Method to Assess Presence of Annual Trend or Unusual Values in Air Quality Data

Prepared for:

Capital Regional District
Environmental Services Department
524 Yates Street
Victoria, BC V8W 2S6

Prepared by:

SENES Consultants Limited
1338 West Broadway, Suite 303
Vancouver, B.C.V6H 1H2

June 2006

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

EXECUTIVE SUMMARY

The Capital Regional District (CRD) tracks air concentrations at local monitoring locations and reports these concentrations on an annual basis. As a component of the annual report, the CRD wishes to assess, and report, on whether there are statistically significant trends in these values from year-to-year or unusual values. A standard methodology is required that will allow the CRD to assess whether these departures from the normal pattern are statistically significant and not the result of random variation. In addition to being statistically defensible, the CRD has indicated that the method should not be difficult or complex to implement, and should be understandable to a wide audience.

Sample data from a monitoring location in the Greater Vancouver Regional District (GVRD) that has a longer time period of measurement than is currently available from the CRD monitoring sites was used to explore the potential magnitude and type of trends that might be expected in the CRD data and the typical degree of variability in annual concentrations from year-to-year. Trends, mostly decreasing concentrations, were apparent in the sample data, and the trends appeared approximately linear over time periods of about 10 years. Normal year-to-year variation was on the order of 5% (for one standard deviation).

The proposed method would be to use linear regression over a 10-year period to assess the magnitude and statistical significance of any potential trends. Confidence intervals on the annual values over the period of the sample data set were developed using the regression method if the trends were statistically significant. Otherwise, the confidence intervals were based on the observed variability in the data set where a statistically significant trend was not found. Annual values that were outside these confidence intervals were considered statistically significant and assessed as outliers. The method has high probability of detecting trends on the order of 2% per year over the 10-year time period. Although the target would be 10-years, the method could be applied once 5 years of data are available, as the method has power to detect trends larger than 2%/y over that time frame.

An important component of the method is visual charting that shows the annual values and the confidence intervals on the annual values over the assessment period. This provides a context on the pattern of concentrations that would be difficult to provide with text or numerical values alone.

The proposed statistical reporting method does not address the consequences of, or reasons for, the trend, or actions to address the observed changes in air quality. The method may detect statistically significant trends that are not of practical importance (e.g. decreasing trends or very small trends in general). Actions arising from the detection of trends, or unusual values, are beyond the scope of a standardized annual assessment of trends and are best addressed, at least in part, by more focussed studies that may include a periodic review of the data or special studies based on specific policy initiatives.

The statistical reporting method suggested by this report is a statistically defensible method of assessing whether changes in air quality are occurring in the CRD. It is expected to be understandable by a wide audience and can be readily implemented in a reproducible manner. The method is applicable for assessing the presence of trends over time periods of about 10 years for a wide range of annual statistics (e.g. mean, 98th percentile, proportion exceeding guidelines).

1.0 INTRODUCTION

The long term air quality monitoring program operating in the Capital Regional District (CRD) is a result of a partnership between the CRD, the British Columbia Ministry of the Environment (MoE) and Royal Roads University. The CRD annually reports on measured air pollutant concentrations at monitoring locations throughout the region. As part of this effort, the CRD would like to provide an assessment on whether changes in air quality are occurring from year-to-year.

The CRD has commissioned SENES to develop a standardized method for assessing whether changes have occurred in the ambient air quality. Air concentrations vary naturally, or normally, from year-to-year in the same way average temperature or rainfall varies. Changes of interest are departures from this normal variation or pattern of air quality. The assessment method was to be based on statistical methods to identify changes that are statistically significant. Other attributes are that the method, and assessment, should be consistently applied, reproducible and understandable by a wide audience.

1.1 FOCUS POLLUTANTS AND GUIDELINES

Table 1.1 shows the six pollutants and averaging periods identified specifically by the CRD for development of an annual reporting method. These pollutants are primarily combustion-related pollutants with the exception of PM_{10}. The table includes the ambient guidelines established by CRD for these pollutants. Since air quality in the CRD is relatively good, some of these guidelines are established at levels that are lower than applicable guidelines used in other jurisdictions. Use of the CRD guidelines is consistent with the 'keeping clean areas clean' component of the Canada Wide Standards (CWS) for air quality.

TABLE 1.1
AIR QUALITY GUIDELINES FOR CAPITAL REGIONAL DISTRICT

Pollutant	Averaging Period	Upper Bound Guideline ($\mu g/m^3$)
CO	8-hour sequential	550
NO_2	1-hour	200
SO_2	24-hour sequential	125
Ozone	8-hour rolling	120
$PM_{2.5}$	24-hour sequential	25
PM10	24-hour sequential	50

The statistics of interest identified for these pollutants were the annual values for mean concentrations (at the averaging periods identified above) and the proportion of measurements during a year exceeding the upper bound guideline.

1.2 DEFINITION OF CHANGE

1.2.1 Natural Variation

Air quality is naturally variable due to the nature of emission sources and meteorological conditions, which vary over many different time scales. Meteorological conditions vary by time-of-day, day-of-the-week and month-of-the-year. Emission rates, and types of emissions, also have similar variability due in part to the meteorology (e.g. release rates from home heating are higher in the winter than in the summer). Emissions that contribute to the monitored concentrations are integrated (averaged) over large spatial scales (e.g. global, regional, city, and local) and the pollutant concentration depends on the relative frequency of winds from these source areas. The measured air concentrations due to these emissions are temporally variable not only due to temporal variability in the emissions themselves but also due to temporal variability in meteorological conditions.

Variation from year-to-year is expected regardless of the variation in emissions. For example, the PM10 concentrations would be expected to be lower in years with wetter conditions than in years with dryer conditions. These normal short-term variations, or changes, in air quality are interesting but are not the focus of this investigation. Normal variation is considered the year-to-year variation that would be present without a change in typical emission rates or typical meteorology.

For these statistical analyses, the primary consideration is the identification of changes in the "normal" air quality as characterized by the annual mean or the annual proportion of measurements over a guideline.

1.2.2 Change in Air Quality

The specific interest is to determine whether there has been a change or trend in measured air quality concentrations from the normal pattern arising from changes in emissions and/or meteorological conditions. The general conceptual (statistical[1]) model is that there is some expected annual value which has "normal" year-to-year variation and that an underlying trend in the expected value might be present.

---

[1] $C(t) = \mu + m(t) + \varepsilon$

Departures from the normal variation can arise if there were a change in one, or more, of the factors that affects air quality. For example, increasing residential development will increase emissions and this, all other factors equal, will result in a trend of increasing air quality. Other departures might be short-term; for example, a forest fire might increase the concentration of PM_{10} and $PM_{2.5}$ during one year but this would be episodic and would not constitute a trend in air quality. However, it is important to identify not only trends (ongoing change) but also the episodic type of change.

1.3 APPROACH

The approach taken to develop and recommend an annual reporting method comprises the following activities:

i) *exploratory review of example air quality data*: This provides indication of the type and magnitude of trends that might be present in CRD air quality and the amount of normal variability from year-to-year that could be anticipated.

ii) *formalize method requirements*: This defines the type of trends or changes that will be assessed by the method, the time periods over which trends are important and the magnitudes of trends that are important to detect.

iii) *selection of statistical method*: the method should fulfil the requirements using statistically defensible techniques and presentations that are understood by a wide audience

iv) *test the detection capability*: using the method and simulated data sets with properties similar to example data, the probability of detecting trends of various magnitudes as a function of normal variation, number of years of data is estimated

v) *example application*: the method is used on the example data sets in order to anticipate the type of reports that would be created using the method with CRD data.

2.0 METHOD REQUIREMENTS

This chapter describes the type of data that will be assessed. The assessment includes exploratory data analyses of air quality measurements from monitoring locations with a longer period of record than is currently available from the CRD monitoring program. The investigation of example data provides information on the type and magnitude of trends anticipated in the CRD data and also provides information on magnitude of normal year-to-year variation. This information is important in defining the method requirements and the selection of an appropriate method.

2.1 MONITORING DATA

The pollutants of concern are measured with continuous monitoring over the year (exception PM_{10}) and therefore reflect actual conditions, apart from analytical uncertainty and time periods when the monitor was not operating. With the exception of PM_{10}, the pollutants are measured on an hourly basis. Data processing[2] is then conducted to determine averages over the appropriate averaging period for each pollutant.

It is unlikely that valid measurements are available for every hour of the year. Missing concentrations from some hours of an averaging period will result in some uncertainty of the actual average during that particular period; regardless, these annual values will be known precisely. More importantly, there are large seasonal variations for some pollutants; therefore, missing measurements over relatively large contiguous periods of time can result in under-or-over estimation of the annual values.

For this study, mean concentrations for 8-hour averaging periods were calculated if at least 7 of the hours had valid concentrations. Annual values were used if at least 80% of the averaging time periods had a measured value. Data quality standards for data completeness used by CRD, or others, may vary from these assumptions without substantial effect on the method.

2.2 EXAMPLE DATA

The current duration of measurement records for the CRD monitoring network is limited, therefore there is little local information on the type and magnitude of trends that will occur or the typical level of variation from year-to-year. Recent data from the Victoria Topaz NAPS station (1999 to 2004) and the T18 monitoring location in GVRD (1988 to 2004) were used to examine the statistical properties of pollutant concentrations over a longer time period and to use these statistical properties in the development and testing of statistical methods.

---

[2] Sometimes referred to as data reduction

The analyses described in the following sections are exploratory in nature and are not intended as a formal assessment of trends in the example data sets.

2.2.1 Time Plot of Annual Values in Example Data

Annual means and 98th percentiles were calculated for the six pollutant measures and time series of these values have been plotted in Figure 2.1. As stated earlier, there is a desire to track trends in both annual mean values and exceedences of the CRD guidelines. As there are very few actual exceedences of the guidelines currently, the exceedence levels themselves may not offer a useful measure in terms of trend analysis. Use of the 98th percentile is suggested here as a potential measure/statistic that may provide supplemental indication on the potential for guideline exceedences. The 98th percentile reflects the higher concentration levels therefore trends in this statistic are more closely related to the potential of exceeding the guideline than a trend in the mean concentration.

It is visually apparent that there have been changes in pollutant concentrations over the time period with some pollutant concentrations continuing to change. Substantial decreases have occurred for some pollutants at the GVRD location, for example, the concentrations of CO and SO_2 have decreased by about a factor of two since the late 1980s. Other pollutants have smaller decreases with, for example, little evidence of a change in PM. An exception is ozone which exhibits an increasing trend in concentration. Trends are less evident at Topaz compared to GVRD due to shorter time period.

The changes depend on both the monitoring location and the statistic under consideration. For example, mean PM10 concentrations at GVRD were relatively constant over the time period however there was a decrease in the 98th percentile concentration. As another example, the mean SO_2 concentration at Topaz is decreasing over the more recent period compared to relatively constant SO_2 at the GVRD station during the same period.

It is interesting that pollutant concentrations at the Topaz location tend to be higher than at the GVRD location with a notable exception being NO_2. Other interesting aspects are that the mean CO at Topaz is lower than the mean CO at the GVRD location; however, the 98th percentile at Topaz is higher than the 98th percentile at GVRD. This implies that CO concentrations within a year are relatively more variable at Topaz compared to GVRD.

The explanations for the patterns are difficult to determine exactly; however, some patterns coincide with changes in emissions that have occurred in the CRD and GVRD. The large decrease in SO_2 at the GVRD monitoring location after 1994 coincides with the closing of refineries and the associated SO_2 emissions. Other factors such as the reduction of sulphur content in diesel fuel and the transition to natural gas from fuel oil for heating would result in reduction of pollutant concentrations and are plausible explanations for some of the decreasing trends.

2.2.2 CRD Guideline Exceedences

Figure 2.2 shows the plots of the proportion of averaging periods that exceed the CRD guidelines. The scales for each graph are shown to allow easy comparison between the pollutants of the proportion of averaging periods that exceed guidelines. With the exception of PM, the proportion of measurements exceeding guidelines is below 1% every year at both GVRD and Topaz. In recent years, there were no averaging periods exceeding the guidelines for CO and SO_2 at both stations with only two years with NO_2 values exceeding the guideline at the GVRD location.

Trends are visually apparent in the data and these trends are similar to the trends observed in the mean and 98th percentile concentrations as might be expected. There are easily apparent decreasing proportions for CO and NO_2 and less apparent, if any, trends for ozone and PM. All values for SO_2 were below the guideline so there are no trends to evaluate.

The normal pattern of proportions are not as smooth as the trends for annual mean or 98th percentiles. This arises in part because the proportions are close to 0 so the relatively variability will be higher than other statistics such as the mean or 98th percentiles. In addition, the annual values have the binomial nature of proportions since the underlying data are discrete variables (e.g. 1 event, 3 events) rather than a continuous variable such as annual mean or 98th percentile. Specifically, the number of events is necessarily a whole number so pattern between years with low numbers of events appears more erratic than a continuous variable such as mean concentration.

FIGURE 2.1
AIR CONCENTRATIONS MEASURED IN EXAMPLE DATA SETS

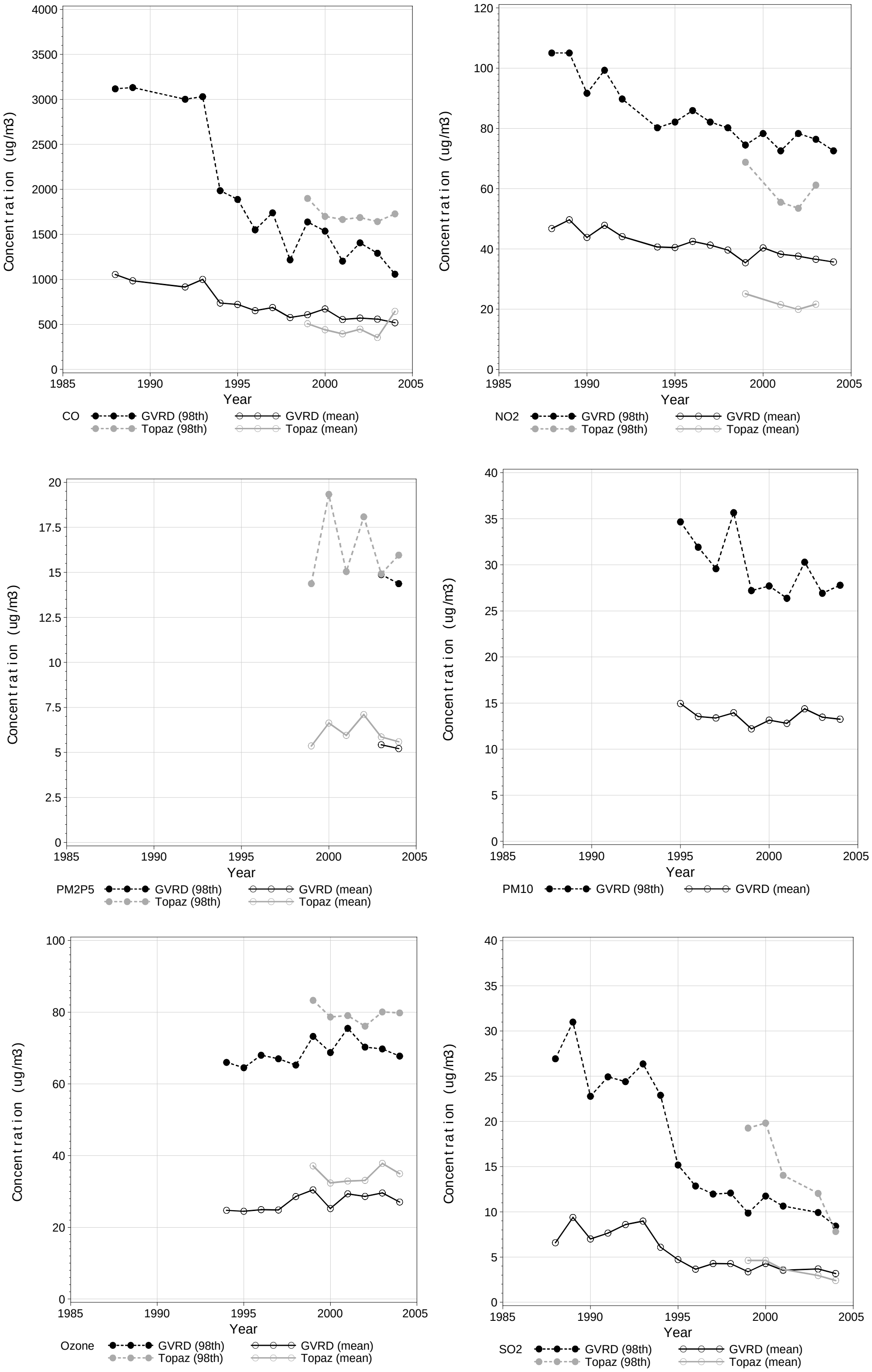
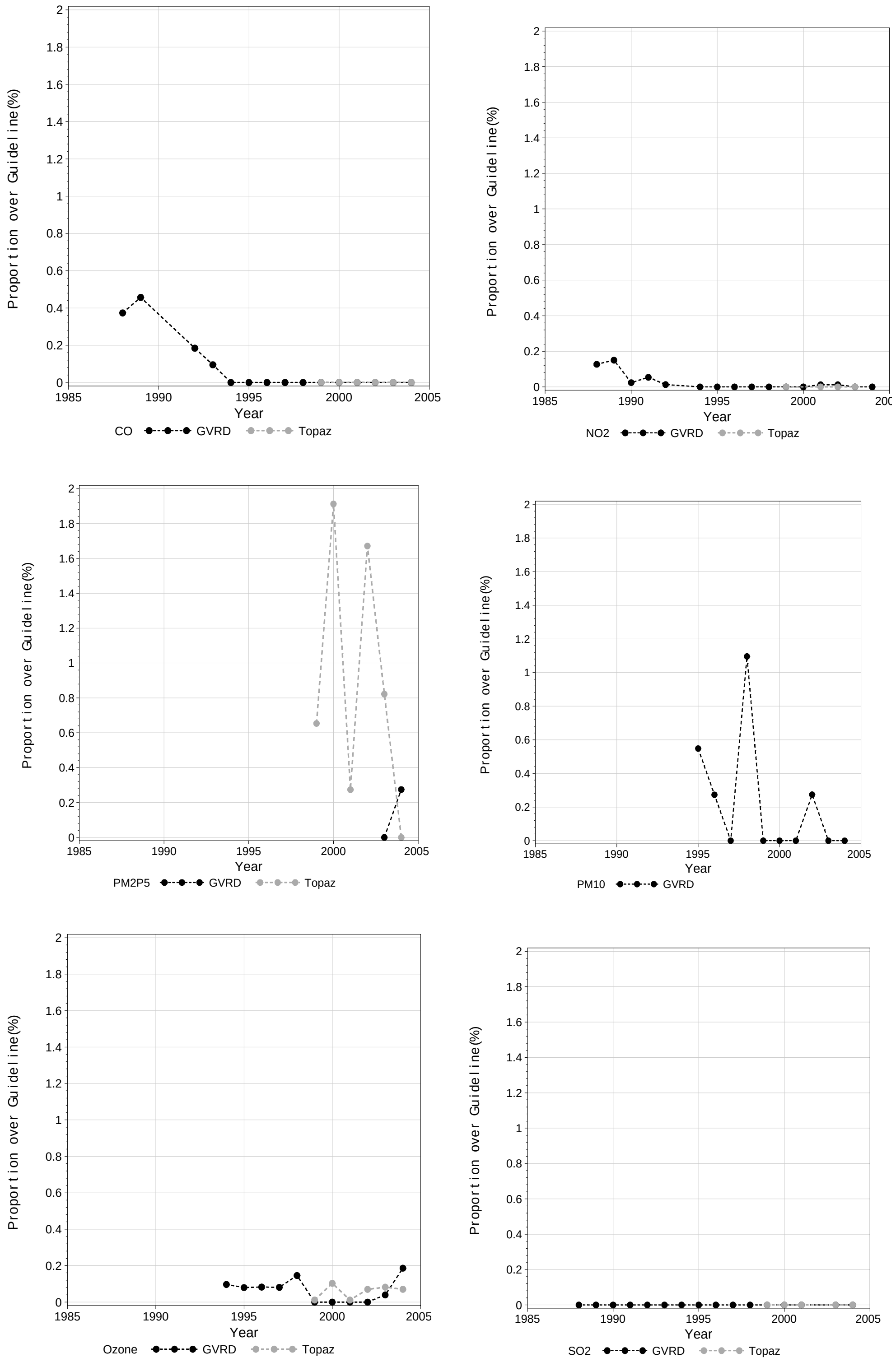


Figure 2.2
Analysis of CRD Guideline Exceedence (with use of both Topaz and GVRD data)



2.2.3 Presence of Trends in Example Data

Simple linear regression analyses were used as an exploratory tool to quantify, approximately, the magnitude of trends in pollutant measures at the GVRD station assuming that the trend was linear over the entire period of data. Table 2.1 shows these results: most, 8 of the 10 time series, showed statistically significant trends over the time period. The table provides the trend in absolute terms, μg/m^3 per year, and in relative terms, % per year. The relative scale allows for comparison of trends across pollutants and between the two annual statistics.

The statistically significant relative trends ranged from about 1.5% per year up to about 8% per year. The two time series with non-significant trends had point estimates of trends with a relative trend of less than 1% per year. It is interesting that the relative trend tended to be larger for the 98th percentiles than for the mean concentrations and the trends in 98th percentiles tended to be more statistically significant than the trends in the mean values.

TABLE 2.1
TREND ESTIMATES FOR PERIOD OF RECORD FOR GVRD EXAMPLE DATA

Pollutant	Trend (ug/m^3 per year)			Relative Trend (% per year)	
	Mean	98th		Mean	98th
CO	-34*	-141*		-4.7*	-7.3*
NO_2	-0.75*	-1.9*		-1.8*	-2.2*
Ozone	0.44*	0.52		1.63*	0.75
PM_{10}	-0.08	-0.73*		-0.6	-2.5*
SO_2	-0.36*	-1.4*		-6.5*	-8*

Note: * indicates a statistically significant trend at the 5% level

2.2.4 Normal (Annual) Levels and Variation

For many of the pollutant measures, the annual values plotted in Figure 2.1 appeared relatively constant since about 2000. The annual statistics for a recent period, 2000 to 2004 inclusive, were summarized to calculate the mean value of the statistic and the variability[3] in the statistic from year to year during these years without a large trend.

Table 2.2 shows the mean value for selected statistics along with the relative year-to-year variability in these statistics. The presence of a statistically significant trend[4], within the five

---

[3] The standard deviation was selected for describing absolute variability in the statistic and the coefficient of variation (the standard deviation divided by the mean) was chosen to describe the relative variability in the statistic.

[4] A simplified screening assessment for the presence of trend was conducted using linear regression and considering the p-value for the slope (trend).

year period, is also indicated and this occurred for a number of pollutants. There were statistically significant trends (decreasing) in CO at the GVRD location for both average (mean) and upper percentiles and for average NO_2. There were statistically significant decreasing trends in SO_2 concentration at Topaz and for upper percentiles at the GVRD location. The amount of variability during this time period includes not only the normal year-to-year variation but also the variation introduced by any trends over the time period.

TABLE 2.2
YEAR-TO-YEAR VARIABILITY (%) FROM 2000-2004

		Central (Average) Tendency		**Higher Percentiles**	
	Location	**Median**	**Mean**	**95th**	**98th**
CO	GVRD	513 (9%) *	575 (10%) *	1047 (13%)**	1299 (14%)
	Topaz	349 (40%)	457 (24%)	1281 (10%)	1685 (2%)
NO_2	GVRD	36 (4%) **	38 (5%) **	68 (4%)	76 (4%)
	Topaz	18 (6%)	21 (5%)	48 (4%)	57 (7%)
SO_2	GVRD	3.3 (10%)	3.7 (13%)	8.1 (14%)*	10 (14%) **
	Topaz	2.4 (28%)**	3.4 (28%)**	10 (33%) **	13 (37%) **
Ozone	GVRD	26 (9%)	28 (7%)	64 (5%)	70 (4%)
	Topaz	33 (8%)	34 (7%)	71 (2%) *	79 (2%)
PM_{10}	GVRD	12 (5%)	13 (4%)	25 (6%)	28 (5%)
$PM_{2.5}$	GVRD				
	Topaz	5.4 (10%)	6.2 (10%)	14 (11%)	17 (12%)

Note: * indicates a statistically significant trend at the 5% level, and ** indicates significance at the 1% level

With the general exception of those measures that exhibit statistically significant trends, the year-to-year variability was typically 10% or less. In fact, the normal variability for NO_2 at GVRD was 5% even though a statistically significant trend was present. The actual variability from year-to-year for that pollutant at that site would be lower than 5%. Overall the estimates of variability in the table tend to overestimate the year-to-year variability due to the presence of non-statistically significant trends.

The relative variability is similar across the various measures with one exception being CO at Topaz. The relative variability in average measures (mean, median) is much larger than the variability in the upper percentiles for this particular station and pollutant. Inspection of the data indicates that the mean value in 2004 was much higher than average concentrations in the other years. This may be an episodic event that contributes to the higher relative uncertainty.

In summary, the relative variability in central tendency and upper percentile measures from year-to-year in the absence of an underlying trend or episodic change is on the order of about 5% for one standard deviation with a likely maximum of 10% or so. This is the "noise" in the time series from which a "signal" of trend or change needs to be estimated.

2.3 DISCUSSION

The exploratory data analyses indicated the following:

i) Many pollutants in the GVRD data indicate the presence of statistically significant trends (changes over time) in the target pollutant concentrations over the recent period (10 – 15 years).
ii) Most of the trends are of decreasing pollutant concentrations.
iii) The trends may be complex, that is, not a constant (straight-line) trend but some trends are potentially levelling off or in some case reversing.
iv) In some cases, the trends are still occurring and are statistically significant over periods of 5 years or less.
v) The normal year-to-year variation, in absence of a trend, is likely on the order of 5% (one standard deviation) with a plausible maximum of about 10% for the pollutants considered.
vi) Some trends of about 1% per year are likely detectable over time periods of 10 years or so.
vii) Trends in the 98th percentile can be different, and easier to detect, than trends in the mean concentration.

Based on the example data, it is likely that pollutant concentration trends will be observable in the CRD monitoring once a number of years of data have been collected. The GVRD data indicate that, in some cases, the recent trends showed decreasing pollutant levels but that these trends may be levelling off. The trends over the next 10 years or so in CRD measurements will not likely be as large as those seen previously at GVRD.

The guidelines of specific interest to CRD involve concentrations over much shorter averaging periods than a year (i.e., 1-hour, 8-hour, 24-hour) and the trend in the proportion of measurements exceeding these guidelines if of interest. This may be accomplished in part by assessing the proportion of measurements exceeding the guideline; however, very few, in some cases no, of the measurements exceed the guidelines. Therefore, using this statistic to assess whether trends are occurring would not identify the trend until guidelines start to be exceeded. Assessing trends in the mean concentration does not guarantee that trends are not occurring in the higher percentiles and that the guideline will not be exceeded in the future. It is apparent that the relative trend in upper percentile concentrations can be different than the trend in the mean concentrations[5].

The assessment of the proportion of measurements that may exceed the guideline must address not only average concentrations but the upper percentile concentrations as well since there are currently few, if any, measurements over the guideline. The 98th percentile is a relatively stable statistic for the pollutant of interest as there are between 7 and about 150 measured values above this statistic each year depending on the averaging period. It is recommended that trends in this statistic be assessed.

The trends can be complex over even a period of 20 years and the trends are not likely to be as simple as a step change or constant linear trend. The actual trend may initially show a decrease due to regulatory or technical improvements (improved automobile efficiency), level off and then begin to increase as development continues to expand. In order to discuss complex trends it may be desirable to establish a reference condition for comparison of current conditions but also to assess the "recent" trend. However, it appears reasonable to assume that trends over a 10-year period are approximately linear.

2.4 METHOD REQUIREMENTS

The statistical requirements of the method include:

i) assess a recent period of data for the presence of a linear (straight-line) trend;
ii) determine whether these trends are statistically significant and quantify the trend, if present;
iii) identify outliers, if present;
iv) provide sufficient probability to detect trends of operational importance[6].

The operational and reporting requirements of the method are:

i) statistically defensible;
ii) understandable by a wide audience;
iii) standardized and reproducible;
iv) implemented with readily accessible software.

---

[5] The variability of concentrations within a year can change over time.

[6] Assumed to be 2% per year.

3.0 STATISTICAL METHODS

This chapter describes a statistically-based method for simultaneously assessing the presence of trends and/or unusual values in air quality concentration over a time period.

3.1 STATISTICAL CONSIDERATIONS

3.1.1 Statistical Significance

The primary approach is to determine if trends or annual values are statistically significant; that is, the observed pattern is unlikely to occur by chance alone. In this approach, the null hypothesis (or assumption) is that no trend or change is occurring and the data are assessed to see if there is evidence against this hypothesis.

Statistical significance has been assessed in the method by considering confidence intervals. For example, if a statistic falls outside of the 95% confidence interval for that statistic, there is less than a 5% chance, or 1 time in 20, (a p-value of 0.05) that the observed value occurred by chance alone. Since this is considered unlikely by chance alone, the pattern would be considered statistically significant at the 5% level.

For example, confidence intervals are estimated for statistics of interest (e.g. the trend, predictions and normal variation). Trends are considered statistically significant if the confidence interval for the trend does not include the value of zero. Annual values are considered statistically significant if they are outside the confidence intervals for normal variation or prediction confidence intervals from the trend analyses.

3.1.2 Statistical Assumptions

The annual values are assumed to be approximately normally distributed about an expected value (e.g. a mean or a trend line if present). Uncertainty in a particular annual value, apart from instrument variation, is minimal since the measurements are continuous and, in effect, provide the actual concentration and not an assessment subject to sampling uncertainty. Measurements of PM10 however, are made on a one in six day schedule; therefore, there will be some sampling uncertainty in the annual value.

The uncertainty in differences between annual values comes from a complex process of meteorological conditions and the pollutants come from many different emission sources such that the measured air concentrations are the result of an integration, or averaging, of multiple emission sources. Under these conditions, the annual statistics are assumed approximately normally distributed for most years. For the case of PM10, the year-to-year variation includes a component of sampling variability in the annual values which is also approximately normally distributed so the year-to-year variation in PM10 will also be approximately normally distributed. However, there may be an usual annual value arising from a very different and, likely, local process (e.g. a forest fire, instrument malfunction). This would cause a departure from the normality, increase the apparent variability and make it more difficult to detect trends or unusual annual values. Statistical methods are available to remove these "outliers".

Trends, if present, are assumed to be approximately linear over time periods on the order of 10 years. This is reasonable for air quality since air concentrations are integrated over a large number of sources and over a large geographic area therefore changes will be gradual unless a monitoring location is highly influenced by a nearby source.

3.2 TRENDS

Many statistical techniques are available for detecting and quantifying trends and these range in complexity from correlation analyses to time series analyses. Correlation analyses and regression techniques were considered in the method development. More complex techniques, such as time series analyses, were not considered due to their specialized statistical software and training requirements and complex interpretation and communication requirements for a standard approach. These methods however may be applicable to detailed investigation of specifically identified patterns of interest.

3.2.1 Correlation Analyses

Correlation analyses calculate a statistic, the correlation coefficient, that quantifies the relationship between two variables. High positive or negative values of the correlation coefficient indicate an association between the two variables. For a time-series of pollutant concentration, the correlation between concentration and year is determined. If the correlation coefficient is statistically significant the conclusion is that a statistically significant trend has occurred. There are a number of methods that could be used with the Pearson and Spearman tests being common choices. While the Spearman test is more general and less sensitive to assumptions about the data, both methods only indicate whether a trend is statistically significant and do not quantify the magnitude of the trend.

3.2.2 Regression Analyses

A large number of regression techniques quantify the magnitude of the trend and indicate the statistical significance of the trend. In addition, regression techniques are recognized, and understood, by a broad audience.

The statistical techniques available for regression include simple linear regression, non-linear, non-parametric and boot-strapping methods among many others. Challenges for regression techniques include the robustness of the selected method to departures from the assumptions and the linearity of the trend.

There are many options available including those that required specialized software or training in statistics and utilizing concepts or approaches that are difficult to explain to a broad audience. Simple linear regression has been selected as the regression method for this study based on assumption of normal distribution, the approximately linear trend over time periods of about 10 years, accessibility of software and training and the broad familiarity of the method. This method can be implemented with many software products and is appropriate when the trend is approximately linear and the deviations between the measured concentrations and the trend line are normally distributed.

3.3 CONTROL CHARTS FOR NORMAL VARIATION

Statistical process control (SPC) methods have been developed by industry as part of total quality management (TQM) initiatives to improve the manufacturing processes. The general method is to characterize the normal variation of a process and establish control limits that describe the range of values. The method is visual and intuitive through graphically charting the values of interest. Once the normal variation has been established, subsequent values are compared to the control limits to determine if the process is still performing normally.

We adopt the SPC principles of quantification of the normal variation of the statistics as confidence intervals on the annual values and visually showing how the annual values compare to this range of natural variation. The confidence intervals provide the statistical significance of the annual values relative to the normal pattern and the visual display aids in the context of the most recent years measurements relative to other measurements at the location.

3.4 COMBINED APPROACH

Air quality concentrations are affected by normal year-to-year variation, trends over time and episodic events. Ideally the system should be robust to the episodic events (outliers) while allowing the detection of trends within the natural year-to-year variation. To this end a combined approach that quantifies trends while also recognizing the presence of outliers is preferred.

3.4.1 Data Reduction and Data Completeness

The hourly data is summarized by the corresponding averaging periods of interest with a criterion for number of hours with measured data. For this study, at least 7 out of 8 hours of an 8-hour averaging period were assumed necessary. The methods may work well with a different criterion; however, this has not been assessed as part of this study.

The values from individual averaging periods are then summarized to annual levels of the following statistics; mean, 98th percentile and the proportion exceeding the guideline. Since many pollutant measures vary by season, a check of completeness of data is made. For this investigation of example data sets, it was assumed that 80% of averaging periods required values before an annual value was used in the method.

3.4.2 Trend Analyses

The first statistical method is to use simple linear regression to fit a statistical (trend) model to the data. The statistical model for trend is

$$c(t) = mt + b + \varepsilon$$

where:

$c(t)$ is the concentration, $\mu g/m^3$, in year t

m is the estimate of linear trend ($\mu g/m^3$ per year)

t is the calendar year

b is the intercept term for the linear trend ($\mu g/m^3$)

ε is the normal year-to-year variation that would be present in the absence of a trend.

Simple linear regression algorithms are generally available; however, we include a check for outliers that, if not addressed, would reduce the ability of the method to detect trends. This is done by a cross-validation procedure which entails re-calculating the regression with each observation, in turn, removed. The studentized residual, which in simple terms is the ratio of the residual to the prediction uncertainty, is calculated for the observation under consideration. High values[7] of the studentized residual indicate the annual value is unusual or an outlier compared to the other data and these observations were excluded from the trend fitting.

---

[7] Annual values with a residual of < -3 or > 3 were excluded. The probability of these occurring by chance alone is about 2 in 1,000.

Regression packages provide the estimate of trend and intercept and confidence intervals for these attributes. The trend is considered statistically significant if the confidence interval does not contain the value of 0.

The approach is also appropriate for detecting trends in the annual proportion as discussed in Appendix B. For more than 5 events per year, the data have similar properties as annual values for mean and 98th percentiles. With smaller number of events the method works adequately for the detection of trends although more sophisticated techniques are available if trends at these proportions of events exceeding guidelines is important.

3.4.3 Confidence Intervals for Annual Values

The next step is to estimate confidence intervals for annual values and the method procedure depends on the whether the trend was statistically significant or not. In both cases, 95% confidence intervals are constructed.

If the trend is statistically significant, confidence intervals of predication of individual observations are provided by most statistical packages. These can be considered as limits of normal year-to-year variation about the estimated trend. These confidence intervals are retained for charting and assessment of unusual annual values.

If the trend is not statistically significant, the mean concentration, not including the outliers identified during the regression, is calculated as is the standard deviation of the concentrations. Upper and lower limits are calculated based on the observed mean, standard deviation and appropriate t-statistic for the selected confidence level.

3.4.4 Assessment of Unusual Values

Annual values are assessed as to whether they are within the confidence intervals. Annual values are assessed as statistically significantly outliers if they fall outside the confidence limits. Since the intervals are for 95% confidence intervals, the annual values so identified are statistically significant outliers at the 5% significance level.

4.0 PERFORMANCE EVALUATION

Actual monitoring data provides site-specific information on the statistical properties of year-to-year variability. Using these data, the performance (power to detect a trend or unusual measurement and the false detection rate) has been simulated as a function of the number of years of data, various magnitudes of (linear) trend and selected confidence levels.

The method provides a suitable probability of detecting a target trend of 2%/y.

4.1 PERFORMANCE ATTRIBUTES AND FACTORS

4.1.1 Statistical Performance Attributes

Performance is evaluated relative to two major attributes; first, the sensitivity of the method to detect a true trend or change in air quality[8]. The second attribute of interest is the false "alarm" rate[9] whereby a change or trend is assessed when, in fact, no change has occurred. Ideally, the method would "always" detect a change, (i.e. 100% sensitivity), and never incorrectly identify that a change occurred (i.e. 0% false alarms).

These objectives are achievable, in principle, if the magnitude of trend is large and there is sufficient data. However in general, there is a compromise between the ideal objectives for the two attributes. Increasing the sensitivity, either through increasing the required probability and/or decreasing the magnitude of change that needs to be detected will result in increasing the chances of incorrectly assessing that a change has occurred.

4.1.2 Factors that Affect Statistical Performance

Factors that are not controllable are properties of the data:

a) the magnitude of the trend present in the pollutant: a larger trend is easier to find than a smaller trend.
b) natural variation from year to year: The larger the variability from the year-to-year, the lower will be the sensitivity to detect a change; that is, the "signal" of a trend can get lost in the "noise" of the natural variation.

---

[8] The formal term would be the statistical power which is the probability of a correct decision of rejecting the null hypothesis when the alternate hypothesis is true.

[9] The alpha decision error rate, or the Type I Error rate, which is the probability of a decision error of rejecting the null hypothesis when it is true.

Factors that are controllable include:

a) Number of years with data: Typically, the probability of detecting a trend increases with increasing amounts of data.
b) Required statistical confidence level. Setting the confidence level for the statistical method establishes the false alarm rate for the method. For example, increasing the required confidence level (and thereby reducing the false alarm rate) will reduce the sensitivity to detect a trend or change of interest.

4.2 PROBABILISTIC SIMULATION

4.2.1 Factors Considered

The performance of the method has been evaluated for the following combinations of factors that affect the performance:

a) Magnitude of the trend (0,1,2,5,10 %/y) Large trends are easier to detect than smaller trends with the values selected to identify probability detection for the target trend of 2% per year as well as trends that are smaller or larger than this target.

b) Normal Year-to-Year variability (5%, 10%). Trends are easier to detect with small variation from year-to-year compared to detecting trends with larger variation from year-to-year. The example data suggest that the natural variation is typically 5% however the variation may range up to about 10%.

c) Trend Length (5, 10, 15 years). Trends are easier to detect with a longer period of record provided the trend is linear over the period. A review of the example data suggests that the trends are reasonably linear over periods of about 10 years. Periods from 5 to 15 years were considered.

d) Confidence Level (95%, 99%). The higher the confidence level, the more difficult it is to detect a trend (e.g. assess as statistically significant) and the lower the false positive rate. A confidence level of 95% is often used; however different confidence levels could be used depending on policy or consequence of decision errors. If it is important not to have false alarms, a confidence level of 99% was considered.

4.2.2 Simulation Procedure

The approach to probabilistic simulation of performance is to conduct several trials of the method on simulated datasets that have the same statistical properties as the monitoring data.

The simulated dataset for a trial is constructed with desired number of years (e.g. 5, 10 or 15). The expected concentration in each year is calculated using the trend that is being simulated (e.g. 2% per year). Random variation is applied to the expected value for each year by using a probabilistic sampling of the year-to-year variation that is being simulated (e.g. 5% or 10%).

The annual method is applied to this simulated data using the confidence level (e.g. 95% or 99%) being considered for the trial. An assessment of the trend is conducted and a record is kept of the assessment results, in particular whether the simulated trend was statistically significant or not.

Two hundred trials (200) were conducted for each combination of trend magnitude, amount of normal variation, length of trend and confidence level. The proportion of these trials was summarized for each combination to estimate the power of the method to detect trends.

4.3 PERFORMANCE RESULTS

Table 4.1 shows a summary of the probability of detecting a trend of 2% /y with trend lengths between 5 and 15 years and year-to-year variability ranging from 5% to the upper bound of 10%. For example, there is 95% probability of detecting the target trend with 10 years of data containing the typical variation of 5% from year-to-year. Detection probabilities for smaller trends will be lower and detection probabilities for large trends with be higher than the values shown in the table for a trend of 2%/y.

TABLE 4.1
PROBABILITY OF DETECTING A 2%/y TREND

Number of Years	5% Variation from Year-to-Year	10% Variation from Year-to-Year
5	0.23 (0.11)	0.24 (0.07)
10	0.95 (0.68)	0.40 (0.21)
15	1.0 (1.0)	0.88 (0.68)

Note: the value in brackets is the probability of detection using 99% confidence intervals in the methods.

4.4 DISCUSSION

The simulation results indicate that the method will be able to detect the 2%/y trend with trend lengths of 10 to 15 years with year-to-year variability approaching the upper bound of 10%. There is some limited power of detecting the target 2%/y trend with 5 years of data if the year-to-year variability is approximately 5%.

The recommendation is to have 10 years of data as a target for applying the statistical method. The probability of detecting a 2%/y trend is high over the range of anticipated year-to-year variability and the assumption of an approximately normal trend over 10 years is stronger than the same assumption over 15 years.

A confidence interval of 95% is recommended as the 99% confidence level appreciably decreases the probability of detection.

5.0 RECOMMENDATION FOR ANNUAL REPORTING

5.1 METHODOLOGY

5.1.1 Assessment of Trend and Outliers

The following paragraphs summarize the statistical method as described in greater detail in Chapter 3.

Data are extracted for the recent period (e.g. 10 years) and the mean, the 98th percentile and the proportion of pollutant values exceeding the average period guideline are summarized. The coverage, percentage of the time periods within the year is determined and the annual statistics are not used if less than 80% of the time periods have valid measurements.

A robust regression method is used to fit a linear (straight-line) trend over a 10 year period of data. The regression method identifies outliers, years with unusually high or low concentrations relative to the general pattern, and estimates the best fitting trend without these outliers. The statistical significance of the trend is assessed based on 95% confidence intervals of the estimated trend.

The method is used to establish 95% confidence limits for the annual values depending on whether the trend was statistically significant or not. If the trend is statistically significant, then the upper and lower limits are developed using the confidence intervals for the regression prediction error.

If the trend is not statistically significant, the mean is assumed constant over the time period. The mean concentration, not including the outliers identified during the regression, is calculated as is the standard deviation of the concentrations. Upper and lower limits are calculated based the observed mean, standard deviation and appropriate t-statistic for the selected confidence level.

Outliers or unusual values are assessed if the annual value is outside the confidence limits.

5.1.2 Visual Charting

Following assessment of trends and outliers, a plot can be prepared showing all annual values, the assessed pattern (trend or constant mean) and the upper and lower confidence limits. The reference lines (e.g. confidence limits for annual values) demonstrate the range of values and patterns. Outliers are shown with a special symbol.

Examples of these charts are shown in a following section.

5.3.3 Interpretation

Interpretation for annual reporting is limited to identifying whether a statistically significant trend was present or a statistical outlier was detected. The following table shows the interpretation for the four possible outcomes using the method.

TABLE 5.1
INTERPRETATION OF FOUR POSSIBLE OUTCOMES OF THE PROPOSED STATISTICAL METHOD

	Value for the Most Recent Year	
Presence of Recent Period Trend	**Not an "Outlier"**	**An "Outlier"**
Trend not statistically significant	*the measured concentration was within normal year-to-year variation*	*the measured concentration was outside (high/low) of the normal range but there is no evidence that a trend is occurring*
Trend statistically significant	*there has been a (increasing/decreasing) trend in concentrations over the recent period*	*there has been a (increasing/decreasing) trend in concentrations over the recent period and the most recent value is a departure (high/low) from this trend*

Given that there are multiple stations with multiple pollutants, a summary tabulation would be useful to many report users. The table would include an indication of whether a trend or outlier was present for that particular pollutant, location and annual statistic. If a trend was present, the trend would be quantified in percentage terms with rounding to one significant figure. The direction of outliers would be provided; however, additional discussion can not be included in the summary table itself due to space and other limitations. Table 5.2 shows suggested entries into a summary table. Conceivably, the most recent annual value could be included in the table so that a comparison of absolute concentrations between monitoring locations or against the guideline could be made.

TABLE 5.2
TABLE ENTRIES FOR SUMMARY ASSESSMENT TABLE

	Annual Value for the Most Recent Year	
Presence of Recent Period Trend	**Not an "Outlier"**	**An "Outlier"**
Trend not statistically significant		*"Outlier"*
Trend statistically significant	Trend rounded to single percentage point (e.g., *3% per yea decreasing*)	trend rounded to single percentage point noting that outlier is present (e.g., *3% per year decreasing [Outlier]*)

There are three classifications to be considered in the layout of the summary table: statistic (i.e. mean, 98th, proportion), pollutant measure (e.g. 1-hour NO_2, 24-hour SO_2, etc) and monitoring location. The layout depends on the user; for example, a resident may prefer to have all pollutants and statistics organized by monitoring location so that all values for a nearby location are in the same area of the table. Some users may want to compare the results for individual statistics between monitoring locations while others may be want to organize by pollutant.

A suggested layout is to organize by pollutant with monitoring locations being rows and the columns indicating the statistic. In this way, comparisons with other locations are easily facilitated by scanning down the rows and all statistics are in the same area of the table. An example of this summary table follows.

5.2 EXAMPLE APPLICATION

The example data from Victoria Topaz and GVRD are used to demonstrate the method. The most recent data, up to 10 years duration, are used after vetting out those years that don't meet the data completeness objectives (i.e. number of missing time periods is too large). The method is applied when there are at least 5 years of data.

5.2.1 Example Summary Table

Table 5.3 summarizes the assessment of the pollutant concentrations measured over the most recent year period (1995-2004) for the GVRD and Topaz stations. Decreasing trends are apparent for CO and NO_2 concentrations at the GVRD location and for SO_2 at both the GVRD and Topaz monitoring locations.

TABLE 5.3
SUMMARY OF TRENDS AND CHANGES FROM TYPICAL LEVELS FOR EXAMPLE DATA

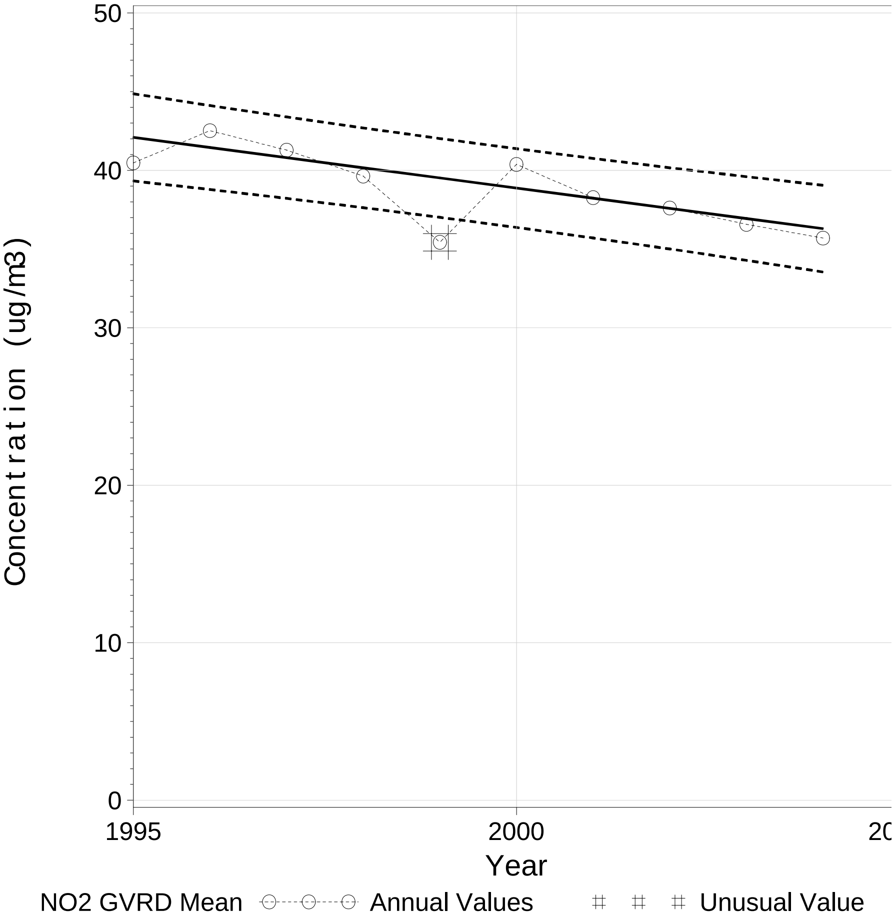
Location	Mean Trend	98th Percentile Trend	Proportion over Guideline
CO			
GVRD	3%/y decreasing	5%/y decreasing	No trend
Topaz	No trend	No trend	No trend
NO_2			
GVRD	2%/y decreasing	1%/y decreasing	No trend
Topaz	n/a	n/a	n/a
Ozone			
GVRD	No trend	No trend	No trend
Topaz	No trend	No trend	No trend
PM_{10}			
GVRD	No trend	2%/y decreasing	No trend
$PM_{2.5}$			
GVRD	n/a	n/a	n/a
Topaz	No trend	No trend	No trend
SO_2			
GVRD	No trend	5%/y decreasing	No trend
Topaz	12%/y decreasing	16%/y decreasing	No trend

5.2.2 Example Charts

Figure 5.1 shows the chart for the pattern of decreasing trend of mean NO_2 concentration observed at the GVRD monitoring location. The regression fitting identified the mean concentration in 1999 as an outlier and the trend was recalculated excluding that value. The trend line calculated by the regression and the upper and lower limits for normal variation about this trend line are shown. The magnitude of the statistically significant trend was a decrease of 2% per year. The measurement from 1999 is an outlier from this pattern.

An interpretation is that "*there has been a decreasing trend in concentrations over the recent period*". The entry for the summary table would read *" 2%/y decreasing"*.

FIGURE 5.1
MEAN NO_2 CONCENTRATIONS AT GVRD MONITORING STATION (1995-2004)



Note: Reference lines show the typical range of concentrations at this site.

Figure 5.2 shows the chart for mean CO concentration at the Topaz at the GVRD monitoring location. The mean concentration measured in 2004 was found to be unusual during the trend fitting. The trend with this value removed was almost statistically significant[10], however, there was insufficient evidence to assess that a trend occurred. Confidence limits were calculated assuming that no change occurred over the time period and these limits and the mean value ware shown on the graph.

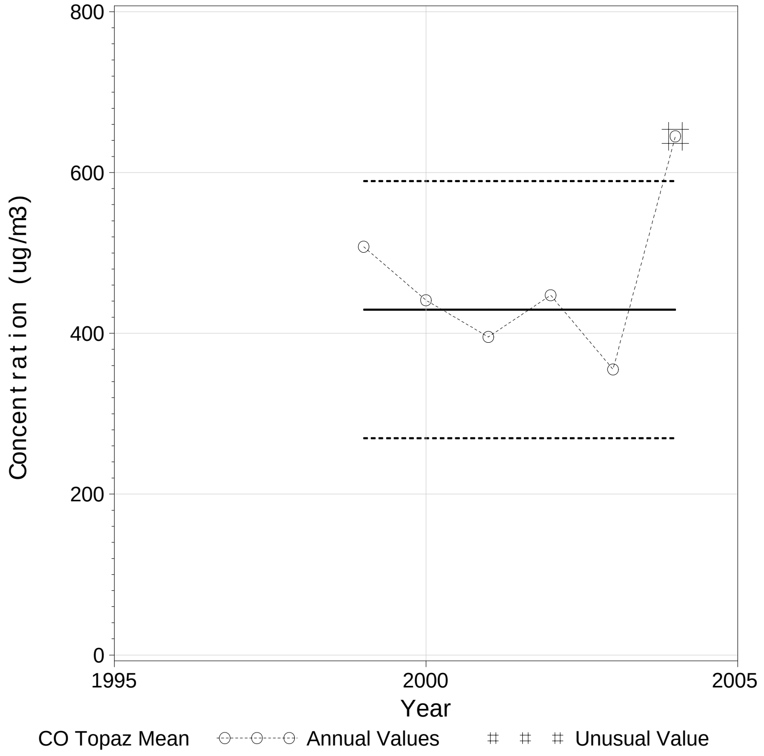
An interpretation is that "*the measured concentration was outside (high/low) of the normal range but there is no evidence of a trend*". The entry to the summary table would be "*Outlier*".

---

[10] A p-value of 0.08 compared to the 0.05 target level

The measurement in 2004 could be the result of a number of causes including episodic events (e.g., construction activities) or a change in instrumentation procedures. Further investigation would be required if there were a need to identify potential causal factors.

FIGURE 5.2
MEAN CO CONCENTRATIONS AT TOPAZ MONITORING STATION (1999-2004)



5.3 IMPLEMENTATION SCHEDULE

The target period for assessment is the last 10 years of data as this provides high probability of detecting a 2%/y trend over the anticipated range of normal variability. However, the methods are able to detect larger trends, and 2% per year trends if the year-to-year variability is small, with a shorter period of record. It is recommended that the method be applied once there is 5 years of data.

5.4 INHERENT LIMITATION OF THE PROPOSED STATISTICAL METHOD

A limitation of the proposed statistical method is that it identifies the magnitude of statistically significant trends and the unusual values but does not address important questions that may arise once statistically significant patterns are identified including:

i) the consequences of these changes;
ii) the reasons the pattern is occurring; and,
iii) actions that will be undertaken because of these changes.

The responses to these types of question are, in general and inherently, beyond the scope of a standardized trend assessment method. In some cases, more detailed analyses that are specific to the pollutant and location could answer some, or part of some, of these questions using more sophisticated and focussed models. However, in many cases the responses and actions will be necessarily driven by qualitative methods and public policy.

It is suggested that the annual analysis provide a basis for identifying concerns that, based on policy and/or a quantitative trigger, would initiate as special study to further investigations. This would include for example, increasing trends that exceeding a particular value or otherwise identified concerns, in addition to the typical statistical analysis (e.g., summarizing maximum, minimum, mean, and percentile concentrations) that is conducted for each annual CRD air quality report.

A periodic report with more additional analyses that provide greater context to the air quality patterns could be conducted on a 3-5 year cycle. This report could assess longer period trends that are more complex than a linear (straight-line) assumed over the 10-year period and in doing so provide more power to detect smaller trends and to show how the trend changes over time. Spatial-temporal trend models could be included to assess whether concentrations at one location were different than the trends at other locations. Such a report would include more detailed results from the analyses than presented in the annual report. Other elements of such a report may include interpretation on causes, consequences and actions based on policy perspective and the results of the more detailed analyses.

6.0 CONCLUSION

A statistically-based method has been proposed that identifies the presence of statistically significant patterns in annual air quality statistics. The method is considered to be robust, has a reasonably high probability of detecting trends of magnitude 2%/year or greater in annual mean and 98th percentiles, and is simple enough so that the method and the outcome of applying the method can be understood by a broad audience. The method uses a 95% confidence interval to assess trends and unusual values (outliers), implying that there is a 5%, or less, chance of incorrectly identifying either a trend over a 10 year period or an unusual value in any given year.

The method uses simple linear regression to identify statistically significant trends. The approach can be implemented with many available software products and is recommended for use with data sets for 10 year time periods. The method has useful probability of detecting trends with less than 10 years of data and is recommended to be implemented once 5 years of data are avialable.

The method is not intended to assess trends over time periods exceeding 15 years or to suggest the importance of, or reasons for, the trends. Such analyses would be part of a more detailed investigation as the magnitude and direction of trends can change substantially over longer time periods. Depending on the nature of trend and the purposes for further investigations, many statistical methods are available.

APPENDIX A SAMPLE APPLICATION OF METHOD

This appendix demonstrates the details of the method for the example applications that are described in the main text.

A.1 MEAN NO$_2$ AT GVRD

A.1.1 ASSESSMENT

Data

The annual mean values were calculated by averaging the averaging-period concentrations of interest. In the case of NO$_2$ these are hourly averaging periods. Each year had averaging period measurements for more than 80% of the possible number of averaging periods so all annual values from the 10-year target period are used.

```
                        Data

                            Averaging
               Annual        Periods
       Year    Value        with Data

       1995     40.5           7902
       1996     42.5           8509
       1997     41.3           8461
       1998     39.6           8587
       1999     35.4           8459
       2000     40.4           8584
       2001     38.3           8556
       2002     37.6           8366
       2003     36.6           8579
       2004     35.7           8538
```

Trend Fitting

An initial regression fitting is conducted on the data to determine if there are any “outliers” in the data using a cross-validation approach that is provided by many statistical packages. The approach is to re-calculate the regression with each observation in turn excluded from the model fitting. The difference, or residual, between the observed annual value and the predicted annual value using a regression model based on the other data is determined. This residual is expressed as a studentized residual such that it is distributed as a standard normal variation (e.g. mean 0 with a standard deviation of 1). High or low values of this standard normal variate are unlikely to occur by chance alone and suggest that the observation is unusual relative to the pattern (i.e. trend) in the data.

For this method application, studentized residuals greater than 3 are assessed as being statistical outliers in the data (although other limits could be used). For the annual mean NO2 at the GVRD location, the 1999 value has a studentized residual of -3.9 and is therefore assessed as an unusual measurement and excluded from the regression modelling.

```
               Cross-validation For Outlier Assessment

                 Annual             Studentized
                 Value    Year      Residual

                   40.5    1995       -0.8
                   42.5    1996        1.1
                   41.3    1997        0.6
                   39.6    1998       -0.0
                   35.4    1999       -3.9
                   40.4    2000        1.3
                   38.3    2001        0.2
                   37.6    2002        0.2
                   36.6    2003       -0.0
                   35.7    2004       -0.2
```

The regression model is fit with the value from 1999 excluded and selected output is shown below. The trend model has an intercept of 1327 μg/m^3 /year with a slope (trend) per year of -0.64 μg/m^3 /year. The statistical significance of this slope is provided by the probability of the observed slope occurring by chance alone if the null hypothesis of the slope was zero. In this case, the probability of an observed slope that is -5.83 standard deviations from the null is 0.0006. Since the probability of 0.0006 is smaller than the criterion for the method (0.05), the trend is assessed as statistically significant. Further details are provided below.

```
                             Analysis of Variance

                                    Sum of           Mean
Source                   DF        Squares         Square    F Value    Pr > F

Model                     1       34.10977       34.10977      33.94    0.0006
Error                     7        7.03515        1.00502
Corrected Total           8       41.14492


             Root MSE              1.00251    R-Square     0.8290
             Dependent Mean       39.16154    Adj R-Sq     0.8046
             Coeff Var             2.55993


                          Parameter Estimates

                         Parameter       Standard
     Variable     DF      Estimate          Error    t Value    Pr > |t|

     Intercept     1    1327.05005      221.06855       6.00      0.0005
     year          1      -0.64409        0.11056      -5.83      0.0006
```

Confidence Intervals and Identification of Outliers

Since the trend is statistically significant, the confidence intervals are based on the prediction intervals for individual observations. The choice for confidence level was 95% for the method. These confidence intervals are a standard output from most statistical packages and are constructed about the trend line. The trend and confidence intervals describe normal variation about the estimated trend and are retained for charting.

Year	Annual Value	Expected Value	Upper 95% Confidence Limit	Lower 95% Confidence Limit	Outlier Assessment
1995	40.5	42.1	44.9	39.3	-
1996	42.5	41.5	44.1	38.8	-
1997	41.3	40.8	43.4	38.2	-
1998	39.6	40.2	42.7	37.6	-
1999	35.4	39.5	42.0	37.0	Below
2000	40.4	38.9	41.4	36.4	-
2001	38.3	38.2	40.8	35.7	-
2002	37.6	37.6	40.2	35.0	-
2003	36.6	36.9	39.6	34.3	-
2004	35.7	36.3	39.1	33.5	-

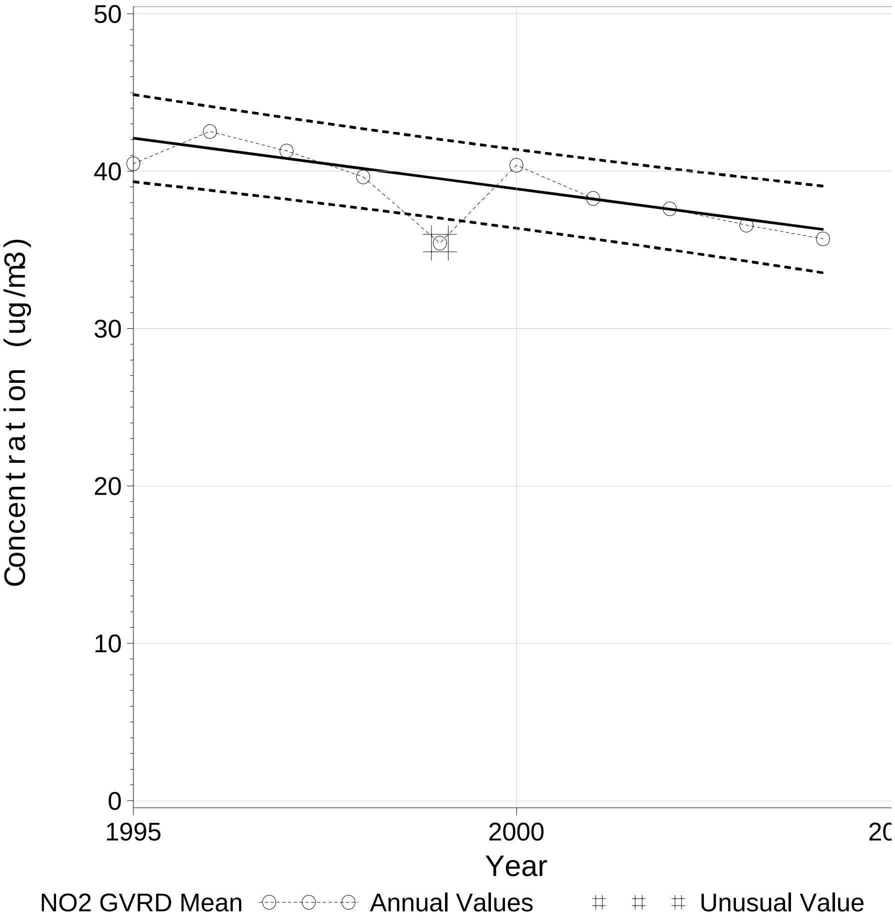
The observed annual values are compared to the confidence intervals to assess outliers. The value of 35.4 µg/m^3 falls outside the confidence interval of 37.0 to 42.0 µg/m^3 for individual observations and is assessed as an outlier than falls below the range of normal variations.

A.1.2 REPORTING

Charting

Since the trend is statistically significant, a chart is created showing the annual values, the estimated trend, the confidence interval about this trend and those annual values which fall outside the normal variation about the trend and are assessed as outliers. It is important that the concentrations axis start at the origin (i.e., 0 µg/m^3) to ensure the appropriate representation of relative variability and magnitude of the trend. Annual mean 1-hour concentrations from 1995 – 2004 are shown in Figure A.1 with a solid line representing the trend and dashed lines representing the confidence intervals, as determined from application of the method.

Figure A.1
Application of the Method to Annual Mean 1-hour NO_2 Concentrations at GVRD Station 18



Summary of Assessment

The trend is expressed in percentage terms as this is generally of more interest than the absolute trend. The mean annual value over the 10 year period, after excluding the outlier in 1999, was 39.2 µg/m^3 so the average trend on a percentage basis would be -2%/y (i.e. 100% times -0.64/39.2).

The 2004 general assessment for these annual values is "*there has been a decreasing trend in concentrations of 2% per year over the last 10 years"*. For a summary table, the entry relative to assessment would be "*2%/y Decrease*". The outlier in 1999 would not need to be mentioned as it did not occur in the most recent year.

A.2 Mean CO at Topaz

A.2.1 ASSESSMENT

Data

The annual mean values were calculated by averaging the averaging-period concentrations of interest. In the case of CO these are 8-hour sequential averaging periods: a total of 1095 (i.e., 8760 hours per year / 8 hour averaging period.) are possible. Each year had averaging period measurements for more than 80% of the possible number of averaging periods so all annual values from the available data are used. The period of record is less than the target of 10 years but is at least 5 years, so the method can be applied to this dataset, although with reduced statistical power.

```
             Data

                  Averaging
        Annual     Periods
Year    Value     with Data

1999      508         997
2000      441         908
2001      396        1030
2002      447        1001
2003      355        1030
2004      645        1023
```

Trend Fitting

The initial regression to identify outliers using cross-validation was conducted. For the annual mean CO at the GVRD location, the 2004 value has a studentized residual of 5.6 and is therefore assessed as an unusual measurement and excluded from the regression modelling.

```
Cross-validation For Outlier Assessment

  Annual           Studentized
  Value    Year    Residual
     508    1999        1.0
     441    2000       -0.0
     396    2001       -0.6
     447    2002       -0.2
     355    2003       -1.7
     645    2004        5.6
```

The regression model is fit with the value from 2004 excluded and selected output is shown below. It can be seen that the probability of an observed slope that is -2.49 standard deviations from the null is 0.0884. Since the probability of 0.0884 is slightly larger than the criterion, for this method, of 0.05, the trend is assessed as not statistically significant. It is noted that the trend was close to being statistically significant.

```
                              Analysis of Variance

                                        Sum of           Mean
        Source                   DF        Squares         Square    F Value    Pr > F

        Model                     1     8942.37662     8942.37662       6.20    0.0884
        Error                     3     4324.76224     1441.58741
        Corrected Total           4          13267


                     Root MSE             37.96824    R-Square     0.6740
                     Dependent Mean      429.42884    Adj R-Sq     0.5654
                     Coeff Var             8.84157


                              Parameter Estimates

                                 Parameter       Standard
             Variable      DF     Estimate          Error    t Value    Pr > |t|

             Intercept      1        60267          24025       2.51      0.0871
             year           1    -29.90381       12.00661      -2.49      0.0884
```

Confidence Intervals and Identification of Outliers

Since the trend is not statistically significant, the assessment is that the trend is zero. In this case, the mean and standard deviation of the data (excluding the value from 2004) are calculated. The 95% confidence intervals for normal variation are calculated using the t-statistic for 95% confidence and the degrees of freedom (4 in this case). Results of the calculation are shown below.

```
                                                      Upper         Lower
                               Number   t-Statistic     95%           95%
      Mean     Standard          of       for 95%     Confidence    Confidence
      Value    Deviation        Data    confidence      Limit         Limit

       429       57.6             5        2.78          589           270
```

The range for the normal variation is constant across the period of data and is retained along with observed annual value for charting purposes. The observed annual values are compared to the confidence intervals to assess outliers. The value of 645 μg/m^3 observed in 2004 falls outside the confidence interval of 279 to 589 μg/m^3 for individual observations and is assessed as an outlier than falls above the range of normal variation, as shown below.

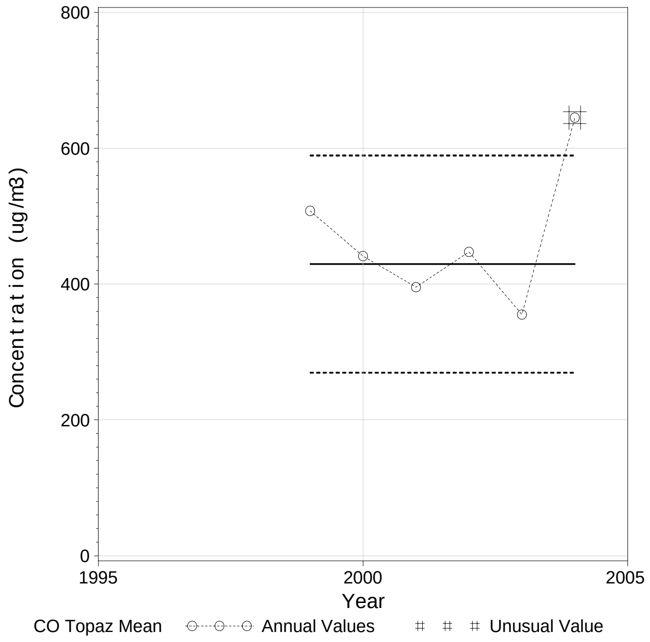
Year	Annual Value	Expected Value	Upper 95% Confidence Limit	Lower 95% Confidence Limit	Outlier Assessment
1999	508	429	589	270	-
2000	441	429	589	270	-
2001	396	429	589	270	-
2002	447	429	589	270	-
2003	355	429	589	270	-
2004	645	429	589	270	Above

A.2.2 REPORTING

Charting

Since the trend was not statistically significant, a chart is created showing the annual values, the mean annual value over the time period, the confidence interval about this mean and those annual values which fall outside the normal variation about the trend and are assessed as outliers. It is important that the concentrations axis start at the origin (i.e., 0 μg/m^3) to ensure the appropriate representation of relative variability and magnitude of the trend.

Figure A.2
Application of the Method to Annual Mean 8-hour CO Concentrations at Victoria Topaz



Summary of Assessment

The estimated trend was not statistically significant therefore is assumed that there is no trend. The most recent value was outside the confidence interval for normal variation of the annul values therefore is assessed as an outliers. The assessment is "*the value in 2004 was above the normal range but there is no evidence that a trend is occurring over the most recent time period"*. The summary table entry would be "*Outlier*".

APPENDIX B TRENDS IN PROPORTION OF AVERAGING PERIODS THAT EXCEED A GUIDELINE LEVEL

The assessment of trend, particularly the trend over time periods of approximately 10 year, is required for the annual proportion of averaging periods that exceed the CRD guideline. The statistical properties of proportion can differ from statistics such as the mean or 98th percentile. While the assumption of normal variation from year-to-year is appropriate in some cases, the current monitoring data from CRD indicate the proportion of measurements that exceed guidelines is rare. This appendix discusses statistical aspects for fitting trends to this type of data. The conclusion is that the method proposed for means and 98th percentiles performs adequately for proportions as well.

B.1 PROPORTION DATA

There are several averaging periods (trials) during each year that are compared to the CRD guidelines. In some cases, the averaging period concentrations will exceed the guidelines (events). For each year, the number of theses events is divided by the number of trials to calculate the proportion of averaging periods exceeding the guideline. It is important to express the comparison to guidelines on a proportion basis rather than number of events for two reasons. First, the number of averaging periods is not the same for different pollutants (e.g., 8760 for one-hour averaging period to a far lower number of trials for PM_{10}). Second, statistical methods for this type of data required consideration not only of the number of events but also the number of trials.

Of interest, is whether this proportion changes over time, that is, if a trend is present, and whether the proportion in a particular year is unusual relative to the general pattern.

B.1.1 Target Trends for Detection

The specification of a trend in proportion of averaging periods exceeding CRD guidelines is more complex than for trends in annual means or annual 98th percentiles where the target was to detect a trend on the order of 2%/y. This complexity arises, in particular, when the number of events and proportion is very small (e.g. 0 or close to 0 events during a year). For example, if the there were on average 2 events per 1000 averaging periods (a proportion of 0.2%), a relative trend of 2% change in number of events would be a change of 0.2 events per year over the 10 year period. This would be a very small change that would be difficult (if not impossible) to detect and, furthermore, would have little consequence from a policy perspective and would be difficult to communicate.

Based on policy and practical importance of trend, the following targets are considered relative to trends in proportion of averaging periods exceeding guidelines:

i) if the average number of events per year is less than 10:
  - a change of 2 events over a 10 year period might be considered important.;

ii) if there are more than 10 events per year on average:
  - a trend of 2% change in number of events per year might be important. With 10 events per year, this would be an overall change of 2 events per year over a 10 year time period.

B.1.2 Reporting of Statistically Significant Trends in Reporting

Reporting the magnitude of statistically significant trends in the proportion exceeding guidelines on a percentage basis per year as was done for the mean and 98th percentile is not recommended for two reasons. First, some confusion can be raised by describing percentage changes on a relative basis for statistic for the absolute change in a statistic that is already on a percentage basis.

For example, consider an overall change of 2 events per year over a 10 year period for a pollutant with an average of 1 event per year based on 1000 trials. The average proportion would be 0.1% (i.e. 1/1000) and the absolute trend in proportion would be 0.02%/y (i.e. 0.2 / 1000). Expressing the trend on a relative basis as provided in for the mean and 98th percentile trends would result in reporting a relative trend of 20% per year (e.g. 0.02%/0.1% x 100%). Reporting a relative trend of 20% per year for an absolute value of 0.1% in the proportion would be confusing to many.

There are two options that may be useful to consider:

i) report the absolute trend in proportion (e.g. 0.02% per year); or,
ii) report the change in number of events per number of trials per year (e.g., 0.2 events/y per 1000 trials.)

The second option is preferred as it contrasts with the reporting of trends suggested for means and 98th percentiles.

B.2 TREND ASSESSMENT

When the proportion is well above zero, that is there are 5 or more events (averaging periods exceeding guidelines) during a year, the assumptions to the method used to assess trend for the mean and 98th percentile are met. However, with fewer events per year, the assumptions and approximations are not as robust.

There are statistical approaches that can be used to fit linear (straight-line) trends to proportion and confidence intervals for these patterns; however, they are not currently accessible for most statistical software. Implementation of these methods requires statistical expertise of specialized techniques that would substantially increase the resources required to assess trends and would be difficult to standardize to a reproducible method. Specifically, there would be subjective inputs to the assessments or probabilistic methods would be required.

Upon review of the example data and consideration of the practical importance of trends when the proportion is small, the use of the simple linear regression method, with slight modifications is considered adequate for the purpose of detecting trends in proportion of events over the CRD guidelines.

B.1.2 Statistical Methods

The logistic regression model is often used to assess trends in data concerning proportions arising from events and trials. The statistical technique allows the fitting of a relationship between proportion and factors of interest while explicitly recognizing the statistical properties of proportions. For example, it ensures that the proportion must be between 0 and 1 and incorporates the variance of the binomial distribution. The fitting of models is actually conducted on the logarithm of the odds-ratio (e.g. $\log (p / (1-p))$ rather than the probability itself. As a result of fitting a linear trend to the logarithm of the odds ratio, the fitted trend in the proportion (p) is not linear (i.e. a straight-line). In addition, readily available software does not provide algorithms to calculate the confidence intervals for predicting individual observations.

There are generalized linear models (GLM) that facilitate the calculation of linear trends in the proportions that can incorporate the binomial nature of count data and are available in commercial software (e.g. PROC GENMOD in SAS). These methods also have difficulties with count data that are zero particularly if the linear trend is such that it tends to go towards negative probabilities. This type of data is present, anticipated, in the CRD monitoring program.

Approaches have been described in the literature for developing confidence intervals from GLM or using Bayesian methods; however, these are currently not standardized and would be time-consuming. These approaches are not considered further relative to standardized annual reporting; however, they may be warranted, in some cases, for specific investigation.

B.2.3 Investigation of Various Methods using Example Data

The performance of the simple linear regression, logistic and GLM general were examined using example data sets. Because the data had very few exceedences of the CRD guideline values, the data were analyzed using exceedences of a value that is 80% of the CRD value, for illustrative purposes.

Figure B.1 shows the proportion of averaging periods above the CRD guideline observed at the Topaz location. The number of events exceeding 80% of the CRD guideline value varied between 1 and 9 with an overall average of about 5 events per year and there were no years with zero events exceeding the guideline. All three statistical methods fit similar trends to data; however, none of the trends were statistically significant. The GLM and logistic regression models had similar p-values of 0.14 and 0.19, respectively. Although not statistically significant at the 0.05 level, these p-values indicate that a trend as large as the one observed is unlikely to have occurred by chance alone 1 time out of 5. The p-value for the linear regression was 0.42. In this case, all three methods could assess the data and all found that the trend was not statistically significant; however the statistical significance was somewhat lower for the linear regression model than the other two methods.

Figure B.1
Proportion (%) of Ozone Averaging Periods
at Topaz That Exceeded 80% of CRD Guideline Value

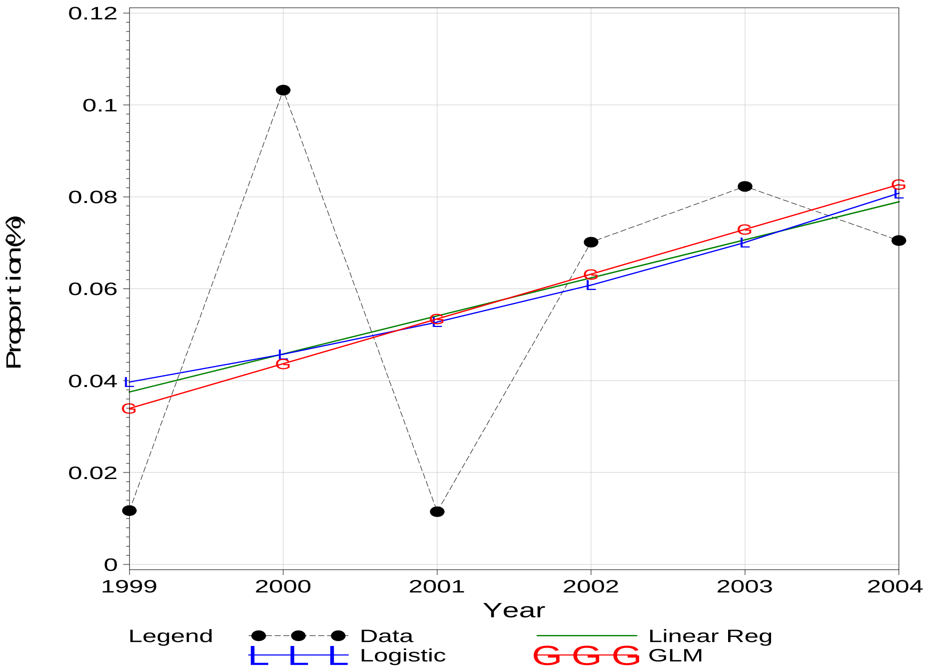
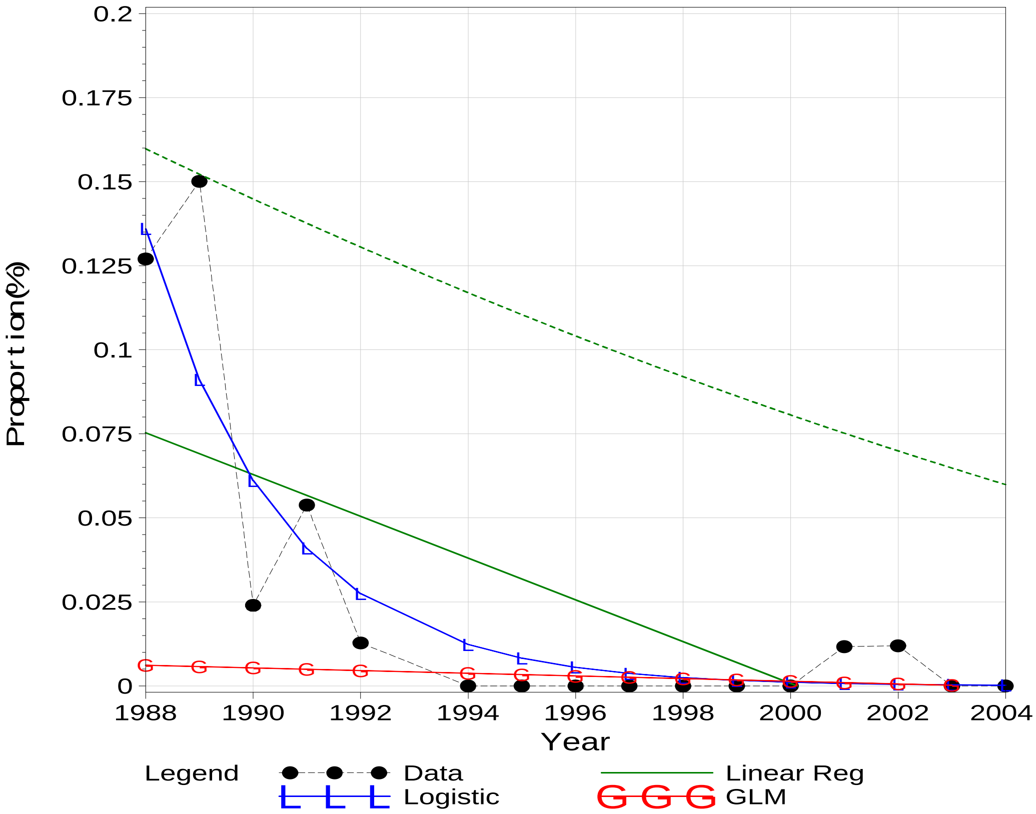


Figure B.2 shows the proportion of NO_2 averaging period measurements at the GVRD stations that exceeded 80% of the CRD guideline value. The numbers of events in the first two years were 11 and 13 and the number exceeding decreased from there to 0 events for most years after 1992, with the exception of 2001 and 2002 when there was one hour that was higher than 80% of the CRD guideline value in each year.

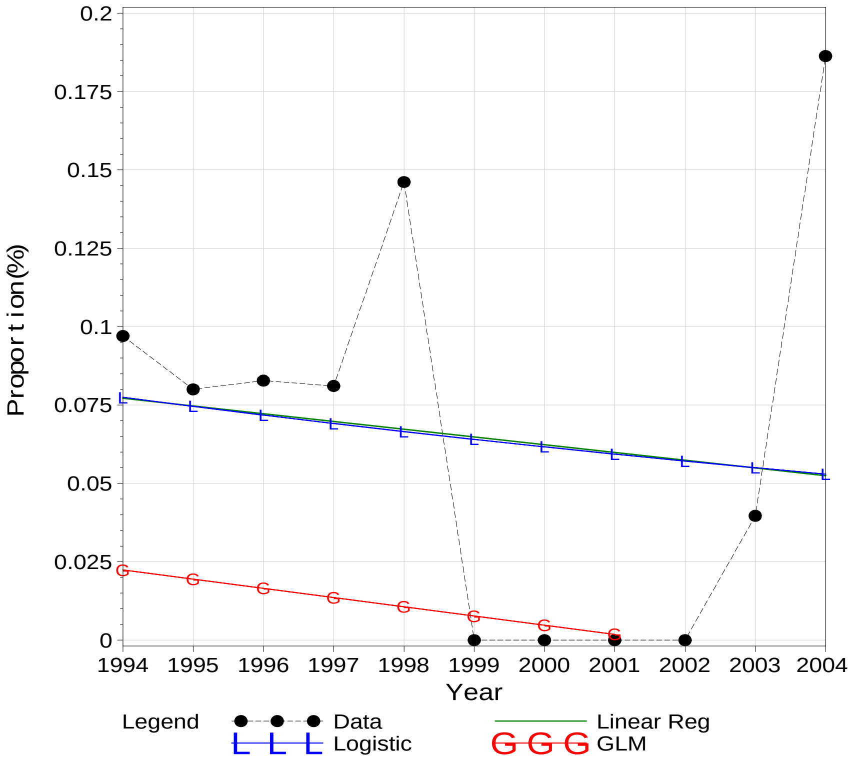
Figure B.2
Proportion (%) of NO_2 Averaging Periods
at GVRD Sample Location That Exceeded 80% of CRD Guideline Value



The logistic regression trend was statistically significant (p-value of <0.0001) and was steeply curved downward during the early part of the time period and then levelled out. The linear regression approximation showed a statistically significant decreasing trend over the time period (p=0.0036); however, the trend crosses the horizontal axis and would predict a negative proportion of averaging periods exceeding the guidelines after 2000. Obviously, this can not be true and is one of the limitations of the approximation with simple linear models when the expected number of events is low (e.g., <5). The GLM model could not accommodate this data with the requirement that the data were binomially distributed and the method produced a nonsensical estimate of trend. Although the GLM model could not fit the data; both the logistic and linear regression models assessed a statistically significant trend. The statistical significance was larger than the regression model.

Figure B.3 shows the proportion of events above 80% of the CRD guideline value for ozone at the GVRD sample location. The number of events ranged between 6 and 11 between 1994 and 1998, was zero from 1999 to 2002, and then reached 14 in 2004. The logistic and linear regression model fit almost the same trend to this data and in both cases the trend was not statistically significant. The GLM method failed to converge, likely due to the extended period of zero events per year.

FIGURE B.3
Proportion (%) OF Ozone Averaging Periods
at GVRD Sample Location That Exceeded 80% of CRD Guideline Value



B.3 PREFERRED APPROACH

Using the example data, the logistic and linear regression models provided similar assessment of the trends in proportion above guidelines; however, the logistic regression tended to find higher statistical significance than the linear regression model. This likely arises in part because it recognizes the binomial nature of the data, but also because it provides more flexibility than the linear regression when fitting a pattern to the data. The GLM method had problems with low numbers of events per year.

The logistic model however does not test, specifically, for a straight-line trend in proportion over the period, and does not usually provide confidence intervals for the individual observations. The simple linear regression approach is preferred as it is consistent with the approach used for mean and 98th percentiles, is readily accessible, and is considered to perform adequately with respect to detection of trends even when there are low numbers of events.

The potential limitations of the simple linear regression approach when the number of events is low (i.e., <5 events per year) could be mitigated through the following modifications to the method:

i) fit the pattern with logistic regression to determine if there is a statistically significant trend in the annual proportion;
ii) if the logistic regression is statistically significant, approximate a linear (straight-line) trend using simple linear regression;
iii) censor expected proportions that are below 1 event per year as less than ("<" the corresponding proportion;
iv) draw the linear trend line as calculated when above 1 event per year and then extend as a horizontal line at the proportion corresponding to one event per year;
v) report and chart lower confidence intervals calculated to be negative proportions as zero.

The determination of the statistical significance of a logistic regression is the least important, and potentially most difficult, of the modifications. It could be considered optional.

A specific caveat would be prudent for the approximation of the magnitude of the trend and the confidence limits when assessing proportions when the expected number of events per year is less than five. In addition, proportions for events that are rolling averages contain substantial serial correlation; for example, if one 8-hour rolling average exceeds the CRD guideline, then the following average will also have high probability of exceeding the guideline as both averages contain 7 of the same hours used to calculate the average. Although this is a further departure from the assumptions of the standard method, the performance will be appropriate for a standard method to assess annual trends in proportion.