<u>Putney Street AQ Sensor Analysis.</u>

This short paper will analyse data found from three E-MOTE AQ sensors placed in Putney High Street in London in 2017. The E-MOTES detected three key pollutants, notably Carbon Monoxide, Carbon Dioxide and Nitric Oxide, and other meteorological factors, such as Temperature and , between July and October 2017.

<u>General Trends 352 Sensor.</u>

I have plotted a time series plot for all three of the pollutants from the '352' sensor over the entire duration of the detection to detect any potential trends or seasonality. Figure 1 contains a graph of each of the time series for carbon monoxide, carbon dioxide and nitric oxide.

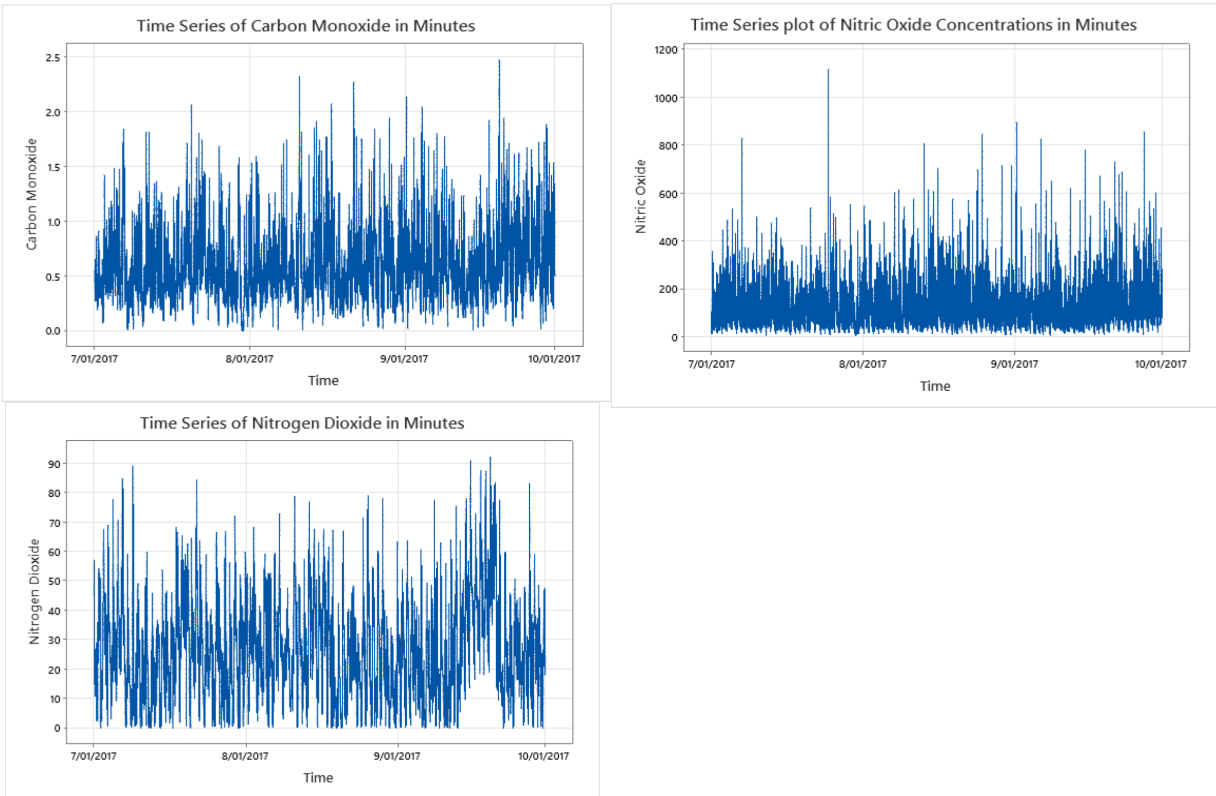

Figure 1: Basic Time Series Plots

All three of the time series plots are relatively stable, with no obvious trends being observed from any of the plots. There are no sudden shifts in the data with the data having clear peaks and troughs but keeping a consistent trend. However, a slight sudden shift can be observed in the levels of Nitrogen Dioxide between the 13th and the 21th of September, which is not observed for the other pollutants. Also, there is an obvious outlier on the nitric oxide time series with a very high amount (1116.14) being detected at 21:54 on the 24th of July. As this rapidly rises from 649.54 the minute before and 840.36 two minutes later, this observation could be a potential error or could be caused by a large plume of pollutant. All of the time series plots seem to show some daily seasonality however it is difficult to observe this when using such a large data set.

Therefore, to judge daily seasonality figure 2 contains a smaller sample of data spanning one week between 00:00 on the 1st of July 2017 and 00:00 on the 8th of July 2017.

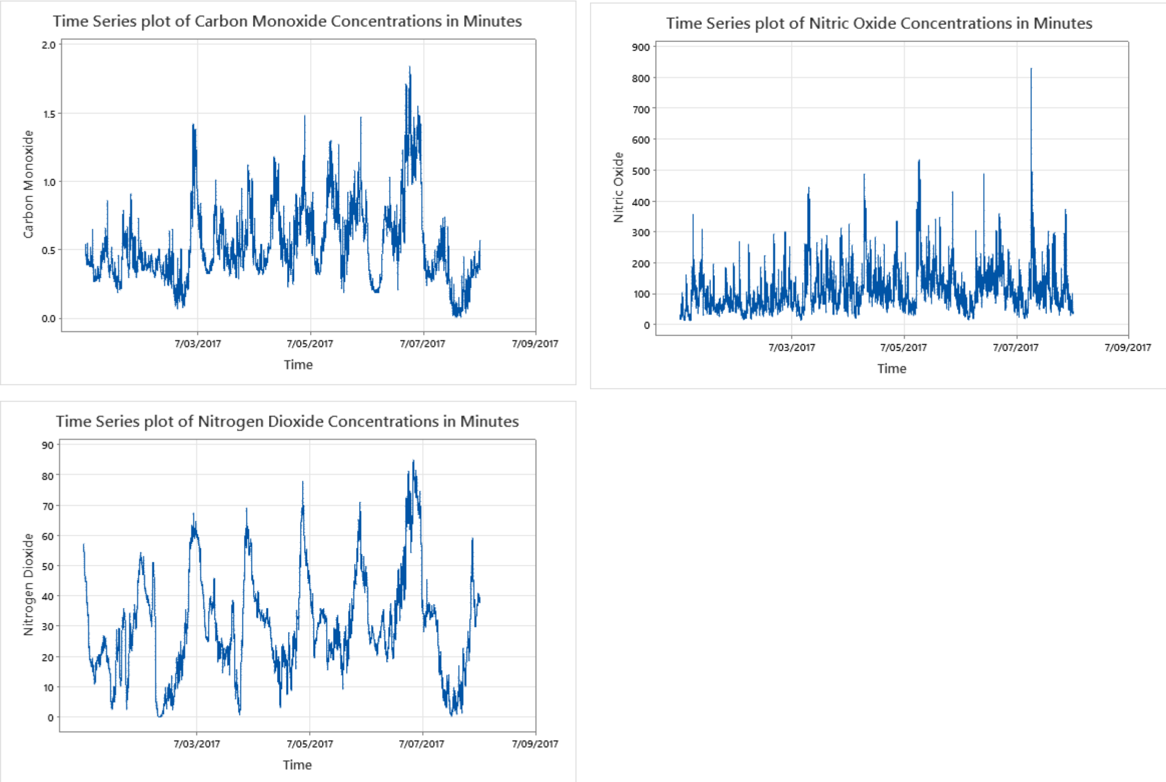

Figure 2: Week Time Series Plots

On these time series plots on figure 2 the seasonality is far simpler to note, with the peaks and troughs of the graphs having irregular magnitudes but occurring at regular intervals every day.

From Figure 2, the carbon monoxide time series plot showed a relatively strong seasonality, with regular peaks and troughs and similar magnitudes. The peaks occurred typically around 7:30 and more significantly around 21:30. The weekend in this sample was the 1st and 2nd and the data from these days showed similar, yet more volatile, intervals but had much smaller magnitudes, this is most likely due to the lack of rush hour work traffic taking place on those days creating less pollutant.

The nitric oxide plot was very volatile with there being many different peaks and troughs within the day. However, although volatile the plot showed similar trends and seasonality each day with the more significant peaks occurring between 6:00 and 9:00. This plot also had peaks and troughs with smaller magnitudes over the weekend.

For the nitrogen dioxide plot there are two peaks each day, occurring from 5:30 – 7:30 and 21:00 - 23:00, with the 21:00-23:00 evening peak being much larger. The troughs of the graph are more irregular however there is a regular trough between 11:30 and 13:30. Similarly to the carbon monoxide plot this time series showed similar, yet more volatile, intervals but had much smaller magnitudes which is likely due to the same reason.

ARIMA Forecasting

ARIMA forecasting can attempted on the time series pollution data to extrapolate the data and make future predictions for the air quality levels. Therefore, using an ARIMA(5,0,0) model an attempt to forecast the levels of each pollutant has been made in figure 3.

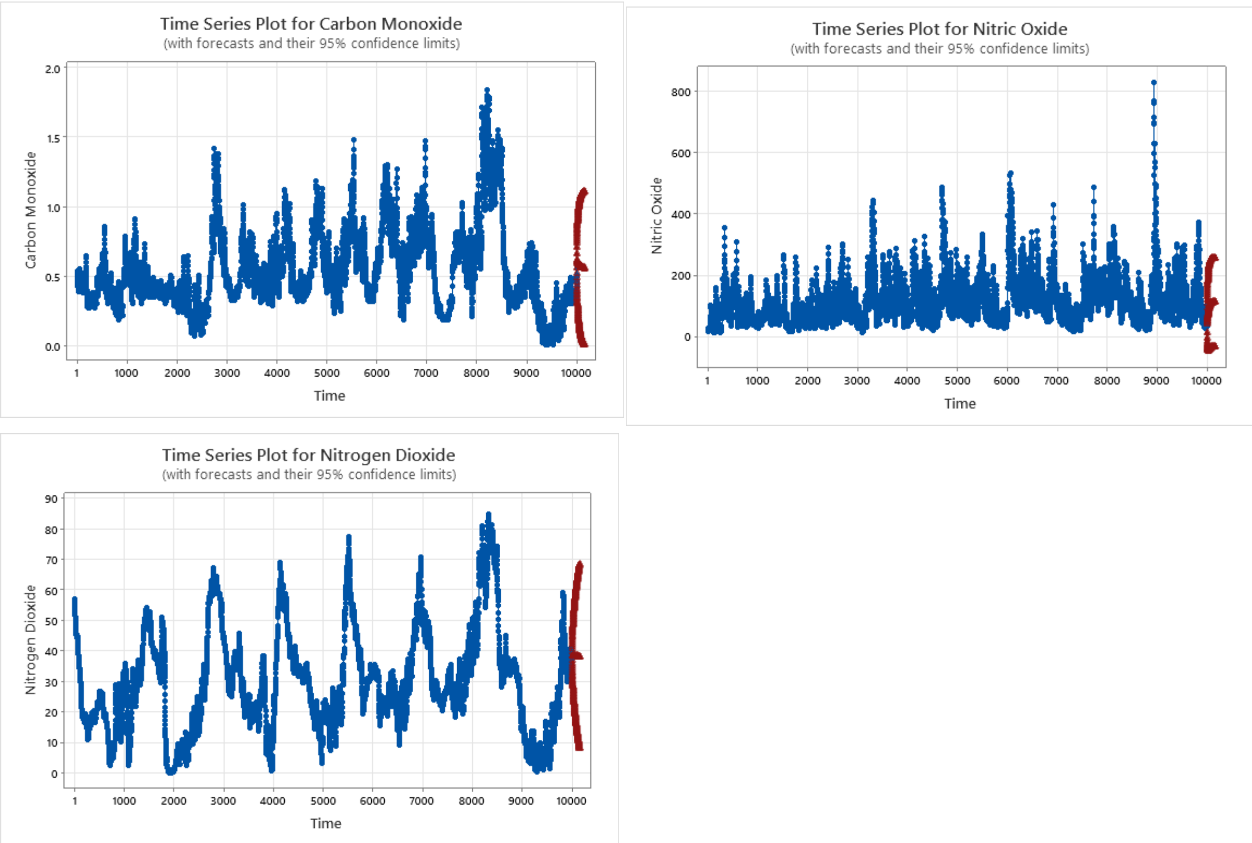

Figure 3:Time Series Plots and Forecasts

Figure 3 outlines the forecasts and the upper and lower limits for each of the pollutants, with the forecasts being outlined by the red lines. They all show the next 150 minutes of data, which if accurate could be used for AQ prediction applications. All three forecasts had relatively large error regions, which is to be expected due to the volatile nature of the concentrations of pollution.

When analysing the ACF, PACF and the residual plots for the data, the quality of the fit of the ARIMA model is inconclusive. However, this method could be refined and used for more seasoned data, to create accurate forecasts for AQ prediction.

Impact of Meteorological Factors on Pollutants

The E-MOTES placed on Putney High Street measured temperature and humidity alongside the three pollutants. Therefore this data could be used to see any potential correlation between each pollutant and factor. To judge this regression analysis has been performed for each pollutant and factor combination.

Figure 3 shows a scatterplot of each pollutant against temperature to see any correlation. For this regression analysis I have initially used the full dataset, however my initial prediction I that the sample may need to be used again as regression doesn't work as well with larger datasets.

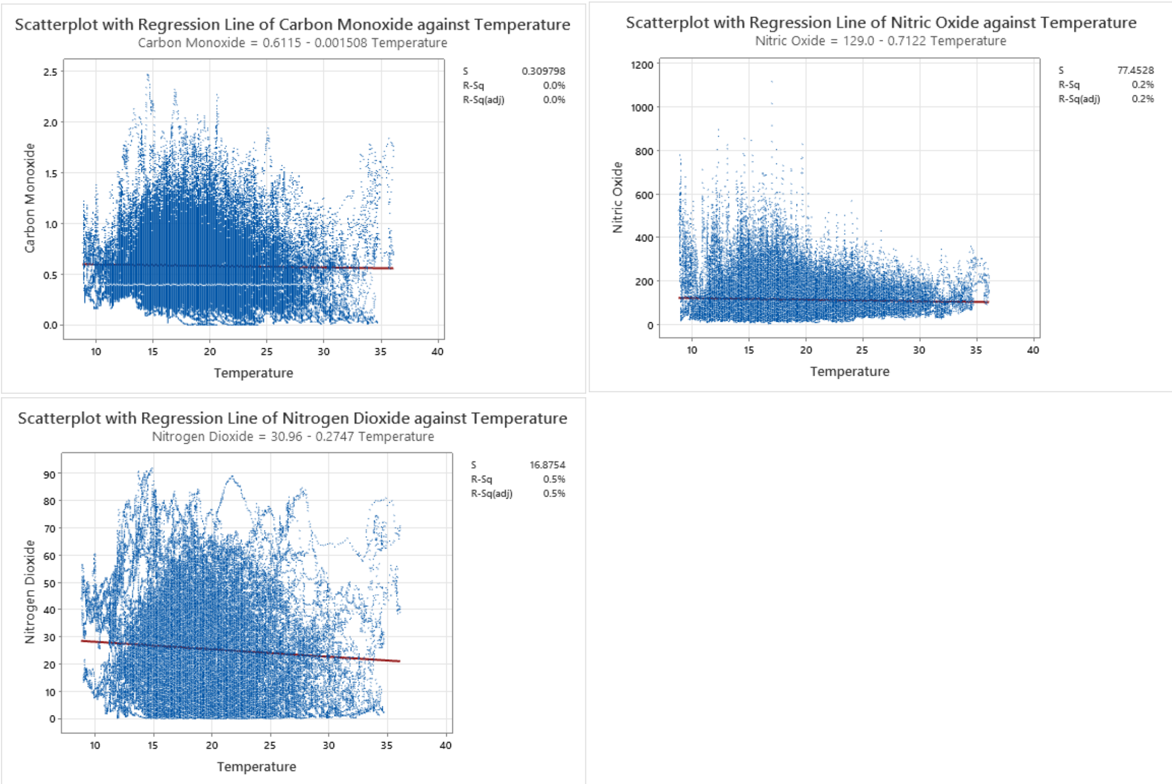

Figure 3: Temperature Regression Plots

All of these three regression plots were very inconclusive as the regression lines had small gradients (0.0015 , 0.71 and 0.27) suggesting that there is no real correlation between the pollutants and temperature. However, the R^2 and S values were all very undesirable with all of the R^2 values being far less than 1%. This would suggest that the regression lines given are a very bad fit and reaffirms that there is no correlation between the pollutants and temperature.

This poor response from regression could be due to the size of the dataset or errors and inaccuracies in the sensors but it is most likely due to there being a large number of uncontrollable variables that also have an effect on the pollutant, as the sensors were tested in the field as opposed to a controlled lab.

Figure 4 shows similar regression plots, but with the pollutants tested against humidity instead of temperature to test their correlation and therefore the effect of humidity on pollution levels.

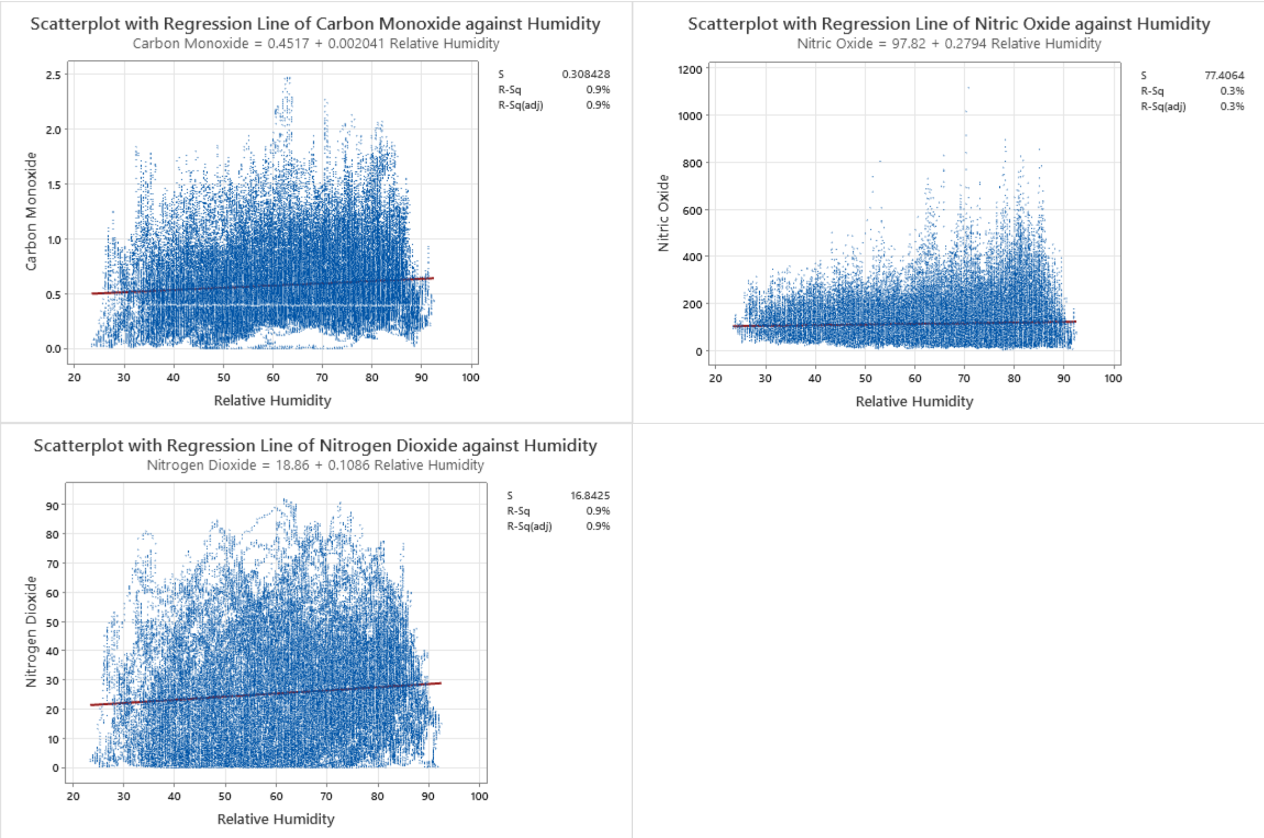

Figure 4: Humidity Regression Plots

Very similarly to the temperature regression plots, the humidity regression plots showed very little correlation between the levels of pollutant and humidity. All of the regression lines had very small gradients and the S and R^2 values were very low. This also suggests that the levels of humidity have little effect on the pollutants. However the poor regression model could be due to the size of the dataset or errors and inaccuracies in the sensors, but it is most likely due to there being a large number of uncontrollable variables, very similar to the temperature plot.

General Trends of 355 and 356 Sensors.

I've plotted a time series graph for the data from the 355 and 356 sensors to judge any potential seasonality or interesting trends.

Figure 5 contains the three time series plots from the 355 sensors for all three pollutants over time.

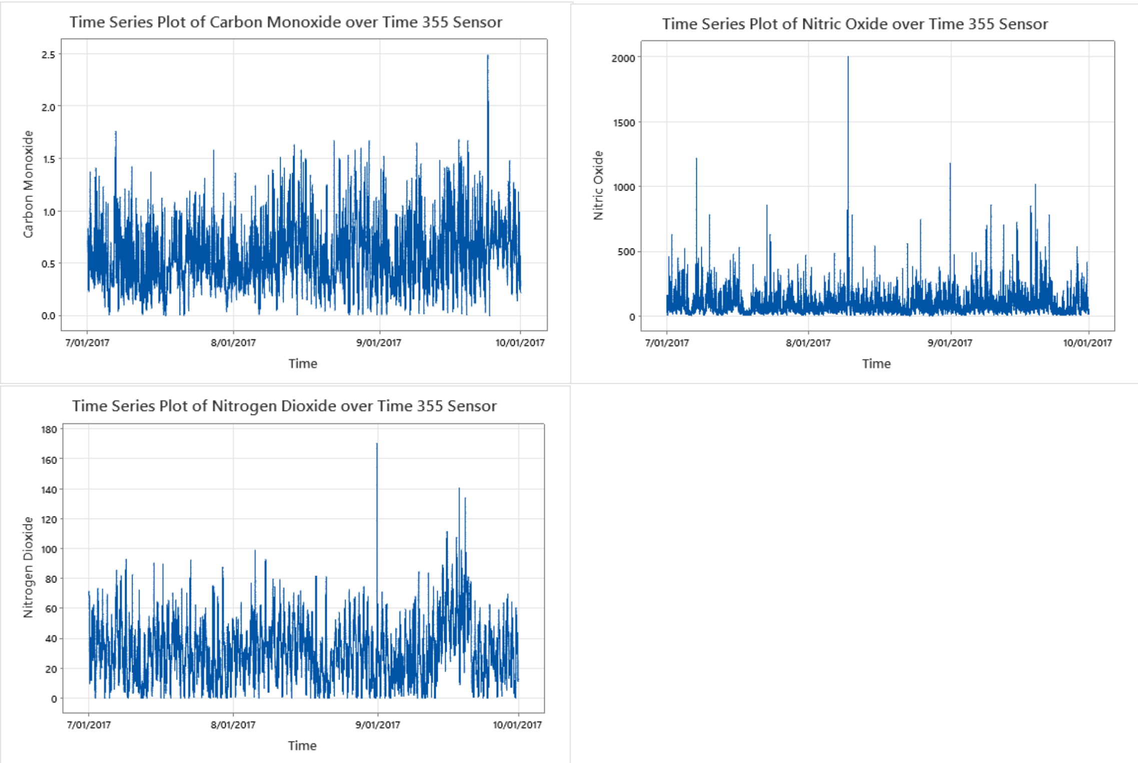

Figure 5: 355 Sensor Time Series plots

There is an unexpected result on all three plots at 00:14 on the 24th of August for the carbon monoxide, 12:10 on the 9th of August for the nitric oxide plot and 16:37 on the 31st of August for the nitrogen dioxide plot. This suggests that there was either an error or just a very large plume of pollutant near the sensor at those times. The nitric oxide spike on the 9th of August is very significant due to its sheer size, however this may be due to an error as the spike isn't registered by any of the other two sensors. Also there is a noteworthy shift in the nitrogen dioxide plot between the 13th and 21st of September, where the pollution levels are higher. This shift is also detected by the 352 sensor which suggests that the shift isn't an error but just a period of high nitrogen dioxide. Therefore, it would be worthwhile noting any changes, i.e. roadworks or events, that occurred between the 13th and 21st of September as it could highlight a potential impacting factor of nitrogen dioxide.

Figure 6 contains similar time series plots but for the 356 sensor for all three of the pollutants over time.

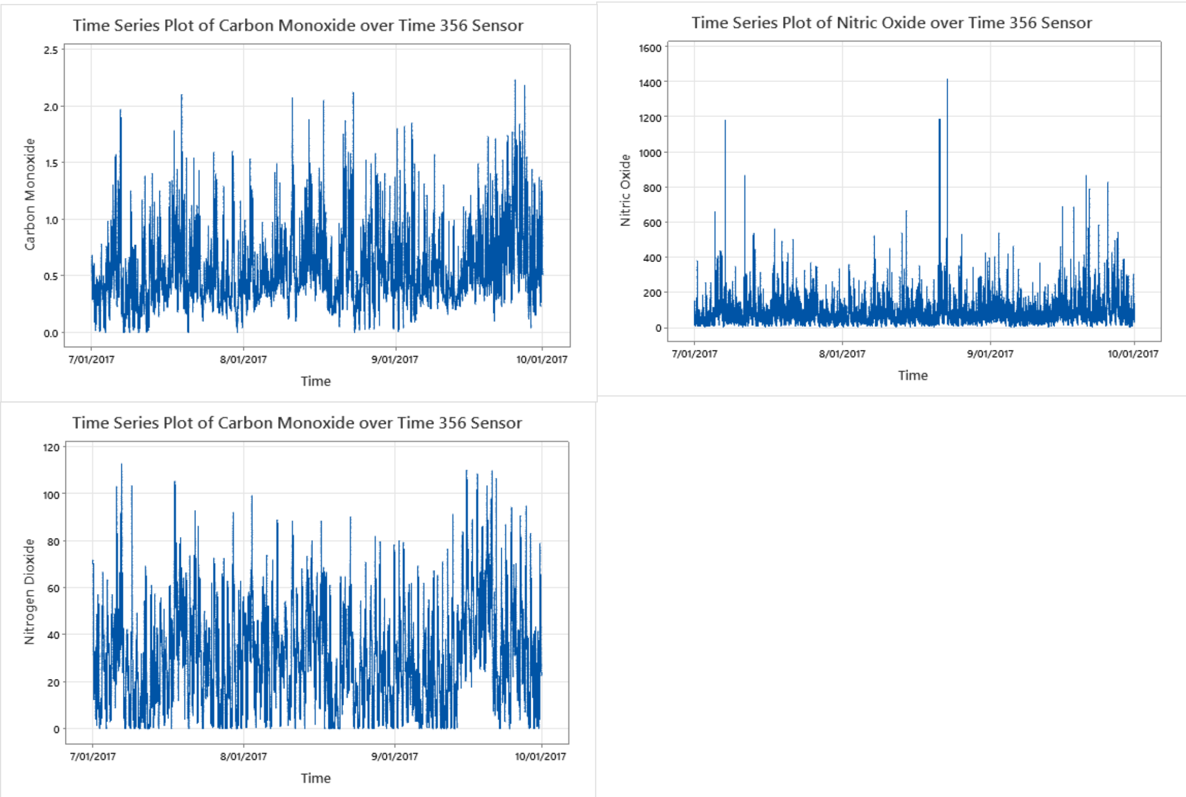

Figure 6: 356 Sensor Time Series plots

All of the pollutants from the 356 sensor showed a very similar trend to that of the 352 and the 355 sensors which indicates that the sensors themselves have high validity as although the sensors are placed apart on a long street, they should demonstrate similar trends. The only real sudden spikes occurred for the nitric oxide pollutant on the 22nd of August at 20:50 and the 21st of August at 6:29. Throughout all three of these analysis' the nitric oxide plots have had the bulk of the sudden spikes, this would suggest that this particular pollutant is more volatile, or the sensors detect more errors from that pollutant, however the former is more likely. The sudden shift in nitrogen dioxide plot between the 13th and 21st of September was also detected by the 356 sensor which further proves its validity.

Effect of Traffic on Pollution on Putney High Street

On the 30th of July between 7:00 and 20:00 there was a bike race, which resulted in Putney High Street being closed. Therefore, this provides a good opportunity to analyse the effect of traffic on the street by comparing that day with the trends previously ascertained in this paper.

Figure 7 contains time series plots for the 30th of July 2017 from the 352 sensor for three of the pollutants over time.

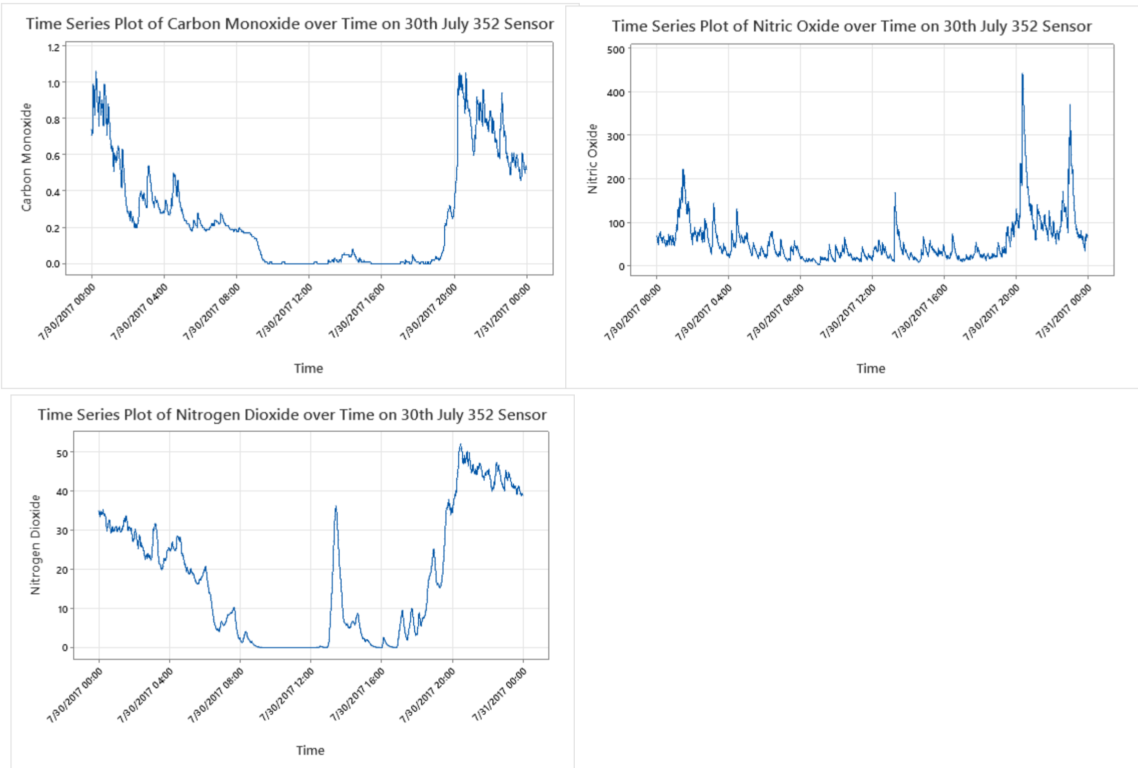

Figure 7: 30th July Trafficless Time Series Plots

From figure 7 it can be immediately noted that the previously noted pattern of spikes between 6:00 and 9:00 and 21:00 and 23:00 but a constant but volatile trend on other times is not followed in the day without traffic. On this day the levels of pollutant dropped and stayed very constantly low for both the carbon monoxide and nitrogen dioxide (except for one spike). The lack of volatility and usual trend when there is no traffic suggests that the both the volatility and the daily seasonality is caused by traffic. Aside from the difference in trend, it is interesting to note that the levels of concentration of pollution in all three pollutants is far lower between 7:00 and 20:00, when there is no traffic. This would suggest that not only does traffic cause the trend of pollution, it also is the main cause of the pollution itself. This would explain why there was no correlation between temperature and humidity and levels of pollution although in theory they should, as the levels of traffic had a much larger impact on the pollution levels and therefore negated the much smaller effect of temperature and humidity.

Conclusion

Overall the E-MOTE senors placed on Putney High Street gave similar (and therefore potentially valid) results, especially for the carbon monoxide and nitrogen dioxide pollutants. The levels of pollution were very volitaile (to be expected with the sporadic nature of the impacting factors) but showed a clear daily seasonality and trend which could be observed from all pollutant data. The effect from temperature and humidity was negligible, however this was most likely due to the changes in temperature and humidity being negated by other changes in road activity. Testing in a controlled environment would be needed for assessment of the effect of temperature and humidity. It was also perceived that road traffic is the largest cause of the daily seasonality of the pollutant and the amount of pollutant itself, which suggests that road traffic is the biggest impacting factor on pollution. Therefore, it is advised that pollution levels are high when traffic is high.