access:UTS

Additional Analysis of ARTC Data on Particulate Emissions in the Rail Corridor.

Prepared by Louise Ryan, Distinguished Professor of Statistics, and Alan Malecki, Doctoral Student, on behalf of accessUTS Pty Ltd for Environmental Protection Authority NSW

This report was prepared by Louise Ryan in good faith exercising all due care and attention, but no representation or warranty, express or implied, is made as to the relevance, accuracy, completeness or fitness for purpose of this document in respect of any particular user's circumstances. Users of this document should satisfy themselves concerning its application to, and where necessary seek expert advice in respect of, their situation. The views expressed within are not necessarily the views of the Environment Protection Authority (EPA) and may not represent EPA policy.

August 2015
Project No: 2015000889

accessUTS Pty Limited
PO Box 123
Broadway NSW 2007 Australia
Tel +61 2 9514 1916 Fax +61 2 9514 1433

ABN 55 098 424 312
accessUTS is a controlled entity of the University of Technology, Sydney

Contents

Additional Analysis of ARTC Data on Particulate Emissions in the Rail Corridor
July 2015

Louise Ryan, Distinguished Professor of Statistics
and
Mr Alan Malecki, Doctoral Student
University of Technology Sydney

1 Executive Summary

This report extends Professor Louise Ryan's February 2014 report [1] for the NSW Enviornment Protection Authority. It describes a further analysis of data from the Australian Rail Track Corporation (ARTC) on particle emissions from coal and other trains in the Hunter rail corridor. This extension is motivated by the release of further data in the form of precipation records and the number of locomotives pulling each train. Precipitation data were made available from a monitoring station in Maitland that recorded rain (in mm) on a daily basis and another monitor in Cessnock that recorded data on a 30 minute basis.

The analysis suggests that the number of locomotives has little influence on the increased particulate levels associated with various types of train passings. One important caveat is that information on the number of locomotives per train is likely to be reported with some error. The analysis also shows that particulate levels are significantly influenced by whether or not it had rained the previous day in Maitland. After accounting for Maitland rainfall, Cessnock rain had no significant influence on particulate levels.

While bearing in mind the ARTC caveat concerning data reliability, the lack of association between particulate levels and numbers of locomotives dispels, to some extent, the hypothesis that diesel exhaust explains a large proportion of the observed increases in particulate levels associated with train passings. The strong association with previous day's rain in Maitland suggests that a key mechanism for the increased particulate levels was stirring up by passing trains of dust particles that had settled previously on the tracks.

Appendix 2 of this report also provides requested additional details requested by EPA in relation to Professor Ryan's February 2014 report.

2 Introduction and Background

In late 2013, Professor Ryan from University of Technology Sydney was asked by the NSW Environmental Protection Authority (EPA) to critique the analysis of data from a study that had been designed to assess the impact of train traffic on particulate levels in the Hunter Valley. The study had involved data collected via a continuous monitoring station that was installed to measure particulate levels in the rail corridor adjacent to tracks carrying different types of trains. Each train passing recorded, along with details such as the type of train (freight, loaded coal, unloaded coal, passenger or unknown), its average speed while passing, the number of locomotives pulling it and the time spent passing the monitor. Data on wind speed and direction were also available, though not for all timepoints. Professor Ryan expressed concerned about the statistical methodology that had been implemented by Katestone, the company that had done the initial analysis. The EPA then asked Professor Ryan to re-analyze the data. She produced a report in February 2014 and subsequently presented the findings to a citizens group in Newcastle. The goal of prior analyses has been to determine:

1. Whether trains operating on the Hunter Valley rail network are associated with elevated particulate matter concentrations; and
2. Whether trains loaded with coal have a stronger association compared with unloaded coal trains or other trains on the network.

Although Professor Ryan's analysis method differed from Katestone's, her final conclusion was consistent with theirs, namely that coal and freight trains were associated with significant increases in the levels of particulates measured in the air. However, there was no indication that loaded coal trains had a stronger association than other kinds of trains. In fact, it appeared that loaded coal trains were associated with lower dust levels than both unloaded coal trains and freight trains. Professor Ryan also felt that there were some indications that diesel exhaust may be a significant contributor and she recommended some additional analysis to explore whether information regarding the number of locomotives on each passing train could help explain the patterns.

This follow-up report has been commissioned to explore if the added data on locomotive numbers has an affect on dust levels. The report also includes information regarding precipitation, which had not been available for the previous analysis.

The same statistical methods as in the previous report by Professor Ryan have been utilized. More details are given presently, but briefly, the analysis was based on regression modelling with a consideration for the likelihood of serial correlation due to the time series structure of the data. The data are analysed at the individual level, and as such there is no loss of information due to aggregation. This permits the use of covariates indicating train type, the number of locomotives pulling it, wind speed, precipitation, as well as additional variables reflecting time of day, day of week and other temporal effects.

The ensuing report presents Professor Ryan's findings. It should be noted that many of the same caveats and concerns apply as for the original report. In particular, use of data from only a single monitoring site made it difficult to generalize the results. That being said, a continuous time series at this single site can still provide useful insight into the questions of interest.

3 The Data

Katestone provided data related to particulate levels measured by the monitoring system, as well as additional data related to passing trains, wind and precipitation.

Particulate data: The monitoring system was set up to measure particulate levels every 6 seconds throughout each day over a 61 day period, yielding a maximum total number of 24*60*60/6 = 14,400 sets of measurements each day. As described in Katestone's final report, however, many of the days collected substantially less than this amount of data. Two days (December 2nd and 16th) had no measurements while an additional three days (December 5th, December 15th, January 1st and January 2nd) had less than 1000 data points (100 minutes of data). Our analysis reports on the subset of 55 days that had at least 1000 monitoring measurements. On these days, the mean number of observations was 10,990 (approximately 18 hours) and 48% of the days had at least 22 hours of monitoring data. As discussed in the Ryan (2014) report, the particulate data were strongly skewed. A log transformation of the data was considered and subsequently implemented as it improved the modelling process.

Wind data: Data on concurrent wind speed and direction were available for approximately 75% of the particulate observations. These observations were spread over 45 days, with an average of 17.5 hours of wind speed and direction data concurrent with particulate data monitoring, per day. The first day of observation (November 30th) had only 867 concurrent wind observations. Thus, we restricted our analyses involving wind speed and direction to the subset of 44 days with at least 1000 data points.

Train data: Over the course of the study period, a total of 5601 trains passed by the monitoring station, with a median of 137 trains per day. Table 1 in the original report shows the breakdown by different train types, along with the duration (in seconds) of each train passing and the average train speed.

Further information was procured by EPA from the ARTC regarding the number of locomotives pulling each train. Unfortunately, data on the number of locomotive pulling the train was available for only 4382, or approximately 78%, of the trains. As shown in the following table, 2067 Empty Coal trains, 1788 Loaded Coal trains, 309 Freight trains, 110 Passenger trains and 108 Unknown trains have locomotive information. Thus, locomotive information was complete or almost complete for all train types except passenger trains. The number of locomotives pulling each train type ranges from 1 locomotive to a maximum of 7.

Table 1: Number of Locomotives by Train Type

	Number Of Locos								
Train Type	1	2	3	4	5	6	7	Missing	Total
Empty Coal	110	827	1057	72	1	0	0	0	2067
Loaded Coal	134	670	927	49	5	2	1	0	1788
Freight	10	121	170	7	1	0	0	0	309
Passenger	78	19	6	4	1	0	2	1216	1326
Unknown	15	53	34	5	1	0	0	3	111

In discussions with the ARTC, it became clear that the locomotive data could not be guaranteed as completely reliable. Hence statistical analyses need to be interpreted somewhat cautiously.

Precipitation data: Precipitation measurements were supplied for each day over the study period at the locations of Cessnock and Maitland. At Cessnock, rain was recorded as a cumulative measurement (in mm) staggered at 30 minute intervals. These measurements could easily be converted into measurements of rainfall reported in successive 30 minute intervals. Although closer to the Metford monitor location, the Maitland rain measurement was not as informative. This is attributed to the single rain measurement supplied indicating the total rainfall for the particular day in Maitland. Thus the analysis of Maitland rain could only be considered at the daily level.

Of the 55 days with adequate particulate monitoring data, rain was observed at Maitland on 20 days and at Cessnock on 19 days. On the days when it rained in Maitland, the average rainfall was 9.2 mm.

We created a number of different rain variables for consideration in subsequent modelling, including whether or not it had rained the previous day in Maitland and Cessnock, the amount of rain at Cessnock during the 30 minute period within which a measurement was taken, as well as whether it had rained in the previous 30 minutes.

Statistical Modelling Strategy

As in our original report, we analysed the data using a variant of linear regression, with outcome variable corresponding to one of the four particulate measures (PM1, PM2.5, PM10 or TSP) transformed to the log scale. The advantage of regression analysis is that it allows for simultaneous adjustment with respect to various confounding factors that may otherwise bias or distort the analysis. For example, loaded and unloaded coal trains were more commonly seen during the early morning and evening hours, compared with passenger trains which tended to be more frequent in daytime hours. Since particulate levels are likely to vary over the course of the day, a naive comparison of particle levels associated with the different train types would be biased. A regression model that includes appropriate terms corresponding to time of day and day of study provides an adjustment that puts train type comparisons on an equal footing. We used an advanced version of linear regression analysis, the so-called generalized additive model, which allows for the flexible modelling of continuous functions using splines and other kinds of functions. All statistical analysis was conducted in the statistical package R[2], using a function called *gam*, which is part of the package mgcv developed by Wood[3]. One of the variables included in the model was a smooth spline function of time of day. Inclusion of this term in the model allowed for the likelihood that there might be a diurnal pattern in the data. Similarly, we included another smooth spline function of day of observation. Inclusion of this term allowed for the possibility that there might be a longer term pattern. We explored the inclusion of day of week indicators, but in the end decided to exclude these since they were not strongly significant and our models included other terms that adjusted for temporal effects.

We first conducted analyses excluding the wind speed and wind direction variables, since these data were not available on all study days. As something of a sensitivity analysis, we repeated our final models on the subset of data where wind data were available.

Exploratory analysis of the residuals from our regression models suggested the presence of strong autocorrelation. This is to be expected in data such as these which represent a long time series of observations measured closely together in time. While there is a variation of the *gam* function available that allows for autocorrelation, we found that because of the magnitude of our analysis datasets, the models were extremely slow to run or did not run at all. Consequently we used a bootstrap [3] to adjust the standard errors computed in our models for autocorrelation. In particular, we used a specific variant, the so called blocked bootstrap[4], which has been developed for use with serially correlated data. In our setting, a particularly simple implementation of the blocked bootstrap was achieved by resampling days. Standard errors were computed based on a total of 50 bootstrap samples, although for computational reasons, the number was reduced to 20 for some sensitivity analyses.

The large size of the analysis dataset meant that each model took several minutes to run and the bootstrap several hours. In order to explore the data and to identify suitable models, we first ran analyses using ordinary linear regression with polynomial terms in time of day and day of study. Our final analyses were then repeated using the more computationally intensive but accurate bootstrapped *gam* function. In general, we found that linear regression models gave qualitatively similar results to the *gam* analyse (aside from standard error estimates) and hence they provided a useful practical approach to exploratory analysis.

We took as the starting point for our models the final models that had been developed and reported on in the first report by Professor Ryan. These models included the following variables:

- For each train type (loaded coal, unloaded coal, freight, passenger, unknown), an indicator of whether or not a train was passing at the time a measurement was taken;

- For each train type (loaded coal, unloaded coal, freight, passenger, unknown), an indicator of whether or not a train had finished passing in the last 5 minutes;
- Smooth terms based on penalized splines in seconds since midnight (this allowed for diurnal patterns);
- Smooth terms based on penalized splines by day (this allowed for day to day variation);
- day of week indicators.

Previous models had also included indicators of whether or not a train would be arriving in the next 1 to 3 minutes. However subsequent discussions suggested that this variable did not really make sense hence it was omitted in this report. To explore the effect of the rain variables as well as the locomotive variables, we added these to our final model, then used variable selection techniques to reduce the model to a more parsimonious model that described the data adequately.

4 Results

We start out with presenting some new plots that show the relationship between the different types of particule measurements. These plots were not included in the original report, but were suggested by some of the members of the Hunter Valley Citizens group when the first set of results were presented to them in 2014. Figure 1 shows a plot that superimposes measurements of PM1, PM2.5, PM10 and TSP. Data is shown at the individual level, with just over a minutes worth of observations plotted. The plots shows very clearly how the different particulate measurements rise and fall together.

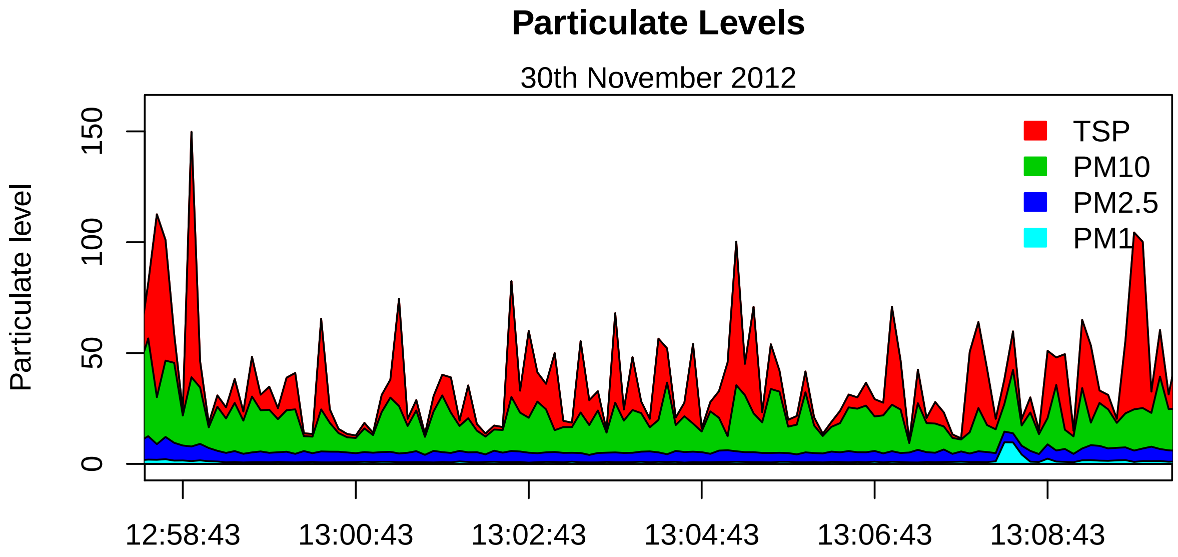


Figure 1: Particulate levels - 10 Minute Period

It is interesting to see in this Figure how TSP and PM10 levels track fairly close together. Figure 2 shows the same plot, but with data transformed to the log scale.

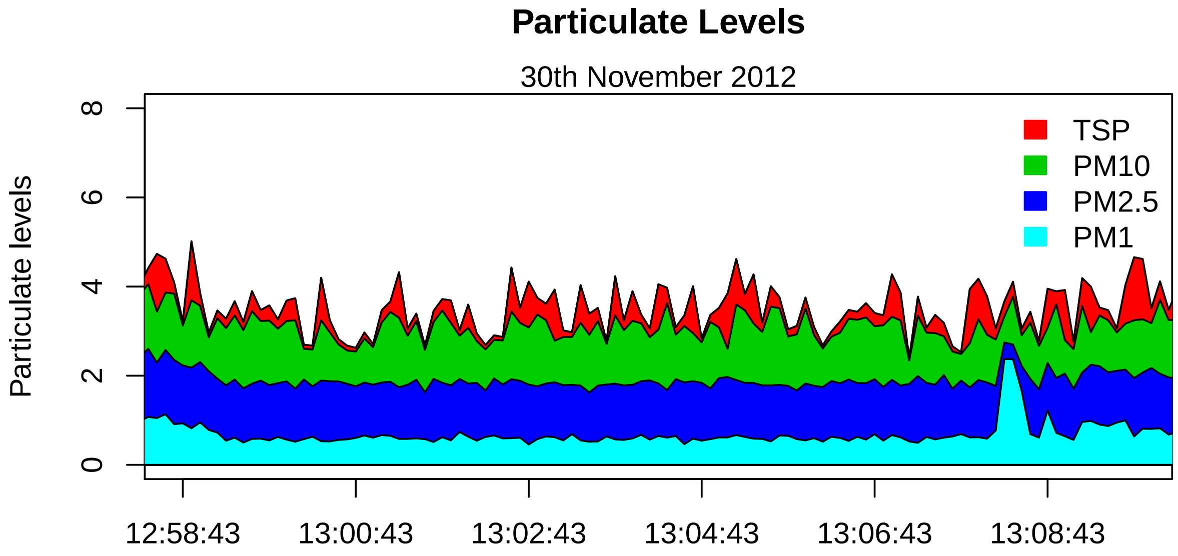


Figure 2: Particulate levels - 10 Minute Period

4.1 Rain

Exploratory analysis suggested relatively little relationship between Cessnock rainfall and measured particulate levels, while there was a fairly strong relationship between particulate levels and Maitland rainfall. This was despite the fact that more detailed information was available on Cessnock rainfall (every 30 minutes) whereas Maitland rainfall was recorded only on a daily level. Figure 3 shows boxplots of logged TSP levels according to whether or not it rained in Maitland on the same day.

Visually the plots suggest that TSP levels are generally slightly lower on days when it rained and also prone to extremes on days when it rained, compared to days when it didn't rain. The mean TSP levels on days where it did not rain at Maitland were 41.02 compared to 35.85 on days where it did. This difference was not statistically significant (p =0.84, based on a block bootstrap). However it turns out that there was a much stronger correlation with whether or not it had rained on the previous day. Figure 3 also shows boxplots of logged TSP levels according to whether or not it had rained in Maitland on the previous day.

This figure shows a much clearer decline in TSP levels when it had rained the previous day. The mean TSP levels on days where it had not rained the previous day at Maitland were 43.05 compared to 32.1 on days where it had. This difference was statistically significant at $p < 0.001$. This will be seen with our subsequently reported regression models. In contrast to the Maitland results, there was relatively little difference in TSP levels according to Cessnock rainfall.

The mean TSP levels on days where it did not rain at Cessnock were 38.78 compared to 39.55 on days where it did. While this difference was not statistically significant (p =0.96, based on a block bootstrap), the effect went away after we simultaneously adjusted for Maitland rainfall. The correlation was also weak when we looked at previous day's rain at Cessnock - see the boxplots in Figure 4. The mean TSP levels on days where

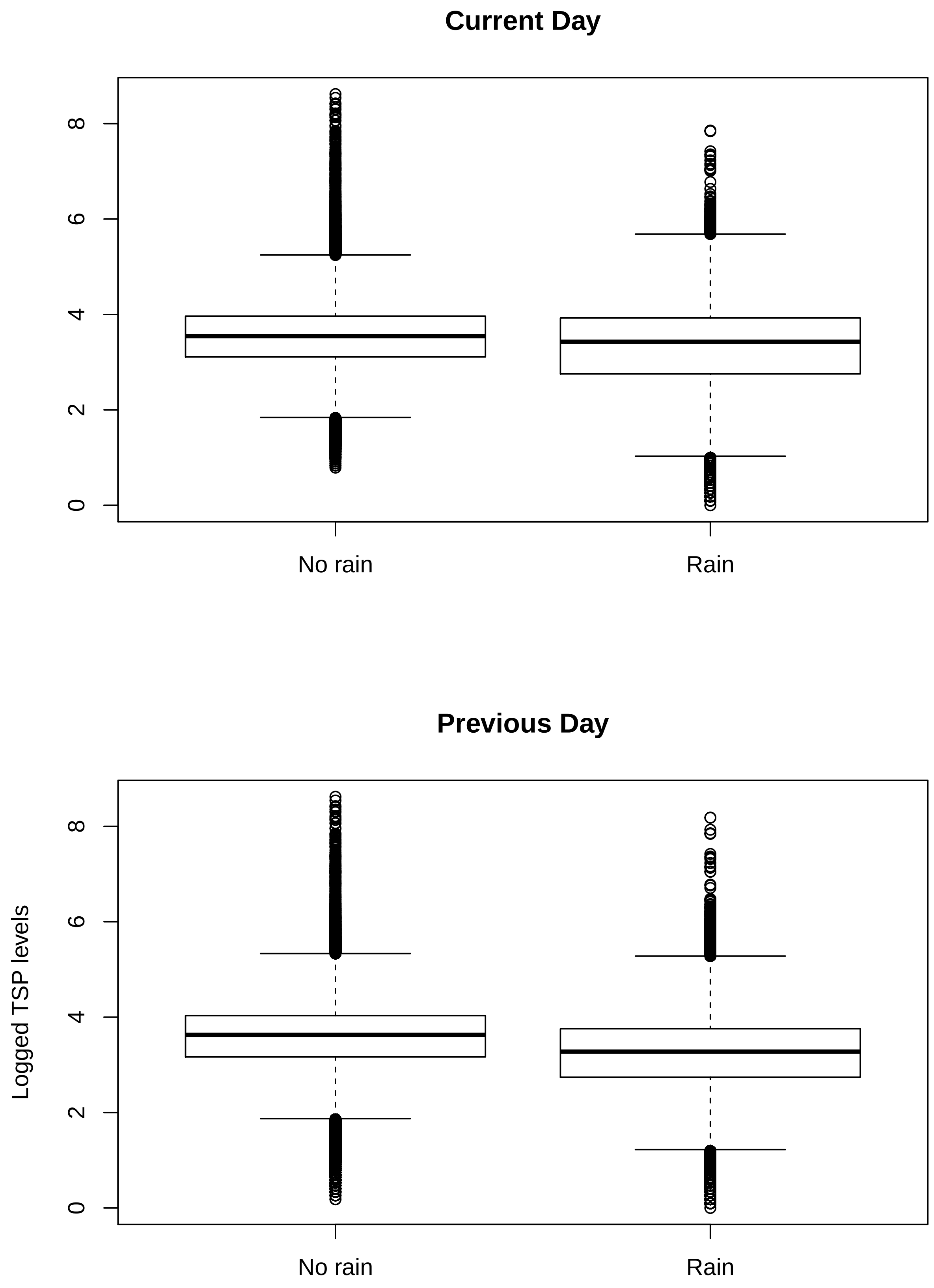
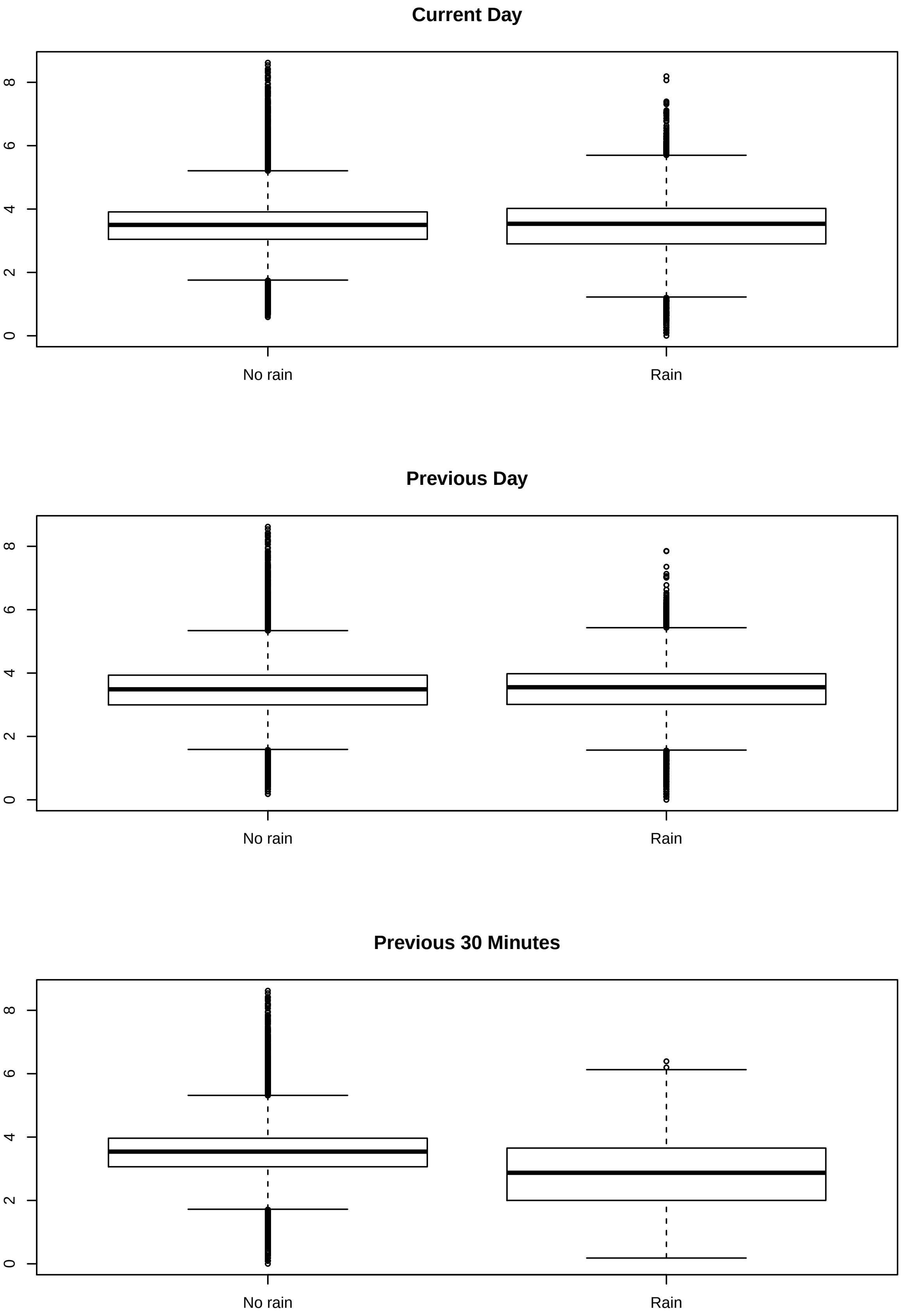


Figure 3: Logged TSP levels in Maitland by rain status



Current Day

Previous Day

Previous 30 Minutes

Figure 4: Logged TSP levels in Cessnock by rain status

it had not rained the previous day at Cessnock were 39.06 compared to 39.08 on days where it had (p =0.84, based on a block bootstrap).

As indicated above, however, while rain status at Maitland was recorded only at a daily level, Cessnock rain was recorded at 30 minute increments. An analysis of these data revealed a slightly different picture. Figure 4 also shows logged TSP levels according to whether or not it had rained in Cessnock within a 30 minute interval. This Figure shows a clearer pattern of lowered TSP levels during the 30 minute periods when it is raining in Cessnock. However, the difference is not statistically significant (p =0.84, based on a block bootstrap).)

We used the modelling strategies described above to build regression models including these rain variables. Appendix 1 shows the results of fitting a regression model that included all the terms described in the previous section, as well as the following rain variables:

- Indicators of whether or not it rained that same day in Cessnock or Maitland (respectively CRainIndDay and MRainInd), as well as the amount of rain that day in both locations (CessnockRain and MaitlandRain);
- Indicators of whether or not it rained the previous day in Cessnock or Maitland (respectively CRainIndPreDay and MRainIndPreDay);
- An indicator of whether or not it rained that same 30 minutes in Cessnock (CRainInd), as well as the amount of rain in those 30 minutes (CrainNow2);
- The amount of rain in the previous 30 minutes in Cessnock (CRainPrevious30Min).

Fifty block bootstraps were used for estimating standard errors. We used standard variable selection techniques to remove those variables that did not significantly improve model fit. After doing this, whether or not it had rained at Maitland on the previous day was the only significant predictor.

4.2 Locomotives

We turn now to the analysis of data on number of locomotives. The following table shows the mean TSP and PM10 levels associated with the passing of various types of train, broken out according to the number of locomotives on the train. Data are not included on passenger trains since very few of these trains had information available regarding the number of locomotives. Data are not included on "other" train types, since there were so few of these. The "0" column refers to the absence of that train type.

Table 2: Mean Particulate levels by Train Types and Number of Locomotives

	Number Of Locos							
Train Type	0	1	2	3	4	5	6	7
<u>TSP</u>								
Empty Coal	38.91	39.99	42.28	43.31	42.63	18.26		
Loaded Coal	38.89	42.12	42.80	42.31	42.10	52.42	28.53	72.56
Freight	39.03	53.27	38.35	49.60	47.00	18.47		
<u>PM10</u>								
Empty Coal	28.97	29.41	31.50	31.94	30.92	11.89		
Loaded Coal	28.95	31.64	31.50	31.51	30.38	41.65	21.58	56.81
Freight	29.06	37.66	28.56	36.30	34.17	15.83		

Note that because there are very few trains with more than three locomotives (see Table 1), it is not possible to interpret anything but the first few columns in the table. But the table does suggest that there is no particular increase in either total particulate levels or PM10 levels with increasing numbers of locomotives. This will be borne out in our subsequent regression analyses.

We also examined some graphical data displays. Figure 5 shows boxplots of logged TSP levels, according to the number of locos on each passing loaded coal train. As in the above table, the value "0" indicates no loaded coal train passing.

The Figure confirms that pattern seen in the table, namely that there is little impact between logged TSP levels and the number of locos on a passing loaded coal train. The slight increase in logged TSP levels associated with the passing of a loaded coal train can be seen from the difference between the case where there are zero locos, compared to all the rest.

Figure 5 also shows similar patterns with regard to logged TSP levels, according to the number of locos on each passing empty coal train and freight train, respectively.

We also looked at plots of logged PM2.5, see Figure 6, according to the number of locos associated with various different kinds of train passings.

We explored the impact of number of locomotives using regression modelling. For each train type, we created a new variable that took the value 0 at times when that train type was not present, and otherwise took the value corresponding to the number of locomotives on the passing train. We added these new variables into our base model that included indicators of whether or not each of the various train types were passing at that moment. Because so few passenger trains had information available regarding the number of locomotives, we ignored the locomotive data for this traintype. Consistent with what was seen in Tables 1 and 2 above, we found that adding information about numbers of locomotives into our regression models yielded a non-significant improvement in model fit. In particular, the regression models (see Appendix 1) shows that the basic train indicators mostly remain significant, but the loco variables had very small coefficients with non-significant associated p-values. Hence we omitted these variables from our final analyses, reported in the next section.

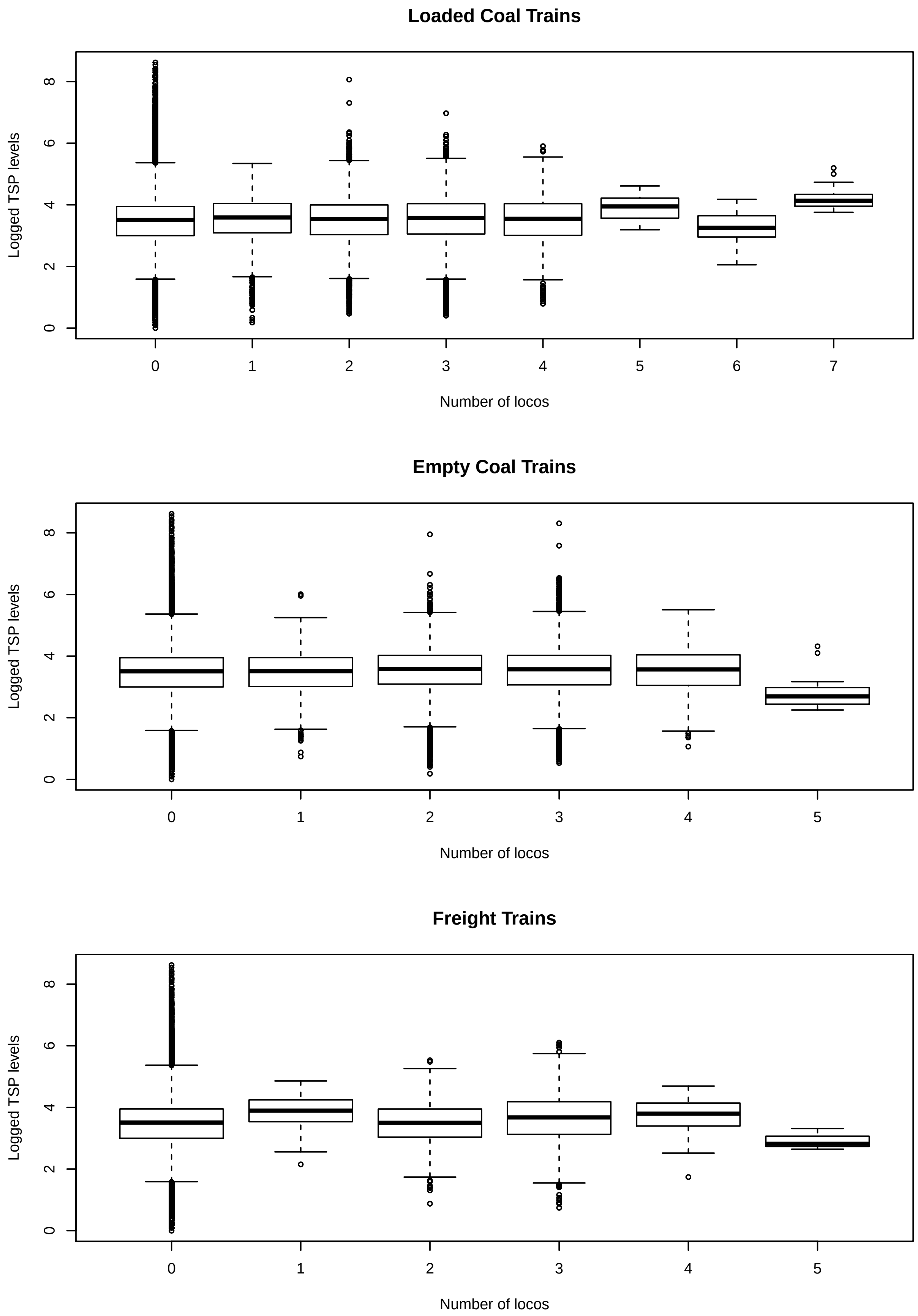


Figure 5: Logged TSP according to number of locos by passing train type

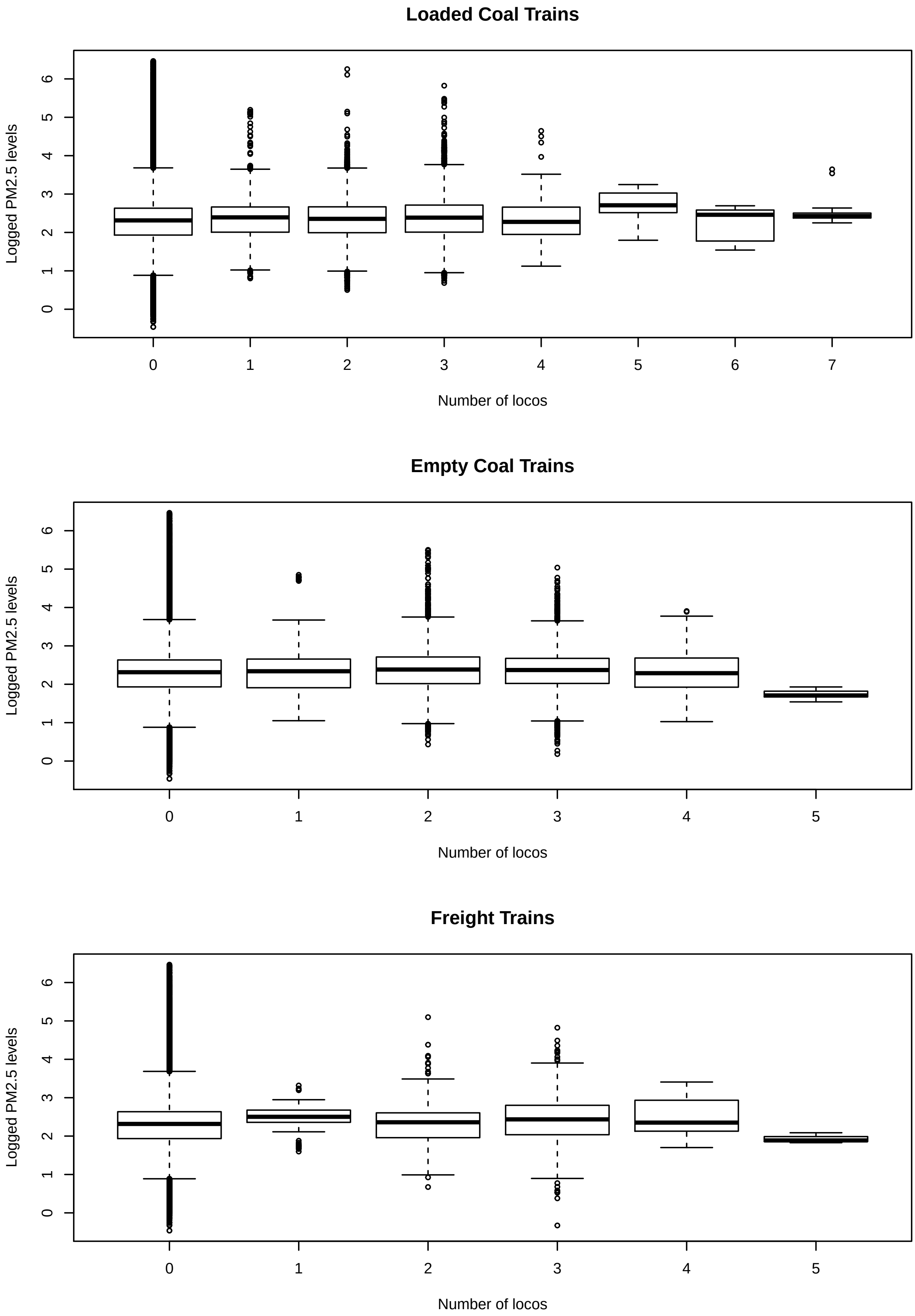


Figure 6: Logged PM2.5 according to number of locos by passing train type

4.3 Final Regression Models

The following Tables show the results of our main regression analysis on logged values of TSP, PM10, PM2.5 and PM1. All four models show similar patterns, namely that particulate levels are higher when any trains are passing, as well as during the five minute period after the trains have passed. The magnitude of increase is similar for freight, loaded coal and unloaded coal trains, and roughly half that magnitude when passenger trains are passing. All models include smooth spline terms in day and seconds since midnight to allow for diernal effects as well as day to day variation. Fifty bootstraps were used to generate standard deviations adjusted for serial correlation.

Table 3: Regression model for log TSP

Variable	Estimate	Std Error	t-value	p-value
Intercept	3.508	0.080	43.80	<0.001
Freight train passing	0.100	0.037	2.729	0.006
Freight passed within 5 min	0.100	0.030	3.316	0.001
Empty coal train passing	0.090	0.017	5.339	<0.001
Empty coal train passed within 5 min	0.113	0.012	9.675	<0.001
Loaded coal train passing	0.073	0.017	4.196	<0.001
Loaded coal passed within 5 min	0.081	0.015	5.387	<0.001
Passenger train passing	0.049	0.018	2.694	0.007
Passenger train passed within 5 min	0.044	0.010	4.542	<0.001
Unknown train type passing	0.124	0.041	3.044	0.002
Unknown train type passed within 5 min	0.074	0.044	1.710	0.087
Rain in Maitland on previous day (Y/N)	-0.303	0.119	-2.548	0.011

Table 4: Regression model for log PM10

Variable	Estimate	Std Error	t-value	p-value
Intercept	3.269	0.083	39.418	<0.001
Freight train passing	0.095	0.036	2.628	0.009
Freight passed within 5 min	0.090	0.029	3.125	0.002
Empty coal train passing	0.085	0.016	5.370	<0.001
Empty coal train passed within 5 min	0.109	0.012	9.271	<0.001
Loaded coal train passing	0.072	0.017	4.186	<0.001
Loaded coal passed within 5 min	0.077	0.015	5.299	<0.001
Passenger train passing	0.037	0.013	2.812	0.005
Passenger train passed within 5 min	0.040	0.010	4.019	<0.001
Unknown train type passing	0.134	0.039	3.410	0.001
Unknown train type passed within 5 min	0.077	0.043	1.797	0.072
Rain in Maitland on previous day (Y/N)	-0.303	0.127	-2.388	0.017

Table 5: Regression model for log PM2.5

Variable	Estimate	Std Error	t-value	p-value
Intercept	2.384	0.075	31.844	<0.001
Freight train passing	0.066	0.031	2.091	0.037
Freight passed within 5 min	0.059	0.023	2.498	0.001
Empty coal train passing	0.073	0.013	5.475	<0.001
Empty coal train passed within 5 min	0.092	0.011	8.321	<0.001
Loaded coal train passing	0.062	0.016	3.974	<0.001
Loaded coal passed within 5 min	0.064	0.013	4.837	<0.001
Passenger train passing	0.029	0.012	2.398	0.017
Passenger train passed within 5 min	0.033	0.010	3.171	0.002
Unknown train type passing	0.146	0.049	2.952	0.003
Unknown train type passed within 5 min	0.081	0.044	1.846	0.065
Rain in Maitland on previous day (Y/N)	-0.307	0.129	-2.376	0.017

Table 6: Regression model for log PM1

Variable	Estimate	Std Error	t-value	p-value
Intercept	1.886	0.018	106.087	<0.001
Freight train passing	0.024	0.010	2.261	0.024
Freight train passed within 5 min	0.021	0.012	1.708	0.088
Empty coal train passing	0.025	0.006	4.507	<0.001
Empty coal train passed within 5 min	0.035	0.005	7.102	<0.001
Loaded coal train passing	0.021	0.007	2.886	0.004
Loaded coal passed within 5 min	0.018	0.004	4.141	<0.001
Passenger train passing	0.005	0.006	0.804	0.422
Passenger train passed within 5 min	0.007	0.005	1.308	0.191
Unknown train type passing	0.058	0.043	1.353	0.176
Unknown train type passed within 5 min	0.034	0.023	1.463	0.144
Rain in Maitland on previous day (Y/N)	-0.084	0.030	-2.798	0.005

All four models suggest that whether or not it rained at Maitland on the previous day has a strongly significant impact on particulate levels. As shown in Appendix 1 and reported in the previous section, we found that after accounting for previous day's rainfall, current day rain in Maitland was not strongly associated with particulate levels, nor was rainfall at Cessnock. This result makes sense since Maitland is quite close to the monitoring site. We also found that there was no interaction between train type and whether or not it had rained yesterday at Maitland. These results all point towards the suggestion that stirring up of dust particles on the tracks and nearby ground is a major driver of the observed increases in particulate levels associated with various train passings.

As discussed in a previous section, we used a variant of linear regression (*gam*) that allowed us to include smooth functions of time in the model as well. This was important in order to adjust for temporal effects that could be influencing results. For example, the following Figure shows that trains are much more likely

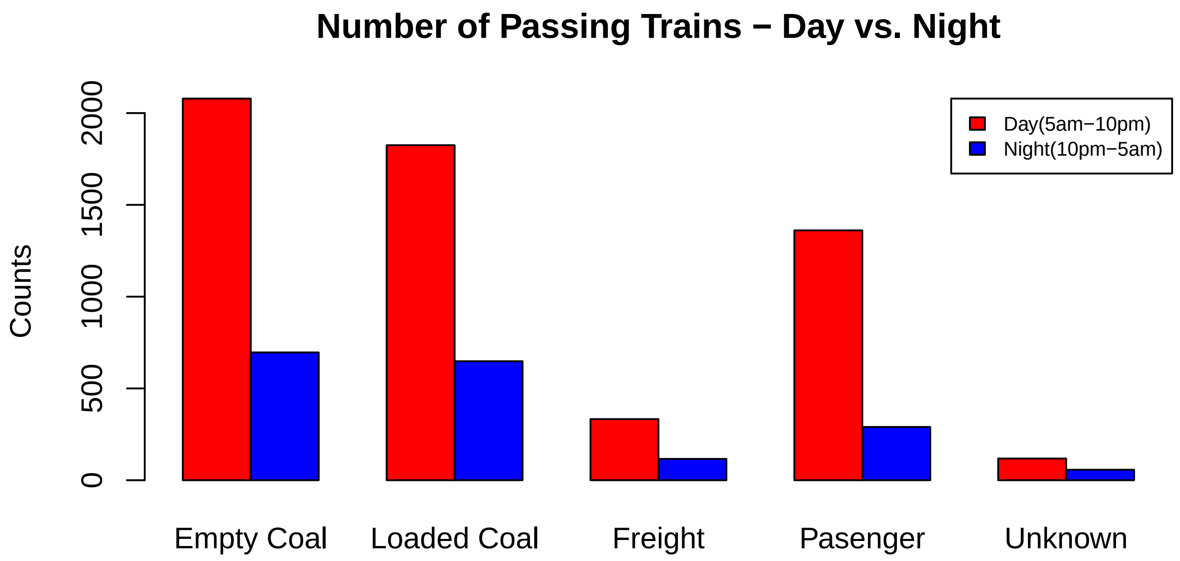


Figure 7: Numbers of various trains passing by day and night

to pass the monitoring station during day time hours, than nighttime hours.

Inclusion of smooth terms in time provides a means of adjusting for this difference. The following plots show the estimated smooth functions of time that have been included in the model.

As indicated in the 2014 report, wind speed and direction data were missing for 151,090 timepoints. It turns out that having the wind blowing towards the monitor is an important predictor of particulate levels. But we can only assess this effect in the reduced dataset where wind data are available. Appendix 1 also reports on regression models that include the variable that indicates whether or not wind was blowing towards the monitor.

5 Conclusion

We have reported on an extended analysis of the Hunter Valley particulate study, taking into account additional data on the number of locomotives on each train as well as data on precipitation. We found that the number of locomotives had little impact on particulate levels. An important caveat is that the ARTC has warned that they do not believe that the locomotive data are entirely accurate. However, this finding does dispel, to some extent, the hypothesis that diesel exhaust explains a large proportion of the observed increases in particulate levels associated with train passings. We found that whether or not it had rained in Maitland the previous day had a significant impact on particulate levels the following day. There was no interaction effect observed: in other words, the impact of previous day's rain was the same, regardless of which type of train was passing. This finding suggests that a key mechanism for the increased particulate levels was stirring up by passing trains of existing dust particles that had settled previously on the tracks and nearby ground.

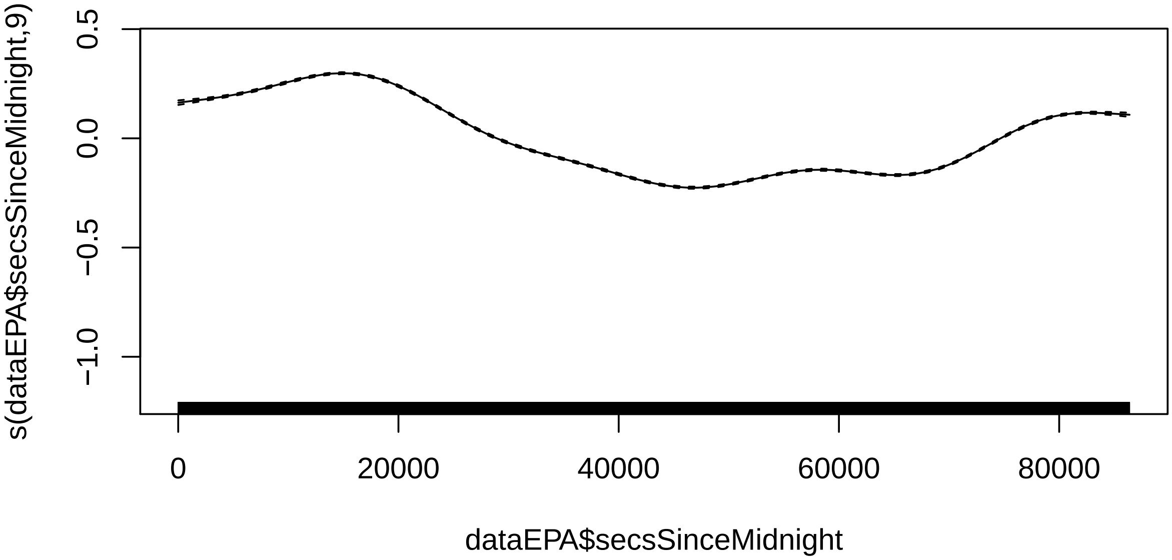


Figure 8: Smooth function of time included in gam model

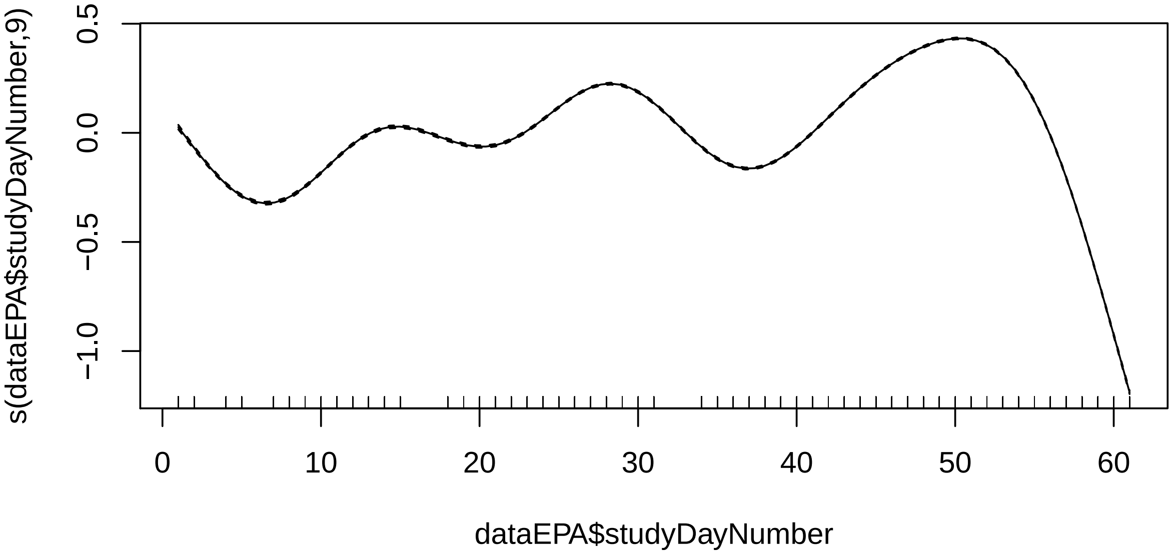


Figure 9: Smooth function of time included in gam model

6 References

[1] Ryan, L and Wand, M. Re-analysis of ARTC data on Particulate Emissions from Coal Trains. Technical Report, University of Technology Sydney

[2] Wood, S (2006). Generalized Additive Models: An Introduction with R. Chapman & Hall/CRC Texts in Statistical Science.

[3] R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

[4] Efron, B. and Tibshirani, R. J. (1994). An introduction to the bootstrap (Vol. 57). CRC press.

7 Appendix 1: Additional Details

This Appendix provides detailed regression output for the models reported in the main body of the report. The appendix also includes outputs of regressions corresponding to various sub-analyses.

Exploring the impact of Rain

```
## [1] "TSP Analysis"
## [1] "number of bootstraps:  50"
##                              Estimate Std. Error t value  Pr(>|t|)
##  [1,] "(Intercept)"         "3.519"  "0.075"    "46.721" "0"
##  [2,] "Freight"             "0.1"    "0.035"    "2.865"  "0.004"
##  [3,] "FreightPassing2"     "0.098"  "0.028"    "3.494"  "0"
##  [4,] "EmptyCoal"           "0.097"  "0.017"    "5.737"  "0"
##  [5,] "EmptyCoalPassing2"   "0.118"  "0.011"    "10.333" "0"
##  [6,] "LoadedCoal"          "0.077"  "0.019"    "4.09"   "0"
##  [7,] "LoadedCoalPassing2"  "0.081"  "0.016"    "5.03"   "0"
##  [8,] "Passenger"           "0.044"  "0.018"    "2.429"  "0.015"
##  [9,] "PassengerPassing2"   "0.038"  "0.01"     "3.953"  "0"
## [10,] "Unknown"             "0.156"  "0.044"    "3.532"  "0"
## [11,] "UnknownPassing2"     "0.097"  "0.039"    "2.473"  "0.013"
## [12,] "CRainInd"            "-0.002" "0.106"    "-0.019" "0.985"
## [13,] "MRainInd"            "-0.385" "0.201"    "-1.919" "0.055"
## [14,] "CessnockRain"        "-0.008" "0.01"     "-0.794" "0.427"
## [15,] "MaitlandRain"        "0.011"  "0.014"    "0.774"  "0.439"
## [16,] "MRainIndPreDay"      "-0.341" "0.112"    "-3.034" "0.002"
## [17,] "CRainIndPreDay"      "0.258"  "0.216"    "1.19"   "0.234"
## [18,] "CRainNow2"           "-0.022" "0.056"    "-0.393" "0.694"
## [19,] "CRainPrevious30Mins" "-0.041" "0.053"    "-0.786" "0.432"
## [20,] "windToward"          "0.125"  "0.039"    "3.227"  "0.001"
```

Number of locomotives does not improve model fit

We fitted a model that included not only the indicators of whether each particular kind of train was passing, but also the number of locomotives on each train. The model structure was the same as reported in the main body of the paper in terms of including smooth terms in day and seconds since midnight. Twenty bootstraps were used for estimating standard errors.

```
## [1] "TSP Analysis"
## [1] "number of bootstraps:  20"
##                            Estimate Std. Error t value  Pr(>|t|)
##  [1,] "(Intercept)"       "3.444"  "0.085"    "40.313" "0"
##  [2,] "Freight"           "0.065"  "0.105"    "0.621"  "0.535"
##  [3,] "FreightLocos"      "0.016"  "0.043"    "0.364"  "0.716"
##  [4,] "FreightPassing2"    "0.099"  "0.028"    "3.468"  "0.001"
##  [5,] "EmptyCoal"          "0.058"  "0.043"    "1.347"  "0.178"
##  [6,] "EmptyCoalLocos"     "0.013"  "0.017"    "0.777"  "0.437"
##  [7,] "EmptyCoalPassing2"  "0.112"  "0.013"    "8.501"  "0"
##  [8,] "LoadedCoal"         "0.062"  "0.032"    "1.957"  "0.05"
##  [9,] "LoadedCoalLocos"    "0.007"  "0.01"     "0.679"  "0.497"
## [10,] "LoadedCoalPassing2" "0.08"   "0.015"    "5.466"  "0"
## [11,] "Passenger"          "0.052"  "0.017"    "3.157"  "0.002"
## [12,] "PassengerPassing2"  "0.044"  "0.007"    "6.757"  "0"
## [13,] "Unknown"            "0.343"  "0.148"    "2.313"  "0.021"
## [14,] "UnknownLocos"       "-0.088" "0.055"    "-1.592" "0.111"
## [15,] "UnknownPassing2"    "0.08"   "0.046"    "1.759"  "0.079"
## [16,] "MRainIndPreDay"     "-0.287" "0.115"    "-2.491" "0.013"
```

Final regression models

```
## [1] "TSP Analysis - wind data not included"
## [1] "number of bootstraps:  50"
##                                Estimate Std. Error t value  Pr(>|t|)
##  [1,] "(Intercept)"            "3.508"  "0.08"     "43.795" "0"
##  [2,] "Freight"                "0.1"    "0.037"    "2.729"  "0.006"
##  [3,] "FreightPassing2"        "0.1"    "0.03"     "3.316"  "0.001"
##  [4,] "EmptyCoal"              "0.09"   "0.017"    "5.339"  "0"
##  [5,] "EmptyCoalPassing2"      "0.113"  "0.012"    "9.675"  "0"
##  [6,] "LoadedCoal"             "0.073"  "0.017"    "4.196"  "0"
##  [7,] "LoadedCoalPassing2"     "0.081"  "0.015"    "5.387"  "0"
##  [8,] "Passenger"              "0.049"  "0.018"    "2.694"  "0.007"
##  [9,] "PassengerPassing2"      "0.044"  "0.01"     "4.542"  "0"
## [10,] "Unknown"                "0.124"  "0.041"    "3.044"  "0.002"
## [11,] "UnknownPassing2"        "0.074"  "0.044"    "1.71"   "0.087"
## [12,] "MRainIndPreDay"         "-0.303" "0.119"    "-2.548" "0.011"
## [13,] "s(secsSinceMidnight).1" "-0.171" "0.146"    "-1.168" "0.243"
```

@

```
## [1] "PM2.5 Analysis - wind data not included"
## [1] "number of bootstraps:  50"
##                                Estimate Std. Error t value  Pr(>|t|)
##  [1,] "(Intercept)"            "2.384"  "0.075"    "31.844" "0"
##  [2,] "Freight"                "0.066"  "0.031"    "2.091"  "0.037"
##  [3,] "FreightPassing2"        "0.059"  "0.023"    "2.498"  "0.012"
##  [4,] "EmptyCoal"              "0.073"  "0.013"    "5.475"  "0"
##  [5,] "EmptyCoalPassing2"      "0.092"  "0.011"    "8.321"  "0"
##  [6,] "LoadedCoal"             "0.062"  "0.016"    "3.974"  "0"
##  [7,] "LoadedCoalPassing2"     "0.064"  "0.013"    "4.837"  "0"
##  [8,] "Passenger"              "0.029"  "0.012"    "2.398"  "0.017"
##  [9,] "PassengerPassing2"      "0.033"  "0.01"     "3.171"  "0.002"
## [10,] "Unknown"                "0.146"  "0.049"    "2.952"  "0.003"
## [11,] "UnknownPassing2"        "0.081"  "0.044"    "1.846"  "0.065"
## [12,] "MRainIndPreDay"         "-0.307" "0.129"    "-2.376" "0.017"
## [13,] "s(secsSinceMidnight).1" "-0.021" "0.101"    "-0.209" "0.835"
```

```
## [1] "TSP Analysis - wind data included"
## [1] "number of bootstraps:  20"
##                            Estimate Std. Error t value  Pr(>|t|)
##  [1,] "(Intercept)"        "3.486"  "0.129"    "27.064" "0"
##  [2,] "Freight"            "0.084"  "0.035"    "2.396"  "0.017"
##  [3,] "FreightPassing2"    "0.028"  "0.023"    "1.213"  "0.225"
##  [4,] "EmptyCoal"          "0.101"  "0.021"    "4.804"  "0"
##  [5,] "EmptyCoalPassing2"  "0.099"  "0.013"    "7.742"  "0"
##  [6,] "LoadedCoal"         "0.088"  "0.02"     "4.443"  "0"
##  [7,] "LoadedCoalPassing2" "0.065"  "0.013"    "4.999"  "0"
##  [8,] "Passenger"          "0.06"   "0.022"    "2.682"  "0.007"
##  [9,] "PassengerPassing2"  "0.043"  "0.012"    "3.66"   "0"
## [10,] "Unknown"            "0.109"  "0.036"    "2.993"  "0.003"
## [11,] "UnknownPassing2"    "0.059"  "0.049"    "1.219"  "0.223"
## [12,] "MRainIndPreDay"     "-0.267" "0.135"    "-1.976" "0.048"
## [13,] "windSpeed"          "0.01"   "0.017"    "0.62"   "0.535"
```

```
## [1] "TSP Analysis - wind data included"
## [1] "number of bootstraps:  20"
##                            Estimate Std. Error t value  Pr(>|t|)
##  [1,] "(Intercept)"        "3.435"  "0.114"    "30.153" "0"
##  [2,] "Freight"            "0.093"  "0.034"    "2.739"  "0.006"
##  [3,] "FreightPassing2"    "0.031"  "0.022"    "1.404"  "0.16"
##  [4,] "EmptyCoal"          "0.105"  "0.021"    "4.992"  "0"
##  [5,] "EmptyCoalPassing2"  "0.102"  "0.013"    "8.032"  "0"
##  [6,] "LoadedCoal"         "0.094"  "0.021"    "4.518"  "0"
##  [7,] "LoadedCoalPassing2" "0.067"  "0.013"    "4.978"  "0"
##  [8,] "Passenger"          "0.061"  "0.022"    "2.808"  "0.005"
##  [9,] "PassengerPassing2"  "0.042"  "0.011"    "3.697"  "0"
## [10,] "Unknown"            "0.115"  "0.036"    "3.209"  "0.001"
## [11,] "UnknownPassing2"    "0.063"  "0.049"    "1.283"  "0.2"
## [12,] "MRainIndPreDay"     "-0.252" "0.136"    "-1.85"  "0.064"
## [13,] "windToward"         "0.12"   "0.033"    "3.697"  "0"
## [1] "PM10 Analysis - wind data included"
## [1] "number of bootstraps:  20"
##                              Estimate Std. Error t value  Pr(>|t|)
##  [1,] "(Intercept)"         "3.215"  "0.121"    "26.488" "0"
##  [2,] "Freight"             "0.083"  "0.034"    "2.46"   "0.014"
##  [3,] "FreightPassing2"     "0.025"  "0.022"    "1.123"  "0.262"
##  [4,] "EmptyCoal"           "0.098"  "0.02"     "4.972"  "0"
##  [5,] "EmptyCoalPassing2"   "0.098"  "0.012"    "7.974"  "0"
##  [6,] "LoadedCoal"          "0.091"  "0.021"    "4.289"  "0"
##  [7,] "LoadedCoalPassing2" "0.064"  "0.014"    "4.734"  "0"
##  [8,] "Passenger"           "0.049"  "0.017"    "2.922"  "0.003"
##  [9,] "PassengerPassing2"   "0.038"  "0.012"    "3.249"  "0.001"
## [10,] "Unknown"             "0.132"  "0.033"    "3.99"   "0"
## [11,] "UnknownPassing2"     "0.063"  "0.049"    "1.29"   "0.197"
## [12,] "MRainIndPreDay"      "-0.256" "0.156"    "-1.64"  "0.101"
## [13,] "windToward"          "0.087"  "0.033"    "2.67"   "0.008"
```

```
## [1] "PM25 Analysis - wind data included"
## [1] "number of bootstraps:  20"
##                              Estimate Std. Error t value  Pr(>|t|)
##  [1,] "(Intercept)"         "2.374"  "0.115"    "20.562" "0"
##  [2,] "Freight"             "0.053"  "0.031"    "1.692"  "0.091"
##  [3,] "FreightPassing2"     "0.003"  "0.021"    "0.151"  "0.88"
##  [4,] "EmptyCoal"           "0.081"  "0.016"    "5.252"  "0"
##  [5,] "EmptyCoalPassing2"   "0.081"  "0.011"    "7.679"  "0"
##  [6,] "LoadedCoal"          "0.073"  "0.019"    "3.808"  "0"
##  [7,] "LoadedCoalPassing2" "0.055"  "0.013"    "4.165"  "0"
##  [8,] "Passenger"           "0.038"  "0.015"    "2.489"  "0.013"
##  [9,] "PassengerPassing2"   "0.035"  "0.013"    "2.733"  "0.006"
## [10,] "Unknown"             "0.156"  "0.044"    "3.583"  "0"
## [11,] "UnknownPassing2"     "0.067"  "0.044"    "1.529"  "0.126"
## [12,] "MRainIndPreDay"      "-0.255" "0.18"     "-1.417" "0.156"
## [13,] "windToward"          "-0.003" "0.032"    "-0.093" "0.926"
```

```
## [1] "PM1 Analysis - wind data included"
## [1] "number of bootstraps:  20"
##                              Estimate Std. Error t value  Pr(>|t|)
##  [1,] "(Intercept)"         "1.876"  "0.028"    "67.537" "0"
##  [2,] "Freight"             "0.027"  "0.012"    "2.255"  "0.024"
##  [3,] "FreightPassing2"     "-0.005" "0.008"    "-0.656" "0.512"
##  [4,] "EmptyCoal"           "0.03"   "0.007"    "4.19"   "0"
##  [5,] "EmptyCoalPassing2"   "0.031"  "0.005"    "5.781"  "0"
##  [6,] "LoadedCoal"          "0.027"  "0.009"    "3.034"  "0.002"
##  [7,] "LoadedCoalPassing2" "0.018"  "0.005"    "3.341"  "0.001"
##  [8,] "Passenger"          "0.009"  "0.007"    "1.302"  "0.193"
##  [9,] "PassengerPassing2"  "0.007"  "0.004"    "1.718"  "0.086"
## [10,] "Unknown"            "0.078"  "0.043"    "1.792"  "0.073"
## [11,] "UnknownPassing2"    "0.04"   "0.024"    "1.662"  "0.097"
## [12,] "MRainIndPreDay"     "-0.078" "0.044"    "-1.763" "0.078"
## [13,] "windToward"         "0.016"  "0.008"    "2.026"  "0.043"
```

8 Appendix 2: Additional Details for 2014 report

The 2014 report by Professor Ryan described a number of sensitivity analyses, but did not provide details. This Appendix provides those details.

Exclusion of times with multiple train passings

There were 12 timepoints where 3 trains were passing simultaneously and 1981 timepoints where two trains were passing. We excluded these total of 1993 timepoints where multiple trains were passing and refit our main model, yeilding the following results:

```
## [1] "TSP Analysis - multiple train passings excluded"
## [1] "number of bootstraps:  20"
##                             Estimate Std. Error t value  Pr(>|t|)
##  [1,] "(Intercept)"         "3.444"  "0.085"    "40.348" "0"
##  [2,] "Freight"             "0.118"  "0.03"     "3.887"  "0"
##  [3,] "FreightPassing2"     "0.099"  "0.028"    "3.474"  "0.001"
##  [4,] "EmptyCoal"           "0.089"  "0.019"    "4.617"  "0"
##  [5,] "EmptyCoalPassing2"   "0.112"  "0.013"    "8.517"  "0"
##  [6,] "LoadedCoal"          "0.077"  "0.017"    "4.566"  "0"
##  [7,] "LoadedCoalPassing2"  "0.08"   "0.015"    "5.449"  "0"
##  [8,] "Passenger"           "0.052"  "0.017"    "2.97"   "0.003"
##  [9,] "PassengerPassing2"   "0.044"  "0.007"    "6.648"  "0"
## [10,] "Unknown"             "0.106"  "0.036"    "2.937"  "0.003"
## [11,] "UnknownPassing2"     "0.081"  "0.045"    "1.783"  "0.075"
## [12,] "MRainIndPreDay"      "-0.287" "0.115"    "-2.491" "0.013"
```

These results suggest that the overall conclusions are unchanged when we exclude these multiple train events.

Adjusting for train speed

We reran the model including the various different trainspeed variables as predictors. The results were as follows:

```
## [1] "TSP Analysis"
## [1] "number of bootstraps:  20"
##                             Estimate Std. Error t value  Pr(>|t|)
##  [1,] "(Intercept)"         "3.49"   "0.085"    "41.14"  "0"
##  [2,] "Freight"             "0.134"  "0.125"    "1.071"  "0.284"
##  [3,] "FreightSpeed"        "-0.002" "0.007"    "-0.28"  "0.779"
##  [4,] "FreightPassing2"     "0.097"  "0.029"    "3.363"  "0.001"
##  [5,] "EmptyCoal"           "0.124"  "0.113"    "1.092"  "0.275"
##  [6,] "EmptyCoalSpeed"      "-0.002" "0.006"    "-0.323" "0.746"
##  [7,] "EmptyCoalPassing2"   "0.111"  "0.013"    "8.57"   "0"
##  [8,] "LoadedCoal"          "0.249"  "0.087"    "2.881"  "0.004"
##  [9,] "LoadedCoalSpeed"    "-0.014" "0.006"    "-2.189" "0.029"
## [10,] "LoadedCoalPassing2" "0.079"  "0.014"    "5.483"  "0"
## [11,] "Passenger"          "0.15"   "0.156"    "0.962"  "0.336"
## [12,] "PassengerSpeed"     "-0.004" "0.006"    "-0.66"  "0.509"
## [13,] "PassengerPassing2"  "0.046"  "0.007"    "6.291"  "0"
## [14,] "Unknown"            "0.034"  "0.153"    "0.221"  "0.825"
## [15,] "UnknownSpeed"       "0.006"  "0.009"    "0.64"   "0.522"
## [16,] "UnknownPassing2"    "0.076"  "0.046"    "1.663"  "0.096"
## [17,] "MRainIndPreDay"     "-0.292" "0.113"    "-2.574" "0.01"
```

Analysis when wind is towards monitor

```
## [1] "TSP Analysis - wind data included"
## [1] "number of bootstraps:  20"
##                                Estimate Std. Error t value  Pr(>|t|)
##  [1,] "(Intercept)"            "3.33"   "0.077"    "43.27"  "0"
##  [2,] "Freight"                "0.094"  "0.037"    "2.573"  "0.01"
##  [3,] "FreightPassing2"        "0.036"  "0.026"    "1.393"  "0.164"
##  [4,] "EmptyCoal"              "0.108"  "0.021"    "5.164"  "0"
##  [5,] "EmptyCoalPassing2"      "0.103"  "0.013"    "8.097"  "0"
##  [6,] "LoadedCoal"             "0.107"  "0.02"     "5.236"  "0"
##  [7,] "LoadedCoalPassing2"     "0.078"  "0.016"    "4.959"  "0"
##  [8,] "Passenger"              "0.062"  "0.021"    "2.89"   "0.004"
##  [9,] "PassengerPassing2"      "0.042"  "0.013"    "3.272"  "0.001"
## [10,] "Unknown"                "0.112"  "0.036"    "3.106"  "0.002"
## [11,] "UnknownPassing2"        "0.05"   "0.051"    "0.983"  "0.326"
## [12,] "windToward"             "0.138"  "0.039"    "3.508"  "0"
## [13,] "s(secsSinceMidnight).1" "-0.153" "0.112"    "-1.373" "0.17"
```

```
## [1] "PM10 Analysis - wind data included"
## [1] "number of bootstraps:  20"
##                                Estimate Std. Error t value  Pr(>|t|)
##  [1,] "(Intercept)"            "3.109"  "0.075"    "41.213" "0"
##  [2,] "Freight"                "0.085"  "0.037"    "2.308"  "0.021"
##  [3,] "FreightPassing2"        "0.029"  "0.026"    "1.14"   "0.254"
##  [4,] "EmptyCoal"              "0.1"    "0.019"    "5.18"   "0"
##  [5,] "EmptyCoalPassing2"      "0.099"  "0.012"    "8.039"  "0"
##  [6,] "LoadedCoal"             "0.104"  "0.02"     "5.113"  "0"
##  [7,] "LoadedCoalPassing2"     "0.076"  "0.016"    "4.84"   "0"
##  [8,] "Passenger"              "0.05"   "0.017"    "2.976"  "0.003"
##  [9,] "PassengerPassing2"      "0.038"  "0.013"    "2.892"  "0.004"
## [10,] "Unknown"                "0.129"  "0.034"    "3.757"  "0"
## [11,] "UnknownPassing2"        "0.05"   "0.049"    "1.016"  "0.31"
## [12,] "windToward"             "0.104"  "0.039"    "2.704"  "0.007"
## [13,] "s(secsSinceMidnight).1" "-0.14"  "0.117"    "-1.194" "0.233"
```

```
## [1] "PM25 Analysis - wind data included"
## [1] "number of bootstraps:  20"
##                                Estimate Std. Error t value Pr(>|t|)
##  [1,] "(Intercept)"            "2.268"  "0.058"    "38.95" "0"
##  [2,] "Freight"                "0.054"  "0.035"    "1.548" "0.122"
##  [3,] "FreightPassing2"        "0.008"  "0.025"    "0.31"  "0.756"
##  [4,] "EmptyCoal"              "0.084"  "0.015"    "5.475" "0"
##  [5,] "EmptyCoalPassing2"      "0.082"  "0.011"    "7.565" "0"
##  [6,] "LoadedCoal"             "0.086"  "0.018"    "4.813" "0"
##  [7,] "LoadedCoalPassing2"     "0.066"  "0.015"    "4.481" "0"
##  [8,] "Passenger"              "0.038"  "0.016"    "2.388" "0.017"
##  [9,] "PassengerPassing2"      "0.034"  "0.014"    "2.441" "0.015"
## [10,] "Unknown"                "0.154"  "0.046"    "3.323" "0.001"
## [11,] "UnknownPassing2"        "0.054"  "0.042"    "1.285" "0.199"
## [12,] "windToward"             "0.013"  "0.038"    "0.348" "0.728"
## [13,] "s(secsSinceMidnight).1" "0.028"  "0.081"    "0.348" "0.728"
```

```
## [1] "PM1 Analysis - wind data included"
## [1] "number of bootstraps:  20"
##                                Estimate Std. Error t value   Pr(>|t|)
##  [1,] "(Intercept)"            "1.843"  "0.013"    "141.011" "0"
##  [2,] "Freight"                "0.027"  "0.012"    "2.242"   "0.025"
##  [3,] "FreightPassing2"        "-0.004" "0.009"    "-0.459"  "0.646"
##  [4,] "EmptyCoal"              "0.031"  "0.007"    "4.315"   "0"
##  [5,] "EmptyCoalPassing2"      "0.032"  "0.005"    "5.759"   "0"
##  [6,] "LoadedCoal"             "0.031"  "0.008"    "3.742"   "0"
##  [7,] "LoadedCoalPassing2"     "0.022"  "0.006"    "3.879"   "0"
##  [8,] "Passenger"              "0.009"  "0.007"    "1.268"   "0.205"
##  [9,] "PassengerPassing2"      "0.007"  "0.005"    "1.547"   "0.122"
## [10,] "Unknown"                "0.077"  "0.045"    "1.722"   "0.085"
## [11,] "UnknownPassing2"        "0.036"  "0.025"    "1.463"   "0.144"
## [12,] "windToward"             "0.021"  "0.01"     "2.158"   "0.031"
## [13,] "s(secsSinceMidnight).1" "0.007"  "0.032"    "0.21"    "0.834"
```