Appendix A

Previous Katestone Studies for Macquarie Generation

A1 Review of monitoring data

In December 2004, Katestone Environmental conducted a pre-feasibility study for Macquarie Generation for a proposed additional power station in the Hunter Valley (Katestone Environmental 2005a). In the pre-feasibility study, the air quality monitoring data were reviewed to understand the existing air quality in the Hunter Valley and to assess the existing capacity in the air shed. This study identified the following major findings:

Over the last 10 years each of the monitoring sites have recorded exceedances of the short-term SO_2 impact assessment criteria. Ground-level concentrations of NO_2 have been well in compliance with the impact assessment criteria. The main identifiable source of the SO_2 exceedances is from both power station operations. Identifying which of the two power stations is responsible for the event is extremely difficult from the information provided.

The proposed new power station has significantly different stack characteristics than the existing power stations and will likely achieve better dispersion capabilities. It is difficult to determine whether the proposed power station will create any additional exceedances of the guideline without further modelling assessments, however, with the following considerations the proposal is plausible:

The coal sulfur content currently used by the power stations may need to be reviewed for any proposed changes in the future. Higher coal sulfur contents may result in additional guideline exceedances and possibly lower the feasibility of an additional power station. Lowering of the coal sulfur content will reduce the number (or likelihood) of exceedances.

The review recommended further assessments into how the existing and proposed power stations will interact, as well as modelling to identify a suitable location for the power station and more information on the feasibility of the proposal.

A2 Model validation for existing operations

The analysis undertaken in this study firstly investigated the performance of the meteorological component of the model. The model's performance was tested for wind speed and temperature at each of the four weather stations in the region (Liddell, Mt Arthur, Ravensworth and Bengalla). Apart from the Bengalla 90 metre tower, all the weather stations are surface stations (10 metre towers), which are strongly influenced by local features and may not be representative of conditions at the height of the power station plumes. Unfortunately no information is available that could be used to verify atmospheric conditions at stack top (or plume height).

The results presented in Appendix 3 of Katestone 2005b indicate the model performed well at all sites in predicting wind speed, direction and temperature. The results obtained in this study are among the best obtained by Katestone Environmental for model performance. The closest agreements with measurements were obtained at the Bengalla and Ravensworth sites. Overall the results show that TAPM performed well for the meteorological component and is suitable for use in this assessment.

The second stage of the validation process considered the model's performance in predicting ground-level concentrations of SO_2 due to the existing power stations. Hourly emissions files were generated for each of the units to accurately represent the time-varying emission rates. The model predicted the robust highest concentration (RHC) well at all sites with slight over-predictions at Mt Arthur and Liddell. The model's performance decreases with the lower percentiles, with TAPM under-predicting the mean and 99th percentile concentrations at most sites, particularly Mt Arthur, Muswellbrook and Ravensworth. This could be due to sources of SO_2 in the region that are not specifically included in the dispersion model, but exist in the region and impact on the lower levels recorded at the monitoring station. For the type of assessment presented in this report it is important that the model adequately represents the peak impacts. The time of peak impacts was also investigated and as reflected in the monitoring the peak impacts all occur during the daytime during light wind convective conditions. Therefore the models performance is considered adequate for this study.

A3 Preliminary Modelling – Local Impacts

Two stages of modelling work have been undertaken to date. These are presented in Katestone Environmental (2005b) and Katestone Environmental (2006a).

The modelling presented in Katestone Environmental (2005b) predicted additional exceedances of the DECC's impact assessment criteria for the proposed Bayswater B Power Station. The modelling undertaken in Katestone Environmental 2006a has adopted more realistic assumptions including future electricity load projections, coal usage and variability in SO_2 emissions and has shown the following:

- The frequency of additional exceedances of the 1-hour average SO_2 criterion is predicted to be less than once per year outside the site boundary with the addition of the Bayswater B Power Station for all of the design configurations tested. The additional exceedances are predicted in unpopulated areas just to the east of Bayswater Power Station.

- Up to two additional exceedances of the 10-minute average criterion are predicted per year for the conventional (twin flue) stack configuration and up to one additional exceedance per year for the Heller stack. The additional exceedances are predicted outside populated areas.

- The predicted frequency of additional exceedances at the nearest towns and monitoring sites is less than one every 16 years for the 1-hour averaging period. One additional exceedance per year is predicted at Lake Liddell for the 10-minute average SO_2 criterion, with the frequency of exceedances predicted to be less than one per year at all other sites.

The above findings are based on various assumptions concerning the fuel used the existing and proposed power stations. These assumptions include:

- Predicted hourly load information for each power station for a projected year.
- Assessment of varying coal sulfur contents for each power station (stochastic emissions modelling).
- High sulfur mineral matter rejected during the coal pulverisation process and therefore not burnt for Liddell and Bayswater B generating units - conservatively estimated to be 30%.

The preliminary modelling presented in Katestone Environmental (2005b) investigated all three potential stack designs for the new power station. The modelling indicated that the two stack option resulted in higher ground-level impacts than the twin-flue and Heller options. Subsequent modelling, Katestone Environmental (2006a), did not include the two stack option.

A4 Inter-Regional Transport

In assessing the impact on air quality of a new development it is important to look at a range of geographical scales. Depending on the size of the development, there may be potential to impact at the local, regional, and possibly inter-regional scale. The inter-regional scale studies considered the potential for emissions of NO_X to be transported into more populated regions and influence the formation of photochemical smog.

In 2002, the CSIRO completed a study (IRTAPS) investigating the impact of the seven coal-fired power stations within and close to the GMR on air quality (Nelson *et al*., 2002a). IRTAPS identified a number of days in which inter-regional transport may have occurred. From all days identified, four case study periods were investigated in more detail. Results are presented in Nelson *et al*. (2002a, 2002b).

The study of the inter-regional transport (IRT) (Katestone Environmental 2006b) of emissions from the Bayswater B Power Station was conducted in conjunction with the CSIRO. Results from IRTAPS that relate to the present state of air quality associated with existing emission sources, formed the basis (or base case) for the study. The findings of IRTAPS were extended via the modification of the emissions inventory to include the Bayswater B Power Station.

The study found that there was no significant change to the predicted peak 1-hour or 4-hour average concentration of ozone (with values differing by less than 0.3%). Based on current modelling results, the proposed expansion of Bayswater Power Station was not found to have a significant effect on ozone levels within the Sydney airshed during the case days investigated, and the inclusion of the Bayswater B Power Station did not lead to additional exceedances of the 1-hour average impact assessment criterion for ozone within the study region.

Appendix B

Evaluation of TAPM Performance

Contents

Tables

Figures

B1 Introduction

In 2005, Katestone Environmental carried out a dispersion model selection study that identified the CSIRO's The Air Pollution Model (TAPM) as the preferred atmospheric plume dispersion model for the Macquarie Generation (MacGen) Bayswater B air quality impact assessment project. At that time, TAPM version 2 was the most up to date model version and this model was evaluated in comparison with the CALMET/CALPUFF modelling suite.

In the time since this study was carried out, two more recent versions of TAPM have been released. For the current impact assessment study for the MacGen Bayswater B project, TAPM version 4 (TAPMv4) has been employed. Consequently, a new study has been carried out to evaluate the performance of the TAPMv4. This evaluation includes a statistical comparison of TAPM predictions against measurements made at MacGen's network of monitoring stations for both meteorological conditions such as wind speed and direction, and ground-level concentrations of sulfur dioxide (SO_2).

It has been assumed that the new versions of the TAPM model have improved the model's overall performance, with this having been evaluated in several papers (Hurley 2005, 2003a, 2003b, Edwards 2004). This model evaluation study, carried out for the Bayswater B project, focuses on the performance of TAPMv4 to simulate the local meteorological conditions, including the distribution of wind speed and directions, in reasonable agreement to the local observations. The evaluation also focuses on the performance of TAPMv4 to predict ground-level concentrations of air contaminants (e.g., SO_2) at discrete sensitive locations where ambient air quality monitoring data is available, and assesses the correlation of the predictions to the observations.

The period modelled and compared with observations is the same as that used in the earlier Katestone Environmental (2005) study, 1 July 2000 to 30 June 2001. The study makes use of the same hourly varying emissions information of SO_2, as used in the 2005 evaluation for the Bayswater and Liddell Power Stations. Monitoring data collected at the five MacGen monitoring stations located at Singleton, Ravensworth, Lake Liddell, Mount Arthur North and Muswellbrook have been correlated with model predictions on a 1-hour average basis.

B2 Methodology

The performance of TAPMv4 was evaluated by in two ways:

1. A comparison of the predicted ground-level concentrations of SO_2 associated with emissions from both Bayswater and Liddell Power Stations with SO_2 measurements collected at the five ambient air quality monitoring stations operated by MacGen in the Hunter Valley.
2. A comparison of predicted and observed meteorological parameters at two surface meteorological monitoring stations operated by MacGen in the Hunter Valley and upper levels winds at the Bengalla Mine tower at three heights above the ground; 20, 40 and 90 metres.

B2.1 Background to Statistics Used

B2.1.1 Comparison of observed and predicted ground-level concentrations of SO_2

The evaluation was undertaken for SO_2 only, as the Bayswater and Liddell Power Stations are expected to be the main contributors to ambient concentrations of SO_2 in the region. Other minor contributors to SO_2 concentrations in the region include traffic, agriculture, mining activities and some minor influence from the more distant industrial facilities in the lower Hunter Valley, Newcastle and the Greater Metropolitan Region (GMR).

For this evaluation study, the same assessment period analysed for the Katestone Environmental (2005) study was evaluated, i.e., 1 July 2000 to 30 June 2001. For this period, 1-hour average measurements from each of the five ambient air quality monitoring stations (Singleton, Ravensworth, Liddell, Mount Arthur and Muswellbrook) were compared to TAPM predictions at each monitoring location. This approach provided for the comparison of two data sets, with a maximum of 8,760 (hours in a year) data points for each site evaluated.

There are three main approaches to the assessment of data for the evaluation of model performance (Chang and Hanna, 2005). This relates to the way the data sets are correlated. Hourly averaged data sets may be evaluated by pairing the data on the following bases:

1. In time and space; where the data at each monitoring location is compared with the hourly prediction on an hour by hour basis
2. In time only; such as the time series of the maximum pollutant concentrations anywhere in the domain of interest (i.e., no penalty is given if the model predicts the maximum concentration at the wrong location)
3. In space only; such as the spatial distribution of the maximum pollutant concentrations over a time period (i.e., no penalty is given if the model predicts the maximum concentration at the wrong time)

Method 1 is regarded as the strictest approach to model evaluation, as it compares the exact hourly prediction at a location with the measurement for that hour for the same location. This method uses a paired t-test, and is typically not used for evaluating the performance of dispersion models because if the wind predictions are only a couple of degrees out, by the time the plume has travelled 5-10 km it may miss the monitoring station completely. This method was carried out in the initial phase of the evaluation in order to analyse the variance between the observed and predicted data sets. The mean and standard deviations of the data sets have been provided as well as the Root Mean Square Error (RMSE), and both the

Systematic (RMSE_s) and Unsystematic (RMSE_u) RMSE values. A description of these statistics is provided in Section B2.1.1.

Method 2 is also a difficult correlation measure as accurate hourly emissions information and background ambient concentrations are required as well as the precise wind conditions, in order to compare predictions and observations on an hour by hour basis. This method provides for the hourly comparison of predicted impacts across the modelling domain rather than at a discrete location.

For air quality impact assessments, the primary concern is for the model to be able to predict the peak impacts in both magnitude and frequency. Consequently, Method 3 is the preferred approach for air dispersion model evaluation. Both predicted and observed data sets are sorted from highest to lowest before comparison. This allows for the comparison of various statistics including the maximum (100th percentile), 99.9th percentile (ninth highest concentration), 99th percentile and the Robust Highest Concentration (RHC). These types of comparisons are called up-paired comparisons.

Two of the three approaches listed above have been used in this evaluation. A preliminary assessment has been carried out using Method 1 to evaluate the model's overall performance in the strictest manner. This analysis provides a degree of thoroughness to the evaluation. As it is regarded the strictest approach, and poor correlation statistics do not infer that the model is entirely unsuitable for use in air quality impact assessment, further evaluation has been carried out using Method 3.

The aim of this correlation analysis, using Method 3, was to determine whether the model was able to predict the right magnitude and frequency of impacts at the selected locations, and for the right reasons. According to the analysis of the ambient monitoring information, the highest ground-level concentrations occur during daytime convective conditions (i.e. light winds and warm sunny days). Consequently, it is important to evaluate whether the model can predict not only the magnitude and frequency of these concentrations, but also during the same meteorological conditions.

B2.1.2 Comparison of observed and predicted wind fields

For the evaluation of the model's performance in simulating the wind fields in the region, a different statistical technique was used. This approach correlated the observed and predicted wind speeds on a time and space basis. The statistics used for this analysis include the mean, standard deviation, Pearson Correlation Coefficient, Index of Agreement, Root Mean Square Error, Systematic Root Mean Square Error, Unsystematic Root Mean Square Error, Skill_E, Skill_V and Skill_R. Each of these statistics is described in Section B2.2. This analysis provides for the evaluation of the model's ability to predict the right wind speeds during each hour of the day.

In order to evaluate the model's ability to predict the right wind direction for each hour of the day, wind speed must be included in the analysis and the entire wind field is broken down into its vector components, U and V. The statistics used for this analysis include the Root Mean Square Error, Systematic Root Mean Square Error, Unsystematic Root Mean Square Error, Index of Agreement, Skill_E, Skill_V, Skill_R and the Mean Absolute Error.

B2.2 Description of Statistics used for the Model Evaluation

B2.2.1 Correlation of Observed and Predicted Ground–level concentrations of Sulfur Dioxide

The following section describes the statistics used in the evaluation of model performance for the prediction of ground-level concentrations.

Root Mean Square Error (RMSE)

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(P_i - O_i)^2}$$

The RSME can be described as the standard deviation of the difference for hourly predicted and observed pairings at a specific point. The RMSE is a quadratic scoring rule, which measures the average magnitude of the error. The difference between predicted and corresponding observed values are each squared and then averaged over the sample. Finally, the square root of the average is taken. Since the errors are squared before they are averaged, the RMSE gives a relatively high weight to large errors. This means the RMSE is most useful when large errors are particularly undesirable. Overall, the RSME is a good overall measure of model performance, but since large errors are weighted heavily (due to squaring), its value can be distorted. RMSE is equal to the unit of the values being analysed i.e., an RMSE of 1.2 for wind speed = 1.2 m/s^{-1}.

Systematic Root Mean Square Error ($RMSE_s$)

$$\mathrm{RMSE_S} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{P}_i - O_i)^2}$$

The $RMSE_s$ is calculated as the square root of the mean square difference of hourly predictions from the regression formula and observation pairings, at a specific point. The regressed predictions are taken from the least squares formula. The $RMSE_s$ estimates the model's linear (or systematic) error. The systematic error is a measure of the bias in the model due to user input or model deficiency, i.e., data input errors, assimilation variables, and choice of model options. The $RMSE_s$ is a metric for the model's accuracy.

Unsystematic Root Mean Square Error ($RMSE_u$)

$$\mathrm{RMSE_U} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{P}_i - P_i)^2}$$

The $RMSE_u$ is calculated as the square root of the mean square difference of hourly predictions from the regression formula and model prediction value pairings, at a specific point. The $RMSE_u$ is a measure of how much of the difference between predictions and observations result from random processes or influences outside the legitimate range of the model. This error may require model refinement, such as new algorithms or higher resolution grids, or that the phenomena being simulated cannot be fully resolved by the model. The $RMSE_u$ is a metric for the model's precision.

Ultimately, for good model performance, the RMSE should be a low value, with most of the variation explained in the observations. Here, the systematic error $RMSE_s$ should approach zero and the unsystematic error, $RMSE_u$, should approach the RMSE since:

$$\mathrm{RMSE}^2 = \mathrm{RMSE}_s^2 + \mathrm{RMSE}_u^2$$

Mean Error and Mean Absolute Error

The Mean Error (ME) is simply the average of the hourly modelled values minus the hourly observed values. It contains both systematic and unsystematic errors and is heavily influence by high and low errors.

The Mean Absolute Error (MAE) measures the average magnitude of the errors in a set of predictions, without considering their direction. It measures accuracy for continuous variables. Expressed in words, the MAE is the average of the absolute values of the differences between predictions and the corresponding observation. The MAE is a linear score, which means that all the individual differences are weighted equally in the average. The MAE and the RMSE can be used together to diagnose the variation in the errors in a set of predictions. The RMSE will always be larger or equal to the MAE; the greater difference between them, the greater the variance in the individual errors in the sample. If the RMSE=MAE, then all the errors are of the same magnitude. Both the MAE and RMSE can range from 0 to ∞. They are negatively-oriented scores, i.e., lower values are better.

Fractional Bias

The fractional bias (FB) refers to the mean systematic difference between C_p and C_o, defined as:

$$\mathrm{FB} = \frac{(\overline{C_o} - \overline{C_p})}{0.5(\overline{C_o} + \overline{C_p})}$$

The FB is used when the data sets show a linear relationship. Consequently, the FB is strongly influenced by infrequently occurring high observed and predicted concentrations. For the FB, good model performance is reflected when the value approaches zero. Chang and Hanna (2004) found that for acceptable performing models, the mean bias is within ±30% of the mean (approximately |FB| < 0.3).

Normalised Mean Square Error

The normalised mean squared error (NMSE) is the squared difference between C_p and C_o, given by:

$$NMSE = \frac{\overline{(C_o - C_p)^2}}{\overline{C_o}\,\overline{C_p}}$$

In a similar manner to the FB, the NMSE is used when the data sets show a linear relationship, and are strongly influenced by infrequently occurring high observed and predicted concentrations. For the NMSE, good model performance is also reflected when the value approaches zero. Chang and Hanna (2004) found that for acceptable performing models, the random scatter is about a factor of two to three of the mean (i.e., approximately NMSE < 1.5).

Factor of 2
The fraction of predictions within a factor of 2 (FAC2) of the observed is defined as:

$$FAC2 = 0.5 \leq \frac{C_p}{C_o} \leq 2.0$$

The FAC2 is a more robust measure than the FB and NMSE because it is not overly influenced by high and low outliers. For the FAC2, good model performance is reflected when the value approaches one. Chang and Hanna (2004) found that for acceptable performing models, the fraction of predictions within a factor of two of observations is about 50% or greater (i.e., FAC2 > 0.5).

Fractional Bias – Ratio of false negatives (FB_{fn}) and false positives (FB_{fp})
The FB_{fn} can be considered as the under-predicting (false-negative) component of the fractional bias, i.e., only those (C_o, C_p) pairs with $C_p < C_o$ are considered in the calculation. Therefore, the value of FB_{fn} represents the percentage of under-predictions that are likely to be false. Similarly, the FB_{fp} can be considered as the over-predicting (false-positive) component of the fractional bias, i.e., only those (C_o, C_p) pairs with $C_p > C_o$ are considered in the calculation. Therefore, the value of FB_{fp} represents the percentage of over-predictions that are likely to be false.

The fractional bias of false negatives is defined as:

$$FB_{fn} = \frac{\frac{1}{2}\sum_i\left[\left|C_{oi} - C_{pi}\right| + \left(C_{oi} - C_{pi}\right)\right]}{\frac{1}{2}\cdot\sum_i\left(C_{oi} + C_{pi}\right)}$$

The fractional bias of false positives is defined as:

$$FB_{fp} = \frac{\frac{1}{2}\sum_i\left[\left|C_{oi} - C_{pi}\right| + \left(C_{pi} - C_{oi}\right)\right]}{\frac{1}{2}\cdot\sum_i\left(C_{oi} + C_{pi}\right)}$$

B2.2.2 Correlation of Observed and Predicted Wind Fields

The following section describes the statistics used in the evaluation of model performance for the simulation of wind speed and direction.

Index of agreement
The Index of Agreement (IOA) is defined as:

$$\mathrm{IOA} = 1 - \frac{\sum_{i=1}^{N}(P_i - O_i)^2}{\sum_{i=1}^{N}(|P_i - O_{mean}| + |O_i - O_{mean}|)^2}$$

The IOA is calculated using a method described in Willmott (1982). The IOA can take a value between 0 and 1, with 1 indicating perfect agreement. The IOA is the ratio of the total RMSE to the sum of two differences, i.e., the difference between each prediction and the observed mean, and the difference between each observation and observed mean. From another perspective, the IOA is a measure of the match between the departure of each prediction from the observed mean and the departure of each observation from the observed mean.

Note: N is the number of observations,
P_i are the hourly model predictions,
O_i are the hourly observations,

O_{mean} is the observed observation mean, and $\hat{P}_i = a + bO_i$ is the linear regression fitted with intercepts a and slope b.

Skill measures

Skill measure statistics are given in terms of a score, rather than in absolute terms. A model's skill can be measured by the difference in the standard deviation of the modelled and observed values.

The Skill_E (se) is indicative of how much of the standard deviation in the observations is predicted to be due to random/natural processes (unsystematic) in the atmospheric boundary layer. i.e., turbulence/chaos. For good model performance, the value for Skill_E should be less than one.

Skill_V (sv) is ratio of the standard deviation of the model predictions to the standard deviation of the observations. For good model performance, the value for Skill_V should be close to one.

SKILL_R (sr) takes into account systematic and unsystematic errors in relation to the observed standard deviation. For good model performance, the value for Skill_E should be less than one.

SKILL_E = (RMSE_U/ STDEV OBS) < 1 shows skill
SKILL_V = (STDEV_MOD/ STDEV _OBS) close to 1 shows skill
SKILL_R = (RMSE/ STDEV _OBS) < 1 shows skill

B3 Results of Correlation Analysis of Observed and Predicted Sulfur Dioxide Concentrations

B3.1 Analysis of variance for data paired in time and space

The first phase of the TAPMv4 performance evaluation was to compare predicted ground-level concentrations of SO_2 at each monitoring location with the measured concentrations for the entire one year dataset. This analysis was performed on the 1-hour average data, and where data capture by the ambient monitoring stations were incomplete, the predicted concentration for that hour was also removed.

Table B1 presents the mean and standard deviation for the observed and predicted data sets at each monitoring location, and the mean error (RMSE). The error can be further described by its systematic and unsystematic components, which are a measure of the model's accuracy and precision.

Table B1 Correlation statistics for air quality predictions for data paired in time and space (in µg/m³)

Parameter	Singleton	Ravensworth	Liddell	Mount Arthur	Muswellbrook
Number of observations (n)	8,747	8,690	8,724	8,760	8,760
Mean_observed	6.1	13.9	8.0	16.3	9.9
Standard deviation_ observed	14.9	34.7	27.7	38.4	28.0
Mean_predicted	3.2	6.6	3.2	7.3	4.4
Standard deviation_ predicted	16.9	25.1	22.4	30.6	23.0
RMSE	20.7	35.9	34.3	43.4	33.5
RMSE_Systematic	12.4	27.1	26.1	31.7	24.7
RMSE_Unsystematic	16.6	23.6	22.3	29.6	22.7

B3.1.1 Interpretation of results for data paired in time and space

As discussed above, a comparison of the means of the predicted and observed data sets is not an appropriate statistic for evaluating their correlation due to the range of the distributions. Rather the standard deviations of the predicted and observed data sets at each location, indicates a similar degree of variation about the mean for each.

The RSME is a key statistic in this analysis as it describes the average error between the observed concentration and the predicted concentration for each hour of the period evaluated (i.e., one year). The range of values of the RSME for the five monitoring stations is 20.7 – 43.4 µg/m^3 which constitutes between 3.6% and 7.6% of the 1-hour average ground-level concentration air quality objective for SO_2. The magnitude of this error is similar to the standard deviation of the observed data and consequently, within the range of the natural variation in the observed data.

The RSME can be further analysed by breaking it down in to its systematic and unsystematic components that describe the accuracy and precision of the model, respectively. This analysis indicates that the accuracy of the model is, on average, within 12.4 to 31.7 µg/m^3 from the observations, while the precision is between 16.6 and 29.6 µg/m^3 around that accuracy.

While the correlation of observed and predicted data sets using a paired in time and space approach is considered strict for the purposes of evaluating air dispersion model performance, the statistics indicate that the model has performed reasonably well at the five monitoring locations that constitute the locations of sensitive receptors in the region. It is interesting to note that the best performed locations are Singleton and Muswellbrook which make up the largest population centres in the local area. These two locations are also at the furthest extents of the modelling domain and downwind of the predominant wind flows from the southeast (Muswellbrook) and the northwest (Singleton).

B3.1.2 Limitations of the paired in time and space analysis

Notwithstanding the issues in pairing model predictions with observations for ground-level concentrations on an hour by hour basis, there are inherent differences in modelled and observed data sets that build in a bias to the results of the correlation analysis. For example, the TAPM model is limited only to computing power in its accuracy and resolution for predicting ground-level concentrations, i.e., TAPM can predict ground-level concentrations in micrograms per cubic metre to twenty-nine decimal places using a 32-bit processor. By comparison, the monitoring equipment used to measure ambient concentrations at the five monitoring stations, in accordance with the Australian Standard 3580.4.1:2008, have a lower detection limit (LDL) of 0.2 parts per hundred million (pphm) (5.7 $\mu g/m^3$ at 0^oC, 101.3 kPa) with a measurement uncertainty of 0.5 pphm (14.3 $\mu g/m^3$ at 0^oC, 101.3 kPa) at a 95% confidence limit. Consequently, the lowest observable measurement with any certainty is likely to be about 20 $\mu g/m^3$.

This can result in a very large difference between the lowest predicted and the lowest observed concentrations that will distort the predicted error of the correlation. As discussed above, the primary concern for air quality modelling is the prediction of the maximum concentrations. The large difference in the observed and predicted concentrations for the lower percentiles, generally when the wind is not blowing in the direction of the location being assessed, will significantly skew the correlation and influence the statistics used to evaluate the relationship. Consequently, some transformation of the data is required to adequately represent the relationship between the highest percentiles of observed and predicted concentrations.

In addition to measurement uncertainty, the standard deviation of the observed data set provides a measure of the natural variability in the observations. This means that for a standard deviation of 30 $\mu g/m^3$, an observation of 0 $\mu g/m^3$ could just as easily be equal to 30 $\mu g/m^3$ or somewhere in between. Consequently, further analysis was carried out on the data by sorting the data from highest to lowest values and omitting all corresponding data points for values below the standard deviation value of the observations. This analysis focussed on the model's performance in predicting the correct magnitude and frequency of the maximum hourly averaged concentrations and is paired in space only.

B3.2 Analysis of variance for data paired in space only

The second phase of the evaluation was carried out using the method described in the USEPA approved BOOT Statistical Model Evaluation Software Package (version 2) (Chang and Hanna, 2005). This approach correlates the magnitude of the impacts and frequency distributions of the observed and predicted data sets by analysing the maximum concentrations, the normalised mean square error, fractional bias and the model's tendency to over- or under-predict. This technique is the most appropriate for air dispersion models as it is important to evaluate whether the model is able to predict the magnitude and frequency of the highest concentrations and to determine whether the model tends to over- or under-predict ground-level concentrations at a particular location.

Table B2 presents the correlation statistics for the evaluation of the model's performance on a paired in space only basis. These correlation statistics assist with the interpretation of the quantile – quantile plots shown in Figure B1 to Figure B5.

Table B2 Correlation statistics for ground-level concentrations of sulfur dioxide for data paired in space only at the five monitoring stations (in µg/m³)

Parameter		Singleton	Ravensworth	Liddell	Mount Arthur	Muswellbrook
Comparison of observed versus predicted statistics						
Highest value	Observed	297	608	684	710	544
	Predicted	469	403	650	686	565
Ninth highest value	Observed	170	390	398	343	313
	Predicted	205	285	372	417	264
RHC	Observed	219	477	479	435	366
	Predicted	268	329	413	507	344
Mean	Observed	41	93	92	100	82
	Predicted	36	64	64	61	51
Standard deviation	Observed	34	74	94	68	62
	Predicted	47	54	84	71	64
Correlation statistics						
Number of observational pairs (n)		739	824	403	992	708
Bias		4.6	28.9	27.6	39.1	30.5
NMSE		0.14	0.21	0.16	0.30	0.25
Correlation coefficient (R)		0.992	0.997	0.995	0.972	0.988
Factor of 2		0.64	0.93	0.73	0.51	0.38
Fractional bias		0.12	0.37	0.35	0.49	0.46
Fractional bias	Ratio of false negatives	0.20	0.37	0.35	0.50	0.46
	Ratio of false positives	0.08	0.00	0.00	0.01	0.01

Table B3 presents the summary of the predicted/observed ratios for the key statistics such as the mean, maximum value (100th percentile), 9th highest value (99.9th percentile), 99th percentile and the Robust Highest Concentration (RHC). The RHC is the mean of the eleven highest concentrations. These statistics illustrate the model's performance in predicting the highest concentrations that are important for the assessment of air quality impacts. Figure B6 presents the data in graphical form.

Table B3 Comparative predicted/observed ratios of ground-level concentrations of sulfur dioxide at the five monitoring stations

Monitoring location	Parameter	Predicted/observed Ratio
Singleton	Mean	0.89
	Maximum	1.58
	RHC	1.22
	99.9th percentile (9th highest)	1.20
	99th percentile	1.10
Ravensworth	Mean	0.69
	Maximum	0.66
	RHC	0.69
	99.9th percentile (9th highest)	0.73
	99th percentile	0.70
Liddell	Mean	0.70
	Maximum	0.95
	RHC	0.86
	99.9th percentile (9th highest)	0.93
	99th percentile	0.71
Mount Arthur	Mean	0.61
	Maximum	0.97
	RHC	1.17
	99.9th percentile (9th highest)	1.22
	99th percentile	0.70
Muswellbrook	Mean	0.63
	Maximum	1.04
	RHC	0.94
	99.9th percentile (9th highest)	0.84
	99th percentile	0.81

The correlation statistics indicate that the models performance at predicting the high events (i.e. those events important for impact assessments) is very good at Liddell, Mount Arthur and Muswellbrook and slightly high at Singleton and slightly low at Ravensworth. The predicted mean concentration at all sites in under-predicted, with a mean error of between 27 µg/m^3 and 39 µg/m^3. This error may be explained in part by the model not containing all of the sources of SO_2 in the region. Previous studies carried out by Katestone Environmental (2005) for MacGen in the study region have found the maximum ground-level concentrations associated with emissions from the existing power stations occur during convective daytime conditions. Consequently, in order to ensure that the model is predicting the important high events during the daytime, an analysis of daytime conditions only has also been undertaken. This is presented in Section B3.3.

B3.2.1 Interpretation of results for data paired in space only

For the assessment of criteria air pollutants such as SO_2, nitrogen dioxide (NO_2), and solid particles, the DECC's Approved Methods (2005) stipulates the predicted maximum (100th percentile) concentration including background for the averaging period being assessed be compared with the air quality objective, while for air toxics, the ninth highest concentration (99.9thpercentile) in isolation should be compared. Consequently, the evaluation has focused on the performance of TAPMv4 to predict the maximum and ninth highest concentrations, and on the correlation of observed and predicted concentrations greater than the natural variability (standard deviation) of the observations.

The results of the correlation analysis indicate the following:

For the comparison of the highest values –

- The predicted/observed ratio indicates the model tends to over-predict the maximum (100th percentile) 1-hour average concentration of SO_2 at Singleton (1.58), while the model tends to under-predict the maximum concentration at Ravensworth (0.66).

- The highest values at Liddell (0.95) and Mount Arthur (0.97) and Muswellbrook (1.04) are predicted very well by the model, where the predicted maximum 1-hour average concentration of SO_2 at each location is within 5% of the observed maximum concentration, and the RHCs are within 17%. This indicates that the model is likely to perform well in the prediction of ground-level concentrations to the north and northwest.

- For the ninth highest value, the percentage difference between predictions and observations at all locations ranges between 7% and 27%.

For the correlation of the data sets –

- The correlation coefficients for observations versus predictions at the five monitoring locations are between 0.972 and 0.997. This indicates a good linear relationship at each location.

- The bias at the five locations ranges between 4.63 µg/m^3 (at Singleton) and 39.13 µg/m^3 (at Mount Arthur) and tends to be in relatively close agreement with the RSME for the entire data set. This indicates that, on average, the error in the model predictions are well within the range of the natural variability of the observations.

- The FAC2 values at Singleton (0.64), Ravensworth (0.93), Liddell (0.73) and Mount Arthur (0.51) are greater than 0.5, indicating acceptable model performance. At Muswellbrook the value is 0.38 and is the result of the model under-predicting the low end of the distribution of ground-level concentrations. This may be largely due to several factors including:
    - the distance between Muswellbrook and the Bayswater and Liddell Power Stations
    - the model's ability to simulate the winds that transport the plume toward this monitoring location
    - the contribution of other background SO_2 sources to observed ground-level concentrations
    - the model's inability to adequately simulate the re-circulation of the plume

- The FB indicates a good model performance at the Singleton (0.12) monitoring location, while the model appears to perform poorly at the other four locations: Ravensworth (0.37), Liddell (0.35), Mount Arthur (0.49) and Muswellbrook (0.46).

- The poor values for the FB and the tendency of the model to under-predict can be explained by the high proportion of false-negatives in the model predictions for Ravensworth (36.9%), Liddell (35.4%), Mount Arthur (49.8%) and Muswellbrook (46.3%), and the very small percentage of false-positives. This indicates that while the model tends to under-predict at these locations, between 35% and 50% of these predictions are false-negatives and may be explained by the natural variability in the observations.

B3.3 Analysis of variance for data paired in space only during the daytime

In order to further investigate the model's performance to predict the highest ground-level concentrations during the right meteorological conditions, i.e., during daytime convective conditions, the correlation analysis was carried out for the daytime in isolation.

Table B4 presents the correlation statistics for the evaluation of the model's performance on a paired in space only basis during the day. These correlation statistics assist with the interpretation of the quantile – quantile plots for the daytime period, shown in Figure B7 to Figure B11.

Table B4 Correlation statistics for ground-level concentrations of sulfur dioxide during the day for data paired in space only at the five monitoring stations (in µg/m³)

Parameter		Singleton	Ravensworth	Liddell	Mount Arthur	Muswellbrook
Comparison of observed versus predicted statistics						
Highest value	Observed	297	608	684	710	544
	Predicted	469	403	650	686	565
Ninth highest value	Observed	163	390	398	334	265
	Predicted	204	279	314	383	264
RHC	Observed	197	477	480	403	357
	Predicted	267	321	394	500	344
Mean	Observed	8	22	11	24	15
	Predicted	5	12	5	12	7
Standard deviation	Observed	18	47	37	48	36
	Predicted	22	32	29	40	31
Correlation statistics						
Number of observational pairs (n)		453	563	230	697	488
Bias		-0.7	38.5	41.7	39.2	33.4
NMSE		0.13	0.23	0.19	0.25	0.21
Correlation coefficient (R)		0.994	0.997	0.994	0.960	0.987
Factor of 2		0.88	1.00	0.72	0.64	0.42
Fractional bias		-0.01	0.40	0.39	0.42	0.42
Fractional bias	Ratio of false negatives	0.11	0.40	0.39	0.44	0.43
	Ratio of false positives	0.12	0.00	0.00	0.02	0.01

Table B5 presents the summary of the predicted/observed ratios for the key statistics such as the mean, maximum value (100th percentile), 9th highest value (99.9th percentile), 99th percentile and the Robust Highest Concentration (RHC). These statistics illustrate the model's performance in predicting the highest concentrations during the day, which are important for the assessment of air quality impacts. Figure B12 presents the data in graphical form.

Table B5 Comparative predicted/observed ratios of ground-level concentrations of sulfur dioxide during the day at the five monitoring stations

Monitoring location	Parameter	Predicted/observed Ratio
Singleton	Mean	0.67
	Maximum	1.58
	RHC	1.36
	99.9th percentile (9th highest)	1.25
	99th percentile	1.14
Ravensworth	Mean	0.52
	Maximum	0.66
	RHC	0.67
	99.9th percentile (9th highest)	0.72
	99th percentile	0.67
Liddell	Mean	0.43
	Maximum	0.95
	RHC	0.82
	99.9th percentile (9th highest)	0.79
	99th percentile	0.69
Mount Arthur	Mean	0.52
	Maximum	0.97
	RHC	1.24
	99.9th percentile (9th highest)	1.15
	99th percentile	0.70
Muswellbrook	Mean	0.50
	Maximum	1.04
	RHC	0.96
	99.9th percentile (9th highest)	0.89
	99th percentile	0.79

B3.4 Interpretation of results for data paired in space only during the daytime

The analysis of unpaired data during the daytime only indicates that the majority of the highest ground-level concentrations occur during the day. There is little difference in the correlation for the highest concentrations between the analysis of the entire modelling period and the daytime period only. This indicates that the model predicts that the highest concentrations will occur during daytime convective conditions, the same conditions illustrated by the observations.

B4 Results of Correlation Analysis of Observed and Predicted Meteorological Fields

B4.1 Surface winds

The evaluation of the performance of TAPMv4 to simulate the wind fields in the local area was carried out through a correlation analysis of the predicted and observed wind speed and direction at two of the ambient meteorological stations operated by MacGen. The two meteorological stations analysed were Liddell and Ravensworth.

The analysis investigated the correlation between the distributions of predicted and observed wind speed at both meteorological stations, and the relationship between both wind speed and direction by breaking the wind fields down in to their vector components, U and V.

Table B6 presents the correlation statistics for the comparison of the observed and predicted wind speed distributions.

Table B6 Correlation statistics for observed versus predicted wind speed

Parameter	Liddell	Ravensworth
Number of observations (n)	7,800	5,832
Mean_observed	2.4	2.6
Mean_predicted	3.5	3.0
Standard deviation_observed	1.6	1.9
Standard deviation _predicted	1.8	1.9
Pearson Correlation Coefficient	0.53	0.71
Index Of Agreement	0.68	0.83
RMSE	1.99	1.52
RMSE_ Systematic	1.23	0.73
RMSE_ Unsystematic	1.57	1.33
Skill_E	0.96	0.68
Skill_V	1.13	0.97
Skill_R	1.22	0.78

Table B7 presents the correlation statistics for the comparison of U and V vector components for comparison of the observed and predicted wind direction and speed.

Table B7 Correlation statistics for observed versus predicted U and V vectors

Parameter	Liddell		Ravensworth	
	U	V	U	V
RMSE	1.91	2.08	1.63	1.63
RMSE_ Systematic	0.31	0.11	0.41	0.53
RMSE_ Unsystematic	1.89	2.08	1.57	1.55
Index Of Agreement	0.86	0.71	0.90	0.83
Skill_E	0.76	1.37	0.61	0.76
Skill_V	1.16	1.74	1.00	1.04
Skill_R	0.77	1.37	0.63	0.81
Mean Absolute Error	1.4	1.6	1.2	1.2

B4.2 Upper level winds

The evaluation of the performance of TAPMv4 to simulate the wind fields in the upper levels was carried out through a correlation analysis of the predicted and observed wind speed and direction at the Bengalla Mines monitoring station.

The analysis investigated the correlation between the distributions of predicted and observed wind speed, and the relationship between both wind speed and direction by breaking the wind fields down in to their vector components, U and V.

Table B8 presents the correlation statistics for the comparison of the observed and predicted wind speed distributions.

Table B8 Correlation statistics for observed versus predicted wind speed at various heights

Parameter	Bengalla 20 m	Bengalla 40 m	Bengalla 90 m
Number of observations (n)	6,149	6,759	5,699
Mean_observed	2.9	3.5	3.9
Mean_predicted	3.6	4.1	4.8
Standard deviation_observed	2.6	2.6	2.8
Standard deviation _predicted	2.1	2.3	2.7
Pearson Correlation Coefficient	0.60	0.60	0.61
Index Of Agreement	0.75	0.76	0.76
RMSE	2.24	2.34	2.59
RMSE_ Systematic	1.49	1.45	1.48
RMSE_ Unsystematic	1.67	1.84	2.13
Skill_E	0.65	0.69	0.76
Skill_V	0.81	0.86	0.96
Skill_R	0.87	0.88	0.93

Table B9 presents the correlation statistics for the comparison of U and V vector components for comparison of the observed and predicted wind direction and speed.

Table B9 Correlation statistics for observed versus predicted U and V vectors

Parameter	Bengalla 20 m		Bengalla 40 m		Bengalla 90 m	
	U	V	U	V	U	V
RMSE	2.50	2.94	2.55	2.88	2.81	3.40
RMSE_ Systematic	0.65	1.76	0.94	1.29	0.64	1.58
RMSE_ Unsystematic	2.41	2.35	2.36	2.57	2.73	3.01
Index Of Agreement	0.80	0.66	0.85	0.71	0.86	0.70
Skill_E	0.93	0.81	0.71	0.94	0.77	0.93
Skill_V	1.24	0.91	1.06	1.11	1.18	1.08
Skill_R	0.96	1.02	0.76	1.06	0.80	1.05
Mean Absolute Error	1.9	2.3	1.9	2.2	2.1	2.6

B4.3 Interpretation of results for wind field correlation analysis

The results of the wind field correlation analysis indicates the following:

In relation to the model's ability to simulate the wind speed –

- At Liddell, the mean predicted wind speed is 1.1 m/s higher than the mean observed wind speed, while the standard deviation for the model predictions is 0.2 m/s greater than the observed standard deviation. At Ravensworth, the mean predicted wind speed is 0.4 m/s higher than the mean observed wind speed, while the standard deviation for the model predictions is equal to the observed standard deviation.

- The upper level Bengalla winds show the predicted wind speed to be 0.6-0.9 m/s higher across the three heights, while the standard deviation for the model predictions is 0.1 to 0.5 m/s greater than the observed standard deviation. This indicates that, on average, the model tends to slightly over-predict the wind speed at both locations. However, the variation around the mean for the observations and predictions is similar.

- The Pearson Correlation Coefficient and IOA at both surface and upper level locations indicate reasonable model performance. While perfect model performance is measured as the value of each measure approaching one, it is generally agreed that an IOA of 0.5 or higher for wind speeds compared on a time and space basis indicates more than reasonable model performance in the simulation of wind speed.

- At Ravensworth, the Skill_R ratio (mean error (RMSE) / observed standard deviation) and Skill_E ratio (unsystematic error (RMSE_u) / observed standard deviation) are both less than one, while the Skill_V ratio (predicted standard deviation / observed standard deviation) is very close to one. This is a measure of good model performance in the simulation of wind speed.

- At Liddell, the Skill_E ratio (unsystematic error (RMSE_u) / observed standard deviation) is slightly less than one, while the Skill_R ratio (mean error (RMSE) / observed standard deviation) is significantly greater than one. The Skill_V ratio (predicted standard deviation / observed standard deviation) is reasonably close to one. This indicates that the model performs less well at Liddell for the simulation of wind speeds.

- At Bengalla, the Skill_R ratio (mean error (RMSE) / observed standard deviation) and Skill_E ratio (unsystematic error (RMSE_u) / observed standard deviation) are both less than one, while the Skill_V ratio (predicted standard deviation / observed standard deviation) is very close to one, particularly at 90 metres. This is a measure of good model performance in the simulation of wind speed at upper levels.

In relation to the model's ability to simulate the wind speed and direction –

- At Ravensworth, the model appears to under-predict the frequency of winds around 1 m/s by about 10%, while over-predicting the frequency of winds between 2 - 5 m/s by about 5%. For wind direction, there appears to be a good agreement in the distributions of observed and predicted winds, with a slight under-prediction of the frequency of winds from the southeast (~ 5%) and a slight over-prediction of the frequency of winds from the northwest (~ 5%).

- At Liddell, the model appears to under-predict the frequency of winds around 1 m/s by about 15%, and between 2 – 3 m/s by approximately 3%, while over-predicting the frequency of winds between 4 - 8 m/s by between 2% and 8%. For wind direction, there appears to be a reasonable agreement in the distributions of observed and predicted winds, with a slight over-prediction of the frequency of winds from the southeast (~ 5%) and a slight over-prediction of the frequency of winds from the northwest (~ 3%). However, the main difference in the predicted versus observed winds at Liddell is illustrated by the thirty degree difference in the peak angle of the north-westerlies. While the observations at Liddell indicate that the highest frequency of winds from the northwest quadrant blow from around 300^o, the model predicts the angle with the highest frequency to be from 330^o. This difference is highlighted in the significant difference in the observed and predicted vector V component discussed below.

- The IOA for both U and V vector components indicate the model performs very well in predicting the right wind speeds from the right directions.

- This is further supported at Ravensworth by the skill measures as the Skill_V is equal to one for the U component and 1.04 for the V component. The Skill_E and Skill_R measures are also significantly less than one, at 0.61 (U) and 0.76 (V), and 0.77 (U) and 0.81 (V), respectively.

- At Liddell, the skill measures for the U component indicate the model performs well in predicting the north-south element of the wind speed and direction, but less well for the V component, or the east-west element of the wind speed and direction. As discussed above, this is related to the 30^o difference in observed versus predicted winds from the northwest quadrant. This may be explained by particularities in the local area that channel winds from the northwest at a slightly more westerly angle at Liddell to that predicted by TAPM.

- At Bengalla, the skill measures for the U component indicate the model performs well in predicting the north-south element of the wind speed and direction, but less well for the V component, or the east-west element of the wind speed and direction.

- The mean absolute error for each of the U and V components at Ravensworth is 1.2, while at Liddell the error is 1.4 and 1.6 for the U and V components, respectively. At Ravensworth the mean absolute error for the U and V components ranged from 1.9 to 2.1 in the U and 2.2 to 2.6 in the V component. This is a reasonably small degree of error between the predictions and observations and can largely be explained by the natural variability in the wind direction when calculating the 1-hour average.

B4.4 Temperature

The evaluation of the performance of TAPMv4 to simulate the temperature was carried out through a correlation analysis of the predicted and observed temperature at the Bengalla Mines monitoring station and at Liddell.

Table B10 presents the correlation statistics for the comparison of the observed and predicted temperature distributions.

Table B10 Correlation statistics for observed versus predicted temperature

Parameter	Liddell	Bengalla 20 m	Bengalla 40 m	Bengalla 90 m
Number of observations (n)	7,796	6,178	6,787	5,724
Mean_observed	17.3	16.8	17.1	15.6
Mean_predicted	18.2	17.6	17.7	16.0
Standard deviation_observed	6.7	7.4	7.3	6.0
Standard deviation _predicted	6.3	6.7	6.6	5.6
Pearson Correlation Coefficient	0.91	0.91	0.91	0.90
Index Of Agreement	0.95	0.95	0.95	0.95
RMSE	2.9	3.1	3.0	2.6
RMSE_ Systematic	1.4	1.4	1.4	1.1
RMSE_ Unsystematic	2.6	2.7	2.7	2.4
Skill_E	0.39	0.37	0.37	0.40
Skill_V	0.93	0.91	0.91	0.93
Skill_R	0.44	0.42	0.42	0.44
Mean Absolute Error	2.0	2.4	5.1	4.6

The results of the temperature correlation analysis indicates the following:

- At Liddell, the mean predicted temperature is 0.9 °C higher than the mean observed temperature, while the standard deviation for the model predictions is 0.4 °C lower than the observed standard deviation. The upper level Bengalla temperatures show the predicted temperature to be 0.4-0.8 °C higher across the three heights, while the standard deviation for the model predictions is 0.4 to 0.7 °C lower than the observed standard deviation. This indicates that, on average, the model tends to slightly over-predict the temperature at both locations and all three heights. However, the variation around the mean for the observations and predictions is similar.

- The Pearson Correlation Coefficient and IOA at both surface and upper level locations indicate very good model performance.

- At Liddell and Bengalla, the Skill_R ratio (mean error (RMSE) / observed standard deviation) and Skill_E ratio (unsystematic error (RMSE_u) / observed standard deviation) are both less than one, while the Skill_V ratio (predicted standard deviation / observed standard deviation) is very close to one. This is a measure of good model performance in the simulation of temperature.

- The mean absolute error for temperature at Liddell is 2.0 °C while for Bengalla the error increased from 2.4 °C at 20 metres to 5.1 °C and 4.6 °C, at 40 metres and 90 metres respectively.

B5 Conclusions

An evaluation of the performance of TAPMv4 has been carried out using statistical techniques. The evaluation has included a comparison of observed and predicted ground-level concentrations of SO_2 at five ambient monitoring stations, and the correlation of observed and predicted meteorological fields at surface and upper level monitoring stations. Several conclusions can be drawn from the evaluation of the TAPMv4 model, including:

- The evaluation of the models ability to simulate the local winds and temperature indicates good general agreement between the predictions and observations at both the surface and upper levels.

- Overall, the model performs well with regard to the prediction of the maximum 1-hour average ground-level concentrations, particularly at Liddell, Mount Arthur and Muswellbrook, and is considered suitable for use in the assessment of criteria pollutants.

- Overall, the model performs reasonably well with regard to the prediction of the ninth highest 1-hour average ground-level concentrations, and is considered suitable for use in the assessment of non-criteria air pollutants.

- For the correlation of ground-level concentrations of SO_2 at the five monitoring stations, the location which performed least well in the prediction of the highest percentiles (maximum, ninth highest and RHC), Ravensworth, is shown to illustrate good skill in the prediction of wind speed and direction.

- The bias illustrated in the correlation statistics at Ravensworth, Liddell, Mount Arthur and Muswellbrook indicate the model's tendency to under-predict ground-level concentrations of SO_2. Further analysis shows that the model tends to under-estimate ground-level concentrations of SO_2 when the observed SO_2 concentration is less than 100 $\mu g/m^3$, and particularly when less than 50 $\mu g/m^3$, by approximately 30 $\mu g/m^3$ - 40 $\mu g/m^3$ for between 35% and 50% of the time. This result is less important for the assessment of impacts from the proposed power stations where the 100th percentile and 99.9th percentile are used and can be largely explained by the issues discussed, including the LDL and measurement uncertainty of the ambient monitors and the contribution of other sources of SO_2 in the region.

B6 References

Chang, J.C. and Hanna, S.R. 2005: Technical description and user's guide for the BOOT statistical model evaluation software package, Version 2.0. http://www.harmo.org/Kit/default.asp

Edwards, M., Hurley P. J, and Physick W. L, 2004, 'Verification of TAPM meteorological predictions using sodar data in the Kalgoorlie region', Aust. Met. Mag., **53**, 29–37.

Hurley P 2005: "The Air Pollution Model (TAPM) Version 3: User Manual". CSIRO Division of Atmospheric Research Internal Paper No. 31. http://www.cmar.csiro.au/e-print/open/hurley_2005b.pdf

Hurley P.J., Physick W.L. Luhar A.K.,and Edwards, M. 2005: 'The Air Pollution Model (TAPM) Version 3', Part 2: Summary of some verification studies', CSIRO Atmospheric Research, Technical Paper No. 72. 39

Hurley P.J., Manins P.C., Lee S., Boyle R., Ng Y. and Dewundege P., 2003a, 'Year-long, high-resolution, urban airshed modelling: Verification of TAPM predictions of smog and particles in Melbourne, Australia', Atmos. Environ., **37**, 1899-1910.

Hurley, P., Physick, W., Cope, M., Borgas, M., and Brace, P., 2003b, 'An evaluation of TAPM for photochemical smog applications in the Pilbara region of WA', Proceedings of the 17th International Clean Air and Environment Conference of CASANZ, Newcastle, Australia, 23 27 November 2003.

Katestone Environmental, 2005b. A report from Katestone Environmental to Macquarie Generation: Preliminary air dispersion modelling of Bayswater B Power Station, June 2005.

Willmott, C.J., 1982: Some comments on the evaluation of model performance. *Bull. Amer. Meteor. Soc.*, **63**, 1309-1313.

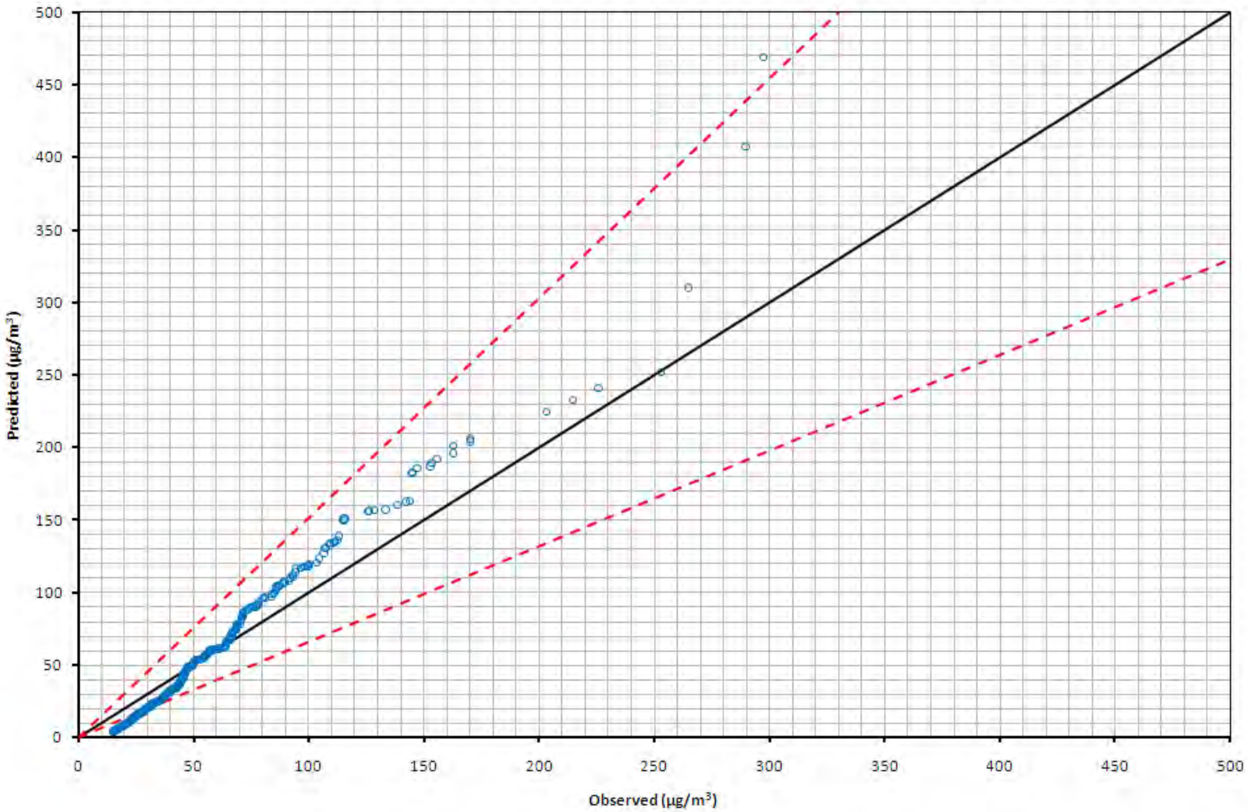

Figure B1 **Correlation of observed versus predicted ground-level concentrations of sulfur dioxide at Singleton**			
Location: Singleton monitoring station	**Period:** 1/07/00 – 30/06/01	**Data source:** Observed vs TAPM predicted	**Units:** µg/m³
Type: Quantile – quantile plot		**Prepared by:** A. Balch	**Date:** August 2009

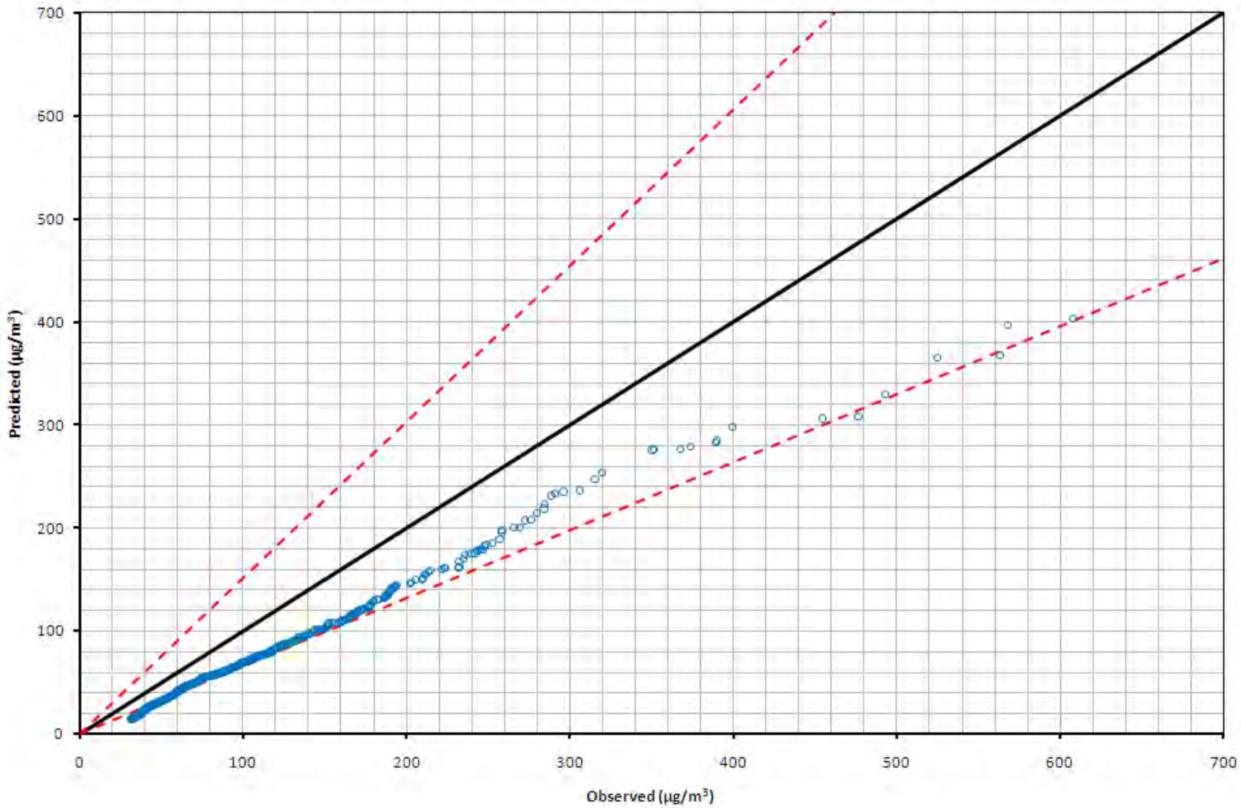

Figure B2 Correlation of observed versus predicted ground-level concentrations of sulfur dioxide at Ravensworth

Location: Ravensworth monitoring station	**Period:** 1/07/00 – 30/06/01	**Data source:** Observed vs TAPM predicted	**Units:** µg/m³
Type: Quantile – quantile plot		**Prepared by:** A. Balch	**Date:** August 2009

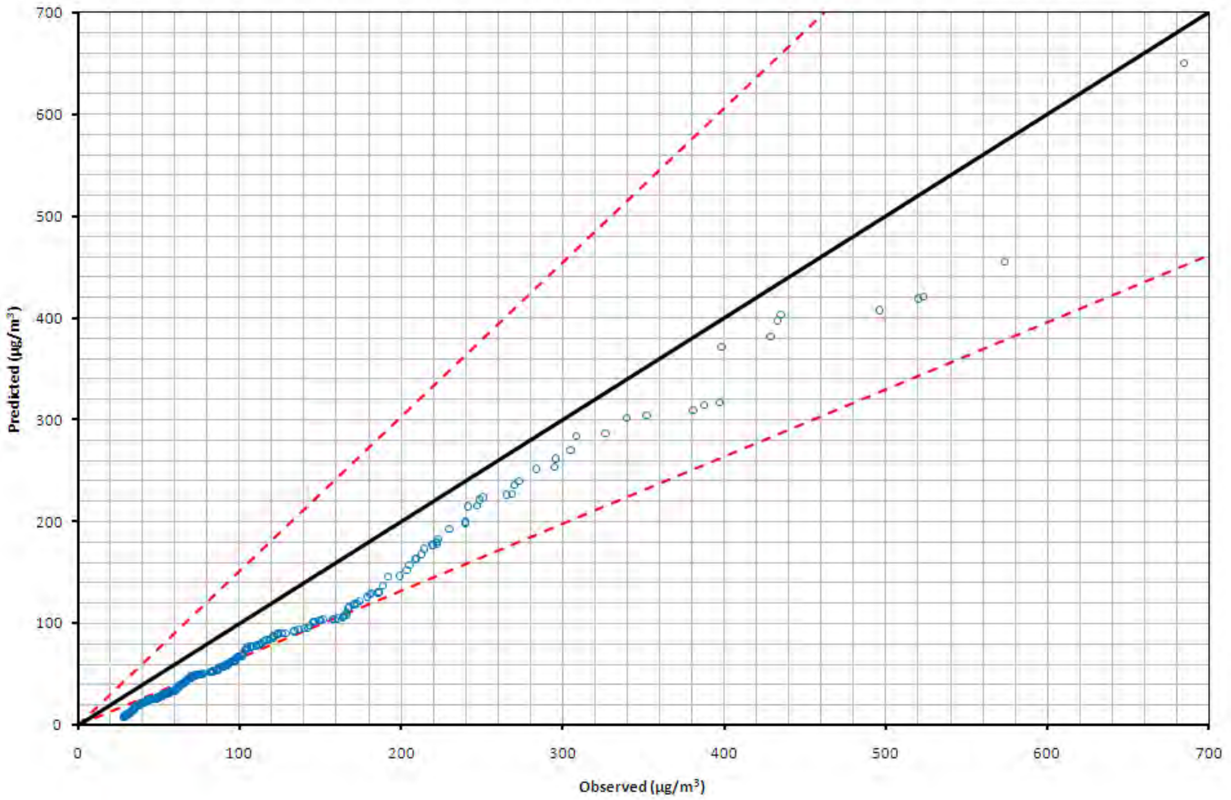

Figure B3 Correlation of observed versus predicted ground-level concentrations of sulfur dioxide at Liddell			
Location: Liddell monitoring station	**Period:** 1/07/00 – 30/06/01	**Data source:** Observed vs TAPM predicted	**Units:** µg/m³
Type: Quantile – quantile plot		**Prepared by:** A. Balch	**Date:** August 2009

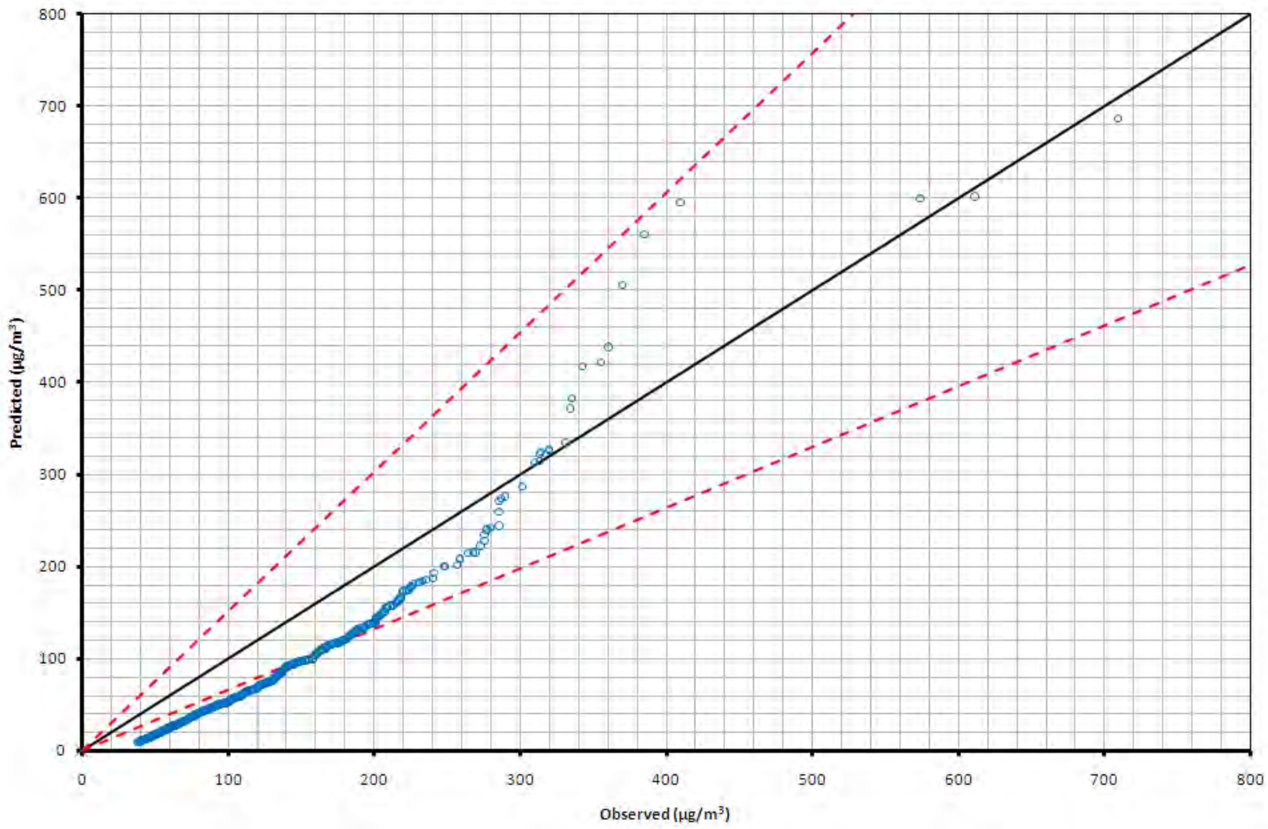

Figure B4 Correlation of observed versus predicted ground-level concentrations of sulfur dioxide at Mount Arthur			
Location: Mount Arthur monitoring station	**Period:** 1/07/00 – 30/06/01	**Data source:** Observed vs TAPM predicted	**Units:** µg/m³
Type: Quantile – quantile plot		**Prepared by:** A. Balch	**Date:** August 2009

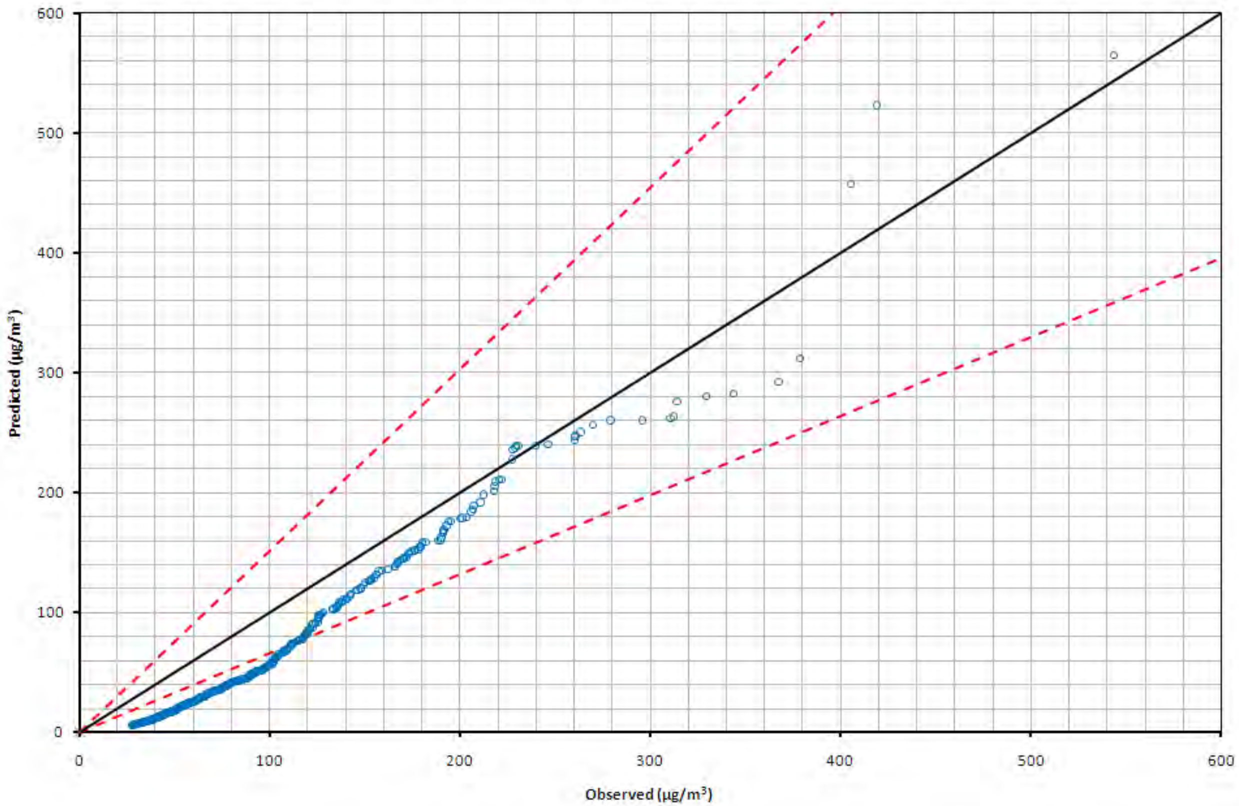

Figure B5 Correlation of observed versus predicted ground-level concentrations of sulfur dioxide at Muswellbrook			
Location: Muswellbrook monitoring station	**Period:** 1/07/00 – 30/06/01	**Data source:** Observed vs TAPM predicted	**Units:** µg/m³
Type: Quantile – quantile plot		**Prepared by:** A. Balch	**Date:** August 2009

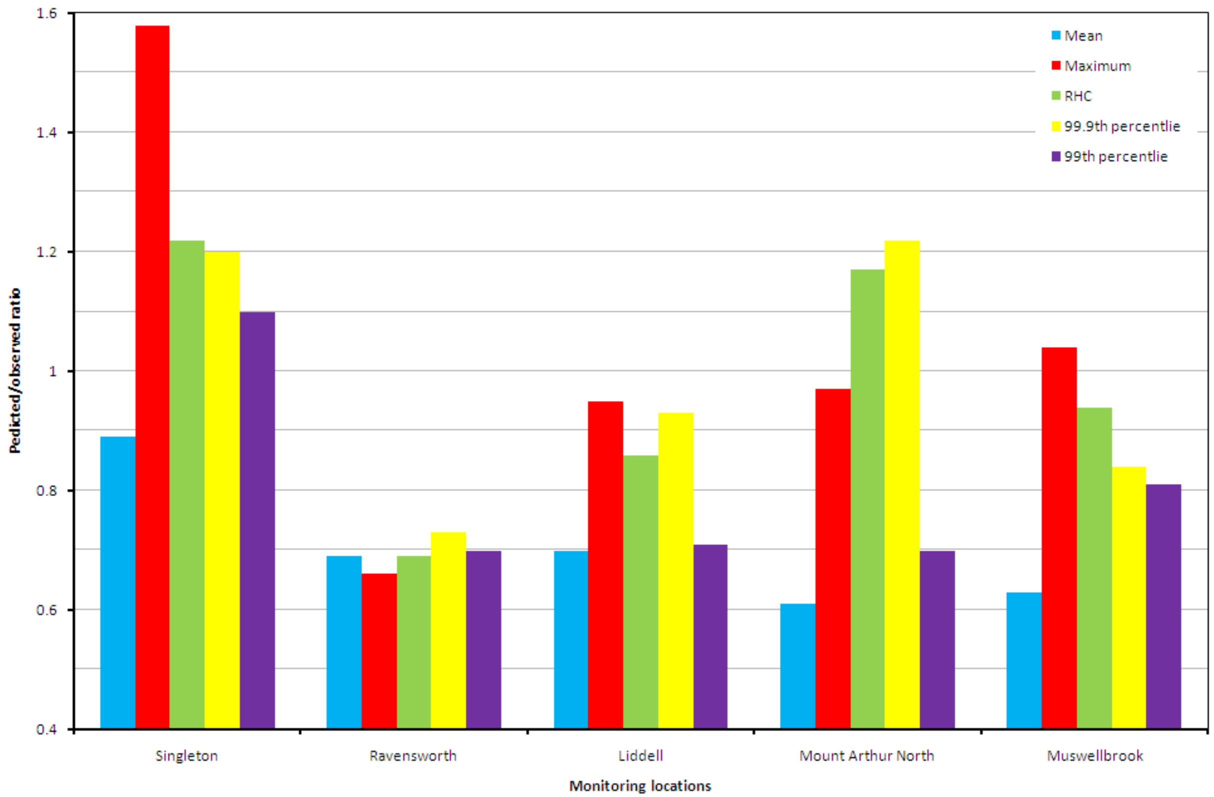

Figure B6 Predicted/observed ratios for the mean, Robust Highest Concentration, 100th, 99.9th and 99th percentiles at each of the five monitoring locations			
Location: Five monitoring station	**Period:** 1/07/00 – 30/06/01	**Data source:** Observed vs TAPM predicted	**Units:** N/A
Type: Histogram		**Prepared by:** A. Balch	**Date:** August 2009

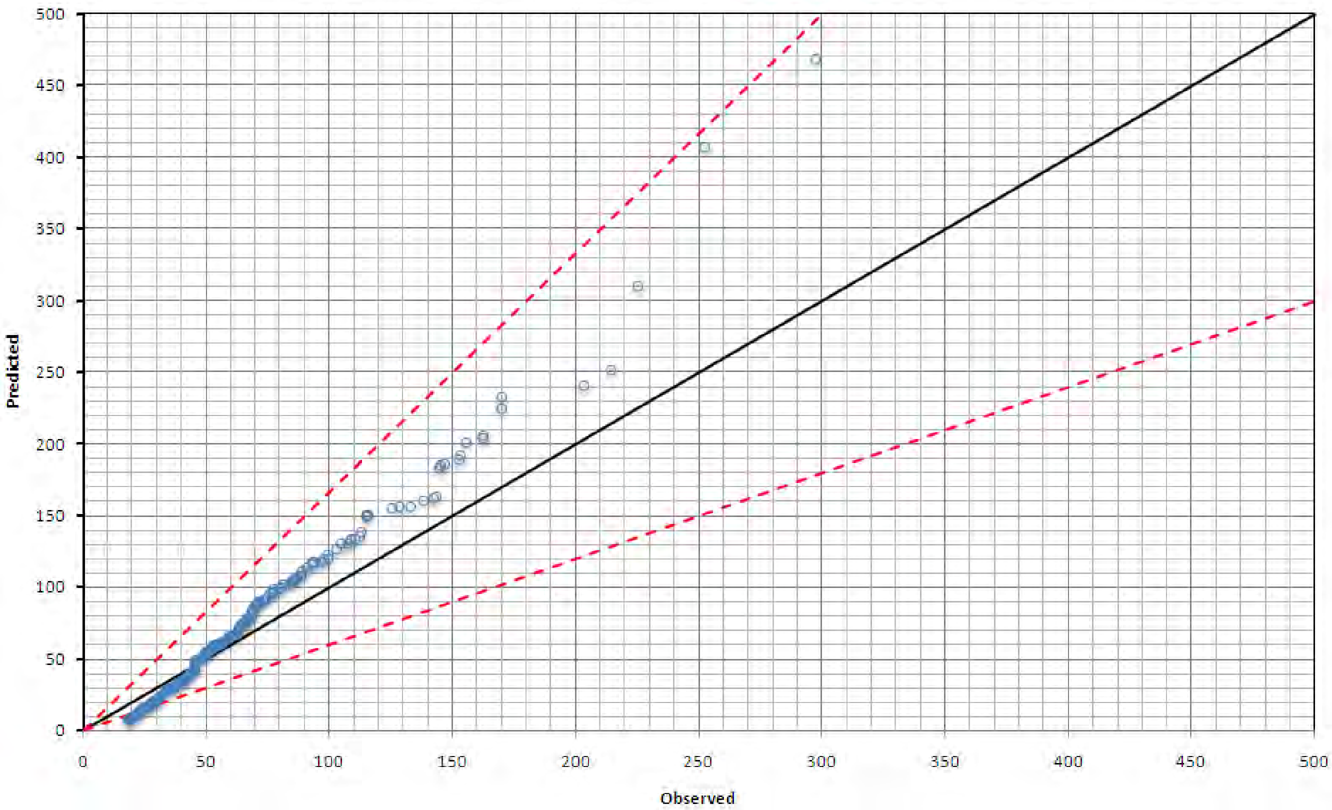

Figure B7 Correlation of observed versus predicted ground-level concentrations of sulfur dioxide at Singleton during the day only			
Location: Singleton monitoring station	**Period:** 1/07/00 – 30/06/01	**Data source:** Observed vs TAPM predicted	**Units:** µg/m³
Type: Quantile – quantile plot		**Prepared by:** A. Balch	**Date:** August 2009

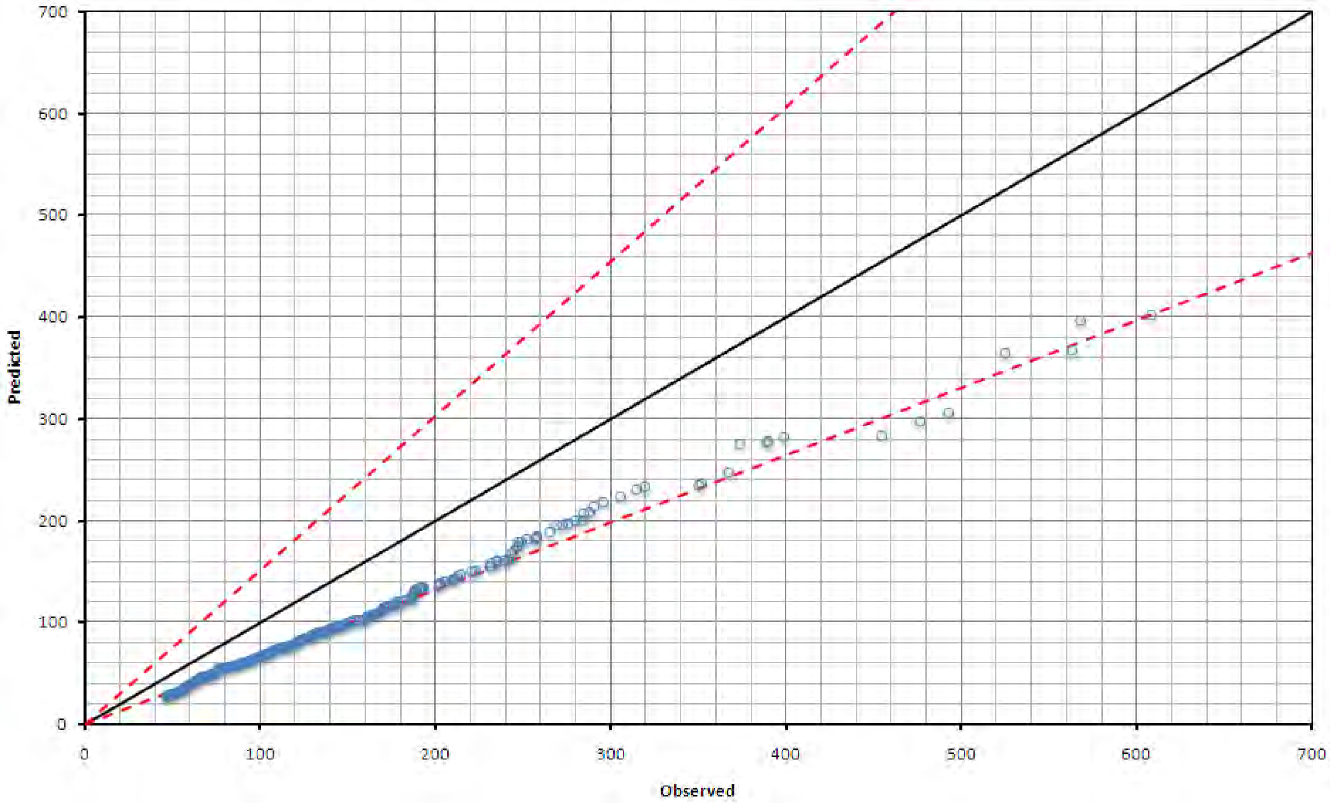

Figure B8 Correlation of observed versus predicted ground-level concentrations of sulfur dioxide at Ravensworth during the day only			
Location: Ravensworth monitoring station	**Period:** 1/07/00 – 30/06/01	**Data source:** Observed vs TAPM predicted	**Units:** µg/m^3
Type: Quantile – quantile plot		**Prepared by:** A. Balch	**Date:** August 2009

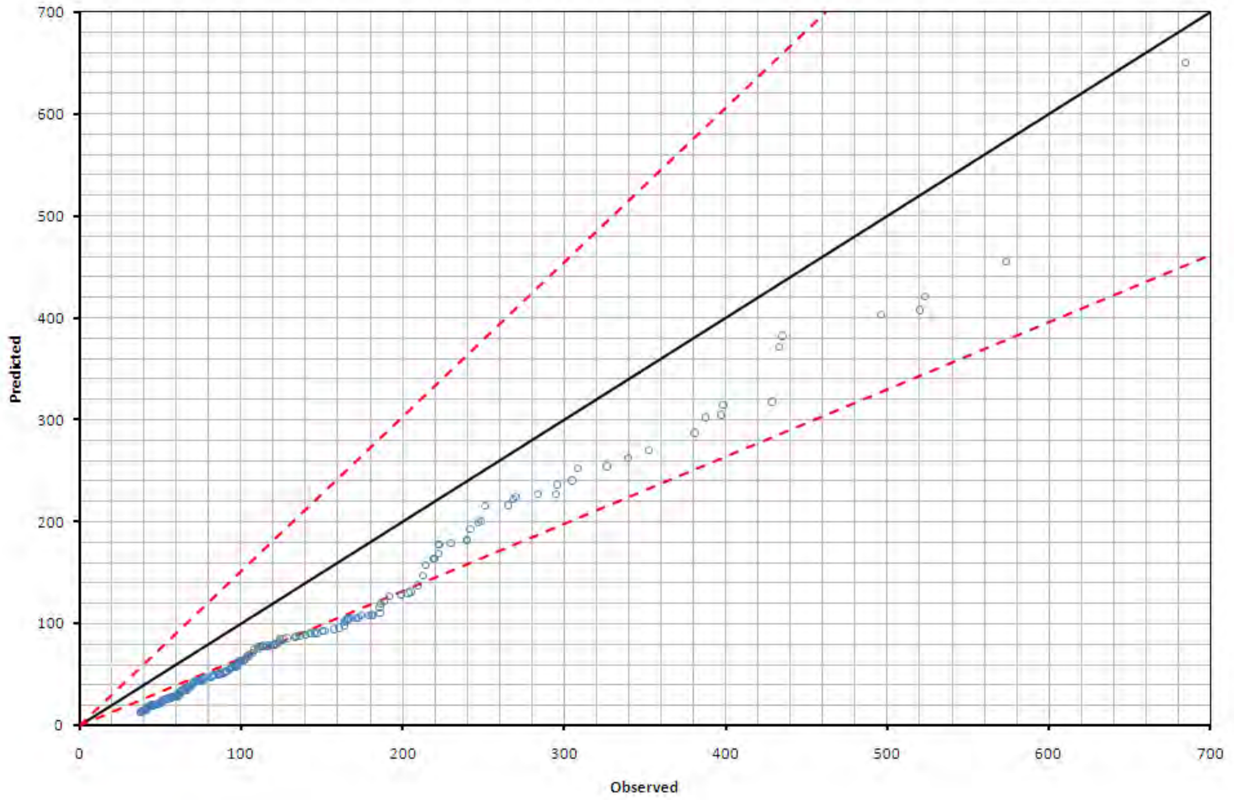

Figure B9 Correlation of observed versus predicted ground-level concentrations of sulfur dioxide at Liddell during the day only			
Location: Liddell monitoring station	**Period:** 1/07/00 – 30/06/01	**Data source:** Observed vs TAPM predicted	**Units:** µg/m³
Type: Quantile – quantile plot		**Prepared by:** A. Balch	**Date:** August 2009

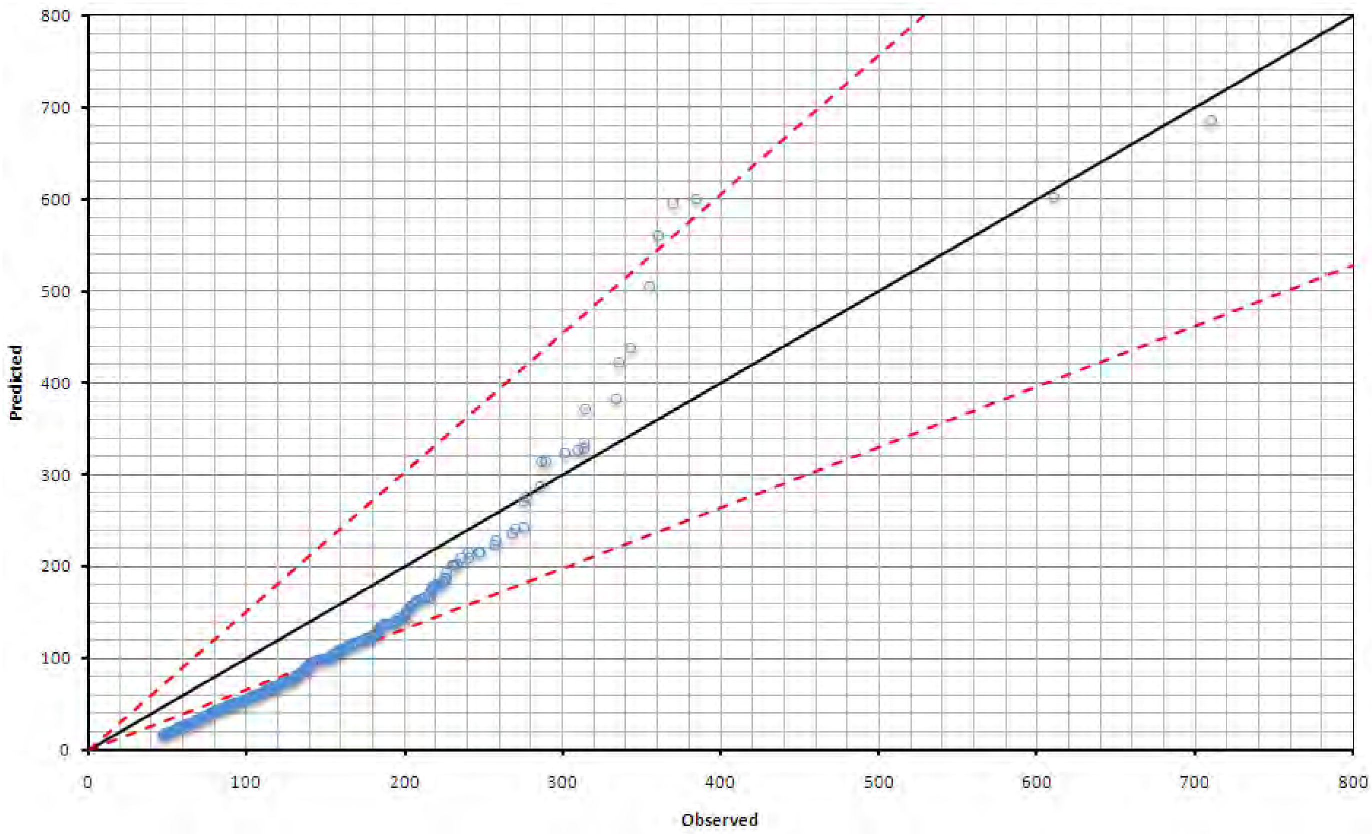

Figure B10	**Correlation of observed versus predicted ground-level concentrations of sulfur dioxide at Mount Arthur during the day only**		
Location: Mount Arthur monitoring station	**Period:** 1/07/00 – 30/06/01	**Data source:** Observed vs TAPM predicted	**Units:** µg/m³
Type: Quantile – quantile plot		**Prepared by:** A. Balch	**Date:** August 2009

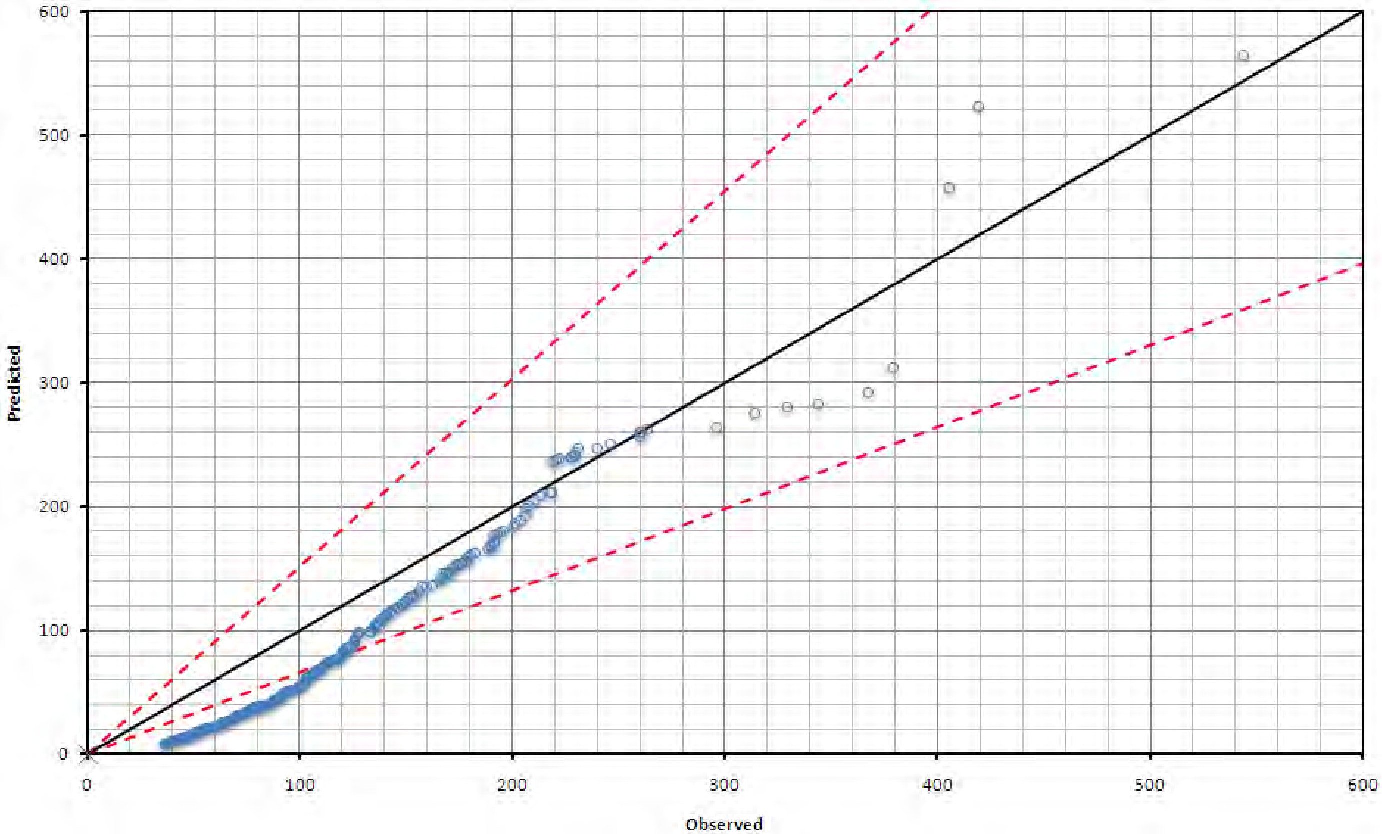

Figure B11 Correlation of observed versus predicted ground-level concentrations of sulfur dioxide at Muswellbrook during the day only			
Location: Muswellbrook monitoring station	**Period:** 1/07/00 – 30/06/01	**Data source:** Observed vs TAPM predicted	**Units:** $\mu g/m^3$
Type: Quantile – quantile plot		**Prepared by:** A. Balch	**Date:** August 2009

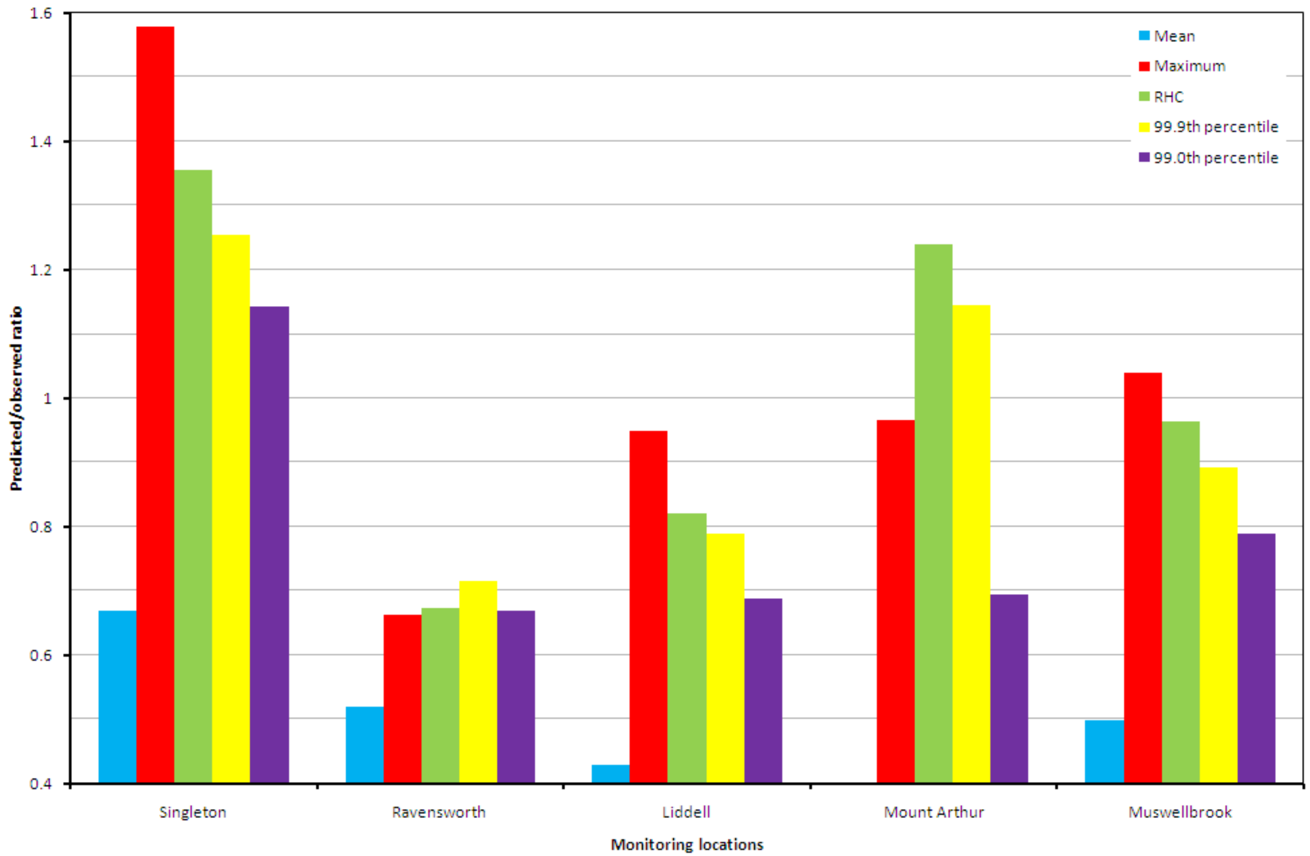

Figure B12	**Predicted/observed ratios for the mean, Robust Highest Concentration, 100th, 99.9th and 99th percentiles at each of the five monitoring locations during the daytime only**		
Location: Five monitoring station	**Period:** 1/07/00 – 30/06/01	**Data source:** Observed vs TAPM predicted	**Units:** N/A
Type: Histogram		**Prepared by:** A. Balch	**Date:** August 2009

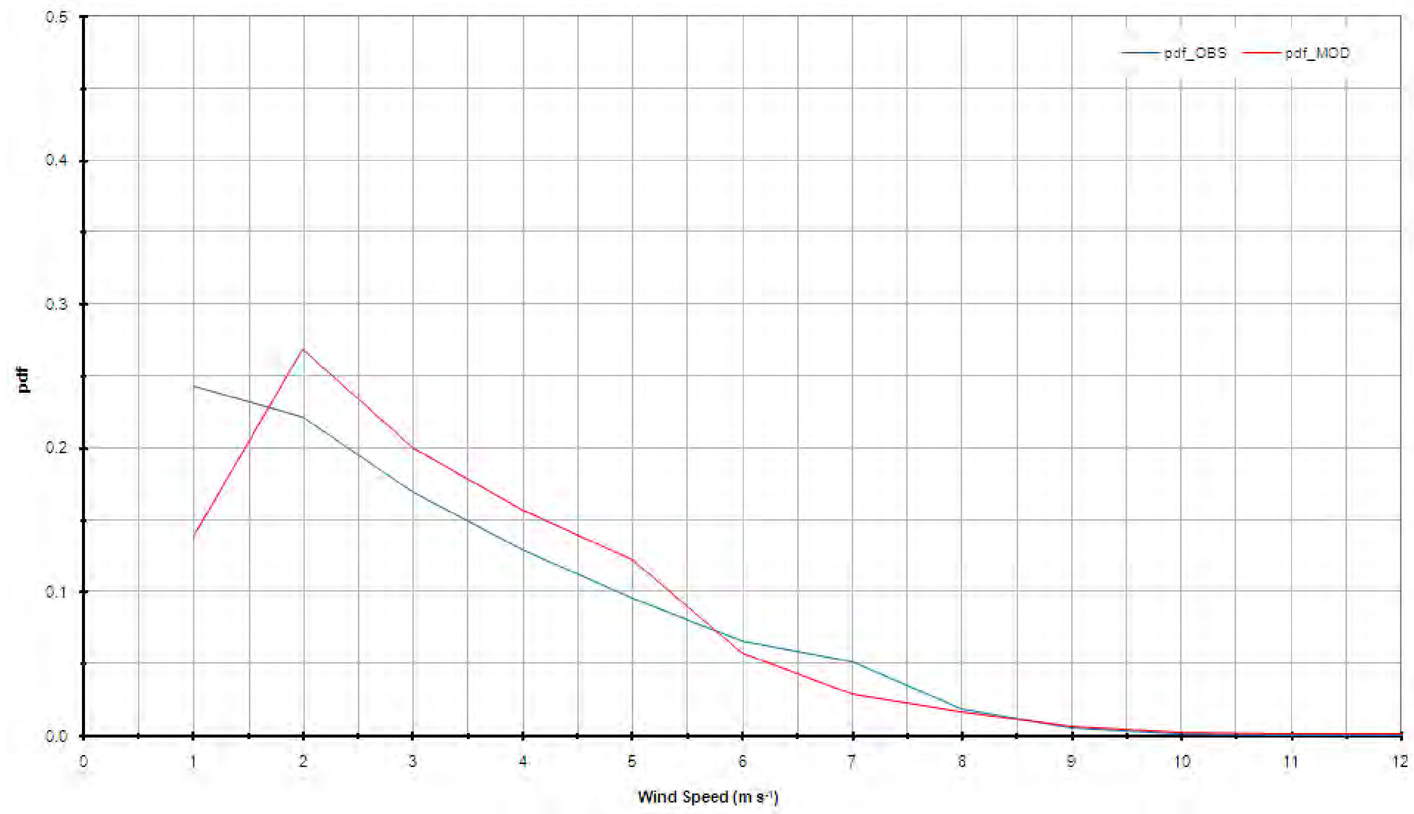

Figure B13 **Comparison of probability density function plots for wind speed at Ravensworth**			
Location: Ravensworth monitoring station	**Period:** 1/07/00 – 30/06/01	**Data source:** Observed vs TAPM predicted	**Units:** Metres per second (m/s)
Type: Probability density function plot		**Prepared by:** A. Balch	**Date:** August 2009

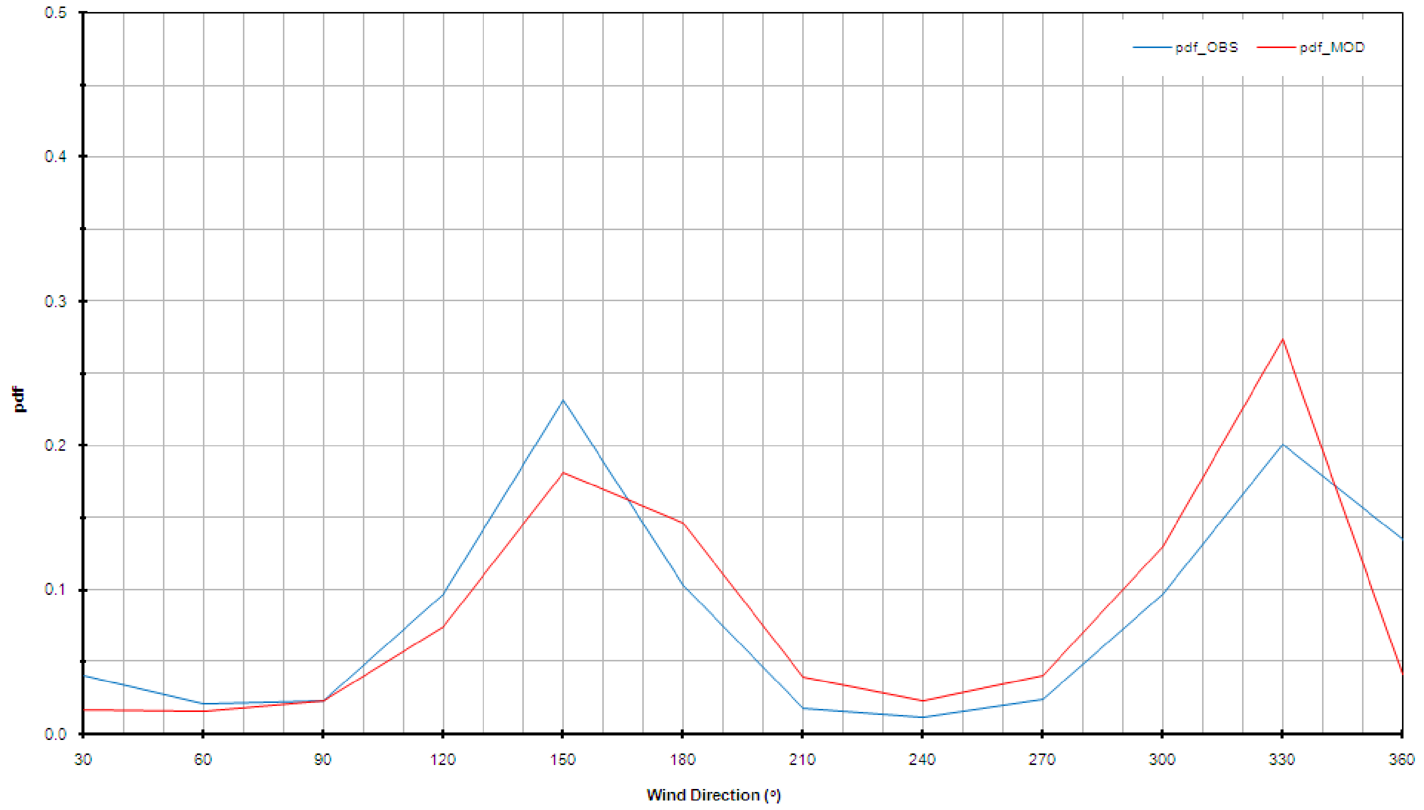

Figure B14 **Comparison of probability density function plots for wind direction at Ravensworth**

Location: Ravensworth monitoring station	**Period:** 1/07/00 – 30/06/01	**Data source:** Observed vs TAPM predicted	**Units:** Degrees (º)
Type: Probability density function plot		**Prepared by:** A. Balch	**Date:** August 2009

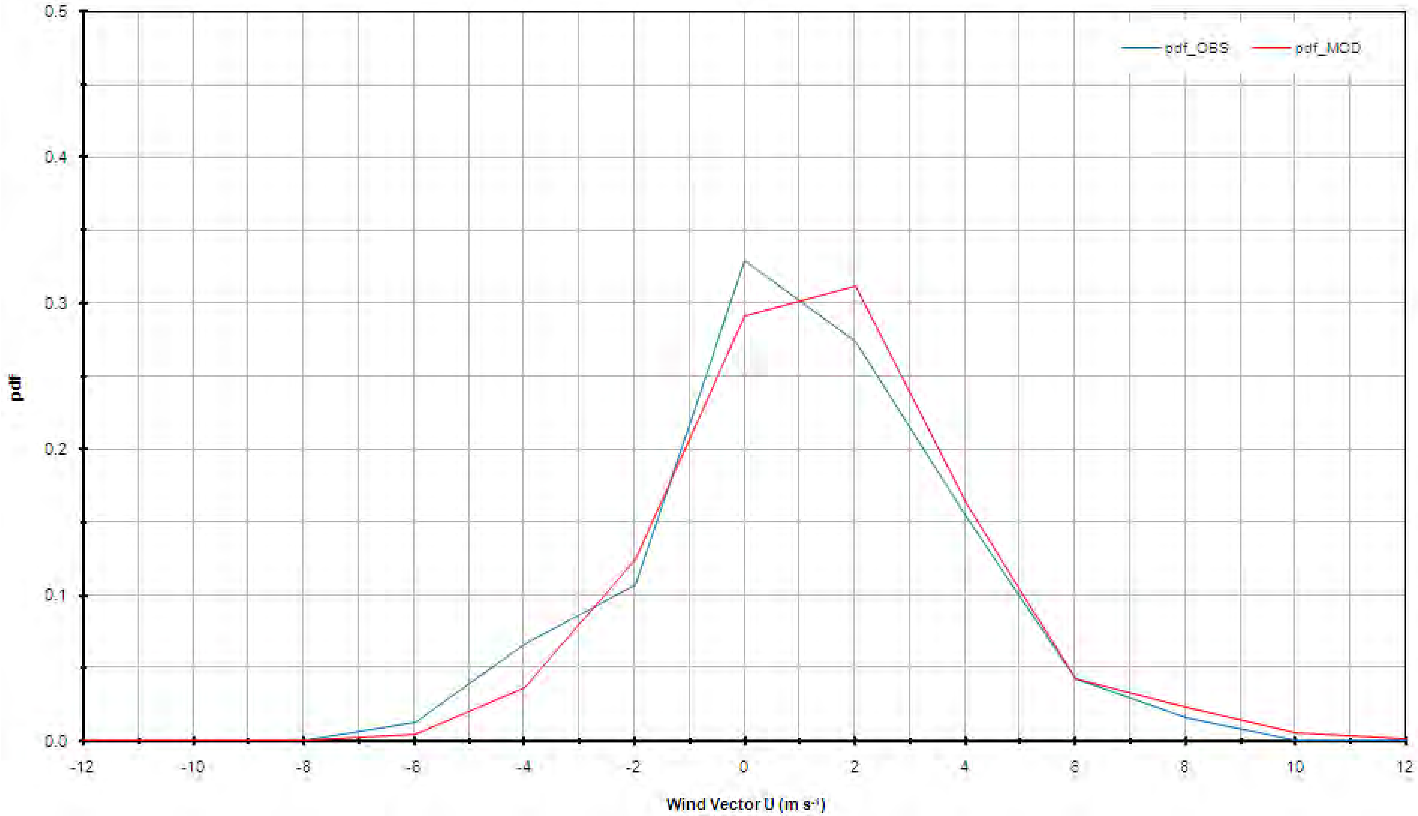

Figure B15	**Comparison of probability density function plots for wind vector U at Ravensworth**		
Location: Ravensworth monitoring station	**Period:** 1/07/00 – 30/06/01	**Data source:** Observed vs TAPM predicted	**Units:** Metres per second (m/s)
Type: Probability density function plot		**Prepared by:** A. Balch	**Date:** August 2009

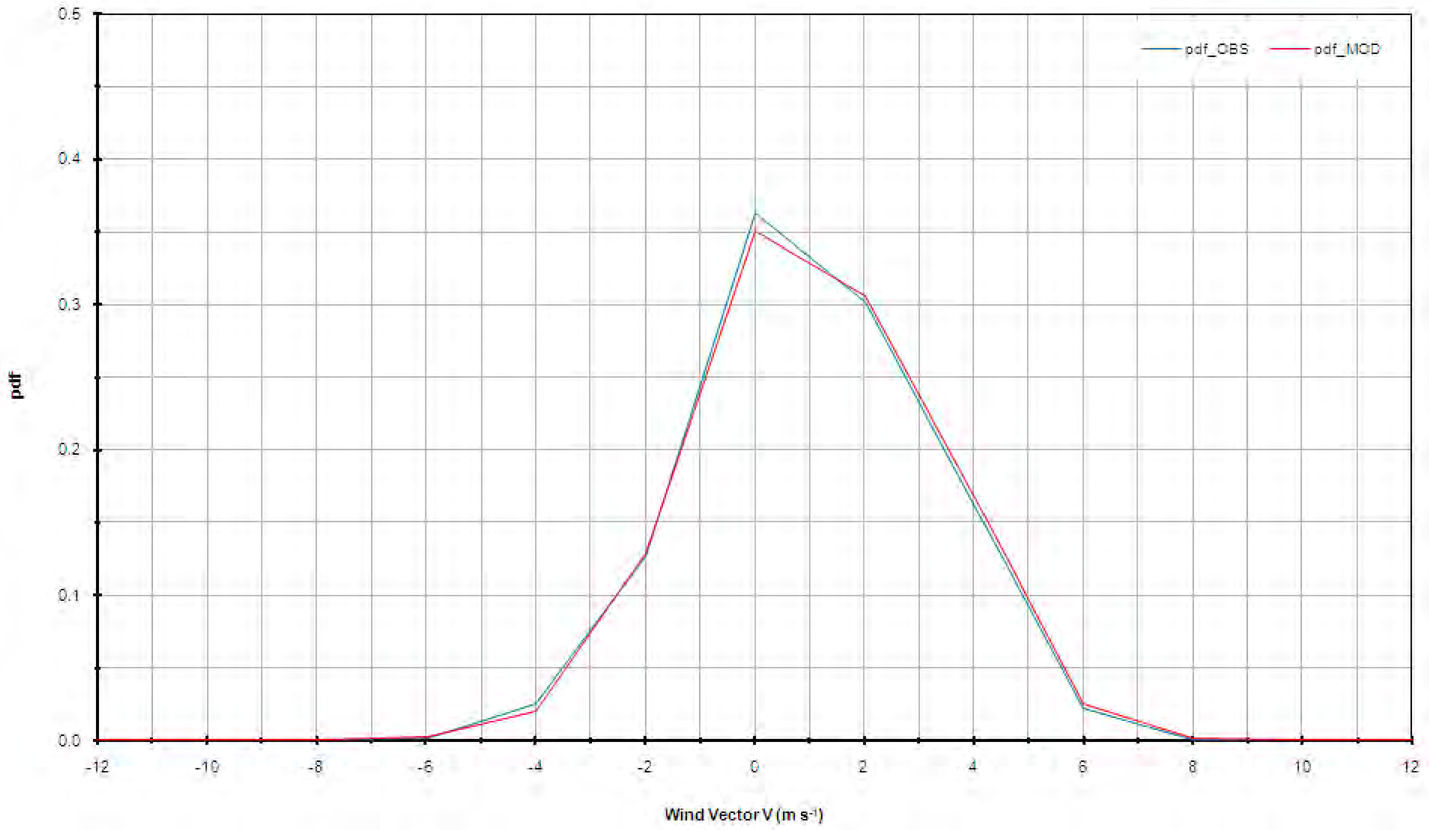

Figure B16 **Comparison of probability density function plots for wind vector V at Ravensworth**			
Location: Ravensworth monitoring station	**Period:** 1/07/00 – 30/06/01	**Data source:** Observed vs TAPM predicted	**Units:** Metres per second (m/s)
Type: Probability density function plot		**Prepared by:** A. Balch	**Date:** August 2009

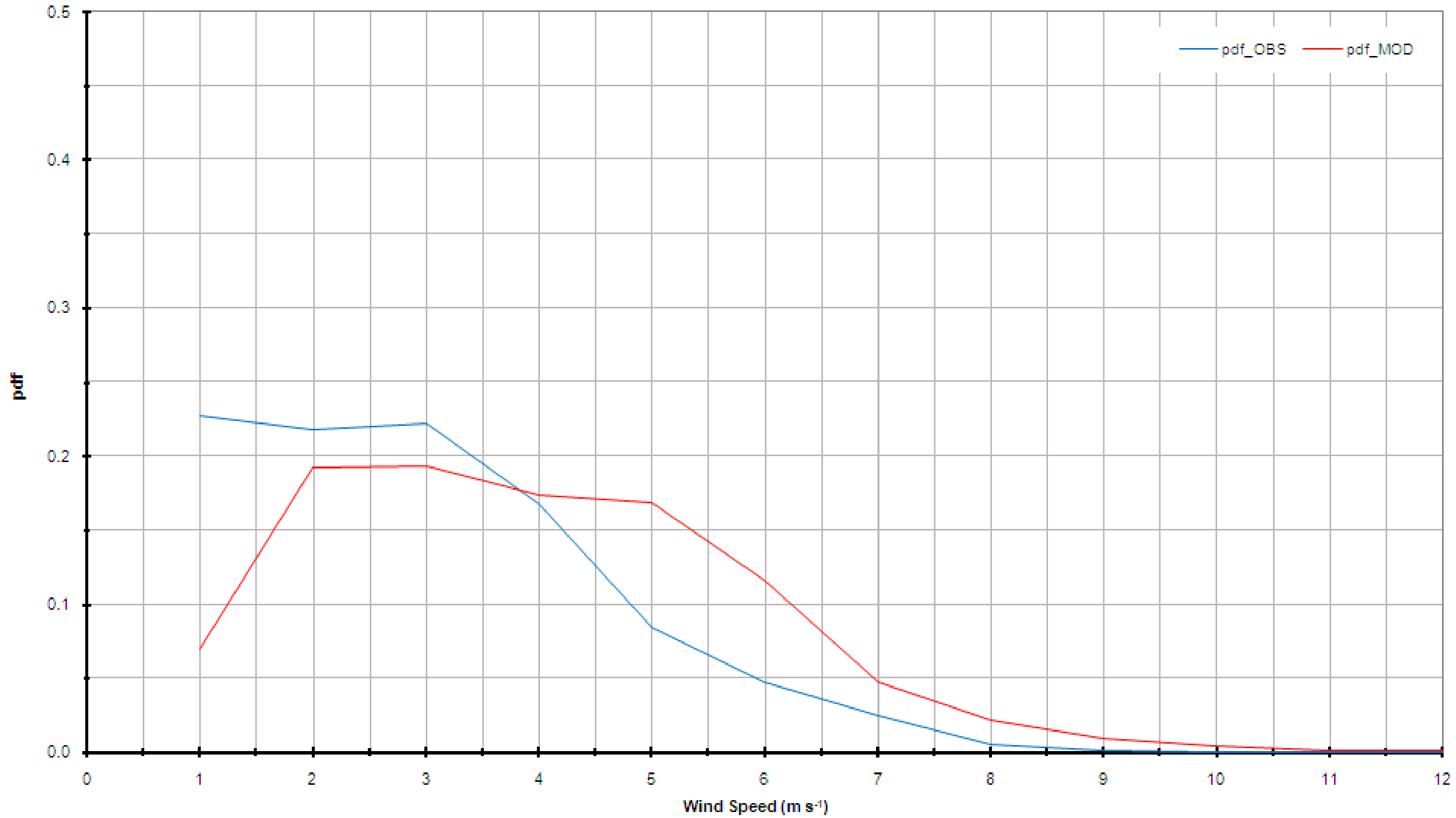

Figure B17 **Comparison of probability density function plots for wind speed at Liddell**			
Location: Liddell monitoring station	**Period:** 1/07/00 – 30/06/01	**Data source:** Observed vs TAPM predicted	**Units:** Metres per second (m/s)
Type: Probability density function plot		**Prepared by:** A. Balch	**Date:** August 2009

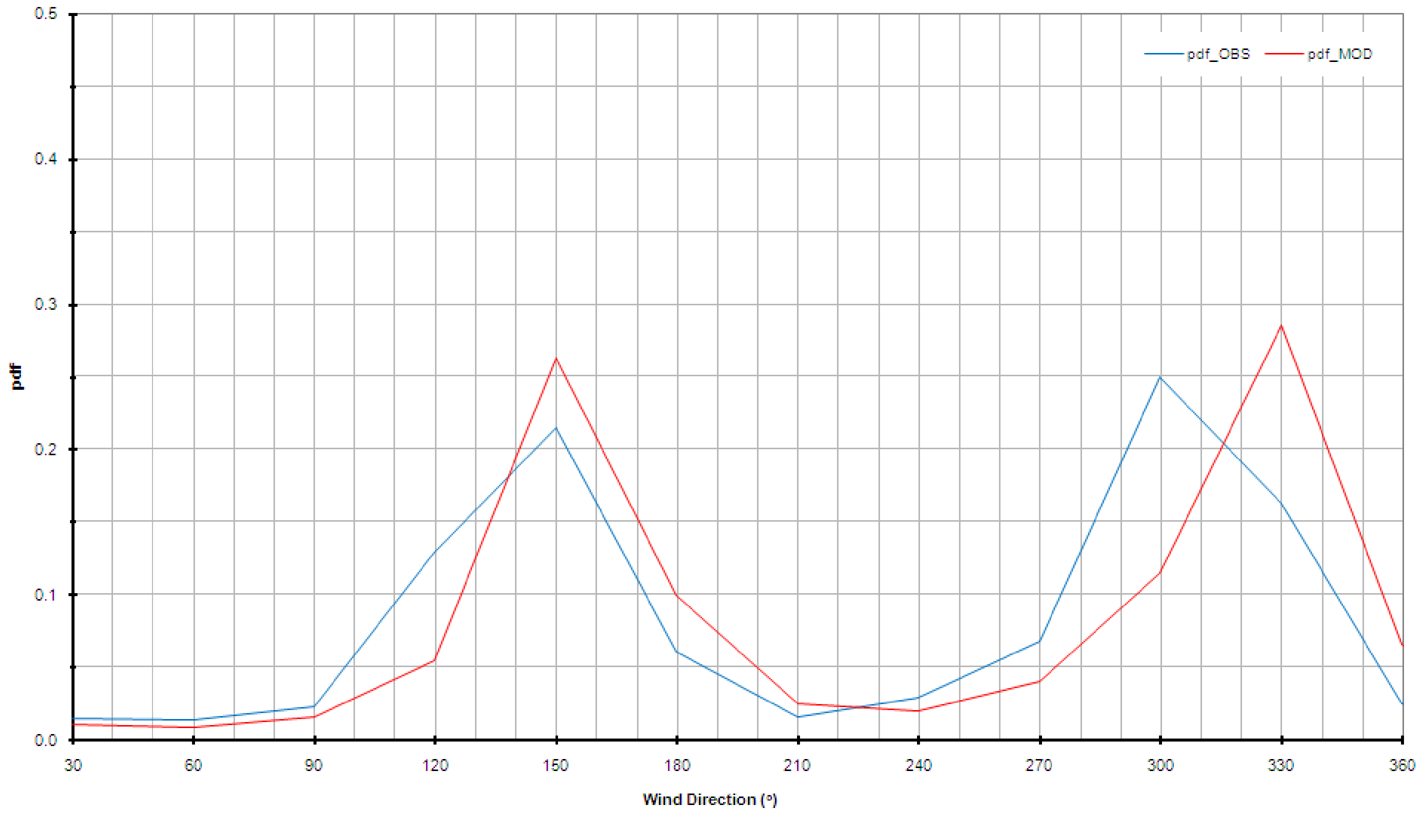

Figure B18 Comparison of probability density function plots for wind direction at Liddell

Location: Liddell monitoring station	**Period:** 1/07/00 – 30/06/01	**Data source:** Observed vs TAPM predicted	**Units:** Degrees (°)
Type: Probability density function plot		**Prepared by:** A. Balch	**Date:** August 2009

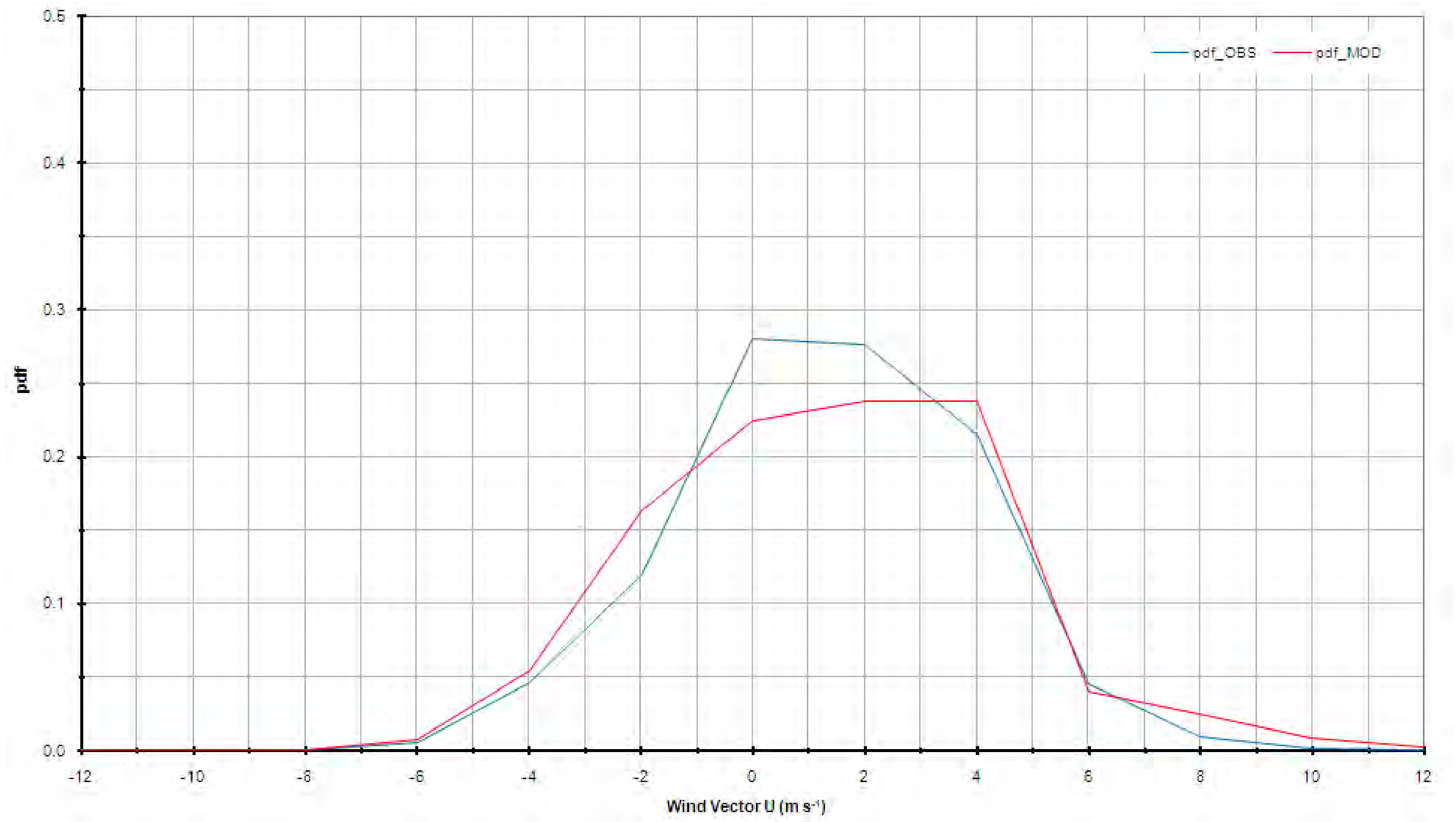

Figure B19	**Comparison of probability density function plots for wind vector U at Liddell**

Location: Liddell monitoring station	**Period:** 1/07/00 – 30/06/01	**Data source:** Observed vs TAPM predicted	**Units:** Metres per second (m/s)
Type: Probability density function plot		**Prepared by:** A. Balch	**Date:** August 2009

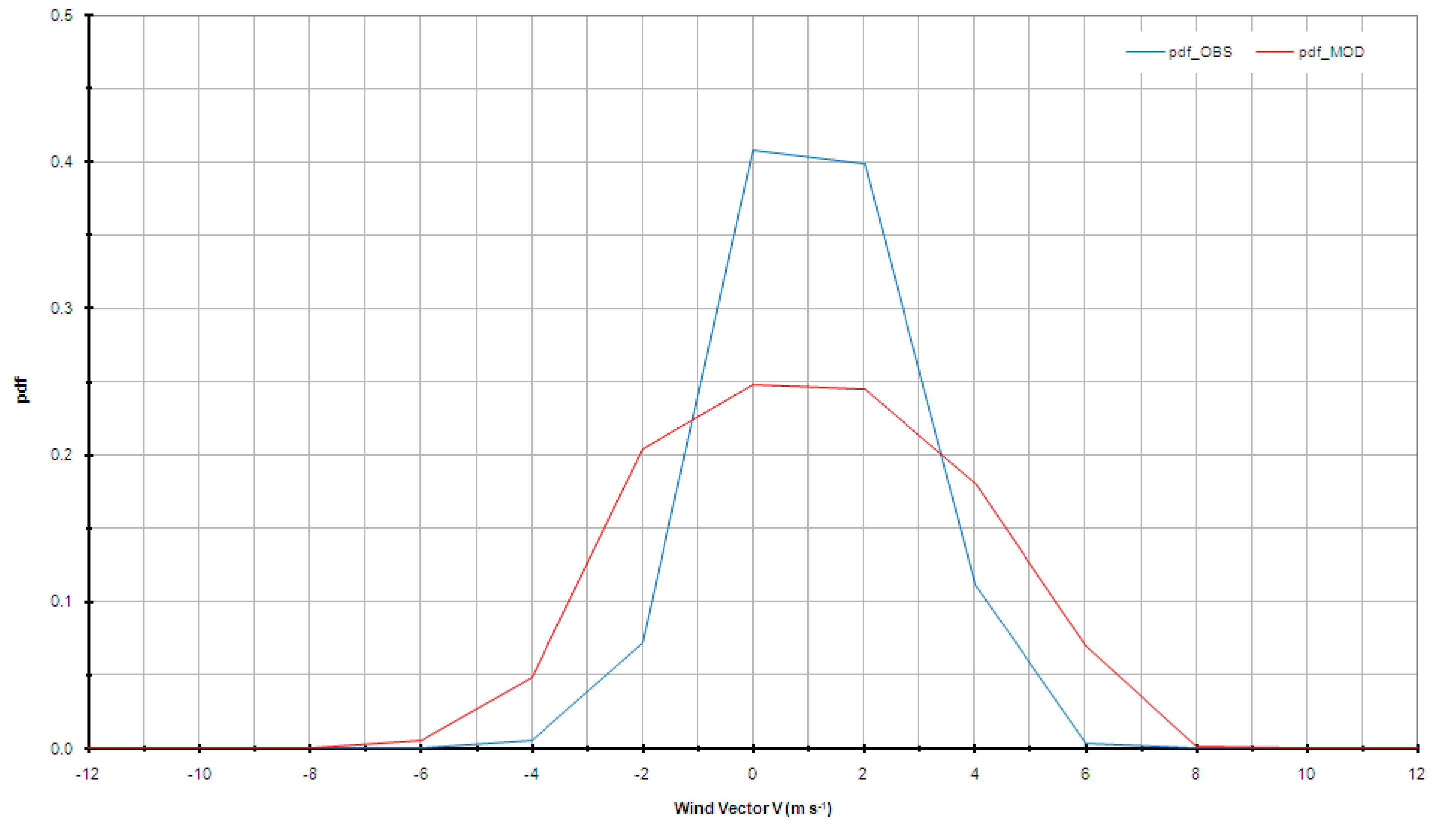

Figure B20	**Comparison of probability density function plots for wind vector V at Liddell**		
Location: Liddell monitoring station	**Period:** 1/07/00 – 30/06/01	**Data source:** Observed vs TAPM predicted	**Units:** Metres per second (m/s)
Type: Probability density function plot		**Prepared by:** A. Balch	**Date:** August 2009

Appendix C

Representative Year Selection

Contents

Tables

Figures

C1 Representative Year Selection

C1.1 Observational Datasets

C1.1.1 Meteorology

Fifteen years of historical monitoring data from 1994 to 2009 at three sites operated by Macquarie Generation were analysed in order to determine a selection of representative meteorological years for dispersion modelling. These sites are

- Mount Arthur North 1994-2007??
- Mitchell Line Rd 2006-2007??
- Lake Liddell 1994-2009
- Ravensworth 1994-2009

The region's predominant wind flows are along the northwest to southeast axis of the valley. There is very little diurnal variation in wind speeds recorded at the Mount Arthur and Liddell sites, while winds at the Ravensworth site are significantly lower during the night compared to the day. Winds from the southeast quadrant dominate summer flows, while winds during winter predominantly flow from the northwest quadrant. Autumn and spring months record a similar frequency of winds from both the northwest and southeast quadrant directions.

The distribution of wind speed and wind direction was determined on an annual basis for all sites. The individual frequencies are then averaged across all years and a climatological baseline is produced. The departure of each year from this baseline can then be quantified and anomalous features extracted. The World Meteorological Organisation (WMO) recommends that when calculating a climatological average, annual datasets with a capture rate less than 80% be omitted from the analysis (WMO 2007). This is due to the fact that 20% of the data is equivalent to an entire season (e.g. summer) and further analysis of these years would adversely skew the results towards a year that is only representative of a particular season or missing an entire seasonal component. Years that have been excluded from the analysis due to the above criteria will be noted for each site. In the selection process those years that contain the highest amount and quality of data for both meteorological variables and pollutant concentrations for all monitoring locations will be given a higher weighting. This will enable a robust and accurate evaluation of the model's performance while ensuring that the each location is well represented in the data and the analysis of potential impacts.

C1.1.2 Pollution

Fifteen years of historical monitoring data from 1994 to 2009 at six sites operated by Macquarie Generation were analysed in order to determine a selection of representative pollution years for dispersion modelling. These sites are

- Mitchell Line Road
- Mount Arthur North
- Lake Liddell
- Ravensworth
- Muswellbrook and
- Singleton

The major source of SO_2 in the region is from the combustion of coal in the Bayswater and Liddell Power Station (refer to Section 7.6). The data indicates the following:

- There have been no exceedances of the annual average impact assessment criteria for SO_2 of 60 µg/m^3 at any monitoring station during the period 1995 to 2009
- There has been one exceedance of the 24-hour average impact assessment criteria for SO_2 of 228 µg/m^3, during the period 1995 to 2009, measured on 12 March 2004 at the Mount Arthur Monitoring Station
- There have been several exceedances of the 1-hour average impact assessment criteria for SO_2 of 570 µg/m^3 during the period 1995 to 2009
- There have been several exceedances of the 10-minute average impact assessment criteria for SO_2 of 700 µg/m^3 during the period 1995 to 2009

There have been no exceedances of the impact assessment criteria for the annual average ground-level concentration of NO_2 of 62 µg/m^3 at any of the monitoring stations during the recording period. The maximum annual average ground-level concentration of NO_2 measured at any of the monitoring stations was 26.7µg/m^3 at Ravensworth (for the March 2006 – February 2007) period, which is 41% of the impact assessment criteria. For the shorter-term 1-hour average, there have been seven exceedances recorded of the impact assessment criteria of 246 µg/m^3. Five of the seven exceedances occurred at the Singleton monitoring site.

C1.2 Methodology

To represent the local and regional variability in meteorological and pollution conditions observed in the Upper Hunter Valley fifteen years of observational data was analysed. The analysis consisted of five stages:

1. Meteorological and pollutant concentration observations such as wind speed, wind direction, SO_2 and NO_2 concentrations (µg/m^3) were converted to frequency space, represented as a probability density function (pdf) for each year in the dataset
2. The average bin frequency was taken as the fifteen year climatological and pollutant concentration baseline distribution, against which each year is assessed for deviations from the average
3. A correlation matrix was designed to determine the degree of departure each year has from the average and between years, where a high correlation (>0.9) shows very little deviation
4. Weighting in the selection process was also given to those years with observed high maximum concentrations and at the largest variety of locations. Where the highest observed year at Muswellbrook may be different from the highest observed year in Singleton, thereby insuring that all sensitive receptor locations are well represented in the selected dispersion modelling scenarios
5. A selection of years representative of the variety of conditions in the region was presented to Macquarie Generation and three representative years were selected

C1.3 Meteorological Results

C1.3.1 Mount Arthur North

The Mount Arthur North monitoring station was missing significant amounts of data (> 20%) for the March-February periods for 1994-1995, 2000-2001, 2002-2003, 2003-2004, 2005-2006 and 2007-2008, and therefore have been excluded from the selection process.

Figure C1 shows the frequency distribution (pdf) of wind speed for all valid years at the Mount Arthur North monitoring site. Mount Arthur North is a relatively high wind site (refer to Section 7.4 for details) due to is higher ground level elevation in relation to Lake Liddell and

Ravensworth. The data shows a distinct increase in the frequency of light winds in 1998-1999 and may be due to equipment failure, as no other year recorded light wind frequencies at that magnitude.

Figure C2 shows the frequency distribution (pdf) of wind direction for all valid years at the Mount Arthur North monitoring site. The variability in wind direction shows a distinct deviation from the average in year 2001-2002 with an increase in the percentage of easterly and westerly winds. The periods 1999-2000 and 2008-2009 saw a slight increase in southerly winds and a decrease in the north westerly component (refer to Section 7.4).

Table C1 shows the correlation matrix of wind speed for all valid years at the Mount Arthur North monitoring station. There is significant deviation from the average and the rest of the years during 1998-1999. All other years show good correlations between years and with the average.

Table C2 shows the correlation matrix of wind direction for all valid years at the Mount Arthur North monitoring station. There is significant deviation from the average and the rest of the years during 1998-1999. All other years show good correlations between years and with the average.

The extreme deviation of the 1998-1999 monitoring year from the climatological average, and the strong correlation of the remaining years between each other, indicates that 1998-1999 is an anomalous year. However, the Mount Arthur North monitoring station has had several issues in regards to data quality and management and has since been relocated to Mitchell Line Road as of 2006. As such, this anomalous year should be viewed with caution and not regarded as a real meteorological phenomenon.

Table C1 Correlation matrix of measured wind speed distributions for all valid years at the Mount Arthur North monitoring site

	1995_1996	1996_1997	1997_1998	1998_1999	1999_2000	2001_2002	2006_2007	2008_2009	AVERAGE
1995_1996	1.00								
1996_1997	0.98	1.00							
1997_1998	0.94	0.95	1.00						
1998_1999	0.25	0.31	0.55	1.00					
1999_2000	0.91	0.94	0.99	0.53	1.00				
2001_2002	0.96	0.97	0.98	0.40	0.98	1.00			
2006_2007	0.98	0.98	0.95	0.28	0.95	0.98	1.00		
2008_2009	0.70	0.78	0.86	0.63	0.90	0.85	0.77	1.00	
AVERAGE	0.89	0.92	0.99	0.63	0.99	0.96	0.91	0.91	1.00

Table C2 Correlation matrix of measured wind direction distributions for all valid years at the Mount Arthur North monitoring site

	1995_1996	1996_1997	1997_1998	1998_1999	1999_2000	2001_2002	2006_2007	2008_2009	AVERAGE
1995_1996	1								
1996_1997	0.92	1.00							
1997_1998	0.93	0.95	1.00						
1998_1999	0.88	0.81	0.93	1.00					
1999_2000	0.82	0.77	0.91	0.99	1.00				
2001_2002	0.79	0.69	0.62	0.57	0.46	1.00			
2006_2007	0.81	0.82	0.94	0.95	0.97	0.41	1.00		
2008_2009	0.59	0.82	0.77	0.55	0.55	0.29	0.72	1.00	
AVERAGE	0.93	0.95	0.99	0.94	0.92	0.64	0.95	0.77	1.00

C1.3.2 Lake Liddell

The Lake Liddell monitoring station was missing significant amounts of data (> 20%) for the year 1994-1995, 2001-2002, 2002-2003 and 2003-2004 and therefore have been excluded from the selection process.

Figure C3 shows the frequency distribution (pdf) of wind speed for all valid years at the Lake Liddell monitoring monitoring station. The data shows very little variability in wind speed between years, and when compared to the climatological average.

Figure C4 shows the frequency distribution (pdf) of wind direction for all valid years at the Lake Liddell monitoring station. The variability in wind direction shows a distinct deviation from the average in the years 2004-2005 and 2008-2009, with a decrease in the frequency of south easterly winds and an increase in north westerly winds. All other years are relatively close the climatological average.

Table C3 shows the correlation matrix of wind speed for all valid years at the Lake Liddell monitoring station. There is no significant deviation from the climatological average in any single year, nor is there any significant deviation between years.

Table C4 shows the correlation matrix of wind direction for all valid years at the Lake Liddell monitoring station. There is no significant deviation from the climatological average in any single year, nor is there any significant deviation between the years. Although 2004-2005 and 2008-2009 appear to show some departure from the average, they still scored relatively high correlation coefficients, 0.89 and 0.84 respectively, indicating that they are not anomalous years.

Table C3 Correlation matrix of measured wind speed distributions for all valid years at the Lake Liddell monitoring site

	1995 - 1996	1996 - 1997	1997-1998	1998-1999	1999-2000	2000-2001	2004-2005	2005-2006	2006-2007	2007-2008	2008-2009	AVERAGE
1995_1996	1.00											
1996_1997	0.98	1.00										
1997_1998	0.95	0.98	1.00									
1998_1999	0.97	0.99	0.99	1.00								
1999_2000	0.91	0.95	0.99	0.98	1.00							
2000_2001	0.87	0.94	0.97	0.96	0.98	1.00						
2004_2005	0.91	0.94	0.99	0.98	0.99	0.97	1.00					
2005_2006	0.99	0.99	0.98	0.99	0.95	0.91	0.95	1.00				
2006_2007	0.94	0.97	0.99	0.99	0.99	0.97	0.99	0.98	1.00			
2007_2008	0.89	0.96	0.96	0.96	0.96	0.99	0.94	0.92	0.96	1.00		
2008_2009	0.92	0.97	0.99	0.99	0.99	0.99	0.99	0.96	0.99	0.97	1.00	
AVERAGE	0.95	0.98	0.99	0.99	0.99	0.98	0.99	0.98	0.99	0.97	0.99	1.00

Table C4 Correlation matrix of measured wind direction distributions for all valid years at the Lake Liddell monitoring site

	1995 - 1996	1996 - 1997	1997-1998	1998-1999	1999-2000	2000-2001	2004-2005	2005-2006	2006-2007	2007-2008	2008-2009	AVERAGE
1995_1996	1											
1996_1997	0.99	1										
1997_1998	0.86	0.84	1									
1998_1999	0.82	0.8	0.99	1								
1999_2000	0.79	0.78	0.96	0.99	1							
2000_2001	0.84	0.81	0.99	0.98	0.96	1						
2004_2005	0.92	0.9	0.85	0.77	0.71	0.84	1					
2005_2006	0.83	0.82	0.92	0.95	0.96	0.93	0.72	1				
2006_2007	0.94	0.92	0.97	0.96	0.95	0.96	0.86	0.96	1			
2007_2008	0.95	0.93	0.96	0.95	0.94	0.95	0.87	0.96	0.99	1		
2008_2009	0.94	0.94	0.69	0.64	0.64	0.66	0.8	0.75	0.83	0.84	1	
AVERAGE	0.95	0.94	0.97	0.95	0.93	0.96	0.89	0.94	0.99	0.99	0.84	1

C1.3.3 Ravensworth

The Ravensworth monitoring station was missing significant amounts of data (> 20%) for the years 1994-1995, 2001- 2002, 2002-2003, 2004-2005, 2006-2007, 2007-2008 and 2008-2009 and, consequently, has been excluded from the selection process.

Figure C5 shows the frequency distribution (pdf) of wind speed for all valid years at the Ravensworth monitoring site. The data shows very little variability in wind speed between years, and when compared to the climatological average.

Figure C6 shows the frequency distribution (pdf) of wind direction for all valid years at the Ravensworth monitoring station. There is no significant deviation from the climatological average in any single year, nor is there any significant deviation between the years.

Table C5 shows the correlation matrix of wind speed for all valid years at the Ravensworth monitoring station. There is no significant deviation from the climatological average in any single year, nor is there any significant deviation between years.

Table C6 shows the correlation matrix of wind direction for all valid years at the Ravensworth monitoring site. There is no significant deviation from the climatological average in any single year, nor is there any significant deviation between years.

Table C5 Correlation matrix of measured wind speed distributions for all valid years at the Ravensworth monitoring site

	1995_1996	*1996_1997*	*1997_1998*	*1998_1999*	*1999_2000*	*2000_2001*	*2002_2003*	*2005_2006*	*AVERAGE*
1995_1996	1.00								
1996_1997	0.99	1.00							
1997_1998	0.99	0.98	1.00						
1998_1999	0.99	0.99	0.99	1.00					
1999_2000	0.99	0.98	0.99	0.99	1.00				
2000_2001	0.99	0.99	0.99	0.99	0.99	1.00			
2002_2003	0.95	0.96	0.98	0.96	0.95	0.97	1.00		
2005_2006	0.99	0.99	0.99	0.99	0.99	0.99	0.96	1.00	
AVERAGE	0.99	0.99	0.99	0.99	0.99	0.99	0.97	0.99	1.00

Table C6 Correlation matrix of measured wind direction distributions for all valid years at the Ravensworth monitoring site

	1995_1996	*1996_1997*	*1997_1998*	*1998_1999*	*1999_2000*	*2000_2001*	*2002_2003*	*2005_2006*	*AVERAGE*
1995_1996	1.00								
1996_1997	0.99	1.00							
1997_1998	0.97	0.95	1.00						
1998_1999	0.97	0.97	0.97	1.00					
1999_2000	0.90	0.87	0.96	0.96	1.00				
2000_2001	0.96	0.95	0.99	0.99	0.97	1.00			
2002_2003	0.91	0.88	0.97	0.95	0.99	0.98	1.00		
2005_2006	0.96	0.97	0.93	0.93	0.83	0.93	0.87	1.00	
AVERAGE	0.98	0.97	0.99	0.99	0.96	1.00	0.96	0.95	1.00

C1.3.4 Inter-site Variability

The calculated climatological average for each site has been compared, following the methods described above. Figure C7 shows the frequency distribution of the average wind speeds for Mount Arthur North, Lake Liddell and Ravensworth. Mount Arthur North stands out as a high wind site due mainly to its elevated position, while Lake Liddell and Ravensworth are very similar in their distribution. Table C7 shows the correlation coefficients for the three sites. Mount Arthur North shows no significant correlation with Lake Liddell or Ravensworth (0.59 and 0.5 respectively). Lake Liddell and Ravensworth on the otherhand show a very high correlation of 0.92.

Table C7 Correlation matrix of the climatological average wind speed distributions for Lake Liddell, Ravensworth and Mount Arthur North

Wind speed	Lake Liddell	Ravensworth	Mount Arthur North
Lake Liddell	1.00		
Ravensworth	0.92	1.00	
Mount Arthur North	0.59	0.50	1.00

Figure C8 shows the frequency distribution of the average wind direction for Mount Arthur North, Lake Liddell and Ravensworth. There appears to be no significant variability in wind direction between the three sites. Mount Arthur North does show a higher frequency of south westerly winds, however all sites show a relatively high correlation (Table C8).

Table C8 Correlation matrix of the climatological average wind direction distributions for Lake Liddell, Ravensworth and Mount Arthur North

Wind direction	Lake Lidell	Ravensworth	Mount Arthur North
Lake Liddell	1.00		
Ravensworth	0.85	1.00	
Mount Arthur North	0.78	0.87	1.00

C1.4 Pollution Results

C1.4.1 Lake Liddell

Significant amounts of data (> 20%) were missing from the years 1994-1995, 1998-1999, 2005-2006 and 2006-2007 for observations of SO_2 and/or NO_2. These years have been excluded from further analysis and are not valid in the selection process.

Figure C9 shows the frequency distribution of SO_2 concentrations at the Lake Liddell monitoring station. The lower end of the distribution (< 200 µg/m^3) has been truncated from the analysis in order to determine high impact years. The distribution of SO_2 measurements shows that hourly concentrations can exceed the impact assessment criteria of 570 µg/m^3 for a one hour average in any given year at the Lake Liddell monitoring station.

Figure C10 shows the frequency distribution of NO_2 concentrations at the Lake Liddell monitoring station. The lower end of the distribution (< 100 µg/m^3) has been truncated from the analysis for the same reasons listed above. NO_2 concentrations are seen to remain well below the impact assessment criteria of 246 µg/m^3 for a one hour average in all valid years.

Correlations of concentration distributions is a not a useful indicator of a representative pollution year due to the fact that it is the high end of the distribution that is of interest and identifying those years that have the highest potential to cause elevated ground-level concentrations at downwind receptors. Therefore the maximum one hour concentration from each year was extracted from the dataset and compared with each other year (Figure C11).

C1.4.2 Ravensworth

Significant amounts of data (> 20%) were missing from the years 1994-1995, 1997-1998, 2004-2005 and 2008-2009 for observations of SO_2 and/or NO_2. These years have been excluded from further analysis and are not valid in the selection process.

Figure C12 shows the frequency distribution of SO_2 concentrations at the Ravensworth monitoring station. The lower end of the distribution (< 200 µg/m^3) has been truncated from the analysis in order to determine high impact years. The distribution of SO_2 measurements shows that hourly concentrations can exceed the impact assessment criteria of 570 µg/m^3 for a one hour average in any given year at the Ravensworth monitoring station.

Figure C13 shows the frequency distribution of NO_2 concentrations at the Ravensworth monitoring station. The lower end of the distribution (< 100 µg/m^3) has been truncated from the analysis for the same reasons listed above. NO_2 concentrations are seen to remain well below the impact assessment criteria of 246 µg/m^3 for a one hour average in all valid years.

As described in Section 1.4.1, the maximum one hour concentration from each year was extracted from the dataset and compared with each other year Figure C14.

C1.4.3 Muswellbrook

A significant amount of data (> 20%) was missing from the years 1994-1995, 1997-1998. These years have been excluded from further analysis and are not valid in the selection process.

Figure C15 shows the frequency distribution of SO_2 concentrations at the Muswellbrook monitoring station. The lower end of the distribution (< 200 µg/m^3) has been truncated from the analysis in order to determine high impact years. The distribution of SO_2 measurements shows that hourly concentrations can exceed the impact assessment criteria of 570 µg/m^3 for a one hour average in any given year at the Muswellbrook site.

Figure C16 shows the frequency distribution of NO_2 concentrations at the Muswellbrook monitoring site. The lower end of the distribution (< 100 µg/m^3) has been truncated from the analysis for the same reasons described above. NO_2 concentrations are seen to remain well below the impact assessment criteria of 246 µg/m^3 for a one hour average in all valid years except 2001-2002 which measured a maximum of 246 µg/m^3.

Figure C17 shows the maximum one hour concentration for each valid year. The years 1996-1997, 1999-2000 and 2001-2002 are shown to have maximum concentrations nearly twice as high as the other years in the dataset.

C1.4.4 Singleton

A significant amount of data (> 25%) was missing from the years 1994-1995, 2005-2006, 2006-2007, 2007-2008 and 2008-2009. These years have been excluded from further analysis and are not valid in the selection process.

Figure C18 shows the frequency distribution of SO_2 concentrations at the Singleton monitoring station. The lower end of the distribution (< 200 µg/m^3) has been truncated from the analysis in order to determine high impact years. The distribution of SO_2 measurements shows that hourly concentrations does not exceed the impact assessment criteria of 570 µg/m^3 for a one hour average in any given year at the Singleton site except 1998-1999 where a maximum 1 hour average of 820 µg/m^3 was recorded.

Figure C19 shows the frequency distribution of NO_2 concentrations at the Singleton monitoring station. The lower end of the distribution (< 100 µg/m^3) has been truncated from the analysis for the same reasons described above. NO_2 concentrations are seen to remain well below the impact assessment criteria of 246 µg/m^3 for a one hour average in all valid years. The years 2002-2003 and 2003-2004 show a slight increase in the frequency one hour averages between 100 µg/m^3 and 150 µg/m^3 in comparison with the other years.

Figure C20 shows the maximum one hour concentration for each valid year. The years 1997-1998, 1998-1999, 2002-2003 and 2003-2004 are shown to have maximum concentrations slightly higher than the other years in the dataset.

C1.4.5 Mount Arthur North

A significant amount of data (> 20%) was missing from the years 1994-1995, 2005-2006, 2006-2007, 2007-2008 and 2008-2009. These years have been excluded from further analysis and are not valid in the selection process.

Figure C21 shows the frequency distribution of SO_2 concentrations at the Mount Arthur North monitoring station. The lower end of the distribution (< 200 µg/m^3) has been truncated from the analysis in order to determine high impact years. The distribution of SO_2 measurements shows that there is an exceedance of the 1-hour average impact assessment criteria for SO_2 of 570 µg/m^3 during the years 1996-1997, 1999-2000, 2001-2002 and 2002-2003.

Figure C22 shows the frequency distribution of NO_2 concentrations at the Mount Arthur North monitoring station. The lower end of the distribution (< 100 µg/m^3) has been truncated from the analysis for the same reasons described above. NO_2 concentrations are seen to remain well below the impact assessment criteria of 246 µg/m^3 for a one hour average in all valid years. The years 2002-2003 shows a slight increase in the frequency of the one hour averages of NO_2 between 100 µg/m^3 and 175 µg/m^3 in comparison with the other years.

Figure C23 shows the maximum one hour concentration for each valid year. The years 1997-1998, 1998-1999, 2001-2002 and 2002-2003 are shown to have maximum concentrations slightly higher than the other years in the dataset.

C1.4.6 Inter-site Variability

The five pollution monitoring sites all show distinct features in their dataset, representing the localised influences each location experiences in regards to the dispersion of pollutants in the atmosphere. In selecting several representative years of SO_2 and NO_2 observations, it was necessary to balance the need to be conservative in respect to sensitive locations and representative of the dataset as a whole. Figure C24 shows the maximum 1-hour average ground-level concentration of SO2 at each monitoring location for valid years only. The data shows that the maximum 1-hour average ground-level concentration does not occur in the same year at each site. For example the year 1999-2000 shows a large range of concentrations; from 276 µg/m3 at Singleton to 1311 µg/m3 at Lake Liddell, while 2000-2001 shows that the maximum concentrations across all sites are relatively close to one another. Maximum 1-hour average ground-level concentrations of NO_2 display a similar tendency as SO_2 that is to be site specific from year to year (Figure C25).

C1.5 Selection of Representative Years

Three years were selected based on the analysis presented above. The selected representative years are:

- March 1 1999 to February 28 2000
- March 1 2000 to February 28 2001
- March 1 2007 to February 29 2008

Figure C24 and Figure C25 show the selected years in boxes. These years were selected as being the most representative of the range of likely impacts to be experienced at sensitive locations while maintaining a conservative element to the assessment and enabling a robust evaluation of the dispersion model's performance. Section C1.3 showed that the inter-annual variability of wind speed and wind direction does not vary significantly from year to year or site to site. This indicates that other meteorological variables, such as the exchanges of surface energy fluxes, boundary layer development or the formation of nocturnal jets are more important to the dispersion of pollutants in the Upper Hunter Valley.

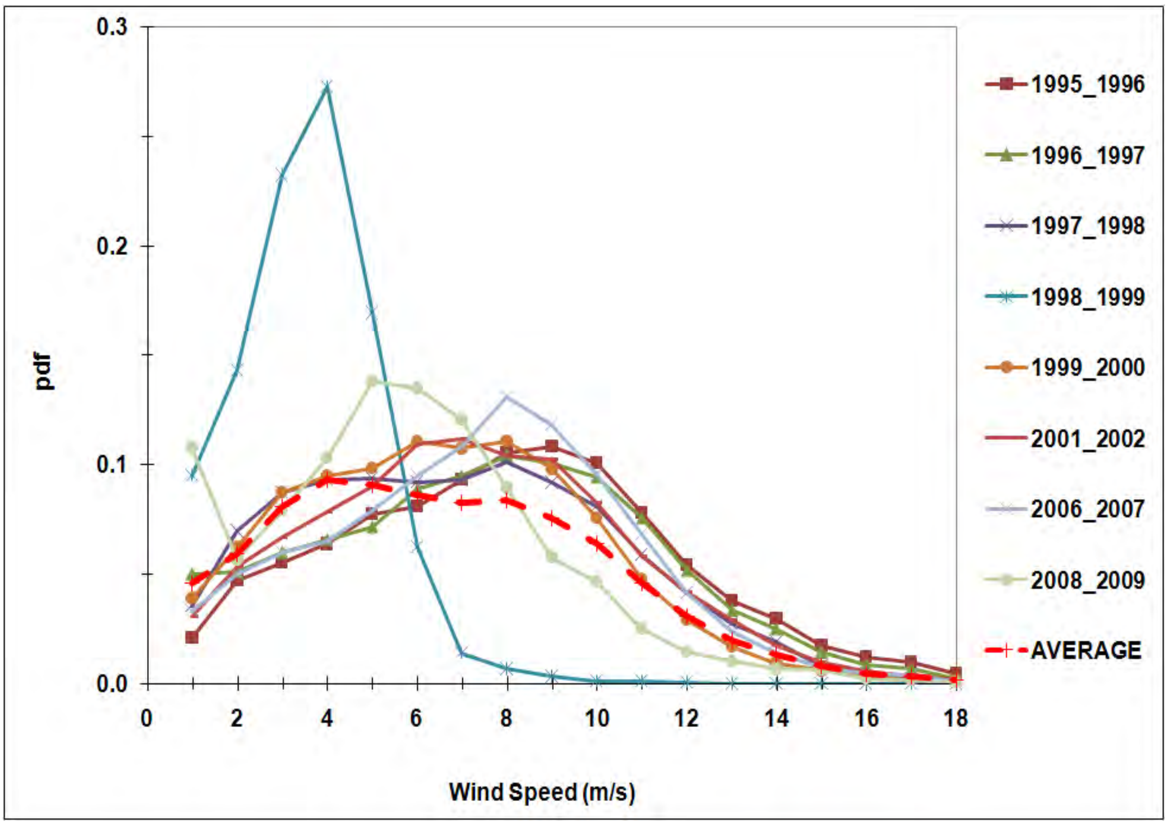

Figure C1	**Frequency distribution (pdf) of wind speed in 1 metre per second (m/s) bins for all valid years at the Mount Arthur North monitoring site**		
Location: Mount Arthur North monitoring site	**Period:** March 1994 to February 2009 (valid years only)	**Data source:** Macquarie Generation	**Units:** m/s
Type: pdf	**Averaging period:** 1-hour	**Prepared by:** Andrew Wiebe	**Date:** August 2009

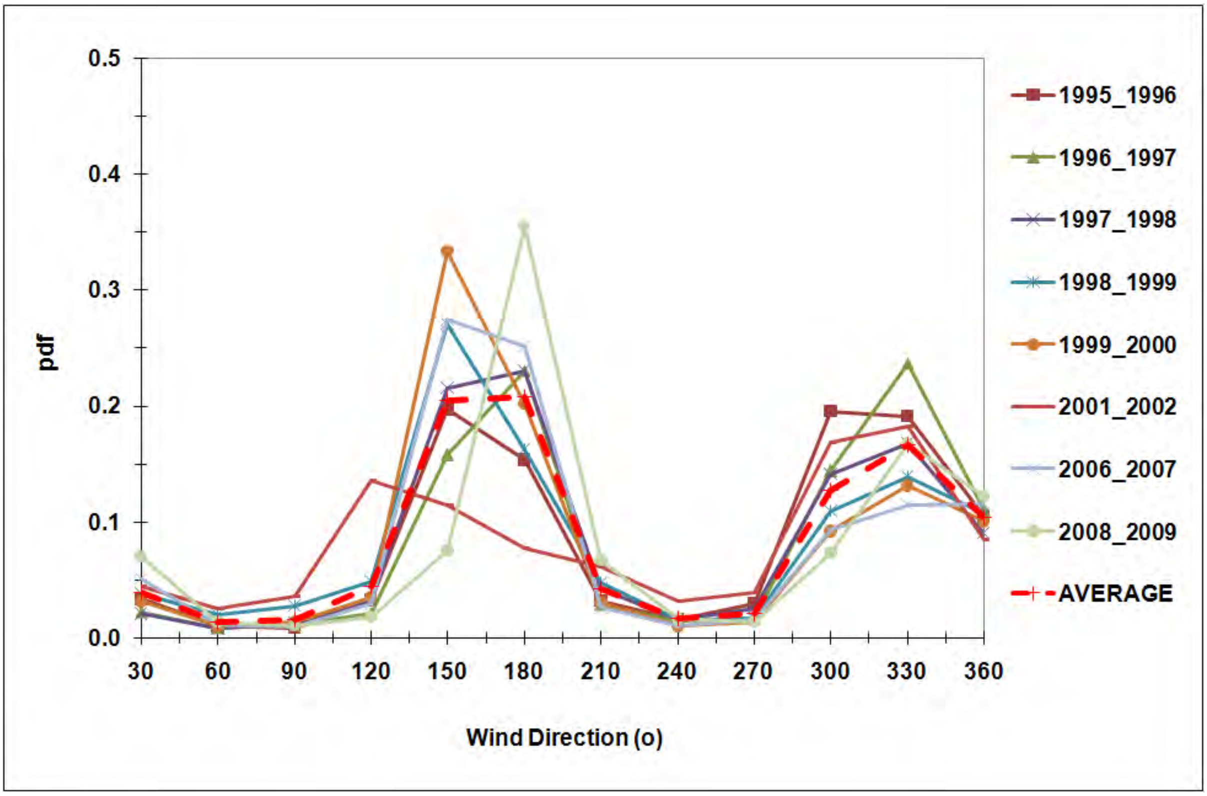

Figure C2 Frequency distribution (pdf) of wind direction in 30 degree (°) bins for all valid years at the Mount Arthur North monitoring site

Location: Mount Arthur North monitoring site	**Period:** March 1994 to February 2009 (excluding March 2003 to February 2005)	**Data source:** Macquarie Generation	**Units:** degrees
Type: pdf	**Averaging period:** 1-hour	**Prepared by:** Andrew Wiebe	**Date:** August 2009

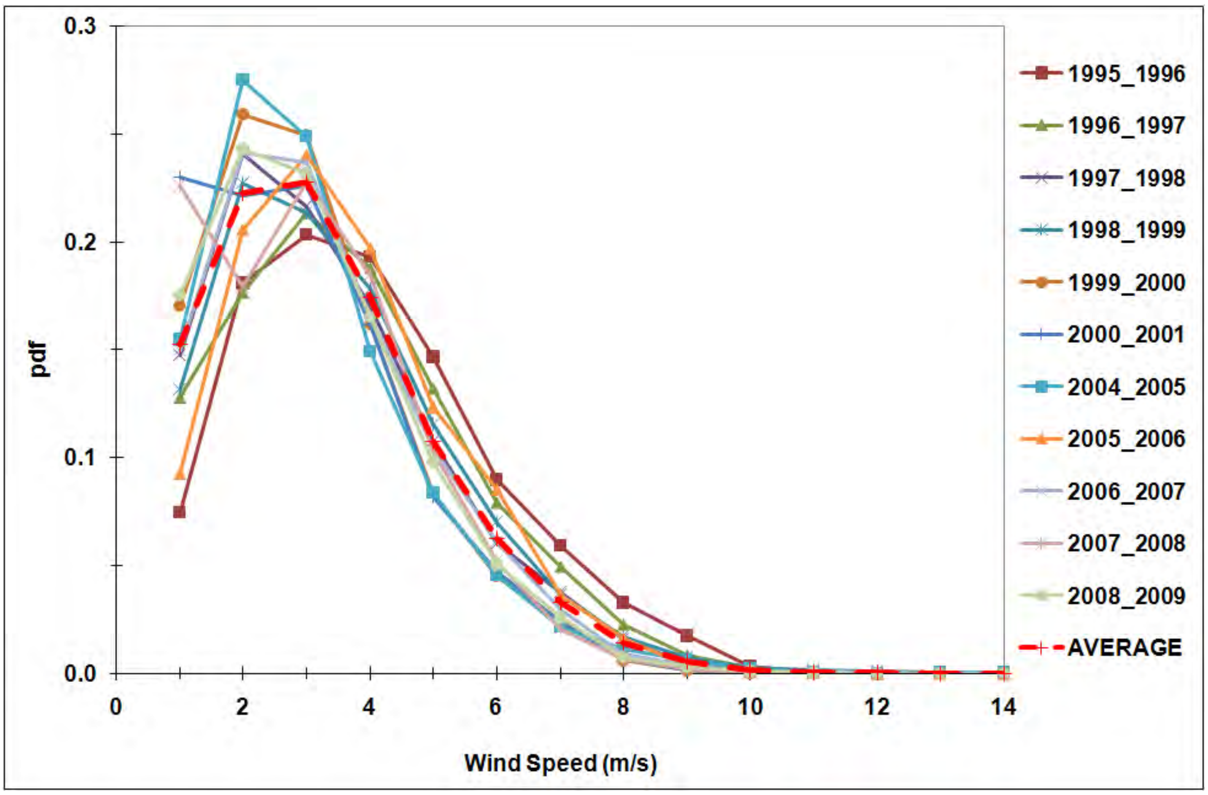

Figure C3 Frequency distribution (pdf) of wind speed in 1 metre per second (m/s) bins for all valid years at the Lake Liddell monitoring site			
Location: Lake Liddell monitoring site	**Period:** March 1995 to February 2009 (excluding March 1994 to February 1995 and March 2001 to February 2004)	**Data source:** Macquarie Generation	**Units:** m/s
Type: pdf	**Averaging period:** 1-hour	**Prepared by:** Andrew Wiebe	**Date:** August 2009

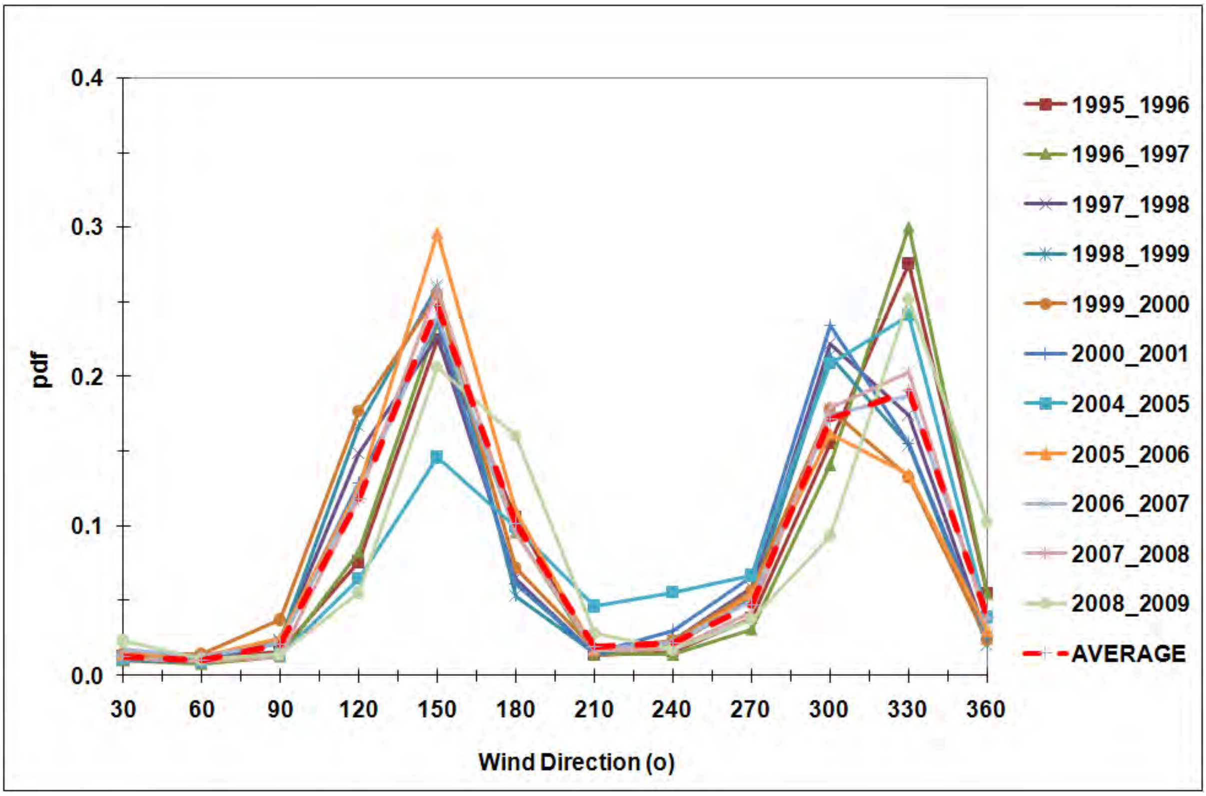

Figure C4 Frequency distribution (pdf) of wind direction in 30 degree (°) bins for all valid years at the Lake Liddell monitoring site

Location: Lake Liddell monitoring site	**Period:** March 1994 to February 2009 (excluding March 1994 to February 1995 and March 2001 to February 2004)	**Data source:** Macquarie Generation	**Units:** degrees
Type: pdf	**Averaging period:** 1-hour	**Prepared by:** Andrew Wiebe	**Date:** August 2009

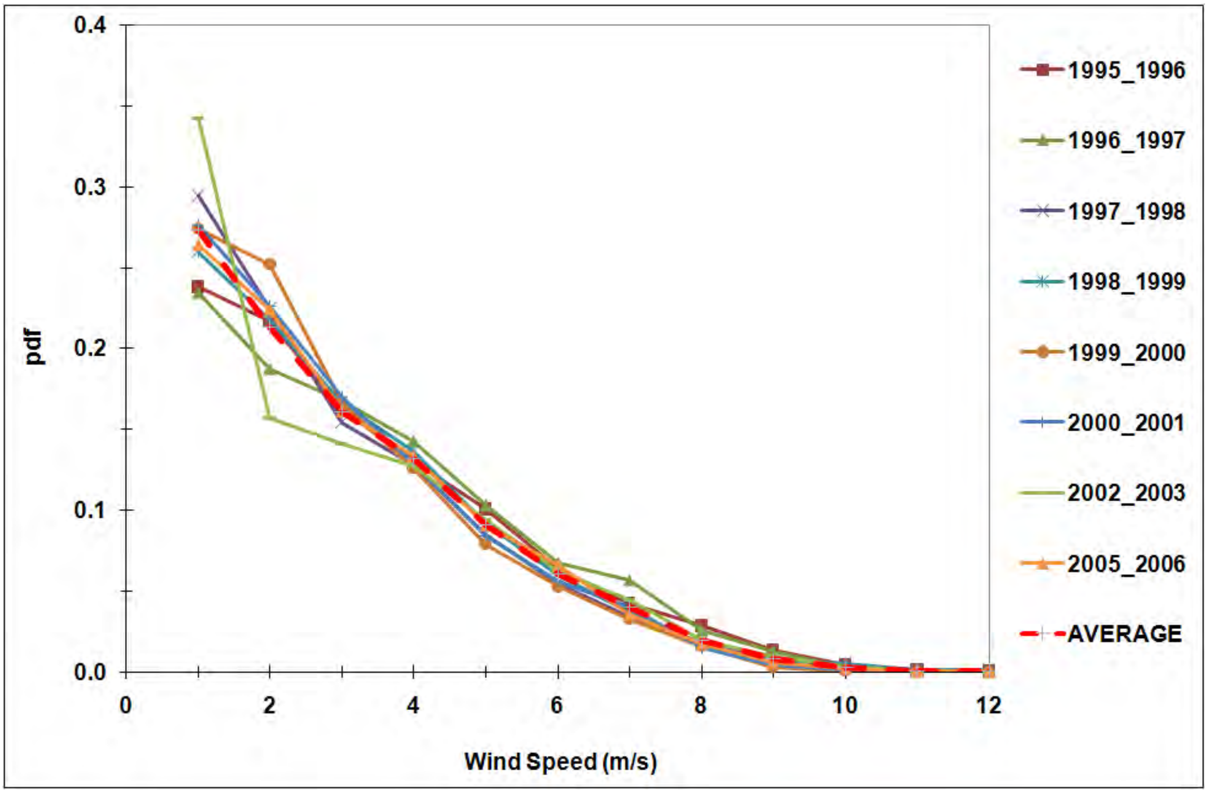

Figure C5	**Frequency distribution (pdf) of wind speed in 1 metre per second (m/s) bins for all valid years at the Ravensworth monitoring site**		
Location: Ravensworth monitoring site	**Period:** March 1995 to February 2009 (excluding 1994 to 1995, 2001 to 2002, 2002 to 2003, 2004 to 2005, 2006 to 2007, 2007 to 2008 and 2008 to 2009)	**Data source:** Macquarie Generation	**Units:** m/s
Type: pdf	**Averaging period:** 1-hour	**Prepared by:** Andrew Wiebe	**Date:** August 2009

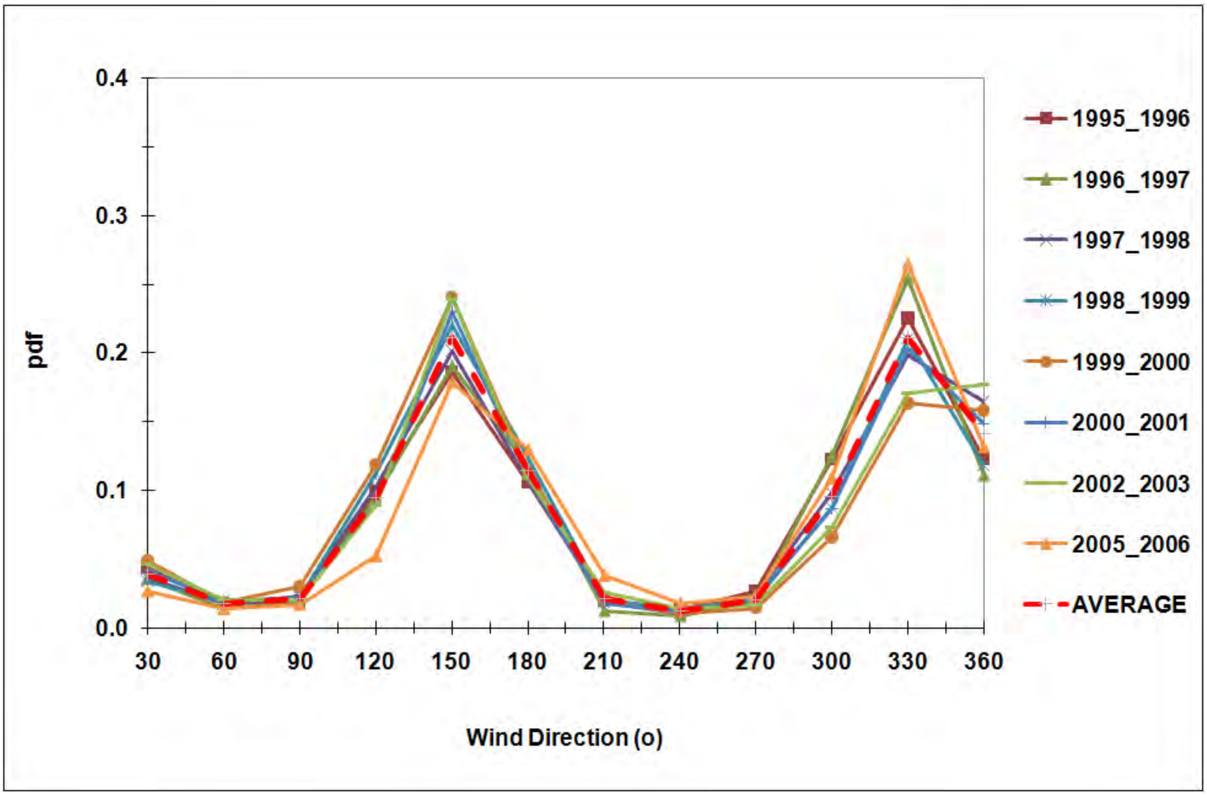

Figure C6 Frequency distribution (pdf) of wind direction in 30 degree (°) bins for all valid years at the Ravensworth monitoring site

Location:	**Period:**	**Data source:**	**Units:**
Ravensworth monitoring site	March 1994 to February 2009 (excluding 1994 to 1995, 2001 to 2002, 2002 to 2003, 2004 to 2005, 2006 to 2007, 2007 to 2008 and 2008 to 2009)	Macquarie Generation	degrees
Type:	**Averaging period:**	**Prepared by:**	**Date:**
pdf	1-hour	Andrew Wiebe	August 2009

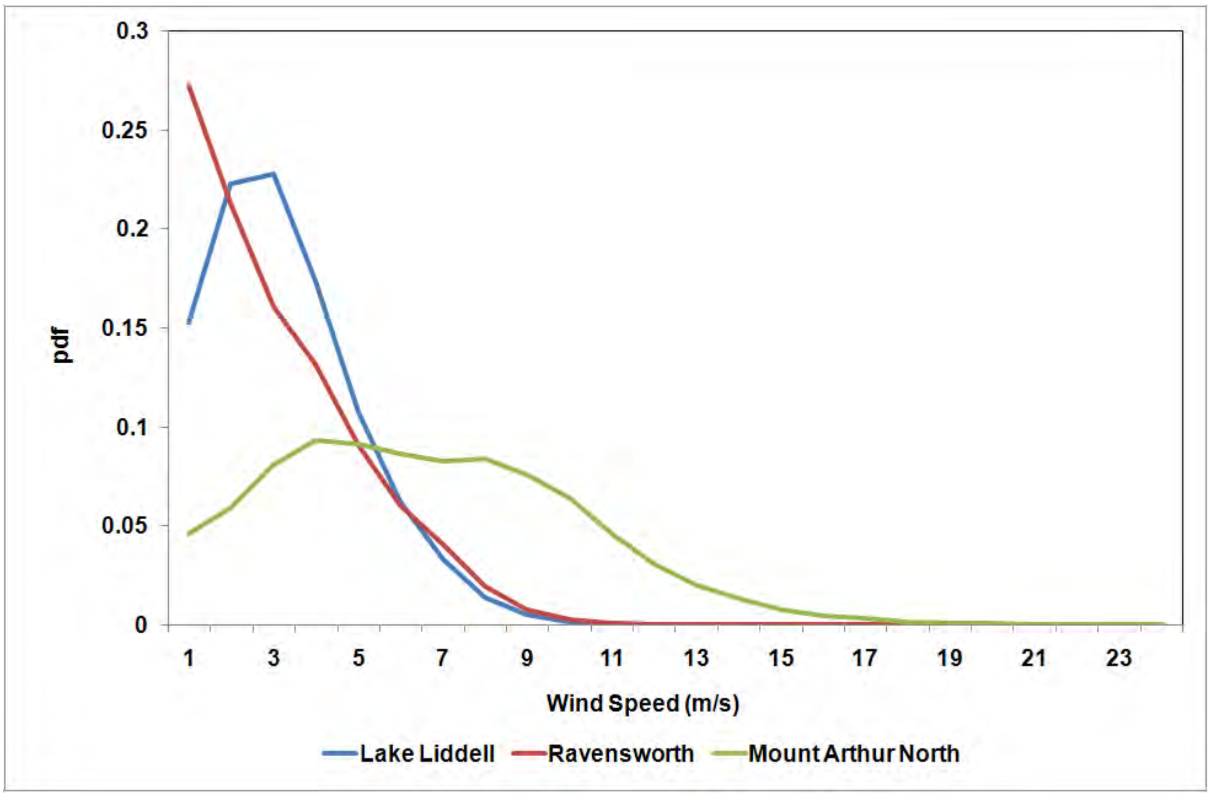

Figure C7	**Frequency distribution (pdf) of wind speed in 1 metre per second (m/s) bins for the climatological average at the Lake Liddell, Ravensworth and Mount Arthur North monitoring site**

Location:	**Period:**	**Data source:**	**Units:**
Lake Liddell, Ravensworth and Mount Arthur North monitoring site	March 1995 to February 2009 (valid years only)	Macquarie Generation	m/s
Type: pdf	**Averaging period:** 1-hour	**Prepared by:** Andrew Wiebe	**Date:** August 2009

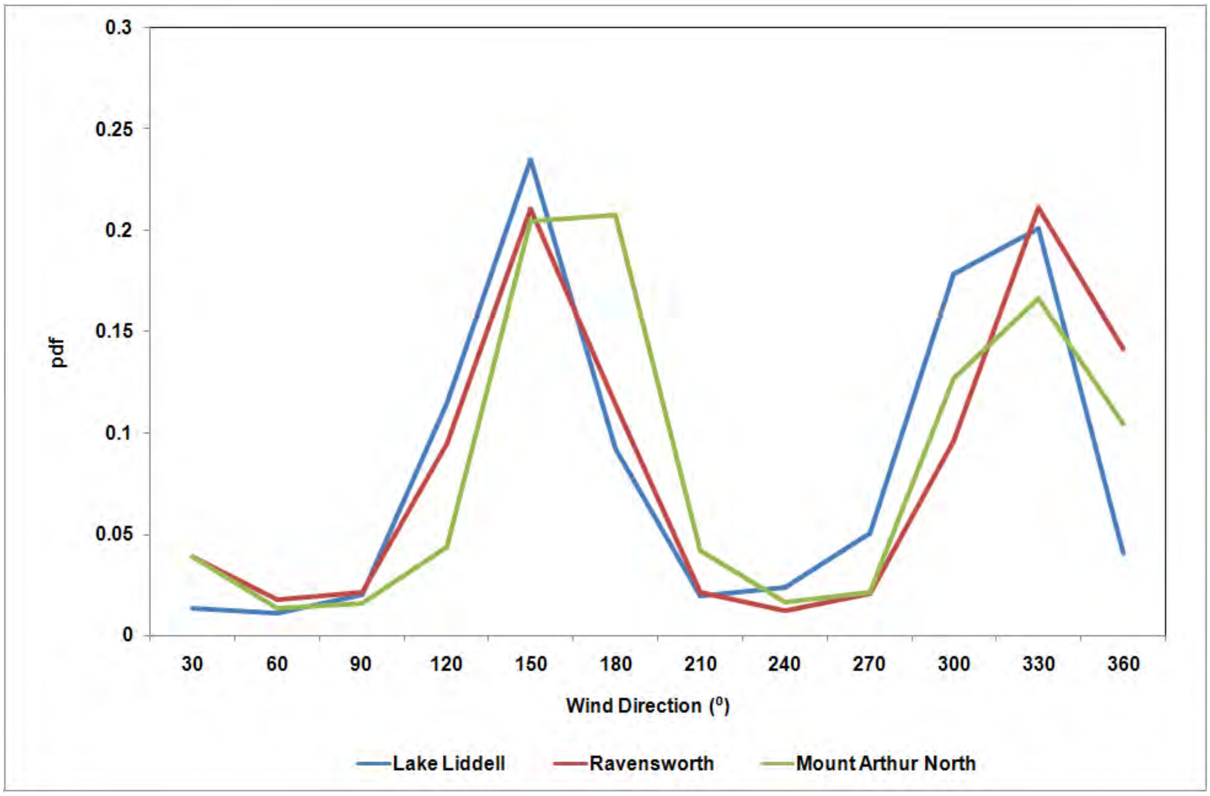

Figure C8	**Frequency distribution (pdf) of wind direction in 30 degree (°) bins for the climatological average at the Lake Liddell, Ravensworth and Mount Arthur North monitoring site**

Location:	**Period:**	**Data source:**	**Units:**
Lake Liddell, Ravensworth and Mount Arthur North monitoring site	March 1995 to February 2009 (valid years only)	Macquarie Generation	degrees
Type:	**Averaging period:**	**Prepared by:**	**Date:**
pdf	1-hour	Andrew Wiebe	August 2009

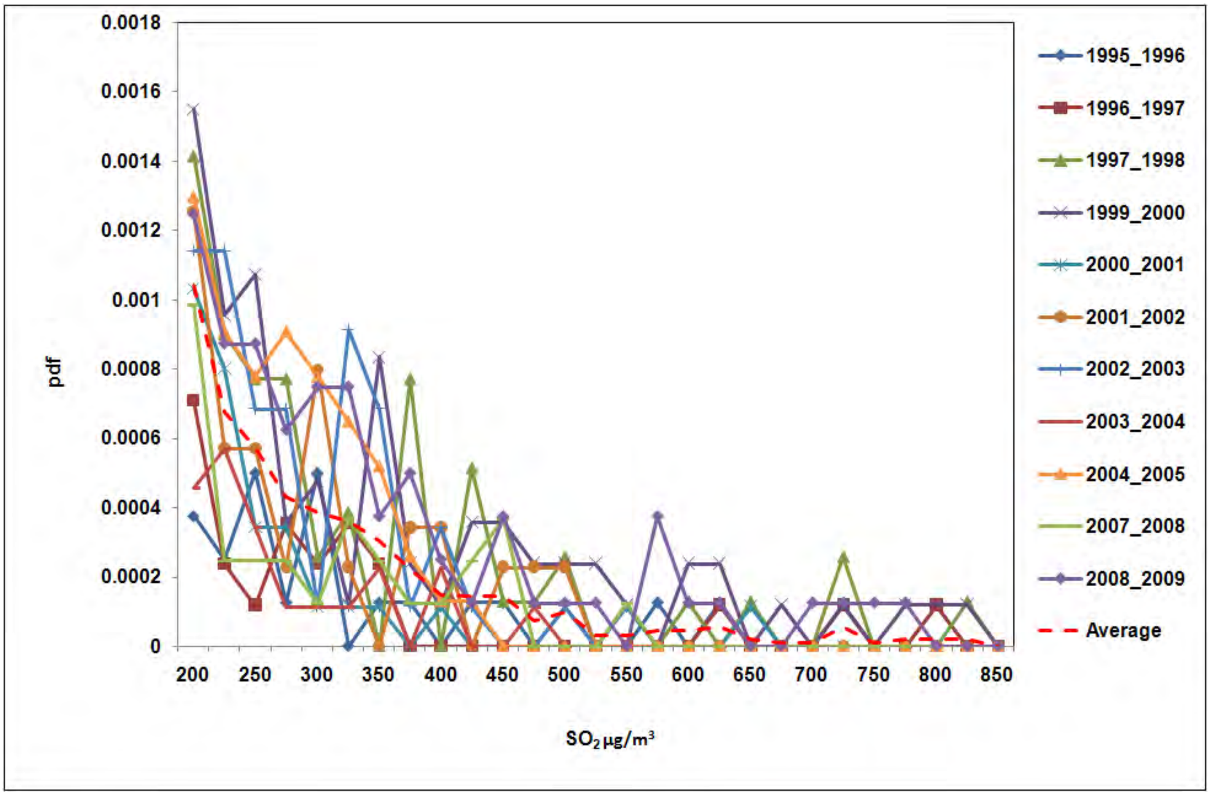

Figure C9	**Frequency distribution (pdf) of Sulfur dioxide (SO_2) in 25 microgram per cubic metre (µg/m³) bins at the Lake Liddell monitoring site for all valid years. Only the high end (> 200 µg/m³) of the distribution is displayed.**

Location:	**Period:**	**Data source:**	**Units:**
Lake Liddell monitoring site	March 1995 to February 2009 (valid years only)	Macquarie Generation	µg/m³
Type: pdf	**Averaging period:** 1-hour	**Prepared by:** Andrew Wiebe	**Date:** August 2009

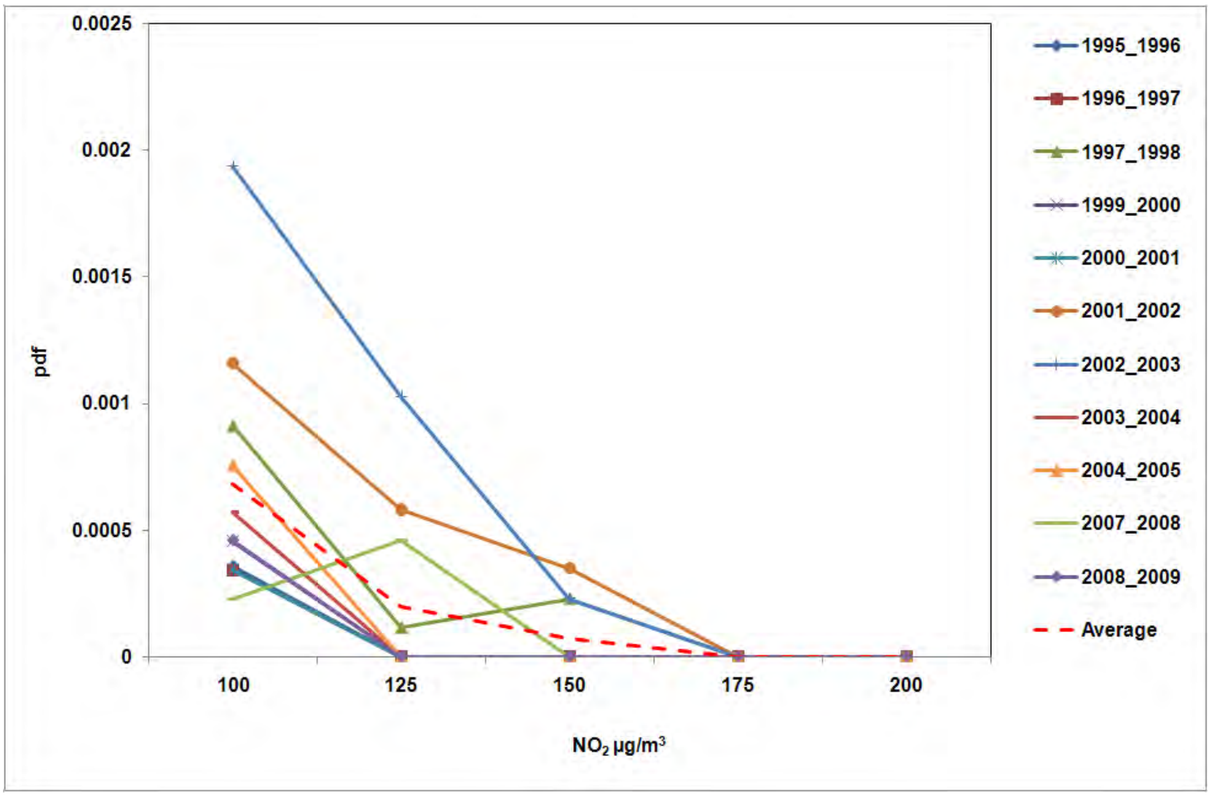

Figure C10 **Frequency distribution (pdf) of Nitrogen dioxide (NO_2) in 25 microgram per cubic metre (µg/m³) bins at the Lake Liddell monitoring site for all valid years. Only the high end (> 100 µg/m³) of the distribution is displayed.**

Location: Lake Liddell monitoring site	**Period:** March 1995 to February 2009 (valid years only)	**Data source:** Macquarie Generation	**Units:** µg/m³
Type: pdf	**Averaging period:** 1-hour	**Prepared by:** Andrew Wiebe	**Date:** August 2009

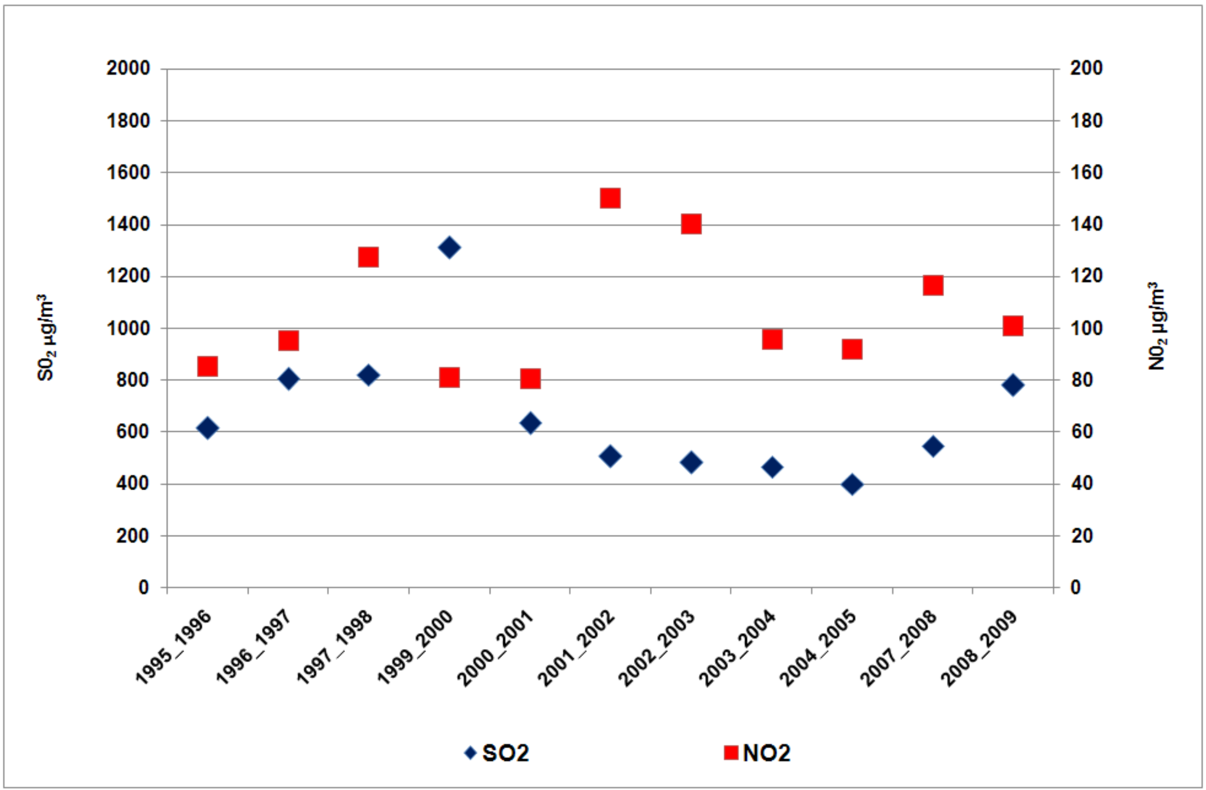

Figure C11	**Maximum 1 hour concentration of Sulfur dioxide (SO_2) and Nitrogen dioxide (NO_2) in the Lake Liddell monitoring site for all valid years.**		
Location: Lake Liddell monitoring site	**Period:** March 1995 to February 2009 (valid years only)	**Data source:** Macquarie Generation	**Units:** µg/m^3
Type: Scatter plot	**Averaging period:** 1-hour	**Prepared by:** Andrew Wiebe	**Date:** August 2009

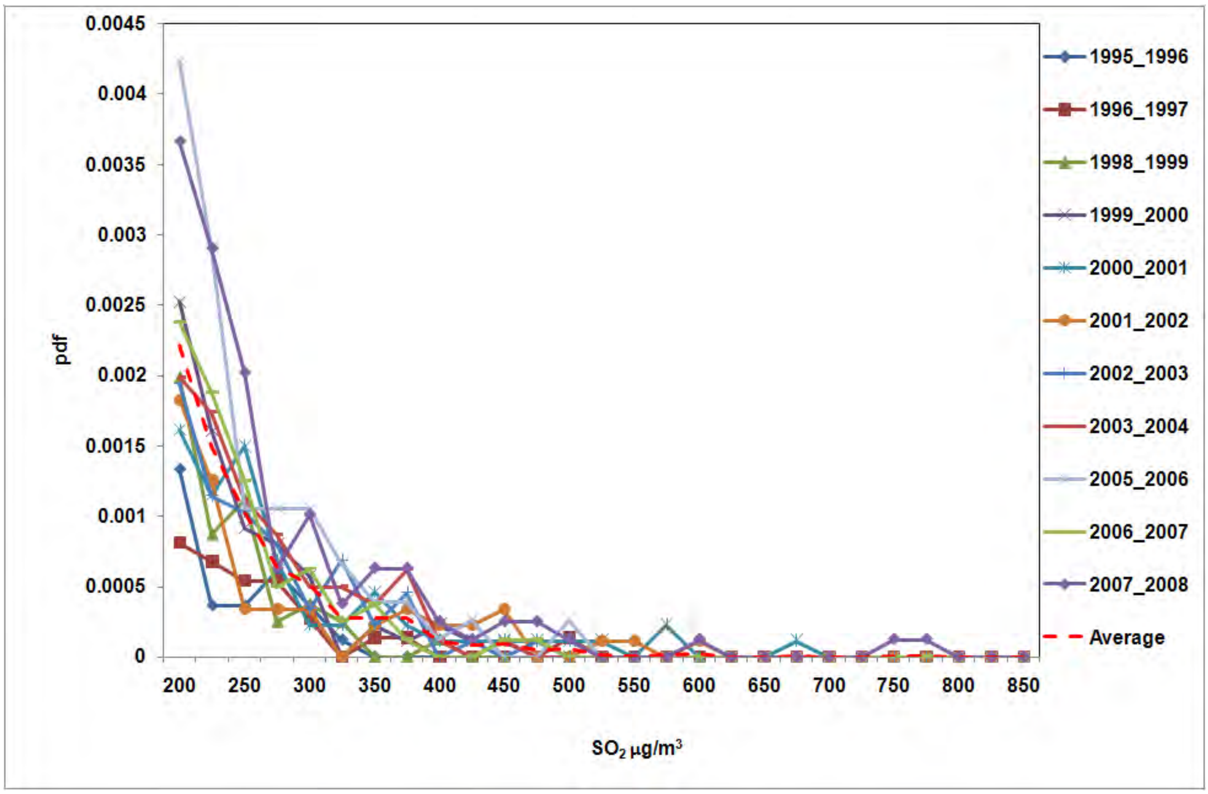

Figure C12	**Frequency distribution (pdf) of Sulfur dioxide (SO_2) in 25 microgram per cubic metre (µg/m³) bins at the Ravensworth monitoring site for all valid years. Only the high end (> 200 µg/m³) of the distribution is displayed.**

Location:	**Period:**	**Data source:**	**Units:**
Ravensworth monitoring site	March 1995 to February 2009 (valid years only)	Macquarie Generation	µg/m³
Type: pdf	**Averaging period:** 1-hour	**Prepared by:** Andrew Wiebe	**Date:** August 2009

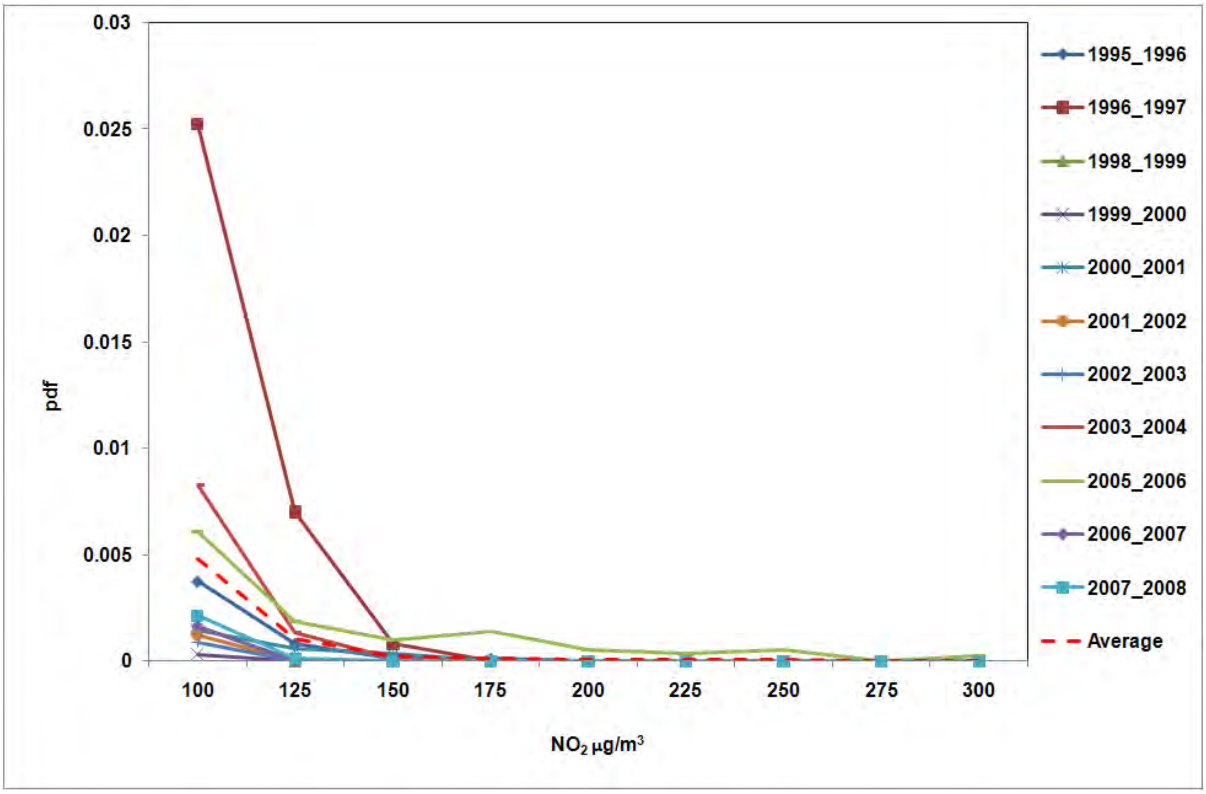

Figure C13	**Frequency distribution (pdf) of Nitrogen dioxide (NO_2) in 25 microgram per cubic metre (µg/m³) bins at the Ravensworth monitoring site for all valid years. Only the high end (> 100 µg/m³) of the distribution is displayed.**		
Location: Ravensworth monitoring site	**Period:** March 1995 to February 2009 (valid years only)	**Data source:** Macquarie Generation	**Units:** µg/m³
Type: pdf	**Averaging period:** 1-hour	**Prepared by:** Andrew Wiebe	**Date:** August 2009

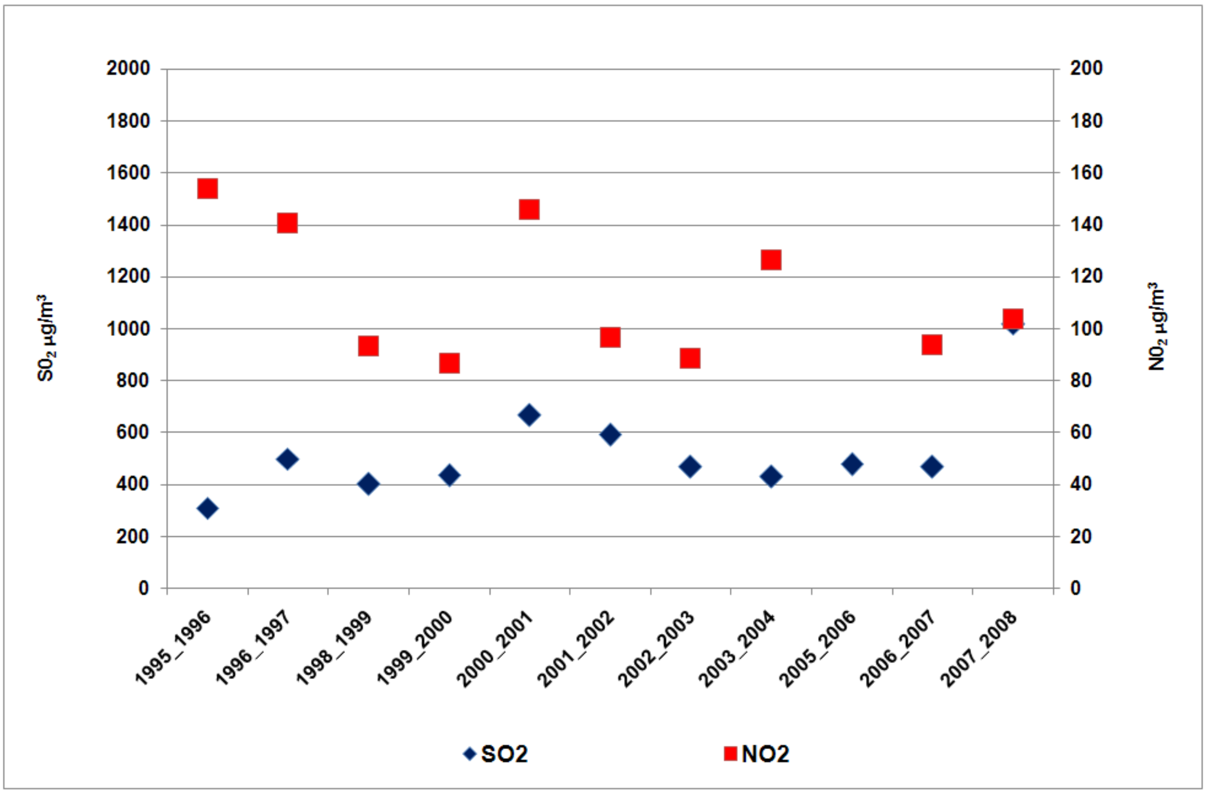

Figure C14	**Maximum 1 hour concentration of Sulfur dioxide (SO_2) and Nitrogen dioxide (NO_2) in the Ravensworth monitoring site for all valid years.**		
Location: Ravensworth monitoring site	**Period:** March 1995 to February 2009 (valid years only)	**Data source:** Macquarie Generation	**Units:** µg/m^3
Type: Scatter plot	**Averaging period:** 1-hour	**Prepared by:** Andrew Wiebe	**Date:** August 2009

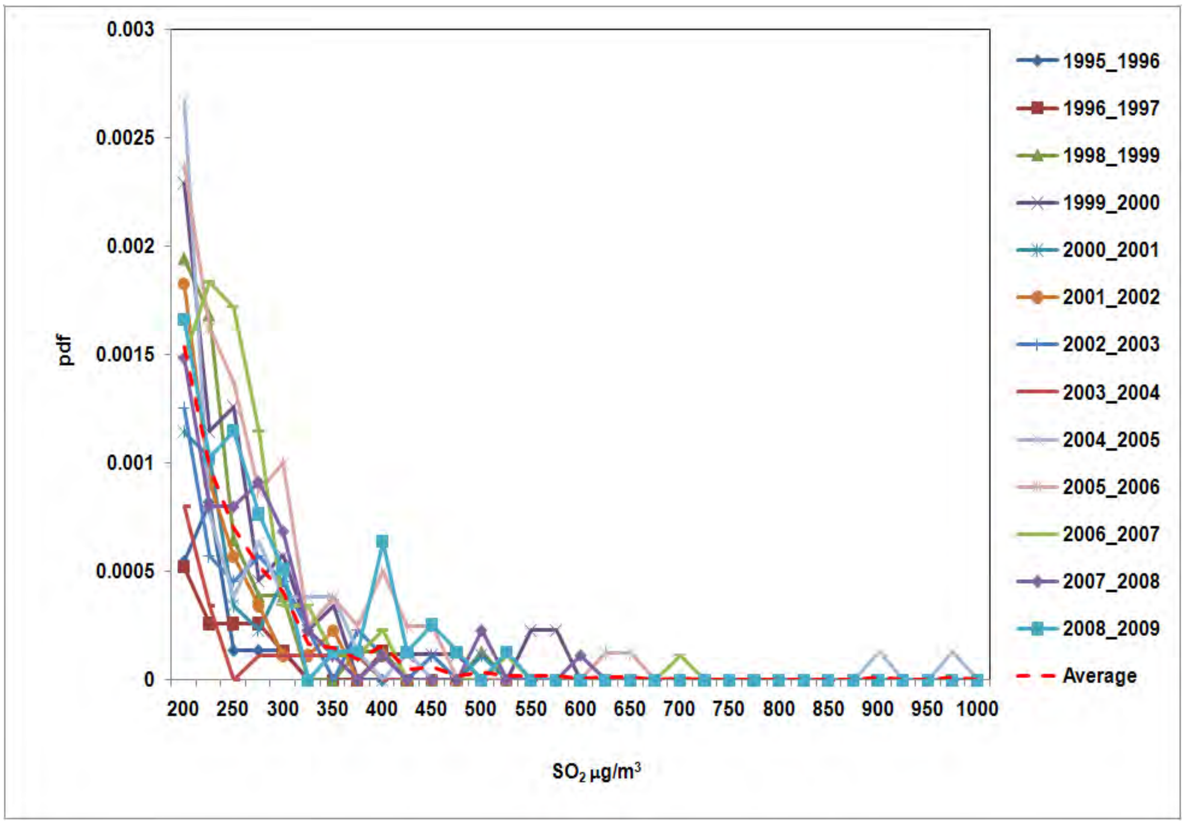

Figure C15	**Frequency distribution (pdf) of Sulfur dioxide (SO_2) in 25 microgram per cubic metre (µg/m³) bins at the Muswellbrook monitoring site for all valid years. Only the high end (> 200 µg/m³) of the distribution is displayed.**

Location:	**Period:**	**Data source:**	**Units:**
Muswellbrook monitoring site	March 1995 to February 2009 (valid years only)	Macquarie Generation	µg/m³
Type: pdf	**Averaging period:** 1-hour	**Prepared by:** Andrew Wiebe	**Date:** August 2009

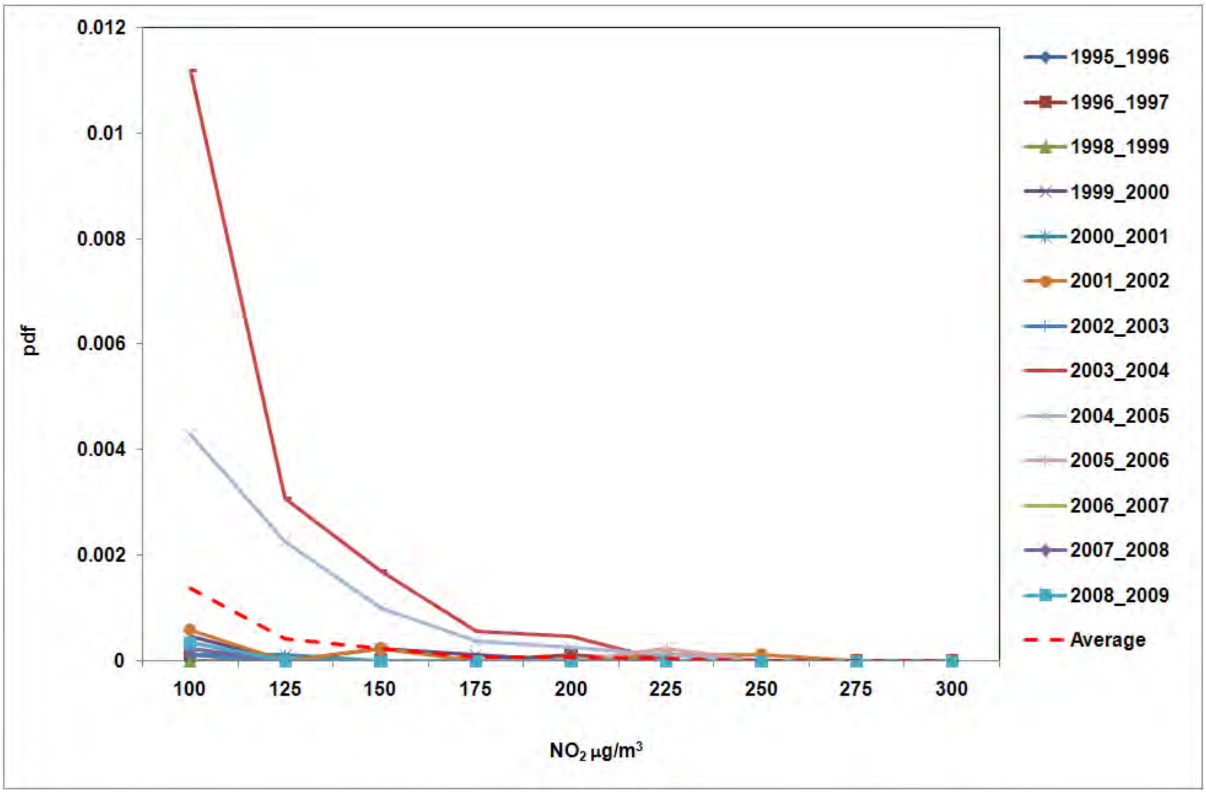

Figure C16	**Frequency distribution (pdf) of Nitrogen dioxide (NO_2) in 25 microgram per cubic metre (µg/m³) bins at the Muswellbrook monitoring site for all valid years. Only the high end (> 100 µg/m³) of the distribution is displayed.**

Location:	**Period:**	**Data source:**	**Units:**
Muswellbrook monitoring site	March 1995 to February 2009 (valid years only)	Macquarie Generation	µg/m³
Type:	**Averaging period:**	**Prepared by:**	**Date:**
pdf	1-hour	Andrew Wiebe	August 2009

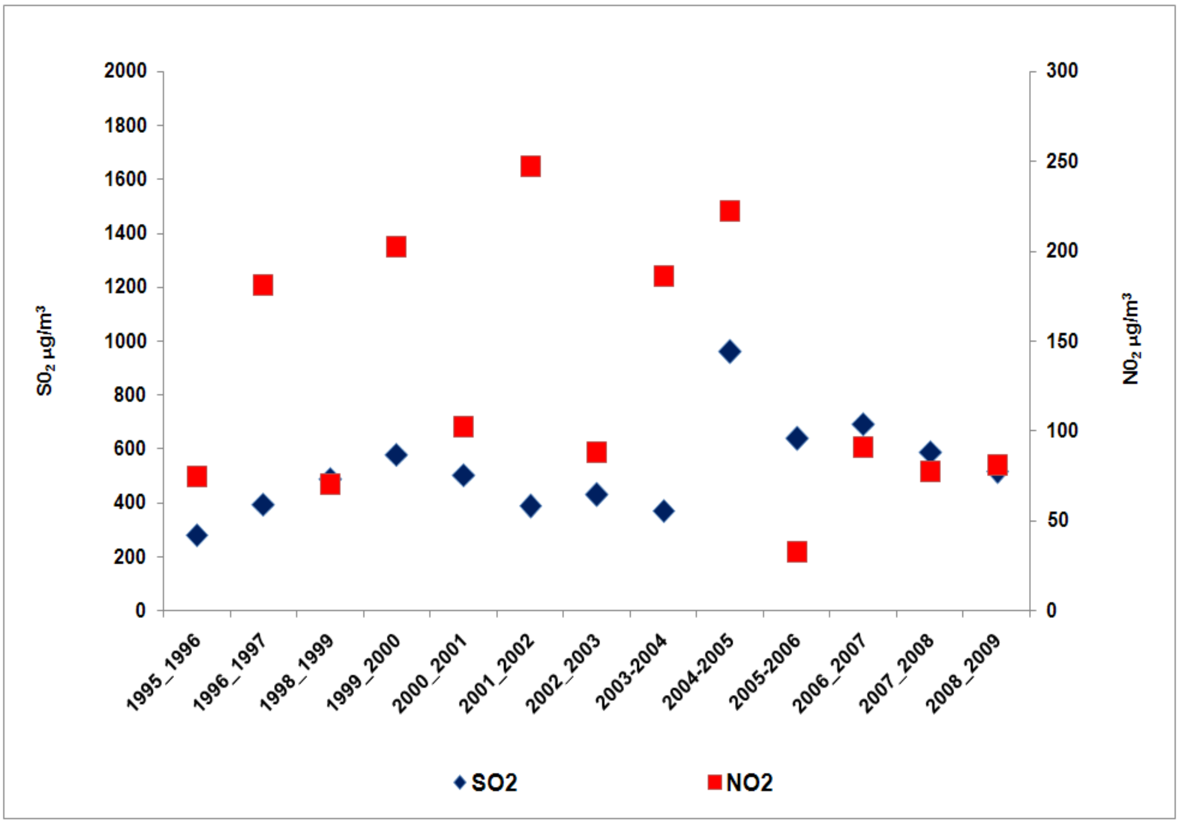

Figure C17	**Maximum 1 hour concentration of Sulfur dioxide (SO_2) and Nitrogen dioxide (NO_2) in the Muswellbrook monitoring site for all valid years.**

Location:	**Period:**	**Data source:**	**Units:**
Muswellbrook monitoring site	March 1995 to February 2009 (valid years only)	Macquarie Generation	µg/m³
Type:	**Averaging period:**	**Prepared by:**	**Date:**
Scatter plot	1-hour	Andrew Wiebe	August 2009

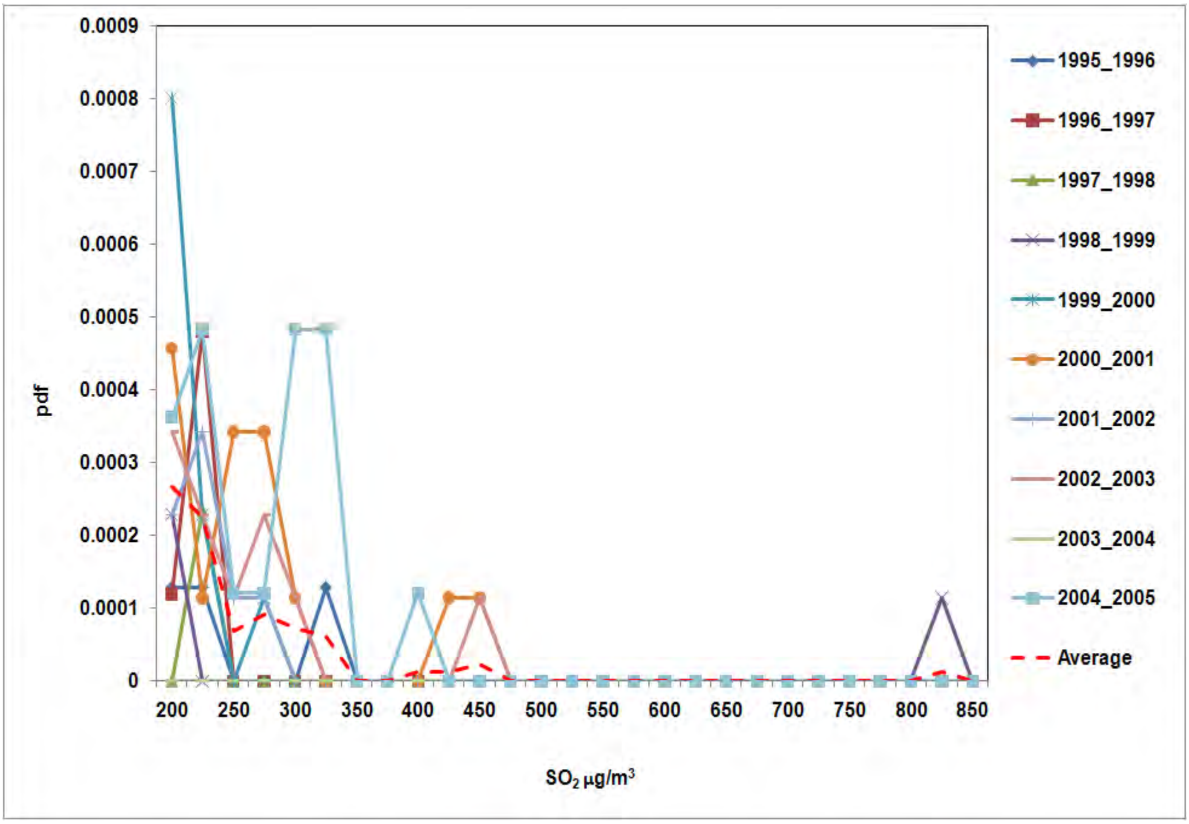

Figure C18	**Frequency distribution (pdf) of Sulfur dioxide (SO_2) in 25 microgram per cubic metre (µg/m³) bins at the Singleton monitoring site for all valid years. Only the high end (> 200 µg/m³) of the distribution is displayed.**

Location:	**Period:**	**Data source:**	**Units:**
Singleton monitoring site	March 1995 to February 2009 (valid years only)	Macquarie Generation	µg/m³
Type: pdf	**Averaging period:** 1-hour	**Prepared by:** Andrew Wiebe	**Date:** August 2009

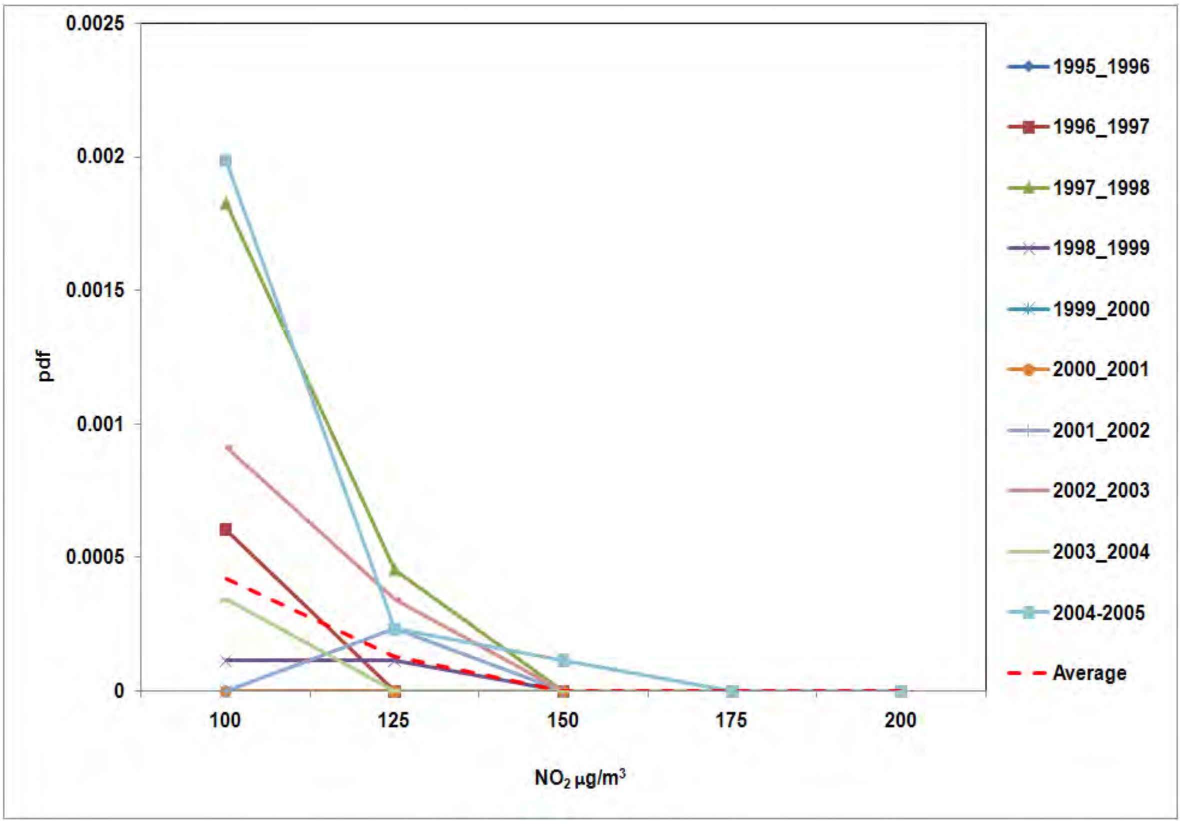

Figure C19 **Frequency distribution (pdf) of Nitrogen dioxide (NO_2) in 25 microgram per cubic metre ($\mu g/m^3$) bins at the Singleton monitoring site for all valid years. Only the high end (> 100 $\mu g/m^3$) of the distribution is displayed.**			
Location: Singleton monitoring site	**Period:** March 1995 to February 2009 (valid years only)	**Data source:** Macquarie Generation	**Units:** $\mu g/m^3$
Type: pdf	**Averaging period:** 1-hour	**Prepared by:** Andrew Wiebe	**Date:** August 2009

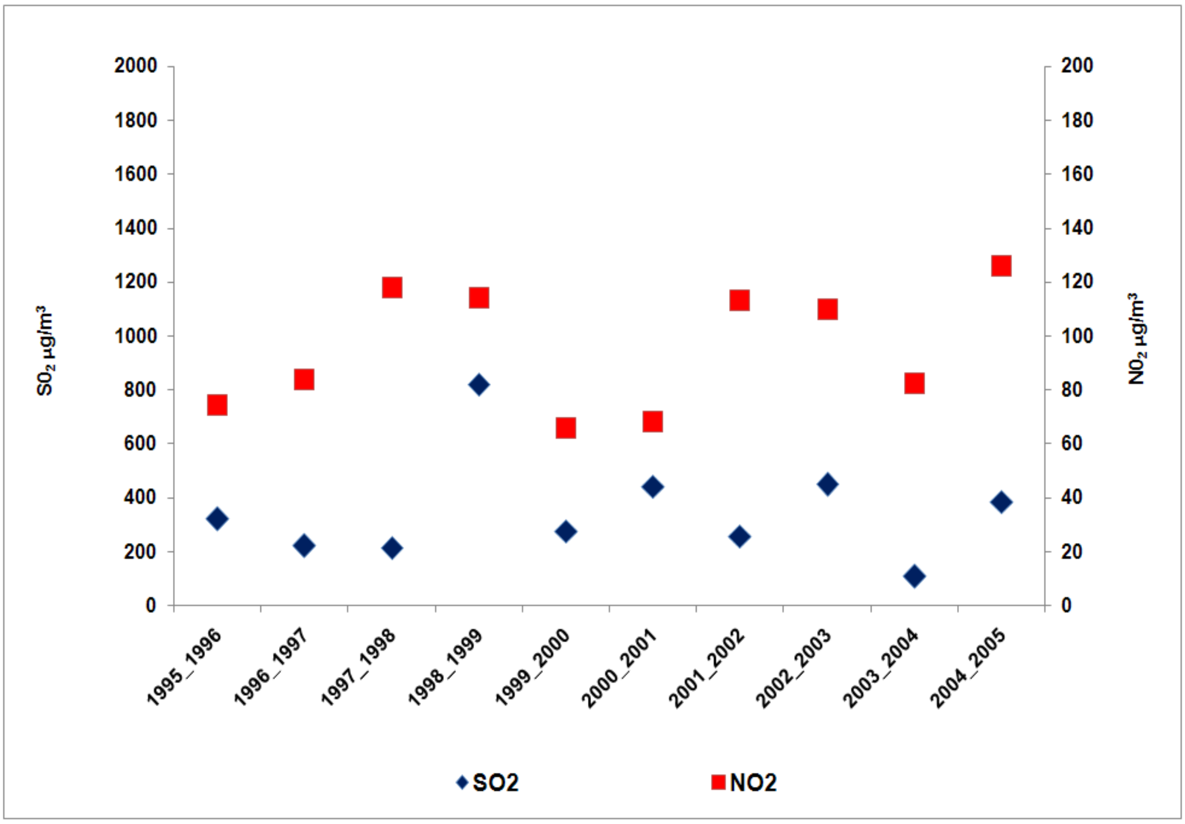

Figure C20	**Maximum 1 hour concentration of Sulfur dioxide (SO_2) and Nitrogen dioxide (NO_2) in the Singleton monitoring site for all valid years.**

Location:	**Period:**	**Data source:**	**Units:**
Singleton monitoring site	March 1995 to February 2009 (valid years only)	Macquarie Generation	µg/m^3
Type:	**Averaging period:**	**Prepared by:**	**Date:**
Scatter plot	1-hour	Andrew Wiebe	August 2009

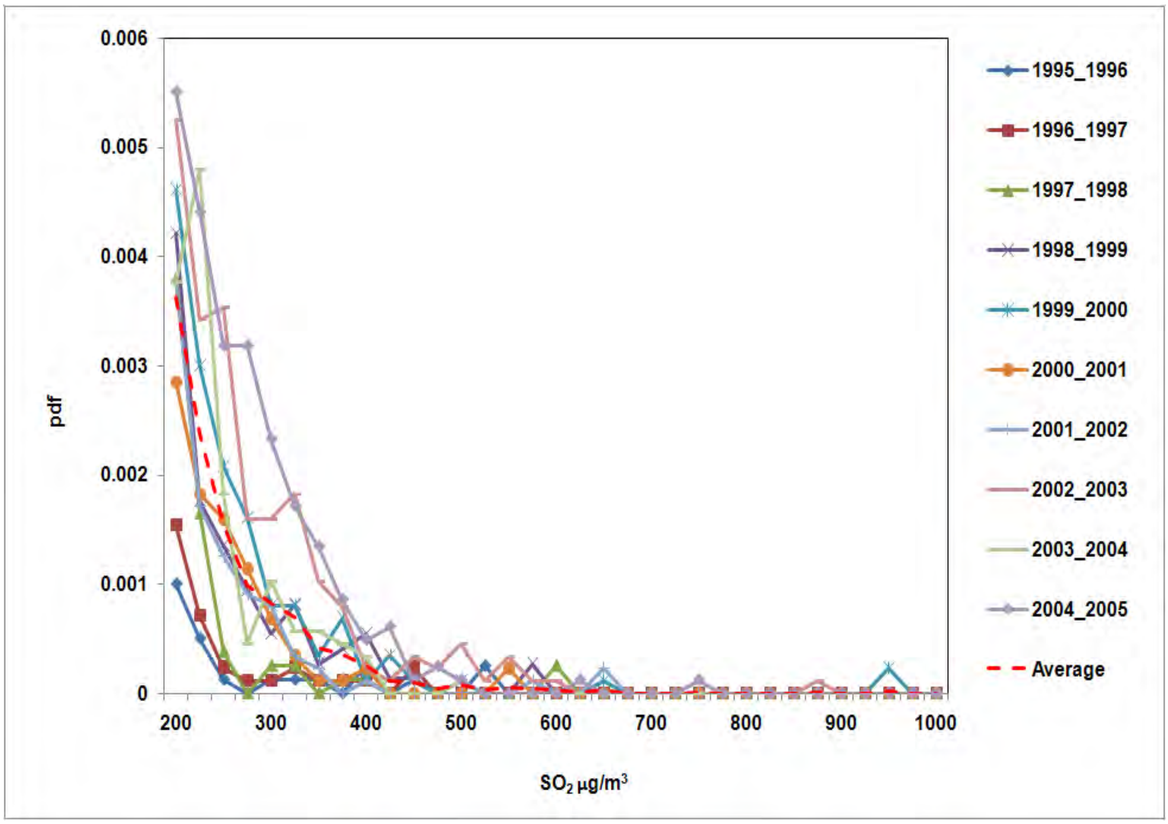

Figure C21	**Frequency distribution (pdf) of Sulfur dioxide (SO_2) in 25 microgram per cubic metre (µg/m³) bins at the Mount Arthur North monitoring site for all valid years. Only the high end (> 200 µg/m³) of the distribution is displayed.**

Location:	**Period:**	**Data source:**	**Units:**
Mount Arthur North monitoring site	March 1995 to February 2009 (valid years only)	Macquarie Generation	µg/m³
Type: pdf	**Averaging period:** 1-hour	**Prepared by:** Andrew Wiebe	**Date:** August 2009

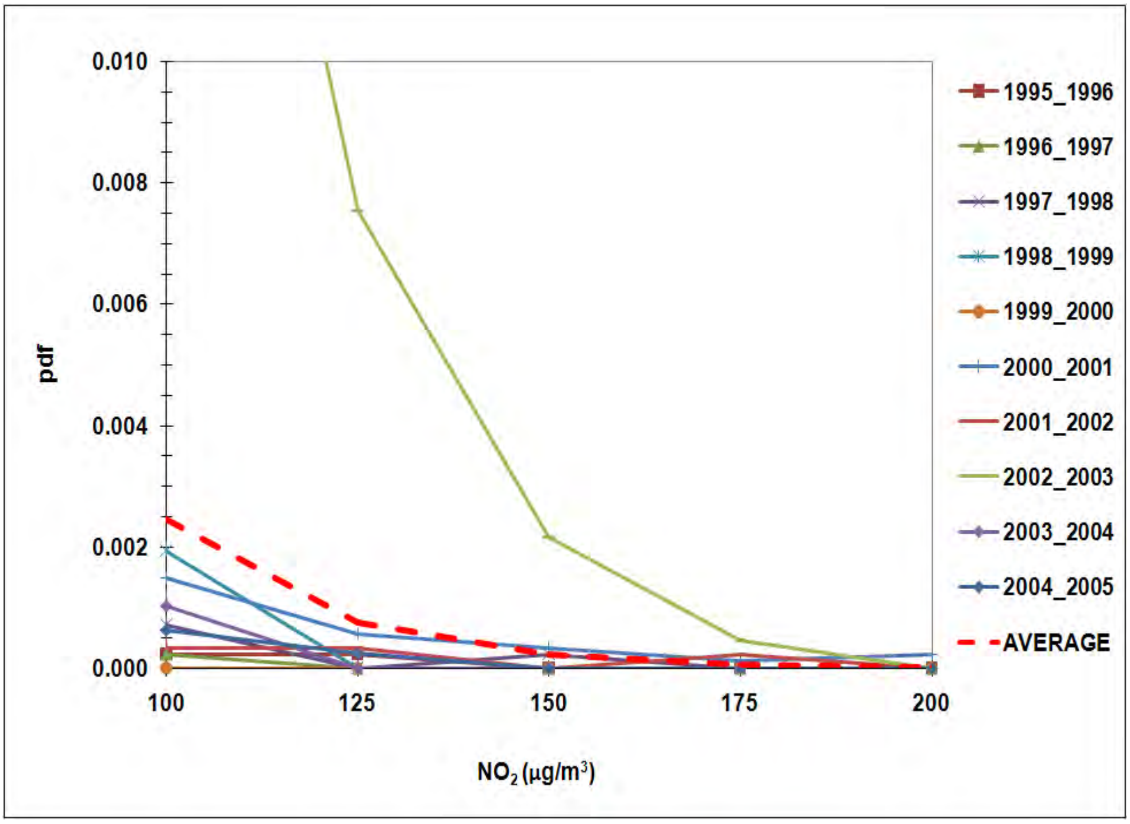

Figure C22Frequency distribution (pdf) of Nitrogen dioxide (NO_2) in 25 microgram per cubic metre (µg/m^3) bins at the Mount Arthur North monitoring site for all valid years. Only the high end (> 100 µg/m^3) of the distribution is displayed.

Location: Mount Arthur North monitoring site	**Period:** March 1995 to February 2009 (valid years only)	**Data source:** Macquarie Generation	**Units:** µg/m^3
Type: pdf	**Averaging period:** 1-hour	**Prepared by:** Andrew Wiebe	**Date:** August 2009

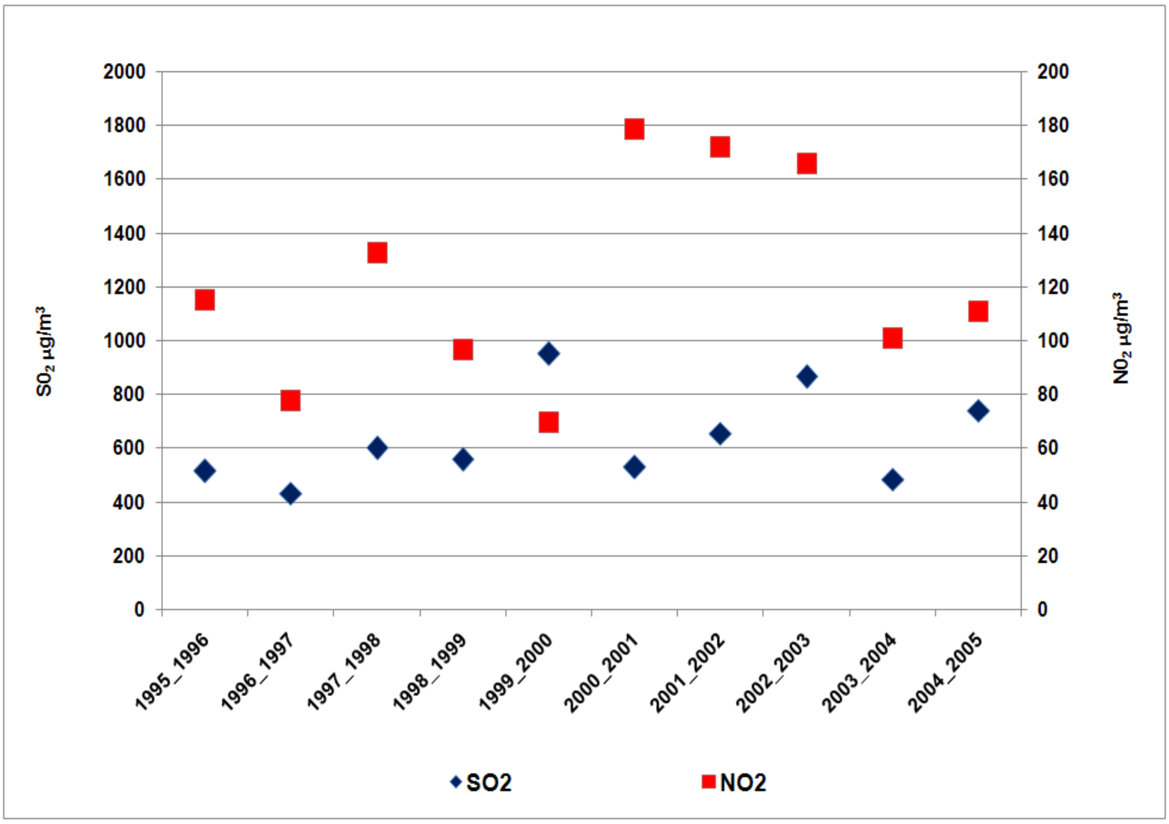

Figure C23	**Maximum 1 hour concentration of Sulfur dioxide (SO_2) and Nitrogen dioxide (NO_2) in the Mount Arthur North monitoring site for all valid years.**		
Location: Mount Arthur North monitoring site	**Period:** March 1995 to February 2009 (valid years only)	**Data source:** Macquarie Generation	**Units:** µg/m^3
Type: Scatter plot	**Averaging period:** 1-hour	**Prepared by:** Andrew Wiebe	**Date:** August 2009

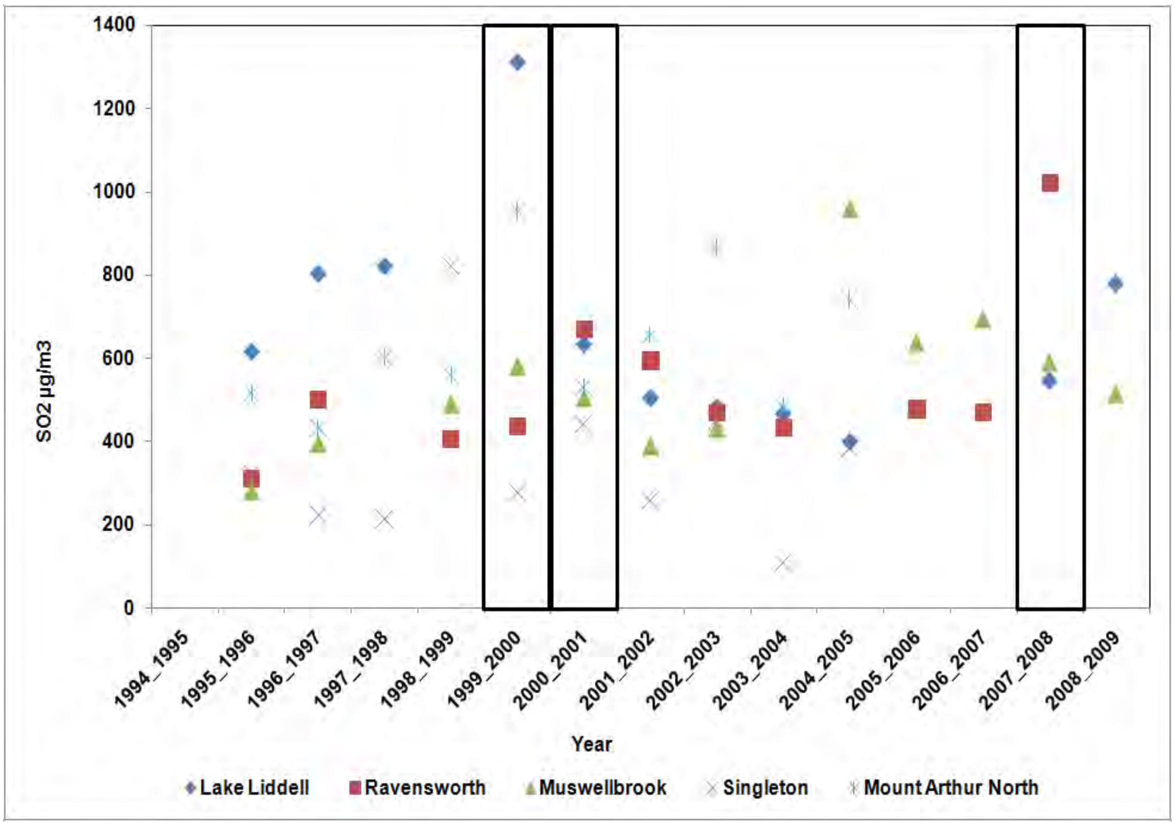

Figure C24	**Maximum 1 hour concentration of Sulfur dioxide (SO_2) at the Lake Liddell, Ravensworth, Muswellbrook and Mount Arthur North monitoring site for valid years only. Boxed points indicate selected years for dispersion modelling**

Location:	**Period:**	**Data source:**	**Units:**
Lake Liddell, Ravensworth, Muswellbrook and Mount Arthur North monitoring sites	March 1994 to February 2009 (valid years only)	Macquarie Generation	µg/m^3
Type:	**Averaging period:**	**Prepared by:**	**Date:**
Scatter plot	1-hour	Andrew Wiebe	August 2009

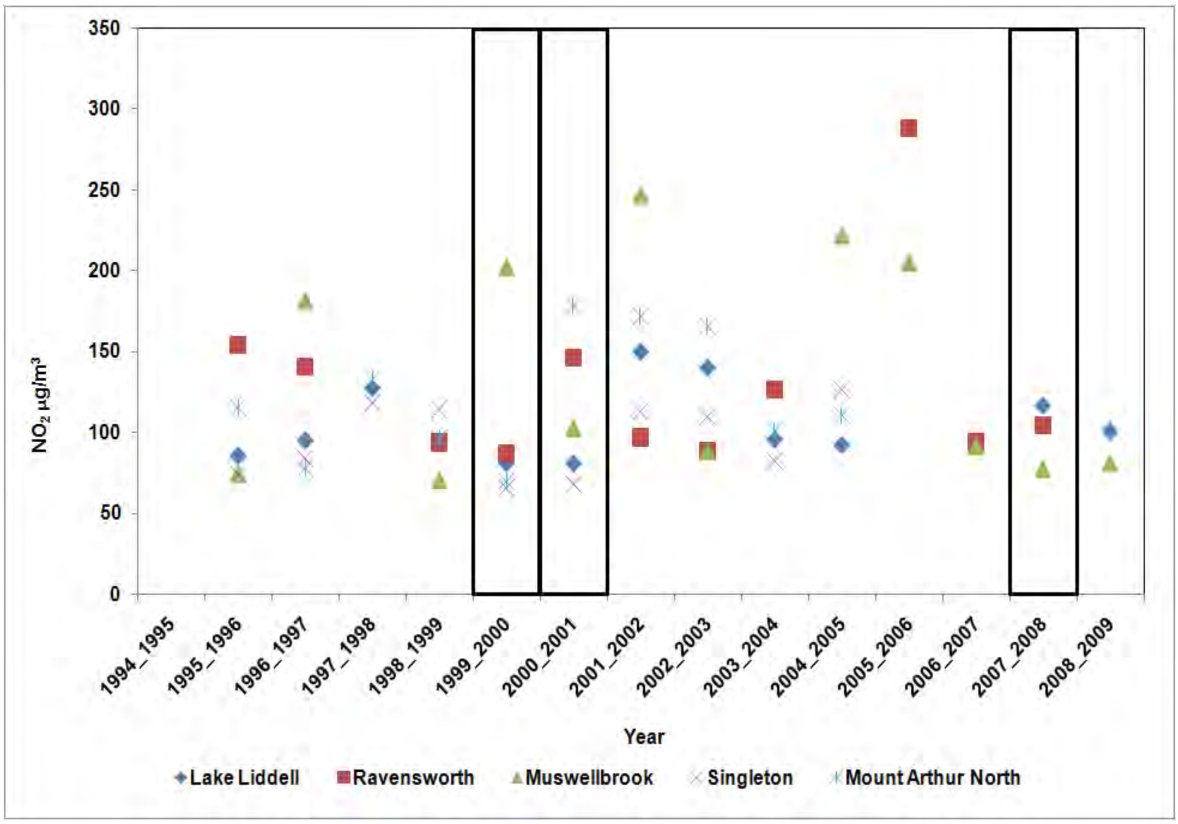

Figure C25	**Maximum 1 hour concentration of Nitogen dioxide (NO_2) at the Lake Liddell, Ravensworth, Muswellbrook and Mount Arthur North monitoring site for valid years only. Boxed points indicate selected years for dispersion modelling**		
Location: Lake Liddell, Ravensworth, Muswellbrook and Mount Arthur North monitoring sites	**Period:** March 1994 to February 2009 (valid years only)	**Data source:** Macquarie Generation	**Units:** µg/m^3
Type: Scatter plot	**Averaging period:** 1-hour	**Prepared by:** Andrew Wiebe	**Date:** August 2009