**ATMOSPHERIC ISSUES & SERVICES BRANCH**

# REPORT


# An Investigation of Daily Maximum Ground-Level Ozone Concentration Forecasting using CART Statistical Techniques

Doug McCollor
Atmospheric Issues & Services Branch
Atmospheric Environment Service
Pacific Region

January, 1993
Report PAES-93-2

# AN INVESTIGATION OF DAILY MAXIMUM GROUND-LEVEL OZONE CONCENTRATION FORECASTING USING *CART* STATISTICAL TECHNIQUES

Doug McCollor
Atmospheric Issues and Services Branch
AES Pacific Region

January, 1993

## *ABSTRACT*

*CART (Classification and Regression Trees), a relatively new statistical analysis and prediction technique in the field of meteorology, was used to study ground-level ozone concentration data obtained during a forecasting experiment in the summer of 1992. Data from six ozone monitoring sites throughout Vancouver and the Lower Mainland were used in the study. It was found that the CART techniques were comparable, but not more favorable, with standard meteorological statistical techniques employed during the original experiment. This study also used time-series representation of the CART results and data. The analysis shows that the CART technique as used in this case relies strongly on the previous day's maximum ozone concentration level as a predictor, resulting at times in a one-day time lag response. The forecast ozone concentration values are also a result of smoothing by the CART statistical technique, diminishing extreme-valued events in the observed data set. As a result, two "poor" air quality events which occurred in the experiment were not predicted by any of the statistical methods studied to date.*

## *INTRODUCTION*

In the summer of 1992 an experiment was carried out to develop and test operating procedures for the production and dissemination of air quality advisories for ground level ozone in conjunction with the Greater Vancouver Regional District (GVRD). In this experiment a series of statistical packages which forecast ground level ozone concentrations at specific sites throughout the Lower Mainland were evaluated using standard statistical techniques (Lord, to be published 1993). A total of four objective statistical approaches and one subjective approach were assessed for ten established monitoring sites throughout the region. The reader should refer to Lord's work for details on this project regarding experiment impetus and design, background, and a complete discussion of statistical approaches. Lord recommends an examination of another statistical technique called CART (which stands for Classification and Regression Trees). The purpose of this report is to explore the CART statistical technique as an approach to forecasting ground-level ozone.

## *STATISTICAL OZONE FORECASTING MODELS*

The statistical methods previously analyzed are as follows. Taylor (1991) developed a set of multiple linear regression (MLR) equations to forecast daily maximum ozone concentrations based on temperature, yesterday's maximum ozone concentration, precipitation, and pressure differences between coast and interior regions in southwestern BC. The equations were developed using a dependent data sample of meteorological and ozone concentration parameters procured over the five year period 1985-1990. Robertson (1992) also developed a set of predictive equations based on MLR techniques as well as a Multiple Discriminant Analysis (MDA) technique using the same five year dependent data set. A fourth predictive model was developed based on a log-normal distribution of maximum ozone concentrations at each site. This model is called the Concord (Ciccone et al., 1992). The fifth method used to predict maximum ozone concentrations entailed subjective analysis and prognosis of synoptic weather patterns over the region using guidance from the Pacific Weather Centre (PWC). Once the meteorological conditions were forecast, the latest ozone concentration trends in the region were examined and a forecast of the maximum ozone concentration for each site was prepared. Results were summarized in a table of continuous variable verification statistics including sample mean, bias, standard deviation (SD), mean absolute error (MAE), root mean square error (RMSE), and reduction of variance (RV) for each statistical forecast model. The pertinent models were compared for each ozone monitoring location. In addition, contingency table statistics were compiled for each statistical model for each location, but were not summarily compared in the initial analysis. The present report expands the continuous variable statistical comparison of Lord to include the CART Regression Tree method. A comparison of all models using appropriate contingency table verification statistics is presented in this report, including both CART routines, Regression Tree and Classification techniques. The data supports a time-series analysis approach and a brief examination of each station's data in this format is presented.

## *CART APPROACH*

CART was used to find predictand classification and regression trees for each location using meteorological and previous-day maximum ozone concentration predictors. In this case the predictors chosen for CART were the same as those used in the other statistical models so that the same dependent data set of predictor values could be used.

In general, CART is able to produce predictand values using two statistical methods. The regression tree method uses the set of predictors to establish a decision tree diagram. CART uses the independent data as a "learning sample" of dependent data events, each consisting of a predictand and predictor values that occurred with the event. This sample is used to establish the statistically "best" decision tree with which to separate out the events into predictand values. When the tree is subsequently used with independent data, an event enters the top of the tree and falls down through a series of "nodes", each with a binary decision to send the event to the left or right, depending on a threshold value of a single predictor or a linear combination of predictors. Upon reaching a "terminal node" at the bottom of the tree, the corresponding predictand value is extracted and used as the forecast daily maximum ozone concentration for each location. Similarly, the classification method establishes a statistically-preferred decision tree composed of binary decision nodes. In this method, when the event decision procedure reaches a terminal node, a final classification is assigned to the event by choosing the category with the maximum number of learning sample events that accumulated there when the tree was constructed. The reader is referred to Burrows (1991) for a detailed meteorological application of the CART technique. For greater detail on CART the reader is referred to Breiman et al. (1984).

In the case at hand, predictors were chosen to be the same as those in the initial statistical models (see Table 1) for the same five year period. For the regression tree approach, CART terminal nodes corresponded to sample ozone concentration levels in ppb. These predictand values were used directly as input parameters in the analysis of continuous variable verification statistics used by Lord and compared directly with the other statistical models. The CART regression tree predictands were also incorporated into contingency table statistics and subsequently compared with the other models. For the classification tree approach, three classification categories were predetermined to correspond to established ozone air quality guidelines (i.e. 0-51 ppb is GOOD, 52-82 ppb is FAIR, and 83 ppb and above is POOR). The results were then tabulated directly under contingency table statistics and compared with the results obtained from the regression tree approach as well as the other models.

| | |
|---|---|
| YVR_YXX05 | Vancouver - Abbotsford pressure difference at 5am PDT |
| YVR_YAZ05 | Vancouver - Tofino pressure difference at 5am PDT |
| YVR_YHE05 | Vancouver - Hope pressure difference at 5am PDT |
| YVR_YYF05 | Vancouver - Penticton pressure difference at 5am PDT |
| YVR_YXX17 | Vancouver - Abbotsford pressure difference at 5pm PDT |
| YVR_YAX17 | Vancouver - Tofino pressure difference at 5pm PDT |
| YVR_YHE17 | Vancouver - Hope pressure difference at 5pm PDT |
| YVR_YYF17 | Vancouver - Penticton pressure difference at 5pm PDT |
| TEMPMX_YVR | Vancouver Maximum Temperature |
| TEMPMX_YHE | Hope Maximum Temperature |
| TEMPMX_YXX | Abbotsford Maximum Temperature |
| PRECIP_YVR | Measurable Precipitation at Vancouver (Y/N) |
| PRECIP_YHE | Measurable Precipitation at Hope (Y/N) |
| PRECIP_YXX | Measurable Precipitation at Abbotsford (Y/N) |
| T1OZHIER | Yesterday's Maximum Ozone Concentration at Location T1 |
| T9OZHIER | Yesterday's Maximum Ozone Concentration at Location T9 |
| T17OZHIER | Yesterday's Maximum Ozone Concentration at Location T17 |
| T15OZHIER | Yesterday's Maximum Ozone Concentration at Location T15 |
| T11OZHIER | Yesterday's Maximum Ozone Concentration at Location T11 |
| T12OZHIER | Yesterday's Maximum Ozone Concentration at Location T12 |
| YVR_TEMPAN | Climate Temperature Anomaly at Vancouver |
| YXX_TEMPAN | Climate Temperature Anomaly at Abbotsford |
| YHE_TEMPAN | Climate Temperature Anomaly at Hope |

**Table 1.**

Table of predictors used in CART approach.

## *ANALYSIS*

In this report, comparisons among the various statistical models were made for each of 6 specific locations within the lower mainland for which at least four model outputs were available. These sites are labeled with their respective GVRD identifier (T##) as well as their geographic name. It should be noted that for the Abbotsford site statistical models were developed on data obtained at Abbotsford Airport (identified as T11); however the monitoring site was moved to a downtown Abbotsford location (T28) prior to the 1992 experiment. According to Lord, this appears to have had little effect on the accuracy of the models. The effect of moving the monitoring site on the forecast performance of the models needs to be further investigated.

### Verification Procedures

The dependent meteorological and ozone concentration data were collected over 153 days of the summer ozone season of 1992. A total of 44 days of accumulated data for five of the sites was available to examine the model output. The sixth site, T17-Richmond South, had 34 days of accumulated data. Forecasts were verified against their matching observations in three separate ways. First, the numerical forecasts were verified using continuous variable statistics. Lord provides detailed results of these verifications. In this report results obtained from the CART Regression Tree approach are compared with those previously obtained and documented (see Table 2). Secondly, Robertson's MDA forecasts, and the numerical forecasts categorized into Good, Fair, and Poor air quality were verified using 3x3 contingency table statistics. This report gives further discussion on this verification method including both CART

Classification and Regression Tree techniques. See Appendix A for the contingency tables for each station. Table 2 holds a summary of the results of the contingency table statistics. Table 4 in this section presents a summary of the best models for each location according to the method of statistical verification. The table is presented for information only since the actual differences between models were small and considered operationally insignificant (Lord). As a third approach, graphical time-series representation of both CART techniques are presented and discussed. The time-series graphs can be found in Appendix B.

**SUMMARY OF CONTINUOUS VARIABLE VERIFICATION STATISTICS**

| LOCATION | FCST MODEL | MEAN | BIAS | SD | MAE | RMSE | RV |
|---|---|---|---|---|---|---|---|
| T1-Robson Sq. | Robertson's MLR | 18 | 0 | 8 | 6 | 8 | -0.06 |
| | Concord's | | -2 | 8 | 6 | 8 | -0.05 |
| | Taylor's MLR | | 1 | 8 | 6 | 8 | 0 |
| | Subjective | | -1 | 8 | 6 | 8 | -0.08 |
| | **CART (Regression)** | | **0** | **8** | **6** | **8** | **-0.03** |
| T9-Rocky Pt. Pk. | Robertson's MLR | 36 | 6 | 11 | 10 | 13 | 0.40 |
| | Concord's | | 9 | 13 | 12 | 15 | 0.12 |
| | Taylor's MLR | | 10 | 12 | 12 | 15 | 0.12 |
| | Subjective | | 8 | 14 | 12 | 16 | 0.07 |
| | **CART (Regression)** | | **10** | **13** | **12** | **16** | **0.00** |
| T12-Chilliwack | Robertson's MLR | 42 | 2 | 16 | 12 | 16 | 0.40 |
| | Concord's | | 2 | 15 | 11 | 15 | 0.48 |
| | Taylor's MLR | | 6 | 15 | 14 | 17 | 0.35 |
| | Subjective | | 5 | 17 | 13 | 17 | 0.30 |
| | **CART (Regression)** | | **2** | **15** | **11** | **15** | **0.47** |
| T15-Surrey East | Robertson's MLR | 40 | 4 | 10 | 9 | 11 | 0.33 |
| | Concord's | | 7 | 12 | 11 | 13 | 0.01 |
| | Taylor's MLR | | 8 | 11 | 11 | 13 | 0.04 |
| | Subjective | | 6 | 12 | 11 | 14 | -0.02 |
| | **CART (Regression)** | | **6** | **12** | **11** | **14** | **0.00** |
| T17-Richmond S. | Robertson's MLR | 34 | 2 | 10 | 9 | 10 | 0.19 |
| | Concord's | | 0 | 11 | 9 | 11 | 0.10 |
| | Taylor's MLR | | 4 | 10 | 8 | 10 | 0.20 |
| | Subjective | | 3 | 11 | 9 | 11 | 0.01 |
| | **CART (Regression)** | | **2** | **11** | **8** | **11** | **0.01** |
| T28-Abbotsford | Robertson's MLR | 40 | 7 | 13 | 12 | 15 | 0.29 |
| | Concord's | | 6 | 12 | 11 | 14 | 0.42 |
| | Taylor's MLR | | 9 | 13 | 13 | 16 | 0.24 |
| | Subjective | | 8 | 14 | 13 | 16 | 0.20 |
| | **CART (Regression)** | | **12** | **16** | **16** | **19** | **-0.19** |

**Table 2.**

A summary of the continuous variable statistics for models tested previously (Lord) is compared with CART Regression statistics from this report. The statistics shown are as follows: the MEAN is the average observed daily 1 hour maximum ozone concentration observed at the site during the 44 forecast days of the experiment, the BIAS is the average error in the forecast, SD is the standard deviation of the errors, the MAE is the mean absolute error, the RMSE is the root mean square error, and the RV is the reduction of variance of the errors. All numbers are reported to the nearest ppb.

## SUMMARY OF CONTINGENCY TABLE VERIFICATION STATISTICS

| LOCATION | FCST MODEL | % | POD | | | FAR | | | BIAS | | | HEIDKE | HEIDKE |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | COR | G | F | P | G | F | P | G | F | P | chance | climatology |
| T1-Robson Sq. | Robertson's MLR | 100 | 1 | - | - | 0 | - | - | 1 | - | - | - | - |
| | Concord's | 100 | 1 | - | - | 0 | - | - | 1 | - | - | - | - |
| | Taylor's MLR | 100 | 1 | - | - | 0 | - | - | 1 | - | - | - | - |
| | Subjective | 100 | 1 | - | - | 0 | - | - | 1 | - | - | - | - |
| | **CART (Regression)** | **100** | **1** | - | - | **0** | - | - | **1** | - | - | - | - |
| | **CART (Classification)** | **100** | **1** | - | - | **0** | - | - | **1** | - | - | - | - |
| T9-Rocky Pt. Park. | Robertson's MDA | 82 | 1 | 0 | - | 0.18 | - | - | 1.22 | 0 | - | 0 | 0 |
| | Robertson's MLR | 91 | 0.92 | 0.88 | - | 0.03 | 0.30 | - | 0.94 | 1.25 | - | 0.72 | 0.50 |
| | Concord's | 86 | 0.86 | 0.88 | - | 0.03 | 0.42 | - | 0.89 | 1.50 | - | 0.62 | 0.25 |
| | Taylor's MLR | 86 | 0.86 | 0.88 | - | 0.03 | 0.42 | - | 0.89 | 1.50 | - | 0.62 | 0.25 |
| | Subjective | 86 | 0.89 | 0.75 | - | 0.03 | 0.40 | 1 | 0.92 | 1.25 | - | 0.60 | 0.25 |
| | **CART (Regression)** | **84** | **0.83** | **0.88** | - | **0.03** | **0.42** | **1** | **0.83** | **1.50** | - | **0.59** | **0.13** |
| | **CART (Classification)** | **82** | **0.94** | **0.25** | - | **0.15** | **0.50** | - | **1.11** | **0.50** | - | **0.24** | **0** |
| T12-Chilliwack | Robertson's MDA | 75 | 0.85 | 0.44 | 0 | 0.09 | 0.56 | 1 | 0.94 | 1 | 3 | 0.37 | -0.10 |
| | Robertson's MLR | 75 | 0.85 | 0.44 | 0 | 0.12 | 0.60 | 1 | 0.97 | 1.11 | 1 | 0.33 | -0.10 |
| | Concord's | 82 | 0.88 | 0.67 | 0 | 0.09 | 0.45 | - | 0.97 | 1.22 | 0 | 0.51 | 0.20 |
| | Taylor's MLR | 73 | 0.74 | 0.78 | 0 | 0.07 | 0.59 | - | 0.79 | 1.89 | 0 | 0.39 | -0.20 |
| | Subjective | 75 | 0.85 | 0.44 | 0 | 0.06 | 0.50 | 1 | 0.91 | 0.89 | 5 | 0.40 | -0.10 |
| | **CART (Regression)** | **86** | **0.88** | **0.89** | **0** | **0.03** | **0.38** | - | **0.91** | **1.44** | **0** | **0.65** | **0.40** |
| | **CART (Classification)** | 75 | **0.88** | **0.33** | **0** | **0.12** | **0.62** | **1** | **1** | **0.89** | **2** | **0.31** | **-0.10** |
| T15-Surrey East | Robertson's MDA | 82 | 0.86 | 0.67 | - | 0.09 | 0.45 | - | 0.94 | 1.22 | - | 0.48 | 0.11 |
| | Robertson's MLR | 89 | 0.94 | 0.67 | - | 0.08 | 0.25 | - | 1.03 | 0.89 | - | 0.64 | 0.44 |
| | Concord's | 82 | 0.83 | 0.78 | - | 0.06 | 0.46 | - | 0.89 | 1.44 | - | 0.52 | 0.11 |
| | Taylor's MLR | 77 | 0.77 | 0.78 | - | 0.07 | 0.53 | - | 0.83 | 1.67 | - | 0.44 | -0.11 |
| | Subjective | 86 | 0.89 | 0.78 | - | 0.06 | 0.36 | - | 0.94 | 1.22 | - | 0.61 | 0.33 |
| | **CART (Regression)** | **82** | **0.89** | **0.56** | - | **0.11** | **0.44** | - | **1** | **1** | - | **0.44** | **0.11** |
| | **CART (Classification)** | 77 | **0.80** | **0.67** | - | **0.10** | **0.54** | - | **0.89** | **1.44** | - | **0.40** | **-0.11** |
| T17-Richmond South. | Robertson's MLR | 88 | 1 | 0 | - | 0.12 | - | - | 1.13 | 0 | - | 0 | 0 |
| | Concord's | 88 | 1 | 0 | - | 0.12 | - | - | 1.13 | 0 | - | 0 | 0 |
| | Taylor's MLR | 91 | 1 | 0.25 | - | 0.09 | 0 | - | 1.10 | 0.25 | - | 0.37 | 0.25 |
| | Subjective | 85 | 0.97 | 0 | - | 0.12 | 1 | - | 1.10 | 0.25 | - | -0.05 | -0.25 |
| | **CART (Regression)** | **85** | **0.97** | **0** | - | **0.12** | **1** | - | **1.10** | **0.25** | - | **-0.05** | **-0.25** |
| | **CART (Classification)** | 76 | **0.87** | **0** | - | **0.13** | **1** | - | **1** | **1** | - | **-0.13** | **-1** |
| T28-Abbotsford | Robertson's MDA | 80 | 0.88 | 0.60 | 0 | 0.06 | 0.40 | 1 | 0.94 | 1 | 3 | 0.51 | 0.18 |
| | Robertson's MLR | 82 | 0.88 | 0.70 | 0 | 0.09 | 0.42 | - | 0.97 | 1.20 | 0 | 0.54 | 0.27 |
| | Concord's | 86 | 0.91 | 0.80 | 0 | 0.06 | 0.33 | - | 0.97 | 1.20 | 0 | 0.65 | 0.45 |
| | Taylor's MLR | 82 | 0.82 | 0.90 | 0 | 0.04 | 0.44 | - | 0.85 | 1.60 | 0 | 0.59 | 0.27 |
| | Subjective | 82 | 0.91 | 0.60 | 0 | 0.06 | 0.40 | 1 | 0.97 | 1 | 2 | 0.55 | 0.27 |
| | **CART (Regression)** | **72** | **0.88** | **0.30** | **0** | **0.06** | **0.50** | **1** | **0.94** | **0.60** | 7 | **0.38** | **-0.09** |
| | **CART (Classification)** | 75 | **0.79** | **0.70** | **0** | **0** | **0.53** | **1** | **0.79** | **1.50** | **3** | **0.48** | **0** |

**Table 3.**

A summary of the contingency table statistics for models tested previously (Lord) is compared with CART Regression and CART Classification statistics from this report. The statistics shown are as follows: the Percent Correct is the total number of correct forecasts compared to the total number of forecasts, POD is the probability of detection (also termed Prefigurance). FAR is the False Alarm Ratio (equivalent to 1 - Post Agreement). Bias is the number of forecasts divided by the number observed for each category. Heidke Skill Score is given with respect to chance and with respect to climatology.

| | VERIFICATION METHOD | |
|---|---|---|
| LOCATION | CONTINUOUS VARIABLE | CONTINGENCY TABLE |
| T1-Robson Square | Taylor's MLR | all models equal |
| T9-Rocky Pt. Park | Robertson's MLR | Robertson's MLR |
| T12-Chilliwack | Concord's<br>CART Regression | CART Regression |
| T15-Surrey East | Robertson's MLR | Robertson's MLR |
| T17-Richmond South | Robertson's MLR<br>Taylor's MLR | Taylor's MLR |
| T28-Abbotsford | Concord's | Concord's |

**Table 4.**

Summary of best models for each location considering two statistical verification methods: Continuous Variable Verification Statistics and Contingency Table Verification Statistics.

**T1 - Robson Square**

Results for this location using both CART techniques were comparable with the other models examined. The continuous variable statistics indicate the CART Regression tree model output was essentially identical to that of Robertson's MLR model with a bias of 0 ppb, standard deviation of 8 ppb, mean absolute error of 6 ppb, and root mean square error of 8 ppb. The reduction of variance was slightly negative at -0.03, approximately in the middle of the range of the other models, with Taylor's MLR giving the best RV result of 0.

Little information was gleaned from the contingency table verification approach since all 44 observations fell in the "good" air quality category. All six models, including the two CART techniques, were 100% correct in forecasting the 44 events into the "good" category with no errors. This was the only one of the six sites that had all events fall into one category.

The time-series representation of the maximum daily ozone concentration levels and CART regression forecasts presents a view of the data that is unavailable in the previous statistical approach. The time-series representation allows a time-dependent correlation to emerge between the predicted and forecast values that is lost in both the continuous variable and contingency table statistics. In the case at hand (see Appendix B for time-series plots), even though all measurements and forecasts of maximum ozone concentration fall under the 52 ppb "good" air quality category, there are clear indications in the data that the previous day's ozone concentration level is a major predictor for the forecast value, at times to the detriment of the forecast. In particular, note the maximum ozone concentration value for the period July 18th to 21st. On July 18th, both observed and forecast values are near 20 ppb. On the 19th, the observed value increased sharply to 47 ppb while the forecast for that day remains at 21 ppb. On the 20th, the forecast value responds to the previous day's increase and reaches a value of 36 ppb, while in fact the observed value for the 20th has fallen off its one day peak and is measured at 20 ppb. The forecast for the 21st follows the same pattern by responding to the previous day's fall in ozone concentration and predicts a value of 21 ppb, while the actual observations have leveled off in the 15-20 ppb range. Therefore the predicted value for July 19th was under forecast by a factor of 2 and the value for the 20th was over forecast by a factor of 2, due to the influence of the previous day's observed ozone concentration at the site. In this particular case, all forecasts and observations were within the same air quality category. However, this one-day time lag evident in the predicted values has strong implications which will be seen at other locations during periods of rapidly changing ozone concentration levels.

**T9 - Rocky Point Park**

CART Regression results exhibited a bias of 10 ppb, a standard deviation of 13 ppb, a mean absolute error of 12 ppb, and a root mean square error of 16 ppb, which position it with the poorest forecast models for this site. The reduction of variance of 0 also places it at the bottom of the list of the available models.

Considering the contingency table verification statistics, Robertson's MLR method had the highest percent correct at 91%, with three of the other four models tied at 86%. In terms of Probability of Detection (POD), all five models ranked above 0.86 in the "good" category. Robertson's MDA ranked a perfect POD of one in the "good" category but fell to zero in the "fair" category. The other models ranked above 0.75 in the "fair" category. In the "poor" category, four models were equivalent in predicting no "poor" episodes with none observed. Only the subjective technique recorded a False Alarm Ration (FAR) of 1 by forecasting a "poor" event which did not occur. Heidke skill score with respect to chance was 0.72 and 0.50 with respect to climatology for Robertson's MLR; these values proved to be the best Heidke skill scores of the experiment for the six stations considered in this report.

Turning to the CART Regression results, with 84% of predictions correct, it ranks one above the poorest model, Robertson's MDA. In terms of POD and FAR the CART regression approach is comparable to the other four models excluding Robertson's MDA. In the important "poor" air quality category, a FAR of 1 puts it in the same rank as the subjective model by over forecasting a "poor" ozone concentration event. The CART regression approach ranks above Robertson's MDA for this location considering HEIDKE skill scores, but falls below the other four models. On the other hand, the CART classification approach fares better in the Probability of Detection category, ranking in the top two in all three categories, "good", "fair", and "poor". However that debt is paid in the False Alarm Ration scores of the "good" and "fair" categories, lying in the poorest two of the models. As with most of the other models, there is no score for "poor", with no predictions or observations in this category.

The time-series plot (Appendix B) of the CART regression data for this location indicates two occasions in which the predicted ozone concentration levels reached the "poor" category above 82 ppb, while in fact observed levels remained "good" to "fair" for all samples. A strong case could be made again by looking closely at the graphical data that the "poor" predictions were the result of increasing ozone concentration observations prior to the missed forecasts. However on August 13th a large increase in observed ozone concentration level to 80 ppb did not result in a high-valued forecast the next day. Therefore a high observed value does not always lead to a high forecast the next day. The previous day's observed value often appears to be a strong predictor of ozone concentration levels, but this is not always the case. At other times, different predictors are given more weight in determining the forecast value. It may prove fruitful to study the relationship among the various predictors and the weight given to each predictor for each day's forecast. This approach is possible in the CART technique.

**T12 - Chilliwack**

The CART Regression technique worked very well for this station. The bias, SD, MAE, and RMSE values were all equivalent to the best models. The RV value was second with a value of 0.47, close behind Concord's RV of 0.48.

The CART Regression technique led all models in the contingency table percent correct category with 86%, with Taylor's MLR the lowest at 73%. Models were similar in the "good" and "fair" categories with POD ratios in the range 0.74 to 0.88 for "good", yet only 0.44 to 0.78 for "fair". T12 Chilliwack was one of two sites to record a single "poor" air quality event in the data sample, yet none of the models were able to produce an accurate forecast of the episode. All models had a POD of zero for the "poor" category. Four of the models had FAR's of 1 indicating they not only missed the "poor" event that did occur, they overforecast at least one "poor" event which did not occur. Returning to the statistics at hand, the CART regression technique led the models in both HEIDKE skill scores, chance and climatology.

The CART Classification technique faired poorly for the data at this site. The POD was zero, as with the other models, and the FAR was one. The Heidke skill scores were in the lowest two of the seven models examined for this site.

It is interesting to note that in the time-series representation of the data for Chilliwack (in Appendix B), the one-day time lag evident at the sites previously examined is not readily apparent. In fact the observed and forecast traces are in phase for much of the record. As previously stated this station did report a category "poor" event in late July. The time series shows that a "fair" reading of 64 on the 29th rose dramatically above the advisory threshold of 82 ppb on the 30th with a measurement of 101 ppb. The CART regression technique indicates a smoothing of the data as the forecast values show much less extreme variations than the actual observations. This "averaging" effect evident in the graphically portrayed data is a by-product of most statistical forecast methods and is particularly troubling when isolated, extreme-valued events characterize the data set.

The CART Classification data is represented by the histogram in Appendix B, figure 3. This method works well in forecasting trends in the data, however the single day time lag is evident in the analysis. Most of the "good" category forecasts are on target, with an overforecast of "poor" on July 18th and 19th. The actual 82 ppb exceedance on July 30th was indicated by a three day trend of "fair" forecasts, but the category three observation was missed.

**T15 - Surrey East**

The maximum daily ozone concentration data for the Surrey East location was clearly best represented by Robertson's MLR model, as indicated by the continuous variable verification statistics, with an RV value of 0.33. The CART regression technique lay back in the pack where the other three initial models were nearly the same. The RV for the CART regression model was exactly zero, showing no skill, as was the case for the other three models (Concord's, Taylor's MLR, and Subjective all lay in the range -0.02 to 0.04).

In terms of contingency table statistics, Robertson's MLR proved itself to be the best model for this site. A percent correct of 89% led the other models. A POD ratio of 0.94 in the "good" category associated with a FAR of 0.08 were good results. Heidke skill scores of 0.64 and 0.44 with respect to chance and climatology, respectively, were the best results of all models for this site. The CART regression model was generally middle of the pack, while the CART classification model faired poorly vying for the worst model for this site with Taylor's MLR.

The time-series plot of the data for Surrey East (see Appendix B) shows no clear, decisive pattern. While no observed or predicted ozone concentration values fell above the 82 ppb threshold, there were three "peaks" in the series in which observed values approached the advisory criterion. In the first case on July 20th, a one-day time lag is evident as the predicted values follow the observed values of the day before. The second (July 31st) and third peaks are in phase with the predicted results, with the third peak, on August 13th, well represented by the CART regression model near a value of 70 ppb.

The graphical representation of the CART classification data shows that predicted values model the trend of actual values, however "fair" episodes are not clearly resolved.

**T17 - Richmond South**

The data for T17 (Richmond South) indicates this station favored "good" air quality condition with occasionally days of "fair" (see Appendix A). The continuous variable verification statistics verify this with a mean value of 34 ppb, second lowest in the sample. As a result all models faired about equally in predicting ozone concentration levels. The CART regression model was similar to the subjective model, which ranked last in the RV statistic with a value of 0.01. However the statistical categories indicate there was very little to differentiate among the five models for this site.

The contingency table verification statistics also show little variation among the models. Taylor's MLR was the best model overall with a 91% correct, a POD value of 1 in the "good" category and no false alarms in the "fair" category. The Heidke skill scores indicate the preference of this model over the others for this location as it recorded the only positive values for chance and climatology scores. The CART regression technique was in the middle of the pack in terms of contingency statistics, while the CART classification technique fell into last place overall. Both CART techniques performed well in the "good" category, but were unable to predict "fair" events, with a POD of zero and a FAR of 1 for both models (similar to the subjective model).

The graphical representation of the data (in Appendix B) shows that predicted values agree with the general trend of observed values, however peak values in the data are not resolved by the model.

**T28 - Abbotsford**

The Concord model led in all continuous variable verification categories, showing it to be the best model to represent T28 (Abbotsford) data. The CART regression faired very poorly in this case, with a bias of 12 compared to 7 for Concord. The RMSE was 19 compared to 14 for Concord, while the RV fell below zero to -0.19 for CART regression; Concord led with an RV of 0.42.

Contingency table verification also shows Concord's model to be the best of the seven examined for this location. POD and FAR statistics for all models are similar in value to other sites for "fair" and "good" categories. However T28 (Abbotsford) was the only other site, after T12 (Chilliwack), to record an exceedance of the 82 ppb "poor" air quality threshold (on the same day as Chilliwack, July 30). All seven models, including CART regression and CART classification, missed the event, each with a POD ratio of zero. Four of the models (Robertson's MDA, Subjective, and both CART techniques) forecast at least one exceedance which did not occur, leading to FAR's of one for each of these models. According to Heidke skill scores, all models were able to predict ozone categorical forecasts better than chance would suggest, but the two CART models were the only two not able to better climatology in this score, with values near zero.

The time series analysis of the data (in Appendix B) shows the three peaks that have been evident in other station's data. Viewing the data, it is clear that the CART regression model was able to resolve the three peaks with a fair degree of success. This is a promising result, however unfortunately there is a phase shift of one day in the pattern which has plagued the results of this experiment to this point.

## *Summary and Conclusions*

As seen in the summary of the best models which was presented in Table 4, the best overall model is Robertson's MLR approach. However different models perform better or worse depending on location. Robertson's MLR proved to be the best model for three of the sites, with Taylor's MLR and Concord's each being the best model for two sites (note there is some overlap because two sites were best represented by two models with essentially identical verification scores). The focus of this report is on the CART technique, and the CART regression model proved best for one site, T12 Chilliwack. The CART classification model performed poorly for all sites.

In studying the contingency table verification statistics for categorical forecasts, it is clearly evident that a great deal of discrepancy exists in model performance broken down into categories. All models performed well in predicting "good" air quality episodes with POD values mostly in the 0.80 to 1.0 range and FAR values around 0.10. Performance deteriorated significantly in accurately predicting "fair" categorical forecasts, with POD values ranging from 0 to 0.90, with an average of 0.46. FAR values for the "fair" category covered the complete range from 0 to 1 with an average of 0.49. These values indicate the models on average give a 50% probability of accurately predicting "fair" ozone concentration levels. However forecast accuracy varies among the models at any particular station; in addition each model's forecast performance varies from station to station. It is difficult to determine whether variations in model performance are a result of particular model characteristics or differences in the effect of meteorological parameters on ozone concentration among the sites. With the complex effects of meteorology, topography, and source characteristics on ozone concentration levels, it is difficult to differentiate between random and systematic sources of error in the forecast models.

The experiment covered 44 days for five of the stations while the sixth, T17 (Richmond South), had 34 days of data, for a total of 254 daily maximum ozone concentration events. Of these, only two fell into the "poor" air quality category, or 0.08% of the total number of events. Of the statistical models tested to date, none have been able to successfully predict the rare occurrence of a "poor" air quality event in the 1992 data set. By plotting the data as a time series, it was seen that a one-day time lag of events between the observed and predicted values was partially responsible for the model's poor performance under the scrutiny of verification statistics. This one-day time lag is due to the influence of the previous day's maximum ozone concentration as an important predictor in the CART linear regression equation. Smoothing of the predicted values was also seen to average the "peaks" of the data series, in certain cases predicting lower values or lower categorical forecasts than observed due to the smoothing of the predictand curve.

## *Recommendations*

The original statistical verification approach by Lord made a series of recommendations concerning repeating the experiment using observed weather and ozone values as inputs. This would separate out the component of the error due to uncertainties in the weather and ozone values from the error residing in the statistical methods themselves. It follows that the CART technique should also be re-evaluated in this same manner, including repeating the experiment using inputs obtained from data available during the evening. By the evening, many of the input parameters to the models are known and so do not need to be forecast. This would reduce the error in the CART forecasts.

Including upper air data as predictors would undoubtedly provide a more complete data set for the CART models. An investigation of model performance including upper level wind, temperature, moisture, and geopotential height and thickness parameters as predictors would prove beneficial.

It may prove useful to study the CART approach in more detail. In this report, the statistically "best" decision tree was chosen to represent predicted values. However, the CART technique allows other decision trees to be chosen which may better represent the data, though at the expense of increased uncertainty in the results.

To provide the most information from the data available it may be desirable to analyze ozone levels averaged over a certain area as opposed to spot forecasts. In Lord's paper he produced a regional forecast using a statistical model and a subjective technique. It may be advisable to break the region into three smaller areas, for example a western section, a central section, and an eastern section (Taylor, personal communication) and average the data in those sections to produce an area forecast.

## *Acknowledgments*

I would like to thank Dr. William Burrows, of Meteorological Services Research Branch in Downsview, for compiling and organizing the CART data.

## *References*

Breiman, L., J. H. Freidman, R. A. Olshen and C. J. Stone, 1984: *Classification and Regression Trees*. Wadsworth & Brooks/Co 358 pp.

Burrows, William R., 1991: Objective Guidance for 0-24-Hour and 24-48-Hour Mesoscale Forecasts of Lake-Effect Snow Using CART. *Weather and Forecasting Bulletin Amer. Meteor. Soc.*, **Vol. 6**, 357-378.

Ciccone, A., et al., 1992: A Statistical Ozone Forecasting Model for the Lower Fraser Valley. Consultant's report prepared for the Air Quality Unit of AISB, Pacific Region, April 1992.

Lord, E. R., to be published 1993: Forecasting Daily Maximum Ground-Level Ozone Concentrations in Greater Vancouver and the Lower Fraser Valley.

Robertson, E., 1992: Statistical Approaches to Forecasting Ground-Level Ozone. Consultant's report prepared for the Air Quality Unit of AISB, Pacific Region, March 1992.

Stanski, Henry R., Laurence J. Wilson and William R. Burrows, 1989: *Survey of Common Verification Methods in Meteorology*. Atmospheric Environment Service Forecast Research Division, 114 pp.

Taylor, E., 1991: Forecasting Ground-Level Ozone in Vancouver and the Lower Fraser Valley of British Columbia. Pacific Region Scientific Services Division Report PAES-91-3, November 22, 1991.

## *Appendix A*

Appendix A contains a list of verification statistics covering the CART Regression and CART Classification techniques.

## VERIFICATION STATISTICS

| | |
|---|---|
| FORECAST MODEL: | CART (Regression) |
| LOCATION: | T1 - Robson Square |

## CONTINUOUS VARIABLE STATISTICS

| | |
|---|---|
| Number of forecasts verified: | 44 |
| Observed sample mean: | 18 |
| Average forecast error: | 0.3 |
| Error variance: | 60 |
| Standard deviation of errors: | 8 |
| Mean absolute error: | 6 |
| Root mean square error: | 8 |
| Reduction of variance: | -0.03 |

## 3X3 CONTINGENCY TABLE STATISTICS

### FORECAST

| OBSERVED | | GOOD (0-51 ppb) | FAIR (52-82 ppb) | POOR (>82 ppb) | TOTAL |
|---|---|---|---|---|---|
| | GOOD (0-51 ppb) | **O** | 0 | 0 | 44 |
| | FAIR (52-82 ppb) | 0 | 0 | 0 | 0 |
| | POOR (>82 ppb) | 0 | 0 | 0 | 0 |
| | TOTAL | 44 | 0 | 0 | 44 |

| | |
|---|---|
| Percent Correct: | 100 |
| Post Agreement of "Good" category: | 1 |
| Post Agreement of "Fair" category: | - |
| Post Agreement of "Poor" category: | - |
| Prefigurance of "Good" category: | 1 |
| Prefigurance of "Fair" category: | - |
| Prefigurance of "Poor" category: | - |
| Bias of "Good" category: | 1 |
| Bias of "Fair" category: | - |
| Bias of "Poor" category: | - |
| Heidke skill score w.r.t. chance: | - |
| Heidke skill score w.r.t. climatology: | - |

## VERIFICATION STATISTICS

FORECAST MODEL: CART (Classification)
LOCATION: T1 - Robson Square

Number of forecasts verified: 44

## 3X3 CONTINGENCY TABLE STATISTICS

### FORECAST

OBSERVED

| | GOOD (0-51 ppb) | FAIR (52-82 ppb) | POOR (>82 ppb) | TOTAL |
|---|---|---|---|---|
| GOOD (0-51 ppb) | 44 | 0 | 0 | 44 |
| FAIR (52-82 ppb) | 0 | 0 | 0 | 0 |
| POOR (>82 ppb) | 0 | 0 | 0 | 0 |
| TOTAL | 44 | 0 | 0 | 44 |

Percent Correct: 100
Post Agreement of "Good" category: 1
Post Agreement of "Fair" category: -
Post Agreement of "Poor" category: -
Prefigurance of "Good" category: 1
Prefigurance of "Fair" category: -
Prefigurance of "Poor" category: -
Bias of "Good" category: 1
Bias of "Fair" category: -
Bias of "Poor" category: -
Heidke skill score w.r.t. chance: -
Heidke skill score w.r.t. climatology: -

## VERIFICATION STATISTICS

| | |
|---|---|
| FORECAST MODEL: | CART (Regression) |
| LOCATION: | T9 - Rocky Point Park |

### CONTINUOUS VARIABLE STATISTICS

| | |
|---|---|
| Number of forecasts verified: | 44 |
| Observed sample mean: | 36 |
| Average forecast error: | 10 |
| Error variance: | 176 |
| Standard deviation of errors: | 13 |
| Mean absolute error: | 12 |
| Root mean square error: | 16 |
| Reduction of variance: | 0.00 |

### 3X3 CONTINGENCY TABLE STATISTICS

FORECAST

| OBSERVED | | GOOD (0-51 ppb) | FAIR (52-82 ppb) | POOR (>82 ppb) | TOTAL |
|---|---|---|---|---|---|
| | GOOD (0-51 ppb) | 30 | 5 | 1 | 36 |
| | FAIR (52-82 ppb) | 0 | 7 | 1 | 8 |
| | POOR (>82 ppb) | 0 | 0 | 0 | 0 |
| | TOTAL | 30 | 12 | 2 | 44 |

| | |
|---|---|
| Percent Correct: | 84 |
| Post Agreement of "Good" category: | 1 |
| Post Agreement of "Fair" category: | 0.58 |
| Post Agreement of "Poor" category: | 0 |
| Prefigurance of "Good" category: | 0.83 |
| Prefigurance of "Fair" category: | 0.88 |
| Prefigurance of "Poor" category: | - |
| Bias of "Good" category: | 0.83 |
| Bias of "Fair" category: | 1.50 |
| Bias of "Poor" category: | - |
| Heidke skill score w.r.t. chance: | 0.59 |
| Heidke skill score w.r.t. climatology: | 0.13 |

## VERIFICATION STATISTICS

FORECAST MODEL: CART (Classification)
LOCATION: T9 - Rocky Point Park

Number of forecasts verified: 44

## 3X3 CONTINGENCY TABLE STATISTICS

### FORECAST

| OBSERVED | | GOOD (0-51 ppb) | FAIR (52-82 ppb) | POOR (>82 ppb) | TOTAL |
|---|---|---|---|---|---|
| | GOOD (0-51 ppb) | 34 | 2 | 0 | 36 |
| | FAIR (52-82 ppb) | 6 | 2 | 0 | 8 |
| | POOR (>82 ppb) | 0 | 0 | 0 | 0 |
| | TOTAL | 40 | 4 | 0 | 44 |

| | |
|---|---|
| Percent Correct: | 82 |
| Post Agreement of "Good" category: | 0.85 |
| Post Agreement of "Fair" category: | 0.50 |
| Post Agreement of "Poor" category: | - |
| Prefigurance of "Good" category: | 0.94 |
| Prefigurance of "Fair" category: | 0.25 |
| Prefigurance of "Poor" category: | - |
| Bias of "Good" category: | 1.11 |
| Bias of "Fair" category: | 0.50 |
| Bias of "Poor" category: | - |
| Heidke skill score w.r.t. chance: | 0.24 |
| Heidke skill score w.r.t. climatology: | 0.0 |

## VERIFICATION STATISTICS

| | |
|---|---|
| FORECAST MODEL: | CART (Regression) |
| LOCATION: | T12 - Chilliwack |

### CONTINUOUS VARIABLE STATISTICS

| | |
|---|---|
| Number of forecasts verified: | 44 |
| Observed sample mean: | 41 |
| Average forecast error: | 2 |
| Error variance: | 224 |
| Standard deviation of errors: | 15 |
| Mean absolute error: | 11 |
| Root mean square error: | 15 |
| Reduction of variance: | 0.47 |

### 3X3 CONTINGENCY TABLE STATISTICS

#### FORECAST

| OBSERVED | | GOOD (0-51 ppb) | FAIR (52-82 ppb) | POOR (>82 ppb) | TOTAL |
|---|---|---|---|---|---|
| | GOOD (0-51 ppb) | 30 | 4 | 0 | 34 |
| | FAIR (52-82 ppb) | 1 | 8 | 0 | 9 |
| | POOR (>82 ppb) | 0 | 1 | 0 | 1 |
| | TOTAL | 31 | 13 | 0 | 44 |

| | |
|---|---|
| Percent Correct: | 86 |
| Post Agreement of "Good" category: | 0.97 |
| Post Agreement of "Fair" category: | 0.62 |
| Post Agreement of "Poor" category: | - |
| Prefigurance of "Good" category: | 0.88 |
| Prefigurance of "Fair" category: | 0.89 |
| Prefigurance of "Poor" category: | 0 |
| Bias of "Good" category: | 0.91 |
| Bias of "Fair" category: | 1.44 |
| Bias of "Poor" category: | 0 |
| Heidke skill score w.r.t. chance: | 0.65 |
| Heidke skill score w.r.t. climatology: | 0.40 |

## VERIFICATION STATISTICS

FORECAST MODEL: CART (Classification)
LOCATION: T12 - Chilliwack

Number of forecasts verified: 44

## 3X3 CONTINGENCY TABLE STATISTICS

### FORECAST

| OBSERVED | | GOOD (0-51 ppb) | FAIR (52-82 ppb) | POOR (>82 ppb) | TOTAL |
|---|---|---|---|---|---|
| | GOOD (0-51 ppb) | 30 | 4 | 0 | 34 |
| | FAIR (52-82 ppb) | 4 | 3 | 2 | 9 |
| | POOR (>82 ppb) | 0 | 1 | 0 | 1 |
| | TOTAL | 34 | 8 | 2 | 44 |

| | |
|---|---|
| Percent Correct: | 75 |
| Post Agreement of "Good" category: | 0.88 |
| Post Agreement of "Fair" category: | 0.38 |
| Post Agreement of "Poor" category: | 0 |
| Prefigurance of "Good" category: | 0.88 |
| Prefigurance of "Fair" category: | 0.33 |
| Prefigurance of "Poor" category: | 0 |
| Bias of "Good" category: | 1 |
| Bias of "Fair" category: | 0.89 |
| Bias of "Poor" category: | 2 |
| Heidke skill score w.r.t. chance: | 0.31 |
| Heidke skill score w.r.t. climatology: | -0.10 |

## VERIFICATION STATISTICS

| | |
|---|---|
| FORECAST MODEL: | CART (Regression) |
| LOCATION: | T15 - Surrey East |

## CONTINUOUS VARIABLE STATISTICS

| | |
|---|---|
| Number of forecasts verified: | 44 |
| Observed sample mean: | 39 |
| Average forecast error: | 6 |
| Error variance: | 151 |
| Standard deviation of errors: | 12 |
| Mean absolute error: | 11 |
| Root mean square error: | 13 |
| Reduction of variance: | 0.00 |

## 3X3 CONTINGENCY TABLE STATISTICS

### FORECAST

OBSERVED

| | GOOD (0-51 ppb) | FAIR (52-82 ppb) | POOR (>82 ppb) | TOTAL |
|---|---|---|---|---|
| GOOD (0-51 ppb) | 31 | 4 | 0 | 35 |
| FAIR (52-82 ppb) | 4 | 5 | 0 | 9 |
| POOR (>82 ppb) | 0 | 0 | 0 | 0 |
| TOTAL | 35 | 9 | 0 | 44 |

| | |
|---|---|
| Percent Correct: | 82 |
| Post Agreement of "Good" category: | 0.89 |
| Post Agreement of "Fair" category: | 0.56 |
| Post Agreement of "Poor" category: | - |
| Prefigurance of "Good" category: | 0.89 |
| Prefigurance of "Fair" category: | 0.56 |
| Prefigurance of "Poor" category: | - |
| Bias of "Good" category: | 1 |
| Bias of "Fair" category: | 1 |
| Bias of "Poor" category: | - |
| Heidke skill score w.r.t. chance: | 0.44 |
| Heidke skill score w.r.t. climatology: | 0.11 |

**VERIFICATION STATISTICS**

FORECAST MODEL: CART (Classification)
LOCATION: T15 - Surrey East

Number of forecasts verified: 44

**3X3 CONTINGENCY TABLE STATISTICS**

**FORECAST**

| OBSERVED | | GOOD (0-51 ppb) | FAIR (52-82 ppb) | POOR (>82 ppb) | TOTAL |
|---|---|---|---|---|---|
| | GOOD (0-51 ppb) | 28 | 7 | 0 | 35 |
| | FAIR (52-82 ppb) | 3 | 6 | 0 | 9 |
| | POOR (>82 ppb) | 0 | 0 | 0 | 0 |
| | TOTAL | 31 | 13 | 0 | 44 |

| | |
|---|---|
| Percent Correct: | 77 |
| Post Agreement of "Good" category: | 0.90 |
| Post Agreement of "Fair" category: | 0.46 |
| Post Agreement of "Poor" category: | - |
| Prefigurance of "Good" category: | 0.80 |
| Prefigurance of "Fair" category: | 0.67 |
| Prefigurance of "Poor" category: | - |
| Bias of "Good" category: | 0.89 |
| Bias of "Fair" category: | 1.44 |
| Bias of "Poor" category: | - |
| Heidke skill score w.r.t. chance: | 0.40 |
| Heidke skill score w.r.t. climatology: | -0.11 |

## VERIFICATION STATISTICS

| | |
|---|---|
| FORECAST MODEL: | CART (Regression) |
| LOCATION: | T17 - Richmond South |

## CONTINUOUS VARIABLE STATISTICS

| | |
|---|---|
| Number of forecasts verified: | 34 |
| Observed sample mean: | 34 |
| Average forecast error: | 2 |
| Error variance: | 126 |
| Standard deviation of errors: | 11 |
| Mean absolute error: | 8 |
| Root mean square error: | 11 |
| Reduction of variance: | 0.01 |

## 3X3 CONTINGENCY TABLE STATISTICS

### FORECAST

OBSERVED

| | GOOD (0-51 ppb) | FAIR (52-82 ppb) | POOR (>82 ppb) | TOTAL |
|---|---|---|---|---|
| GOOD (0-51 ppb) | 29 | 1 | 0 | 30 |
| FAIR (52-82 ppb) | 4 | 0 | 0 | 4 |
| POOR (>82 ppb) | 0 | 0 | 0 | 0 |
| TOTAL | 33 | 1 | 0 | 34 |

| | |
|---|---|
| Percent Correct: | 85 |
| Post Agreement of "Good" category: | 0.88 |
| Post Agreement of "Fair" category: | 0 |
| Post Agreement of "Poor" category: | - |
| Prefigurance of "Good" category: | 0.97 |
| Prefigurance of "Fair" category: | 0 |
| Prefigurance of "Poor" category: | - |
| Bias of "Good" category: | 1.10 |
| Bias of "Fair" category: | 0.25 |
| Bias of "Poor" category: | - |
| Heidke skill score w.r.t. chance: | -0.05 |
| Heidke skill score w.r.t. climatology: | -0.25 |

## VERIFICATION STATISTICS

FORECAST MODEL: CART (Classification)
LOCATION: T17 - Richmond South

Number of forecasts verified: 34

## 3X3 CONTINGENCY TABLE STATISTICS

### FORECAST

| OBSERVED | | GOOD (0-51 ppb) | FAIR (52-82 ppb) | POOR (>82 ppb) | TOTAL |
|---|---|---|---|---|---|
| | GOOD (0-51 ppb) | 26 | 4 | 0 | 30 |
| | FAIR (52-82 ppb) | 4 | 0 | 0 | 4 |
| | POOR (>82 ppb) | 0 | 0 | 0 | 0 |
| | TOTAL | 30 | 4 | 0 | 34 |

| | |
|---|---|
| Percent Correct: | 76 |
| Post Agreement of "Good" category: | 0.87 |
| Post Agreement of "Fair" category: | 0 |
| Post Agreement of "Poor" category: | - |
| Prefigurance of "Good" category: | 0.87 |
| Prefigurance of "Fair" category: | 0 |
| Prefigurance of "Poor" category: | - |
| Bias of "Good" category: | 1 |
| Bias of "Fair" category: | 1 |
| Bias of "Poor" category: | - |
| Heidke skill score w.r.t. chance: | -0.13 |
| Heidke skill score w.r.t. climatology: | -1 |

## VERIFICATION STATISTICS

| | |
|---|---|
| FORECAST MODEL: | CART (Regression) |
| LOCATION: | T28 - Abbotsford |

### CONTINUOUS VARIABLE STATISTICS

| | |
|---|---|
| Number of forecasts verified: | 44 |
| Observed sample mean: | 40 |
| Average forecast error: | 12 |
| Error variance: | 243 |
| Standard deviation of errors: | 16 |
| Mean absolute error: | 16 |
| Root mean square error: | 19 |
| Reduction of variance: | -0.19 |

### 3X3 CONTINGENCY TABLE STATISTICS

**FORECAST**

| OBSERVED | | GOOD (0-51 ppb) | FAIR (52-82 ppb) | POOR (>82 ppb) | TOTAL |
|---|---|---|---|---|---|
| | GOOD (0-51 ppb) | 29 | 2 | 2 | 33 |
| | FAIR (52-82 ppb) | 2 | 3 | 5 | 10 |
| | POOR (>82 ppb) | 0 | 1 | 0 | 1 |
| | TOTAL | 31 | 6 | 7 | 44 |

| | |
|---|---|
| Percent Correct: | 73 |
| Post Agreement of "Good" category: | 0.94 |
| Post Agreement of "Fair" category: | 0.50 |
| Post Agreement of "Poor" category: | 0 |
| Prefigurance of "Good" category: | 0.88 |
| Prefigurance of "Fair" category: | 0.30 |
| Prefigurance of "Poor" category: | 0 |
| Bias of "Good" category: | 0.94 |
| Bias of "Fair" category: | 0.60 |
| Bias of "Poor" category: | 7 |
| Heidke skill score w.r.t. chance: | 0.38 |
| Heidke skill score w.r.t. climatology: | -0.09 |

## VERIFICATION STATISTICS

FORECAST MODEL: CART (Classification)
LOCATION: T28 - Abbotsford

Number of forecasts verified: 44

## 3X3 CONTINGENCY TABLE STATISTICS

**FORECAST**

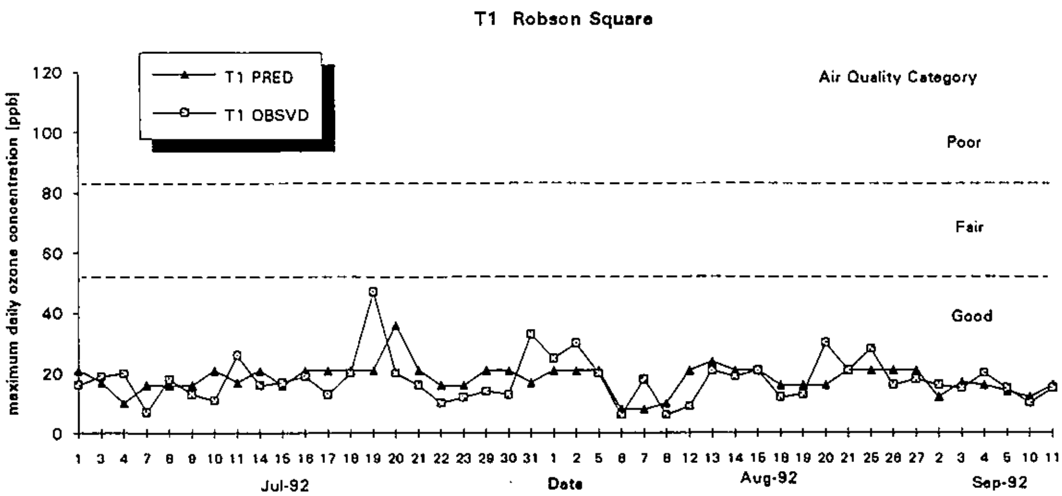
**OBSERVED**

| | GOOD (0-51 ppb) | FAIR (52-82 ppb) | POOR (>82 ppb) | TOTAL |
|---|---|---|---|---|
| GOOD (0-51 ppb) | 26 | 7 | 0 | 33 |
| FAIR (52-82 ppb) | 0 | 7 | 3 | 10 |
| POOR (>82 ppb) | 0 | 1 | 0 | 1 |
| TOTAL | 26 | 15 | 3 | 44 |

Percent Correct: 75
Post Agreement of "Good" category: 1
Post Agreement of "Fair" category: 0.47
Post Agreement of "Poor" category: 0
Prefigurance of "Good" category: 0.79
Prefigurance of "Fair" category: 0.70
Prefigurance of "Poor" category: 0
Bias of "Good" category: 0.79
Bias of "Fair" category: 1.50
Bias of "Poor" category: 3
Heidke skill score w.r.t. chance: 0.48
Heidke skill score w.r.t. climatology: 0

*Appendix B*

Time series representation of predicted and observed CART Regression and CART Classification methods.



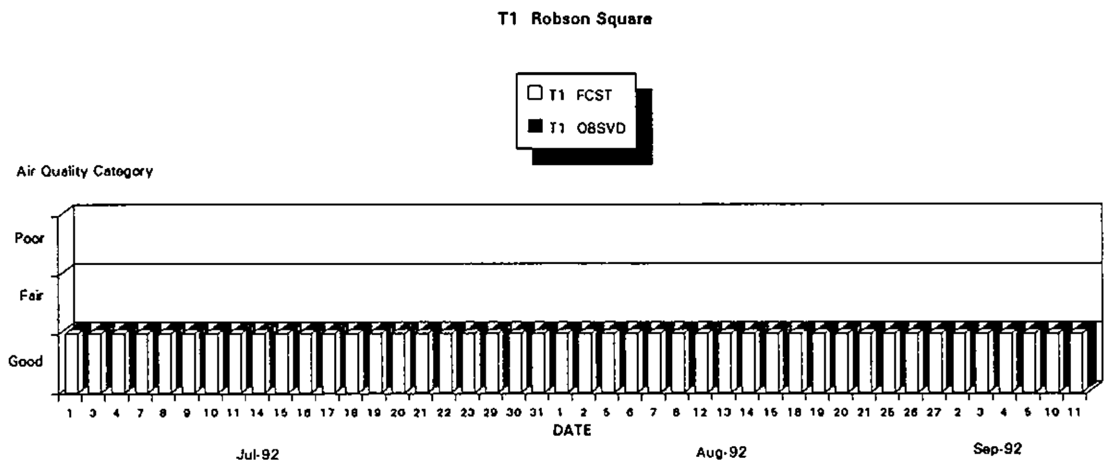
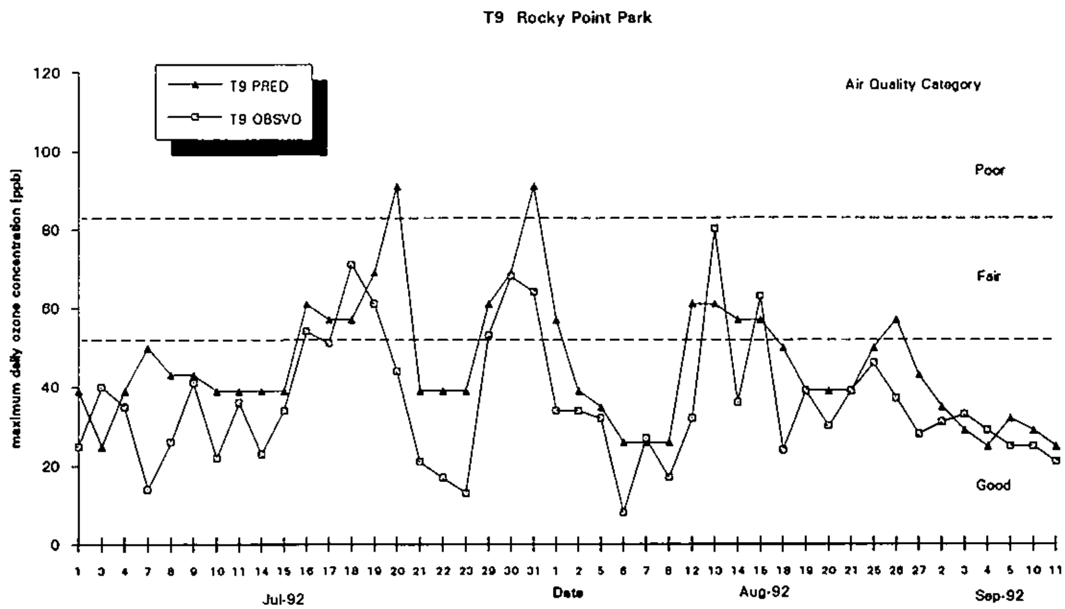




Fig 1. Time series representation of CART regression (top) and CART classification (bottom) methods of predicting maximum daily ozone concentrations for GVRD station T1 - Robson Square.



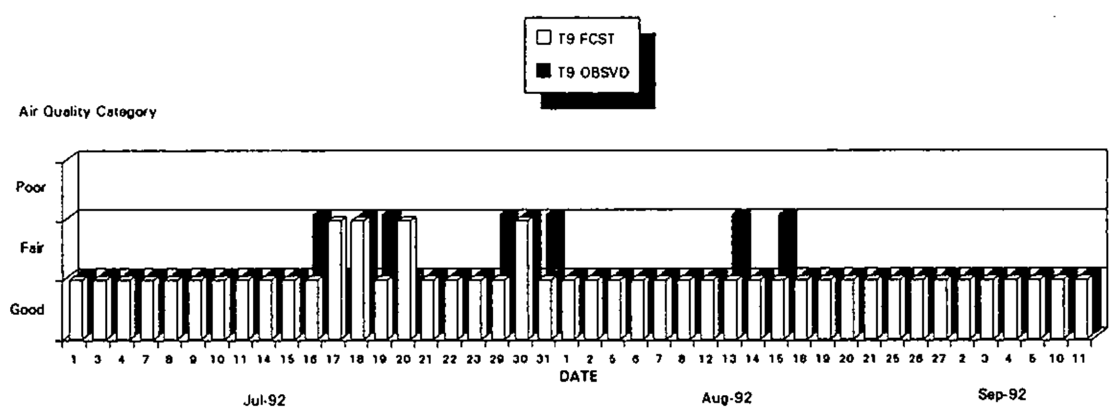
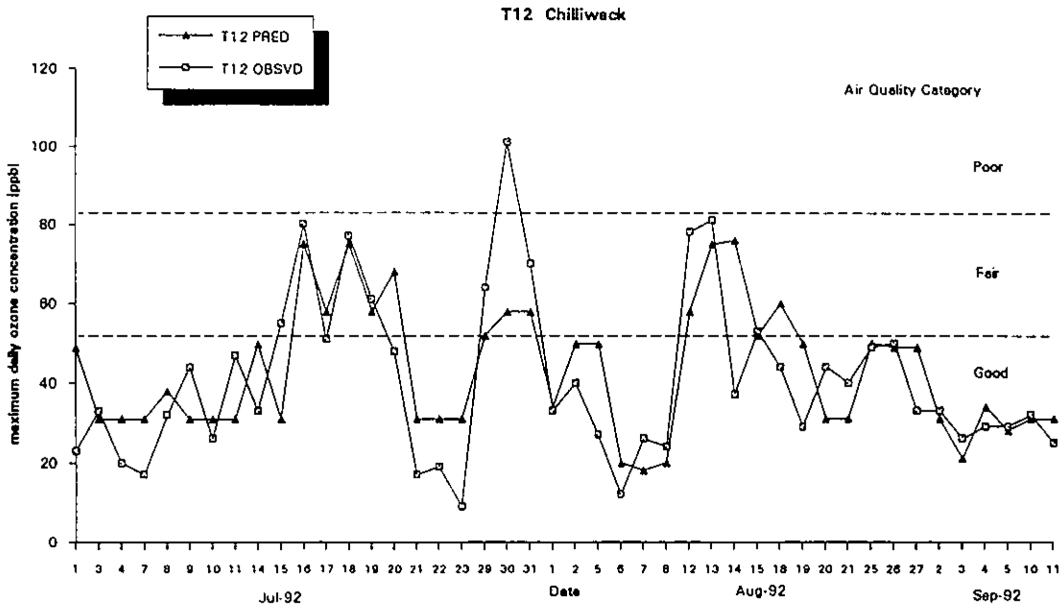




Fig 2. Time series representation of CART regression (top) and CART classification (bottom) methods of predicting maximum daily ozone concentrations for GVRD station T9 - Rocky Point Park.



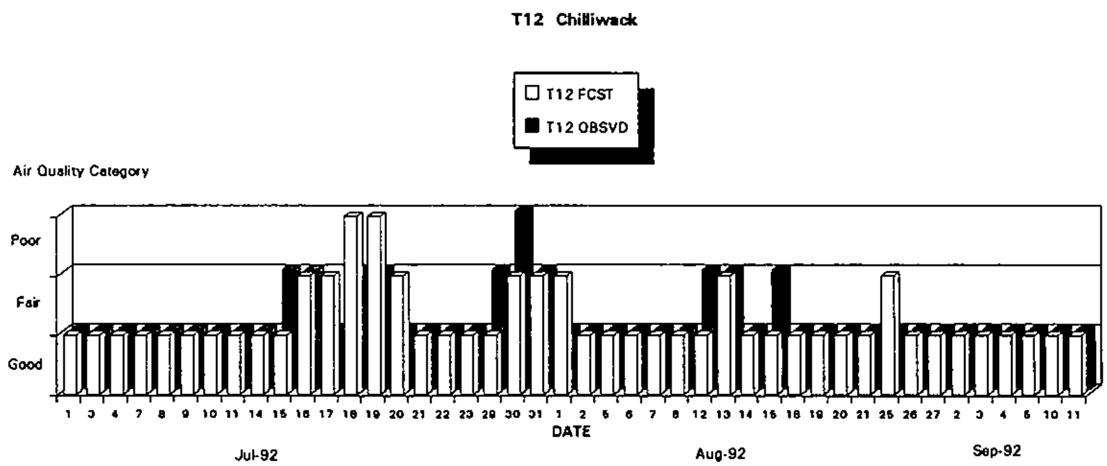
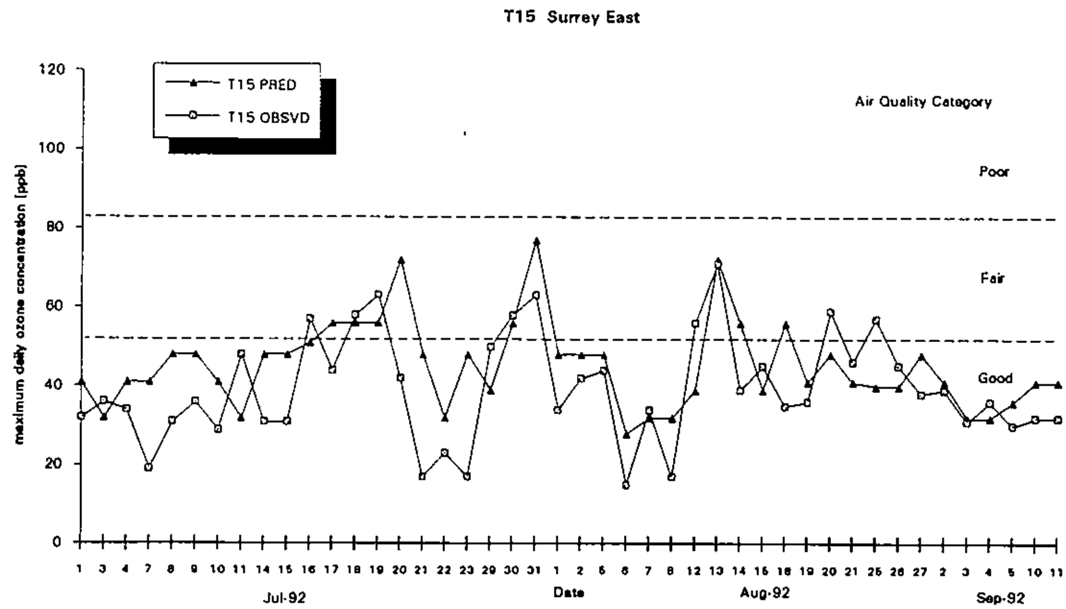




Fig 3. Time series representation of CART regression (top) and CART classification (bottom) methods of predicting maximum daily ozone concentrations for GVRD station T12 - Chilliwack.



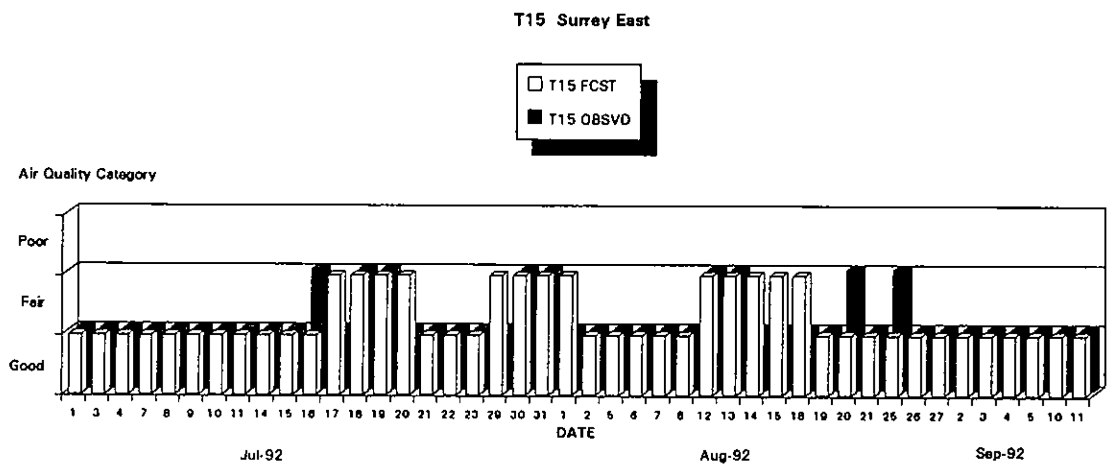
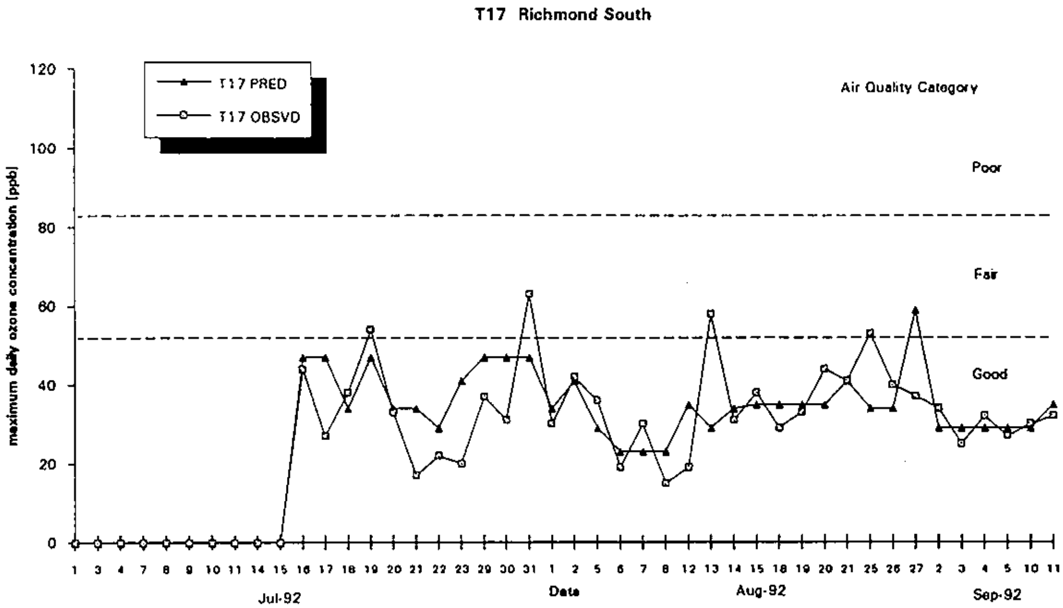




Fig 4. Time series representation of CART regression (top) and CART classification (bottom) methods of predicting maximum daily ozone concentrations for GVRD station T15 - Surrey East.



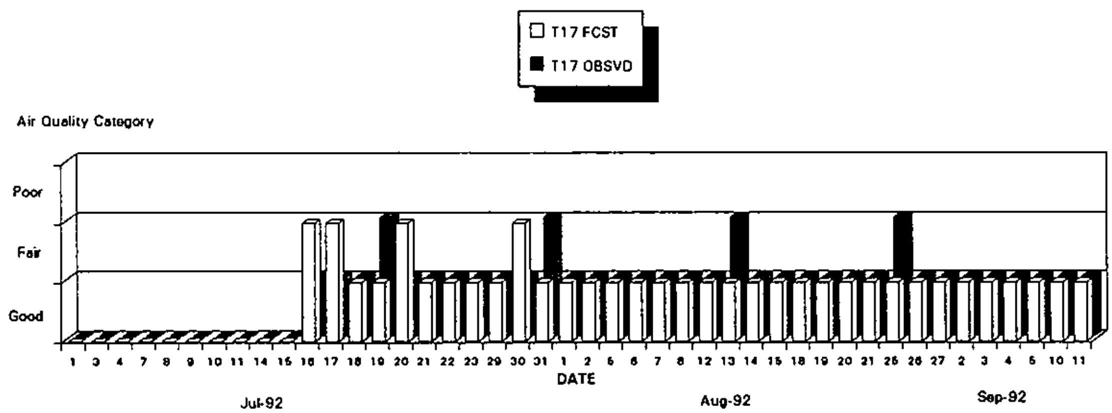




Fig 5. Time series representation of CART regression (top) and CART classification (bottom) methods of predicting maximum daily ozone concentrations for GVRD station T17 - Richmond South. Note the first 10 days of data for this station are missing.

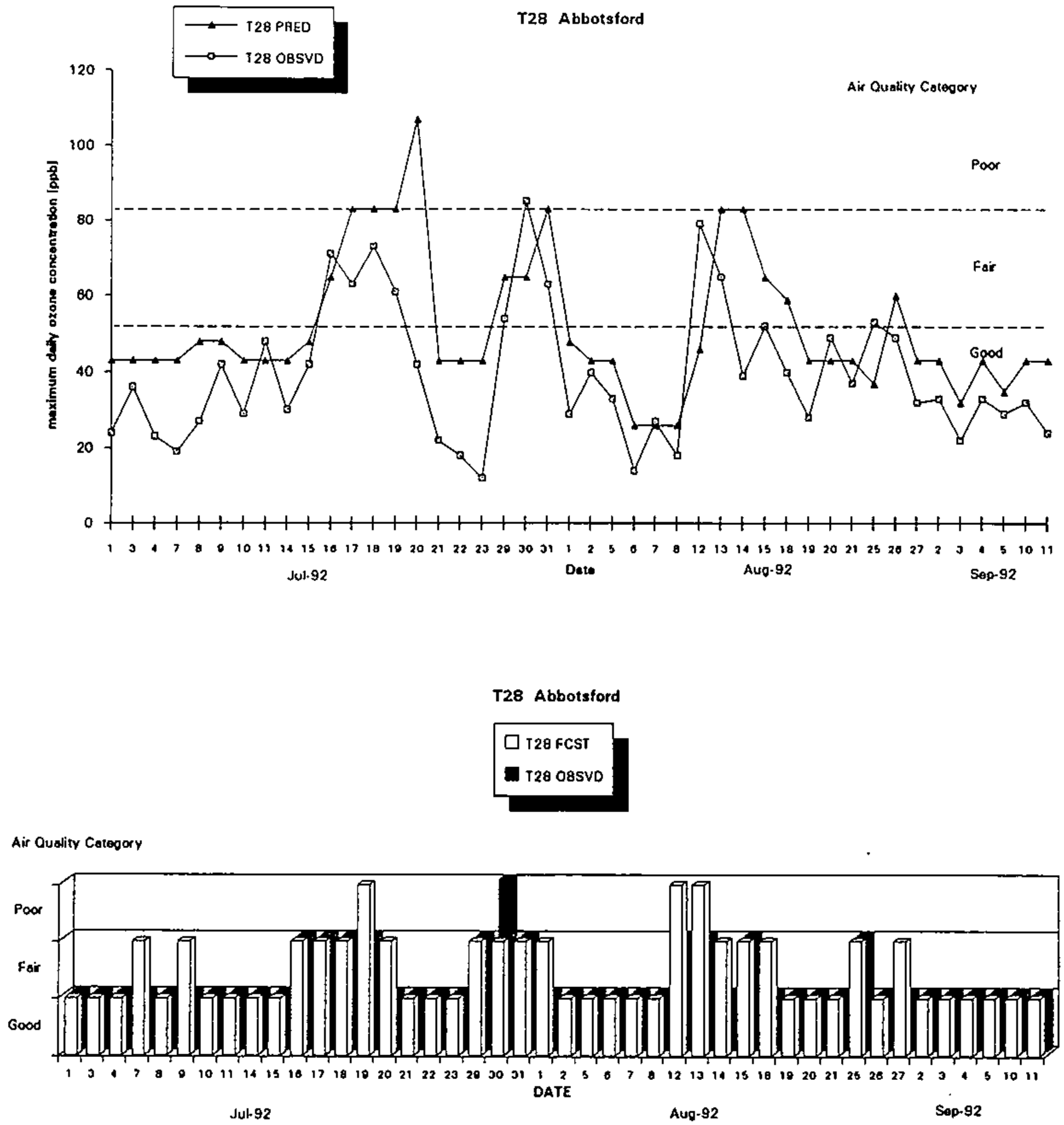



Fig 6. Time series representation of CART regression (top) and CART classification (bottom) methods of predicting maximum daily ozone concentrations for GVRD station T28 - Abbotsford.