FEASIBILITY STUDY INTO ACCURATELY MODELLING SULPHUR DIOXIDE FLUX IN THE LOWER ATMOSPHERE

by

GARETH PRITCHARD

A dissertation submitted to the University of Glasgow for the degree of Masters of Science

2009

Declaration

This thesis has been composed by myself and it has not been submitted in any previous application for a degree. The work reported within was executed by myself, unless otherwise stated.

Acknowledgements

This thesis could not have been written without the support and back-up of many people around me, who I would wish to thank below.

Firstly, I would like to thank my supervisor Professor Marian Scott for her never-ending patience and support throughout this lengthy writing process. Her contacts and knowledge throughout were immensely useful throughout and this thesis would never have come to an end if it wasn't for her extra drive and determination (and ability to give a kick up the backside if necessary!). I would like to thank both Ron Smith and Mhairi Coyle at C.E.H for their expert knowledge in the world of environmental data and their ability to make time for me to ask questions and to provide data and advice when necessary. I would like to thank Simone Giannerini for his knowledge of chaotic behaviour and the use of his program used to estimate Lyapunov Exponents as well. Finally, the staff in the Statistics department were also of great use throughout my time and I would like to say thank you to all of them.

On a more personal level I would like to thank the rest of the postgraduate students in my department. There are too many to list them all, but I would like to give special mention to Claire, Helen, Marco, Ally, Sarah, Jim, Eleni and James most of all for the friendship and support throughout my time in the department. The open and friendly atmosphere displayed by everyone throughout made working at the uni a fantastic opportunity and one that I do not regret.

Finally away from the department, I would like to thank my mum, dad and sister for their support, especially towards the end of the writing up. Thank you so much. Also to Billy, Scott, Daniel, Barry and Gow I would like to say thank you to them for all their friendship. I couldn't hope to have better friends, as well as the many, many people that would take far too long to mention. However, one last person who should get a name-check is my wonderful girlfriend Donna. She has been a constant support through the lengthy process of this thesis and her encouragement was of great help to me throughout.

Abstract

Atmospheric pollutants are of concern for both their effects on human health and on plants and crops. Since the 1960s monitoring networks have been created, linked to international protocols regulating emissions of pollutants such as sulphur and nitrogen and also to validation studies of large-scale atmospheric transport models. One such monitoring site in the UK is at Auchencorth Moss, close to Edinburgh, where routine half-hourly measurements of sulphur dioxide are made. The time series shows a large amount of variation, and it is of interest to explore any trend in the pollutant level along with any presence of seasonal and diurnal cycles and to draw comparisons with pollutant transport model predictions. However, before carrying out such analysis, it is necessary to investigate the sources of variation. This thesis will consider the nature of the calculation of the sulphur dioxide flux, based on three simultaneous concentration measurements corrected for stability height. The need to calculate a slope estimate based on three points led to some difficulties and these were looked at to see whether these were creating difficulty when it came to modelling the fluxes. It was concluded that there were a high proportion of fluxes calculated using slope estimates with high R^2 values and so any difficulty might lie in the actual data themselves rather than any technicalities in the calculations used to define the flux. From there, each variable involved in the calculation of the flux was studied, using approaches such as signal-to-noise ratios and sensitivity analysis. From these it was seen where most variation was occurring. Signal-to-noise ratio techniques did not work very well with the very low data measurements collected, which was disappointing but the values collected were generally very low suggesting a large level of noise in the data. Sensitivity analysis helped to show where most of the variation lay. Using a sampling based method it was shown that most of the variation lay in the gas concentrations themselves rather than any of the other variables involved in the calculation of the flux. This led to the conclusion that the gas concentrations rather than anything else were contributing to the difficulty of modelling sulphur dioxide fluxes. This suggested that there might be a possibility that there was no problem in the data collection approach or calculations of a flux, but perhaps the data itself was too variable to be modelled.

Chaos theory offers a different approach to the analysis of time-series and this thesis explores the use of the Lyapunov exponent to investigate chaotic behaviour over

different aggregated timescales. The chaos definition used was the popular "Sensitivity based on initial conditions" approach favoured by most people in this field. Looking at how quickly two data points placed very closely together could diverge after a certain time period would show whether any predictions made would be highly susceptible to any variation would be a very useful finding. Using three different techniques gave disappointing results however. The techniques all produced results which were sometimes conflicting with each other and none of which gave any convincing argument for, or against, the existence of chaos. This led to two potential conclusions. One being that the data were very noisy, but predictable underneath this, or methods of estimating chaotic behaviour can be flawed. This thesis also looks at how Extreme Value Analysis can be used on very noisy environmental time series and how useful it can be in explaining the behaviour of the larger values measured. In this study there were some large peaks in each of the years when looking at a time series analysis. These values were studied separately from the data using Generalised Extreme Value theory and the General Pareto Distribution. The Pareto distribution approach was concluded to give the better insight into the data. This was shown to model the extreme values reasonably well though both options could be taken as valid from these approaches. Finally the measured and modelled data (collected from a Europe-wide model) were compared and analysed to see how well they compare and what techniques from each of the previous analyses can be used to bring them closer together. These tended to show that the two data sets (modelled and measured) did not match up particularly well. Techniques such as a Bland-Altman analysis and many comparison diagnostic tests were analysed to see whether there were differences between the two. Even when some findings from earlier chapters were applied to the data such as applying a minimum R^2 to any slope estimates did not help.

Contents

Chapter 3 – Looking into Chaos

Chapter 4 – Extreme Value Analysis

List of Tables

List of Figures

Chapter 1 - Introduction

Motivation

Science has contributed to many different leaps forward in technology and thinking throughout time. However, as progress has been made, there have been many high-profile stories about the damage man has made to the planet. Scientific research can be used in order to quantify how much damage has occurred, and make predictions of what might happen in the future in order that governments and leaders can plan accordingly. This might be anything from the amount of food generated (see the OECD-FAO Agricultural Outlook 2008) to even population estimates for humans (see the UN Report, World Population to 2300, 2004)

In environmental science it can be very useful to be able to model natural and man-made effects that happen. This can be anything from the growth/shrinking of an animal population to the growth rate of a flower or plant. This allows scientists and other environmental analysts a greater degree of understanding of what goes on in these systems, and helps them to explain why events may be occurring. The main reason for modelling something though, is for the opportunity to estimate what might happen in the future. For high-profile issues such as global warming, populations close to extinction etc. the advantages of having an accurate model that can predict what may happen in 1,5, 10…years into the future can prove invaluable.

Scare stories about ice-caps melting and climate change are well documented in the press, and treaties like the Kyoto agreement are set up to try and reduce the amount of (mainly) CO_2 in the atmosphere. Being able to predict the levels of gasses in the atmosphere could therefore prove very useful. This is especially useful for working out where gasses are spreading to, since they can be carried long distances through the air and so there may be high levels in quite "random" places, away from power stations or other things that may cause these gasses to be emitted.

Another example of a treaty set up is the LRTAP Convention (Long-Range Transboundary Air Pollution). This was created in 1979 in order to try and protect the environment. This was mainly done by setting protocols, many with so-called 'critical values'. These are a threshold that companies/governments are not allowed to exceed when they produce harmful gasses such as SO_2 or NO_2 These two

chemicals in particular are very harmful to the environment, as they are known to be the main causes of acid rain, which can cause damage to vegetation, animals and humans (through building corrosion, poisoning water etc.)

Some of these gas levels are modelled by EMEP (Co-operative Programme for Monitoring and Evaluation of the Long-Range Transmission of Air Pollutants in Europe). This programme *"provides governments and subsidiary bodies under the (Long Range Transboundary Air Pollution) LRTAP Convention with qualified scientific information to support the development and further evaluation of the international protocols on emission reductions negotiated within the Convention."* (http://www.emep.int/emep_description.html)

The EMEP programme has three main elements. They collect emission data. They study environmental data and they attempt to model the data in order that future predictions can be made. EMEP also teams up with many scientists and scientific task forces in order that its results can be verified and checked.

The programme uses a series of modelled meteorological data along with modelled gas concentrations to predict a flux or rate of SO_2. These have been based upon data collected for over 40 years in some sites around Europe. From these data a model has been made and updated through the years, from one that modelled daily values for 100km squares through most of Europe, to a more improved one that now models values over a 50km square. The model used a Lagrangian method at first that was since updated to an Eulerian approach in order to model over the smaller areas (the new model is described at length in Bartnicki et.al (1998)). One thing to make mention of though is the volume of variables that need to be modelled in order to work out a predicted flux. There will still be some concern at how much a model can be expected to correlate with measured data over such a large area and this will be of interest to study and gain conclusions from.

Below shows a map of Europe along with a grid showing the 50km regions that northern Europe has been split into. The EMEP model predicts daily fluxes for each one of these squares. This is shown in Figure 1.1:

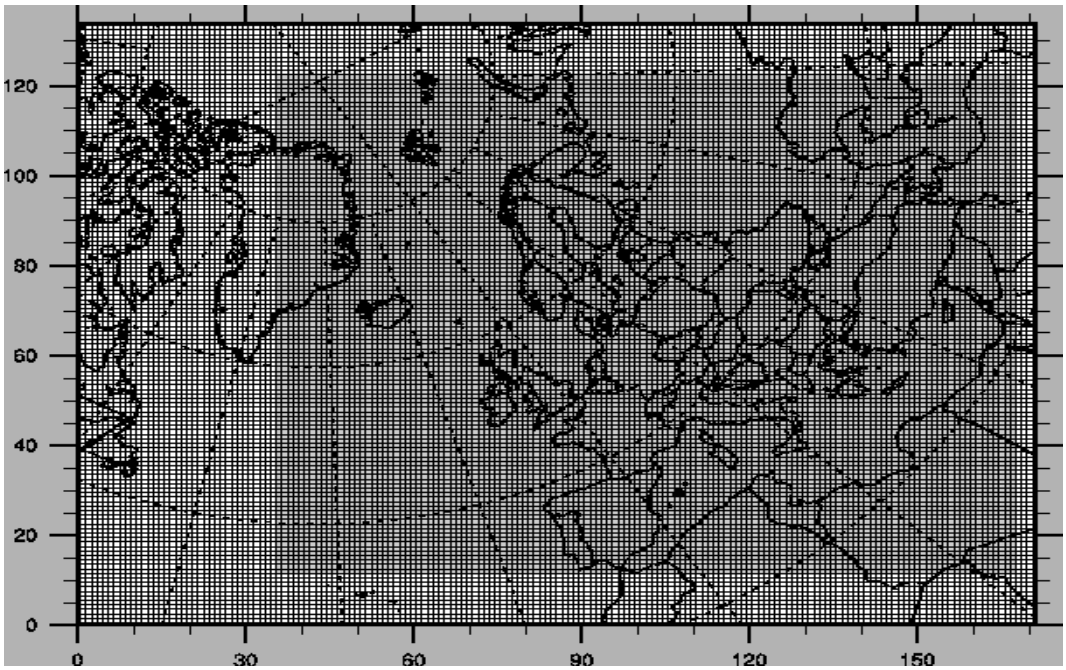


Figure 1.1: Domain of the EMEP model divided into the 50km 'squares' – the small squares are the grid cells

It can be seen that these squares look small, but since each of them are actually 50km by 50km it can be seen that it may be difficult to provide one value that may match the levels of gas concentration that are actually measured. Weather conditions for instance may be very localised and so provide differences between two sites in the same square.

The Centre for Ecology and Hydrology, a nationwide organisation that studies areas such as biodiversity, water, biogeochemistry, environmental informatics, climate change and sustainable economies, collects measurements of different chemicals routinely every half hour over different areas in Britain. Their research mainly focuses on understanding the world and importantly, the repercussions of human activity on the world. This thesis will use data collected and recorded at their station in Edinburgh.

Most of the focus will be on one particular site in Eastern Scotland called Auchencorth Moss. Different variables are taken from a measuring tower, including gas concentrations at different heights, and then calculations are made in order to produce a flux measurement (this will be described in more detail in Chapter 2). A flux is defined as the amount that flows through a unit area at unit time. This flux is then compared to the value that the EMEP model calculates it to be at that particular grid location. As mentioned already these data are modelled at a daily level, and so

the measured data will have to be aggregated up to daily levels in order to compare between them.

There is much literature around and many (simple and non-simple) techniques for immediately comparing two sets of data, however it would be useful to use techniques which help to give more information to explain more about any differences that may occurs between these two data sets, rather than just producing a simple p-value that shows the model validity/invalidity . For instance it will be useful to see if there are techniques for quantifying the quality of the data and if a model-measurement comparison can be improved in order that they match up more closely (and hence create a higher level of trust in future levels produced)

This thesis will attempt to show the difficulties in working with environmental data and produce techniques which, while used in different contexts, may not have been widely applied to environmental systems, in order to see if these explain some of the behaviour that can be seen in environmental time series. It will also be useful to look at some of the patterns and trends that occur when monitoring data.

1.2 Scope of Thesis

This thesis contains six chapters. In this chapter, a brief overview is given of the kind of data that will be worked with throughout the rest of the thesis, along with the main problems that may be reached in the course of attempting to model sulphur dioxide fluxes. Chapter 2 will look at the measured fluxes and look at them in more detail to see ways of assessing the quality of the data that has been collected in order that they can be compared to the model. From this chapter, conclusions will be reached about which years appear to give "better" results and this will become useful in later chapters – especially when the modelled and measured data are looked at together. This chapter will also contain a sensitivity analysis to see which particular variable(s) are having most influence over the calculated flux values.

In Chapter 3, the measured data is looked at in terms of whether it shows signs of chaotic behaviour. The most popular technique is to calculate Lyapunov Exponents.

These assess whether the data shows signs of being very sensitive to small changes, which make it impossible to predict what will happen in the future, without being certain of what value is being measured currently. Two ways of estimating these exponents will be looked at based on Giannerini and Rosa (2004) and Nychka et.al. (1992) respectively and compared to each other before concluding whether the data is chaotic or not.

With variable time series data it can also be useful to look at the more extreme values that may be measured. Chapter 4 will start looking more closely at these values that have been measured and will look to see if these show any pattern to them, by using classical Extreme Value Theory. Both the Generalised Extreme Value (GEV) family and the Generalised Pareto Distribution (GPD) will be looked at in order to see if the data fits either or both of them. These are the two most common way of analysing these sets of values and have been used in several other environmental studies. This will be helpful in helping to ascertain whether the values furthest away from the modelled data could still be explained.

Chapter 5 will bring all the conclusions from previous chapters together in order to see how closely the modelled and measured data that has been collected match up to each other. By using information gained from Chapters 2,3 and 4, these will all be applied to the data to see if there are any ways of validating the model using different subsets of the measured data. Also this chapter will discuss the spatial aspects of the modelled and measured data since the model is for a 50km by 50km square and it is being compared against just one measuring station.

Finally Chapter 6 will conclude all the results from the chapters and discuss the findings from this thesis and any future work.

Chapter 2 - Sulphur Dioxide Flux Measurements

2.1 Introduction

A flux is described as: "The rate of flow of fluid, particles, or energy through a given surface" (www.dictionary.com). In this study, this will be the rate at which SO_2 moves through the air. As mentioned in Chapter 1, the values of interest are those of flux due to *dry deposition.* The following picture helps to show exactly what dry deposition is (Figure 2.1)

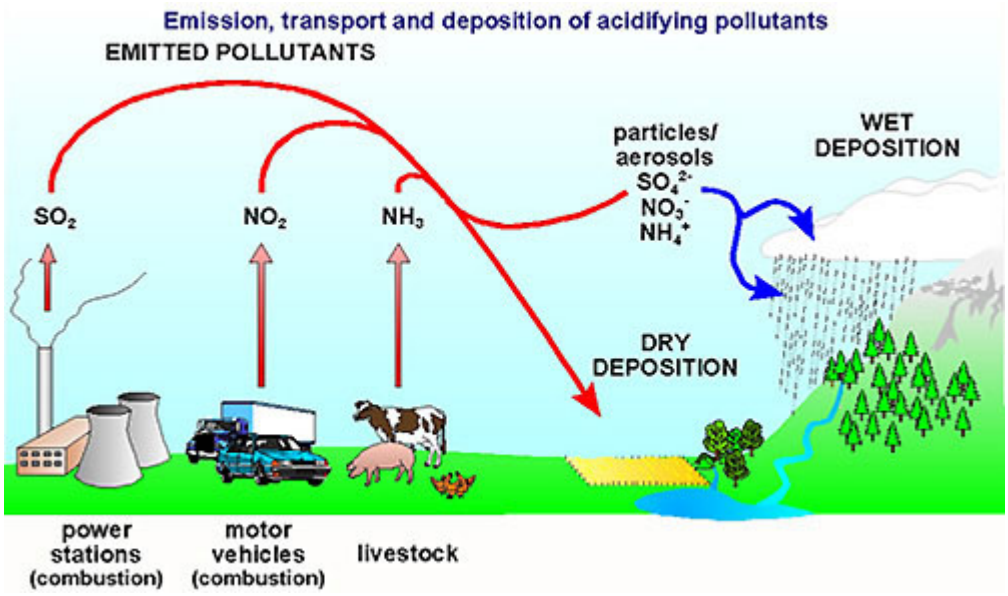


Figure 2.1: Picture taken from http://www.ceh.ac.uk/aboutceh/sections/edin_pollution.htm. This shows how dry deposition differs from wet deposition and how it is deposited.

Figure 2.1 distinguishes between dry deposition and wet deposition. Wet deposition is produced by rainfall, whereas dry deposition occurs from the transfer of a pollutant

to the surface by any other means (air currents etc.). SO_2 can be blown for many, many miles before it is deposited (either wet or dry), so it can be difficult to predict just how much will be measured in specific locations.

At Auchencorth Moss, meteorological and physical measurements are taken half-hourly every day from a measurement tower on the site. The tower is pictured below (Figure 2.2)

Figure 2.2: A picture of the measuring tower at Auchencorth Moss taken from http://www.ceh.ac.uk/aboutceh/sections/edin_pollution.htm.

The tower has devices to measure gas concentrations from three heights on the tower. Other things measured each half-hour include the air temperature, the wind direction and the wind speed. The friction velocity (a reference wind velocity using the air density along with the horizontal and vertical wind speeds) is measured by a sonic anemometer.

Since the calculation of a sulphur dioxide flux depends on a number of measured and theoretical variables, the properties of an SO_2 flux will depend on the attributes of these input variables. Analysing these parameters in detail will help to assess the uncertainty of the measurements made in comparison to the 'true' value that exists in the environment at that particular time and in the model comparison.

Once a sensitivity study has been completed, an important step will be to ask questions about the 'quality' of the fluxes and methods will be shown to deal with the problems that could occur. By applying certain statistical procedures to the data, the data quality can be measured in quantative ways. Goodness-of-fit tests will also be useful since the flux calculation relies on a slope estimate from three points (this will

be expanded upon in pages 23 and 24). Other techniques will also be used to analyse potential problems that may affect the model-measurement comparisons that will be shown later on in Chapter 5.

Further analysis of the measurements will include temporal aggregation at different scales - from the half-hourly measurements into longer time periods such as days and months - to see if there are patterns in these longer data series. Also analysing differences between day and night could help to see if the growth of plants underneath the measuring tower in the canopy, affects the quality of the gas concentrations that are measured. Additionally, looking at the distribution and spread of the gas concentrations and the heights at which the measurements are taken (once they have been stabilised- this is described below) will help to show again if there appears to be a reasonable level of consistency. It would probably be expected that the heights should not change very often and so if they are, then it may help to ask why.

This chapter underpins much of the statistical analysis in Chapters 3 and 4, and where the measured data are compared to the modelled in Chapter 5. Also some of the later techniques involved in identifying a signal and assessing variation will be useful when looking at the chaotic behaviour that may be present (explained more in Chapter 3)

2.2 Estimation of the Flux

In order to estimate an SO_2 flux at Auchencorth Moss, certain variables need to be calculated. The technique used at this site is the 'eddy correlation' method (Monteith & Unsworth 1990). This has the advantage of measuring a flux directly, rather than other techniques which infer it rather than measure it. Two of these techniques are discussed by Griffith and Galle (2000) and Leuning et.al (1999). The multi-stage technique used here is described below.

The first step is to calculate a length L (the Monin-Obukhov length) using the formula

$$L = \frac{(-u^*)^3 c_p \rho T}{kgH} \qquad (2.1)$$

- u* is the friction velocity. This is necessary in finding out how much of the turbulence measured is caused by wind and not heat flux**.** If u* is measured at less than 0.08 however it is treated as missing as the wind speed is regarded as too low and measurements will be too uncertain. Friction velocity is measured by a sonic anemometer.
- ρ is a constant air density (=1246 gm^{-3}),
- c_p is another constant (=1.01). This is a basic property of matter. It is the quantity of heat required to raise the temperature of 1kg of matter by $1^oK.$,
- T is the ambient temperature (in Kelvin) averaged over two heights on the measuring tower.
- k is von Karman's constant, a constant of proportionality (0.41),
- g is the acceleration due to gravity (9.87ms^{-1})
- H is the rate of heat transfer per unit area. This explains how much turbulence is caused by heat radiating from the surface. This is measured at the tower.

L is routinely filtered to remove any unreasonable values. If the absolute value of L falls below 2, then the value is rejected and treated as missing. The reason for this is that the atmosphere has become either too stable or unstable at this point for the micrometeorological methods to be applied (Monteith and Unsworth 1990). This, as well as the u* filter mentioned earlier are the only two filters applied.

The second step in calculating the fluxes involves calculating the stabilised corrected height (SCH) using the formulae listed below in (2.2) to (2.6)

$$\psi_{Ha(i)} = \frac{-5.2(gasht_i - d)}{L} \qquad (2.2)$$

$$\chi_i = \left(1 - \frac{16(gasht_i - d)}{L}\right)^{\frac{1}{4}} \qquad (2.3)$$

$$\psi_{Hb(i)} = 2\log\frac{(1+\chi_i^{\,2})}{2} \qquad (2.4)$$

$$z\psi_{H(i)} = \begin{cases} \psi_{Ha(i)} \ldots\ldots L > 0 \\ \psi_{Hb(i)} \ldots\ldots otherwise \end{cases} \qquad (2.5)$$

$$SCH_i = \log(gasht_i - d) - z\Psi_{H(i)} \qquad (2.6)$$

$$i = 1,..,3$$

- gasht is the heights that the gas concentrations were measured at in metres above the ground; these are constant over a year but can change at the beginning of a new year depending on the height that the foliage underneath the tower may be expected to reach.
- 'd' is a constant, which is worked out as approximately 70% of the canopy height (Campbell 1977). This is used as a measure of the 'zero plane', i.e. the height at which the wind speed is zero. (Monteith and Unsworth p.113-117)
- SCH values are the Stabilised Corrected Heights (in metres above the ground)

Before the flux calculation can be made, one more filter is applied to the data. The heights and concentrations should only be used when/if the wind is not being interfered with by the measuring hut. When the wind is blowing directly over the measuring hut then this will mean that an unreliable measurement could be taken and so this means that any measurements, collected while the wind direction is between 60 and 170 degrees, are rejected.

The $z\Psi_{H(i)}$ values (from 2.5) are used to calculate the SCH's. These are necessary in order that the wind-speed profiles are more or less linear. In micrometeorology it is assumed that the atmosphere is stable and so the wind-speed profile and log(height) can be related linearly. In a lot of real cases though this cannot be automatically assumed and so the profiles are linearised to account for this. The SCH's are calculated as shown in equation (2.6).

Finally in stage three, the flux can be calculated. Each gas concentration (at each height) has a corresponding stability corrected height (SCH) as mentioned previously. At each time point, there are therefore three concentrations and 3 SCH's. This is then treated as a regression problem using gas concentration as the response variable and the SCH as the explanatory variable. From these three points, a least squares estimate is used in order to fit the best linear model for each set of three 'points' as shown in equation (2.7)

$$gasconc_i = m(SCH_i) + c + \varepsilon_i \qquad (2.7)$$

- m being the gradient or slope of the best fit line
- c being the intercept

i = 1,..,3

Figure 2.3 shows this pictorially. The three points are shown as the small circles and the piecewise lines drawn in bold. The best-fit line has been drawn on as a dotted line.

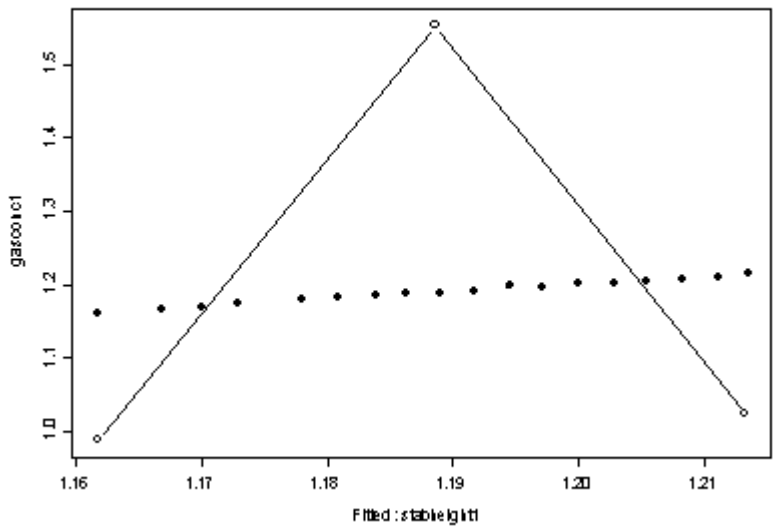


Figure 2.3: Plot showing gas concentrations against stabilised heights, for one half hourly measurement from 1996 with a dotted line showing the best-fit line through them. Solid lines indicate the piecewise lines.

Now, finally the flux value can be calculated as shown below (2.8)

$$\text{Flux} = -k \times (u^*) \times m \qquad (2.8)$$

Where k and u* are as before in equation (2.1) and m is the gradient of the best-fit line from (2.7). The product of these three quantities provides a flux measurement for SO_2 levels.

Possible problems with the quality assurance of the data could be:

1. Although filters have been applied to three of the variables (u*, Wind Direction and L), are these taking out all unreliable measurements, or are they taking out too many values leaving a data set too sparse to analyse properly? Are these missing measurements all occurring at one time or over more "random" time periods?
2. Can a regression model be trusted to be reliable when it is only based on three points? If values at one of these heights have been poorly measured, this would surely give a poor fit and therefore an untrustworthy slope measurement, meaning a poorly calculated flux value.
3. There are many different variables (measured and theoretical) used to calculate the flux measurement. How much variation will each of these produce in the flux?

The first problem is probably of least concern that needs to be looked at in more detail. In order to check this though, some brief analysis of the gas concentrations that are removed will be carried out, in order to see if these show any considerable differences to the filtered data. This will be in Section 2.5. Section 2.6 will show a chi-squared analysis to analyse whether there are any differences in the amount of missing fluxes at particular half hourly periods. Because the concentrations are measured and not derived they should show if the filter applied to L, u* and Wind Direction will make any difference.

The second and third problems will be looked at in more detail in sections 2.3 and 2.4.

2.3 Evaluating the Best Linear Model

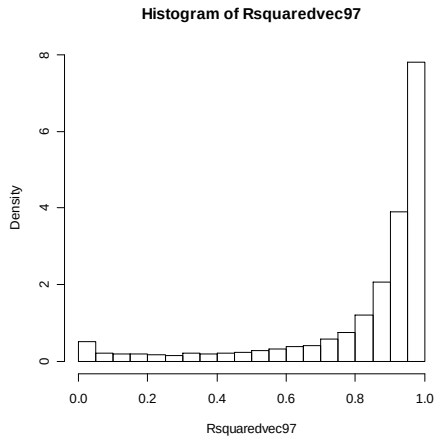
With only three points to fit a best-fit line, there is a worry that a particular flux could be influenced by one outlying measured value. If the three points do not look like they may lie on a straight line then perhaps a flux measurement should not be calculated since the flux is directly related to the gradient of the best fit line.

2.3.1 Analysing the Goodness-of-Fit of the Straight Line

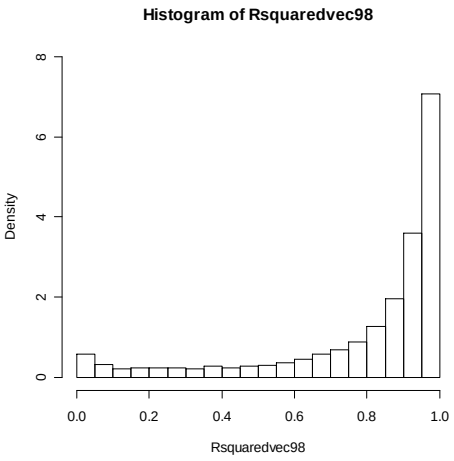
For each of the years, the R^2 value for each best-fit line has been evaluated. This has been performed on the half hourly data so that the data can be checked for quality at their measured format before they are aggregated to the hourly or daily measurements. The 17520/17568 (depending on whether it is a leap year or not) R^2 values for the half-hourly data are shown in Figure 2.4:

Year	Min	1st Quartile	Median	Mean	3rd Quartile	Max	Numbers missing
1997	0.000	0.800	0.925	0.823	0.975	1.000	3213
1998	0.000	0.738	0.913	0.795	0.971	1.000	2986
1999	0.000	0.587	0.829	0.720	0.950	1.000	3233
2000	0.000	0.740	0.967	0.797	0.978	1.000	2840
2001	0.000	0.289	0.709	0.605	0.927	1.000	3126

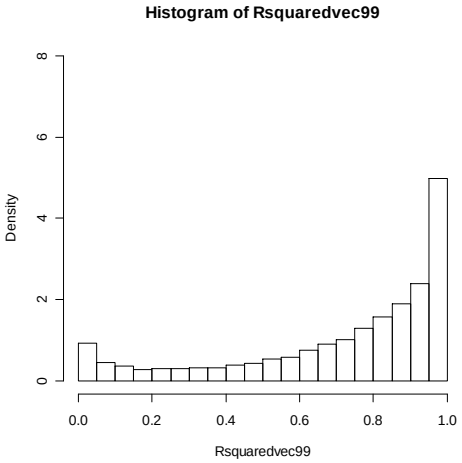
Table 2.1: The summary statistics for each year of the half hourly R^2 values. The missing values include the filtered out values.



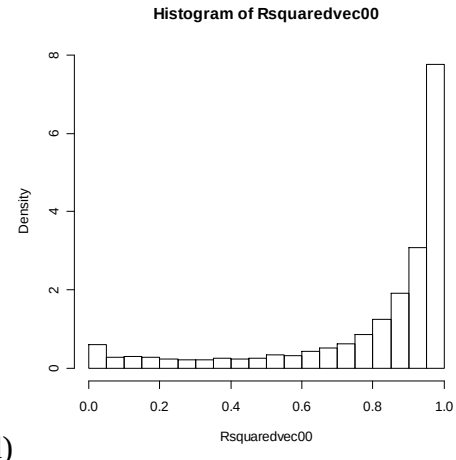
(a)



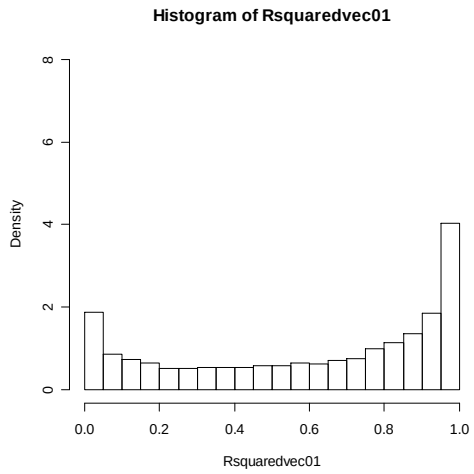
(b)



(c)



(d)



(e)

Figure 2.4: Density histograms of R^2 vs. a probability frequency for the years, (a) 1997 (b) 1998, (c) 1999 (d) 2000 and (e) 2001. The y-values multiplied by the bar-widths (0.05) sum to 1.

At first glance the histograms all seem relatively similar and show that the three points are providing a reasonable straight line in most cases (over 75% of values for 4 of the five years are showing values above 50% and over half the values for all years are above the same figure). Certainly 1997-2000 all look reasonably similar from the graphs above. However, it should be noted that 1999 is slightly different and perhaps should be analysed with a little more care as there are more lower R^2 values than in other years. However, over three quarters of the data gives a better than 50% R^2 value here (from Table 2.1) so there is still reason to believe that the best fit line (and hence the flux measurements that come directly from it), is fitting the three points well in most cases.

1999 however does show a few, very low R^2 values. These are more noticeable than in the other years (ignoring 2001 for now) when looking at Figure 2.4. These could be scrutinised to find out when they are occurring. For instance, if they are all occurring together in time, then perhaps the machine had a fault in it that day, and those results should maybe be discounted. If they are scattered then perhaps this could just be put down to a single measurement error and this shouldn't cause too many problems especially when averaging them for a daily value (however this will be looked at).

In 2001 however, there does seem to be slightly more of a problem. Certainly looking at the 1st quartile shows that a quarter of the R^2 data points fall under 30%. This would appear to be a problem as these low values affect at least 25% of the flux calculations. When comparing the modelled and measured data later on-this may have to be thought about if the comparisons prove to be worse for the 2001 data.

2.3.2 R^2 simulation

A filter on the data points could possibly be something to think about. This would allow an acceptance of a flux value only when the best-fit line was "good enough" (below is a discussion of what may be seen as acceptable). With three points, a potential method might be to look at the middle value and use that, to see if the three points are 'collinear enough' to believe a straight-line model could be fitted to the data. Figure 2.5 shows pictorially how this could be applied.

It could be useful to consider a simulation study, in order that three points could be chosen and then varied in order to see how much that would change an R^2 value and to assist in interpreting an 'acceptable' R^2 level as an additional filter.

.

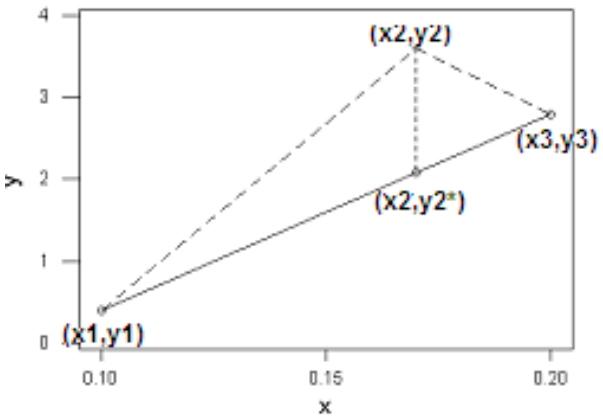


Figure 2.5: Plot using simulated data showing three points (x_1, y_1), (x_2, y_2) and (x_3, y_3). Solid line is line connecting (x_1, y_1) to (x_3, y_3) with a point marked at the x_2 co-ordinate. Dashed lines are the piecewise lines.

Figure 2.5 shows the basis of a simulation that was applied using three arbitrary points. Firstly three points were chosen that lay in a straight line (these are (x_1, y_1), (x_2, y_2*) and (x_3, y_3) in Figure 2.5), then the 2nd point (i.e. y_2^*) was moved perpendicular to the x-axis, and R^2 values were calculated from the best fitted line for (x_1,y_1), (x_2, y_2) and (x_3,y_3). These were then plotted against the standardised differences between y_2 and y_2^*. The differences were standardised by dividing the vertical distance between (x_2, y_2) and (x_2, y_2*) by y_2* (=1.6). For example, the standardised distance on the second line (0.15625) was obtained by dividing 0.25 by 1.6. The results obtained are shown in Table 2.2 and Figure 2.6.

y_2 value	Distance from (0.15, 1.6)	Standardised Distance	R^2
1.6	0	0	100%
1.85	0.25	0.15625	98.56%
1.93	0.33	0.20625	97.54%
2.1	0.5	0.3125	94.53%
2.35	0.75	0.46875	88.48%
2.6	1	0.625	81.20%
3.1	1.5	0.9375	65.75%
3.6	2	1.25	51.92%
4.1	2.5	1.5625	40.87%
4.6	3	1.875	32.43%
5.1	3.5	2.1875	26.07%
5.6	4	2.5	21.26%

Table 2.2: The simulated distances of the second point from the perfect fitted line and the resultant R^2 that it produces.

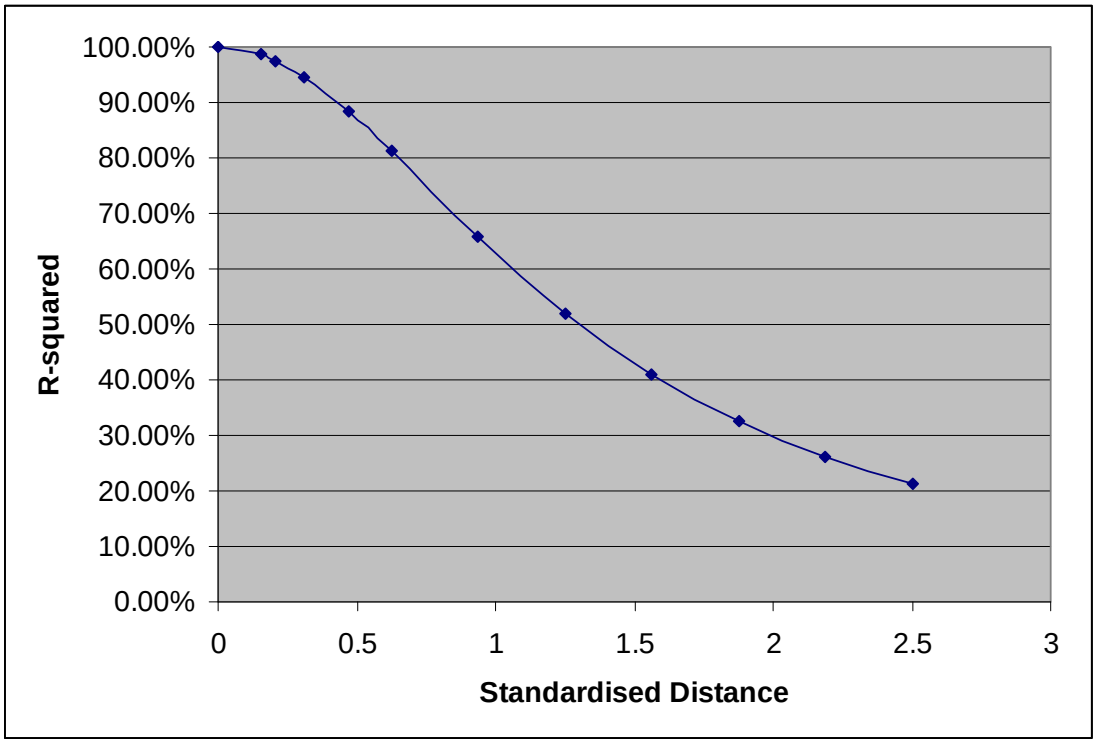


Figure 2.6: Plot of R^2 values against corresponding standardised distance

Figure 2.6 shows that a "middle value" with a standardised distance of 2.5 will reduce the R^2 value to nearly 20%. This may prove to be a reasonable threshold so that any very poorly fitted flux values will not be calculated. It could of course be made even stricter. Table 2.2 and Figure 2.6 show how far away the middle point would be in these cases.

Some of the real measured values were used to see what sort of distances (and distance*) were being produced. An example is shown in Figure 2.7. This shows the last half hourly set of values taken from 1997

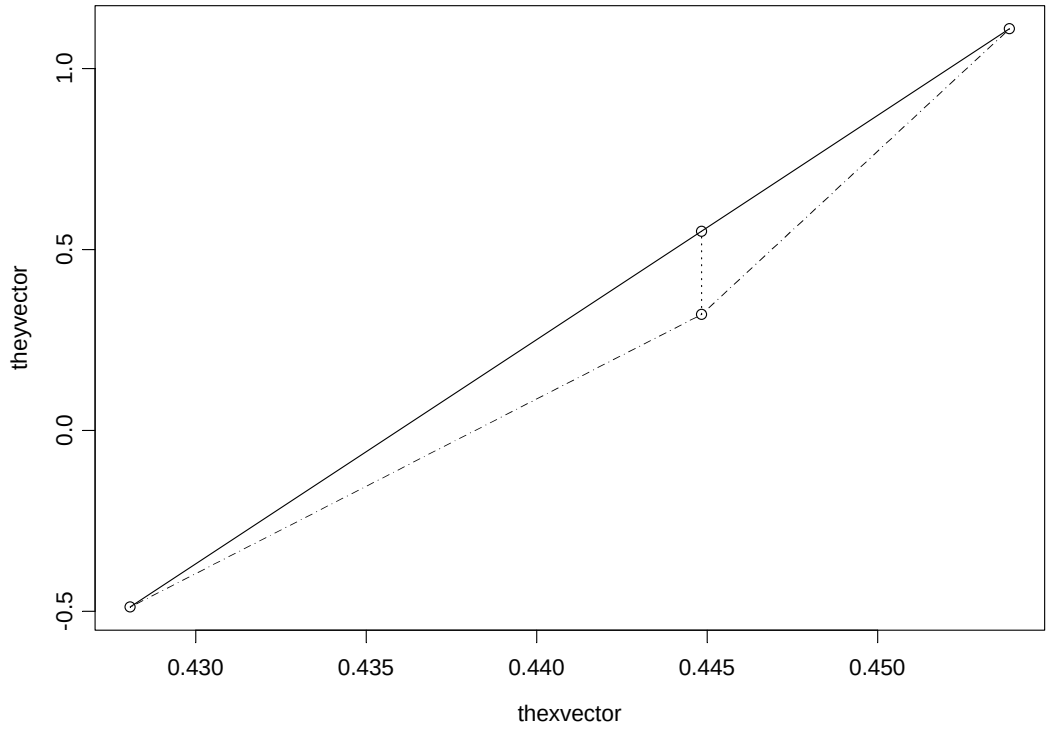


Figure 2.7: Showing a real set of three values from 1997, along with the distance between the real middle point and where the middle point would be if the first and last points were connected

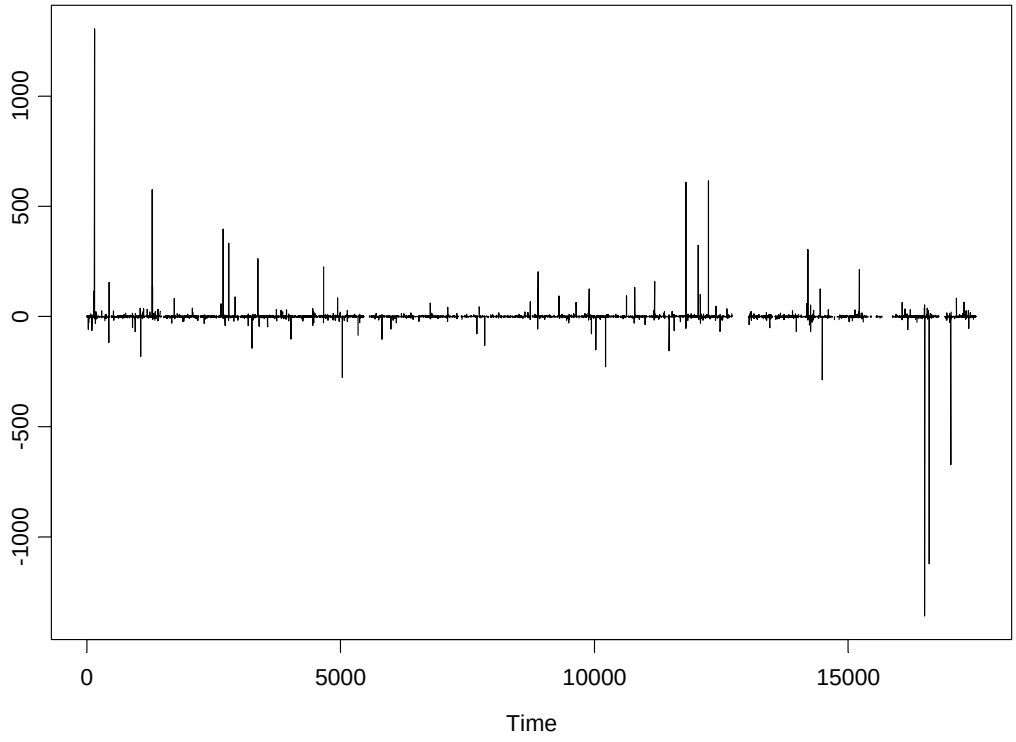


Figure 2.8: Time Series showing the Distance* values throughout 1997, using the half hourly values.

Distance*	No. of Distance* Smaller	% of Distance* Smaller	R^2 at that Distance*
0	0	0	1
0.15625	623	4.36	0.9857
0.20625	804	5.63	0.9754
0.31250	1185	8.30	0.9453
0.46875	1731	12.13	0.8848
0.62500	2319	16.25	0.8120
0.93750	4247	29.75	0.6575
1.25000	9324	65.32	0.5192
1.56250	11563	81.00	0.4087
1.87500	12334	86.40	0.3243
2.18750	12712	89.5	0.2607
2.50000	12968	90.84	0.2126

Table 2.3 The summary statistics of the Distance* from the 1997 half hourly data.

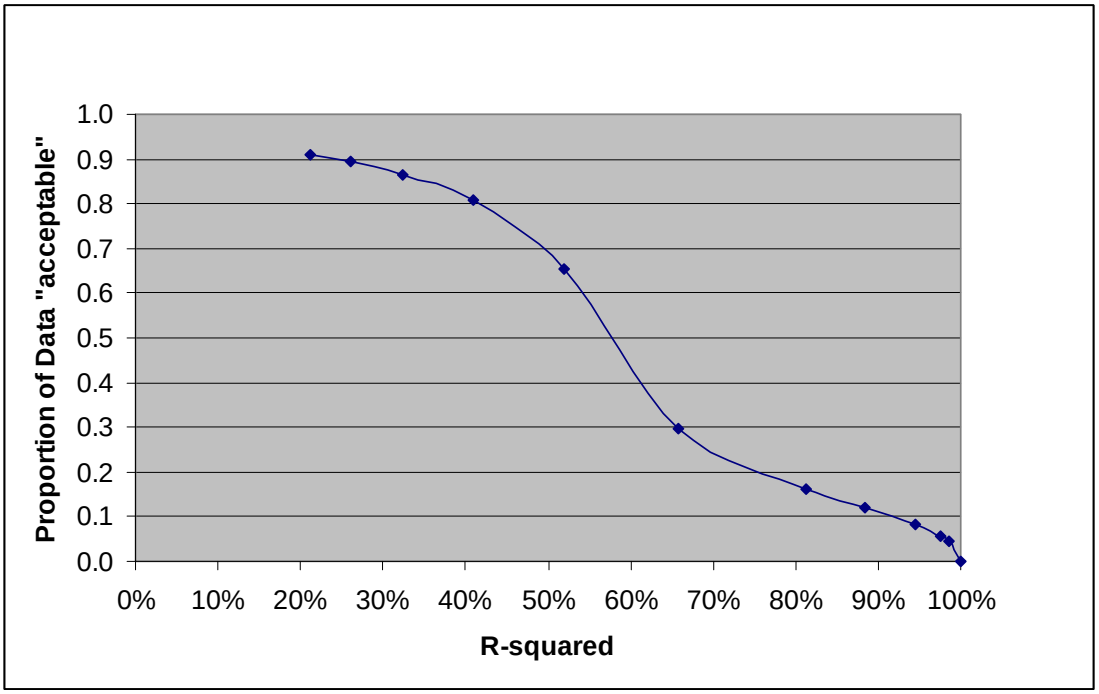
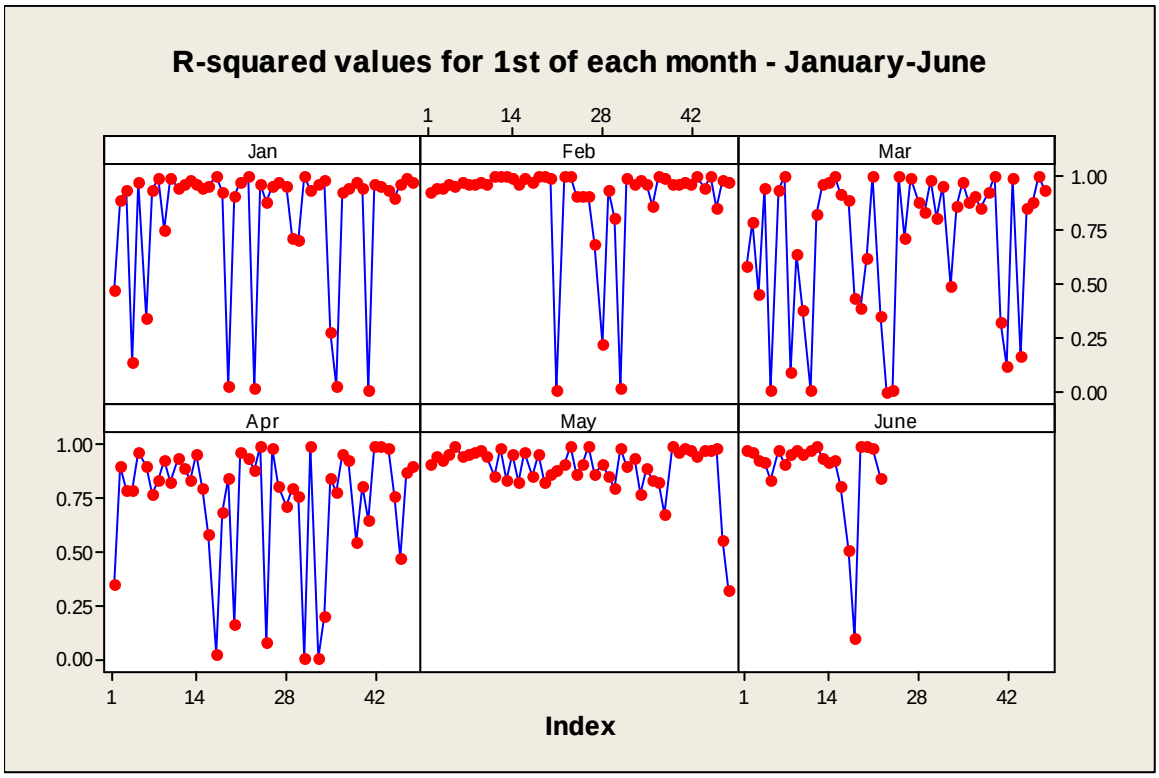


Figure 2.9: Using Table 2.2 to plot the percentage of data points with a lower distance than at each threshold (Col 2 of Table 2.2) and the corresponding R^2 value (Col 4 of Table 2.2)

Table 2.3 shows that there are still approximately 65% of available data that have an R^2 of over 50% and in fact over 90% of the points give a better than 20% R^2 figure as was mooted as a potential threshold on page 30. Were the bar to be set at 30% (say), then there would be between approximately 86 to 90% of the data points included. This doesn't appear to be a great difference in the number of rejected values. In fact Figure 2.9 shows that the small differences in the percentage of data rejected as the R^2 value decreases from around 30-40%. Therefore the 30% value will be used later on in Chapter 5 when the modelled and measured data are compared against each other.

2.3.3 R^2 values Throughout the Day

The R^2 values may vary throughout the day. Perhaps winds are strongest at particular points during the day and so may be carrying more of the pollutant. Since SO_2 comes from factories, maybe there will be certain points of each day in which there is more pollutant in the air. In order to look at this Figure 2.10 shows the half hourly R^2 values from the first day of each month in 1997.



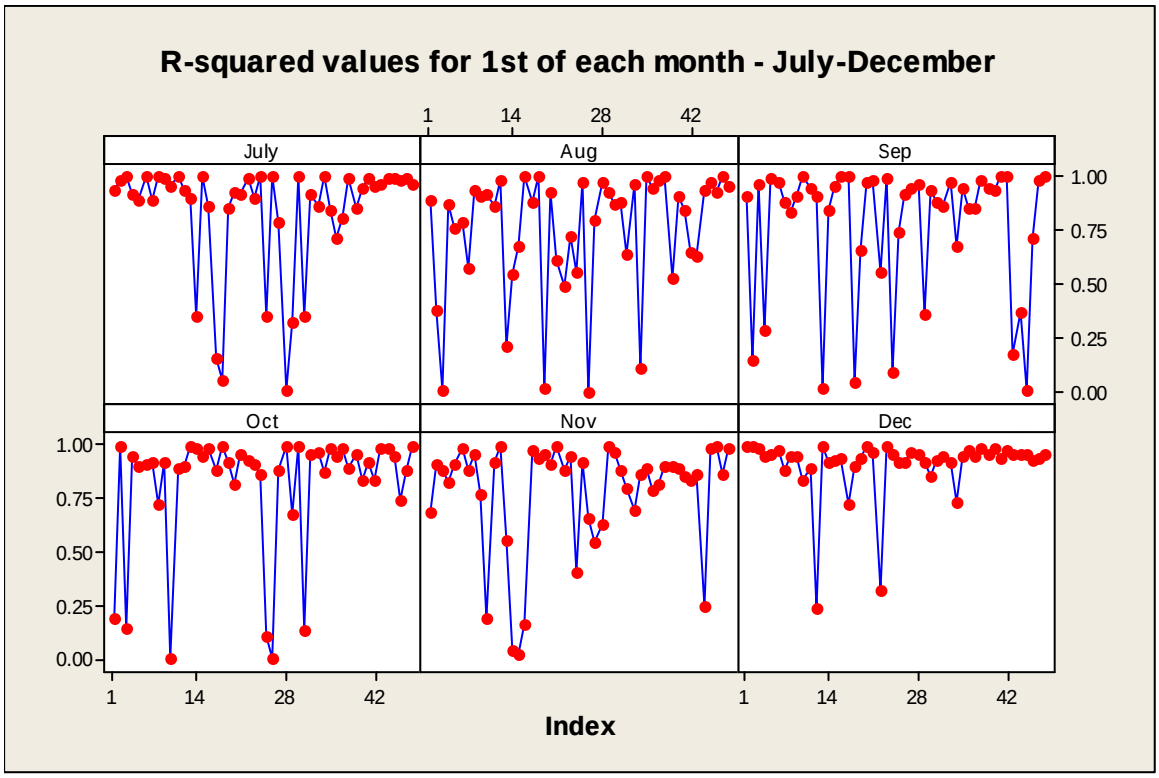


Figure 2.10: Time Series of the half hourly R^2 values over the 1st day of each month from 1997.

Figure 2.10 shows how the R^2 values vary throughout the day. These twelve days are typical of the values calculated over each of the years and as can be seen there is no real "worst time" of the day in each of the plots above. This (lack of) pattern is repeated throughout the 5 years in total so there is nothing to look at more in depth with regard to this.

2.3.4 R^2 Values for Day and Night in Summer and Winter

It was thought that perhaps there could be a difference between the day and night calculations. Perhaps any activity in the canopy below during the day could affect the quality of the three gas concentrations that are being measured. This can be checked by looking subjectively at some plots of day and night fluxes and comparing any differences between them.

	Day Median	Night Median	Difference (Day - Night)	CI for Difference	P-value
1997	0.91844	0.94233	-0.01435	(-0.017, -0.012)	<0.0001
1998	0.90404	0.92847	-0.01075	(-0.014, -0.008)	<0.0001
1999	0.81672	0.85024	-0.01512	(-0.020, -0.010)	<0.0001
2000	0.91494	0.91468	0.00020	(-0.0015, 0.002)	0.7864
2001	0.71341	0.73286	-0.00522	(-0.011, -0.0002)	0.0387

Table 2.4: The results from a Mann-Whitney analysis on the difference between the day and night R^2 values (negative values indicating night giving better fits)

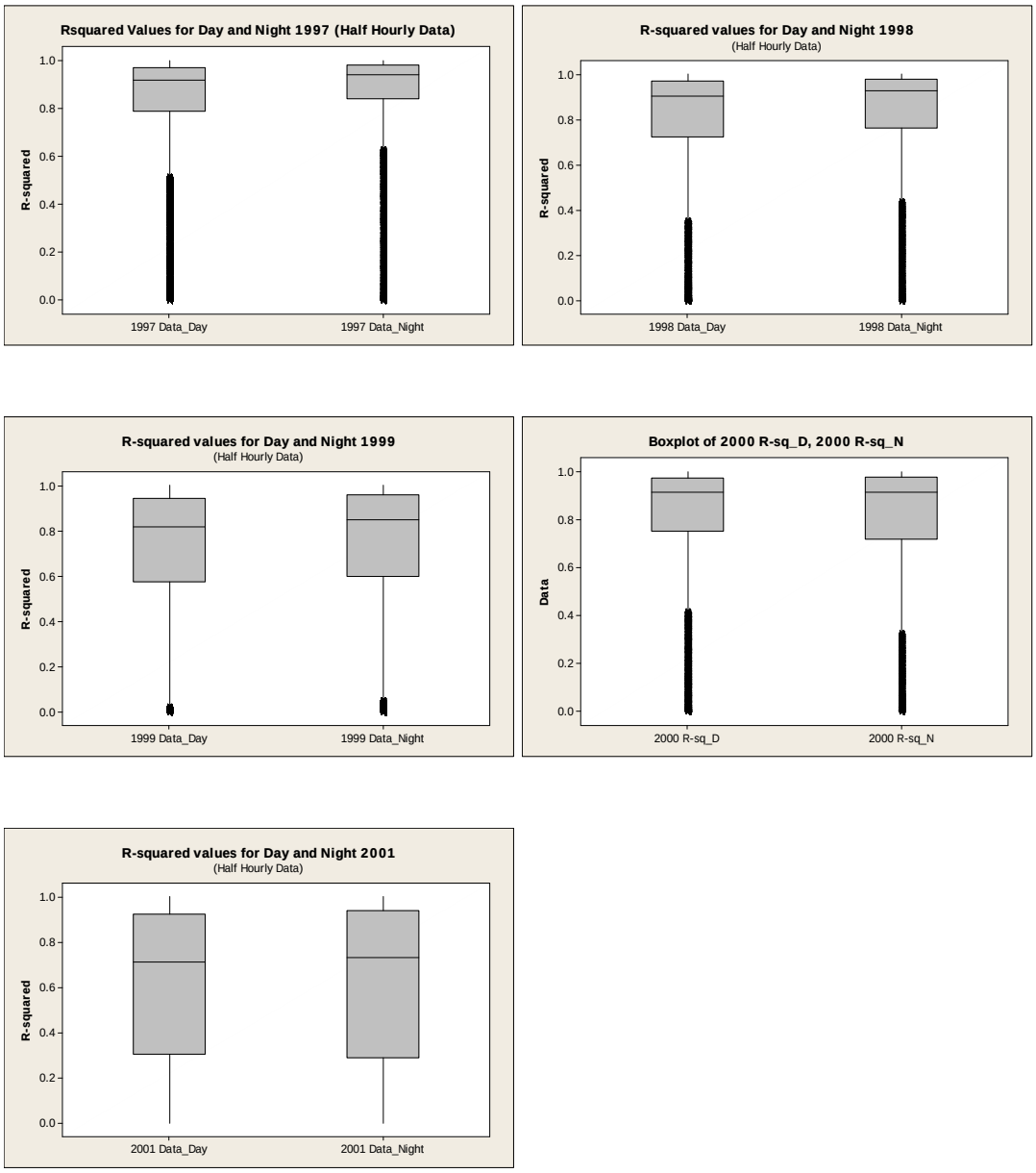


Figure 2.12: The Daytime R^2 values against the R^2 values at night, using the half hourly fluxes from 1997-2001.

Figure 2.12 shows that there appears to be little difference subjectively between each of the years – day or night. Because there are so many data points though, it is difficult to tell entirely. A more formal analysis between the day/night R^2 values can be implemented by performing Mann-Whitney tests on these R^2 values with the hypotheses below:

- H_0 will be that there is no difference in the values of the R^2 between day and night.
- H_1 that there is some difference.

Mann-Whitney tests will be used as the distributions of the R^2 values are not normally distributed, so the 2-sample t-test would not be suitable. The Mann-Whitney tests only require the variance and shape of the distributions to be the same and from the plots in Figure 2.12 this seems reasonable.

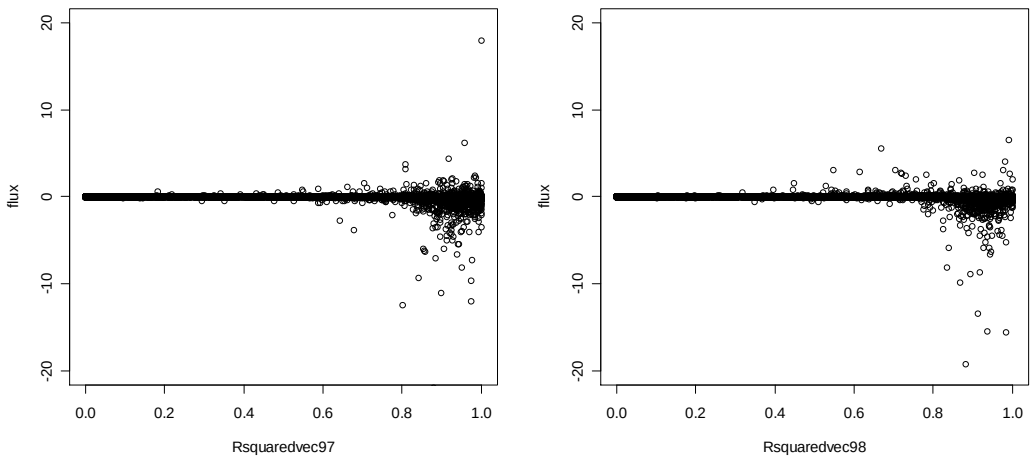
These tests show that (bar 2000), the straight lines, as measured by the R^2, are better fits at night to the data than during the days. However, the differences are very small for every year, so it seems unlikely that the measurements will be highly affected by poorer measurements during daylight hours. Clearly with 17000+ data points for each year, the confidence intervals are very tight.

2.3.5 Comparing the Fluxes and the R^2 values

It would be of interest to see if there is any pattern between the flux values and the corresponding R^2 value for that particular value. Then it can be seen if the higher R^2 values coincide with high or low fluxes or whether the goodness of fit has no bearing on the flux value that has been derived.

Five plots showing the flux values against the R^2 values again for each year are shown in Figure 2.13

1997 1998



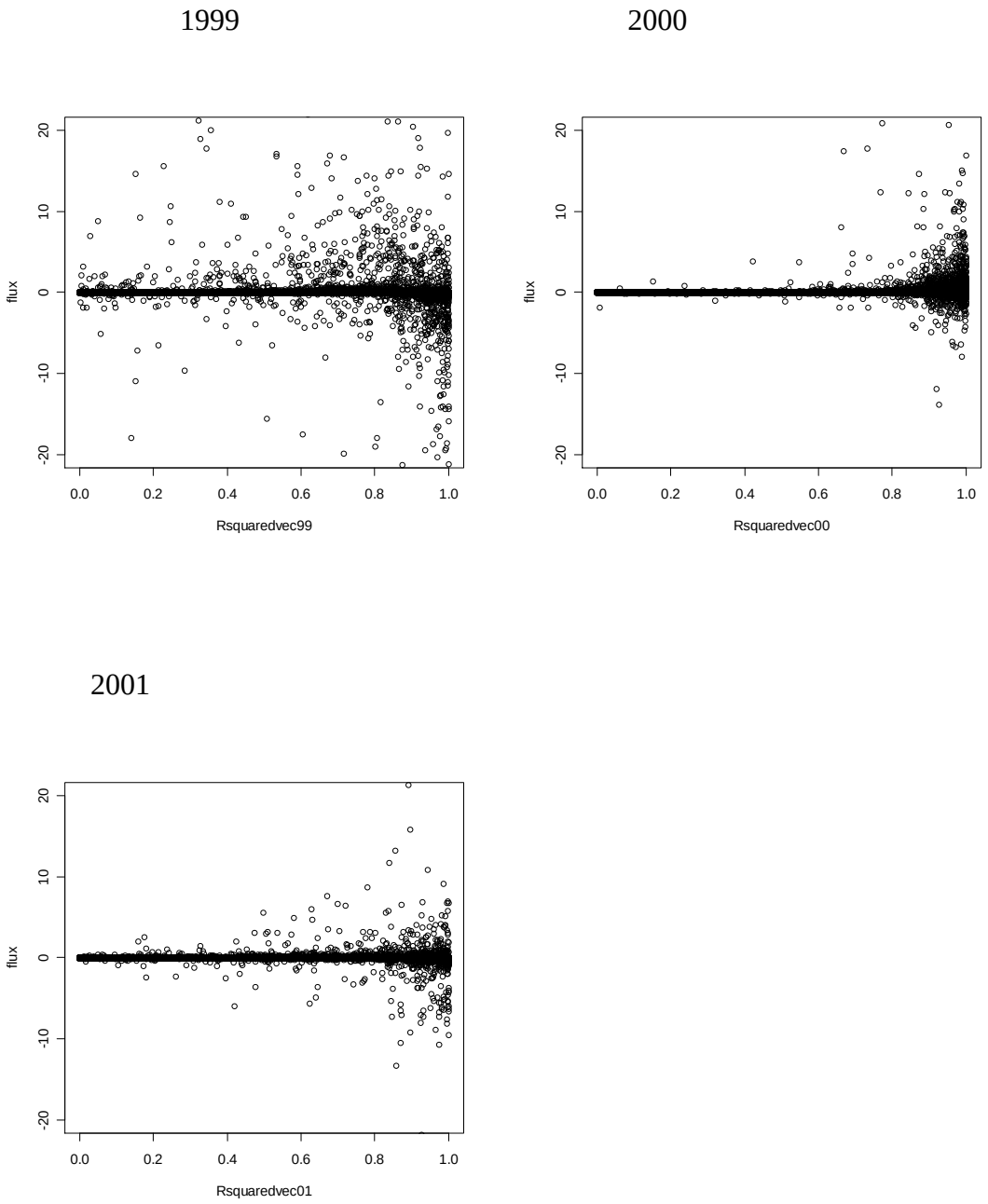


Figure 2.13 Plots of the R^2 value vs. the corresponding flux value for each half hour, from 1997-2001. Graphs have all been set to a max of 20 and min of –20 on the flux axis.

Subjectively from Figure 2.13, the plots look pretty similar in shape for 1997 and 1998, and appear to have more scatter towards the right hand side (i.e. higher R^2 values) for the final 3 years. Because most of the fluxes are very small, it is difficult to see any obvious relationship that may be between these two variables. It should be

noticed though that there appears to be very few high positive or negative fluxes at the low end of the R^2 scale. This could suggest that the high fluxes are producing better fits in general. However it is dangerous to assume causality here, as it could also be that the better fitted data are producing higher fluxes. This seems to make more sense as three points which don't have a high R^2 value may have a very flat line as the best-fit estimate.

Since it appears to be the smaller fluxes that are producing the lower R^2 values, it might be useful to look at just how many of the slope estimates (since the flux is directly calculated from this), are significantly different from zero. It may be that if these are removed from the data, (since a slope estimate of zero will produce a flux of zero too), then this will improve the R^2 values in general and give fluxes that can be accepted more readily.

By taking the slope estimate and creating a 95% confidence interval as the estimate ± 1.96* the standard error of the slope should give an idea of whether the slopes are actually significantly different from zero. Table 2.5 shows this:

Year	**No. of (non-missing) measurements**	**No. of significant slopes**	**% of Significant slopes**
1997	14307	10729	74.99
1998	14534	10119	69.62
1999	14290	7750	54.23
2000	14728	10323	70.09
2001	14394	6028	41.88

Table 2.5: The number and percentage of approximate significant slopes that have been obtained in each year

From Table 2.5 it can be seen that the data in three of the years appear to give around 70% 'good' slopes (i.e. ones that are significantly greater or less than zero). 1999 and 2001 seem to be quite low however with just over half of the slopes being significant. This may prove to be useful if the 1999 and 2001 years are the worst in comparison to the modelled data

In 2001 the number of significant slopes drops to a very low value in comparison to the other years. This may help to explain why there are a lot of low R^2 values compared to other years in 2001, since many of these slopes are not significantly different from zero. This could therefore be another filter to think about when calculating fluxes. This will be explored more in Chapter 5 when the model/measurements are looked at further.

To see if this alters the pattern of R squared values, the table below (Table 2.6) shows a summary of the R^2 values for each year, but only taking into account the significant slopes.

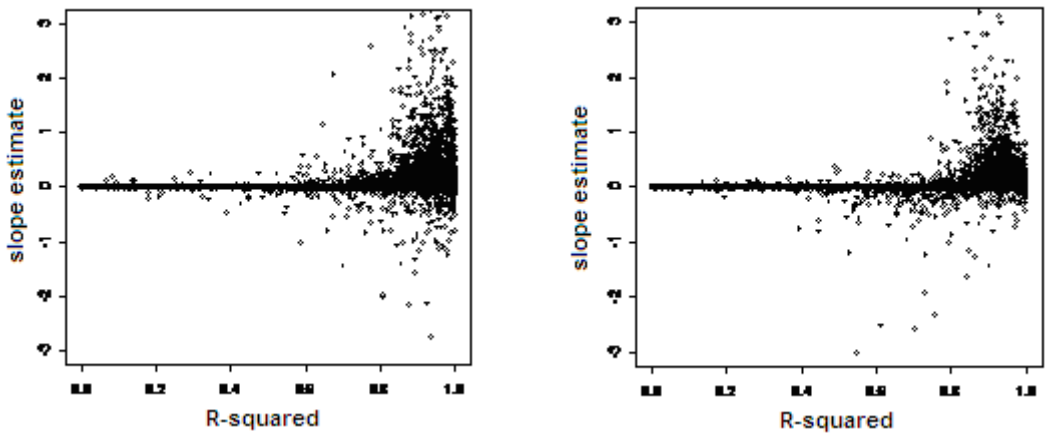
R^2 values for sig slopes	Min	Q1	Med	Mean	Q3	Maximum
1997	0.800	0.9072	0.953	0.9401	0.9850	1.000
1998	0.8001	0.9051	0.9514	0.9383	0.9839	1.000
1999	0.800	0.8709	0.9424	0.9287	0.9840	1.000
2000	0.800	0.9072	0.9588	0.9422	0.9887	1.000
2001	0.8001	0.8866	0.9468	0.9316	0.9843	1.000

Table 2.6: Summary of R^2 values collected between the years 1997 and 2001

Table 2.6 shows that the significant slopes give far better slope estimates, as we would expect, (based on Figure 2.12) than the non-significant ones. Perhaps if only these (significant slopes) are analysed, then these will give more accurate flux measurements. This should be something to take into account when comparing the model and measured data in Chapter 5. This possibly also shows that the reason the 1999 data seems to vary more is because the number of slopes that are actually providing a significant flux measurement is very small.

Also useful could be looking at the actual slope estimates against the R^2 values. Since the flux values are obtained by multiplying the slope estimates by a constant and u* which can change for each half hourly period there is a possibility that the patterns may look slightly differently for these. These plots are shown in Figure 2.14.

From Figure 2.14 it can be seen from this that some of the slope estimates are positive and some are negative. From (2.8) it can be seen that a positive slope estimate will lead to a negative flux. This is defined as an upward flux, which as the name suggests would occur when more SO_2 is coming up from the ground rather than down through the air. This may be the case on rather still days where SO_2 is not moving much through the lower atmosphere. Also the fluxes look slightly different to the fluxes vs. R-squared that were pictured in Figure 2.12. It can be seen that 1999 is the only year that appears to have more slope estimates below 0, than above it, which seems odd when compared to the other six years of data. Differences with the final 3 years data look possible too. During the years 1997 and 1998 it looks like most of the slope estimates are greater than 0. However in the final three years (99-01) there appear to be more negative values. Looking at the spread of the data also seems to show that 1999 and 2001 seem to have quite a small spread in comparison to the other years. This may suggest that these years may be easier to model and so perhaps give more accurate results. This should perhaps be taken into consideration when looking at the years.



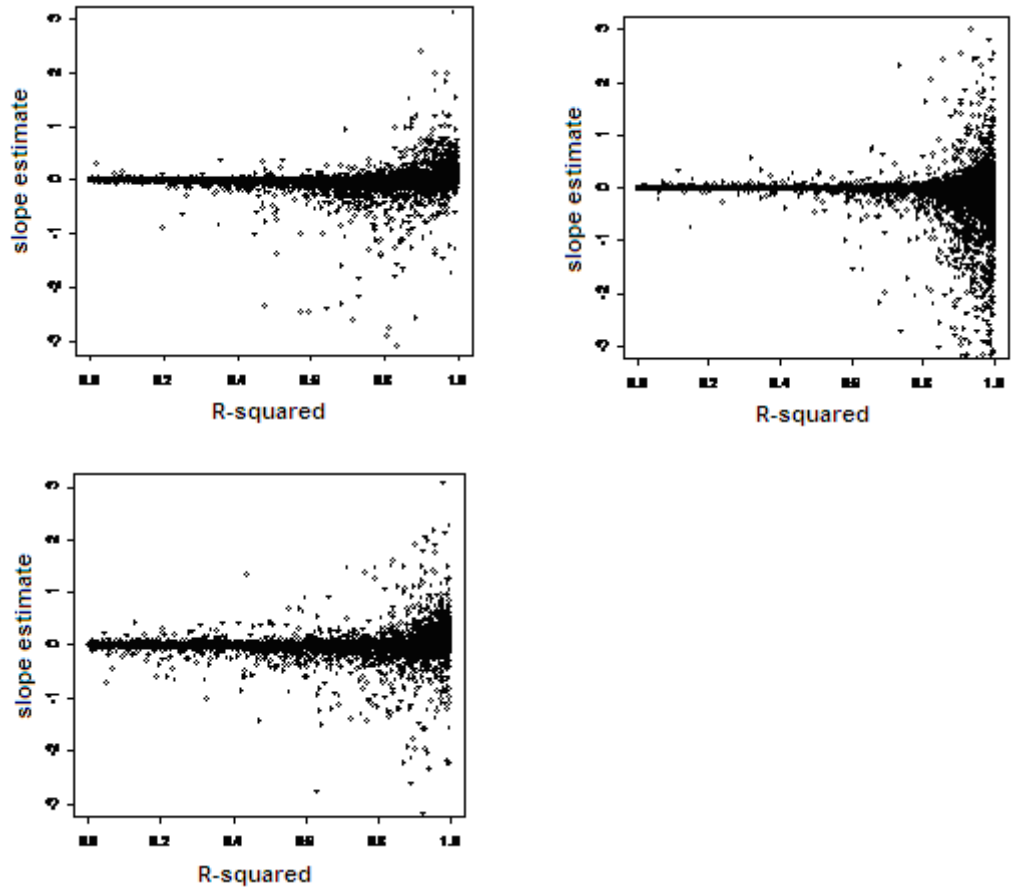


Figure 2.14: Slope estimates against R^2 for the years 1997-2001

2.3.6 U, Gas Concentrations and Stability Corrected Heights*

In the previous sections, u* was used in the flux calculation along with the slope estimates, and the gas concentrations and stability corrected heights were used to create the slope estimates.. Since these are obviously then of great importance to any calculation of a flux measurement, then these should probably be analysed in a similar way to the earlier analyses.

2.3.6.1 Looking at the Friction Velocity (u*)

Firstly u* shall be looked at as this is a simple straight measurement from the tower. Firstly Figure 2.15 shows the histograms of u* over the 5 year period 1997-2001:

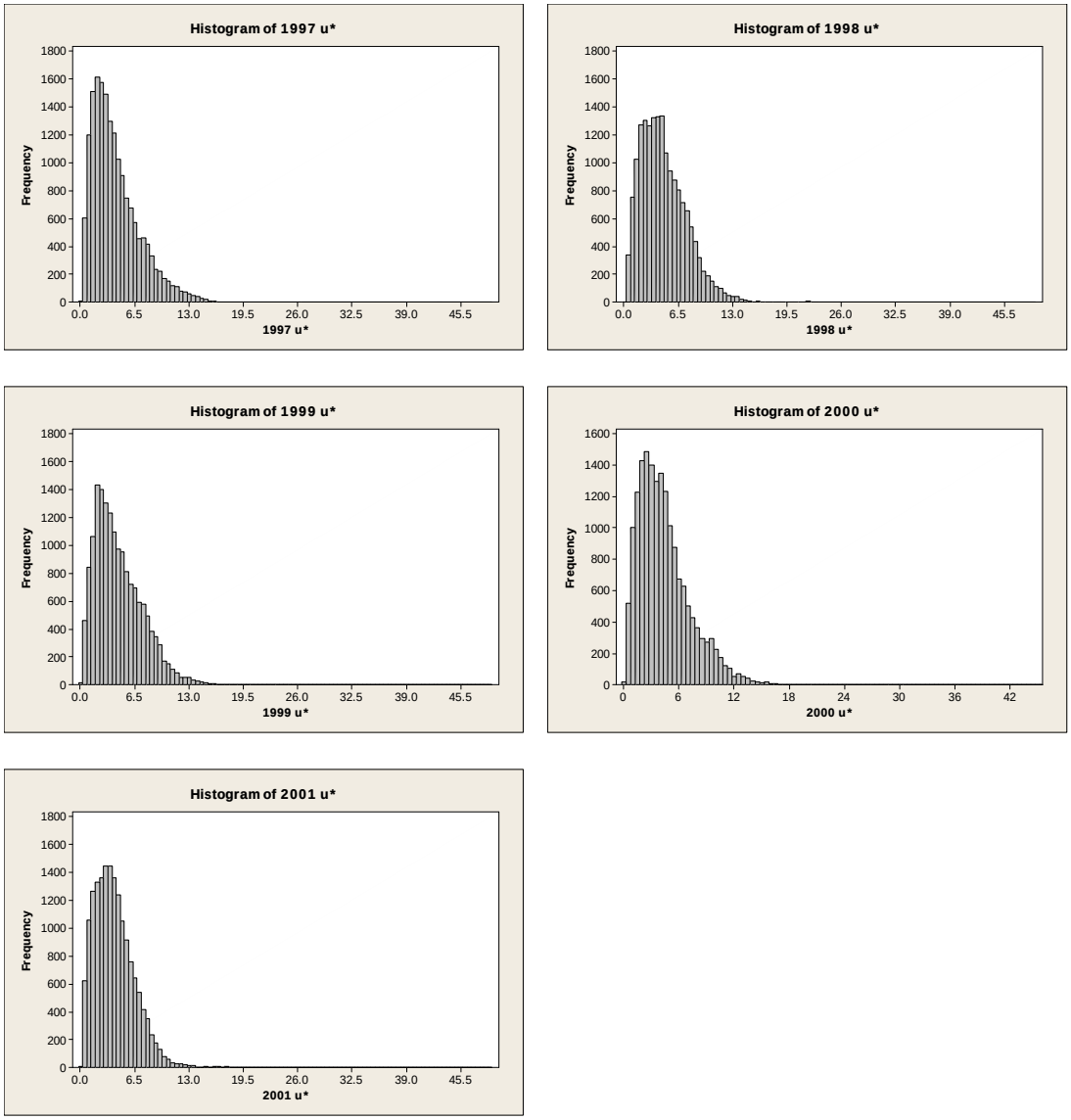


Figure 2.15: u* half-hourly measurements from 1997 to 2001

As can be seen in most of the years the u* distributions are slightly right-skewed but reasonably similar throughout the 5 years. The assumption of normality would be a risky one in this case. There also seems to be perhaps an exponential or chi-squared distribution to the figures, and appears to show that there are no real extreme values and hence shouldn't affect the fluxes negatively. From the plot it can be reasonably assumed that u* appears to remain similarly distributed throughout the 5 years that are being looked at. So there certainly appears to be no obvious change or difference when looking at the friction velocity of the air throughout this time period.

2.3.6.2 Gas Concentrations and Stability Corrected Heights

Table 2.7 and 2.8 contain the summary statistics for the SCHs and secondly the gas concentrations (at each of the 3 different measurement heights):

From these summary statistics, a few things stand out. Firstly, some of the maximum values can be seen to be very large when compared to the rest of the data in both the gas concentrations and the SCH's. These data will be analysed in more detail in Chapter 4, so it is important here to look if the data appears normally distributed, if these extreme values are not taken into consideration.

Year	Gas Heights (m)	Min	1st Quartile	Median	Mean	3rd Quartile	Max
1997	2.82	-5.414	0.910	1.015	1.296	1.167	73.138
	1.21	-5.437	0.065	0.110	0.197	0.172	29.534
	0.43	-5.491	-1.122	-1.108	-1.100	-1.100	7.630
1998	3.05	-5.703	1.021	1.100	1.365	1.240	90.298
	1.16	-5.726	0.035	0.064	0.138	0.115	32.106
	0.35	-5.786	-1.394	-1.387	-1.379	-1.375	6.159
1999	3.05	-5.291	1.030	1.097	1.497	1.231	236.564
	1.16	-5.319	0.039	0.063	0.181	0.111	84.646
	0.35	-5.394	-1.393	-1.387	-1.370	-1.376	18.532
2000	3.05	-4.989	0.959	1.097	1.416	1.268	65.364
	1.46	-5.012	0.248	0.315	0.429	0.393	29.943
	0.71	-5.048	-0.522	-0.491	-0.459	-0.456	12.798
2001	2.82	-4.492	0.919	1.016	1.379	1.206	76.967
	1.21	-4.530	0.070	0.111	0.234	0.188	31.105
	0.43	-4.616	-1.119	-1.107	-1.082	-1.084	8.108

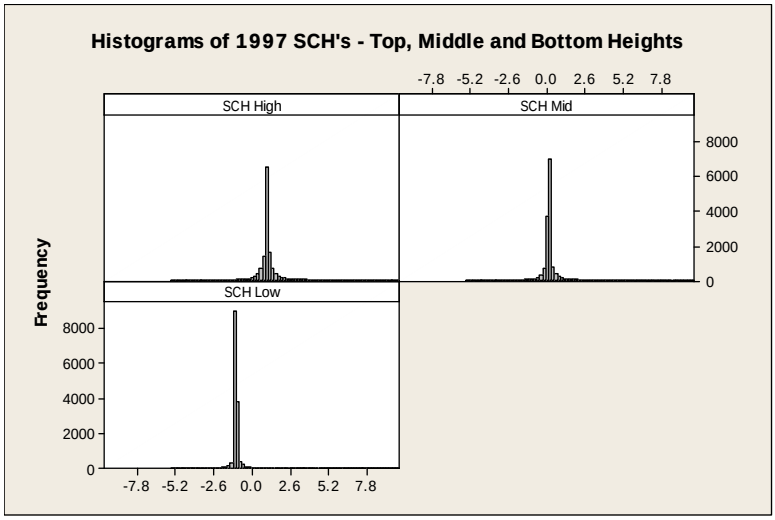
Table 2.7: The summary statistics for the stability corrected heights over the years 1995-2001 (in half-hourly measurements). The gas heights show the three heights in metres that measurements were taken at in that particular year. These have only been calculated if there is a corresponding gas concentration.

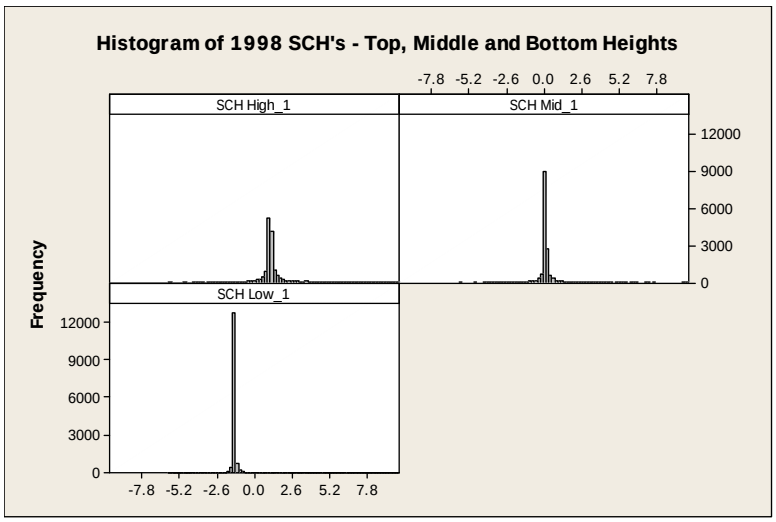
Year	Gas Heights	Min	1st Quartile	Median	Mean	3rd Quartile	Max	No.missing
1997	2.82	0.0000	0.100	0.214	0.751	0.610	42.341	3213
	1.21	0.0000	0.100	0.201	0.667	0.561	38.336	3213
	0.43	0.0000	0.100	0.193	0.612	0.518	35.208	3213
1998	3.05	0.00015	0.123	0.201	0.546	0.374	54.339	2986
	1.16	0.00004	0.126	0.195	0.494	0.351	47.103	2986
	0.35	0.00003	0.123	0.187	0.452	0.328	42.654	2986
1999	3.05	-0.162	0.064	0.124	0.451	0.300	26.939	3230
	1.16	-0.150	0.070	0.132	0.477	0.318	30.560	3230
	0.35	-0.161	0.068	0.129	0.464	0.304	29.108	3230
2000	3.05	-0.061	0.054	0.113	0.500	0.278	67.172	2840
	1.46	-0.040	0.061	0.121	0.546	0.299	68.988	2840
	0.71	-0.052	0.061	0.121	0.567	0.301	69.800	2840
2001	2.82	-0.072	0.054	0.114	0.458	0.324	26.878	1229
	1.21	-0.052	0.063	0.125	0.488	0.359	27.395	1232
	0.43	-0.041	0.060	0.120	0.467	0.341	25.921	1231

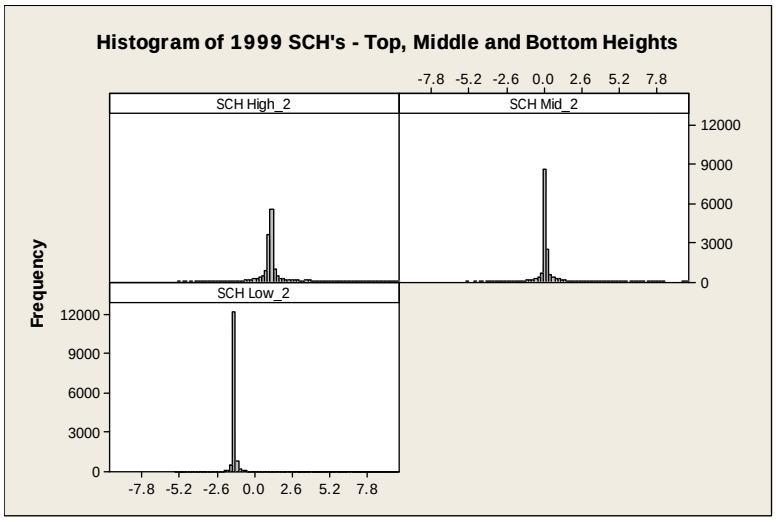
Table 2.8: The summary statistics for the gas concentration measurements over the years 1995-2001 (in half hourly measurements). The gas heights show the three heights in metres that measurements were taken at in that particular year

Figures 2.15 and 2.16 show the histograms of the stability corrected heights and the gas concentrations.

Stability Corrected Heights







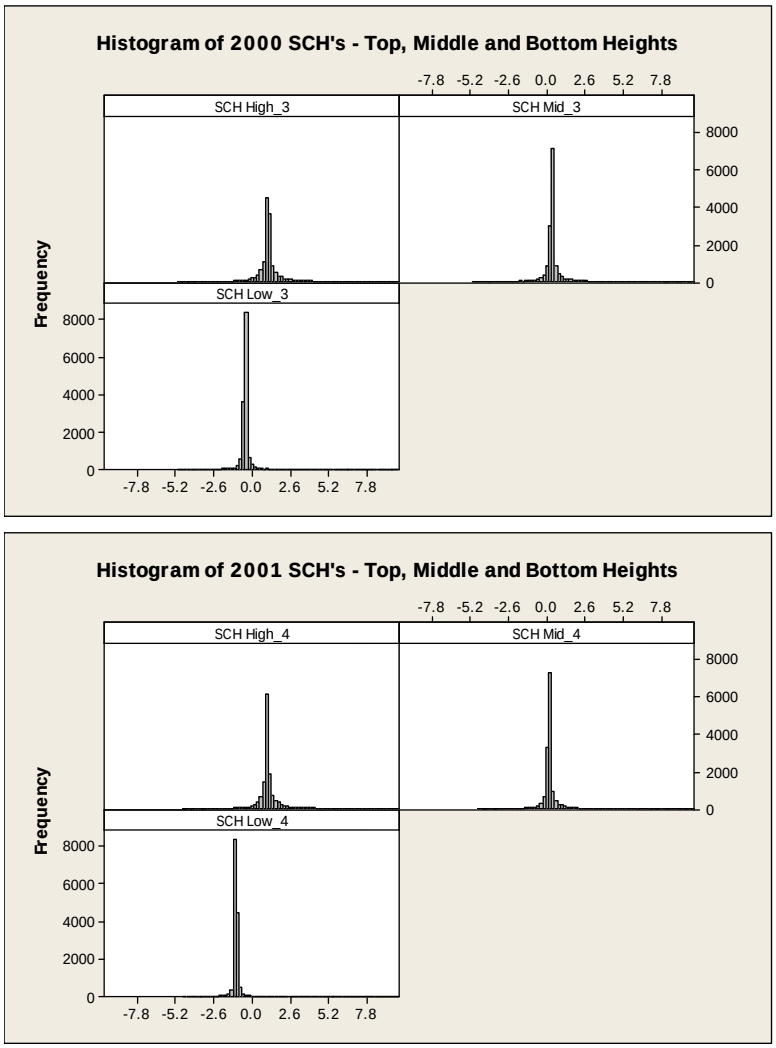
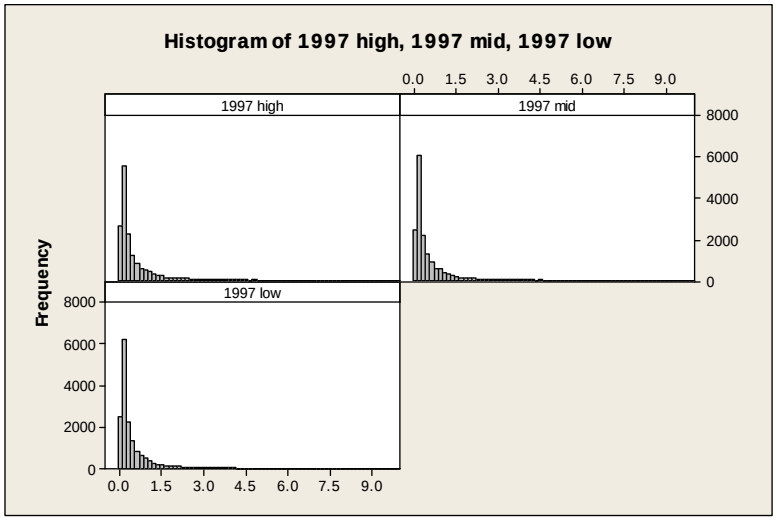
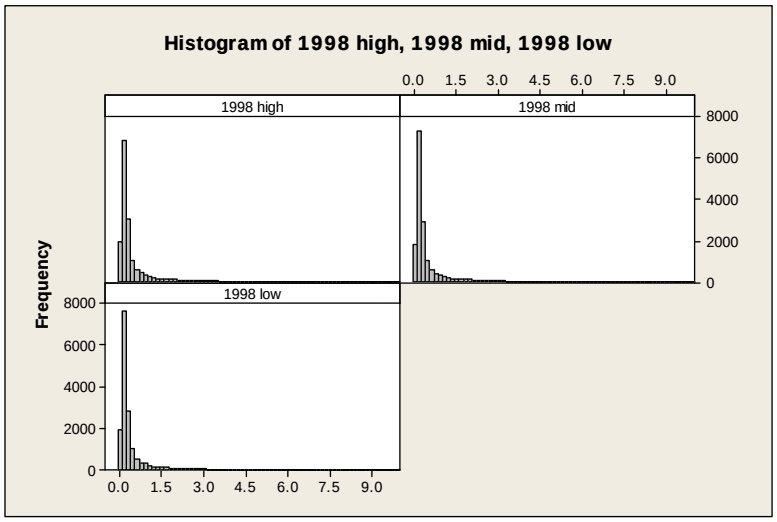
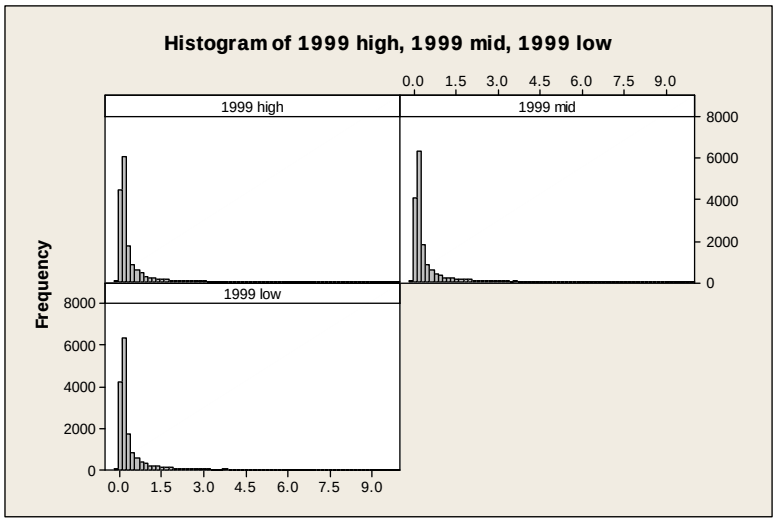


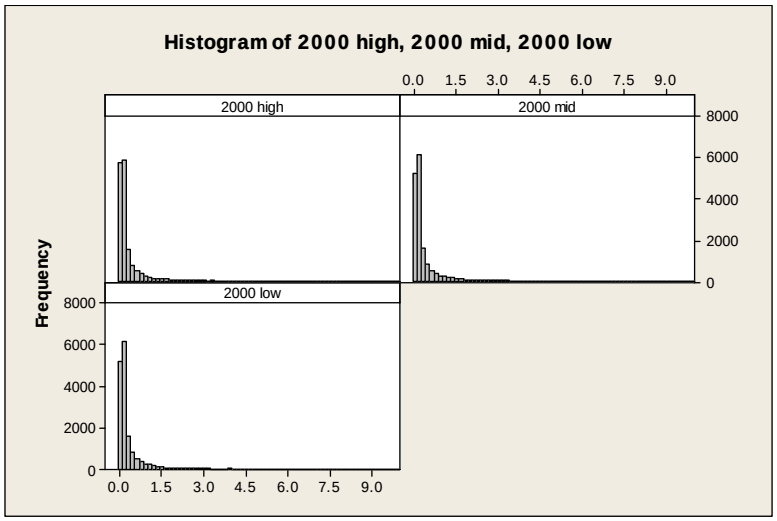
Figure 2.16: Distributions of the stability corrected heights from 1997-2001. Top left plot is for the highest height, top right corresponds to the middle height and the bottom left is from the lowest height

Gas Concentrations









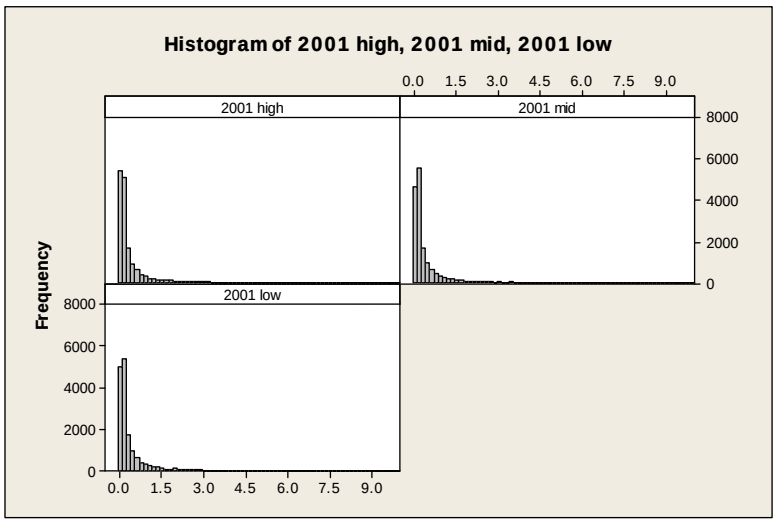


Figure 2.17 Distributions of the Gas Concentrations from 1997-2001. Top left equals the concentration at the highest height, top right equals the concentration at the middle height and the bottom left equals the concentration at the lowest height.

Looking at the SCH's first in Figure 2.16, the data look reasonably normal at all three heights. Only 1999 looks to possibly deviate from this, but there would still be a strong argument to be made for a normal distribution to be a reasonable approximation to that as well. These histograms have been shown on a –10 to 10 scale so the extreme data doesn't affect the distribution of the bulk of the data.

In Figure 2.17 the gas concentrations can be seen to appear to follow a similar distribution to each other in each year 1997-2001. They all look possibly exponential in distribution, but there doesn't seem to be much change in any of the years, once the extreme data is looked at separately.

We can look at a Sensitivity Analysis in order to see where the variation in the fluxes might be coming from. This will be looked at in Section 2.5.

So far it would appear that the main concern in these data are the number of non-

significant slopes that are being calculated, coupled with a reasonably high number of low R^2 values when using the slope estimate to try and estimate a flux. It has been shown (Table 2.4, for instance) that these appear to be highly related and so by only using the significant slopes it would appear that the slope estimates calculated would potentially be closer to the "real" value of the flux. Looking at some of the other variables that are used in the flux doesn't appear to have shown anything that would adversely affect the level of confidence that a flux calculation would provide. However it is still clear when looking at the fluxes throughout the year that they still appear difficult to model. Since this may be down to random noise it would be useful to perform some sort of Signal to Noise calculation so that it can be seen how easy or difficult it is going to be to predict what is going to happen in the future.

2.4 Signal- to- Noise Ratios

2.4.1 Introduction

Signal-to-Noise Ratios are used often to ascertain how good a signal is. As the name suggests it is simply a ratio of the level of a signal to the level of background noise. The bigger the Ratio, the better the signal is. It is commonly used with regard to radio signals, but can also be used in medicine for looking at cells and other measurements as well. It is useful to put a value on how much noise there is around a signal.

2.4.2 Estimating the SNRs

The fluxes having now been looked at, along with the measurements used in the calculation of them it might be useful to start looking at how easy or difficult it might be to extract a signal that would be used to predict future observations. At the start of this chapter a question was asked about how easy it might be to predict a flux value. By looking at the fluxes and the variables used to calculate the flux it can be seen which particular variables appear to be most constant and which seem to deviate more. It would be useful to look at how much signal there is in comparison to the levels of noise. For this purpose, Signal-to-Noise Ratios (SNRs) are useful. There are

different ways of estimating these, one is discussed in Rout & Mittal (2006), a "time-local, inband signal-to-noise ratio" is estimated in Mellinger and Clark (2006). Lots of the time SNRs are used in the measurement of sound/light waves etc, and not used as often in environmental systems (including the two previously mentioned above). Pauluzzi (2000), however, has produced an estimate which can be applied to the environmental data in this context. This is defined in (2.9)

$$\hat{\rho} = \frac{\frac{1}{2}\sqrt{6M_2^2 - 2M_4}}{M_2 - \frac{1}{2}\sqrt{6M_2^2 - 2M_4}} \qquad (2.9)$$

where

$$M_2 = \frac{1}{N}\sum_{n=1}^{N} |y_n|^2 \qquad M_4 = \frac{1}{N}\sum_{n=1}^{N} |y_n|^4$$

where $\hat{\rho}$ is the SNR estimate and $\{y_n\}$ $1 \le n \le N$ are the data points.

One problem when using this estimate is that sometimes the value '$6M_2^2 - 2M_4$' can turn out to be negative when small values for $\{y_n\}$ are used. This will obviously mean that an estimate for ρ will not exist. However for larger values this should not be a problem.

Applying these to the 48 half hourly fluxes that are derived each day would help to show how much signal there is present in the fluxes. It would also be useful at the same time to look at the inputs that are used to derive fluxes (the gas concentrations, wind friction velocities and the Monin-Obukhov lengths (L)). Table 2.9 shows the summary statistics of the SNR estimate, for each of the input variables and the flux, from the years 1997-2001. It can be seen that in a lot of occasions, this estimate doesn't produce values for some variables, but there are interesting results from the ones that are calculated:

	N	Min	Q1	Median	Mean	Q3	Max
1997 Flux	37	0.03	0.34	0.62	0.79	1.07	2.33
U*	353	0.28	4.28	8.27	12.39	15.26	86.80
L	31	0.32	2.72	10.51	$2.2*10^{14}$	59.46	$1.7*10^{15}$
Wind Dir	319	0.04	8.02	62.34	166.9	198.3	3280
SO_2 – high	174	0.10	1.13	2.71	$5.9*10^{13}$	6.42	$1.7*10^{15}$
SO_2 – middle	187	0.07	1.30	3.26	$5.5*10^{13}$	8.03	$1.7*10^{15}$
SO_2 – low	189	0.05	1.52	3.55	$5.4*10^{13}$	8.30	$1.7*10^{15}$
1998 Flux	39	0.16	0.36	1.03	1.03	1.51	2.71
U*	340	0.10	5.89	10.15	$1.6*10^{13}$	18.63	$1.7*10^{15}$
L	23	0.31	1.52	7.283	$2.9*10^{14}$	58.5	$1.7*10^{15}$
Wind Dir	326	0.02	21.27	88.46	$1.6*10^{13}$	243.2	$1.7*10^{15}$
SO_2 – high	217	0.14	1.96	6.13	$1.9*10^{14}$	15.86	$1.7*10^{15}$
SO_2 – middle	219	0.05	2.08	7.41	$1.9*10^{14}$	22.46	$1.7*10^{15}$
SO_2 - low	225	0.04	2.32	7.49	$1.8*10^{14}$	28.2	$1.7*10^{15}$
1999 Flux	32	0.17	0.46	0.90	1.02	1.51	2.81
U*	341	0.42	5.91	10.22	$9.6*10^{14}$	21.58	$1.7*10^{15}$
L	21	1.52	$1.7*10^{15}$	$1.7*10^{15}$	$1.5*10^{15}$	$1.7*10^{15}$	$1.7*10^{15}$
Wind Dir	320	0.09	18.62	112.8	$1.02*10^{14}$	296.4	$1.7*10^{15}$
SO_2 – high	185	0.25	1.24	2.61	$1.5*10^{14}$	7.33	$1.7*10^{15}$
SO_2 – middle	190	0.17	1.60	3.17	$1.5*10^{14}$	8.20	$1.7*10^{15}$
SO_2 - low	191	0.07	1.38	3.00	$1.5*10^{14}$	7.63	$1.7*10^{15}$
2000 Flux	22	0.05	0.23	0.62	0.81	1.11	2.99
U*	360	0.11	4.84	8.89	$1.9*10^{13}$	16.40	$1.7*10^{15}$
L	11	0.24	0.91	1.55	$6.0*10^{14}$	$1.7*10^{15}$	$1.7*10^{15}$
Wind Dir	318	0.16	16.50	77.31	$2.1*10^{13}$	187.12	$1.7*10^{15}$
SO_2 – high	150	0.03	1.39	2.47	∞	15.58	∞

SO_2 – middle	151	0.17	1.45	3.16	∞	7.92	∞
SO_2 - low	153	0.01	1.24	2.95	∞	6.99	∞
2001 Flux	28	0.11	0.47	0.74	0.81	0.98	1.89
U*	359	0.008	5.16	9.99	$7.2*10^{13}$	18.5	$1.7*10^{15}$
L	17	0.71	$1.7*10^{15}$	$1.7*10^{15}$	$1.51*10^{15}$	$1.7*10^{15}$	$1.7*10^{15}$
Wind Dir	321	0.21	19.27	85.37	$8.1*10^{13}$	235.2	$1.7*10^{15}$
SO_2 – high	160	0.09	1.48	2.37	$2.1*10^{14}$	6.19	$1.7*10^{15}$
SO_2 – middle	159	0.13	1.60	3.36	$2.1*10^{14}$	7.23	$1.7*10^{15}$
SO_2 - low	161	0.15	1.63	3.22	$2.3*10^{14}$	7.01	$1.7*10^{15}$

Table 2.9: SNRs for the daily flux values along with the input parameters from 1997-2001

It can be seen here that in many cases the signal-to-noise ratio for the flux was not able to be calculated. This was because the formula used generated negative values for the '$6M_2^2 - 2M_4$' part- however for the most part this was not the case, so this could generally be ignored and only the non-missing results were used.

This is obviously very disappointing and suggests that the estimation technique isn't particularly suited to the low measurements that are occurring in this data set. The lack of options in an environmental setting for this sort of calculation however do not leave many other options open to performing this kind of analysis with the data so the results from this analysis will be studied, however scarce they may be.

Some of the means are very large. This is because some of the SNRs calculated had very small values for '$6M_2^2 - 2M_4$' which led to some extremely large values being calculated which increased the means by vast amounts! This is why the median value was looked at as a reasonable figure for the average here.

From these results (and others that were generated) it could be seen that the SNR values were very low. In the case of the flux values most of the SNRs lay between 0 and 1. Other literature on this topic can use SNRs of up to 1000 or 2000, to give an indication of just how small these values are.

Looking at the input variables it can be seen that u* appears to give the "best" signal

with the higher SNRs, whereas the values for L were very scarce. However it would appear that most of the noise in the input variables appears to be coming from the gas concentrations themselves. As the concentrations are required in order to produce the slope estimate then it would be useful to see how much this will affect the variation that the fluxes are showing. The other measurements all have higher medians and quartiles for each of the 5 years, so from this analysis it may appear that the concentrations may be having the most affect on the variation in the fluxes. In the next section it will be useful to have a look at how much variation it has compared to other input variables.

2.5 Sensitivity Analysis

It is clear that the fluxes are highly variable which will cause problems when trying to model them. Since these are derived from a model, it might be useful to look at a Sensitivity Analysis (SA) of the flux based on input parameters. This will show how much the variation of a model can be apportioned to the variation in the input parameters to the model. (Saltelli et.al 2000).

Different ways of estimating the sensitivity of a model can be found; the adjoint method (Hier-Majumber et al. 2006), Fuzzy-number based methods (Dou et al. 1995) and second-order reliability methods (Unlu et al. 1995) to mention a few. However a global sampling based method will be used here from Saltelli et al (2000) as they have applied techniques to environmental data beforehand and their approach looks suitable for this case.

In general the model is written as:

$$y = f(x_1,..., x_n) \tag{2.10}$$

By treating the output as y and the input vectors as x_k, distributions for the inputs can be assigned ($D_1,\ldots,D_k$), then values for each $\mathbf{x}_k$ can be produced and from these $y_1,\ldots,y_k$ can be derived and it can be seen which parameters contribute to most variation in the output.

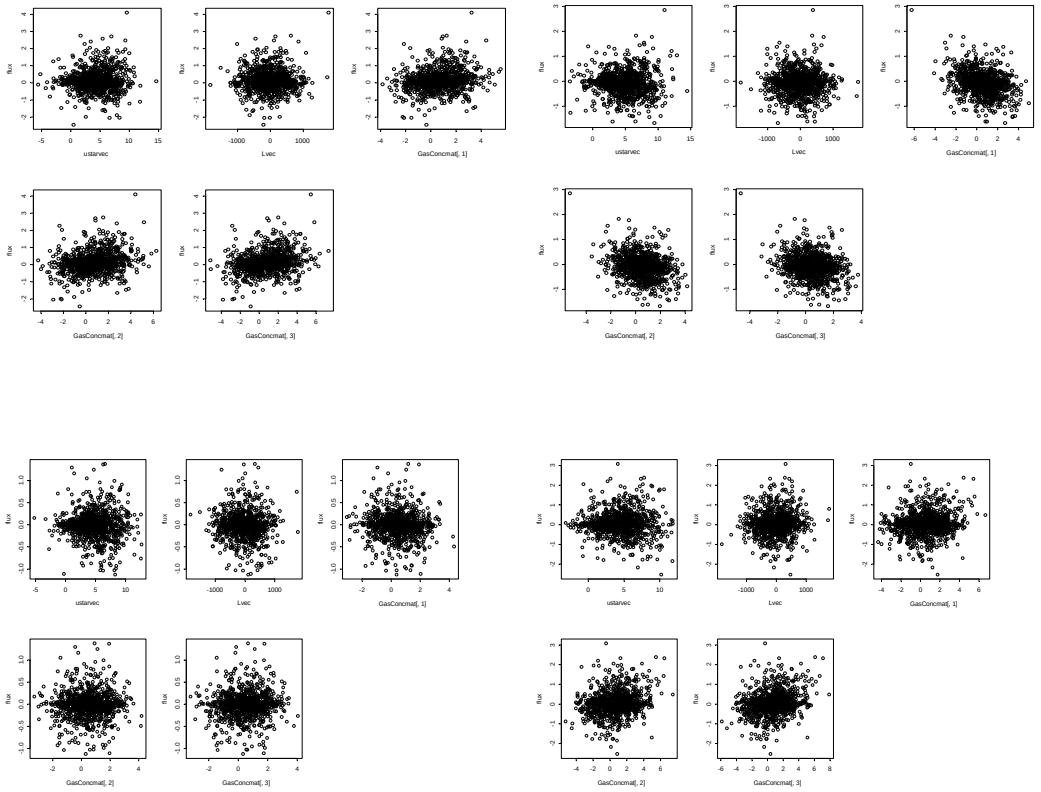
One thing that should be mentioned here is that a linear model (2.10) is being used to try and analyse a multiplicative model (2.8). A natural technique to improve this

would be to take a logarithm of (2.8), however since the fluxes can be negative (as well as the slope estimates) then this is not possible. However the model fitted from (2.10) should give us a good indication of which variables in the model are contributing most to the variation in the model.

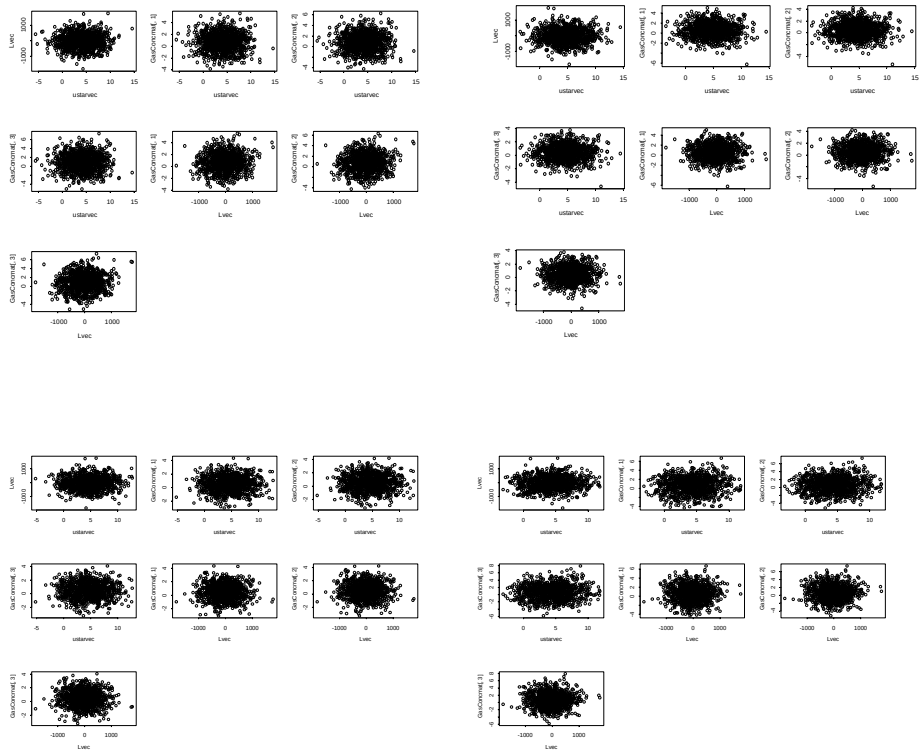
The flux is derived from five input parameters, namely 3 gas concentrations, the Monin-Obukhov length (L) and the wind friction velocity (u*). Both L and u* are approximately normally distributed and so these were used to sample from in order to generate values for the SA. With regard to the gas concentrations, it was seen as unrealistic to think of these being generated from independent distributions when they seemed very highly correlated. So these were sampled from a multivariate normal distribution, which took into account the correlations of the concentrations against each other (see Figure 2.19(b) and(c)). This was done in R and meant that the gas concentrations maintained their dependence to each other. 1000 values were sampled for each of the input parameters, which were then used to produce 1000 fluxes.

The flux values were then plotted against each of the input parameters in order to see what relationships there were (if any) between each of them and the output.

(a)



(b)



(c)

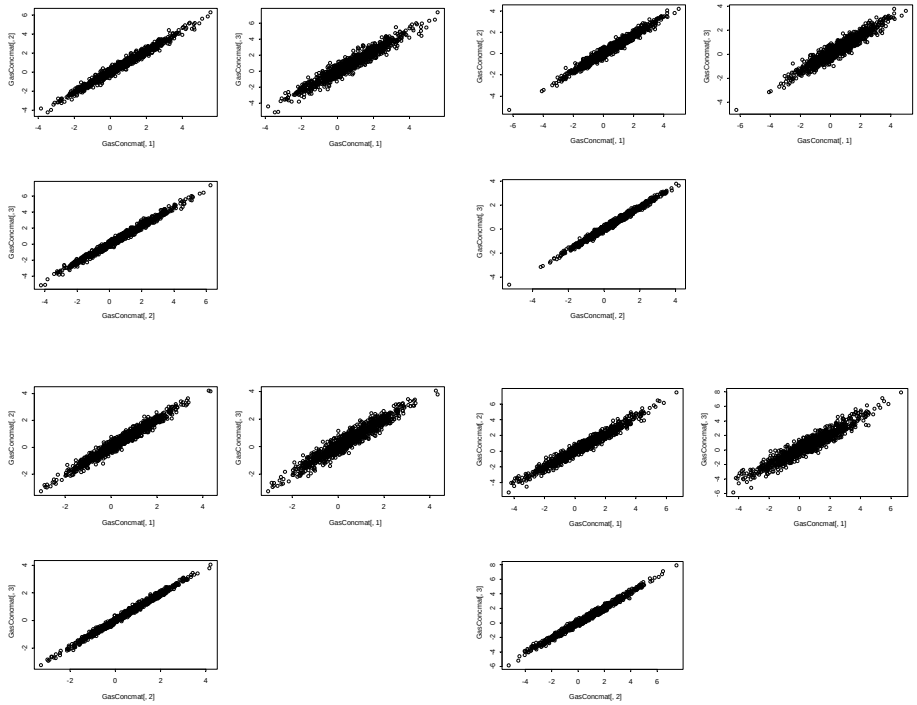


Figure 2.18: Plots from SA from 1997,1998, 1999 and 2000: a) A plot showing (from left to right) flux against u*, L, then the 3 gas concentrations. b) Plots showing the input variables against each other firstly L and the 3 Gas concentrations against u*, then the three gas concentrations against L. c) The Gas Concentrations are plotted against each other (top v. middle, top v. bottom, middle v. bottom)

Figure 2.19 shows very similar pictures of the data. It looks like there is not any obvious relationship between the flux and the input parameters independently from figure (a). Figure (b) doesn't show any relationship between any of the parameters, as we would expect since the input parameters were independently selected. Figure (c) shows the high levels of correlation between the gas concentrations at the three heights.

In order to see which input parameters were contributing to most variation in the flux, a regression was performed on the data to produce a model of the form:

$$flux = \beta_0 + \sum_{j=1}^{5} \beta_j x_j + \varepsilon \qquad (2.10)$$

where x_1=u*

x_2=L

$x_3 \ldots x_5$= Gas Concentrations

This produced a model, from which the coefficients of each parameter could be looked at in order that it could be seen which of them contributed most to the variation in the data. However, it is more useful to look at Standardised Regression Coefficients which give a better indication of which variable is causing most variation (Saltelli et al 2000). These give a coefficient that has been normalised to have mean 0, standard deviation 1. So it gives a level of importance based on moving each variable away from its expected value by a fixed fraction of its standard deviation.

Table 2.10 below shows the results from 1997-2000

Input Variable	Year	Co-eff	SRC	p-value
u*	1997	0.0121	0.021	0.065
	1998	-0.0032	-0.020	0.5299
	1999	-0.0033	-0.037	0.2840
	2000	0.0039	0.020	0.5502
L	1997	<0.0001	0.013	0.5809
	1998	<0.0001	-0.010	0.7554
	1999	<0.0001	-0.024	0.4820
	2000	<0.0001	0.013	0.6969
GasConc1	1997	-0.3432	-0.964	0.0171
	1998	-0.1905	-0.640	0.1809
	1999	-0.4443	-2.004	<0.0001
	2000	-0.3433	-1.106	0.0012
GasConc2	1997	0.0661	-0.401	0.8093
	1998	-0.5065	-1.456	0.2435
	1999	0.2309	1.077	0.1440
	2000	-0.1299	-0.453	0.6176
GasConc3	1997	0.2877	1.635	0.0366
	1998	0.7036	1.820	0.0240
	1999	0.2002	0.915	0.0788
	2000	0.4576	1.689	0.0055

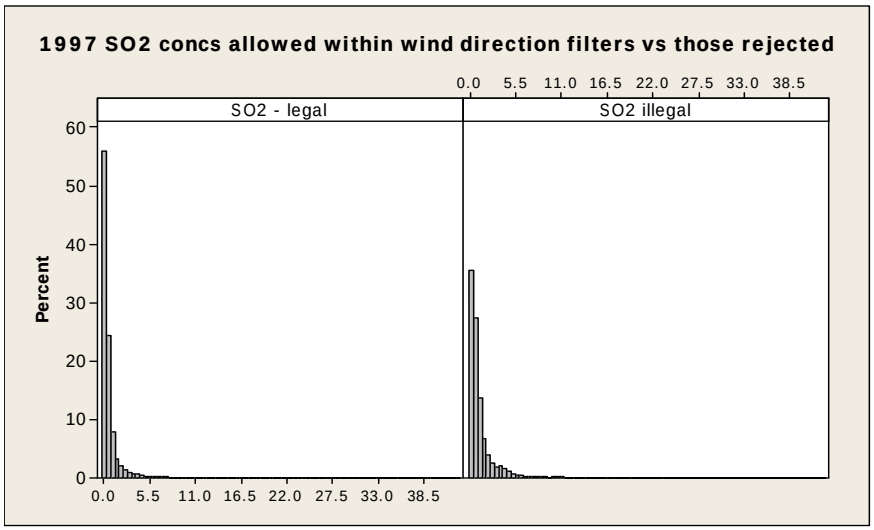
Table 2.10: The regression coefficients for a linear model on the daily 1997-2000 data, along with Standardised Regression Coefficients and p-value for each term.

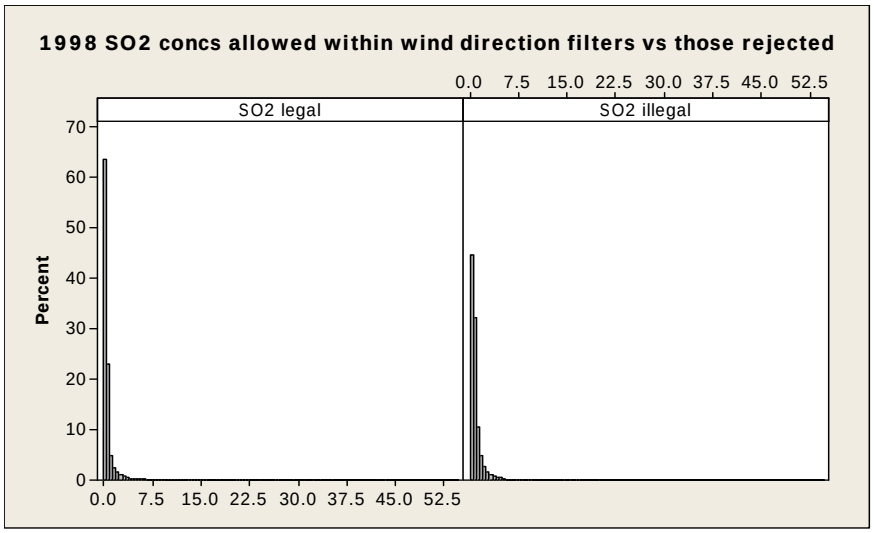
From Table 2.10 it appears that GasConc3 tend to contribute most to the variation of the flux (i.e. the one taken at the lowest point on the tower), however it would probably be more useful to look at the three concentrations together as they are so highly correlated that it would be difficult to imagine that one would remain the same if the others increased. The gas concentration SRCs still look much bigger than the coefficients that u* and L both give, suggesting that the gas concentrations are still the biggest contributor to variation in the fluxes, (with SRCs over 100 times as big as the

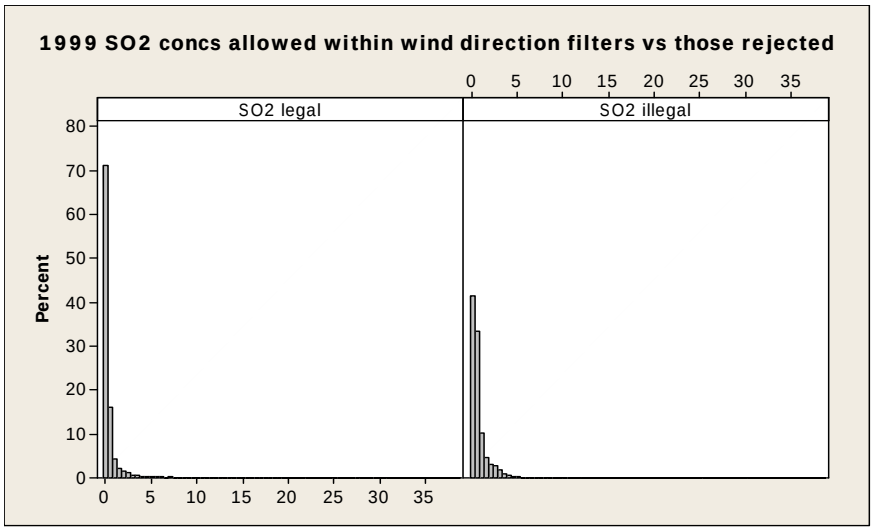
Monin-Obukov length and wind friction velocity). This would lead to a belief that these input parameters are not very important and there should be a good link between the fluxes and the gas concentrations, as one might expect.

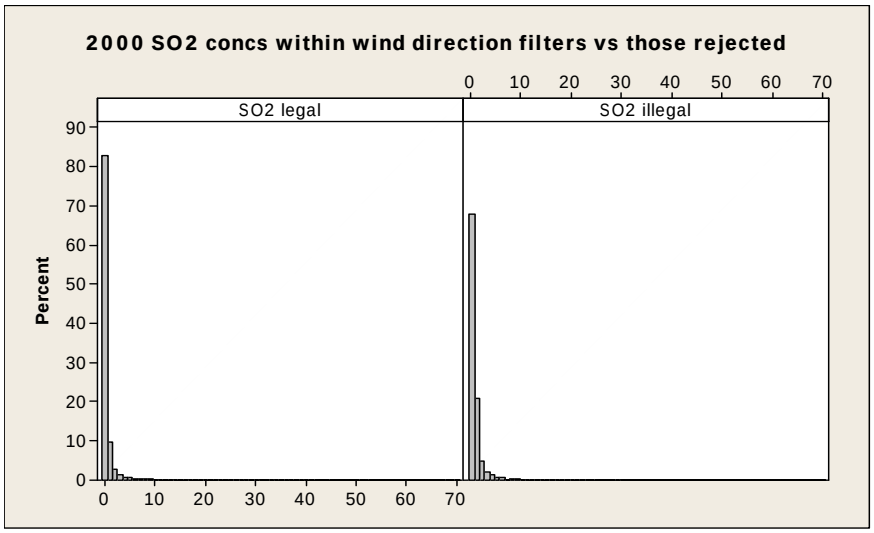
2.6 Filters Applied and Missing Data

As was mentioned in Section 2.2 it will be important to look at whether the values that are being rejected by the filters would show any unusual behaviour or differences with the data that have been allowed into the flux calculations. As can be seen from Figure 2.20, the wind filters that have been applied do not appear to affect the concentrations that have been measured, as the two distributions seems reasonably equal for each year 1997-2001. There perhaps appears to be some slightly larger concentrations in the samples that have been rejected but certainly it would appear nothing that would unduly affect the results in this case.









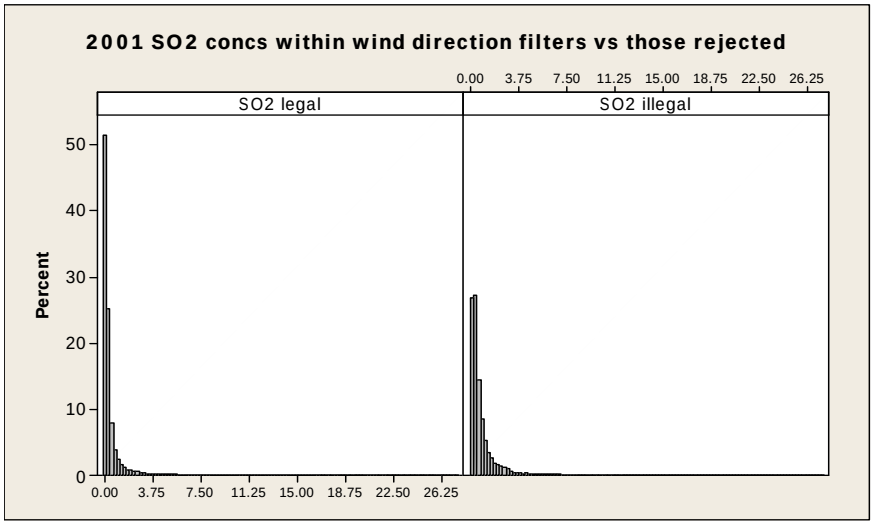


Figure 2.19: The half hour SO_2 concentrations distributed throughout each year 1997-2001. These have been split into both the concentrations not filtered, and filtered by the wind direction filter.

The other thing that was mentioned in Section 2.2 was to look at whether the data that were missing occurred at any particular times of the day. If one part of the day were showing more missing data than another then this would be of concern as this could be affecting the quality of the average daily fluxes that are being calculated. In order to test this, the number of times a value was found to be missing at each half hourly period of the day, throughout the year was studied. These results are shown in Table 2.10.

It can be seen from this table that there is no clear time of the day when more or less missing values are occurring so therefore it seems reasonable to assume that these data are missing at random throughout the five years of data collected. This would indicate that data is being collected fairly evenly from all times of the day and night throughout the year and therefore there isn't any complications over certain times having very scarce data.

	No of times missing							No of times missing						
Time of Day	0	1	2	3	4	5	Time of Day	0	1	2	3	4	5
1	85	122	108	45	3	3	25	60	139	111	45	8	3
2	82	149	89	40	5	1	26	64	136	112	42	9	3
3	72	145	104	37	6	2	27	69	144	99	43	8	3
4	77	146	99	36	7	1	28	68	141	96	46	14	1
5	82	142	98	30	14		29	72	137	95	52	7	3
6	88	135	87	45	9	2	30	51	150	113	38	11	3
7	74	148	96	35	12	1	31	70	139	102	38	14	3
8	81	142	102	31	8	2	32	67	141	105	40	11	2
9	78	135	104	39	10		33	81	115	117	44	7	2
10	66	161	86	41	12		34	70	138	107	40	9	2
11	77	148	94	40	6	1	35	70	127	111	42	15	1
12	72	151	94	40	9		36	68	140	106	32	18	2
13	72	137	99	45	11	2	37	70	130	112	42	11	1
14	74	131	113	36	11	1	38	81	128	98	44	14	1
15	80	122	114	39	10	1	39	78	134	93	48	11	2
16	83	133	101	37	11	1	40	72	139	101	41	13	
17	78	130	98	48	8	4	41	64	141	110	37	13	1
18	74	142	93	44	8	5	42	74	133	103	44	12	
19	63	143	101	45	12	2	43	90	130	93	42	9	2
20	56	151	100	49	5	5	44	71	150	102	34	8	1
21	64	129	112	51	8	2	45	63	162	88	41	12	
22	71	134	98	48	13	2	46	68	139	105	43	9	2
23	76	121	110	41	15	3	47	83	122	97	50	13	1
24	65	143	114	33	9	2	48	70	147	100	43	5	1

Table 2.11: Showing the number of times that missing data occurred in each half hourly period over the 5 year period. The rows indicate which half hour of the day it is, the columns are the number of times over the 5 years that a value was missing

2.7 Conclusion + Discussion

It has been seen in this chapter that in order to calculate fluxes many variables have to be measured including wind speeds, measurement heights and temperature to name three, as well as the gas concentrations. From these at least 8 separate calculations need to be applied to gas concentrations in order to produce a flux measurement from these. This can possibly lead to variation and quality and this has been what this chapter has explored.

Producing a best-fit line from 3 points on a graph, means that there is a larger margin for error and each measurement has a very high influence over the gradient of the best-fit slope, which is necessary to calculate the flux.

From the analysis of the seven years (1995-2001), by producing R^2 values (which it should be noted do not tell anything about how "true" the measurements are, only

how well a slope, based on three points, fits the data), it has been seen that for most years, a lot of the slope calculations appear to fit the data pretty well. This could give some confidence in the flux values that have been calculated from these. However, for the two years 1999 and 2001 there are a few potential problematic points that could have enough influence to affect a comparison between the modelled and measured data that will occur in later chapters.

Other problems have been shown to be that there are a reasonable percentage of the slopes who when combined with their standard error to produce some 95% confidence intervals, overlap zero, and therefore (since the flux is directly linked via multiplication) meaning that the flux for that particular half-hour could be zero. In 2001 especially it was shown that only 6987 of a potential 17520 observed slope values were significantly above/below zero. This tied in with the fact that there were more lower R^2 values in that particular year than any other.

Looking at the R^2 values in some more detail, especially during times of the day and year did not show very much. The hypothesis that perhaps 'better' measurements would be achieved at night when no plants were growing in the canopy, where the measuring tower was, appeared to be unfounded by analysing some time series plots and the R^2 values that could be directly compared against each other.

Since the flux is obtained directly from the slope measurement, the two variables (gas concentration and stability corrected height) that were used to measure this were analysed. Some of the maximum values that were produced seemed a little too high, but these were very few and doesn't look as if it would affect the data too much.

Finally, the other variables used in the flux measurement were looked at in order to see how much variation there appeared to be in them. Most had small variation, however it may be something to think about if the results show disagreement with the modelled data.

From these preliminary results that showed some variance in the flux results some Signal to Noise Ratios were looked at and found to be low. The SNR results however were estimated using a technique which did not produce a value for (in some cases) a majority of the input parameters. This meant that little could be analysed from these results. This led onto a Sensitivity analysis on the flux and the inputs that it receives. Using a technique applied to other environmental studies it was found that most of the variation in the fluxes was coming from the gas concentrations, with the other

variables lending little or no influence on this.

It should also be mentioned that most of this chapter has focussed primarily on the half hourly data. When the data is compared to the model in Chapter 5, this will primarily be the daily data which is used. However, if the half hourly data has been filtered to only allow “good” values then this should produce daily values that we can then make comparisons with the modelled data without having to worry about any measurement or calculation errors.

The next chapter will move from the data quality analysis to look at whether the data are able to be predicted at all, with or without these findings above.

Chapter 3 – Looking into Chaos

3.1 Introduction

It would be useful to know if the data, measured at Auchencorth Moss, are predictable. This can be achieved by looking to see if the series we have has a chaotic aspect to it. Below is an overview of what chaos means, how it can be defined, and how it can be employed in environmental situations such as the one that is being looked at here.

In addition, a useful aspect of looking at the data in this way will mean that it can be seen if the data are chaotic at the daily level. This will be useful to know when it comes to comparing the modelled and measured data, as it will give us knowledge of whether any model will fit the data well or not. From looking at time series of the data earlier on it could be seen that it will be difficult for any deterministic model to fit the data accurately. If the data are too unpredictable or very sensitive to very small changes then a completely different approach will need to be looked at instead. It will be useful to see the methods used to define chaos and see if aggregating the data up to daily levels (or even diurnal or hourly scales) will alter the levels of chaos that may or may not be present, i.e. identifying a temporal scale to work in.

3.2 Chaos

3.2.1 Introduction to Chaos

Chaos is a topic, which can provoke a great deal of disagreement in many people's minds. Even those who believe that chaotic behaviour can and does exist in environmental/natural systems, can have very different beliefs about what it is that they define as chaos. A quick Internet search immediately reveals that chaos is:

- A state of extreme confusion and disorder (www.cogsci.princeton.edu/cgi-bin/webwn)
- (physics) a dynamical system that is extremely sensitive to its initial conditions (www.cogsci.princeton.edu/cgi-bin/webwn)
- Chaos is the breakdown of predictability, or a state of disorder (cf Chaos

Discussion, also Chaos is Everywhere Discussion). (www.shodor.org/interactivate/dictionary/c.html)

- Chaos theory states that things are not really random, just complex. (http://www.webslave.dircon.co.uk/alife/chaos.html)
- Complicated patterns that are not truly random. Chaos is a cryptic form of order, what a random-number generator produces. There is, as the phrase goes, "a sensitive dependence on initial conditions." Because chaos was defined in a paradoxical way ("It may look random, but it's merely chaotic"), it is a term often misused or misunderstood (faculty.washington.edu/wcalvin/bk9/bk9gloss.htm)

There are many more definitions listed, but these are just five, showing some views that are held about chaos.

Definitions of chaos in time series are also a matter of some debate. According to Tong (1990) a loose description of chaos can be when

randomness is described by a strictly deterministic equation.

Tong defines randomness as the definition of chaos. Rapp and Schmah (2000) claim that he falls into one of two main groups of statisticians. There are the “randomness finders” and the “rule finders” and Tong falls into the former. This group use the degree of randomness as a measure of the complexity (chaos) of the time series, and total randomness shows a higher complexity. The "rule finders" believe however that complexity lies between these two extremes and in fact periodic and completely random series (i.e. both extremes) are both least complex. (Gu et al. 2004).

Babovic and Keijzer (1999) define chaos as a continuous power spectrum that does not contain any dominant frequency.

A common method of defining chaos is defining something known as Sensitive Dependence on Initial Conditions (SDIC). This technique is explained by Ellner (2000) as:

a small uncertainty about the system state now, producing a large uncertainty about what the state will be a while from now.

This basically means that unless a measurement is incredibly precise, then it will be very difficult/impossible to give a (worthwhile) prediction about the state at a future time point.

As can be seen, some of these definitions do not differ from each other greatly and should perhaps not be treated as entirely separate definitions. However, hopefully what has been shown in the paragraphs above is that there is a lot of variation when it comes to even defining what chaos actually is. Therefore, it should be imagined, that there will be a few ways of analysing real data as being chaotic, some which possibly disagree with others.

This review hopes to bring together the ways in which chaos and environmental time-series can be linked together and how chaotic behaviour can be assessed. There will be a few methods of attempting to calculate levels of chaos in a system. Both reasons for a particular method, and limitations of the same method will both be discussed. If it is possible then methods will be compared against each other as best they can.

3.2.2 Chaos in Environmental Systems

Even when restricting analysis to environmental systems there is still some dispute, as to how to measure whether chaos exists and how to quantify it. According to Turchin and Ellner (2000)

Despite an intensive theoretical and empirical investigation during the ensuing two decades, however, we do not have a widely accepted example of chaos in a field population.

Examples of these differing methods include Hastings et.al 1993 which looks at using Poincare Maps, Ellner and Turchin 1995 which uses first derivative estimates and Zimmer 1999 who talks about various ways of estimating Lyapunov Exponents.

Many scientists have attempted to apply chaos theory to environmental systems in order to explain them. Populations of voles living in Europe (Turchin and Ellner 2000), sunspot indexes and concentrations of carbon dioxide data at the South Pole (Giannerini and Rosa 2004) and forecasting river discharges (Babovic and Keijzer

1999) are just a few of the differing systems under which chaotic behaviour can manifest itself.

The following sections will help to show some of the differing methods by which chaos can be measured. Some of the examples mentioned above may be referred to later.

3.2.3 Arguments Against Using Chaos Theory Techniques

Before using some of these techniques to look at the data at Auchencorth Moss it would be wise to see arguments that are used by people who disagree with it. For example there are some arguments that suggest that chaos, whether it may exist or not, could be impossible to measure using the techniques shown above. Timmer et al. (2000) claims

"*For the calculation of the Lyapunov spectrum we had applied an algorithm that is nowadays known to be able to yield positive Lyapunov exponents even for white noise. Therefore, we now doubt the validity of these former results."*

This paper asserts that, even at a very small scale, noise can disable techniques used for ascertaining whether chaotic behaviour exists. If there is a stochastic element to the data then chaos can be falsely identified. Hence it will be very important to use chaotic techniques that will not be susceptible to normal variation in data patterns. It must be able to distinguish between noisy data which has an underlying model, and one which does not contain any predictable behaviour. If this can be achieved then Timmer et al's comments will be moot. It will be important that the techniques that are applied to the data must be able to differentiate between white noise and non random variation.

3.2.4 How is Chaos Assessed

As mentioned previously, chaos has a few definitions, which form the basis of a measure of chaotic behaviour. In this section some of the more commonly used definitions will be listed and commented on. Also, for some, there will be methods of estimating the level of chaos present in a system, so that meaningful values can be

computed from real-life time series etc.

The first way (and most popular way) of assessing how chaotic a set of data is, is by using Lyapunov Exponents. These use as the definition of chaos, the SDIC definition and work out how much a slight perturbation of a point (x(t) to x*(t) say) can grow into a large uncertainty over time (x(t+m) to x*(t+m))

In order to compute Lyapunov exponents, there must exist a time series of the form below (3.1).

$$X(t+1) = F(X(t), \varepsilon(t)) \tag{3.1}$$

Where F is the function that takes the value of the time series from X(t) to X(t+1), and ε is a random noise function.

Local and global Lyapunov exponents are then defined as follows in (3.2) and (3.3) respectively:

Local: $$\lambda_m(t) = \frac{1}{m}\log \| J(t+m-1)...J(t+1)J(t)U_0 \| \tag{3.2}$$

Global: $$\lambda = \lim_{m\to\infty} \frac{1}{m}\log \| J(t+m-1)...J(t+1)J(t) \| \tag{3.3}$$

where *m* is the duration of the time interval over which the exponent is measured, *J(t)* is the first derivative matrix (Jacobian matrix) of *F(X(t))* and U_0 is a vector of length 1 in the direction of the initial perturbation. ||..|| means the norm in this case.

Quite simply, if the Lyapunov exponents for the data are positive then there is evidence of chaos in the system. Wolff et al. (2004) states

"For one-dimensional chaotic systems, Lyapunov exponents have an important practical use: it is a necessary condition for the existence of chaos that the Lyapunov exponent be positive."

These exponents are the most widely used for calculating and quantifying chaotic

behaviour of systems. Ellner uses them to calculate the chaos element in a moth population time series, Frazier and Kockelman (2004) use them to look for chaos in transport systems and the analysis of the CO_2 data by Giannerini and Rosa (2004) also use Lyapunov exponents to analyse their data.

3.2.5 Estimating Lyapunov Exponents

From many time series however, it can be difficult to calculate Lyapunov exponents exactly, because of the difficulty in knowing what the underlying model (F) actually is, so there have been developed methods of estimating the values in time series.

The first of these is the simplest to implement. Firstly "flybys" are picked out of the data. These are times (t_1, t_2) where $\|X(t_1)- X(t_2)\|$ falls below some threshold value. These are treated as perturbations of the state at time t_1 and then the exponents are estimated by (3.4)

$$L_m(t) = \frac{1}{m}\log\frac{\| X(t_1+m) - X(t_2+m) \|}{\| X(t_1) - X(t_2) \|} \qquad (3.4)$$

By letting m tend to infinity, the global Lyapunov exponent can also be estimated from (3.4)

Obviously, this is a very simple way of calculating values of Lyapunov exponents at each time increment, and therefore can seem quite appealing for this reason. However, it has a major flaw, which make it certainly not the best approach to try. This is that this method does not take into account any noise that may be present in the data. Unless the system is 100% deterministic, a series has a large chance of diverging in time due to the presence of noise, that is prevalent in many real life (and especially in environmental) scenarios. Both this method and the following method are mentioned in more detail in Ellner (2000). (This concern is expanded upon in Section 3.2.5)

The second method is a more robust method. This involves estimating a function for the data - typically a non-linear model in the form shown in (3.1). From this estimate of F (F* say), estimates of J(t) can be produced by differentiating F* and

then equations (3.2) and (3.3) will be used with the estimates of each of these variables.

The advantages for this method are numerous when compared to the first method. Firstly, white noise should be accounted for when choosing the function F*, and as such the Lyapunov Exponent estimate should be more accurate. Secondly, the observed trajectory is still used here. This method does not test the model for chaos, it tests the real data for chaos. Doing the analysis this way means there is no added error from a simulated time series.

The negative side of this approach is that, it can sometimes be very difficult to find a "good" model for F. This method can also throw up odd Lyapunov Exponents too when many data points are used. (Tempkin and Yorke 2004)

A third method to estimate the exponents is described in Giannerini and Rosa (2004) where the LE estimation comes from the evolution in time of the distance separating initially close points. This is defined as the Maximum Characteristic Lyapunov Exponent (MLCE). This formula is shown below (3.5):

$$S(\nu) = \frac{1}{T}\sum_{i=1}^{T} \ln\left(\frac{1}{nf \min} \sum_{x_j \in U(x_i,\varepsilon)} \| x_{i+\nu} - x_{j+\nu} \|\right) \tag{3.5}$$

where

- T is the number of points (x_i) involved in the calculation,
- nfmin is the number of neighbours of each point that are closer than ε and have a temporal separation greater than nmin Kantz and Schreiber (1997) discuss choosing sensible values of ε and nmin. As long as both are not too small generally this shouldn't cause great concern.

This gives approximate straight lines over each of the values of x_i representing the evolution of the logged mean distance. The average along the trajectory of these lines, gives an estimated Global Lyapunov Exponent.

As shown by Giannerini and Rosa this estimator has two main advantages, as it do not require modelling, and the computation is easier than calculating Jacobian estimators. The problems with this are that there are no theoretical results for consistency, as it is a fairly new technique and for an asymptotic variance.

A method for estimating Lyapunov exponents using the R-language has been devised by Nychka et al. (1992). The LENNS program (Lyapunov Estimates for Noisy Nonlinear Systems) estimates global Lyapunov exponents from time series data. This programs uses a series of FORTRAN programs and calculates the exponent using methods that Nychka et al (1992) expressed.

The program uses the assumption that the data are of the form:

$$x_t = f(x_{t-1}, x_{t-2}, ..., x_{t-d}) + \sigma e_t \qquad (3.6)$$

where

1 $x_t \in R$ and e_t are independently identically distributed variables with zero mean and unit variance
2 f is a non-linear (in most cases) function,
3 d is the embedding dimension. This quantifies how far into the past the model looks for an explanation in changes to the current population

LENNS estimates the function $\hat{f}$ and uses this along with the data $\{x_t\}$ to produce a dominant/global Lyapunov estimate. It uses the Lyapunov exponent definition of:

$$\lambda = \lim_{m \to \infty} \frac{1}{m} \log \| J_m J_{m-1} ... J_1 \| \qquad (3.7)$$

where λ is the global Lyapunov exponent, m is the time delay and J_t is the Jacobian (1st derivative) matrix of f in Equation (3.6). By estimating this function, $\hat{f}$ an estimated Lyapunov exponent can be produced, with $\hat{J}_k$ being derived from $\hat{f}$.

Once this has been calculated, the LE can be analysed by simply seeing if it is positive (signifying chaos) or negative-(signalling non-chaotic behaviour).

This method is designed for small data sets (the user's guide recommends data sets of 500 or less)- however this was written in 1992 so larger data sets can be used instead, although will be very slow. It will be very useful for applying to the daily averages for each year. Since the program runs many models (the manuscript quotes

"if you have to shoot blind, shoot often") , this is a very slow process to select the best models for each time delay (from (3.7)), embedding dimension (as shown in (3.6)) and smoothing parameters (that determine the complexity of the model) to fit to the data set.

The program uses two methods of calculating Lyapunov exponents, one using singular value decomposition (SVD) (which is discussed in McCaffrey et.al *(*1992*)* and the other using QR factorisation (QR), (discussed in Arbarbanel *(*1992)). The reason that two values are calculated is because the SVD estimate has been shown to be positively biased when estimated. The QR value has less bias in most cases.

The LENNS authors write this about the program itself:

"*The program runs a lot of potential models for each of three separate parameters. Firstly there is a time delay (L) that runs from 1 through to 12. There is also the embedding dimension (d) used for each estimate which runs from 1 through 10. And finally k signifies the number of "hidden units" that could be used in the model which runs from 1 through to 8.*

The program outputs the 20 'best' potential fits, for each combination of these parameters, from L=d=k=1 to L=12, d=10, k=8. By fitting 250 parameter sets from a Uniform distribution, the program calculates the RMS (root mean square) for each fit and saves the lowest 20 of these."

From these 20 values the program also produces two values for working out the "best" model for each combination of parameters. The GCV (Generalised Cross Validation) method and the BIC (Bayesian Information Criteria) method are both used for reasons explained below. The BIC is useful in that it explains the goodness of fit of a particular model whilst penalising extra parameters (more so than the Aikaike Information Criterion does) (Schwarz 1978). The GCV is very computationally efficient- when compared to other "leave one out methods" and also relies on less assumptions (for example the Gaussian distribution of errors.

Firstly, the GCV is calculated using:

$$V_c = \left\{ \frac{RMS}{(1 - p\frac{c}{n})} \right\}^2 \qquad (3.8)$$

where p is the number of parameters, c is a fixed constant and n is the number of data points. The standard GCV is worked out by taking c=1.

The BIC uses:

$$BIC = \frac{1}{2}\left\{ 1 + \log(2\pi) + 2\log(RMS) + p\frac{\log n}{n} \right\} \qquad (3.9)$$

It has been shown that in most cases the GCV criterion should be preferred unless linearity has been rejected as the BIC can tend to overfit noisy linear data.

Now a "best" Lyapunov Exponent can be estimated for each particular combination of parameters and analysed.

Techniques recommended for producing summaries of this data are:

- plots of the "best" Lyapunov exponent estimate against the time delay, (as in Giannerini 2004)
- scatterplots of estimated LE vs. GCV or BIC
- plot of estimated LE vs. d for the L of the single best fit.

3.3 Chaos Analysis of Auchencorth Moss Data

3.3.1 Introduction

Two of the methods described above will be used in order to look at the data, the LENNS method, which was described at length in the previous section, and the MLCE method that was described in (3.5). The flyby method (3.4) in Section 3.2.5

looks like it can be too easily affected by noise so will not be used here.

Having two programs will allow a qualitative check on the sulphur dioxide data that is being looked at here. It also means that the "chaoticness" of a system can be analysed at a number of different time scales and it may be of some interest to see if the different years show differences in whether chaos is present or not and also if certain times of the year appear to show common patterns of chaos/non-chaos. Having two different programs calculating estimates for the Lyapunov exponents also means that comparisons can hopefully be achieved and therefore any spurious results can be flagged and corrected.

Firstly, it would be useful to look at the descriptive statistics of the daily measured data from 1996-2001. This is shown in Table 3.1:

	N	N*	Mean	Median	TrMean	StDev	SEMean	Min	Max	Q1	Q3
1997	355	10	-0.09	-0.02	-0.05	0.53	0.03	-9.28	1.26	-0.08	-0.005
1998	340	25	-0.14	-0.02	-0.04	1.53	0.08	-28.05	0.97	-0.06	-0.002
1999	344	21	0.39	0.01	0.02	3.62	0.20	-9.61	50.81	-0.004	0.02
2000	343	23	-1.23	-0.02	-0.09	17.19	0.93	-315.44	15.93	-0.11	0.002
2001	344	21	0.002	0.01	0.01	0.13	0.01	-1.99	0.50	-0.003	0.02

Table 3.1: The summary statistics for years 1997-2001

As can be seen it looks like there may be some problems with missing data. To apply the data to the program these missing values will have to be imputed. Looking at the histograms in Figure 3.1 along with the fact the data set contains very few missing values, it looks reasonable to impute the missing data by taking random samples from the Normal distribution (and using the sample mean and variance that has been calculated from the non-missing observations).

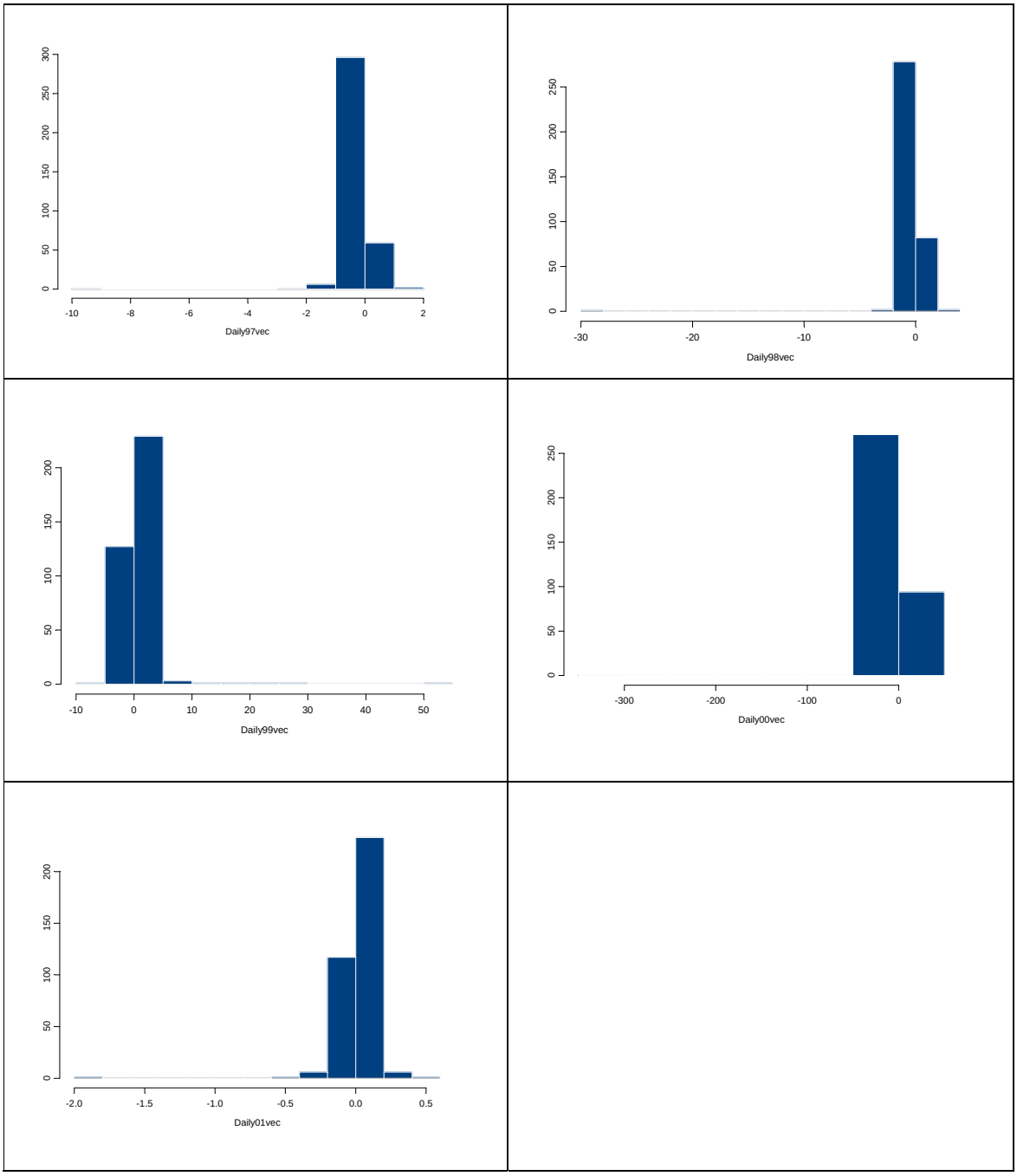


Figure 3.1: The daily data from 1997 to 2001

The new samples are shown below in Table 3.2 for the same years in order to show that they do not differ too much from the original “real” data

	N	Mean	Median	TrMean	StDev	SEMean	Min	Max	Q1	Q3
1997	365	-0.09	-0.02	-0.05	0.54	0.03	-9.28	1.26	-0.08	-0.004
1998	365	-0.14	-0.02	-0.05	1.53	0.08	-28.05	4.49	-0.07	-0.001
1999	365	0.41	0.01	0.05	3.62	0.19	-9.61	50.81	-0.01	0.03
2000	366	-1.22	-0.03	-0.14	16.64	0.87	-315.44	15.93	-0.16	0.001
2001	365	0.002	0.006	0.006	0.13	0.01	-1.99	0.50	-0.004	0.02

Table 3.2: The summary statistics for years 1997-2001 with imputed values.

3.3.2 Chaos Estimation Results

The LENNS program can then start working on these data in their revised format. Below are the results of the daily data from 1996 – 2001. These have been plotted so as to show the best model (based on the lowest logged GCV) for each time lag (L) in the data set. The different lines in the plot correspond to different embedding dimensions. These are shown in Figure 3.2:

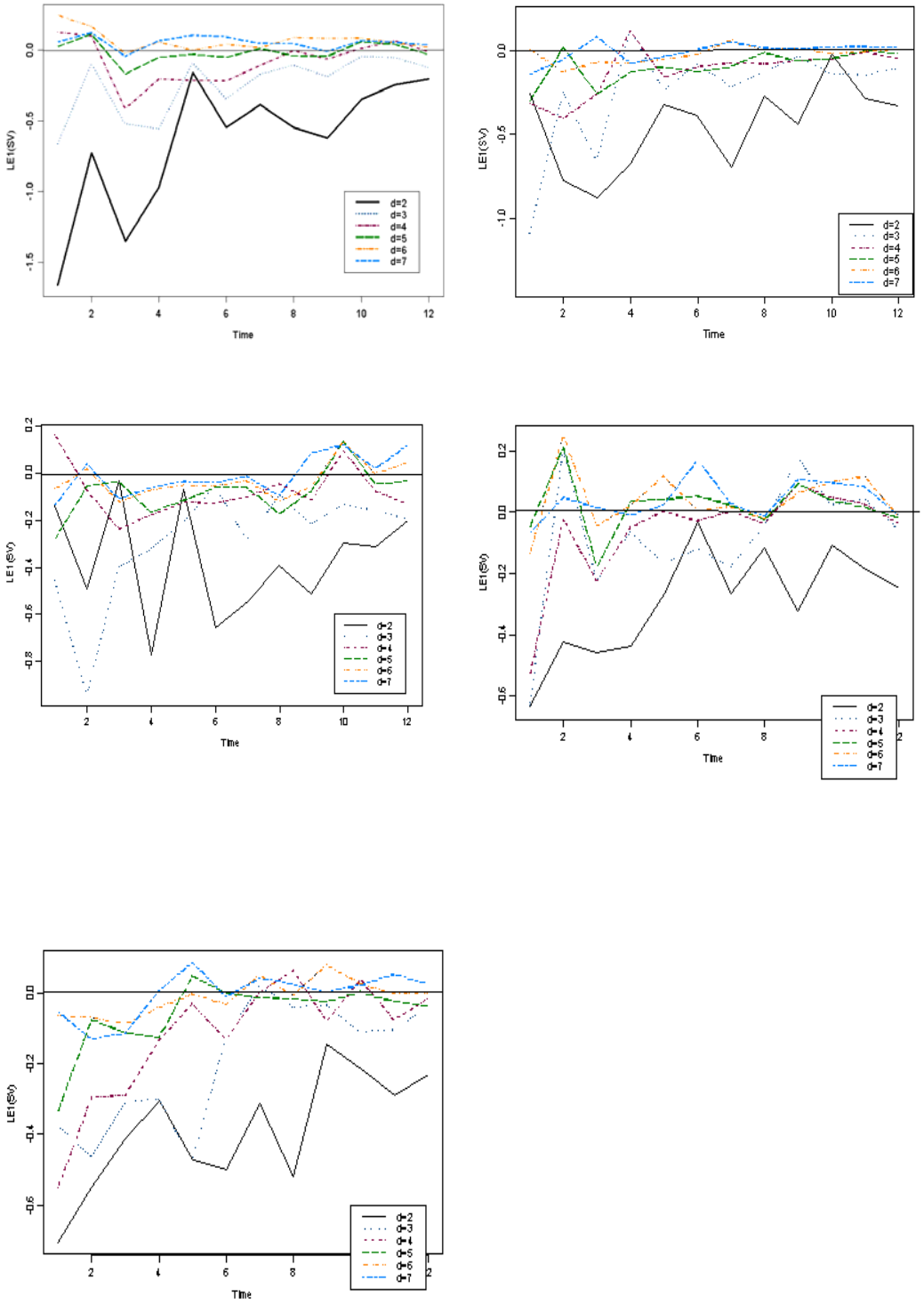


Figure 3.2: Daily LEs against time lags (1-12days). Each graph represents year of daily data from 1997-2001. There is a line drawn across at LE=0.

There are some important things to be inferred from these. Firstly it can be seen that most of the data points lie under the line at LE=0 indicating that there is non-chaotic

behaviour at most of these points. It can also be seen that in general most of the lines seem to follow a similar pattern with the exception of d (the embedding dimension) being equal to 2. In 2000 perhaps it can be seen that there may be more chaotic behaviour perhaps than in the other years as the plot shows more of the LEs are above zero.

The LEs obtained by the LENNS program give some indication of the level of chaotic behaviour or non-chaotic behaviour at each time lag, corresponding to each embedding dimension. It looks like when the embedding dimension is set to 2 this appears to give very different results than the rest of the dimensions (3-7).
It would be of interest to look at the time series of the five years of data all together and see what kind of chaos this produced even with a larger time lag to it. By increasing the time lag to, say, 24 instead of 12, this produces the graph below (Figure 3.3) along with the table listing the LE estimates at each time lag (Table 3.3):

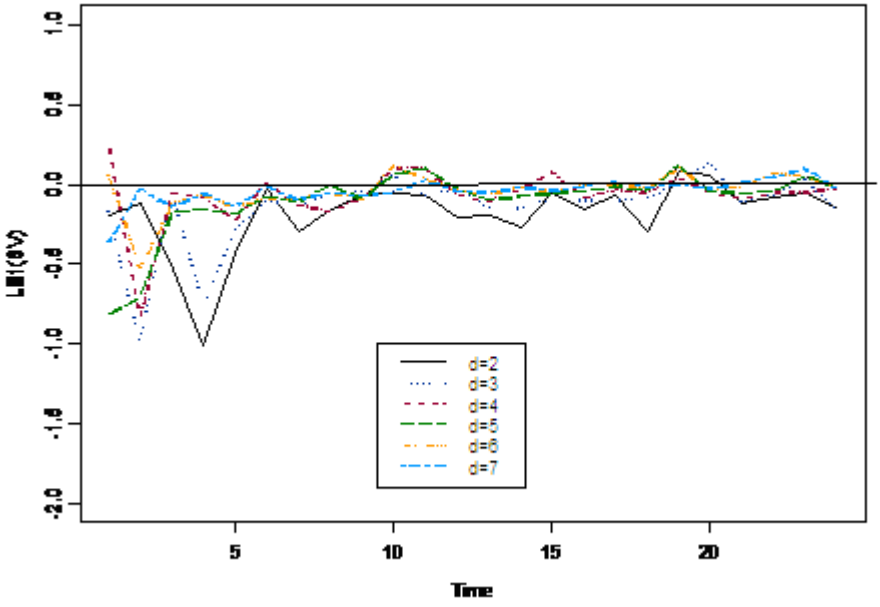


Figure 3.3: Time lag (x-axis) against Lyapunov Exponent (y-axis) for 5 years 1997-2001, with embedding dimensions 2-7.

	d=2	d=3	d=4	d=5	d=6	d=7
Time Lag=1	-0.195	-0.171	0.251	-0.820	0.055	-0.364
Time Lag=2	-0.120	-0.996	-0.836	-0.704	-0.534	-0.029
Time Lag=3	-0.538	-0.041	-0.060	-0.173	-0.122	-0.144
Time Lag=4	-1.005	-0.768	-0.081	-0.162	-0.073	-0.057
Time Lag=5	-0.433	-0.292	-0.226	-0.187	-0.148	-0.142
Time Lag=6	-0.036	-0.108	0.008	-0.073	-0.089	-0.011
Time Lag=7	-0.296	-0.139	-0.134	-0.109	-0.089	-0.101
Time Lag=8	-0.152	-0.079	-0.175	-0.004	-0.056	-0.054
Time Lag=9	-0.063	-0.048	-0.085	-0.094	-0.099	-0.080
Time Lag=10	-0.057	0.059	0.091	0.061	0.113	-0.048
Time Lag=11	-0.068	-0.78	0.104	0.100	0.034	0.024
Time Lag=12	-0.206	-0.038	-0.069	-0.027	-0.042	-0.044
Time Lag=13	-0.197	-0.147	-0.112	-0.099	-0.067	-0.058
Time Lag=14	-0.271	-0.160	-0.037	-0.078	-0.020	-0.019
Time Lag=15	-0.062	-0.094	0.080	-0.055	-0.049	-0.044
Time Lag=16	-0.152	-0.107	-0.086	-0.051	-0.010	-0.018
Time Lag=17	-0.072	-0.109	-0.040	-0.010	-0.014	0.019
Time Lag=18	-0.290	-0.089	-0.056	-0.043	-0.022	-0.024
Time Lag=19	0.075	-0.007	0.040	0.117	0.101	-0.006
Time Lag=20	0.055	0.136	-0.051	-0.045	-0.037	-0.033
Time Lag=21	-0.114	-0.117	-0.095	-0.059	-0.020	0.011
Time Lag=22	-0.082	-0.071	-0.057	-0.048	0.064	0.045
Time Lag=23	-0.053	0.029	-0.055	0.053	0.043	0.095
Time Lag=24	-0.153	-0.162	-0.029	-0.029	-0.028	-0.023

Table 3.3: The Lyapunov Exponents that are in Figure 3.5, with d going from 2 to 7.

Again this is showing that there is not much chaotic behaviour about the daily time series. This appears to add more credence to the theory that the data is just very noisy, and therefore it would probably be wise to try and filter out this "noise" and try and extract the signal that is being picked out by the LENNS program.

In order to verify the results of this method, it would be useful to look at another method of estimating Lyapunov exponents from the data. The method explained in Giannerini and Rosa (Equation 3.5) will also be applied to the data and then these results can be checked against one another.

	1997	1998	1999	2000	2001
d=2	-0.0139	-0.0413	-0.01207	0.00518	-0.00448
d=3	-0.00388	-0.0385	0.02142	-0.00284	-0.00363
d=4	-0.00783	-0.0351	0.02222	0.00128	0.00616
d=5	-0.00392	-0.0320	0.02164	0.00506	0.00439
d=6	-0.00648	-0.0277	0.02316	0.00945	0.0135
d=7	0.00411	-0.0237	0.02465	0.00791	0.0178

Table 3.4: The LEs for the different embedding dimensions (2-7) for each of the years 1997-2001 using the MLCE method

It can be seen that these two methods both show different results- the MLCE program is estimating very small slopes and is showing chaotic behaviour in the final 3 years as shown by Table 3.4. Though both programs show very small values that fluctuate between positive and negative values for different embedding dimensions.

There would be a strong case for giving more weight to the LENNS result as it is searching through many models to pick out the best ones, and also is designed to deal with a lot of noise. It was seen in Chapter 2 that the flux data looks very noisy (Section 2.4 showed very small Signal-to-Noise ratios). It may not be very useful on very large sample sizes but whilst looking at the daily values it is certainly reasonable to use these results.

3.3.3 Checking Modelled Results for Chaos

A good approach to go for next would be to look at how the LENNS program and the MLCE program work on the modelled EMEP daily results that have been described in Chapter 1. Since these data points have been modelled then it would be expected that these would show no chaos in them. However it has been seen that the model is a very complicated one, bringing many different modelled data into it. It would also be interesting to see whether or not the LEs are similar to the measured data for each of the years concerned. Table 3.5 and Figure 3.4 show the MLCE method followed by the LENNS method for each of the four years 1997 and 1999-2001 so that they can be compared to one another.

	1997	1998	1999	2000	2001
d=2	0.0122	N/A	0.00721	0.0100	0.00494
d=3	0.00888	N/A	0.00971	0.0155	0.00962
d=4	0.00736	N/A	0.0165	0.0188	0.0184
d=5	-0.00106	N/A	0.0352	0.0337	0.0292

Table 3.5: Lyapunov Exponents estimated by the MLCE program for the EMEP modelled data from 1997-2001 (with embedding dimensions 2-5)

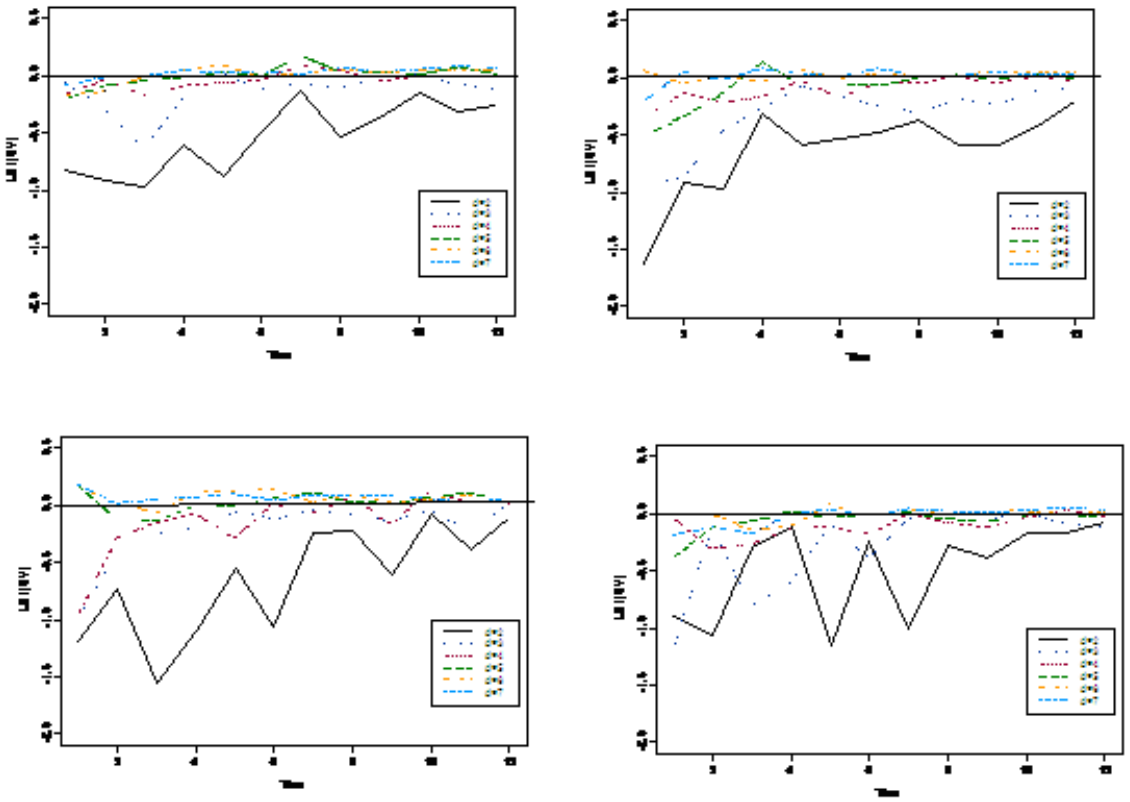


Figure 3.4: LENNS time-series plots for 1997,1999,2000, 2001 showing LEs for EMEP modelled data, d=2,…7. Time goes from 1-12 days on x axis. A line is drawn at LE=0

Both of these methods show that the values mostly remain around zero for most of the time intervals, which would be expected since the data points do come from a model. The MLCE method tends to show small positive values for the EMEP data. This is slightly concerning since the data is modelled and therefore should show no levels of chaos in its results. The LENNS method stays below or around zero for nearly all values of d and so again looks a more reliable measure of the chaos/non-chaos in the system.

3.3.4 Refining the Timescale

3.3.4.1 Diurnal Fluxes

It would be interesting to know if refining the timescale to a finer scale would change the level of chaos detected in the data. Perhaps it may be interesting to look at the diurnal cycle of fluxes. This can be done crudely by dividing the year into 6 months of "winter" (October-March inclusive) and 6 months of "summer" (April-September inclusive). During the winter the assumption will be that "day" falls between 8am and 6pm and night at all other times. During the summer- day will be taken from 6am till 8.30pm and night for the rest of the time. As has been mentioned this is just a crude estimation, but it should be enough to pick out any changes in the Lyapunov analysis over the same years as previously measured. Again, there will be missing data to impute into the data set, so it will be useful to see how much or little effect this looks to have on the data set that will be used in the LENNS program.
In order to study this, the diurnal data will take the average value from each "night period" and "day period" to create 730 (or 732) data points (i.e. 365 days and nights).

Running this on LENNS and the MLCE program give these results shown in Figure 3.5:

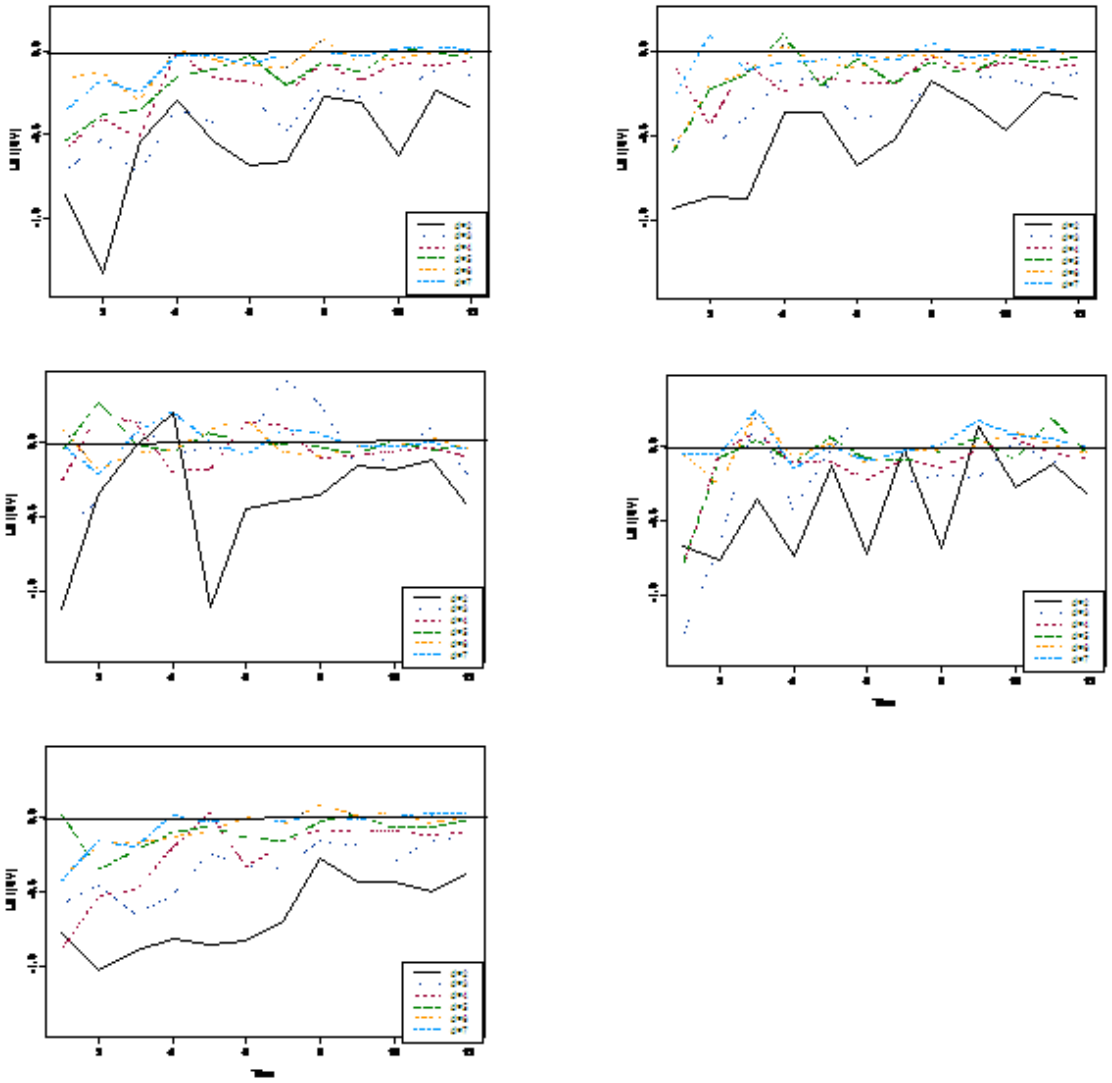


Figure 3.5: Plots showing the diurnal time series plots with time lag on x-axis and Lyapunov Exponents on y-axis from years 1997-2001 using LENNS program. The time goes from 1-12 on the x axis and a line is shown at LE=0

These methods again show differences between the conclusions they provide. The first point to make when looking at the MLCE method is that taking a slope estimate from this looks as if it will not be useful since there does not appear to be an obvious linear relationship on any of the plots.

When taking these slope estimates though they all seem to show chaotic behaviour over the series as a whole.

The LENNS analysis shows something different with most of the exponents lying in the non-chaotic "half" of the graph. Therefore these methods appear to contradict each other here.

In summary, after looking at the three different time periods above using the two methods (MLCE and LENNS) it seems that they do not give similar results (except for perhaps the modelled data). Since LENNS is designed to work on noisy data sets, then the results from this should perhaps be looked at with more importance than the MLCE data. From Section 3.3 this looks reasonable.

3.3.4.2 Hourly Data

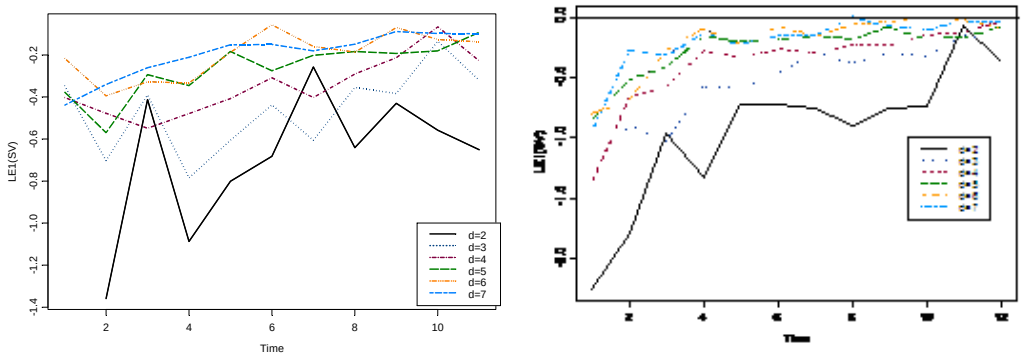


Figure 3.6: The hourly time series plots with time lag on x-axis and Lyapunov Exponents on y-axis from years 1997 and 2000 using LENNS program. There is a line at LE=0 on the second plot but the first stays below 0.

The LENNS program was also used on the hourly flux data. When Nychka et.al designed the LENNS program they were expecting that it should be used on small data sets, since it tests so many models for each time lag, embedding dimension and hidden units. When it was attempted on these hourly data sets (with over 8000 values), it was only able to produce plots for the two years pictured. The hourly values in fact show more negative values than in the daily data. However it does appear to suggest again that the data can be modelled.

3.4 Conclusion

This chapter has shown the advantages and disadvantages of measuring, or at least trying to measure chaotic signals in a data series. The advantages in the main are that, if chaos is found to be present, then time will not be wasted in an attempt to search for a (non-existent) model. Ways of calculating these chaotic signals come from the estimation of Lyapunov exponents. Three methods are described in this paper. However, there are arguments against the use of these methods, as papers have pointed out that stochastic noise can signal chaotic behaviour even in non-chaotic systems. By using the LENNS method that Nychka et.al provide though, noise can be accounted for by the use of a neural network method that applies lots of potential models to the data and chooses the best ones.

From analysis of the data, these methods have produced some disappointing results however. In most cases it appears that the methods for estimating levels of chaos in systems are not producing consistent results and this probably lends credence to Timmer et.al.'s criticism levelled in Section 3.3.2 earlier. Most of the "noise" that exists in the data, coupled with the very low measurements made are making it extremely difficult for any of the techniques to pick out the signal. The fact that the MLCE technique and the LENNS technique appeared to produce different sets of results when analysing the data suggests that it would be very dangerous to make any firm sets of conclusions about whether there is an existence of chaotic behaviour in the data or not.

Although the results obtained from here have been discouraging, it has not ruled out the chances that the measured data are able to be modelled. The next chapter will look at the fact that part of the reason the two data sets are struggling to match may be the very high values that are being measured in some half hourly periods of the day. It may be useful to look at some of these results separately in order to see if there is anything interesting about them.

Chapter 4 – Extreme Value Analysis

4.1 Introduction

It has been seen in the previous analyses of the data that there appears to be a number of observations which could be considered to be "extreme". It would be of interest to study these extreme values in more detail to see if there is any particular model lying behind these that could explain their occurrences.

Many people started looking at Extreme value theory (EVT) in the 1970s in order that particular families of models might potentially explain unusually high or low observations. Many of these have been used on a variety of real-life examples. Smith (1990) writes on modelling extremes in a ground level ozone situation, Chan and Gray (2006) modelled electricity spot prices, whereas Coles (2001) modelled closing prices of the Dow Jones Index, engine failure times, sea levels and daily rainfall. Coles et.al. (2003) also used EVT to show how looking at maximum rainfall levels could be used to predict future extreme rainfall, using a flood in Venezuela in 1999 to show how it would work. Salmon (2004) shows how EVT can be used to predict housing market crashes. Fernandez (2007) uses an approach which isolates the extreme values to compare between 10 exchange rates to see if they have similarities in their extremes as they fluctuate throughout the year and Soja and Starkel (2007) look at the clustering patterns in the extreme rainfalls of the Himalayas. Clustering of extremes is an interesting issue and one that could be useful for the flux data.

All of these sets of data have been analysed in a very similar fashion, using the same families of models and as such the two main methods along with the family of distributions that are used are detailed below. These are the Generalised Extreme Value Distribution and the Generalised Pareto Distribution. The former uses a series of block maxima and uses these as the "extreme" values to which the model then fits, whereas the latter allows the use of the raw data and picks values higher than a

suitable threshold in order to model these. Both methods will be introduced and applied to the Auchencorth flux values and from these results (and suitable diagnostic plots) it can be decided which of the methods appears to be more appropriate.

4.2 Generalised Extreme Value Distribution

Looking at this method firstly, it is required that the data are put in the form of a series of maximum values (4.1)

$$M_n = \max\{X_1,...,X_n\} \tag{4.1}$$

where $X_1, \ldots X_n$ is a sequence of independent random variables which have a common distribution F. The distribution of M_n can be derived in theory for all values of n, but only in terms of F^n which is unknown. Using classical techniques to estimate F are usually not suitable here as small discrepancies in the estimate can lead to larger discrepancies in F^n. This is why Coles (2001) and others suggest the use of a family of models which are estimated on the extreme data alone. However, firstly M_n has to be "normalised" in order that it does not degenerate to a point mass. The normalisation is shown as:

$$M_n^* = \frac{M_n - b_n}{a_n} \tag{4.2}$$

for sequences of constants $\{a_n > 0\}$ and $\{b_n\}$. By choosing these carefully the difficulties shown above should not arise. Therefore a family of models is chosen for M_n^* rather than M_n.

It can be shown that if there are the sequences $\{a_n\}$ and $\{b_n\}$ then M_n^* belongs to one of the following families:

(a) $G(z) = \exp\left\{-\exp\left[-\left(\frac{z-b}{a}\right)\right]\right\}$

(b) $$G(z)=\begin{cases}0 & z\le b\\ \exp\left\{-\left(\dfrac{z-b}{a}\right)^{-\alpha}\right\} & z>b\end{cases}$$

(c) $$G(z)=\begin{cases}\exp\left\{-\left[-\left(\dfrac{z-b}{a}\right)^{\alpha}\right]\right\} & z<b\\ 1 & z\ge b\end{cases} \qquad (4.3)$$

for parameters a>0, b and α>0. These are individually known as the Gumbel, Frechet and Weibull distributions respectively.

There are two weaknesses to having three models for the extreme values. Firstly, there needs to be a technique in order to choose which particular model should be used in order to estimate the relevant parameters. Then from this any subsequent analysis would have to assume this decision to be correct and would not allow any uncertainty of this choice.

Therefore it is far better to reformulate the models above into one single family of models:

$$G(z)=\exp\left\{-\left[1+\xi\left(\frac{z-\mu}{\sigma}\right)\right]^{\frac{-1}{\xi}}\right\} \qquad (4.4)$$

valid on the set $\left\{z:1+\xi\frac{(x-\mu)}{\sigma}>0\right\}$,

1 The scale parameter $\sigma>0$

2 The location parameter $-\infty<\mu<\infty$

3 The shape parameter $-\infty<\xi<\infty$

G(z) is defined as the Generalised Extreme Value (GEV) family of distributions. It is relatively simple to check that this family contains all 3 distributions shown in (4.3), by choosing ξ>0 and ξ<0 for the Frechet and Weibull distributions. By using the limit $\xi\to 0$ for the case ξ=0, leads to the Gumbel Distribution.

By combining these data into one family of models, the problems listed above will disappear as some appropriate inference on an estimate for ξ will immediately show which particular model is most suitable. Also the uncertainty in the estimate will give a convenient measure of the uncertainty in the model choice.

All that remains now is to decide on a method of estimating each parameter in the GEV model. This can be done by Maximum Likelihood Estimation (MLE). One problem with this can be at the end points of the GEV distribution, (Smith 1985), who showed that because these end points are a function of the parameter values, then $\mu-\sigma/\xi$ is an upper end point of the distribution when $\xi<0$ and a lower end point when $\xi>0$. Smith managed to simplify this to 3 cases:

1 $\xi>-0.5$, this gives regular ML estimators that have the usual asymptotic properties.
2 $-1 < \xi < -0.5$ This generally gives ML estimators but they do not have the usual asymptotic properties.
3 $\xi < -1$ ML estimators are unlikely to be obtainable. (4.5)

However, the final two situations are not often encountered as they cover distributions of data with short upper tails. Certainly in the Auchencorth Moss data this should not be a problem.

In order to calculate the MLE's for each of the three parameters in the GEV distribution shown in (4.4), the log likelihood to be maximised is shown below in (4.6):

$$l(\mu,\sigma,\xi) = -m\log\sigma - \left(1+\frac{1}{\xi}\right)\sum_{i=1}^{m}\log\left[1+\xi\left(\frac{z_i-\mu}{\sigma}\right)\right] - \sum_{i=1}^{m}\left[1+\xi\left(\frac{z_i-\mu}{\sigma}\right)\right]^{\frac{-1}{\xi}} \quad (4.6)$$

assuming that

$$1+\xi\left(\frac{z_i-\mu}{\sigma}\right)>0 \text{ for i = 1,…,m and } \xi \neq 0$$

If $\xi = 0$ then a different log-likelihood is obtained from the Gumbel Distribution

$$l(\mu,\sigma) = -m\log\sigma - \sum_{i=1}^{m}\left(\frac{z_i-\mu}{\sigma}\right) - \sum_{i=1}^{m}\exp\left\{-\left(\frac{z_i-\mu}{\sigma}\right)\right\} \tag{4.7}$$

The GEV distribution provides a model for the distribution of block maxima. In the case of the Auchencorth data this seems suitable as it means that the daily/weekly/monthly maximum values could all be analysed which could prove useful in trying to see if the extreme values are following this particular distribution.

After the MLE estimates for the three parameters $(\hat{\mu},\hat{\sigma},\hat{\xi})$ are obtained, there needs to be a way of firstly checking how well this particular model and the data agree. This can be achieved by analysing both probability plots and quantile plots.
Once the model choice has been verified, it would be useful to produce a returns level plot. The returns level is obtained by inverting the GEV distribution (4.4) as shown below in (4.8):

$$z_p = \begin{cases} \mu - \frac{\sigma}{\xi}\left[1-\{-\log(1-p)\}^{-\xi}\right] & \xi \neq 0 \\ \mu - \sigma\log\{-\log(1-p)\} & \xi = 0 \end{cases} \tag{4.8}$$

This return level (z_p) is exceeded by the maximum in a particular year with probability p.

This is useful as z_p can be plotted against $\log y_p$ - which in the second (Gumbel) case will give a linear plot. Else, if $\xi<0$ or $\xi>0$, the plot will converge to $\mu-\sigma/\xi$ or have no finite bound respectively.

4.3 Generalised Pareto Distribution

There are other methods rather than only using block maxima in order to model the extreme values in a data set. One technique that is widely used involves using a complete data set and only modelling the values that occur above a particular threshold.

It would be relatively simple to define a function for modelling particular values over a certain threshold using basic probability as shown in (4.9):

$$\Pr\{X > u + y \mid X > u\} = \frac{1 - F(u+y)}{1 - F(u)}, \quad y > 0 \tag{4.9}$$

However this would require knowing F. In practical applications this is generally not the case, and so approximations for F should be made, (similar to the GEV distribution).

Additionally, the GEV distribution (4.10) can be used, but altered in order to find a distribution function for (X-u) conditional on X>u, which gives approximately:

$$H(y; \tilde{\sigma}, \xi) = 1 - \left(1 + \frac{\xi y}{\tilde{\sigma}}\right)^{\frac{-1}{\xi}} \tag{4.10}$$

which is defined on $\left\{ y : y > 0 \text{ and } \left(1 + \frac{\xi y}{\tilde{\sigma}}\right) > 0 \right\}$

where $\tilde{\sigma} = \sigma + \xi(u - \mu)$

(4.10) is defined as the Generalised Pareto Family (GPF). This implies that if the block maxima can be approximated by the GEV distribution then the excesses can be modelled by the GPF. In fact ξ is equal in both the GEV and GPF cases, and the same three cases shown in (4.10) also apply here.

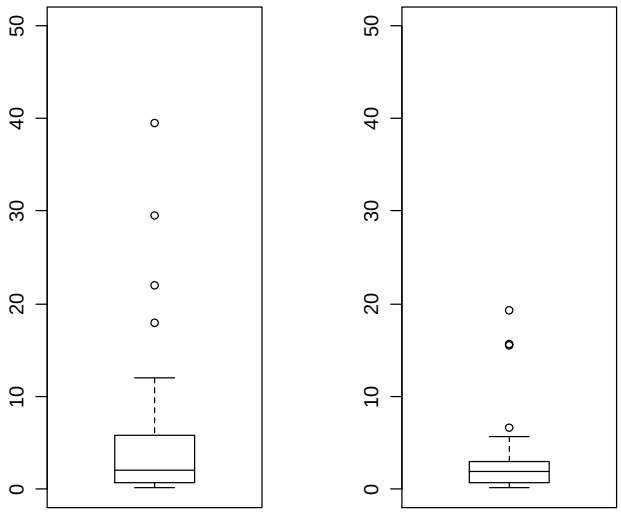
The only difficulty that arises here is the choice of the threshold. If the threshold is chosen to be too low, then the model may be biased, due to violation of the

asymptotic basis of the model. If it too high, then there will be too few data points, leading to high variance in each of the parameter estimates. Adopting low thresholds is the standard procedure in real life examples however and by estimating parameters for a range of threshold values will allow the influence of the threshold value to be made clear.

4.4 Analysing the Auchencorth Data

4.4.1 Analysis using the GEV Family

The daily maximum values from Auchencorth Moss will be used first to try and fit a GEV model. This is suitable since the SO_2 appears to have no seasonal variation



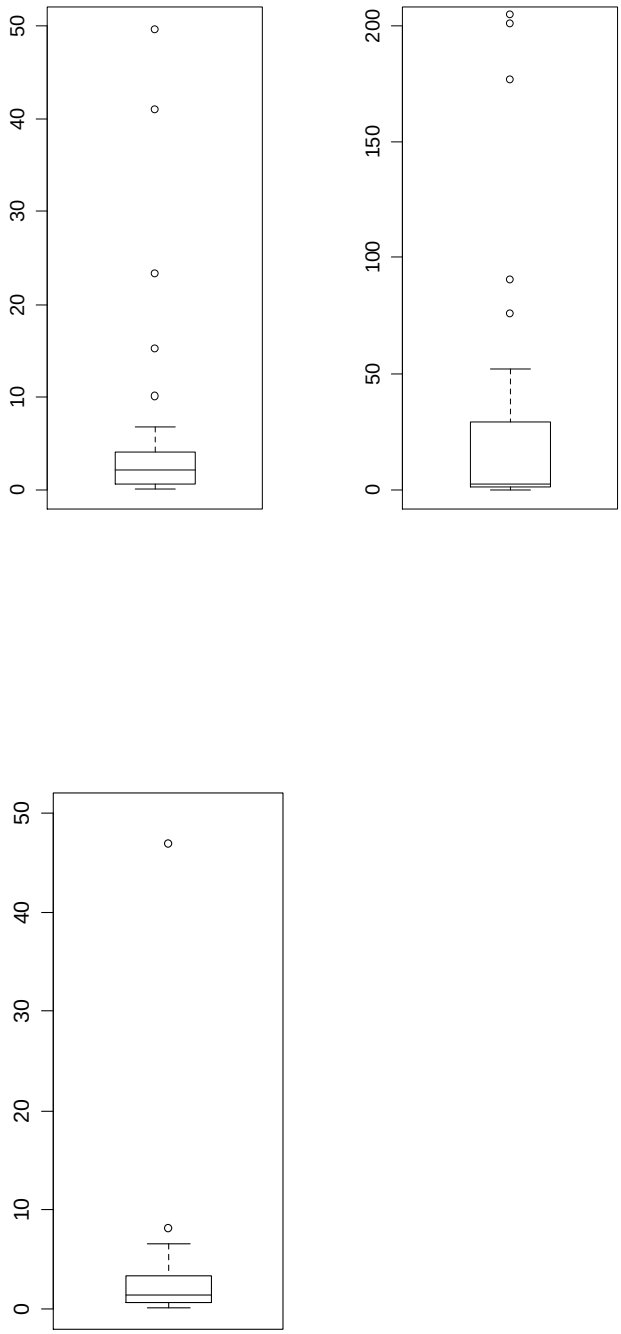


Figure 4.1: Boxplots of the weekly maximum values for years 1997-2001

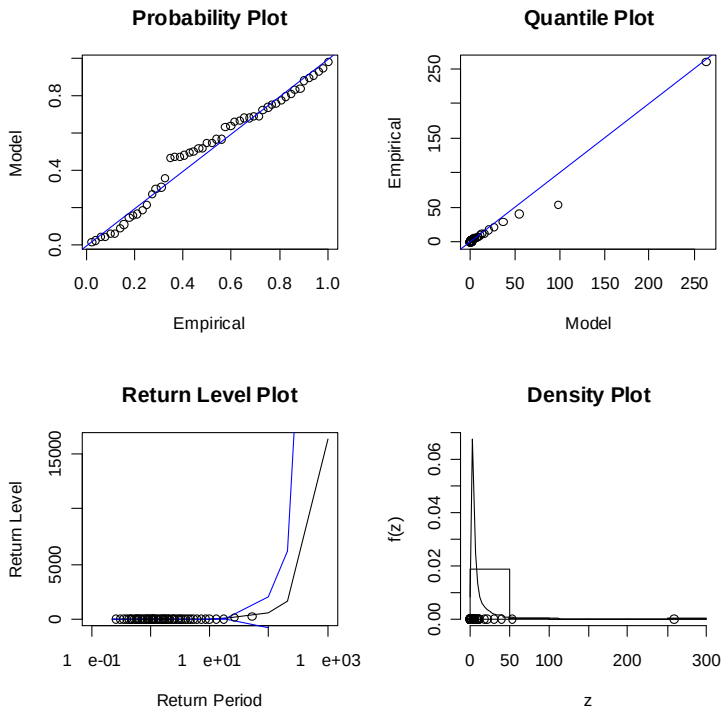
From the GEV distribution the values obtained for each of the three parameters along with the negative log-likelihood from a Maximum Likelihood Estimate calculation, are shown in Table 4.1:

Year	Negative Log-Likelihood	$\hat{\mu}$(std error)	$\hat{\sigma}$(std error)	$\hat{\xi}$(std error)
1997	140.006	0.993 (0.249)	1.448 (0.391)	1.399 (0.275)
1998	123.356	0.853 (0.190)	1.155 (0.281)	1.223 (0.230)
1999	141.797	0.935 (0.228)	1.411 (0.377)	1.424 (0.246)
2000	196.197	1.581 (0.441)	0.996 (0.189)	2.003 (0.265)
2001	100.177	0.920 (0.177)	2.003 (0.265)	0.7140 (0.216)

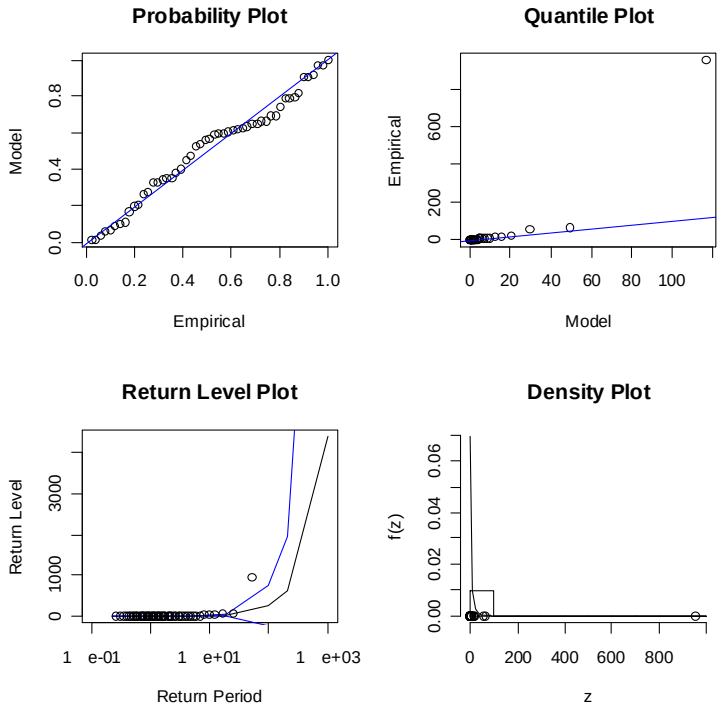
Table 4.1: MLE of each parameter of the GEV distribution described in (4.4) along with their standard errors and the negative log-likelihood of the model.

Coles (2001) suggests a group of diagnostic plots should be produced in order to check the model that has been used. These are shown for the data in Table 4.1, in Figure 4.2:

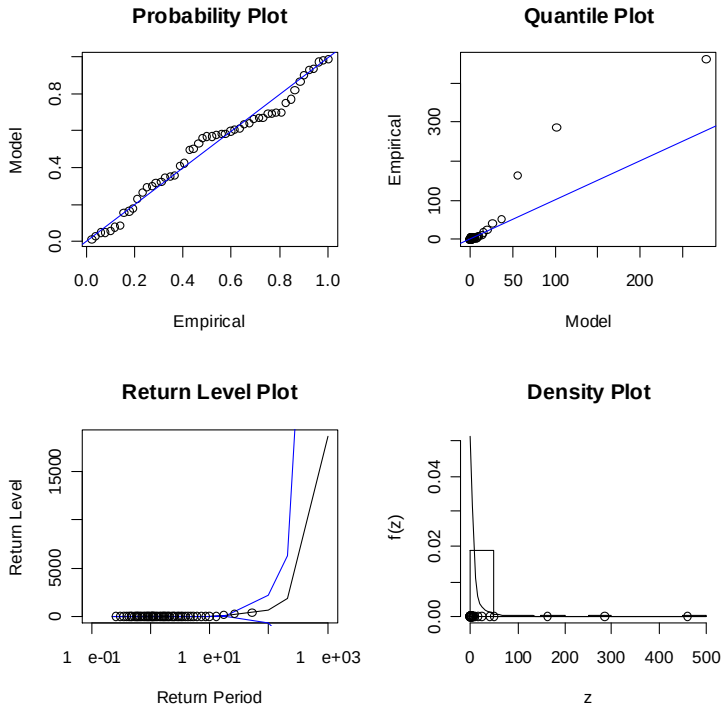
1997



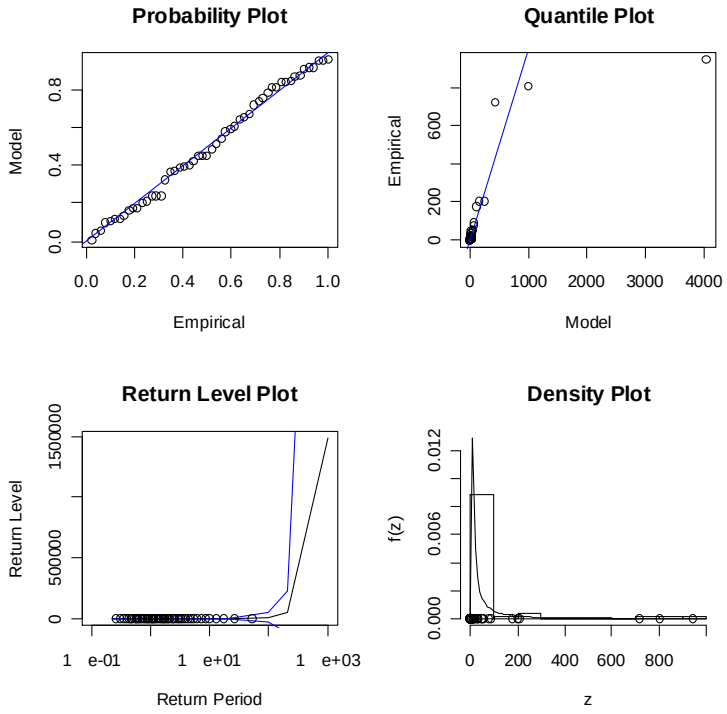
1998


Probability Plot
Model
Empirical
0.0
0.4
0.8
0.0
0.2
0.4
0.6
0.8
1.0
Quantile Plot
Empirical
Model
0
200
600
0
20
40
60
80
100
Return Level Plot
Return Level
Return Period
0
1000
3000
1 e-01
1 e+01
1 e+03
Density Plot
f(z)
z
0.00
0.02
0.04
0.06
0
200
400
600
800

1999


Probability Plot
Model
Empirical
0.0
0.4
0.8
0.0
0.2
0.4
0.6
0.8
1.0
Quantile Plot
Empirical
Model
0
100
300
0
50
100
200
Return Level Plot
Return Level
Return Period
0
5000
15000
1 e-01
1 e+01
1 e+03
Density Plot
f(z)
z
0.00
0.02
0.04
0
100
200
300
400
500

2000



2001

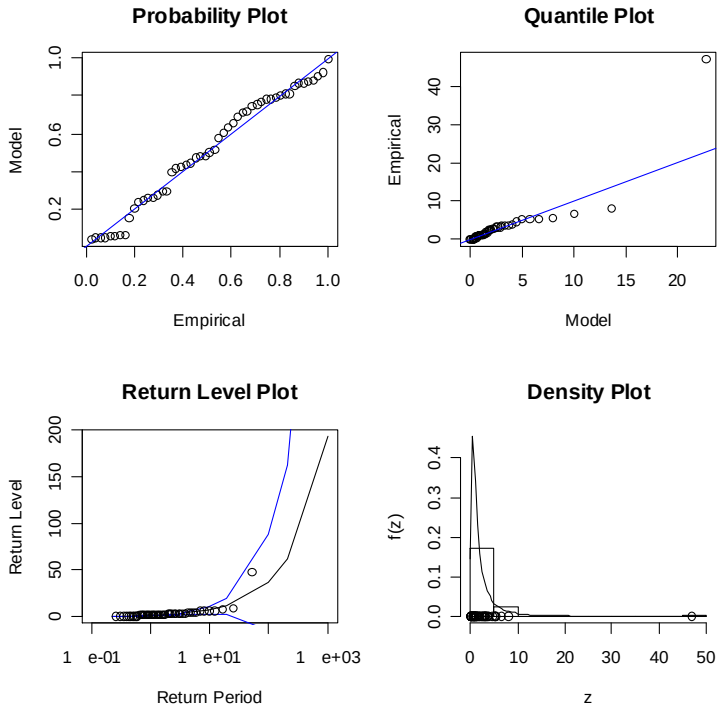


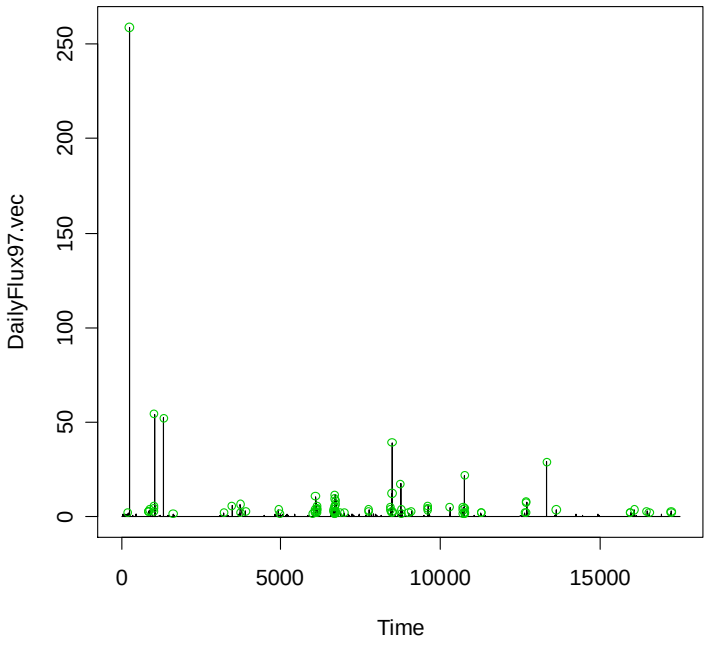
Figure 4.2: Probability Plot, Quartile Plot, Return Level Plot and Density Plot for the MLE of the GEV model fitted using parameters in Table 4.1 for years 1997-2001

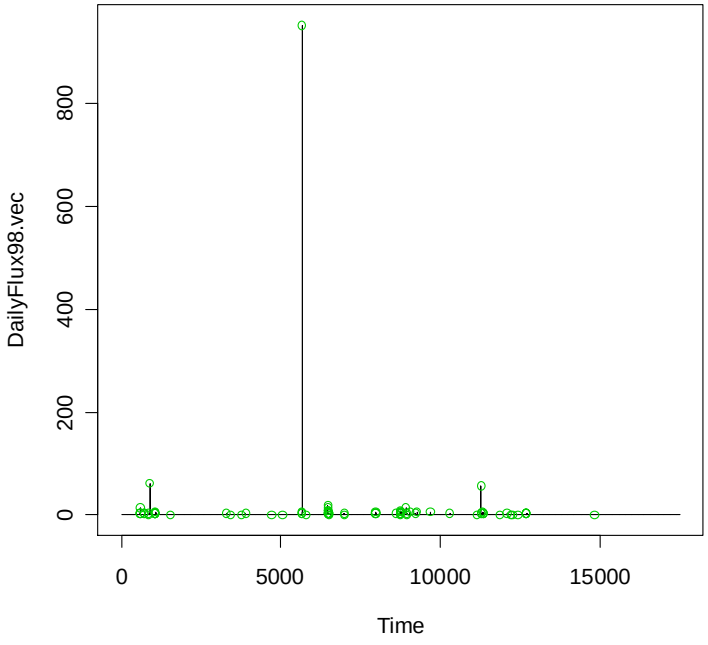
It can be seen from Table 4.1 that each of the parameters for the GEV distribution for all 5 years are significantly different from zero. It can also be seen that ξ is positive in each of the five years and as such it can be assumed that these ML estimators have the usual asymptotic properties in this case. (4.5).

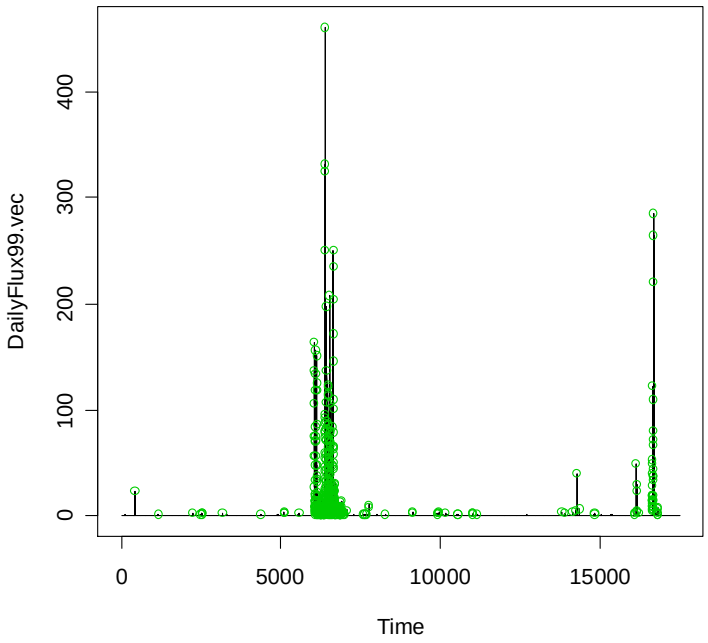
From the diagnostic plots, while the probability plots appear to fit the straight line reasonably well, the quartile plots for 1998-2001 appear to deviate a lot from the line. These in most cases appear to be due to one point in particular being so much bigger than the rest (a fact that is reflected in the density plots also produced). This gives some doubt to a GEV model possibly being suitable for modelling the extreme data that has been collected- and perhaps suggests that a different one should be used instead. Looking at the data once more, perhaps logarithms of the data could be used in order to fit the model slightly better. This may be something to think about in the future.

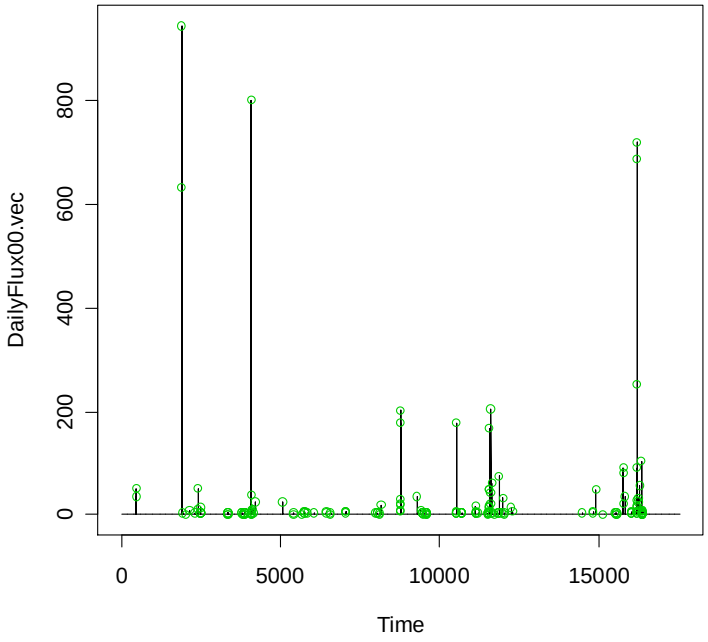
4.4.2 Analysis Using the GPF distribution

Using the GPF, a decision should be made on the choice of the threshold as mentioned previously. From Table 4.2 it can be seen that a threshold choice of $2\mu gSm^{-3}$, means that most of the data is filtered out and only a small percentage remains to be analysed. In this example the half-hourly data for each year will be analysed. As missing data can prove to be a problem – any missing values have been replaced with a value below the threshold so as not to affect the data used to calculate the parameters in the Pareto distribution. It will also be useful though to see what effect a change of this threshold will have on the parameters estimated. Figure 4.3 firstly shows the data that has been collected and the points above $2\mu gSm^{-3.}$ This will also be useful in that it can be seen whether the extreme values are clustering together or are appearing at "random" points.


DailyFlux97.vec
0
50
100
150
200
250
0
5000
10000
15000
Time


DailyFlux98.vec
0
200
400
600
800
0
5000
10000
15000
Time


DailyFlux99.vec
0
100
200
300
400
0
5000
10000
15000
Time


DailyFlux00.vec
0
200
400
600
800
0
5000
10000
15000
Time

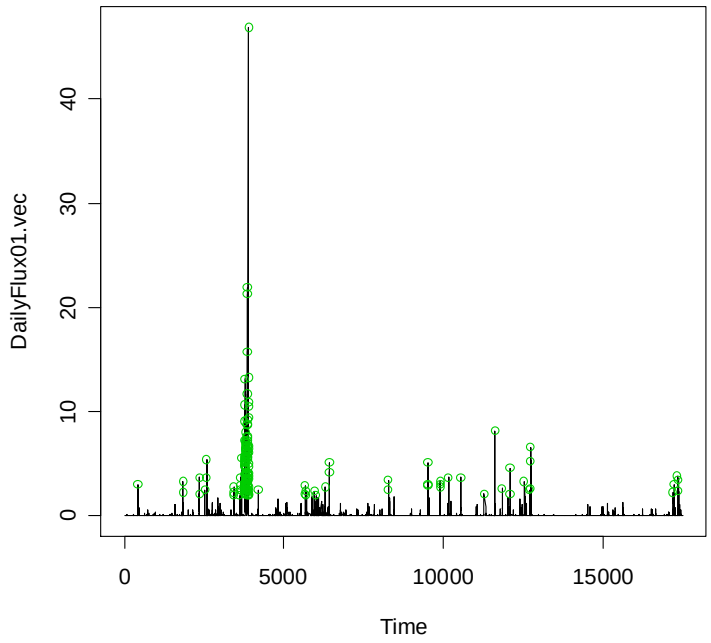


Figure 4.3: Time Series plots of the daily flux values. Any missing values have been imputed with the value 0.01 since only values above 2μgSm^{-3} will be considered in the Pareto model. The 'extreme' points are indicated in green.

From Figure 4.3 it can be observed that the extreme points (based on the definition given here) appear to be spread throughout the year in each of the 5 years that are being looked at. 1999 is more difficult to analyse since (as it has been shown in previous chapters and in Table 4.2) there are a lot of higher fluxes that fall into the extreme category than in any of the other years. Section 4.5 will look more at whether the extreme data are clustering anyway, so for now an attempt will be made to model the extreme values.

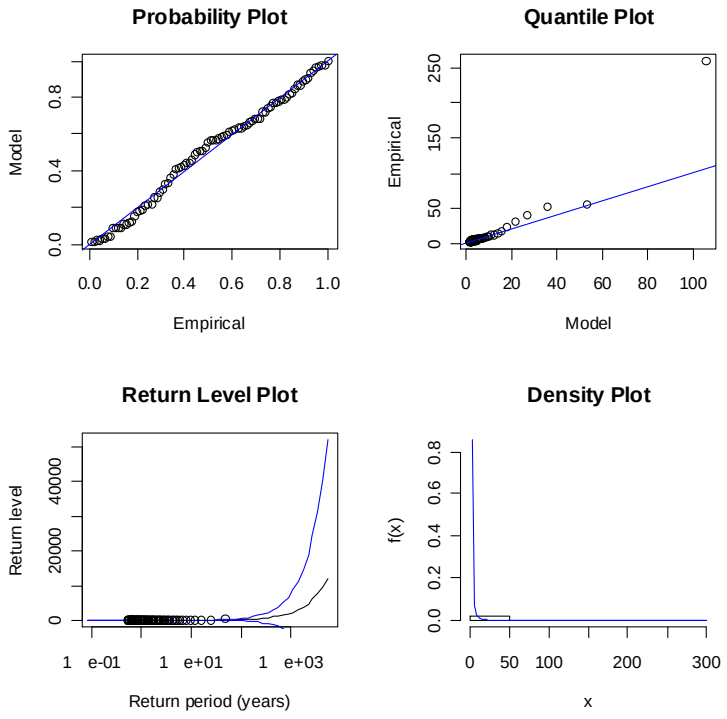
From the Pareto Distribution the values obtained for each of the two parameters along with the negative log-likelihood from an MLE calculation and the percentage of points above the chosen threshold of 2μgSm^{-3}.

Year	% of data above threshold	Negative Log-Likelihood	$\hat{\sigma}$ (std error)	$\hat{k}$ (std error)
1997	0.5%	190.15	1.171 (0.259)	1.002 (0.221)
1998	0.5%	165.09	1.067 (0.221)	0.973 (0.205)
1999	4%	2528.10	4.120 (0.319)	1.095 (0.079)
2000	1%	594.60	1.574 (0.252)	1.644 (0.186)
2001	0.89%	332.82	2.539 (0.289)	0.188 (0.083)

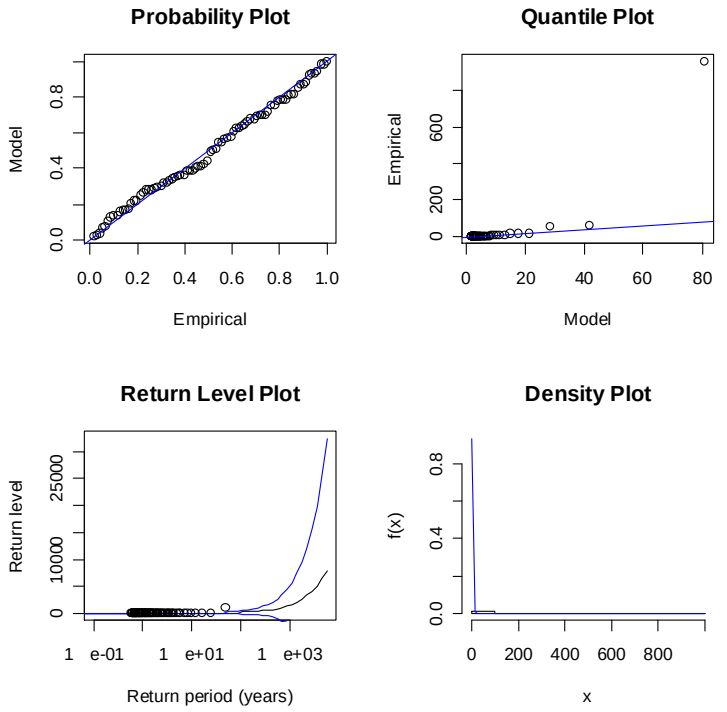
Table 4.2: MLE of each parameter of the Generalised Pareto Distribution along with their standard errors as well as the negative log-likelihood of the model and the percentage of data above the threshold (of 2 in this case).

The same diagnostic plots that were produced in Figure 4.2 can be reproduced for the GPD, in Figure 4.4:

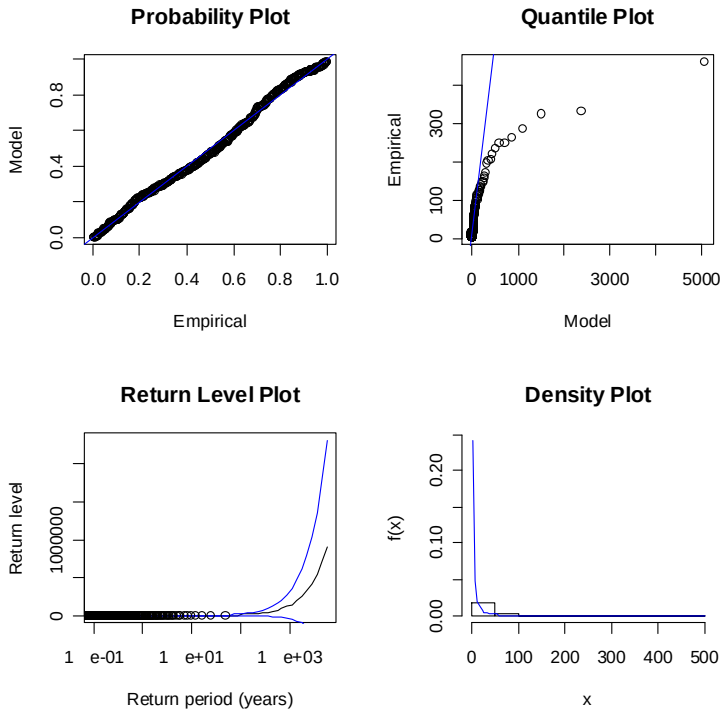
1997



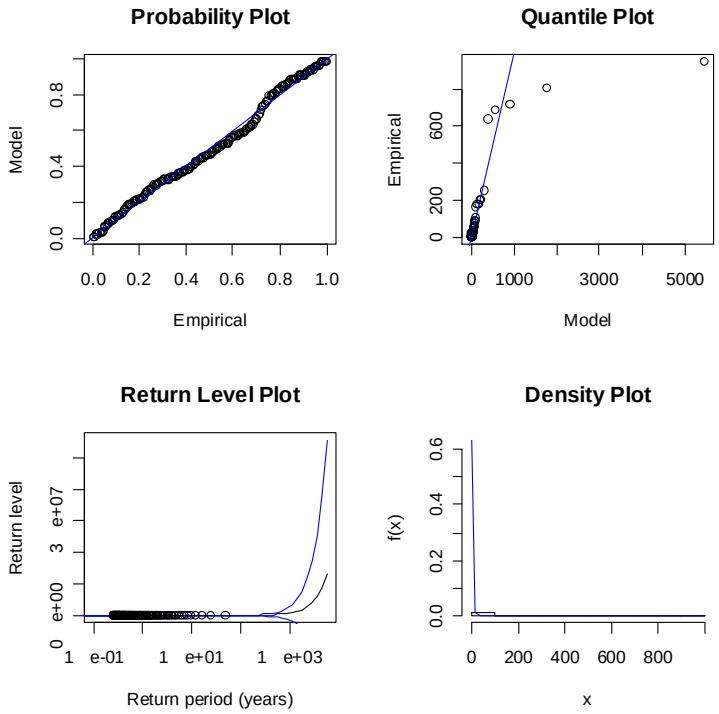
1998


Probability Plot
Model
Empirical
Quantile Plot
Empirical
Model
Return Level Plot
Return level
Return period (years)
Density Plot
f(x)
x

1999


Probability Plot
Model
Empirical
Quantile Plot
Empirical
Model
Return Level Plot
Return level
Return period (years)
Density Plot
f(x)
x

2000



2001

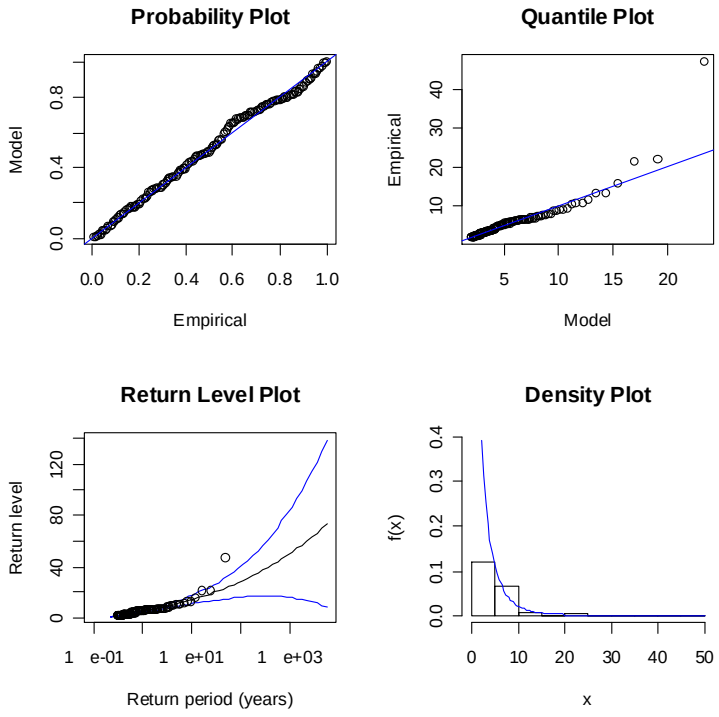


Figure 4.4: Diagnostic plots (same as Figure 4.2) for the GPD on fluxes > 2 for each year 1997-2001

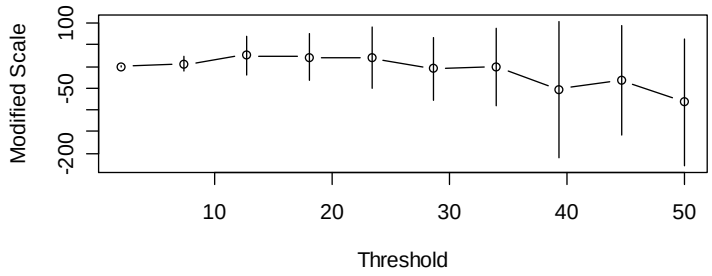
These diagnostic plots look like they show some improvement from the GEV plots.

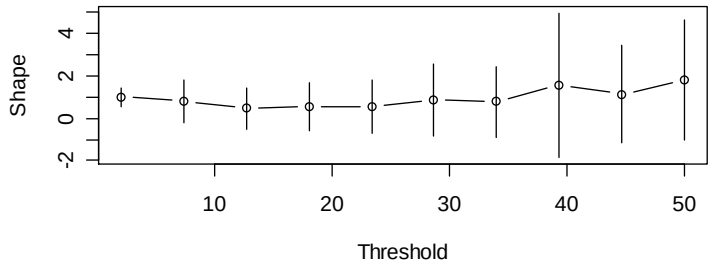
The probability plot looks like it follows the normal line a lot better, but there are still some points on the quantile plot which are of some concern. Again because of the skewness in the extremes it means that the density plot looks similar to the plots shown in Figure 4.2. Subjectively though the GPD plots look better.
One important point to consider here is that these plots are produced when the threshold is 2. It would be useful to see what sensitivity these models have to the choosing of the threshold, since clearly a model that changes significantly depending on the threshold choosing, may be difficult to analyse results from. The number of exceedences are shown in Table 4.3 and the parameter estimates are graphically summarised in Figure 4.5:

Threshold	1997	1998	1999	2000	2001
1	247	204	922	372	263
2	88	81	720	192	157
3	51	41	596	123	108
4	34	28	483	92	76
5	24	20	400	73	62
6	21	14	347	61	42
7	15	11	313	57	24
8	13	11	288	55	17
9	12	8	269	54	13
10	10	7	245	51	10
11	10	7	231	50	7
12	9	7	221	48	6
13	7	7	214	48	6
14	7	6	212	44	4
15	7	6	194	44	4

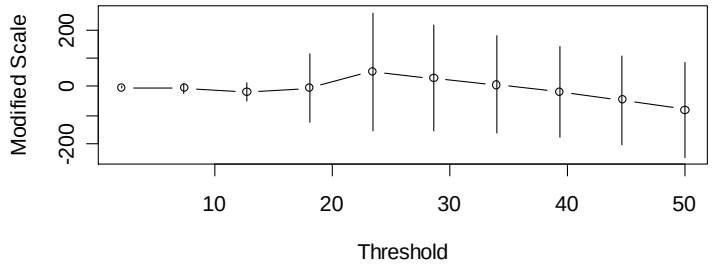
Table 4.3: Number of points exceeding the threshold for each year 1997-2001 using the half hourly data

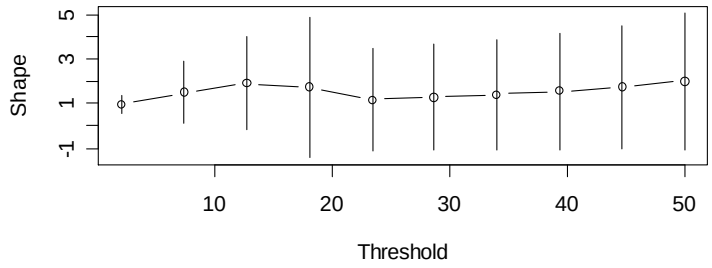
(a)


Modified Scale
-200
-50
100
10
20
30
40
50
Threshold

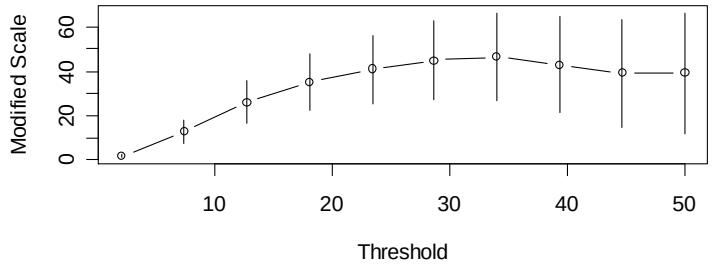

Shape
-2
0
2
4
10
20
30
40
50
Threshold

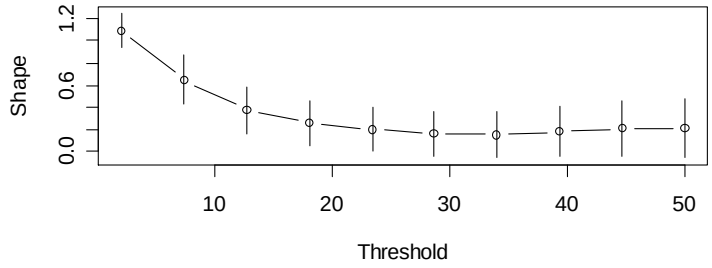
(b)


Modified Scale
-200
0
200
10
20
30
40
50
Threshold

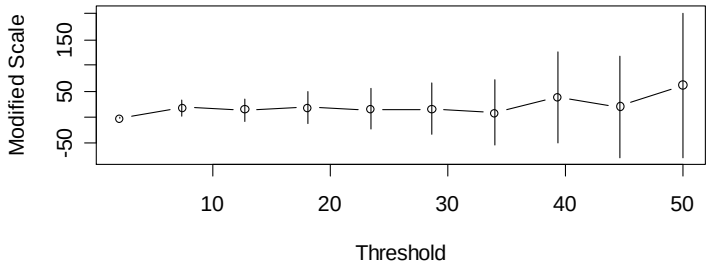

Shape
-1
1
3
5
10
20
30
40
50
Threshold

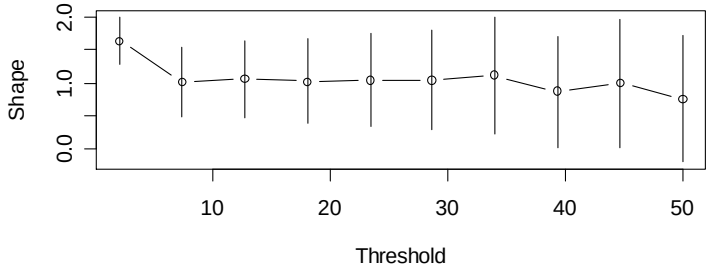
(c)


Modified Scale
0
20
40
60
10
20
30
40
50
Threshold


Shape
0.0
0.6
1.2
10
20
30
40
50
Threshold

(d)


Modified Scale
-50
50
150
10
20
30
40
50
Threshold


Shape
0.0
1.0
2.0
10
20
30
40
50
Threshold

(e)

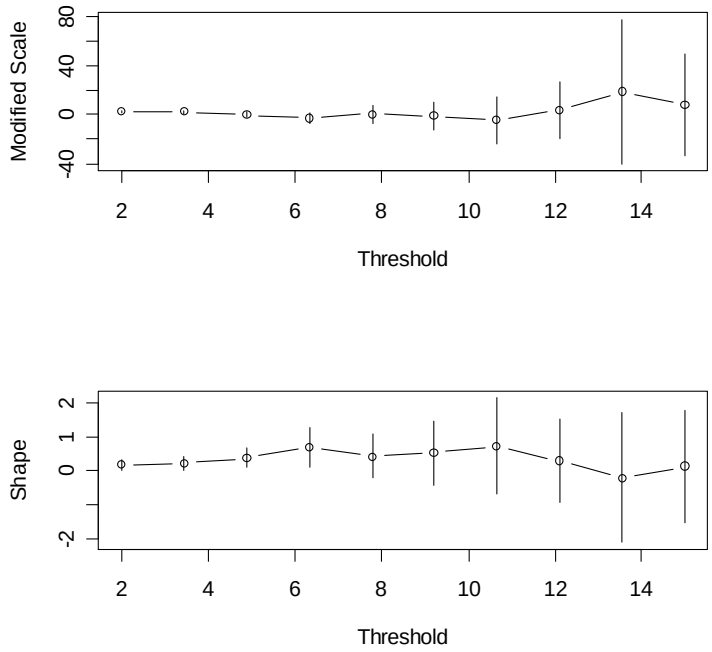


Figure 4.5: Plots showing the parameter estimates for the scale (μ) parameter and the shape (ξ) parameter respectively for the years (a) 1997, (b) 1998 (c) 1999 (d) 2000 and (e) 2001 daily maximum flux values

It can be seen from the graphs in Figure 4.5 that for almost all the years, the choice of threshold doesn't appear to have too much of an effect on the parameter estimates in the GPF. However in 1999, it can be seen that the parameter estimates only start to stay even, when the threshold flux value is approximately $20\mu gSm^{-3}$. Obviously as the exceedance level increase, the number of data values will decrease and so the variation around each parameter estimate will increase, but the estimate stays approximately the same. Therefore there would be a strong argument for keeping the threshold at $2\mu gSm^{-3}$, since it gives more data points to work with and appears to have little effect on the shape and scale parameters, bar 1999. This will be looked at separately from the rest when it is analysed further in Chapter 5, in order that sensible

conclusions can be produced from it.

4.5 Clustering

4.5.1 Introduction

It can be seen from the plots in Figure 4.3 that it may be that the extreme values are mostly clustered together and therefore are not independent. It would be interesting to study this fact further since this may help to understand whether the extreme data may have been the result of any natural weather events (say gale force winds or torrential rain). Therefore it would be of interest to see if the extreme data fall into clusters or whether it appears that they are just occurring independently throughout the years. Ferro and Segers (2003) suggest the use of an extremal index in order to measure the level of clustering that can be found in a data set.

The extremal index is defined as follows: Firstly let $\xi_1,\ldots, \xi_n$ be a strictly stationary sequence of random variables that have marginal distribution F, a right end point $\omega = \sup\{x : F(x) < 1\}$ and a tail function $\bar{F}$= 1-F. Then if $M_{k,l}$ is defined to be $\max\{\xi_i : i = k+1,...,l\}$ for integers $0 \leq k < l$ then $\xi_1,\ldots, \xi_n$ has extremal index $\theta \in [0,1]$ if for every $\tau>0$ there exists a sequence $u_1,\ldots,u_n$ such that as $n \to \infty$

1 $n\bar{F}(u_n) \to \tau$

2 $P(M_{0,n} \leq u_n) \to \exp(-\theta\tau)$

From this it can be shown (Leadbetter et al (1983)) that if $\theta = 1$ then there is no clustering in the extreme data, and if $\theta < 1$ then exceedences will tend to cluster. Obviously F is difficult to ascertain from real-life data sets so there is a way in which the extremal index can be estimated.

This estimation involves choosing a threshold u as in the Pareto distribution explained previously and defining N as:

$$N = \sum_{i=1}^{n} I(\xi_i > u)$$

where $\xi_{1,\ldots,}\ \xi_n$ is a sample of data. So N is the number of exceedences of u, and $1 \leq S_1 < \ldots < S_N \leq n$ are the exceedence times. Then define $T_i = S_{i+1} - S_i$ as the inter-exceedence times (for i=1,…, N-1).

From these simple definitions Ferro and Segers show two estimates of θ can be made which are useful in different circumstances, (i.e. when the maximum exceedence time is above 2 or below 2) as shown below in (4.11)

$$\tilde{\theta}_n(u) = \begin{Bmatrix} 1 \wedge \hat{\theta}_n(u) & if\ \max\{T_i : 1 \leq i \leq N-1\} \leq 2 \\ 1 \wedge \theta_n^*(u) & if\ \max\{T_i : 1 \leq i \leq N-1\} > 2 \end{Bmatrix} \tag{4.11}$$

where

$$\hat{\theta}_n(u) = \frac{2\left(\sum_{i=1}^{N-1} T_i\right)^2}{(N-1)\sum T_i^2} \tag{4.12}$$

and

$$\tilde{\theta}_n^*(u) = \frac{2\left\{\sum_{i=1}^{N-1}(T_i - 1)\right\}^2}{(N-1)\sum_{i=1}^{N-1}(T_i - 1)(T_i - 2)} \tag{4.13}$$

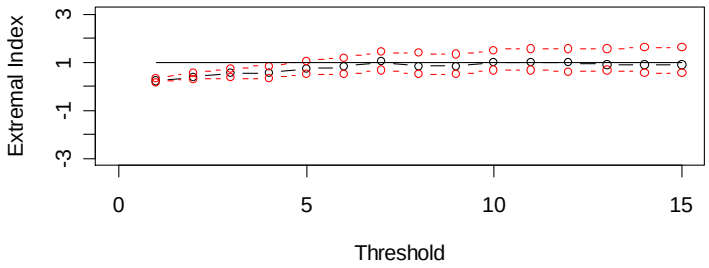
4.5.1.1 Bootstrapping Intervals

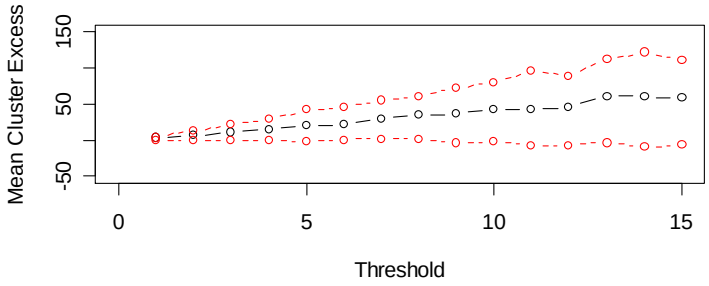
It would be very useful if a measure of confidence could be calculated from the extremal indexes described above. A method for doing this would be bootstrapping.

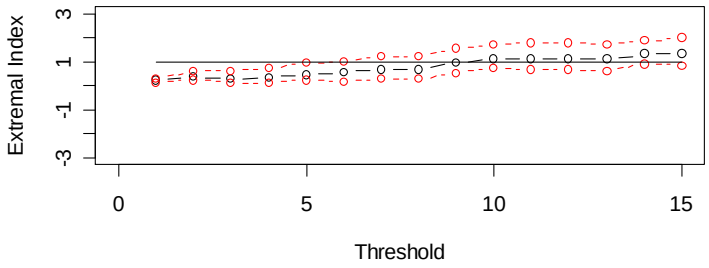
Bootstrapping requires the calculation of many estimates for θ. This can be achieved by taking random samples from {T} (with replacement) in order to calculate estimates of θ. Once 100 (say) estimates have been produced, then a standard 95% confidence interval can be fitted around $\tilde{\theta}_n^*(u)$ using the mean and standard deviation from the 100 estimates. This technique helps to show how much variation is actually in our estimate.

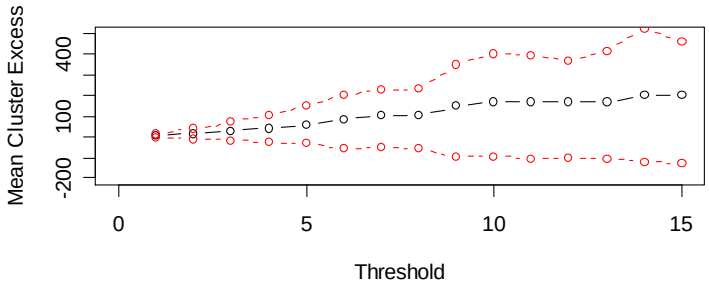
4.5.2 Using Flux Data

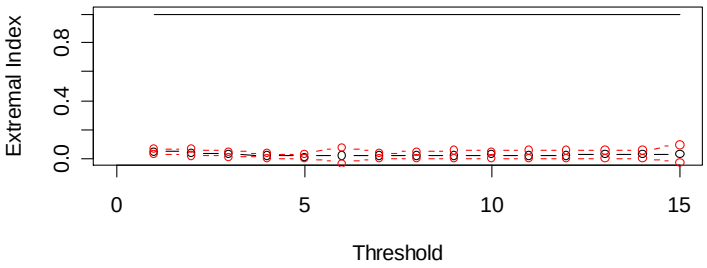
Now that a way to estimate the extremal index has been shown in the previous section, it can be applied to the Auchencorth data set. It will be applied to the half hourly flux data for each year. As well as this – the mean cluster excess value will be calculated along with the indices. These are calculated by summing the exceedences (after subtracting the thresholds) and then calculating the mean. These will be shown along with the extremal indexes with thresholds chosen between 0 and 15. As well as this, a table will show how many exceedences occur at each particular choice of threshold. Around the extremal indexes and mean cluster excesses, bootstrapped confidence intervals are applied as calculated in the above section.

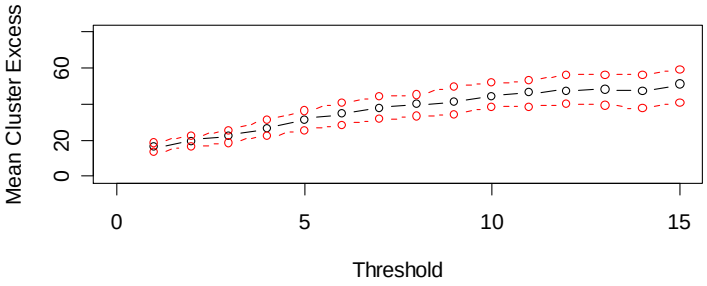

Extremal Index
-3
-1
1
3
0
5
10
15
Threshold

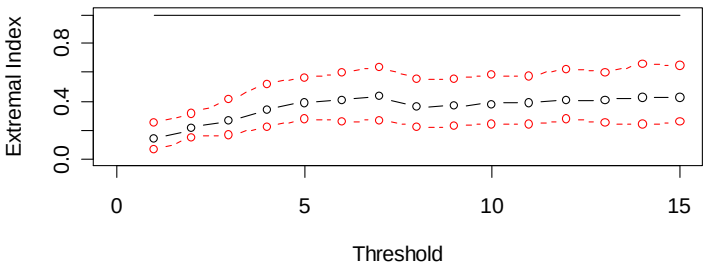

Mean Cluster Excess
-50
50
150
0
5
10
15
Threshold

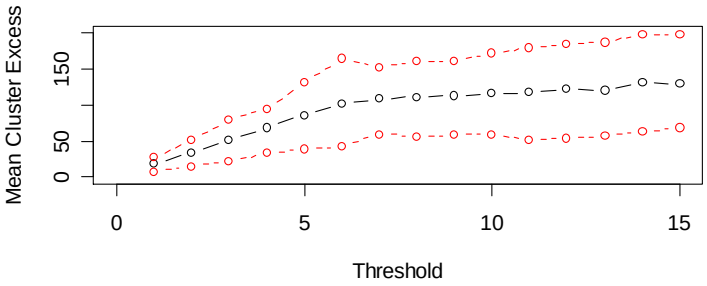

Extremal Index
-3
-1
1
3
0
5
10
15
Threshold


Mean Cluster Excess
-200
100
400
0
5
10
15
Threshold


Extremal Index
0.0
0.4
0.8
0
5
10
15
Threshold


Mean Cluster Excess
0
20
60
0
5
10
15
Threshold


Extremal Index
0.0
0.4
0.8
0
5
10
15
Threshold


Mean Cluster Excess
0
50
150
0
5
10
15
Threshold

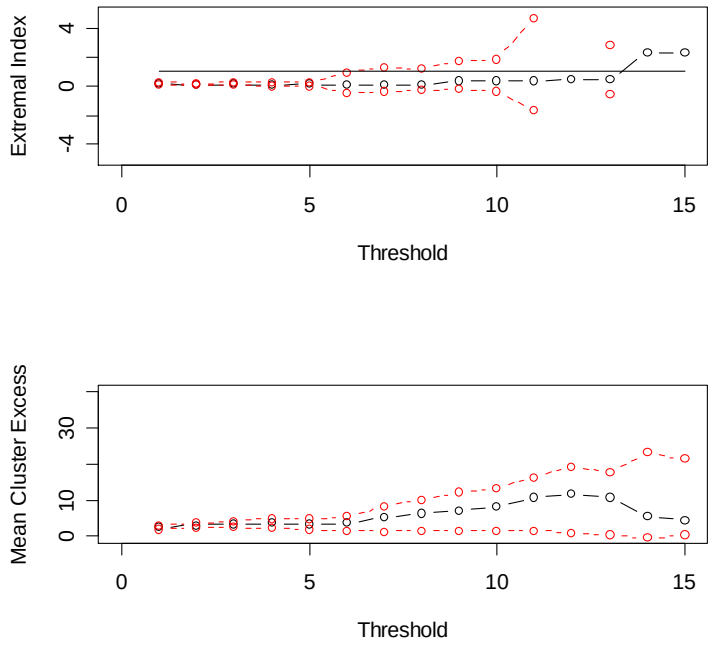


Figure 4.6: Pairs of plots showing firstly the extremal indexes as well as a horizontal line at θ=1, and below the mean cluster excesses for each year 1997-2001. Also shown in red are the bootstrapped confidence intervals for each year.

It can clearly be seen from Figure 4.6, the estimator for the extremal index starts to produce wide intervals when there are a small number of exceedences – and in some cases this pushes the estimator to a value bigger than 1, which, as shown in 4.5.1 is not a reasonable value.

It appears that 1997 and 1998 don't appear to show much evidence clustering of points. As the threshold goes up, the index quickly rises to a value close to 1. However in 1999 and 2000 there appears to be an obvious shift. Table 4.3 shows that the number of exceedences is certainly larger than the other three years and this high number may help to show why the extremal index is so low as there are many of the values close to each other. Figure 4.3 appears to back this up as there appears to be certain close groups of extreme values throughout these years. In 2001 the number of exceedences appears to revert back to a level similar to 97 and 98, however this time θ appears to stay reasonably low until late on when the small number of points appear to take it quite high up past 1. It could be concluded from this that 1997, 1998 and

2001s extreme values may be harder to explain, since they seem to appear at more interspersed times than in 1999 or 2000. This suggests that the data in 1999 and 2000 might be more difficult to model, since the extreme values appear to be as a result of (say) one big event, that the EMEP model may not have been programmed to take account of. In the three former 'non-clustering' years, it appears that the extremes are perhaps just occurring, maybe due to one spurious result that has been corrected quickly. This may be useful in Chapter 5 if the 1999 and 2000 data appear harder to fit the model to.

4.6 Conclusion

From the EVT that has been applied to the daily and weekly data it appears that, using the Generalised Extreme Value Distribution is not the best approach as (using the weekly maximums for this example), there are too many very low values still, and some very high values in these. This is evident when looking at the Q-Q plot especially of the data where the points appear to deviate a lot from the normal line.

Therefore, it is perhaps more relevant to look at the results that are given from the Generalised Pareto Family since this only applies to the values over a certain threshold. There are fewer concerns when the diagnostic plots are looked at under this family and so these look like they will give better results for the analysis. The choice of the thresholds isn't a particularly big issue either as it has been shown that the values for the model remain stable when the threshold is altered. This is with the possible exception of 1999 which has more extreme values at the value of $2\mu gSm^{-3}$

This suggests that the extreme values in the daily fluxes can be modelled. The parameter values for the GPF each year are contained in Table 4.2.

Whether the extreme data were clustering or not, is another thing that can be looked at. By estimating extremal indexes it can be inferred that in three of the years (1997, 1998 and 2001) there does not appear to be clustering in the extreme values and it could be inferred that these results are nothing more than some local weather conditions (for example). In 1999 and 2000 however the extreme values are tending to cluster together which may become a problem in the next chapter. The next chapter will finally look at the modelled data against the measured data and compare the two sets with each other to see what differences there are between them.

Chapter 5 – Comparing the Measured and Modelled Data

5.1 Introduction

Now that the data have been thoroughly analysed and there has been a study performed into whether or not the measured data could be chaotic, it would be useful to look at how the modelled data should be validated against the measured data. Using information that has been obtained from the previous chapters will help in finding out reasons for any differences that may lie in the (EMEP) modelled and (Auchencorth) measured data.

A useful approach would be to look firstly at what the ideal situation would be for a model-measurement comparison and what sort of statistical analysis would be performed, then look at ways to possibly estimate these approaches from a real life setting like the one that is being studied here.

From this it can be deduced whether there appears to be any similarities between the modelled and measured data and if not, whether there may be a pattern to any differences. Whether the measured data appear to show higher or lower values in general will give a clear indication of whether there appears to be any bias prevalent. It will also be interesting to see if using the Event Analysis discussed in Chapter 2 (2.5) will help to show why the differences are occurring.

The modelled data comes from EMEP in a daily format so comparing it at different timescales to see where differences may lie would be useful, but at the moment only by averaging the daily modelled data to create weekly/monthly values is the only way that the data sets can be looked at in a pairwise setting. However, if the data is looked at for some particular days (perhaps the ones that appear to show greatest displacement in terms of the daily values) could be looked at by seeing how the half-hourly values vary throughout the day to make sure they are not being unfairly influenced by a large “spurious” value.

After all this, it should be noted that the uncertainty, that is going to exist due to comparing a fixed point against a value that is an average over 50km squares, may

contribute to any differences and some analyses into quantifying this may come in useful. This will lead into the next chapter where this will be discussed at more length.

5.2 Methods of Comparing Modelled and Measured Data

There are many different ways in which it can be measured how well two series of data match each other. One of the simplest methods would be to just plot the two series on a normal x-y plot and see how well it can be fitted by a line of equality. One way of measuring the goodness-of-fit could be the R-squared value. However this could be a false result as two variables could be well correlated despite not being equal, so there should be other methods looked at. Bland and Altman (1986) suggest a method where the difference between the measured and modelled data can be plotted against the average of the two values, and conclusions can be made from the shape of the points that are plotted. Stohl et.al (1998) shows more quantitative results by measuring a number of different statistics, ranging from R-squared to looking at over-predictions and under-predictions. It is important when thinking about well a model fits to measured data, not to just look at how small the residuals are (say) but also how well the model reflects the shape of the time series of the measured data. It is important that a model captures seasonality and any trends in the data, before it can be shown to be an accurate one.

5.3 Comparisons of the raw daily measurements against the EMEP modelled values

The first thing that it would be good to look at would be simply the measured values against the EMEP values at the daily level. This will be able to show whether there are differences between the data sets. Figure 5.1 shows these for the years 1997, 1999, 2000 and 2001 along with a line of equality in order that it can be seen how well the points match each other.

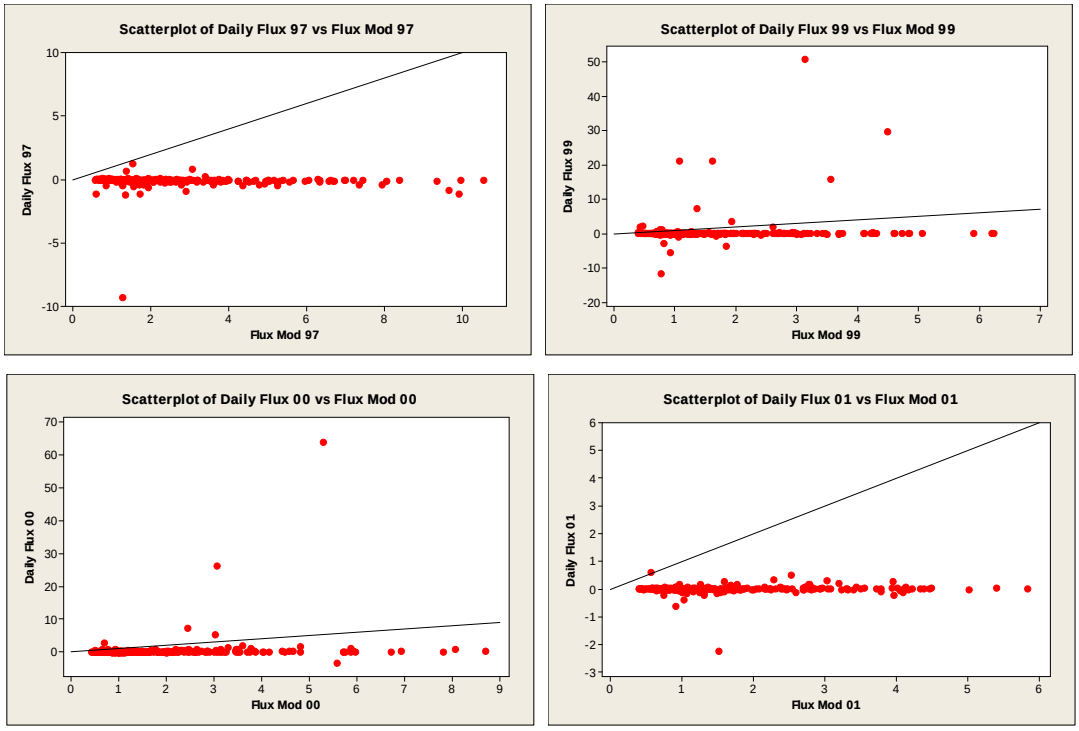


Figure 5.1 Plots showing the relationship between the modelled (x axis) and measured (y axis) data for the years 1997, 1999, 2000 and 2001. A line of equality has been put onto each plot.

These data appear to not match the modelled data particularly well in any year. This is reasonably obvious from Figure 5.1. However, it can be seen that the range of points in the measured data is a lot larger than the modelled. This makes it rather difficult to see whether the smaller measured values are matching the modelled in any way. Therefore it would be useful to see whether the distributions of the data are similar, even if the actual values are not. It would be useful to see the data in a time series format so that the shape of the data can also be subjectively analysed. These are looked at in Figure 5.2:

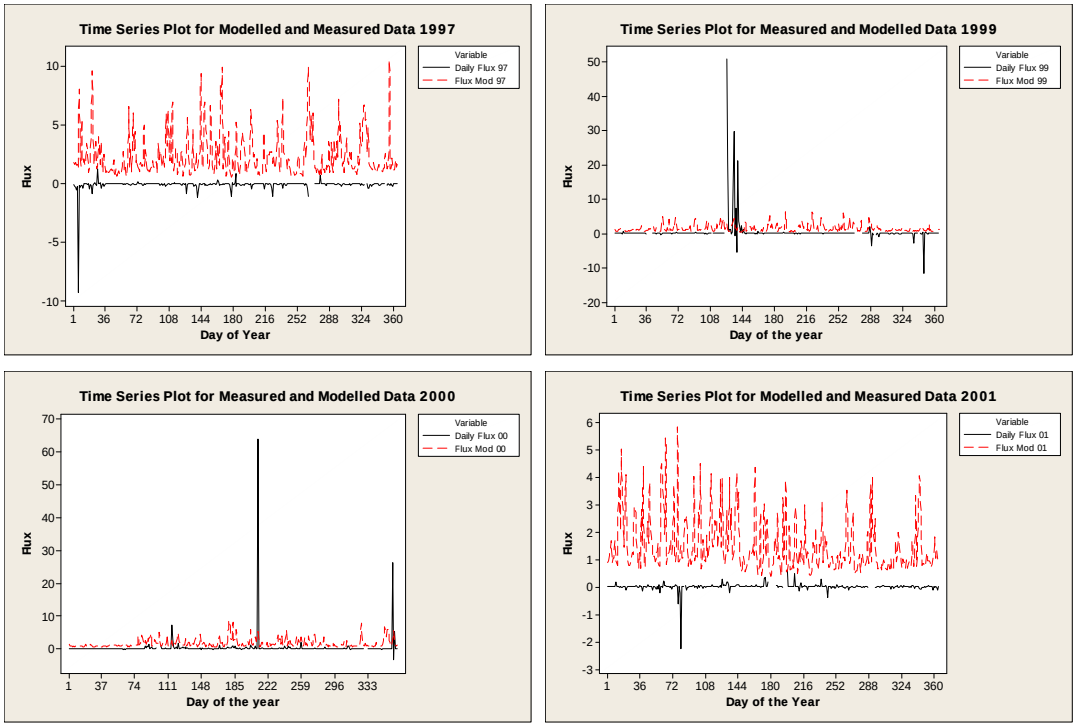


Figure 5.2: Plots showing the time series of the modelled series (in red) and the measured series (in black)

Because the measured series is commonly larger than the modelled series it would appear to be more useful to look at a zoomed in version of each of these plots, in order to see the shape of the measured data more clearly. These are considered in Figure 5.3:

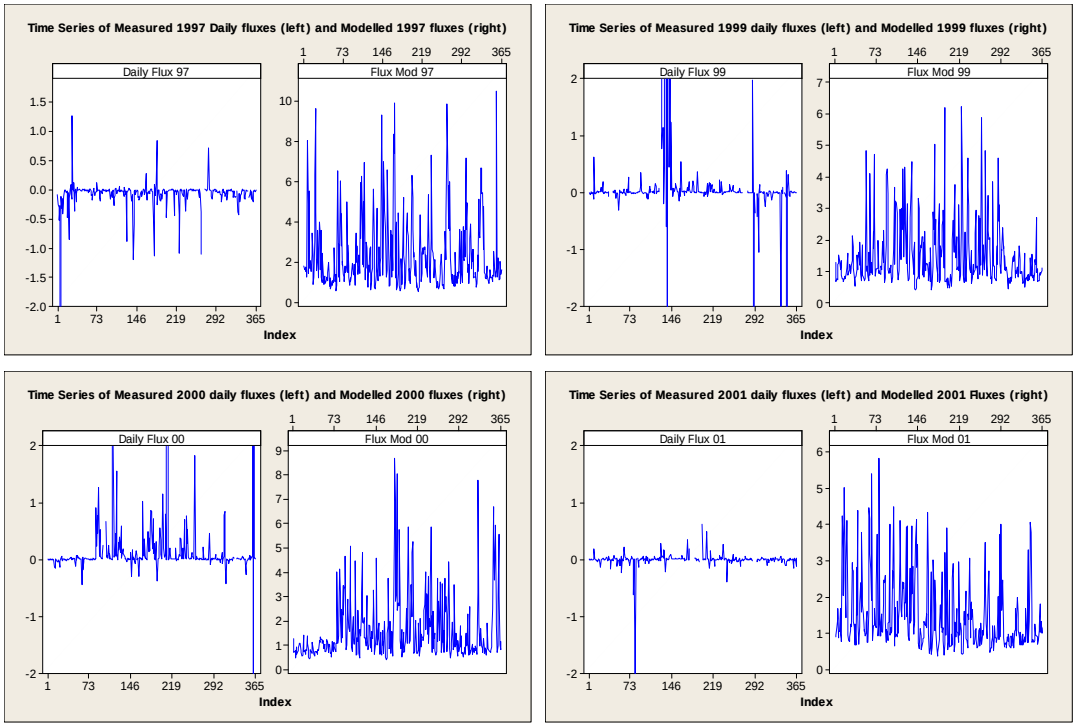


Figure 5.3: Time Series plots of Figure 5.2, but these show the shapes of both time series more clearly. The measured data appears on the left and the modelled on the right

It can be seen from Figure 5.3 that the two sets of data do seem very different. As well as not matching particularly well size-wise, it can be seen that the measured data doesn't appear to follow the shape of the modelled data either, never mind the size. This is merely a subjective opinion, and so it would be useful if formal statistics could be calculated in order to measure how well each of the measured data matches the model.

There are a number of ways of looking at how similar or different measured and modelled data are. As mentioned in Section 4.1, Stohl et.al (1998) recommends a variety of different methods which are described below:

1 The Bias (B) where $B = \frac{1}{N}\sum_{i=1}^{N}(P_i - M_i)$ where P_i and M_i are the predicted values from the model and the measured values respectively, and N is the number of paired points.

2 The Fractional Bias (FB) where $FB = \frac{2B}{(\bar{P}+\bar{M})}$ where $\bar{P}$ and $\bar{M}$ are the mean values of the modelled and measured data

3 The Normalised Mean Square Error (NMSE) where

$$NMSE = \frac{1}{N}\sum_{i=1}^{N}\frac{(P_i - M_i)^2}{\bar{P}\,\bar{M}}$$

4 The Spearman rank-order correlation coefficient (r_s)

5 The percentage of modelled predictions that agree within a factor of 2 with the measurements (FA2)

6 The percentage of modelled predictions that agree within a factor of 5 with the measurements (FA5)

7 The number of overpredictions which is measured as a percentage in order to tell whether the model tends to overpredict or underpredict (FOEX) where

$FOEX = 100(\frac{N_{(P_i > M_i)}}{N} - 0.5)$. This will always be between –50% and +50%

For the four years for which the modelled daily data and the measured data can be compared with each other each of these have been calculated and added in Table 5.1:

	Bias	Frac Bias	NMSE	r_s	FA2	FA5	FOEX
1997	2.399	2.121	-42.947	-0.448	0.56%	0.85%	50%
1999	1.092	1.146	24.659	0.051	2.33%	4.36%	45.93%
2000	1.244	1.247	25.114	0.302	1.96%	9.24%	48.32%
2001	1.476	1.990	1111.72	0.028	0.29%	0.29%	49.71%

Table 5.1: This contains 7 different ways in which the modelled and measured data can be compared.

The first thing to notice is that for all 4 years the values look very low, for each particular category. Before some of these can be interpreted a general overview appears to show that 1997 seems to be giving far better results than the other three

years, but still shows nowhere near any sort of strong similarities. It can be seen for instance that the Pearson correlation figures are all very low for each year, (especially for 2001 which shows an incredibly small correlation coefficient). These back up the plots above that show the points not falling anyway close to a line of equality. The FOEX figures all show what has already been seen previously, that the measured data is consistently higher than the modelled data. Some of these methods for evaluating models are more sensitive than others. The FA2 and FA5 values can be very sensitive especially when values fall around 0. This is a problem in this data set since a lot of measured (and modelled values) are around zero. It would be useful if a technique could be applied to see the main problem could be between the modelled and measured data.

5.4 Bland Altman Analysis

Bland and Altman (1986) discuss many of the problems that face trying to compare two sets of data. They discuss some of the problems that have been mentioned early (high correlations not meaning that two sets of data are close to each other for one). They use an approach featured below:

The Bland-Altman plot can help to see this graphically. By plotting the differences (modelled – measured) against the average deposition (using both the modelled and measured values to get this), the points can be compared against the mean value of the differences. 95% of the points should lie between the (mean (d) ± 2*standard deviation (s)), assuming the differences are normally distributed. Bland and Altman suggest that if the difference between d+2s and d-2s is not "clinically important" then the two methods can be said to agree, since they could be used interchangeably. If this difference is too large then it can be said that the two methods certainly cannot. Though this analysis will not be dealing with any “clinical” issues, this can still be used for the sulphur dioxide comparisons in a similar way. With the large extremes left in these plots were not useful at all, so the filter of removing measured values above 2μgSm^{-3} was applied. These plots are shown in Figure 4.4:

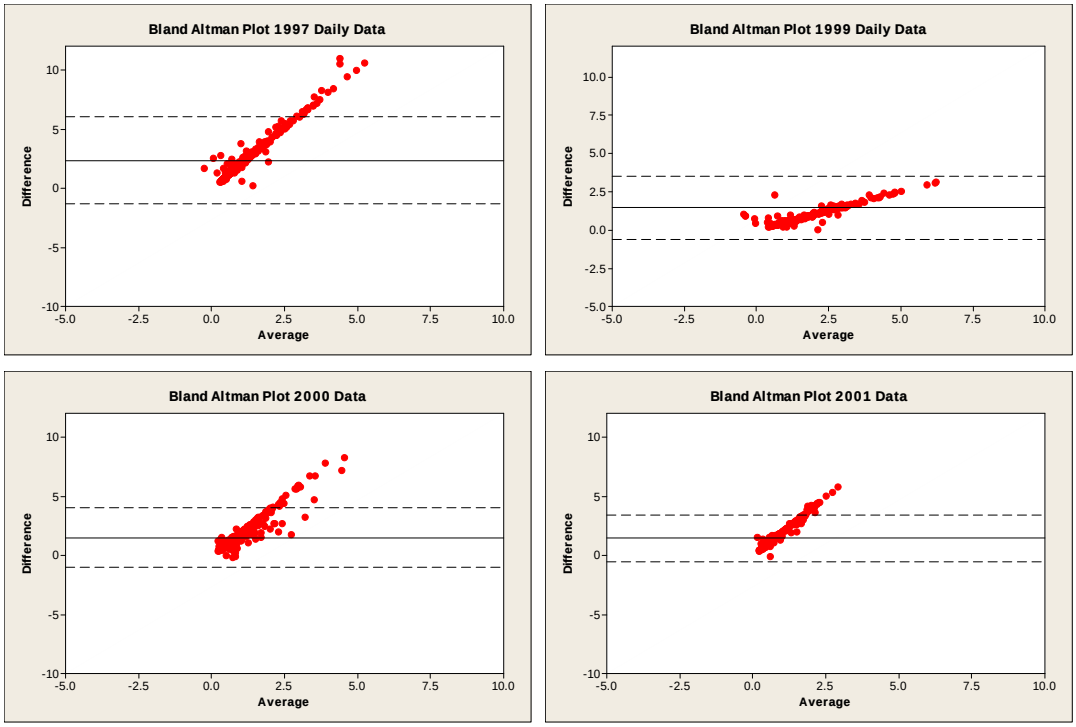


Figure 5.4: Bland-Altman plots showing the averages of the modelled and measured data against the differences for years 1997, 1999, 2000 and 2001. Lines are drawn at the mean ± 2* standard deviation.

It can be seen that in each plot the data appears to follow a straight line. This isn't too surprising as the measured data is a lot bigger than the modelled data in the previous plots. However it is most important to look at the two "boundaries" for each year. It can be seen clearly that in all 4 years, the difference between d+2s and d-2s is too large for the methods to be used interchangeably and so as has been seen previously the two sets of data in their raw format do not match each other well enough.

When the R^2 values above 30% were taken into account (by removing the ones below), this also unfortunately did not make a great deal of difference. Because the daily values were aggregated from the raw half hour values, most of the results ended up averaging to nearly/exactly the same value as before. Only in the extreme cases (which were shown above in the Bland-Altman analysis) were the means changed at all.

5.5 Spatial Aspects

As was discussed in previous chapters, part of the problem with comparing these values is that one is taken from a point measurement inside a spatial area of 50km by 50km. It may be that the local weather conditions make it impossible to believe that the average given by the model will be accurate to every location inside the square. The amount of sulphur dioxide in the air can vary over very small areas so it would be difficult to believe that the same levels should be expected over such a wide area. It is difficult to quantify the level of variation that might be expected from one of these grid squares, but it is easy to imagine that it might be rather large.

5.6 Conclusions

From this chapter it has been shown, very clearly that comparisons between modeled and measured data in a natural environment can be very difficult indeed. Especially when using a data set with what has been shown to have such high levels of noise. This chapter has looked at how difficult it is to compare a noisy data set with a model that models over a large area. The two do not compare well against each other for any of the 4 years looked at. It would appear that the levels of noise are the main cause behind this.

Common techniques to compare between the two sets of values were looked at, along with techniques such as Bland-Altman plots in order to graphically see what differences there were between the two data sets, along with several statistical calculations in order to show more quantitatively how well/badly the data sets compared with one another. All of the techniques applied to the data sets showed that there was a clear disparity between the two sets of results.

It was studied whether removing some of the data that was deemed less reliable from the analysis in previous chapters, would allow a better comparison, however it ended up showing nothing different here.

Hence the main conclusion that can be taken from this chapter is that there are major pitfalls when looking at two sets of data taken in very different ways. There are ways to make it as clean as possible but these still might end up showing differences between them. One thing that should be made clear is that the levels of noise in the measured data, as shown in Chapter 2, mean that it is almost impossible to blame the

model in this, as it is very difficult to be confident about the measured data.

Chapter 6 – Final Conclusions and discussion

The previous chapters have shown the many difficulties and problems that accompany trying to fit an accurate model to routinely measured, high frequency sulphur dioxide fluxes at a single monitoring station (Auchencorth Moss) which might then be used to allow verification of large scale atmospheric transport models. The measurements have a complex structure, and can be impacted by weather conditions and other environmental situations which can change very easily over even small areas and in short time periods. For this setting, further issues arise since many of the modelled and measured fluxes are very low and can be affected by errors in the measuring equipment and even human error. These together combined to make it very difficult to define an underlying model taking account of this additional variation. The flux calculation requires several different measurements from the tower. Gas concentrations are taken from three heights and several other input variables are needed in order to calculate a flux. The fact that slope estimates are required from three data points for the flux also mean that the data quality can be reduced by one false measurement from the many variables that are collected.

From the earlier chapters it was shown that using only the gas concentrations that had significant slope estimates to produce the measured fluxes (removing the more poorly fitted models (defined by the R^2 values)) reduced the data by up to 10-15% but gave slope estimates that were based on better fitted models . The values which were removed also tended to have the lowest flux values, which makes sense as the low sulphur dioxide fluxes calculated generally came from the low slope estimates that were obtained from a linear model which did not fit the three gas concentrations particularly well and ended up merely producing a very flat line and hence slope estimates of small magnitude. It was also checked in detail whether the fluxes being calculated were being affected by time of day or seasonality. These analyses did not provide any significant results to explain the variability. However, the reassurance provided by these analyses was important for the final comparison between the modelled and measured data at since ‘unusual values’ and any time of day effects did not need to be accounted for. Sensitivity analysis helped to show that most of the variation in the flux came from the gas concentrations rather than the other input variables in a quantitative sense. This suggested that problems in modelling fluxes might come from problems in predicting the raw concentrations.

Chapter 3 explored the advantages and disadvantages of using ideas concerning chaotic behaviour to help model the measured fluxes. However the results turned out to be disappointing – despite an extensive check of the different methods of estimating Lyapunov Exponents. Three methods were discussed in some detail, and two of them were applied to the data sets with differing results. The results gave some indication of chaotic behaviour, but this was dependent on scale. As Timmer et. al suggest, it is very difficult to find a method that can identify between a chaotic system and one that has a small signal hidden by large amounts of white noise and unfortunately this analysis did not allow us to distinguish between these two different explanations in a consistent or reliable way.

As is commonly the case in environmental time series, the next set of analyses focussed on the extreme values which might represent episodes of air pollution. The extreme value analysis was very useful and helped to show that the extreme values in these data sets could be modelled by classic extreme value theory technique. Techniques that looked at the clustering of the values along with the raw values themselves were used. Two techniques worked reasonably well on the extreme values – the Generalised Extreme Value Theory and the Generalised Pareto Distribution - the Pareto distribution especially was shown to not have too much variation even when the threshold value for what constituted an "extreme" value was altered for 4 of the 5 years. Analysing the extreme values also allowed an opportunity to explore the effect on subsequent analysis of their removal and whether the final model measurement comparison was improved.. Further work on modelling the extreme values would be recommended as there were indication of results which could have possibly been analysed further (e.g. by comparing them with data sets from other sites in the local area or further afield perhaps).

The final chapter showed the results of the model measurement comparison. The model in this case was a large scale atmospheric transport model (EMEP), which effectively provides predictions at a grid scale (maybe give the dimensions). The issue was whether single monitoring station results could be useful in model verification. An initial comparison showed considerable disparity between the modelled and measured data using basic scatter plots compared with lines of equality. Applying the screening techniques that had been studied in Chapter 2 did not bring the data sets closer to each other, and there did not seem to be much relationship between the modelled and measured data sets at all. Different statistical techniques

such as Bland-Altman plots and statistical measures of agreement all showed a distinct lack of agreement between the two sets. Even when the extreme values were taken out the two sets of data still differed by large amounts, and using the Bland-Altman techniques, the "limits of agreement" were much too far apart (in comparison to the size of many of the measured means) to be confident about these at all.

However one thing that was not studied in too much detail was the fact that a point estimate was compared against a model which generates a spatial average for a large area. It would be useful to look at this in more detail to see whether a model over such an area should be expected to fit well against one point inside it. This is definitely one more area that could be looked at in more detail.

While many of the results in this thesis were disappointing since they did not improve the model- measurement comparisons, nor indeed help explain the differences between the modelled and measured data to any great length, there were some interesting findings. Overall, two areas of further work seemed the most promising, these are application of chaos to environmental time series, and the further analysis of extremes.

The thesis showed the difficulty in assessing chaotic behaviour in a data set that has a small signal obscured behind lots of noise. The use of Extreme Value Theory showed some interesting results which could be taken further forward to see whether these are common in other data sets, especially where episodic behaviour (high pollution events) are especially important.

Bibliography

Abarbanel, H.D.I. (1992). Lyapunov exponents in chaotic systems: Their importance and their evaluation using observed data. *Modern Physics* Letters B

Babovic, V and Keijzer, M. (1999) Forecasting of River Discharges in the Presence of Chaos and Noise, *Coping With Floods: Lessons Learned from Recent Experiences, Kluwer*.

Bartnicki, J., Krzysztof, O., Jonson, J.E., Berge, E., Unger, S. (1998) Description of the Eulerian Acid Deposition Model *Transboundary Acidifying Air Pollution in Europe, EMEP/MSC-W, Status Report 1/98, Norwegian Meteorological Institute, Oslo, Norway*

Campbell, G.S. (1977) An introduction to Environmental Biophysics. *Springer-Verlag*.

Chan, K.F., Gray, P. (2006) Using Extreme Value Theory to Measure Value-at-Risk for Daily Electricity Spot Prices, *International Journal of Forecasting* **22** 283-300

Coles, S. (2001) An Introduction to Statistical Modeling of Extreme Values, *Springer Series in Statistics*

Coles, S., Pericchi, L.R., Sisson, S (2003) A fully probabilistic approach to extreme rainfall modelling, *Journal of Hydrology*, **273** 35-50

Dou, C.H., Woldt, W., Bogardi, I., Dahab, M. (1995) Steady-state Groundwater Flow Simulation with Imprecise Parameters, *Water Resource* **31**(11) 2709-2719

Ellner, S.P., Turchin, P. (1995) Chaos in a Noisy World: new methods and evidence from time-series analysis, *American Naturalist* **145** 343-375

Ellner, S.P. (2000) Defining Chaos for Real, Noisy Data: Local Lyapunov Exponents and Sensitive Response to Perturbations, *Chaos in Real Data* 1-32

Ferro, C.A.T. and Segers, J. (2003) Inference for Clusters of Extreme Values. *Journal of the Royal Statistical Society*, **Series B 65**, 545-556

Frazier, C., Kockelman, K.M. (2004) Chaos Theory and Transportation Systems: An Instructive Example.

Giannerini, S and Rosa, R. (2004) Assessing Chaos in Time Series: Statistical Aspects and Perspectives, *Linear and Nonlinear Dynamics in Time Series* **8(2)** Article 11

Griffith, D.W.T., Galle, B. (2000). Flux Measurements of NH_3, N_2O and CO_2 using dual beam FTIR spectroscopy and the flux-gradient technique, *Atmospheric Environment* **34** 1087-1098

Gu, F., Shen, E., Meng, X., Cao, Y., Cai Z. (2004) Higher Order Complexity of Time Series (2004), *International Journal of Bifurcation and Chaos* **14(8)** 2979-2990

Hastings, A., Hom, C.L., Ellner, S.P., Turchin, P., Godfray, H.C.J. (1993) Chaos in Ecology: Is Mother Nature a strange attractor? *Annual Review of Ecology and Systematics* **24** 1-33

Hier Majumder, C.A., Travis, B.J., Belanger, E., Richard, G., Vincent, A.P., Yuen, D.A. (2006) Efficient Sensitivity Analysis for Flow and Transport in the Earth's Crust and Mantle, *Geophysics J. International* **166** 907-922

Huysmans, M., Madarasz, T., Dassargues, A. (2006) Risk Assessment of Groundwater Pollution using Sensitivity Analysis and a Worst Case Scenario Analysis, *Environmental Geology* **50** 180-193

Kantz, H., Schreiber, T. (1997) Nonlinear Time Series Analysis, *Cambridge University Press*

Leadbetter, M. R., Lindgren, G. and Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*. Springer

Leuning, R., Baker, S.K., Jamie, I.M., Hsu, C.H., Klein, L., Denmead, O.T., Griffith, D.W.T. (1999) Methane Emission from Free-Ranging Sheep: A Comparison of Two Measurement Methods, *Atmospheric Environment* **33** 1357-1365

Marcotte, Denis. (1995) Generalized cross-validation for covariance model selection, *Mathematical Geology* **27(5)** 659-672

Mellinger, D.K., Clark, C.W. (2006) MobySound: A Reference Archive for Studying Automatic Recognition of Marine Mammal Sounds, *Applied Acoustics* **67 (11-12)** 1226-1242

Monteith J.L., Unsworth M.H. (1990), Principles of Environmental Physics, *Butterworth-Heinemann Ltd*

Nychka, D., Ellner, S., Bailey, B. LENNS Package for R (1992) *http://www.cgd.ucar.edu/~nychka/*
OECD-FAO Agricultural Outlook 2008-2017, *OECD/FAO* (2008)

Pauluzzi, D.R., Beaulieu, N.C. (2000) A Comparison of SNR Estimation Techniques for the AWGN Channel, *IEEE Trans. On Communications*, **48:10**, 1681-1691

Rapp, P.E., Schmah, T.I. (2000) Dynamical analysis in clinical practice, *Chaos in Brain?* 52-62

Rout, B.K., Mittal, R.K. (2006) Tolerance Design of Robot Parameters Using Taguchi Method, *Mechanical Systems and Signal Processing* **20** 1832-1852

Saltelli, A., Chan, K., Scott, E.M. (2000) Sensitivity Analysis, *Wiley Series in Probability and Statistics*

Schwarz, G., (1978) Estimating the dimension of a model. *Ann. Stat*. **6:** 461–464

Smith, R.L (1985) Maximum Likelihood Estimation in a class of non-regular cases,

Biometrika, **72** 67-90

Smith, R.L. (1989) Extreme Value Analysis of Environmental Time Series: An Application in Ground-Level Ozone, *Statistical Science* **4(4)** 367-377

Stohl, A., Hittenberger, M., Wotawa, G. (1998) Validation of the lagrangian particle dispersion model FLEXPART against large-scale tracer experiments. *Atmospheric Environment* **24** 4245-4264

Tempkin, J.A., Yorke, J.A. (2004) Spurious Lyapunov Exponents Computed from Data

Timmer, J., Haussler, S., Lauk, M., Lucking, C.-H (2000) Pathological tremors: Deterministic chaos or nonlinear stochastic oscillators? *Chaos* **10(1)** 278-288

Tong, H (1990) Non-linear Time Series: A Dynamical System Approach 57-58

Turchin, P. and Ellner, S.P (2000) Living on the Edge of Chaos: Population Dynamics of Fennoscandian Voles, *Ecology* **81(11)** 3099-3116

Unlu, K., Parker, J.C., Chong, P.K. (1995) Comparison of three uncertainty-analysis methods to assess impacts on groundwater of constituents leached from land-disposed waste. *Hydrogeol* **J 3(2)** 4-18

Wolff, R., Yao. Q., Tong, H. (2004) Statistical Tests for Lyapunov Exponents of Deterministic Systems, *Linear and Nonlinear Dynamics in Time Series* **8(2)** Article 10

World Population to 2300 (2004), *UN Department of Economics and Social Affairs, Population Division*

Zimmer, C. (1999) Life After Chaos, *Science* **284** 83-86