Senseable City Lab :.:: Massachusetts Institute of Technology

This paper might be a pre-copy-editing or a post-print author-produced .pdf of an article accepted for publication. For the definitive publisher-authenticated version, please refer directly to publishing house's archive system

Applied Geography xxx (xxxx) 102943

Contents lists available at ScienceDirect

Applied Geography

journal homepage: www.elsevier.com/locate/apgeog

Improving NO_2 prediction by integrating tree diversity, urban form, and scale sensitivity through mobile monitoring

Ye Tian [a, b, e], An Wang [a], Simone Mora [a, d, *], Priyanka deSouza [c], Xiaobai Yao [e], Fábio Duarte [a], Hui Lin [b], Carlo Ratti [a]

[a] Senseable City Laboratory, Department of Urban Studies and Planning, Massachusetts Institute of Technology, 9-250 77 Massachusetts Avenue, Cambridge, MA, 02139, USA
[b] School of Geography and Environment, Jiangxi Normal University, 99 Ziyang Avenue, Nanchang, Jiangxi, CN 330022, China
[c] Department of Urban and Regional Planning, University of Colorado Denver, 1201 Larimer St, Denver, CO, 80204, USA
[d] Department of Computer Science, Norwegian University of Science and Technology
[e] Department of Geography, University of Georgia, 210 Field St., Athens, GA, 30602, USA

ARTICLE INFO

Handling Editor: J Peng

Keywords:
Air pollution
Urban forest
Urban form
Lacunarity
Opportunistic mobile monitoring

ABSTRACT

Air pollution is a major threat to public health. However, two issues have not been adequately addressed in most conventional Land Use Regression models for air pollution prediction: 1). A combination of urban forest involvement and urban form representation; 2). Scale sensitivity analysis of model variables. Here, we apply lacunarity to investigate the spatial sensitivity of predictors, incorporate 2-D and 3-D urban form to comprehensively characterize the urban environment, and examine the tree diversity impacts on air pollution distribution using unique NO_2 datasets collected through opportunistic mobile monitoring in the Bronx, New York, and Oakland, California. We find that lacunarity-optimized models could reduce the computation burden by extracting the upper limits of the spatial heterogeneity of predictors while keeping the model accuracy simultaneously. Furthermore, there are synthetic effects between the urban form and tree diversity on NO_2 distribution, and such effect directions could be non-monotonic. Finally, although the increase in tree diversity could facilitate the reduction of regional NO_2 concentration, it is essential to seek a balance between tree diversity and tree dominance to effectively improve air quality on the city scale. The findings are useful for environmental scientists striving for better air quality and urban planners caring for the well-being of cities.

Glossary

AI	Aggregation Index
ANN	Artificial Neural Network
AQI	Air Quality Index
AR	Aspect Ratio
CA	Class Area
CDC	Centers for Disease Control and Prevention
DBH	Diameter at Breast Height
DCAS	Department of Citywide Administrative Services
ED	Edge Density
EDF	Environmental Defense Fund
EPA	Environmental Protection Agency
H_Var	Height Variance
LPI	Largest Patch Index
LSI	Landscape Shape Index
LUR	Land Use Regression
MDI	Mean Decrease in Impurity
MIT	Massachusetts Institute of Technology
MLR	Multiple Linear Regression
NY	New York
NYC	New York City
PD	Patch Density
PDP	Partial Dependence Plot
RF	Random Forest
RH	Relative Humidity
RMSE	Root Mean Squared Error
SHAP	SHapley Additive exPlanations

* Corresponding author.
E-mail addresses: yetian92@mit.edu (Y. Tian), an_wang@mit.edu (A. Wang), moras@mit.edu (S. Mora), priyanka.desouza@ucdenver.edu (P. deSouza), xyao@uga.edu (X. Yao), fduarte@mit.edu (F. Duarte), huilin@cuhk.edu.hk (H. Lin), ratti@mit.edu (C. Ratti).

https://doi.org/10.1016/j.apgeog.2023.102943
Received 15 December 2022; Received in revised form 21 February 2023; Accepted 19 March 2023
0143-6228/© 20XX

CORRECTED PROOF

Note: Low-resolution images were used to create this PDF. The original images will be used in the final composition.

SHI	Shannon Index
SII	Simpson Index
SVF	Sky View Factor
TAZ	Traffic Analysis Zone
UFORE	Urban FORest Effects

1. Introduction

Air pollution is associated with increased mortality and a wide range of serious diseases (Kheirbek et al., 2016). A challenge in predicting urban air pollution is ascribed to the spatiotemporal heterogeneity of pollutants distribution and complex urban microenvironments (Apte et al., 2017). Due to their high cost, existing stationary regulatory monitoring networks are sparse and cannot capture granular variations of pollution (Do et al., 2020). In recent years, mobile air quality monitoring has emerged as an effective practice to map hyperlocal air pollution in dense urban environments (DeSouza et al., 2020).

High-resolution mobile monitoring provides unique opportunities for environmental scientists to understand the interactions between air quality, human activities, and urban physical environments from a holistic perspective of urban sustainability at a fine scale (Cummings et al., 2021). Although LUR (Land Use Regression) is the most widely-used tool for air quality data interpretation and prediction due to the relatively lower level of demand for input data requirements compared to dispersion and chemical transport models (Messier et al., 2018; Saha et al., 2019), the current LUR may neglect the combination of three essential aspects of urban environments: 1). Tree diversity, which plays a pivotal role in ameliorating air quality and improving urban ecosystem health (Roeland et al., 2019). The One Health initiative, which emphasizes the connection between the health of people, animals, and the environment, recognizes the key role that trees play in regulating urban ecology (CDC, 2022); 2). Scale sensitivity of predictors, which lacks widely recognized theoretical guidelines regarding the optimum buffer distances for variable aggregation during model construction; 3). Urban form, particularly in 3-D, which can better characterize the influence of the physical environment on the atmosphere wind field with higher explanatory power in estimating intra-urban air pollutants than their 2-D counterparts (Ke et al., 2022; Tian et al., 2022).

The role of tree diversity is essential in supporting ecosystem services and air quality improvement, particularly for the abatement of airborne particulate pollution (Manes et al., 2014). For example, the UFORE (Urban FORest Effects) model, developed by the U.S. Forest Service to quantify multiple urban forest ecosystem services, identified tree diversity as one of the most effective methods of improving local air quality (Saunders et al., 2011). To illustrate, Sicard et al. (2018) devised a novel Species Air Quality Index (S-AQI) of suitability to air quality improvement for tree diversity and suggested that city planners should select species with an S-AQI > 8. Manes et al. (2014) considered tree diversity in different climatic conditions to confirm the crucial role of trees in supporting significant ecosystem services to improve air quality. However, how and to what extent tree diversity can impact air pollution distribution at different spatial scales is still largely underexplored.

3-D urban form, consisting of both horizontal and vertical elements, can directly influence urban ventilation and the dispersion of pollutants (Peng et al., 2021), and their influence has also been verified in different physical environments, such as street canyons, neighborhoods, and communities, through idealized experiment models (Hang et al., 2012; Yuan et al., 2019). For example, SVF (Sky View Factor) can influence urban ventilation resistance, shape ventilation corridors, and determine the dispersion of air pollutants (Fang & Zhao, 2022). Tang et al. (2013) used building heights and geometry to enhance the estimation of land use-related variables and the pollution dispersion fields for long-term air pollutants. Edussuriya et al. (2011) found that air pollution concentrations were impacted by the AR (Aspect Ratio), building volume, and building height variation.

Furthermore, tree diversity and urban form also have synthetic effects on air quality (Pugh et al., 2012). For example, deposition and dispersion are tightly coupled to the 3-D urban form and the synoptic-scale flow. As the porosity of the barrier increases, the effective path-length decreases, and the opportunity for the removal of particles by deposition increases (Tong et al., 2016). The AR significantly affects pollutant dispersion because of alterations in airflow patterns (Zhong et al., 2016). When horizontal length scales and AR are small and residence times are short, there is little opportunity for deposition to become effective. Moreover, green walls in street canyons with a large AR may make appreciable differences in ground-level concentrations due to deeper or narrower street canyons (Pugh et al., 2012). Notwithstanding, limited studies incorporated both tree diversity and 3-D urban form factors into the air quality research.

Additionally, most prior studies relied on arbitrary multiple buffer distances, which could not interpret the spatial scale sensitivity, were prone to UGCoP (Uncertain Geographic Content Problem) that biases the final results due to the different analysis units (Kwan, 2012), and be time-consuming for model construction. Therefore, researchers applied lacunarity, which measures the spatial heterogeneity of predictors depending on the image's texture at varying scales (Plotnick et al., 1996). To illustrate, objects with low lacunarity are homogeneous and translationally invariant because of the same gap sizes. In contrast, high-lacunarity objects have a wide range of gap sizes which are usually heterogeneous and translationally variant (Plotnick et al., 1993). As a result, for variable aggregation, increasing the buffer distance beyond a certain level would not result in any substantive impact on the statistical results, because the buffer distance surpasses the maximum level of the repeating patterns in the landscape (Roces-Diaz et al., 2015). Therefore, lacunarity could be used to define the upper limit of spatial variability when quantifying the landscape patterns (Labib et al., 2020).

Our study targets bridging the current research gaps in urban air pollution prediction, including the underrepresentation of urban forestry information, the lack of urban form involvement, and the inefficiency of the traditional buffer sizing method. To illustrate, we integrate tree diversity and 3-D urban form into existing 2-D information-dependent LUR tools to identify the most influential predictors and to sheds light on the interactions between air pollution, urban form, and tree diversity, which demonstrates our multi-disciplinary efforts to attain optimal health for the entire urban ecological system. Hyperlocal NO_2 levels were collected through an opportunistic mobile monitoring campaign in the Bronx, NY. To explore the impacts of urban form and tree diversity in different cities, we transferred our air quality modeling methodology to Oakland, California, a city on the U.S. west coast with a distinct urban environment from the Bronx, to access the model's generalization performance. We also conducted the lacunarity analysis to determine the upper bounds of predictors' buffer sizes while keeping the fixed buffer sizing as a benchmark. Models developed using lacunarity-optimized and conventional fixed buffers were contrasted to reveal the effects of spatial heterogeneity on air quality prediction. The proposed lacunarity method is highly scalable and transferrable to other air quality prediction applications. Our study is not only relevant to environmental scientists and ecologists striving for better air quality but to urban planners and decision-makers to strategically manage urban forests and physical environments.

CORRECTED PROOF

2. Methods

2.1. Research areas

Fig. 1 displays the mobile NO_2 (451,322 mobile measurements) in the Bronx, which has the highest asthma hospitalization rate in New York City, with a dense road network and other noxious land uses that are significant sources of air pollution (Gorai et al., 2014). The NO_2 data has a mean value of 9.45 ppb (parts per billion) and a standard deviation of 3.96 ppb collected between September 10th to December 17th, 2021 as part of the *City Scanner* research initiative at the Massachusetts Institute of Technology (senseable.mit.edu/cityscanner). Five vehicles operated by NYC Department of Citywide Administrative Services (NYC-DCAS), including four road maintenance trucks and one sanitation vehicle, were equipped with City Scanner low-cost sensing nodes with Alphasense NO_2-A4 electrochemical gas sensors, high-accuracy GPS, and temperature and humidity sensors (Mora et al., 2019, pp. 743–748). More details can be seen in Table S1 (Mobile monitoring) in Supporting Information. The NO_2 mobile measurements in Oakland refer to Apte et al. (2017).

2.2. Research design

For the data preparation, the mobile monitoring NO_2 data was first calibrated against a reference air quality station in the Bronx and was aggregated by the road segment to reduce the systematic noise and biases (Section 2.2.1). Then, we calculated the lacunarity value for different land use types and street trees. Tree diversity was characterized by SHI (Shannon Index), the richness and evenness degree of tree species, SII (Simpson Index), the dominance of tree species, and DBH (Diameters at Breast Height) (Section 2.2.2). Next, we calculated the 2-D (Section 2.2.3) and 3-D (Section 2.2.4) urban form factors to characterize the urban landscape comprehensively. The upper bound for predictor aggregation was calculated from lacunarity values (Section 2.2.5). Combined with other covariates, these variables were used to predict NO_2 through two modeling techniques: lacunarity-optimized and conventional LURs. After feature selection, we constructed the RF-(Random Forest)-, ANN (Artificial Neural Network)-, and MLR (Multiple Linear Regression)- LUR (Land Use Regression) models to analyze their correlation and interpret the results through feature importance, influential directions, and marginal effects (Section 2.2.6). Fig. 2 shows the research flowchart. Table 1 shows all the predictors according to the potential NO_x emission sources and factors affecting their distribution and dispersion (Apte et al., 2017; Messier et al., 2018).

2.2.1. NO_2 preprocessing

Before using the mobile monitoring NO_2 for further analysis, we calibrate the data using minute-by-minute reference EPA (Environmental Protection Agency) station data. Multiple calibration algorithms are tested, including linear and non-linear ones, to best reproduce the reference data. Calibration algorithm performances are stringently evaluated using 10% hold-out samples 100 times. Calibrated *City Scanner* NO_2 data is highly correlated with the reference signals using the best-performing algorithm (average external $R^2 = 0.92$), as shown in Fig. S1. To further increase the transferability of the research, we also construct the model using the NO_2 data in Oakland, CA, which was collected by Google Street View vehicles from May 28th, 2015 to December 21st, 2017 with 38,252 mobile measurements (Apte et al., 2017). All calibrations were performed "through the probe" by connecting the calibration system to the car at the sample inlet and maximizing the temporal correlations between the individual pollutants on a daily basis (Brantley et al., 2014). To remove the impacts of relative humidity on NO_2 measurements, values in both Bronx and Oakland were removed for relative humidity above 90% due to the growth of the hygroscopic particle, leading to mass overestimations (Goin et al., 2021). To keep compatible and alleviate the impact of the "Spike" signal from low-cost sensors (e.g. keep reporting high values for approximately 30s when a truck pass by), all NO_2 was processed and aggregated by the midpoint of each road segment and used median value to represent the NO_2 at 30m spacing to mitigate the measurement uncertainty at point locations in time as reported in previous research (Apte et al., 2017; DeSouza et al., 2020; Goin et al., 2021).

2.2.2. Tree diversity

Tree diversity is critical in determining pollutant removal in the urban environment. We extract 66,294 and 38,564 trees along the streets with 129 and 195 species in the Bronx and Oakland, respectively (Fig. S2: Spatial distribution of tree species in the Bronx and Oakland). To further explore how tree diversity may affect NO_2 concentration, we summarize the median NO_2 value around each tree species (Section 3.1.2). Previous ecological studies strongly recommended using both SHI (Shannon Index) and SII (Simpson Index) to account for the most complete mathematical description of species richness and relative abundance (Galle et al., 2021). The SHI was initially proposed by Shannon (1948) to quantify the level of the mixture and can be extended to measure tree species entropy, which increases as the community's richness and evenness increase. The SII was introduced by Simpson (1949) to measure the degree of dominance, such as the dominant distribution for specific tree species. Both of them can be used to measure tree diversity. More details about SHI and SII can be seen in

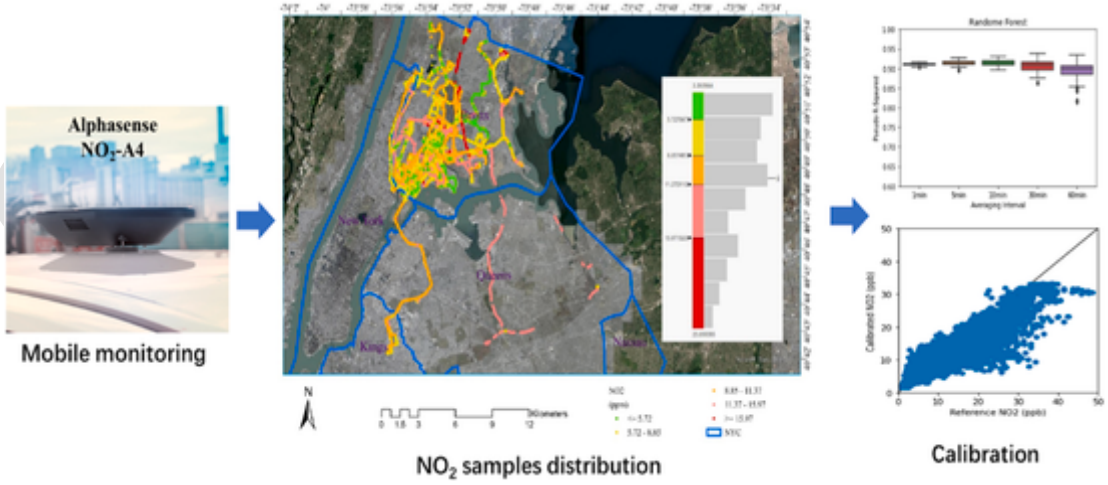

Fig. 1. NO_2 concentrations in the Bronx were collected in 2021 as part of a mobile sampling experiment.

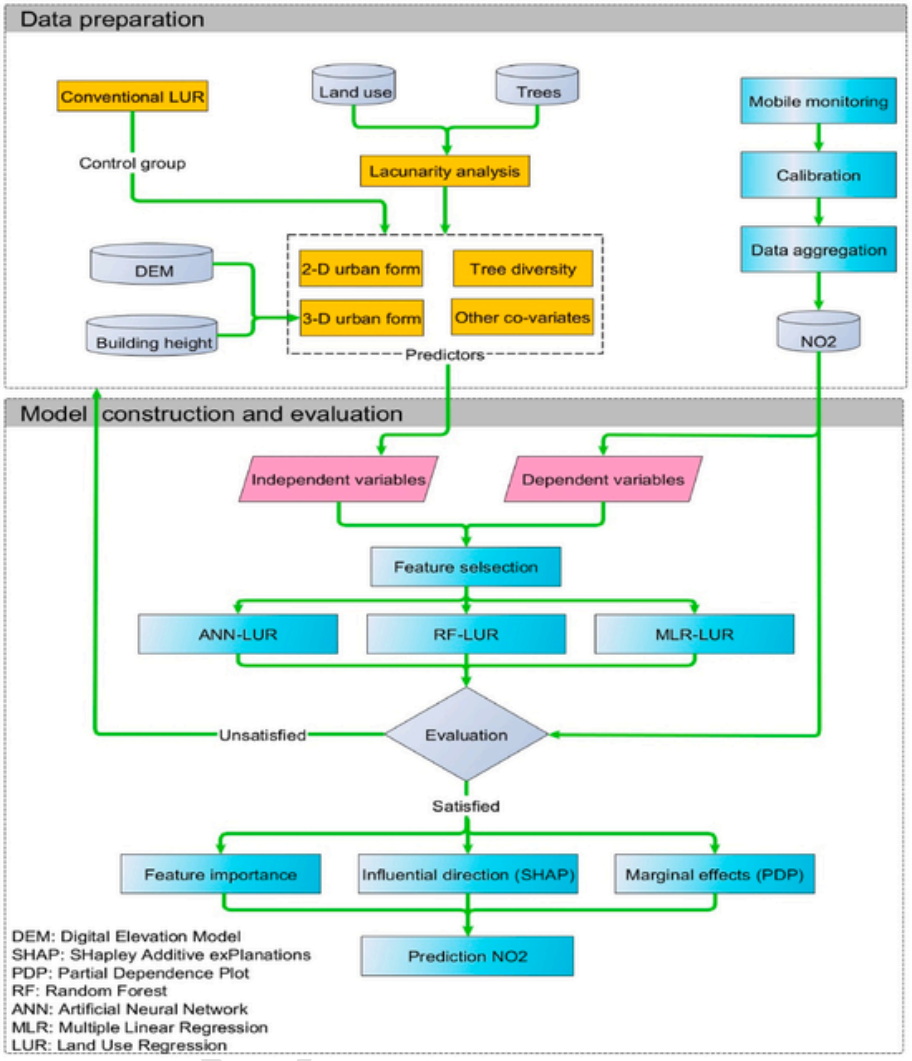

Fig. 2. Research flowchart.

Supporting Information Section S1: Tree diversity. The DBH (Diameter at Breast Height) is also aggregated based on the median value of diameters for all tree species.

2.2.3. 2-D urban form metrics

To comprehensively characterize the urban environment, we calculate 2-D urban form metrics through the integration of land use and landscape metrics. Specifically, we calculate six landscape metrics, including CA (Class Area), ED (Edge Density), PD (Patch Density), LPI (Largest Patch Index), LSI (Landscape Shape Index), and AI (Aggregation Index), for each land use type (Fig. S3: land use in Bronx and Oakland). Table S2 describes each metric in detail (McGarigal, 2015). These metrics capture the spatial coverage, fragmentation, patch dominance, and shape complexity of land use patterns (Liu et al., 2018b) that could shape the spatiotemporal dispersion of air pollutants in the urban environment (Cho & Choi, 2014; Rodríguez et al., 2016). In addition to using each metric as an independent covariate, we also add their interaction terms as predictors (Tian et al., 2020a; Tian & Yao, 2022). This gives us 24 urban form metrics (four land uses times six landscape metrics) in total. In this study, we apply each term with a combined meaning. For example, AI_Water_3000 means the Aggregation Index for water calculated within the 3,000m buffer size, AR_100 indicates the Aspect Ratio calculated within the 100m buffer size, and so forth.

2.2.4. 3-D urban form metrics

We apply SVF (Sky View Factor), H_Var (Height Variance), and AR (Aspect Ratio) to capture 3-D urban form factors given their synthetic effects with tree diversity on air pollution distribution (Pugh et al., 2012). More technical details of 3-D urban form metrics can be found in the Supporting Information Sections from S2 to S4.

To illustrate, the SVF quantifies the openness to the sky of a given location, which reflects the degree of urban ventilation affecting air pollution dispersion (Liu et al., 2016). SVF ranges between 0 and 1 where close to 1 indicates that almost the entire hemisphere is visible (e.g., planes and peaks), and close to 0 means almost no sky is visible (e.g., deep sinks and valleys) (Zakšek et al., 2011). The H_Var describes the variation of the building height, which impacts the distribution of wind flow and the corresponding ventilation corridor (Liu et al., 2016). The AR represents the ratio of building height to street width, which gives an idea of the change of surface-air interfaces due to the vertical dimensions of buildings compared to the flat terrain (Grimmond & Oke, 1999), which plays an essential role in the interaction between tree diversity, urban form, and air pollution distribution. For example, air quality improves with a decrease in AR due to a larger vertical exchange of air pollutants between street roofs suggesting that an open central street can result in better air quality (Shen et al., 2017).

Table 1
Summary of model predictors.

Predictors	Indices	Sources	Resolution
Tree diversity	Shannon Index Simpson Index Diameters at Breast Height	OpenTrees.org	10m
2-D urban form	Landscape metrics for different land uses	Open Street Map	30m
3-D urban form	Sky View Factor Building Height Variance Aspect Ratio	Microsoft Building Footprint Data	30m
Meteorological factors	Wind Speed Wind Direction Temperature Precipitation Relative Humidity Sea Level Pressure Feel Temperature	Iowa Environmental Mesonet	Station measurement with every 5-min interval
Socioeconomic factors	Population Density Median Household Income	U·S. Census Braeuer	Block level
	Traffic	Annual Average Daily Traffic	Daily
Others	Nearest distances to the landfill, industry, waterway, bus stations	Open Street Map	–

(The building and land use data are polygon-based Shapefile data. We calculated the average building area for the Bronx (159.03 m^2 for 698,806 polygons) and Oakland (176.55 m^2 for 161,415 polygons). Therefore, we aggregate them by the grid size of 30m (900 m^2 includes almost 6 buildings for each grid) for urban form calculation. Likewise, the trees are point-based Shapefile data, and we calculate the average tree interval (6.22m for 493,205 trees) in the Bronx and (8.83 m for 38,564 trees) and Oakland. Thus, we calculate tree diversity by the grid size of 10m (100 m^2), and each grid could include around 11 to 16 trees).

2.2.5. Lacunarity

Compared with previous studies using a series of arbitrary buffer distances to construct the model predictors (Kerckhoffs et al., 2022; Zhang et al., 2022), we apply lacunarity, as proposed by Mandelbrot (1982), to describe the texture of the landscape. In this study, we calculate the lacunarity value for land uses and street trees that have proven to be correlated with NO_2 distribution (Apte et al., 2017) and select the flattened value of lacunarity as the upper bound of the buffer distance (Section 3.1.1). This way, we can capture the maximum ranges of spatial heterogeneity of predictors and reduce the computation burden, particularly for predictors aggregation at large buffer distances. Consequently, it provides a comprehensive and less resource-intensive solution to address scale effects for model construction (Labib et al., 2020). Following Equation (1) is used to calculate lacunarity in this study:

$$\wedge(r) = \frac{Var\,[S(r)]}{E^2\,[S(r)]} + 1 \quad (1)$$

where $\wedge(r)$ is the lacunarity for the box size r, S(r) is the number of pixels that fall in the box, and the box mass. In this study, the minimum box size is 1 and the maximum box size is half of the image dimensions (maximum value of image length or width). The box size value where the lacunarity values flatten out refers to the raster data transformed from heterogeneity to homogeneity from this distance, and it will be regarded as the scale upper limits of the appropriate buffer distance.

2.2.6. Model construction

We conduct two sets of models: the conventional LUR model, using a series of fixed buffer distances as in previous studies (Kerckhoffs et al., 2022; Zhang et al., 2022), and lacunarity-optimized, using a series of the same buffer distances but setting the upper bounds according to the flattened point of lacunarity value. The basic analysis units are the different buffer regions constructed around each mobile measurement. Previous researchers have found that RF (Random Forest) models have been at least as accurate and, in many cases, more accurate than linear models (Tian et al., 2020a). Therefore, this study combined RF and LUR for model construction. More details can be seen in Fig. S4: The procedures for constructing the RF model. To compare the model performance, we apply the same strategy in ANN (Artificial Neural Network) and MLR (Multiple Linear Regression) models. Considering potential multicollinearity between the covariates, we use the Greedy Stepwise algorithm to filter correlated covariates and select the subset of the most significant covariates as candidates in the RF-LUR model. To avoid overfitting, we apply the 10-fold cross-validation to tune the model parameters. Besides the R^2 and RMSE (Root Mean Squared Error), we plot feature importance, SHAP (SHapley Additive exPlanations) value, and PDP (Partial Dependence Plot) to further interpret the model. To illustrate, feature importance is computed as the decrease in node impurity weighted by the probability of reaching that node, namely MDI (Mean Decrease in Impurity). It counts the times a feature is used to split a node weighted by the number of samples it splits (Perrier, 2015). SHAP value was proposed by Lundberg and Lee (2017) as a united approach to explaining the output of any machine learning model because it can: 1). show how much each predictor contributes, positively or negatively, to the target variable; 2). get its own set of SHAP values for each observation; 3). be calculated for any tree- or non-tree-based models. The PDP shows the predictors' marginal effects on the predicted outcome (Friedman, 2001), and it demonstrates whether the relationship between the target and a feature is linear, monotonic, or more complex and how the average prediction changes as the specific predictor changes.

3. Results

3.1. Descriptive results

3.1.1. Lacunarity analysis

Fig. 3 shows the lacunarity value for different land uses and street trees. To illustrate, the "flattened point" (where the value flattens out and the gradient approaches zero, indicating the maximum value of heterogeneity pattern) would be used as the bounded buffer sizes (labeled as vertical lines) for the lacunarity-optimized model. For example, in the Bronx, the value of the flattened point for "Buildings" is 34, so the upper bound is 1020m (34 × 30m, 30m represents the spatial resolution of box size). Therefore, we select 100m, 300m, 500m, and 1020m as the buffer distances for calculating urban form metrics for "Buildings" in the lacunarity-optimized model. If the product of lacunarity value and the box size is larger than 5000m, we set 5000m as the upper bound because it has been used as the maximum potential influencing range in previous air quality models (Cowie et al., 2019; Naughton et al., 2018). The procedures for detecting the flattened value and selecting upper bounded buffer sizes keep the same in Oakland. Details can be seen in Table S3 (Selected features for lacunarity-optimized models in the Bronx and Oakland).

As can be seen from Fig. 3, in the Bronx, the flattened value for green space (198) and water (200) is higher than that for roads (18) and buildings (34), which indicates significant gaps, uneven distribution, and a higher level of spatial heterogeneity for green space and water than buildings and roads. The roads have the smallest flattened value compared to other land uses, suggesting that road networks have relatively homogeneous spatial distribution in the urban environment. This reflects the relatively dense distribution of road networks and the sparse distribution of water and green space in the Bronx. Likewise, in Oakland, the distribution of lacunarity values for land uses and street trees maintains a similar trend where buildings and roads are more homogeneously distributed than the greenspace and water.

CORRECTED PROOF

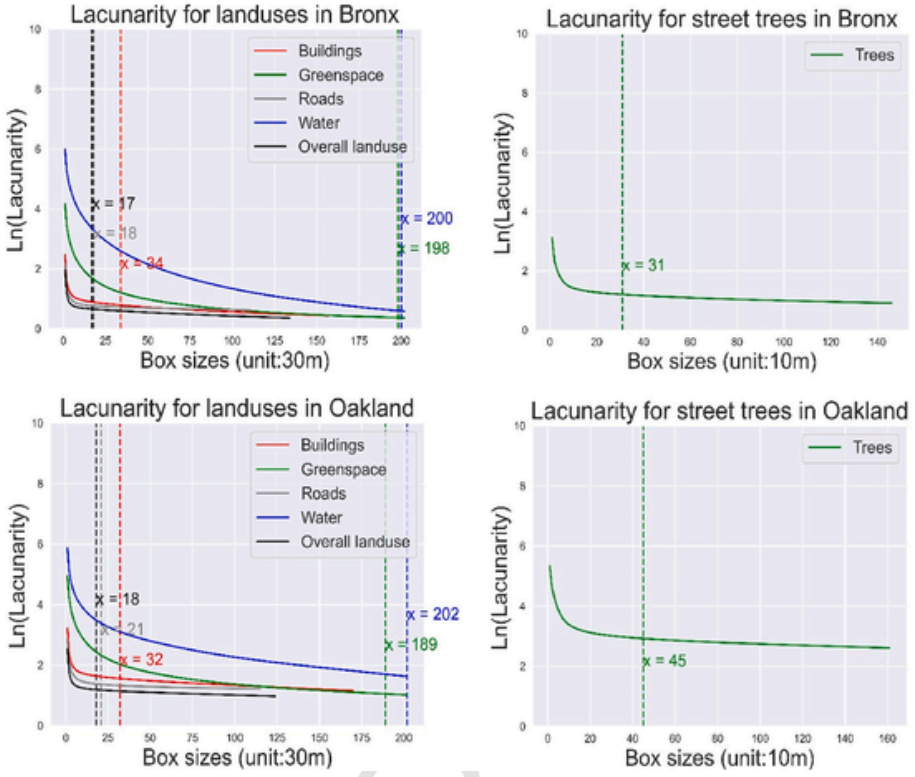

Fig. 3. Lacunarity-bounded buffer sizes in Bronx and Oakland
(The "Overall landuse" means we combine different land use types into a single image. The moving window sizes for land uses and street trees are 30m and 10m, respectively).

3.1.2. *Tree diversity and NO_2 distribution*

We calculate the median value of NO_2 concentration around deciduous, semi-deciduous, and evergreen trees aggregated based on the buffer distances of 310m (31 × 10m) and 450m (45× 10m), the flattened value of street trees (Fig. 3). The distribution of NO_2 concentration for different tree species is shown in Fig. 4a and b. As can be seen, deciduous and semi-deciduous trees are surrounded by the widest range of NO_2 concentration in the Bronx and Oakland, respectively, and the median values of NO_2 concentration for deciduous and evergreen trees are close in both Bronx and Oakland. In addition, the deciduous trees also have a higher average level of NO_2 concentration than evergreen trees in both Bronx (9.73 ppb and 9.71 ppb for the deciduous and evergreen trees, respectively) and Oakland (3.61 ppb and 3.37 for the deciduous and evergreen trees, respectively) Oakland. Additionally, the Bronx has an average higher level of NO_2 concentration (9.72 ppb) than Oakland (3.67 ppb) for all trees, which could be attributed to the fact that the Bronx has a higher percentage of deciduous trees (89.92%) than Oakland (38.97%). This accords with previous findings that evergreen trees were more effective in reducing annual NOx concentration than deciduous trees (Tiwari & Kumar, 2020).

In Fig. 4 (c) and (d), we summarize the 5th to 95th percentile value for the median NO_2 concentration for each tree (point-based), and the corresponding values were matched to certain tree species to explore their differences in NO_2 concentration distribution. In the Bronx, the "Callery pear", a deciduous tree with oval leaves, has the highest level of NO_2 concentration (10.28 ppb). The "northern red oak", regarded as "one of the most handsome, cleanest, and stateliest trees in North America" (Foundation, 2022), enjoys a relatively lower level of NO_2 concentration (9.45 ppb). In Oakland, the worst air quality is found around "Pistacia Chinensis" (11.92 ppb), with deciduous leaves and absent terminal leaflet, while the "Ligustrum lucidum", an evergreen tree with opposite and glossy dark green leaves, is surrounded by the lowest level of NO_2 concentrations (1.98 ppb). In addition, the SII (Simpson Index) and SHI (Shannon Index) augment with the increase of the buffer distances in both cities. Details can be seen in Table S4 (Summary statistics of tree diversity in Bronx and Oakland). The findings suggest that it is necessary to incorporate tree diversity into air quality models as the NO_2 concentration varies around different tree species.

3.2. *Model performance*

We find that RF-LUR overall outperforms ANN-LUR and MLR-LUR for both lacunarity-optimized and conventional LUR models in Bronx and Oakland with higher R^2 and lower RMSE. More details can be seen in Table S5. Therefore, we further compared the model performance of the RF-LUR models in Fig. 5. As can be seen, the performance of the lacunarity-optimized models is comparable to that of the conventional LUR models in both Bronx (R^2 = 0.51 vs. R^2 = 0.52) and Oakland (R^2 = 0.79 for both models). However, the number of predictors has significantly reduced from 236 in conventional LUR models (Table S6: Summary statistics of predictors for conventional LURs in the Bronx) to 158 in the Bronx and 161 in Oakland (Table S3). It demonstrates that the lacunarity analysis is advantageous in reducing or avoiding the repetitive aggregation for predictors, especially for calculating predictors with large buffer distances, which is time- and labor-intensive. Therefore, instead of applying a series of arbitrary buffer distances, we can calculate the lacunarity value first to find the upper bounds of the buffer distance for each predictor and leave out the distances that are larger than that bounds.

Besides, the R^2 of RF-LUR has reduced to 0.47 and 0.48 without considering tree diversity in lacunarity-optimized and conventional LUR models, respectively, in the Bronx, and reduced to 0.77 for both models in Oakland, which confirms the necessity of incorporating tree diversity to improve model accuracy. Likewise, the model performance decreased when we ignore the urban form predictors (R^2 reduce to 0.47 and 0.48 for lacunarity-optimized and conventional LUR models, re-

UNCORRECTED

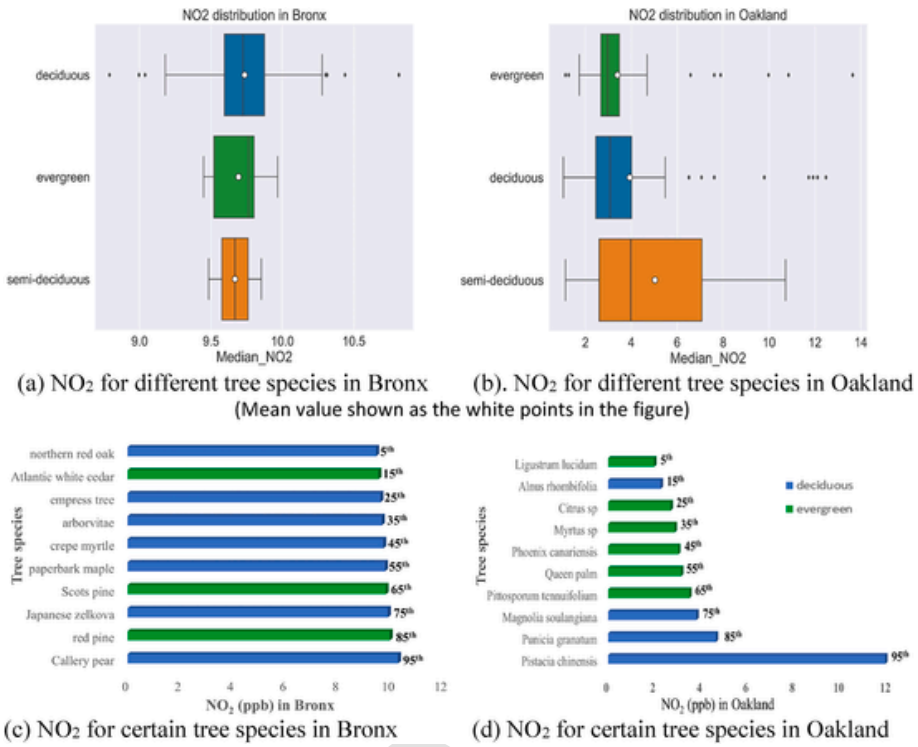

Fig. 4. NO_2 concentration around different tree species in Bronx and Oakland
((a) and (b) show the median NO_2 concentration for all deciduous, semi-deciduous, and evergreen trees in Bronx and Oakland, respectively. (c) and (d) indicate NO_2 concentration levels ranging from the 5th to 95th percentile around specific tree species in Bronx and Oakland, respectively).

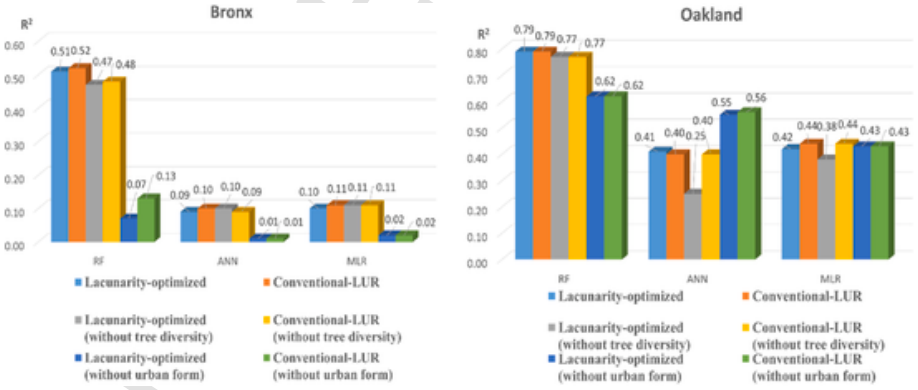

Fig. 5. Model performance in Bronx and Oakland in different groups.

spectively, in the Bronx, and reduce to 0.62 for both models in Oakland).

To make a comparison, we explore the correlation between each predictor (Fig. S5). We find that NO_2 concentration shows a low correlation with the selected variables, which demonstrates that the relationship between NO_2 concentration and the predictors is almost nonlinear. It also reconfirms that MLR performed the worst compared with RF and ANN models.

3.2.1. Feature importance

Given the higher accuracy of RF-LUR models, we explored the selected feature importance shown in Fig. 6. The selected features from conventional LUR models strongly align with the features in lacunarity-optimized models, which confirms the premise that the lacunarity value could capture the maximum range of spatial heterogeneity. In other words, most selected features are captured within the upper bounded buffer sizes limited by lacunarity values. For example, the flattened value of lacunarity for the building is 34 (refers to 1020m of upper limit) and for roads is 18 (refers to 540m). Therefore, the buffer distances for final selected features in the conventional LUR model in the Bronx for buildings and roads are all less than 1020m (e.g., LPI_Buildings_300, Largest Patch Index of buildings calculated within 300m buffer size) and 540m (e.g., LSI_Roads_300, Largest Shape Index of roads calculated within 300m buffer sizes). In Oakland, the lacunarity value for the street trees is 45 (refers to 450m). Therefore, the maximum buffer distances in conventional LUR models for selected SII (Simpson Index) and SHI (Shannon Index) are 500m (e.g., SII_100, SII_500, and SHI_500), which is close to 450m. Those findings are also consistent with roads, water, and green space, suggesting that the lacunarity value could be an essential indicator to extract the upper limit of the spatial heterogeneity of predictors and facilitate reducing the computation burden.

For selected predictors, in the Bronx, the 2-D urban form has the highest feature importance in both lacunarity-optimized (84.60%) and

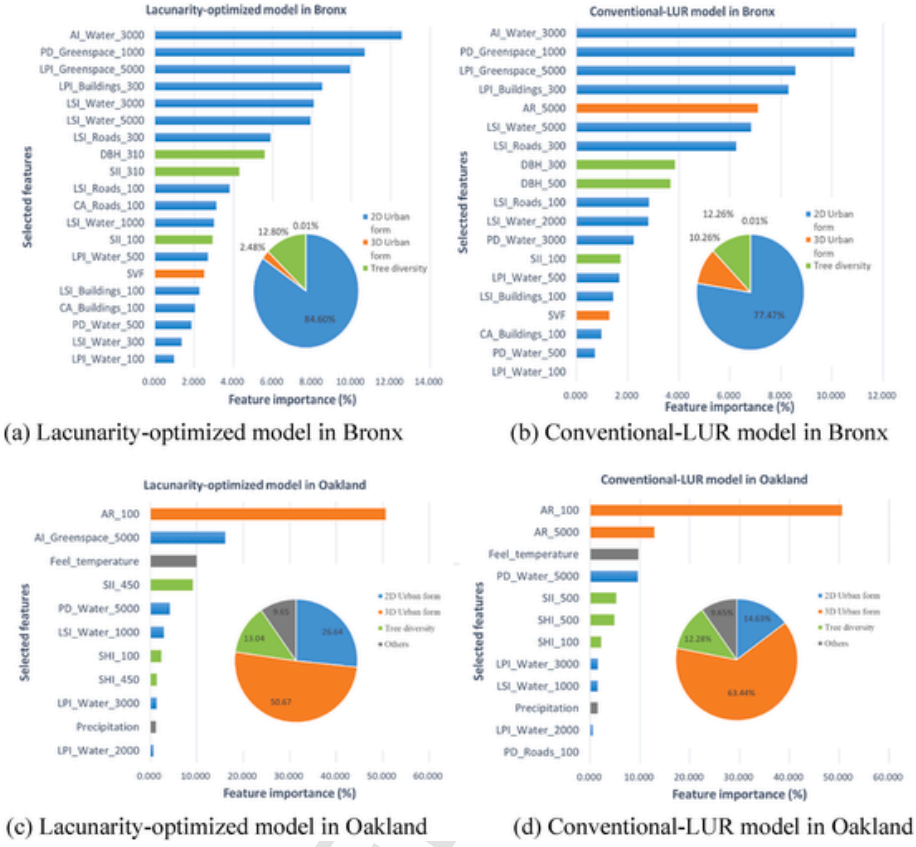

Fig. 6. Feature importance for all selected features in lacunarity-optimized and conventional LUR model in Bronx and Oakland.

conventional LUR (77.47%) models and tree diversity shows higher feature importance in the lacunarity-optimized model (12.80%) than the conventional LUR model (12.26%). In Oakland, 3-D urban form is the most significant feature in both models (50.67% and 63.44% in lacunarity-optimized and conventional LUR models, respectively). Noticeably, tree diversity accounts for similar level of feature importance in the lacunarity-optimized model (13.04%) than in the conventional LUR model (12.28%). Therefore, applying lacunarity value to set the boundary for tree diversity aggregation in lacunarity-optimized models could better reveal the contribution of tree diversity to NO_2 distribution than the conventional LUR models in both Bronx and Oakland. Besides, the aggregation (e.g. AI, Aggregation Index) and dominance (e.g. PD, Patch Density) of water and green space at large buffer distances (e.g., 1000m, 3000m, and 5000m) account for a higher level of feature importance than other features in both models in the Bronx, which could be attributed to the fact that urban green space has the potential to mitigate air pollution in high-density urban areas (Neft et al., 2016). In Oakland, the AR (Aspect Ratio), SII, and SHI show higher feature importance than the green space and water, suggesting the essential synthetic effects between urban form and tree diversity on the distribution of NO_2 concentration. Moreover, the fragmentation degree of buildings (e.g., LPI_Building_300 in the Bronx) and building-street organization (e.g., AR_100 in Oakland) are also associated with heat transfer, air circulation, and consequently the diffusion of pollutants in urban environments (Liu et al., 2018a). The different impacts of urban form on NO_2 concentration between the Bronx and Oakland could be explained by their different city sizes and local climate. For example, Oakland (202.1 km^2) is almost twice the Bronx (110 km^2). Therefore, they have different levels of energy consumption, production activities, and traffic volume, and the dominant factors could vary for the two cities as well (Sun et al., 2022). For local climate, in the Bronx, the summers are warm, humid, and wet, and the winters are very cold, snowy, and windy with temperature varying from −3 °C to 30 °C (Spark, 2022b). However, Oakland has a Mediterranean climate with long, comfortable, and arid the summers and short, cold, and wet winters with temperatures varying from 7 °C to 24 °C (Spark, 2022a). Different humidity, temperature, and wind situation would also alter the urban form-air quality relationship (Tian et al., 2020b).

3.2.2. SHAP (SHapley additive exPlanations) value

To further interpret the potentially influential directions of the selected features on the prediction results, we plotted the SHAP value for the selected features shown in Fig. 7. In the Bronx, the aggregation and dominance of water (e.g., AI_Water_3000 and LSI_Water_5000) and the roads (e.g., LSI_Roads_300) in both lacunarity-optimized and conventional LUR models show relatively positive associations with NO_2 concentration (higher values distributed on the right side of the X-axis). In contrast, PD_Water_5000 in the lacunarity-optimized model shows negative associations, suggesting that urban watershed distributions at large buffer distances may have non-monotonic effects on NO_2 concentration. Likewise, the dominance of green space (e.g., LPI_Greenspace_5000 in the lacunarity-optimized model and LSI_Greenspace_1000 in the conventional LUR model) is also negatively related to NO_2 concentration. In Oakland, SII_450 (lacunarity-optimized model) SII_500 (conventional LUR model), LSI_Water_1000, and Feel_temperature (both models) show positive correlations, whereas SHI_450 (lacunarity-optimized model) and SHI_500 (conventional LUR model) displays a negative association, which reconfirms that the lacunarity-optimized model could capture the same crucial information as conventional LUR models. However, the LPI_Building_300 in the Bronx and AR_100 in Oakland show non-monotonic effects on NO_2 distribution. In sum, the increase in the dominance of specific tree species (larger SII) is associated with a higher level of NO_2 concentration, while planting more diverse tree species to increase the richness (larger SHI) relates to

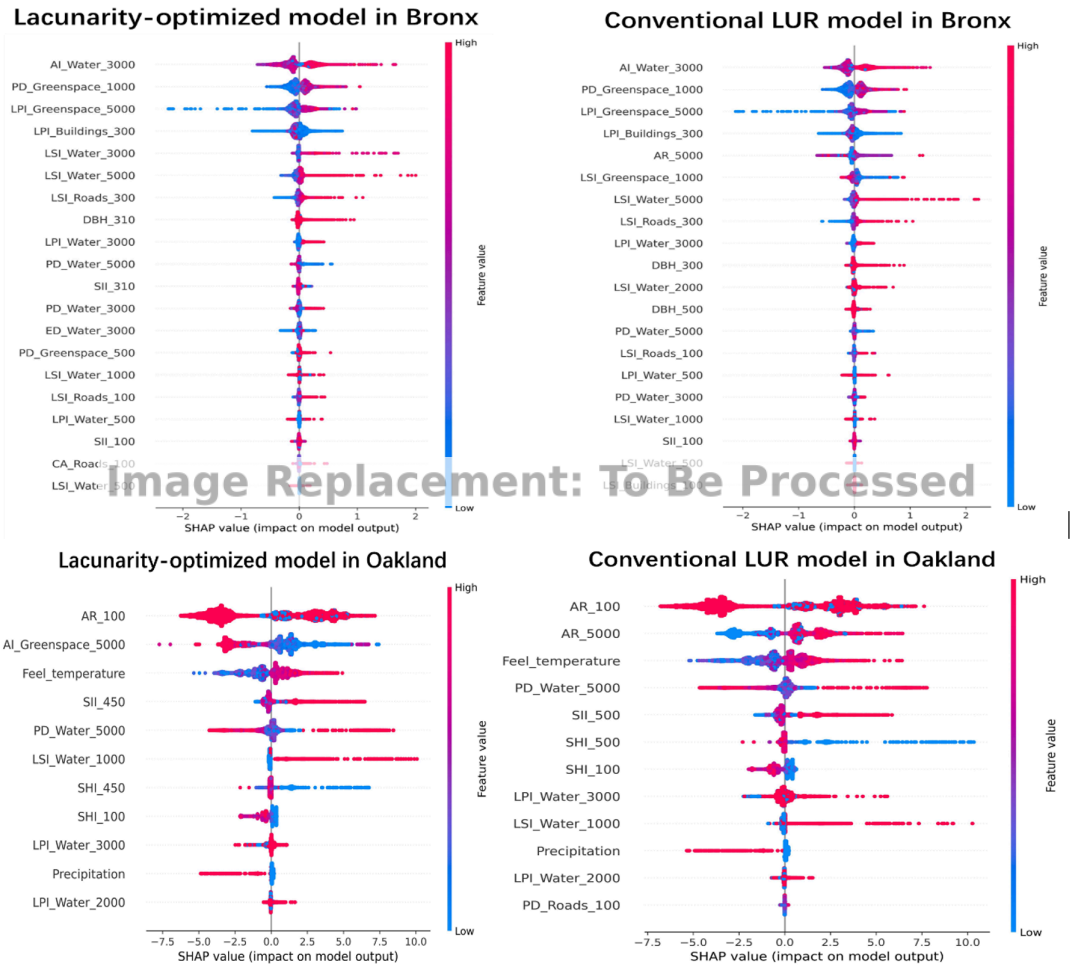

Image Replacement: To Be Processed

Fig. 7. SHAP value for lacunarity-optimized and conventional LUR models in Bronx and Oakland for all selected features.

better air quality. This manifests the multifaced effects of tree diversity on NO_2 concentration as in previous studies. For example, on the one hand, trees could facilitate the widespread deposition of various gases and particles and influence microclimate and air turbulence to improve air quality (Neft et al., 2016). On the other hand, trees could produce wind-dispersed pollen and emit a range of gaseous substances (e.g., NOx) that take part in photochemical reactions to deteriorating air quality (Grote et al., 2016).

3.2.3. PDP (Partial Dependence Plot) distribution

As tree diversity shows a non-monatomic correlation on NO_2 concentration, it is necessary to investigate how and to what extent tree diversity relates to NO_2 concentration on the city scale. In Fig. 8, we plotted the PDP by selecting tree diversity indices with higher feature importance (SII_100 and SII_310 in the Bronx and SII_450 and SHI_450 in Oakland based on lacunarity-optimized models) to explore the marginal effects between tree diversity and NO_2 concentration. In the Bronx, the NO_2 concentration is relatively lower when SII_100 is lower than 0.80, and the SII_310 ranges between 0.82 and 0.92 (purple clusters). In addition, the increase of SII_100 is associated with lower NO_2 concentration, whereas the SII_310 has complicated impacts. Too high (>0.92) or too low (<0.79) of SII_310 would both relate to the higher NO_2 concentration (>9.91 ppb). In Oakland, the NO_2 concentration is relatively lower when SHI_450 is higher than 0.70 and SII_450 is lower than 0.95. The impacts of SHI_450 and SII_450 themselves on NO_2 concentration are almost monotonic negative (values change homogeneously from high to low from a bottom-up perspective) and positive (values change homogeneously from low to high from a left-right perspective), respectively, which accords with the findings in Fig. 7. Therefore, the level of tree diversity (both SII and SHI) needs to be controlled within limited ranges to improve the air quality effectively instead of going to extremes by planting more diverse tree species or simply one specific species on the city scale. We also find that SVF, AR, and SII have evident synthetic effects on the NO_2 concentration in the Bronx according to the PDP between the tree diversity, 3-D urban form, and NO_2 concentration in Fig. S6.

3.2.4. Prediction comparison

To compare the prediction results between lacunarity-optimized and conventional LUR models, we created the fishnet with the grid size of 510m (17 × 30m) and 540m (18× 30m), the lacunarity value of "Overall landuse" in Fig. 3, in the Bronx and Oakland, respectively (Fig. 9). The meteorological factors were input as the average value during the collection period for Bronx (September 10th to December 17th, 2021) and Oakland (May 28th, 2015 to December 21st, 2017), respectively. In the Bronx, the spatial distribution of NO_2 concentration for the two models is similar with higher values located in the southern and eastern parts. Besides, the prediction results from conventional LUR models show an overall higher NO_2 concentration (mean value of 10.37 ppb) and larger variance (standard deviation of 0.34 ppb) than the lacunarity-optimized (mean value of 10.15 ppb and standard deviation of 0.14 ppb) model. In Oakland, higher NO_2 concentration is almost distributed in the southern part for the two models. Besides, the lacunarity-optimized model has a higher average value (22.10 ppb vs.

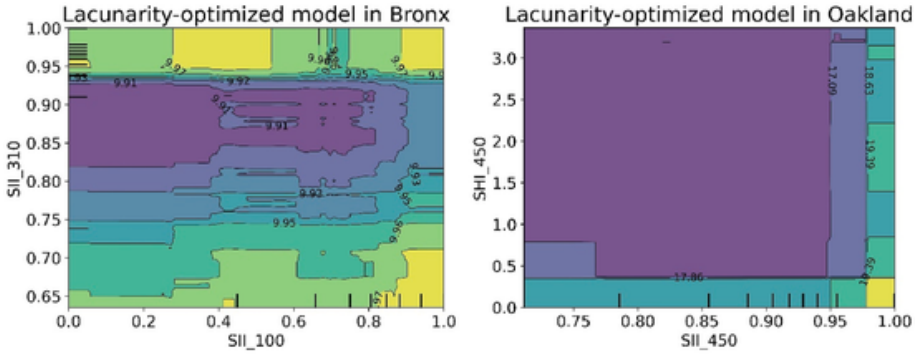

Fig. 8. PDP in Bronx and Oakland based on the lacunarity-optimized model
(Values in PDP figures indicate the level of NO_2 concentrations. Low to high NO_2 concentrations indicate the gradual change of colors from purple to green and yellow. To better explore the marginal effects between tree diversity indices on NO_2 concentration, we need to single out the tree diversity indices with higher feature importance. According to Fig. 6, in the Bronx, only SII_100 (1.72%) has been selected in conventional LUR models. However, SII_100 and SII_310 account for 5.35% of feature importance in total in the lacunarity-optimized model. Therefore, we choose SII_100 and SII_310 in the Bronx through the lacunarity-optimized model to plot PDP. In Oakland, SII_450 and SHI_450 account for 10.17% and 10.58% in total for the conventional and lacunarity-optimized models, respectively. Therefore, we select SII_450 and SHI_450 in the lacunarity-optimized model to plot PDP in Oakland. The DBH is not considered here because no relevant DBH parameters have been selected in Oakland for both models.).

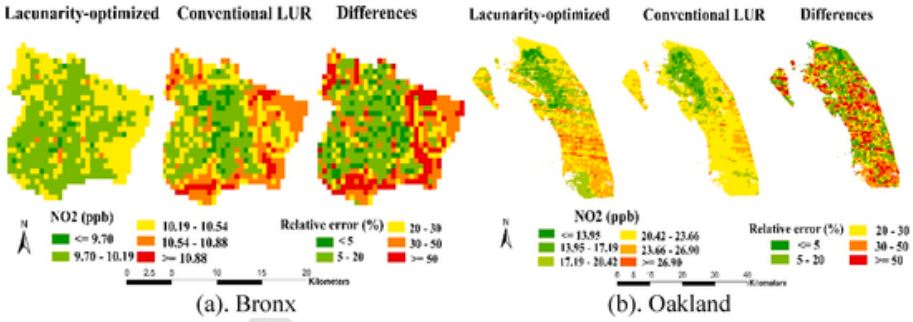

Fig. 9. Predicted NO_2 concentration based on lacunarity-optimized and conventional LUR models and their relative differences in Bronx (a) and Oakland (b).

21.84 ppb) and lower variation (3.23 ppb vs. 2.60 ppb) than the conventional LUR model.

In addition, the locations with higher NO_2 concentrations have larger relative errors ($\geq$ 50%) between the two models in both Bronx and Oakland, indicating that the model may not estimate the higher level of NO_2 concentration reliably partly due to the transported pollutants from nearby emission sources. Therefore, future models could incorporate more accurate and detailed NOx emission inventories to reduce the bias in these hotspot regions. Besides, the absolute prediction error is lower in the Bronx (from −0.50 to 1.03 ppb) than that in Oakland (−8.99 to 12.71 ppb), which could be attributed to the fact that urban form-air quality relationship is also modulated by other factors, such as seasonality (Liu et al., 2018c), city sizes (Fan et al., 2017), and development stages (Liang & Gong, 2020) among different cities.

4. Discussion

In this study, we incorporate tree diversity and 2-D and 3-D urban form metrics to capture the spatiotemporal change of NO_2 concentration in both Bronx, NY, and Oakland CA in the U.S. through mobile monitoring. As urban trees have non-monotonic impacts on the change of air pollution concentration, this study tries to combine both urban forest and morphology quantitatively to further explore their correlation.

Even though deciduous trees are surrounded by a higher level of NO_2 concentration than semi-deciduous and evergreen trees in both Bronx and Oakland, the influential direction of tree diversity (Fig. 6) indicates the diversified tree species are associated with an overall lower NO_2 concentration. Notwithstanding, the level of tree diversity should not be unlimited high. According to the non-monotonic and marginal effects of tree diversity on NO_2 distribution (Fig. 7), the diversity level of tree species should be controlled in limited ranges to effectively control air pollution. Therefore, it is promising to plant more semi-deciduous or evergreen trees to reduce air pollution regionally, but also necessary to strike a balance between the diversity and dominance of tree species to improve air quality on the city scale. The findings could help improve the urban air quality and urban ecological health for a better realization of the One Health initiative.

We also optimized the model construction process by introducing the lacunarity analysis and considering the maximum range of spatial heterogeneity for different land use and distribution of tree species. The final results demonstrate that lacunarity-optimized models could reduce the computation burden by extracting the upper limits of the spatial heterogeneity of predictors while keeping the model accuracy simultaneously (Fig. 5). This illustrates that it would be unnecessary to include as many buffer distances as possible because increasing the buffer distance beyond a certain level for variable aggregation would not result in any substantive impact on the statistical analysis due to the homogenous urban context as the buffer distance surpasses that of the repeating patterns in the landscape (Hoechstetter et al., 2011; Myint & Lam, 2005; Roces-Diaz et al., 2015). Our lacunarity approach can efficiently determine land use feature-specific upper bounds of buffer sizes, therefore, saving computational resources for extracting information in large areas while capturing the same crucial information and keeping a similar level of accuracy as the conventional way simultaneously.

This study also reveals that tree diversity shows higher feature importance in the lacunarity-optimized model (12.80% and 13.04% in Bronx and Oakland, respectively) than the conventional LUR model

(12.60% and 12.28% in Bronx and Oakland, respectively). It suggests, on the one hand, that urban forests play a non-ignorable role in NO_2 concentration distribution; on the other hand, quantifying the maximum range of spatial heterogeneity of street trees could better reveal their potential influence on the NO_2 concentration. Therefore, urban planners and ecologies need to work together to customize the positions for tree planting under different land use patterns.

We acknowledge several limitations in our study. First of all, the current study reflects the association between tree diversity, urban form, and NO_2 concentration, but the causality could be further analyzed through Rubin or Pearl's structural casual models with modulators, such as traffic intensity and collection periods. Therefore, future studies could classify the mobile data by the TAZ (Traffic Analysis Zone) or the measurement time (e.g., weekdays and weekends or day-night division) (Montazeri et al., 2021). Secondly, the upper bounds of buffer distances obtained from lacunarity analysis could be further validated by analyzing the influential range of emission sources through deterministic models, such as Gaussian plume dispersion (Gibson et al., 2013). Thirdly, to represent air quality comprehensively, more types of air pollutants (e.g., $PM_{2.5}$, PM_{10}, and O_3) need to be added to represent the urban emission inventories. In addition, other predictors, such as Frontal Area Index, Leaf Area Index, Crown Volume Fraction, tree porosity, tree spacing, and phenology should be incorporated to better reflect the effects of the urban forest and physical environments on air quality. Fourthly, more techniques, such as the semi-variogram (Zhou & Lin, 2019), can be experimented with to address the spatial dependence of mobile monitoring observation points to reduce the bias in regression analysis. Finally, during model construction, we applied a series of fixed buffer distances from 100m to 5000m. However, previous studies found that larger particles were more sensitive to wind factors (Tian et al., 2022). Thus, it deserves future research endeavors to create adaptive buffer distances for more fine-tuned models to explore the synthetic effects of local meteorology and land uses on NO_2 distributions.

5. Conclusions

It is promising to leverage opportunistic mobile monitoring and investigate the associations between a wide range of tree diversity, urban form, and NO_2 concentration distribution on hyperlocal scales. The results further explained the rationale of the non-monotonic effects of urban trees on NO_2 concentration by finding that it is necessary to seek a balance between tree diversity and tree dominance to overall improve air quality on the city scale. Besides, the lacunarity-optimized method we proposed could quantify the spatial heterogeneity of predictors for feature extraction objectively, as they do not rely on predefined landscape units or patches and thus can effectively deal with the complexity of the urban environment with comparable model performance with conventional LUR models. Moreover, quantifying the ranges of spatial heterogeneity of street trees during model construction could better reveal the contribution of tree diversity to NO_2 distribution in both the Bronx and Oakland. Finally, there are synthetic effects between the urban form and tree diversity on NO_2 distribution, and such effect directions could be non-monotonic as well, which reconfirms the non-linear interactions between urban form, tree diversity, and air pollution distribution. Therefore, future studies could explore the detailed mechanism of how urban form and tree diversity impact NO_2 generation, diffusion, and purification in urban settings.

Declaration of competing interest

All co-authors have seen and agree with the contents of the manuscript and there is no conflict of interest to report.

Acknowledgment

The authors thank all other members of the MIT Senseable City Lab Consortium (FAE Technology, MipMap, Samoo Architects & Engineers, GoAigua, DAR Group, Ordinance Survey, RATP, Anas S.p.A., ENEL Foundation, Università di Pisa, KTH (Sweden), ITB (Indonesia), UTEC (Peru), Politecnico di Torino, SMART (Singapore), AMS Institute, and the cities of Laval, Curitiba, Stockholm, Amsterdam, Helsingborg) for funding this research, and special thanks to Joshua Apte and his group's work to provide NO_2 data for Oakland.

The authors thank the New York City Mayor's Office of the Chief Technology Officer, the Department of Health and Mental Hygiene, the Department of Citywide Administrative Services, as well as the EDF - Environmental Defense Fund, for providing the sensing fleet and providing feedback on the research work.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.apgeog.2023.102943.

References

Apte, J.S., Messier, K.P., Gani, S., et al. (2017). *High-resolution air pollution mapping with Google street view cars: Exploiting big data. Environmental Science & Technology*, *51*(12), 6999–7008.

Brantley, H., Hagler, G., Kimbrough, E., et al. (2014). *Mobile air monitoring data-processing strategies and effects on spatial air pollution trends. Atmospheric Measurement Techniques*, *7*(7), 2169–2183.

CDC. *One health* Available at: https://www.cdc.gov/onehealth/index.html. 2022.

Cho, H.-S., & Choi, M. (2014). *Effects of compact urban development on air pollution: Empirical evidence from Korea. Sustainability*, *6*(9), 5968–5982.

Cowie, C.T., Garden, F., Jegasothy, E., et al. (2019). *Comparison of model estimates from an intra-city land use regression model with a national satellite-LUR and a regional Bayesian Maximum Entropy model, in estimating NO2 for a birth cohort in Sydney, Australia. Environmental Research*, *174*, 24–34.

Cummings, L.E., Stewart, J.D., Reist, R., et al. (2021). *Mobile monitoring of air pollution reveals spatial and temporal variation in an urban landscape. Frontiers in Built Environment*, *7*, 648620.

DeSouza, P., Anjomshoaa, A., Duarte, F., et al. (2020). *Air quality monitoring using mobile low-cost sensors mounted on trash-trucks: Methods development and lessons learned. Sustainable Cities and Society*, *60*, 102239.

Do, T.H., Tsiligianni, E., Qin, X., et al. (2020). *Graph-deep-learning-based inference of fine-grained air quality from mobile IoT sensors. IEEE Internet of Things Journal*, *7*(9), 8943–8955.

Edussuriya, P., Chan, A., & Ye, A. (2011). *Urban morphology and air quality in dense residential environments in Hong Kong. Part I: District-level analysis. Atmospheric Environment*, *45*(27), 4789–4803.

Fang, Y., & Zhao, L. (2022). *Assessing the environmental benefits of urban ventilation corridors: A case study in hefei*. China: Building and Environment, 108810.

Fan, C., Tian, L., & Li, J. (2017). *Research progress of impacts of urban form on air quality*: *Vol. 12*. Urban Development.

Foundation, A.D. *Northern red OakQuercus rubra* Available at: https://www.arborday.org/trees/treeguide/TreeDetail.cfm?Itemid=2451. 2022.

Friedman, J.H. (2001). *Greedy function approximation: A gradient boosting machine. Annals of Statistics*, 1189–1232.

Galle, N.J., Halpern, D., Nitoslawski, S., et al. (2021). *Mapping the diversity of street tree inventories across eight cities internationally using open data. Urban Forestry and Urban Greening*, *61*, 127099.

Gibson, M.D., Kundu, S., & Satish, M. (2013). *Dispersion model evaluation of PM2. 5, NOx and SO2 from point and major line sources in Nova Scotia, Canada using AERMOD Gaussian plume air dispersion model. Atmospheric Pollution Research*, *4*(2), 157–167.

Goin, D.E., Sudat, S., Riddell, C., et al. (2021). *Hyperlocalized measures of air pollution and preeclampsia in Oakland, California. Environmental Science & Technology*, *55*(21), 14710–14719.

Gorai, A.K., Tuluri, F., & Tchounwou, P.B. (2014). *A GIS based approach for assessing the association between air pollution and asthma in New York State, USA. International Journal of Environmental Research and Public Health*, *11*(5), 4845–4869.

Grimmond, C., & Oke, T.R. (1999). *Aerodynamic properties of urban areas derived from analysis of surface form. Journal of Applied Meteorology and Climatology*, *38*(9), 1262–1292.

Grote, R., Samson, R., Alonso, R., et al. (2016). *Functional traits of urban trees: Air pollution mitigation potential. Frontiers in Ecology and the Environment*, *14*(10), 543–550.

Hang, J., Li, Y., Sandberg, M., et al. (2012). *The influence of building height variability on pollutant dispersion and pedestrian ventilation in idealized high-rise urban areas. Building and Environment*, *56*, 346–360.

Hoechstetter, S., Walz, U., & Thinh, N.X. (2011). *Adapting lacunarity techniques for gradient-based analyses of landscape surfaces. Ecological Complexity*, *8*(3), 229–238.

Ke, B., Hu, W., Huang, D., et al. (2022). *Three-dimensional building morphology impacts on*

CORRECTED PROOF

PM2. 5 distribution in urban landscape settings in Zhejiang, China. Science of the Total Environment, 826, 154094.
Kerckhoffs, J., Khan, J., Hoek, G., et al. (2022). *Mixed-effects modeling framework for Amsterdam and copenhagen for outdoor NO2 concentrations using measurements sampled with Google street View cars*. Environmental science & technology.
Kheirbek, I., Haney, J., Douglas, S., et al. (2016). *The contribution of motor vehicle emissions to ambient fine particulate matter public health impacts in New York city: A health burden assessment. Environmental Health, 15*(1), 89.
Kwan, M.-P. (2012). *The uncertain geographic context problem. Annals of the Association of American Geographers, 102*(5), 958–968.
Labib, S., Lindley, S., & Huck, J.J. (2020). *Scale effects in remotely sensed greenspace metrics and how to mitigate them for environmental health exposure assessment. Computers, Environment and Urban Systems, 82*, 101501.
Liang, L., & Gong, P. (2020). *Urban and air pollution: A multi-city study of long-term effects of urban landscape patterns on air quality trends. Scientific Reports, 10*(1), 1–13.
Liu, Y., Fang, X., Cheng, C., et al. (2016). *Research and application of city ventilation assessments based on satellite data and GIS technology: A case study of the yanqi lake eco-city in huairou district, beijing. Meteorological Applications, 23*(2), 320–327.
Liu, X., Hu, G., Chen, Y., et al. (2018a). *High-resolution multi-temporal mapping of global urban land using Landsat images based on the Google Earth Engine Platform. Remote Sensing of Environment, 209*, 227–239.
Liu, Y., Wu, J., Yu, D., et al. (2018b). *The relationship between urban form and air pollution depends on seasonality and city size. Environmental Science and Pollution Research.* https://doi.org/10.1007/s11356-018-1743-6. 1-14.
Liu, Y., Wu, J., Yu, D., et al. (2018c). *The relationship between urban form and air pollution depends on seasonality and city size. Environmental Science and Pollution Research, 25*(16), 15554–15567.
Lundberg, S.M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Proceedings of the 31st international conference on neural information processing systems* (pp. 4768–4777).
Mandelbrot, B.B. (1982). *The fractal geometry of nature*. New York: WH freeman.
Manes, F., Silli, V., Salvatori, E., et al. (2014). *Urban ecosystem services: Tree diversity and stability of PM10 removal in the metropolitan area of Rome. Annali di Botanica, 4*, 19–26.
McGarigal, K. (2015). *FRAGSTATS help*.
Messier, K.P., Chambliss, S.E., Gani, S., et al. (2018). *Mapping air pollution with Google Street View cars: Efficient approaches with mobile monitoring and land use regression. Environmental Science & Technology, 52*(21), 12563–12572.
Montazeri, A., Lilienthal, A.J., & Albertson, J.D. (2021). *A spatial land use clustering framework for investigating the role of land use in mediating the effect of meteorology on urban air quality. Atmospheric Environment X, 12*, 100126.
Mora, S., Anjomshoaa, A., Benson, T., et al. (2019). *Towards large-scale drive-by sensing with Multi-Purpose City scanner nodes*. 2019 IEEE 5th World Forum on Internet of Things (WF-IoT). IEEE.
Myint, S.W., & Lam, N. (2005). *A study of lacunarity-based texture analysis approaches to improve urban image classification. Computers, Environment and Urban Systems, 29*(5), 501–523.
Naughton, O., Donnelly, A., Nolan, P., et al. (2018). *A land use regression model for explaining spatial variation in air pollution levels using a wind sector based approach. Science of the Total Environment, 630*, 1324–1334.
Neft, I., Scungio, M., Culver, N., et al. (2016). *Simulations of aerosol filtration by vegetation: Validation of existing models with available lab data and application to near-roadway scenario. Aerosol Science and Technology, 50*(9), 937–946.
Peng, Y., Gao, Z., Buccolieri, R., et al. (2021). *Urban ventilation of typical residential streets and impact of building form variation. Sustainable Cities and Society, 67*, 102735.
Perrier, A. *Feature importance in random forests* Available at: https://alexisperrier.com/datascience/2015/08/27/feature-importance-random-forests-gini-accuracy.html. 2015.
Plotnick, R.E., Gardner, R.H., Hargrove, W.W., et al. (1996). *Lacunarity analysis: A general technique for the analysis of spatial patterns. Physical Review, 53*(5), 5461.
Plotnick, R.E., Gardner, R.H., & O'Neill, R.V. (1993). *Lacunarity indices as measures of landscape texture. Landscape Ecology, 8*(3), 201–211.
Pugh, T.A., MacKenzie, A.R., Whyatt, J.D., et al. (2012). *Effectiveness of green infrastructure for improvement of air quality in urban street canyons. Environmental Science & Technology, 46*(14), 7692–7699.
Roces-Diaz, J.V., Diaz-Varela, R.A., Alvarez-Alvarez, P., et al. (2015). *A multiscale analysis of ecosystem services supply in the NW Iberian Peninsula from a functional perspective. Ecological Indicators, 50*, 24–34.
Rodríguez, M.C., Dupont-Courtade, L., & Oueslati, W. (2016). *Air pollution and urban structure linkages: Evidence from European cities. Renewable and Sustainable Energy Reviews, 53*, 1–9.
Roeland, S., Moretti, M., Amorim, J.H., et al. (2019). *Towards an integrative approach to evaluate the environmental ecosystem services provided by urban forest. Journal of Forestry Research, 30*(6), 1981–1996.
Saha, P.K., Li, H.Z., Apte, J.S., et al. (2019). *Urban ultrafine particle exposure assessment with land-use regression: Influence of sampling strategy. Environmental Science & Technology, 53*(13), 7326–7336.
Saunders, S., Dade, E., & Van Niel, K. (2011). An Urban Forest Effects (UFORE) model study of the integrated effects of vegetation on local air pollution in the Western Suburbs of Perth, WA. *19th International congress on modelling and simulation (MODSIM2011)*. Perth, Australia: Modelling and Simulation Society of Australia and New Zealand Inc. 19th International congress on modelling and simulation (MODSIM2011).
Shannon, C.E. (1948). *A mathematical theory of communication. The Bell system technical journal, 27*(3), 379–423.
Shen, J., Gao, Z., Ding, W., et al. (2017). *An investigation on the effect of street morphology to ambient air quality using six real-world cases. Atmospheric Environment, 164*, 85–101.
Sicard, P., Agathokleous, E., Araminiene, V., et al. (2018). *Should we see urban trees as effective solutions to reduce increasing ozone levels in cities? Environmental Pollution, 243*, 163–176.
Simpson, E.H. (1949). *Measurement of diversity. Nature, 163*(4148). 688-688.
Spark, W. *Climate and average weather year round in Oakland* Available at: https://weatherspark.com/y/541/Average-Weather-in-Oakland-California-United-States-Year-Round. 2022a.
Spark, W. *Climate and average weather year round in the Bronx* Available at: https://weatherspark.com/y/24500/Average-Weather-in-The-Bronx-New-York-United-States-Year-Round. 2022b.
Sun, J., Zhou, T., & Wang, D. (2022). *Relationships between urban form and air quality: A reconsideration based on evidence from China's five urban agglomerations during the COVID-19 pandemic. Land Use Policy, 118*, 106155.
Tang, R., Blangiardo, M., & Gulliver, J. (2013). *Using building heights and street configuration to enhance intraurban PM10, NOx, and NO2 Land use regression models. Environmental Science & Technology, 47*(20), 11643–11650.
Tian, Y., deSouza, P., Mora, S., et al. (2022). *Evaluating the meteorological effects on the urban form–air quality relationship using mobile monitoring*. Environmental science & technology.
Tian, Y., & Yao, X. (2022). *Urban form, traffic volume, and air quality: A spatiotemporal stratified approach. Environment and Planning B: Urban Analytics and City Science, 49*(1), 92–113.
Tian, Y., Yao, X.A., Mu, L., et al. (2020a). *Integrating meteorological factors for better understanding of the urban form-air quality relationship. Landscape Ecology, 35*(10), 2357–2373.
Tian, Y., Yao, X.A., Mu, L., et al. (2020b). *Integrating meteorological factors for better understanding of the urban form-air quality relationship. Landscape Ecology, 35*, 2357–2373.
Tiwari, A., & Kumar, P. (2020). *Integrated dispersion-deposition modelling for air pollutant reduction via green infrastructure at an urban scale. Science of the Total Environment, 723*, 138078.
Tong, Z., Baldauf, R.W., Isakov, V., et al. (2016). *Roadside vegetation barrier designs to mitigate near-road air pollution impacts. Science of the Total Environment, 541*, 920–927.
Yuan, M., Song, Y., Huang, Y., et al. (2019). *Exploring the association between the built environment and remotely sensed PM2. 5 concentrations in urban areas. Journal of Cleaner Production, 220*, 1014–1023.
Zakšek, K., Oštir, K., & Ž, K. (2011). *Sky-view factor as a relief visualization technique. Remote Sensing, 3*(2), 398–415.
Zhang, A., Xia, C., & Li, W. (2022). *Exploring the effects of 3D urban form on urban air quality: Evidence from fifteen megacities in China. Sustainable Cities and Society, 78*, 103649.
Zhong, J., Cai, X.-M., & Bloss, W.J. (2016). *Coupling dynamics and chemistry in the air pollution modelling of street canyons: A review. Environmental Pollution, 214*, 690–704.
Zhou, S., & Lin, R. (2019). *Spatial-temporal heterogeneity of air pollution: The relationship between built environment and on-road PM2. 5 at micro scale. Transportation Research Part D: Transport and Environment, 76*, 305–322.

CORRECTED PROOF