CHICKEN EGG QUALITY ASSESSMENT FROM VISIBLE/NEAR INFRARED OBSERVATIONS

By

Nicolas Abdel-Nour

Department of Bioresource Engineering

McGill University

Montreal, Quebec, Canada

A thesis submitted to the Graduate and Postdoctoral Office in partial fulfillment of the requirements for the degree of Master of Science

October 2008

ABSTRACT

Egg is a fragile component within the human diet. Important changes occur in egg during storage. Prediction of these changes is critical in order to grade the eggs upon their quality and freshness. The objectives of this study were to evaluate the application of visible and near infrared spectroscopy as a non-destructive method for the assessment of egg quality and freshness. Therefore, visible and near infrared transmittance spectral data ranging from 350 to 2500 nm was collected with the help of a radiospectrometer on 360 freshly laid eggs. A partial least squares model was built in order to link the spectral data with the most widely used destructive methods, namely Haugh Units and albumen pH in terms of egg quality and the number of storage days in terms of egg freshness.

The ability of maximum R^2 method to select the relevant wavelengths in order to build a partial least squares (PLS) predictive model was investigated in the first part of the study. The results showed that this method improved the predictive ability of the model. Coefficient of determination (R^2) and root mean square error of cross validation (RMSECV) were calculated in order to select sets of wavelengths to build the model with the best predictive ability.

The second part of the study was based on building calibration models for predicting egg freshness in terms of number of storage day and egg quality in terms of Haugh Units and albumen pH. The results showed that the models had good predictive ability and R^2 for number of storage days, Haugh Units and albumen pH were 0.89, 0.79 and 0.90, respectively. RMSECV for these three parameters were 1.65, 5.05 and 0.06, respectively.

RÉSUMÉ

L'oeuf est un composant fragile dans le regime alimentaire humain. Des changements importants arrivent dans loeuf pendant le stockage. La prediction de ces changements eat ctitique pour classer les oeufs selon leur qualité et leur fraîcheur. Les objectifs de cette étude étaient d'évaluer l'application méthode basée sur la spectroscopie visible et infra-rouge proche comme une method non destructive pour l'évaluation de la qualité et la fraîcheur des oeufs. Donc, la transmission visible et infra rouge proche des données spectrales aux limites de 350 à 2500 nm ont été exécutées à l'aide d'un radiosectromètre sur 360 oeufs récemment pondus. Un modèle des moindres carrées partiels (MCP) a été construit afin de lier les données soectrakes avec les méthodes destructives les plus utilisées, à savoir Unité de Haugh at le pH d'albumen en termes de qualité d'oeufs et le nombre de jours de stoclage en termees de fraîcheur d'oeufs.

La première étude a traité de la capacité de la méthode maximum R^2 à choisir les longueurs d'onde appropriées afin d'établir un modèle des moindres carrés partiels (MCP). Les résultats ont révélé combien cette méthode a été un bon outil dans le choix des longueurs d'onde instructives et dans l'amélioration de la capacité prédictive du modèle. Le coefficient de détermination (R^2) et les erreurs de la racine carrée moyenne (ERCM) ont été calculés afin de choisir des ensembles de longueurs d'onde, lesquels aident le mieux à construire le modèle qui possède la meilleure capacité prédictive.

La seconde étude a visé l'établissement des modèles prédictifs de la fraîcheur d'oeufs en fonction du nombre des jours de stockage et de leur qualité en fonction de l'unité de Haugh (UH) et du pH de l'albumen. Les résultats ont prouvé que les modèles construits ont eu une bonne capacité predictive: pour le nombre de jours de stockage a été 0.89, l'unité Haugh a été 0.79 et le pH d'albumen a été 0.90. A noter que les ERCM pour ces trois paramètres ont été 1.65, 5.05 et 0.06, respectivement.

AKNOWLEDGMENTS

I would like to express my most and special thanks to my thesis supervisor, Dr. Michael Ngadi for his continuous and important guidance and encouragement as well as for giving me the opportunity to experience the graduate students and enriched my knowledge.

Special gratitude goes to my thesis co-supervisor, Dr. Shiv Prasher for his support, advice and motivation.

A special thanks to Dr. Saab Abi Saab and Dr. Marie Abi Saab who encouraged me and gave me some advices.

I owe thanks to Wassim El-Khoury for his help and support.

I would like to thanks Miss Nida Chaudhary for helping to improve my skills in many fields.

I would like also to specially thank Miss Simona Nemes for her encouragement and advice.

I am very grateful to Mariana Monroy for her help and advice in the lab's instrument.

I also owe an appreciation to Dr. Youssef Karimi Zindhasty for his valuable help and advice.

A special thanks to Akinbode Adedeji for his help in the laboratory work.

I would like to thank Arun Muthukumaran for his help.

Thanks to all the wonderful people that I met at Macdonald Campusand made my experience so satisfactory.

Foremost, I would like to express my gratitude to my family who supported and encouraged me through all my study.

To my parents

CONTRIBUTION OF AUTHORS

This thesis was written according to the rules and regulations of the Faculty of Graduate Studies and Research of McGill Guidelines for a Manuscript Based Thesis, where-in it is stated the following:

"*As an alternative to the traditional thesis format, the thesis can consist of a collection of papers of which the student is an author or co-author. These papers must have a cohesive, unitary character making them a report of a single program of research. The structure for the manuscript-based thesis must conform to the following:*

1. Candidates have the option of including, as part of the thesis, the text of one or more papers submitted, or to be submitted, for publication, or the clearly-duplicated text (not the reprints) of one or more published papers. These texts must conform to the "Guidelines for Thesis Preparation" with respect to font size, line spacing and margin sizes and must be bound together as an integral part of the thesis. (Reprints of published papers can be included in the appendices at the end of the thesis.)

2. The thesis must be more than a collection of manuscripts. All components must be integrated into a cohesive unit with a logical progression from one chapter to the next. In order to ensure that the thesis has continuity, connecting texts that provide logical bridges preceding and following each manuscript are mandatory.

3. The thesis must conform to all other requirements of the "Guidelines for Thesis Preparation" in addition to the manuscripts.

The thesis must include the following:

1. *a table of contents;*
2. *a brief abstract in both English and French;*

3. *an introduction which clearly states the rationale and objectives of the research;*
4. *a comprehensive review of the literature (in addition to that covered in the introduction to each paper);*
5. *a final conclusion and summary;*

4. As manuscripts for publication are frequently very concise documents, where appropriate, additional material must be provided (e.g., in appendices) in sufficient detail to allow a clear and precise judgement to be made of the importance and originality of the research reported in the thesis. "

Manuscripts based on the thesis:

1. Abdel-Nour, N, M. Ngadi, S. Prasher, and Y. Karimi. 2008. Combined Maximum R^2 and partial least squares method for wavelengths selection and analysis of spectroscopic data. Applied Spectroscopy. (Submitted for publication).
2. Abdel-Nour, N., M. Ngadi, S. Prasher, and Y. Karimi. 2008. Prediction of egg freshness and albumen quality using visible/near infrared spectroscopy. Presented at "76e Congrès de l'ACFAS".
3. Abdel-Nour, N., M. Ngadi, S. Prasher, and Y. Karimi. 2008. Prediction of egg freshness and albumen quality using visible/near infrared spectroscopy. Food and Bioprocess Technology. (Submitted for publication).

All the sample collection, experimental works, data analysis, and the preparation of the manuscripts were completed by the candidate, under the supervision of Dr. Michael Ngadi, thesis supervisor and Dr. Shiv Prasher, thesis co-supervisor. Dr. Youssef Karimi provided additional help and support.

TABLE OF CONTENTS

LIST OF FIGURE

LIST OF TABLES

LIST OF SYMBOLS

A	The *m* by *p* matrix of absorbance values
a	The intercept
a_i	The common shift in multiple linear regression
b	The slope
b	Path-length of light
b_i	The common amplification in multiple linear regression
CFIA	Canadian Food Inspection Agency
C	Carbon
c	Concentration of the absorbing material
CO_2	Carbon dioxide
E	The *n* by *N* matrix of residual
e_{ik}	The residuals in multiple linear regression
E_A	The matrix of residual spectra
EN	Electronic nose
F	The *f* by *p* matrix of eigenvectors
F	the *n* by *M* matrix of residual
FT-NIR	Fourier transform near-infrared
H	Hydrogen
h	The observed height of albumen in millimetres
HU	Haugh Unit
IR	Infrared
KANN	Kohonan artificial neural network
LV	Latent variable
LR H NMR	Low-resolution proton nuclear magnetic resonance
MAXR	Maximum R^2

MAXR-PLS	Maximum R^2 -partial least squares
MLR	Multiple linear regression
MRI	Magnetic resonance imaging
MSC	multiplicative scatter correction
NIR	Near infrared
P	The *N* by *p* matrix of loading
PCA	Principal component analysis
PLS	Partial least squares
PLS1	Partial least squares, type 1
PRESS	Prediction residual error sum of squares
PRESS(r)	Prediction residual error sum of squares obtained with r factors
Q	The *M* by *p* matrix of loading
r	Correlation coefficient
R^2	Coefficient of determination
RMSE	Root mean square error
RMSECV	Root mean square error for cross-validation
RMSEP	Root mean square error for prediction
RPD	Ration of performance to deviation
S	The *m* by *f* matrix of scores
s	Second
T	the *n* by *p* matrix of the extracted latent variables
U	the *n* by *p* matrix of the extracted latent variables
VIS	Visible
VIS/NIRS	Visible/Near Infrared spectroscopy
w	The weight of the egg in grams
$\hat{y}_i$	The prediction of the concentration of interest in calibration sample *i*
$\hat{y}_{i/i}$	The predicted concentration values
y_i	The measured value in calibration sample *i*

Greek

$A(\lambda)$	Absorbance at wavelength λ
$\varepsilon(\lambda)$	Molar absorption coefficient at wavelength λ

CHAPTER 1

GENERAL INTRODUCTION

The egg industry is facing challenges of intensive production with decreased labour. There has been an increase in the number of laying hens and diets of hens have improved. These factors have led to an increase in production of eggs at a lower cost. However, this industry needs accurate and reliable information about the egg in order to grade it precisely and to provide quality to the consumers that meets their requirements with respect to egg quality and standards (Kemps et al., 2006). The quality of an egg is defined by its internal and external attributes. The internal attributes consist of quality of the albumen and yolk parts. The most widely used methods for internal egg quality measurement are albumen pH and Haugh Units (HU) (Haugh, 1937). External quality is dependent on the egg shell quality and is related to the existence of cracked shell or stains.

In practice, egg quality can be assessed by candling-visual inspection. This is the simplest and most widely used method but it is subject to human error. Objective evaluation can be done by evaluating microbiological, physical and chemical quality by the use of laboratory methods. These methods are time consuming, need sample destruction and use of chemical compounds. In addition, they cannot be applied on-line. Thus, the development of accurate, on-line and non-destructive techniques is needed.

Much research has been carried out with the aim of developing a non-destructive method for egg quality assessment. Kuchida et al. (1999) used computer image analysis for the prediction of albumen to yolk ratio and found that this method can predict the ratio without breaking the egg. Hutchison et al. (1992) used magnetic resonance imaging (MRI) and reported that this technique can be used for assessment of the microanatomy of eggs. Dutta et al. (2003) used an electronic nose (EN) system employing four tin-oxide odour sensors for the determination of egg freshness. They reported that they could divide the eggs into three groups based on freshness with an accuracy of 95%. Berardinelli et al. (2005) used Fourier transform near-infrared (FT-NIR) spectroscopy to evaluate the height of the thick albumen and found a good correlation between spectral and measured data.

Visible/Near-infrared spectroscopy (VIS/NIRS) is one of the most powerful techniques for quantitative and qualitative analysis of foods because valuable chemical, moisture and other descriptions of the constituent parts of an item can be provided by spectral data (Casasent and Chen, 2003). As an example, quantitative information about water and protein content can be obtained by analysis of NIR absorption spectra (Giangiacomo and Dull, 1986). Advances in modern multivariate algorithms and improvement in the speed of microprocessors facilitate the extraction of valuable information about chemical and physical composition of foods using this technique.

VIS/NIRS has many advantages over traditional chemical methods. It is quick, non-destructive, accurate, reliable, contactless and economical (De Ketelaere et al., 2004). It has the potential to measure different components of foods at the same time. However, it is highly dependent on time-consuming calibration procedures and data analysis (Buning-Pfaue, 2003). Nevertheless, once the instrument is calibrated, measurements can be obtained in an average time of 0.25 s per measurement.

In the past few years, VIS/NIRS has been used increasingly in assessment of the internal quality of agricultural products (Osborne et al., 1993; Anderson and Walker, 2003; Park et al., 2001). Research carried out includes the determination of dry matter in onions, quality characteristics of mandarin, quality of kiwi fruit, fat and moisture content in foods (Birth et al., 1985; Gómez et al., 2006; Slaughter and Crisosto, 1998; Cozzolino, 2005; Miralbes, 2003).

With respect to egg quality and freshness, research has been carried out to study the usefulness of VIS/NIR spectroscopy for the assessment of egg quality and freshness. Kemps et al. (2006) used VIS/NIRS for the prediction of quality of eggs and then built a partial least squares model to link the spectral data with the measured HU and albumen pH. The correlation coefficient between the measured and the predicted HU and albumen pH had a maximum value of 0.86. Kemps et al. (2007) then combined the transmittance spectroscopy with low resolution magnetic resonance spectroscopy (LR H NMR) in order to improve this predictive model.

Current literature lacks thorough exploration on the usefulness of the NIR region in determination of egg quality in terms of the most widely used destructive method, namely HU and albumen pH. This is also the case with respect to egg freshness in terms of number of storage days.

The development of a rapid VIS/NIRS analytical technique for the assessment of egg freshness and quality is required. In addition, there is a lack of an automated method for the selection of relevant wavelengths. Development of such a method will lead to a robust predictive model due to exclusion of the uninformative wavelengths.

1.1 REFERENCES

Anderson, N.M., and P.N. Walker. 2003. Measuring fat content of ground beef stream using on-line Visible/NIR Spectroscopy. Transactions of the ASAE. 46(1): 117-124.

Berardinelli,A., A. Giunchi, A. Guarnieri, F. Pezzi, and L. Ragni. 2005. Shell egg albumen height assessment by FT-NIR Spectroscopy. Transactions of the ASAE. 48(4): 1423-1428.

Birth, G.S., G.G. Dull, W.T. Renfore, and S.J. Kays. 1985. Non-destructive spectrometric determination of dry matter in onions. Journal of the American Society for Horticultural Science. 110(2): 297-303.

Buning-Pfaue, H. 2003. Analysis of water in food by Near-Infrared Spectroscopy. Food Chemistry. 82: 107-115.

Casasent, D., and Chen, X.W. 2003. Waveband selection for hyperspectral data; optimal feature selection. The International Society for Optical Engineering. Optical Pattern Recognition XIV. pp. 259-270.

Cozzolino, D., I. Murray, A. Chree, and J.R. Scaife. 2005. Multivariate determination of free fatty acids and noisture in fish oils by partial least-squares Regression and Near-Infrared Specroscopy. Food Science and Technology. 38 (8): 821-828.

De Ketelaere, B., F. Bamelis, E. Decuypere, and J.G. De Baerdemaeker. 2004. Non-destructive measurements of the egg quality. World's Poultry Science Journal. 60: 289-302.

Dutta, R., E.L. Hines, J.W. Gardner, D.D. Udrea, and P. Boilot. 2003. Non-destructive egg freshness determination: an electronic nose based approach. Measurement Science and Technology. 14: 190-198.

Giangiacomo, R., and Dull, G.G. 1986. NIR spectrophotometric determination of individual sugars in aqueous mixtures. Journal of Food Science. 51: 679-683.

Gómez, A.H., Y. He, and A.G. Pereira. 2006. Non-destructive measurement of acidity, soluble solids and firmness of satsuma mandarin using VIS/NIR-Spectroscopy techniques. Journal of Food Engineering. 77:313-319.

Haugh, R.R. 1937. A new method for determining the quality of an egg, U.S. Egg Poultry. 39: 27-49.

Hutchison, M.J., A. Lirette, R.J. Etches, R.A. Towner, and E.G. Janzen. 1992. An assessment of egg yolk structure using magnetic resonance imaging. Poultry Science. 71: 2117-2121.

Kemps, B J, F. R. Bamelis, B. De Ketelaere, K. Mertens, K. Tona, E.M. Decuypere, and J.G. De Baerdemaeker. 2006. Visible transmission spectroscopy for assessment of egg qulaity. Journal of the Science of Food and Agriculture. 86:1399-1406.

Kemps, B.J., B. De Katelaere, F.R. Bamelis, K. Mertens, K. Tona, E.M. Decuypere, J.G. De Baerdemaeker, and F. Schwägelet. 2007. Albumen freshness assessment by combining visible Near-Infrared Transmission and Low-Resolution Proton Nuclear Magnetic Resonance Spectroscopy. Journal of Poultry Science. 86: 752-759.

Kuchida, K., M. Fukaya, S. Miyoshi, M. Suzuki, and S. Tsuruta. 1999. Nondestructive prediction method for albumen:yolk ratio in chicken eggs by computer image analysis. Poultry Science. 78: 909-913.

Miralbes, C. 2003. Prediction chemical composition and alveograph parameters on wheat by Near-Infrared Transmittance Spectroscopy. Journal of Agriculture and Food Chemistry. 51: 6335-6339.

Osborne, B.G., T. Fearn, and P.H. Hindle. 1993. Aplications of Near Infrared Spectroscopy in food and beverage. *In"*Practical NIR spectroscopy with applications in food and beverage analysis". (DDavid Browning ed.). Harlow: Longman Scientific and Technical.

Park, B., Y.R. Chen, W.R. Hruschunka, S.D. Shackellford, and M. Koohmaraie. 2001. Principal component regression of near-infrared reflectance spectra for beef Tenderness prediction. Trans ASAE. 44(3): 609-615.

Slaughter, D.C., and C.H. Crisosto. 1998. Non-destructive internal quality assessment of kiwifruit using Near-Infrared Spectroscopy. Seminars in Food Analysis. 3: 131-140.

CHAPTER 2

GENERAL OBJECTIVES

This study had the general objective to investigate the usefulness of VIS/NIRS in the prediction of egg freshness in terms of the number of storage days and the egg quality in terms of the most widely destructive method used, namely the Haugh Units (HU) and the albumen pH. The specific objectives were:

1. To create a new automated method for selection of relevant wavelengths used to build a partial least squares (PLS) predictive model.
2. To investigate the ability of the combination of maximum R^2 with PLS regression as an automated method for the wavelength selection.
3. To build a VIS/NIRS based calibration model to predict the HU, albumen pH and number of storage days.
4. To test and validate the ability of each model on prediction of the mentioned quality and freshness parameters.

CHAPTER 3

LITERATURE REVIEW

In this chapter relevant literature on chicken egg quality and application of visible and near infrared (VIS/NIR) spectroscopy for egg quality is reviewed. First the basics of chicken egg quality are introduced. Then a general introduction on destructive and non-destructive methods used for the assessment of egg quality is discussed. After, an introduction of the principles of VIS/NIR spectroscopy is given and then related research on these techniques and their use for chicken egg quality prediction are explored. Finally, different methods for analyzing spectral information are discussed.

3.1 CHICKEN EGG

Egg is an inexpensive but very nutritious component within the human diet. It is one of the few foods that are used widely worldwide and are healthy and safe for consumers. It is a vehicle for reproduction and can be raw material for many products in food processing plants.

3.1.1 Chicken egg constituents

A complicated series of processes leads to the egg of the laying hen. In general the egg needs 2 weeks to be formed. The major parts of the chicken egg are: yolk, albumen, membrane and egg shell.

The yolk of the egg is formed in three stages during 10 to 12 days before being laid. Albumen is formed during a few hours and has 4 layers. The percentage of the total albumen varies between eggs due to many factors. Shell membrane is constituted of inner and outer membranes separated by the air cell; their role is defence against bacterial invasion. Egg shell is constituted of calcium carbonate, magnesium carbonate, calcium phosphate, organic matter and protein. Egg shell is formed in a distinct pattern allowing gas exchange.

Chicken egg constituents are shown in figure 3.1.

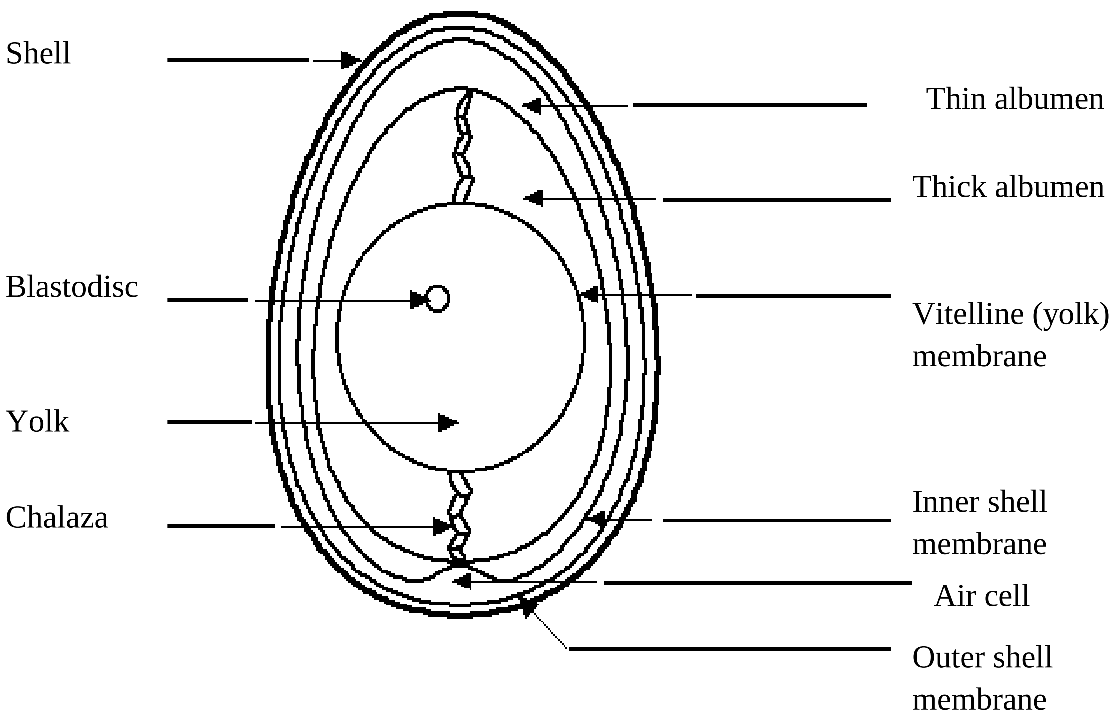


Figure 3.1 chicken egg constituents (adapted from http://www.enchantedlearning.com/subjects/birds/label/chickenegg2/label.IF)

Composition of the egg as well as percentage of each layer is shown in Table 3.1.

Table 3.1

Composition of chicken egg (Stadelman and Cotterrill, 1995)

Main part	Layer	% of weight of egg
Yolk	Germinal disk (blastoderm) Latebra Light yolk layer Dark yolk layer Vitelline membrane	30 - 33
Albumen	Outer thin albumen Outer thick albumen Inner thin albumen Inner thick albumen Chalaza	60
Membrane	Outer shell membrane Inner shell membrane Air cell	0 - 1
Shell	Cuticle Spongy (calcareous) layer Mammillary layer	9 - 12

3.1.2 Chicken Egg Quality

Eggs are fragile commodities and the quality begins to decline as soon as it is laid. Many factors affect egg quality, for instance: hen age induced moulting, climate, environmental effects and nutrition. Thus, egg quality is highly unpredictable in these products.

Egg quality is based on characteristics of the egg that affect its acceptability to the consumer. Each of the main components of the egg (shell, albumen and yolk) has a natural variability which is not in line with the modern consumers' requirements (De Katelaere et al., 2004). Nowadays, concern about egg quality is growing steadily (Kemps et al., 2006).

The overall quality of the chicken egg is determined by the egg shell quality and egg internal quality. Both of them are of paramount importance to the egg industry. The appearance of the egg is important for consumer appeal. In fact, egg shell quality is based on egg size, egg specific gravity, shell color, shell breaking strength, shell deformation, shell weight, percentage shell, shell thickness and shell ultra structure (Roberts, 2004). For table eggs, shells must be strong enough to prevent failure during packing and/or transportation (Narushin et al., 2004). For hatching eggs, shells must be initially thick and strong to preserve the embryo and then it must become thin and weak later during incubation in order to allow the gas exchange as well as easier cracking when hatching (Narushin and Romanov, 2002). Interior egg quality is based on albumen quality, yolk quality and the presence of blood or meat spots (Jacob et al., 2000).

3.1.3 Egg grading

The purpose of the egg grading is to sort the eggs into categories based on size or weight, quality factors of the shell and such internal portions of the egg as the albumen, yolk, air cell and possible abnormalities (Stadelman, 1995). The eggs are divided into four groups in terms of the interior and exterior quality: Canada A, Canada B, Canada C and Canada Nest Run. Table 3.2 summarizes the qualifications of each grade according to the Canadian Food Inspection Agency (CFIA) standards.

Table 3.2

Summary of Canadian standards for individual egg quality (Egg Regulations, 2008)

	Specifications for each Grade			
	Canada A	Canada B	Canada C	Canada Nest Run
Shell	Clean Uncracked Normal in Shape	Clean Uncracked Slightly abnormal in shape	Cracked but not leaking Stain spots less than 1/3 of the shell surface	10% cracked 5% of eggs with dirt where the surface is more than 160 mm^2 3% of eggs are leakers
Air cell	5 mm or less in depth	8 mm or less in depth	Over 3/16 inches in depth Unlimited movement and free or bubbly	
Albumen	Reasonably firm	Clear Reasonably firm	Meat spots or blood spots less than 3mm in diameter	
Yolk	Indistinct Outline Round and reasonably well centered	Distinct outline Very slight degree of germ development	Prominent outline Definitely oblong in shape	

3.2 EGG WHITE OR ALBUMEN

The albumen constitutes 60% of the egg weight, 12% of the albumen is solids, 10.2% is protein, 1.0% is carbohydrate and 0.68% ash (Froning, 1998; Watkins, 1995). Egg white protein and characteristics are shown in Table 3.3.

Table 3.3

Egg white proteins and characteristics (Froning, 1998; Li-Chan et al., 1996)

Protein	% of albumen proteins	Characteristics
Ovalbumin	54	Phosphoglycoprotein
Ovotranferrin (Conalbumin)	12	Binds metallic ions
Ovomucoid	11	Inhibits trypsin
Ovomucin	3.5	Sialoprotein, Viscous
Lyzozyme	3.4	Lyzes proteins
Globulins	8.0	---
Ovoinhibitor	1.5	Inhibits serine proteases
Ovoglycoprotein	1.0	Sialoprotein
Ovoflavoprotein	0.8	Binds riboflavin
Ovomacroglobulin	0.5	Strongly antigenic
Cystatin	0.05	Inhibits thiol proteases
Avidin	0.05	Binds biotin

3.2.1 Albumen quality

Albumen quality has a major influence on overall interior egg quality and thinning of the albumen can point to a quality loss. An egg with a good albumen quality should be free from internal blemishes such as blood spots, pigment spots and meat spots. The albumen consists of thick and thin material, which in the fresh egg alternatively surrounds the yolk sphere in three concentric layers, the thin layer, the thick fibrous layer, and the outer thin layer (Mathews, 1986). Thick albumen is a gel and thin albumen is a fluid (Brooks and Hale, 1959). The thick albumen forms a capsule around the yolk that is impenetrable in fresh eggs. During storage, the gelatinous structure of thick albumen changes its physical and chemical characteristics and gradually breaks down into a clear liquid, loosing consistency (Robinson and Monsey, 1972). Albumen quality is measured in terms of Haugh Unit (HU) calculated from the weight of the intact egg and the height of the albumen (Haugh, 1937).

3.2.2 Factors affecting albumen quality

Many factors are reported by the researchers to affect the albumen quality: storage time, temperature, hen age, strain of bird, nutrition, disease. These factors interact in affecting the quality.

3.2.2.1 Effect of storage time and temperature

Egg storage time and conditions are the major factors that affect albumen quality. After the egg is laid, the carbon dioxide (CO_2) evaporates through the shell causing an increase of albumen pH. This loss is faster at higher temperature. The increase in albumen pH can be a reason for the change in viscosity of the albumen. Kemps et al. (2007) reported that 65% of the variation in HU is accounted by an increase in albumen pH. With storage, albumen pH increased and albumen height decreased (Li-Chan and Nakai, 1989). These changes led to a decrease in the HU.

Scott and Silversides (2000) reported that there was no effect of storage time on eggshell weight. However, the principal changes with storage were the decrease in albumen and egg weights. Schäfer et al. (1999) reported that with time, the isoelectric point of ovalbumin becomes slightly acidic and this change is in accordance with the formation of S-ovalbumin. They concluded that these changes are related to temperature rather than storage time. Kröckel et al. (2005) reported that the microbial growth in the egg increases with age. This growth is related strikingly with storage temperature and time.

3.2.2.2 Effect of hen strain and age

Silversides and Scott (2001) reported that the strain of the bird affects the quality of the egg and they found that the eggs from ISA-Brown hens were larger than those from ISA-White hens, with more shell and albumen but less yolk. In addition, when considered as a percentage of the egg, the shell and albumen decreased with increasing age of the hen. Coutts and Wilson (1990) reported that HU decreases by around 1.5 to 2 units for each month of lay with increasing age. Kröckel et al. (2005) found that the age of the hen has an effect on the bacterial stability of the eggs and they reported that this stability is lower in the middle of the laying cycle.

3.2.2.3 Nutrition and disease

Bird nutrition affects the albumen quality. Balnave et al. (2000) reported that the quality increases with increasing dietary protein and amino acid. Franchini et al. (2002) found that a supplementation of ascorbic acid can increase the quality.

Infectious bronchitis virus is the main disease that affects albumen quality (Spackman, 1987). Infectious bronchitis impairs the synthesis of albumen proteins in the magnum of the oviduct (Butler et al., 1972).

3.2.3 Causes of decreasing albumen quality

Many mechanisms lead to the liquefaction of egg white including: protease enzymes, depolymerisation by hydroxyl ion at increasing pH values, reduction by thiol type reducing agents and interaction with lysozyme. Proteolytic enzymes, hydroxyl ion and disulphide bond depolymerise ovomucin and hence lead to the liquefaction of the albumen (Wells and Belyavin, 1987). However, these substances are not solely responsible for the natural liquefaction of the egg white gel or the natural depolymerisation of ovomucin (Beveridge and Nakai, 1975). In addition, in view of the existence of O-glycosidically linked trisaccharides, specifically in β-ovomucin, enzymatic hydrolysis of this glycosidic link may also be responsible for the liquefaction of the thick egg white gel (Robinson, 1987).

This chemical reaction takes place during the natural liquefaction of thick egg white gel at the relatively high pH value of 9.2 in egg white. The destruction of the gelatinous nature of thick egg white can occur due to ovomucin-lysozyme interaction as the pH of the albumen changes after being laid (Robinson and Monsey, 1972).

3.3 EGG PRODUCTION IN CANADA

Egg production in Canada for different purposes from January 2006 until March 2008 is shown in Table 3.4.

Table 3.4
Chicken egg production in Canada (Statistics Canada, 2008)

	2006	2007	2008 (until September)
	Quantity in thousands of dozens		
Sold for consumption	508,395	504,555	378,096
Sold for hatching	59,282	60,344	45,703
Home consumption	1,495	1,495	1053
Leakers and rejects	11,085	11,051	8434
Total disposition of the eggs	580,257	577,444	433,280

3.3.1 Egg Production Challenges

Egg is an important part of human diet. It is also an important raw material and ingredient in food processing. Egg quality is critical from the consumer, hatchery and processing plant points of views. Consumers expect safe and healthy products. Hatcheries need to increase egg hatchability rates whereas egg processors seek good quality raw materials in order to minimize economic losses, optimize plant process operations and maintain quality standards. However, loss of eggs resulting from leakers and rejects can reach 2% of total production as can be seen in Table 3.4.

A major problem facing the egg industry is food safety. This problem came out after the demonstration in 1988 that contents of intact eggs can be contaminated by *Salmonella enteritidis* (S.e.). This contamination can lead to food poisoning incidents. After such contamination, consumers come to believe that most of the eggs are contaminated. They avoid using recipes that contain raw or undercooked eggs.

Another major problem is to predict egg quality with higher efficiency. Nowadays, the egg quality is assessed using candling. Candling has the advantage of being non-destructive and rapid. However, since each egg must be examined regardless of the type of equipment used, the increasing throughput of modern egg sorting machines makes this quality inspection type more difficult and less reliable (De Ketelaere et al., 2004).

A fast and non-destructive quality assessment tool with the modern information technology has multiple advantages: in the packing plant, such fast technologies allow for measuring the quality of an individual egg instead of estimating the quality by sampling some eggs out of a large batch. These techniques also allow the measurement of more quality attributes and guarantee a better overall quality for each individual egg (De Ketelaere et al., 2004).

3.4 CHICKEN EGG QUALITY EVALUATION

In the packing plant, eggs are sorted into different classes based on their quality. For this purpose, these plants use destructive and non-destructive methods. Quality indicators such as dirt, cracks, meat spots and blood spots are evaluated, making the classification subject to human error.

3.4.1 Destructive methods

The most widely used and accepted measure of albumen quality is the Haugh Units (Haugh, 1937). This method is based on the weight of the intact egg and the albumen height of the broken egg. When a fresh egg is carefully broken on to a smooth flat surface, the yolk is in a central position surrounded by the thick albumen. When a stale egg is broken, the yolk is often displaced to one side; the thick albumen has become thinner and the surface of the albumen increases resulting in a decreasing of the albumen height leading to a decrease of HU (Kemps et al., 2006). Silversides and Villeneuve (1994) criticized the Haugh method and they reported that there was a weak relation between the albumen height and egg weight and they concluded that HU was dependant on the albumen height and independent from the egg or albumen weight until the third week of storage thus, when all the data were combined the HU correction for egg weight is unnecessary.

Since pH changes with time, it can be used as a measure of albumen freshness. This change is due to the loss of the carbon dioxide (CO_2) through the pores of the eggshell. The albumen pH is dependent on the equilibrium between dissolved carbon dioxide, bicarbonate ions, carbonate ions and proteins (Li-Chan et al., 1995). Albumen refractive index can be used as an indication for the freshness of the egg. The conversion of ovalbumin into S-ovalbumin is also an interesting indicator of egg aging (Donovan and Mapes, 1976). Furosine, an indicator of the initial steps of a Maillard reaction, produced by acid hydrolysis of the Amadori compounds, can be a very reliable index of egg freshness when determined in the albumen (Hildago et al., 1995; Hidalgo et al., 2006).

3.4.2 Non-destructive methods

Since candling is a simple, rapid and non-destructive method, it is the most widely used technique. The basic of this technique is to shine a light through an egg in order to assess its conditions. Accurate candling is best done in a darkened room. Candling equipment can range from a simple unit to a sophisticated instrument. Each egg has to be examined irrespective of the type of instrument used. In freshly laid eggs, the yolk is seen as a faint shadow. With time the egg quality decreases, the yolk casts darker shadow since it floats closer to the shell. A darker shadow can also be seen when eggs contained enlarged yolks or when the vitelline membrane is weak (Jacob et al., 2000). Blood spots can also be detected as a dark, red-colored spots in the egg. The limitation of this method is that when the throughput increases up to 120,000 eggs per hour it becomes less accurate (De Ketelaere, 2004). Candling method is shown in Figure 3.3.

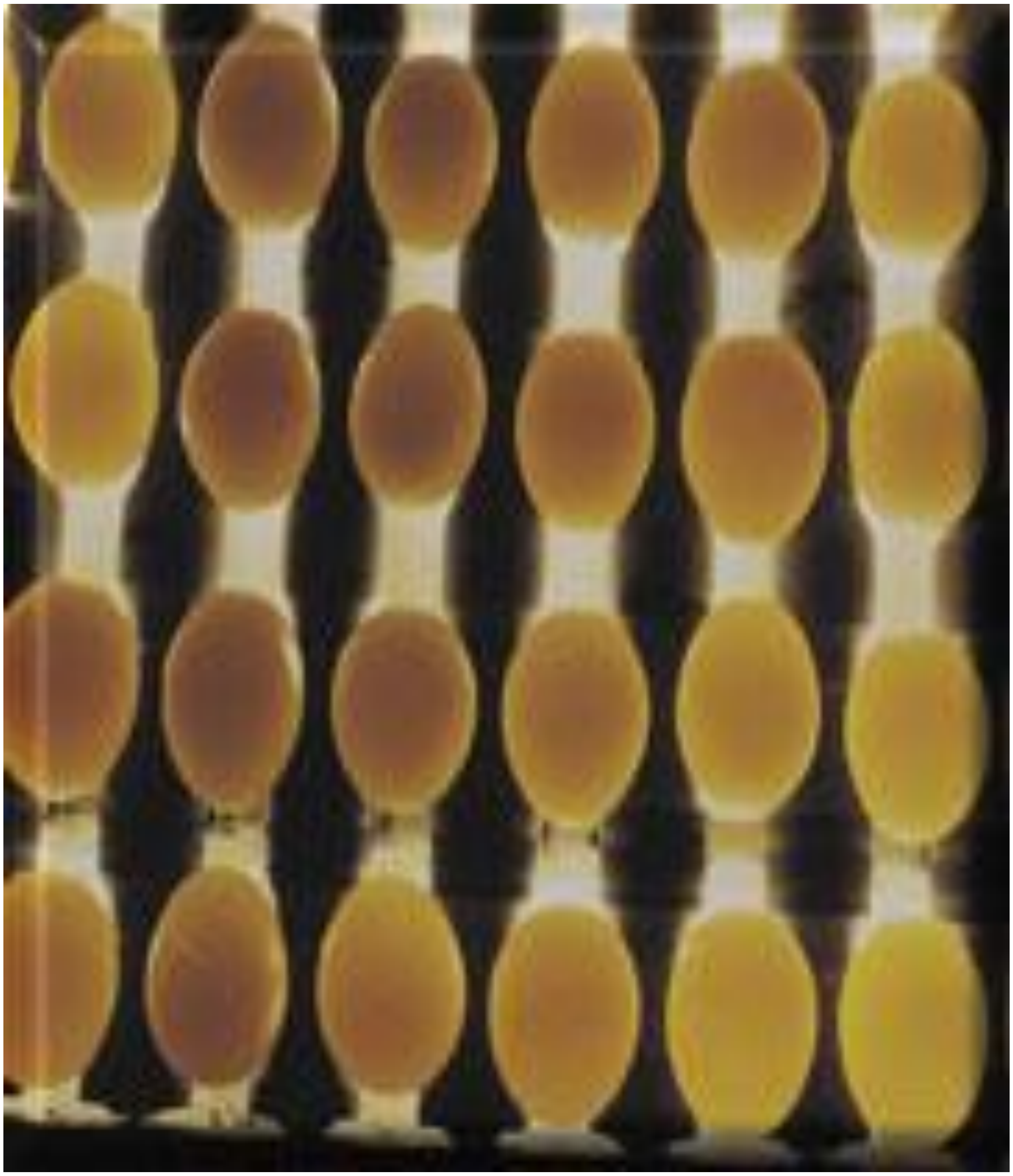


Figure 3.2 Candling eggs (adapted from http://www.eggs.ab.ca/egg_industry/ EGGSLIGHT2.jpg)

Kuchida et al. (1999) proposed computer image analysis as a non-destructive method for prediction of the yolk:albumen ratio. They reported that there were a significant correlation coefficient of 0.85 between observed and predicted values and hence the classification of the yolk:albumen ratio by a non-destructive method is feasible. Another non-destructive method based on the torsion pendulum can provide an estimate of the ratio of thick to thin albumen in the egg (Tsarenko, 1973).

Karaoui et al. (2006) used front face fluorescence spectroscopy for the evaluation of thick and thin egg albumen. They reported that the Maillard reaction can be used as an indicator to differentiate between fresh and aged eggs. Ragni et al. (2007) examined the dielectric characterization of the chicken eggs during storage, for this purpose they measured the dielectric constant and loss factor over the frequency range 20-1800 MHz. Open-ended coaxial probe on thick albumen and yolk of eggs was used after 1, 2, 4, 8 and 15 days of storage. They found that the maximum increment during storage time was registered for dielectric constant at 20 MHz.

Schwägele et al. (2001) used low-resolution nuclear magnetic resonance spectroscopy to assess the quality of chicken egg. Eggs were placed in a homogeneous magnetic field to excite the hydrogen nuclei in the sample. They used the longitudinal and transversal relaxation times to measure the changes in the eggs. They found that during the first week of storage the transversal relaxation time shows an exponential decrease for all temperatures whereas the transversal relaxation time decreased linearly, especially at higher temperatures due to liquefaction of albumen. Dutta et al. (2003) used an electronic nose based system. For this technique they used four tin-oxide odour sensors to determine the egg freshness. Völgyi (2000) investigated the usefulness of using microwave sensors for the assessment of egg freshness. He found that this method is a useful tool for observing the temporal variations in the subsurface properties of the eggs and reported that there was a variation of the chemical and physico-chemical measurements of ammonia content and viscosity with age. His plans included the development of a microwave device for the automatic selection of the old eggs. Berardinelli et al. (2005) tried FT-NIR spectroscopy and they found that their model was able to predict albumen height with a maximum error of ±1mm for 80% of the samples. They concluded that the transmission measurements of the thick albumen portion confirmed good correlation between predicted and observed values but this technology is better for a group of eggs rather than with a single egg. Laghi et al. (2005) investigated the feasibility of using proton nuclear magnetic resonance relaxation to understand the quality loss of hen egg during the first few days of storage. They found that the low-field is the best relaxation time indicator of albumen quality. Kröckel et al. (2005) successfully used LR H NMR as a non destructive method to evaluate the relation between microorganisms and the physico-chemical changes in egg.

3.5 VISIBLE/NEAR INFRARED SPECTROSCOPY

3.5.1 General introduction

3.5.1.1 History

Even though the use of NIR spectroscopy has been growing steadily during the last two decades, this technique is old. In fact, William Herschel in 1800 noticed the NIR spectrum when his thermometer registered temperatures even beyond the red portion of visible region. It was not until 1881 that the importance of hydrogen bonds in producing NIR absorptions was realised (Abney and Festing, 1881). Bands characteristics of C-H bonds were found in 1905. The introduction of the assignment of NIR spectral bands of organic compounds to specific groups was between 1922 and 1929 (Ellis, 1929). In 1960s the first NIR instrument for on-line measurement appeared in market and one of the earliest on-line application was for the inspection of flour in 1980s (Osborne, 1986; Osborne et al., 1989). It has been widely used for the determination of internal quality of foods (Williams and Norris, 1987). Twenty one years ago, a single instrument allowed scanning the visible and the NIR spectrum (McCaig, 2002) where the visible region is the range of the wavelengths between 400 and 700 nm and the NIR region is the range of the wavelengths between 700 and 2500 nm. This instrument allowed the measurement of pigments such as chlorophyll in immature canola seeds with selection of relevant wavelengths associated with the pigments (Williams and Sobering, 1993). Using this NIR technique, insects in single wheat kernels can be identified, water in meat, sucrose, fructose and glucose in fruit juices can be predicted (Maghirang et al., 2003, Norris, 1983, Osborne et al., 1993a). The analysis of NIR absorption spectra provides quantitative information about constituents like water and proteins in a scanned sample (Giangiacomo and Dull, 1986).

Nowadays, due to the development of mathematical multivariate data analysis and due to the improvement in the speed of microprocessors, quality of several kinds of agricultural products can be determined using optical measurements on laboratory scale and for on-line processes. Over the last two decades, near-infrared spectroscopy (NIRS) use has increased and has been recognized as a non-destructive method for the estimation of a number of food constituents and properties by measuring the amount of light transmitted or reflected. NIRS is a convenient tool not only for characterizing foods, but also for quality assessment and process control (Andres et al., 2007).

3.5.2 Theory of NIR spectroscopy

Spectroscopy is the study of the interaction between electromagnetic radiation and atoms, molecules, or other chemical species (Mohan, 2004). The electromagnetic spectrum varies from shorter wavelengths to longer wavelengths as shown in Figure 3.3. The spectrum is divided into several regions including visible and infrared. The visible region covers the wavelengths ranging from 0.4 to 0.7 μm.

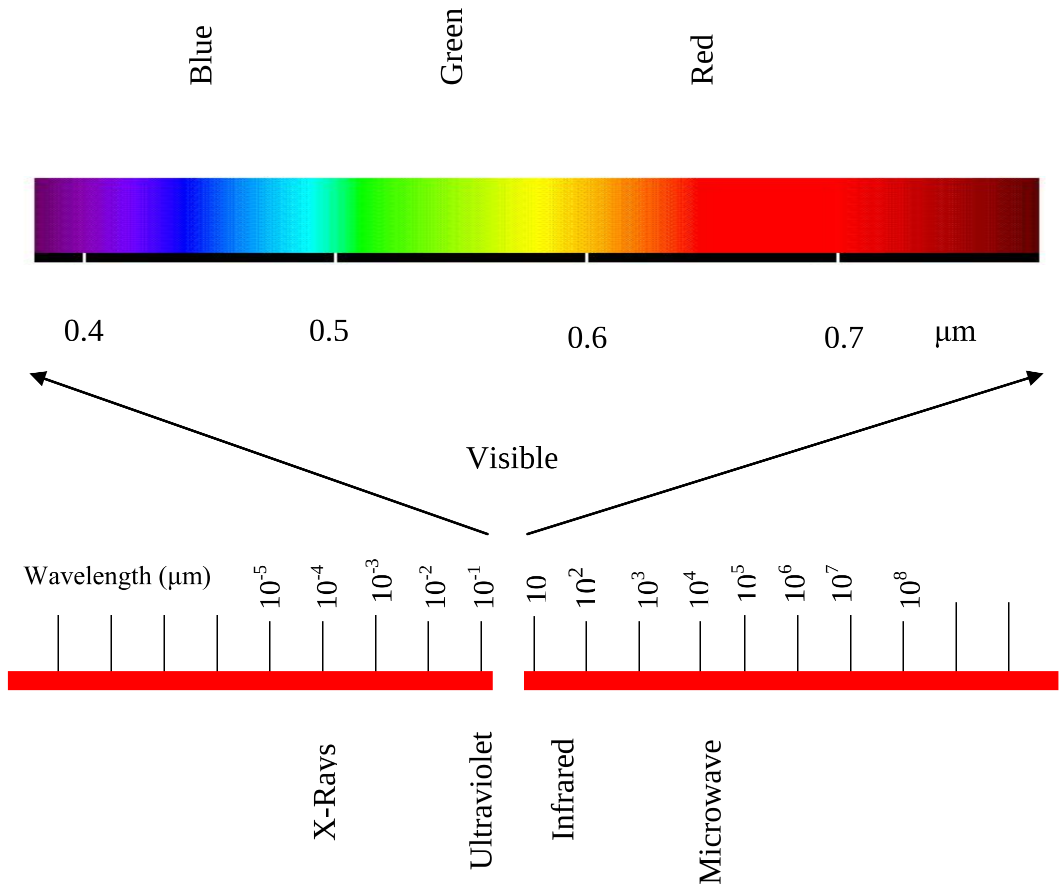


Figure 3.3 Schematic of electromagnetic spectrum

Infrared refers to the spectral region that extends from 700 to 10^6 nm in wavelength. This region is divided into three regions: near infrared (NIR), middle infrared (mid-IR or IR) and far infrared. The smaller the wavelengths are, the higher their energy is. Division and characteristics of each region are shown in Table 3.5.

Table 3.5

Characteristics of infrared region and their divisions (Osborne et al., 1993b)

Region	Characteristic transitions	Wavelength range (nm)	Wavenumber range (cm^{-1})
Near infrared (NIR)	Overtones combinations	700 - 2500	14300 - 4000
Middle infrared (IR)	Fundamental vibrations	2500 – 5 x 10^4	4000 – 200
Far infrared	Rotations	5 x 10^4 - 10^6	200 - 10

The ground state or the lowest energy state is preferred by all atoms. When a molecule absorbs radiation energy, its state becomes the excited state and creates an absorption spectrum. The vibration caused by the absorption of IR radiation by the covalent bonds of a molecule could be stretching or bending. Stretching is when the vibration causes a continuous change in the interatomic distance along the axis of bond between the two atoms whereas bending is when the vibration involves a change in bond angle. There are 4 types of bending: scissoring, rocking, wagging and twisting (Osborne et al., 1993b). To initiate the vibration of a molecule, the IR radiation must be at the same frequency as the vibration frequency of that molecule.

Different types of bonds in the molecule need different absorption bands or characteristic absorption in IR region where the fundamental vibrations occur. These vibrations then lead to overtones and combination bands observed in the NIR region. In order to choose the corresponding bands in the NIR region, the knowledge of IR absorption wavelength for each group is a necessity. Bellamy (1975) defined the position of the fundamental bands in IR region. Wetzel (1983) reported that the results of dividing the wavelengths of the fundamental vibration by 2, 3 and 4 for the first, second and third overtones can define the position of fundamental vibration. Absorption of light atoms like hydrogen occurs in the NIR region which mainly arises from overtones and combinations of vibration of the hydrogenic bonds. These bands are broad compared to the information given in mid-IR region, because the extensive overlapping bands observed in NIR region make it difficult to relate a band to specific groups for quantitative and qualitative purposes. Sophisticated multivariate algorithms can solve this problem by correlating spectral changes to a specific component in a material. The difficulty of preparation samples in IR spectroscopy makes the use of VIS/NIRS more common when working with thick samples because the fundamental absorption bands are stronger in the NIR region. During the last decades, the VIS/NIRS has been widely used for many reasons: Simple instrumentation, no sample preparation, no use of hazardous solvents and it is fast and accurate. These reasons make the VIS/NIRS a powerful instrument for non-destructive analysis.

3.6 VIS/NIR SPECTRAL ANALYSIS OF FOOD

The major components of food (water, proteins, carbohydrates and lipids) contain the overtones and combination for these molecules' fundamental vibrations particularly those involving hydrogen (Osborne, 1993a). Each of these components has a specific response in the visible and infrared region. These responses can be used for the prediction or classification of food components. The spectral data is analysed by correlating the changes in the response and the corresponding component. VIS/NIR technique has been widely used for qualitative and quantitative assessment of food and agri-food products. It has been used for the quality assessment of crossbred material from wheat breeding programs since the late 1970s. It is used to determine the protein content of wheat. Concerning milk, analysers based on this technique have been used since 1960s to analyse protein, moisture, fat and lactose in a wide range of dairy products. NIRS is also used in the meat and fish industry. It allows determining protein, fat and moisture contents of meat products and whole fish. This technique allows grading of fruits and vegetables by size, shape and color and has the ability to detect the ripeness of fruits. It is also used in authentication of food and can establish whether the claimed sample composition is genuine or not (Downy, 1996). The use of VIS/NIRS in the assessment of food quality and safety in the last 2 decades is shown in Figure 3.4.

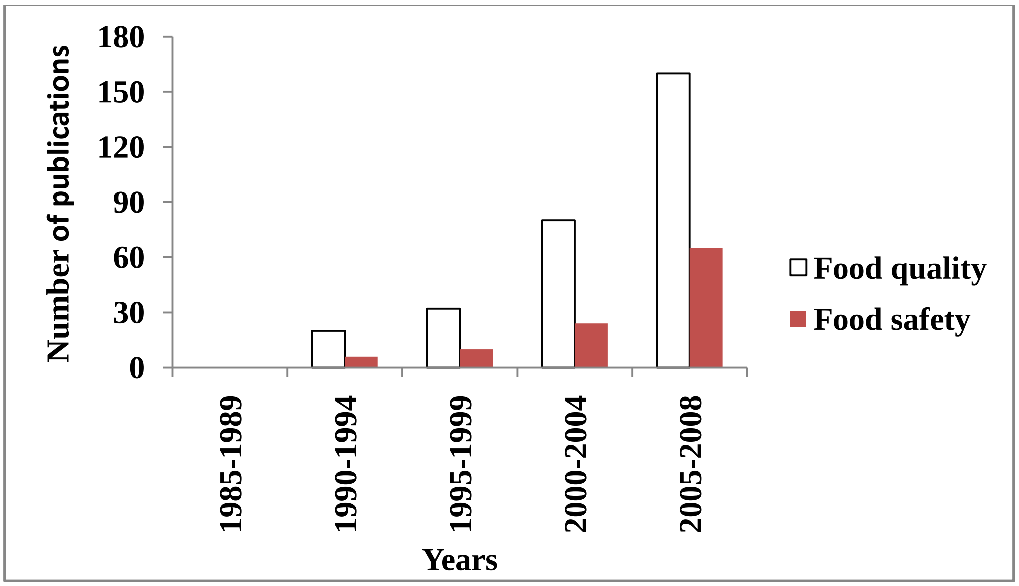


Figure 3.4 Use of Visible/Near Infrared spectroscopy in the assessment of food quality and safety from 1985 until 2008

3.6.1 VIS/NIR spectral analysis of egg

Spectroscopy has the advantage of being a fast and non-contact method. The contactless nature of this technique makes it suitable for egg quality assessment since contact can be minimized for hygienic reasons and a large number of eggs can be graded in a short period of time with no sample preparation (De Ketelaere, 2004).

The use of NIRS has also been evaluated for the prediction of shell pigmentation, blood and meat spots and hatching eggs (Wei and Bitgood, 1989; Narushin et al., 2004; Gielen et al., 1979; Das and Evans, 1992a; Das and Evans, 1992b; Bamelis et al., 2002).

Diverse results have been obtained from the studies related to prediction of egg albumen quality using NIRS. Kemps et al. (2006) investigated the feasibility of using visible transmission spectroscopy as a non-destructive method to assess the freshness of an egg. The spectral data of 600 white shelled eggs were compared with the pH and the HU. A partial least squares (PLS, Type 1) model was built to predict the HU and the pH of the albumen based on the transmission spectral data. Their results have shown that light transmission spectrum of an egg provides quantitative information about egg freshness and the correlation coefficients between the predicted and measured value were 0.84 and 0.87 for HU and pH of the albumen, respectively. Kemps et al. (2007) combined the visible and Near-infrared transmission spectroscopy with LR H NMR to investigate whether the assessment of albumen freshness will improve or not. They concluded that combining the two spectroscopic techniques did not improve this assessment when compared to the use of the transmission spectroscopy alone. Schmilovitch et al. (2002) used NIR spectral data and could predict the number of days after laying, the size of the air chamber, the weight loss and pH of eggs value with an $R^2 > 0.90$. This high value refers to group means not to the individual egg. Liu et al. (2007) used transmission spectroscopy for the measurement of internal quality of chicken eggs. They found that the relevant transmittance spectral data related to egg freshness is between 400 and 600 nm.

There is no known study on the link between spectral data and important physical quality attributes such as albumen viscosity. This is due to the available HU alternative measurement technique and the complex nature of albumen viscosity measurements when a classical rheological approach is used (De Ketelaere et al., 2003; Tung et al., 1970; Pitsilis et al., 1984; Lucisano et al., 1996).

3.7 VIS/NIR SPECTRAL DATA ANALYSIS METHODS

VIS/NIR spectroscopy contains vital spectral response information that provides detailed chemical, moisture, and other descriptions of constituent parts of an item (Casasent and Chen, 2003). Feature extraction, which is reduction of data dimensionality by extracting features from original spectral space or transformed feature space, plays a key role in the success of hyperspectral target detection and classification (Cheriyadet and Bruce, 2003).

3.7.1 General view

VIS/NIRS spectral data analysis requires calibration with a set of samples of known composition (training set) to build a model that provides a correlation between the spectral data and the component. Various algorithms can be used to build the predictive model such as Beer's law or a more complicated multivariate analysis such as partial least squares (PLS) and principal component analysis (PCA).

3.7.2 Beer's law

Beer's law is the base of quantitative analysis in absorption spectroscopy. It cites that the absorption rate at a certain wavelength is proportional to the concentration of the absorbing material. There is a linear relationship between the absorbance intensity and the concentration of the component of interest. This relationship is shown in Equation (3.1):

$$A(\lambda) = \varepsilon(\lambda) b c \qquad (3.1)$$

Where $A(\lambda)$ is the absorbance at wavelength λ, $\varepsilon(\lambda)$ is the molar absorption coefficient at the particular wavelength, b is the path-length and c is the concentration of the absorbing material. $\varepsilon(\lambda)$ can be determined by measuring the absorbance of a calibration standard with known concentration of the component. At a selected wavelength for a given component this coefficient is constant. However, this law does not include the effects of more than one component and cannot model the interactions between the components. To solve this problem, advanced multivariate calibration methods have been developed such as PLS and PCA (Sedman, 2000).

3.7.3 Principal component analysis (PCA)

Principal component analysis (PCA) is a multivariate spectral decomposition technique that reduces the spectral data depending on its variability. It uses projections to reduce a large amount of variables into a new set of variables. The new variables obtained are principal components. If, for example, the main variation in the calibration set spectral data is due to the concentration of the constituents of the sample, a new set of spectra are calculated representing the changes in all the wavelengths and relating concentration of the components in a sample. The new set is used to build the calibration model. The original calibration spectra can be reconstructed by the use of the eigenvectors: each eigenvector is multiplied by a different scaling constant and summed up with the results with the mean spectrum until the original spectrum is reached. The difference between the reconstructed spectra and the original one is called the spectral residual. The scaling constants are called scores. Scores are then regressed against the component concentration data (Thermo Galactic, 2003).

In PCA, the training data is reduced into two matrices: eigenvectors (spectra) and scores. With *m* samples, *p* wavelengths used in calibration and *f* eigenvectors, the model will be calculated using Equation (3.2):

$$A = SF + E_A \tag{3.2}$$

Where A is the ($m \times p$) matrix of absorbance values, S is the ($m \times f$) matrix of scores for all spectra, F is the ($f \times p$) matrix of eigenvectors and E_A is the matrix of residual spectra.

3.7.4 Partial least squares (PLS)

Partial least squares (PLS) is another spectral decomposition technique and it is similar to PCA but it has the advantage of performing the decomposition on both spectral and concentration data. It can be used to build a predictive model when a large number of factors exist and are highly collinear. A restricted number of factors are constructed from the combination of original spectral data and regression on the factor scores and used to build the predictive model (Osborne et al., 1993c). Two sets of eigenvectors are generated: a set of spectral loadings that represent the common variation in spectral data and a set of spectral weights corresponding to the changes of spectral response due to the difference of concentration in samples. The spectra with higher concentration of components are weighed more than those containing lower concentration (Thermo Galactic, 2003).

The computation of PLS is performed by adding factors to the regression model until the procedure stopped. Performance of the calibration set will improve by adding factors but with more factors than required, the model ends up with an overfitting and the predictive ability decreases (Osborne, 1993c). A remedial action against the overfitting is cross validation (Wold, 1978) where the need to choose a set of samples for the validation is avoided. In this technique, the same spectra used for training set are used for calibrated set. This means that with m number of calibration samples, the calibration is performed $m-1$ times, predicting the m^{th} sample (Sedman, 2000). To select the number of latent variables (LV), the most commonly used method is the prediction of residual error sum of squares (PRESS). It is plotted against the number of factors. The plot falls down to a minimum which corresponds to the best number of factors.

3.7.5 Multiplicative scatter correction (MSC)

Multiplicative scatter correction (MSC) is a pre-treatment method for the spectral data. It is used in order to remove background of the machine and noise from the data. This technique rotates the spectra to remove some of this effect; MSC rotates the entire spectrum so it can fit almost the mean spectrum (Osborne 1993c). An ideal or reference spectrum is required for the MSC which can be the average of the calibration set. When an MSC technique is applied for a calibration set, it must be applied to the samples taken before using this data in the prediction model.

3.7.6 Selection of informative wavelengths

Multivariate analysis such as PLS and PCA are good tools for working with a large amount of data, such as VIS/NIR spectral data, by building the predictive model with the entire spectra. However, reducing data by choosing the relevant wavelengths is a good strategy to improve the predictive ability of the model and to reduce calculation time.

Several methods are used for choosing the informative wavelengths. One method is to use the beta coefficient in PLS algorithm also known as linear correlation coefficient. The plot of beta coefficient shows the region where the wavelengths are correlated to the concentration of the component of interest. Positive value of beta coefficient at a particular wavelength indicates that the concentration increased with an increase in absorbance at that wavelength. Maghirang et al. (2003) used the beta coefficient to find the important wavelengths for the detection of insects in wheat kernels. Perez-Mendoza et al. (2003) used the same method in order to classify flours with or without insect fragments. Uddin and Okazaki (2004) used Multiple Linear Regression (MLR) to select the informative wavelengths in order to classify fresh and frozen-thawed fish.

Goel (2003) used the maximum R^2 (MAXR) method in the SAS software to choose the best MLR models for the estimation of various crop biophysical parameters. The MAXR method attempts to find the best one-variable method, two-variable model, and so on. It starts with the best one-variable model, to which a second variable is added in order to build the two-variable model that has the best R^2. Variables in the model are replaced one by one, until the highest R^2 is obtained. This procedure then continues for the best three, four-variable models and so on.

Bengalore et al. (1996) coupled a genetic algorithm to select the relevant wavelengths and PLS to build the predictive model. Du et al. (2004) used changeable size moving window PLS (CSMWPLS) and searching combination moving window PLS (SCMWPLS) methods. The results showed that by combining these two methods, the predictive model was improved. Todeschini et al. (1999) used Kohonen artificial neural networks (K-ANN) to select the relevant wavelengths and found an improvement in the model.

3.8 CONCLUDING REMARKS

It can be concluded from the above review that NIRS and VIS/NIRS can be used for predicting egg quality in terms of HU and albumen pH and freshness in terms of number of storage days. However, there is a lack of information on the use of the NIR region for predicting internal quality of eggs.

It can also be concluded that the VIS/NIR spectral data contains important information about the components of interest of the sample. The selection of relevant and informative wavelengths is of utmost importance in order to build a robust predictive model for the egg quality and freshness assessment.

3.9 REFERENCES

Abney, W., and E.R. Festing. 1881. On the influence of the atomic grouping in the molecules of organic bodies on their absorption in the Infra-red region of the spectrum. Philosophical Transactions of the Royal Society. 172: 887-918.

Andres, S., I. Murrat, E.A. Navajas, A.V. Fisher , N.R. Lambe, and L. Bunger. 2007. Prediction of sensory characteristics of lamb meat samples by Near-Infrared Reflectance Spectroscopy. Meat Science. 76: 509-516.

Balnave, D., R.J. Gill, X. Li, and W. Bryden. 2000. Responses of Issa Brown laying hens to a pre-layer diet containing additional calcium and to dietary protein and lysine concentrations during lay. Australian Journal of Agricultural Research Science. 79: 106-112.

Bamelis, F., K. Tona, J.G. De Baerdemaeker, and E.M. Decuypere. 2002. Detection of early embryonic development in chicken eggs using visible light transmission. British Poultry Science. 43: 922-928.

Bangalore, A.S., R.E. Shaffer, and G.W. Small. 1996. Genetic algorithm-based method for selecting wavelengths and model size for use with partial least-squares regression: application to near-infrared spectroscopy. Analytical chemistry. 68: 4200-4212.

Bellamy, L.J. 1975. The infra-red spectra of complex molecules. (Chapman and Halled.). London, England.

Berardinelli, A., A. Giunchi, A. Guarnieri, F. Pezzi, and L. Ragni. 2005. Shell egg albumen height assessment by FT-NIR spectroscopy. Transactions of the ASAE. 48(4): 1423-1428.

Beveridge, T., and S. Nakai. 1975. Effect of sulphydryl blocking on the thinning of egg white. Journal of Food Science. 40: 864-868.

Brooks, J., and H.P. Hale. 1959. The mechanical properties of the thick white of the hen's egg. Biochemistry Biophysics Acta. 32:237-250.

Butler, E.J., M.J. Curtis, A.W. Pearson, and J.S. McDougall. 1972. Effect of infectious bronchitis on the structure and composition of egg albumen. Journal of the Science and Agriculture. 23:359-369.

Casasent, D., and X.W. Chen. 2003. Waveband selection for hyperspectral data; optimal feature selection. The International Society for Optical Engineering. Optical Pattern Recognition XIV. Proceedings of the SPIE. 5106: 259-270.

Cheriyadat, A., and L.M. Bruce. 2003. Why principal component analysis is not an appropriate feature extraction method for hyperspectral data. IEEE. 3420-3422.

Coutts, J.A., and G.C. Wilson. 1990. Egg quality handbook. Queensland Department of Primary Industries, Australia.

Das, K., and M.D. Evans. 1992a. Detecting fertility of hatching eggs using machine vision II: Histogram characterization method. Transactions of the ASAE. 35(4):1135-1341.

Das, K., and M.D. Evans. 1992b. Detecting fertility of hatching eggs using machine vision II: Neural network classifiers. Transactions of the ASAE. 35(6):2035-2041.

De Ketelaere, B., F. Bamelis, E. Decuypere, and J. G. De Baerdemaeker. 2004. Non-destructive measurements of the egg quality. World's Poultry Science Journal. 60: 289-302.

Donovan, J.W., and C.J. Mapes. 1976. A differential scanning calorimetric study of conversion of ovalbumin to S-ovalbumin in eggs. Journal of the Science of Food and Agriculture. 27: 197-204.

Downy, G. 1996. Authentication of food and food ingredients by near infrared spectroscopy. Journal of Near Infrared Spectroscopy. 4:47-61.

Du, Y.P., Y.Z. Liang, J.H. Jiang, R.J. Berry, and Y. Ozaki. 2004. Spectral regions selection to improve prediction ability of PLS models by changeable size moving window partial least squares and searching combination moving window partial least squares. Analytica Chimica Acta. 501: 183-191.

Dutta, R., E.L. Hines, J.W. Gardner, D.D. Udrea, and P. Boilot. 2003. Non-destructive egg freshness determination: an electronic nose based approach. Measurement Science and Technology. 14:190-198.

Egg Regulations. 2008. Egg Regulations. C.R.C., c.284.

Ellis, J.W. 1929. Molecular absorption spectra of liquids below 3 μ.. Transactions of the Faraday Society. 25: 888-898.

Franchini, A., F. Sirri, N. Tallarico, G. Minelli, N. Iaffaldano, and A. Meluzzi. 2002. Oxidative stability and sensory and functional properties of eggs from laying hens fed supranutritional doses OF vitamins E and C. Journal of Poultry Science. 81:1744-1750.

Froning, G.W. 1998. Recent advances in egg products research and development. In “Egg processing workshop”. (Riverside and Modesto ed.). University of California, USA.

Giangiacomo, R., and G.G. Dull. 1986. NIR spectrophotometric determination of individual sugars in aqueous mixtures. Journal of Food Science. 51: 679-683.

Gielen, R.M.A.M., L.P. De Jong, and H.M.M. Kerjvkiet. 1979. Electro-optical blood-spot detection in intact eggs. IEEE Transactions on instrumentation and measurements. IM-28(3): 177-183.

Goel, P.K. 2003. Hyper-spectral remote sensing for weed and nitrogen stress detection. Ph.D. Thesis. McGill University. Montreal. QC. pp. 69-137.

Haugh, R.R. 1937. A new method for determining the quality of an egg. U.S. Egg Poultry. 39: 27-49.

Hildago, A., M. Rossi, and C. Pompei. 1995. Furosine as a freshness parameters of shell eggs. Journal of Agriculture and Food Chemistry. 43: 1673-1677.

Hildago, A., M. Rossi, and C. Pompei. 2006. Estimation of equivalent egg age through furosine analysis. Journal of Food Chemistry. 94: 608-612.

http://www.eggs.ab.ca/egg_industry/EGGSLIGHT2.jpg. Visited at 28-8-2008
http://www.enchantedlearning.com/subjects/birds/label/chickenegg2/label.IF. Visited at 28-08-2008.

Jacob, J.P, R.D. Miles, and F.B. Mather. 2000. Egg quality. University of Florida, IFAS extension.

Karaoui, R., B. Kemps, F. Bamelis, B. De Katelaere, K. Merten, R. Schoonheydt, E. Decuypere, and J. G. De Baerdemaeker. 2006. Development of a raid method based on front face fluorescence spectroscopy for the monitoring of egg freshness: 1-evolution of thick and thin egg albumens. Europe Food Researche Technology. 223:303-312.

Kemps, B., F. Bamelis, B. De Ketelaere, K. Mertens, K. Tona, E.M. Decuypere, and J.G. De Baerdemaeker. 2006. Visible transmission spectroscopy for assessment of egg qulaity. Journal of the Science of Food and Agriculture. 86:1399-1406.

Kemps, B., B. De Ketelaere, F. Bamelis, K. Mertens, E.M. Decuypere, J.G. De Baerdemaeker, and F. Schwägelet. 2007. Albumen freshness assessment by combining visible Near-Infrared Transmission and Low-Resolution Proton Nuclear Magnetic Resonance Spectroscopy. Journal of Poultry Science. 86: 752-759.

Kröckel, L., R. Poser, and F. Schwägele. 2005. Low Resolution Proton Nuclear Resonance Spectroscopy: Application for the determination of the internal quality of intact EGGS microbiological aspects. Fleischwirtschaft. 85(10): 108-112.

Kuchida, K., M. Fukaya, S. Miyoshi, M. Suzuki, and S. Tsuruta. 1999. Non-destructive prediction method for yolk:albumen ratio in chicken eggs by computer image analysis. Journal of Poultry Science. 78: 909-913.

Laghi, L., M.A. Cremonini, G. Placucci, S. Sykora, K. Wright, and B. Hills. 2005. A proton NMR relaxation study of hen egg quality. Magnetic Resonance Imaging. 23: 501-510.

Li-Chan, E., and S. Nakai. 1989. Biochemical basis for the properties of egg white. Critical Reviews in Poultry Biology. 2(1): 21-59.

Li-Chan, E. C.Y., W.D. Powrie, and S. Nakai. 1995. The chemistry of eggs and egg products. In ``Egg Science and Technology. (W.J.S.O.J. Cotterill, ed). Food products Press. An Imprint of The Hawarth Press, Inc. New York. London.

Liu, Y., Y. Ying, A. Ouyang, and Y. Li. 2007. Measurement of internal quality in chicken eggs using visible transmittance spectroscopy technology. Food Control. 18: 18 – 22.

Lucisano, M., A. Hidalgo, E.M. Comelli, and M. Rossi. 1996. Evolution of chemical and physical albumen characteristics during the storage of shell eggs. Journal of Agricultural and Food Chemistry. 44:1235-1240.

Maghirang, E.B., F.E. Dowell, J.E. Baker, and J.E. Throne. 2003. Automated detection of single wheat kernels containing live or dead insects using Near-Infrared reflectance spectroscopy. Transactions of the ASAE. 46(4): 1277-1282.

Mathews, W.W. 1986. Atlass of descriptive embryology. Macmillan Publishing Co., New York, p.108.

McCaig, T.N. 2002. Extending the use of Visible/near-infrared reflectance spectrophotometers to measure color of food and agricultural products. Food Research International. 35: 731-736.

Mohan, J. 2004. General introduction to electromagnetic spectrum and molecular spectroscopy. In “Organic Spectroscopy: Principles and Applications”. (J. Mohan ed.) Alpha Science International Ltd.

Narushin, V.G., and M.N. Romanov. 2002. Egg physical characteristics and hatchability. World's Poultry Science Journal. 58: 297-303.

Narushin, V.G., T.A. Van Kempen, M.J. Wineland, and V.L. Christensen. 2004. Comparing infrared spectroscopy and egg size measurements for predicting eggshell quality. Journal of Biosystems Engineering. 87:367-373.

Norris , K.H. 1983. Multivariate analysis of raw materials. *In*"Chemistry and World Food Supplies: The New Frontiers ". (L.W. Shemilt ed.). Pergamon press, New York, USA.

Osborne, B.G. 1986. The use of near infrared reflectance spectroscopy in the control of flour protein content. Journal of Chemical Technology and Biotechnology. 36: 364-366.

Osborne, B.G., S.C.W. Hook, and P.I. Maris. 1989. Recent developments in the analysis of cereals by near-infrared reflectancce/transmittance. In "Proceedings of ICC 89 Symposium. Wheat End-use Properties". (H. Salovaara Ed.) University of Helsinki. Finland.

Osborne, B.G., T. Fearn, and P.H. Hindle. 1993a. Aplications of near infrared spectroscopy in food and beverage. *In*"Practical NIR spectroscopy with applications in food and beverage analysis". (DDavid Browning ed.). Harlow: Longman Scientific and Technical.

Osborne, B.G., Fearn, T., and P.H. Hindle. 1993b.Theory of near infrared spectrophotmetry. *In*"Practical NIR spectroscopy with applications in food and beverage analysis". (DDavid Browning ed.). Harlow: Longman Scientific and Technical.

Osborne, B.G., Fearn, T., and P.H. Hindle. 1993c.Near infrared calibration II. *In*"Practical NIR spectroscopy with applications in food and beverage analysis". (DDavid Browning ed.). Harlow: Longman Scientific and Technical.

Perez-Mendoza, J., J.E. Thhrone, F.E. Dowell, and J.E. Baker. 2003. Detection of insect fragments in wheat flour by Near-Infrared Spectroscopy. Journal of Stored Products Research. 39(3): 305-312.

Pitsillis, J.G., D.B. Brooker, H.V. Walton, and O.J. Cotteril. 1984. Rheological properties of liquid egg white. Transactions of the ASAE. 27: 300-304.

Ragni, L., A. Al-Shami, G. Mikhaylenko, and J. Tang. 2007. Dielectric characterization of hen eggs during storage. Journal of Food Engineering. 82: 450-459.

Roberts, J. 2004. Factors affecting egg internal quality and egg shell quality in laying hens. Journal of Poultry Science, 41:161-177.

Robinson, D.S., and J.B. Monsey. 1972. Changes in the composition of ovomucin during liquefaction of thick white. Journal of the Science of Food and Agriculture. 23: 29-38.

Robinson, D.S. 1987. The chemical basis of albumen quality. *In* "Egg Quality-Current Problems And Recent Advances" (B.M. Freeman Ed.). Carfax Publishing Company. Abingdon, England.

Schäfer, A., W. Drewes, and F. Schwägele. 1999. Effect of storage temperature and time in egg white protein. Nahrung. 43(2): 86-89.

Schmilovitch Z., A. Hoffman, H. Egoza and E. Klein, 2002. Determination of egg freshness by NNIRS (near-near infrared spectroscopy), presented at EurAgEng, Budapest, paper No.02-AP-023.

Schwägele, f., R. Poser, and L. Krockel. 2001. Application of low-resolution NMR spectroscopy of intact eggs-measurement of quality determining physical characteristics. Fleischwirtschaft. 81: 103-106.

Scott, T.A., and F. G. Silversides. 2000. The effect of storage and strain of hen on egg quality. Journal of Poultry Science. 79: 1725-1729.

Sedman, 2000. Application of Fourier Transform Infrared Spectroscopy in the Analysis of Edible Fats and Oils. *Ph.D.Thesis*. McGill University. Montreal. QC. pp. 37-90.

Silversides, F.G., and T.A. Scott. 2001. Effect if storage and layer age on quality of eggs from two lines of hens. Journal of Poultry Science. 80:1240-1245.

Silversides, F.G., and P. Villeneuve. 2000. Is the Haugh Unit correlation for egg weight valid for eggs stored at room temperature? Journal of Poultry Science. 73: 50-55.

Spackman, D., 1987. The effect of disease on egg quality. *In* "Egg quality-current problems and recent advances". Butterworths, London, Boston, Dorban, 255-282.

Stadelman, W.J. 1995. Quality identification of shell eggs. *In* "Egg science and tecnology". (W.S.O.J. Cotterill, ed.). Food products Press. An Imprint of the Haworth Press, Inc. New York. London.

Statistics Canada. 2008. Poultry and egg statistics. Catalogue number: 23-015-X.

Thermo Galactic. 2003. PLS Plus IQ^{TM}. User's Guide. Salem, NH, USA.

Tsarenko, P.P. 1973. Apparatus for determination of egg quality. USSR Patent. 290-439.

Todeschini, R., D. Galvagni, J.L. Vílchez, M. Del Olmo, and N. Navas. 1999. Kohonen artificial neural networks as a tool for wacelength selection in multicomponent spectrofluorometric PLS modelling: application to phenol, *o*-cresol, *m*-cresol and *p*-cresol mixtures. Trends in Analytical Chemistry. 18:93-98.

Tung, M.A., J.F. Richards, and B.C. Morrison. 1970. Rheology of fresh, aged and gamma irradiated egg white. Journal of Food Science. 35:872-874.

Uddin, M., and E. Okazaki. 2004. Classification of fresh and frozen-thawed fish by Near-infrared Spectroscopy. Journal of Food Science. 69(8): C665-C668.

Völgyi, F. 2000. Specifying the freshness of eggs using microwave sensors. Subsurface Sensing Technologies and Applications. 1 (1): 119-139.

Watkins, B.A. 1995. The nutritive value of the egg. *In* ”Egg Science and Technology” (W.S.O.J. Cotterill, ed). Food products Press. An Imprint of the Haworth Press, Inc. New York, London.

Wei, R., and J.J. Bitgood. 1989. A new objective measurement of eggshell color. 1. A test for potential usefulness of two color measuring devices. Poultry Science. 69: 1175-1780.

Wetzel, D.L. 1983. Near-Infrared Reflectance Analysis of Major Components in Foods. *In* “Instrumental analysis of Foods-Recent Progress”. (G. C. and G. Inglett ed.). Academic Press Inc. New York, USA.

William, P.C., and K.H. Norris. 1987. Near infrared technology in the agricultural and food industries. American Association of Cereal Chemists Inc., St. Paul, Minnesota, USA.

Williams, P.C., and D.C. Sobering. 1993. Comparison of commercial near-infrared transmittance and reflectance instruments for analysis of whole grains and seeds. Journal of Near Infrared Spectroscopy. 1: 25-32.

Wold, S. 1978. Cross-validatory estimation of the number of components in factor and principal components models. Technometrics. 20: 397-405.

PREFACE TO CHAPTER 4

According to the literature, the partial least squares (PLS) regression is the most widely used model for the analysis of spectral data where a highly overlapped data is present. However, selection of relevant wavelengths and their use in the model is of utmost importance since it improves the predictive ability of the model and reduces its complexity.

Fresh 360 laid eggs collected from the same flock and stored over a period of 16 days. They were stored at 18°C and 55% relative humidity. The spectral data was linked to the measured albumen pH and Haugh Units with the PLS regression models. Maximum R^2 (MAXR) technique was used in order to choose the relevant wavelengths. The aim of this research was to evaluate the power of MAXR techniques to choose the important wavelengths.

Research papers based on the chapter:

Abdel-Nour, N., M. Ngadi, S. Prasher, and Y. Karimi. 2008. Combined maximum R^2 and partial least squares method for wavelengths selection and analysis of spectroscopic data. Applied Spectroscopy. (Submitted for publication).

CHAPTER 4

COMBINED MAXIMUM R^2 AND PARTIAL LEAST SQUARES METHOD FOR WAVELENGTH SELECTION AND ANALYSIS OF SPECTROSCOPIC DATA

4.1 ABSTRACT

The selection of wavelengths in multivariate analysis is of utmost importance in order to build a strong and robust predictive model. The aim of this research was to investigate the feasibility of an automated selection of sets of relevant wavelengths in Visible/Near Infrared spectroscopy by combining the maximum R^2 (MAXR) method with partial least squares (PLS) regression (MAXR-PLS) to build a PLS predictive model. The data used to test this method were derived from the determination of albumen pH and Haugh Units (HU) as tools for testing the egg freshness. For this purpose, 360 eggs were stored during 16 days under a temperature of 18°C and a relative humidity of 55%. For each egg, the VIS/NIR transmission spectra and the two most widely used methods for the assessment of egg quality namely the HU and the albumen pH were performed. A PLS model was built using the full spectra and compared with the models built by selected wavelengths using MAXR-PLS method. Using the mentioned method, the correlation coefficients between the measured and predicted values were up to 95% and the root mean square error for cross-validation (RMSECV) were 0.05 and 5.05 for pH and HU, respectively. In addition, this method reduces the complexity of the models by

reducing the latent variables (LV). Despite the complexity of the spectral data, the Maximum R^2 method leads to a robust predictive model that uses the informative wavelengths.

4.2 INTRODUCTION

Visible (VIS) and Near Infrared (NIR) spectroscopy is a powerful, fast and non-destructive technique that is increasingly used for measuring a large number of chemical and physical properties of agricultural products. Successful applications have been reported for the determination of soluble solids in cherry and apricot (Carlini et al., 2000), dry matter in onions (Birth et al., 1985) and egg freshness (Kemps et al., 2006). However, extracting the useful information from spectral data is the fundamental challenge. Therefore, there is a need for automated techniques to extract relevant information. The partial least squares (PLS) method is often used in spectroscopy to analyze spectral data containing overlapping absorption peaks, interference effects from diffuse light scatter and noise from the hardware used to collect the data (Osborne et al., 1997). The choice of wavelengths should be adequate to build a strong model using PLS with good predictive ability and excluding those wavelengths that are irrelevant to the model.

The choice of the wavelengths is required to establish a calibration model with minimum errors in prediction. The benefits gained from wavelength selection are the stability of the model to the collinearity in multivariate spectra as well as the interpretability of the relationship between the sample composition and the model (Jiang et al., 2002). Bangalore et al. (1996) studied the feasibility of coupling genetic algorithm methods for the selection of wavelengths with partial least squares regression for analysing spectral data. They found that the results obtained after selection were better than those obtained with no spectral range selection. Du et al. (2004) used the changeable size window partial least squares and searching combination moving window partial least squares and they found that the combination of these two methods improved the prediction ability of the PLS model. Todeschini et al. (1999) proposed the use of Kohonen artificial neural networks (K-ANN) for selecting a set of wavelengths. The results have shown that the predicting ability of the PLS model was improved. Perez-Mendoza et al. (2003) used beta coefficient for the selection of important wavelengths to build a PLS model for the classification of flours with and without insect fragments. Ventura et al. (1998) used the multiple linear regression (MLR) procedure to select the best wavelengths for determination of soluble solids in apple.

The aim of this research was to develop a method with the capacity to select the informative and relevant wavelengths. With this method we attempted to combine the maximum R^2 (MAXR) method with PLS regression (MAXR-PLS). The specific objectives of this study were to select sets of wavelengths and eliminate the uninformative wavelengths, improve the predictive ability of the PLS model by using the optimal set of wavelengths and decrease the complexity of the model by decreasing the latent variables (LV) used in the model.

4.3 THEORY AND ALGORITHM

4.3.1 Partial least squares (PLS)

PLS is a quantitative spectral decomposition technique that is advantageous as it performs the decomposition on both spectral and concentration data. Two sets of eigenvectors are generated when the calibration spectral data are processed using the PLS method. These are a set of spectral loadings corresponding to the common variation in spectral data and a set of spectral weights corresponding to the changes in spectra due to the differences in concentration. The method assigns a set of scores for spectral data and for concentration data. As a result, the spectra containing higher concentration of the components are weighted more than those with low concentration (Thermo Galactic, 2003).

Having 2 data sets (blocks of variables): *N*-dimensional and *M*-dimensional space of variables, the PLS models the relations between these two blocks of variables. After observing n data samples from each block of variables, PLS decomposes the (*n x N*) matrix of zero-mean variables *X* and the (n x M) matrix of zero-mean variables *Y* into the form in Equations (4.1) and (4.2):

$$X = TP^t + E \quad (4.1)$$

$$Y = UQ^t + F \quad (4.2)$$

Where T and U are (*n x p*) matrices of the *p* extracted latent variables, P is the (*N x p*) matrix, Q is the (*M x p*) matrix. Q and P represent the matrices of loadings. E is the (*n x N*) matrix and F is the (*n x M*) matrix where E and F represent the matrices of residuals (Rosipal and Krämer, 2006).

Validation of the PLS calibrated model is performed by leave-one-out cross-validation technique where the same spectra used as training set are predicted back against the same model. This means that with m numbers of calibration samples, the model is built with m-1 samples and the m^{th} sample is predicted as unknown sample. The selection of latent variables (LV) number when building the model is of paramount importance: more variables selected cause an overfitting, while less variables causes underfitting. From the statistical point of view, the number of samples must be equal or more than five times the number of LVs. The prediction residual error sum of squares (PRESS) is a commonly used method for the selection of the LV's number. It is plotted against the number of factors. The plot falls down to a minimum corresponding to the best number of LVs on PLS calibration model. It is calculated using the Equation (4.3):

$$\text{PRESS(r)} = \sum_{i=1}^{n} (\hat{y}_i - y_i)^2 \quad (4.3)$$

Where $\hat{y}_i$ is the prediction of the concentration of interest in calibration sample i, y_i is the measured value in calibration sample i, and PRESS(r) is the PRESS value obtained with r factors.

4.3.2 Multiplicative scatter correction (MSC)

In order to remove background and noise from NIR spectra, MSC has been proposed as a pre-treatment technique before the PLS calculation. The MSC technique is a transformation that allows the removal of amplification and offset effect from the spectra. In principle, this estimation should be applied only on the part of the spectrum that is influenced by light scattering. In practice the whole spectrum is sometimes used. For the MSC an ideal or reference spectrum is required. As an estimate to that ideal, the average of the calibration set can be used. Each spectrum is regressed in the set-mean spectrum; the effect of scatter is responsible for variations along a straight line whereas deviations from the line correspond to the absorption by the sample components (Blanco et al., 1997). The main multiple linear regression is cited in Equation (4.4):

$$X_{ik} = a_i + b_i r_k + e_{ik} \qquad (4.4)$$

Where a_i represents the "common shift" and is related to proportional additive effect, b_i represents the "common amplification" and is related to multiplicative effect, e_{ik} represents the residuals and is related to the chemical information.

The corrected spectrum is calculated by using Equation (4.5):

$$X_{ij}(MSC) = \frac{(Xij - a)}{b} \quad j = 1, 2, ..., p \qquad (4.5)$$

4.3.3 Maximum R^2 (MAXR)

MAXR technique is used in order to choose sets of wavelengths which contain information about the variables. Goel et al. (2003) used the MAXR criterion with PROC REG procedure of SAS software to choose the best model for estimations of various crop biophysical parameters. The MAXR technique finds the best one-variable, best two-variable and so on which produces the model with highest R^2 value. It begins by finding the best one-variable model. A second variable is added to the model in a manner to increase the R^2 values and to build the two-variable model. The best two-variable model is chosen by removing one variable in the model and replacing it one by one, with each of the other variables until it produces the best R^2. In the same way, it builds the best three-variable and four-variable model and so on. This process is called "compare and switch".

4.3.4 Root mean square error based on cross-validation

The root mean square prediction error based on cross-validation (RMSECV) is an estimate of the standard error in prediction and is calculated by using Equation (4.6):

$$\text{RMSECV} = \sqrt{1/N \sum_{i=1}^{N} (\hat{y}_i - y_i)^2} \tag{4.6}$$

Where $\hat{y}_i$ is the predicted concentration value of the i-th calibration sample and $\hat{y}_i$ is the measured value of the i-th calibration sample.

The RMSECV is used, in addition to R^2, to compare the predictive models of the different sets of wavelengths. The RMSECV indicates the mean difference between the measured and predicted values (Kobayshi and Salam, 2000). The predictive ability of the model will be better with smaller RMSECV values: the smaller the RMSECV value for a model the stronger the prediction ability of the model will be (Todeshini et al., 1999).

4.4 MATERIAL AND METHODS

4.4.1 Visible/Near-infrared spectroscopic data

VIS/NIR transmission data of 360 intact white-shell leghorn eggs were obtained using a spectroradiometer (FieldSpec® Pro, Analytical Spectral Devices, Boulder, CO, USA) in 2151 wavebands. The spectroradiometer measures transmittance at wavelengths from 350 to 2500 nm with 1 nm increment. Four halogen lamps 500 watts each were used as a light source. A white chamber was placed between the light and the samples to avoid heating up the samples and to uniformly distribute the light. A pure white standard was used for calibration. Each egg was scanned 3 times and averaged. After the transmission spectra were determined, the eggs were weighed (±0.01mg) and broken. The albumen height was measured using a mounted digital Vernier Caliper (Marathon Watch Company Ltd., On, Ca). These measurements allowed determination of the Haugh Units (Haugh, 1937) using Equation (4.7):

$$HU = 100 \log_{10} (h\text{-}1.7w^{0.37}+7.6) \tag{4.7}$$

Where *HU* is the Haugh Units, *h* is the observed height of albumen in millimetres and *w* is the weight of the egg in grams.

After separating the albumen from the yolk, the albumen pH was determined using a pH-meter (Accumet Basic AB15 pHmeter, Fisher Scientific, USA).

4.4.2 Method description

The flowchart that described the method used for selecting the wavelengths is shown in Figure 4.1.

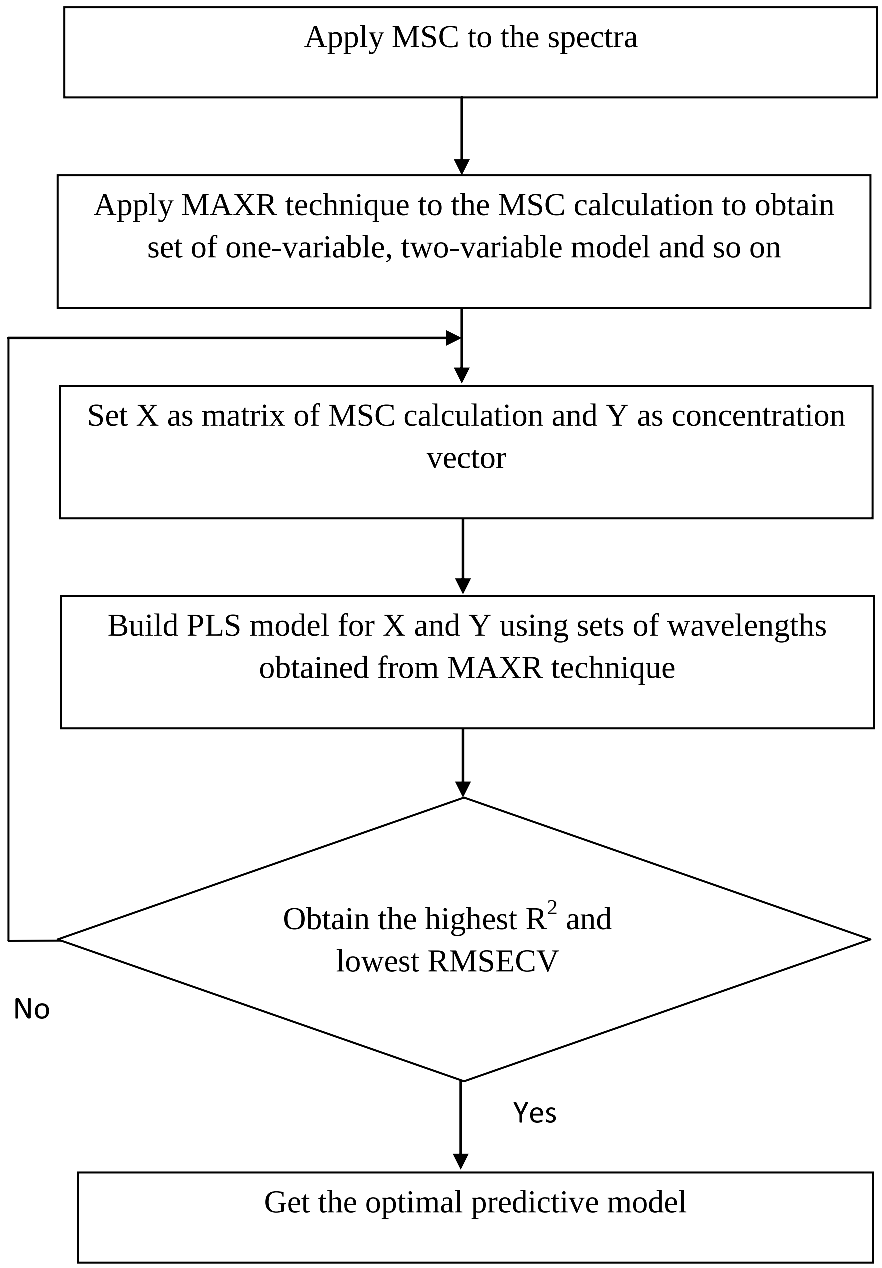


Figure 4.1 Flowchart for execution of selection wavelengths. Abbreviations used are: MSC is multiplicative scatter correction, MAXR is the maximum R^2 method, PLS is the partial least square and RMSECV is the root mean squared error of cross validation.

In the beginning, full MSC is applied to the whole spectra, and then the MAXR technique was used to choose the appropriate sets of wavelengths. After that, a PLS model was built using the different sets of wavelengths. Finally the models were tested to determine their robustness and good predictive ability.

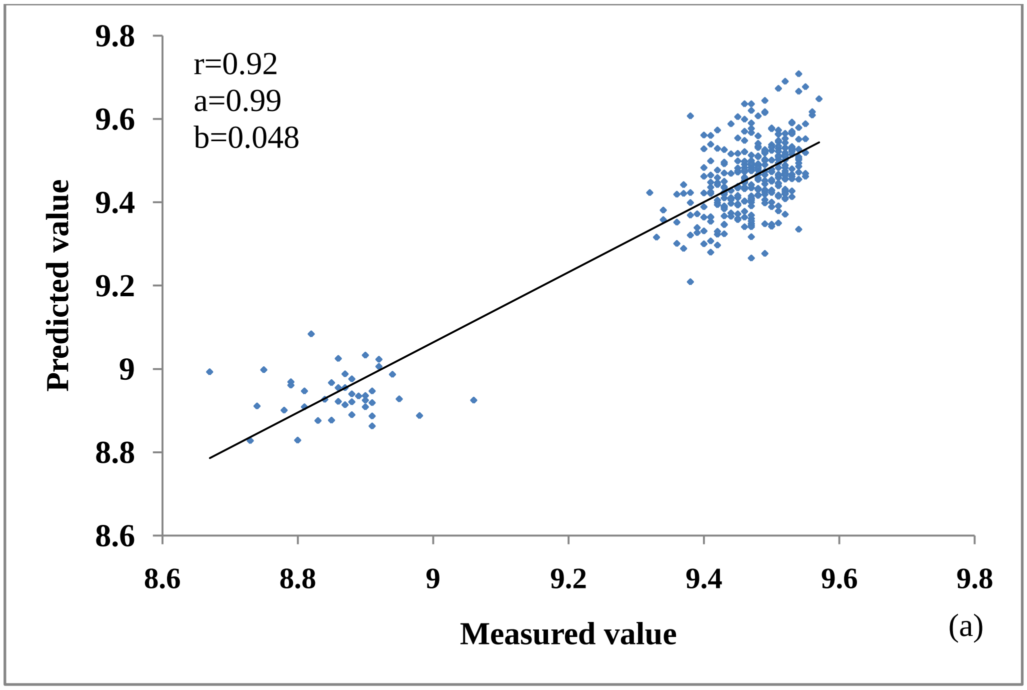
4.4.3 Data processing

VIS/NIR Spectral data were analysed using SAS® (version 9.1, 2003) statistical software package and The Unscrambler® (version 9.7, 2007).

Firstly, full multiplicative scatter correction (MSC) was performed using The Unscrambler program. MAXR technique in SAS program was used to identify the most informative wavelengths. After choosing the sets of relevant wavelengths, the VIS/NIR spectral data were linked to HU and albumen pH by a Partial Least Squares, type 1 (PLS1) regression model. Multivariate analysis of the samples was performed in The Unscrambler program. The models obtained were validated by using the leave-one-out cross-validation method. In the leave-one-out method, all data are used to build the model except the one used to test this model. The slope and the intercept of regression lines were calculated and compared with their ideal values of 0 and 1, respectively using a statistical program called IRENE® (version Beta 1.00, 2003) (Fila et al., 2003).

4.5 RESULTS AND DISCUSSIONS

The transmittance measurements from 350 to 400 as well as from 1751 to 2500 nm were not included in the analysis because of extreme noise. Sets of wavelengths containing less than 8 wavelengths gave low coefficient correlations and R^2 as well as higher RMSECV. Therefore the PLS models was judged not to have a strong predictive ability. Thus, following results focused on the sets containing 8, 9, 10 and full spectra 1350 wavelengths.



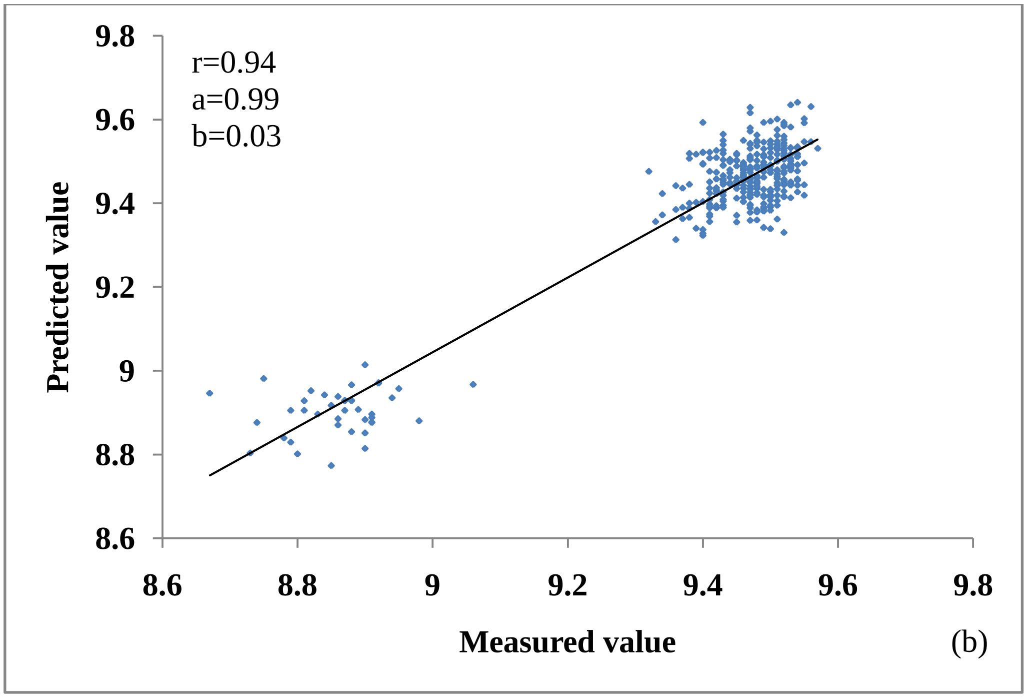
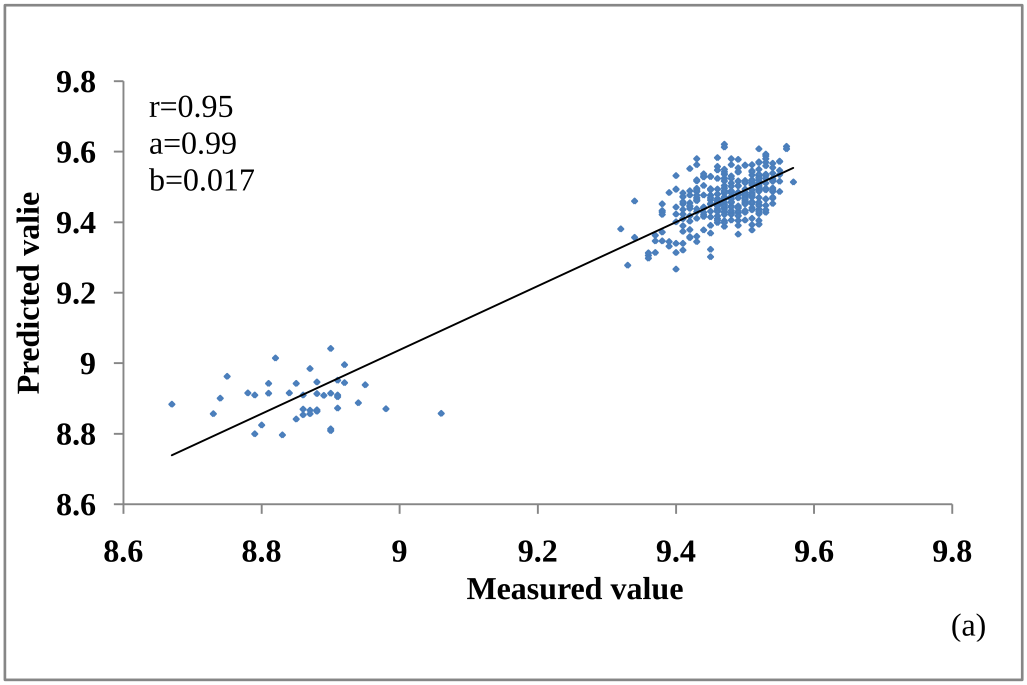


Figure 4.2 Relationship between measured and predicted albumen pH obtained with (a) 8 and (b) 9 wavelengths using MAXR-PLS method.



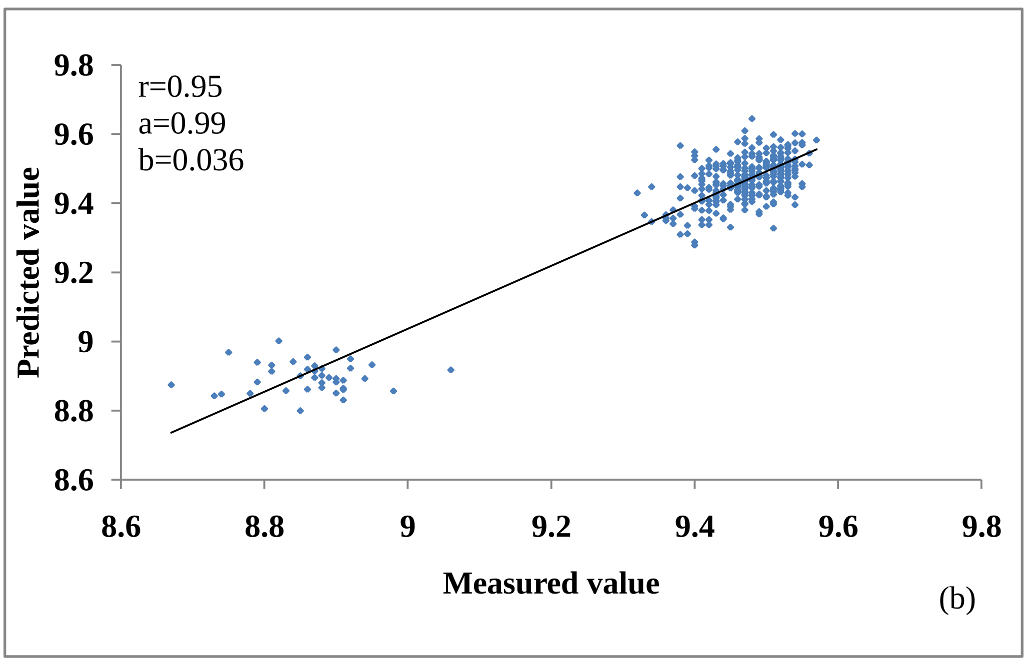


Figure 4.3 Relationship between measured and predicted albumen pH obtained with (a) 10 wavelengths and (b) full spectra using MAXR-PLS method.

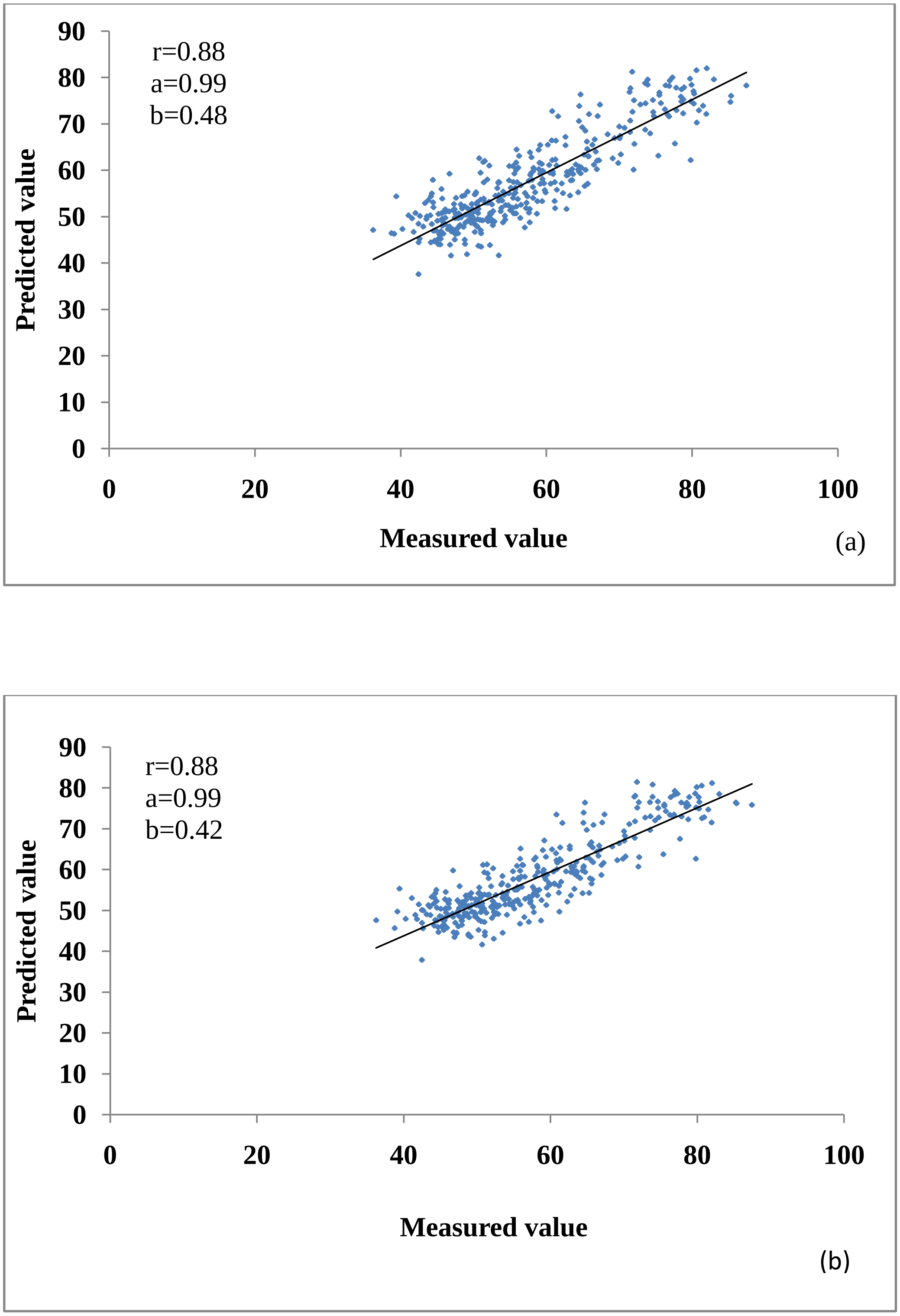
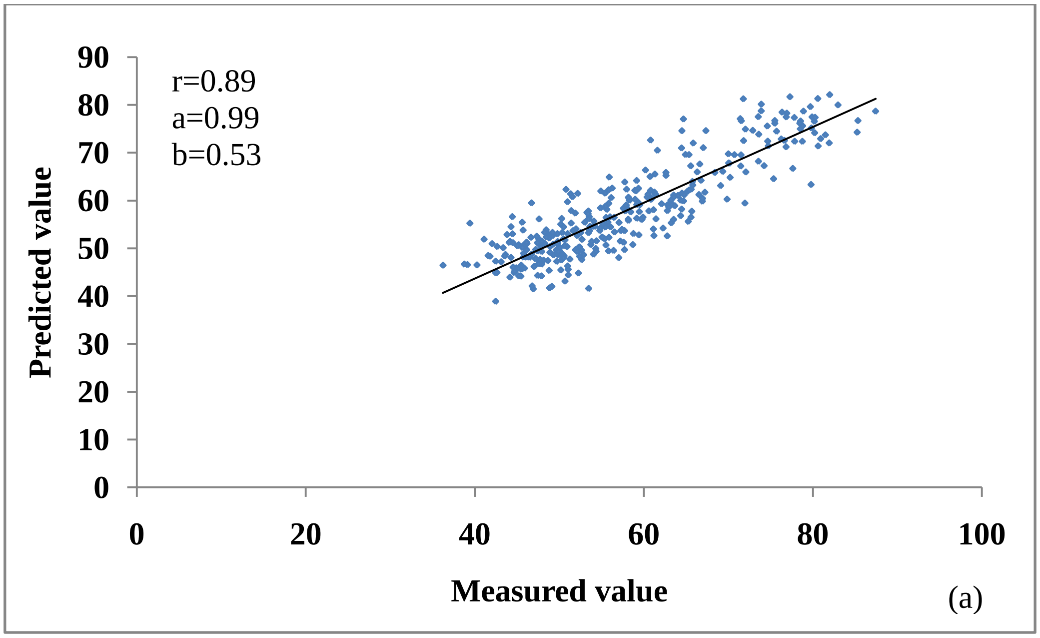


Figure 4.4 Relationship between measured and predicted Haugh Units obtained with (a) 8 and (b) 9 wavelengths using MAXR-PLS method.



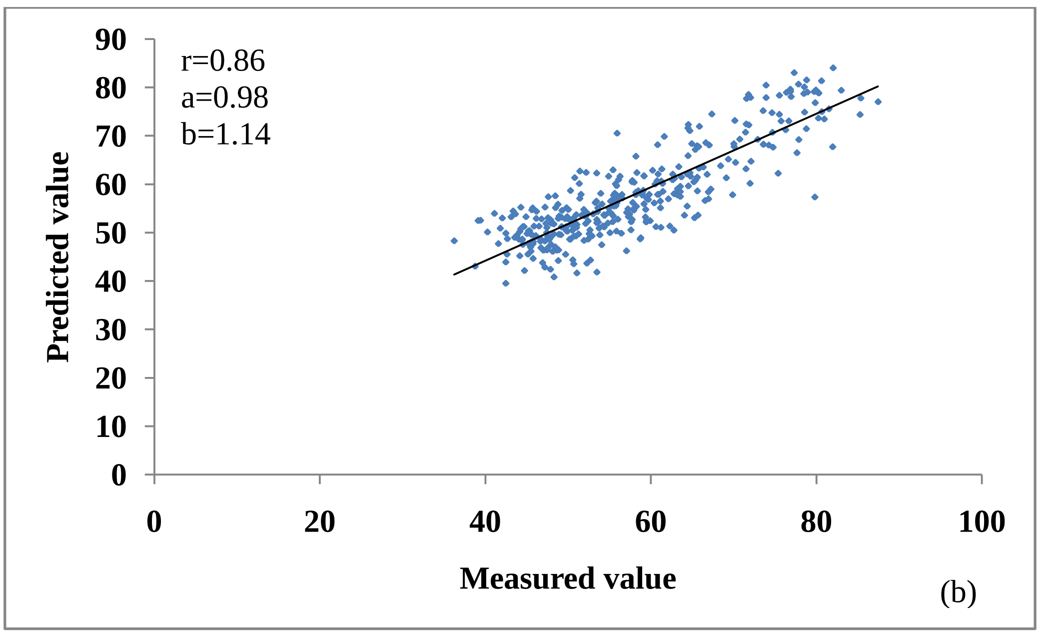


Figure 4.5 Relationship between measured and predicted Haugh Units obtained with (a) 10 wavelengths and (b) full spectra using MAXR-PLS method.

Figures 4.2, 4.3, 4.4 and 4.5 show the plot of predicted versus measured for albumen pH and HU, respectively with different number of wavelengths chosen using MAXR-PLS method and the full spectra. The parameters *r, b* and *a* are the correlation coefficient, slope and intercept, respectively obtained by least-squares regression. Figures 4.2 and 4.3 show that there are 2 groups of scatter due to the difference of temperature between day 0 and the other days. It also shows that the best result obtained for *r*, *a* and *b* was by using 10 wavelengths in building the PLS model. The results obtained with 10 wavelengths and full spectra for building the model are the same in terms of the coefficient correlation and the slope but using the relevant wavelengths improves the intercept. As a result, the PLS could be built with 10 wavelengths instead of the full spectra with improved robustness.

As a comparison between the sets of wavelengths and the full spectra, Figure 4.3 shows that by using MAXR-PLS method, the 3 parameters *r*, *a* and *b* have numerically improved. On the other hand, it can be seen that the best coefficient correlation is obtained by using 10 wavelengths whereas the best intercept is obtained by using 9 wavelengths. Addition of wavelengths to the model constituted with 10 wavelengths for the prediction of albumen pH and HU did not improve these three parameters. The reason for this is that the presence of uninformative wavelengths decreases the predictive ability. As a result, the MAXR-PLS method improved the PLS model built with selected wavelengths for the prediction of HU and decreased the complexity of the model for prediction of albumen pH.

For both the HU and the albumen pH for all the sets of wavelengths used, and after tested in the IRENE program using Least Squares method, the results showed that there was no statistical difference between the value of slope and intercept obtained from the predictive model and their ideal (0 and 1 for the intercept and slope, respectively).

The number of wavelengths used in MAXR, PLS and MAXR-PLS methods is shown in Tables 4.1 and 4.2. Every set of wavelength chosen used

MAXR-PLS method is used to build PLS models with the calibration set and then the prediction set to test the performance of these models. The calculated root mean square error of cross validation (RMSECV) and the R^2 are also listed.

Table 4.1

Prediction results for albumen pH obtained with maximum R^2 (MAXR), partial least squares (PLS) and MAXR-PLS methods

Method	Number of wavelengths	Latent Variable number	R^2	RMSECV
MAXR	8	-	0.84	0.07
	9	-	0.88	0.06
	10	-	0.89	0.06
PLS	All spectra (1350)	7	0.90	0.06
MAXR-PLS	8	6	0.84	0.08
	9	6	0.89	0.07
	10	7	0.90	0.06

Table 4.2

Prediction results for Haugh Units (HU) obtained with maximum R^2 (MAXR), partial least squares (PLS) and MAXR-PLS methods

Method	Number of wavelengths	Latent Variable number	R^2	RMSECV
MAXR	8	-	0.78	5.17
	9	-	0.78	5.10
	10	-	0.79	5.06
PLS	All spectra (1350)	6	0.74	5.51
MAXR-PLS	8	6	0.78	5.14
	9	6	0.78	5.11
	10	7	0.79	5.05

From these 2 tables, it can be observed that when the number of wavelengths increases from 8 to 10, the R^2 increases whereas the RMSECV decreases. The number of LVs in PLS and MAXR-PLS methods increases with an increasing number of wavelengths. These results are similar to those obtained by Todeschini et al. (1999) who found that when the number of wavelengths decreased, the dimension of the model decreased but the predictive ability was still high. The minimum value of RMSECV and the maximum value of R^2 for the albumen pH and HU are obtained using 10 wavelengths with 7 and 6 LVs, respectively in the MAXR-PLS methods. These results are in contradiction to Kemps et al. (2006) who found that the relevant information concerning egg freshness (albumen pH and HU) are detected in 6 wavelengths. Table 4.1 shows that the model used to predict the pH built from PLS regression had better R^2 and RMSECV than the model built using MAXR method, whereas, Table 4.2 showed the opposite for HU. For the reasons cited before, the multivariate analysis is used often in spectroscopy due to its ability to overcome the collinearity problems. Combining PLS and MAXR methods could be useful. Therefore, the results obtained from these two tables showed that the PLS model can be built with selective wavelengths using MAXR method and not necessarily with all the spectra. In addition, the robustness and the predictive capability of the model can be improved with selective wavelengths.

Table 4.2 showed that constructing a predictive model for HU with all spectra required 6 LVs and the RMSECV value was 5.51 whereas using the MAXR-PLS method required 7 LVs and the RMSECV value was 5.05. Therefore, the inclusion of uninformative wavelengths can lead to collinearity and the PLS model becomes less stable than the model based on informative wavelengths. It is evident that the whole original spectral matrix does not allow satisfactory predictions for samples. Thus, the use of PLS method alone with all spectral, the VIS/NIR spectroscopy cannot be used to accurately classify the HU for eggs. The proposed MAXR-PLS method selected the most informative wavelengths and provided a robust predictive model with only 10 wavelengths.

4.6 CONCLUSION

In this article, the usefulness of a new method named as MAXR-PLS for the building of a PLS predictive model was investigated. The results presented above demonstrated that this method is a good tool for this purpose. For the HU and albumen pH, the results showed an improvement in prediction with 10 wavelengths compared to those obtained with the full spectrum. The RMSECV for the model built with 10 wavelengths and the whole spectra for the prediction of HU were 5.05 and 5.51, respectively. The R^2 of the model predicting the HU for the selected wavelengths and the full spectra were 0.89 and 0.86, respectively. As well, there was a numerical improvement of the correlation coefficient, slope and intercept. By eliminating the less informative wavelengths and selecting the relevant ones using this method, the PLS models can improve the predictive ability, decrease the model complexity and reducing the time of analysis.

4.7 ACKNOWLEDGEMENTS

The authors will like to acknowledge the Poultry Industry of Canada for funding this study.

4.8 REFERENCES

Bangalore, A.S., R.E. Shaffer and G.W. Small. 1996. Genetic algorithm-based method for selecting wavelengths and model size for use with partial least-squares regression: application to near-infrared spectroscopy. Analytical Chemistry. 68: 4200-4212.

Birth, G.S., G.G. Dull, W.T. Renfore, and S.J. Kays. 1985. Nondestructive spectrophotometric determination of dry matter in onions. Journal of the American Society for Horticultural Science. 110(2): 297-303.

Blanco, M., J. Coello, H. Iturriaga, S. Maspoch, and C. De La Pezula. 1997. Effect of data preprocessing methods in near-infrared diffuse reflectance spectroscopy for the determination of the active compound in a pharmaceutical preparation. Applied Spectroscopy. 51(2): 240-246.

Carlini, P., R. Massantini, and F. Mencarelli. 2000. Vis-NIR measurement of soluble solids in cherry and apricot by PLS regression and wavelength selection. Journal of Agriculture and Food Chemistry. 48: 5236-5242.

Du, Y.P., Y.Z. Liang, J.H. Jiang, R.J. Berry, and Y. Ozaki. 2004. Spectral regions selection to improve prediction ability of PLS models by changeable size moving window partial least squares and searching combination moving window partial least squares. Analytica Chimica Acta. 501: 183-191.

Fila, G., A. Bellochi, M. Acutis, and M. Donatelli. 2003. IRENE: a software to evaluate model performance. European Journal of Agronomy. 18: 269-372.

Goel, P.K. 2003. Hyper-Spectral remote sensing for weed and nitrogen stress detection. *Ph.D.* Thesis. McGill University. Montreal. QC. pp. 238-256.

Haugh, R.R. 1937. The Haugh Unit for measuring egg quality. US Egg Poultry Magazine. 43:552-555, 572-573.

IRENE, 2003. Integrated Resources for Evaluating Numerical Estimates. Ver. beta 1.00. Bologna, Italy.

Jiang, J., R.J. Berry, H.W. Siesler, and Y. Ozaki. 2002. Wavelength selection in multicomponent spectral analysis by moving window partial least-squares regression with applications to mid-infrared and near-infrared spectroscopic data. Analytical Chemistry. 74: 3555-3565.

Kemps, B.J., F.R. Bamelis, B. De Katelaere, K. Mertens, K. Tona, E.M. Decuypere, and J.G. De Baerdemaeker. 2006. Visible transmission spectroscopy for the assessment of egg freshness. Journal of the Science of Food and Agriculture. 86: 1399-1406.

Kobayshi, K. And M.U. Salam. 2000. Comparing simulated and measured values using mean square deviation and its components. Agronomy Journal. 92: 345-352.

Osborne, S.D., R.B. Jordan, and R. Künnemeyer. 1997. Method of wavelength selection for partial least squares. Analyst. 122: 1531-1537.

Perez-Mendoza, J., J.E. Throne, F.E. Dowell, and J.E. Baler. 2003. Detection of insect fragments in wheat flour by near-infrared spectroscopy. Journal of Stored Products Research. 39(3): 305-312.

Rosipal, R. and N. Krämer. 2006. Overview and recent advances in partial least squares. SLSFS 2005, LNCS 3940, pp. 34-51.

SAS. 2003. SAS User's Guide: statistics. Ver 8.2. Cary, N.C.: SAS Institute, Inc.

Thermo Galactic. 2003. PLS Plus IQ™. User's Guide. Salem, NH, USA.

The Unscrambler. 2007. The Unscrambler User's Guide: ver. 9.7. Woodbridge, N.J., USA: CAMO Software AS.

Todeschini, R., D. Galvagni, J.L. Vílchez, M. Del Olmo, and N. Navas. 1999. Kohonen artificial neural networks as a tool for wacelength selection in multicomponent spectrofluorometric PLS modelling: application to phenol, *o*-cresol, *m*-cresol and *p*-cresol mixtures. Trends in Analytical Chemistry. 18:93-98.

Ventura, M., A. De Jager, H. De putter, and F.P.M.M. Roelofs. 1998. Non-destructive determination of soluble solids in apple fruit by near infrared spectroscopy. Postharvest Biology and Technology. 14(1): 21-28.

PREFACE TO CHAPTER 5

The development of combined maximum R^2 method with partial least squares (PLS) model for the selection of informative wavelengths and analysis of visible and near infrared spectral data was explained in Chapter 4.

In Chapter 5, spectral observations from both the visible and near infrared region of 360 eggs stored for 16 days at 18°C and 55% relative humidity were performed. The combined maximum R^2 method with PLS model was applied in order to link spectroscopic data to the most widely used destructive methods for the assessment of egg quality, namely the HU and albumen pH, and to the number of storage days. The aim of this research was to investigate the potential of VIS/NIRS in the prediction of egg freshness in terms of storage days and egg quality in terms of HU and albumen pH.

Research papers based on this chapter:

Abdel-Nour, N., M. Ngadi, S. Prasher, and Y. Karimi. 2008. Prediction of egg freshness and albumen quality using visible near infrared spectroscopy. Food and Bioprocess Technology. (Submitted for publication).

CHAPTER 5

PREDICTION OF EGG FRESHNESS AND ALBUMEN QUALITY USING VISIBLE/NEAR INFRARED SPECTROSCOPY

5.1 ABSTRACT

Important changes occur in egg during storage, leading to loss of quality. Prediction of these changes is critical in order to monitor egg quality and freshness. The aim of this research was to evaluate application of visible (VIS) and near infrared (NIR) spectroscopy as a rapid, non-destructive and on-line technique for the assessment of egg quality. Three hundred and sixty intact white-shelled eggs freshly laid by the same flock of hens fed with a standard feed were obtained. They were put under controlled conditions of temperature and humidity (T=25°C and RH=55%) for 16 days of storage. On days 0, 2, 4, 6, 8, 10, 12, 14 and 16, 40 eggs were selected and the transmission spectral data was obtained in the range from 350 to 2500 nm. The non-destructive spectral data was compared to egg sample's Haugh Units (HU) and albumen pH in terms of quality and to the number of storage days in terms of freshness. A partial least squares predictive model was developed and used to link the destructive assessment methods and the number of storage days with the spectral data. The correlation coefficient between the measured and predicted values of HU, albumen pH and number of storage days were up to 0.94, R^2 was up to 0.90 and the root mean square error values for the validation (RMSEV) were 5.05, 0.06 and 1.65, respectively. These results showed that the VIS/NIR transmission spectroscopy is a good tool for the assessment of egg freshness

and albumen quality and can be used as a non-destructive method for the prediction of HU, albumen pH and number of storage days. In addition, the relevant information about these parameters was in the VIS and NIR ranging from 411 to 1729 nm.

5.2 INTRODUCTION

Eggs are very important components in the human diet because they are affordable and nutritious. Fresh eggs are fragile and their quality declines with time. However, the number of storage days is not the only factor that influences egg quality. One of the most important contributors to egg quality is the albumen.

Albumen quality is not consistent and varies amongst eggs. It is influenced by many factors such as temperature, environmental conditions and hen age as well as freshness. One of the best quality indicators of storage conditions and time is albumen thinning. When a freshly laid egg is broken onto a smooth flat surface, the yolk is generally in a central position surrounded by thick albumen. However, the thick albumen becomes thinner with time resulting in a larger area and displacing the yolk to one side (Karoui et al., 2006). Many theories have been suggested to explain albumen thinning or liquefaction but there is no single reason that explains this phenomenon. This liquefaction could occur from protease enzymes, depolymerisation by hydroxyl ions at increasing pH values or reduction by thiol type reducing agents and the interaction of ovomucin-lysosyme complex. After being laid, the pH of albumen changes and leads to a destabilization of ovomucin-lysozyme interaction (Robinson and Monsey, 1972). Proteolytic enzymes, hydroxyl ions and disulphide bonds depolymerise ovomucin, leading to the albumen thinning (Wells and Norris, 1987).

Several methods are used for the assessment of egg quality. These methods are divided into two groups: destructive and non-destructive methods. The most widely used destructive method is the Haugh Units (Haugh, 1937), which is based on the relationship between the weight of the intact egg and the albumen height. After being laid, as time passes, the weight of the intact egg decreases due to the loss of water from the egg and the albumen height decreases due to changing of albumen viscosity. This decrease in egg weight and albumen height leads to an increase in the HU. Albumen pH is another parameter used to estimate egg quality. In freshly laid eggs, the albumen has a pH range of 7.6 to 7.9 and has a cloudy appearance due to the presence of carbon dioxide (Dutta et al., 2003). Within a short period of time, albumen pH increases to 9.5 due to the release of carbon dioxide (Lapao et al., 1999). The buffering capacity of the albumen becomes weak between a pH ranging from 7.5 to 8.5. This weakness can lead to a rapid increase in pH during the first few days of storage (Karoui et al., 2006).

The main disadvantage of destructive methods is the need to perform measurements directly on the albumen (Kemps et al., 2007). An example of non-destructive methods is Fourier transform near-infrared (FT-NIR) spectroscopy. Berardinelli et al. (2005) used FT-NIR for the assessment of shell egg and albumen height. The authors predicted the height of thick albumen in 80% of the samples with a maximum error of ±1mm.

VIS/NIR spectroscopy is increasingly used for testing the quality of many agricultural products because it is rapid, non-invasive and could be used on-line. Successful use of this technique has been reported when assessing the internal quality of fruit and vegetables. Research has been carried out to determine the dry matter in onions (Birth et al., 1985), the quality characteristics of mandarin (Gómez et al., 2006) and the quality of kiwi fruit (Slaughter and Crisosto, 1998). Liu et al. (2007) measured visible transmittance spectra and found that the inspection of egg freshness in the range of 400-600 nm is feasible. Kemps et al. (2006) examined the usefulness of the visible transmission spectra in predicting egg freshness. These authors built a Partial Least Square (PLS) regression model in order to link the spectral data with the measured albumen pH and HU. They reported that the correlation coefficients between the measured and predicted albumen pH and HU were 0.86 and 0.82, respectively. Furthermore, they have reported that the relevant information concerning egg freshness was in the range of 570 and 750 nm. Kemps et al. (2007) combined the Visible/Near-infrared transmission and low-resolution proton nuclear magnetic resonance spectroscopy in order to improve the assessment of albumen quality. Their results showed that the prediction with the transmittance spectroscopy alone was better than when the two types of spectroscopy were combined. However, in their studies the spectra ranged from 200 to 1100 nm.

The objective of this research was to investigate the usefulness of the VIS/NIR transmittance spectroscopy (400-2500 nm) as a non-destructive method for the prediction of egg freshness in terms of number of storage days and egg quality in terms of albumen pH and HU. More specific objective is to determine an alternative non-destructive and accurate method that can provide the same information provided by destructive methods.

5.3 MATERIAL AND METHODS

5.3.1 Sample preparation and spectra collection

A total of 360 intact white-shelled leghorn eggs were used for this experiment. These eggs were collected from hens of the same flock fed with a standard ration. The collected eggs were stored under controlled conditions at 18°C and 55% relative humidity in order to cause a variation in albumen pH and HU. Forty eggs were chosen at day 0, 2,4,6,8,10,12,14 and 16 of storage. For each egg, VIS/NIR transmission data were obtained using a spectroradiometer (FieldSpec® Pro, Analytical Spectral Devices, Boulder, CO, USA) in 2151 wavebands. The spectroradiometer measures transmittance at wavelengths ranging from 350 to 2500 nm with 1 nm increment. Four halogen lamps, 500 watts each, were used as a light source. A white chamber between the light source and the sample was used to uniformly distribute the light and avoid heating up the samples. A spectroradiometer sensor was placed underneath the egg and only the light transmitted through the sample was transported to the spectroradiometer. The spectral data of the egg was obtained by averaging 3 measurements.

After the transmission spectra were determined, each egg was weighed (±0.01mg) and broken carefully on a flat surface. The albumen height was measured using a mounted digital Vernier Caliper (Marathon Watch Company Ltd., On, Ca). These measurements allowed the determination of the HU using Equation (1):

$$HU = 100\log_{10}(h\text{-}1.7w^{0.37}+7.6) \qquad (1)$$

Where HU is the Haugh Units, h is the observed height of albumen in millimetres and w is the weight of the egg in grams.

After separating the albumen from the yolk, the albumen pH was determined using a pH-meter (Accumet Basic AB15 pH-meter, Fisher Scientific, USA). The pH-meter was calibrated using buffer solutions at pH 7 and 10.

5.3.2 Wavelengths selection

The analysis of spectral data is complicated due to the presence of several compounds in eggs. Selection of relevant wavelengths relating to the compound of interest and avoiding interference of others should be a target in terms of building a robust predictive model. Therefore, maximum R^2 (MAXR) method was used in this study to select the informative wavelengths (Abdel-Nour et al., 2008).

5.3.3 Data processing

To analyze VIS/NIR spectral data SAS® version 9.1 (SAS, 2003) statistical software package and The Unscrambler® version 9.7 (The Unscrambler, 2007) were used.

In order to remove background noise from VIS/NIR spectral data, full multiplicative scatter correction (MSC) was applied to the spectral data using The Unscrambler program. After choosing the relevant sets of wavelengths by MAXR, VIS/NIR spectral data were linked to number of days, measured HU and albumen pH by a PLS regression model. Multivariate analysis was performed using The Unscrambler program. The developed models were validated by using the leave-one-out cross validation method. The predictive models were evaluated by calculating the root mean square error (RMSE), the ratio of performance to deviation (RPD), coefficient of determination (R^2), the slope and the intercept of the regression line. The slope and the intercept of regression lines were calculated and compared to their ideal values of 0 and 1, respectively, using IRENE® version Beta 1.00 statistical program (IRENE, 2003) (Fila et al., 2003). To avoid the over/under fitting, the Latent Variables (LV) number was calculated using the prediction residual error sum of squares (PRESS) method.

5.4 RESULTS AND DISCUSSIONS

The transmittance measurements from 350 to 400 nm as well as from 1751 to 2500 nm were not included in this analysis because of extreme noise. The average spectral response of the eggs for 8 measurement days is illustrated in Figure 5.1. Although, VIS/NIR spectra are highly overlapped as shown in Figure 5.1, however, the difference in the transmittance values among the days can be seen in this figure. This demonstrates that the selection of relevant wavelengths is important and a good strategy to avoid the inclusion of uninformative wavelengths in the predictive model. Even though, prediction of quality and freshness in shell eggs is complicated, selection of informative wavelengths made it possible.

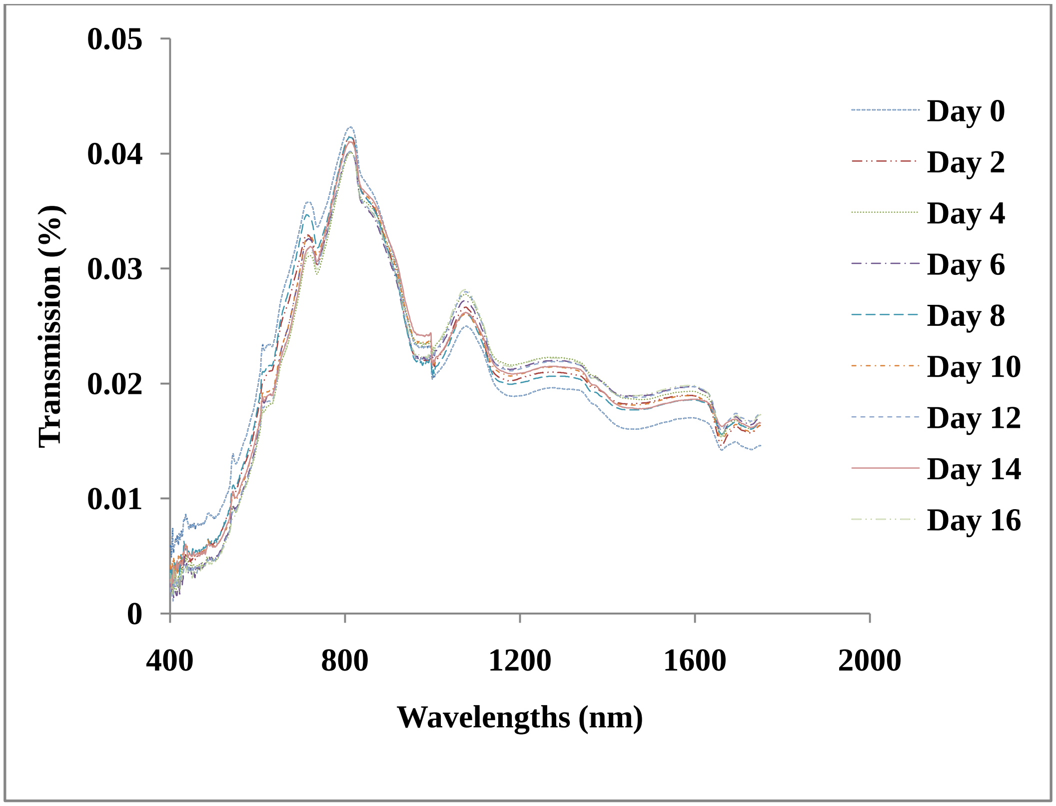


Figure 5.1 Mean of multiplicative scatter correction (MSC) transmission spectra for different days

After the transmission spectra were pre-processed using MSC, the spectral data was set as the independant X matrix and the concentration vector as the dependant Y matrix, with the wavelengths selected using the MAXR method to build the PLS predictive model. The predictive versus measured values of albumen pH, Haugh Units and number of days of storage are shown in Figures 5.2, 5.3 and 5.4, respectively. The parameters *r*, *a* and *b* are the correlation coefficient, the slope and the intercept, respectively. The value of slope and intercept were tested against their ideal values of 0 and 1, respectively. Results showed that there was no significant difference between them. Variation between the eggs laid at the same time and stored under the same conditions is shown in Figures 5.4.

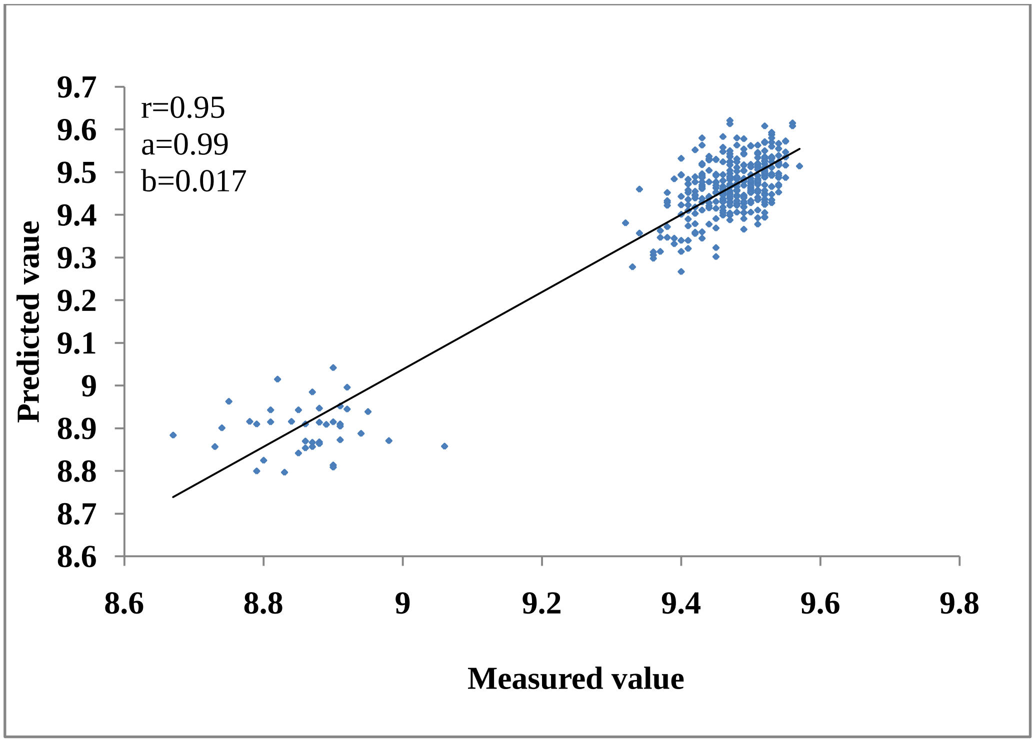


Figure 5.2 Relationship between measured and predicted albumen pH

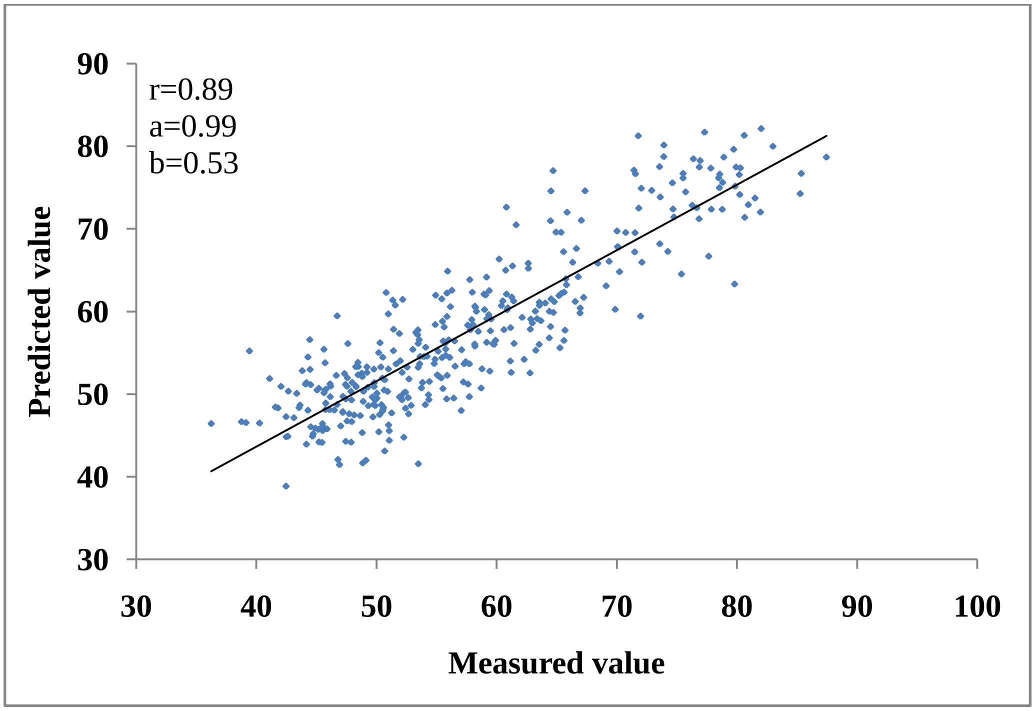


Figure 5.3 Relationship between measured and predicted Haugh Units

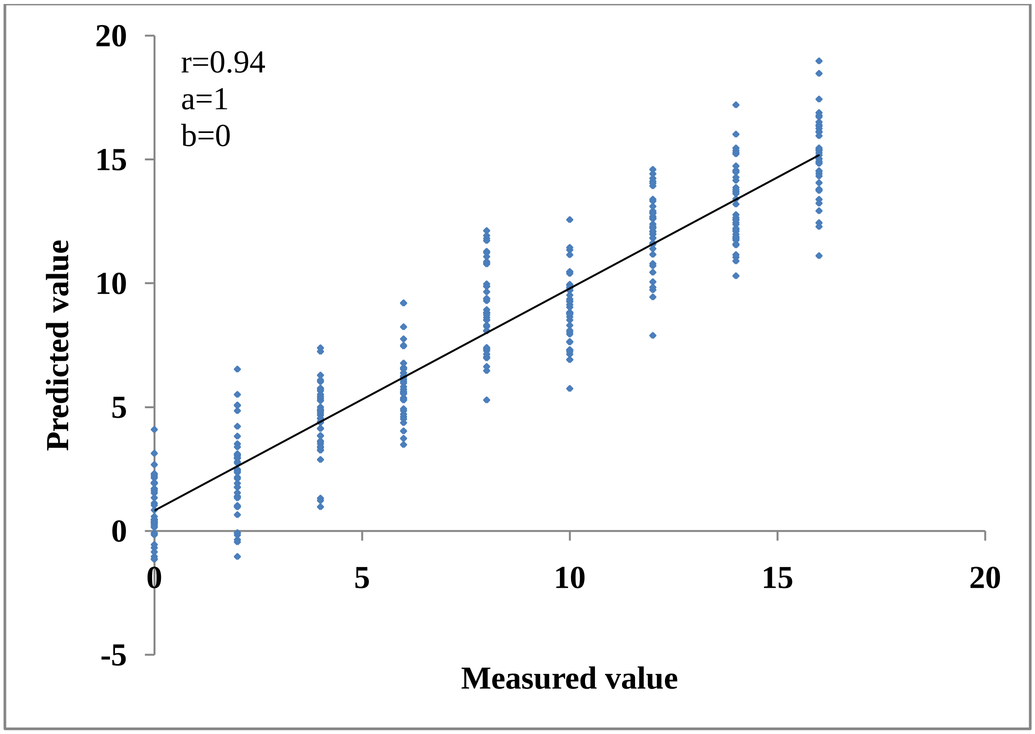


Figure 5.4 Relationship between the measured and predicted number of storage days

It can be seen from Figure 5.2 that there are 2 groups of scatter due to the difference between the albumen pH. One group for freshly laid eggs another group for all other storage days. This confirmes rapid increase in albumen pH in first day after laying.

As a comparison, the lowest correlation coefficient was found in Figure 5.3 for the measured and the predicted HU. This low value could be attributed to the measurement error and not to transmission spectral analysis. This was in agreement with the reports of Kemps et al. (2006). These authors have shown that the place where the height of the thick albumen is measured is crucial but difficult to determine objectively. It can be seen in Figure 5.4 that in the first 2 days of storage, there was a negative prediction. There were 9 eggs predicted negatively at day 0 and 4 eggs at day 4. Considering, the high correlation coefficient (r=0.94) between the predicted and measured values it can be concluded that the negative prediction could be due to the cloudiness of the albumen during the first few days after being laid which results from the presence of the carbon dioxide in the albumen (Dutta et al., 2003).

Wavelengths and latent variables used to build the predictive PLS model based on the MAXR method are shown in Table 5.1. Ten wavelengths were used for building the PLS model and 7, 6 and 6 latent variables (LV) for the albumen pH, HU and number of storage days, respectively.

Table 5.1

Number of wavelengths, number of latent variables and wavelengths used using MAXR-PLS method for albumen pH, Haugh Units (HU) and number of days

	Number of wavelengths used	Wavelengths used	Latent variables
Albumen pH	10	499, 816, 904, 936, 977, 1129, 1196, 1345, 1595, 1719	7
Haugh Units	10	415, 435, 444, 474, 484, 661, 1225, 1319, 1370, 1728	6
Number of storage days	10	411, 478, 484, 530, 531, 661, 1359, 1402, 1635, 1729	6

As shown in Table 5.1, 10 wavelengths from the visible and near infrared region for all the 3 predictive models were used to build the PLS predictive model. These results are not in agreement with those of Kemps et al.

(2006) where the PLS predictive model was built with 6 wavelengths and 4 LVs. This difference can be attributed to the instruments used, experimental set up or the light source. Consequently, for building the predictive model for the albumen pH, 1 wavelength in the visible region was used and the others are in the NIR region where the NIR spectra correspond to vibrational modes involving O-H chemical bonds (Karoui et al., 2006). However, the predictive model for HU and number of storage days required 6 wavelengths in the visible region and 4 wavelengths in the near infrared region. Albumen becomes thinner over time which leads to a change in the transmitted spectra. The HU and the number of storage days under controlled conditions are related to the viscosity and cleanliness which requires more wavelengths in the visible range to be used to build the predictive model than those used to build that of albumen pH. For example, Maillard or browning reactions occurs due to aging. They have the absorption light between 600 and 700 nm, accounting for the use of wavelength 661 in the predictive models of HU and the number of storage days.

The R^2 and (RMSE) for calibration and validation sets as well as the ratio of performance to deviation (RPD), which is equal to the standard deviation divided by the RMSEV are shown in Table 5.2.

Table 5.2

Prediction results for albumen pH, Haugh Units and number of storage days

	R^2		RMSE		RPD
	Validated	Calibrated	Validated	Calibrated	
Albumen pH	0.90	0.91	0.06	0.06	3.32
Haugh Units	0.79	0.79	5.05	4.90	2.16
Number of storage days	0.89	0.90	1.65	1.64	4.92

The predictive ability of the models built was high as shown in Table 5.2. The RPD can be criterion for testing the prediction ability of the model. Although, Kemps et al. (2006) reported that an RPD of 2 for HU means that the eggs could be subdivided into 2 groups in terms of internal quality. It is hard to classify the eggs based on RPD. This study showed that the value of RPD can be more than 2. This is in agreement with the results of McGlone et al. (2002) who reported an RPD value of 3. However, it may be concluded that for the RPD values more than 2, eggs can be classified in two groups in terms of internal quality. The R^2 for calibrated and validated sets for the predictive model of HU was 0.79, which is lower than the R^2 for albumen pH and number of storage days. As mentioned earlier, this could be due to the measurement error.

Prediction of the albumen pH, HU and number of storage days can be done by using wavelengths from the VIS and NIR region up to 1729 as shown in Tables 5.1 and 5.2. This result is not in agreement with the studies carried out by Kemps et al. (2006) who have shown that the relevant information in terms of albumen pH and HU is restricted to the interval between 570 and 750 nm and Kemps et al. (2007), who found that the spectral region between 500 to 900 nm is valuable for the prediction of albumen freshness. These differences can be due to the methods of selection of relevant wavelengths or due to the range of spectral data studied. In overall comparison with previous work done by Kemps et al. (2006), the correlation coefficients were improved. These authors reported that the correlation coefficients for HU and albumen pH were 0.84 and 0.86, respectively. Six wavelengths in the visible region were used to build the predictive model. Whereas, in this work, the correlation coefficients for HU and albumen pH were 0.89 and 0.95, respectively, and 10 wavelengths in the visible and near infrared region were used. The improvement in accuracy of prediction and the difference in number of wavelengths used can be attributed to the method of selection of relevant wavelengths used for building the predictive model and to the instrument used. However, the wavelengths in this study ranged from 350 to 2500 nm, whereas, in their study the wavelengths ranged from 200 to 1100 nm.

5.5 CONCLUSION

In this research, the ability of VIS/NIR spectroscopy to assess egg freshness in terms of number of storage days and albumen quality in terms of albumen pH and HU were investigated. The results presented above have shown that the transmission spectral data of the egg contains information about egg quality and freshness. A suitable predictive model was built with PLS regression model using informative wavelengths selected with MAXR method. The developed model was used to relate the spectral data to the albumen pH, HU and number of storage days. For albumen pH, HU and the number of storage days, the R^2 for the validation set were 0.90, 0.79 and 0.89, respectively, and the low values of RMSEV of 0.06, 5.05 and 1.65, for the same, suggesting that the VIS/NIR spectroscopy can be used as a non-destructive method for the assessment of egg quality and freshness.

5.6 ACKNOWLEDGEMENTS

The authors would like to acknowledge the Poultry Industry of Canada for funding this study.

5.7 REFERENCES

Abdel-Nour, N, M. Ngadi, S. Prasher, and Y. Karimi. 2008. Combined Maximum R^2 and partial least squares method for wavelengths selection and analysis of spectroscopic data. (Submitted to Applied Spectroscopy).

Berardinelli,A., A. Giunchi, A. Guarnieri, F. Pezzi, and L. Ragni. 2005. Shell egg albumen height assessment by FT-NIR Spectroscopy. Transactions of the ASAE, 48(4): 1423-1428.

Birth, G.S., G.G. Dull, W.T. Renfore and S.J. Kays. 1985. Non-destructive spectrometric determination of dry matter in onions. Journal of the American Society for Horticultural Science, 110(2): 297-303.

Dutta, R., E.L. Hines, J.W. Gardner, D.D. Udrea, and P. Boilot. 2003. Non-destructive egg freshness determination: an electronic nose based approach. Measurement Science and Technology, 14: 190-198.

Fila, G., A. Bellochi, M. Acutis, and M. Donatelli. 2003. IRENE: a software to evaluate model performance. European Journal of Agronomy 18: 269-372.

Gómez, A.H., Y. He, and A.G. Pereira. 2006. Non-destructive measurement of acidity, soluble solids and firmness of Satsuma mandarin using VIS/NIR-spectroscopy techniques. Journal of Food Engineering, 77: 313-319.

Haugh , R.R. 1937. The Haugh Unit for measuring egg quality. US Egg Poultry Magazine. 43: 552-555.

IRENE, 2003. Integrated resources for evaluating numerical estimates. Ver. beta 1.00. Bologna, Italy.

Karoui, M., B. Kemps, F. Bamelis, B. De Katelaere, E. Decuypere, and J. De baerdemaeker. 2006. Methods to evaluate egg freshness in research and industry: a review. European Food Research Technology, 222: 727-732.

Kemps, B.J., F.R. Bamelis, B. De Katelaere, K. Mertens, K. Tona, E.M. Decuypere, and J.G. De Baerdemaeker. 2006. Visible transmission spectroscopy for the assessment of egg freshness. Journal of the Science of Food and Agriculture 86: 1399-1406.

Kemps, B.J., B. De Katelaere, F.R. Bamelis, K. Mertens, K. Tona, E.M. Decuypere, J.G. De Baerdemaeker, and F. Schwägelet. 2007. Albumen freshness assessment by combining visible near-infrared transmission and low-resolution proton nuclear magnetic resonance spectroscopy. Journal of Poultry Science, 86: 752-759.

Lapao, C., L.T. Gamma, and M. Caveiro Soares. 1999. Effects of broiler breeder age and length of egg storage on albumen characteristics and hatchability. Journal of Poultry Science, 78: 640-645.

Liu, Y., Y. Ying, A. Ouyang, and Y. Li. 2007. Measurement of internal quality in chicken eggs using visible transmittance spectroscopy technology. Food Control, 18: 18-22.

McGlone, V.A., R.B. Jordan, and P.J. Martinson. 2002. Vis/NIR estimation at harvest of pre- and post-storage quality indices for ‘Royal Gala’ apple. Postharvest Biology and Technology. 25 : 135-44.

Robinson, D.S. and J.B. Monsey. 1972. Changes in the composition of ovomucin during liquefaction of thick white. Journal of the Science of Food and Agriculture, 23:29-38.

SAS. 2003. SAS User's guide: statistics. Ver. 8.2 Cary, N.C.: SAS Institute, Inc.

Slaughter, D.C., and C.H. Crisosto. 1998. Non-destructive internal quality assessment of kiwifruit using near-infrared spectroscopy. Seminars in Food Analysis, 3: 131-140.

The Unscrambler. 2007. The Unscrambler User's Guide: ver. 9.7 Woodbridge, N.J., USA: CAMO Software AS.

Wells, P.C., and K.H. Norris. 1987. Egg quality-Current problem and recent advances. *In* "Egg Quality-Current Problems and Recent Advances" (B.M. Freeman Ed.). Carfax Publishing Company. Abingdon, England.

CHAPTER 6

SUMMARY AND CONCLUSIONS

6.1 SUMMARY

This study investigated the ability of the maximum R^2 method to select relevant wavelengths and to use them to build a partial least squares predictive model. The usefulness of the visible/infrared spectroscopy to assess the quality and egg freshness: number of storage days, Haugh Units (HU) and albumen pH were also investigated.

To achieve these objectives, 360 freshly laid eggs from the same flock fed with a standard ration were collected. Eggs were stored at T=18°C and 55% relative humidity for a duration of 16 days. Forty samples were collected on day 0, 2, 4, 6, 8, 10, 12, 14 and 16. Spectral data was collected for the samples and internal quality assessed using the most widely used destructive methods, namely HU and albumen pH.

6.2 CONCLUSIONS

The conclusions cited below are in the same order as presented in the thesis:

(1) The ability of the maximum R^2 (MAXR) method to select important wavelengths in visible/Near infrared (VIS/NIR) spectral data was investigated. The sets of wavelengths selected were used to build partial least squares (PLS) predictive models for the HU and albumen pH. The robustness of the models was tested by calculating the root mean squared error for the prediction (RMSEP) and the correlation coefficients between the predicted and the measured values. The results showed that using this method the correlation coefficients were up to 95% and the RMSEP were 0.05 and 5.05 for

albumen pH and HU, respectively. This method reduced the complexity of the model and improved its predictive ability.

(2) The use of VIS/NIR spectroscopy of assessment of egg quality and freshness was investigated. Transmission spectral data was obtained in the range of 350 to 2500 nm and linked with the number of storage days, albumen pH and HU by the mean of partial least squares (PLS) predictive models. The correlation coefficients between the measured and predicted number of storage days, HU and albumen pH were up to 0.9, the R^2 was up to 0.9 and RMSEP were 1.65, 5.05 and 0.06, respectively. These results showed that the VIS/NIR spectroscopy can be used as a non-destructive method for the assessment of egg quality.