Using air quality data to track progress toward PM_{10} standards: Case study - Christchurch 1999 – 2006

NIWA Client Report: CHC2007-135
November 2007

NIWA Project: PCCA075

Using air quality data to track progress toward PM_{10} standards: Case study - Christchurch 1999 – 2006

Tim Appelhans[1]
*Jeff Bluett[2]
Katie Dey[2]
Gavin Fisher[3]
Andy Sturman[1]
Emily Wilton[4]

1. University of Canterbury
2. NIWA
3. Endpoint Ltd
4. Environet
*Corresponding Author

Prepared for

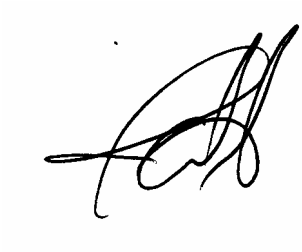
Foundation for Research, Science and Technology

NIWA Client Report: CHC2007-135
November 2007
NIWA Project: PCCA075

National Institute of Water & Atmospheric Research Ltd
10 Kyle Street, Riccarton, Christchurch
P O Box 8602, Christchurch, New Zealand
Phone +64-3-348 8987, Fax +64-3-348 5548
www.niwa.co.nz

Contents

Reviewed by:



Jeff Smith

Approved for release by:

Guy Coulson

Executive Summary

Air pollution from particulates (as PM_{10}) represents a significant problem in many parts of New Zealand. Most large urban areas, and many small ones, have exceedences of the National Environmental Standard value for 24-hour concentrations of PM_{10} of 50 $\mu g\ m^{-3}$.

The high concentrations are due to emissions from various sources, and these need to be reduced. However, the relationship between emissions and concentrations is complex, since the weather and climate factors that lead to poor dispersion and high concentrations are highly variable. For a given level of emissions, some areas that are exposed and windy do not get exceedences, whilst others that are sheltered can experience unacceptably high values.

In order to formulate suitable mitigation strategies it is necessary to understand the relationship between the emissions and the weather in order to verify that mitigation actions are working, and to assess trends. This cannot be done by examining the monitoring results alone, since a particularly bad year might have higher concentrations even though the emissions are going down, or conversely a particularly good year might see lower concentrations that are due solely to the weather rather than any emissions reduction.

Objectives

This research has been undertaken to investigate in detail the complex relationship between emissions, weather and PM_{10} concentrations. The analysis has been conducted on monitoring data from central Christchurch over the 8 year period 1999 to 2006. Christchurch was chosen because Environment Canterbury has operated an active series of mitigation policies, and it was postulated that the analysis could help determine how effective these have been. The objective was not to test these policy options, but simply to analyse the monitoring results to increase understanding. The PM_{10} monitoring data from Christchurch are also of high quality and well suited to the analysis.

A key feature of this research is that it has been undertaken separately, and independently, by three separate collaborating groups. The work is part of the Foundation for Research Science and Technology programme “Protecting New Zealand's Clean Air”, and represents a combined effort to see if any, or each, of the three methods can produce useful results.

The data required were supplied to each of the groups by Environment Canterbury and has undergone a significant degree of quality control. The data used included (1) the hourly values of PM_{10} concentration at the Coles Place monitoring site in St Albans, and (2) weather parameters from the same site, comprising air temperature (at two heights), wind speed, wind direction, and relative humidity. The objective of each of the research groups was to see if any of these parameters, or combinations of them, could explain the substantial year-to-year variation found in the basic monitoring data (shown in Figure S1), and whether the impact of their variability could be removed

from inter-annual variations in air pollution concentrations (leaving a time series affected solely by changes in emissions).

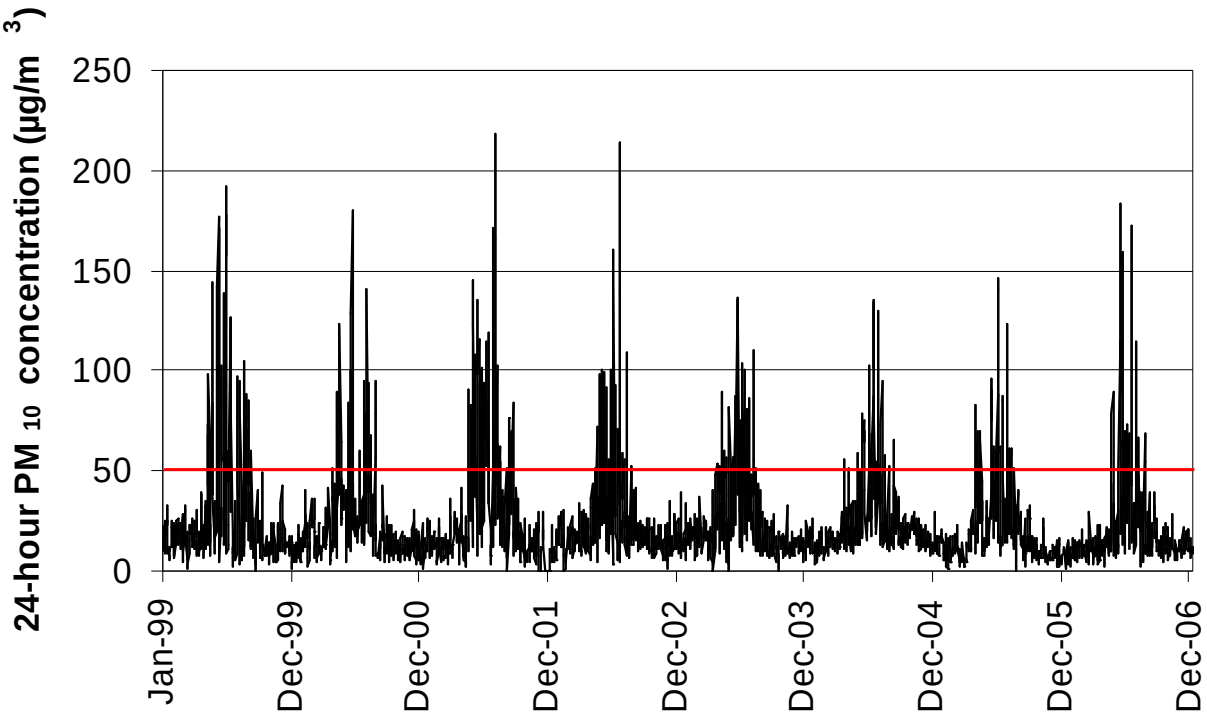


Figure S1: Time series of the 24-hour PM_{10} concentrations at Coles Place, 1999-2006 (the standard value is also shown by the red horizontal line).

Methodology

The first approach (by NIWA and Environet) consisted of a "regression tree analysis". This is a sophisticated model that takes all of the weather variables, along with the hourly PM_{10} concentrations, and progressively tries to make the best 'fit' to see which of the weather variables can best explain the variability in the PM_{10}.

The second approach (by Endpoint) was a "simple correlation analysis". This method examines the correlation between 24-hour PM_{10} concentrations and selected weather variables that might influence the concentrations, such as wind speed, number of calm periods, and air temperatures.

The third approach (by Canterbury University) was a "complex regression analysis". This method uses filtering and data transformation techniques to remove the influence of weather variables, and identify the underlying concentration variations due to emissions variations.

The basic results of the three approaches are shown in Figures S2, S3 and S4.

Results

Results from the three approaches are summarised.

Regression tree analysis: The regression tree analysis produces a series of relationships (regression trees) between the concentrations and various combinations of the weather variables. It then applies a range of statistical tests to determine which variables have the strongest influence on the concentrations. These variables can then be factored out of the basic concentration time series to indicate the longer term trend. The result is shown in Figure S2 for the percentage of winter days that experience exceedences of the standard. The results indicate a decreasing trend in frequency of exceedences since the peak in 2001.

Simple correlation analysis: Correlations were examined between PM_{10} concentrations and a number of weather parameters, including wind speed, air temperature, number of cold periods, number of calm periods, and temperature profiles. This was only carried out on daily data and focused on the winter months only. The results showed that the best relationship occurred between air temperature and concentration, shown in Figure S3. The simple correlation analysis indicates that over this 8 year period, for a given set of weather conditions, the air pollution concentrations have been decreasing, reflecting a lower level of emissions.

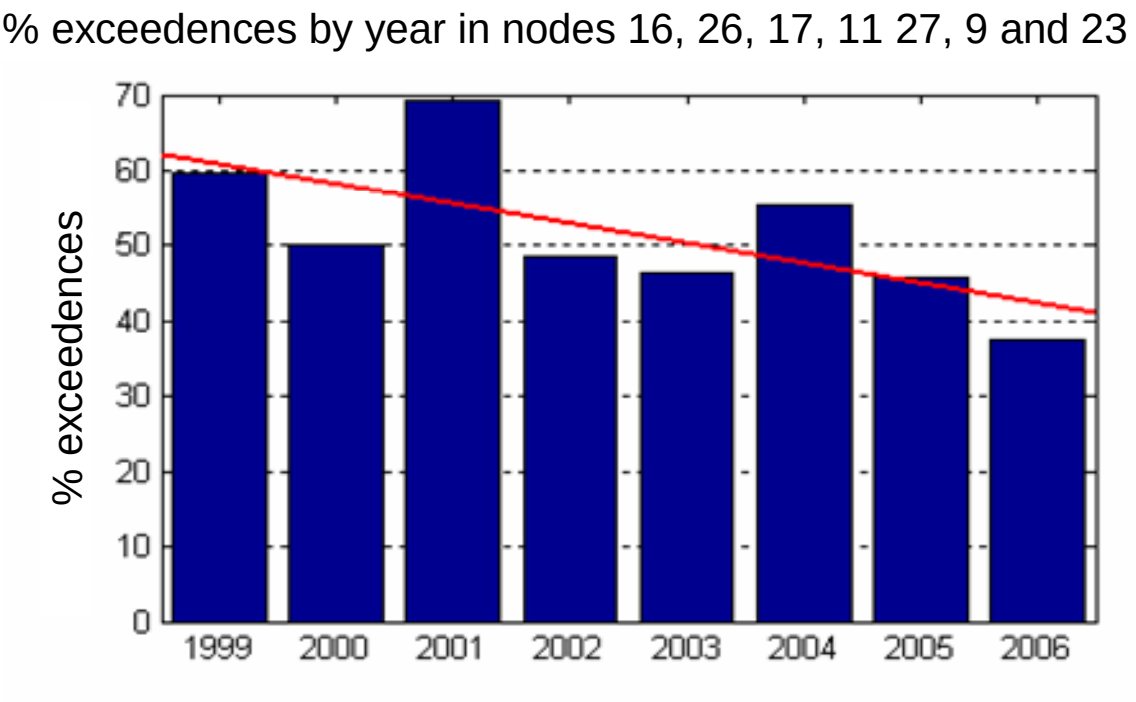


Figure S2: Regression tree analysis results showing the percentage of similar winter days experiencing an exceedence of the 24-hour PM_{10} standard.

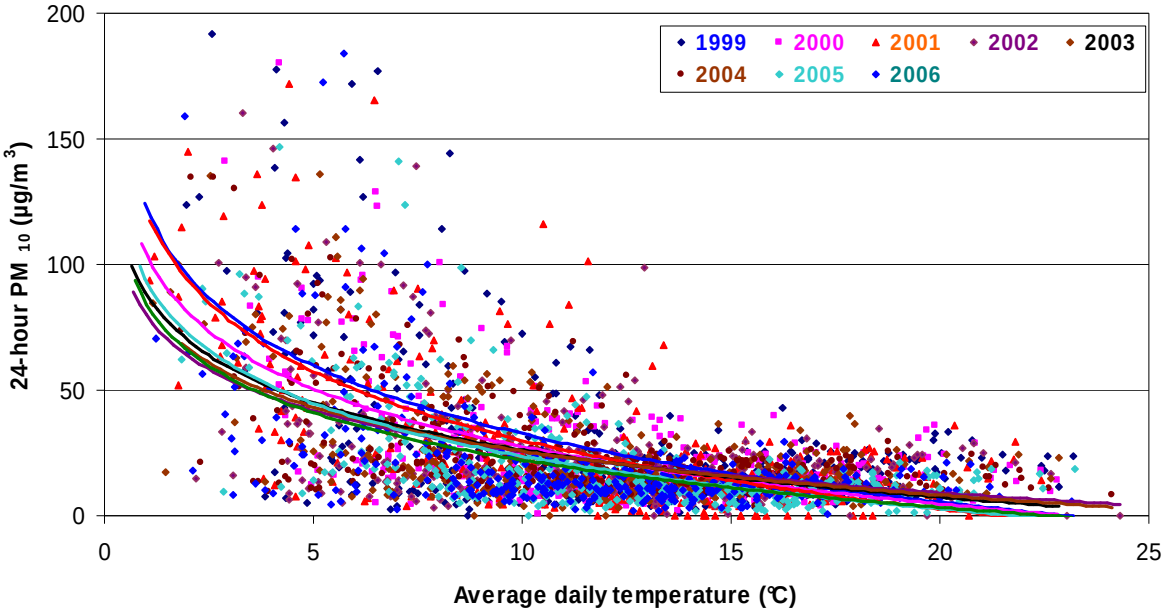


Figure S3: Yearly trends in 24-hour average PM_{10} with respect to average daily temperature (1999-2006). The curves fitted to the more recent years tend to be lower, indicating lower air pollution for a given air temperature**.**

Complex regression analysis: The focus of this part of the analysis was on data from the 5pm to midnight period, since this is when most of the air pollution emissions from home heating occur. Evening averages of PM_{10} and weather variables (5 pm – 12 am) were calculated from the hourly data. Following careful analysis of the distributions, average PM_{10} concentrations were transformed using the natural logarithm ($\log_e$) and wind speed using the square root. The effects of seasonal variations of temperature were then removed using a multi-stepped process, and then a multiple linear regression performed that removes the influence of the weather variables. The residuals between observed and predicted were then used to identify any trends that could not be explained by the weather variations, and finally the evening time series was smoothed using a 365 day moving average filter. The results are shown in Figure S4 – both the original data set, and the one with the influence of weather variables removed. It can be inferred from these results that emissions increased up to 2001/2002, and have decreased since.

Discussion

Each of these three analytical approaches, using the same input data have produced consistent results. It is not surprising that they all show some reduction in PM_{10} concentrations in central Christchurch, since even a cursory examination of the time series data (Figure S1) hints at this. What is more relevant is that each has shown (a) a very clear quantification of the influence of the weather and climate – with the main determinant being air temperature, and (b) a long term average decrease in peak concentrations. Although not analysed in detail, the studies indicate a reduction in the peak concentrations over the 8 year period of 3-4% per year that is largely independent of weather factors.

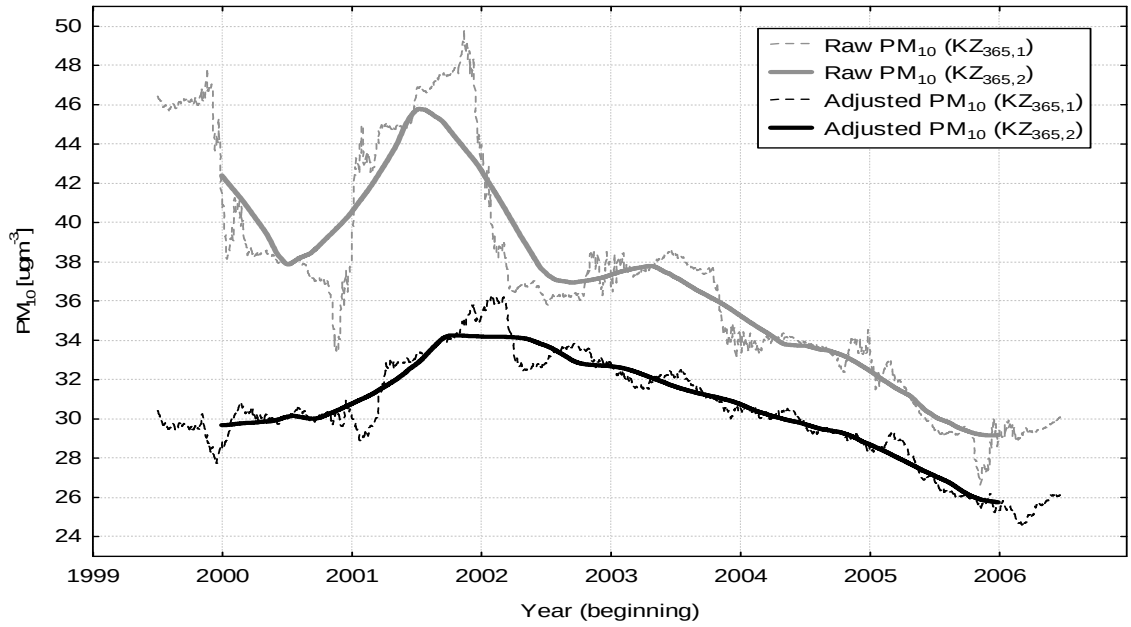


Figure S4: Smoothed trend lines of the average 5-12pm PM_{10} concentrations with the influence of weather variables removed.

It is not possible to make any solid predictions of future pollution concentrations, since factors influencing the emissions of PM_{10} may change. However, if this rate of decrease was linear, and if it was to be maintained – and these are both very big 'ifs' – central Christchurch will be closer to (if not meeting) the National Environmental Standard by 2013. The certainty of this outcome could be improved if this study was repeated every two to three years.

The research, whilst indicating some promising outcomes, must be interpreted cautiously. The data period is not very long and contains a significant degree of variability. A full analysis of the statistical significance of the results has not been carried out, and what was completed for the regression tree analysis showed that the results frequently did not meet stringent statistical significance criteria. It should also be noted that the main focus of research so far has been on controlling for the influence of weather variables on average concentrations. However, the National Environmental Standards refer to exceedences of specific values over certain time periods. It is the number of these exceedences that must be reduced by 2013 – this may, or may not, be related to a reduction in long-term average concentrations. The rationale for initially focusing on average values is that it allows an assessment of general trends of air pollution concentrations using more robust statistical techniques. Future work on analysis of the frequency distribution of hourly and 24-hour average concentrations, and non-parametric analysis of exceedences will strengthen the results presented here.

There is also the issue of the unexplained variability in the results, which indicates that there are other factors influencing the concentrations that have not been taken account of here. These include the influence of temperature inversions, more complex wind flows around the city, and just how often people use their solid fuel burners because of non-weather factors (e.g. week verses weekend use), fuel prices, special events, and unusual cold snaps, etc.

Summary

The PM_{10} concentrations measured in central Christchurch over the 8 years from 1999-2006 have been analysed using three different and independent techniques. Each has shown that differences in the weather from year-to-year can explain a significant amount of the variability in measured PM_{10}. The strongest relationship is with air temperature (colder weather means higher concentrations), but also with wind speed (calmer winds means higher concentrations). Other factors such as inversion strength also influence the concentrations.

Each of the studies suggests that there has been a reduction in PM_{10} emissions over the period, particularly since 2001. Although not a statistically strong result, it can be inferred from the studies that a reduction in emissions has averaged about 3-4% per year over 1999 to 2006.

Any, or all, of these techniques can also be applied to other areas in New Zealand to assess the influence of weather on PM_{10} concentrations, which will be different in different regions.

1. Introduction

1.1. Background

Assessing trends in air quality has long been problematic for air quality practitioners because of the variable impact of meteorology from year-to-year. In urban areas of New Zealand, the frequency of calm winds and temperature inversions affects the number of high pollution days. Meteorological factors also affect the magnitude of concentrations measured.

Assessing trends in air quality data in New Zealand is also limited in many areas by availability of data, with many datasets of limited duration or based on one day in three sampling regimes. Notwithstanding these difficulties, monitoring of trends in contaminant concentrations is a necessary task for councils to track the effectiveness of policy options in managing air quality.

The importance of evaluating trends in air contaminants, in particular in average concentrations of PM_{10}, has increased with the introduction of National Environmental Standards (NES) for air quality. The NES specify a limit of 50 $\mu g\ m^{-3}$ (24-hour average) for PM_{10} with only one allowable exceedence per year. They include restrictions on councils' ability to grant resource consents for air discharges in or affecting non-complying airsheds. The regulations include restrictions on issuing new resource consents from 2013 and a prohibition on granting resource consents for significant PM_{10} discharges if a linear reduction in PM_{10} concentrations from 2005 to 2013 is not met. The latter is referred to as the straight-line path (SLiP) (Fisher, Kuschel and Mahon, 2006) to compliance by 2013. Thus the ability of regulators to track trends in PM_{10} concentrations between 2005 and 2013 is a very important tool.

To date there has been limited investment in developing tools which enable regional councils to assess the trends in air pollution concentrations with time. This project will help plug this knowledge gap and assist regional councils in their task of moving toward compliance with the NES. The detailed analysis described in this report has been applied to data from Christchurch, as an example, with the objective of demonstrating the methodology for use in any region of New Zealand.

1.2. Aim and objective

The aim of this report is to provide regional councils with tools that will enable the use of air quality monitoring data to assess:

- Compliance with a region's SLiP
- The effectiveness of a region's air quality management strategy

The objective of this study is to develop and compare methods for assessing trends in PM_{10} emissions using monitored PM_{10} concentrations and meteorological data. Christchurch is used as an example because of its well-known air pollution problem and available data.

This objective can be distilled to address trends in emissions that may be of interest to environmental regulators:

- Trends in average PM_{10} emissions from year to year and more specifically from winter to winter.
- Trends in emissions on days when a PM_{10} concentration of 50 $\mu g\ m^{-3}$ is being exceeded: i.e. year to year trend in the number of NES exceedences.
- Trends in emissions on days when maximum PM_{10} concentrations are measured: i.e. year to year trends of peak 24-hour pollution events.

These distilled objectives can be achieved by sequentially refining the analysis of the data set to progressively reduce the meteorological influence on PM_{10} concentrations, i.e. consider more tightly defined meteorological events (e.g., a lower "low" wind speed classification, or a greater proportion of calm hours). Different approaches to evaluating medium to long term trends might be developed depending on which of these objectives is deemed most important.

It should also be noted that the main focus of research so far has been on controlling for the influence of weather variables on average concentrations. However, the National Environmental Standards refer to exceedences of specific values over certain time periods. It is the number of these exceedences that must be reduced – this may, or may not, be related to a reduction in long-term average concentrations. The rationale for initially focusing on average values is that it allows an assessment of general trends of air pollution concentrations using more robust statistical techniques. Future work on analysis of the frequency distribution of hourly and 24-hour average

concentrations, and non-parametric analysis of exceedences will strengthen the results presented here.

1.3. Multi-pronged approach to problem solving

Three project groups were established, each with the brief of developing a method to control for the influence of weather variables on average PM_{10} concentrations and to use this method to facilitate an assessment of emission trends in Christchurch over the period 1999 to 2006. The general approaches were as follows:

- Group 1: Regression tree analysis based on hourly and 24-hour PM_{10} and meteorological data.
- Group 2: Simple correlation and regression of 24-hour PM_{10} data and weather parameters.
- Group 3: Complex regression and filtering techniques by extracting the influence of specific weather parameters using hourly data.

The methods and outputs were compared and their effectiveness and the overall consistency of the results were considered. Methods and results were also compared with a study undertaken by Canesis in 2004 (Environment Canterbury, 2004). The outcomes of the Canesis study are summarised in Section 2.

1.4. Structure and content of the report

The report is structured to provide easy to follow and concise information. The content of the report has been selected to provide relevant and practical guidance to air quality resource managers, scientists, analysts and technicians. The report is structured:

- Background and overview of the study – Section 1
- Review of previous investigation into Christchurch PM_{10} trends – Section 2
- Summary of regression tree analysis – Section 3
- Summary of simple correlation and regression analysis – Section 4
- Summary of complex regression analysis – Section 5

- Discussion and comparison of results – Section 6

- Conclusions, gaps and recommendations for future work – Section 7

Details of the method and the full set of results from the regression tree, simple correlation and regression, and complex regression analyses are provided in Appendices A, B and C respectively.

1.5. Monitoring site, monitoring methods, datasets and definitions

The datasets used in the study were provided by Environment Canterbury and consisted of all data collected at the Coles Place (St Albans, Christchurch) air quality monitoring site. A map and photograph of the Coles Place monitoring site are shown in Figure 1.1 and Figure 1.2 respectively.

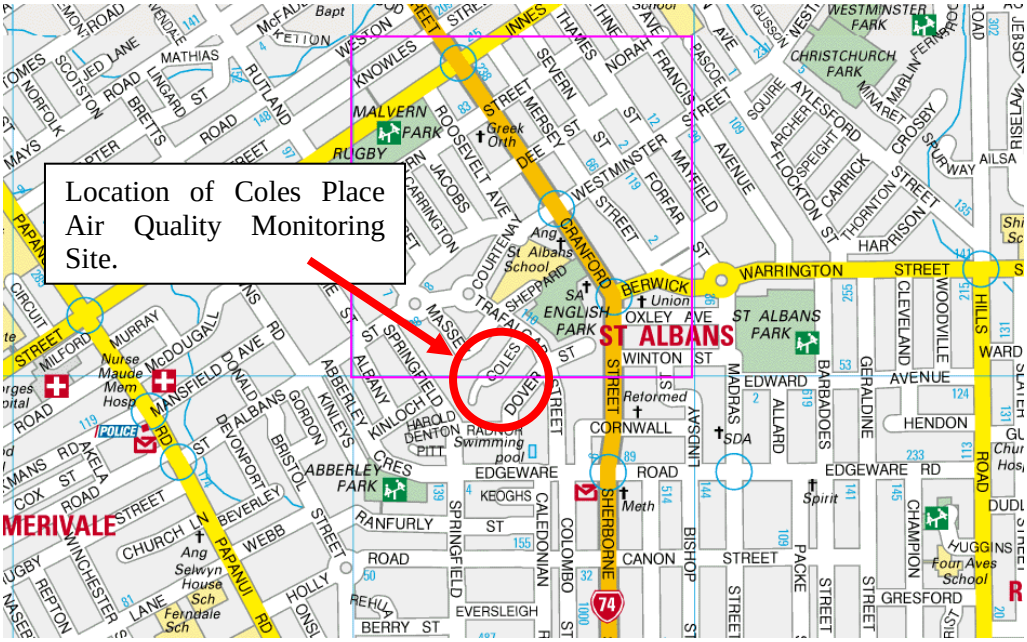


Figure 1.1: Map showing the location of the Coles Place monitoring site.

The Coles Place monitoring site was commissioned by Environment Canterbury (ECan) in August 1998. It was established as an alternative to the historic Packe St site (also located in St Albans), which was situated on a block of land that was to be redeveloped. As the Packe St site was able to be used until 2002, a comparison of concentrations measured at both sites could be made during the overlap period. Concentrations of PM_{10} varied little between these sites.

In Christchurch, PM_{10} has been monitored continuously using the Rupprecht and Patashnick Co., Inc. Tapered Element Oscillating Microbalance (TEOM) particulate

monitors since 1994. The operation of these instruments is automated and they can be interrogated remotely, with data available every ten minutes. This method of measurement heats the air sample to minimise water affecting the particle weight, but volatile particulate is lost during heating and the total mass of particulate is reduced. From 1999, PM_{10} was monitored with a TEOM operated with an inlet temperature of 40°C to reduce the loss of the volatile fraction of PM_{10}. A trial was carried out in 2003 and 2004 whereby a filter dynamic measurement system (FDMS) was integrated with a TEOM operating at 30°C (TEOM-FDMS). Concentrations from the TEOM-FDMS were found to be very similar to those from the High Volume Sampler. As a result of this trial, in 2005 Environment Canterbury adopted the TEOM-FDMS for the purpose of reporting PM_{10} concentrations under the requirements of the National Environmental Standards.

Figure 1.2: Photograph of the Coles Place monitoring site.

To avoid inconsistencies between the analyses, the following definitions were agreed between the three project groups:

- A PM_{10} exceedence equals a 24-hour average PM_{10} concentration greater than 50 µg m^{-3} when measured from midnight to midnight using the FDMS or data adjusted for FDMS equivalence.

- The equations for adjusting 24-hour average TEOM data measured at 40°C (1999 to 2004) to FDMS equivalence were:

  TEOM@40>=44, FDMS equivalent = (TEOM@40+3.15)/0.75

  TEOM@40<44, FDMS equivalent = (TEOM@40-2.23)/0.74

- 1-hour average TEOM data measured at 40°C (1999-2004) were adjusted to FDMS equivalence using the method described in Appendix A.

- The 24-hour average period from midnight to midnight is calculated from hourly data using the hours 01:00 on reporting date to 00:00 on the day following the reporting date.

- The annual peak PM_{10} concentration is the highest 24-hour average PM_{10} concentration measured between January and December for the reporting year, adjusted to FDMS equivalence.

- The annual average PM_{10} concentration is the average of the 24-hour average PM_{10} concentrations from 1 January to 31 December for the reporting year, adjusted to FDMS equivalence.

The regression tree and complex regression studies used 1-hour average PM_{10} concentrations measured at Coles Place over the years 1999 to 2006. The 1999 to 2003 data were measured using a TEOM operated at 40°C. The 2004 to 2006 data were recorded using a TEOM-FDMS. The TEOM(40) data (1999 to 2003) were adjusted to be TEOM-FDMS equivalent measurements using the method detailed in Appendix A.

The regression tree and simple correlation studies also used 24-hour average PM_{10} concentrations measured at Coles Place over the years 1999 to 2006. The 1999 to 2004 data were measured using a TEOM operated at 40°C. The 2004 to 2006 data were recorded using a TEOM-FDMS. The TEOM(40) data (1999 to 2003) were adjusted to be TEOM-FDMS equivalent measurements using the methodology established by Environment Canterbury and as detailed in bullet point 2 above.

1.6. Limitations of the study

Emissions of PM_{10} vary hour-to-hour, day-to-day, month-to-month and year-to-year. Separating out the temporal variability in PM_{10} emissions from the variability occurring as a result of different meteorological impacts is a very complex task.

This study utilised one of the longest and most comprehensive PM_{10} data sets available in New Zealand. A three pronged approach to problem solving was employed to ensure the results presented are as robust as practically possible. However, there are limitations to this study which introduce some uncertainty into the results and prevent a definitive solution being provided. For example despite the data record being one of the longest available in New Zealand it is constrained to eight years.

The limitations of the data and methods used in this study are noted in the relevant sections of the report and potential enhancements are suggested in the report's conclusions. While all care has been taken to produce robust and useful findings, the conclusions presented in the report must be kept in context of the study's limitations.

1.7. Trends in PM_{10} concentrations – Christchurch 1999 to 2006

Figure 1.3 shows the annual 1-hour average PM_{10} concentrations as measured at the Coles Place monitoring site for 1999 to 2006. Data capture at the site is very good, with each year having at least 95% of the available data. A total of approximately 69,000 hours of data is presented in Figure 1.3. The blue boxes in Figure 1.3 represent the inter-quartile range and the white diamond the median value.

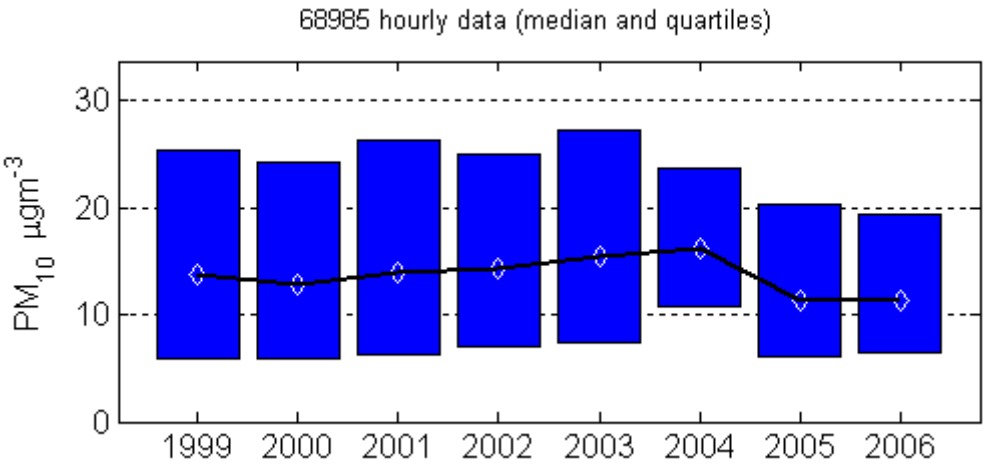


Figure 1.3: Trends in hourly average PM_{10} concentrations at Coles Place, Christchurch 1999-2006.

Figure 1.4 shows the annual 24-hour average PM_{10} concentrations as measured at the Coles Place monitoring site for 1999 to 2006.

The statistical significance (95% confidence interval) of the year-to-year differences in 24-hour average PM_{10} concentrations was investigated using the Kruskel-Wallis test (a nonparametric version of the classical one-way ANOVA). Details of the Kruskel-Wallis (K-W) method and test results are presented in Appendix A.

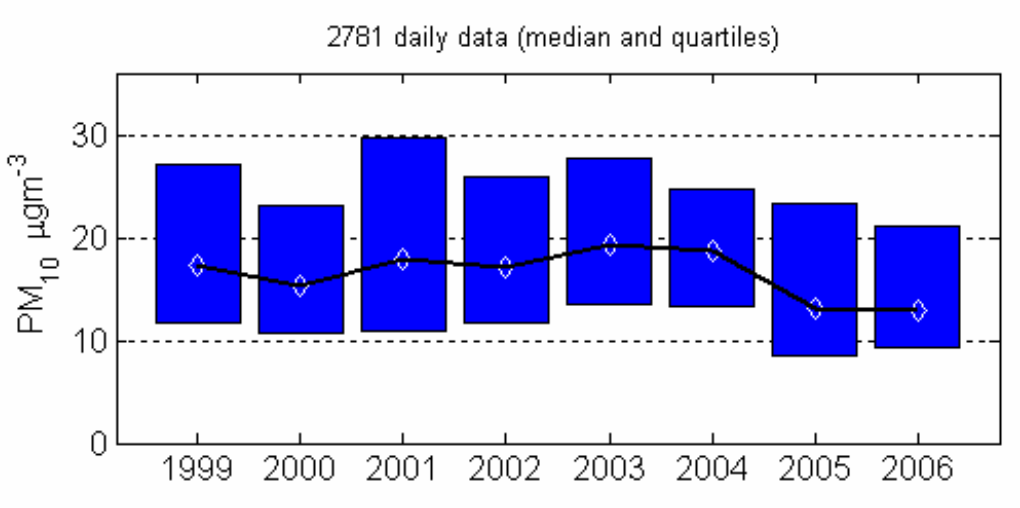


Figure 1.4: Trends in 24-hourly average PM_{10} concentrations at Coles Place, Christchurch 1999-2006.

The results of the K-W test show that there is a significant increase in both 1-hour and 24-hour concentrations from 1999 until peak values are reached in 2004. Following this peak there is a significant decrease to 2005. The 2006 1-hour and 24-hour concentrations are not significantly different from 2005. It is important to note that the year-to-year trends demonstrated in Figure 1.3 and Figure 1.4 do not account for the year-to-year variation in meteorology.

2. Previous assessment of trends in Christchurch's PM_{10} concentrations

Canesis Network Ltd were engaged by ECan to assess the effect of its air quality management policy of reducing the emission from solid-fuel burning domestic space heating devices. To extract information on changing domestic emissions from ECan's ambient air quality data, Canesis employed a new measure of air quality - "*qualifying evenings*". The results of this study are presented in full in a report entitled *An Alternative Measure of Air Quality* (Environment Canterbury, 2004), while the key points from that report are summarised below.

2.1. PM_{10} and meteorological data

The Canesis report was produced in 2004 and at that time ECan had a PM_{10} data record that ran from 1996 to 2003. There were two inherent issues with the PM_{10} data set, that Canesis had to address before being able to use it as a continuous record. The data were recorded:

- Using TEOMs operating with varying inlet temperatures: 30, 40 and 50 ºC
- At two different sites: Packe Street (1996 to 1999) and Coles Place (1999 to 2003).

To assess the trends in particulate emissions it is highly desirable to have data recorded at one site, using one method. Details on how data from the two sites measured by TEOMs with different inlet temperatures were synthesised into a continuous series are provided in Section 3 of the Canesis report. Very briefly, the PM_{10} data collected by the TEOM at various inlet temperatures were all transformed by regression analysis to TEOM(40) equivalent data. The site-to-site differences were accounted for by defining relationships of wind speed, temperature and inversion strength between the two sites. The Packe Street PM_{10} data were then mapped to the Coles Place site using the site-to-site meteorological relationships.

The result of these data transformations was to create TEOM(40) equivalent PM_{10} and meteorological data sets that were specific to the Coles Place site for the years 1996 to 2003.

2.2. Qualifying evenings

The meteorological data were then used to define "*qualifying evenings*" which would facilitate the evaluation of the effect of emissions from solid fuel domestic space heaters on ambient air quality data. Canesis assumed that emissions from solid fuel domestic space heaters exert the most influence on ambient air quality, and are therefore most easily measured when there is:

- A temperature inversion – this traps pollution within a shallow vertical layer.
- Little wind – limits dispersion (dilution) of pollution
- Low temperatures (cold) – people light fires and the PM_{10} in the urban atmosphere is emitted mainly from home heating.

These three meteorological criteria resulted in Canesis focusing their analysis on data monitored between 6pm to midnight over the months April to September.

Qualifying evenings were defined using three meteorological variables:

- Temperature measured at 1 m
- Wind speed
- Temperature inversion – difference in temperature measured at 1 and 10 metres above ground level

A qualifying evening had to have all thee variables lower than "threshold values". However, Canesis found the process of defining threshold values "a bit subjective". They investigated 40 different definitions of qualifying evenings using combinations and permutations of threshold values. At the end of this process, threshold values for a "reference scenario" were chosen:

- Temperature measured at 1 m had to be below $9^{o}C$
- Wind speed had to be lower than 1.5 m s^{-1}
- Temperature inversion (temperature at 1 m minus temperature at 10 m) had to be less than $-0.3^{o}C$

Only PM_{10} data from qualifying evenings (when reference scenario thresholds are met) were used in the analysis. For qualifying evenings, the hourly PM_{10} values between 6 pm to midnight were used to find a mean and median value. This PM_{10} value was taken as representative of that particular evening. By concentrating on the representative PM_{10} values, Canesis suggested it should be possible to extract information regarding emissions from solid-fuel domestic heating.

2.3. Normalising data

Averaged PM_{10} over the qualifying evenings was plotted for each year. The scatter of data within years was found to be large. Canesis concluded that the large range was due in part to varying weather conditions. Scatter within qualifying evenings was reduced by normalising PM_{10} with respect to weather conditions (specifically wind speed and temperature measured at 1 m), thus reducing scatter in the data and increasing the data's (trends) significance. Details of the normalisation procedure are

provided in Section 4.2 of the Canesis report. Canesis's normalisation procedure also accounts for year-to-year variation in climatic conditions.

The normalising procedure removed between 5% and 40% of the scatter within each year, depending on the thresholds used to define qualifying nights. The causes of the remaining scatter are unknown.

2.4. Results

Canesis found a "*significant*" downward trend — via a least squares linear fit — of qualifying evening PM_{10} values over the period 1996 to 2003. Canesis's findings were independent (within reason) of the combination of variables used to define qualifying evenings and independent of whether mean or median PM_{10} values were used in the analysis.

2.5. Conclusions

Canesis concluded that:

- Emissions from solid fuel domestic space heaters appear to be reducing by between 2 and 6% every year.

- Measuring the change in emissions from solid fuel domestic space heaters using normalised qualifying evening PM_{10} data was robust, since a qualitatively similar downward trend was observed for all reasonable definitions of "qualifying" and independent of techniques used (average, median, etc.).

- It is important to understand the causes of the remaining scatter in PM_{10} values and further reduce it, as this will further improve confidence in the year-to-year downward trend.

3. Regression tree analysis

3.1. Introduction

The method used to analyse the impact of meteorology on PM_{10} concentrations was to allow for statistical groupings of the relationship between PM_{10} and meteorological variables using a regression tree analysis. An overview of the regression tree method is provided in Appendix A. A full technical description of the method can be found in

De'ath and Fabricius (2000). Very briefly, the regression tree identifies which meteorological variables cause the most variation in the PM_{10} concentrations. The tree model then clusters the PM_{10} data into groups with similar "*predictor variables*". In the Coles Place PM_{10} hourly data set the strongest predictor variables included month, hour of day, one metre air temperature and wind speed. The PM_{10} data from a particular group (with similar predictor variables) is then broken down into year and subjected to trend analysis. The year-to-year differences in PM_{10} were tested for statistical significance using the K-W test.

3.2. Method

A four staged approach was used, with the regression trees fitted using Matlab (Mathworks version 2006a).

1. **Complete 1-hour average PM_{10} data set.** The entire hourly average PM_{10} data (68,985 hours) were clustered using a regression tree. This created groups of PM_{10} data defined by similar meteorological and temporal variables. Within each node some of the effect that varying meteorological and temporal parameters have on PM_{10} concentrations had been removed (or controlled). The data contained in the group with the highest PM_{10} values (1604 hours) was broken down by year and the year-to-year variation in hourly average PM_{10} concentrations examined.

2. **High concentration hourly average PM_{10} data:** The subset of hourly average PM_{10} data representing high concentrations (from stage 1) were subjected to a second and more refined regression tree analysis. This created sub-groups within the high pollution subset within which some further effect of meteorology on PM_{10} had been removed. The data contained in a combination of these high pollution nodes (1054 hours) were broken down by year and the year-to-year variation in hourly average PM_{10} concentrations examined.

3. **Complete 24-hour average PM_{10} data:** The entire 24-hour average PM_{10} data set (2781 days) was clustered using a regression tree. The data contained in the group with the highest PM_{10} values (206 days) was broken down by year and the year-to-year variation in 24-hour average PM_{10} concentrations examined.

4. **Trends in exceedences of 50 µg m^{-3}:** The entire 24-hour average PM_{10} data set (2781 days) was clustered using a regression tree. The top seven

meteorological classifications (490 days) were selected to provide a range of PM_{10} concentrations from extreme to marginal (above 45 µg m^{-3}) NES exceedences. The data contained within these seven groups were broken down by year and the year-to-year trends in the proportion of days producing an NES exceedence was examined.

The 1999-2006 PM_{10} and meteorological datasets were fairly complete. Less than 2% of the hourly data was missing and less than 5% of the daily data was missing. In the hourly dataset, if either the PM_{10} or any of the meteorological data was missing, the hour was discarded. In the 24-hour dataset, days were only used if 75% of their hourly data was available.

3.3. PM_{10} concentration trend analysis

3.3.1. Identifying and grouping high hourly average PM_{10} data

The hourly average PM_{10} data set contains a large number (68,985 hours) of data and has a skewed distribution, with many low values and fewer high values. With a skewed data set such as this one, there is a risk that the influence of the fewer high value data will be out of proportion to their number. This issue can be dealt with by reducing the magnitude of difference between high and low values. A logarithmic transformation could have been used. However, to retain the importance of the high values, and because tree models do not require a normal distribution, a square root transformation was chosen. The transformed data set was then subjected to a regression tree analysis. Detailed results from the regression tree analysis can be found in Appendix A.

The main predictor variables used to categorise the PM_{10} data were temperature, wind speed, month of the year, time of day and temperature inversion. Some of these predictor variables, such as month of year, temperature and time of day are likely to be segregating data based on variations in PM_{10} emissions.

The regression tree gave a coefficient of determination (r^2) of 0.38. This indicates that around 38% of the variation in PM_{10} concentrations are accounted for by the relationships between the predictor variables (both meteorology and temporal) and PM_{10} concentrations described by the regression tree. This shows that the regression tree analysis is reducing some of the impact that year to year variation in meteorological conditions have on PM_{10} concentrations.

The high pollution group was defined by the following predictor variables: temperature measured at 1 m was less than 4.6°C, hours of the day were between 6 pm to 3 am and a temperature inversion (warmer temperature at 10 m than at 1 m) was at least -1°C. It is important to note that the data contained in the high pollution hourly subset were defined by critical meteorological conditions and not by PM_{10} values. However, the 1604 hours of data in the high pollution group identified by the regression tree had a mean value just above 144 µg m^{-3}, which is approximately four times higher than the next group.

3.3.2. Time trend analysis on high pollution hourly data

The high pollution hourly PM_{10} concentrations subset was disaggregated by year and the resulting year-to-year trend is displayed in Figure 3.1. The blue boxes in Figure 3.1 represent the inter-quartile range and the white diamond the median value. There were at least 120 data points for each year.

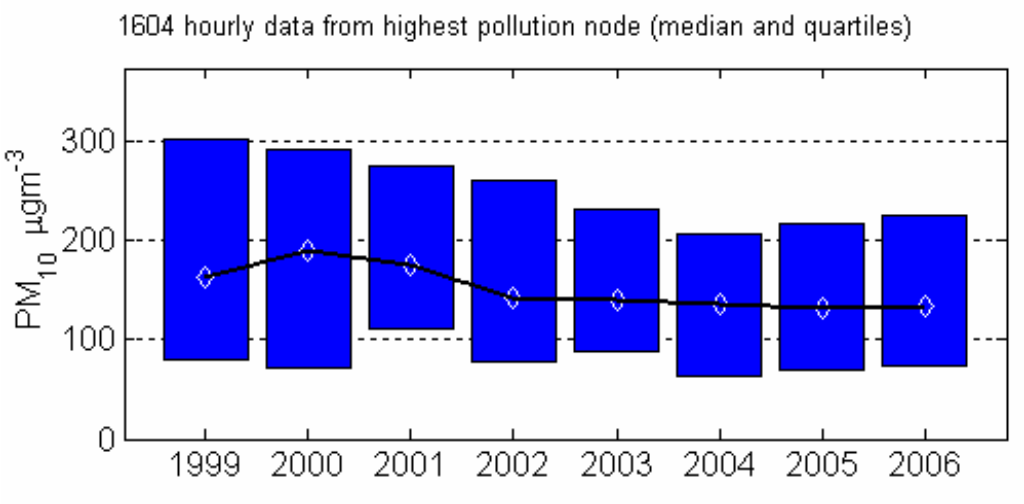


Figure 3.1: Year-to-year variation in hourly average PM_{10} concentrations within the high pollution sub-group.

Figure 3.1 shows that within the high pollution group the highest hourly average PM_{10} concentrations tended to be recorded in 2000. Between 2000 and 2004 there has been a general decrease in concentrations, but the decrease has slowed or even stopped in 2005 and 2006. The year-to-year differences were explored using the K-W test. The results of the K-W analysis show that the decrease in hourly average PM_{10} concentrations between 2000 and 2004 is statistically significant (95% confidence interval (CI)). The concentrations recorded in 2005 and 2006 were not significantly different to those recorded in 2004 (See Figure A3). The decrease in median values of

1-hour average PM_{10} concentrations displayed in Figure 3.1 over the period 1999 to 2006 is 30 μg m^{-3}. This represents an annual average decrease of 2.3%.

To assess the effectiveness of the tree model in reducing the meteorological variability, a multiple linear model was fit to the high pollution hourly subset of data. The percentage variation explainable by meteorological predictors, using a multiple linear regression (MLR) model reduced from 29% on the complete dataset to 18% on the high pollution subset. This shows that the refining of the dataset has succeeded in removing some of the variability in PM_{10} due to meteorology. More detail is provided in Appendix A.

3.3.3. Identifying and grouping extreme hourly average PM_{10} data

In an attempt to remove as much of the PM_{10} variability due to meteorology as possible, a second regression tree analysis was conducted on the untransformed hourly average high pollution data group (1604 hours). Data transformation was not necessary because of the revised data range and the resulting frequency distribution of the high pollution category. The regression tree analysis sub-divided the high pollution data group into nine separate meteorological or seasonal groupings.

The extreme high pollution group identified by the second regression tree contained 1054 hours of data with a mean value of about 225 μg m^{-3}. The extreme high pollution group was defined by the following predictor variables: wind speed less than 1.17 m s^{-1}, months April to August.

3.3.4. Time trend analysis on extreme high pollution data

The extreme high pollution hourly average PM_{10} concentrations were disaggregated by year (resulting in at least 79 data points for each year). The resulting year-to-year trend is displayed in Figure 3.2.

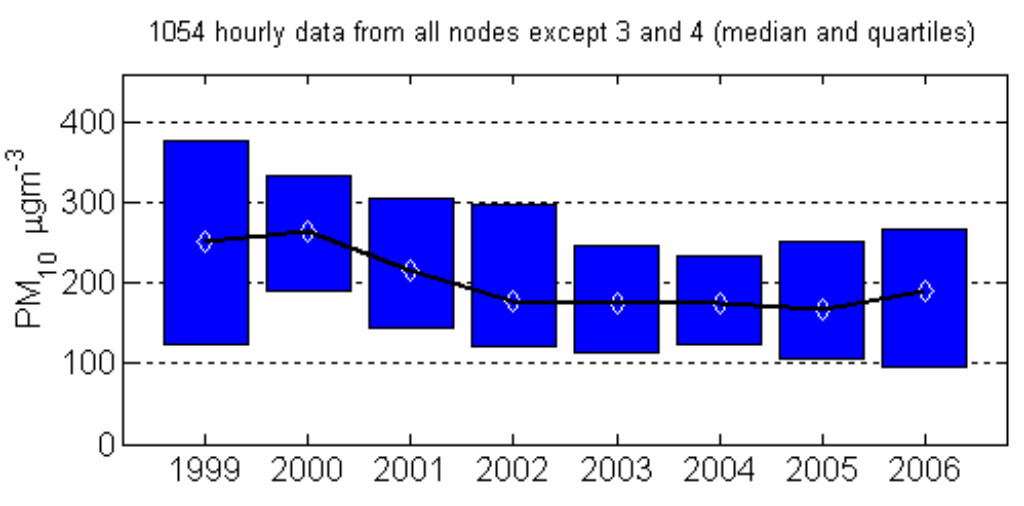


Figure 3.2: Year-to-year variation in hourly average PM_{10} concentrations within the extreme high pollution sub-group.

Figure 3.2 shows a very similar result to that displayed in Figure 3.1. Within the extreme high pollution group the highest hourly average PM_{10} concentrations tended to be recorded in 2000. Between 2000 and 2004 there has been a general decrease in concentrations, but the decrease has slowed or even stopped in 2005 and 2006. The year-to-year differences were explored using the K-W test. The results of the K-W analysis show that decrease in hourly average PM_{10} concentrations between 2000 and 2004 is statistically significant (95% CI). The concentrations recorded in 2005 and 2006 were not significantly different to those recorded in 2004 (See Figure A7). The decrease in median values of 1-hour average PM_{10} concentrations displayed in Figure 3.2 over the period 1999 to 2006 is 62 µg m^{-3}. This represents an annual average decrease of 3.1%.

The regression tree for the high pollution sub-group produced a coefficient of determination (r^2) of 0.33, compared to 0.38 for the complete data set. This shows that the regression tree analysis for the high pollution subgroup is further reducing the impact that year-to-year variation in meteorological conditions has on PM_{10} concentrations.

A multiple linear regression model was also fit to the extremely high pollution subset. This model concluded that weather variables explained only 9% of the variation in the extreme pollution data, and indicates that the further refining of the high pollution subset has succeeded in removing more of the variability in PM_{10} due to meteorology. More detail is provided in Appendix A.

3.3.5. Identifying and grouping the high 24-hour average PM_{10} data

The 24-hour average PM_{10} data set contains 2,781 days of data and has a skewed distribution, with many low values and fewer high values. The square root transformed data set was subjected to a regression tree analysis. Detailed results from the regression tree analysis can be found in Appendix A.

The main predictor variables used to categorise the 24-hour PM_{10} data were wind speed, month of the year, one metre air temperature, and temperature inversion. Wind speed was the most important predictor in the model and contributes just over half of the explanatory power of the model.

The regression tree gave a coefficient of determination (r^2) of 0.65. This indicates that around 65% of the variation in PM_{10} concentrations was accounted for by the relationships between the predictor variables (both meteorology and temporal) and PM_{10} concentrations described by the regression tree.

The tree grouped the 24-hour data into 15 nodes, three of which were identified as high pollution days. The 3 high pollution groups identified by the regression tree contained 206 days of data with a mean value of about 80 µg m^{-3}. The highest pollution group had a mean value of 90 µg m^{-3} and was defined by the following predictor variables: wind speeds lower than 1.2 m s^{-1}, winter months (May, June, July and August), one metre air temperature less than 8.1°C and a temperature inversion of at least -0.1°C. The second and third highest pollution groups had mean values of 80 and 70 µg m^{-3} respectively, and were defined by very similar predictor variables, but with slightly higher wind speeds. The 3 high pollution groups were aggregated to create the high pollution 24-hour subset.

3.3.6. Time trend analysis of high pollution 24-hour average PM_{10} concentrations

The high pollution 24-hour PM_{10} data were disaggregated by year (resulting in at least 14 data points per year). The resulting year-to-year trend in 24-hour average PM_{10} concentrations is displayed in Figure 3.3.

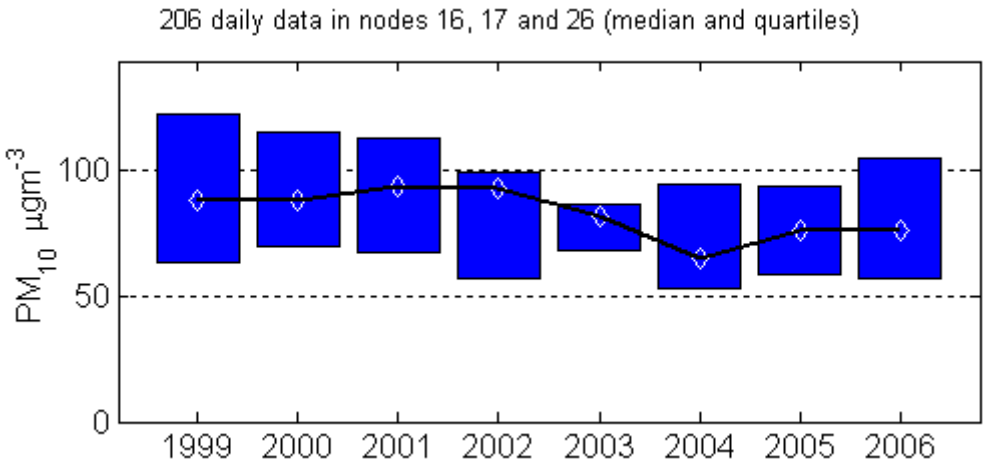


Figure 3.3: Year-to-year variation in 24-hour average PM_{10} concentrations within the 3 high pollution groups.

Figure 3.3 shows that within the high pollution subset, the highest 24-hour average PM_{10} concentrations tended to be recorded in 2001 and 2002. Between 2002 and 2004 there has been a general decrease in concentrations, but the decrease has stopped or even reversed in 2005 and 2006. The year-to-year differences were explored using the K-W test. The results of the K-W analysis show that the decrease in 24-hourly average PM_{10} concentrations between 2001 and 2004 was not statistically significant (95% CI). Also, the concentrations recorded in 2005 and 2006 were not significantly different to those recorded in 2004. (See Figure A10). The decrease in median values of 24-hour average PM_{10} concentrations displayed in Figure 3.3 over the period 1999 to 2006 is 11.3 µg m^{-3}. This represents an annual average decrease of 1.6%.

A second regression tree analysis performed on the high pollution group did not assist in further evaluating the impact of meteorology because of the limited number of days within this classification (206 data points). However, the percentage variation explainable by meteorological predictors, using a MLR (multiple linear regression) model reduced from 61% with the complete dataset to 28% with the high pollution subset. This shows that the refining of the dataset has succeeded in removing some of the variability in PM_{10} due to meteorology. More detail is shown in Appendix A.

3.3.7. Trends in exceedences of 50 µg m^{-3}

PM_{10} data for the period 1999 to 2006 breaches the NES concentration of 50 µg m^{-3} (24-hour average) on 286 of the 2781 sample days (~10%). The majority of these exceedences (69%) occurred within the three high pollution groups analysed in

Section 3.3.5. These groups typically represent meteorological conditions during which the worst case PM_{10} concentrations occur.

To assess the changes in the number of NES exceedences associated with a reduction in PM_{10} emissions requires the capture of less extreme meteorological conditions, i.e., those days currently or previously resulting in concentrations around 50 µg m^{-3}. To increase the capture of these "*marginal days*", additional groups from the regression tree analysis undertaken in Section 3.3.5 were added. The top seven meteorological classifications from the regression tree were selected as being representative of marginal days. This combination of groups contained 490 days (18% of the days) and contained 253 exceedences (88% of the total exceedences recorded). This grouping of the 24-hour average PM_{10} data (extreme and marginal days) was disaggregated by year and the percentage of days greater than 50 µg m^{-3} was calculated for each year. Figure 3.4 shows the year-to-year variation of the percentage of marginal days with PM_{10} concentrations of greater than 50 µg m^{-3} (24-hour average).

Figure 3.4 indicates that in 1999 around 60% of the PM_{10} concentrations on marginal days resulted in NES breaches compared with just less than 40% during 2006. The data displayed in Figure 3.4 suggest that over the period 1999 to 2006 there has been an annual average decrease of 2.8% in the chance that a marginal day will produce an exceedence of the PM_{10} NES. The most probable explanation for this is a decrease in PM_{10} emissions.

3.4. Discussion

The approach used in this section to evaluate trends in PM_{10} concentrations with time involved three steps: (1) grouping PM_{10} data based on similar predictor variables (meteorological conditions, month of year and time of day), (2) disaggregation of these groups into year of monitoring, and (3) analysis of the resulting year-to-year trend. Trend analyses were undertaken using both 1-hour and 24-hour concentrations.

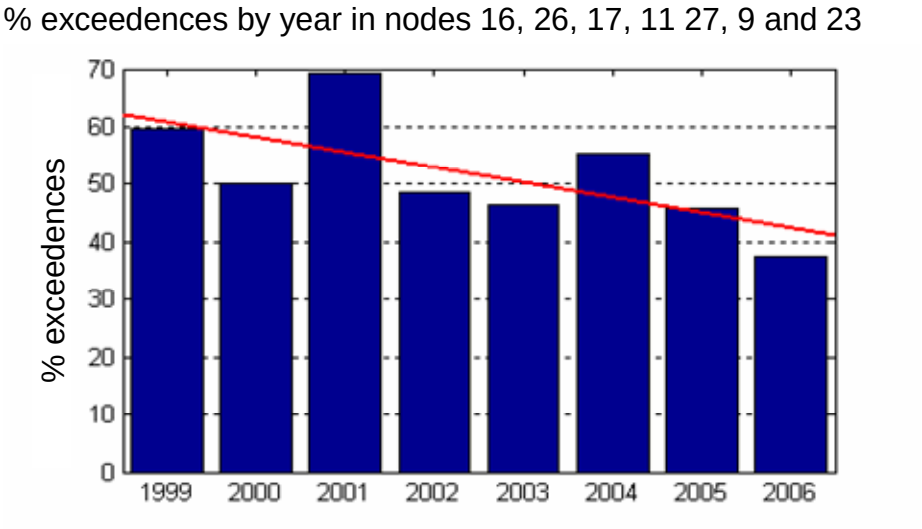


Figure 3.4: Year-to-year variation of the percentage of marginal days with PM_{10} concentrations of greater than 50 µg m^{-3} (24-hour average).

The regression tree analysis identified that the highest 1-hour PM_{10} concentrations occurred most frequently when wind speed was 1 m s^{-1} or less, one metre air temperatures were below 2°C, a temperature inversion existed, the time was between 18:00 to 03:00 and the seasons were late autumn and winter. The predictor variables objectively identified by the regression tree very closely matched those that would intuitively be expected.

From the almost 70,000 hours of 1-hour average PM_{10} data recorded between 1999 and 2006, a high pollution sub-set was identified. This high pollution group contained approximately 1600 hours of data (2.3% of total) with a mean value of 144 µg m^{-3}. An additional analysis of the high pollution data set was undertaken to produce an extreme high pollution group. This extreme high pollution group contained approximately 1000 hours of data (1.5% of total) with a mean value of 225 µg m^{-3}. Trend analysis was undertaken on both the high and extremely high pollution groups, and both groups of data showed very similar trends. The highest 1-hour average PM_{10} concentrations tended to be recorded in 2000. Between 2000 and 2004 there has been a significant decrease in concentrations, but the decrease has slowed or even stopped in 2005 and 2006.

These results suggest that there was a general decrease in PM_{10} emissions over the period 1999 to 2004, but the decrease has slowed or perhaps even halted in more recent years. However, it is possible that in these later years, some of the impact of meteorology is not accounted for by the relationships described by the regression trees. It should also be remembered that not all effects of meteorology will have been removed and the effect of the remaining meteorology is difficult to quantify.

The regression tree analysis found the highest 24-hour PM_{10} concentrations to occur most frequently when: wind speeds were lower than 1 m s^{-1}, winter months, ground temperature was less than 8°C and a temperature inversion existed. From the almost 3,000 days of PM_{10} data recorded over 1999 to 2006 a high pollution sub-set was identified. This high pollution group contained approximately 200 hours of data (7% of total) with a mean value of approximately 80 µg m^{-3}. Trend analysis was undertaken of this high pollution group. The highest 24-hour average PM_{10} concentrations tended to be recorded in 2001. Between 2001 and 2004 there has been a general decrease in concentrations, but the decrease has slowed or even stopped in 2005 and 2006. This is a very similar trend to that observed in the 1-hour average data, although in this example the difference in concentrations between 2001 and 2004 was not statistically significant. The trends observed in the 24-hour average PM_{10} concentrations suggest that emissions of PM_{10} over the period 2001 to 2004 may have decreased, but not significantly so.

An initial reaction to the 24-hour average result could be that it is inconsistent with that observed in the 1-hour average data. However the emissions from home heating do not occur consistently over a 24-hour period. Most of the heating activity (and emissions) occurs in the evening (4pm to 10pm) with a smaller, but still important, activity period between 06:00 and 10:00. Therefore the effect of any home heating emission reduction strategy is going to be most strongly observed on an hour-to-hour time scale. When the time scale of observations is increased from 1-hourly to 24-hour the total amount of PM_{10} reduction remains the same, but its effect will be smoothed over a longer period and therefore relatively small and harder to identify. So it is possible that a significant reduction in PM_{10} emissions is occurring, although anticipated reduction in 24-hour average PM_{10} concentrations is not observed as strongly as anticipated due to the effect of the longer period averaging period.

Another very significant variable impacting on the 24-hour average approach is the simplification of the meteorological parameters (24-hour averages) used in the model. In particular, it is not helpful to use 24-hour averages of wind speed and temperature inversions because of the significant variability in these parameters on days when PM_{10} concentrations are elevated. High pollution days in Christchurch are characterised by strong negative temperature gradients, cold temperatures and low wind speeds during the evening. During the daytime, however, the temperatures are warmer than average, temperature gradients more strongly positive than average, and wind speeds can be higher. Thus the extremes in meteorology observed during the evening are not well reflected in a 24-hour average. Refining the meteorological variables used in the analysis of trends in 24-hour average PM_{10} could significantly improve this analysis.

Decrease in the proportion of "marginal days" with PM_{10} concentrations that just exceed the NES, provides further evidence that PM_{10} emissions were decreasing over the period in consideration. This decrease in the proportion of days that exceed the NES is only apparent within a broader range of meteorological data classifications (rather than worst case). This is because the PM_{10} concentrations within the worst case meteorological classification are typically much higher than 50 µg m^{-3} and even relatively large reductions in PM_{10} concentrations are unlikely to take them below the NES. The proportion of NES exceedences within the high pollution group of data is therefore only likely to change once significant emissions reductions are achieved.

The most relevant trend to monitor, in terms of meeting the NES, is that associated with changes in 24-hour average PM_{10} concentrations under worst case meteorological conditions. In essence these trends can be observed by considering the upper quartile data in Figure 3.3. As indicated above, tracking this group of data relative to the number or proportion of exceedences does not illustrate reductions in concentrations because of the absolute difference between existing peak concentrations and 50 µg m^{-3}. It is possible, however, that this will be a useful tool for illustrating trends in exceedences at some stage in the future, when a greater reduction in concentrations has been achieved.

Summary of Key Findings: Regression Tree Analysis

This analysis has used all of the hourly measurements of air pollution and weather from the Coles Place monitoring site over the 8 years from 1999 to 2006 (70,000 hours). The air pollution measurements are then all examined in groups to see which ones occurred under similar weather. In this way, a 'tree' of relationships or 'predictor variables' is built up. These trees may have many variables in them, and each can be assigned a level of importance. In other words, the method identifies which weather variables are most important for explaining the amount of air pollution. Once these relationships have been discovered from the regression trees, they can be used to say something about the underlying trends in emissions.

The highest concentrations tended to occur during winter (May, June, July and August) evenings, when the air was still and cold. A group of high air pollution measurements (about 2000 hours) that occurred under very similar conditions were identified. These high pollution measurements were then clustered according to the year within which they were measured. Finally, a year-to-year comparison was undertaken to see if there had been any changes over time.

This study suggests that the highest air pollution measurements were recorded in 2000. Between 2000 and 2004 there appears to have been a significant decrease in air pollution concentrations. However, the concentrations measured in 2005 and 2006 are very similar to those in 2004. The decrease in concentrations seems to have slowed or stopped in recent years. The analysis suggests that emissions in 2000 and 2001 were the highest and they then decreased over 2002, 2003 and 2004, but appear to have levelled off in 2005 and 2006. There is an overall decrease through the 8 year period though.

4. Simple correlation analysis

4.1. Introduction

There is an obvious and direct relationship between the weather and the occurrence of air pollution. This is strong in Christchurch, but occurs to some extent throughout the country. The most direct relationships are with winds (*….when it's windy air pollution emissions get blown away…*), and with temperature (*…when it's cold people burn more in home heating appliances, emit more, and air pollution emissions increase….*). However, there are also more subtle relationships, such as the occurrence of inversions – which often coincide with cold temperatures and calm periods. This part of the research examines the relationship between air pollution (as PM_{10}) and weather variables using simple correlation methods. It attempts to establish the relationship between the key weather variables and occurrences of high pollution levels, in order to understand the year-to-year variability in pollution that is seen in the monitoring.

This section is a summary, with the highlights of the results and more detailed analysis given in Appendix B.

4.2. Method

The methodology is straightforward and simple. The basic PM_{10} and weather data series from the central Christchurch monitoring site for the period 1999 to 2006 has been used. This has been cleaned, corrected and analysed on an hour by hour basis for the whole period (as described in Appendix B). Since high levels of air pollution in Christchurch generally only occur in winter, only the winter data have been used (May, June, July and August). The analysis has also been conducted just using the core winter months of June and July.

Simple correlations were made between the ambient pollution data of interest (peak concentrations of PM_{10} and frequency of exceedences of the PM_{10} standard) and the weather variables (average daily temperatures and occurrences of calm wind periods). These were then analysed for each year in the 8-year study period to assess relationships and trends.

The relationship between the weather variables and pollution was then analysed to assess how this had been changing over the study period, and see if any inferences could be made about trends in Christchurch's air pollution.

4.3. Results

Table 4.1 shows a summary of the basic cross-correlations. It shows a high level of correlation between core winter average near-surface air temperature and average PM_{10} concentrations (-0.79), and between the average hours of calms and concentration (-0.75). There is a weaker correlation with 10 m air temperatures (-0.63), and with average 24-hour wind speed (-0.54). There is little correlation with the low level temperature difference (-0.26), and none with humidity (0.07).

Table 4.1: Correlation of average 24-hour PM_{10} with various 24-hour weather conditions for the months of June and July 1999-2006.

June and July of:	Average PM_{10} (Coles Place) ($\mu g\ m^{-3}$)	Average temp. at 1m (ºC)	Average temp. at 10m (ºC)	Average temp. difference (1m-10m) (ºC)	Wind speed (ms^{-1})	Average hours of calms (/day)	Average percent of calm winds (%)	Average RH (%)
1999	58.0	6.7	7.9	-1.2	2.2	13.0	54.1%	85.1
2000	38.3	8.4	9.2	-0.8	2.3	11.5	48.1%	74.1
2001	64.8	5.6	6.2	-0.6	1.8	15.5	64.4%	72.0
2002	48.7	6.4	6.5	-0.1	1.8	14.0	58.4%	79.1
2003	49.9	6.8	7.4	-0.6	1.9	14.0	58.5%	77.0
2004	51.0	6.9	7.6	-0.7	2.0	14.0	58.5%	75.7
2005	46.2	6.9	7.0	-0.1	2.0	13.8	57.4%	78.1
2006	50.8	5.7	na	na	1.9	na	na	na
Correlation		**-0.79**	**-0.63**	**-0.26**	**-0.54**	**0.75**	**0.75**	**0.07**

The correlation between PM_{10} and temperatures and calms is not independent. Cold temperatures and calm winds can themselves be correlated, although not necessarily completely equivalent in producing a high PM_{10} values. For instance, there could be cold windy days with low PM_{10}, or relatively warm calm days with high PM_{10}. The correlation with vertical temperature difference (being a crude indicator of inversions) is not at all strong (-0.26). Temperature inversions are transient in nature, and vary in depth, which has a significant effect on the ambient PM_{10} concentration on an hour by hour basis. These transient events are not able to be resolved and accounted for when using 24-hour averages, as has been done here.

Figure 4.1 shows all of the daily PM_{10} concentrations plotted against average temperature and frequency of calms. This shows that high pollution (here indicated by days where the standard was exceeded) generally occur only when the temperature is low, and the percentage of calms is high. There were no exceedences when the average temperature was above 14°C and very few when there was less than 10-12

hours of calms per day (here a 'calm' is defined as an average hourly wind speed of less than 2.0 m s^{-1}).

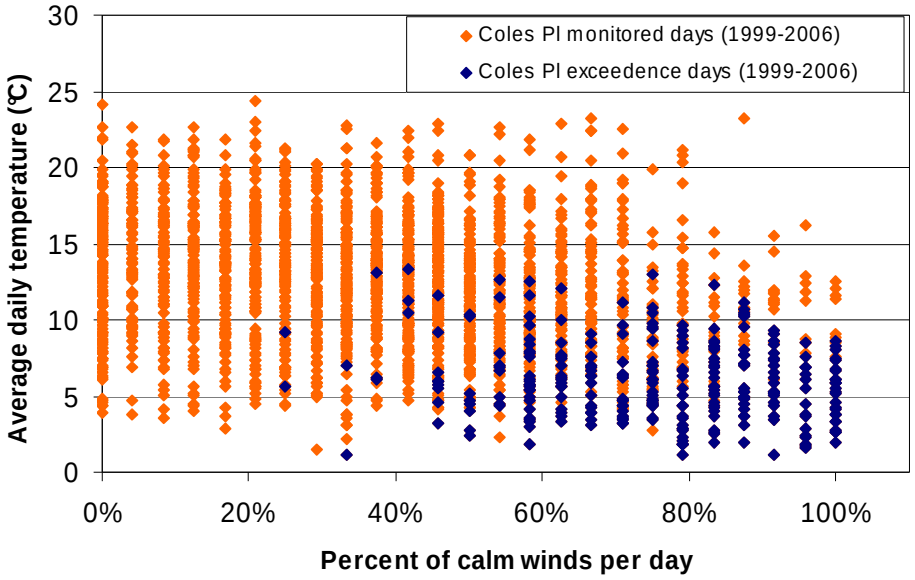


Figure 4.1: Comparison of monitored days and days which exceeded the NES for Coles Pl (1999-2006) using average daily temperature and percentage of calms.

The number of 'cold' days (with temperature < 5 ºC) varies significantly from a low of 13 in 2000 to a high of 42 in 2001 (Figure 4.2). The average PM_{10} concentration for this subset of days also differs from year-to-year, with some hint of a downward trend.

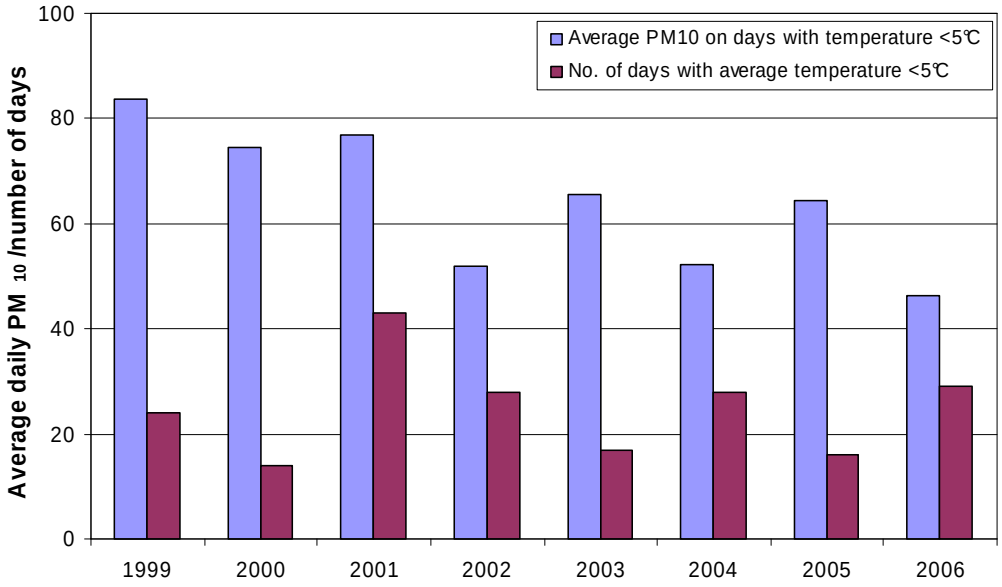


Figure 4.2: Average PM_{10} concentrations on days with an average temperature below 5ºC, for all months 1999-2006.

The pattern is similar using data from 'cool' days, as shown in Figure 4.3. Here instead of using days with < 5 ºC, the days of 6-9 ºC were arbitrarily selected. There is a similar variability in the number of days (44 to 75), but less in the average PM10 concentration, again with some indication of a slight downward trend. This result suggests that there is some defined relationship between the weather and air pollution that can be adjusted for temperature variation so that emissions trends can be assessed.

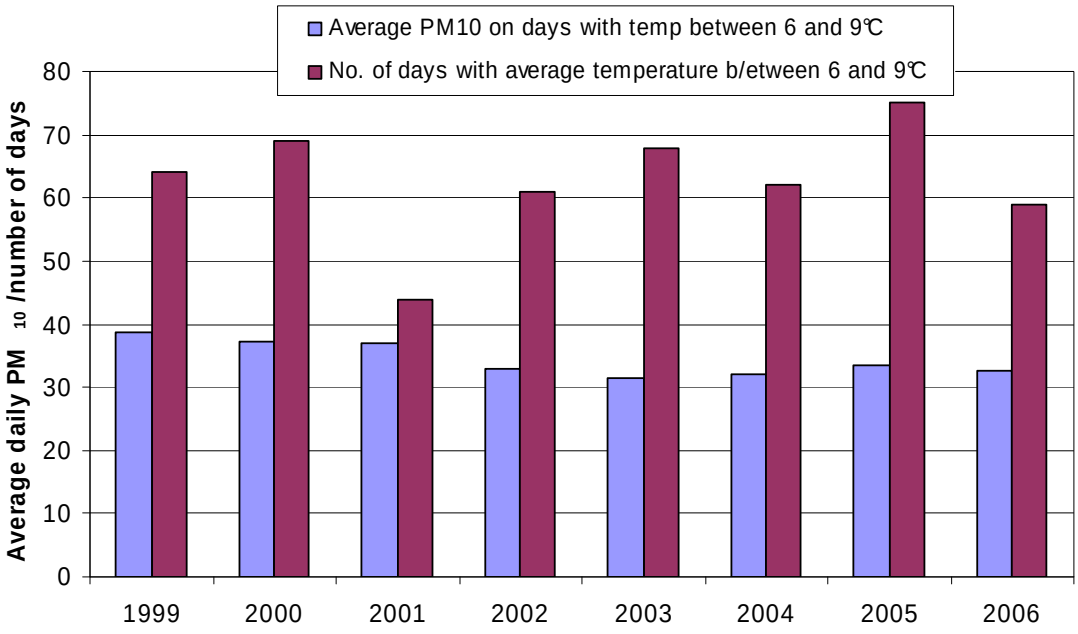


Figure 4.3: Average PM_{10} concentrations on days with an average temperature between 6 and 9 ºC, for all months 1999-2006.

4.3.1. Correlation model

For the winters of each of the 8 years covered, the daily average temperature and the PM_{10} concentrations were plotted together, as shown in Figure 4.4. A logarithmic curve was fitted to the relationship. Despite the large scatter, a trend in the form of this relationship emerges. That is, for a given temperature, the amount of pollution associated with that temperature appears to be less in more recent years.

This analysis was also conducted for the average temperature over the June/July period, and for the frequency of calms. The results are similar, but not presented here (see Appendix B, Figure B12).

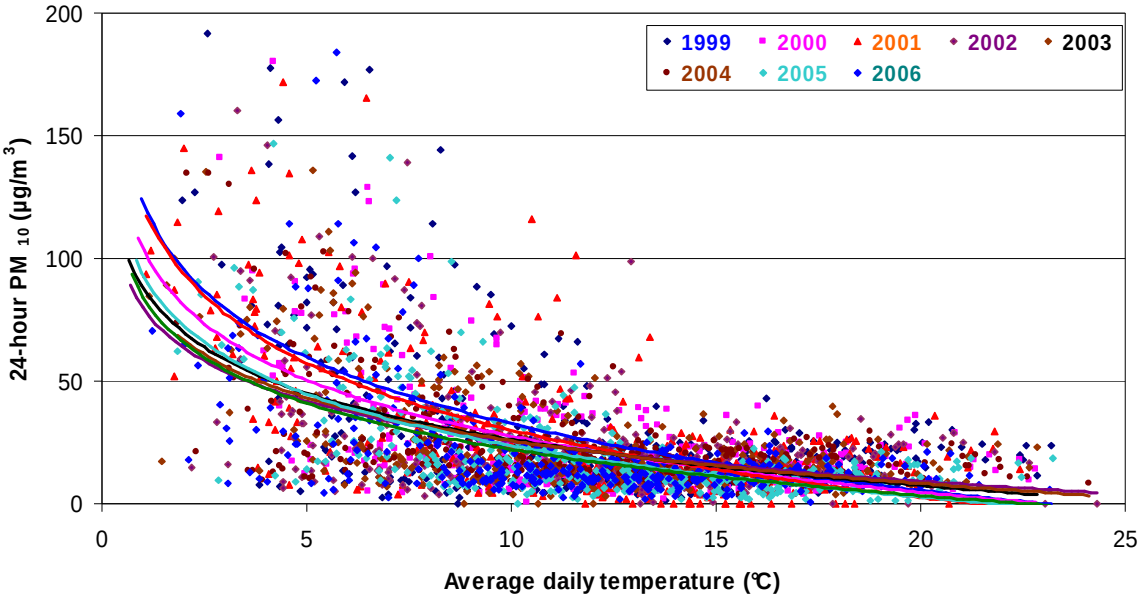


Figure 4.4: Yearly trends in PM_{10} with respect to average daily temperature (1999-2006).

4.4. Discussion

This analysis shows some level of correlation between daily weather averages and PM_{10}, but it is by no means strong. As noted earlier, the relationship is likely to be influenced by other weather factors that have not been analysed – such as inversions. Good temperature inversion data are frequently not available for most cities and towns of New Zealand. Other factors that may influence the relationship are secondary factors (such as variations in the price or availability of fuel) and more subtle weather features (such the length of cold periods, where longer periods of cold weather may cause people to burn more than would be expected by the temperatures alone).

Given the results shown, it is tempting to conclude that they provide evidence of a systemic reduction in PM_{10} emissions in the Christchurch airshed. The data may well indicate this, but the results need to be interpreted with caution, since (a) there is substantial variability and the statistical significance of the relationship has not been determined, (b) the weather relationships are crude, omitting important factors such as inversion extent and strength, (c) some secondary drivers have not been accounted for, such as fuel prices or particularly cold, or warm, periods.

Summary of Key Findings: Simple Correlation Analysis

This analysis shows that much of the year-to-year variability in the air pollution occurring in Christchurch in winter-time can be explained by simple weather variables. The strongest relationship occurs with temperature – the winters with colder average temperatures experience higher air pollution. This fairly obvious result occurs consistently and for a range of temperatures. Higher air pollution also occurs in winters that have more periods of light winds or calms.

The year-to-year variation in these weather parameters can mask any long term trends in emissions. When variations in the temperature and calms are partially accounted for, there are indications that the emissions in Christchurch have decreased over the period 1999 to 2006, resulting in lower pollution concentrations for a given weather pattern. However, this result is not strong, and the trend occurs over several years with any particular year capable of having higher or lower pollution levels than the previous years, due to the high year-to-year variability in the weather.

5. Complex regression analysis

5.1. Introduction

The methodology used in the complex regression analysis is to some extent based on the method outlined by Wise and Comrie (2005). The mentioned paper investigated meteorologically adjusted trends of PM_{10}, ozone and a number of meteorological variables in the southwest of the United States of America. The Kolmogorov-Zurbenko filter (KZ filter) used in their paper was first introduced to investigations of ambient air quality by Rao and Zurbenko (1994) to effectively separate different frequencies within a time series. Numerous studies have subsequently confirmed its usefulness in achieving various objectives within air quality research (e.g. Hogrefe et. al. 2003; Ibarra-Berastegi et. al. 2001; Porter et. al. 2001; Yang and Miller 2002; Anh, Duc and Azzi 1997; Eskridge et al. 1997).

In this study, however, the above mentioned method (Wise and Comrie 2005) needed to be modified to some extent. The application of multiple linear regression and subsequent residual analysis using the KZ filter assumes constant emissions, which, in the case of PM_{10} emissions in Christchurch is not appropriate. The preliminary approximation of constant emissions is described below.

5.2. Method

A PM_{10} time series (FDMS equivalent; see Appendix A for details on data preparation) from Coles Place, St Albans (Appendix C, Section 3) was analysed in order to evaluate meteorological influences over the period 1999–2006, inclusive. PM_{10} concentrations were averaged on a daily basis for the hours between 5 pm and 12 am as it was found to be the main peak time for concentrations (see Figure 5.1).

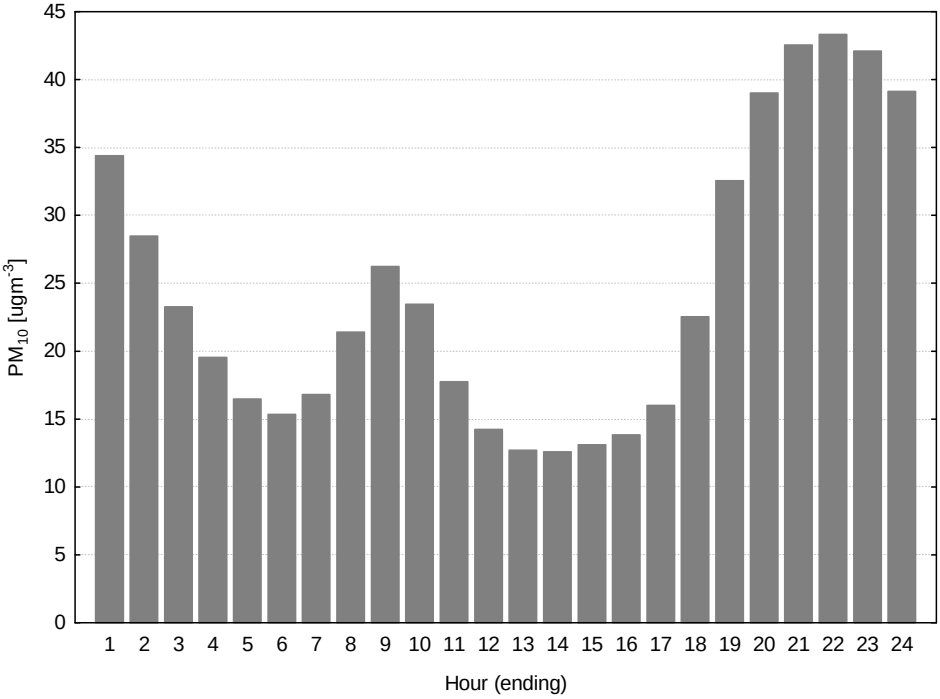


Figure 5.1: Mean diurnal PM_{10} concentrations for Coles Place, St Albans (1999–2006). Values shown are hourly averages ending on the hour given.

Wind speed, 1m air temperature and the vertical temperature difference between 1m and 10m, all obtained from the same site as the PM_{10} concentrations, were chosen to be representative of meteorological conditions. Wind speed and temperature difference represent the intensity of horizontal and vertical mixing within the atmospheric boundary layer, respectively. These meteorological variables were averaged over the same period as PM_{10}.

The identified diurnal maximum of concentrations in the evening hours reflects the fact that by far the largest contribution to PM_{10} concentrations measured in Christchurch originates from home heating devices such as log burners. This makes emissions very variable as they are highly dependent on temperature behaviour and can therefore not be considered as constant over time (in comparison to emissions from traffic which show only little variation over the course of a year). This is underlined by the fact that PM_{10} exceedences only occur in winter. In other words, air temperature is a key cause of PM_{10} emissions in Christchurch, whereas wind speed and temperature difference modify concentrations.

Following the methodology described by Wise and Comrie (2005), multiple linear regression analysis was performed to identify (and subsequently remove) meteorological influences from the dataset. This approach, however, assumes that emissions are approximately constant over time and that resulting concentrations are

being modified only by weather conditions. As outlined above, PM_{10} concentrations in Christchurch cannot be considered constant due to the influence of temperature. Therefore, this dependency had to be removed prior to the regression.

According to Ott (1990) pollutant concentrations generally show log normality and this was evident in this case. Therefore, PM_{10} concentrations were transformed using the natural logarithm. Afterwards, each year was split into a winter and a summer season (April–September and October–March, respectively). For each season of each year, evening PM_{10} concentrations were regressed against temperature and the calculated dependency was corrected using the following formula, retaining the maximum variation by recalculating the observed residuals to represent deviations from a zero trend line (i.e. the overall mean of PM_{10}):

$$corr.\ log_e\ PM_{10} = (((log_e\ PM - (a + b \cdot TEMP)) / 2) \cdot \sqrt{3})) + avg.\ PM$$

with *corr.* log_e *PM* = natural logarithm of temperature-corrected PM_{10} concentration, log_e *PM* = natural logarithm of raw PM_{10} concentrations, *TEMP* = 1m air temperature, *avg. PM* = mean raw evening PM_{10} concentration for each season, *a* = intercept of the calculated regression and *b* = slope of the calculated regression. After this, the dataset was rejoined into a continuous time series.

Figure 5.2 shows the inverse relationship between log_e *PM* and *TEMP* for the whole time series before the correction was applied. The shape of the cluster of points does not indicate a linear dependency between the two variables. It was found that this is due to the fact that the observed negative correlation was significantly higher during the winter seasons (i.e. at lower temperatures), while some years even showed a positive relationship in the summer seasons (Appendix C, Section 1).

This also led to a residual weak negative correlation after the corrected seasons were rejoined to create a complete time series (Appendix C, Section 2), which was removed in the same way as outlined above for individual seasons and resulted in a completely temperature independent time series of PM_{10} concentrations (Figure 5.3).

This temperature corrected data set was then taken as the input time series for the subsequent multiple linear regression analysis to identify, quantify and subsequently remove meteorological influences on evening PM_{10} concentrations in Christchurch (Appendix C, Section 4).

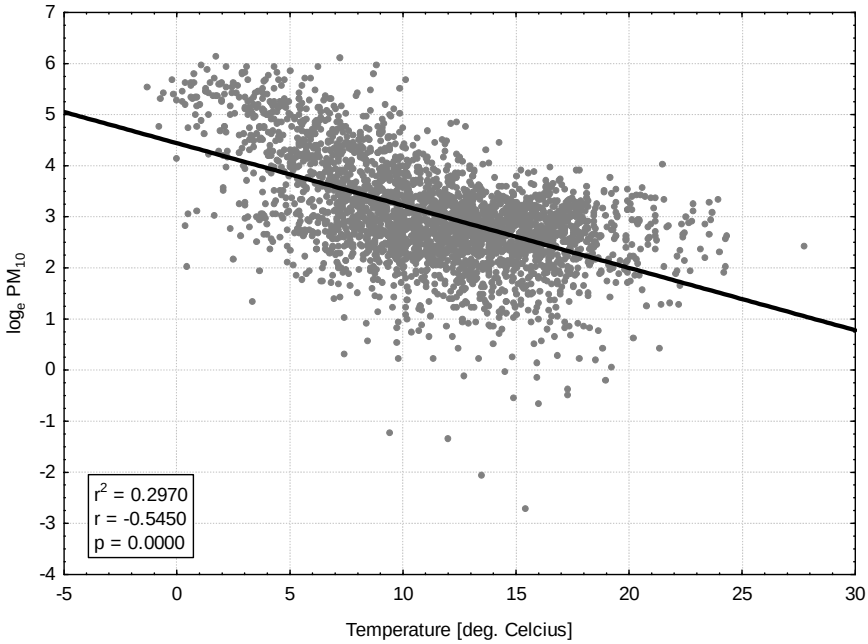


Figure 5.2: The relationship between 1 m air temperature and the natural logarithm of PM_{10} recorded between 5 pm and 12 am at Coles Place, St. Albans.

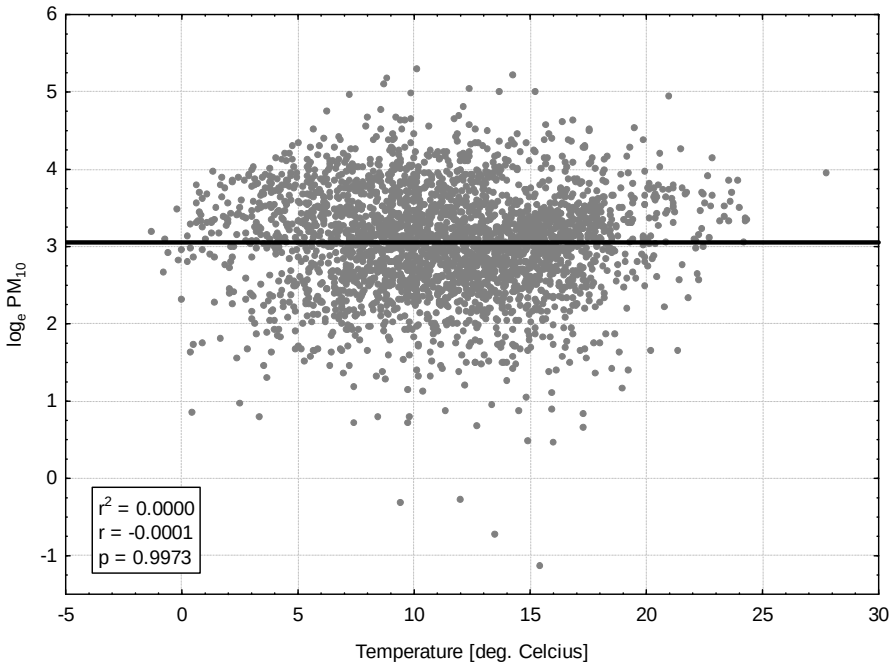


Figure 5.3: PM_{10} concentrations recorded between 5 pm and 12 am at Coles Place, St. Albans showing no dependency on temperature.

To remove the remaining meteorological influences of horizontal and vertical mixing conditions, multiple linear regression analysis was performed. The corrected PM_{10} concentrations were selected as the dependent variable and regressed against the

independent variables of wind speed (which was transformed using the square root to approximate a normal distribution) and temperature difference. The regression was able to explain about 20% of the remaining variance in the PM_{10} concentrations (r^2 = 0.193).

Residual analysis was performed to investigate the variation in concentrations which remained unexplained by the above mentioned meteorological influences. The residuals can be considered as revealing variations within PM_{10} concentrations due to factors other than meteorology and thus represent a better approximation of the behaviour of emissions.

To make the residuals comparable to concentrations they were added to the overall mean of the raw (but temperature corrected) measurements. This step is necessary as residuals, by definition, only represent deviations from a calculated series of values – in this case the calculated series of optimal predictions of PM_{10} concentrations based on variations in wind speed and temperature difference – and therefore fluctuate around a zero line (i.e. their sum equals zero).

This new dataset can now be understood as adjusted PM_{10} concentrations, where meteorological influences, namely 1 m air temperature, wind speed and presence and strength of a temperature inversion have been removed (Appendix C, Section 5).

5.3. Results

A simple moving average filter (based on the Kolmogorov-Zurbenko or KZ filter) was applied to both the raw and the adjusted PM_{10} data set. A window size of 365 days (evenings) was chosen to average out seasonal fluctuations and two repeated iterations were run to facilitate interpretation of the trend. The result of the first iteration is referred to as $KZ_{365,1}$ and the second as $KZ_{365,2}$. Each iteration truncates half the size of the chosen window length at each end of the time series. So, the first filter run effectively cuts one year off the time series and the second run two years.

The comparison of the resulting trends is shown in Figure 5.4. The dotted lines show smoothed hourly PM_{10} concentrations where seasonal fluctuations have been removed ($KZ_{365,1}$). The solid lines are further smoothed (removing fluctuations smaller than 1.4 years) to aid interpretation of the long term behaviour of PM_{10} concentrations ($KZ_{365,2}$).

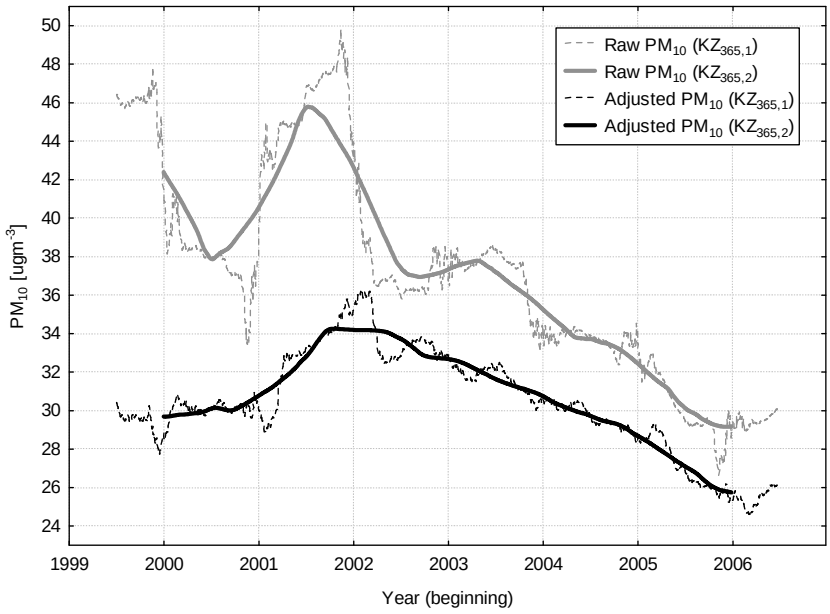


Figure 5.4: Comparison of the trend of raw (grey) and adjusted (black) hourly PM_{10} concentrations recorded between 5 pm and 12 am at Coles Place, St Albans. The dates indicate the beginning of the year.

The raw trend for the evening hours follows the trend observed in daily averages (for comparison refer to Aberkane, Harvey and Webb 2005) reflecting general winter conditions (primarily influenced by variations in mean temperatures throughout the winter months). This becomes particularly evident when comparing average concentrations for 2000 and 2001, reflecting mild and cold winter conditions respectively.

The adjusted trend suggests an increase in PM_{10} emissions with a peak in late 2001 to early 2002. Afterwards, emissions appear to decrease steadily until autumn 2006. Thereafter, $KZ_{365,1}$ indicates a slight increase towards winter 2006, but fluctuations are apparent throughout the entire trend line and no solid conclusions should be drawn from this.

Figure 5.5 provides a summary of the adjusted time series plotted as a box plot showing mean concentrations along with their standard deviation for each year. Similar to Figure 5.4, a decrease in emissions is observed after a peak in 2001. It can be assumed that 2001 was a rather exceptional year as indicated by the wide span of the standard deviation. Since 2002, variability within each year has decreased and somewhat stabilised in recent years (with 2006 again indicating a slight increase). A similar trend is shown Appendix C, Section 6 (showing the median and quartiles for each year), where the span of the upper quartile with respect to the median is somewhat wider than the lower quartile in earlier years compared to recent years. This

is especially true for the winter months May to August as shown in Appendix C (Section 7). However, the trends in medians for both the overall yearly data, as well as the winter subsets, are rather different from the trends depicted by the mean values (Figure 5.5 and Appendix C, Section 8). For the median values, no clear trend can be established for the winter months. The whole year data show a delay in the peak in emissions, which does not occur until 2003. The trends in the mean values for both winter and whole year data (Figure 5.5 and Appendix C, Section 8) show similar patterns and furthermore, follow the trend described by the moving average filter (Figure 5.4).

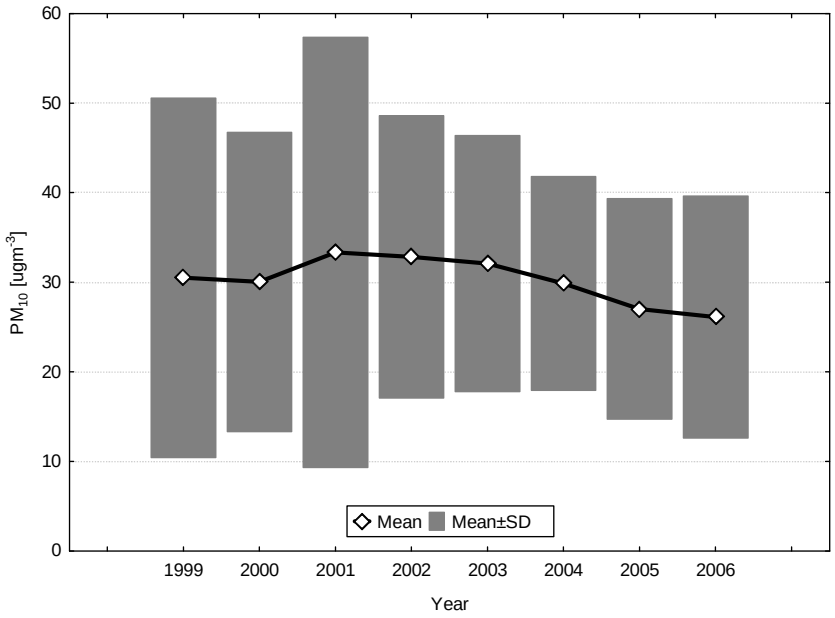


Figure 5.5: Box plot of the final adjusted annual times series of PM_{10} concentrations showing the mean and standard deviation for each year.

Nevertheless, the trend lines presented only cover a period of 6 to 7 years ($KZ_{365,2}$ and $KZ_{365,1}$ respectively) and it needs to be emphasised that a robust interpretation of long term behaviour is inappropriate considering this short time span. No clear evidence can be found to explain the behaviour of emissions as described by the adjusted trend, but it may reflect the effectiveness of Environment Canterbury's air quality management strategy.

5.4. Discussion

The analysis described above revealed a trend for PM_{10} emissions which peaks in 2001/2002 and shows a decrease since then. This, however, does not agree entirely with Environment Canterbury's emissions inventories, which identify a peak in emissions in 1999 and a decrease in 2002. Furthermore, as described earlier, the

statistical approach used in this study assumes constant emissions which are subsequently modified by meteorological conditions. The creation of a time series that is independent of temperature behaviour (which can be interpreted as a rough proxy for seasonality) certainly supports the approximation of emissions being constant in time. However, it is a matter of fact that the use of home heating devices is limited to the cold half of the year in Christchurch and therefore, emissions will always be subject to seasonal variation. As a result of this, any analysis of the complete time series will always be limited in its accuracy. The utilisation of a running mean filter with a window that is larger than the seasonal fluctuations certainly reduces the impact of this uncertainty. Nevertheless, a level of uncertainty about the accuracy of the results due to the methodology remains. A quantification of the expected bias seems impossible at this stage, although further modification of the analysis might provide means to overcome this problem. However, for the time being, the results presented can be considered as a good approximation of emission behaviour in the vicinity of Coles Place derived from a time series of average evening PM_{10} concentrations.

Summary of Key Findings: Complex Regression Analysis

A meteorologically adjusted time series of evening PM_{10} concentrations was created which is independent of temperature, as well as of vertical and horizontal mixing. This adjusted data set can be seen as a representation of PM_{10} emissions as the above mentioned meteorological influences have been removed. The adjusted trend suggests an increase in emissions with a peak in late 2001 to early 2002, while they appear to decrease steadily thereafter.

6. Discussions

The PM_{10} concentrations measured in central Christchurch over the 8 years from 1999-2006 have been analysed using three different and independent techniques. Each has shown that differences in the weather from year-to-year can explain a significant amount of the variability in measured PM_{10}. The strongest relationship is with air temperature (colder weather means higher concentrations), but also with wind speed (calmer winds means higher concentrations). Other factors such as inversion strength also influence the concentrations.

The data and resources required to undertake each of three analytical techniques used in this study are compared in Table 6.1.

Despite the different techniques used, each of the analyses indicates that there has been a reduction in PM_{10} emissions over the period, particularly since 2001. Although not a statistically strong result, the studies also indicate that this reduction has been of the order of 3-4% per year over 1999 to 2006. Although these results cover a different

period than that considered (1996 to 2003) in the Canesis investigation (Environment Canterbury, 2004), the results are broadly consistent.

Table 6.1: Comparison of the three analytical techniques used in this study

Analysis Technique	Input data required	Resources needed to undertake Analysis
Regression Tree	1-hour average PM_{10}, hour of day and meteorological data 24-hour average PM_{10} and meteorological data	Knowledge of advanced statistical methods and access to advanced statistical software.
Simple Correlation	24-hour average PM_{10} and meteorological data	Knowledge of basic statistical methods and access to basic statistical software.
Complex Regression	1-hour average PM_{10} and meteorological data for 18:00 to 0:00 each day	Knowledge of advanced statistical methods and access to advanced statistical software.

7. Conclusion

7.1. Achievement of the aim and objective of the study

The aim of this project was to provide regional councils with tools which will enable the use of air quality monitoring data to assess:

- Compliance with the region's SLiP
- The effectiveness of the region's air quality management strategy

The objective of this study is to develop and compare methods for assessing trends in PM_{10} emissions using monitored PM_{10} concentrations and meteorological data. The trends teased out from the monitoring data can be compared to the region's SLiP and the effectiveness of the region's air quality management strategy assessed.

Each of these three analyses used the same input data and produced consistent results. Although not analysed in detail, the studies all indicate a reduction in the peak over the 8 year period of around 3-4% per year. Each analysis showed (a) a very clear quantification of the influence of the weather and climate – with the main determinant being air temperature, and (b) a long term average decrease in peak concentrations.

The research, whilst indicating some promising outcomes, must be interpreted cautiously. The data period is not very long and contains a significant degree of

variability. A full analysis of the statistical significance of the all results has not been carried out, and what was completed for the regression tree analysis showed that the results frequently did not meet stringent statistical significance criteria.

There is also the issue of the unexplained variability in the results, which indicates that there are other factors influencing the concentrations that have not been taken account of here. These include the influence of temperature inversions, more complex wind flows around the city, and just how often people use their solid fuel burners because of non-weather factors, such as fuel prices, special events, unusual cold snaps, etc.

Given the positive outcomes it is concluded that both the aim and objective of the study have been met.

7.2. Potential enhancements

The attempts to achieve the aims of this study have highlighted a number of gaps in the available data. These are in three broad areas – air pollution data, weather and climate data, and data on social and economic drivers.

7.2.1. Air pollution data

The PM_{10} monitoring record from Christchurch is relatively good, being probably the best in New Zealand – which is one of the main reasons for selecting Christchurch as the study area. It is of good quality, with few gaps, and well verified. Analyses of this nature always desire longer periods, but this cannot be regarded as a shortfall.

A potential drawback is the areal representativeness of the PM_{10} data. The Coles Place site in St Albans is located just north of the city centre and reasonably representative of central Christchurch, but there may be variations across the city that need to be better understood. For instance, although the weather over Christchurch is reasonably uniform (compared to most other large urban centres in New Zealand), there are features that are known to affect air pollution, such as drainage flows off the Port Hills and Canterbury Plains, and coastal sea breeze effects. These are not fully reflected in the PM_{10} data for Coles Place, and may be important factors in understanding the relationships between weather and air pollution, and their spatial variation. An equivalent record in terms of quality and length is needed for 2-3 other sites in Christchurch.

A second issue is that this research has focused mostly on evaluating trends in PM_{10} concentration, rather than the frequency of PM_{10} exceedences. It is the latter that are

the prime focus of the National Environmental Standards (NES). It is therefore not yet possible to provide definitive guidance to regional councils regarding their ability to meet the requirements of the NES by 2013. This will be the focus of continued work which will investigate the nature of the frequency distributions of the air quality data and their significance for prediction of future numbers of NES exceedences.

7.2.2. Weather and climate data

Again, the basic meteorological data are good and generally representative of the weather and climate of Christchurch. However, throughout the research, one major limiting factor has been identified – the lack of suitable data on inversions.

The occurrence, height and strength of inversions have a very strong influence on air pollution concentrations because of their effect on vertical mixing. For instance, in a simplistic way, if X kilograms of PM_{10} emissions is mixed into an inversion capped layer 50m deep, with calm winds, then it will produce a concentration of Y μg m^{-3}. However, should this mixed layer be only 25m deep (and with all other factors remaining constant), then the resulting concentration will be 2 times Y μg m^{-3}. The real atmosphere is not this simple, but the point is made that inversion characteristics can have a dramatic effect on pollution concentrations.

Data on inversions are not available for the simple reason that it is very hard to obtain them. The basic methodology is to have a tall tower that is fitted with temperature sensors every few metres. Such facilities are difficult to arrange and expensive to maintain. Some spot data can be obtained using tethered or free rising balloons, as has been done many times in Christchurch (McKendry et al. 2004, Corsmeier et al. 2006). However, whilst these methods give excellent data on specific events, they cannot be sustained through every night during winter, for several years.

Some benefit may be gained by incorporating the effect of a time lag between meteorological variables (such as temperature) and pollution: i.e. the temperature of the preceding hour may be a better predictor of air pollution than the temperature for the same hour in which the pollution was monitored.

Remote sensing devices (such as acoustic sounders and lidars) have good data capabilities, but these have problems of their own, often in terms of cost to maintain them over the required periods.

Some attempt has been made to use the differential temperature measurements available, based on sensors at 1m and 10m above the ground. This has not proved to

be a very strong indicator, and does not correlate well with 24-hour average pollution concentrations (which is not surprising). More sophisticated methods might show a better relationship, but these have been beyond the scope of this study.

In addition, there may be other weather factors that have not been studied here that influence pollution concentrations, such as more detailed airflow and temperature features around the city.

7.2.3. Social and economic data

Finally, it is obvious that there will be variable emissions over Christchurch during any particular period because of social and economic factors. These might include particular types of weather – such as long very cold snaps – that cause people to burn more than just the temperature alone might indicate. They might also include factors such as the price or availability of fuel – particularly electricity – that mean people use more wood. This occurred dramatically in 1992 during major power outages, resulting in the highest ever air pollution year. There might also be other subtle factors that have been shown in some overseas studies to affect the amount of air pollution independently of the weather – such as special public events, school holidays, flu epidemics, wild fires, and so on.

7.3. Ongoing and future research

When undertaking trend analyses with data sets which contain a relatively low number of years (say less than 10), there is a risk that any long term trend contained in the data series will be obscured by single event aberration and/or short term changes. In trend analyses, data sets covering longer periods of time are likely to provide results that are more robust and contain greater certainty. In this project, eight years of data (1999 to 2006) were analysed. The results presented in this report would benefit from being revisited biannually when an additional two years of data become available. This repeated review process would add certainty to the direction and magnitude of the trends in PM_{10} concentration illustrated in this report.

As mentioned in Section 7.2.1, it should also be noted that the main focus of research so far has been on controlling for the influence of weather variables on average concentrations. However, the National Environmental Standards refer to exceedences of specific values over certain time periods. It is the number of these exceedences that must be reduced by 2013 – this may, or may not, be related to a reduction in long-term average concentrations. The rationale for initially focusing on average values is that it allows an assessment of general trends of air pollution concentrations using more

robust statistical techniques. Future work on analysis of the frequency distribution of hourly and 24-hour average concentrations, and non-parametric analysis of exceedences will strengthen the results presented here.

It would be of interest and value to expand this study to locations beyond Christchurch. The methods used in this study work best in locations that have simple source profiles (one dominant polluter), non-complex terrain and uncomplicated meteorology. However, the single most important factor which will determine the success of an investigation such as this, is the availability of a long term PM_{10} data set. Given these criteria, target locations for expanding the study could include Timaru and Nelson. Informative results may also be obtained from Wellington and the Hutt Valley. Preliminary investigations into Auckland, which has a relatively complex source profile and varied terrain, suggest that undertaking a trend analysis of PM_{10} concentrations in this region would be more of a challenge.

8. Acknowledgements

This work was conducted under the Foundation for Research, Science and Technology programme "Protecting New Zealand's Clean Air", contract CO1X0405.

Cover Photograph. Mike Freeman, Environment Canterbury.

The data used have been freely supplied by Environment Canterbury, and in particular Teresa Aberkane of Environment Canterbury has greatly assisted the study with her support, co-operation and enthusiasm.

This report has benefited from reviews and feedback by:

- Teresa Aberkane - Environment Canterbury
- Vicky Lucas - Environment Canterbury
- Tim Mallett - Environment Canterbury
- Clive Marsh -Statistician
- Louise Wickham - Ministry for the Environment

The authors are also greatly indebted to Professor Peter Brimblecombe of the University of East Anglia, UK for his review of the report.

9. References

Aberkane, T.; Harvey, M.; Webb, M. (2005). Annual Ambient Air Quality Monitoring Report 2004. Christchurch, Environment Canterbury (ECan)**:** 58p.

Anh, V.; Duc, H.; Azzi, M.(1997). Modeling anthropogenic trends in air quality data. *Journal of the Air & Waste Management Association 47*: 66-71.

Corsmeier, U.; Kossmann, M.; Kalthoff, N; and Sturman, A. (2006). Temporal evolution of winter smog within a nocturnal boundary layer at Christchurch, New Zealand. *Meteorology and Atmospheric Physics 91*:, 129-148.

De'ath, G.A.; Fabricius B.K. (2000). Classification and regression trees: a powerful yet simple technique for ecological data analysis. *Ecology 81*: 3178-3192.

Environment Canterbury (ECan). (2004). An alternative measure of air quality. Canesis Network Ltd Report to Environment Canterbury Report U04/76.

Fisher, G.W.; Kuschel, G.; Mahon, K. (2006). Straight line and curved line paths (SLiPs and CLiPs), developing the targets and predicting the compliance. http://www.niwascience.co.nz/ncces/air_quality/chc2006_5.pdf

Eskridge, R.E.; Ku, J.Y.; Rao, S.T.; Porter, P.S.; Zurbenko, I.G. (1997). Separating different scales of motion in time series of meteorological variables. *Bulletin of the American Meteorological Society 78*: 1473-1483.

Hogrefe, C.; Vempatyb, S.; Rao, S.T.; Porter, P.S. (2003). A comparison of four techniques for separating different time scales in atmospheric variables. *Atmospheric Environment 37*: 313-325.

Ibarra-Berastegi, G.; Madariaga, I.; Elías, A; Agirre, E.; Uria, J. (2001) Long-term changes of ozone and traffic in Bilbao. *Atmospheric Environment 35*: 5581-5592.

McKendry, I.G.; Sturman, A.P.; Vergeiner, J. (2004). Vertical profiles of particulate matter size distributions during winter domestic burning in Christchurch, New Zealand. *Atmospheric Environment 38*: 4805-4813.

Ott, W.R. (1990). A Physical Explanation of the Lognormality of Pollutant Concentrations. *Journal of the Air & Waste Management Association 40*: 1378-1383.

Porter, P.S.; Rao, S.T.; Zurbenko, I.G.; Dunker, A.M.; Wolff, G.T. (2001). Ozone Air Quality over North America: Part II—An Analysis of Trend Detection and Attribution Techniques. *Journal of the Air & Waste Management Association 51***:** 283-306.

Wise, E.K.; Comrie, A.C. (2005). Extending the Kolmogorov-Zurbenko Filter: Application to Ozone, Particulate Matter, and Meteorological Trends. *Journal of the Air & Waste Management Association 55*: 1208-1216.

Yang, J.; Miller, D.R. (2002). Trends and variability of ground-level O3 in Connecticut over the period 1981-1997. *Journal of the Air & Waste Management Association 52*: 1354-1361.

Appendix A: Detailed report on the regression tree analysis

Al. Calibration of hourly TEOM(40) PM_{10} to TEOM-FDMS equivalent

The original correlation between the unadjusted hourly TEOM(40) data and the TEOM-FDMS data is high (R^2 = 0.89), but there is a tendency for TEOM(40) to under-measure by about 18% at higher concentrations. To correct for this, a regression tree model was fit to predict the TEOM(40) adjustment required based on the TEOM(40) measurement itself, the time of year and other meteorological predictors. This resulted in a tree defining seven data groups and additive adjustments to the raw TEOM(40) data.

Because of the large range of TEOM(40) values in one of the groups, it was felt that a multiplication factor would be more appropriate for this group than an additive adjustment, so this change was made. Table A1 lists the final seven adjustment rules. To adjust a TEOM(40) value, the temperature at 1m and the month of the measurement are required. Read down the 'criteria' column until a matching criterion is found, then apply the adjustment in the 'adjustment' column.

Table A1: Adjustment rules for one hour average TEOM(40) data.

Data group name	Criteria	Adjustment
Very high PM	TEOM(40) >= 77	multiply by 1.23
Summer, low PM	TEOM(40) < 18 and December-March	add 0.96
Non-summer, cool at 1m	TEOM(40) < 77, April-November, temp1m < 3.5	add 14.36
Summer, mid PM	TEOM(40) between 18 and 40, December-March	add -5.69
Summer, high PM	TEOM(40) between 40 and 77, December-March	add -24.77
April-May or August-November, warm at 1m	TEOM(40) < 77, temp1m > 3.5 April-May or August-November	add 4.66
June-July, but warm at 1m	TEOM(40) < 77, temp1m > 3.5, June-July	add 8.91

The adjustment has improved the correlation from R^2 = 0.89 to 0.93. More importantly, the RMA best-fit line now under-measures only slightly by 1-3% at higher values compared to a larger under-measurement of 18% before adjustment.

A2 Kruskal-Wallis test for significant differences

The Kruskal-Wallis (K-W) test can be used to test for significant differences between groups of skewed data, like pollution data. The test does not assume that the data come from a normal distribution; in fact it converts all values to ranks before analysis, thereby creating a uniform distribution. The K-W routine tests the hypothesis that all groups have the same median rank against the alternative that the median ranks are different. It returns a p-value for the likelihood that the observed differences could occur purely by chance. If more than two groups are being tested, a particular p value can be set and the significance of the difference between different groups can be displayed graphically.

A3 Regression trees method

Meteorology is known to have a large effect on atmospheric PM_{10}, so it is difficult to look for long term trends in PM_{10} without considering and controlling for the effect of meteorology. Regression trees were used because unlike linear regression models, they can handle both continuous and categorical predictors in the same model. They also do not assume linear relationships, instead fitting threshold relationships to the data. Interactions between predictors are also handled automatically.

The regression trees were fit using Matlab (Mathworks version 2006a). Because of the skewed nature of PM_{10} concentration data, the PM_{10} values were transformed if necessary before the model was fit. The following 7 predictors were allowed in each model: temperature at 1m, temperature at 10m, temperature difference (1m to 10m), wind speed, solar radiation, (all 1-hour average values), month (categorical predictor), hour of day (categorical predictor).

The tree model creates classes grouping similar PM_{10} values. The class membership criteria are specified using the (mostly meteorological) predictor data. The tree can be pruned back or left at full size. Either way it produces a number of meteorologically specified classes, with relevance to PM_{10}. More meaningful annual comparisons could then be made between the PM_{10} values within each meteorological class.

A4 Hourly analysis

A4.1 Regression trees based on complete hourly dataset to create high pollution subsets

Figure A1 shows a graphical representation of the regression tree performed on the full hourly dataset. The tree can be interpreted by starting with all the data in the top group. Any data with temperature at 1m < 4.6°C (cold) follow the left branch and all other data go right. The subset of (cold) data in the left hand group is now split on hour. Any cold data between the hours of 3 am and 6 pm go left and all other (cold) data go right. The other splits can be read in a similar way. The 6 final groups are shown with the average (square root) PM_{10} values and the group number.

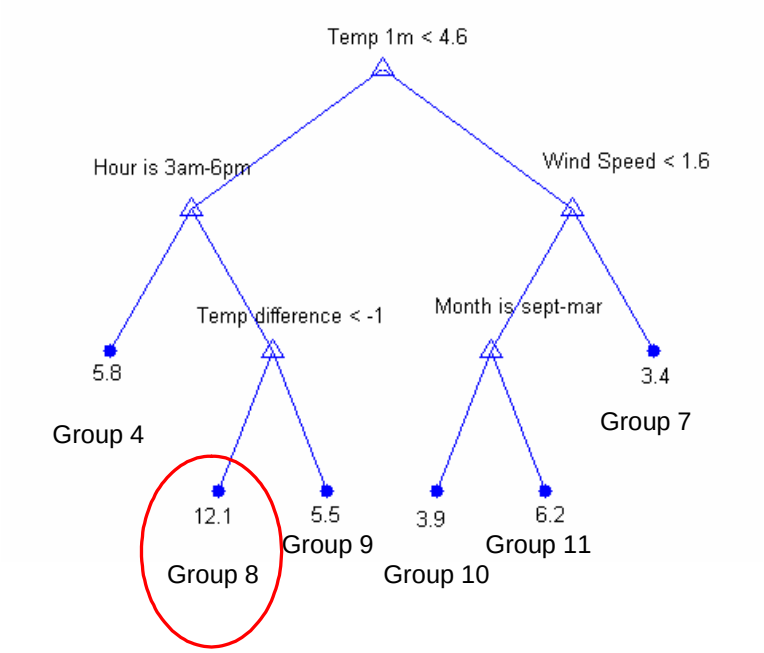


Figure A1: Regression tree separating PM_{10} concentrations into different meteorological and seasonal classifications. Blue dots mark final data groupings and are labelled with average square root of PM_{10} concentration and group number. The red ellipse marks the group designated as high pollution.

Temperature at 1m was the most important predictor in this model, contributing about 45% to the explanatory power of the model. Hour was next, contributing about 20%.

Figure A2 shows the range of PM_{10} values within each of the 6 groups produced by the tree. Group 8 (calm night time conditions with temperature inversion) contained most of the high PM_{10} values and will be termed the 'high pollution' subset. Most of the lowest PM_{10} concentrations are recorded in groups 7 and 10, characterised by warmer conditions occurring either near summer or with wind.

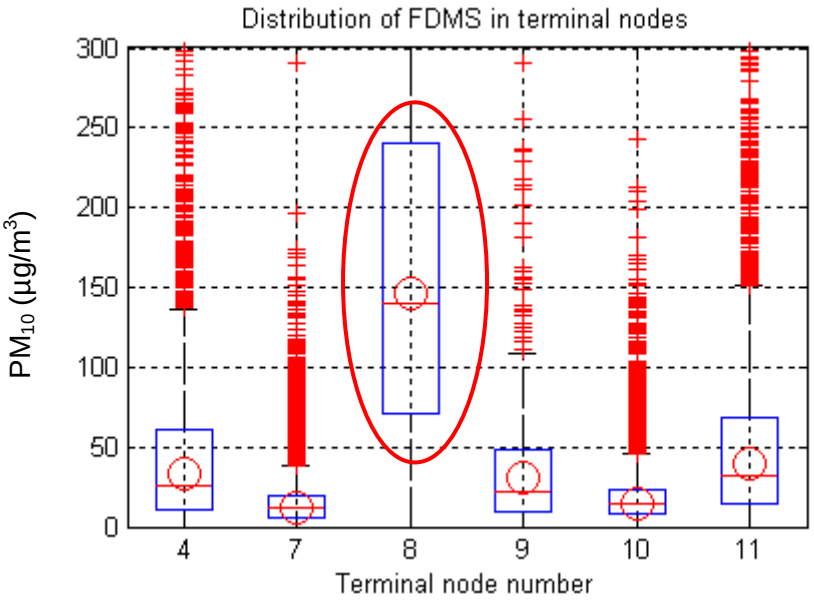


Figure A2: Distribution of PM_{10} within the groups produced by the tree model. Horizontal red lines inside the blue boxes indicate the median, red circles indicate the mean, and blue boxes encompass 50% of the data in each group. Black whiskers extend to 1.5 times the inter-quartile range and data outside that are marked by red crosses. The large red ellipse marks the group designated as high pollution.

A4.2 Significance of time trend results based on the high pollution hourly dataset

Figure 3.1 in the main report shows PM_{10} trends through time for data in the high hourly pollution subset (group 8). Figure A3 shows the significance of these differences using the Kruskal-Wallis significance test. Differences are significant where the bars do not overlap, i.e. the reduction in pollution between 2001 and 2003 is significant, but that between 2001 and 2002 is not.

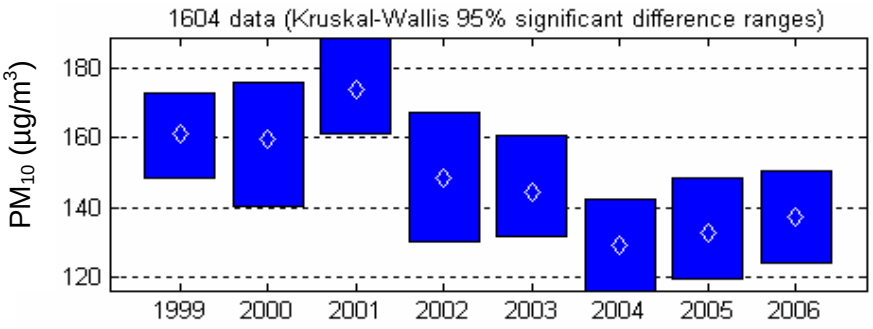


Figure A3: Significance of annual differences in the high pollution hourly subset (group 8) using the Kruskal-Wallis test. The white diamonds represent the median concentration of group 8.

In order to further reduce the amount of meteorological variability in this subset of data, a multiple linear model was fit to the data. This model showed that 18% of the variability within the high pollution subset could still be accounted for by meteorological and seasonal effects. The model can be used to predict a pollution value based solely on the meteorological predictor variables. It was then possible to compare the actual pollution with the meteorologically-predicted value. In theory, if emissions are higher, the observed pollution value is likely to be higher than that predicted using meteorology, and vice-versa. Figure A4 shows that the median actual pollution was higher than the predicted pollution in 1999-2002, but that the actual and predicted values have been quite similar since then. This suggests a potential decrease in PM_{10} emissions from 2001 to 2004.

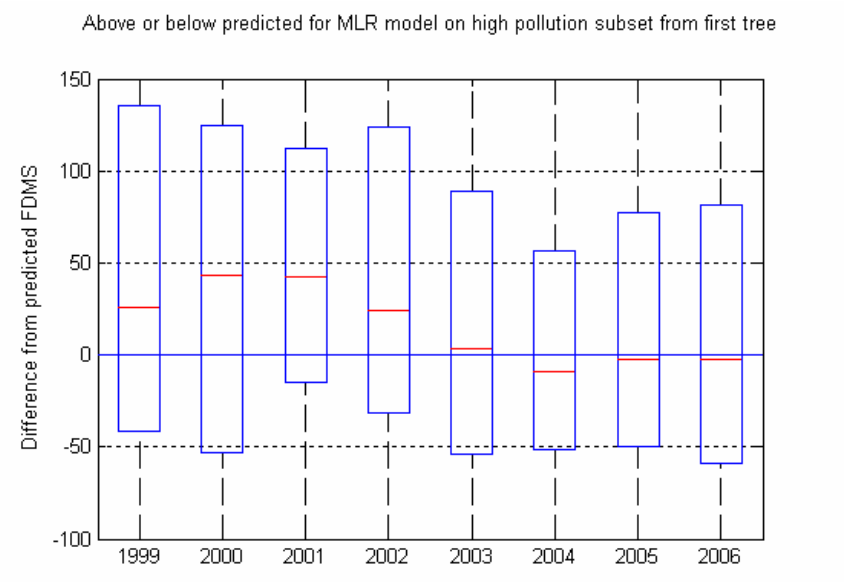


Figure A4: Comparison of actual pollution with meteorologically-predicted pollution (using a multiple linear regression model) based on a subset of high pollution hourly data (group 8).

A4.3 Regression tree based on high pollution hourly data to create extremely high pollution subset

Figure A5 shows the regression tree fit to the high pollution hourly data.

Figure A6 shows the range of PM_{10} values within each of the groups produced by the tree. Groups 3 and 4 contain most of the low pollution data and were separated early in the tree. Data from all groups except 3 and 4 will be termed the 'extremely high pollution' subset.

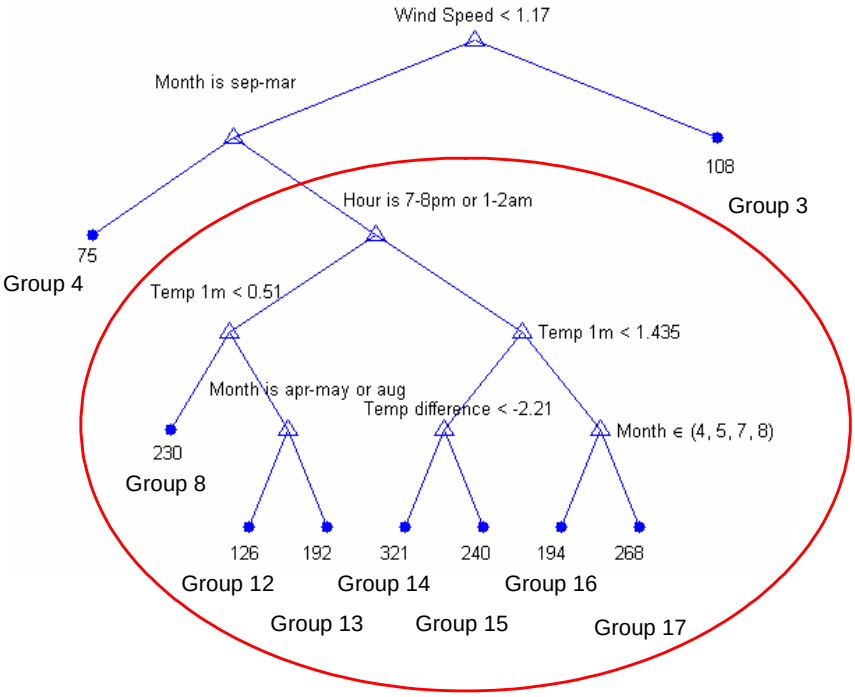


Figure A5: Regression tree fit to high pollution hourly data. Blue dots mark final data groupings and are labelled with average PM_{10} concentration and group number. The red ellipse marks the groups designated as extremely high pollution.

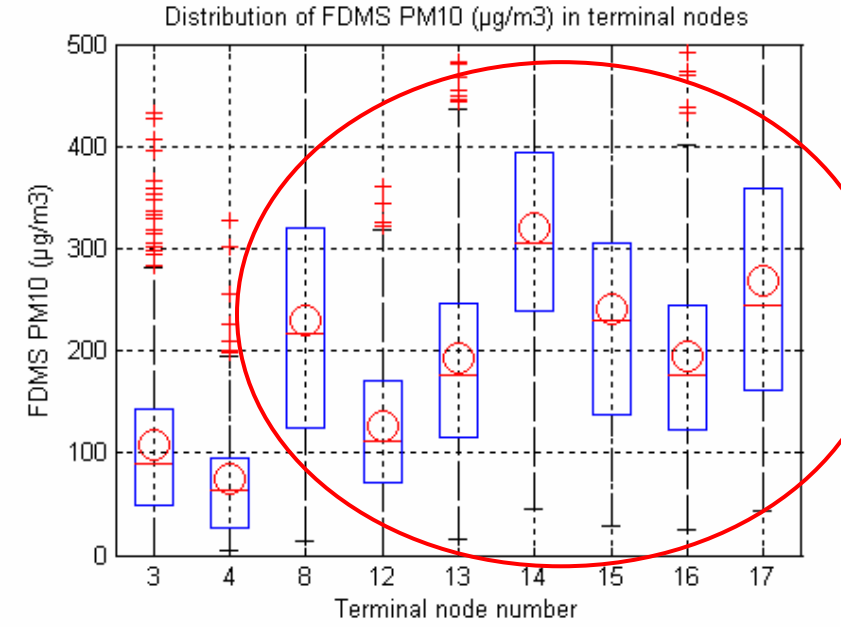


Figure A6: Distribution of hourly PM_{10} within the groups produced by the tree model. Horizontal red lines within the blue boxes indicate the median, red circles indicate the mean, and blue boxes encompass 50% of the data in each group. Black whiskers extend to 1.5 times the inter-quartile range and data outside that are marked by red crosses. The large red ellipse marks the group designated as extremely high pollution.

A4.4 Significance of time trend results based on extreme high pollution hourly data

Figure 3.2 in the main report shows hourly PM_{10} trends through time for data in the extremely high pollution groups. Figure A7 shows the significance of these differences using the Kruskal-Wallis significance test.

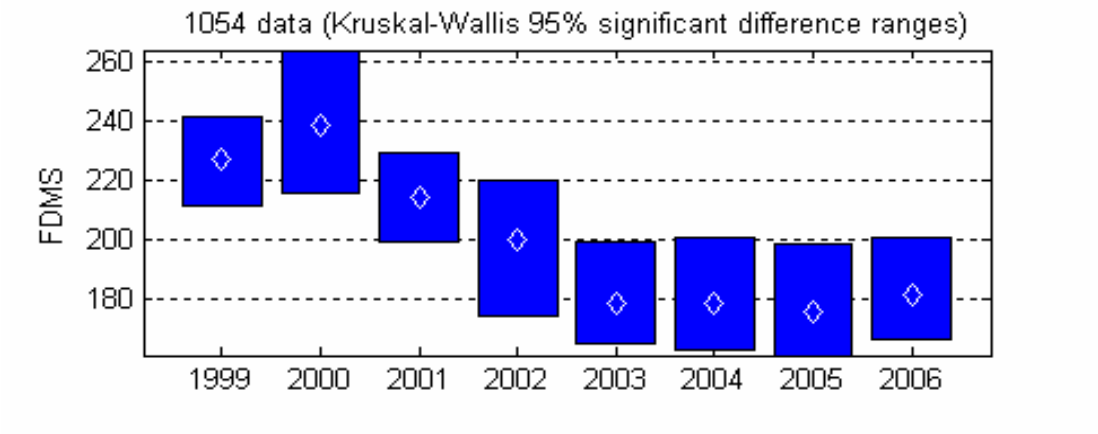


Figure A7: Significance of differences in hourly PM_{10} in the high pollution group (14) using Kruskal-Wallis test.

A multiple linear regression model was fit to this subset of data to see how much additional meteorological variability could be removed. The model could explain only 9% of the variability in the dataset, although it is not likely that all of the remaining 91% of variability can be attributed to emissions.

A5 24-hour analysis

A5.1 Regression tree on complete 24-hour dataset to create high pollution subset

Figure A8 shows the regression tree fit to the complete 24-hour dataset. Wind speed is the most important predictor in the model and contributes just over half of the explanatory power of the model. Nearly all of the October-March data fall into groups 18 and 19 (low pollution groups).

Figure A9 shows the range of PM_{10} values within each of the groups produced by the tree. The groups containing most of the highest pollution data are groups 16 and 17 (calm, cold days with temperature inversion). Group 26 also represents fairly high pollution. It contains slightly windier winter days that are cold, with a temperature inversion of at least -0.3 degrees.

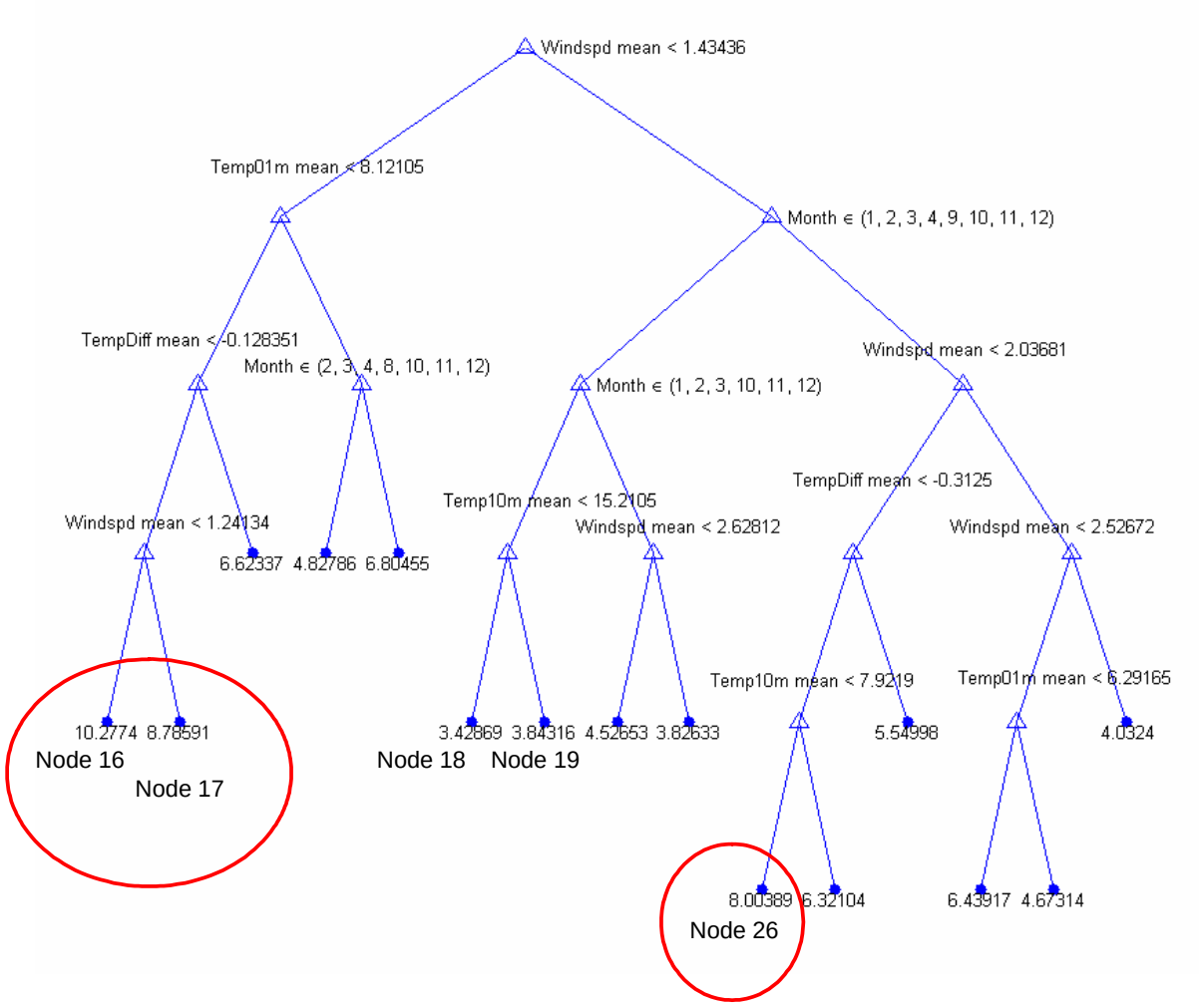


Figure A8: Regression tree fit to full daily dataset. Blue dots mark final data groupings and are labelled with average square root of PM_{10} concentration. Groups of interest are also labelled with the group number. The red circles mark the groups designated as high pollution.

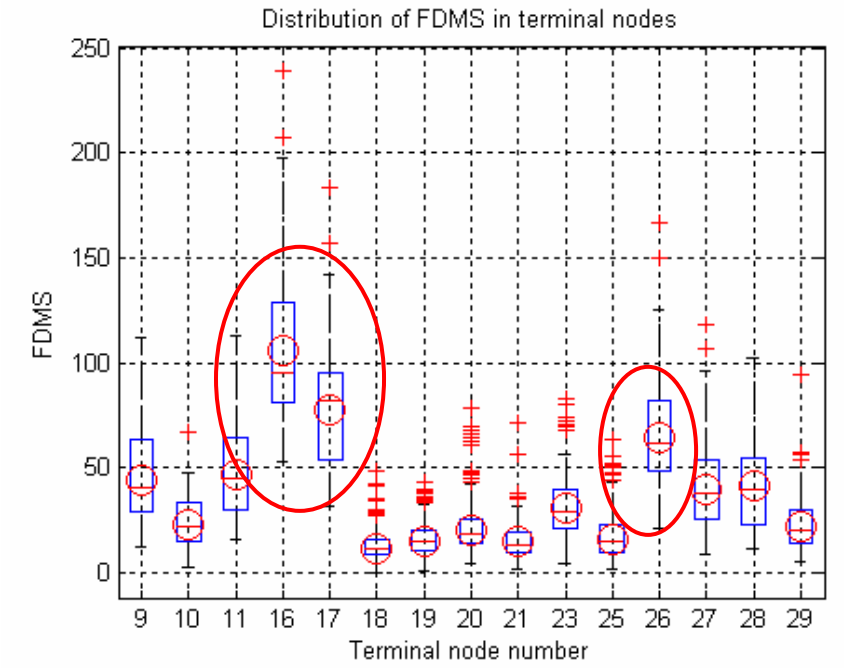


Figure A9: Distribution of PM_{10} within the groups produced by the tree model. Horizontal red lines within the blue boxes indicate the median, red circles indicate the mean, blue boxes encompass 50% of the data in each group. Black whiskers extend to 1.5 times the inter-quartile range and data outside that are marked by red crosses. The ellipses mark the groups designated as high pollution.

A5.2 Significance of time trend results for high pollution 24-hour data

Figure 3.1 in the main report shows PM_{10} trends through time for high pollution daily data. Figure A10 shows the significance of these differences using the Kruskal-Wallis significance test.

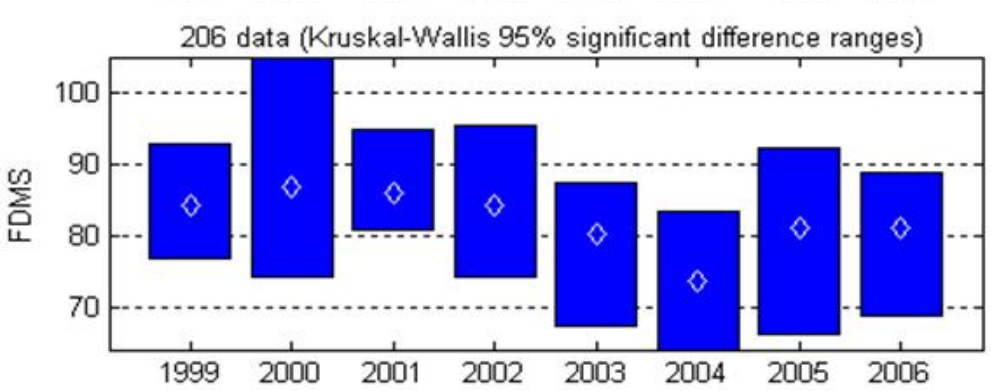


Figure A10: Significance of differences in high pollution daily PM_{10} using the Kruskal-Wallis test.

In order to further reduce the amount of meteorological variability in this subset of data, a multiple linear model was fit to the data. Figure A11 shows that the median actual pollution was higher than the predicted pollution in 2000-2001, and lower than predicted from 2002 to 2004.

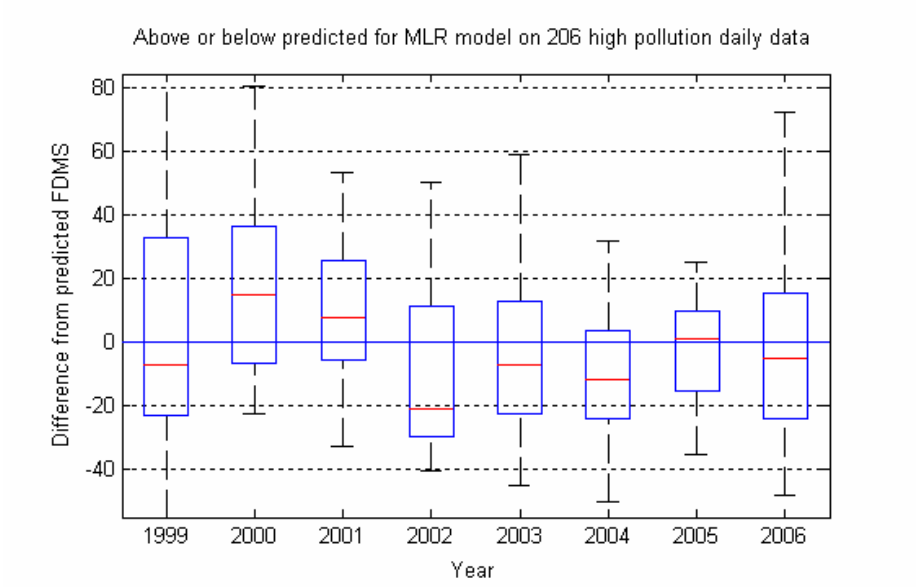


Figure A11: Comparison of actual pollution with meteorologically-predicted pollution (using a multiple linear regression model) for a subset of high pollution daily data.

Figure A11 can be compared with Figure 3.3 in the main report. Removing this extra 28% of variance has resulted in the positions of the maximum and minimum moving back by two years. The shape of the graphs are somewhat similar with an early rise and fall (peaking in 2000-2002, falling to a minimum in 2002-2004), followed by a very slight rise up till 2006.

Appendix B: Detailed report on the simple correlation analysis of PM_{10} and weather

B1 Objective

The objective of this component of the research is to examine the relationships between PM_{10} concentrations and various weather factors in Christchurch. The trends in ambient concentrations are assessed by examining the weather and climate conditions over the study period, allowing inferences to be made about trends in emissions.

B2 Summary Conclusions

This is a simplified graphical correlation analysis, examining straightforward relationships between temperature, wind and monitored PM_{10} concentrations over the years 1999 to 2006, which shows that:

The relationship between weather factors and the winter time PM_{10} concentrations appears to show concentrations have been decreasing over the 7 year period (but not uniformly)

This suggests that emissions from wood burners in the area around the Coles Place monitoring site in winter have been decreasing – i.e. for a given weather condition the resulting PM_{10} concentration is not as great in recent years as it was in the past.

The data and analysis techniques are too noisy to quantify this further at this stage.

B3 Background

Weather conditions affect pollutant concentrations both through changing people's behaviour (i.e. *"...is it cold enough to fire up the burner…?"*) and also by changing the rate of dispersion (i.e. *"…is it windy enough to blow the smoke away..?").* For example, cold days have a direct effect on home heating emissions, so that it is usually assumed that the colder the day, the more domestic home heating emissions occur. There is some evidence that the effect is more subtle than this rather obvious one, as some people may not start their burners until there have been two or three consecutive cold days. Also, different people have different temperature thresholds, and finally, the 'type' of cold weather may influence heating decisions, with still clear day-time

cold weather being less of a trigger than cold wet nights. So temperature alone may not be very highly correlated with emissions beyond the obvious general seasonal factors.

Other weather conditions such as wind speed affect pollutant concentrations. High winds increase dispersion, while long periods of light winds can result in decreased dispersion and elevated pollutant levels.

A more subtle and difficult to analyse factor is the occurrence of inversion layers. These are due to complex meteorological factors and their extent or frequency is not measured directly. Their details are inferred from various other measurements, and they can have a strong influence on the trapping of air pollution.

The study conducted here attempts to elicit primary relationships between PM_{10} and the simple weather variables of temperature and wind speed. It does not include an analysis of the secondary factors – such as periods of cold weather that may induce people to use their home heating appliances more.

B4 Methodology

Creating one continuous data set from 1999 to 2006 necessitated combining data from various monitors (Table B1). It is acknowledged that combining these data sets introduces variability. However, the analysis below focuses more on long-term trends in PM_{10} concentrations and the impact of weather variability upon them. The data set analysed combines monitoring data from a TEOM@40[1] and two FDMS TEOM@30 monitors. In general, FDMS TEOM data were preferred, although earlier monitoring years 1999 to 2002 consisted of only TEOM@40 data. To account for the tendency for TEOM@40 monitors to under-measure at higher concentrations, and to achieve consistence between the monitors, a correction factor was applied, as follows: if the TEOM@40 measured 44 µg m^{-3} or higher, the FDMS equivalent = (TEOM@40 + 3.15) /.75, and if TEOM@40 measured less than 44 µg m^{-3}, the FDMS equivalent = (TEOM@40 -2.23) /.74. This correction factor was recommended by Environment Canterbury.

[1] These figures refer to the inlet temperature of the TEOM (tapered element oscillating microbalance) instrument used to measure ambient PM_{10}. The filter is heated to prevent moisture entering the instrument and can be set between 30°C and 50°C. Heating significantly above ambient temperatures tends to evaporate some of the volatile component of the PM_{10}, producing lower values that need to be corrected to agree with the reference method. The FDMS is a recent adaptation that splits the flow, drying and heating to 30°C and chilling to capture volatiles otherwise lost. Concentrations from this method agree well with the reference method.

Table B1: Overview of PM_{10} monitoring at Coles Place, St. Albans, Christchurch.

Year	PM_{10} data sets[1]			Data set used for analysis[2]
1999	TEOM@40°			Corrected TEOM@40°
2000	TEOM@40°			Corrected TEOM@40°
2001	TEOM@40°			Corrected TEOM@40°
2002	TEOM@40°			Corrected TEOM@40°
2003	TEOM@40°	FDMS TEOM@30°		Primary: FDMS TEOM@30 Secondary: Corrected TEOM@40°
2004	TEOM@40°	FDMS TEOM@30°#11	FDMS TEOM@30°#12	P rimary: FDMS #11 TEOM@30 Secondary: : FDMS #12 TEOM@30 Tertiary: Corrected TEOM@40°
2005	TEOM@40°	FDMS TEOM@30°#11	FDMS TEOM@30°#12	P rimary: FDMS #11 TEOM@30 Secondary: : FDMS #12 TEOM@30 Tertiary: Corrected TEOM@40°
2006	FDMS TEOM@30°			FDMS TEOM@30°

[1] Datasets do not always represent complete or near complete years. Monitoring is often intermittent.

[2] Datasets were often combined to account for missing data. Primary denotes the main dataset used. Secondary and Tertiary data was used when primary data were not available.

B5 Note on Calms

This study makes reference to calm winds or periods of calm winds. For the purposes of this study calm winds were designated as those with hourly averages less than 2 m s^{-1}. This is a relatively arbitrary designation, although it is based on the data plotted in Figures B1 and B2 which show the 1-hour PM_{10} concentration versus the hourly wind speed. As can be seen from the figures, wind conditions less than 2 m s^{-1} result in reduced dispersion, leading to a higher number of hours with elevated concentrations of PM_{10}.

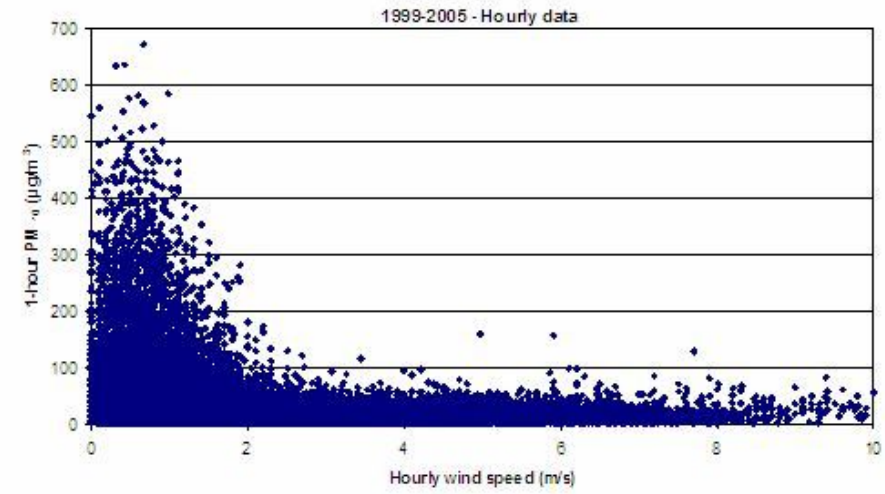


Figure B1: 1-hour PM_{10} versus hourly wind speed (0-10 m s^{-1}) at Coles Place, 1999-2005.

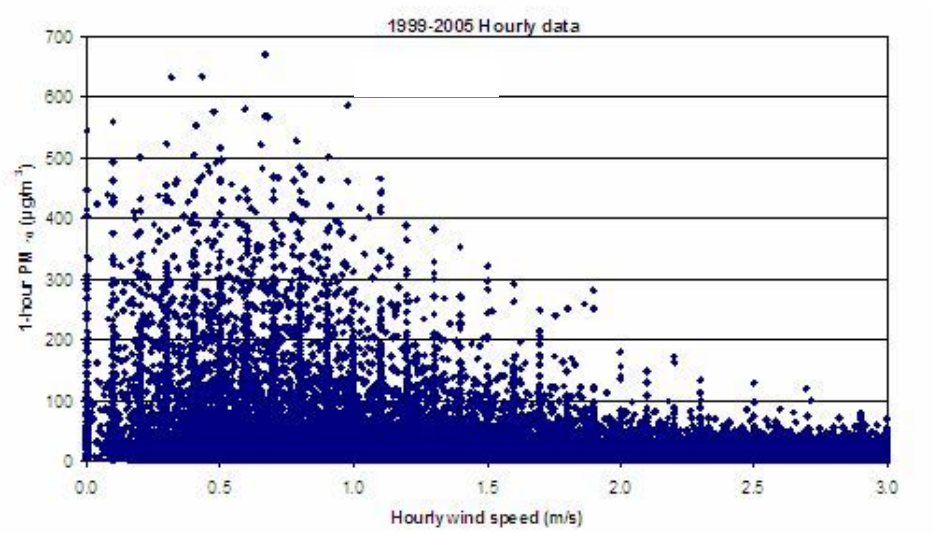


Figure B2: 1-hour PM_{10} versus hourly wind speed focusing on lower range of wind speeds (0-3 m s^{-1}) at Coles Place, 1999-2005.

Clearly, there is a range of choices that can be made when defining calms. The accepted meteorological definition is 0.5 m s^{-1}. This may be acceptable for air pollution analysis, but from the data in Figure B3 it is seen that a significant number of high pollution hours can occur at wind speeds above this. By the time the wind speed is 2.5 m s^{-1} or greater, the concentrations definitely decrease. The value of 2.0 m s^{-1} was chosen to encapsulate all the high pollution concentration hours (concentrations above 200 µg m^{-3}) in the data set. Similar results are obtained with choices of 1.5, 1.0, or 0.5 m s^{-1}, although the amount of data is less, and therefore the statistical results less significant.

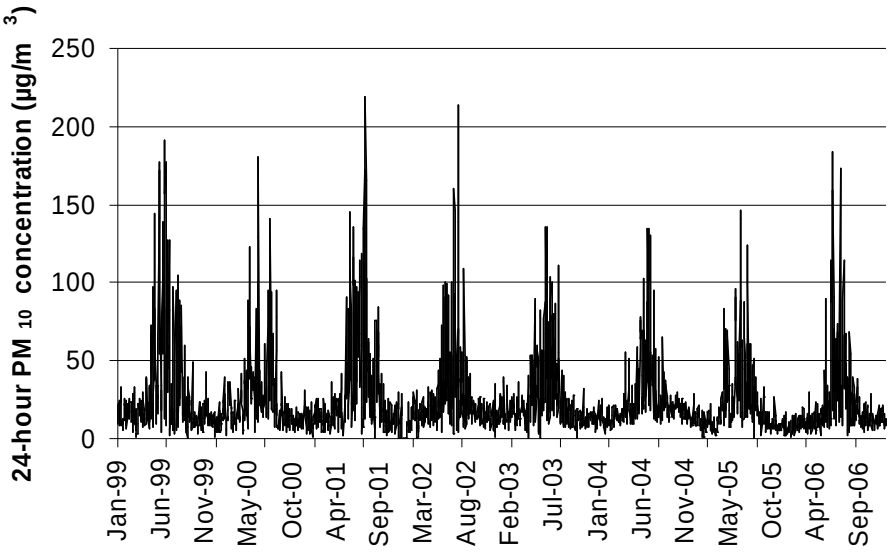


Figure B3: Historical 24 hour PM_{10} record for Christchurch using data from Coles Place (1999-2006).

B6 Results

Simple correlations and plotted relationships are used to determine the interaction between PM_{10} and weather conditions in Christchurch by examining PM_{10} data monitored at Coles Place from 1999 to 2006. The basic corrected hourly data series is shown in Figure B4, while Table B2 shows the correlation of various winter weather conditions with the average PM_{10} for June and July. Temperature and the number of calm winds show the highest correlation.

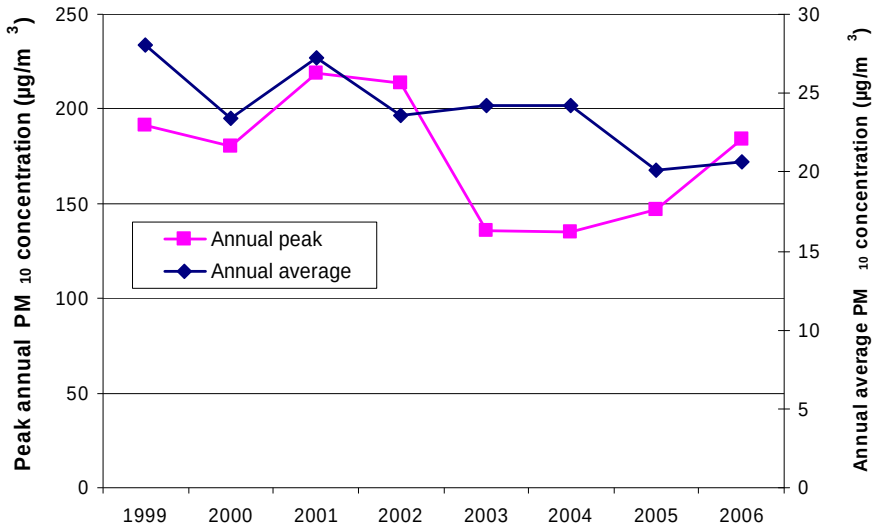


Figure B4: Annual average PM_{10} concentration (µg m^{-3}) and annual peak PM_{10} concentration (µg m^{-3}) for Coles Place (1999-2006).

Table B2: Correlation of average 24-hour PM_{10} with various weather conditions for the months of June and July.

June and July of:	Average PM_{10} (Coles) (µg m^{-3})	Average temp at 1m (ºC)	Average temp at 10m (ºC)	Average temp difference (1m-10m) (ºC)	Wind speed (m s^{-1})	Average hours of calms (/day)	Average percent of calm winds (%)	Average RH (%)
1999	58.0	6.7	7.9	-1.2	2.2	13.0	54.1%	85.1
2000	38.3	8.4	9.2	-0.8	2.3	11.5	48.1%	74.1
2001	64.8	5.6	6.2	-0.6	1.8	15.5	64.4%	72.0
2002	48.7	6.4	6.5	-0.1	1.8	14.0	58.4%	79.1
2003	49.9	6.8	7.4	-0.6	1.9	14.0	58.5%	77.0
2004	51.0	6.9	7.6	-0.7	2.0	14.0	58.5%	75.7
2005	46.2	6.9	7.0	-0.1	2.0	13.8	57.4%	78.1
2006	50.8	5.7	na	na	1.9	na	na	na
Correlation		**-0.79**	**-0.63**	**-0.26**	**-0.54**	**0.75**	**0.75**	**0.07**

The correlation between PM_{10} and temperatures and calms is not independent. Cold temperatures and calm winds can themselves be correlated, although not necessarily completely equivalent in producing high PM_{10} values. For instance, there could be cold windy days with low PM_{10}, or relatively warm calm days with high PM_{10}. The correlation with vertical temperature difference (being a crude indicator of inversions) is not at all strong (-0.26). Temperature inversions are transient in nature, and vary in depth, which has a significant effect on the ambient PM_{10} concentration on an hour by hour basis. These transient events are not able to be resolved and accounted for when using 24-hour averages, as has been done here.

Figure B4shows an overall downward trend in both annual average and annual peak PM_{10} concentration despite year-to-year fluctuations. To some extent these fluctuations may be due to emissions reductions, although actual trends cannot be determined without examining weather conditions.

Figure B5 shows that the majority of exceedences occur on cold days with calm winds (The results do not change significantly if 1.5 or 2.5 m s^{-1} are used to define 'calm', but they do if lesser or higher speeds are used). Few exceedences occur on days when calm wind periods occur for less than 50% of the time. Also, most exceedences occur when the average daily temperature is below 10ºC.

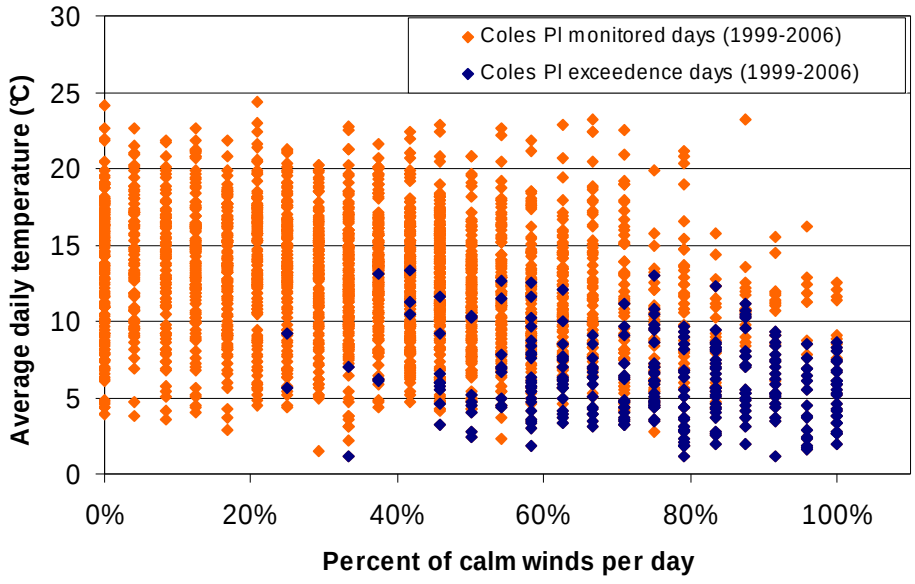


Figure B5: Comparison of monitored days and days which exceeded the NES for Coles Pl (1999-2005) using average daily temperature and percentage of calms.

Figure B6 shows the time of year and temperature at which exceedences occurred for 2001. The data show that the majority of exceedences occur during the winter months. In order to determine why some cold, winter days do not result in an exceedence, an

additional figure is presented to examine wind conditions during June and July. Figure B7 shows that the magnitude of the 24-hour PM_{10} concentration is closely related to the hours of calm wind periods experienced that day. A similar pattern exits for all the years in the study (not shown).

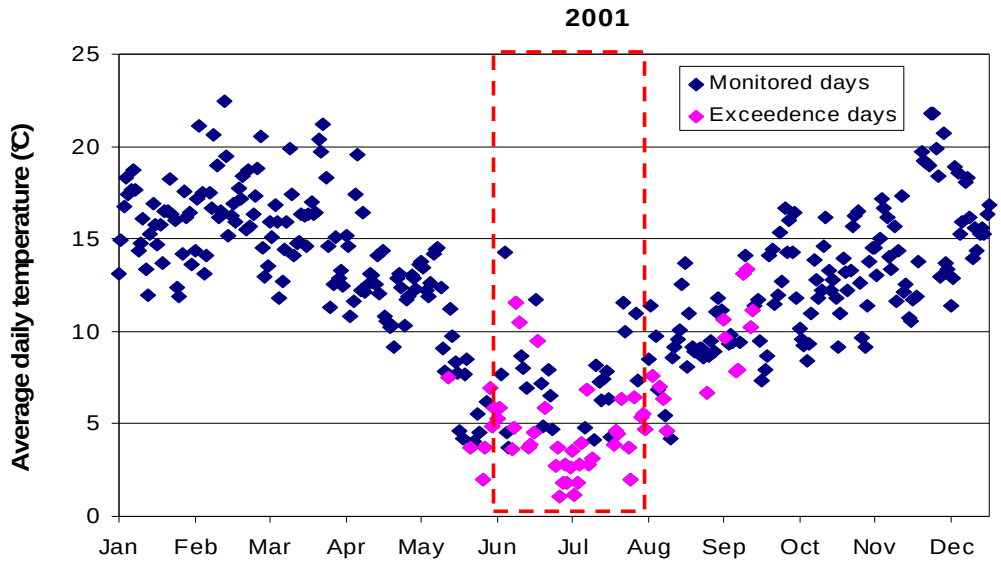


Figure B6: Monitored days and exceedence days for 2001. The red box indicates the dates chosen for further analysis as seen in the figure below.

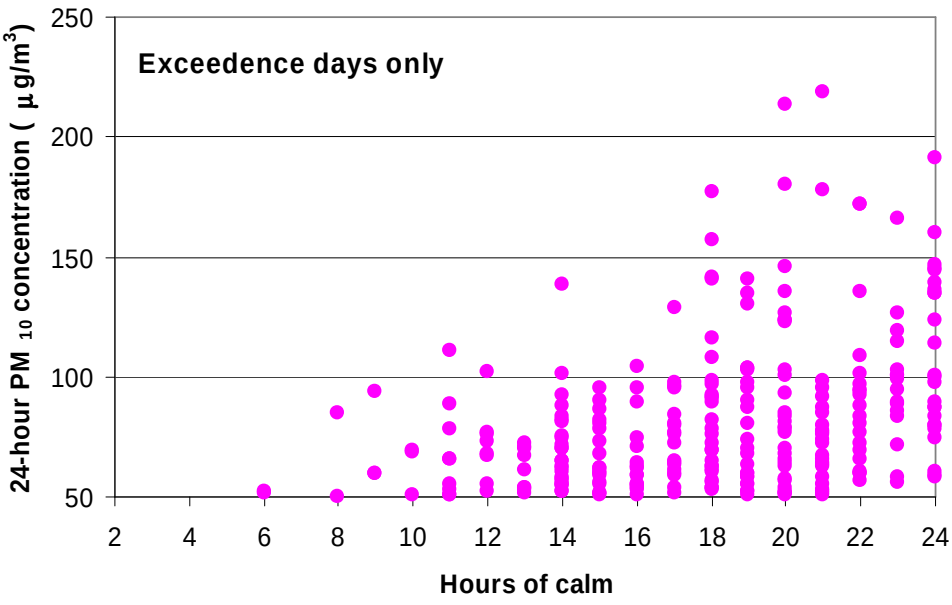


Figure B7: Correlation of calm wind conditions and 24-hour PM_{10} concentrations for exceedence days, 1999 to 2006.

There is not a particularly strong relationship between temperature and the average number of calms winds. However, 2001 experienced both the coldest temperature and the highest number of hours of calm winds. This year also experienced the second highest PM_{10} concentrations (only just lower than 1999).

B6.1 Trends

Figures B8, B9 and B10 show days with similar weather conditions throughout the 8 years studied, providing a simplified way to 'de-trend' the PM_{10}. Here the 'weather' is determined only by air temperature and wind speed. Ideally, it should include other parameters that are known to affect pollution concentrations – particularly the presence and strength of temperature inversions. However, data on inversions are not readily available, as this is a difficult parameter to measure. Some further work is being undertaken using limited temperature profile data as an indicator of possible inversions.

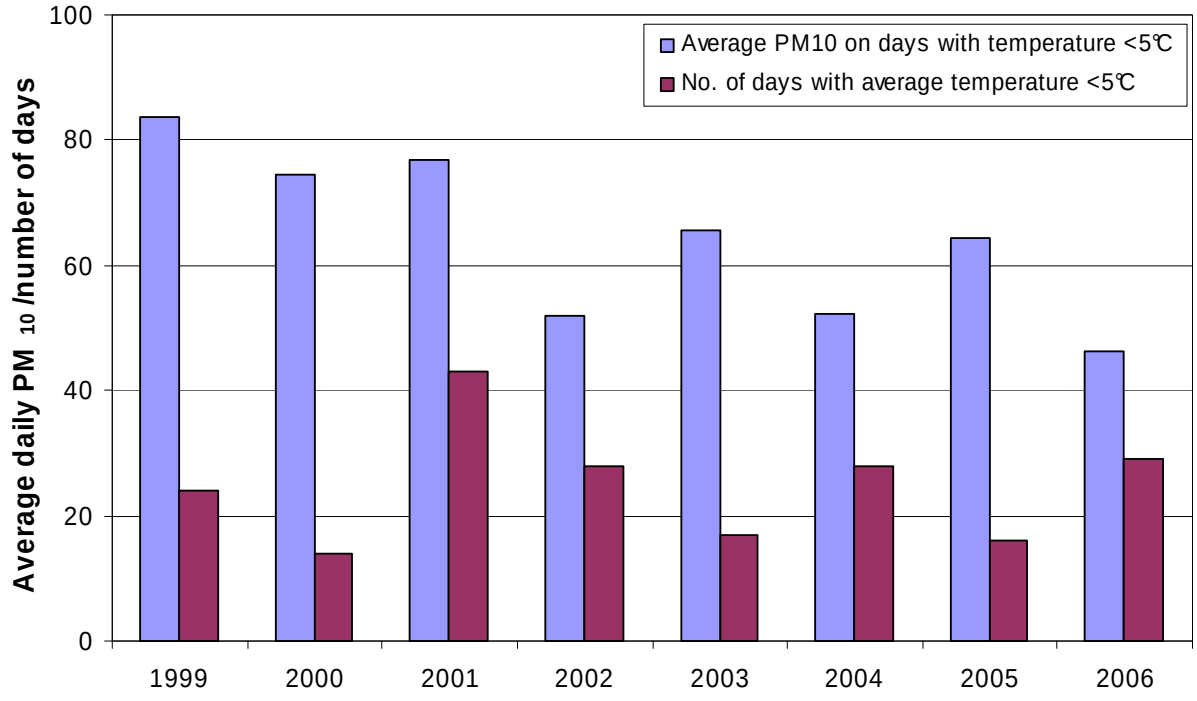


Figure B8: Trend analysis: average PM_{10} concentrations on days with an average temperature below 5ºC.

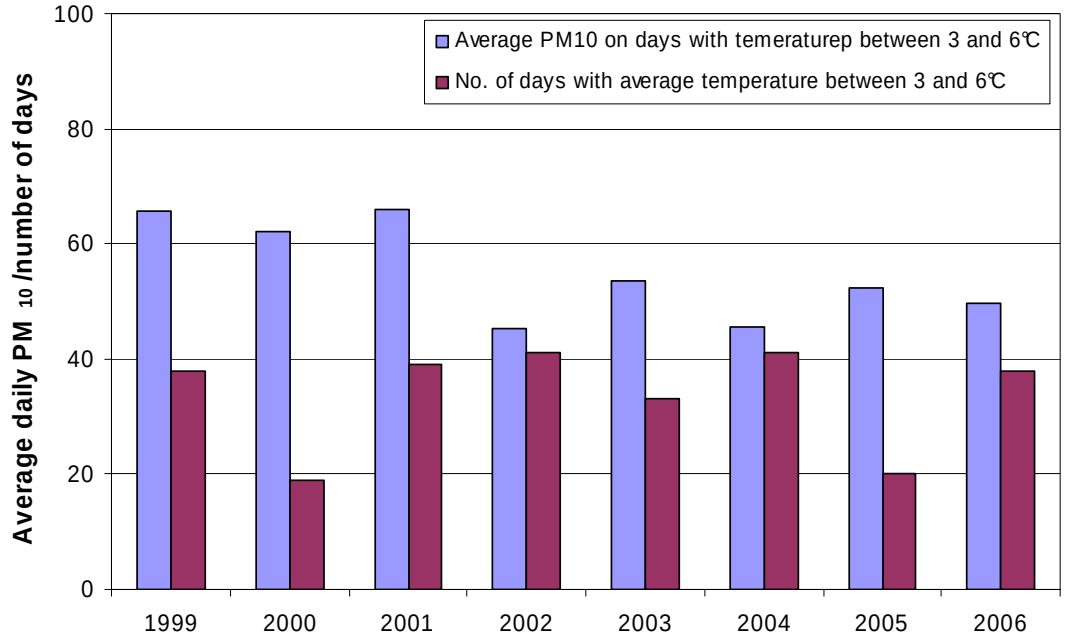


Figure B9: Trend analysis: average PM_{10} concentrations on days with an average temperature between 3 and 6ºC.

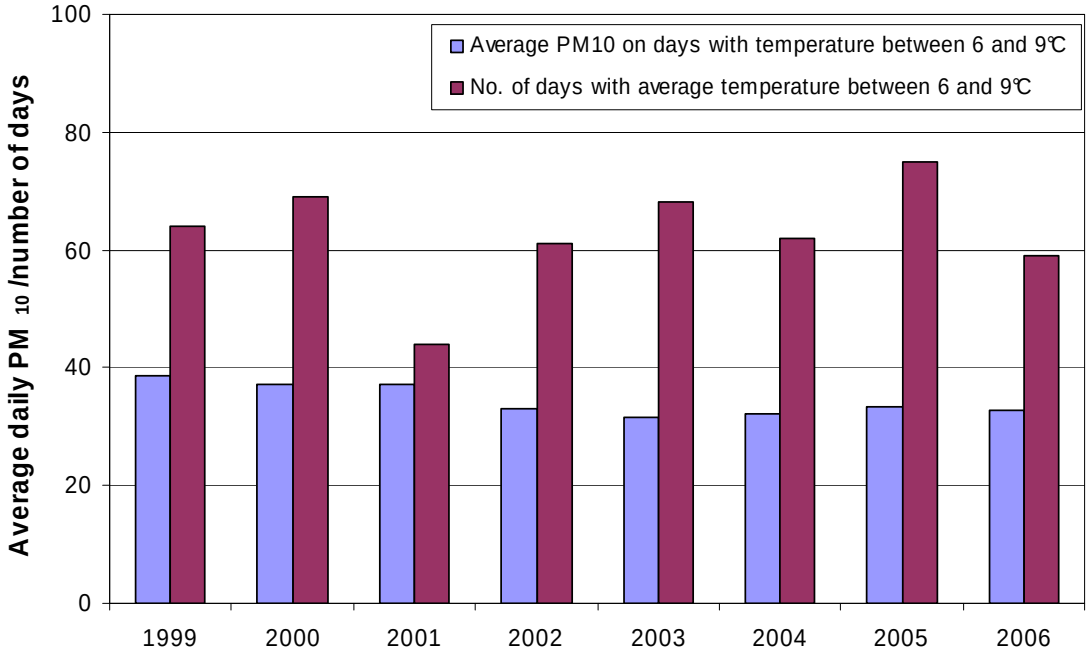


Figure B10: Trend analysis: average PM_{10} concentrations on days with an average temperature between 6 and 9ºC.

B6.2 Trends for PM_{10} on cold days

Each of Figures B8, B9 and B10 shows that for all the choices of the definition of "cold' (being <5 ºC, 3-6 ºC, and 6-9ºC) there is a discernable trend for lower average PM_{10} concentrations in the more recent years. These temperature choices are arbitrary. This is not uniform from year-to-year and cannot be used to make any prediction about what might occur in future years. It does show a level of correlation that implies emissions are reducing, but does not discount the possibility that more subtle weather features have an influence.

B6.3 Trends for exceedences

Figure B11 shows a simple correlation between the number of exceedence days and the average temperature for the two winter months of June and July. In the most general way, colder winters tend to have a higher number of exceedences (for instance, 2001 was particularly cold and had the highest number of exceedences, conversely 2000 was relatively warm and had a lower number of exceedences). However, 2006 was as cold as 2001 (on average) but had far fewer exceedences. This may be due to emissions reductions, or may be the result of a more subtle feature in the weather that is not explained by average temperature.

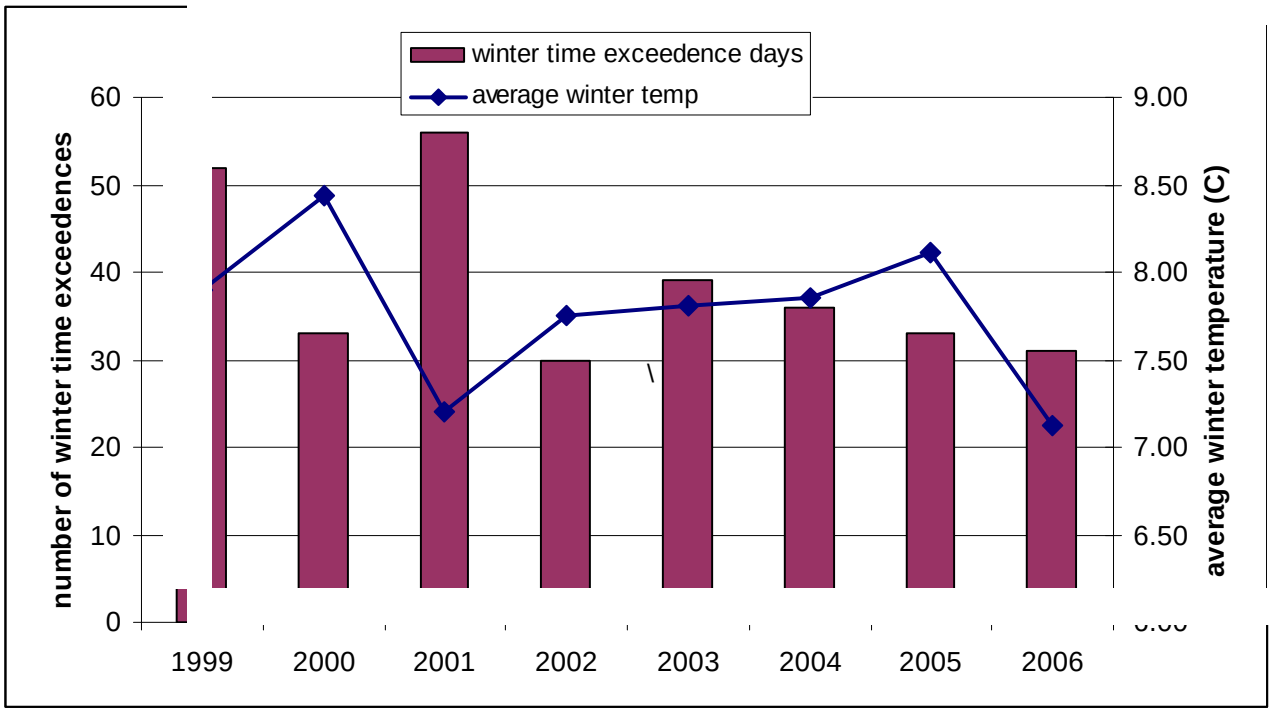


Figure B11: Winter-time exceedences and average temperatures.

B7 Correlation Model

Figures B12, B13 and B14 show daily PM_{10} concentrations as a function of temperature and calms for each year in the study. Figure B14 does not contain data for 2006 as it was unavailable at the time the analysis was undertaken. These curves are basic logarithmic fits to the relationship between daily temperature and daily PM_{10} concentration. They indicate in a simple way the expectation of a particular concentration occurring on a given day of a particular temperature. The logarithmic fit was chosen (as opposed to linear or polynomial) as it gives a better fit to the data. The data show a large degree of scatter and must be interpreted with caution.

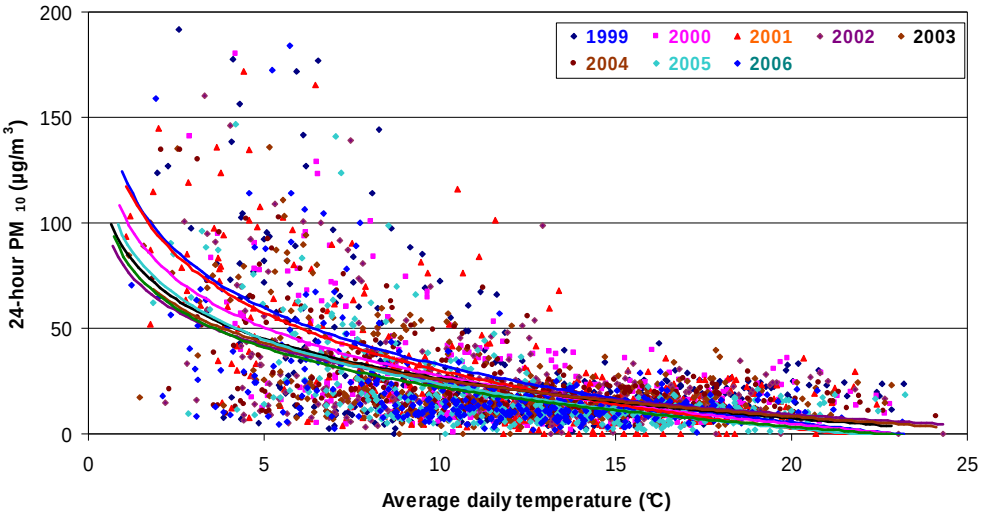


Figure B12: All days plotted, showing annual differences in the relationship between PM_{10} and temperature using logarithmic regression.

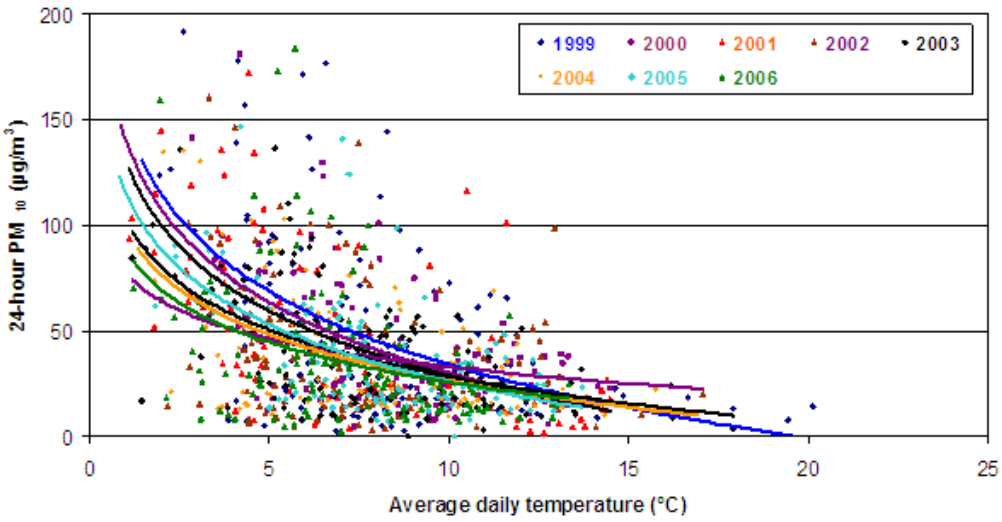


Figure B13: Only June and July, showing annual differences in the relationship between PM_{10} and temperature using logarithmic regression.

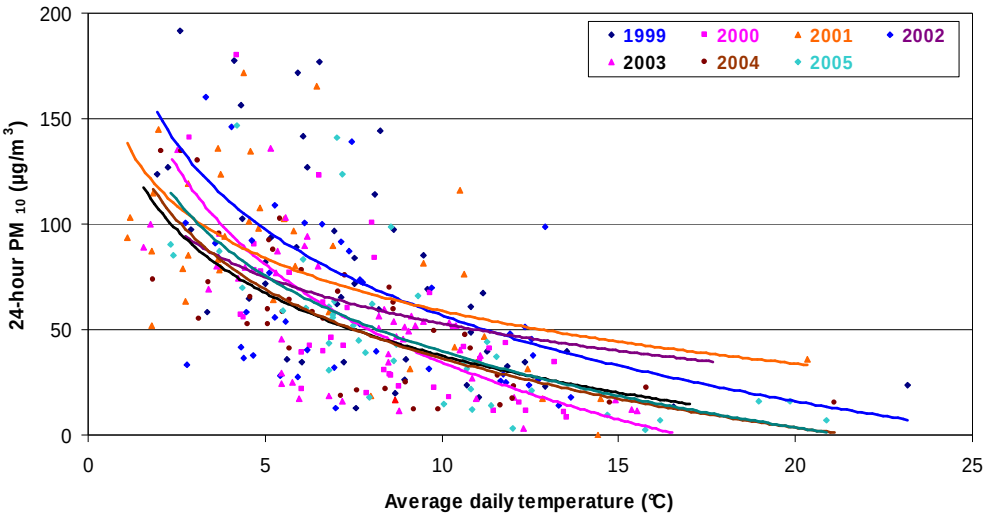


Figure B14: All days with daily calms 75% or greater (16 hours), showing annual differences in the relationship between PM_{10} and temperature using logarithmic regression

These results indicate that, overall, recent years such as 2003, 2004 and 2005 have experienced lower PM_{10} concentrations than earlier years during conditions of similar temperatures, or similar frequencies of calm periods.

Although these figures are very noisy, with a relatively large scatter in the data, they do indicate a decreasing slope in the relationship between PM_{10} and temperature (especially when the coincident occurrence of calms is also taken into account). In other words, for a given set of weather conditions (winter temperatures essentially), the PM_{10} concentration seems to be lower in more recent years (e.g. 2004 and 2005) than in previous years (e.g. 1999, 2000) suggesting a lower level of emissions in the more recent past. This is analysed in more detail in the next section.

B7.1 Correlation overview

The results presented so far are difficult to interpret due to the large scatter in the data. The data were transformed in various ways in order to test the fit of various relationships between parameters. The best fit was obtained with a logarithmic transformation of the concentrations (shown in Figure B15).

These relationships are explored further, where the coefficients of the logarithmic relationships are plotted for each year. These are simply called “Factor A” and “Factor B”, from the logarithmic fits of the form:-

24-hour PM_{10} = -A x Ln(weather parameter) + B

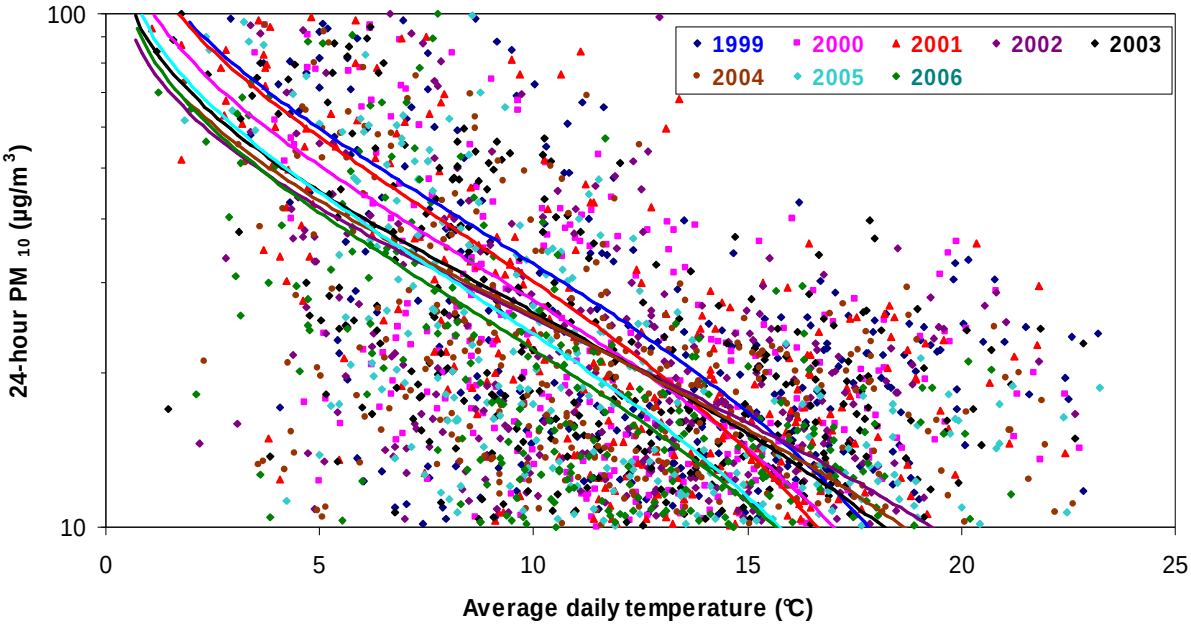


Figure B15: All days plotted, showing annual differences in the relationship between PM_{10} and temperature using logarithmic regression, with the concentrations plotted on a logarithmic axis.

The nature of the correlations are summarised further in Figures B16, B17 and B18. These reflect the form of the relationships plotted in Figures B12, B13 and B14.

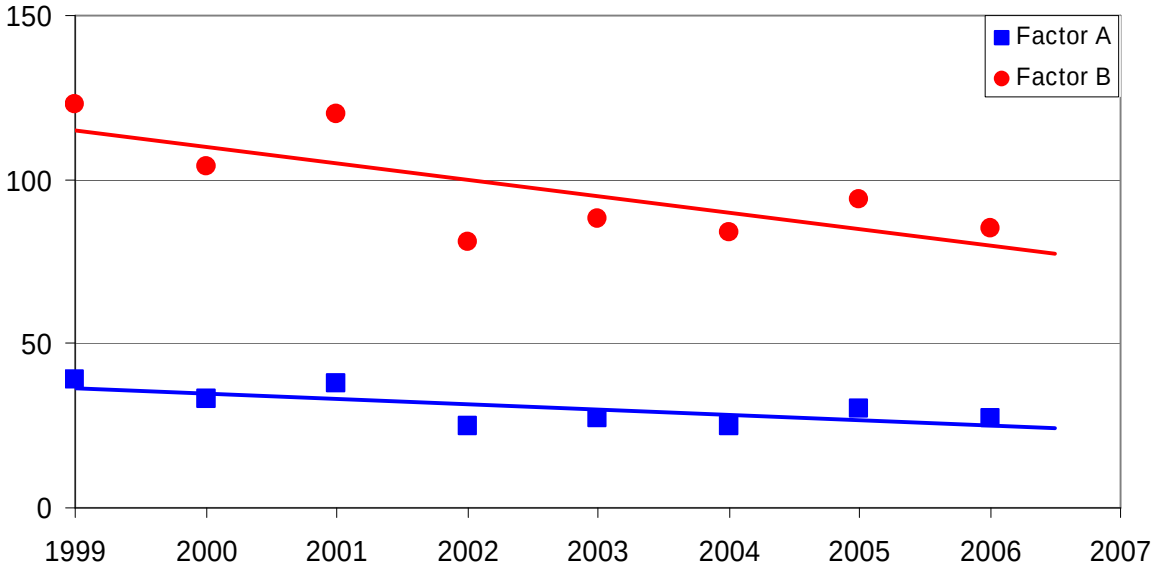


Figure B16: Summary of the regression relationship between winter temperatures and average PM_{10}, where A and B are parameters from the logarithmic equation described previously.

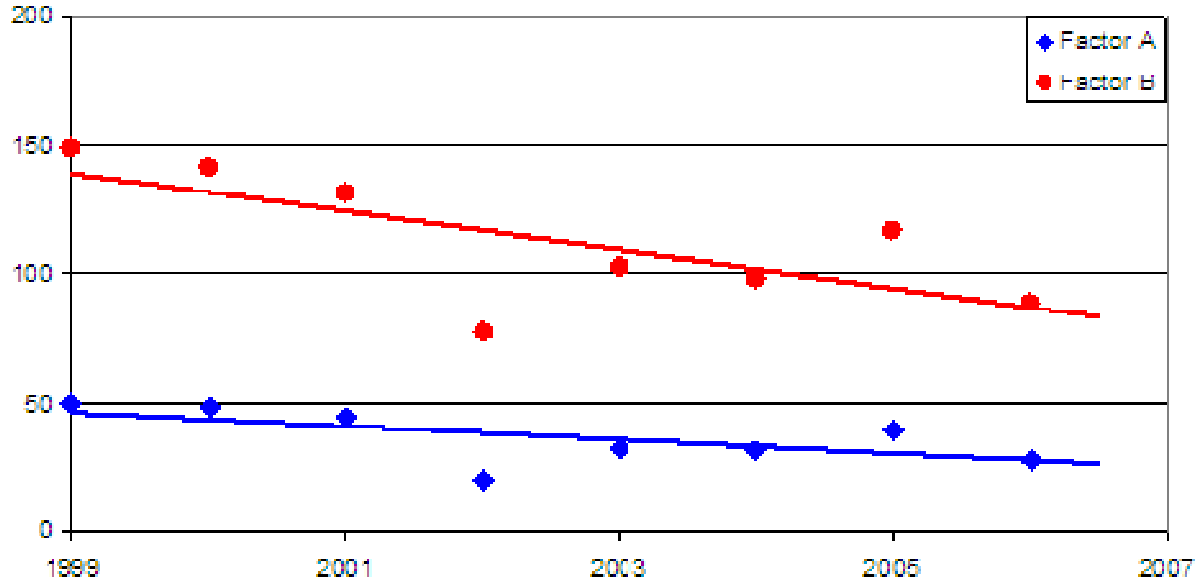


Figure B17: Summary of the regression relationship between June and July temperatures and average PM_{10}, where A and B are parameters from the logarithmic equation described previously.

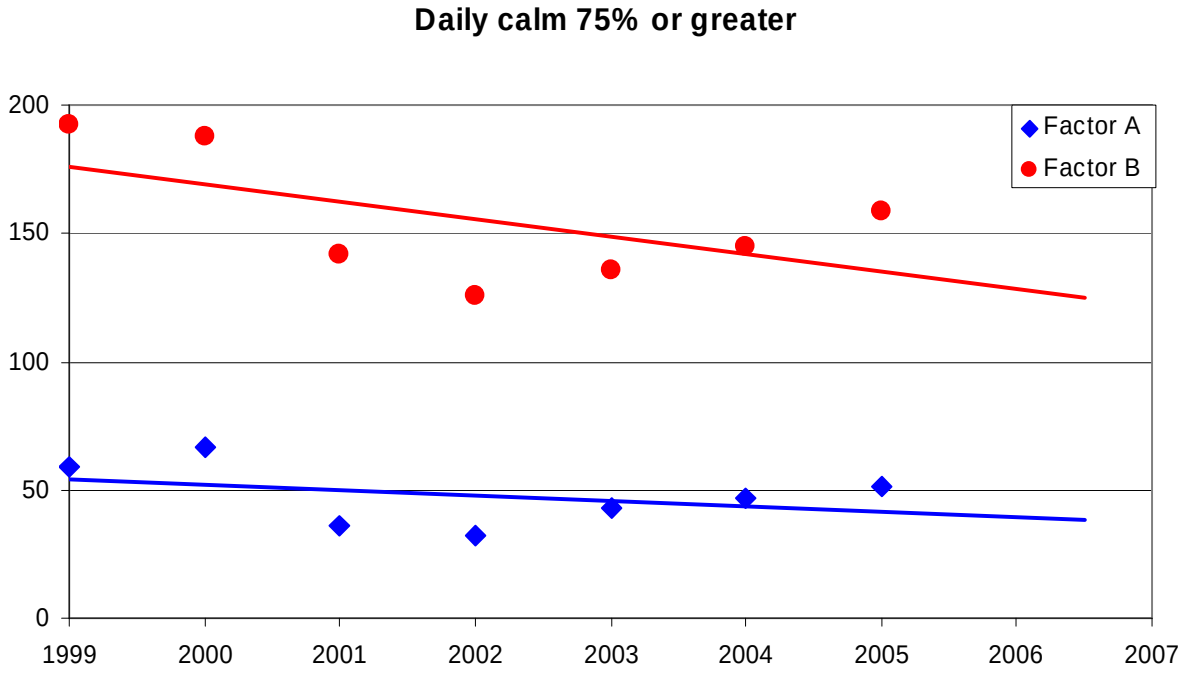


Figure B18: Summary of the regression relationship between calm periods and average PM_{10}, where A and B are parameters from the logarithmic equation described previously.

These figures thus highlight the way the relationship between emissions and concentrations has changed from year-to-year, under given weather scenarios (based on air temperatures and number of calm periods).

This is not a realistic relationship, but a simplistic way of describing these curves. "Factor A" and "Factor B" do not represent physical variables, since they are simply descriptors of an arbitrary transformation of the relationship between daily temperatures and averaged concentrations. Alternative fits could have been used, such as:

$$\text{24-hour } PM_{10} = A \times \exp(B \times \text{average daily temperature})$$

This is also a reasonably good fit to the data and can produce slightly more realistic parameters, with "A" being an estimate of the PM_{10} concentration at 0°C and "B" being the change in temperature leading to a 1/e reduction in PM_{10} concentration. The simplistic logarithmic factors do, however, give some indication of the way this relationship has been changing from year-to-year over the period, shown in Figure B18.

Both factors have been decreasing, but with some degree of year-to-year variability, as yet unexplained. It is important to note that these factors do not represent any particular geophysical parameters – they are statistical descriptors of the data analysed. This analysis was also conducted for the average temperature over the whole winter period, and for the frequency of calms.

There is considerable year-to-year variation. The 2001 year stands out, having been a particularly cold year with a higher than normal amount of calm periods. The relationship is strongest with average temperature. It is likely that some of the variation in these figures is due to weather variables that have not been included in the analysis – such as inversions.

B7.2 Summary

The key findings of the analysis are:

- PM_{10} concentrations are strongly dependent on the season, with 24-hour concentrations peaking in winter with values in excess of twice the standard (summer time PM_{10} concentrations rarely exceed 20 µg m^{-3} and almost never exceed 30 µg m^{-3}).

- Over the winter months of June and July from 1999 to 2006, exceedence of the 50 µg m^{-3} standard occurred on 42% of all days on average

- Ambient air temperature exhibits the strongest correlation with the number of exceedences, with the next strongest correlation being that of the number of periods of calms winds. Colder weather generally results in higher emissions, and higher concentrations.

- Wintertime PM_{10} concentrations are heavily dependent on the wind conditions, with more calm periods generally resulting in higher concentrations.

- The relationship between emissions and concentrations for given temperature conditions appears to have been weakening between 1999 and 2006, suggesting that emissions may have been reducing over this period.

When some of the year-to-year weather variability is accounted for, there is some indication that emissions have been reducing over the last few years.

This has been a straightforward graphical analysis with simple correlation calculations. It is important to understand the complex nature of the relationships being analysed through visual inspection, as well as using advanced statistical techniques. A natural extension to the analysis is to now use more powerful statistical techniques in order to (a) quantify in more detail the relationships discovered, and (b) give an idea of the statistical significance of the relationship.

B8 Discussion

This analysis shows some level of correlation between weather and PM_{10}, but it is by no means strong. As noted earlier, the relationship is likely influenced by other weather factors than have not been analysed – such as inversions. Some attempts will be made to do this, but good data are simply not available. They do not account for the secondary variations that must occur for 'colder' versus 'warmer' winters.

B8.1 Auckland

Some preliminary results for Auckland are less encouraging, but this is due to several specific factors where Auckland is different from other New Zealand cities: (1) Auckland is geographically and meteorologically more complex than Christchurch, (2)

PM_{10} concentrations and the number of exceedences in Auckland are not as great, (3) exceedences in Auckland occur not only due to home heating, but also a significant number occur on moderately – or even very – windy days and are thought to be due to dust and sea spray, and (4) in Auckland a principal contributor is vehicle emissions rather than the home heating emissions that dominate in Christchurch.

The ultimate aim of this work is to try to incorporate future predictions of New Zealand weather from climate models and assess what might be the effects on air quality for several years into the future.

B8.2 Speculation

Given the results shown in Figures B16, B17 and B18, it is tempting to conclude that there is evidence of a systemic reduction in PM_{10} emissions in the Christchurch airshed. The data may well indicate this, but the results need to be interpreted with caution, since (a) there is substantial variability and the statistical significance of the relationship has not been calculated, (b) the weather relationships are crude, omitting important factors such as inversion extent and strength, and (c) some secondary drivers have not been accounted for, such as fuel prices or particularly cold, or warm, periods.

Appendix C: Detailed report on the complex regression analysis of PM_{10} and weather

C1 Introduction

This appendix serves as an elaboration of the analysis presented in Section 5 of the main report. The following section provides a detailed step by step overview of the methodology used, followed by a schematic presentation of the methodology in Section C2. The final section presents additional figures which were not incorporated in the main report.

C2 Detailed overview of methodology (step by step)

1. Calculation of evening averages of PM_{10} and weather variables (5 pm – 12 am) from hourly data.

2. Transformation of PM_{10} concentrations using the natural logarithm ($\log_e$) and of wind speed using the square root (sqrt).

3. Creation of subsets of $\log_e$ PM_{10} and meteorological variables (sqrt wind speed, temperature at 1m, temperature difference between 1m and 10m) for each year.

4. Separation of each year into summer and winter subsets (November – April and May – October, respectively).

5. Regression of $\log_e$ PM_{10} against temperature for each seasonal subset (Figure C1).

6. Removal of correlation via recalculation of $\log_e$ PM_{10} values through application of following formula:

$$corr.\ log_e\ PM_{10} = (((log_e\ PM - (a + b \cdot TEMP)) / 2) \cdot \sqrt{3})) + avg.\ PM$$

with $corr.\ log_e\ PM$ = natural logarithm of temperature corrected PM_{10} concentration, $log_e\ PM$ = natural logarithm of raw PM_{10} concentrations, $TEMP$ = 1m air temperature, $avg.\ PM$ = mean raw PM_{10} concentration for

each season, a = intercept of the calculated regression and b = slope of the calculated regression.

7. Rejoining of seasonal subsets into a complete series (Figures C2 and C3).

8. Removal of remaining temperature correlation by repeating step 6 (Figures 5.3 and C4).

9. Multiple linear regression using temperature corrected $\log_e PM_{10}$ as dependent variable and square root wind speed and temperature difference as independent variables.

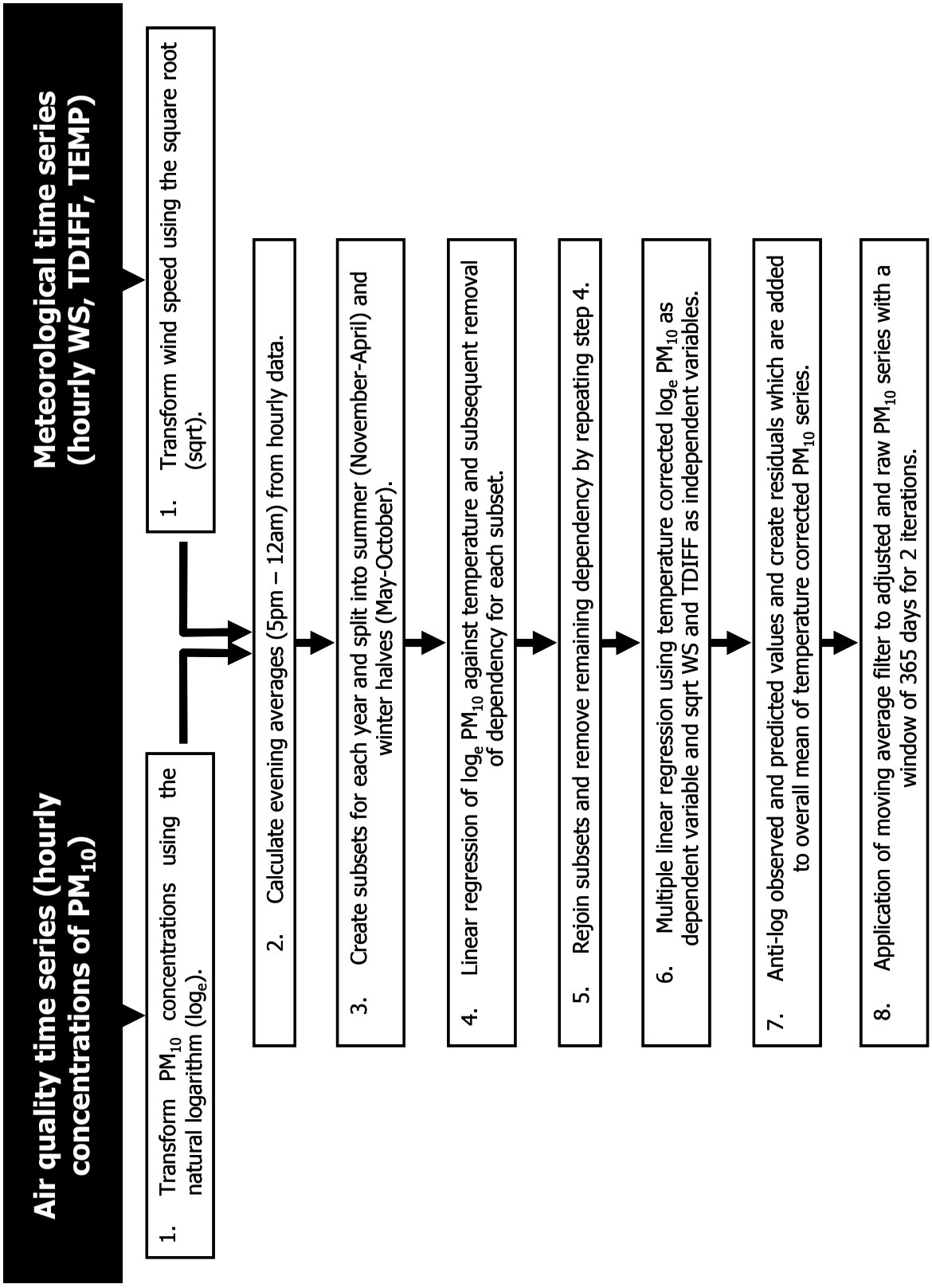
10. Anti-log of observed and predicted values and calculation of residuals (observed minus predicted).

11. Adding of residuals to overall mean of temperature corrected PM_{10} to produce adjusted series (Figure C5).

12. Application of moving average filter with a window of 365 days for 2 iterations (Figure 5.4).

13. Creation of annual bar graphs for comparison with results from other techniques (Figures C6, C7 and C8).

C3 Schematic presentation of methodology



C4 Figures displaying results (as referred to in Section 5 of the main report)

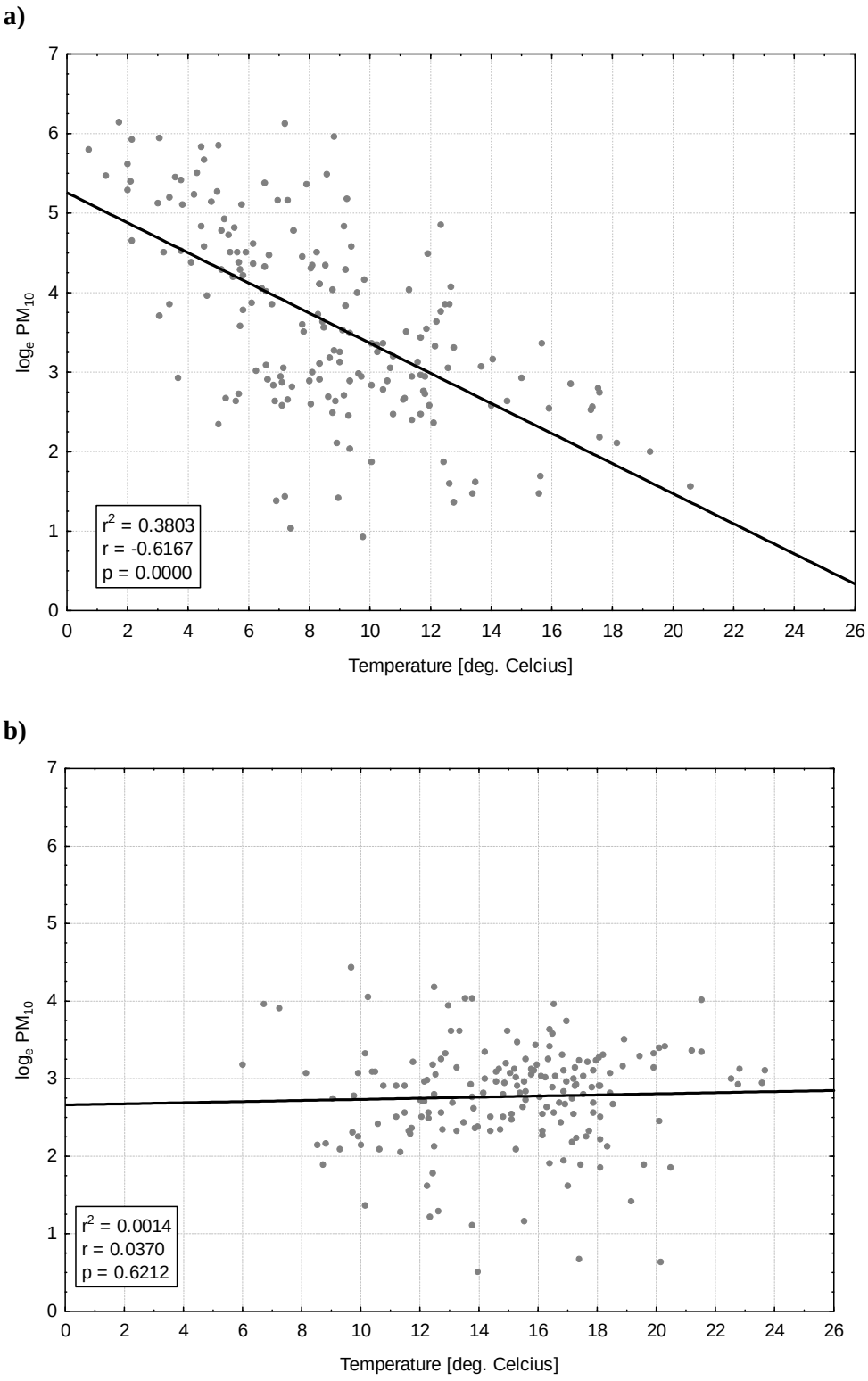


Figure C1: Log$_e$ PM$_{10}$ as a function of temperature at Coles Place, St. Albans in a) winter, and b) summer 1999.

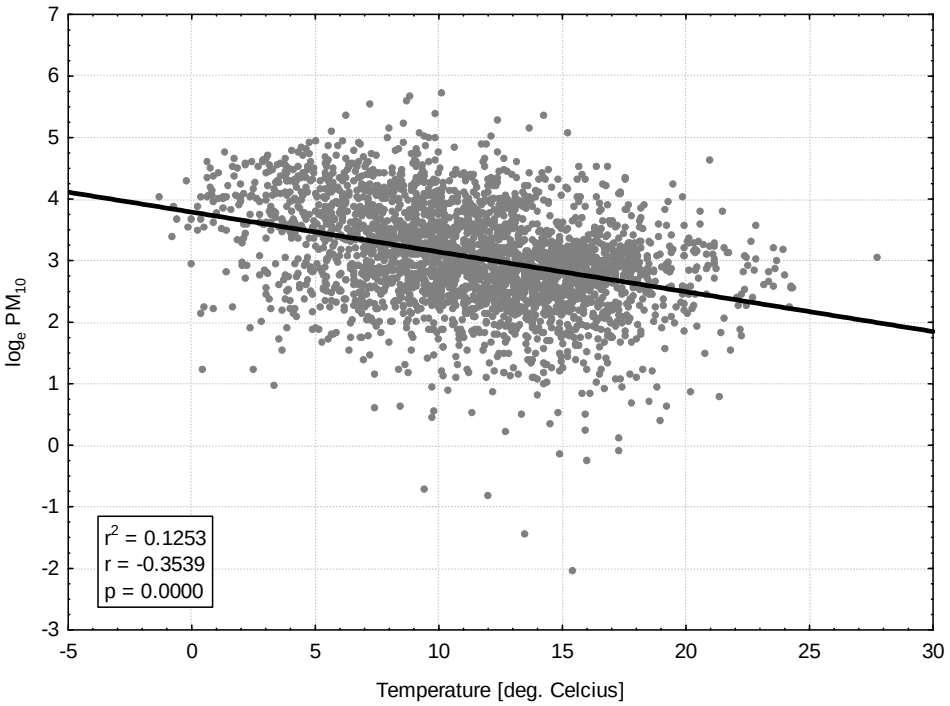


Figure C2: Scatter plot of rejoined corrected time series for 1999-2006 showing remaining temperature dependency.

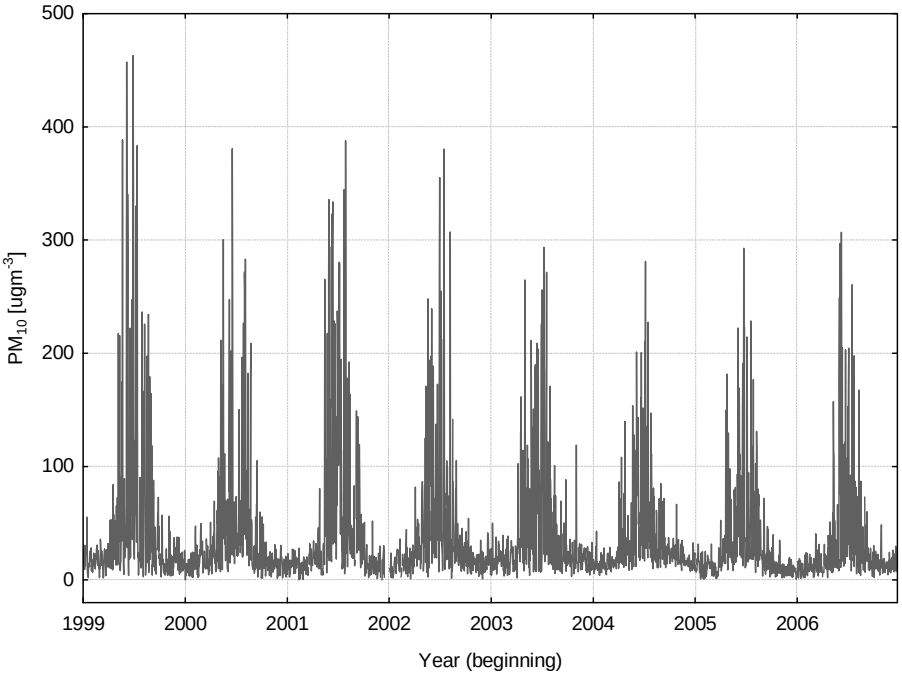


Figure C3: Raw time series of PM_{10} concentrations for Coles Place, St Albans (1999 – 2006).

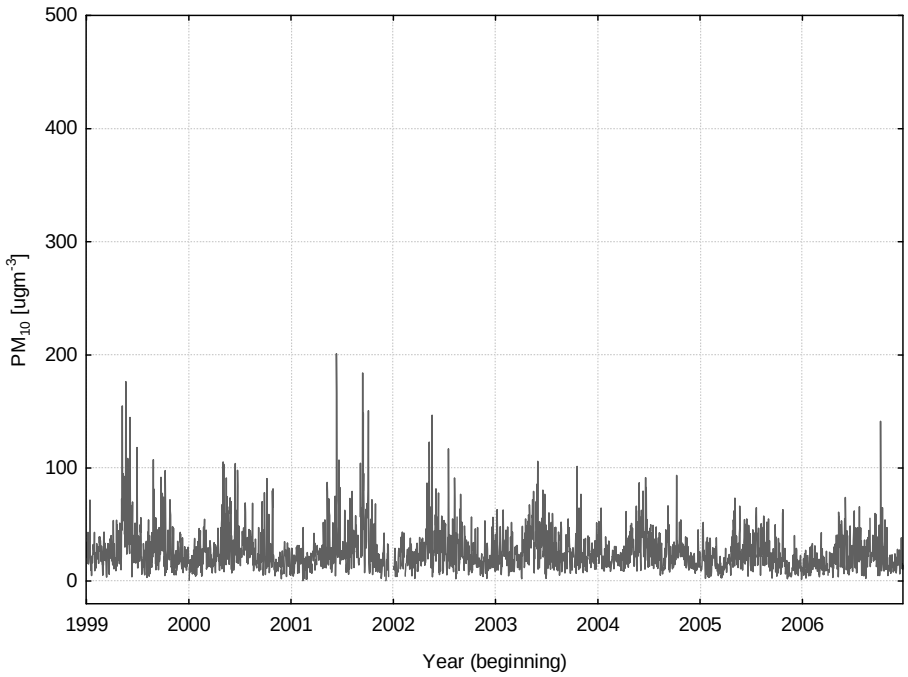


Figure C4: Temperature corrected time series of PM_{10} concentrations for Coles Place, St Albans (1999 – 2006).

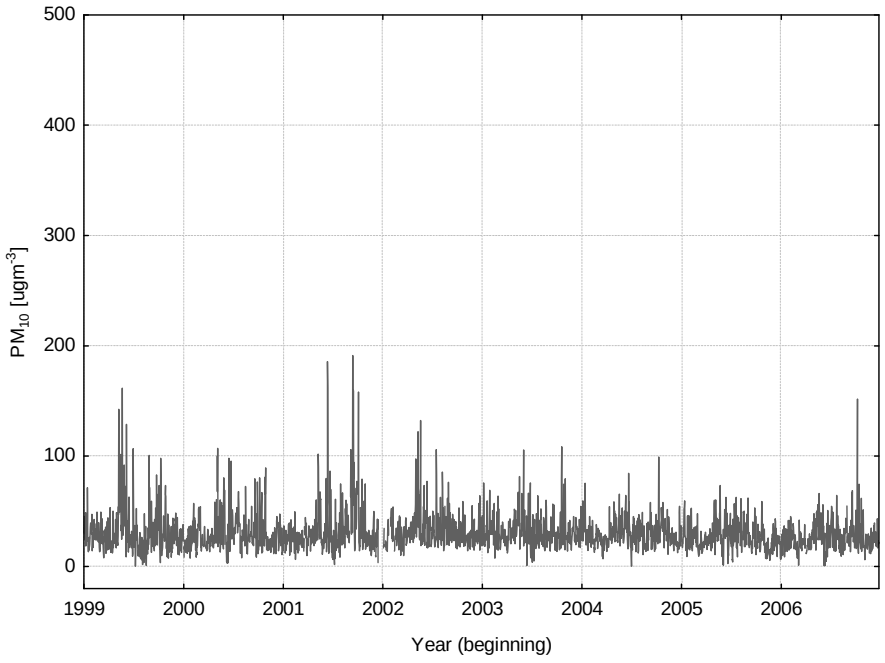


Figure C5: Meteorologically adjusted time series of PM_{10} concentrations for Coles Place, St Albans (1999 – 2006).

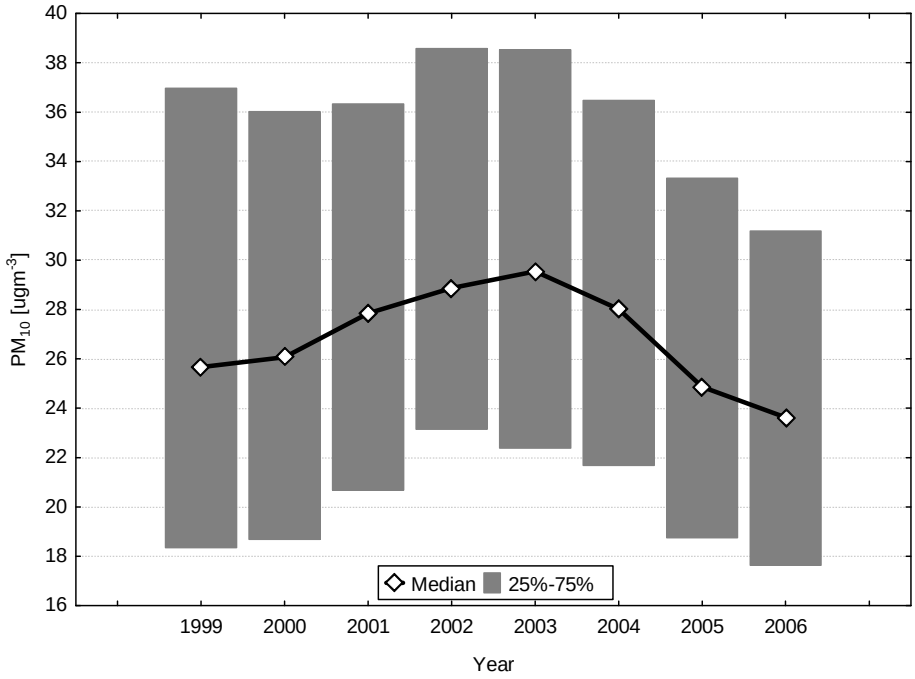


Figure C6: Box plot of the meteorologically adjusted time series of PM_{10} concentrations showing the median and 25% - 75% percentiles for each year.

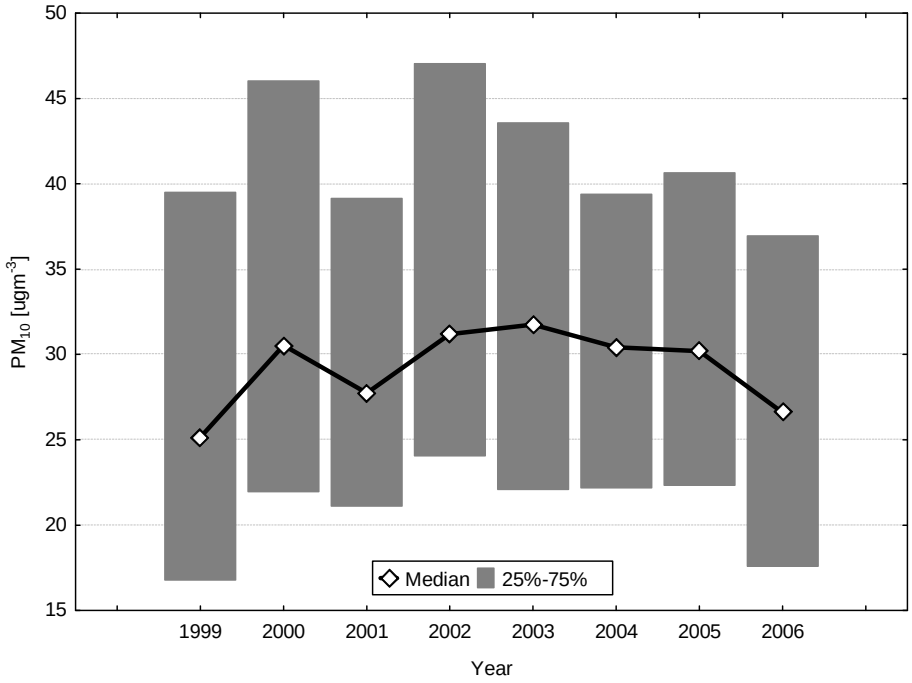


Figure C7: Box plot of the meteorologically adjusted time series of PM_{10} concentrations showing the median and 25% - 75% percentiles for each winter only (May – August).

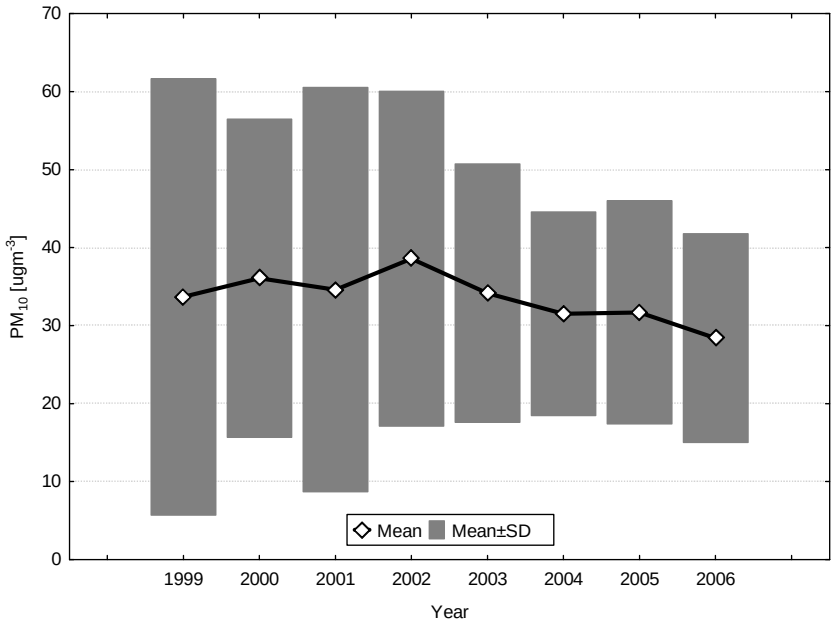


Figure C8: Box plot of the final adjusted times series of PM_{10} concentrations showing the mean and standard deviation for each winter only (May – August).