Traffic Tracking and Air Quality: A Holistic Approach to Predicting Traffic-Related Air Pollution.

by

Laura Ahinee Deveer

BSc., Kwame Nkrumah University of Science and Technology, 2021

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

MASTER OF APPLIED SCIENCE

in the Department of Civil Engineering

University of Victoria

Traffic Tracking and Air Quality: A Holistic Approach to Predicting Traffic-Related Air Pollution.

by

Laura Ahinee Deveer

BSc., Kwame Nkrumah University of Science and Technology, 2021

Supervisory Committee

Dr Laura Minet, Supervisor

Department of Civil Engineering

Dr Rishi Gupta, Departmental Member

Department of Civil Engineering

Abstract

On-road transportation has long been a significant contributor to air pollution in cities. Over the past few decades, the transportation sector, which has been targeted by major regulations, has undergone substantial changes. These changes include a shift in vehicle fleet composition and both natural and artificial alterations to traffic patterns. Despite the importance of on-road transportation for urban mobility, it remains a major source of air pollution and a public health challenge. Therefore, it is essential to accurately measure and model the temporal and spatial distribution of traffic-related air pollution. In that way, targeted implementations can be made to combat the adverse health effects associated to the exposure to these pollutants. In this thesis, we investigate the potential of low-cost methods to accurately estimate air pollutant concentrations. To this end, we employed modern traffic tracking technologies with low-cost sensors and machine learning techniques. The research addresses the effectiveness of leveraging these technologies to understand the factors and interactions that influence air quality. Chapter 2 predicts real-time air pollutant concentrations with high accuracy across pollutants using traffic videos and machine learning algorithms. The results reveal the superior performance of non-linear models over linear models. In addition, the Shapley additive explanation plots employed in this study effectively captured the intricate relationships between pollutants and their predictors. Chapter 3 examines the influences of traffic, particularly from cruise ship activities, on local air quality in James Bay, Victoria, BC. Results indicate that both emissions from traffic and cruise ship activities affected air quality. The integration of low-cost sensors with these traffic tracking technologies proves crucial for accurate air quality analysis and allows for context-specific and real-time assessments, providing valuable insights for policy makers and urban planners.

Table of Contents

List of Tables

List of Figures

Acknowledgments

I would like to express my deepest gratitude to my supervisor, Dr Laura Minet, for her guidance, encouragement, and expertise throughout the course of this research. It has been a pleasure to be part of the CleanAir lab and participate in these amazing projects. I am also thankful to my committee member, Dr Rishi Gupta, for his insightful feedback and support.

A heartfelt thanks to my colleagues and friends who provided both moral and technical support, making this journey more manageable. Finally, I am forever indebted to my family for their unwavering encouragement and understanding, without which this thesis would not have been possible.

CHAPTER 1

Introduction

1.1 Background

Over the years, the development of transportation systems, specifically improved road networks, has encouraged higher volumes of vehicles on the road, which have led to poorer air quality. In urban areas, traffic is one of the major contributors to outdoor air pollution (Guarnieri and Balmes 2014; Sanchez et al. 2020). Vehicles emit particles and gaseous pollutants directly into the ambient air, as either exhaust emissions or non-exhaust emissions. Exhaust emissions are pollutants that are released into the air as a result of fuel combustion in engines and include pollutants such as carbon monoxide (CO), nitrogen oxides (NO, NO_2, and NOx), which are produced when nitrogen in the air reacts with oxygen at high temperatures, along with particulate matter (particularly $PM_{2.5}$ and PM_{10}). Non-exhaust emissions, on the other hand, come from road dust suspension, tire and break wear, and evaporative emissions (such as volatile organic compounds) (Frosina et al. 2018). Traffic-related air pollution (TRAP) describes the complex mixture of these pollutants emitted from vehicles and contributes significantly to the degradation of air quality.

Reports from Health Canada (2021) estimated that on-road traffic contributed to 28% of all CO emissions and 21% of all NO emissions in Canada, making them the largest contributor among all transportation classes. Additionally, it was reported that volatile organic compounds contributed to 7% of emissions, while less than 2% were from particulate matter. Given that about 4 out of 10 people in Canada live within 250 meters of high-traffic roadways, human exposure to TRAP is associated with a wide range of adverse health effects (Health Canada 2021). These effects include, but are not limited to, cancer, premature mortality, respiratory issues, and neurological and cognitive disorders (Allen et al. 2017; Han et al. 2017; Thurston et al. 2017). It is estimated that every year, TRAP contributes to approximately 1,200 premature deaths, 210,000 asthma symptom days and 2.7 million acute respiratory symptom days in Canada. Furthermore, the socioeconomic cost of health impacts of TRAP was estimated at $9.5B per year, based on 2015 currency (Health Canada 2021).

Khan et al. (2018)'s review of exposure assessment for TRAP revealed several techniques that have been employed. These approaches range from traditional methods, including ground-level monitoring and satellite remote sensing, to deterministic and stochastic modelling techniques. For instance, previous studies have relied on measurements from high-end fixed air quality monitoring stations, which are considered representative of a given area (Elliot et al. 2016; Santiago et al. 2021). In Canada, under the National Air Pollution Surveillance program established in 1969, there are 286 monitoring sites in 203 communities located in every province and territory (Environment and Climate Change Canada 2022). Given Canada's size, it is evident from these statistics that only few stations are deployed per city, mostly in the densely populated

areas or near city centers. The high maintenance costs, limited spatial representativeness, and low temporal resolution data from these air quality monitoring stations have led to a clear shift towards other methodologies for exposure assessments (H. Z. Li et al. 2019; Santiago and Martilli 2010). Satellite remote sensing, also used as a ground-level measurement technique, has been criticized for the high costs involved in obtaining the equipment and uncertainties in the data retrieved, as it does not necessarily represent the ground truth of air quality. These drawbacks have led to the emergence of air quality monitors equipped with low-cost gas sensors, which are cost-effective and commercially available (Karagulian et al. 2019; Snyder et al. 2013). The underlying concepts of low-cost sensors, on which air quality monitors are based, comprise a cluster of technologies that provide numerous advantages such as device compactness, low power demand, and minimal need for maintenance (Zimmerman et al. 2018). Furthermore, the automatic operation of air quality monitors allows for ease of deployment in multiple areas, resulting in a much denser air quality monitoring network (Si et al. 2020).

Other studies have also used modelling approaches including deterministic models, which mathematically explore the relationship between variables by using the knowledge of physical and biological mechanisms (Mueller and Mallard 2011; Wang et al. 2014; Wu et al. 2014). While these established theories offer useful insights into TRAP and their predictors, they come with complex prior knowledge requirements, unreliable and limited data, and various usage limitations (Chen et al. 2013). They do not sufficiently explain the nonlinearity of the many factors related to pollutant formation. On the other hand, machine learning models have the advantage of handling non-linear features, which results in more accurate estimations. In addition, they can be applied without needing an in-depth understanding of the dynamic and chemical processes between air pollutants and other relevant atmospheric variables (X. Li et al. 2019; Liu et al. 2018).

An important aspect of estimating TRAP is traffic data; however, it is quite challenging to gather due to factors like real-time data availability, high equipment cost, and complexity of traffic patterns. Modern traffic data collection employs a variety of technologies, from simple inductive loop sensors and GPS devices to intelligent transportation systems, Bluetooth and Wi-Fi digital radio transmitters, and surveillance videos (Forehead and Huynh 2018). These technologies provide cost savings, better coverage, and increased accuracy compared to more labor-intensive data collection methods like manual counting.

1.2 Research Questions

The research questions that guided the chapters in this thesis include:

- How can low-cost traffic tracking technologies, such as video surveillance and tube counters, improve real-time prediction of TRAP?
- How can machine learning techniques be utilized to understand the features that influence air quality and accurately predict air pollution?

- How do traffic influences, such as those from cruise ship activities, contribute to air quality?

1.3 Summary of Aims

In the context of integrating low-cost techniques for estimating TRAP, this thesis aims to utilize novel methods to accurately estimate TRAP. In Chapter 2, we employ traffic data extracted from video surveillance using object detection techniques, alongside air pollutant concentrations data obtained from low-cost sensors, as well as machine learning algorithms, to predict air pollutant levels. Chapter 3 investigates the relationship between traffic induced from cruise ship activities on the local air quality in James Bay, Victoria BC. To this end, air pollutant concentration data was collected using low-cost sensors, and traffic data was gathered from tube counters, allowing for comprehensive correlation analysis. The results presented in these chapters emphasize the value of novel data sources and their combinations in deepening our understanding of TRAP and its influence on air quality. By leveraging these innovative approaches, the research provides valuable insights into the dynamics between traffic patterns and air pollution, offering potential pathways for effective air quality management.

CHAPTER 2

Research Paper 1 – 'Real-Time Air Quality Prediction Using Traffic Videos and Machine Learning'

This chapter is submitted for publication. Entitled 'Real-Time Air Quality Prediction Using Traffic Videos and Machine Learning'. Its main author is also the author of this thesis, having done its research, calculations and writing in consultation with Dr. Laura Minet as the primary advisor. All section, equation, and reference numbering has been modified to integrate it with this thesis.

Abstract

Given the wide range of health effects associated with traffic related air pollution (TRAP), accurately estimating its concentrations is crucial. Traditional techniques rely on statistical models that tend to oversimplify the complex nature of air pollutant concentrations, whereas machine learning (ML) techniques yield better accuracies. However, required data inputs, particularly complex traffic data, are costly and mostly not collected real-time. This study develops ML models to accurately predict air pollutant concentrations by extracting traffic variables necessary for predicting TRAP solely from traffic videos. Fine particulate matter ($PM_{2.5}$), nitrogen dioxide (NO_2) and ozone (O_3) concentrations are recorded by a low-cost air quality sensor placed alongside a traffic camera. Traffic information, including volumes and speeds of different vehicle classes, is extracted using object detection and tracking algorithms. Emissions estimates are modelled in the Motor Vehicle Emission Simulator (MOVES). Air pollutant prediction models are developed using Extreme Gradient Boosting , random forest, and multiple linear regression models across six different predictor combinations. Our optimal models accurately predict real-time $PM_{2.5}$, NO_2, and O_3 concentrations with R-squared values (R^2) of 0.97, 0.98, and 0.94, respectively. The Shapley additive explanation (SHAP) is employed to identify the importance of the features impacting pollutant levels. Temperature and relative humidity have the most substantial impact on pollutant levels, followed by background concentrations. Traffic data, although not as influential as the other predictors, was relevant to the changes in pollutant levels. This study demonstrates a cost-effective approach with high

accuracies in predicting real-time air quality using a low-cost and maintenance tool: a video camera.

2.1 Introduction

Studies have identified on-road transportation as one of the most significant sources of air pollution in cities (Fei, Hsiao, and Chen 2024; Ho, Hsiao, and Chen 2023; Lei et al. 2024; Martín-Baos et al. 2022; Song et al. 2019). Pollutants such as carbon monoxide (CO), volatile organic compounds (VOCs), particulate matter (e.g. PM_{10} and $PM_{2.5}$) and nitrogen oxides (NO_x) account for a majority of traffic-related air pollution (TRAP) and are associated with adverse health effects including asthma, impaired lung function and increased cardiovascular morbidity and mortality (Ohlwein et al. 2019; Pan, Yao, and Yang 2016; Thurston et al. 2017). With 12 deaths per 100,000 people recorded, air pollution is the 13th leading mortality risk factor in Canada and exposure to $PM_{2.5}$, ozone (O_3) and nitrogen dioxide (NO_2) led to 15,300 premature deaths in 2016 (Health Canada 2021). For this reason, it is crucial for public health authorities to accurately estimate the impacts of emissions from on-road transport on ambient air quality in order to propose solutions to minimize their health effects (Martín-Baos et al. 2022).

The accurate estimation of air pollutant concentrations at a refined spatio-temporal resolution can be challenging due to the complex spatial and temporal dynamic changes of air pollutant composition (Zhang et al., 2022). Approaches to predicting air pollutant concentrations are divided into two main perspectives: deterministic (or mechanistic) and statistical (Hoshyaripour et al., 2016; Wang et al., 2017; Yang & Huang, 2024). Deterministic models, such as the Community Multiscale Air Quality (CMAQ) (United States Environmental Protection Agency. 2022) and the Weather Research and Forecasting model coupled with Chemistry (WRF- Chem) (Powers et al. 2017), rely on theoretical physical and chemical principles to simulate the dispersion and chemical reactions of air pollutants and estimate atmospheric air pollutant concentrations. Although these models offer valuable insights into the complexities of air pollution prediction, they depend heavily on complex input information, which restricts their usability (Feng et al. 2015). To address this shortfall, several studies have evaluated effective alternatives to deterministic models through statistical approaches, including classical statistical and machine learning models, which utilize probability, statistics, and stochastic processes for forecasting air pollutant concentrations (Chen et al., 2018 ; Kamińska, 2018; Shams et al., 2020, 2021; Wang et al., 2020). Classical models, for example Auto Regressive Integrated Moving Average (ARIMA), combine the moving averages and autoregressive components to model historical data. However, they require a large amount of statistical data for accurate predictions, and their inability to capture non-linear relationships limits their prediction accuracy (Jian et al., 2012; Zhang et al., 2018). In recent decades, traditional machine learning (ML) approaches, such as support vector regression (Li et al., 2019; Moazami et al., 2016), random forest (Kamińska, 2018; Liu et al., 2018), and artificial neural networks (ANN) (Alimissis et al. 2018; Bai et al. 2016; Feng et al. 2015), have been employed to overcome the limitations of classical models by capturing the non-linear relationships between pollutants and their wide range of

spatial and temporal predictors. In their study, Wang et al. (2020) compared the prediction performance of a land use regression model, which is a multilinear regression model, and two ML models (ANN and gradient boost). Their study highlighted the superior performance of the ML techniques for predicting $PM_{2.5}$ and black carbon concentrations. Moreover, ML models have been found to capture well spatial and temporal variations of pollutant concentrations and can handle multiple air pollutants simultaneously (Masmoudi et al. 2020; Wen et al. 2019).

Despite the extensive studies employing ML models for air pollution prediction, effectively extracting different types of data, including detailed traffic information such as flow, speed and composition, which are important predictors of TRAP, remains a difficult task (Wen et al. 2019). Traditional traffic detectors, such as inductive loop-based, ultrasonic and geomagnetic detectors, can only capture specific points of traffic flow data, and require installation and maintenance, which limits their effectiveness in terms of accuracy, costs and ability to cover wide areas (Chen et al. 2021). The use of deep learning to extract traffic information from various visual sources has proven to be more efficient and economical in obtaining real-time traffic data (Kim 2022). Previous studies that have implemented this approach used data from unconventional sources, such as Google Images, which are updated infrequently and cannot provide real-time information. The rise of computer vision techniques has led to the utilization of video-based vehicle detection to analyze TRAP (García-González et al., 2021; Hong et al., 2020). For instance, Song et al. (2019) used such techniques to study the impact of vehicle patterns across different urban functional zones on air pollutant concentrations, but faced challenges such as time-consuming image labelling and low-camera quality. Fei et al. (2024) addressed some of these limitations with transfer learning on public datasets but struggled with occlusion and detection of certain vehicle types. The widespread deployment of low-cost video surveillance has served as an alternative by generating large datasets for more context-specific models (Fedorov et al., 2019). Although these alternative approaches have been effective, gaps remain in optimizing models for extracting complex traffic information, such as accurately distinguishing between similar vehicle types, and efficiently processing image data. This study aims to address these gaps by leveraging custom datasets specific to the study area and improved image processing methods.

This study adopts a low-cost approach to predicting $PM_{2.5}$, NO_2 and O_3 concentrations using solely traffic cameras. The study utilises a custom dataset to identify a wide range of vehicle classes and more complex and detailed traffic variables that cannot be captured by traditional detectors. Furthermore, the study explores the interactions between various predictors to understand their influence in predicting air pollution, addressing a challenge evident in most studies that use ML models. In this regard, we developed three models: multiple linear regression models, random forest models and eXtreme Gradient Boosting Decision Tree models, using different combinations of predictors, including traffic volumes, traffic speeds, vehicle classes, background concentration levels, and meteorology. All data required for the development of the predictive models were acquired by leveraging a low-cost air quality monitor in combination

with traffic information (e.g., vehicle count, speed, type) obtained from a static traffic camera and associated emissions estimated from the U.S. Environmental Protection Agency (EPA) Motor Vehicle Emission Simulator (MOVES) model. Our study aims to demonstrate the effectiveness of ML techniques coupled with computer vision to estimate real-time TRAP concentrations, as opposed to previous studies that have relied on traditional approaches for monitoring TRAP and on statistical models, which fail to provide real-time information.

2.2 Methodology

Figure 2.2 provides a comprehensive overview of the methods employed in the study. Firstly, atmospheric data, including fine particulate matter ($PM_{2.5}$), nitrogen dioxide (NO_2) and ozone (O_3) concentrations, and temperature and relative humidity, were collected using a sensor-based air quality monitor and calibrated. Traffic information was extracted by employing an annotated custom dataset developed for Victoria (BC), Canada, and applied to two main algorithms to detect and track vehicles. Second-by-second speeds were also extracted and used to estimate vehicular emissions. Finally, the calibrated air pollutant concentrations, traffic volumes, speeds, emission estimates, meteorology, and background air pollutant concentrations were used to develop air pollutant prediction models.

2.2.1 Study area and data collection

The data collection took place along Gorge Rd in Victoria between October 2023 and April 2024. A map of the study location is shown in Figure 2.1. Gorge Road, which is a key route for local traffic, runs through the City of Victoria and parts of the District of Saanich, connecting various neighbourhoods. It provides access to downtown Victoria, residential areas, and recreational spots along the Gorge Waterway.

Minute-by-minute $PM_{2.5}$, NO_2, and O_3 concentrations, as well as meteorological data (i.e., temperature, relative humidity, and barometric pressure) were collected using the CTair air quality monitor, a low-cost sensor developed by Scentroid Inc. (Anon 2021). The instrument can be mounted outdoors on a wall or pole and sends gathered data via a 3G internet connection to a cloud-based software (SIMS3). The CTair, with its inlet pointing outward, was strategically positioned alongside a traffic camera, both mounted on a pole at approximately 3 meters above street level. Traffic videos were recorded with the Miovision camera (Miovision Technologies Inc. 2023) on non-rainy days for two weekdays in October 2023 and two weeks in March-April 2024 between 9:00 am and 6:00 pm. The videos had a resolution of 720 by 480 pixels and were relatively low-resolution, meaning that we could not distinguish details such as license plate

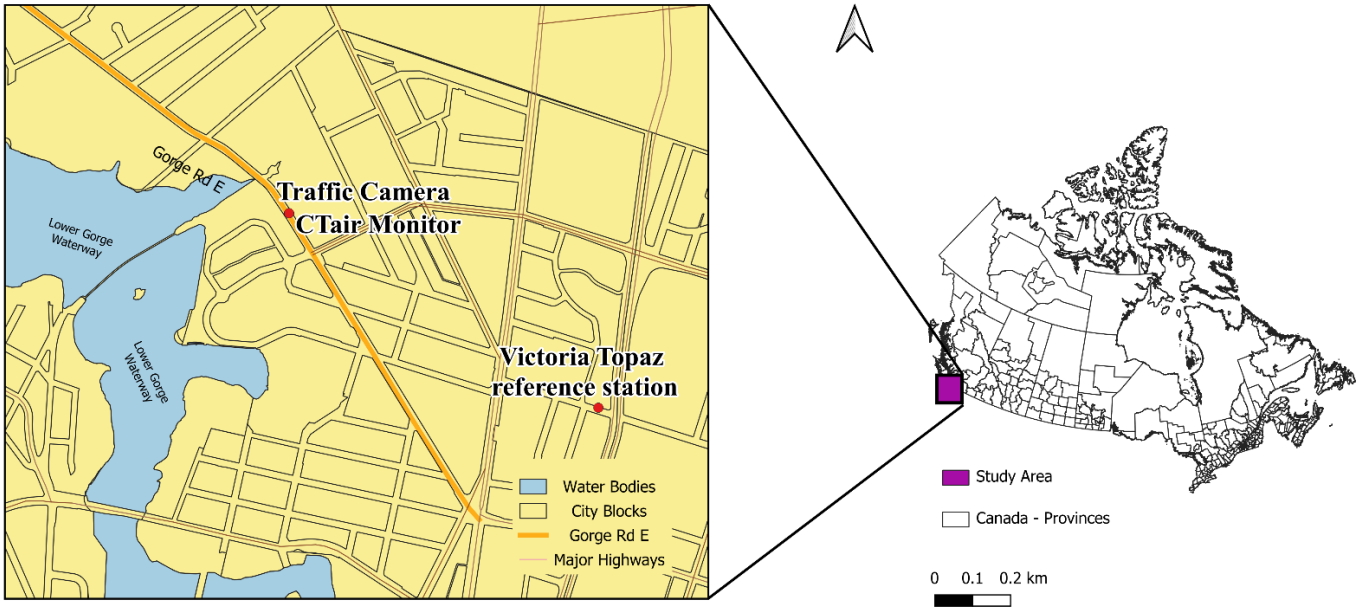


Figure 2.1. Map of the study area featuring Gorge Rd.

numbers, driver and passenger faces, or specific vehicle characteristics. Air pollutant data were recorded continuously throughout the monitoring period (October 2023 – April 2024). In March-April 2024, the SmartLinx (SLX) - Speed developed by Liveable Cities (Liveable Cities Smartlinx 2023) for measuring speeds was also deployed by the City of Victoria alongside the rest of the instrumentation.

2.2.2 Calibration of the Air Quality Monitor

One common drawback of low-cost sensors is their proneness to cross-sensitivity with other air pollutants (Castell et al. 2017; Kumar et al. 2015; Liang 2021; Spinelle et al. 2015). For example, O_3 has been observed to undergo oxido-reduction reactions in the presence of NO_2, which can interfere with measured concentrations (Mead et al. 2013). In addition, the performance of PM sensors that operate using light-scattering technology is influenced by the particle's composition, size and shape distribution, temperature, and relative humidity, leading to both long and short-term drift phenomena (Wang et al. 2015). To address these measurement inaccuracies, several studies have explored alternative approaches to calibrating low-cost sensors. Their results highlight the need to calibrate each sensor individually, since calibration performance highly depends on unique combinations of concentration levels, calibration techniques, pollutants studied, and local environmental factors (e.g., temperature and relative humidity) (Feinberg et al. 2018; Kelly et al. 2017).

The CTair was initially calibrated using data collected between January and March 2023 with data from the reference monitoring station located at Victoria Topaz, approximately 2.4

kilometers from the study area (Figure 2.1). At the Victoria Topaz station, the BAM (Beta Attenuation Monitor) instrument continuously measures minute-by-minute $PM_{2.5}$, O_3, NO_2, temperature, and relative humidity, following the provincial government's guidelines for operation and maintenance. Data collected throughout the colocation period (a total of 47,000 datapoints) were cleaned and filtered, which included the removal of missing values (114) from the BAM measurements which were associated with instrument breakdown events. Multilinear regression (MLR), neural network (NN), and random forest (RF) models were used to calibrate the low-cost sensors against the BAM instrument through an iterative process to minimize the error and loss between the actual and predicted values of the pollutant concentrations. Temperature, relative humidity, $PM_{2.5}$, NO_2, and O_3 concentrations recorded by the CTair were used as input features, and the $PM_{2.5}$, NO_2, and O_3 concentrations measured by the BAM at Topaz reference station as targets for the machine learning models' supervision. The models were then evaluated based on three performance metrics: coefficient of determination (R^2) (Equation 9), mean absolute error (MAE) (Equation 10) , and root mean squared error (RMSE) (Equation 11). The performances of the three models were compared, and the model with the minimum error for each of the pollutants analyzed was selected as the best model and applied to the concentrations collected by the CTair on Gorge Rd.

The next stage was to estimate background concentration levels. To achieve this, the spline of minimums methodology (Brantley et al. 2014; Fei et al. 2024; Larson et al. 2017) was employed to resolve measured concentrations collected at Victoria Topaz into background concentrations. This approach, adopted from Shairsingh et al. (2019), is a time-series based technique where smoothed data is applied to determine the minimum values. In this study, this method was applied in two steps: 1) separating smoothed measurements into discrete 5-minute windows and determining the minimum value in each window; and 2) fitting a smooth thin plate regression line to interpolate between the minimum concentrations. Measurements collected at the Topaz Station during the sampling period and thus corrected were then considered as background concentrations.

2.2.3 Extraction of Traffic Information

Development of a custom dataset

A total of 190 hours of videos recorded from MioVision traffic cameras deployed throughout Victoria were sourced from the City of Victoria to develop a custom dataset of annotated on-road vehicles images. Each frame of traffic video was treated as an image. The low-quality images were initially annotated using the Roboflow platform (Dwyer, B., Nelson, J., and Solawetz, J. 2022) into 10 different classes: passenger cars, passenger trucks, light commercial trucks, taxis, transit buses, garbage trucks, coach buses, cyclists, long haul trucks, and short haul trucks (shown in the Supplementary Information Figure S2.1). The data annotation involved drawing bounding boxes for each vehicle, assigning the appropriate label, and preparing the images in a format interpretable by the object detection model. Furthermore, data augmentation techniques

such as flipping, cropping, rotating, and blurring were applied to the dataset to increase the sample size, create additional variations of the dataset, and mitigate issues with occlusion and viewpoint variation (Shorten, Khoshgoftaar, and Furht 2021). This technique is also known to increase the robustness of object detection models by improving their capacity to be generalizable and more accurate in the prediction of real-world data.

Detection of vehicles in traffic videos

The convolutional neural network-based object detection algorithm You Only Look Once (YoloV8) (Khan et al. 2022) was employed to process the traffic video recordings. YoloV8 is widely popular for object detection tasks due to its high accuracy and efficient performance (Guo et al. 2023). The algorithm conducts real-time object detection in one stage by generating bounding boxes around detected objects and predicting the class probability of each object (Umair et al. 2021). The initial dataset comprised approximately 4,000 images, which were used for the initial training of the model, with 80% allocated for training and 20% for validation. After annotating the initial dataset, we utilized the auto-annotate function on the Roboflow platform to increase the size of the dataset. This function leverages a trained model to automatically label new datasets, thereby optimizing the annotation process. The final model was developed using a total 10,964 images. The vehicle detection model's performance was assessed using three accuracy measure indices:

- Precision (Equation 1) is the proportion of true positive predictions relative to the total number of predictions made by the model, including both true and false positives. True positives refer to a correct prediction where the model accurately identifies a positive instance, while false positives refer to an incorrect prediction where the model mistakenly identifies negative instances as positive.

$$\text{Precision} = \frac{(True\ Positive)}{(True\ Positive + False\ Positive)} \quad (1)$$

- Recall (Equation 2) quantifies the ratio of accurately predicted actual positives by the model, taking into account both true positives and false negatives. False negatives refer to incorrect predictions where the model fails to identify a positive instance, classifying them as negative.

$$\text{Recall} = \frac{(True\ Positive)}{(True\ Positive + False\ Negative)} \quad (2)$$

- Mean average precision mAP metric (Equation 3) provides the overall accuracy of the model, balancing both precision and recall.

$$\text{mAP} = \frac{1}{N}\sum_{i=1}^{N} AP \quad (3)$$

where :

- N is the number of classes
- AP is the average precision for class i, calculated as the area under the precision recall curve for that class.

Tracking and counting of vehicles in videos

To track the vehicles in traffic videos, multiple object tracking was conducted using the trained YoloV8 model in conjunction with the object tracker ByteTracker (Zhang et al. 2022). ByteTracker is a relatively new technique for efficient real-time object tracking and utilises one-shot detection-based method by combining object tracking and detection into a single model, as opposed to other trackers such as the SORT (Simple Online and Realtime Tracking) and DeepSORT (Abouelyazid 2023). The algorithm leverages deep learning to assign unique identification numbers (track IDs) to vehicles, ensuring that each vehicle maintains their respective track IDs as they move through the frames in the video. This state-of-the-art framework has been evaluated on several benchmark datasets, compared to the SORT and DeepSORT algorithms, and has demonstrated high accuracies and efficiencies (Ng, Goh, and Tee 2023; Zhang et al. 2022).

Vehicle counting begins after the detection phase and continues through the tracking phase. We drew a virtual horizontal line (shown in the Supplementary Information Figure S2.2) on the video footage and recorded the crossing of each vehicle class over the line in each direction of the road. We then validated the vehicle counting system by conducting Pearson correlation analyses between the model results, manual counts and counts obtained from the City of Victoria's counting equipment. We used 120 randomly selected one-minute video clips from the total traffic video dataset to assess the performance of the model. Pearson correlation analyses were conducted between the counts from the vehicle detection, manual and automatic counts.

Speed estimation

The final phase consisted of estimating the vehicle speeds from the traffic videos by measuring the travelling distance in a given unit of time. To ensure accurate speed estimation, it was necessary to correct the relative distortion associated with the use of cameras. This involved applying perspective transformation to remove relative distortion along the way and calculate the distance traveled by vehicles based on their coordinates across different frames. This approach employs a linear transformation matrix to adjust the image distance to the actual distance, thereby reducing the error in speed estimation by transforming each frame into a plain sheet so that the position of any vehicle can be determined from the pixel values (Afifah, Nasrin, and Mukit 2018). The homography matrix (H), which is a 3x3 matrix, contains the mapping points that transform coordinates from the source image to the target image (desired perspective) (Aburaddaha 2024). To determine the homography matrix, we used the fixed distances of streetlights along with the measured width of the road as reference points. We then identified the

corners (points) of the road by drawing a virtual polygon on the road (presented in the Supplementary Information Figure S2.3), and those points were used as the target. The matrix's structure is demonstrated as in Equation (4).

$$H = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \quad (4)$$

To apply the homography equation for transforming a coordinate from the source image to the corresponding point in the transformed image, we used Equation (5).

$$k' = H \cdot k \quad (5)$$

where :

- $k' = (x', y', 1)$ are the coordinates of the point in the transformed image,
- H is the 3×3 homography matrix,
- $k = (x, y, 1)$ are the coordinates of the corresponding point in the source image.

The distance between the points is then calculated using the Euclidean distance (Δd) formula (Equation 6).

$$\Delta d = \sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2} \quad (6)$$

where:

- (x_i, y_i) represents the horizontal and vertical coordinates of the target vehicle in frame i,
- (x_{i+1}, y_{i+1}) represents the horizontal and vertical coordinates of the target vehicle in the following frame i+1.

The speed is finally calculated as in Equation 7:

$$\text{Speed (km/hr)} = \frac{distance\ (m)}{time\ (s)} \times 3.6 \quad (7)$$

Relative and absolute error values were calculated by comparing the speed model results with the actual speed data obtained from the City of Victoria's equipment.

All phases of the traffic information extraction including detection, tracking, counting, and speed estimation were implemented in Python language using the OpenCV and Supervision library.

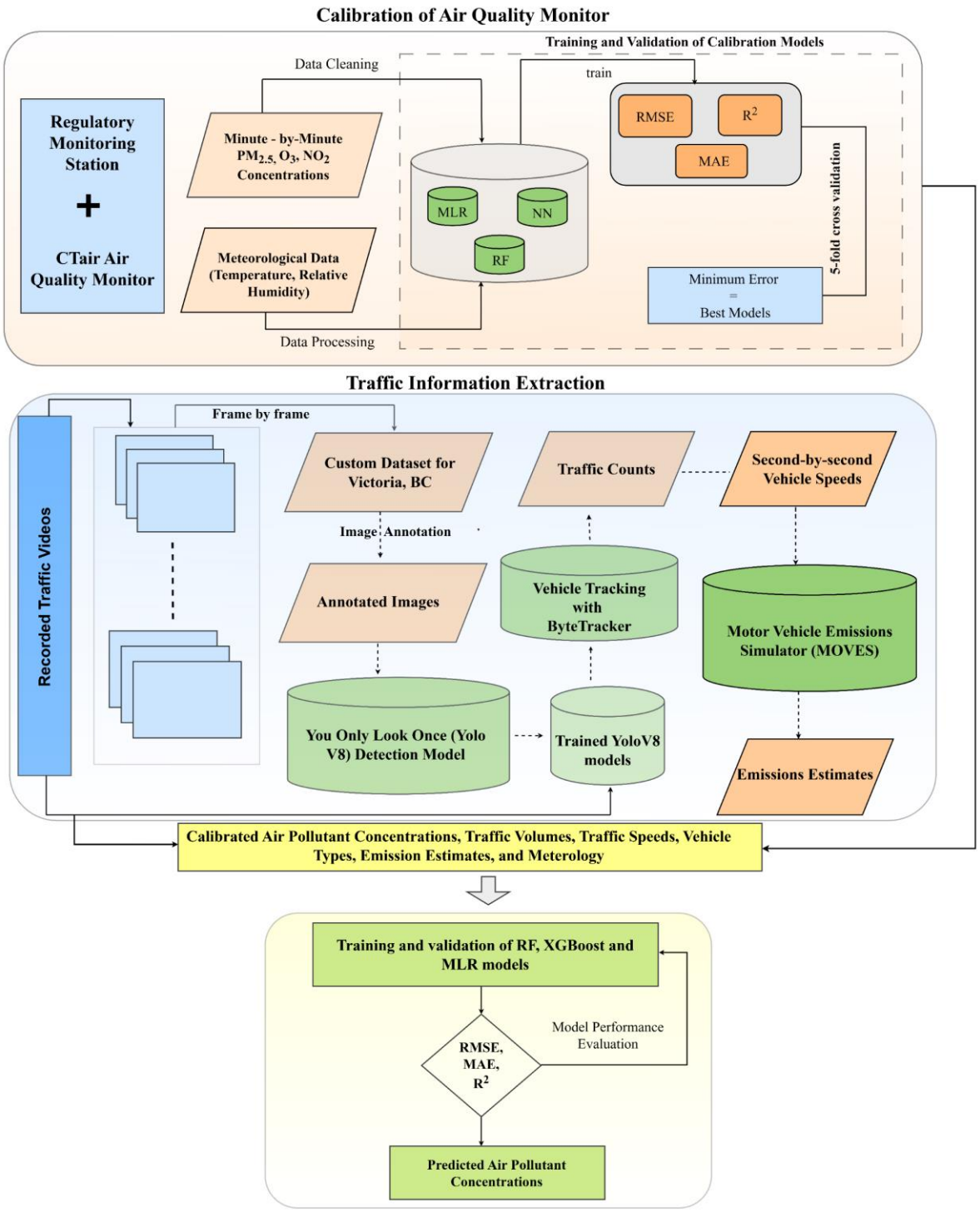


Figure 2.2. Methodology Framework including data collection, CTair air quality monitor calibration, custom dataset development, vehicle detection and tracking, emissions estimation, and machine learning models development. MLR: multi-linear regression; NN: neural network; RF: random forest; RMSE: root mean squared error; R2: coefficient of determination; MAE: mean absolute error.

2.2.4 Emissions Estimation

In this study, we used the U.S. EPA's Motor Vehicle Emission Simulator (MOVES) (USEPA 2023) to estimate exhaust, evaporative, brake and tire wear emission factors for all types and ages of on-road vehicles (Le Hong and Zimmerman 2021). MOVES is used to estimate vehicle emissions under various conditions specified by the user, such as vehicle type, vehicle speed, and emission process. In this study, we generated emission factors in grams per kilometers (g/km) for $PM_{2.5}$, NO_2, NO, VOC, sulfur dioxide (SO_2), and CO. We used the default input parameters in MOVES to derive emission factors for different types of vehicles (i.e., passenger cars, passenger trucks, light commercial trucks, transit buses, long haul trucks and short haul trucks) from the past 30 years across different road types and speed bins. After obtaining these factors, we calculated emissions for each combination of pollutant, vehicle type, and road type for a typical weekday, using vehicle fleet data from the Insurance Corporation of British Columbia (ICBC) (ICBC 2020). Since the vehicle types in the ICBC dataset differed from those in MOVES, we matched them using "Body style" available in the ICBC dataset. We then used average speed distribution data, obtained from previously processed traffic videos (see Paragraph 2.2.2), to estimate the appropriate emission factors for the average speed on the road link and the appropriate vehicle category. These values were multiplied by the length of the road link and the number of vehicles of each category on it. The result was an aggregated inventory of emissions for each pollutant on the road link considered. Detailed methodology employed for the emission estimates are presented in Section 2.8.1 in the Supplementary Information.

2.3 Air Pollutant Concentration Predictive Models

We employed three different machine learning models to predict the minute-by-minute air pollutant concentrations measured by the CTair air quality monitoring station using the traffic parameters extracted from the videos. The models - multiple linear regression model (MLR), random forest model (RF), and eXtreme Gradient Boosting Decision Tree (XGBoost) - were implemented in the scikit-learn statistical package in Python. For each of these models, we constructed a total of six different versions based on different combinations of features (predictors), to examine how these combinations impacted the predictions. The explanatory variables considered for the air pollutant and emissions predictions in each of the models are presented in Table 2.1. This comprehensive approach was essential for analyzing the feature importance and interaction effects and to help identify which factors were the most influential in determining pollutant levels.

Table 2.1. Feature Combinations for Air Pollutant Concentrations and Emissions Prediction Models.

Explanatory Variable	Description	Measure Type	Predictor Set 1	Predictor Set 2	Predictor Set 3	Predictor Set 4	Predictor Set 5	Predictor Set 6
Traffic Volumes	Volume of traffic on the road	Count	Included	Included	**Not Included**	Included	**Not Included**	**Not Included**
Vehicle Classes	Categories of vehicles (e.g., passenger cars, passenger trucks)	Categorical	Included	Included	Included	Included	Included	Included
Speed	Speed of vehicles	Km/hr	**Not Included**	Included	**Not Included**	Included	**Not Included**	**Not Included**
Background concentrations (bkg)	Background levels of pollutants	$PM_{2.5}$ (µg/m^3), O_3 (ppb), NO_2 (ppb)	Included	Included	Included	Included	Included	**Not Included**
Meteorology from Regulatory Station (Topaz)	(Temperature, Relative humidity)	T (°C), RH (%)	Included	Included	Included	**Not Included**	**Not Included**	Included
Traffic Emissions	Emissions from traffic sources	g	**Not Included**	**Not Included**	Included	**Not Included**	Included	Included

2.3.1 Multiple Linear Regression (MLR) Model

Multiple Linear Regression (MLR) models are useful for examining the relationships between dependent and independent variables, particularly with small sample sizes, using least squares fitting. They are a form of supervised machine learning that predicts numeric values (dependent variables) from a set of features (Ouma, Okuku, and Njau 2020). The model calculates a weighted sum of input features plus a bias (intercept) and assumes independence among variable components to form predictions (Ahmed and Mohamed 2021). The MLR equation is expressed as follows:

$$y_{target} = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots\ldots\ldots \beta_n x_{ni} + s_i \quad (8)$$

where:

- y_{target} is the observed value of the dependent variable,
- $x_1, x_2, \ldots, x_n$ are the explanatory variables,
- $\beta_0, \beta_1, \ldots\ldots, \beta_n$ are the regression coefficients.

In this context, y_{target} represents air pollutant concentrations, with the predictors including traffic volumes, meteorology, speeds, vehicle types, emission estimates, and background concentrations.

2.3.2 Random Forest (RF) Model

Random forests (RF) are a robust ensemble learning method which is used for classification, regression, and other similar tasks. The model works by constructing multiple decision trees during the training stages and aggregates their outputs to make final predictions, by either taking the mode for tasks regarding classification or taking the mean for regression tasks (Liu et al. 2019). The bagging technique is employed in this model, and this involves training each tree on a different subset and then sampling them with replacement. An advantage of RFs over other machine learning models is their ability to improve performance over individual trees by reducing variance and overfitting (Mustakim et al. 2022). To determine the degree of importance of each variable, RF randomly shuffles (permutes) the values of the variable to break the relation with the response and then measures the decrease in model accuracy. By averaging the difference in accuracy across all the trees in the forest, the importance of a variable can be determined. The model uses this technique to provide insights into which features significantly influence the model's predictions (Liaw and Wiener 2002). Using our predictor and target variables from the traffic information extraction and calibration results, we employed the grid search method to adjust hyperparameters of the RF to improve the performance of each model. The set of hyperparameters examined in this study included the number of estimators, maximum depth, minimum sample split, minimum samples leaf and maximum features.

2.3.3 EXtreme Gradient Boosting Decision Tree (XGBoost) Model

The eXtreme Gradient Boosting Decision Tree (XGBoost) model, proposed by Chen & Guestrin. (2016), is an iterative tree-based boosting framework designed to enhance the performance of the gradient boosting algorithm. XGBoost is versatile, capable of handling both linear and non-linear regression problems, and efficiently integrates software and hardware optimization techniques while utilizing fewer computing resources compared to other models (Ma et al. 2020). Additionally, XGBoost can handle categorical and sparse features and utilizes regularization techniques to manage complex data. Although both decision trees, the main distinction between RFs and XGBoost lies in their tree construction. While the trees for RFs are built independently of each other, each new tree is added to enhance the predictions of the previously built trees in the XGBoost model (Pan 2018). The main objective of the XGBoost model is to learn and fit the residual error between the prediction results of the previous tree and the actual value, thereby assessing the impact of each feature on the model's output. For this study, hyperparameter tuning was conducted using a grid search method to optimize the XGBoost algorithm's performance. Parameters such as learning rate, maximum tree depth (max_depth), number of trees (n_estimator), column subsampling ratio (colsample_bynode), and minimum child weight (min_child_weight) were adjusted.

2.4 Model Performance Evaluation

For each of the models and feature combinations, we randomly split the data into 80% training and 20% testing datasets, then applied a five-fold cross validation (CV) to the training dataset.

Cross-validation involved randomly dividing the data into five folds, training the model on four folds, and evaluating its performance on the remaining fold. The process was repeated five times, resulting in five models from the five datasets. The prediction performance of all three models (multiple linear regression, random forests and eXtreme Gradient Boosting Decision Tree) were compared using three metrics as described in Equations (9) to (11).

$$\text{Coefficient of determination: } R^2 = 1 - \frac{\sum_{i=1}^{N}(X_i - y_i)^2}{\sum_{i=1}^{N}(X_i - \bar{x})^2} \quad (9)$$

$$\text{Mean absolute error: MAE} = \frac{1}{N}\sum_{i=1}^{N} | \ x_i - y_i| \quad (10)$$

$$\text{Root mean squared error: RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - y_i)^2} \quad (11)$$

where:

- y_i is the observed pollutant concentrations,
- x_i is the i-th value of the predicted concentrations,
- $\bar{x}$ are the average measurements of the given model,
- N is the number of the number of samples.

The best models for each of the pollutants were selected from the comparisons. To analyse the feature importance and understand which features contribute the most to the model's predictions, we used the SHapley Additive exPlanations (SHAP). This method generates Shapley values which provide a unified measure of the importance of a model's features (Lundberg and Lee 2017).

2.5 Results

2.5.1 Calibration of the Air Pollution Sensors

After analysing the raw measurements collected from the 8-week collocation period, we observed significant differences in peak values by the CTair compared to those recorded by the reference station, suggesting the presence of systematic bias or variability. Specifically, the measurements from the CTair tended to overestimate $PM_{2.5}$ and O_3 concentrations while they underestimated NO_2 concentrations. The differences in measurements underscored the importance of calibrating the CTair monitor under local conditions for accurate measurements. Following the assessment, the random forest (RF) model exhibited superior performance for all pollutants compared to the other models in predicting $PM_{2.5}$, NO_2, and O_3 concentrations, achieving R^2 values of 0.74, 0.69 and 0.94 respectively. Figures S2.4-S2.6 in the Supplementary Information present the time series of the concentrations measured by the CTair and BAM instruments before and after calibration by the three models. The neural network (NN) model

also performed well, with R^2 values of 0.53 for $PM_{2.5}$ and 0.80 for O_3. In contrast, the multiple linear regression (MLR) model had the poorest performance, with an R^2 value of 0.39 for $PM_{2.5}$ and 0.53 for O_3. A general summary of the calibration results is presented in Table 2.2.

Table 2.2 Performance of calibration models for air pollutant concentrations recorded by the CTair based on performance metrics.

Pollutant	Model	MSE[a]	R^2 [b]
$PM_{2.5}$ ($\mu g/m^3$)	MLR[c]	3.49	0.39
	NN[d]	2.65	0.53
	RF[e]	1.50	0.74
O_3 (ppb)	MLR	19.51	0.68
	NN	6.61	0.89
	RF	3.63	0.94
NO_2 (ppb)	MLR	10.21	0.52
	NN	15.86	0.25
	RF	6.62	0.69

[a]: Mean Square Error

[b]: Coefficient of determination

[c:] Multiple linear Regression

[d:] Neural Network

[e:] Random Forest

Overall, higher performances were observed in predicting O_3 concentrations compared to $PM_{2.5}$ and NO_2, possibly due to higher data variability in the recorded $PM_{2.5}$ and NO_2 concentrations. Notably, both NN and RF models demonstrated their ability to handle peak values observed in the CTair air quality data. Prior to inputting measurements into the air pollutant prediction model, all data from the CTair monitor were corrected with the RF models developed.

2.5.2 Traffic Information

The YoloV8 model was trained on a total of 10,964 custom dataset of images representing vehicles in Victoria. The object detection model demonstrated high performance, with strong overall accuracy and a mean average precision across all classes of 0.973 (as shown in Figures S2.7 and S2.8 of the Supplementary Information). The high accuracy of the model indicates its effectiveness in accurately identifying the vehicles in the dataset. Specifically, most classes achieved perfect or near-perfect precision and recall rates, with all vehicle classes achieving

accuracy greater than 92%. After detection, the trained model was applied to the traffic videos, and the traffic counts were conducted in conjunction with ByteTracker. The Pearson correlation coefficient between the model's counts and the manual counts, and between the model's counts and city's counts, were 0.91 and 0.88, respectively. For the validation of the speed model, we obtained an average error rate of 4.02%.

2.5.3 Descriptive Statistics

We obtained a dataset of 4,460 records corresponding to minute-by minute data of: 1) average $PM_{2.5}$, NO_2, and O_3 concentrations and background levels; 2) total traffic volume; 3) average $PM_{2.5}$, NO_2, NO, VOC, SO_2, and CO emission estimates; 4) vehicle counts for each category; 5) average vehicle speeds. Table 2.3 presents the results obtained from the descriptive analysis of this study's dataset. In the context of traffic information, the data revealed clear differences in the pattern among different vehicle classes extracted from the object detection model. In particular, passenger trucks were the most present, followed by passenger cars and light commercial trucks. Subsequently, taxis and transit buses showed moderate intensities which indicated their presence was steady but not overwhelming. In contrast, the lowest traffic volumes recorded were coach buses and garbage trucks. The distribution of vehicle classes per hour by day is presented in Figure S2.9 in the Supplementary Information. Generally, the traffic variables showed less variability over time compared to the pollutant concentrations. Regarding the emissions data, the dataset exhibited high variability and significant right-skewness across the pollutants, with NO_2 emissions having a skewness of 3.5 and VOC emissions having a skewness of 4.22. These infrequent but intense emission patterns were also consistent across the other emissions like $PM_{2.5}$ and NO, which underscored the presence of rare but significant peaks in the area. The correlation between each of the explanatory and target variables were further analyzed, and the results are detailed in Table S2.1 of the Supplementary Information.

Table 2.3 Statistical distribution of the dataset.

Variables	Mean	Median	Minimum	Maximum	Standard Deviation
$PM_{2.5}$ conc. [a] ($\mu g/m^3$)	5.25	5.15	3.31	9.25	0.95
NO_2 conc. [a] (ppb)	9.05	7.82	2.15	16.12	3.94
O_3 conc. [a](ppb)	23.15	18.38	17.74	38.36	7.99
$PM_{2.5}$ bkg[b]($\mu g/m^3$)	4.34	4.00	0.00	9.01	2.36
NO_2 bkg[b](ppb)	5.84	4.08	0.20	23.30	4.61
O_3 bkg[b](ppb)	20.44	19.20	1.5	43.60	5.67
$PM_{2.5}$ Emi.[c] (g)	0.05	0.06	0.00	0.69	0.06
CO Emi.[c] (g)	2.64	2.62	0.00	14.43	1.29
NO_2 Emi.[c] (g)	0.17	0.15	0.00	2.04	0.25
NO Emi.[c] (g)	1.10	1.20	0.00	14.52	1.35
VOC Emi.[c] (g)	0.19	0.21	0.00	2.59	0.20
SO_2 Emi.[c] (g)	0.00	0.00	0.00	0.01	0.00
Traffic Volume	4.64	4.00	0.00	28.00	3.95
Speed (Km/hr)	44.44	45.12	0.00	125.01	16.02
Temperature (^oC)	19.27	20.19	8.00	25.61	3.04
Relative Humidity (%)	56.50	54.88	27.12	75.58	10.65

[a]: Concentrations

[b]: Background concentration levels

[c]: Emissions

2.5.4 Performance of Air Pollutant Concentration Predictions across Model Variations

The performance of the prediction models was evaluated using the key performance metrics: R^2, RMSE, and MAE. This section focuses on the prediction of air pollutant concentrations developed with the multiple linear regression model (MLR), random forest model (RF), and eXtreme Gradient Boosting Decision Tree (XGBoost) models, each applied to the different combinations of predictors which covered:

- Predictor Set 1: traffic volumes, background concentrations, and meteorological data
- Predictor Set 2: traffic volumes, speed, background concentrations, and meteorological data
- Predictor Set 3: traffic emissions, background concentrations, and meteorological data

- Predictor Set 4: traffic volumes, speed, and background concentrations
- Predictor Set 5: traffic emissions and background concentrations
- Predictor Set 6: traffic emissions and meteorological data

Scatter plots were employed to compare the consistency between the prediction performance of the three machine learning models across the different combination of features, as they reflect the linear relationship between the observed and predicted data. The model results for the best performing models and feature combinations for each pollutant are shown in Figure 2.3. All other scatter plots are shown in Figures S2.10-S2.15 of the Supplementary Information.

In terms of the overall accuracy, the best performing models across all pollutants were found among the non-linear models, specifically the RF and XGBoost models. When we compared the MLR model performance for $PM_{2.5}$, which recorded its maximum R^2 with Predictor Set 3 (i.e., 0.49), the RF and XGBoost models achieved much higher R^2 values (i.e., 0.97 and 0.98 respectively). The RF model attained its highest R^2 using features from Predictor Set 1 (i.e., traffic volumes, background concentrations, and meteorological data), while the XGBoost model achieved its highest R^2 with Predictor Set 3 (i.e., traffic emissions, background concentrations, and meteorological data). A similar trend was observed for NO_2, where the XGBoost model exhibited the highest performance with Predictor Sets 2 (i.e., traffic volumes, speed, background concentrations, and meteorological data) and 3, and RF earned the highest R^2 with Predictor Set 1, similar to the results for $PM_{2.5}$. In the case of O_3 predictions, the superior performance of the XGBoost model was recorded with Predictor Set 3 features, and RF performed best with Predictor Set 1. The MLR model on the other hand, achieved its highest R^2 with Predictor Set 2. Predictor Sets 4 and 5 resulted in the lowest R^2 values for all models, indicating that these predictor combinations are less effective for NO_2 predictions.

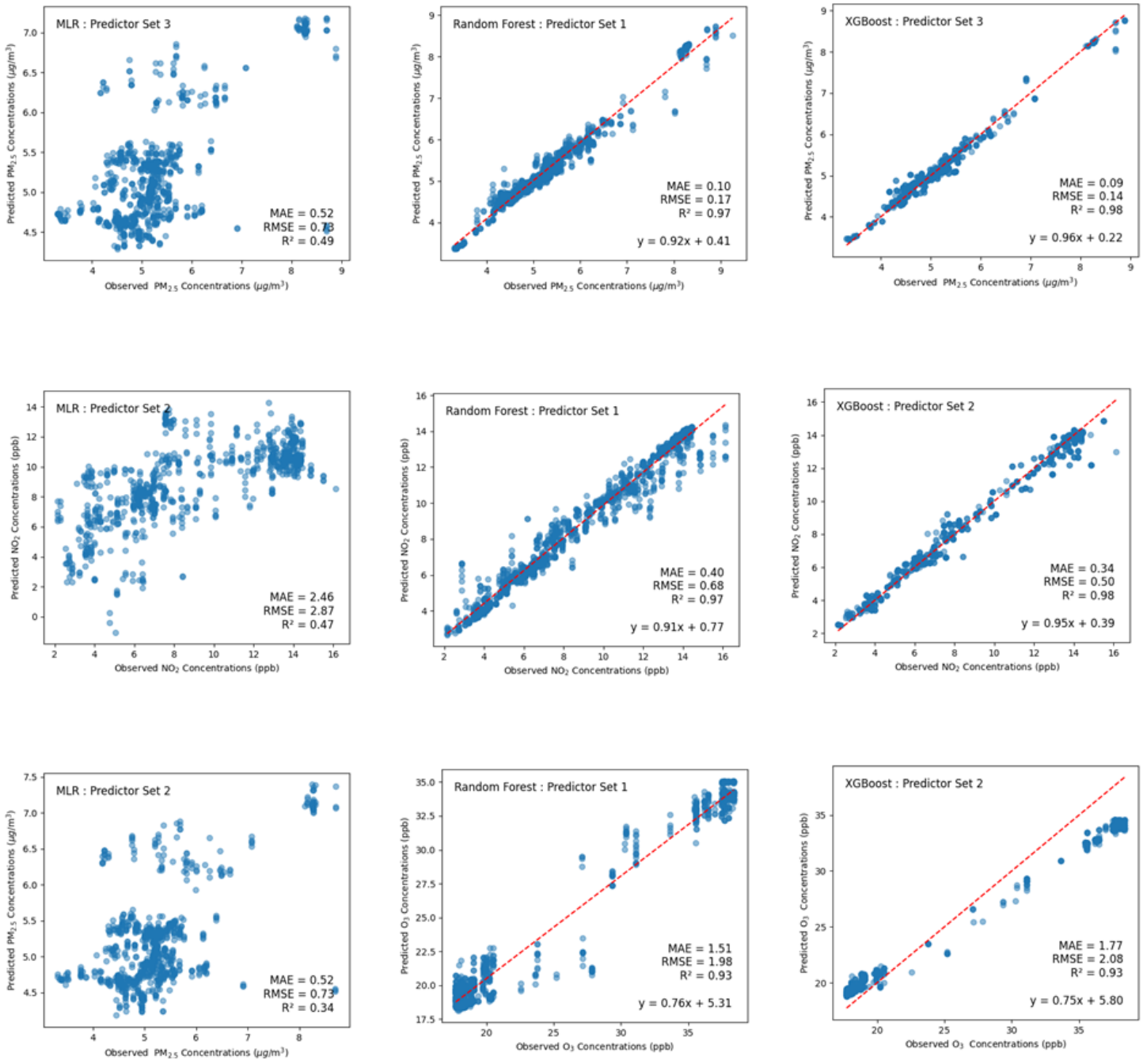


Figure 2.3. Scatter plots of observed and predicted $PM_{2.5}$, NO_2, and O_3 concentrations using best performing Predictor sets from random forest , eXtreme Gradient Boosting Decision Tree, and multiple linear regression models.

These metrics illustrate the ability of non-linear models to effectively capture the complex interactions between predictors and pollutant concentrations, in contrast to the limited ability of the MLR model. Further analysis into the error metrics revealed that there were significant reductions in mean absolute error (MAE) and root mean square error (RMSE) when moving from the linear to the non-linear models. We observed that for $PM_{2.5}$, the RF and XGBoost models consistently outperformed the MLR model across all the feature combinations. Specifically, RF achieved a reduction in MAE by 81.8% and RMSE by 78.2%, while XGBoost reduced MAE and RMSE by 78.8% and 78.1%, respectively. As for the error metrics for NO_2, RF showed the greatest reductions in MAE by 86.7% and RMSE by 80.4%. XGBoost also performed well, reducing MAE by 86.2% and RMSE by 82.6%. For O_3, which recorded the lowest reductions among all the pollutants, RF achieved a decrease in MAE by 57.5% and RMSE by 53.4%. XGBoost reduced MAE by 53.8% and RMSE by 54.3% compared to the MLR model. Generally, across all the pollutants, Predictor Set 4 showed the smallest reductions in MAE and RMSE compared to the linear model for both RF and XGBoost, which was expected since these sets were part of the sets which recorded the lowest R^2 values.

Further analysis into the variability of the model performance across the different predictor sets revealed that the RF model demonstrated the most consistency in model performance with Predictor Sets 1, 2, and 3 for $PM_{2.5}$, NO_2, and O_3 models, which indicated a stable model behavior. In contrast, the MLR model's performance had high variability, particularly with Predictor Sets 4 and 5, and this highlighted its sensitivity to changes in predictor variables. This variability analysis further emphasizes the robustness of non-linear models in maintaining high predictive accuracy across different predictor combinations.

2.5.5 Feature Interpretation and Analysis of Machine Learning Models

Figures 2.4 – 2.6 show the effects of the features on the predicted $PM_{2.5}$, NO_2 and O_3 concentrations using the XGBoost models for the feature combinations. These effects were measured using the SHAP plot, which provides insights into how each of the features contributes to the model's output. Positive SHAP values indicate that a feature contributes to increasing the predicted value, while negative SHAP values mean that the feature contributes to decreasing the predictive value of the model. In the same way, the dots on the plots represent each feature's attribution value and are colored based on the feature's values - high impacts are colored red, and low impacts are colored blue. When the dots are stacked vertically (or close to zero), it means the feature does not significantly alter the outcome of the model for different observations.

Initial analysis using the SHAP plots developed for the non-linear models showed that XGBoost and RF model plots were similar, however, XGBoost models explained the feature importances better, as demonstrated by their higher variability in SHAP values. Therefore, the results presented will focus on the XGBoost model's performance, with the RF SHAP plots shown in Figures S2.16 – S2.26 in the Supplementary Information.

The SHAP plots for Predictor Sets 1, 2, and 3 for the $PM_{2.5}$, NO_2, and O_3 models showed that temperature, relative humidity, and background levels were the features with the most influence on pollutant levels. These features had a wide range of SHAP values and suggested that they were critical in determining the air pollutant levels. Generally, we observed from the SHAP plots that higher temperatures and relative humidity both increased $PM_{2.5}$ levels, indicating that warm, moist conditions were conducive to elevate $PM_{2.5}$ levels. With respect to NO_2 and O_3 concentrations, the higher the temperatures, the higher the NO_2 and O_3 levels. However, higher relative humidity was associated with lower NO_2 and O_3 levels. In contrast, background NO_2 levels were not as impactful to the NO_2 model's performance in the presence of meteorological conditions as they were for $PM_{2.5}$, and O_3 levels. For instance, across Predictor Sets 1 and 2, the SHAP values for NO_2 background levels were lower than those for temperature and relative humidity which was not the case for $PM_{2.5}$ and O_3 levels, whose SHAP plots consistently showed that high background levels had a direct and strong impact on their predictions, regardless of the presence of other features.

A closer look at the traffic-related features, such as traffic volume and the vehicle types, revealed relatively lower SHAP values, which indicated that they had less impact compared to meteorological and background levels. For instance, when we included average speed data in Predictor Set 2, although it added a new dimension, where lower speeds led to higher $PM_{2.5}$, NO_2, and O_3 levels, they remained less influential overall across all the pollutants.

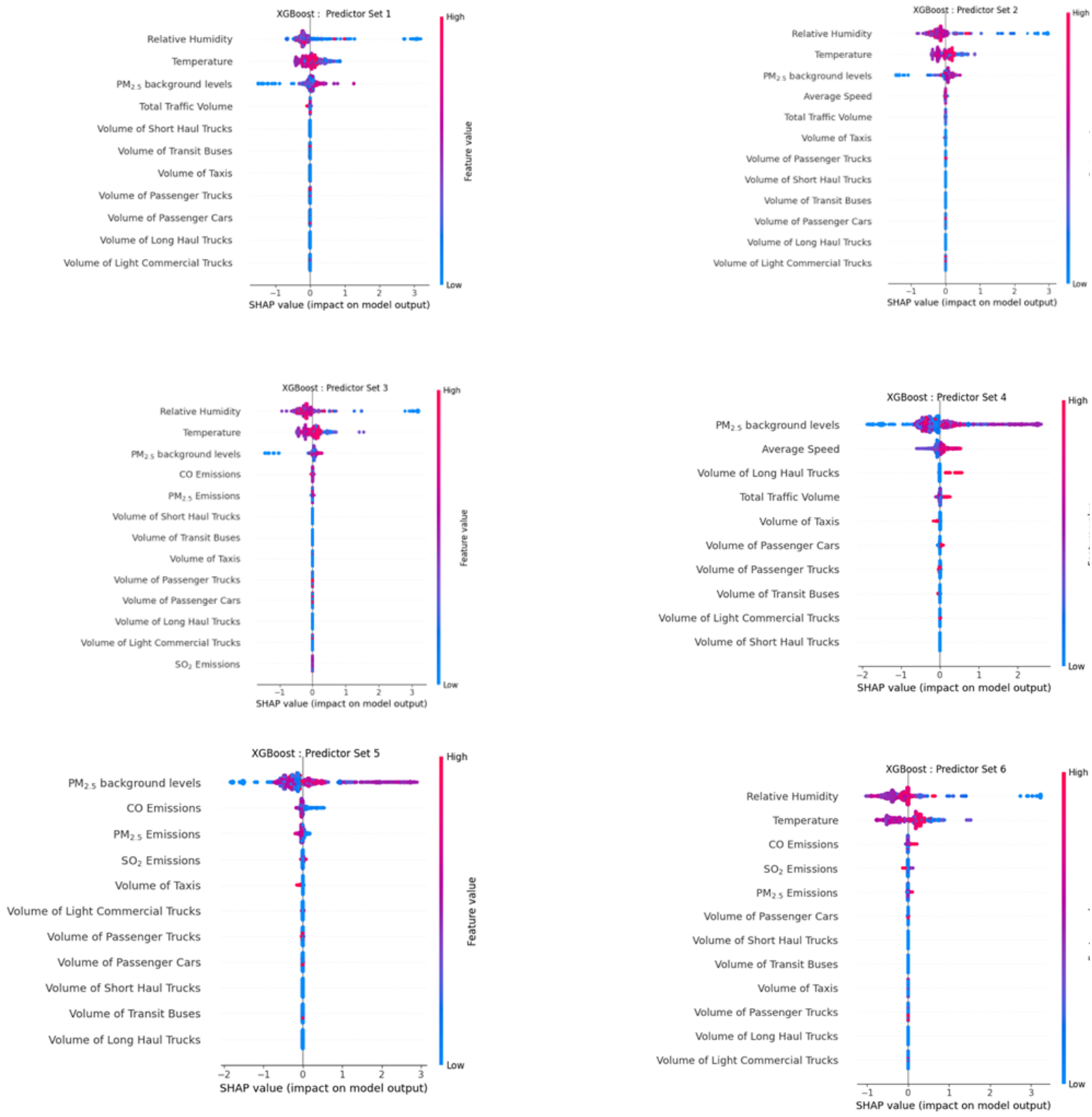


Figure 2.4. SHAP summary plots of eXtreme Gradient Boosting Decision Tree model for $PM_{2.5}$ concentrations using Predictor Sets 1-6.

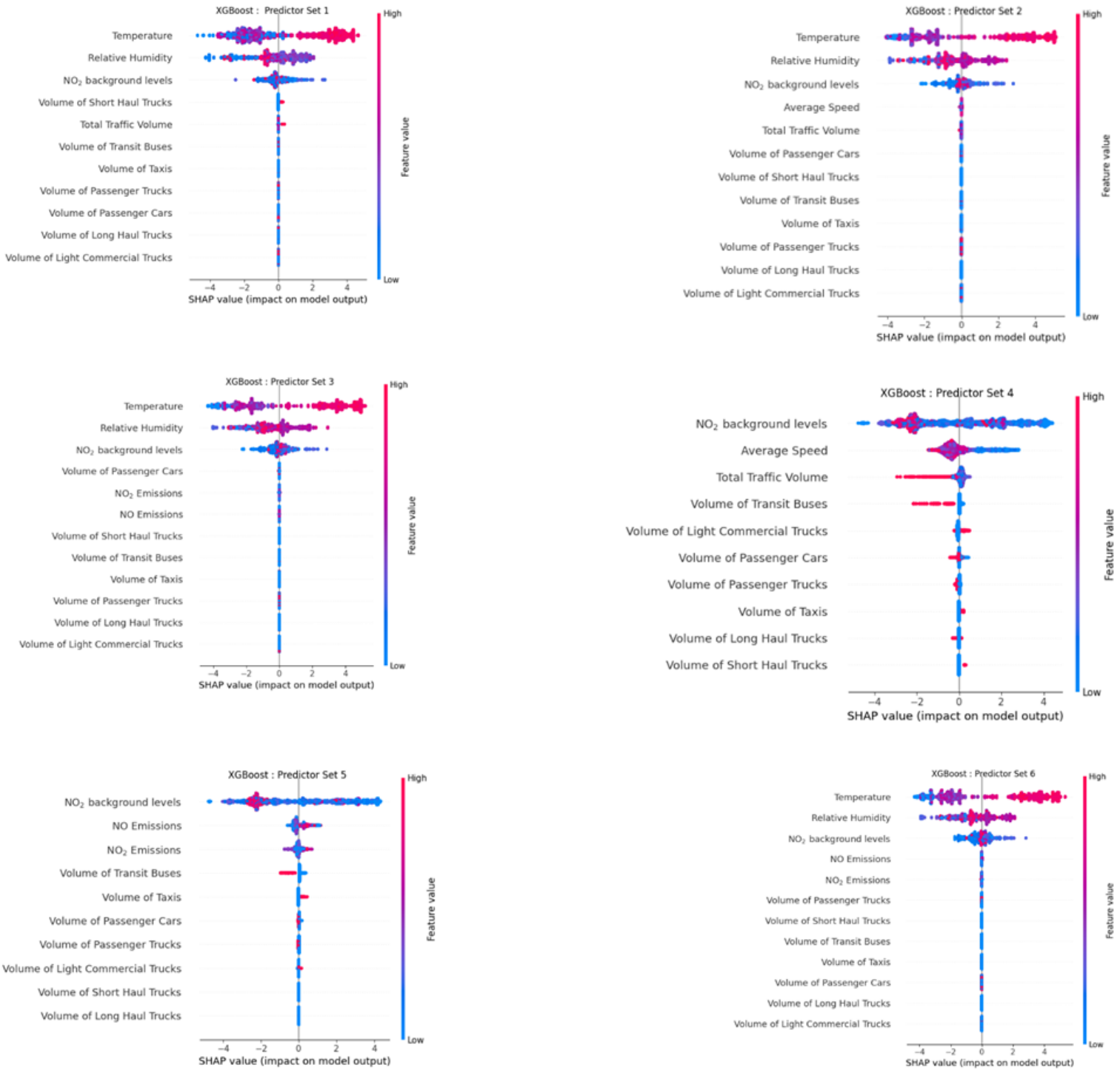


Figure 2.5. SHAP summary plots SHAP summary plots of eXtreme Gradient Boosting Decision Tree model for NO_2 concentrations using Predictor Sets 1-6.

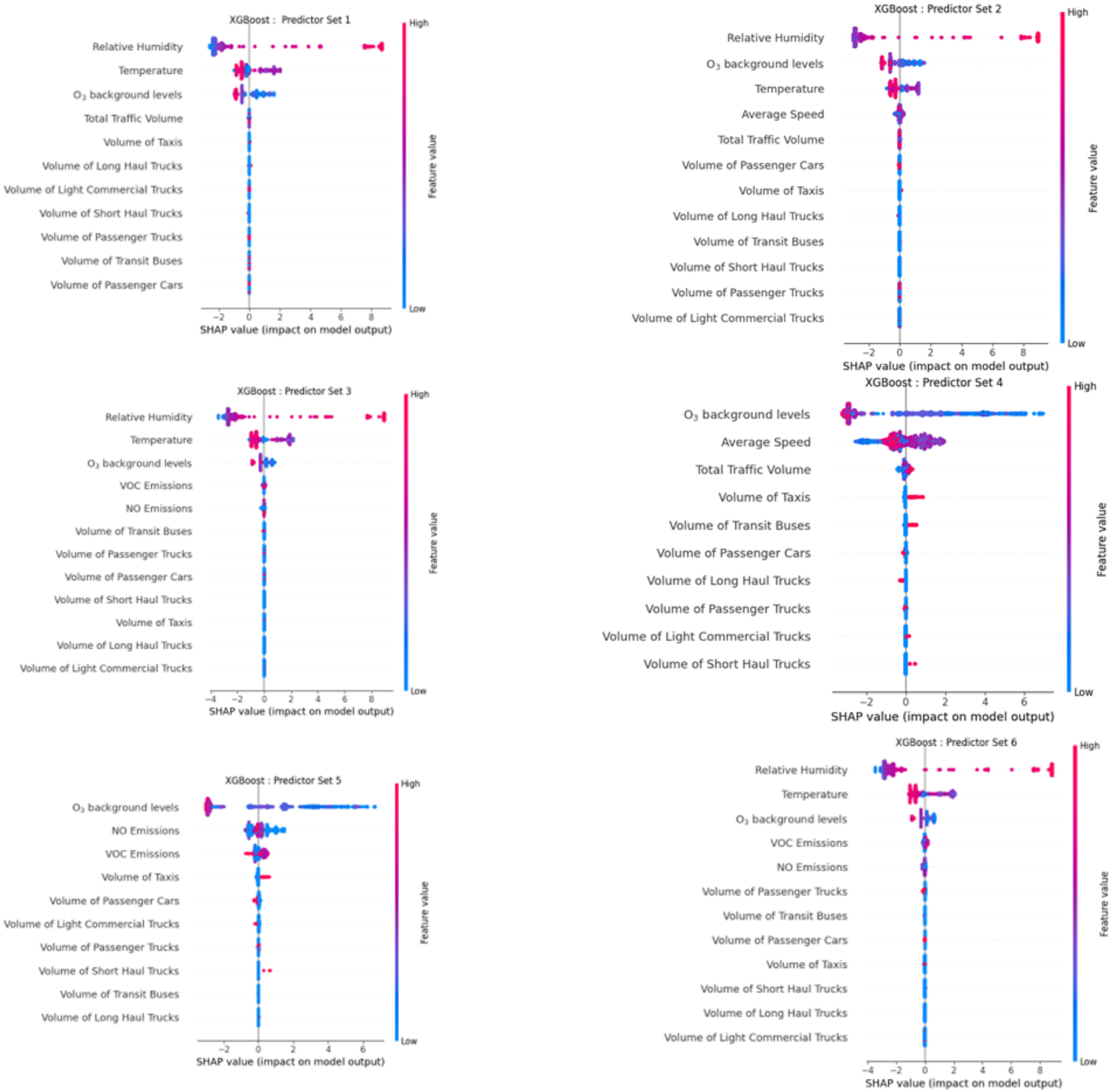


Figure 2.6. SHAP summary plots of SHAP summary plots of eXtreme Gradient Boosting Decision Tree model for O_3 concentrations using Predictor Sets 1-6.

Conversely, when we removed meteorological conditions and focused more on traffic data, such as speed and traffic volume in Predictor Set 4, we observed a strong influence from traffic,

especially from average speed, where lower speeds led to higher $PM_{2.5}$, NO_2, and O_3 levels, even though background concentrations still remained relevant. Here, the volume of long-haul trucks also had substantial impact on the $PM_{2.5}$ levels, while total traffic volume and transit buses had the highest influence on NO_2, and O_3 concentrations. Ultimately, Predictor Set 4 underscored the importance of specific traffic dynamics in pollution patterns. Predictor Set 3, which was the first set to include emissions, also highlighted notable impacts of CO and $PM_{2.5}$ emissions on $PM_{2.5}$ concentrations, NO and NO_2 emissions on NO_2 concentrations, and NO and VOC emissions on O_3 concentrations, linking direct emission sources to pollutant concentration levels. We observed high SHAP values for all emissions included in the predictions of each pollutant, demonstrating their notable influence in increasing the model's predictive power. With the major impacts emissions data had on air pollutant concentrations, we utilized Predictor Sets 5 and 6 to observe the model's behavior on the other two most important features: meteorological conditions (Predictor Set 6) and background levels (Predictor Set 5). Our analysis of their SHAP plots revealed that meteorological conditions, especially relative humidity combined with emissions had more impact on $PM_{2.5}$ and O_3 levels, while temperature, combined with emissions had more impact on NO_2 levels.

2.6 Discussion

The study's analysis revealed that for air pollutant concentrations predictions, non-linear models (RF and XGBoost) consistently outperformed the MLR model. Specifically, across all pollutant concentrations, the XGBoost model emerged as the most performant in terms of R^2 values, highlighting its ability to capture the complex interactions between predictors compared to the linear models. Non-linear models' ability to handle the intricate non-linear interactions in the data lies in their architecture. In that, ensemble learning used in these models significantly improve the accuracy and robustness of their predictions. Our findings are consistent with previous research emphasizing the effectiveness of non-linear models in the prediction of air pollutants. For instance, studies by Ren et al. (2020) and J. Li et al. (2022) reported higher prediction accuracies with RF and XGBoost models compared to linear models, particularly for complex interactions between predictors. Similarly, research by Chen et al. (2018) highlighted the robustness of non-linear models for PM concentrations, aligning with our results where RF outperformed other models for $PM_{2.5}$ predictions.

The SHAP value method identified variations in the SHAP values between RF and XGBoost models, reflecting how different these models capture the sensitivity of features. XGBoost showed a higher sensitivity to meteorological conditions and background conditions, as evidenced by its higher SHAP values for these features. The model also captures feature interactions more distinctively, and this indicates a more nuanced understanding of the combined effects of different predictors. XGBoost's ability to capture these feature interactions was due to its advanced modelling as they are able to capture the complex interactions between air pollutant levels. XGBoost's outcome of higher SHAP values shows that it is able to provide accurate predictions by considering the combined effects of multiple predictors, rather than evaluating

each predictor in isolation. Generally, our analysis of RF and XGBoost's predictions of $PM_{2.5}$, NO_2 and O_3 indicated that meteorological data was a crucial predictor for all pollutants, but it has the most significant impact on $PM_{2.5}$ and O_3 concentrations. This is likely due to the complex interactions between weather conditions and these pollutants.

Overall, across all pollutants and predictor sets, traffic data did not have major influence on the predictions compared to the other features. It should be noted that our study area, Victoria, is generally not a traffic-driven city, which could explain these results. Victoria has relatively moderate traffic conditions, has a well-developed public transportation system and extensive bike lanes, which reduce the dependency on cars. It is also a coastal city with relatively low buildings and thus few urban canyons, which favors the dispersion of TRAP. This methodology could still be applied to other traffic-intensive cities, where traffic data might be more relevant in the predictions. We were still able to develop additional insights throughout the models with the predictor sets that focused on traffic data. Specifically, we observed the $PM_{2.5}$ concentrations were more heavily influenced by traffic emissions and speed. In contrast, NO_2 concentrations showed the highest accuracy with traffic volume data alone. Regarding O_3, a distinct behavior was observed, where both traffic volumes and emissions contributed significantly to the pollutant's predictions. Higher traffic speeds are associated with reduced idling and smoother traffic flow and seems in our case to be leading to lower $PM_{2.5}$ and NO_2 concentrations. Since the formation of O_3 is influenced by precursor emissions (NO and VOCs) and meteorological conditions, the inclusion of these emissions enables the model to better capture the complex chemical reactions and interactions, thus improving prediction accuracy. In terms of vehicle types and their impacts, we observed that the inclusion of traffic speed and emissions showed varying degrees of influence across all pollutants. Major contributors for $PM_{2.5}$ were long-haul trucks, taxis, and passenger cars. For NO_2, the vehicle types with the most influence were short-haul trucks, passenger trucks, and transit buses and for O_3 emissions, passenger cars and taxis were the most significant. Across all the pollutants, traffic speed generally decreased the impact of vehicle types, especially heavy-duty and medium-duty vehicles, on $PM_{2.5}$ and NO_2 concentrations. When emissions data were introduced, the substantial contributions of heavy-duty vehicles to $PM_{2.5}$ and NO_2 levels were highlighted, while the role of lighter vehicles in O_3 formations were emphasized. It should be noted that the impact of emissions was more pronounced than that of average speeds and volumes. The high SHAP values for emissions data were expected due to their direct contribution to pollutant levels and their strong and immediate influence. Furthermore, each pollutant had specific influences: $PM_{2.5}$ was most influenced by meteorological data and emissions, NO_2 by a combination of meteorological data, emissions and traffic dynamics. And O_3 by meteorological conditions and emissions data. Other studies have indicated the influence of meteorological data on O_3 production, and their results are consistent with this study. Higher SHAP values for meteorological data, such as temperature, is associated with the phenomenon where increased temperature, when solar radiation is strong, enhances O_3 production. Conversely, when solar radiation is weaker, the effect on O_3 production diminishes (Li et al. 2022; Otero, Rust, and Butler 2021).

In this study, traffic data including traffic speeds, and 10 different vehicle classes extracted from video-based vehicle detection was used in conjunction with air pollutant concentration data. This represents a novel approach to analyzing the effects of certain features on air quality in cities. Compared to the use of traditional detectors for collecting data, vehicle classification using computer vision is economical and efficient and has been proven in this study to provide adequate information on air quality in cities. Although the present study had many advantages, it is also associated with a set of limitations. The absence of real emissions data at the location prevented the validation of the results derived from the MOVES model, which could potentially impact the accuracy of predictions. Additionally, the study location did not capture adequate volumes of all vehicle classes, such as coach buses, short and long-haul trucks, which led to underrepresentation in the models, and limited the generalizability and completeness of the analysis. While the non-linear models like XGBoost were able to effectively capture the feature interactions, they might not generalize well to different regions and are more complex, making it challenging to understand their data interpretation processes. Furthermore, the models did not fully account for seasonal, diurnal variations and extreme events, which could have affected the prediction accuracy. While this study has provided a basis for understanding the different features and combinations which affect $PM_{2.5}$, NO_2, and O_3 levels, future research could explore the integration of additional predictors, such as land use data and socioeconomic factors, to further enhance the models' accuracy and interpretability. Additionally, studies examining the temporal dynamics of feature impacts on pollutant concentrations and emissions would be valuable. This could help in understanding how seasonal variations and long-term trends might affect model's performance and feature importance. Ultimately, the high performance of the prediction models and the insights gained from the feature importances contributing to the predictions obtained from this study clearly demonstrate the ability to accurately predict air quality using data from low-cost sensors and traffic cameras alone.

2.7 Conclusion

This study used traffic videos combined with machine learning techniques to predict air pollutant concentrations. We leveraged the Yolov8 object detection model and the ByteTracker for multi-class vehicle detection and tracking to retrieve traffic counts and speeds, which were then used as inputs into the MOVES EPA model for the estimation of $PM_{2.5}$, NO_2, CO, SO_2, NO, and VOC emissions. Overall, 4,460 records corresponding to minute-by-minute data of traffic volume, traffic speeds, vehicle classes, air pollutant emissions, and meteorological conditions were analyzed to investigate the influence of different feature combinations on local air quality. Three different machine learning models were compared to analyze the effectiveness of each model in air pollution analysis. While linear models are simple, robust, and fast, they failed to capture the underlying complex relationships between pollutants and their predictors, which led to low accuracy in predicting pollutant concentrations. Our findings emphasize the critical role of traffic-related variables, background pollutant levels, and meteorological factors in air quality prediction, hence providing valuable insights for targeted interventions and policymaking. This

study shows that we can use a simple tool such as a traffic camera combined with information from a regulatory station to predict air quality.

2.8 Supplementary Information

Figure S2.1. Categories of images annotated for the custom dataset development.

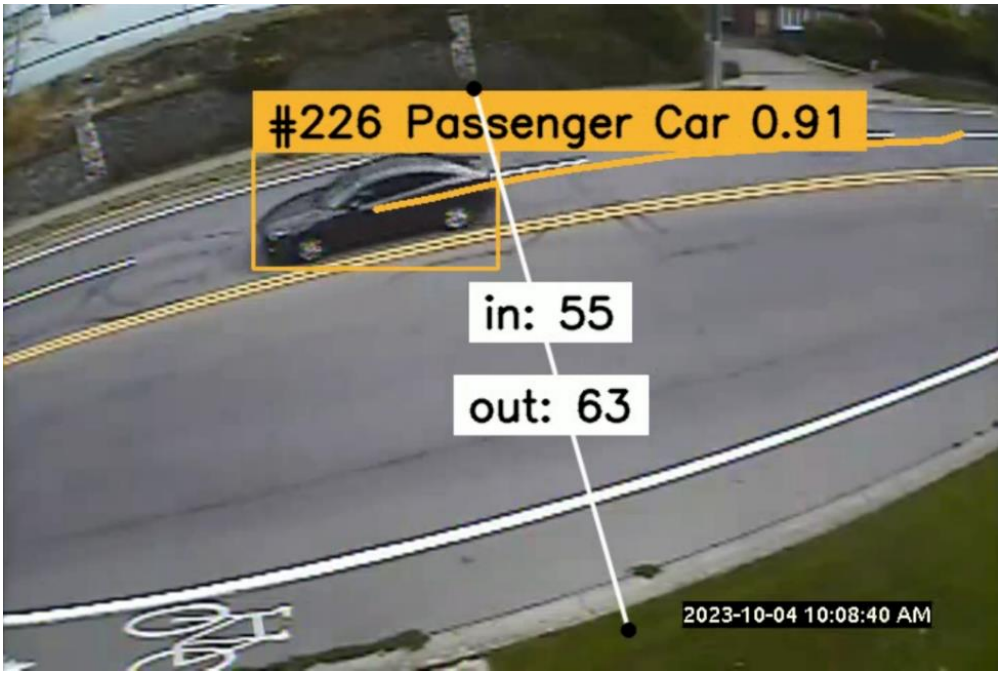


Figure S2.2. Virtual line drawn on video frame to facilitate vehicle counting.

Figure S2.3. Polygon drawn on the road to estimate image coordinates of the road.

2.8.1 Emission Estimates

In this study, we used the U.S. EPA's Motor Vehicle Emission Simulator (MOVES) to estimate exhaust, evaporative, brake and tire wear emission factors for all types and ages of on-road vehicles. MOVES is used to estimate vehicle emissions under various conditions specified by the user, such as vehicle type, vehicle speed, and emission process. We used the default input parameters in MOVES to derive emission factors for different types of vehicles (i.e., passenger cars, passenger trucks, light commercial trucks, transit buses, long haul trucks and short haul trucks) from the past 30 years across different road types and speed bins. In this study, we generated emission factors in grams per kilometers (g/km) for $PM_{2.5}$, NO_2, NO, VOC, sulfur dioxide (SO_2), and CO . After obtaining these factors, we calculated emissions for each combination of pollutant, vehicle type, and road type for a typical weekday, using vehicle fleet data from the Insurance Corporation of British Columbia (ICBC) (ICBC 2020).

2.8.2 Data Sources and Vehicle Categorization

We utilized the vehicle fleet data from the Insurance Corporation of British Columbia (ICBC) to categorize the vehicles in our study area, Victoria. The ICBC data included a comprehensive breakdown of vehicles into categories including passenger cars, taxi, passenger trucks, light commercial trucks, transit buses, long haul trucks and short haul trucks. We focused on the following categories provided in the ICBC dataset:

- Passenger Vehicles : This included sedans, hatchbacks, SUVs, and similar vehicles used for personal transportation.
- Commercial Vehicles : This included trucks, buses, and other vehicles used for commercial purposes.

We initially aligned these categories with the MOVES model and mapped the ICBC vehicle types to MOVES categories based on the "Body Style" attribute. Example of how the mapping was performed is shown in Table S2.1:

Table S2.1. Vehicle Mapping with ICBC dataset.

ICBC Vehicle Category	ICBC Body Style	MOVES Vehicle Category
Passenger Vehicles	Car	Passenger Car
	Station Wagon	Passenger Car
	Taxi	Passenger Car
	SUV	Passenger Truck
	Pickup Truck	Passenger Truck
Commercial Vehicles	Van	Light Commercial Truck
	Medium Truck	Short- haul Truck
	Bus	Transit Bus
	Heavy Truck	Long- haul Truck

2.8.3 Emission Calculation

After establishing the mapping of vehicles with the ICBC dataset, we were able to allocate the emission factors obtained from MOVES to the vehicle categories, taking into account the age distribution and fuel type of the vehicles in Victoria. We then incorporated the average speed distribution data, obtained from previously processed traffic videos (see Paragraph 2.2.2), to estimate the appropriate emission factors for the average speed on the road link and the appropriate vehicle category. These values were multiplied by the length of the road link and the number of vehicles of each category on it. For each combination of pollutant, vehicle type, and road type, emissions were calculated using:

$$E_{ijk} = EF_{ij} \text{ X } L_k \text{ X } N_{ij} \qquad \text{(T1)}$$

Where:

- E_{ijk} is the emission of a pollutant for vehicle type i on the road link.
- EF_{ij} is the emission factor for vehicle type i and speed bin j.
- L_k is the length of road link k
- Nij is the number of vehicles of type i on road link k.

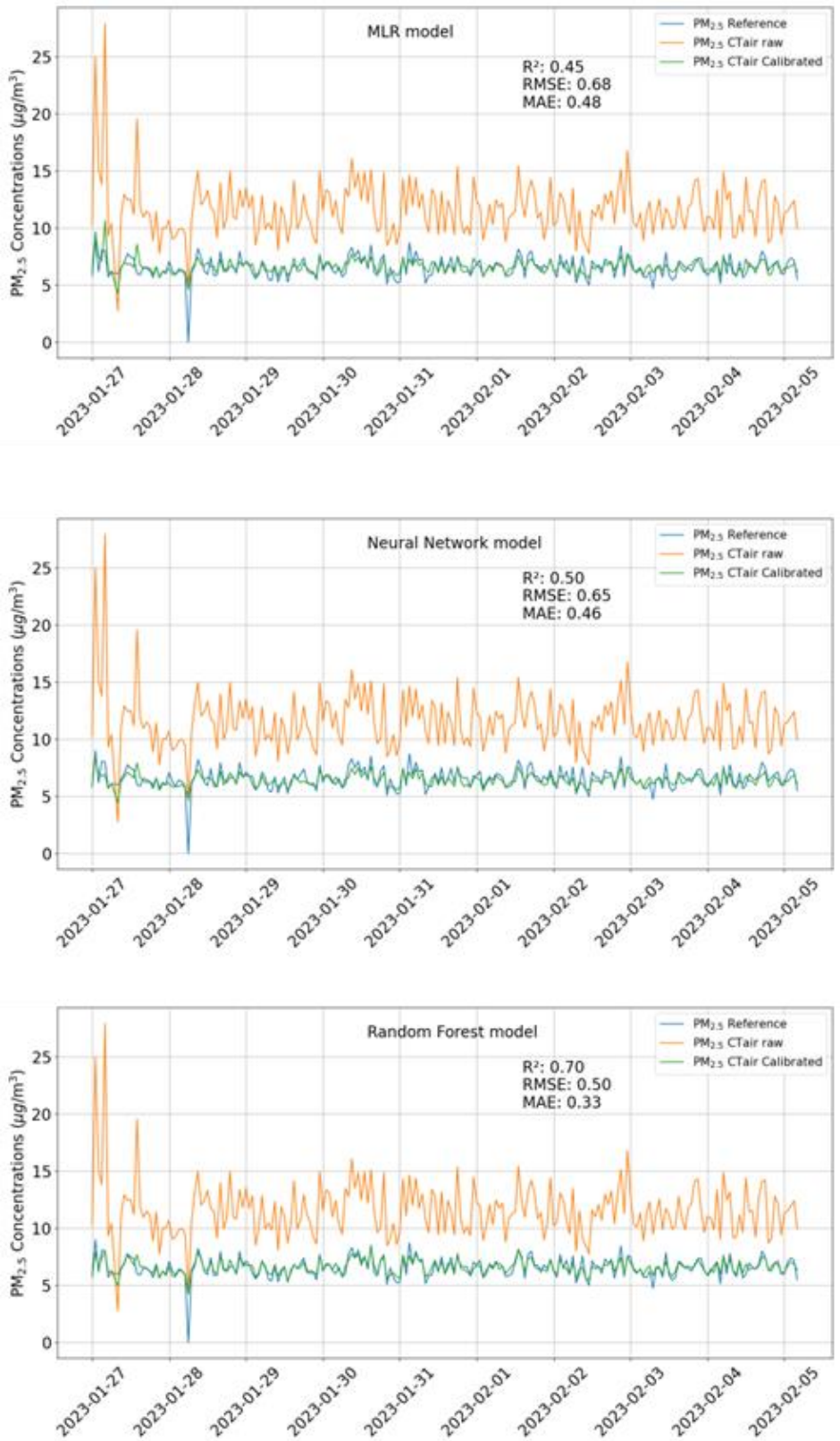


Figure S2.4. Time series of hourly $PM_{2.5}$ trends before and after calibration with MLR ,random forest, and neural network models.

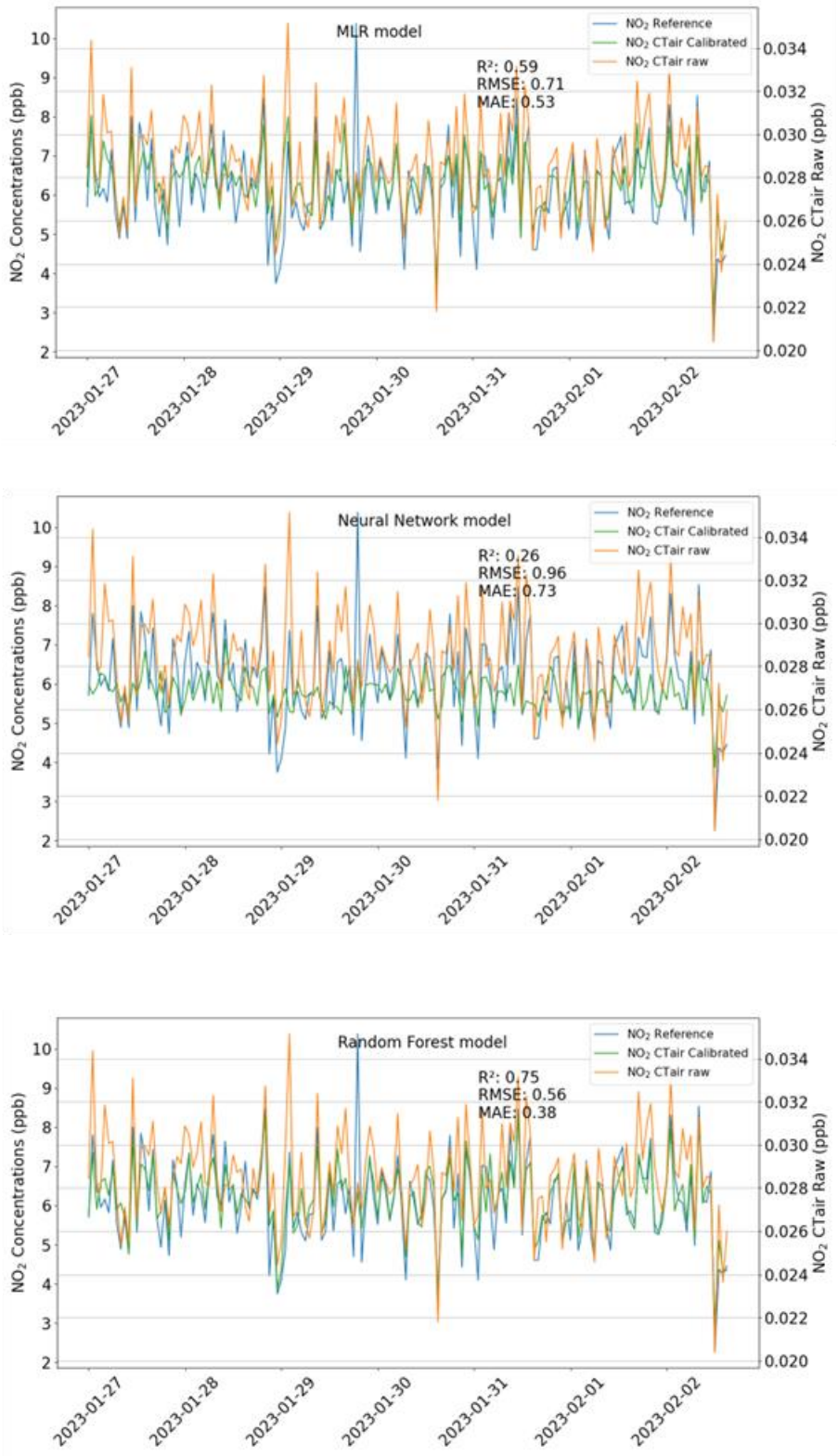


Figure S2.5. Time series of hourly NO_2 trends before and after calibration with MLR, random forest, and neural network models.

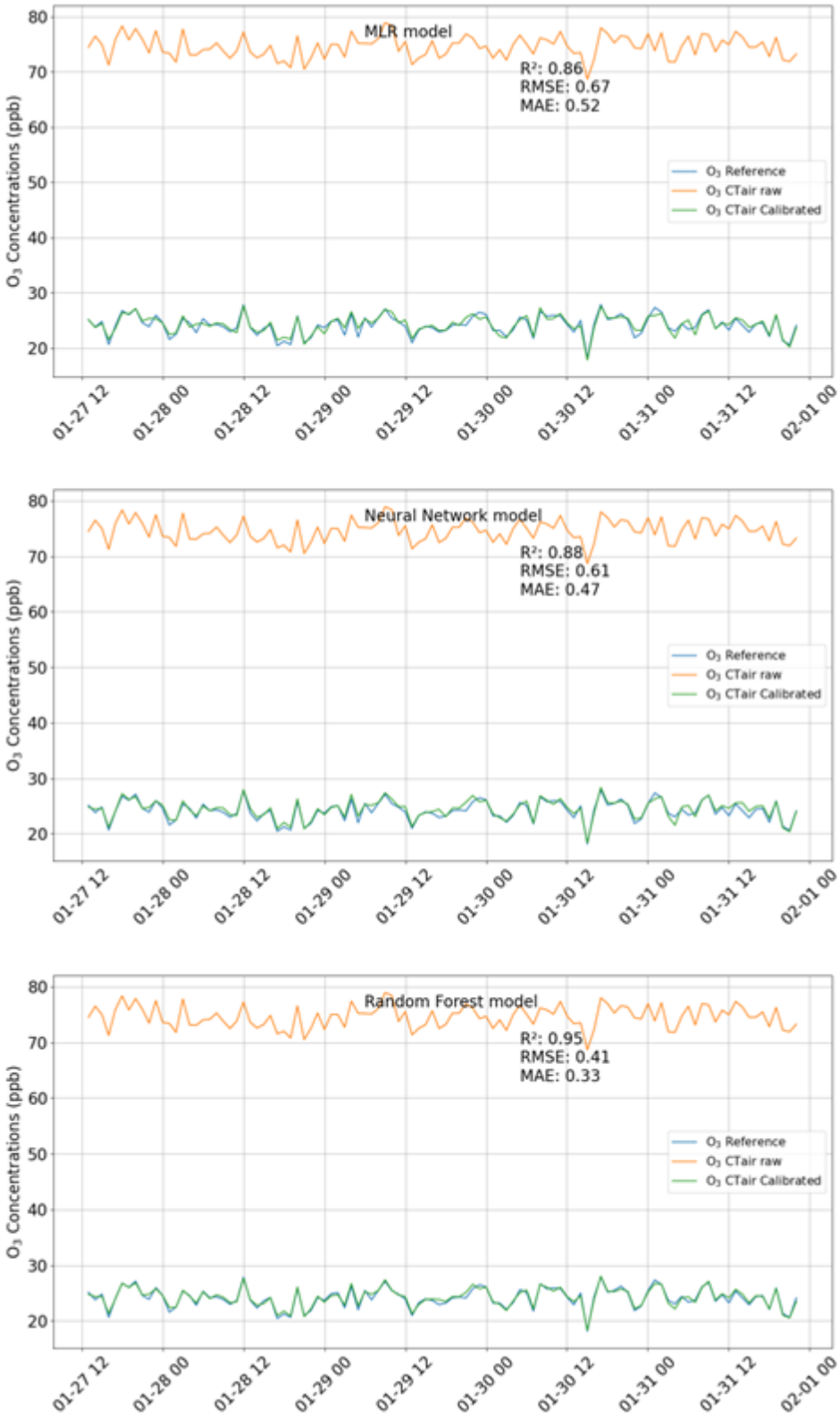


Figure S2.6. Time series of hourly O_3 trends before and after calibration with MLR ,random forest, and neural network models.

Figure S2.7. Vehicle classification by object detection model. Each frame shows the detected objects with their respective classes, along with their confidence scores represented inside the colored boxes.

2.8.4 Vehicle Detection Model Performance

Figure S2.8 presents the precision-recall curves for the vehicles detected by the model , such as “Garbage truck” , “Light Commercial truck” , “Passenger truck” , “Passenger car”, and others. Each curve demonstrates the performance of the object detection model for the respective class. The numbers next to each class in the legend represent the Average Precision (AP) score for that class, indicating the model’s precision across the different recall levels. The mean Aveage Precision (mAP) at 0.5 for all classes is 0.973, highlighting the overall high accuracy of the model in correctly identifying objects while minimizing false positives across the range of recall values.

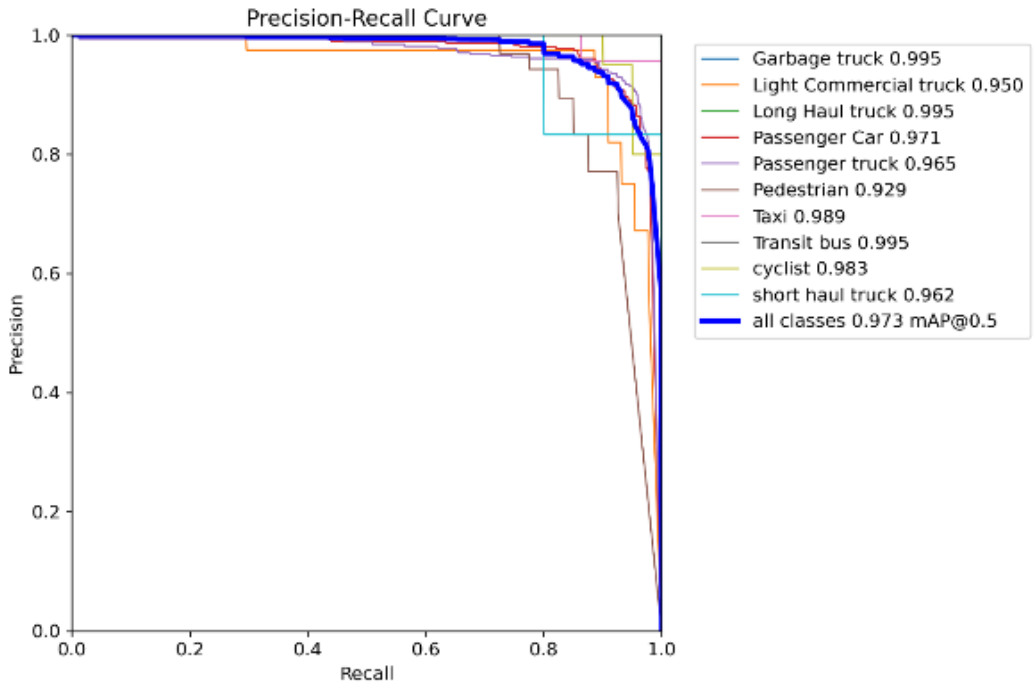


Figure S2.8. Precision-Recall Curve for Vehicle Classification.

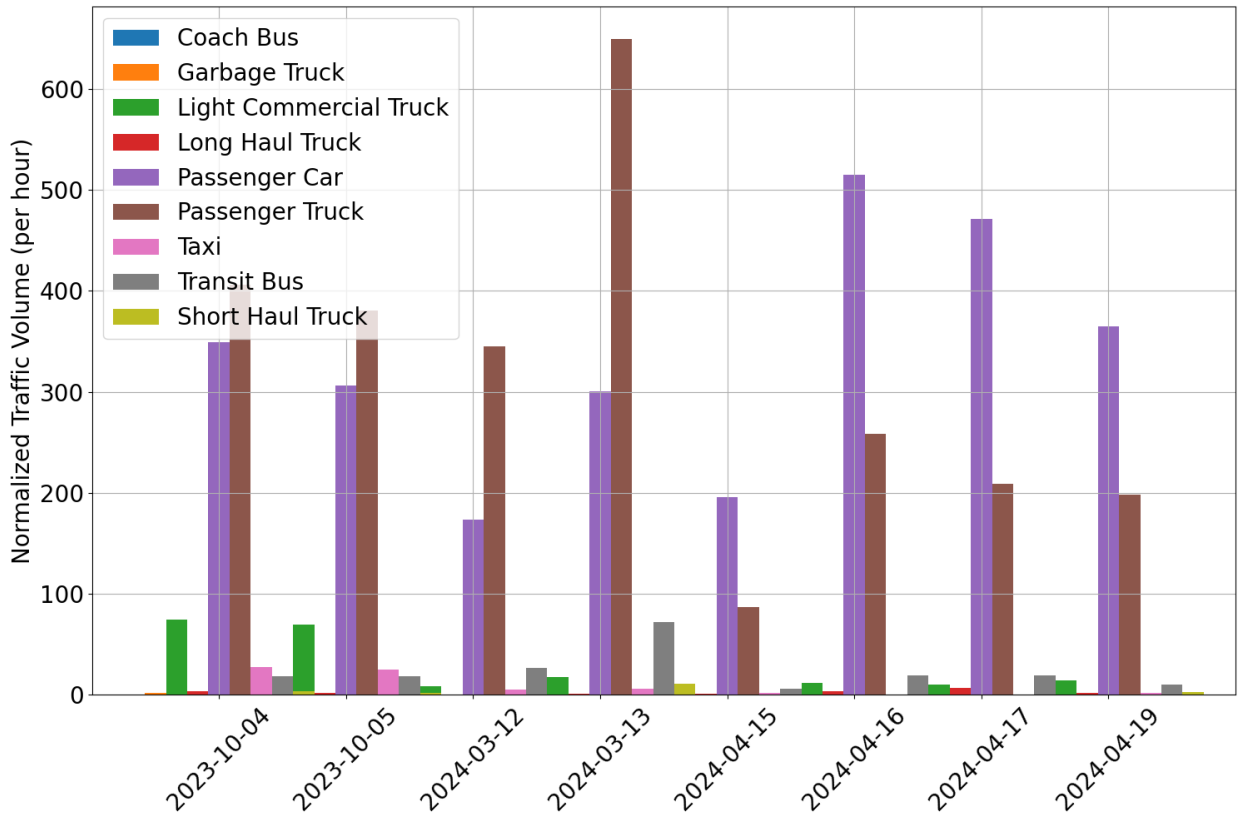


Figure S2.9. Distribution of Vehicle Classes per Hour by Day at Study location.

Table S2.2 Features with corresponding correlations.

Features correlated with $PM_{2.5}$	Pearson Correlation	Features correlated with NO_2	Pearson Correlation	Features correlated with O_3	Pearson Correlation
$PM_{2.5}$ Background levels	0.181	**NO_2 Background levels**	-0.222	**O_3 Background levels**	-0.180
Temperature	-0.204	**Temperature**	0.634	**Temperature**	-0.383
Relative Humidity	-0.563	**Relative Humidity**	-0.071	**Relative Humidity**	0.806
Total Counts	0.019	**Total Counts**	-0.110	**Total Counts**	0.003
Speed	0.158	**Speed**	-0.085	**Speed**	-0.116
Light Commercial Truck	-0.021	**Light Commercial Truck**	0.064	**Light Commercial Truck**	0.002
Long Haul Truck	0.101	**Long Haul Truck**	0.067	**Long Haul Truck**	-0.081
Passenger Car	0.012	**Passenger Car**	-0.027	**Passenger Car**	-0.023
Passenger Truck	0.007	**Passenger Truck**	-0.026	**Passenger Truck**	-0.024
Taxi	-0.090	**Taxi**	0.013	**Taxi**	0.122
Transit Bus	0.030	**Transit Bus**	-0.076	**Transit Bus**	-0.004
Short Haul Truck	-0.004	**Short Haul Truck**	0.037	**Short Haul Truck**	0.013
SO_2 Emissions	-0.027	**NO_2 Emissions**	-0.011	**NO Emissions**	0.002
CO Emissions	-0.099	**NO Emissions**	0.010	**VOC Emissions**	0.001
$PM_{2.5}$ Emissions	-0.044				

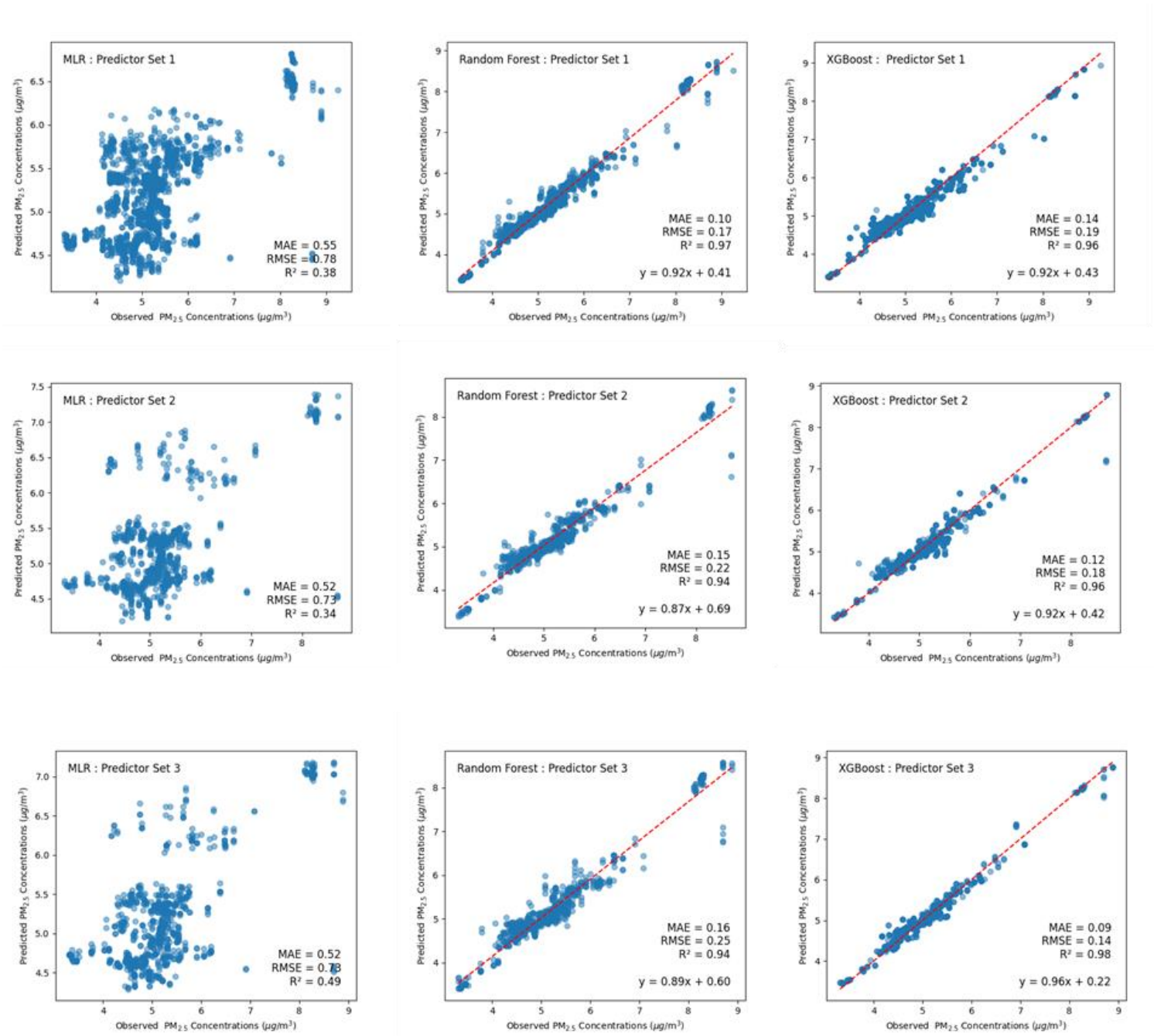


Figure S2.10. Scatter plots of observed and predicted $PM_{2.5}$ using multiple linear regression , random forest , and XGBoost with Predictor Sets 1-3.

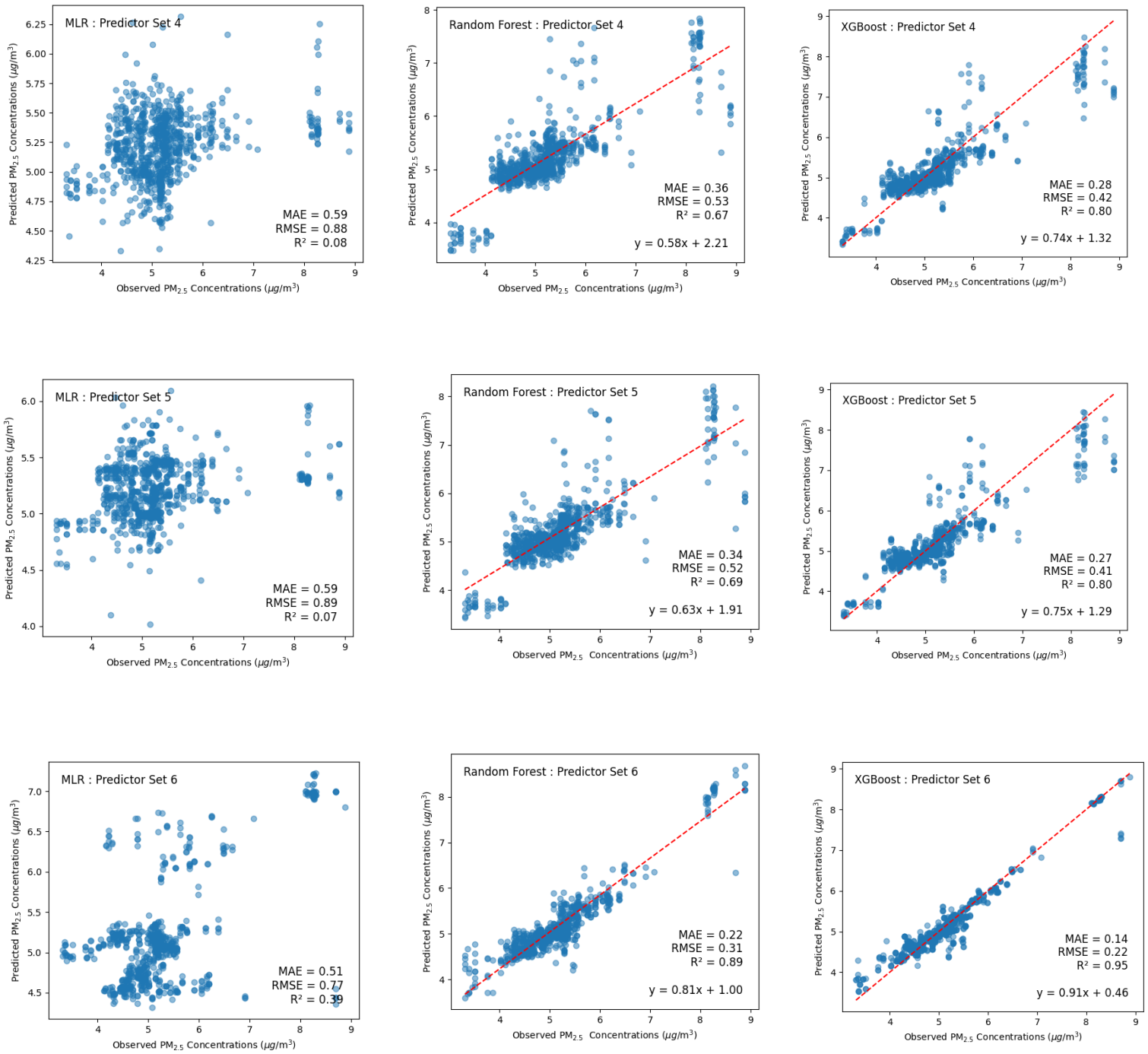


Figure S2.10 Scatter plots of observed and predicted $PM_{2.5}$ using multiple linear regression , random forest , and XGBoost with Predictor Sets 4-6.

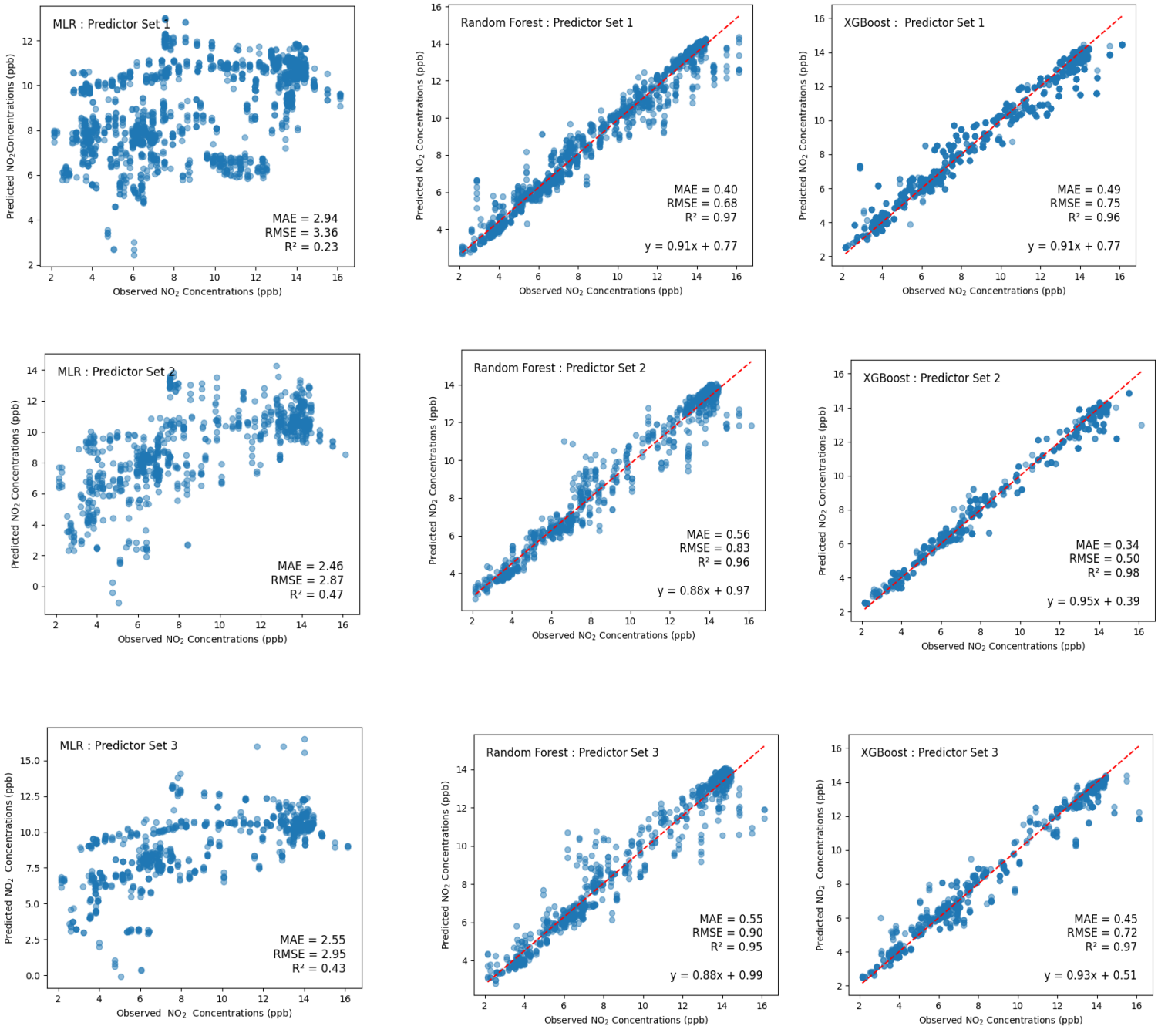


Figure S2.11. Scatter plots of observed and predicted NO_2 using multiple linear regression , random forest , and XGBoost with Predictor Sets 1-3.

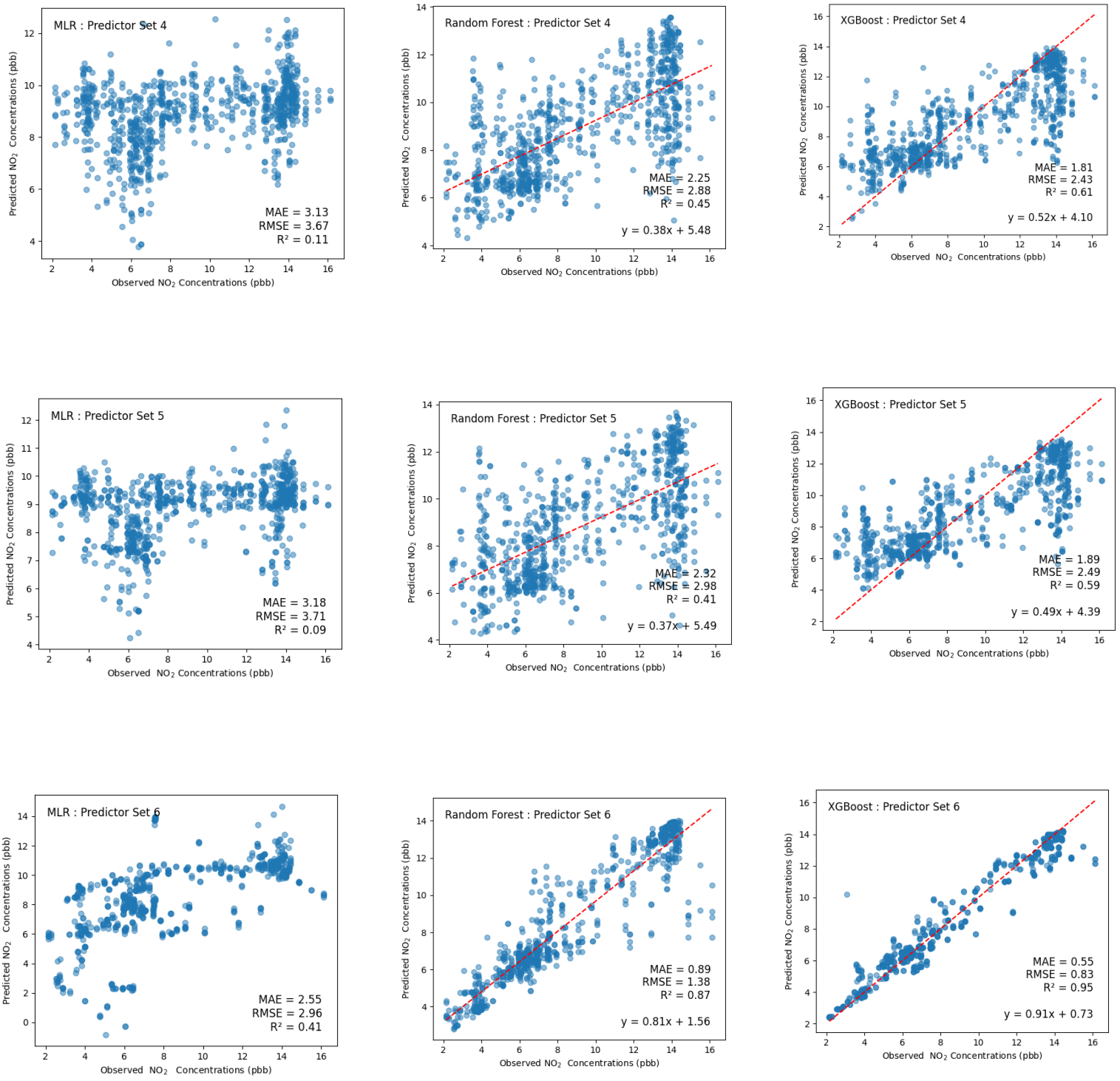


Figure S2.12. Scatter plots of observed and predicted NO_2 using multiple linear regression , random forest , and XGBoost with Predictor Sets 4-6.

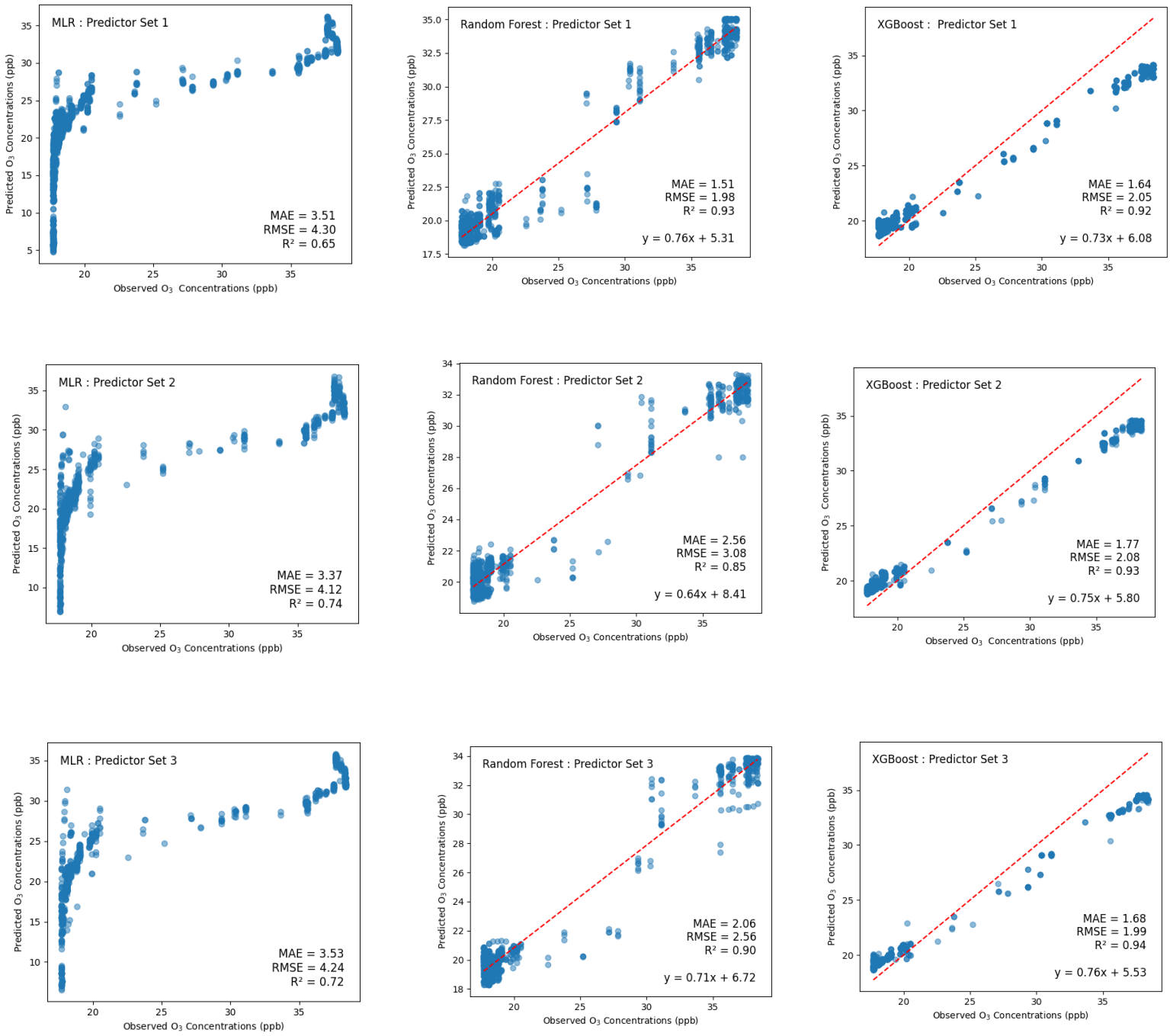


Figure S2.13. Scatter plots of observed and predicted O_3 using multiple linear regression , random forest , and XGBoost with Predictor Sets 4-6.

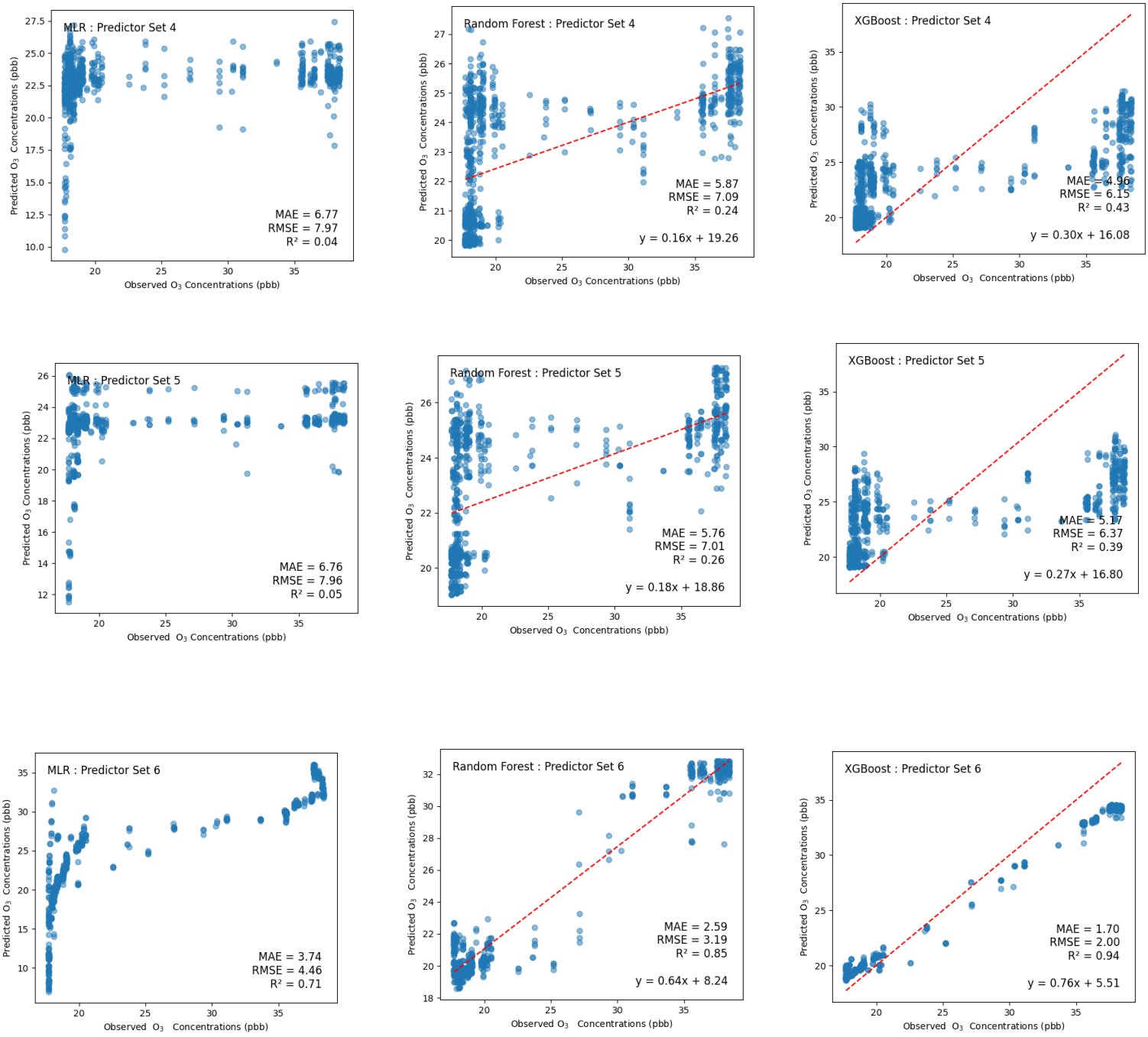


Figure S2.14. Scatter plots of observed and predicted O_3 using multiple linear regression , random forest , and XGBoost with Predictor Sets 4-6.

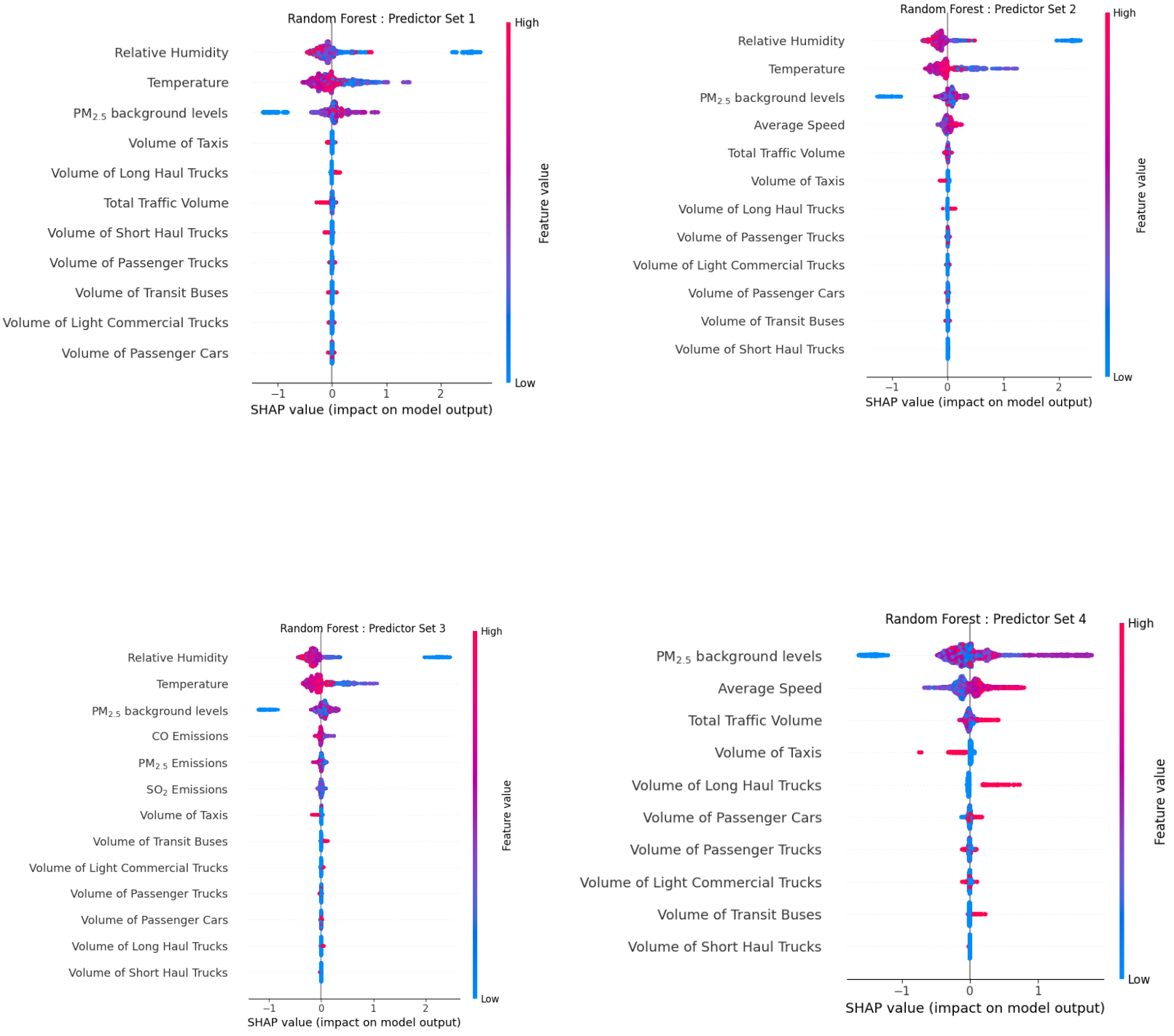
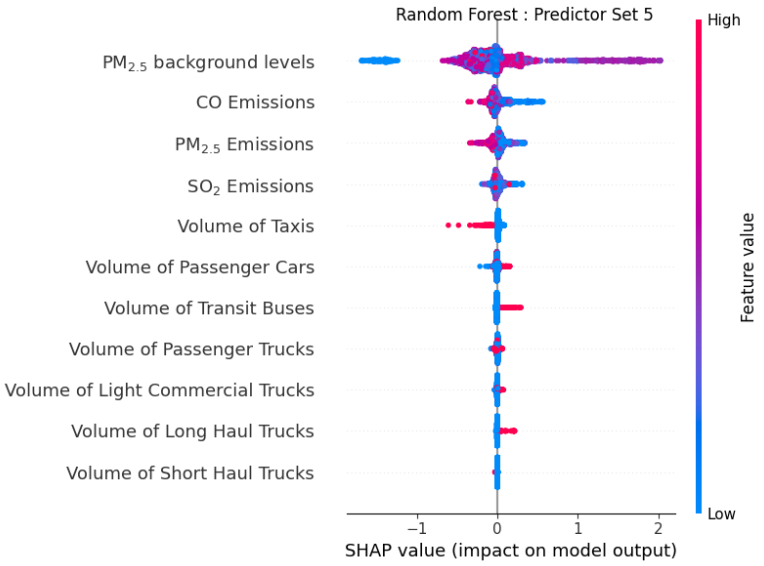
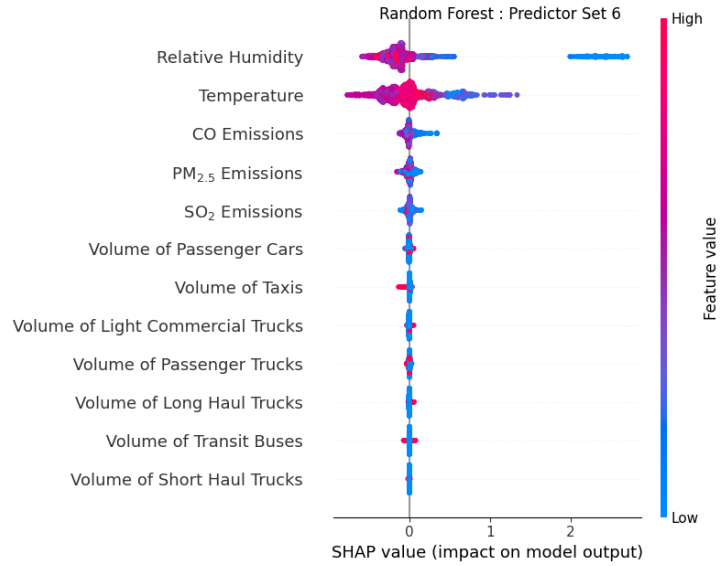
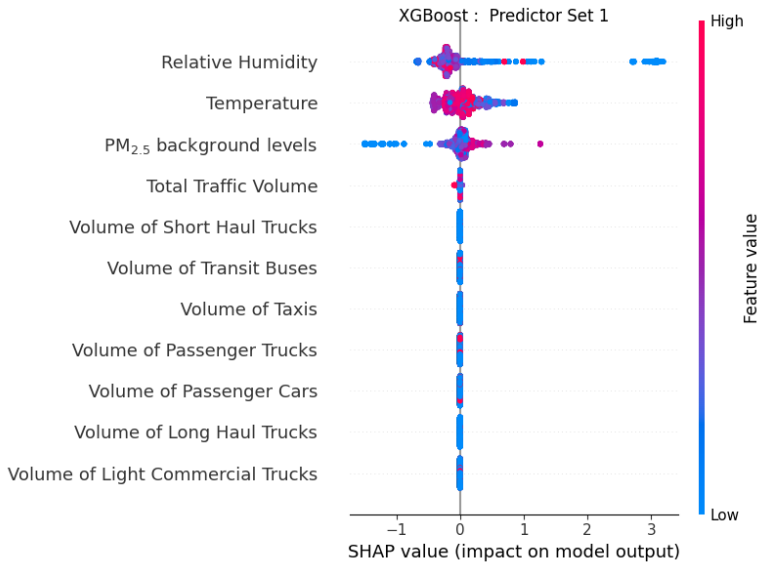


Figure S2.15. SHAP summary plot for $PM_{2.5}$ with Random Forest with Predictor Sets 1-4.







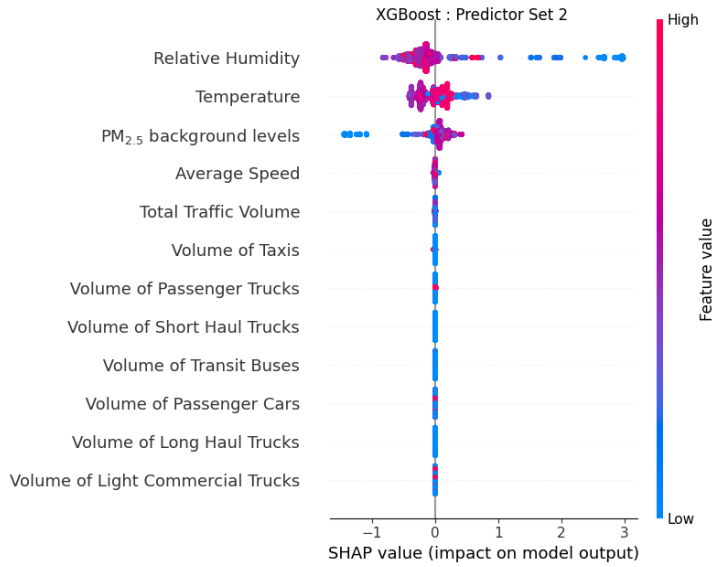


Figure S14a

Figure S2.16. SHAP summary plot for $PM_{2.5}$ with XGBoost with Predictor Sets 1-2.

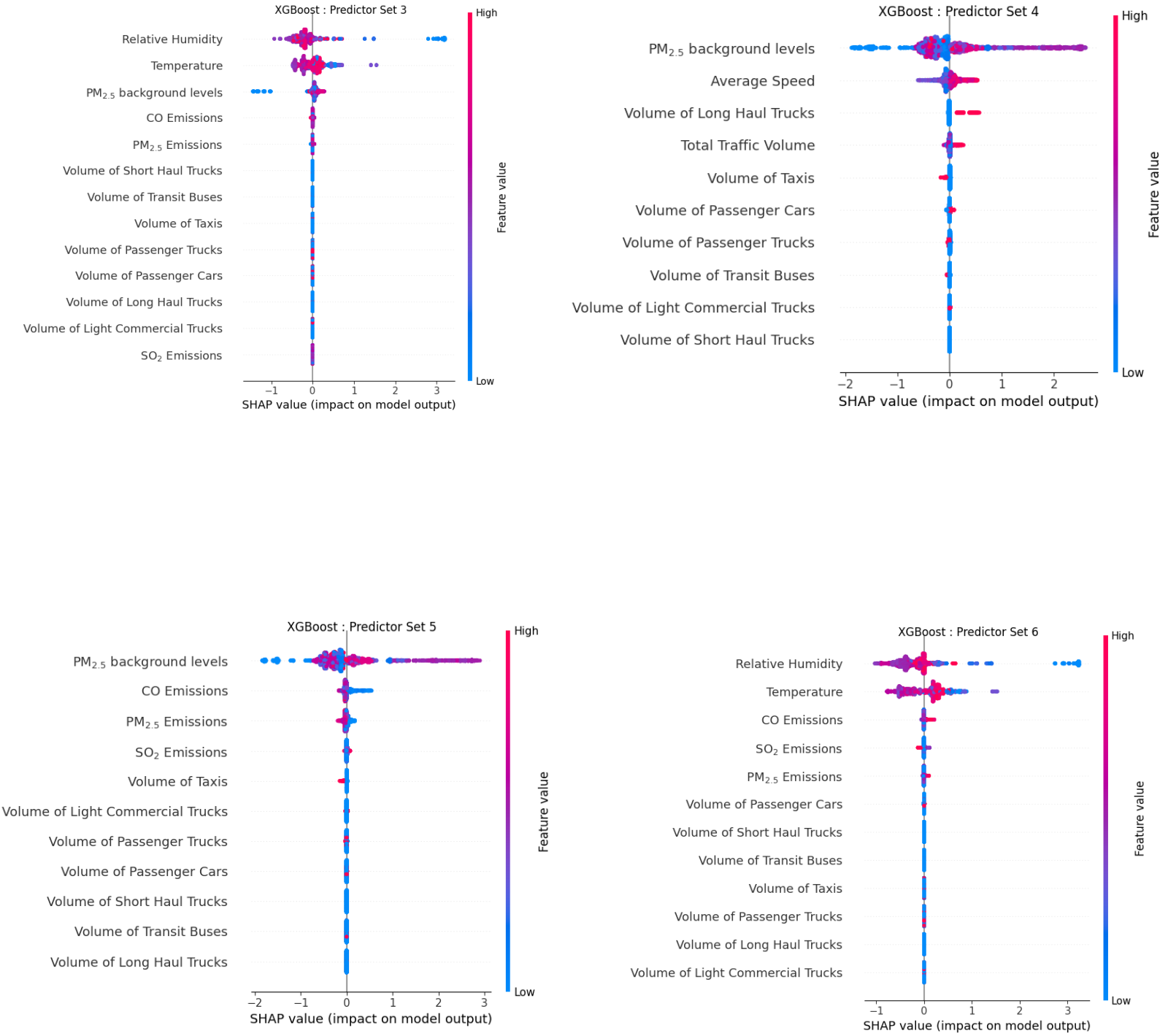
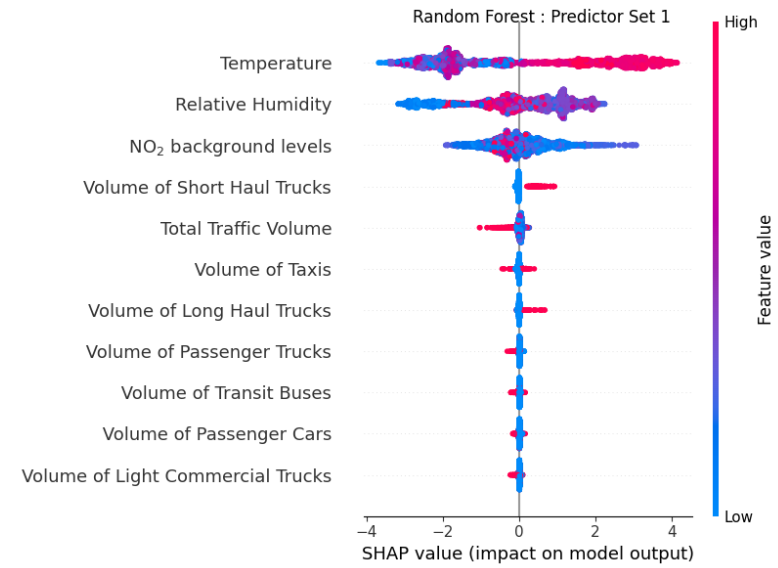
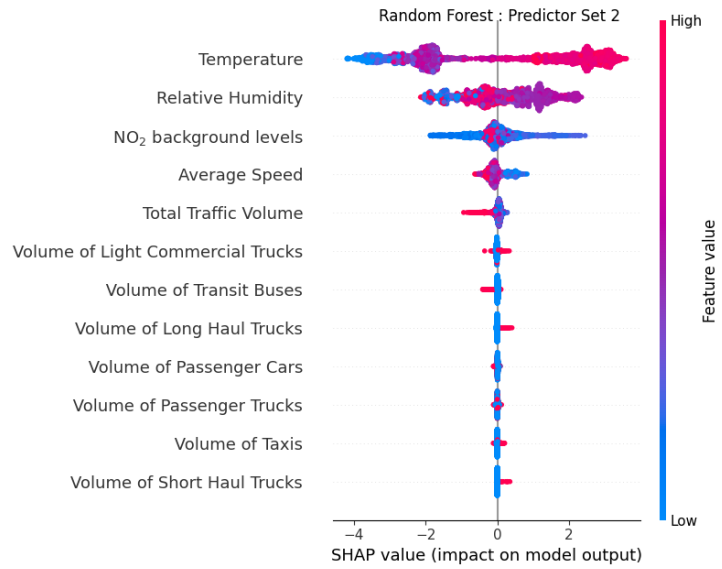
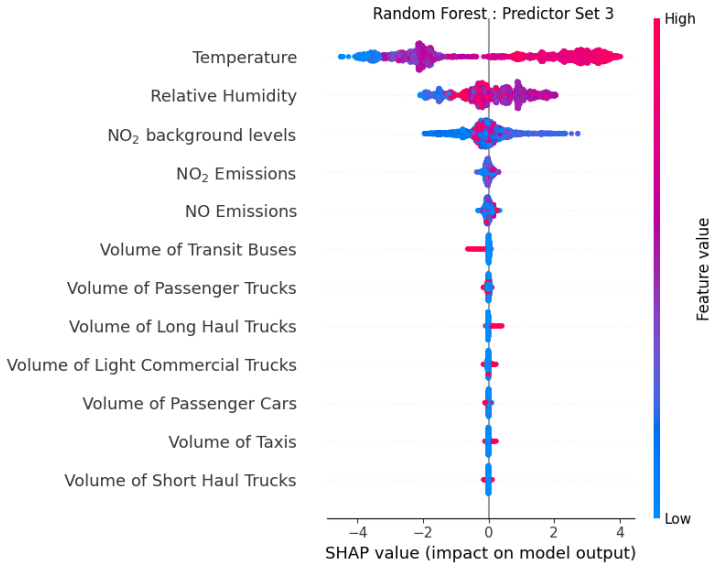


Figure S2.17. SHAP summary plot for $PM_{2.5}$ with XGBoost with Predictor Sets 3-6.







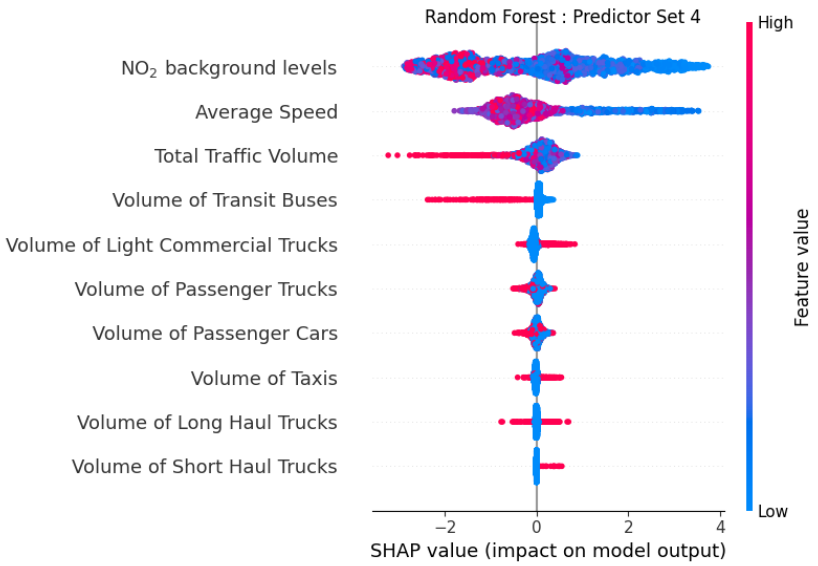
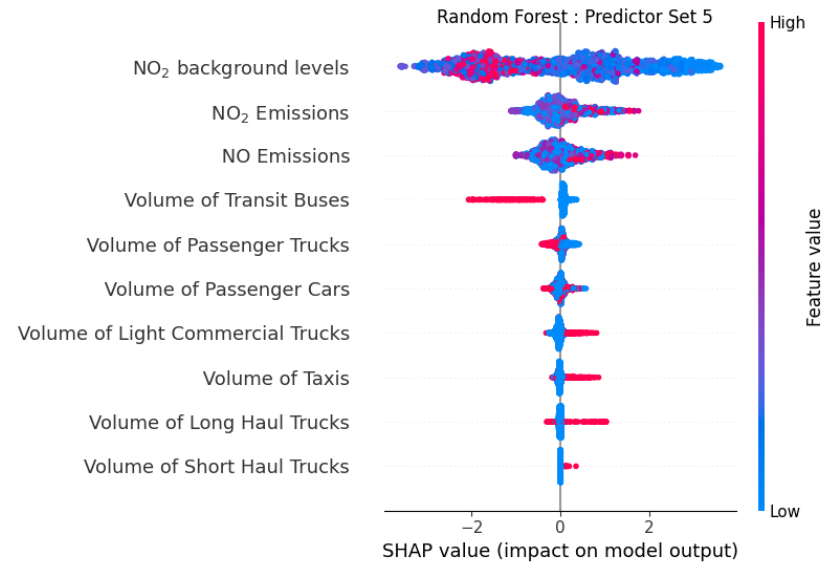


Figure S2.18. SHAP summary plot for NO_2 with Random Forest with Predictor Sets 1- 4.



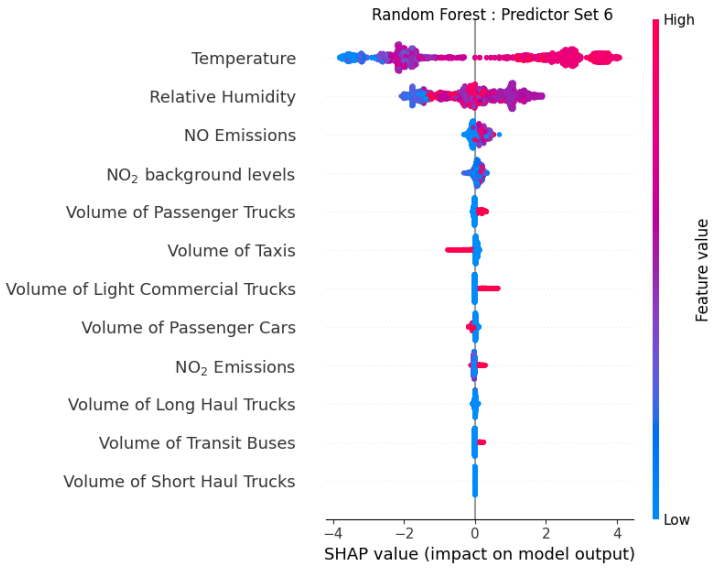
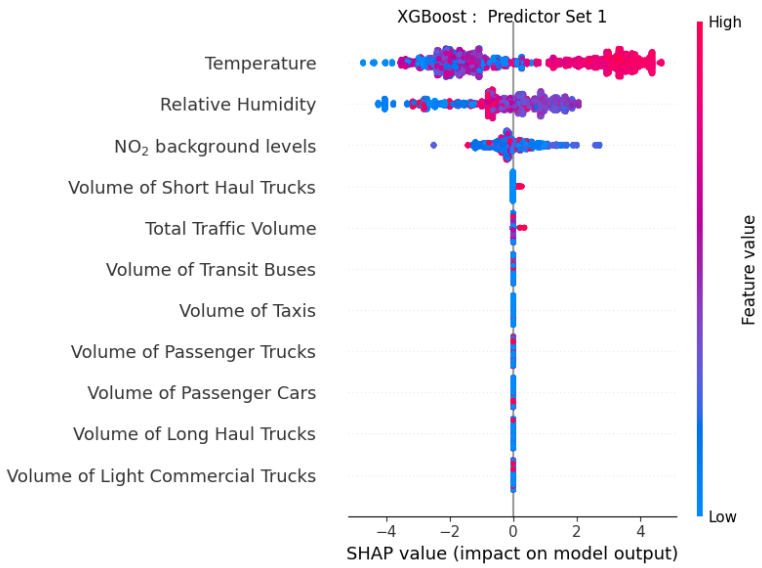


Figure S2.19. SHAP summary plot for NO_2 with Random Forest with Predictor Sets 5- 6.



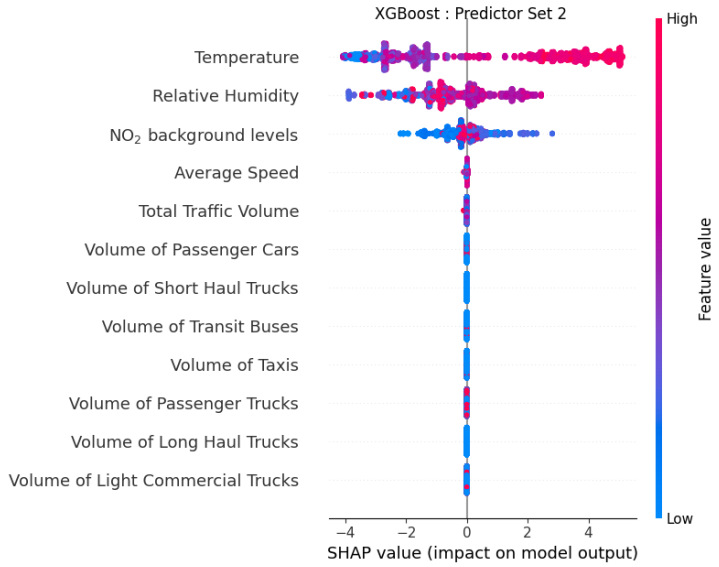


Figure S2.20. SHAP summary plot for NO_2 with XGBoost with Predictor Sets 1- 2.

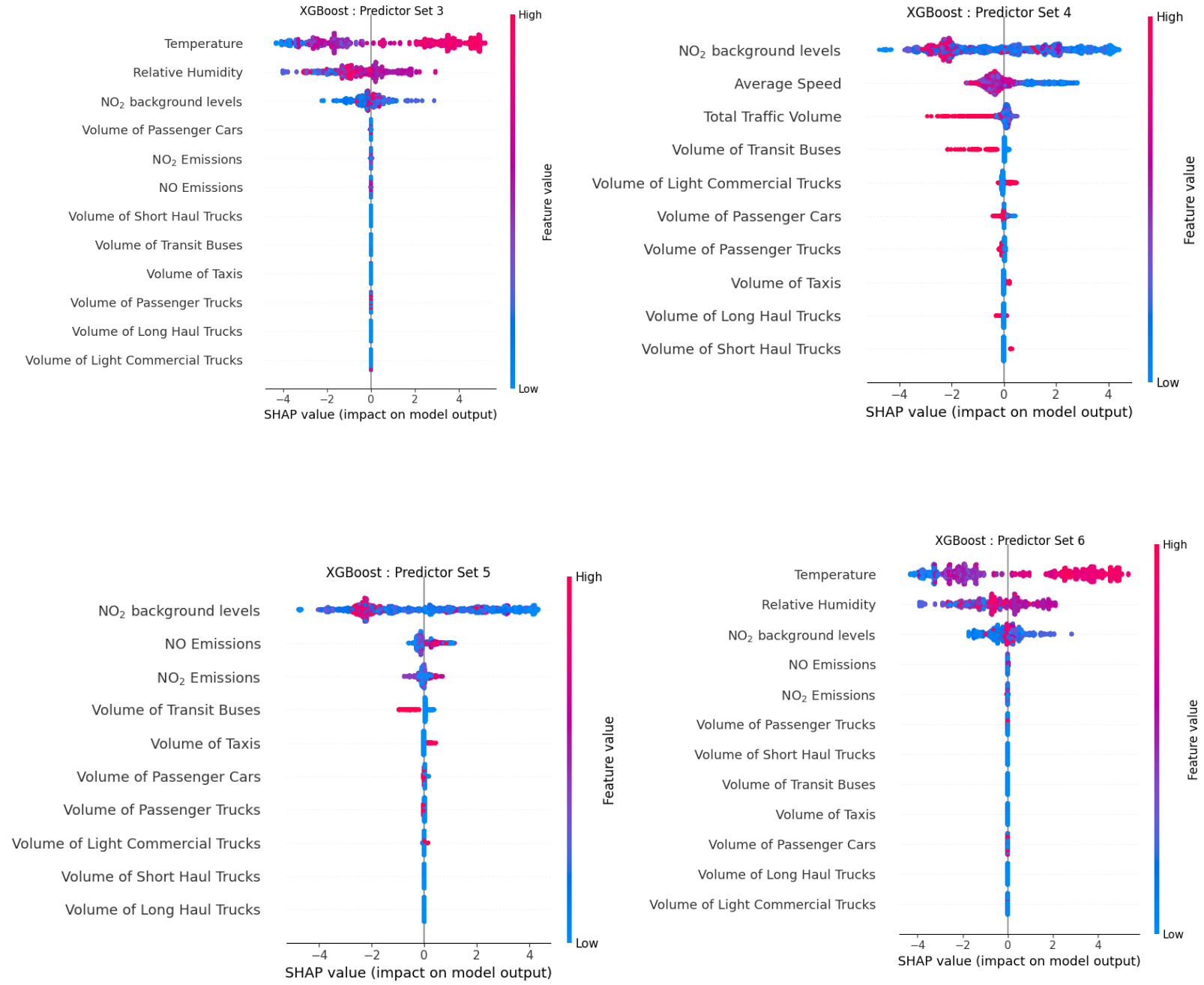
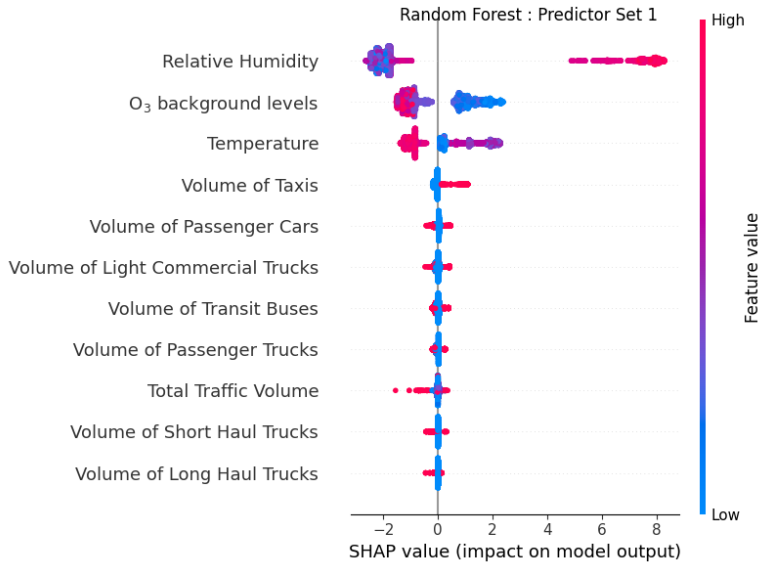
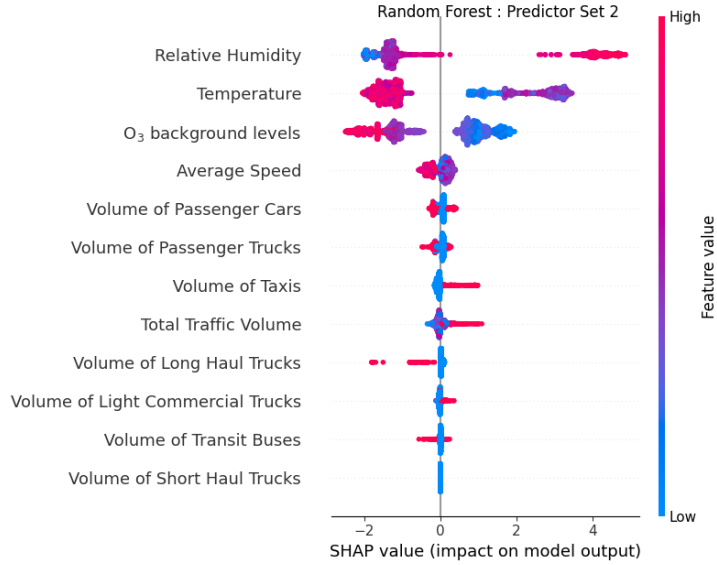
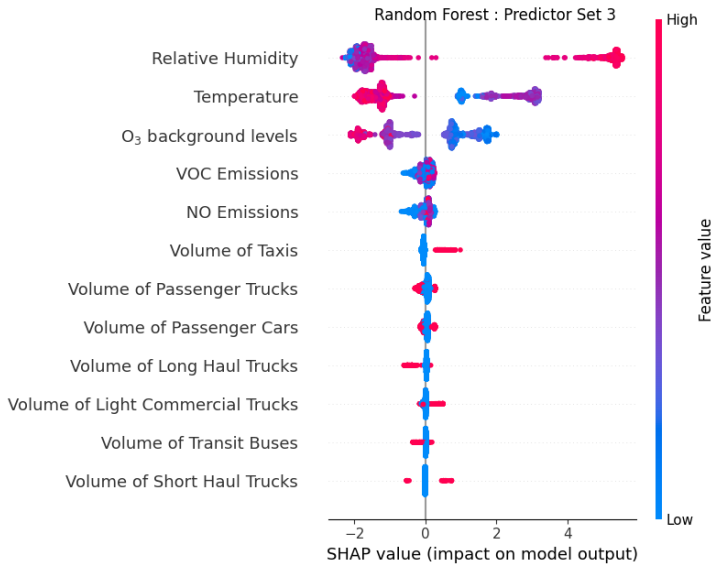


Figure S2.21. SHAP summary plot for NO_2 with XGBoost with Predictor Sets 3- 6.







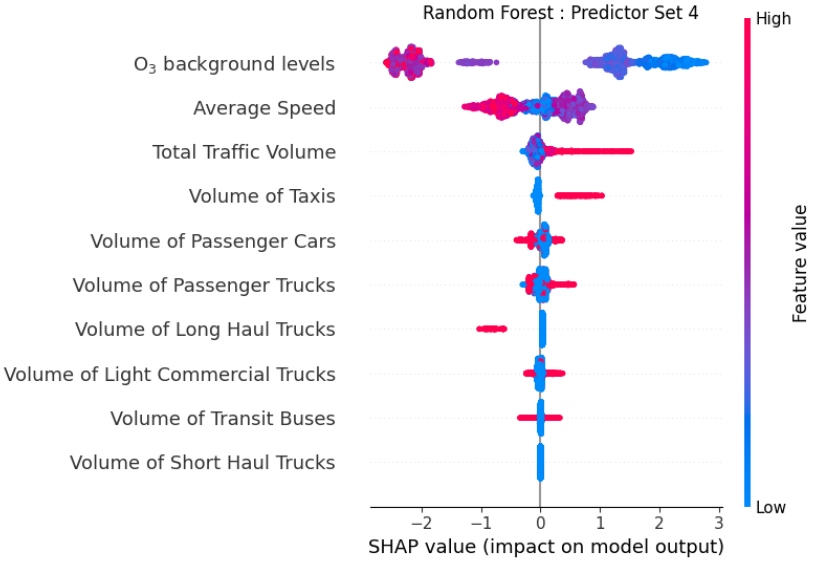
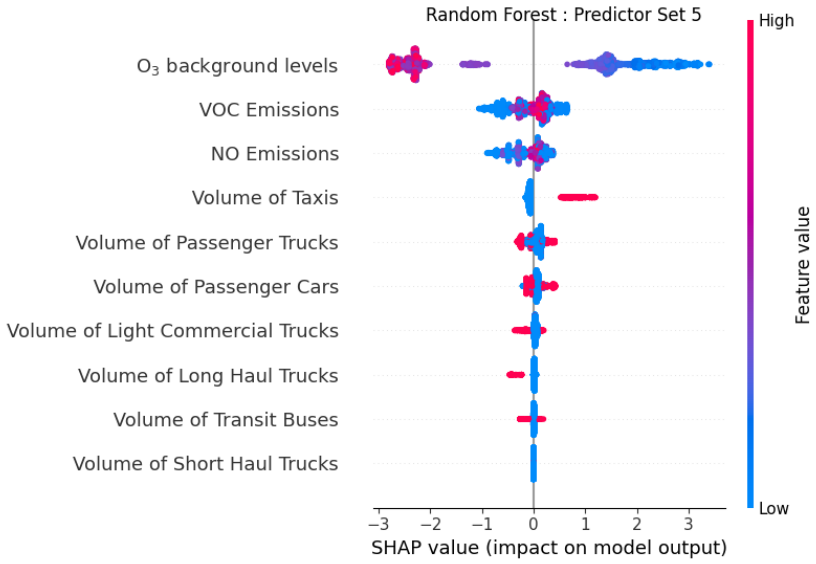


Figure S2.22. SHAP summary plot for O_3 with Random Forest with Predictor Sets 1- 4.



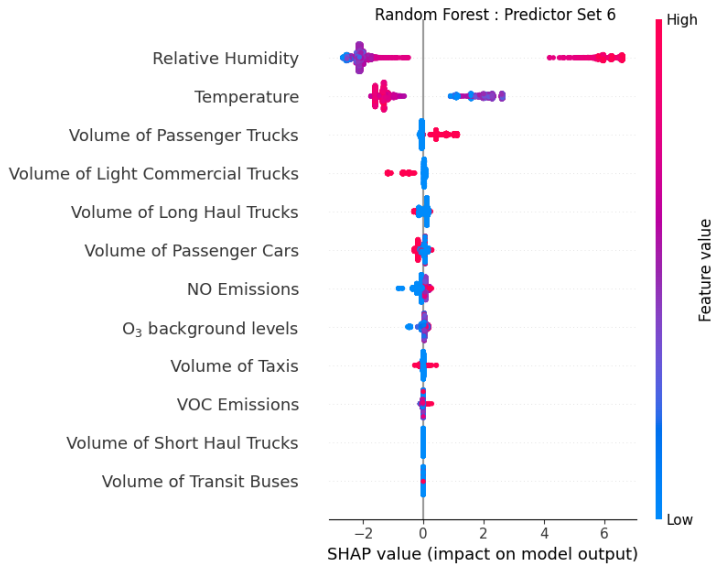
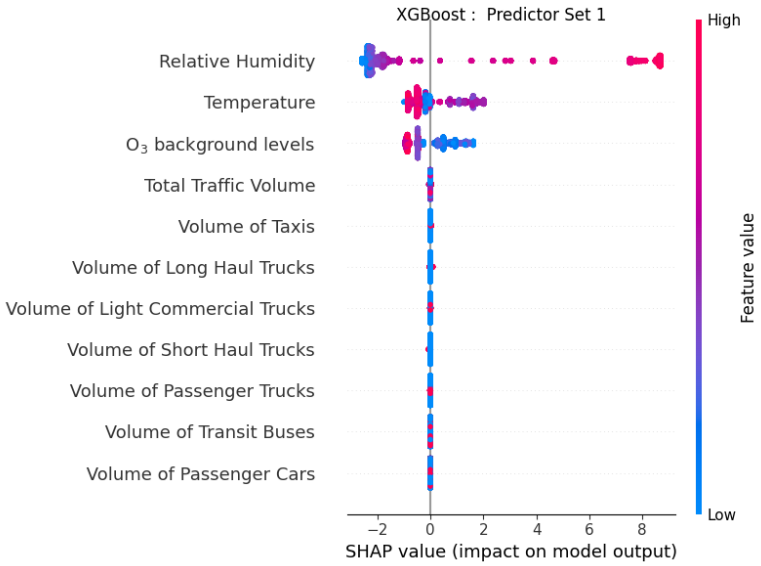


Figure S2.23. SHAP summary plot for O_3 with Random Forest with Predictor Sets 5- 6.



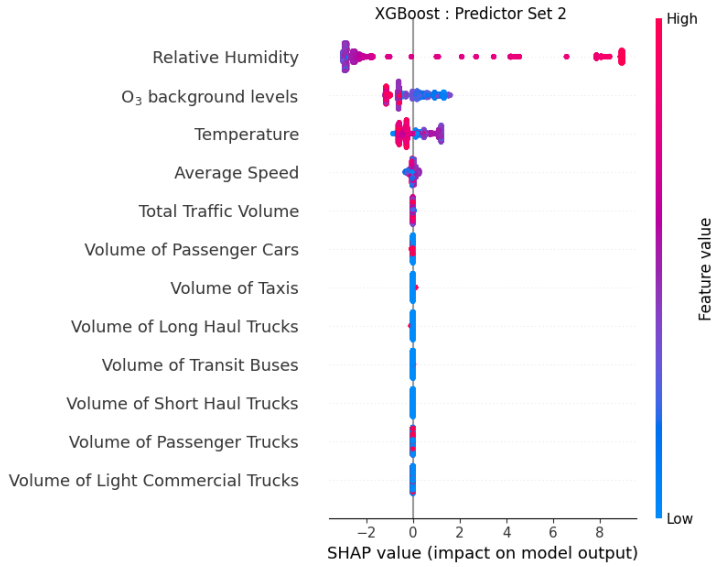
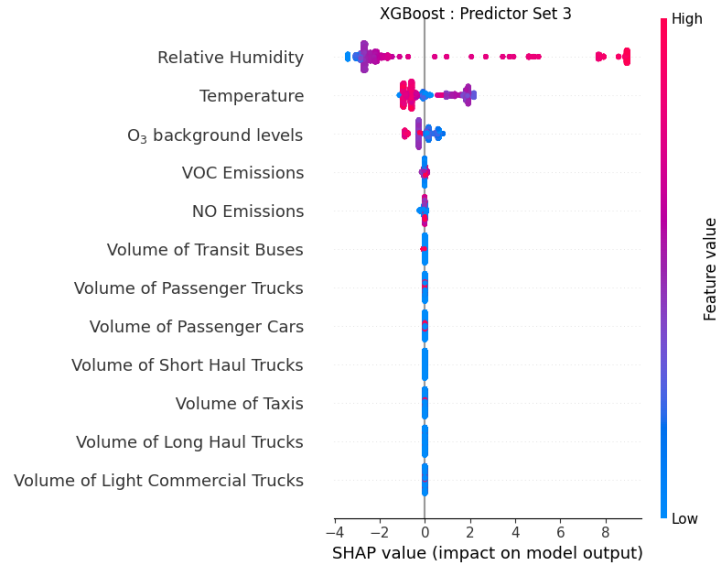
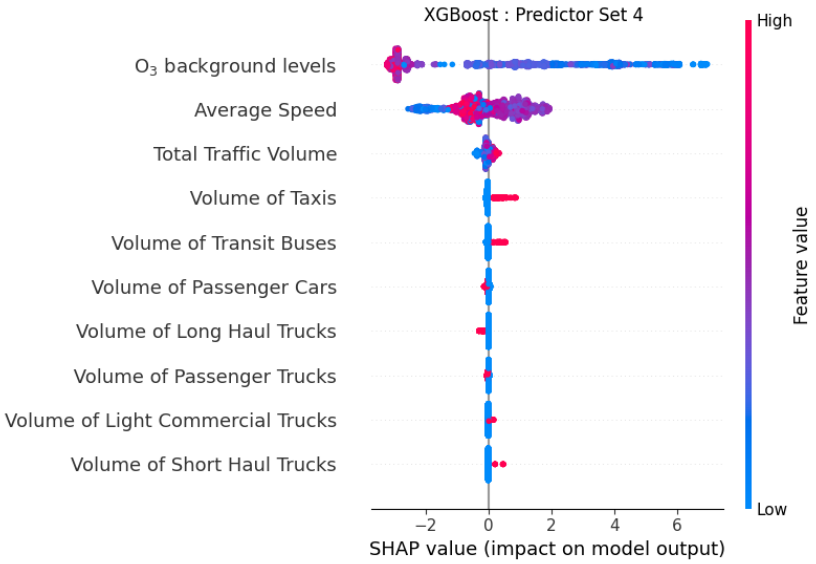
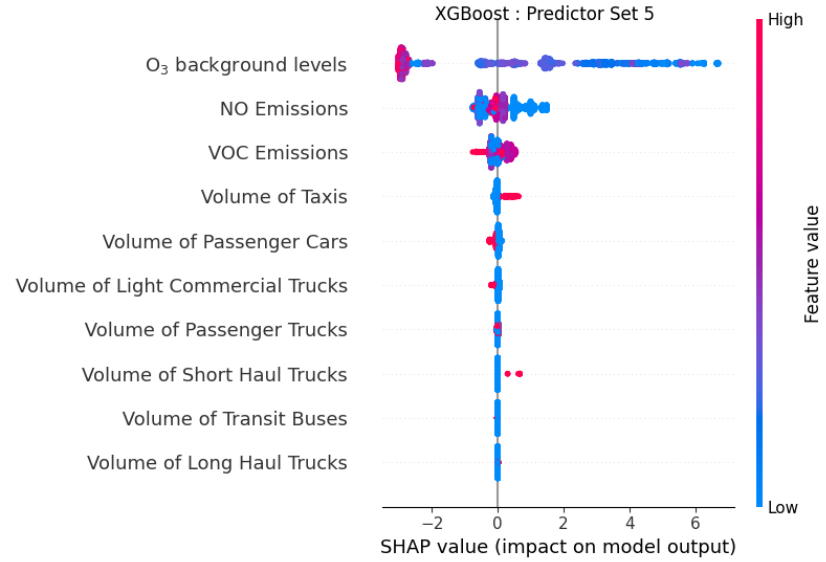


Figure S2.24. SHAP summary plot for O_3 with XGBoost with Predictor Sets 1- 2.







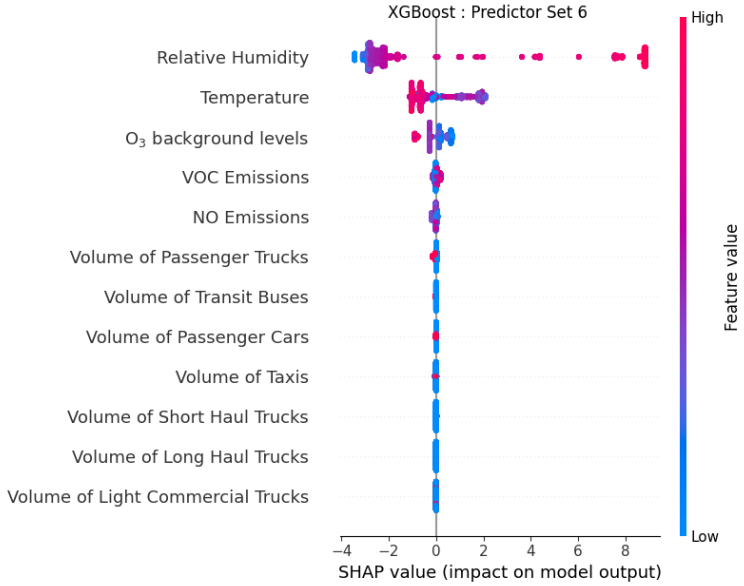


Figure S2.25. SHAP summary plot for O_3 with XGBoost with Predictor Sets 3- 6.

CHAPTER 3

Research Paper 2: 'Beyond the Horizon: A Study on the Influence of Cruise Ship Activities on Traffic Patterns and Local Air Quality'

This chapter is prepared as a submitted for publication. Entitled 'A Study on the Influence of Cruise Ship Activities on Traffic Patterns and Local Air Quality'. Its main author is also the author of this thesis, having done its research, calculations and writing in consultation with Dr. Laura Minet as the primary advisor. All section, equation, and reference numbering has been modified to integrate it with this thesis.

Abstract

The cruise ship industry, the highest growing sector in maritime transport, has gained attention for its contribution to local and global emissions. While existing studies have focused on emissions from cruise ships, little attention is given to activities induced by cruise ships, particularly their influence on on-road traffic. Passenger's reliance on numerous coach buses and taxi rides heightens transportation activities in the area, contributing to an increase in the local air pollution from vehicular traffic. This study investigates the impact of cruise-ship induced traffic on the local air quality in James, Bay neighborhood of Victoria, British Columbia. Air quality monitoring stations equipped with sensors were placed in volunteer households during the cruise ship season to measure carbon monoxide (CO), nitrogen dioxide (NO_2), ozone (O_3) and fine particulate matter ($PM_{2.5}$) concentrations. A combination of air pollutant concentrations, cruise ship data, and traffic modeling were used to analyze the local air quality impact. Results indicated that both cruise ship activity and traffic data share influences on pollutant levels, with NO emissions primarily stemming from traffic, particularly buses and passenger vehicles. Cruise ship activity indirectly affects air quality by influencing local traffic patterns.

3.1 Introduction

The cruise ship industry represents the highest growing asset in maritime transport, boasting an average annual growth rate of 7.37% over the last 30 years globally. In Canada, the industry, which has over 900 harbours and ports across the world's longest coastline, contributes more than $4 billion annual input into the economy and directly and indirectly generates an average of

30,000 jobs. Projections for the industry's revenue indicate an anticipated increase of 8.32% with number of cruise users peaking to 990,000 by 2027 which equates to $1,532 average revenue by user (Transport Canada 2022). The industry's growth has been associated with various factors such as new technologies, tourist activities to satisfy consumer demands, and port availability (Lloret et al. 2021a).

The rise in cruise ship numbers over the years has spurred considerable research to analyse their impacts. The industry has been linked to a broad range of externalities including political, socio-cultural and environmental costs and benefits (Calatayud, Sánchez González, and Marquez 2022). Economically, the industry generates revenue from docking and parking fees, as well as boosts of local food/entertainment sectors (Hoogkamer 2013). Socio-culturally, cruise ship impacts can lead to loss of neighbourhood, overcrowded public spaces and contribute to homogenization effect where port facilities are dominated by cruise operators, affecting quality of life in local communities (McCarthy 2018). Environmental impacts are associated with increased pollution of air, water (both fresh and marine), soil and land cover, noise contamination and resulted health effects to passengers, crew, residents living near cruise ports and working in shipyards (Lloret et al. 2021b). Studies investigating the emission impacts from cruise ships on both global and economic scales have highlighted that shipping activities and vessels are influencing factors of pollution in cities due to the proximity of ports to urban areas (Perdiguero and Sanz 2020). Exposure to air pollution is one of the major causes of diseases and premature death and represents the largest environmental risk factor to human health (WHO 2016). Air pollutants like nitrogen oxides (NO_x), sulphur oxides (SO_x), particulate matter ($PM_{2.5}$ and PM_{10}), and carbon monoxide (CO) pose serious health risks including respiratory infections, lung cancer, stroke, heart disease, and cardiovascular conditions (IAPH, 2007).

Numerous studies have delved into the impact of cruise ship activity on local air quality in port surroundings (Eckhardt et al. 2013; Kennedy 2019; Lloret et al. 2021b; Mölders, Gende, and Pirhalla 2013). Murena et al.(2018) estimated that approximately 70% of ship emissions occur within 400 km from coasts in Naples, Italy. Perdiguero & Sanz (2020) conducted a comprehensive examination, considering factors such as cruise arrivals and port stays, on air quality across the entire city of Barcelona. Results from their study indicate that emissions of NO_x, CO, PM_{10} and SO_2 from cruise ships led to a decline in air quality throughout the city. A similar investigation in Victoria, BC, demonstrated elevated concentrations of SO_2, NO and $PM_{2.5}$, particularly during weekends coinciding with the presence of ships (Poplawski et al. 2011).

While existing literature has predominantly focused on the direct emissions from cruise ships and their influence on local air quality, very little attention has been paid to the induced activities stemming from the presence of these ships and their numerous passengers, such as increased vehicular traffic in port cities. Indeed, the influx of passengers from cruise ships often places substantial strain on cities, particularly in cases where they all arrive simultaneously (International Transport Forum 2016). Passengers typically rely on numerous coach buses and

taxi rides, which heightens transportation activities in the area and contributes to increases in traffic-related air pollution (e.g., NO_x, O_3, PM, CO). Calatayud et al. (2022) examined the correlation between cruise ship activity and traffic congestion, revealing a 12% increase in urban congestion associated with cruise ship activity in areas surrounding the port. Traffic-related air pollution is known to have effects on human health, and it is imperative that it is studied in conjunction with cruise activities. Taking Victoria, BC, as a testbed, this study aims to bridge the gap in literature by not only focusing on the emissions and traffic congestion from cruise activity but also investigating the significant relationship between traffic-related air pollutant concentrations and the presence of cruise ships.

3.2 Study Area

The hub for cruise ship arrivals in Victoria is the Ogden Point terminal, situated in the southwest corner of James Bay. James Bay is a residential neighbourhood that borders downtown Victoria, with a total area of 239 hectares and surrounded by water on three sides (Bendle 2012). Notably, the Ogden terminal is strategically located within 500 m of the nearest residential dwellings and has four deep sea berths ranging from 240 to 335 meters. The harbour is open 24 hours a day, seven days a week, year-round, and generates revenue through various cruise activities, including cargo ship maintenance and berthing of commercial vessels. The terminal welcomed a record 326 cruise ships and hosted 970,000 cruise visitors to Victoria in 2023, marking a 26% increase in passenger arrivals compared to 2018. Additionally, the area serves as a busy transportation hub

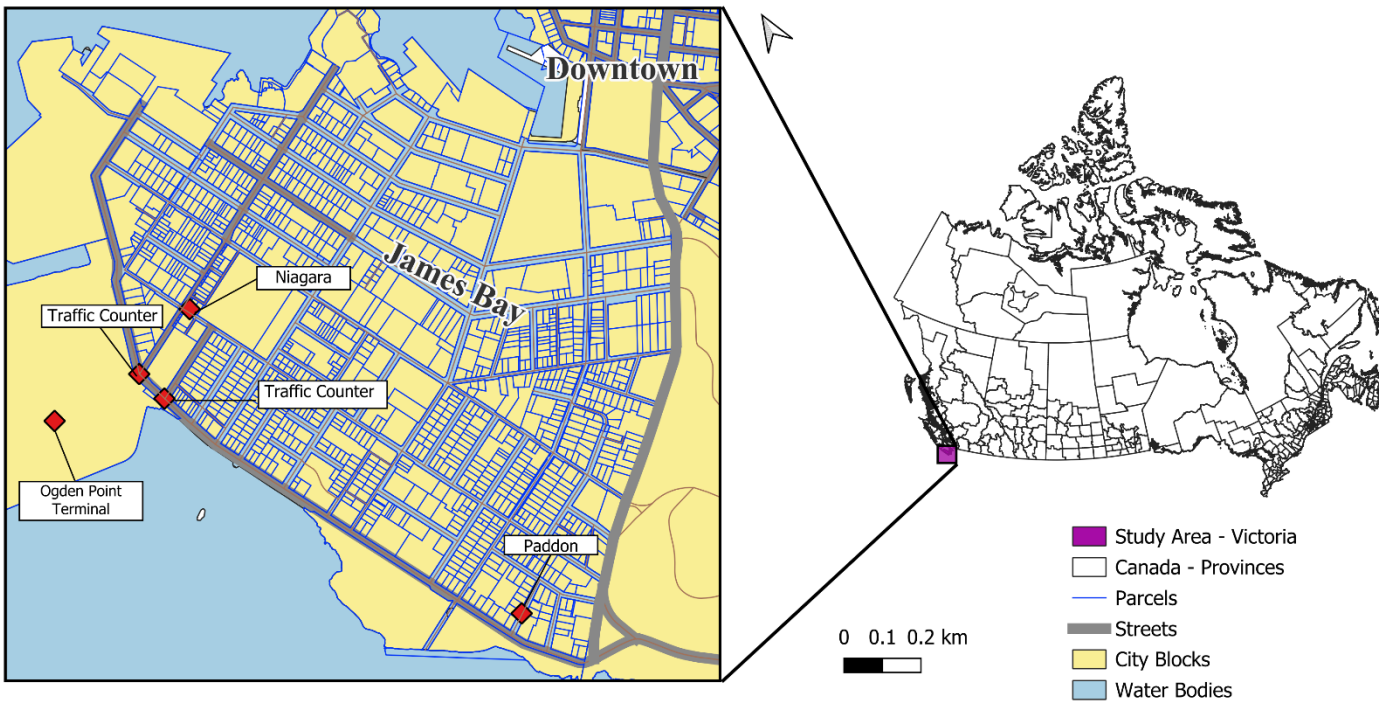


Figure 3.1. Location of cruise ship terminal, air quality monitors and traffic tube counters.

for buses, taxis, and other tour operators, facilitating the conveyance of passengers to tourist destinations. Most of the cruise ships which visit Victoria are from the United States (U.S), and they use this port as a stop in order to comply with the Passenger Vessel Services Act (PVSA) of 1886. This law requires that ships which are flagged outside the U.S and will travel between U.S ports make a stop at a foreign port. This makes Victoria a convenient location for most Alaska-bound cruises originating from or ending in U.S ports such as Seattle and San Francisco. As a result, the ships rarely stay overnight in Victoria, and typically arrive either late afternoon or evening and depart a few hours later. While this limited time at the port means Victoria may not be the primary destination for many cruises, it still remains a popular stop due to its historic sites and attractions. The cruise ship season typically spans from April to October. Aside cruise ships and vehicular traffic during this season, other emission sources in the neighbourhood of James Bay include float planes, helijet services, and passenger ferries. However, these services normally operate north of the area and have relatively stable daily schedules. This allows us to analyze peaks in air pollutant concentrations in relation to cruise ship activities and transportation services conveying passengers.

3.3 Methodology

In this study, we investigated the extent to which cruise ship induced traffic impact air quality in the James Bay neighbourhood of Victoria, BC. To explore the research question, a comprehensive dataset encompassing cruise ship activity, traffic information, and air pollutant concentrations was collected during the 2023 cruise ship season (April – October). We conducted correlation analyses to understand the relationship between cruise ship schedule, traffic, and air pollutant levels in the area.

3.3.1 Instrumentation

We conducted a data collection campaign which involved measuring nitric oxide (NO), nitrogen dioxide (NO_2), fine particulate matter ($PM_{2.5}$), and ozone (O_3) concentrations at two volunteer households in James Bay between June and November 2023. To do this, we employed the low-cost sensor air quality monitor Real-time Affordable Multi-Pollutant RAMP (Anon 2019) in the front yard of two volunteer households. The volunteer households were situated close to Niagara Street and Paddon Avenue (as shown in Figure 3.1), which are located at 400 m and 1,500 m from the Ogden Terminal, respectively. To account for the traffic density at the Niagara location, we placed tube counters for collecting traffic volumes of different vehicle classes on Dallas Road west of Montreal Street (indicated in Figure 3.1), which is located approximately 230 m from Niagara. For the traffic through the Paddon location, the tube counter was placed at Dallas Rd West of Dock Street (indicated in Figure 3.1), which is located at approximately 1,100 m from Paddon. The counter on Dallas Rd experiences the most traffic all year round, from Ogden Point to downtown Victoria. This is because the shortest route to downtown passes through this location, hence it attracts high traffic throughout the year.

The decision to place the air pollutant monitors at both locations were due to several factors. Firstly, Niagara Street is very close to Ogden Point terminal and is a major residential throughfare in the James Bay neighborhood. This proximity means that the residents here are likely to experience direct impact of emissions from the cruise ships docked at the terminal and vehicles leaving from there. Since tourists may want to access nearby parks and other attractions, we also expect that this location will experience high volumes of traffic. Additionally, the second air pollutant monitor at Paddon Avenue, shares similar characteristics with Niagara but is situated in a quieter residential area, although farther from Ogden Point Terminal. The sensor was placed in the front yard of the volunteer's house, which is at the intersection with Paddon Avenue. Despite its quieter setting, we expect the sensor located at this location to pick up higher peaks in pollutant concentration, due to it close proximity to Dallas Rd, which has higher traffic volumes. The strategic placement of the sensors and tube counters ensures that the study captures both direct emissions from cruise ships and the vehicular traffic they generate, providing a comprehensive understanding of the air quality impacts in James Bay.

3.3.2 Calibration of the RAMP Sensor

One common drawback of low-cost sensors is the issue of cross-sensitivity with other air pollutants (Castell et al. 2017; Kumar et al. 2015; Liang 2021; Spinelle et al. 2015). For example, O_3 has been observed to undergo oxido-reduction reactions in the presence of NO_2, which can interfere with measured concentrations (Mead et al. 2013). It was therefore essential that we collocated the measurements from the low-cost sensors with measurements from the fixed regulatory monitoring station in Victoria. All the RAMP sensors used for the data collection in this study were initially setup at Victoria Topaz between May and June, and the data measured during this period were used for the collocation. We employed the methodology detailed in Chapter 2 (Section 2.2.1) for the calibration of these sensors. In this case, the minute-by-minute measurements of NO, NO_2, $PM_{2.5}$, and O_3 from the RAMP sensors were calibrated using a hybrid model combining linear and random forest models to ensure accurate measurements. The results from the calibration model were then evaluated based on performance metrics including coefficient of determination (R^2), mean absolute error (MAE), and root mean squared error (RMSE). In addition, we accounted for background concentrations using the continuous air quality monitoring station located in Colwood which is approximately 15,000 m west from the Ogden terminal. The Colwood monitoring station is situated in an area which is less influenced by local emissions from cruise ships and traffic. Hence, by comparing the measurements from the Colwood Station with the data collected from the RAMP sensors at Niagara Street and Paddon Avenue, we were able to identify the baseline levels of air pollutants which are not associated with the local sources such as traffic and emissions.

3.3.3 Traffic Data

As mentioned in the previous paragraph, tube counters were placed close to the volunteer households to estimate the traffic intensity. This process involved conducting traffic counts for two weeks from August 23 and September 4, to retrieve traffic detailed data on traffic volume

and vehicle classification. We strategically positioned the tube counters on Dallas Road west of Montreal Street and Dallas Road west of Dock Street to capture the traffic relevant to Niagara and Paddon Avenue, respectively. Total traffic volumes from these counters were used as the value for traffic density for the analysis. We collected the traffic data continuously over the two-week period, and it included 15-minute intervals, which provided insights into the peak daily traffic patterns. The tube counters also allowed us to identify to the different vehicle types present at each of the locations. These vehicle types included passenger cars (including those with recreational trailers), 2-axle pick-ups (predominantly vans, and RVs, including those with recreational trailers), buses, motorcycles, vehicles > 3-axles, and unclassified vehicles.

3.3.4 Cruise Ship Activity

Our analysis also drew information from the cruise ship schedules, which were obtained from the Greater Victoria Harbour Authority from June to November 2023. Data on the number of cruise ships docked at the Ogden Point and number of passengers was retrieved from this dataset. We first categorized the data by day and noted the total number of ships present at the port each day. In addition, we recorded the number of passengers on board the ships to provide an estimate of the human traffic associated with each docking event.

3.3.5 Correlation Analysis

We conducted correlation analyses in Python by employing Pearson correlation coefficients to quantitatively assess the strength and direction of linear associations between cruise ship activity, air pollutant concentrations, and traffic. The correlation coefficient values, which range from 1 to -1, were used to interpret the results of the analysis. For instance, when the correlation value is positive, this meant that the variables were significant positively correlated, and when the value is negative, this indicated that there was a significant negative correlation between the variables, however, a value of 0 indicated that were no significant relationship between the variables. To implement this analysis, we first retrieved the above background pollutant concentrations (difference between background and observed concentrations), tube counter data, and cruise ship schedules. This process involved handling missing values and aligning the datasets based on timestamps to facilitate accurate comparison. For assessing the impact of cruise ships, we initially analyzed the daily cruise ship and passenger numbers gathered from June to November with the air pollutant concentrations. Since the traffic data was only collected for two weeks in August and September, the correlation analysis discussed in the study was conducted with the daily traffic data, daily cruise ship activity data, and daily air pollutant concentrations for that period of time.

3.4 Results and Discussions

3.4.1 Cruise Ship Activity Over Entire Summer Vs Two Weeks

Figure 3.2 presents the daily cruise ship activity over the entire data collection period from mid-June to November. We can see distinct patterns in the ship activity throughout this time frame.

Daily passenger data is also shown in Figures S3.1(entire summer) and S3.2 (two weeks) in the Supplementary Information. At the beginning of the data collection in mid- June to July, we observed frequent activity with peaks reaching up to three ships per day. This increased activity

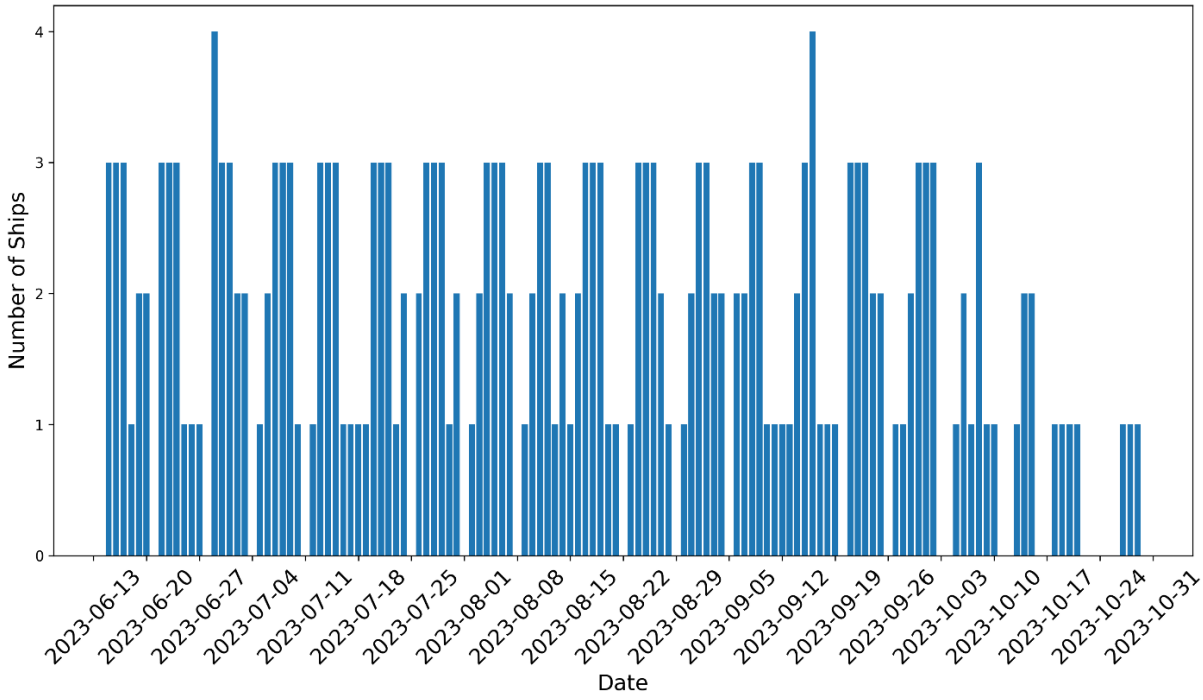


Figure 3.2. Daily cruise ship numbers for two weeks concurrent with traffic data.

continues into August and September, although we observed some fluctuations, but generally showed reduced activity in September compared to August. We observed that October was the month with the most decline in cruise ship numbers and more days where there was no cruise ship. This was expected as the cruise ship season was ending during this month. The highest cruise activity was often on Fridays and Saturdays, with an average of three ships per day (shown in Figure S3.3 in the Supplementary Information). Thursdays were more moderate days with two ships per day but the lowest days with cruise ship activities were Mondays, Wednesdays, and Sundays with no cruise ships on Tuesday.

When we focused our analysis on the two weeks concurrent with traffic data (presented in Figure 3.3), on average there were about two to three cruise ships arriving daily and the passenger numbers averaged around 5600 per day (shown in Figure S3.2), with high variability during the observation period. Although the number of cruise ships during this period was consistent with the overall dataset average of 2 ships per day, passenger numbers were notably higher by approximately 14% compared to the entire dataset average of 4,940 passengers per day. Our observations suggest that there was a late summer surge in passenger traffic, which aligned with the peak tourism season of late August. The cruise ship numbers recorded for each day of the

week is shown in Figure S3.4 in the Supplementary Information. The highest number of cruise ships arriving on a single day was 3, occurring on 5 days throughout this period. The peaks in the passenger numbers, typically falling on Thursdays, Fridays, and Saturdays, reinforcing the peak activity period during late August.

3.4.2 Air Pollutant Concentrations (Observed and Background) Over Entire Summer Vs Two Weeks

Measurements from the RAMP sensors were calibrated to ensure accuracy in further analysis. The results from the calibration for each of the sensors are presented in Table S3.1 in the Supplementary Information. The R^2 values obtained for the sensors which were deployed at the Niagara location were 0.87, 0.74, 0.88, and 0.69 for NO, NO_2, O_3, and $PM_{2.5}$ respectively and that for the Paddon location were 0.9, 0.69,0.84, and 0.60 for NO, NO_2, O_3 and $PM_{2.5}$ respectively. After calibration, the RAMP sensors were deployed at the volunteer households to

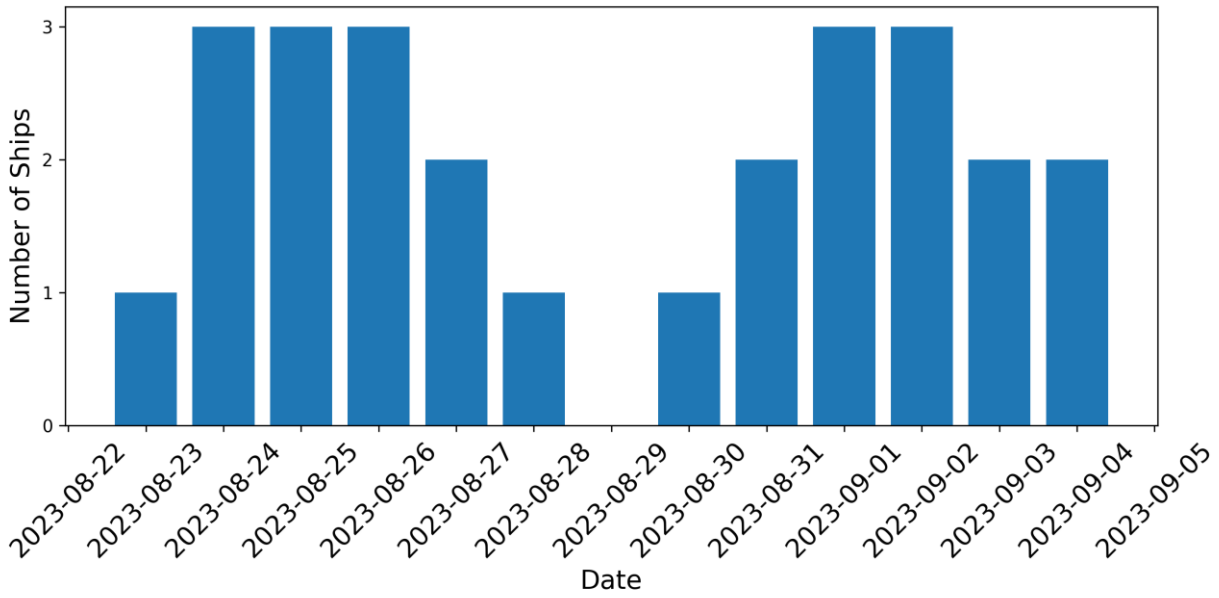


Figure 3.3. Daily cruise ship numbers for two weeks concurrent with traffic data.

Table 3.1 Descriptive Statistics of hourly average pollutant concentrations at Niagara and Paddon for the whole summer.

	Niagara				Paddon			
Pollutants	**Mean**	**Minimum**	**Maximum**	**Standard Deviation**	**Mean**	**Minimum**	**Maximum**	**Standard Deviation**
NO (ppb)	4.4	0.1	158.7	11.9	2.5	0.0	41.7	3.4
NO_2 (ppb)	8.6	1.2	66.9	6.1	7.4	0.7	21.5	3.6
$PM_{2.5}$ ($\mu g/m^3$)	9.0	3.3	62.5	5.8	8.4	3.8	38.4	4.8
O_3 (ppb)	17.3	0.2	36.1	8.5	19.5	0.5	42.6	9.1

Table 3.2 Descriptive Statistics of hourly average pollutant concentrations at Niagara and Paddon for two weeks.

	Niagara				Paddon			
Pollutants	**Mean**	**Minimum**	**Maximum**	**Standard Deviation**	**Mean**	**Minimum**	**Maximum**	**Standard Deviation**
NO (ppb)	2.8	0.3	30.1	3.4	2.6	0.2	15.4	2.9
NO_2 (ppb)	9.4	3.2	22.0	3.5	8.1	2.3	19.7	2.9
$PM_{2.5}$ ($\mu g/m^3$)	16.5	5.3	39.1	8.1	14.9	5.6	32.8	7.4
O_3 (ppb)	17.8	0.4	35.7	9.4	21.7	2.4	42.6	10.2

Table 3.3 Descriptive Statistics of background concentration recorded for the whole summer and the two weeks.

	Entire Summer				Two Weeks (August 23 – September 4)			
Pollutants	**Mean**	**Minimum**	**Maximum**	**Standard Deviation**	**Mean**	**Minimum**	**Maximum**	**Standard Deviation**
NO (ppb)	0.6	0.1	7.3	0.9	1.1	0.1	49.4	2.7
NO_2 (ppb)	4.3	0.4	16.2	3.3	3.8	0.2	22.9	3.4
$PM_{2.5}$ ($\mu g/m^3$)	12.6	-4.0	43.0	9.7	6.3	-7.0	59.0	7.1
O_3 (ppb)	24.2	1.6	61.7	13.1	21.2	0.2	61.7	11.1

collect the pollutant concentration levels, with additional background air pollutant data collected at the Colwood station. Our analysis of the air pollutant concentrations at both locations revealed distinct differences between the observed pollutant data and background concentrations over the entire data collection period (June – November) and the two weeks concurrent with traffic data (August 23 – September 4). Descriptive statistics for the hourly concentrations at both Niagara

and Paddon for the entire summer and the two weeks concurrent with traffic data are presented in Table 3.1 and 3.2, respectively. Time series for the daily pollutant levels for the whole summer are shown in Figure S3.4 – S3.5 in the Supplementary Information.

For the entire data collection period at the Niagara location, NO, NO_2, and $PM_{2.5}$ showed high variability particularly prominent in June, July, and October. The hourly average concentrations for NO and NO_2 were around 4.40 ppb and 8.64 ppb, respectively, with $PM_{2.5}$ averaging 9.04μg/m³. In contrast, O_3 levels at Niagara had an average concentration of 17.30 ppb, which demonstrated a different pattern influenced by atmospheric conditions.

A similar observation was made at the Paddon location, where the measurements recorded showed high variability in NO, NO_2, and $PM_{2.5}$, with fluctuations during the same month as the Niagara location. The average concentrations for this location were however slightly lower, with NO at 2.52 ppb, NO_2 at 7.37 ppb, and $PM_{2.5}$ at 8.43 μg/m³, O_3 levels at Paddon were slightly higher with an average of 19.47 ppb. At both locations, Paddon and Niagara, $PM_{2.5}$ were low and stable around 5 μg/m³ from June to mid-July. We however saw a noticeable increase (about 10 μg/m³) for $PM_{2.5}$ levels at Niagara in early August, and although there were fluctuations after this, we observed the highest peak for this pollutant in late August. For Paddon, a similar pattern was made where the $PM_{2.5}$ levels were stable until late July, then rose in mid-August with continued fluctuations in September and October. NO levels, on the other hand, showed high variability with frequent peaks, while the NO_2 concentrations remained relatively stable at the Niagara locations. O_3 also had peaks in late August and early September.

Descriptive statistics for the hourly average background concentrations for the data collection period and the two weeks have been displayed in Table 3.3. The comparison between the observed and background concentrations highlighted the substantial impact of localized activities on the air quality at both the Niagara and Paddon locations. In that, over the whole summer, at Niagara, NO levels were on average, 76% higher than the background levels, while NO_2 concentrations were 56% more. $PM_{2.5}$ concentrations were also 30% higher than the background measurements. Similarly, at the Paddon location, we observed that average NO concentrations were 137% higher, NO_2 were 94% higher, and $PM_{2.5}$ were 32% higher than the measurements recorded at the Colwood station. The background data, which represented the regional air quality showed much lower concentrations for these pollutants and suggested that these areas were influenced greatly by the local emissions. In contrast, O_3 levels were 22% higher at Niagara and 8% higher at Paddon in the background data compared to the observed measurements, which are likely due to photochemical reactions driven by sunlight and temperature. These reactions can overshadow the effects of O_3 precursors such as NO, hence lead to lower observed ozone levels in urban areas compared to the background measurements.

During the period of August 23 to September 4, concurrent with the traffic data, we observed unique trends at both locations. Time series for the hourly pollutant levels during this period are shown in Figure 3.4. At Niagara and Paddon, NO levels averaged 2.79 ppb and 2.59 ppb,

respectively. The measurements were slightly lower compared to the whole summer average, yet they still remained substantially higher than the background level of 0.63 ppb.

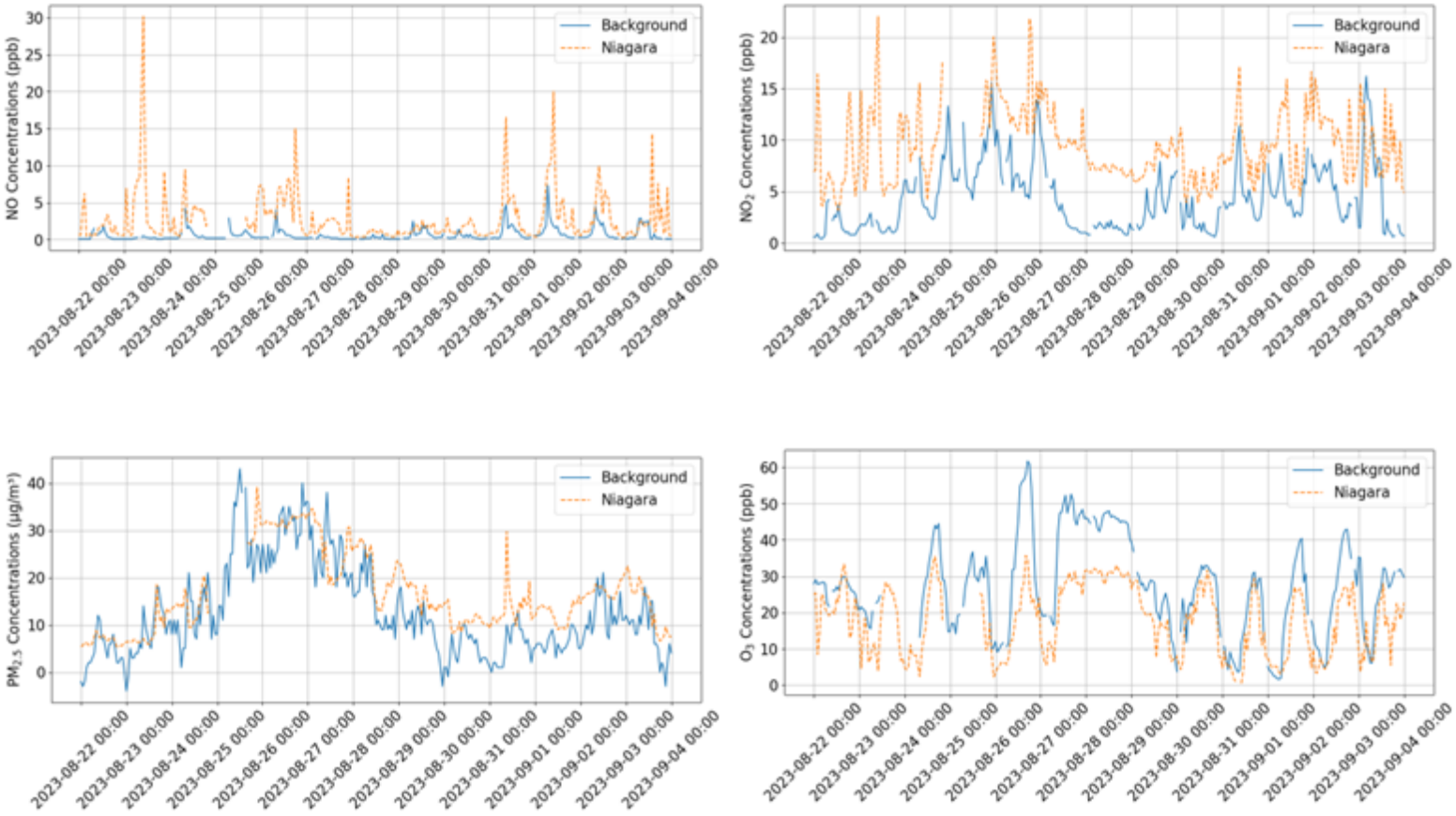


Figure 3.4. Average hourly NO, NO_2 , $PM_{2.5}$, and O_3 concentrations measured at Niagara (observed) and Colwood (background).

NO_2 concentrations remained elevated at an average of 9.44 ppb at Niagara and 8.12 ppb at Paddon, also slightly higher than the average for the whole summer and still higher than the background levels which averaged 4.34 ppb at this time. With respect to $PM_{2.5}$ concentrations , we observed they rose significantly, averaging 16.47 µg/m³ at Niagara and 9.15µg/m³ at Paddon, an 84% at Niagara and 28% increase at Paddon. Again, they were also higher that the background levels (12.64 µg/m³). O_3 levels however remained stable with an average of 17.78 and 21.70 ppb, at Niagara and Paddon, respectively. Similarly, their background levels were also much higher (27.23 ppb) during this period. The differences between pollutant concentrations at Niagara and Paddon can be attributed primarily to their relative proximity to the cruise ships. Niagara, being closer to the cruise ship terminal, is more directly impacted by emissions from the ships themselves, and likely leads to higher concentrations of the pollutants. In contrast, Paddon, being further from the cruise ships, is less exposed to these direct emissions, which resulted in comparatively lower pollutant levels. These observations from the two weeks data concurrent

with traffic provide a focused picture of the localised emissions, which makes it an ideal period for detailed analysis to understand the impact of the specific local activities on the air quality.

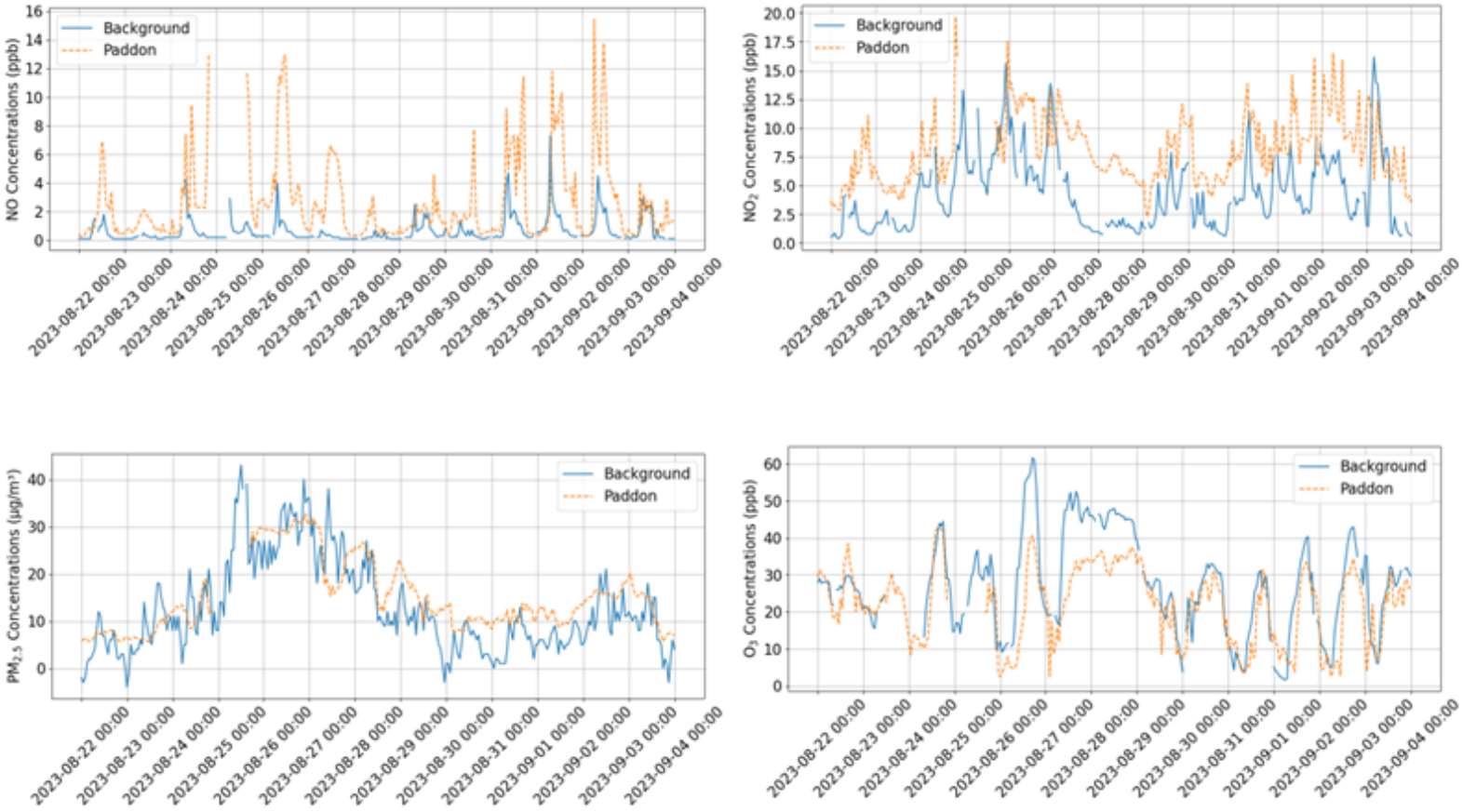


Figure 3.5. Average hourly NO, NO_2 , $PM_{2.5}$, and O_3 concentrations measured at Paddon (observed) and Colwood (background).

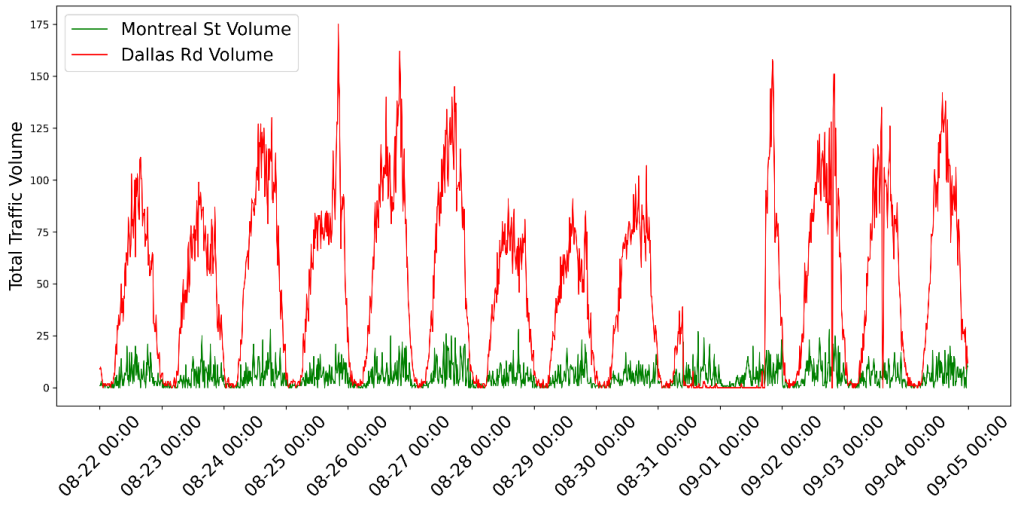


Figure 3.6. Hourly traffic volume from tube counters placed on Montreal Street and Dallas Rd

3.4.3 Traffic Data Collected Over the Two Weeks

The traffic dataset included 15-minute traffic volume and vehicle types recorded with the tube counters set up close to the two locations. There were variabilities in the hourly traffic data recorded from the tube counters at the two locations. The hourly total traffic volume for both tube counters is shown in Figure 3.6 along with the vehicle classification distribution in Figures S3.6 and S3.7 in the Supplementary Information. The traffic flow from the counter close to Paddon (Dallas Rd) was higher than the flow from the counter close to Niagara (on Montreal Street), which is expected since most vehicles in that area are either from/going to downtown Victoria, and the shortest route is through this area. We observed that there were peaks in the traffic volume toward the end of August and early September. An emphasis on the vehicle types further highlighted the differences observed between Montreal St and Dallas Rd. At the Montreal St, the counts showed higher volumes and fluctuations for passenger cars, with a notable peak on August 30. Other vehicle classes, including buses, vans and RVs, buses, vehicles with more than three axles, and unclassified vehicles, exhibit relatively stable but much lower volumes compared to passenger cars.

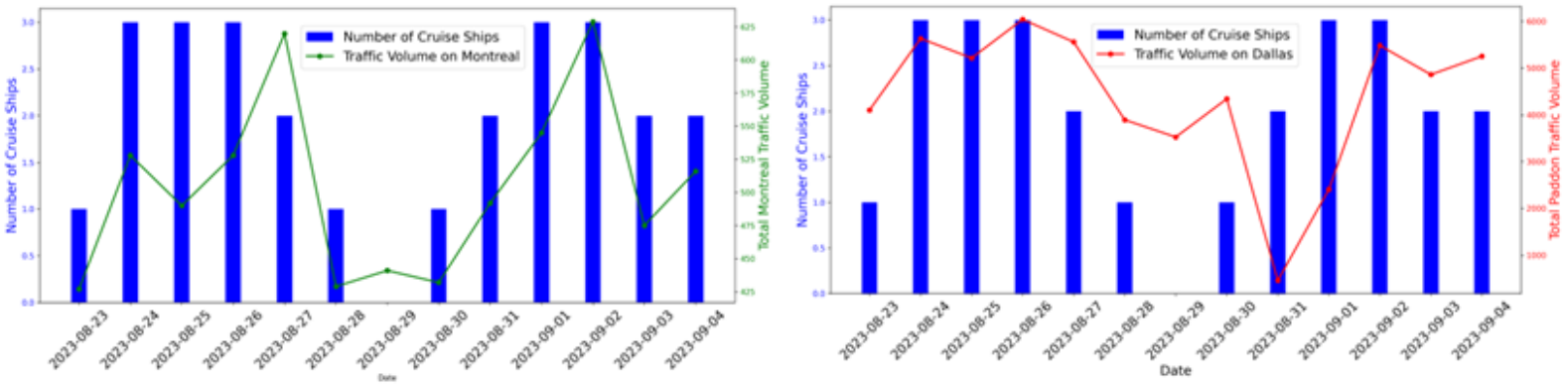


Figure 3.7. Daily cruise ship and traffic pattern at Montreal St and Dallas Rd (August 23 – September 4,2023).

In addition, we focused the analysis on the dynamics between daily cruise ships and traffic patterns. Our analysis revealed that at Montreal (shown in Figure 3.7), traffic volume peaks aligned with the higher number of cruise ships on specific dates, such as August 27 and September 2, which suggests that cruise ship arrivals contributed to the increased traffic. However, we observed inconsistencies on days like August 24 – August 26, and September 3, where the traffic volume does not spike despite the high number of cruise ships. Similarly, the traffic at Dallas Rd (shown in Figure 3.7) showed even more alignment with the cruise ship numbers but were not always consistent with the cruise ship numbers. Dallas Rd also experienced significant traffic peaks on August 24 – August 26, and September 1, which coincided with the high cruise ship numbers recorded on those days. However, on August 31 , there was a sharp drop in traffic volume, despite the increase in cruise ship numbers.

3.5 Correlation Analysis

Pearson correlation coefficients were employed to assess the relationship between cruise ship numbers, traffic patterns, and air pollutant concentrations. We initially analyzed the correlation between the cruise ship activities and traffic patterns across the different vehicle types counted at Montreal (the counter for Niagara) and Dallas (the counter for Paddon). The results are presented in Figure 3.8. At both locations, there was a notable correlation between the number of ships and traffic volume.

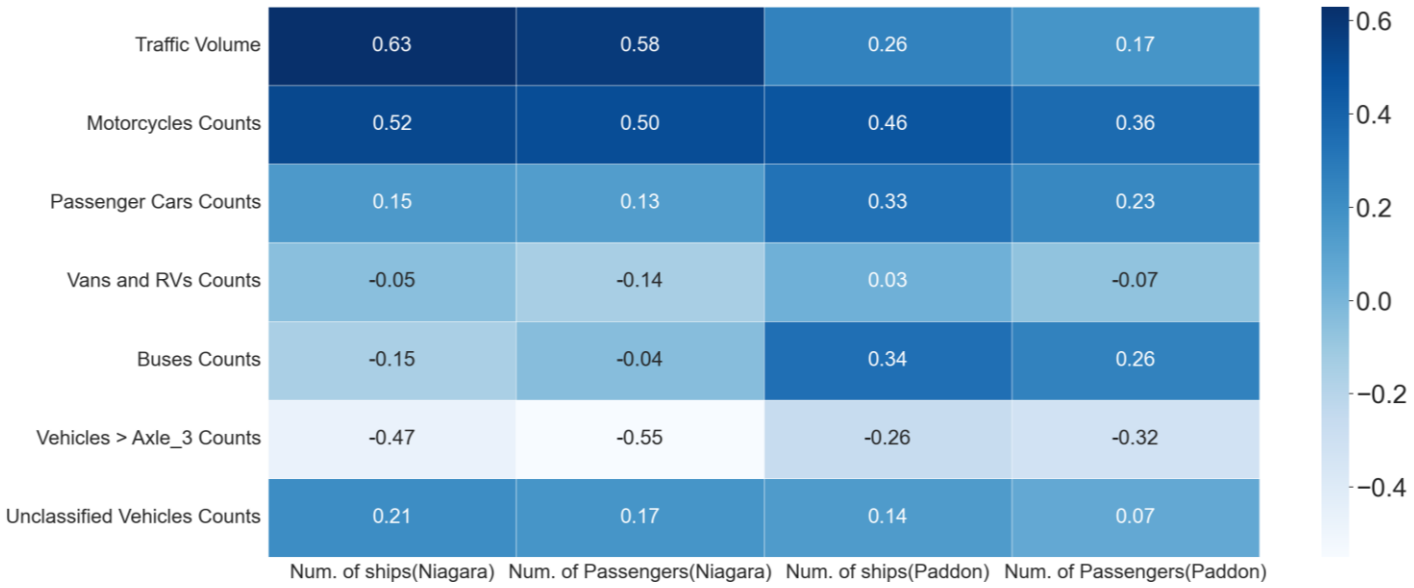


Figure 3.8. Heatmap showing correlation results between cruise ship, traffic volume, and vehicle classes at Niagara and Paddon.

Our analysis indicated that at Montreal (tube counter close to Niagara) , cruise ship activity significantly impacted the traffic volume. However, this strong correlation was primarily influenced by the use of smaller, flexible transportation modes such as motorcycles and unclassified vehicles. These correlations are likely due to tourist's preference for these vehicles for convenience. In addition, given that Montreal Street is very close to the Ogden Point, and it being situated in a residential area, it's logical that tourists and locals will favor the smaller vehicles for short trips, or accessing nearby attractions. With respect to the other vehicle classes, the results indicated a negative correlation with buses, which suggested that bus traffic in minimal on Montreal Street. This is because most buses are likely to use Dallas Road, which is a more direct route from Ogden Point to downtown Victoria and other major destinations. Additionally, larger buses are less unlikely to divert onto Montreal Street than Dallas Road, explaining the weak correlation observed. Interestingly, although the correlation between cruise ship activity and traffic flow at Dallas (counter close to Paddon) were positive, they were less pronounced compared to Montreal. This might be due to Dallas Road's preference as a primary thoroughfare; hence it accommodates diverse traffic which are beyond just cruise-related activities, thereby diluting the direct correlation with cruise ship numbers. However, cars still

played a notable role in passenger transport along this route. Overall, we observed that the correlation at Montreal highlights the strong impact of cruise activity on vehicle types that are not typically used by passengers from cruise ships, which suggests that these passengers prefer modes of transport that offer more flexibility and ease of access, which was contrary to the observations made at Dallas Road. motorcycles and passenger.

Figure 3.9. Healtmap showing correlation results between cruise ship,traffic volume, and vehicle classes at Niagara.

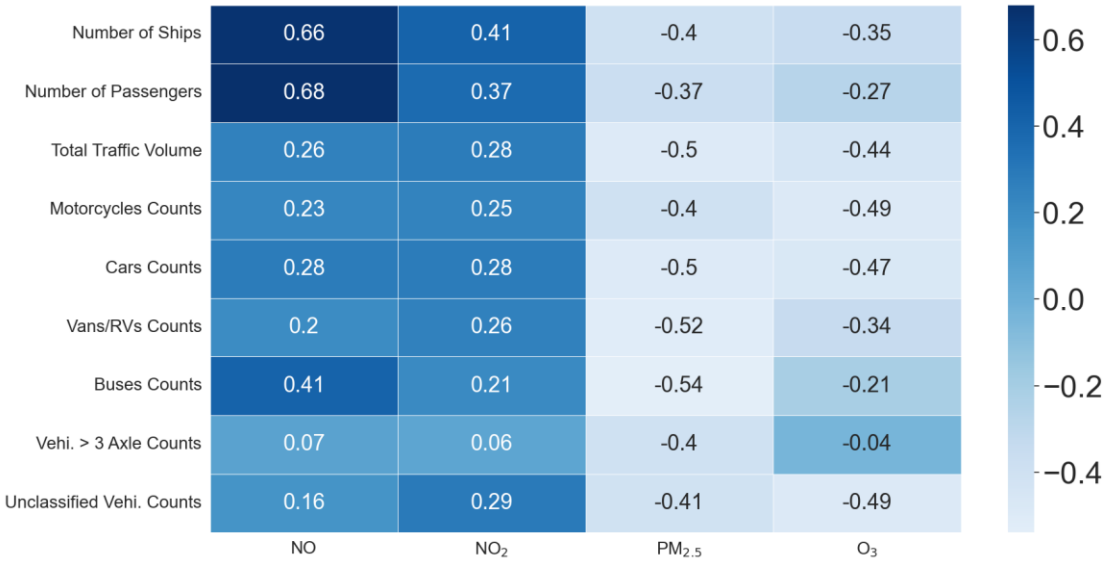


Figure 3.10. Heatmap showing correlation results between cruise ship,traffic volume, and vehicle classes at Paddon.

Figures 3.9 and 3.10 show the heat maps with the correlation results obtained at both locations. At the Niagara location, which is closer to the Ogden Point Terminal, the correlation analysis showed that NO and NO_2 levels were the most highly correlated with cruise ship activity. The positive correlations between NO and NO_2 levels align with the trends of cruise ship and passenger numbers. For instance, on August 24, August 25 and September 01 (shown in Figure 3.3), there were three cruise ships at the port and that led to high levels of NO and NO_2. These spikes suggest that emissions from these vessels contributed directly to the increase in NO and NO_2 levels. In contrast, we observed negative impacts of cruise ship numbers with the O_3 and $PM_{2.5}$ concentrations. On high cruise ship activity days, lower O_3 levels and even lower $PM_{2.5}$ levels were recorded, such as on August 24 and August 25. However, this was not the case for days with no cruise ship activity, such as August 29. This negative correlation was expected for O_3, since O_3 is a secondary air pollutant that forms from the reaction of VOCs and NO_x, and often exhibits trends that are opposite to NO_x concentrations. When we focused the analysis on the Paddon location, which is farther from the terminal, it was interesting to see that the correlation analysis revealed even stronger correlations for NO and NO_2 with cruise ship activity compared to Niagara. These higher spikes coincided with the days where we experienced higher cruise ship numbers as specified for Niagara (Figure 3.3). Our results suggest that despite the distance of this location, the influence of cruise ships was still apparent. This is likely because the air quality monitor at Paddon is directly facing the sea, hence providing a clearer path for emissions from the ships, unlike the monitor at Niagara, which is placed among residential buildings that might obstruct and diffuse the emissions. For $PM_{2.5}$ and O_3, the correlation coefficient was also negative but at a slightly higher state than Niagara. Our finding are consistent with similar studies like Poplawski et al. (2011), who conducted a comprehensive study on air quality in Victoria, BC, and attributed higher levels of NO_2 with the presence of cruise ships due to emissions from ships' engines, which continue running while docked. Furthermore, their study observed that $PM_{2.5}$ levels showed less variation between cruise ships and non-cruise ship days. Anastasopolos et al.(2021), in their study to assess the air quality in Canadian ports, also observed significant reductions in $PM_{2.5}$, and no significant trends in O_3 and NO_2 levels especially at the Victoria port.

Our analysis also revealed substantial differences between the two locations in terms of the traffic correlation with the air pollutant concentration levels. We utilized the counts from the counter placed on Montreal Street for the Niagara location and used the counts placed on Dallas Road for the Paddon correlation analysis. Our analysis showed that the total volume on Montreal Street had weak positive correlations with NO and $PM_{2.5}$ measurements at the Niagara location, which indicates that while the traffic density contributes somewhat to these pollutants, the impact is not substantial. Although NO_2 concentrations were only slightly influenced, these pollutant levels were impacted a bit higher than NO and $PM_{2.5}$ levels. In contrast, O_3 had a weak negative correlation with the traffic volume, which implied that increased traffic volume led to lower O_3 levels. Among the specific vehicle classes, buses and vans/RVs were the most contributors to the pollutant levels at Niagara. Particularly, buses, vans/RVs showed strong positive correlations

with NO_2 and NO which is likely due to their frequent use for transporting cruise passengers. Vans/RVs on the other hand, had a negative correlation with O_3 while motorcycles and passenger cars showed weaker correlations , indicating they had less impact on the air quality at Niagara. In addition, heavy-duty vehicles, such as 3-axle vehicles did not have any significant impact on the NO_2, O_3, and $PM_{2.5}$ levels. These observations aligned with high traffic days, such as August 25 and September 2 (see Figure 3.7).

A different observation was made with the traffic impact at Paddon compared to Niagara. At Paddon, both NO and NO_2 concentrations had stronger positive correlation with traffic volume than Niagara, and $PM_{2.5}$ and O_3 had negative correlations with traffic volume. The higher traffic volume on Dallas Road (where the counter for Paddon is situated), due to its connection to downtown Victoria, signified the impact of traffic on air pollutant concentrations. This correlation highlighted that NO and NO_2 are direct emissions from vehicles, while $PM_{2.5}$ and O_3, being more secondary pollutants, have their presence more dispersed, reducing their impacts. During the higher traffic days in the observation period, specifically on August 25, September 1 and September 02 (see Figure 3.7), there were increased levels of NO, NO_2, $PM_{2.5}$, and O_3 concentrations. We associated these peaks to buses, passenger cars, vans and RVs which were correlated positively with NO and NO_2 but negatively correlated with $PM_{2.5}$ and O_3 levels. These results were in agreement with Rossi et al.(2020), whose study found strong positive correlation between NO, NO_2, and NOx concentrations and suggested that vehicle counts directly led to the higher concentrations of these nitrogen oxides.

3.6 Conclusion

In this study, we aimed to assess the relationship between traffic patterns, cruise ship activity, and air pollutant concentrations. Our findings revealed similarities in correlation patterns between cruise ship activity and traffic data, suggesting shared influences on pollutant concentrations. Additionally, variations in the magnitude of correlation between pollutants indicated the potential differences in emission sources and the varying impacts of cruise ship activity and traffic om specific pollutants. While NOx emissions are associated with both cruise ships and traffic, traffic was identified as the primary source, particularly from buses at Niagara and passenger cars and RVs' vans at Paddon. These vehicles influenced NO and NO_2 concentrations and highlighted the role of local traffic in increasing levels of NOx pollution. Notably, the prominence of buses and cars as significant contributors to pollutant concentrations aligns with expectations, as these transportation modes serve as primary means of conveyance to tourist destinations. The correlation analysis also showed that cruise ship activity influences local traffic patterns, and indirectly affect air quality. This combined impact underscores the need to consider both direct emissions from ships and indirect emissions associated with traffic in strategies considered for air quality management. While the study provided valuable insights, a few limitations should be noted. Firstly, the observation period we used for the correlation analysis was relatively short, covering a few weeks. Secondly, the study did not account for detailed cruise ship data such as exact arrival and departure times, fuel type or ship size , which

all could further explain the cruise ship impacts on air quality. Additionally, the analysis relied on pearson correlation coefficients, which, while useful are not able to capture the complexity of interactions between multiple variables and potential non-linear relationships. Future studies may benefit from extending observation periods for longer durations, especially for traffic data, and using more detailed cruise ship data such as arrival and departure times. Additionally, employing more complex analysis techniques such as machine learning or dispersion models can provide a comprehensive assessment of the relationships between cruise ship activity and air quality. By capturing data before, during, and after the cruise ship seasons, researchers can also investigate temporal variations and identify any seasonal trends or fluctuations in pollutant concentrations.

3.7 Supplementary Information

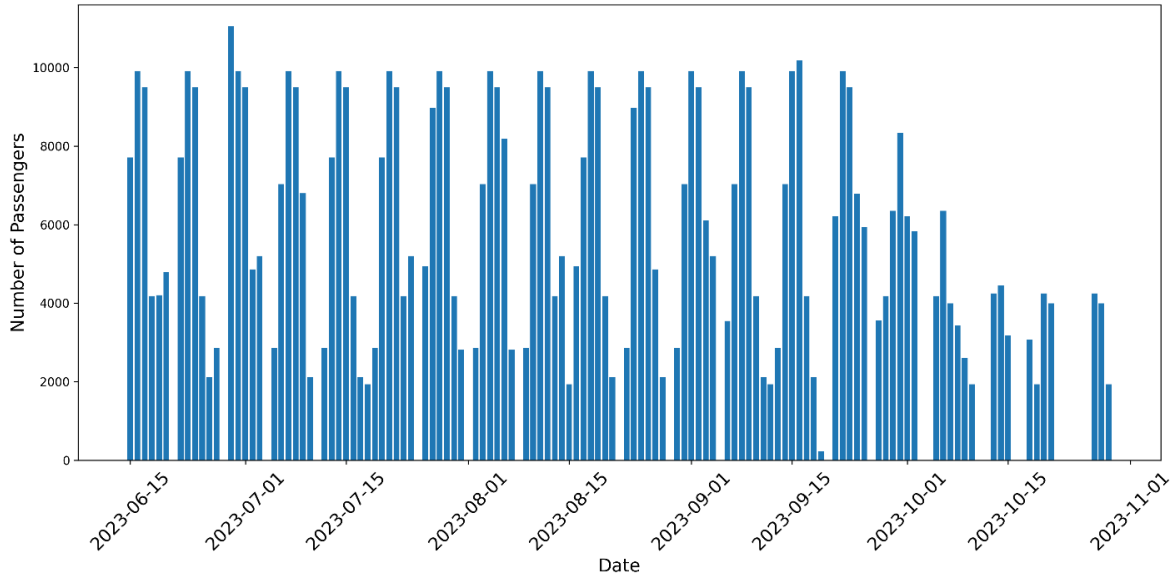


Figure S3.1. Average daily passenger numbers for the whole summer.

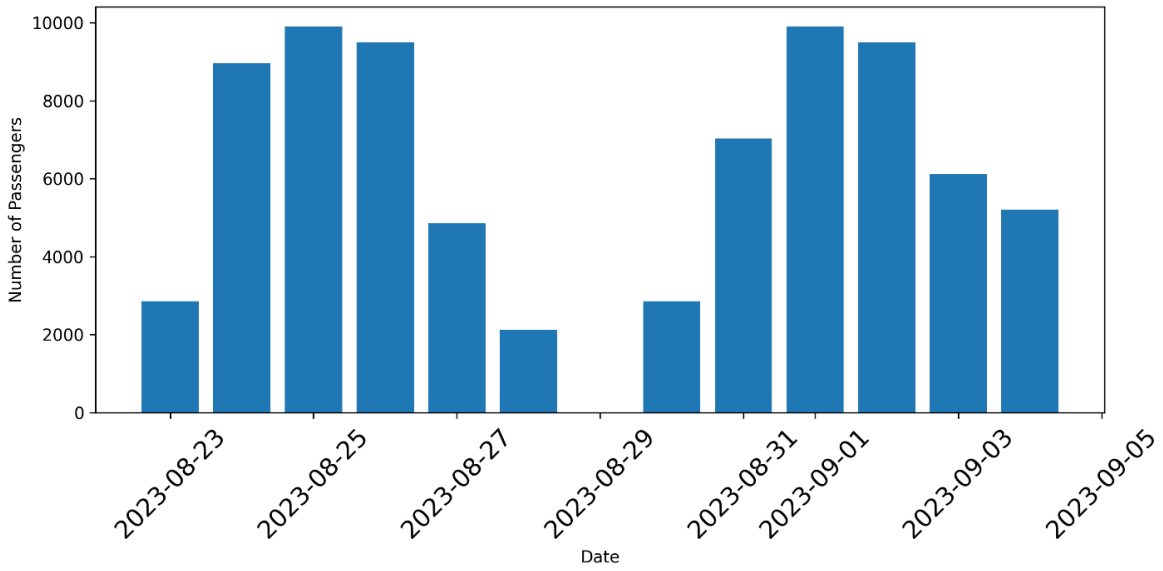


Figure S3.2. Average daily passenger numbers for two weeks concurrent with traffic data.

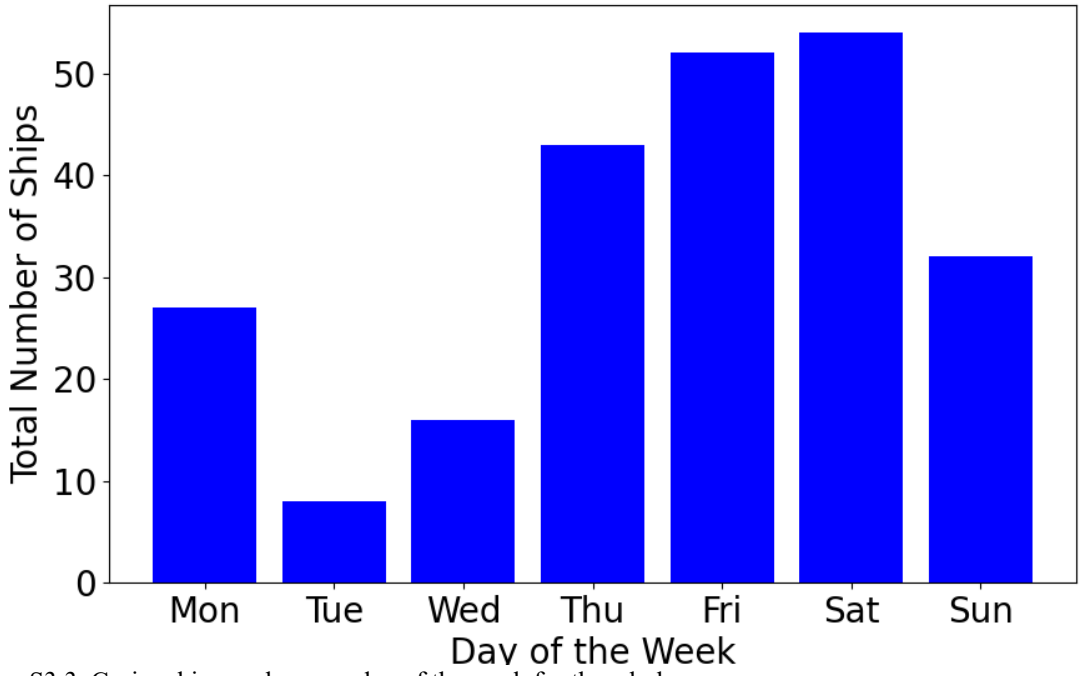


Figure S3.3. Cruise ship numbers per day of the week for the whole summer.

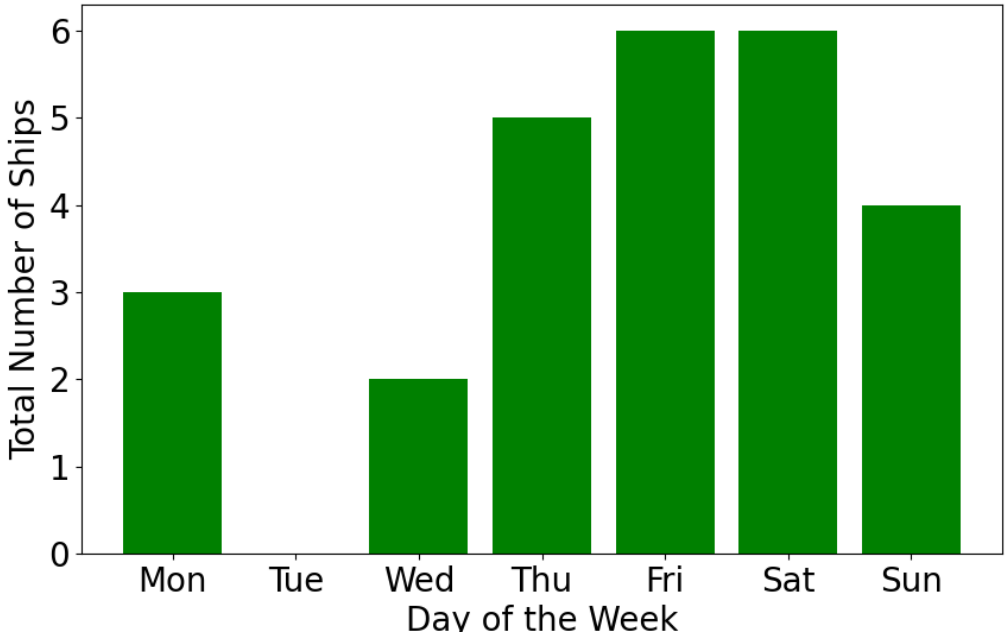


Figure S3.4. Cruise ship numbers per day of the week for two weeks concurrent with traffic data.

Table S3.1. Performance of Calibration Models for sensors deployed at volunteer households

	Niagara			Paddon		
Pollutants	**R^2 [a]**	**RMSE [b]**	**MAE [c]**	**R^2**	**RMSE**	**MAE**
NO (ppb)	0.87	2.60	0.98	0.9	1.74	0.97
NO_2 (ppb)	0.74	3.81	2.27	0.69	3.23	2.11
O_3 (ppb)	0.88	3.33	2.47	0.84	3.09	2.24
$PM_{2.5}$ (µg/m3)	0.69	5.24	3.46	0.60	4.72	2.99

[a]: Coefficient of determination

[b]: Root Mean Square Error

[c]: Mean Absolute Error

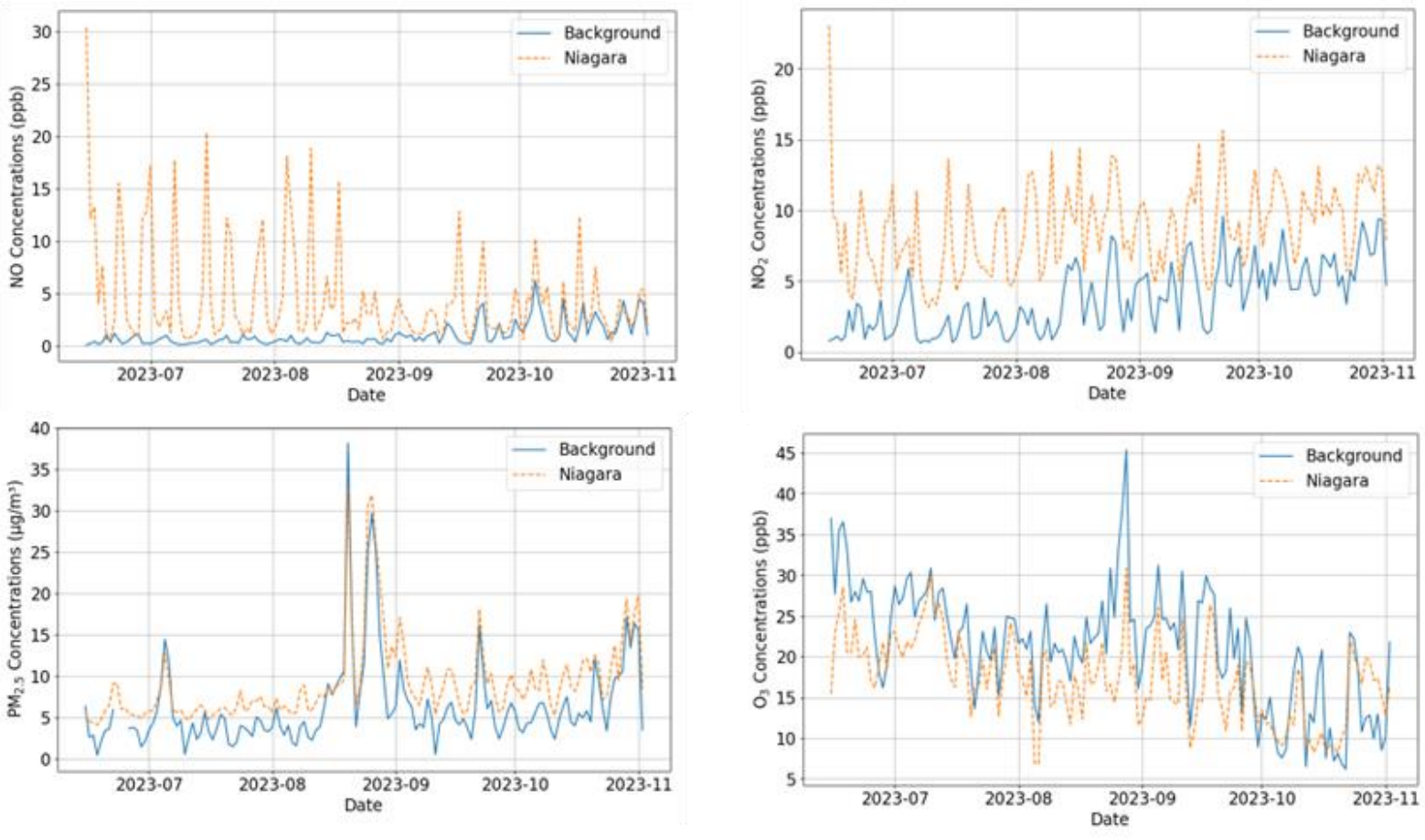


Figure S3.5. Average daily NO, NO_2 , $PM_{2.5}$, and O_3 concentrations measured at Niagara (observed) and Colwood (background) for the whole summer.

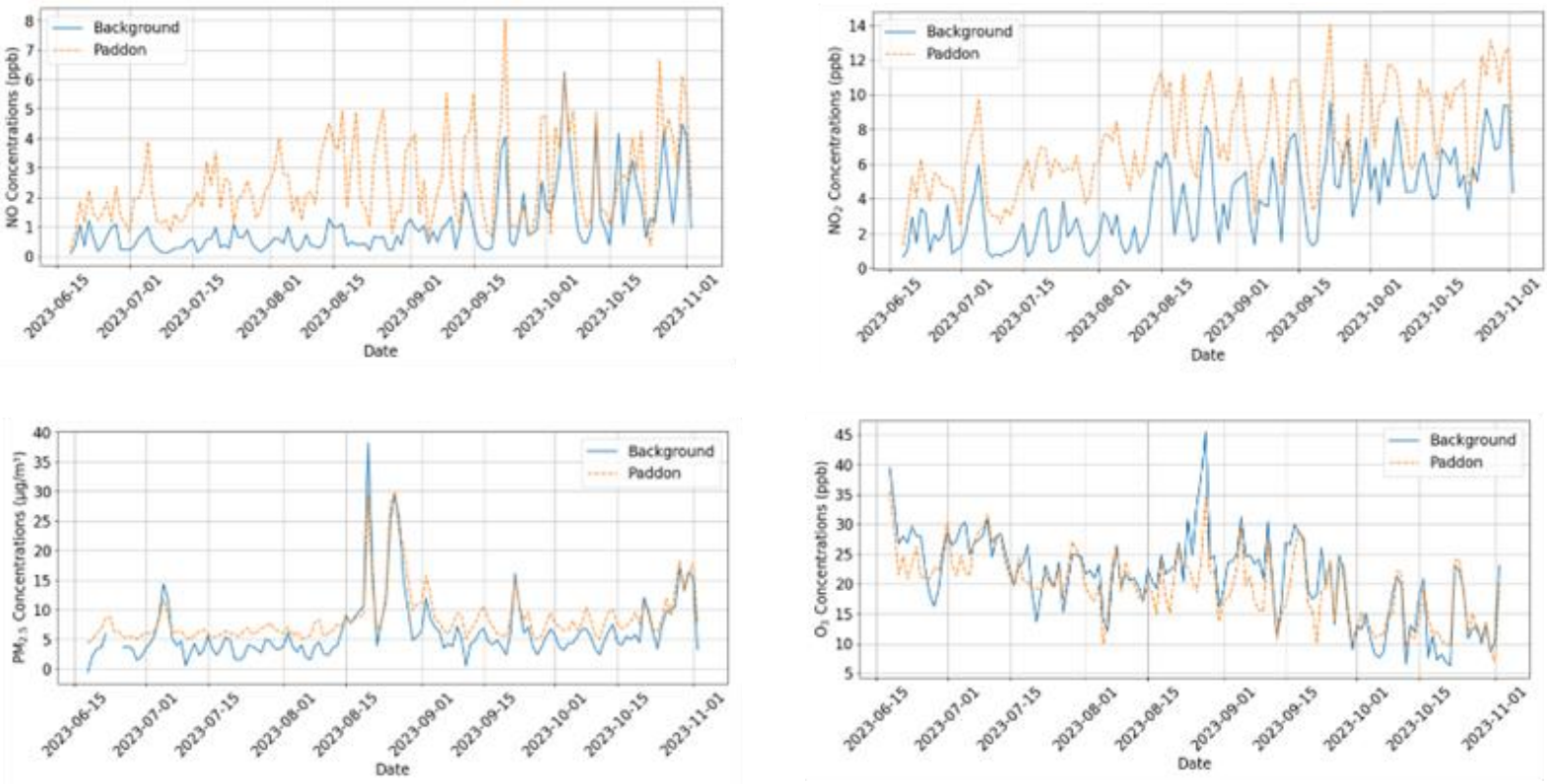


Figure S3.6.Average daily NO, NO_2 , $PM_{2.5}$, and O_3 concentrations measured at Paddon (observed) and Colwood (background) for the whole summer.

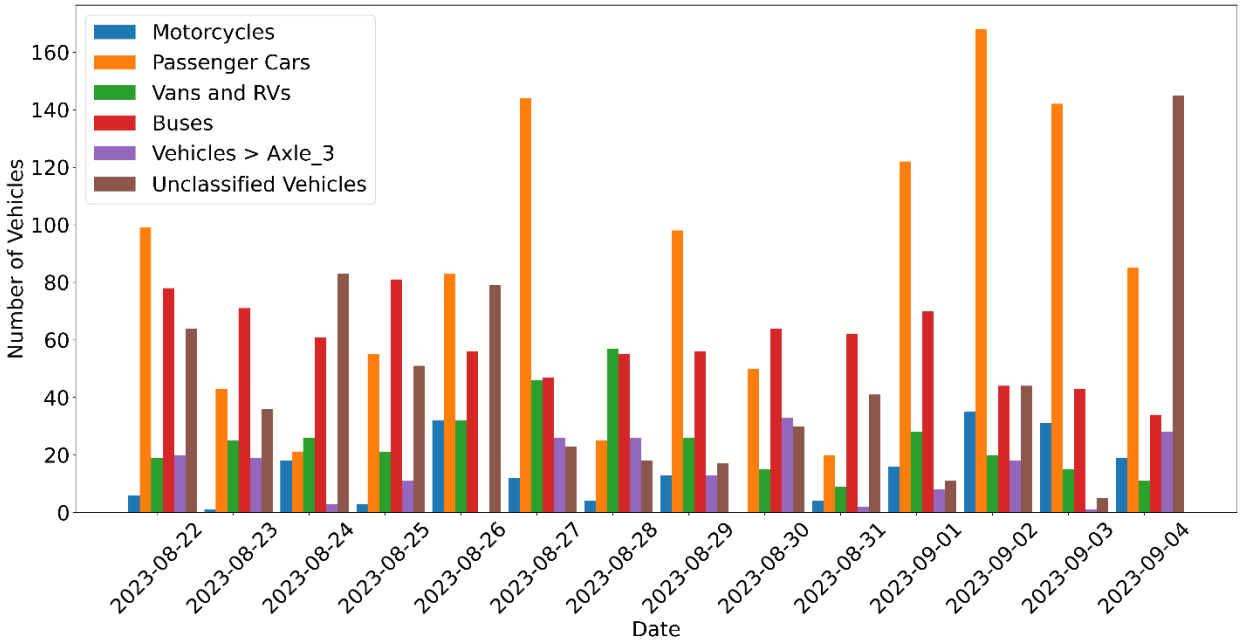


Figure S3.7. Vehicle distribution on Montreal Street.

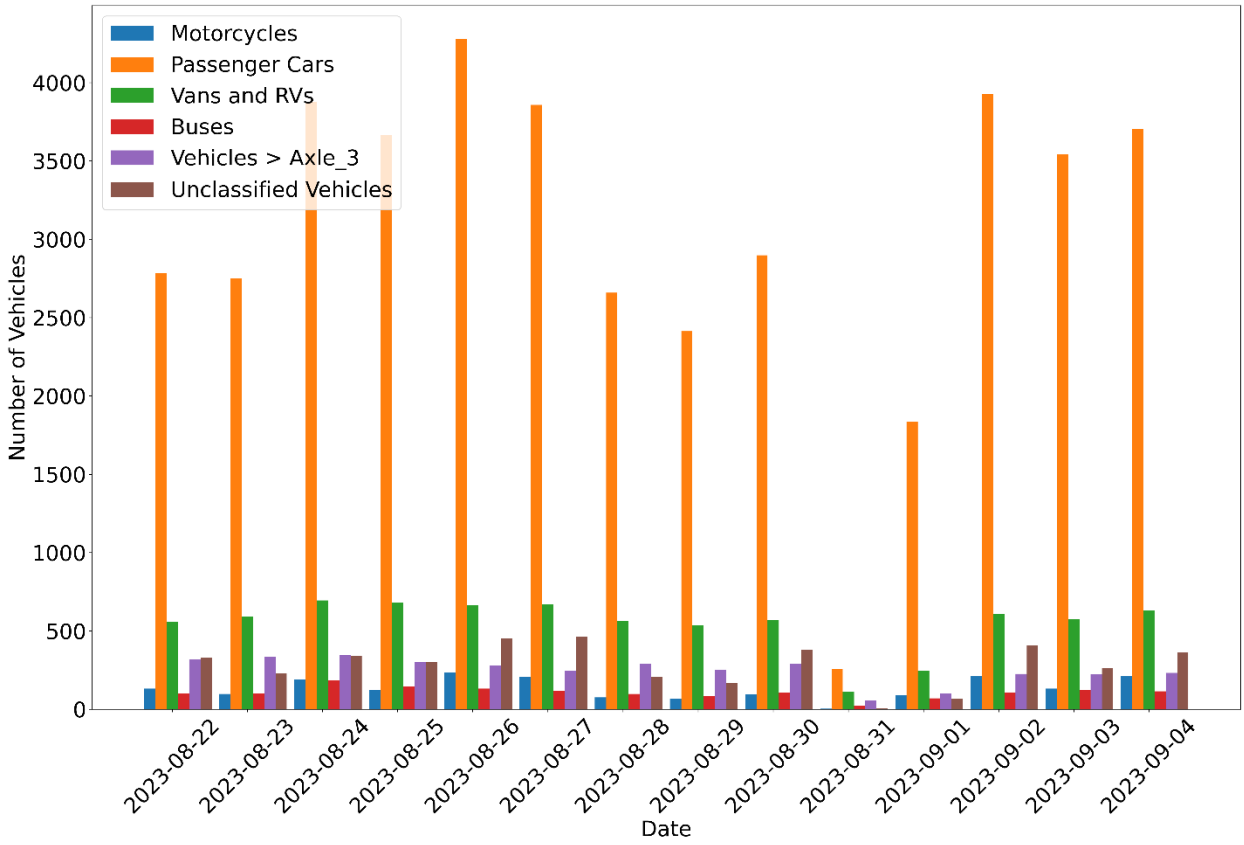


Figure S3.8. Vehicle distribution on Dallas Road.

CHAPTER 4

Summary of Thesis Findings

The research highlighted in this thesis explored innovative, low-cost methods for predicting traffic-related air pollution (TRAP) by leveraging advanced traffic technologies with low-cost sensors and machine learning techniques. Through this research, we were able to address key questions on the effectiveness of these technologies in enhancing real-time air quality predictions and understanding the factors that influence air quality.

By combining video surveillance with YoloV8 object detection model and ByteTracker for multi-class vehicle detection and tracking in **Chapter 2**, we demonstrated the substantial benefits of using these technologies to accurately extract traffic information. These low-cost technologies provided crucial inputs, such as traffic counts and speeds, which, integrated in the MOVES EPA model to estimate vehicle emissions, enabled accurate estimation of pollutants such as $PM_{2.5}$, NO_2, CO, SO_2, NO, and VOCs. The comparison of linear and non-linear models in **Chapter 2** allowed us to emphasize the superior performance of non-linear models compared to linear models, and their essential role in capturing the intricate relationships between pollutants and their predictors. Specifically, the non-linear models explained the features that influence air quality levels in our study area. The combinations of different predictors including traffic volume, traffic speeds, vehicle classes, air pollutant emissions, and meteorological data, provided valuable insights into the key factors. The findings from this study underscored that these features, especially meteorological data, background concentrations and traffic emissions, when incorporated into predictive models, enhances their accuracy.

The research presented in **Chapter 3** explored the secondary influences of traffic, particularly those that stem from the presence of cruise ships, on local air quality in James Bay, Victoria, BC. The study indicated that both emissions from traffic and cruise ship activities play crucial roles in shaping the air quality. The correlation between cruise ship schedules and the increased traffic volumes, particularly from buses and passenger vehicles, showed that cruise ships were indirectly responsible for the local air quality by altering traffic patterns.

In summary, traffic tracking technologies emerged as a pivotal factor in accuracy of the air quality analysis conducted in this thesis. Combining data from low-cost air pollutant sensors with the detailed traffic information from video surveillance and tube counters enables a more context -specific and real-time analysis of the air quality in different areas of Victoria. This thesis underscored the critical importance of integrating various data sources and advanced technologies to improve the estimation of traffic-related air pollution. By leveraging traffic tracking technologies, low-cost sensors, and machine learning models, we can achieve more accurate and temporally resolved air quality estimates. These findings provide valuable insights for policy makers and urban planners and highlight the necessity for targeted interventions that address both direct the indirect emission sources. The holistic approach presented in this research

offers a robust framework for future studies and practical applications aimed at mitigating the adverse effects of traffic-related air pollution.

Although this study made important strides in understanding TRAP, a few limitations were encountered. Firstly, the study focused on a relatively small geographic area with specific traffic patterns, which limited the generalizability of the findings to other urban environments with different traffic and meteorological conditions. The limited temporal resolution of the data also restricted the ability of the study to capture longer-term air quality trends and seasonal variations. Furthermore, the study did not fully account for the complexity of meteorological phenomena, such as wind patterns which could have a significant impact on the pollutant dispersion and concentration. While MOVES EPA model was employed to estimate vehicle emissions, the lack of more precise, direct emission measurements was another limitation. Moreover, detailed information on cruise ships – such as fuel type, and engine conditions was unavailable, which constrained the accuracy of our cruise ship-related findings.

Building on the findings of this research, future work could address some of these limitations while expanding the scope of the research. One promising direction would be to broaden the geographic focus of the study, and apply the methodologies developed in this study to a wider range of cities with varying traffic patterns and air quality challenges. Such studies could reveal the robustness and transferability of this low-cost approach to different urban conditions. In addition, future studies could extend the temporal resolution to include longer periods, capturing seasonal and yearly trends to better understand the long-term dynamics of TRAP.

Another area for future research is improving the granularity of the data, especially regarding the cruise ship activities. Detailed data on cruise ship schedules, engine technologies, fuel types, and sizes would allow for more accurate assessment of their environmental impacts. Furthermore, integrating more comprehensive meteorological data, such as wind speeds and atmospheric pressure could significantly enhance the model's accuracy in predicting pollutant dispersion.

To further improve prediction accuracy, future studies could also explore the use of advanced machine learning models, such as deep learning approaches that can handle large datasets and more complex, non-linear relationships. Additionally, the collaboration with municipalities and environmental agencies could lead to the deployment of these tracking technologies at a city-wide scale, providing policy makers and urban planners with the data needed to make informed decisions about reducing TRAP.

By addressing these limitations and expanding the scope, future research can develop a more robust, flexible, and widely applicable framework for estimating and mitigating traffic-related air pollution.

References

Abouelyazid, Mahmoud. 2023. “Comparative Evaluation of SORT, DeepSORT, and ByteTrack for Multiple Object Tracking in Highway Videos.”

Aburaddaha, Abdalmalek. 2024. “Leveraging Perspective Transformation for Enhanced Pothole Detection in Autonomous Vehicles.”

Afifah, Fatima, Sharmin Nasrin, and Abdul Mukit. 2018. “Vehicle Speed Estimation Using Image Processing.” 48(1).

Ahmed, Hiyam Abobaker Yousif, and Sondos W. A. Mohamed. 2021. “Rainfall Prediction Using Multiple Linear Regressions Model.” Pp. 1–5 in *2020 International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCCEEE)*. Khartoum, Sudan: IEEE.

Alimissis, A., K. Philippopoulos, C. G. Tzanis, and D. Deligiorgi. 2018. “Spatial Estimation of Urban Air Pollution with the Use of Artificial Neural Network Models.” *Atmospheric Environment* 191:205–13. doi: 10.1016/j.atmosenv.2018.07.058.

Allen, J. L., C. Klocke, K. Morris-Schaffer, K. Conrad, M. Sobolewski, and D. A. Cory-Slechta. 2017. “Cognitive Effects of Air Pollution Exposures and Potential Mechanistic Underpinnings.” *Current Environmental Health Reports* 4(2):180–91. doi: 10.1007/s40572-017-0134-3.

Anastasopolos, Angelos T., Uwayemi M. Sofowote, Philip K. Hopke, Mathieu Rouleau, Tim Shin, Aman Dheri, Hui Peng, Ryan Kulka, Mark D. Gibson, Paul-Michel Farah, and Navin Sundar. 2021. “Air Quality in Canadian Port Cities after Regulation of Low-Sulphur Marine Fuel in the North American Emissions Control Area.” *Science of The Total Environment* 791:147949. doi: 10.1016/j.scitotenv.2021.147949.

Anon. 2019. “Sensit Ramp.”

Anon. 2021. “Scentroid CTair.”

Bai, Yun, Yong Li, Xiaoxue Wang, Jingjing Xie, and Chuan Li. 2016. “Air Pollutants Concentrations Forecasting Using Back Propagation Neural Network Based on Wavelet Decomposition with Meteorological Conditions.” *Atmospheric Pollution Research* 7(3):557–66. doi: 10.1016/j.apr.2016.01.004.

Bendle, Johannes. 2012. “PLANNING SENIOR-FRIENDLY NEIGHBOURHOODS: A COMPARATIVE ANALYSIS OF JAMES BAY AND FERNWOOD, VICTORIA, BRITISH COLUMBIA.”

Brantley, H. L., G. S. W. Hagler, E. S. Kimbrough, R. W. Williams, S. Mukerjee, and L. M. Neas. 2014. “Mobile Air Monitoring Data-Processing Strategies and Effects on Spatial Air Pollution Trends.” *Atmospheric Measurement Techniques* 7(7):2169–83. doi: 10.5194/amt-7-2169-2014.

Calatayud, Agustina, Santiago Sánchez González, and Jose Maria Marquez. 2022. “Using Big Data to Estimate the Impact of Cruise Activity on Congestion in Port Cities.” *Maritime Economics & Logistics* 24(3):566–83. doi: 10.1057/s41278-021-00198-3.

Castell, Nuria, Franck R. Dauge, Philipp Schneider, Matthias Vogt, Uri Lerner, Barak Fishbain, David Broday, and Alena Bartonova. 2017. “Can Commercial Low-Cost Sensor Platforms Contribute to Air Quality Monitoring and Exposure Estimates?” *Environment International* 99:293–302. doi: 10.1016/j.envint.2016.12.007.

Chen, Chen, Bin Liu, Shaohua Wan, Peng Qiao, and Qingqi Pei. 2021. “An Edge Traffic Flow Detection Scheme Based on Deep Learning in an Intelligent Transportation System.” *IEEE Transactions on Intelligent Transportation Systems* 22(3):1840–52. doi: 10.1109/TITS.2020.3025687.

Chen, Gongbo, Shanshan Li, Luke D. Knibbs, N. A. S. Hamm, Wei Cao, Tiantian Li, Jianping Guo, Hongyan Ren, Michael J. Abramson, and Yuming Guo. 2018. “A Machine Learning Method to Estimate PM2.5 Concentrations across China with Remote Sensing, Meteorological and Land Use Information.” *Science of The Total Environment* 636:52–60. doi: 10.1016/j.scitotenv.2018.04.251.

Chen, Tianqi, and Carlos Guestrin. 2016. “XGBoost: A Scalable Tree Boosting System.” Pp. 785–94 in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. San Francisco California USA: ACM.

Chen, Yuanyuan, Runhe Shi, Shijie Shu, and Wei Gao. 2013. “Ensemble and Enhanced PM10 Concentration Forecast Model Based on Stepwise Regression and Wavelet Analysis.” *Atmospheric Environment* 74:346–59. doi: 10.1016/j.atmosenv.2013.04.002.

Dwyer, B., Nelson, J., and Solawetz, J. 2022. “Roboflow (Version 1.0).”

Eckhardt, S., O. Hermansen, H. Grythe, M. Fiebig, K. Stebel, M. Cassiani, A. Baecklund, and A. Stohl. 2013. “The Influence of Cruise Ship Emissions on Air Pollution in Svalbard – a Harbinger of a More Polluted Arctic?” *Atmospheric Chemistry and Physics* 13(16):8401–9. doi: 10.5194/acp-13-8401-2013.

Elliot, Alex J., Sue Smith, Alec Dobney, John Thornes, Gillian E. Smith, and Sotiris Vardoulakis. 2016. “Monitoring the Effect of Air Pollution Episodes on Health Care Consultations and Ambulance Call-Outs in England during March/April 2014: A Retrospective Observational Analysis.” *Environmental Pollution* 214:903–11. doi: 10.1016/j.envpol.2016.04.026.

Environment and Climate Change Canada. 2022. *National Air Pollution Surveillance (NAPS) Program*.

Fedorov, Aleksandr, Kseniia Nikolskaia, Sergey Ivanov, Vladimir Shepelev, and Alexey Minbaleev. 2019. “Traffic Flow Estimation with Data from a Video Surveillance Camera.” *Journal of Big Data* 6(1):73. doi: 10.1186/s40537-019-0234-z.

Fei, Yu-Hsuan, Ta-Chih Hsiao, and Albert Y. Chen. 2024. “Adapting Public Annotated Data Sets and Low-Quality Dash Cameras for Spatiotemporal Estimation of Traffic-Related Air Pollution: A Transfer-Learning Approach.” *Journal of Computing in Civil Engineering* 38(3):04024006. doi: 10.1061/JCCEE5.CPENG-5667.

Feinberg, Stephen, Ron Williams, Gayle S. W. Hagler, Joshua Rickard, Ryan Brown, Daniel Garver, Greg Harshfield, Phillip Stauffer, Erick Mattson, Robert Judge, and Sam Garvey. 2018. “Long-Term Evaluation of Air Sensor Technology under Ambient Conditions in Denver, Colorado.”

Feng, Xiao, Qi Li, Yajie Zhu, Junxiong Hou, Lingyan Jin, and Jingjie Wang. 2015. “Artificial Neural Networks Forecasting of PM2.5 Pollution Using Air Mass Trajectory Based Geographic Model and Wavelet Transformation.” *Atmospheric Environment* 107:118–28. doi: 10.1016/j.atmosenv.2015.02.030.

Forehead, H., and N. Huynh. 2018. “Review of Modelling Air Pollution from Traffic at Street-Level - The State of the Science.” *Environmental Pollution* 241:775–86. doi: 10.1016/j.envpol.2018.06.019.

Frosina, Emma, Luca Romagnuolo, Antonella Bonavolontà, Assunta Andreozzi, Adolfo Senatore, Francesco Fortunato, and Pino Giliberti. 2018. “Evaporative Emissions in a Fuel Tank of Vehicles: Numerical and Experimental Approaches.” *Energy Procedia* 148:1167–74. doi: 10.1016/j.egypro.2018.08.025.

García-González, Jorge, Miguel A. Molina-Cabello, Rafael M. Luque-Baena, Juan M. Ortiz-de-Lazcano-Lobato, and Ezequiel López-Rubio. 2021. “Road Pollution Estimation from Vehicle Tracking in Surveillance Videos by Deep Convolutional Neural Networks.” *Applied Soft Computing* 113:107950. doi: 10.1016/j.asoc.2021.107950.

Guarnieri, Michael, and John R. Balmes. 2014. “Outdoor Air Pollution and Asthma.” *The Lancet* 383(9928):1581–92. doi: 10.1016/S0140-6736(14)60617-6.

Guo, Junmei, Haitong Lou, Haonan Chen, Haiying Liu, Jason Gu, Lingyun Bi, and Xuehu Duan. 2023. “A New Detection Algorithm for Alien Intrusion on Highway.” *Scientific Reports* 13(1):10667. doi: 10.1038/s41598-023-37686-w.

Han, Xiao, Yuqin Liu, Hong Gao, Jianmin Ma, Xiaoxuan Mao, Yuting Wang, and Xudong Ma. 2017. “Forecasting PM2.5 Induced Male Lung Cancer Morbidity in China Using Satellite Retrieved PM2.5 and Spatial Analysis.” *Science of The Total Environment* 607–608:1009–17. doi: 10.1016/j.scitotenv.2017.07.061.

Health Canada. 2021. *Health Impacts of Air Pollution in Canada: Estimates of Morbidity and Premature Mortality Outcomes – 2021 Report*.

Ho, Yu-Hsuan, Ta-Chih Hsiao, and Albert Y. Chen. 2023. “Emission Analysis of Electric Motorcycles and Assessment of Emission Reduction With Fleet Electrification.” *IEEE Transactions on Intelligent Transportation Systems* 24(12):15369–78. doi: 10.1109/TITS.2023.3272385.

Hong, Kris Y., Pedro O. Pinheiro, and Scott Weichenthal. 2020. “Predicting Outdoor Ultrafine Particle Number Concentrations, Particle Size, and Noise Using Street-Level Images and Audio Data.” *Environment International* 144:106044. doi: 10.1016/j.envint.2020.106044.

Hoogkamer, Lauren Perez. 2013. “ASSESSING AND MANAGING CRUISE SHIP TOURISM IN HISTORIC PORT CITIES: CASE STUDY CHARLESTON, SOUTH CAROLINA.”

Hoshyaripour, G., G. Brasseur, M. F. Andrade, M. Gavidia-Calderón, I. Bouarar, and R. Y. Ynoue. 2016. “Prediction of Ground-Level Ozone Concentration in São Paulo, Brazil: Deterministic versus Statistic Models.” *Atmospheric Environment* 145:365–75. doi: 10.1016/j.atmosenv.2016.09.061.

ICBC. 2020. “Vehicle Polpulation.”

International Transport Forum. 2016. “Cruise Shipping and Urban Development: The Case of Venice.”

Jian, Le, Yun Zhao, Yi-Ping Zhu, Mei-Bian Zhang, and Dean Bertolatti. 2012. “An Application of ARIMA Model to Predict Submicron Particle Concentrations from Meteorological Factors at a Busy Roadside in Hangzhou, China.” *Science of The Total Environment* 426:336–45. doi: 10.1016/j.scitotenv.2012.03.025.

Kamińska, Joanna A. 2018. “The Use of Random Forests in Modelling Short-Term Air Pollution Effects Based on Traffic and Meteorological Conditions: A Case Study in Wrocław.” *Journal of Environmental Management* 217:164–74. doi: 10.1016/j.jenvman.2018.03.094.

Karagulian, Federico, Maurizio Barbiere, Alexander Kotsev, Laurent Spinelle, Michel Gerboles, Friedrich Lagler, Nathalie Redon, Sabine Crunaire, and Annette Borowiak. 2019. “Review of the Performance of Low-Cost Sensors for Air Quality Monitoring.” *Atmosphere* 10(9):506. doi: 10.3390/atmos10090506.

Kelly, K. E., J. Whitaker, A. Petty, C. Widmer, A. Dybwad, D. Sleeth, R. Martin, and A. Butterfield. 2017. “Ambient and Laboratory Evaluation of a Low-Cost Particulate Matter Sensor.” *Environmental Pollution* 221:491–500. doi: 10.1016/j.envpol.2016.12.039.

Kennedy, Ryan David. 2019. “An Investigation of Air Pollution on the Decks of 4 Cruise Ships.”

Khan, Ishtiaq Rasool, Syed Talha Abid Ali, Asif Siddiq, Muhammad Murtaza Khan, Muhammad Usman Ilyas, Saleh Alshomrani, and Susanto Rahardja. 2022. “Automatic License Plate Recognition in Real-World Traffic Videos Captured in Unconstrained Environment by a Mobile Camera.” *Electronics* 11(9):1408. doi: 10.3390/electronics11091408.

Khan, Jibran, Matthias Ketzel, Konstantinos Kakosimos, Mette Sørensen, and Steen Solvang Jensen. 2018. “Road Traffic Air and Noise Pollution Exposure Assessment – A Review of Tools and Techniques.” *Science of The Total Environment* 634:661–76. doi: 10.1016/j.scitotenv.2018.03.374.

Kim, HyungJun. 2022. “Multiple Vehicle Tracking and Classification System with a Convolutional Neural Network.” *Journal of Ambient Intelligence and Humanized Computing* 13(3):1603–14. doi: 10.1007/s12652-019-01429-5.

Kumar, Prashant, Lidia Morawska, Claudio Martani, George Biskos, Marina Neophytou, Silvana Di Sabatino, Margaret Bell, Leslie Norford, and Rex Britter. 2015. “The Rise of Low-Cost Sensing for Managing Air Pollution in Cities.” *Environment International* 75:199–205. doi: 10.1016/j.envint.2014.11.019.

Larson, Timothy, Timothy Gould, Erin A. Riley, Elena Austin, Jonathan Fintzi, Lianne Sheppard, Michael Yost, and Christopher Simpson. 2017. “Ambient Air Quality Measurements from a Continuously Moving Mobile Platform: Estimation of Area-Wide, Fuel-Based, Mobile Source Emission Factors Using Absolute Principal Component Scores.” *Atmospheric Environment* 152:201–11. doi: 10.1016/j.atmosenv.2016.12.037.

Le Hong, Zoe, and Naomi Zimmerman. 2021. “Air Quality and Greenhouse Gas Implications of Autonomous Vehicles in Vancouver, Canada.” *Transportation Research Part D: Transport and Environment* 90:102676. doi: 10.1016/j.trd.2020.102676.

Lei, Jiashuo, Chao Yang, Qingyan Fu, Yuan Chao, Jie Dai, and Quan Yuan. 2024. “An Approach of Localizing MOVES to Estimate Emission Factors of Trucks.” *International Journal of Transportation Science and Technology* 13:229–42. doi: 10.1016/j.ijtst.2023.02.002.

Li, Hugh Z., Peishi Gu, Qing Ye, Naomi Zimmerman, Ellis S. Robinson, R. Subramanian, Joshua S. Apte, Allen L. Robinson, and Albert A. Presto. 2019. “Spatially Dense Air Pollutant Sampling: Implications of Spatial Variability on the Representativeness of Stationary Air Pollutant Monitors.” *Atmospheric Environment: X* 2:100012. doi: 10.1016/j.aeaoa.2019.100012.

Li, Jiangtao, Xingqin An, Qingyong Li, Chao Wang, Haomin Yu, Xinyuan Zhou, and Yangli-ao Geng. 2022. “Application of XGBoost Algorithm in the Optimization of Pollutant Concentration.” *Atmospheric Research* 276:106238. doi: 10.1016/j.atmosres.2022.106238.

Li, Xiaoli, Aorong Luo, Jiangeng Li, and Yang Li. 2019. “Air Pollutant Concentration Forecast Based on Support Vector Regression and Quantum-Behaved Particle Swarm Optimization.” *Environmental Modeling & Assessment* 24(2):205–22. doi: 10.1007/s10666-018-9633-3.

Liang, Lu. 2021. “Calibrating Low-Cost Sensors for Ambient Air Monitoring: Techniques, Trends, and Challenges.” *Environmental Research* 197:111163. doi: 10.1016/j.envres.2021.111163.

Liaw, Andy, and Matthew Wiener. 2002. “Classification and Regression by randomForest.” 2.

Liu, Huixiang, Qing Li, Dongbing Yu, and Yu Gu. 2019. “Air Quality Index and Air Pollutant Concentration Prediction Based on Machine Learning Algorithms.” *Applied Sciences* 9(19):4069. doi: 10.3390/app9194069.

Liu, Ying, Guofeng Cao, Naizhuo Zhao, Kevin Mulligan, and Xinyue Ye. 2018. “Improve Ground-Level PM2.5 Concentration Mapping Using a Random Forests-Based Geostatistical Approach.” *Environmental Pollution* 235:272–82. doi: 10.1016/j.envpol.2017.12.070.

Liveable Cities Smartlinx. 2023. “Smartlinx.”

Lloret, Josep, Arnau Carreño, Hrvoje Carić, Joan San, and Lora E. Fleming. 2021a. “Environmental and Human Health Impacts of Cruise Tourism: A Review.” *Marine Pollution Bulletin* 173:112979. doi: 10.1016/j.marpolbul.2021.112979.

Lloret, Josep, Arnau Carreño, Hrvoje Carić, Joan San, and Lora E. Fleming. 2021b. “Environmental and Human Health Impacts of Cruise Tourism: A Review.” *Marine Pollution Bulletin* 173:112979. doi: 10.1016/j.marpolbul.2021.112979.

Lundberg, Scott M., and Su-In Lee. 2017. “A Unified Approach to Interpreting Model Predictions.”

Ma, Jinghui, Zhongqi Yu, Yuanhao Qu, Jianming Xu, and Yu Cao. 2020. “Application of the XGBoost Machine Learning Method in PM2.5 Prediction: A Case Study of Shanghai.” *Aerosol and Air Quality Research* 20(1):128–38. doi: 10.4209/aaqr.2019.08.0408.

Martín-Baos, José Ángel, Luis Rodriguez-Benitez, Ricardo García-Ródenas, and Jun Liu. 2022. “IoT Based Monitoring of Air Quality and Traffic Using Regression Analysis.” *Applied Soft Computing* 115:108282. doi: 10.1016/j.asoc.2021.108282.

Masmoudi, Sahar, Haytham Elghazel, Dalila Taieb, Orhan Yazar, and Amjad Kallel. 2020. “A Machine-Learning Framework for Predicting Multiple Air Pollutants’ Concentrations via Multi-Target Regression and Feature Selection.” *Science of The Total Environment* 715:136991. doi: 10.1016/j.scitotenv.2020.136991.

McCarthy, John. 2018. “Responsible Cruise Tourism and Regeneration: The Case of Nanaimo, British Columbia, Canada.” *International Planning Studies* 23(3):225–38. doi: 10.1080/13563475.2018.1428539.

Mead, M. I., O. A. M. Popoola, G. B. Stewart, P. Landshoff, M. Calleja, M. Hayes, J. J. Baldovi, M. W. McLeod, T. F. Hodgson, J. Dicks, A. Lewis, J. Cohen, R. Baron, J. R. Saffell, and R. L. Jones. 2013. “The Use of Electrochemical Sensors for Monitoring Urban Air Quality in Low-Cost, High-Density Networks.” *Atmospheric Environment* 70:186–203. doi: 10.1016/j.atmosenv.2012.11.060.

Miovision Technologies Inc. 2023. “Miovision Traffic Camera.”

Moazami, Saber, Roohollah Noori, Bahman Jabbarian Amiri, Bijan Yeganeh, Sadegh Partani, and Salman Safavi. 2016. “Reliable Prediction of Carbon Monoxide Using Developed Support Vector Machine.” *Atmospheric Pollution Research* 7(3):412–18. doi: 10.1016/j.apr.2015.10.022.

Mölders, Nicole, Scott Gende, and Michael Pirhalla. 2013. “Assessment of Cruise–Ship Activity Influences on Emissions, Air Quality, and Visibility in Glacier Bay National Park.” *Atmospheric Pollution Research* 4(4):435–45. doi: 10.5094/APR.2013.050.

Mueller, Stephen F., and Jonathan W. Mallard. 2011. “Contributions of Natural Emissions to Ozone and PM $_{2.5}$ as Simulated by the Community Multiscale Air Quality (CMAQ) Model.” *Environmental Science & Technology* 45(11):4817–23. doi: 10.1021/es103645m.

Murena, F., L. Mocerino, F. Quaranta, and D. Toscano. 2018. “Impact on Air Quality of Cruise Ship Emissions in Naples, Italy.” *Atmospheric Environment* 187:70–83. doi: 10.1016/j.atmosenv.2018.05.056.

Mustakim, Nurul A’isyah, Ahmad Zia Ul-Saufie, Wan Nur Shaziayani, Norazian Mohamad Noor, and Sofianita Mutalib. 2022. “Prediction of Daily Air Pollutants Concentration and Air Pollutant Index Using Machine Learning Approach.” *Pertanika Journal of Science and Technology* 31(1):123–35. doi: 10.47836/pjst.31.1.08.

Ng, Jin Jie, Kah Ong Michael Goh, and Connie Tee. 2023. “Traffic Impact Assessment System Using Yolov5 and ByteTrack.” *Journal of Informatics and Web Engineering* 2(2):168–88. doi: 10.33093/jiwe.2023.2.2.13.

Ohlwein, Simone, Ron Kappeler, Meltem Kutlar Joss, Nino Künzli, and Barbara Hoffmann. 2019. “Health Effects of Ultrafine Particles: A Systematic Literature Review Update of Epidemiological Evidence.” *International Journal of Public Health* 64(4):547–59. doi: 10.1007/s00038-019-01202-7.

Otero, Noelia, Henning W. Rust, and Tim Butler. 2021. “Temperature Dependence of Tropospheric Ozone under NOx Reductions over Germany.” *Atmospheric Environment* 253:118334. doi: 10.1016/j.atmosenv.2021.118334.

Ouma, Yashon O., Clinton O. Okuku, and Evalyne N. Njau. 2020. “Use of Artificial Neural Networks and Multiple Linear Regression Model for the Prediction of Dissolved Oxygen in Rivers: Case Study of Hydrographic Basin of River Nyando, Kenya.” *Complexity* 2020:1–23. doi: 10.1155/2020/9570789.

Pan, Bingyue. 2018. “Application of XGBoost Algorithm in Hourly PM2.5 Concentration Prediction.” *IOP Conference Series: Earth and Environmental Science* 113:012127. doi: 10.1088/1755-1315/113/1/012127.

Pan, Long, Enjian Yao, and Yang Yang. 2016. “Impact Analysis of Traffic-Related Air Pollution Based on Real-Time Traffic and Basic Meteorological Information.” *Journal of Environmental Management* 183:510–20. doi: 10.1016/j.jenvman.2016.09.010.

Perdiguero, Jordi, and Alex Sanz. 2020. “Cruise Activity and Pollution: The Case of Barcelona.” *Transportation Research Part D: Transport and Environment* 78:102181. doi: 10.1016/j.trd.2019.11.010.

Poplawski, Karla, Eleanor Setton, Bryan McEwen, Dan Hrebenyk, Mark Graham, and Peter Keller. 2011. “Impact of Cruise Ship Emissions in Victoria, BC, Canada.” *Atmospheric Environment* 45(4):824–33. doi: 10.1016/j.atmosenv.2010.11.029.

Powers, Jordan G., Joseph B. Klemp, William C. Skamarock, Christopher A. Davis, Jimy Dudhia, David O. Gill, Janice L. Coen, David J. Gochis, Ravan Ahmadov, Steven E. Peckham, Georg A. Grell, John Michalakes, Samuel Trahan, Stanley G. Benjamin, Curtis R. Alexander, Geoffrey J. Dimego, Wei Wang, Craig S. Schwartz, Glen S. Romine, Zhiquan Liu, Chris Snyder, Fei Chen, Michael J. Barlage, Wei Yu, and Michael G. Duda. 2017. “The Weather Research and Forecasting Model: Overview, System Efforts, and Future Directions.” *Bulletin of the American Meteorological Society* 98(8):1717–37. doi: 10.1175/BAMS-D-15-00308.1.

Ren, Xiang, Zhongyuan Mi, and Panos G. Georgopoulos. 2020. “Comparison of Machine Learning and Land Use Regression for Fine Scale Spatiotemporal Estimation of Ambient Air Pollution: Modeling Ozone Concentrations across the Contiguous United States.” *Environment International* 142:105827. doi: 10.1016/j.envint.2020.105827.

Rossi, Riccardo, Riccardo Ceccato, and Massimiliano Gastaldi. 2020. “Effect of Road Traffic on Air Pollution. Experimental Evidence from COVID-19 Lockdown.” *Sustainability* 12(21):8984. doi: 10.3390/su12218984.

Sanchez, Kristen A., Margaret Foster, Mark J. Nieuwenhuijsen, Anthony D. May, Tara Ramani, Joe Zietsman, and Haneen Khreis. 2020. “Urban Policy Interventions to Reduce Traffic Emissions and Traffic-Related Air Pollution: Protocol for a Systematic Evidence Map.” *Environment International* 142:105826. doi: 10.1016/j.envint.2020.105826.

Santiago, J. L., R. Borge, B. Sanchez, C. Quaassdorff, D. De La Paz, A. Martilli, E. Rivas, and F. Martín. 2021. “Estimates of Pedestrian Exposure to Atmospheric Pollution Using High-Resolution Modelling in a Real Traffic Hot-Spot.” *Science of The Total Environment* 755:142475. doi: 10.1016/j.scitotenv.2020.142475.

Santiago, J. L., and A. Martilli. 2010. “A Dynamic Urban Canopy Parameterization for Mesoscale Models Based on Computational Fluid Dynamics Reynolds-Averaged Navier–Stokes Microscale Simulations.” *Boundary-Layer Meteorology* 137(3):417–39. doi: 10.1007/s10546-010-9538-4.

Shairsingh, Kerolyn K., Cheol-Heon Jeong, Jonathan M. Wang, Jeffrey R. Brook, and Greg J. Evans. 2019. “Urban Land Use Regression Models: Can Temporal Deconvolution of Traffic Pollution Measurements Extend the Urban LUR to Suburban Areas?” *Atmospheric Environment* 196:143–51. doi: 10.1016/j.atmosenv.2018.10.013.

Shams, Seyedeh Reyhaneh, Ali Jahani, Saba Kalantary, Mazaher Moeinaddini, and Nematollah Khorasani. 2021. “The Evaluation on Artificial Neural Networks (ANN) and Multiple Linear Regressions (MLR) Models for Predicting SO2 Concentration.” *Urban Climate* 37:100837. doi: 10.1016/j.uclim.2021.100837.

Shams, Seyedeh Reyhaneh, Ali Jahani, Mazaher Moeinaddini, and Nematollah Khorasani. 2020. "Air Carbon Monoxide Forecasting Using an Artificial Neural Network in Comparison with Multiple Regression." *Modeling Earth Systems and Environment* 6(3):1467–75. doi: 10.1007/s40808-020-00762-5.

Shorten, Connor, Taghi M. Khoshgoftaar, and Borko Furht. 2021. "Text Data Augmentation for Deep Learning." *Journal of Big Data* 8(1):101. doi: 10.1186/s40537-021-00492-0.

Si, Minxing, Ying Xiong, Shan Du, and Ke Du. 2020. "Evaluation and Calibration of a Low-Cost Particle Sensor in Ambient Conditions Using Machine-Learning Methods." *Atmospheric Measurement Techniques* 13(4):1693–1707. doi: 10.5194/amt-13-1693-2020.

Snyder, Emily G., Timothy H. Watkins, Paul A. Solomon, Eben D. Thoma, Ronald W. Williams, Gayle S. W. Hagler, David Shelow, David A. Hindin, Vasu J. Kilaru, and Peter W. Preuss. 2013. "The Changing Paradigm of Air Pollution Monitoring." *Environmental Science & Technology* 47(20):11369–77. doi: 10.1021/es4022602.

Song, Jinchao, Chunli Zhao, Tao Lin, Xinhu Li, and Alexander V. Prishchepov. 2019. "Spatio-Temporal Patterns of Traffic-Related Air Pollutant Emissions in Different Urban Functional Zones Estimated by Real-Time Video and Deep Learning Technique." *Journal of Cleaner Production* 238:117881. doi: 10.1016/j.jclepro.2019.117881.

Spinelle, Laurent, Michel Gerboles, Maria Gabriella Villani, Manuel Aleixandre, and Fausto Bonavitacola. 2015. "Field Calibration of a Cluster of Low-Cost Available Sensors for Air Quality Monitoring. Part A: Ozone and Nitrogen Dioxide." *Sensors and Actuators B: Chemical* 215:249–57. doi: 10.1016/j.snb.2015.03.031.

Thurston, George D., Howard Kipen, Isabella Annesi-Maesano, John Balmes, Robert D. Brook, Kevin Cromar, Sara De Matteis, Francesco Forastiere, Bertil Forsberg, Mark W. Frampton, Jonathan Grigg, Dick Heederik, Frank J. Kelly, Nino Kuenzli, Robert Laumbach, Annette Peters, Sanjay T. Rajagopalan, David Rich, Beate Ritz, Jonathan M. Samet, Thomas Sandstrom, Torben Sigsgaard, Jordi Sunyer, and Bert Brunekreef. 2017. "A Joint ERS/ATS Policy Statement: What Constitutes an Adverse Health Effect of Air Pollution? An Analytical Framework." *European Respiratory Journal* 49(1):1600419. doi: 10.1183/13993003.00419-2016.

Transport Canada. 2022. *The Government of Canada and Industry Announce New Environmental Measures for Cruise Ships*. Ottawa.

Umair, Muhammad, Muhammad Umar Farooq, Rana Hammad Raza, Qian Chen, and Baher Abdulhai. 2021. "Efficient Video-Based Vehicle Queue Length Estimation Using Computer Vision and Deep Learning for an Urban Traffic Scenario." *Processes* 9(10):1786. doi: 10.3390/pr9101786.

United States Environmental Protection Agency. 2022. "CMAQ (Version 5.3)."

USEPA. 2023. "Motor Vehicle Emission Simulator: MOVES4."

Wang, An, Junshi Xu, Ran Tu, Marc Saleh, and Marianne Hatzopoulou. 2020. “Potential of Machine Learning for Prediction of Traffic Related Air Pollution.” *Transportation Research Part D: Transport and Environment* 88:102599. doi: 10.1016/j.trd.2020.102599.

Wang, Deyun, Shuai Wei, Hongyuan Luo, Chenqiang Yue, and Olivier Grunder. 2017. “A Novel Hybrid Model for Air Quality Index Forecasting Based on Two-Phase Decomposition Technique and Modified Extreme Learning Machine.” *Science of The Total Environment* 580:719–33. doi: 10.1016/j.scitotenv.2016.12.018.

Wang, Yang, Jiayu Li, He Jing, Qiang Zhang, Jingkun Jiang, and Pratim Biswas. 2015. “Laboratory Evaluation and Calibration of Three Low-Cost Particle Sensors for Particulate Matter Measurement.” *Aerosol Science and Technology* 49(11):1063–77. doi: 10.1080/02786826.2015.1100710.

Wang, ZiFa, Jie Li, Zhe Wang, WenYi Yang, Xiao Tang, BaoZhu Ge, PinZhong Yan, LiLi Zhu, XueShun Chen, HuanSheng Chen, Wei Wand, JianJun Li, Bing Liu, XiaoYan Wang, Wei Wand, YiLin Zhao, Ning Lu, and DeBin Su. 2014. “Modeling Study of Regional Severe Hazes over Mid-Eastern China in January 2013 and Its Implications on Pollution Prevention and Control.” *Science China Earth Sciences* 57(1):3–13. doi: 10.1007/s11430-013-4793-0.

Wen, Congcong, Shufu Liu, Xiaojing Yao, Ling Peng, Xiang Li, Yuan Hu, and Tianhe Chi. 2019. “A Novel Spatiotemporal Convolutional Long Short-Term Neural Network for Air Pollution Prediction.” *Science of The Total Environment* 654:1091–99. doi: 10.1016/j.scitotenv.2018.11.086.

Wu, Q. Z., W. S. Xu, A. J. Shi, Y. T. Li, X. J. Zhao, Z. F. Wang, J. X. Li, and L. N. Wang. 2014. “Air Quality Forecast of PM₁₀ in Beijing with Community Multi-Scale Air Quality Modeling (CMAQ) System: Emission and Improvement.” *Geoscientific Model Development* 7(5):2243–59. doi: 10.5194/gmd-7-2243-2014.

Yang, Fan, and Guangqiu Huang. 2024. “An Optimized Decomposition Integration Model for Deterministic and Probabilistic Air Pollutant Concentration Prediction Considering Influencing Factors.” *Atmospheric Pollution Research* 15(7):102144. doi: 10.1016/j.apr.2024.102144.

Zhang, Bo, Yi Rong, Ruihan Yong, Dongming Qin, Maozhen Li, Guojian Zou, and Jianguo Pan. 2022. “Deep Learning for Air Pollutant Concentration Prediction: A Review.” *Atmospheric Environment* 290:119347. doi: 10.1016/j.atmosenv.2022.119347.

Zhang, Lanyi, Jane Lin, Rongzu Qiu, Xisheng Hu, Huihui Zhang, Qingyao Chen, Huamei Tan, Danting Lin, and Jiankai Wang. 2018. “Trend Analysis and Forecast of PM2.5 in Fuzhou, China Using the ARIMA Model.” *Ecological Indicators* 95:702–10. doi: 10.1016/j.ecolind.2018.08.032.

Zhang, Yifu and Sun, Peize and Jiang, Yi and Yu, Dongdong and Weng, Fucheng and Yuan, Zehuan and Luo, Ping and Liu, Wenyu and Wang, and Xinggang. 2022. “Multi-Object Tracking by Associating Every Detection Box.”

Zimmerman, Naomi, Albert A. Presto, Sriniwasa P. N. Kumar, Jason Gu, Aliaksei Hauryliuk, Ellis S. Robinson, Allen L. Robinson, and R. Subramanian. 2018. “A Machine Learning Calibration Model Using Random Forests to Improve Sensor Performance for Lower-Cost Air Quality Monitoring.” *Atmospheric Measurement Techniques* 11(1):291–313. doi: 10.5194/amt-11-291-2018.