EA-Thames

Data Mining of Calculations for the Control of Emissions of Organic Compounds

B E A Fisher, Environment Agency
S Jensen and **P Waldron**, both SPSS(UK) Ltd

ENVIRONMENTAL POLICY – RISK AND FORECASTING
REPORT NO. 27

MARCH 2003

ENVIRONMENT AGENCY

Copies of this report are available from:

The Marketing Officer
Environmental Policy - Risk and Forecasting
Environment Agency
Kings Meadow House
Kings Meadow Road
Reading RG1 8DQ

tel. 0118 953 5262
fax 0118 953 5265
email riskandforecasting@environment-agency.gov.uk

www.environment-agency.gov.uk/science

Title	Series	Version
Data mining of calculations for the control of emissions of organic compounds	Environmental Policy – Risk and Forecasting Report No. 27	Final
Principal author	**Signature**	**Date**
B E A Fisher		
Approved by	**Signature**	**Date**
R Willows		
J Irwin		
P Madden		
Distribution	Public Domain	
Copy Number	N/A	

EXECUTIVE SUMMARY

This report describes the application of data mining techniques for understanding better the relationships between variables within a data set. Data mining techniques may be applied to a wide range of data sets, though they have mainly found favour within the business community where there may be no underlying scientific explanation for a relationship between two variables. In this application the results of complex atmospheric chemistry calculations describing the ozone producing potential of volatile organic compounds (VOCs) downwind of a major industrial source has been used as the input data set. This is an interesting application, but was chosen mainly as a convenient data set on which to test the techniques. The full results of the tests have been provided and so are only of value to a specialist reader wishing to gain some insight into the application of data mining techniques, and to see an example of their use. The full power of the technique may not have been exploited, and there may be further information that could be extracted from the data set.

In this example, data mining was not able to reduce the data to a simple, robust form, which would have enabled it to be used for regulatory applications. However it did demonstrate underlying similarities between the behaviour of individual VOCs, which could be understood in terms of their chemical classification e.g. aromatics and alkenes. Data mining is easy to use given the availability of a data mining package. Given the widespread availability of good quality environmental data it would appear that there could be other possibilities for exploiting data mining within the Environment Agency data sets. It may have a role in data reduction, and in exploring underlying relationships, testing data consistency or generating environmental quality criteria. For example successful data mining could result in a simplified classification of VOCs in the Agency's Pollution Inventory.

CONTENTS

1. INTRODUCTION

Industrial compounds, particularly volatile organic carbons (VOCs), can potentially create ozone under differing conditions. In a related study (Derwent and Nelson, 2003), estimates have been derived for 120 volatile organic compound (VOC) species of the excess ozone production in ppb integrated over the downwind environment for a 1 tonne $hour^{-1}$ industrial emission source. This is the integrated downwind ozone production, or IDOP. This ozone production could also be expressed as an ozone-equivalent emission ceiling, or OEC, for each emitted VOC species in tonne $hour^{-1}$, which when exceeded, would lead to ozone concentrations in the downwind environment on average over 5 ppb higher than in the absence of that industrial VOC source. The IDOP approach offers a substantial advance on the "photochemical ozone creation potential" approach previously applied in ozone assessments, because the focus moves from one of emission equivalence to ozone production in the downwind environment. This raises the question as to whether industry sectors should be given bulk VOC emission reduction targets, or whether a strategy focused on individual species would be more cost-effective.

On a site-by-site basis, these results should be able to point to which processes in a large industrial complex produce the largest ozone impacts in the downwind environment. In particular, the IDOP procedure proposed should enable the identification of the most important VOC species in process emissions, and allow for the development of a cost-effective ozone control strategy. It is a "risk-based" way of weighting sources. A "risk-based" strategy can then be compared with the flat-rate, across-the-board proposals inherent within the UN ECE Gothenburg Protocol and the EU National Emission Ceilings Directive. Since the calculations involve the detailed atmospheric chemistry of a large number of volatile organic compounds, the question arises as to whether many of these compounds behave in broadly similar ways regarding their potential to form ozone. Indeed many models of ozone formation rely on calculations of a small number of representative VOCs. The purpose of this work is to see whether it may be possible to draw such simplifications from the complex calculations contained in this earlier work. The ability to classify a compound into high-ozone/low-ozone production categories would facilitate regulation of VOC output, and enable the Environment Agency to provide more efficient and specific recommendations for VOC reduction.

1.1 Data Mining

Data mining concerns extracting patterns from data sets. These patterns can be used to gain insight into complex data. The objective of this study was to cluster 120 volatile organic compounds based upon the results of complex modelling, and to describe the clusters by a series of operational rules and confidence levels enabling a VOC to be allocated easily to a cluster.

1.2 Input Data Sets

Four data sets were provided, corresponding to Tables 3.1, 3.2, 3.4 and 3.5 of the ozone reactivity study (Derwent and Nelson, 2003) each containing ozone production results from 120 VOCs. Within each data set, each column indicates the ozone production under a specific condition. The differences in the data sets arise from differences in the conditions under which ozone production from the VOCs was calculated.

In the first two data sets (Tables 3.1 and 3.2 of the ozone reactivity study), there were four sets of conditions, while in Tables 3.4 and 3.5 there were only two sets of conditions upon which to cluster ozone production behaviour. In the latter case ozone production under the two conditions was very different, so it was of interest to see if changing the condition brought similar or proportional increases, or decreases in ozone. This makes the clustering behaviour rather straightforward, while under conditions of Table 3.1 and 3.2 it would not be so obvious that individual VOCs were behaving in a similar way. Hence tests on Table 3.1 and Table 3.2 data presents a more severe test of the data mining methods. Table 3.6 of the ozone reactivity report was not used, as the results for each VOC under the two conditions considered were very similar. The data set was not treated as a whole, which would then have contained 9 different sets of conditions and would have been a test of clustering under a wider range of conditions. The data sets are listed in Appendix 1, for completeness and summarised Table 1.

Table 1 Size of Data Sets Analysed

Data set reference in ozone reactivity study	Number of chemicals	Number of conditions (=columns)	Number of missing values
Table 3.1	120	4	21 values missing under conditions 3 and 4
Table 3.2	120	4	21 values missing under conditions 2 and 4
Table 3.4	120	2	Complete
Table 3.5	120	2	20 values missing under conditions 2

Some preliminary results using the Clementine 7.0 Data Mining Workbench (for more details of methods see later) are presented in Appendix 3. These show examples of classifications using a Kohonen network, a form of neural network clustering. The data is mapped onto a 2-dimensional grid, whose size 3x3 cells, or 3x1 cells, is chosen by the user. Each VOC has been assigned a position in the (x,y) plane and those which are close together, are assumed to belong to the same cluster. The values of x and y for each VOC represents some non-linear combination of the corresponding values in the input data set, but do not have a physical meaning, and are not listed in the appendix. Instead chemicals within the same cluster are given the same classification e.g. 00 or 01. Another proprietary algorithm "Two Step" has been applied. This is not able to handle input data sets with missing values. Hence it was only applied to cases in Table 3.1 for which all four conditions had non-zero values. It has the advantage that the user does not have to decide on the number of clusters, only a maximum and minimum number. The algorithm selects an optimum number, between 2 and 15 clusters. In the case of Table 3.1, it decided that there were only two clusters.

The preliminary analysis showed that data mining could be applied, but it did not suggest a simple way of reducing the results of 120 VOCs into a simpler form. However these preliminary results demonstrated that certain chemical categories of VOCs, namely aromatics and alkenes, tended to produce the greatest ozone and had a tendency to cluster together. In the analysis the chemical categories were used in displaying results. One also sees that separating data into clusters may not be very useful practically. Instead a simple set of rules, which would tell one that all VOCs, which

satisfied some simple conditions, would behave similarly, would be a practical way of reducing large data sets generated by computer runs into manageable proportions.

1.3 Data preparation

Volatile organic compounds can be divided into classes according to the structure of the skeleton of carbon atoms in the compound and the way they are bonded. This classification of VOCs into classes, such as alkanes, alkenes etc, is part of the normal nomenclature of VOCs. It was used here to see whether there was any relationship between the clustering of VOC compounds produced by their calculated integrated downwind ozone production and the classification by VOC class. The data mining was performed on the results of the earlier calculations without introducing expert judgement.

Data mining was undertaken by making use of the Clementine 7.0 package, available from SPSS(UK) Ltd. This is a data mining workbench that enables the user to quickly develop predictive models of complex data and deploy them to improve decision making. Data mining is concerned with extracting patterns from data sets. These patterns can be used to gain insight into complex data, aspects of a system's operations, and to predict outcomes for future situations as an aid to decision-making. For example will a certain segment of the population ignore a mail shot or respond to it? Will a process give high, medium, or low yield on a batch of raw material? Clementine contains a number of algorithms and in normal application a range of methods would be tried without considering whether an algorithm is especially suited to a problem

See5 (Windows 98/Me/2000/XP) developed by Rulequest Research, and its Unix counterpart *C5.0*, are sophisticated data mining tools for discovering patterns that delineate categories, assembling them into classifiers, and using them to make predictions. They form part of the Clementine package and have been used in this application to aid interpretation. *See5/C5.0* classifiers are expressed as decision trees, or sets of 'if-then rules', that are generally easier to understand than neural networks. This process is known as profiling. The VOC category information (alkanes, alkenes etc) was used to interpret the clusters produced by the clustering algorithms.

2. MODELLING

2.1 Algorithms

Each data set based on Tables 3.1, 3.2, 3.4 and 3.5 of the ozone reactivity study was modelled individually, taking the four, or two sets of conditions, as the inputs. Two types of clustering techniques were applied for comparison purposes: K-means and Kohonen networks. The clusters were then profiled using *C5.0* decision tree modelling. Clustering with just two conditions does not lead much scope for generalisation. Ideally one should cluster the complete data set under the full set of 9 different conditions (not all the conditions in Tables 3.1, 3.2, 3.4 and 3.5 are distinct). The full power of the method may not therefore have been tested.

K-means clustering is a non-hierarchical method that works by defining a fixed, analyst-determined number of clusters i.e. the number of clusters has to be selected beforehand, and then iteratively minimises an objective function. In this way it splits a set of objects into a selected number of groups. The objective function is defined as mean square distance from each member of a cluster to the centroid of the cluster. The centroid's position is recalculated every time a case is added to the cluster, and this addition and centroid adjustment continues until all the cases are grouped into the final required number of clusters. It works best when the variance of each cluster is similar. The initial clusters are arbitrarily chosen.

Kohonen networks, a type of self-organised mapping, are a type of neural network, which performs clustering. There are two layers of neurons, the input and output, with the data introduced to the input layer. All of the input neurons are connected to all of the output neurons and these connections have strengths or weights associated with them. The output map is a two-dimensional grid of neurons that describes the final relationship between the inputs and outputs, and the clusters are described by an output neuron's connections to the inputs. When the network is fully trained, records that are similar should appear together in the output map and records that are vastly different should appear far apart. In this project 3 x 3 unit maps were used after trials with a few mapping arrangements.

The *C5.0* algorithm builds decision trees, or sets of if-then rules. When *C5.0* builds a decision tree, it uses an entropy-based measure that uses the concept of "information gain" to determine optimal splits. Confidence limits are based upon the mis-classification rate of a particular rule in relation to the number of training cases used in the tree. Rule sets are created by extracting the leaves, or cluster, and listing the rules that maps to a particular "leaf" of a tree, or cluster.

Table 2 Summary of Data Sets and Analysis Techniques

Data sets produced in Derwent and Nelson (2003)'s study	**Analysis techniques used**
Table 3.1 Table 3.2 Table 3.4 Table 3.5	Kohonen clustering 3x3 network K-mean clustering 6 clusters defined C5.0 decision trees

2.2 Results

For each data set, the following charts represent clusters using each of the two techniques, with the clusters overlaid by compound category to aid interpretation. The first four figures are the Kohonen clusters represented in 2-dimensional space, overlaid by compound category, to see if compounds within the same category cluster together. The results have been printed in colour to be understandable.

The following figures compare the Kohonen clusters to the more traditionally used K-mean clustering. Each cluster is represented by a horizontal bar, and according to the

Figure 1 Kohonen clusters derived from Table 3.1 of reactivity study

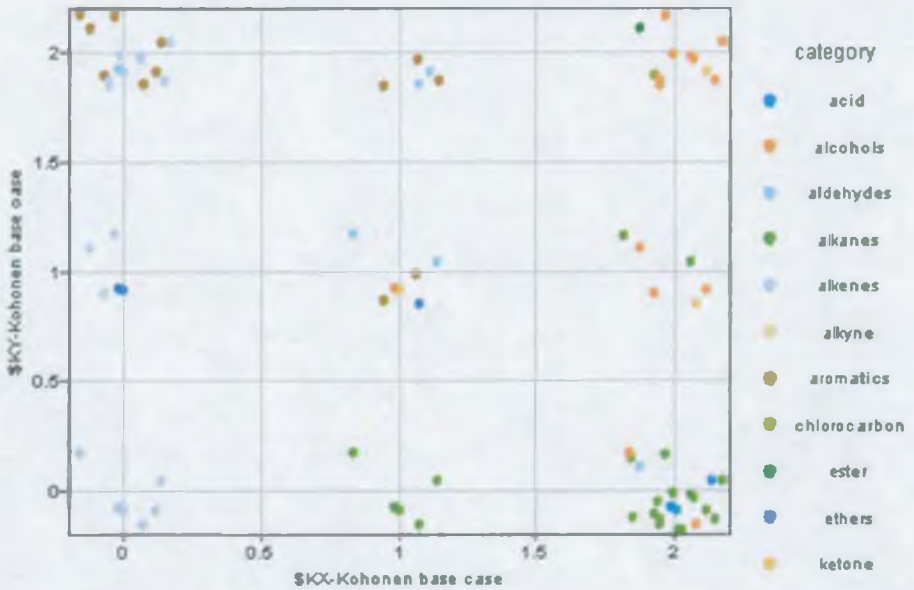


Figure 2 Kohonen clusters derived from Table 3.2 of reactivity study

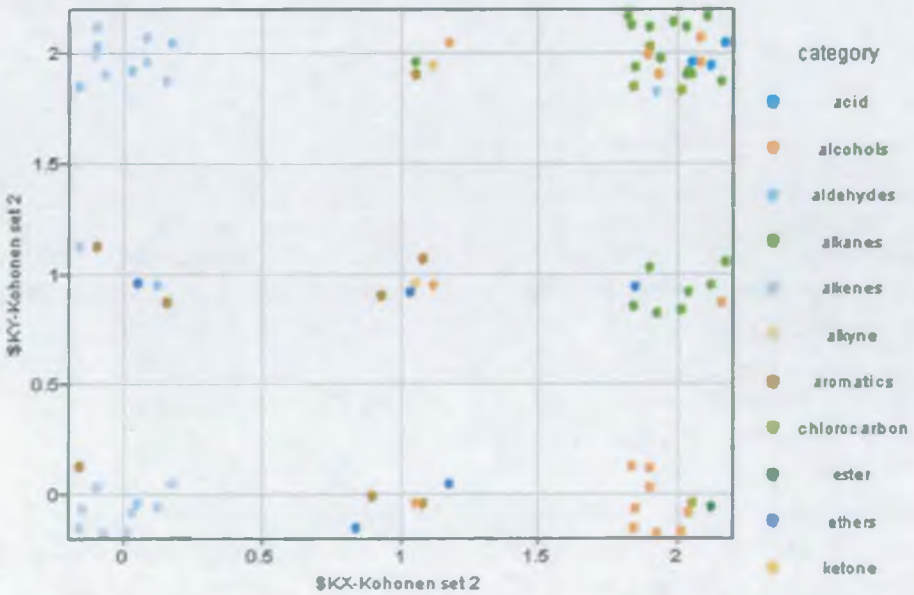


Figure 3 Kohonen clusters derived from Table 3.4 of reactivity study

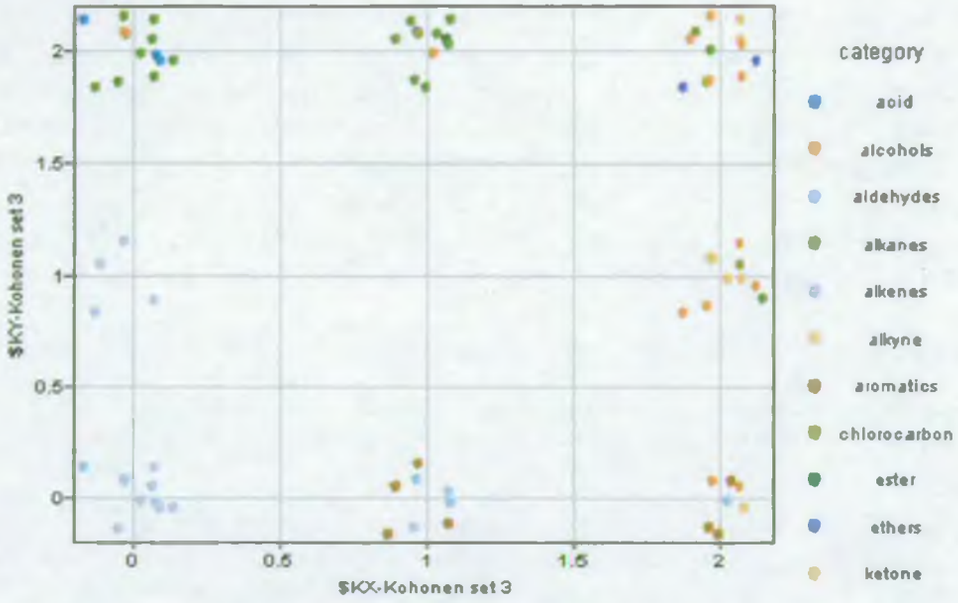
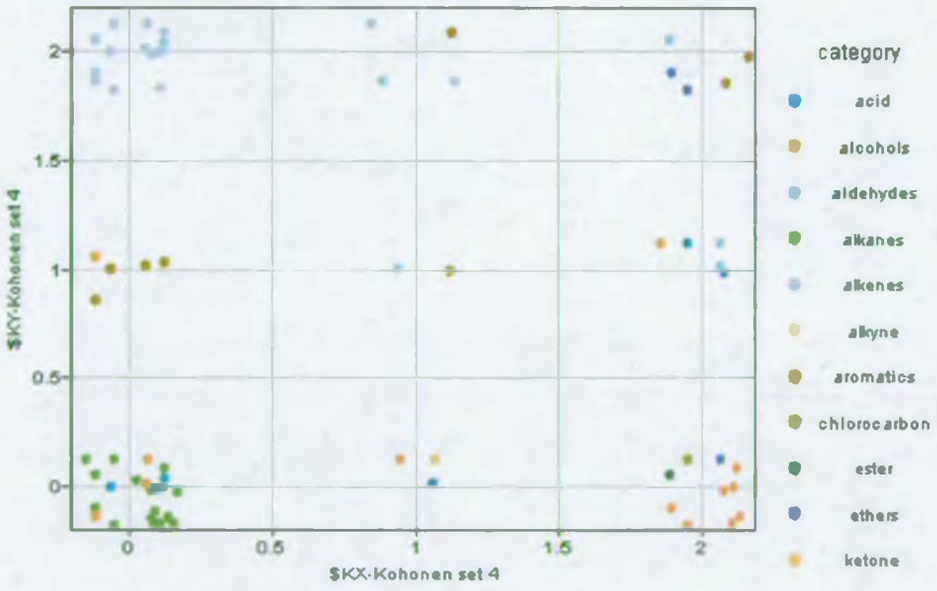


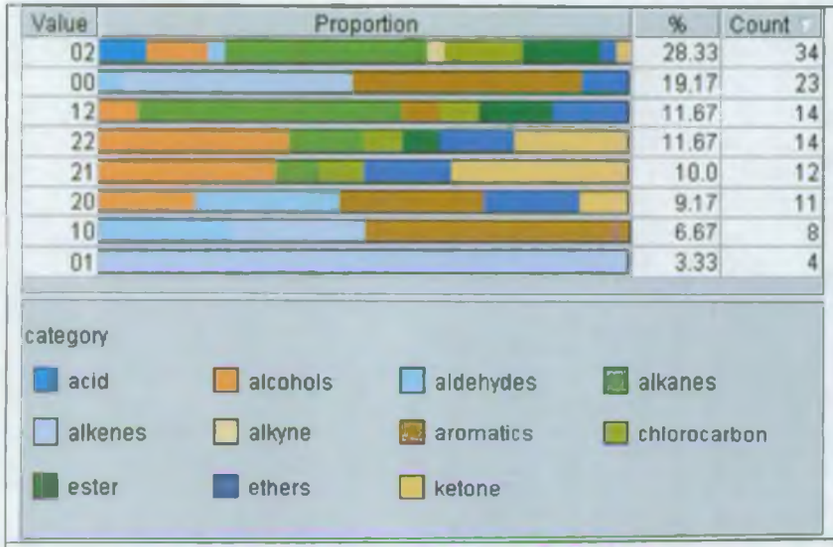
Figure 4 Kohonen clusters derived from Table 3.5 of reactivity study



algorithm settings, there will be a maximum of 9 clusters in the Kohonen results. The K-mean algorithm was set to create 6 clusters. The results of each algorithm were overlaid by VOC compound category (as defined by Appendix 2), to see if the category was related to the behaviour of each compound under the different conditions. If a category predicted a compound perfectly, each bar (cluster) would have a single colour, or a colour would be found only on one of the cluster-bars.

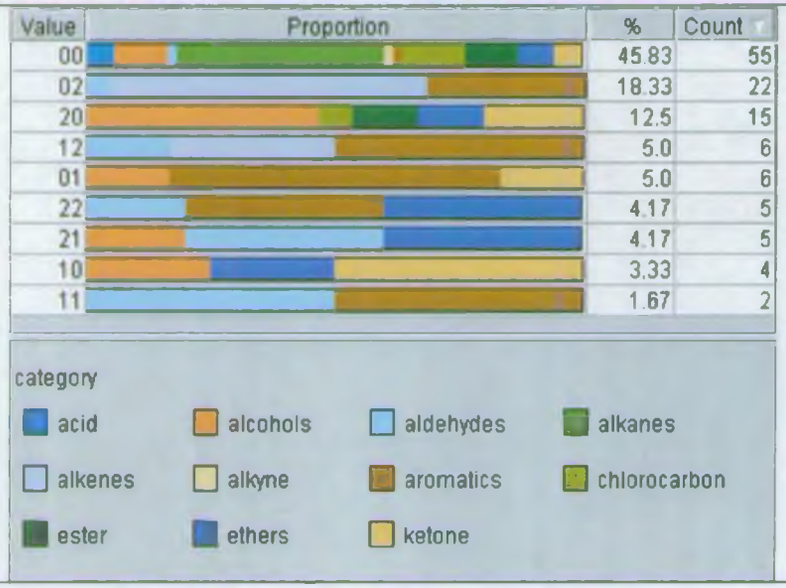
The results do not demonstrate an obvious match between the category and cluster. The different data sets (Table 3.1, 3.2 through to Table 3.4, 3.5) behaved differently according to the number of clusters produced, as well as the relationship of cluster to category.

Figure 5. Comparison of clusters by Kohonen and K-means

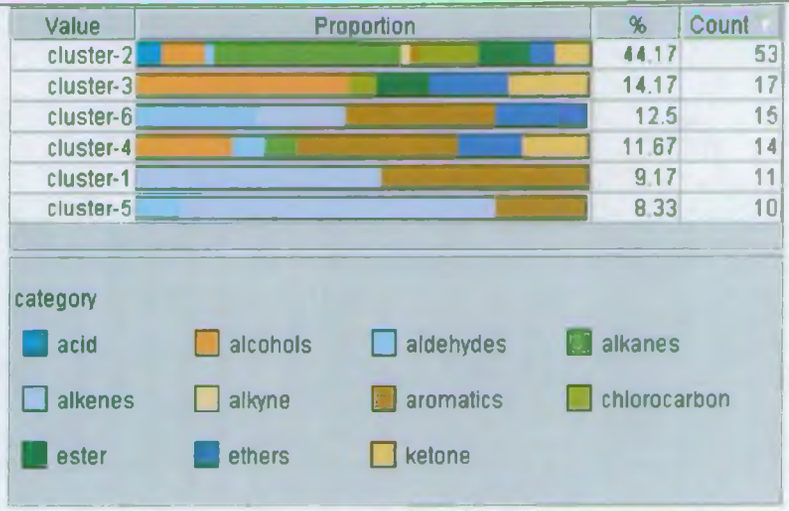


E. Table 3.4 data – Kohonen 3x3 clusters

F. Table 3.4 data - K-means, 6 clusters requested



G. Table 3.5 data – Kohonen 3x3 clusters



H. Table 3.5 data - K-means, 6 clusters requested

2.3 Cluster profiling results

Profiles of the clusters were developed using *C5.0* to better understand the results, and to see if the results could be used operationally. The rule sets were developed directly from the data, and the graphical decision trees were produced secondarily for visualisation purposes. The results of the two are in fact slightly different for some of the data sets, because of the way the confidence limits are treated by the algorithm. The directly generated rule sets are more accurate for use as rule sets, and so are left in this document. Note the colours used in the rule sets have a different meaning to those adopted previously, and refer to the clusters, not to the VOC category.

Profiling results are presented at the end of this document. The presentation of the results are shown as a rule set and decision tree for each model, so for the Table 3.1 data, there is a rule set and decision tree for the Kohonen clusters, followed by a rule set and decision tree for the K-mean clusters. Sets 2 to 4 follow the same structure, hence the 16 pages.

The profiling results are left in a raw state. Some of the numbers in the rule set and decision tree require further explanation. The rules refer to each cluster in turn. The rules for each cluster are numbered sequentially. The two numbers in brackets refer to the number of cases captured by the rule and to the confidence that a cluster is captured by the rule. 1.0 means perfect confidence, 0 means there is no confidence in the rule. *Treatment* number refers to, or is equivalent to *condition* number or *column* number in the data tables. Each cluster is defined by constraints on conditions, either on 4 conditions, or on 2 conditions. The hope, not fulfilled in this case, is that each cluster would be described by a simple constraint. The decision tree contains the number of cases captured by a constraint, and the number of clusters within each constraint, or node. The treatments or conditions, which define the constraint, refer to simple limits on the ozone production.

2.4 Evaluation

The analyses were performed independently of expert knowledge of the data sets, so these results should, at best, be considered a first run, rather than definitive. Although not producing definitive conclusions they are sufficient to show the power of the methods. This report is a supplement to the Derwent and Nelson (2003) study. It demonstrates that the results from these calculations cannot be classified in a simple form, and that individual VOC ozone production valucs need to be retained in regulatory assessments.

A potential refinement to this analysis would be to discretise the ozone production values into high/medium/low values, in order to make qualitative assessments about the production potential of the VOCs. Discretisation would also improve the interpretation and the use of the rule sets. Currently the rule sets say, for example, if ozone production under condition 1 is greater than 0.68, and ozone production under condition 2 is less than 1.56, then the VOC compound falls into cluster 00. If the treatment values were discretised, the rulesets would then say, for example, if ozone production under condition 1 is high and ozone production under condition 2 is low, then the compound falls into cluster 00. In addition, cluster 00 would contain most of those compounds that have high production under condition 1, yet low production under condition 2.

3. CONCLUSIONS

The application of data mining using the Clementine package is seen to be a useful tool for evaluating output from complex models. In operation a range of techniques can be applied to a problem, on a trail and error basis, to see if relationships within the data can be discovered. Other statistical techniques are available in the package. A standard procedure for data mining has been produced, known as CRISP-DM (Cross Industry Standard Process for Data Mining), (SPSS, 2000), with a view to aiding systematic business and commercial use. It can also be applied to process modelling, as demonstrated here. It has potential uses for evaluating a wide range of environmental data sets.

For this application, the input data sets cannot be reduced with confidence to a simpler form, although many of the relationships seen from the data mining are explicable in terms of what is known about types of volatile organic compounds (VOCs). With regard to their ozone producing potential, it appears necessary to assess each VOC individually. However as this remains a complex, and not necessarily robust procedure, not easily applied in regulation, further efforts should be directed towards a practical regulatory approach, following the ideas demonstrated by data mining techniques.

Data mining methods may be criticised as a purely statistical technique, representing modelling without understanding. In this example the data represents the ozone production downwind of an industrial source under a range of conditions ranked according to the single species of VOC emitted from the source. An attempt to explain the results in terms of simple photochemistry is given in Appendix 4. It is seen that for toluene a consistent relationship is not found. It is concluded that an alternative method of classification, simple interpretation of results using expert judgement, is also not easily applied, nor easily generalised, although it results in some better insights.

4. ALGORITHM REFERENCES

Kohonen networks

Kohonen T., 1997. *Self-Organising maps*, Springer-Verlag, Berlin. See also Neural Networks Research Centre (NNRC). http://www.cis.hut.fi/: The latest research and how it applies to applications. http://www.cis.hut.fi/~sami/thesis/node18.html: A useful introduction to self-organising maps and how they relate to K-means.

K-means clustering

Arabie P. and Hubert H., 1994. *Cluster Analysis in Marketing Research*. In Advanced Methods of Marketing Research, edited by Richard P. Bagozzi, Blackwell Publishers. A useful recent review of cluster analysis.

C5.0 rule induction/decision trees

Quinlan, R. 1993. *Programs for Machine Learning*. Morgan Kaufmann Publishers, San Mateo. Detailed description of C4.5 with source code listing. The Rulequest website has some comments on C5.0 versus C4.5. http://www.rulequest.com/see5-comparison.html and http://www.rulequest.com/see5-win.html

General references

SPSS, 2000. CRISP-DM 1.0 Step-by-step data mining guide

Derwent R G and Nelson N, 2003. Development of a reactivity index for the control of the emissions of organic compounds, Environment Agency R&D Report P4-105. Copies of this report are available from the Environment Agency R&d Dissemination Centre c/o WRc, Frankland Road, Swindon, Wilts SN5 8YF
Email publications@wrcplc.co.uk

Table 3.1 Data. Rule set from Kohonen clustering (profiling of Figure 5A)

- Rules for 00 - contains 1 rule(s)
  - Rule 1 for **00** (6, 1.0)
    - if treatment 1 > 0.28
    - and treatment 2 <= 5.6100001
    - then **00**
- Rules for 01 - contains 1 rule(s)
  - Rule 1 for **01** (4, 0.75)
    - if treatment 1 > 0.16
    - and treatment 1 <= 0.27000001
    - then **01**
- Rules for 02 - contains 1 rule(s)
  - Rule 1 for **02** (31, 0.645)
    - if treatment 2 > 5.6100001
    - then **02**
- Rules for 10 - contains 2 rule(s)
  - Rule 1 for **10** (5, 1.0)
    - if treatment 2 <= 2.72
    - and treatment 2 > 2.4200001
    - and treatment 1 > -0.050000001
    - then **10**
  - Rule 2 for **10** (5, 0.8)
    - if treatment 2 <= 2.4200001
    - and treatment 1 > -0.050000001
    - and treatment 2 > 1.6799999
    - then **10**
- Rules for 11 - contains 1 rule(s)
  - Rule 1 for **11** (10, 0.8)
    - if treatment 2 > 3.55
    - and treatment 2 <= 5.6100001
    - and treatment 1 <= 0.28
    - then **11**
- Rules for 12 - contains 1 rule(s)
  - Rule 1 for **12** (5, 1.0)
    - if treatment 2 > 5.6100001
    - and treatment 1 <= 0.16
    - then **12**
- Rules for 20 - contains 2 rule(s)
  - Rule 1 for **20** (32, 0.875)
    - if treatment 2 <= 1.6799999
    - then **20**
  - Rule 2 for **20** (89, 0.449)
    - if treatment 2 <= 5.6100001
    - then **20**
- Rules for 21 - contains 1 rule(s)
  - Rule 1 for **21** (10, 0.5)
    - if treatment 1 <= -0.050000001
    - and treatment 2 > 2.4200001
    - then **21**
- Rules for 22 - contains 1 rule(s)
  - Rule 1 for **22** (15, 0.6)
    - if treatment 2 > 2.72
    - and treatment 2 <= 3.55
    - and treatment 1 <= 0.28
    - then **22**
- Default: 20

Table 3.1 Data. Decision tree from Kohonen clustering (profiling of Figure 5A)

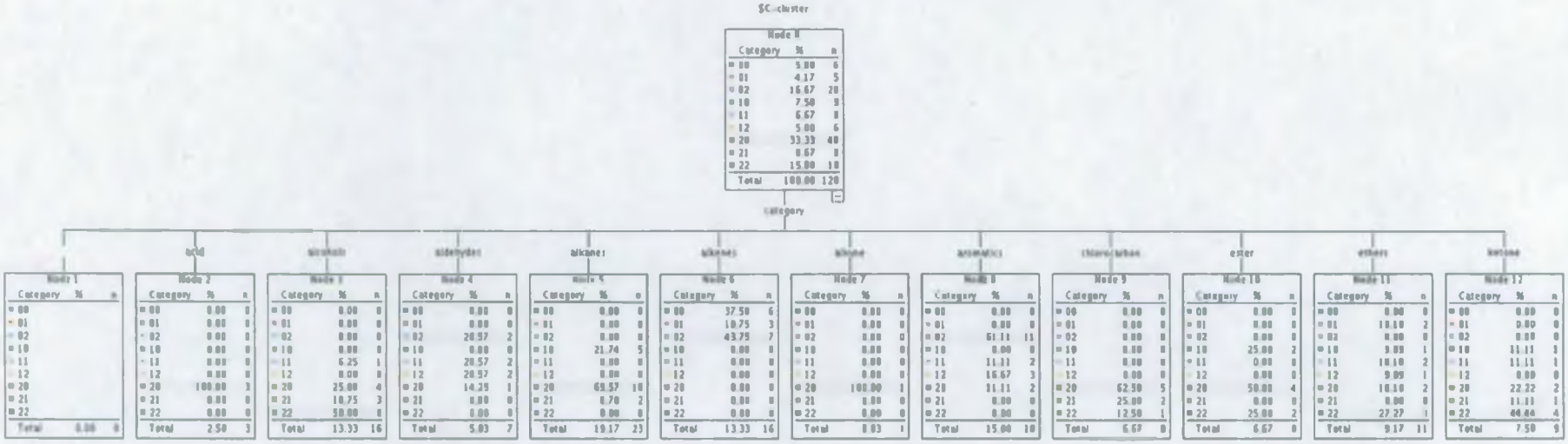


Table 3.1 Data. Rule set from K mean cluster analysis (profiling of Figure 5B)

- Rules for cluster-1 - contains 1 rule(s)
  - Rule 1 for **cluster-1** (15, 0.667)
    - if treatment 2 > 9.6000004
    - then **cluster-1**
- Rules for cluster-2 - contains 1 rule(s)
  - Rule 1 for **cluster-2** (86, 0.616)
    - if treatment 1 <= 0.15000001
    - then **cluster-2**
- Rules for cluster-3 - contains 1 rule(s)
  - Rule 1 for **cluster-3** (6, 1.0)
    - if treatment 1 > 0.15000001
    - and treatment 2 <= 4.5100002
    - then **cluster-3**
- Rules for cluster-4 - contains 1 rule(s)
  - Rule 1 for **cluster-4** (28, 0.893)
    - if treatment 2 > 2.72
    - and treatment 1 <= 0.15000001
    - then **cluster-4**
- Rules for cluster-5 - contains 1 rule(s)
  - Rule 1 for **cluster-5** (5, 1.0)
    - if treatment 2 > 9.6000004
    - and treatment 3 <= 0.41999999
    - then **cluster-5**
- Rules for cluster-6 - contains 1 rule(s)
  - Rule 1 for **cluster-6** (34, 0.324)
    - if treatment 1 > 0.15000001
    - then **cluster-6**
- Default: cluster-2

Table 3.1 Data. Decision tree from K mean cluster analysis (profiling of Figure 5B)

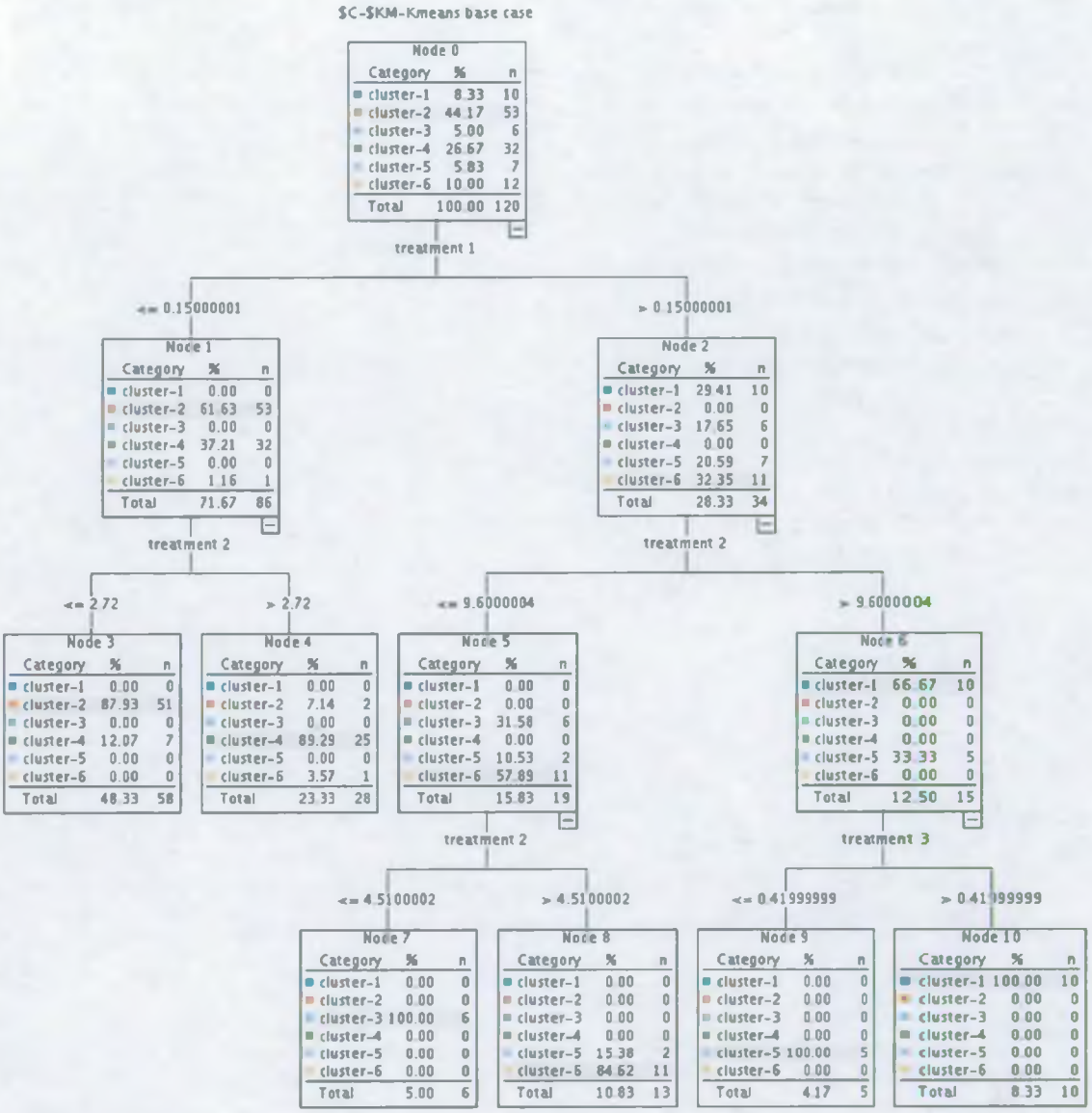


Table 3.2 Data. Rule set from Kohonen clustering (profiling of Figure 5C)

- Rules for 00 - contains 1 rule(s)
  - Rule 1 for **00** (26, 0.769)
    - if treatment 1 > 6.4899998
    - then **00**
- Rules for 01 - contains 1 rule(s)
  - Rule 1 for **01** (6, 1.0)
    - if treatment 1 > 6.4899998
    - and treatment 3 <= 11.66
    - then **01**
- Rules for 02 - contains 1 rule(s)
  - Rule 1 for **02** (10, 1.0)
    - if treatment 4 <= 0.20999999
    - and treatment 3 > 4.9000001
    - then **02**
- Rules for 11 - contains 1 rule(s)
  - Rule 1 for **11** (8, 0.625)
    - if treatment 3 > 4.9000001
    - and treatment 1 <= 6.4899998
    - and treatment 4 > 0.20999999
    - then **11**
- Rules for 20 - contains 1 rule(s)
  - Rule 1 for **20** (13, 0.769)
    - if treatment 3 > 3.9300001
    - and treatment 3 <= 4.9000001
    - then **20**
- Rules for 21 - contains 1 rule(s)
  - Rule 1 for **21** (17, 0.765)
    - if treatment 3 <= 3.9300001
    - and treatment 1 > 2.04
    - then **21**
- Rules for 22 - contains 1 rule(s)
  - Rule 1 for **22** (73, 0.521)
    - if treatment 3 <= 4.9000001
    - then **22**
- Default: 10

Table 3.2 Data. Decision tree from Kohonen clustering (profiling of Figure 5C)

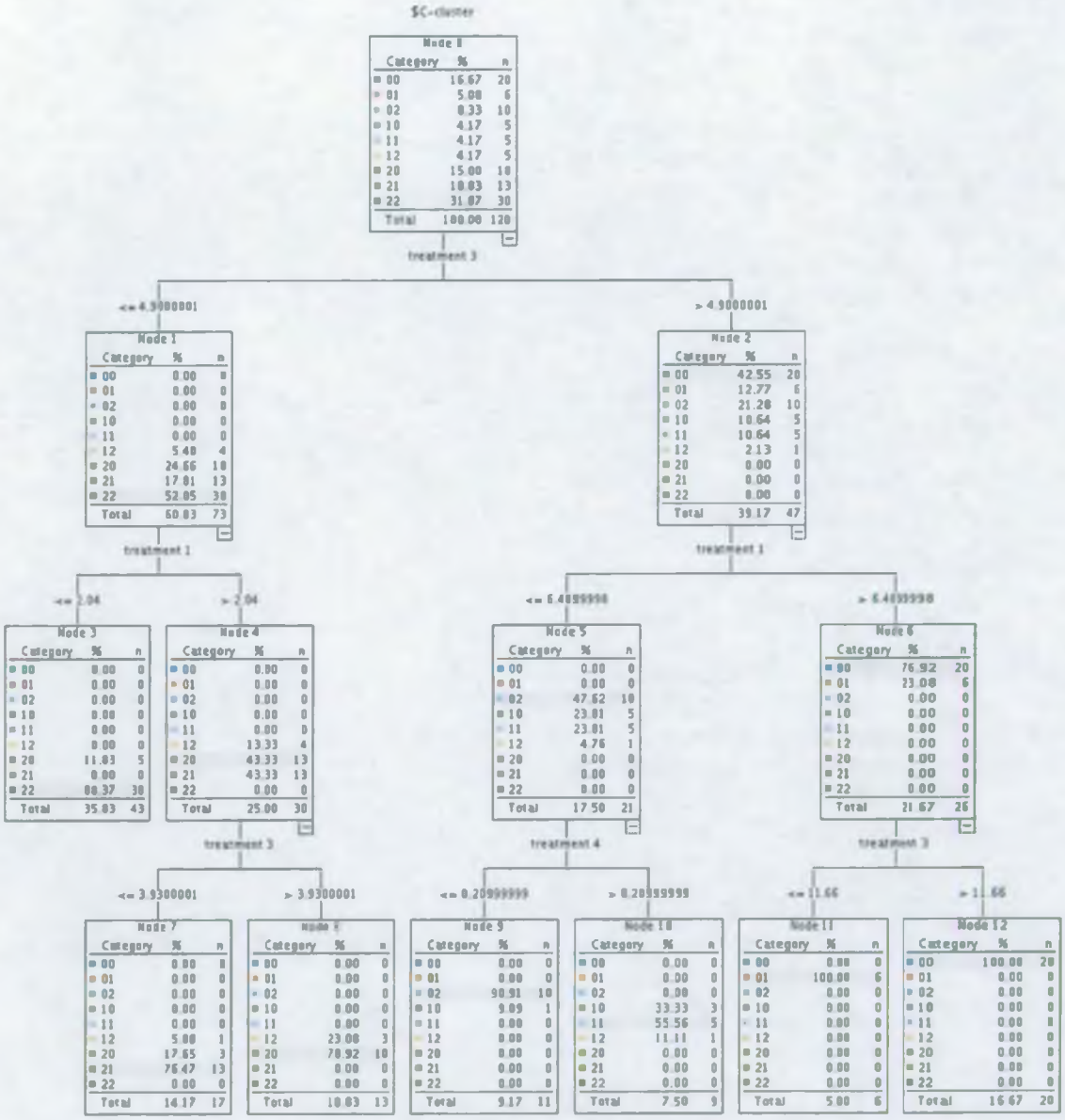


Table 3.2 Data. Rule set from K mean cluster analysis (profiling of Figure 5D)

- Rules for cluster-1 - contains 1 rule(s)
  - Rule 1 for **cluster-1** (31, 0.548)
    - if treatment 3 > 9.71
    - then **cluster-1**
- Rules for cluster-2 - contains 1 rule(s)
  - Rule 1 for **cluster-2** (23, 0.826)
    - if treatment 1 <= 1.0700001
    - then **cluster-2**
- Rules for cluster-3 - contains 1 rule(s)
  - Rule 1 for **cluster-3** (89, 0.494)
    - if treatment 3 <= 9.71
    - then **cluster-3**
- Rules for cluster-4 - contains 1 rule(s)
  - Rule 1 for **cluster-4** (8, 1.0)
    - if treatment 3 > 15.06
    - then **cluster-4**
- Rules for cluster-5 - contains 1 rule(s)
  - Rule 1 for **cluster-5** (6, 1.0)
    - if treatment 3 > 9.71
    - and treatment 1 <= 5.21
    - then **cluster-5**
- Rules for cluster-6 - contains 1 rule(s)
  - Rule 1 for **cluster-6** (11, 1.0)
    - if treatment 2 > 5.8899999
    - and treatment 3 <= 9.71
    - then **cluster-6**
- Default: cluster-3

Table 3.2 Data. Decision tree from K mean cluster analysis (profiling of Figure 5D)

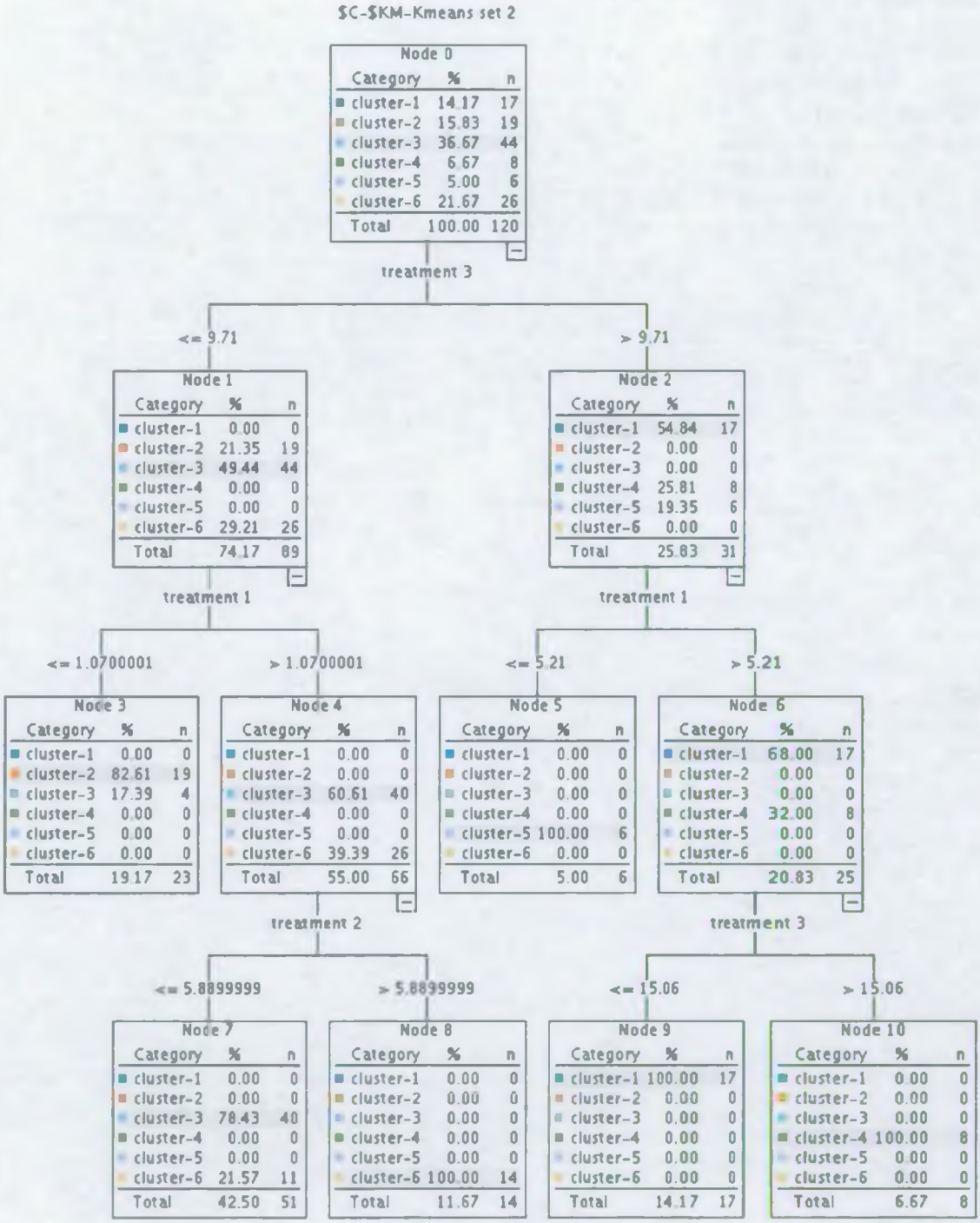


Table 3.4 Data. Rule set from Kohonen clustering (profiling of Figure 5E)

- Rules for 00 - contains 1 rule(s)
  - Rule 1 for **00**
    - if treatment 2 > 43.290001
    - then **00**
- Rules for 01 - contains 1 rule(s)
  - Rule 1 for **01**
    - if treatment 1 > 5.1199999
    - and treatment 2 <= 25.370001
    - then **01**
- Rules for 02 - contains 1 rule(s)
  - Rule 1 for **02**
    - if treatment 2 <= -3.8099999
    - then **02**
- Rules for 10 - contains 1 rule(s)
  - Rule 1 for **10**
    - if treatment 2 <= 43.290001
    - and treatment 1 > 8
    - and treatment 2 > 25.370001
    - then **10**
- Rules for 12 - contains 1 rule(s)
  - Rule 1 for **12**
    - if treatment 2 <= 4.5300002
    - and treatment 2 > -3.8099999
    - then **12**
- Rules for 20 - contains 1 rule(s)
  - Rule 1 for **20**
    - if treatment 1 > 5.1199999
    - and treatment 1 <= 8
    - then **20**
- Rules for 21 - contains 1 rule(s)
  - Rule 1 for **21**
    - if treatment 2 > 16.35
    - and treatment 1 <= 5.1199999
    - then **21**
- Rules for 22 - contains 1 rule(s)
  - Rule 1 for **22**
    - if treatment 2 > 4.5300002
    - and treatment 2 <= 16.35
    - and treatment 1 <= 5.1199999
    - then **22**
- Default: 02

Table 3.4 Data. Decision tree from Kohonen clustering (profiling of Figure 5E)

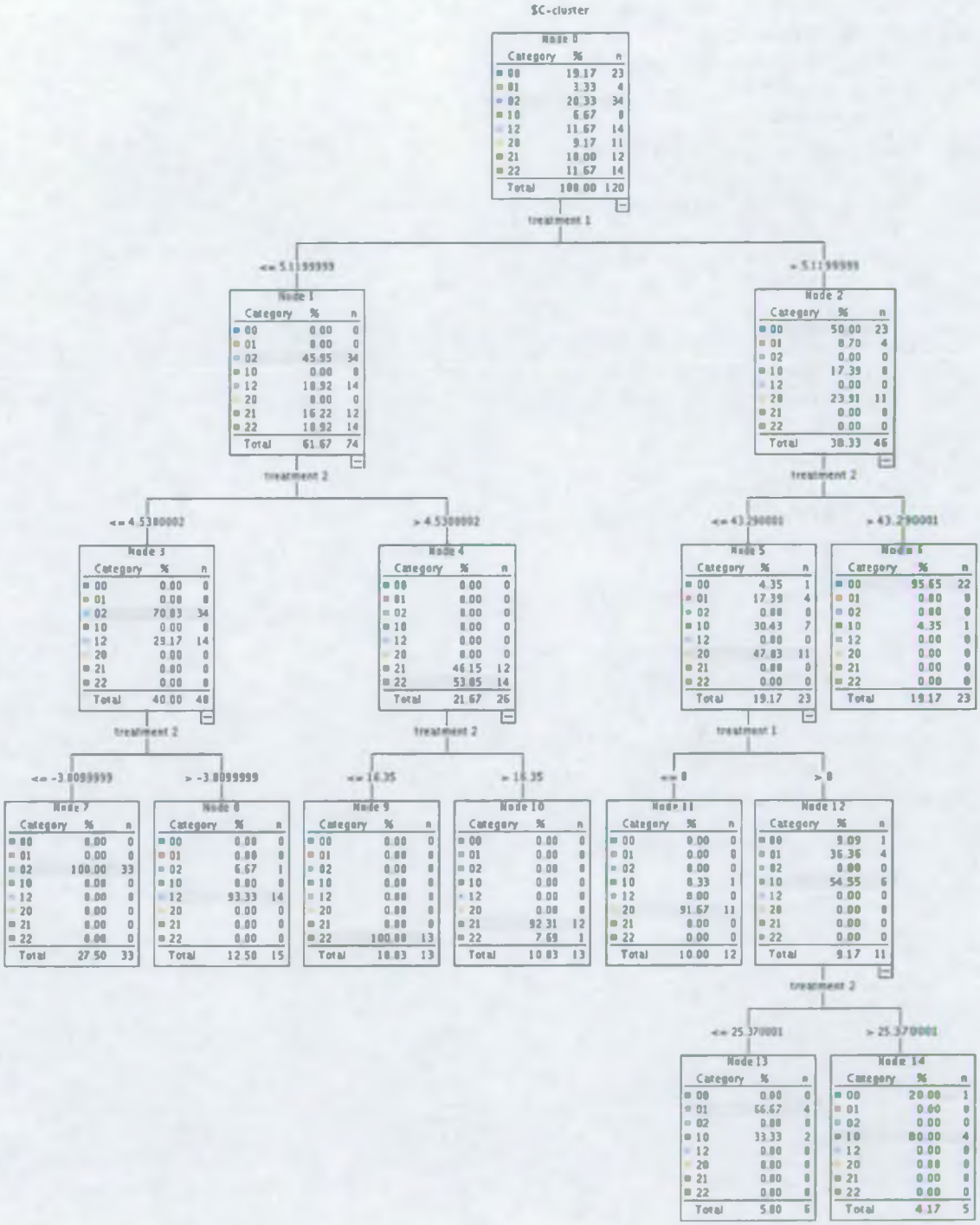


Table 3.4 Data. Rule set from K mean cluster analysis (profiling of Figure 5F)

- Rules for cluster-1 - contains 1 rule(s)
  - Rule 1 for **cluster-1** (16, 1.0)
    - if treatment 1 > 5.1199999
    - and treatment 1 <= 10.14
    - then **cluster-1**
- Rules for cluster-2 - contains 2 rule(s)
  - Rule 1 for **cluster-2** (43, 1.0)
    - if treatment 2 <= 0.050000001
    - then **cluster-2**
  - Rule 2 for **cluster-2** (74, 0.581)
    - if treatment 1 <= 5.1199999
    - then **cluster-2**
- Rules for cluster-3 - contains 1 rule(s)
  - Rule 1 for **cluster-3** (5, 0.8)
    - if treatment 2 > 60.259998
    - then **cluster-3**
- Rules for cluster-4 - contains 1 rule(s)
  - Rule 1 for **cluster-4** (31, 1.0)
    - if treatment 1 <= 5.1199999
    - and treatment 2 > 0.050000001
    - then **cluster-4**
- Rules for cluster-5 - contains 1 rule(s)
  - Rule 1 for **cluster-5** (6, 1.0)
    - if treatment 1 > 5.1199999
    - and treatment 2 <= 25.370001
    - then **cluster-5**
- Rules for cluster-6 - contains 1 rule(s)
  - Rule 1 for **cluster-6** (19, 1.0)
    - if treatment 1 > 10.14
    - and treatment 2 > 25.370001
    - and treatment 2 <= 60.259998
    - then **cluster-6**
- Default: cluster-2

Table 3.4 Data. Decision tree from K means cluster analysis (profiling of Figure 5F)

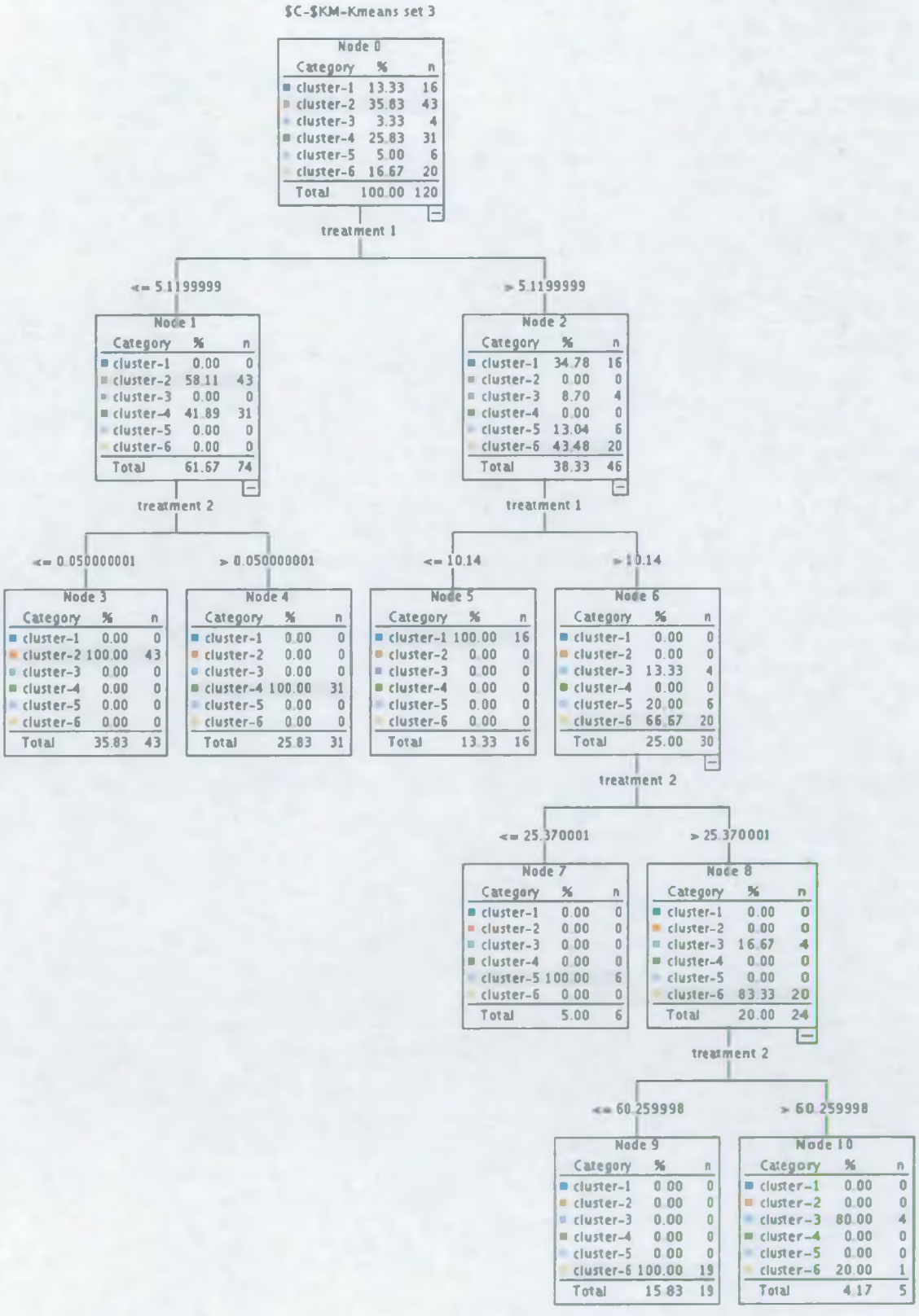


Table 3.5 Data. Rule set from Kohonen clustering (profiling of Figure 5G)

- Rules for 00 - contains 1 rule(s)
  - Rule 1 for **00** (55, 1.0)
    - if treatment 2 <= 5.0100002
    - then **00**
- Rules for 01 - contains 1 rule(s)
  - Rule 1 for **01** (6, 1.0)
    - if treatment 1 > 5.1199999
    - and treatment 2 <= 13.65
    - then **01**
- Rules for 02 - contains 1 rule(s)
  - Rule 1 for **02** (22, 1.0)
    - if treatment 2 > 27.469999
    - then **02**
- Rules for 10 - contains 1 rule(s)
  - Rule 1 for **10** (4, 1.0)
    - if treatment 2 > 5.0100002
    - and treatment 1 <= 5.1199999
    - then **10**
- Rules for 11 - contains 1 rule(s)
  - Rule 1 for **11** (2, 1.0)
    - if treatment 2 > 13.65
    - and treatment 1 <= 6.5799999
    - then **11**
- Rules for 12 - contains 1 rule(s)
  - Rule 1 for **12** (6, 1.0)
    - if treatment 1 > 11.44
    - and treatment 2 <= 27.469999
    - then **12**
- Rules for 21 - contains 1 rule(s)
  - Rule 1 for **21** (2, 1.0)
    - if treatment 1 <= 8.3599997
    - and treatment 2 > 13.65
    - and treatment 1 > 6.5799999
    - then **21**
- Rules for 22 - contains 1 rule(s)
  - Rule 1 for **22** (5, 1.0)
    - if treatment 1 > 8.3599997
    - and treatment 1 <= 11.44
    - then **22**
- Default: 20

Table 3.5 Data. Decision tree from Kohonen clustering (profiling of Figure 5G)

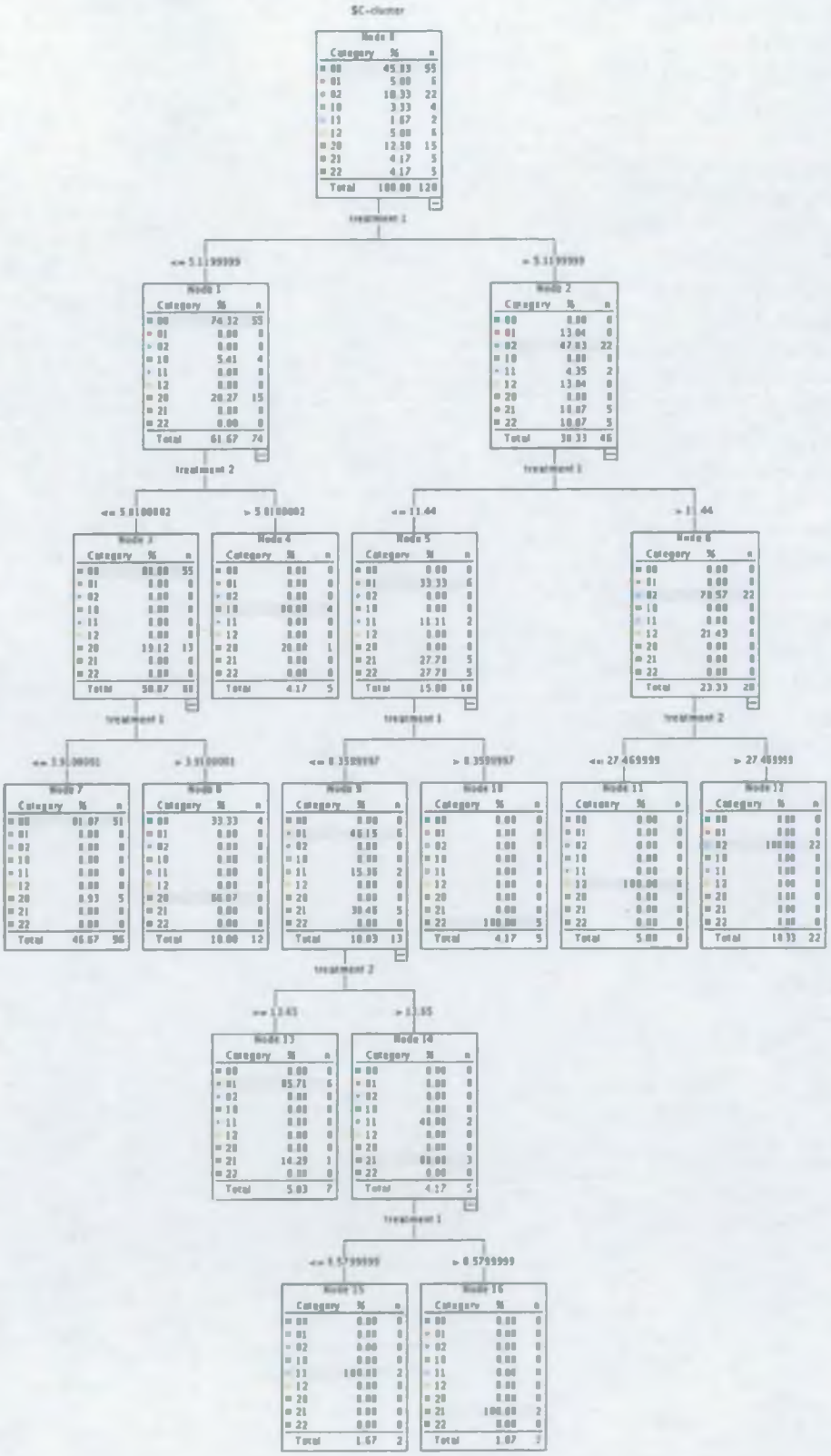
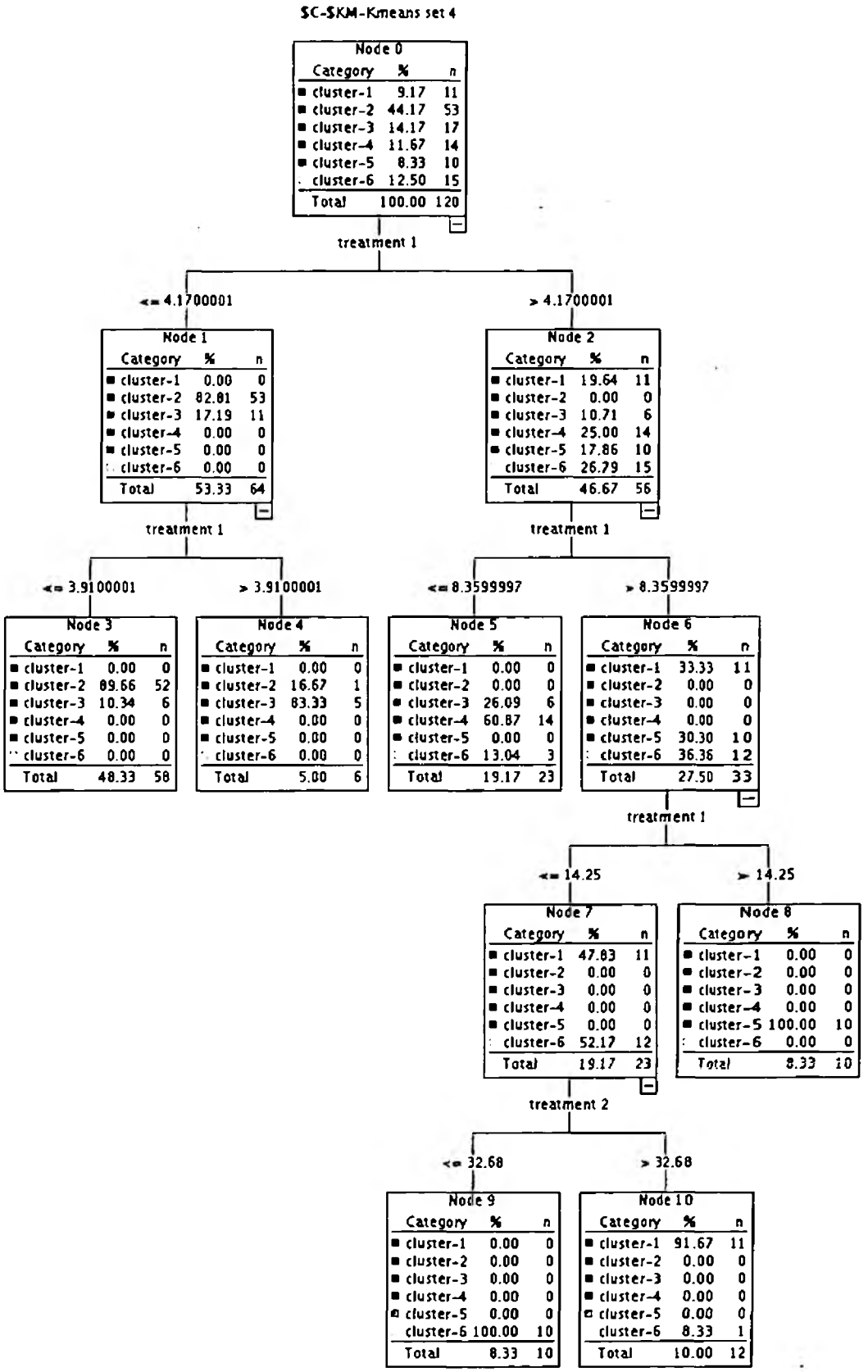


Table 3.5 Data. Rule set from K mean cluster analysis (profiling of Figure 5G)

- Rules for cluster-1 - contains 1 rule(s)
  - Rule 1 for **cluster-1** (11, 1.0)
    - if treatment 2 > 32.68
    - and treatment 1 <= 14.25
    - then **cluster-1**
- Rules for cluster-2 - contains 1 rule(s)
  - Rule 1 for **cluster-2** (58, 0.897)
    - if treatment 1 <= 3.9100001
    - then **cluster-2**
- Rules for cluster-3 - contains 1 rule(s)
  - Rule 1 for **cluster-3** (6, 0.833)
    - if treatment 1 > 3.9100001
    - and treatment 1 <= 4.1700001
    - then **cluster-3**
- Rules for cluster-4 - contains 1 rule(s)
  - Rule 1 for **cluster-4** (23, 0.609)
    - if treatment 1 > 4.1700001
    - and treatment 1 <= 8.3599997
    - then **cluster-4**
- Rules for cluster-5 - contains 1 rule(s)
  - Rule 1 for **cluster-5** (10, 1.0)
    - if treatment 1 > 14.25
    - then **cluster-5**
- Rules for cluster-6 - contains 2 rule(s)
  - Rule 1 for **cluster-6** (10, 1.0)
    - if treatment 1 > 8.3599997
    - and treatment 2 <= 32.68
    - and treatment 1 <= 14.25
    - then **cluster-6**
  - Rule 2 for **cluster-6** (23, 0.522)
    - if treatment 1 > 8.3599997
    - and treatment 1 <= 14.25
    - then **cluster-6**
- Default: cluster-2

Table 3.5 Data. Decision tree from K mean cluster analysis (profiling of Figure 5G)



APPENDIX 1: INPUT DATA SETS USED IN THE ANALYSIS

Data Set 3.1 Integrated ozone production of 120 chemicals under 4 sets of conditions listed in Table 3.1 of the ozone reactivity study

1	1,3,5-trimethylbenzene	1.02	13.55	1.15	8.95
2	3,5-dimethylethylbenzene	0.87	12.86	0.95	8.61
3	3,5-diethyltoluene	0.83	12.14	0.88	8.33
4	1,2,4-trimethylbenzene	0.79	11.82	0.87	7.97
5	isoprene	0.78	12.30	0.77	8.66
6	1,2,3-trimethylbenzene	0.74	11.44	0.88	7.81
7	cis but-2-ene	0.69	3.94	0.10	-0.35
8	2-methylbut-2-ene	0.68	2.04	0.61	0.03
9	propylene	0.64	11.91	0.29	6.85
10	but-1-ene	0.58	10.86	0.17	6.37
11	1,3-butadiene	0.56	11.88	0.67	8.79
12	m-xylene	0.55	11.99	0.74	8.71
13	trans pent-2-ene	0.55	3.66	0.07	-0.18
14	ethylene	0.54	13.97	0.08	9.46
15	cis pent-2-ene	0.51	3.66	0.06	-0.18
16	2-methylbut-1-ene	0.47	10.44	0.03	6.94
17	formaldehyde	0.46	11.28	0.66	8.92
18	butylene	0.45	12.12	0.61	10.31
19	p-xylene	0.45	9.60	0.52	7.23
20	trans hex-2-ene	0.44	3.13	0.49	0.68
21	cis hex-2-ene	0.44	3.14	0.47	0.73
22	m-ethyltoluene	0.42	9.46	0.54	7.00
23	3-methylbut-1-ene	0.40	8.73	-0.05	5.11
24	trans but-2-ene	0.39	10.42	0.08	6.66
25	o-xylene	0.38	9.41	0.54	6.88
26	pent-1-ene	0.36	8.40	0.01	4.50
27	i-butyraldehyde	0.28	8.63	0.47	6.31
28	p-ethyltoluene	0.28	7.64	0.42	5.82
29	propionaldehyde	0.27	6.49	0.42	4.78
30	hex-1-ene	0.27	6.93	0.41	4.54
31	o-ethyltoluene	0.16	7.34	0.37	5.55
32	toluene	0.16	5.83	0.23	4.65
33	pentanaldehyde	0.15	5.08	0.27	3.61
34	butyraldehyde	0.13	5.68	0.31	3.98
35	acetaldehyde	0.10	4.51	0.27	3.17
36	ethylbenzene	0.09	6.11	0.29	4.85
37	propylbenzene	0.09	5.17	0.21	4.26
38	i-pentane	0.08	2.57	0.09	2.06
39	n-pentane	0.07	1.97	0.09	1.44
40	i-butane	0.05	2.99	0.15	2.54
41	2-methylpentane	0.04	2.68	0.09	1.95
42	i-butanol	0.03	4.35	0.18	3.65
43	methyl-i-butylketone	0.03	5.61	0.26	4.30
44	methylpropylketone	0.02	3.41	0.17	2.59

45 2,2-dimethylbutane	0.02	1.35	0.04	0.96
46 n-butanol	0.00	3.39	0.14	2.66
47 dimethylether	0.00	2.71	0.10	2.50
48 n-butane	0.00	1.73	0.05	1.37
49 1-methoxy-2-propanol	0.00	3.61	0.17	3.09
50 2-butoxyethanol	0.00	3.89	0.13	3.02
51 methylethylketone	-0.01	2.30	0.09	1.93
52 ethane	-0.02	0.57	0.04	0.47
53 styrene	-0.03	1.20	0.01	-0.37
54 3-methylpentane	-0.03	2.72	0.11	2.06
55 s-butyl acetate	-0.03	2.58	0.09	2.05
56 i-propylbenzene	-0.03	4.97	0.20	3.91
57 n-hexane	-0.04	1.68	0.05	1.00
58 i-propyl acetate	-0.04	2.36	0.06	1.97
59 trans dichloroethylene	-0.06	2.69	0.13	2.50
60 trichloroethylene	-0.06	1.88	0.09	1.71
61 propane	-0.06	1.20	0.06	0.97
62 2-methylhexane	-0.07	1.96	0.06	1.14
63 neopentane	-0.07	1.34	0.07	1.10
64 2,3-dimethylbutane	-0.07	3.13	0.12	2.33
65 3-methylhexane	-0.07	1.85	0.07	1.07
66 cyclohexanol	-0.08	3.53	0.19	2.96
67 methyl-t-butylether	-0.08	1.31	0.03	1.16
68 acetylene	-0.10	0.52	0.02	0.50
69 ethanol	-0.10	1.93	0.06	1.75
70 propanoic acid	-0.12	0.85	0.04	0.83
71 n-decane	-0.12	-0.15	-0.02	-0.66
72 n-octane	-0.12	0.54	0.01	-0.13
73 n-propyl acetate	-0.14	1.94	0.05	1.48
74 n-nonane	-0.14	0.13	-0.01	-0.47
75 s-butanol	-0.14	2.66	0.11	2.32
76 acetone	-0.15	0.51	0.00	0.52
77 cyclohexane	-0.15	2.42	0.07	1.87
78 cyclohexane	-0.15	2.42	0.07	1.87
79 cis dichloroethylene	-0.15	2.98	0.08	2.76
80 ethyl acetate	-0.15	1.43	0.04	1.20
81 methyl chloride	-0.15	0.04	0.02	0.01
82 n-heptane	-0.15	1.07	0.04	0.44
83 cyclohexane	-0.16	2.41	0.12	1.85
84 diacetone alcohol	-0.17	1.74	0.09	1.50
85 methylene dichloride	-0.17	0.21	0.01	0.20
86 diethylketone	-0.17	1.99	0.10	1.68
87 formic acid	-0.17	0.33	0.02	0.24
88 methyl chloroform	-0.18	0.11	0.02	0.05
89 n-undecane	-0.18	-0.22	-0.07	-0.86
90 benzene	-0.18	1.88	0.07	1.55
91 tetrachloroethylene	-0.19	0.12	0.01	0.10
92 i-propanol	-0.19	1.71	0.04	1.66
93 n-butyl acetate	-0.20	1.34	0.05	0.90
94 methanol	-0.22	1.21	0.05	1.12
95 acetic acid	-0.22	1.07	0.01	0.94

96 t-butanol	-0.22	1.01	0.02	0.81
97 methyl acetate	-0.24	0.60	0.00	0.49
98 n-dodecane	-0.27	-0.39	-0.09	-0.96
99 benzaldehyde	-0.33	-2.30	-0.21	-3.30
100 methyl-i-propylketone	-0.15	2.59		
101 hexan-2-one	0.11	3.55		
102 hexan-3-one	0.00	2.89		
103 methyl-t-butylketone	-0.06	1.92		
104 n-propanol	0.02	3.01		
105 3-pentanol	-0.13	2.50		
106 2-methylbutanol	0.03	3.43		
107 3-methylbutanol	0.07	4.97		
108 3-methylbutan-2-ol	-0.05	2.88		
109 2-methylbutan-2-ol	-0.12	0.68		
110 ethylene glycol	-0.02	2.98		
111 propylene glycol	0.02	3.50		
112 diethylether	0.22	8.87		
113 diisopropylether	0.26	9.48		
114 ethyl-t-butylether	-0.10	3.54		
115 2-methoxyethanol	0.04	5.21		
116 2-ethoxyethanol	0.08	6.23		
117 1-butoxypropanol	-0.06	3.47		
118 methyl formate	-0.24	0.42		
119 t-butyl acetate	-0.12	0.53		
120 chloroform	-0.18	0.16		

Data Set 3.2 Integrated ozone production of 120 chemicals under 4 sets of conditions listed in Table 3.2 of the ozone reactivity study

1 ethylene	13.97	21.30	18.61	3.30
2 1,3,5-trimethylbenzene	13.55	17.20	14.25	1.20
3 3,5-dimethylethylbenzene	12.86	16.64	13.77	1.09
4 isoprene	12.30	20.44	17.06	1.42
5 3,5-diethyltoluene	12.14	16.35	13.53	1.03
6 butylene	12.12	23.93	20.95	3.38
7 m-xylene	11.99	17.25	14.36	1.46
8 propylene	11.91	20.66	18.54	2.28
9 1,3-butadiene	11.88	19.87	16.86	2.15
10 1,2,4-trimethylbenzene	11.82	15.76	13.09	1.05
11 1,2,3-trimethylbenzene	11.44	15.32	12.81	0.96
12 formaldehyde	11.28	17.73	15.36	2.56
13 but-1-ene	10.86	18.33	16.29	1.96
14 2-methylbut-1-ene	10.44	20.22	18.37	2.63
15 trans but-2-ene	10.42	14.53	12.45	2.20
16 p-xylene	9.60	14.44	12.14	1.18
17 diisopropylether	9.48		11.44	
18 m-ethyltoluene	9.46	14.06	11.90	1.13
19 o-xylene	9.41	13.98	11.81	0.87
20 diethylether	8.87		10.42	
21 3-methylbut-1-ene	8.73	16.52	15.06	1.37
22 i-butyraldehyde	8.63	12.94	11.56	0.36
23 pent-1-ene	8.40	14.81	13.39	1.28
24 p-ethyltoluene	7.64	11.73	10.14	0.94
25 o-ethyltoluene	7.34	11.04	9.71	0.89
26 hex-1-ene	6.93	12.97	11.66	0.91
27 propionaldehyde	6.49	9.73	8.83	0.07
28 2-ethoxyethanol	6.23		7.52	
29 ethylbenzene	6.11	9.66	8.36	0.95
30 toluene	5.83	9.16	7.94	0.94
31 butyraldehyde	5.68	8.96	8.00	-0.24
32 methyl-i-butylketone	5.61	7.84	7.28	0.75
33 2-methoxyethanol	5.21		6.35	
34 propylbenzene	5.17	8.54	7.39	0.87
35 pentanaldehyde	5.08	8.39	7.45	-0.29
36 i-propylbenzene	4.97	7.90	6.89	0.71
37 3-methylbutanol	4.97		6.72	
38 acetaldehyde	4.51	6.65	6.03	-0.30
39 i-butanol	4.35	6.67	6.07	0.97
40 cis but-2-ene	3.94	11.25	13.73	-0.06
41 2-butoxyethanol	3.89	5.61	5.12	0.80
42 trans pent-2-ene	3.66	9.30	12.99	-0.13
43 cis pent-2-ene	3.66	9.30	12.82	-0.15
44 1-methoxy-2-propanol	3.61	5.23	4.82	0.98
45 hexan-2-one	3.55		4.90	
46 ethyl-t-butylether	3.54		4.31	
47 cyclohexanol	3.53	5.23	4.86	0.78

48 propylene glycol	3.50		4.39	
49 1-butoxypropanol	3.47		4.07	
50 2-methylbutanol	3.43		4.85	
51 methylpropylketone	3.41	5.08	4.69	0.33
52 n-butanol	3.39	4.92	4.49	0.66
53 cis hex-2-ene	3.14	8.43	11.89	-0.39
54 2,3-dimethylbutane	3.13	4.70	4.26	0.55
55 trans hex-2-ene	3.13	8.43	11.89	-0.39
56 n-propanol	3.01		4.06	
57 i-butane	2.99	4.23	3.91	0.86
58 cis dichloroethylene	2.98	4.27	3.93	0.86
59 ethylene glycol	2.98		4.17	
60 hexan-3-one	2.89		4.05	
61 3-methylbutan-2-ol	2.88		3.93	
62 3-methylpentane	2.72	3.87	3.60	0.62
63 dimethylether	2.71	3.75	3.54	1.01
64 trans dichloroethylene	2.69	3.90	3.58	0.79
65 2-methylpentane	2.68	3.84	3.53	0.53
66 s-butanol	2.66	3.84	3.59	0.76
67 methyl-i-propylketone	2.59		3.76	
68 s-butyl acetate	2.58	3.73	3.48	0.68
69 i-pentane	2.57	3.62	3.37	0.68
70 3-pentanol	2.50		3.51	
71 cyclohexane	2.42	3.56	3.37	0.62
72 cyclohexane	2.42	3.56	3.37	0.62
73 cyclohexane	2.41	3.49	3.29	0.54
74 i-propyl acetate	2.36	3.40	3.19	0.68
75 methylethylketone	2.30	3.80	3.53	0.31
76 2-methylbut-2-ene	2.04	8.92	12.46	-0.94
77 diethylketone	1.99	3.15	2.94	0.28
78 n-pentane	1.97	2.79	2.61	0.47
79 2-methylhexane	1.96	2.86	2.66	0.28
80 n-propyl acetate	1.94	2.66	2.48	0.48
81 ethanol	1.93	2.83	2.65	0.45
82 methyl-t-butylketone	1.92		2.94	
83 trichloroethylene	1.88	2.66	2.47	0.57
84 benzene	1.88	2.74	2.55	0.47
85 3-methylhexane	1.85	2.75	2.54	0.23
86 diacetone alcohol	1.74	2.97	2.74	0.28
87 n-butane	1.73	2.42	2.27	0.49
88 i-propanol	1.71	2.61	2.44	0.61
89 n-hexane	1.68	2.44	2.25	0.30
90 ethyl acetate	1.43	2.04	1.92	0.40
91 2,2-dimethylbutane	1.35	2.02	1.87	0.30
92 n-butyl acetate	1.34	2.00	1.86	0.30
93 neopentane	1.34	1.94	1.79	0.40
94 methyl-t-butylether	1.31	1.92	1.80	0.41
95 methanol	1.21	1.86	1.75	0.40
96 styrene	1.20	5.89	6.58	0.39
97 propane	1.20	1.68	1.58	0.41
98 n-heptane	1.07	1.67	1.54	0.09

99 acetic acid	1.07	1.43	1.33	0.38
100 t-butanol	1.01	1.38	1.32	0.31
101 propanoic acid	0.85	1.28	1.22	0.32
102 2-methylbutan-2-ol	0.68		1.04	
103 methyl acetate	0.60	0.79	0.75	0.21
104 ethane	0.57	0.76	0.71	0.21
105 n-octane	0.54	0.91	0.84	-0.11
106 t-butyl acetate	0.53		0.74	
107 acetylene	0.52	0.87	0.78	0.19
108 acetone	0.51	0.94	0.91	0.16
109 methyl formate	0.42		0.44	
110 formic acid	0.33	0.37	0.37	0.14
111 methylene dichloride	0.21	0.33	0.34	0.13
112 chloroform	0.16		0.19	
113 n-nonane	0.13	0.41	0.36	-0.24
114 tetrachloroethylene	0.12	0.20	0.18	0.10
115 methyl chloroform	0.11	0.16	0.17	0.10
116 methyl chloride	0.04	0.07	0.08	0.07
117 n-decane	-0.15	0.11	0.08	-0.33
118 n-undecane	-0.22	-0.09	-0.11	-0.42
119 n-dodecane	-0.39	-0.28	-0.28	-0.49
120 benzaldehyde	-2.30	-2.10	-2.43	-1.63

Data Set 3.4 Integrated ozone production of 120 chemicals under 2 sets of conditions listed in Table 3.4 of the ozone reactivity study

1 ethylene	18.61	87.80
2 butylene	20.95	82.80
3 formaldehyde	15.36	78.34
4 diisopropylether	11.44	67.06
5 2-methylbut-1-ene	18.37	66.81
6 diethylether	10.42	60.26
7 trans but-2-ene	12.45	56.00
8 m-xylene	14.36	55.15
9 but-1-ene	16.29	52.14
10 1,3-butadiene	16.86	52.13
11 1,3,5-trimethylbenzene	14.25	50.87
12 3,5-diethyltoluene	13.53	50.40
13 3,5-dimethylethylbenzene	13.77	50.21
14 p-xylene	12.14	49.98
15 m-ethyltoluene	11.90	49.58
16 propylene	18.54	49.01
17 1,2,4-trimethylbenzene	13.09	48.22
18 1,2,3-trimethylbenzene	12.81	46.84
19 3-methylbut-1-ene	15.06	46.35
20 pent-1-ene	13.39	46.17
21 p-ethyltoluene	10.14	45.13
22 o-xylene	11.81	44.67
23 hex-1-ene	11.66	43.55
24 o-ethyltoluene	9.71	43.29
25 ethylbenzene	8.36	42.58
26 i-butyraldehyde	11.56	41.80
27 propylbenzene	7.39	41.74
28 toluene	7.94	40.97
29 2-ethoxyethanol	7.52	40.80
30 i-propylbenzene	6.89	36.68
31 propionaldehyde	8.83	36.49
32 isoprene	17.06	36.24
33 methyl-i-butylketone	7.28	35.37
34 3-methylbutanol	6.72	35.33
35 pentanaldehyde	7.45	34.98
36 butyraldehyde	8.00	33.60
37 i-butanol	6.07	33.49
38 2-methoxyethanol	6.35	33.21
39 acetaldehyde	6.03	27.53
40 styrene	6.58	26.99
41 hexan-2-one	4.90	25.37
42 2-methylbutanol	4.85	25.02
43 methylpropylketone	4.69	24.05
44 propylene glycol	4.39	22.16
45 ethylene glycol	4.17	21.40
46 cis dichloroethylene	3.93	21.13

47 2-butoxyethanol	5.12	20.08
48 hexan-3-one	4.05	19.93
49 n-butanol	4.49	19.69
50 methyl-i-propylketone	3.76	18.85
51 2,3-dimethylbutane	4.26	18.77
52 methylethylketone	3.53	17.63
53 ethyl-t-butylether	4.31	16.76
54 trans dichloroethylene	3.58	16.35
55 cis hex-2-ene	11.89	14.68
56 trans hex-2-ene	11.89	14.68
57 n-propanol	4.06	14.17
58 methyl-t-butylketone	2.94	13.70
59 cyclohexanol	4.86	13.31
60 trans pent-2-ene	12.99	13.26
61 1-methoxy-2-propanol	4.82	13.13
62 cis pent-2-ene	12.82	12.94
63 diacetone alcohol	2.74	11.68
64 cis but-2-ene	13.73	11.17
65 diethylketone	2.94	10.54
66 i-butane	3.91	10.26
67 3-methylbutan-2-ol	3.93	7.75
68 s-butyl acetate	3.48	7.00
69 1-butoxypropanol	4.07	6.77
70 2-methylbut-2-ene	12.46	6.76
71 2-methylpentane	3.53	6.40
72 s-butanol	3.59	5.30
73 3-pentanol	3.51	4.53
74 3-methylpentane	3.60	3.85
75 i-propyl acetate	3.19	3.66
76 cyclohexane	3.29	2.19
77 i-pentane	3.37	1.11
78 trichloroethylene	2.47	0.05
79 benzene	2.55	-0.66
80 3-methylhexane	2.54	-1.24
81 benzaldehyde	-2.43	-1.73
82 ethanol	2.65	-1.93
83 cyclohexane	3.37	-2.10
84 cyclohexane	3.37	-2.10
85 n-propyl acetate	2.48	-2.11
86 2-methylhexane	2.66	-2.87
87 dimethylether	3.54	-3.02
88 n-pentane	2.61	-3.81
89 2,2-dimethylbutane	1.87	-4.81
90 n-hexane	2.25	-5.05
91 i-propanol	2.44	-5.44
92 n-butyl acetate	1.86	-5.61
93 n-butane	2.27	-5.74
94 ethyl acetate	1.92	-7.39
95 methanol	1.75	-7.74
96 neopentane	1.79	-7.87
97 n-heptane	1.54	-8.42

98 methyl-t-butylether	1.80	-9.73
99 acetone	0.91	-9.76
100 propane	1.58	-10.42
101 n-octane	0.84	-10.50
102 t-butanol	1.32	-11.35
103 n-nonane	0.36	-11.76
104 acetic acid	1.33	-12.07
105 n-decane	0.08	-12.36
106 propanoic acid	1.22	-12.54
107 2-methylbutan-2-ol	1.04	-12.59
108 n-undecane	-0.11	-12.67
109 n-dodecane	-0.28	-12.96
110 acetylene	0.78	-13.57
111 t-butyl acetate	0.74	-14.15
112 methyl acetate	0.75	-14.48
113 ethane	0.71	-14.67
114 methyl formate	0.44	-15.87
115 formic acid	0.37	-16.07
116 methylene dichloride	0.34	-16.10
117 tetrachloroethylene	0.18	-16.64
118 chloroform	0.19	-16.65
119 methyl chloroform	0.17	-16.79
120 methyl chloride	0.08	-17.08

Data Set 3.5 Integrated ozone production of 120 chemicals under 2 sets of conditions listed in Table 3.5 of the ozone reactivity study

1 cis but-2-ene	13.73	48.40
2 cis pent-2-ene	12.82	44.46
3 trans pent-2-ene	12.99	44.44
4 1,3,5-trimethylbenzene	14.25	43.57
5 2-methylbut-2-ene	12.46	41.52
6 3,5-dimethylethylbenzene	13.77	41.23
7 propylene	18.54	40.40
8 isoprene	17.06	40.27
9 cis hex-2-ene	11.89	39.95
10 trans hex-2-ene	11.89	39.95
11 3,5-diethyltoluene	13.53	39.25
12 1,2,4-trimethylbenzene	13.09	37.61
13 1,2,3-trimethylbenzene	12.81	37.25
14 butylene	20.95	34.91
15 but-1-ene	16.29	34.52
16 1,3-butadiene	16.86	34.26
17 m-xylene	14.36	34.14
18 2-methylbut-1-ene	18.37	32.68
19 3-methylbut-1-ene	15.06	32.45
20 formaldehyde	15.36	29.24
21 pent-1-ene	13.39	28.94
22 ethylene	18.61	28.03
23 o-xylene	11.81	27.47
24 m-ethyltoluene	11.90	26.63
25 p-xylene	12.14	26.57
26 i-butyraldehyde	11.56	25.48
27 hex-1-ene	11.66	25.03
28 propionaldehyde	8.83	23.63
29 trans but-2-ene	12.45	23.46
30 butyraldehyde	8.00	21.98
31 p-ethyltoluene	10.14	20.34
32 pentanaldehyde	7.45	19.91
33 o-ethyltoluene	9.71	18.93
34 styrene	6.58	17.32
35 acetaldehyde	6.03	17.28
36 methyl-i-butylketone	7.28	13.65
37 ethylbenzene	8.36	12.32
38 toluene	7.94	11.25
39 propylbenzene	7.39	9.78
40 i-propylbenzene	6.89	9.26
41 methylpropylketone	4.69	9.24
42 i-butanol	6.07	6.84
43 methylethylketone	3.53	6.32
44 2-butoxyethanol	5.12	5.35
45 n-butanol	4.49	5.26
46 2,3-dimethylbutane	4.26	5.01

47 cyclohexanol	4.86	4.83
48 benzaldehyde	-2.43	4.53
49 diacetone alcohol	2.74	4.37
50 1-methoxy-2-propanol	4.82	4.20
51 2-methylpentane	3.53	3.35
52 3-methylpentane	3.60	3.13
53 cis dichloroethylene	3.93	2.88
54 s-butanol	3.59	2.78
55 3-methylhexane	2.54	2.69
56 cyclohexane	3.29	2.68
57 i-butane	3.91	2.67
58 s-butyl acetate	3.48	2.62
59 trans dichloroethylene	3.58	2.55
60 i-pentane	3.37	2.50
61 2-methylhexane	2.66	2.48
62 cyclohexane	3.37	2.18
63 cyclohexane	3.37	2.18
64 i-propyl acetate	3.19	2.13
65 n-hexane	2.25	2.04
66 ethanol	2.65	2.02
67 n-pentane	2.61	1.99
68 n-propyl acetate	2.48	1.84
69 dimethylether	3.54	1.76
70 benzene	2.55	1.66
71 trichloroethylene	2.47	1.62
72 2,2-dimethylbutane	1.87	1.56
73 n-heptane	1.54	1.56
74 n-butyl acetate	1.86	1.53
75 n-butane	2.27	1.52
76 i-propanol	2.44	1.40
77 n-octane	0.84	1.19
78 ethyl acetate	1.92	1.18
79 acetone	0.91	1.10
80 neopentane	1.79	1.06
81 methanol	1.75	1.00
82 methyl-t-butylether	1.80	0.96
83 n-nonane	0.36	0.89
84 propane	1.58	0.82
85 n-decane	0.08	0.73
86 t-butanol	1.32	0.67
87 n-undecane	-0.11	0.64
88 acetic acid	1.33	0.58
89 n-dodecane	-0.28	0.57
90 propanoic acid	1.22	0.54
91 acetylene	0.78	0.41
92 methyl acetate	0.75	0.28
94 methylene dichloride	0.34	0.02
95 formic acid	0.37	0.01
93 ethane	0.71	0.00
99 diethylketone	2.94	0.00
100 methyl-i-propylketone	3.76	0.00

96 tetrachloroethylene	0.18	-0.06
97 methyl chloroform	0.17	-0.09
98 methyl chloride	0.08	-0.14
101 hexan-2-one	4.90	
102 hexan-3-one	4.05	
103 methyl-t-butylketone	2.94	
104 n-propanol	4.06	
105 3-pentanol	3.51	
106 2-methylbutanol	4.85	
107 3-methylbutanol	6.72	
108 3-methylbutan-2-ol	3.93	
109 2-methylbutan-2-ol	1.04	
110 ethylene glycol	4.17	
111 propylene glycol	4.39	
112 diethylether	10.42	
113 diisopropylether	11.44	
114 ethyl-t-butylether	4.31	
115 2-methoxyethanol	6.35	
116 2-ethoxyethanol	7.52	
117 1-butoxypropanol	4.07	
118 methyl formate	0.44	
119 t-butyl acetate	0.74	
120 chloroform	0.19	

APPENDIX 2: CLASSIFICATION OF VOCS

ALCOHOLS
- 34 3-methylbutanol
- 37 i-butanol
- 42 2-methylbutanol
- 44 propylene glycol
- 45 ethylene glycol
- 49 n-butanol
- 57 n-propanol
- 59 cyclohexanol
- 63 diacetone alcohol
- 67 3-methylbutan-2-ol
- 72 s-butanol
- 73 3-pentanol
- 91 i-propanol
- 95 methanol
- 102 t-butanol
- 107 2-methylbutan-2-ol

ALKANES
- 51 2,3-dimethylbutane
- 66 i-butane
- 71 2-methylpentane
- 74 3-methylpentane
- 76 cyclohexane
- 77 i-pentane
- 80 3-methylhexane
- 83 cyclohexane
- 84 cyclohexane
- 86 2-methylhexane
- 88 n-pentane
- 89 2,2-dimethylbutane
- 90 n-hexane
- 93 n-butane
- 96 neopentane
- 97 n-heptane
- 100 propane
- 101 n-octane
- 103 n-nonane
- 105 n-decane
- 108 n-undecane
- 109 n-dodecane
- 113 ethane

ALKENES
- 1 ethylene
- 2 butylene
- 5 2-methylbut-1-ene
- 7 trans but-2-ene
- 9 but-1-ene
- 16 propylene
- 19 3-methylbut-1-ene
- 20 pent-1-ene
- 23 hex-1-ene
- 32 isoprene
- 55 cis hex-2-ene
- 56 trans hex-2-ene
- 60 trans pent-2-ene
- 62 cis pent-2-ene
- 64 cis but-2-ene
- 70 2-methylbut-2-ene

ALDEHYDES
- 3 formaldehyde
- 26 i-butyraldehyde
- 31 propionaldehyde
- 35 pentanaldehyde
- 36 butyraldehyde
- 39 acetaldehyde
- 81 benzaldehyde

AROMATICS
- 8 m-xylene
- 10 1,3-butadiene
- 11 1,3,5-trimethylbenzene
- 17 1,2,4-trimethylbenzene
- 18 1,2,3-trimethylbenzene
- 12 3,5-diethyltoluene
- 13 3,5-dimethylethylbenzene
- 14 p-xylene
- 15 m-ethyltoluene
- 21 p-ethyltoluene
- 24 o-ethyltoluene
- 25 ethylbenzene
- 27 propylbenzene
- 28 toluene
- 22 o-xylene
- 30 i-propylbenzene
- 40 styrene
- 79 benzene

ETHERS
- 4 diisopropylether
- 29 2-ethoxyethanol
- 6 diethylether
- 38 2-methoxyethanol
- 47 2-butoxyethanol
- 53 ethyl-t-butylether
- 61 1-methoxy-2-propanol
- 69 1-butoxypropanol
- 82 ethanol
- 87 dimethylether
- 98 methyl-t-butylether

KETONES
- 33 methyl-i-butylketone
- 41 hexan-2-one
- 43 methylpropylketone
- 48 hexan-3-one
- 50 methyl-i-propylketone
- 52 methylethylketone
- 58 methyl-t-butylketone
- 65 diethylketone
- 99 acetone

ASSORTED ACIDS
- 104 acetic acid
- 106 propanoic acid
- 115 formic acid

ESTERS
- 68 s-butyl acetate
- 75 i-propyl acetate
- 85 n-propyl acetate
- 92 n-butyl acetate
- 94 ethyl acetate
- 111 t-butyl acetate
- 112 methyl acetate
- 114 methyl formate

ALKYNES
- 110 acetylene

CHLOROCARBONS
- 46 cis dichloroethylene
- 54 trans dichloroethylene
- 78 trichloroethylene
- 116 methylene dichloride
- 117 tetrachloroethylene
- 118 chloroform
- 119 methyl chloroform
- 120 methyl chloride

APPENDIX 3: PRELIMINARY CLUSTERING RESULTS

Preliminary attempts to cluster chemicals based on the results in Table 3.1 and Table 3.2 of the ozone reactivity study are given below.

Preliminary clustering of Table 3.1

Description	Kohonen cluster 1x3	Description	TwoStep
1,2,3-trimethylbenzene	00	1,2,3-trimethylbenzene	1
1,2,4-trimethylbenzene	00	1,2,4-trimethylbenzene	1
1,3,5-trimethylbenzene	00	1,3,5-trimethylbenzene	1
1,3-butadiene	00	1,3-butadiene	1
2-ethoxyethanol	00	2-methylbut-1-ene	1
2-methylbut-1-ene	00	3,5-diethyltoluene	1
2-methylbut-2-ene	00	3,5-dimethylethylbenzene	1
3,5-diethyltoluene	00	3-methylbut-1-ene	1
3,5-dimethylethylbenzene	00	but-1-ene	1
3-methylbut-1-ene	00	butylene	1
but-1-ene	00	butyraldehyde	1
butylene	00	ethylbenzene	1
diethylether	00	ethylene	1
diisopropylether	00	formaldehyde	1
ethylbenzene	00	hex-1-ene	1
ethylene	00	i-butyraldehyde	1
formaldehyde	00	i-propylbenzene	1
hex-1-ene	00	isoprene	1
i-butyraldehyde	00	methyl-i-butylketone	1
isoprene	00	m-ethyltoluene	1
m-ethyltoluene	00	m-xylene	1
m-xylene	00	o-ethyltoluene	1
o-ethyltoluene	00	o-xylene	1
o-xylene	00	pent-1-ene	1
pent-1-ene	00	pentanaldehyde	1
p-ethyltoluene	00	p-ethyltoluene	1
propionaldehyde	00	propionaldehyde	1
propylene	00	propylbenzene	1
p-xylene	00	propylene	1
trans but-2-ene	00	p-xylene	1
1-butoxypropanol	01	toluene	1
1-methoxy-2-propanol	01	trans but-2-ene	1
2-butoxyethanol	01	1-methoxy-2-propanol	2
2-methoxyethanol	01	2,2-dimethylbutane	2
2-methylbutanol	01	2,3-dimethylbutane	2
3-methylbutan-2-ol	01	2-butoxyethanol	2
3-methylbutanol	01	2-methylbut-2-ene	2
3-pentanol	01	2-methylhexane	2
acetaldehyde	01	2-methylpentane	2
butyraldehyde	01	3-methylhexane	2
cis but-2-ene	01	3-methylpentane	2

cis hex-2-ene	01	acetaldehyde	2
cis pent-2-ene	01	acetic acid	2
ethylene glycol	01	acetone	2
ethyl-t-butylether	01	acetylene	2
hexan-2-one	01	benzaldehyde	2
hexan-3-one	01	benzene	2
i-butanol	01	cis but-2-ene	2
i-propylbenzene	01	cis dichloroethylene	2
methyl-i-butylketone	01	cis hex-2-ene	2
methyl-i-propylketone	01	cis pent-2-ene	2
methylpropylketone	01	cyclohexane	2
methyl-t-butylketone	01	cyclohexane	2
n-propanol	01	cyclohexane	2
pentanaldehyde	01	cyclohexanol	2
propylbenzene	01	diacetone alcohol	2
propylene glycol	01	diethylketone	2
toluene	01	dimethylether	2
trans hex-2-ene	01	ethane	2
trans pent-2-ene	01	ethanol	2
2,2-dimethylbutane	02	ethyl acetate	2
2,3-dimethylbutane	02	formic acid	2
2-methylbutan-2-ol	02	i-butane	2
2-methylhexane	02	i-butanol	2
2-methylpentane	02	i-pentane	2
3-methylhexane	02	i-propanol	2
3-methylpentane	02	i-propyl acetate	2
acetic acid	02	methanol	2
acetone	02	methyl acetate	2
acetylene	02	methyl chloride	2
benzaldehyde	02	methyl chloroform	2
benzene	02	methylene dichloride	2
chloroform	02	methylethylketone	2
cis dichloroethylene	02	methylpropylketone	2
cyclohexane	02	methyl-t-butylether	2
cyclohexane	02	n-butane	2
cyclohexane	02	n-butanol	2
cyclohexanol	02	n-butyl acetate	2
diacetone alcohol	02	n-decane	2
diethylketone	02	n-dodecane	2
dimethylether	02	neopentane	2
ethane	02	n-heptane	2
ethanol	02	n-hexane	2
ethyl acetate	02	n-nonane	2
formic acid	02	n-octane	2
i-butane	02	n-pentane	2
i-pentane	02	n-propyl acetate	2
i-propanol	02	n-undecane	2
i-propyl acetate	02	propane	2
methanol	02	propanoic acid	2
methyl acetate	02	s-butanol	2
methyl chloride	02	s-butyl acetate	2

methyl chloroform	02	styrene	2
methyl formate	02	t-butanol	2
methylene dichloride	02	tetrachloroethylene	2
methylethylketone	02	trans dichloroethylene	2
methyl-t-butylether	02	trans hex-2-ene	2
n-butane	02	trans pent-2-ene	2
n-butanol	02	trichloroethylene	2
n-butyl acetate	02	1-butoxypropanol	
n-decane	02	2-ethoxyethanol	
n-dodecane	02	2-methoxyethanol	
neopentane	02	2-methylbutan-2-ol	
n-heptane	02	2-methylbutanol	
n-hexane	02	3-methylbutan-2-ol	
n-nonane	02	3-methylbutanol	
n-octane	02	3-pentanol	
n-pentane	02	chloroform	
n-propyl acetate	02	diethylether	
n-undecane	02	diisopropylether	
propane	02	ethylene glycol	
propanoic acid	02	ethyl-t-butylether	
s-butanol	02	hexan-2-one	
s-butyl acetate	02	hexan-3-one	
styrene	02	methyl formate	
t-butanol	02	methyl-i-propylketone	
t-butyl acetate	02	methyl-t-butylketone	
tetrachloroethylene	02	n-propanol	
trans dichloroethylene	02	propylene glycol	
trichloroethylene	02	t-butyl acetate	

Preliminary clustering of Table 3.2

Description	Kohonen cluster 3x3	Description	Kohonen cluster 3x1	Description	TwoStep
2-methylbut-2-ene	00	2,2-dimethylbutane	00	1-methoxy-2-propanol	1
acetaldehyde	00	2,3-dimethylbutane	00	2,2-dimethylbutane	1
butyraldehyde	00	2-methylbutan-2-ol	00	2,3-dimethylbutane	1
cis but-2-ene	00	2-methylhexane	00	2-butoxyethanol	1
cis hex-2-ene	00	2-methylpentane	00	2-methylbut-2-ene	1
cis pent-2-ene	00	3-methylhexane	00	2-methylhexane	1
pentanaldehyde	00	3-methylpentane	00	2-methylpentane	1
propionaldehyde	00	acetic acid	00	3-methylhexane	1
trans hex-2-ene	00	acetone	00	3-methylpentane	1
trans pent-2-ene	00	acetylene	00	acetaldehyde	1
2,3-dimethylbutane	01	benzaldehyde	00	acetic acid	1
cyclohexanol	01	benzene	00	acetone	1
methylethylketone	01	chloroform	00	acetylene	1
methylpropyl-ketone	01	cis dichloroethylene	00	benzaldehyde	1
n-butanol	01	cyclohexane	00	benzene	1
styrene	01	cyclohexane	00	butyraldehyde	1
2,2-dimethylbutane	02	cyclohexane	00	cis but-2-ene	1
2-methylhexane	02	diacetone alcohol	00	cis dichloroethylene	1
3-methylhexane	02	diethylketone	00	cis hex-2-ene	1
acetic acid	02	dimethylether	00	cis pent-2-ene	1
acetone	02	ethane	00	cyclohexane	1
acetylene	02	ethanol	00	cyclohexane	1
benzaldehyde	02	ethyl acetate	00	cyclohexane	1
benzene	02	formic acid	00	cyclohexanol	1
diacetone alcohol	02	i-butane	00	diacetone alcohol	1
diethylketone	02	i-pentane	00	diethylketone	1
ethane	02	i-propanol	00	dimethylether	1
ethanol	02	i-propyl acetate	00	ethane	1
ethyl acetate	02	methanol	00	ethanol	1
formic acid	02	methyl acetate	00	ethyl acetate	1
i-propanol	02	methyl chloride	00	ethylbenzene	1
methanol	02	methyl chloroform	00	formic acid	1
methyl acetate	02	methyl formate	00	i-butane	1
methyl chloride	02	methylene dichloride	00	i-butanol	1
methyl chloroform	02	methylethylketone	00	i-pentane	1
methylene dichloride	02	methylpropylketone	00	i-propanol	1
methyl-t-butylether	02	methyl-t-butylether	00	i-propyl acetate	1
n-butane	02	n-butane	00	i-propylbenzene	1
n-butyl acetate	02	n-butanol	00	methanol	1
n-decane	02	n-butyl acetate	00	methyl acetate	1
n-dodecane	02	n-decane	00	methyl chloride	1
neopentane	02	n-dodecane	00	methyl chloroform	1
n-heptane	02	neopentane	00	methylene dichloride	1
n-hexane	02	n-heptane	00	methylethylketone	1

n-nonane	02	n-hexane	00	methyl-i-butylketone	1
n-octane	02	n-nonane	00	methylpropylketone	1
n-pentane	02	n-octane	00	methyl-t-butylether	1
n-propyl acetate	02	n-pentane	00	n-butane	1
n-undecane	02	n-propyl acetate	00	n-butanol	1
propane	02	n-undecane	00	n-butyl acetate	1
propanoic acid	02	propane	00	n-decane	1
t-butanol	02	propanoic acid	00	n-dodecane	1
tetrachloroethylene	02	s-butanol	00	neopentane	1
trichloroethylene	02	s-butyl acetate	00	n-heptane	1
diethylether	10	styrene	00	n-hexane	1
diisopropylether	10	t-butanol	00	n-nonane	1
hex-1-ene	10	t-butyl acetate	00	n-octane	1
i-butyraldehyde	10	tetrachloroethylene	00	n-pentane	1
o-ethyltoluene	10	trans dichloroethylene	00	n-propyl acetate	1
p-ethyltoluene	10	trichloroethylene	00	n-undecane	1
2-butoxyethanol	11	1-butoxypropanol	10	pentanaldehyde	1
i-butanol	11	1-methoxy-2-propanol	10	propane	1
i-propylbenzene	11	2-butoxyethanol	10	propanoic acid	1
methyl-i-butylketone	11	2-methoxyethanol	10	propionaldehyde	1
propylbenzene	11	2-methylbut-2-ene	10	propylbenzene	1
2-methylpentane	12	2-methylbutanol	10	s-butanol	1
3-methylpentane	12	3-methylbutan-2-ol	10	s-butyl acetate	1
cis dichloroethylene	12	3-methylbutanol	10	styrene	1
cyclohexane	12	3-pentanol	10	t-butanol	1
cyclohexane	12	acetaldehyde	10	tetrachloroethylene	1
cyclohexane	12	butyraldehyde	10	toluene	1
dimethylether	12	cis hex-2-ene	10	trans dichloroethylene	1
i-butane	12	cyclohexanol	10	trans hex-2-ene	1
i-pentane	12	ethylene glycol	10	trans pent-2-ene	1
i-propyl acetate	12	ethyl-t-butylether	10	trichloroethylene	1
s-butanol	12	hexan-2-one	10	1,2,3-trimethylbenzene	2
s-butyl acetate	12	hexan-3-one	10	1,2,4-trimethylbenzene	2
trans dichloroethylene	12	i-butanol	10	1,3,5-trimethylbenzene	2
1,2,3-trimethylbenzene	20	i-propylbenzene	10	1,3-butadiene	2
1,2,4-trimethylbenzene	20	methyl-i-butylketone	10	2-methylbut-1-ene	2
1,3,5-trimethylbenzene	20	methyl-i-propylketone	10	3,5-diethyltoluene	2
1,3-butadiene	20	methyl-t-butylketone	10	3,5-di-methylethylbenzene	2
2-methylbut-1-ene	20	n-propanol	10	3-methylbut-1-ene	2
3,5-diethyltoluene	20	pentanaldehyde	10	but-1-ene	2
3,5-dimethyl-	20	propylbenzene	10	butylene	2

ethylbenzene					
3-methylbut-1-ene	20	propylene glycol	10	ethylene	2
but-1-ene	20	trans hex-2-ene	10	formaldehyde	2
butylene	20	1,2,3-trimethylbenzene	20	hex-1-ene	2
ethylene	20	1,2,4-trimethylbenzene	20	i-butyraldehyde	2
formaldehyde	20	1,3,5-trimethylbenzene	20	isoprene	2
isoprene	20	1,3-butadiene	20	m-ethyltoluene	2
m-ethyltoluene	20	2-ethoxyethanol	20	m-xylene	2
m-xylene	20	2-methylbut-1-ene	20	o-ethyltoluene	2
o-xylene	20	3,5-diethyltoluene	20	o-xylene	2
pent-1-ene	20	3,5-dimethylethylbenzene	20	pent-1-ene	2
propylene	20	3-methylbut-1-ene	20	p-ethyltoluene	2
p-xylene	20	but-1-ene	20	propylene	2
trans but-2-ene	20	butylene	20	p-xylene	2
2-ethoxyethanol	21	cis but-2-ene	20	trans but-2-ene	2
2-methoxyethanol	21	cis pent-2-ene	20	1-butoxypropanol	
3-methylbutanol	21	diethylether	20	2-ethoxyethanol	
ethylbenzene	21	diisopropylether	20	2-methoxyethanol	
toluene	21	ethylbenzene	20	2-methylbutan-2-ol	
1-butoxypropanol	22	ethylene	20	2-methylbutanol	
1-methoxy-2-propanol	22	formaldehyde	20	3-methylbutan-2-ol	
2-methylbutan-2-ol	22	hex-1-ene	20	3-methylbutanol	
2-methylbutanol	22	i-butyraldehyde	20	3-pentanol	
3-methylbutan-2-ol	22	isoprene	20	chloroform	
3-pentanol	22	m-ethyltoluene	20	diethylether	
chloroform	22	m-xylene	20	diisopropylether	
ethylene glycol	22	o-ethyltoluene	20	ethylene glycol	
ethyl-t-butylether	22	o-xylene	20	ethyl-t-butylether	
hexan-2-one	22	pent-1-ene	20	hexan-2-one	
hexan-3-one	22	p-ethyltoluene	20	hexan-3-one	
methyl formate	22	propionaldehyde	20	methyl formate	
methyl-i-propylketone	22	propylene	20	methyl-i-propylketone	
methyl-t-butylketone	22	p-xylene	20	methyl-t-butylketone	
n-propanol	22	toluene	20	n-propanol	
propylene glycol	22	trans but-2-ene	20	propylene glycol	
t-butyl acetate	22	trans pent-2-ene	20	t-butyl acetate	

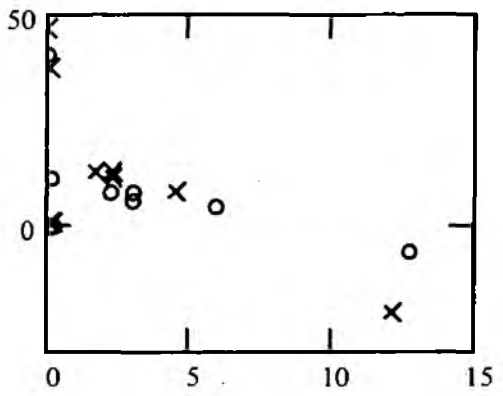
APPENDIX 4: INTERPRETATION OF RESULTS IN TERMS OF THE RATIO OF VOC TO NOx

On many occasions the results of complex calculations of ozone production are interpreted in terms of the ratio of VOC to NOx.

It is interesting to note from Table 3.1, that a 25 fold increase in VOC emission rate (from 0.4 to 10 t/h) does not lead to a 25 fold increase in ozone for most of the VOCs. One can start to attempt to group the VOC behaviour. The first group of highly reactive VOCs produce ozone, but the response is sub-linear at the higher VOC emission rate. There is a group of VOC species (cis but-2-ene, 2-methyl-2-ene, trans pent-2-ene, cis pent-2-ene, trans-2-hex-2-ene and cis hex-2-ene), which shows a different behaviour. For 10t/h emission rate representing the larger ratio, R, of VOC to NOx, there is still a small positive ozone response. For a 50% reduction in the NOx and VOC in the background air, in which R is much greater than 1, there is a small positive response for 0.4t/h. response. This latter case corresponds reasonably closely to the future situation when the UN ECE Gothenburg Protocol and the EU National Emissions Ceilings Directive are fully implemented. For 10t/h representing the largest R ratio considered, the ozone production is small and negative. The cases represent approximately situations of increasing VOC to NOx ratio, R, and for VOCs in this group, the production is becoming independent of VOC emissions and "NOx limited".

An attempt has been made to explain results in terms of the ratio of VOC to NOx. This cannot be defined precisely, but is defined approximately as the ratio R of VOC to NOx on a molecular basis, entering the system. For two reactive VOC species, toluene and 1,2,3-trimethylbenzene, the integrated downwind ozone production has been compared with R for each of the conditions described by the input data. This is plotted below. For 1,2,3-trimethylbenzene there is a weak indication of an inverse trend over the nine conditions considered, corresponding to the four columns of Table 3.1, the third column of Table 3.2, Table 3.3, and the second columns of Table 3.4, Table 3.5 and Table 3.6 of Derwent and Nelson (2003).

Figure A4.1 *The integrated downwind ozone production in ppb (y-axis) is plotted against an estimate of the ratio R of VOC to NOx concentration near the start of the calculation (x-axis) for two VOCs, 1,2,3-trimethylbenzene and toluene. R is not defined very precisely for these calculations.*



Ratio R of VOC to NOx concentration
× denotes the 1,2,3-trimethylbenzene and 0 the toluene relationship

For toluene a consistent relationship is also not found. It is concluded that a simple interpretation of results using some expert judgement is not possible, although some consistent trends seem to occur in the data set.

FEEDBACK FORM

To help us improve the usefulness of our reports and guidance notes, we would be grateful if you would take a few moments to complete and return this brief questionnaire.

1. Dissemination

(a) How did you hear about this report?

_____

_____

_____

(b) How did you obtain a copy (if different from above)?

_____

_____

_____

2. Content

How useful did you find the material in this report?

Excellent [ ] Good [ ] Satisfactory [ ] Poor [ ]

3. Presentation

How clear was the presentation of the material?

Excellent [ ] Good [ ] Satisfactory [ ] Poor [ ]

4. Follow-up

If you contacted **Risk & Forecasting** staff about this document, how helpful was the response?

Excellent [ ] Good [ ] Satisfactory [ ] Poor [ ]

5. More information (optional)

Would you be interested in receiving notification of similar reports, guidance notes and other information in the future? If so, please provide your email address here. (It will not be disclosed to a third party.)

_____

CONTINUED OVERLEAF

6. Your own details (optional)

Name _____

Agency department / Organisation _____

_____

7. Any other comments?

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

Please return this form to:

The Marketing Officer
Environmental Policy – Risk and Forecasting
Environment Agency
Kings Meadow House
Kings Meadow Road
Reading RG1 8DQ

Or fax to: 0118 953 5265