

Non-Destructive Evaluation of Apple Maturity Using an Electronic Nose System

Lakshmi Prasad Pathange

Thesis submitted to the faculty of

Virginia Polytechnic Institute and State University

In partial fulfillment of the requirement for the degree of

Master of Science

in

Biological Systems Engineering

Parameswarakumar Mallikarjunan, Chair

Richard P Marini

Sean O' Keefe

David H. Vaughan

February 6th, 2003

Blacksburg Virginia

Keywords: Electronic Nose, Maturity Indices, Statistical Analysis, Apples, Gala, York

Copyright 2003, Lakshmi Prasad Pathange

ABSTRACT

Non-Destructive Evaluation of Apple Maturity

Using an Electronic Nose System

Lakshmi Prasad Pathange

Virginia Tech, 2003

Academic Advisor:

Dr. Kumar Mallikarjunan

The apple growers and packaging houses are interested in methods that can evaluate the quality of apples non-destructively. Harvested fruits are a mixture of immature, mature, and over mature fruits, thereby posing a great problem in deciding their end use and storage time. It is expected that the technique developed from the present project could be effectively used to classify the harvested fruit into immature, mature and over mature apples, rapidly and non-destructively. It would also help the growers to predict the optimum dates to harvest the fruits.

York and Gala were the varieties of apples that were used in this study and were obtained from Virginia Tech College of Agriculture and Life Sciences Kentland Farm. Apples were harvested at different times resulting in different maturity groups (immature, mature and ripe). Gala apples were harvested on three dates with an interval of 10 days, while York apples were harvested on four dates with an interval of 14 days. They were stored at 0°C until sampled. For each harvest date, the experiments were conducted in two sets (10 each) on two consecutive days. First the ethylene levels were measured, followed by gas chromatograph and electronic nose. Then the maturity indices were measured.

Three maturity indices, starch index, firmness and soluble solids were used as the three variables for the statistical analysis to identify and categorize the fruits into three maturity categories referred as immature, mature and over mature fruits. Apples were also categorized into three maturity groups based on the emanation levels of the aroma compounds evolved from the fruits. Then electronic nose sensor responses were categorized into the above maturity categories, and their effectiveness was determined using a statistical procedure called Discriminant Analysis (DA).

From the DA cross validation results the correct classification percentage for Gala and York apples into maturity groups was 95%. The Electronic nose sensor's effectiveness to categorize the same observations based on sensor responses in to the above classified maturity categories was 83% correct in case Gala apples and 69% for York apples. The EN sensors response data were analyzed by the EN system software and the correct classification percentage for Gala was 83% and for York was 81%. Aroma-based categorization for Gala apples was 100% correct, while the electronic nose for the same analysis was 80%.

Based on the three physical parameters, an objective evaluation of maturity could be accomplished. Principal Component Analysis, Canonical Discriminant Analysis and DA results demonstrated that the electronic nose could be used to classify apples into three identified maturity-based groups. The EN sensors (Gala apples), could also classify the apples into aroma-based categories. Thus, it can be concluded that the EN system holds promise as non-destructive evaluation technique to determine the maturity of an apple.

Acknowledgements

I am very grateful to my academic advisor, Dr. Parameswarakumar Mallikarjunan, for his consideration, support, valuable guidance, time and encouragement, which has immensely helped me to complete my Master's degree at Virginia Tech. I sincerely thank Dr. Richard P. Marini, who served on my graduate committee and with whom I had very valuable discussions. I sincerely thank Dr. Sean O' Keefe for his valuable guidance and serving on my committee. I also thank Dr. David H. Vaughan for his valuable comments and support as a member of my graduate committee.

I convey my special thanks to Ms. Harriet Williams for her invaluable help and constant support, without which my experiments would have been incomplete. I am grateful to Dr. Thangaraju Murugesan, for his invaluable support in completing my experiments. I sincerely thank Dr. John Perumpral for all the help he has extended to me in completing my study. I would like to thank Mr. John Chandler, Mr. Donald L. Sowers, Mr. Leon D. Alley and Mr. Aubrey Murden for their kind assistance. I would also like to thank all the Biological Systems Engineering staff members for all the co-operation they have extended during my study.

I would also like to thank Dr. Chenming Zhang and Dr. David R. Bevan for their invaluable guidance in my summer project. I am grateful to all the graduate students for who have made me feel at home and also shared my joys during the course of my study.

Last but not the least, due to lack space; I would like to express my gratitude to every one who has helped to reach the present stage of my life. I am especially indebted to my parents, without whose constant encouragement, support and love, I would not have achieved anything.

Table of Contents

Abstract	ii
Acknowledgements	iv
Table of Contents	v
List of Tables	x
List of Figures	xiii
Chapter 1. Introduction.....	1
1.1. Introduction	1
1.2. Significance.....	4
1.3. Hypothesis.....	5
1.4. Over all objectives of the Research	5
Chapter 2. Literature Review	6
2.1 Introduction.....	6
2.2. Fruit quality.....	6
2.3. Quality assessment.....	7
2.4. Maturity.....	8
2.5. Maturity indices	9
2.5.1. Ethylene	9
2.5.2. Starch-iodine test	9
2.5.3. Fruit firmness.....	10
2.5.4. Soluble Solids	10
2.5.5. Color	11

2.5.6. Titratable acidity	12
2.6. Variation in apples	12
2.7. Currently available non-destructive methods to measure maturity	13
2.8. Apple flavor, aroma and fruit quality	14
2.9. Apple aroma as a basis for developing a non destructive technique	15
2.10. Electronic nose technology	18
2.11. Statistical Analysis.....	20
2.11.1. Introduction to Multivariate analysis	20
2.11.2. Principal Component Analysis	21
2.11.3. MANOVA.....	22
2.11.4. Discriminant Analysis.....	22
2.11.5. Variable Selection Procedure	25
2.11.6. Canonical Discriminant Analysis	25
2.11.7. Canonical Correlation Analysis	26
Chapter 3 - Materials and Methods	30
3.1. General.....	30
3.2. Measurement of maturity indices.....	30
3.3. Measurement of ethylene levels.....	32
3.4. Measurement aroma compounds	32
3.5. Identification of aroma compounds	33
3.6. Electronic nose Settings.....	33
3.7. Experimental Procedure.....	34
3.8. Data Description	35

3.9. Statistical Procedure.....	35
Chapter 4. Results and Discussion for Gala apples.....	53
4.1. Introduction.....	55
4.2. Gala- Statistical Analysis of Maturity Indices.....	55
4.2.1. Exploratory analysis.....	55
4.2.2. Inferential analysis.....	56
4.3. Statistical Analysis of Electronic Nose Sensor Response Data.....	57
4.3.1. Exploratory analysis.....	57
4.3.2. Inferential analysis.....	57
4.4. Electronic Nose System Software Analysis.....	58
4.5. Canonical Correlation Analysis.....	59
4.6. Analysis of pH and ethylene effect on the maturity classification.....	60
4.7. Harvest Date Effect.....	61
4.8. Analysis With Reduced Number of Sensors.....	61
4.9. Summary.....	62
Chapter 5. Results and Discussions for York Apples.....	79
5.1. Introduction.....	79
5.2. York- Statistical Analysis of Maturity Indices.....	79
5.2.1. Exploratory analysis.....	79
5.2.2. Inferential analysis.....	80
5.3. Statistical Analysis of Electronic Nose Sensor Response Data.....	81
5.3.1. Exploratory analysis.....	81

5.3.2. Inferential analysis	82
5.4. Electronic Nose System Software Analysis.....	82
5.5. Canonical Correlation Analysis	83
5.6. Analysis ethylene effect on the maturity classification	84
5.7. Harvest Date Effect.....	84
5.8. Analysis with Reduced Number of Sensors.....	84
5.9. Summary	85
Chapter 6 - Correlation of EN Sensor Responses of Gala and York	102
6.1. Introduction.....	102
6.2. CCA on 32 Sensors.....	102
6.3. CCA on significant sensors.....	102
Chapter 7- Chromatography Analysis.....	108
7.1. Introduction.....	108
7.2. Gala Aroma -Statistical Analysis.....	108
7.3. EN sensor response, maturity indices (including ethylene) and five aroma compounds data	111
7.4. York-statistical analysis.....	112
7.5. Comparison of Gal and York chromatograms.....	112
7.6. Summary	113
Chapter 8 Conclusions and Recommendations	132
8.1. Conclusions.....	132
8.2. Recommendations.....	133
References	134

Appendices	144
A. Conversion of mole% into milliliter	144
B. Gala and York raw data.....	146
C. Color analysis.....	151
D. Correlation matrices of maturity indices.....	153
Vita	156

LIST OF TABLES

- 2.1.** The values of maturity indices (Royal Gala apples) such as firmness, soluble solids and titratable acidity, measured in three different conditions. Data obtained from Cliff et al. (1998).....27
- 2.2.** The mean values of maturity indices (Fuji apples) such as firmness, soluble solids and titratable acidity, measured in three different years. Data obtained from Blankenship et al. (1997).....28
- 2.3.** The values of pH and corresponding TA values are reported. Data was taken from Keener et al. (1999). Two samples for each of the cultivars were reported28
- 4.1.** Mean values (with standard deviation) of maturity indices for Gala apples harvested on three dates64
- 4.2.** Classification table obtained after performing DA for Gala apples. Columns and rows represent the maturity categories (1, 2, and 3). The values represent the number of fruits placed in that category. The groups were categorized based on maturity indices.....65
- 4.3.** Mahalanobis distance to determine the similarity of groups of Gala apples with unknown maturity to a set of values from 3 groups of apples based on maturity indices. The left column indicates the groups and the right column indicates the distance65
- 4.4.** Classification table obtained after performing DA for Gala apples. Columns and rows represent the maturity categories (1, 2, and 3). The values represent the number of fruits placed in that category. The categorization was performed on EN sensor response data66

4.5.	Mahalanobis distance to determine the similarity of groups of Gala apples with unknown maturity to a set of values from 3 groups of apples based on maturity indices. The left column indicates the groups and the right column indicates the distance between them.....	66
4.6.	Classification table obtained from the EN software analysis for Gala apples. Columns and rows represent the maturity categories (1, 2, and 3). The values represent the number of fruits placed in that category. The categorization was performed on EN sensor response data.....	67
5.1.	Mean values (with standard deviation) of maturity indices for York apples harvested on three dates	87
5.2.	Classification table obtained after performing DA for York apples. Columns and rows represent the maturity categories (1, 2, and 3). The values represent the number of fruits placed in that category. The groups were categorized based on maturity indices.....	88
5.3.	Mahalanobis distance to determine the similarity of groups of York apples with unknown maturity to a set of values from 3 groups of apples based on maturity indices. The left column indicates the groups and the right column indicates the distance between them.....	88
5.4.	Classification table obtained after performing DA for York apples. Columns and rows represent the maturity categories (1, 2, and 3). The values represent the number of fruits placed in that category. The categorization was performed on EN sensor response data	89
5.5.	Mahalanobis distance to determine the similarity of groups of Gala apples with unknown maturity to a set of values from 3 groups of apples based on maturity indices. The left column indicates the groups and the right column indicates the distance between them.....	89
5.6.	Classification table obtained from the EN software analysis for York apples. Columns and rows represent the maturity categories (1, 2, and 3). The values represent the number of	

fruits placed in that category. The categorization was performed on EN sensor response data.....	90
7.1. Classification table obtained after performing DA for Gala apples. Columns and rows represent the maturity categories (1, 2, and 3). The values represent the number of fruits placed in that category. The groups were categorized based on aroma compounds emanation levels.....	114
7.2. Mahalanobis distance to determine the similarity of groups of Gala apples with unknown maturity to a set of values from 3 groups of apples based on maturity indices. The left column indicates the groups and the right column indicates the distance	114
7.3. Classification table obtained after performing DA for Gala apples. Columns and rows represent the maturity categories (1, 2, and 3). The values represent the number of fruits placed in that category. The categorization was performed on EN sensor response data ..	115
7.4. Mahalanobis distance to determine the similarity of groups of Gala apples with unknown maturity to a set of values from 3 groups of apples based on maturity indices. The left column indicates the groups and the right column indicates the distance between them...	115

LIST OF FIGURES

2.1. Gala-starch index Chart	29
3.1. The Instron universal testing machine with the 11 mm penetrometer probe plunger used to measure flesh firmness.....	38
3.2. Digital refractometer used to measure soluble solids concentration and the digital balance used to weigh fruit	39
3.3. Examples of immature Gala dipped in iodine solution to evaluate starch hydrolysis. Starch is present in the dark areas, whereas light areas (non-stained) indicate starch hydrolysis.....	40
3.4. The pH meter was used to measure the pH of the juice extracted from the apple after grinding the flesh. The reference electrode (in this combination electrode) was completely immersed.....	41
3.5. The tristimulus colorimeter used to measure the apple surface color.....	42
3.6. Fisher- Hamilton Gas Partitioner was used to measure ethylene	43
3.7. Gas Chromatograph used to measure volatile aroma compounds present in the headspace gas of the apple sample	44
3.8. The Electronic Nose System (Cyrano 320) with sensors exposed to the headspace gas of the apple in the bottle.....	45
3.9. Basic settings adjusted for each method, depending on the sample. Baseline purge time, Sample exposure time and purge time are three basic settings.....	46

3.10. The scrolling strip chart of the mature apple, when to exposed to the 32 sensors. This figure illustrates the real-time responses for all 32 sensors. The Y-axis is represented by sensor responses, while the X-axis, is represented by time in seconds	47
3.11. The basic method settings used for evaluating the Gala apple headspace gas. The canonical algorithm was used with no normalization. All sensors were switched on	48
3.12. The basic method settings used for evaluating the York headspace gas. The main difference between the Gala and York basic settings was the purge timings	49
3.13. Smell print of the immature Gala apple stored in the EN system memory	50
3.14. Smell print of the immature York apple stored in the EN system memory.....	51
3.15. From the sensors responses, the Euclidean distance for each observation to the centroid of the class are calculated and stored in the training set	52
4.1. Scatter plot of the first two principle components illustrating three clusters. The 1-60 represents the observations. Prin1 is represents starch and firmness variation. Prin2 represents primarily represents soluble solids.	68
4.2. Chi-Squared probability plot for multivariate normal. The X-axis (expected) represents Chi-Square Quantile. The Y-axis (dsq) indicates the mahalanobis D-square distance	69
4.3. Scatter plot of the first two principle components. The categories 1, 2 and 3 represent immature, mature and over mature groups respectively. Prin1 is represents starch and firmness. Prin2 represents soluble solids.....	70
4.4. Scatter plot of the first two canonical variables after the categorization of the three clusters based on Fig. 4.1 into three maturity categories. The categories 1, 2 and 3 represent immature, mature and over mature groups, respectively. Can1 is represents starch and firmness. Can2 represent primarily by firmness	71

4.5. Q-Q plot to examine graphically if the observations are from a multivariate normal distribution. Data were obtained with electronic nose. The X-axis (expected) represents Chi-Square Quantile. The Y-axis (dsq) indicates the mahalanobis D-square distance	72
4.6. Scatter plot of first two principle components after the Electronic nose sensor response data were categorized into three maturity groups (based on Fig 4.2). The categories 1, 2 and 3 represent immature, mature and over mature groups, respectively. Prin1 represents the response of all sensors. Prin2 represents most of the sensors responses	73
4.7. Scatter plot of the first two canonical variables after the EN sensor response data into the three maturity categories. The categories 1, 2 and 3 represent immature, mature and over mature groups respectively. Can1 represents the response of all sensors. Can2 represents most of the sensors responses	74
4.8. Scatter plot of the first two factors. PCA Plot generated by the EN system software analysis after the sensors response data were grouped into three maturity groups with 57 observations. The primary use of the PCA by the software is to detect outliers. No information about the axis is available in the software. The first two classes represent immature, next two classes represents mature and the last two classes over mature groups. 1-10 represents En sensor responses in a particular class.....	75
4.9. EN system software 3D Canonical projection plots. Three clusters are supposed to be obtained from the above CDA plot. No information about the axis is available in the software.....	76
4.10. CCA plot –1, illustrates a linear relationship between the first canonical function (Chr1) of the maturity indices data and the first canonical function (Sens1) of the sensor response data	77

4.11. CCA plot –2, illustrates almost poor relationship between the second canonical function (Chr2) of the maturity indices data and the second canonical function (Sens2) of the sensor response data	78
5.1. Scatter plot of the first two principle components illustrating three clusters. The 1-60 represents the observations. Prin1 is represents starch, firmness and soluble solids variation. Prin2 represents primarily represents soluble solids	91
5.2. Scatter plot of the first two principle components. The categories 1, 2 and 3 represent immature, mature and over mature groups respectively. Prin1 is represents starch, firmness and soluble solids variation. Prin2 represents primarily represents soluble solids.....	92
5.3. Chi-Squared probability plot for multivariate normal. The X-axis (expected) represents Chi-Square Quantile. The Y-axis (dsq) indicates the mahalanobis D-square distance	93
5.4. Scatter plot of the first two canonical variables after the categorization of the three clusters based on Fig. 5.1 into three maturity categories. The categories 1, 2 and 3 represent immature, mature and over mature groups, respectively. Can1 represents starch, firmness and soluble solids variation. Can2 represents firmness and starch.....	94
5.5. Scatter plot of first two principle components after the Electronic nose sensor response data were categorized into three maturity groups (based on Fig 5.2). The categories 1, 2 and 3 represent immature, mature and over mature groups, respectively. Prin1 represents the response of all sensors. Prin2 represents most of the sensors responses	95
5.6. Q-Q plot to examine graphically if the observations are from a multivariate normal distribution. Data were obtained with electronic nose. The X-axis (expected) represents Chi-Square Quantile. The Y-axis (dsq) indicates the mahalanobis D-square distance	96

- 5.7.** Scatter plot of the first two canonical variables. CDA plot after the EN sensor response data were grouped into the three maturity categories. The categories 1, 2 and 3 represent immature, mature and over mature groups respectively. Can1 represents the response of all sensors. Can2 represents most of the sensors responses.....97
- 5.8.** Scatter plot of the first two factors. PCA Plot generated by the EN system software analysis after the sensors response data were grouped into three maturity groups with 57 observations. The primary use of the PCA by the software is to detect outliers. No information about the axis is available in the software. The first two classes represent immature, next two classes represents mature and the last two classes over mature groups. 1-10 represents En sensor responses in a particular class.....98
- 5.9.** Scatter plot of the first two factors. CDA Plot generated by the EN system software analysis after the sensors response data were grouped into three maturity groups with 57 observations. No information about the axis is available in the software. The first two classes represent immature, next two classes represents mature and the last two classes over mature groups. 1-10 represents En sensor responses in a particular class.....99
- 5.10.** CCA plot –1, illustrates a linear relationship between the first canonical function (Chr1) of the maturity indices data and the first canonical function (Sens1) of the sensor response data. Sens1 represents all the sensors, while Chr1 represents starch, firmness and soluble solids100
- 5.11.** CCA plot –2, illustrates almost poor relationship between the second canonical function (Chr2) of the maturity indices data and the second canonical function (Sens2) of the sensor response data. Sen2 represents non of the sensors, while Chr2 represents soluble solids..101

- 6.1.** CCA plot –1, illustrates that there seems to be no correlation between the first canonical function (Sens1) and the first canonical function (Chr1). Sens1 represents responses all sensors of Gala, while Chr1 represents responses of all sensors of York.....104
- 6.2.** CCA plot –2, illustrates that there seems to be no correlation between the first canonical function (Sens2) and the first canonical function (Chr2). Sens2 is represents responses all sensors of Gala, while Chr2 represents responses of all sensors of York.....105
- 6.3.** CCA plot –1, illustrates that there is a correlation between the first canonical function (Sens1) and the first canonical function (Chr1). Sens1 represents responses of significant sensors 6, 14, 16, 17 and 23 of Gala, while Chr1 represents responses of significant sensors 2, 3, 8, 9, 16, 24 and 27 of York106
- 6.4.** CCA plot –2 illustrates a modest correlation between the first canonical function (Sens2) and the first canonical function (Chr2) of the significant sensors of Gala. Sens2 represents responses of significant sensors 5, 16, 17 and 23 of Gala, while Chr1 represents responses of significant sensors 2, 3, 8, 9, 16, 24 and 27 of York107
- 7.1.** PCA plot illustrates the three clusters. Prin1 represents all aroma compounds emanation levels. Prin2 is represented only few other significant aroma compounds emanation levels.116
- 7.2.** Typical chromatogram of volatiles compounds released from immature Gala apples, 10 days before optimum harvest maturity. No significant peaks were observed117
- 7.3.** Typical chromatogram of volatile compounds released from mature Gala apples, 10 days before optimum harvest maturity. In order to maintain consistency, five significant peaks were considered. The retention times of the five volatile compounds were 8.2, 9.7, 10.1, 11.3 and 12.6 minutes. They were identified using the MS library (GC/MS) as hexyl butyrate, hexyl 2-methylbutyrate, butyl hexanoate, hexyl hexanoate and apha-farnesene 118

7.4. Typical chromatogram of volatile compounds released from over mature Gala apples. The chromatogram represents the optimum physiological maturity. The above five significant (figure-7.2) volatile compounds were considered	119
7.5. PCA plot after the categorization of the three clusters based on figure into three aroma-based categories. The categories 1, 2 and 3 represent pre-climacteric (immature), climacteric (mature) and post climacteric (over mature) groups respectively	120
7.6. Chi- Squared probability plot for multivariate normal. The X-axis (expected) represents Chi-Square Quantile. The Y-axis (dsq) indicates the mahalanobis D-square distance	121
7.7. CDA plot after the categorization of the three clusters based on Fig. 7.5 into three aroma-based categories. The categories 1, 2 and 3 represent immature, mature and over mature groups respectively. Prin1 represents all five aroma compounds emanation levels. Prin2 is represented three significant aroma compounds emanation levels.....	122
7.8. Scatter plot of first two principle components after the Electronic nose sensor response data were categorized into three maturity groups (based on Fig 4.2). The categories 1, 2 and 3 represent immature, mature and over mature groups, respectively. Prin1 represents the response of all sensors. Prin2 represents most of the sensors responses	123
7.9. Scatter plot of the first two canonical variables after the EN sensor response data into the three maturity categories. The categories 1, 2 and 3 represent immature, mature and over mature groups respectively. Can1 represents the response of all sensors. Can2 represents most of the sensors responses	124
7.10. CCA plot –1, illustrates a linear correlation between the first canonical function (Sens1) and the first canonical function (Chr1). Sens1 represents sensors (6 highly correlated sensors) and Chr1 represents a1, a3 and a4	125

7.11. CCA plot –2, illustrates almost no strong correlation between the second canonical function (Chr2) of the chromatographic data and the second canonical function (Sens2) of the sensor response data. Sens2 represents all sensor responses while Chr2 represents soluble a3 and a5.....126

7.12. CCA plot –1, illustrates a poor correlation between the first canonical function (Chr1) of the chromatographic data and maturity indices and the first canonical function (Sens1) of the sensor response data. Chr1 represents starch, firmness and five aroma compounds. Sens1 represents only one sensor (sensor 14).....127

7.13. CCA plot –2, illustrates a linear correlation between the second canonical function (Chr2) of the chromatographic data and maturity indices and the second canonical function (Sens2) of the sensor response data. Chr2 represents all maturity indices, while Sens2 represents all significant sensors.....128

7.14. Exploratory PCA plot for York chromatographic analysis. Only two discernable groups could be obtained.129

7.15. Typical chromatogram of volatile compounds released from over mature Gala apples130

7.16. Typical chromatogram of volatile compounds released from over mature York apples....131

Chapter 1

1.1. Introduction

One of the key features of globalization is that products produced in one part of the world can be effectively marketed in any other part of the world. Two factors, namely quality and supply, play a pivotal role in the marketability of any product in the global market. These two important issues affect the US apple industry: 1) Steady increase in the worldwide apple production and China accounting for loin's share of the increase. It is predicted that by 2005 there will be approximately a 30 % increase in the world apple production, from 53,165 thousand metric tons in 1997 to 68,319 thousand metric tons in 2005 (Rourke, 1998). 2) Steady increase in the fresh supplies of better quality from major southern hemisphere countries like Argentina and Chile (Warner, 1991).

Most of the fresh produce is stored in refrigerated units of modern supermarkets. Few large multinational corporations such as Wal-Mart, Tesco etc, own a large number of such firms and thus dominate the food industry. These firms often acknowledge both the professional opinion and the public demand, in determining price and quality of the products. Consequently these firms demand assurances from packinghouses and growers to supply better quality apples (Rourke, 1998).

Apples are an excellent source of dietary fiber and are free from fat, cholesterol and sodium. They are rich in phytonutrient antioxidants and in mineral boron. Recent research on health qualities of apples have linked it with a range of health benefits, such as reduction in risk for heart disease, cancer, lung problems, thrombotic stroke, etc

(Miller, 2001). Consumers are health conscious and are motivated by health messages. In recent years the apple industry has been emphasizing the health qualities of apples. Such emphasis may lead to an increase in demand for apples. Consequently consumers are willing to pay a premium price for good quality apples that are attractive in appearance, crisp, full flavored in taste and available through out the year.

The total U.S. utilized apple production in 2000 was 5.1 million metric tons with a farm-gate value of 1.3 billion dollars (U.S. Dept. of Commerce, 2000). In the same year the fresh fruit consumption was 59% and the processed fruit was 39%, while the remaining 2% was not marketed (U.S. Apple Association, 2002). Currently, random samples are used to determine the quality of the apple population. Consequently apple lots having high percentage of defects are discarded and lots having high percentage of inferior quality are sent for processing. Thus even the good quality apples that could have been sold for premium price, are discarded or processed, incurring a substantial loss to the seller (Keener et al., 1999).

The above-stated problems arise due to the fact that sometimes harvested fruits are a mixture of mature, immature and over mature fruits, and that the quality of an apple fruit depends primarily on its level of maturity at the time of harvest. Even though the external appearance of an immature fruit may look perfect to harvest, store and sell, due to their pre-climacteric physiological condition, these apples do not ripen normally, and thus their taste is strongly impaired due to lack of full-flavor compounds (Brackmann and Strief, 1994). On the other hand over mature fruits have shorter storage life, soften rapidly, develop storage disorders such as off flavor, lack of firm texture, and are

unattractive in appearance. Some apples also have internal defects like water core and internal browning.

Generally, consumers primarily decide an apple's quality based on its external factors such as, size, color, appearance and sometimes fruit response to thumb pressure. Some of them even purchase an apple based on its reputation, the brand name and the seller. But only by consuming the apple, can internal quality be evaluated. Thus, marketing immature or over mature apples may negatively effect the reputation of the product. Therefore, growers should take care to harvest fruit at optimum maturity. Until now, there have been no non-destructive techniques to determine the maturity of an apple. Currently apple maturity of an apple is detected by traditional techniques such as °Brix (soluble solids), starch conversion index, firmness or pressure testing and apple surface ground color. All the above-mentioned testing methods except color are destructive in nature, and so only sample tests are conducted to assess the quality of the whole population. Since all the above-mentioned indices vary considerably from apple to apple, they cannot be used for pre- and post harvest quality assessment of individual apples.

Thus there is an urgent need to develop a non-destructive technique to determine the quality of apples and to classify them into pre-climacteric, climacteric or post climacteric fruits. Such a technology could also be used to distinguish the inferior quality fruits used for juice worth \$40/t and premium quality fruits worth \$200/t (U.S. Dept. of Commerce, 1994) from the rejected apple lots (Keener et al., 1999). Predicting optimal

harvest dates for apples may be improved by including additional maturity indices that are non-destructive in nature.

Electronic nose system is a sensor-based technology, which considers the total headspace volatiles and creates a unique smell print. Unlike gas chromatography, electronic nose does not resolve the sample volatiles into its individual components, but responds to the whole set of volatiles in a unique digital pattern. These patterns are signature of the particular set of aromatic compounds. For each process or application of interest, a database of such digitized patterns is created, called the training set, then any unknown sample with its unique volatile, digital pattern is compared with the existing training set database.

Preliminary studies from our research group demonstrated that electronic nose (Cyranose nose 320) has potential for evaluating apple maturity (Pathange et al., 2002), and it could distinguish between mature and over mature York apples. The present research was conducted to determine the effectiveness of the electronic nose system on Gala and York apples.

1.2. Significance

Apple growers and packinghouses are interested in methods which can non-destructively evaluate the quality of apples. Since harvested fruit is sometimes a mixture of mature, immature and over mature fruits, thereby posing a great problem in deciding its end use and storage time. It is expected that the technique developed from the present project would be effectively used to classify the harvested fruit into mature and

immature, and to predict the optimum harvest dates. It can also be used as a maturity index for determining maturity level of individual apples, before and after harvest.

1.3. Hypothesis

Electronic nose system can be used effectively to non-destructively classify apples, based on the maturity indices and volatile compounds, as pre-climacteric (immature), climacteric (mature) or post climacteric (over mature).

1.4. Overall objectives

- 1) Use conventional maturity indices to classify apples as immature, mature and over-mature.
- 2) Use aroma compounds emanation levels, to classify apples as pre-climacteric, climacteric and post climacteric.
- 3) To determine if electronic nose can properly classify apples into these three groups.
- 4) To determine if electronic nose sensor responses can be used as a non-destructive maturity indicator.

Chapter 2- Literature Review

2.1 Introduction

The quality of an apple depends on its maturity at harvest. Several maturity indices are used to assess the maturity of apples, but most of these indices are destructive, and so only a sample population is used to assess the maturity of the total population. It is already demonstrated in literature that volatile compounds released from apples may be used as a potential maturity index.

Recent technological developments in chemosensory technology have enabled researchers to develop a class of instruments known as electronic noses, which generate a unique digital image for each complex vapor mixture. Currently, Electronic noses technology is being investigated to study its applicability, to a wide variety of problems including the evaluation of apple maturity.

2.2. Fruit quality

Apple fruit quality can be defined in many ways. Most of the definitions are related to the characteristic features that develop during the post harvest life of the fruit (Knee and Smith, 1989). Some of the quality attributes determining apple quality are color, flavor, aroma, size and texture (Vangdal, 1985). If fruit is destined for a long-term storage (such as cold or controlled atmosphere storage), then fruit quality can be equated as physiological maturity or harvest maturity (Kingston, 1991). If fruit is destined for immediate consumption, the fruit quality primarily depends on the consumer's perceptions and preferences and is usually referred as "mature" or "ripe." According to

Watada et al. (1984) the edible quality is referred as commercial maturity and the physiological maturity is referred as horticultural maturity.

2.3. Quality assessment

According to Kingston (1991) suitable maturity indices and their desirable values, for both horticultural and commercial maturity were established, after several years (or seasons) of fruit quality evaluation programs. In case of the commercial maturity, the main quality indices were firmness, soluble solids and titratable acidity; for horticultural maturity, the harvested fruits should have higher starch and titratable acidity levels than commercial mature apples, and should have lower firmness and ethylene evolution levels than commercially mature apples. Some of the values of the maturity indices for different cultivars are listed in Table-2.1, Table 2.2 and Table 2.3.

Quality assessment is a complex problem at both commercial and horticultural levels. In the supermarket, consumers primarily evaluate the apple quality based on external factors such as, size, color and firmness and their experience. But only after consumption of the fruit, final judgment about the fruit quality can be made (Kingston, 1991). Thus the burden of quality assurance is on the provider of the fruit, because consumers rely on fruit reputation and brand name. Since all the techniques that determine quality parameters (firmness, starch, etc) are destructive, even horticultural maturity cannot be determined for each fruit; rather random sampling of the population is done to determine the horticultural maturity or quality. There are also problems relating the final (consumer acceptable) quality of the fruit to the harvest quality of the fruit

(Knee and Smith, 1989). To ascertain apple quality before and after harvest, rapid non-destructive maturity evaluation techniques are needed.

2.4. Maturity

Fruits capable of ripening after being detached from the tree are referred as mature (physiological) fruits. Fruits that have not yet reached physiological maturity are referred as immature. Fruits, whose quality indices such as full-flavor, aroma, texture, and juiciness, acidity are acceptable for immediate consumption, are referred as ripe (over mature) or commercially mature. In climacteric fruits such as apples, there is a marked sudden increase in physiological processes namely the respiration rate (carbon dioxide evolution) and the ethylene evolution rate, as the fruit matures. Concomitantly there are physical and chemical changes that follows the physiological changes such as hydrolyzation of starch to sugars, drop in chlorophyll levels, changes in skin and flesh color, drop in pH levels, changes in seed color, softening of cementing material between the cells and enhanced emanation of aroma compounds.

Since the age of the fruit (maturity level) has a critical effect on the quality (Brookfield et al., 1993), the apples should be harvested when the quality criteria are best satisfied. Since the maturity determines the rate of quality loss (flavor and firmness, and green color) (Tugwell, 1998), harvesting fruit at an optimal physiological condition (harvest maturity) ensures the fruit quality at a later stage by enhancing a number of quality characteristics, such as an extended shelf life, a slower rate of decline in firmness, acidity and color (Smith, 1984). Thus, maturity is the key for good apple quality (Tugwell, 1998).

2.5. Maturity indices

2.5.1. Ethylene

Ethylene is a naturally occurring plant hormone, which aids in plant growth and development. In climacteric fruits such as apples, the accelerated production of ethylene gas is believed to be the primary stimulating agent for the onset of ripening. The increase in ethylene production rate can be so dramatic that there can be a 100-fold increase in just two days. Thus maturity can be defined as that stage when the fruit has entered the crucial climacteric stage. Depending on the ethylene production levels, the immature stage of the fruit is referred as pre-climacteric and the ripe stage of the fruit is referred as post-climacteric (Kupferman, 1986). Depending on the levels of ethylene and carbon dioxide released from an apple, the climacteric stage can be determined (Song and Bangerth, 1996). Thus ethylene production levels can be used as a maturity index to classify the apples into immature and mature categories.

2.5.2. Starch-iodine test

Starch accumulates in apple flesh during fruit growth and is hydrolyzed into sugars, as the fruit matures (Smith et al., 1979). Starch hydrolysis is evaluated with a starch iodine test. Fruits are cut horizontally and the cut surface is dipped in a solution of iodine and potassium iodide, and the pattern of blue-black stain is compared with charts for that cultivar (Fig. 2.1) and rated on a scale of 1-9 (Smith et al., 1979). Factors other than maturity may influence starch index. Climatic conditions, cultivars (Phillips and Poapst, 1952) and seasons (Poapst et al., 1959) all influence starch pattern. According to

Wills et al. (1980) the organoleptic changes in the fruit do not correlate very well with the changes in starch index value, so using starch index values alone should not be used to determine apple maturity (Blanpied, 1974; Smock, 1948).

2.5.3. Fruit firmness

According to Kingston (1991), as fruit ripen the cementing material between cells, the middle lamella, starts to dissolve; accompanied by changes in the cell sap result in the softening of the fruit. A penetrometer can be used to measure the softening (firmness) of the fruit by recording the resistance of the peeled fresh fruit by inserting a plunger of a known diameter (generally, 11mm). According to Vangdal (1982) fruit firmness is highly correlated to overall fruit quality. Several factors affect firmness changes apart from maturity. According to Blanpied et al. (1978) fruit firmness values decrease as fruit size increases. They also reported that firmness can be affected by nitrogen fertilization, fruit position in canopy, water core, and temperature of the fruits. They also reported that fruit firmness values vary considerably between seasons and orchards. In some cultivars there was no real difference between commercial immature and commercial mature fruits because rate of change of firmness was very slow in those cultivars. Thus, Kingston (1991) concluded that firmness values by themselves were inconsistent and was an unreliable maturity index.

2.5.4. Soluble Solids

As apple fruits mature, starch is hydrolyzed into sugars. Instead of measuring sugar content by an established cumbersome chemical analysis, a much easier

measurement of soluble solids is generally employed. A refractometer was used to measure the soluble solids concentration of juice extracted from fruit. Generally the soluble solids concentration increase as the fruit matures and so can be used as an index for determining maturity of the apple fruit (Kingston, 1991). However, soluble solids are influenced by many factors other than maturity. Soluble solids vary with position of the fruit in the canopy (Shaw and Rowe, 1982), individual orchards they are grown in (Reid et al., 1982), fungicide application (Rouchaud et al., 1983) and particular season they are grown (Ingle and D'Souza, 1989). According to Harman and Watkins (1981) soluble solids also take in account total organic acids, whose variation pattern do not always coincide with that of the sugars (soluble solids) and thus soluble solids by itself should not be used as the sole guide for evaluating fruit maturity.

2.5.5. Color

Apples start to lose green skin color as the fruit starts to ripen, due to lower levels of chlorophyll production, and thus other pigments start to appear on the skin (Fiddler, 1973). Concentration of chlorophyll or other pigments can be measured analytically, but a much simpler method using color cards is employed. These cards are similar to a starch index chart, were developed for each cultivar, and represent a distinct stage of color (pigment) development (Olsen et al., 1986). This method is a subjective assessment of color and a much better objective assessment of the fruit color can be performed using a tristimulus colorimeter (Kingston, 1991).

Ground color is used as the maturity index, rather than red blush color that develops at later stages of maturity. This is due to the fact that development of red blush

color depends mainly on the environmental factors such as, temperature and the amount of sun exposure before harvest (Olsen and Martin, 1980). But even ground color is influenced by all the above-mentioned factors that affect red blush color formation. Thus it alone cannot be used as maturity indicator.

2.5.6. Titratable acidity

As fruit gradually ripen the total organic acid concentration gradually declines (Mann and Singh, 1986). Generally the titratable acidity is determined by neutralizing the acid in the juice with sodium hydroxide solution. Titratable acidity is an important parameter that affects apple flavor (a combination of sugars, aroma compounds, astringent compounds and acids). Apples having titratable acidity above or below a certain level are unacceptable to the consumers.

Gradual changes in titratable acidity during ripening are not exhibited by all the apple cultivars. Factors affecting titratable acidity also include nitrogen fertilization (Hikasa et al., 1986), the season and orchards in which they are grown (Knee and Smith, 1989) and position of the fruit in the canopy (Robinson et al., 1983). So titratable acidity by itself is not a reliable maturity index and should be used in conjunction with other maturity indices mentioned above (Kingston, 1991).

2.6. Variation in apples

Harvested fruit sometimes consist of a mixture of immature, mature and over mature fruits, thereby posing a great problem in deciding their end use. This is due to the presence of within-tree variability in apple fruits. Though the factors that effect such

variation are poorly understood, a few of these factors are listed below. Tree shade affects the size of the fruit, development of red color and concentration of soluble solids (Jackson, 1980). On the other hand, levels of nitrogen application decrease the concentration of soluble solids, titratable acidity and fruit firmness but increase the green coloration in the fruit (Hikasa et al., 1986; Olsen et al., 1986). Fruiting lateral orientation (Tustin et al., 1988) and the age of the wood (Volz et al., 1994) can affect the size of the fruit. Firmness and size of the fruit were greatly influenced by spur vigor (Volz et al., 1995) and differential rate of flowers to reach anthesis (Kingston, 1991) both cause variation in maturity. Apart from within- tree variation, other factors that influence the individual fruit quality indices are the growing season, which greatly influences the physiological maturity (Volz et al., 1995) and the particular orchard where the fruit was grown (Knee and Smith, 1989).

2.7. Currently available non-destructive methods to measure maturity

Researchers have attempted to develop reliable techniques to measure apple maturity, non-destructively. Abbot et al. (1968) concluded that acoustic resonance test could be used to measure the textural (firmness) suitability for harvest, storage and subsequent consumption. Shmulevich et al. (1996) had listed some of the promising non-destructive techniques, Muramatsu et al. (1999) concluded that his proposed usage of remote sensing technology to assess apple fruit textural changes with a laser doppler vibrometer was more reliable than all the previously non-destructive techniques reported by Shmulevich et al. (1996). On the other hand, Peirs et al. (2000) concluded that optimal harvest date could be predicted by non-destructively measuring internal quality

parameters such as soluble solids and acidity with VIS/NIR –spectroscopy. But since all the above-mentioned non-destructive techniques measure only one or at most two, physical or chemical parameters, they cannot be used as a comprehensive maturity index.

2.8. Apple flavor, aroma and fruit quality

Apple flavor can be defined as a complex composition of aroma, taste and texture of a fruit. Dimick and Hoskin (1983) stated that research to understand the complex nature of apple flavor had started as early as the beginning of nineteenth century. During the ripening process, the sensory quality attributes such as flavor and texture result from a number of pre- and post-harvest factors (Dirnick et al., 1989). According to Paillard (1982) there are two factors, external and the internal that influence flavor formation in apple fruits. The external factors are pre-harvest factors (soil/hydroponic culture, fertilization and climate/irrigation), harvest date (maturity) and post- harvest factors (storage time, storage conditions such as temperature, humidity and gas composition). The internal factors are genetic (cultivar type) and metabolic regulation (ethylene respiration). It has been mentioned in the literature that during maturation, harvest and storage, a number of volatile compounds are released, which contribute to the flavor and aroma of the apple. Since aroma is a primary factor affecting the flavor, aroma evaluation can be used as a criterion to determine the flavor quality (Dirnick et al., 1989).

According to Dimick and Hoskin (1983) the presence of ester compounds with molecular weight between 100 and 130 can be considered as one of the primary requirement for aroma in an apple fruit. According to White (1950), the eight alcoholic group compounds constituted 92% of the total volatile compounds, four carbonyl group

compounds constituted 6%, while the rest were constituted by numerous ester group compounds. A complex mixture of the above-mentioned three groups volatile compounds, results in formation of characteristic apple aroma (Young et al., 1996). Flath et al., (1967) demonstrated through organoleptic evaluation that volatile compounds like hexanol, trans-2-hexenal, and ethyl -2- methyl butyrate were responsible for aroma in Delicious apple essences. William et al., (1977) identified 4-methoxyallyl benzene as the compound that contributed to the aniseed (Spice-like) aroma, in apple the cultivars they tested.

From their gas chromatography-olfactometry analysis, Young et al. (1996) suggested that there were four major aroma compounds, namely 2-methylbutyl acetate, butyl acetate, hexyl acetate and butanol that were associated with the flavor of Royal Gala. They demonstrated graphically how the above mentioned four flavor compounds were related to the nine sensory attributes including overall aroma, red apple aroma, sweet aroma, acid aroma, overall flavor, red apple flavor, sweet flavor, acid flavor and characteristic apple flavor. Out of the four compounds, 2-methyl butyl acetate had the most important effect on the sensory attributes, effecting eight out of nine attributes, followed by butanol.

2.9. Apple aroma as a basis for developing a non destructive technique

Brackmann et al., (1994) reported that generally apple aroma production levels concomitantly increase with climacteric respiration and reach the maximum levels 2-3 weeks later. Song and Bangerth's (1996) experimental results indicate a similar evolution pattern between ethylene, respiration and total aroma production levels in Golden

Delicious. Although there are more than 250 compounds that contribute to the full flavor formation in apple fruits (Dimick and Hoskin, 1983), it was concluded that there were only few compounds (impact compounds) that had decisive impact on the sensory quality of the apples (Cunningham et al., 1986; Song and Bangerth, 1996). Song and Bangerth's (1996) results indicated that the production pattern of four volatile compounds (impact compounds) butyl acetate, hexylacetate, 2-methylbutylacetate and ethyl-2-methylbutanoate, was similar to that of ethylene evolution, respiration rate and total aroma production patterns. Dirnick et al., (1989) could predict the optimum harvest time for storage apples, by employing linear regression between the logarithm of butyl acetate concentration and the picking date.

According to Brown et al., (1965), maximum production of certain volatile compounds from apples was concomitant with the respiratory climacteric. On the other hand, some compounds did not emanate until much later, during the ripening process. They demonstrated that aroma compounds varied with cultivar and with age of the apple. They concluded from the chromatograms, which varied in aroma production, that aroma compounds could be used to develop critical criteria to determine apple maturity and quality. Thus apples of similar physiological condition would have similar chromatograms that could be used to determine fruit age.

According to Brackmann and Streif (1994) ethylene and aroma production depends on the cultivar since they are genetic characters; aroma production of Gravenstein cultivar was 33 times more than that of Granny Smith. Thus the aroma production levels could be used to identify apple cultivars.

Consequently, maturity evaluation based on aroma can be used to develop a consistent and reproducible nondestructive technique to evaluate apple quality from harvest to consumer. Though aromatic compounds could be used as a potential maturity index, this information has not been used significantly to develop a maturity indicator (Young et al., 1999), because both trained and sensory panels and gas chromatography techniques are time consuming, complicated and expensive.

Electronic nose (EN) technology, which simulates the human nose, can overcome some of the difficulties associated with classical flavor measurement (Young et al., 1999). Tin oxide based sensors were used by Simon et al., (1996) to monitor blueberry flavor. Benady et al., (1992) related the data derived from electronic senses to various ripeness indices such as slip pressure, and classical volatile measurements in melons. Data from sensory panels were correlated to the electronic nose data that registered gases from the degradation reactions in tomatoes (Simon et al., 1996).

Young et al. (1999) demonstrated that electronic nose technology using metal oxide sensors could be used as a potential maturity indicator to predict the harvest date for Royal Gala apples. However, the sensors used were sensitive to moisture and were also associated with sensor drift, and they performed their experiments on apple tissue (destructive methods). The present research will be conducted primarily with newer sensing technologies, which uses conducting polymers that less sensitive to moisture variation. In addition to predicting harvest date, the present analysis could be performed to evaluate the maturity of apples by grouping them according to maturity.

2.10. Electronic nose technology

To address the ongoing revolution in analytical chemistry, wherein there is continuous size reduction in analytical instruments and their labor-intensive procedures characterized by a gradual transformation of the complex information science into a decision-making science, a class of instruments known as electronic noses has been developed. They are so named because they are designed to mimic human olfactory processes (Cyrano Sciences, Inc., 2001). Though Electronic nose systems were developed to imitate the human nose, the mechanism and know-how of actual functioning of the human nose has yet to be discovered. According Van Deventer et al. (2001), currently it has not been possible to design and develop equipment that can act as duplicate the human nose. Though GC/MS and GC-olfactory techniques were used in the past, aroma of a particular sample is a complex mixture compounds, and no amount of statistical calculations or multiple sniff ports could yield the exact smell print of the sample (Van Deventer et al., 2001).

Electronic nose systems consist of an array of chemical sensors which respond to the volatile flavors from a sample (Bartlett et al., 1997) in a unique pattern (Haugen and Kvaal, 1998). Though electronic nose is not a substitute for human sensory panels, which are most reliable and sensitive in measuring aroma, it can be used as a rapid, automated and objective alternative to detect measure and monitor aroma.

The electronic nose systems that are commercially available are based on conducting polymers, quartz microbalance or metal oxide (Van Deventer et al., 2001). Van Deventer et al., (2001) concluded that conducting polymer sensors were best at

discriminating the retained solvents in printed food packaging. The present study uses a conducting polymer-based system from Cyrano Sciences (Cyrano 320) with an array of 32 non-specific sensors. Each conducting polymer sensor is composed of a pair of electrical contacts, connected by a composite film of non-conducting polymer and conductive carbon particles. Upon exposure to volatile compounds, the composite film swells, breaking the original conductive pathways and alters the resistance between the electrical contacts. This variation in the resistance of the exposed sensor is registered as the sensor output in the electronic nose system. Each of the 32 sensors in the Cyranose 320 are made with a unique polymeric material and, when exposed to a particular vapor mixture each sensor reacts in a different but reproducible manner producing a "smellprint" (combination of resistances of all sensor) for each volatile mixture (Cyrano Sciences, Inc., 2001). A database of smellprints or the digital images of a chemical vapor mixture is created by training the electronic nose system. Then using a prediction algorithm, such as a multivariate technique (PCA, CDA, etc), a model can be developed. When a new unknown vapor mixture is to be identified, the EN system digitizes the vapor mixture and compares this digital image with the previously established database (model) of smellprints in its memory. The unique feature of the EN system is that its response takes into account all the characteristic features (chemical and physical properties) of a sample, but does not provide information about the composition of the complex mixture. Thus this system can be used when the decision about a chemical vapor of a sample is more important than its contents, such as a spoiled vs. non-spoiled food sample, age of a fruit, type of cheese etc. (Cyrano Sciences, Inc., 2001).

Brackmann et al. (1994) reported that CO₂ and ethylene levels were closely related to that of ethylene and aroma compounds in apple. Since electronic nose system accounts for the whole headspace gas, the various aroma compounds and their concentrations in addition to ethylene and carbon dioxide which characterize physiological maturity can be used as a maturity index because these three factors are related to apple maturity. According to Young et al. (1999) the EN analysis was approximately 40 times more sensitive than the headspace/gas chromatography. Additional advantages of EN (Cyarnose 320) include portability, and economics compared to other commercially available systems.

2.11. Statistical Analysis

2.11.1. Introduction to Multivariate analysis

Multivariate techniques are usually employed to summarize large amounts of data, with many independent variables that may be related, and with few response variables. These techniques are often used to understand the relationship among 1) the independent variables, 2) the experimental units, and 3) both independent variables and the experimental units. Some of the multivariate techniques that are most commonly used as exploratory analysis (trying to explore the relationships among the variable) are, Principal Component Analysis (PCA), Factor Analysis (FA) and Canonical Correlation Analysis (CCA). Commonly used techniques for comparing group means are Multivariate Analysis of Variance (MANOVA) and Canonical Variate Analysis (CVA). In order to predict group membership, Discriminant Analysis (DA), Canonical

Discriminant Analysis (CDA) and cluster analysis (CA) are commonly used (Johnson, 1998).

2.11.2. Principal Component Analysis

Principal Component Analysis (PCA) is an exploratory multivariate technique. PCA involves a mathematical operation that determines the transformation of a set of predictable variables (possibly correlated) into a (smaller) set of new uncorrelated variables called principal components. Some of the objectives of PCA are: 1) To discover the true dimensions of the data set and thus reduce the dimensionality of the data set; 2) if possible, to interpret and identify meaningful underlying principles and variables respectively; 3) To screen the data for any outliers or clusters present in the data; and 4) To reduce the number of variables and to detect structure in the relationships between variables.

Since the uncorrelated principal components contain almost the entire information that is contained in the original variables, their values (principal component scores) can be used as an input for Discriminant analysis. This becomes a necessity when the sample size of the experimental units is smaller than the size of original variables and in such cases inversion of variance–covariance matrix cannot occur, thus discriminant analysis fail to work. The unique feature of these principal components is that the first accounts for the most variability in the data, the second component accounts for the most of the remaining variability in the data, and each succeeding component takes accounts for less variability in the data. The appropriate number of principal components (i.e., true dimensionality of the data) is determined. The variance extracted by each factor is called

the eigenvalue. The most widely used method for determining how many factors to retain is to retain only factors with eigenvalues greater than 1.0.

2.11.3. MANOVA

One important condition that must be verified before using multivariate methods is that the experimental units should be independent. That is the values of variables measured on one experimental unit should not have any influence on the values of the variables measured on any other experimental unit. It is also assumed that the multivariate data, on which above mentioned multivariate methods are employed, is considered to be a random sample from a multivariate normal distribution.

As in ANOVA, the MANOVA is used to compare the means of many populations, but it considers all variables simultaneously. This method can be used as an exploratory technique because if a researcher concludes that there was a significant difference between the categories, then further analysis with CDA or DA could be used to ascertain the effectiveness of the classification. Some of the most popular MANOVA testing procedures include Roy's test, Lawley and Hotelling's test, Pillai's test, Wilks' likelihood ratio test and Roy's second test.

2.11.4. Discriminant Analysis

Discriminant analysis is used primarily to answer three basic questions: 1) is the number of sensors and the sensor data obtained from the training set useful for building a model to classify the apples into its maturity level or stage? 2) Can the model classify correctly the unknown apples of varying maturity levels? 3) If not, what is the

miscalculated percentage? Discriminant analysis also known as classification analysis is a multivariate method for classifying observations into appropriate categories (apples into appropriate maturity levels) (Johnson, 1998).

The concept of discriminant analysis is analogous to regression analysis, as the goal of the latter being to predict the value of the dependent variable, while that of the former being to predict the category of the individual observation (Johnson, 1998). The main difference is that multivariate (discriminant analysis) approach is used when the variables are not independent. This condition violates the assumption of regression (Marini, 2003).

According to Johnson (1998) there are four nearly equivalent ways to develop a discriminant rule to classify observations into categories (Likelihood Rule, Linear Discriminant Function Rule, Mahalanobis Distance Rule and Posterior Probability Rule). There are three different methods which be can used to verify or estimate the probability of the correct classification of the observations and are described in detail below (Johnson, 1998).

1) *Resubstitution Method*

The resubstitution method employs a discriminant rule to the same data which were used to develop the rule and check how many observations were correctly classified by the rule into the correct categories. This method presumes that if a rule cannot classify properly on the original data used to build the rule, then there is a poor chance of it doing well with a new data set. The major drawback with this method is its overestimation of

the probabilities, when it classifies correctly. In SAS this method can be invoked using the DATA = option (lists).

2) *Holdout Method*

This method uses a holdout set or a test data set, where we know which observation belongs to which particular category, and the hold out data set is not used to develop the discriminant rule. The major drawback for this method is that one has to sacrifice the hold out data in order to build the discriminant rule, thus not being able to develop the best possible discriminant rule. In SAS this method can be invoked using the DATA = option (testdata).

3) *Cross-Validation Method*

Lachenbruch (1968) first proposed the cross-validation method, also known as jackknifing. This is the preferred method when compared to the above two discriminant rules. The first observation vector is holdout and the remaining data is used to construct the discriminant rule, then the rule is used to classify the first observation, and then it checks whether the observation is correctly classified into the particular category. In the next step, the second observation vector is removed, but the first observation is replaced back into the original data, and then the discriminant rule is constructed. The rule thus developed is used to classify the second observation and thus check whether the observation is classified correctly. Thus the same process is continued for the entire data set and also noting down the category it is being classified. It is claimed that this method

is almost unbiased. In SAS this method can be invoked using the DATA = option (crosslists).

2.11.5. Variable Selection Procedure

Since the number of variables involved in this study is high (32), a variable selection procedure is used to reduce the number of variables, which are really necessary for effective discrimination of the data. The three types of variable selection procedures are Forward Selection Procedure, Backward Elimination Procedure and Stepwise Selection Procedure. Johnson (1998) recommends the Stepwise Selection Procedure when the number of variables exceeds 15.

2.11.6. Canonical Discriminant Analysis

Canonical Discriminant Analysis (CDA) is a dimension reduction technique that creates new canonical variables by taking special linear combinations of the original response variables. The canonical variables of the CDA, in some sense, are similar to principal components of the PCA. The principal advantage of CDA is its ability to allow the researcher to visualize the observations, which are classified into the different categories, in 2-D or 3-D space. Another advantage of CDA is that the output from a PCA can be used as an input for the CDA, thus the data visualized. If possible, one can attempt to interpret the canonical variables (Johnson, 1998).

2.11.7. Canonical Correlation Analysis

CCA is generally performed when there is a need to compare groups of variables. It helps in reducing the dimensionality of the data. CCA can be used to summarize the underlying relationship between groups of variables by creating new variables from the existing groups of variables. These new variables are called canonical functions. While performing the CCA, the optimum number of canonical functions, can be known, only after performing a preliminary CCA. Generally the option NCAN=2 is used to limit the number of canonical functions generated to two. Interpretation of canonical functions is generally considered to be difficult (Johnson, 1998).

Table 2.1. The values of maturity indices (Royal Gala apples) such as firmness, soluble solids and titratable acidity, measured in three different conditions. Data obtained from Cliff et al. (1998).

Stage of the apple	Harvest date	Flesh firmness (N)	Soluble solids (°Brix)	Titratable acidity
	09/5/1995	76.1	12.3	0.526
Harvest	09/19/1995	66.5	12.3	0.447
	09/5/1995	56.4	13.4	0.372
Post storage in air*	09/19/1995	53.3	13.7	0.357
	09/5/1995	61.2	13.1	0.452
5.0% O₂	09/19/1995	58.9	13.0	0.39

* Post-storage- CA storage with 1.5% CO₂

Table 2.2. The mean values of maturity indices (Fuji apples) such as firmness, soluble solids and titratable acidity, measured in three different years. Data obtained from Blankenship et al. (1997).

Year	Firmness (N)	Soluble solids (%)	Starch index (1-9)
1991	81.8	16.2	8.4
1992	70.7	17	6.7
1993	68.5	13.1	5.8

Table 2.3. The values of pH and corresponding TA values are reported. Data was taken from Keener et al. (1999). Two samples for each of the cultivars were reported.

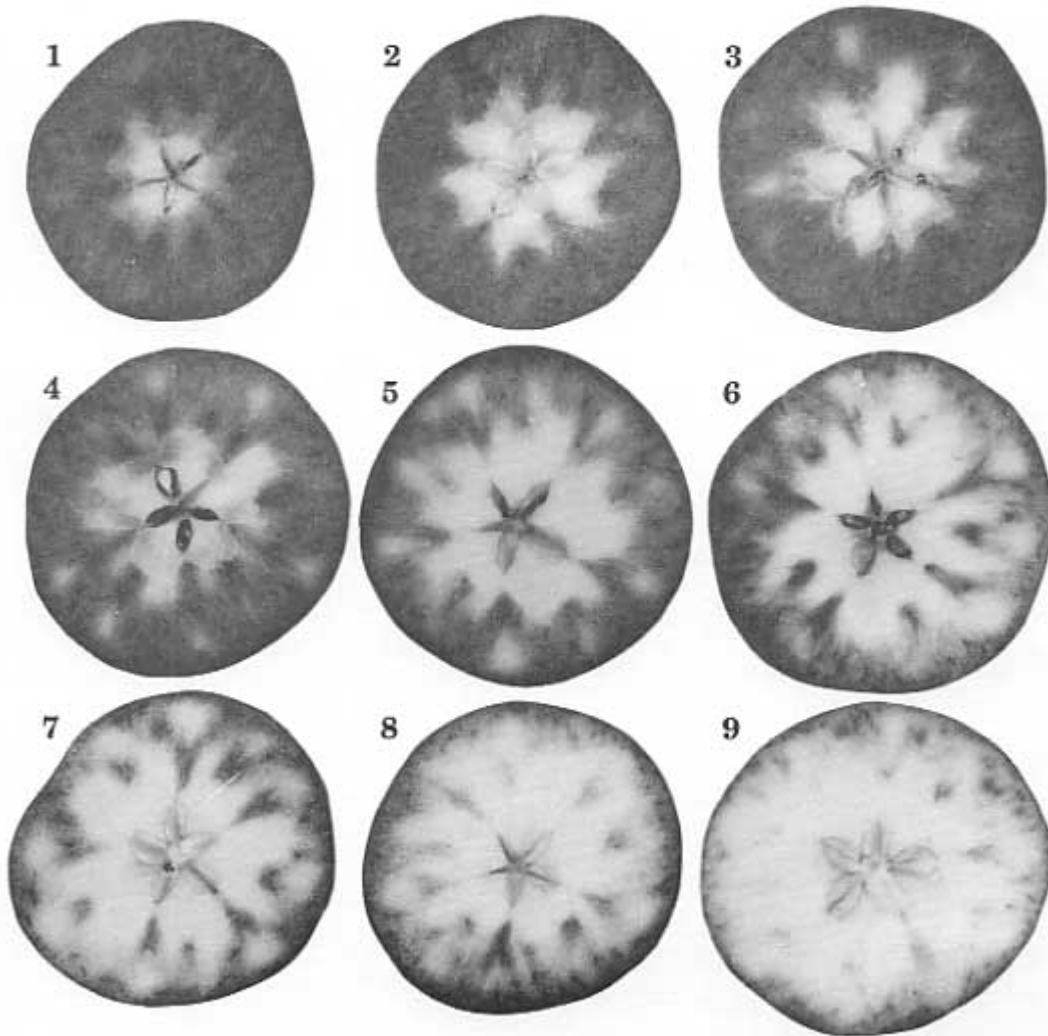
Cultivars	pH	TA (mg/g)
	3.12	4.05
Golden Delicious	3.24	2.53
	3.87	1.3
Delicious	3.83	1.23
	3.09	3.76
Granny smith	3.07	4.05



Fruit Program

Gala Starch Chart

Wesley R. Autio



Issued by the University of Massachusetts Cooperative Extension, Robert G. Helgesen, Director, in furtherance of the Acts of May 8 and June 30, 1914; University of Massachusetts, United States Department of Agriculture, and Massachusetts counties cooperating. University of Massachusetts Cooperative Extension offers equal opportunity in programs and employment. F-106:7/91-500

Figure 2.1. Gala-starch index Chart.

Chapter 3 - Materials and Methods

3.1. General

Gala and York apples were obtained from the Virginia Tech College of Agriculture and Life Sciences Kentland Farm. Apples were harvested at different times to obtain different maturity groups (immature, mature and ripe). Gala were harvested on Aug. 12 and 22, and Sept. 02, 2002 and York were harvested Sept. 22, October 7 and 22, and Nov. 02, 2002. For each harvest date 20 apples (15 apples for York on Sept. 22) were stored at 0°C for no more than four days. The apples from each harvest date were allowed to warm to ambient temperature overnight before experiments were conducted on 10 fruit samples on two consecutive days. The two sets used for sampling were randomly selected. Headspace evaluation (electronic nose, gas chromatograph, and gas partitioner) was performed on one day and maturity indices were measurement within 24 hours. Individual apples were placed in a 1.5 liter glass bottle for approximately one day. The headspace gas from the glass bottle was injected into a gas partitioner, a gas chromatograph and then was exposed to the electronic nose, successively.

3.2. Measurement of maturity indices

Firmness

Flesh firmness was measured using the Instron Universal Testing Machine (Model 1011, Instron Corp., Canton, MA) (Fig. 3.1), with a transducer capacity of 50 Kg. Firmness was measured on York after peeling. The probe used was a standard 11mm penetrometer head. The crosshead speed was 15mm/min. Measurements were made at

two positions per fruit, one on the equator and the other perpendicular to it. The maximum peak force was measured in each case.

Puncture test

This test was similar to that of the firmness test. The only difference was the skin of the apple under investigation was not removed. This was performed only on Gala.

Total Soluble solids

The undiluted expressed apple juice's was used to measure total soluble solid concentration in °Brix was measured using a digital refractometer (Model ABBE MARK II, Reichert Inc., Buffalo, NY) (Fig. 3.2).

Starch

Each fruit was cut in half, and the cut surface was immersed in a solution of 10 g of potassium iodide and 2.5 g of Iodine crystals. Using a starch index chart, a starch index value (1-9) was assigned to each fruit (Fig. 3.3).

Titrateable acidity

From literature (Keener, 1999), titrateable acidity (TA) and pH were correlated at least for Golden Delicious. So instead of TA, pH of the juice was measured using an electrode pH meter (Model AR15, Fisher-Scientific, Pittsburgh, PA) (Fig. 3.4).

Surface color

The greenest region of each fruit was measured with a tristimulus colorimeter (Model CR-300, Minolta Co., Ramsey, NJ) (Fig. 3.5), and values of L (Lightness), C (Chroma) and H (Hue angle) were calculated.

3.3. Ethylene measurement

A Fisher-Hamilton Gas Partitioner (Model 29, Diversified Equipment Company, Inc., Lorton, VA) (Fig. 3.6) was used for measuring ethylene in the headspace gas. A 1 ml syringe was used to draw the headspace gas from the glass bottle. Two consecutive columns in the gas partitioner were used for the experiments. The column no.1 was 6 ft x 1¼ in aluminum packed with 30% DEHS (Di-2-ethylhexylsebacate) on 60-80 mesh chromosorb P. The column no.2 used was 6-½ ft x 3-1/16 in aluminum packed with 40-60 mesh molecular sieve 13X. The run time was for 5.5 minutes. Thermal conductivity cell was used as the detector. The flow rate of the helium gas was 35 ml/min. The instrument was calibrated before each set of experiments, with a standard mixture of air and ethylene (1.15mole % of ethylene or 1.15 µl ethylene/ml of air) supplied by Airgas Specialty Gases (Theodore, AL). Using Appendix-A the units of ethylene measurements were converted to µl ethylene/ml of air.

3.4. Measurement of aroma compounds

Solid Phase Microextraction (SPME) technique was used in order to collect the aroma compounds present in the headspace gas. Polymethylsiloxane coated SPME fiber was used to analyze Gala, while carbowax coated SPME fiber was used for York. A

Hewlett-Packard Gas Chromatograph (Model 5890, Hewlett-Packard, Palo Alto, CA) (Fig. 3.7) was used with HP-5 column (25m * 0.32mm * 1.05 μ m) to detect the volatile compounds in the headspace gas. The carrier gas was helium with a flow of 1 ml/min in a splitless condition. The injection temperature was 280°C. FID (Flame Ionization Detector) was used with a temperature 300°C. A temperature program was used starting at 50°C (for 1.5 min holding time) and raising temperature 50°C/min for 5 minutes. The SPME fiber was exposed to headspace gas for 10 minutes, to attain equilibrium between the fiber and the headspace volatile. Then the fiber was retained in the injection port throughout the whole run.

3.5. Identification of aroma compounds

A Hewlett-Packard Gas Chromatograph (Model 6890, Hewlett-Packard, Palo Alto, CA) with a HP-1 column (25 m * 0.2m m * 0.33 μ m) (HP-1 Methyl Siloxane) was used to identify the volatile compounds in the headspace gas. Helium was the carrier gas with a flow of 4 ml/min. The injection temperature was 250°C. A MSD (Mass Spectrometer Detector) was used. The Gradient elution technique was used starting at 50°C (for 1.5 min holding time) and raising temp 50°C/min for 5 minutes. The SPME fiber was exposed to headspace gas for 10 minutes, to attain equilibrium between the fiber and the headspace volatile. Then the fiber was retained in the injection port through out the whole run.

3.6. Electronic nose Settings

The Electronic Nose System (Cyrano 320 (Model 320, Cyrano Sciences Inc., Pasadena, CA) (Fig. 3.8)) that was used. The EN system had 32 sensors. Preliminary experiments were conducted in order to develop a suitable method for the evaluation of apple aroma. The pictorial basic method settings in general, are shown in Fig. 3.9. Fig. 3.10 illustrates the scrolling strip chart, which allows researchers to follow the real-time response for all sensors. Fig. 3.10 shows a mature apple sensor response in real time, during the training of the EN system. The method settings for Gala and York are shown in Fig. 3.11 and 3.12 respectively. Every class from the selected method was chosen to train the electronic nose. The resulting smell print stored in the electronic nose, from exposure during the training is shown in Fig. 3.13 for Gala and Fig. 3.14 for York. Fig. 3.15 illustrates the trained set in the Cyranose 320.

3.7. Experimental Procedure

An apple was placed in the bottle for 26 hours for Gala. This time was required to measure ethylene levels, even for immature fruits. After the Gala results were analyzed, it was released that there might be an autocatalytic effect of ethylene on the apples, so the time for equilibrium should be reduced. To compare York results with those of Gala, 24 hours of equilibrium time was chosen.

Ethylene levels were measured by injecting 1ml of the headspace gas into the gas partitioner. The gas partitioner was calibrated before every set of experiments, using a standard ethylene concentration of 1.15 μl ethylene/ml of air (Airgas Specialty Gases,

Theodore, AL). Simultaneously the SPME fiber was exposed to the headspace gas in the bottle for 10 minutes. The fiber was then retained in the injection port of the gas chromatograph throughout the whole run of 15 minutes. Consecutively the electronic nose sensors were exposed to headspace gas.

After the headspace evaluation, the apples were removed from the bottle and weighed with an electronic balance. The ground color was measured using the tristimulus calorimeter. Firmness was measured on two sides of each fruit, with a penetrometer probe 11 mm in diam. The fruit was then cut in half and the cut surface was immersed in an iodine solution to evaluate the starch index. Soluble solids concentration in the juice was measured using a digital refractometer. The fruit was cut and ground to extract juice and pH of the fruit (only Gala) was measured with a pH meter.

3.8. Data Description

Maturity indices and electronic nose sensor response data were obtained from 60 of Gala and 75 York apples. The primary maturity indices that were used to define maturity were, starch index value, flesh firmness, soluble solids concentration and ground color.

3.9. Statistical Procedure

Principal Component Analysis (PCA), MANOVA, Canonical Discriminant Analysis (CDA), Discriminant Analysis (DA) and Canonical Correlation Analysis (CCA) were used to analyze maturity indices and EN response data for both Gala and York.

Electronic nose (EN) system software was a MATLAB program (Cyrano Sciences Inc., 2002). The algorithm used for both cultivars was Canonical. The data generated with the electronic nose sensors, were analyzed by three statistical procedures (PCA, DA and CDA) to verify the model developed with electronic nose software. Using statistical procedures in addition to the software analysis was important because the EN system software had a significant constraint that only six classes were available per method, and only 10 observations could be entered per and this allowed verification of the model that was developed by electronic nose system software.

All 32 sensors were used for the analysis. The primary interest was to establish that electronic nose could classify the apples into the same groups which were based on the maturity, firmness (puncture test), starch, soluble solids and color. All four maturity indices were used for Gala. Ethylene and pH (Gala) were not considered because ethylene was not internal ethylene and pH was not the actual titratable acidity. There was little information in the literature indicating if the two indices are related to maturity. But a subsequent separate analysis was performed including the pH and ethylene values.

SAS's PROC PRINCOMP (Cary, North Carolina) was used to identify outliers and grouping structure in the data. The number of eigenvalues was selected by evaluating cumulative explained variation. The effect of maturity level was evaluated with the MANOVA option of PROC GLM. In addition, a macro (Friendly, 1998) was used to check the condition for multivariate normality for the pooled data set. The CDA option PROC CANDISC was used to confirm the conclusions from PCA and MANOVA. DA was used to identify the misclassified observations and to determine the percentage of

correct classifications of apples in the maturity groups. The DA option in PROC DISCRIM supports three methods to verify the model. Of the three methods, the cross-validation method was used because the number of observations used for the model was relatively low when compared to the number of variables used in the model. Canonical correlation analysis was performed with PROC CANCELL to summarize the relationships that exist between the maturity indices and sensor response. The same procedure was used to analyze the chromatographic data.

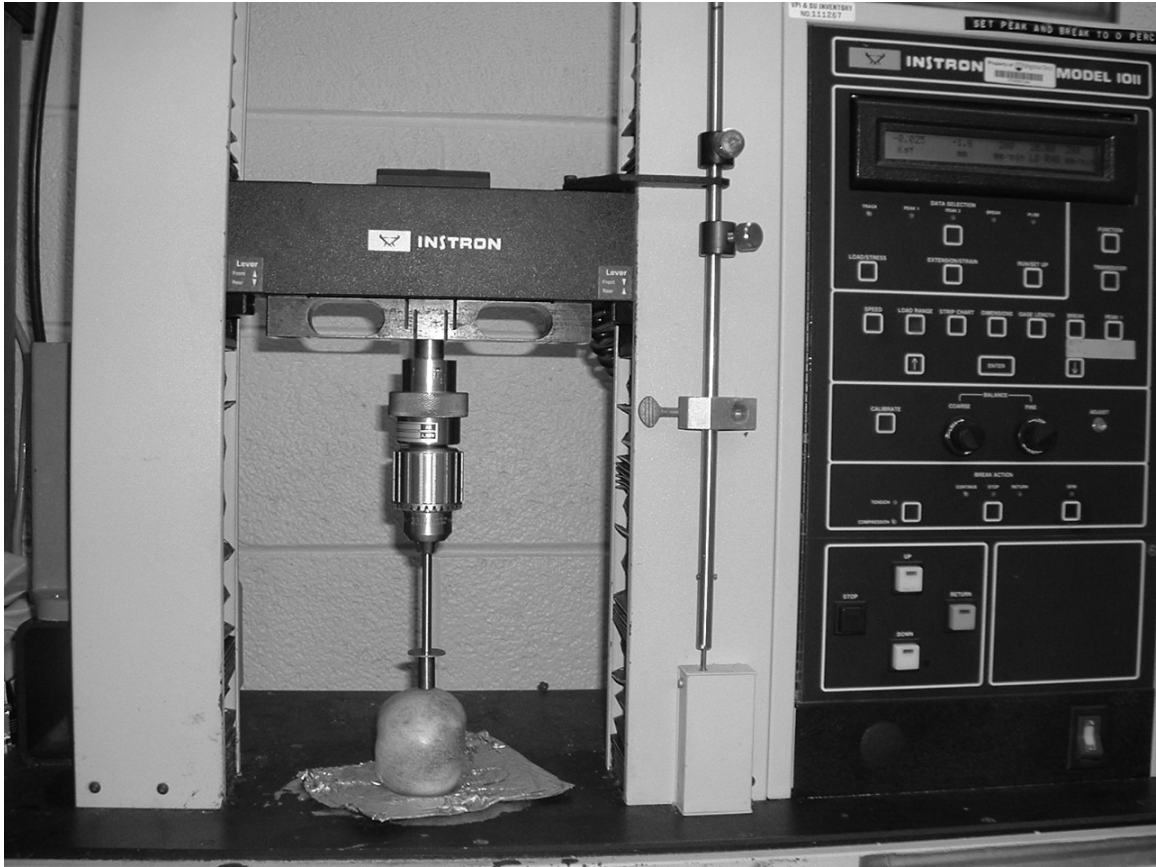


Figure-3.1. The Instron universal testing machine with the 11 mm penetrometer probe plunger used to measure flesh firmness.



Digital refractometer

Digital balance

Figure 3.2. Digital refractometer used to measure soluble solids concentration and the digital balance used to weigh fruit.



Figure 3.3. Examples of immature Gala dipped in iodine solution to evaluate starch hydrolysis. Starch is present in the dark areas, whereas light areas (non-stained) indicate starch hydrolysis.

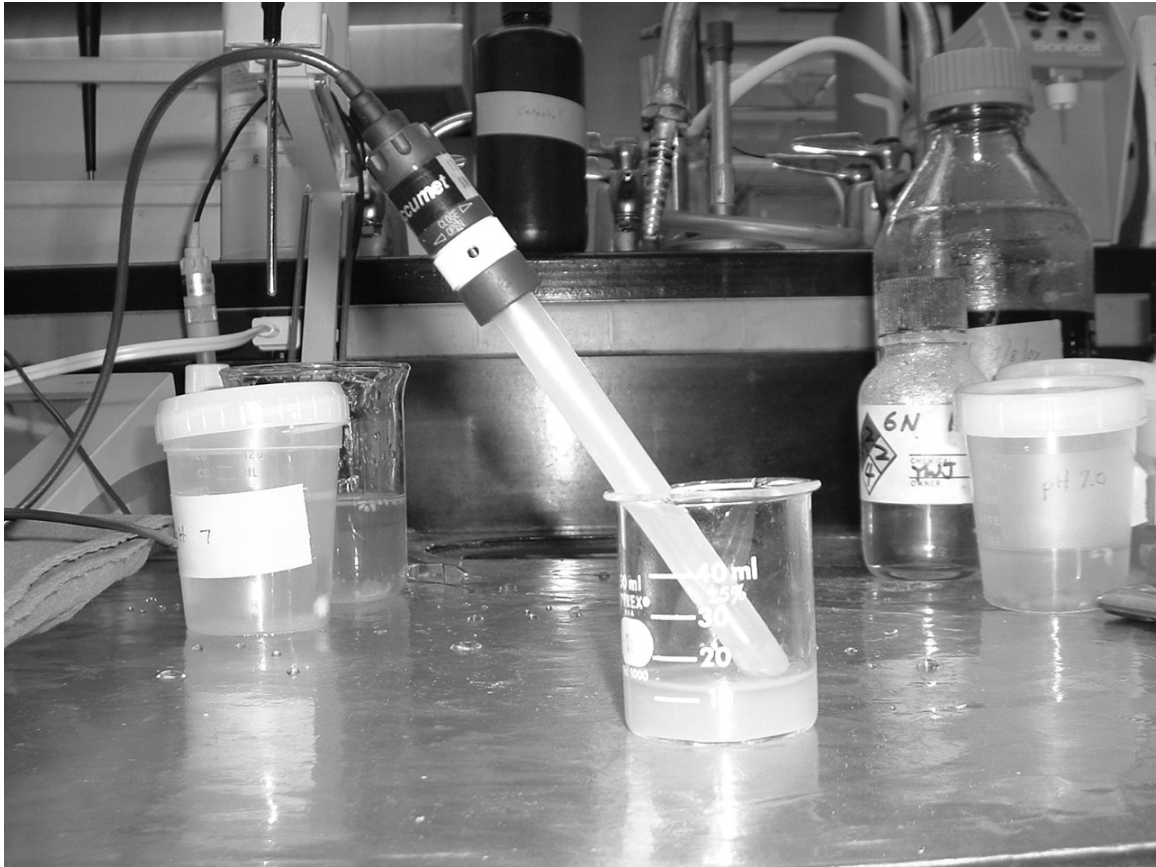


Figure 3.4. The pH meter was used to measure the pH of the juice extracted from the apple after grinding the flesh. The reference electrode (in this combination electrode) was completely immersed.



Figure 3.5. The tristimulus colorimeter used to measure the apple surface color.

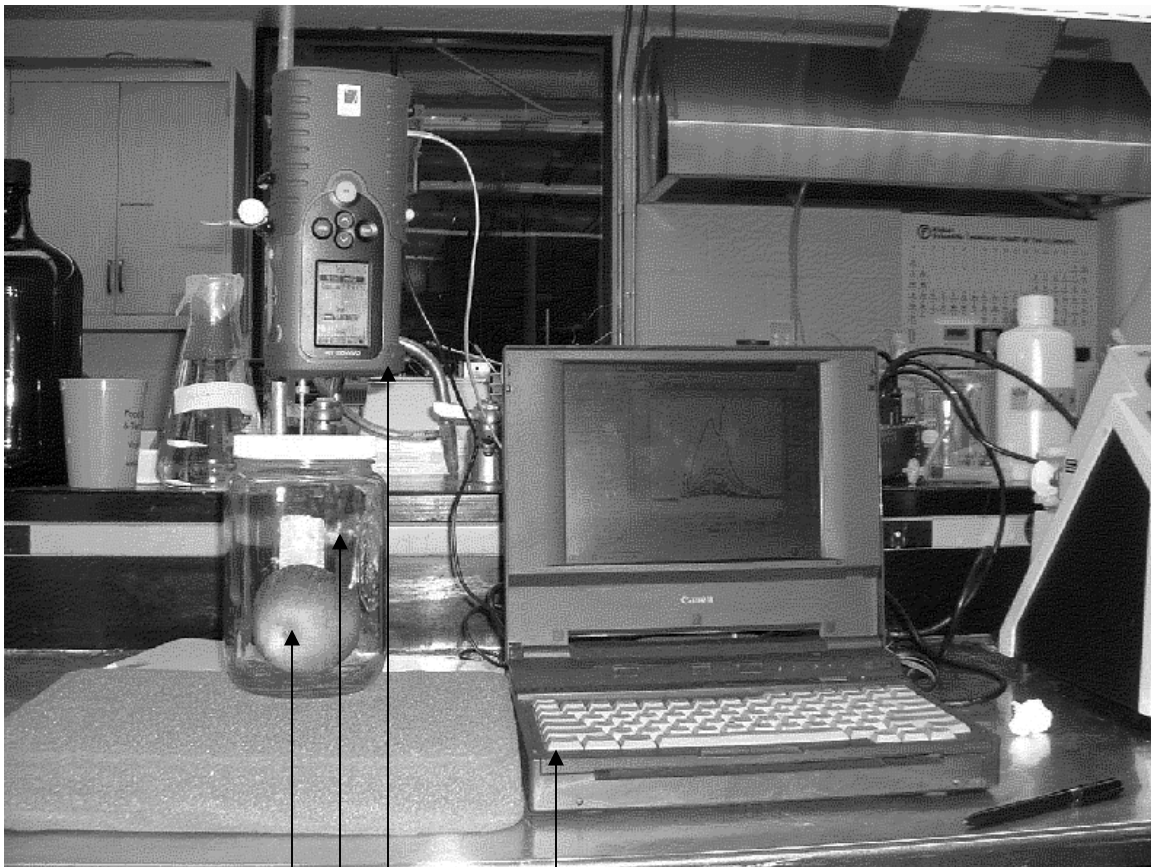


Figure 3.6. Fisher-Hamilton Gas Partitioner was used to measure ethylene.



SPME fiber held in the injection port

Figure 3.7. Gas Chromatograph used to measure volatile aroma compounds present in the headspace gas of the apple sample.



Apple sample

PC for data management

Cyrano 320

Headspace gas collected in the bottle

Figure 3.8. The Electronic Nose System (Cyrano 320) with sensors exposed to the headspace gas of the apple in the bottle.

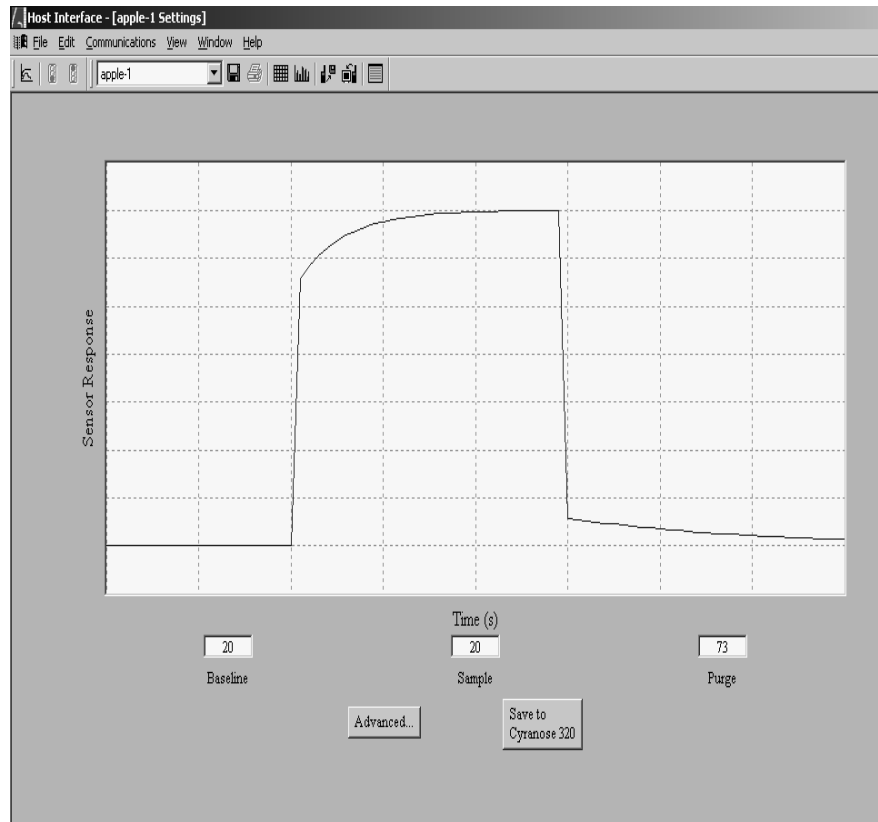


Figure 3.9. Basic settings adjusted for each method, depending on the sample. Baseline purge time, Sample exposure time and purge time are three basic settings.

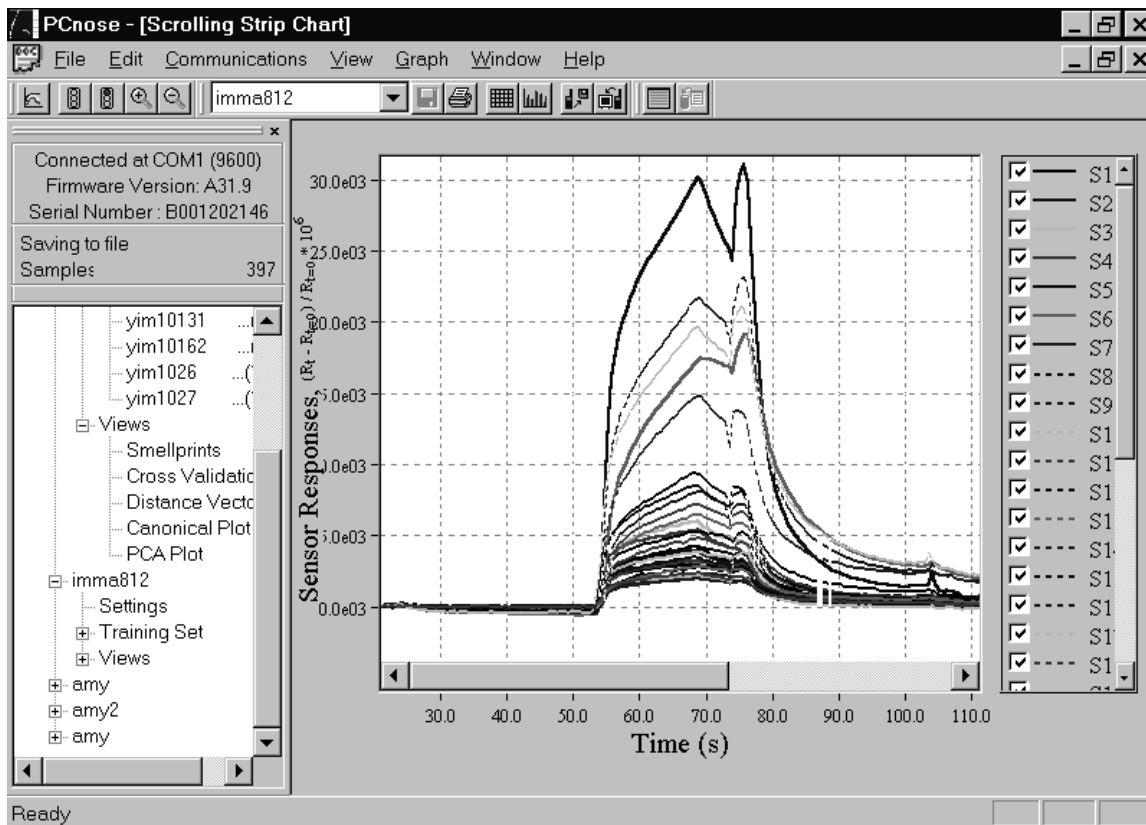


Figure 3.10. The scrolling strip chart of a mature apple. This figure illustrates the real-time responses for all 32 sensors. The Y-axis is represented by sensor responses, while the X-axis represents time in seconds.

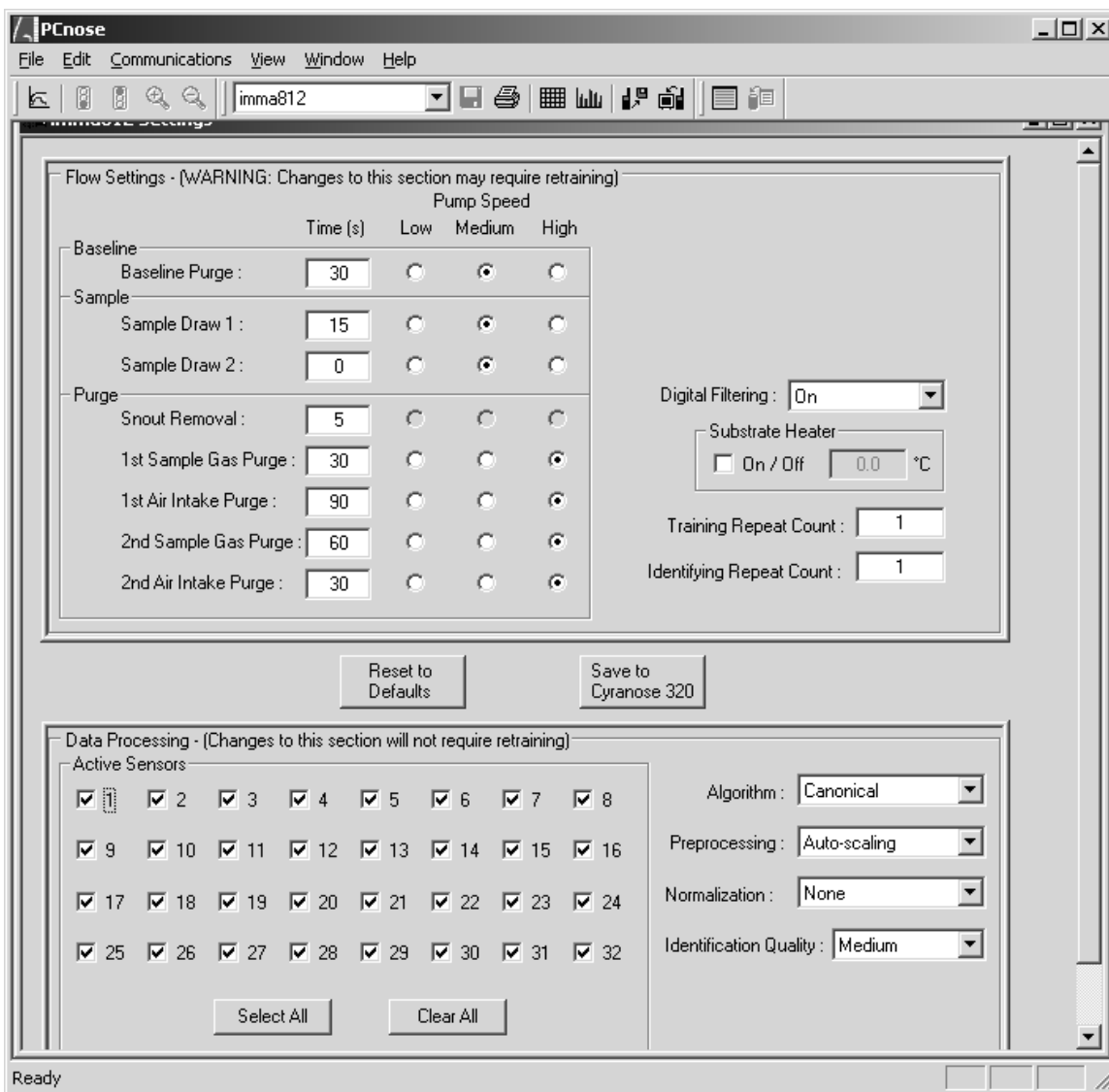


Figure 3.11. The basic method settings used for evaluating the Gala apple headspace gas. The canonical algorithm was used with no normalization. All sensors were switched on.

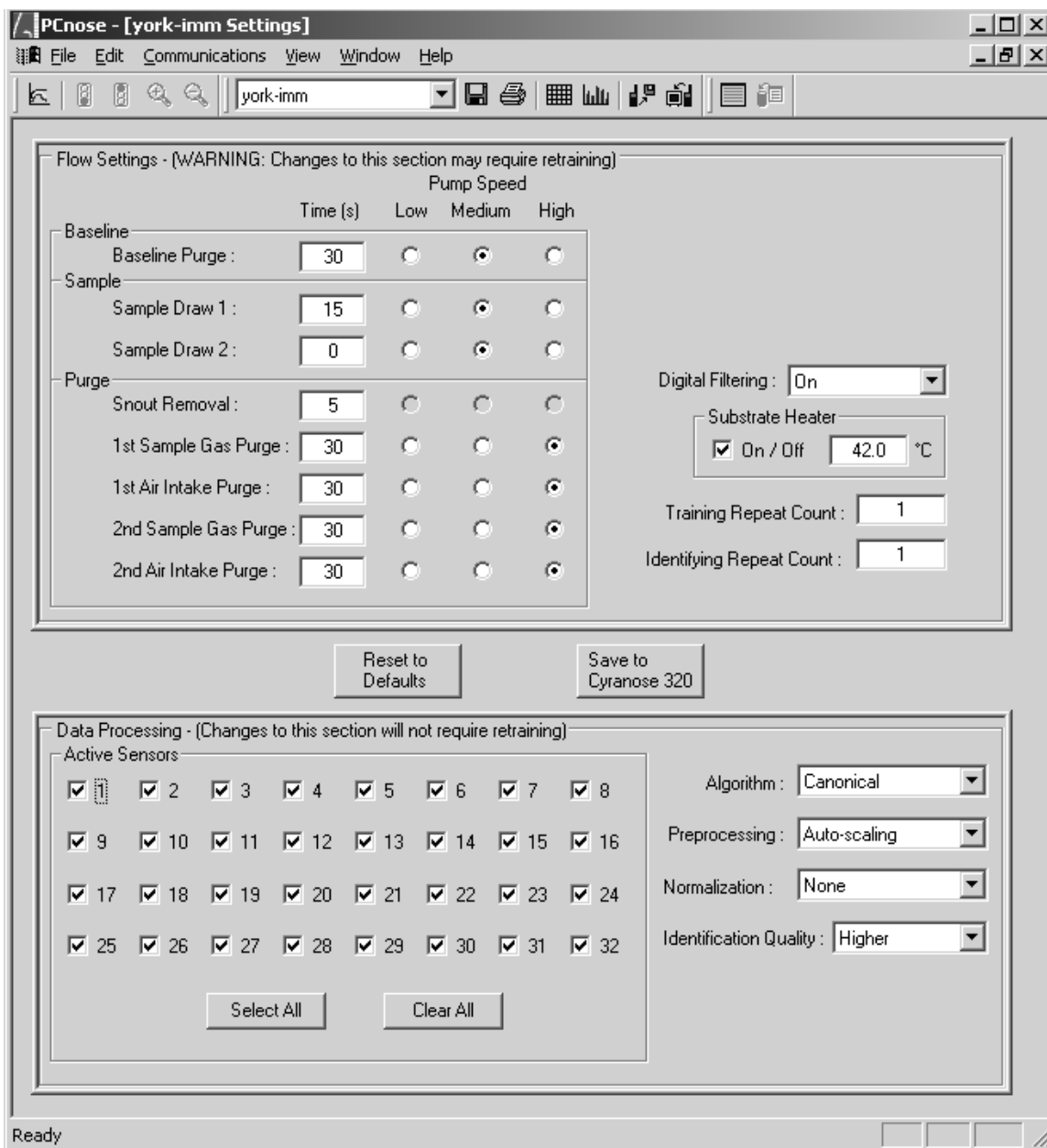


Figure 3.12. The basic method settings used for evaluating the York headspace gas. The main difference between the Gala and York basic settings was the purge timings.

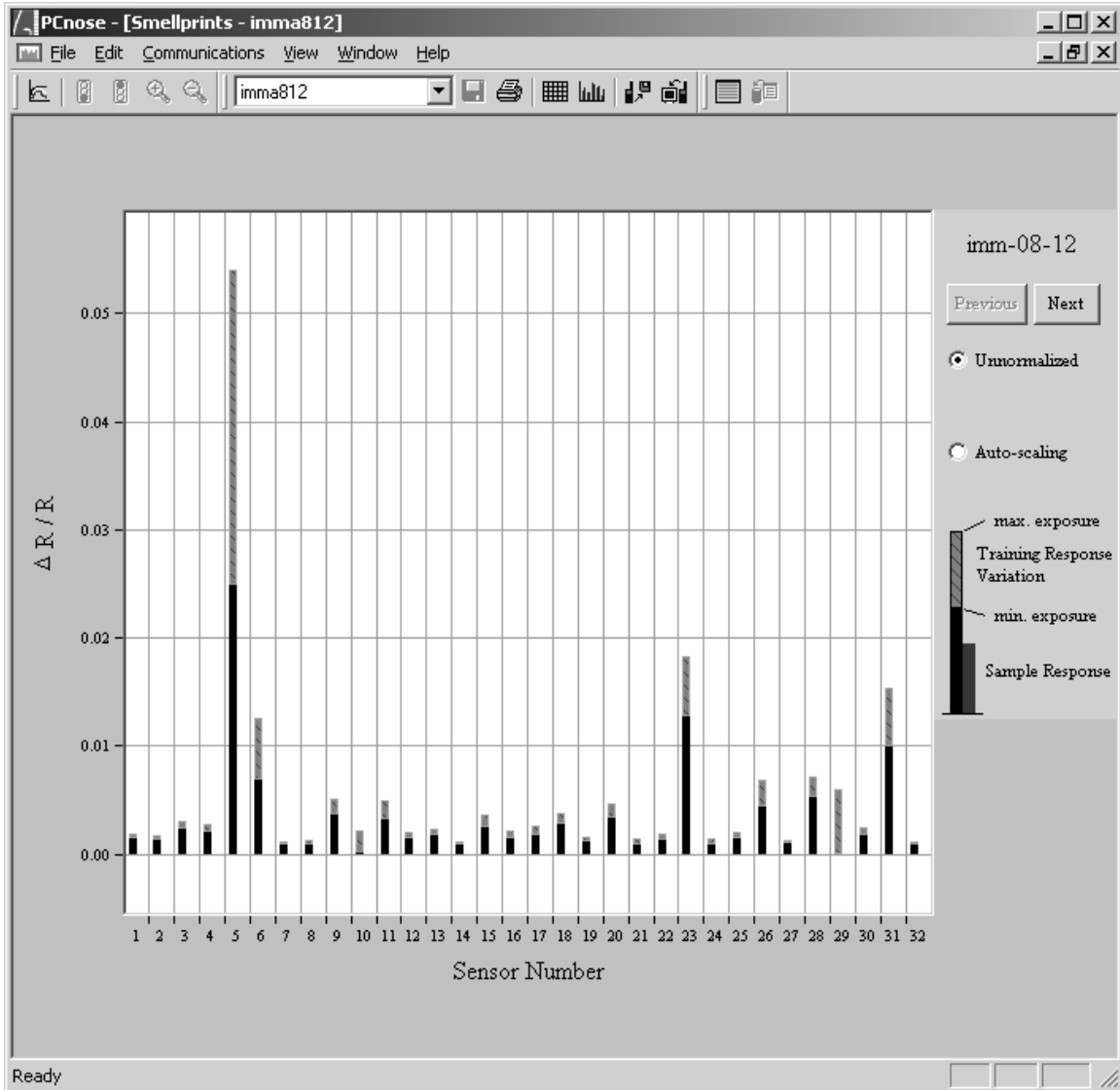


Figure 3.13. Smell print (or digital image) of immature Gala apples stored in the EN system memory.

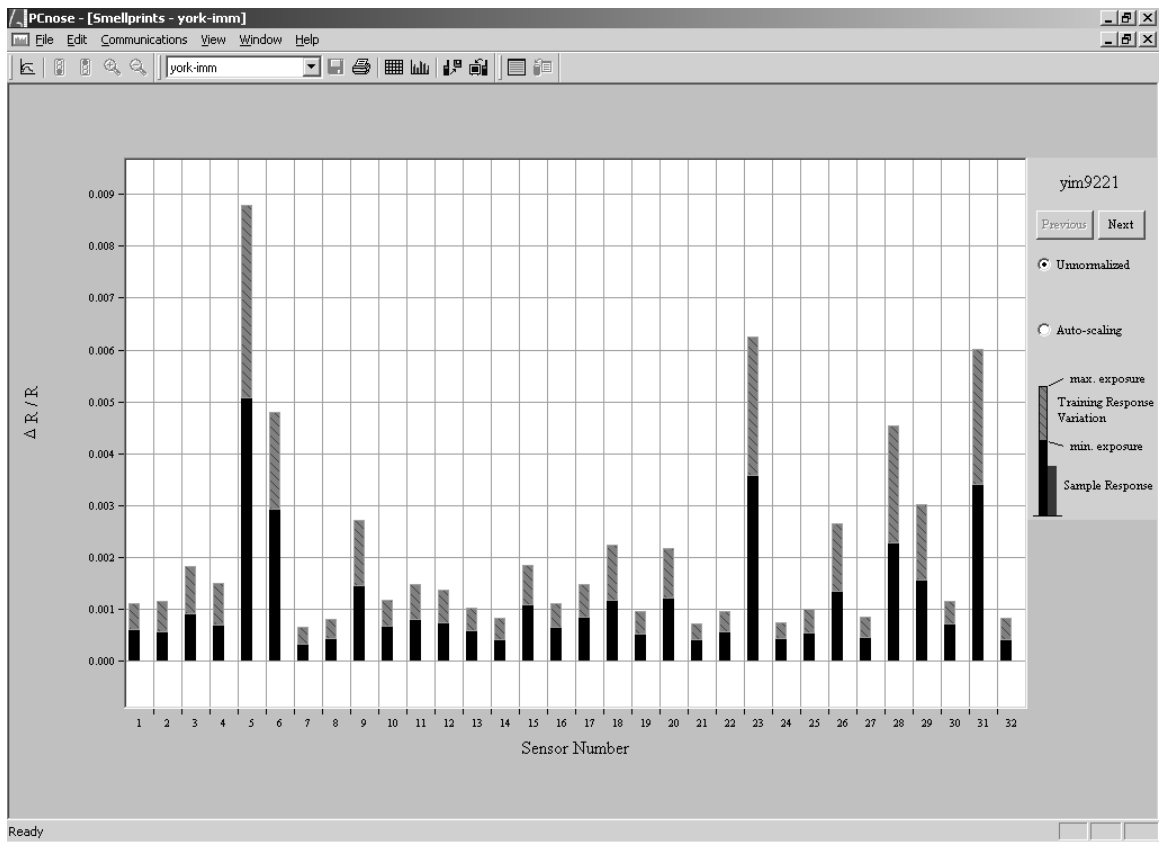


Figure 3.14. Smell print (or digital image) of immature York apples stored in the EN system memory.

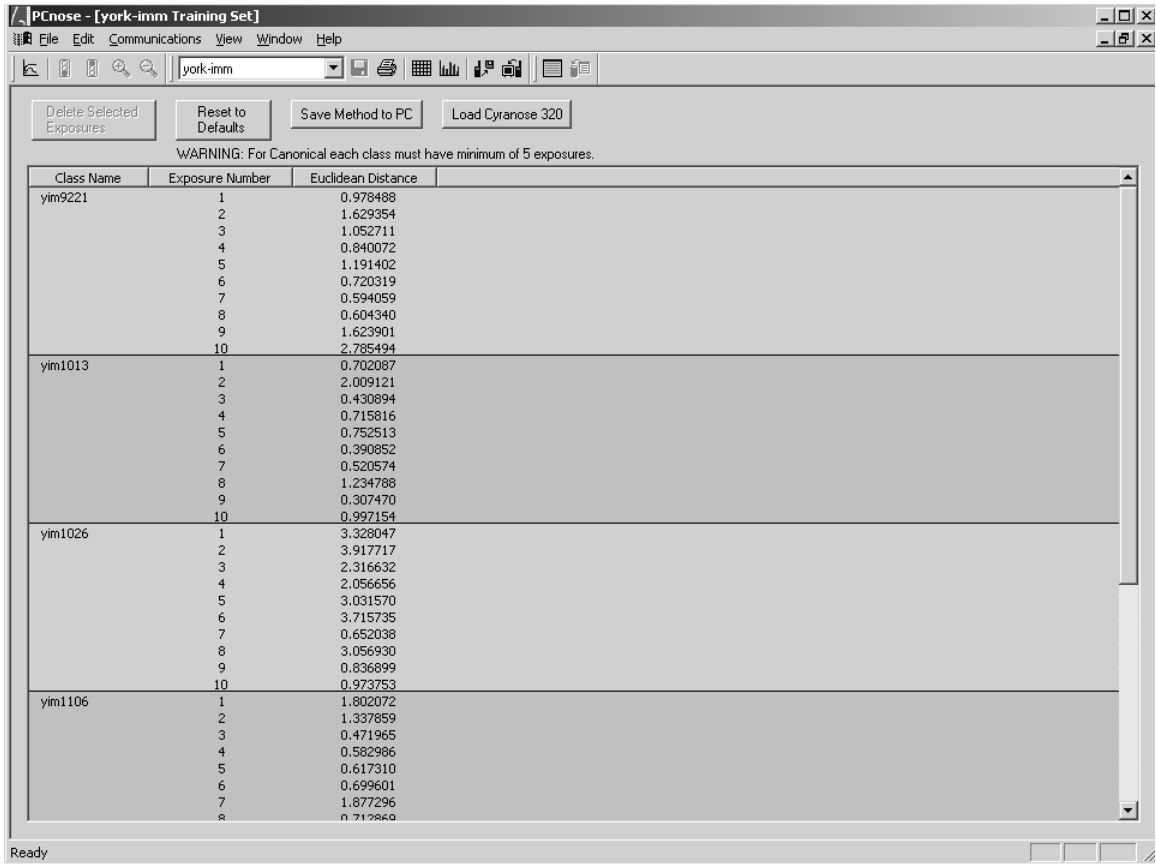


Figure 3.15. From the sensors responses, the Euclidean distance for each observation to the centroid of the class was calculated and stored in the training set.

Chapter 4 - GALA –Results and Discussions

4.1. Introduction

Six maturity indices namely, starch, puncture strength, soluble solids, color, ethylene and pH were measured on each apple. The experimental results of the five maturity indices except color are listed in Appendix-B. The mean values of the maturity indices are given in Table 4.1. Puncture strength values were compared with firmness values because there were no puncture strength values reported in literature for Gala apples. We expected higher numbers for puncture strength than that for firmness because puncture strength included skin resistance to the applied force. As expected the mean puncture strength values for second harvest date (Table 4.1) were higher than the reported values of firmness (Table 2.1). The starch values for first and second harvest dates were very consistent with the starch values reported in the literature (Table 2.1). Though the soluble solids concentration was slightly higher (only by 2%) in case of first harvest date (Table 4.1), the second harvest date values were very consistent with the reported values (Table 2.1). Though the color was measured for all 60 apples, the data were not used in the subsequent analyses due to a calibration error in the equipment (Appendix-C). The pH values were consistent with the data reported by Keener et al. (1999) (Table 2.3). Soluble solids often increase as apple mature, this was not the case in this study (Table 4.1). Values of pH varied little with harvest date. Such results are disappointing but not unexpected because soluble solids are influenced by many environmental factors, and also the 20-apple sample is generally inadequate for estimating all the maturity indices.

The ethylene values reported by Walsh et al., (1993) were of the range of 1-2 $\mu\text{l}/\text{kg}/\text{hr}$ for 5 hours, while the values represented in Table 4.1 when converted $\mu\text{l}/\text{kg}$ units (Appendix-B), the values had an average ethylene value of 0.16ml/kg for 26 hours (6.14 $\mu\text{l}/\text{kg}/\text{hr}$). Such exceptional high values may be due to the autocatalytic effect of ethylene. The trend of ethylene evolution was not unexpected because Walsh et al., (1993) concluded that the ethylene evolution rate was less before harvest (when compared to storage) and attributed the trend to the “parent-plant inhibition of ripening effect” displayed by the parent plant.

Exploratory and inferential analyses were employed. The exploratory analysis was used to explore the presence of any grouping structure in the data. The inferential analysis was used to derive meaningful inferences from the grouped structure. Since there was no established objective method to classify apples into three maturity groups (immature, mature and over mature), statistical analyses such as PCA, CDA and DA were performed on maturity indices to identify and categorize the observations into three maturity groups. Then EN sensor response data was categorized into the above-maturity based categories. The efficiency of EN sensors in classifying the data into maturity groups was evaluated by performing statistical analysis on the EN sensor response data. EN system software was also used to evaluate the efficiency of the sensors. Simultaneously statistical analysis was used to validate the EN system software results. In addition, extra statistical procedures were employed to understand two factors (pH and ethylene) that are affected by maturity; and also to understand the correlation between the maturity indices and the EN sensors.

4.2. Gala- Statistical Analysis of Maturity Indices

4.2.1. Exploratory analysis

The three maturity indices, starch index, puncture strength and soluble solids were used as the three variables for the statistical analysis to identify and categorize the data into three maturity categories. Apples were numbered from 1 to 60 and PCA was performed on these numbered observations to classify the apples into three maturity categories (Fig. 4.1). Three clusters were identified from a plot of the first two principle components. These groups are referred to as immature, mature and over mature groups. Further analysis was performed to improve the classification, which was originally 92% correct. By running a DA on the three groups, the misclassified observations were identified, and the categories were reorganized. A classification table was developed from the DA the classification of the reorganized groups, and 5 % of the observations were misclassified (Table-4.2).

The results from the PCA of the above-classified data (maturity-based categories) are given below. There were no apparent outliers in the data. From the correlation matrix (Appendix-D), puncture strength and starch index were highly correlated ($r = -0.81$). Based on eigenvalues the first two principal components were selected because together they accounted for 93% of the total variation. The first principle component (Prin1) that explained approximately 63% of the total variation was primarily related to starch and puncture strength. Prin2 was related to soluble solids.

MANOVA was performed and the Q-Q plot (Fig.4.2) indicates that the data are from a multivariate normal distribution. Results from the MANOVA also indicate that population means differed significantly (Wilks' Lambda $F = 35.62$, $P < 0.0001$).

4.2.2. Inferential analysis

The scatter plot of the first two principal components (Fig. 4.3) indicates that Prin1 axis was represented by the starch and puncture strength variation scale, while Prin2 axis was related to soluble solids variation scale. PCA and Prin1 indicate apples in the immature category (represented as 1) had lower starch values and higher puncture test values. The mature group (represented as 2) had intermediate starch and puncture strength values, and the over mature group (represented as 3) have high values of starch and low values of puncture strength. The soluble solids for all three categories was scattered all along the Prin2 axis.

The first eigenvalue from the CDA explained 100% of the total variation, so the first canonical component (Can1) was used. From the analysis most of the Can1 variation could be explained by starch and puncture strength. Thus confirming the interpretation from PCA that starch and puncture strength were the most important maturity indices for classifying Gala into maturity groups. For Can2 the major contribution was from firmness. The scatter plot of the first two canonical variables (Fig. 4.4) illustrates three clusters along the Can1 axis, thus confirming the inference drawn from the PCA and CDA that three maturity groups were indeed present. The classification table from DA shows that 5% of the apples were misclassified (Table 4.2).

4.3. Statistical Analysis of Electronic Nose Sensor Response Data

4.3.1. Exploratory analysis

The apples were separated into groups based on maturity indices (sect. 4.2.2), so the present analysis was performed on the electronic nose sensor responses to determine the efficiency of the electronic nose for categorizing the apples into the same groups. PCA was performed on EN data to determine if there were any clusters that would represent the grouping (maturity) structure in the data. From the PCA correlation matrix, most of the sensors were highly correlated ($r > 0.85$) to each other, while correlation (r) among few sensors was as low as 0.28. All sensors contributed almost equally to prin1, confirming the non-specific nature of these sensors. From the eigenvalues, the first four principal components together explained, as much as 98.5% of the total variation, and so these four were the most important principal components. The Q-Q plot (Fig. 4.5) indicates that the observations are from a multivariate normal distribution. MANOVA indicated that the maturity groups were not all equal (Wilks' Lambda $F = 3.7$, $P < 0.0001$). Thus it can be concluded that maturity influenced electronic nose sensor responses.

4.3.2. Inferential analysis

The scatter plot of the first two principal components group the data into 3 clusters. Maturity level 1 was discriminated by Prin2 and maturity levels 2 and 3 were discriminated by Prin1 (Fig. 4.6). Since the sensors were non-specific, there could not be any meaningful interpretation of the axes in the PCA plot. The first four principal

components from the PCA were used as input for CDA and DA. The first two eigenvalues produced by CDA explained 100%, so the first and second canonical component (Can1 and Can2) explained all the variation. From the analysis of Can1, the important contributors were Prin1, Prin2 and Prin4, but inference concerning these axes in terms of sensors could not be deduced due to the non-specific nature of the sensors. The scatter plot of the first 2 canonical variables (Fig. 4.7) indicates three clusters, thus verifying the inference drawn from the PCA that electronic nose can classify the apples into maturity categories. The classified table from DA (Table 4.4) shows that 17% of the apples were misclassified. The interclass separation (Table 4.5) between the three groups using EN sensor responses was less than the interclass separation between the three groups generated using maturity indices. An interclass distance of 5 or more is considered as a good indicator of good separation (Cyrano 320 User's Manual, 2000), the EN sensors were able to separate the apples into three groups.

4.4. Electronic Nose System Software Analysis

The same electronic nose sensor data that was used in earlier statistical analyses (PCA, CDA and DA) were used as the raw data for the EN software (MATLAB) analysis. The main constraint with this software was that the number of classes available per method was only six, and only 10 observations can be entered into each class. Therefore, the 60 observations for Gala apples were divided in 6 classes with each maturity based-category being represented by two consecutive classes in the EN system software. The immature group had 21 observations; one observation was deleted for a total of 20 observations. Only 19 apples in the over mature group resulting a total of 59

observations instead of 60 observations. The analysis was performed using the canonical algorithm (programmed in MATLAB). The analysis results are presented in Table 4.6 and Fig. 4.8 and 4.9 (DA Table, PCA plot and CDA plot, respectively). Seventeen percent of the apples were misclassified (Table 4.6). The interclass distance between the groups seemed to be close, but since two classes represented one particular group, it was hard to derive a conclusion. To duplicate the EN software, the DA (SAS) was performed with the above data (59 observations), and the correct classification percentage was 85%.

4.5. Canonical Correlation Analysis

The canonical correlation analysis (CCA) was performed to determine if the electronic nose sensors were correlated with the maturity indices. From the correlation matrix many sensors were highly correlated ($r > 0.70$) to starch and puncture strength. Soluble solids were poorly correlated ($r < 0.20$) with the remaining all maturity indices. The likelihood ratio test indicated that three canonical correlation were statistically significant, so NCAN=2 option was not used for CCA. From the CCA, the first canonical correlation function (Chr1) generated for the maturity indices was highly correlated to puncture strength and starch, while Chr2 was highly correlated to soluble solids. Twelve EN sensors were highly correlated to the first canonical function (Sens1) and 8 sensors were correlated to Sens2. From CCA correlation analysis, it could be concluded that the sensors that were highly correlated ($r > -0.7$) to chr1 (starch and puncture strength) were sensors 1, 2, 3, 4, 8, 9, 12, 18, 19, 27, 28 and 29. Chr2 was poorly correlated to all the three maturity indices ($r < 0.25$). The CCA plot-1 (Fig. 4.10) illustrates a strong linear correlation between the sensors (canonical function1, Sens1)

and the maturity indices (canonical function1, Chr1), indicating that the twelve sensors were linearly correlated with puncture strength and starch. From the CCA plot –2 (Fig. 4.12) indicates a poor correlation, indicating the remaining 8 sensors were not correlated to non of the three maturity indices.

4.6. Analysis of pH and ethylene effect on the maturity classification

To investigate whether pH could be used as a maturity index, for the above-classified groups, pH was appended as an extra maturity index and the statistical analyses (PCA, CDA and DA) were rerun. From correlation matrix for all five maturity indices (including pH and ethylene), pH was poorly correlated to all four maturity indices ($r < -0.24$). From the DA analysis the correct classification percentage came down from 93% to 88%. Thus it can be concluded that pH may not be a good indicator of maturity for Gala apples. This supports the results of Keener et al. (1999), where only Golden Delicious had a statistically significant relationship ($P < 0.009$), between TA and pH, out of the three apple cultivars Granny Smith, Golden Delicious and Delicious.

Similar analyses were rerun twice, by including total ethylene levels, and then ethylene levels adjusted for fruit weight. Both analyses reduced the maturity based classification percentage to 88% (total ethylene) and 92% (adjusted ethylene). From correlation matrix, the total ethylene levels were modestly correlated with puncture strength ($r = -0.36$) and starch ($r = 0.29$), and were not correlated with soluble solids ($r < -0.15$).

Consequently, pH and ethylene were both appended to the other maturity indices. From the DA analysis, the classification percentage came down to 87%. A CCA analysis was performed between all six maturity indices (including ethylene and pH) and the electronic nose sensor responses. The sensor responses were highly correlated with firmness ($r > -0.7$) and starch ($r > 0.7$), modestly correlated with ethylene ($r > 0.45$) and poorly correlated with soluble solids ($r < 0.1$) and pH (almost all $r < 0.1$) (Appendix-C).

4.7. Harvest Date Effect

In order to ascertain the harvest date effect, the data were classified based on the harvest date into three groups. To evaluate the data even through the electronic nose software, the 3 harvest dates were grouped into 6 classes, two consecutive classes representing one harvest date. From the EN software analysis, the correct classification percentage was 76%. DA (for the above 60 observations) grouped the fruit into correct harvest date approximately 83% of the time.

4.8. Analysis with Reduced Number of Sensors

To investigate if there was any moisture effect in the electronic nose sensor response data, the moisture sensitive sensors (sensors 5, 6, 23 and 31) were removed and the analysis was performed. From DA, there was no significant difference in the output, with the DA classification percentage reduced by 3% from 83% to 80%. So it can be concluded that there was no significant effect of moisture on the sensor responses.

The stepwise (significance level for elimination (SLE) = 0.40 and significance level for staying (SLS) = 0.15) method of the STEPDISC procedure in the DA was run

using all 32 sensors, in order to reduce the number of sensors that were statistically significant. Consequently, 26 sensors were removed by the statistical procedure, leaving only 6 sensors (5, 6, 14, 16, 17, and 23), which had significant effect on the classification of the data into three maturity groups. From DA, there was no significant difference in the output, with the DA classification percentage reduced by 3% from 83% to 80%.

4.9. Summary

From the PCA of the observations (1-60), three categories were selected which were based on maturity indices namely, starch, and puncture strength and soluble solids. Further analysis such as PCA and CDA illustrated the presence of three clusters (Fig. 4.3 and 4.4). From the PCA plot (Fig. 4.3), the principal axis Prin1 can be interpreted as the starch index and puncture test variation scale, and Prin2 represented soluble solids variation scale. Thus the three clusters can be divided along the Prin1 axis explaining approximately 63% of the total variation. From DA cross-validation results, the correct classification percentage was 95%. The Three maturity based-categories were named as immature, mature and over mature groups consisting of 21, 20 and 19 observations respectively.

The Electronic nose sensor responses were grouped into the above-maturity (maturity indices) categories. From PCA and CDA results, the electronic sensor effectiveness to classify the apples into the above maturity groups was determined. From the DA cross-validation results, the correct classification percentage was 83%. The same electronic nose sensor response data were run using the EN system software (MATLAB). From the cross validation results the correct classification percentage was 83%. The

above six classes were also statistically evaluated, and from DA, the correct classification percentage improved to 85%.

From the CCA, 12 sensor responses were correlated with the starch and puncture test values. It was concluded that pH was not a good indicator of maturity for Gala apples. The efficiency of electronic nose system software to classify the apples based on harvest data was 76%, while the same analysis by statistical DA was 83%. The moisture sensitive sensors (5, 6, 23 and 31) were removed, but there was no significant effect on the classification percentage. The STEPDISC procedure in DA was performed to reduce the number of sensors, and the remaining sensors that were statistically significant were sensors 5, 6, 14, 16, 17, and 23.

Table 4.1. Mean values (with standard deviation) of maturity indices for Gala apples harvested on three dates.

Harvest date	Puncture strength (N)	Soluble solids ($^{\circ}$Brix)	Starch index (1-9)	pH	Ethylene (ml /kg of fruit for 26 hr)
08/12/02	137.4 \pm 16.5	14.7 \pm 1.2	2.4 \pm 0.1	3.8 \pm 0.9	0.17 \pm 0.06
08/22/02	126.1 \pm 13.1	14.9 \pm 0.9	5.0 \pm 2.0	3.6 \pm 0.1	0.16 \pm 0.09
09/02/02	98.0 \pm 15.8	14.3 \pm 1.2	7.1 \pm 1.7	3.7 \pm 0.1	0.2 \pm 0.07

Table 4.2. Classification table obtained after performing DA for Gala apples. Columns and rows represent the maturity categories (1, 2, and 3). The values represent the number of fruits placed in that category. The groups were categorized based on maturity indices.

Actual Maturity Group	Predicted classification group		
	1	2	3
1	21	0	0
2	2	18	
3	0	1	18

Table 4.3. Mahalanobis distance to determine the similarity of groups of Gala apples with unknown maturity to a set of values from 3 groups of apples based on maturity indices. The left column indicates the groups and the right column indicates the distance.

From Maturity	Mahalanobis Distance
1-2	4.5
1-3	28.1
2-3	10.2

Table-4.4. Classification table obtained after performing DA for Gala apples. Columns and rows represent the maturity categories (1, 2, and 3). The values represent the number of fruits placed in that category. The categorization was performed on EN sensor response data.

Actual Maturity Group	Predicted classification group		
	1	2	3
1	16	5	0
2	2	16	2
3	0	1	18

Table-4.5. Mahalanobis distance to determine the similarity of groups of Gala apples with unknown maturity to a set of values from 3 groups of apples based on maturity indices. The left column indicates the groups and the right column indicates the distance between them.

From Maturity	Mahalanobis Distance
1-2	4.4
1-3	21.6
2-3	10.8

Table 4.6. Classification table obtained from the EN software analysis for Gala apples. Columns and rows represent the maturity categories (1, 2, and 3). The values represent the number of fruits placed in that category. The categorization was performed on EN sensor response data.

Actual Maturity Group	Predicted classification group		
	1	2	3
1	15	5	0
2	2	17	2
3	0	1	19

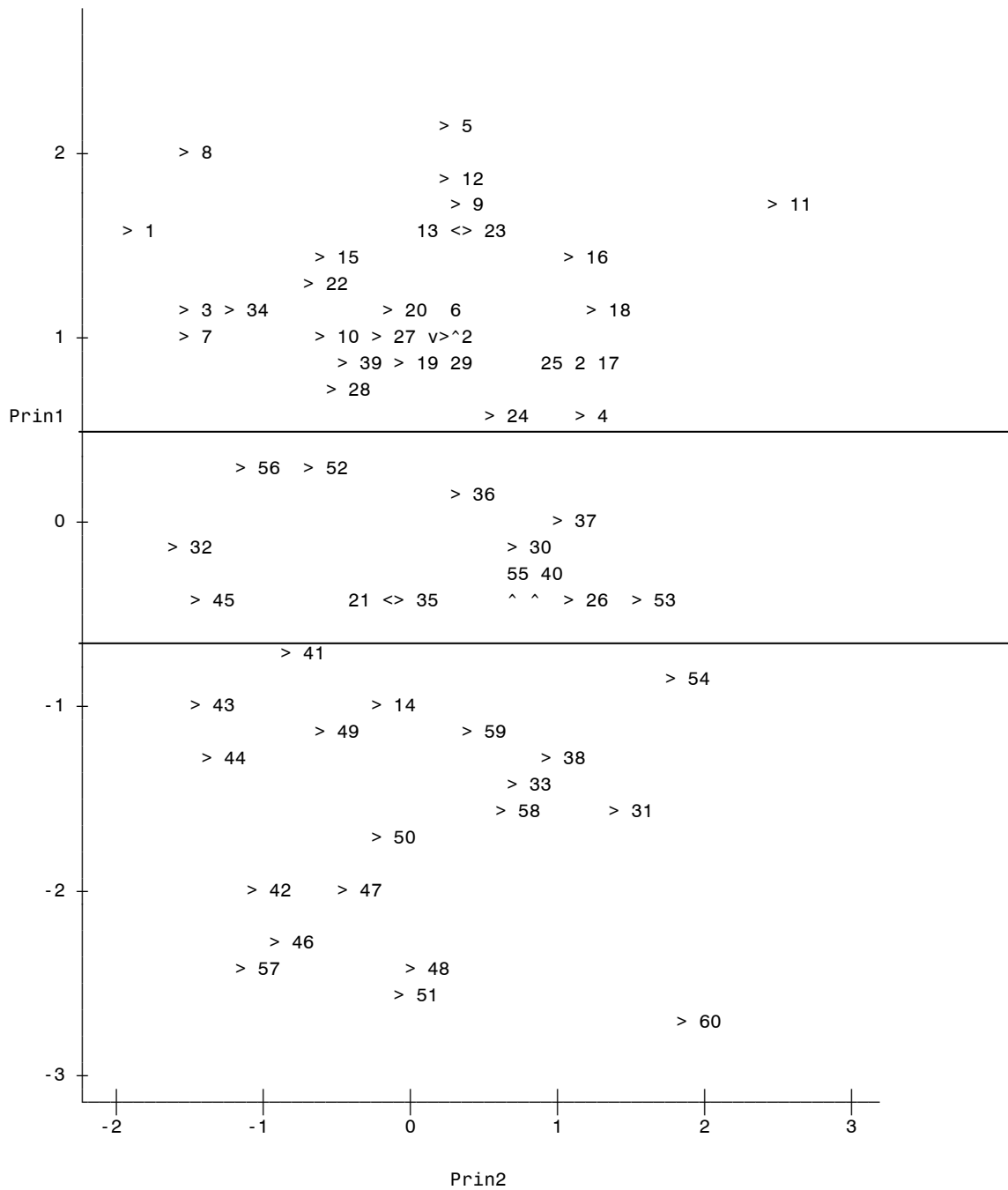


Figure 4.1. Scatter plot of the first two principle components illustrating three clusters. The 1-60 represents the observations. Prin1 is represents starch and firmness variation. Prin2 represents primarily represents soluble solids.

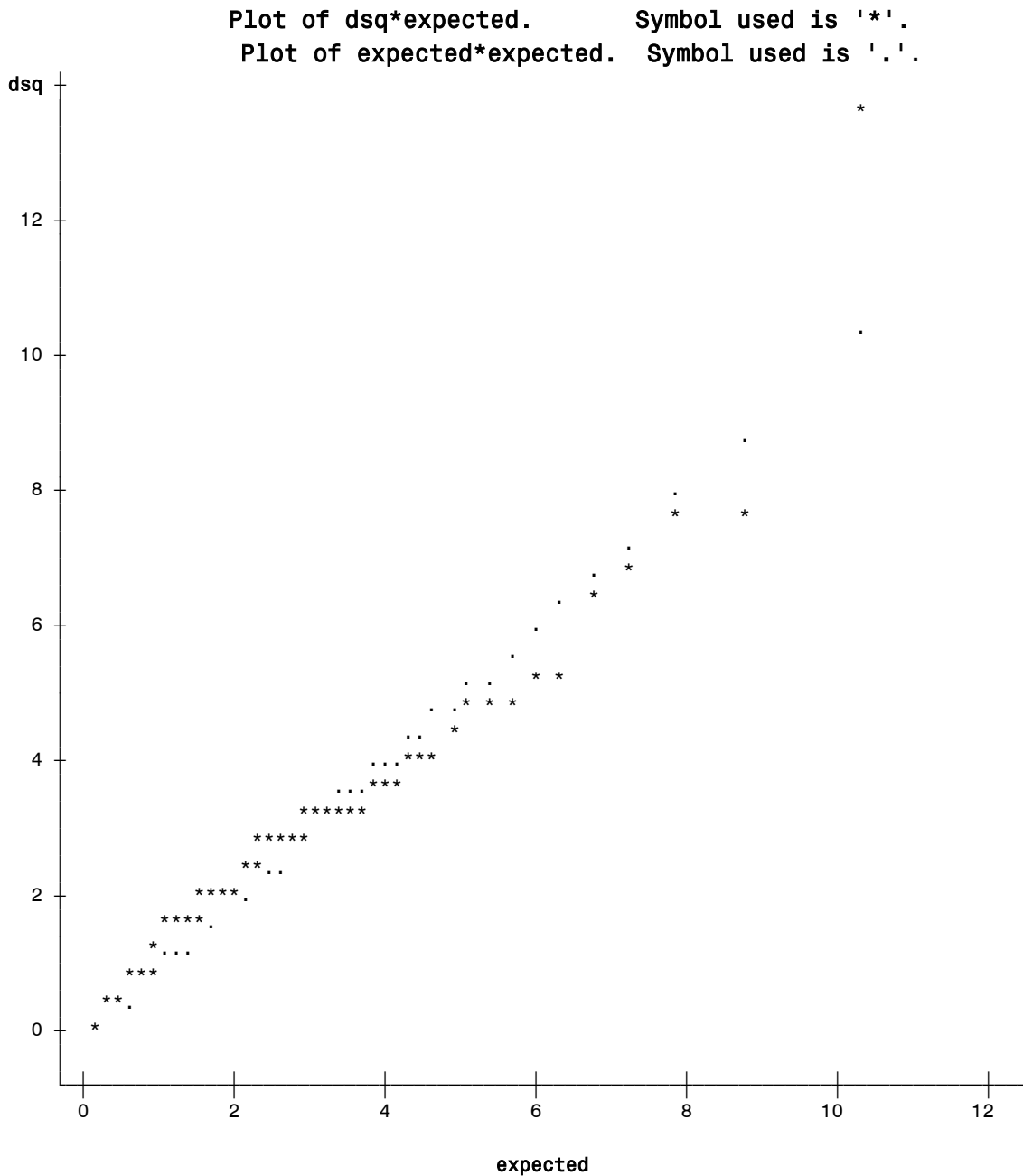


Figure 4.2. Chi-Squared probability plot for multivariate normal. The X-axis (expected) represents Chi-Square Quantile. The Y-axis (dsq) indicates the mahalanobis D-square distance.

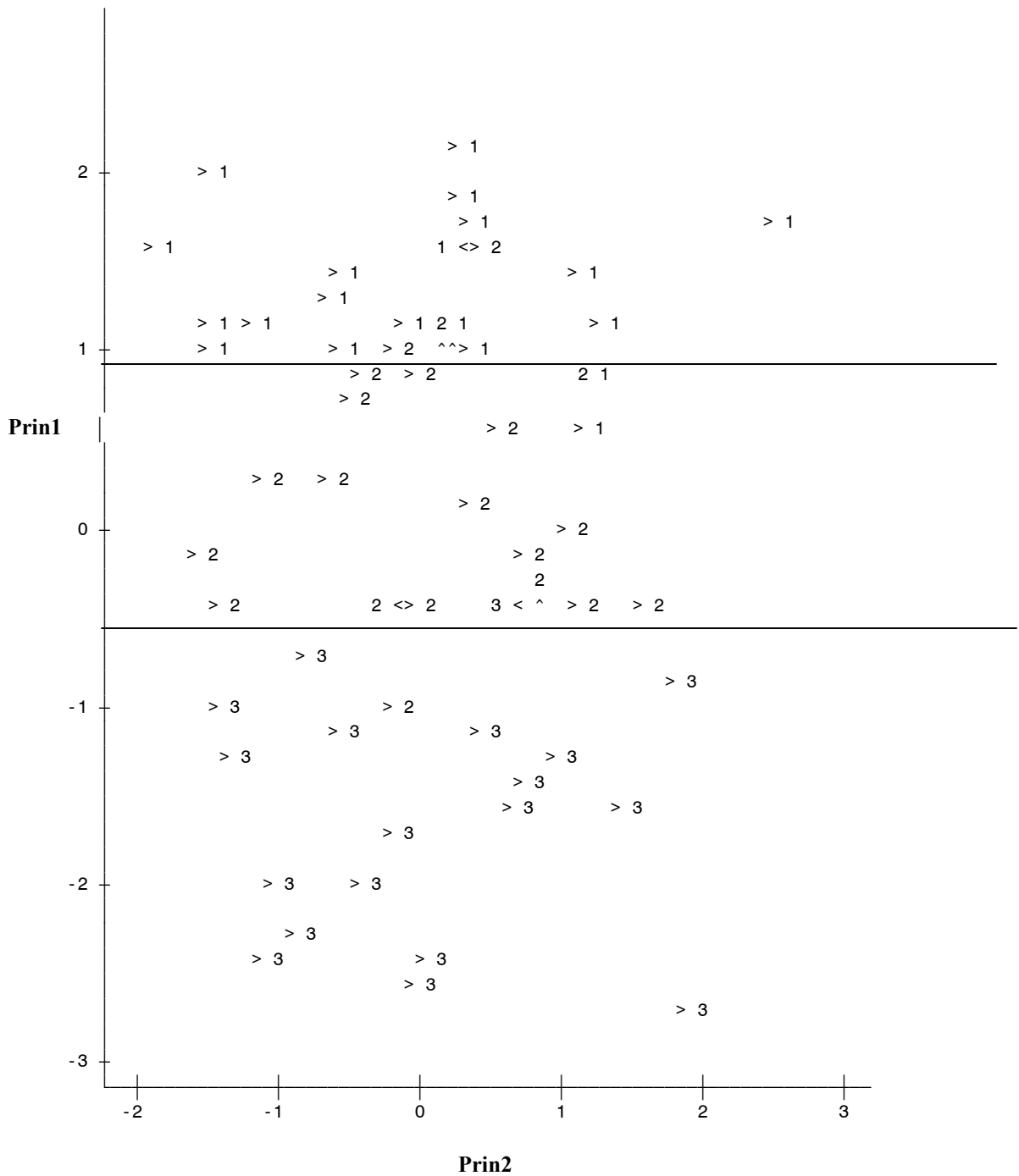
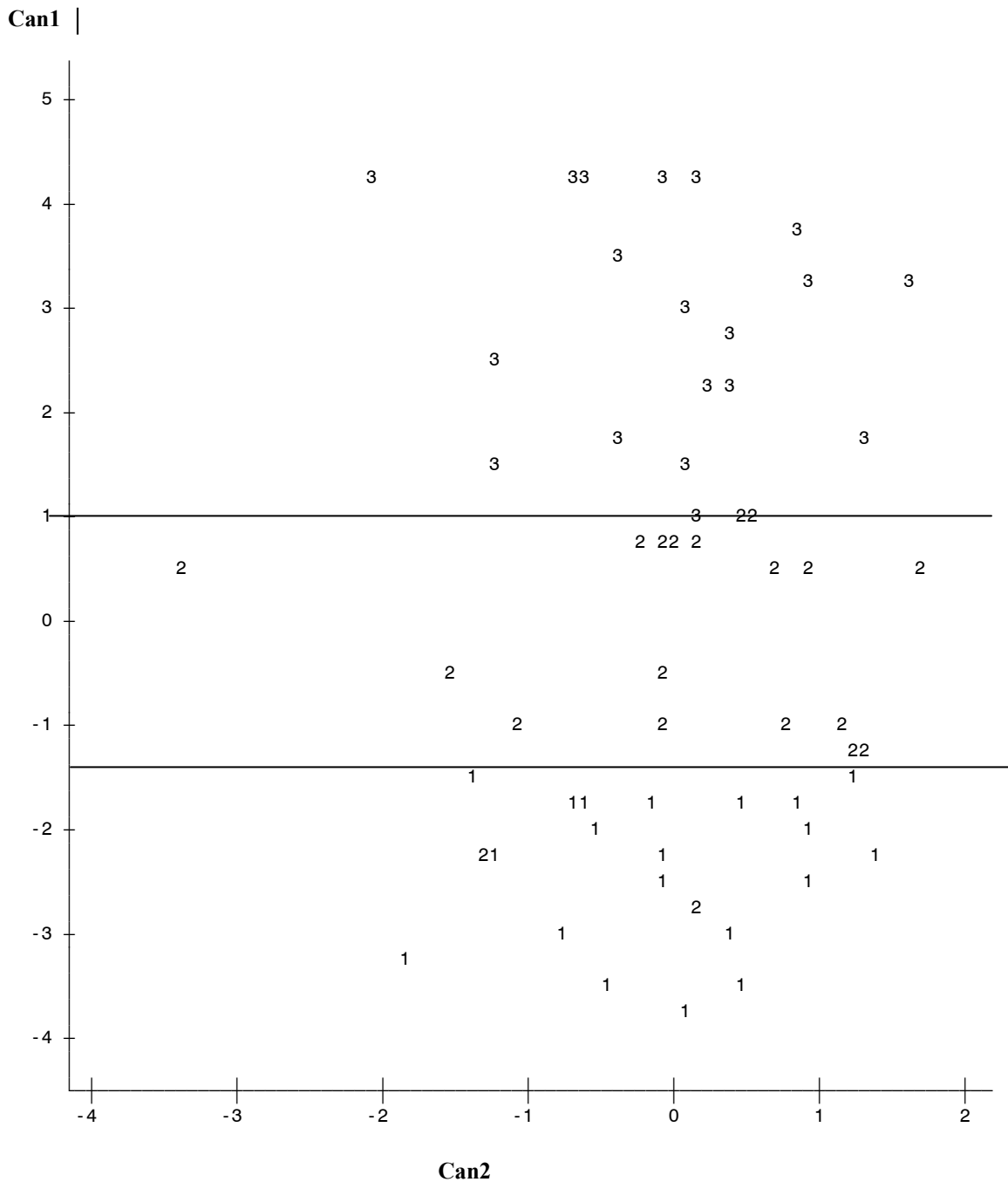
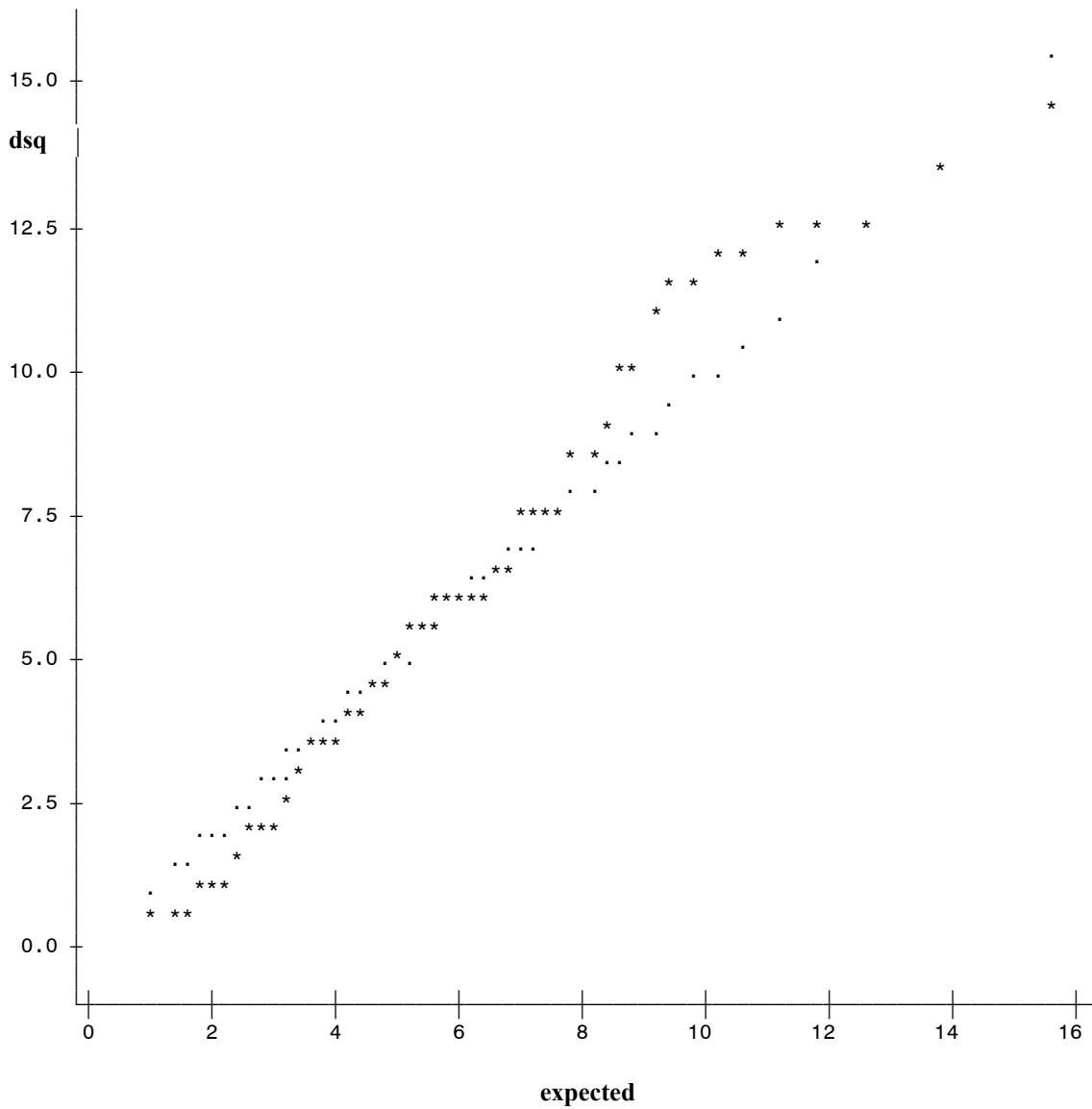


Figure 4.3. Scatter plot of the first two principle components. The categories 1, 2 and 3 represent immature, mature and over mature groups respectively. Prin1 is represents starch and firmness. Prin2 represents soluble solids.



NOTE: 1 obs hidden.

Figure 4.4. Scatter plot of the first two canonical variables after the categorization of the three clusters based on Fig. 4.1 into three maturity categories. The categories 1, 2 and 3 represent immature, mature and over mature groups, respectively. Can1 is represents starch and firmness. Can2 represent primarily by firmness.



NOTE: 34 obs hidden.

Figure 4.5. Q-Q plot to examine graphically if the observations are from a multivariate normal distribution. Data were obtained with electronic nose. The X-axis (expected) represents Chi-Square Quantile. The Y-axis (dsq) indicates the mahalanobis D-square distance.

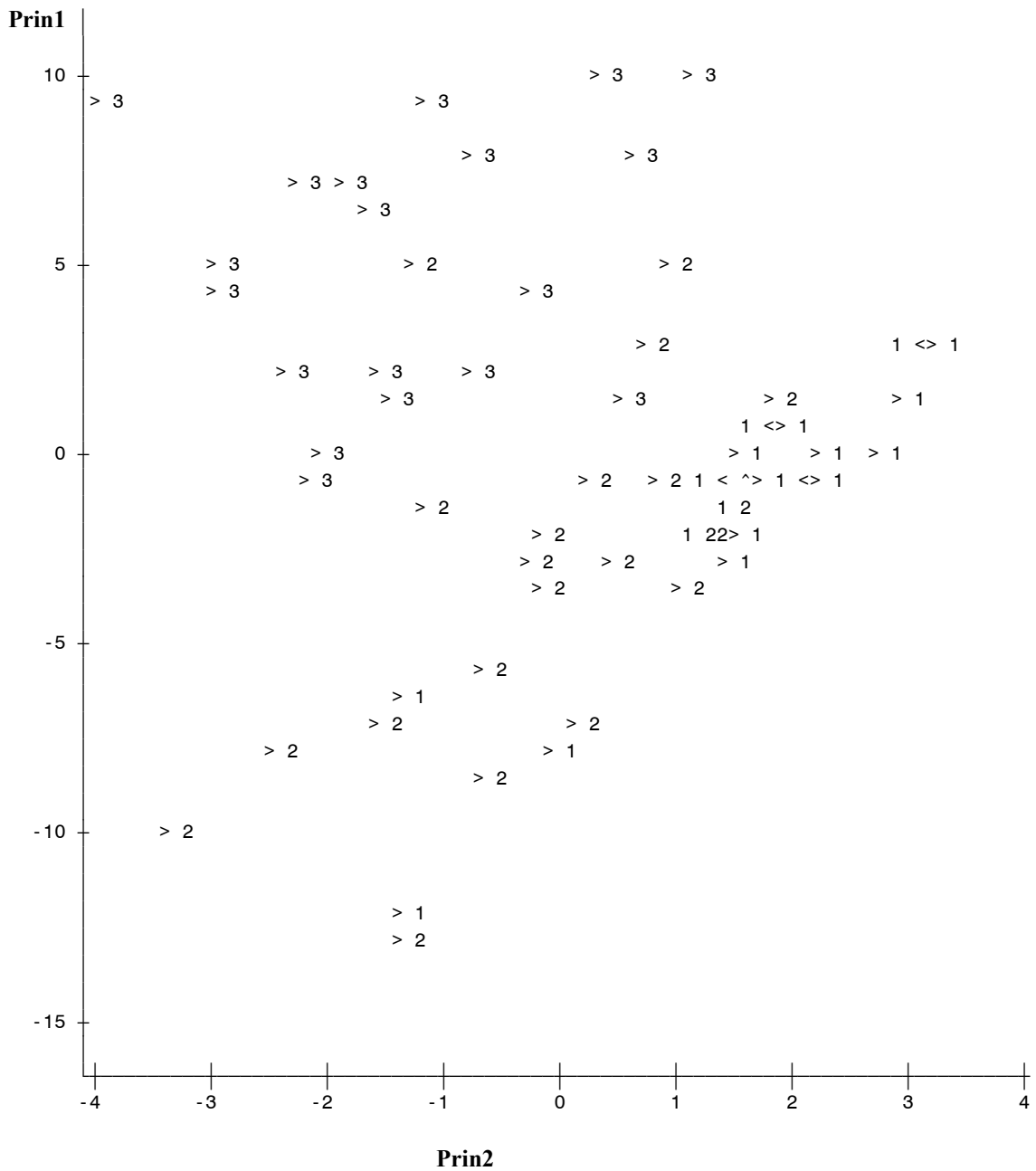
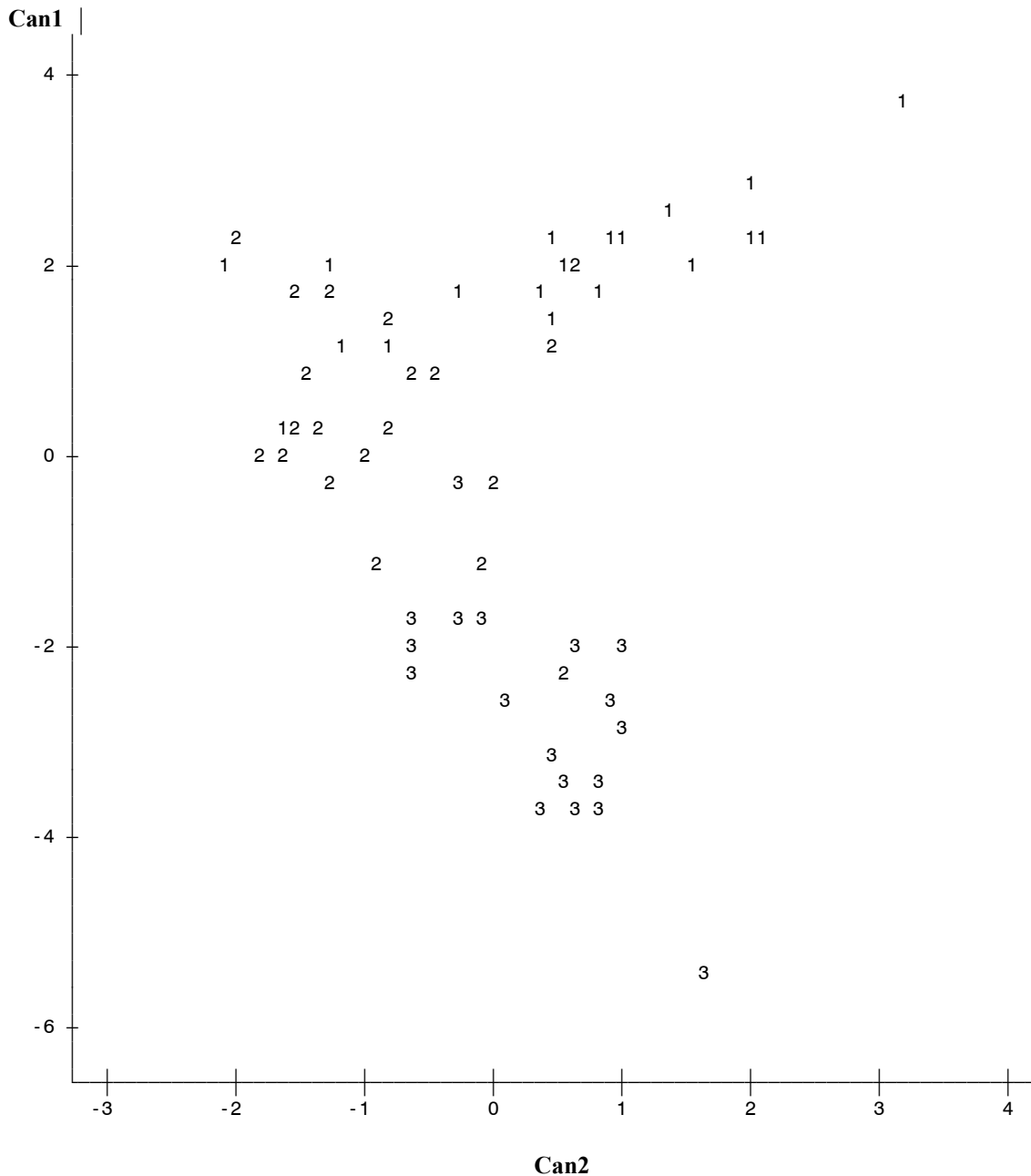


Figure 4.6. Scatter plot of first two principle components after the Electronic nose sensor response data were categorized into three maturity groups (based on Fig 4.2). The categories 1, 2 and 3 represent immature, mature and over mature groups, respectively. Prin1 represents the response of all sensors. Prin2 represents most of the sensors responses.



NOTE: 3 obs hidden.

Figure 4.7. Scatter plot of the first two canonical variables after the EN sensor response data into the three maturity categories. The categories 1, 2 and 3 represent immature, mature and over mature groups respectively. Can1 represents the response of all sensors. Can2 represents most of the sensors responses.

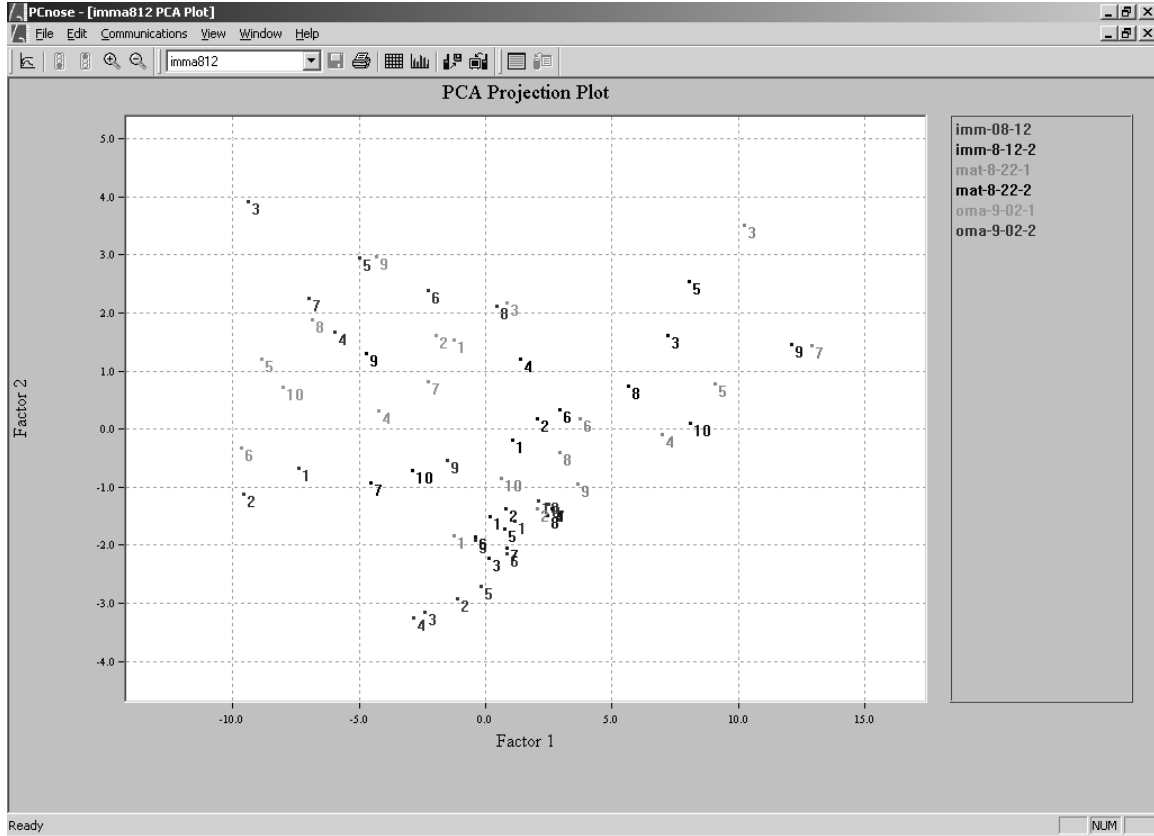


Figure 4.8. Scatter plot of the first two factors. PCA Plot generated by the EN system software analysis after the sensors response data were grouped into three maturity groups with 57 observations. The primary use of the PCA by the software is to detect outliers. No information about the axis is available in the software. The first two classes represent immature, next two classes represents mature and the last two classes over mature groups. 1-10 represents En sensor responses in a particular class.

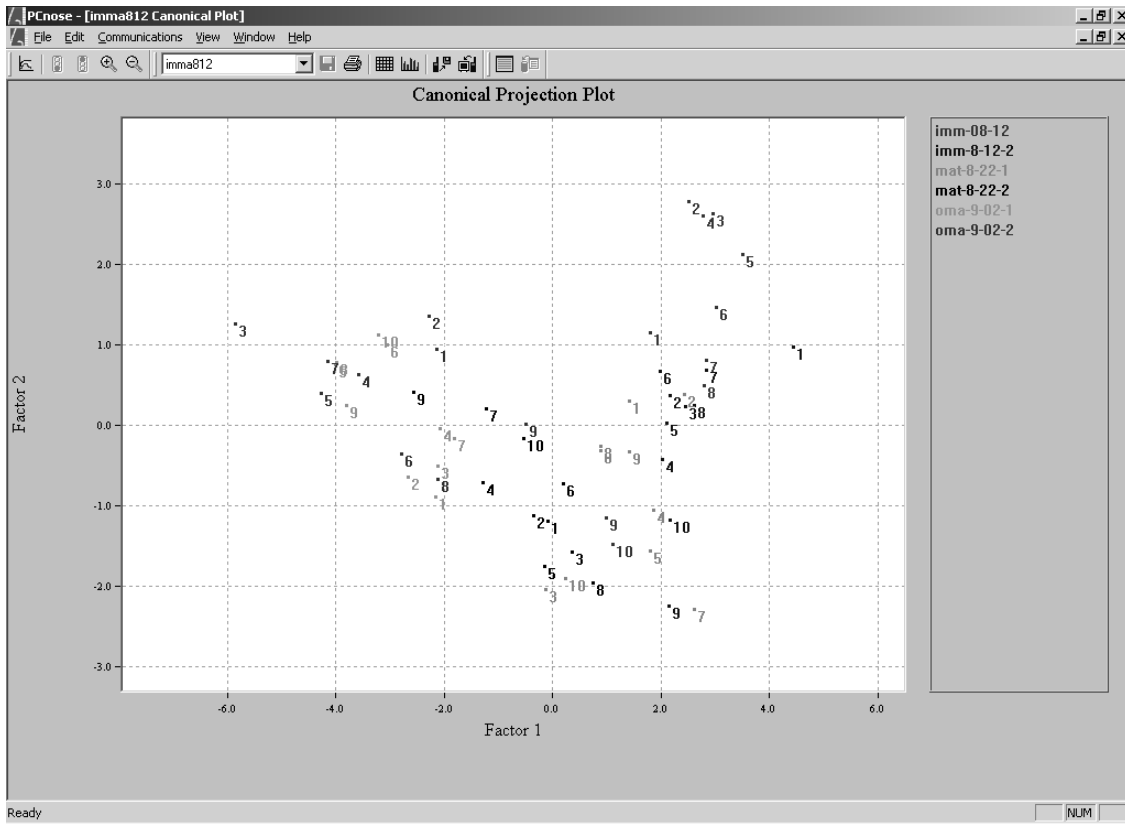
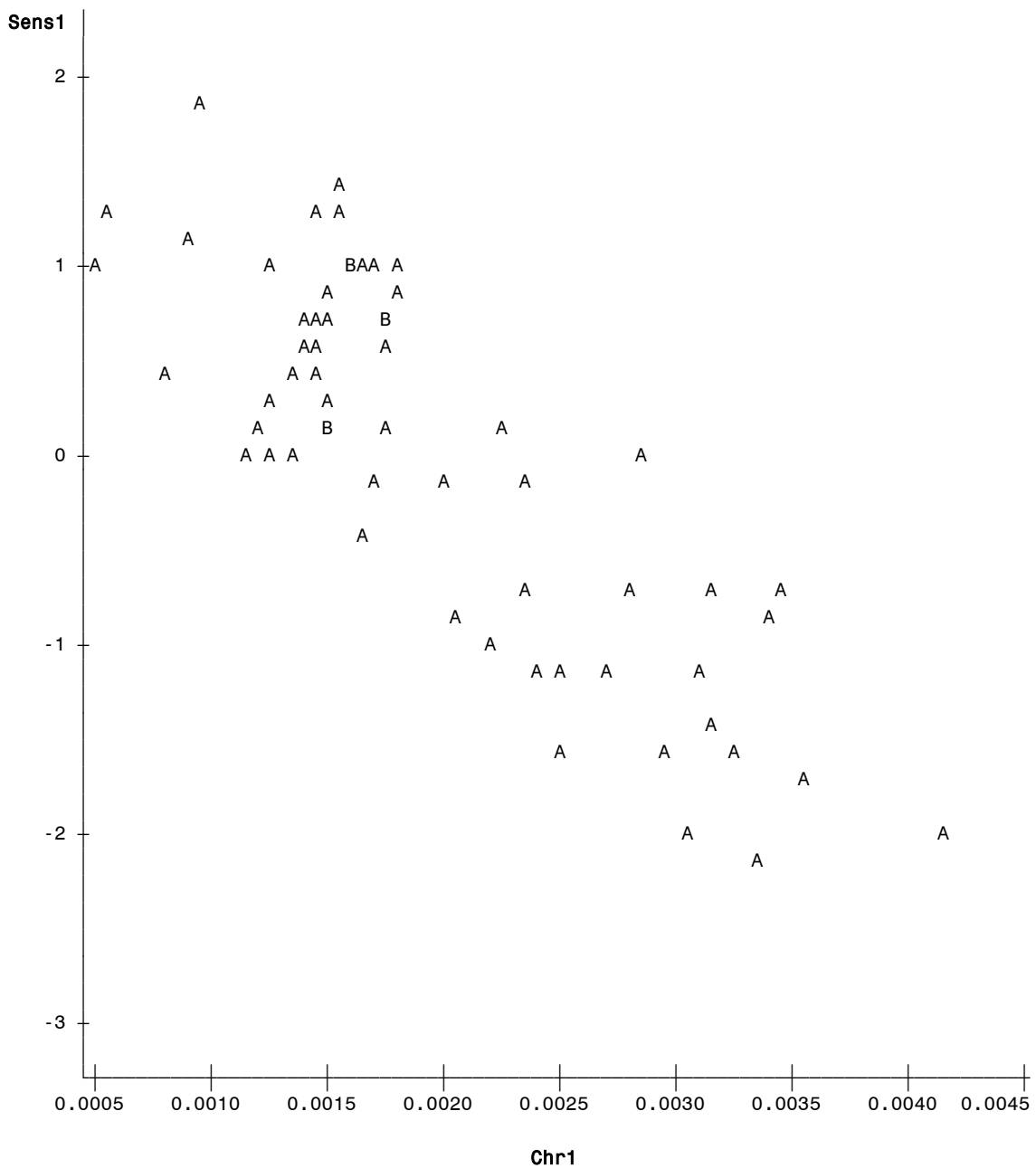


Figure 4.9. EN system software 3D Canonical projection plots. Three clusters are supposed to be obtained from the above CDA plot. No information about the axis is available in the software.



Legend: A = 1 obs, B = 2 obs, etc.

Figure 4.10. CCA plot -1, illustrates a linear relationship between the first canonical function (Chr1) of the maturity indices data and the first canonical function (Sens1) of the sensor response data.

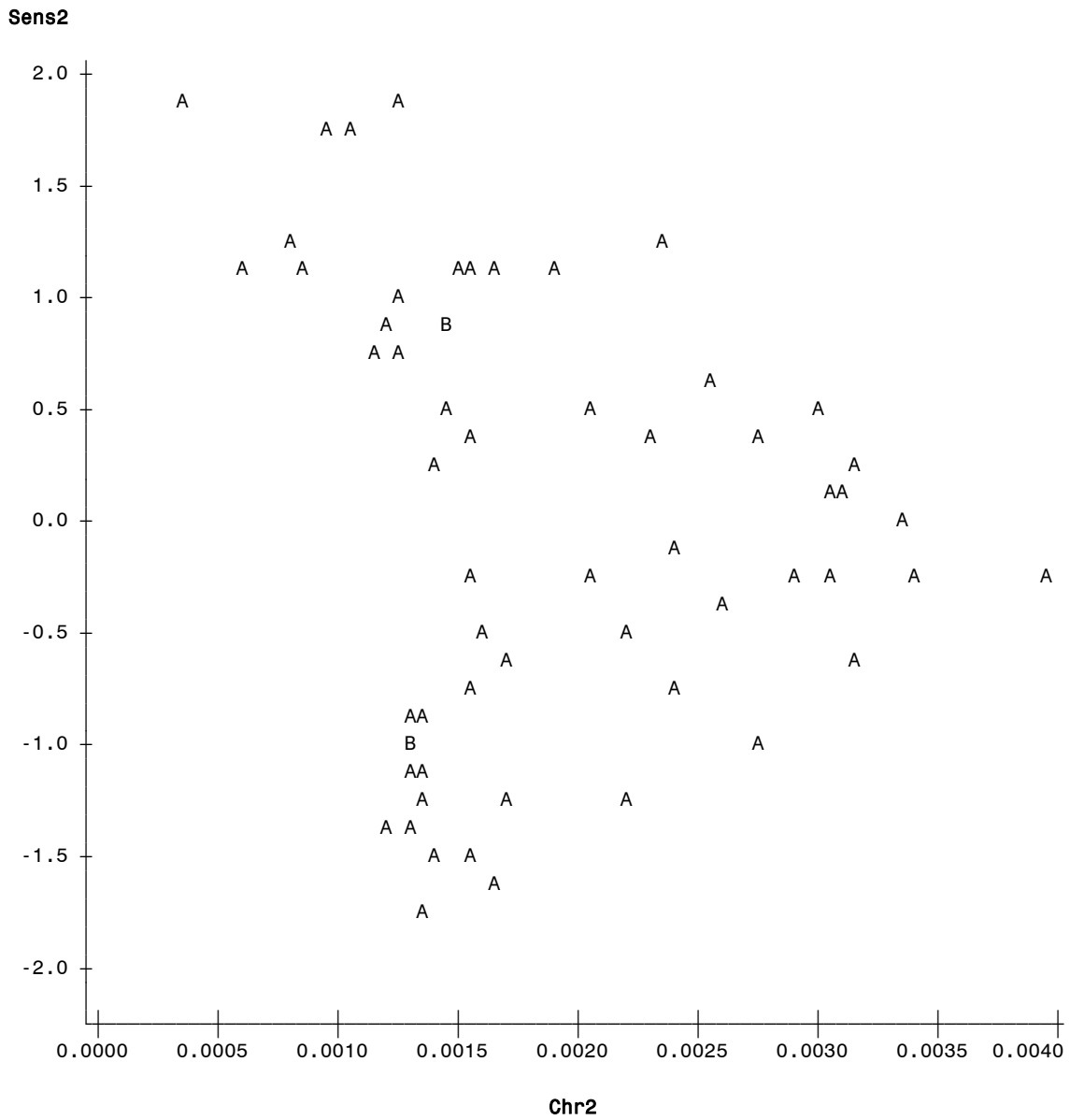


Figure 4.11. CCA plot -2, illustrates almost poor relationship between the second canonical function (Chr2) of the maturity indices data and the second canonical function (Sens2) of the sensor response data.

Chapter 5 - York– Results and Discussions

5.1. Introduction

Starch, firmness, soluble solids, color and ethylene were measured and the mean values are presented in Table 5.1. The ethylene concentrations were in the expected range; mean concentrations levels 0.09ml/kg for 24 hours and 0.15ml/kg for 24 hours (Table 5.1) for the first and last harvest date respectively. York data analysis was similar to that of data Gala analysis, only differences were that instead of puncture test values, firmness values were used and that the pH analysis was not performed. The experimental results of the four maturity indices except color are listed in Appendix-B. Starch and soluble solids mean values increased, flesh firmness values decreased, from first to fourth harvest date. Though the color was measured for all 75 apples, they were not used in the subsequent analysis due to a calibration error in the equipment (Appendix-C). The notable differences were that, instead of puncture test values, firmness values were used and that the pH analysis was not performed. The experimental results of the four maturity indices except color are listed in Appendix-A. Starch and soluble solids mean values increased, flesh firmness values decreased, from first to fourth harvest date.

5.2. York- Statistical Analysis of Maturity Indices

5.2.1. Exploratory analysis

The York apples were numbered from 1-75 (only 15 samples were used from first harvest date). PCA was performed on the numbered data and the maturity indices were the four variables. From the PCA plot (Fig. 5.1), three categories were identified and

were named immature, mature and over mature (Fig. 5.2). Each group had 31, 27 and 17 observations, respectively.

PCA was performed on the classified data (based on maturity indices) to understand the underlying principles of variation. From the correlation matrix (Appendix-D), firmness and starch were modestly correlated ($r = -0.51$). The next best correlation was between soluble solids and firmness ($r = 0.30$), followed by soluble solids and starch ($r = -0.36$). From PCA, the variation in Prin1 was mostly accounted by starch, firmness and soluble solids. Prin2 variation was mainly accounted by soluble solids. Thus the Prin1 axis could be interpreted as starch, firmness and soluble solids variation scale and Prin2 axis can be interpreted as the soluble solids variation scale.

The Q-Q plot (Fig. 5.3) indicates that the observations are from a multivariate normal distribution. MANOVA indicates that the maturity groups were not equal (Wilks' Lambda $F = 32.19$, $P < 0.0001$). Thus it can be concluded that there was a significant effect of the maturity on the maturity indices.

5.2.2. Inferential analysis

From the PCA plot (Fig. 5.2) it can be inferred that immature apples (represented as 1) had low starch ratings and soluble solids but high firmness values. Mature apples (represented as 2) had intermediate values of starch, firmness and soluble solids. Over mature apples (represented as 3) had high starch ratings and soluble solids with low firmness values.

The first three principal components from the PCA were used as input for CDA and DA. From CDA, since the first eigenvalue explained as much as 81.5% of the total variation, it can be used as the principal axis along which most of the information resides. From the CDA, starch, firmness and soluble solids variation can be interpreted along the Can1 axis. Can2 represents firmness and starch variation. Fig. 5.4 (CDA Plot) illustrates three clusters can be classified along the Can1 axis, thus verifying the inference drawn from the PCA. From the DA the cross validation results (Table 5.2), the classification was 95% correct and their interclass distance is given in Table 5.3. Separation between the three groups was very high (Table 5.3).

5.3. Statistical Analysis of Electronic Nose Sensor Response Data

5.3.1. Exploratory analysis

Statistical analysis on the electronic nose sensor data was performed to determine the efficiency of the electronic nose system software to categorize the apples into the maturity categories. PCA was performed on these data to identify outliers and to determine any clusters that would represent the grouping structure (maturity) in the data. Most of the sensors were highly correlated ($r > 0.95$) to each other and their contribution to prin1 was almost equal, thus each sensor contributed equally to the variation in data, confirmed the non specificity of these sensors. From the eigenvalues the first four principal components contributed to 99.7% of the total variation, so these four components were selected as the input data for CDA and DA. The PCA plot (Fig. 5.5), illustrates that there were three main clusters. Fig 5.6 indicates the data were multivariate normal. MANOVA was also performed to determine if indeed there was a maturity

effect. The Wilks' Lambda F value was 9.92 with a P value <0.0001 , so there is a significant effect of maturity on the sensor responses.

5.3.2. Inferential analysis

From the PCA, only four principal components were used as input for CDA and DA. From CDA, the first two eigenvalues contributed to as much as 100%, so the first and second canonical component (Can1 and Can2) explained all the variation. Prin1 and Prin2 accounted for most of the variation in Can1, and Prin2 and Prin3 accounted for most of the variation in Can2. Inference about what they represent in terms of sensors could not be deduced due to the non-specificity of the sensors. Fig. 5.7 (CDA Plot) illustrates three clusters or groups, thus confirming the inference drawn from the PCA. From the DA cross validation results (Table 5.4), the classification was approximately 69% correct. The interclass separation between the three groups using an EN sensors response was less than the interclass separation between the three groups, generated using maturity indices (Fig. 5.5). Since the interclass separation distance was less than 5, we could conclude that the separation was fair.

5.4. Electronic Nose System Software Analysis

The same maturity based electronic nose sensor data that was earlier used in the above statistical analysis was also used as the input for the electronic nose system software. York data consisted of 75 observations, but due to the previously mentioned constraint in the electronic nose software the first 20 observations were from the immature group, the second 20 observations were from the mature group and the final 17

observations were from the over mature group were selected, amounting to a total of 57 observations. These observations were divided into 6 classes with each maturity category being represented by two consecutive classes in the electronic nose system software. The analysis was performed using the Canonical algorithm. The results are presented in following Fig. 5.8 and 5.9, respectively. From the EN cross validation results (Table. 5.6), the classification was approximately 81 % correct. To duplicate the analysis, the same data (57 observations and 6 classes) were used as the input for the statistical DA analysis, and the correct classification percentage was approximately 83%.

5.5. Canonical Correlation Analysis

The CCA was performed to investigate if there were strong correlations between the sensors and the maturity indices. All sensors were highly correlated to starch ($r > 0.65$) and firmness ($r < -0.5$) and modestly correlated to soluble solids ($r < 0.33$). From the likelihood ratio test for statistical significance, only two canonical correlation were statistically significant, so the NCAN=2 option was used for CCA. The first canonical function or variable (Chr1) created for the maturity group was highly correlated to starch ($r = 0.96$) and modestly correlated to firmness ($r = -0.67$) and soluble solids ($r = 0.57$). And Chr2 was highly correlated to soluble solids ($r = -0.83$). Thus, all sensors were highly correlated ($r > 0.68$) to Chr1 (represents starch, firmness and soluble solids). For Chr2 (represents soluble solids), there were no sensors that correlated ($r < 0.1$). The given CCA plot-1 (Fig. 5.10) shows a strong correlation between the sensors canonical function1 (Sen1) and the maturity canonical function1 (Chr1), indicating that all sensors in the electronic nose were linearly correlated to all three maturity indices But for CCA

plot-2 (Fig. 5.11) the correlation is almost zero between Sen2 and Chr2, indicating that non of the sensors were correlated to soluble solids

5.6. Analysis ethylene effect on the maturity classification

The above analyses (sect. 5.2) were rerun by including ethylene levels (adjusted levels), in addition to the three standard maturity indices. From the DA, classification percentage was reduced by only 4% from 95% to 91%. Thus it can be concluded that the ethylene could be a potential maturity index when it is appended with other standard maturity indices. From the correlation matrix, ethylene levels and soluble solids were modestly correlated ($r = 0.334$), the next best correlation was with starch ($r = 0.21$) and firmness ($r = -0.21$).

5.7. Harvest date Effect

Unlike Gala, York apples were harvested four times and therefore the given data was classified into four harvest groups 1, 2, 3, 4 respectively. Due to the electronic nose software constraint all the observation could not be considered for analysis. From each harvest date the first 10 observations were considered ($N = 40$). From the EN software cross validation analysis, the 40 observations were classified accurately (100% correct) into their respective categories. DA correctly classified the apples 70% correct.

5.8. Analysis with Reduced Number of Sensors

The moisture sensitive sensors 5, 6, 23 and 31, were removed and the analysis was rerun. From the statistical DA, the correct classification percentage was remained at

69%. This indicated that the moisture had little or no impact on the sensor response analysis. The stepwise (SLE = 0.4 and SLS = 0.15) method in the STEPDISC procedure of DA was used to reduce the number of sensors that were statistically significant. From the STEPDISC procedure, 7 sensors (2, 3, 8, 9, 16, 24, and 27) were retained in the model, removing the other non-significant 25 sensors. Further analysis was performed on these 7 variables (sensors), and from the statistical DA, the correct classification percentage was 80%, which was 11% higher than the classification percentage with all 32 sensors.

5.9. Summary

The data were separated into three categories (Fig. 5.1) based on starch, firmness, soluble solids and color. The PCA and CDA plots (Fig. 5.2 and 5.3) illustrate the three clear clusters. Prin1 axis can be interpreted as the variation scale for both starch and firmness values, while, Prin2 can be interpreted as color and soluble solids. These results support accepted fact that the starch and firmness values were more correlated than that of soluble solids and color to the maturity of the apple fruit. From the DA, the cross validation results were 95% correct. The three categories were named as immature, mature and over mature categories consisting of 31, 27 and 17 observations respectively.

This grouping was used to categorize the electronic nose sensor responses into three categories. The effectiveness of the sensor responses classification percentage was 69% correct. For the EN software the classification percentage (6-groups) was 81%. The correct classification percentage was 83% for DA (SAS).

From CCA, all sensors (Sens1) were significantly correlated with the maturity indices. As a result there were no specific sensors that could be correlated with the maturity indices. Appending ethylene to other maturity indices before performing statistical analysis did not improve the classification percentage. Electronic nose could effectively classify 100% correct based on harvest date, but DA was correct by only 70%. By removing the moisture sensitive sensors, the classification percentage was not significantly changed. The reduced sensors after the STEPDISC procedure were 2, 3, 8, 9, 16, 24, and 27.

Table 5.1. Mean values (with standard deviation) of maturity indices for York apples harvested on three dates.

Harvest date	Firmness (N)	Soluble solids (°Brix)	Starch index (1-9)	Ethylene (ml /kg of fruit for 26 hr)
09/22/02	95.8±7.2	11.7±0.7	1.5±0.6	0.09±0.7
10/07/02	90.5±12.8	13.4±1.3	2.0±0.9	0.12±0.06
10/22/02	78.3±7.1	13.3±0.9	3.0±1.3	0.12±0.04
11/02/02	77.6±6.5	14.0±1.2	4.5±1.6	0.15±0.05

Table 5.2 Classification table obtained after performing DA for York apples. Columns and rows represent the maturity categories (1, 2, and 3). The values represent the number of fruits placed in that category. The groups were categorized based on maturity indices.

Actual Maturity Group	Predicted classification group		
	1	2	3
1	29	2	0
2	1	26	0
3	0	1	16

Table-5.3. Mahalanobis distance to determine the similarity of groups of York apples with unknown maturity to a set of values from 3 groups of apples based on maturity indices. The left column indicates the groups and the right column indicates the distance between them.

From Maturity	Mahalanobis Distance
1-2	7.7
1-3	26.4
2-3	7.7

Table 5.4. Classification table obtained after performing DA for York apples. Columns and rows represent the maturity categories (1, 2, and 3). The values represent the number of fruits placed in that category. The categorization was performed on EN sensor response data.

Actual Maturity Group	Predicted classification group		
	1	2	3
1	26	5	0
2	9	15	5
3	0	4	13

Table 5.5. Mahalanobis distance to determine the similarity of groups of Gala apples with unknown maturity to a set of values from 3 groups of apples based on maturity indices. The left column indicates the groups and the right column indicates the distance between them.

From Maturity	Mahalanobis Distance
1-2	1.9
1-3	11.9
2-3	6.8

Table 5.6. Classification table obtained from the EN software analysis for York apples. Columns and rows represent the maturity categories (1, 2, and 3). The values represent the number of fruits placed in that category. The categorization was performed on EN sensor response data.

Actual Maturity Group	Predicted classification group		
	1	2	3
1	16	4	0
2	5	14	1
3	0	4	13

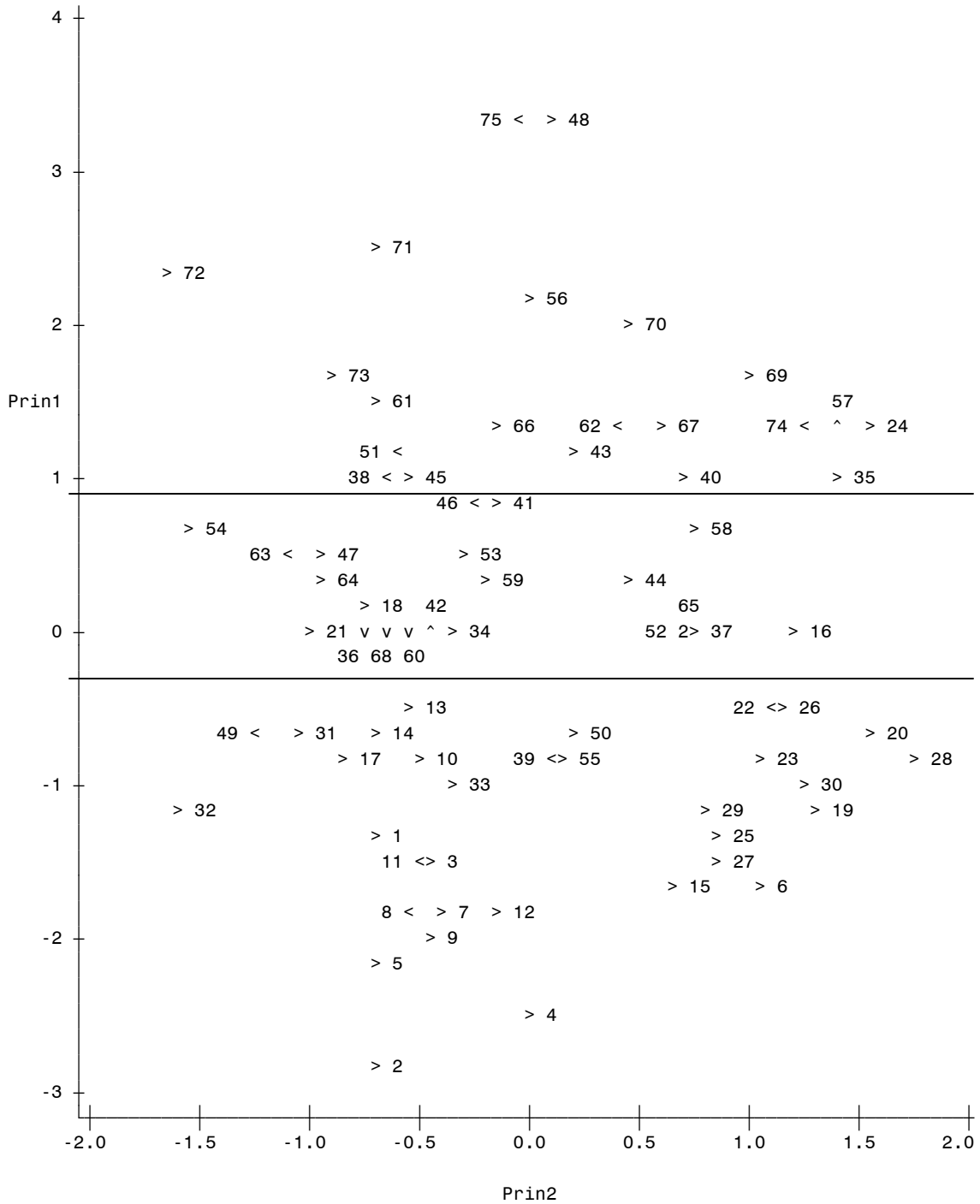


Figure 5.1. Scatter plot of the first two principle components illustrating three clusters.

The 1-60 represents the observations. Prin1 is represents starch, firmness and soluble solids variation. Prin2 represents primarily represents soluble solids.

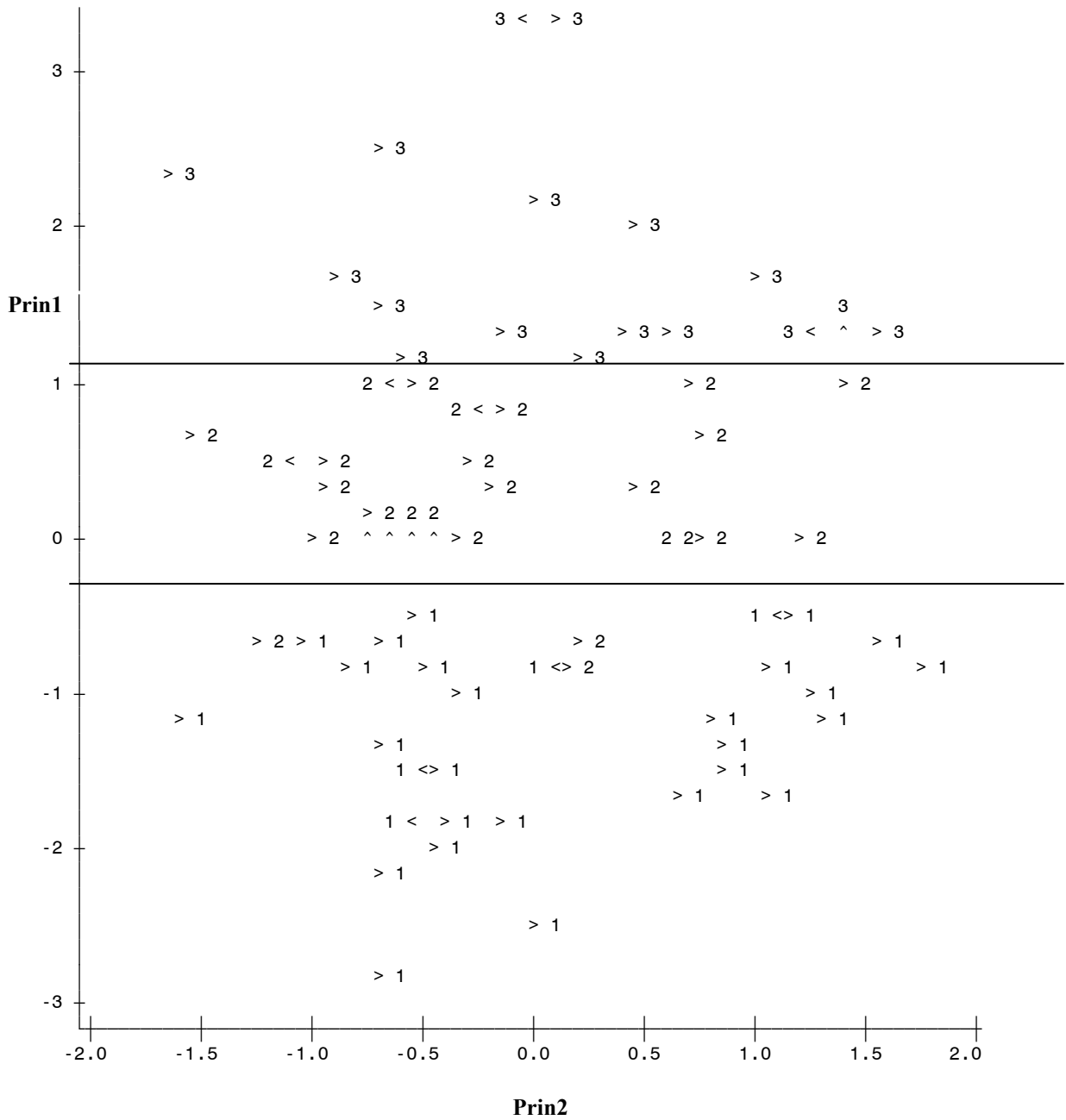
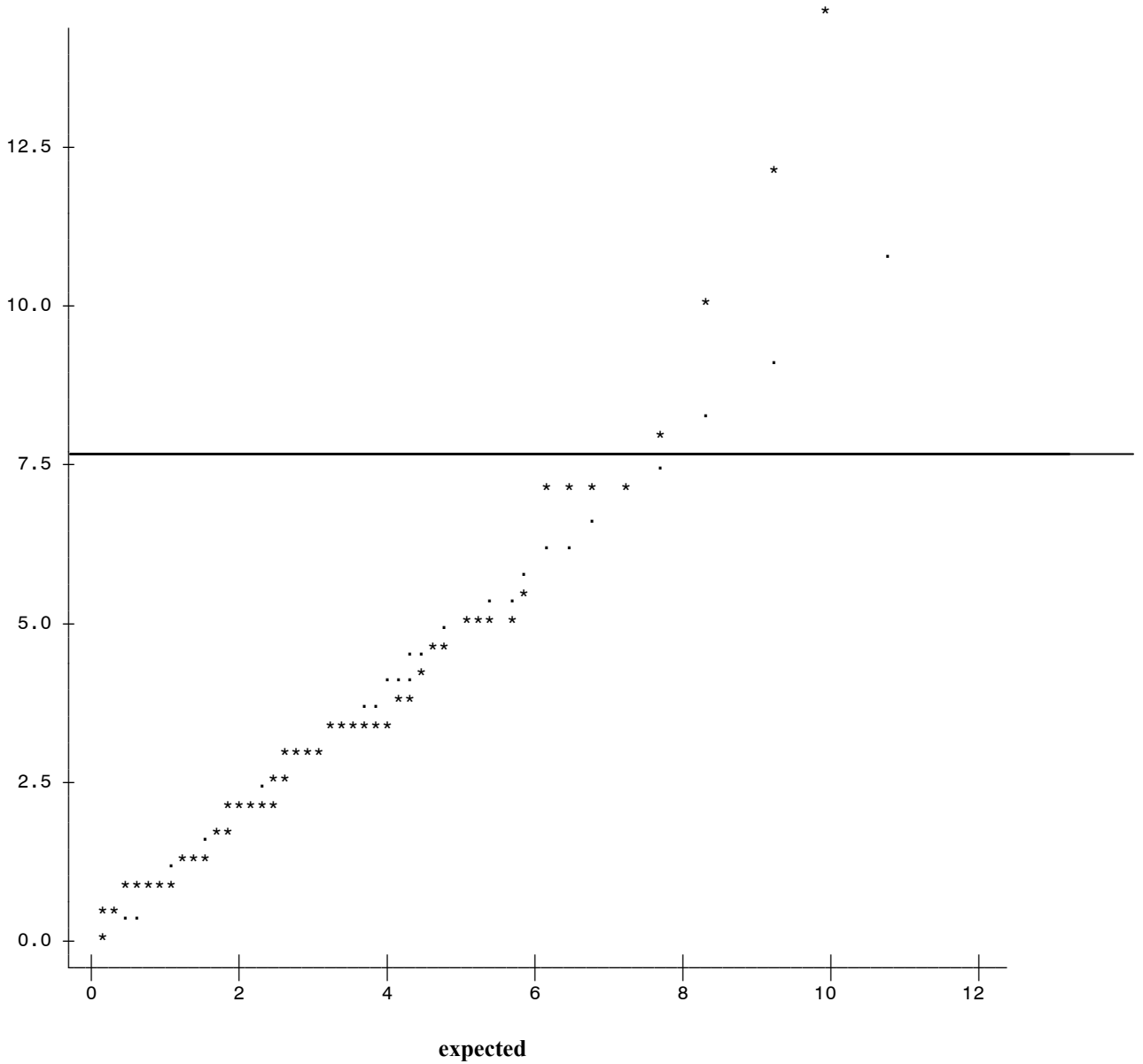


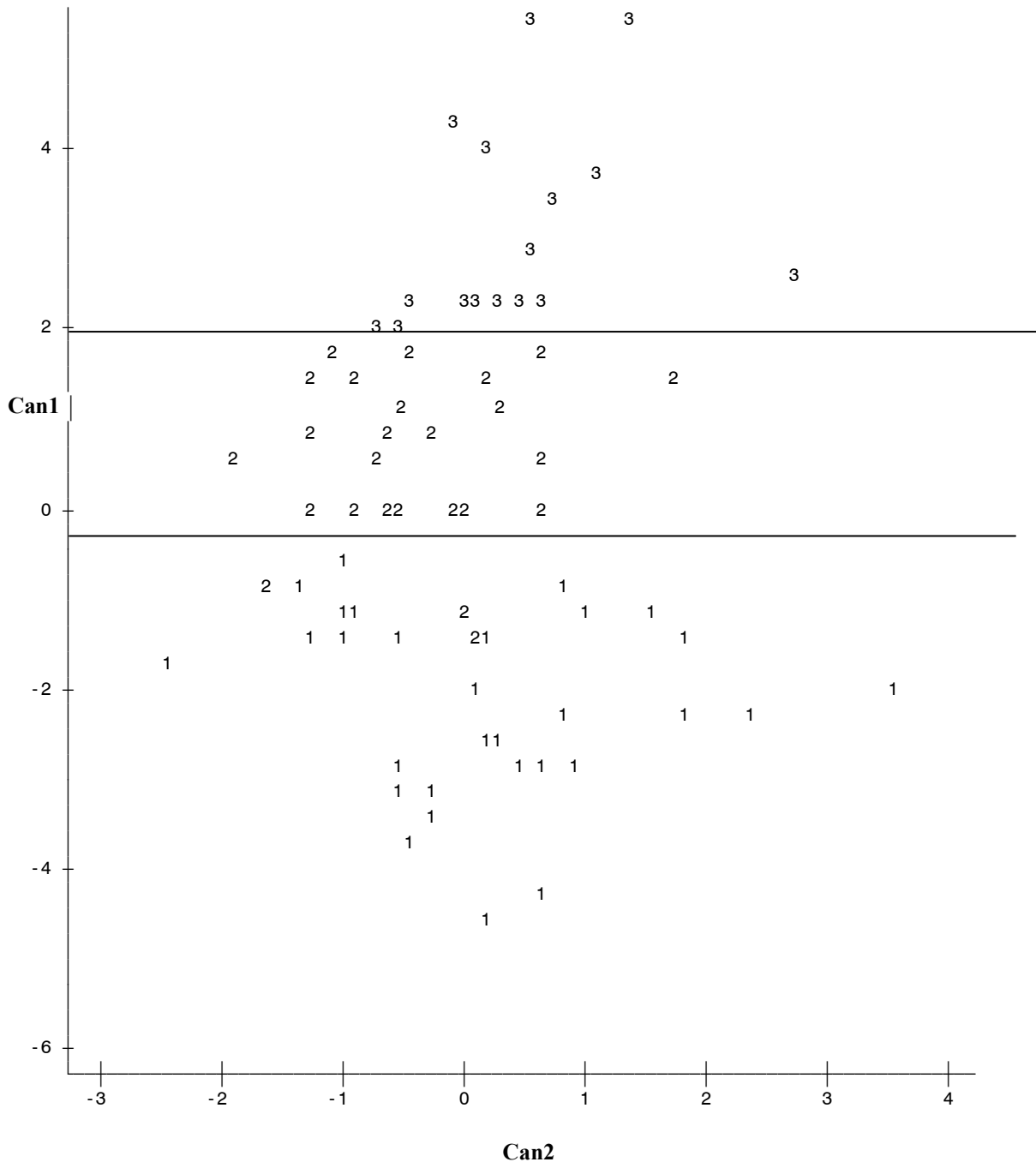
Figure-5.2. Scatter plot of the first two principle components. The categories 1, 2 and 3 represent immature, mature and over mature groups respectively. Prin1 is represents starch, firmness and soluble solids variation. Prin2 represents primarily represents soluble solids.

plot of dsq*expected. Symbol used is '*'.
Plot of expected*expected. Symbol used is '.'.



NOTE: 79 obs hidden.

Figure 5.3. Chi-Squared probability plot for multivariate normal. The X-axis (expected) represents Chi-Square Quantile. The Y-axis (dsq) indicates the mahalanobis D-square distance.



NOTE: 4 obs hidden.

Figure 5.4. Scatter plot of the first two canonical variables after the categorization of the three clusters based on Fig. 5.1 into three maturity categories. The categories 1, 2 and 3 represent immature, mature and over mature groups, respectively. Can1 represents starch, firmness and soluble solids variation. Can2 represents firmness and starch.

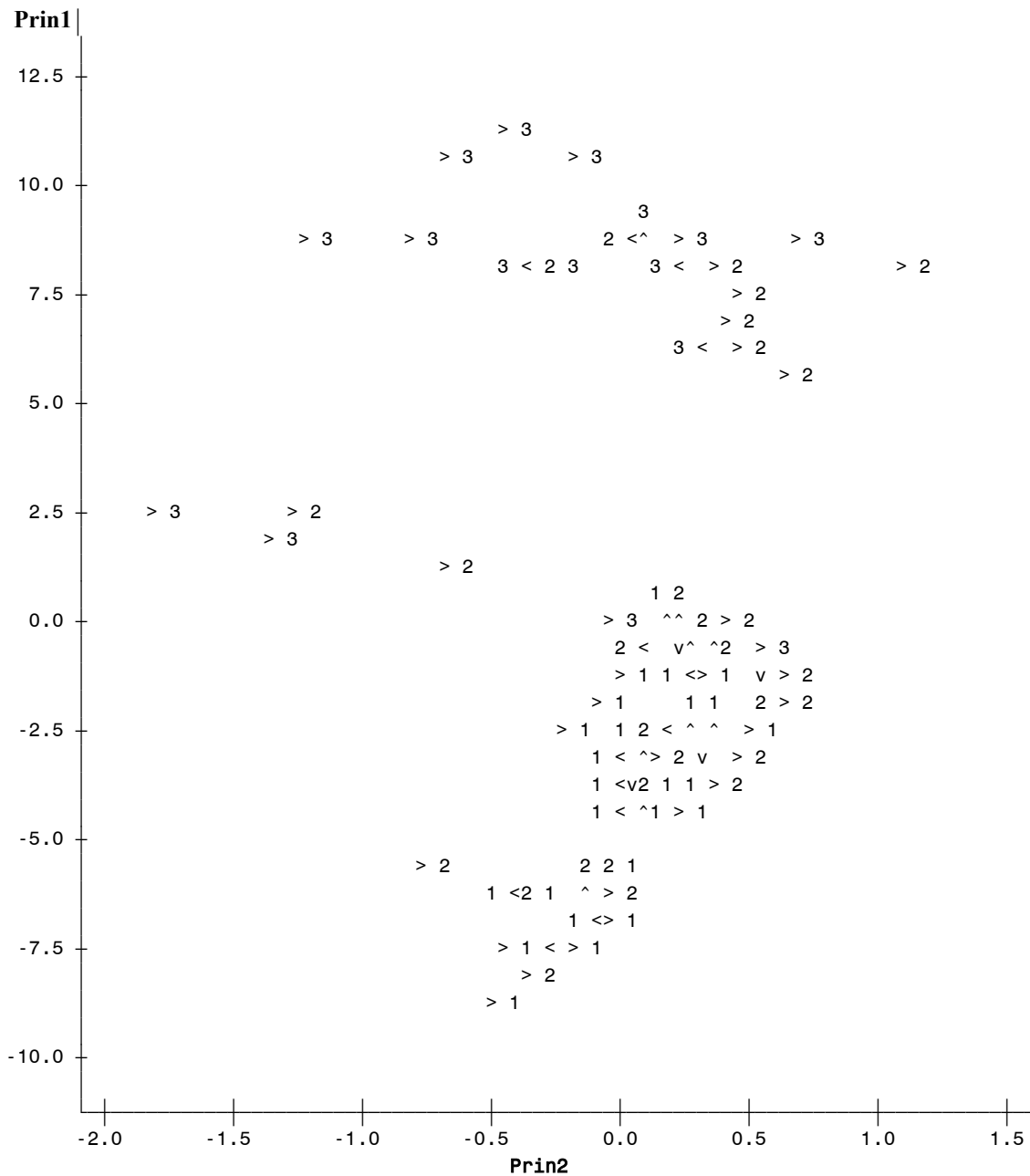
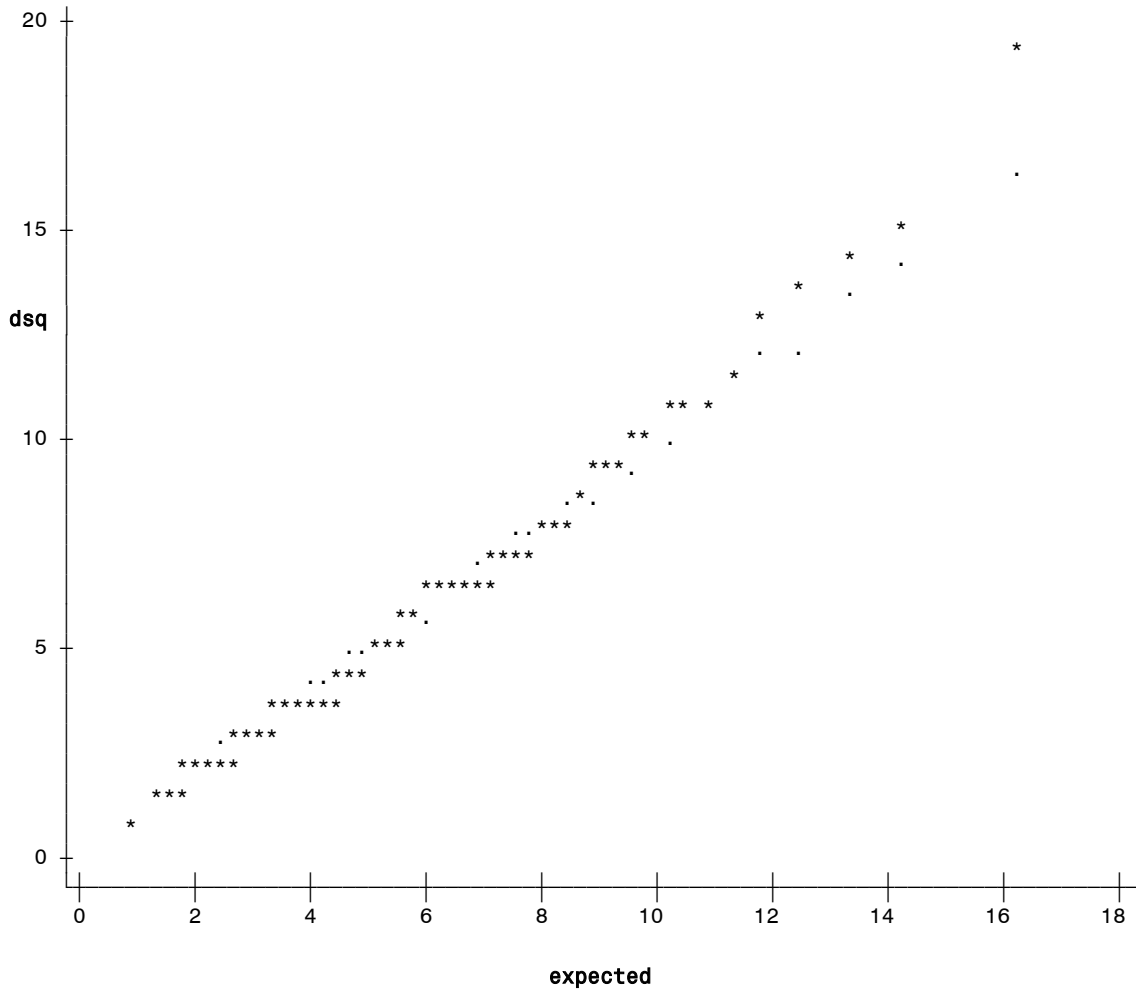


Figure 5.5. Scatter plot of first two principle components after the Electronic nose sensor response data were categorized into three maturity groups (based on Fig 5.2). The categories 1, 2 and 3 represent immature, mature and over mature groups, respectively. Prin1 represents the response of all sensors. Prin2 represents most of the sensors responses.



NOTE: 77 obs hidden.

Figure 5.6. Q-Q plot to examine graphically if the observations are from a multivariate normal distribution. Data were obtained with electronic nose. The X-axis (expected) represents Chi-Square Quantile. The Y-axis (dsq) indicates the mahalanobis D-square distance.

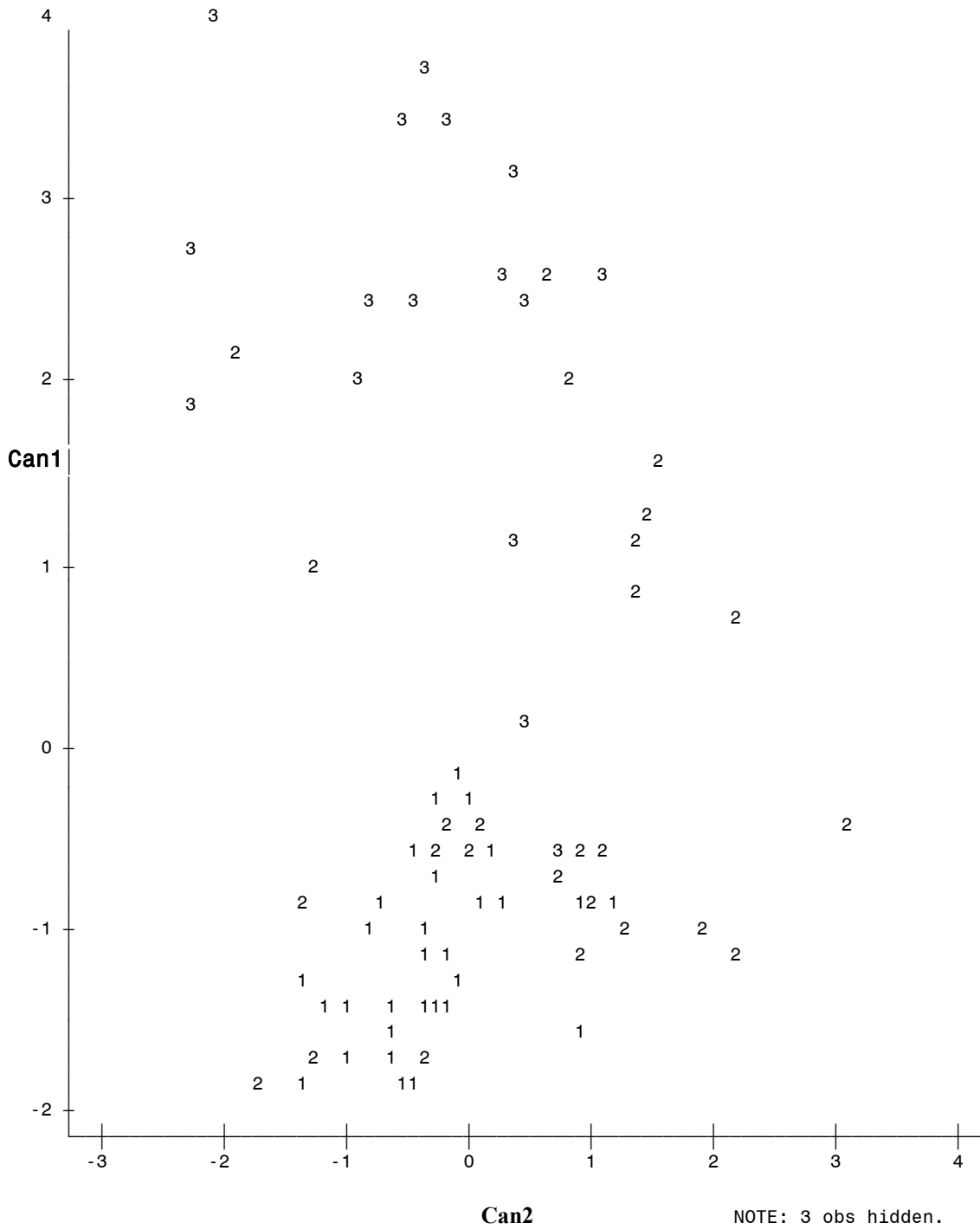


Figure 5.7. Scatter plot of the first two canonical variables. CDA plot after the EN sensor response data were grouped into the three maturity categories. The categories 1, 2 and 3 represent immature, mature and over mature groups respectively. Can1 represents the response of all sensors. Can2 represents most of the sensors responses.

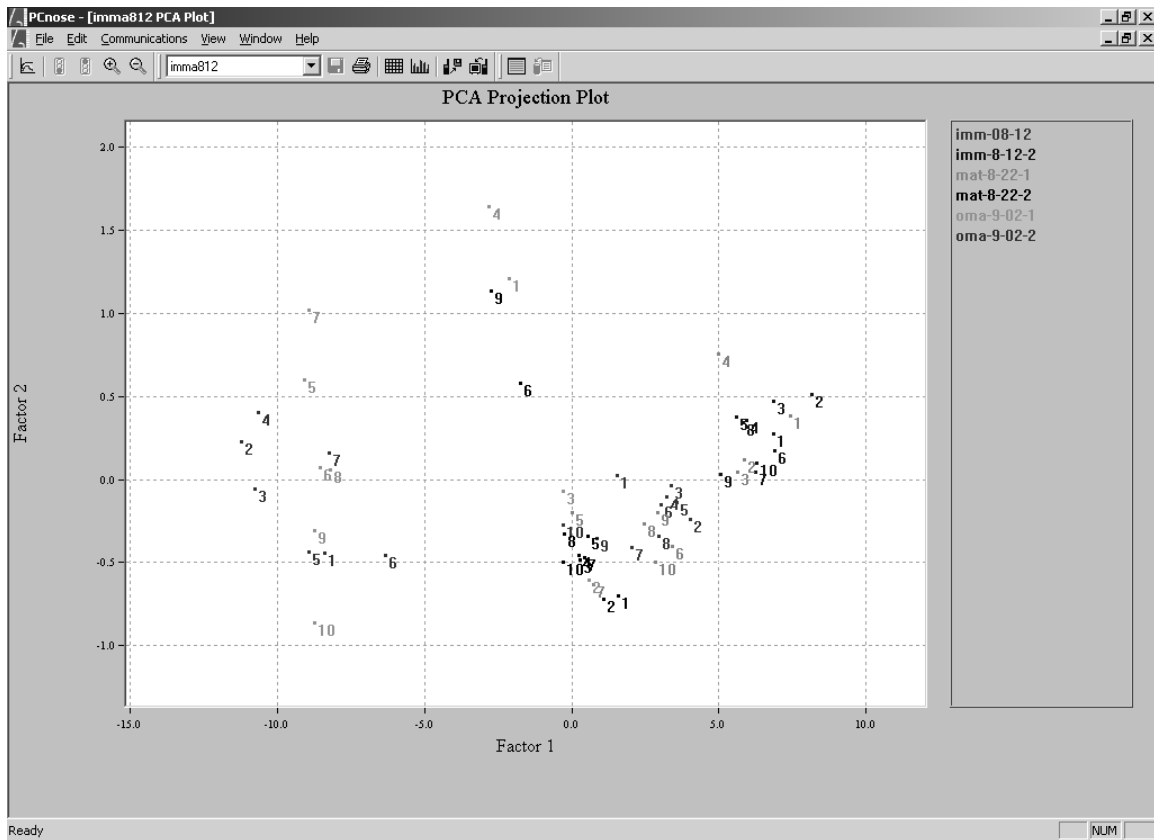


Figure 5.8. Scatter plot of the first two factors. PCA Plot generated by the EN system software analysis after the sensors response data were grouped into three maturity groups with 57 observations. The primary use of the PCA by the software is to detect outliers. No information about the axis is available in the software. The first two classes represent immature, next two classes represents mature and the last two classes over mature groups. 1-10 represents En sensor responses in a particular class.

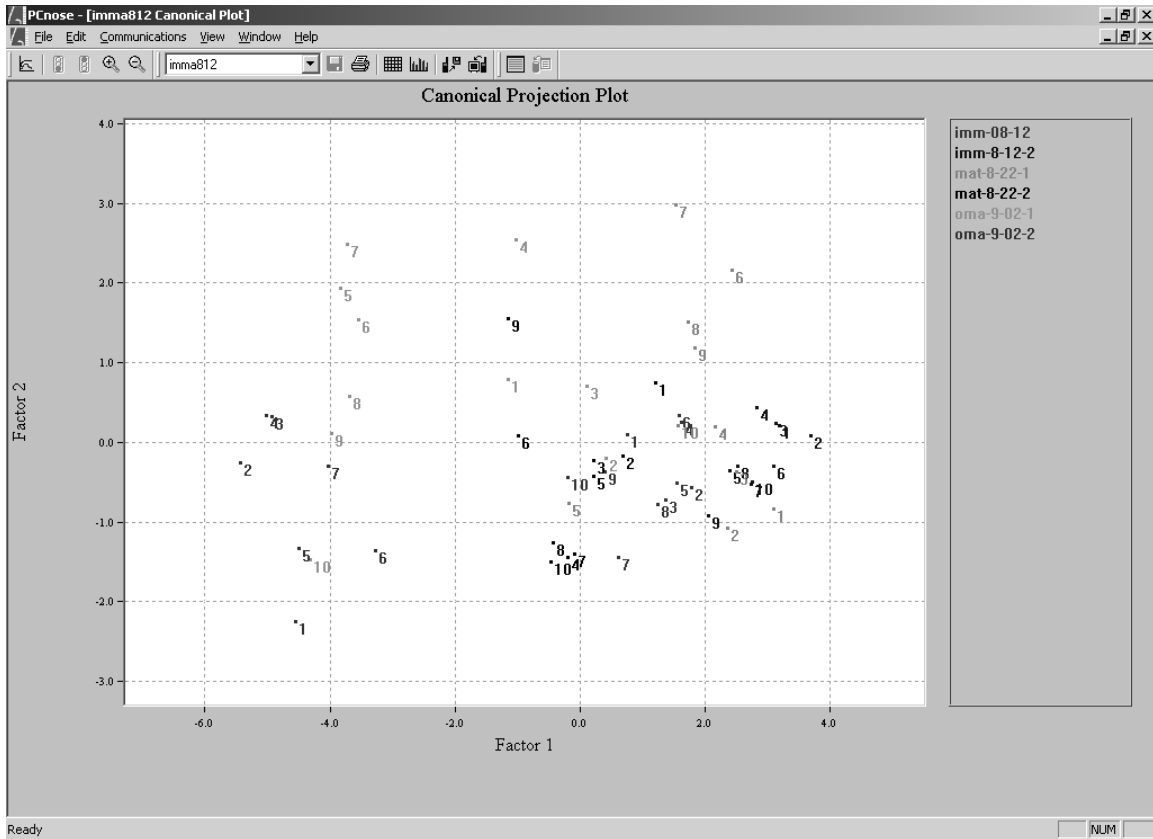


Figure 5.9. Scatter plot of the first two factors. CDA Plot generated by the EN system software analysis after the sensors response data were grouped into three maturity groups with 57 observations. No information about the axis is available in the software. The first two classes represent immature, next two classes represents mature and the last two classes over mature groups. 1-10 represents En sensor responses in a particular class.

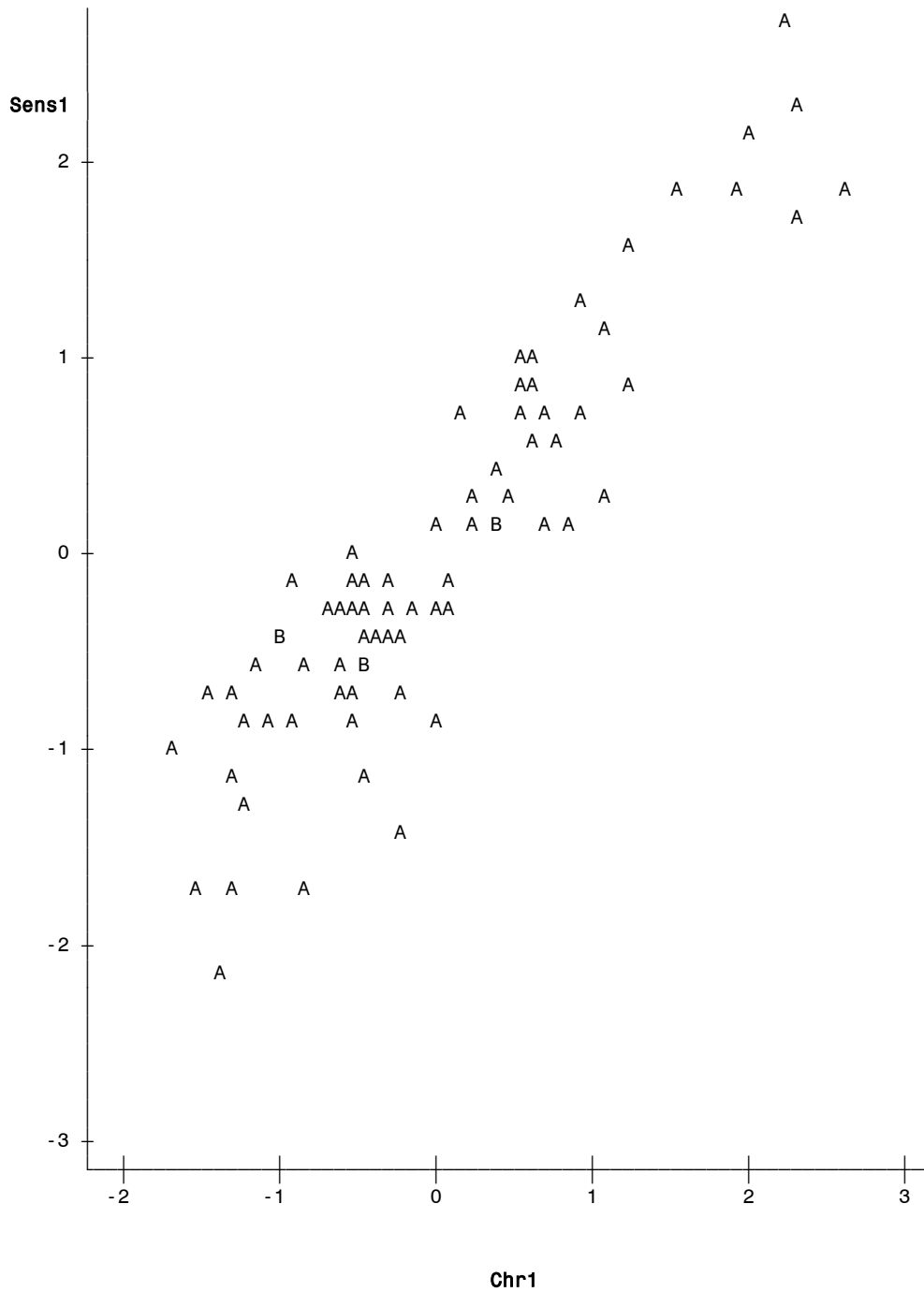


Figure 5.10. CCA plot -1, illustrates a linear relationship between the first canonical function (Chr1) of the maturity indices data and the first canonical function (Sens1) of the sensor response data. Sens1 represents all the sensors, while Chr1 represents starch, firmness and soluble solids.

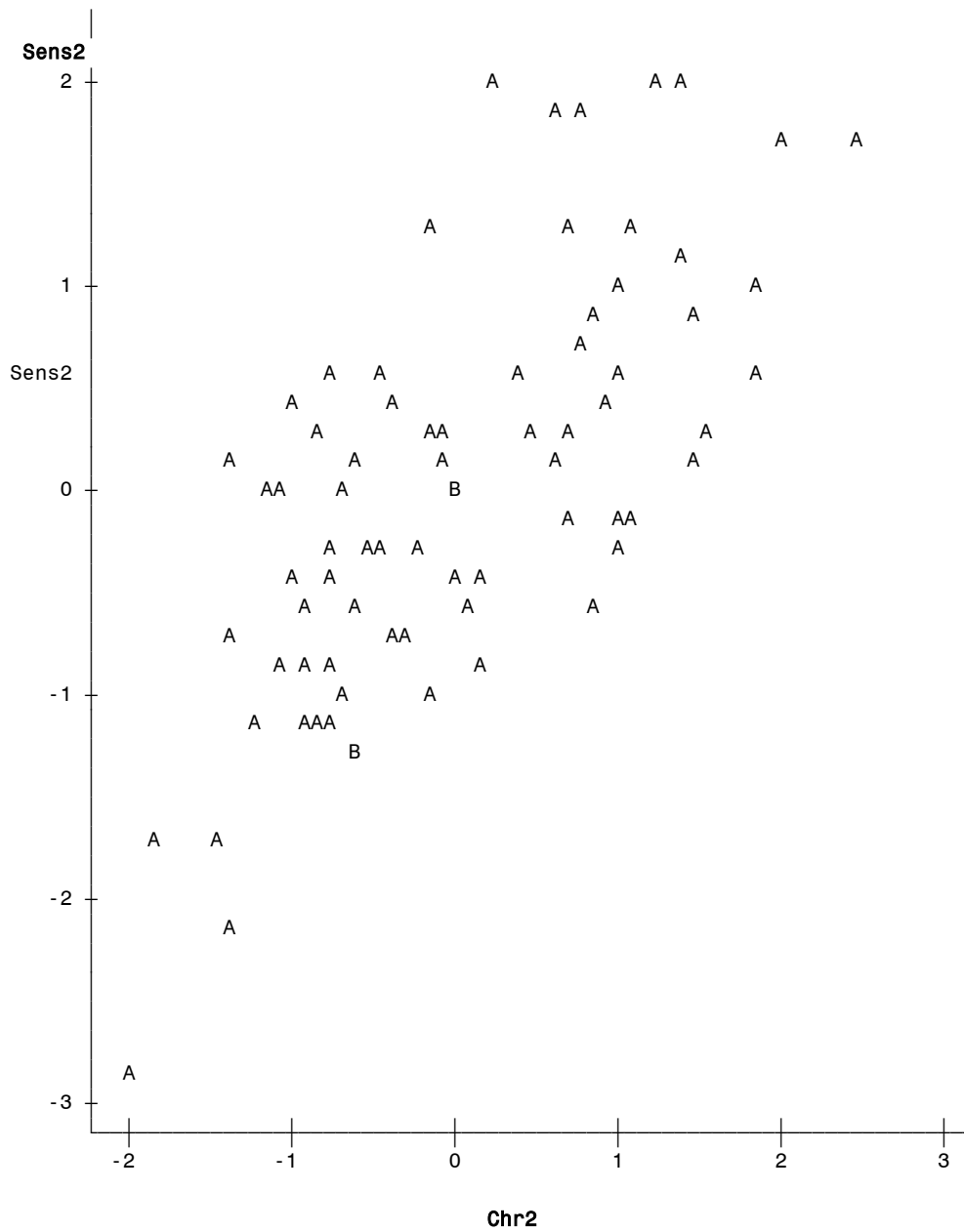


Figure 5.11. CCA plot -2, illustrates almost poor relationship between the second canonical function (Chr2) of the maturity indices data and the second canonical function (Sens2) of the sensor response data. Sen2 represents non of the sensors, while Chr2 represents soluble solids.

Chapter 6 - Correlation of EN Sensor Responses of Gala and York

6.1. Introduction

Only sensor 16 was significant for both apple cultivars. So all-32 sensors of Gala and York were analyzed by CCA. And then CCA was performed to analyze the significant EN sensors.

6.2. CCA on 32 Sensors

Though the EN sensor response correlations among themselves in case of Gala ($r = 0.7-0.9$) and York ($r > 0.95$) were very high, the correlation of sensor responses of Gala with that of York responses were of modest range ($r = -0.1-0.75$). There were some sensors for Gala, which were negatively correlated with the sensor responses for York. From the canonical Plot (Figure- 6.1 and 6.2), we can see there is no strong correlation between the sensors of the two cultivars, indicating that the sensor responses for the two cultivars were not in the same pattern.

6.3. CCA on significant sensors

To discover which of the sensor responses were important and different, a CCA was done between the significant (reduced) sensors of Gala (sensors 5, 6, 14, 16, 17 and 23) and York (sensors 2, 3, 8, 9, 16, 24 and 27). From the correlation matrix, both the sensor responses were highly correlated among themselves (Gala ($r = 0.7-0.9$) and York ($r > 0.95$)). Two (5 and 14) sensors from Gala were highly correlated with all the

significant sensors of York. While the correlation for remaining four sensors' (sensors 6, 16, 17 and 23) with the other entire seven sensors of the York data, was relatively low.

Sensors 6, 14 (highly), 16, 17 and 23 contributed for the first canonical function Sens1, while sensors 5 (highly) and 16, 17 and 23 (moderately) contributed to the second canonical function Sens2 for Gala. All significant sensors contributed the first and second canonical functions Chr1 and Chr2 for York. From the CCA plots (Fig. 6.3 and 6.4), we can conclude that the sensor responses were correlated, but nothing specific about the underlying sensors could be inferred (due to the non-specific nature).

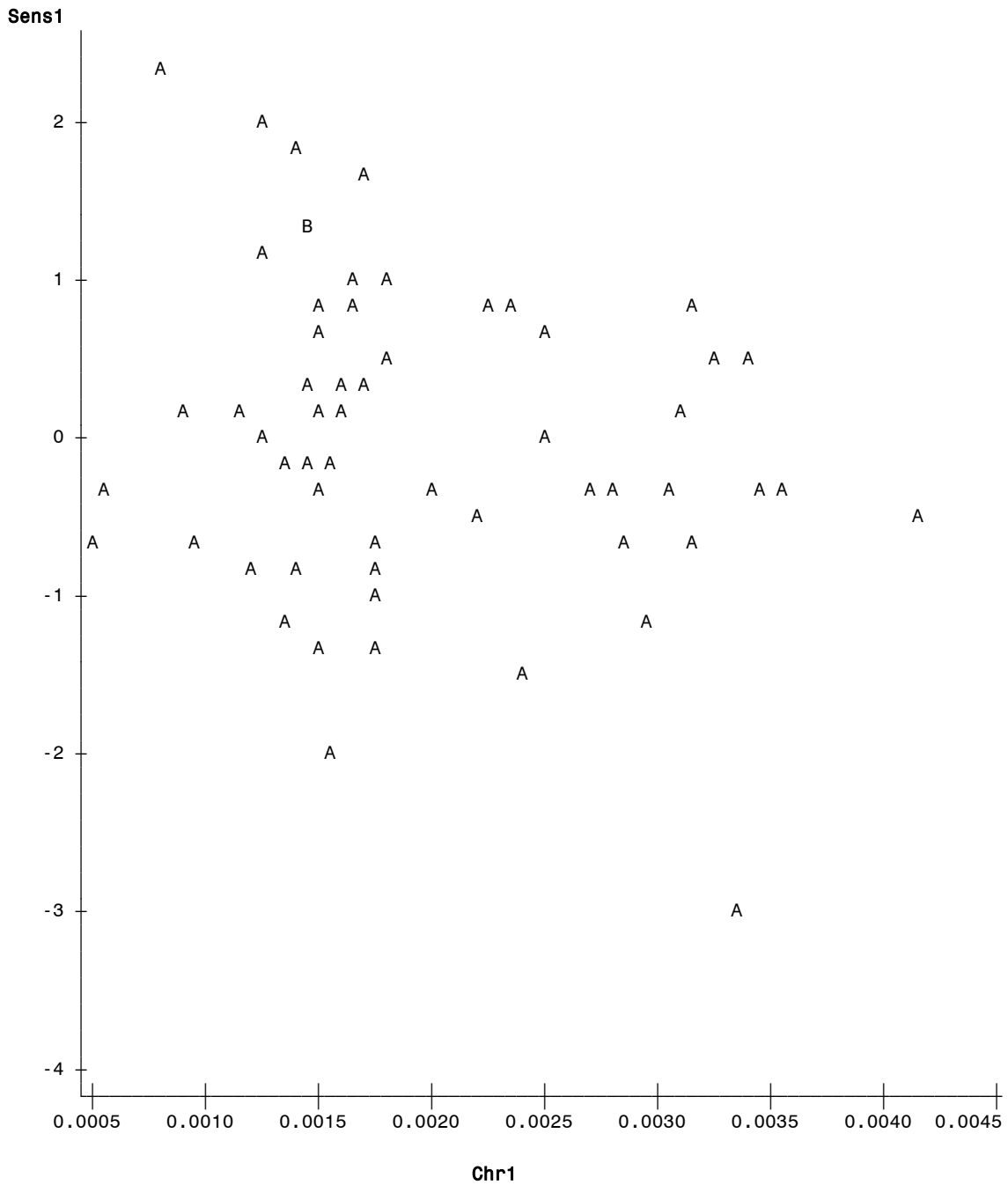


Figure 6.1. CCA plot -1, illustrates that there seems to be no correlation between the first canonical function (Sens1) and the first canonical function (Chr1). Sens1 represents responses all sensors of Gala, while Chr1 represents responses of all sensors of York.

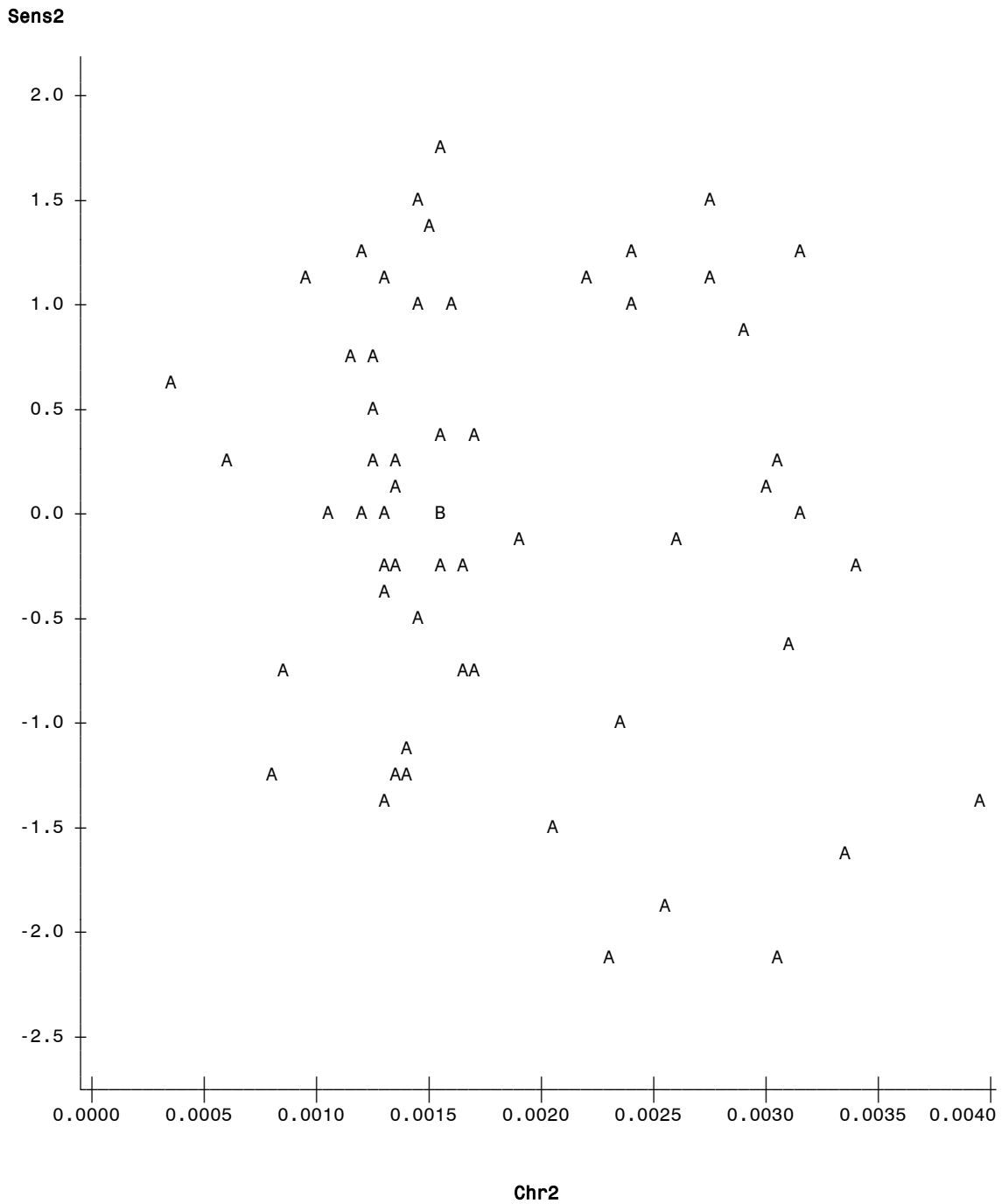


Figure-6.2. CCA plot -2, illustrates that there seems to be no correlation between the first canonical function (Sens2) and the first canonical function (Chr2). Sens2 is represents responses all sensors of Gala, while Chr2 represents responses of all sensors of York.

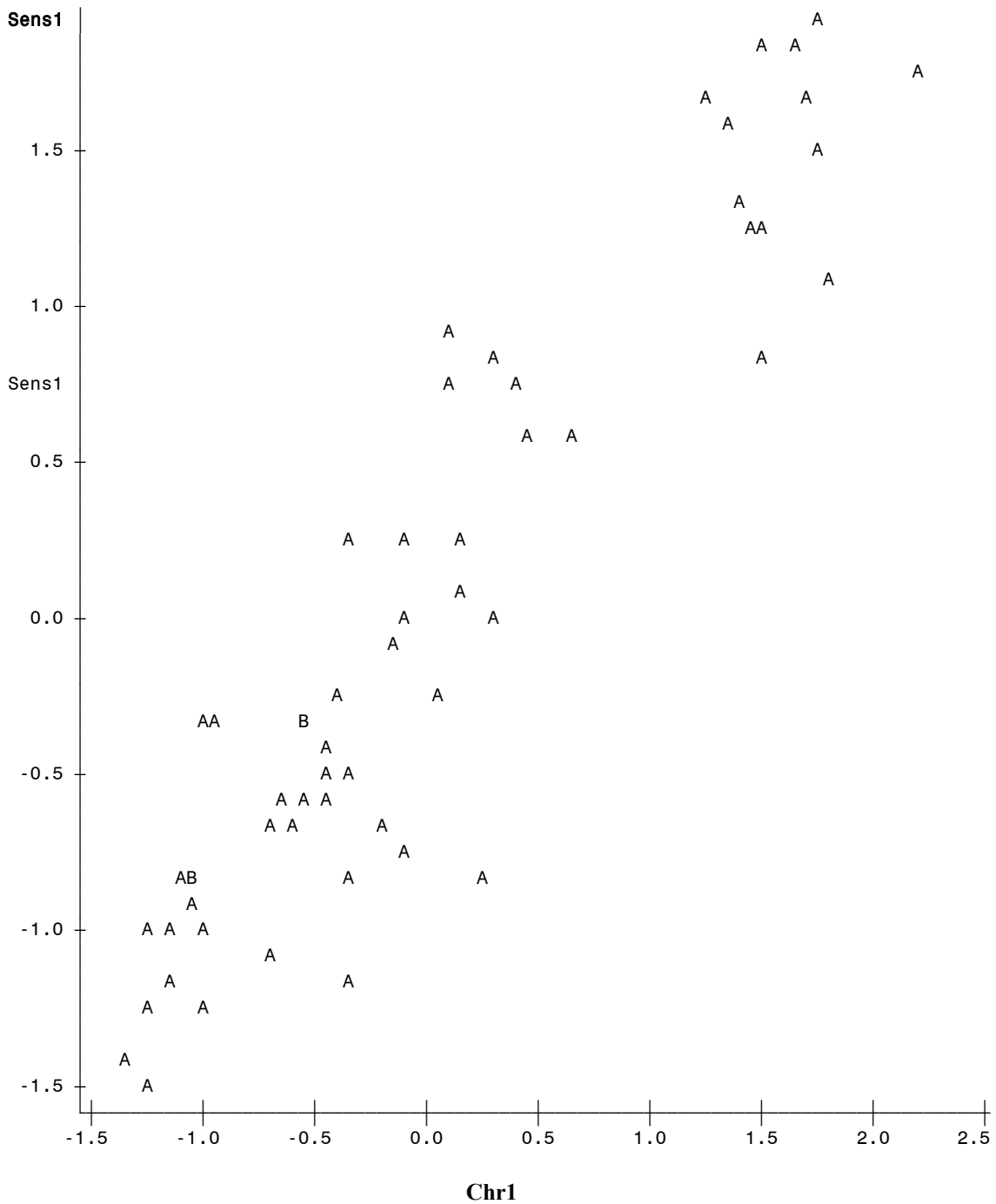


Figure-6.3. CCA plot -1, illustrates that there is a correlation between the first canonical function (Sens1) and the first canonical function (Chr1). Sens1 represents responses of significant sensors 6, 14, 16, 17 and 23 of Gala, while Chr1 represents responses of significant sensors 2, 3, 8, 9, 16, 24 and 27 of York.

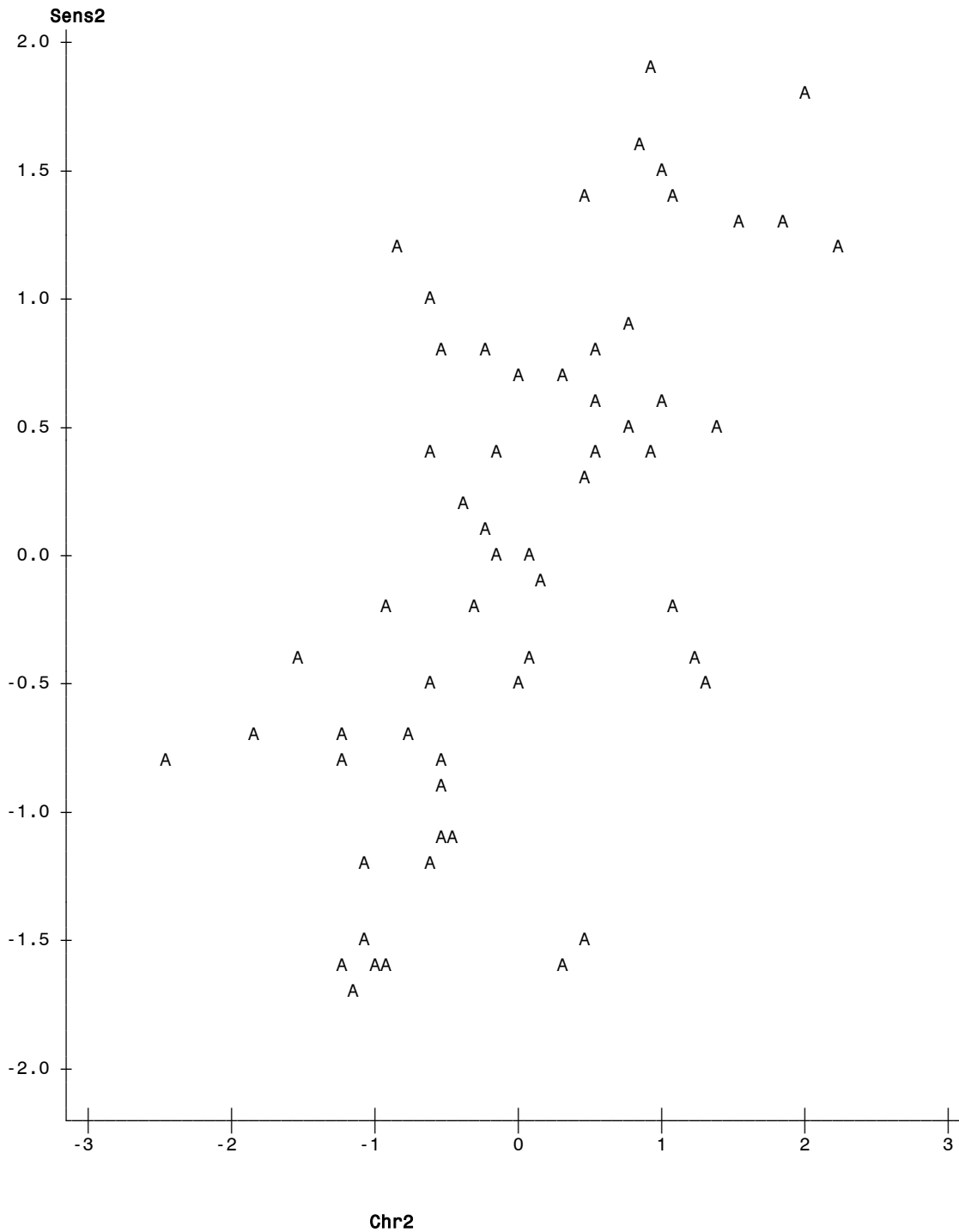


Figure-6.4. CCA plot -2 illustrates a modest correlation between the first canonical function (Sens2) and the first canonical function (Chr2) of the significant sensors of Gala. Sens2 represents responses of significant sensors 5, 16, 17 and 23 of Gala, while Chr1 represents responses of significant sensors 2, 3, 8, 9, 16, 24 and 27 of York.

Chapter 7- Chromatography Analysis

7.1. Introduction

EN sensors, when exposed to the headspace gas of an apple sample, respond to a complex mixture of gases, including carbon dioxide, ethylene, aromatic compounds and water. An experiment was conducted to determine if the aroma profile of the headspace gas also changes with fruit maturity. The efficiency of EN sensors to separate apples into aroma-based categories was evaluated. Correlation analysis was performed between the aroma compounds and the sensors.

7.2. Gala Aroma -Statistical Analysis

The chromatographic peaks for Gala were numbered from 1 to 60. From the PCA plots (Fig 7.1), three clusters were identified. The first set of chromatograms had negligible amount of aroma compound peaks (Fig. 7.2) and thus represents pre-climacteric (immature) apples. For computational reasons these peaks, whose area counts were not significant, were assigned a value of zero. The second set of chromatograms (Fig. 7.3) had significant aroma peaks and was considered as the set that represents climacteric or mature set of apples. The third set of chromatograms (Fig. 7.4), were not much different from the second group, but were different either in the levels of the aroma compounds or the number of aroma peaks. These apples were considered to be over mature. The total area counts of the aromatic compounds were taken into consideration; the chromatographic data were grouped into the previous maturity categories. From the DA analysis, the correct classification percentage was 71%.

An independent analysis of chromatographic data, based on the three climacteric levels was performed. To maintain consistency in the aroma compounds, five significant aroma compounds were selected which represented the significant peaks in the chromatograms. The five aroma compounds were hexyl butyrate, hexyl 2-methylbutyrate, butyl hexanoate, hexyl hexanoate and alpha-farnesene. These aroma compounds were confirmed with the MS library analysis. Further work is needed to confirm the aroma compound identity. Mattheis et al. (1998) evaluated the volatile ester compounds from intact Royal Gala apples and identified all the above four ester compounds. Song and Bangerth (1996) identified alpha-farnesene in Golden Delicious.

PCA analysis was performed on the chromatographic data. For convenience the aroma compounds were named a1, a2, a3, a4 and a5 for hexyl butyrate, hexyl 2-methylbutanoate, butyl hexanoate, hexyl hexanoate and alpha-farnesene respectively. From the correlation matrix (Appendix-D), a1 was highly correlated to a2 ($r = 0.85$) and a4 ($r = 0.79$), and a4 was highly correlated to a2 ($r = 0.90$). Correlations among the remaining variables were of modest ($r = 0.50-0.70$). From the eigenvalues the first four principal components explained 98.5% of the total variation. All aroma components contributed almost equally to Prin1. So Prin1 axis can be interpreted as the total amount of aroma evolution scale from the apples. From the PCA plot (Fig. 7.5), and based on the Prin1 axis, all the observations can be classified into the three stages of maturity. Fruits in the pre-climacteric group had low amounts of aroma compounds, climacteric fruits had intermediate levels, and post-climacteric fruits had high levels of aroma production.

MANOVA was performed and the data were from a multivariate normal distribution (Fig. 7.6). Observation 51 was an outlier and was subsequently removed. For the subsequent analysis only 59 observations were considered. Results from a MANOVA indicated that maturity affected aroma (Wilks' lambda $F = 52.13$, $P < 0.0001$). The first four principal components were the input data for CDA and DA. The CDA plot (Fig. 7.7), verified conclusions from PCA, and the canonical plot provide a better separation of groups. DA categorized the emanation levels of the five compounds perfectly (100%) (Table 7.1). The interclass separation was higher (Table 7.2) than the interclass separation based on maturity indices, indicating that categorization with chromatographic data was more a reliable. These results agree with Brackmann et al. (1994) who reported that apple aroma production increases with climacteric respiration and reaches a maximum 2-3 weeks later.

Electronic nose sensors response data was evaluated statistically. The electronic sensor data was classified based on aroma and PCA was performed. Due to the non-specific nature of sensors, the principal axes in the PCA plot (Fig. 7.8) could not be interpreted in terms of sensors, but a grouping structure was obvious in the PCA and CDA plot (Fig. 7.9). DA indicated that 80% of the apples were classified correctly (Table 7.3). The interclass distances were higher than 5, indicating good separation of categories (Table 7.4). The DA was only 80% correct indicating that the electronic nose sensors, in addition to responding to aroma compounds, might also be responding to other maturity induced compounds such as ethylene, water vapor, carbon dioxide and other varying aroma compounds. Such conclusion may be valid since electronic nose sensors technology is not yet completely understood.

From the CCA and from Sens1 and Chr1 analysis, sensors 1, 2, 4, 8, 12, 14 and 27 were highly correlated to a1, a3 and a4 aroma compounds. While from Sens2 and Chr2 analysis, most of the 32 sensors were moderately correlated to a3 and a5 aroma compounds. The CCA plot (Fig. 7.10) indicates that sensor data and chromatographic data are correlated, indicating that the seven sensors might be responding to the aroma compounds a1, a3 and a4. On the other hand Fig. 7.11 indicates that the sensor data is poorly correlated to chromatographic data, indicating that the 32 sensors response to the aroma compounds a3 and a5 was not very high.

7.3. EN sensor response, maturity indices (including ethylene) and five aroma compounds data

Final analysis was done with all 60 sixty observations with 16 variables (6 significant sensors, 5 maturity indices (including pH and ethylene) and 5 aroma compounds). All 32 sensors were not considered because the analysis might be biased towards the EN sensors. From the CCA, sensor 14 (out of six significant sensors) was highly ($r = 0.8$) correlated with all aroma compounds, starch and puncture strength. CCA plot-1 (Fig. 7.12) illustrates a poor correlation between Sens1 (represents only one EN significant sensors 14, out of the six significant sensors) and Chr1 (represents starch, firmness and five aroma compounds) because out of six sensors, only one sensor is correlated with Chr1. On the hand, Sens2 represents all six sensors, while Chr2 represented all maturity indices. CCA plot-2 (Fig. 7.13) illustrates a linear correlation between Sens2 and Chr2, indicating correlation between EN sensors responses and maturity indices. From the DA,

the correct classification percentage was 83%. Thus it can be concluded that EN sensors, maturity indices including aroma compounds were measuring the maturity of the apple.

7.4. York- statistical analysis

The analysis with York chromatographic data resulted only in two groups (Fig 7.14).

There was a problem with the glass liner in the gas chromatograph, for the first two harvest dates. By the time we realized the defect, the third harvest measurements were in progress, and during the analysis the fiber was stripped off. Data from the first three harvests are questionable. Therefore, exploratory PCA was performed and the observations were categorized into two groups. Since these results were not expected statistical analysis was terminated.

7.5. Comparison of Gala and York chromatograms

The chromatograms of Gala and York for the final harvest are shown in Fig. 7.15 and 7.16. Though the chromatograms look similar, they vary both in number of aroma compounds and their respective concentration levels. This comparison can help partially explain the difference in significant sensors for the two cultivars.

7.6. Summary

Gala apples were divided into three categories, based on chromatographic data. To maintain consistency, five aroma compounds (hexyl butyrate, hexyl 2-methylbutyrate, butyl hexanoate, hexyl hexanoate and alpha-farnesene) were selected. The three groups were considered as pre-climacteric, climacteric and post-climacteric fruits. From the PCA

plots, all five-aroma compounds equally contributed to Prin1. Thus Prin1 could be interpreted as the total aroma production scale. From the Prin1, it can be concluded that pre-climacteric fruits had low levels of total aroma, climacteric fruits had intermediate levels and the post-climacteric fruits had high levels of aroma. From the DA, the aroma-based classification was 100% correct. The aroma indices were also classified into maturity-based categories, and DA was performed, the correct classification percentage was 71%. The electronic sensor responses were also classified into aroma-based categories, in order to determine the effectiveness of the electronic nose sensor response classification. The correct percentage was 80%.

CCA was, it can be concluded that the EN sensor technology, maturity indices based measurements and chromatographic data were all correlated (at least by the correlation value of 0.7). The correct classification percentage was 83% when all 16 variables were taken into consideration.

Table 7.1. Classification table obtained after performing DA for Gala apples. Columns and rows represent the maturity categories (1, 2, and 3). The values represent the number of fruits placed in that category. The groups were categorized based on aroma compounds emanation levels.

From Maturity	1	2	3
1	26	0	0
2	1	17	0
3	0	0	16

Table-7.2. Mahalanobis distance to determine the similarity of groups of Gala apples with unknown maturity to a set of values from 3 groups of apples based on maturity indices. The left column indicates the groups and the right column indicates the distance

From Maturity	1	2	3
1	0	22.94366	66.67160
2	22.94366	0	20.55312
3	66.67160	20.55312	0

Table 7.3. Classification table obtained after performing DA for Gala apples. Columns and rows represent the maturity categories (1, 2, and 3). The values represent the number of fruits placed in that category. The categorization was performed on EN sensor response data.

From Maturity	1	2	3
1	21	5	0
2	3	12	2
3	0	2	14

Table 7.4. Mahalanobis distance to determine the similarity of groups of Gala apples with unknown maturity to a set of values from 3 groups of apples based on maturity indices. The left column indicates the groups and the right column indicates the distance between them.

From Maturity	1	2	3
1	0	6.83572	27.62271
2	6.83572	0	8.99409
3	27.62271	8.99409	0

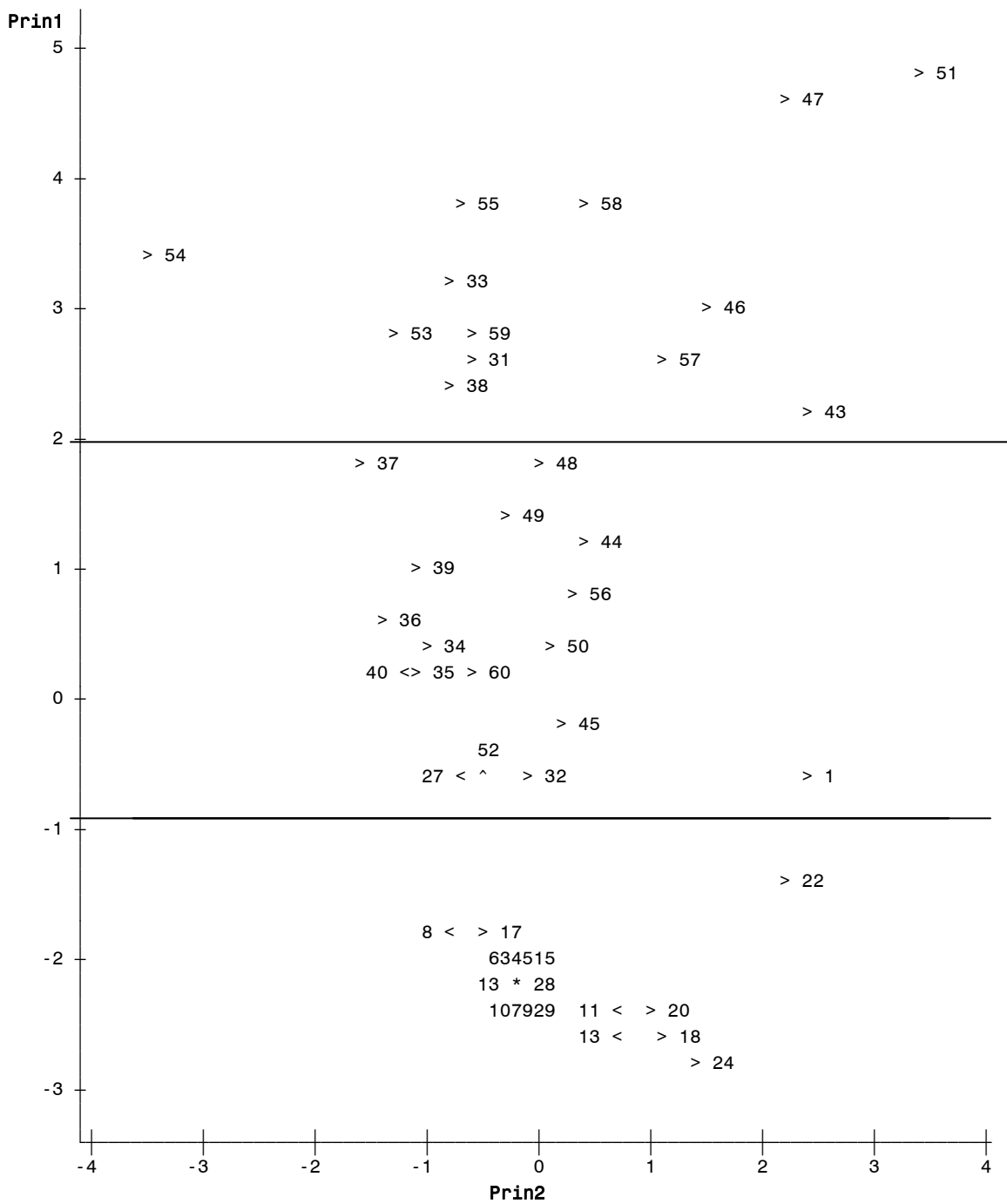


Figure 7.1. PCA plot illustrates the three clusters. Prin1 represents all aroma compounds emanation levels. Prin2 is represented only few other significant aroma compounds emanation levels.

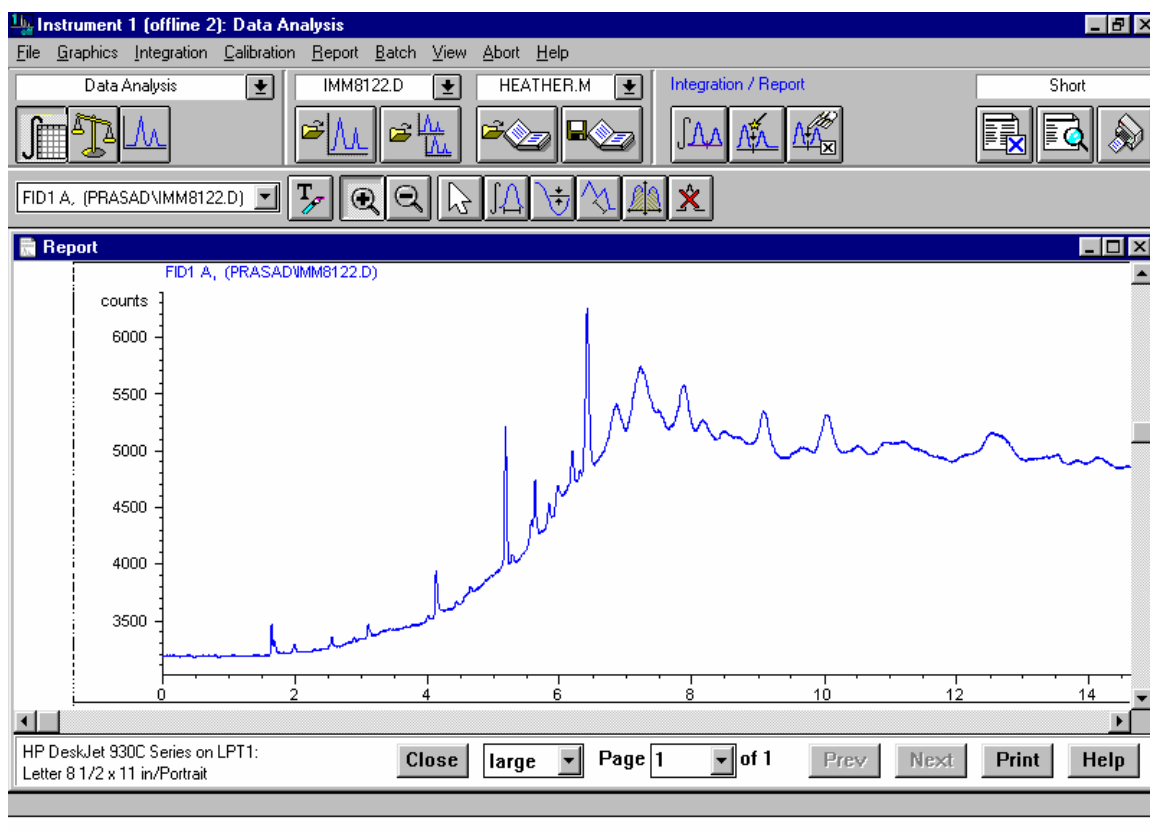


Figure 7.2. Typical chromatogram of volatiles compounds released from immature Gala apples, 10 days before optimum harvest maturity. No significant peaks were observed.

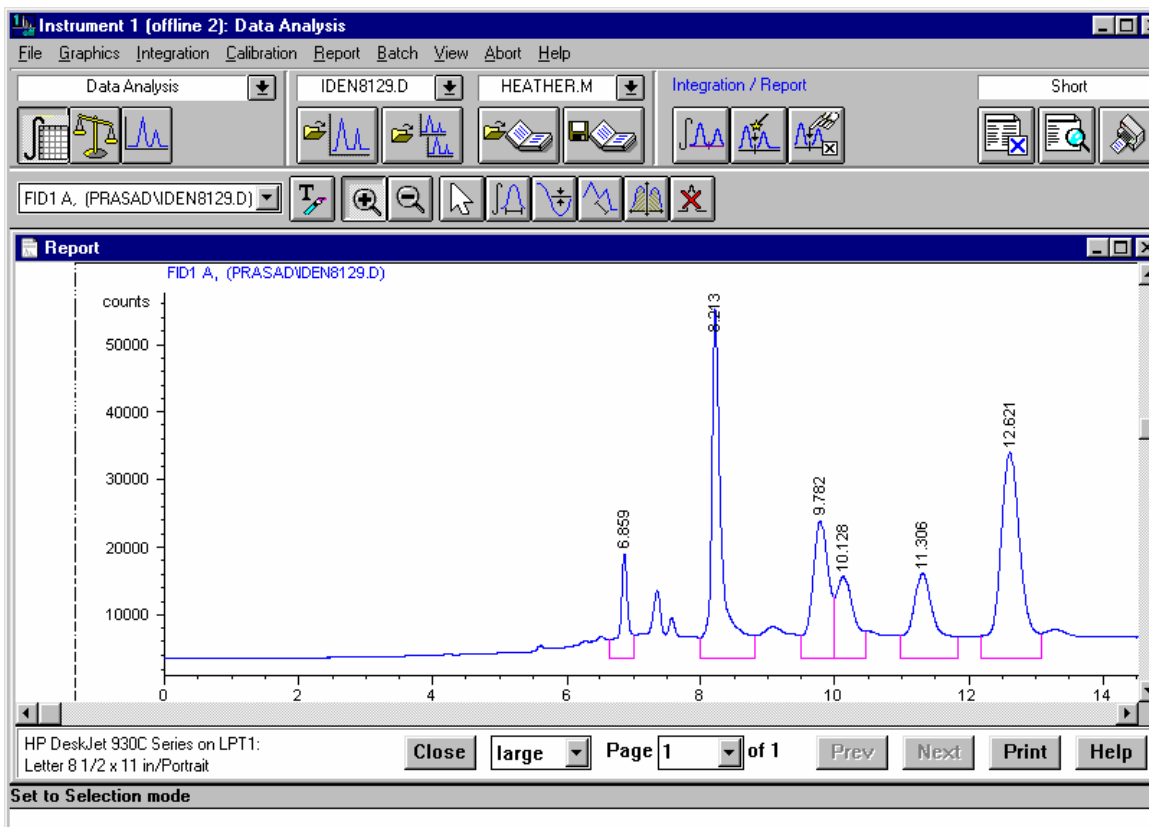


Figure 7.3. Typical chromatogram of volatile compounds released from mature Gala apples, 10 days before optimum harvest maturity. In order to maintain consistency, five significant peaks were considered. The retention times of the five volatile compounds were 8.2, 9.7, 10.1, 11.3 and 12.6 minutes. They were identified using the MS library (GC/MS) as hexyl butyrate, hexyl 2-methylbutyrate, butyl hexanoate, hexyl hexanoate and alpha-farnesene.

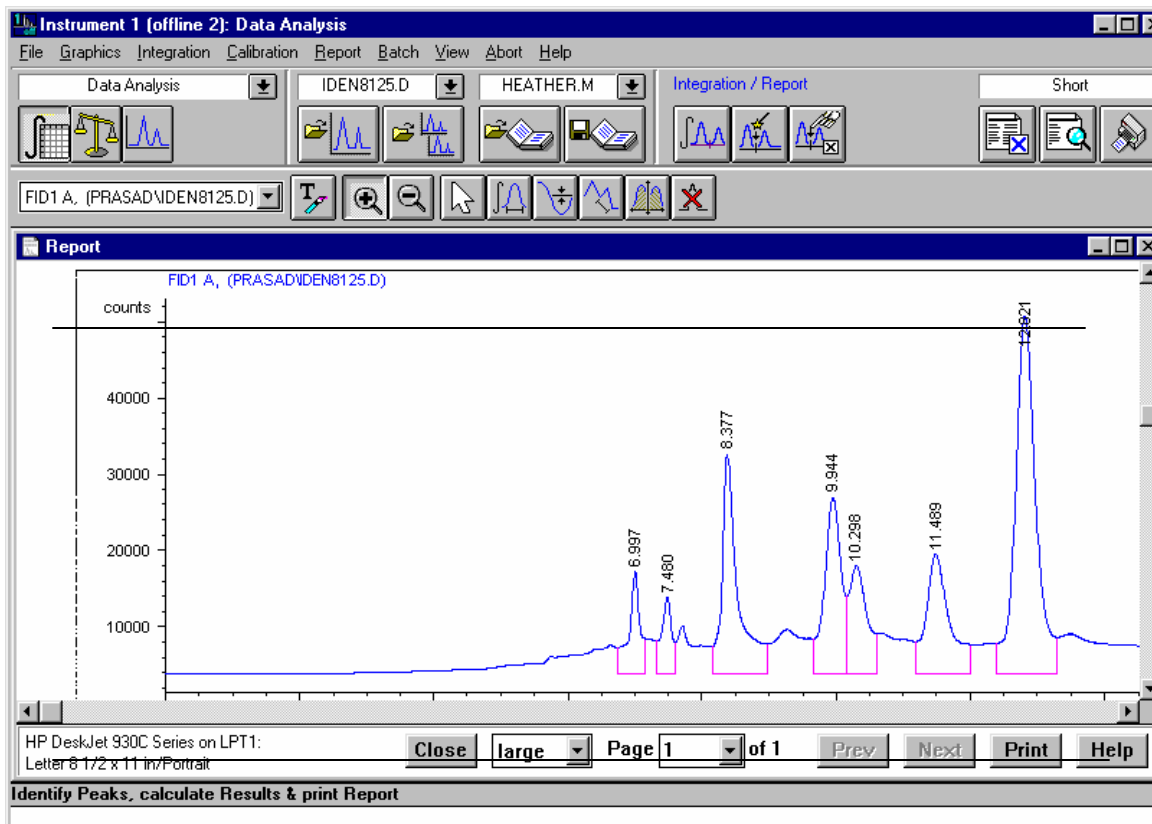


Figure 7.4. Typical chromatogram of volatile compounds released from over mature Gala apples. The chromatogram represents the optimum physiological maturity. The above five significant (figure-7.2) volatile compounds were considered.

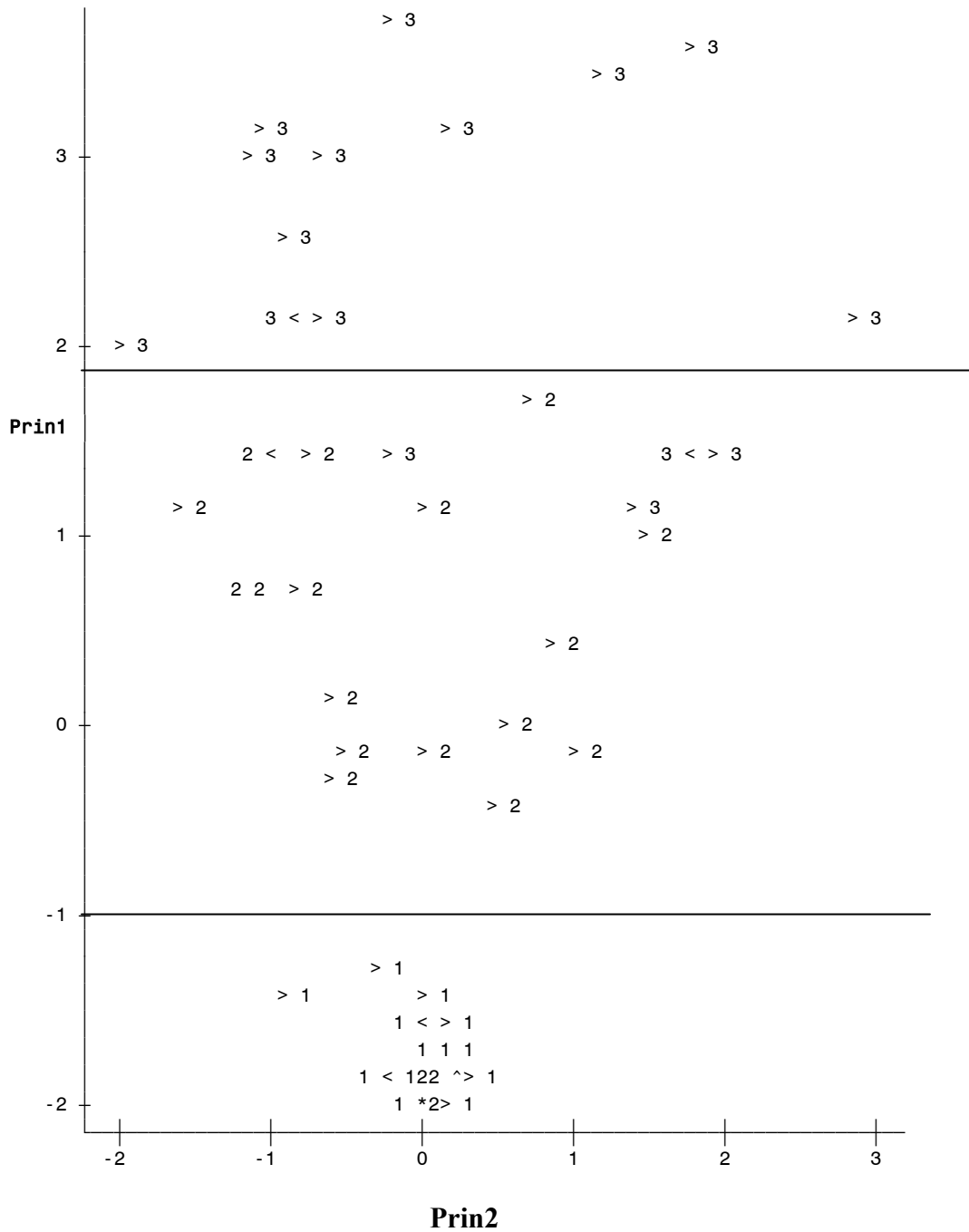
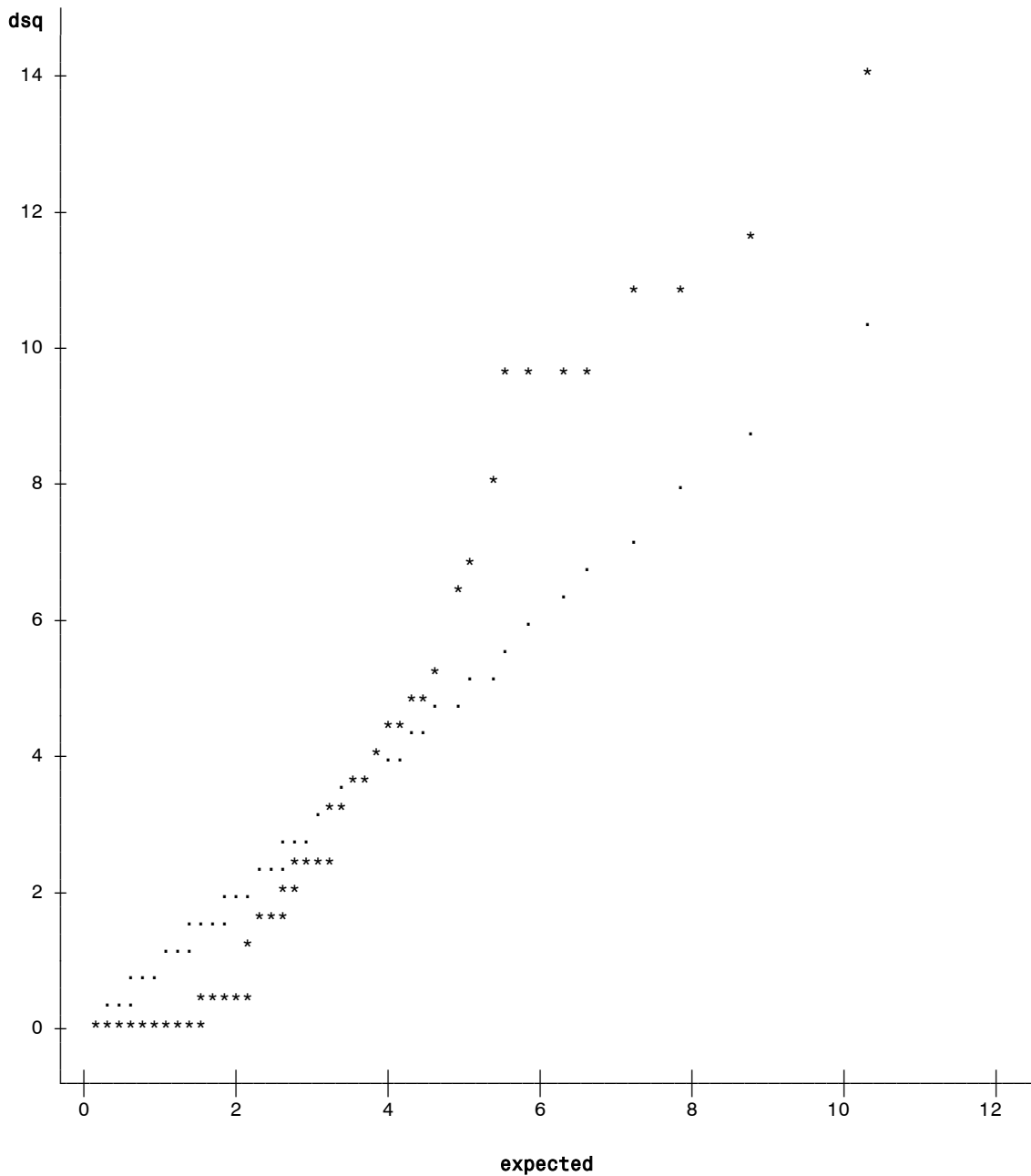


Figure 7.5. PCA plot after the categorization of the three clusters based on figure into three aroma-based categories. The categories 1, 2 and 3 represent pre-climacteric (immature), climacteric (mature) and post climacteric (over mature) groups respectively.



NOTE: 32 obs hidden.

Figure-7.6. Chi-Squared probability plot for multivariate normal. The X-axis (expected) represents Chi-Square Quantile. The Y-axis (dsq) indicates the mahalanobis D-square distance.

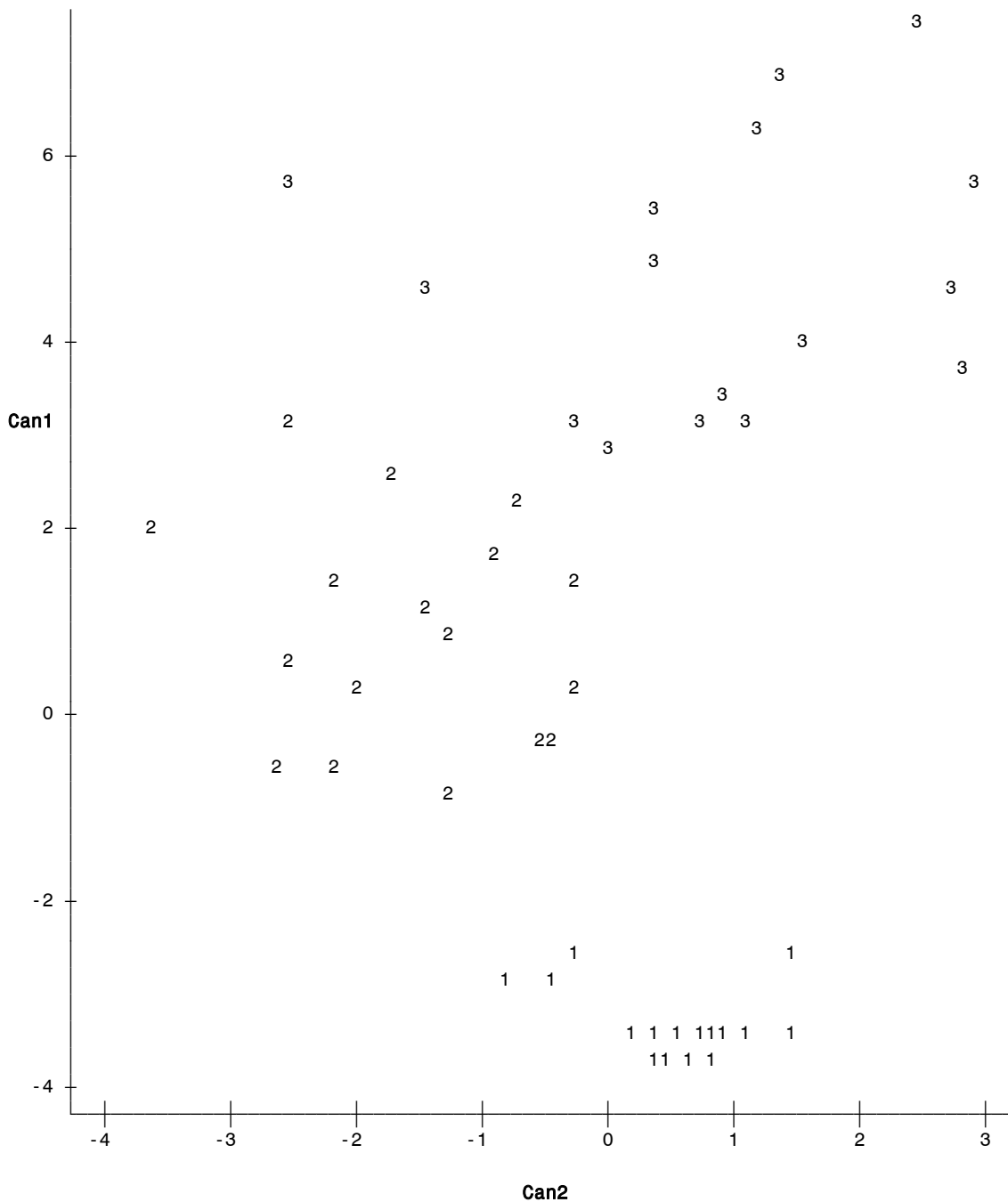


Figure-7.7. CDA plot after the categorization of the three clusters based on Fig. 7.5 into three aroma-based categories. The categories 1, 2 and 3 represent immature, mature and over mature groups respectively. Prin1 represents all five aroma compounds emanation levels. Prin2 is represented three significant aroma compounds emanation levels.

Prin1 |

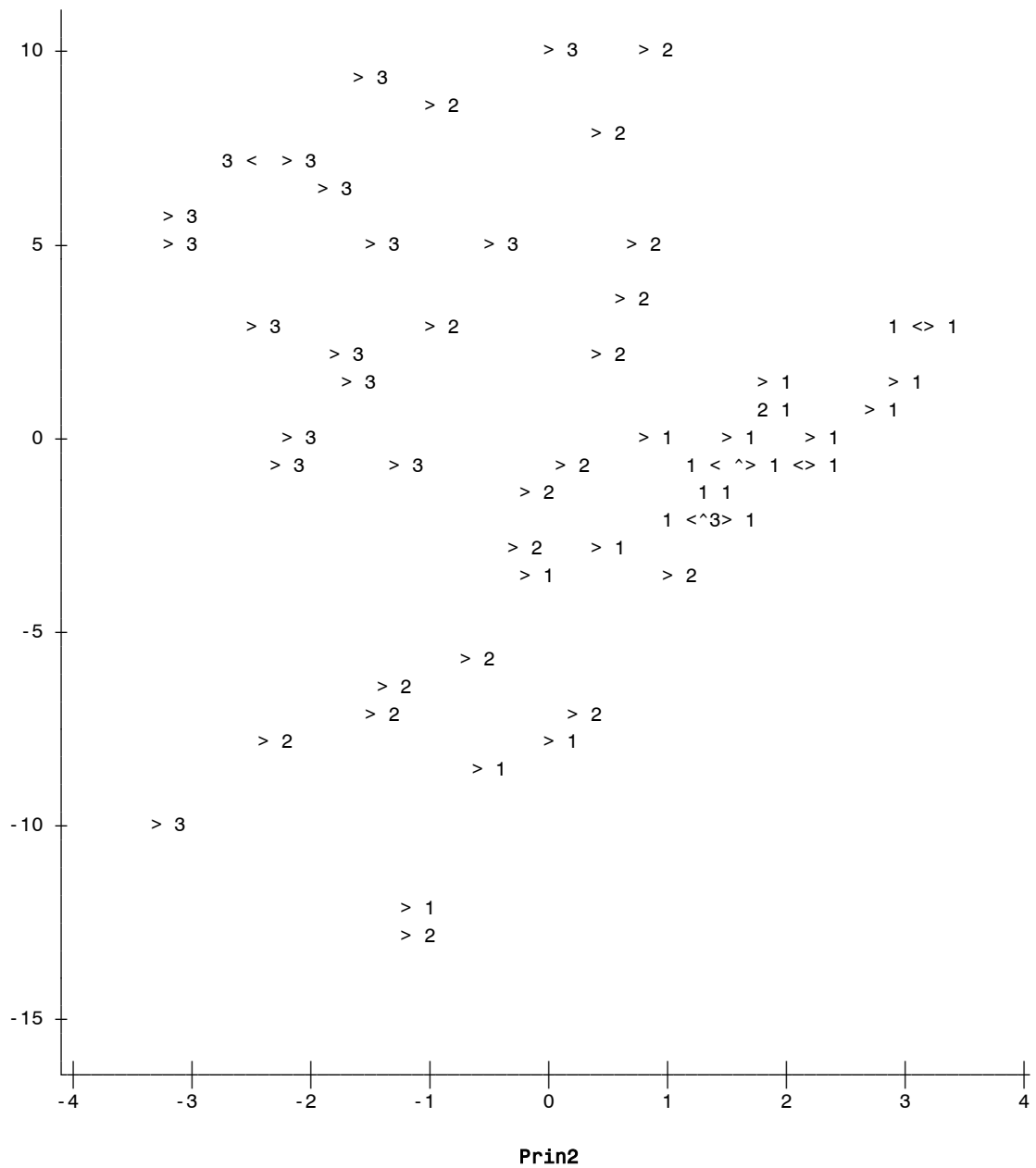


Figure-7.8. Scatter plot of first two principle components after the Electronic nose sensor response data were categorized into three maturity groups (based on Fig 4.2). The categories 1, 2 and 3 represent immature, mature and over mature groups, respectively. Prin1 represents the response of all sensors. Prin2 represents most of the sensors responses.

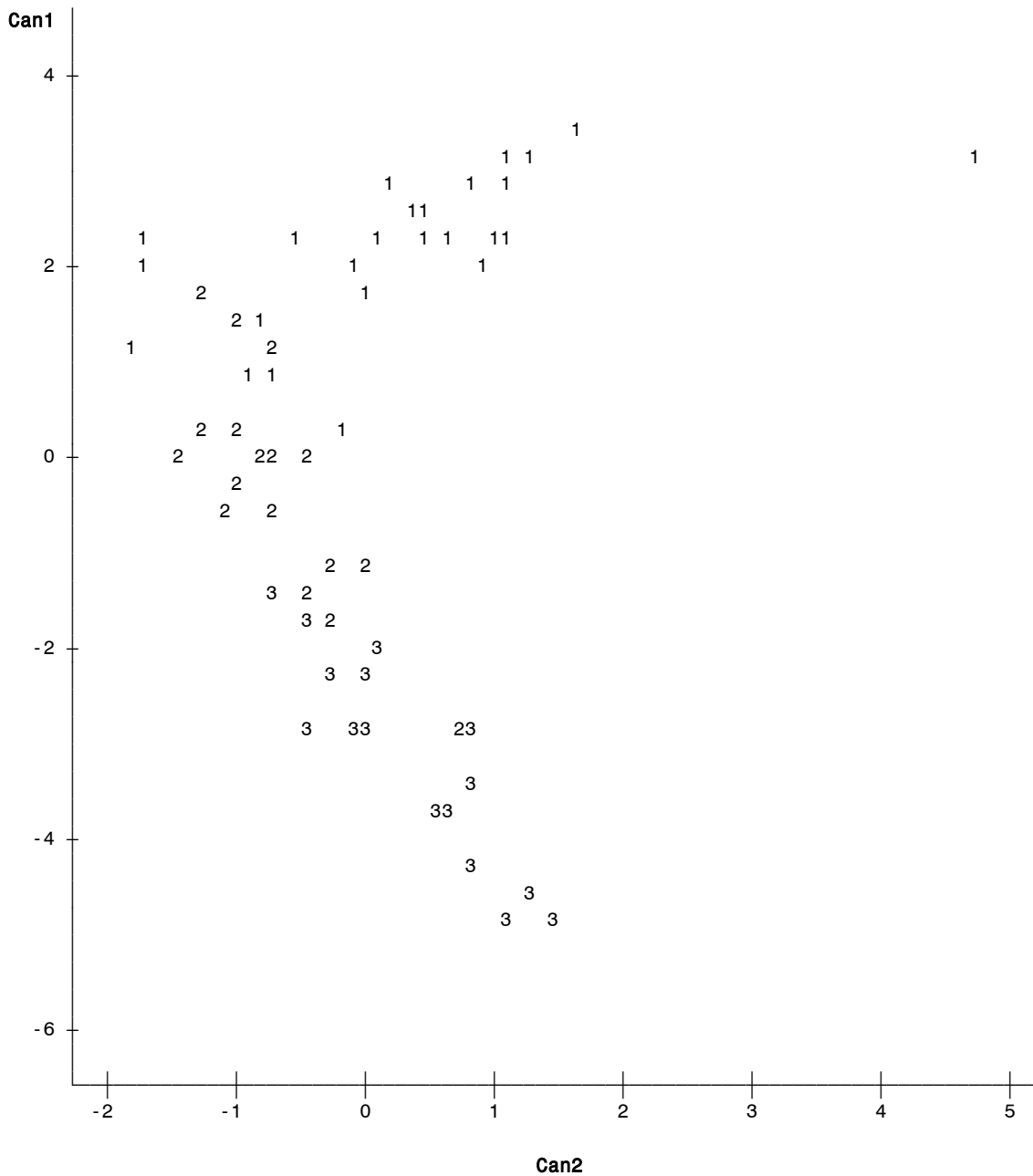


Figure-7.9. Scatter plot of the first two canonical variables after the EN sensor response data into the three maturity categories. The categories 1, 2 and 3 represent immature, mature and over mature groups respectively. Can1 represents the response of all sensors. Can2 represents most of the sensors responses.

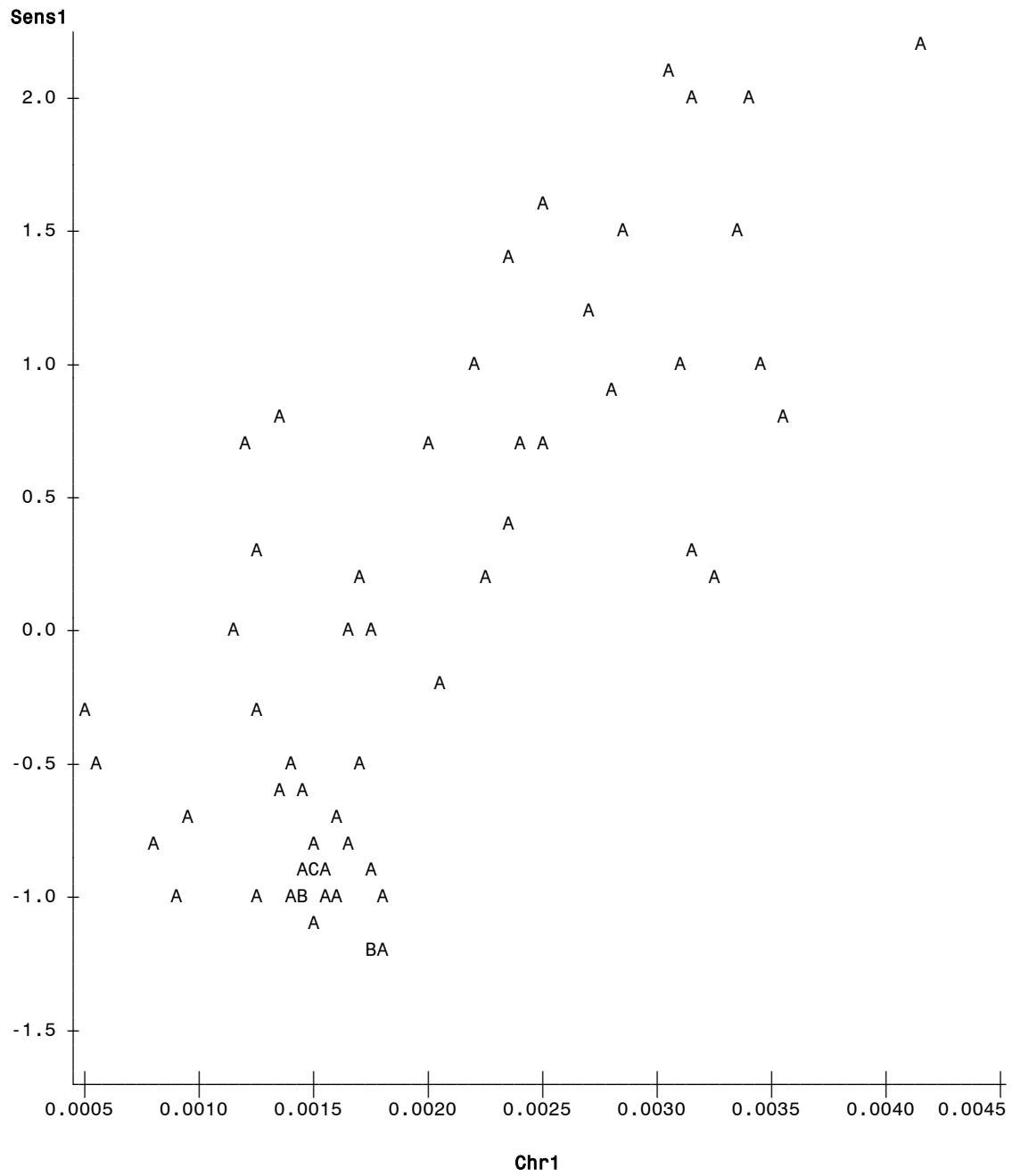


Figure-7.10. CCA plot -1, illustrates a linear correlation between the first canonical function (Sens1) and the first canonical function (Chr1). Sens1 represents sensors (6 highly correlated sensors) and Chr1 represents a1, a3 and a4.

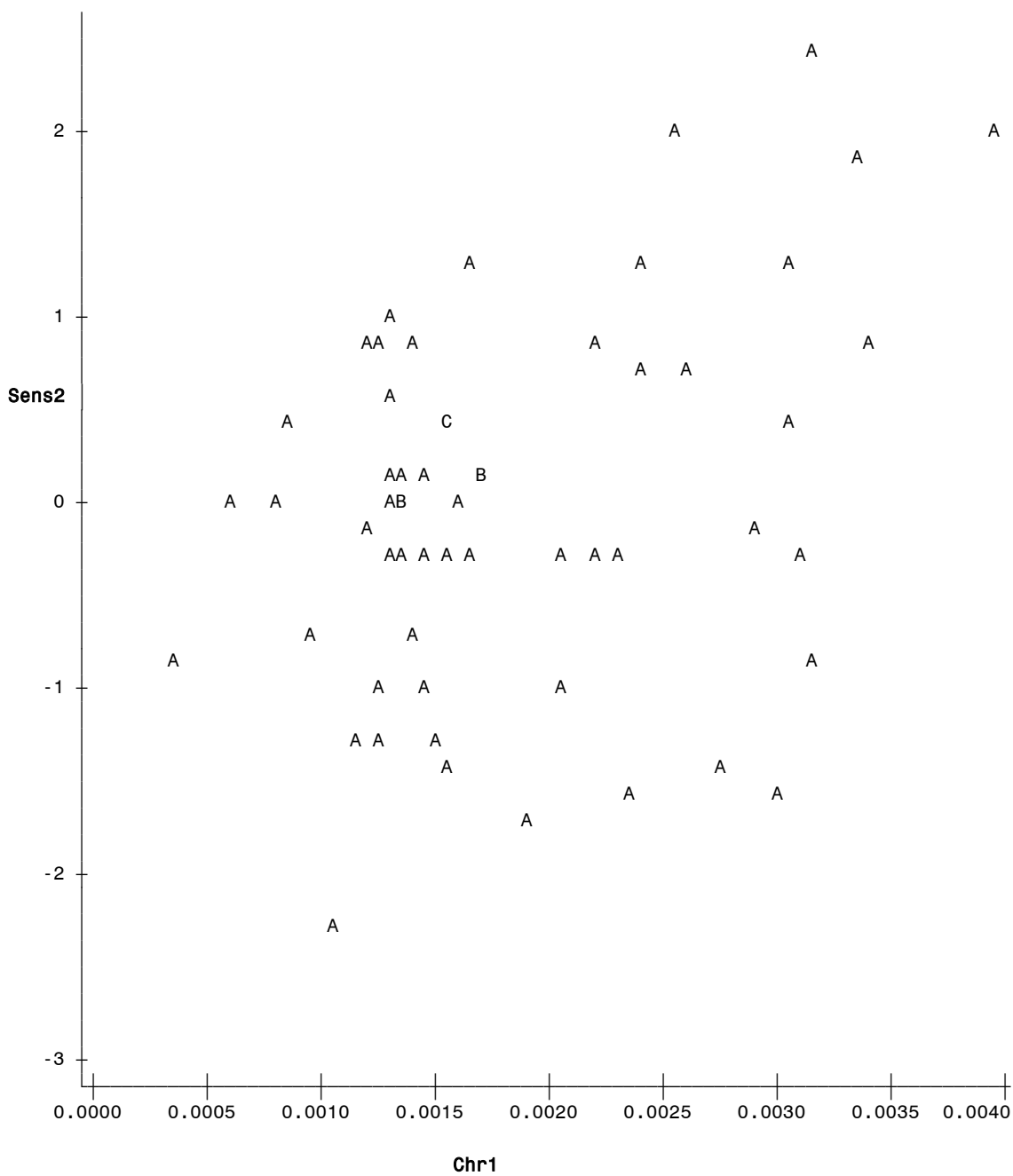


Figure-7.11. CCA plot -2, illustrates almost no strong correlation between the second canonical function (Chr2) of the chromatographic data and the second canonical function (Sens2) of the sensor response data. Sens2 represents all sensor responses while Chr2 represents soluble a3 and a5.

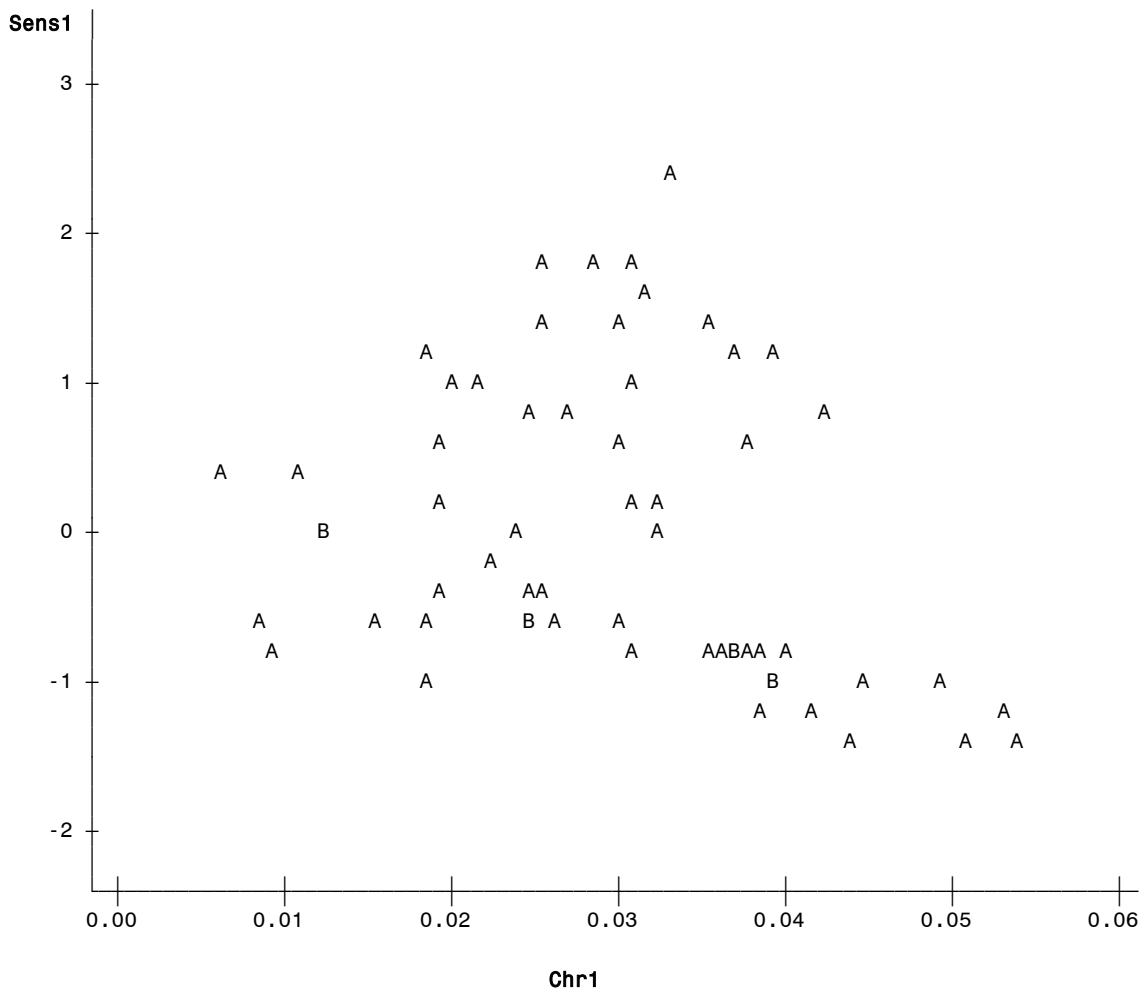


Figure-7.12. CCA plot -1, illustrates a poor correlation between the first canonical function (Chr1) of the chromatographic data and maturity indices and the first canonical function (Sens1) of the sensor response data. Chr1 represents starch, firmness and five aroma compounds. Sens1 represents only one sensor (sensor 14).

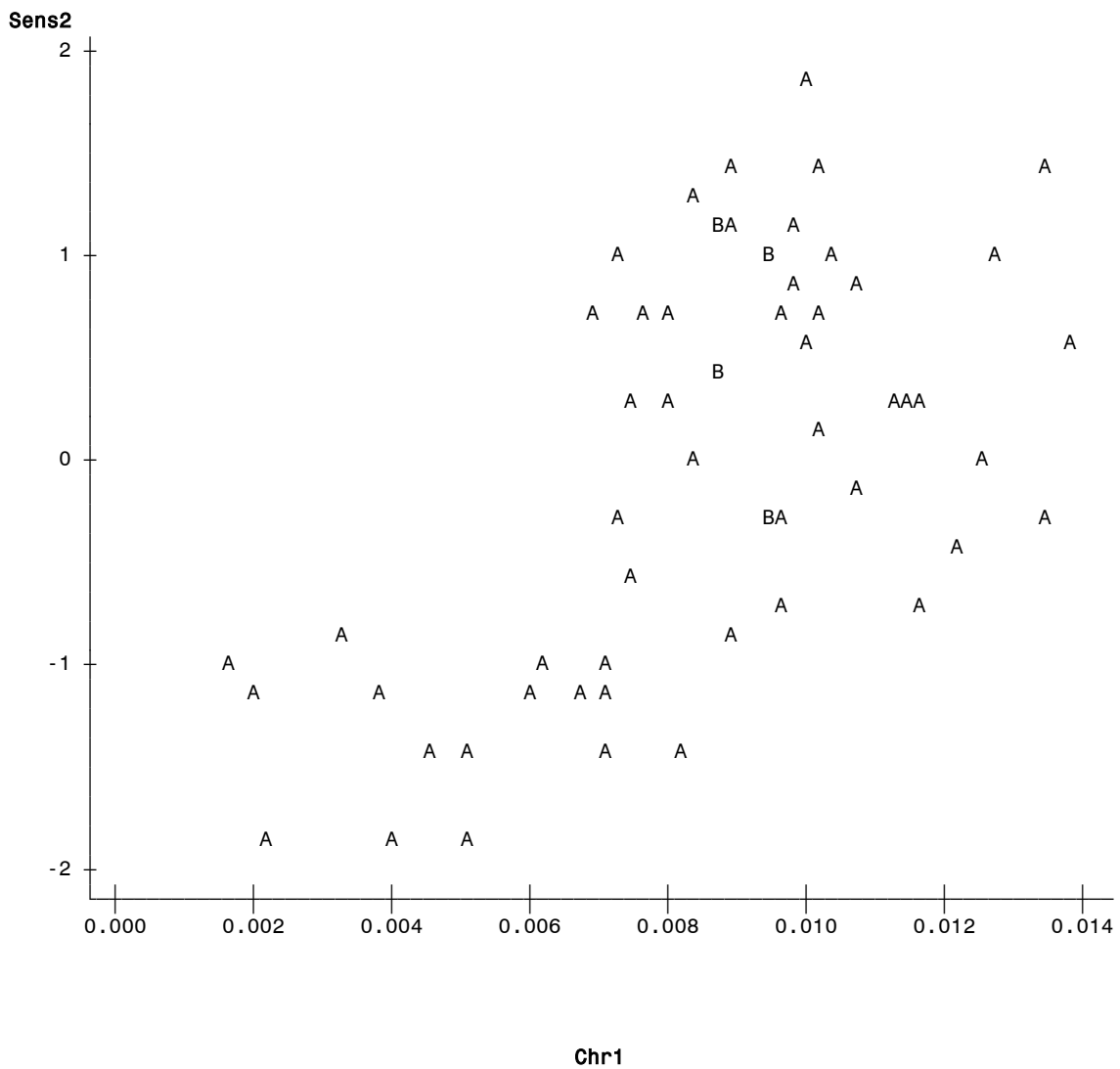


Figure-7.13. CCA plot -2, illustrates a linear correlation between the second canonical function (Chr2) of the chromatographic data and maturity indices and the second canonical function (Sens2) of the sensor response data. Chr2 represents all maturity indices, while Sens2 represents all significant sensors.

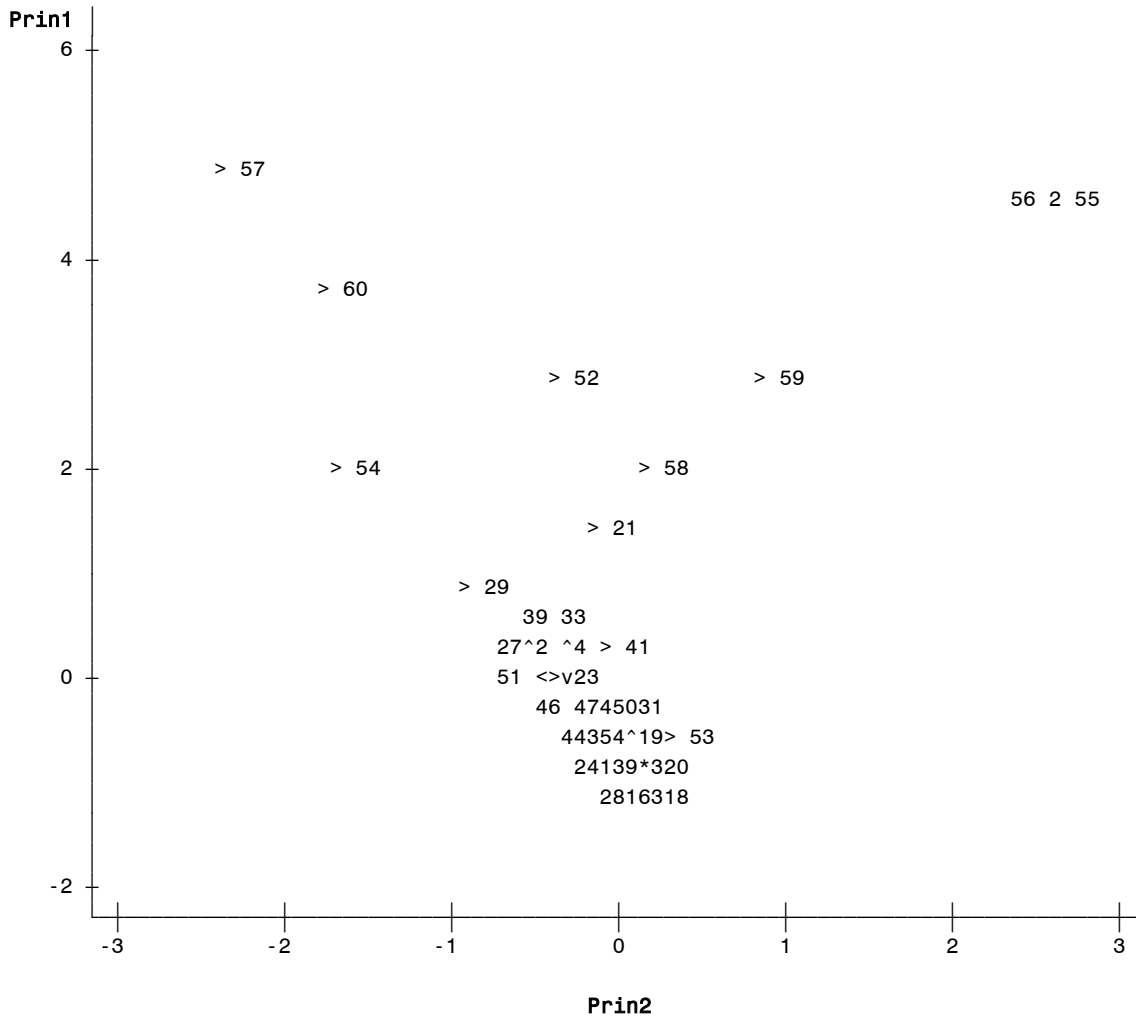


Figure-7.14. Exploratory PCA plot for York chromatographic analysis. Only two discernible groups could be obtained.

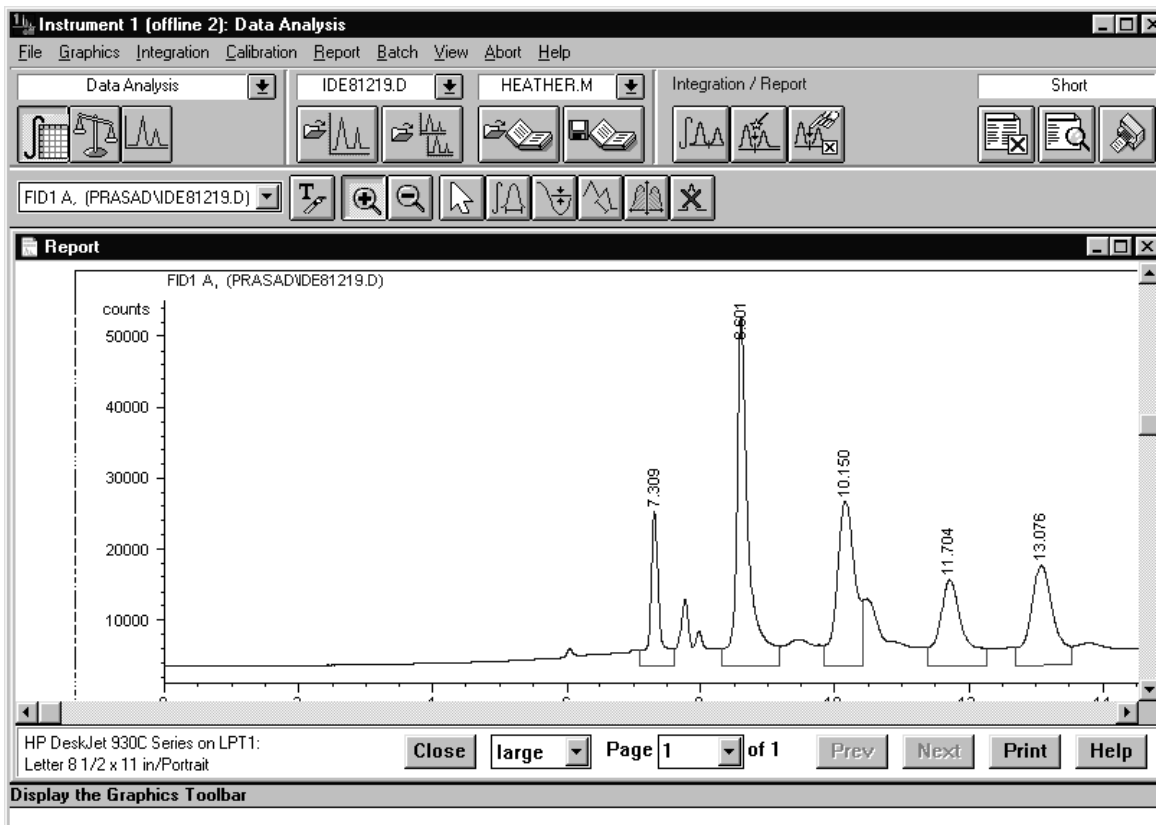


Figure 7.15. Typical chromatogram of volatile compounds released from over mature Gala apples.

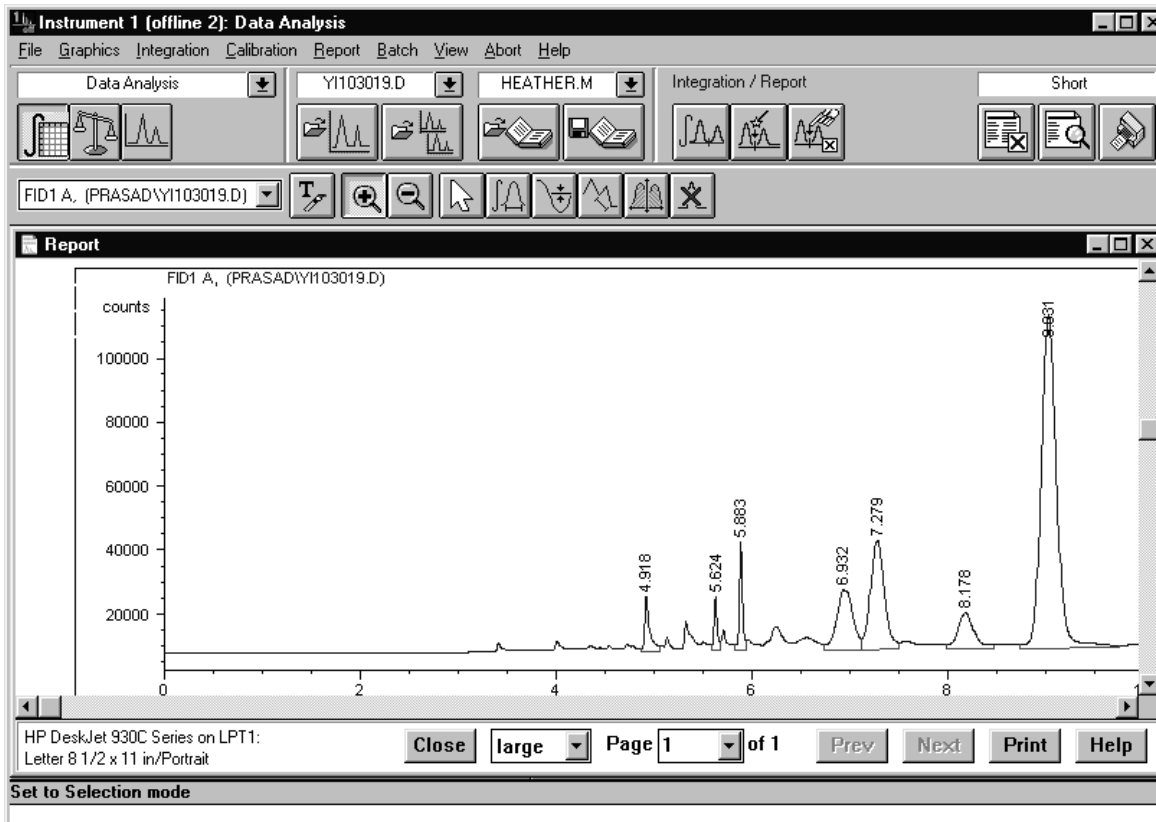


Figure 7.16. Typical chromatogram of volatile compounds released from over mature York apples.

Chapter 8

8.1. Conclusions

- 1) Based on the four physical parameters, namely starch, firmness (puncture test in case of Gala), soluble solids and color, an objective evaluation of maturity could be accomplished.
- 2) PCA, CDA and DA results demonstrated that the electronic nose could be used to classify apples into three identified maturity-based groups.
- 3) EN sensors could also demonstrate that they can classify the apples by harvest date and, thus, can be used to detect optimum harvest date.
- 4) Except for the classification based on harvest date for York apples, the statistical DA was more efficient in correct classification than the EN system software analysis.
- 5) From the CCA of EN responses between and York and Gala apples, the EN system might be used to identify the specific variety of the apples under consideration.
- 6) Based on aroma (five volatile compounds), three categories could be classified, namely pre-climacteric, climacteric and post-climacteric groups.
- 7) The EN sensors in case of Gala apples showed better classification percentage in case of aroma-based categories (80%) than the maturity based categories (77%).
- 8) Thus, it can be concluded that the EN system holds promise as non-destructive evaluation technique to determine the maturity of an apple.

8.2. Recommendations

- 1) The developed training set could not be used to identify the apples using the EN system, since the age of the apple was not the same original age of the apples used to build the training set. Since the training sets were generated in the present project, the reliability of this database (training set) should be tested in the coming years.
- 2) Since most of the apples are stored before the final use, the available non-destructive technique (using a EN system) should be used to evaluate the maturity during storage, in order to understand the factors that affect maturity.
- 3) A more comprehensive knowledge of sensors would help to determine the underlying mechanisms that occur in sensor responses. Such an understanding would provide to upgrade the EN system to evaluate both maturity and quality of the apples, rapidly and efficiently.

References

Abbott, J. A., M. F. Childers, G. S. Bachman, J. V. Fitzgerald, and F. J. Matusik. 1968. Acoustic vibration for detecting textural quality of apples. *J. Amer. Soc. Hort. Sci.* 93: 725-737.

Barlett, P. N., J. M. Elliot, and J. W. Gardner. 1997. Electronic noses and their application in the food industry. *Food Technol* 51: 44-48.

Benady, M., J. E. Simon, D. J. Charles, and G. E. Miles. 1992. Determining melon ripeness by analysing headspace gas emissions. *Assoc. Soc. Agric. Eng. Paper* 92: p-6055, cited by Young et al. (1999).

Blankenship, M. S., M. Parker, and C.R. Unrath. 1997. Use of Maturity indices for predicting postharvest firmness of 'Fuji' apples. *HortScience*. 32: 909-910.

Blanpied, G. D. 1974. A study of indices for earliest acceptable harvest of 'Delicious' apples. *J. Amer. Soc. Hort. Sci.* 99: 537-539.

Blanpied, G. D., and W. J. Bramlage, D. H. Dewey, R. L. Labelle, L. M. Massey, G. E. Mattus, W. C. Stiles, and A. E. Watada. 1978. A standardized method for collecting apple pressure test data. *New York's Food and Life Sciences Bulletin* No. 74. 8 p. New York state College of Agriculture and Life Sciences, Ithaca, NY, cited by Kingston (1991).

Brackmann A. and J. Strief. 1994. Ethylene, CO₂ and aroma volatiles produced by apple cultivars. *Acta Horticulturae* 368: 51-58.

Brookfield, P. L., J. G. Hughes, and C. B. Watkins. 1993. Fruit maturity and quality of selectively picked 'Royal Gala' apples. *N. Z. J. Crop. Hort. Sci.* 21:183-187.

Brown, D. S., J. R. Buchanan, and J. R. Hicks. 1965. Volatiles from apple fruits as related to variety, maturity, and ripeness. *J. Amer. Soc. Hort. Sci.* 88: 98-104.

Cliff, M. A., O. L. Lau, and M. C. King. 1998. Sensory characteristics of controlled atmosphere-and air stored 'Gala' apples. *Journal of Food Quality.* 21: 239-249.

Cunningham, D. G., T. E. Acree, J. Barnard, R. M. Butts, and P. A. Braell. 1986. Charm analysis of Apple volatiles, *Food Chemistry* 19: 137-147.

Cyrano Sciences, Inc. 2000. *Cyranose 320 User's Manual*. Cyrano Sciences, Inc. Pasadena, CA.

Cyrano Sciences, Inc. 2001. [http:// www.Cyranoscience.com](http://www.Cyranoscience.com). Viewed in December 2001.

Cyranose Sciences, Inc. 2002. Telephonic conversation with the technical service personnel, Cyrano Sciences, Inc. Talked in December 16, 2002.

Dimick, P. S., and J. C. Hoskin. 1983. Review of apple flavor-state of art. *CRC Critical Reviews in Food Science and Nutrition.*18: 387-409.

Dirnick, P., H. De Pooter, and N. Schamp. 1989. Aroma development in ripening fruits. p. 23-34. In: R. Teranishi, R. G. Buttery, and F. Shahidid (eds.), Flavor Chemistry: Trends and Developments. ACS Symp. Ser. 388, Amer. Chem. Soc., Washington, D.C.

- Fidler, J. C. 1973. The optimum date of harvest. p.10-16. In: F. C. Fidler, B. G. Wilkinson, K.L. Edney, and R.O. Sharples (eds), The Biology of Apple and Pear Storage. Commonwealth Agr. Bureaux, Farnham Royal, Slough, England, cited by Kingston (1991).
- Harman, J. E., and C. B. Watkins. 1981. Fruit testing: Use of refractometers. Aglink HPP212, Media Services, MAF, Wellington, New Zealand, cited by Kingston (1991).
- Hikasa, Y., H. Muramatsu, and T. Minegishi. 1986. Effects of nitrogen application on the growth, yield and fruit quality of dwarf apple trees (in Japanese). Bull. Hokkaido Prefectural Agr. Expt. Sta. 55: 23-31 [Hort. Abstr. 58:63; 1988], cited by Kingston (1991)
- Haugen, J. E., and K. Kvaal. 1998. Electronic nose and artificial neural network. Meat-Sci. 49 (suppl.): S273-286.
- Inge, M., and M. C. D'Souza. 1989. Fruit characteristics of 'Red Delicious' apple strains during maturation and storage. J. Amer. Soc. Hort. Sci. 114: 776-780.
- Jackson, J. E. 1980. Light interception and utilization by orchard systems. Hort. Rev. 2: 208-267.
- Johnson, D. E. 1998. p.1-8, 93-105, 113-119, 217-164, 494-511. Applied Multivariate Methods for Data Analysis, Duxbury Press, Brooks/Cole Publishing Company, Pacific Grove, CA.
- Kingston, C. M. 1991. Maturity indices for apple and pear. Hort. Rev. 13: 407-432.

Knee, M., and S. M. Smith. 1989. Variation in quality of apple fruits stored after harvest on different dates. *J. Hort. Sci.* 64: 413-419.

Keener, K. M., R. L. Strohine, and J.A. Nyenhuis. 1999. Evaluation of low field (5.3 – MHz) proton magnetic resonance measurement of D_w and T_2 as methods of non-destructive quality evaluation of apples. *J. Amer. Soc. Hort. Sci.* 124: 289-295.

Kupferman, E. M. 1986. Sampling for accurate analysis. p. 17-18. Apple Maturity Program Handbook, Apple Maturity Program, Wenatchee, WA, cited by Kingston (1991).

Lachenburg, P.A., and M.R. Mickey. 1968. Estimation of Error Rates in Discriminant Analysis. *Technometrics*, 10 no.1. p. 1-11, cited by Johnson (1998).

Mann, S. S., and B. Singh. 1986. Studies on changes during development and maturation of pear fruit. p. 251-254. In: T. R. Chadha, V. P. Bhutani, and J. L. kaul (eds), Advances in Research on Temperate Fruits. Dr. Y. S. Parmar Univ. of Horticulture and Forestry, Solan, India.

Mattheis J.P., D.A. Buchanan, and J.K. Fellman. 1998. *J. Amer. Soc. Hort. Sci.* 123: 426-432.

Marini, R. P. 2003. Discussion with Dr. Richard P. Marini on February 20, 2003, Department of Horticulture, Virginia Polytechnic Institute and State University.

Miller, J. 2001. USApple focusing on health qualities of apples. *The Produce News*. 104 (36). <http://theproducenews.com/storydetail.cfm?ID=1472>. Viewed in January 2003.

Friendly, M. 1991. Robust Multivariate Outlier Detection. SAS System for Statistical Graphics, First Edition. SAS Institute Inc., Cary, NC, USA.

Muramatsu, N., N. Sakurai, N. Wada, R. Yamamoto, K. Tanaka, T. Asakura, Y. I.

Takano. 1999. Remote sensing of fruit textural changes with a laser doppler vibrometer. *J. Amer. Soc. Hort. Sci.* 125: 120-127.

Olsen, K. L., and G. C. Martin. 1980. Influence of apple bloom date on maturity and storage quality of 'Starking Delicious' apples. *J. Am. Soc. Hort. Sci.* 105:183-186.

Olsen, K. L., R. D. Bartram, and M. W. Williams. 1986. 'Golden Delicious' apples: Productivity and quality with adequate but modest leaf nitrogen levels. p. 11-16. *Apple Maturity Program Handbook*, Apple Maturity Program Wenatchee, WA, cited by Kingston (1991).

Pathange, L.P., P. Mallikarjunan, S. F. O' Keefe, and R. P. Marini. 2002. Electronic nose technology: a non-destructive technique to evaluate apple quality. Abstract presented at the Annual Meeting & Food Expo, Institute of Food Technologists, Anaheim, CA June 15-19, 2002: 91D-2.

Paillard, N. M. W. 1982. Factors influencing flavor formation in fruits. In: P. Schreier (Editor), Flavour Walter de Gruyter, Berlin, New York. p. 479-507, cited by Dirnick et al. (1989).

Peirs, A., J. Lammertyn, K. Ooms, B.M. Nicolai. 2000. Prediction of the optimal picking date of different apple cultivars by means of VIS/NIR- spectroscopy. *Postharvest Biology and Technology* 21: 189-199.

Phillips, W. R., and P.A. Poapst. 1952. Storage of apples. Horticultural Div. Exptal. Farm Service. Publication 776. Canada Dept. Agr., cited by Kingston (1991).

Poapst, P. A., G.M. Ward, and W.R. Phillips. 1959. Maturation of 'McIntosh' apples in relation to starch loss and abscission. *Can. J. Plant. Sci.* 39:257-263, cited by Kingston (1991).

Reid, M. S., C. A. S. Padfeild, C. B. Watkins, and J. E. Harman. 1982. Starch iodine pattern as a maturity index for 'Granny Smith' apples. 1. Comparison with flesh firmness and soluble solids content. *New Zealand J. Agr. Res.* 25: 229-237, cited by Kingston (1991).

Robinson, T. L., E. J. Seeley, and B. H. Barritt. 1983. Effect of light environment and spur age on 'Delicious' apple fruit size and quality. *J. Amer. Soc. Hort. Sci.* 108: 855-861.

Rouchard, J., C. Moons, and J. A. Meyer. 1983. Effects of selected fungicide treatments on the sugar content of 'Golden Delicious' and 'Jonagold' apples. *Scientia Hort.* 20:161-168.

Rourke, D. A. 1998. World apple marketing dynamics. U.S. paper presented at the 41st annual IDFTA Conference, February 21-25, 1998, Pasco, Washington.

Shaw, S. J., and R. N. Rowe. 1982. Sources of variability on apple maturity index data. *New Zealand Ag. Sci.* 16:51-53.

Shmulevich, I., N. Galili, and D. Rosenfeld. 1996. Detection of fruit firmness by frequency analysis. *Trans. Amer. Soc. Agr. Eng.* 39:1047-1055, cited by Muramatsu et al. (1999).

Simon, J. E., A. Hetzron, B. Bordelon, G. E. Miles, and D.J. Charles. 1996. Electronic Sensing of aromatic volatiles for sorting of blueberries. *J. Food Sci.* 61: 967-999, 972.

Smith, R. B., E. C. Loughheed, and E. W. Franklin, and I. McMillan. 1979. The starch iodine test for determining stage of maturation of apples. *Can. T. Plant Sci.* 59:725-735.

Smith, S. M. 1984. Improvement of aroma of Cox's orange pippin apples stored in low oxygen atmospheres. *J. Hortic. Sci.* 59: 515-522.

Smock, R. M. 1948. A study of maturity indices for 'McIntosh' apples. *Proc. Amer. Soc. Hort. Sci.* 45: 77-86, cited by Kingston (1991).

Song, J., and F. Bangerth. 1996. The effect of harvest date on aroma compound production from 'Golden Delicious' apple fruit and relationship to respiration and ethylene production. *Postharvest Biology and Technology* 8: 259-269.

Tugwell, B. L., 1998. Maturity the key to quality for apples, pears and cherries. *Proceedings of the 29th National Cherry Growers of Australia and the 53rd National*

Australian Apple and Pear Growers Association Conference. 16-20th August 1998.
Adelaide, SA. p.65-66.

Tutsin, D. S., P. M. Hirst, and I. J. Warrington. 1988. Influence of orientation and position of fruiting lateral on canopy light penetration, yield, and fruit quality of 'Granny Smith' apple. *J. Amer. Soc. Hort. Sci.* 113: 693-699.

U.S. Apple Association. 2002. <http://www.usapple.org/consumerinfo>. Viewed in August 2002.

U.S. Department of Commerce. 1994. Statistical abstracts of the U.S. section 23-Agriculture 114: 689-689, cited by Keener et al (1999).

U.S. Department of Commerce. 2000. Statistical abstracts of the U.S. section 17-Agriculture No.835. p. 540-540.

Van Deventer D. 2001. Discrimination of retained solvent levels in printed food-packaging using electronic nose systems. Masters Thesis, Virginia Polytechnic Institute and State University.

Vangdal, E. 1982. Eating quality of pears. *Acta. Agr. Scand.* 33: 135-139, cited by Kingston (1991).

Vangdal, E. 1985. Quality criteria for fruit for fresh consumption. *Acta. Agr. Scand.* 35: 41-47.

Volz, R. K., T.B. Ferguson, E.B. Hewett, and D.J. Woolley. 1994. Wood age and leaf area influences fruit size and mineral composition of apple fruit. *J. Hort. Sci.* 69: 385-395, cited by Volz et al. (1995).

Volz, R. K., J.W. Palmer, and H.M. Gibbs. 1995. Within-tree variability in fruit and maturity for 'Royal Gala' apple. *Acta Horticulturae* 379: 67-71.

Walsh, C. S., and S. A. Altman. 1993. Measurement of the ethylene evolution rate of apples during maturation, and its correlation with some-used maturity indices. *Acta Horticulturae (ISHS)* 343: 51-52.

Warner, G. 1991. Researcher wants flavor test at harvest. *The Good Fruit Grower* 46: 16-16.

Watada, A. E., R. C. Herner, A. A. Kader, R. J. Romani, G. L. Staby. 1984. Terminology for the description of developmental stages of horticultural crops. *HortScience*.19: 20-21.

Williams, A. A., O. G. Tucknott, and M. J. Lewis. 1977. 4-methoxyallyl benzene: an important aroma component of apples, *J. Sci. Food Agric.* 28: 185-190, cited by Dimick and Hoskin (1983).

Wills, R.B.H., P.A. Bambridge, and K.J. Scott. 1980. The use of flesh firmness and other objective tests to determine consumer acceptability of 'Delicious' apples. *Austral. J. Expt. Agr. Anim. Husb.* 20: 252-256, cited by Kingston (1991).

White, J. W. Jr., 1950. Composition of a volatile fraction of apples. *Food Res.* 15:68, cited by Dimick and Hoskin (1983).

Young, H., J. M. Gilbert, S. H Murray, and R. D. Ball. 1996. Casual effects of aroma compounds on royal gala apple flavors. *J. Sci. Food Agric.* 71: 329-336.

Young, H, K. Rossiter, M. Wang, and M. Miller. 1999. Characterization of Royal Gala apple aroma using electronic nose technology- potential maturity indicator. *J. Agric. Food. Chem.* 47: 5173-5177.

Appendix-A

Conversion of the mole % to milliliters

1) 1 mole of any gas = 22.4 l

100 mole of air contains = 1.15 mole of ethylene (standard – given by Airgas, Theodore, AL)

So 1 l of air contains = 0.0115 l of ethylene (standard)

1 ml of air contains = 0.001151 ml of ethylene

Average ethylene in apples in immature apples was 2.95 % mole calculated by calibration of the gas partitioner by standard ethylene.

1 ml of air contains = $2.95 * 0.01151 * 5$ (where 5 is weight factor, since the fruit Weight was in grams)

= 0.16 ml / kg of fruit for 26 hours.

1 ml of air contains = 6.14 $\mu\text{L}/\text{kg}/\text{hr}$.

Appendix-B

Maturity indices data of York and Gala apples

Gala maturity indices data

Obs	Puncture strength (N)	Soluble solids (°Brix)	Starch	pH	Ethylene
1	145.75	12.5	2	3.61	0.142394
2	128.184	15	2	3.85	0.155982
3	134.309	13	2	3.86	0.201541
4	122.157	16	3	3.86	0.172185
5	154.007	15	1	3.77	0.112366
6	134.015	15	3	3.83	0.0987
7	127.547	13	2	3.81	0.091823
8	149.205	13	1	3.63	0.214207
9	150.136	15	2	3.76	0.184729
10	138.033	14	3	3.83	0.217778
11	146.02	17.5	2	3.78	0.222259
12	146.657	15	1	3.9	0.283806
13	155.967	15	3	3.71	0.079578
14	79.919	14.5	4	3.72	0.263518
15	139.062	14	2	3.85	0.241265
16	132.006	16	1	3.87	0.127668
17	131.418	16	3	3.94	0.212225
18	150.626	16	4	3.73	0.078141
19	142.639	14.5	4	3.86	0.23868
20	140.091	14.5	3	3.96	0.112838
21	116.326	14.4	6	3.91	0.165295
22	143.57	13.9	3	3.8	0.102935
23	146.608	15.1	2	3.52	0.11151
24	129.507	15.3	4	3.69	0.150334
25	132.496	16	3	3.79	0.065401
26	116.424	15.8	6	3.77	0.083187
27	143.864	14.4	4	3.63	0.154326
28	134.358	14	4	3.57	0.103221
29	125.783	14.9	2	3.59	0.103937
30	110.152	15.5	4	3.72	0.279651
31	106.967	16.1	9	3.69	0.334541
32	119.56	12.8	5	3.76	0.068067
33	112.504	15.3	9	3.58	0.120102

34	140.434	13.3	3	3.5	0.279248
35	117.453	14.5	6	3.56	0.231831
36	135.24	14.9	6	3.54	0.075916
37	128.772	15.7	6	3.45	0.279984
38	105.154	15.6	8	3.52	0.263844
39	139.699	14.1	4	3.57	0.222597
40	118.237	15.6	6	3.47	0.029311
41	116.571	13.6	7	3.58	0.174535
42	84.035	13.4	8	3.64	0.106359
43	98.49	13	6	3.67	0.126425
44	100.842	13	7	3.62	0.162499
45	107.8	13	5	3.66	0.161388
46	84.182	13.5	9	3.74	0.153588
47	95.942	14	9	3.71	0.185493
48	79.282	14.6	9	3.76	0.191332
49	102.655	13.9	7	3.58	0.167162
50	94.325	14.3	8	3.53	0.113089
51	77.714	14.5	9	3.85	0.36134
52	122.402	13.9	4	3.57	0.153634
53	117.992	16.4	6	3.51	0.252423
54	113.484	16.6	7	3.67	0.307314
55	116.767	15.44	6	3.44	0.239001
56	113.925	13.4	3	3.57	0.179964
57	80.164	13.3	9	3.88	0.174926
58	89.817	15.3	7	3.74	0.160362
59	96.285	15.1	6	3.45	0.302294
60	68.747	16.7	9	3.88	0.298878

York maturity indices data

Obs		Starch Firmness (kg-f)	Soluble solids °Brix	Ethylene
1	3	98.637	11.3	0.128862
2	1	104.664	10.4	0.091118
3	2	96.824	11.5	0.110672
4	1	106.477	11.4	0.095683
5	1	96.187	10.9	0.070521
6	1	103.782	13.2	0.102016
7	1	93.786	11.6	0.153303
8	1	94.521	11.3	0.090915
9	1	96.971	11.3	0.078792
10	2	87.171	12.1	0.079269
11	2	96.432	11.5	0.137828
12	1	96.04	11.8	0.144031

14	2	81.144	12.4	0.264491
15	2	83.643	12	0.184233
17	1	100.156	12.8	0.207356
19	1	82.663	14.8	0.144207
20	2	83.3	11.8	0.07135
22	2	70.658	12.7	0.111725
23	2	106.526	13.7	0.09314
25	2	100.695	14.5	0.106116
26	3	77.91	12.1	0.064466
27	2	95.011	14.1	0.098724
28	1	93.835	13.9	0.095883
29	2	74.823	16.1	0.195221
30	3	109.711	13	0.067738
32	2	96.432	14.1	0.173957
33	1	101.283	13.1	0.190978
39	2	103.782	14.6	0.124897
50	1	94.619	13.4	0.100643
55	4	114.219	13.6	0.091251
13	2	79.87	11.7	0.099119
16	1	75.852	10.8	0.067125
18	1	83.251	12.3	0.077066
21	2	76.881	13	0.109021
31	4	89.67	15.4	0.295332
34	3	79.821	12.4	0.10415
36	2	86.534	14.1	0.126935
37	4	74.382	13.1	0.076491
40	1	84.966	12.9	0.095716
41	2	72.324	14.9	0.096621
42	4	79.87	13.6	0.071518
44	2	76.783	12.8	0.053909
45	3	73.01	14.3	0.094968
46	2	78.547	14.1	0.076339
47	5	81.193	13.1	0.130827
49	3	73.255	13.6	0.151303
51	4	78.743	12.4	0.162108
52	6	62.377	15.6	0.173878
53	2	77.812	11.5	0.128265
54	2	89.523	13	0.130417
58	4	72.961	13.3	0.233314
59	2	85.211	14.1	0.156232
60	3	76.832	13.3	0.14151
61	4	69.972	12	0.133662
63	2	91.679	12.8	0.149646
64	6	75.558	14.6	0.127552
65	3	78.547	15.9	0.151908
68	3	82.173	14.6	0.013147
24	4	85.848	13.1	0.166615
35	4	87.465	12.5	0.129904

38	5	73.892	13.3	0.152112
43	4	77.714	14.6	0.149299
48	4	75.803	12.3	0.088875
56	3	72.765	12.5	0.133467
57	2	85.211	14.1	0.124409
62	5	78.743	13.9	0.133178
66	4	78.988	14.8	0.165831
67	3	80.997	12.5	0.135354
69	4	77.42	15.6	0.156237
70	5	75.362	15.1	0.326728
71	6	65.611	14.1	0.170967
72	7	66.885	12.8	0.163156
73	8	88.494	12.8	0.089613
74	3	77.42	15.7	0.11233
75	7	67.326	15.3	0.176182

Appendix-C

Color analysis

Color had three parameters L, C and H that were equally important, and thus all three had to be included in the analysis. But such an analysis might be biased towards color because the remaining three maturity indices were represented by only one parameter. PCA was performed using the 3 color parameters as the three variables. Only the first principal component was used to represent color in subsequent analysis. It was later discovered that the color values were incorrect and so the color index was removed from the subsequent analysis of maturity indices.

Appendix-D

Correlation matrices of maturity indices

Gala correlation matrix of five maturity indices

Correlation Matrix	Soluble solids	Firmness	pH	Starch	Ethy	A1	A2	A3	A4	A5
Sol. solids	1.00	0.06	0.02	0.04	0.21	-0.05	0.08	-0.19	-0.04	0.31
Firmness	0.06	1.00	0.13	-0.81	-0.26	-0.48	-0.48	-0.71	-0.58	-0.29
pH	0.2	0.13	1.00	-0.36	0.04	-0.47	-0.51	-0.09	-0.53	-0.43
Starch	0.04	-0.81	-0.36	1.00	0.17	0.69	0.69	0.67	0.71	0.48
Ethylene	0.21	-0.26	0.08	0.02	1.00	0.18	0.38	0.19	0.30	0.42
A1	-0.05	-0.48	-0.47	0.69	0.18	1.00	0.85	0.54	0.79	0.51
A2	0.08	-0.48	-0.51	0.69	0.38	0.85	1.00	0.57	0.90	0.70
A3	-0.19	-0.71	-0.09	0.67	0.19	0.54	0.57	1.00	0.67	0.26
A4	-0.04	-0.58	-0.53	0.71	0.30	0.79	0.90	0.67	1.00	0.70
A5	0.31	-0.29	-0.43	0.48	0.42	0.51	0.70	0.26	0.70	1.00

* A1, A2, A3, A4, A5 represents aroma compounds hexyl butyrate, hexyl 2-methylbutanoate, butyl hexanoate, hexyl hexanoate and alpha-farnesene, respectively.

York correlation matrix of four maturity indices

Correlation Matrix	Starch	Firmness	Soluble solids	Ethylene
Starch	1.00	-0.51	0.36	0.30
Firmness	-0.51	1.00	-0.30	-0.30
Soluble solids	0.36	-0.30	1.00	0.48
Ethylene	0.21	-0.21	0.33	1.00

VITA

The author, Lakshmi Prasad Pathange, was born on January 27, 1978, in Andhra Pradesh, India. He graduated with a Bachelor of Technology degree in Food Processing and Preservation Technology from College of Technology, Osmania University, in May 2000. He began pursuing his Master of Science degree in Biological Systems Engineering at Virginia Polytechnic Institute and State University in January 2001.