Distribution Agreement

Signature:

______________________________ 4/15/17

Sarita Mohanty Date

Identifying land use and meteorological factors associated with the ratio of personal $PM_{2.5}$ exposures versus ambient concentration in a panel of college students

By

Sarita Mohanty

Master of Public Health

Biostatistics

Howard Chang, Ph.D.

Committee Chair

Donghai Liang, Ph.D

Committee Member

Identifying land use and meteorological factors associated with the ratio of personal $PM_{2.5}$ exposures versus ambient concentration in a panel of college students

By

Sarita Mohanty

B.A.

Purdue University

2014

Thesis Committee Chair: Howard Chang, Ph.D.

An abstract of

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University

in partial fulfillment of the requirements for the degree of
Master of Public Health

in Biostatistics

2020

Abstract

Identifying land use and meteorological factors associated with the ratio of personal $PM_{2.5}$ exposures versus ambient concentration in a panel of college students

By Sarita Mohanty

Exposure to air pollution, such as fine particulate matter $PM_{2.5,}$ have been associated with many adverse health outcomes. A large portion of the urban population live within a close distance to major roadways. Given that traffic is a major source of $PM_{2.5}$, there is an increased interest in understanding personal exposure to $PM_{2.5}$ at fine spatiotemporal resolution. This study is motivated by a dataset of personal $PM_{2.5}$ measurements from wearable sensors at 15 second intervals with geo-location information, and hourly meteorological data from stationary monitors every hour. We built linear regression models and random forest models for predicting the ratio between personal exposure to $PM_{2.5}$ and background ambient $PM_{2.5}$ levels. Both approaches identified several land use and meteorological factors, including distance to highway, traffic counts, temperature, and relative humidity.

Identifying land use and meteorological factors associated with the ratio of personal $PM_{2.5}$ exposures versus ambient concentration in a panel of college students

By

Sarita Mohanty

BA

Purdue University

2014

Thesis Committee Chair: Howard Chang, Ph.D.

A thesis submitted to the Faculty of the

Rollins School of Public Health of Emory University

in partial fulfillment of the requirements for the degree of
Master of Public Health
in Biostatistics

2014

Acknowledgements

This thesis marks the culmination of two challenging but rewarding years at the Rollins School of Public Health, and I am indebted to many who have helped and supported me along the way.

First and foremost, I would like to thank my advisor, Howard Chang who has been an incredible mentor and has always challenged me to think beyond my comfort zone. The regular encouragements and infectious enthusiasm have not only helped me complete my thesis but also have shaped me into a better statistician. Secondly, I would like to thank Donghai Liang for being a wonderful knowledge repository and providing me with insightful feedback on my thesis.

I would also like to thank the friends I have made in the Biostatistics department who have shared laughs with me and given me hope in hopeless situations. Additionally, I am grateful to the Biostatistics faculty who have been inspired me with their dedication to the subject and who have been a strong support system.

Finally, I would like to express my unending gratitude to my parents and my brother, who have been a source of comfort, support, and affection from across the world. It has been difficult to be away from home for so long, but I hope I am making you proud.

Tables of Contents

1. Introduction

Outdoor air pollution is a mixture of many different pollutants that originate from natural and anthropogenic sources. Anthropogenic sources are increasing due to greater needs in transportation, power generation, industrial activity, and domestic heating and cooking. Fine particles with aerodynamic diameters less than 2.5 micrometers ($PM_{2.5}$) is a common air pollutant that is often used as an indicator of air quality. Cohort and case-control studies across Europe, North America and Asia have shown that there are positive relationships between high levels of $PM_{2.5}$ and adverse health outcomes such as lung cancer, asthma, and cardiovascular diseases (Sarnat et al., 2018) (Loomis et al., 2013) Moreover, the World Health Organization declared air pollution to be one of the major environmental health risks and estimated 4.2 million premature deaths in 2016 to be due to outdoor air pollution (Guan et al., 2019). An estimated 30 – 45% of the population in large North American cities live within a range of up to 300 to 500 m from a major road, an environment that is highly affected by traffic-related air pollution (Sarnat et al., 2018).

As the concern around exposure to harmful air pollutants grows in the world, there is a greater need to understand personal exposure to air pollution on a real-time basis and at a geographically localized scale. The conventional way of measuring outdoor air pollution is by using stationary monitors. These monitors are highly reliable, accurate and measure many kinds of pollutants. However, the instruments are large, heavy and very expensive. The locations of these monitors are often based on careful considerations to capture specific sources (e.g. traffic) or measure background levels. As a result, they are sparsely distributed and cannot capture the highly heterogenous spatial variations within cities, especially near roadways. The stationary monitors may also lead to substantial measurements errors for the use in epidemiology study as shown by Liang et al (Liang et al., 2018). Finally, most stationary monitors provide readings every hour or every day. Therefore, the air pollution maps created by the stationary monitors have low spatial and temporal resolutions (Yi et al., 2015)

Reliance on low spatiotemporal resolution of air quality data will limit the public's awareness of their personal exposure and therefore their personal health risks. As stated previously, air pollution varies greatly with a myriad of factors. Research has shown that pollution concentrations within one street can vary greatly over a few meters and a few seconds (Tiwary et al., 2011). Moreover, research has also shown that personal measurements of black carbon, a pollutant that reflect vehicular combustion, can be vastly different from ambient roadside monitoring (Sarnat et al., 2018)

To remedy this issue, there has been development of small, portable monitors that can be used in various ways to measure pollutants and can also be available at low cost. Mobile monitors can either be small devices that measure as they are in motion or can be stationary portable monitors that are easily moved between locations. Mobile monitors are especially useful in potentially capturing the spatiotemporal variability in air pollution that results from traffic volume, distance from sources, pollutant chemistry, and local weather patterns (Moltchanov et al., 2015). Since pollution levels can vary in small spatiotemporal scales, dense networks offer an increased spatial flexible that cannot be found in stationary monitors (Brantley et al., 2014).

Development of low-cost air quality sensors is a rapidly evolving technology area. Sensors are now available commercially in a wide variety of designs and capabilities and cost between $100 - $2500 (Nathan & Scobell, 2012). The MicroPEM is an example sensor for characterizing personal P.M. exposure that is wearable and can provide measurements every 15 seconds. The MicroPEM also contains a Teflon filter that can be gravimetrically weighed to calibrate real time data (Yan et al., 2018)(RTI International, 2016). These wearable air pollution monitors allow individuals to understand their exposure to air pollution while doing certain activities or being at certain locations and thereby remedy their lifestyle if needed (Sarnat et al., 2018). It also allows for researchers to quantify the health effects of air pollution in a more personalized manner.

This thesis describes an analysis of MicroPEM data with high spatial and temporal resolutions collected from 34 students of the Georgia Institute of Technology during the period September 22nd 2014 – December 5th 2014. The main aims are to develop models for the ratio of personal exposure to $PM_{2.5}$ versus ambient background level using predictors such as meteorological conditions, time of day, and distance to major roadways. We focus on the ratio to help elucidate factors that can contribute to error associated with use of background levels measured at standard fixed location monitors instead of actual personal exposures.

2. Methods

2. 1 The Dorm Room Inhalation to Vehicular Exposure (DRIVE) Study

The Dorm Room Inhalation to Vehicle Emissions (DRIVE) study was conducted to measure traditional single-pollutant and novel multipollutant traffic indicators along the busiest and most congested highway artery in the geographic core of Atlanta. Personal exposure data were collected through field sampling on the Georgia Institute of

Technology's (GIT) campus. The GIT campus is adjacent to one of the most heavily trafficked highway arteries (I-75/I-85 highway connector). Moreover, there are also many smaller roadways that surround the GIT campus that could also contribute towards $PM_{2.5}$. The DRIVE study recruited 34 students who lived in the two different dormitories – one near the highway connector (Near dorm) and one that was farther away (Far Dorm). The Near dorm was approximately 20 m west of the connector highway while the far dorm was approximately 1.4 km west of the connector highway (Sarnat et al., 2018) The subjects only walked on weekdays (Monday – Friday) and therefore, no data were collected on weekends.

The GIT map below shows a sample of the recorded locations of all the participants, the two dormitories, as well as the connector highway and secondary roads surrounding campus. The buildings on campus are also marked.

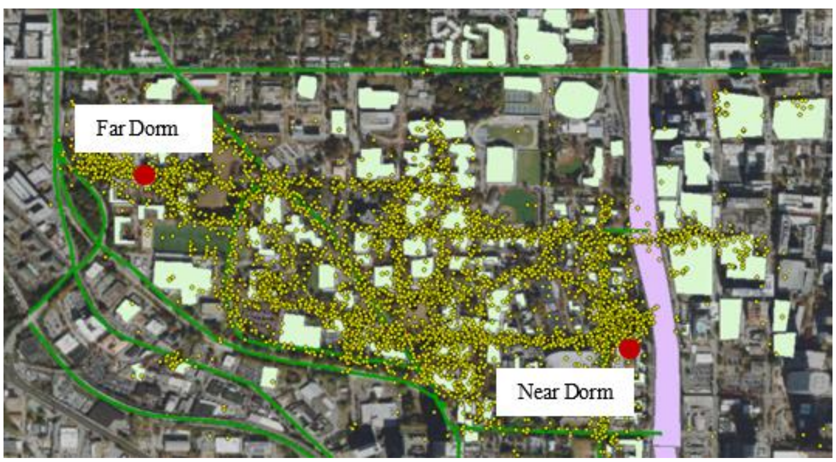

Figure 1: Map of GIT campus - Red dots indicate the location of the dorm, Light green polygons represent the buildings in and around the campus, Purple polygon represents the connector highway, the Green lines represent the secondary roads in and around the campus

2.1.1 Personal Exposure Data

Continuous personal exposure monitoring was conducted on days when students were on campus between September 22nd, 2014 to December 5th, 2014. Each week, about six students participated in two consecutive 48-hour personal exposure sampling sessions where they were asked to carry a 3lbs personal sampling pack that recorded $PM_{2.5}$ at one-minute intervals. The measurements were done via the nephelometry by the MicroPEMs that contained 37-mm Teflon Filters, which collected particles through the sampling inlets at a 0.5-L/min flow rate. (Sarnat et al., 2018) Measurements were bias-corrected using gravimetric analysis of the 48- hour integrated Teflon filter within the personal sampling devise.

2.1.2 Location Data

Each subject's position was recorded using global positioning system (GPS) trackers that were attached to the side of the personal sampling pack. When a participant stayed within a radius of about 6m for longer than 5 minutes, their locations were recorded as waypoints. GPS measurements were recorded every second when the participant was moving.

2.1.3 Meteorological and Background $PM_{2.5}$ Data

There are several stationary monitors around the GIT campus. For the purposes of this analysis, we obtained meteorological data from the roadside monitor (RDS). The RDS monitor is 10m away from the closest lane of the 15 – lane connector highway and the area surrounding the monitor was a campus parking lot. The meteorology monitor measured the temperature, rain volume, relative humidity, wind speed, wind direction at every hour. We also obtained hourly $PM_{2.5}$ measurement from the Jefferson Street (JST) monitor, which is located 2.4 km from the highway, to reflect the background level. The two monitors are shown in Figure 2.

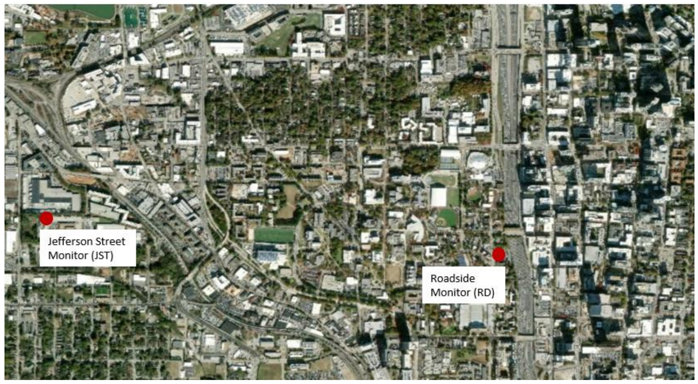

Figure 2: Map of GIT campus - Red dots indicate the location of the stationary monitors that were used in this study

2.1.3 Traffic Count Data

The hourly traffic counts data for the connector highway was taken from the Georgia Department of Transportation's database. However, there was no traffic count data available between 10/22/2014 to 11/11/2014. Since traffic volumes remained consistent for all the days, we assigned the hourly mean of traffic counts for the missing days from available data.

2.2 Development of the Analytic Dataset

In order to build spatiotemporal models for personal exposure to $PM_{2.5}$, we first linked various datasets to a common spatial and temporal scale. Since the $PM_{2.5}$ measurements were taken every minute, the GPS locations were recorded every 15 seconds and the meteorological variables were recorded every hour, we decided to aggregate the GPS data to the minute-scale. Hourly background $PM_{2.5}$ and meteorological measurements were assigned to for the corresponding minute.

For the days between 10/22/2014 to 11/11/2014, the meteorological measurements from the RDS monitor were missing. As a result, we used a regression-based imputation technique to predict the temperature, relative humidity, and wind speed values for the missing days. The predictors for the regression for each of the variables were the respective measurements from two additional meteorological monitors in the study region. Table 1 shows the linear models that were used to impute the meteorological variable. Our ability to impute hourly temperature and relative humidity is excellent. Since wind direction is measured in degrees, we did not use the same regression-based imputation technique and instead used the mean of the hour from all the available days to impute the values.

Table 1: Linear models used for regression-based imputation on selected meteorological variables

Meteorological Variable	**Linear Model**	**R^2**
Temperature	$Temp_{RDS} = -0.057 + 0.386Temp_{JST} + 0.627Temp_{RF}$	0.996
Relative Humidity	$RH_{RDS} = 6.627 + 0.236RH_{ATL} + 0.656RH_{JST}$	0.961
Wind Speed	$WSPD_{RDS} = 0.649 + 0.090WSPD_{ATL} + 0.334WSPD_{JST} + 0.172WSPD_{RF}$	0.328

Furthermore, if there was no location data for a minute for a participant, their location was treated to be the same as their last recorded location. Finally, there were several time points for which traffic counts were not available. However, looking at Figure 3, we find there is a distinct pattern on traffic flow on the connector highway. Therefore, for the missing hours, we use the mean of the hour from all the available days to impute their values.

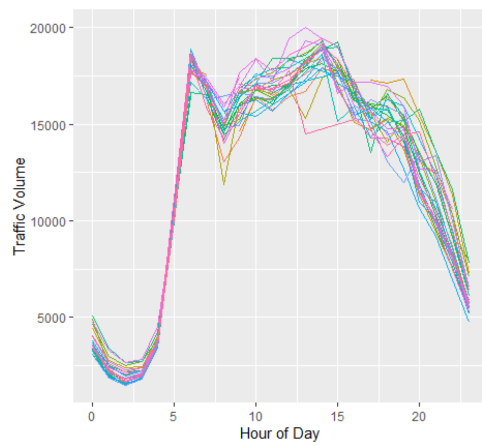

Figure 3: Daily Weekday Traffic Volume trends – each color represents a day in the study

2.3 Derived Variables

Several derived variables were created that were used as predictors for personal exposure to $PM_{2.5}$

Indoor Indicator - Buildings were manually marked using ArcMap's default basemap. If the GPS measurements overlapped with the buildings, the indoor indicator takes the value 1 at the time.

Upwind Indicator - To assess if the wind was blowing from the highway towards participant (downwind), the wind field were plotted on ArcMap (Figure 4) and visually assessed to create a rule for points being upwind or downwind. For example, points to the west of the highway were assigned to be downwind if the wind direction was greater than 180^{o} while the points to the east of the highway.

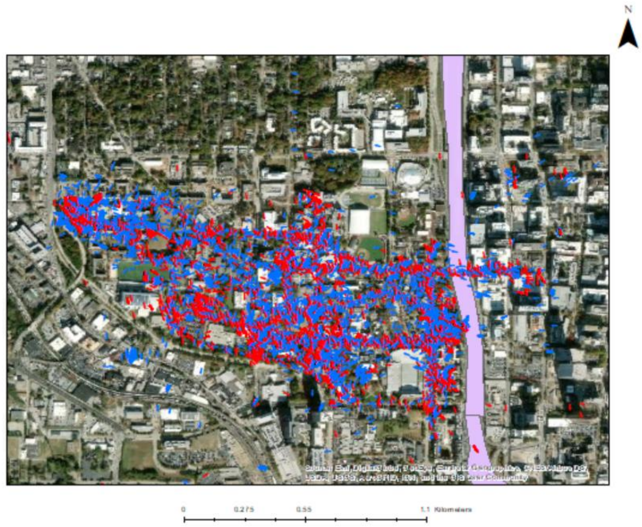

Figure 4: Upwind vs. Downwind – The blue arrows represent downwind points and the red arrows represent upwind points

Time Intervals – To assess how time of day affects personal exposure, each day was divided into five intervals as follows:

T1 (Morning): 7am – 10am

T2 (Midday): 10am – 4pm

T3 (Evening): 4pm – 8pm

T4 (Night): 8pm – 12am

T5 (Early Morning): 12am – 7am

Secondary Road – Besides the connector highway, there are several other streets inside or around the main GIT campus. We created an indicator variable for whether there was a secondary road within a 50m buffer area.

Distance to Connector Highway – This was calculated using ArcMap by spatially joining the GPS measurements with the highway polygon. The closest distance between the points and the highway was calculated in meters.

2.4 Data Analysis

The outcome variable for this analysis, RatioPM, was defined to be the ratio between the 15-second personal $PM_{2.5}$ measurement as recorded by the MicroPEM sensor and the hourly $PM_{2.5}$ measurement as recorded by the JST monitor. The purpose of using this outcome variable rather than the personal $PM_{2.5}$ measurement is to help identify predictors that are associated with differences between personal exposures and background levels. Specifically, we wanted to evaluate various variables may impact the error associated with the use of ambient $PM_{2.5}$ levels to assign exposure.

Linear Regression Model (LR)

We first constructed a linear regression model by fitting a full model with RatioPM as the outcome and all the predictors. Then, using the regression subset procedure in R (package: leaps) we selected the following linear model based on the maximum adjusted r^2 value

$$\begin{aligned} RatioPM = {} & \mathrm{T1} + \mathrm{T2} + \mathrm{T3} + \mathrm{T4} + \text{Temperature} + \text{Relative Humidity} \\ & + \text{Wind Speed} + \text{Traffic Count} + \text{Indoors} \\ & + \text{Distance from Highway} + \text{Upwind} + \text{Secondary Road} + \varepsilon \end{aligned}$$

Random Forest Regression Model (RF)

To create a RF model, we used the *Ranger* package in R with a split rile that was based on minimizing variance. We used the same predictors that we included in the LR model. We set the number of trees to be 500 and number of variables to possibly split at in each node to be 5. The importance of each variable in the model was based on the variance of the response variable.

2.5 Prediction Grid Calculations

To better visualize model predictions over the study region, we designed a 100 x 100 grid on top of the GIT campus to be able to make predictions (Figure 6). Each grid measured an area of 18m x 18m. Predictions from the models were made for each hour for each grid centroid.

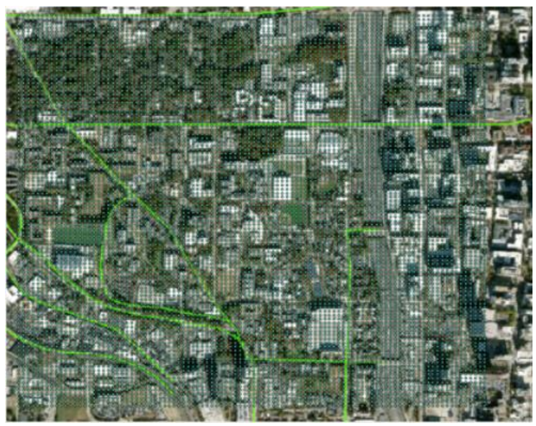

Figure 6: Grid centroids of 100 x 100x. Each grid is 18m x 18m and the total area on which RatioPM was predicted is

3. Results

The summary table for the predictors included in the dataset is shown in Table 2.

Table 2: Summary of variables measured or derived in the study, N=34

	Mean (SD)
Personal $PM_{2.5}$ ($\mu g/m^3$)	8.34 (5.71)
Jefferson St. $PM_{2.5}$ ($\mu g/m^3$)	10.64 (6.09)
$PM_{2.5}$ Ratio	0.95 (0.74)
Hourly Temperature (Farenheit)	61.13 (12.98)
Hourly Relative Humidity (%)	64.67 (17.44)
Hourly Windspeed (mph)	2.09 (1.38)
Hourly Traffic Count	12534 (5711)
Indoors (Count/% of Total)	27130 (23.5%)
Downwind (Count/% of Total)	65065 (56.2%)
Secondary Road (Count/% of Total)	30150 (26.1%)
Distance from Connector Highway (m)	465.62 (480.80)

The table on the next page shows the variables stratified by each within-day time period (Table 3).

Table 3: Summary of variables measured or derived in the study stratified by Time, Mean (SD), N=34

	7 – 10am	10 – 4pm	4pm – 8pm	8pm – 12am
Personal $PM_{2.5}$ ($\mu g/m^3$)	8.48 (5.80)	7.98 (5.17)	8.22 (5.19)	8.96 (6.81)
Jefferson St. $PM_{2.5}$ ($\mu g/m^3$)	14.19 (7.71)	7.59 (4.46)	9.14 (4.27)	12.15 (5.91)
Ratio $PM_{2.5}$	0.70 (0.50)	1.30 (1.00)	1.04 (0.69)	0.81 (0.56)
Temperature (Farenheit)	57.12 (11.54)	67.80 (11.34)	65.57 (12.37)	59.74 (11.77)
Relative Humidity (%)	74.61 (14.03)	54.54 (16.31)	53.54 (14.67)	64.46 (13.45)
Windspeed (mph)	1.98 (1.18)	2.99 (1.06)	2.13 (1.41)	1.67 (1.33)
Traffic Count	15940 (1017)	17559 (922)	15456 (892)	9715 (2683)
Indoors (Count/% of Total)	2888 (21.1%)	8105 (26.7%)	4992 (24.2%)	4390 (22.8%)
Downwind (Count/% of Total)	7859 (57.3%)	17133 (56.5%)	11773 (57.2%)	10705 (55.5%)
Secondary Road (Count/% of Total)	3433 (25.0%)	6337 (20.9%)	4979 (24.2%)	5836 (30.3%)
Distance from Connector Highway (m)	463.55 (456.08)	494.88 (415.54)	481.32 (475.97)	448.40 (509.26)

3.1 LR Model

The coefficients and their associated p-values for the predictors in the LR model are shown in Table 4.

Table 4: LR Model Predictors, R^2 = 0.202

	Estimate	95% Confidence Interval	p-value
Intercept (T5)	2.078	(2.047, 2.110)	*
T1 (7am – 10am)	-0.073	(-0.091, -0.056)	*
T2 (10am, - 4pm)	0.448	(0.428, 0.468)	*
T3 (4pm – 8pm)	0.196	(0.178, 0.214)	*
T4 (8pm – 12am)	-2.611×10^{-4}	(-0.014, 0.013)	0.286
Temperature (Fahrenheit)	-0.013	(-0.013, -0.012)	*
Relative Humidity (%)	-8.263×10^{-3}	(-8.539×10^{-3}, -7.987×10^{-3})	*
Wind Speed (mph)	0.057	(0.053, 0.060)	*
Traffic Counts	-1.480×10^{-6}	(-2.658×10^{-6}, (-3.012×10^{-7}	*
Indoors	-0.097	(-0.107, -0.087)	*
Distance from Connecter (meters)	-5.409×10^{-5}	(-6.280×10^{-5}, -4.538×10^{-5})	*
Downwind	-0.054	(-0.062, -0.045)	*
Secondary Road	0.037	(0.028, 0.047)	*

* Indicates p-value < 0.01

3.2 RF Model

As mentioned previously, the variable importance of the RF model was based on the variance of the response. Figure 7 shows the ranking of the variables in the model. We see that the relative humidity is an important variable in all three models.

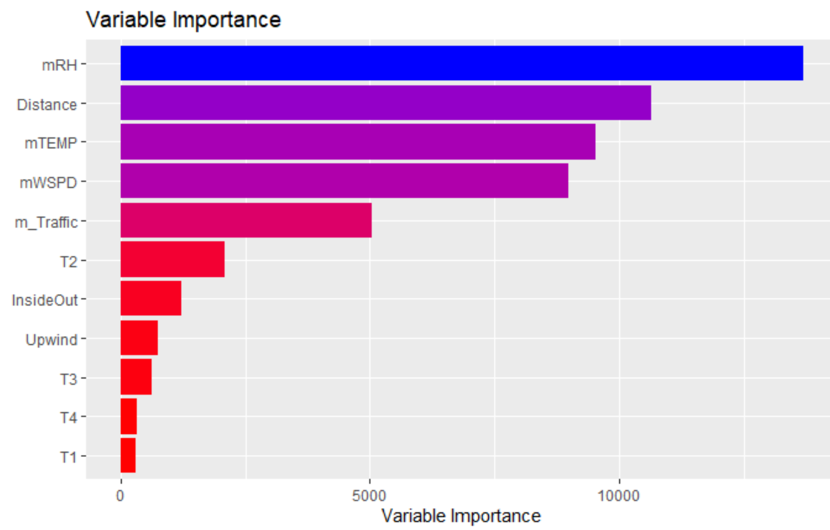

Figure 7: Variable Importance – The RF model shows relative humidity to be the most important variable in the model, followed by Distance from Highway, Temperature, and Wind Speed

3.3 Predictions based on LR and RF Models

We predicted in-sample RatioPM using both the LR and the RF model and compared them to the observed RatioPM. Figure 5 shows that the RF model is better able to capture extreme values and provides estimates that are more similar in distribution to the observed RatioPM.

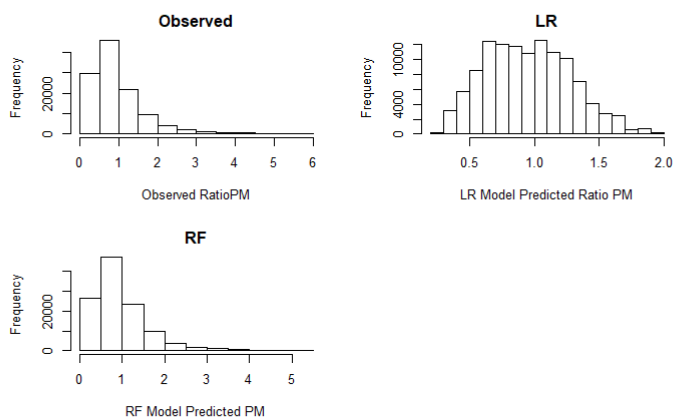

Figure 5: Distributions of observed RatioPM, LR Model Predicted RatioPM and RF Model Predicted RatioPM

Using the two models, we created predictions for each hour on the grid shown in Figure 6. The average for each of the time period (T1 – T4) for each of the model is shown below in Figure 8.

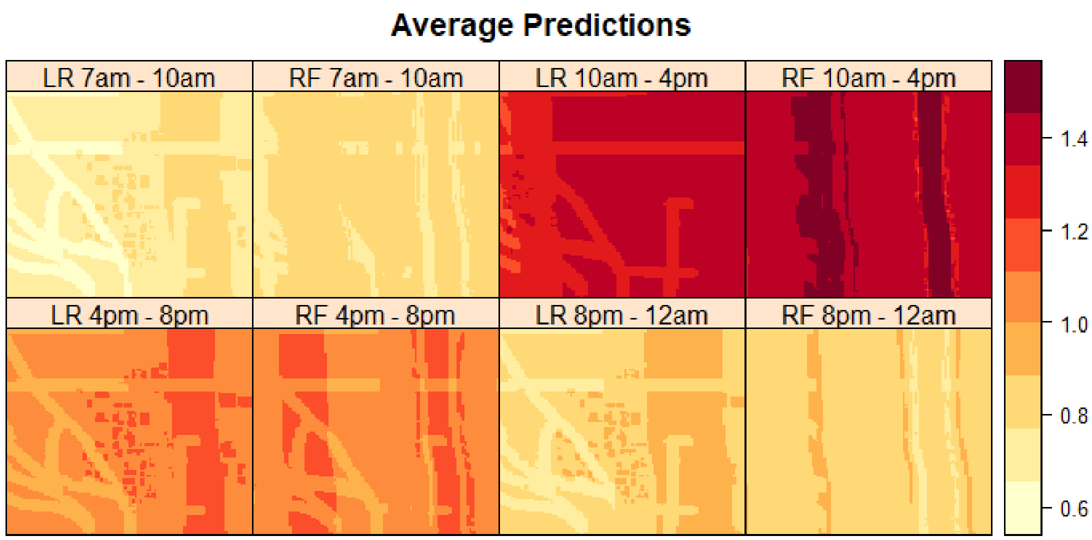

Figure 8: Predictions of the ratio between personal PM2.5 exposures and ambient concentration by time periods based on LR (top panel) and RF models

Table 4 below shows the average RatioPM for all the 10,000 grids for each of the time periods.

Table 4: Average Predictions of RatioPM by Time, Mean (SD)

	LR	RF
7 – 10am	0.725 (0.056)	0.795 (0.032)
10 – 4pm	1.338 (0.056)	1.414 (0.044)
4 – 8pm	1.065 (0.056)	1.068 (0.052)
8 – 12am	0.843 (0.056)	0.848 (0.034)

The LR and the RF model produce some interesting results. First, we see that ratio of personal exposure to ambient $PM_{2.5}$ is the highest in the time period from 10am to 4pm. This is in line with the observed data as well since the mean RatioPM for 10am – 4pm is 1.30 (Table 3). It is interesting to note that the personal exposure of the subjects remains consistent through the different time periods. However, the background $PM_{2.5}$ levels as measured by the JST monitor appear to change quite dramatically between the time periods, thereby leading to changes in the RatioPM.

From the linear model, we can see the direction of each of the predictors in relationship to the response. Among the continuous variables, temperature, relative humidity, traffic counts, and distance from highway appear to be negatively related with RatioPM, while wind speed appears to be positively related with RatioPM. Among the categorical variables, we see that being indoors, and downwind are negatively related to RatioPM, while within an 50m-buffer of secondary road is positively related. All these variables are significant predictors in the model. It is interesting to see that traffic count is negatively related to RatioPM. This means that with a higher volume of traffic, the ratio between personal exposure and ambient $PM_{2.5}$ decreases. This could be explained by the fact that an increase in traffic in the connector highway creates higher level of ambient air pollution and therefore reduces the ratio of personal exposure to ambient $PM_{2.5}$. Similarly, being within a 50m proximity of secondary road leads to an increase in the RatioPM.

Distance from highway is negatively correlated with RatioPM. This is in line with our hypothesis that as an individual move farther away from the highway, the individual's RatioPM lowers. Given the large volume of traffic on the connector highway, an individual who is very close to the highway is likely to be exposed to $PM_{2.5}$ at a much higher level than the ambient $PM_{2.5}$ levels on campus.

Moreover, we see that the random forest model generally produces slightly higher predictions for RatioPM. Given, the model's inherent ability to account for extreme values, the model is better able to predict higher values. The RF model also shows relative humidity to be the most important predictor in RatioPM. Cheng et al. shows that an increase in relative humidity is positively correlated with $PM_{2.5.}$ They also showed that high humidity conditions were favorable in the formation of air pollution episodes with high $PM_{2.5}$ concentrations in Bejing, China (Cheng et al., 2015)This implies that an increase in relative humidity increases ambient air pollution and therefore we should expect to see a smaller RatioPM.

In the linear regression model, we see that the RatioPM for participants who were indoors was higher in all time periods except for between 10-4pm. Given that the ambient air pollution was assumed to be constant throughout the entire area for the hour, this indicates that personal exposure to $PM_{2.5}$ was higher indoors in those time periods. This could indicate some potential sources of indoor air pollution that the students were exposed to in the campus's buildings.

4. Discussion

4.1 Limitations

Limitations of this study include the relatively small area of the GIT campus. Second, we used GPS points to identify subject's locations and to determine key predictor variables such as whether they were indoors and their distance from the highway. However, GPS data is known to exhibit spatial measurement error, which could impact our results. Moreover, our methodology in creating an indicator variable for indoors or outdoors can be improved on, since we visually identified parts of the GIT campus as buildings (Figure 1).

Adherence to the study protocol by participants in the study is unknown. Although the participants were asked to wear the sampling pack during all times when they were moving in the 48-hour period, it is possible that they did not completely adhere to this given the extra 3lbs weight. We also know that a person breathes at a higher rate when they are more active. Therefore, a person who is walking fast is likely to intake more $PM_{2.5}$ than another person who is idle in the same location and in the same hour. This difference will not be captured by the microPEM.

4.2 Future Work

Based on our results, we see that there is a significant difference in RatioPM within the different time periods within the day. The time periods were created based on general knowledge on when college students might be outside and when they would be active. As a next step, we would create separate models for each of the time periods and determine which variable is the most critical in predicting RatioPM. We can also build similar models using the absolute personal $PM_{2.5}$ measurement as the response variable. It would be interesting to observe if personal $PM_{2.5}$ had the same covariate as the RatioPM.

Moreover, the RF model appears to predict lower RatioPM in areas that are right adjacent to the highway. It is worth investigating why that might be since we would hypothesize that the RatioPM would gradually decrease as a person moved away from the highway. One possible explanation may be the sparse data points near on and very near the highway and future analysis may consider restricting the training dataset only to observations west of the highway.

Citations:

Brantley, H. L., Hagler, G. S. W., Kimbrough, E. S., Williams, R. W., Mukerjee, S., & Neas, L. M. (2014). Mobile air monitoring data-processing strategies and effects on spatial air pollution trends. *Atmospheric Measurement Techniques*, *7*(7), 2169–2183. https://doi.org/10.5194/amt-7-2169-2014

Cheng, Y., He, K. bin, Du, Z. Y., Zheng, M., Duan, F. K., & Ma, Y. L. (2015). Humidity plays an important role in the PM2.5 pollution in Beijing. *Environmental Pollution*, *197*, 68–75. https://doi.org/10.1016/j.envpol.2014.11.028

Guan, Y., Johnson, M. C., Katzfuss, M., Mannshardt, E., Messier, K. P., Reich, B. J., & Song, J. J. (2019). Fine-Scale Spatiotemporal Air Pollution Analysis Using Mobile Monitors on Google Street View Vehicles. *Journal of the American Statistical Association*, *1459*. https://doi.org/10.1080/01621459.2019.1665526

Liang D, Golan R, Moutinho JL, Chang HH, Greenwald R, Sarnat SE, Russell AG, Sarnat JA. Errors associated with the use of roadside monitoring in the estimation of acute traffic pollutant-related health effects. Environmental research. 2018 Aug 1;165:210-9.

Loomis, D., Grosse, Y., Lauby-Secretan, B., Ghissassi, F. el, Bouvard, V., Benbrahim-Tallaa, L., Guha, N., Baan, R., Mattock, H., & Straif, K. (2013). The carcinogenicity of outdoor air pollution. *The Lancet Oncology*, *14*(13), 1262–1263. https://doi.org/10.1016/S1470-2045(13)70487-X

Moltchanov, S., Levy, I., Etzion, Y., Lerner, U., Broday, D. M., & Fishbain, B. (2015). On the feasibility of measuring urban air pollution by wireless distributed sensor networks. *Science of the Total Environment*, *502*, 537–547. https://doi.org/10.1016/j.scitotenv.2014.09.059

Nathan, A. J., & Scobell, A. (2012). How China sees America. *Foreign Affairs*, *91*(5), 1–5. https://doi.org/10.1017/CBO9781107415324.004

RTI International. (2016). *MicroPEM^{TM}—PM2.5 Personal Exposure Monitor*. 1–2. https://tinyurl.com/y5bs32xk

Sarnat, J. A., Russell, A., Liang, D., Moutinho, J. L., Golan, R., Weber, R. J., Gao, D., Sarnat, S. E., Chang, H. H., Greenwald, R., & Yu, T. (2018). Developing Multipollutant Exposure Indicators of Traffic Pollution: The Dorm Room Inhalation to Vehicle Emissions (DRIVE) Study. *Research Report (Health Effects Institute)*, *April*(196), 3–75. https://doi.org/Submitted

Tiwary, A., Robins, A., Namdeo, A., & Bell, M. (2011). Air flow and concentration fields at urban road intersections for improved understanding of personal exposure. *Environment International*, *37*(5), 1005–1018. https://doi.org/10.1016/j.envint.2011.02.006

Yan, B., Pitiranggon, M., Ross, J., Engeland, T. A., Stelter, A., & Chillrud N, S. (2018). Validation of Miniaturized Particulate-Matter Real-Time Samplers for Characterizing Personal Polycyclic Aromatic Hydrocarbon Exposure. *Journal of Analytical & Bioanalytical Techniques*, *09*(02). https://doi.org/10.4172/2155-9872.1000403

Yi, W. Y., Lo, K. M., Mak, T., Leung, K. S., Leung, Y., & Meng, M. L. (2015). A survey of wireless sensor network based air pollution monitoring systems. In *Sensors (Switzerland)* (Vol. 15, Issue 12). https://doi.org/10.3390/s151229859