Chapter 4

Conclusions and Further Discussion

Various modelling approaches have been taken to estimate and then test for statistical evidence of trends in the daily average midnight to midnight PM_{10} concentrations. For all of the models considered (which have different underlying assumptions) statistical evidence for a (decreasing) trend has been found at over the 1% significance level. As you will likely be aware evidence for statistical significance does not necessarily mean the result is of any practical relevance. However, ample evidence has been found for the practical significance of these trend estimates as well, see for example Figures 3.5, 3.19 and 3.21.

4.1 Will Christchurch Meet NES in 2013?

Section 3.6 investigated the effect of extrapolating various forms of the trend estimates to 2013, when the NES specifies each airshed should observe at most one exceedance of $50\mu g/m^3$ per year. **The following discussion assume the estimated trends will be solely due to the emission changes**.

It is very clear from Figure 3.23 that if ECan does nothing further in terms of emission reduction strategies (based on the information provided in the sample of data), then Christchurch City will not reach the NES target, on the presumption that the past PM_{10} concentrations and meteorology are representative of that possible in 2013. The moderate case in Figure 3.23.2 essentially considers if no more reduction in emissions are enacted, then it is also very unlikely the target will be met in 2013.

The most optimistic case in Figure 3.23.4 essentially considers what happens if the reductions in emissions since the proposed Coal Ban Hearings and Clean Heat policies continue to give the same rate of reductions through to 2013, even then Christchurch City is very unlikely to meet the NES target by 2013. Despite the subjectivity involved in the trend extrapolations, sufficient evidence is provided of a need for further emission reduction strategies/policies to be considered.

Despite the model not directly considering the impact of emissions on the daily average midnight to midnight PM_{10} concentrations, any reduction in emissions will lead to smooth variation in the PM_{10} concentrations which is captured by the smooth trend models considered in this project. Information on the emissions and/or solid fuel burner conversions/removals could be incorporated into the model to further understand the impact of emissions on the concentrations (and therefore projected impacts of emission reduction policy changes), however this data was not available at a fine enough resolution to provide indicative results for this project. Future work could consider incorporating this information into the models considered, or how they vary in relation to the estimated trends.

4.2 Dangers with Smooth Trends

The evidence presented for the trend significance are based on the assumption that the models are correct, or at least a reasonable approximation to the truth. Despite the moderate complexity of the model considered in this project, we know that "All models are wrong, but some are useful" - George Box. We have constructed numerous statistically and physically sensible models, which all provide a consistent story, which capture all the dominant features (e.g. effects of meteorology on mean and variance, autocorrelation and smooth trend) of the variability in observed PM_{10} concentrations provided from 1995-2007. We therefore expect the models will provide a reasonable approximation to the truth, and reliable trend estimation and testing.

However, the user must be aware of all the assumptions underlying each model which are outlined and discussed in detail within Chapter 3. A key assumption is the idea of a slowly varying or smooth trend in the PM_{10} concentrations (either in the mean on the log scale for the GLM/GAM/GAMM models, and for each quantile in the quantile regression model). The slowly varying trend estimate will pick up all sources of smooth variation in the PM_{10} concentrations that is not explained by the meteorological terms in the model.

Causal explanations of the estimated smooth trend must be carefully considered. Certainly the emission reductions from policies like Clean Heat and other EECA schemes will contribute to the smooth trend. However, any other causal sources of smooth variation will also contribute to the estimated trend.

Any slowly varying socio-economic (and other anthropogenic) factors will contribute to the smooth trends. As we know the representation of the meteorological impacts on PM_{10} will be wrong (but will hopefully be a good approximation). Unfortunately, the slowly varying nature of climatic movements (including the possibility of climate change) may also contribute to the smooth trend estimate. Although the climate variation will be represented in the meteorological variables, as the model for their impact on the PM_{10} concentrations is only an approximation to the truth, any lack of representivity could show itself as a smooth variation due to longer term climatic variation/change.

Slowly varying changes to the measurement equipment (including maintenance, slowly going out of calibration, etc.), vegetation around the site, local emission sources/airflows will also contribute to the trend estimate. There will likely be other sources we have not listed as well of course.

The features of the trend estimates (i.e. peaks, troughs and decreasing trend) do not seem to correlate strongly with the site/machine sources of the hourly PM_{10} concentrations used. The dominant sources are:

- PST50 - 1995-1997;
- PST30 - early winter 1998;

- CPT30 - late winter 1998;
- CPT40 - 1999-2003;
- CPFDM - 2003 onwards;

which do not seem to be strongly related to the features of estimated trends:

- increase 1995-1997;
- peak 1997-1998;
- general decrease 1998 onwards;
- smaller peaks 2001-2 and 2004-5;
- sharp decrease from 2006.

This provides some further confidence that the crude calibrations of the measurements between the sites/machine in Chapter 1 have not unduely influenced the trend estimates. Strong confidence was also built when the models were rebuilt using only the data from Coles Place since 2000 (i.e. from only two machines TEOM FDMS and TEOM40) and providing consistent trend estimates for this period. For example, compare Figure 3.10 to Figure 3.11.

4.3 Consistency With Previous Results

The Canesis approach by Marsh and Wilkins (2004) considered only 'qualifying evenings'. The trend estimate for the meteorological corrected PM_{10} concentrations for their chosen thresholds is given in Figure 11 of their report, which is reproduced in Figure 4.1. The Canesis approach is rather different to ours, as it had slightly different goals to this project as it aims to develop an index for domestic heat sources only, by focusing on qualifying evenings. Further it is difficult to compare the consistency of the yearly trend estimates, as their yearly estimates vary substantially depending on the thresholds chosen. Canesis only had available data from 1996-2003.

It seems most appropriate to compare their 'normalised' trend estimates with our estimates. The increase from 1996-7 is consistent with our estimates. The decrease in 1998 from Canesis is not consistent. The peak in 1999 is strongly inconsistent with our estimates. The general decline, with slight peak in 2001 is consistent with our approach. The overall decrease from 1996-2003 observed by Canesis (for almost all thresholds is consistent with our overall decrease in this time.

Figure 4.2 reproduces the key trend estimates from the previous approaches of NIWA in Appelhans *et al.* (2007) given by Figures S2, S3 and S4. Our estimates are consistent with those presented in Figure S2, with only our estimate for 2005 being relatively larger than NIWA's.

Figure S3 shows the results from a simple log-linear model and appear to show the years with highest PM_{10} concentration down to smallest: 1999, 2001, 2000, 2005, 2003, 2006, 2004

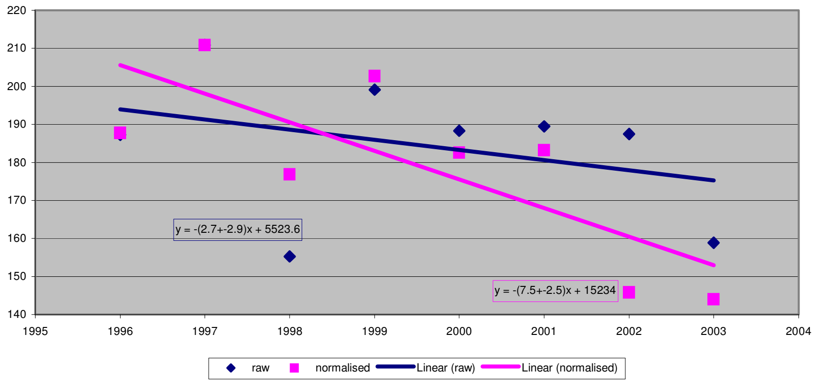

Figure 11. Raw (blue) and normalised (pink) average PM10 values over the years 1996-2003 for the Reference Scenario. The linear trend is steeper for the normalised PM10 values. The errors are quoted at the 1-sigma level, and are errors of the linear fit only (ie, the standard deviation of Figure 10 is not taken into consideration).

Figure 4.1: Figure 11 reproduced (with permission from ECan) from Marsh & Wilkins (2004) report from Canesis.

and 2002. Our trend estimates from the GAMM+AR(1) model exhibit the highest to lowest PM_{10} concentrations: 1999, 2001, 2000, 2002, 2005, 2004, 2003 and 2006. There are clearly quite a number of similarities between the estimates from our models and the log-linear approach by NIWA. The key differences are the estimates for 2006 and 2002.

In Figure S4 it makes sense to only compare their 'Adjusted' values with our trend estimates. The peak in Figure S4 in 2001 followed by a decline is consistent with our estimates. However, we have found strong evidence for a further peak in 2004-5 which is not observed in the NIWA Figure S4. The low value in 2000 followed by an increase in 2001 is consistent with our trend estimates.

The lack of fully consistent results with the Canesis and NIWA approaches is not surprising given they have focussed on a rather different subset of the data (even if you ignore the previously highlighted concerns around their approaches). We have focused directly on the daily average midnight to midnight PM_{10} concentrations, as this is the quantity of direct interest for the concern to meet the NES.

The strong consistency within the estimated trends using the broad spectrum of modelling approaches we have taken (including the initial quantile regression approach, which is allowing the data to speak for themselves with few model constraints) and the rigor of the methodology from a statistical and physical viewpoint leads us to feel confident in the conclusions drawn.

4.4 Are They Overly Complex Models?

In determining the selection of terms to be included in the model we have considered not just statistical testing criteria but also the physical relevance of the terms (in discussion with ECan staff). The extensive exploratory analysis and testing confirm the usefulness of the terms included. We have also guarded against overfitting in the statistics used for the model choice.

Figure 4.2.1 *NIWA Figure S2*

% exceedences by year in nodes 16, 26, 17, 11 27, 9 and 23

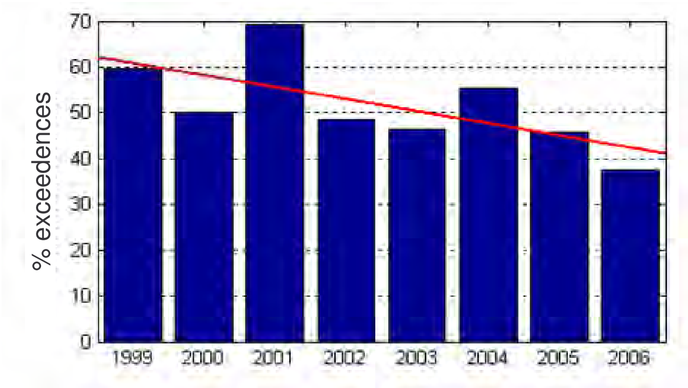

Figure S2: Regression tree analysis results showing the percentage of similar winter days experiencing an exceedence of the 24-hour PM_{10} standard.

Figure 4.2.1 *NIWA Figure S3*

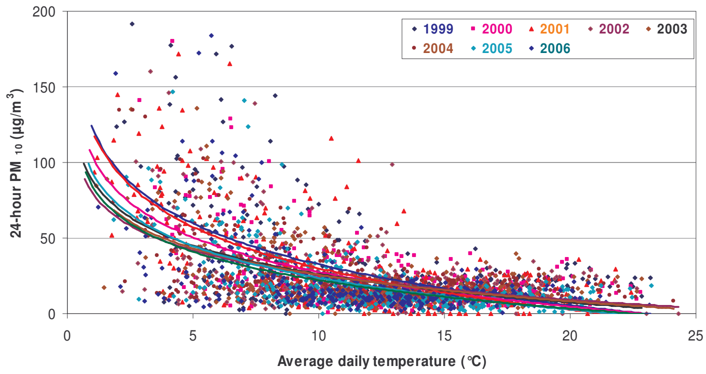

Figure S3: Yearly trends in 24-hour average PM_{10} with respect to average daily temperature (1999-2006). The curves fitted to the more recent years tend to be lower, indicating lower air pollution for a given air temperature**.**

Figure 4.2.1 *NIWA Figure S4*

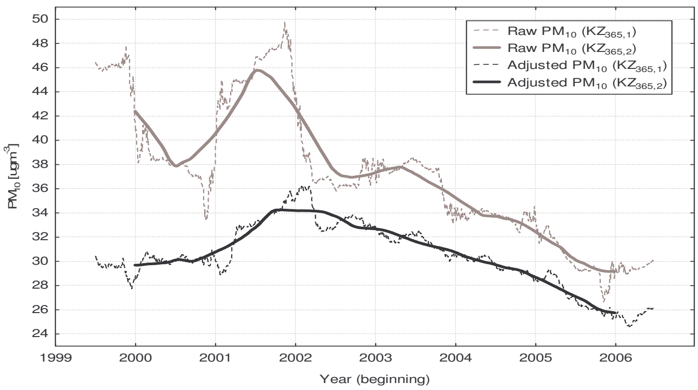

Figure S4: Smoothed trend lines of the average 5-12pm PM_{10} concentrations with the influence of weather variables removed.

Figure 4.2: Figure S2-S4 reproduced (with permission from ECan) from Appelhans *et al.* (2007) by NIWA.

In the final models we have included 21 meteorological terms (plus the intercept) thus using a total of 22 degrees of freedom. We have considered various trend models with different numbers of degrees of freedom (in the range 7-12 degrees of freedom being used), with the final GAMM model with AR(1) component (in some sense the best model considered) having smooth terms with an equivalent number of degrees of freedom of 7 (plus an extra one for the AR(1) correlation parameter), gives a total degrees of freedom used up of around 30. Given that we have around 1500 observations used to fit the model (after leaving out days with missing values), this is a very small number of degrees of freedom being used. Especially given the model manages to explain close to 70% of the variability in daily average PM_{10} concentrations (using R^2 performance statistic as typically used in standard linear regression models).

From a physical viewpoint we are also not surprised as to the number of terms needed. The PM_{10} concentrations are believed to be predominantly due to home heating based emissions. There are many factors that go into a householder's choice as to whether to start the solid fuel home heating. Clearly the temperature of the house will be critical, which will be influenced air temperature and amount of sun received inside the house during the day (cloud base and amount). The colder temperatures outside will correlate with other meteorological conditions (e.g. no rain, higher humidity). For a given set of emissions, other meteorological variables will influence the PM_{10} concentrations, e.g. wind speed and temperature differential (inversion). Hence, all the meteorological terms included in the models, are not only statistically but also physically justifiable.

4.5 Further Work

Various choices and approximations have been made in producing the trend estimates (including test of significance) and forecasts of exceedances in 2013. The choices have been made to produce physically and statistically justifiable estimates in a reasonable timescale. In the ideal world fine tuning of these estimates by further investigations and extensions of these models would be considered. However, the decisions that have been made to rationalise the modelling required within the available timescales have been made to ensure the final results are robust and reliable. This section aims to highlight some further areas for continuing research in this application.

4.5.1 Raw Hourly Data

We have taken the approach of pooling together the PM_{10} concentrations from various sites/machines to produce 2 constructed daily average PM_{10} concentration series (including series provided by ECan). Many questions were raised in Chapter 1 during the processing of this data and calibration to Coles Place TEOM FDMS equivalents.

Firstly, in Figure 1.2 we observed substantial differences between the hourly measurements from different sites/machines. The large differences could potentially have a string influence on the daily average concentrations. Therefore, future work should investigate these potential outliers further.

We have protected against a substantial influence from these potential outliers in this work by considering refitting the model with over 100 potential outliers and influential points in Section 3.4, from which no substantive change in conclusions drawn were found.

4.5.2 Calibration

There is still an outstanding question as to the appropriateness of the calibration models of PM_{10} concentrations and meteorological variability between the sites/machines. In this report we have utilised a simple linear regression based calibration technique.

The principle reason why simple linear regression is not formally appropriate (so will provide only a reasonable approximation) is that standard linear regression estimation methods assume the explanatory variables (e.g. Coles Place TEOM40 in this case) are measured with no error, so that the only source of error is in the response variable (e.g. Coles Place FDMS in this case). In this application, both the explanatory and the response variables are subject to error, which is often referred to as an 'errors-in-variables' problem.

The measurements are also correlated (e.g. correlation in time) with each other and the variance of them does change over time (often known as heteroscedasticity). These features violate the assumptions underlying simple linear regression when using ordinary least squares to fit the model.

Many alternative techniques (and formally more appropriate) are available e.g. orthogonal regression (appropriate when variance of errors in explanatory and the response variables is the same) or geometrical mean regression (sometimes known as standardised major axis regression, or reduced major axis regression). However, these techniques are not easily adapted to situation with correlations in the measurements over time. Further, inference for these models is rather challenging, especially given the PM_{10} measurements are far from normally distributed.

However, these more advanced techniques frequently lead to similar calibration equations, and visual inspection of those found using simple linear regression appear to be appropriate.

However, it is also clear from the research of Bluett *et al.* that a deeper investigation of how the calibration may change over time and in relation to site changes and meteorological conditions should be considered. It also became clear in Figures 1.3 and 1.5 that our calibration equations and those used by ECan using previous subsets of the daily data are substantially different, providing further evidence for the need to investigate how the calibration equation my change over time or in relation to the meteorological conditions.

We have protected against issues arising from the calibration equation by comparing the results from all 3 constructed series, which use different sources of data and calibration equations. Further, in Section 3.4 we also examined the trend estimates when only data from Coles Place since 2000 (actually a small amount of data from Packe Street machines could slip through from infilling the missing hourly measurements), from which the trend estimates were essentially the same.

A simple further analysis which would confirm consistency of the trend estimates in this report independently of the calibration issue, would be to apply the modelling approach to each machine/site PM_{10} concentrations series separately rather than try and combine them into one near complete trend estimate. This simple analysis could be extended to an advanced statistical modelling approach to avoid the calibration problem. The data from each site/machine could be kept as they are (not calibrated) and construct a form of random effects or hierarchical model to explain the different impact of the meterological/trend impacts on each series individually, but allowing pooling of information from the all sites/machines together. This approach would save the issue of having to calibrate them in advance. This latter advanced approach is a PhD level research problem.

4.5.3 Which of three series is most reliable?

It is clear from Chapter 3 that the FDMBAM and ECANCP series, gave very similar trend estimates to each other. Further, as noted in Section 3.4 after around 100 potential outliers and influential observations were removed from these two series, the trend estimates were left almost unchanged. This provides some confidence in the reliability of the daily averages in these two series. Given that the ECan provided daily PM_{10} concentration data (ECANCP) has been used for numerous projects and manually checked for obvious outliers by the dedicated ECan staff, it is likely this particular series is the most reliable in terms of fewest outliers. It is not appropriate, of course, to comment as to the reliability of the different calibrations used for the three series, as this requires a much more substantial investigation as suggested in the previous section.

The FDMT30 series on the other hand had been substantially impacted by outliers, leading to some clear differences in the trend estimates. Section 3.4 showed that if these outliers were removed the resultant trend estimates were much more similar to those of FDMBAM and ECANCP. Therefore, this series is clearly the most unreliable as it stands. However, it has the potential to be the most reliable overall the calibration of the hourly measurements between sites/machines is based on the complete set of overlapping measurements (and does not use the BAM measurements as used in the FDMBAM series which are known to be less reliable), whereas the ECan provided dataset only uses a subset of the daily observations for the defining the calibration equations. Many of the outliers in the FDMT30 series have been identified as due to missing hourly measurements, so if these could be remedied then this series could be made more reliable.

Of course in the ideal world, the safest option is to always consider all 3 series in tandem, to compare the consistency of the results.

4.5.4 Missing Values

A further issue for future consideration is the use of imputation/interpolation methods to fill-in the missing PM_{10} concentration and meteorological measurements. There are many pre-existing methods for imputation which could be appropriate. Further work should investigate the usefulness and robustness of these various techniques.

We have presumed the missing measurements are missing completely at random, and if this is the case they will not have an impact on the trend estimates and test.

An ongoing issue which is known to have adversely affected the FDMT30 daily average PM_{10} series, is the missing measurements within a day. Substantial biases in the daily averages can be introduced depending on which hourly measurements are missing. We observed in Figure 3.10 (and estimates from other models) that the trend estimates using the FDMT30 series were substantially different to the for FDMBAM and ECANCP, which we believe is predominantly due the issue of missing hourly measurements. After exclusion of outliers (many of which were based on less than a complete set of hourly measurements within a day) in Section 3.4 the trend estimate using the FDMT30 series became much more similar to those from the FDMBAM and ECANCP series.

Future work should consider either interpolation/imputation of the missing measurements, or exclusion of these daily measurements based on too few hourly measurements. In particular, a focus of manual inspection of the possible outliers should focus on the very large differences in the 3 constructed series highlighted in Figure 1.5 and those which are exceedances of the NES target level of $50.5\mu g/m^3$.

4.5.5 Modelling Assumptions

The GLM/GAM/GAMM models (normal family with log link) considered in this report allow the mean of the distribution of measurements to vary in a log-linear form with the meteorological and trend components. The variability of the error distribution is assumed to be constant, i.e. does not increase with mean PM_{10} level. As discussed in Section 3.1 alternative family distributions (e.g. gamma), and possibly link functions, could also be appropriate. Further investigation of these models could be worthwhile. The gamma family was considered and applied to the ECANCP series, with similar conclusions drawn about the trend as using the normal family in this report, so was not considered further.

Future work should consider whether the shape of the distribution changes possibly as a function of the meteorological variables and trend terms. However, our subjective impression from the quantile regression results in Section 3.5 is that it is only limited there is evidence for this, although no formal test has been carried out. If this approach is taken there are various statistical models which could be used, e.g. vector GAM's (VGAM's) or advanced quantile regression models.

An extreme value approach which would focus in on just the high exceedance days may also be a suitable approach. However, these are relatively more complex, closer to research type models suitable for Masters or PhD level study.

4.5.6 Explanatory Variables

In the quantile regression modelling in Section 3.5 we assumed that all the meteorological terms useful for predicted PM_{10} concentrations in the GLM/GAM/GAMM models are appropriate for modelling all quantiles. This assumption is not necessarily appropriate, in fact other meteorological terms may be more appropriate for high quantiles versus low quantiles. A formal model selection for the quantile regression model would be appropriate if the results from this model are to be used for anything over than an exploratory technique. Recent research has also considered measures of dependence between observations in the tails of the a distribution (i.e. high exceedance days). We considered a lag autocorrelation in the GAMM, but have not considered more complex tail dependence. However, the time available for this project did no permit examination of these measures of tail dependence.

In Chapter 2 we explored the usefulness of lagged meteorological variables for determining PM_{10} concentrations. Further exploratory work (unreported) has confirmed the statistical relevance of various lagged terms, however these would need further investigation and validation of their practical and statistical relevance. Chapter 2 also highlighted the need to consider possible meteorological interaction terms.

The wind direction was not included in the modelling in the report. There are many issues which would need to be resolved if the wind direction is to be utilised. Firstly, how do we calibrate the wind directions between the Packe Street and Coles Place sites? What daily statistics of the wind speed are appropriate to consider? A crude standard deviation of the wind speed was considered in selecting terms for the GLM/GAM/GAMM models, but was not selected for the final model. Clearly as the wind direction is a circular statistic an alternative circular measure of average/variability would be more appropriate. Expert knowledge of the meteorology of Christchurch and how the wind direction may provide information on the weather and therefore impact on PM_{10} concentrations should also be considered.

Given that the majority of PM_{10} emissions for high exceedance days will be due to home heat sources (in the evening), it may also make sense to include variables based on the evening meteorology, or possibly that at 17:00-18:00. The reason being that it is this time that people return home from work and make the decision whether to start the solid fuel burner or not. Once started, most homeowners would continue its operation all evening. However, if it is slightly warmer when they arrive home, they may make the decision to not start the sold fuel burner. However, in the timescales available for this project it was not possible consider these evening related statistics.

References

Appelhans, T., Bluett, J., Dey, K., Sturman, A. & Wilton, E. (2007). Using air quality data to track progress toward PM_{10} standards: Case study Christchurch 1999-2006. NIWA CHC2007-135.

Bluett, J., Wilton, E. Franklin, P., Dey, K. Aberkane, T., Petersen, J. & Sheldon, P. (2007). PM_{10} in New Zealand's urban air: A comparison of monitoring methods. NIWA CHC2007-059.

Marsh, C. & Wilkins, A.H. (2004). An alternative measure of air quality. Canesis, CLR 04/01.

Resource Management Regulations (2004). National Environmental Standards Relating to Certain Air Pollutants, Dioxins, and Other Toxics. SR 2004/309

Scarrott, C.J. (2008). Statistical review of two reports on particulate pollution in Christchurch. University of Canterbury.