URBAN FORM AND AIR QUALITY: A FINE-SCALE LOOK AT 22 MAJOR U.S. CITIES

by

Evan Blomquist

A thesis submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Master of Science in
Earth Science

Charlotte

2023

Approved by:

_____
Dr. Gang Chen

_____
Dr. Matthew Eastin

_____
Dr. Jean-Claude Thill

ABSTRACT

EVAN BLOMQUIST. Urban Form and Air Quality: A Fine-Scale Look at 22 Major U.S. Cities.
(Under the direction of DR. GANG CHEN)

Common pollutants present in cities, such as nitrogen dioxide (NO_2), can cause a variety of adverse human health outcomes. Urban form (e.g., land cover and its spatial patterns) is a crucial factor contributing to air quality. There have been efforts investigating the effects of urban form on air quality. However, urban form is typically quantified at coarse resolution (e.g., 30 m) that is too broad to capture the spatial patterns of land cover in highly heterogeneous urban environments. Moreover, most of the efforts are limited to one single or only a few cities, leading to inconsistent or even contradictory findings among the studies. This project proposes utilizing 1-m resolution land cover maps to accurately capture fine-scale urban patterns in 22 major cities across the U.S., investigate their relationship with NO_2, and analyze the relationship variation across seasons and urban neighborhoods over diverse geographic regions in the country. The study bridges a gap in understanding the effects of highly heterogeneous urban land cover and its spatial patterns on air quality and can help urban planners and practitioners make informed decisions in the development of smart and sustainable cities.

Keywords: Air Quality, Urban Form, Fine-scale, Multi-city, Nitrogen Dioxide (NO_2), Remote Sensing, Sustainability

ACKNOWLEDGEMENTS

I would like to acknowledge the following individuals for their invaluable contributions to this research. First and foremost, I would like to thank my master's advisor, Dr. Gang Chen, for providing me with guidance and support throughout the project. I am also grateful to my committee members, Dr. Jean-Claude Thill and Dr. Matthew Eastin, for their insightful feedback and critical evaluation of my work. In addition, I extend my gratitude to Shitij Govil, who provided me with assistance on data retrieval and calculations. Without his help, this research would not have been possible. I would also like to thank the Geography and Earth Sciences department at UNC Charlotte for providing resources and support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. Introduction

Urban areas accommodate over 56% (4.4 billion) of the world's population and are expected to add another 2.2 billion people by 2050 (United Nations, 2023). The urban atmosphere is a "giant chemical reactor" (Seinfeld, 1989), which contains a variety of harmful pollutants such as nitrogen dioxide (NO_2) and fine particulate matter 2.5 (PM 2.5 – particles less than 2.5 micrometers in diameter). World Health Organization (WHO) estimates that more than 80% of the world's population currently residing in cities are exposed to air quality levels that exceed guidelines (WHO, 2022). In all types of ecosystems throughout the world, urban areas produce about 78% of carbon and airborne pollutants. This adversely affects roughly 50% of the world's population (Liang & Gong, 2020). NO_2 is one of the most common and severe pollutants and can cause problems with lung function/development and cardiovascular systems (Gulia et al., 2015; Siddiqui et al., 2020). It has been shown that long-term exposure to NO_2 can increase the risk of mortality from cardiovascular diseases by 11%, respiratory diseases by 3%, and cancer by 2% (Eum et al, 2019). There has also been evidence to suggest that people of lower income or socio-economic status are associated with an increased risk of exposure to this pollutant along with other traffic related air impurities, and that children living in households with increased traffic density 200 m from their home are more likely to develop lung related issues such as asthma (Cakmak et al., 2016).

A wealth of studies have investigated individual key factors (e.g., transportation and weather) affecting air pollution or more specifically NO_2 concentrations (Ikram et al., 2015; Salas et al., 2021; Yang et al., 2021); however, sustainable and smart urban development necessitates a holistic approach to address this issue. For example, a busy highway increases pollution, although it can be reduced by nearby urban forests (Roeland et al., 2019). The spatial

patterns of buildings vary from one city/neighborhood to another, which could result in significant differences in road networks, traffic, and hence, air pollution levels (McCarty and Kaza, 2015). From the lenses of urban planning, air pollution is directly or indirectly affected by urban physical elements (e.g., roads, buildings and trees) and their spatial arrangements/patterns, which together are referred to as urban form. Aiming to support informed decision-making in urban sustainable development, there have been growing efforts exploring the relationship between urban form and air pollution (e.g., Bechle et al. 2017; Ku, 2020; Kang et al., 2019; Li and Zhou, 2019). For example, Bechle et al. (2017) studied 1274 urban areas globally using satellite estimates. They found that contiguity, vegetation, and urban morphology had statistically significant relationships with NO_2 concentrations, suggesting that properly designed urban form can reduce pollution exposure. However, they also found that climate, city population, and country-level income significantly impacted NO_2 concentrations.

While urban form was found to have an evident impact on NO_2 levels, low or medium-resolution satellite sensors, such as 30 m Landsat and 500 m MODIS, were typically used to delineate and analyze urban land cover and spatial patterns (e.g., Chen et al., 2013; Liang et al., 2020). The reality is that urban areas, especially those intensively developed big cities, are of high spatial heterogeneity and fragmentation. Relatively small geographic entities (e.g., street trees and neighborhood roads) tend to be ignored in data of a coarse resolution, from which the spatial patterns of land cover are likely to be over-generalized. See an example in Figure 1, which shows a comparison of land cover at 1 m versus 30 m resolution over the same urban region. As argued by Bereitschaft & Debbage (2013), fragmentation and continuity play a major role affecting air pollution, but using a coarse scale (e.g., 30 m) cannot accurately identify the impact. We note that fine-scale urban form has recently been introduced to studying air

pollution, such as Urban Atlas (2-4 m resolution) used by Ajtai et al. (2020). The challenge is that most of those efforts were limited. Case and regional studies are highly beneficial, however, it leads to a concern about how generalizable the findings are. Some of the findings are not consistent, or even contradict each other. For example, Li and Zhou (2018) found that low urban compactness had a significant correlation with poor air quality, while low urban compactness was found to decrease non-point source emissions in studies by Fan et al. (2018) and Lu and Liu (2016). Different from their findings, Bechle et al. (2011) observed that compactness was not a significant predictor of NO_2 concentration, and it was reconfirmed in Bechle et al. (2017).

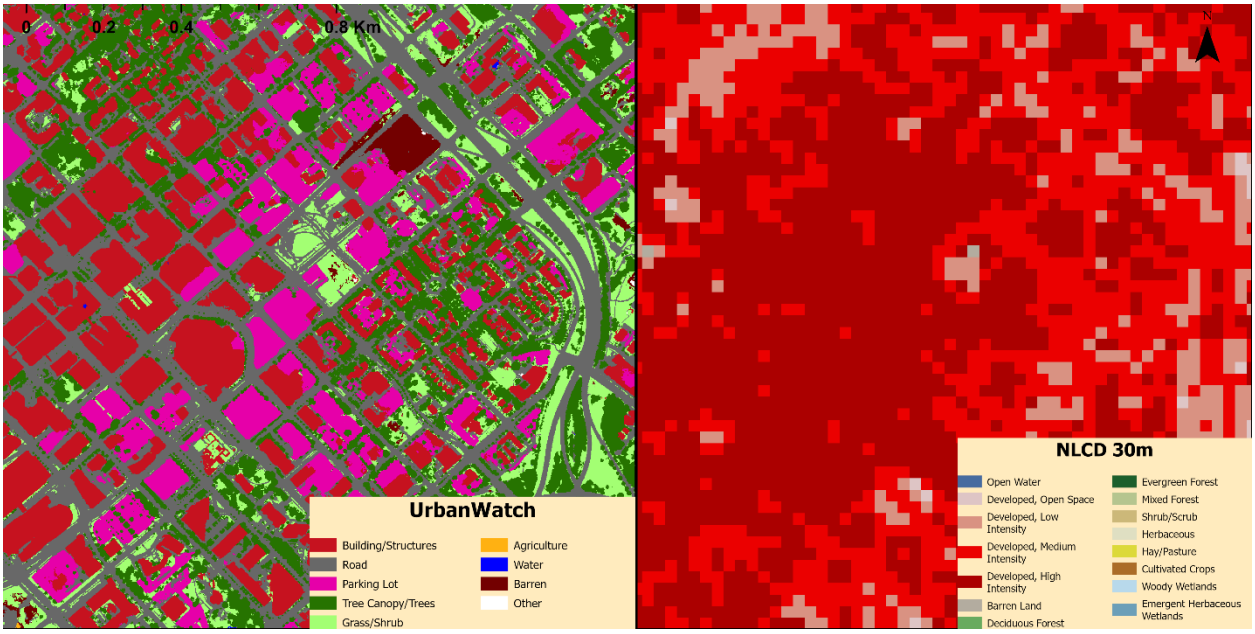

Figure 1: Different urban land cover and spatial patterns represented in 1-meter [from UrbanWatch develop by Zhang et al. (2022)] versus 30-meter land cover (from National Land Cover Database – NLCD developed by the Multi-Resolution Land Characteristics Consortium) over the same urban region in Charlotte, North Carolina, United States

Based on the above considerations, this study aims to investigate the relationship between fine-scale urban form (1 m resolution) and air quality (with an emphasis on NO_2). To ensure a strong generalization ability of our findings to inform decision-making over broad regions, we

captured 1 m resolution urban form over multiple cities, i.e., 22 major cities across the U.S. Through statistical modeling using Boosted Regression Trees (BRT), we would like to answer three specific questions in this communication: (i) *Is the relationship between fine-scale urban form and NO_2 concentration consistent across cities?* Urban form has high variation in land cover and spatial patterns. Here we used a range of variables to quantify various aspects of urban form, and evaluated how consistent the relationships are between those variables and air quality. (ii) *How is the relationship affected by the spatial extent (i.e., buffer zone) at which urban form is captured?* Urban land cover patterns are highly variable when neighborhoods are observed within buffer zones of different sizes. It could cause inconsistent results of the studied relationship. Through this multi-city study, we evaluated whether buffer zone plays a consistent role in understanding such a relationship. (iii) *How does the change in season affect the relationship between fine-scale urban form and NO_2 concentration?* Season is known to affect air quality (e.g., Jayamurugan et al., 2013; Ikram et al., 2015). For example, increased emissions from heating and transport, as well as the cold and dense air/lack of sunlight in the winter tend to increase concentrations of NO_2 and lessen its diurnal variation (Boersma et al., 2009; Roberts-Semple, Song, and Gao, 2012; Kendrick et al., 2015). At the fine scale, however, seasonal effect may be amplified or curtailed to variably affect microclimate and pollutant concentration. Our study is one of the first to address those questions through a fine-scale look at the urban form across a number of U.S. cities that have a broad and diverse geographical distribution.

2. Literature Review

2.1 Study Areas

When looking at air quality as well as land use patterns, spatial variability is important as well as a challenge in most cases. In Table 1, you can see study areas used in previous similar studies. As can be seen, studies tend to limit their analysis to either a couple of major cities or one region, restricting the generalization of their findings. Ku (2020) looked at different urban metrics that could potentially correlate either positively or negatively with air pollutant variables. While five cities were studied, it was limited to a small geographical region in Taiwan. While this most likely gives very detailed information about the areas involved in the analysis, these patterns could change if observed somewhere else. In another analysis done by the authors McCarty et al. (2015), data from the entire United States were used. However, their analyses relied on limited and unbalanced air quality monitoring stations which causes biased results in urban regions.

Table 1: Description of the study areas used in some recent studies.

Title	Author	Study Area
Support tools for land use policies based on high resolution regional air quality modelling	Ajtai et al. (2020)	Bucharest Functional Urban Area, Romania
Does urban forestry have a quantitative effect on ambient air quality in an urban environment?	Irga, Burchett, and Torpy (2015)	Sydney, Australia
Influence of urban vegetation on air pollution and noise exposure – A case study in Gothenburg, Sweden	Klinberg et al. (2017)	Gothenburg, Sweden
Analysis of the NO2 tropospheric product from S5P TROPOMI for monitoring pollution at city scale	Prunet et al. (2020)	Paris, Milan, Madrid, and Athens
Effect of Land Use and Cover Change on Air Quality in Urban Sprawl	Zou et al. (2016)	Changsha, Zhuzhou and Xiangtan (Hunan Province)
Accounting for spatial effects in land use regression for urban air pollution modeling	Bertazzon et al. (2015)	Calgary, Canada

The impacts of urbanization on air quality over the Pearl River Delta in winter: roles of urban land use and emission distribution	Chen et al. (2013)	Pearl River Delta Region
The impacts of road traffic on urban air quality in Jinan based GWR and remote sensing	Wang et al. (2021)	Jinan, China

2.2 Data

In past studies one of the most common data sources for air quality has been that of ground station monitoring. These ground stations can be implemented at stationary points on the ground and can track PM 2.5 as well as trace gasses such as NO_2, Ozone (O_3), and sulfur dioxide (SO_2). These stations are one of the most accurate ways to measure air quality at the surface and have been implemented for decades to record these observations. This can be seen in studies like Eum et al. (2019), where ground station monitors were used to conduct a wide-scale study across the US to assess the effects of long-term exposure to NO_2 on elderly populations. The authors obtained their data from the daily NO_2 data from the Air Quality System (AQS) publicly available from the EPA from December of 2000 to December of 2008. While they did find that air quality did increase cardiovascular health risks, it is widely known that air quality monitoring stations show much less spatial variability in the data that is recorded. This is due to them usually being installed at a fixed point, and areas that do not have these stations cannot be included in analysis. Typically, these monitoring stations are installed in the denser part of urban areas to see whether the city is following EPA guidelines. There has also been recent advances in the area of consumer-grade air quality devices that can be placed in different areas by citizen scientists. These are optical station monitors that can record readings such as PM 2.5 and weather conditions such as temperature and relative humidity. While these devices are great for making air quality monitoring networks within cities, and according to a study done by Stavroulas et al

(2020), these devices were seen to offer stability when used within an urban area. It does beg the question on whether everyday people are able to perform required maintenance. Leading to an issue when trying to define a generalized relationship on a large scale.

In recent years, another popular data source has been satellite imagery specifically designed to report trace gas values in the atmosphere. These generally differ from surface observations, as you are viewing the average values of the tropospheric column. What this basically means is you are looking at averaged values of an entire atmospheric column that is above the surface, which can lead to less accurate and skewed results. The resolution of the images also tends to be coarse and can have pixels that represent multiple kilometers. While these are definite downsides, there is still value in using satellite imagery to measure trace gas pollutants. They typically provide a much wider range of spatial variation in the data collected, which is valuable despite the fact the data is less accurate. While ground station monitoring is beneficial for small-scale studies, satellite imagery has the clear advantage when trying to observe patterns across an entire landscape that may or may not have air quality stations. Two popular satellite platforms that observe trace gas values at the tropospheric level are OMI sensor and the TROPOMI sensors. Both platforms observe NO_2, SO_2, O_3, as well as aerosol optical depth or AOD. The TROPOMI sensor is a nadir viewing spectrometer that is aboard the European Space Agency's Sentinel 5p (Veefkind et al., 2012). Unlike the older platform OMI, which was developed by NASA, this sensor has a much higher spatial resolution of 7 km (5.6 km as of 6 August 2019) in the along-track, and about 3.6 km in the across-track of the swath. A study done by Geffen et al. (2019), compared the NO_2 and NO (nitrogen dioxide) readings collected from the TROPOMI sensor against the same readings recorded by the OMI sensor. Due to them having about the same recording time, it was simple to draw comparisons between the

two platforms. It had been found in their study that both were good at picking up large hotspots, but the TROPOMI sensor was able to better pick up smaller hotspots. Therefore, this sensor has been gaining popularity for use in current studies. One example of such studies is one conducted by Ghahremanloo et al. (2021), where the impact of COVID-19 on air pollution in East Asia was observed. Using the nitrogen dioxide data provided, the authors were able to determine that due to less frequent use of roadways and travel, NO_2 had shown a decrease than in previous years allowing for less polluted air. Because of the resolution of the sensor, urban air quality can be better observed.

There are also some scenarios where portable monitoring are also used. These studies usually take place in one location. This is a great method for sampling, as it allows for spatial variability as well as accuracy. It is, however, resource-demanding, as well as time-consuming. Usually in this case the data must be collected personally/manually, meaning that the time reference is only for as long as you record data. Depending on how the instrument is being used, the data may also not be continuous or different users might observe different results or have biased results due to the instrument being used, making this data source to be slightly inappropriate if looking at large timeframes or any scenario where gaps in data would skew the results. This technique was employed by Irga, Burchett, and Torpy (2015) in a study they conducted on the effects of urban forestry on ambient air quality. In this study, the authors chose 11 monthly air samples that were randomly chosen based on a radius around 100m from the chosen sample site. This negated temporal non-independence, but one must ask how much of the data recordings were skewed due to human error. Recordings were taken 30m away from any road, but with the samples being taken from non-stationary points could account for some data loss.

Due to the nature of the topic being observed in this study, it is important to also review popular data sources for that of land use/land cover (LULC). These provide classifications of what is present within a landscape and allows for urban land use to be quantified as a value. Typically, these work by assigning a value, or classification to a given grouping of pixels. The size of this pixel can change however depending on the dataset, and what techniques were employed during the classification of the raster dataset. There are several resolutions that can be employed; however, most datasets tend not to go finer than resolutions of 10m and can go all the way up to 30m resolutions. This can cause classes to get averaged together and does not allow for patterns to be observed clearly in highly heterogenous locations such as cities (see an example in Fig. 1). When looking at urban form, the landscape metrics used to study habitat fragmentation may also be used when characterizing the urban area. Urban landscape metrics, however, involve using patches of contiguous urban areas, unlike traditional landscape metrics that use natural areas as their focus. After patches are identified different landscape metrics such as number, mean patch area, and other variables can be calculated. These urban metric patterns can be used to measure growth at a city level, but its use case has been limited to certain geographies (Bereitschaft & Debbage, 2013; Buyantuyev, Wu, & Gries, 2010; Seto & Fragkias, 2005).

2.3 Models

In terms of models, there are various types that can be employed depending on the work being done. Process-based models can consider atmospheric or meteorological data to predict air quality values, while statistical regression is able to observe a relationship between what is present on the ground and how they interact with key urban pollutants. OLS (ordinary least square) regression is a simple model that is commonly used in this case scenario. This model is a

form of global model, which means that it uses all training data to train a single classifier or variable and can be applied equally to different areas of interest (Fang et al., 2015). OLS regression is represented by equation (1) as shown below.

$$y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \varepsilon,\ \varepsilon \sim N(O, \delta^2) \qquad (1)$$

where x_i and y are respectively independent and dependent variables; k is the number of independent variables; β_0 is the intercept; β_i is the parameter estimate for the variable x_i; ε is just simply the error term. The parameter estimates, β_i, are assumed to be spatially stationary (Fang et al., 2015). A form of OLS regression was used during the study of Ku, 2020. In this case, the author used multiple linear regression (MLR), which is an extension of OLS. Ordinarily, OLS regression is only able to use one explanatory variable. MLR, on the other hand, can look at multiple variables that might explain the relationship to the independent variable.

Another popular model used by researchers is known as geographically weighted regression, also known as GWR. This model is numerically intensive and requires a good amount of computing power but is generally more accurate than OLS at predicting spatial variables. Since GWR can consider spatial heterogeneity of a landscape, results tend to be more accurate than that of other simpler forms of linear regression that do not take this into account such as OLS (Bertazzon et al., 2015; Wang et al., 2021; Lu and Liu, 2016). Specific requirements for this model include the calculation of *n* local linear regression equations, which use a distance-based weighting scheme. This allows for the dataset to be divided into *n* subsamples, one for each georeferenced dataset (Griffith, 2008). This model is overall very similar to the implementation of OLS; however it considers non-stationary variables such as

climate and air pollution by allowing model parameters to change over space (Su et al., 2012; Gao and Li, 2011). It has been implemented by many researchers and has been noted by Robinson et al. (2013), to outperform OLS. Due to this fact, Wang et al. (2021) used GWR in their study of urban road networks and traffic in relation to air quality to observe the relationship between the variables observed. It should also be noted, however, that GWR is not the end all be all model to use for every circumstance that observes spatial data. In previous studies, there are two major limitations that come along with using GWR that has been ignored by a good number of researchers. The first issue that arises during the use of this model is the multiple testing issue, which makes significance at the local level questionable. The second limitation is that GWR models, produce for all variables, a single optimized bandwidth. This gives the assumption that all factors affect air quality at the same spatial scale (Fotheringham, Yue, & Li, 2019; da Silva & Fotheringham, 2016). This assumption is questionable, as it is simply not the case under most circumstances. It is important to remember, that while it is possible to create generalizations, pollutants and trace atmospheric gases are highly variable depending on the area or region, and it is more efficient to observe patterns than it is to try and precisely predict these values based off land use forms.

Recently newer forms of models have been implemented to account for the complex nature of the relationship between air quality and urban morphology, due to them being ultimately non-linear (Edussuriya, Chan, and Malvin, 2014). This requires machine learning models like random forests (RF) or boosted regression trees (BRT) to be used, and their use in previous studies has started to gain traction. Elith et al. (2008) provided a working guide on how to use BRT with ecological as well as landscape patterns. Throughout their study, this model was touted to be more stable and better at predicting characteristics than traditional learning models.

It also fares better than random forest models since there is no need for complex data transformations and easily handles complex relationships, without the risk of overfitting that usually comes with the territory of using an RF model. RF is based on one single tree instead of sequentially generated trees that are used by BRT. The boosting algorithm that this form of model uses is iterative and develops a final model by progressively adding trees as well as constantly re-weighting the data to compensate for poorly predicted relationships from the previous tree (Leathwick et al., 2006). This averaging technique by bagging or boosting trees significantly improves the performance of the model being run (Hastie et al., 2001)

2.4 Findings:

It is generally accepted that both PM 2.5 as well as NO_2 are both heavily influenced by traffic patterns as well as roadway networks within urban areas. Both PM 2.5 and NO_2 are created by both industry as well as exhaust from automobiles. This is seen in studies such as a review on urban air quality written by Gulia et al. (2015) in which they conducted an analysis on current policies regarding mitigation practices. In the UK alone, road transport has been shown to be the largest contributing factors to NO_x and PM (particulate matter) emissions. NO_2 being a trace gas greatly impacted by urbanization as well as industry. Thermal powerplants as well as heavily urbanized areas in India have been shown to cause hotspots for air pollution, mostly NO_2 (Siddiqui et al., 2020). In these hotspots, some of the more common anthropogenic factors include domestic combustion, agricultural waste burning, incomplete fuel combustion, industrial pollutants, power plants, and construction activity (Siddiqui et al., 2020). This is also seen in the study written by Lu and Liu (2016), where the cities with higher levels of urbanization and traffic congestion had the most amount of NO_2 as well as SO_s (sulfur dioxide). This study was done on 287 cities in China, meaning that this is most likely a generalized pattern. PM 2.5 is a

surface level emission that has a strong correlation with exhaust gasses within urban areas. Coal combustion is another source of this pollutant, but it has been seen that traffic is the main polluter within cities, especially during peak hours that cause traffic jams (Wang et al., 2021). There are also other metrics and landcover types that can also influence air pollution levels in urban areas. In their study, Ku et al. (2020), determined that nitrogen dioxide was negatively correlated with metrics associated with vegetation and forests. This metric was primarily land use percentage or PLAND, meaning that higher percentages of vegetation in a landscape caused air quality to decrease over time and buildings did not really have any significant impact on pollution levels. The drawback of this study, however, is that it cannot be generalized since only 30 ground point monitoring stations were used in a comparatively small region to collect their air quality data. This begs the question of whether the findings from this study can significantly contribute to better urban planning methods. For example, McCarty and Kaza. (2016), stated that fragmentation due to roadway networks in urban areas is an important consideration when analyzing land use change and air quality. There is also a great degree of questioning that needs to be done on the resolution of these land cover datasets. Typically, finer resolutions tend to fare better due to the heterogeneity found within cities. If the datasets being used do not pick up the full scope of land use within a city, it is difficult to draw any significant conclusions from this data. McCarty and Kaza (2016), even admitted that the resolution of the dataset could have influenced their findings.

3. Methods

3.1 Study area

Cities in the southeast typically have vastly different climatological conditions than cities in the northeast or western parts of the country. If one is in Charlotte or Raleigh, these are categorized as mixed-humid, whereas cities further west like Denver are in cold climates; these areas are going to have different NO_2 patterns.

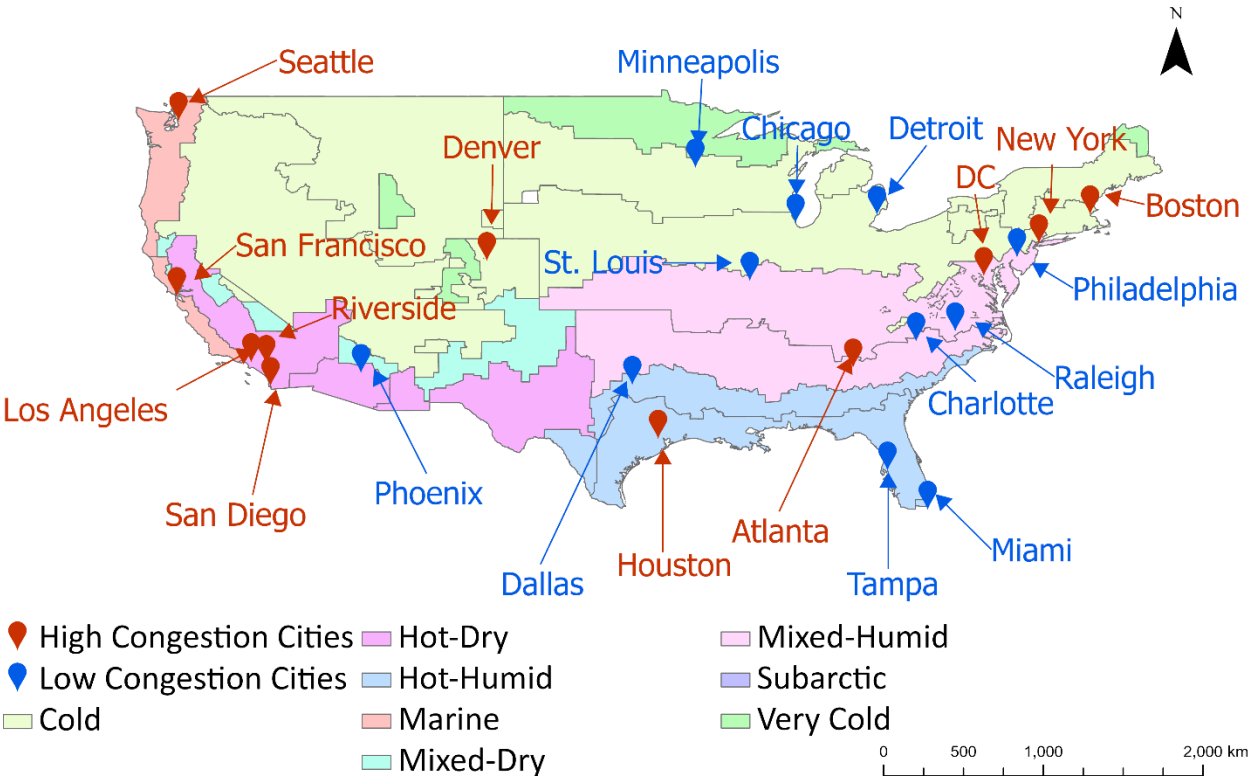

Figure 2: Shows the study area and cities overlayed with regional climate zones. The U.S. Energy Information Administration provided climate zone data.

These 22 cities (recorded in Figure 2) were chosen because they span multiple different climate zones throughout the United States. These climate zones are listed as cold, hot-dry, hot-humid, marine, mixed-dry, subarctic, and very cold. NO_2 reacts differently in the atmosphere depending on climatic conditions; choosing such a wide variety of cities from different climates allows for a

generalized approach for analysis. Data providing the climate zone information was retrieved from the U.S. Energy Information Administration.

3.2 Data sources

3.2.1 Fine-scale urban form at 1-meter resolution

The land use/landcover (LULC) dataset used for this study is the UrbanWatch database developed by Zhang et al. (2022). This dataset provides 1-meter land cover classifications at 91.25% accuracy for 22 urban areas within the United States and is still being expanded into other cities. Figure 3 shows a comparison between UrbanWatch and other commonly used databases.

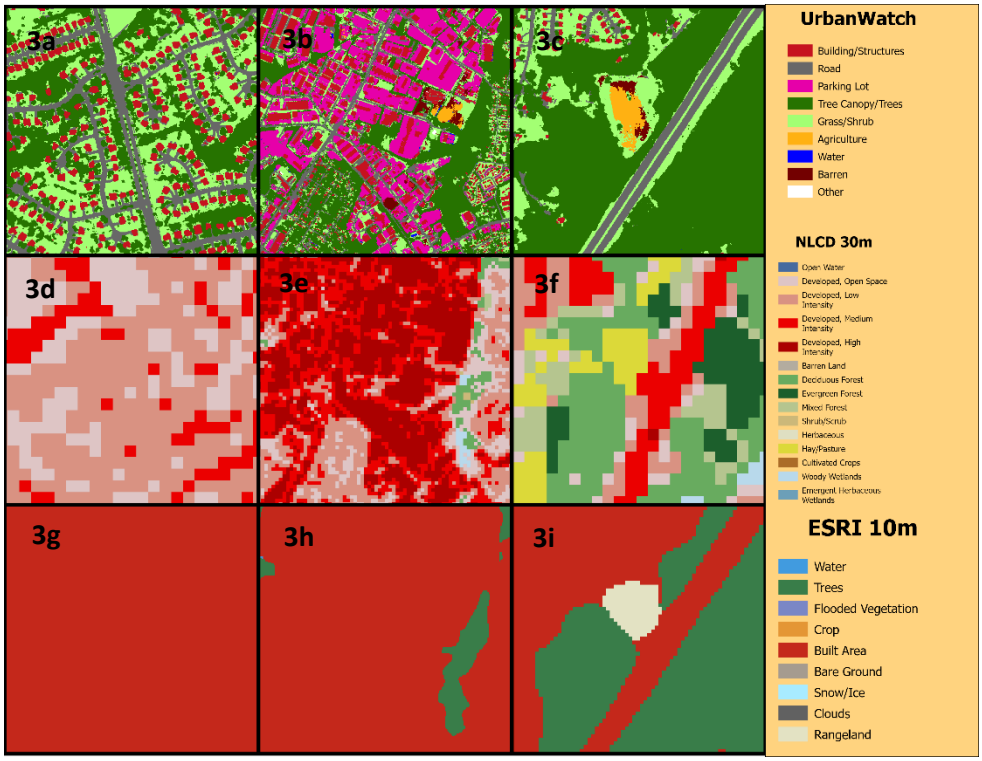

Figure 3:Figure 3a-c shows a representation of the 1-meter data from UrbanWatch compared to the NLCD 30-meter dataset shown in 3d-f as well as ESRI 10m 3g-i in the same geographic area.

There is a clear advantage to using a 1m dataset when studying urban landscapes. Cities are often dense and highly heterogeneous, making it difficult to see the differences between land classes when using a traditional data source like NLCD. The classes that are described by the UrbanWatch data are as follows; Buildings/Structures, Roads, Parking Lots, Tree Canopy, Grass/Shrub, Agriculture, Water, Barren, and lastly Other.

Zhang et al 2022 used 1-m imagery retrieved from NAIP (National Agriculture Imagery Program) covering these cities from the USGS Earth Explorer portal during leaf on seasons taken from the years 2014-2017. These classifications were then applied to these images using Fine resolution, Large-area Urban Thematic information Extraction (FLUTE), which is a form of semi-supervised classification. The generated dataset just shows the land cover data for the urban centers within city limits. This means that there is no information available for suburb or rural areas.

3.2.2 TROPOMI NO_2 dataset

Due to the spatial scope of this study, data collected from ground station sites would have been insufficient. Satellite data must be used to create a generalized observation across a broad area, such as the TROPOMI sensor aboard the Sentinel 5-precursor satellite. This sensor has a spatial resolution of 7.2x3.6 km^2. However, as of 6 August 2019, data is recorded in 5.6x3.6 km^2 pixels (Geffen et al., 2020). The satellite also passes over every day roughly at noon.

Regardless of this being one of the smallest resolutions currently available for column density values of NO_2, the coarseness of this dataset would not be helpful for this study. Physical oversampling was used to pre-process the data retrieved from the TROPOMI sensor to help solve this problem. Oversampling is the process of creating a level 3 data product that is much finer than the original level 2 product. This was done by importing the downloaded daily TROPOMI

data into Python for each of the studied seasons using a provided code from the authors Sun et al. (2018) found on GitHub. This created a referenced GeoTiff image for fall, winter, summer, and spring. Allowing for the data to be refitted to finer grid sizes by treating the pixels inside the rectangular FOV (field of view) on the ground as a sensitivity scale (Sun et al., 2018); making refitting the data to a grid of .01° (roughly 1km on the ground) possible. While the data was still courser than our LULC dataset, this new resolution allows a better comparison between the two datasets. Figure 4 shows how the dataset looks after it has gone through the process and become a level 3 data product.

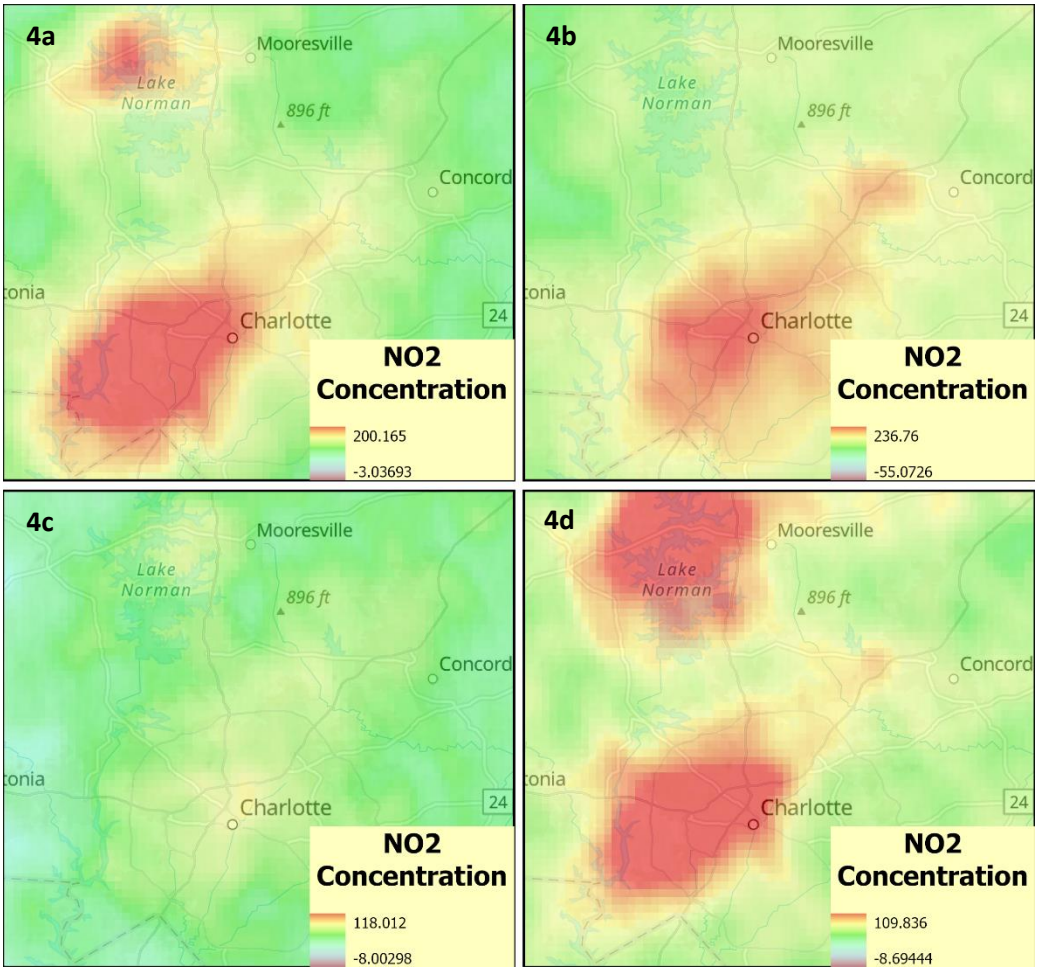

Figure 4:Map showing NO_2 concentration in Charlotte, North Carolina. 4a shows concentrations in the fall, 4b shows concentrations in the winter, 4c shows concentrations for the spring, and 4d shows concentrations for summer. Units are in $\mu mol/m^2$.

To account for seasonal variations in NO_2 due to weather patterns and regional climates, collecting observations from each season, Fall (September-November) 2019, Winter (December-February).2020, Spring (March-May) 2020, and Summer (June- August) 2020. With NO_2 concentrations being highly dependent on geographic and meteorological patterns, it is crucial to account for these shifts in the analysis as much as possible. It is also important to note that there was a global pandemic towards the start of spring 2020, causing lockdowns. Due to this, traffic patterns were not the same, so there might be some variation with the data collected from this time.

3.2.3 Congestion Data and Index Creation

As can be seen from Figure 2, each of the 22 cities in which the observations were taken had been sorted into areas of high and low congestion. This technique accounted for the fact that NO_2 is heavily influenced by traffic patterns within cities (Gulia et al., 2015; Siddiqui et al., 2020). Separating the study area and creating an index allowed for the congestion levels to be used as a control within the study. For areas of high congestion, this allowed them to be compared with areas of low congestion and see if this had any impact on the model results.

Data from the Texas A&M Transportation Institute Mobility Division was used to create the congestion index for city rankings. This dataset contains base statistics for multiple types of congestion metrics used to measure these values across 101 urban areas in the U.S., ranging from 2017-2020. In their paper, Rahman et al. (2020) determined that the three of the metrics mentioned earlier, TTIR, COEX, and CODU, were all directly correlated/connected to city congestion levels. These values could be used to observe how highly congested a city is. TTIR is a ratio of average daily travel time in peak hours to travel time in off-peak hours, COEX is a

percentage of how many lane miles are congested, and CODU is a measure of how many hours peak rush times last (Rahman et al., 2020).

Since these variables are equally tied to congestion levels, principal components analysis (PCA) could create an index of congestion using the values connected to those metrics. The values used to generate the index were taken from 2018. This was because this year was the most recent these values could be collected from.

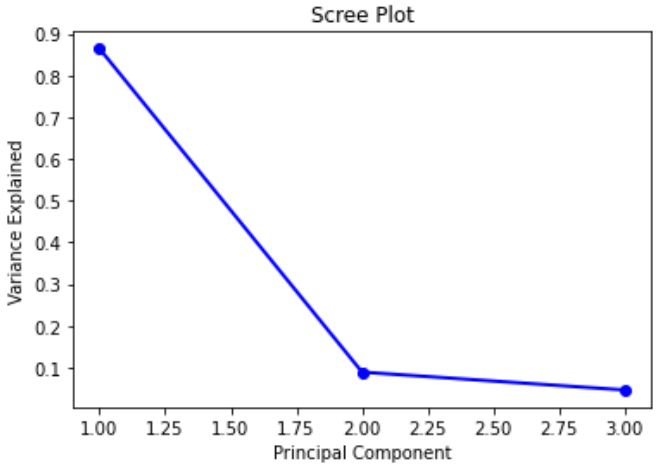

Figure 5:The scree plot explains the variance collected by each component. This plot shows that the first component explains most of the variance.

As seen in Figure 5, the first component of these three metrics combined explains most of the variance in the model. Reducing the data to a one-dimensional index would not cause much data loss. The index allowed for cities to be ranked from 1-22. The highest congested cities start at one and end in the city with the lowest amount at 22. The ranks given for each city can be seen in Table 2; any value greater than 11 was categorized as low congestion, and the remaining as high congestion. To see the ranks of each city used in this study, refer to Table 4. To use PCA data for all three of these metrics were retrieved from the year 2018, as this was the most recent year in which these metrics were recorded for each city.

Table 2: Shows the rank given to each city based on their congestion levels.

Urban Area	**Rank**
San Francisco- CA	1
Los Angeles- CA	2
San Diego CA	3
Riverside- CA	4
Washington- DC	5
Seattle- WA	6
Houston- TX	7
New York- NY	8
Denver- CO	9
Atlanta- GA	10
Boston- MA	11
Dallas- TX	12
Phoenix- AZ	13
Chicago- IL	14
Philadelphia- PA	15
Miami- FL	16
Minneapolis- MN	17
Detroit- MI	18
Charlotte- NC	19
Tampa- FL	20
Raleigh- NC	21
St. Louis- MO	22

3.3 Urban Form Extraction

One of the main objectives of this project is to define the relationship between urban form and NO_2. To calculate the urban form, first the cities had to be sampled using random sampling techniques. The number of sample points were determined based on the size of the city, but the maximum number was set at eight to make sure observed areas would not overlap. After the random points were generated for each city, circular buffers were drawn around them in two different radii sizes of 1500m and 500m, then extracted using the extract by mask tool in ArcGIS Pro. This was largely done to test relationship consistencies at different spatial scales (Ku, 2020).

This study will use five urban metrics calculated using the FRAGSTATS v4.2.1 software as the independent/explanatory variables. FRAGSTATS is a free tool that can quantify aspects of an area called landscape metrics and assists in identifying spatial patterns (Benxin and Lannone, 2020). Landscape metrics used overall were Patch Density (PD), Largest Patch Index (LPI), Edge Density (ED), Mean Patch Area (AREA_MN), and Aggregation Index (AI). Urban landscape metrics use patches of contiguous urban areas, unlike traditional ones that use natural areas as their focus. After patches are identified, different landscape metrics can be calculated. These urban metric patterns can measure growth at a city level, but their use case has been limited to specific geographies (Bereitschaft & Debbage, 2013; Buyantuyev, Wu, & Gries, 2010; Seto & Fragkias, 2005). Table 3 shows the metrics and a short description relevant to air quality/pollution.

Table 3: Describes the potential landscape metrics that may be used in each model and its relevance to air quality. These are referenced from the help guide for FRAGSTATS and reworded in the context of air quality.

Metric Definitions	
Metric	Description
PD (Patch Density)	The PD metric is relevant to air quality as it provides information on the spatial distribution of land cover types that could impact air pollution levels. PD is calculated by dividing the number of patches of a specific land cover type by the total landscape area in square meters, and then multiplying by 10,000 and 100 to convert it into 100-hectare units. This metric can be useful in understanding the potential sources of air pollution across a landscape. For example, a higher PD value for urban land cover types could suggest a more fragmented landscape with a greater number of smaller urban patches, potentially leading to increased emissions of pollutants from traffic or industrial activities.
LPI (Largest Patch Index)	The LPI metric is relevant to air quality as it can provide insights into the potential impact of dominant land cover types on air pollution levels. LPI is calculated by dividing the area (in square meters) of the largest patch of a specific land cover type by the total landscape area, and then multiplying it by 100 to convert it into a percentage. This metric can be useful in understanding the influence of the largest patch of a particular land cover type on the overall landscape. For instance, a high LPI value for vegetated land cover types could suggest that a large portion of the landscape is dominated by a vegetation, potentially swaying pollutant concentration levels.
ED (Edge Density)	The ED metric is relevant to air quality as it can help identify potential sources of air pollution associated with different land cover types. ED is calculated by dividing the sum of the lengths of all edge segments involving a specific land cover type by the total landscape area, and then multiplying it by 10,000 to convert it into hectares. This metric can be useful in understanding the amount of edge between different land cover types, which can affect the exchange of pollutants and other environmental factors. For instance, a higher ED value for urban land cover types could

	suggest that there are more edges between urban areas and natural or agricultural land cover types, potentially leading to increased air pollution from traffic or other urban activities. Additionally, if a landscape border is present, ED includes only true edge segments between patches of different land cover classes, while if a landscape border is absent, a user-specified proportion of boundary segments and internal background edge segments are included in the calculation.
AREA_MN (mean area)	The mean patch size (AREA_MN) metric is relevant to air quality as it can provide insights into the potential impact of different land cover types on air pollution levels. AREA_MN at the class level is calculated based on the number of patches and total class area. This metric can be useful in understanding the distribution and size of patches of a specific land cover type, which can affect air quality in different ways. For example, a lower AREA_MN value for forested land cover types could suggest a more fragmented landscape with smaller forest patches, potentially leading to decreased air quality due to reduced filtration of pollutants and less carbon sequestration. It is important to note that although AREA_MN is derived from the number of patches, it does not convey information about the specific number of patches present, which could be relevant for understanding the spatial distribution of pollutants.
AI (Aggregation Index)	The AI (aggregation index) metric is relevant to air quality as it can provide information on the spatial arrangement of a specific land cover type and its potential impact on air pollution levels. AI is calculated by dividing the number of like adjacencies involving a specific class by the maximum possible number of like adjacencies involving the corresponding class, which is achieved when the class is maximally clumped into a single, compact patch. The resulting value is then multiplied by 100 to convert it into a percentage. This metric can be useful in understanding the degree to which a particular land cover type is clumped or dispersed across a landscape, which can affect the movement and concentration of pollutants. For instance, a higher AI value for urban land cover types could suggest a more compact and contiguous landscape with higher concentrations of pollution, potentially leading to negative impacts on air quality and human health. Conversely, a lower AI value for natural or agricultural land cover types could suggest a more dispersed and heterogeneous landscape that allows for better filtration of pollutants and improved air quality.

Also aiming to study which landscape metric associated with what class type has the most considerable impact on a city, the landscape metric database was broken up by land class once the metrics were calculated. This ultimately was done to ensure that we were seeing a clear picture of how the spatial characteristics of each class type impacted the concentration of NO_2.

Table 4: Shows the class type and associated landscape metrics used in the seasonal models. These metrics will be used as the independent variable in the models and will change based on the land class type.

Class Metrics	
Class Type	Associated Metrics
Structures	PD, LPI, AREA_MN, AI
Trees	PD, ED, AREA_MN, AI
Roads	PD, ED, AREA_MN, AI

3.4 Statistical modeling

BRT is a technique that aims to improve the performance of a model. This is done by fitting many models and combining them for prediction (Elith et al., 2008). Some hyperparameters first need to be tuned when using this model. These parameters are the learning rate, number of trees, and tree depth. When determining these, Elith et al. (2008) recommend using at least 1000 trees and a learning rate between .1-.0001. In their working guide, they used these parameters for a similar purpose as an example for ecological and landscape analysis, making this suitable for our study as well. The tree depth or complexity must also be set; for this purpose, a value of 1-10 was used. Tree depth refers to the number of splits that are within a tree, and more complex interactions occur at larger depths. Finding balance can be tricky, however, because larger tree depths can cause model instability (Elith et al., 2008). This is why the range of 1-10 was used for this parameter. This allows for complex interactions to be observed without compromising stability. Unlike linear models, BRT modeling is able to handle multicollinearity better than traditional models like OLS regression. However, VIF scores were still calculated,

and any metric that had above a value of 5 for each class was dropped. That is how the final metrics in Table 4 were determined. These are also the metrics that will be included as the independent variables for each specific associated class type, and will be compared with NO_2 levels as the dependent variable. BRT, unlike random forest models, requires no need for data transformations or elimination of outliers and can easily handle complex relationships. Random forests can also lead to greater bias because results are based on one single tree instead of multiple. Boosting, on the other hand, grows the trees by sequentially modeling the residuals, leading to less bias. BRT also has no significance value (P-value) or degrees of freedom (Elith et al., 2008). This process also introduces stochastic gradient boosting to the model which allows it to be more accurate without a high risk of being overfitted. BRT also has been shown to be more stable than random forests (Yang et al., 2016; Friedman, 2002).

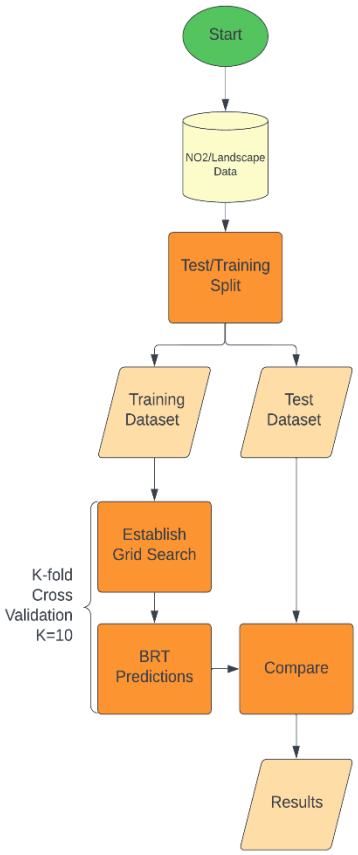

Figure 6: Shows an outline of the BRT model used for this study

3.4.1 Model validation and calibration

Model verification was done by using k-folds cross-validation k=10. Cross-validation was used on a training dataset where 80% of the data was randomly selected to determine the best hyperparameter combination by selecting the model with the lowest root mean squared error (RMSE) using a grid search method. This was done to ensure that the model was not being overfitted by the parameters, and that the model was as accurate as possible when run on the test dataset for validation and comparison.

The test dataset retained 20% of the data that was also randomly selected. Figure 6 shows an outline of the model, and Table 5 shows the parameters used for the grid search.

Table 5: Shows the values used for each parameter in the grid search used for the BRT modeling.

Grid Search Parameters	
Parameter	Value
Number of trees	1,000-10,000; spaced by 100
Tree Depth/Complexity	1-10
Learning Rate/Shrinkage	.1-.0001

This study will have 24 models broken up by season, class type, and buffer radius size (1500m and 500m). In order to account for traffic as a control, the congestion rankings from each individual city will be included as an independent variable. This was done to control any outside factors that may come into play, as there has been evidence to suggest that these factors can influence NO_2 levels within cities(Gulia et al., 2015; Siddiqui et al., 2020; Lu and Liu, 2016; Wang et al., 2021; Ku, 2020; McCarty and Kaza, 2015).

4 Results

4.1 Seasonal NO_2 Concentrations

When looking at the seasonal NO_2 concentrations, the season with the highest average values was winter (e.g., Figure 7). This trend was noticed in all buffer radius sizes, leading to the conclusion that conditions in the winter may be very favorable for the formation of NO_2.

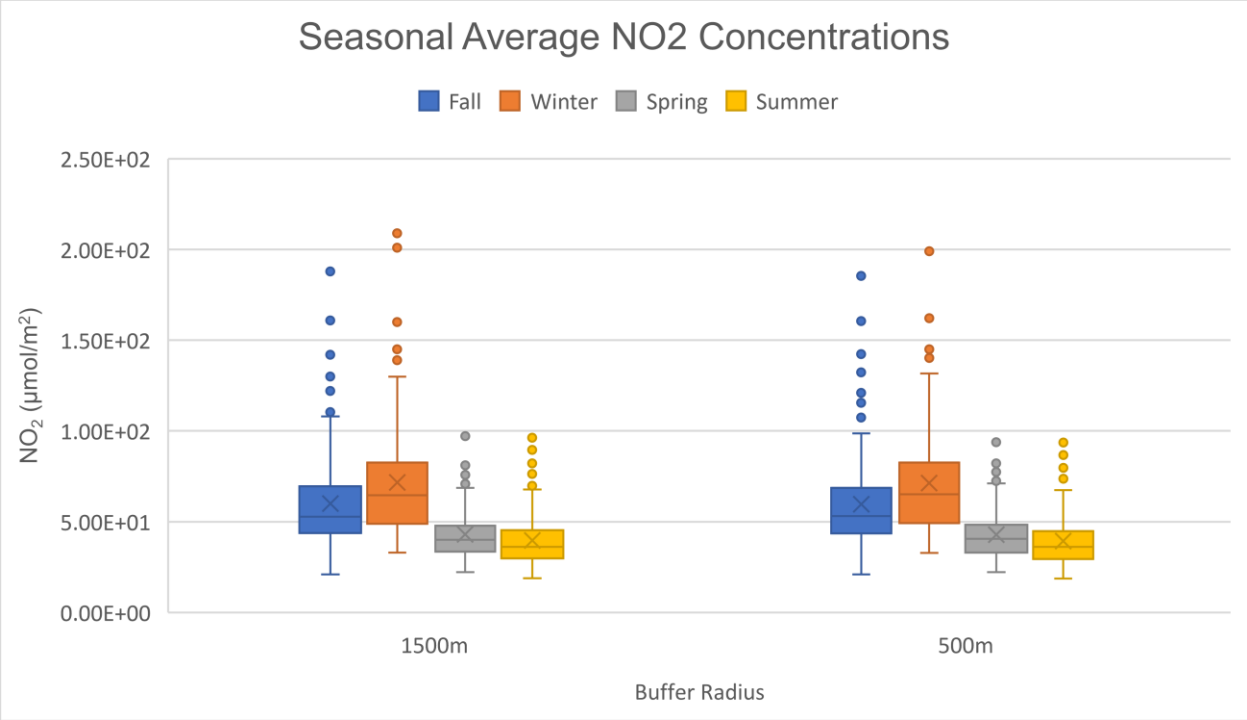

Figure 7: Shows the average seasonal concentrations of NO2. A trend of winter having the highest concentrations out of all the seasons, regardless of buffer radius, can be seen.

Out of all the seasons, winter also seems to have the largest spread of data and interquartile range (IQR). Summer and spring are consistent with their NO_2 readings and appear to be very similar. Fall appears to have the most similar spread of data to winter.

4.2 Influence of Urban Land Class and Form on NO_2 Concentrations

This section describes how landscape metrics can impact the concentration of NO_2 differently depending on which urban land class is observed in the model. After the model was

run it was observed across the board, the land classes that had the most considerable impact on these concentrations were urban buildings, trees, and roads.

4.2.1 Buildings

Table 6: This table shows the difference between the model results for observations collected in each buffer radius

Buildings					
1500m			**500m**		
Season	**R-Squared**	**RMSE ($\mu mol/m^2$)**	**Season**	**R-Squared**	**RMSE ($\mu mol/m^2$)**
Fall	.31	26.28	Fall	.07	28.78
Winter	.50	27.95	Winter	.41	32.09
Spring	.39	13.16	Spring	.16	15.27
Summer	.49	8.33	Summer	.06	11.21

The observations from both buffer radii appear to be similar, however there are some differences to note. Table 6 shows that the R-Squared values appear to be performing better in the 1500m buffer radius than in that of 500m. This can be seen with fall where the R-Squared value within a 1500m radius is .31 as opposed to in 500m meters where it falls to .07. The root mean squared errors (RMSEs) also appear to be lower in the 1500m observation models. This indicates that there is less overfitting from results produced from these observations, and that there is more stability using these parameters. A consistent trend emerged with these results where winter had the highest R-Squared value, .50 and .41 respectively, and summer had the lowest RMSEs values which were 8.33 $\mu mol/m^2$ and 11.21 $\mu mol/m^2$.

Table 7: Lists all of the independent variables used in the model and lists their importance in terms of how much they impacted NO_2 values. Variable importance is in terms of percentage, meaning that variables that have a value of 100 are the most impactful, while 0 means there was no impact. These percentages are also based on each singular value, so multiple metrics can have a value of 100 or 0.

Buildings Variable Importance

1500m Buffer Radius				
Variable	**Fall (%)**	**Winter (%)**	**Spring (%)**	**Summer (%)**
Congestion Ranking	100	100	100	84.8
PD	20.6	8.97	17.7	33.6
LPI	10.6	0	3.52	3.56
AREA_MN	48.7	39	30	100
AI	0	1.01	0	0
500m Buffer Radius				
Congestion Ranking	100	100	100	100
PD	4.21	3.33	24.1	3.06
LPI	0	0	0	0
AREA_MN	37.2	7.66	46.2	46.2
AI	8.94	9.76	10.9	10.9

With variable importance, as listed in Table 7, other than the congestion ranking AREA_MN is typically listed as the second most important variable. Congestion ranking holds importance throughout all seasons except for in summer at the 1500m scale, where AREA_MN has a value of 100 and congestion only has an importance value of 84.8.

4.2.2 Trees

The trees land cover class had similar results to that of buildings, however there was less of a distinction between the two buffer radius sizes. The largest R-Squared values were in the winter with .47 at the 1500m buffer scale, and .55 at the 500m buffer scale.

Table 8: This table shows the difference between the model results for observations collected in each buffer radius for the trees landcover class.

Trees					
1500m			**500m**		
Season	**R-Squared**	**RMSE (µmol/m^2)**	**Season**	**R-Squared**	**RMSE (µmol/m^2)**
Fall	.39	26.39	Fall	.40	26.88
Winter	.47	29.1	Winter	.55	27.92
Spring	.30	14.28	Spring	.40	12.73
Summer	.41	9.14	Summer	.43	8.1

Table 8 also shows that the lowest error values for RMSE are once again in the summer. 9.14 µmol/m^2 is recorded for the 1500m buffer observations, and 8.1 µmol/m^2 for 500m.Instead of using LPI as one of the variables in the model, this was replaced by ED for trees. The variable with the largest amount of importance is shown to be congestion rank, with PD being the second most important variable in most cases. Table 9 lists these values, and PD is one of the main variables impacting NO_2 at both scales.

Table 9: Lists all the independent variables used in the model for trees and lists their importance in terms of how much they impacted NO_2 values. Variable importance is in terms of percentage, meaning that variables that have a value of 100 are the most impactful, while 0 means there was no impact. These percentages are also based on each singular value, so multiple metrics can have a value of 100 or 0.

Trees Variable Importance

1500m Buffer Radius				
Variable	**Fall (%)**	**Winter (%)**	**Spring (%)**	**Summer (%)**
Congestion Ranking	100	100	100	100
PD	39.7	27.3	68.3	57.5
ED	29.9	19.5	52.5	24.4
AREA_MN	14.1	39.8	20.1	0
AI	0	0	0	3.14
500m Buffer Radius				
Congestion Ranking	100	100	100	100
PD	18.3	22.8	31	37.6
ED	45.7	18.1	.2	0
AREA_MN	10.1	0	5.5	8.6
AI	0	0.8	0	3.7

4.2.3 Roads

The same metric variables used in the models for tree cover were also used for roads (PD, ED, AREA_MN, & AI). The R-Squared values do not show the same consistencies for winter having the highest in both buffer radii. In Table 9 winter had the highest R-Squared of .68 in 1500m buffer radii, but is dropped down to third highest in a 500m buffer with a value of .50. Summer remained the season with the lowest RMSEs. In 1500m summer showed a value of 9.20 $\mu mol/m^2$ and 8.96 $\mu mol/m^2$ at 500m.

Table 10: This table shows the difference between the model results for observations collected in each buffer radius for the roads landcover class.

Roads					
1500m			500m		
Season	**R-Squared**	**RMSE ($\mu mol/m^2$)**	**Season**	**R-Squared**	**RMSE ($\mu mol/m^2$)**
Fall	.44	23.68	Fall	.67	22.05
Winter	.68	22.82	Winter	.50	28.30
Spring	.54	11.29	Spring	.65	9.88
Summer	.14	9.20	Summer	.35	8.96

Table 11: Lists all the independent variables used in the model for trees and lists their importance in terms of how much they impacted NO_2 values. Variable importance is in terms of percentage, meaning that variables that have a value of 100 are the most impactful, while 0 means there was no impact. These percentages are also based on each singular value, so multiple metrics can have a value of 100 or 0.

Roads Variable Importance				
1500m Buffer Radius				
Variable	**Fall (%)**	**Winter (%)**	**Spring (%)**	**Summer (%)**
Congestion Ranking	100	100	100	100
PD	3.3	6.5	3.8	0
ED	19	7.1	47	33
AREA_MN	7.3	22	2.4	25
AI	0	0	0	4.1
500m Buffer Radius				
Congestion Ranking	100	100	100	100
PD	0	10.9	0	0
ED	2.90	0	27.1	32.6
AREA_MN	6.23	19.1	10.4	45.4
AI	17	6	3.06	43.1

There were no consistencies across all seasons for variable importance in terms of metric values, but congestion ranking did stay the most important variable throughout all given model parameters. There does seem to be a bit of agreement that ED and AREA_MN show importance in most seasons and both buffer radii. The exception being AI with a value of 17% importance during the fall within a 500m buffer. It is also clear that the 500m buffer radius observations had less consistency than 1500m buffer radius observations. At the 1500m buffer scale, ED (edge density) was second most important throughout all seasons besides winter in which AREA_MN took its place instead. There was no consistent variable seen for all seasons at the 500m scale.

5 Discussion

5.1 Fine-scale consistencies across studies

To develop the relationship between NO_2 and fine-scale urban form, we used data from 22 urban areas across the U.S. Due to the relationship between NO_2 and traffic congestion already being well established (Agudelo-Castañeda et al., 2013; Gualtieri et al., 2015; Lähde et al., 2014; Madrazo 2018); lockdown protocols during COVID-19 caused a decrease in NO_2 concentrations in many areas (Ghahremanloo et al., 2020; Rossi et al., 2020). Having congestion rankings used as a control in the model allowed us to generalize the data between each city and look at other patterns that may have developed with urban form.

When we observe the landcover dealing with buildings, the most important factors contributing to increasing NO_2 levels appear to be most related to the size of buildings or how much of the landscape they cover. Partial dependence is a way to show the relationship of non-linear models such as boosted regression trees. It shows the relationship of a single independent variable and the predicted based on the other variables that are in the model. If we note the partial dependence plots shown in Figure 8, one can see that the variables of importance associated with this characteristic increase the NO_2 levels to a point and then eventually level off. It has been shown in previous studies that large built structures can dampen both wind direction and speed, indirectly causing the concentration of NO_2 to rise because of poor ventilation in the area (Chen et al., 2021; Peng et al., 2021; Mao et al., 2022). It is difficult to say if this is what is happening with the mean area of the landscape, as there is no way to determine the 3-D shape of a structure from the land cover dataset used. This finding is like studies that have shown the effect of landscape fragmentation within an urban landscape (McCarty and Kaza, 2015; Fan et

al., 2014). However, using 1 m data like the UrbanWatch dataset allows for individual patches of buildings to be isolated in the landscape when running calculations. This is advantageous as it allows for the individual 2D shape of buildings within an observed landscape to be analyzed. Typically, with a course resolution dataset, such as the 30 m NLCD, these buildings will get averaged into one patch that can also include vegetation or even roadways.

It is widely accepted that urban forests and larger green spaces are a key part of reducing NO_2 levels within cities (Baró et al., 2015; Manes et al., 2016; Song et al., 2016). However, it was also demonstrated by Gong et al. (2022) that these trees seem to be absorbing NO_2 disproportionately. Here in this study, we may be seeing the same trends across each of the 22 cities being studied. When looking at what impacted NO_2 levels the most in terms of tree canopy, clustering appears to play a major part. Using partial dependence (Figure 9), however, it can be shown that this relationship is clearly not linear. It appears that as clustering in the observed landscape increases, that the so do NO_2 levels overall. If we look more closely though, this clustering can also decrease concentrations of this pollutant as well. Different vegetation types can have varying effects on NO_2 at large (Dai et al., 2023), so these values could be dependent on the species of tree. Due to the nature of the dataset being used for landcover, low-level vegetation such as shrubbery and grass can be excluded from the model, but there still may be variation between the species of the tree itself. This cannot be observed as there is no information on vegetation species within the landcover database.

Roadway networks are also typically difficult to capture within the landscape when running metric-based calculations. Typically, the full shape of this land cover classification cannot be determined as it will get averaged into other urban land classifications. This is problematic, as it was shown that congestion levels and city-wide traffic patterns greatly impact

the amount of NO_2 within a city. Being able to set a defined border to observe this land use classification allows for the shape of these networks to be fully calculated. Roadways are an essential contributing factor of NO_2 due to the congestion and daily travel from vehicles. Congestion and traffic patterns cause about 50% of all NO_2 emissions (Basaric et al., 2014; Frey et al., 2010; Shon et al., 2011). Salas et al. (2021) noted that traffic restrictions put into place in Madrid had dramatically reduced the NO_2 levels in areas that were previously well above the threshold for acceptable limits. Based on the importance of variables, besides congestion ranking, attributes that contributed to NO_2 levels the most dealt mainly with size and continuity within the landscape. It can be noted in Figure 10 that these metrics do also appear to be increasing the amount of NO_2 present within these cities. Again, the relationship is not linear, but these metrics consistently increase levels of this pollutant overall. Considering that traffic and congestion are relevant with rising NO_2 levels, this result is consistent (Basaric et al., 2014; Frey et al., 2010; Shon et al., 2011). Extreme traffic conditions suggest an immediate impact on NO_2 concentrations; however, we note that urban form affects traffic patterns (Zhou and Gao, 2020), leading to an indirect impact on NO_2 emissions.

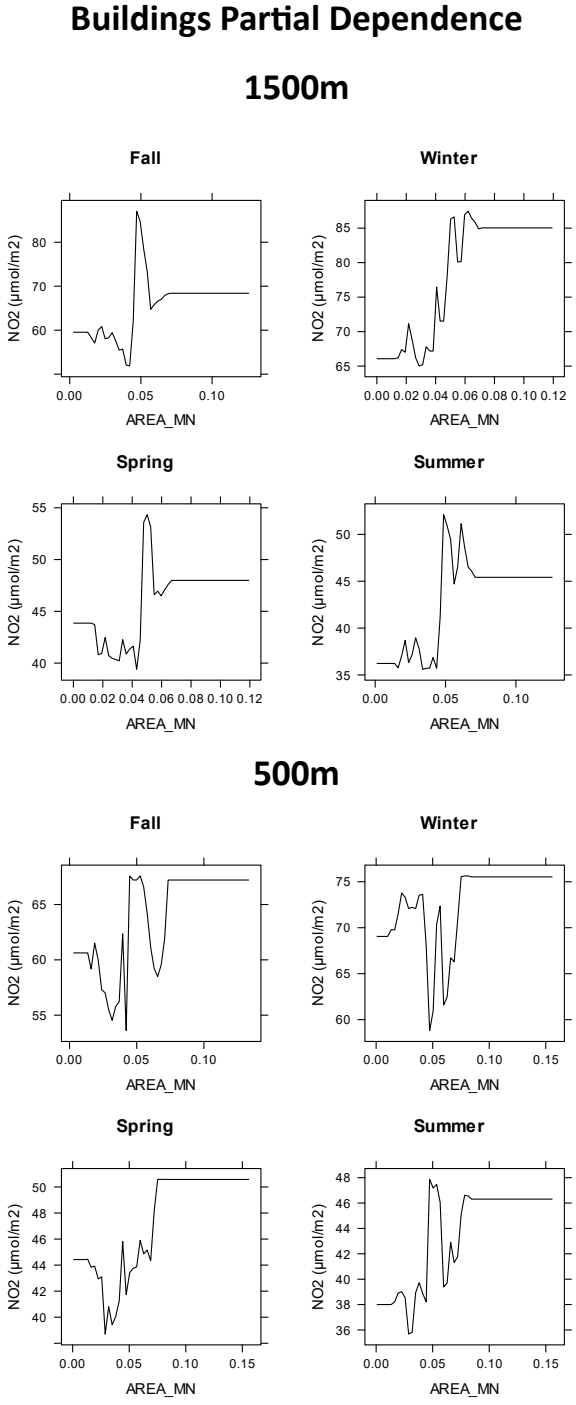

Figure 8: Shows the partial dependence of the variable AREA_MN throughout all seasons and both buffer radius scales. Partial dependence is based on how the individual variable is effecting NO_2 in comparison to everything included in the model

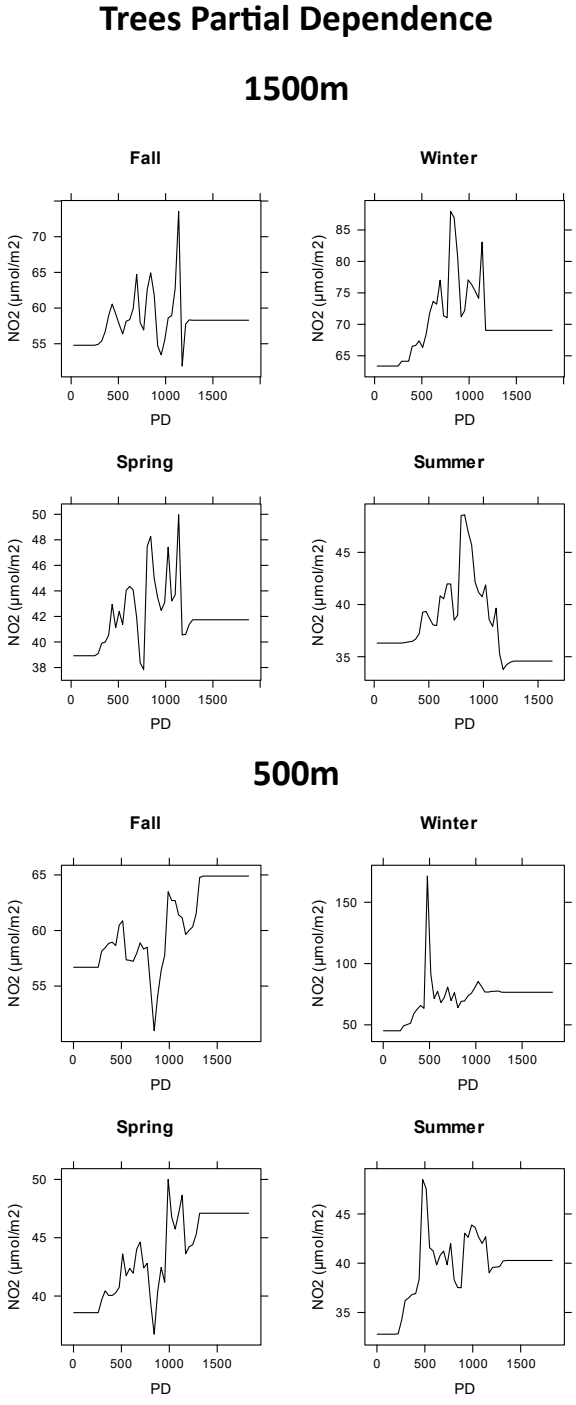

Figure 9: Shows the partial dependence of the variable PD throughout all seasons and both buffer radius scales. Partial dependence is based on how the individual variable is effecting NO_2 in comparison to everything included in the model

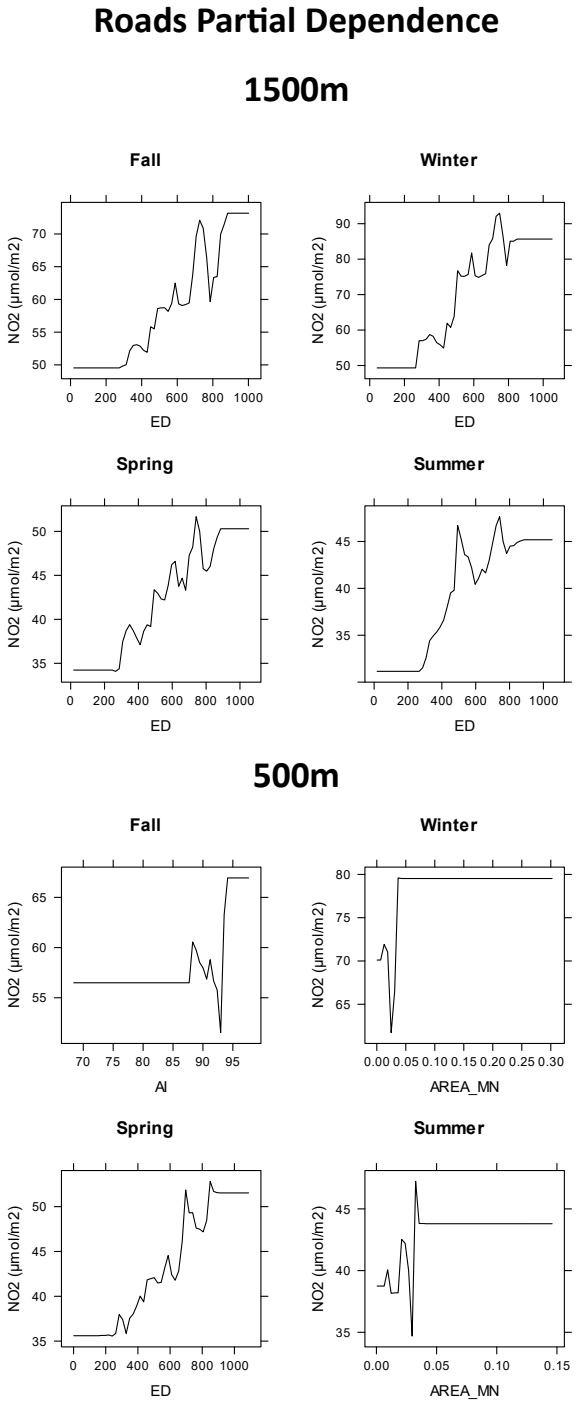

Figure 10: Shows the partial dependence of the variable of importance throughout all seasons and both buffer radius scales. Partial dependence is based on how the individual variable is effecting NO_2 in comparison to everything included in the model

5.2 Buffer Radius Size

The differences between individual buffer sizes within the landscape appear to minimal, however, 1500m did seem to perform better in terms of having higher R-Squared values and lower RMSEs overall. The most difference that could be seen was in the analysis of the buildings' land use classification. Here, R-Squared results clearly saw higher values for the 1500m buffer radius when compared to the 500m buffer radius. Landscape metrics are based on the size of the landscape from which they are calculated and are highly scalable (Wu et al., 2002). Whether or not a 1500m or 500m buffer radius will perform better than the other may be highly dependent on the observed city because of the scalable nature of these metrics. With buildings being the one to show the difference between buffer sizes the clearest, a buffer with a radius of 500m may overestimate what is in the landscape and not allow for other urban landforms to be observed well enough. Given the analysis conducted by this study, it appears that there really is not enough of a difference between these two observational scales to give a clear answer of which is better for all classification types. Ku (2020) stated that 1500m performed the best out of all buffer types tested in their study, however, they used 30m land use data and performed analysis using multiple regression. City size and urban population has also been shown to be a contributing factor to NO_2 concentrations within cities (Lamsal et al., 2013). Without using a finer-scale land cover dataset to determine the values of these metrics in highly heterogenous environments, it questions the accuracy of any other study trying to build a relationship between urban form and air quality, specifically NO_2.

5.3 Consistencies across seasons

With NO_2 being highly variable, seasonal fluctuations were observed to see whether there was any season consistently performing better than the other. It was also important to observe whether there were any seasonal consistencies between variable importance for each of the class types. At first glance there does not seem to be any major seasonal patterns noted, however, winter appeared to produce the highest R-Squared values overall. With winter also appearing to have the highest concentrations of NO_2, this observation is most likely not a coincidence. In other studies, it has been consistently shown that winter can produce higher concentrations of this pollutant than any other season (Boersma et al., 2009; Kendrick et al., 2015). One of the main pathways NO_2 is removed from the atmosphere is through exposure to sunlight and photochemical destruction, which allows for ozone (O_3) to be formed towards the end of this process (Pancholi et al., 2018). In the winter, there is typically less solar radiation during the day in the northern hemisphere. This is added on top of more people using heating services during this time of year, which have a biproduct of NO_2 (Bozkurt et al., 2018).

There also appeared to be a consistent trend of summer having the lowest RMSE value when compared to the other seasons. This indicates that there is a better model fit during this time of year. Temperatures and synoptic weather conditions, such as wind direction and speed, have a tremendous impact on the levels of NO_2 within a city. It has been shown in studies that urban heat islands (UHIs) can influence air pollutant concentrations. In this context, a UHI is where temperatures in urban areas are higher than in rural or vegetated land (Lai, 2009). It is unclear as to if these varying temperatures in the summer are causing these better fits, but again, summer typically has higher averages of solar radiation in the northern hemisphere. This may allow for NO_2 to be removed easier during this time, allowing for a better fit.

5.4 Challenges

In this study, a couple of challenges were faced. The main one being the resolution of the tropospheric NO_2 data. The use of physical oversampling allowed us to re-grid the data to a grid of about 1km in resolution. Oversampling is a useful tool, and in this case refers to the creation of a level 3 data product that is at a finer scale than the original dataset. This brought the resolution of the NO_2 datasets closer to the resolution of 1 m, but there may still be inconsistencies due to the large differences in resolution size. There was not enough coverage available from ground-based monitoring stations, meaning that this was still the only option regardless. Another challenge that comes along with using satellite-based measurements is the temporal scale that is observed. The specific satellite used for this study, the Sentinel 5p, only passes over each location once a day at around noon. This means that there is no continuous recording of NO_2 concentrations at the ground. NO_2 is significantly impacted by diurnal cycles, as it is removed through the environment through photochemical pathways (Boersma et al., 2009; Chen et al., 2021; Pancholi et al., 2018). Not being able to observe these cycles is a downside, as there could have been a pattern that we missed if this was able to be recorded.

It is also worth noting that there was a major health crisis in the form of COVID-19 within the time frame of data collection. There was a marked decrease in motor vehicle use during this time that could cause NO_2 concentrations to be overall skewed for some locations (Ghahremanloo et al., 2021).

6 Conclusion

In this study, we sought to answer three research questions about the relationship between fine-scale urban form and air quality. First, we explored whether this relationship is consistent across cities. Our findings suggest that the impact of urban form on air quality varies across cities and depends on factors such as traffic congestion and land classification. For example, we observed that buildings had a relatively large impact on NO_2 concentrations in cities with low congestion, while roads and their continuity in the landscape held importance when this land classification was analyzed. Model results were also extensively less stable when used on areas of high congestion, having lower R-Squared values and higher amounts of error.

Second, we investigated how the relationship is affected by the spatial extent or buffer zone at which urban form is captured. Our analysis showed that the choice of buffer radius can affect the stability and accuracy of model predictions. We found that observations collected from a 1500m buffer radius had slightly better model results and higher R-Squared values compared to observations collected from a 500m buffer radius. However, we also observed that 500m buffers can perform better for certain land classifications and in some congestion levels. Therefore, urban planners and designers should be aware of the limitations and trade-offs associated with different buffer sizes when making decisions about urban form and air quality.

Finally, we examined how the relationship changes with the season. Our study found that winter had the highest overall concentration of NO_2, likely due to the lack of solar radiation, which is one of the main pathways for NO_2 removal from the atmosphere. Additionally, we observed that there are different consistencies between importance of variables depending on the season and which buffer radius was being used for analysis.For example, there were some

inconsistencies with different metrics of importance that were seen regardless of landcover type that were more associated with observations collected within a 500m buffer radius that were not seen when using a 1500m buffer. Seasonality also caused inconsistent variable importance when looking at land covers such as roads or trees.

Overall, our study contributes to a better understanding of the complex relationship between fine-scale urban form and air pollution. It highlights the need for city-specific approaches to reduce air pollution and emphasizes the importance of considering both temporal and spatial scales in urban planning and design decisions aimed at improving air quality. It also shines a light on the importance congestion levels can have on a city when making these design decisions.

Works Cited

Ajtai, N., et al. (2020). Support Tools for Land Use Policies Based on High Resolution Regional Air Quality Modelling. *Land Use Policy*, 95, 103909. https://doi.org/10.1016/j.landusepol.2019.03.022

Agudelo-Castañeda, D.M., et al. (2013). Measurement of Particle Number and Related Pollutant Concentrations in an Urban Area in South Brazil. *Atmospheric Environment*, 70, 254-262. https://doi.org/10.1016/j.atmosenv.2013.01.029

Baró, F., et al. (2015). Mismatches Between Ecosystem Services Supply and Demand in Urban Areas: A Quantitative Assessment in Five European Cities. *Ecological Indicators*, 55, 146-158. https://doi.org/10.1016/j.ecolind.2015.03.013

Basaric, V., et al. (2011). Effects of Traffic on NO2 and PM10 Emissions in Novi Sad. *Polish Journal of Environmental Studies*, 23(5), 1837-1842.

Bechle, M. J., Millet, DB, & Marshall, J.D. (2011). Effects of income and urban form on urban NO2: Global evidence from satellites. *Environmental Science & Technology*, 45.11, 4914-4919. https://doi.org/10.1021/es103866b

Bechle, M. J., Millet, DB, & Marshall, J.D. (2013). Remote Sensing of Exposure to NO2: Satellite Versus Ground-Based Measurement in a Large Urban Area. *Atmospheric Environment*, 69, 345-353. https://doi.org/10.1016/j.atmosenv.2012.11.046

Bechle, M. J., Millet, DB, & Marshall, J.D. (2017). Does urban form affect urban NO2? Satellite-based evidence for more than 1200 cities. *Environmental Science & Technology*, 51.21, 12707-12716. https://doi.org/10.1021/acs.est.7b01194

Bereitschaft, B., and Keith, D. (2013). Urban Form, Air Pollution, and CO2 Emissions in Large U.S. Metropolitan Areas. *The Professional Geographer*, 65.4, 612-635. https://doi.org/10.1080/00330124.2013.799991

Bertazzon, S., Johnson, M., Eccles, K., & Kapplan, G.G. (2015). Accounting for spatial effects in land use regression for urban air pollution modeling. *Spatial and spatio-temporal epidemiology*, 14, 9-21. https://doi.org/10.1016/j.sste.2015.06.002

Boersma, K. F., et al. (2009). Validation of urban NO2 concentrations and their diurnal and seasonal variations observed from the SCIAMACHY and OMI sensors using in situ surface measurements in Israeli cities. *Atmospheric Chemistry and Physics*, 9.12, 3867-3879. https://doi.org/10.5194/acp-9-3867-2009

Bozkurt, Z., Üzmez, Ö. Ö., Döğeroğlu, T., Artun, G., & Gaga, E. O. (2018). Atmospheric concentrations of SO2, NO2, ozone and VOCs in Düzce, Turkey using passive air samplers: Sources, spatial and seasonal variations and health risk estimation. *Atmospheric Pollution Research*, 9.6, 1146-1156. https://doi.org/10.1016/j.apr.2018.05.001

Cakmak, S., et al. (2016). The modifying effect of socioeconomic status on the relationship between traffic, air pollution and respiratory health in elementary schoolchildren. *Journal of Environmental Management*, 177, 1-8. https://doi.org/10.1016/j.jenvman.2016.03.051

Chen, B., & Lannone III, B.V. (2020). [FR431] FRAGSTATS: A Free Tool for Quantifying and Evaluating Spatial Patterns: FOR362/FR431, 6/2020. *EDIS*, 2020.6, 9-9. doi.org/10.32473/edis-fr431-2020

Chen, B., Yang, S., & Xiang-De, X. (2014). *Theoretical and Applied Climatology*, 117.1-2, 29-39. doi: 10.1007/s00704-013-0982-1.

Chen, G., Li, R., Guoqian Z. (2021) Impacts of urban geometry on outdoor ventilation within idealized building arrays under unsteady diurnal cycles in summer. *Building and Environment*, 206, 108344. https://doi.org/10.1016/j.buildenv.2021.108344

Dai, A., et al. (2023). Effect of different plant communities on NO2 in an urban road greenbelt in Nanjing, China. *Scientific Reports*, 13.1, 3424. https://doi.org/10.1038/s41598-023-30488-0

Edussuriya, P., Chan, A., Malvin, A. (2014). Urban morphology and air quality in dense residential environments: Correlations between morphological parameters and air pollution at street-level. *Journal of Engineering Science and Technology*, 9.1, 64-80.

Elith, J., Leathwick, J.R., Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, 77.4, 802-813, https://doi.org/10.1111/j.1365-2656.2008.01390.x

Eum, K.D. (2019). Long-Term NO2 Exposures and Cause-Specific Mortality in American Older Adults. *Environment International*, 124, 10-15. https://doi.org/10.1016/j.envint.2018.12.060

Fan, C. (2018). Examining the impacts of urban form on air pollutant emissions: Evidence from China. *Journal of environmental management*, 212, 405-414. https://doi.org/10.1016/j.jenvman.2018.02.001

Fan, C., & Myint, S. (2014). A Comparison of Spatial Autocorrelation Indices and Landscape Metrics in Measuring Urban Landscape Fragmentation. *Landscape and urban planning*, 121, 117-128. https://doi.org/10.1016/j.landurbplan.2013.10.002

Fang, C., et al. (2015). Estimating the impact of urbanization on air quality in China using spatial regression models. *Sustainability*, 7.11, 15570-15592. https://doi.org/10.3390/su71115570

Friedman, J. H. (2002). Stochastic gradient boosting. *Computational statistics & data analysis*, 38.4, 367-378. https://doi.org/10.1016/S0167-9473(01)00065-2

Gao, J., Li, S. (2011). Detecting spatially non-stationary and scale-dependent relationships between urban landscape fragmentation and related factors using Geographically Weighted Regression. *Applied Geography*, 31.1, 292-302. https://doi.org/10.1016/j.apgeog.2010.06.003

Geffen, J. (2020). S5p TROPOMI NO2 slant column retrieval: Method, stability, uncertainties and comparisons with OMI. *Atmospheric Measurement Techniques*, 13, 1315-1335. https://doi.org/10.5194/amt-13-1315-2020

Ghahremanloo, M., et al. (2021). Impact of the COVID-19 Outbreak on Air Pollution Levels in East Asia. *The Science of the total environment*, 40.11, 2751-2769. https://doi.org/10.1016/j.scitotenv.2020.142226

Griffith, D.A. (2008). Spatial-filtering-based contributions to a critique of geographically weighted regression (GWR). *Environment and Planning*, 40.11, 2751-2769. doi:10.1068/a38218

Gulia, S., et al. (2015). Urban Air Quality Management-A Review. *Atmospheric pollution research*, 6.2, 286-304. https://doi.org/10.5094/APR.2015.033

Gualtieri, G., (2015). A statistical model to assess air quality levels at urban sites. *Water, Air, & Soil Pollution*, 226, 1-15. https://doi.org/10.1007/s11270-015-2663-4

Hastie, T., Friedman, J., & Tibshirani, R. (2009). Random forests. *The elements of statistical learning: Data mining, inference, and prediction*. In Hastie, T., Friedman, J., & Tibshirani, R., *The Elements of Statistical Learning* (pp. 587-604). New York- Springer.

Ikram, M., Yan, Z., Liu, Y., Qu, W. (2015). Seasonal effects of temperature fluctuations on air quality and respiratory disease: a study in Beijing. *Natural Hazards*, 79, 833-853. https://doi.org/10.1007/s11069-015-1879-3

Irga, P.J., Burchett, M.D., Torphy, F.R. (2015). Does Urban Forestry Have a Quantitative Effect on Ambient Air Quality in an Urban Environment?. *Atmospheric environment*, 120, 173-181. https://doi.org/10.1016/j.atmosenv.2015.08.050

Jayamurugan, R., et al. (2013). Influence of temperature, relative humidity and seasonal variability on ambient air quality in a coastal urban area. *International Journal of Atmospheric Sciences*, 2013. https://doi.org/10.1155/2013/264046

John H. Seinfeld (1989). Urban Air Pollution: State of the Science. *Science*, 246, 745-752.

Kang, J.E., Yoon, D.K., & Bae, H.J. (2015). Evaluating the effect of compact urban morphology on air quality in Korea. *Environment and Planning B: Urban Analytics and City Science*, 122, 133-141. https://doi.org/10.1177/2399808317705880

Kendrick, C.M., Koonce, P., & George, L.A. (2015). Diurnal and seasonal variations of NO, NO2 and PM2. 5 mass as a function of traffic volumes alongside an urban arterial. *Atmospheric Environment*, 122, 133-141. https://doi.org/10.1016/j.atmosenv.2015.09.019

Klinberg, J., et al. (2017). Influence of urban vegetation on air pollution and noise exposure–a case study in Gothenburg, Sweden. *Science of the Total Environment*, 599, 1728. https://doi.org/10.1016/j.scitotenv.2017.05.051

Ku, C.A. (2020). Exploring the Spatial and Temporal Relationship between Air Quality and Urban Land-Use Patterns Based on an Integrated Method. *Sustainability*, 12.7, 2964. https://doi.org/10.3390/su12072964

Lähde, T., et al. (2014). Mobile particle and NOx emission characterization at helsinki downtown: Comparison of different traffic flow areas. *Aerosol and Air Quality Research*, 14.5, 1372-1382. doi: 10.4209/aaqr.2013.10.0311

Lai, L.W., Cheng, W.L. (2009). Air quality influenced by urban heat island coupled with synoptic weather patterns. *Science of the total environment*, 407.8, 2724-2733. https://doi.org/10.1016/j.scitotenv.2008.12.002

Lamsal, L.N., et al. (2013). Scaling relationship for NO2 pollution and urban population size: a satellite perspective. *Environmental science & technology*, 47.14, 7855-786. https://doi.org/10.1021/es400744g

Leathwick, J. R., et al. (2006). Variation in demersal fish species richness in the oceans surrounding New Zealand: an analysis using boosted regression trees. *Marine ecology progress series*, 321, 267-281. doi:10.3354/meps321267

Lee, J.H., et al. (2014). Land use regression models for estimating individual NOx and NO2 exposures in a metropolis with a high density of traffic roads and population. *Science of the total environment*, 472, 1163-1171. https://doi.org/10.1016/j.scitotenv.2013.11.064

Li, F., & Zhou, T. (2019). Effects of urban form on air quality in China: An analysis based on the spatial autoregressive model. *Cities*, 89, 130-140. https://doi.org/10.1016/j.cities.2019.01.025

Lian, 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, 1-3. https://doi.org/10.1038/s41598-020-74524-9

Liang, Z., Wei, F., Wang, Y., Huang, J., Jian, H., Sun, F., & li.., S. (2022) The context-dependent effect of urban form on air pollution: A panel data analysis. *Remote Sensing.*, 14.1, 7. https://doi.org/10.3390/rs12111793

Liu, Y., He, L., Qin, W., Lin, A., & Yang, Y. (2022). The effect of urban form on PM2. 5 concentration: Evidence from china's 340 prefecture-level cities. *Remote Sensing* 14.1, 7., https://doi.org/10.3390/rs14010007

Lu, C., & Liu, Y. (2016). Effects of China's Urban Morphology on Urban Air Quality. *Urban Studies*, 53.12, 2607. doi: 10.1177/0042098015594080

Madrazo, J., Clappier, A., Cuesta, O., Belecázar, L.C., González, Y., Bolufe, J., Sosa, C., Carillo, E., Manso, R., Canciano, J., & Golay, F. (2019) Evidence of traffic-generated air pollution in Havana. *Atmósfera*, 32.1, 15-24. doi: 10.20937/ATM.2019.32.01.02

Mao, S., Zhou, Y., Gao, W., Jin, Y., Zhao, H., Luo, Y., Chen, S., Chen, X., Zhang, G., Lun, F., Pan, Z., An, P. (2022). Ventilation capacities of chinese industrial cities and their influence on the concentration of NO2. *Remote Sensing*, 14.13, 3348. https://doi.org/10.3390/rs14143348

Manes, F., Marando, F., Capotori, G., Blasi, C., Salvatori, E., Fusaro, L., Ciancarella, L., Mircea, M., Marchetti, M., Chirici, G., Mufano, M. (2016). Regulating ecosystem services of forests in ten Italian metropolitan cities: air quality improvement by PM10 and O3 removal. *Ecological indicators*, 67, 425-440. https://doi.org/10.1016/j.ecolind.2016.03.009

McCarty, J., Kaza, N. (2015). Urban form and air quality in the United States. *Landscape and Urban Planning*, 139, 168-179. https://doi.org/10.1016/j.landurbplan.2015.03.008

Novotny, E. V., Bechle, M. J., Millet, Marshall, J. D. (2011). *Environmental science & technology*, 45.10, 4407-4414. https://doi.org/10.1021/es103578x

Pancholi, P., Kumar, A., Bikundia, D. S., and Chourasiya, S. (2018). An observation of seasonal and diurnal behavior of O3–NOx relationships and local/regional oxidant (OX = O3 + NO2) levels at a semi-arid urban site of western India. *Sustainable Environment Research*, 28.2, 79-89. https://doi.org/10.1016/j.serj.2017.11.001

Peng, Y., Gao, Z., Buccolieri, R., Shen, J., & Ding, W. (2021). Analysis of the NO2 tropospheric product from S5P TROPOMI for monitoring pollution at city scale. *Sustainable Cities and Society*, 67, 102735. https://doi.org/10.1016/j.scs.2021.102735

Prunet, P., Lezeaux, O., Camy-Peyret, C., & Thevenon, H. (2020) Analysis of the NO2 tropospheric product from S5P TROPOMI for monitoring pollution at city scale. *City and Environment Interaction*, 8, 100051. https://doi.org/10.1016/j.cacint.2020.100051

Rahman, Md. M., Najaf, P., Fields, M.J., Thill., J.C. (2020). Traffic congestion and its urban scale factors: Empirical evidence from American urban areas. *International journal of sustainable transportation*, 16.5, 406-421. https://doi.org/10.1080/15568318.2021.1885085

Roberts-Semple, D., Song, F., Gao, Y. (2012). Seasonal characteristics of ambient nitrogen oxides and ground–level ozone in metropolitan northeastern New Jersey. *Atmospheric Pollution Research*, 3.2, 247-257. https://doi.org/10.5094/APR.2012.027

Roeland, S., Moretti, M., Amorim, J. H., Branquinho, C., Fares, S., Morelli, F., Niinemets, Ü., Paoletti, E., Pinho, P., Sgrigna, G., Stojanovski, V., Tiwary, A., Sicard, P., & Calfapietra, C. (2019). Towards an integrative approach to evaluate the environmental ecosystem services provided by urban forest. *Journal of Forestry Research*, 30, 1981-1996. https://doi.org/10.1007/s11676-019-00916-x

Rossi, R., Cecceto, R., & Gastaldi, M. (2020). Effect of road traffic on air pollution. Experimental evidence from COVID-19 lockdown. *Sustainability*, 12.21, 8984. https://doi.org/10.3390/su12218984

Salas, R., Perez-Villadoniga, M. J., Prieto-Rodriguez, J., & Russo, A. (2021). *Transportation Research Part D: Transport and Environment*, 16.5, 406-421. https://doi.org/10.1016/j.trd.2020.102689

Seto, K. C., Fragkias, M. (2005). Quantifying Spatiotemporal Patterns of Urban Land-Use Change in Four Cities of China with Time Series Landscape Metrics. *Landscape Ecology*, 20.7, 871-888. https://doi.org/10.1007/s10980-005-5238-8

Shi, K., Shen, J., Wang, L., Ma, M., Cui, Y. (2020). A multiscale analysis of the effect of urban expansion on PM2. 5 concentrations in China: Evidence from multisource remote sensing and statistical data. *Buildings and Environment*, 174, 106778. https://doi.org/10.1016/j.buildenv.2020.106778

Shon, Z. H., Kim, K. H., Song, S. K. (2011). Long-term trend in NO2 and NOx levels and their emission ratio in relation to road traffic activities in East Asia. *Atmospheric Environment*, 45.18, 3120-3131. https://doi.org/10.1016/j.atmosenv.2011.03.009

Siddiqui, A., Halder, S., Chauhan, P., Kumar, P. (2020). COVID-19 Pandemic and City-Level Nitrogen Dioxide (NO2) Reduction for Urban Centres of India. *Journal of the Indian Society of Remote Sensing*, 48.7, 999-1006. https://doi.org/10.1007/s12524-020-01130-7

Song, C., Lee, W. K., Choi, H. A., Kim, J., Jeon, S. W., & Kim, J. S. (2016). Spatial assessment of ecosystem functions and services for air purification of forests in South Korea. *Environmental Science & Policy*, 63, 27-34. https://doi.org/10.1016/j.envsci.2016.05.005

Song, W., Jia, H., Li, Z., Tang, D., & Wang, C. (2019). Detecting urban land-use configuration effects on NO2 and NO variations using geographically weighted land use regression. *Atmospheric Environment*, 197, 166-176. https://doi.org/10.1016/j.atmosenv.2018.10.031

Stavroulas, I., Grivas, G., Michalopoulos, P., Liakakou, E., Bougiatioti, A., Kalkavouras, P., Fameli, K. M., Hatzianastassiou, N., Mihalopoulos, N., & Gerasopoulos, E. (2020). Field evaluation of low-cost PM sensors (Purple Air PA-II) under variable urban air quality conditions, in Greece. *Atmosphere*, 11.9, 926. https://doi.org/10.3390/atmos11090926

Su, S., Xiao, R., Zhang, Y. (2012). Multi-scale analysis of spatially varying relationships between agricultural landscape patterns and urbanization using geographically weighted regression. *Applied Geography*, 32.2, 360-375. https://doi.org/10.1016/j.apgeog.2011.06.005

Sun, K., Zhu, L., Cady-Pereira, K., Miller, C. C., Chance, K., Clarisse, L., Coheur, P. F., Abad, G. G., Huang, G., Liu, X., Damme, M. V., Yang, K., Zondlo, M. (2018). A physics-based approach to oversample multi-satellite, multispecies observations to a common grid. *Atmospheric Measurement Techniques*, 11.12, 6679-6701. https://doi.org/10.5194/amt-11-6679-2018

United Nations, *68% of the world population projected to live in urban areas by 2050, says UN*, *Department of Economic and Social Affairs*, Retrieved April 18, 2023 from https://www.un.org/development/desa/en/news/population/2018-revision-of-world-urbanization-prospects.html

Veefkind, J.P., Aben, I., McMullan, K., Forster, H., de Vries, J., Otter, G., Claas, J., Eskes, H. J., de Haan, J. F., Kleipool, Q., van Weele, M., Hasekamp, O., Hoogeveen, R., Landgraf, J., Snel,

R., Tol, P., Ingmann, P., Voors, R., Kruizinga, B., Vink, R., & Levet, P.F. (2012). TROPOMI on the ESA Sentinel-5 Precursor: A GMES mission for global observations of the atmospheric composition for climate, air quality and ozone layer applications. *Remote sensing of environment*, 120, 70-83. https://doi.org/10.1016/j.rse.2011.09.027

Wang, Q., Feng, H., Feng, H., Yue, Y., Li, J., Ning, E. (2021). The Impacts of Road Traffic on Urban Air Quality in Jinan Based GWR and Remote Sensing. *Scientific reports*, 11.1, 15512. https://doi.org/10.1038/s41598-021-94159-8

Wu, J., Shen, W., Sun, W., & Tueller, P. T. (2002). Empirical patterns of the effects of changing scale on landscape metrics. *Landscape ecology*, 17, 761-782. https://doi.org/10.1023/A:1022995922992

World Health Organization (2022, 19 December). *Ambient (outdoor) air pollution*. WHO. https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health

Xu, M., Sbihi, H., Pan, X., & Brauer, M. (2019). Local variation of PM2. 5 and NO2 concentrations within metropolitan Beijing. *Atmospheric Environment*, 200, 254-263. https://doi.org/10.1016/j.atmosenv.2018.12.014

Yang, R. M., Zhang, G. L., Liu, F., Lu, Y. Y., Yang, F., Yang, F., Yang, M., Zhao, Y. G., & Li, D.C. (2016). Comparison of boosted regression tree and random forest models for mapping topsoil organic carbon concentration in an alpine ecosystem. *Ecological Indicators*, 60, 870-878. https://doi.org/10.1016/j.ecolind.2015.08.036

Yang, J., Wen, Y., Wang, Y., & Seinfeld, J. H. (2021). From COVID-19 to future electrification: Assessing traffic impacts on air quality by a machine-learning model. *Proceedings of the National Academy of Sciences*, 118.26, e2102705118. https://doi.org/10.1073/pnas.2102705118

Yasin, M. Y., Yusoff, M. M., Abdullah, J., Noor, N. M., & Noor, N. M. (2021). Urban sprawl literature review: Definition and driving force. *Geografia*, 17.2, 116-128. https://doi.org/10.17576/geo-2021-1702-10

Zhang, Y., Chen, G., Myint, S. W., Zhou, Y., Hay, G. J., Vukomanovic, J., Meentemeywer, R. K. (2022). UrbanWatch: A 1-Meter Resolution Land Cover and Land Use Database for 22 Major Cities in the United States. *Remote sensing of environment*, 278, 113106. https://doi.org/10.1016/j.rse.2022.113106

Zhou, H., & Gao, H. (2020). The impact of urban morphology on urban transportation mode: A case study of Tokyo. *Case Studies on Transport Policy*, 8.1, 197-205. https://doi.org/10.1016/j.cstp.2018.07.005

Zou, B., Xu, S., Sternberg, T., & Fang, X. (2016). Effect of land use and cover change on air quality in urban sprawl. *Sustainability*, 8.7, 677. https://doi.org/10.3390/su8070677