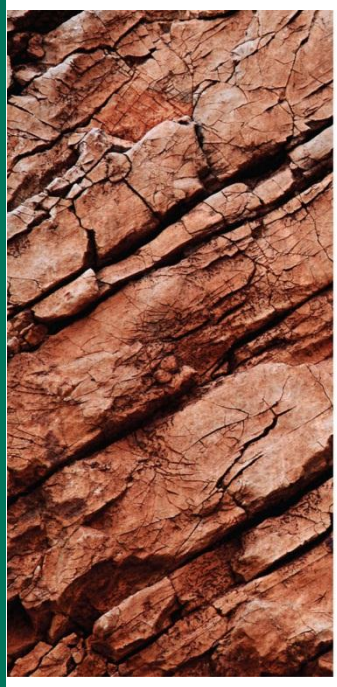
27 September 2012

REPORT

LITERATURE REVIEW AND RESEARCH PLAN

Towards Improving the Prediction of Extreme Pollution Concentrations - Accounting for Uncertainty and Randomness

Submitted to:
NIWA Ltd
41 Market Place
Private Bag 99940
Auckland 1149
New Zealand



Report Number.
0678104507_001_R_Rev0
_2012

Golder Associates

Table of Contents

TABLES

APPENDICES

EXECUTIVE SUMMARY

The aim of the research described in this report is to improve the modelling and prediction of exceedences of the National Environmental Standard (NES) for PM_{10} and disseminate practical guidance on appropriate methods to modellers and regulators. The report takes the first steps towards fulfilling this aim by reviewing the scientific literature on extreme air pollution events and their modelling, and presenting a methodology for a New Zealand-specific case study which incorporates concepts of model sensitivity and uncertainty analysis.

The literature review is concerned with measures of high-end pollution concentrations associated with extreme events, such as NES exceedences. The measures include the ground-level concentration (glc) itself, the number of target exceedences per year and, equivalently, the probability of exceedence. These measures are often not reproduced or predicted well by dispersion models, yet they are of prime importance to regulators such as Regional Councils.

There are several potential reasons for the apparent poor model performance discussed in the review. These may due to uncertainties in the modelling, from incomplete or inappropriate input information or from poor model performance. Investigating *uncertainty* in model predictions leads to more *certainty*, as confidence limits can be placed on the predictions. Without confidence limits, there is no indication as to the reliability of the predictions.

A key aspect to account for is the stochastic nature of pollution concentrations; the apparent randomness which can follow standard probability distributions, especially in the upper tail of the concentration distribution. Methods for accounting for this in modelled and observed data sets are discussed here.

This review discusses the sensitivity of model results to external parameters. A sensitivity analysis is a worthwhile step to determine the parameters which need investigating most closely if model performance is to be improved. Following this, the review discusses the next step, which is an uncertainty analysis, quantifying the likely changes in output concentrations given the likely range of input parameters. Essentially, three types of uncertainty have been identified and studied in the literature. These are (1) uncertainties in data sources, such as emissions, meteorology and ambient air quality data, (2) uncertainties due to the formulation of model dynamics, and (3) unresolved features such as turbulent fluctuations in meteorology and consequently in concentrations. The total uncertainty may be couched in standard ways in terms of model performance statistics, such as fractional bias, normalized mean-squared error and root-mean-square error, index of agreement and factor-of-2 measures. These statistical measures may also be applied to the various components of the model uncertainty.

The information gained from the uncertainty analysis can be used to improve model performance. Inverse modelling techniques can be used to evaluate the known uncertainties together, and optimize inputs such as emissions. This can then lead to a best-fit combination of all sources of information and general reduction in error-bars. A technique for doing this is outlined in the review.

The initial uncertainty analysis focuses on model performance for average conditions, not extremes. Indeed, no unified theory apportioning uncertainty to its various sources for extreme concentrations was found.

The review then outlines methods which focus on the tail-end of the distribution of air pollution concentrations. These include the use of suitable model performance measures, such as the relative percentile error and probability of exceedence, and normalized 'error' quantities applying only to the highest concentrations. Also, extreme value theory is outlined in the review. This allows the fitting of statistical distributions to the extreme observed and modelled concentrations, and derived quantities such as the daily maximum of 1-hour data (useful for SO_2) or the 24-hour maximum and other percentiles through the year (useful for PM_{10}). The extreme value theory can also incorporate re-sampling methods to determine confidence intervals of observed concentrations, with which model results may be compared.

Following from the literature review, the approaches are summarized and brought together in a research plan which focuses on a New Zealand based case study of uncertainties around modelling PM_{10} in an urban airshed.

1.0 INTRODUCTION

1.1 Aims of the Report

This report explores issues associated with use of dispersion modelling techniques to predict the concentrations and frequency of extreme air pollution events, including exceedences of ground-level concentration (glc) targets. The probability of exceedence (POE) of the NES for PM_{10} currently, or in future years under given air quality management scenarios is of interest to local authorities (MfE 2004b; MfE 2011). Knowledge of the likelihood of NES exceedences, and the potential number of them, would assist Regional Councils in their development of plans to manage air quality, and would assist in their assessment of the future effectiveness of such plans. Indeed, the NES make specific reference to the number of permitted exceedences in future years, providing for split targets in some regions (MfE 2011). Airsheds currently with more than 10 exceedences a year must meet three exceedences by 1 September 2016, and one exceedence by 1 September 2020. Airsheds with fewer than 10 exceedences a year must meet one exceedence by 1 September 2016.

At present it is reasonably common for Regional Councils to use airshed models or other air dispersion models to assess current urban PM_{10} concentrations or make predictions of future levels. The models can be used to predict whether the NES will be complied with, and to predict trends between now and 2020. The dispersion model is one of a number of air quality management tools at Councils' disposal, another key tool being the emissions inventory. However, dispersion models can have difficulties in predicting high concentrations. Whilst they can represent the magnitude of concentration reasonably well, they do not generally give good estimates of the number of exceedences. This may be because models are designed to produce 'ensemble-average' estimates of glcs under some predefined 'typical' conditions, whereas exceedences actually occur under extreme or unusual meteorological conditions and/or emissions scenarios. The high concentrations appear random and are not predictable by commonly-used deterministic modelling techniques. A way of increasing the value of dispersion models as airshed management tools is through a closer examination of the sources of uncertainty associated with model predictions, and accounting for variations from average meteorological and emission conditions using statistical techniques. The apparently random component of a distribution of glcs is modelled in a probabilistic way, and it is then appropriate to determine a confidence interval for the expected number of exceedences, rather than just a single estimate of this number.

A successful prediction of the likelihood air-pollution exceedences using a dispersion model includes the quantification of uncertainties associated with model results. A distribution of results, rather than a single number, is produced, from which follows the POE. Two corollaries of this are (i) model sensitivity to changes in external parameters needs to be known to determine how uncertainties in the parameters lead to uncertainties in model predictions, and (ii) the model performance needs to be quantified, as this leads also contributes to uncertainties in model predictions. Internationally, the evaluation of model uncertainty is recommended by regulators, if not often carried out. Many environmental agencies throughout the world promote and provide guidance on model uncertainty analysis. It is mentioned in the MfE's modelling good-practice guide (MfE 2004a) but rarely, if ever, carried out in New Zealand.

The ultimate goal of this work is to attain a better understanding of and improved prediction of conditions leading to air pollution exceedences. Working towards this goal, this report discusses techniques used to provide sensitivity and uncertainty analyses of model predictions, examines model performance measures used for pollution levels ranging from means to extremes, and considers the variety of sources of error and uncertainty associated with numerical dispersion modelling.

It should be noted that a complete picture of uncertainties includes those associated with emissions inventory development, meteorological monitoring, ambient air quality monitoring, and source apportionment analysis, for example. Thus the concepts presented here, as applied to dispersion modelling, are also appropriate to monitoring and other forms of data-gathering, and any modelling and analysis techniques.

This report[1] is the result of a 'pilot' project and takes the first steps towards the abovementioned goal. It presents a description of the issues at hand, reviews the relevant scientific literature, and a puts forward a plan for applying the concepts and methods to New Zealand-specific cases.

1.2 Giving Regulators More Certainty by Assessing Uncertainty

This section introduces some statistical concepts related to high-end air-pollution concentrations, and provides a link between the theoretical ideas and their applicability to regulators' needs.

Models can perform well for average conditions, and often give a reasonable estimate of the maximum concentration. However, knowledge of the number of exceedences requires a good model simulation of the high-end tail of the concentration distribution, including the upper percentiles. The high concentrations may occur under calm meteorological conditions, which are difficult to model, or they may be due to higher than usual emissions, which are not accounted for in an emissions inventory designed for typical conditions. Higher than average emissions themselves may occur under unusual meteorological conditions, or there may be variation due to human behaviour in the operation of domestic fires, for example. Incorporating these effects deterministically would need a good model of hour-by-hour changes in meteorology during low-wind conditions (and their effect on pollution concentration), and knowledge of actual emissions from day-to-day. These are stringent requirements which, if satisfied, would still not remove all sources of uncertainty. Therefore an alternative approach is examined here, in which sources of uncertainty due to unresolved processes can be represented by probability distributions as if they were random.

To reiterate, the use of the word 'random' here refers to components of dispersion processes which are not resolved by measurements or modelling, and which contain some uncertainty. The following points provide a focus for the review carried out in this report:

i) Air pollution glcs arise through a combination of deterministic and stochastic processes.

ii) There is an apparent 'random' component to observed air pollution glcs.

iii) The random component is larger for extreme pollution events than for average conditions.

iv) Glcs have been found to follow well-known probability distributions, but the distributions followed by the mid-range of concentrations are *different* to those followed by the high tail-ends (Rao et al. 1985; Lu and Fang 2003; Sharma et al. 1999).

This project is concerned with improving the model simulation of the upper tail of air pollution distributions for regulatory purposes, which is related to sensitivity and uncertainty analysis, and involves a statistical view of aspects of the modelling processes.

In New Zealand, predicted maximum concentrations from, say, a proposed industrial development, are reported with no formal determination of confidence limits. In the past, the calculation of confidence limits may have been daunting, requiring excessive computational resources for an ensemble of model simulations, or advanced statistical analysis. However, an analysis of uncertainty is expected in many overseas jurisdictions, which leads to a probabilistic expression of results. An example of a deterministic expression is the following:

"In 2013 City X will experience seven PM_{10} exceedences of 50 $\mu g/m^3$, and in 2016 it will not be compliant with the NES limit of three exceedences",

The equivalent probabilistic expression is as follows:

"In 2013 City X will experience between five and eight PM_{10} exceedences (95% confidence interval), and by 2016 it has a 30% chance of being compliant with the NES limit of three exceedences".

[1] This report is subject to the limitations listed in Appendix A.

(Statements about 'City X' as a whole may be refined to give a more detailed spatial distribution of the likelihood of compliance (see Colvile et al. 2002)). Presentation of results in a probabilistic manner can increase confidence in decision-making, and provides guidance on the need for remedial action. That is, if the confidence interval lies entirely in exceedence of the Standard, action is required; if the confidence interval lies entirely below the Standard, no action is required; if the confidence interval straddles the Standard, carry out further investigations to increase certainty. In spite of results being expressed in an apparently *less* certain manner (they include a confidence interval where previously they did not), regulators can be given *more* certainty.

1.3 Structure of the Report

As mentioned above, predictions of compliance likelihood and questions on distribution of high-end concentrations are statistical in nature. They require examination of sensitivities and uncertainties in all aspects of the modelling process, from input data, to model formulation, to data used for validation. To fulfil the aims of this project, Sections 2.0 and 3.0 discuss the issue of sensitivity to changes in model inputs, and outline some methods for determining the model output uncertainty. Such methods have the potential to provide an insight into the behaviour of the modelling system, and to provide a firmer basis – through increased understanding of the limitations of the model results – for air quality decision making.

Section 4.0 extends the examination of uncertainty to exceedences and extreme concentrations. It also briefly discusses methods for improving model performance (and optimizing model inputs). Section 5.0 provides a plan for investigating New Zealand-specific cases and makes some recommendations on assessing uncertainty in New Zealand's regulatory setting. Some concluding remarks are provided in Section 6.0. A list of the papers reviewed is given in Section 7.0.

2.0 SENSITIVITY ANALYSIS

Model sensitivity can be defined simply as the rate of change of outputs with respect to inputs. For example, it might be the percentage change of peak modelled PM_{10} in suburbs P, Q and R for each percentage change in domestic-fire emission rates in those suburbs. This is not to determine ambient PM_{10} changes under different emissions scenarios, but to eventually determine how robust the modelled glcs are to uncertainties in emission rates. As a simple example:

(a) Given the calculated model sensitivity is 0.5 % change in peak PM_{10} per 1 % change in emissions, and
(b) Given that emissions in suburb P are certain to within 10%, then
(c) The predicted peak PM_{10} in suburb P can be said to be certain to within 5%.

In this example, (a) is the calculation of model sensitivity, and the information from (b) allows a basic assessment of the model uncertainty in (c). A similar analysis can be carried out for less direct measures of pollution (such as ozone or secondary-particulate concentrations) and for model-configuration parameters in general.

The rate of change of modelled concentration with respect to specific input parameters mentioned in (a) may be calculated by simply re-running the model which those parameters independently varied, or 'perturbed'. This would usually have to be done through multiple model runs, as the dependence of output glcs on input parameters can be complex. A traditional example would be the dependence of output ozone concentration on its chemical precursors oxides of nitrogen and volatile organic gases.

Sensitivity analyses are more widely carried out than uncertainty analyses, and can be used to rank the influence of parameters in terms of the model sensitivity to them. Categories of sensitivity analysis can be defined as follows (Borrego et al. 2008):

1) **Source-based**, where there are many output variables, and few input parameters. Sensitivity of the output variables can be determined by small set of model runs in which the input parameters are varied. The method is known as 'direct', or 'tangent'.
2) **Receptor-based**, where there are few output variables, many input parameters. Inverse modelling techniques (such as 'adjoint' methods (Schlunzen and Sokhi 2008))) are more efficient than direct methods in this case, but the same measures of sensitivity are calculated as for the source-based approach.

The sensitivity analysis is usually centred on a 'base case' with variations on this configuration being tested. Also, it is usually a linear analysis, extrapolating outputs along straight lines through the base case and the variations. This is appropriate to small parameter variations from the base case. Non-linear analyses are possible, so that sensitivity to a slightly range of perturbations can be determined (Yegnan et al. 2002; Williamson et al. 2004).

3.0 UNCERTAINTY ANALYSIS

3.1 Introduction

Uncertainty analysis may be carried out to provide confidence intervals around model outputs, based on the level of certainty associated with components of the modelling process (including the model itself). This work focuses on quantifying the amount of uncertainty numerically; verbal descriptions may use either 'certainty' or 'uncertainty' to refer to essentially the same thing. A 95% confidence interval means we are 95% *certain* that the results lie within that interval; we are, however, *uncertain* as to where the results lie exactly.

Uncertainty analysis is an extension on sensitivity analysis in two main respects. Firstly, it is not restricted to small perturbations about a base case. Secondly, the likely range of inputs is accounted for, and a likely range of outputs is calculated statistically. For example, if confidence intervals for each input have been estimated, the modelled concentrations can also be given a confidence interval. A full uncertainty analysis is a generalization of the example given in Section 2.0.

In broad terms, 'uncertainty' in model predictions arises from incomplete knowledge or shortage of information. This may relate to the models, their formulation, their representations of complex physical processes, model parameters, constants and associated data sources (Denby et al. 2010; Koracin et al. 2007; Rao 2005; Sax and Isakov 2003). An uncertainty analysis attempts to quantify the sources of uncertainty and their effects on the modelling outcomes.

Uncertainty analysis appears to not be widely carried out for regulatory purposes. However, there is much in the scientific literature on it, dating back at least thirty years (see Hanna 1977). There appears not to be any standardized approach promoted by regulatory bodies, but they provide directives and guidance nonetheless. For instance, the European Harmonization Initiative has provided guidance on uncertainty analysis in response to European Union (EU) directives, as has the Forum for Air Quality Modelling in Europe (FAIRMODE), the Joint Research Centre of the European Commission and the Air4EU programme (see Schlunzen & Sokhi 2008; Miranda et al. 2008; Denby et al. 2010; Flemming & Stern 2007). The USEPA atmospheric modelling and analysis division carries out some uncertainty analysis, although it is not promoted in a regulatory sense. However, the USEPA support centre for air quality modelling does supply the dispersion model SCIPUFF in addition to the usual regulatory models, which simulates concentration variability in addition to 'ensemble-mean' concentrations (see Weil et al. 1992). The MfE modelling good-practice guide (MfE 2004a) acknowledges the importance of uncertainty analysis, but provides no specific advice. All in all, regulations and guidance appear to be sketchy, aside from those promoted in the EU. Even in the EU, there are directives, but there is little guidance. Further work following from that presented in this report should be aimed at providing guidance to regulators and modellers in New Zealand.

3.2 Sources of Uncertainty

As mentioned already, there are many sources of uncertainty. Examples are shown in Table 1, divided into three main types. The list is synthesized from the work of several authors (Rao 2005; Ching et al. 2006; Dabberdt and Miller 2000; Koracin et al. 2007; Denby et al. 2010; Builtjes and collaborators 2006; Flemming and Stern 2007). Uncertainties or errors in the aspects listed here contribute to uncertainty in model results, and quantifying the relationships between them is a main focus of this report. Later, model performance statistics will be discussed, as they measure the mismatch between model outputs and data used for validation. Even if the total error is small, that may be due to cancellation of errors due to individual sources of uncertainty. Ideally, the individual sources of uncertainty should be quantified; they will then propagate through to a measure of the uncertainty in the model outputs, which will be present even if the model results appear to be in close agreement with observed ambient data.

In Table 1, term 'stochastic' refers not only to true randomness, but also to the essentially deterministic processes which are not resolved by the model or by measurements (for instance, 'sub-grid scale' turbulence and its effect on pollution concentration).

Table 1: Sources of model uncertainty.

Type	Category	Examples
Type 1[&]	Data Sources and Model Inputs	Emissions inventory, projected emissions; model grid structure and resolution, meteorological and ambient pollution data accuracy and representativeness; geographical data; baseline and boundary information; chemistry; model input parameters.
Type 2	Model Formulation	Conceptual, physical, mathematical and numerical formulation of the model – applied to meteorology, dispersion and chemistry, and other model schemes[#].
Type 3	Stochastic Variability	Atmospheric turbulence, effects unresolved at the model's smallest spatial and temporal scales; effects averaged out in the data; unpredictable effects such as human behaviour.

[&]The uncertainty 'types' have been defined for the purpose of this report, and do not necessarily adhere to any accepted convention.

[#]"Schemes" refers to a model's formulation of such things as plume rise, stability, dispersion, chemistry, etc.

Note that this categorization is somewhat subjective – no study addresses every aspect, and some examples may be categorized differently depending on the application in question. It should be noted that although all of the types can be quantified, only those of Type 1 and Type 2 are reducible by improvements in model formulation and execution. Uncertainties of Type 3 cannot be reduced or removed (without changing to a different kind of model). Type 3 is probably a reasonable definition of 'stochastic' variability, and that any irreducible uncertainties in the dispersion modelling would fall into this category.

The next sections discuss the uncertainty types in turn, describing methods for evaluating their effects on model results.

3.3 Uncertainties in Data Sources and Model Inputs (*Type 1*)

3.3.1 Background

This type of uncertainty is probably the most straightforward to address, and a significant amount of effort has been devoted to it in the scientific literature. Following Rao (2005), for example, there are two steps to the evaluation of Type 1 uncertainties:

Step 1 – assign probability distributions to model inputs, incorporating uncertainties in parameters, measurement errors, and accounting for correlations among parameters. These may be expressed as means and standard deviations.

Step 2 – propagate the probabilities through to produce a probability distribution of model outputs. Again, this may be expressed as means, standard deviations, and therefore confidence intervals in the model predictions.

3.3.2 Step 1 – probability distributions of model inputs

Statistical models for pollution concentration and their application in dispersion modelling were examined in the 1980s by Jakeman and collaborators (Taylor et al. 1986; Jakeman et al. 1986; Jakeman et al. 1988). Pollution concentrations were found to follow known probability distributions, such as exponential, gamma, lognormal and Weibull. Hybrid deterministic/stochastic models were proposed, in which the pollution extremes were modelled as the stochastic component. Although the techniques have not become commonplace, the concept of probability distributions of pollution extremes – indicating some randomness in actual pollution levels measured – was beginning to be appreciated. This work probably helped promote the notion of specifying model inputs in terms of statistical parameters. For example, Hanna et al. (1998) provide some numerical uncertainty ranges for a selection of input variables, with a view to these being put to use by the air quality community. These include emission rates, initial and boundary concentrations, meteorological conditions and chemical reaction rates.

The key papers by Hanna et al. (1998) and Rao (2005) discuss the importance of expert judgement in estimating uncertainty limits for input parameters. The smaller the amount of information available (that is, the lower the certainty regarding input data), the higher the level of expert judgment required. Consultation of several experts may be needed, whose opinions may differ wildly. Rao (2005) discusses at length expert judgment and its use in developing reliable confidence intervals for model inputs.

3.3.3 Step 2 – propagation of input uncertainties to output confidence intervals

Several techniques have been applied to the calculation of model output confidence intervals, based on given uncertainties in input data and other parameters. They may essentially be thought of as either ‘analytical’ or ‘numerical’.

For simple models (for instance, Gaussian plume dispersion models) with a small number of parameters and relatively simple relationships between inputs and outputs, analytical techniques may use input uncertainties to provide confidence intervals around model outputs. For Gaussian models, Rao (2005) provides examples of analysis deriving output means and confidence intervals from given means and standard deviations of input parameters.

For many dispersion models – in which concentrations are not given by the simple Gaussian formula – numerical techniques must be used, in which the model is run many times, varying the model configuration each time. Known as Monte Carlo methods, the model is re-run with changed input parameters, which are chosen to be within their pre-determined ranges of uncertainty. Examples abound in the literature, and Monte Carlo methods have been applied to a range of dispersion problems from point-source emissions to traffic in street canyons to whole urban airsheds (Irwin et al. 1987; Tatang et al. 1997; Hanna et al. 1998; Bergin et al. 1999; Bergin and Milford 2000; Dabberdt and Miller 2000; Hanna et al. 2001; Vardoulakis et al. 2002; Sax and Isakov 2003; Rodriguez and Dabdub 2003; Rao 2005; Rodriguez et al. 2007; Hanna et al. 2007; Miranda et al. 2008).

Monte Carlo methods are theoretically straightforward to apply. However, the number of runs which may be carried out is limited by available computational resources, and also the model case study length. The literature cited above employs between a few tens and three thousand ensemble model runs. Techniques employed to choose a few representative parameter combinations from ranges in each of several parameters include Latin hypercube sampling (Rodriguez et al. 2007). In this, each parameter is divided into

a range of intervals and a parameter from each interval is used exactly once in the ensemble of runs[2]. In addition, the probability distributions (or means and confidence intervals) of the input parameters need to be accounted for, so that the range of model outputs obtained is also given confidence intervals. This aspect is addressed in Latin hypercube sampling by dividing the input parameters into ranges of equal probability. Alternatively, input parameters values may be chosen at the known confidence limits of the parameter (Dabberdt & Miller 2000).

[Aside: a Monte Carlo simulation can be used to fit a polynomial response surface of model outputs to a range of input parameters, such that outputs can be generated from further combinations of input parameters without re-running the original dispersion model (Tatang et al. 1997; Hanna et al. 1998; Hanna et al. 2001). However, this has limited use in uncertainty analysis, and should not be used to extrapolate far from the ranges of parameters used in its determination.]

3.4 Analysis of Uncertainties due to Model Formulation (*Type 2*)

Uncertainties in the formulation of dynamical, physical or chemical processes in dispersion models and their effects on model results have not been addressed in the literature as often as uncertainties in data and other inputs. There are several possible reasons for this; probabilities and confidence intervals associated with model schemes are nearly impossible to judge, accounting for uncertainties in certain of the model schemes may not properly account for all model errors and biases, or it may be difficult to separate the effects of input uncertainties (Type 1). Note that implementation of Monte Carlo methods mentioned above does not address uncertainties in the models themselves.

Work in the literature addresses uncertainties in model formulation by either carrying out multiple runs of the same model, but with different physical or chemical parameterizations (Dabberdt and Miller 2000), or running an ensemble of different models each configured the same way (Galmarini et al. 2004 (two papers); Vautard et al. 2009). In either case, variability in model outputs gives a measure of dispersion model uncertainty (for fixed input choices).

Running ensembles of models does not provide a measure of uncertainty in the formulation of an individual model, and in most applications, dispersion modelling is carried out using the single model most fit for the particular purpose. However, the 'total' uncertainty may be estimated through comparison of model outputs with monitoring data, and statistical techniques may be employed to estimate the model uncertainty as a residual error from the total uncertainty and the other Types (1 and 3) of uncertainties (for example, Koracin et al. 2007). This is discussed further in Section 3.6.

3.5 Analysis of Uncertainties due to Stochastic Variability (*Type 3*)

Stochastic variability occurs on shorter spatial and temporal scales than can be resolved either by the model or by observations. 'Stochastic' essentially means random, and although the variability is governed by deterministic laws of physics, it *appears* to be random and is amenable to statistical techniques. Essentially, this variability in either meteorology or pollution concentrations is due to fluctuations in wind velocity due to atmospheric turbulence. Examples include fluctuations which manifest as plume diffusion (in the so-called 'inertial sub-range' of scales), or plume meandering due to larger eddies (if they are unresolved by the meteorological model). Observed fluctuations are seen to be larger as the averaging time gets shorter, and so their effects must be parameterized if longer averaging periods are used. For a meandering plume, the standard deviation of pollutant concentration may equal or exceed the mean concentration, when the averaging time is greater than the meander time-scale.

[2] For example, if there are two parameters whose ranges are divided into ten intervals, there are potentially one hundred combinations of input parameters. If the two sets of ranges are displayed as 10x10 matrix, a Latin square marks only ten combinations, with each row and column containing exactly one sample.

It is useful to define the ensemble mean and the ensemble standard deviation of the pollutant concentration as taken from the average over an ensemble of possible concentrations which depend on small-scale unresolved properties. This is useful for modelling, in that models are designed to simulate ensemble-averaged quantities. Then, the ensemble standard deviation of the pollutant concentration is a representation of the pollutant's 'stochastic' variability. Nearly all dispersion models simulate the ensemble-mean concentration, and parameterize the variability using flux-gradient diffusion relationships or apply dispersion curves. These may be referred to as 'first-order' closure models. Only one model in relatively common use (SCIPUFF) explicitly simulates the concentration standard deviation as a time-dependent variable (second-order closure). The review of Weil et al. (1992) discusses natural variability and its simulation in dispersion models. Irwin et al. (2007) examined in detail the relationship between observed and modelled variability using tracer experiment data and a Monte Carlo trajectory model. They found that variability in plume trajectory can be larger than the plume width, and plume centre line concentrations varied by a factor of two among the ensemble of model runs.

3.6 Analysis of Total Uncertainty (*All Uncertainty Types*)

This section discusses the combined effects of the three uncertainty types. Modellers generally assess model performance by comparing model results with observations, which implicitly includes all of the sources of uncertainty. The following relation shows qualitatively how the various sources of uncertainty are linked (based on Rao (2005)):

$$\left[\left(C_o - C_p\right)^2\right] = \left[\left(\delta C_p\right)^2\right] + \left[(\delta C_o)^2\right] + \left[d^2\right] \quad + \quad \sigma_c^2 \qquad (1)$$

Total ***Term (a)*** ***Term (b)*** ***Term (c)*** ***Term (d)***

Square brackets denote ensemble means of the quantities shown. The concentrations are 'individual' observations or model predictions, and are time and location dependent. The stochastic uncertainty, σ_c^2, is the ensemble variance of the observations. Equation (1) (which is Rao's Equation (3)) links the sources of uncertainty in observations and model predictions. The total uncertainty is the sum of the uncertainty types already mentioned, where according to the definitions used in Table 1, Terms (a) and (b) are defined above as Type 1, Term (c) is Type 2, and Term (d) is Type 3.

The terms in Equation (1) and their definitions are explained in Table 2.

Table 2: Components of total uncertainty.

Uncertainty type	Equation (1) term	Formula (or part thereof)	Definition
1	(a)	$\delta C_p = C_p - C_{pa}$	Model error due to input data errors
		C_p	Modelled concentration
		C_{pa}	Modelled concentration (in the absence of input data errors)
1	(b)	$\delta C_o = C_o - C_{oa}$	Measurement error
		C_o	Observed concentration
		C_{oa}	True concentration (in the absence of measurement error)
2	(c)	$d = (C_{oa} - C_{pa})$	Model bias
3	(d)	σ_c^2	Stochastic Variability

Note that terms $\mathbf{C_{oa}}$ and $\mathbf{C_{pa}}$ are what would be observed or modelled if there were no errors. These can never be known for a single observation or model realisation, and estimates made of these indirectly through $\mathbf{\delta C_o}$ and $\mathbf{\delta C_p}$. The model bias **d** in the absence of other errors must also be estimated.

The observed quantities may be thought of as single realisations from an ensemble of possible concentrations that may have occurred under the conditions as far as they were known, and their ensemble mean taken from those single realisations. However, as there is only one possible model realisation for each set of input conditions, there is no ensemble mean. It is appropriate to think in terms of an ensemble mean of $\mathbf{C_o}$, but not $\mathbf{C_p}$.

In theory, a minimum in total uncertainty should occur at a particular level of model complexity, with a higher total uncertainty both higher *and* lower levels of model complexity. It is neither the most complex model, nor the least complex, which produces the most certain results (Rao (2005)). Increasing the number of meteorological parameters decreases Terms (c) and (d), but increases Terms (a) and (b). A more complex model has the potential to perform better and leave less of the variability unresolved. However, it will usually have larger data requirements and these have the potential to introduce more uncertainty.

Equation (1) refers to concentrations at a specific time and location, and partitions the squared difference between modelled and observed glcs. Averaging this over a whole time series at a point and taking the square root leads to an equation which partitions the root-mean-square error of the model. The same may be done for other model-performance measures such as biases, correlations and the index of agreement (though their formulas may be complex). Hence the following is true:

(i) Relationships similar to equation (1) may be constructed which partition model-performance statistics into different components of uncertainty (or 'errors').

(ii) The total uncertainty, as measured by model-performance statistics, can be examined and attributed to different sources of error (this is a re-wording of (i)).

(iii) The existence of such a relationship means that not every component needs to be calculated or estimated explicitly. One component (but no more than one) can be estimated as a residual of the others. This is done in the example shown in Section 3.6.2.

The next sections discuss model-performance statistics in general, and show an example of how the total uncertainty may be broken down into components.

3.6.1 Model performance statistics

There are many model performance statistics which compare model outputs with observations, to give a measure of the total uncertainty. Rao (2005) considers the fractional bias (FB), mean model bias ([d]), normalized mean square error (NMSE) and figure of merit in space (FMS).

Borrego et al. (2008) provide a list of statistics (see their Table 1), pointing out that most have problems with undue influence from extremes. For instance, FB and NMSE are influenced by high concentrations. Also, the geometric-mean bias (MG) and geometric variance (VG) are unduly influenced by near-zero concentrations. According to these authors, the most 'useful' parameters are the index of agreement (IOA) and correlation coefficient, the root mean square error (RMSE) and NMSE, and the FB.

Chang and Hanna (2004) give a list of suitable model-performance measures, including typical magnitudes which would represent good model performance, based on other studies (see their Section 2.10). For instance, they conclude that good model performance is indicated if (i) at least 50% of model predictions are within a factor of 2 of observations, (ii) the mean bias is within 30% of the mean, and random scatter[3] is within a factor of two to three of the mean. The authors also present bootstrapping techniques to produce confidence intervals for the model performance measures.

[3] This is Chang and Hanna's phrase, but is not explicitly defined by them.

There are several other 'standard' papers on air quality model performance statistics (for example Willmott 1981; Willmott 1982; Cox and Tikvart 1990).

3.6.2 Partitioning total model uncertainty (an example)

Koracin et al. (2007) present a statistical approach for partitioning total model uncertainty into its separate components due to the models used and to various sources of input information. This is based on an assessment of magnitudes of individual error components. It calculates statistical measures of the individual error components and combines them together. The approach uses assumptions regarding the probability distributions of the error components so that they may be combined mathematically. (As to whether the approach is labelled 'analytical' or 'numerical' as defined in Section 3.3.3, it has elements of both).

The measure of total model uncertainty used by Koracin et al. (2007) is the fractional error, as defined in Equation (2).

$$E_t = \frac{(C_p - C_o)}{1/2\,(C_p + C_o)} \qquad (2)$$

E_t is partitioned similarly to the total uncertainty in Equation (1), except that the meteorological and dispersion model uncertainty components are treated separately. Each uncertainty component is also represented by a fractional error. The fractional errors are estimated individually and then combined into the total uncertainty, using statistical techniques such as maximum likelihood estimate. The total uncertainty is a thus a function of the individual components, expressed in a general form in Equation (3).

$$E_t = f(E_e, E_m, E_a, E_d, E_s) \qquad (3)$$

Note that in contrast to the mean squared error components defined in Equation (1), the components of E_t can be positive or negative, and in combination they could cancel each other out and lead to E_t being zero. In this case the modelled concentrations may match observed concentrations, but the analysis shows that there are still errors or uncertainties. Descriptions of the individual terms are given in Table 3.

Table 3: Sources of uncertainty in the Koracin et al. (2007) model.

Symbol	Name	Description
E_e	Emission errors.	Derived from variability in source emission rate.
E_m	Measurement errors.	Derived from uncertainties in air pollution measurements; includes fluctuations in background concentrations.
E_a	Atmospheric model errors.	Depends on meteorological model parameter uncertainties and model performance.
E_d	Dispersion model errors.	Depends on dispersion model parameter uncertainties and model performance.
E_s	Stochastic component.	Due to unresolved effects of atmospheric turbulence.

Koracin et al. (2007) provide a theoretical description of the stochastic model, describing how to estimate the errors listed here in terms of their likely statistical distributions. Assuming an underlying statistical distribution, such as a normal distribution, allows estimation of the basic parameters for that distribution, namely, its mean and standard deviation. Having calculated these parameters for each error component, they can then be combined. The methods were applied to an atmospheric tracer experiment lasting several days and its simulation using a model. In this case, all terms in Equation (3) were estimated except $\mathbf{E_d}$, which was calculated as a residual term. The authors found the largest uncertainty in the stochastic component of the modelling, through the testing of several atmospheric turbulence schemes.

Whilst the details of how the estimates of uncertainty were calculated are not given by Koracin et al. (2007), the paper demonstrates a practical example of how the partitioning of total model error could be carried out. Partitioning the total error into its components – as a logical extension of the presentation of model performance statistics – has a number of benefits, as follows:

a) The most important contributions to uncertainty in the model results can be identified. These may not be due to the dispersion model itself, but could be measurement error or emissions variability, for example.

b) The largest components on the total error can be targeted for reduction of their uncertainties to obtain the best reduction in total error.

c) Obtaining the best performance from the modelling system should lead to more confidence in the model as a predictive tool for alterative scenarios where there are no data.

We would therefore suggest that the error partitioning is a worthwhile exercise. Work following on from this project will be aimed at testing some approaches and providing practical advice to modellers and regulators on how to incorporate uncertainty analysis into air quality policy decisions.

3.7 Reducing Uncertainties – Inverse Modelling

This section outlines a type of approach which has the potential improve model predictions through the reduction in magnitude of the components of uncertainty.

The incorporation of a statistical component in dispersion modelling leads to results whose uncertainty is bounded by confidence intervals. The confidence intervals arise through the propagation of model-input uncertainties through the modelling process. The input uncertainties are quantified in advance of dispersion modelling through estimates and assumptions using expert judgement or other standard procedures. The modelling proceeds in a 'forwards' direction, from input data, through the model, and to model results and their comparison with observations.

As an extension of forward modelling techniques, inverse modelling combines the various sources of input uncertainty with prior knowledge of expected outputs (along with *their* uncertainties) *and reduces all of the uncertainties*. For instance, emissions and their uncertainties (expressed as standard deviations) may be input to a dispersion model which outputs concentrations and their associated standard deviations. Subsequently, a comparison with *observed* concentrations (and their standard deviations) can be used to improve the input emissions estimates and reduce *their* uncertainties. Essentially, a 'Bayesian' statistical approach is used to optimize emissions, and the original concentration predictions from the forward model are improved upon (Bayes 1763).

The extended procedure is referred to as 'inverse modelling' in the sense that concentrations are used to update emissions estimates (or infer them from scratch if there were no prior information). It may involve running models backwards in time, if this is more computationally efficient[4].

[4] This is equivalent to the choice between source-based and receptor-based modelling mentioned in Section 2.0, and depends on the relative numbers of emission source parameters and concentration data.

A successful application of inverse modelling techniques should reduce uncertainties in the emissions estimates *and* predicted concentrations. This occurs through the updated emissions leading to new modelled concentrations being closer to those observed. The potential to reduce all of the uncertainties arises because the model is being used to relate all of the variables. More certainty arises as use is made of the extra knowledge that the variables are indeed related, when the prior estimates of emissions and the observed concentrations do not account for this.

The review paper by Rao (2007) discusses several inverse-modelling techniques, including the use of back trajectories, adjoint methods, tangent models, the Kalman filter and variational data assimilation. The Kalman filter was used by Gilliland et al. (2010) to model sources of ammonia in continental USA, and by Jorquera and Castro (2010) for urban pollution episodes. However, Rao (2007) warns against the application of Kalman filter techniques to systems which are not linear. Bayesian techniques (Tarantola 1987) to estimate source strengths were applied by Gimson and Uliasz (2003) for agricultural methane emissions. Bayesian Monte Carlo methods were used by Bergin and Milford (2000). In general, outputs need not simply be modelled pollutant glcs, but may be derived quantities, such as the percentage change in ozone given percentage change in NO_x and VOC emissions.

The following points should be noted regarding inverse modelling:

1) The technique does not simply change emissions so that the model closely reproduces observed concentrations. Rather, it accounts for quantified uncertainties in emissions, model performance and ambient concentrations to produce a ‘best-fit’ solution (with associated confidence intervals) through all of the variables.

2) Independent estimates of variability in emissions – due to human behaviour, for example – could be incorporated into the analysis to improve the modelling of extreme concentrations under unusual emissions.

3) Other factors leading to high concentrations, such as extreme meteorological conditions, may also be accounted for in the analysis.

In addition to being a useful tool for reducing uncertainties in general, inverse modelling has the potential to improve the modelling of high-end concentrations and exceedences of target concentrations (though this has not been tested yet). The optimization of emissions could lead to their adjustment from average emission levels to worst-case emission levels, and lead to an improvement in modelled concentrations specifically at the high end. Modelling of high end concentrations is discussed in the Section 4.0.

3.8 Summary

This section of the report has focused on dispersion modelling uncertainty, describing its various sources, outlining some methods for quantifying the various components of the total uncertainty, and describing approaches which may potentially reduce the uncertainties and improve model predictions of ground-level concentrations.

Two key extensions of the techniques already mentioned, which are relevant to management of NES exceedences in New Zealand are discussed in the following sections of this report.

The first extension is the application of uncertainty analysis to model performance statistics relevant to high-end concentrations, including target exceedences. These performance statistics are not often encountered in New Zealand.

The second extension is the application of methods to measures of high-end concentration, such as maximum concentration or probability of exceedence. These are not performance statistics – which quantify errors and biases – as they relate to predictions under scenarios for which there are no observational data. However, an uncertainty analysis of them may still be carried out.

The aim here is to show how to give confidence intervals to measures of predicted high-end-concentration, which are partitioned according to uncertainties in the modelling process.

4.0 EXCEEDENCES AND EXTREME CONCENTRATIONS

4.1 Introduction

The discussion so far on model sensitivity and uncertainty has not specifically considered extreme concentrations or other measures such as the number of exceedences of air quality targets. Model performance indicators mentioned above – such as fractional bias or index of agreement – relate to the full range of conditions which are modelled and measured. The models are deterministic, in that the modelled concentration at one instant in time is determined with complete certainty from the modelled concentration at the previous instant in time. However, as pointed out above, there is a statistical, apparently random component to the pollution concentration, more apparent at the higher concentrations that occur rarely.

This section focuses on the modelling of extreme concentrations, by introducing specific model performance indicators used for high-end concentrations, and by discussing the stochastic nature of extreme concentrations.

4.2 Model Performance Indicators for Extreme Concentrations

Borrego et al. (2008) list some model performance indicators for extreme concentrations, based on USEPA and EU guidance. Like those indicators mentioned in Section 3.6.1, these also apply to the total model uncertainty, not individual types of uncertainty. Examples of model statistics for extremes are listed in Table 4. The USEPA indicators were developed for photochemical ozone modelling, biased on a criterion 1-hour ambient concentration of 60 ppb, and their descriptions are taken verbatim from USEPA (1996). They may be easily adapted to other air pollutants of relevance in New Zealand. The relative percentile error (RPE) should not be sensitive to outliers, and the selected percentile corresponds to a specified number of allowable exceedences. Hence this measure is highly applicable to the NES for PM_{10}.

The POE is obtained from the probability distribution of concentrations, and can be calculated from model results and presented as a spatial map (Walker 2007). The POE is not a performance measure, but a useful statistic which can be derived from model results.

The model performance indicators discussed in this section specifically apply to high-end concentrations. Quantifying model performance in simulating such concentrations is a step towards the analysis and partitioning of uncertainties, and the provision of confidence intervals for the expected number of exceedences of air quality targets such as the NES for PM_{10}.

The absolute value of an expression is indicated by vertical bars, thus distinguishing $NERROR_{60}$ from $NBIAS_{60}$. Also, these measures can be used for the 24-hour average PM_{10}, calculating the error and bias over occasions when the observed concentration exceeds 50 $\mu g/m^3$.

Table 4: Model statistics for high-end concentrations. (All of these, except the POE, are measures of model performance relative to observations).

Description	Source of information	Symbol	Formula	Indicative range of acceptable performance
Normalized accuracy of domain-wide maximum 1-hr concentration unpaired in space and time.	(USEPA 1996)	A_u	$\mathbf{(Pred_{peak} - Obs_{peak})/Obs_{peak}}$	+/-15% to 20%
Mean normalized bias of all predicted and observed concentration pairs where the observed concentrations exceed 60 ppb.	(USEPA 1996)	$NBIAS_{60}$	**Mean (Pred – Obs)/Obs** where Obs > 60ppb. Paired in space and time.	+/-5% to 15%
Mean normalized error of all predicted and observed concentration pairs where the observed concentrations exceed 60 ppb.	(USEPA 1996)	$NERROR_{60}$	**Mean |Pred – Obs|/Obs** where Obs > 60ppb. Paired in space and time.	+/-30% to 35%
Relative percentile error	(Flemming & Stern 2007; Borrego et al. 2008)	RPE	**|Pred – Obs|/Obs**. Unpaired, averaged over receptors for selected percentiles.	Not specified
Probability of exceedence	(Walker 2007)	POE	Integral of probability density upwards from the criterion value.	Not specified

4.3 Stochastic Approach for Extreme Concentrations

The notion of apparent randomness in pollution concentrations and the necessity of a statistical approach to their analysis and modelling were introduced in Section 3.2 of this report. The findings that pollution concentrations followed known probability distributions were mentioned in passing in Section 3.3.2. As also mentioned in Section 3.3.2, hybrid deterministic/stochastic models were proposed in the mid-1980s (Taylor et al. 1986; Jakeman et al. 1986; Jakeman et al. 1988). At the same time, Rao et al. (1985) proposed methods for determining the statistical parameters for extreme modelled and observed concentrations, and using these to assess model performance.

The work by Rao et al. (1985) should be considered a fundamental paper on extreme concentrations, being the first to apply the statistical theory to the analysis of air quality modelling results. The concepts presented are still relevant, and the research recommended in this report should pay close attention to them.

The key messages which may be gleaned from the paper by Rao et al. (1985) are listed as follows:

a) The tail of a distribution of concentrations (observed or modelled) can follow an exponential-like distribution. Several possible distributions are mentioned in this paper and elsewhere in the literature.

b) The extreme concentrations – for instance, the highest, m^{th}-highest for small m – then follow a specific exponential distribution known as a Gumbel distribution. This is true no matter what the original distribution was, provided it was an exponential type. The shape and location parameters of the Gumbel distribution can be found by several methods, including maximum likelihood estimation and least-squares techniques. This is the essence of the 'exact' EVT.

c) Other measures of the extreme concentration – for instance, the maximum hourly concentration for each day – also follow a Gumbel distribution. However, this is only true in the limit of large sample sizes, where the sample in question is a fixed-period observational data set. This is the essence of the 'asymptotic' EVT.

d) From observational data, the maximum concentration each day can be calculated, followed by the mean and median of those maxima (over a long time series). The mean and median are measures of model simulation of extreme concentrations. A probability distribution can be determined for the observed maximum using bootstrap re-sampling, leading to calculated confidence intervals for the observed maximum. Model performance can then be considered satisfactory if the model's maximum concentration is within the confidence interval for the observations.

e) Model performance for other upper-percentile concentrations may be assessed in the same way as outlined in (d), by comparing the modelled percentile with the confidence interval for the observed percentile.

Further points are worth noting from the above, as follows:

- The EVT fits a distribution to the tail end of the concentration distribution, and gives the probability of occurrence of *any* high value, *even if it did not occur in the original data sample.* It indicates that *higher concentrations than those observed so far may still occur in the future* and gives their likelihood, or frequency. Statistical parameters must be found in this case using the EVT, not using bootstrapping, as the re-sampling cannot reproduce the extremes in the original data.

- When the EVT does not apply – for example, to derived parameters such as the daily maximum – bootstrap re-sampling *must* be used to determine confidence intervals on the mean and median of such parameters (but not their extremes). The re-sampling has the advantage that no assumptions about any underlying statistical distribution are needed.

- The re-sampling procedure in (d) and (e) calculates confidence intervals for the daily maximum for *observations only*. This is because the observed concentrations are a single realisation of a set of possible concentrations that may have occurred under the same external conditions (of emissions and meteorology). The measurements *could* have been different under the same conditions. However, model results represent an ensemble-mean of those possible realisations over the observation period, and it is inappropriate to present them in terms of other possible realisations. Hence, a single set of model results for upper-percentile concentrations should be compared with observations and their confidence intervals for those percentiles.

- There are some restrictions on the data to which the EVT is applied. They should be independent and identically distributed, and tests should be carried out to ensure that these conditions are satisfied.

- The re-sampling analysis may be carried out for model performance measures pertaining to upper percentiles, as well as the upper percentiles themselves.

Other authors have more recently applied the ideas of Rao et al. (1985) to observed concentration exceedences in urban areas (Lu and Fang 2003; Sharma et al. 1999). In essence, they use historical data sets (and hence knowledge of how many exceedences there have been in the past) and statistical techniques to determine the expected number of exceedences in any year.

The stochastic approach outlined here thus changes the nature of the model-predicted high-end concentrations, allowing model predictions to be extended to a probability distribution for the upper-percentile concentrations. For example, rather than saying the maximum predicted concentration is 45 µg/m^3, it might be found that the concentrations follow a statistical distribution whose 95th percentile is 45 µg/m^3, and whose 99th percentile is 50 µg/m^3. This would mean a POE of 50 µg/m^3 of 1%, although the modelled maximum is only 45 µg/m^3. If the concentrations were 24-hour averages, this would mean possibly 3 or 4 exceedences per year are predicted in general, even though there may be fewer occurring in the year actually modelled.

The above bulleted points follow from the work of Rao et al. (1985). Several more general points are worth noting:

- The approach as outlined by Rao et al. (1985) essentially accounts for statistical variability, or uncertainties of Type 3.
- The statistical approach does not excuse poor model performance. The example given shows a small POE when the deterministic part of the model shows no exceedences. If there were actually a high number of exceedences observed, and/or model predictions of far fewer (even using EVT), then model performance should be investigated.
- The approach does not incorporate (i) input data uncertainties or (ii) model formulation issues. Addressing these Type 1 and Type 2 uncertainties (following approaches outlined in previous sections) could increase the predicted POE. This would occur through, say, the possibility of unusual emission conditions outside of those presented in inventory inputs.
- Arguably, obtaining the best model performance (that is, reducing Type 2 uncertainties, especially biases) should take precedence over the consideration of Type 1 and Type 3 uncertainties.

This section has outlined in general terms some of the principles of a stochastic approach to modelling high-end concentrations, POE and the number of exceedences. The approach may appear to be quite complex to an air quality scientist with little expertise in statistics; the application of statistical principles to issues in air quality requires either collaboration between air quality specialists and statisticians, or significant expertise in both fields. The work of Rao et al. (1985) is now close to three decades old and the literature search carried out for the current project has not yielded a significant body of work following on from it.

However, the current application of air quality targets worldwide to measures of maximum permissible air pollution levels needs to pay some attention to the apparent random nature of air pollution, which is particularly manifest at the upper end of the ranges of observed concentrations which are observed. This amounts to accounting for uncertainty and providing regulatory bodies with confidence intervals for peak air pollution levels. As pointed out in the above, regulatory bodies produce guidance which recommends that uncertainty analysis is carried out, although they rarely give specific guidance. The regulatory environment in New Zealand is no exception to this: the NES for air quality and the ambient air quality guidelines are based on peak concentrations, and there is a sentiment among some Regional Councils that uncertainties should be accounted for and that research is required on methods to do this.

5.0 DISCUSSION AND RECOMMENDATIONS

5.1 Summary of Literature Review

This report has surveyed the scientific literature on the incorporation of uncertainty into dispersion model predictions. This is a necessary component of on-going work to improve the modelling of exceedences of air quality targets. Reasons are given as follows:

i) Exceedences are in the upper tail of a statistical distribution of concentrations, necessitating a statistical approach.

ii) Exceedences may occur under unusual conditions, such as higher than average emissions, or specific meteorological situations which are difficult to model.

iii) The variation of emissions about the average is known statistically at best, as they are a reflection of human behaviour. The variation can be incorporated into modelling results through an uncertainty analysis.

iv) There are many sources of uncertainty, which should be accounted for quantitatively as part of the decision-making purposes, in a statistically robust manner.

The above sections of this report deal with some aspects of uncertainty analysis, as addressed in the literature. The papers reviewed deal with some aspects of uncertainty analysis, but not all. For instance, there appears to be no theory at present which unifies uncertainty apportionment and the stochastic techniques which are applied to extreme concentrations, which would be important for assessing sources of uncertainty in conditions under which air quality targets may be breached. This may be because such a theory would be intractable – that is yet to be investigated.

To meet the aims of this project a complete framework for modelling the POE would incorporate sensitivity analysis, then an analysis of uncertainty of basic types, further broken down into specific sources of uncertainty, and applied to extreme concentrations. In tandem with this would be an effort to reduce those uncertainties through inverse modelling or other optimization processes, which would be applied to the model and associated data.

5.2 Aspects to Address in Current and Future Work

It is generally found that dispersion models which are purely deterministic do not perform well for extreme concentrations and guideline exceedences. The models themselves are tuned for average or ‘typical’ conditions, and their inputs – such as emissions or modelled meteorology – are not representative of extreme conditions. There are several ways to improve this situation, discussed in this report. A suitable approach may be summarized as follows:

i) Examine model-performance measures for averages and extremes, to determine whether the model is performing well for these conditions.

ii) Incorporate uncertainties in model formulation and data sources, account for turbulent variability by modelling concentration fluctuations explicitly and present outputs as probabilities and confidence intervals, rather than a single answer which contains no indication of its certainty.

iii) Incorporate a statistical representation of high-end concentrations into a deterministic model. This may employ advanced techniques, such as extreme value theory.

iv) Apply an inverse modelling approach to the confidence intervals determined in (ii) to optimize model configuration and improve overall performance.

Items (ii) and (iii) address separate modelling issues, namely, uncertainties and the representation of extreme concentrations. The key challenge is to develop a framework to link these, quantifying uncertainty under average conditions, extending the concept to extreme conditions, and improving model performance at the tail end of the concentration distribution. Having developed the framework, it will be important to provide practical implementation guidelines for regulators and the modelling community in New Zealand.

Section 5.2.1 suggests some lines of investigation following from this literature review.

5.2.1 Improving the assessment and prediction of exceedences of the National Environmental Standard (NES) for PM_{10}

Several avenues of enquiry have been identified in this review. This section contains a methodology for investigating, quantifying and improving the modelling of exceedences of the National Environmental Standard (NES) for PM_{10}, by way of a case study of urban airshed modelling. A set of proposed tasks is itemized in Table 5, with cross references to the appropriate sections in this report.

Golder has carried out several investigations in recent years regarding PM_{10} and $PM_{2.5}$ in cities, using urban airshed models (Golder 2009; Golder 2011; Golder 2012a; Golder 2012b). These included examination of population exposure to PM_{10} and $PM_{2.5}$ under current and projected future emissions, hot-spots of pollution, inter-suburb transport, and emissions reductions required to attain the NES (or World Health Organization guidelines for $PM_{2.5}$). Whilst providing useful guidance to air quality managers on the magnitude of high-end concentrations, they do not generally predict the number of exceedences of the NES well. Also, sensitivity of high-end concentrations to the wind speed has been found. The models are based on emissions inventories, which do not incorporate the day-to-day variability that may lead to NES exceedences. It is anticipated that the main focus for improving the modelling of exceedences of the NES for PM_{10} should be on sensitivities and uncertainties around wind speed and emission rates. These aspects provide a useful initial focus for investigations, reflected in Table 5.

Through each of the items in Table 5 it is anticipated that there would be new challenges to overcome in the evaluation of model uncertainty. These include the following:

- The need for expert judgement on the reliability of meteorological and emissions data (item 4).
- The modelling of concentration variance and an understanding of what it represents (item 5).
- The estimation of uncertainties in the model itself (Type 2) – without carrying out an ensemble of runs with different model options, nor running an ensemble of different models. TAPM includes meteorological and dispersion model components, with different sources of uncertainty in each. Type 2 uncertainties may need to be determined as a residual of total model error; Type 1 and Type 3 uncertainties (item 6).
- At the time of writing, an approach to the optimization of the meteorological model has not been developed (item 7).

Items 8 to 12 on extreme value statistics should be straightforward.

Although Items 3 onwards have been carried out to some extent internationally, they would be new to New Zealand. At this stage their outcomes are difficult to predict, and it is likely that further research would be required to address them thoroughly and defensibly as works of science. Once the techniques have been successfully applied in the New Zealand context, it will be important to disseminate results to the air quality community in New Zealand, and provide practical guidance to modellers and regulators for future use (item 13).

Table 5: Improving the modelling of exceedences of the National Environmental Standard (NES) for PM10. Case-study methodology for urban airshed modelling of PM_{10}.

Item	Description	Section
1	Run The Air Pollution Model (TAPM) for Hawke's Bay; winter months; several years; emissions inventory and projections; run model with and without data assimilation. This has been carried out under contract to the Hawke's Bay Regional Council (Golder 2012a).	N/A
2	Calculate model performance measures for 'average' conditions at air quality monitoring sites.	3.6.1
3	Calculate sensitivity to wind speed and emission rates.	2.0
4	Calculate uncertainty of Type 1, due to uncertainties in wind speed and emission rates.	3.3
5	Calculate uncertainty of Type 3, due to turbulent fluctuations.	3.5
6	Calculate uncertainty of Type 2, due to model errors, as a residual from items 2, 4 and 5.	3.4
7	Apply inverse-modelling techniques to optimize meteorology and emissions and reduce uncertainties in concentrations.	3.7
8	Calculate model-performance measures for extreme concentrations.	4.2
9	Use extreme value theory to derive probability distributions for extreme concentrations.	4.3
10	Compare modelled and observed probability distributions for extremes.	4.3
11	Determine models of the highest, 2^{nd}-, 4^{th}-, 11^{th}- highest, etc., 24-hour PM_{10} concentrations (relevant to the NES for PM_{10}), derived from observations.	4.3
12	Assess the fit of model results to the observed distributions.	4.3
13	Provide practical guidance on uncertainty analysis to modellers and regulators.	N/A

6.0 CONCLUSION

This work was carried out under contract to NIWA, as part of the research programme "*Healthy Urban Atmospheres*" (FRST contract C01X0813). The proposed work included a literature review, followed by a focus on the improvement of model predictions of air quality exceedences through investigation into the way emissions information is presented as an inventory.

The literature review has considered a number of general technical issues related to the model prediction of exceedences of the NES for PM_{10}. The issues include model performance measures for extreme concentrations, sources of uncertainty, extreme value statistics, concentration probability distribution and optimization. For each, methods for addressing and quantifying them have been outlined.

Optimization of emissions information, as originally proposed, turns out to be one of many stages in the development of a full uncertainty analysis. A plan has been presented which applies the techniques to a New Zealand-specific practical example. This relates to dispersion of PM_{10} in a city, simulated using an urban airshed model.

The successful application of methods for assessing uncertainty to examples specific to the New Zealand regulatory arena, and the subsequent provision of useful guidance to regulators and modellers alike, will result in significant advancements and improvements in dispersion modelling practices in New Zealand, and provide more certainty to regulators in their air quality decision making.

7.0 REFERENCES

Bayes, T., 1763. An Essay towards Solving a Problem in the Doctrine of Chances. By the Late Rev. Mr. Bayes, F. R. S. Communicated by Mr. Price, in a Letter to John Canton, A. M. F. R. S. *Philosophical Transactions of the Royal Society of London*, 53, pp.370–418. Available at: http://rstl.royalsocietypublishing.org/cgi/doi/10.1098/rstl.1763.0053 [Accessed July 21, 2011].

Bergin, M.S. et al., 1999. Formal Uncertainty Analysis of a Lagrangian Photochemical Air Pollution Model. *Environmental Science & Technology*, 33(7), pp.1116–1126. Available at: http://pubs.acs.org/doi/abs/10.1021/es980749y.

Bergin, M.S. & Milford, J.B., 2000. Application of Bayesian Monte Carlo analysis to a Lagrangian photochemical air quality model. *Atmospheric Environment*, 34(5), pp.781–792. Available at: http://linkinghub.elsevier.com/retrieve/pii/S1352231099003465.

Borrego, C. et al., 2008. Procedures for estimation of modelling uncertainty in air quality assessment. *Environment International*, 34(5), pp.613–620. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18234341 [Accessed August 5, 2011].

Builtjes, P., 2006. Air4EU Project. Available at: http://www.air4eu.nl/reports_products.html.

Chang, J.C. & Hanna, S R, 2004. Air quality model performance evaluation. *Meteorology and Atmospheric Physics*, 87(1-3), pp.167–196. Available at: http://www.springerlink.com/index/10.1007/s00703-003-0070-7 [Accessed July 19, 2011].

Ching, J., Herwehe, J. & Swall, J., 2006. On joint deterministic grid modeling and sub-grid variability conceptual framework for model evaluation. *Atmospheric Environment*, 40(26), pp.4935–4945. Available at: http://linkinghub.elsevier.com/retrieve/pii/S1352231006000641 [Accessed August 15, 2011].

Colvile, R.N. et al., 2002. Uncertainty in dispersion modelling and urban air quality mapping. *Environmental Science and Policy*, 5, pp.207–220.

Cox, W. & Tikvart, J., 1990. A Statistical Procedure for Determining the Best Performing Air Quality Simulation. *Atmospheric Environment*, 24A, pp.2387–2395.

Dabberdt, W.F. & Miller, E., 2000. Uncertainty, ensembles and air quality dispersion modeling: applications and challenges. *Atmospheric Environment*, 34(27), pp.4667–4673. Available at: http://linkinghub.elsevier.com/retrieve/pii/S1352231000001412.

Denby, B. et al., 2010. *Guidance on the use of models for the European Air Quality Directive*,

Flemming, J. & Stern, R., 2007. Testing model accuracy measures according to the EU directives—examples using the chemical transport model REM-CALGRID. *Atmospheric Environment*, 41(39), pp.9206–9216. Available at: http://linkinghub.elsevier.com/retrieve/pii/S1352231007006838 [Accessed August 7, 2011].

Galmarini, S., Bianconi, R., Klug, W., et al., 2004. Ensemble dispersion forecasting - Part I: concept, approach and indicators. *Atmospheric Environment*, 38(28), pp.4607–4617. Available at: http://linkinghub.elsevier.com/retrieve/pii/S1352231004004959 [Accessed July 11, 2011].

Galmarini, S., Bianconi, R., Addis, R., et al., 2004. Ensemble dispersion forecasting-Part II: application and evaluation. *Atmospheric Environment*, 38(28), pp.4619–4632. Available at: http://linkinghub.elsevier.com/retrieve/pii/S1352231004004960 [Accessed June 27, 2011].

Gilliland, A.B. et al., 2003. Seasonal NH3 emission estimates for the eastern United States based on ammonium wet concentrations and an inverse modeling method. *Journal of Geophysical Research*, 108(D15), p.4477.

Gimson, N.R. & Uliasz, M., 2003. The determination of agricultural methane emissions in New Zealand using receptor-oriented modelling techniques. *Atmospheric Environment*, 37(28), pp.3903–3912.

Golder, 2012a. *Airshed Modelling of Exposure to Particulates in the Hawke's Bay Region*, Report prepared by Golder Associates (NZ) Limited for the Hawke's Bay Regional Council, June 2012.

Golder, 2009. *Contributions to PM10 levels in Napier and Hastings from surrounding urban and rural areas - airshed modelling re-analysis*, Report prepared by Golder Associated (NZ) Limited for the Hawke's Bay Regional Council, April 2009.

Golder, 2012b. *Development of an air quality model and meteorological data sets for the Nelson-Richmond urban area*, Report prepared by Golder Associates (NZ) Limited for Nelson City Council and Tasman District Council, February 2012.

Golder, 2011. *Population exposure to particulates in Auckland - urban airshed modelling*, Report prepared by Golder Associated (NZ) Limited for the Auckland Council, March 2011.

Hanna, S R, 1977. Air quality model evaluation and uncertainty. *Bulletin of the American Meteorological Society*, 38, pp.406–412.

Hanna, S R et al., 2001. Uncertainties in predicted ozone concentrations due to input uncertainties for the UAM-V photochemical grid model applied to the July 1995 OTAG domain. *Atmospheric Environment*, 35, pp.891–903.

Hanna, S R, Chang, J.C. & Fernau, M.E., 1998. Monte Carlo estimates of uncertainties in predictions by a photochemical grid model (UAM-IV) due to uncertainties in input variables. *Atmospheric Environment*, 32(21), pp.3619–3628. Available at: http://linkinghub.elsevier.com/retrieve/pii/S1352231097004196.

Hanna, Steven R. et al., 2007. Uncertainties in Air Toxics Calculated by the Dispersion Models AERMOD and ISCST3 in the Houston Ship Channel Area. *Journal of Applied Meteorology and Climatology*, 46(9), pp.1372–1382. Available at: http://journals.ametsoc.org/doi/abs/10.1175/JAM2540.1 [Accessed August 3, 2011].

Irwin, J S et al., 1987. Relating error bounds for maximum concentration estimates to diffusion meteorology uncertainty. *Atmospheric Environment*, 21(9), pp.1927–1937. Available at: http://linkinghub.elsevier.com/retrieve/pii/0004698187901533.

Irwin, John S., Petersen, William B. & Howard, S.C., 2007. Probabilistic Characterization of Atmospheric Transport and Diffusion. *Journal of Applied Meteorology and Climatology*, 46(7), pp.980–993. Available at: http://journals.ametsoc.org/doi/abs/10.1175/JAM2515.1 [Accessed August 7, 2011].

Jakeman, A.J., Simpson, R.W. & Taylor, J.A., 1988. Modeling distributions of air pollutant concentrations - III. the hybrid deterministic-statistical distribution approach. *Atmospheric Environment*, 22(I), pp.163–174.

Jakeman, A.J., Taylor, J.A. & Simpson, R.W., 1986. Modeling distributions of air pollutant concentrations - II. estimation of one and two parameter statistical distributions. *Atmospheric Environment*, 20(12), pp.2435–2447.

Jorquera, H. & Castro, J., 2010. Analysis of urban pollution episodes by inverse modeling. *Atmospheric Environment*, 44(1), pp.42–54. Available at: http://linkinghub.elsevier.com/retrieve/pii/S1352231009008346 [Accessed August 7, 2011].

Koracin, D. et al., 2007. A statistical approach for estimating uncertainty in dispersion modeling: An example of application in southwestern USA. *Atmospheric Environment*, 41, pp.617–628.

Lu, H.-C. & Fang, G.-C., 2003. Predicting the exceedances of a critical PM10 concentration — a case study in Taiwan. *Atmospheric Environment*, 37, pp.3491–3499.

MfE, 2004a. *Good practice guide for atmospheric dispersion modelling*, Report ME522. Wellington, New Zealand: Ministry for the Environment.

MfE, 2011. Resource Management (National Environmental Standards for Air Quality) Regulations 2004. Available at: http://www.mfe.govt.nz/laws/standards/air-quality/review/index.html.

MfE, 2004b. Resource Management National Environmental Standards relating to certain air pollutant, dioxins and other toxics regulations.

Miranda, A.I. et al., 2008. Determination of model uncertainty. *Joint Report of COST Action 728 (Enhancing Mesoscale Meteorological Modelling Capabilities for Air Pollution and Dispersion Applications) and GURME (GAW Urban Research Meteorology and Environment Project).*, pp.12–20.

Rao, K.S., 2007. Source estimation methods for atmospheric dispersion. *Atmospheric Environment*, 41(33), pp.6964–6973. Available at: http://linkinghub.elsevier.com/retrieve/pii/S1352231007004086 [Accessed August 7, 2011].

Rao, K.S., 2005. Uncertainty Analysis in Atmospheric Dispersion Modeling. *Pure and Applied Geophysics*, 162(10), pp.1893–1917. Available at: http://www.springerlink.com/index/10.1007/s00024-005-2697-4 [Accessed August 7, 2011].

Rao, S.T. et al., 1985. Resampling and extreme value statistics in air quality model performance evaluation. *Atmospheric Environment*, 19(9), pp.1503–1518. Available at: http://www.sciencedirect.com/science/article/pii/0004698185902884 [Accessed August 5, 2011].

Rodriguez, M.A. et al., 2007. Air quality impacts of distributed power generation in the South Coast Air Basin of California 2: Model uncertainty and sensitivity analysis. *Atmospheric Environment*, 41(27), pp.5618–5635. Available at: http://linkinghub.elsevier.com/retrieve/pii/S1352231007001975 [Accessed August 7, 2011].

Rodriguez, Marco A & Dabdub, D., 2003. Monte Carlo uncertainty and sensitivity analysis of the CACM chemical mechanism. *Journal of Geophysical Research*, 108(D15), p.4443. Available at: http://www.agu.org/pubs/crossref/2003/2002JD003281.shtml [Accessed August 7, 2011].

Sax, T. & Isakov, V., 2003. A case study for assessing uncertainty in local-scale regulatory air quality modeling applications. *Atmospheric Environment*, 37(25), pp.3481–3489. Available at: http://linkinghub.elsevier.com/retrieve/pii/S1352231003004114 [Accessed August 5, 2011].

Schlunzen, K.H. & Sokhi, R.S., 2008. Overview of Tools and Methods for Meteorological and Air Pollution Mesoscale Model Evaluation and User Training. *Joint Report of COST Action 728 (Enhancing Mesoscale Meteorological Modelling Capabilities for Air Pollution and Dispersion Applications) and GURME (GAW Urban Research Meteorology and Environment Project)*.

Sharma, P., Khare, M. & Chakrabarti, S., 1999. Application of extreme value theory for predicting violations of air quality standards for an urban road intersection. *Transportation Research Part D*, 4(3), pp.201–216. Available at: http://www.sciencedirect.com/science/article/pii/S1361920999000061 [Accessed August 7, 2011].

Tarantola, A., 1987. Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation. , p.613.

Tatang, M.A. et al., 1997. An efficient method for parametric uncertainty analysis of numerical geophysical models. *Journal of Geophysical Research*, 102(D18), pp.21,925–21,932.

Taylor, J., Jakeman, A. & Simpson, R., 1986. Modeling distributions of air pollutant concentrations - I. identifications of statistical models. *Atmospheric Environment*, 20(9), pp.1781–1789.

USEPA, 1996. Compilation of Photochemical Models' Performance Statistics for 11/94 Ozone SIP Applications. , p.50.

Vardoulakis, S. et al., 2002. Model sensitivity and uncertainty analysis using roadside air quality measurements. *Atmospheric Environment*, 36(13), pp.2121–2134. Available at: http://linkinghub.elsevier.com/retrieve/pii/S1352231002002017.

Vautard, R. et al., 2009. Skill and uncertainty of a regional air quality model ensemble. *Atmospheric Environment*, 43(31), pp.4822–4832. Available at: http://linkinghub.elsevier.com/retrieve/pii/S1352231008009072 [Accessed June 24, 2011].

Walker, S.-E., 2007. Model and observation uncertainty analysis towards data assimilation on the urban scale. , p.37.

Weil, J.C., Sykes, R.I. & Ventakram, A., 1992. Evaluating Air-Quality Models: Review and Outlook. *Journal of Applied Meteorology*, 31, pp.1121–1145.

Williamson, Derek G, Graettinger, Andrew J & Yegnan, Anand, 2004. Methods for incorporating uncertainty into dispersion and deposition models. *Environmental Progress*, 23(2), pp.158–167. Available at: http://doi.wiley.com/10.1002/ep.10016 [Accessed August 5, 2011].

Willmott, C.J., 1981. On the validation of models. *Physical Geography*, 2, pp.184–194.

Willmott, C.J., 1982. Some comments on the evaluation of model performance. *Bulletin of the American Meteorological Society*, 63, pp.1309–1313.

Yegnan, A, Williamson, D G & Graettinger, A J, 2002. Uncertainty analysis in air dispersion modeling. *Environmental Modelling & Software*, 17, pp.639–649.

APPENDIX A

Report Limitations

REPORT LIMITATIONS

This Document has been provided by Golder Associates (NZ) Ltd ("Golder") subject to the following limitations:

(i). This Document has been prepared for the particular purpose outlined in Golder's proposal and no responsibility is accepted for the use of this Document, in whole or in part, in other contexts or for any other purpose.

(ii). The scope and the period of Golder's Services are as described in Golder's proposal, and are subject to restrictions and limitations. Golder did not perform a complete assessment of all possible conditions or circumstances that may exist at the site referenced in the Document. If a service is not expressly indicated, do not assume it has been provided. If a matter is not addressed, do not assume that any determination has been made by Golder in regards to it.

(iii). Conditions may exist which were undetectable given the limited nature of the enquiry Golder was retained to undertake with respect to the site. Variations in conditions may occur between investigatory locations, and there may be special conditions pertaining to the site which have not been revealed by the investigation and which have not therefore been taken into account in the Document. Accordingly, additional studies and actions may be required.

(iv). In addition, it is recognised that the passage of time affects the information and assessment provided in this Document. Golder's opinions are based upon information that existed at the time of the production of the Document. It is understood that the Services provided allowed Golder to form no more than an opinion of the actual conditions of the site at the time the site was visited and cannot be used to assess the effect of any subsequent changes in the quality of the site, or its surroundings, or any laws or regulations.

(v). Any assessments made in this Document are based on the conditions indicated from published sources and the investigation described. No warranty is included, either express or implied, that the actual conditions will conform exactly to the assessments contained in this Document.

(vi). Where data supplied by the client or other external sources, including previous site investigation data, have been used, it has been assumed that the information is correct unless otherwise stated. No responsibility is accepted by Golder for incomplete or inaccurate data supplied by others.

(vii). The Client acknowledges that Golder may have retained subconsultants affiliated with Golder to provide Services for the benefit of Golder. Golder will be fully responsible to the Client for the Services and work done by all of its subconsultants and subcontractors. The Client agrees that it will only assert claims against and seek to recover losses, damages or other liabilities from Golder and not Golder's affiliated companies. To the maximum extent allowed by law, the Client acknowledges and agrees it will not have any legal recourse, and waives any expense, loss, claim, demand, or cause of action, against Golder's affiliated companies, and their employees, officers and directors.

(viii). This Document is provided for sole use by the Client and is confidential to it and its professional advisers. No responsibility whatsoever for the contents of this Document will be accepted to any person other than the Client. Any use which a third party makes of this Document, or any reliance on or decisions to be made based on it, is the responsibility of such third parties. Golder accepts no responsibility for damages, if any, suffered by any third party as a result of decisions made or actions based on this Document.

At Golder Associates we strive to be the most respected global company providing consulting, design, and construction services in earth, environment, and related areas of energy. Employee owned since our formation in 1960, our focus, unique culture and operating environment offer opportunities and the freedom to excel, which attracts the leading specialists in our fields. Golder professionals take the time to build an understanding of client needs and of the specific environments in which they operate. We continue to expand our technical capabilities and have experienced steady growth with employees who operate from offices located throughout Africa, Asia, Australasia, Europe, North America, and South America.

Africa	+ 27 11 254 4800
Asia	+ 852 2562 3658
Australia & New Zealand	+ 61 7 3721 5400
Europe	+ 44 356 21 42 30 20
North America	+ 1 800 275 3281
South America	+ 55 21 3095 9500

solutions@golder.com
www.golder.com

AUCKLAND

Tel +64 9 486 8068
Fax +64 9 486 8072

Level 2
Takapuna Business Park
4 Fred Thomas Drive
Takapuna
Auckland 0622

PO Box 33-849
Takapuna 0740

TAURANGA

Tel +64 7 928 5335
Fax +64 7 928 5336

78 Maunganui Road
Tauranga 3116

PO Box 13611
Tauranga Central
Tauranga 3141

HAMILTON

Tel +64 7 928 5335
Fax +64 7 928 5336

Ruakura AgResearch Centre
Room 31 in the Homestead
East Street
Hamilton 3214

PO Box 19-479
Hamilton 3244

NELSON

Tel +64 3 548 1707
Fax +64 3 548 1727

Level 3
295 Trafalgar Street
Nelson 7010

PO Box 1724
Nelson 7040

CHRISTCHURCH

Tel +64 3 377 5696
Fax +64 3 377 9944

Level 1
132 Tuam Street
Christchurch 8011

PO Box 2281
Christchurch 8140

DUNEDIN

Tel +64 3 479 0390
Fax +64 3 474 9642

Level 9A
John Wickliffe House
265 Princes Street
Dunedin 9016

PO Box 1087
Dunedin 9054

ARROWTOWN

Tel +64 3 409 8106
Fax +64 3 409 8106

11 Arrow Lane
Arrowtown 9302

Golder Associates