ANALYSIS OF AIRBORNE PARTICULATE MATTER CONCENTRATIONS DURING RAINSTORMS IN BOSTON, MASSACHUSETTS

by

Andrew Shapero

Senior Honors Thesis
Department of Civil and Environmental Engineering
Tufts University
Medford, Massachusetts
Dr. John Durant (Advisor)
Dr. Elena Naumova (Reader)

May 6, 2014

ABSTRACT

Even though there is precipitation in Boston an average of 127 days out of the year, there have been few efforts to measure particulate air pollution during rainstorms in Boston. During rainstorms, particle number concentration (PNC), which is an air pollutant associated with cardiovascular disease, and meteorological conditions both change. It is unclear, however, if the rainfall itself or the changes in meteorological conditions cause PNC to change. The objectives of this study were: (1) to quantify the changes in PNC, temperature, atmospheric pressure, relative humidity, wind speed, and wind direction during rainstorms; (2) to construct a multivariate linear regression model to determine how PNC is associated with rainstorm duration; and (3) to determine which meteorological conditions are associated with changes in PNC during rainstorms. The study was performed at the Massachusetts Department of Environmental Protection (MassDEP) Air Pollution Monitoring Site in Roxbury, Massachusetts. PNC and meteorological conditions (temperature, atmospheric pressure, relative humidity, wind speed, wind direction, and rainfall rate) were measured at five-minute intervals from September 9, 2013 through March 12, 2014. A multivariate linear regression model and a time series of rainstorms were constructed. Average PNC during rainy weather was measured to be 8% less than average PNC during dry weather ($p<0.001$). In general, PNC levels decreased throughout rainstorms at a rate of 15%/hour. A multivariate linear regression model estimated that rainfall was associated with PNC removal at an average rate of 12%/hour. A comparison of the simple linear regression model and the multivariate linear regression model indicated that the decrease in PNC associated with rainfall duration is associated with the rainfall itself and not other meteorological conditions.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Page

ABBREVIATIONS

CAFEH: Community Assessment of Freeway Exposure and Health

CPC: Condensation particle counter

gam: generalized additive model

lnPNC: The natural logarithm of particle number concentration

MassDEP: Massachusetts Department of Environmental Protection

MLR: Multivariate linear regression model for the natural logarithm of particle number concentration

nm: nanometers

PNC: Particle number concentration (particles/cm^3 of air)

SLR: Simple linear regression model for the natural logarithm of particle number concentration

UFP: Ultrafine particles

1.0 INTRODUCTION

1.1 Health Effects of Traffic-Related Air Pollution

Exposure to traffic-related air pollution is associated with pulmonary and cardiovascular disease. For example, lung disease is more prevalent in children who live near major highways than in children who do not live near major highways (Van Vliet et al., 1997), and traffic-related particulate matter has been shown to be associated with decreased heart-rate variability (Adar et al., 2007).

Highways are associated with many different air pollutants, but one class of pollutants of concern is particulate matter. Recently, researchers have begun to investigate the association between traffic-related particulate air pollution and cardiovascular disease (Brugge et al., 2007).

Particle number concentration (PNC) is a measure of airborne fine and ultrafine particles that have an aerodynamic diameter between 4 and 3,000 nanometers (nm).

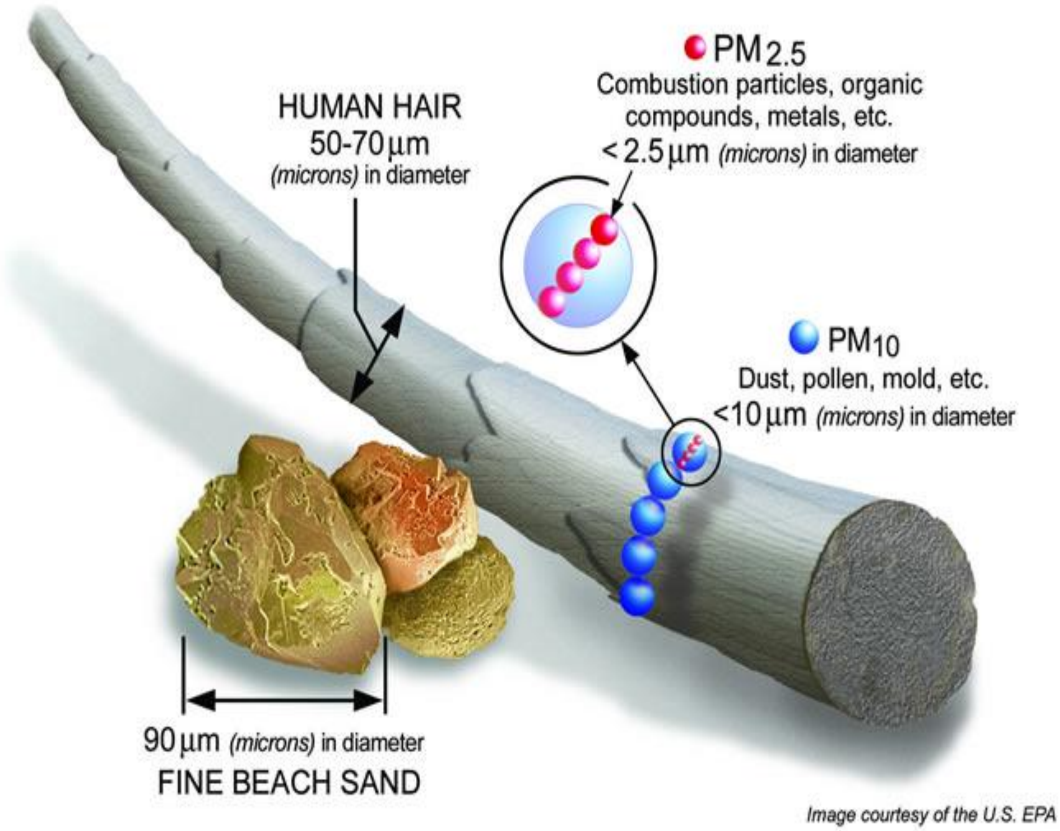

Figure 1: *Diagram of particulate matter sizes. The particles labeled* $PM_{2.5}$ *are 2,500 nm in diameter. They are roughly the same size as the largest particles counted in PNC measurements, which are 3,000 nm.*

Ultrafine particles (UFP) (aerodynamic diameter less than 100 nm) (Figure 2) are of particular concern because the particles are so small that they penetrate deep into the lungs and can enter the bloodstream (EPA, 2014).

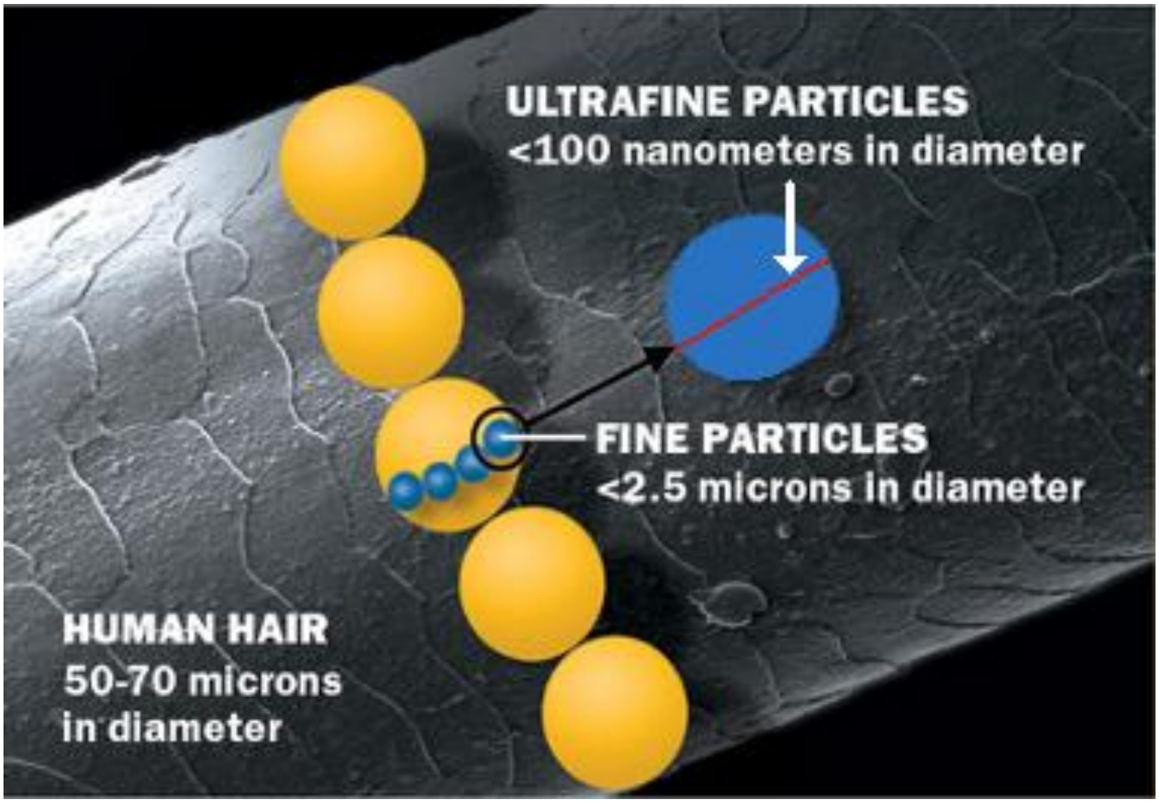

Figure 2: *Diagram of ultrafine particle size. Ultrafine particles make up a significant component of the particles counted in PNC measurements. Source: epa.gov.*

To establish the relationship between UFP and cardiovascular disease, accurate human exposure assessment of airborne particles is necessary (Padro-Martinez et al., 2012).

1.2 The CAFEH Study

In Boston, the Community Assessment of Freeway Exposure and Health (CAFEH) study is trying to associate particle number concentration (PNC) and cardiovascular disease through a multi-disciplinary approach. As part of the CAFEH study, ambient PNC measurements are being made in near-highway areas and blood samples from study participants are being tested for biomarkers that indicate inflammation and possible risk of cardiovascular disease.

The air pollution measurements are used to create models to estimate ambient PNC throughout the Boston area. By creating models of human exposure to PNC, the CAFEH researchers will determine if biomarkers that indicate inflammation are associated with elevated PNC (Brugge et al., 2013).

1.3 Previous Ambient PNC CAFEH Models

Before human exposure to PNC can be estimated, ambient PNC throughout the Boston area must be modeled. To do this, a multivariate linear regression model has been developed that uses temporal, spatial, and temporal-spatial variables to estimate PNC (Patton et al., 2014). The model estimates the natural logarithm of PNC (lnPNC) as a function of temperature, wind speed, wind direction, location relative to major highways, distance from highways and major roads, day of the week, and traffic volume.

One weakness of the ambient PNC model developed by Patton et al. (2014) is that data collection took place during very few rainy days. Therefore, the model does not consider the effects of rainfall on PNC; however, the effect of rain on pollutant concentrations should not be ignored. There is precipitation an average of 127 days in Boston each year (climate-zone.com, 2014). Because data collection did not take place during many rainstorms, the ambient PNC model ignored a potentially significant aspect of Boston's climate.

If rain does in fact affect ambient PNC, then it may also affect human exposure to the pollutant. The addition of rainfall into the models could provide a more accurate estimation of human exposure to PNC and would therefore provide a more accurate quantification of the association between PNC and cardiovascular disease. This thesis addresses the potential effects of rainfall on PNC in the Boston area.

1.4 Literature Review of Rainfall and Particulate Air Pollution

It is well understood that precipitation can remove particles from the atmosphere; therefore, the scavenging of airborne particles must be addressed in air quality models (Seinfeld and Pandis, 1998). As raindrops fall downward towards the ground, they collide with airborne particles and deposit the particles on the ground in a process known as washout or wet deposition (Nieto et al., 1994; Jacob, 1999; Andronache et al., 2006).

The removal of particles during rainfall is evidenced by the sky appearing to be clear after a rainstorm. As particulate concentrations decrease, visibility increases because of the decrease in the scattering and absorption of light by the particles. The particles counted in the PNC range are actually expected to reduce visibility more than are particles of other sizes. Particles that range from 100 to 1,000 nm are most effective at scattering light because particles of this size range are about the same size as the wavelengths of light (Cooper and Alley, 2011). The reduction in particle concentrations during rainstorms is visible, but it has not been quantified in a way that is useful for ambient PNC modeling.

Equations exist for the theoretical removal of particles during rainfall if the particle and raindrop sizes are known (Andronache et al., 2006). But PNC is a measurement of a range of particles sizes; therefore, theoretical particle removal equations based on discrete particle sizes cannot be used. Even though theoretical particle removal equations cannot be applied to the models, such equations can still help the understanding of how PNC might be affected by rainfall.

Theoretical removal equations indicate that particles of different sizes are removed with different efficiencies. In general, small particles are removed from the atmosphere less efficiently than are large particles (Nieto et al. 1994). The fluid dynamics process of impaction may be the reason that large particles are scavenged more efficiently than small particles. From the frame of reference of the falling raindrop, particles follow a streamline around the raindrop. Small particles remain in the streamlines and do not collide with the raindrop. The inertia of the larger particles causes them to collide with the water droplet, and the larger particles are consequently removed from the atmosphere as the raindrop falls to the ground (Cooper and Alley 2011; Duhanyan 2011).

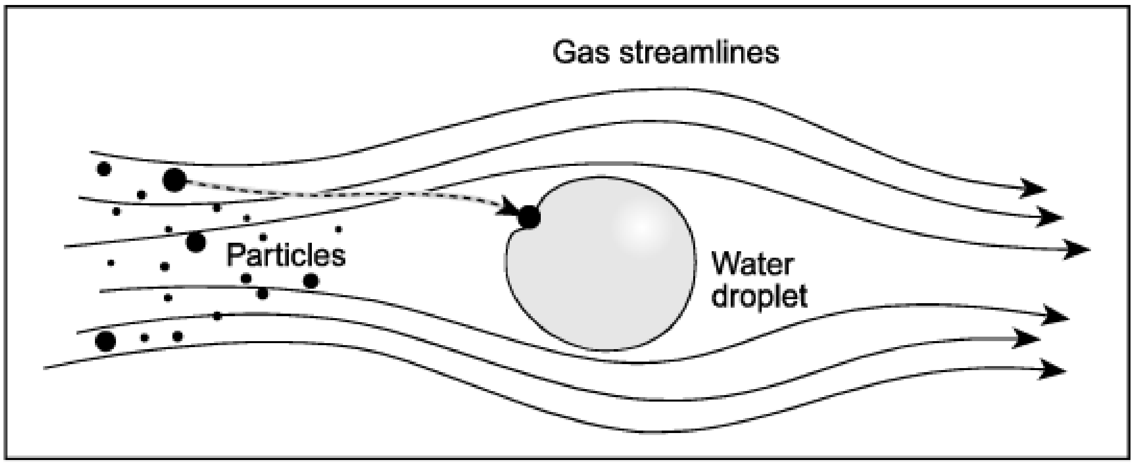

Figure 3: *Diagram of the physical process of inertial impaction. Small airborne particles may not be removed during rainfall because their mass is so small that they do not collide with the raindrops. Source: U.S Environmental Protection Agency in collaboration with North Carolina State University (NCSU).*

Nieto et al. has found that coarse particle modes (with an average diameter of about 6,000 nm) are removed very effectively by rainfall. But finer particle modes (with an average diameter of about 300 nm) are not significantly affected by raindrop scavenging. Based on the Nieto et al. results, some of the particles in the PNC range (4 to 3,000 nm) should be removed, but other particles in the PNC range should not be removed.

Even though theoretical removal rates cannot be applied to the models, basic theory could be incorporated into a multivariate linear regression model that is similar to the existing ambient PNC models. Regardless of what the removal rate is, the literature suggests that the decrease in particulate concentrations should follow an exponential decay function, such that

$$PNC_t = PNC_0 e^{-\lambda t} \qquad \text{Equation 1}$$

(Duhanyan et al., 2011). Lambda (λ) is the scavenging coefficient, the variable PNC_0 is the initial PNC at the beginning of a rainstorm, and PNC_t is PNC after a storm duration of time t. Duhanyan et al. (2011) provides summary tables of λ values for different pollutants. The scavenging coefficient is likely to be dependent on a variety of factors, including rainfall rate. Rainfall rate is believed to be positively correlated with scavenging coefficients; that is, as rainfall rate increases, pollutant removal increases. The effect of rainfall rate on a scavenging coefficient will not be discussed in this paper, but the effects of rainfall rate on PNC will be incorporated into the analysis.

Equation 1 describes how rainfall removes PNC, but it does not take into account meteorological conditions that might be changing during a rainstorm. For example, meteorological conditions, such as wind speed, for example, could be changing significantly during rainstorms. Previous studies have shown that particulate pollution concentrations can increase during rainfall because rainfall is not the only factor acting on particle concentrations during a rainstorm (Andronache 2006).

The changes in meteorological conditions during rainstorms can be modeled by simple linear regression models, such that

$$x_i = \beta_{s,i} t + \varepsilon, \qquad \text{Equation 2}$$

where the meteorological conditions x_i were modeled as functions of the coefficients $\beta_{simple,i}$, storm duration t, and an error term ε. Herein, I will refer to $\beta_{simple,i}$ as $\beta_{s,i}$.

Because many factors might affect PNC during a rainstorm, a multivariate linear regression model can be used so that conditions can be held constant as the association between PNC and rainfall is calculated. The multivariate linear regression model can incorporate a scavenging coefficient and can estimate lnPNC as a function of meteorological conditions, temporal variables, and storm duration, such that

$$lnPNC = \beta_0 + \beta_{multivariate,1}x_1 + \beta_{multivariate,2}x_2 + \ \dots + \beta_{multivariate,k}x_k + \lambda t + \varepsilon,$$

Equation 3

where β_0 is the intercept of the model, x_i are the meteorological and temporal variables, $\beta_{multivariate,i}$ are the coefficients for the meteorological and temporal variables, and ε is the error term. Herein, I refer to $\beta_{multivariate,i}$ as $\beta_{m,i}$.

As far as I know, no other study has incorporated a scavenging coefficient into a multivariate linear regression model. Additionally, the literature does not explain how changes in temperature, atmospheric pressure, relative humidity, wind speed, and wind direction during rainstorms might affect particulate concentrations.

1.6 Incorporation of Rainfall into Linear Regression

Because rainfall might affect the association between PNC and cardiovascular disease, rainfall should be incorporated into a model that is similar to the existing CAFEH ambient PNC model. The existing CAFEH ambient PNC model uses multivariate linear regression to estimate PNC as a function of temporal, spatial, and temporal-spatial variables (Patton et al., 2014). If the proper rainfall data were collected, a similar multivariate linear regression model could be constructed. The new model could incorporate a scavenging coefficient into a multivariate linear regression model to estimate lnPNC.

Manipulation of Equation 2 reveals that

$$lnPNC_t = lnPNC_0 - \lambda t. \qquad \text{Equation 4}$$

Holding all other variables constant, lnPNC will decrease at a rate of λ/minute during a rainfall event. If the change in PNC associated with rainstorm duration (while holding all other variables constant) can be determined, then a scavenging coefficient will have been calculated.

2.0 OBJECTIVES, HYPOTHESES, AND PRELIMINARY RESEARCH

2.1 Objectives

The objectives of this study were: (1) to quantify the changes in PNC, temperature, atmospheric pressure, relative humidity, wind speed, and wind direction during rainstorms; (2) to construct a multivariate linear regression model to determine how PNC is associated with rainstorm duration; and (3) to determine which meteorological conditions are associated with changes in PNC during rainstorms.

The models produced as a part of this study are not intended to replace the existing CAFEH model. The existing CAFEH model considers spatial variables, not just meteorological and temporal variables, because it is used to estimate human exposure to PNC in various locations throughout the Boston area. The CAFEH model is also meant to apply for the entire Boston area, but my model is only meant to apply to a specific location in Boston. The estimation of human exposure at one specific location is not of much interest. But one of the outcomes of this study, a wet scavenging coefficient for PNC in the Boston area, is of interest. Assuming that concentrations, particle size distribution, and raindrop characteristics at the site of my experiments are representative of the Boston area, then the calculated scavenging coefficient could be applied to the existing CAFEH models.

To meet the objectives of the study, the hypotheses in the following section were tested. Hypotheses were made based on the existing CAFEH models, the existing literature, and my own preliminary research. The preliminary research only considers data points that were recorded during conditions during which rainfall was possible. Data points for which temperature was below freezing and for which relative humidity was below 65% were not considered (see Section 3.0 for more details).

2.2 Hypotheses and Preliminary Research

2.2.1 Temperature hypotheses and preliminary research

Based on the previous CAFEH model, PNC is inversely related to temperature (Patton et al., 2014). Likewise, other studies have also found that decreased temperature is associated with increased particle concentrations (von Bismarck-Osten et al., 2013). Although some of the increase in particle concentrations during cold weather may be due to increased use of heating systems and thus increased emissions (e.g., von Bismarck-Osten et al., 2013), colder weather is conducive to aerosol condensation and new particle formation (Olivares et al., 2007). Particle nucleation theory indicates that the rate of particle formation is inversely related to temperature (Jacobson 1999). Therefore, particle concentrations can be expected to be greater during cold weather than during warm weather because of both particle formation rates and emission rates.

Based on my own preliminary research, rainy weather is associated with higher average temperature than dry weather ($p<0.001$) (Figure 4).

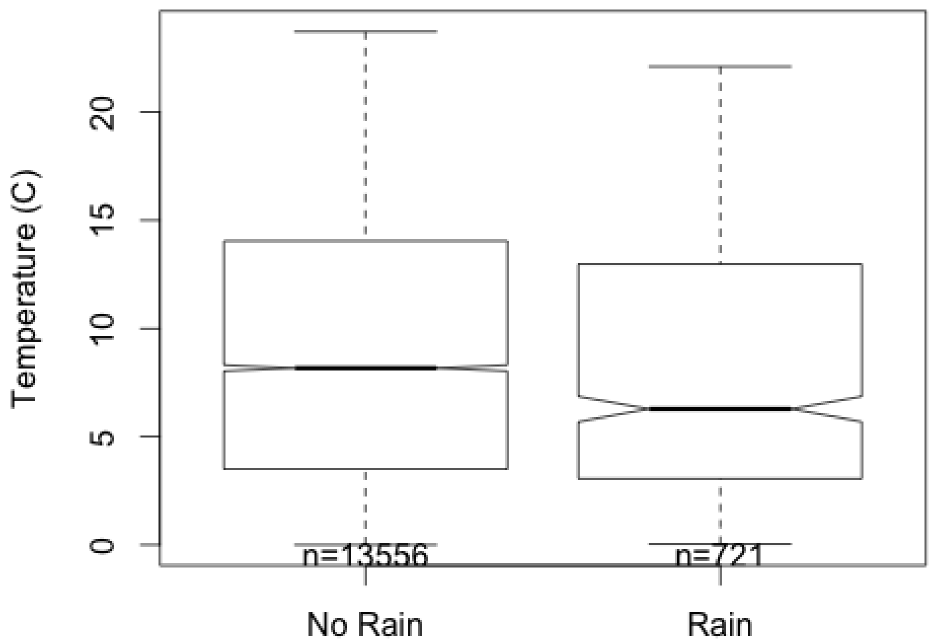

Figure 4: *Boxplots of air temperature during periods of rain (> 0 mm/hour) and no rain between September 2013 and March 2014 at Roxbury, Massachusetts. Data points were limited to conditions with temperatures above freezing and above 65% relative humidity. Within the box represents 25% to 75% (or the interquartile) of the data points. Methods for collection and processing of the data used for preliminary analyses can be found in Section 3.0. n=number of five-minute data points.*

Based on the existing CAFEH model, a literature review, and my own preliminary research, the hypotheses for temperature are:

- H1: $\beta_{m, temperature} < 0$. In the multivariate regression model, the coefficient for temperature is less than zero.
- H2: $\beta_{s, temperature} < 0$. In a simple linear regression model, temperature decreases throughout the duration of rainstorms.

If both hypotheses cannot be rejected, then the changes in temperature associated with rainstorms can be expected to be associated with increasing PNC. Changes in temperature during rainstorms, therefore, could mask the potential decrease in PNC during rainfall.

2.2.2 Atmospheric pressure hypotheses and preliminary research

While the changes in temperature during rainstorms are expected to be associated with changes in lnPNC, atmospheric pressure, on the other hand, is not expected to significantly affect lnPNC. Based on the CAFEH models, atmospheric pressure will not have a significant effect on PNC (Patton et al., 2014). Similarly, in the models done by von Bismarck-Osten et al. (2013), atmospheric pressure covaries with other parameters and is not significantly correlated with particle concentrations.

Based on my preliminary research, atmospheric pressure is generally lower during rainy weather than during dry weather ($p<0.001$) (Figure 5). Basic meteorology principles also indicate that rainy weather is associated with lower average atmospheric pressure than dry weather (Ahrens 2008).

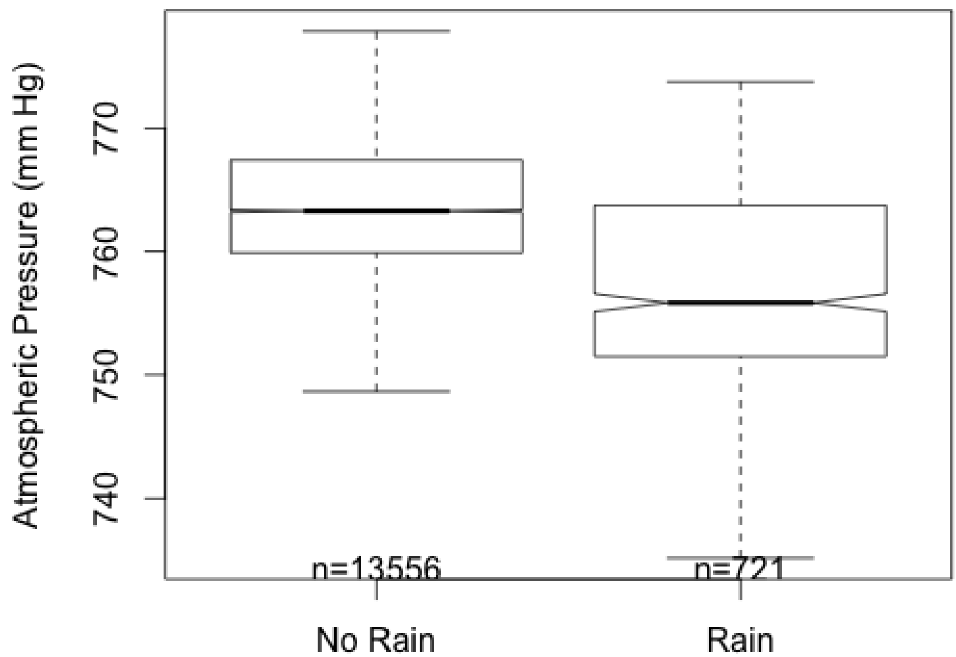

Figure 5: *Boxplots of atmospheric pressure during periods of rain (> 0 mm/hour) and no rain between September 2013 and March 2014 at Roxbury, Massachusetts. Data points were limited to conditions with temperatures above freezing and above 65% relative humidity. Within the box represents 25% to 75% (or the interquartile) of the data points. Methods for collection and processing of the data used for preliminary analyses can be found in Section 3.0. n=number of five-minute data points.*

Based on the existing CAFEH model, a literature review, and my own preliminary research, the hypotheses for atmospheric pressure are:

- H3: $\beta_{m,\ pressure} = 0$. In the multivariate linear regression model, the coefficient for atmospheric pressure is not different from zero.
- H4: $\beta_{s,\ pressure} < 0$. In a simple linear regression model, atmospheric pressure will decrease throughout the duration of a rainstorm.

If hypothesis H3 cannot be rejected, then the changes in atmospheric pressure during rainstorms, regardless of what those changes are, will not be expected to be associated with significant changes in PNC. Based on the hypotheses, changes in atmospheric pressure during rainstorms are not expected to mask or enhance the potential decrease in PNC during rainfall.

2.2.3 Relative humidity hypotheses and preliminary research

Similar to atmospheric pressure, relative humidity is also not expected to significantly affect lnPNC during rainstorms. Based on the CAFEH model, relative humidity will not have a significant effect on PNC (Patton et al., 2014). Von Bismarck-Osten et al. (2013) also found that relative humidity covaries with other parameters in particle concentration models but is not significantly correlated with particle concentrations.

Based on my preliminary research (and as is commonly accepted), average relative humidity is generally higher during rainy weather than during dry weather ($p<0.001$) (Figure 6).

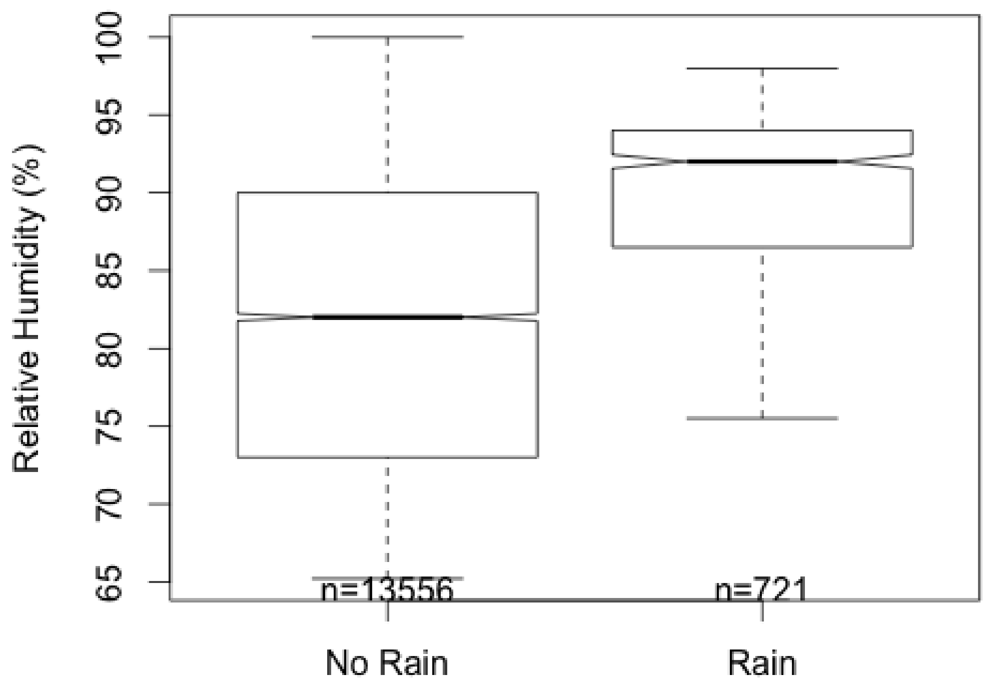

Figure 6: *Boxplots of relative humidity during periods of rain (> 0 mm/hour) and no rain between September 2013 and March 2014 at Roxbury, Massachusetts. Data points were limited to conditions with temperatures above freezing and above 65% relative humidity. Within the box represents 25% to 75% (or the interquartile) of the data points. Methods for collection and processing of the data used for preliminary analyses can be found in Section 3.0. n=number of five-minute data points.*

Based on the existing CAFEH models, a literature review, and my own preliminary research, the hypotheses for relative humidity are:

- H5: $\beta_{m,\ humidity} = 0$. In the multivariate linear regression model, the coefficient for relative humidity is not different from zero.

- H6: $\beta_{s,\ humidity} > 0$. In a simple linear regression model, relative humidity will increase throughout the duration of a rainstorm.

If hypothesis H5 cannot be rejected, then the changes in relative humidity during rainstorms, regardless of what those changes are, will not be expected to be associated with significant changes in PNC. Based on the hypotheses, the changes in relative humidity during rainstorms are not expected to mask or enhance the potential effects of rainfall on PNC.

2.2.4 Wind speed hypotheses and preliminary research

While changes in atmospheric pressure and relative humidity were not expected to have significant effects on lnPNC, increases in wind speed are commonly accepted to cause decreases in atmospheric pollutant concentrations. Increased wind speeds are conducive to increased mixing and dilution and therefore lower pollutant concentrations (Ahrens 2008). Based on the previous CAFEH model, wind speed is inversely related to PNC (Patton et al., 2014). And from my preliminary research, rainy weather is associated with significantly higher wind speeds than dry weather (Figure 7).

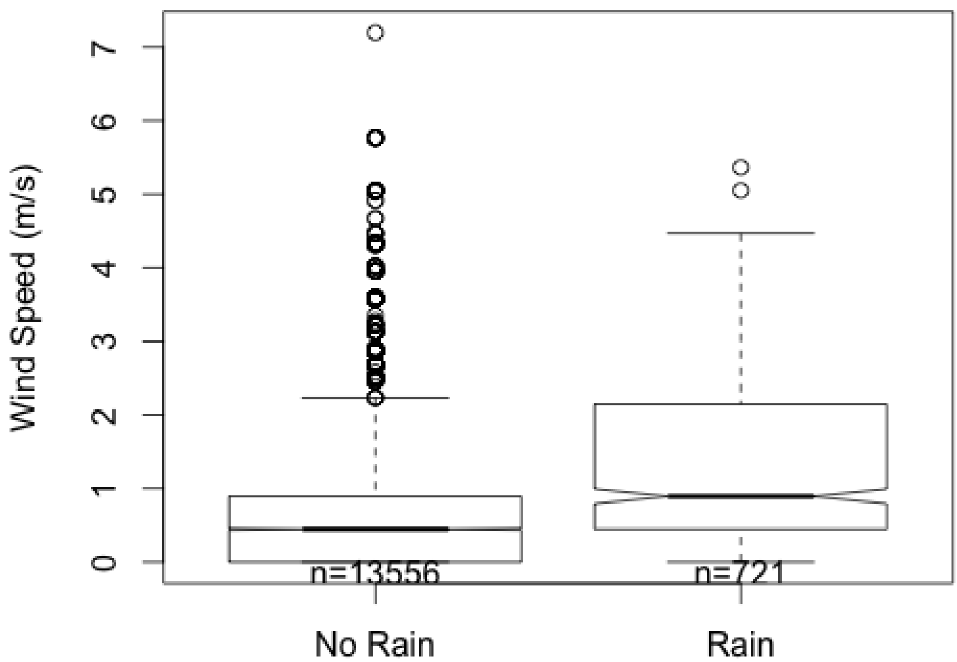

Figure 7: *Boxplots of wind speed during periods of rain (> 0 mm/hour) and no rain between September 2013 and March 2014 at Roxbury, Massachusetts. Data points were limited to conditions with temperatures above freezing and above 65% relative humidity. Within the box represents 25% to 75% (or the interquartile) of the data points. Methods for collection and processing of the data used for preliminary analyses can be found in Section 3.0. n=number of five-minute data points.*

Based on the existing CAFEH model, a literature review, and my own preliminary research, the hypotheses for wind speed are:

- H7: $\beta_{m,\ wind\ speed} < 0$. In the multivariate linear regression model, the coefficient for wind speed is less than zero.
- H8: $\beta_{s,\ wind\ speed} > 0$. In a simple linear regression model, wind speed will increase throughout the duration of a rainstorm.

If both of the hypotheses cannot be rejected, then the changes in wind speed during a rainstorm can be expected to be associated with a decrease in lnPNC. Changes in wind speed during rainstorms, therefore, could enhance the potential decrease in PNC due to rainfall.

2.2.5 Wind direction hypotheses and preliminary research

The CAFEH model indicates that changes in wind direction are associated with significant changes in PNC (Patton et al., 2014). Patton et al. (2014) transformed wind direction by using trigonometric functions so that wind direction fit into the multivariate linear regression model, and a similar procedure was followed for this study (see Section 3.3.3).

Different wind directions are expected to be associated with different PNC because wind direction determines the source of air pollution. For example, PNC will probably be greater if the sampling location is downwind of a major source, such as a highway, than if the sampling location is upwind of a major source. The sampling site was west of highway I-93 (see Figure 9), so winds from the east could be expected to be associated with higher PNC than winds from other directions.

It is unclear how wind direction might change during a storm, but for the sake of having a hypothesis to test, it will be assumed that wind direction will change significantly during rainstorms.

Based on the existing CAFEH model, a literature review, and my own preliminary research, the hypotheses for wind speed are:

- H9: $\beta_{\text{m, sine(wind direction)}} \neq 0$, $\beta_{\text{multivariate, cosine(wind direction)}} \neq 0$. In the multivariate linear regression model, the coefficients for wind direction are significantly different from zero.

- H10: $\beta_{s,\ sine(wind\ direction)} \neq 0$, $\beta_{simple,\ cosine(wind\ direction)} \neq 0$. Wind direction will change significantly throughout the duration of a rainstorm.

If the hypotheses cannot be rejected, then the changes in wind direction during rainstorms can be expected to be associated with changes in lnPNC.

2.2.6 Particle number concentration hypotheses and preliminary research

Based on a literature review of the effects of rainfall on particulate air pollution concentrations, PNC is expected to decrease throughout the duration of a storm. My preliminary research also indicates that rainy weather is associated with lower PNC than dry weather ($p<0.001$) (Figure 8).

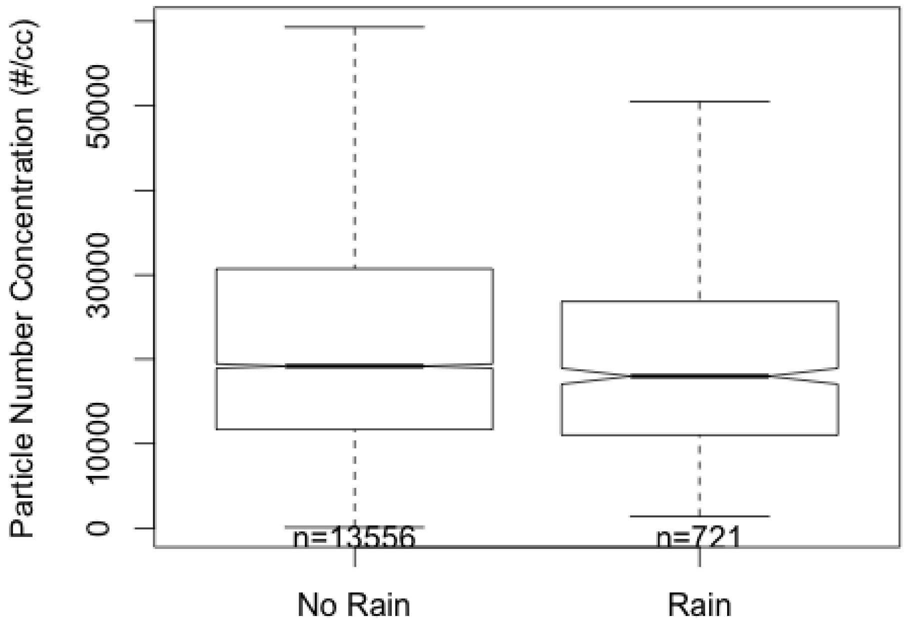

Figure 8: *Boxplots of PNC during periods of rain (> 0 mm/hour) and no rain between September 2013 and March 2014 at Roxbury, Massachusetts. Data points were limited to conditions with temperatures above freezing and above 65% relative humidity. Within the box represents 25% to 75% (or the interquartile) of the data points. Methods for collection and processing of the data used for preliminary analyses can be found in Section 3.0. n=number of five-minute data points.*

The difference in lnPNC during rainy weather is expected to be due to the effects of rainfall and possibly because of the effects of meteorological changes during rainstorms.

Therefore, the hypotheses for lnPNC are:

- H11: $\beta_{m,\ storm\ duration} < 0$. In the multivariate linear regression model, the coefficient for storm duration is less than zero.

- H12: $\beta_{s,\ storm\ duration} < 0$. In a simple linear regression model, the natural log of PNC will decrease throughout the duration of a rainstorm.
- H13: $\beta_{m,\ storm\ duration} \neq \beta_{s,\ storm\ duration.}$ The coefficient for storm duration in the multivariate linear regression models will be significantly different from the coefficient for storm duration in the simple linear regression model.

If hypothesis H11 cannot be rejected, then rainfall is associated with significant removal of PNC. If hypothesis H12 cannot be rejected, then lnPNC is decreasing significantly throughout rainstorms. And if hypothesis H13 cannot be rejected, then meteorological conditions are associated with significant changes in lnPNC during rainstorms.

It is unclear how the changes in meteorological conditions during rainstorms will affect PNC. Some conditions (wind speed, for example) are expected to be associated with decreasing lnPNC, but others (temperature) are expected to be associated with increasing lnPNC during rainstorms. It should be noted that a significant change in a meteorological condition is not necessarily associated with a significant change in PNC. For example, if the simple linear regression model indicates that wind speed increases significantly throughout the duration of a rainstorm, wind speed may not have actually changed significantly five minutes into the storm. It is during longer rainstorms that meteorological conditions may have actual significant effects on PNC.

Changing meteorological conditions during rainstorms could possibly be enhancing or masking the effect of rainfall on lnPNC. Multivariate linear regression will need to be used to determine if the change in lnPNC associated with rainfall is associated with the rainfall itself or with other meteorological conditions.

3.0 METHODS

3.1 Data Collection

Because rainstorms generally occur within a time frame of a few hours, hourly aggregated data might not depict changes in PNC and meteorological conditions during a rainstorm with enough precision. Therefore, I chose to collect data with a temporal resolution of five minutes or less.

A condensation particle counter (TSI model 3873 Environmental Particle Counter) (CPC) and a Davis weather station were installed at the Massachusetts Department of Environmental Protection (MassDEP) fixed monitoring site in Roxbury, Massachusetts (Figure 9). The CPC was set so that it recorded ambient PNC measurements every thirty seconds. The weather station was set so that it recorded temperature, atmospheric pressure, relative humidity, wind speed, wind direction, rainfall rate, and other meteorological variables every five minutes. A finer temporal resolution would have been preferable, but data storage was an issue.

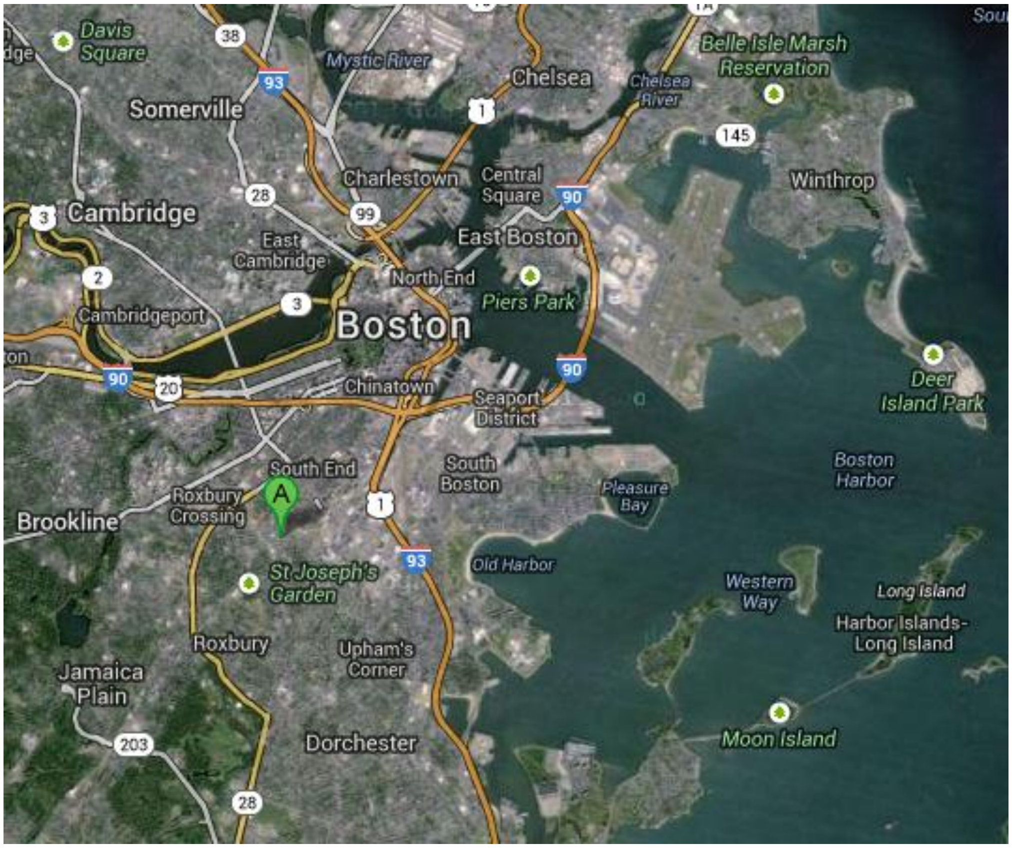

Figure 9: *Location of the Roxbury fixed site within the Boston area. Source: Google Maps.*

Data collection began on September 9, 2013, and data were collected through March 12, 2014. Throughout this time period, approximately 20% of all possible data points were not collected because of inconsistencies in traveling to the sampling site for data collection.

Data collection occurred every week. Data collection from the CPC needed to occur every two weeks, but data collection from the weather station needed to occur once a week. The weather station memory could only hold eight days worth of data. There were a few instances where the weather station data was not collected in time, and therefore, some of the data was lost.

Data from the CPC was collected by using a flash drive. When the PNC data was moved from the flash drive to a laptop computer, the data on the flash drive needed to be deleted to make room for new data.

The CPC needed to be replaced every six weeks because the wick in the CPC is not usable after six weeks. Also, deionized water needed to be added to the CPC each week.

Data from the weather station was collected by using a USB link cable. When the weather data was transferred from the weather station to a laptop computer, the data on the weather station was deleted. The data on the weather station did not necessarily need to be deleted because the weather station simply rewrites over the last data point in the weather station memory. However, the weather station data was deleted after it was transferred to a laptop computer so that the data could not be downloaded again. Data that was not deleted could potentially be downloaded multiple times and would then create duplicates of data points.

3.2 Quality Control

There were certain data points that needed to be removed from the dataset. The CPC records errors if there is an issue with CPC when data was recorded. All data points with recorded errors were deleted from the dataset. An R script provided by Allison Patton was used to remove data points with errors.

The analyses for the boxplots and time series were also limited to conditions for which rainfall was possible; therefore, data points for which temperature was below freezing (0°C) were excluded from the dataset. Data points were also excluded if relative humidity was below 65%, the lowest humidity in the dataset when rainfall was observed.

The data used to make the regression models was limited even further. Only data points associated with rainfall were used for these analyses. Only data points that occurred directly before rainfall or during rainfall were used to make the regression models.

3.3 Data Analysis

3.3.1 Resolving temporal resolution issues

Once the data were collected, the statistical software R was used to process and analyze the data. The software R was used to merge the two time series—the PNC time series and the weather time series. Because the PNC dataset and the weather dataset did not have the same temporal resolution, the data were aggregated by the largest time step—five minutes.

For each hour, 12 five-minute time intervals were created so that the PNC data and the weather data could be paired up. Five-minute intervals occur on the hour, five minutes past the hour, ten minutes past the hour, and so on. The weather data was aggregated so that each five-minute interval represented the five minutes directly before that time. For example, a data point that was recorded by the weather station at 1:02 AM was placed in the data bin that was stamped "1:05 AM" and fell into the first five-minute interval of the hour; similarly, a data point that was recorded at 1:33 AM was placed into the data bin that was stamped "1:35 AM" and fell into the seventh five-minute interval of the hour.

The PNC data was also aggregated so that each five-minute interval represented the five-minute interval directly before that time. However, because there were multiple data points in each five-minute interval, the data were aggregated by the median lnPNC value. There are ten lnPNC data points within each five-minute interval, and the median of those ten lnPNC data points represented the entire five-minute interval. Once the temporal resolution of the two datasets was uniform, the datasets were merged into one dataset that contained PNC and meteorological data for each five-minute interval.

3.3.2 Transforming the wind direction variable

The wind direction variable also needed to be transformed because lnPNC is not expected to vary linearly with wind speed. Instead, lnPNC is expected to vary cyclically with wind direction. Therefore, a sine function and a cosine function were used to describe the variations in wind direction. Just as the hour and month variables were transformed, the wind direction degrees variable was converted to two variables—sine(wind direction degrees$\times \frac{\pi}{180}$) and cosine(wind direction degrees$\times \frac{\pi}{180}$). Within the trigonometric functions, a multiplicative factor of $\frac{\pi}{180}$ was used to convert the wind direction degrees to radians.

3.3.3 Creating the binary rain variable

While some variables had to be transformed, others had to be created from the existing dataset. Once the data were aggregated into a time series of lnPNC, temperature, atmospheric pressure, relative humidity, wind speed, wind direction, and rainfall rate, a binary (dummy) variable for rainfall was created from the rainfall rate variable. Data points were either dry weather data points or rainy weather data points. All data points with a rainfall rate of greater than zero were considered to be data points associated with rain; all data points with a rainfall rate equal to zero were considered to be data points associated with dry weather.

3.3.4 Creating the rainfall lag variables

Because duration of rainfall, and not just whether or not it is raining, is an important factor in determining the effect of rainfall on PNC (see Equation 1), a rainfall duration variable was produced to include in the models. To create the rainfall duration variable, a series of rainfall lag variables were created within the dataset. Because they are created from the rainfall variable, the rainfall lag variables are also binary variables. The first rainfall lag variable simply indicates whether or not the previous data point in the time series was a dry weather data point or a rainy weather data point. In other words, the first rainfall lag variable identifies whether or not there was rainfall in the previous five-minute interval. The second rainfall lag variable identifies whether or not there was rainfall two five-minute intervals ago. Many rainfall lag variables were created for each data point so that it was possible to recognize long consecutive periods of rainfall or dry weather.

3.3.5 Creating the rainfall duration variable

Once the rainfall lag variables were created, the storm duration variable was created. If a data point was a dry weather data point, then the storm duration was set equal to zero. If a data point was a wet weather data point, then the storm duration was set equal to five minutes. If a data point was a wet weather data point *and* the first rainfall lag variable indicated rainfall, then the storm duration was updated and set equal to ten minutes. Otherwise, the storm duration variable was set equal to the existing storm duration. This process was completed until the storm duration variable reached a value of 120 minutes. Essentially, the code for the analysis performed a series of repeated tests to determine if the previous five-minute intervals had any rainfall.

There was little data for storms of duration greater than two hours. I did not want the model to apply to durations of greater than two hours because I did not want to assume the behavior of lnPNC when there was so little data. Therefore, the rainfall duration variable was limited to 120 minutes.

3.4 Linear Regression

3.4.1 Simple linear regression

Once all of the necessary variables were properly created and transformed, the linear regression process could begin. To test the hypotheses in Section 2, the trends in meteorological conditions and lnPNC during rainfall periods were determined using simple linear regression models (Equation 2).

Simple linear regression models were built for temperature, atmospheric pressure, humidity, wind speed, and both wind direction functions. A simple linear regression model for lnPNC as a function of storm duration was also built.

The simple linear regression models were built so that they could be used in conjunction with the multivariate linear regression models to ultimately demonstrate how meteorological conditions are associated with PNC during a rainstorm. The multiplication of the sign of the coefficient from the simple linear regression and the sign of the coefficient from the multivariate linear regressions indicate how the changes in meteorological conditions are expected to affect PNC during rain events.

3.4.2 Multivariate linear regression

A multivariate linear regression model (MLR) was also created. The MLR estimated lnPNC as a function of temperature, atmospheric pressure, humidity, wind speed, wind direction, day of the week, hour of the day, month of the year, rainfall rate, and rainstorm duration.

4.0 RESULTS

4.1 Summary of Data

Overall, there were over 39,000 collected data points that contained information on lnPNC, temperature, atmospheric pressure, humidity, wind speed, wind direction, rainfall rate, and rainfall duration. Of those data points, about 14,000 had both a temperature above freezing and a relative humidity above 65%, and of those, about 720 (approximately 5%) were associated with rainfall. Because each data point represents five minutes, the dataset indicates that there were about 60 hours of rainfall within the study period (September 9, 2013 to March 12, 2014). Five percent of the data points used for the boxplot and time series analyses were associated with rainfall, but only about 1% of all the collected data points were associated with rainfall. In other words, from September 9, 2013 to March 12, 2014, it was raining only 1% of the time. However, 5% of the data points selected for the analysis were associated with rainfall.

Even though rainfall occurred in only 1% of the data points, there was still rainfall very frequently. In Boston, there is rainfall approximately every three days (Figure 10).

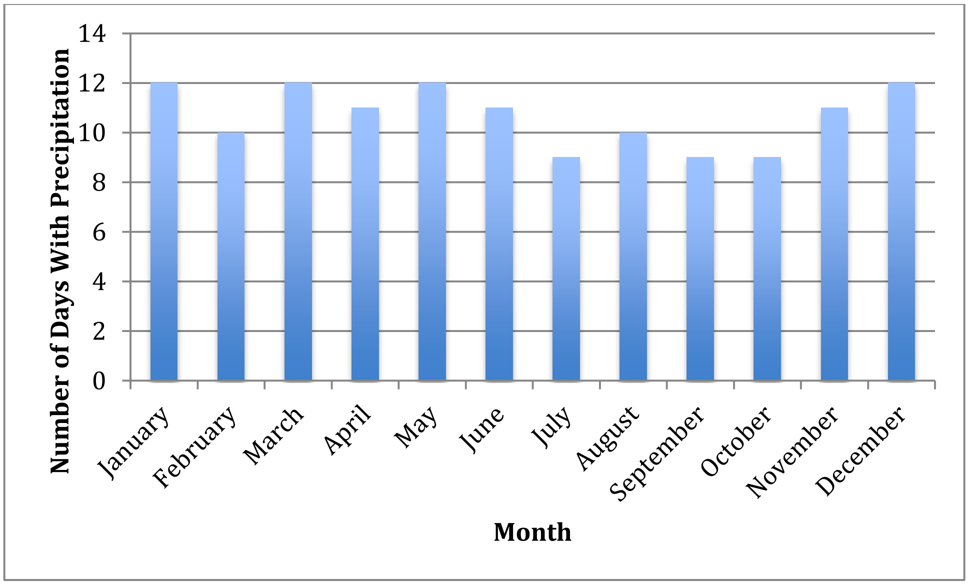

Figure 10: *Number of days with precipitation for each month in Boston, Massachusetts. Source: climate-zone.com.*

Thus, even though there is precipitation on many days within the study period, there is little rainfall on each individual day.

While 14,000 data points were used to produce the boxplots and the time series, only about 800 points were used to make the linear regression models. As stated above, there were only 720 data points associated with rainfall. These 720 data points make up about 80 rainfall events. The dry weather data points that were recorded five-minutes before the beginning of each rainfall event were included in the dataset used for linear regression. Therefore, about 800 data points were used to construct the linear regression models.

Because the analysis is so focused on changes in PNC and meteorology over time, it is helpful to see the overall time series for PNC and meteorological conditions. Figure 11 shows median lnPNC measurements aggregated throughout the five-minute intervals.

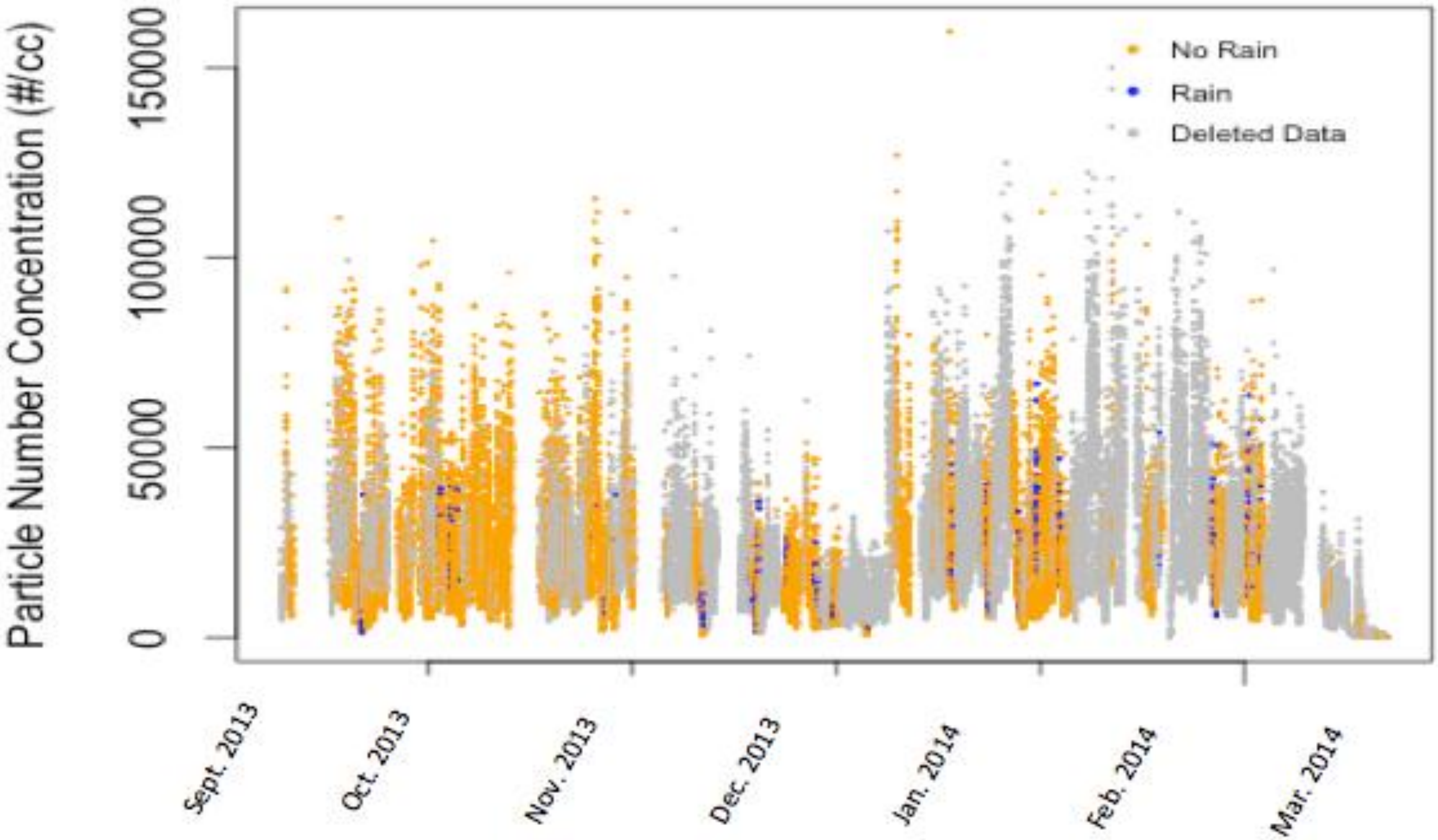

Figure 11: *Time series of five-minute lnPNC from September 2013 to March 2014. Data points were deleted from the dataset either because the air temperature was less than 0°C or because relative humidity was less than 65%. Blank spaces represent times for which data points were lost during the data collection process.*

Figure 12 shows the time series of temperature during the study period.

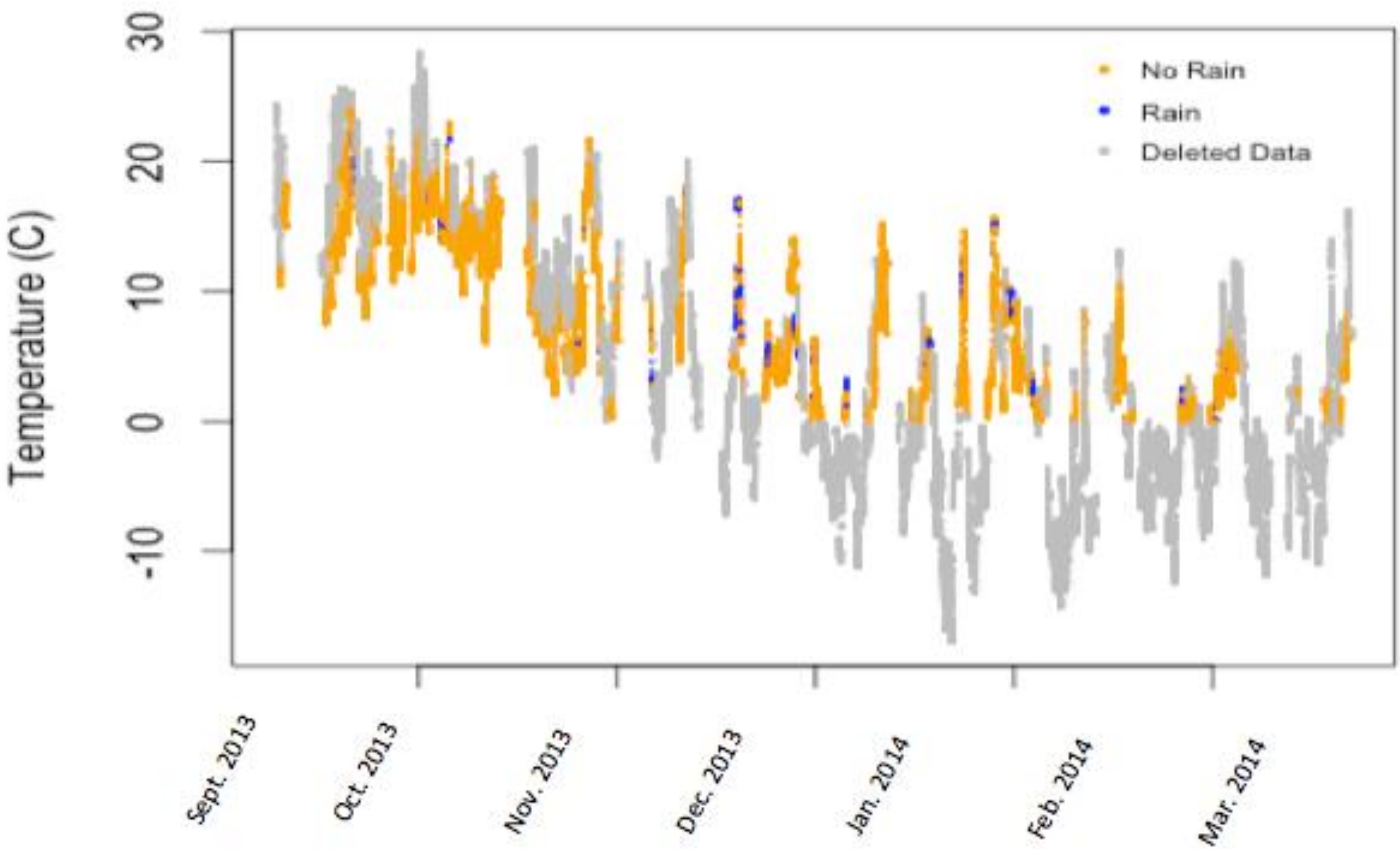

Figure 12: *Time series of temperature from September 2013 to March 2014. Data points were deleted from the dataset either because the air temperature was less than 0°C or because relative humidity was less than 65%. Blank spaces represent times for which data points were lost during the data collection process.*

Figure 13 shows the time series of atmospheric pressure during the study period.

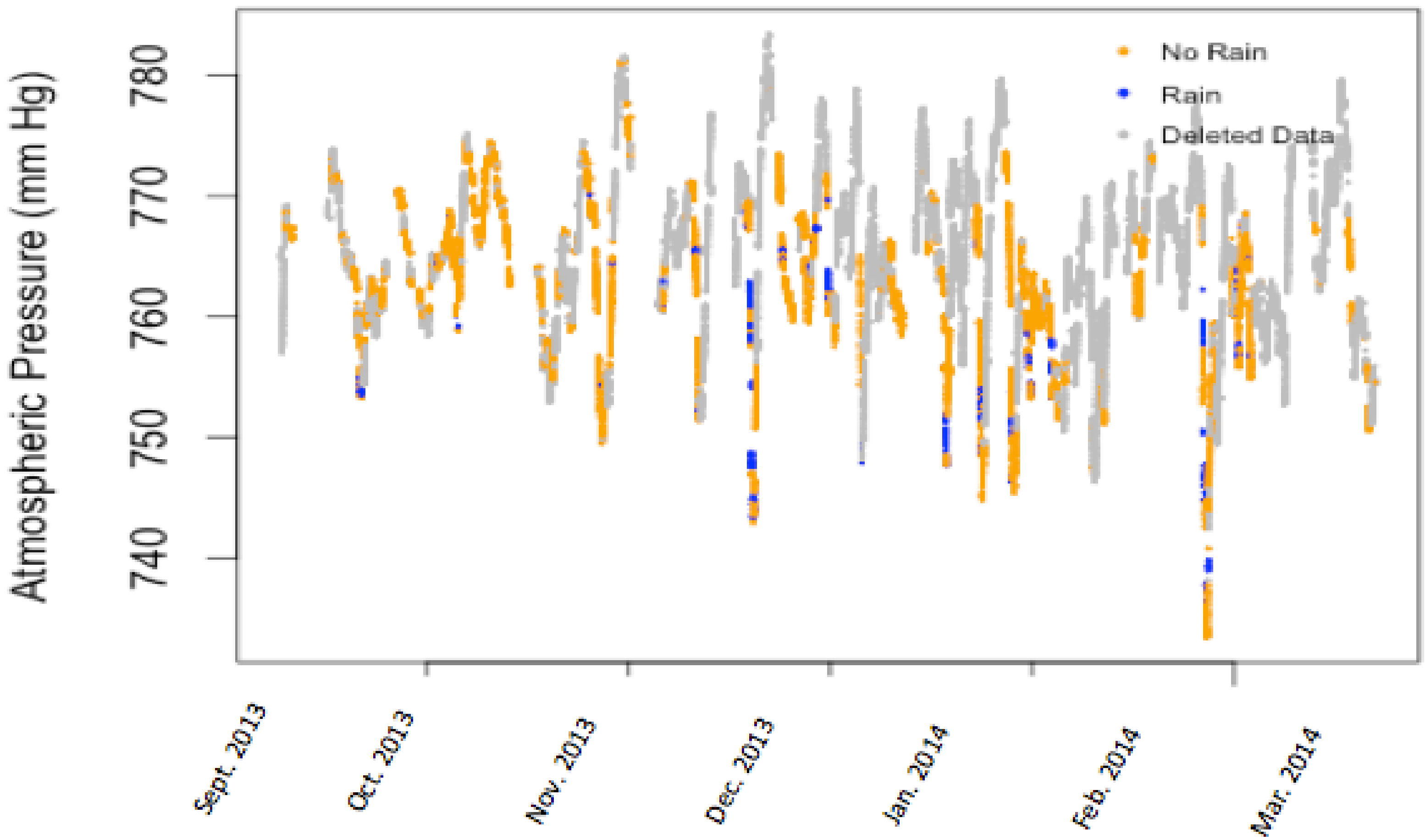

Figure 13: *Time series of atmospheric pressure from September 2013 to March 2014. Data points were deleted from the dataset either because the air temperature was less than 0°C or because relative humidity was less than 65%. Blank spaces represent times for which data points were lost during the data collection process.*

Figure 14 shows the time series of relative humidity during the study period.

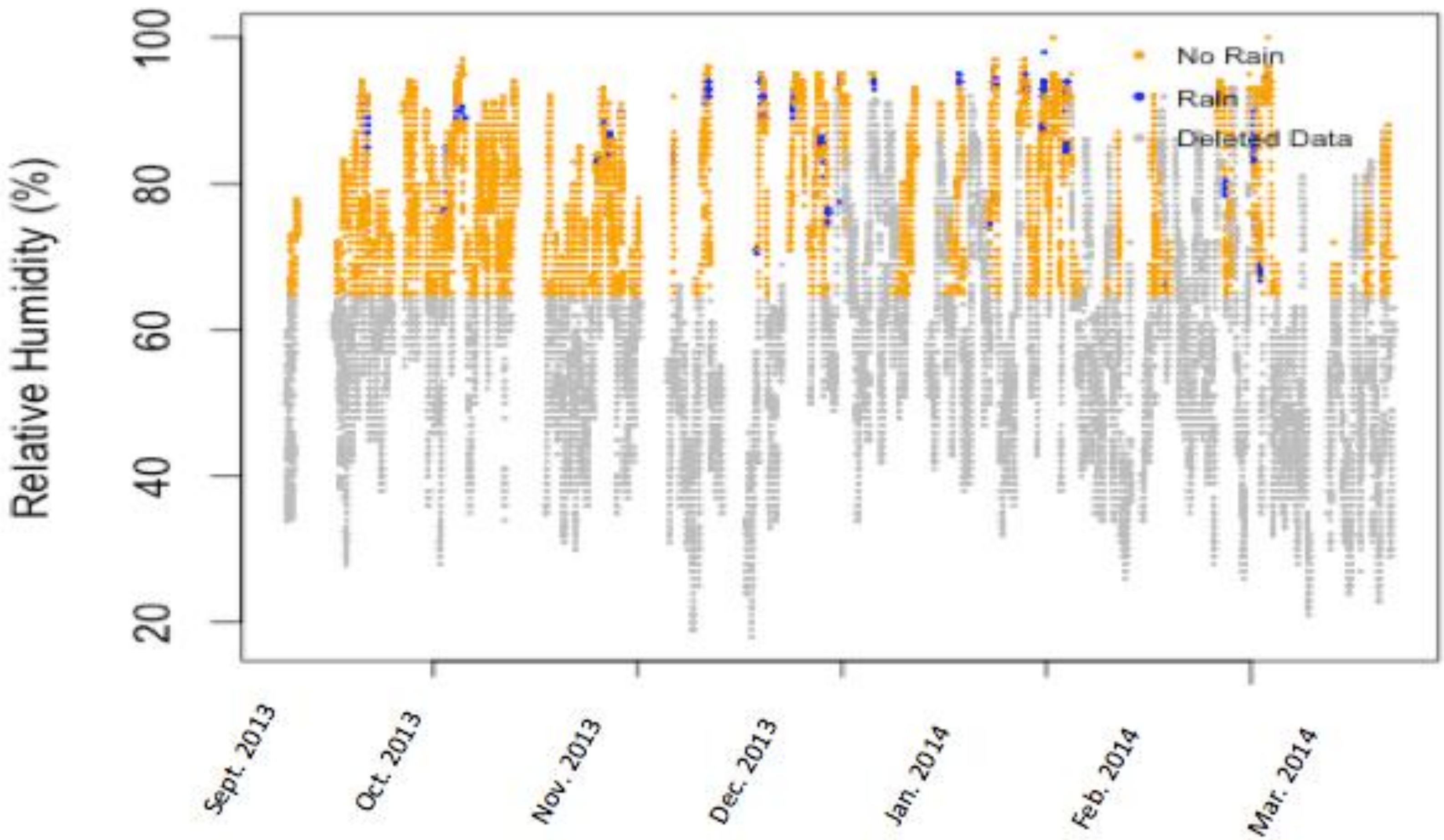

Figure 14: *Time series of relative humidity from September 2013 to March 2014. Data points were deleted from the dataset either because the air temperature was less than 0°C or because relative humidity was less than 65%. Blank spaces represent times for which data points were lost during the data collection process.*

Figure 15 shows the time series of wind speed during the study period.

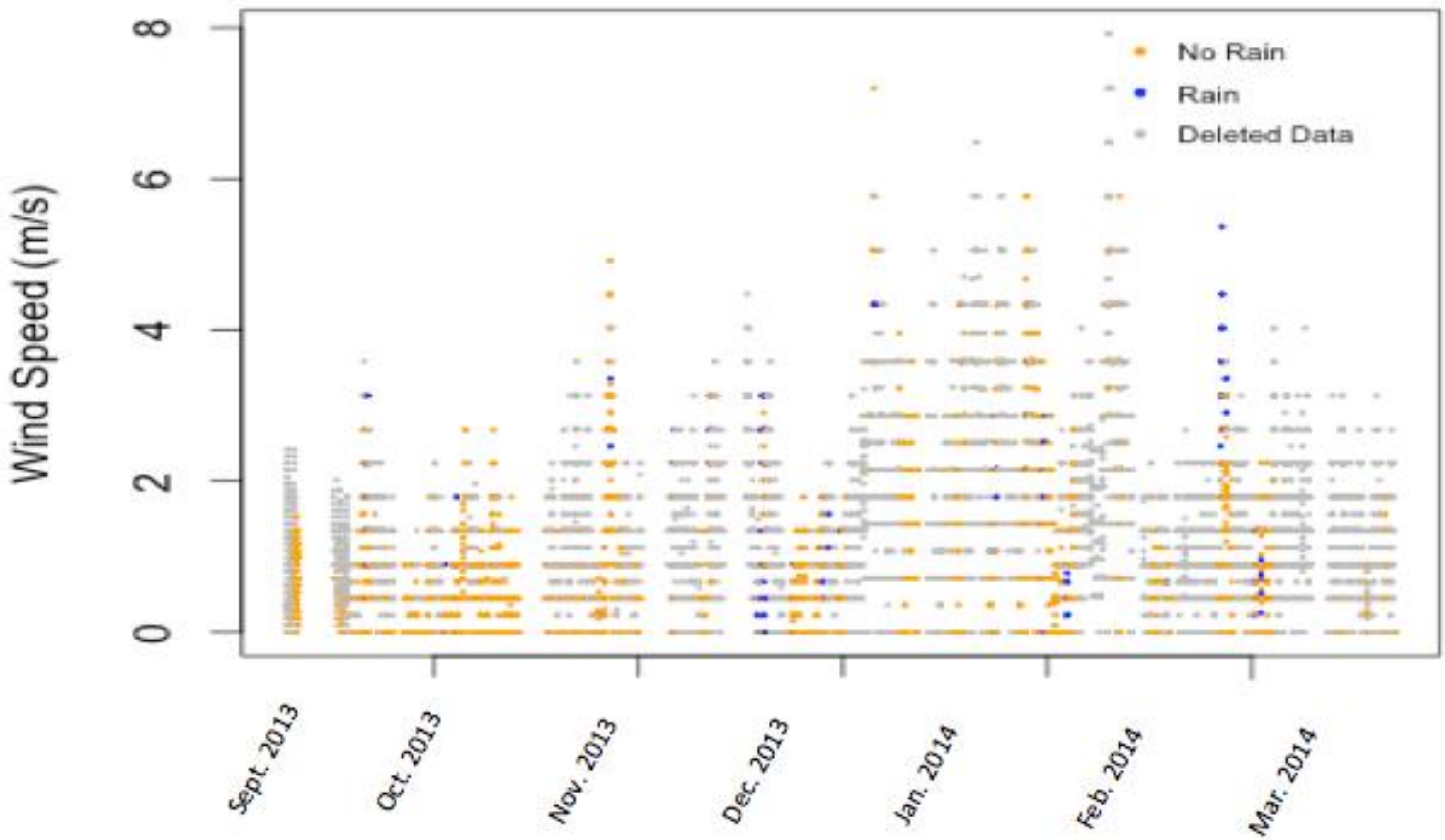

Figure 15: *Time series of wind speed from September 2013 to March 2014. Data points were deleted from the dataset either because the air temperature was less than 0°C or because relative humidity was less than 65%. Blank spaces represent times for which data points were lost during the data collection process.*

Figure 16 shows the time series of wind direction during the study period.

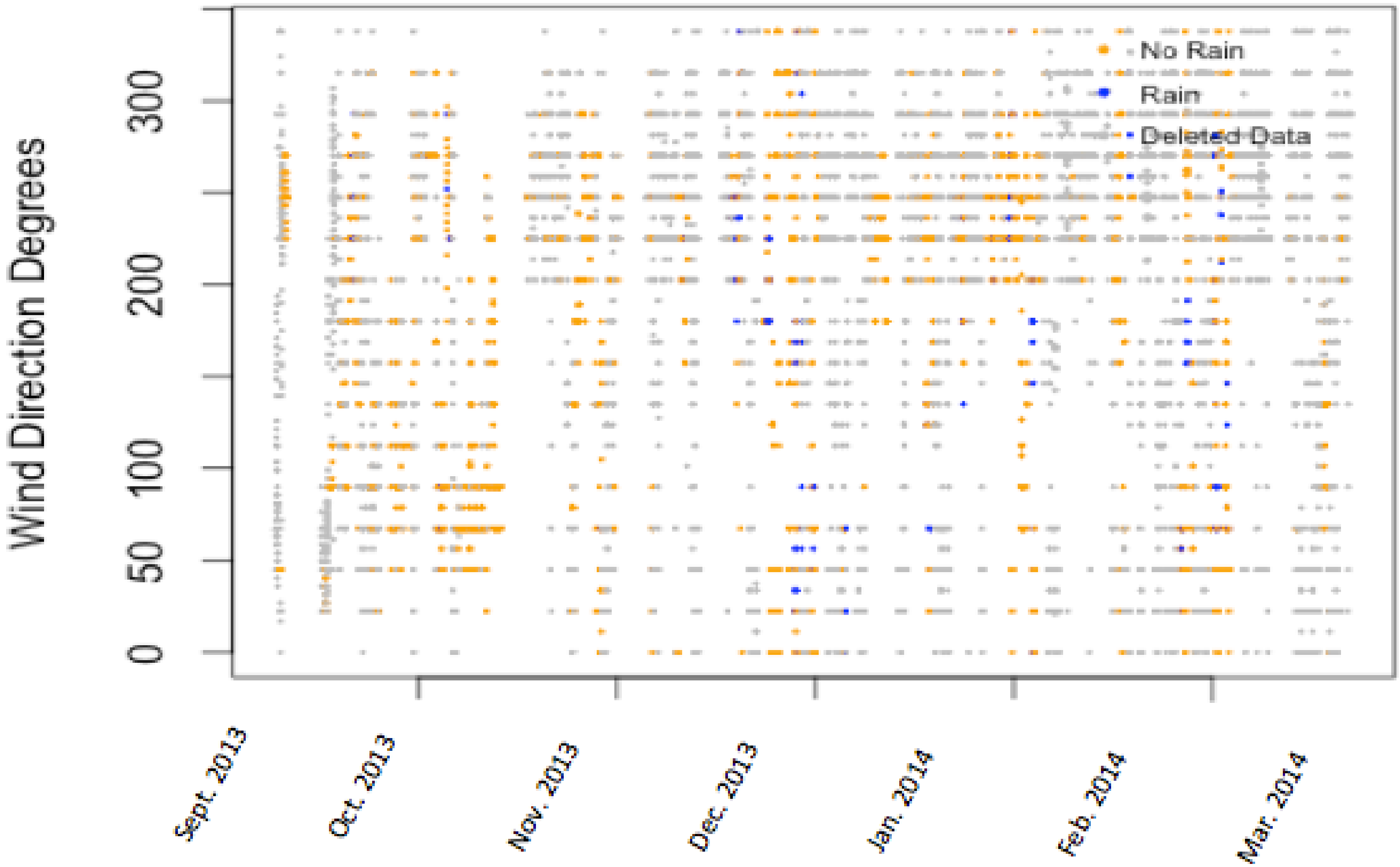

Figure 16: *Time series of wind direction from September 2013 to March 2014. Data points were deleted from the dataset either because the air temperature was less than 0°C or because relative humidity was less than 65%. Blank spaces represent times for which data points were lost during the data collection process.*

Figure 17 shows the above PNC, temperature, and atmospheric pressure plots zoomed in to the first week of October.

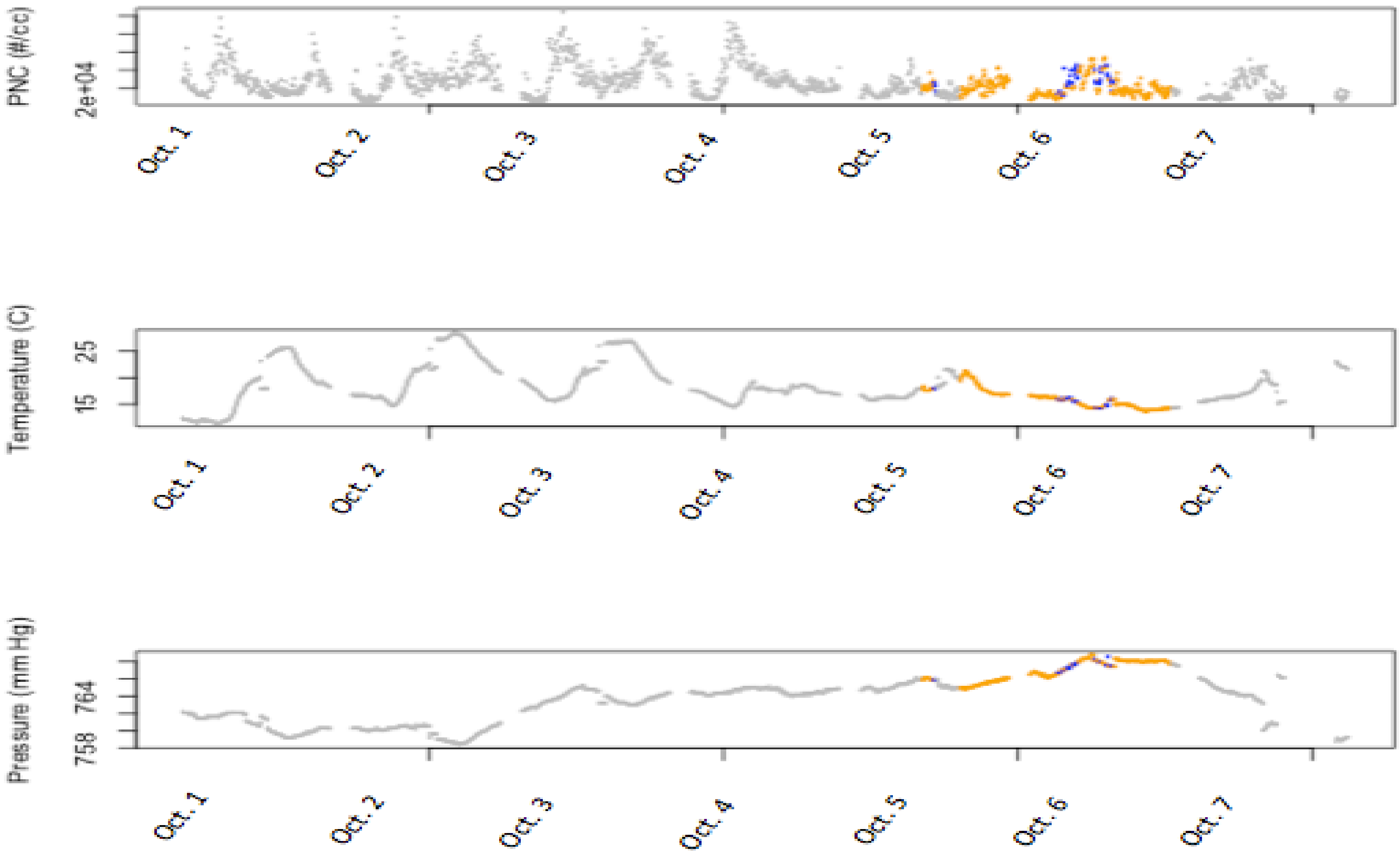

Figure 17: *Time series of PNC, temperature, and atmospheric pressure from October 1, 2013 to October 7, 2013. Data points were deleted from the dataset either because the air temperature was less than 0°C or because relative humidity was less than 65%. Blank spaces represent times for which data points were lost during the data collection process.*

Figure 18 shows the above PNC, relative humidity, and wind speeds plots zoomed in to the first week of October.

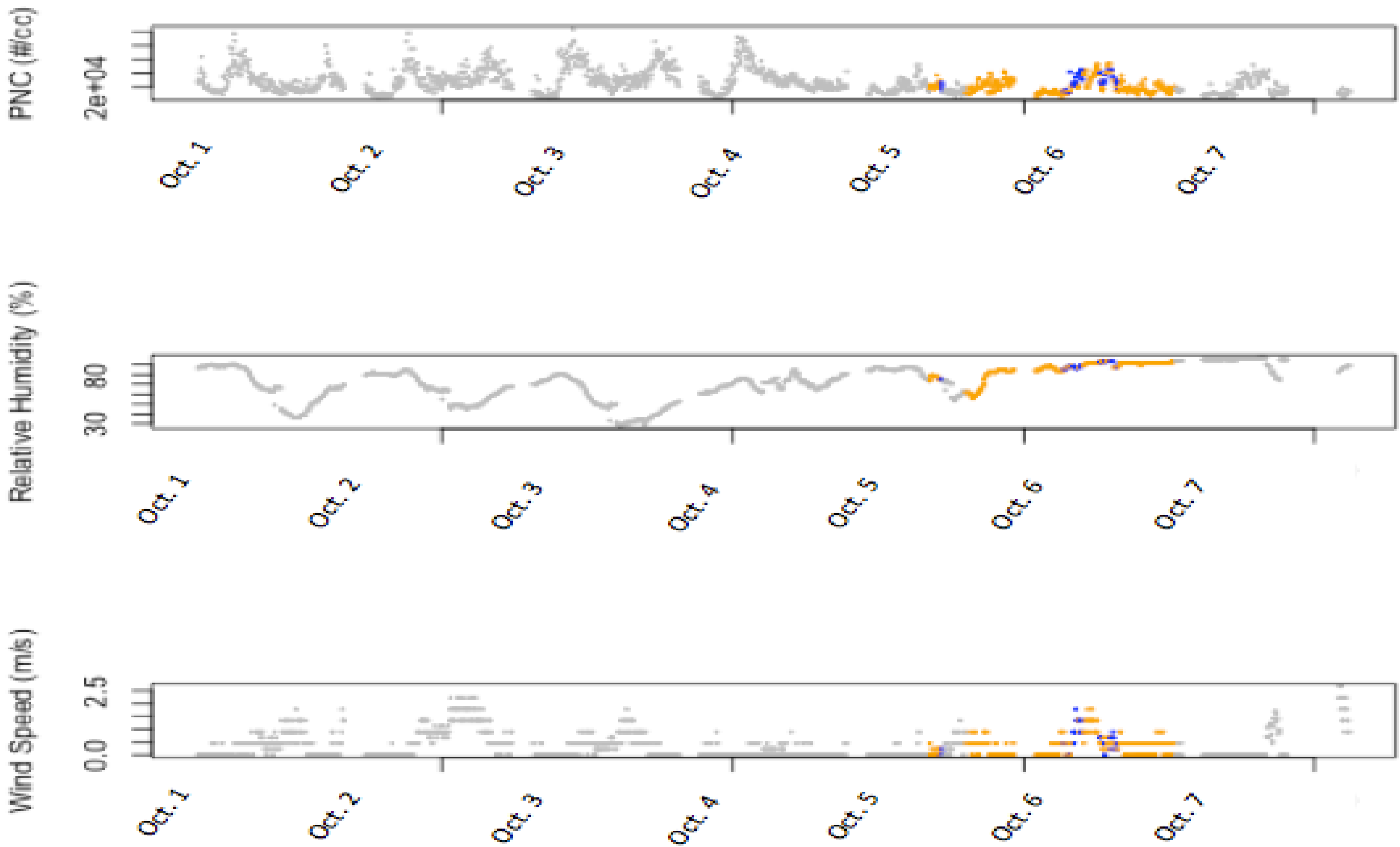

Figure 18: *Time series of PNC, relative humidity, and wind speed from October 1, 2013 to October 7, 2013. Data points were deleted from the dataset either because the air temperature was less than 0°C or because relative humidity was less than 65%. Blank spaces represent times for which data points were lost during the data collection process.*

Figure 19 shows the above PNC, wind speed, and wind direction plots zoomed in to the first week of October.

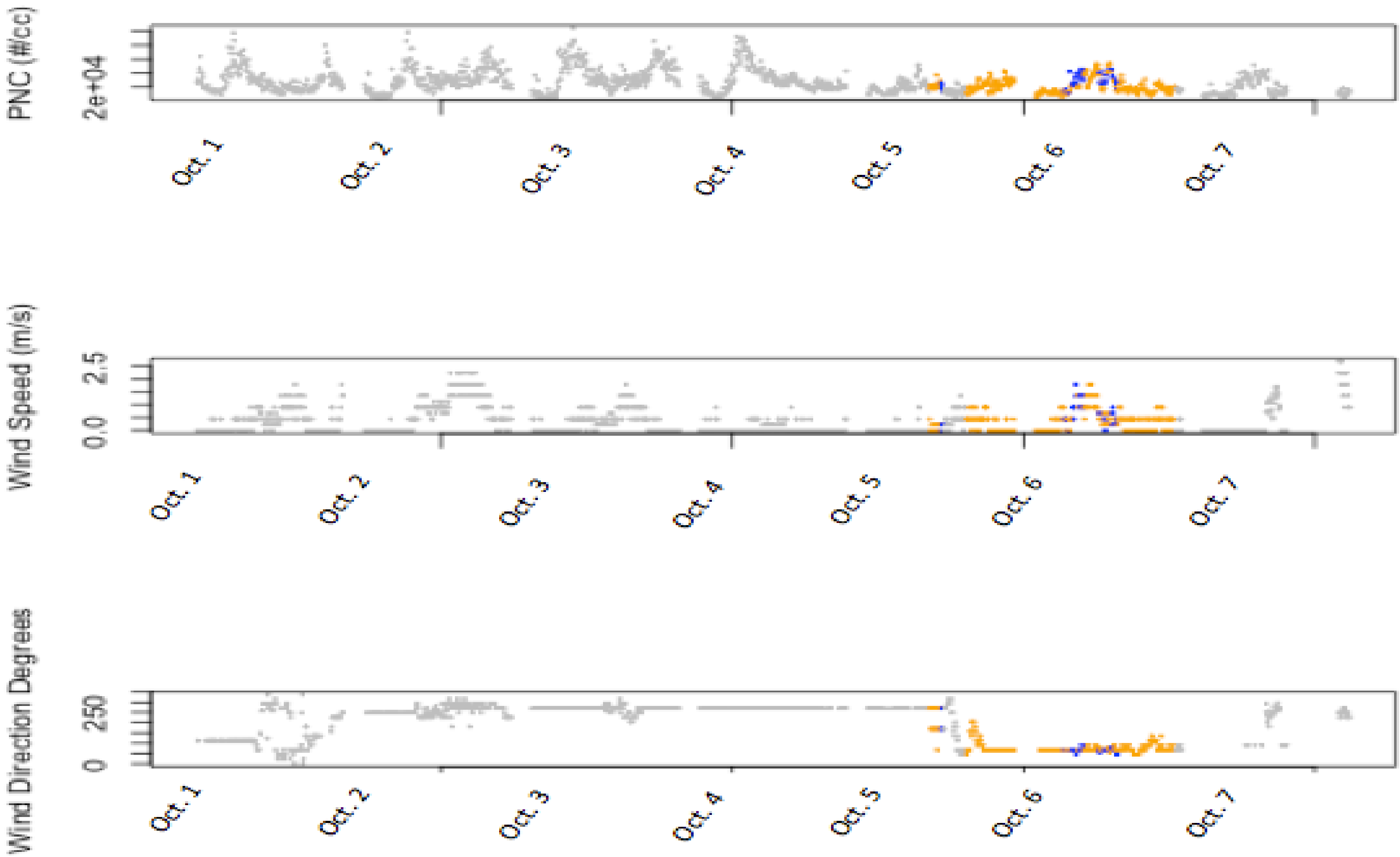

Figure 19: *Time series of PNC, wind speed, and wind direction from October 1, 2013 to October 7, 2013. Data points were deleted from the dataset either because the air temperature was less than 0°C or because relative humidity was less than 65%. Blank spaces represent times for which data points were lost during the data collection process.*

The time series plots simply help the visualization of the distribution of lnPNC and meteorological conditions throughout the study period. One other distribution that is helpful to visualize is that of rainstorm duration. Figure 20 is a cumulative distribution function for rainstorm duration. The figure excludes data points for which storm duration is equal to zero. Ninety-five percent of the data points used for the analysis had a rainstorm duration equal to zero.

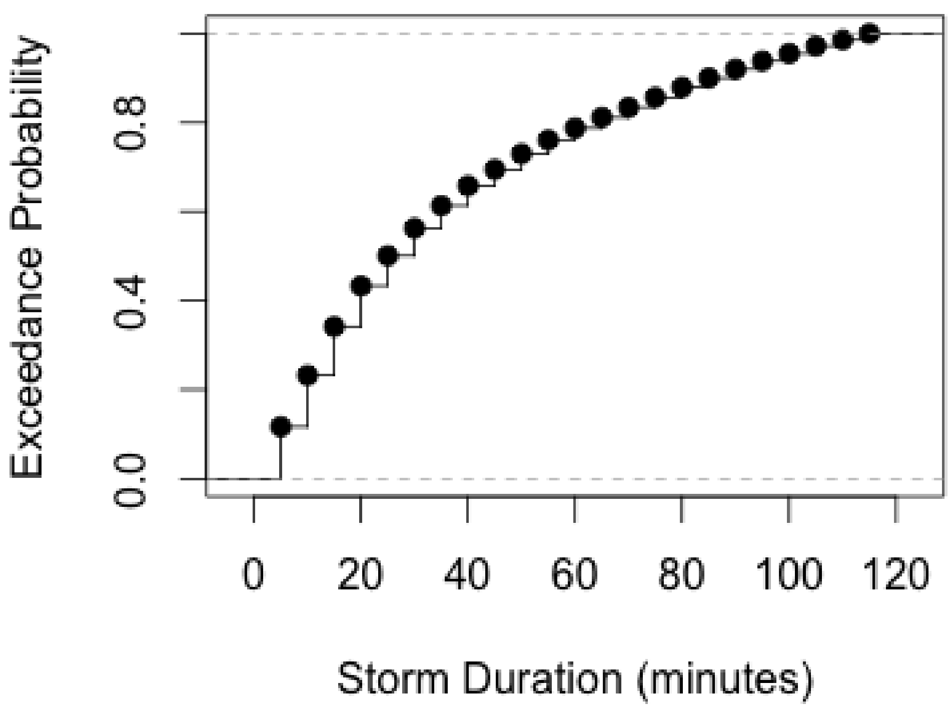

Figure 20: *Cumulative distribution function for storm duration.*

The cumulative distribution function of rainstorm durations shows what percentage of storms exceeded a certain storm duration. For example, the figure indicates that about 50% of storms are over 25 minutes long, and only about 30% of storms are over an hour long. By knowing the number of rainstorms of each length within a given time period and by knowing the decrease in PNC associated with rainstorms of each duration, the researchers could calculate the potential overestimation of human exposure to PNC.

4.2 Simple Linear Regression Models

The simple linear regression models estimate meteorological conditions and lnPNC as functions of storm duration. Simple linear regression models were built to estimate how the average of each meteorological condition changes throughout a rainstorm. The results are summarized in Table 1.

Table 1: *Simple linear regression models indicating the relationships of lnPNC and meteorology as functions of rainfall duration.*

Variable as a Function of Rainfall Duration	**Explanatory Variable**	**Coefficient Estimate**	**Standard Error**	**P Value**
lnPNC	(Intercept)	9.791345	0.035226	<0.001
	Duration of Rainfall Events	**-0.002570**	0.000744	<0.001
Temperature (°C)	(Intercept)	8.452003	0.304121	<0.001
	Duration of Rainfall Events	**-0.006007**	0.006423	0.35
Atmospheric Pressure (mm Hg)	(Intercept)	758.1	0.409238	<0.001
	Duration of Rainfall Events	**-0.036575**	0.008643	<0.001
Relative Humidity (%)	(Intercept)	87.590012	0.361324	<0.001
	Duration of Rainfall Events	**0.036970**	0.007631	<0.001
Wind Speed (m/s)	(Intercept)	1.046350	0.059793	<0.001
	Duration of Rainfall Events	**0.006148**	0.001263	<0.001
Sine(Wind Direction)	(Intercept)	0.0562907	0.0413445	0.17
	Duration of Rainfall Events	**-0.0040748**	0.0008672	<0.001
Cosine(Wind Direction)	(Intercept)	0.0086616	0.0316248	0.78
	Duration of Rainfall Events	**-0.0002495**	0.0006634	0.71

Figure 21 is a visualization of the simple linear regression model for lnPNC.

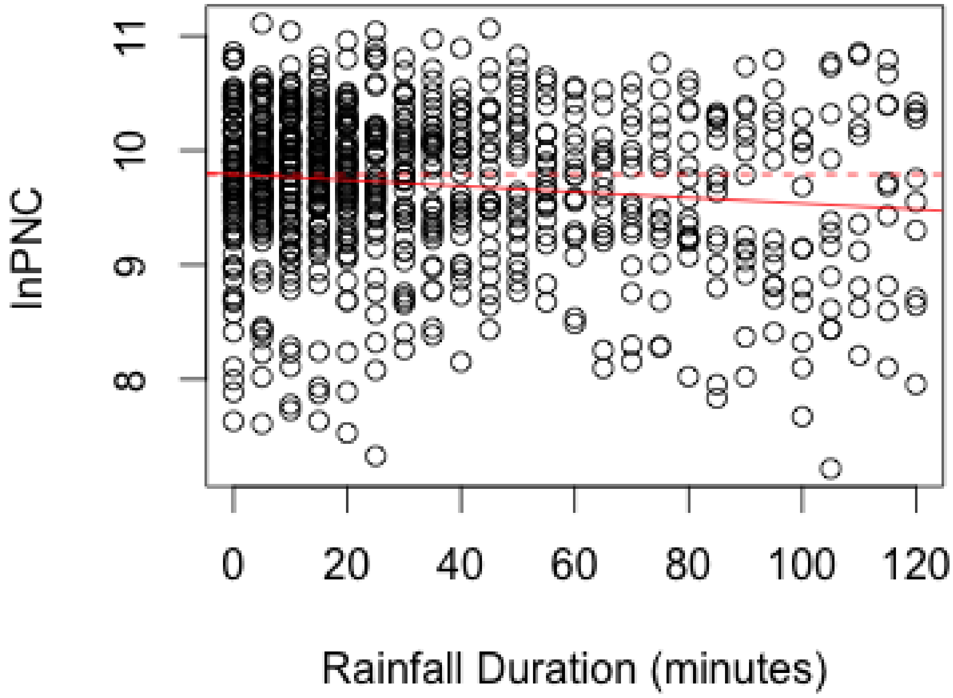

Figure 21: *Plot of lnPNC as a function of rainfall duration. The solid red line is the best-fit line, and the dashed red line is a horizontal line to provide contrast for the best-fit line.*

The simple linear regression models indicate that during rainstorms, average temperature does not change significantly, average atmospheric pressure decreases, average humidity increases, average wind speed increases, and wind direction changes significantly. Also, ignoring changes in meteorological conditions, lnPNC decreases significantly during rainstorms at an average rate of 15%/hour.

4.3 Multivariate Log-Normal Particle Number Concentration Models

The simple linear regression model for lnPNC (SLR) indicates that lnPNC decreases during rainstorms, but the SLR does not identify if the decrease is associated with the rainfall itself or other meteorological conditions. A multivariate linear regression model (MLR) can adjust for other meteorological conditions and calculate the change in lnPNC associated with rainfall.

From the data that were collected, the MLR for estimating lnPNC, as is summarized in Table 2, was created.

Table 2: *Multivariate linear regression model for lnPNC.* $R^2=0.43$.

Variable		**Coefficient Estimate**	**Standard Error**	**P value**
(Intercept)		11.4904774	3.0054082	<0.001
Temperature (°C)		0.0256359	0.0069738	<0.001
Atmospheric Pressure (mm Hg)		-0.0028601	0.0037555	0.44
Wind Speed (m/s)		-0.1839512	0.0242608	<0.001
Cosine(Wind Direction $\times \frac{\pi}{180}$)		0.0812415	0.0389747	<0.05
Sine(Wind Direction $\times \frac{\pi}{180}$)		0.0417929	0.0394797	0.29
Relative Humidity (%)		0.0039582	0.0042150	0.35
Weekday (Relative to Friday)	Monday	-0.0864106	0.0843200	0.31
	Tuesday	0.3649103	0.0837628	< 0.001
	Wednesday	-0.0305426	0.0778930	0.70
	Thursday	0.0873006	0.0963593	0.37
	Saturday	-0.1492197	0.0829728	0.07
	Sunday	0.1500268	0.0784896	0.06
Cosine(Hour$\times \frac{360}{24} \times \frac{\pi}{180}$)		-0.4128481	0.0381934	<0.001
Sine(Hour$\times \frac{360}{24} \times \frac{\pi}{180}$)		-0.1978448	0.0326674	<0.001
Cosine(Month$\times \frac{360}{12} \times \frac{\pi}{180}$)		-0.1952537	0.1033359	0.06
Sine(Month$\times \frac{360}{12} \times \frac{\pi}{180}$)		0.5386900	0.0644698	<0.001
Rainfall Rate (mm/hour)		0.0016957	0.0034092	0.62
Storm Duration (minutes)		-0.0019427	0.0006127	<0.01

The model indicates how meteorological conditions, temporal variables, and storm duration are associated with average lnPNC. From Table 2, many of the hypotheses from Section 2.0 can be tested. The results of the hypothesis tests are summarized in Table A.1.

As hypothesized, wind speed was inversely related to lnPNC, and atmospheric pressure and relative humidity are not correlated with changes in lnPNC. Wind direction also had a significant effect on lnPNC, as indicated by the significant coefficient for cosine(wind direction). Contrary to my expectations, temperature was positively correlated with lnPNC. Also, storm duration and lnPNC were inversely related, indicating that rainfall duration was associated with significant decrease in lnPNC. Adjusting for meteorological conditions, the MLR indicates that PNC decreases at a rate of 12%/hour during rainstorms.

4.4 Expected Changes in lnPNC Due to Meteorology During Rainstorms

By using the MLR in conjunction with the simple linear regression models, it can be determined if meteorological conditions are potentially enhancing or masking the association between lnPNC and rainfall duration.

When using the MLR and the simple linear regressions, changes in wind speed during rainfall are expected to be associated with decreasing lnPNC. Therefore, the changes in wind speed during rainfall could potentially enhance the association between lnPNC and rainfall duration that was observed in the SLR.

The sine of wind direction is expected to decrease throughout rainstorms, and according to the MLR, the sine of wind direction is positively correlated with lnPNC. Therefore, the changes in wind direction during rainstorms are expected to be associated with decreasing lnPNC, and the change in wind direction could potentially enhance the association between lnPNC and rainfall duration in the SLR.

All other measured meteorological conditions (temperature, atmospheric pressure, relative humidity, cosine of wind direction) either did not change significantly during rainstorms or were not significantly correlated with lnPNC.

During rainfall, all the changes in meteorology indicate that lnPNC can be expected to be decreasing. The combination of these meteorological changes could potentially enhance the association between PNC and rainfall duration observed in the SLR. Ultimately, the coefficient for storm duration in the MLR indicates whether or not the rainfall itself is associated with a decrease in PNC.

5.0 DISCUSSION

5.1 Comparison of Linear Regression Models

The first PNC model—the simple linear regression of lnPNC as a function of storm duration—is essentially a best-fit line (Figure 21). The SLR indicates that average PNC decreases significantly at an average rate of 15%/hour over the course of rainstorms. However, the SLR does not indicate whether rainfall or other meteorological conditions are associated with the decrease in PNC during rainstorms; therefore, the coefficient from the SLR was compared to the coefficient from the MLR.

The MLR was adjusted for temperature, atmospheric pressure, relative humidity, wind speed, wind direction, rainfall rate, day of the week, month, and hour of the day, and it explains the variability in lnPNC better than does the SLR. The MLR indicates that PNC decreases at a rate of 12%/hour during rainfall, but this decrease is not statistically significantly different from the decrease indicated by the SLR.

Because the other explanatory variables are held constant as the coefficient for storm duration is calculated, the decrease in lnPNC during rainstorms is associated with the rainfall itself in the MLR. Because the coefficients for rainfall duration in the SLR and the MLR are the statistically same, the observed decrease in PNC in the SLR is entirely associated with the rainfall and not with other meteorological conditions. If, for example, the magnitude of the coefficient in the SLR were statistically significantly greater than the magnitude of the coefficient in the MLR, then some of the decrease in PNC during rainstorms would be associated with conditions besides rainfall, and those conditions would be enhancing the observed effect of rainfall on PNC. Because the coefficients are *not* statistically significantly different, it is rainfall, and not other meteorological conditions, that is associated with the decrease in PNC.

Even though the changes in wind speed and wind direction during rainstorms were *expected* to be associated with decreasing PNC, meteorological conditions other than rain were not associated with changes in PNC during rainfall. The magnitude of the rainfall duration coefficient in the SLR was slightly larger than the magnitude of the rainfall duration coefficient in the MLR, but the difference between the two was not significantly different. Perhaps if more data were collected, the differences between the two rainfall duration coefficients would become statistically significant.

5.2 Determining a Scavenging Coefficient from the Models

The MLR indicates that, controlling for meteorological conditions and temporal variables, rainfall is associated with lnPNC removal at an average rate of -0.0019/minute. Therefore, $\lambda_{MLR} = 0.0019$**.** This scavenging coefficient indicates that it would take about 365 minutes of rainfall to reduce PNC by one half.

Washout of particles is dependent on rainfall rate, which is, of course, not constant. Studies suggest that scavenging coefficients are likely to be greater for rainstorms with a higher rainfall rate than for rainstorms with a lower rainfall rate (Andronache et al., 2006). Therefore, it is important to recognize that the scavenging coefficient provided by the MLR is an *average* scavenging coefficient.

However, this might not apply to small particles. Considering that the vast majority of particles counted in PNC measurements are very small, the average scavenging rate may not have a large variance. A study of aerosol scavenging by precipitation found that rainfall rate significantly affects the scavenging rate for large particles, but rainfall rate does not significantly affect scavenging rates for small particles (100 nm to 1,000nm) (Zhang et al. 2004). Therefore, the variance of the scavenging coefficient calculated in the MLR may be rather small, and the scavenging coefficient can apply for most types of rainfall.

5.3 Comparison With the Previous CAFEH Model

Some of the coefficients in the models were surprising. The results of the simple linear regression models for lnPNC and meteorology during rainfall were not surprising, but one of the coefficients in the MLR was surprising. The coefficient for temperature had a positive sign, even though lnPNC and temperature are commonly accepted to be inversely correlated (Patton et al., 2014; von Bismarck-Osten et al., 2013). It could be that particle concentrations and temperature are inversely correlated with temperature during dry weather, but the opposite is true during rainy weather. During dry weather, particle formation rates are inversely correlated with temperature (Jacobson, 1999), but perhaps rainfall hinders particle formation during cold weather.

5.4 Weaknesses of the Model

My research indicated that between September 2013 and March 2014 there were about 80 rainstorms. This value is a little greater than expected. Assuming that there are 127 days of precipitation in Boston and that the days of precipitation are uniformly distributed (which might not be such a bad assumption in Boston, seeing as Boston receives about the same amount of precipitation days each month (Figure 10)), 74 days of precipitation would be expected for the study period (climate-zone.com, 2014).

But considering that snow was not considered for this study, the expected days of rainfall during the study period would be less than 74. However, multiple rainstorms can occur during one day and considering the weaknesses of the storm duration function in the model, multiple "rainstorms" can occur within one actual rainstorm.

The chosen definition of a rainstorm required a rainstorm to have consecutive five-minute intervals with rain intensity greater than 0 mm/hour. Thus, if there were a rainstorm that had just a brief break, it would be split into two different rainstorms. For example, if there were two hours of rain and then five minutes of no rain and then two more hours of rain, the average person would consider this to be one rainstorm. Sometimes rainstorms just stop for a few minutes. Throughout the first two hours of the rainstorm, PNC would be expected to be decreasing. Once the rain started again, PNC for the first five minutes of the second two-hour block of rain should not be expected to be the same as PNC for the first five minutes of the first two-hour block of rain. PNC should be expected to be significantly reduced for the beginning of the second two-hour block of rain. But the storm duration function does not recognize that the two two-hour storms are essentially one four-hour storm.

Not only does the model not take into account small breaks in rainstorms, but it also does not take into account the rebound period that might occur after a storm. PNC five minutes after the end of a long rainstorm should not be expected to be the same as the average PNC associated with dry weather.

Another factor that was not considered in the analysis was autocorrelation. At first, incorporating autocorrelation into the model was attempted. However, because data points were removed from the analysis, the time series was not perfectly continuous. Because the time series was not perfectly continuous, autocorrelation could not be incorporated into the model.

Not only do weaknesses exist within the data analysis, but weaknesses also exist within the data collection. One issue with the study is that data collection did not take place throughout an entire year, but rather just from September 9, 2013 to March 12, 2014. Rainstorms in the summer may affect PNC differently than rainstorms in the fall or spring. A full year of data would be very helpful for more accurately determining how PNC changes during rainstorms.

6.0 CONCLUSION

A simple linear regression model indicated that PNC does in fact decrease during rainstorms. During rainstorms, meteorological conditions changed significantly. Atmospheric pressure decreased, relative humidity increased, wind speed increased, and wind direction changed significantly. When the model was adjusted for weather conditions, the PNC removal associated with rainfall duration did not change significantly, indicating that the decrease in PNC during rainfall is not associated with meteorological conditions other than rain. Future CAFEH models could adjust for rainfall, and the studies might find that rainfall is associated with PNC removal. But it has not yet been determined if rainfall will significantly reduce overall human exposure to PNC. Rainfall duration is associated with a modest PNC decrease, but rainfall only occurred during 1% of the study period. The CAFEH researchers should determine if a modest decrease in PNC during only 1% of total time of exposure is enough to significantly affect overall human exposure.

7.0 ACKNOWLEDGEMENTS

I would like to thank all the people who have helped me with my thesis. Dr. John Durant and Dr. Elena Naumova have guided me through my research as my thesis committee. Allison Patton has taught me so much about data analysis in R and has been helping me for almost three years. Matt Simon has helped me with my data collection and with guidance for my research. The Tufts Exposure Science Group, consisting of Dr. Doug Brugge, Kevin Lane, Wig Zamore, Hanaa Rohman, and Luz Padro-Martinez, has provided valuable input for my study. John Lane has assisted me with my data collection in Roxbury. And finally, all of my friends who have helped me with my data collection have been so supportive throughout the process. Thank you to Grace House, Brandon Wheeler, Jamie Norton, and Bobby McShane. The Tufts Civil and Environmental Engineering Department, the John A. Cataldo Scholarship, and the National Institute of Health Sciences provided funding.

REFERENCES

Adar, S., Gold, D., Coull, B., Stone, P., & Suh, H. (2007). Focused exposures to airborne traffic particles and heart rate variability in the elderly. *Epidemiology, 18*, 95-103.

Ahrens, C. D. (2008). *Meteorology today* (9 ed.). Pacific Grove: Brooks Cole.

Andronache, C., Gronholm, T., Laasko, L., Phillips, V., & Venalainen, A. (2006). Scavenging of ultrafine particles by rainfall at a boreal site: observations and model estimations, *Atmospheric Chemistry and Physics, 6,* 4739–4754.

Brugge, D., Durant, J. L., & Rioux, C. (2007). Near-highway pollutants in motor vehicle exhaust: a review of epidemiologic evidence of cardiac and pulmonary health risks. *Environmental health: a global access science source*, *6*, 23. doi:10.1186/1476-069X-6-23

Brugge, D., Lane, K., Padro-Martinez, L., Stewart, A., Hoesterey, K., Weiss, D., . . . Mwamburi, M. (2013). Highway proximity associated with cardiovascular disease risk: the influence of individual-level confounders and exposure misclassification. Environmental Health, 12(1), 84.

Climate Information for - Massachusetts - Northeast - United States - Climate Zone. N.p., n.d. Web. 31 Mar. 2014. <http://www.climate-zone.com/climate/united-states/Massachusetts/Boston/>.

Cooper, C. D., & Alley, F. C. (2009). *Air pollution control: a design approach* (4 ed.). Prospect Heights (IL): Waveland press.

Duhanyan, N., & Roustan, Y. (2011). Below-cloud scavenging by rain of atmospheric gases and particulates. *Atmospheric Environment*, *45*(39), 7201–7217. doi:10.1016/j.atmosenv.2011.09.002

EPA. Environmental Protection Agency, n.d. Web. 31 Mar. 2014. <http://www.epa.gov/pm/health.html>.

EPA. Environmental Protection Agency in collaboration with North Carolina State University (NCSU, n.d. Web. 30 April. 2014.

Jacob, D. J. (1999). *Introduction to atmospheric chemistry*. Princeton, N.J.: Princeton University Press.

Jacobson, M. Z. (1999). *Fundamentals of atmospheric modeling*. Cambridge, UK: Cambridge University Press.

Nieto, P.J., Garcia B., Diaz, J.M., & Brana, M.A. (1994). Parametric study of selective removal of atmospheric aerosol by below-cloud scavenging, *Atmospheric Environment, 28*(14), 2335–2342.

Olivares, G., Johansson, C., Ström, J., Hansson, H.-C. (2007). The role of ambient temperature for particle number concentrations in a street canyon. *Atmospheric Environment 41*, 2145-2155.

Padró-Martínez, L. T., Patton, A. P., Trull, J. B., Zamore, W., Brugge, D., & Durant, J. L. (2012). Mobile monitoring of particle number concentration and other traffic-related air pollutants in a near-highway neighborhood over the course of a year. *Atmospheric environment (Oxford, England : 1994)*, *61*, 253–264. doi:10.1016/j.atmosenv.2012.06.088

Patton, A. P., Collins, C., Naumova, E. N., Zamore, W., Brugge, D., & Durant, J. L. (2014). An hourly regression model for ultrafine particles in a near-highway urban area. *Environmental science & technology*, *48*(6), 3272–80. doi:10.1021/es404838k

Seinfeld, J., & Pandis, S. (1998). *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change*. New York: Wiley.

Vliet, P. V., Knape, M., Hartog, J. d., Harsemma, H., & Brunekreef, B. (1997). Motor vehicle exhaust and chronic respiratory symptoms in children living near freeways. *Environmental Research*, *74*, 122-132.

Von Bismarck-Osten, C., Birmili, W., Ketzel, M., Massling, A., Petäjä, T., & Weber, S. (2013). Characterization of parameters influencing the spatio-temporal variability of urban particle number size distributions in four European cities. *Atmospheric Environment*, *77*, 415–429. doi:10.1016/j.atmosenv.2013.05.029

Zhang, L., Michelangeli, D. V, & Taylor, P. a. (2004). Numerical studies of aerosol scavenging by low-level, warm stratiform clouds and precipitation. *Atmospheric Environment*, *38*(28), 4653–4665. doi:10.1016/j.atmosenv.2004.05.042

Zwack, L. M., Paciorek, C. J., Spengler, J. D., & Levy, J. I. (2011). Modeling spatial patterns of traffic-related air pollutants in complex urban terrain. *Environmental health perspectives*, *119*(6), 852–9. doi:10.1289/ehp.1002519

APPENDIX

Table A.1: *Comparison of the hypotheses for the multivariate linear regression model and the results for the multivariate linear regression model (MLR).*

	Hypothesis	**Result**	**Reject/Do Not Reject Hypothesis**
H1	$\beta_{m\ temperature} < 0$	$\beta_{m\ temperature} > 0$	Reject
H3	$\beta_{m,\ pressure} = 0$	$\beta_{m,\ pressure} = 0$	Do Not Reject
H5	$\beta_{m,\ humidity} = 0$	$\beta_{m,\ humidity} = 0$	Do Not Reject
H7	$\beta_{m\ wind\ speed} < 0$	$\beta_{m\ wind\ speed} < 0$	Do Not Reject
H9	$\beta_{m,\ sine(wind\ direction)} \neq 0$, $\beta_{m,\ cosine(wind\ direction)} \neq 0$	$\beta_{m,\ sine(wind\ direction)} = 0$, $\beta_{m,\ cosine(wind\ direction)} \neq 0$	Reject
H11	$\beta_{m,\ storm\ duration} < 0$	$\beta_{m,\ storm\ duration} < 0$	Do Not Reject

Table A.2: *Expected change in lnPNC due to meteorological changes during rainstorms based on the multivariate linear regression (MLR) coefficients and the simple linear regression coefficients.*

Meteorological Condition	**Sign of Multivariate Linear Regression Coefficient**	**Sign of Simple Linear Regression Coefficient During Rainstorm**	**Expected Change in lnPNC**
Temperature	+	0	0
Atmospheric Pressure	0	—	0
Relative Humidity	0	+	0
Wind Speed	—	+	—
Sine(Wind Direction)	0	—	+
Cosine(Wind Direction)	+	—	—

Natural Logarithm of Particle Number Concentration

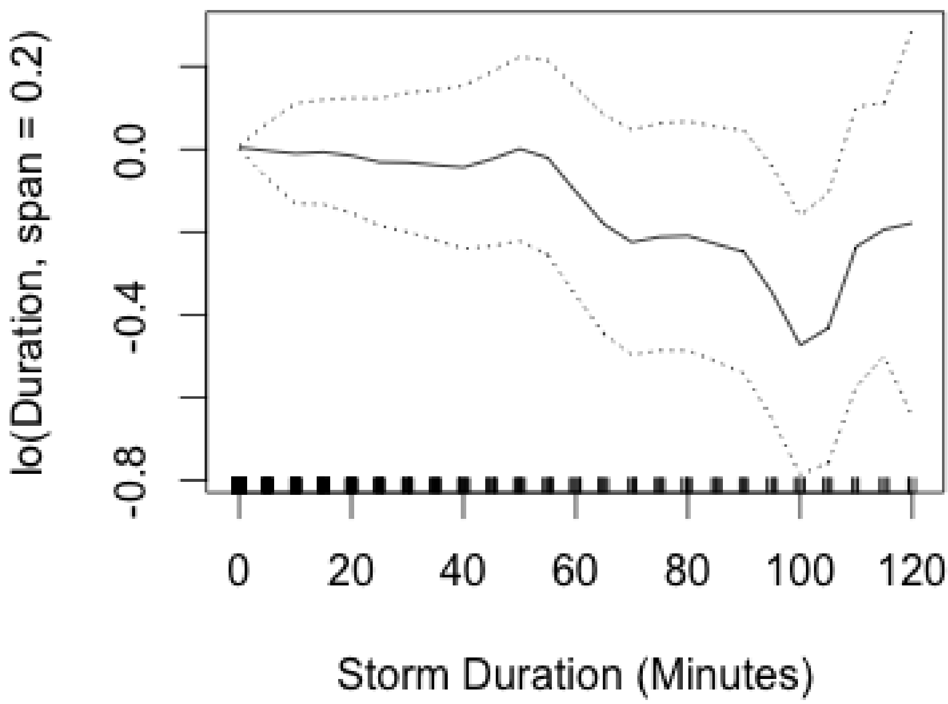

Figure A.1: *Gam plot of lnPNC as a function of storm duration. The dotted lines represent the standard error. The tick marks at the bottom of the figure represent each data points. Large black boxes represent many tick marks and therefore a high density of data points.*

Temperature

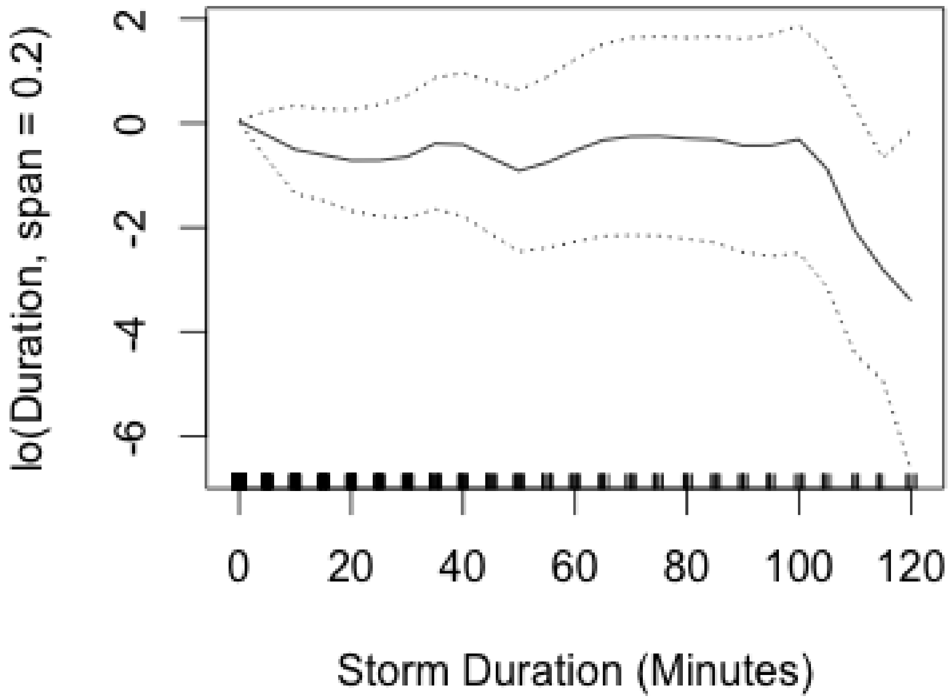

Figure A.2: *Gam plot of temperature as a function of storm duration. The dotted lines represent the standard error. The tick marks at the bottom of the figure represent each data points. Large black boxes represent many tick marks and therefore a high density of data points.*

Atmospheric Pressure

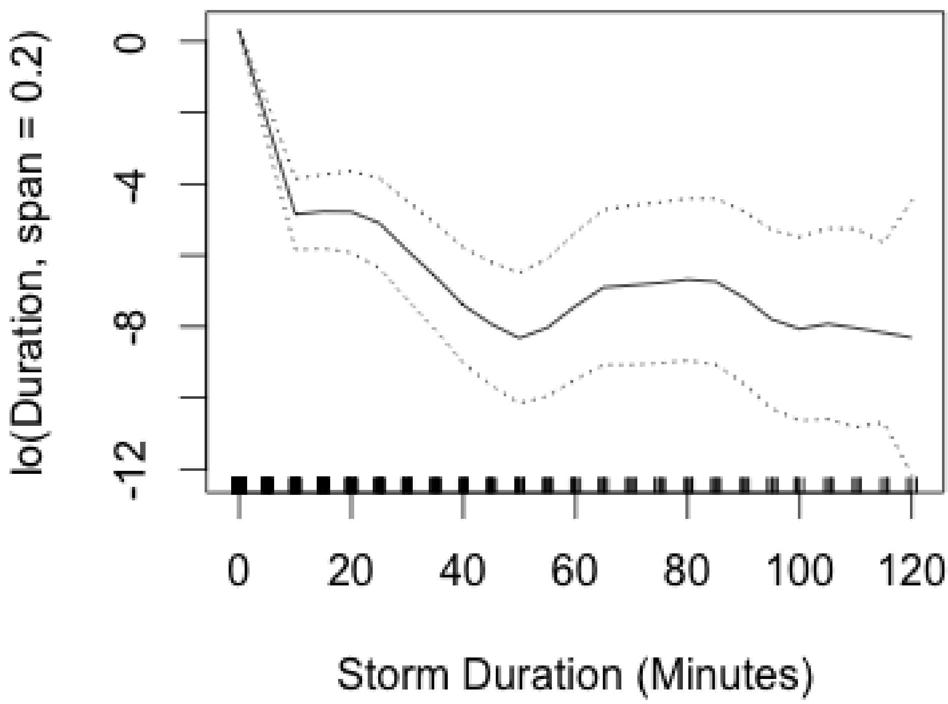

Figure A.3: *Gam plot of atmospheric pressure as a function of storm duration. The dotted lines represent the standard error. The tick marks at the bottom of the figure represent each data points. Large black boxes represent many tick marks and therefore a high density of data points.*

Relative Humidity

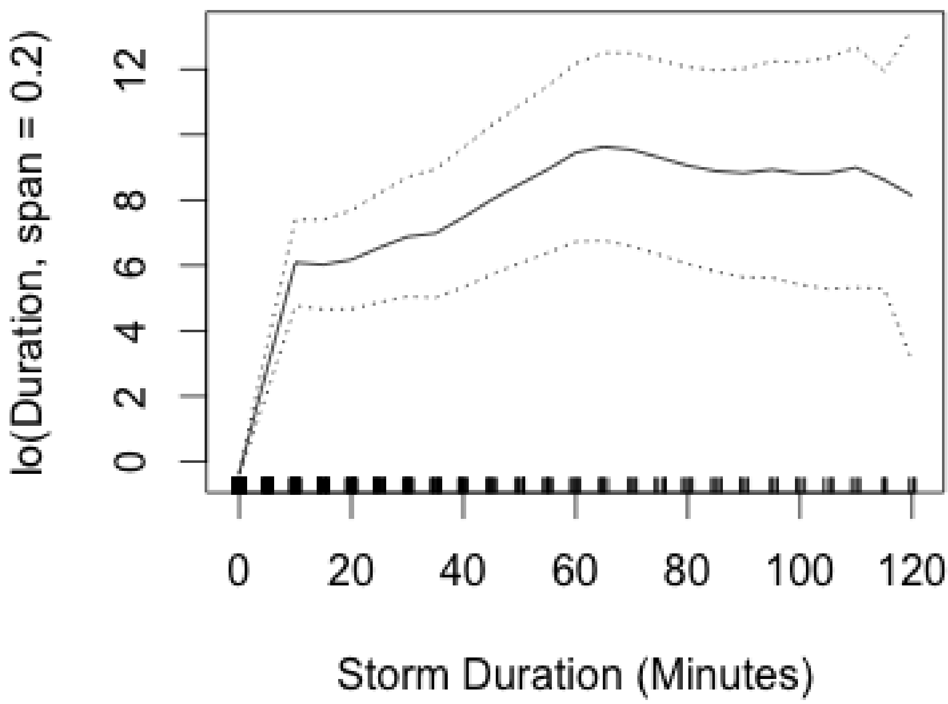

Figure A.4: *Gam plot of relative humidity as a function of storm duration. The dotted lines represent the standard error. The tick marks at the bottom of the figure represent each data points. Large black boxes represent many tick marks and therefore a high density of data points.*

Wind Speed

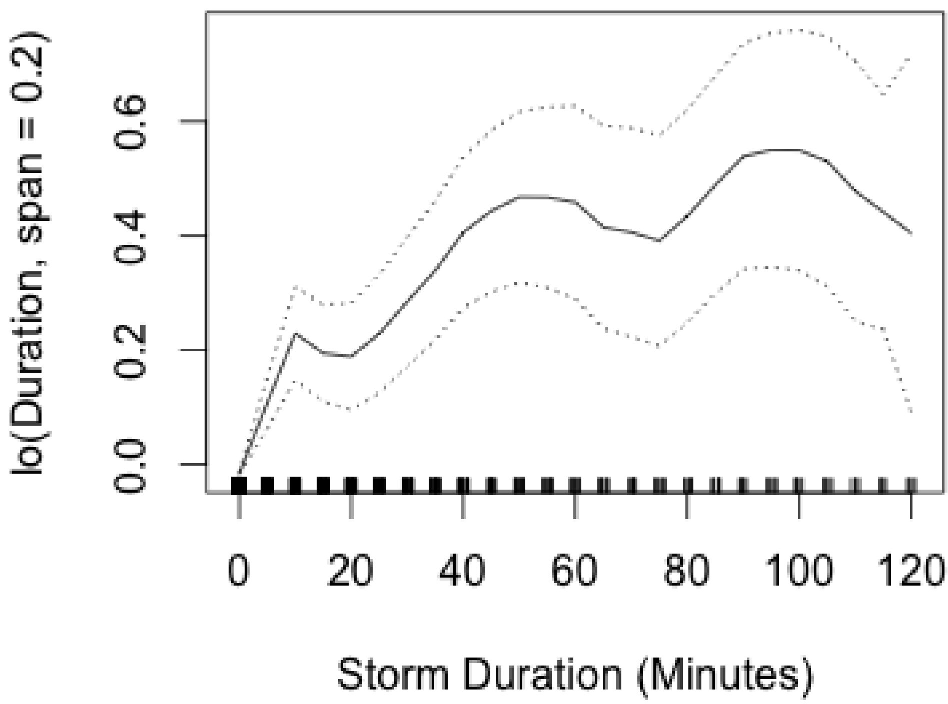

Figure A.5: *Gam plot of wind speed as a function of storm duration. The dotted lines represent the standard error. The tick marks at the bottom of the figure represent each data points. Large black boxes represent many tick marks and therefore a high density of data points.*

Sine(Wind Direction)

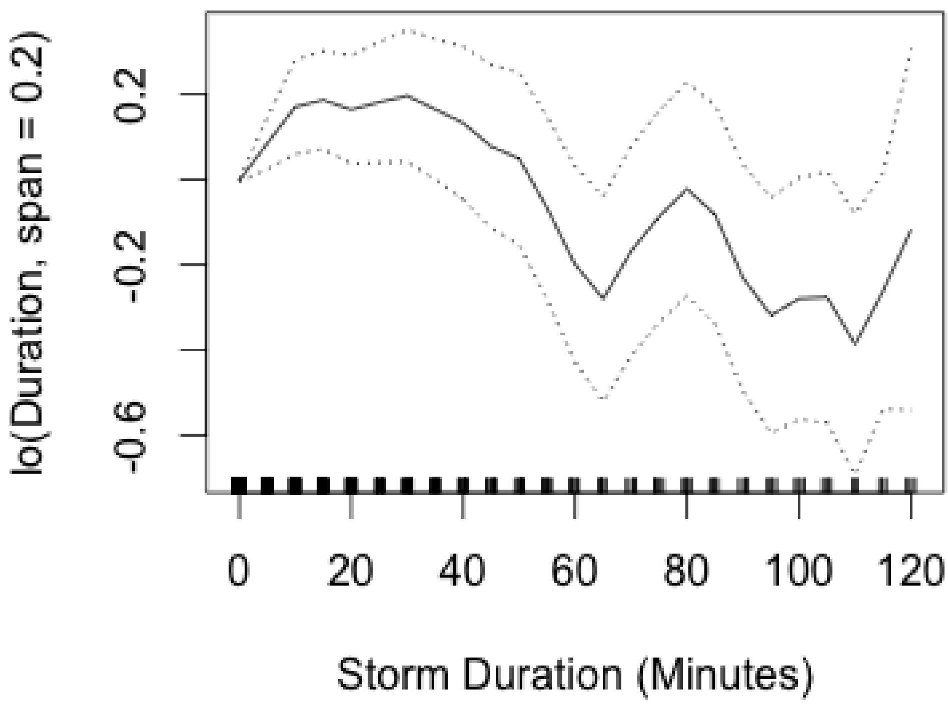

Figure A.6: *Gam plot of sine(wind direction) as a function of storm duration. The dotted lines represent the standard error. The tick marks at the bottom of the figure represent each data points. Large black boxes represent many tick marks and therefore a high density of data points.*

Cosine(Wind Direction)

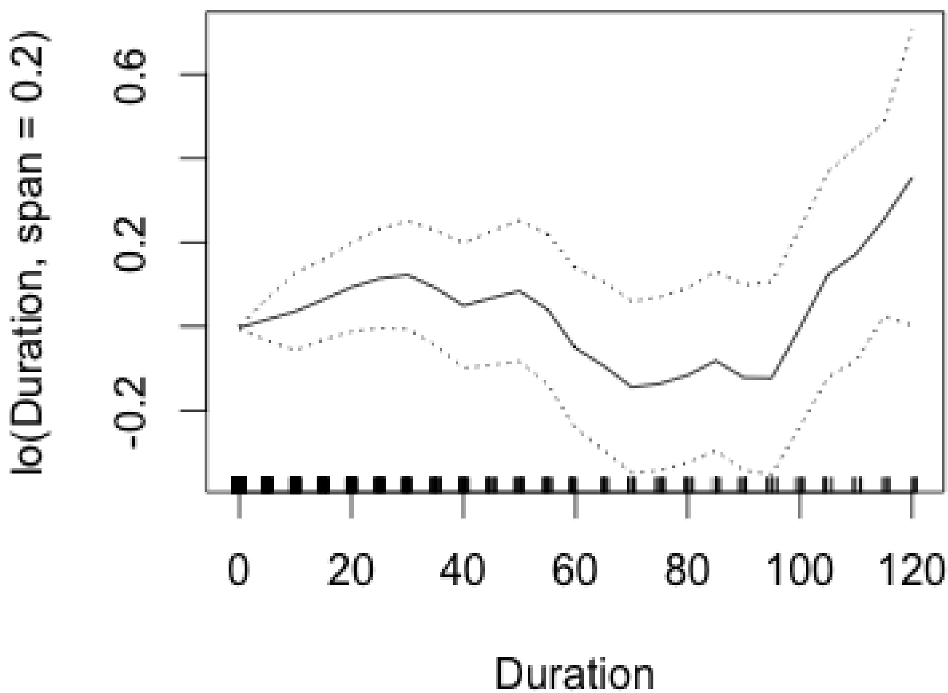

Figure A.7: *Gam plot of cosine(wind direction as a function of storm duration. The dotted lines represent the standard error. The tick marks at the bottom of the figure represent each data points. Large black boxes represent many tick marks and therefore a high density of data points.*