

**DEVELOPMENT OF FAST AND SIMPLE
ANALYTICAL METHODS FOR THE
DETERMINATION OF HONEY ADULTERATION
AND FORGERY BASED ON CHEMOMETRIC
MULTIVARIATE DATA ANALYSIS BY USING
MOLECULAR SPECTROSCOPY**

**A Thesis Submitted to
the Graduate School of Engineering and Sciences of
İzmir Institute of Technology
in Partial Fulfillment of the Requirements for the Degree of**

MASTER OF SCIENCE

in Chemistry

**by
Başak BAŞAR**

**December 2016
İZMİR**

We approve the thesis of **Başak BAŞAR**

Examining Committee Members:

Prof. Dr. Durmuş ÖZDEMİR

Department of Chemistry, İzmir Institute of Technology

Doç. Dr. Hasan ERTAŞ

Department of Chemistry, Ege University

Prof. Dr. Figen KOREL

Department of Food Engineering, İzmir Institute of Technology

28 December 2016

Prof. Dr. Durmuş ÖZDEMİR

Supervisor, Department of Chemistry,
İzmir Institute of Technology

Prof. Dr. Ahmet Emin EROĞLU

Head of the Department of Chemistry

Prof. Dr. Bilge KARAÇALI

Dean of the Graduate School of
Engineering and Sciences

ACKNOWLEDGEMENTS

There are many people who made this thesis possible and I am pleased to thank them here.

I am very much thankful to my advisor, Prof. Dr. Durmuş ÖZDEMİR for his valuable guidance, patience and encourage during the whole period of my master thesis and for everything I've learned from him in the last two and a half years. I think I was very lucky to be a member of chemometrics research group.

Besides my advisor, I would like to thank to rest of my thesis committee: Prof. Dr. Figen KOREL and Assoc. Prof. Dr. Hasan ERTAŞ.

I am pleased to acknowledge TUBITAK for the financial support.

Thanks to Ordu Apiculture Research Institute for their technical support and providing pure and adulterated honey samples.

Special thanks to the members of Chemometrics Research Team, Ayten Ekin MEŞE and Gün Deniz AKKOÇ, for their patience and helps and being with me all the time when I need.

My sincere thanks to İpek HARMANLI, Buket KAN, Pınar BAŞAR and Özlem KARABOĞA for their motivations and support every time.

Finally, I would like to thank you my parents, Zahide BAYRAK and Reşat BAŞAR, for their endless support and confidence throughout my life.

ABSTRACT

DEVELOPMENT OF FAST AND SIMPLE ANALYTICAL METHODS FOR THE DETERMINATION OF HONEY ADULTERATION AND FURGERY BASED ON CHEMOMETRIC MULTIVARIATE DATA ANALYSIS BY USING MOLECULAR SPECTROSCOPY

Honey is one of the most valuable and expensive nutrition due to its health effects on human body. In recent years, honey adulteration has become an important problem and is a subject of many publications. There exists various analytical methods for determination of honey adulteration with $^{13}\text{C}/^{12}\text{C}$ isotope ratio mass spectrometry (IR-MS) being the most common. However, one of the recent studies indicates that different honey types depending on geographical and botanical origin may have significantly different $^{13}\text{C}/^{12}\text{C}$ isotope ratios rendering this method questionable. Thus, development of an analytical method for qualitative and quantitative determination of forgery and adulteration of honey without tedious and complicated sample preparation while being relatively simple and fast new analytical methods became a must. In this study, Fourier Transform Infrared spectroscopy coupled with Attenuated Total Reflectance and Fourier Transform Near Infrared spectroscopy based chemometrics multivariate calibration models were developed for the quantitative determination of honey adulteration. To simulate adulteration scenarios, artificially adulterated honey samples were prepared by adding beet sugar, corn syrup, glucose and sucrose with various concentrations to pure honey samples. Two different multivariate calibration methods namely Genetic Inverse Least Squares and Partial Least Squares were used and the applicability of these methods have been evaluated with an independent validations and test set composed of FTIR spectra of more than 100 pure honey samples along with the adulterated samples. Standard error of cross validation and standard error of prediction values for honey content of the samples were found 2.52% and 2.19% (w/w %), respectively.

ÖZET

BALDAKİ SAHTECİLİĞİN VE TAĞŞIŞIN BELİRLENMESİ İÇİN KEMOMETRİK VERİ ANALİZİNE DAYALI BASİT VE HIZLI MOLEKÜLER SPEKTROSKOPİK METOTLARIN GELİŞTİRİLMESİ

Bal mali değeri yüksek ve insan sağlığı üzerinde önemli etkileri olan bir besindir. Son yıllarda, bazı üreticiler tarafından kar oranını yükseltmek amacı ile en çok tağşışe maruz kalan besinlerden de biri haline gelen bal ile ilgili çeşitli analitik çalışmalar yaygın olarak yürütülmektedir. Söz konusu bu sahteciliklerde yaygın olarak mısır şurubu ve pancar şekeri gibi yapay tatlandırıcılar kullanılmakla beraber, kullanılan analitik yöntemlerle tağşışi nitel ve nicel olarak belirlemek mümkün hale getirilmektedir. Bu analitik yöntemlerden, izotop oranı kütle spektrometresi (IR-MS) yaygın olarak kullanılmakta ve bu yöntemde $^{13}\text{C}/^{12}\text{C}$ oranı baz alınmaktadır. Ancak son dönemlerdeki çalışmalara göre farklı coğrafik ve botanik orijinlere sahip saf bal örneklerinin de birbirinden farklı $^{13}\text{C}/^{12}\text{C}$ izotop oranlarına sahip olabildiği, dolayısıyla kantitatif olarak tağşış eden miktarı tespitinin oldukça zor olduğu vurgulanmıştır. Bu sebeple nicel ve nitel olarak tağşışli bal örneklerini tespit edebilmek amacı ile hızlı ve kolay uygulanabilir analitik yöntemler geliştirilmesi gerektiği sonucuna varılmıştır.

Bu çalışmada, tağşışli bal örneklerinin miktarsal tayini için Fourier dönüşümlü kızılötesi spektroskopisi ve Yakın kızılötesi spektroskopisi kullanılarak alınan spektral verilere kemometrik çok değişkenli kalibrasyon metotları uygulanarak yeni bir analitik metot geliştirilmesi hedeflenmiştir. Çalışmada, pancar şekeri, mısır şurubu, glikoz ve sakkaroz saflığı bozan tatlandırıcılar olarak kullanılmış ve farklı veri setlerinde kullanılan bu maddelerle çeşitli tağşış senaryoları hazırlanmıştır. Genetik algoritma tabanlı ters en küçük kareler ve kısmi en küçük kareler yöntemi kullanılarak çok değişkenli kalibrasyon modelleri oluşturulmuş ve bu modeller üretici ve marketlerden temin edilen 100 adet saf bal örneği ile test edilmiştir. Bal bileşenine ait kalibrasyon ve validasyon setlerine ait hata değerleri %2.52 ve %2.19 aralığında tanımlanmıştır.

TABLE OF CONTENTS

LIST OF FIGURES	viii
LIST OF TABLES.....	xi
CHAPTER 1.INTRODUCTION	1
1.1. Scope of Thesis	2
1.2. Literature Review	2
CHAPTER 2.INSTRUMENTATION	8
2.1. Infrared Spectroscopy (IR).....	8
2.1.2. Attenuated Total Reflectance (ATR)	9
2.2. Near Infrared Spectroscopy (NIR)	11
CHAPTER 3.MULTIVARIATE CALIBRATION METHODS.....	12
3.1. Classical Least Squares (CLS)	12
3.2. Inverse Least Squares (ILS)	13
3.3. Partial Least Squares (PLS).....	15
3.4. Genetic Inverse Least Squares (GILS).....	17
3.4.1. Algorithm Options.....	18
3.4.2. Initialization of Gene Pool	18
3.4.3. Evaluation of the Genes in the Population	19
3.4.4. Selecting Parent Genes to be Breed	20
3.4.5. Cross-Over of the Selected Genes.....	21
3.4.6. Replacing Offsprings with Parents.....	21
3.4.7. Prediction of the New Data and Frequency Plots	21
CHAPTER 4.EXPERIMENTATION	23
4.1. Sample Collection	23
4.2. Sample Preparation	25
4.3. Data Collection.....	34
4.4. Data Processing & Chemometrics Model Construction	35
4.5. Design of Data Sets	36
CHAPTER 5.RESULTS AND DISCUSSION.....	42

5.1. The First Adulteration Scenario (Beet Sugar and Water were Used as Adulterants)	44
5.2. The Second Adulteration Scenario (Corn Syrup and Water were Used as Adulterants)	52
5.2.1. FTIR-ATR Results of Second Adulteration Scenario	52
5.2.2. FTNIR Results of Second Adulteration Scenario	57
5.3. The Third Adulteration Scenario (Corn Syrup, Beet Sugar and Water were Used as Adulterants)	64
5.4. The Fourth Adulteration Scenario (Glycose, sucrose and Water Used as Adulterants)	81
5.5. The Combined Adulteration Scenario	92
 CHAPTER 6.CONCLUSION	 102
 REFERENCES	 103

LIST OF FIGURES

<u>Figure</u>	<u>Page</u>
Figure 2.2. Schematic illustration of the ATR principle.....	10
Figure 3.1. Schematic representation of the matrices used in PLS method along with their sizes.....	16
Figure 3.2. A sample plot of number of components vs. PRESS values.....	17
Figure 3.3. Schematic illustration of roulette wheel selection method.....	20
Figure 4.1. The map of Turkey in which black areas show the geographical regions of authentic honey samples used in this study.	23
Figure 4.2. A small subset of the collected honey samples	24
Figure 4.3. Examples of synthetically prepared adulterated honey samples.	25
Figure 4.4. Percent concentration of the honey, beet sugar and water contents (w/w %) of the samples in the first scenario.	27
Figure 4.5. Percent concentration of the honey, corn syrup and water contents (w/w %) of the samples in the second scenario.....	30
Figure 4.6. Percent concentration of the honey, corn syrup, beet sugar and water contents (w/w %) of the samples in the third scenario.....	32
Figure 4.7. Percent concentration of honey, glyucose, sucrose and water contents (w/w%) of the samples in the fourth scenario.....	34
Figure 5.1. The spectra of 6 different botanical origins that are thyme, chestnut, pine, flower-pine, flower and cotton-flower honey collected by using FTIR spectroscopy coupled with ATR accessory.	42
Figure 5.2. The spectra of 139 honey samples that belong to different botanical and geographical origin by using FTIR spectroscopy coupled with ATR accessory.	43
Figure 5.3. FTIR-ATR spectra of 6 authentic and 69 adulterated honey samples prepared with two adulterants (beet sugar and water).....	44
Figure 5.4. Actual versus predicted plot of honey, beet sugar and water contents resulting from GILS in the first scenario.	45
Figure 5.5. Actual versus predicted plots of honey, beet sugar and water contents resulted from PLS in the first scenario.....	47
Figure 5.6. FTIR-ATR spectra of 50 authentic honey samples which were introduced to the multivariate calibration models to test their success.	49
Figure 5.7. Honey content prediction of 50 authentic honey samples with both GILS and PLS methods.....	49

Figure 5.8. Beet sugar and water contents prediction of 50 authentic honey samples with both GILS and PLS methods.....	51
Figure 5.9. FTIR-ATR spectra a total of 65 authentic and adulterated honey samples prepared with corn syrup.....	52
Figure 5.10. Actual versus predicted plot of honey and corn syrup contents resulted from GILS in the second scenario.	53
Figure 5.11. Actual versus predicted plot of honey and corn syrup contents resulting from PLS model in the second scenario.	54
Figure 5.12. FTIR-ATR spectra total of 100 authentic and commercial honey samples.....	55
Figure 5.13. Honey content prediction of 100 authentic and commercial honey samples with both GILS and PLS methods.	56
Figure 5.14. Prediction of corn syrup content of 100 authentic and commercial honey samples with both GILS and PLS methods.	57
Figure 5.15. FTNIR spectra of the 65 authentic and adulterated honey samples.	58
Figure 5.16. Actual versus predicted plots of honey and corn syrup resulting from GILS in the second scenario with FTNIR spectroscopy.....	59
Figure 5.17. Actual versus predicted plots of honey and corn syrup contents resulting from PLS in the second scenario with FTNIR spectroscopy.....	60
Figure 5.18. FTNIR spectra of 35 authentic and commercial honey samples.....	61
Figure 5.19. Honey content prediction of 35 authentic and commercial honey samples with both GILS and PLS methods with FTNIR spectroscopy.....	62
Figure 5.20. Corn syrup content prediction of 35 authentic and commercial honey samples with both GILS and PLS models using FTNIR spectroscopy.	63
Figure 5.21. FTIR spectra of 103 adulterated and authentic honey samples prepared with corn syrup, beet sugar and water in the third adulteration scenario.	64
Figure 5.22. Fingerprint region of FTIR spectra between 1500 and 900 cm^{-1} wavenumber range of 103 adulterated and authentic honey samples prepared with corn syrup, beet sugar and water in the third adulteration scenario.	65
Figure 5.23. Actual vs. predicted plots of honey, corn syrup, beet sugar and water contents obtained from GILS method in the fourth scenario with FTIR spectroscopy.....	66
Figure 5.24. Frequency distribution plots that were obtained by performing GILS method for honey, corn syrup, beet sugar and water contents to develop multivariate calibration models.....	68
Figure 5.25. Actual vs. predicted plots of honey, corn syrup, beet sugar and water contents resulted from PLS method in the fourth scenario with FTIR spectroscopy.....	70

Figure 5.26. Honey content prediction of 100 authentic and commercial honey samples with both GILS and PLS methods with FTIR spectroscopy in the third adulteration scenario.....	72
Figure 5.27. Predicted corn syrup, beet sugar and water contents of 100 authentic and commercial honey samples with both GILS and PLS methods with FTIR spectroscopy.....	73
Figure 5.28. Comparison of prediction ability of the developed GILS and PLS models with corn syrup, beet sugar, water and honey contents.....	75
Figure 5.29. Standard residual and normal probability plots of the calibration, independent validation and test sets obtained from GILS method in the third adulteration scenario.	79
Figure 5.30. Standard residual and normal probability plots of the calibration, independent validation and test sets obtained from PLS method in the third adulteration scenario.	80
Figure 5.31. FTIR-ATR spectra of total of 58 authentic and adulterated honey samples prepared with using glucose, sucrose and water as adulterants in the fourth adulteration scenario.	81
Figure 5.32. Actual versus predicted plots of honey, glucose, sucrose, total glucose, total sugar, fructose and water contents resulted from GILS in the fourth scenario.	83
Figure 5.33. Actual versus predicted plots of honey, glucose, sucrose, total glucose, total sugar, fructose and water contents resulted from PLS in the fourth scenario.	86
Figure 5.34. Honey content prediction of 100 authentic and commercial honey samples with both GILS and PLS models.	89
Figure 5.35. Predicted concentrations of glucose, sucrose, water, total glucose, total sugar and fructose content in 100 authentic and commercial honey samples with GILS method.....	90
Figure 5.36. Predicted concentrations of glucose, sucrose, water, total glucose, total sugar and fructose content in 100 authentic and commercial honey samples with PLS method.....	91
Figure 5.37. Actual versus predicted plots of honey, total sugar and water contents resulted from GILS in the combined adulteration scenario.	96
Figure 5.38. Actual versus predicted plots of honey, total sugar and water contents resulted from PLS in the combined adulteration scenario.	98
Figure 5.39. Honey content prediction of 100 authentic and commercial honey samples with both GILS and PLS methods using FTIR spectroscopy in the combined adulteration scenario.	99
Figure 5.40. Total Sugar and water predictions of 100 authentic and commercial honey samples with both GILS and PLS models.	100

LIST OF TABLES

<u>Table</u>	<u>Page</u>
Table 2.1. Spectral Regions of Infrared Spectroscopy	8
Table 4.1. Percent composition of 75 pure, binary and ternary adulterated samples prepared with pure honey, beet sugar and pure water.	26
Table 4.2. Percent composition of 65 pure, binary and ternary adulterated samples prepared with pure honey, corn syrup and water.....	29
Table 4.3. Percent composition of 109 pure, binary, ternary and quaternary adulterated samples prepared with pure honey, corn syrup, beet sugar and water.....	31
Table 4.4. Percent composition of 58 pure and quaternary adulterated samples prepared with pure honey, glyucose, sucrose and water.....	33
Table 4.5. Calibration and independent validation sets of the first scenario.	37
Table 4.6. Calibration and independent validation sets of the second scenario.	38
Table 4.7. Calibration and independent validation sets of the third scenario.	40
Table 4.8. Calibration and independent validation sets of the fourth scenario.....	41
Table 5.1. Standard error of cross validation (SECV), standard error of prediction (SEP), maximum and minimum values of the components (Max and Min) and correlation coefficient (R^2) of GILS models belong to the first adulteration scenario.....	46
Table 5.2. The number of principle components (PC), standard error of cross validation (SECV), standard error of prediction (SEP) along with maximum and minimum operating ranges of the contents (Max and Min) and correlation coefficients (R^2) of PLS models belonging to the first adulteration scenario.....	48
Table 5.3. Standard error of cross validation (SECV), standard error of prediction (SEP), maximum and minimum operating range of components (Max and Min) and correlation coefficient (R^2) of GILS models belong to second adulteration scenario.....	54
Table 5.4. The number of principle components (PC), standard error of cross validation (SECV), standard error of prediction (SEP), maximum and minimum operating range of components (Max and Min) and correlation coefficient (R^2) of PLS models belonging to the second adulteration scenario.....	55

Table 5.5. Standard error of cross validation (SECV), standard error of prediction (SEP), maximum and minimum operating range of components (Max and Min) and correlation coefficient (R^2) of GILS models belong to second adulteration scenario with FTNIR spectroscopy.	59
Table 5.6. The number of principle components (PC), standard error of cross validation (SECV), standard error of prediction (SEP), maximum and minimum operating range of components (Max and Min) and correlation coefficient (R^2) of PLS models belong to second adulteration scenario with FTNIR spectroscopy.....	60
Table 5.7. Standard error of cross validation (SECV), standard error of prediction (SEP), maximum and minimum operating range of components (Max and Min) and correlation coefficient (R^2) of GILS models belong to the fourth adulteration scenario with FTIR-ATR spectroscopy.	67
Table 5.8. Number of principle components (PC), standard error of cross validation (SECV), standard error of prediction (SEP), maximum and minimum operating range of components (Max and Min) and correlation coefficient (R^2) of GILS models belong to the fourth adulteration scenario with FTIR-ATR spectroscopy.	71
Table 5.9. Predictions of binary mixtures of corn syrup with water and beet sugar with water from GILS method.....	74
Table 5.10. Predictions of binary mixtures of corn syrup with water and beet sugar with water from PLS method.....	74
Table 5.11. Honey, corn syrup, beet sugar and water contents of seven suspicious honey samples predicted by GILS method.	77
Table 5.12. Honey, corn syrup, beet sugar and water contents of seven suspicious honey samples predicted by PLS method.	77
Table 5.13. Standard error of cross validation (SECV), standard error of prediction (SEP), maximum and minimum operating range of contents (Max and Min) and correlation coefficient (R^2) of GILS models belong to the fourth adulteration scenario.	85
Table 5.14. The number of principle components (PC), Standard error of cross validation (SECV), standard error of prediction (SEP), maximum and minimum operating range of contents (Max and Min) and correlation coefficient (R^2) of PLS models belong to the fourth adulteration scenario.	88
Table 5.15. Percent compositions of 120 pure, binary, ternary and quaternary adulterated samples prepared with pure honey, corn syrup, beet sugar, glucose, sucrose and water.....	93

Table 5.16. Standard error of cross validation (SECV), standard error of prediction (SEP), maximum and minimum operating range of contents (Max and Min) and correlation coefficient (R^2) of GILS models belong to the combined adulteration scenario.	97
Table 5.17. The number of principle components (PC), standard error of cross validation (SECV), standard error of prediction (SEP), maximum and minimum operating range of contents (Max and Min) and correlation coefficient (R^2) of PLS models belong to the combined adulteration scenario.	99

CHAPTER 1

INTRODUCTION

Honey is produced from nectar of flowers and various types of plants by honeybees. It is a natural sweetener and it has a positive effect on human health. The nectars of flowers have been collected and transformed by honeybees with their own specific enzymes and left into the comb to mature and ripen. Honey has been considered to be one of the most valuable and widely consumed foods due to its antioxidant effect and the other beneficial properties for centuries, all over the world.

Due to the variety of geo-climate conditions, a great deal of plant and flower from different botanical origin has been grown in Turkey. Therefore, beekeeping has been done commonly and various types of honey have been collected around the world. Turkey has been known as the second largest honey producer after China and they are followed by Ukraine, USA and Russia. However, Turkey took the fourth place in exportation of honey to the world market.

The nutritional value of honey and gradually increase of consumer awareness for the consumption of natural and healthy food products in recent years makes it one of the most expensive food products. Therefore, honey has become one of the most adulterated foods with artificial sweeteners such as beet sugar, corn syrup, glucose, sucrose, etc. Adulteration has been described as a type of modification of foods and it is carried out by adding artificial materials to the pure ones to decrease their manufacturing cost and increase the amount of the product with cheaper substitutes.

On the other hand, these artificial sweeteners may lead to formation of toxic compound, 5-hydroxymethylfurfural (HMF) during the addition period as they require heating the honey and artificial sugar (such as, corn syrup, beet sugar and etc.) solutions in order to obtain a homogenous fake product which cannot be easily detected by the ordinary consumer. In addition, various pH and temperature conditions may promote the formation of HMF which carries a potential risk on human health as it is known to be a carcinogenic substance. To be more specific, this carcinogenic compound (HMF), which is an intermediate Maillard reaction (Hostalkova, Klingelhofer, and Morlock 2013), is more commonly observed in sugar rich-nutrients with artificial sweeteners,

especially under acidic conditions. Also, as the storage time increases, HMF formation is observed even in low temperatures.

Although HMF formation may also occur even in authentic honey samples, addition of artificial sweeteners to honey results in increased formation of this compound. As these facts highlights, to be able to distinguish both qualitatively and quantitatively natural honey from adulterated ones is essential for both protecting human health as well as to protect honest produces from an unfair price competition in local markets where the official control is not very strict. In addition, as it is valuable export product, it is important to ensure trustable commerce for the countries reputation.

1.1. Scope of Thesis

The aim of this study is to develop a simple method for the qualitative and quantitative determination of honey adulteration by using Fourier Transform Infrared Spectroscopy (FTIR) coupled with chemometrics multivariate calibration techniques. For this purpose, in addition to authentic honey samples, synthetically adulterated honey samples were prepared in laboratory and their FTIR-ATR spectra were collected to develop multivariate calibration models with Partial Least Squares (PLS) and Genetic Algorithm based Inverse Least Square (GILS) methods. The results of both GILS and PLS were compared from various adulteration scenarios such as binary, ternary and quaternary mixtures of synthetic sweeteners with pure honey.

1.2. Literature Review

In recent years, a great deal of studies has been carried out to distinguish adulterated honey samples from authentic ones. Adulterated honey samples includes various types of artificial sweeteners (sugar syrups, corn syrup, glyucose, sucrose, etc.) and in order to detect adulterated honey samples analytical techniques can be used as an alternative to the wet chemistry such as mass spectrometric, chromatographic and molecular spectroscopic methods. Among them, the most widely used one is isotope ratio mass spectrometry (IR-MS). Additionally, gas chromatography (GC), gas chromatography coupled with mass Spectrometry (GC-MS), liquid chromatography coupled with mass spectrometry (LC-MS), high performance liquid chromatography (HPLC), high performance liquid chromatography coupled with mass spectrometry

(HPLC-MS), nuclear magnetic resonance spectroscopy (NMR), high performance thin layer chromatography (HPTLC), Fourier transform infrared spectroscopy (FTIR), Raman spectroscopy, near infrared spectroscopy (NIR) and fluorescence spectroscopy, have been performed to detect adulterated honey samples.

Stable carbon isotope ratio mass spectroscopy (IR-MS) is based on $^{13}\text{C}/^{12}\text{C}$ ratio and $^{13}\text{C}/^{12}\text{C}$ ratio between honey and its protein fraction give qualitative and quantitative information about honey adulteration. C_3 (sugarbeet) and C_4 (sugarcane) plants belong to different photosynthetic cycles. Whereas authentic honey samples have characteristic properties of C_3 (sugarbeet) plants, artificial sweeteners have C_4 (sugarcane) plants characteristics. In literature, Simsek et al. were studied to detect adulteration in commercially available honey samples with isotope ratio mass spectrometry and elemental analyzer (EA-IRMS). A total of 31 authentic honey samples from different botanical origin and 43 commercial samples were collected for this study. The results of this study reveals that, more or less than 10 samples were found to be adulterated honey samples (Simsek, Bilsel, and Goren 2012). In another study, high fructose corn syrup (HFCS) and sucrose were used as artificial sweeteners and they were added to the authentic honey samples gathered from Brazil, Canada, Argentina and USA. The detection performance of adulterated samples was indicated with their detection limits and 6 Brazilian honey were found out as adulterated samples by Padovan and his research team (Padovan 2003). Another study was carried out by Cinar and his team in Turkey. In their study artificially adulterated honey samples were prepared by using one type of mono floral honey (honeydew honey) and high fructose corn syrup was added in different concentrations to the pure honey as adulterant. These adulterated samples were analyzed by using isotope ratio mass spectroscopy (IR-MS). Although 100 pure honeydew honey samples studied in this project, none of them had found significantly different from each other in terms of their isotope ratio between honey and its proteins (Cinar, Eksi, and Coskun 2014). Unlike previously mentioned studies with IR-MS method, Cengiz et al. aimed to validate their method developed by IRMS. Therefore, limit of detection (LOD) and limit of quantification (LOQ) of this method were determined. In order to test the predictive ability of the method 13 commercial honey samples were collected one of them was found out as frustrated honey sample (Cengiz, Durak, and Ozturk 2014). Guler et al. studied with 100 authentic and adulterated honey samples. In order to achieve adulterated honey samples honeybee colonies were fed

with synthetic sugars namely, high fructose corn syrup (HFCS), bee feeding syrup (BFS), glucose monohydrate sugar (GMS) and sucrose sugar (SS). These adulterated honey samples were evaluated in terms of their sugar type, sugar level and interaction of sugar type and the *P*-value of sugar level were found to be lower than 0.001 ($P < 0.001$) so this factor was defined as significant. Depending on their results, they finally promoted that isotope ratio mass spectrometry method could not detect indirect adulteration of honey, efficiently (Guler et al. 2014). Starch-based sugar syrup, sucrose syrup (SS), glucose syrup (GS) and high fructose corn syrup (HFCS) were used to prepare adulterated honey samples in an another study. The synthetically adulterated honey samples were analyzed with isotope ratio mass spectrometer coupled with elemental analyzer (EA-IRMS). Eventually they concluded that C_4 sugar syrups (GC and HFCS) could be detected but C_3 could not by EA-IRMS method (Tosun 2013).

Detection of adulterated honey samples with IR-MS method is a tedious, hard, expensive and time consuming method. Due to these drawbacks, alternative analytical methods were required and chromatographic and spectroscopic methods were also used to reveal adulterated honey samples. Chromatographic techniques have generally been performed coupled with mass spectrometry and they are called as GC-MS, HPLC-MS etc. Ruiz Matute and his research team was developed a method with GC-MS and their artificial honey samples were composed of high fructose inulin syrup (HFIS) in different concentrations (5%, 10%, 20% of HFIS) and pure honey and in all of the prepared samples inulotriose, fructose, inulobiose, kestoses, sucrose, dianhydrase of fructose (DAFs) were detected. Inulotriose was determined as the best marker for HFIS, and they said that among their test samples which were various types of pure honey, this marker was not detected in any of them by analyzing with G-MS method. (Ruiz-Matute et al. 2010). In another publication, the same research group was studied by using GC and GC-MS methods to investigate the detection of honey adulteration as well as comparing the performances of these analytical techniques. For this purpose, 20 different nectar and honeydew honey samples and 6 different syrups were collected. A method depended on yeast (*saccharomyces cerevisiae*) was developed and the adulterated samples which included adulterants in low concentration (about 5% (w/w %)) could be found out (Ruiz-Matute et al. 2007). In addition, Ultra-High Performance Liquid Chromatography coupled with time of flight mass spectrometry (UHPLC-TOF-MS) was used to detect adulterated honey samples in another study and their purpose

was defined as developing a method which could be used to detect more than one adulterant simultaneously to control quality and safety of the honey products (Du et al. 2015). Xue et al. developed a rapid, simple and effective method to figure out the adulterated samples with high- performance liquid chromatography with diode array detection (HPLC-DAD). Rice syrup was used as adulterant and with the help of NMR and MS methods the marker was found to be 2-acetylfuran-3-glucopyranoside (AFGP). The honey samples adulterated with 10% rice syrup could be detected by using the developed HPLC-DAD method (Xue et al. 2013). Another research group used high performance liquid chromatography coupled with isotope ratio mass spectrometry (HPLC-IRMS) for the same purpose. In this study determining ^{13}C isotope ratio of individual sugars (sucrose, glucose and fructose) was aimed. The values of ^{13}C in honey samples and the values of ^{13}C in protein were compared and a strong correlation between the ^{13}C ratio of individual sugars were observed. Since low level detection of adulteration was possible for the adulterated samples prepared with sugar, the method was declared as more sensitive method among the other IR-MS methods (CABANero, Recio, and Ruperez 2006).

On the other hand, vibrational spectroscopic techniques (Near Infrared, Mid Infrared, Raman Spectroscopy etc.) were required for testing food authenticity and quality control analyses in food industry. However, these analytical methods should be used along with special accessories and spectral data processing with chemometrics. Mishra et al. aimed to detect of Indian honey samples adulterated with jiggery syrup by using NIR spectroscopy. Data compression was processed by PCA method and a calibration model was developed by partial least square (PLS) method. Based on their results NIR Spectroscopy was defined as a successful and simple method for detection of honey adulteration (Mishra et al. 2010). Bazar et al. was also studied with NIR Spectroscopy in order to reveal differences in water spectral pattern among adulterated honey samples. In order to prepare the adulterated honey samples, four different authentic samples were mixed and high fructose corn syrup (HFCS) was used as adulterant in different concentration ranges. They used partial least squares (PLS) method to develop multivariate calibration model and they proved that NIR spectroscopy coupled with PLS method could detect adulterated honey samples successfully (Bazar et al. 2016). Rios-Corripio et al. was investigated honey adulteration with standard sugar solutions (glucose, fructose and sucrose) and also with cheap syrups (corn inverted and cane sugar) with NIR spectroscopy. By using the data of the

authentic and adulterated honey samples, 2D and 3D PCA score plots were generated and the differences between authentic and adulterated samples were determined resulting from this PCA scores. To determine the optimum concentration range of honey ingredients PLS was used as multivariate calibration method. Standard error of cross validation (SECV) values were found in 0.377–0.583 (w/w %) and standard error of prediction (SEP) values were also found in 1.550–3.150 (w/w %) for standard sugar solutions (Rios-Corripio, Rojas-López*, and Delgado-Macuil 2012). In addition, low field Nuclear Magnetic Resonance spectroscopy (LF ^1H NMR) and physicochemical analytical methods were performed by Ribeiro and his team to reveal the artificial samples prepared with HFCS. By using bi-exponential fitting, NMR data were analyzed and the physicochemical parameters were compared with the physicochemical parameters which were water activity, PH, and color. The results showed that adulterated samples, which were composed of different concentration of corn syrups, could be differentiated from the authentic ones due to their physical properties (Ribeiro et al. 2014). In 2012, Li and his research team were performed Raman spectroscopy for this purpose. The adulterants were selected to be high fructose corn syrup and maltose syrup and the adulterated samples were created by standard addition method, thus the concentration of these samples changed from 10% to 40% (w/w %). Adaptive iteratively reweighted penalized least square (airPLS) was used for baseline correction and PLS-LDA method was applied for classify the samples. The research team said that they could be separated the adulterated and non-adulterated honey samples by using Raman spectroscopy coupled with PLS-LDA method.(Li et al. 2012).

On the other hand, FTIR spectroscopy in conjunction with attenuated total reflectance (ATR) accessory is known to be another widely used analytical technique for this aim. Kelly at al. used 580 authentic honey samples in their study and fully inverted beet syrup, high fructose corn syrup, partially invert cane syrup, dextrose and beet sucrose was added to the pure honey samples in various mass percent to prepare adulterated honey samples. ATR-FTIR spectroscopy was used to analyze all of the samples in 4000-800 cm^{-1} region and SIMCA and PLS methods were used as multivariate calibration methods. While the adulterated samples with fully inverted beet syrup could not determine successfully, they indicated that none of the samples placed in the test set included HFCS (Kelly, Petisco, and Downey 2006). Besides, three different adulterants (corn syrup, high fructose corn syrup, inverted sugar) were used to prepare synthetic samples from authentic honey samples they were collected four

different geographical regions in Mexico. FTIR spectroscopy was used along with PLS method and an independent validation set were predicted by the developed calibration model, also these authentic honey samples were classified by using SIMCA and they claimed that these prediction and classification results were done correctly (Gallardo-Velázquez et al. 2009). Sivakesava et al. used FTIR-ATR spectra coupled with PLS method to detect cane medium invert sugar in pure honey samples belonged to three different botanical origins. LDA and CVA methods were developed to classify adulterated honey samples (Sivakesava and Irudayaraj 2001b). FTIR-ATR spectroscopic technique and chemometric multivariate calibration methods were combined to cane sugar adulteration in authentic honey samples. On the basis of data compressing principle component analysis PLS, LDA, CVA models were developed and validated. Additionally two different ANN methods (BPN, RBFN) were applied to these data sets and as a result of this study they found out that LDA was better than BPN method (Irudayaraj, Xu, and Tewari 2003). In a different study, adulterated samples were prepared with corn syrup, glyucose, fructose and sucrose contents. FTIR-ATR spectroscopy was used in conjunction with PLS method. According to the multivariate calibration models they aimed to predict the level of artificial sugar into the pure honey samples. At the end of the multivariate data analysis, they demonstrated that using FTIR spectroscopy to detect corn syrup adulteration in honey has good potential due to lower correlation (Sivakesava and Irudayaraj 2001c). Same research group performed another study by using inverted beet sugar as an adulterant. Prepared adulterated samples were scanned by FTIR-ATR spectrometry. They were selected 950-1500 cm^{-1} spectral region and applied PLS (first derivative) to develop multivariate calibration models. They demonstrated that FTIR spectroscopy was suitable for rapid detection of inverted beet sugar in honey samples (Sivakesava and Irudayaraj 2001a). Kelly et al. purposed detection of honey adulteration with FTIR spectroscopy coupled with attenuated total reflectance (ATR) accessory. A total of 320 (both authentic and adulterated with D-glucose and D-fructose solutions) honey samples were collected and prepared and their spectra recorded by FTIR-ATR spectroscopy. Data processing was studied using k-nearest neighbors (kNN) and partial least squares (PLS) regression on first derivative spectra. They promoted this methods were successful (Kelly, Downey, and Fouratier 2004).

CHAPTER 2

INSTRUMENTATION

2.1. Infrared Spectroscopy (IR)

Molecular spectroscopy has been used to investigate the interaction of electromagnetic waves and matters, with Infrared Spectroscopy being one of the most popular molecular spectroscopic techniques. Infrared region a wide spectral region and falls between visible (VIS) and microwave region, 12.800 to 10 cm^{-1} . This region is subdivided into three parts, Near Infrared region (NIR), Middle Infrared region (Mid-IR) and Far Infrared region (FIR), due to different instrumentation and application areas (Skoog and West 1980). These three regions and the most used infrared region are given in Table 2.1 in terms of wavenumber (ν), wavelength (λ), and frequencies (ν).

Table 2.1. Spectral Regions of Infrared Spectroscopy

Region	Wavelengths (λ), μm	Wavenumbers (ν), cm^{-1}
Near	0.75 to 2.5	12800 to 4000
Middle	2.5 to 50	4000 to 200
Far	50 to 1000	200 to 10
Most Used	2.5 to 15	4000 to 670

Before Fourier Transform was introduced, which was in 1980s, mid-IR region instruments were dispersive type instruments which allows for structural identification only. Quantitative analyses, however, became possible with the invention of FTIR spectroscopy which uses Michelson interferometer. Furthermore, spectral region of mid-IR can be divided in to two parts, functional region (4000- 1500 cm^{-1}) and fingerprint region (1500- 500 cm^{-1}). Mid-IR region where the spectral information obtained from vibrational and rotational stretching modes represents the chemical identity of a compound while the absorbance values are linearly correlated with the concentration. An illustration for the instrumentation of FTIR spectroscopy is given in Figure 2.1.

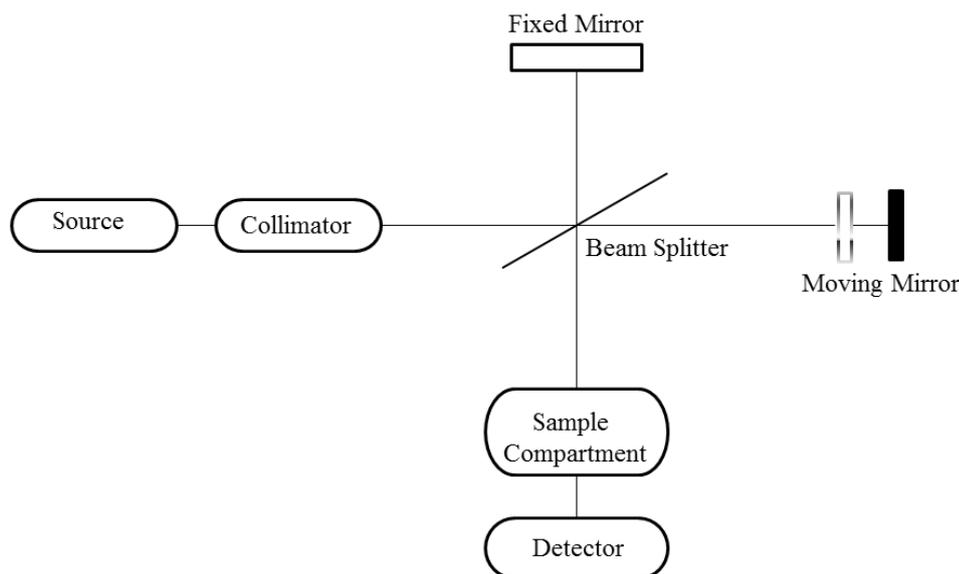


Figure 2.1. Simple block diagram for the instrumentation of FTIR Spectroscopy.

As an alternative to wet chemistry, FTIR spectroscopy being a non-destructive, easy, and cheap analytical technique which provides both qualitative and quantitative information makes it an invaluable tool for chemists. Various accessories can be attached to the FTIR instruments, such as, attenuated total reflectance (ATR) or grazing angle total reflectance (GATR) which allows analyses of samples with little or no sample preparation. (Karoui, Downey, and Blecker 2010).

Spectra obtained by FTIR spectroscopy can be processed with chemometrics methods which allows for more successful calibration and classification. Thus, this spectroscopic technique has the potential to fulfill the industrial needs for food quality and authenticity analyses making it suitable for this study.

2.1.2. Attenuated Total Reflectance (ATR)

Attenuated total reflectance (ATR) accessory is an accessory that is used along with FTIR spectroscopy. ATR accessory allows the analyses of solids, liquids and gel forms with no additional sample preparation procedures. There are various types of crystals used in ATR accessories, such as Zinc Selenide (ZnSe), Germanium and Diamond etc. Even though ZnSe crystal is widely used, cheaper and available for samples that are in liquid or gel forms, diamond crystal is known as the most successful ATR crystal due to its robustness and reliability. In addition, diamond crystal used in

ATR appears to be more expensive than the others, but it compensates this drawback by providing longer life-time. The crystal is positioned parallel on the upper surface of the ATR accessory and a schematic illustration of ATR accessories is given in Figure 2.2.

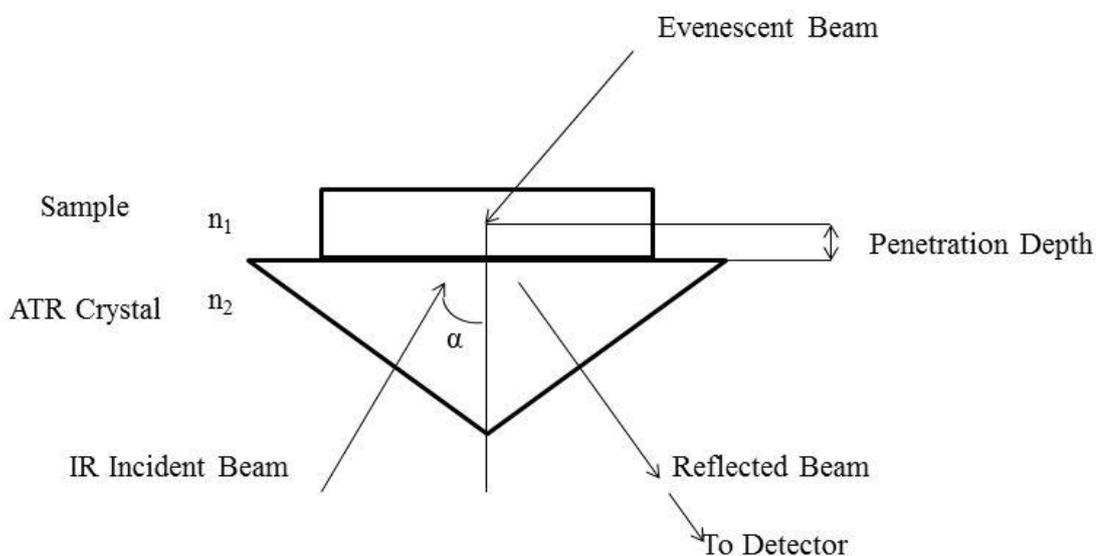


Figure 2.2. Schematic illustration of the ATR principle.

As illustrated in Figure 2.2, first the incident beam is passes through the crystal which creates an evanescent wave. This wave then penetrates the sample a few microns ($0.5 \mu - 5 \mu$) and is reflected to crystal. The beam continues to bounce between crystal and sample several times and finally exists the crystal to reach the detector. In addition, the number of reflection of the beam through each surface of the crystal changes depending on both the coming angle of the incident beam and the thickness and length of the ATR crystal. In order to obtain a successful IR spectrum having good contact between the sample and the crystal is a must. Moreover, if the sample to be measured is in solid form, application of some pressure might be necessary which can be achieved by an apparatus called pressure clump providing controllable amount of pressure on the sample.

2.2. Near Infrared Spectroscopy (NIR)

The near infrared region starts at the upper boundary of the visible region and spreads over a wide range area from 770 nm to 2500 nm. NIR spectroscopy has similarities with UV-vis spectroscopy in terms of optical systems (mirrors and sample holders). The reason of the interaction between the analyzed sample and the beam that comes from the source is overtones and combination of the fundamental vibrational transitions that occurs in mid-IR region around $3000\text{-}1700\text{ cm}^{-1}$. Since the bands are overtones and combination bands, they have low molecular absorptivity and 0.1 % detection limit. According to the quantum mechanics approach, energy transition between more than one vibrational energy states is forbidden. Even though these overtones are theoretically forbidden in quantum mechanics, these oscillations are seen due to their inharmonic character.

Furthermore, the absorption intensity of a NIR spectrum is weaker than a FTIR spectrum, since the probability of the mentioned forbidden transitions is less than the fundamental vibrational transitions. In addition, the absorption peaks of near infrared spectroscopy are wider than the peaks obtained from FTIR spectroscopy, because of various rotational transition states that exist between two energy states. The lower signal to noise ratio and the broad peaks makes employment of chemometric methods necessary. Moreover, NIR radiation has an important advantage in the quantitative analyses of food and biomedical samples containing water, fats and proteins. Although both diffuse reflection and transmission measurements can be used in NIR region, diffuse reflection measurements are most widely used(Skoog and West 1980).

CHAPTER 3

MULTIVARIATE CALIBRATION METHODS

Multivariate calibration takes advantage of multiple variables for constructing models to predict the interested properties of new samples unlike the univariate calibration which relies on a single variable. For instance, a spectral peak that is assumed to be linearly proportional with the concentration of a compound might be interfered with a peak of another compound. Also, in some cases the spectrum may be so complex that choosing a single wavenumber is not possible. Another point is the complementary information from the detector that are responses in other wavenumbers can be taken into account by using multivariate calibration techniques in order to enhance the predictive power of the model.

3.1. Classical Least Squares (CLS)

Classical least squares (CLS) is a calibration technique that is based on Lambert-Beer's law which explains the absorbance value as a function of concentration is shown in equation 3.1.

$$\mathbf{A} = \mathbf{C} \cdot \mathbf{K} + \mathbf{E}_A \quad (3.1)$$

Where \mathbf{A} is $m \times n$ matrix of absorbance values, \mathbf{C} is $m \times l$ matrix of concentration values of calibration set. Here, l is the number of components in a sample, m is the number of samples in calibration set and n is the number of wavelengths in a given spectrum. The unknown in this equation is the \mathbf{K} matrix which is $l \times n$ matrix of absorptivity coefficients that relates the absorbance values at n number of wavelengths to the concentrations of l components of the m number of calibration samples. The term \mathbf{E}_A stands for the absorbance residuals that are not fitted by the model equation. The least square solution of \mathbf{K} is given as (Equation 3.2.).

$$\mathbf{K} = (\mathbf{C}' \cdot \mathbf{C})^{-1} \cdot \mathbf{C}' \cdot \mathbf{A} \quad (3.2)$$

Here C' is the transpose of C matrix and superscript -1 on the upper right corner of the parenthesis stands for matrix inversion. Once the K matrix is obtained concentrations of the components in a given unknown sample can be predicted by Equation 3.3 as:

$$\mathbf{c} = (\mathbf{K} \cdot \mathbf{K}')^{-1} \cdot \mathbf{K} \cdot \mathbf{a} \quad (3.3)$$

Here \mathbf{c} is $l \times 1$ vector of component concentrations in a given unknown sample and \mathbf{a} is a $n \times 1$ matrix of absorbance values from the spectrum of the unknown sample. Despite of being quite straight forward and computationally inexpensive, there are some limitations of the CLS method. In order to have a successful model all the species that has a significant impact on the spectrum in the calibration samples must be known and included into the \mathbf{C} matrix, since the computed coefficients are actually based on complete composition of the samples. This is, however, not always feasible especially in the cases where some minor unknown components might present but can not be added to the concentration matrix. Thus, complex organic compounds containing thousands of major and minor components make it impossible for the CLS models predict a single property such as glucose in honey.

3.2. Inverse Least Squares (ILS)

Inverse Least Squares is based on inverse Lambert-Beer's law and assumes the concentration is a function of absorbances (Equation 3.4).

$$\mathbf{C} = \mathbf{A} \cdot \mathbf{P} + \mathbf{E}_C \quad (3.4)$$

Where \mathbf{C} and \mathbf{A} is the same as in the CLS. Here \mathbf{P} is $n \times l$ matrix of regression coefficients which relate the absorbance values to the concentrations of the components in the calibration set and \mathbf{E}_C is the matrix of concentration residuals. Similar to CLS, the solution of \mathbf{P} is calculated as (Equation 3.5) using pseudo-inverse

$$\mathbf{P} = (\mathbf{A}' \cdot \mathbf{A})^{-1} \cdot \mathbf{A}' \cdot \mathbf{C} \quad (3.5)$$

Unlike CLS, the ILS model has an advantage of modelling one component at a time as given in Equation 3.6.

$$\mathbf{p} = (\mathbf{A}' \cdot \mathbf{A})^{-1} \cdot \mathbf{A}' \cdot \mathbf{c} \quad (3.6)$$

Where \mathbf{p} is a $n \times 1$ vector of regression coefficient for the component being modelled and \mathbf{c} is a $m \times 1$ vector of concentrations of the component modeled in the calibration set. Once the \mathbf{p} vector is obtained, the predictions of an unknown sample are determined by using Equation 3.7.

$$c = \mathbf{a} \cdot \mathbf{p} \quad (3.7)$$

Among the calibration techniques, the reason that makes ILS more common is the ability to construct a model for single component without the knowledge of interfering species due to the coefficients that are to be multiplied with the absorbances rather than the concentrations. Another advantage is the assumption of the source of errors which are assumed to be in absorbances in CLS while ILS assumes they are caused by the concentration values which is a more common case because of the increasing precision and accuracy of analytical instruments compared to the personal errors during the sample preparation step. The human involving procedures such as dilution and sample preparation are another common source of errors on the concentrations side.

However, when the explanatory variables of the calibration set are spectral data that has typically more than 1000 variables, the downsides of the ILS methods become evident as there are so many collinear information on a given spectra. Therefore, the absorbance readings of close wavelengths (wavenumbers) are somewhat correlated causing a problem called multicollinearity. For highly correlated data, there might be an infinite number of solutions which solves \mathbf{P} or \mathbf{p} that minimizes the prediction errors. While this assures the best fit to the calibration set, the model may fail to predict the concentration of further samples. In other words, the model may overfit to the training set. Leaving out the absorbances at the wavenumbers which are expected to be irrelevant to the concentrations can be performed via feature selection methods such as genetic algorithms (GA) and least absolute shrinkage and selection operator (LASSO) in order to prevent the multicollinearity problem. Another very common solution to this issue is to using factor based projection methods such as principle component regression (PCR) and partial least squares (PLS) which are based on principle component analysis (PCA). A brief description of the Genetic algorithms and PLS methods is reported in following section.

3.3. Partial Least Squares (PLS)

The projection methods are very common tools for chemometricians because of the way the multicollinearity problem is handled while they also allow for multicomponent predictions. In principal components analysis (PCA), the data is projected to a new space in a way that there is no correlation among these new variables in this space. This means that the original absorbance data matrix (**A**) is decomposed into two smaller matrixes so called scores (**T**) and loadings (**B**). Another property of PCA projected data is the variables are sorted by the how much variance they explain from the original data. This allows using only few first projected variables (scores) to account for the most of the variance while leaving others out helps removing the instrumental noise. These chosen variables (first few column of scores), then, can be used in a regression model. The new data is projected to this space using the loadings which has the same number of columns as scores.

Similar to PCA, partial least squares (PLS) is a projection method where it is not only projects absorbances but also projects concentrations in a way that the covariance between them is maximized. Since the number of latent variables can be adjusted as choosing the number of principle components in PCA, PLS also eliminates the multicollinearity problem. A big advantage of PLS over PCA is the maximization of the covariance ensures the projected variables contains the information for prediction of responses while in PCA even if the most of the variance is explained it may not account for the information which explains the responses. In PLS, errors are assumed to be distributed evenly between absorbances and concentrations. The model equations for PLS are given in Equation 3.8 and 3.9 for absorbances and concentrations, respectively.

$$\mathbf{A} = \mathbf{T} \cdot \mathbf{B} + \mathbf{E}_A \quad (3.8)$$

$$\mathbf{c} = \mathbf{T} \cdot \mathbf{r} + \mathbf{e}_c \quad (3.9)$$

Here, **A** matrix has the same dimensions given in CLS and ILS. **T** is mxh matrix of scores and **B** is hxn matrix of loadings coming from PCA decomposition. The matrix \mathbf{E}_A is now different than which is given in the CLS as it comes from PCA but the size is identical to the one given in CLS. The term **c** is again $mx1$ vector of calibration concentrations of the component being modeled and **r** is $hx1$ vector of PLS regression

coefficients obtained by iterative solutions of Equations 3.8 and 3.9, consecutively. Figure 3.1 shows a schematic diagram of the Equations 3.8 and 3.9.

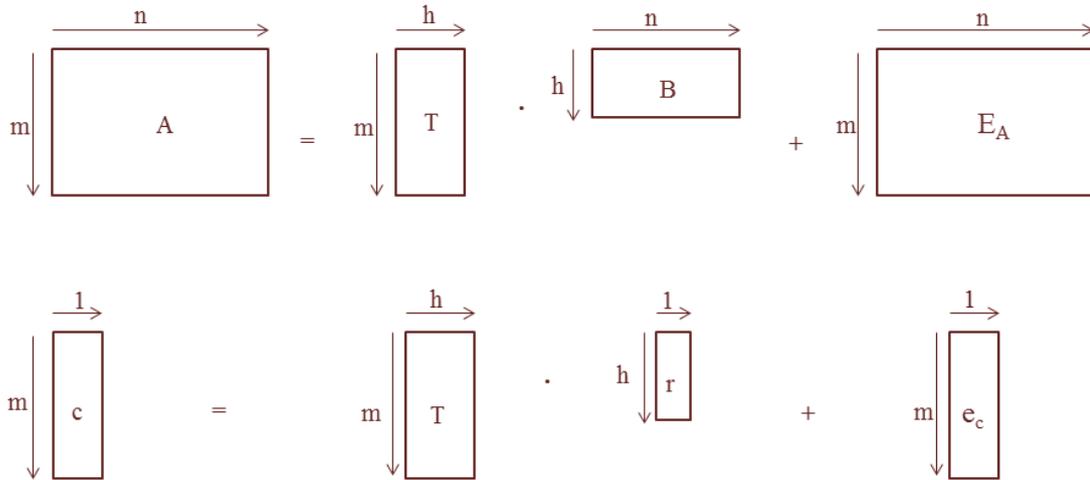


Figure 3.1. Schematic representation of the matrices used in PLS method along with their sizes.

Selection of optimal number of components (Number of PC: h) is a vital procedure for a successful model. Cross-validation can be used for plotting the number of components vs. standard error of cross validation (SECV). The point where the SECV starts to increase is usually selected as the optimal number of components since addition of the other components is likely to make the model less generalized and prone to overfitting (Equation 3.10).

$$PRESS = \sum_{i=1}^m (\hat{c}_i - c_i)^2 \quad (3.10)$$

Where \hat{c} is predicted component concentration. An example for the plot of $PRESS$ values vs. number of principle components is illustrated in Figure 3.2.

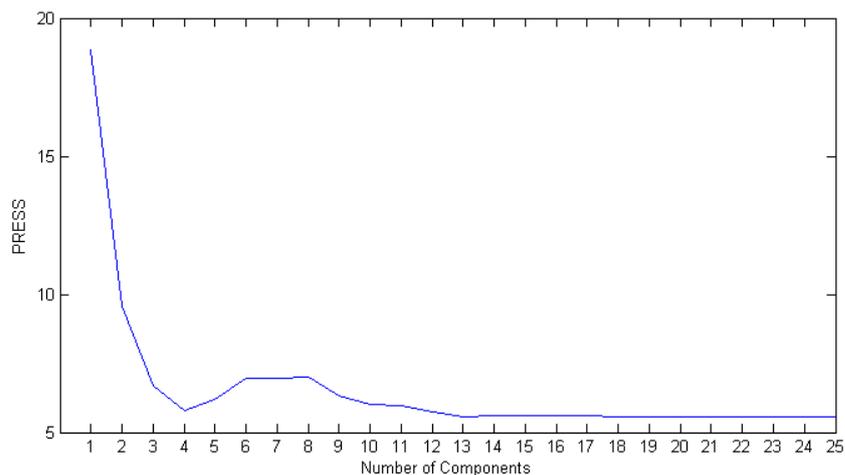


Figure 3.2. A sample plot of number of components vs. PRESS values.

As can be seen from Figure 3.2, the optimum number of PLS component for the given example should be selected as 4.

3.4. Genetic Inverse Least Squares (GILS)

Genetic algorithm (GA) is a variable selection method whose goal is to iteratively find the best combination of predictor variables that are well correlated with the response variables. The idea behind the GA is mimicking the evolution in nature where the genes of a population changes toward increasing the adaptation of the individuals. In chemometrics, this “fitness” criteria is defined as how successful a calibration or classification model is. As the number of variables increase calculating the fitness of each combination of them gets computationally very expensive quickly. In this manner, GA can be considered as a computationally cheap approach to the global minima problem.

Genetic inverse least squares (GILS) combines the genetic algorithm for feature selection and inverse least squares for calibration. Feature selection by the means of genetic algorithm helps constructing better models by excluding the variables containing noises and irrelevant information. Genetic algorithm was inspired by Darwin’s Evolution Theory and aims to mimic the nature where the individuals who are the fittest for their environment are more likely to breed survive and pass their genetic information to the next generation thus eventually increasing the fitness of the whole population.

The number of variables in spectral data is generally too many (1000+) and may contain random noises or other peaks that interferes with the information carrying peaks. Thus, combining GA with ILS provides noise reduction as well as a workaround for the collinearity problem. The success of the Genetic Inverse Least Squares(GILS) in constructing calibration models has been shown before (ÖZDEMİR and Öztürk 2004).

Genetic algorithm consists of five main steps: initialization of gene pool, evaluation of gene population, selection of genes, cross-over of genes, and replacing off-springs with parents.

The terms used in GILS and their definitions are listed below:

- ➔ Gene: Combination of absorbances at specific wavenumbers
- ➔ Fitness: Inverse of standard error of cross validation
- ➔ Cross-over: Exchange of half of the variables between two genes.
- ➔ Mutation: Excessive or deficient number of variables in a gene that may result after many iterations.

3.4.1. Algorithm Options

The minimum and maximum number of variables as well as number of genes is adjustable. As the number of genes increases the computational time increases as well however since more set of variables are selected the final model is usually gets better. Before accepting a gene to the gene pool, it is subjected to a fitness function that returns a correlation coefficient which is then compared to the threshold value defined by user. Another option, that is the number iterations, changes how many times the cross-over among the genes occurs thus increases the model's accuracy as well as the computational time. Finally, one can change the number of runs to alter how many best genes that will contribute to the averaged prediction, to be found.

3.4.2. Initialization of Gene Pool

The first step is to select random number of variables, which should not be more than the number of samples to prevent multicollinearity problem, to construct a gene. Then, the fitness and R^2 value of this gene is calculated by the means of cross validation. The cross-validation method used in this study is leave-one-out cross

validation where each sample is left out and predicted by the model constructed with the rest of the samples. If the resulting R^2 value is not less than the threshold value (e.g. 0.5), then this gene is accepted to the population. This procedure is repeated until desired number of genes is added to the population. A typical gene which is used in GILS is given in Equation 3.11.

$$G = [A_{462} A_{2719} A_{1999} A_{582} A_{1330}] \quad (3.11)$$

There are 5 variables that are randomly selected by the genetic algorithm in the gene given and the subscripts are for the wavenumbers where absorbance values used. Once a gene is selected it is then tested by determining the R^2 value of the model (Equation 3.12).

$$R^2 = \left(\frac{n \sum xy - (\sum x)(\sum y)}{\sqrt{n(\sum x^2) - (\sum x)^2} \sqrt{n(\sum y^2) - (\sum y)^2}} \right)^2 \quad (3.12)$$

Here x stands for the actual concentration and y stands for the predicted concentration of the samples in the calibration set.

3.4.3. Evaluation of the Genes in the Population

When a gen is selected as a potential candidate for the initial gene pool, it is success is then evaluated by using a fitness function that is inversely proportional to the Standard Error of Cross Validation (SECV) as given in Equation 3.13.

$$\text{Fitness} = \frac{1}{SECV} \quad (3.13)$$

Where the SECV can be calculated using Equation 3.14:

$$SECV = \sqrt{\frac{\sum_{i=1}^m (c_i - \hat{c}_i)^2}{m-2}} \quad (3.13)$$

In the SECV calculations, m represents the number of samples, c represents the actual concentration and \hat{c} refers to the predicted concentration. After determining the fitness value of each gene, the genes are sorted by their fitness value in decreasing order.

3.4.4. Selecting Parent Genes for Breeding

Breeding involves the exchange of information between two genes, thus, to mimic nature, the genes are selected as pairs. To achieve this, there are several established algorithms such as tournament selection, top-down selection and roulette wheel selection. In this study roulette wheel selection algorithm was used. Figure 3.3 shows schematic representation of roulette wheel selection where each slot represents the particular gene's fitness value. Thus the gene with the highest fitness has the largest are.

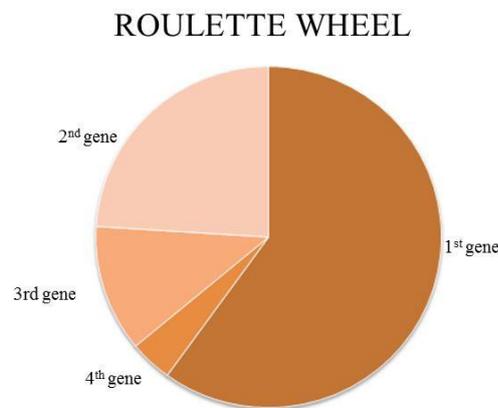


Figure 3.3. Schematic illustration of roulette wheel selection method.

In roulette wheel selection, each gene is placed on the wheel with the space they occupy in the wheel is determined by how fit they are. This provides a higher probability to be selected for the genes with larger fitness when the wheel is spun. This procedure is repeated until the number of selected genes is equal to the number of genes where some genes may be selected multiple times.

3.4.5. Cross-Over of the Selected Genes

In this step, single point cross over was applied for breeding where the first half of the variables in a gene is replaced by the last half of the other gene as illustrated in the following example. This results in two new genes represented as NEWG1 and NEWG2 shown below.

Parents:

G1 = [A₅₁₈, A₆₂₆₃ # A₂₇₁₇, A₅₄₄₄]

G2 = [A₉₉₉, A₆₆₆, A₃₃₃ # A₁₅₄₇, A₁₉₂₆, A₅₇₅₁]

Offsprings:

NEWG1 = [A₅₁₈, A₆₂₆₃, A₁₅₄₇, A₁₉₂₆, A₅₇₅₁]

NEWG2 = [A₂₇₁₇, A₅₄₄₄, A₉₉₉, A₆₆₆, A₃₃₃]

The # symbol refers to where the gene is divided. After the cross-over of all selected gene pairs in order of their section, their fitness values are calculated.

3.4.6. Replacing Offspring with Parents

Firstly, regardless of their fitness, the offspring genes are replaced with the parent genes in each iteration. Secondly, the fittest gene that is obtained from the cross-over step is compared to the fittest gene of the previous iteration. If any of the offspring in the new generation is fitter than the previous fittest one, the new fittest one is saved to be compared in the next iteration. The whole cycle of breeding and cross over procedure is repeated until a predefined number of iteration has been reached. The steps starting from selection of initial genes to the end iterations is called a run. At the end of each run, the best gene is saved and is then used to construct a model.

3.4.7. Prediction of the New Data and Frequency Plots

In the prediction step, the models constructed with each run are used to predict the concentrations of independent validation set and any other test set if exists. These concentration predictions are then averaged for all the runs to obtain a final result. This averaging effect is expected to generate more accurate predictions and usually provides a higher weight to the absorbance readings at wavenumbers containing more

information regarding to the concentrations. This can also be observed by looking at the frequency plots which show how many times a variable is used in the best genes of each run.

CHAPTER 4

EXPERIMENTATION

4.1. Sample Collection

In this study, a total of 139 authentic and commercial pure honey samples, which belong to different botanical and geographical origin, were collected at three different harvest seasons in Turkey (2014-2015-2016). The map shown in Figure 4.1 illustrates the geographical regions of honey that were used in this study. All of these honey samples were stored at room temperature until they were analyzed. In addition, corn syrup and beet sugar were purchased from local markets to be used as adulterants for preparation of synthetic samples. Among the 139 honey samples 39 of them were chosen according to their botanical and geographical origins in order to introduce maximum variability into calibration sets. Therefore, these reserved pure honey samples were used to prepare synthetically adulterated honey samples in our laboratory and the rest of 100 samples were used to test the developed multivariate calibration models.



Figure 4.1. The map of Turkey in which black areas show the geographical regions of authentic honey samples used in this study.

At the beginning of the study 87 authentic honey (*flower, honeydew, thyme, chestnut, cotton-flower, citrus honey*) samples as comb honey and filtered honey were obtained from beekeepers and commercial producers from the locations indicated on Figure 4.1 in 2014 and 2015 seasons. In addition, a total of 16 commercial brand honey samples (*Balkaşık filtered flower honey, Marmaris filtered flower honey, Balparmak filtered flower honey, Yeşil Tire filtered honeydew honey, Carrefour filtered flower honey, Marmaris filtered honeydew honey, Kipa filtered honeydew honey, Balkovan filtered flower honey, Yeşil tire filtered flower honey, Migros filtered flower honey, Tansaş filtered honeydew honey, Fer filtered flower honey, Carrefour filtered honeydew honey, Honeybana filtered honeydew honey, Pepe Anavarza filtered honeydew honey, Egebal filtered honeydew honey*) were purchased from local markets. Moreover, in order to increase the diversity of geographical and botanical origin of samples, a total of 23 authentic honey samples (*lavender honey, sloe-canola-sunflower mixture honey, mixed canola honey, linden honey, acacia honey, mixed trefoil honey, purple thyme honey, pure citrus honey, flower honey and honeydew honey*) from Ordu Apiculture Research Institute (Republic of Turkey Ministry of Food Agriculture and Livestock) were added to sample pool in the 2015 season. In 2016 harvest season, a total of 13 commercial brand honey samples (*theme honey, pine honeydew honey, theme-thistle, honeydew, heather honey and citrus honey*) were collected from local producers located in Marmaris and Datça regions. Thus a total of 139 ($87+16+23+13=139$) authentic and commercial honey samples were used in this study.



Figure 4.2. A small subset of the collected honey samples

In order to prepare adulterated honey samples, simple sugars (beet sugar and high fructose corn syrup); inverted sugars (glycose and sucrose) were used to prepare different adulteration scenarios. In all cases, water was used as a solvent to dissolve solid sugar and to obtain homogenous mixtures as well as being an adulterant.

4.2. Sample Preparation

In this study, four different adulteration scenarios were designed using different combination of artificial sweeteners. The components of the various adulteration scenarios will be given in detail later in this section. In order to enhance the diversity of adulterated samples both in terms of botanical and geographical origin, several pure honey samples were used to prepare mono floral synthetic mixtures. In addition, these pure honey samples were then mixed to prepare multifloral and multiregional honey stocks. These stock mixtures were prepared by using equal mass percentages of each pure honey sample used to produce particular stock mixture. Examples of synthetically prepared adulterated honey samples are shown in Figure 4.2.

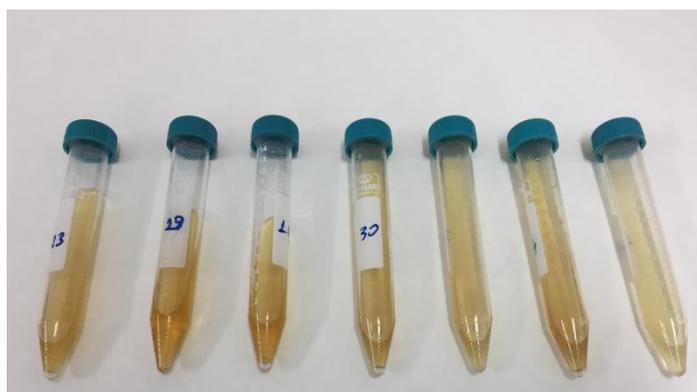


Figure 4.3. Examples of synthetically prepared adulterated honey samples.

While the first scenario included only beet sugar and water as adulterants, the second one contained only corn syrup as adulterant. Adulterated samples in the third scenario were prepared with corn syrup, beet sugar and water. Finally, the adulterants of the fourth scenario were glycose, sucrose and water.

Firstly, a simple adulteration scenario was designed where beet sugar and water were used as adulterants. Due to the very high viscosity of honey, dissolving solid beet sugar in honey is almost impossible, thus a sugar-water stock solution in 50% (w/w %) was prepared. Then, this stock solution was used to prepare adulterated honey samples. The concentration profile of 75 binary, ternary and pure honey samples used in the first scenario is shown in Table 4.1.

Table 4.1. Percent composition of 75 pure, binary and ternary adulterated samples prepared with pure honey, beet sugar and pure water.

No	Honey (w/w %)	Beet Sugar (w/w %)	Water (w/w %)	No	Honey (w/w %)	Beet Sugar (w/w %)	Water (w/w %)
1--7	100.00	0.00	0.00	42	18.13	40.94	40.94
8	36.71	0.00	63.29	43	68.90	15.55	15.55
9	32.50	0.00	67.50	44	8.51	45.74	45.74
10	47.36	0.00	52.64	45	87.58	6.21	6.21
11	47.15	0.00	52.85	46	83.17	8.41	8.41
12	47.62	0.00	52.38	47	35.93	32.04	32.04
13	20.18	0.00	79.82	48	38.13	30.94	30.94
14	18.40	0.00	81.60	49	65.99	17.01	17.01
15	7.79	0.00	92.21	50	43.94	28.03	28.03
16	17.48	0.00	82.52	51	10.98	44.51	44.51
17	81.02	0.00	18.98	52	14.12	42.94	42.94
18	56.70	0.00	43.30	53	62.12	18.94	18.94
19	68.25	0.00	31.75	54	48.83	25.59	25.59
20	24.55	0.00	75.45	55	34.83	32.59	32.59
21	31.46	0.00	68.54	56	77.32	11.34	11.34
22	59.42	20.29	20.29	57	38.02	30.99	30.99
23	64.52	17.74	17.74	58	46.92	26.54	26.54
24	6.01	47.00	47.00	59	31.50	34.25	34.25
25	6.44	46.78	46.78	60	4.40	47.80	47.80
26	73.15	13.43	13.43	61	59.88	20.06	20.06
27	42.91	28.55	28.55	62	4.20	47.90	47.90
28	92.80	3.60	3.60	63	15.46	42.27	42.27
29	97.18	1.41	1.41	64	88.59	5.70	5.70
30	73.47	13.26	13.26	65	40.53	29.73	29.73
31	53.58	23.21	23.21	66	75.75	12.12	12.12
32	74.43	12.79	12.79	67	25.15	37.42	37.42
33	93.20	3.40	3.40	68	58.65	20.68	20.68
34	42.04	28.98	28.98	69	12.10	43.95	43.95
35	62.68	18.66	18.66	70	89.80	5.10	5.10
36	73.28	13.36	13.36	71	87.43	6.28	6.28
37	66.64	16.68	16.68	72	31.90	34.05	34.05
38	59.05	20.47	20.47	73	4.12	47.94	47.94
39	29.78	35.11	35.11	74	43.05	28.47	28.47
40	69.19	15.40	15.40	75	46.59	26.71	26.71
41	66.99	16.50	16.50				

As mentioned above, in order to provide maximum variability, adulterated samples were prepared by using with 6 different mono floral honey samples. Also, a multi floral stock was prepared from these 6 mono floral honey samples. Therefore, a total of 7 honey stocks were used to prepare 68 ($7 \times 7 = 68$) binary and ternary adulterated samples. As seen from the Table 4.1, the first entry stands for the samples from 1 to 7 which are those pure honey samples. Among the remaining 68 samples, 14 of them are binary honey and water samples and 54 of them are ternary honey, beet sugar and water samples in which each sample contains equal amount beet sugar and water as the stock is 50% beet sugar-water (w/w %) solution. The ternary mixtures were prepared in a way that the samples from 8 to 13 were prepared by using only one type mono floral honey and beet sugar-water stock solution. Similarly, every other block that contains 7 ternary mixtures was also prepared with the remaining 5 mono floral honey samples. Finally, the last 12 ternary samples in the table were prepared with the poli floral stock honey and beet sugar-water mixture. Figure 4.4 shows the concentration changes for the components of the first scenario.

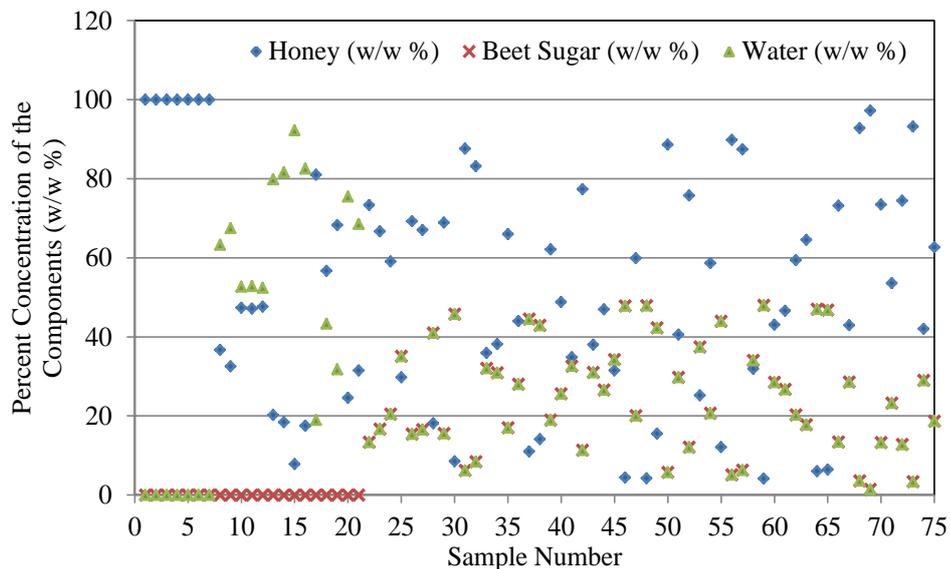


Figure 4.4. Percent concentration of the honey, beet sugar and water contents (w/w %) of the samples in the first scenario.

As seen from the Figure 4.4, the concentration of pure honey was changed from 4% to 100% (w/w %), beet sugar from 0% to 48% (w/w %) and water from 0% to 92% (w/w %).

The second scenario was also another simple adulteration scenario, because just two adulterants, corn syrup and water, were used to prepare synthetically adulterated honey samples. In this scenario, only one type of mono floral, flower honey, was used for the preparation of all adulterated honey samples. Because preparing a homogenous mixture of corn syrup and honey was difficult, corn syrup was diluted with water, as 50% (w/w %) corn syrup-water solution and this diluted stock solution was used for the preparation of the 42 adulterated samples (honey, corn syrup and water). Unlike the previous adulteration scenario, 9 binary mixtures of corn syrup and water samples were prepared at various concentrations (10-90 w/w %) and 1 corn syrup stock. In addition, 12 pure honey samples from various botanical (flower and honeydew) and geographical origin were also used for the construction of calibration models. As a result, a total of 65 samples were gathered in order to prepare calibration and independent validation sets. These samples were also used in FTNIR analyses by diluting them to 1:10 (w/w %) with water. The concentration profile of the adulterated and pure honey samples is shown in Table 4.2. Figure 4.5 shows the concentration changes for the components of the second scenario.

Table 4.2. Percent composition of 65 pure, binary and ternary adulterated samples prepared with pure honey, corn syrup and water.

No	Honey (w/w %)	Corn Syrup (w/w %)	Water (w/w %)	No	Honey (w/w %)	Corn Syrup (w/w %)	Water (w/w %)
1--12	100.00	0.00	0.00	39	77.00	11.50	11.50
13	0.00	9.97	90.03	40	70.88	14.56	14.56
14	0.00	20.02	79.98	41	85.18	7.41	7.41
15	0.00	30.07	69.93	42	82.34	8.83	8.83
16	0.00	40.22	59.78	43	71.28	14.36	14.36
17	0.00	50.01	49.99	44	87.17	6.42	6.42
18	0.00	60.15	39.85	45	75.28	12.36	12.36
19	0.00	69.96	30.04	46	64.05	17.97	17.97
20	0.00	79.82	20.18	47	71.85	14.08	14.08
21	0.00	90.29	9.71	48	63.02	18.49	18.49
22	0.00	100.00	0.00	49	87.18	6.41	6.41
23	78.30	10.85	10.85	50	64.90	17.55	17.55
24	84.18	7.91	7.91	51	72.78	13.61	13.61
25	59.94	20.03	20.03	52	80.17	9.91	9.91
26	84.74	7.63	7.63	53	74.15	12.92	12.92
27	72.30	13.85	13.85	54	67.23	16.39	16.39
28	73.62	13.19	13.19	55	74.49	12.76	12.76
29	65.80	17.10	17.10	56	68.38	15.81	15.81
30	81.62	9.19	9.19	57	76.21	11.89	11.89
31	76.39	11.80	11.80	58	87.22	6.39	6.39
32	99.02	0.49	0.49	59	96.43	1.79	1.79
33	48.56	25.72	25.72	60	81.21	9.39	9.39
34	69.55	15.22	15.22	61	92.87	3.57	3.57
35	89.25	5.38	5.38	62	85.47	7.26	7.26
36	68.68	15.66	15.66	63	75.70	12.15	12.15
37	64.23	17.89	17.89	64	59.78	20.11	20.11
38	73.01	13.50	13.50	65	96.04	1.98	1.98

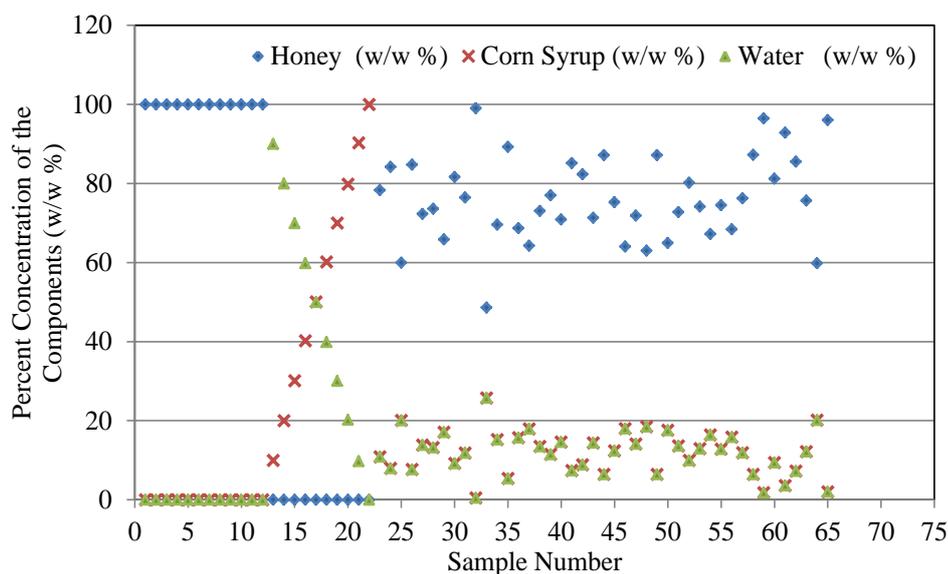


Figure 4.5. Percent concentration of the honey, corn syrup and water contents (w/w %) of the samples in the second scenario.

As can be seen in Figure 4.5, the concentration of pure honey was changed from 0% to 100% (w/w %), corn syrup from 0% to 100% (w/w %) and water from 0% to 90% (w/w %).

In the third scenario, corn syrup, beet sugar and water were used as adulterants and 7 mono floral pure honey samples and their mixture that was prepared as stock were used to prepare 74 adulterated samples. Among these pure honey samples 4 of them were collected directly from the bee hives and they were filtrated firstly and the combs were removed after centrifuging. Moreover, 8 more pure honey samples were also added to the data set in order to enhance the variability of honey types. Furthermore; to use beet sugar as adulterant in this data set, it was dissolved with water as 50% (w/w %) sugar-water solution, however; the corn syrup was used as it is in the binary mixtures with honey and no further dilution was done. As in the second scenario 9 binary mixtures of corn syrup-water and 2 corn syrup stocks were also included in the data set. In addition, 8 binary mixtures of beet sugar-water (10-80 w/w %) were prepared for the model set. As a result, a total of 108 (7+1+74+9+2+8=109) were used to generate calibration and independent validation sets. The concentration profiles of the samples used in the third scenario are shown in Table 4.3.

Table 4.3. Percent composition of 109 pure, binary, ternary and quaternary adulterated samples prepared with pure honey, corn syrup, beet sugar and water

No	Honey (w/w %)	Corn Syrup (w/w %)	Beet Sugar (w/w %)	Water (w/w %)	No	Honey (w/w %)	Corn Syrup (w/w %)	Beet Sugar (w/w %)	Water (w/w %)
1-16	100.00	0.00	0.00	0.00	63	71.19	24.36	2.22	2.22
17	64.35	35.65	0.00	0.00	64	48.70	23.28	14.01	14.01
18	80.07	19.93	0.00	0.00	65	59.70	10.11	15.09	15.09
19	72.91	27.09	0.00	0.00	66	53.59	19.90	13.25	13.25
20	74.81	25.19	0.00	0.00	67	85.04	4.83	5.07	5.07
21	60.46	39.54	0.00	0.00	68	57.04	5.58	18.69	18.69
22	94.36	5.64	0.00	0.00	69	79.42	5.36	7.61	7.61
23	86.33	13.67	0.00	0.00	70	83.97	6.80	4.62	4.62
24	79.85	20.15	0.00	0.00	71	63.87	23.15	6.49	6.49
25	66.23	33.77	0.00	0.00	72	60.35	29.78	4.94	4.94
26	67.70	32.30	0.00	0.00	73	58.21	11.77	15.01	15.01
27	69.50	30.50	0.00	0.00	74	59.76	14.31	12.96	12.96
28	82.71	17.29	0.00	0.00	75	52.15	23.94	11.95	11.95
29	69.45	30.55	0.00	0.00	76	69.52	26.33	2.07	2.07
30	86.45	13.55	0.00	0.00	77	58.17	20.06	10.89	10.89
31	0.00	9.97	0.00	90.03	78	50.81	28.46	10.36	10.36
32	0.00	20.02	0.00	79.98	79	68.84	10.19	10.49	10.49
33	0.00	30.07	0.00	69.93	80	79.73	11.37	4.45	4.45
34	0.00	40.22	0.00	59.78	81	59.23	26.50	7.13	7.13
35	0.00	50.01	0.00	49.99	82	71.19	24.31	2.25	2.25
36	0.00	60.15	0.00	39.85	83	64.75	16.23	9.51	9.51
37	0.00	69.96	0.00	30.04	84	61.88	12.65	12.73	12.73
38	0.00	79.82	0.00	20.18	85	54.72	23.02	11.13	11.13
39	0.00	90.29	0.00	9.71	86	55.32	18.34	13.17	13.17
40	0.00	100.00	0.00	0.00	87	89.32	7.07	1.80	1.80
41	0.00	0.00	10.20	89.80	88	68.62	14.89	8.25	8.25
42	0.00	0.00	19.97	80.03	89	64.27	7.34	14.19	14.19
43	0.00	0.00	30.05	69.95	90	85.75	3.30	5.48	5.48
44	0.00	0.00	39.99	60.01	91	69.29	4.35	13.18	13.18
45	0.00	0.00	49.87	50.13	92	77.44	9.92	6.32	6.32
46	0.00	0.00	60.06	39.94	93	62.75	8.50	14.38	14.38
47	0.00	0.00	70.08	29.92	94	74.92	0.13	12.47	12.47
48	0.00	0.00	77.75	22.25	95	56.36	10.76	16.44	16.44
49	94.56	0.00	2.72	2.72	96	44.66	12.68	21.33	21.33
50	89.36	0.00	5.32	5.32	97	65.07	3.39	15.77	15.77
51	84.70	0.00	7.65	7.65	98	41.42	23.62	17.48	17.48
52	99.08	0.00	0.46	0.46	99	45.86	13.79	20.17	20.17
53	47.36	0.00	26.32	26.32	100	53.83	20.39	12.89	12.89
54	30.43	0.00	34.79	34.79	101	55.98	7.63	18.19	18.19
55	26.64	0.00	36.68	36.68	102	51.66	16.86	15.74	15.74
56	77.84	0.00	11.08	11.08	103	69.24	19.17	5.79	5.79
57	87.90	0.00	6.05	6.05	104	48.65	19.02	16.17	16.17
58	79.60	0.00	10.20	10.20	105	60.35	21.76	8.95	8.95
59	74.58	0.00	12.71	12.71	106	59.55	26.29	7.08	7.08
60	97.56	0.00	1.22	1.22	107	69.20	10.46	10.17	10.17
61	35.51	0.00	32.24	32.24	108	79.15	11.82	4.51	4.51
62	58.80	9.17	16.01	16.01	109	0.00	100.00	0.00	0.00

As seen in Table 4.3, the first entry represents the 15 mono floral and 1 polifloral honey samples and the rest of 92 (108-16=92) samples are binary, ternary and quaternary mixtures of honey, corn syrup, beet sugar and water. Among these samples, 14 of them starting from the second entry are the binary mixtures containing honey and corn syrup. Following 19 samples were binary mixtures of corn syrup-water and beet sugar-water solutions that do not contain any honey. After these, following 13 samples are ternary mixtures of honey, beet sugar and water. The remaining 47 samples are quaternary synthetically adulterated honey samples which contain honey, corn syrup, beet sugar and water. Figure 4.6 shows the concentration changes for the components of the third scenario.

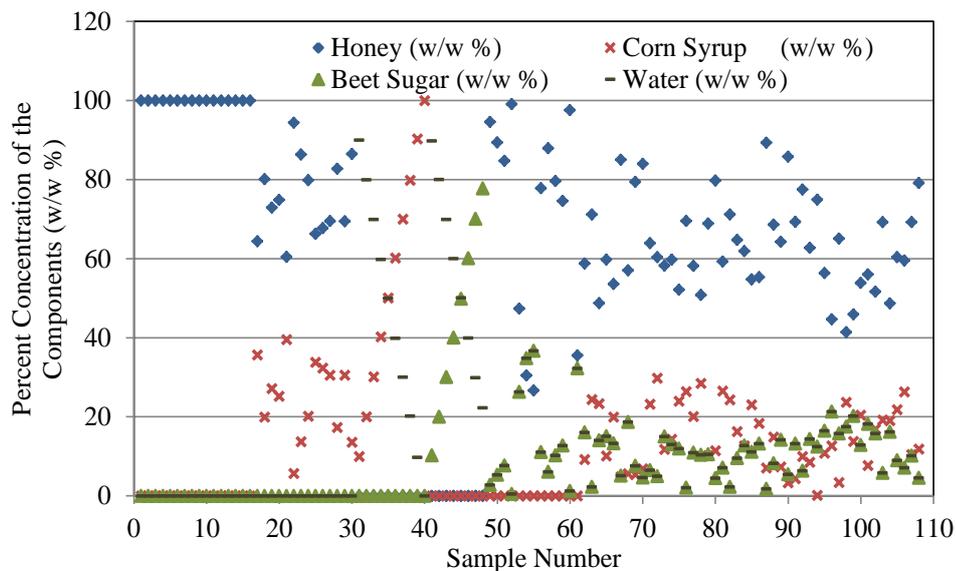


Figure 4.6. Percent concentration of the honey, corn syrup, beet sugar and water contents (w/w %) of the samples in the third scenario.

As seen in Figure 4.5, the concentration of pure honey was changed from 0% to 100% (w/w %), corn syrup from 0% to 100% (w/w %), beet sugar from 0% to 78% (w/w %) and water from 0% to 90% (w/w %).

The adulteration scenarios described up to this point consisted of corn syrup, beet sugar and water, which are thought to be widely used and can be easily obtained from the local markets to adulterate the authentic honey samples. However, there is no guarantee about the purities of corn syrup and beet sugar even though there is a label on the package which states that it is 100% beet sugar. So, in addition to the previous adulteration scenarios designed with corn syrup and beet sugar, the fourth one was

designed with using analytical grade glucose and sucrose as adulterants. In order to prepare adulterated honey samples with these sweeteners, they were dissolved in water to obtain %50 (w/w %) solutions and left in heat bath at approximately 30°C to achieve homogenous solutions. The concentration profile of the samples that were prepared to design the fourth adulteration scenario is shown in Table 4.4.

Table 4.4. Percent composition of 58 pure and quaternary adulterated samples prepared with pure honey, glucose, sucrose and water.

No	Honey (w/w %)	Glucose (w/w %)	Sucrose (w/w %)	Water (w/w %)	No	Honey (w/w %)	Glucose (w/w %)	Sucrose (w/w %)	Water (w/w %)
1-8	100.00	0.00	0.00	0.00	34	80.47	7.45	2.32	9.76
9	68.08	0.58	15.38	15.96	35	70.94	9.44	5.09	14.53
10	60.48	9.00	10.76	19.76	36	75.52	6.16	6.08	12.24
11	53.61	8.66	14.54	23.20	37	83.64	5.65	2.53	8.18
12	46.65	7.02	19.65	26.67	38	81.34	7.05	2.28	9.33
13	71.59	5.04	9.17	14.20	39	78.26	4.84	6.03	10.87
14	65.77	0.67	16.44	17.11	40	94.59	0.34	2.36	2.70
15	50.13	5.84	19.09	24.94	41	67.91	7.89	8.16	16.05
16	71.09	5.57	8.88	14.46	42	60.07	0.26	19.71	19.96
17	75.07	7.07	5.39	12.46	43	89.06	3.80	1.67	5.47
18	64.30	3.56	14.29	17.85	44	45.58	7.75	19.46	27.21
19	67.01	2.19	14.31	16.50	45	89.58	1.51	3.70	5.21
20	57.46	4.78	16.49	21.27	46	65.58	6.55	10.66	17.21
21	87.68	0.34	5.82	6.16	47	81.79	1.16	7.95	9.11
22	55.78	3.66	18.45	22.11	48	83.42	4.28	4.01	8.29
23	85.69	0.78	6.37	7.16	49	88.67	0.35	5.32	5.66
24	72.92	6.86	6.68	13.54	50	77.99	9.65	1.36	11.00
25	69.54	0.53	14.70	15.23	51	55.38	2.73	19.58	22.31
26	64.48	9.67	8.09	17.76	52	53.43	4.17	19.12	23.29
27	73.04	1.93	11.55	13.48	53	68.48	3.52	12.24	15.76
28	68.62	8.87	6.82	15.69	54	95.39	0.38	1.93	2.30
29	64.98	9.39	8.12	17.51	55	44.87	8.87	18.69	27.56
30	93.04	0.45	3.03	3.48	56	78.47	7.40	3.37	10.77
31	57.45	5.95	15.32	21.28	57	48.82	5.80	19.79	25.59
32	83.39	7.14	1.16	8.31	58	73.03	7.97	5.51	13.48
33	68.94	2.84	12.69	15.53					

As given in Table 4.4, the first entry stands for 7 mono floral honey samples and 1 stock which is obtain from equal amount of these 7 mono floral honeys. In this adulteration scenario, a total of 50 (58-8=50) adulterated honey samples were prepared using sucrose and glucose solutions. While preparing the 50 adulterated samples, 6 samples were prepared with each mono floral honey giving a total of 42 samples and the last 8 of them were prepared with the stock honey mixture. The concentration changes for the components of the fourth scenario are shown in Figure 4.7.

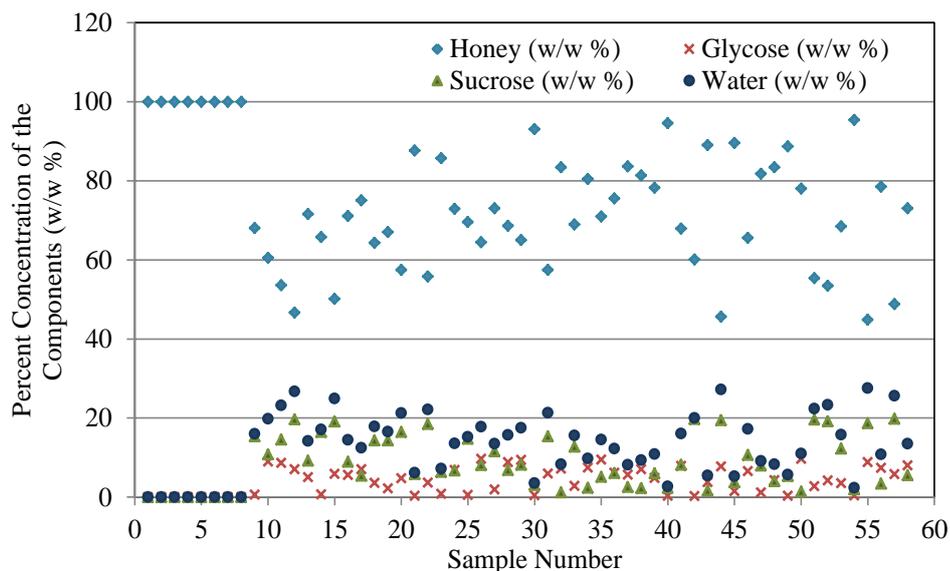


Figure 4.7. Percent concentration of honey, glucose, sucrose and water contents (w/w%) of the samples in the fourth scenario.

Figure 4.7 shows that the concentration of pure honey was changed from 40% to 100% (w/w %) whereas glucose from 0% to 10% (w/w %), sucrose from 0% to 20% (w/w %) and water from 0% to 30% (w/w %).

4.3. Data Collection

In this study, Fourier Transform Infrared Spectroscopy combined with three reflection diamond attenuated total reflectance (ATR) accessory and Near Infrared spectroscopy were used to carry out spectroscopic analyses and two different FTIR spectrometers which belong to same company were performed during these analyses. One of them was Perkin Elmer Spectrum 100 FTIR Spectrometer that was only used for the first data set, the other one was Perkin Elmer Frontier (FTIR/ NIR) Spectrometer. In the instruments, Cr-Ni wire was used as a light source. DTGS (Deuterated Triglycine Sulphate) was used as a detector while OptKBr was used as a beam splitter with a Ne-He laser in these instruments. FTIR spectra of both adulterated and pure honey samples were scanned at the range between 4000-600 cm^{-1} spectral regions with spectral resolution of 4 cm^{-1} after air background was taken. Each spectrum was recorded as $\log(1/R)$. In addition, Perkin Elmer Frontier (FTIR/ NIR) Spectrometer was used at NIR

mode for analyzing the samples that were belonged to the second adulteration scenario. Unlike FTIR mode, NIR TGS detector was used in NIR mode. Before analyzing in NIR range, these samples were diluted to 1:10 % (w/w %). FT-NIR spectra of adulterated and pure samples were scanned at the range between 1000-2500 nm spectral region with 64 cm^{-1} resolution, 0.5 scan speed and number of accumulation was set to 20. Moreover in NIR mode of the beam splitter of this instrument was CaF_2 , whereas the window was KBr. While the spectra scanned 1mm quartz cuvettes were used and air background was taken before analyzing the samples.

4.4. Data Processing & Chemometrics Model Construction

Microsoft Excel (MS Office 2010, Microsoft Corporation) program was used in order to separate the data to obtain calibration set, independent validation set and a set to be predicted. While calibration sets were used to develop chemometrics multivariate calibration models, the independent validation sets were used to determine the predictive power of these models. In addition to this, the rest of the data sets which are to be predicted were consisted of pure honey samples and commercial honey samples. In order to construct calibration models, genetic inverse least square (GILS) and partial least square (PLS) methods which were both coded in Matlab programming language (MATLAB R2016a-MathWorks Inc., Natick, MA) were used. Leave-one-out cross validation was used in both GILS and PLS in order to avoid overfitting of the models and also determine optimum number of PLS components. In the case of GILS method, the algorithm was set to run with 30 genes where each gene represents collection of randomly selected variables whose maximum size depends on the number of calibration samples. The variables are selected from the whole spectral range with an initial selection criterion of R^2 having a value of at least 0.50. The program was set to run 100 times in which the number of iteration was kept to 50 in each run. At the end, the best gene with the lowest standard error of cross validation (SECV) for the calibration set were selected to build final model for each run, resulting in a total of 100 best models. These models were then used to predict independent validation set and the standard error of prediction (SEP) were determined.

4.5. Design of Data Sets

In the first scenario, which composed of pure honey, beet sugar and water, On the basis of variation in concentration of the components (pure honey, beet sugar and water) 75 samples were divided into calibration and independent validation data sets. Therefore; out of 75 samples, 45 of them were selected randomly for the calibration set and the remained 30 samples were introduced to the model as independent validation set as seen Table 4.5. Here, 6 pure mono floral pure honey samples and 1 stock from the mixture of these 6 were reserved for the calibration set.

Table 4.5. Calibration and independent validation sets of the first scenario.

Calibration Data Set							
No	Honey (w/w %)	Sugar (w/w %)	Water (w/w %)	No	Honey (w/w %)	Sugar (w/w %)	Water (w/w %)
1*7	100.00	0.00	0.00	27	7.79	0.00	92.21
8	64.52	17.74	17.74	28	10.98	44.51	44.51
9	6.44	46.78	46.78	29	62.12	18.94	18.94
10	42.91	28.55	28.55	30	34.83	32.59	32.59
11	36.71	0.00	63.29	31	17.48	0.00	82.52
12	32.50	0.00	67.50	32	81.02	0.00	18.98
13	73.47	13.26	13.26	33	46.92	26.54	26.54
14	74.43	12.79	12.79	34	4.40	47.80	47.80
15	42.04	28.98	28.98	35	4.20	47.90	47.90
16	47.36	0.00	52.64	36	88.59	5.70	5.70
17	47.15	0.00	52.85	37	75.75	12.12	12.12
18	66.64	16.68	16.68	38	58.65	20.68	20.68
19	29.78	35.11	35.11	39	89.80	5.10	5.10
20	66.99	16.50	16.50	40	31.90	34.05	34.05
21	47.62	0.00	52.38	41	43.05	28.47	28.47
22	20.18	0.00	79.82	42	56.70	0.00	43.30
23	8.51	45.74	45.74	43	68.25	0.00	31.75
24	83.17	8.41	8.41	44	24.55	0.00	75.45
25	38.13	30.94	30.94	45	31.46	0.00	68.54
26	18.40	0.00	81.60				
Validation Data Set							
No	Honey (w/w %)	Sugar (w/w %)	Water (w/w %)	No	Honey (w/w %)	Sugar (w/w %)	Water (w/w %)
1	59.42	20.29	20.29	16	65.99	17.01	17.01
2	6.01	47.00	47.00	17	43.94	28.03	28.03
3	73.15	13.43	13.43	18	14.12	42.94	42.94
4	92.80	3.60	3.60	19	48.83	25.59	25.59
5	97.18	1.41	1.41	20	77.32	11.34	11.34
6	53.58	23.21	23.21	21	38.02	30.99	30.99
7	93.20	3.40	3.40	22	31.50	34.25	34.25
8	62.68	18.66	18.66	23	59.88	20.06	20.06
9	73.28	13.36	13.36	24	15.46	42.27	42.27
10	59.05	20.47	20.47	25	40.53	29.73	29.73
11	69.19	15.40	15.40	26	25.15	37.42	37.42
12	18.13	40.94	40.94	27	12.10	43.95	43.95
13	68.90	15.55	15.55	28	87.43	6.28	6.28
14	87.58	6.21	6.21	29	4.12	47.94	47.94
15	35.93	32.04	32.04	30	46.59	26.71	26.71

In the second scenario, that composed of pure honey, corn syrup and water, the concentration of pure honey and corn syrup were changed from 0% (w/w %) to 100% (w/w %) while water concentration was changed between 0% and 90% (w/w %). Among these 65 samples, 45 of them were selected randomly for the calibration set and

the remained 20 samples were introduced to the model as independent validation set as seen Table 4.6.

Table 4.6. Calibration and independent validation sets of the second scenario.

Calibration Data Set							
No	Honey (w/w %)	Corn Syrup (w/w %)	Water (w/w %)	No	Honey (w/w %)	Corn Syrup (w/w %)	Water (w/w %)
1x8	100.00	0.00	0.00	27	74.15	12.92	12.92
9	68.68	15.66	15.66	28	67.23	16.39	16.39
10	64.23	17.89	17.89	29	74.49	12.76	12.76
11	73.01	13.50	13.50	30	68.38	15.81	15.81
12	77.00	11.50	11.50	31	76.21	11.89	11.89
13	70.88	14.56	14.56	32	87.22	6.39	6.39
14	85.18	7.41	7.41	33	96.43	1.79	1.79
15	82.34	8.83	8.83	34	81.21	9.39	9.39
16	71.28	14.36	14.36	35	92.87	3.57	3.57
17	87.17	6.42	6.42	36	85.47	7.26	7.26
18	75.28	12.36	12.36	37	75.70	12.15	12.15
19	64.05	17.97	17.97	38	59.78	20.11	20.11
20	100.00	0.00	0.00	39	96.04	1.98	1.98
21	71.85	14.08	14.08	40	0.00	60.15	39.85
22	63.02	18.49	18.49	41	0.00	50.01	49.99
23	87.18	6.41	6.41	42	0.00	79.82	20.18
24	64.90	17.55	17.55	43	0.00	100.00	0.00
25	72.78	13.61	13.61	44	0.00	69.96	30.04
26	80.17	9.91	9.91	45	0.00	40.22	59.78
Validation Data Set							
No	Honey (w/w %)	Corn Syrup (w/w %)	Water (w/w %)	No	Honey (w/w %)	Corn Syrup (w/w %)	Water (w/w %)
1x3	100	0.00	0.00	12	76.39	11.80	11.80
4	78.30	10.85	10.85	13	99.02	0.49	0.49
5	84.18	7.91	7.91	14	48.56	25.72	25.72
6	59.94	20.03	20.03	15	69.55	15.22	15.22
7	84.74	7.63	7.63	16	89.25	5.38	5.38
8	72.30	13.85	13.85	17	0.00	30.07	69.93
9	73.62	13.19	13.19	18	0.00	9.97	90.03
10	65.80	17.10	17.10	19	0.00	20.02	79.98
11	81.62	9.19	9.19	20	0.00	90.29	9.71

In the data set of third scenario, that composed of pure honey, corn syrup, beet sugar and water, pure honey and corn syrup concentration were changed from 0% (w/w %) to 100% (w/w %) while sugar concentration was changed between 0% and 80% (w/w %). In addition, water concentration was changed between 0% and 90% (w/w %).

Moreover; out of 109 samples which were given in Table 4.3, 103 of them were used as calibration and independent validation data set and remain 6 samples that were binary mixtures of sugar-water and corn syrup-water solutions (sample number 34, 36, 38, 44, 46 and 109 given in Table 4.3) were used to test the prediction accuracy of the model. Therefore; among 103 samples, 73 of them were randomly selected and placed in calibration data and the rest of them (30 samples) were introduced to the model as independent validation data. This data arrangement is shown in Table 4.7.

Table 4.7. Calibration and independent validation sets of the third scenario.

Calibration Data Set									
No	Honey (w/w %)	Corn Syrup (w/w %)	Sugar (w/w %)	Water (w/w %)	No	Honey (w/w %)	Corn Syrup (w/w %)	Sugar (w/w %)	Water (w/w %)
1x13	100	0.00	0.00	0.00	44	87.90	0.00	6.05	6.05
14	64.35	35.65	0.00	0.00	45	62.75	8.50	14.38	14.38
15	80.07	19.93	0.00	0.00	46	74.92	0.13	12.47	12.47
16	72.91	27.09	0.00	0.00	47	56.36	10.76	16.44	16.44
17	74.81	25.19	0.00	0.00	48	44.66	12.68	21.33	21.33
18	60.46	39.54	0.00	0.00	49	65.07	3.39	15.77	15.77
19	94.36	5.64	0.00	0.00	50	41.42	23.62	17.48	17.48
20	86.33	13.67	0.00	0.00	51	45.86	13.79	20.17	20.17
21	79.85	20.15	0.00	0.00	52	53.83	20.39	12.89	12.89
22	66.23	33.77	0.00	0.00	53	55.98	7.63	18.19	18.19
23	67.70	32.30	0.00	0.00	54	51.66	16.86	15.74	15.74
24	69.50	30.50	0.00	0.00	55	69.24	19.17	5.79	5.79
25	71.19	24.31	2.25	2.25	56	48.65	19.02	16.17	16.17
26	64.75	16.23	9.51	9.51	57	60.35	21.76	8.95	8.95
27	61.88	12.65	12.73	12.73	58	59.55	26.29	7.08	7.08
28	54.72	23.02	11.13	11.13	59	69.20	10.46	10.17	10.17
29	55.32	18.34	13.17	13.17	60	79.15	11.82	4.51	4.51
30	89.32	7.07	1.80	1.80	61	0.00	9.97	0.00	90.03
31	68.62	14.89	8.25	8.25	62	0.00	20.02	0.00	79.98
32	64.27	7.34	14.19	14.19	63	0.00	30.07	0.00	69.93
33	85.75	3.30	5.48	5.48	64	0.00	50.01	0.00	49.99
34	69.29	4.35	13.18	13.18	65	0.00	69.96	0.00	30.04
35	77.44	9.92	6.32	6.32	66	0.00	90.29	0.00	9.71
36	94.56	0.00	2.72	2.72	67	0.00	100	0.00	0.00
37	89.36	0.00	5.32	5.32	68	0.00	0.00	10.20	89.80
38	84.70	0.00	7.65	7.65	69	0.00	0.00	19.97	80.03
39	99.08	0.00	0.46	0.46	70	0.00	0.00	30.05	69.95
40	47.36	0.00	26.32	26.32	71	0.00	0.00	49.87	50.13
41	30.43	0.00	34.79	34.79	72	0.00	0.00	70.08	29.92
42	26.64	0.00	36.68	36.68	73	0.00	0.00	77.75	22.25
43	77.84	0.00	11.08	11.08					
Validation Data Set									
No	Honey (w/w %)	Corn Syrup (w/w %)	Sugar (w/w %)	Water (w/w %)	No	Honey (w/w %)	Corn Syrup (w/w %)	Sugar (w/w %)	Water (w/w %)
1x3	100	0.00	0.00	0.00	17	74.58	0.00	12.71	12.71
4	82.71	17.29	0.00	0.00	18	97.56	0.00	1.22	1.22
5	69.45	30.55	0.00	0.00	19	35.51	0.00	32.24	32.24
6	86.45	13.55	0.00	0.00	20	63.87	23.15	6.49	6.49
7	58.80	9.17	16.01	16.01	21	60.35	29.78	4.94	4.94
8	71.19	24.36	2.22	2.22	22	58.21	11.77	15.01	15.01
9	48.70	23.28	14.01	14.01	23	59.76	14.31	12.96	12.96
10	59.70	10.11	15.09	15.09	24	52.15	23.94	11.95	11.95
11	53.59	19.90	13.25	13.25	25	69.52	26.33	2.07	2.07
12	85.04	4.83	5.07	5.07	26	58.17	20.06	10.89	10.89
13	57.04	5.58	18.69	18.69	27	50.81	28.46	10.36	10.36
14	79.42	5.36	7.61	7.61	28	68.84	10.19	10.49	10.49
15	83.97	6.80	4.62	4.62	29	79.73	11.37	4.45	4.45
16	79.60	0.00	10.20	10.20	30	59.23	26.50	7.13	7.13

In the fourth scenario that composed of pure honey, glucose, sucrose and water, 45 of the samples given in Table 4.4 were randomly selected as calibration set and the remaining 13 samples were used for the independent validation set. Multivariate calibration models with GILS and PLS methods were generated with these calibration and independent validation sets. Table 4.8 shows the concentration profiles of these calibration and independent validation sets.

Table 4.8. Calibration and independent validation sets of the fourth scenario.

Calibration Data Set									
No	Honey (w/w %)	Glucose (w/w %)	Sucrose (w/w %)	Water (w/w %)	No	Honey (w/w %)	Glucose (w/w %)	Sucrose (w/w %)	Water (w/w %)
1x8	100	0.00	0.00	0.00	27	68.94	2.84	12.69	15.53
9	68.08	0.58	15.38	15.96	28	80.47	7.45	2.32	9.76
10	60.48	9.00	10.76	19.76	29	70.94	9.44	5.09	14.53
11	53.61	8.66	14.54	23.20	30	83.64	5.65	2.53	8.18
12	71.59	5.04	9.17	14.20	31	81.34	7.05	2.28	9.33
13	65.77	0.67	16.44	17.11	32	78.26	4.84	6.03	10.87
14	50.13	5.84	19.09	24.94	33	67.91	7.89	8.16	16.05
15	75.07	7.07	5.39	12.46	34	60.07	0.26	19.71	19.96
16	64.30	3.56	14.29	17.85	35	89.06	3.80	1.67	5.47
17	67.01	2.19	14.31	16.50	36	89.58	1.51	3.70	5.21
18	87.68	0.34	5.82	6.16	37	65.58	6.55	10.66	17.21
19	55.78	3.66	18.45	22.11	38	81.79	1.16	7.95	9.11
20	85.69	0.78	6.37	7.16	39	88.67	0.35	5.32	5.66
21	69.54	0.53	14.70	15.23	40	77.99	9.65	1.36	11.00
22	64.48	9.67	8.09	17.76	41	55.38	2.73	19.58	22.31
23	73.04	1.93	11.55	13.48	42	68.48	3.52	12.24	15.76
24	64.98	9.39	8.12	17.51	43	95.39	0.38	1.93	2.30
25	93.04	0.45	3.03	3.48	44	44.87	8.87	18.69	27.56
26	57.45	5.95	15.32	21.28	45	48.82	5.80	19.79	25.59
Validation Data Set									
No	Honey (w/w %)	Glucose (w/w %)	Sucrose (w/w %)	Water (w/w %)	No	Honey (w/w %)	Glucose (w/w %)	Sucrose (w/w %)	Water (w/w %)
1	46.65	7.02	19.65	26.67	8	94.59	0.34	2.36	2.70
2	71.09	5.57	8.88	14.46	9	45.58	7.75	19.46	27.21
3	57.46	4.78	16.49	21.27	10	83.42	4.28	4.01	8.29
4	72.92	6.86	6.68	13.54	11	53.43	4.17	19.12	23.29
5	68.62	8.87	6.82	15.69	12	78.47	7.40	3.37	10.77
6	83.39	7.14	1.16	8.31	13	73.03	7.97	5.51	13.48
7	75.52	6.16	6.08	12.24					

CHAPTER 5

RESULTS AND DISCUSSION

As mentioned in previous chapters, in this study, FTIR spectroscopy coupled with three reflection diamond attenuated total reflectance (ATR) accessory and FTNIR spectroscopy were used to collect spectral data and these collected data have been used to develop multivariate calibration models. In this thesis, four main adulteration scenarios and one combined adulteration scenario (gathering the prepared samples in the main adulteration scenarios) have been carried out. While FTIR spectroscopy was used for all adulteration scenarios, FTNIR spectroscopy was only performed in the second adulteration scenario as an alternative spectroscopic technique. Detailed description of these scenarios was given in Chapter 4.

As it is explained in detail in Chapter 2, FTIR spectroscopy has been used to characterize functional groups, bonding types and nature of compounds. Despite having a complex chemical structure, food and agricultural products could also be identified by FTIR spectroscopy. To investigate any spectral difference among pure honey samples that belongs to different botanical and geographical origin, 6 mono floral honey samples from various botanical origins were selected and their FTIR-ATR spectra are shown in Figure 5.1.

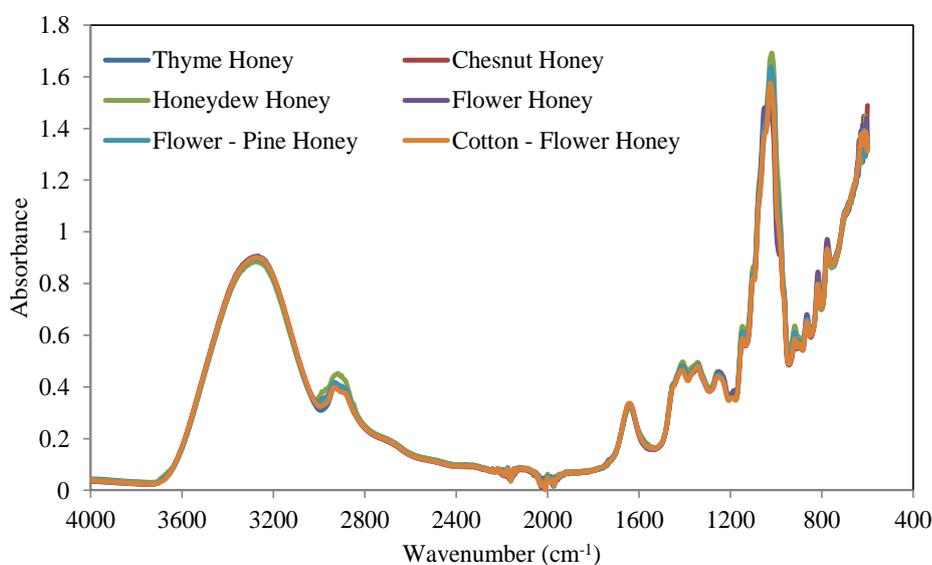


Figure 5.1. The spectra of 6 different botanical origins that are thyme, chestnut, pine, flower-pine, flower and cotton-flower honey collected by using FTIR spectroscopy coupled with ATR accessory.

As can be seen in Figure 5.1, all of the honey samples have different botanical origins but their spectral features look very similar. However, closer examination of the peaks around $3000\text{-}2800\text{ cm}^{-1}$ and $1600\text{-}1200\text{ cm}^{-1}$ indicates absorbance differences among the various botanical origins due to the some physical (such as, viscosity and density) and chemical composition differences. With the same perspective, a total of 139 spectra collected from different geographical area and also belong to various botanical origin are shown in Figure 5.2.

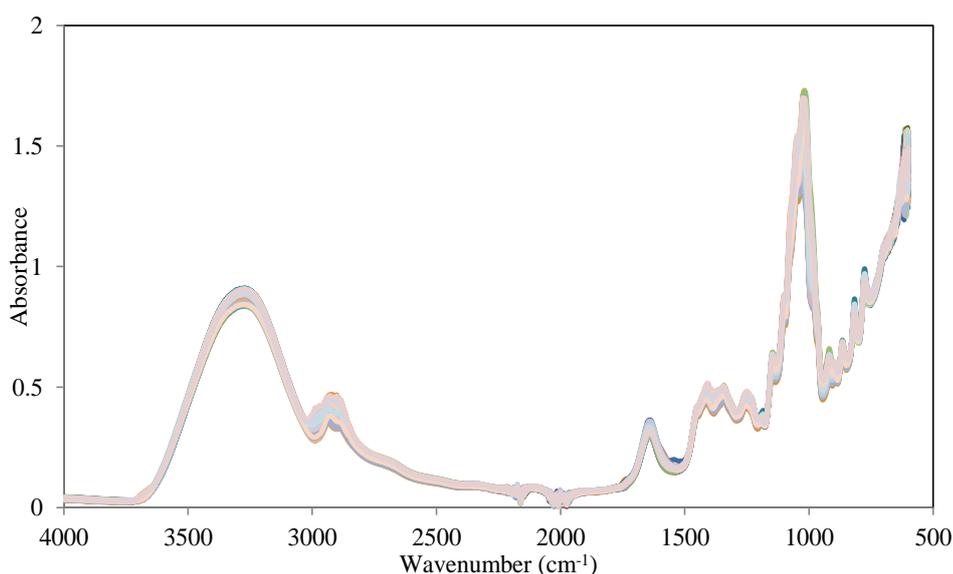


Figure 5.2. The spectra of 139 honey samples that belong to different botanical and geographical origin by using FTIR spectroscopy coupled with ATR accessory.

As can be seen in Figure 5.2, very small differences based on the intensity of each authentic honey samples, which are assumed to be non-adulterated, can be noticed between $4000\text{-}600\text{ cm}^{-1}$ spectral range. To prepare adulterations scenarios that cover the small variability caused by the geographical and botanic origins, some of these authentic honey samples are used to develop a chemometrics multivariate calibration models while the rest of them are used to test the calibration models. The number of honey sample that were added to calibration and independent validation sets in the model building steps are varied as explained in Chapter 4.

5.1. The First Adulteration Scenario (Beet Sugar and Water were Used as Adulterants)

The first adulteration scenario designed as a simple one. Only beet sugar and water were used to prepare the adulterated honey samples in this scenario. This scenario was composed of a total of 75 authentic and adulterated honey samples (Table 4.1). FTIR-ATR spectra of the 75 honey samples were shown in Figure 5.3.

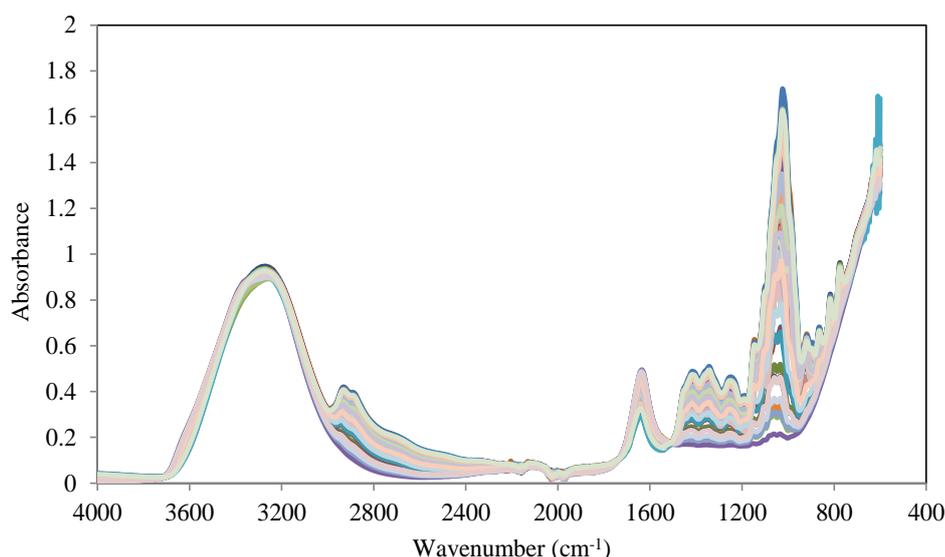


Figure 5.3. FTIR-ATR spectra of 6 authentic and 69 adulterated honey samples prepared with two adulterants (beet sugar and water).

As can be seen in Figure 5.3, there have been relatively larger variations compared to the FTIR spectra of authentic samples shown in Figure 5.2 among the intensities of the FTIR-ATR spectra of 75 samples and these variations have been more apparent especially in the range of $3000\text{--}2400\text{ cm}^{-1}$ and $1500\text{--}800\text{ cm}^{-1}$ wavenumber. The reason of this variability can be explained with the existence of different percentage of beet sugar and water used as adulterant in synthetically prepared honey samples. While the visible variations in spectra may provide qualitative evidence of whether an adulteration exist, in order to determine the amount of adulteration quantitatively, development of chemometrics models are necessary.

As described in Table 4.5, out of 75 samples on the basis of honey concentration distribution, 45 of them were introduced as calibration set and the remaining samples were assigned as independent validation data set (details about selecting these samples

into calibration and independent validation sets were given in Chapter 4). By using calibration data set, multivariate calibration models were developed with GILS and PLS for three components namely honey, beet sugar and water contents while the independent validation set was used to validate these models. In order to avoid overfitting leave-one-out cross validation was used. In the case of PLS modelling optimum number of PC was determined based on this leave-one-out cross validation approach. Actual vs. GILS predicted concentration plots of honey, beet sugar and water contents have been obtained and these plots are given in Figure 5.4.

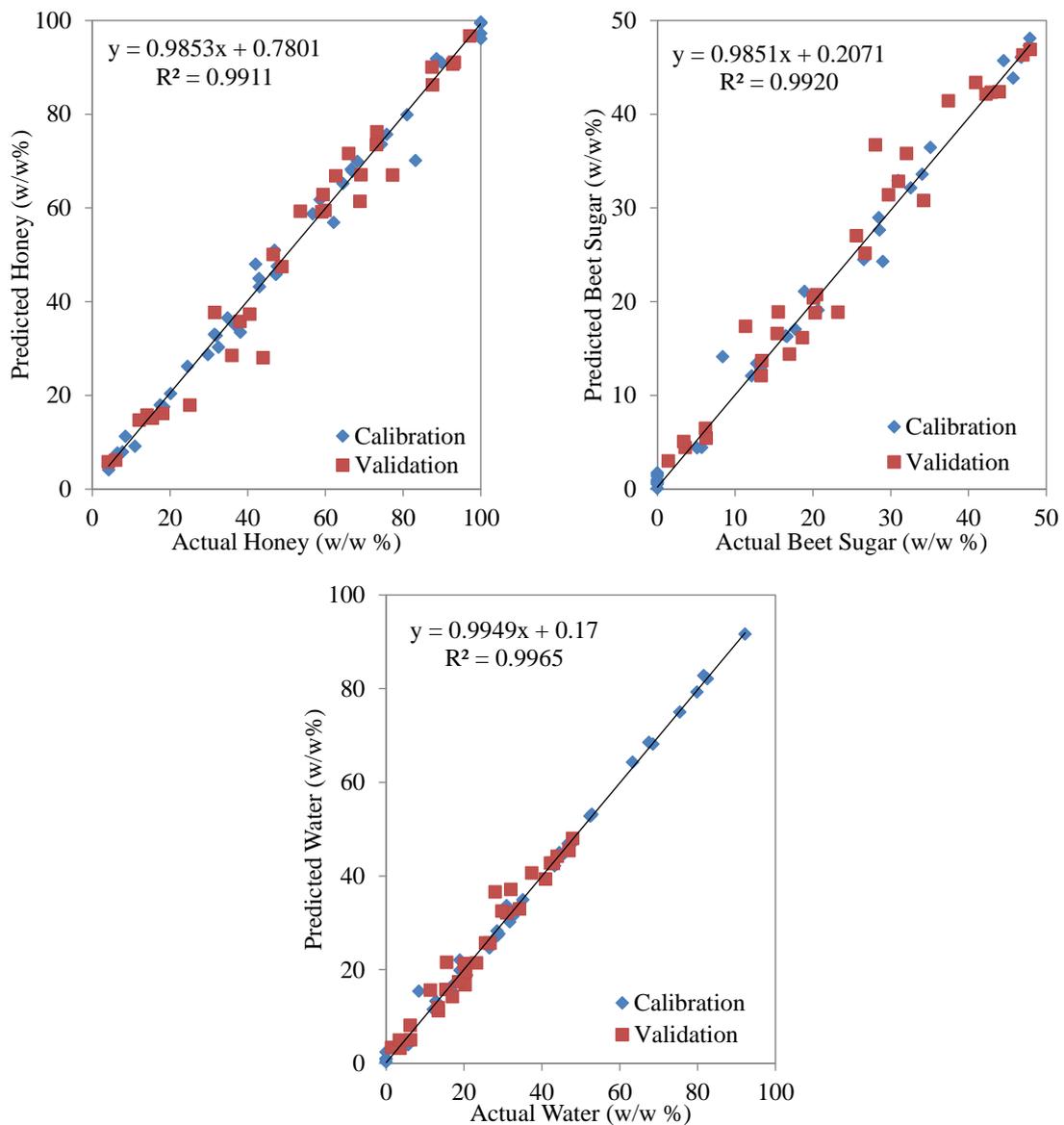


Figure 5.4. Actual versus predicted plot of honey, beet sugar and water contents resulting from GILS in the first scenario.

As can be seen in Figure 5.4, correlation coefficients (R^2) of honey, beet sugar and water are 0.9911, 0.9920 and 0.9965, respectively. Due to having high R^2 value, each model can be regarded as successful for the calibration set. However, this does not guarantee the same success for the independent validation set. Therefore, evaluation of standard error of cross validation (SECV) and standard error of prediction (SEP) values for the independent validation set in the dynamic range of each contents can be helpful for the performances of all these models. The values of SECV, SEP and R^2 are shown in Table 5.1 with the concentration range of each content.

Table 5.1. Standard error of cross validation (SECV), standard error of prediction (SEP), maximum and minimum values of the components (Max and Min) and correlation coefficient (R^2) of GILS models belong to the first adulteration scenario.

	SECV (w/w %)	SEP (w/w %)	Max	Min	R^2
Honey	2.96	4.94	100	5	0.9911
Beet Sugar	1.50	2.77	50	0	0.9920
Water	1.58	2.36	95	0	0.9965

As seen in Table 5.1, the values of SECV and SEP have been reported with the same unit of the concentrations of calibration and independent validation samples (w/w %) and the care must be taken while evaluating the magnitudes of these values as they depend on the dynamic range of calibration set (Max-Min). The SEP value for the honey model appears to be the largest value, but this could be a misleading conclusion if the range of honey is ignored. In fact, the relative magnitude of SEP value for the beet sugar is as large as the SEP value of honey model. In addition to that, the dynamic ranges of pure honey and water contents are similar; however, the SECV and SEP values of honey model are higher and R^2 is lower than water model.

On the other hand, PLS method was also performed as a reference multivariate calibration method to GILS with the same calibration data set and validated using the same independent validation data set that was used in GILS. The resulted PLS predicted concentrations vs actual concentrations of all three contents are given in Figure 5.5.

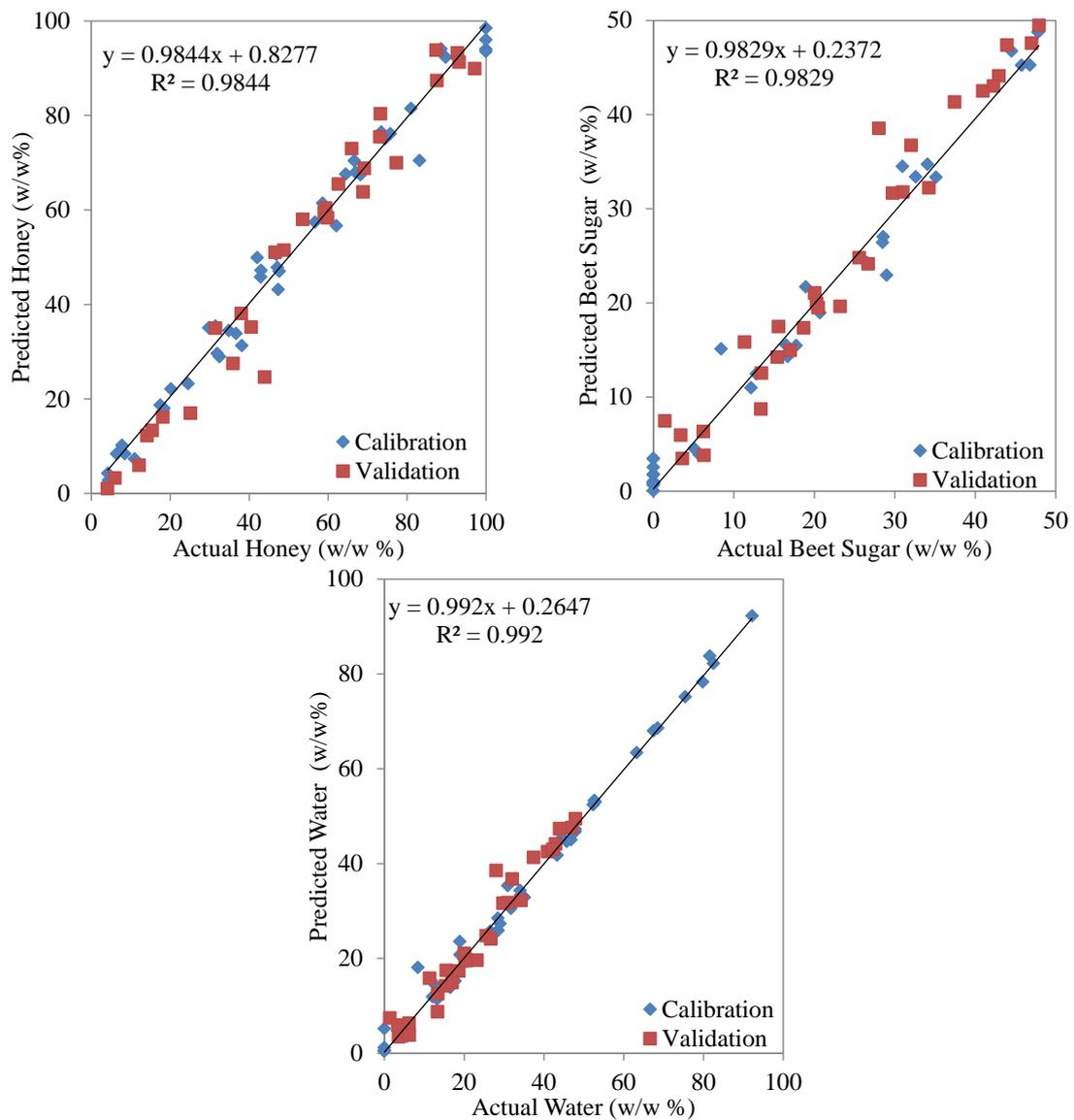


Figure 5.5. Actual versus predicted plots of honey, beet sugar and water contents resulted from PLS in the first scenario.

As can be observed in Figure 5.5, correlation coefficients of honey, beet sugar and water are found as 0.9844, 0.9829 and 0.9920, respectively. When GILS and PLS methods have been compared in terms of their R^2 , no significant differences were observed. Likewise, the number of principle component (PC), R^2 , SECV and SEP values have been shown individually along with their maximum and minimum (Max and Min) operating ranges in Table 5.2.

Table 5.2. The number of principle components (PC), standard error of cross validation (SECV), standard error of prediction (SEP) along with maximum and minimum operating ranges of the contents (Max and Min) and correlation coefficients (R^2) of PLS models belonging to the first adulteration scenario.

	Number of PC	SECV (w/w %)	SEP (w/w %)	Max	Min	R^2
Honey	7	3.92	5.65	100	5	0.9844
Beet Sugar	7	2.18	2.61	50	0	0.9829
Water	4	2.37	3.16	95	0	0.992

When the SECV and SEP values given in Table 5.2 are compared with the values from GILS it is seen that somewhat larger SECV values were obtained with PLS. However, the SEP values for honey and beet sugar were quite similar, but SEP value for water was larger than the GILS model. While the optimum number of PC for honey and beet sugar was 7, it is found to be 4 for water model and this could be the reason for the larger SEP value. As in the development of GILS models, the full FTIR-ATR spectra have been used while developing calibration models with PLS, but GILS has the advantage of reducing the number of variables to the size of calibration set thereby eliminating uncorrelated spectral features. However, no variable reduction was applied with PLS although it is possible to combine the same genetic algorithm to the PLS algorithm. In fact a genetic algorithm based PLS method was also developed, this version of the PLS was not used in this study as the aim was to compare the performance of the GILS with the standard PLS method.

After completing the modeling process with both GILS and PLS, the final models were also tested with a secondary independent data set which is composed of 50 pure honey samples that belong to different botanical and geographical origins. The FTIR-ATR spectra of the 50 authentic honey samples are demonstrated in Figure 5.6. Although, all of these samples were known to be pure honey samples, the models for the honey, beet sugar and water contents were executed to predict each of these components.

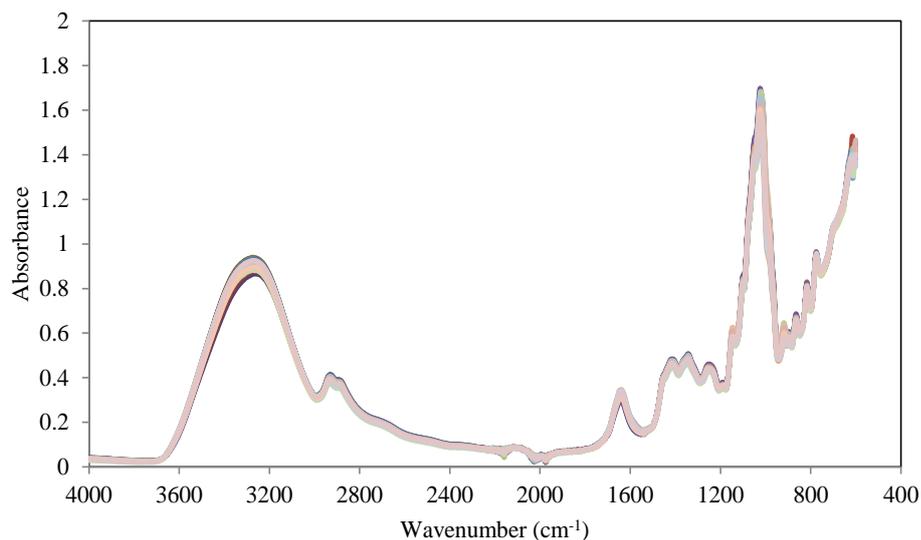


Figure 5.6. FTIR-ATR spectra of 50 authentic honey samples which were introduced to the multivariate calibration models to test their success.

As can be seen from the Figure 5.6, FTIR spectra of 50 authentic honey samples do not possess any significant visible differences in their spectral intensities as observed. The concentration predictions in terms of honey contents of these 50 samples by GILS and PLS methods are given in Figure 5.7.

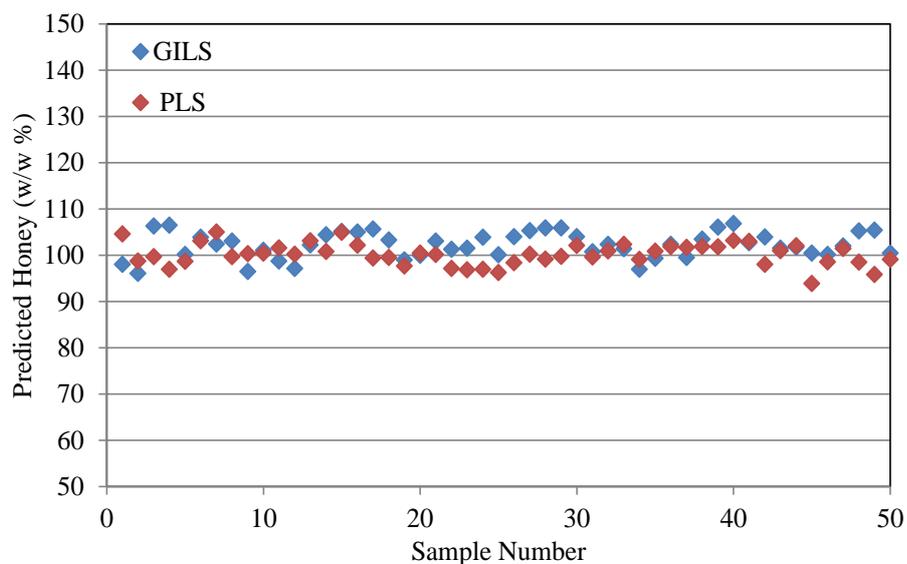


Figure 5.7. Honey content prediction of 50 authentic honey samples with both GILS and PLS methods.

As Figure 5.7 shows, the concentration of honey content in 50 samples can be predicted between the range of 90% and 110% (w/w %) with GILS and PLS models, so the prediction ranges of two calibration methods for honey content can be determined as $\pm 6.85\%$ (w/w %). These samples have been known as non-adulterated honey, so it is expected that the predictions of honey content concentrations of all samples to be as close as possible to 100% (w/w %). The deviations from 100% among the predictions are expected due to the reported SEP values of these models which were 4.9% for GILS and 5.65% for PLS. Out of the 50 samples 38 of the GILS predictions were within the SEP value of the 4.94% where as 49 of the PLS predictions were within the SEP value of the 5.65%. The largest absolute deviation for the GILS model was 6.85% where as for the PLS model was 6.10% (w/w %). These result demonstrated that both models were able to generate successful prediction for the secondary independent test set.

On the other hand, the prediction results of adulterants (beet sugar and water) obtained by GILS and PLS models are given in Figure 5.8, respectively.

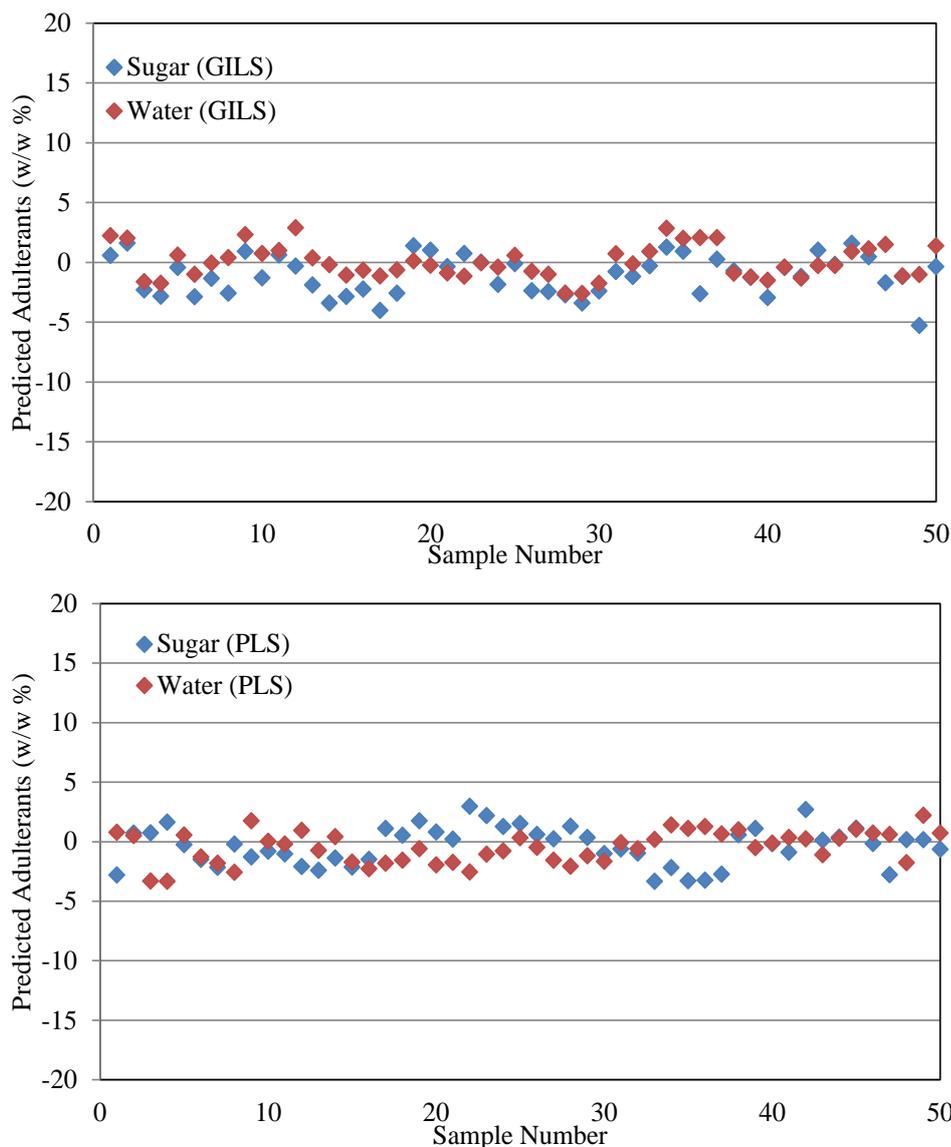


Figure 5.8. Beet sugar and water contents prediction of 50 authentic honey samples with both GILS and PLS methods.

Both beet sugar and water contents can be predicted with ± 5 (w/w %) deviation using GILS and PLS models. In fact the largest absolute deviation for beet sugar was 5.29% for GILS but it was only one sample and the rest of the samples were deviated within the models SEP values. When the water model predictions were compared it was found that the GILS prediction demonstrated a maximum absolute deviation of 2.80% where as for the PLS predictions resulted in a maximum absolute deviation of 3.37% (w/w %). When the predictive performances of models for each adulterant as well as honey content are taken into consideration, both GILS and PLS models appear to be sufficiently good for determination of adulterants quantitatively.

5.2. The Second Adulteration Scenario (Corn Syrup and Water were Used as Adulterants)

In the second adulteration scenario corn syrup was used as adulterant and water was used as solvent. For this, 50% (w/w %) corn syrup-water solution was prepared as the stock solution. In this adulteration scenario, there are a total of 65 authentic and adulterated honey samples and the concentration profile of these samples were given in Table 4.2. Unlike the previous adulteration scenario, to enhance the performance of the multivariate calibration models, a total of 10 corn syrup solutions with varying amount of water were included in the adulteration scenario as shown in Table 4.2. To develop calibration models, two different molecular spectroscopic methods, FTIR and FTNIR spectroscopies were used in this adulteration scenario.

5.2.1. FTIR-ATR Results of Second Adulteration Scenario

In this part FTIR spectroscopy coupled with three reflection diamond ATR accessory was used and the spectra of all 65 samples are illustrated in Figure 5.9.

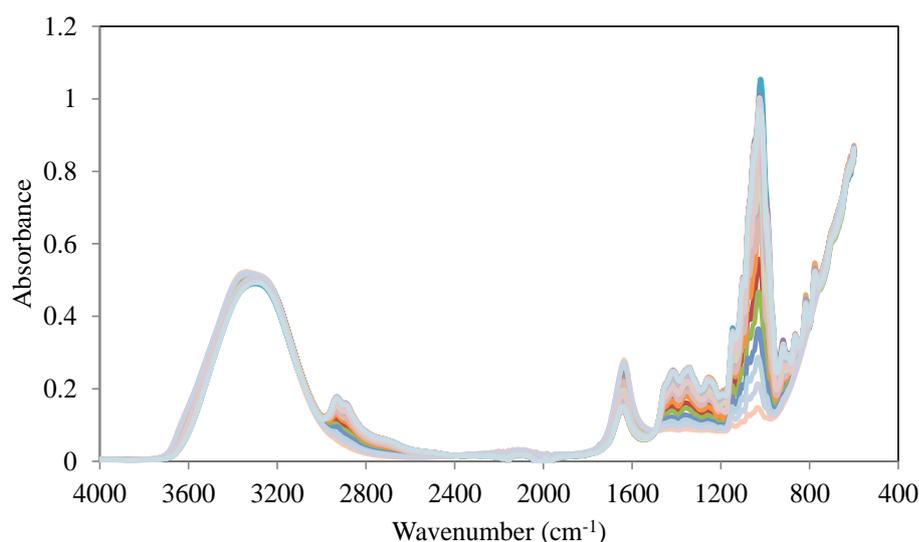


Figure 5.9. FTIR-ATR spectra a total of 65 authentic and adulterated honey samples prepared with corn syrup.

As can be seen from Figure 5.9, small variations among the spectral intensities of all 65 samples are visible and this absorbance differences becomes more apparent especially in the fingerprint region ($<1500\text{ cm}^{-1}$) because of the differences in the concentration of the corn syrup.

As in the previous scenario, the multivariate calibration models were developed with GILS and PLS methods. As mentioned in Table 4.6, out of 65 samples, 45 of them were selected randomly and used for modelling and the rest of them were used as independent validation data set. Actual concentrations vs. GILS predicted concentrations plots of authentic honey and corn syrup contents are given in Figure 5.10.

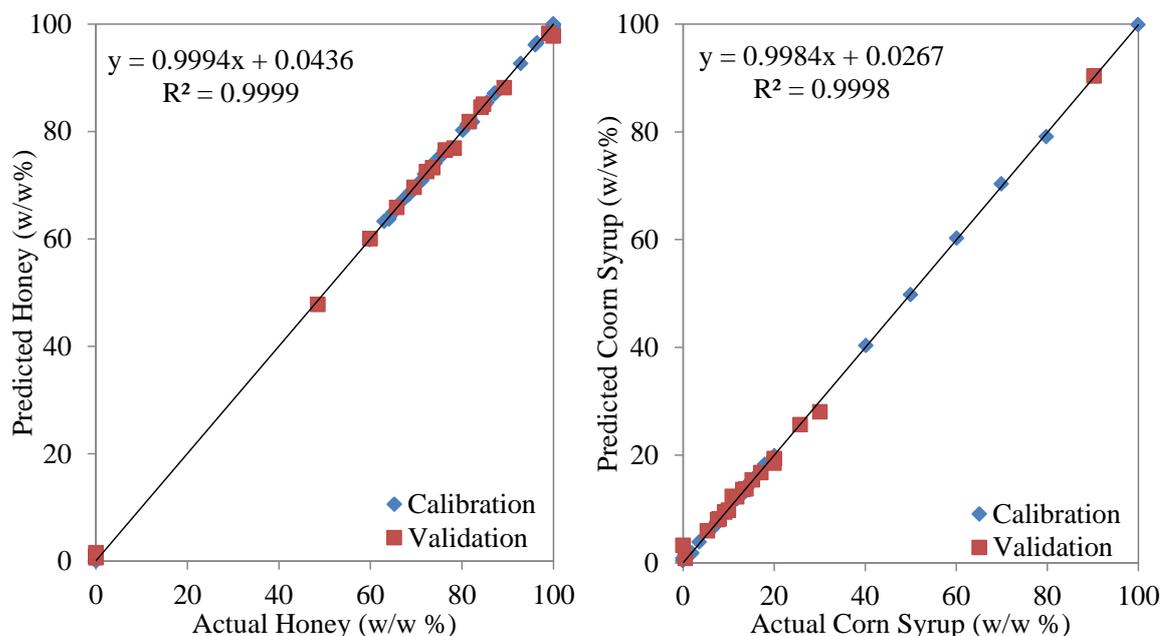


Figure 5.10. Actual versus predicted plot of honey and corn syrup contents resulted from GILS in the second scenario.

As shown in Figure 5.10, by looking at the correlation coefficients, it is possible to say that both models are successful. The R^2 values of the each content are more than 0.99. In order to compare the predictive success of these models SECV, SEP values should be considered with their dynamic ranges. SECV, SEP and R^2 values are given in detail along with their dynamic ranges in Table 5.3.

Table 5.3. Standard error of cross validation (SECV), standard error of prediction (SEP), maximum and minimum operating range of components (Max and Min) and correlation coefficient (R^2) of GILS models belong to second adulteration scenario.

	SECV (w/w %)	SEP (w/w %)	Max	Min	R^2
Honey	0.24	1.43	100	0	0.9999
Corn Syrup	0.30	1.19	100	0	0.9998

The results in Table 5.3 shows that despite of having wide operating range, both methods have very small SECV and SEP values with R^2 values more than 0.99.

In addition, PLS model was performed with the same calibration and independent validation data sets in order to compare the prediction performance with GILS models. Actual concentrations vs. PLS predicted concentrations plots of honey and corn syrup contents are shown in Figure 5.11.

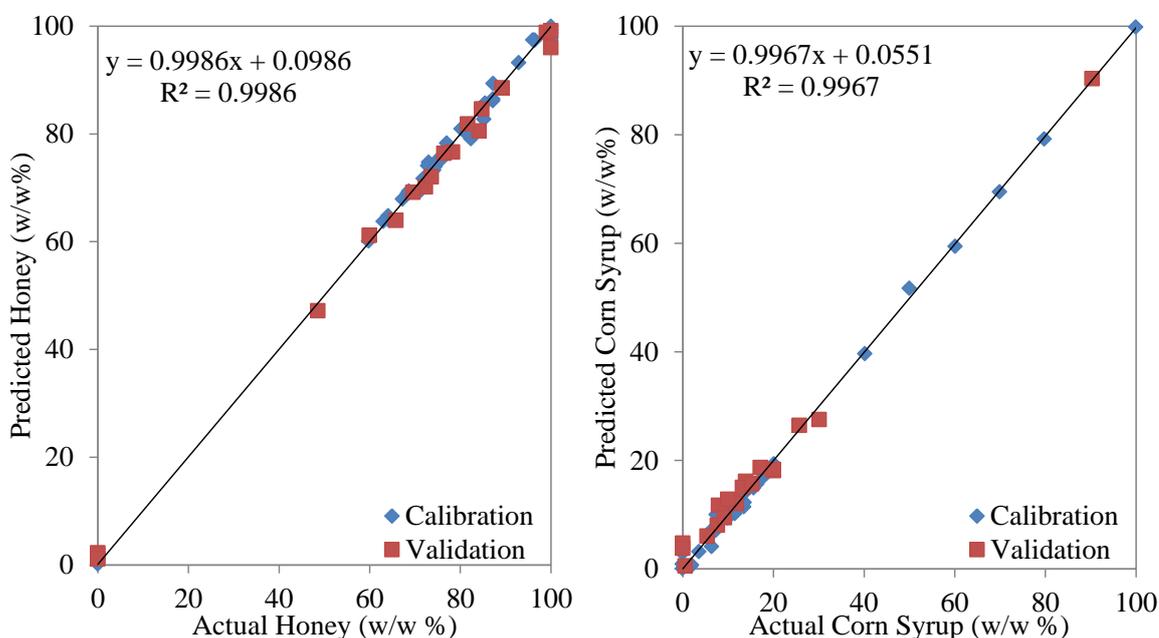


Figure 5.11. Actual versus predicted plot of honey and corn syrup contents resulting from PLS model in the second scenario.

As seen in Figure 5.11, R^2 values of both models are 0.9986 and 0.9967 and these values are found to be slightly smaller than GILS models. In order to evaluate the performance of these PLS models in depth, the number of PC, SECV, SEP and R^2 values are given in Table 5.4 along with their dynamic ranges.

Table 5.4. The number of principle components (PC), standard error of cross validation (SECV), standard error of prediction (SEP), maximum and minimum operating range of components (Max and Min) and correlation coefficient (R^2) of PLS models belonging to the second adulteration scenario.

	Number of PC	SECV (w/w %)	SEP (w/w %)	Max	Min	R^2
Honey	8	1.16	1.83	100	0	0.9967
Corn Syrup	8	1.28	2.11	100	0	0.9967

As shown in Table 5.4, both honey and corn syrup models were developed with 8 principle components. There are significant differences between SECV values of GILS and PLS models but the SEP values of both methods are somewhat closer to each other. This indicates that GILS models were partially over fitted but the SEP values of GILS were still lower than PLS.

Besides, in order to test the success of these GILS and PLS models 100 authentic and commercial samples that were collected in three different harvest seasons (explained in Chapter 4.1) were introduced to the calibration models and their honey and corn syrup contents were predicted. The FTIR-ATR spectra of these 100 honey samples is shown in Figure 5.12.

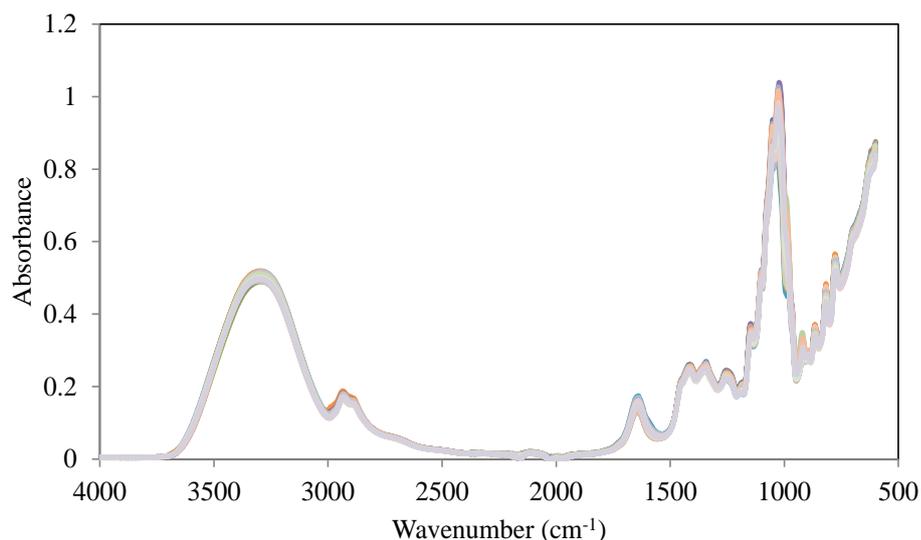


Figure 5.12. FTIR-ATR spectra total of 100 authentic and commercial honey samples.

Because these 100 samples were expected to be authentic honey samples whose FTIR-ATR spectra are shown in Figure 5.12, there is no visible difference among their spectral information, similar to the other authentic honey samples. In addition, these 100

authentic and commercial honey samples are also used in the following adulteration scenarios as secondary test samples.

The results of prediction contents of 100 pure and commercial honey samples are given in Figure 5.13 and Figure 5.14 to compare the predictive performance of GILS and PLS methods.

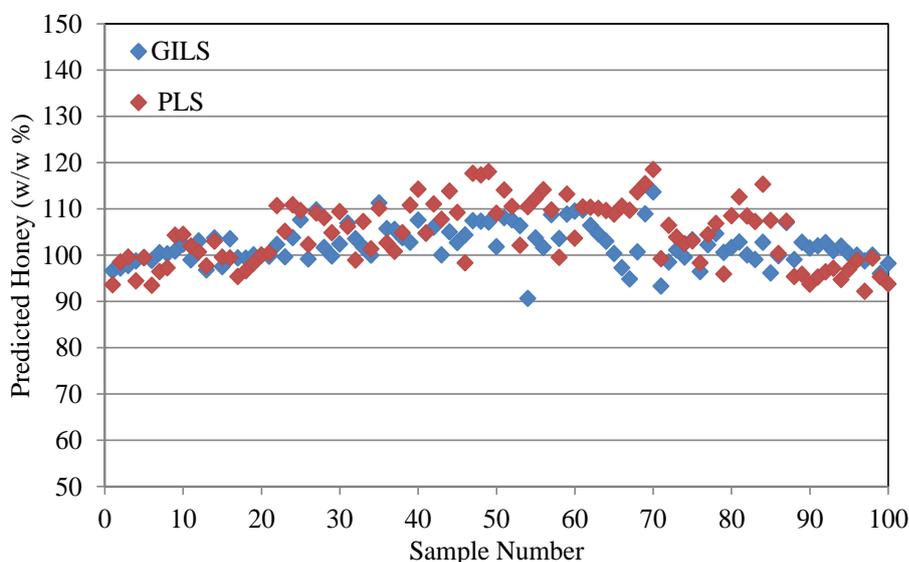


Figure 5.13. Honey content prediction of 100 authentic and commercial honey samples with both GILS and PLS methods.

The GILS and PLS prediction of honey content of 100 commercial and authentic honey samples given in Figure 5.13 indicates that there is a curved prediction trend among the 100 samples with both GILS and PLS. The possible explanation of this curved trend could be associated with the time of the spectra collection since not all of these 100 samples FTIR spectra were taken at the same time due to the different collection seasons. As seen from the predictions, except two samples, honey content of 98 samples was predicted between 90% and 110% (w/w %) by using GILS model. So the prediction range of honey content is found $\pm 10\%$ (w/w %). While the GILS predictions appear to be fluctuated evenly in the range of 90-110%, the PLS predictions are in the range of 90%-120% (w/w %) with most of the predictions being positively biased.

Furthermore, the corn syrup content of these samples was also investigated by using the same models. Predicted corn syrup concentrations of all 100 samples resulting from GILS and PLS models are shown in Figure 5.14.

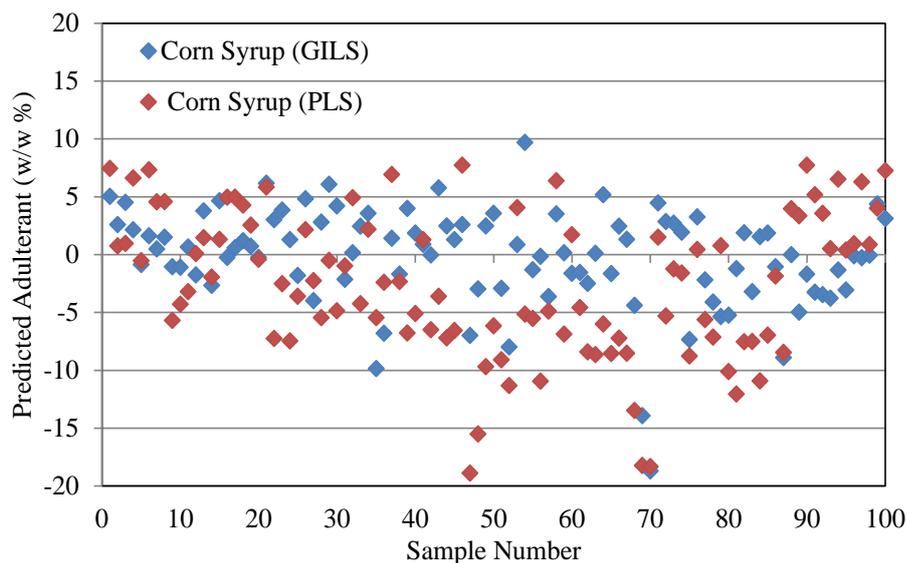


Figure 5.14. Prediction of corn syrup content of 100 authentic and commercial honey samples with both GILS and PLS methods.

As observed in Figure 5.14, the prediction of corn syrup content of the mentioned 100 authentic and commercial honey samples ranges between -10% and 10% (w/w %). However, 2 samples predicted by GILS and 9 samples by PLS appear to be the outliers and found on the negative side. As a result of this, prediction performance of PLS model have been interpreted to be lower than GILS model and this can backed up with their SEP values. In addition, as this scenario was only based on two component mixture (honey and corn syrup) there is no independence of the components of the mixtures since their sum adds up to 100%. This situation forces the models to be over fitted which can be seen in their SECV values.

5.2.2. FTNIR Results of Second Adulteration Scenario

In this part of the study, FTNIR spectroscopy was used to develop multivariate calibration models namely GILS and PLS. The 65 samples prepared for the second adulteration scenario, which had been used for FTIR analyses, were diluted to 1:10 with water. The FTNIR spectra of 65 samples are given in Figure 5.15.

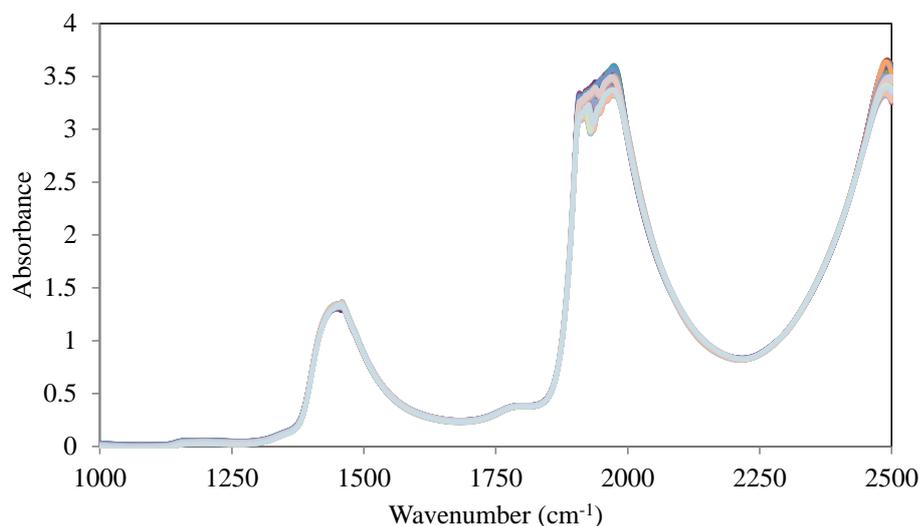


Figure 5.15. FTNIR spectra of the 65 authentic and adulterated honey samples.

As shown in Figure 5.15, even though the samples are diluted 1:10 and the cuvette used was quite narrow (1mm) there exists some absorbance values around 3.5, which is the saturation point, especially on the right side of 2000 cm^{-1} . However, existence of this absorbance may affect the linear correlation negatively according to the Beer's law, the variable selection property of GILS algorithm is expected to provide the models that avoids that range.

The multivariate calibration models with FT-NIR spectra were developed by using data set having same samples numbers in FTIR analysis as mentioned in Chapter 5.2.1. Resulting predicted concentrations from GILS model versus the actual concentrations are shown in Figure 5.16.

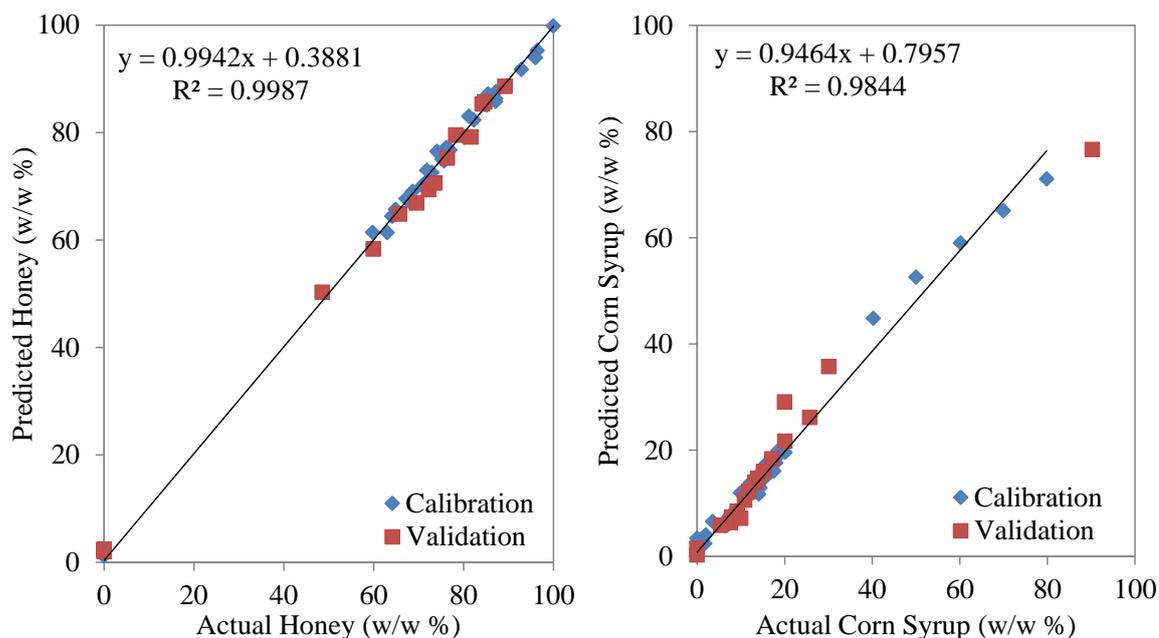


Figure 5.16. Actual versus predicted plots of honey and corn syrup resulting from GILS in the second scenario with FTNIR spectroscopy.

As can be seen in Figure 5.16, although same samples in calibration and independent validation sets were used to develop multivariate calibration models, corn syrup model are found to be more scattered around the regression line. As a result of this, the regression coefficient for corn syrup was found to be 0.9844 while for honey content it was 0.9987. To interpret these models in detail, SECV, SEP and R^2 values are given along with their dynamic ranges in Table 5.5.

Table 5.5. Standard error of cross validation (SECV), standard error of prediction (SEP), maximum and minimum operating range of components (Max and Min) and correlation coefficient (R^2) of GILS models belong to second adulteration scenario with FTNIR spectroscopy.

	SECV (w/w %)	SEP (w/w %)	Max	Min	R^2
Honey	1.15	4.24	100	0	0.9987
Corn Syrup	2.36	4.71	100	0	0.9844

As seen in Table 5.5, although GILS models of honey and corn syrup contents have smaller SECV values, their SEP values are much greater especially when compared with the FTIR results. Moreover, these GILS models developed by using FTNIR spectroscopy have smaller R^2 than the ones of FTIR.

Additionally, PLS model was also applied as reference multivariate calibration method to the same calibration and independent validation data sets. The actual vs. predicted plots of PLS models for honey and corn syrup contents are shown in Figure 5.17.

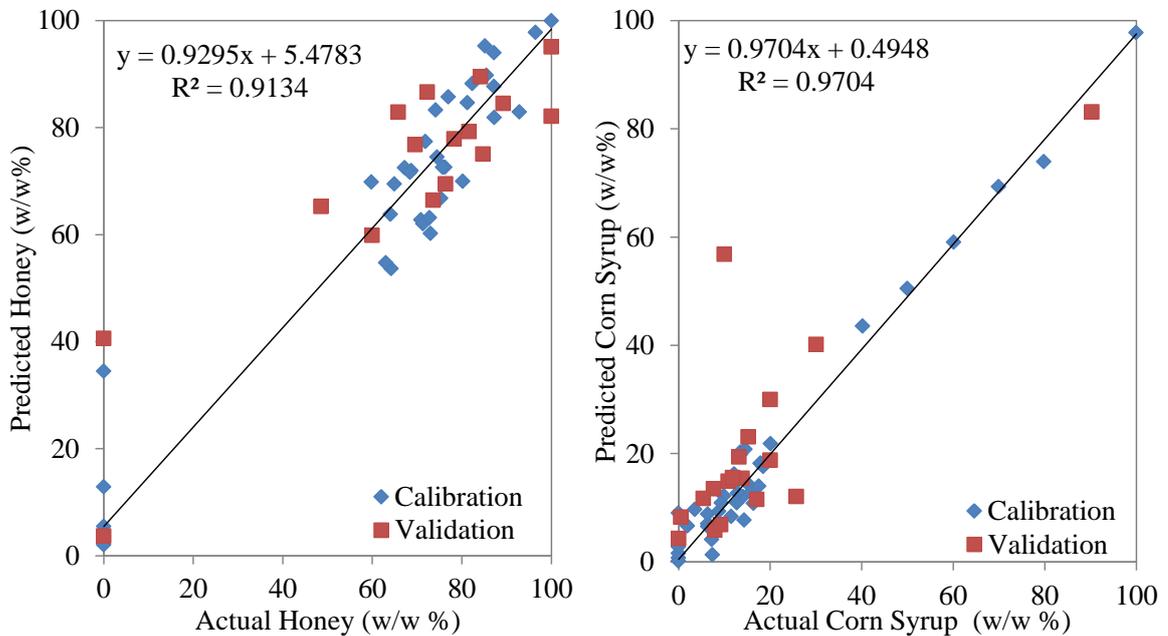


Figure 5.17. Actual versus predicted plots of honey and corn syrup contents resulting from PLS in the second scenario with FTNIR spectroscopy.

Actual concentration vs. PLS predicted concentration plots of honey and corn syrup contents, which are given in Figure 5.17, show that PLS model is not successful as GILS models, because R^2 values of honey is 0.9134 whereas R^2 of corn syrup is 0.9704. Number of PCs, SECV and SEP are given with their operating ranges to examine the performance of these models in Table 5.6.

Table 5.6. The number of principle components (PC), standard error of cross validation (SECV), standard error of prediction (SEP), maximum and minimum operating range of components (Max and Min) and correlation coefficient (R^2) of PLS models belong to second adulteration scenario with FTNIR spectroscopy.

	Number of PC	SECV (w/w %)	SEP (w/w %)	Max	Min	R^2
Honey	7	9.08	16.52	100	0	0.9134
Corn Syrup	15	3.82	12.43	100	0	0.9704

As shown in Table 5.6, both honey and corn syrup PLS models have greater SECV and SEP values while they have smaller R^2 values than GILS models. The reason of having weak correlations in this model can be explained by the saturation in the FTNIR spectra which is used as is while modelling. SECV and SEP values of honey and corn syrup models, which were developed with 7 and 15 principle components respectively, suggest that using these models for prediction of honey and corn syrup contents is not suitable.

In this part of the study, 20 commercial and 15 authentic honey samples were introduced to the models to test their successes rather than using all 100 samples due to avoid time consuming dilution procedure without prior performance knowledge and the FTNIR spectra of these 35 samples is given in Figure 5.18.

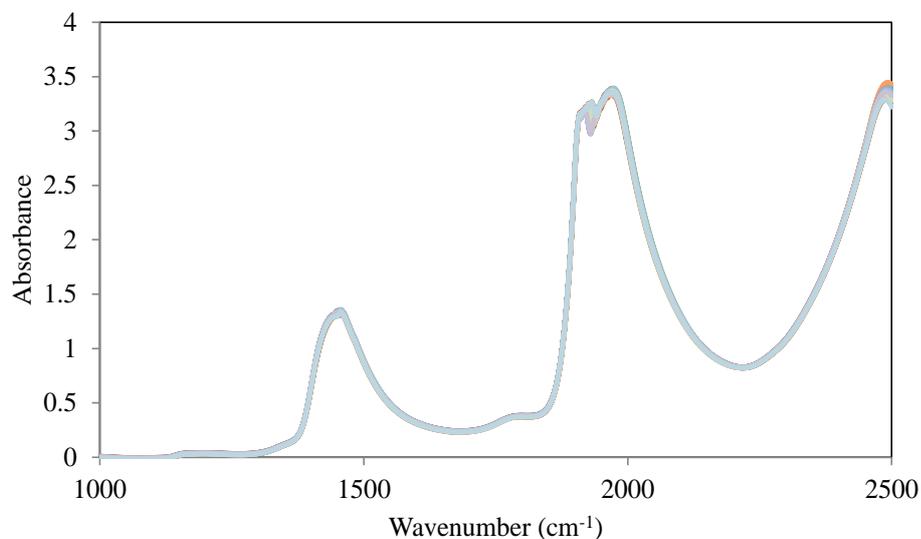


Figure 5.18. FTNIR spectra of 35 authentic and commercial honey samples.

FTNIR spectra of 35 authentic and commercial honey samples were shown in Figure 5.18. There are no significant differences among the absorbance values of these honey samples.

The results of prediction honey contents of 35 pure and commercial honey samples are given in Figure 5.19 and prediction corn syrup contents are given in Figure 5.20 to compare the predictive performance of GILS and PLS methods.

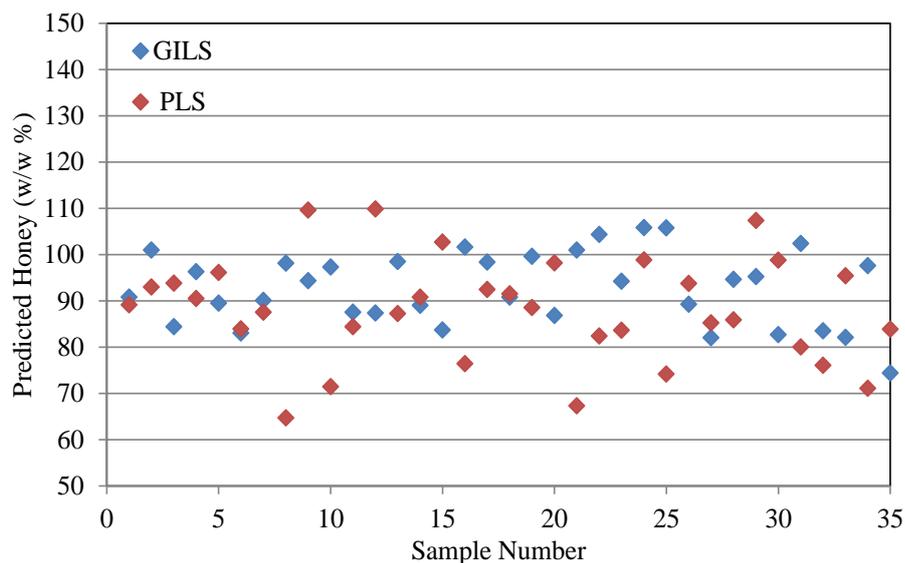


Figure 5.19. Honey content prediction of 35 authentic and commercial honey samples with both GILS and PLS methods with FTNIR spectroscopy.

As seen in Figure 5.19, the prediction ability of these models that were developed by using FTNIR spectroscopy is weaker than the FTIR spectroscopy. In this study, 10 out of 35 samples were predicted to be below 90% using GILS model whereas using PLS the 20 out of 35 samples were predicted below 90% with relatively lower concentrations. Thus, when these two models are compared, honey predictions are worse for the PLS model. Then, the same 35 samples' spectra were used to predict their corn syrup contents. The prediction results of GILS and PLS methods for corn syrup contents are shown in Figure 5.20.

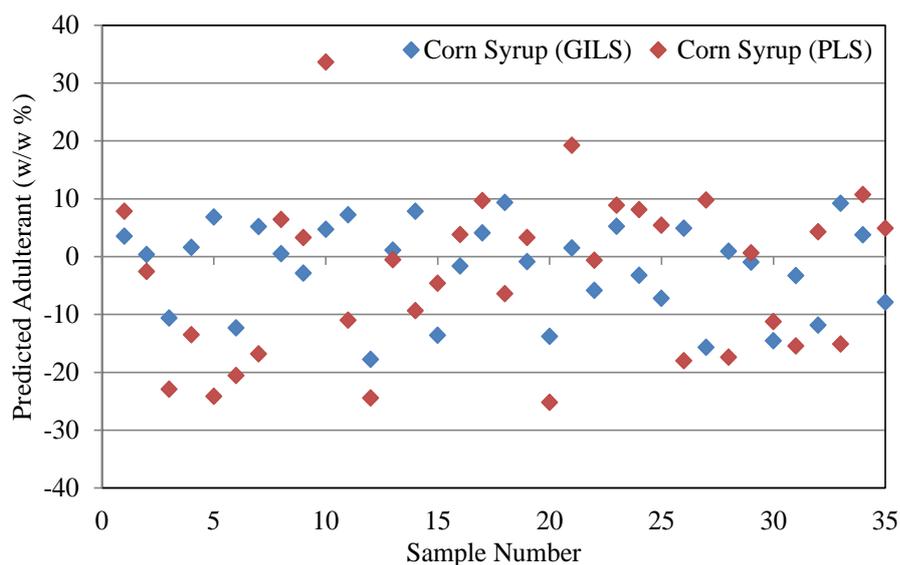


Figure 5.20. Corn syrup content prediction of 35 authentic and commercial honey samples with both GILS and PLS models using FTNIR spectroscopy.

As can be seen in Figure 5.20, while FTNIR spectroscopy was carried out for data collection, prediction of corn syrup content with GILS and PLS methods were spread over a wider range than the results of FTIR spectroscopy. While the upper limit of GILS predicted samples was found to be 10% (w/w %), the lower limit was spread out to -20% (w/w %). However, the predicted concentrations of PLS were found in wider concentration range and it changed from 25% to 35% (w/w %).

Consequently, when the models developed with FTIR and FTNIR spectra are compared, it is obviously seems that the models with the results of FTIR spectroscopy are more reasonable than FTNIR. In addition, honey samples must be diluted for FTNIR analysis as they are very viscose which makes it almost impossible to take transmission measurement. With the help of ATR accessory it is a very simple task to collect FTIR spectra by applying a single drop of the sample regardless of the viscosity as long as homogeneity is provided. Moreover, even with 1:10 dilution of the sample NIR measurement still requires to use very narrow cuvettes such as 1 mm or 2 mm thickness and this makes difficult to clean cuvette from sample to sample each time.

5.3. The Third Adulteration Scenario (Corn Syrup, Beet Sugar and Water were Used as Adulterants)

In this adulteration scenario more adulterants were used than the previous ones and design of this scenario was shown in Table 4.3. A total of 109 adulterated and authentic samples were prepared for this adulteration scenario. Among them, 103 authentic and adulterated honey samples were used in the calibration and independent validation data sets and selection of samples was done randomly and shown in Table 4.7. The remaining 6 samples, which are to simulate a special case where no honey is present, were assigned as a minor test set to evaluate the prediction ability of the developed GILS and PLS models. Similar to previous scenarios, FTIR-ATR spectroscopy was used for spectral analyses. FTIR-ATR spectra of these samples are illustrated in Figure 5.21.

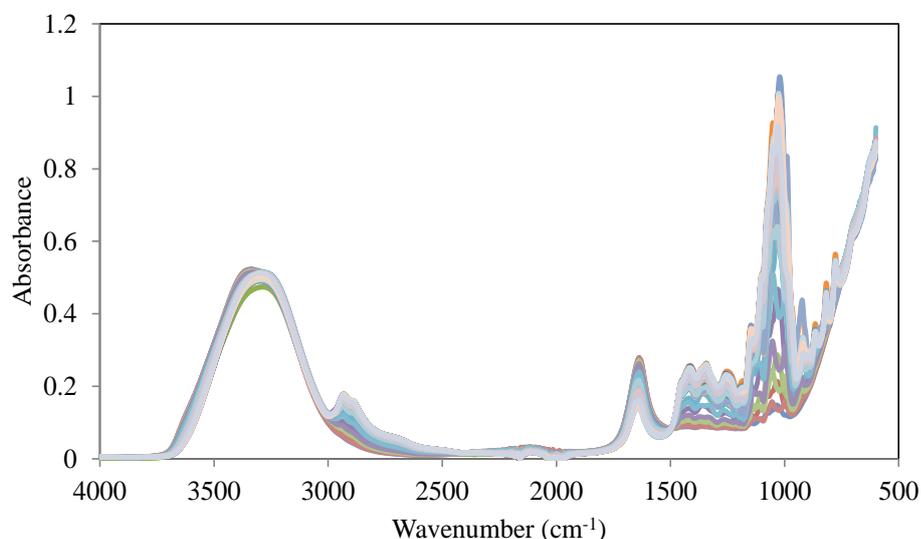


Figure 5.21. FTIR spectra of 103 adulterated and authentic honey samples prepared with corn syrup, beet sugar and water in the third adulteration scenario.

Even though full spectra of the 103 samples are shown in Figure 5.21, when the wavenumber range narrowed down to 1500 – 900 cm^{-1} interval, the differences among them become more apparent and the narrow region FTIR-ATR spectra are shown in Figure 5.22.

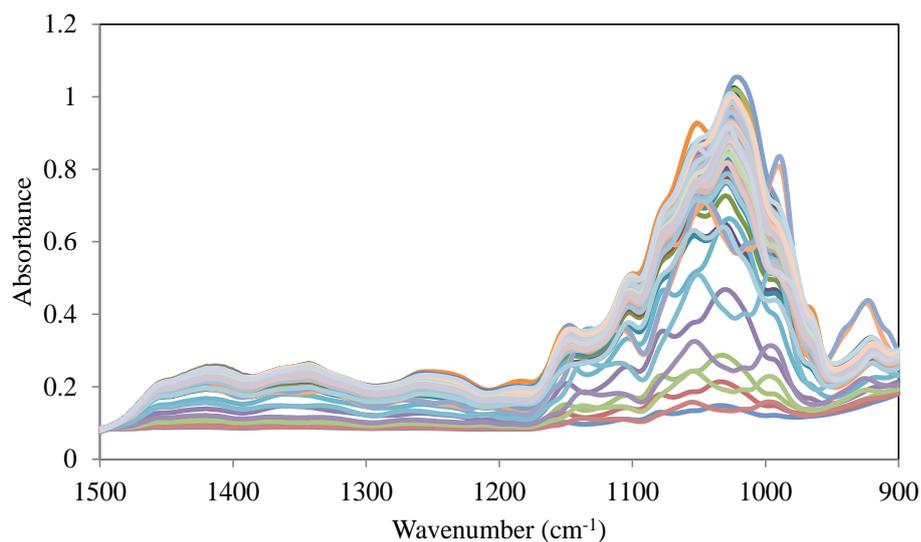


Figure 5.22. Fingerprint region of FTIR spectra between 1500 and 900 cm^{-1} wavenumber range of 103 adulterated and authentic honey samples prepared with corn syrup, beet sugar and water in the third adulteration scenario.

As demonstrated by Figure 5.22, when wavenumber range was narrowed down, intensity differences between the adulterated samples became visible. These differences are primarily due to the fact that adulterated honey samples were prepared with adding different amount of adulterants to the pure honey samples. Additionally, a number of 13 samples do not have any pure honey and they prepared with only beet sugar or corn syrup diluted with water in various concentrations to improve the operating ranges of these adulterants while modeling process. Furthermore, GILS and PLS exploits these spectral differences to identify the concentration of adulterants.

Like the previous adulteration scenarios, calibration models with GILS and PLS were performed and the calibration models were developed for each content (honey corn syrup, beet sugar and water), individually. The plots that show the actual concentrations vs. GILS predicted concentrations are shown in Figure 5.23.

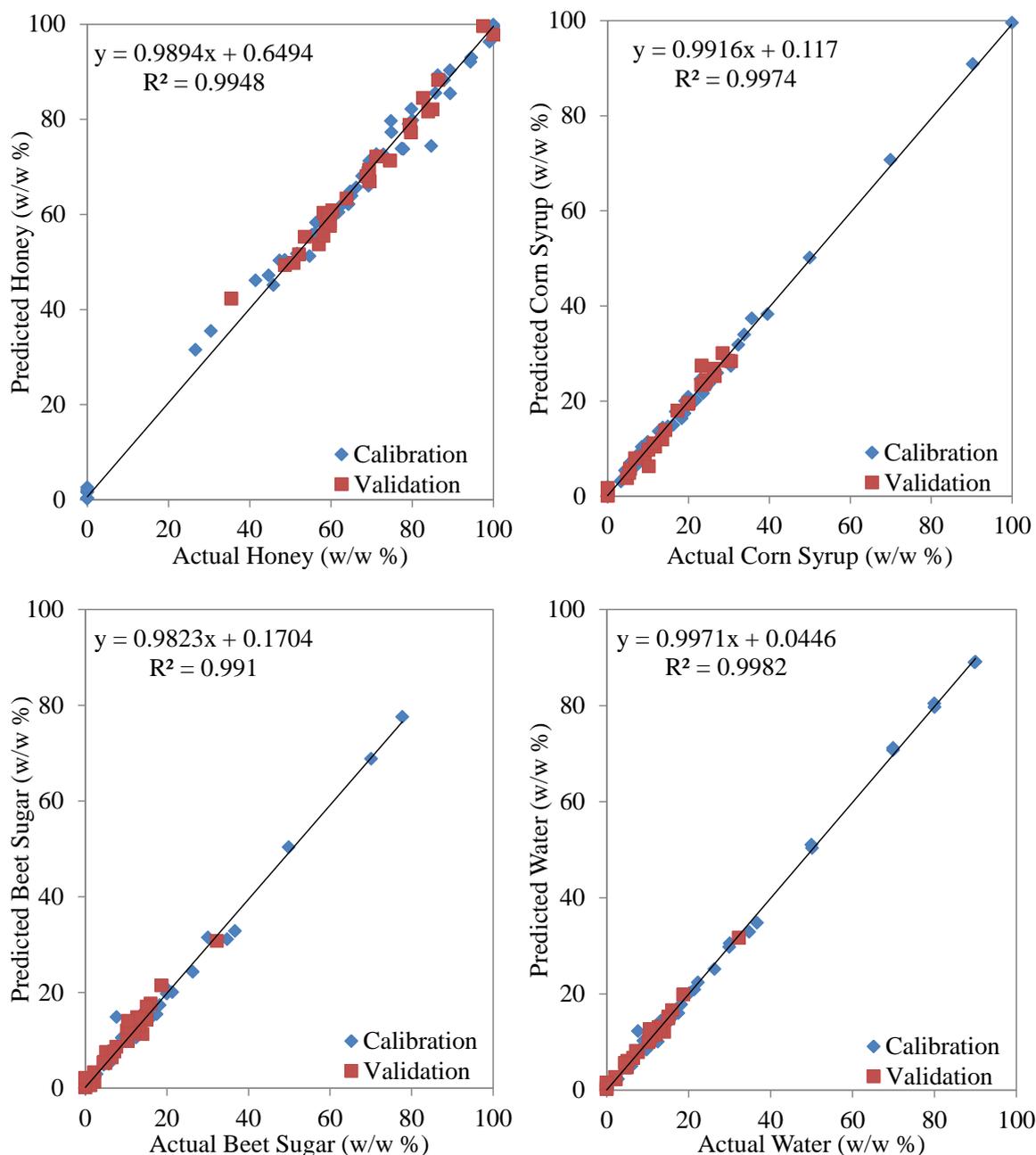


Figure 5.23. Actual vs. predicted plots of honey, corn syrup, beet sugar and water contents obtained from GILS method in the fourth scenario with FTIR spectroscopy.

The plots given in Figure 5.23 indicates that operating range of honey and corn syrup contents were set between 0% and 100% (w/w %), whereas beet sugar was changed between 0% and 80% (w/w %) and water was between 0% and 90% (w/w %). Additionally, all R^2 values of average predictions resulting from 100 best models used are over 0.99. To compare these models in detail SECV, SEP and R^2 values of each content are given with corresponding dynamic ranges in Table 5.7.

Table 5.7. Standard error of cross validation (SECV), standard error of prediction (SEP), maximum and minimum operating range of components (Max and Min) and correlation coefficient (R^2) of GILS models belong to the fourth adulteration scenario with FTIR-ATR spectroscopy.

	SECV (w/w %)	SEP (w/w %)	Max	Min	R^2
Honey	2.52	2.19	100	0	0.9946
Corn Syrup	0.98	1.64	100	0	0.9976
Beet Sugar	1.43	1.54	80	0	0.9910
Water	0.97	0.90	90	0	0.9982

According to the statistical parameters given in Table 5.7, dynamic ranges of all the models for honey, corn syrup, beet sugar and water contents are almost between 0 % and 100 % (w/w %) and this allows the comparison of their SECV and SEP values. The SECV value of honey content was calculated as 2.52 (w/w %) while corn syrup content was 0.98 (w/w %), furthermore the SEP value of honey content was calculated as 2.19 (w/w %) while corn syrup content was 1.64 (w/w %). These results indicate that corn syrup model is slightly better than honey model. Moreover, the concentration of beet sugar change between 0 % and 80 % (w/w %), while the concentration of water content change between 0 % and 90 % (w/w %). Despite having a slightly narrower operating range, SECV and SEP values of beet sugar model were higher than water content.

The fact that GILS method is operated with a genetic algorithm for variable selection and model optimization, it was decided to generate frequency distribution of the most survived variables within the spectral range. The selection frequency of each variable is obtained by counting how many times these variables are used in all “runs”. The plot which contains frequency over average spectra can be helpful to identify which wavenumbers contributes to the modelling most since the genetic algorithm aims to find best set of variables by an iterative process. These plots are also useful in some cases to evaluate the performance of genetic algorithm by comparing the frequency of the selected variables to the entire spectrum of honey in order to determine whether intensely selected regions correspond the characteristic component information. Figure 5.24 shows the frequency plots for each model.

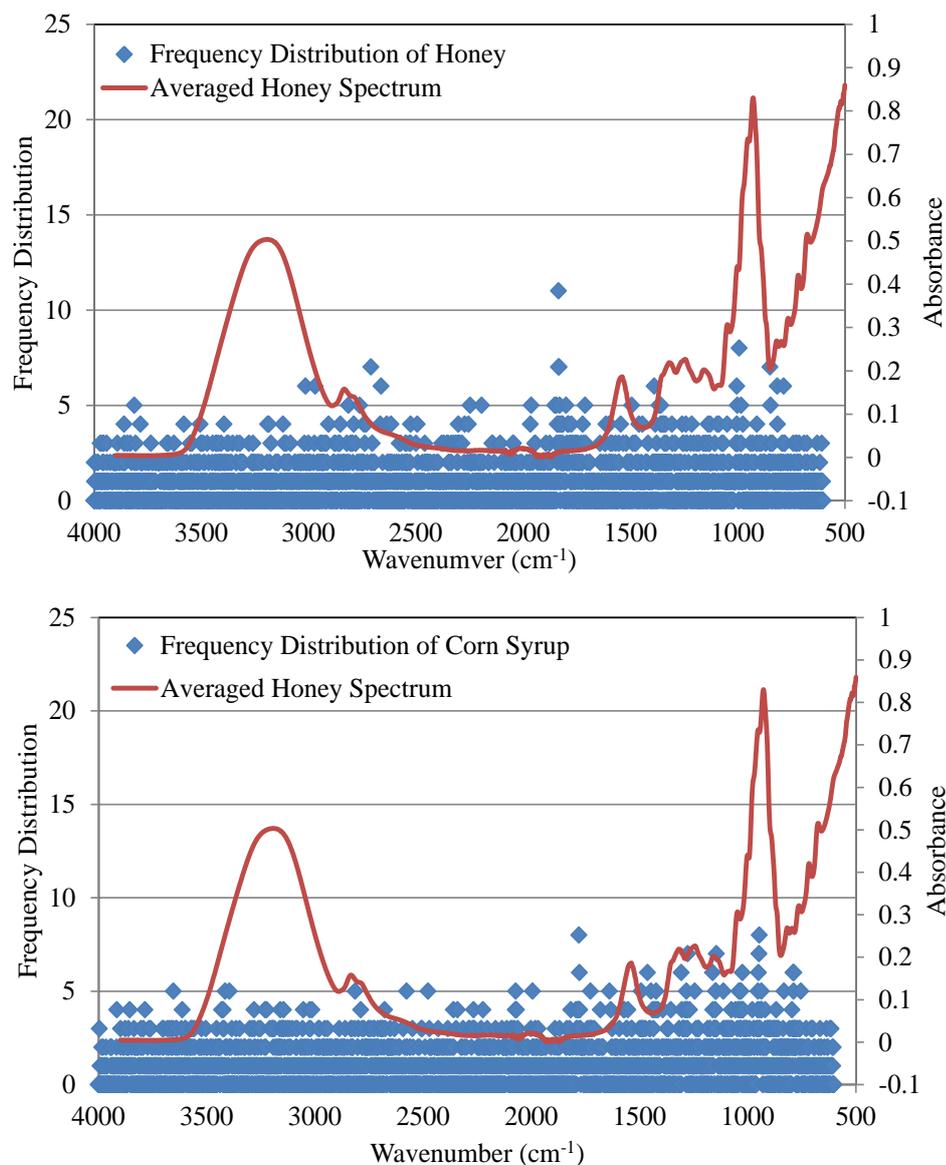


Figure 5.24. Frequency distribution plots that were obtained by performing GILS method for honey, corn syrup, beet sugar and water contents to develop multivariate calibration models.

(cont. on next page)

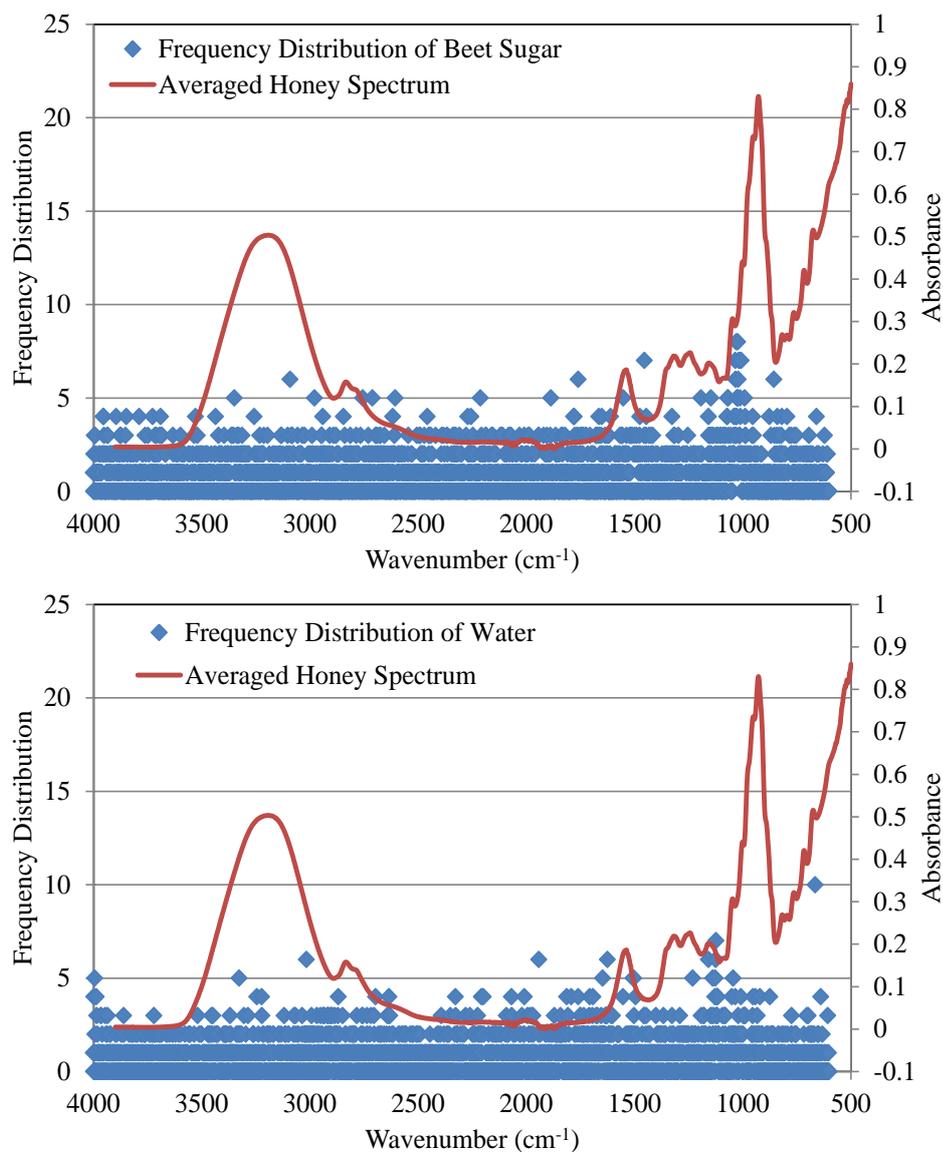


Figure 5.24. (cont.)

As can be seen in Figure 5.24, the predominant wavenumbers in frequency distribution of water model matches with the fingerprint of water indicating a successful variable selection of GA (Maréchal 2011). Thus, the wavenumbers with high selection frequencies corresponding to the honey, corn syrup, and beet sugar models have some similarities especially around 1000 cm^{-1} wavenumber range. In addition, high selection frequency around $1800\text{-}1600\text{ cm}^{-1}$ wavenumber range which corresponds carbonyl bonds was observed for the honey model whereas the selection frequency for corn syrup and beet sugar at this region is lower. These results demonstrate the ability of the genetic algorithm to extract relevant information from the whole spectral range for the component being modeled.

In addition, PLS models developed by using the same operating ranges with the previous GILS model were used as a reference calibration method as in previous scenarios. Actual vs. predicted plots of all four components obtained by PLS methods are given in Figure 5.25.

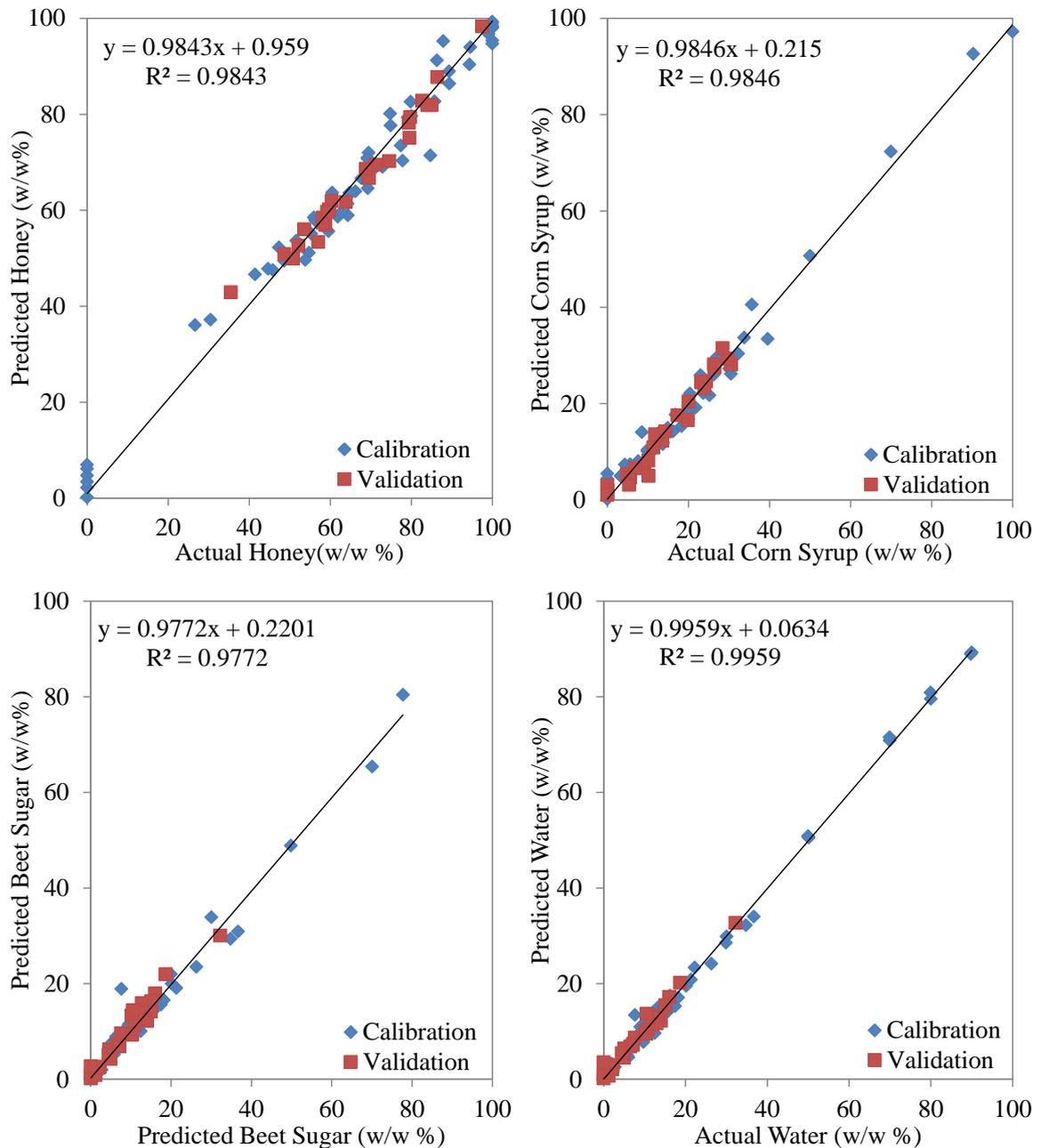


Figure 5.25. Actual vs. predicted plots of honey, corn syrup, beet sugar and water contents resulted from PLS method in the fourth scenario with FTIR spectroscopy.

Among the calibration plots that are shown in Figure 5.25, the correlation coefficients in descending order are: water, corn syrup, honey and beet sugar. In order to interpret calibration models, SECV, SEP, R^2 values their operating ranges along with number of principle components (PC) are shown in Table 5.8.

Table 5.8. Number of principle components (PC), standard error of cross validation (SECV), standard error of prediction (SEP), maximum and minimum operating range of components (Max and Min) and correlation coefficient (R^2) of GILS models belong to the fourth adulteration scenario with FTIR-ATR spectroscopy.

	Number of PC	SECV (w/w %)	SEP (w/w %)	Max	Min	R^2
Honey	9	4.27	2.49	100	0	0.9843
Corn Syrup	10	2.43	2.55	100	0	0.9846
Beet Sugar	8	2.26	1.66	80	0	0.9772
Water	7	1.47	1.18	90	0	0.9959

The results given in Table 5.8 indicates that the R^2 values of honey, corn syrup, beet sugar and water contents, which are 0.9843, 0.9846, 0.9772 and 0.9959 respectively, are relatively smaller than GILS models. While the PLS models of honey and corn syrup were developed with 9 and 10 number of principle components between the dynamic range of 0 % to 100 % (w/w %), the number of PCs for beet sugar was 8 and its operating range was between 0 % and 80 % (w/w %) and the number of PCs for water was 7 and its operating range was between 0 % and 90 % (w/w %). The SECV value of honey content was calculated as 4.27(w/w %) and 2.43 (w/w %) for the corn syrup. Moreover, the SECV value of beet sugar was calculated as 2.26 (w/w %) and the SECV value of water was 1.47 (w/w %). Besides, when the SEP values of all four contents were compared, the smallest SEP value belongs to water content predictions which is 1.18 (w/w %), whereas honey content has the highest one as 2.49 (w/w %).

Similar with the FTIR results that were given in the second adulteration scenario (Chapter 5.2.1), the models generated for honey, corn syrup, beet sugar and water were also tested with the same secondary test set described in section 5.2.1 which consist of 100 authentic and commercial honey samples, whose FTIR-ATR spectra had been shown in Figure 5.12. The predicted honey content concentrations by using the GILS and PLS methods are given in Figure 5.26.

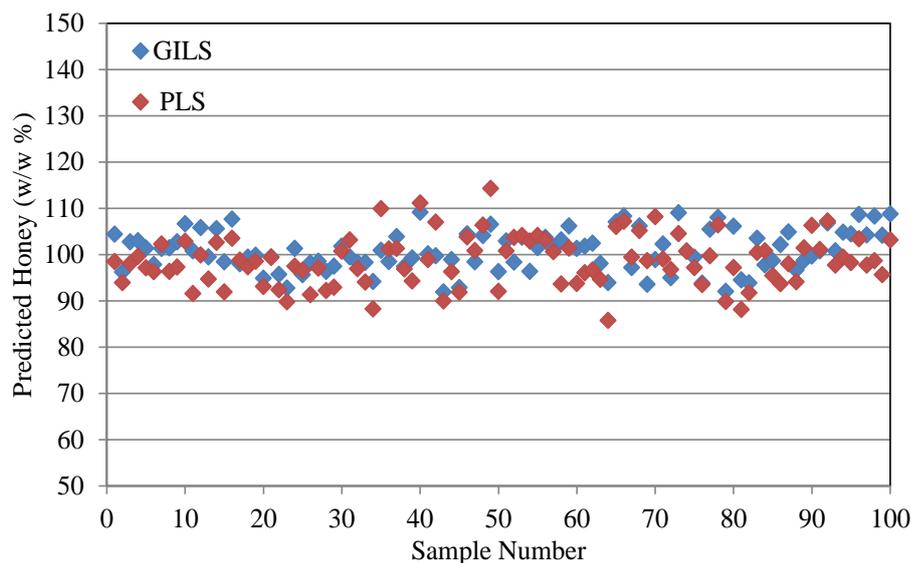


Figure 5.26. Honey content prediction of 100 authentic and commercial honey samples with both GILS and PLS methods with FTIR spectroscopy in the third adulteration scenario.

As can be seen from Figure 5.26, honey content of a total of 100 authentic and commercial honey samples were predicted. According to these results, honey concentration of the majority of the samples were predicted between 90% and 110% (w/w %), so the variability in the predicted values ranges around $\pm 10\%$ (w/w %) for both GILS and PLS models. In addition, when the graph is examined in detail, only the predictions of 6 samples by PLS were detected out of this range, whereas there is no sample predicted by GILS falls outside of this range.

Also, Figure 5.27 shows the prediction results of adulterants (corn syrup, beet sugar and water) with GILS and PLS models, respectively.

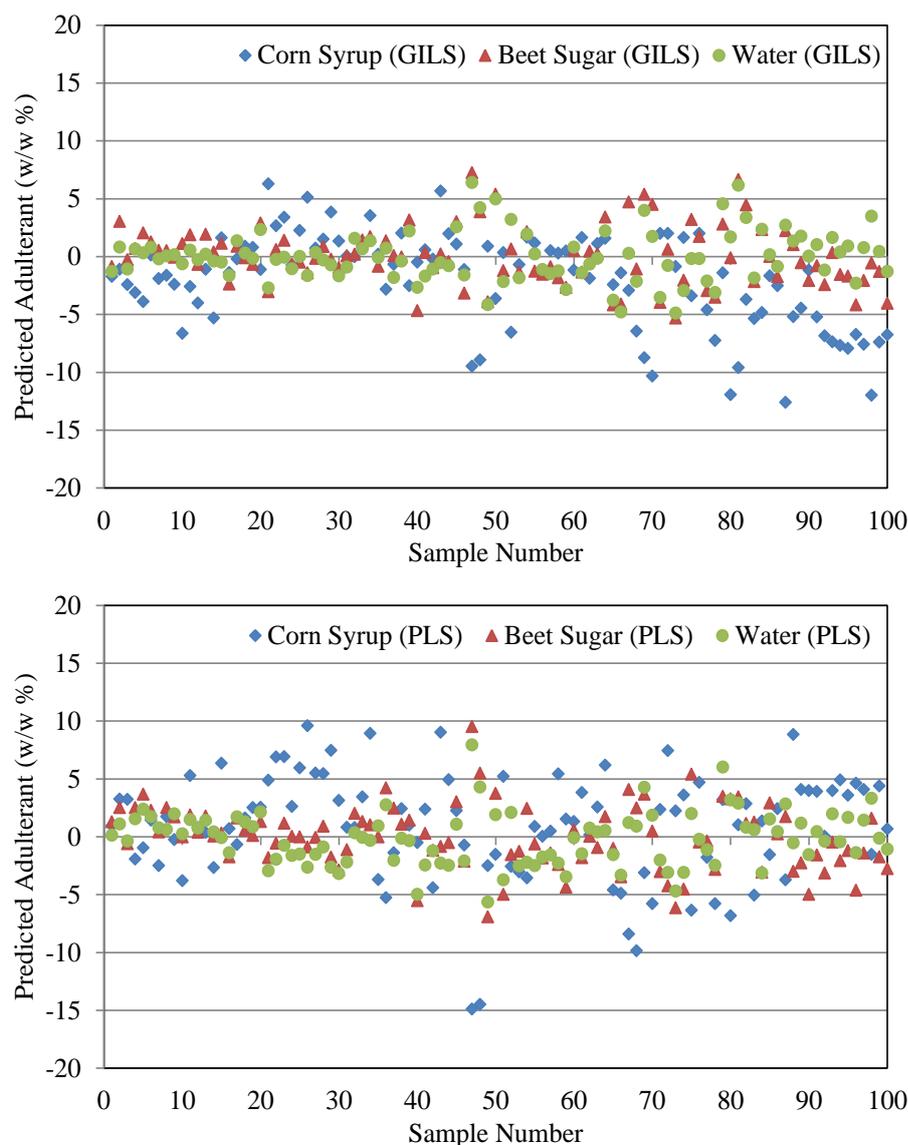


Figure 5.27. Predicted corn syrup, beet sugar and water contents of 100 authentic and commercial honey samples with both GILS and PLS methods with FTIR spectroscopy.

For the sake of comparison, the prediction results of corn syrup, beet sugar and water contents were given together in the same scale of ± 20 (w/w %) in Figure 5.27. However, the models generated for beet sugar and water resulted in the predictions mostly ranging from -5% to +5% deviations around 0% of these components as this test set contains only the pure honey samples. On the other hand, the model generated for the corn syrup by both GILS and PLS gave some prediction values which are outside of this $\pm 5\%$ (w/w %) interval. In addition, predicted corn syrup concentrations of more than 10 samples with PLS model were found more than 5% whereas more than 20 samples predicted with GILS were observed below to -5%.

Overall, the predictions of the 76 samples using PLS model for corn syrup resulted in $\pm 5\%$ (w/w %) interval. However, among the remaining 24 samples, 16 of them were deviated to positive side while 8 of them were deviated to negative side. Thus, corn syrup content of the majority of 100 samples were predicted between $\pm 5\%$ (w/w %) interval.

Quantitative prediction ability of the calibration models was evaluated by investigating the content of pure honey sample up to this step. In addition, the developed calibration models were used to predict concentrations of binary solutions which contain corn syrup-water and beet sugar-water contents. As illustrated in Table 4.3, from the total of 109 samples, 103 of them were assigned as calibration and independent validation data set where the 6 samples containing no honey were also predicted. These 6 samples contain 40%, 60%, 80% and 100% corn syrup and 60% and 80% beet sugar. They were introduced to GILS and PLS models developed in the third adulteration scenario to predict their contents in terms of corn syrup, beet sugar, water and honey concentrations. The results of quantitative determination for the contents of 6 samples obtained using the GILS models are given in Table 5.9 and the results obtained using the PLS models are given in Table 5.10.

Table 5.9. Predictions of binary mixtures of corn syrup with water and beet sugar with water from GILS method.

GILS Prediction	Corn Syrup		Beet Sugar		Water		Honey	
	Actual	Predicted	Actual	Predicted	Actual	Predicted	Actual	Predicted
1	40.22	38.96	0.00	-0.13	59.78	59.82	0.00	0.92
2	60.15	60.96	0.00	0.70	39.85	40.90	0.00	-2.38
3	79.82	79.62	0.00	-1.19	20.18	21.06	0.00	-0.63
4	100.00	97.84	0.00	-0.64	0.00	-1.24	0.00	4.96
5	0.00	-0.79	39.99	42.13	60.01	61.10	0.00	-2.24
6	0.00	-2.64	60.06	59.22	39.94	36.12	0.00	9.04

Table 5.10. Predictions of binary mixtures of corn syrup with water and beet sugar with water from PLS method.

PLS Prediction	Corn Syrup		Beet Sugar		Water		Honey	
	Actual	Predicted	Actual	Predicted	Actual	Predicted	Actual	Predicted
1	40.22	35.19	0.00	0.53	59.78	61.05	0.00	7.03
2	60.15	61.64	0.00	0.88	39.85	40.35	0.00	0.54
3	79.82	82.10	0.00	0.91	20.18	20.72	0.00	-1.85
4	100.00	91.27	0.00	-3.00	0.00	-2.09	0.00	12.53
5	0.00	0.92	39.99	45.32	60.01	61.18	0.00	-11.01
6	0.00	-2.18	60.06	60.65	39.94	36.43	0.00	4.63

As seen in Table 5.9 and 5.10, the first four samples had been prepared as corn syrup-water solutions and they included 40.22%, 60.15%, 79.82% and 100% (w/w %) corn syrup content and the last two samples had been prepared as beet sugar-water solutions and they included 39.99% and 60.06% (w/w %) beet sugar content. In order to evaluate these results in terms of the developed GILS and PLS prediction ability; graphical illustrations of each component along with their actual, GILS predicted and PLS predicted concentrations are given in Figure 5.28.

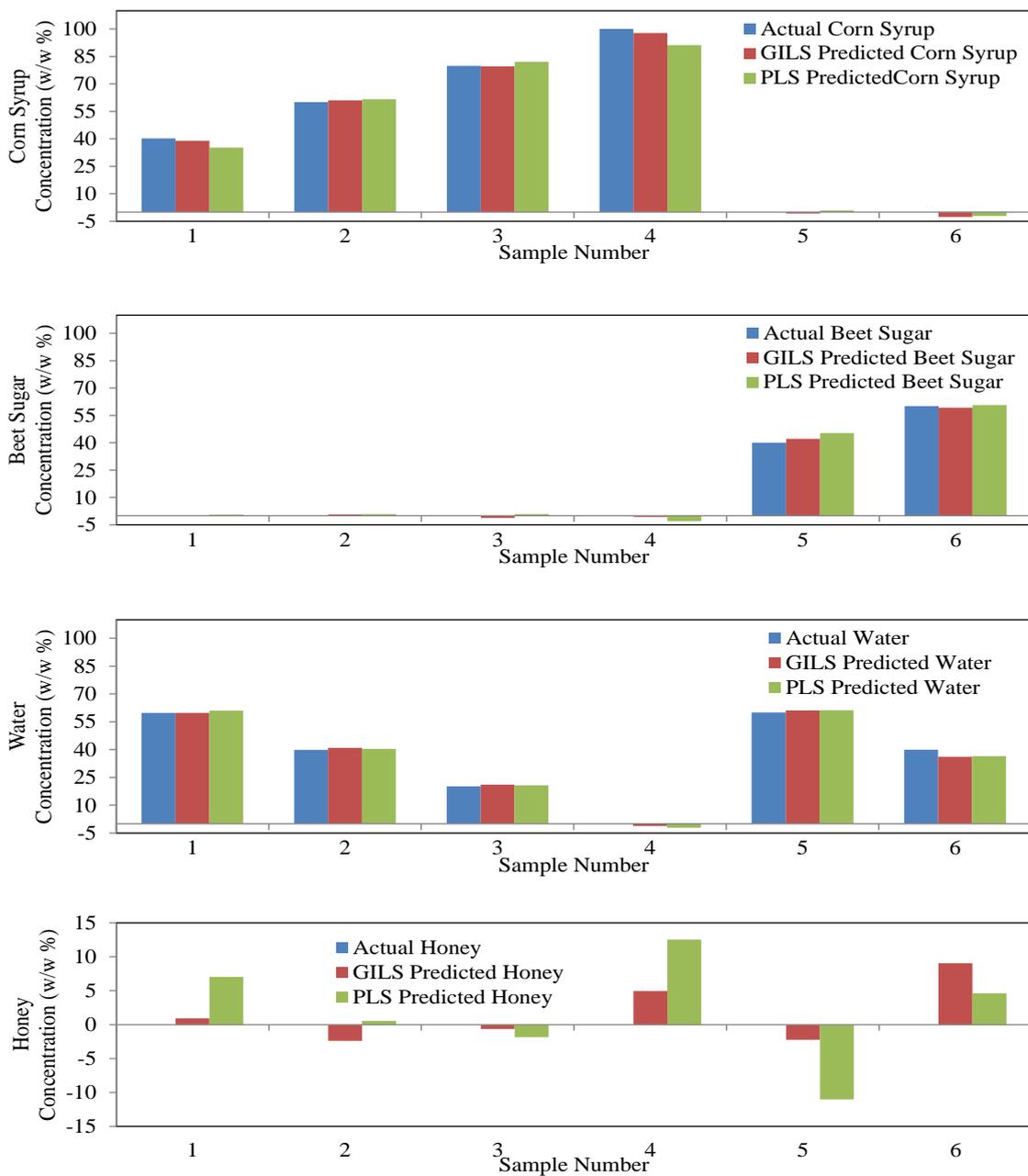


Figure 5.28. Comparison of prediction ability of the developed GILS and PLS models with corn syrup, beet sugar, water and honey contents.

As seen from Figure 5.28, the content of these six synthetically prepared corn syrup-water and beet sugar-water solutions were predicted by using the developed GILS and PLS models. The first graph represents distribution of actual concentration, GILS and PLS predicted concentration of corn syrup content for each of the six samples. Here only the one sample (sample number 6) which contains no corn syrup was predicted as -2.6% (w/w %) with GILS method whereas predicted values of sample number 1 and 4 were deviated from the actual value as 5% and 9%, respectively, in the negative direction. The results of the remaining samples predicted by both methods were in good agreement with the actual values. When the results of the beet sugar model were examined, it is seen that the GILS predictions for all the samples were better than the corresponding SEP value of the model. However, PLS predictions of beet sugar for the third and fourth samples were deviating around 3% to 5% (w/w %) from the actual value in the negative direction. Additionally, results of water content predictions demonstrated that the sample number 6 was predicted with a somewhat larger deviation (approximately 4%) by both GILS and PLS models but the other 5 sample prediction results were in good agreement with the actual values.

Although these 6 samples had no honey content, the models generated for honey by both GILS and PLS were also used to evaluate the predictive ability. As seen from the honey content prediction plot, sample number 4 and 5 were predicted above $\pm 5\%$ (w/w %) interval by using PLS method, whereas only the sample number 6 was predicted outside of $\pm 5\%$ (w/w %) range by using GILS method.

In addition, seven suspicious (possibly adulterated) honey samples, which were received from Ordu Apiculture Research Institute were used to test the prediction ability of the corresponding multivariate calibration models in this scenario. Honey, corn syrup, beet sugar and water content of these seven samples were predicted quantitatively by using the developed GILS and PLS models and their results are given in Table 5.11 and 5.12, respectively.

Table 5.11. Honey, corn syrup, beet sugar and water contents of seven suspicious honey samples predicted by GILS method.

GILS Prediction	Honey (w/w %)	Corn Syrup (w/w %)	Beet Sugar (w/w %)	Water (w/w %)	Total Prediction Amount
1	26.93	-5.20	30.52	49.03	101.27
2	10.93	86.94	-0.82	-1.25	95.80
3	9.10	96.67	-2.89	-4.42	98.46
4	77.04	-11.38	23.44	15.02	104.12
5	81.90	3.46	10.30	3.35	99.02
6	79.28	7.15	6.75	8.36	101.54
7	98.22	9.76	2.42	2.01	112.41

As shown in Table 5.11, honey content of the first three samples were predicted lower than %30 (w/w %), thus these samples appear to be highly adulterated. The beet sugar and water predictions of the first sample indicate that it is most probably adulterated with beet sugar and water while the following two samples were mostly corn syrup and honey. The honey content of the remaining four samples (4-7) were found to be ranging from 80% to 95% (w/w %) and the adulterants were ranged between 5% and 20% (w/w %). These results should be assessed with the SECV and SEP values of the developed models since these predictions also have some error which is as high as SEP values of the models. Moreover, honey, corn syrup, beet sugar and water contents of the same seven samples were predicted with the developed PLS models in order to compare the prediction ability of these two calibration methods and their results are shown in Table 5.12.

Table 5.12. Honey, corn syrup, beet sugar and water contents of seven suspicious honey samples predicted by PLS method.

PLS Prediction	Honey (w/w %)	Corn Syrup (w/w %)	Beet Sugar (w/w %)	Water (w/w %)	Total Prediction Amount
1	46.16	-20.64	27.54	47.52	100.59
2	16.16	83.56	-1.11	-0.94	97.66
3	24.10	80.21	-3.92	-5.33	95.06
4	86.71	-24.83	25.16	12.45	99.49
5	98.47	-10.31	10.01	1.84	100.01
6	84.98	5.44	5.28	5.15	100.84
7	85.39	17.26	-1.96	-0.47	100.22

The PLS predicted results shown in Table 5.12 indicate similar results with GILS predictions for the first three samples. Thus, it can be concluded that these samples (1-3) are adulterated honey samples. On the other hand, the remaining four samples (4-7) appear to be pure mostly honey (with at least 80% honey content) and around 20% corn syrup, beet sugar and water as possible adulterants.

Up to this point three different scenarios were investigated for the possible adulteration cases. Among them this third scenario would be the most likely adulteration attempt which may have been detected in real life. The models generated in this scenario demonstrated that it is possible to determine the amount of such adulterant namely corn syrup and beet sugar in pure honey. This judgement so far has been made based on SECV and SEP values of the models. On the other hand, it would also be helpful to evaluate standard residuals for the data used in the model building step (calibration and independent validation set) and the secondary test set. Figure 5.29 shows the standard residual plots for the models obtained with GILS along with the normal probability plot of these standard residuals. Similarly, standard residual plots for the models generated with PLS method were illustrated in Figure 5.30.

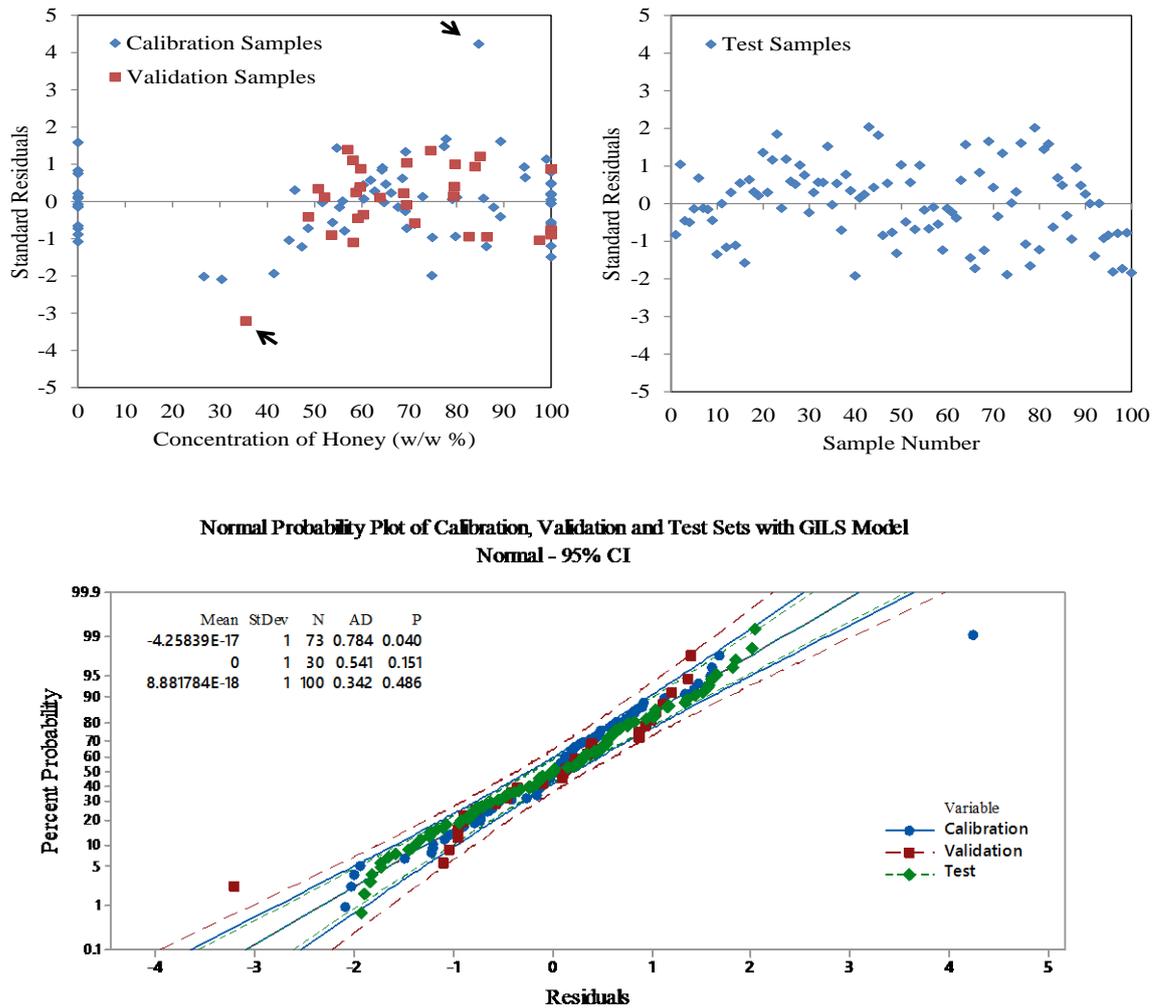


Figure 5.29. Standard residual and normal probability plots of the calibration, independent validation and test sets obtained from GILS method in the third adulteration scenario.

The upper left graph in Figure 5.29 shows the standard residual plot of the samples in the calibration and independent validation sets. The standard residual plot for the secondary test set is shown in the upper right graph. The third graph in the same figure illustrates the normal probability distribution of all three sets at 95% confidence level. As seen from the standard residuals plots of calibration and independent validation sets there is one sample in each set that are outside of the $\pm 3\sigma$ interval. The samples which are out of these boundaries can be assumed to be out of normal distribution with 95% confidence. Therefore, these two samples seem to be outliers. The standard normal probability plots also support this conclusion as the same samples in the calibration and independent validation sets were seen outside of the 95% confidence

interval boundaries of each set. In addition to these, P -values of normal probability plots indicates that the standard residuals of the calibration set is the only one that deviates from normal distribution as one of the sample is far from the confidence boundaries. On the other hand, there is also one sample in the independent validation set but the P -value (0.151) is above α value of 0.05 (at 95% confidence level), thus this deviation is not significant at the given confidence level. Finally, P -value of secondary test set is 0.486 which is almost 10 times larger than α value.

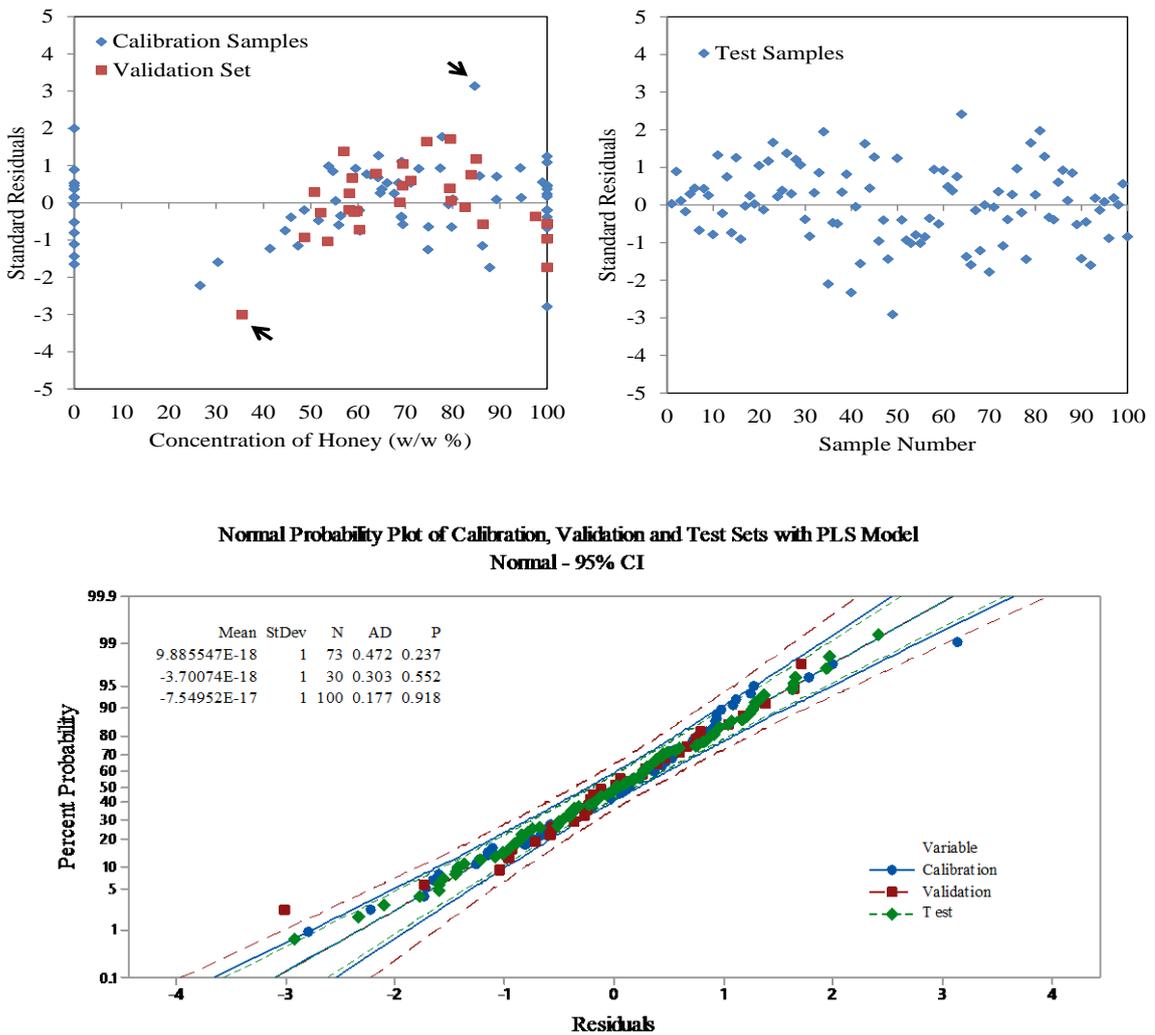


Figure 5.30. Standard residual and normal probability plots of the calibration, independent validation and test sets obtained from PLS method in the third adulteration scenario.

According to the standard residual plots of PLS predicted calibration and independent validation set, there is one sample from each set that is not in the $\pm 3\sigma$ range while there are no outliers in the predictions of the secondary test set. Normal probability plot for all three sets shows that the residuals follow normal distribution as can be seen from the P -values which are all greater than α value (0.05). The same two samples that are mentioned in Figure 5.29 are also identified as outliers, but their deviations are not high as given in the GILS models.

5.4. The Fourth Adulteration Scenario (Glucose, sucrose and Water Used as Adulterants)

In the fourth adulteration scenario, analytical grade glucose, sucrose and water were used as adulterant. The concentration profiles of the adulterated honey samples were explained in Table 4.4. Like the other adulteration scenarios, FTIR-ATR spectroscopy was used for data collection and the corresponding spectra are shown in Figure 5.31.

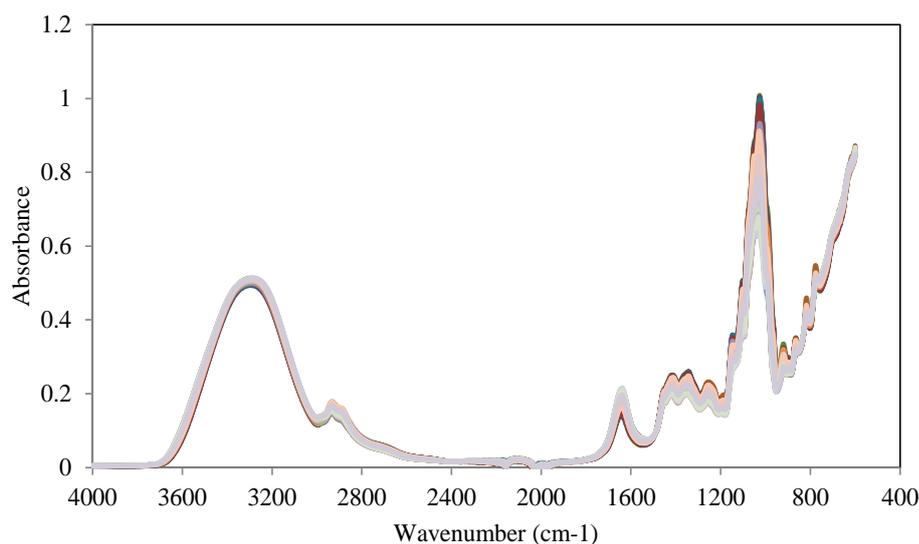


Figure 5.31. FTIR-ATR spectra of total of 58 authentic and adulterated honey samples prepared with using glucose, sucrose and water as adulterants in the fourth adulteration scenario.

As seen in Figure 5.31, FTIR-ATR spectra of these 58 authentic and adulterated honey samples have smaller spectral differences in terms of their absorbance values when compared to the adulteration scenarios created with corn syrup and beet sugar. These smaller differences can be explained by the relatively narrow concentration ranges of the adulterants namely glucose and sucrose. Although the concentration of corn syrup and beet sugar were changed from 0% to 90% in the previous adulteration scenarios, the concentration of glucose and sucrose ranged between 0% and 20% (w/w %).

As it is known from the studies in literature, sucrose is a disaccharide which can be decomposed to two isomeric monosaccharides namely glucose and fructose in acidic medium (Carmen and Gomez 2015). Thus, the acidity of the synthetically adulterated samples, which were prepared in this scenario with sucrose-water and glucose-water stock solutions, were measured and the pH values were found to be in acidic region (pH=5.5). Therefore, this scenario requires the modeling of three additional components which are total glucose (glucose + one half of the sucrose), fructose (one half of the sucrose) and total sugar (glucose + sucrose). As a result, a total of seven contents were modelled to cover all the possibilities.

Here, multivariate calibration models were developed by using GILS and PLS methods as in previous scenarios. A total of 58 authentic and adulterated honey samples were prepared whose concentration profiles were given in Table 4.4. As shown in Table 4.8, a total of 45 samples were selected for developing calibration models and rest of them were used to validate these models. Actual vs. GILS predicted concentrations are given in Figure 5.32 for each content which are honey, glucose, sucrose, total glucose, total sugar, fructose and water.

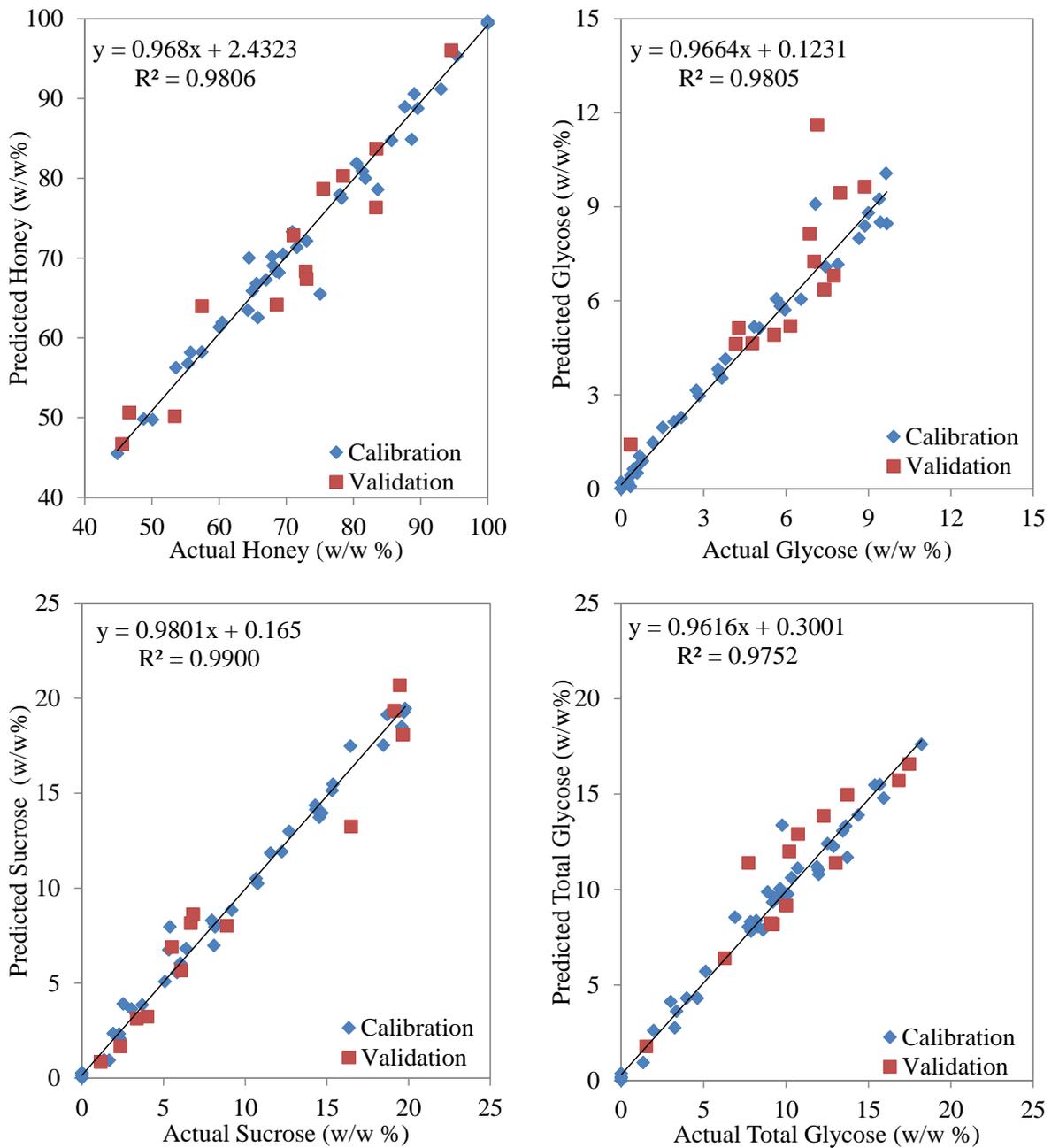


Figure 5.32. Actual versus predicted plots of honey, glucose, sucrose, total glycose, total sugar, fructose and water contents resulted from GILS in the fourth scenario.
(cont. on next page)

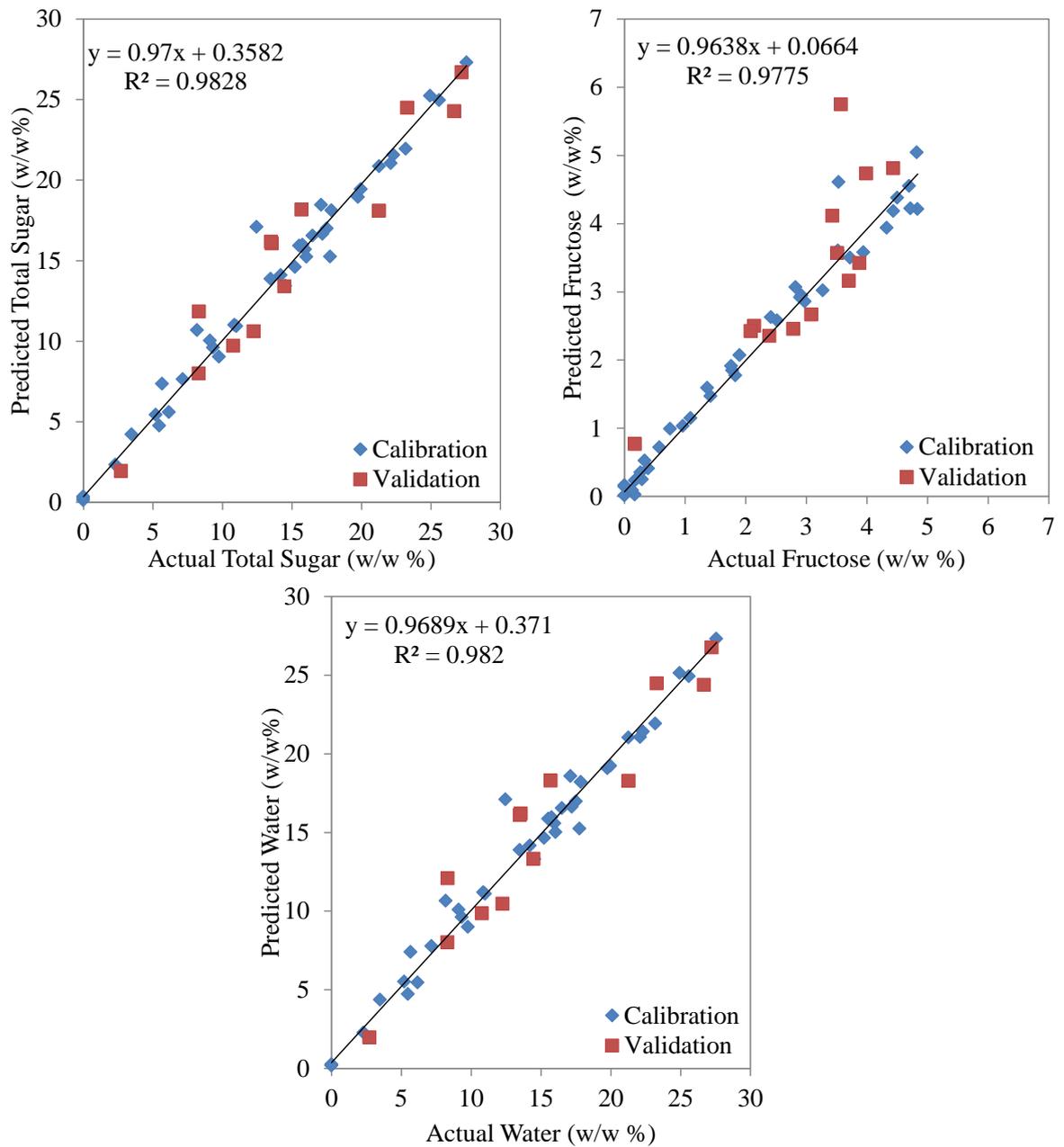


Figure 5.32. (cont.)

As shown in Figure 5.32, the R^2 values of all seven calibration models are found to be lower than 0.99, except the model of sucrose content. Furthermore, the prediction of validation samples appears to be highly deviated from the actual values. The calculated values of SECV, SEP and R^2 are given with the operating ranges each content in Table 5.13.

Table 5.13. Standard error of cross validation (SECV), standard error of prediction (SEP), maximum and minimum operating range of contents (Max and Min) and correlation coefficient (R^2) of GILS models belong to the fourth adulteration scenario.

	SECV (w/w %)	SEP (w/w %)	Max	Min	R^2
Honey	2.31	4.03	100	45	0.9806
Glycose	0.49	1.52	10	0	0.9805
Sucrose	0.68	1.36	20	0	0.9879
Total Glycose	0.85	1.59	20	0	0.9883
Total Sugar	1.09	2.06	30	0	0.9828
Fructose	0.26	0.75	10	0	0.9775
Water	1.11	2.09	30	0	0.9820

Although the SECV values of the components glycose, sucrose, total glycose and fructose appears to be below 1%, corresponding SEP values of the same components are much larger compared to the previous scenarios described above. For example, the SEP values of glycose and fructose are almost three times larger than their SECV values indicating that these models were not able to generate successful predictions for the independent validation set. These results explain the large spread for the validation samples given in Figure 5.32. As a result, all the GILS models are mostly over fitted. One possible reason for this overfitting could be the small number of samples in the calibration set. In addition, it is observed that honey has highest SECV and SEP values among all other components, it also has the widest dynamic range. Thus, the SECV and SEP values of honey should be compared to the others on the basis of the dynamic concentration ranges.

Similar to the other adulteration scenarios, PLS modelling was also performed with the same calibration and independent validation data sets. The resulting models were used to compare the performance of PLS and GILS calibration methods. Actual vs. PLS predicted concentrations plots of honey, glycose, sucrose, total glycose, total sugar, fructose and water contents are shown in Figure 5.33.

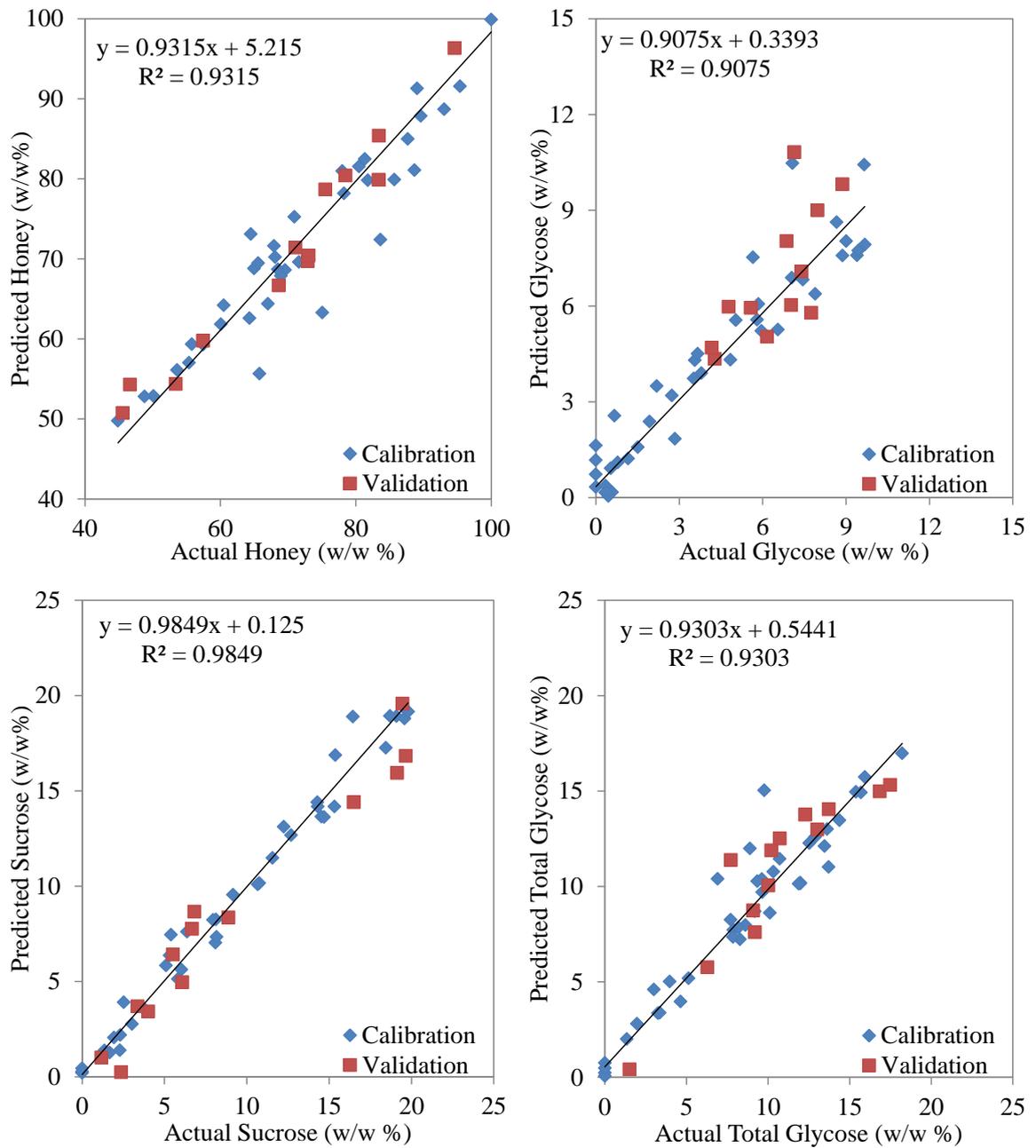


Figure 5.33. Actual versus predicted plots of honey, glucose, sucrose, total glucose, total sugar, fructose and water contents resulted from PLS in the fourth scenario.

(cont. on next page)

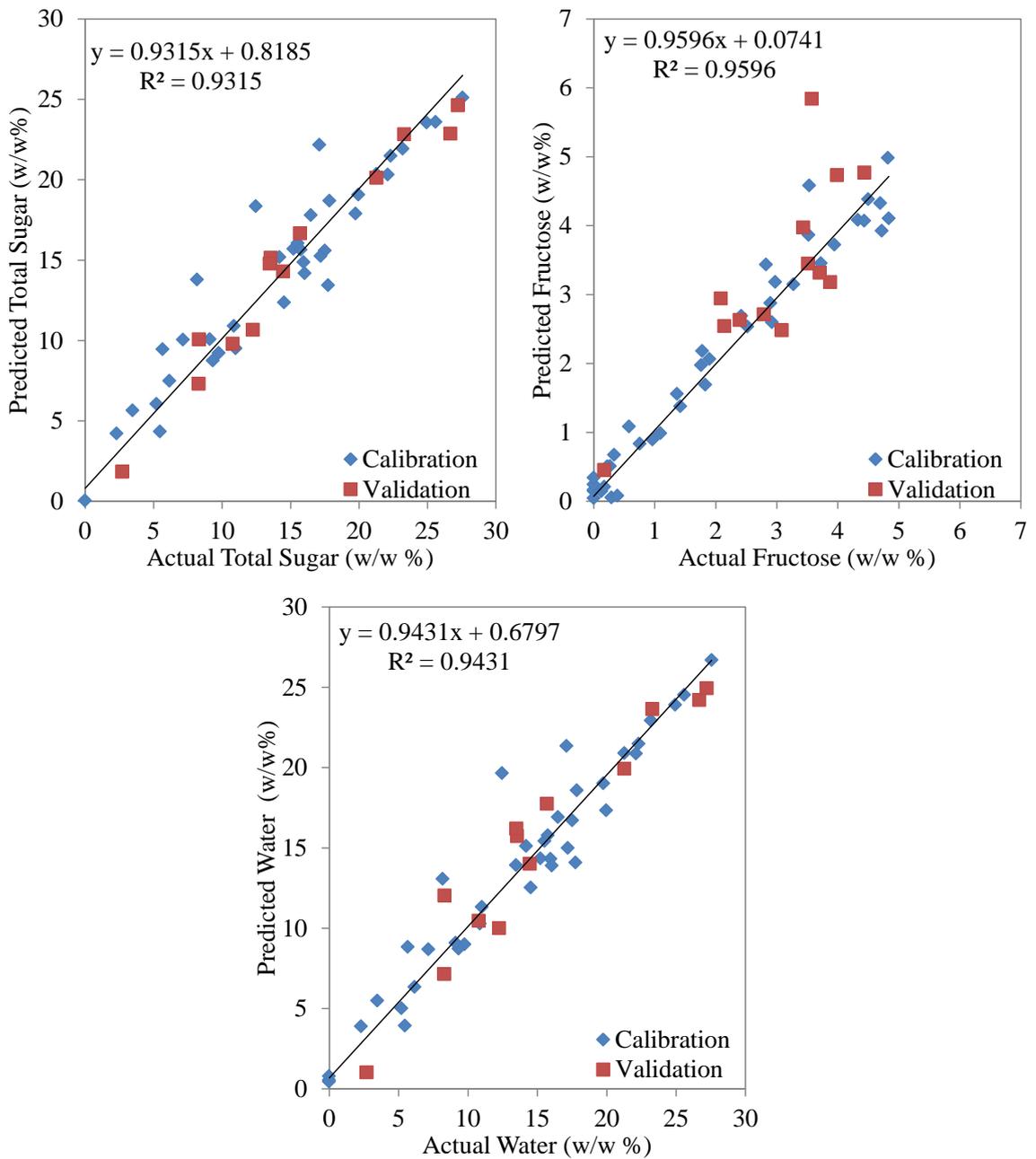


Figure 5.33. (cont.)

As given in Figure 5.33, the highest correlation coefficient among all calibration models belong to sucrose model and it is found to be 0.9846 followed by water and total sugar models whose R^2 values are 0.9431 and 0.9315, respectively. In order to compare all seven models in term of their SECV, SEP and R^2 values, these values are given in Table 5.14, along with the number of PCs.

Table 5.14. The number of principle components (PC), Standard error of cross validation (SECV), standard error of prediction (SEP), maximum and minimum operating range of contents (Max and Min) and correlation coefficient (R^2) of PLS models belong to the fourth adulteration scenario.

	Number of PC	SECV (w/w %)	SEP (w/w %)	Max	Min	R^2
Honey	3	4.32	3.34	100	45	0.9315
Glycose	5	1.06	1.40	10	0	0.9075
Sucrose	8	0.83	1.62	20	0	0.9849
Total Glycose	5	1.41	1.62	20	0	0.9303
Total Sugar	3	1.09	2.16	30	0	0.9315
Fructose	8	0.35	0.79	10	0	0.9596
Water	5	1.97	2.02	30	0	0.9431

When the results given in Table 5.14 compared with the results of GILS given in Table 5.13, the most obvious difference is that R^2 values resulting from GILS models being much higher than the ones for the PLS models. As seen, the correlation coefficients of GILS models are changing from 0.97 to 0.99, whereas these values are ranging from 0.90 to 0.98 for the PLS models. Correspondingly, this difference is also reflected to the SECV values of both methods. A close examination of SECV values obtained with GILS reveals that they are much smaller than the ones obtained from the PLS models. In addition, when the ratio of the SEP over SECV values (not shown here) of both methods were compared, it is clear that these ratio values for GILS models are mostly larger than two whereas in the case of PLS they are more like around one. Thus, it is expected that the PLS models should give better prediction values for the secondary test set. The same 100 authentic and commercial honey samples used in previous adulteration scenarios were introduced to the GILS and PLS models in order to obtain predicted honey and adulterant concentrations and the results are given in Figure 5.34.

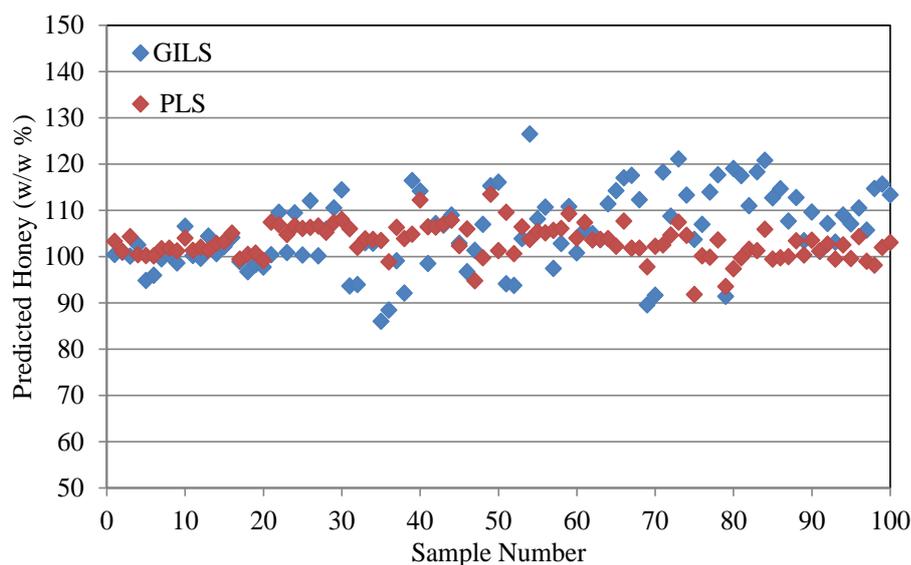


Figure 5.34. Honey content prediction of 100 authentic and commercial honey samples with both GILS and PLS models.

As given in Figure 5.34, when the PLS and GILS predicted honey concentrations were compared, it is evident that the GILS predictions are more scattered especially for the sample numbers from 65 to 90 where there are more than 15 prediction values which are above 110% (w/w %) honey content. Additionally, the prediction value for sample number 54 is around 126% (w/w %), whereas the same sample were predicted as 104% (w/w %). In contrast to the GILS results, the PLS model generated much better predictions and only two samples which are sample number 40 and 49 were predicted above 110% (w/w %) and the prediction of these samples with GILS were also above 110% (w/w %). The reason for the lack of success observed in GILS results would be possible overfitting as mentioned above. Furthermore the SECV and SEP values of the GILS model which are 2.31% and 4.03% (w/w %) also supports this claim where SECV and SEP value of PLS model are 4.32% and 3.34% (w/w %), respectively.

In addition, the same 100 sample were introduced GILS and PLS models for the other six components (glycose, sucrose, total glycose, total sugar, fructose and water) and their concentrations were predicted. While GILS predicted concentrations of glycose, sucrose and water are given in the upper graph, the concentration predictions of total glycose, total sugar and fructose are given in the lower graph in Figure 5.35 in order to clarify the display of the results. Similarly, PLS predicted concentrations of the same six components are shown as two different graphs in Figure 5.36.

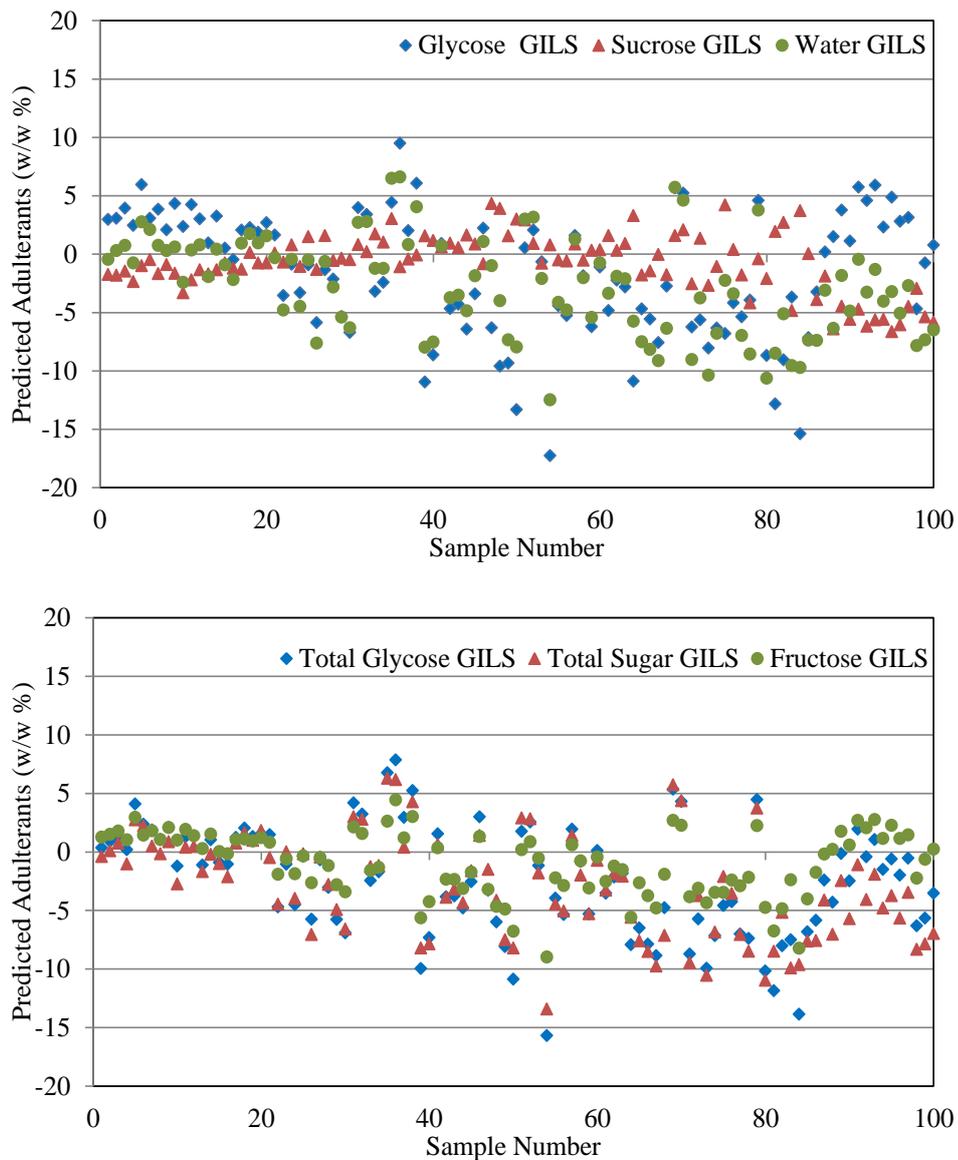


Figure 5.35. Predicted concentrations of glycose, sucrose, water, total glycose, total sugar and fructose content in 100 authentic and commercial honey samples with GILS method.

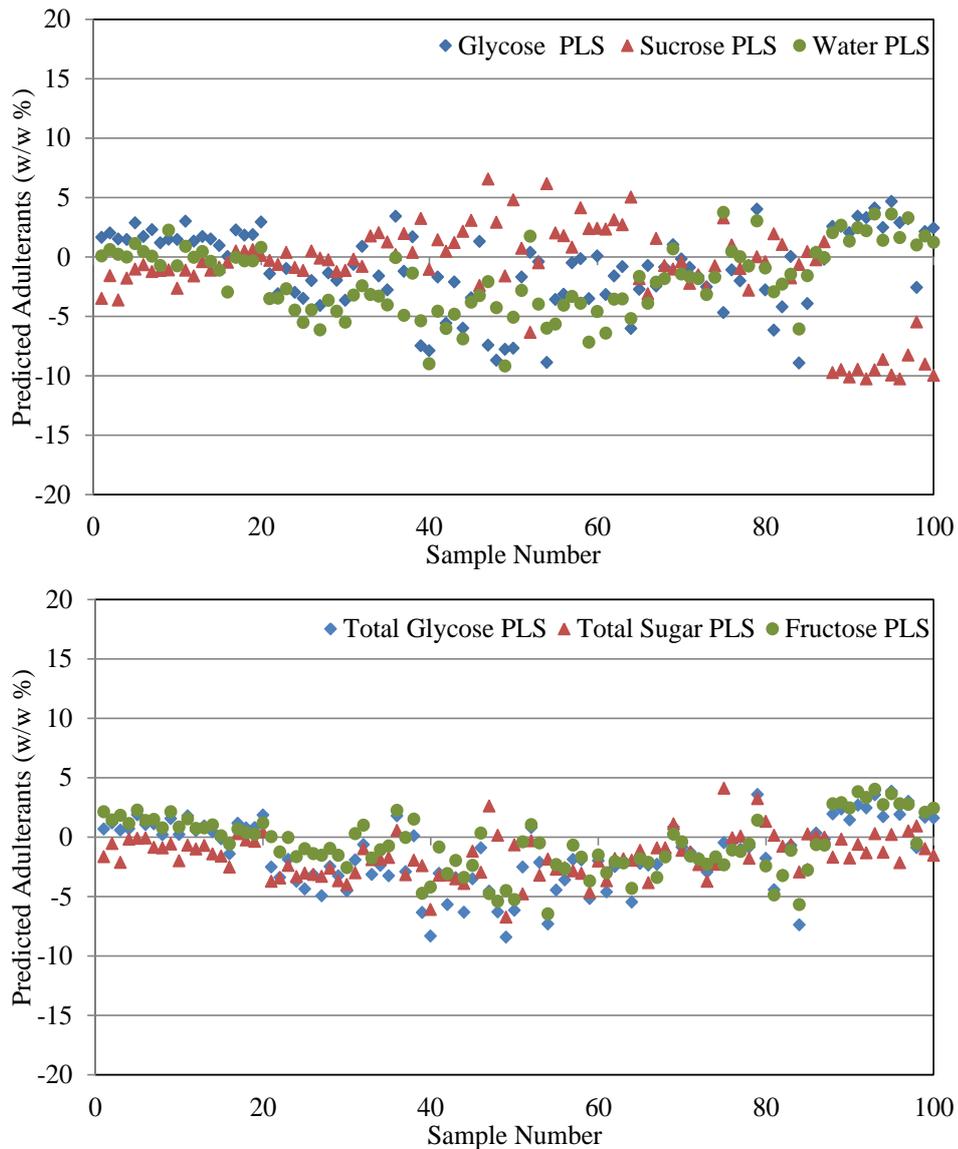


Figure 5.36. Predicted concentrations of glycose, sucrose, water, total glycose, total sugar and fructose content in 100 authentic and commercial honey samples with PLS method.

As illustrated in Figure 5.35 and 5.36, GILS predictions are found to be more scattered than PLS predictions, especially in glycose content. The reason of this scattering can be explained by SECV and SEP values of glycose model which is shown in Table 5.11 for GILS and in Table 5.12 for PLS. On the other hand, when compared with the predicted concentrations of total glycose, total sugar and fructose models, PLS predictions are also found to be in narrower interval than GILS predictions. In summary, the prediction performance of the multivariate calibration models for this fourth scenario are found to be less successful than the models developed in previous scenarios.

5.5. The Combined Adulteration Scenario

This adulteration scenario was constructed by gathering the adulterated samples that were prepared for the previous scenarios. The selected synthetically adulterated samples were composed of various types of pure honey as well as water and different artificial sweeteners which are corn syrup, beet sugar, glyucose, sucrose. In this combined scenario, there are a total of 120 (112 adulterated honey and 8 authentic honey) samples and the concentration profile of these samples are given in Table 5.15.

Table 5.15. Percent compositions of 120 pure, binary, ternary and quaternary adulterated samples prepared with pure honey, corn syrup, beet sugar, glucose, sucrose and water.

No	Honey (w/w %)	Corn Syrup (w/w %)	Beet Sugar (w/w %)	Water (w/w %)	Glucose (w/w %)	Sucrose (w/w %)	Total Sugar (w/w %)
1*8	100.00	0.00	0.00	0.00	0.00	0.00	0.00
9	64.35	35.65	0.00	0.00	0.00	0.00	35.65
10	80.07	19.93	0.00	0.00	0.00	0.00	19.93
11	72.91	27.09	0.00	0.00	0.00	0.00	27.09
12	82.71	17.29	0.00	0.00	0.00	0.00	17.29
13	74.81	25.19	0.00	0.00	0.00	0.00	25.19
14	60.46	39.54	0.00	0.00	0.00	0.00	39.54
15	69.45	30.55	0.00	0.00	0.00	0.00	30.55
16	94.36	5.64	0.00	0.00	0.00	0.00	5.64
17	86.33	13.67	0.00	0.00	0.00	0.00	13.67
18	79.85	20.15	0.00	0.00	0.00	0.00	20.15
19	67.70	32.30	0.00	0.00	0.00	0.00	32.30
20	86.45	13.55	0.00	0.00	0.00	0.00	13.55
21	69.50	30.50	0.00	0.00	0.00	0.00	30.50
22	58.80	9.17	16.01	16.01	0.00	0.00	25.19
23	71.19	24.31	2.25	2.25	0.00	0.00	26.56
24	64.75	16.23	9.51	9.51	0.00	0.00	25.74
25	71.19	24.36	2.22	2.22	0.00	0.00	26.59
26	48.70	23.28	14.01	14.01	0.00	0.00	37.29
27	61.88	12.65	12.73	12.73	0.00	0.00	25.39
28	59.70	10.11	15.09	15.09	0.00	0.00	25.20
29	53.59	19.90	13.25	13.25	0.00	0.00	33.15
30	54.72	23.02	11.13	11.13	0.00	0.00	34.15
31	85.04	4.83	5.07	5.07	0.00	0.00	9.89
32	55.32	18.34	13.17	13.17	0.00	0.00	31.51
33	89.32	7.07	1.80	1.80	0.00	0.00	8.87
34	57.04	5.58	18.69	18.69	0.00	0.00	24.27
35	68.62	14.89	8.25	8.25	0.00	0.00	23.14
36	64.27	7.34	14.19	14.19	0.00	0.00	21.54
37	79.42	5.36	7.61	7.61	0.00	0.00	12.97
38	85.75	3.30	5.48	5.48	0.00	0.00	8.78
39	69.29	4.35	13.18	13.18	0.00	0.00	17.53
40	77.44	9.92	6.32	6.32	0.00	0.00	16.24
41	83.97	6.80	4.62	4.62	0.00	0.00	11.42
42	94.56	0.00	2.72	2.72	0.00	0.00	2.72
43	89.36	0.00	5.32	5.32	0.00	0.00	5.32
44	79.60	0.00	10.20	10.20	0.00	0.00	10.20
45	74.58	0.00	12.71	12.71	0.00	0.00	12.71
46	99.08	0.00	0.46	0.46	0.00	0.00	0.46
47	47.36	0.00	26.32	26.32	0.00	0.00	26.32
48	97.56	0.00	1.22	1.22	0.00	0.00	1.22
49	77.72	0.00	11.14	11.14	0.00	0.00	11.14
50	30.43	0.00	34.79	34.79	0.00	0.00	34.79
51	26.64	0.00	36.68	36.68	0.00	0.00	36.68
52	77.84	0.00	11.08	11.08	0.00	0.00	11.08

(cont. on next page)

Table 5.15 (cont.)

No	Honey (w/w %)	Corn Syrup (w/w %)	Beet Sugar (w/w %)	Water (w/w %)	Glycose (w/w %)	Sucrose (w/w %)	Total Sugar (w/w %)
53	35.51	0.00	32.24	32.24	0.00	0.00	32.24
54	20.37	0.00	39.81	39.81	0.00	0.00	39.81
55	62.75	8.50	14.38	14.38	0.00	0.00	22.87
56	63.87	23.15	6.49	6.49	0.00	0.00	29.64
57	60.35	29.78	4.94	4.94	0.00	0.00	34.72
58	56.36	10.76	16.44	16.44	0.00	0.00	27.20
59	44.66	12.68	21.33	21.33	0.00	0.00	34.01
60	58.21	11.77	15.01	15.01	0.00	0.00	26.78
61	65.07	3.39	15.77	15.77	0.00	0.00	19.16
62	59.76	14.31	12.96	12.96	0.00	0.00	27.28
63	52.15	23.94	11.95	11.95	0.00	0.00	35.90
64	45.86	13.79	20.17	20.17	0.00	0.00	33.96
65	69.52	26.33	2.07	2.07	0.00	0.00	28.41
66	53.83	20.39	12.89	12.89	0.00	0.00	33.28
67	58.17	20.06	10.89	10.89	0.00	0.00	30.94
68	50.81	28.46	10.36	10.36	0.00	0.00	38.83
69	55.98	7.63	18.19	18.19	0.00	0.00	25.83
70	51.66	16.86	15.74	15.74	0.00	0.00	32.60
71	69.24	19.17	5.79	5.79	0.00	0.00	24.97
72	48.65	19.02	16.17	16.17	0.00	0.00	35.18
73	60.35	21.76	8.95	8.95	0.00	0.00	30.71
74	68.84	10.19	10.49	10.49	0.00	0.00	20.67
75	59.55	26.29	7.08	7.08	0.00	0.00	33.37
76	79.73	11.37	4.45	4.45	0.00	0.00	15.82
77	69.20	10.46	10.17	10.17	0.00	0.00	20.63
78	59.23	26.50	7.13	7.13	0.00	0.00	33.64
79	79.15	11.82	4.51	4.51	0.00	0.00	16.34
80	68.08	0.00	0.00	15.96	0.58	15.38	15.96
81	53.61	0.00	0.00	23.20	8.66	14.54	23.20
82	46.65	0.00	0.00	26.67	7.02	19.65	26.67
83	71.59	0.00	0.00	14.20	5.04	9.17	14.20
84	50.13	0.00	0.00	24.94	5.84	19.09	24.94
85	71.09	0.00	0.00	14.46	5.57	8.88	14.46
86	64.30	0.00	0.00	17.85	3.56	14.29	17.85
87	67.01	0.00	0.00	16.50	2.19	14.31	16.50
88	57.46	0.00	0.00	21.27	4.78	16.49	21.27
89	55.78	0.00	0.00	22.11	3.66	18.45	22.11
90	85.69	0.00	0.00	7.16	0.78	6.37	7.16
91	72.92	0.00	0.00	13.54	6.86	6.68	13.54
92	69.54	0.00	0.00	15.23	0.53	14.70	15.23
93	73.04	0.00	0.00	13.48	1.93	11.55	13.48
94	68.62	0.00	0.00	15.69	8.87	6.82	15.69
95	64.98	0.00	0.00	17.51	9.39	8.12	17.51
96	93.04	0.00	0.00	3.48	0.45	3.03	3.48
97	57.45	0.00	0.00	21.28	5.95	15.32	21.28

Table 5.15 (cont.)

No	Honey (w/w %)	Corn Syrup (w/w %)	Beet Sugar (w/w %)	Water (w/w %)	Glycose (w/w %)	Sucrose (w/w %)	Total Sugar (w/w %)
98	83.39	0.00	0.00	8.31	7.14	1.16	8.31
99	68.94	0.00	0.00	15.53	2.84	12.69	15.53
100	80.47	0.00	0.00	9.76	7.45	2.32	9.76
101	75.52	0.00	0.00	12.24	6.16	6.08	12.24
102	81.34	0.00	0.00	9.33	7.05	2.28	9.33
103	78.26	0.00	0.00	10.87	4.84	6.03	10.87
104	94.59	0.00	0.00	2.70	0.34	2.36	2.70
105	60.07	0.00	0.00	19.96	0.26	19.71	19.96
106	89.06	0.00	0.00	5.47	3.80	1.67	5.47
107	45.58	0.00	0.00	27.21	7.75	19.46	27.21
108	89.58	0.00	0.00	5.21	1.51	3.70	5.21
109	65.58	0.00	0.00	17.21	6.55	10.66	17.21
110	81.79	0.00	0.00	9.11	1.16	7.95	9.11
111	83.42	0.00	0.00	8.29	4.28	4.01	8.29
112	77.99	0.00	0.00	11.00	9.65	1.36	11.00
113	55.38	0.00	0.00	22.31	2.73	19.58	22.31
114	53.43	0.00	0.00	23.29	4.17	19.12	23.29
115	68.48	0.00	0.00	15.76	3.52	12.24	15.76
116	95.39	0.00	0.00	2.30	0.38	1.93	2.30
117	44.87	0.00	0.00	27.56	8.87	18.69	27.56
118	78.47	0.00	0.00	10.77	7.40	3.37	10.77
119	48.82	0.00	0.00	25.59	5.80	19.79	25.59
120	73.03	0.00	0.00	13.48	7.97	5.51	13.48

As seen from Table 5.15, there are 7 contents for the development of multivariate calibration models. Among them corn syrup, beet sugar, glycose, sucrose and water were used to adulterate pure honey samples. While both corn syrup and beet sugar were modelled separately in the third scenario where the samples contained honey, corn syrup beet sugar and water, it is not reasonable to do the same thing in this scenario, since the adulterated samples contain glycose not only from beet sugar, but also from the commercial analytical grade glycose. Similarly, there are two sources of fructose in which one from commercial analytical grade sucrose and the other one from corn syrup as it could be high fructose corn syrup. Therefore, it is better to model total sugar in adulterated samples. Thus, in this scenario only three models which are honey, total sugar and water were generated. While developing calibration models among the 120 samples, 8 pure and 72 adulterated honey samples were selected as calibration set and the rest of them (40 samples) were used as independent validation data set. Actual vs. GILS predicted concentrations of these contents were shown in Figure 5. 37.

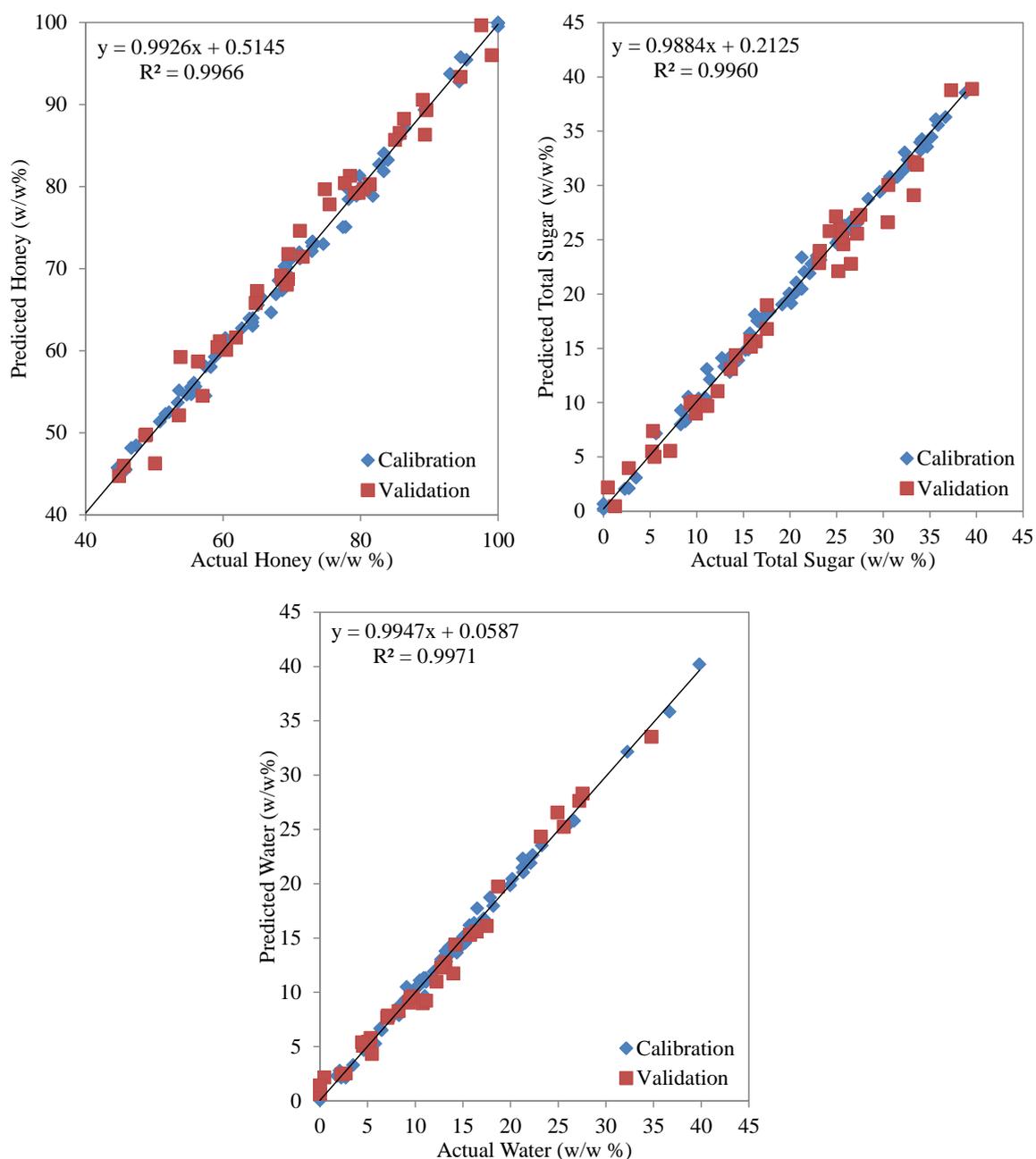


Figure 5.37. Actual versus predicted plots of honey, total sugar and water contents resulted from GILS in the combined adulteration scenario.

As seen from Figure 5.37, the correlation coefficients are greater than 0.99 for all the three models indicating that the developed models can be considered to be quite successful for the calibration set. In terms of independent validation set, the predicted values seem to be in good agreement as seen from these plots. In order to evaluate the performance of established models in detail, SECV, SEP and R^2 values are given with the operating ranges each of the contents in Table 5.16.

Table 5.16. Standard error of cross validation (SECV), standard error of prediction (SEP), maximum and minimum operating range of contents (Max and Min) and correlation coefficient (R^2) of GILS models belong to the combined adulteration scenario.

	SECV (w/w %)	SEP (w/w %)	Max	Min	R^2
Honey	1.03	2.05	100	45	0.9966
Total Sugar	0.71	1.57	40	0	0.9960
Water	0.48	1.01	40	0	0.9971

As given in Table 5.16, when the operating ranges of each content is taken into consideration, the highest SECV and SEP value belongs to the honey content as expected while SECV and SEP values of water content are almost one half of the values for honey content. Since the dynamic range of water is only from 0% to the 40% (w/w %), makes the SECV and SEP values of all the three models comparable. Moreover, PLS method was also performed on the same calibration set to obtain corresponding models for the three components and these models were used to predict the same independent validation set. The plots of actual vs. PLS predicted concentrations are given in Figure 5.38.

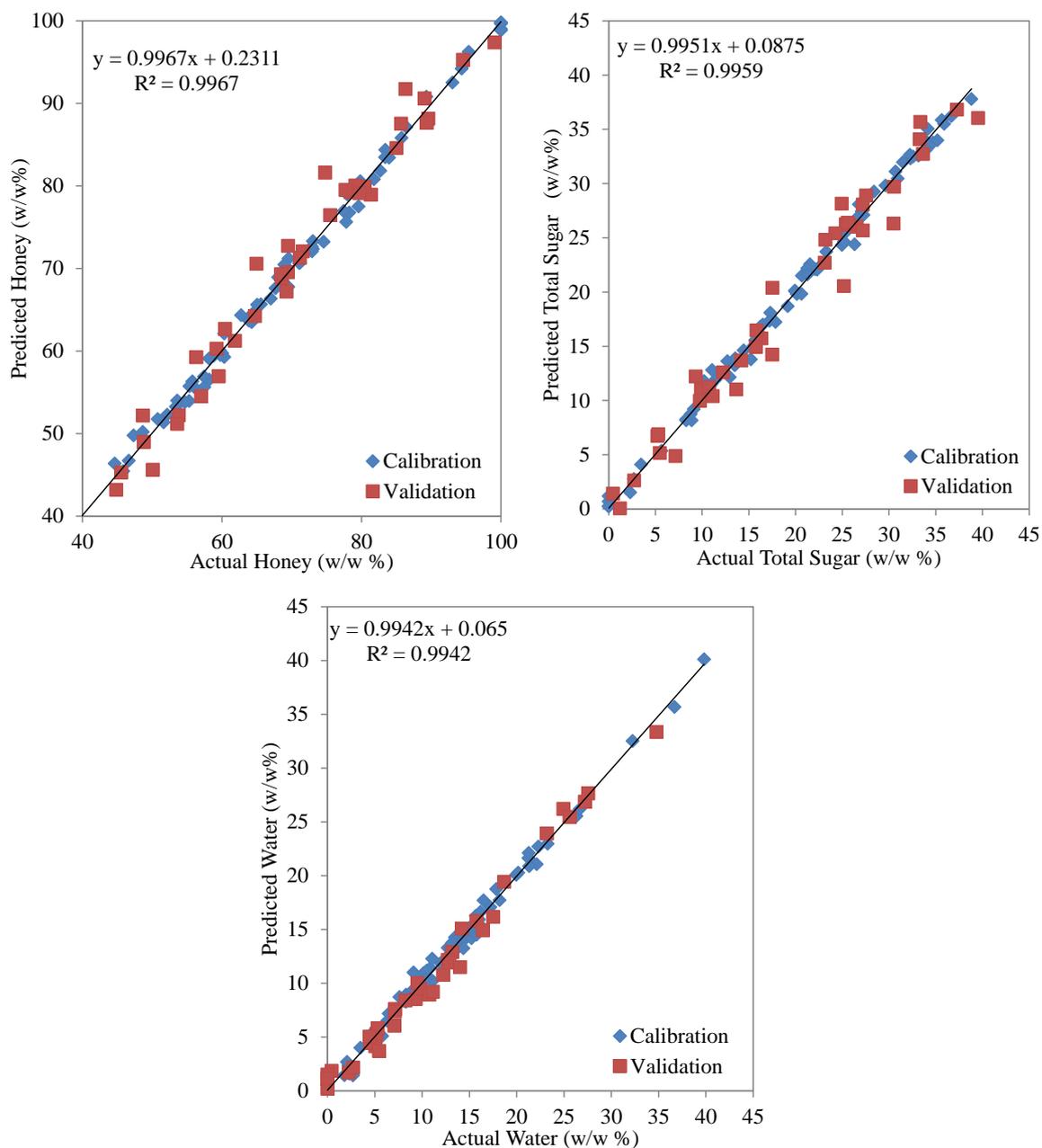


Figure 5.38. Actual versus predicted plots of honey, total sugar and water contents resulted from PLS in the combined adulteration scenario.

As seen in Figure 5.38, the correlation coefficients obtained from PLS method are close to the ones calculated for GILS models. Thus, in order to compare the models generated with both methods, the SECV, SEP and R^2 results of PLS models are given along with the number of principle components in Table 5.17.

Table 5.17. The number of principle components (PC), standard error of cross validation (SECV), standard error of prediction (SEP), maximum and minimum operating range of contents (Max and Min) and correlation coefficient (R^2) of PLS models belong to the combined adulteration scenario.

	Number of PC	SECV (w/w %)	SEP (w/w %)	Max	Min	R^2
Honey	12	1.02	2.39	100	45	0.9967
Total Sugar	13	0.71	1.84	40	0	0.9959
Water	9	0.69	1.17	40	0	0.9942

As can be seen in Table 5.17, in combined adulteration scenario, PLS models of honey, total sugar and water contents were established with 12, 13 and 9 number of principle components, respectively. Furthermore, when the SECV and SEP values are compared, honey content has the highest value than the others similar to the GILS models' SECV and SEP values. A total of 100 authentic and commercial honey samples, which were used in the previous scenarios as test set whose FTIR spectra had been shown in Figure 5.12, were introduced to the developed PLS models in order to predict their honey, total sugar and water contents. The predicted concentration results of honey contents by performing the developed GILS and PLS models are shown in Figure 5.39.

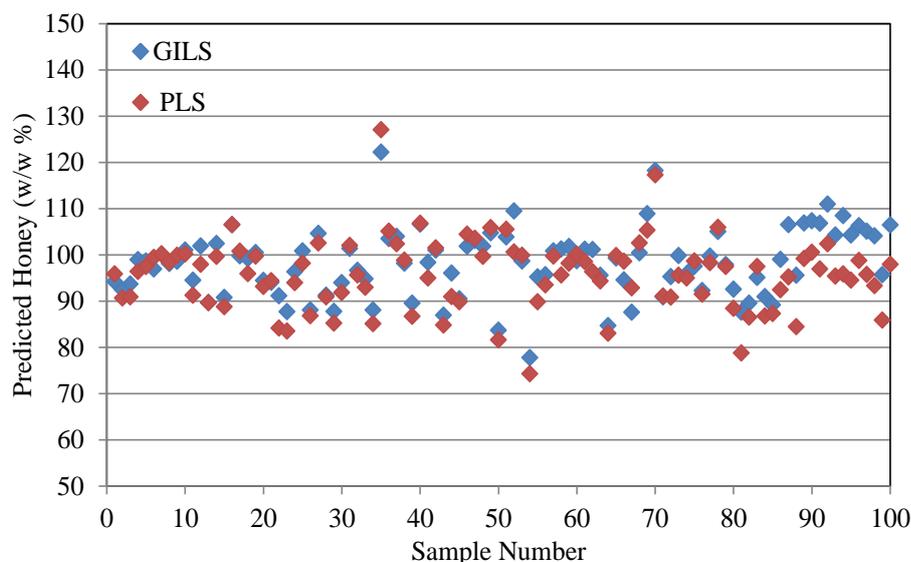


Figure 5.39. Honey content prediction of 100 authentic and commercial honey samples with both GILS and PLS methods using FTIR spectroscopy in the combined adulteration scenario.

As seen in Figure 5.39, among a total of 100 samples that were to be predicted, the honey content concentrations of about 80 of them were estimated in the same range, which is between 90% and 110% (w/w %) for both GILS and PLS. However, predicted concentrations of the remaining samples are more scattered and they are between 75% and 130% (w/w %). Moreover, the same 100 samples were also introduced to the developed PLS and GILS calibration models for determination of adulterants which are total sugar and water. The prediction results of the concentration of total sugar and water contents are illustrated in Figure 5.40.

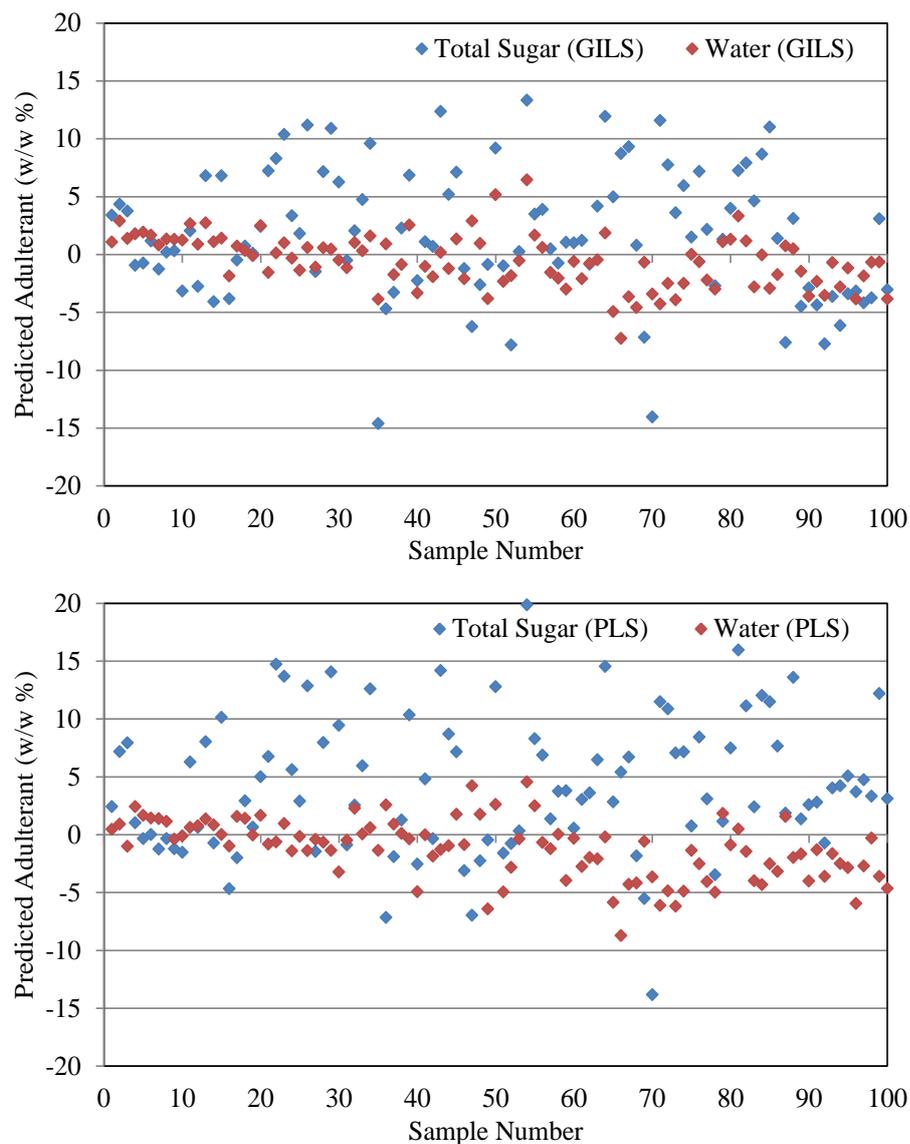


Figure 5.40. Total Sugar and water predictions of 100 authentic and commercial honey samples with both GILS and PLS models.

Figure 5.40 shows that, although predicted results obtained with GILS models appear to be slightly better than PLS model, the predicted concentration results of the total sugar content is spread out to the positive side of the $\pm 20\%$ prediction interval with both developed models. Moreover, despite having more accurate prediction results than total sugar content, water content has also scattered prediction results than the previous scenarios. Figure 5.40 shows that, the GILS predictions for total sugar concentrations appears to be better. The predictions, however, are spread in a range of $\pm 20\%$ with the more predictions being on the positive side for both models. Even though the water predictions are found to better than the predictions of total sugar concentrations, the scattering still exists with a range of $\pm 5\%$ (w/w).

CHAPTER 6

CONCLUSION

In this study, a new, fast and simple analytical method have been developed to determine adulterated honey samples prepared with different adulterants. For this purpose, molecular spectroscopic techniques, namely FTIR and FTNIR spectroscopies were used along with two different chemometric multivariate calibration methods which are GILS and PLS.

The four main and one combined honey adulteration scenarios were prepared with various artificial sweeteners and authentic honey samples belonged to different botanical and geographical origins. Among them, the results of the third adulteration scenario were found to be more successful and reliable than the others. In this adulteration scenario three adulterants namely corn syrup, beet sugar and water was used in different concentrations to prepare the adulterated honey samples. The success of this scenario was proved by the highest R^2 values that were found to be over 0.99 for GILS models and optimum SECV values between 0.97% - 2.52% (w/w%) and SEP values between 0.90% - 2.19% (w/w %) values obtained by GILS and PLS models that were developed for each of the four components. In addition, in order to test the prediction ability, 100 authentic and commercial honey samples were set to the each developed models as secondary test set as well as 6 adulterated samples, which include only binary mixtures of corn syrup-water and beet sugar water in different concentrations, were also predicted by the developed model as a minor test set. According to prediction results it is observed that both GILS and PLS methods can be used successfully to differentiate authentic and adulterated honey samples if a proper adulteration scenario has been designed.

On the other hand, when the results of remaining adulteration scenarios are interpreted, their prediction abilities are less successful than the mentioned third scenario. This result can be explained by the concentration ranges and the types of the adulterants that were used while preparing the adulterated samples. Moreover, the number of sample in these scenarios might be considered as insufficient.

REFERENCES

- Douglas A. Skoog, F. James Holler, Stanley R. Crouch-Principles of Instrumental Analysis sixth edition-Brooks Cole (2006)
- Bazar, G., R. Romvari, A. Szabo, T. Somogyi, V. Eles, and R. Tsenkova. 2016. "NIR detection of honey adulteration reveals differences in water spectral pattern." *Food Chem* 194:873-80. doi: 10.1016/j.foodchem.2015.08.092.
- Cabanero, ANA I, Jose L Recio, and Mercedes Ruperez. 2006. "Liquid chromatography coupled to isotope ratio mass spectrometry: a new perspective on honey adulteration detection." *Journal of agricultural and food chemistry* 54 (26):9719-9727.
- Carter, S. and Gomez, C. "Kinetics of the Inversion of Sucrose" *California State Science Fair*, April 2, 2015
- Cengiz, Mehmet Fatih, M. Zeki Durak, and Musa Ozturk. 2014. "In-house validation for the determination of honey adulteration with plant sugars (C4) by Isotope Ratio Mass Spectrometry (IR-MS)." *LWT - Food Science and Technology* 57 (1):9-15. doi: 10.1016/j.lwt.2013.12.032.
- Cinar, S. B., A. Eksi, and I. Coskun. 2014. "Carbon isotope ratio ($^{13}\text{C}/^{12}\text{C}$) of pine honey and detection of HFCS adulteration." *Food Chem* 157:10-3. doi: 10.1016/j.foodchem.2014.02.006.
- Du, B., L. Wu, X. Xue, L. Chen, Y. Li, J. Zhao, and W. Cao. 2015. "Rapid Screening of Multiclass Syrup Adulterants in Honey by Ultrahigh-Performance Liquid Chromatography/Quadrupole Time of Flight Mass Spectrometry." *J Agric Food Chem* 63 (29):6614-23. doi: 10.1021/acs.jafc.5b01410.
- Gallardo-Velázquez, Tzayhri, Guillermo Osorio-Revilla, Marlene Zuñiga-de Loa, and Yadira Rivera-Espinoza. 2009. "Application of FTIR-HATR spectroscopy and multivariate analysis to the quantification of adulterants in Mexican honeys." *Food Research International* 42 (3):313-318. doi: 10.1016/j.foodres.2008.11.010.
- Guler, A., H. Kocaokutgen, A. V. Garipoglu, H. Onder, D. Ekinici, and S. Biyik. 2014. "Detection of adulterated honey produced by honeybee (*Apis mellifera* L.) colonies fed with different levels of commercial industrial sugar (C(3) and C(4) plants) syrups by the carbon isotope ratio analysis." *Food Chem* 155:155-60. doi: 10.1016/j.foodchem.2014.01.033.
- Hostalkova, A., I. Klingelhofer, and G. E. Morlock. 2013. "Comparison of an HPTLC method with the Reflectoquant assay for rapid determination of 5-hydroxymethylfurfural in honey." *Anal Bioanal Chem* 405 (28):9207-18. doi: 10.1007/s00216-013-7339-6.

- Irudayaraj, J, R Xu, and J Tewari. 2003. "Rapid determination of invert cane sugar adulteration in honey using FTIR spectroscopy and multivariate analysis." *Journal of food science* 68 (6):2040-2045.
- Karoui, Romdhane, Gerard Downey, and Christophe Blecker. 2010. "Mid-infrared spectroscopy coupled with chemometrics: A tool for the analysis of intact food systems and the exploration of their molecular structure– Quality relationships– A review." *Chemical reviews* 110 (10):6144-6168.
- Kelly, J Daniel, Cristina Petisco, and Gerard Downey. 2006. "Application of Fourier transform midinfrared spectroscopy to the discrimination between Irish artisanal honey and such honey adulterated with various sugar syrups." *Journal of agricultural and food chemistry* 54 (17):6166-6171.
- Kelly, JF Daniel, Gerard Downey, and Vanessa Fouratier. 2004. "Initial study of honey adulteration by sugar solutions using midinfrared (MIR) spectroscopy and chemometrics." *Journal of agricultural and food chemistry* 52 (1):33-39.
- Li, Shuifang, Yang Shan, Xiangrong Zhu, Xin Zhang, and Guowei Ling. 2012. "Detection of honey adulteration by high fructose corn syrup and maltose syrup using Raman spectroscopy." *Journal of Food Composition and Analysis* 28 (1):69-74. doi: 10.1016/j.jfca.2012.07.006.
- Maréchal, Yves. 2011. "The molecular structure of liquid water delivered by absorption spectroscopy in the whole IR region completed with thermodynamics data." *Journal of Molecular Structure* 1004 (1):146-155.
- Mishra, S., U. Kamboj, H. Kaur, and P. Kapur. 2010. "Detection of jaggery syrup in honey using near-infrared spectroscopy." *Int J Food Sci Nutr* 61 (3):306-15. doi: 10.3109/09637480903476415.
- Özdemir, D., and Öztürk, B. 2004. "Genetic multivariate calibration methods for near infrared (NIR) spectroscopic determination of complex mixtures." *Turkish Journal of Chemistry* 28 (4):497-514.
- Padovan, G. 2003. "Detection of adulteration of commercial honey samples by the ¹³C/¹²C isotopic ratio." *Food Chemistry* 82 (4):633-636. doi: 10.1016/s0308-8146(02)00504-6.
- Ribeiro, Roberta de Oliveira Resende, Eliane Teixeira Mársico, Carla da Silva Carneiro, Maria Lúcia Guerra Monteiro, Carlos Conte Júnior, and Edgar Francisco Oliveira de Jesus. 2014. "Detection of honey adulteration of high fructose corn syrup by Low Field Nuclear Magnetic Resonance (LF 1H NMR)." *Journal of Food Engineering* 135:39-43. doi: 10.1016/j.jfoodeng.2014.03.009.
- Rios-Corripio, M. A., M. Rojas-López*, and R. Delgado-Macuil. 2012. "Analysis of adulteration in honey with standard sugar solutions and syrups using attenuated total reflectance-Fourier transform infrared spectroscopy and multivariate methods." *CyTA - Journal of Food* 10 (2):119-122. doi: 10.1080/19476337.2011.596576.

- Ruiz-Matute, A. I., S. Rodríguez-Sánchez, M. L. Sanz, and I. Martínez-Castro. 2010. "Detection of adulterations of honey with high fructose syrups from inulin by GC analysis." *Journal of Food Composition and Analysis* 23 (3):273-276. doi: 10.1016/j.jfca.2009.10.004.
- Ruiz-Matute, Ana I, Ana C Soria, Isabel Martínez-Castro, and ML Sanz. 2007. "A new methodology based on GC-MS to detect honey adulteration with commercial syrups." *Journal of agricultural and food chemistry* 55 (18):7264-7269.
- Simsek, Adnan, Mine Bilsel, and Ahmet C. Goren. 2012. "¹³C/¹²C pattern of honey from Turkey and determination of adulteration in commercially available honey samples using EA-IRMS." *Food Chemistry* 130 (4):1115-1121. doi: 10.1016/j.foodchem.2011.08.017.
- Sivakesava, S, and J Irudayaraj. 2001a. "Detection of inverted beet sugar adulteration of honey by FTIR spectroscopy." *Journal of the Science of Food and Agriculture* 81 (8):683-690.
- Sivakesava, S, and J Irudayaraj. 2001b. "Prediction of inverted cane sugar adulteration of honey by Fourier transform infrared spectroscopy." *JOURNAL OF FOOD SCIENCE-CHICAGO*- 66 (7):972-978.
- Sivakesava, S, and J Irudayaraj. 2001c. "A rapid spectroscopic technique for determining honey adulteration with corn syrup." *Journal of Food Science* 66 (6):787-791.
- Skoog, Douglas A, and Donald M West. 1980. *Principles of instrumental analysis*. Vol. 158: Saunders College Philadelphia.
- Tosun, M. 2013. "Detection of adulteration in honey samples added various sugar syrups with ¹³C/¹²C isotope ratio analysis method." *Food Chem* 138 (2-3):1629-32. doi: 10.1016/j.foodchem.2012.11.068.
- Xue, X., Q. Wang, Y. Li, L. Wu, L. Chen, J. Zhao, and F. Liu. 2013. "2-acetylfuran-3-glucopyranoside as a novel marker for the detection of honey adulterated with rice syrup." *J Agric Food Chem* 61 (31):7488-93. doi: 10.1021/jf401912u.