Non-destructive Measurement of Tomato Quality using Visible and Near-infrared Reflectance Spectroscopy

By

Limei Chen

Department of Bioresource Engineering
Macdonald Campus, McGill University
Sainte Anne-de-Bellevue, Quebec, Canada

August, 2008

A thesis submitted to McGill University in partial fulfilment of the requirements for the degree of Master of Science

ABSTRACT

Limei Chen M.Sc. (Bioresource Engineering)

Non-destructive Measurement of Tomato Quality using Visible and Near-infrared Reflectance Spectroscopy

Experiments were conducted to assess the feasibility of determining the quality attributes of tomato (*Lycopersicon esculentum* Mill cv 'DRK 453' and 'Trust') based upon visible/near-infrared reflectance (VIS/NIR) spectroscopy. A partial least squares regression (PLS) method was used to build prediction models.

Excellent prediction performance was achieved for lycopene content (LC), colour value a*/b* ratio, tomato colour index (TCI), and firmness. Coefficient of determination (R^2) for each of the parameters was respectively 0.96, 0.99, 0.99, and 0.97. All these R^2 were significant at 1% level. The root mean square errors of prediction (RMSEP) for all the parameters were low indicating the high quality of the fit of the prediction models. The values were 2.15, 0.06, 1.52, and 1.44 for LC, a*/b* ratio, TCI, and firmness, respectively. However, the models for prediction of titratable acidity, soluble solids content (SSC) and acid-Brix ratio showed relatively poor reliability, with R^2 value of 0.49, 0.03 and 0.65, and RMSEP of 0.43, 0.15 and 0.08, respectively.

Further, a model built by the PLS2 method showed good performance in simultaneously predicting a*/b* ratio, TCI, firmness, and LC of tomato, with R^2 values of 0.99, 0.99, 0.97, and 0.92, and RMSEP of 0.06, 1.75, 1.44, and 3.03, respectively. Once again here all the R^2 values were significant at 1% level.

RÉSUMÉ

Limei Chen M.Sc. (Génie des Bioressources)

Évaluation non destructive de la qualité de la tomate par spectroscopie de réflectance dans le visible et le proche infrarouge

Des essais visant à évaluer la faisabilité d'utiliser la spectroscopie de réflectance dans le visible et le proche infrarouge (VIS/PIR) pour déterminer certaines caractéristiques contribuant à la qualité de la tomate (*Lycopersicon esculentum* Mill. cv. 'DRK 453' et 'Trust') ont été menés. Une analyse de régression partielle par les moindres carrés a servi à bâtir des modèles de prédiction.

D'excellentes prédictions ont été obtenues pour la teneur en lycopène (TL), la valeur chromatique a*/b*, l'indice de couleur de la tomate (ICT), et la fermeté. Les coefficients de détermination (R^2) pour chacun de ces paramètres ont été de 0.96, 0.99, 0.99 et 0.97. Tous ces R^2 ont été significatifs à un niveau de 1%. L'erreur-type de prédiction (ETP) a été petite pour tous ces paramètres, indiquant un très bon degré d'ajustement des modèles. Des valeurs d'ETP de 2.15, 0.06, 1.52 et 1.44 ont respectivement été obtenues pour le TL, le rapport a*/b*, l'ICT, et la fermeté. Cependant, les modèles visant à prédire l'acidité totale, la teneur en solides solubles et le rapport acide-Brix se sont montrés peu fiables avec des valeurs respectives de R^2 de 0.49, 0.03 et 0.65 et de ETP de 0.43, 0.15 et 0.08.

De plus, un modèle multivariable bâti par une méthode de régression partielle par des moindres carrés (PLS2) s'est montrée très performant pour la prédiction simultanée du rapport a*/b*, de l'ICT, de la fermeté et de la TL avec des valeurs respectives de R^2 de 0.99, 0.99, 0.97 et 0.92 et de ETP de 0.06, 1.75, 1.44 et 3.03. Comme auparavant toutes les valeurs de R^2 ont été significatives à un niveau de 1%.

ACKNOWLEDGEMENTS

My profound thanks go to my esteemed supervisor Dr. G.S.V. Raghavan, James McGill Professor in the Department of Bioresource Engineering, McGill University, for his valuable guidance, consistent support and constructive suggestions throughout this project. I greatly appreciate his encouragement and help, without which it would not have been possible to fulfil my study.

I am highly grateful to my co-supervisor Dr. Clément Vigneault, Research Scientist from Horticulture R&D Centre of Agriculture and Agri-food Canada, Saint-Jean-sur-Richelieu, Quebec, Canada who introduced me to the challenges of this project, provided the facilities in his laboratory and gave me valuable guidance. I take this opportunity to express my great gratitude to him.

I am deeply indebted to Dr. Denis Charlebois and Dr. Marie-Thérèse Charles for sharing their knowledge and experience with me without reservation, allowing me to use their laboratory facilities and giving me consistent academic support. I sincerely thank Dr. Ning Wang for offering me the opportunity to pursue my study at McGill University, and Dr. Alain Clément for sharing his research experience.

I sincerely thank Dr. George Dodds for helping me with the grammar check of my thesis and the translation of the abstract in French. I would like to thank Dominique Roussel, Bernard Goyette, Jérôme Boutin, and Jianbo Lu for their help with my laboratory work. I would also like to thank my fellow students and friends Jagadeesh Lingegowdaru, Perumal Rajkumar, Zhenfeng Li, Nicolas Abdel-Nour, Simona Nemes, Satyanarayan Dev, Mick Wu, and Ruiming Chen for their help, suggestions and encouragement. I sincerely appreciate the help of Susan Gregus, Abida Subhan and Trish Singleton during my stay.

I take this opportunity to express my gratitude to the Natural Sciences and Engineering Research Council of Canada (NSERC) for the financial support of my study at McGill.

I have always appreciated my parents, Yinglu Chen and Xiaojuan Li, for their love and support.

CONTRIBUTION OF AUTHORS

The reported work presented here was performed by the candidate and supervised by Dr. G.S.V.Raghavan, James McGill Professor, Department of Bioresource Engineering, Macdonald Campus of McGill University. Dr. C.Vigneault, Research Scientist from Horticulture R&D Centre of Agriculture and Agri-food Canada, Saint-Jean-sur-Richelieu, Quebec, Canada, co-supervised my work. Dr. D.Charlebois and Dr. M.T.Charles, Research Scientists from Horticulture R&D Centre of Agriculture and Agri-food Canada, Saint-Jean-sur-Richelieu, Quebec, Canada, provided guidance and laboratory facilities during the study.

The authorships of the papers are as follows:

1. **Assessment of Lycopene, Acidity and Soluble Solids Content of Tomatoes through VIS/NIR Spectroscopy**

   L.Chen, D.Charlebois, M.T.Charles, G.S.V.Raghavan, C.Vigneault

2. **Modeling of Physical and Chemical Attributes of Tomatoes using VIS/NIR Spectroscopy**

   L.Chen, G.S.V.Raghavan, C.Vigneault, D.Charlebois, M.T.Charles

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

NOMENCLATURE

*a**	chromaticity coordinate (redness or greenness)
ABR	acid-Brix ratio
b*	chromaticity coordinate (blueness or yellowness)
ºBrix	degree Brix (unit of soluble solids content)
CIE	Commission Internationale de l'Éclairage
DOR	days of ripening
g	gram (unit of mass)
kg	kilogram (unit of mass)
L*	chromaticity coordinate (lightness)
LC	lycopene content
LV	latent variables
mg	milligram (unit of mass)
ml	milliliter (unit of volume)
MLR	multiple linear regression
mm	millimetre (unit of length)
MSC	multiple scatter correction
N	newton (unit of force)
NIR	near-infrared
nm	nano meter (unit of length)
PC	principal components
PCA	principal component analysis
PCR	principal component regression
PLS	partial least squares
r	correlation coefficient
R^2	coefficient of determination
RH	relative humidity
RMSEC	root mean square error of calibration
RMSECV	root mean square error of cross validation
RMSEP	root mean square error of prediction
RPM	round per minute
SDR	standard deviation ratio
SEP	standard error of prediction
SSC	soluble solids content
TA	titratable acidity
TCI	tomato colour index
Tg	teragram (unit of mass)
TSS	total soluble solids
VIS	visible

CHAPTER I

GENERAL INTRODUCTION

1.1 Introduction

Tomato (*Lycopersicon esculentum* Mill.) is one of the most widely produced fruits in the world. In 2007, world production of tomato was 126 Tg (FAOSTAT). The production quantities of all continents and countries with over 5 Tg are shown in Table 1.1. Tomato is considered high nutritional food because it is low in fat, calories and cholesterol-free, and rich in vitamins A and C. Additionally, tomato is also important sources of phytochemicals such as carotenoids, mainly lycopene and β-carotene. Many studies have shown that lycopene may have a protective effect against carcinogens in the liver, brain, colon, breast, cervix and prostate, therefore preventing or delaying certain types of cancer (Bramley, 2000). In addition, lycopene has a preventive effect against coronary heart disease (Manson et al., 1993).

With such intensive tomato production, it is important to develop efficient analytical methods for quality evaluation and sorting. Visible/near-infrared (VIS/NIR) reflectance spectroscopy has been established as a non-destructive analytical technique for determining chemical constituents and quality parameters in many agricultural produces and transformed products. It is gaining attention in the field of postharvest quality evaluation of fruits, owing to its many distinct advantages of non-destructiveness, quick measurement time, simplicity of sample preparation, chemical-free measurement, and simultaneous measurement of multiple attributes.

Although several studies have investigated the application of VIS/NIR spectroscopy to the evaluation of tomato fruit and produce quality, few of these have achieved good results in terms of predicting lycopene content and firmness.

1.2 Hypothesis

It is hypothesized here that tomato properties, including lycopene content, acidity, soluble solids content, colour, and firmness, can be quantified using VIS/NIR

spectroscopy and chemometric analysis. The NIR spectroscopic method is rapid, non-destructive and requires no hazardous chemicals. Hence, developing a new assay based upon VIS/NIR spectroscopy is very important in the fields of food analysis and postharvest technology.

1.3 Objectives

1. To study the feasibility of determining quality attributes of intact tomato fruits based upon VIS/NIR reflectance spectroscopy;
2. To establish calibration models for predicting physico-chemical properties of tomato, including lycopene content, titratable acidity, soluble solids content, colour, and firmness.

Table 1.1: Tomato production quantity of continents and countries (over 5 Tg) in 2007

Continent	Production Quantity (Tg)	Country	Production Quantity (Tg)
Asia	67.65	China	33.50
		Turkey	9.92
		India	8.59
		Iran	5.00
America	22.88	USA	11.50
Europe	20.50	Italy	6.03
Africa	14.51	Egypt	7.55
Oceania	0.56		
Total	126.10		

(Source: FAOSTAT)

CHAPTER II

LITERATURE REVIEW

Discovery of the near-infrared (NIR) spectrum was reported in 1800 by William Herschel during his measurements of the heat energy of solar emissions beyond the red portion of the visible spectrum (Davies, 2000). However, it is only since 1970 that NIR technology has progressed rapidly, owing to the development of modern instruments. VIS/NIR spectroscopy has recently become an increasingly important non-destructive analytical technique in food science. VIS/NIR spectroscopy has many advantages over chemical and other instrumental methods of food analysis. Its quick response time, the simplicity of sample preparation, chemical-free and non-invasive measurement, easy use in process control and grading systems, and simultaneous measurement of multiple attributes have made this technique expand into the field of food production and processing where frequent food quality evaluation is necessary. However, there are also some disadvantages to NIR spectroscopy which limit its application in food science. For instance, the price of NIR instruments is high, model building linking VIS/NIR spectra and quality attributes is complicated, such models are usually only used for a limited time, and little research has been made on model transfer. Although numerous studies have investigated the application of VIS/NIR spectroscopy within the food industry, much more exploration is needed to overcome limitations to its practical application.

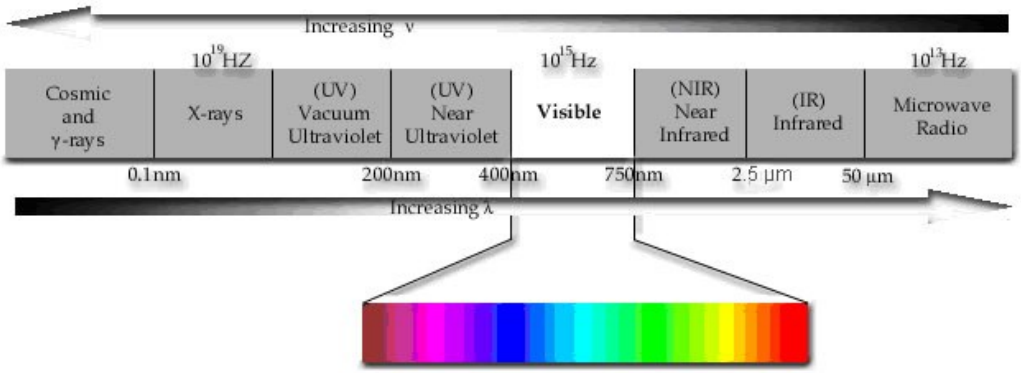
Grown worldwide, tomato *(Lycopersicon esculentum* Mill.) is a well-known, and highly nutritious fruit. In 2007, world production was 126 Tg (FAOSTAT), most of which was processed into products, such as ketchup, tomato sauce and tomato juice, but a significant volume was still sold fresh. Given the magnitude of tomato production, the use of NIR spectroscopy for measuring the quality of tomato has been a matter of serious consideration.

In light of the above, this chapter reviews VIS/NIR spectroscopy, chemometrics, application of NIR spectroscopy and quality attributes of tomatoes.

2.1 Principles of VIS/NIR spectroscopy

2.1.1 Chemical principles

The near-infrared spectrum is located between the infrared and the visible range covering the region of the electromagnetic spectrum from 750 to 2500 nm (Figure 2.1). The impact of NIR region on matter involves the response of O-H, N-H, C-H, and S-H molecular bonds. When organic molecules are irradiated with NIR frequencies, these bonds are subject to vibrational energy changes. The NIR absorption of polymers originates from the overtones and combination tones of these bonds' stretching vibrations and stretching-bending combinations. Thanks to modern instrumentation, it is possible to explain the NIR spectra by assignment of the positions of the bands to bonds involving hydrogen. The absorption intensity decreases as the overtone increases. Large numbers of overtones and combination bands result in broader and weaker NIR bands compared to mid-infrared bands. However, it may be difficult to assign NIR bands because of the overlapping of bands and complicated combinations of vibrational modes. VIS/NIR spectroscopy has a number of advantages when a precise and proper spectral analysis of VIS/NIR spectra is made. For example, it is quick, suitable for non-destructive analysis, and it allows multiple analyses relating to different properties from a single scan.



(Photo: http://www.baylor.edu/bucas/index.php?id=37025)

Figure 2.1: Light spectrum showing the VIS/NIR region

2.1.2 Physical principles

When radiation interacts with a sample, the incident radiation will be remitted, absorbed or transmitted, and the relative contribution of each phenomenon depends on the chemical constitution and physical parameters of the sample. The remission is considered to have two elements: (i) reflection (either specular or diffuse) induced by the surface of the sample, and (ii) scatter resulting from multiple refractions at phase changes inside the material. The cell wall interfaces which induce abrupt changes in refractive index are the main elements of scatter in fruit and vegetables (McGlone et al., 1997).

2.2 NIR Instrumentation

2.2.1 NIR Spectrometers

An NIR spectrometer consists of a light source (usually a tungsten halogen lamp), sample presentation accessory, monochromator, detector, and optical components, such as lenses, collimators, beam splitters, integrating spheres and optical fibers.

According to the type of monochromator, spectrometers are classified as diode instruments, such as emitting diode arrays (EDA), photodiode detector arrays (PDA) and laser diode spectrometers; filter instruments, including fix-filter, wedge-interference filter, tilting-filter, acousto-optical tunable filter (AOTF), and liquid crystal tunable filter (LCTF) spectrometers; scanning monochromator instruments in which gratings or prisms are used to separate the frequencies, producing spectra with equally spaced data across the full range from 750 to 2500 nm; Fourier transform NIR (FT-NIR) spectrometers using an interferometer to generate modulated light, where the time domain signal of the light reflected or transmitted by the sample onto the detector can be converted into a spectrum via a fast Fourier transform; and Hadamard-transform spectrometer. The most popular detectors are lead sulfide (PbS)-, indium gallium arsenide (InGAS)-, and silicon-based devices.

2.2.2 Trends

Because of their high acquisition speed and operation without moving parts, which enables them to be mounted on fruit grading lines, there is a clear trend towards PDA systems in which a fixed grating focuses the dispersed radiation onto an array of silicon (350-1100nm) or InGAS (1100-2500nm) photodiode detectors,. Compact portable and hand-held instruments continue to attract attention. In Japan, the Kubota and FANTEC companies have developed a series of portable instruments to measure the properties of many kinds of fruits, such as apple (*Malus domestica* Borkh.), orange (*Citrus sinensis* (L.) Osbeck), peach (*Prunus persica* (L.) Batsch), melon (*Cucumis melo* L.) and watermelon (*Citrullus lanatus* (Thunb.) Matsum. & Nakai) in the field. Saranwong et al. (2003a) used the FT20, a portable device from FANTEC company, to determine the sugar value of mangoes (*Mangifera indica* L.) and obtained a result of standard error of prediction (SEP)=0.40 and R=0.98 which was similar to the research-grade NIRS6500 (FOSS). A European consortium of research institutes created a portable glove-shaped apparatus which was equipped with various miniaturized sensors for measuring soluble solids content (SSC), internal colour and maturity (Hernandex Sanchez et al., 2003). Other applications of portable devices have been studied by Walsh et al. (2000), Temma et al. (2002a), Miller and Zude-Sasse (2004), and Zude et al. (2006). As there is a bright future for applications of portable devices, more research is required in this area.

2.2.3 Measurement setup

Three measurement setups are predominantly used to obtain NIR spectra: reflectance mode, transmittance mode and interactance mode. Since penetration of NIR radiation into product tissue decreases exponentially with depth (Lammertyn et al., 2000), it is important to choose the measurement configuration. Fraser et al. (2001) found a penetration depth of at least 25 mm in the 700-900nm range, but less than 1 mm in the 1400-1600 nm range. Later, they found that the skin was the main obstacle for light penetration when they evaluated the quality of mandarin (*Citrus reticulata*) fruit (Fraser et al. 2003). Compared to reflectance mode and interactance mode, transmittance mode carries information about the core of the product.

2.3 Chemometrics

Due to the large amount of information hidden in NIR spectral data and the fact that it is difficult to assign specific absorption bands to specific functional groups, multivariate statistical techniques (also called chemometrics) are required to extract the information. Chemometrics in NIR spectroscopy analysis includes three aspects including spectral preprocessing, calibration model building and model transfer.

2.3.1 Spectral preprocessing

The data acquired from a NIR spectrometer contain irrelevant information such as scattering effects, instrumental noises and so on, which have no bearing on sample-related parameters. Spectral preprocessing is used to remove this information so that regression techniques can properly handle the data and generate reliable and accurate calibration models. Several methods have been developed for this purpose, such as smoothing, multiplicative scatter correction (MSC) transformation, standard normal variate (SNV) transformation, and derivative (Naes et al., 2004). Several smoothing techniques have been proposed to remove random noise from NIR spectra, including moving average and Savitzky-Golay algorithm (Naes et al., 2004). MSC is used to compensate for additive (offset shifts) and/or multiplicative (amplification) effects in spectral data; moreover, a number of similar effects can be successfully treated with MSC, such as path length problems, interference, etc. Extended multiplicative scatter correction (EMSC) works in a similar way; in addition, it allows compensation for wavelength-dependent spectral effects, such as chemical interference effects (Martens and Stark, 1991). The results of SNV are similar to those of MSC. The practical difference is that SNV standardizes each spectrum using only the data from that spectrum. The choice between SNV and MSC is a matter of taste. The derivative process attempts to correct for baseline effects in spectra. The Savitzky-Golay method is the most frequently used algorithm to do derivation, and first- and second-order derivatives are most popular. In food analysis, many

calibration models of NIR spectra obtained from the derivative method gave good results (Lammertyn et al., 1998).

2.3.2 Calibration models

Recently, most calibration models have been built by multivariate regression techniques trying to establish a relationship between Y-variables and X-variables. In spectroscopy, the X- and Y-variables can be denoted as spectra and quality attributes of interest. A good regression model should extract all the relevant information from the spectra, and avoid over fitting. There are a variety of approaches to multivariate regression, in which multiple linear regression (MLR), principal component regression (PCR) and partial least squares regression (PLS) are the main used methods.

MLR is a method used for relating the variations in a Y-variable to the variations of several X-variables. MLR is applied when the number of X-variables is smaller than the number of samples, and the X-variables are linearly independent. However, due to the high collinearity of the NIR spectra, MLR models usually do not perform well (Naes et al., 2004). PCR is a two-step method, which first performs a principal component analysis (PCA) on the X-variables and then uses the principal components (PCs) as predictors in a MLR. PLS carries out both the X- and Y-matrices simultaneously to find the latent variables in X which will best predict the latent variables in Y. These latent variables are referred to as PLS-components (PCs). The PCs are ordered according to their significance for predicting the Y-variables, so the first PC is the most relevant, the second the next most relevant, and so on. PLS1 and PLS2 are two type methods of PLS, the difference being that PLS1 predicts one Y-variable at a time while PLS2 predicts several Y-variables simultaneously. However, if the Y-variables are independent, PLS1 may provide better prediction (Wold et al., 2001). PCR and PLS1 are the methods most frequently chosen when strong collinearity exists in X-variables, when there is noise in the data, or when there are numerous X-variables, which is the case for NIR spectral data. With respect to PLS1 and PCR, PLS1 usually gives results similar to those of PCR, but uses fewer PCs.

2.3.3 Models transfer

The reliability of property measurement by NIR spectroscopic method depends on the calibration model, however the model accuracy might dramatically decrease when using a different instrument. It is important to share model libraries and realize the model transfer to improve the practical application of NIR technology. Several transfer techniques (also known as instrumental standardization) have been reported and discussed. Sjoblom et al. (1998) reported that better results were obtained with orthogonal signal correction (OSC). While Greensill et al. (2001) compared a number of techniques for calibration model transfer between diode array systems, including piecewise direct standardization (PDS), OSC, finite impulse response technique, model updating and a wavelet transform-based standardization technique (WT). The best performance was obtained after transformation by WT and model updating. Although there are many methods for calibration model transfer, the accuracy of transferred models cannot reach the level of the original models.

2.4 Application of VIS/NIR spectroscopy in quality analysis of horticultural produce

Since the 1990s, a number of studies have measured the quality attributes of fruits and vegetables using VIS/NIR spectroscopy. However, most research has been carried out on fruits such as apple and mandarin, and only a few reports have focused on vegetables. With respect to the attributes of interest in produce, most papers focused on soluble solids content, and other characteristics such as firmness, pH, acidity, colour, dry matter, chemical content, and maturity.

2.4.1 Application to apples

Many studies (Moons et al., 1997; Cho et al., 1998; Lammertyn et al., 1998; Ventura et al., 1998; Lu et al., 2000; Peirs et al., 2000, 2001, 2002, 2003a, 2003b, 2005; McGlone et al., 2002a, 2003b, 2005; Temma et al., 2002a, 2002b; Clark et al., 2003; Park et al., 2003; McGlone and Martinsen, 2004; Walsh et al., 2004; Liu and Ying, 2005; Liu et al., 2006; Xing et al., 2006; Xing and De Baerdmaeker, 2007) have reported the determination of different attributes of apples.

For SSC, the reported root mean square error of prediction (RMSEP) were mostly in the range of 0.4-0.7% Brix. Liu and Ying (2005) generated models which gave a RMSEP of 0.0043 and 0.0678 for titratable acidity (TA) and pH respectively. As for firmness, Moons et al. (1997) and Lu et al. (2000) both reported poor accuracy, whereas Park et al. (2003) obtained relatively better results for firmness of apples (SEP=7.02, R^2=0.786). Peirs et al. (2000) found a method to predict the maturity of apples, with their best model having an r of 0.90 and SEP of 7.4 days. Xing and De Baerdmaeker (2007) obtained an accuracy of over 95% in detecting bruise spots on the surface of apples. Clark et al. (2003) detected brown hearts in "Braeburn" apples using NIR spectroscopy with R^2 values ranging from 0.69 to 0.91 and RMSEP from 7.9% to 15.4%.

2.4.2 Application to tomatoes and its products

With respect to the non-destructive analysis of quality of tomatoes and tomato products, some papers have been published (Slaughter et al., 1996; Hong and Tsou, 1998; Goula and Adamopoulos, 2003; Jha and Matsuoka, 2004; Khuriyati et al., 2004; Pedro and Ferreira, 2005, 2007; Baranska et al., 2006; Shao et al., 2007; Clement et al., 2008).

The model developed by Slaughter et al. (1996) predicted SSC of intact tomatoes by VIS/NIR interactance mode, with r=0.89 and SEP=0.33 ºBrix. Hong and Tsou (1998) reported calibration models, obtained by using second derivative preprocessing and the MLR method, to predict total soluble solids (TSS), TA and colour (a/b) of tomatoes with SEP values of 0.34, 0.06 and 0.09 and R^2 of 0.96, 0.94 and 0.98, respectively. Goula and Adamopoulos (2003) studied NIR spectroscopic methods to measure the percent moisture, sugars, acidity, proteins and salts of tomato juice; the best model, established using MLR, gave correlation coefficients for all parameter over 0.95. Jha and Matsuoka (2004) developed a calibration model to predict acid-Brix ratio (ABR) in tomato juice using PLS over wavelengths ranging from 1059.5 to 1124.8 nm. The model with the best performance had SEP and r value of 0.009 and 0.92, respectively. A good prediction of the dry matter of tomato was obtained (SEP=0.36 and R^2=0.96) by Khuriyati et al. (2004). Pedro and Ferreira (2005)

presented a nondestructive method for determining total solids, SSC, lycopene content and β-catotene content of tomato concentrate products. A splitting approach for spectra selection, MSC pretreatment and PLS were applied to achieve optimal prediction abilities, with RMSEP and r values for total solids 0.4157, 0.9998; SSC 0.6333, 0.9996; lycopene 21.5779, 0.9996; β-carotene 0.7296, 0.9981, respectively. In 2007 they studied the feasibility of calibrating different properties of tomato products by PLS2. They established a very good model to predict four properties — total solids, total sugars, glucose and fructose — with resulting SEP of 2.67, 18.69, 11.60, and 13.45, respectively. Baranska et al. (2006) compared FR-Raman, ATR-IR and NIR spectroscopy methods for measuring lycopene and β-carotene content in tomatoes and tomato products. The best prediction of quality was achieved using IR spectroscopy. Shao et al. (2007) evaluated the application of NIR in measuring the quality characteristics of "Heatwave" tomatoes, including SSC, pH and firmness [indicated by compression force (F_c) and puncture force (F_p)]. The RMSEP and r value obtained were 0.19% Brix and 0.90, 0.09 and 0.83, 16.017N and 0.81, and 1.18N and 0.83, for SSC, PH, F_c and F_p respectively. Clement et al. (2008) proposed an approach of factor analysis to evaluate the ripening and taste of tomatoes. A new variable, called "tomato maturity stage (TMS)", which is related to colour, lycopene content, firmness, TA, pH, and SSC was proposed. The regression model to predict TMS was obtained with an RMSEP of 0.259 and R^2 of 0.93. They also presented a model to discriminate varieties of pink tomato types and field-grown tomatoes from other varieties. However, they found it impossible to measure the gustatory index which is linked to electrical conductivity (EC), SSC, TA, and pH by VIS/NIR spectroscopy.

2.5 Tomato

Tomato (*Lycopersicon esculentum* Mill.) is a tropical fruit native to Central, South, and southern North America from Mexico to Peru. It is now grown worldwide with thousands of cultivars and it is the second most consumed culinary vegetable in the world after potato (Solanum tuberosum L.) (Gould, 1992). Tomatoes are

considered a highly nutritional food because they are low in fat, calories and cholesterol. Additionally, they are rich in vitamins A and C, lycopene, β-carotene (Mangels et al., 1993) and other antioxidants (Davies and Hobson, 1981).

The fruit is mainly composed of water, soluble solids (SS), insoluble solids, organic acids, micronutrients. Soluble solids are mainly sugars, including equal amounts of glucose and fructose with a small amount of sucrose, and minerals (mainly K, Ca, Mg and P). Insoluble solids are mainly constituted of fibers, like hemicelluloses, celluloses and pectins. Usually tomato presents 4.5-8.5% total solids which include soluble and insoluble solids, excluding seeds and skin (Gould, 1992). Titratable acids are composed primarily of organic acids, such as citric and malic acid. Micronutrients include vitamins, phytochemicals, etc.

2.5.1 Quality of tomatoes

Tomato maturity is usually assessed by fruit colour, firmness and flavour (Dorais et al., 2001; Batu 2004). Fruit colour is probably the most important attribute that determines overall quality. The maturity of tomatoes are traditionally classified in six stages based on the external colour change of the fruit from green to red, which are mature-green, breaker, turning, pink, light-red and red-ripe. Colour evolution during fruit ripening is mainly related to the breakdown of chlorophyll and synthesis of lycopene, which is responsible for the red colour and constitutes 75-83% of the total pigment content at full ripeness, whereas β-carotene occupies only 3-8.4% of total carotenoids (Gould, 1992; Abushita et al., 1997; Raffo et al., 2002). The colour is strongly dependent on cultivar and storage conditions (López et al., 2003). Gómez et al. (2001) also found that the a*/b* ratio was better than a* in distinguishing varieties.

Besides colour, firmness is one of the most important quality attributes to consumers (Tijskens & Evelo, 1994). Marketable fruits should have firmness values above 1.45 N mm^{-1} (Batu, 2004). The exact molecular changes that lead to fruit softening are still unknown. However, it is known that number of cell wall hydrolytic enzymes contribute to tissue softening and lessening of intercellular adhesion (Fisher and Bennett, 1991).

The SS and TA are important components of flavour. Fruits high in both acids and sugars have excellent flavour, while tart fruits have low sugar content and bland fruits have low acidity. The large variation in acid content has a much greater impact on tomato flavour than the limited variation in sugar content (Saltveit, 2005).

2.5.2 Importance of tomato phytochemicals in human health

Phytochemicals, also known as phytonutrients, are non-nutritive plant chemicals which are associated with the prevention of certain chronic diseases, including cardiovascular diseases, cancers, diabetes, osteoporosis and vision diseases. Phytochemicals contained in tomatoes are mainly carotenoids, of which lycopene predominantly contributes and also there is also a small amount of β-carotene.

Lycopene is a red carotenoid mainly found in tomato, watermelon and other red fruits, and it has been recognized as the most effective carotenoid. It has attracted attention due to its effect as a natural antioxidant. Many studies have shown that lycopene might have a protective effect against many types of cancer, such as liver cancer, breast cancer, cervical cancer and prostate cancer (Clinton, 1998; De Stefani et al., 2000). In addition, lycopene has a preventive effect against coronary heart disease (Manson et al., 1993). β-carotene has the highest provitamin A activity. In addition, hundreds of studies have shown that β-carotene may decrease risks of cancer and heart disease (Ziegler, 1991).

2.5.3 Non-destructive methods for measuring tomato quality

Many studies have proposed non-destructive methods for measuring internal quality of fruits and vegetables. Most technologies are focused on spectroscopy and spectroscopic imaging, including nuclear magnetic resonance (NMR), Raman spectroscopy, fluorescence spectroscopy, magnetic resonance imaging (MRI), laser-scattering imaging, etc.

Many studies have used NMR spectroscopy to determine of quality attributes of tomato fruit, such as firmness and ripeness (Chen et al., 1989; Stroshine et al., 1991; Kim et al., 1994). MRI, based on the principles of NMR, has been used to monitor fruit maturity and detect internal defects of tomatoes (Ishida et al., 1989; Pech et al., 1990; Saltveit, 1991). Baranski et al. (2005) illustrated Raman spectroscopy and

showed that Raman mapping could be used to measure the carotenoid distribution and content in tomatoes. Lai et al. (2007) showed the feasibility of using fluorescence spectroscopy and imaging to identify pigments involved in ripening, thereby detecting the stage of maturity and fruit damage. Tu et al. (2000) reported that a laser-scattering image system has the potential to evaluate the ripeness of tomatoes.

Besides spectroscopic methods, other technologies have also been studied. Gómez et al. (2006) evaluated the capacity of an electronic nose to monitor the change in volatile production associated with ripeness states for tomato and found that it was possible to differentiate and to classify the different tomato maturity states by this technology. Schotte et al. (1999) studied the acoustic impulse-response technique to evaluate firmness of tomatoes.

2.6 Conclusion

Application of NIR spectroscopy as a non-destructive analysis method has shows great promise in the field of food analysis which includes quality evaluation of fresh fruits and vegetables. The survey of literature reveals that most research on evaluating quality of horticultural produce by this technology has focused on measurement of SSC of fruits, especially apples. However, limited studies have examined the feasibility of using NIR spectroscopy to assess quality of tomato, one of the most widely produced and consumed fruits. Hence, it is of great interest to study the possibility of using this method to measure important characteristics of tomatoes, including SSC, TA, lycopene content, firmness and colour.

CHAPTER III

ASSESSMENT OF LYCOPENE, ACIDITY AND SOLUBLE SOLIDS CONTENT OF TOMATOES THROUGH VIS/NIR SPECTROSCOPY

3.1 Abstract

Non-destructive models based on visible/near-infrared (VIS/NIR) reflectance spectroscopic technique have been evaluated for predicting physiological properties of two varieties of tomatoes (*Lycopersicon esculentum* Mill cv. 'DRK 453' and 'Trust'), including lycopene content, soluble solids content (SSC), titratable acidity (TA) and acid-Brix ratio (ABR). Partial least squares (PLS) regression analysis was performed on the spectral data to build prediction models. Various spectral windows within the 400-2350 nm spectral range and pre-processing methods including multiple scatter correction (MSC) and Savitzky-Golay first derivative were assessed in optimizing the model for each parameter. Excellent prediction performance was achieved for lycopene content, which was R^2=0.96 and root mean square error of prediction (RMSEP)=2.15 mg kg^{-1}. The models of TA (RMSEP = 0.43 mg ml^{-1}, R^2 = 0.49), SSC (RMSEP = 0.15 ºBrix, R^2 = 0.03) and ABR (RMSEP = 0.08, R^2 = 0.65) gave relatively poor reliability.

3.2 Introduction

Tomato is one of the most widely produced and consumed fruits in the world. In 2007, world production of tomato was 126 Tg (FAOSTAT), most of which was processed into products such as ketchup, tomato sauce and tomato juice, though a significant volume was sold fresh. Tomatoes are favored by many people because they are low in fat, calories, cholesterol-free, and tasty. Additionally, the tomato is rich in vitamins A and C, lycopene, β-carotene (Mangels et al., 1993) and other antioxidants

(Davies and Hobson, 1981).

Maturity of tomatoes is usually assessed by their colour, firmness and flavor (Dorais et al., 2001; Batu, 2004). However, at present tomatoes are sorted mostly based on their external appearance, such as size, colour and surface defects (Abbott, 1999). Beyond these organoleptic parameters, flavour and nutritional attributes of tomatoes should also be included, such as soluble solids content, acidity and lycopene content for evaluation.

Several studies have found that content of sugar, acids and their interactions were highly related to flavour quality of tomatoes (Stevens et al., 1979; Hobson and Bedford, 1989). Simandle et al. (1966) reported that taste panel scores were correlated with SSC in tomatoes. According to Jones and Scott (1983), tomato flavour could be improved by increasing SSC and acidity. SSC is traditionally determined by refractometry, while acidity is usually measured by titration using phenolphthalein as indicator. These standard analytical methods are destructive and are not applicable to continuous systems.

Lycopene, the major carotenoid of ripe tomato fruit, has been found to be largely responsible for tomato's beneficial health effects (Gerster, 1997; Rao and Agarwal, 1999). Lycopene acts as a potent antioxidant to protect cells against oxidative damages and thereby decreasing the risk of chronic diseases (Rao and Agarwal, 1999). In addition, many studies have indicated that lycopene may have a protective and preventive effect against many types of cancer, such as prostate, breast, cervical and liver cancer (Gerster, 1997; Clinton, 1998; De Stefani et al., 2000) and coronary heart disease (Manson et al., 1993). Beyond its organoleptic role, lycopene could also be seen as a nutraceutical component of tomato. Therefore its concentration in tomatoes could be used in the evaluation and sorting processes. Conventional quantification methods for lycopene are usually laborious, require destructive sampling and employ hazardous organic solvents (Adsule and Dan 1979; Fish et al., 2002). Consequently it would be advantageous to develop a non-destructive and quick method to assess the lycopene content of tomatoes.

Since its introduction in the early 1970s, the use of near-infrared (NIR)

spectroscopy as a non-destructive analytical technique has progressed rapidly in a number of fields. Its rapid adoption owes much to its distinct advantages: quick response time, simplicity of sample preparation, chemical-free measurement, and simultaneous measurement of multiple attributes. Food chemistry has greatly benefited from these developments which allow the determination of a series of properties, such as soluble solid content (SSC), acidity and dry matter, in different food matrices.

SSC, an important characteristic of fruits and vegetables, has attracted the greatest attention from researchers. Good prediction results have been achieved for a wide range of produce using VIS/NIR spectroscopic technology, including apple (Lammertyn et al., 1998; Walsh et al., 2004; Nicolaï et al., 2006), mandarin (Kawano et al., 1993; McGlone et al., 2003a; Gómez et al., 2006), mango (Saranwong et al., 2001, 2003a, 2003b), peach (Slaughter, 1995; Walsh et al., 2004), melon (Long and Walsh, 2006), kiwifruit (*Actinidia deliciosa* C.F.Liang.& A.R.Ferguson) (McGlone et al., 2002b; McGlone and Kawano, 1998), etc. Many studies also reported the NIR technology on acidity prediction of different fruits (Moons et al., 1997; Schmilovitch et al., 2000; McGlone et al., 2002a, 2003a; Saranwong et al., 2003b; Liu and Ying, 2005). However, compared to SSC, it seems more difficult to predict acidity based on this method. It is probably because the concentration of acids in most fruits and vegetables is too low to affect the NIR spectrum significantly (Nicolaï et al., 2007).

Several studies have investigated the applications of NIR spectroscopy to the evaluation of tomato fruit and product quality. The first publication on tomato quality parameters is by Slaughter et al. (1996) in which they predicted SSC of tomato fruits. Hong and Tsou (1998) reported using an multiple linear regression (MLR) method to develop calibration models to predict total soluble solids (TSS), TA and colour value (a/b ratio) of tomato with a standard error of prediction (SEP) of 0.34, 0.06, 0.09 and R^2 of 0.96, 0.94, 0.98 respectively. Goula and Adamopoulos (2003) have determined moisture, sugars, TA, salts, and proteins in tomato juice. Jha and Matsuoka (2004) have calibrated the acid-Brix ratio of various tomato juices using PLS. A good result for calibration of tomato dry matter was obtained by Khuriyati et al. (2004). Pedro and

Ferreira (2005) developed PLS calibration models which performed very well in determining total solids, SSC, lycopene and β-carotene of tomato concentrate products. Later in 2007, they reported a good model to simultaneously predict four further properties: total solids, total sugars, glucose and fructose content. Shao et al. (2007) evaluated the application of VIS/NIR in measuring the quality characteristics of tomato "Heatwave", including SSC, pH and firmness and achieved satisfying results. Clément et al. (2008) proposed an approach of factor analysis to evaluate the ripening and taste of tomatoes. A regression model to predict "tomato maturity stage (TMS)", which is related to colour, lycopene content, firmness, TA, pH, and SSC, was obtained and had an RMSEP of 0.259 and R^2 of 0.93. Nevertheless, they found it is impossible to measure the gustatory index, which is linked to electrical conductivity (EC), SSC, TA, and pH, using VIS/NIR spectroscopy. However, most of these studies on fresh tomatoes were focused on measurement of SSC or acidity, whereas research about prediction of lycopene content by VIS/NIR spectroscopy requires investigation.

3.3 Objectives

In light of all the research performed, the present study on VIS/NIR spectroscopy was undertaken and had the following objectives.

1. To study the feasibility of determining the internal quality attributes of intact tomato fruits based upon VIS/NIR reflectance spectroscopy;
2. To establish calibration models for predicting physico-chemical properties of tomato, including lycopene content, TA, SSC, and acid-Brix ratio.

3.4 Materials and methods

3.4.1 Sample preparation

Tomato fruits (*L. esculentum* Mill.) of two varieties (cv. 'DRK 453' and cv. 'Trust') were harvested from a commercial greenhouse located in Saint-Damase, QC, Canada. The experiment was repeated three times. On each occasion 30 mature green tomatoes (15 'DRK 453' and 15 'Trust') were stored and allowed to ripen at 16ºC and

90-93% RH. Six tomato fruits (3 from each variety) were subjected to spectroscopic measurements and reference physico-chemical analyses at 1, 5, 8, 12 and 16 days of ripening (DOR), allowing most tomato maturity stages to be sampled. The tomatoes were washed with distilled water and dried thoroughly before spectroscopic measurements.

3.4.2 Acquisition of spectra

All spectral measurements were performed using a spectroradiometer (FieldSpec® Pro FSP 350-2500P/A110000, Analytical Spectral Devices Inc., Boulder, CO) coupled with a contact probe of 10mm spot size (Analytical Spectral Devices Inc., Boulder, CO). For each fruit, reflectance spectra (350-2500 nm) were taken at six (1 and 5 DOR) or four (8, 12 and 16 DOR) equidistant positions around the equator. For each reflectance spectrum, 10 scans were averaged at any given time/position, for a total of 60 or 40 scans per fruit per sampling. Reflectance, *R*, was calculated as the ratio of the visible and near-infrared energy reflected from the sample surface to a standard reference (Spectralon disk). The signals were preprocessed with ViewSpec Pro V2.14 (Analytical Spectral Devices Inc., Boulder, CO).

3.4.3 Reference analyses

After spectral measurements, each fruit was homogenized with a Waring blender for 1 min at maximum speed and the resultant slurry filtered through two layers of cotton cloth. The filtered tomato juice was used for chemical reference analyses.

3.4.3.1 Lycopene content

Lycopene content was determined according to the reduced volumes of organic solvents method of Fish et al. (2002). A 0.6 g (determined to the nearest 0.01 g) aliquot of tomato juice filtrate was placed in a 40 ml amber screw-top vial containing 5 ml of 0.05% (w/v) butylated hydroxytoluene (BHT) in acetone, 5 ml of 95% USP grade ethanol, and 10 ml of hexane. The sample was extracted on an orbital shaker at 180 RPM for 15 minutes on ice. After shaking, 3 ml of deionized water was added to the vial and the sample was shaken for an additional 5 minutes on ice. The vial was then left at room temperature for 5 minutes to allow for phase separation. The

absorbance of the upper (hexane) phase was measured by spectrophotometer (Biochrome, Ultrospec 3100 Pro, UK) at 503 nm, blanked against hexane solvent. The lycopene content was estimated from an average of two aliquots of the same homogenate using equation 3.1:

$$\text{Lycopene}\ (\text{mg kg}^{-1}) = \frac{(A_{503} * 31.2)}{\text{Weight of sample used (g)}} \qquad (3.1)$$

3.4.3.2 Titratable acidity

TA was determined using an automatic titrator (Titrino 719S, Metrohm, Switzerland) with 2 ml of tomato juice filtrate diluted in 30 ml of distilled water. The tomato juice extract was titrated by 0.1N sodium hydroxide (NaOH) to pH 8.1. Titratable acidity was expressed as g citric acid/mL tomato juice. The readings were averaged from duplicate measurements.

3.4.3.3 Soluble solids content

SSC was determined with a hand-held digital refractometer (AR200, Reichert Analytical Instruments, Depew, NY) operating at room temperature. A few drops of tomato juice extract were placed on a dry and clean refractometer prism and a reading was immediately taken. Total soluble solids were expressed as ºBrix.

3.4.4 Data analysis

Experimental tomato fruit were split into two sets: for the first and second batches of tomato fruits, 60 samples were used for calibration and a further 24 tomato fruits served as an external validation set. Data analysis was carried out using "The Unscrambler v9.7" (CAMO Inc., Woodbridge, NJ), a statistical software package for multivariate calibration.

All fruit reflectance measurements were transformed to absorbance (log (1/R)) values to obtain linear correlations of the NIR values with physico-chemical parameters measured by reference methods (e.g., lycopene content, TA, SSC, and acid-Brix ratio). Given that noise, that would affect the accuracy of calibration, was clearly present at the extremities of all reflectance spectra, only the range of 400 nm to 2350 nm was taken into consideration.

3.4.4.1 Wavelength range selection

To improve calibration results (McGlone et al., 2002a, 2002b, 2003a) several spectral windows of exclusively visible, exclusively NIR or combined VIS-NIR wavelengths were used to assess various attributes. The best wavelength ranges were selected based on the root mean square error of prediction (RMSEP) of models.

For lycopene content, five windows (A: 400-750nm, B: 400-2350nm, C: 400-1300nm, D: 450-1000nm, E: 750-1300nm) were chosen. For TA, five windows (A: 400-750nm, B: 400-2350nm, C: 400-1300nm, D: 750-2350nm, E: 1200-1800nm) were chosen. For SSC, four windows (A: 400-750nm, B: 400-2350nm, C: 400-1000nm, E: 750-1500nm) were chosen. For acid/Brix, four windows (A: 400-750nm, B: 400-2350nm, C: 700-1000nm, D: 750-2350nm) were chosen. These ranges were chosen on the basis of prior research on various fruits and Martens' automatic uncertainty test which shows significant X variables (Westad and Marten, 2000).

3.4.4.2 Spectra pre-processing

The data obtained from the VIS/NIR spectroradiometer contained irrelevant information such as scattering effects and instrumental noise, which would have no bearing in assessing sample characteristics. Pre-processing was used to diminish such irrelevant information and to obtain reliable and accurate calibration models. To test the influence of this pre-processing on the quality of calibration models, two types of methods, multiple scatter correction (MSC) and Savitzky-Golay first derivative, were used on the best wavelength range selected for each parameter.

MSC was used to compensate for additive (offset shifts) and multiplicative (amplification) effects in spectral data and to deal with path length problems (Martens and Naes, 1989). Due to the fresh fruit's scattering of light, light does not always travel the same distance in the sample before it is detected. A longer light-travel path results in a lower relative reflectance value, since more light is absorbed. This kind of variation is eliminated by MSC. First order derivative was applied using the Savitzky-Golay algorithm to correct for baseline effects in spectra.

3.4.4.3 Multivariate calibration

A PLS regression method was used to build the calibration models. In general, principal component regression (PCR) and PLS are the methods most often chosen for NIR spectral analysis. PLS usually gives similar results as PCR, but uses fewer PCs (De Jong, 1993). PLS is a two-step method. Firstly the original independent information (X-variables) is projected onto a small number of latent variables (LV) to simplify the relationship between X-variables and Y-variables as the relationship is concentrated on the smallest number of LVs (Naes, et al., 2004).

$$X = TP + E \quad (3.2)$$

$$Y = UQ + F \quad (3.3)$$

where T and U are the scores of the X and Y matrices, P and Q are the loadings of the X and Y matrices, and E and F are the residuals for matrices X and Y. In the second step T and U are processed by linear regression. Thus PLS not only estimates component concentrations, but also assesses chemical and physical properties from VIS/NIR spectra (Lammertyn et al., 1998; Gómez et al., 2006).

Full cross validation was performed on the calibration samples to determine the optimal number of PCs and also to validate the models. With this method, one sample is left out from the calibration data set and the model is calibrated with the remaining data points. Then the value for the left-out sample is predicted and the prediction residual is computed. The process is repeated until every observation has been left out of the calibration set once; then all prediction residuals are combined to calculate the root mean square error of cross validation (RMSECV).

3.4.4.4 Selection of the best models

Choosing the best models should be based on a number of parameters, including RMSECV, root mean square error of calibration (RMSEC), the difference between RMSECV and RMSEC, the correlation coefficient (r) of validation, and the number of PCs. A good model should have a low RMSEC, a low RMSECV or RMSEP, a high correlation coefficient but a small difference between RMSEC and RMSEP. A large difference between RMSEC and RMSEP would indicate that too

many PCs are used in the model and noise is being modeled (Lammertyn et al., 1998). Moreover, a relatively low number of PCs is generally desirable.

RMSEC and RMSECV or RMSEP are defined using equation 3.4 and equation 3.5:

$$\text{RMSEC} = \sqrt{\frac{1}{I_c}\sum_{i=1}^{I_c}(\hat{y}_i - y_i)^2} \tag{3.4}$$

$$\text{RMSECV or RMSEP} = \sqrt{\frac{1}{I_p}\sum_{i=1}^{I_p}(\hat{y}_i - y_i)^2} \tag{3.5}$$

where,

$\hat{y}_i$ = predicted value of the i^{th} observation,

y_i = measured value of the i^{th} observation,

I_c = number of observations in the calibration set,

I_p = number of observations in the validation set,

The standard deviation ratio (SDR), defined as the deviation ratio of the response variable to the RMSEP (McGlone and Kawano 1998), was used to evaluate the quality of models. An SDR value between 2.5 and 3 or above corresponds to good and excellent prediction accuracies.

3.4.4.5 External validation

By means of full cross validation, the validation samples originate from the same sample set as calibration, so that spectral variability is included. It is necessary to validate models with an independent set of samples. After the optimal calibration model for each parameter was selected, it was applied to the external validation set. The prediction results, such as R^2, RMSEP and SDR, were calculated to assess the accuracy of the model.

3.5 Results and Discussion

3.5.1 Quality characteristics of tomato

The summary of the statistical results of all quality attributes of the calibration and validation sample sets are listed in Table 3.1 and the changes in these quality

attributes for calibration samples of both varieties are illustrated in Figure 3.1. The mean (± standard deviation) lycopene content, SSC, TA and acid-Brix ratio (ABR) of the validation set, measured by reference methods were 14.08 mg kg^{-1} (±11.14 mg kg^{-1}), 4.70 ºBrix (±0.16 ºBrix), 3.81 mg ml^{-1} (±0.56 mg ml^{-1}), and 0.81 (±0.12), respectively. As expected, lycopene content increased during ripening, while TA and ABR tended to decrease slightly. However, changes in SSC presented a complex pattern. SSC increased until 8 days of ripening (DOR), at which tomatoes were at about the pink stage, decreased at 12 DOR, then increased slightly again for variety 'DRK 453' but remained constant for variety 'Trust'.

3.5.2 VIS/NIR spectra

Figure 3.2 shows the mean raw reflectance spectrum of one tomato at wavelengths ranging between 350 and 2500 nm. The absorbance spectra (400-2350 nm) of tomatoes of the two different varieties tested at different stages of maturity (Figure 3.3) show the spectra to be quite similar, with some characteristic peaks and valleys revealing the fruit characteristics.

For variety 'DRK 453' (Figure 3.3a), it is clear that the variation between the mean spectrum of mature green tomatoes (1 DOR) and that of other maturity stages is significant across the entire wavelength range. Variations among the spectral curves at the four maturity stages after mature green are only apparent between 400 to 700 nm. A continuous absorbance decrease in the visible region (from 400 to 750 nm) with two small peaks at 560 nm, except for the spectrum of tomato at 1 DOR, and 675 nm can be observed, followed by a fluctuating increase in the whole NIR region until 2350 nm, with a maximum at about 1930 nm. The absorption curve for tomato at 1 DOR has an obvious peak in the visible range around 675 nm. High absorbance in this region is indicative of red absorbing pigments, particularly chlorophyll which gives the fruit its green colour. After the 675 nm peak, the spectrum remains relatively flat until 900nm. There is a prominent peak at 980 nm which is due to absorption by water and carbohydrates. This corresponds to the strong 960-990 nm absorption band of water, which dominates fruit components (Williams and Norris, 2001). Another two peaks appear at 1197 nm and 1448 nm, related to strong water absorption bands from

1125 to 1270 nm and from 1440 to 1485 nm. The absorbance curve pattern for tomatoes is very similar to that of other fruits, such as apple, cherry (*Prunus cerasus* L.*)*, mandarin, and orange (Lu et al., 2000; Lu, 2001; Gómez et al., 2006; Cayuela, 2008). Tomatoes at 1 DOR, had the lowest lycopene content but the highest levels of TA, SSC and acid/Brix value (Figure 3.1), had lower absorbance in most wavelength ranges except for 400-505 nm, 590-720 nm, and 1830-2350nm Comparatively, tomatoes at 16 DOR had the highest lycopene content,lowest TA, and greater absorbance in the 400-615 nm and 710-1380 nm ranges.

The spectra of calibration samples of the variety "Trust" (Figure 3.3b) showed almost the same trends and characteristics as "DRK 453", but differences between mean spectra for mature green tomatoes *vs.* those at other stages were not as distinctly different in "Trust" as in "DRK 453".

Calibration samples' raw absorbance spectra (400-2350nm; Figure 3.4a) and the same spectra preprocessed by MSC (Figure 3.4b) or S. Golay first derivative (Figure 3.4c) show that consistent offset shifts and baseline shifts initially exist in the spectra, due to light scattering. It is apparent that the peak and valley positions of the preprocessed spectra correspond to those of the original spectra; however, it can be seen that the MSC method eliminates the baseline shifts and makes the peaks and valleys clearer, while the Savistzky-Golay first derivative method differentiates the overlapped peaks in the spectra.

3.5.3 Prediction of quality parameters

3.5.3.1 Lycopene content

Performance of PLS models for lycopene content prediction in different spectral ranges was evaluated using the RMSECV statistic (Figure 3.5). Compared to other spectral windows, spectral window D (450-1000 nm) offered the best results in predicting lycopene content (RMSECV=2.57). Any ranges including visible wavelengths were distinctly superior to those restricted to NIR wavelengths (window E). This is attributable to the fact that lycopene is the principal red pigment of tomato fruit and absorbs light between 600-750 nm. Additionally, in order to test if 450-1000nm was the optimal spectral range, a further study was made by altering the

lower and upper wavelength limit on window D at intervals of 50 nm (Figure 3.6). The lowest RMSECV values along each curve were achieved at wavelengths of 450nm and 1000 nm respectively. Predictive performance sharply deteriorated as the lower limit of the window extended beyond 650 nm, thus excluding the chlorophyll absorbance band (Figure 3.6a). If wavelengths from 900-1000 nm were omitted, the model quality deteriorated (Figure 3.6b).

Two preprocessing methods, MSC and Savitzky-Golay first derivative, were applied on the optimal spectral range to find the best calibration model for lycopene content prediction. Since model statistics and a graphical comparison of all models (Table 3.2 and Figure 3.7a), showed model 2 to have the lowest RMSECV (2.57), with 8 PC and no data preprocessing, this model was chosen as the optimal calibration model (R^2=0.94, and SDR=4.14). When the model was used to predict a further 24 external samples, the prediction results were also excellent (R^2 = 0.96, RMSEP = 2.15, SDR=5.18; Table 3.3, Figure 3.8a). This result was superior to that obtained by Baranska et al. (2006) using NIR spectroscopy to quantify lycopene in tomato fruits and related products (R^2=0.85 and SECV=91.19).

3.5.3.2 Titratable acidity

A comparison of RMSECVs for PLS regression models used in TA prediction with different spectral ranges (Figure 3.9), showed the lowest value was obtained for NIR window E (1200-1800 nm). Consequently, this band was chosen as the best range for TA prediction. There were no further improvements, but rather deterioration when MSC and first derivative methods were used on the spectral data (Table 3.2 and Figure 3.7b). Therefore, the best result of multivariate regression (model 6) obtained between NIR measurements and the TA of tomatoes was R^2=0.69, RMSECV=0.31 and SDR=1.93 with 8 PCs. When the PLS model 6 was applied to predict 24 prediction samples, the prediction results (Table 3.3 and Figure 3.8b), showed an R^2 of 0.49, RMSEP of 0.43 and SDR of 1.30. Apparently this technique was not adequate in predicting TA of unseen tomato fruits.

The one possible reason leading to the difficulty for acidity prediction of fruit is that the covalent bond between carbon and oxygen in the acid functional group

–COOH has very low absorptivity when compared to the bond of C-H or O-H. Another reason for the poor prediction could be the low concentration of acid and small SD of the reference data. In previous literature, difficulties in predicting acidity of fruits were also encountered (McGlone et al., 2003a, Liu and Ying, 2005; Shao et al., 2007; Cayuela, 2008).

3.5.3.3 Soluble solids content

A comparison of the accuracies of models calibrated on different spectral windows (Figure 3.10) showed a lower RMSECV for the full wavelength range than for the visible or NIR ranges singly or in combination. Therefore preprocessing methods and multivariate calibration were applied on the full wavelength range (400-2350 nm) instead of on a limited spectral range for SSC prediction.

The calibration statistics for PLS modeling of SSC prediction (Table 3.2) and differences in model performance, depicted graphically (Figure 3.7c), indicate that calibration results were improved by the MSC methods. Model 10 appeared to have the lowest RMSECV (0.17) and highest R^2 (0.65), with 7 PCs. However, the optimal model, with its poor accuracy resulted in an even poorer prediction for the external validation (Table 3.3 and Figure 3.8c) with an R^2 of 0.03, RMSEP of 0.15 and SDR of 1.07.

Although, SSC had been successfully predicted by VIS/NIR spectroscopic methods in various fruits including apple and kiwifruit, statistics of models for non-destructive measurement of tomato SSC by spectroscopic methods have generally been poorer (Peiris et al., 1998; Walsh et al., 2004; Shao et al., 2007). Hence, it was not unexpected to obtain such unsatisfactory results when the reference SSC values of calibration samples were analyzed (Figure 3.1c). During later maturity stages, SSC values of both varieties fluctuated without showing any trend. Although SSC usually increases with tomato ripeness stage, it tends to decrease at the red-ripe stage for some cultivars (Renquist and Reid, 1998). Thus we failed to correlate the SSC values of fruits with their light absorption over the entire spectral range employed. Additionally, the SDs of two sample sets (0.29 for calibration set and 0.16 for validation set), were relatively small compared to those of samples in other studies, making prediction

more difficult (Walsh et al., 2004). According to McGlone et al. (2007), it was impossible for spectral analysis to differentiate between soluble and insoluble carbohydrate absorbance bands, therefore, it was only possible to predict fruit SSC after the fruit had fully ripened and had no remaining starch.

3.5.3.4 Acid-Brix ratio

The performance of PLS regression models in selected spectral windows were compared (Figure 3.11). The spectral range from 700-1000 nm (window C) was selected as the optimal wavelength range for predicting ABR because the RMSECV for this range was the lowest (RMSECV=0.068). The calibration and validation statistics for the PLS modeling of ABR are shown in Table 3.2. Instead of being improved, the predictive ability of model 13 was worsened by both preprocessing methods (Figure 3.7d). Therefore, model 13, with R^2=0.63, RMSECV=0.068 and SDR=1.65, was considered as the best model among all the models tested. When model 13 was used to predict the external validation sample set, values of R^2=0.65, RMSEP=0.077 and SDR=1.52 were obtained (Table 3.3, Figure 3.8d). It is not surprising to achieve an unsatisfactory result for ABR prediction, because neither TA nor SSC had been predicted successfully and the absorption of light in the selected spectral range was relatively low.

3.6 Conclusion

By means of PLS regression methods, calibration models based on VIS/NIR spectral reflectance measurements were established to predict lycopene content, soluble solids content, titratable acidity and acid-Brix ratio of tomato fruits. It was possible to use a non-destructive technique to quantify the lycopene content of tomatoes, whereas accuracies of prediction for other properties were not satisfactory. The statistics of the best model for each parameter were as follows:

i. for lycopene content, R^2=0.93 and RMSEP=2.87 with 4 PCs;

ii. for TA, R^2=0.33 and RMSEP=0.51 mg ml^{-1} with 6 PCs;

iii. for SSC, R^2=0.03 and RMSEP=0.15 ºBrix with 7 PCs; and

iv. for ABR, R^2=0.65 and RMSEP=0.077 with 4 PCs.

In a further study, the use of different measurement modes will be investigated for predicting SSC and TA, such as interactance and transmittance modes. Additionally, a wider range of tomato varieties should be included in new research to develop robust models.

Table 3.1: Statistical analysis of the calibration and validation sample sets, including the data range, mean and standard deviation (S.D.)

Characteristic	Calibration (Number of Sample: 60)			Validation (Number of Sample: 24)		
	Range	Mean	S.D.	Range	Mean	S.D.
Lycopene (mg/kg)	0-41.15	14.51	10.63	0-36.62	14.08	11.14
SSC (ºBrix)	4.10-5.30	4.70	0.29	4.30-5.00	4.70	0.16
TA (mg/ml)	3.16-5.89	4.10	0.60	2.99-4.92	3.81	0.56
Acid/Brix	0.67-1.18	0.88	0.11	0.62-1.02	0.81	0.12

Table 3.2: Results of calibration and full-cross validation of the models

Characteristic	Model	Wavelength range (nm)	Preprocessing method	No. of PCs	Calibration		Full-cross validation			Outlier No.
					RMSEC	R^2	RMSECV	R^2	SDR	
Lycopene (mg/Kg)	model 1	400-2350	-	8	1.97	0.97	2.96	0.92		0
	model 2	450-1000	-	8	1.88	0.97	2.57	0.94	4.14	0
	model 3	450-1000	MSC	2	3.42	0.89	3.62	0.89		0
	model 4	450-1000	S.Golay 1st derivative	4	2.04	0.96	2.82	0.93		0
TA (mg/ml)	model 5	400-2350	-	8	0.26	0.78	0.36	0.59		1
	model 6	1200-1800	-	8	0.25	0.80	0.31	0.69	1.93	1
	model 7	1200-1800	MSC	5	0.32	0.66	0.37	0.57		1
	model 8	1200-1800	S.Golay 1st derivative	3	0.31	0.68	0.40	0.48		1
SSC (^{o}Brix)	model 9	400-2350	-	9	0.11	0.85	0.18	0.63		0
	model 10	400-2350	MSC	7	0.12	0.84	0.17	0.65	1.71	0
	model 11	400-2350	S.Golay 1st derivative	3	0.14	0.78	0.24	0.34		0
Acid/Brix	model 12	400-2350	-	2	0.07	0.60	0.07	0.56		0
	model 13	700-1000	-	4	0.06	0.68	0.07	0.63	1.65	0
	model 14	700-1000	MSC	2	0.07	0.55	0.08	0.49		0
	model 15	700-1000	S.Golay 1st derivative	4	0.06	0.70	0.09	0.44		0

Table 3.3: Results of external validation of the optimal models

Characteristic	Model	Wavelength range (nm)	Preprocessing method	No. of PCs	RMSEP	R^2	SDR
Lycopene	model 2	450-1000	-	8	2.15	0.96	5.18
TA	model 6	1200-1800	-	8	0.43	0.49	1.30
SSC	model 10	400-2350	MSC	7	0.15	0.03	1.07
Acid/Brix	model 13	700-1000	-	4	0.08	0.65	1.52

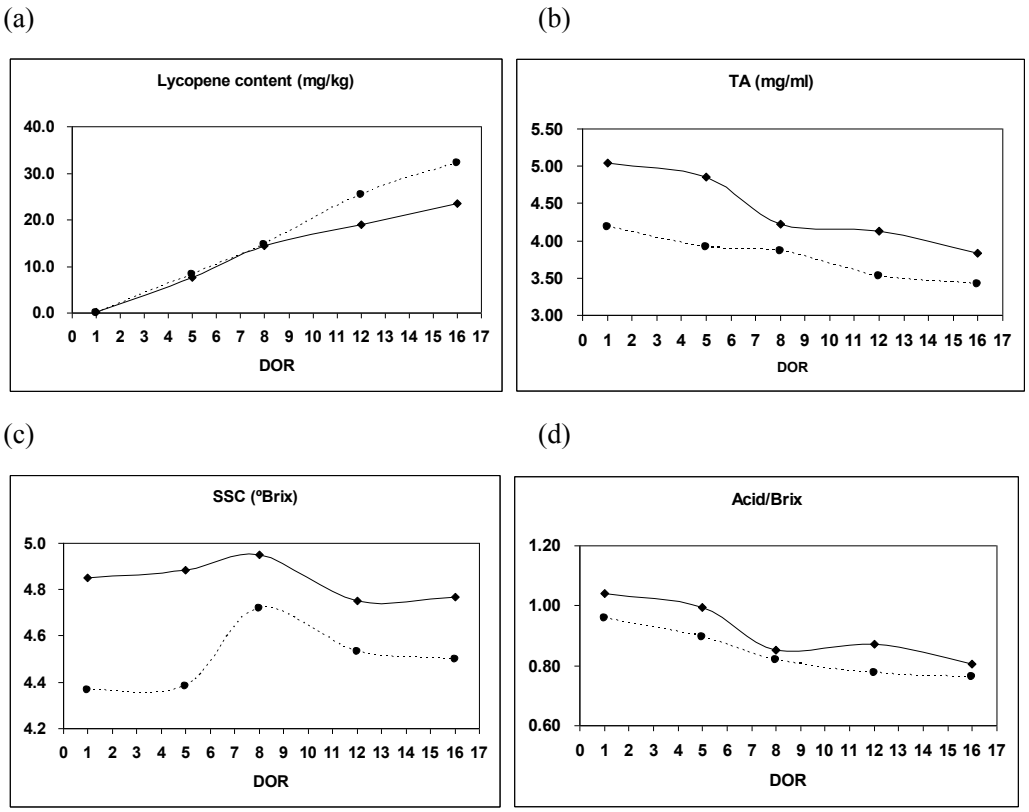


Figure 3.1: Changes of quality properties of tomatoes vs. day of ripening (DOR): (a) lycopene content; (b) titratable acidity; (c) soluble solids content; (d) acid/Brix ratio. The solid lines indicate tomatoes of cv. 'DRK 453'; the dotted lines indicate tomatoes of cv. 'Trust'.

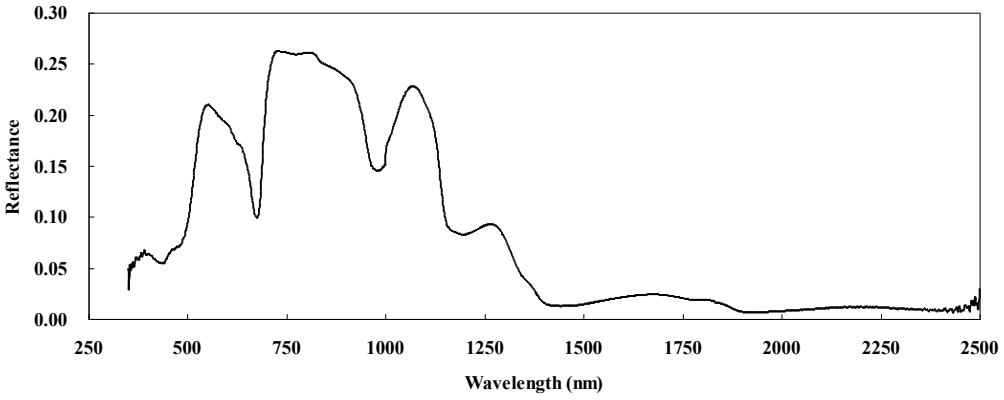
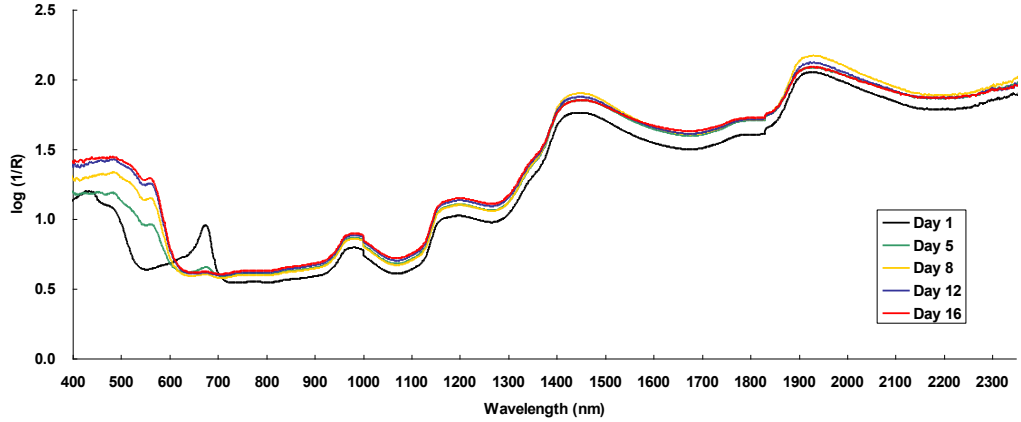


Figure 3.2: Original reflectance spectrum of one tomato (cv. 'DRK 453') at 1 day of ripening (350-2500nm).

(a)



(b)

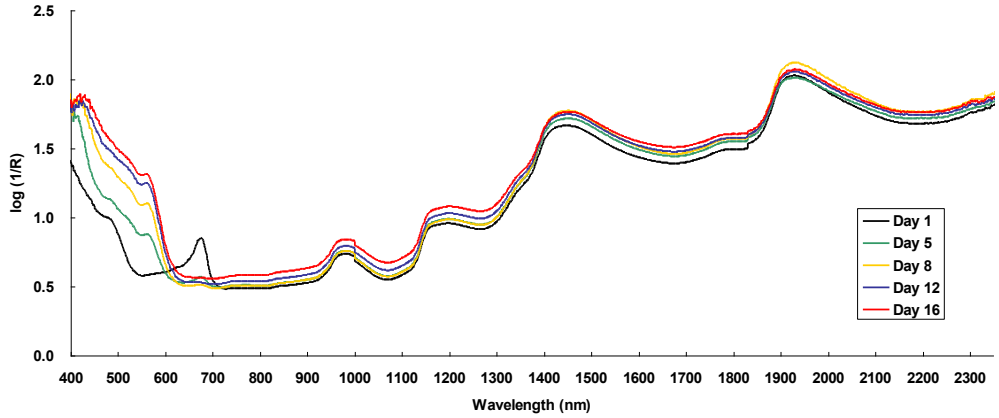
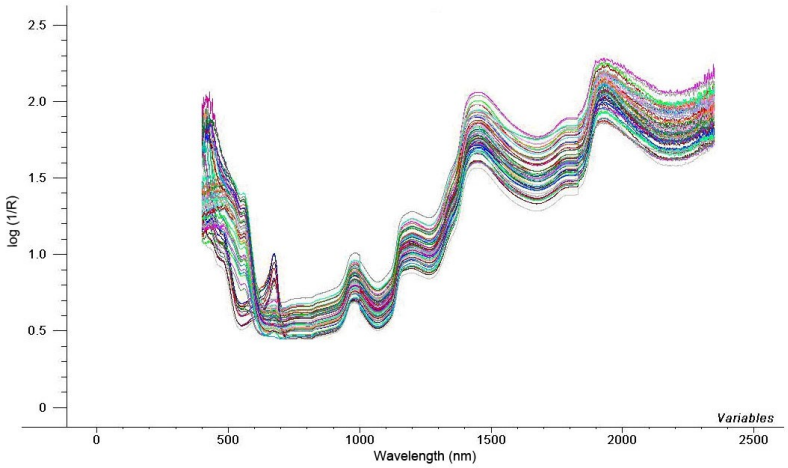
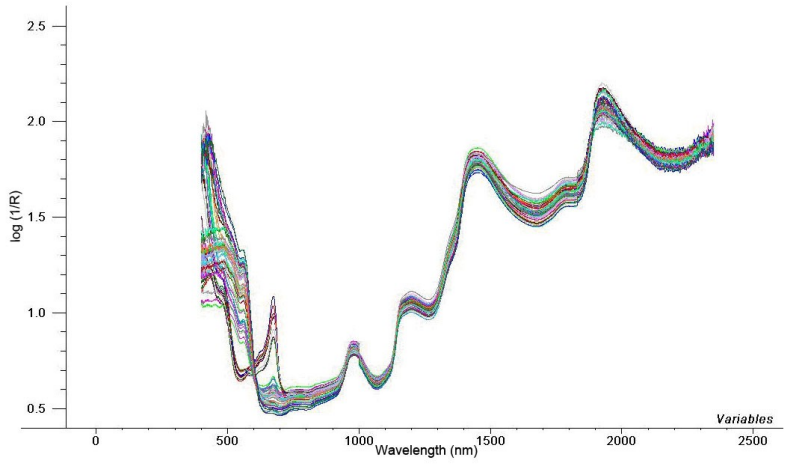


Figure 3.3: Absorbance spectra of tomatoes of two varieties measured at 1, 5, 8, 12 and 16 days of ripening (400-2350 nm): (a) cv. 'DRK453' ; (b) cv. 'Trust'.

(a)



(b)



(c)

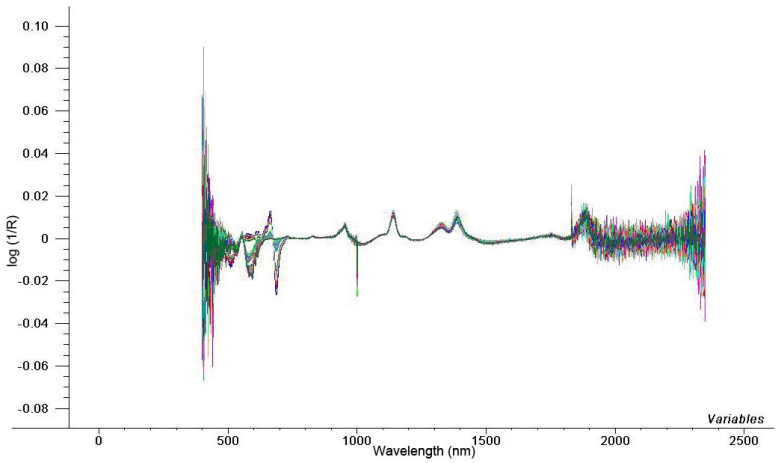


Figure 3.4: (a) Absorbance (log (1/R)) spectra of all tomatoes of calibration set. Preprocessed spectra by (b) MSC and (c) Savitzky-Golay first derivative.

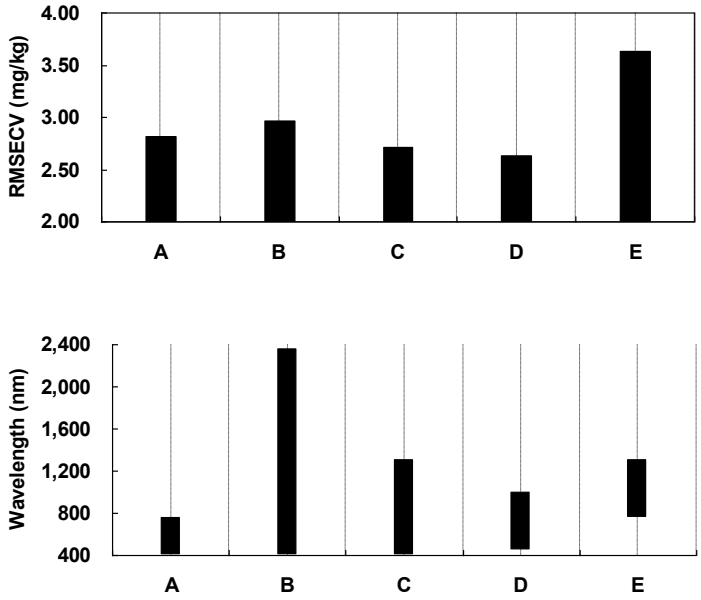
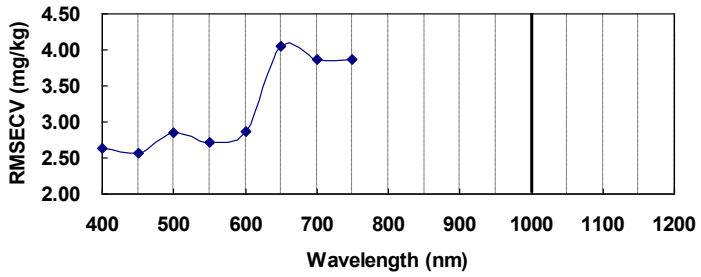


Figure 3.5: Root mean square error of full-cross validation (RMSECV) for lycopene content prediction vs. the spectral window. The vertical bars in the bottom chart indicate the wavelength range for each spectral window.

(a)



(b)

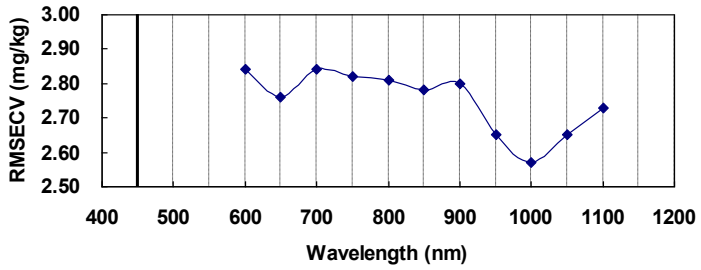
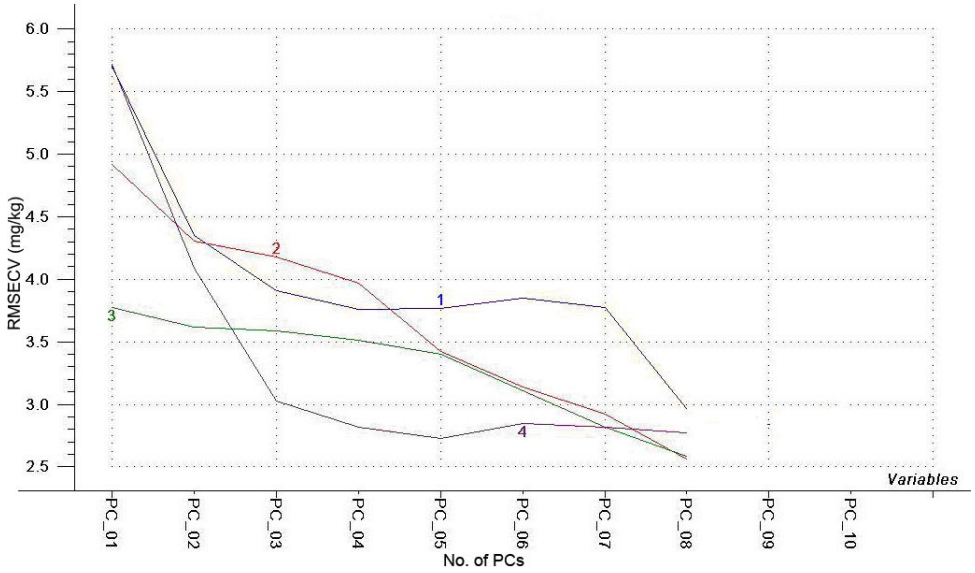
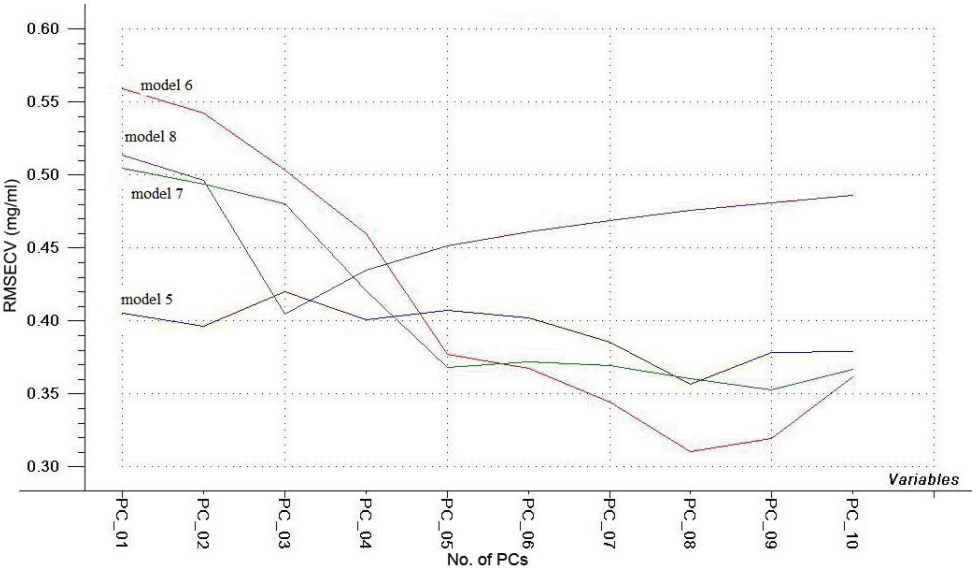


Figure 3.6: Root mean square error of full-cross validation (RMSECV) for lycopene content prediction vs. the spectral window: (a) upper wavelength limit is fixed; (b) lower wavelength limit is fixed. The bold vertical lines indicate the position for the fixed upper and lower limits in (a) and (b), respectively.

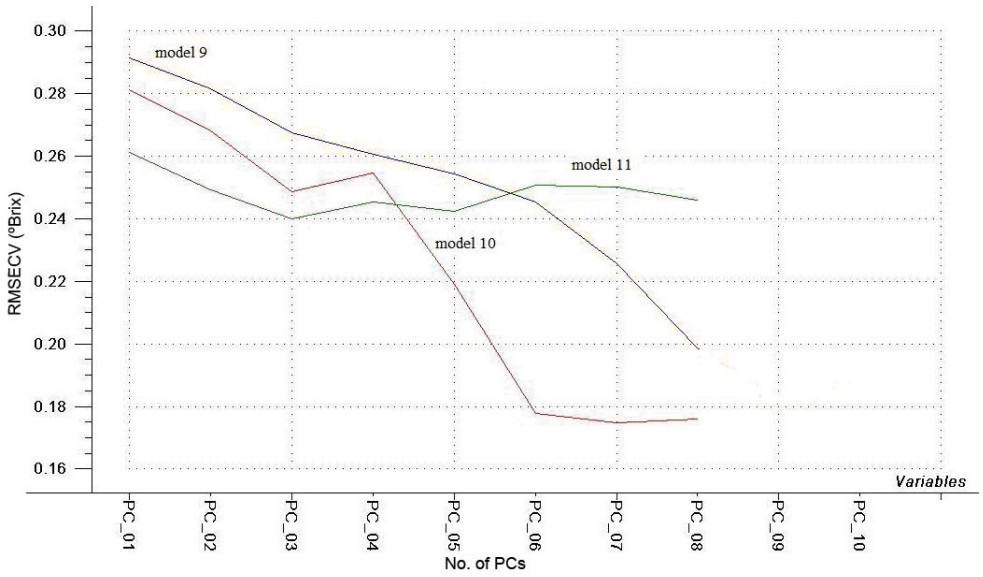
(a)


RMSECV (mg/kg)
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
1
2
3
4
Variables
PC_01
PC_02
PC_03
PC_04
PC_05
PC_06
PC_07
PC_08
PC_09
PC_10
No. of PCs

(b)


RMSECV (mg/ml)
0.60
0.55
0.50
0.45
0.40
0.35
0.30
model 6
model 8
model 7
model 5
Variables
PC_01
PC_02
PC_03
PC_04
PC_05
PC_06
PC_07
PC_08
PC_09
PC_10
No. of PCs

(c)



(d)

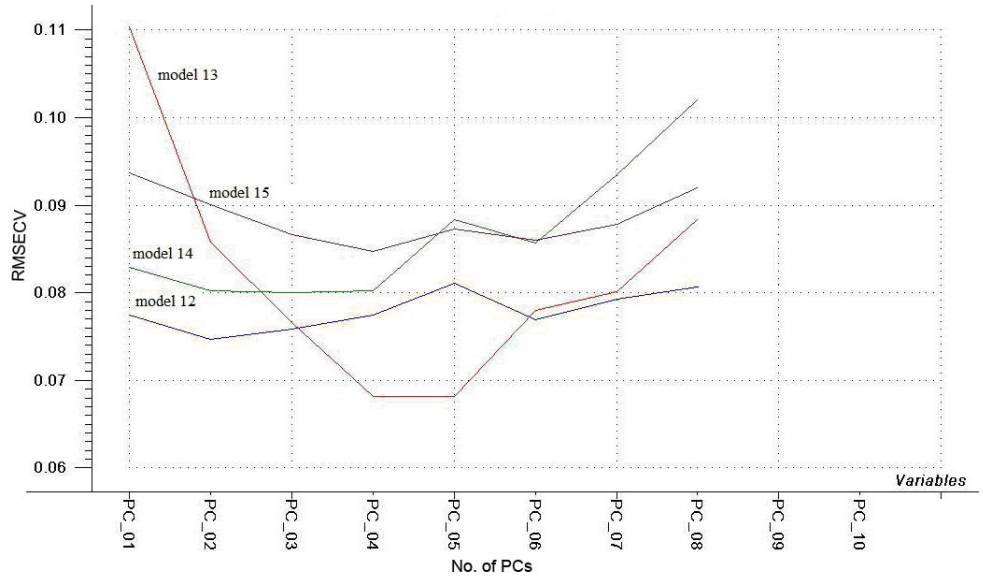
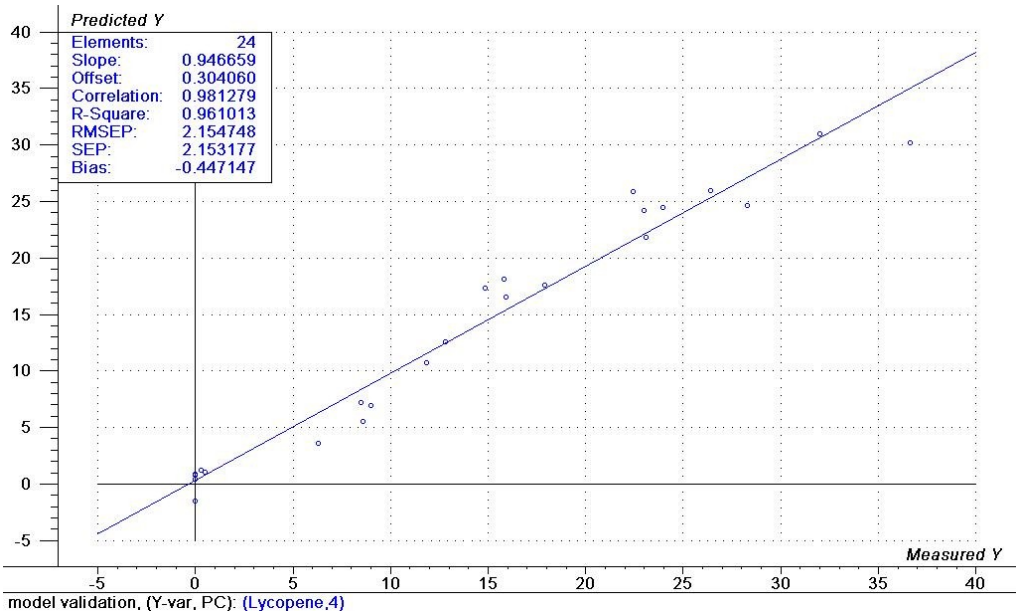
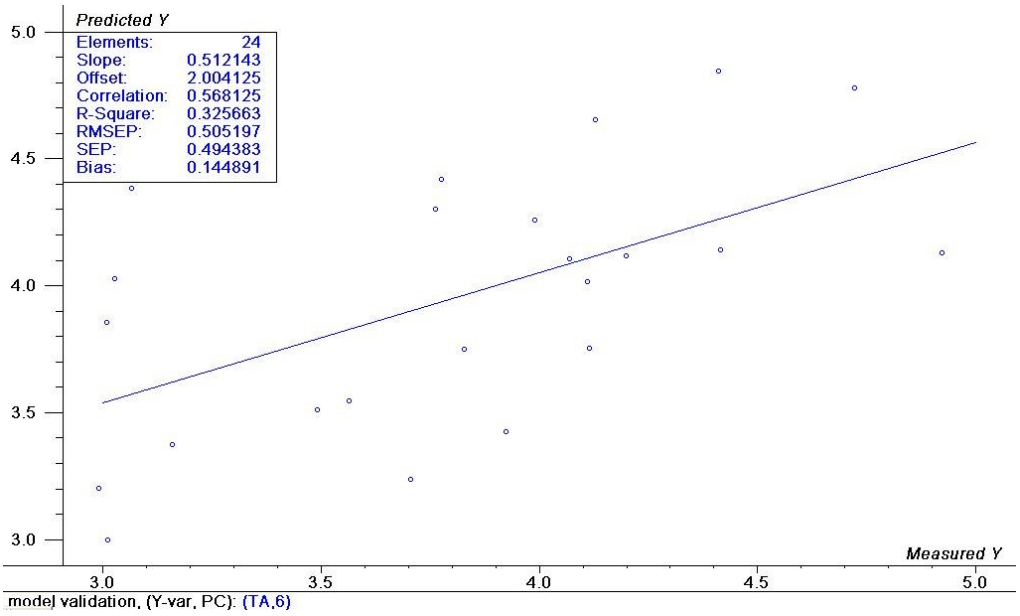


Figure 3.7: Root mean square error of full-cross validation (RMSECV) of models for each property vs. PLS components: (a) lycopene content; (b) titratable acidity; (c) soluble solids content; (d) acid/Brix ratio.

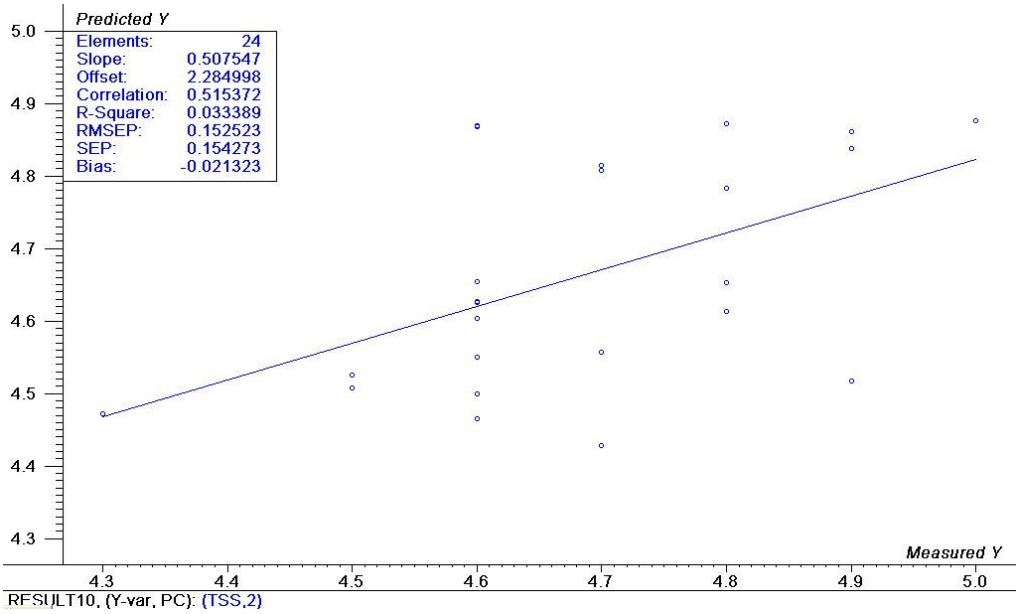
(a)


Predicted Y
Elements: 24
Slope: 0.946659
Offset: 0.304060
Correlation: 0.981279
R-Square: 0.961013
RMSEP: 2.154748
SEP: 2.153177
Bias: -0.447147
Measured Y
model validation, (Y-var, PC): (Lycopene,4)

(b)


Predicted Y
Elements: 24
Slope: 0.512143
Offset: 2.004125
Correlation: 0.568125
R-Square: 0.325663
RMSEP: 0.505197
SEP: 0.494383
Bias: 0.144891
Measured Y
model validation, (Y-var, PC): (TA,6)

(c)



(d)

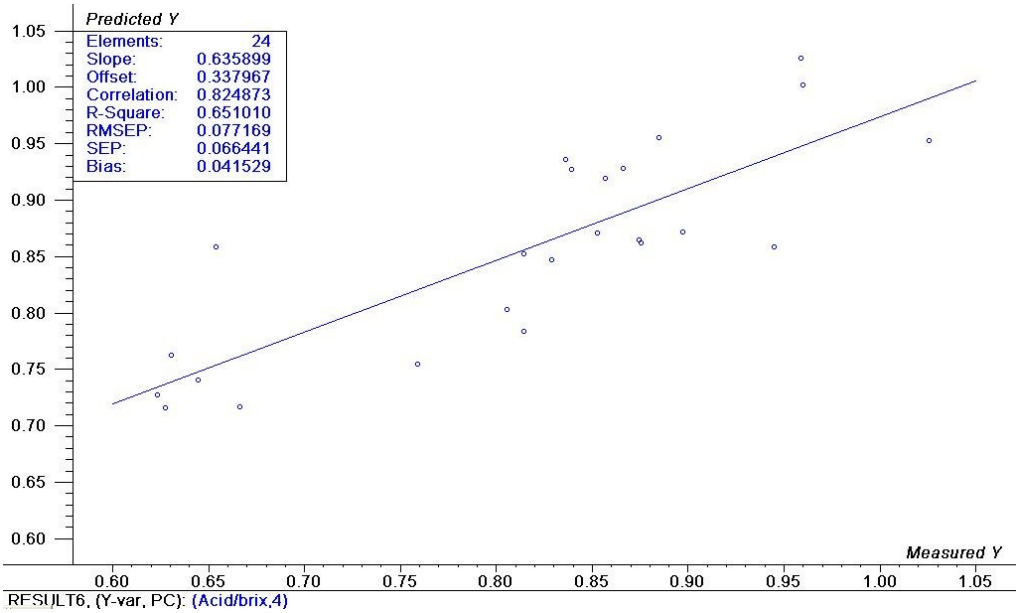


Figure 3.8: The predicted vs. the measured values of the properties of the validation set for the optimal models: (a) lycopene content; (b) titratable acidity; (c) soluble solids content; (d) acid/Brix ratio.

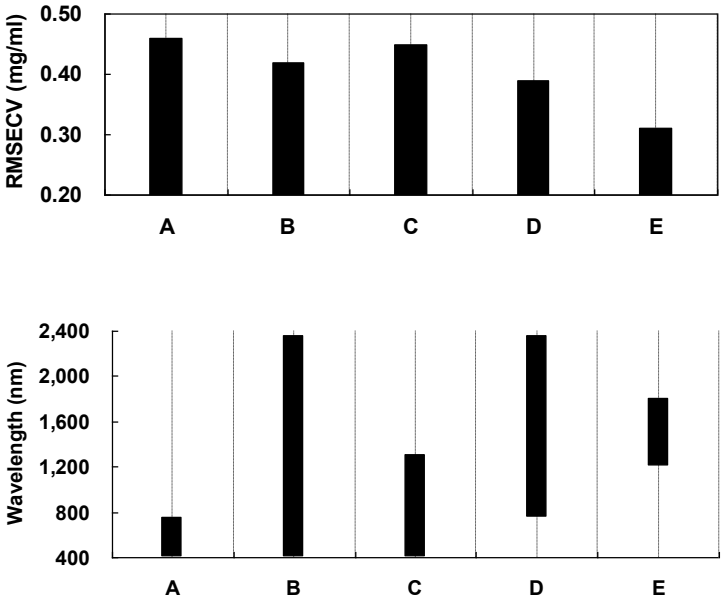


Figure 3.9: Root mean square error of full-cross validation (RMSECV) for titratable acidity prediction vs. the spectral window. The vertical bars in the bottom chart indicate the wavelength range for each spectral window.

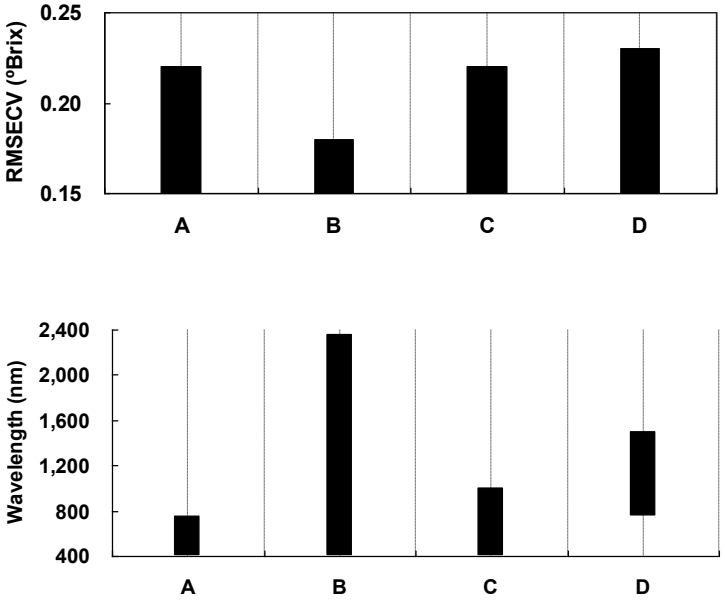


Figure 3.10: Root mean square error of full-cross validation (RMSECV) for soluble solids content prediction vs. the spectral window. The vertical bars in the bottom chart indicate the wavelength range for each spectral window.

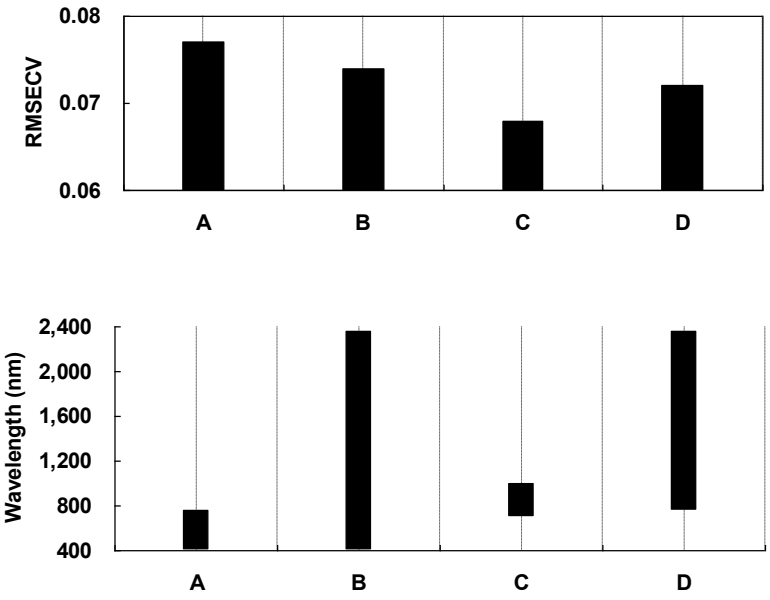


Figure 3.11: Root mean square error of full-cross validation (RMSECV) for acid/Brix ratio prediction vs. the spectral window. The vertical bars in the bottom chart indicate the wavelength range for each spectral window.

CONNECTING TEXT

The visible/near-infrared (VIS/NIR) reflectance spectra of two varieties of tomato (variety 'DRK 453' and 'Trust') at 1, 5, 8, 12 and 16 days of ripening were measured. The feasibility of predicting the nutraceutical and organoleptic parameters, including lycopene content, soluble solids content (SSC), titratable acidity (TA), and acid-Brix ratio (ABR), using a partial least square 1 (PLS1) method was studied and the results were discussed in Chapter III. Further, the prediction models for physical parameters of tomato, including colour and firmness, were established and validated. Colour, firmness and lycopene content were simultaneously calibrated using a PLS2 method. The results are presented and discussed in Chapter IV.

CHAPTER IV

MODELING OF PHYSICAL AND CHEMCIAL ATTRIBUTES OF TOMATOES USING VIS/NIR SPECTROSCOPY

4.1 Abstract

Non-destructive models employing visible/near-infrared (VIS/NIR) reflectance spectroscopy were developed to predict characteristics of two varieties of tomato (*Lycopersicon esculentum* Mill cv. 'DRK 453' and 'Trust'). Tomato properties including colour, firmness and lycopene content were also measured by conventional methods. Partial least squares (PLS) regression analysis was performed on the spectral data to build prediction models. Various spectral windows within the 400-2350 nm spectral range and pre-processing methods including multiple scatter correction (MSC) and Savitzky-Golay first derivative were assessed in optimizing the model for each parameter. R^2 values of 0.99, 0.99 and 0.97, and root mean square errors of prediction (RMSEP) of 0.06, 1.52 and 1.44 N for individual prediction models of colour values a*/b*, tomato colour index (TCI) and firmness, respectively. Further, a model built by the PLS2 method showed excellent performance in simultaneously predicting colour, firmness and lycopene content of tomato. The R^2 values of the PLS2 model for a*/b*, TCI, firmness and lycopene content were 0.99, 0.99, 0.97, and 0.92, respectively, while their RMSEP values were 0.06, 1.75, 1.44 N, and 3.03 mg kg^{-1}.

4.2 Introduction

Skin colour and firmness are the two most important attributes of tomato for customer evaluation (Tijskens and Evelo, 1994). Therefore, the perception of fruit colour is probably the very first quality evaluation that determines consumers' buying decision (Garrett et al., 1960). This is also the case because colour is also an indicator of tomato ripeness. Colour evolves during fruit ripening from green to red and is

mainly related to the breakdown of chlorophyll and synthesis of lycopene (Hobson and Davies, 1971). Tomato maturity is traditionally classified in six stages, based on the external colour changes of the fruit: mature-green, breaker, turning, pink, light-red and red-ripe (Gould, 1992). According to CIELAB L*, a*, b* color space, an increase in a* corresponds to a change from green to red, a more positive b* value represents increasing yellowness and L* represents lightness of the colour. The tomato colour index (TCI) relates these three parameters in a single-value criterion (Richardson and Hobson, 1987). Although a* is suitable to assess red colour development, colour changes in tomato are usually recorded as a*/b* ratio (Tijskens and Evelo, 1994) since it was proven to be a better index than a* alone or the TCI in distinguishing varieties (Gómez et al., 2001).

Firmness of tomatoes may be the final index by which the purchase decisions are made (Gormley and Egan, 1978). Softening takes place during storage and ripening of tomatoes. The exact molecular changes that drive fruit softening are still unknown. However, it is known that number of cell wall hydrolytic enzymes contribute to tissue softening and intercellular adhesion (Fisher and Bennett, 1991). Neither unripe nor overripe fruits are desirable.

Colour is commonly measured by chromameter which can give accurate values about colour, while different instruments for measuring firmness of fruits and vegetables have been used, such as penetrometers and universal testing machines. Not only the instrumental methods for firmness measurement are destructive, but also measurements for each attributes have to be done separately. One of the advantages of NIR spectroscopic method for quality analysis is that it would potentially be possible to measure different attributes simultaneously.

The use of NIR spectroscopy in assessing firmness has been already studied for various fruits, including apple (Moons et al., 1997; Lammertyn et al., 1998; Lu et al., 2000; McGlone et al., 2002a; Park et al., 2003), kiwifruit (McGlone and Kawano, 1997); mandarine (Gómez et al., 2006), mango (Schmilovitch et al., 2000), pear (Nicolaï et al., 2008), cherry (Lu, 2001), tomato (Shao et al., 2007). Moons et al. (1997) reported a good correlation of 0.96, but the prediction error was high (standard

error of prediction (SEP)=9.3 N). Schmilovitch et al. (2000) compared MLR and PLS methods in predicting mango firmness and found that MLR gave a better result (R^2=0.8226 and SEP=17.14). Lu (2001) developed calibration models with good results to predict firmness of two varieties of sweet berries. Park et al. (2003) obtained a result of R^2=0.786 and SEP=7.02 using principal component regression (PCR) and a spectral range from 400nm to 1800nm. Gómez et al. (2006) established their best calibration model for firmness of mandarine to have a correlation of 0.83 and RMSEP of 8.53 based on PLS. However, some researchers consider that VIS/NIR spectroscopy is not a robust method for firmness measurement (McGlone and Kawano, 1997; Lu et al., 2000; McGlone et al., 2002a, Nicolai et al., 2008).

The potential of VIS/NIR spectroscopy in predicting the quality attributes of tomatoes and tomato products has been evaluated by several researchers (Slaughter et al., 1996; Hong and Tsou, 1998; Goula and Adamopoulos, 2003; Jha and Matsuoka, 2004; Khuriyati et al., 2004; Pedro and Ferreira, 2005, 2007; Baranska et al., 2006; Shao et al., 2007; Clément et al., 2008). However, only one paper has been published about predicting the firmness of tomatoes (Shao et al., 2007). They investigated VIS/NIR spectroscopy to measure the firmness of tomato variety "Heatwave" which was also determined by two reference methods, including a compression test and a Magness-Taylor puncture test. Relatively good results were obtained for the two types of firmness indices, which are r=0.81 and SEP=16.017N for compression force and r=0.83 and SEP=1.18N for puncture force. More research to develop calibration models for a wider range of tomato varieties is needed.

4.3 Objectives

In light of all the research performed, the present study on VIS/NIR spectroscopy was undertaken with the following objectives.

1. To evaluate the feasibility of determining the quality attributes of intact tomato fruits based on VIS/NIR spectroscopy;
2. To establish calibration models for predicting colour and firmness of tomato;
3. To build a calibration model by PLS2 method to simultaneously predict

physico-chemical properties of tomato, including colour, firmness and lycopene content.

4.4 Materials and methods

4.4.1 Sample preparation

Tomato fruits (*L. esculentum* Mill.) of two varieties (cv. 'DRK 453' and cv. 'Trust') were harvested from a commercial greenhouse located in Saint-Damase, QC, Canada. The experiment was repeated three times. On each occasion 30 mature green tomatoes (15 'DRK 453' and 15 'Trust') were stored and allowed to ripen at 16ºC and 90-93% RH. Six tomato fruits (3 from each variety) were subjected to spectroscopic measurements and reference physico-chemical analyses at 1, 5, 8, 12 and 16 days of ripening (DOR), allowing most tomato maturity stages to be sampled. The tomatoes were washed with distilled water and dried thoroughly before spectroscopic measurements.

4.4.2 Acquisition of spectra

All spectral measurements were performed using a spectroradiometer (FieldSpec® Pro FSP 350-2500P/A110000, Analytical Spectral Devices Inc., Boulder, CO) coupled with a contact probe of 10mm spot size (Analytical Spectral Devices Inc., Boulder, CO). For each fruit, reflectance spectra (350-2500 nm) were taken at six (1 and 5 DOR) or four (8, 12 and 16 DOR) equidistant positions around the equator. For each reflectance spectrum, 10 scans were averaged at any given time/position, for a total of 60 or 40 scans per fruit per sampling. Reflectance, *R*, was calculated as the ratio of the near-infrared energy reflected from the sample surface to a standard reference (Spectralon disk). The signals were preprocessed with ViewSpec Pro V2.14 (Analytical Spectral Devices Inc., Boulder, CO).

4.4.3 Reference analyses

After the spectral measurements, each fruit was subjected to colour and firmness measurements at the same positions around the equator.

4.4.3.1 Colour

Based on the Commission Internationale de l'Éclairage (CIE), the colour of tomatoes was determined in terms of the colour values L*, a* and b* using a Minolta chromameter CR-400 (Minolta Co., Osaka, Japan). The three colour values were used to calculate the tomato colour index (TCI) according to the equation 4.1 (Richardson and Hobson, 1987).

$$TCI = \frac{2000\,a^*}{L^*(a^{*2} + b^{*2})^{0.5}} \tag{4.1}$$

4.4.3.2 Firmness

A universal testing machine (Lloyd Instrument, LRX, Fareham, UK) equipped with a 100 N load cell and fitted with a standard 11 mm diameter hemispherical-tip probe driven downwards at a speed of 25 mm min^{-1} to a depth of 5.5 mm was used in firmness testing. The firmness was expressed as the peak force and recorded in N.

4.4.3.3 Lycopene content

After the firmness measurements, each fruit was homogenized with a Waring blender for 1 minute at maximum speed and the resultant slurry filtered through two layers of cotton cloth. The tomato juice extract was used for lycopene content analysis.

Lycopene content was determined according to the reduced volumes of organic solvents method of Fish et al. (2002). Absorbance of the hexane upper phase was measured at 503 nm against a pure hexane blank, using a spectrophotometer (Biochrome, Ultrospec 3100 Pro, UK). The lycopene content of each tomato was estimated from an average of two aliquots of the same homogenate using equation 4.2:

$$\text{Lycopene}\ (\text{mg kg}^{-1}) = \frac{(A_{503} * 31.2)}{\text{Weight of sample used (g)}} \tag{4.2}$$

4.4.4 Data analysis

Tomato samples were split into two sets. For the first and second batches of tomato fruits, 60 samples were used for calibration and the third batch had 24 tomato fruits which served as an external validation set. Data analysis was carried out using

"The Unscrambler v9.7" (CAMO Inc., Woodbridge, NJ), a statistical software package for multivariate calibration.

All fruit reflectance measurements were transformed to absorbance (log (1/R)) values to obtain linear correlations of the NIR values with physico-chemical parameters measured by reference methods (e.g., a*/b*, TCI and firmness). Given that noise was clearly present at the extremities of all reflectance spectra, and would affect the accuracy of calibration, only spectral data in the range of 400 nm to 2350 nm were taken into consideration.

4.4.4.1 Wavelength range selection

To improve calibration results when assessing various tomato fruit attributes (McGlone et al., 2002a, 2002b, 2003a) several spectral windows were tested from exclusively visible to exclusively NIR wavelengths, as well as various VIS/NIR combinations. The best wavelength ranges were selected based on the root mean square error of prediction (RMSEP) of models.

For a*/b* ratio, four windows (A: 400-750nm, B: 450-600nm, C: 400-2350nm, D: 400-1500nm) were chosen. For TCI, four windows (A: 400-750nm, B: 430-700nm, C: 400-2350nm, D: 430-1400nm) were chosen. For firmness, four windows (A: 400-750nm, B: 400-2350nm, C: 500-1100nm, D: 750-1100nm) were chosen. These ranges were chosen on the basis of prior research on various fruits and Martens' automatic uncertainty test which shows significant variables (Westad and Marten, 2000).

4.4.4.2 Spectra preprocessing

The data obtained from the NIR spectrometer contained information not relevant to predicting sample information, such as scattering effects and instrument noise. Pre-processing was used to diminish such irrelevant information and obtain reliable and accurate calibration models. To test the influence of this pre-processing on the quality of calibration models, two types of pre-processing methods, MSC and Savitzky-Golay first derivative, were used on the best wavelength range selected for each parameter.

MSC was used to compensate for additive (offset shifts) and multiplicative

(amplification) effects in spectral data and deal with path length problems (Martens and Naes, 1989). Light scattering occurs in fresh fruit, as light does not always travel the same distance in the sample before returning to the detector. A longer light travel path corresponds to a lower relative reflectance value, since more light is absorbed. This kind of variation was eliminated by MSC. First order derivative pre-processing was applied using the Savitzky-Golay algorithm to correct for baseline effects in spectra.

4.4.4.3 Multivariate calibration

A PLS regression method was used to build the calibration models. In general, PCR and PLS are the methods most often chosen for NIR spectral analysis. PLS usually gives similar results as PCR, but uses fewer PCs (De Jong, 1993). PLS is a two-step method. Firstly the original independent information (X-variables) is projected onto a small number of latent variables (LV) to simplified the relationship between X-variables and Y-variables as the relationship is concentrated on the smallest number of LVs (Naes, et al., 2004).

$$X = TP + E \tag{4.3}$$

$$Y = UQ + F \tag{4.4}$$

where T and U are the scores of the X and Y matrices, P and Q are the loadings of the X and Y matrices, and E and F are the residuals for matrices X and Y. In the second step T and U are processed by linear regression. Thus PLS not only estimates component concentrations, but also assesses chemical and physical properties from VIS/NIR spectra (Lammertyn et al., 1998; Gómez et al., 2006).

The PLS2 method was used to build a calibration model to simultaneously predict a*/b* ratio, TCI, firmness, and lycopene content, with spectral data within the wavelength range of 450-1000 nm.

Full cross validation was performed on the calibration samples to determine the optimal number of PCs and also to validate the models. With this method, one sample is left out from the calibration data set and the model is calibrated with the remaining data points. Then the value for the left-out sample is predicted and the prediction residual is computed. The process is repeated until every observation has

been left out of the calibration set once; then all prediction residuals are combined to calculate the root mean square error of cross validation (RMSECV).

4.4.4.4 Selection of the best models

Choosing the best models should be based on a number of parameters, including RMSECV, root mean square error of calibration (RMSEC), the difference between RMSECV and RMSEC, the correlation coefficient (r) of validation, and the number of PCs. A good model should have a low RMSEC, a low RMSECV or RMSEP, a high correlation coefficient but also a small difference between RMSEC and RMSEP. A large difference indicates that too many PCs are used in the model and noise is being modelled (Lammertyn et al., 1998). Moreover, a relatively low number of PCs is generally desirable. RMSEC and RMSECV or RMSEP are defined using equation 4.5 and equation 4.6:

$$\text{RMSEC} = \sqrt{\frac{1}{I_c}\sum_{i=1}^{I_c}(\hat{y}_i - y_i)^2} \tag{4.5}$$

$$\text{RMSECV or RMSEP} = \sqrt{\frac{1}{I_p}\sum_{i=1}^{I_p}(\hat{y}_i - y_i)^2} \tag{4.6}$$

where

$\hat{y}_i$ = is the predicted value of the i^{th} observation,

y_i = is the measured value of the i^{th} observation,

I_c = is the number of observations in the calibration set,

I_p = is the number of observations in the validation set.

The, standard deviation ratio (SDR) is defined as the deviation ratio of the response variable to the RMSEP (McGlone and Kawano 1998). This was also used as a parameter to evaluate the quality of models. An SDR value between 2.5 and 3 or above corresponds to good and excellent prediction accuracies, respectively.

4.4.4.5 External validation

By means of full cross validation, the validation samples originate from the same sample set as the calibration, so spectral variability is included. It is necessary to

validate models with an independent set of samples. After the optimal calibration model for each parameter was selected, it was applied to the external validation set. The prediction results, such as R^2, RMSEP and SDR, were calculated to assess the accuracy of the model.

4.5 Results and Discussion

4.5.1 Quality characteristics of tomato

The summary of the statistics for all quality attributes of calibration and validation sample sets are listed in Table 4.1 and the changes in these quality attributes of calibration samples of both varieties are illustrated in Figure 4.1. The mean (± standard deviation) colour value a*/b* ratio, TCI and firmness of the validation set, measured by reference methods were 0.68 (±0.64), 21.64 (±20.38) and 16.33 N (±8.81 N), respectively. As can be seen, colour value a*/b* ratio, TCI and firmness showed a linear correlation with the ripening stages of the tomatoes. Colour value a*/b* ratio and TCI increased while fruit firmness decreased as the fruits ripened.

4.5.2 VIS/NIR spectra

Figure 4.2 shows the average original reflectance spectrum of one tomato in the wavelength range between 350 and 2500 nm. The absorbance spectra of tomatoes of different maturity stages for both varieties from 400nm to 2350 nm are shown in Figure 4.3a and Figure 4.3b, respectively. The shapes of the spectra are quite consistent and some peaks and valleys revealing the characteristics of the fruit are in evidence.

For variety ‘DRK 453’, important differences exist between the mean spectrum of mature green tomatoes and those at later ripening stages over the entire wavelength range. Variations among the spectral curves of maturity stages after mature green are only apparent between 400 to 700 nm, Except for the spectrum of mature green tomatoes (measured at DOR 1), a continuous decrease in absorbance in the visible region (400-750 nm) with two small peaks at 560 nm and 675 nm can be observed, followed by a fluctuating increase in the whole NIR region until 2350 nm,

with a maximum at about 1930 nm. The absorption curve of mature green tomatoes has an obvious peak in the visible range around 675 nm, which corresponds to the chlorophyll absorbance band. After this peak, the absorbance spectrum remains relatively flat until 900 nm. A prominent peak occurs at 980 nm, attributable to absorption by water and carbohydrates and corresponds with a strong absorption band of water (960-990 nm), a dominant component of fruit (Williams and Norris, 2001). Another two peaks occurred at 1197 nm and 1448 nm, related to strong water absorption bands from 1125 to 1270 nm and from 1440 to 1485 nm. Basically the absorbance curve pattern for tomatoes is very similar to that of other fruits, such as apple, cherry, mandarin, and orange (Lu et al., 2000; Lu, 2001; Gómez et al., 2006; Cayuela, 2008). Within the wavelength range of 400-1300 nm, except for 600-715 nm, the more mature tomatoes were, the higher was the absorbance. The change in the average absorbance with storage time was consistent with that of colour value a*/b*, TCI and firmness.

As for the spectra of calibration samples of variety 'Trust', they showed similar trends and characteristics as those of DRK 453, except that spectral differences between mature green and riper tomatoes more muted in 'Trust' than 'DRK 453'.

Calibration samples' raw absorbance spectra (400-2350nm; Figure 4.4a) and the same spectra preprocessed by MSC (Figure 4.4b) or S. Golay first derivative (Figure 4.4c) show that consistent offset shifts and baseline shifts initially exist in the spectra, due to light scattering. It is apparent that the peak and valley positions of the preprocessed spectra correspond to those of original spectra. However, it can be seen that the MSC method eliminates the baseline shifts and make the peaks and valleys clearer, whereas the Savitzky-Golay first derivative method differentiate overlapped peaks in the spectra. Taken together, these observations suggest that the possibility exists of relating the spectra of tomatoes to their quality attributes, particularly in the visible wavelength range of 400-700 nm.

4.5.3 Prediction of quality parameters

4.5.3.1 Colour value a*/b* ratio

The lowest value of RMSECVs for a*/b* ratio prediction by PLS regression

models was obtained for a visible range window of 450-600 nm (window B; Figure 4.5). Windows within the visible range were distinctly superior to the full range window or a combined VIS/NIR range (window C and D; Figure 4.5). Colours perceived by the human eye are due to the reflection of visible light ranging from 380 to 750 nm from objects, and it is essential to include this range to predict fruit colour. However, there are several water absorption bands in the NIR range which are much stronger than the absorption bands of pigment components (Williams and Norris, 2001). Therefore, in the present study, the NIR range was excluded to avoid its influence on colour prediction. Additionally, a further study was made to find the optimal wavelength range by altering the lower and upper wavelength limit on window B. Figure 4.6 shows that the lowest RMSECV value in each plot was achieved at wavelengths of 450nm and 600 nm respectively. So window B was chosen as the best range for a*/b* prediction.

A comparison of the performances of the PLS-built model without pretreatment and with MSC and Savitzky-Golay first derivative pre-treatments (Table 4.2, Figure 4.7a) shows that the lowest RMSECV was achieved with model 4, with 2 PCs. Model 4, calibrated by data pretreated by first derivative on a range of 450-600 nm, was selected as the best model (R^2=0.99, RMSECV=0.06 and SDR=8.89). The prediction accuracy for external validation was also excellent (R^2=0.99 and RMSEP=0.06; Table 4.3, Figure 4.8a).

4.5.3.2 Tomato colour index

Performance of PLS models built on different spectral ranges for TCI prediction was evaluated by their RMSECVs (Figure 4.9). The model derived from spectral window D (430-1400nm) generated a lower RMSECV (2.02) than those obtained with other spectral windows. Further study to test if window D was the optimal wavelength range was done by altering its lower and upper wavelength limits. The RMSECVs for every range were plotted (Figure 4.10), and showed that the lowest RMSECV values in each plot was achieved for wavelengths of 430nm and 1400 nm respectively.

Coupling of PLS models for TCI with MSC or Savitzky-Golay first derivative

pre-processing methods showed no improvement in predictive performance over the PLS model alone (Table 4.2, Figure 4.7b). Model 6 was the optimal model in predicting TCI ($R^2 \geq 0.99$), RMSECV = 2.02 and SDR = 8.84, and contained three PCs. Used in a predictive capacity (validation), Model 6 performed well (R^2=0.99 and RMSEP=1.52; Table 4.3, Figure 4.8b). The calibration and validation statistics for PLS modelling of TCI prediction are shown in Table 4.2.

4.5.3.3 Firmness

A PLS model with 4 PCs and using spectral window C (500-1100 nm) showed the lowest RMSECV (2.24; Figure 4.11), and thus a slightly better performance than the visible range (window A) or full wavelength range (window B), and clearly superior performance to the NIR range (window D). In order to test if 500-1100 nm was the optimal spectral range for firmness prediction, the lower and upper wavelength limits on window C were shifted, as shown in Figure 4.12. The lowest RMSECV value appeared at wavelengths boundaries of 500nm and 1100 nm, respectively. So window C was chosen as the best range for firmness prediction. Predictive performance of the model sharply deteriorated as the lower limit of the window was set beyond 700 nm which excluded the chlorophyll absorbance band at 680 nm (Figure 4.12a).

Two preprocessing methods, MSC and Savitzky-Golay first derivative, were applied to the spectral range of 500-1100 nm to find the best calibration model for firmness prediction (Table 4.2, Figure 4.7c). Model 10, with 4 PCs, showed the lowest RMSECV (2.24) and gave a good performance for the calibration set (R^2=0.90, SDR=3.10). Furthermore, results of external validation were also excellent (R^2=0.97 and RMSEP=1.44 N; Figure 4.8c). This model's performance in predicting tomato firmness was superior to that of Shao et al. (2007) whose model performance statistics were r = 0.82, RMSEP=15.80 N and SDR=1.54. The present result is also much better than those achieved for other fruits: $0.38 < r < 0.58$ for apple firmness using wavelengths from 800 to 1700 nm (Lu et al., 2000), and a R^2 of 0.22 and 0.79 for the same parameter of Gala apple and Red Delicious apple, respectively, with wavelengths of 400-1800 nm (Park et al., 2003).

4.5.3.4 Quality parameters prediction using PLS2

Correlations existing among all quality parameters (*y*-variables) under study, including L*, a*, b*, a*/b*, TCI, firmness, and lycopene content, are shown in a correlation loading plot (Figure 4.13). Variables proximal to one another in the loading plot have a high positive correlation, while variables diagonally opposite to each other are negatively correlated. The values of a*, a*/b*, TCI and lycopene were positively correlated, as were L* and firmness; while firmness was negatively correlated with TCI, lycopene and a*/b*. However, b* varied independently. Correlation coefficients of properties amongst each other (Table 4.4.) show that variables TCI, firmness and lycopene content presented high correlation with a*/b* ratio, a property that could be determined quite accurately by the VIS/NIR spectroscopic method. Therefore, it should be possible to predict a*/b*, TCI, firmness, and lycopene content simultaneously by the PLS2 method.

A pre-test to find an optimal wavelength range was made over different spectral ranges, with the range of 450-1100 nm being found optimal for firmness and lycopene content prediction, which were relatively hard to predict from spectroscopic data. Model 13, built on non-preprocessed data, showed an excellent performance on all properties (Table 4.5). Model 14 showed a better performance in predicting firmness but poorer for other parameters, and model 15 showed a better performance in predicting lycopene content, but a poorer prediction of firmness. Therefore, model 13 was chosen as the best PLS2 model. The prediction results of external validation were also excellent for prediction of colour, firmness and lycopene content (Table 4.5, Figure 4.14).

4.6 Conclusion

By means of PLS regression methods, calibration models based on VIS/NIR spectral reflectance measurements were established to predict a*/b* ratio, TCI and firmness of tomato fruits, respectively. The model's excellent performance indicates that it is possible to use a non-destructive technique to analyze these physico-chemical

properties of tomato. Also, a suitable PLS2 calibration model was obtained for simultaneous determination of colour, firmness and lycopene content of the fruit. In a further study, we should investigate a wider range of tomato varieties to make the models more robust.

Table 4.1: Statistical analysis of the calibration and validation sample sets, including the data range, mean and standard deviation (S.D.)

Characteristic	Calibration (Number of samples: 60)			Validation (Number of samples: 24)		
	Range	Mean	S.D.	Range	Mean	S.D.
a*/b*	−0.36-1.47	0.73	0.56	−0.45-1.50	0.68	0.64
TCI	−13.92-45.37	24.16	17.85	−14.95-44.37	21.64	20.38
Firmness (N)	7.65-34.33	15.16	6.95	7.67-39.98	16.33	8.81

Table 4.2: Results of calibration and full-cross validation of the models

Characteristic	Model	Wavelength range (nm)	Preprocessing method	No. of PCs	Calibration		Cross-validation			Outlier No.
					RMSEC	R^2	RMSECV	R^2	SDR	
a*/b*	model 1	400-2350	-	4	0.08	0.98	0.09	0.97		1
	model 2	450-600	-	3	0.06	0.99	0.07	0.98		1
	model 3	450-600	MSC	2	0.08	0.98	0.09	0.97		1
	model 4	450-600	1st derivative	2	0.06	0.99	0.06	0.99	8.89	1
TCI	model 5	400-2350	-	4	2.24	0.98	2.53	0.97		0
	model 6	430-1400	-	3	1.84	0.99	2.02	0.99	8.84	0
	model 7	430-1400	MSC	4	2.15	0.98	2.49	0.98		0
	model 8	430-1400	1st derivative	4	1.80	0.99	2.29	0.98		0
Firmness (N)	model 9	400-2350	-	4	2.16	0.90	2.44	0.88		0
	model 10	500-1100	-	4	1.84	0.93	2.24	0.90	3.10	0
	model 11	500-1100	MSC	2	2.17	0.90	2.33	0.89		0
	model 12	500-1100	1st derivative	1	2.39	0.88	2.53	0.87		0

Table 4.3: Results of external validation of the optimal models

Characteristic	Model	Wavelength range (nm)	Preprocessing method	No. of PCs	RMSEP	R^2	SDR
Colour a*/b*	model 4	450-600	S.Golay 1st derivative	2	0.06	0.99	10.49
TCI	model 6	430-1400	-	3	1.52	0.99	13.41
Firmness	model 10	500-1100	-	4	1.44	0.97	6.12

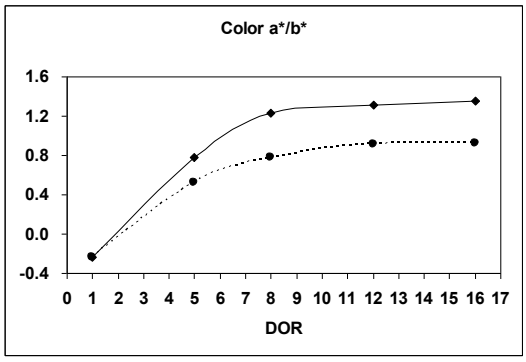
Table 4.4: Correlation coefficients of the properties of tomatoes

	a*/b*	TCI	Firmness	Lycopene
a*/b*	1.00			
TCI	0.98	1.00		
Firmness	0.83	0.63	1.00	
Lycopene	0.56	0.89	0.66	1.00

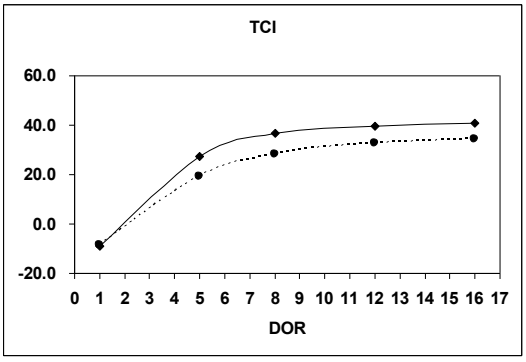
Table 4.5: Results of calibration and full-cross validation of the models built using PLS2 and results of external validation of the optimal model

Model	Preprocessing method	No. of PCs	Characteristic	Calibration		Cross-validation			External validation		
				RMSEC	R^2	RMSECV	R^2	SDR	RMSEP	R^2	SDR
Model 13		6	Colour a*/b*	0.05	0.99	0.05	0.99	10.57	0.06	0.99	8.89
			TCI	2.25	0.98	2.49	0.98	7.17	1.75	0.99	11.65
			Firmness	1.85	0.93	2.24	0.90	3.10	1.44	0.97	4.83
			Lycopene	2.75	0.93	3.17	0.91	3.35	3.03	0.92	3.51
Model 14	MSC	6	Colour a*/b*	0.06	0.99	0.07	0.98	8.00			
			TCI	2.18	0.98	2.52	0.98	7.08			
			Firmness	1.70	0.94	2.00	0.92	3.48			
			Lycopene	2.99	0.92	3.44	0.90	3.09			
Model 15	1st derivative	6	Colour a*/b*	0.05	0.99	0.06	0.99	9.03			
			TCI	1.93	0.98	2.50	0.98	7.14			
			Firmness	2.24	0.89	2.60	0.86	2.67			
			Lycopene	1.99	0.96	2.77	0.93	3.84			

(a)



(b)



(c)

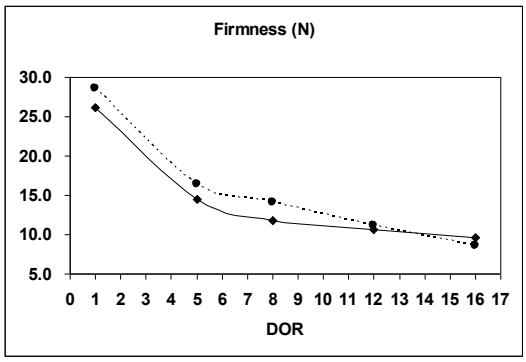


Figure 4.1: Changes of quality properties of tomatoes vs. day of ripening (DOR): (a) colour value a*/b* ratio; (b) tomato colour index; (c) firmness. The solid lines indicate tomatoes of cv. 'DRK 453'; the dotted lines indicate tomatoes of cv. 'Trust'.

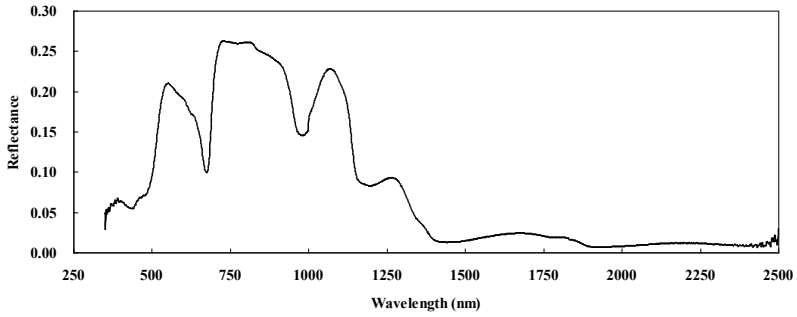


Figure 4.2: Original reflectance spectrum of one tomato (cv. DRK '453') at 1 day of ripening (350-2500nm).

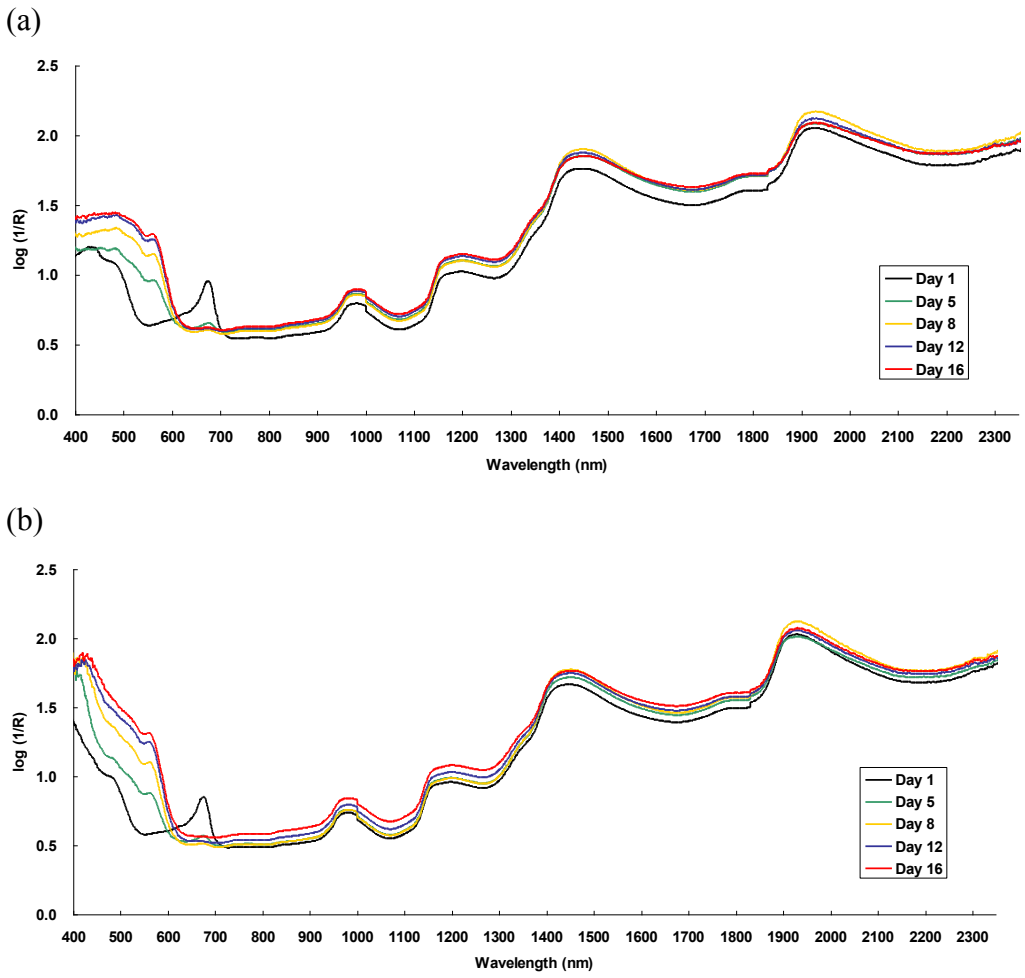
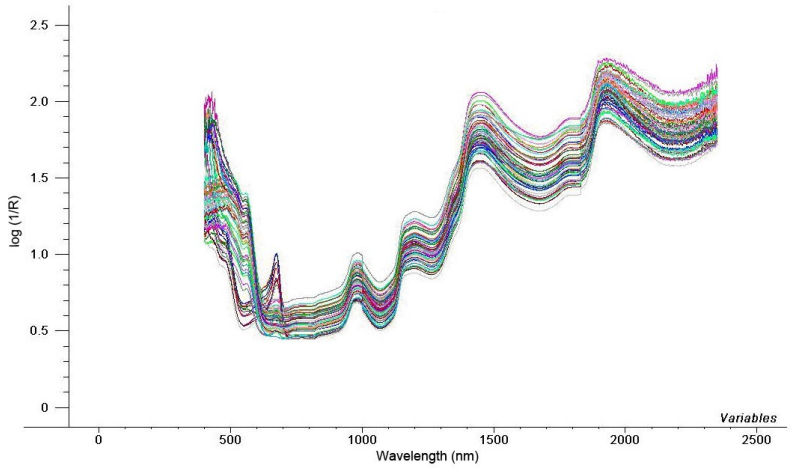
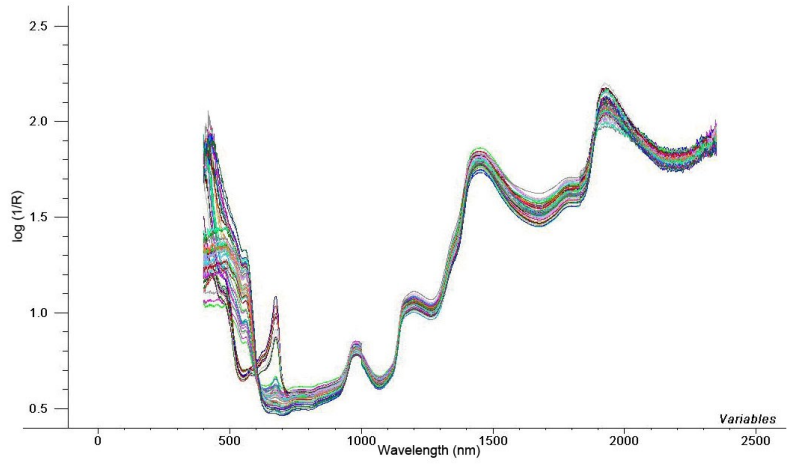


Figure 4.3: Absorbance spectra of tomatoes of two varieties measured at 1, 5, 8, 12 and 16 days of ripening (400-2350nm): (a) cv. 'DRK453' ; (b) cv. 'Trust'.

(a)



(b)



(c)

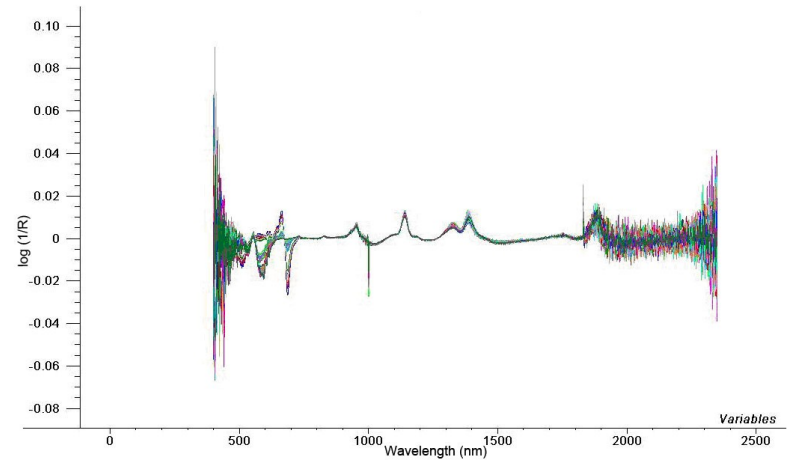


Figure 4.4: (a) Absorbance (log (1/R)) spectra of all tomatoes of calibration set. Preprocessed spectra by (b) MSC and (c) Savitzky-Golay first derivative.

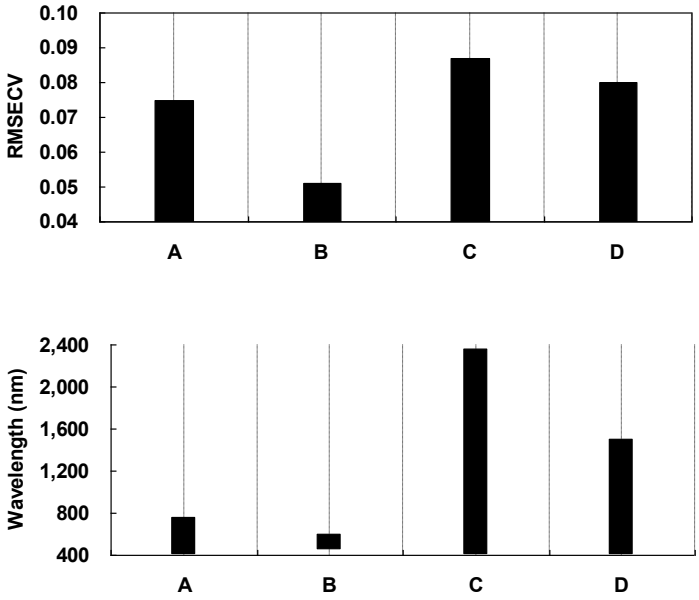
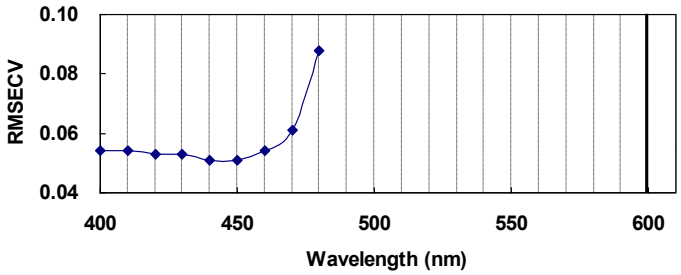


Figure 4.5: Root mean square error of full-cross validation (RMSECV) for colour value a*/b* ratio prediction vs. the spectral window. The vertical bars in the bottom chart indicate the wavelength range for each spectral window.

(a)



(b)

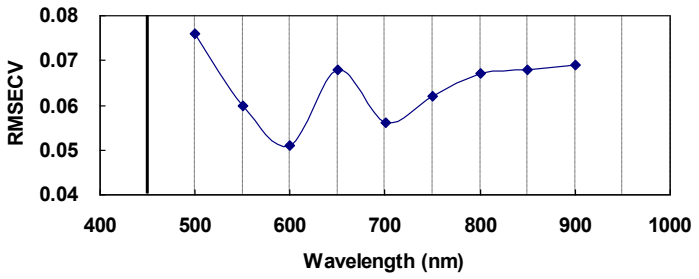
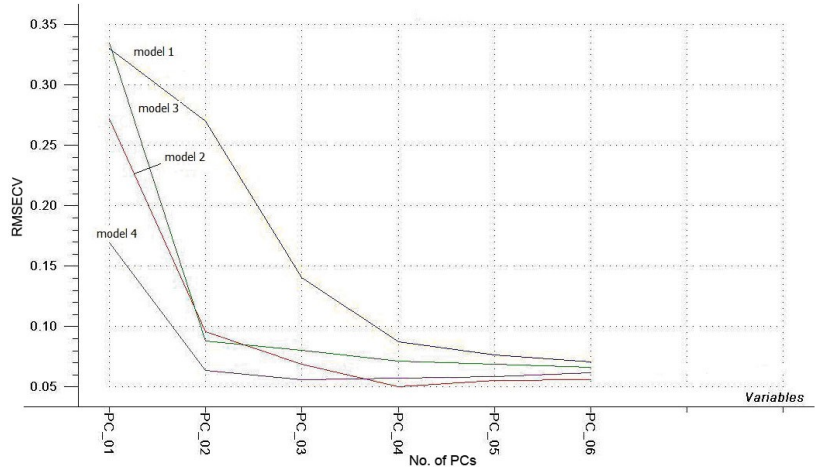
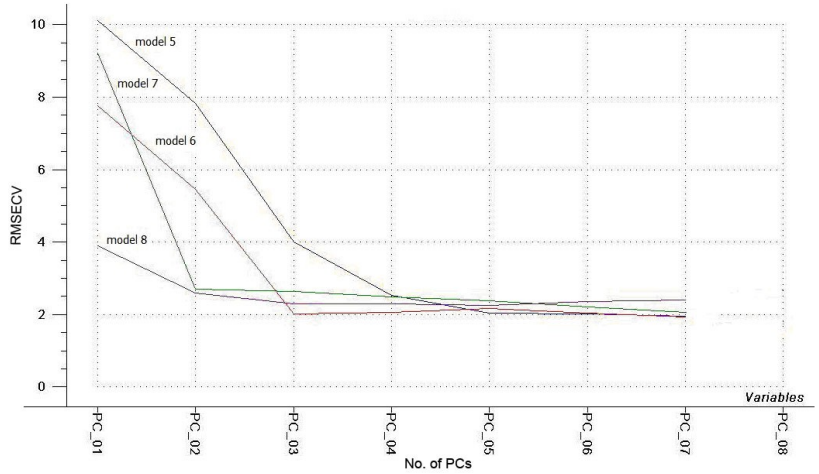


Figure 4.6: Root mean square error of full-cross validation (RMSECV) for colour value a*/b* ratio prediction vs. the spectral window: (a) the upper wavelength limit is fixed; (b) the lower wavelength limit is fixed. The bold vertical lines indicate the position for the fixed upper and lower limits in (a) and (b), respectively.

(a)



(b)



(c)

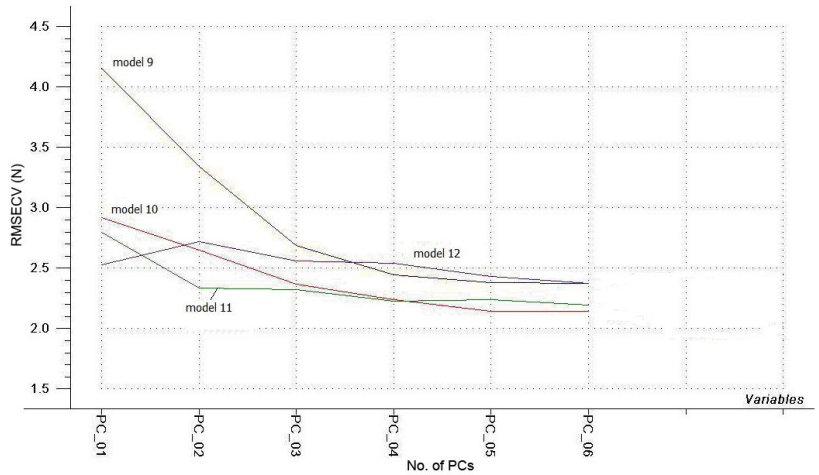
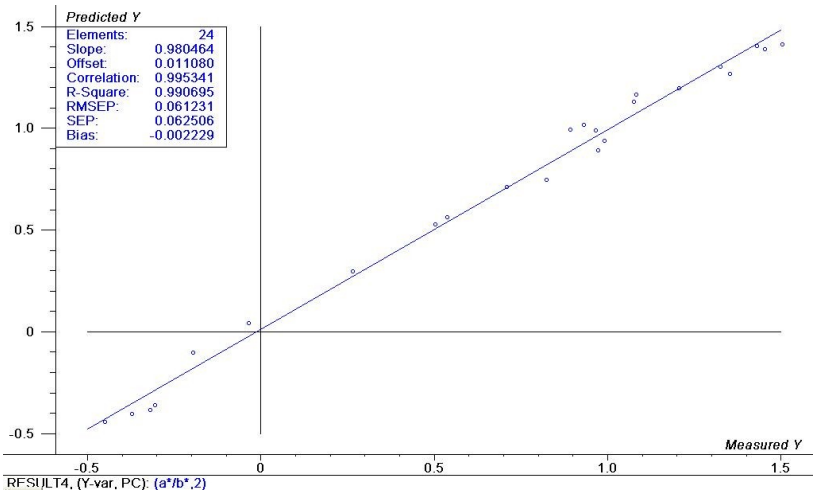
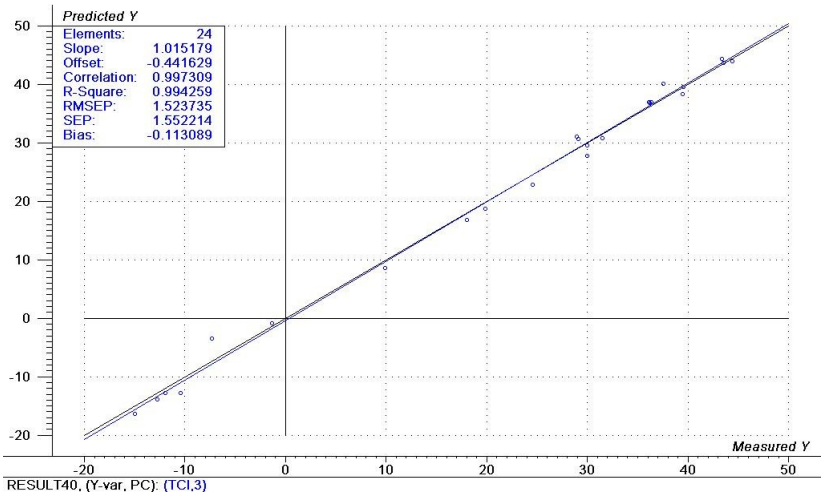


Figure 4.7: Root mean square error of full-cross validation (RMSECV) of models for each property vs. PLS components: (a) colour value a*/b* ratio; (b) tomato colour index; (c) firmness.

(a)



(b)



(c)

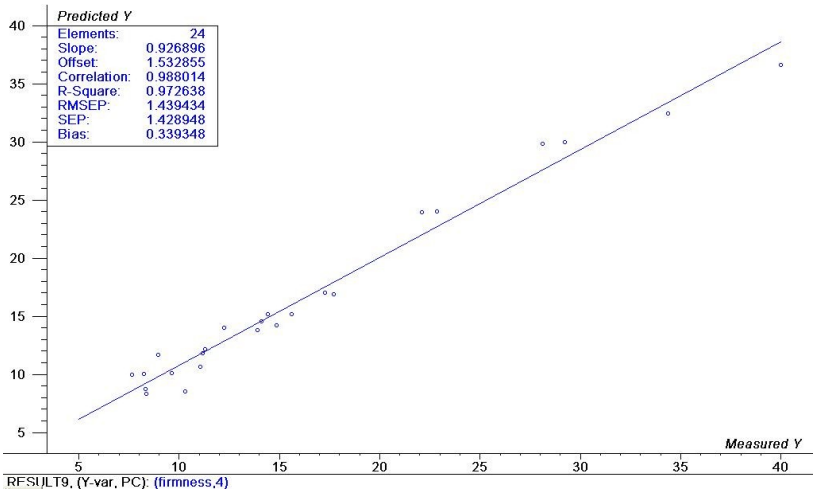


Figure 4.8: The predicted vs. the measured values of the properties of the validation set for the optimal models: (a) colour value a*/b* ratio; (b) tomato colour index; (c) firmness.

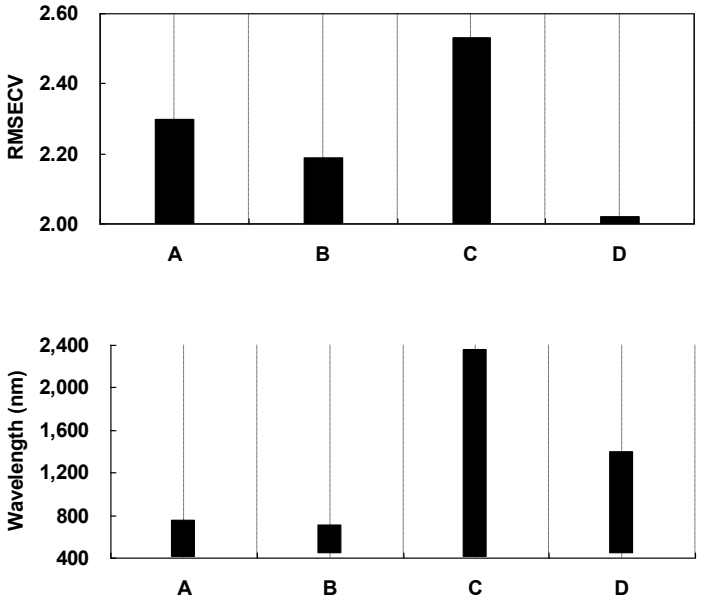
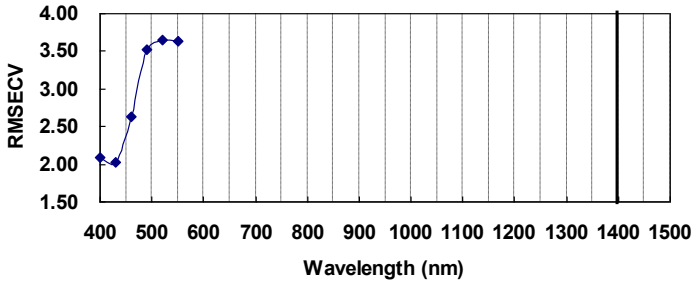


Figure 4.9: Root mean square error of full-cross validation (RMSECV) for tomato colour index prediction vs. the spectral window. The vertical bars in the bottom chart indicate the wavelength range for each spectral window.

(a)



(b)

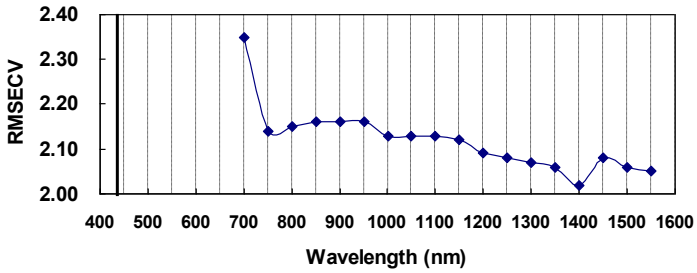


Figure 4.10: Root mean square error of full-cross validation (RMSECV) for tomato colour index prediction vs. the spectral window: (a) the upper wavelength limit is fixed; (b) the lower wavelength limit is fixed. The bold vertical lines indicate the position for the fixed upper and lower limits in (a) and (b), respectively.

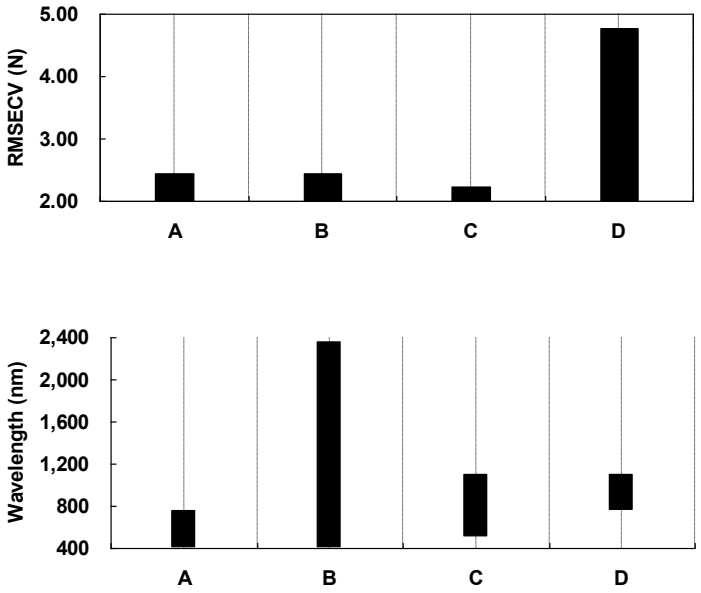
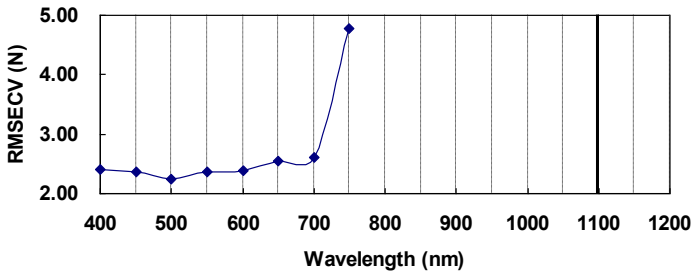


Figure 4.11: Root mean square error of full-cross validation (RMSECV) for firmness prediction vs. the spectral window. The vertical bars in the bottom chart indicate the wavelength range for each spectral window.

(a)



(b)

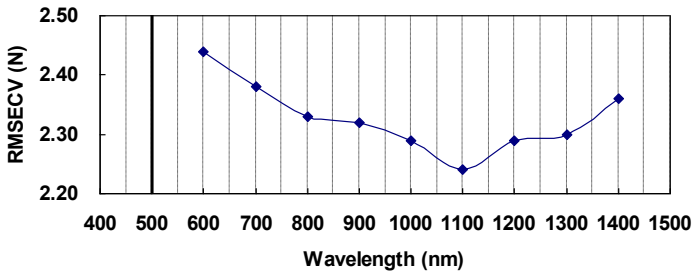


Figure 4.12: Root mean square error of full-cross validation (RMSECV) for firmness prediction vs. the spectral window: (a) the upper wavelength limit is fixed; (b) the lower wavelength limit is fixed. The bold vertical lines indicate the position for the fixed upper and lower limits in (a) and (b), respectively.

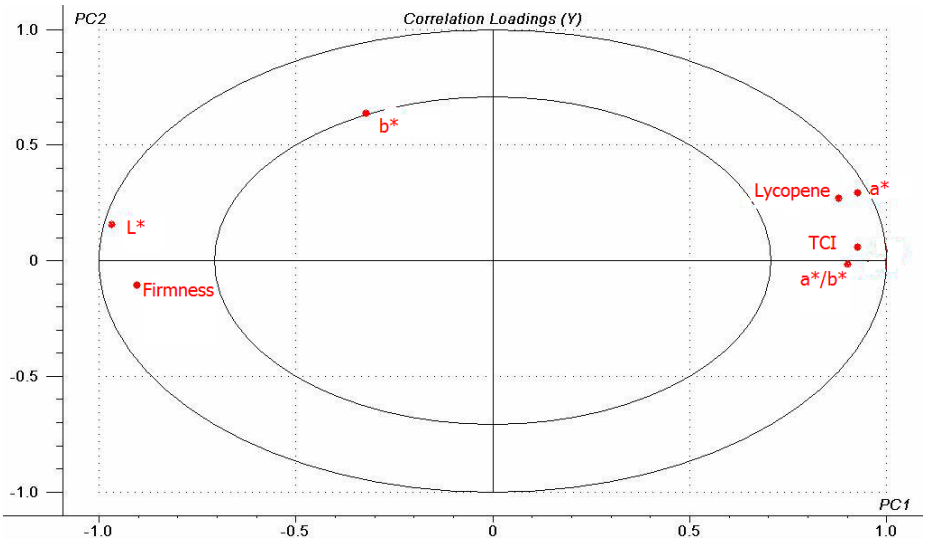
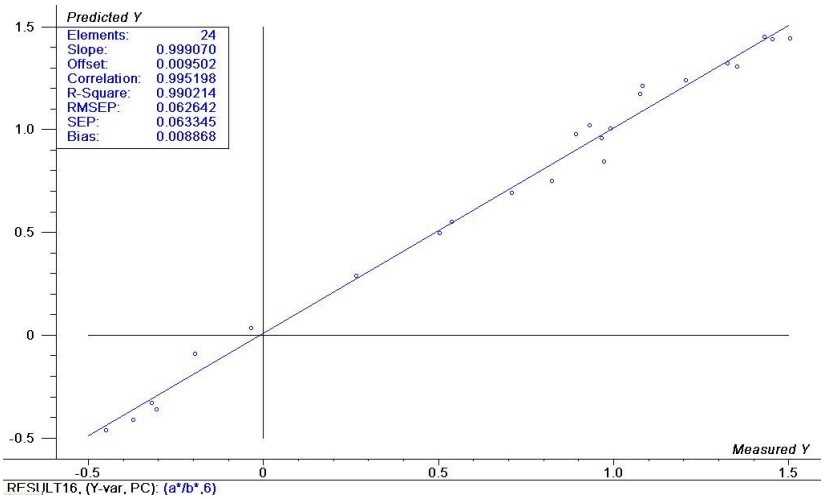
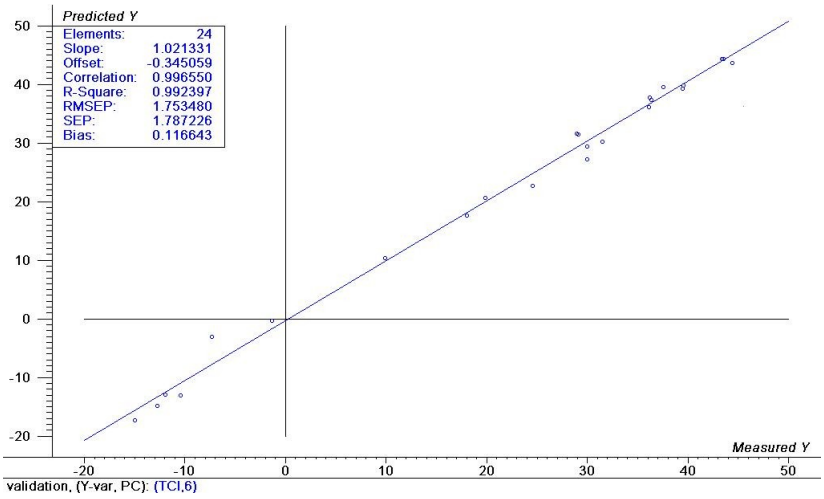


Figure 4.13: Correlation loadings of properties under study.

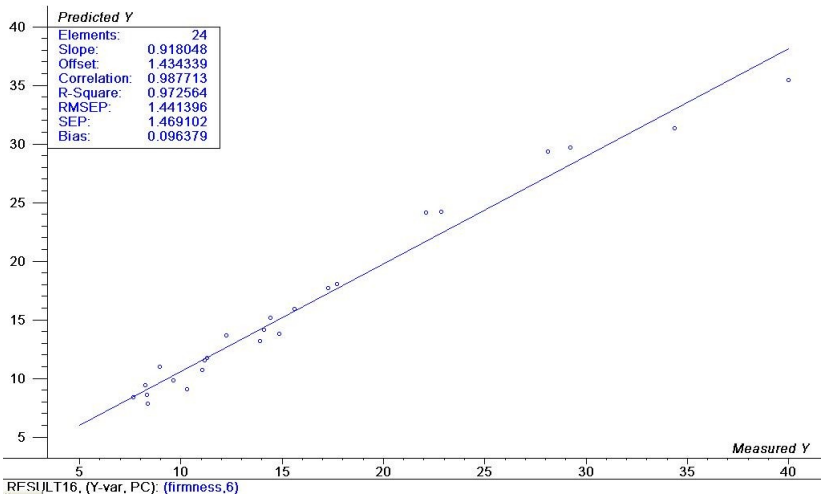
(a)



(b)



(c)



(d)

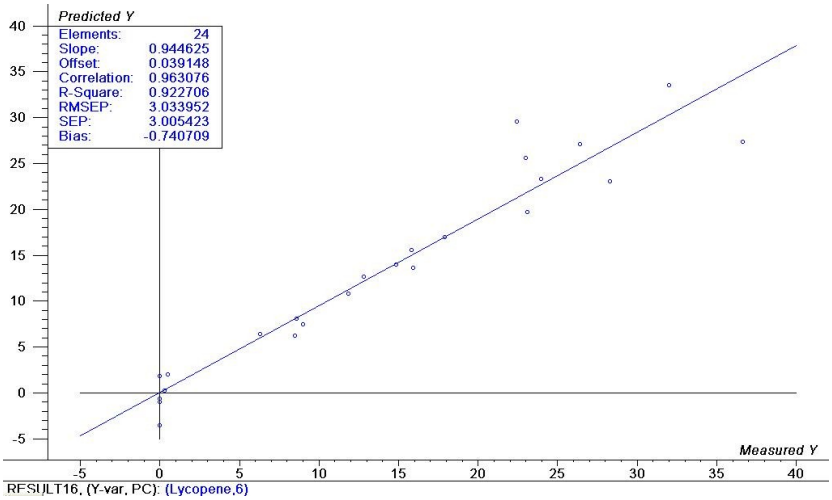


Figure 4.14: The predicted vs. the measured values of the properties of the validation set for the model 36 built using PLS2: (a) colour value a*/b* ratio; (b) tomato colour index; (c) firmness; (d) lycopene content.

CHAPTER V

GENERAL SUMMARY AND CONCLUSIONS

Tomatoes are one of the most widely produced and consumed fruits in the world. They are favored by many people mainly because they are low in fat, calories and cholesterol-free; rich in vitamins A and C. Additionally, tomatoes are also an important source of antioxidant - lycopene. Lycopene is known to have a potential protective effect against certain types of cancer and coronary heart disease. Visible/near-infrared (VIS/NIR) spectroscopy as a non-destructive analytical technique has been widely used with many agricultural products. The present study was undertaken to investigate the application of VIS/NIR spectroscopy for measuring quality attributes of tomatoes.

By means of partial least squares (PLS) regression method, calibration model based on VIS/NIR spectral reflectance measurements for each quality parameter of tomato fruits, including lycopene content, soluble solids content (SSC), titratable acidity (TA), acid-Brix ratio (ABR), colour value a*/b* ratio, tomato colour index (TCI), and firmness, was established. The model results indicate that it is possible to use this non-destructive technique to determine the lycopene content, a*/b* ratio, TCI, and firmness, however, accuracies of prediction for SSC, TA and ABR were not satisfactory.

Various wavelength ranges within 400-2350 nm and pre-processing methods methods including multiple scatter correction (MSC) and Savitzky-Golay first derivative were assessed to optimize the model for each parameter. The best spectral range for lycopene content, a*/b* ratio, TCI, and firmness was found within 450-1000 nm, 450-600 nm, 430-1400 nm, and 500-1100 nm, respectively. Except for a*/b* ratio, no pre-processing method improved the predictive ability.

The statistics of the best model for each parameter are as followed: for lycopene content, R^2=0.93 and RMSEP=2.87 mg kg^{-1} with 4 PCs; for TA, R^2=0.33 and RMSEP=0.51 mg ml^{-1} with 6 PCs; for SSC, R^2=0.03 and RMSEP=0.15 ºBrix with 7 PCs; for ABR, R^2=0.65 and RMSEP=0.077 with 4 PCs; for a*/b* ratio, R^2=0.99 and RMSEP=0.06 with 2 PCs; for TCI, R^2=0.99 and RMSEP=1.52 with 3 PCs; and for firmness, R^2=0.97 and RMSEP=1.44 N with 4 PCs.

A suitable PLS2 calibration model was obtained at the wavelength range of 450-1000 nm for simultaneously calibrating colour, firmness and lycopene content of the tomato. The R^2 values of the PLS2 model with 6 PCs for a*/b*, TCI, firmness and lycopene content were 0.99, 0.99, 0.97, and 0.92, respectively, while its RMSEP values were 0.06, 1.75, 1.44 N, and 3.03 mg kg^{-1}.

The study was conducted on two varieties of tomatoes (cv. 'DRK 453' and 'Trust') and results may not be applicable to other varieties. In a further study, a wider range of tomato varieties can be investigated to make the models more robust. Additionally, different measurement setups such as transmittance mode should be test to compare with the present measuring mode (reflectance mode).

REFERENCES

Abbott, J.A., 1999. Quality measurement of fruits and vegetables. *Post-harvest Biol. Technol.* 15: 207–225.

Abushita, A.A., Hebshi, E.A., Daood, H. G. and Biacs, P. A., 1997. Determination of antioxydant vitamins in tomatoes. *Food Chem.* 60: 207–212.

Adsule, P.G. and Dan, A., 1979. Simplified extraction procedure in the rapid spectrophotometric method for lycopene estimation in tomato. *J. Food Sci. Technol.* 16: 216.

Baranska, M., Schutze, W. and Schulz, H., 2006. Determination of Lycopene and *β*-Carotene Content in Tomato Fruits and Related Products: Comparison of FT-Raman, ATR-IR, and NIR Spectroscopy. Anal. Chem. 78: 8456-8461.

Baranski, R., Baranska, M. and Schulz, H., 2005. Changes in carotenoid content and distribution in living plant tissue can be observed and mapped in situ using NIR-FT-Raman spectroscopy. *Planta*. 222: 448-457.

Batu, A., 2004. Determination of acceptable firmness and colour values of tomatoes. *Journal of Food Engeering,* 61: 471–475.

Bramley, P.M., 2000. Is lycopene beneficial to human health? *Phytochemistry*. 54:233-236.

Cayuela, J.A., 2008. Vis/NIR soluble solids prediction in intact oranges (*Citrus sinensis* L.) cv. Valencia Late by reflectance. *Postharvest Biology and Technology*. 47: 75–80

Chen, H. and De Baerdemaeker, J., 1990. Resonance frequency and firmness of tomatoes during ripening. Proceedings of the 22nd International Conference on Agricultural Mechanization, Zaragoza, Spain, March 27-30, 1: 61-68.

Chen. P., McCarthy, M.J. and Kauten, R., 1989. NMR for internal quality evaluation of fruits and vegetables. *Trans. ASAE*, 32: 1741-1753.

Cho, R.K., Sohn, M.R. and Kwon, Y.K., 1998. New observation of nondestructive evaluation for sweetness in apple fruit using near infrared spectroscopy. *J. Near Infrared Spectrosc.* 6: A75–A78.

Clark, C.J., McGlone, V.A. and Jordan, R.B., 2003. Detection of brownheart in 'Braeburn' apple by transmission NIR spectroscopy. *Postharvest Biol. Technol.* 28: 87–96.

Clément, A., Dorais, M. and Vernon, M., 2008. Multivariate Approach to the Measurement of Tomato Maturity and Gustatory Attributes and Their Rapid Assessment by Vis-NIR Spectroscopy. *J. Agric. Food Chem.* 56: 1538–1544.

Clinton, S.K., 1998. Lycopene: chemistry, biology, and implications for human health and disease. *Nutrition Reviews*. 56:35–51.

Davies, A.M.C., 2000. William Herschel and the discovery of near infrared. *Spectrosc. Eur.* 12: 10–16.

Davies, J. N. and Hobson, G. E., 1981. The constituents of tomato fruit-the influence of environment, nutrition, and genotype. *Critical Reviews in Food Science and Nutrition.* 15 (3): 205–80.

De Jong, S., 1993. PLS fits closer than PCR. *J. Chemometr.* 7: 551-557.

De Stefani, E., Oreggia, F., Boffetta, P., Deneo-Pellegrini, H., Ronco, A. and Mendilaharsu, M., 2000. Tomatoes, tomato-rich foods, lycopene and cancer of the upper respiratory tract: A case control in Uruguay. *Oral Oncology* 36:47–53.

Dorais, M., Papadopoulos, A. P. and Gosselin, A., 2001. Greenhouse tomato fruit quality. *Hortic. Re*V. 26: 239–319.

FAOSTAT online, http:// faostat.fao.org/default.aspx

Fish, W.W., Perkins-Veazie, P. and Collins, J.K., 2002. A quantitative assay for lycopene that utilizes reduced volumes of organic solvents. *J. Food Comp. Anal.* 15: 309-317.

Fisher, R.L. and Bennett, A.B., 1991. Role of cell wall hydrolases in fruit ripening. *Annual Review of Plant Physiology and Plant Molecular Biology*. 42: 675-703

Fraser, D.G., Kunnemeyer, R., McGlone, V.A., and Jordan, R.B., 2001. Near infra-red (NIR) light penetration into an apple. *Postharvest Biol. Technol.* 22: 191-194.

Fraser, D.G., Jordan, R.B., Kunnemeyer, R. and McGlone, V.A., 2003. Light distribution inside mandarin fruit during internal quality assessment by NIR spectroscopy. *Postharvest Biol. Technol.* 27: 185–196.

Garrett, A., Ammerman, G., Desrosier, N., and Fields, M., 1960. Effect of color on marketing of fresh tomatoes. *J. Am. Soc. Hortic. Sci.* 76: 555-559.

Gerster, H., 1997. The potential role of lycopene for human health. *Journal of the American College of Nutrition*. 16: 109-126.

Gómez, A.H., Hu, G., Wang, J. and Pereira, A.G., 2006. Evaluation of tomato maturity by electronic nose. *Computers and Electronics in Agriculture.* 54: 44–52.

Gómez, R., Costa, J., Amo, M., Alvarruiz, A., Picazo, M. and Pardo, J.E., 2001 Physicochemical and colorimetric evaluation of local varieties of tomato grown in SE Spain. *J. Sci. Food Agric.* 81: 1101–1105.

Gormley, L. and Egan, S., 1978. Firmness and colour of fruit at some tomato cultivars from various sources during storage. *Journal of Science of Food and Agriculture.* 29: 534–538.

Goula, A.M. and Adamopoulos, K.G., 2003. Estimating the composition of tomato juice products by near infrared spectroscopy. *J. Near Infrared Spectrosc.* 11: 123-136.

Gould, W.A., 1992. Tomato Production, Processing & Technology, 3rd ed.; CTI Pub. Inc.: Baltimore.

Greensill, C.V., Wolf, P.J., Spiegelman, C.H. and Walsh, K.B., 2001. Calibration transfer between PDA-based NIR spectrometers in the NIR assessment of melon soluble solids content. *Appl. Spectrosc.* 55: 647–653.

Hernandez Sanchez, N., Lurol, S., Roger, J.M. and Bellon-Maurel,V., 2003. Robustness of models based on NIR spectra for sugar content prediction in apples. *J. Near Infrared Spectrosc.* 11: 97–107.

Hobson, G.E. and Davies, J.N., 1971. The tomato. In: The Biochemistry of Fruits and Their Products. Hulme, A.C., Ed. New York: Academic Press, Vol. 2, pp. 437-482.

Hobson, G.E. and Bedford, L., 1989. The composition of cherry tomatoes and its relation to consumer acceptability. *J. Horticultural Sci.* 64(3): 321–329.

Hong, T.L. and Tsou, S.C.S., 1998. Determination of tomato quality by near infrared spectroscopy. *J. Near Infrared Spectrosc.*6: A321-A324.

Ishida, N., Kobayashi, T., Koizumi, M. and Kano, H., 1989. H-NMR imaging of tomato fruits. *Agric. Biol. Chem.* 53: 2363-2367.

Jha, S.N. and Matsuoka, T., 2004. Non-destructive determination of acid–brix ratio of tomato juice using near infrared spectroscopy. *International Journal of Food Science and Technology.* 39: 425–430.

Jones, R.A. and Scott, S.J., 1983. Improvement of tomato flavor by genetically increasing

sugar and acid contents. *Euphytica.* 32: 845–855.

Kawano, S., Fujiwara, T. and Iwamoto, M., 1993. Nondestructive determination of sugar content in Satsuma mandarin using near infrared (NIR) transmittance. *J. Jpn. Soc. Hortic. Sci.* 62: 465–470.

Khuriyati, N., Matsuoka, T. and Kawano, S., 2004. Precise near infrared spectral acquisition of intact tomatoes in interactance mode. *J. Near Infrared Spectrosc.* 12: 391–395.

Kim, S-M., McCarthy, M.J. and Chen, P., 1994. Feasibility of tomato quality grading and sorting using magnetic resonance. Paper No. 9446519. ASAE Conf., Atlanta, GA.

Lai, A., Santangelo, E., Soressi, G.P. and Fantoni, R., 2007. Analysis of the main secondary metabolites produced in tomato (*Lycopersicon esculentum*, Mill.) epicarp tissue during fruit ripening using fluorescence techniques. *Postharvest Biology and Technology* 43: 335–342.

Lammertyn, J., Nicolaï, B., Ooms, K., De Smedt, V. And De Baerdemaeker, J., 1998. Non-destructive measurement of acidity, soluble solids, and firmness of Jonagold apples using NIR spectroscopy. *Trans. ASAE.* 41: 1089–1094.

Lammertyn, J., Peirs, A., De Baerdemaeker, J. and Nicolaï, B., 2000. Light penetration properties of NIR radiation in fruit with respect to non-destructive quality assessment. *Postharvest Biol. Technol.* 18: 121–132.

Liu, Y. and Ying, Y., 2005. Use of FT-NIR spectrometry in non-invasive measurements of internal quality of 'Fuji' apples. *Postharvest Biol. Technol.* 37: 65–71.

Liu, Y., Ying, Y., Yu, H. and Fu, X., 2006. Comparison of the HPLC method and FT-NIR analysis for quantification of glucose, fructose and sucrose in intact apple fruits. *J. Agric. Food Chem.* 54: 2810–2815.

Long, R.L.,Walsh, K.B., 2006. Limitations to the measurement of intact melon total soluble solids using near infrared spectroscopy. Aust. J. Agric. Res. 57: 403–410.

López, M. D., Contreras, M. and Fernández-Alba, A. R., 2003. Time evolution of tomato quality parameters versus storage conditions. *Acta Hortic.* 604: 619–624.

Lu, R., Guyer, D.E. and Beaudry, R.M., 2000. Determination of firmness and sugar content of apples using near-infrared diffuse reflectance. *Journal of Texture Studies.* 31: 615-630.

Lu, R., 2001. Predicting firmness and sugar content of sweet cherries using near-infrared

diffuse reflectance spectroscopy. *Trans. ASAE.* 44(5): 1265–1271.

Mangels, A.R., Holden, J.M., Beecher, G.R., Forman, M.R., and Lanza, E. 1993. Carotenoids in fruits and vegetables: an evaluation of analytic data. *Journal American Dietetic Association.* 93: 284–296.

Manson, J.E., Gaziano, J.M., Jonas, M.A. and Hennekens, C.H., 1993. Antioxidants and cardiovascular disease: A review. *Journal of the American College of Nutrition* 12:426.

Martens, H. and Stark, E., 1991. Extended multiplicative signal correction and spectral interference subtraction-new preprocessing methods for near-infrared spectroscopy. *J. Pharm. Biomed. Anal.* 9: 625–635.

Martens, H. and Næs, T., 1989. Multivariate Calibration. John Wiley & Sons Ltd., Chichester, Great Britain. pp. 345-349.

McGlone, V.A., Abe, H. and Kawano, S., 1997. Kiwifruit firmness by near infrared light scattering. *J. Near Infrared Spectrosc.* 5: 83–89.

McGlone, V.A. and Kawano, S., 1998. Firmness, dry-matter and soluble-solids assessment of postharvest kiwifruit by NIR-spectroscopy. *Postharvest Biol. Technol.* 13: 131–141.

McGlone, V.A., Jordan, R.B. and Martinsen, P.J., 2002a. Vis/NIR estimation at harvest of pre- and post-storage quality indices for 'Royal Gala' apple. *Postharvest Biol. Technol.* 25: 135–144.

McGlone,V.A., Jordan, R.B., Seelye, R., and Martinsen, P.J., 2002b. Comparing density and NIR methods for measurement of Kiwifruit dry matter and soluble solids content. *Postharvest Biol. Technol.* 26: 191–198.

McGlone, V.A., Fraser, D., Jordan, R.B. and Kunnemeyer, R., 2003a. Internal quality assessment of mandarin fruit by vis/NIR spectroscopy. *J. Near Infrared Spectrosc.* 11: 323–332.

McGlone, V.A., Jordan, R.B., Seelye, R. and Clark, C.J., 2003b. Dry-matter—a better predictor of the post-storage soluble solids in apples? *Postharvest Biol. Technol.* 28: 431–435.

McGlone, V.A. and Martinsen, P.J., 2004. Transmission measurements on intact apples moving at high speed. *J. Near Infrared Spectrosc.* 12: 37–43.

McGlone, V.A., Martinsen, P.J., Clark, C.J., Jordan, R.B., 2005. On-line detection of

Brownheart in Braeburn apples using near infrared transmission measurements. *Postharvest Biol. Technol.* 37: 142–151.

McGlone, V.A., Clark, C.J. and Jordan, R.B., 2007. Comparing density and VNIR methods for predicting quality parameters of yellow-fleshed kiwifruit (*Actinidia chinensis*) *Postharvest Biol. Technol.* 46: 1–9.

Miller, W.M. and Zude-Sasse, M., 2004. NIR-based sensing to measure soluble solids content of florida citrus. *Appl. Eng. Agric.* 20: 321–327.

Moons, E., Dubois, A., Dardenne, P. And Sindic, M., 1997. Nondestructive visible and NIR spectroscopy for the determination of internal quality in apple. Proceedings from the Sensors for Non-destructive Testing International Conference and Tour, 18–21 February 1997, Orlando, FL. NRAES (Northeast Reg. Agric. Eng. Serv.), Ithaca, NY, pp. 122–132.

Naes, T., Isaksson, T., Fearn, T. and Davies, T., 2004. A User-friendly Guide to Multivariate Calibration and Classification. NIR publications, Charlton, Chichester, UK.

Nicolaï, B.M., Theron, K.I. and Lammertyn, J., 2006. Kernel PLS regression on wavelet transformed NIR spectra for prediction of sugar content of apple. *Chemom. Intell. Lab. Syst.* 85: 243–252.

Nicolaï, B.M., Beullens, K., Bobelyn, E., Peirs, A., Saeys, W., Theron, K.I., and Lammertyn, J., 2007. Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. *Postharvest Biology and Technology.* 46: 99–118.

Park, B., Abbott, J.A., Lee, K.J., Choi, C.H. and Choi, K.H., 2003. Near-infrared diffuse reflectance for quantitative and qualitative measurement of soluble solids and firmness of Delicious and Gala apples. *Trans. ASAE*. 46: 1721–1731.

Pech, J.C., Latche. A., Andrieu. M.H., Raynal. J. and Pradere, J., 1990. In vivo monitoring of fruit ripening by proton NMR imaging. Abst. XXIII Int. Hort. Congress. Florence Italy. 1: 635.

Pedro, A.M.K. and Ferreira, M.M.C., 2005. Nondestructive determination of solids and carotenoids in tomato products by near-infrared spectroscopy and multivariate calibration. *Analytical Chemistry.* 77: 2505-2511.

Pedro, A.M.K. and Ferreira, M.M.C., 2007. Simultaneously calibrating solids, sugars and

acidity of tomato products using PLS2 and NIR spectroscopy. *Analytica Chimica Acta.* 595: 221–227.

Peiris, K.H.S., Dull, G.G., Leffler, R.G., Kays, S.J., 1998. Near-infrared (NIR) spectrometric technique for nondestructive determination of soluble solids content in processing tomatoes. *American Society for Horticultural Science.* 123(6): 1089-1093.

Peirs, A., Lammertyn, J., Nicolaï, B. and De Baerdemaeker, J., 2000. Non-destructive quality measurements of apples by means of NIR-spectroscopy. *Acta Hort.* 517:435-440.

Peirs, A., Ooms, K., Lammertyn, J. and Nicolaï, B.M., 2001. Prediction of the optimal picking date of different apple cultivars by means of VIS/NIR-spectroscopy. *Postharvest Biol. Technol.* 21: 189–199.

Peirs, A., Scheerlinck, N., Touchant, K. and Nicolaï, B.M., 2002. Comparison of Fourier transform and dispersive near infrared reflectance spectroscopy for apple quality measurements. *Biosyst. Eng.* 81 (3): 305–311.

Peirs, A., Scheerlinck, N. and Nicolaï, B.M., 2003a. Temperature compensation for near infrared reflectance measurement of apple fruit soluble solids contents. *Postharvest Biol. Technol.* 30: 233–248.

Peirs, A., Scheerlinck, N., De Baerdemaeker, J. and Nicolaï, B.M., 2003b. Starch index determination of apple fruit by means of a hyperspectral near infrared reflectance imaging system. *J. Near Infrared Spectrosc.* 11: 379–389.

Peirs, A., Schenk, A. and Nicolaï, B.M., 2005. Effect of natural variability among apples on the accuracy of VIS-NIR calibration models for optimal harvest date predictions. *Postharvest Biol. Technol.* 35: 1–13.

Raffo, A., Leonardi, C., Fogliano, V., Ambrosino, P., Salucci, M., Gennaro, L., Bugianesi, R., Giuffrida, F. and Quaglia, G., 2002. Nutritional value of cherry tomatoes (*Lycopersicon esculentum* cv. Naomi F1) harvested at different ripening stages. *J. Agric. Food Chem.* 50: 6550–6556.

Rao, A.V. and Agarwal, S., 1999. Role of lycopene as anti-oxidant carotenoid in the prevention of chronic diseases: a review. *Nutrition Research.* 19: 305-323.

Renquist, A.R. and Reid, J.B., 1998. Quality of processing tomato (*Lycoperscion esculentum*) fruit from four bloom dates in relation to optimal harvest timing. *N. Z. J. Crop Hortic.*

Sci. 26: 161–168.

Richardson, C. and Hobson, G.E., 1987. Compositional changes in normal and mutant tomato fruit during ripening and storage. *Journal of the science of food and agriculture.* 40: 245–252.

Saltveit, M.E., 1991. Determining tomato fruit maturity with non-destructive in vivo nuclear magnetic resonance imaging. *Postharv. Biol. Technol.* 1: 153-159.

Saltveit, M.E., 2005. Fruit ripening and fruit quality. In: Hervelink, E. (ed.), *Tomatoes*, CAB International, Wallingford, UK, pp. 145-170.

Saranwong, S., Sornsrivichai, J. and Kawano, S., 2001. Improvement of PLS calibration for Brix value and dry matter of mango using information from MLR calibration. *J. Near Infrared Spectrosc.* 9: 287–295.

Saranwong, S., Sornsrivichai, J. and Kawano, S., 2003a. Performance of a portable near infrared instrument for Brix value determination of intact mango fruit. *J. Near Infrared Spectrosc.* 11: 175–181.

Saranwong, S., Sornsrivichai, J. and Kawano, S., 2003b. On-tree evaluation of harvesting quality of mango fruit using a hand-held NIR instrument. *J. Near Infrared Spectrosc.* 11: 283–293.

Schmilovitch, Z., Mizrach, A., Hoffman, A., Egozi, H., and Fuchs, Y., 2000. Determination of mango physiological indices by near-infrared spectrometry. *Postharvest Biol. Technol.* 19: 245–252.

Schotte, S., De Belie, N. and De Baerdemaeker, J., 1999. Acoustic impulse-response technique for evaluation and modelling of firmness of tomato fruit. *Postharvest Biology and Technology.* 17: 105–115.

Shao, Y., He, Y., Gómez, A.H., Pereir A.G., Qiu, Z. and Zhang, Y., 2007. Visible/near infrared spectrometric technique for nondestructive assessment of tomato 'Heatwave' (Lycopersicum esculentum) quality characteristics. *Journal of Food Engineering.* 81: 672–678

Simandle, P.A., Brogdon, J.L., Sweeney, J.P., Mobley, E.O., and Davis, D.W. 1966. Quality of six tomato varieties as affected by some compositional factors. *P. Amer. Soc. Hort. Sci.* 89: 532–538.

Sjoblom, J., Svensson, O., Josefson, M., Kullberg, H., and Wold, S., 1998. An evaluation of orthogonal signal correction applied to calibration transfer of near infrared spectra. *Chemometrics and Intelligent Laboratory Systems.* 44: 229-244.

Slaughter, D.C., 1995. Non-destructive determination of internal quality in peaches and nectarines. *Trans. ASAE* 38: 617–623.

Slaughter, D.C., Barrett, D. and Boersig, M., 1996. Nondestructive Determination of Soluble Solids in Tomatoes using Near Infrared Spectroscopy. *Journal of Food Science.* 61 (4): 695-697.

Stevens, M.A., Kader, A.A. and Albright, M., 1979. Potential for increasing tomato flavor via increased sugar and acid content. *J. Amer. Soc. Hort. Sci.* 104(1): 40–42.

Stroshine, R.L., Cho, S.I., Wai, W-K., Krutz, G.W. and Baianu, I.C., 1991. Magnetic resonance sensing of fruit firmness and ripeness. Paper No. 91-6565. ASAE Conf., Chicago, IL.

Temma, T., Hanamatsu, K. and Shinoki, F., 2002a. Development of a portable near infrared sugar-measuring instrument. *J. Near Infrared Spectrosc.* 10: 77–83.

Temma, T., Hanamatsu, K. and Shinoki, F., 2002b. Measuring the sugar content of apples and apple juice by near infrared spectroscopy. Opt. Rev. 9: 40–44.

Tijskens, L.M.M. & Evelo, R.G., 1994. Modelling colour of tomatoes during postharvest storage. *Postharvest Biology and Technology.* 4: 85-98.

Tu, K., Jancsók, P., Nicolaï, B. and De Baerdemaeker, J., 2000. Use of laser-scattering imaging to study tomato-fruit quality in relation to acoustic and compression measurements. *International Journal of Food Science & Technology.* 35: 503-510.

Ventura, M., De Jager, A., De Putter, H. and Roelofs, F.P.M.M., 1998. Non-destructive determination of soluble solids in apple fruit by near infrared spectroscopy (NIRS). *Postharvest Biol. Technol.* 14: 21–27.

Walsh, K.B., Golic, M. and Greensill, C.V., 2004. Sorting of fruit using near infrared spectroscopy: application to a range of fruit and vegetables for soluble solids and dry matter content. *J. Near Infrared Spectrosc.* 12: 141–148.

Walsh, K.B., Guthrie, J.A. and Burney, J.W., 2000. Application of commercially available, low-cost, miniaturised NIR spectrometers to the assessment of the sugar content of intact fruit. *Aust. J. Plant Physiol.* 27: 1175–1186.

Westad, F. and Marten, H., 2000. Variable selection in near infrared spectroscopy based on significance testing in partial least squares regression. *J. Near Infrared Spectrosc*. 8 (2): 117–124.

Williams, P. and Norris, K.H., 2001. Variable affecting near infrared spectroscopic analysis. In: Williams, P., Norris, K.H. (Eds.), Near infrared Technology in the Agriculture and Food Industries, second ed. The American Association of Cereal Chemists, St. Paul, MN, pp. 171–185.

Wold, S., Sjostrom, M. and Eriksson, L., 2001. PLS-regression: a basic tool of chemometrics. *Chemom. Intell. Lab. Syst.* 58, 109–130.

World Processing Tomato Council (WPTC) online, http://www.wptc.to/index.aspx 05/02/2008

Xing, J., Bravo, C., Moshou, D., Ramon, H. and De Baerdemaeker, J., 2006. Bruise detection on 'Golden Delicious' apples by vis/NIR spectroscopy. *Comp. Electron. Agric.* 52: 11–20.

Xing, J. and De Baerdemaeker, J., 2007. Fresh bruise detection by predicting softening index of apple tissue using VIS/NIR spectroscopy. *Postharvest Biology and Technology.* 45: 176–183.

Ziegler, R.G., 1991. Vegetables, fruits, and carotenoids and the risk of cancer. *American Journal of Clinical Nutrition.* 53:S251–S259.

Zude, M., Herold, B., Roger, J.M., Bellon-Maurel, V. and Landahl, S., 2006. Nondestructive tests on the prediction of apple fruit flesh firmness and soluble solids content on tree and in shelf life. *J. Food Eng.* 77: 254–260.