fits almost as high as the full model. In addition, more intra-urban air pollution variation is shown by the GSV-based models.

Our work has limitations and could be further improved in several ways. First, the local hotspots identified by the GSV-based models should be compared against ground truth to verify the authenticity of hotspots. Traditional EPA regulatory surface network is sparsely distributed thus is not dense enough for the verification purposes. In recent years, the emergence of low-cost, portable air pollution sensors is changing the paradigm of air pollution monitoring,[40] and may be used for verifying local hotspots with proper data quality assurance and quality control. Another potential source is mobile monitoring campaigns which deploy field-grade sensors on mobile platforms. mobile monitoring campaigns generally have higher data quality than low-cost sensor monitoring network. Second, the spatial resolution of the Landsat imagery is coarse and thus shows weaker ability to capture spatial variation of air pollution. Other remote sensing imagery with higher spatial resolution is available and can be investigated for future study. Third, models with higher temporal resolution can be developed to fully leverage the potential of remote sensing products, including monthly, daily, or even hourly air pollutant models. Street view imagery may still be used to provide additional street level information to enhance the remote sensing models. The impact of buffer size of street view derived features could be explored aiming at a balance between the model fits and spatial coverage for model prediction.

4.5 Supporting Information

Table 4-S1. Feature aggregation levels for feature importance scores ranking.

Features	Level 3 name	Level 2 name	Level 1 name
building	GSV-building	GSV-built environment	GSV
canopy	GSV-canopy		
house	GSV-house		
hovel	GSV-hovel		
pier	GSV-pier		
signboard	GSV-signboard		
skyscraper	GSV-skyscraper		
swimming pool	GSV-swimming pool		
wall	GSV-wall		
bicycle	GSV-bicycle	GSV-human	
person	GSV-person		
animal	GSV-animal	GSV-natural	
earth	GSV-earth		
hill	GSV-hill		
land	GSV-land		
mountain	GSV-mountain		
rock	GSV-rock		
sand	GSV-sand		
sky	GSV-sky		
bridge	GSV-bridge	GSV-transport network	
dirt track	GSV-dirt track		
path	GSV-path		
road	GSV-road		
runway	GSV-runway		
sidewalk	GSV-sidewalk		
streetlight	GSV-streetlight		
traffic light	GSV-traffic light		
airplane	GSV-airplane	GSV-transport vehicles	
boat	GSV-boat		
bus	GSV-bus		
car	GSV-car		
minibike	GSV-minibike		
ship	GSV-ship		
truck	GSV-truck		
van	GSV-van		
field	GSV-field	GSV-vegetation	
flower	GSV-flower		
grass	GSV-grass		
palm	GSV-palm		
plant	GSV-plant		
tree	GSV-tree		

fountain lake river sea water waterfall	GSV-fountain GSV-lake GSV-river GSV-sea GSV-water GSV-waterfall	GSV-water	
ndvi_min ndvi_max	Landsat-ndvi	Landsat-ndvi	
ndwi	Landsat-ndwi	Landsat-ndwi	
bci	Landsat-bci	Landsat-bci	
blue	Landsat-blue	Landsat-blue	
green	Landsat-green	Landsat-green	
ndbai	Landsat-ndbai	Landsat-ndbai	
nir	Landsat-nir	Landsat-nir	
red	Landsat-red	Landsat-red	
swir1	Landsat-swir1	Landsat-swir1	
swir2	Landsat-swir2	Landsat-swir2	
tirs1	Landsat-tirs1	Landsat-tirs1	
AOD_047 AOD_055	AOD	AOD	AOD
OMI	OMI	OMI	OMI
Year	Year	Year	Year

Table 4-S2. Descriptive statistics of GSV and Landsat-derived features within 250 m of NO_2 EPA monitors stratified by land use types. GSV features are applied with a scale of 10000.

Land use type	Feature	Min.	Max.	Mean	Std.
Urban	GSV-built environment	9.60	2828.83	650.42	475.95
	GSV-human	0.00	21.41	1.98	2.70
	GSV-natural	2286.94	5846.59	4076.49	801.98
	GSV-transport network	1998.68	3766.21	3046.07	368.60
	GSV-transport vehicles	1.51	760.65	196.76	145.65
	GSV-vegetation	141.55	4497.32	1877.17	949.47
	GSV-water	0.00	71.76	3.51	10.32
	Landsat-bci	-1375.41	365.32	-33.68	183.73
	Landsat-blue	-20.14	167.43	87.79	28.37
	Landsat-green	23.34	210.86	120.91	31.79
	Landsat-ndbai	-649.03	-447.21	-522.36	31.02
	Landsat-ndvi	89.43	1387.65	716.19	294.26
	Landsat-ndwi	-619.76	-24.75	-315.39	117.73
	Landsat-nir	73.03	391.01	265.42	46.37
	Landsat-red	20.56	245.39	127.77	40.70
	Landsat-swir1	55.43	326.94	220.33	40.56
	Landsat-swir2	41.40	269.12	159.88	39.16
	Landsat-tirs1	298242.19	329874.70	314778.84	5815.78
Suburban	GSV-built environment	5.67	2459.78	439.05	343.83
	GSV-human	0.00	14.05	1.19	1.78
	GSV-natural	2069.00	6182.15	4394.27	793.76
	GSV-transport network	1253.67	3818.58	2981.99	497.93
	GSV-transport vehicles	0.00	714.55	125.59	118.61
	GSV-vegetation	436.00	6224.01	1939.08	1106.00
	GSV-water	0.00	65.17	3.07	9.14
	Landsat-bci	-355.34	285.22	-62.09	167.41
	Landsat-blue	26.14	175.46	81.63	28.12
	Landsat-green	52.07	219.57	118.10	33.80
	Landsat-ndbai	-590.21	-418.86	-513.04	31.52
	Landsat-ndvi	194.53	1529.98	773.44	342.28
	Landsat-ndwi	-677.13	-1.32	-344.29	138.38
	Landsat-nir	153.03	408.79	280.51	47.68
	Landsat-red	43.14	246.28	125.36	45.07
	Landsat-swir1	132.44	332.70	229.19	43.03
	Landsat-swir2	74.53	270.18	164.29	44.03
	Landsat-tirs1	299587.67	327103.79	313959.32	5835.45
Rural	GSV-built environment	1.25	447.57	111.35	113.78
	GSV-human	0.00	6.40	0.57	1.37
	GSV-natural	2809.11	6812.29	4905.22	882.51
	GSV-transport network	1409.37	3695.23	2545.31	673.25
	GSV-transport vehicles	0.00	268.30	27.74	50.29

GSV-vegetation	251.99	5734.83	2347.88	1344.61
GSV-water	0.00	356.46	16.75	62.89
Landsat-bci	-371.90	185.47	-169.74	167.58
Landsat-blue	16.06	132.08	64.31	30.43
Landsat-green	41.20	187.16	103.36	38.72
Landsat-ndbai	-595.69	-382.06	-494.36	42.51
Landsat-ndvi	251.03	1523.11	948.35	358.42
Landsat-ndwi	-637.83	-127.36	-418.75	131.42
Landsat-nir	169.15	384.81	297.14	47.99
Landsat-red	28.56	253.87	110.93	56.42
Landsat-swir1	121.09	416.88	247.12	64.65
Landsat-swir2	58.74	348.54	167.36	70.13
Landsat-tirs1	299015.13	325483.55	311980.16	6803.73

Table 4-S3. Descriptive statistics of GSV and Landsat-derived features within 250 m of $PM_{2.5}$ EPA monitors stratified by land use types. GSV features are applied with a scale of 10000.

Land use type	Feature	Min.	Max.	Mean	Std.
Urban	GSV-built environment	4.33	3686.86	617.17	483.79
	GSV-human	0.00	62.03	2.03	4.04
	GSV-natural	1778.21	6394.66	3975.23	753.10
	GSV-transport network	1641.51	4010.38	3006.75	415.22
	GSV-transport vehicles	0.20	813.18	173.60	134.49
	GSV-vegetation	141.55	5654.00	2101.36	1017.43
	GSV-water	0.00	72.09	2.87	8.92
	Landsat-bci	-1375.66	365.34	-80.90	169.95
	Landsat-blue	-20.14	183.43	79.45	27.07
	Landsat-green	23.34	231.97	112.33	30.28
	Landsat-ndbai	-656.54	-443.33	-522.45	28.34
	Landsat-ndvi	89.42	1493.57	806.57	299.46
	Landsat-ndwi	-635.63	-24.41	-351.53	122.60
	Landsat-nir	73.01	373.49	271.72	44.57
	Landsat-red	20.55	270.23	116.56	39.10
	Landsat-swir1	55.42	326.92	216.64	36.58
	Landsat-swir2	41.39	288.67	152.02	35.17
	Landsat-tirs1	298241.95	331181.25	313464.12	5675.47
Suburban	GSV-built environment	7.59	2711.55	487.58	416.03
	GSV-human	0.00	14.05	1.19	1.66
	GSV-natural	827.38	6102.75	4249.20	737.45
	GSV-transport network	1455.77	3892.14	2956.67	465.62
	GSV-transport vehicles	0.45	852.18	133.82	128.48
	GSV-vegetation	228.67	6442.39	2080.31	1069.79
	GSV-water	0.00	221.61	3.09	18.00
	Landsat-bci	-355.33	285.21	-110.01	147.49
	Landsat-blue	30.10	175.45	73.95	24.59
	Landsat-green	56.22	219.56	110.30	29.33
	Landsat-ndbai	-586.45	-433.37	-514.27	26.91
	Landsat-ndvi	206.34	1555.31	868.30	315.20
	Landsat-ndwi	-677.11	41.10	-379.06	125.65
	Landsat-nir	142.47	408.79	288.53	44.90
	Landsat-red	39.29	246.27	114.44	40.04
	Landsat-swir1	133.89	324.62	223.88	36.51
	Landsat-swir2	73.70	270.19	155.56	39.13
	Landsat-tirs1	297899.98	331782.21	312644.82	5876.71
Rural	GSV-built environment	0.00	447.57	128.66	119.51
	GSV-human	0.00	16.00	0.65	2.22
	GSV-natural	1729.00	9366.00	4607.21	1206.08
	GSV-transport network	288.00	3695.23	2408.42	706.01
	GSV-transport vehicles	0.00	268.30	31.40	49.87

GSV-vegetation	276.10	6569.43	2786.78	1482.38
GSV-water	0.00	11.80	1.53	2.93
Landsat-bci	-391.36	164.36	-236.99	123.77
Landsat-blue	18.55	115.66	53.31	22.04
Landsat-green	45.49	186.08	89.52	28.94
Landsat-ndbai	-561.58	-406.24	-499.13	35.16
Landsat-ndvi	251.03	1698.29	1129.65	323.80
Landsat-ndwi	-757.60	-251.33	-482.55	118.92
Landsat-nir	169.16	457.90	323.20	50.39
Landsat-red	28.32	253.88	89.00	45.87
Landsat-swir1	152.77	381.87	229.95	52.77
Landsat-swir2	63.62	348.54	143.49	56.41
Landsat-tirs1	296371.65	327090.87	308207.16	6476.05

Table 4-S4. Summary of model performance for models with different types of independent variables using spatial hold-out cross validation. Model evaluation metrics included spatial hold-out 10-fold R^2, mean absolute error (MAE), and root mean squared error (RMSE). The difference of model fits by random and spatial hold-out CV was also shown in the table.

Air pollutant	Model name	Spatial Hold-out CV			Difference		
		R^2	MAE	RMSE	R^2	MAE	RMSE
NO_2	GSV-only	0.49	2.92	14.90	0.41	-1.72	-11.88
	Landsat-only	0.47	3.05	15.50	0.35	-1.36	-10.25
	GSV+OMI	0.54	2.76	13.50	0.36	-1.60	-10.71
	Landsat+OMI	0.59	2.68	12.10	0.27	-1.19	-8.03
	GSV+Landsat+OMI	0.57	2.70	12.60	0.34	-1.57	-10.02
$PM_{2.5}$	GSV-only	0.34	1.46	3.82	0.34	-0.47	-1.98
	Landsat-only	0.27	1.52	4.23	0.30	-0.35	-1.72
	GSV+AOD	0.38	1.43	3.60	0.33	-0.47	-1.90
	Landsat+AOD	0.33	1.47	3.88	0.26	-0.31	-1.48
	GSV+Landsat+AOD	0.41	1.40	3.43	0.31	-0.46	-1.81

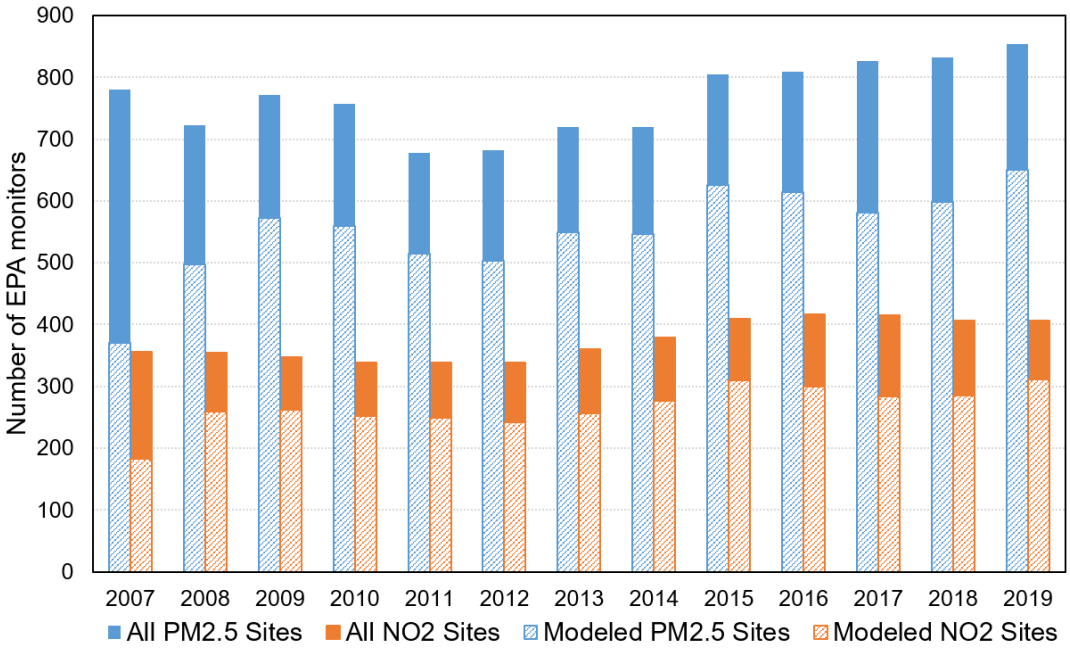

Figure 4-S1. The number of $PM_{2.5}$ and NO_2 EPA monitors used or excluded for model development each year from 2007 to 2019. Bars with strips indicate monitors used for modeling while solid bars indicate monitors excluded for modeling. The full bars are all available EPA monitors.

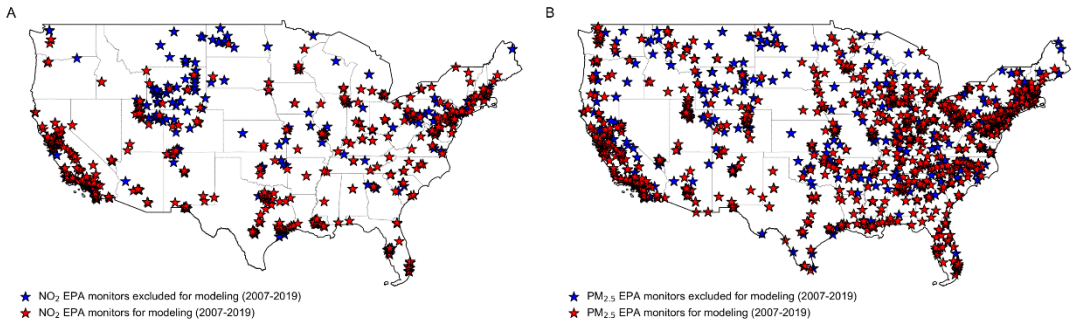

Figure 4-S2. Spatial distribution of NO_2 and $PM_{2.5}$ EPA monitors during 2007-2019. Red stars represent EPA monitors used for model development while blue stars represent EPA monitors excluded for modeling. Monitors without GSV images within 250m were excluded for modeling due to an incomplete set of model independent features.

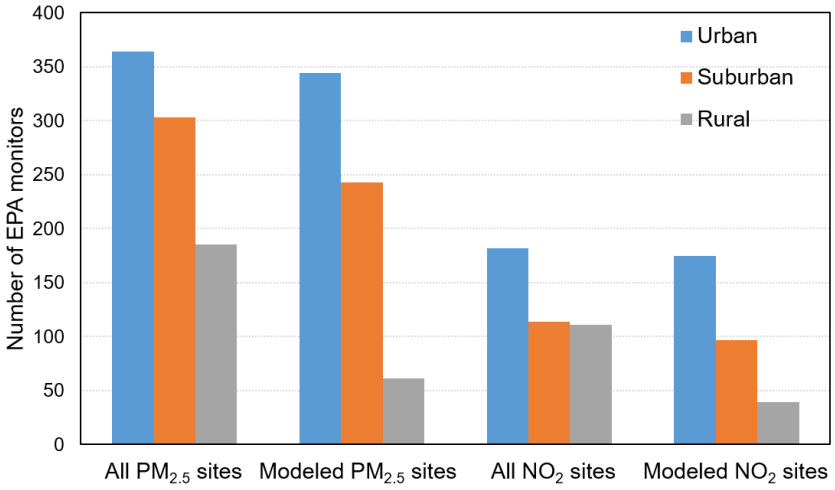

Figure 4-S3. Distribution of NO_2 and $PM_{2.5}$ EPA monitors by urban, suburban, and rural using 2019 for illustration. The urbanicity classification is based on the locale classification provided by the National Center for Education Statistics.

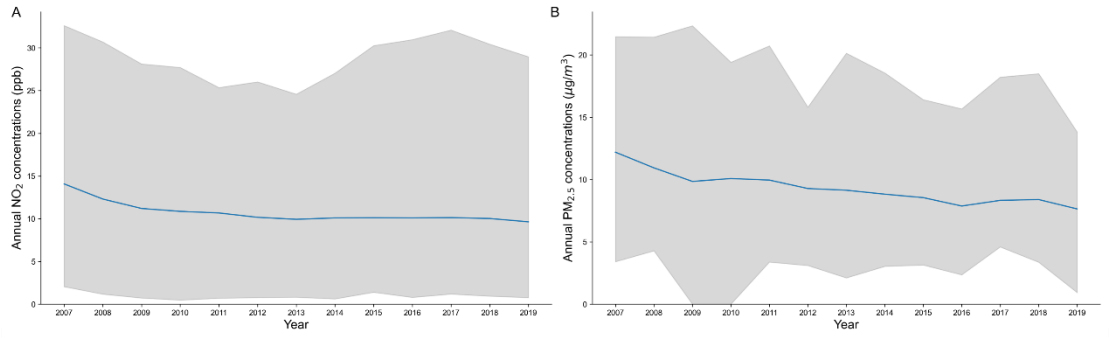

Figure 4-S4. The decreasing trends of annual NO_2 and $PM_{2.5}$ concentrations during 2007-2019 from the EPA regulatory network. The blue line is the mean of annual concentrations from all modeled EPA monitors. The grey area represents the minimum and maximum annual concentrations for each year.

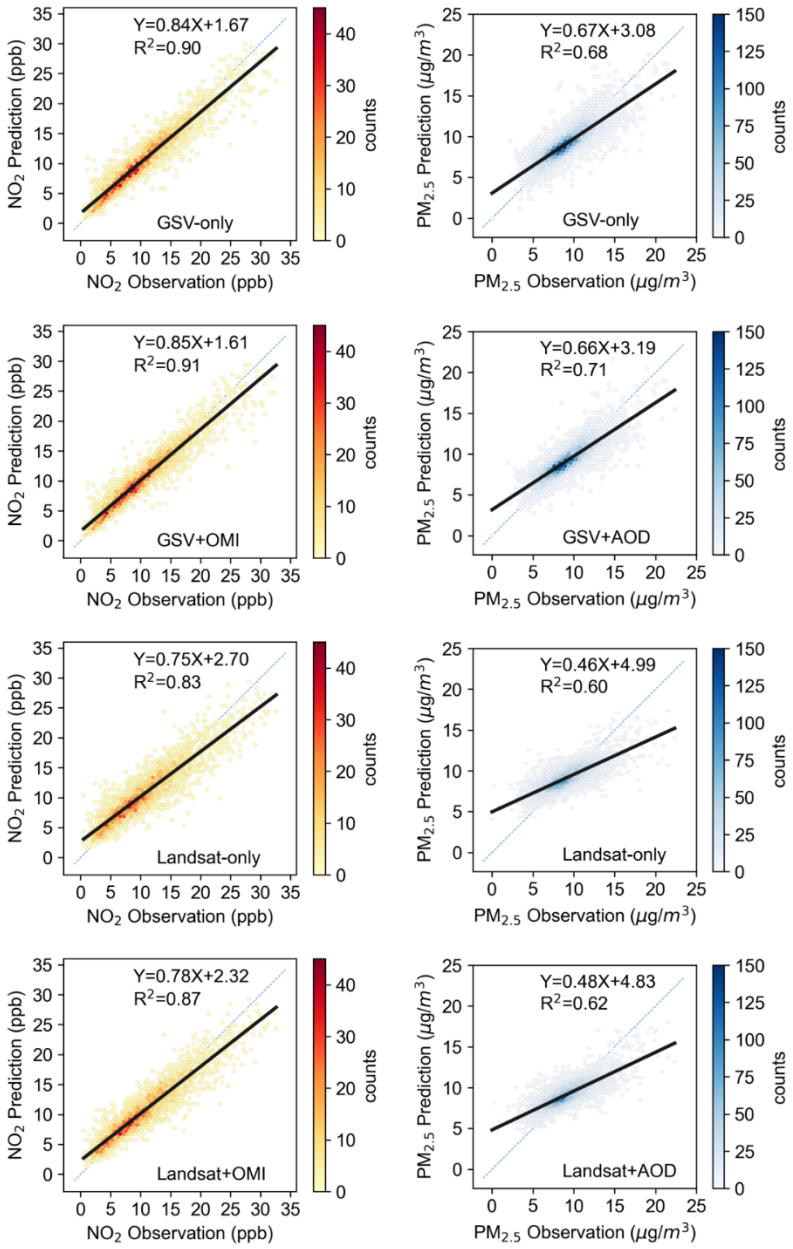

Figure 4-S5. Number of available NO_2-GSV pairs and GSV images in different years using scenario 1 and 5 for illustration. The left axis represents the number of GSV images found within 500m of each NO_2 monitoring site in each year while the color of the stripes represents its frequency. Darker colors indicate more EPA sites with the same amount of GSV images. The right axis represents the number of NO_2-GSV pairs for model development in each year.

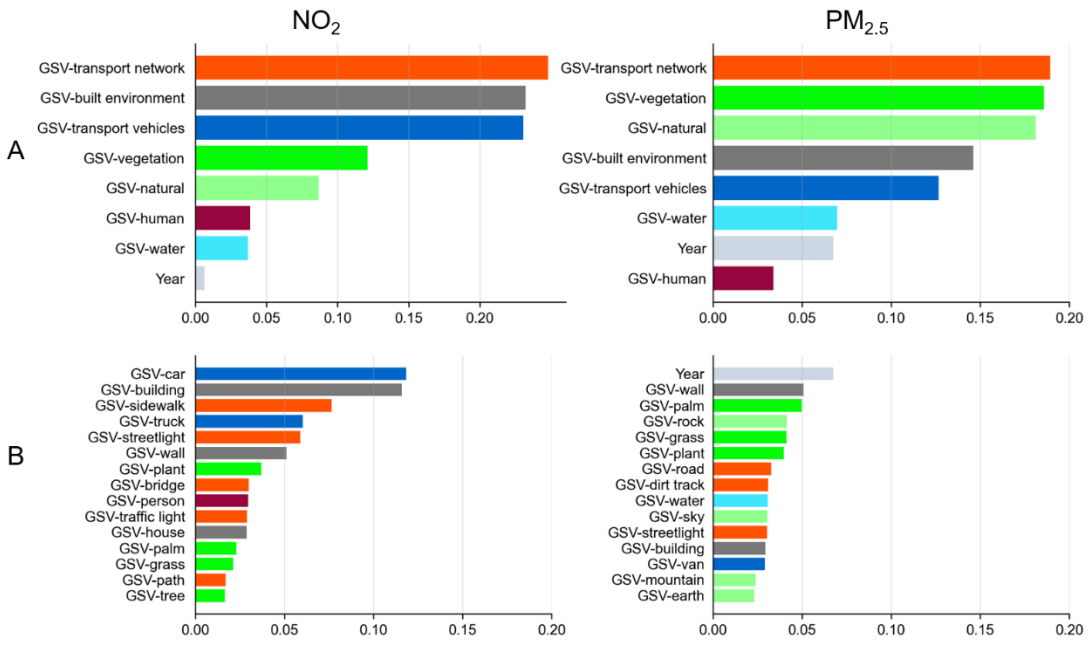

Figure 4-S6. Ranking of feature importance scores for the NO_2 and $PM_{2.5}$ GSV-only models. (A) The ranking of feature importance scores based on aerosol estimates, aggregated GSV and Landsat feature groups (i.e., aggregation level 2). (C) Top 15 important features based on feature scores of aerosol estimates, individual GSV and Landsat features (aggregation level 3). The details of the aggregation levels are listed in **Table 4-S1**.

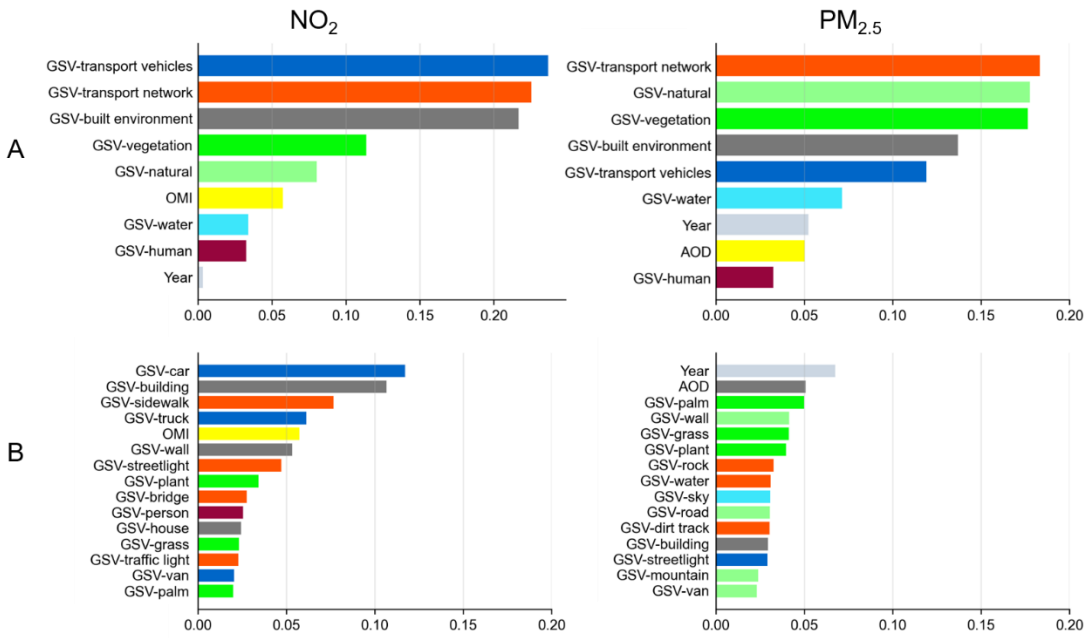

Figure 4-S7. Ranking of feature importance scores for the NO_2 GSV+OMI and $PM_{2.5}$ GSV+AOD models. (A) The ranking of feature importance scores based on aerosol estimates, aggregated GSV and Landsat feature groups (i.e., aggregation level 2). (C) Top 15 important features based on feature scores of aerosol estimates, individual GSV and Landsat features (aggregation level 3). The details of the aggregation levels are listed in **Table 4-S1**.

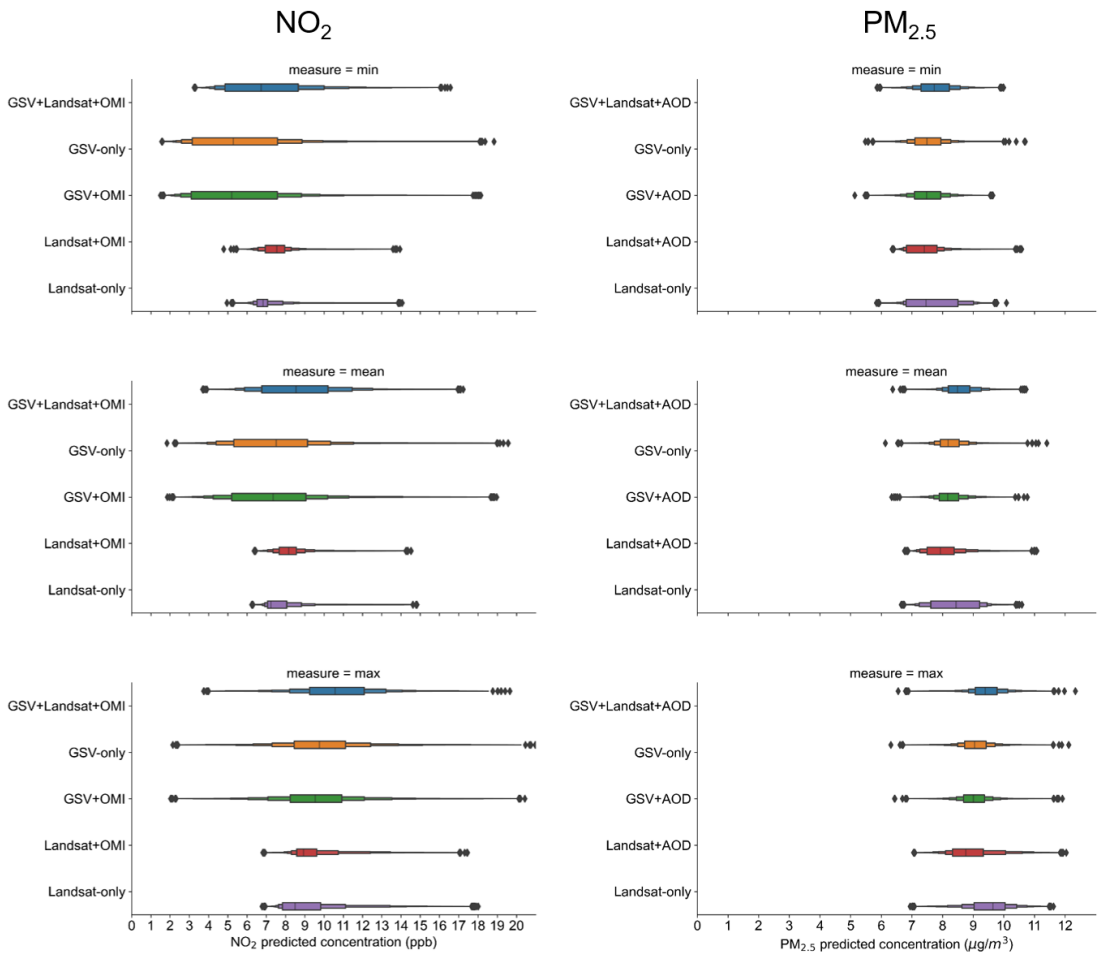

Figure 4-S8. Statistical distribution of predicted NO_2 and $PM_{2.5}$ annual concentrations at census tract level by different image-based models using 2019 for illustration. The left panel shows results for NO_2, and the right panel shows results for $PM_{2.5}$. The first, second, and third rows represent the distribution of minimum, mean, and maximum annual concentrations, respectively.

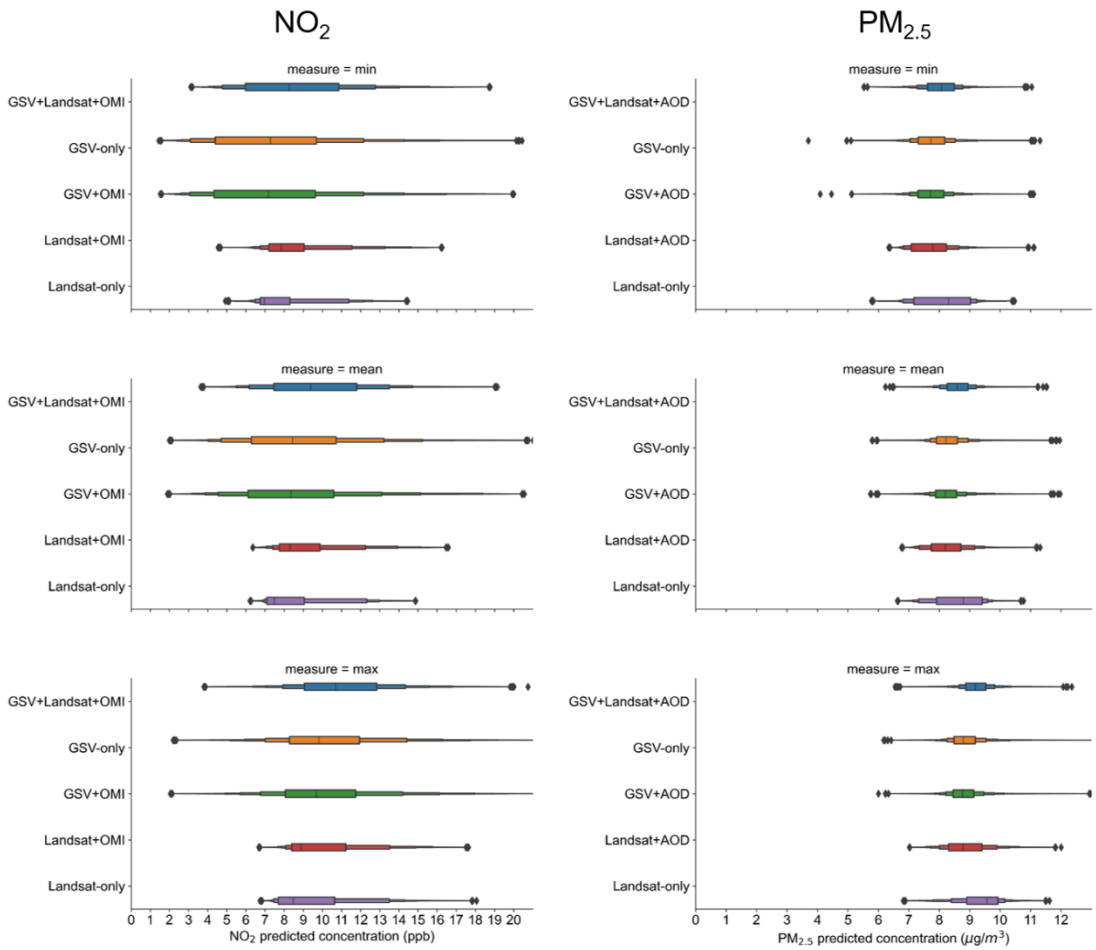

Figure 4-S9. Statistical distribution of predicted NO_2 and $PM_{2.5}$ annual concentrations at census block group level by different image-based models using 2019 for illustration. The left panel shows results for NO_2, and the right panel shows results for $PM_{2.5}$. The first, second, and third rows represent the distribution of minimum, mean, and maximum annual concentrations, respectively.

4.6 References

1. Burnett, R.; Chen, H.; Szyszkowicz, M.; Fann, N.; Hubbell, B.; Pope III, C. A.; Apte, J. S.; Brauer, M.; Cohen, A.; Weichenthal, S. Global estimates of mortality associated with long-term exposure to outdoor fine particulate matter. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115* (38), 9592-9597.

2. Pope, C. A.; Ezzati, M.; Dockery, D. W. Fine-Particulate Air Pollution and Life Expectancy in the United States. *N. Engl. J. Med.* **2009**, *360* (4), 376-386, Article. DOI: 10.1056/NEJMsa0805646.

3. Manisalidis, I.; Stavropoulou, E.; Stavropoulos, A.; Bezirtzoglou, E. Environmental and health impacts of air pollution: a review. *Front. Public Health* **2020**, *8*, 14.

4. Fuller, R.; Landrigan, P. J.; Balakrishnan, K.; Bathan, G.; Bose-O'Reilly, S.; Brauer, M.; Caravanos, J.; Chiles, T.; Cohen, A.; Corra, L. Pollution and health: a progress update. *The Lancet Planetary Health* **2022**, *6* (6), e535-e547.

5. Al-Kindi, S. G.; Brook, R. D.; Biswal, S.; Rajagopalan, S. Environmental determinants of cardiovascular disease: lessons learned from air pollution. *Nat. Rev. Cardiol.* **2020**, *17* (10), 656-672.

6. Anenberg, S. C.; Mohegh, A.; Goldberg, D. L.; Kerr, G. H.; Brauer, M.; Burkart, K.; Hystad, P.; Larkin, A.; Wozniak, S.; Lamsal, L. Long-term trends in urban NO_2 concentrations and associated paediatric asthma incidence: estimates from global datasets. *The Lancet Planetary Health* **2022**, *6* (1), e49-e58.

7. Li, Y.; Hong, T.; Gu, Y.; Li, Z.; Huang, T.; Lee, H. F.; Heo, Y.; Yim, S. H. Assessing the spatiotemporal characteristics, factor importance, and health impacts of air pollution in Seoul by integrating machine learning into land-use regression modeling at high spatiotemporal resolutions. *Environ. Sci. Technol.* **2023**, *57* (3), 1225-1236.

8. Ryan, P. H.; LeMasters, G. K. A review of land-use regression models for characterizing intraurban air pollution exposure. *Inhal. Toxicol.* **2007**, *19* (sup1), 127-133.

9. 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.; et al. Mapping air pollution with Google Street View cars: Efficient approaches with mobile monitoring and land use regression. *Environ. Sci. Technol.* **2018**, *52* (21), 12563-12572, Article. DOI: 10.1021/acs.est.8b03395.

10. 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.; Hamburg, S. P. High-resolution air pollution mapping with Google Street View cars: Exploiting big data. *Environ. Sci. Technol.* **2017**, *51* (12), 6999-7008, Article. DOI: 10.1021/acs.est.7b00891.

11. De Hoogh, K.; Chen, J.; Gulliver, J.; Hoffmann, B.; Hertel, O.; Ketzel, M.; Bauwelinck, M.; van Donkelaar, A.; Hvidtfeldt, U. A.; Katsouyanni, K. Spatial $PM_{2.5}$, NO_2, O_3 and BC models for Western Europe–Evaluation of spatiotemporal stability. *Environ. Int.* **2018**, *120*, 81-92.

12. Wang, M.; Beelen, R.; Eeftens, M.; Meliefste, K.; Hoek, G.; Brunekreef, B. Systematic evaluation of land use regression models for NO_2. *Environ. Sci. Technol.* **2012**, *46* (8), 4481-4489.

13. Larkin, A.; Geddes, J. A.; Martin, R. V.; Xiao, Q. Y.; Liu, Y.; Marshall, J. D.; Brauer, M.; Hystad, P. Global land use regression model for nitrogen dioxide air pollution. *Environ. Sci. Technol.* **2017**, *51* (12), 6957-6964, Article. DOI: 10.1021/acs.est.7b01148.

14. Liao, K.; Huang, X. H.; Dang, H. F.; Ren, Y.; Zuo, S. D.; Duan, C. S. Statistical approaches for forecasting primary air pollutants: A review. *Atmosphere* **2021**, *12* (6),

686, Review. DOI: 10.3390/atmos12060686.

15. Pedersen, M.; Giorgis-Allemand, L.; Bernard, C.; Aguilera, I.; Andersen, A. M. N.; Ballester, F.; Beelen, R. M. J.; Chatzi, L.; Cirach, M.; Danileviciute, A.; et al. Ambient air pollution and low birthweight: a European cohort study (ESCAPE). *Lancet Resp. Med.* **2013**, *1* (9), 695-704, Article. DOI: 10.1016/s2213-2600(13)70192-9.

16. Xu, X. H.; Ha, S. U.; Basnet, R. A review of epidemiological research on adverse neurological effects of exposure to ambient air pollution. *Front. Public Health* **2016**, *4*, 31, Review. DOI: 10.3389/fpubh.2016.00157.

17. Hayes, R. B.; Lim, C.; Zhang, Y. L.; Cromar, K.; Shao, Y. Z.; Reynolds, H. R.; Silverman, D. T.; Jones, R. R.; Park, Y.; Jerrett, M.; et al. $PM_{2.5}$ air pollution and cause-specific cardiovascular disease mortality. *Int. J. Epidemiol.* **2020**, *49* (1), 25-35, Article. DOI: 10.1093/ije/dyz114.

18. Eeftens, M.; Beelen, R.; de Hoogh, K.; Bellander, T.; Cesaroni, G.; Cirach, M.; Declercq, C.; Dedele, A.; Dons, E.; de Nazelle, A.; et al. Development of land use regression models for $PM_{2.5}$, $PM_{2.5}$ absorbance, PM_{10} and PM_{coarse} in 20 European study areas; Results of the ESCAPE project. *Environ. Sci. Technol.* **2012**, *46* (20), 11195-11205, Article. DOI: 10.1021/es301948k.

19. Shen, Y.; de Hoogh, K.; Schmitz, O.; Clinton, N.; Tuxen-Bettman, K.; Brandt, J.; Christensen, J. H.; Frohn, L. M.; Geels, C.; Karssenberg, D. Europe-wide air pollution modeling from 2000 to 2019 using geographically weighted regression. *Environ. Int.* **2022**, *168*, 107485.

20. Chen, J.; de Hoogh, K.; Gulliver, J.; Hoffmann, B.; Hertel, O.; Ketzel, M.; Weinmayr, G.; Bauwelinck, M.; van Donkelaar, A.; Hvidtfeldt, U. A. Development of Europe-wide models for particle elemental composition using supervised linear regression and random forest. *Environ. Sci. Technol.* **2020**, *54* (24), 15698-15709.

21. Di, Q.; Kloog, I.; Koutrakis, P.; Lyapustin, A.; Wang, Y.; Schwartz, J. Assessing $PM_{2.5}$ exposures with high spatiotemporal resolution across the continental United States. *Environ. Sci. Technol.* **2016**, *50* (9), 4712-4721.

22. Li, Y.; Sha, Z.; Tang, A.; Goulding, K.; Liu, X. The application of machine learning to air pollution research: A bibliometric analysis. *Ecotoxicol. Environ. Saf.* **2023**, *257*, 114911.

23. Gardner-Frolick, R.; Boyd, D.; Giang, A. Selecting data analytic and modeling methods to support air pollution and environmental justice investigations: A critical review and guidance framework. *Environ. Sci. Technol.* **2022**, *56* (5), 2843-2860.

24. Lu, T.; Marshall, J. D.; Zhang, W.; Hystad, P.; Kim, S.-Y.; Bechle, M. J.; Demuzere, M.; Hankey, S. National empirical models of air pollution using microscale measures of the urban environment. *Environ. Sci. Technol.* **2021**, *55* (22), 15519–15530.

25. Chen, G.; Li, S.; Knibbs, L. D.; Hamm, N. A.; Cao, W.; Li, T.; Guo, J.; Ren, H.; Abramson, M. J.; Guo, Y. A machine learning method to estimate $PM_{2.5}$ concentrations across China with remote sensing, meteorological and land use information. *Sci. Total Environ.* **2018**, *636*, 52-60.

26. Lefevre, S.; Tuia, D.; Wegner, J. D.; Produit, T.; Nassar, A. S. Toward seamless multiview scene analysis from satellite to street level. *Proc. IEEE* **2017**, *105* (10), 1884-1899, Article. DOI: 10.1109/jproc.2017.2684300.

27. Qi, M.; Hankey, S. Using street view imagery to predict street-level particulate air pollution. *Environ. Sci. Technol.* **2021**, *55* (4), 2695-2704, Article. DOI: 10.1021/acs.est.0c05572.

28. Qi, M.; Dixit, K.; Marshall, J. D.; Zhang, W.; Hankey, S. National Land Use Regression Model for NO_2 Using Street View Imagery and Satellite Observations. *Environ. Sci. Technol.* **2022**, *56* (18), 13499-13509.

29. Zhao, H. S.; Shi, J. P.; Qi, X. J.; Wang, X. G.; Jia, J. Y.; Ieee. Pyramid Scene Parsing Network. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* **2017**, 2881-2890. DOI: 10.1109/cvpr.2017.660.

30. Lamsal, L. N.; Krotkov, N. A.; Marchenko, S. V.; Joiner, J.; Oman, L.; Vasilkov, A.; Fisher, B.; Qin, W.; Yang, E.-S.; Fasnacht, Z.; et al. OMI/Aura NO_2 Tropospheric, Stratospheric & Total Columns MINDS Daily L3 Global Gridded 0.25 degree x 0.25 degree, NASA Goddard Space Flight Center, Goddard Earth Sciences Data and Information Services Center (GES DISC), Accessed: [Aug. 1, 2021], 10.5067/MEASURES/MINDS/DATA301. **2020**. (acccessed Aug 1, 2021).

31. Demuzere, M.; Hankey, S.; Mills, G.; Zhang, W.; Lu, T.; Bechtel, B. Combining expert and crowd-sourced training data to map urban form and functions for the continental US. *Sci. Data* **2020**, *7* (1), 1-13.

32. Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* **2017**, *202*, 18-27.

33. Huang, K.; Bi, J.; Meng, X.; Geng, G.; Lyapustin, A.; Lane, K. J.; Gu, D.; Kinney, P. L.; Liu, Y. Estimating daily PM2. 5 concentrations in New York City at the neighborhood-scale: Implications for integrating non-regulatory measurements. *Sci. Total Environ.* **2019**, *697*, 134094.

34. Stafoggia, M.; Bellander, T.; Bucci, S.; Davoli, M.; De Hoogh, K.; De'Donato, F.; Gariazzo, C.; Lyapustin, A.; Michelozzi, P.; Renzi, M. Estimation of daily PM_{10} and $PM_{2.5}$ concentrations in Italy, 2013–2015, using a spatiotemporal land-use random-forest model. *Environ. Int.* **2019**, *124*, 170-179.

35. Ren, X.; Mi, Z.; Georgopoulos, P. G. Comparison of Machine Learning and Land Use Regression for fine scale spatiotemporal estimation of ambient air pollution:

Modeling ozone concentrations across the contiguous United States. *Environ. Int.* **2020**, *142*, 105827.

36. Kim, S. Y.; Bechle, M.; Hankey, S.; Sheppard, L.; Szpiro, A. A.; Marshall, J. D. Concentrations of criteria pollutants in the contiguous US, 1979-2015: Role of prediction model parsimony in integrated empirical geographic regression. *PLoS One* **2020**, *15* (2), e0228535, Article. DOI: 10.1371/journal.pone.0228535.

37. Young, M. T.; Bechle, M. J.; Sampson, P. D.; Szpiro, A. A.; Marshall, J. D.; Sheppard, L.; Kaufman, J. D. Satellite-based NO_2 and model validation in a national prediction model based on universal kriging and land-use regression. *Environ. Sci. Technol.* **2016**, *50* (7), 3686-3694.

38. Hart, J. E.; Yanosky, J. D.; Puett, R. C.; Ryan, L.; Dockery, D. W.; Smith, T. J.; Garshick, E.; Laden, F. Spatial modeling of PM_{10} and NO_2 in the continental United States, 1985–2000. *Environ. Health Perspect.* **2009**, *117* (11), 1690-1696.

39. World Bank. Urban population (% of total population). *World Development Indicators, The World Bank Group*. (acccessed 01/10/2023).

40. Snyder, E. G.; Watkins, T. H.; Solomon, P. A.; Thoma, E. D.; Williams, R. W.; Hagler, G. S. W.; Shelow, D.; Hindin, D. A.; Kilaru, V. J.; Preuss, P. W. The changing paradigm of air pollution monitoring. *Environ. Sci. Technol.* **2013**, *47* (20), 11369-11377, Editorial Material. DOI: 10.1021/es4022602.

Chapter 5 Conclusion

This dissertation investigates the feasibility of utilizing open-source imagery datasets and advanced machine learning algorithms to extract image-based features for empirical air quality modeling. Specifically, street view imagery and remote sensing imagery are leveraged through computer vision and machine learning to capture air pollution variation at both local (chapter 2) and national scale (chapters 3 and 4). The design of the three studies is based on an incremental approach. The first study serves as a pilot study performed at a single city scale with the main purpose of establishing the street view-based modeling pipeline. The second study expands to a national scale with a focus of model development on its own. The third study further aims for model application to generate air pollution prediction surfaces as the final products. All of the three studies have demonstrated the potential of combing public imagery datasets and machine learning to provide an innovative empirical modeling approach for estimating street-level patterns of air pollution.

5.1 Key findings

The first study develops a single-city air quality model to predict street-level particulate air pollution solely based on street view imagery. A set of GSV features were extracted from GSV images through a deep learning model (i.e., PSPNet). Traditional step-wise linear regression models (commonly applied in the LUR studies) were developed solely based on the GSV features to directly compare with previous LUR models that relied on traditional GIS-derived variables. Three feature selection methods were conducted to exclude irrelevant and spurious features for model development: (1) theory-driven feature selection, (2) data-driven feature selection, and (3) integrated feature selection.

The results from the first study suggest that solely using GSV-derived feature can achieve good model performance. Adjusted R^2 (10-fold CV R^2) of BC and PN models with integrated feature selection was 0.57-0.64 (0.50-0.57) and 0.65-0.73 (0.61-0.66), respectively. Compared to LUR models with traditional GIS-derived variables, the GSV-based air quality modeling approach achieved similar model performance and reasonably consistent spatial pattern for model prediction. More importantly, the GSV-based models identified additional local hotspots that traditional GIS-derived models did not. The results suggest that GSV imagery, processed with computer vision techniques, is a promising data source to develop empirical air quality models with high spatial resolution and consistent predictor variables processing protocol.

The second study extended the previous effort from a single-city scale to the national scale and developed GSV-based national NO_2 models across multiple years (2007-2019). Besides the GSV features, remote sensing derived air quality estimate was also considered, i.e., the satellite observations of NO_2 from OMI. Unlike the single-city study which developed the step-wise linear regression models, machine learning models (i.e., random forest regression) were applied for NO_2 air pollution modeling. The impact of image availability on the model performance was systematically evaluated.

The results from the second study suggest that GSV-derived features still show good model performance in explaining air pollution variation at national scale. Analysis of the feature importance score showed that among GSV features, transport vehicles, transport network, and built environment variables contributed most to the NO_2 air pollution. Adding satellite-derived aerosol estimate (i.e., OMI) can further boost moder performance, especially when GSV images are not sufficient for modeling. However, the addition effect decreased significantly as more GSV images being included into the

model. Based on model prediction surfaces for several sample cities, we found that compared to conventional LUR approach, the GSV-based models suggested enhanced ability to capture intra-urban NO_2 variability. At last, we developed image-unlimited and image-limited models and the results suggest that a parsimonious image sampling strategy may be sufficient and most cost-effective for model development and application.

Due to the limitations of GSV image availability in spatial and temporal dimensions, the third study explores the potential of solely remote sensing based features. For comparison, this work applies GSV-derived features, satellite derived aerosol estimates, and Landsat-derived features to develop random forest models for both national NO_2 and $PM_{2.5}$ air pollution through 2007-2019. Different combinations of image-based predictor variables are tested and evaluated. The ultimate goal of this study is to generate national NO_2 and $PM_{2.5}$ air pollution maps at high spatial resolution across multiple years.

The results from the third study find that for both pollutants, the full models (i.e., using all three types of predictors) achieved the highest model fit compared to other model configurations. Models incorporating GSV features generally had improved model fit as compared to models solely using remote sensing derived predictors. Although the difference of model fit was relatively small among models, the model prediction surfaces display increased spatial variation. An examination of model predictions within US urbanized areas suggests that the GSV-based models captured both the highest and lowest pollutant concentrations while remote sensing features tended to smooth the air pollution variations. The results suggest that GSV features may have the capability to better capture fine-scale air pollution variability – a hypothesis that we aim to verify in further studies (e.g., compare against mobile monitoring data).

5.2 Limitations and Implications for future studies

This dissertation provides image-based empirical air quality modeling approaches using advanced machine learning techniques and open-source imagery data (i.e., street view and satellite imagery). The proposed modeling approach is promising as it only involves data sources which are uniformly available on the large spatial scale making it possible to establish a consistent data collection, processing, and modeling protocol across large geographies regardless of political boundaries. Still, these innovative image-based predictors have their own limitations.

The major limitation of the GSV variables is that the availability of GSV images varies among different locations and time periods. Most GSV images are distributed along the road network or locations with more human activities. Urban areas are more likely to have a denser network of GSV imagery compared to rural areas. But even in those developed regions, the updating frequency of GSV imagery also varies. This spatio-temporal instability leads to model uncertainty. Compared to GSV imagery, satellite imagery has higher temporal resolution with stable updating. However, satellite-derived features have coarser spatial resolution which are incapable of capturing local hotspots or intra-urban variation of air pollution. In addition, as no computer vision techniques are flawless, misclassification occurs and may lead to errors in image-based features.

A few improvements could be made in future studies. First, more advanced modeling approaches may be investigated. For example, currently PSPNet is used to extract GSV features from the GSV images before model development. In the future, deep learning techniques may be directly applied to build air quality models. Second, more spatial-resolved air quality data may be collected for modeling. Current work only modeled annual air pollution concentrations from EPA sites, which have limited spatial

coverage. As low-cost sensor-based monitoring campaigns are surging. With proper data quality control, data from such campaigns may be leveraged to play an important role in capturing intra-urban variability of air pollution. Third, remote sensing products with higher spatial resolution have been emerging in recent years. Only remote sensing-based models may witness a significant improvement when utilizing these new products.

Finally, the results of this dissertation may be applied to provide an important evidence base for national and global exposure assessment, environmental injustice analysis and other epidemiology studies. Future studies may further expand our current effort. For example, to date, the application of machine learning in air quality modeling has seen a surge but few have served health studies. The insights ML approaches can provide into exposure and health assessment, especially for disadvantaged populations, remains to be studied. Future work may combine our high-resolved air pollution estimates with health data to perform exposure assessment of air pollution to aid human health studies. The predicted air pollutant concentrations can be applied to estimate exposure disparities among different demographic groups and subsequently linked to mortality.