# Understanding Air Quality Data, Traffic, and Weather Parameters Collected from Near-Road Stations


Ayla Moretti[1,2], Jill Luo[2], Guoyuan Wu[2], Brandon Feenstra[1,2], Kanok Boriboonsomsin[2], Matthew Barth[2]

*1. University of California, Riverside – Department of Chemical and Environmental Engineering 2. College of Engineering – Center for Environmental Research and Technology (CE-CERT)*


UCR College of Engineering- Center for Environmental Research & Technology

## Introduction

Previous research studies have established relationships between air quality, and traffic and weather parameters based on hourly averaged data. In this study, we were able to obtain higher resolution (i.e., 5-minute averaged) measurement data of fine particulate matter ($PM_{2.5}$), nitric oxide (NO), nitrogen dioxide ($NO_2$), and nitrogen oxides ($NO_x$) from two near-road air monitoring stations managed by South Coast Air Quality Management District. The objective of this study is to gain a better understanding of the relationships between near-road pollutant concentrations and traffic speed, traffic flow, weather parameters, and time of day at a higher time resolution using four months worth of air quality measurement data.

## Data Collection

Data was collected from January 2018 through April 2018 for the following locations:

**Ontario SR-60 Near Road (60NR) Air Monitoring Station**

Located at 2330 S Castle Harbour Pl, Ontario, CA 91761; approximately 10 meters North of California State Route 60 (SR-60) between the Grove Ave and Vineyard Ave exits.

**Long Beach I-710 Near Road (710NR) Air Monitoring Station**

Located at 5895 Long Beach Blvd. Long Beach, CA 90805; approximately 20 meters East of Interstate 710 (I-710) between the exits of W. Del Amo Boulevard and Long Beach Boulevard.


**Satellite images of the SCAQMD near-roadway air monitoring stations selected for this study (Sources: Google Maps and PeMS)**

## Data Preparation

The raw data obtained from the near-road air monitoring stations and PeMS required further data processing: examining of outliers, averaging values, and removing missing values. The pearson correlation coefficients was calculated to examine the linear relationship between any two numerical variables. This was done to identify potential multicollinearity issues among the variables – the results indicated that the explanatory variables were not linearly related with each other.

Box-cox transformation was performed to transform non-normal concentration values to a normal-distribution shape. Lambda of 0.5 was applied for $PM_{2.5}$ concentrations ($\mu g/m^3$) and 0.2 was applied for NO, and $NO_x$ concentrations (ppb). The Box-cox transformation did not improve $NO_2$ distribution and therefore was not applied to $NO_2$ concentration values.

## Modeling Method

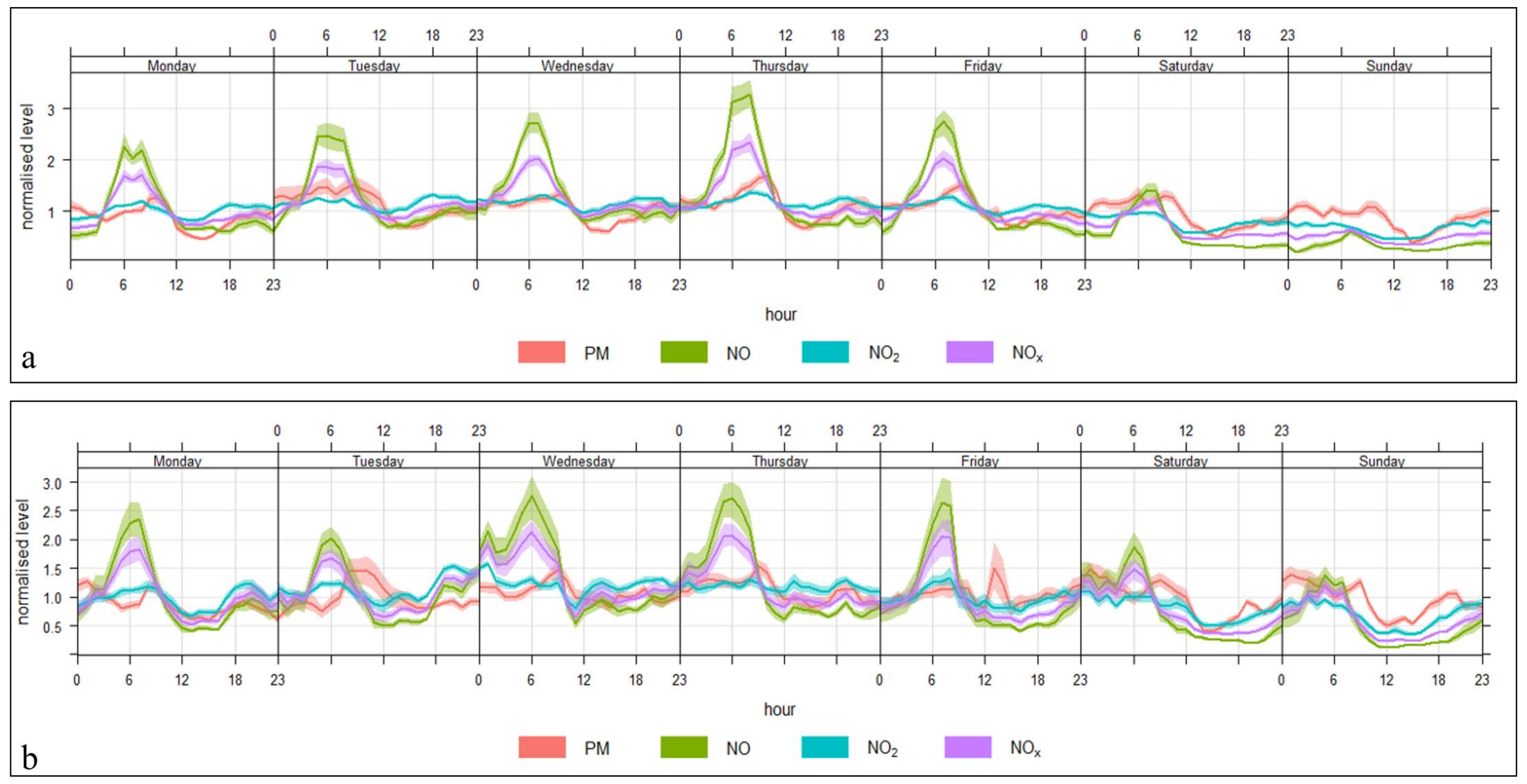
**Multiple Linear Regression (MLR)**

MLR is the simplest multivariate regression method that models the linear relationship between the explanatory variables on the observed traffic, and weather parameters on $PM_{2.5}$, NO, $NO_2$, and $NO_x$ concentration. The general equation for the MLR model can be written as:

$$y = \beta_0 + \sum_i \beta_i * x_i + \varepsilon_i$$

where y represents the estimated model output; $\beta_0$ is the intercept; $\beta_i$ is the regression coefficient associated with the i-th variable; $x_i$ is the value of the i-th variable; and $\varepsilon_i$ is an independent, normally distributed, random error with zero mean and constant variance.

## Results




**Normalized pollutant concentration based on day of week and time of day: a) SR-60 and b) I-710**

### 5-minute Averaged Data

| 5-min Averaged - Whole Day | SR-60 | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | $PM_{2.5}$ | | NO | | $NO_2$ | | $NO_x$ | |
| | $\beta_i$ | p-value | $\beta_i$ | p-value | $\beta_i$ | p-value | $\beta_i$ | p-value |
| **Intercept** | **2.49E+00** | <2e-16 | **2.80E+00** | < 2e-16 | **8.65E+01** | < 2e-16 | **3.16E+00** | < 2e-16 |
| **RH** | **7.37E-03** | <2e-16 | **5.37E-04** | 9.78E-08 | **-1.37E-01** | < 2e-16 | **-7.76E-04** | < 2e-16 |
| **Temperature** | **2.05E-02** | <2e-16 | **-5.05E-03** | < 2e-16 | **-2.97E-02** | 0.000217 | **-3.58E-03** | < 2e-16 |
| **Pressure** | 9.26E-03 | 0.14 | 2.48E-03 | 0.13867 | 5.69E-02 | 0.332666 | 2.08E-03 | 0.098657 |
| **Wind Speed** | **2.80E-01** | <2e-16 | **4.04E-02** | < 2e-16 | **3.02E+00** | < 2e-16 | **5.11E-02** | < 2e-16 |
| **Wind Direction** | **-1.08E-03** | <2e-16 | **-8.25E-04** | < 2e-16 | **-2.22E-02** | < 2e-16 | **-6.11E-04** | < 2e-16 |
| **Speed West A** | -5.77E-02 | 0.11 | 3.79E-03 | 0.69396 | **-2.49E+00** | 1.57E-13 | **-1.83E-02** | 0.011598 |
| **Speed West B** | -3.38E-02 | 0.297 | **-6.32E-02** | 2.52E-13 | **7.35E-01** | 0.015134 | **-2.30E-02** | 0.000391 |
| **Flow West A** | -4.18E-06 | 0.99 | 1.55E-04 | 0.07816 | 3.06E-03 | 0.322101 | **1.63E-04** | 0.014213 |
| **Flow West B** | -2.43E-04 | 0.515 | **5.76E-04** | 6.81E-09 | **-1.92E-02** | 3.30E-08 | 5.91E-05 | 0.429529 |
| **Speed East A** | **-2.37E-01** | <2e-16 | **-5.53E-02** | 1.42E-14 | **-7.94E-01** | 0.00163 | **-4.11E-02** | 3.38E-14 |
| **Speed East B** | **3.66E-01** | <2e-16 | **2.56E-02** | 0.00101 | **-1.92E+00** | 2.04E-12 | 4.80E-03 | 0.411926 |
| **Flow East A** | **7.43E-03** | <2e-16 | **3.18E-03** | < 2e-16 | **7.77E-02** | < 2e-16 | **2.36E-03** | < 2e-16 |
| **Flow East B** | **-7.61E-03** | <2e-16 | **-3.09E-03** | < 2e-16 | **-6.51E-02** | < 2e-16 | **-2.19E-03** | < 2e-16 |
| **Timestamp** | **-3.41E-05** | <2e-16 | **-1.02E-05** | < 2e-16 | **-3.46E-04** | < 2e-16 | **-8.50E-06** | < 2e-16 |
| **Multiple R-squared** | 0.1457 | | 0.2916 | | 0.2306 | | 0.2708 | |
| **Adjusted R-squared** | 0.1452 | | 0.2913 | | 0.2302 | | 0.2704 | |
| **p-value** | < 2.2e-16 | | < 2.2e-16 | | < 2.2e-16 | | < 2.2e-16 | |
| **Degrees of Freedom** | 26803 | | 26803 | | 26803 | | 26803 | |
| **Residual Standard Error** | 1.294 | | 0.345 | | 12.09 | | 0.2598 | |

Variables in boldface are statistically significant at 5% α-level

For SR-60 the speed and flow in the East direction is generally significant for all pollutants, and the speed and flow in the West direction is generally significant for NO, $NO_2$ and $NO_x$.

| 5-min Averaged - Whole Day | I-710 | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | $PM_{2.5}$ | | NO | | $NO_2$ | | $NO_x$ | |
| | $\beta_i$ | p-value | $\beta_i$ | p-value | $\beta_i$ | p-value | $\beta_i$ | p-value |
| **Intercept** | **4.96E+00** | < 2e-16 | **4.46E+00** | < 2e-16 | **7.32E+01** | < 2e-16 | **4.06E+00** | < 2e-16 |
| **RH** | **1.08E-02** | < 2e-16 | **-4.41E-03** | < 2e-16 | **-1.55E-01** | < 2e-16 | **-3.53E-03** | < 2e-16 |
| **Temperature** | **1.20E-02** | 5.80E-15 | **-1.91E-02** | < 2e-16 | -8.03E-03 | 0.6074 | **-1.13E-02** | < 2e-16 |
| **Pressure** | - | - | - | - | - | - | - | - |
| **Wind Speed** | **3.82E-01** | < 2e-16 | **1.77E-01** | < 2e-16 | **5.84E+00** | < 2e-16 | **1.65E-01** | < 2e-16 |
| **Wind Direction** | **3.01E-04** | 0.0029 | **2.05E-03** | < 2e-16 | **3.87E-02** | < 2e-16 | **1.36E-03** | < 2e-16 |
| **Speed North A** | **-9.00E-02** | 1.20E-06 | 1.71E-03 | 0.7774 | **-3.94E-01** | 0.0373 | **1.48E-02** | 0.002428 |
| **Speed North B** | -1.86E-02 | 0.3353 | **-6.51E-02** | < 2e-16 | **-8.80E-01** | 7.94E-06 | **-5.68E-02** | < 2e-16 |
| **Flow North A** | 2.51E-05 | 0.7495 | **-1.15E-04** | 6.77E-06 | **-3.33E-03** | 3.30E-05 | **-6.87E-05** | 0.000907 |
| **Flow North B** | **4.67E-04** | 0.0284 | **-7.13E-04** | < 2e-16 | **-4.06E-02** | < 2e-16 | **-8.04E-04** | < 2e-16 |
| **Speed South A** | 1.48E-02 | 0.3108 | **-2.40E-02** | 4.72E-07 | **-1.48E+00** | < 2e-16 | **-2.07E-02** | 8.57E-08 |
| **Speed South B** | **-1.56E-01** | 5.48E-09 | **-1.07E-01** | < 2e-16 | **-2.07E+00** | 3.62E-14 | **-8.40E-02** | < 2e-16 |
| **Flow South A** | **-1.65E-03** | < 2e-16 | **8.44E-04** | < 2e-16 | **1.71E-02** | < 2e-16 | **6.72E-04** | < 2e-16 |
| **Flow South B** | **1.41E-03** | 1.95E-09 | **2.09E-04** | 0.00621 | **1.03E-02** | 1.73E-05 | **1.57E-04** | 0.011014 |
| **Timestamp** | **-5.08E-05** | < 2e-16 | **-1.91E-05** | < 2e-16 | **-3.95E-04** | < 2e-16 | **-1.46E-05** | < 2e-16 |
| **Multiple R-squared** | 0.137 | | 0.3236 | | 0.2528 | | 0.3022 | |
| **Adjusted R-squared** | 0.1366 | | 0.3232 | | 0.2524 | | 0.3019 | |
| **p-value** | < 2.2e-16 | | < 2.2e-16 | | < 2.2e-16 | | < 2.2e-16 | |
| **Degrees of Freedom** | 25372 | | 25372 | | 25372 | | 25372 | |
| **Residual Standard Error** | 1.362 | | 0.4433 | | 13.91 | | 0.3588 | |

Variables in boldface are statistically significant at 5% α-level

For I-710 the speed and flow in the South direction is generally significant for all pollutants, and the speed and flow in the North direction is generally significant for NO, $NO_2$, and $NO_x$.

### Morning vs Afternoon

| Adjusted $R^2$ values | | $PM_{2.5}$ | NO | $NO_2$ | $NO_x$ |
|---|---|---|---|---|---|
| SR-60 | Morning | 0.13 | 0.3628 | 0.2799 | 0.3443 |
| | Afternoon | 0.1659 | 0.1947 | 0.396 | 0.3055 |
| I-710 | Morning | 0.1788 | 0.4241 | 0.2853 | 0.3904 |
| | Afternoon | 0.1223 | 0.3412 | 0.3113 | 0.3176 |

**Adjusted $R^2$ values for morning and afternoon for SR-60 and I-710**

Generally NO and $NO_x$ have a higher $R^2$ value when looking only at the morning, and $NO_2$ tends to have a higher $R^2$ value in the afternoon. $PM_{2.5}$ has a higher $R^2$ value in the morning for I-710 and in the afternoon for SR-60. In some cases, such as $NO_2$, breaking the model down based of morning and afternoon showed an improved $R^2$ for both when compared to the whole day model.

## Conclusions and Future Work

In this study, 15 variables and more than 26,000 rows of data were collected, including weather parameters, traffic speed, and traffic volume near the air monitoring stations. MLR models were applied to the data to examine the relationship among the weather parameters, traffic data, time of day, and near-freeway air pollutant concentration. Generally, $NO_x$ concentrations can be better explained by the selected variables than $PM_{2.5}$. Additionally time of day (morning vs afternoon) showed to have a significant impact in some cases.

Future work includes gathering more data to gain a better understating on how congestion affects air quality pollutants and including truck flow data into the models. With more near-road air quality data, we hope to better explain the concentration variations and predict the near-road pollutant concentrations in the future.

## Acknowledgements


The authors would like to acknowledge the Center for Advancing Research in Transportation Emissions, Energy, and Health (CARTEEH) for funding this research, and the South Coast Air Quality Management District (SCAQMD) for providing air pollution and weather data from the monitoring stations.