Executive Summary:

In this project our group set out to determine what effect, if any, an air monitor’s distance from the roadway had on the measured air-pollution levels of carbon monoxide, nitrogen dioxide, and fine particulate matter (PM$_{2.5}$). The group was also searching to see if we could identify pollution emitters using wind data. "Fine particles," such as those found in smoke and haze, are 2.5 micrometers in diameter and smaller. We obtained yearly, hourly, and daily data for each pollutant from 1999-2010 from the U.S. Environmental Protection Agency (USEPA). For both carbon monoxide and nitrogen dioxide the air monitors that are less than 30 meters from the road had higher concentrations than the other monitors located further from the road. For fine particulate matter, the monitors located 30 to 100 meters from the road were significantly different from monitors located closer to the road (less than 30 meters) or monitors located further than 100 meters from the road. Regression models were developed which examined the air pollution concentration as a function of elevation, the annual average daily traffic count, the distance the monitor is located from the roadway, the interaction of traffic count and distance from the roadway and the effect of year, which reflects air pollution controls implemented over time. The explanatory variables: elevation, the interaction of traffic count and distance from the roadway and the effect of year are significant for all three pollutants (p<.001). The annual average daily traffic count and distance from the roadway were also highly significant for NO$_2$ (p<0.0001). Finally, the distance from the roadway was also highly significant for PM fine (p<0.0001).
Carbon monoxide and nitrogen dioxide have higher levels in the winter months. A majority of CO monitors reported their smallest concentration at 0.3 parts per million. Only NO\textsubscript{2} was found to have a day of the week effect with weekend days having lower concentrations than weekdays at all three set back distances from the roadway.

When examining the wind data our group was able to use hourly wind speed multiplied by the hourly concentration of each sites individual pollutants to create a wind graph showing pollutant readings at degree intervals of 10. Using Google Images of the individual sites our team was able to see that our wind graphs correctly identified primary and secondary pollutant sources for all 8 case study sites. Each graph showed that the main sources of pollution were roadways within ~300 meters of the road, showing that the pollution readings measured by the air monitors were coming from roadways and automobile traffic, which is the main goal of these roadside monitoring devices.

Our final analysis examines each pollutant at multiple sites with different setback distances in 3 Cities. Are the air monitoring sites decreasing at the same rate at each monitor with a different setback distance? Upon scaling each data point by it’s original starting value we were able to graph and do calculations on points of varying distances and values. After doing a comparison test of each pollutant by distance, we found that their was no significant difference in the overall means of each pollutant between the 3 different distance groups of <30 meters, between 30 and 100 meters and >100 meters. In short, there is no difference in the decreasing rate of pollutants at a significance level of .05.

**Introduction and Background:**

In this project our group set out to answer several questions posed to us by our client, Mr. David Mintz of the U. S. Environmental Protection Agency (USEPA). He asked us to investigate the
possible impact of the distance a monitoring site is from nearest major road on changing pollution levels over time. Are pollution levels improving at faster rate near major roadways than air monitoring sites located further from the roadway? We will try to answer the following questions regarding air pollution near roadways:

- What is impact of setback distance and traffic volume on air pollution monitors?
- Is the difference (or change) in concentration greater on certain days of the week?
- Is there a weekend/weekday effect due to changes in traffic patterns?
- Is the diurnal (hourly) pattern different due to rush hour traffic?
- Are we able to pinpoint pollution sources using pollution levels and wind data?
- Is it possible to determine if each pollutant is decreasing or increasing at the same rate for differing distances and within different cities?

The most important part of this project was to first become thoroughly informed of the pollutants we would be examining. The three pollutants we examined for each monitoring site were nitrogen dioxide, carbon monoxide, and PM$_{2.5}$. All three of these pollutants are important to monitor as they have been linked to an increase in respiratory symptoms, hospitalization for heart or lung diseases, and even premature death. Also, all of these pollutants have at least some connection to the automobile industry and carbon monoxide and nitrogen dioxide and are directly emitted from cars and trucks via exhaust. Fine particulate matter can be directly emitted from sources such as forest fires, or they can form when gases emitted from power plants, industries and automobiles react in the air. All three pollutants are related directly or indirectly to emissions from cars and trucks.
The next step our team took was to examine how and where the USEPA air monitoring sites are placed. The highest numbers of air monitoring sites are placed in highly populated areas that have heavy traffic patterns. These are the larger, more heavily populated cities around the United States. The air monitoring sites are placed according to Federal monitoring regulations and are operated for long periods of time making it possible to evaluate trends over a decade or longer. The air monitoring data from these sites are compared with the annual average traffic data, wind speed and wind direction, and several other factors to develop appropriate regression models.

One final bit of information that is important to the initial research of the project is the understanding that there are 10 different regions as defined by the USEPA. These 10 regions encompass all 50 states, the US Virgin Islands, and Puerto Rico. Each of these regions is broken up geographically and contains several states. Some regions contain US territories as well. Our group decided to choose case study sites from as many of the 10 different EPA regions to see how pollutants behaved in different parts of the country.

Objectives:

The principal objective is to determine what effect, if any, an air monitor’s distance from the roadway had on the measured air-pollution levels of carbon monoxide, nitrogen dioxide, and fine particulate matter. Are concentrations at “near-road” cites changing at a different rate than other monitors? Are day of the week, time of day, season of the year, and year to year changes, different based upon distance from the roadway? Is there a time of day effect on air pollution levels, such as the morning rush hour? What’s the effect of increasing emissions from power plants on nitrogen dioxide and PM$_{2.5}$ levels, while gearing up for the start of the work week on Monday when work resumes in downtown office buildings? The hourly data will allow us to
examine diurnal patterns and see if there are similar patterns for carbon monoxide, nitrogen dioxide and fine particulate matter. Do they all increase during rush hour?

We were interested in the effect of distance from the road on the air monitor. We set out to see how much the air pollutant levels are influenced by distance and how different distances affect their rates of change in air pollution levels. Further, how is it affected by the average annual daily traffic count (AADT), which gives you the average number of cars that pass by a roadway every day? Since all three of the pollutants should be influenced by how much traffic is in their vicinity, we are trying to determine if roads with very high traffic counts have significantly higher pollutant levels. We are interested in the impact of meteorological factors such as wind speed and direction, and other factors, such as the elevation of the site from sea level. Finally we are interested in seeing if each pollutant is decreasing at the same rate within different cities. Since each and every site will have different AADT, and different distances from the road, scaling the data in each city should allow us to see if each site in a city is decreasing at different rates.

**Methods and Analysis:**

To complete our project it was suggested by Mr. Mintz, USEPA, that we choose 5-10 different cities around the United States to use as case studies for our project. To select these cities we had several criteria. The first of which was that we had a “triple site” which we considered a site that recorded all three pollutants. We also wanted to be able to find the distance from the nearest road for each site selected, have the AADT count for each of those roads as well as the elevation. Finally each site must have measured air quality data from 1999 to 2010. The sites we selected were located in:

Elizabeth, NJ

Baltimore, MD
Philadelphia, PA
Denver, CO
Sacramento, CA
Houston, TX
Tucson, AZ
Charlotte, NC

To factor in distance, we used the distance between each air-monitoring station and the closest road to the monitor. We also used government records to determine elevation. As for wind data, we collected data from an online archive, called www.wunderground.com. This site provides us with wind speed and direction (in 10 degree increments ranging from 0-360) for every hour. This data is drawn from the local airport in each city so this wind speed data will have a slight variation from the true wind data at each of the 8 sites mentioned above. To analyze the data, we performed the following functions:

- Stepwise Regression Analysis
- Routine Regression Analysis
- Correlation Analysis
- T Tests
- F-Tests
- Distance Comparison Tests

To test the yearly data we grouped the 8 case study sites into 3 different groups based on distance between the monitor and the nearest road. Those groups were “near-road” sites that were 0-30 meters from the road, “middle” sites that were between 31-100 meters from the road, and finally, “far” sites that were greater than 101 meters from the road. We ended up having 4 near-road sites, 2 middle sites, and 2 far sites. Once this was done we collected the yearly data for each site. To measure CO we took the
annual second maximum 8-hour average, for NO$_2$ we took the 98$^{th}$ percentile of the yearly 1-hour averages, and for PM$_{2.5}$ we took the annual arithmetic mean of the 24-hour averages. This information resulted in the following graphs, which in order are CO, NO$_2$, and PM$_{2.5}$:

Using stepwise regression analysis, it was determined that a number of factors were important in determining the pollution levels of an area. Currently, these factors are:

- Distance from the road
- Elevation of the city
- Traffic of the road
- The interaction/combination of proximity to the road and the road’s traffic
Currently, the strongest predictors for each pollutant seem to be:

- The elevation term when examining CO (SAS output on the far left below)
- Elevation, AADT, and the AADT*Distance terms for NO₂ (SAS output on the far right below)
- Elevation, AADT*Distance, and Distance terms for PM₂.₅ (SAS output on the next page)

These significance levels were <0.0001, which is much smaller than an α level of 0.05, widely considered an indicator of significance. This suggests that elevation is a very important factor in all three pollutants, and that air quality improves as the distance from the road increases, but only to a certain point if the road is too busy. Currently, it can be assumed that pollution levels will be diluted or lowered at a distance, and if possible that monitoring stations should be as close as possible to them. We should also be wary of placing houses too close to the roadways.

| Variable | DF  | Parameter Estimate | Standard Error | t Value | Pr > |t|  | Variable | DF  | Parameter Estimate | Standard Error | t Value | Pr >|t|  |
|----------|-----|--------------------|---------------|--------|------|---|    |----------|-----|--------------------|---------------|--------|------|---|
| Intercept| 1   | 504.52983          | 46.95216      | 10.75  | .0001|   | Intercept| 1   | 3300.56219         | 395.67150     | 8.54   | .0001|
| elv      | 1   | 9.00000000000      | 0.000015015   | 5.67   | .0001|   | elv      | 1   | 6.01226            | 0.000035      | 9.06   | .0001|
| audit    | 1   | -0.00001081        | 0.00000805    | -1.25  | 0.2147|   | audit    | 1   | 0.000000000000000  | 0.000000312   | 11.81  | .0001|
| audit_dist| 1  | -1.1579E+7         | 3.403000E+6   | -3.40  | 0.0010|   | audit_dist| 1  | -0.000000000000000  | 2.875678E+7   | -11.56 | .0001|
| dist     | 1   | 0.000000000000000  | 0.000101      | 0.78   | 0.4399|   | dist     | 1   | 6.92061            | 0.000055      | 3.35   | 0.0012|
| YEAR     | 1   | -0.25013           | 0.02942       | -8.68  | .0001|   | YEAR     | 1   | -1.65503           | 0.15099       | -8.42  | .0001|
Daily data were then examined. The graphs below show the daily maximum hourly value reported each day for the year of 2010. The graphs show all eight sites and display the monitoring sites which are less than 30 meters away with a red dot, a blue dot if the monitor is between 30-100 meters and with a black dot if the monitor is over 100 meters away from the nearest road. Carbon monoxide is shown having higher levels in the winter months. It also showed a majority of CO monitors reporting their smallest concentration at 0.3 parts per million. We also contrasted the three groups based on the distance from the road and each of the three groups was considered different with p-value less than .05. For both carbon monoxide and nitrogen dioxide the air monitors that are less than 30 meters from the road had higher concentrations than the other monitors located at further distances from the road. For fine particulate matter the graph shows that there are no clear differences over the year.
Data were also examined to determine if there were significant differences among the days of the week. Lower levels were expected on the weekend due to less traffic. The box plots below are shown for the four sites that are less than 30 meters from the road. Upon examining graphs and performing statistical tests, both carbon monoxide and fine particulate matter showed no significant differences among the days of the week. For nitrogen dioxide, however, both Saturday and Sunday were found to have lower levels than the weekdays. It is expected that the amount of traffic may be the cause, but it could also be due in part to power plants, which have lower
emissions on weekends. This process was conducted again utilizing data collected from monitors located between 30 and 100 meters from the nearest road and for monitors located over 100 meters from the nearest road. Data showed the same trends as the first analysis, where the carbon monoxide and fine particulate matter showed no differences among the days of the week, while nitrogen dioxide levels had much lower levels on the weekends.

Hourly data were then examined. The graphs below show the value reported each hour from 1999 to 2011 for NO\textsubscript{2} and CO. The fine particulate matter hourly data varied from city to city.
so the graphs are only consistent from 2007 to 2011. There are 2 different types of graphs with
the plain blue ones showing a single city's data for the time period and the multicolored one
showing all cities for the time period. The graphs that show all eight sites display the monitoring
sites which are less than 30 meters away with a red dot, a blue dot if the monitor is between 30-
100 meters and with a black dot if the monitor is over 100 meters away from the nearest road.
Carbon monoxide is shown having higher levels in the winter months. This graph shows a
majority of CO monitors reporting their smallest concentration at 0.25 parts per million. This
value is considered to be half of the monitor's minimum detectable of .5 parts per million. The
horizontal white lines occurring in the NO₂ and CO graphs were due to the monitors only reading
to the 2nd decimal place and thus rounding the concentration readings. By observing the graphs, it
is easy to see the decreasing trend in pollution levels throughout the years. Also by looking at the
cumulative graphs the difference in pollution levels are visible with the monitors closest to the
road having consistently higher readings.
We also contrasted the three groups based on the distance from the road and each of the three groups was considered different with p-value less than .05. For both carbon monoxide and nitrogen dioxide the air monitors that are less than 30 meters from the road had higher concentrations than the other monitors located at further distances from the road. For fine particulate matter the graph shows that there are no clear patterns occurring. The results of the significance test at a .05 level are shown below:

<table>
<thead>
<tr>
<th>Contrast</th>
<th>DF</th>
<th>Contrast SS</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>PM$_{2.5}$ Contrast Analysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>between 30 and 100 vs less than 30</td>
<td>1</td>
<td>355992.1681</td>
<td>355992.1681</td>
<td>4255.41</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>between 30 and 100 vs over 100</td>
<td>1</td>
<td>255890.7157</td>
<td>255890.7157</td>
<td>3058.83</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>less than 30 vs over 100</td>
<td>1</td>
<td>344.9556</td>
<td>344.9556</td>
<td>4.12</td>
<td>0.0423</td>
</tr>
<tr>
<td>CO Contrast Analysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between 30 and 100 vs less than 30</td>
<td>1</td>
<td>572.989487</td>
<td>572.989487</td>
<td>3389.07</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>between 30 and 100 vs over 100</td>
<td>1</td>
<td>2224.023050</td>
<td>2224.023050</td>
<td>13154.5</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>less than 30 vs over 100</td>
<td>1</td>
<td>266.407359</td>
<td>266.407359</td>
<td>1575.73</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>NO$_2$ Contrast Analysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>between 30 and 100 vs less than 30</td>
<td>1</td>
<td>8859229.80</td>
<td>8859229.80</td>
<td>51978.7</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>between 30 and 100 vs over 100</td>
<td>1</td>
<td>18466903.78</td>
<td>18466903.78</td>
<td>108349</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>less than 30 vs over 100</td>
<td>1</td>
<td>938838.52</td>
<td>938838.52</td>
<td>5508.33</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

One of the questions we need to ask is: “Are we measuring the air impacted by over the road, or just the air of the general area?” Obviously, all air monitors do both, but if it does not measure more the air coming from the road and less of the air of the general area, we cannot guarantee that we’ve got the real measurements of the road. As mentioned earlier, we obtained wind speed data from [www.wunderland.org](http://www.wunderland.org), and we have managed to implement wind data into the analysis. The two meteorological variables that affect the pollution dispersed from roadways to the air monitors are wind speed and wind direction.
To utilize the data, we created a variable called “wsxpol,” an abbreviation of wind speed times pollution. Multiplying the number of the pollution measurement by the wind speed creates Wsxpol. For example, if we had CO measurement of 5 parts per million going at 5 miles per hour, the wsxpol value would be 25. To determine the direction of the major sources of pollution, we created graphs featuring wsxpol as the vertical axis and wind direction in degrees as the horizontal axis. To elaborate on the degrees for the horizontal axis, we used a scale from 0 to 360. If the degree was 0 or 360 it was coming from the North. If the degrees were 90, it was considered coming from the East. Due South was 180, and due West was 270. When we plot our pollution data on our earlier graph, we will find plumes of pollution, where the higher numbers congregate in one or more general directions. We can assume that these plumes are in fact due to a source that is emitting most of the measured pollution.

With the direction of our pollution source marked, we then go to Googlemaps, and take a look at the aerial view of the site. For example, if the wsxpol data suggest that we have a major pollution emitter to the South. Below are the wind plumes from Philadelphia along with the map image showing the site as a yellow dot in the upper left hand quadrant of the image.
Our final analysis focused on decreasing yearly rates. For this final analysis the team was seeking to determine if each pollutant, CO, NO₂, and PM₂.₅, were decreasing in similar rates across the country. For this experiment we focused on 3 different case study cities:

Phoenix, Arizona

Philadelphia, Pennsylvania

Sacramento, California

These three sites were chosen because each of them contained one site from each of our 3 previously determined distance categories, which also monitored all three pollutants of interest. By comparing the findings from these 9 sites other external factors such as distance from the road, AADT, and wind speed are negated. Upon gathering the 9 sites yearly pollutant averages from 1999-2000, we scaled each data point in an effort to normalize the data. To do this the team divided each yearly pollution value by the starting years value. This made each pollutant begin at
1.0 and go up or down compared on the variance from year to year. Our group called this scaling data by COper, NO2per, and PM2.5per. The resulting graphs for Philadelphia are below:

**Carbon Monoxide**

![Carbon Monoxide Graph](image1)

**Nitrogen Dioxide**

![Nitrogen Dioxide Graph](image2)

**PM$_{2.5}$ Fine Particulate**

![PM$_{2.5}$ Fine Particulate Graph](image3)
After scaling this data down, the team used SAS analysis software to test if there were any differences between the distances of each pollutant at a significance level of .05. The findings show that only one of the comparison tests between the closest nitrogen dioxide readings and those greater than 100 meters had a significantly different mean at a significant level of .05. The other 8 comparison tests show that there is no noticeable difference in the means at varying distances. The values of the t-test findings are shown below:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Error</th>
<th>t Value</th>
<th>Pr &gt;</th>
<th>t</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>COper</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 vs 2</td>
<td>0.03985240</td>
<td>0.06605807</td>
<td>0.60</td>
<td>0.5479</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 vs 3</td>
<td>0.11309472</td>
<td>0.06871205</td>
<td>1.65</td>
<td>0.1034</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 vs 3</td>
<td>0.07324232</td>
<td>0.06605807</td>
<td>1.11</td>
<td>0.2706</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NO2per</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 vs 2</td>
<td>-0.05775682</td>
<td>0.03818177</td>
<td>-1.51</td>
<td>0.1340</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 vs 3</td>
<td>-0.09846947</td>
<td>0.03971579</td>
<td>-2.48</td>
<td>0.0151</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 vs 3</td>
<td>-0.04071264</td>
<td>0.03818177</td>
<td>-1.07</td>
<td>0.2893</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PM2.5per</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 vs 2</td>
<td>-0.01464136</td>
<td>0.03462286</td>
<td>-0.42</td>
<td>0.6734</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 vs 3</td>
<td>-0.01372756</td>
<td>0.03601388</td>
<td>-0.38</td>
<td>0.7040</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 vs 3</td>
<td>0.00091381</td>
<td>0.03462286</td>
<td>0.03</td>
<td>0.9790</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Conclusions:**

We have shown that the distance from the road does indeed have an impact on air pollution levels, but it is not the sole factor. The average annual daily traffic count and elevation in combination with distance from the road are the factors that influence the pollution levels. When examining the hourly pollution data the team found an overall decreasing trend through the years for each of the three pollutants. This was significant at a .05 significance level. The windspeed * pollution level graphs allowed the team to identify primary and secondary pollution sources for
each of the cities that we investigated. After examining 3 different cities, with each having three monitoring sites, the team found that only the NO$_2$ readings between close monitors compared to far away monitors were significant at an alpha level of 0.05. The air pollution monitoring devices the EPA uses to track pollution near roadways are doing their job, getting readings from primary roadways. The one overlaying positive that our group has found is that the regulations the EPA has put in place to limit pollution from automobile traffic are working, shown by the decreasing trends seen at the hourly, daily, and yearly levels.