NEAR-INFRARED TECHNIQUE IN DEVELOPING AN ONLINE SUGAR CONTENT MONITORING SYSTEM OF SUGARCANE ON THE ELEVATOR CONVEYOR

KITTISAK PHETPAN

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF ENGINEERING IN AGRICULTURAL ENGINEERING
FACULTY OF ENGINEERING
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG
2019

KMITL-2019-EN-D-108-044

NEAR-INFRARED TECHNIQUE IN DEVELOPING AN ONLINE SUGAR CONTENT MONITORING SYSTEM OF SUGARCANE ON THE ELEVATOR CONVEYOR

KITTISAK PHETPAN

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF ENGINEERING IN AGRICULTURAL ENGINEERING
FACULTY OF ENGINEERING
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG
2019
KMITL-2019-EN-D-108-044

เทคนิคอินฟราเรดย่านใกล้ในการพัฒนาระบบติดตามความหวานของอ้อยแบบออนไลน์บนสายพานลำเลียง

กิตติศักดิ์ เพ็ชรพันธ์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรดุษฎีบัณฑิต
สาขาวิชาวิศวกรรมเกษตร
คณะวิศวกรรมศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
พ.ศ.2562
KMITL-2019-EN-D-108-044

NEAR-INFRARED TECHNIQUE IN DEVELOPING AN ONLINE SUGAR CONTENT MONITORING SYSTEM OF SUGARCANE ON THE ELEVATOR CONVEYOR

KITTISAK PHETPAN

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF ENGINEERING IN AGRICULTURAL ENGINEERING
FACULTY OF ENGINEERING
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG
2019
KMITL-2019-EN-D-108-044

หัวข้อวิทยานิพนธ์	เทคนิคอินฟราเรดย่านใกล้ในการพัฒนาระบบติดตามความหวานของอ้อยแบบออนไลน์บนสายพานลำเลียง
นักศึกษา	ว่าที่ร้อยตรี กิตติศักดิ์ เพ็ชรพันธ์
รหัสประจำตัว	58601044
ปริญญา	วิศวกรรมศาสตรดุษฎีบัณฑิต
สาขาวิชา	วิศวกรรมเกษตร
พ.ศ.	2562
อาจารย์ที่ปรึกษาวิทยานิพนธ์	รศ.ดร. ปานมนัส ศิริสมบูรณ์
อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม	ดร. วสุ อุดมเพทายกุล

บทคัดย่อ

ด้วยแนวโน้มการทำเกษตรสมัยใหม่ที่มุ่งเน้นการปรับปรุงในเรื่องของผลผลิตต่อพื้นที่และคุณภาพ โดยเฉพาะอย่างยิ่งในอุตสาหกรรมอ้อยได้มีความเจริญก้าวหน้ามากขึ้น หลักสำคัญสำหรับประเด็นนี้คือการจัดการด้านต่างๆ ผ่านการเข้าถึงปัจจัยการผลิตตลอดจนความแปรปรวนของผลผลิตและคุณภาพภายในแปลง และด้วยความต้องการวัดหรือติดตามคุณภาพภายในแปลงทำให้มีการเสนอเทคนิคเชิงสเปกโทรสโกปีของคลื่นแสงอินฟราเรดย่านใกล้ว่าเป็นเทคนิคที่มีความเป็นไปได้ในการพัฒนาระบบติดตามคุณภาพอ้อยแบบออนไลน์ซึ่งสามารถติดตั้งบนรถตัดอ้อยเพื่อทำการวัดขณะเก็บเกี่ยวอ้อย

วัตถุประสงค์หลักของวิทยานิพนธ์นี้คือ การศึกษาวิจัยเพื่อพิสูจน์ความเป็นไปได้ของการใช้เทคนิคอินฟราเรดย่านใกล้ในการพัฒนาระบบติดตามความหวานของอ้อยที่กำลังถูกลำเลียงอยู่บนชุดลำเลียง การศึกษาถูกแบ่งออกเป็น 4 ส่วนหลัก ส่วนแรก คือการศึกษาอิทธิพลของพันธุ์อ้อยต่อการทำนายความหวาน การศึกษานี้ใช้อ้อยสายพันธุ์เชิงการค้าจำนวน 2 สายพันธุ์ (พันธุ์ลำปาง 92-11 และขอนแก่น 3) โดยผลการศึกษาพบว่าสายพันธุ์อ้อยที่แตกต่างกันส่งผลต่อการทำนายความหวานของอ้อย ด้วยเหตุนี้จึงเสนอการสร้างแบบจำลองชนิด global ซึ่งเป็นการสร้างแบบจำลองจากข้อมูลทั้งสองสายพันธุ์ขึ้น และสามารถให้ผลการทำนายความหวานอ้อยเป็นที่น่าพอใจ ส่วนที่สองของงานวิจัยนี้คือ การนำเสนอระบบต้นแบบในการวัดความหวานของอ้อยแบบออนไลน์บนสายพานลำเลียงโดยใช้เทคนิคอินฟราเรดย่านใกล้ และทำการประเมินสมรรถนะแบบจำลองที่ใช้ทำนายตัวอย่างอ้อยขณะกำลังถูกลำเลียง ระบบต้นแบบประกอบด้วย 2 ส่วนหลักคือ ชุดลำเลียงอ้อยซึ่งจำลองการทำงานในสภาวะจริงของระบบลำเลียงของรถตัดอ้อย และระบบการตรวจวัดสัญญาณเชิงแสง (สเปกตรัม) แบบจำลองถูกพัฒนาด้วยสมการถดถอยกำลังสองน้อยที่สุดบางส่วน (PLSR) โดยให้ค่าสัมประสิทธิ์การตัดสินใจ (R^2) และค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ยของการทำนายเท่ากับ 78.5% และ 0.30 องศาบริกซ์ ตามลำดับ ซึ่งเป็นการบ่งชี้ถึงความเป็นไปได้สำหรับการ

ใช้เทคนิคอินฟราเรดย่านใกล้ในการวัดความหวานอ้อยแบบออนไลน์ ผลของความเป็นไปได้นี้ส่งผลถึงการศึกษาส่วนที่สามในการค้นหาว่าจะดีกว่าหรือไม่ หากพัฒนาแบบจำลองโดยไม่ใช้ย่านแสงที่มนุษย์สามารถมองเห็นได้ อีกทั้งยังศึกษาอิทธิพลของระดับการลำเลียงอ้อยบนระบบลำเลียงว่ามีผลต่อการทำนายความหวานหรือไม่ ผลการศึกษาพบว่า การพัฒนาแบบจำลองเฉพาะช่วงคลื่นเนียร์อินฟราเรดย่านใกล้มีความเหมาะสมมากกว่าแบบจำลองที่มีช่วงครอบคลุมย่านแสงที่มนุษย์มองเห็นได้ โดยมีความคลาดเคลื่อนกำลังสองเฉลี่ยของการทำนายกลุ่มตัวอย่างที่เป็นอิสระจากการพัฒนาแบบจำลอง (RMSEP) เท่ากับ 0.31 องศาบริกซ์ สำหรับการศึกษาอิทธิพลของระดับของอ้อยบนระบบลำเลียง พบว่า ระดับของอ้อยที่ต่างกันส่งผลต่อการทำนายความหวานของอ้อย และควรเป็นปัจจัยที่ต้องมีการพิจารณาสำหรับการพัฒนาแบบจำลองในอนาคต ส่วนสุดท้ายของการศึกษาวิจัยสำหรับวิทยานิพนธ์นี้ คือ การประเมินวิธีการต่างๆ เพื่อใช้สำหรับการคัดแยกข้อมูลสเปกตรัมที่ไม่ต้องการโดยใช้การวิเคราะห์องค์ประกอบหลัก (PCA) การวิเคราะห์การจัดกลุ่ม (cluster analysis) และดัชนีทางสถิติ ผลการศึกษาพบว่า การประยุกต์ใช้ดัชนีทางสถิติเพื่อคัดแยกสเปกตรัมที่ไม่ต้องการให้ผลดีที่สุด โดยมีความถูกต้องในการคัดแยกเท่ากับ 89.19% นอกจากนี้แล้ว สเปกตรัมที่ผ่านการคัดแยกด้วยวิธีดังกล่าวถูกนำไปสร้างแบบจำลองร่วมกับค่าความหวาน และให้ค่าสัมประสิทธิ์ของการตัดสินใจ (R^2) และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของการพิสูจน์ภายใน (RMSECV) เท่ากับ 0.768 และ 0.43 องศาบริกซ์ ตามลำดับ

จากผลการศึกษาข้างต้น วิทยานิพนธ์นี้เป็นงานวิจัยแรกที่ยืนยันถึงความเป็นไปได้ในการใช้เทคนิคอินฟราเรดย่านใกล้สำหรับการวัดสัญญาณเชิงแสง (สเปกตรัม) ในโหมดการสะท้อนกลับโดยไม่สัมผัสตัวอย่าง เพื่อทำนายปริมาณความหวานของอ้อยขณะกำลังถูกขนถ่ายอยู่บนระบบลำเลียงของรถตัดอ้อย การประยุกต์ใช้ดัชนีทางสถิติในการคัดเลือกสเปกตรัมอ้อย และการพัฒนาแบบจำลองโดยใช้คลื่นแสงอินฟราเรดย่านใกล้ในช่วง 700-900 nm ได้ถูกนำเสนอสำหรับการพัฒนาระบบติดตามความหวานของอ้อยแบบออนไลน์บนรถตัดอ้อย อย่างไรก็ตาม เนื่องจากอิทธิพลของสายพันธุ์อ้อยและระยะห่างระหว่างตำแหน่งการติดตั้งเซ็นเซอร์กับระดับของตัวอย่างอ้อยบนระบบลำเลียงที่ต่างกันส่งผลต่อการทำนายความหวานของอ้อย ดังนั้นการพัฒนาแบบจำลองสำหรับการใช้งานจริงในอนาคตจึงต้องมีการเพิ่มเติมความหลากหลายของข้อมูลจากอ้อยสายพันธุ์ต่างๆ อีกทั้งมีความจำเป็นสำหรับการศึกษาวิจัยต่อยอดเพื่อค้นหาระยะการติดตั้งเซ็นเซอร์สำหรับการวัดสเปกตรัมที่เหมาะสม

Thesis	Near-infrared technique in developing an online sugar content monitoring system of sugarcane on the elevator conveyor
Student	Acting Sub Lt. Kittisak Phetpan
Student ID.	58601044
Degree	Doctor of Engineering
Program	Agricultural Engineering
Year	2019
Thesis Advisor	Assoc. Prof. Dr. Panmanas Sirisomboon
Co - Thesis Advisor	Dr. Vasu Udompetaikul

ABSTRACT

Modern agricultural production focusing on the improvement of yield and quality aspects, particularly in sugarcane production, is on the rise, without an enlargement of the areas used for growing crops. A key factor is management through the access of production elements, as well as the variation of such aspects across a given field. With the need for real-time monitoring of quality in the field, the near-infrared (NIR) spectroscopic approach has been promoted as being promising for developing an online system to be installed on sugarcane harvesters.

In this thesis, the research study for the possibility of using the NIR spectroscopic technique to develop an online sugar content monitoring system for the transfer of sugarcane on the elevator conveyor of harvesters was proven after being divided into 4 main sections. The study regarding the influence of sugarcane variety on sugar content prediction using NIR spectroscopy was firstly performed. Two commercial sugarcane varieties (LK92-11 and Khonkaen 3) were used for the study, and it was found that the variety of sugarcane affects sugar content prediction using shortwave near-infrared spectroscopy. Based on this result, a global model was proposed to develop the model and gave satisfying results for sugar content prediction. Afterward, a prototype online sugar content measuring system for sugarcane using the NIR spectroscopic technique was proposed and used to assess the performance of the calibration model for real-time evaluation. This lab-scale prototype consisted of two main parts – a cane billet elevator and a spectral acquisition system. The system can detect the spectra of the cane billets conveyed

at a speed of 2 m/s using an integration time of 14 ms. Partial least squares (PLS) modeling was used to correlate the obtained spectra with the soluble solids content (SSC). With this modeling, coefficients of determination (R^2) and root mean square error of prediction (RMSEP) of 78.5% and 0.30 ºBrix, respectively, were statistical parameters indicting the possibility for the online SSC measurement of the sugarcane billets on the elevator. This possibility lead to the third part of this research study. An exploration into the advantages of modeling cane-SSC predictions without the visible spectral range was conducted, as well as into whether it is possible to develop the calibration by neglecting the influence of different cane levels on the elevator. The result identified that there is no need to include the visible region for the cane-SSC prediction model, showing the ability for independent external sample prediction in an error term of RMSEP of 0.31 °Brix. The influence of different levels of cane delivery on the conveyor was found to affect the cane-SSC prediction, and should be considered for modeling in the future. For the final part of this thesis, the spectral classification methods were evaluated for use as a technique in screening non-sugarcane spectra. Spectral classification using principal component analysis (PCA), cluster analysis and a proposed statistical index was conducted. The adoption of statistical index was the most successful, showing 89.19% for overall accuracy. Furthermore, PLS modeling employing the spectra obtained from the best approach was also performed. This exhibited R^2 and an error for internal validation (root mean square error of cross-validation, RMSECV) of 0.768 and 0.43 ºBrix, respectively.

Based on these findings, this thesis confirmed the possibility for using the near-infrared spectroscopic technique in sugar content evaluation of sugarcane being transferred on the elevator conveyor. Non-contact reflectance was the mode for spectral measurement. Sugarcane spectral selection using the statistical index and calibration modeling using near-infrared spectral range of 700-900 nm were proposed for developing the online sugar content monitoring system of sugarcane moving on the conveyor of the sugarcane harvester. Since, sugarcane variety and levels of cane delivery on the elevator affect sugar content prediction. So, modeling with diversity of sugarcane varieties is needed for real use. In addition, digging into the rather suitable position of the optic to be installed is required for further study.

Acknowledgements

The success of this thesis came about due to many sources of assistance and contributions. First of all, I would like to express my deepest gratitude to both of my advisors, Assoc. Prof. Dr. Panmanas Sirisomboon and Dr. Vasu Udompetaikul, for their valuable time and advice, as well as their encouragement throughout the study period. I cannot thank you enough for such important learning experiences. Particularly, I wish to acknowledge the financial support from the Royal Golden Jubilee PhD scholarship (PHD/0102/2558) of Thailand research fund (TRF) for tuition fee, monthly expenses, as well as international research experience.

I would like to thank the Near Infrared Spectroscopy Research Center for Agricultural Products and Food (www.nirsresearch.com) and the Precision Agriculture Laboratory at King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand, for laboratory space and instruments. Also, many thanks are due to Mr. Ampol Jongsomboonpokha for providing sugarcane samples and for allowing me access to his sugarcane harvester.

Special thanks are also extended to Prof. Shrinivasa Upadhyaya for giving me an opportunity to join international research at University of California, Davis, USA. With this chance, I greatly appreciate Dr. Irwin Donis-Gonzalez for having me as a part of his programs, extending my knowledge under his supervision.

I express my gratitude to all faculties of the department of agricultural engineering at King Mongkut's Institute of Technology Ladkrabang (KMITL). They provided not only academic relationships but also everything for living, motivation and support. Many thanks are also due to all of my friends who took the time to support me in my hard research.

Finally, I would like to dedicate all successes of this thesis to my family. They have always given me unconditional love and support, both physically and mentally, throughout my studies and my life in general.

Kittisak Phetpan

Table of contents

Page

Table of contents (continued)

Page

Table of contents (continued)

Page

Table of contents (continued)

Page

List of tables

List of figures

List of figures (continued)

List of figures (continued)

Figure **Page**

List of figures (continued)

Figure **Page**

Abbreviations

ASM	Assisted scanning method
Bias	Average error
CCS	Commercial cane sugar
DSM	Direct scanning method
FOV	Field-of-view
LLM	Linear learning machine
LV	Latent variable of PLS models
MA	Moving average smoothing
MLR	Multiple linear regression
MSC	Multiplicative scatter correction
NIR	Near-infrared
PA	Precision agriculture
PC	Principal component
PCA	Principal component analysis
PCR	Principal component regression
PLSR	Partial least squares regression
Pol	Polarization
R^2	Coefficient of determination
RMSECV	Root mean square error of cross-validation
RMSEP	Root mean square error of prediction
RPD	Standard deviation to standard error of prediciton
SD	Standard deviation
SECV	Standard error of cross-validation
SIMCA	The soft independent modeling of class analogies
SNV	Standard Normal variate
SSC	Soluble solids content (°Brix)
SVD	Singular value decomposition
SVMs	Support vector machines
SW-NIR	Shortwave near-infrared
Vis/SW-NIR	Visible and shortwave near-infrared

Chapter 1
Introduction

1.1 Background

Agriculture provides the basic essentials for humanity around the world. The food we eat, the beverages we drink or even the fuels we use can illustrate the importance of agriculture to the quality of our life. One of the agricultural products that has a dominant influence on modern life is sugarcane. It is not only used in sugar production but also used to produce alternative fuels in the form of ethanol (Cookson, 2012). Sugar production worldwide amounted to roughly 194 million metric tons in 2017/2018 (USDA, 2018). The number one sugar producing country in the world is Brazil, yielding almost 40 million metric tons, whilst the top five territories are composed of Brazil, India, EU countries, Thailand and China (Walton, 2019). Furthermore, this crop is also important for the energy situation due to the fact that reserves of fossil fuels are decreasing and will eventually be depleted. The use of sugarcane for producing bioethanol to slow down the use of fossil fuels is one potential solution; hence, Brazil was announced to be a pioneer in producing bioethanol from sugarcane, outputting more than 20 billion liters of ethanol per year whilst expected to reach 50 billion liters per year by 2020 (Cookson, 2012). Subsequently, these factors point towards the sugarcane industry's importance to modern life.

The increase in global population, expected to reach 8.6, 9.8 and 11.2 billion in 2030, 2050 and 2100, respectively (United Nations, 2017), directly affects the demand for essential items needed to maintain modern standards of living. Of course, this means that the demand for sugarcane would be increasing as well. As a result, increasing yield and maintaining quality in sugarcane production without enlarging the areas used for growing crops is also a chief aim in commercial crop production. To achieve these goals, access to the site-specific information for the management of field input and activities is the best option as the variation in the yield and quality within farmland is still a big barrier. The spatial variation of yield can reach 30 – 45% in terms of the coefficient of variation (Bramley and Quabba, 2001),

whereas that of commercial cane sugar (CCS), i.e., the primary industrial sugarcane quality index, can reach up to 6.5 units within a single 8.8 ha sugarcane block (Kingston and Hyde, 1995).

Precision agriculture (PA) is a valuable management tool, gathering technologies to manage and improve the agricultural production processes through the recognition of production factors, soil fertility as well as the variation in yield and quality across a given field. Of course, PA technology has already been applied in the sugarcane industry. Accessing the site-specific yield outcome of sugarcane across the field and construct variability maps is not an arduous process, but still presents challenges for getting the site-specific quality (Bramley, 2009). In the future, if the information on both spatial variations across the field can be accessed, farm management practices will take action against this information and a fair payment system to the growers will be implemented (Nawi et al., 2013a). In addition, the information could be beneficial to the sugarcane industry in dealing with individual small-holding farming and cooperative farming on a large scale.

The need for real-time monitoring sensors or systems in sensing the quality responses of sugarcane in the field is obvious. Hence, exploration of techniques that can measure cane quality in the field has been conducted broadly, such as electronic refractometer (Mccarthy and Billingsley, 2002), microwave (Nelson, 1987; Klute, 2007; Shah and Joshi, 2010) and spectroscopic techniques (Nawi et al., 2013a; Nawi et al., 2013b). From these techniques, the spectroscopic approach has been developed continuously with an expectation of real-time senor detection in field use. Nawi et al. (2013a), who is probably the first pioneer in conducting research using the spectroscopic technique for sugar content measurements in cane billets based on surface scanning with a non-contact strategy, claimed that the spectroscopic technique looked promising as it showed good results with a coefficient of determination (R^2) and root means square error of predictions (RMSEP) of 0.91 and 0.721 °Brix, respectively. Later, Nawi et al. (2014) determined that if the sugar content monitoring system has been developed, the optimum location to install the sugar content monitoring system will be on the elevator of sugarcane harvester. This location is an open section in which the cane billets could be

scanned directly, with approximately 90% of trash already blown out by the primary extractor. In addition, the system could be placed at any point along the elevator.

The spectroscopic technique involves the utilization of the electromagnetic spectrum, covering the wavelength range from approximately 350-2500 nm (Nawi et al., 2013b). The first half of this range, 350-1100 nm, is defined as an area of visible and shortwave near-infrared (Vis/SW-NIR), ascribed to the third and fourth overtones of O-H and C-H stretching modes (Walsh et al., 2000) and has been claimed to be suitable for development as a sensor for monitoring sugarcane quality during harvesting (Nawi et al., 2014). Unfortunately, this is purely theoretical and there has been no research conducted to demonstrate whether or not the spectroscopic technique is possible to be developed as an online sugar content monitoring system for sugarcane being transferred on the elevator conveyor of sugarcane harvesters.

1.2 Research objectives

The goal of this thesis was to investigate the possibility of using the spectroscopic technique in developing an online sugar content monitoring system of sugarcane billets to be installed on the elevator conveyor of sugarcane harvester. The research was related to the following specific objectives:

1. To preliminarily study the influence of sugarcane variety on sugar content prediction using near-infrared spectroscopy.
2. To propose a prototype online sugar content measuring system for sugarcane using the near-infrared spectroscopic technique and to assess the performance of the calibration model for real-time evaluation.
3. To develop the partial least squares (PLS) regression, exploring whether it was better in modeling sugar content prediction of sugarcane billets moving on the elevator outside of the visible spectral range (450-700 nm), and whether it was possible to develop the calibration model by neglecting the influence of different cane levels on the elevator.
4. To evaluate the spectral classification methods for use as a technique in screening non-sugarcane spectra.

1.3 Navigation of the thesis

This thesis is organized into seven chapters. A brief discussion of every chapter is presented below.

Chapter 1 presents the research background initiating this thesis, as well as defining the research objectives.

Chapter 2 involves theories and literature reviews, beginning with an overview of the sugarcane industry consisting of the importance of the sugarcane industry, sugarcane botany, current sugarcane harvesting and quality measurement systems, as well as existing technology for sugarcane industry. Introduction of near-infrared spectroscopy and its application in the sugarcane industry are then given. At the end of this chapter, chemometrics is presented. In this part, all approaches that were applied for spectral pre-processing, classification and pattern recognition as well as modeling are also explained theoretically.

Chapter 3, the first study of this thesis, presents the influence of sugarcane variety on sugar content prediction using near-infrared spectroscopy. The need for the application within in-field quality measurement is also proposed. The outcome of this chapter leads to setting up the scope of the thesis, which aims at the investigation of the spectroscopic application for developing an online sugar content monitoring system for sugarcane.

Chapter 4 proposes a prototype of an online sugar content monitoring system for sugarcane, simulating the circumstances of movement and conveying speed of the cane billets as being transferred on the elevator conveyor of sugarcane harvester. This chapter also presents the initial assessment as to whether the calibration models can be used to predict in real-time the sugar content of sugarcane.

Chapter 5 focuses on modeling the optimal partial least squares (PLS) regression for real use. An appropriate spectral range and the influence of different levels of cane billets on an elevator are the issues focused upon. Furthermore, model evaluation by predicting unknown samples is also presented in this chapter.

Chapter 6 mainly has to do with the spectral filtration process of the online soluble solids content (SSC) measuring system of sugarcane. Exploration and evaluation of

strategies in distinguishing between sugarcane and non-sugarcane (slat and floor) spectral detection is the focus here. In addition, the calibration model constructed from the output obtained from the optimal strategy is also exhibited.

Chapter 7 summarizes the main conclusions and also recommends further studies.

1.4 References

Bramley R.G.V. and Quabba R.P. 2001. "Opportunities for improving the management of sugarcane production through the adoption of precision agriculture - An Australian perspective." **Proceedings of the 24th Congress of the International Society of Sugar Cane Technologists.** 38-46.

Bramley R. 2009. "Lessons from nearly 20 years of precision agricultural research, development, and adoption as a guide to its appropriate application." **Crop and Pasture Science.** 60, 197-217.

Cookson C. 2012. A tank of sugar: how Brazil runs on biofuel. **FT Magazine**, United Kingdom

Kingston G. and Hyde R.E. 1995. "Intra-field variation of commercial cane sugar (CCS) values." **Proceedings of the Australian Society of Sugar Cane Technologists.** 17, 30-38.

Klute U. 2007. "Microwave measuring technology for the sugar industry." **International Sugar Journal.** 109(1308), 1-6.

Mccarthy S. and Billingsley J. 2002. "A sensor for the sugar cane harvester topper." **Sensor Review.** 22(3), 242-246.

Nawi N.M., Chen G. and Jensen T. 2013b. Visible and shortwave near infrared spectroscopy for predicting sugar content of sugarcane based on a cross-sectional scanning method. Journal of Near Infrared Spectroscopy, 21, 289–297.

Nawi N.M., Chen G. and Jensen T. 2014. "In-field measurement and sampling technologies for monitoring quality in the sugarcane industry: a review." **Precision Agriculture.** 15, 684–703.

Nawi N.M., Chen G., Jensen T. and Mehdizadeh S.A. 2013a. "Prediction and classification of sugar content of sugarcane based on skin scanning using visible and shortwave near infrared." **Biosystems Engineering.** 115(2), 154-161.

Nelson S.O. 1987. “Potential agricultural applications for RF and microwave energy.” **Transactions of the ASAE.** 30(3), 818-831.

Shah S. and Joshi M. 2010. “Modeling microwave drying kinetics of sugarcane bagasse.” **International Journal of Electronics Engineering.** 2(1), 159–163

United Nations. 2017. **News: World population projected to reach 9.8 billion in 2050, and 11.2 billion in 2100**. [Online]. Available: https://www.un.org/development/desa/en/news/population/world-population-prospects-2017.html. Accessed on 4 January 2019.

United States Department of Agriculture. 2018. **Sugar: World Markets and Trade.** [Online]. Available: https://apps.fas.usda.gov/psdonline/circulars/sugar.pdf. Assessed on 6 April, 2019.

Walsh K.B., Guthrie J.A. and Burney J.W. 2000. “Applicaion of commercially available, low cost, miniaturused NIR spectrometers to the assessment of the sugar content of intact fruit.” **Australian Journal of Plant Physiology.** 27(12), 1175-1186.

Walton J. 2019. **The 5 Countries that Produce the Most Sugar.** [Online]. Available: https://www.investopedia.com/articles/investing/101615/5-countries-produce-most-sugar.asp. Assessed on 6 April, 2019.

Chapter 2
Theories and literature reviews

2.1 Overview of the sugarcane industry

2.1.1 Importance of the sugarcane industry

Sugarcane is an economically important crop for many countries in the tropics and subtropics. It is used as a main raw material to produce sugar worldwide and is also an increasingly important source of raw material for ethanol biofuel production (Weeks, 2017). Added-value products consisting of sugar, molasses, ethanol, and energy are the overall throughput for the sugarcane industry (Moore, 2013). The main product of this industry is sugar, as global production in 2017/2018 is up 2.8 million tons to 194.6 million (Agriculture, 2018). Approximately 80% of the production from over 70 countries uses sugarcane as a raw material (Walton, 2019). The top five major producers in the worldwide sugar trade are Brazil, India, the European Union (EU), Thailand and China. Among the top producers, Brazil is the largest sugar producer, accounting for almost 52% in market share as well as increasing ethanol production to support increasing domestic demand for alternative fuel (Walton, 2019). For the 2018/2019 sugar overview, global production is forecast to be down by 9 million tons, with the roughly 8-million-ton drop in Brazil being caused by unfavorable weather, low global sugar price and more sugarcane being diverted towards ethanol production (Agriculture, 2018). Thailand's sugarcane industry produces around 15 million metric tons of sugar, which is a record level. This is caused by expanded land areas shifted from cassava production (Walton, 2019). Also, increasing demand for biofuels in the country is expected to be supported by this increase in sugar production. These identify the importance of this industry influencing not only the worldwide sugar trade but also the alternative fuels side.

2.1.2 Sugarcane botany

Sugarcane, *Saccharum officinarum,* is a perennial tropical grass of the family Gramineae and the genus Saccharum (Godshall, 2003). The period around 12-24 months after planting is usually used for harvesting, depending on varieties and

climate (Cuadra et al., 2012). The sugarcane plant consists of a stalk, leaf and root system. Its entire plant is shown in Figure 2.1. An important part of sugarcane is the stalk, used for propagation and for sugar production. The sugarcane stalk, shown in Figure 2.2, is circular in cross-section and is differentiated into segments called joints that comprise a node and an internode. The node is comprised of a lateral bud posited in the axil of the leaf, a root band, and growth ring. Difference in varieties and plant health leads to several types of nodes and internodes illustrated in Figure 2.3.

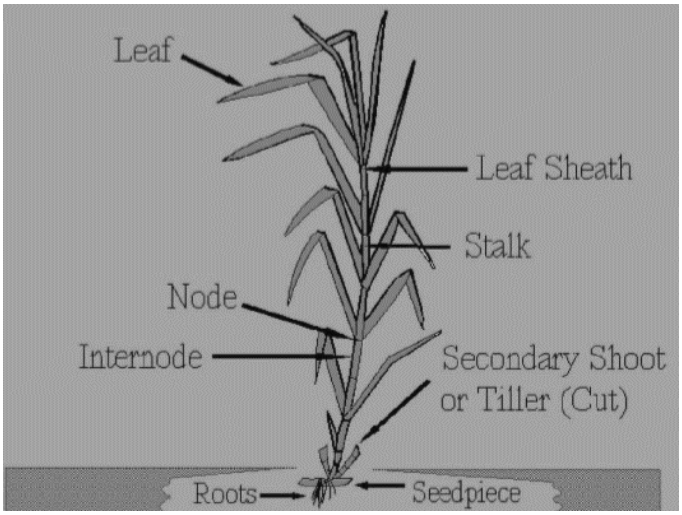


Figure 2.1 The sugarcane plant (Sandhu, 2016).

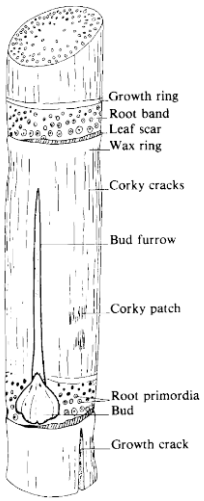


Figure 2.2 Stem of sugarcane (James, 2004).

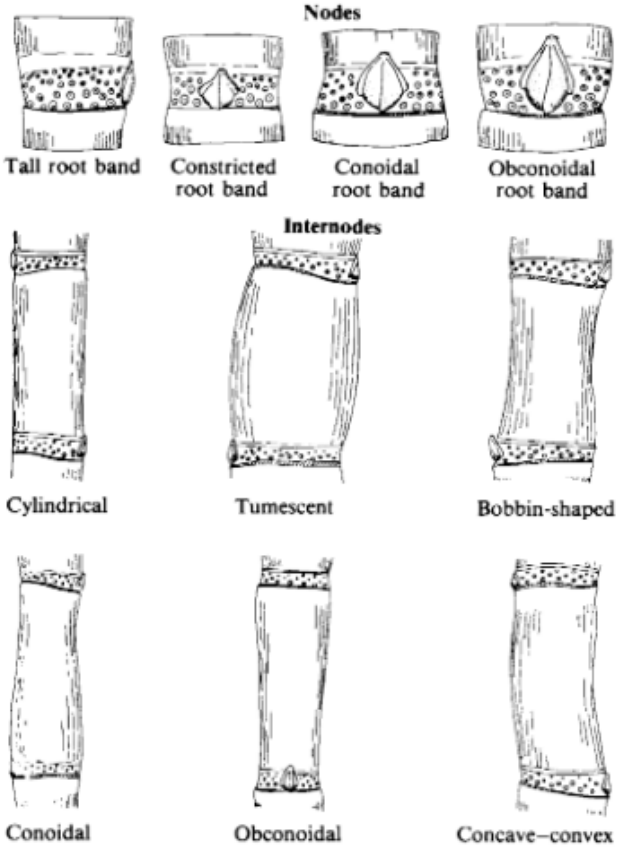


Figure 2.3 Types of sugarcane nodes and internodes (James, 2004).

Growing new sugarcane commercially is done by cutting the stem or stalk into sections that contain one or more buds (James, 2004). After planting, roots originate and develop to supply moisture to the growing buds that germinate and produce more shoots, forming a clump (Nawi, 2014). The primary stem grows from auxiliary buds, whereas the secondary and tertiary stems spring up at the base of the primary, as illustrated in Figure 2.4. Sugarcane is known as the ratoon crop that can regrow from stubble and can be harvested every year for up to 6 years. However, dropping in productivity usually starts from the fourth to fifth crops (Cuadra et al., 2012). Land preparation is done after the last ratoon cycle for the next crop cycle or to grow new crops.

General characteristics of sugarcane stalk vary in hardness, length and diameter, dependent on factors of varieties and growing conditions (James, 2004). Furthermore, there are differences in wax coating and colors, inherent to the nature of this plant and dependent on the varieties. Variation in color is caused by different

levels of two basic pigments from red and blue anthocyanins and green chlorophyll. Whenever yellow is present within the stalk, it arises out of the absence of both anthocyanins and chlorophyll (James, 2004). The internode covered by its leaf sheath causes soft-toned colors and is not alike when exposed to sunlight. These are the common physical characteristics of sugarcane stalks. For the stalk composition, the mature stalk typically contains around 13% sugar (90% sucrose), 12% fiber and 75% water (Henry, 2010). A high sugar content appears at the lower segments of the stalk. Being very rich in sucrose make this plant highly attractive for sugar production worldwide.

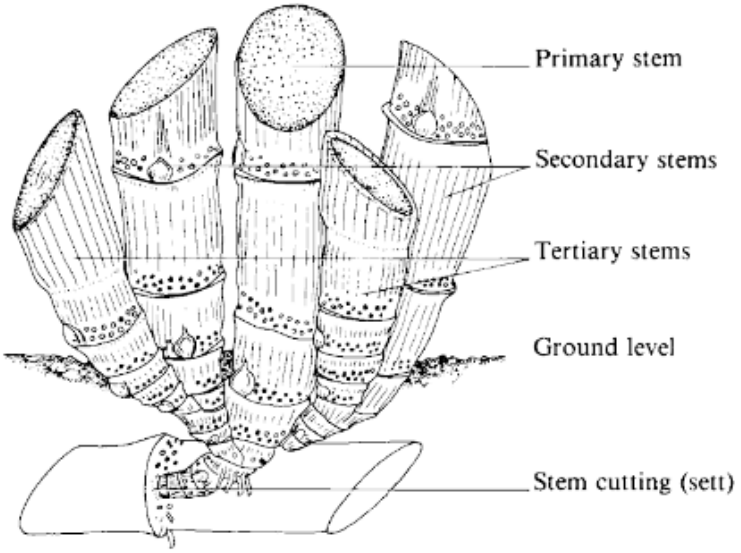


Figure 2.4 The formation of tillers (James, 2004).

2.1.3 Current sugarcane harvesting

Sugarcane is a seasonal crop, typically harvested at peak maturity at around 11 to 12 months after planting, dependent on varieties. Various harvesting practices are used in the sugarcane industry. Manual and mechanical harvestings can currently be found (Netafim's Agriculture Department, 2019). Based on how the sugarcane is presented to be harvested, there are two practices i.e. harvesting with pre-harvest burning (burnt cane harvesting) and without burning (green cane harvesting) (Weekes, 2004; Ma, 2014).

Manual harvesting is a very labor-intensive activity and requires skilled labors to avoid loss of cane and sugar yields. Concerns with regards to fatigue due to excessive stress and exposure to harmful pests from plantations, as well as an increase in harvesting labor being diverted to other industries, is the main issue for this harvesting type (Ma, 2014). With this concern, an assessment of the manual harvesting of the sugarcane in Sudan was conducted to be compared with mechanical harvesting. The results showed that manual sugarcane harvesting (0.2 USD/ton) is more expensive than mechanical means (0.1 USD/ton); payoff for the labor represents 74.14% of the total cutting cost, whereas factor of infield cane losses present 4.72% and 4.22% for the manual harvesting and mechanical harvesting systems, respectively (Ahmed, 2015). This suggests that adoption of mechanical harvesting is unavoidable for the sugarcane industry worldwide. This method can harvest one hectare of sugarcane within 1 hour using only 3.3-4.2 machines, whereas 850-1000 men are required for manual harvesting (Yadav, 2002). This causes continuous growth for the use of sugarcane harvesters, which can be generally grouped into whole stalk harvesters and chopper harvesters (Ma, 2014). In addition, harvesting sugarcane using mechanical methods can reduce “Green House Gas” emissions emitted from the burning necessary in manual harvesting (Figure 2.5) (Braunbeck, 1999). With the increase in mechanical sugarcane harvesters, the chopper harvester (Figure 2.6) is more agreeable to use than stalk harvesters (Figure 2.7).

Figure 2.5 Manual sugarcane harvesting from burned fields (DeFrantz, 2019).

Figure 2.6 Sugarcane chopper harvester working in the field.

Figure 2.7 Whole stalk sugarcane harvester (FOTMA, 2019).

The operation of the chopper harvester and the overview of its components shown in Figure 2.8 have been demonstrated by Cox (2002) as follows: “it begins with chopping the leafy matter off at the top of the sugarcane stalk using the topper. The sugarcane is then knocked over slightly by the knockdown and fin rollers and is cut at the ground level by the base cutters. The feed rollers then pull the whole stalks of sugarcane up into the choppers which cut the sugarcane into billets for 200-300 mm long. At this point, the primary extractor fan removes extraneous matter such as leaf and dirt, present amongst the billets. The billets then fall into the elevator and are lifted to a height suitable for delivering into the vehicle alongside the harvester. For some additional extraneous matter, it is removed by a secondary extractor fan.”

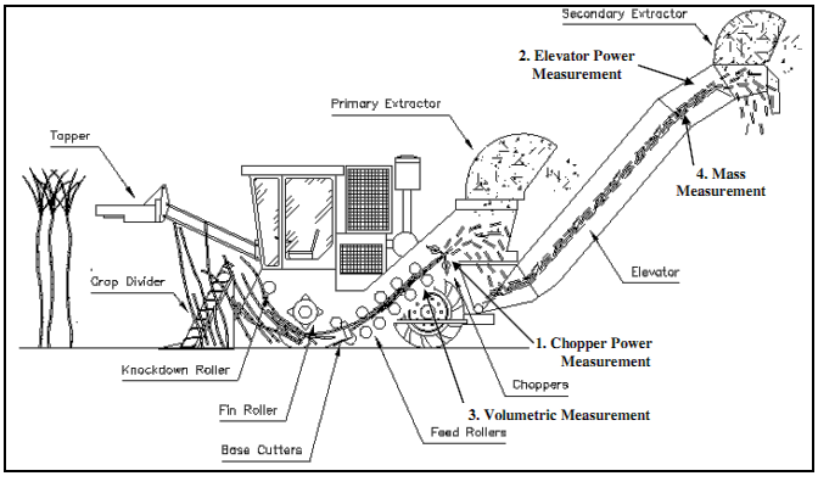


Figure 2.8 View of the components and operation of the chopper sugarcane harvester (Cox, 2002)

The merits and weaknesses of either harvesting with pre-harvest burning or harvesting without burning are the subject of debate. Manual cane harvesting focuses on burning the cane before harvesting because of the ease of work and the greater productivity achievable despite the increased levels dirt and dust (Weekes, 2004). The burning practice can remove about 80% of cane leafy materials without causing significant damage to the interior of the cane stalk (Braunbeck, 1999; Weekes, 2004). The reduce in leafy matters also leads to reduced waste in sugar production (Sopa, 2010). Furthermore, burning before harvesting avoids dangers from pests e.g. snakes and insects entering the fields of the harvesting laborers (Weekes, 2004). However, burning the cane causes negative impacts on sugar yield, owing to saccharose deterioration which leads to a reduced rate of sugar extraction. Moreover, cane burning emissions have become one of the most problematic environmental issues worldwide, increasing Green House Gases and also affecting the respiratory system of populations in nearby communities (Cancado et al., 2006). These concerns have caused sugarcane producers worldwide to realize the effects of burning sugarcane before harvesting, prompting them to switch to green cane harvesting (without burning) instead (Ma, 2014). The harvesting practice through which sugarcane is harvested without any preconditioning is a challenging matter, owing to the increased amount of extraneous materials such as tops and leafy materials (Ma, 2014). However, chopper harvesters are currently able to separate the extraneous materials

using two extractor fans and have been adopted extensively by chief sugarcane producers such as Australia, Brazil, USA and Thailand. Use of the chopper harvester blows out the residual trash, usually spreading evenly over the field. This may provide a more sustainable option for sugarcane production compared to burning the cane before harvesting, as it helps control weeds, reduces soil moisture loss and decreases soil erosion, as well as helping to add organic matter to the soil (Braunbeck, 1999), improving soil fertility and productivity (Núñez, 2008). Green cane harvesting must also account for deterioration, thus the period between harvesting and processing must also be minimized in the same way as it is for burnt cane (Weekes, 2004).

2.1.4 Quality measurement systems for the sugarcane industry

Sugarcane is one of commercial plants that is traded based on its quality. In the sugarcane industry, commercial cane sugar (CCS) is used as a quality index for trading between growers and sugar mills. CCS refers to the percent of recoverable sucrose from fresh cane and is derived from the measurements of soluble solids content (ºBrix), Pol and fiber contents of the sugarcane. Hence, it is a primary factor of economic value for this industry besides the sugarcane yield. The standard CCS could be calculated as the following equations (BSES, 2001):

$$°Brix\ in\ cane = °Brix\ in\ juice \times \frac{100-(\%\ fiber+3)}{100} \quad (2.1)$$

$$Pol\ in\ cane = Pol\ in\ juice \times \frac{100-(\%\ fiber+5)}{100} \quad (2.2)$$

$$Impurities\ in\ cane = °Brix\ in\ cane - Pol\ in\ cane \quad (2.3)$$

$$CCS = Pol\ in\ cane - 0.5 \times (impurities) \quad (2.4)$$

Note: 3 and 5 are correction factors for °Brix and pol measurements, whereas impurities are non-recoverable sugars.

2.1.4.1 Soluble solids content (%SSC or °Brix)

°Brix is the concentration of total soluble solids containing sucrose as a major part, fructose and glucose in a solution or juice (Naderi-Boldaji, 2016). It can be determined using a refractometer that measures the refractive index, monitoring the amount of light bending through liquids. The concept is that the more dissolved solids in water, the slower light travels through it. This produces more of the "bending" effect on light, showing the refracted angle on a scale of the refractometer. The Brix scale is defined in g solute per 100 g solution (ProSciTech, 2014).

2.1.4.2 Polarization (%Pol)

%Pol is a measure of sucrose concentration in juice by using a polarimeter which transmits a polarized light through the juice. The concept is a measure of the rotation angle of the polarized light to calculate the %Pol (Naderi-Boldaji, 2016). It is defined in g solute per 100 g solution).

2.1.4.3 Fiber

Fiber is one of sugarcane's stem components, both dry and insoluble (BSES, 2001; Reddy, 2015). The percentage of fiber is the ratio of dried fiber to the initial weight of a stalk sample (Naderi-Boldaji, 2016). A standard method for fiber determination within the industry is performed by either the whole stalk or prepared cane method (BSES, 2001).

2.1.5 Existing technology for sugarcane industry

2.1.5.1 The existing sugarcane yield monitoring techniques

With the importance of monitoring sugarcane yield variation across the field, several attempts have been conducted to accomplish yield mapping, discussed below:

1) Chopper power measurement

The measure of chopper power is one of the mass flow rate sensing techniques that depends on the assumption of the power required for chopping the sugarcane into billets. It is proportional to the mass flow rate of sugarcane through the choppers (Cox, 2002). Figure 2.9 shows the chopper system of the sugarcane harvester. Yield variation could be estimated by post calibrating the data of mass flow rate with the known yield of the field. Cox (2002) concluded that the

advantages of this technique were that it was the simple sensor design, that the sensor components can be purchased off the shelf and that it was reliable as no part was in contact with crop material. However, some disadvantages were that it was not a direct indication of mass flow rate, and that crop conditions (such as burnt/green cane, moisture content and fiber content) and machine conditions (such as blade sharpness) affected the measurements. This technique was first studied for monitoring sugarcane yield by Cox et al. (1996). They claimed that the system produced a linear output line with a correlation of 0.96. However, Jensen et al. (2010) assessed the commercially available chopper power measurement sensor and concluded that it did not adequately perform as a yield monitor of sugarcane.

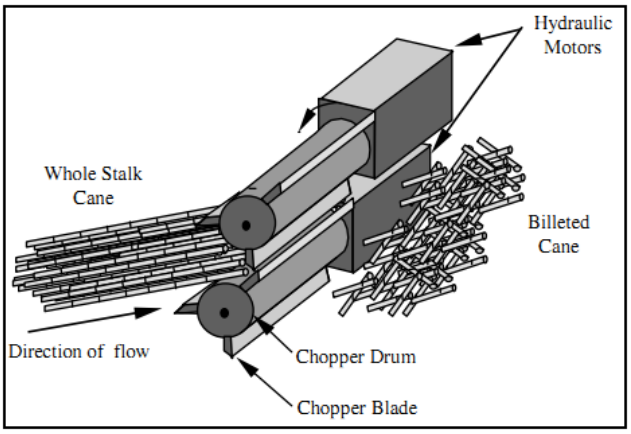


Figure 2.9 The chopper system of the sugarcane harvester (Cox, 2002).

2) Elevator power measurement

The elevator of a sugarcane harvester (Figure 2.10) is a part used to deliver sugarcane billets from the chopper system into the haul-out vehicles. It is driven by two hydraulic motors at the top of the elevator. Cox (2002) explained that with around 3 vertical meters of the billets lifted, energy was required to overcome gravity and friction on the elevator floor. These effects could indicate the sugarcane flow rate through the elevator that was proportional to the power required to move the billets. The hydraulic power can be measured using the same method as the chopper power measurement. Moreover, he explained that most merits were similar to the measure at the chopper position. This technique was a more direct technique of measuring mass flow than the system measured at the chopper. However, it

required the correct operation of the elevator system. Different harvesting conditions such as wet/dry field conditions and green/burnt sugarcane can also change the friction of the billets over elevator floor. Cox et al. (1996) researched this technique to monitor sugarcane yield and reported a linear output line with a correlation of 0.95. However, Jensen et al. (2010) also evaluated the sensor based on the change in pressure across the elevator motor and concluded that the harvester speed had a minimal effect on this technique. Although the result was better than the measurement of the chopper power, it was not enough to be used in practical application.

Figure 2.10 Elevator of the chopper sugarcane harvester (Cox, 2002).

3) Volume measurement using the feed roller operation

Figure 2.11 shows the volume measurement using the feed roller operation on the sugarcane harvester. It consists of a set of cylinders that feed the sugarcane stalks from the base cutters to the chopper system. The feed rollers operate in pairs: the top rollers are adjustable, depending on thickness of sugarcane passing through, whereas the bottom rollers are fixed in position (Cox, 2002). The volume of sugarcane moving through the feed rollers could be measured from the opening or separation between the top and bottom rollers. Most merits are similar to previous techniques but this one is cheap to manufacture and able to measure flow rate before sugarcane losses occur via the extractor fans. However, this works based on the assumptions of sugarcane movement through the feed rollers with constant velocity and width, which may not always be valid. This technique has been

evaluated by Jensen et al. (2010) who concluded that it provided the best result and was potentially a viable sugarcane yield monitor with some cautions.

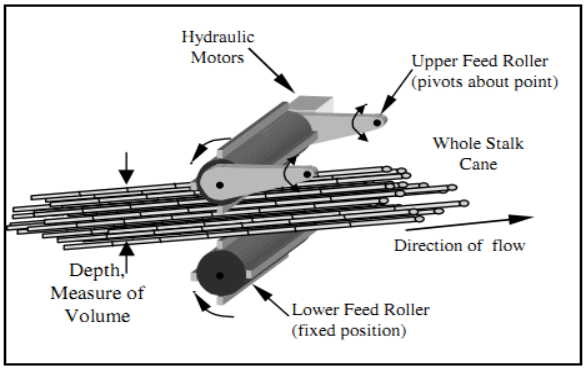


Figure 2.11 Volume measurement using the feed rollers operation (Cox, 2002).

4) Mass measurement using weighing pad

This is the direct measure of the sugarcane flow rate, using a weighing scale plate mounted at the upper portion of the elevator to measure the sugarcane flow rate across the elevator. The concept design has been explained by Cox (2002) as follows "The weigh pad consisted of a plate mounted in the elevator floor, hinged at one end and supported with a load cell on the other end (Figure 2.12). The flights of the elevator pushed the sugarcane billets over the plate to register the mass on the load cell." Calibration was not affected by crop conditions and it was simple to calibrate using single weight. However, there were some disadvantages, i.e. it was sensitive to mechanical noise, had a relatively complex sensor design and contacted with the crop materials (Cox, 2002). Some problems with accumulation of soil, sediment, and debris were also observed in the gap between the weighing scale and elevator floor that causeed locking of the weighing plate to the elevator floor (Price, 2011).

This technique has been investigated with placed, trial and strategic differences:

Benjamin (2002) tested a weighing scale without any support systems and reported the result with a correlation coefficient of 0.966, average errors of 11.05%. The results indicated that different varieties of sugarcane influenced the reading values.

Pagnano and Magalhaes (2001) tested this technique with some support systems (accelerometers and low pass filter system). They reported the range of errors from 0.35 to 4.02% in laboratory tests and 0.83 to 28.66 % in field tests, and discussed that the range of errors was probably from the starting and stopping of the harvester during harvest.

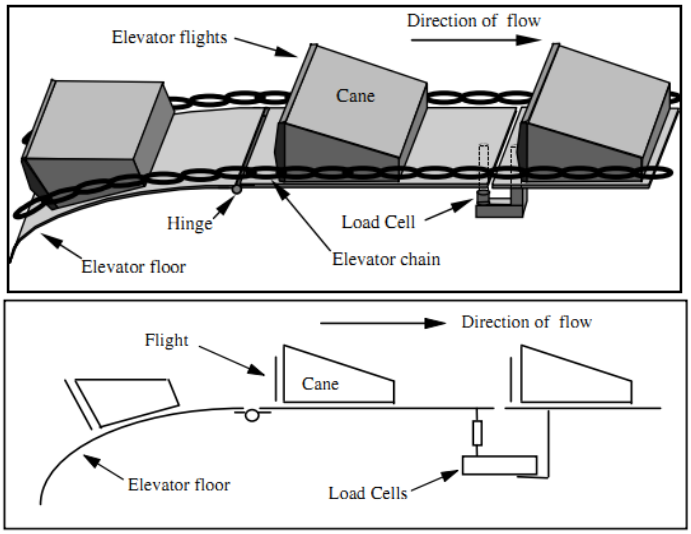


Figure 2.12 Conceptual design of elevator weigh pad (Cox, 2002).

5) Sugarcane flow measurement using optical sensor

Typically, this technique works well for monitoring yield in some crops such as peanut and cotton. Hence, it has been studied by Price et al. (2011) who adapted to the sugarcane harvester for yield monitoring (Figure 2.13). A duty-cycle type manner with three fiber optic sensors mounted in the elevator floor was used to estimate sugarcane yield. Their result gave the linear relationship between the optical sensor response and the actual sugarcane yields with a coefficient of determination (R^2) of 0.98 and the average prediction error of 7.5% for a field test. However, they reported that the error decreased according to the increase of the harvested area (tonnage) with an estimated error of 0.03% for 57.8 Mg loads. From the results, they claimed that the sugarcane variety, harvester speed, harvested distance, and moving direction of the harvester did not affect the sugarcane yield prediction.

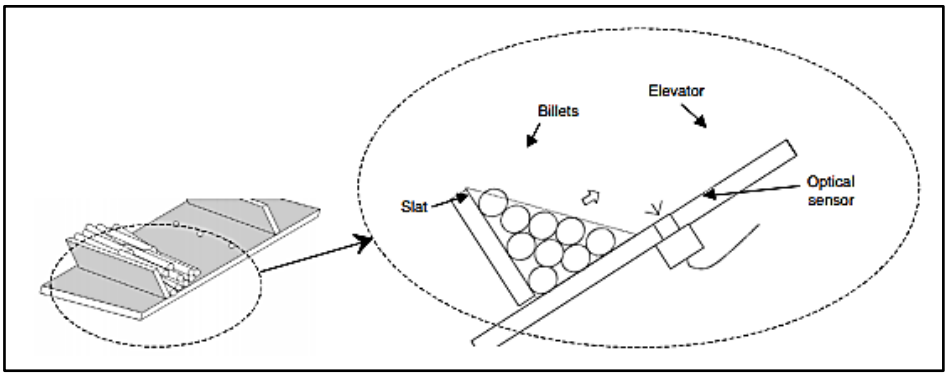


Figure 2.13 Strategy to detect billets from the elevator floor using optical sensors (Price, 2011)

6) Sensing sugarcane volume using ultrasonic sensors

This technique is the invention of Harvestmaster (Juniper Systems Inc.) and is used for potato harvesting. According to Price et al. (2007), they reviewed with the use of this technique on sugarcane harvesters, mounting an overhead ultrasonic sensor set over the yield conveyor to sense the volume of sugarcane on each slat. When used for several hours a day, the calibration started to drift. In this case, they did not know why this problem occurred and concluded that a possible reason was that the sensors became dirty, as well as the effects of air and temperature.

2.1.5.2 The idea of monitoring sugarcane quality during harvesting

The quality of sugarcane, which is the measurement in terms of sugar content, is one of the components that is used for the basis of the payment system to the growers. Nowadays, the quality information of each field could randomly be determined to be representative of the field. With the spatial variability of sugarcane quality, the idea of monitoring sugarcane quality across the field appears. Many more benefits of the sugarcane quality measurement in the field have been exemplified by Nawi et al. (2014), as there could potentially be a rapid in-field quality measurement system during harvesting which could be used to optimize the sugarcane quality. Hence if monitoring of the quality during harvesting occurs, site-specific information is obtained. To monitor the sugarcane quality during harvest, some sensors or instruments need to be mounted at any positions of the harvester. These have been conducted to look for alternative techniques for measuring the quality in the field, such as:

1) Electronic refractometer

It was developed by McCarthy and Billingsley (2002) and was used to determine the optimum topping height of sugarcane stalk during harvesting. It has been determined by Nawi et al. (2014) that juice samples deposited onto the sensor during harvesting were insufficient. They also reviewed the results and reported the poor results of this system, finding that a large amount of trash and leaf materials hindered the freshly topped stalks from wiping across the sensor.

2) Microwave applications

It was applied in the sugarcane industry and has been reviewed by Nawi et al. (2014). They found that several researchers such as Nelson (1987), Klute (2007) and Shah and Joshi (2010) have measured the sugar content but it was still popular for laboratory use.

3) Near infrared (NIR) spectroscopic method

This technique works well for predicting the sugar content in many agricultural materials from solid samples form such as in apples (Huang, 2010), guavas (Hsieh, 2005) and pineapples (Chia, 2012). In the sugarcane industry, the attempts at predicting the sugar content from stalk form have been conducted based on cross-sectional scanning (Nawi et al., 2013a) and skin scanning (Nawi et al., 2013b). The best model predicted samples based on cross-sectional scanning and skin scanning gave a coefficient of determination (R^2) of 0.87 and 0.91, and root means square error of prediction (RMSEP) of 1.45 and 0.721 °Brix, respectively. Their results implied that the NIR approach was the most promising technique for monitoring sugarcane quality in the term of sugar content.

With the interest of spectroscopic approach, Nawi et al. (2014) reviewed the possible sample forms and their measurement locations on the harvester when using this technique as an on-the-go measurement system. Their reviews divided the possible sample forms for practical application of this technique into three types i.e. leaf samples, juice samples, and stalk samples. The possible locations where each type could be measured on the harvester are shown in Figure 2.14. For leaf samples, the NIR spectroscopic technique could be used to sense chlorophyll status in corn (Tumbo et al., 2002) and to measure nitrogen content in sugarcane leaf (Abdel-Rahman et al., 2010). Interestingly, good correlation between a specific wavelength

for chlorophyll absorption at 680 nm and sugar content in melons was obtained according to the report of Sugiyama (1999). As such, this showed the possibility of whether chlorophyll/nitrogen-response scanning in sugarcane leaf is required for sugar content prediction. Hence, installation of a spectrometer on the harvester was proposed, to be mounted at the topper to sense the sugarcane leaf and predict for sugar content (Nawi et al., 2014). In case of juice samples, Nawi et al. (2014) proposed that the chopper drum was probably the best place for installation of the spectrometer to measure the sugarcane quality from juice samples without any sample preparation. However, shortage of juice samples for the measurement, the presence in large amounts of trash and limited space caused this place to be dismissed for the sensor installation. Stalk samples for sugarcane quality measurement have been grouped by Nawi et al. (2014) into sub-three forms i.e. a standing stalk, cut stool (a living stump of the cane on the ground after harvesting) and billets. At first, they suggested that in front of the crop divider was the best place to install a spectrometer to scan the standing stalk while the harvester was moving forwards. However, they found that the presence of trash, dead leaf and bud might be an obstacle for NIR scanning. Secondly, they illustrated that installing a spectrometer on the base cutter to scan the stool for measuring the sugarcane quality was possible. However, this strategy might encounter the moist surfaces of cut sugarcane which was dirty and soiled. As such, this could cause the inaccurate measurement. Finally, the quality measurement on billet samples was the strategy that seemed to be the most promising. This approach has been proposed in their reviews, suggesting that the measurement could be done by scanning the billet samples on the elevator conveyor because around 75-90% of trash was blown out by the primary extractor according to the report of Mailander et al. (2010).

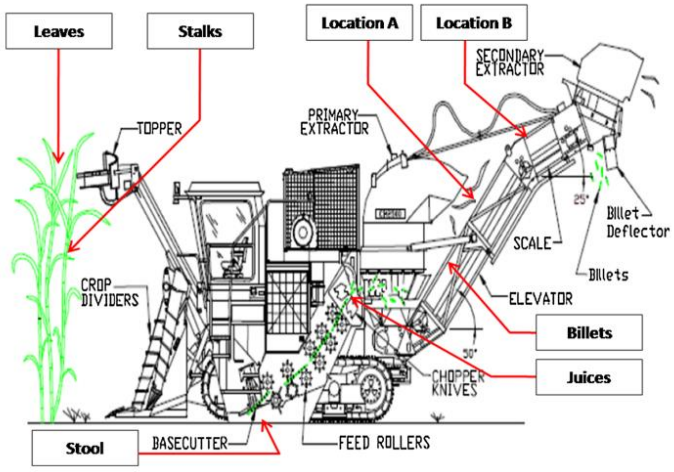


Figure 2.14 Possible locations where each sample type could be obtained (Caryn, 2002).

Two potential methods for scanning the cane billets on the harvester have been proposed by Nawi et al. (2014). Their ideas were to directly scan the sugarcane billets moving on the elevator (called the direct scanning method, DSM) and to scan the billet samples which were supplied by a sampling mechanism (called the assisted scanning method, ASM). In Figure 2.14, they suggested that the location A angled at 50° to the ground of the elevator was suitable for the ASM due to the fact that billets tend to group on the slat. For location B, they reported that it was the suitable location for the DSM as billets between two slats would be distributed evenly. For the DSM, it has been proposed that the sensor should be mounted in the middle of a slat, whereas a light source should be installed at an angle of 45° from the sensor as shown in Figure 2.15. The ASM idea relied on a vacuum system to assist the sampling mechanism, which could be mounted at the location A as shown in Figure 2.14. For more details, they described as follows: “At 50° slope of the elevator, billets tend to group on the slat wall as shown in Figure 2.16. Hence, the empty space between each slat appears. A vacuum nozzle can be placed at the top of the slat tip and was assumed that only one billet located at the place would be vacuumed. Once vacuumed, the billet goes into the measurement chamber. It is positioned at the first metering device in a horizontal orientation to ensures that only one billet enters the measurement chamber at one time. The second metering device holds the sample for scanning then returns it back onto the elevator.”.

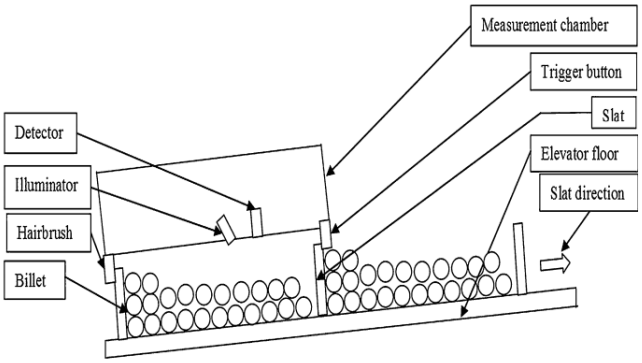


Figure 2.15 Schematic diagram for DSM (Nawi, 2014).

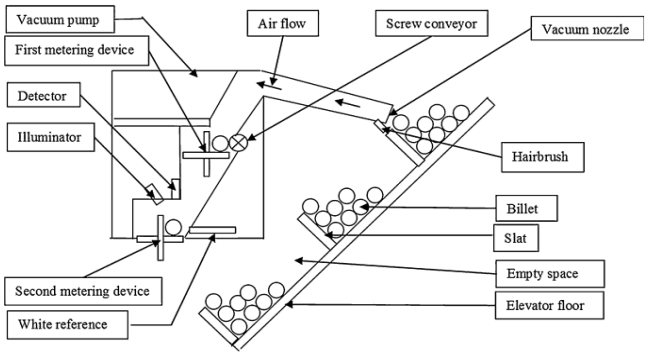


Figure 2.16 Schematic diagram for ASM (Nawi, 2014).

2.2 Introduction of near-infrared spectroscopy and its application

2.2.1 Theory of near-infrared spectroscopy

Infrared energy is the electromagnetic energy of molecular vibration, defining the NIR energy band as 780 to 2,500 nm for convinience, as the infrared (or mid-infrared) is 2,500 to 4,000 nm; and for the far-infrared is 4,000 to 10,000 nm (Workman, 2015). However, the NIR region of the electromagnetic spectrum was also defined by The American Society of Testing and Materials (ASTM) as the wavelength range of 780–2,526 nm corresponding to the wavenumber range of 12820–3959 cm^{-1} (Burns and Ciurczak, 2001; Siesler et al., 2002; Reich, 2005). As shown in Figure 2.17, the NIR light is a small range, roughly extending from the red edge of visible light to 2500 nm. Its energy causes molecular chemical bonds to vibrate at specific frequencies. This causes the bonds absorb the energy and to excite themselves to a higher energy level, resulting in transitions between vibrational energy states that relates to overtones and combinations of fundamental vibrations of C–H, N–H, O–H

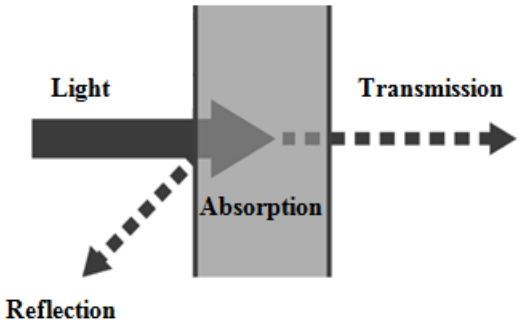


Figure 2.18 Interaction of radiation with matter.

2.2.1.2 Absorption of radiation

If equation 2.5 is considered only by focusing on absorbed intensity in order to measure either the transmitted or reflected intensity, then the absorption of radiation by a sample involves the loss of energy through radiation. However, Osborne et al. (1993) mentioned the Beer-Lambert law which related the light attenuation of the transmitted radiation when traveling through the sample as follows:

The fraction dI/I of intensity I absorbed by an infinitesimal thickness of sample was proportional to the number of molecules dn in that thickness, or

$$-dI/I = kdn \tag{2.6}$$

Take integrate into equation 2.6

$$-\int_{I_0}^{I_T} \frac{dI}{I} = k\int_{0}^{n} dn$$

$$-\ln(I_T - I_o) = k(n-0)$$

$$ln\, I_0 - ln\, I_T = kn$$

$$ln(I_0/I_T) = kn \tag{2.7}$$

When the transmittance, T was I_T/I_0, where I_0 was the intensity of the incident and I_T that of the transmitted intensity, n was the number of molecules in the path of the beam, therefore:

$$ln\frac{1}{T} = kn \tag{2.8}$$

However, an absorbance (A) equation was written by logarithm using base 10, so that:

$$\frac{1}{T} = e^{kn}$$

$$log\frac{1}{T} = log\, e^{kn} = kn\ log\ e = 0.434kn$$

$$A = log\frac{1}{T} = kn \tag{2.9}$$

where $k \approx 0.434k$, since n was the proportional to the concentration (c) of molecules in the sample and the thickness (b) which the radiation passed, and the constant a was called the absorptivity.

$$A = log\frac{1}{T} = log(I_0/I_T) = abc \tag{2.10}$$

and equation 2.10 became

$$A = abc \tag{2.11}$$

2.2.1.3 Reflection of radiation

Reflective behaviors of the incident radiation on the samples depend on the various factors such as the matter surface and the different particles within the sample. As such, figure 2.19 shows the reflective behavior of both smooth and rough surfaces. Osborne et al. (1993) reported that a sample which was opaque and non-absorbing would reflect in a manner similar to the law of mirror, whereas if the sample surface was rough, the beam of light would reflect and diffuse in many different directions.

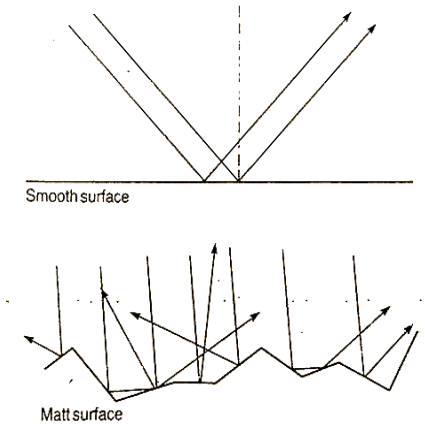


Figure 2.19 Surface effect-specular reflectance (Osborne, 1993).

Moreover, there are some cases in which the radiation can be transmitted through the sample surface and then absorbed according to the Beer Lambert law; specifically, this can be found in solid samples such as powder, rice and so on. These behaviors were also explained by Osborne et al. (1993), wo said that the radiation faced the discrete particles within the sample and disseminated in all directions; this phenomenon was discovered by Tyndall (1869) and is known as 'scattering' (Figure 2.20).

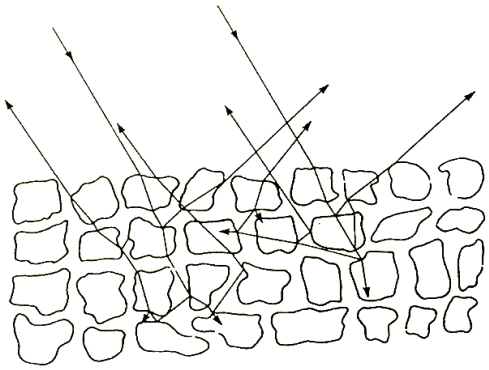


Figure 2.20 Diffuse (body) reflectance (Osborne, 1993).

The scattering effect is difficult to describe mathematically for particulate samples. However, Osborne et al. (1993) mentioned the theory which was proposed by Kubelka and Munk (1931) which states that the absorbance of a completely opaque layer of infinite thickness depends on its scattering and absorption, as per the following equation:

$$\frac{(1-R_\infty)^2}{2R_\infty} = \frac{k}{s} \quad (2.12)$$

where R_∞ was the reflectance of the infinitely thick layer, s was the scattering constant and k was the absorption constant. According to the Beer-Lambert law (Kortum et al., 1963; Birth and Zachariah, 1976; Osborne et al., 1993), the absorption coefficient k in equation 2.10 is in fact equal to concentration (c) multiplied by the absorptivity (a) as below:

$$\frac{(1-R_\infty)^2}{2R_\infty} = \frac{ac}{s} \quad (2.13)$$

In actuality, Osborne et al. (1993) referred to the Norris's study in 1983, mentioning that he measured the diffuse reflectance of a non-absorbing standard in order to compare with that of the sample. The result of conversion to the common logarithm produced a nearly linear relationship with concentration of sample as following ways:

$$log(R'/R) = log\,1/R + log\,R' \sim ac/s \quad (2.14)$$

where R' was the reflectance of the standard and R that of the sample $(R' > R)$. Since the sample which was non-absorbing ($R' \to 1$), $log\,R'$ was therefore constant and was ignored. It might be rewritten as below:

$$log\,1/R \sim \frac{ac}{s} \quad (2.15)$$

2.2.2 Sample presentation

Presenting a sample to a NIR instrument is an important factor as it affects the spectral response. The four main sample presentations (Figure 2.21) consist of transmission, reflection, transflection and interaction. Lighting at one side of the sample and the detection of the transmitted part of such light from the other side is called transmission, which is suitable for liquids. However, the transmittance mode has also been applied for solid samples, for instance to determine the sugar content of mandarins (Kawano, 1993). Whenever a light source is mounted at a specific angle, typically 45° from a detector to avoid specular reflection, to illuminate at the

sample surface and diffusely reflected part is detected from the surface, this is reflection measurement, more suitable for an opaque sample. Transflection is the combination of transmission and reflection techniques. Light is passed through the sample layer and then scattered back from a reflector to the detector. As such, the pathlength of this measurement is 2 times. In the case of interaction, an interaction probe with a concentric outer ring for emitting light and an inner portion for receiving light is used, contacted with the surface sample (Siesler, 2008).

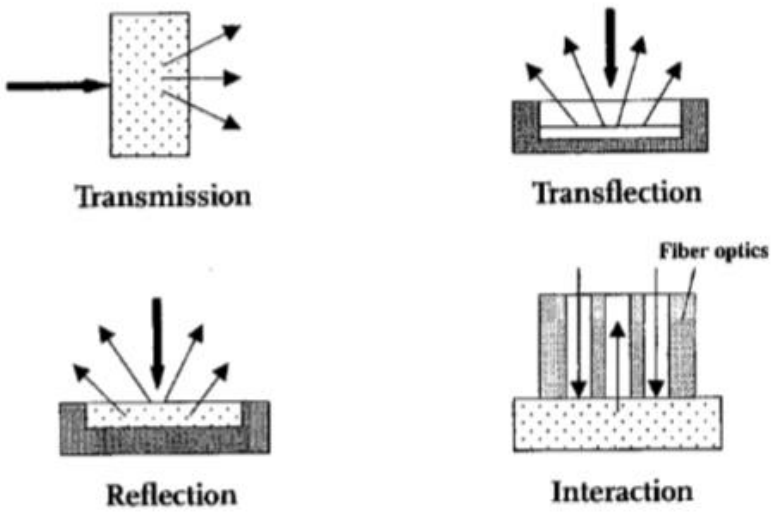
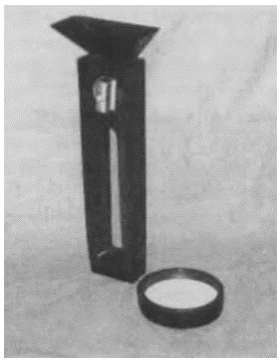
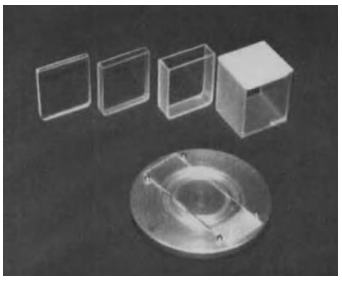


Figure 2.21 Sample presentations of transmission, reflection, transflection and interaction (Siesler, 2008)

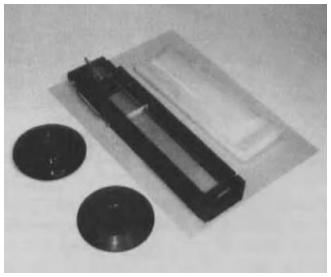
There are several kinds of sample holders, developed by the manufactures and adopted for spectral measurement. The holders (Figure 2.22) are designed for whole grains, single kernel, powdered samples, pastes, liquid and fruit, mostly used for laboratory circumstance (Siesler, 2008). For online applications and non-contact measurement, a fiber optic probe is suitable.



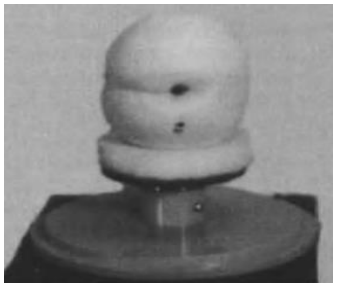
a



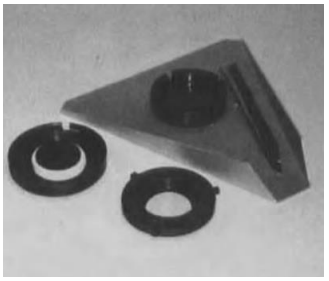
B



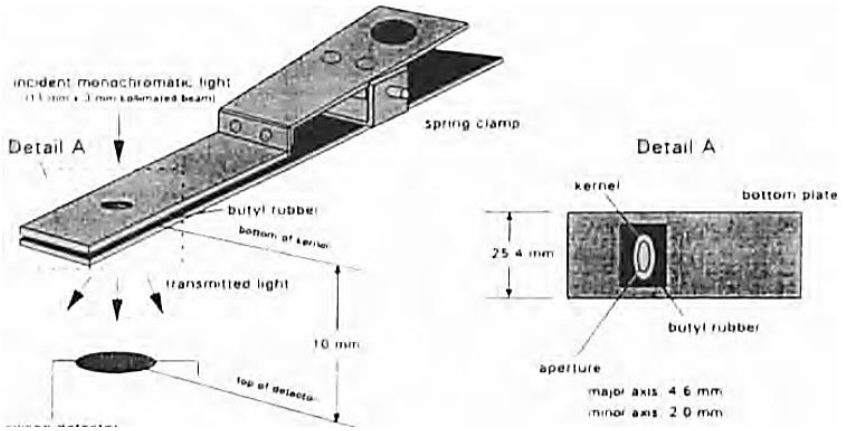
c



D



e



f

Figure 2.22 Sample holders for different types of sample; a) whole grain, b) liquid, c) pastes, d) fruit, e) powdered sample and f) single kernel (Siesler, 2008)

2.2.3 Near-infrared (NIR) instrumentation

An NIR spectrometer usually includes a light source (usually a tungsten halogen light bulb), sample presentation interface, monochromator, detector, and optical components (lenses, collimators, beam splitters, integrating spheres as well as optical fiber) (Nicolaï et al., 2007). All spectrometers are used to collect spectral data, characterizing changes in absorption intensity as a function of frequency or wavelength. Conveniently, common spectrometers could be classified by the types of monochromator. Filter instruments are the simplest and cheapest type, containing a limited number of wavelengths. They can be both a wheel holding the number of absorptions and interference filters (Nicolaï et al., 2007). This type of instrument is suitable for applications where the exact absorption region is known, e.g. protein, moisture and oil in agricultural samples (Osborne, 2006). In a scanning monochromator instrument, a grating is used to separate the individual frequencies or wavelengths of NIR radiation. Rotation of this wavelength separator causes the individual wavelengths to reach a detector (Nicolaï et al., 2007). The Fourier Transform (FT) Spectrophotometer operates by using an interferometer to generate modulated light. The time domain signal of the light reflected or transmitted from the sample is converted into a spectrum via a fast Fourier transform (Nicolaï et al., 2007). A photodiode array (PDA) spectrophotometer involves employing an array of photodiodes as a detector. A fixed grating is used to disperse the radiation onto the array (Nicolaï et al., 2007). This type of the instrument is very useful when rapid online measurements are required thanks to having no moving parts for wavelength selection, leading to the ability to scan at high speed (Osborne, 2006). Also, its low-cost and partibility are other advantages of this instrument. Based on this, the PDA spectrometer normally equipped with a Si detector that works in the electromagnetic region of 350-1100 nm is under considerable attention for field use applications (Nawi, 2014).

2.2.4 NIR spectroscopic applications in the sugarcane industry

NIR application was conducted and was pioneered to be used as a rapid method for sugarcane juice analysis a long time ago. Such analysis determined °Brix, pol, fiber, and moisture content (Nawi, 2014). This prompted researchers in this area

to apply NIR spectroscopy for both qualitative and quantitative measurements (Madse-li, 2003; Valderrama, 2007; O'Shea, 2011). The NIR spectroscopy has been developed for analysis of molasses, mixed juice and massecuites in a laboratory space (Schaffler, 1997; Simpson, 2010) and that of sugarcane, bagasse and sugar in factories with online applications (Staunton, 2006). Furthermore, the several studies regarding the spectroscopic applications in the sugarcane industry have been summarized in Table 2.1 by Nawi (2014), who conducted the NIR spectroscopic technique to determine sugarcane quality from stalk samples and claimed that the technique was possible to develop as a sensor for monitoring sugarcane quality on the harvester.

Table 2.1 Typical application of spectroscopic methods in the sugarcane industry (Nawi, 2014).

Wavelength, nm	Measurement mode	Sample form	Prediction accuracy
NIR (1445 - 2348)	Reflectance	Fibrated	°Brix (R^2 = 0.91) CCS (R^2 = 0.91) Fiber content (R^2 = 0.89) Pol (R^2 = 0.96)
NIR (1445 - 2348)	Transflectance	Clarified juice	°Brix (R^2 = 0.97) CCS (R^2 = 0.97) Pol (R^2 = 0.98)
MIR (8000 - 12500)	Reflectance	Raw juice	Pol (R^2 = 0.98)
NIR (1111 - 2222)	Transmittance	Raw juice	Pol (R^2 = 0.96)
NIR (1111 - 2500)	Reflectance	Fibrated	Pol (R^2 for the calibration model = 0.93)
NIR (1100-2500)	Reflectance	Fibrated	Pol (R^2 = 0.96) °Brix (R^2 = 0.97) Fiber content (R^2 = 0.90)
NIR (1100 - 2500)	Transmittance	Clarified juice	°Brix (R^2 = 0.99) Pol (R^2 = 0.99)
NIR (1100 - 2498)	Reflectance	Fibrated	Pol (SEP = 0.21%)
VNIR (600 – 1100)	Transmittance	Stalk	Pol (RMSEP = 1.1% Pol)

2.3 Chemometrics

Chemometrics is the chemical discipline that uses mathematical and statistical methods to design or select optimal measurement procedures and experiments, and to provide maximum chemical information or to extract relevant information from chemical data (Otto, 2017). Analysis of complex mixtures or determination of an interest in complex materials like food and agricultural products requires the use of chemometrics (Nicolaï et al., 2007). As water is the most important chemical constituent of these materials and is the main absorber in near-infrared (NIR) radiation, this NIR spectrum is dominated by water, and as such, causes information of interest to be masked. This leads to the NIR spectrum being highly convoluted. Here, challenging spectral characteristics complicated by scattering effects, tissue heterogeneities, instrumental noise, ambient effects as well as other sources of variability, are not considered. These factors definitely affect the response of NIR spectrum. Multivariate statistical techniques, also called chemometrics, are required to extract the information of interest in the spectrum (Nicolaï et al., 2007). This involves modeling techniques paired with spectral preprocessing (Geladi, 1995).

2.3.1 Spectral preprocessing

Spectral preprocessing is the mathematical manipulation of data prior to the main analysis. It is used to remove or minimize irrelevant information in the spectra for improving the model accuracy (Andrade-Garda, 2009). The common pretreatments can be divided into two main groups for smoothing or differentiation and for spectral normalization (Tomuta, 2018).

2.3.1.1 Smoothing and differentiation

The smoothing approach is proposed to remove random noise from spectra. Two common techniques of smoothing well-known and widely used are the moving average filter and the Savitzky-Golay filter (Andrade-Garda, 2009). The moving average relies on running along the spaced data or spectrum of a predefined window in which the filter width is defined (Figure 2.23). The calculation of filtering raw signals or raw spectral data is expressed as the following equation (Otto, 2017):

$$y_k^* = \frac{1}{2m+1} \sum_{j=-m}^{j=m} y_{k+j} \tag{2.16}$$

where y_k^* is filtered value, k is the index for actual data point, $2m+1$ identifies the window size or filter width, and m is the number of variables used to set the window size.

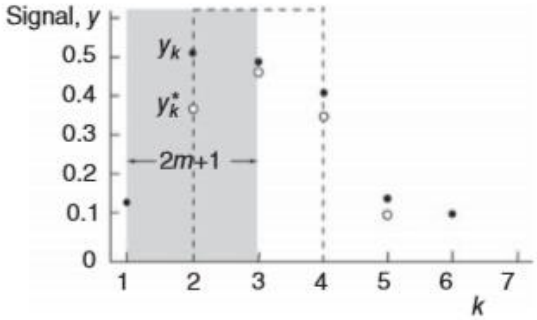


Figure 2.23 Moving-average filter for a filter width of $2m+1=3$ (Otto, 2017).

In case of the Savitzky-Golay method or polynomial smoothing, the filtering or smoothing result is very effective as raw data points filtered by this method are weighted differently, whereas all data are weighted by the same factor, $1/(2m+1)$, in case of the moving average filter (Otto, 2017). Equation (2.17) expresses the calculation of this method (Varmuza, 2016).

$$x_j^* = \frac{1}{N} \sum_{h=-k}^{k} c_h x_{j+h} \tag{2.17}$$

where x_j^* is the smoothed value, N is the normalization factor, k is the number of adjacent values at each side of j, c_h are the filter coefficients depending on the used polynomial degree.

Differentiation or the derivative is used to enhance spectral resolution and also to remove both additive and multiplicative effects in the spectra (Tomuta, 2018). If the first order is used, called first derivative, it can remove the baseline spectral behavior, whereas if the second order is used or second derivative, both

baseline and linear trend in the spectral can be removed. The basic concept for derivation explained by (Rinnan, 2009) is that the difference between two subsequent spectral data points is determined for the first derivative, whereas calculating the difference between two successive points is the second derivative.

$$x_i' = x_i - x_{i-1} \tag{2.18}$$

$$x_i'' = x_i' - x_{i-1}' = x_{i-1} - 2 \cdot x_i + x_{i+1} \tag{2.19}$$

where x_i' refers the first-order derivative spectra and x_i'' denotes the second-order derivative spectra at wavelength i.

2.3.1.2 Spectral normalization

Normalization is a mathematical tool used to correct the multiplicative effects, the shifts and the trends in baseline, which can be subdivided into two main groups. The first is simple normalization methods (mean, maximum and range normalizations and standard normal variate) that require only the information from the spectrum itself to be normalized, whereas the other requires the reference spectra to be corrected, multiplicative scatter correction (Tomuta, 2018). Each brief concept is as follows:

$$x_{mean.nor} = \frac{x_i}{x_{mean}} \tag{2.20}$$

where $x_{mean.nor}$ denotes the absorbance of mean normalization at any given variable, x_i is original absorbance value at given wavelength and x_{mean} is averaged spectral data.

$$x_{max.nor} = \frac{x_i}{x_{max}} \tag{2.21}$$

where $x_{max.nor}$ is maximum normalization value at any given variable, x_i is original absorbance value at given wavelength and x_{max} is the maximum absolute value of the spectral data.

$$x_{range.nor} = \frac{x_i}{x_{i(max)} - x_{i(min)}} \quad (2.22)$$

where $x_{range.nor}$ is range normalization value at any given variable, x_i is original absorbance value at given wavelength and x_{max} and x_{min} are the maximum and minimum absolute values of the spectral data, respectively. Standard normal variate (SNV) starts with mean centering and consists of dividing mean-centered spectra by the standard deviation over the spectral intensities, causing each spectrum to return a mean and a variance of 0 and 1, respectively (Tomuta, 2018). Equation (2.23) expresses the calculation of the SNV.

$$x_{corr} = \frac{x_i - \bar{X}}{SDev(X)} \quad (2.23)$$

where x_{corr} is the normalized value at any given wavelength, x_i is raw absorbance value at given wavelength, whereas $\bar{X}$ and $SDev(X)$ is the average and the standard deviation of the spectrum. For multiplicative scatter correction (MSC), its basic concept is the same as that for the simple normalization methods such as SNV except that reference signal or spectrum is required (Rinnan, 2009). This causes their results to be quite similar. The MSC calculation is comprised of two steps, starting with the estimation of correction coefficients and then the correction of the spectrum using such coefficients (Tomuta, 2018) as follows:

$$x_{org} = b_0 + b_{ref} \cdot x_{ref} \quad (2.24)$$

$$x_{corr} = \frac{x_{org} - b_0}{b_{ref}} \quad (2.25)$$

where x_{corr} is corrected spectrum, x_{ref} is reference or averaged spectrum, x_{org} is original sample spectrum, whereas b_{ref} and b_0 are slope and intercept obtaining from the estimation.

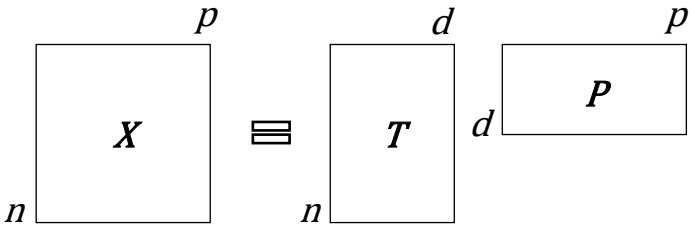
2.3.2 Pattern recognition and classification

Pattern recognition or grouping analytical data can typically be divided into two main groups, i.e. unsupervised and supervised methods. The unsupervised

methods are the strategy to classify the data with no supervisor in the sense of known membership, such as principal component analysis (PCA) and cluster analysis (Otto, 2017). On the contrary, if the membership of objects is known then this is a supervised method called pattern recognition, for instance, linear learning machine (LLM), discriminant analysis, the soft independent modeling of class analogies (SIMCA), and support vector machines (SVMs) (Otto, 2017). However, only PCA and cluster analysis applied for spectral classification in this thesis are provided in concept and theory.

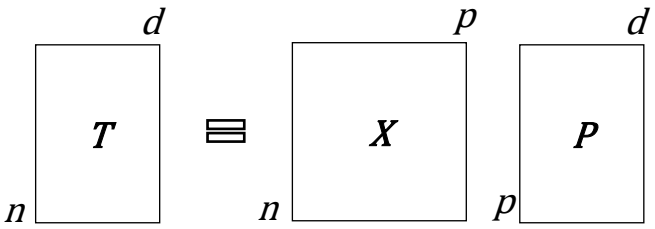
Principal component analysis (PCA) is one of the projection methods normally used as the dimension-reduction technique. The key idea of PCA is to decompose the original matrix (X) by a product of the score (T) and loading (P) matrices, which can be formulated as follows:

$$X = TP^T \tag{2.26}$$



where X is consisted of n rows (observations) and p columns (variables). n represents the number of principal components (PCs). With this concept, projection of the original data matrix (X) onto the lower dimensional space (T) is considered as the PCs determination. This changes the equation (2.26) to be:

$$T = XP \tag{2.27}$$



Each new score t (PC) is estimated as linear combinations of the variables x with the loading p. The first vector or PC explains the most variance in the original matrix. The elements of each PC are as follows (the first PC is expressed as an example):

$$\begin{aligned} t_{11} &= x_{11}p_{11} + x_{12}p_{21} + \cdots + x_{1p}p_{p1} \\ t_{21} &= x_{21}p_{11} + x_{22}p_{21} + \cdots + x_{2p}p_{p1} \\ &\vdots \\ t_{n1} &= x_{n1}p_{11} + x_{n2}p_{21} + \cdots + x_{np}p_{p1} \end{aligned}$$

Plotting the elements of two scored vectors of the matrix (T) is used to interpret for discrimination.

The cluster analysis belonging to unsupervised pattern recognition is the measure of similarity of the objects, decided based on their distance from each other. The shorter the distance between objects, the more similar they are (Otto, 2017). The Euclidean distance determination, widely used and adopted in this thesis, is demonstrated graphically and expressed according to:

$$d_{ij} = \left[\sum_{k=1}^{K} \left|x_{ik} - x_{jk}\right|^2\right]^{1/2} \tag{2.28}$$

where K is the number of variables; i, j are the indices for object i and j.

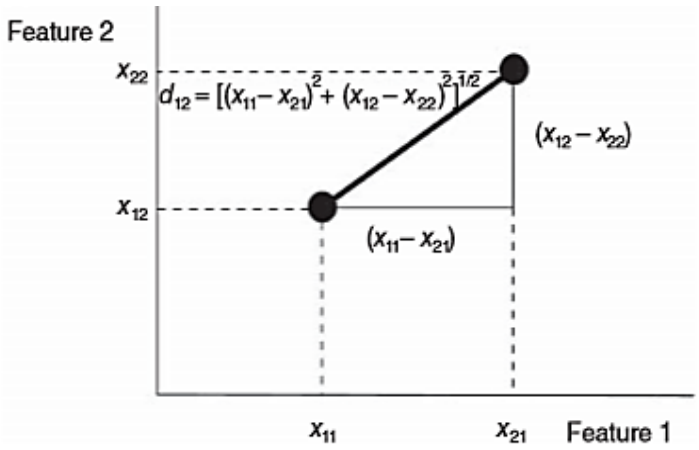


Figure 2.24 Euclidean distance for two features (Otto, 2017).

2.3.3 Modeling

Several methods are used for modeling, such as multiple linear regression (MLR), principal component regression (PCR) and partial least square regression (PLS). However, only the PLS used in this thesis is focused on in this context.

2.3.3.1 Introduction to partial least square regression (PLS)

The aim of the common regression is to model one response, y, by a set of the predictor variables, X. To do so, the multiple linear regression (MLR) works well if X - predictors are fairly few and uncorrelated (Wold et al., 2001). However, when a number of predictors are too large and strongly correlated to each other, MLR is inappropriate (Tobias, 1995).

PLS regression (PLSR) is well known for analyzing data with numerous predictors, including being collinear, and also simultaneously modelling several responses, y (Wold et al., 2001). Abdi (2003) has defined that the PLS generalizes and combines features from principle component analysis (PCA) and multiple linear regression (MLR), i.e. it tries to simultaneously decompose X and y to find a few "new" variables (called latent variables, LVs) that explain as much as possible of the covariance between X and y. These LVs are called $X - scores$ and are orthogonal to each other, eliminating the multicollinearity problem. They are used as the predictors of y. As such, the regression coefficient (b) will be calculated and used to predict $\hat{y}$ values. In order to understand the details of PLS calculation and how it works, the following sections need to be studied, consisting of eigenvalues and eigenvectors, singular values decomposition and how it works, and the PLSR algorithms and its calculation.

2.3.3.2 Introduction to eigenvalues and eigenvectors

A is a square symmetric matrix. An eigenvector of A is a vector v with elements which are not all zero, whereas λ is an unknown scalar (Reyment, 1996).

$$Av = \lambda v \tag{2.29}$$

another form of equation (2.29) is

$$Av - \lambda v = 0$$

or

$$(A - \lambda I)v = 0 \tag{2.30}$$

From equation (2.30), which represents a system of homogeneous equations, it could be implied that the vector v is orthogonal to all row vectors of $(A - \lambda I)$. However, the equation (2.30) could not be solved normally by pre-multiplying both sides of the equation by $(A - \lambda I)^{-1}$due to the trivial solution of v = 0. Hence, a nontrivial solution can be obtained if $(A - \lambda I)$ is singular, i.e. its determinant must be zero. Therefore, setting the determinant of $(A - \lambda I)$ to zero, known as the characteristic equation, is the first step to solve the eigenvalues λ and eigenvectors v, thus:

$$|A - \lambda I| = 0 \tag{2.31}$$

If a $p - by - p$ matrix A is considered, it is going to have p eigenvalues, $\lambda_1, \lambda_2, \ldots, \lambda_p$. The results of determinant calculation will be a polynomial of order, p, thus

$$|A - \lambda I| = \lambda^p + c_1\lambda^{p-1} + c_2\lambda^{p-2} + \ldots + c_p = 0 \tag{2.32}$$

From equation (2.30), it can be rewritten as:

$$(A - \lambda_i I)v_i = 0 \; ; \quad (i = 1, \ldots, p) \tag{2.33}$$

The eigenvector v_i associated with eigenvalue λ_i can be obtained by solving this equation. However, Reyment and Joreskog (1996) explained that a unique solution cannot be obtained for an eigenvector. Therefore, eigenvectors are always normalized to be of unit length. Of course, the eigenvectors corresponding to different eigenvalues are orthogonal.

For multiple dimensional data, it is always placed as the element of the matrix. If the eigenvalues $\lambda_i (i = 1, \ldots, p)$ are the elements of a diagonal matrix D and the eigenvectors are the columns of the matrix, then the equation (2.29) becomes in the form as:

$$AV = VD \tag{2.34}$$

Let

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pp} \end{bmatrix}$$

$$V = \left[v_1 v_2 \cdots v_p \right]$$

and

$$D = \begin{vmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_p \end{vmatrix}$$

With the concept of orthogonal matrices, Reyment and Joreskog (1996) described how if the minor product moment of a matrix is an identity matrix, then the matrix is said to be orthogonal. Hence, it can be implied that the column vectors of the matrix V, which are the normalized vectors, are orthogonal to each other. Of course, the matrix V is orthonormal, so that

$$V'V = VV' = I; \quad V' = V^T$$

The post-multiplication and pre-multiplication of eq. (2.36) by V' gives

$$A = VDV^T \tag{2.35}$$

and

$$D = V^T AV \tag{2.36}$$

Another way to write the eq. (2.35) that is

$$A = \lambda_1 v_1 v_1^T + \lambda_2 v_2 v_2^T + \cdots + \lambda_p v_p v_p^T \tag{2.37}$$

which is the weighted sum of matrices $v_i v_i^T \ (i = 1, \ldots, p)$ of rank 1. Each term in eq. (2.37) is orthogonal to all other terms. The equation (2.35) or (2.37) is called the spectral decomposition of A.

2.3.3.3 Introduction to singular value decomposition (SVD) and how it works

If A is a square symmetric matrix, the eigenvectors corresponding to distinct eigenvalues can be normally solved by using the characteristic equation, eq. (2.31), to obtain the eigenvalues and plugging λ back into the system of equation, eq. (2.30) to obtain the eigenvectors. In a practical case, a real matrix A is rectangular. Of course, either AA^T or A^TA is calculated to become a square symmetric matrix to solve the eigenvalues and eigenvectors. When this information is placed as the elements of the matrix form as eq. (2.36), the form of $A = VDV^T$ always appears. Subsequently, instead of using either AA^T or A^TA to represent the original matrix A, both $AA^T$and A^TA should be used.

The SVD is the technique that consists of calculating the eigenvalues and eigenvectors of AA^T and A^TA. Gruber (2014) expressed the definition of the SVD as being the algorithm for decomposing any matrix into the product of an orthogonal matrix, a diagonal matrix with nonzero diagonal elements arrange from highest to lowest. The SVD theorem is as follows:

$$A = U\sum V^T \tag{2.38}$$

where U is an $n \times n$ orthogonal matrix.
Σ is a $n \times p$ diagonal matrix.
V^T is an $p \times p$ orthogonal matrix.

From eq. (2.38), when a $n - by - p$ matrix A is transformed to be the square symmetric matrix by taking AA^T, it is expressed in the form as follows:

$$AA^T = U\sum V^T V\sum^T U^T = U\sum\sum^T U^T = U\sum^2 U^T \tag{2.39}$$

V disappeared in eq. (2.39) because V^TV is the identity. On the other hand, taking A^TA it is expressed as follows:

$$A^TA = V\sum^T U^T U\sum V^T = V\sum\sum^T V^T = V\sum^2 V^T \tag{2.40}$$

In eq. (2.40), U is not present, as $U^T U$ is also the identity. When the equation (2.39) and (2.40) are compared, the $\sum^2$ term appears in both equations. From the eq. (2.38), therefore, it can be concluded that the eigenvectors of AA^T build up the columns of U, whereas that of $A^T A$ build up the columns of V. In the case of eigenvalues from AA^T or $A^T A$ in $\sum^2$, it is the same and is in the term of the square. As such, the singular values in $\sum$ are square roots of eigenvalues. It is the diagonal entries of $\sum$ the matrix and ranked from highest to lowest.

2.3.3.4 PLSR algorithms and its calculation

The idea of PLSR is to decompose both the predictor matrix X and response matrix Y like in principal analysis (PCA):

$$X = TP^T Y = UQ^T \tag{2.41}$$

and then perform regression between T and U (Ng, 2013). Therefore, the algorithm widely used for calculating the PLSR model is "Nonlinear iterative partial least squares (NIPALS)" (Wold et al., 2001). It works with the original matrix X and Y. Moreover, the PLSR algorithms can be considered according to a number of Y variables used in the calibration, i.e. PLS 1 algorithm for one variable and PLS 2 algorithm for two or more responses. This context will be referred to PLS 1 only.

The NIPALS algorithm on PLS 1 starts with centering matrix X and y. The process is done as follows:

$$\tilde{x}_{ij} = x_{ij} - \bar{x}_i, \qquad \tilde{y}_{1j} = y_{1j} - \bar{y}_1$$

where $\bar{x}_i$ is the average of column i

To do the algorithm with a single y variable (PLS 1), it is non-iterative. The calculations of PLSR model are as follows:

Let X is a $n \times p$ matrix and y is an $n \times 1$ vector.

$$X = \begin{bmatrix} \tilde{x}_{11} & \tilde{x}_{12} & \cdots & \tilde{x}_{1p} \\ \tilde{x}_{21} & \tilde{x}_{22} & \cdots & \tilde{x}_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ \tilde{x}_{n1} & \tilde{x}_{n2} & \cdots & \tilde{x}_{np} \end{bmatrix}, \qquad y = \begin{bmatrix} \tilde{y}_{11} \\ \tilde{y}_{21} \\ \vdots \\ \tilde{y}_{n1} \end{bmatrix}$$

Step 1: Calculate $X-weight$, w, which is the vector reflecting the covariance structure between the variable X and response y. Several strategies can be performed to calculate, for instance:

1) Take the SVD on $X^T y$ to get the first column vector in the matrix U, such that

$$X^T y = U \sum V^T \tag{2.42}$$

2) Calculate from the equation of

$$w_i = X^T y / \|X^T y\|; \qquad i = 1,2,\dots,r \tag{2.43}$$

Step 2: Calculate $X-scores$, t, or latent variables (LVs).

$$t_i = X w_i; \qquad i = 1,2,\dots,r \tag{2.44}$$

The $X-scores$, t, are normalized and can be written in matrix form as:

$$\underset{n\times 1}{\begin{bmatrix} t_i \end{bmatrix}} = \underset{n\times p}{\begin{bmatrix} X \end{bmatrix}} \underset{p\times 1}{\begin{bmatrix} w_i \end{bmatrix}}; \qquad i = 1,2,\dots,r \tag{2.45}$$

Each t is estimated as linear combinations of the variables x with the weight w, such that:

$$\begin{aligned} t_{1i} &= w_{1i}\tilde{x}_{11} + w_{2i}\tilde{x}_{12} + \cdots + w_{pi}\tilde{x}_{1p} \\ t_{2i} &= w_{1i}\tilde{x}_{21} + w_{2i}\tilde{x}_{22} + \cdots + w_{pi}\tilde{x}_{2p} \\ &\vdots \qquad\qquad \vdots \qquad\qquad \vdots \\ t_{ni} &= w_{1i}\tilde{x}_{n1} + w_{2i}\tilde{x}_{n2} + \cdots + w_{pi}\tilde{x}_{np} \end{aligned}$$

Step 3: Calculate the $X-loading$, p from the equation of

$$p = X^T t \tag{2.46}$$

Step 4: Calculate the $y-loading$, q from

$$q = y^T t \tag{2.47}$$

In the case of a single y variable, q is set to 1. When $X - scores$, t, and $X - loading$, p is the PLS estimation to the calibration spectra X, the residual in the spectra is computed by the next step. Similarly, the calculation of residual in y is also computed by next step.

Step 5: Calculate the residual of X and y.

$$E = X - t_i {p_i}^T \tag{2.48}$$

$$F = y - t_i {q_i}^T \tag{2.49}$$

The procedure is re-iterated from the step 1-5 by starting with the new residuals until either a number of $i(i = 1,2, \dots, r)$ are completed or X becomes a null matrix. To easily understand, every parameter is placed in the matrix form and rewritten to look as a multiple regression model, such that

$$y = TQ^T = XB \tag{2.50}$$

$$XB = TQ^T \tag{2.51}$$

$$XB = XWQ^T \tag{2.52}$$

However, Mevik and Wehrens (2007) showed that one complication was that columns of matrix Wcould not be compared directly, as it was derived from successively deflated matrices $\boldsymbol{E}$ and $\boldsymbol{F}$. Therefore, they suggested an alternative way to represent the weights, in such that all columns relate to the original X matrix, was given by

$$R = W(P^T W)^{-1} \tag{2.53}$$

and $T = XR$

$$XB = XRQ^T \tag{2.56}$$

The regression coefficient matrix can be determined as

$$B = RQ^T \tag{2.57}$$

or

$$\underset{p\times 1}{\left[B\right]} = \left[\quad \underset{p\times r}{R} \quad\right]\left[\quad \underset{r\times 1}{Q} \quad\right]$$

So, the predicted value matrix is calculated as

$$\hat{y} = XB = X_{pred}RQ^T \tag{2.58}$$

or

$$\underset{n\times 1}{\left[\hat{y}\right]} = \left[\quad \underset{n\times p}{X_{pred}} \quad\right]\underset{p\times r}{\left[R\right]}\underset{r\times 1}{[Q]}$$

2.4 References

Abdi H. 2003. “Partial Least Squares (PLS) Regression.” In M. Lewis-Beck, A. Bryman, & T. Futing (Eds.), **Encyclopedia of Social Sciences Research Methods:** Thousand Oaks.

Abdel-Rahman E.M., Ahmed F.B. and van den Berg M. 2010. “Estimation of sugarcane leaf nitrogen concentration using in situ spectroscopy.” **Journal of Applied Earth Observation and Geoinformation International**. 12, 52–57.

Ahmed A.E. and Alam-Eldin A.O.M. 2015. “An assessment of mechanical vs manual harvesting of the sugarcane in Sudan – The case of Sennar Sugar Factory.” **Journal of the Saudi Society of Agricultural Sciences**. 14(2), 160-166.

Andrade-Garda J. 2009. **Basic chemometric techniques in atomic spectroscopy**. Cambridge: RSC Publishing.

Benjamin C.E. 2002. “**Sugar cane yield monitoring system.**” USA: Louisiana State University.

Birth G.S. and Zachariah G.L. 1976. “Spectrophotometry of agricultural products, in Quality Detection in Foods Gaffney.” **American Society of Agricultural Engineering**, 1(76), 6-11.

Braunbeck O., Bauen A., Rosillo-Calle F. and Cortez, L. 1999. “Prospects for green cane harvesting and cane residue use in Brazil.” **Biomass and Bioenergy**, 17(6), 495-506.

BSES. 2001. The laboratory manual for Australian sugar mills. **Analytical methods and tables.** 2, Australia: BSES Limited.

Burns D.A. and Ciurczak E.W. 2001. “Principles of near-infrared spectroscopy.” In **Handbook of Near-Infrared Analysis**. New York: Marcel Dekker Inc. 7-18.

Cancado J.E., Saldiva P.H., Pereira L.A., Lara L.B., Artaxo P., Martinelli L.A., Arbex M.A., Zanobetti A. and Braga A.L.F. 2006. “The impact of sugar cane-burning emissions on the respiratory system of children and the elderly.” **Environmental Health Perspectives.** 114(5), 725-729.

Caryn B., Mailander M. and Price R. 2002. “**Sugar cane yield monitoring system.**” Baton Rouge, LA: Agricultural and Biological Engineering Louisiana State University.

Chia K.S., Rahim H.A. and Rahim R.A. 2012. “Prediction of soluble solids content of pineapple via non-invasive low cost visible and shortwave near infrared spectroscopy and artificial neural network.” **Biosystems Engineering**, 113, 158–165.

Cox G.J., Harris H., Pax R. and Dick R. 1996. “Monitoring cane yield by measuring mass flow rate through the harvester.” **Proceedings of the Australian Society of Sugar Cane Technology**, 152-157.

Cox G.J. 2002. “**A yield mapping system for sugar cane chopper harvesters**.” University of Southern Queensland.

Cuadra S.V., Costa M.H., Kucharik C.J., Darocha H.R., Tatsch J.D., Inman-Bamber G., Da Racha R.P., Leite C.C. and Cabral O.M.R. 2012. “A biophysical model of Sugarcane growth.” **GCB Bioenergy**, 4(1), 36-48.

DeFrantz T.F. 2019. “**CANE: a responsive environment dancework.**” https://sites.duke.edu/cane/sugar-cane-past/. Accessed May 04 2019.

FOTMA (2019). 4GZ-125 **"Whole Stalk Sugarcane Combine Harvester."** http://www.combine-harvesters.com/4GZ-125-Whole-Stalk-Sugarcane-Combine-Harvester-pd6765416.html.

Geladi P. and Dåbakk E. 1995. "An Overview of Chemometrics Applications in near Infrared Spectrometry." **Journal of Near Infrared Spectroscopy**, 3(3), 119-132, doi:10.1255/jnirs.63.

Godshall M.A. and Legendre B.L. 2003. "SUGAR | Sugarcane." In B. Caballero (Ed.), **Encyclopedia of Food Sciences and Nutrition (Second Edition)** (pp. 5645-5651). Oxford: Academic Press.

Gruber M.H.J. 2014. "**Matrix Algebra for Linear Models.**" USA: John Wiley & Sons.

Henry R. and Kole C. 2010. **"Basic information on the sugarcane plant."**

Hsieh C. and Lee Y. 2005. "Applied visible/near-infrared spectroscopy on detecting the sugar content and hardness of pearl guava." **Applied Engineering in Agriculture,** 21(6), 1039–1046.

Huang M. and Lu R. 2010. "Optimal wavelength selection for hyperspectral scattering prediction of apple firmness and soluble solids content." **Transaction of the ASABE,** 53(4), 1175–1182.

James G.L. 2004. "An introduction to sugarcane" In G.L. James (Ed.), **Sugarcane** (Second Edition ed., pp. 1-18). USA: Blackwell Science Ltd.

Jensen T., Baillie C., Bramley R., Bella L.D., Whiteing C. and Davis R. 2010. "Assessment of Sugarcane Yield Monitoring Technology for Precision Agriculture." **Proceedings of the Australian Society of Sugar Cane Technology**, 32, 410-423.

Kawano S., Fujiwara T. and Iwamoto M. 1993. "Nondestructive Determination of Sugar Content in Satsuma Mandarin using Near Infrared (NIR) Transmittance." **Journal of the Japanese Society for Horticultural Science**, 62(2), 465-470.

KhanAcademy. 2019. "**Light: Electromagnetic waves, the electromagnetic spectrum and photons.**"

Klute U. 2007. "Microwave measuring technology for the sugar industry." **International Sugar Journal** 109(1308), 1-6.

Kortum G. 1963. "Principles and thchniques of diffuse-reflectance spectroscopy." **Angewandke Chemie International Edition Spectroscopy,** 2(7), 333-341.

Ma S., Karkee M., Scharf P. and Zhang Q. 2014. "**Sugarcane harvester technology: A critical overview.**" (Vol. 30).

Madsen Ii L.R., White B.E. and Rein P. (2003). "**Evaluation of a near infrared spectrometer for the direct analysis of sugar cane.**" (Vol. 23).

Mailander M., Benjamin C., Price R. and Hall S. 2010. "Sugar cane yield monitoring system." **Applied Engineering in Agriculture**, 26 (6), 965–969.

McCarthy S. and Billingsley J. 2002. "A sensor for the sugar cane harvester topper." **Sensor Review**, 22(3), 242-246

Mevik B.H. and Wehrens R. 2007. "The pls Package: Principal Component and Partial Least Squares Regression in R." **Journal of Statistical Software**, 18(2).

Moore P.H., Paterson A.H. and Tew T. 2013. "**Sugarcane : Physiology, Biochemistry, and Functional Biology.**" Ames, Iowa :: Wiley Blackwell.

Naderi-Boldaji M., Ghasemi-Varnamkhasti M., Fazeliyan-Dehkordi M., Ahmad Mireei S. and Hosseinzadeh B. 2016. "A non-destructive sensing technique for sugarcane quality in terms of commercial cane sugar (CCS)." Paper presented at the **Proceedings of the 3rd Iranian International NDT Conference**, Olympic Hotel, Tehran, Iran, Feb 21-22, 2016.

Nawi N.M. 2014. "**Development of new measurement methods to determine sugarcane quality from stalk samples.**" University of Southern Queensland.

Nawi N.M., Chen G. and Jensen T. 2013a. "Visible and shortwave near infrared spectroscopy for predicting sugar content of sugarcane based on a cross-sectional scanning method." **Journal of Near Infrared Spectroscopy**, 21, 289–297.

Nawi N.M., Chen G. and Jensen T. 2014. "In-field measurement and sampling technologies for monitoring quality in the sugarcane industry: a review." **Precision Agriculture**, 15(6), 684–703.

Nawi N.M., Chen G., Jensen T. and Mehdizadeh S. 2013b. "Prediction and classification of sugar content of sugarcane based on skin scanning using visible and shortwave near infrared." **Biosystems Engineering**, 115, 154–161.

Nelson S.O. 1987. "Potential agricultural applications for RF and microwave energy." **Transactions of the ASAE**, 30(3), 818-831.

Netafim's Agriculture Department, 2019. **“Harvesting Management.”** http://sugarcanecrops.com/agronomic_practices/harvesting_management/. Accessed May 3 2019.

Ng K.S. 2013. **“A Simple Explanation of Partial Least Squares.”**

Nicolaï B.M., Beullens K., Bobelyn E., Peirs A., Saeys W., Theron K.I. and Lammertyn J. 2007. “Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review.” **Postharvest Biology and Technology**, 46(2), 99-118, doi:https://doi.org/10.1016/j.postharvbio.2007.06.024.

Núñez O. and Spaans E. 2008. “Evaluation of green-cane harvesting and crop management with a trash-blanket.” **Sugar Tech**, 10(1), 29-35, doi:10.1007/s12355-008-0005-1.

O'Shea M., Staunton S.P., Donald D. and Simpson J. 2011. “Developing laboratory near infra-red (NIR) instruments for the analysis of sugar factory products.” **Proceedings of the Australian Society of Sugar Cane Technology,** 33, 1-8.

Osborne B.G. 2006. “Near-infrared Spectroscopy in Food Analysis.” **Encyclopedia of Analytical Chemistry,** John Wiley & Sons, Ltd.

Osborne B.G., Fearn T., Hindle P.H. and Hindle P.T. 1993. “**Practical NIR Spectroscopy with Applications in Food and Beverage Analysis”** Longman Scientific & Technical.

Otto M. 2017. “**Chemometrics: Statistics and Computer Application in Analytical Chemistry”** (Third Edition ed.). Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA.

Pagnano N.B. and Magalhaes P.S. 2001. “Sugarcane Yield Measurement.” In Blackmore, S., & Grenier, G. (Eds.), (pp. 839-833): 3rd **European Conference on Precision Agriculture.**

Price R.R., Johnson R.M., Viator R.P., Larsen J. and Peters A. 2011. “Fiber Optic Yield Monitor for a Sugarcane Harvester.” **American Society of Agricultural and Biological Engineers,** 54(1), 31-39.

Price R.R., Larsen J. and Peters A. 2007. “Development of an Optical Yield Monitor for Sugarcane Harvesting.” **American Society of Agricultural and Biological Engineers.**

ProSciTech. 2014. **“Basic principles of refractometry: Measuring substances dissolved in water and certain oils.”** https://laboratoryresource.com/?navaction=getitem&id=174. Accessed May 08 2019.

Reddy N. and Yang Y. 2015. “Fibers from Sugarcane Bagasse.” In N. Reddy, & Y. Yang (Eds.), **Innovative Biofibers from Renewable Resources** (pp. 29-30). Berlin, Heidelberg: Springer Berlin Heidelberg.

Reich G. 2005. **“Near-infrared spectroscopy and imaging: Basic principles and pharmaceutical applications.”** 57, 1109–1143.

Reyment R.A. and Jvreskog K.G. 1996. **“Applied Factor Analysis in the Natural Sciences.”** UK: Cambridge University Press.

Rinnan A., van den Berg F. and Engelsen S.B. 2009. “Review of the most common pre-processing techniques for near-infrared spectra.” **Trends in Analytical Chemistry**, 28(10), doi:10.1016/j.trac.2009.07.007.

Sugarcane Botany: **“A Brief View”** (2016). The University of Florida, Institute of Food and Agricultural Sciences UF/IFAS http://edis.ifas.ufl.edu/sc034. Accessed April 23, 2019.

Schaffler K.J. and De Gaye M.T.D. 1997. “Rapid near infrared estimation of multicomponents in mixed juice and final molasses: the possibility of day-to-day control of raw sugar factories using NIR.” In **Proceedings of the International Society of Sugar Cane Technologists,** 71, 153-160.

Shah S. and Joshi M. 2010. “Modeling microwave drying kinetics of sugarcane bagasse.” **International Journal of Electronics Engineering,** 2(1), 159–163.

Siesler H.W., Ozaki Y., Kawata S. and Heise H.M. 2002. “Origin of near-infrared absorption bands” In **Near-Infrared Spectroscopy: Principles, Instruments, Applications.** (pp. 11–41). Weinheim Wiley-VCH Verlag GmbH.

Siesler H.W., Ozaki Y., Kawata S. and Heise H.M. 2008. **“Near-Infrared Spectroscopy: Principles, Instruments, Applications”** Weinheim Wiley-VCH Verlag GmbH.

Simpson R. and Naidoo Y. 2010. “Progress in improving laboratory efficiencies using near infrared spectroscopy (NIRS).” In **Proceedings of the International Society of Sugar Cane Technologists**, 27, 1-8.

Sopa C. 2010. “A Study of Sugarcane Leaf-Removal Machinery during Harvest.” **American Journal of Engineering and Applied Sciences,** *3,* doi:10.3844/ajeassp.2010.186.188.

Staunton S.P. and Wardrop K. 2006. “Development of an on-line bagasse analysis system using NIR spectroscopy.” In **Proceedings of the Australian Society of Sugar Cane Technologists,** 28, 446–453.

Sugiyama J. 1999. “Visualization of sugar content in the flesh of a melon by near-infrared imaging.” **Journal of Agriculture and Food Chemistry,** 47, 2715–2718.

Tobias R.D. “An Introduction to Partial Least Squares Regression.” In, 1995: **SUGI Proceedings.**

Tomuta I., Porfire A., Casian T. and Gavan A. 2018. “Multivariate Calibration for the Development of Vibrational Spectroscopic Methods.” In M. Stauffer (Ed.), “**Calibration and Validation of Analytical Methods - A Sampling of Current Approaches:”** IntechOpen.

Tumbo S.D., Wagner D.G. and Heinemann P.H. 2002. “On-the-go sensing of chlorophyll status in corn.” **Transactions of the ASAE,** 45, 1207–1215

Sugar:World Markets and Trade (2018). https://apps.fas.usda.gov/psdonline/circulars/sugar.pdf. Accessed 6 April, 2019.

Valderrama P., Braga J.W.B. and Poppi R.J. 2007. “Validation of multivariate calibration models in the determination of sugar cane quality parameters by near infrared spectroscopy.” **Journal of the Brazilian Chemical Society,** 18, 259-266.

Varmuza K. and Filzmoser P. 2016. **“Introduction to Multivariate Statistical Analysis in Chemometrics.”** CRC Press.

Investopedia 2019. **“The 5 Countries that Produce the Most Sugar.”** https://www.investopedia.com/articles/investing/101615/5-countries-produce-most-sugar.asp. Accessed 6 April, 2019.

Weekes D. 2004. “Harvest Management.” In G. James (Ed.), **Sugarcane** (Second Edition ed., pp. 160-180). USA: Blackwell Science Ltd.

Weeks D.P. 2017. “Chapter Four - Gene Editing in Polyploid Crops: Wheat, Camelina, Canola, Potato, Cotton, Peanut, Sugar Cane, and Citrus.” In D. P. Weeks, & B. Yang (Eds.), **Progress in Molecular Biology and Translational Science**, 149, 65-80, Academic Press.

Wold S., Sjostrom M. and Eriksson L. 2001. “PLS-regression: a basic tool of chemometrics.” **Chemometrics and Intelligent Laboratory Systems**, 58 109–130.

Workman J. 2015. **“An Introduction to Near Infrared Spectroscopy.”** https://www.spectroscopynow.com/details/education/sepspec1881education/an-introduction-to-near-infrared-spectroscopy.html. Accessed 6 April, 2019.

Workman J. and Weyer L. 2007. **“Practical Guide to Interpretive Near-Infrared Spectroscopy.”** CRC Press, Inc.

Yadav R.N.S., Sharma M.P., Kamthe S.D., Tajuddin A., Yadav S. and Tejra R.K. 2002. “Performance evaluation of sugarcane chopper harvester.” **Sugar Tech,** 4(3), 117-122.doi:10.1007/BF02942692.

Chapter 3

A study of the influence of sugarcane variety on sugar content prediction using shortwave near-infrared spectroscopy

This chapter describes a preliminary study regarding the influence of sugarcane variety on sugar content prediction using shortwave near-infrared spectroscopy. An overview of the scientific issues of this chapter is shown in Figure 3.1. Inside the chapter, two commercial sugarcane varieties (LK92-11 and Khonkaen 3) were used. Partial least squares (PLS) calibrations based on two spectral ranges (605-1070 nm and 800-900 nm) were firstly constructed and then compared to each other. The spectroscopic model from each variety of data was called either LK92-11 model or Khonkaen 3 model, while that of the pooling data was called global model. The optimum spectral range for the calibration was selected and applied for investigating the influence of sugarcane variety. Based on these results, the need for the application on in-field quality measurement is proposed.

* This chapter constituted the international conference article: Phetpan K, Udompetiakul V, Sirisomboon P. “A study of the influence of sugarcane variety on sugar content prediction using shortwave near-infrared spectroscopy.” **The 9th Thai Society of Agricultural Engineering International Conference: TSAE 2016**, 8-10 September 2016, pp.76-80.

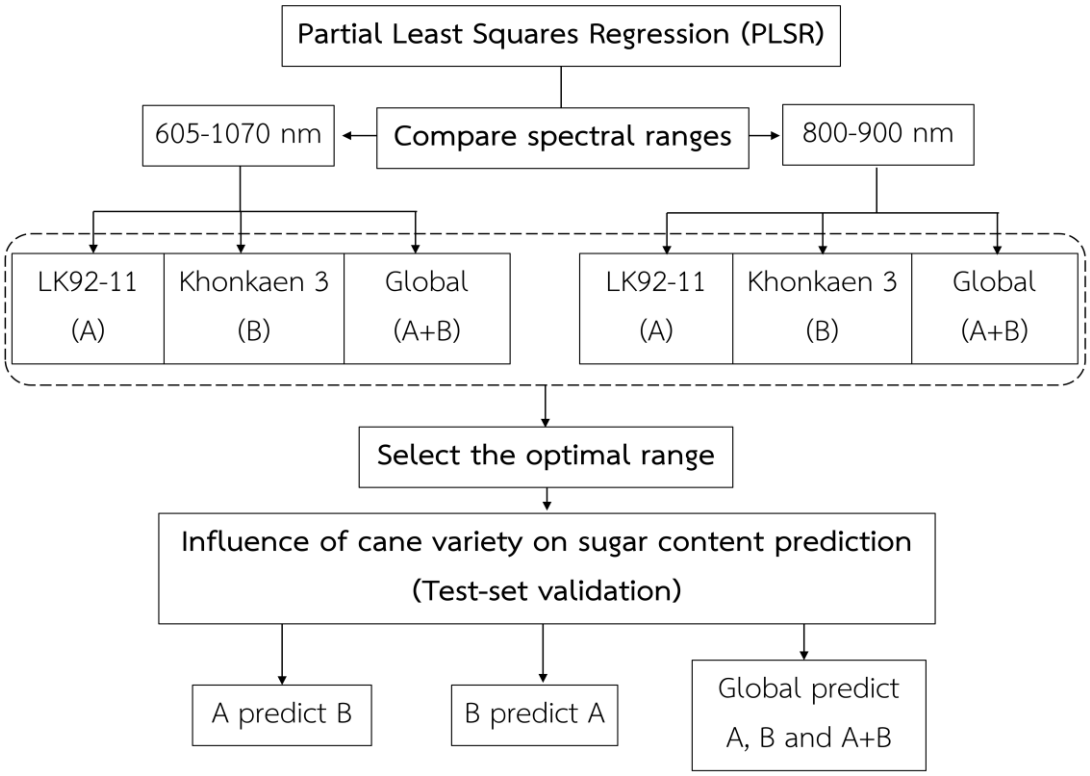


Figure 3.1 Overview of research procedures for studying the influence of sugarcane variety on sugar content prediction

3.1 Introduction

In attempting to improve sugarcane productivity, there is a growing interest in the adoption of precision agriculture (PA) technology (Wrigley and Moore, 2006; Nawi et al., 2013a). PA is a tool that leads to higher farm profits by providing site-specific variability data for crop input management (Wrigley and Moore, 2006). However, for the situation of current PA application in sugarcane industry, Bramley (2009) cited by Nawi et al. (2013b) mentioned that PA could be able to only monitor the yield but had no ability to measure its quality. Of course, not only the yield map is useful for sugarcane field management, but also the quality map. In addition, measurement of quality level leads to many more advantages such as optimizing the harvest schedules for maximum returns in the crushing season (Staunton et al., 2011; Nawi et al., 2014) and helping the industry to assess the performance of varieties and crop classes (Lawes and Lawn, 2005; Nawi et al., 2014).

In Thailand, sugar content is used as a qualitative indicator of sugarcane. With the need for in-field measurement, alternative technologies which were reviewed by

Nawi et al. (2014) are electronic refractometer, microwave, and portable spectrometer. One of the most promising technologies is the portable spectrometer known as spectroscopic methods.

Spectroscopic methods are well known as a fast, low-cost and non-destructive technique and have been applied in the sugarcane industry to predict sugar content in several forms of sample such as clarified juice (Berding et al., 1991), stalk cross-section and stalk skin (Nawi et al., 2013a; Nawi et al., 2013b) with good prediction performance. However, their experiments were still based on laboratory circumstance. Hence, Nawi et al. (2014) also reviewed regarding in-field measurement for monitoring quality in the sugarcane industry and concluded that using the portable visible and shortwave near-infrared (Vis/SW-NIR) spectrometer and scanning the samples in form of the stalk is the best approach for field uses. Moreover, they recommended the need to be carried out for in-field measurement that was the influence of sugarcane variety. Therefore, the aim of this study was to investigate the influence of sugarcane variety on sugar content prediction of sugarcane using shortwave near-infrared (SW-NIR) spectroscopy.

3.2 Materials and Methods

3.2.1 Sample preparation

The sugarcane stalks of 2 commercial varieties, LK 92-11 and Khonkaen 3, were collected from Nakhon Ratchasima Sugarcane Experiment Station, Nakhon Ratchasima, Thailand at maturity stage of 11 months after planting. They were grown under the same crop management scheme. At 11-month state, the physical appearance of the sugarcane stem is similar to that of the harvesting stage at around 12 months (Cane and Sugar Promotion Center Region 4, 2016) with a better range of sweetness for the experiment. Fifty-two internode samples were collected for each variety.

3.2.2 Spectral acquisition

A portable near-infrared instrument (FQA-NIRGUN, Fantec, Japan) with the interaction mode from 600 to 1100 nm at 2 nm intervals was used to acquire NIR spectra of the samples. The influence of ambient light on the acquired spectra was

eliminated by using a protecting cover while conducting the NIR scanning. Each internode was scanned by the NIR instrument at its middle position, then was cut into an individual piece. The experiment was performed at room temperature (25±1 °C).

3.2.3 Soluble solids (°Brix) measurement

After the spectral acquisition, each internode was squeezed to extract its juice for °Brix measurement. The juice from each one was kept in an aluminum container, stirred and immediately poured onto a digital refractometer (Pocket PAL-1 Refractometer, ATAGO, Japan). Typically, standard sweetness quality of sugarcane is defined by CCS value (Commercial Cane Sugar) which is derived from °Brix, Pol, and fiber content (Nawi et al., 2013a; Nawi et al., 2013b). Since, further work of this research has been planned to propose quality spatial-variability maps of the field with simple indicators, only °Brix value is adequate to be the qualitative indicator for mapping.

3.2.4 Development of calibration and validation models

Prior to calibration, the spectra data used for model development were pre-treated using a savitsky-golay first and second derivative, multiplicative scatter correction (MSC) and standard normal variate (SNV) for the optimal performance of spectra. MSC was found to be the best one and used for calibration. The Unscrambler X 10.3 (Camo, Norway) was used for the spectral pre-processing and for model development.

Two ranges of spectra were used for calibration. One is a range of 605-1070 nm which is the full range of the spectrometer. The other is that of 800-900 nm, which was indicated by Travers et al. (2014) that it appeared more relevant to soluble solids content prediction. So, partial least squares (PLS) calibration for sugar content based on these two spectral ranges were compared. The spectroscopic model from each variety data was called either LK92-11 model or Khonkaen 3 model, while that of the pooling data was called global model. In this context, full cross-validation was used to evaluate the quality of each model, as indicated by the coefficient of determination (R^2) and root mean square error of cross-validation (RMSECV). From the comparison, the spectral range of 800-900 nm gave better

quality for all models. So, this range was then used to construct the PLS model for studying the Influence of sugarcane variety on sugar content prediction.

External validation was used as a key to study the influence of sugarcane variety. The original data of each variety was randomly split into two sets; two-thirds of data was used for the calibration set and one-thirds of data for the test set. LK92-11 and Khonkaen 3 models were re-developed and then validated using the test set data. The global model was re-constructed using both calibration sets and then validated using three test-set data obtained from LK92-11, Khonkaen 3 and global (LK92-11 and Khonkaen 3) were combined. Of course, the influence of sugarcane variety on sugar content prediction using SW-NIR models was considered by R^2 value and root mean square error of prediction (RMSEP).

3.3 Results and Discussion

The summary of statistical information for calibration and test set data of internode samples is shown in Table 3.1. Figure 3.2 shows the averaged spectrum of the two sugarcane varieties. There were two ranges of spectrum used for calibration. Its raw absorbance spectra having full (605-1070 nm) and selected (800-900 nm) ranges are shown in Figure 3.2 (a) and (c), respectively. The obvious difference of spectrum for different sugarcane variety could be seen in the range of 605 to 730 nm (Figure 3.2 (a)). This difference was probably caused by the existence of a chlorophyll absorption peak ~680 nm (Travers et al., 2014). MSC method was used to pre-treating of the raw spectra. Shapes of full and selected ranges pre-treated are shown in Figure 3.2 (b) and (d), respectively.

A comparison of spectroscopic models constructed from 2 ranges is shown in Table 3.2. The optimum results were obtained from the range of 800-900 nm for all models giving the R^2 values of 0.787 to 0.865 and RMSECV values of 0.57 to 0.74 °Brix. In addition, a number of latent variables (LVs) used for PLS calibrations with the range of 800-900 nm were distinctly less than that of 605-1070 nm. This was done to ensure that the selected spectral range is relevant to the sugar content prediction according to Travers et al. (2014).

Table 3.1 Summary of statistical information of internode samples

Data set	models	Cross-validation					Test set				
		N	Max (°Brix)	Mean (°Brix)	Min (°Brix)	SD	N	Max (°Brix)	Mean (°Brix)	Min (°Brix)	SD
Calibration	LK92-11	52	26.20	22.53	19.90	1.66	-	-	-	-	-
	Khonkaen 3	52	25.90	22.21	19.90	1.30	-	-	-	-	-
	Global	104	26.20	22.37	19.90	1.50	-	-	-	-	-
Test set	LK92-11	35	26.20	22.68	19.90	1.82	17	25.50	22.17	19.90	1.63
	Khonkaen 3	35	25.90	22.23	19.90	1.14	17	25.10	22.20	19.90	1.14
	Global	70	26.20	22.45	19.90	1.52	34	25.50	22.19	19.90	1.45

Note: SD = Standard deviation

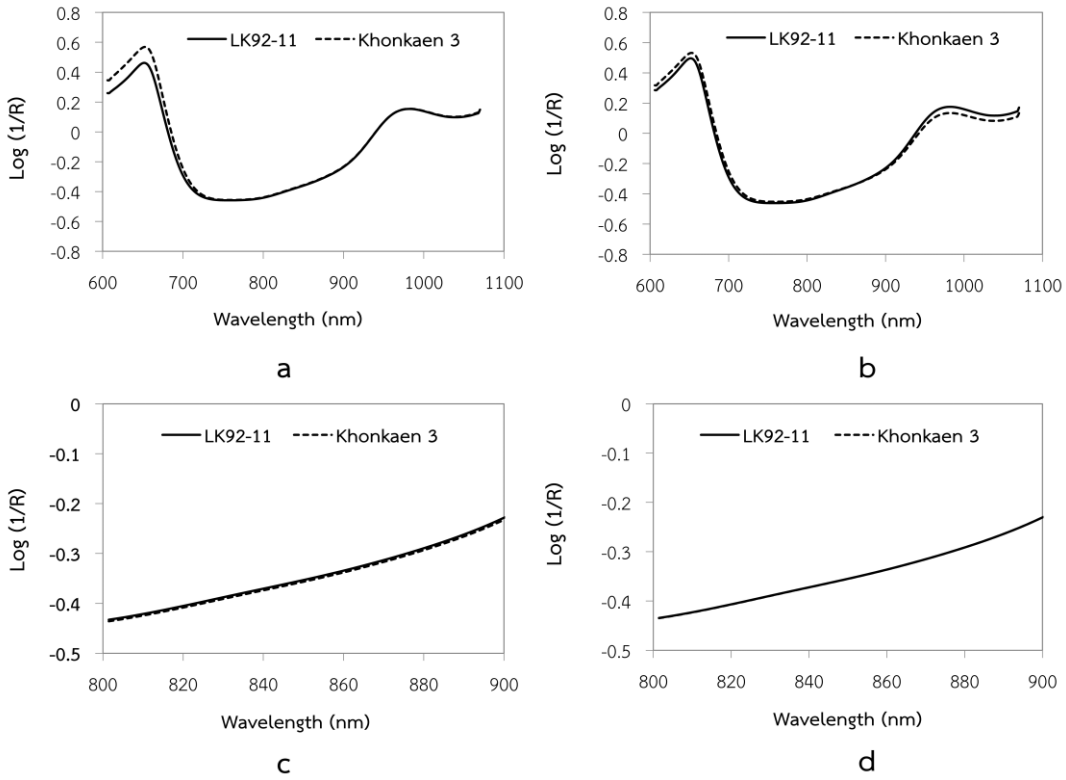


Figure 3.2 Averaged spectrum of two sugarcane varieties for (a) raw absorbance spectra of 605-1070 nm; (b) raw absorbance spectra of 605-1070 nm pre-treated with MSC; (c) raw absorbance spectra of 800-900 nm and (d) raw absorbance spectra of 800-900 nm pre-treated with MSC

The external validation technique was used to study the influence of sugarcane variety on sugar content prediction. The result in Table 3.3 indicated that the LK92-11 model and Khonkaen 3 prediction models could not be used in the

cross variety. However, the global model gave satisfied results with R^2 values of 0.756, 0.732 and 0.741, and RMSEP values of 0.62, 0.82 and 0.72 °Brix for predicting of LK92-11, Khonkaen 3 and Global data, respectively. The fact that reflectance spectra of fruit contain information on both the absorbance and scattering properties of the tissue (Nicolai et al., 2008). This means that not only chemical components are related to the reflectance but also small suspended particles as well as the microstructure of the fruit. Based on the results in this part, the scattering properties of the tissue of different sugarcane varieties is probably the reason affecting the sugar content prediction.

Table 3.2 The results of spectroscopic models constructed from two ranges of spectra

Model	Range (nm)	Pretreatment	N	LVs	R^2_c	RMSEC (°Brix)	R^2_{cv}	RMSECV (°Brix)
LK92-11				8	0.806	0.73	0.683	0.95
Khonkaen 3	605-1070	MSC	52	5	0.846	0.51	0.781	0.62
Global				10	0.811	0.65	0.692	0.83
LK92-11				3	0.865	0.61	0.823	0.71
Khonkaen 3	800-900	MSC	52	3	0.852	0.5	0.811	0.57
Global				3	0.787	0.69	0.76	0.74

Table 3.3 The results of external validation for studying the influence of sugarcane variety on sugar content prediction

Model	Range (nm)	Pretreatment	Calibration set				Test set		
			N	LVs	R^2_c	RMSECV (°Brix)	N	R^2_p	RMSEP (°Brix)
A predict B	800-900	MSC	35	3	0.884	0.77	17	0.478	1.14
B predict A				2	0.764	0.59		0.43	0.94
Global predict A			70	3	0.793	0.77	34	0.756	0.62
Global predict B								0.732	0.82
Global predict A+B								0.741	0.72

Note: A = LK92-11, B = Khonkaen 3, Global = Pooling of LK92-11 and Khonkaen 3

3.4 Conclusions

This study has demonstrated that the specific variety spectroscopic models could not predict sugar content of the other variety. This confirmed that variety of sugarcane affects sugar content prediction using shortwave near-infrared spectroscopy. However, a global model gave satisfying results for sugar content prediction and needs for the application on the in-field quality measurement.

3.5 References

Berding N., Brotherton G.A. and Skinner J.C. 1991. "Near infrared reflectance spectroscopy for analysis of sugarcane from clonal evaluation trials: II. Expressed juice." **Crop Sciences.** 31, 1024-1028.

Bramley R.G.V. 2009. "Lessons from nearly 20 years of precision agriculture research, development, and adoption as a guide to its appropriate application." **Crop & Pasture Science** 60, 197-217.

Cane and Sugar Promotion Center Region 4. 2016. "**Classification of cane variety.**" Available at: http://oldweb.ocsb.go.th/udon/Udon4/p10_smai.htm. Accessed on 9 July 2016.

Lawes R.A. and Lawn R.J. 2005. "Applications of industry information in sugarcane production systems." **Field Crops Research.** 92, 353-363.

Nawi N.M., Chen G. and Jensen T. 2013a. "Visible and shortwave near-infrared spectroscopy for predicting sugar content of sugarcane based on a cross-section scanning method." **Journal of Near Infrared Spectroscopy.** 21: 289-297.

Nawi N.M., Chen G. and Jensen T. 2014. "In-field measurement and sampling technologies for monitoring quality in the sugarcane industry: a review." **Precision Agriculture.** 15, 684-703.

Nawi N.M., Chen G., Jensen T. and Mehdizadeh S.A. 2013b. "Prediction and classification of sugar content of sugarcane based on skin scanning using visible and shortwave near infrared." **Biosystems Engineering.** 115, 154-161.

Nicolai B.M., Verlinden B.E., Desmet M., Saevels S., Saeys W., Theron K., Cubeddu R., Pifferi A. and Torricelli A. 2008. "Time-resolved and continuous wave NIR

reflectance spectroscopy to predict soluble solids content and firmness of pear." **Postharvest Biology and Technology.** 47, 68–74.

Staunton S., Donald D. and Pope G. 2011. "Estimating sugarcane composition using ternary growth relationships." **Proceedings of the Australian Society of Sugar Cane Technology**. 33, 1-8.

Travers S., Bertelsen M.G., Petersen K.K. and Kucheryavskiy S.V. 2014. "Predicting pear (cv. Clara Frijs) dry matter and soluble solids content with near infrared spectroscopy." **LWT-Food Science and Technology.** 59, 1107-1113.

Wrigley T. and Moore S. 2006. "Public Environment Report." **Canegrowers**. Brisbane, Australia.

Chapter 4

An online visible and near-infrared spectroscopic technique for the real-time evaluation of the soluble solids content of sugarcane billets on an elevator conveyor

This chapter proposes a prototype online detection system based on the visible and near-infrared spectroscopic (Vis/SW-NIR) technique for the real-time evaluation of the soluble solids content (SSC) of sugarcane billets on an elevator conveyor. Interestingly, movement characteristics and conveying speed of the cane billets while being scanned by the spectroscopic sensor was simulated under real practice circumstances. The prototype developed in the thesis works based on direct scanning method (DSM), which was proposed by Nawi et al. (2014). The DSM is simpler than the assisted scanning method (ASM) that needs some other assisted equipment. Although, the DSM was navigated to be conducted at the location B on the elevator conveyor of the sugarcane harvester, the prototype of this research was focused on the location A where it has no cover over the conveyor of the real harvester and is more suitable to deal with.

Calibration models were also constructed and assessed whether it could be used to predict the SSC of sugarcane billets being transferred on the conveyor. The overview of modeling process is shown below (Figure 4.1).

* This chapter constituted the publication article: Phetpan K, Udompetiakul V, Sirisomboon P. "An online visible and near-infrared spectroscopic technique for the real-time evaluation of the soluble solids content of sugarcane billets on an elevator conveyor." **Computers and Electronics in Agriculture**, 154 (2018), pp. 460-466.

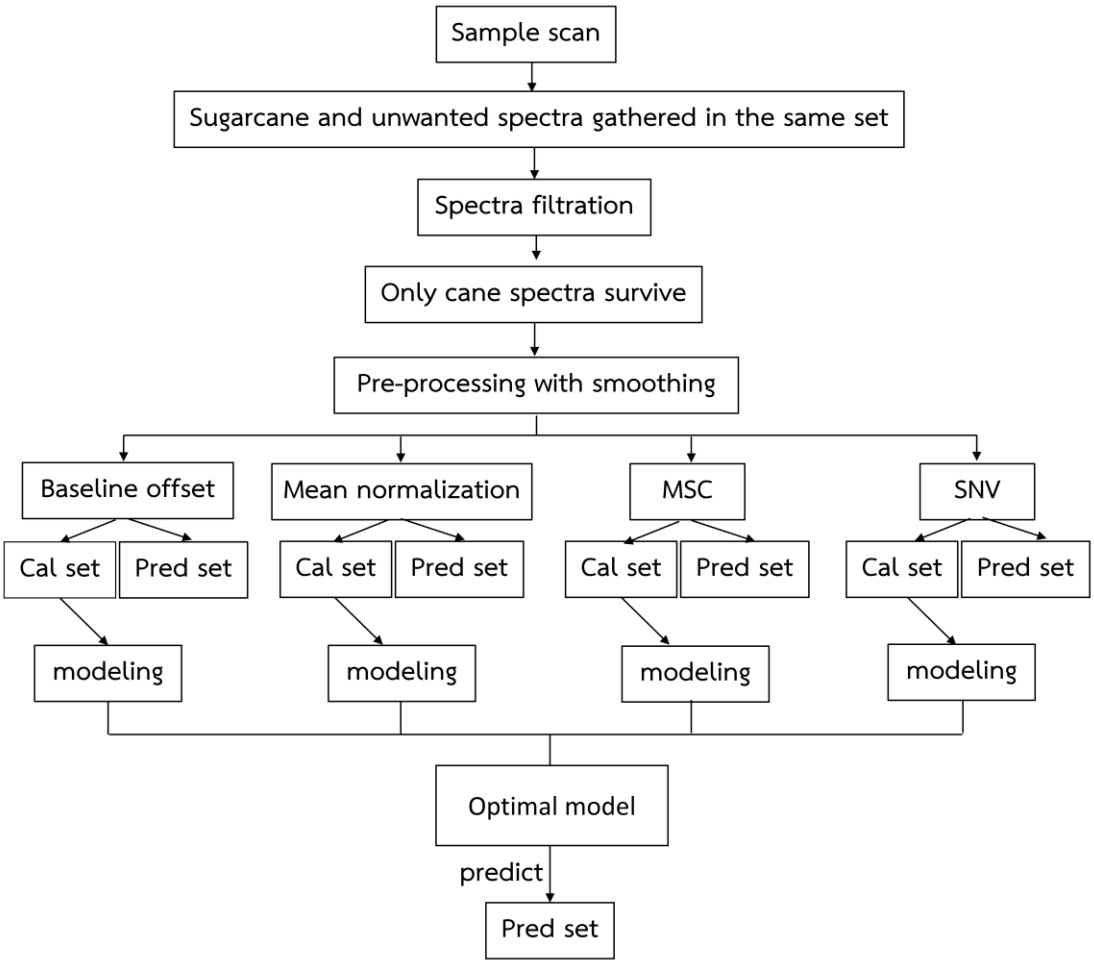


Figure 4.1 Overview of modeling process for predicting the sweetness (SSC) of sugarcane billets moving on the elevator

4.1 Introduction

Sugarcane is one of the principal raw materials used in sugar production worldwide. It is an economically important crops, with a global export value of around 27 billion USD in 2017 (International Trade Centre, 2017). Brazil is the world's largest sugarcane producer and sugar exporter, followed by Thailand and France. Brazil has been a pioneer in the use of sugarcane to produce alternative fuels in the form of ethanol. It produces more than 20bn liters of ethanol a year from sugar and is expected to reach 50bn liters a year by 2020 (Cookson, 2012).

Demand for cane sugar derived ethanol fuels have been consistently growing in other parts of the world. This has prompted efforts to achieve improvements in agricultural productivity in terms of crop yield and quality without necessarily

increasing the amount of farmland being employed. The primary industrial sugarcane quality index is defined by the commercial cane sugar (CCS) parameter which is the percent of recoverable sucrose from fresh cane. This measure of sugar content is used for pricing and trading between growers and sugar mills (Dixon and Johnson, 1988). Variation in the sugarcane CCS parameter is dependent on numerous factors including; cultivar, age, moisture, nutrient and temperature (Naderi-Boldaji et al., 2016). Lawes et al. (2000) studied the spatial variability in CCS, finding that it was up to 9 units in an individual field. They noted that site-specific management of the field input and activities could help minimize the CCS variation by maximizing the yield and/or quality of the production in the less productive areas.

Precision agriculture (PA) are all-encompassing tools or technologies to improve the management of agricultural production processes through the recognition of limiting production factors, soil fertility as well as the variation in yield and quality across a given field. It is one of the means for crop management strategies (Bramley, 2009). Nevertheless, its application requires the generation of a spatial variability map to perform the necessary site-specific management of industrial farms. The key issue is developing an efficient, cost effective ways of determining PA in such situations.

Bramley (2009) reviewed PA applications in the sugarcane industry and found that to monitor the sugarcane yield and to construct variability maps was not necessarily an arduous process. However, the monitoring of crop quality during harvesting, and the production of the subsequent map was still going on the research and still based on the laboratory scale. This prompted the development of several field based techniques for rapid sugar cane quality measurement including; electronic refractometers (Mccarthy and Billingsley, 2002), microwaves (Nelson, 1987; Klute, 2007; Shah and Joshi, 2010) and spectroscopic techniques (Nawi et al., 2013a, 2013b). The refractometer based method required a mechanism for squeezing cane juice prior to measurement, while the non-destructive microwave and spectroscopic methods proved to be useful in the laboratory but were unproved in field use. This prompted Nawi et al. (2014) to propose that a spectroscopic method had real potential for field use if installed on the elevator conveyor used during sugarcane harvesting. This would allow for the direct measurement of the soluble solids

content (SSC, °Brix) of sugarcane, without the requiring the direct analysis of the sugarcane juice.

The spectroscopic technique of Nawi et al. (2013a) utilizes an area of the electromagnetic spectrum that covers the wavelength range from around 350-2500 nm. Visible and shortwave near-infrared spectroscopy (Vis/SW-NIR), 350-1100 nm, appear to be promising bands as these are typically ascribed to the third and fourth overtones of O-H and C-H stretching modes (Walsh et al., 2000) that are present in sugar molecules. In addition, the method has been used to determine the value of SSC for the sugarcane stalk (Nawi et al., 2013b) and it has been claimed that it has the potential to be developed into a sensor for monitoring sugarcane quality during harvesting (Nawi et al., 2014).

In an effort to develop an online sugarcane quality measurement, a prototype system for SSC prediction was developed and tested. The goals of this work are twofold (1) to propose an online detection system of the sugarcane billets based on the Vis/SW-NIR technique which could be installed on the elevator conveyor of the sugarcane harvester; (2) to evaluate the performance of the system for online SSC measurement of the sugarcane billets on the elevator conveyor.

4.2 Materials and methods

4.2.1 Sample preparation

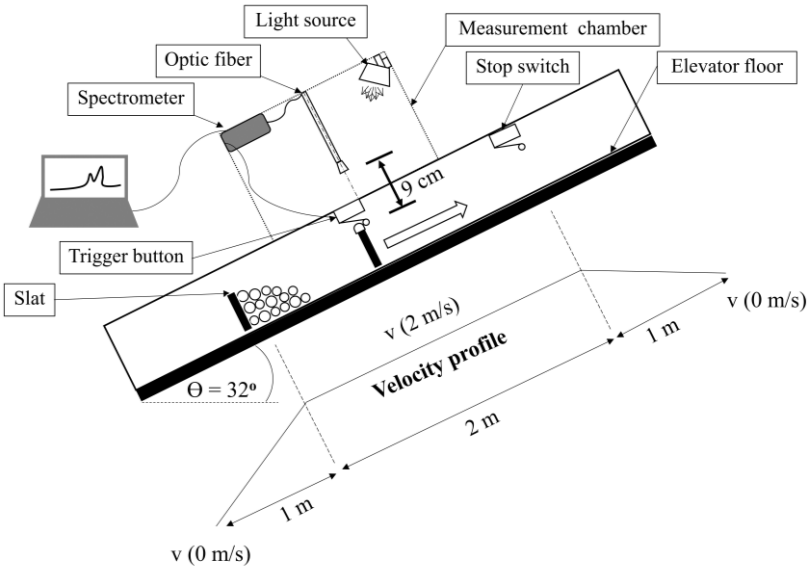
Fifty clumps of sugarcane were collected from fields in Suphan Buri, Thailand, in February 2017. Each clump is composed of approximately 20 stalks which were cut into billets with approximate length of 20 centimeters. The cane variety was Khonkaen 3, one of the most common commercial varieties in Thailand. The sugarcane were collected at approximately 11 and 12 months after planting in order to cover the different maturity stages in commercial harvesting period of this sugarcane variety.

4.2.2 On-line detection system and spectral acquisition

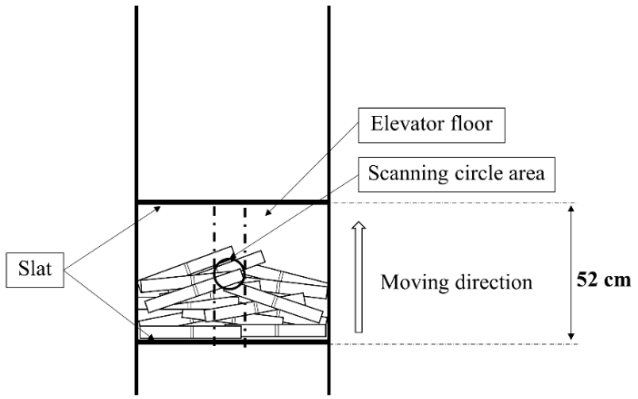
An online measurement system (Figure 4.2) had been developed in this study for monitoring the value of SSC in sugarcane billets. The system consisted of two main parts – a cane billet elevator and a spectral acquisition system. The elevator

used in the study was built with the same dimensions as those of a common sugarcane harvester (John Deere 3520). It was set at an angle of 32° from the ground. The elevator was divided into three distinct regions for controlling the velocity speed. The starting and ending regions, each being one meter in length, were used for ramping up and down the speed to and from the stationary state. The middle region, which was two meters in length, was set to deliver the sugar cane at a constant conveying speed. Two slats were mounted to the conveyor chain. The upper one was used to trigger the acquisition switch and the lower one was used for moving the billets along the elevator. The sample movement was set according to the velocity profile as shown in Figure 4.2a. The acquisition system was used to collect the spectral data of the billet group as it moved along the elevator. A chamber was built around the acquisition device to allow for the installation of a constant light source and to protect interference from environment light. Four halogen lamps (Aluline Halogen 12V/50W R111, Royal Philips, Holland) were mounted at a distance of 60 cm and at 45° away from the elevator floor. A Vis/SW-NIR spectrometer (AvaSpec-2048-USB2, Avantes BV, Netherlands) was installed, operating in the spectral range of 350-1100 nm with the spectral resolution of 2.4 nm. Spectral measurement was done in reflectance mode as the modeling with reflectance spectra performed better than that with absorbance spectra (Nawi et al., 2013b). The amount of light reflected from the samples was collected using a 25° field-of-view (FOV) of the optic fiber that was fitted to the spectrometer and was fixed at 9 cm over the top edge of the slat on the elevator. At that distance, the scanning area covered a circle of approximately 4 cm diameter (Figure 4.2b). The spectrometer was set to scan the spectra using an external trigger.

An appropriate integration time must be specified to achieve the optimum system sensitivity. Based on our chamber environment, the integration time was set to 14 ms, yielding approximately 90% full-scale Analog-to-Digital Converter (ADC) of the reference material reflectance. Each triggering of the device led to the collection of 19 spectral scans (one set of scans), with the integration time set to cover the spectral measurement between 2 slats (a distance of 52 cm) with a conveying speed of 2 m/s (typical speed of the elevator of the sugarcane harvester and the speed used in this study).



a)



b)

Figure 4.2 Scheme of an online measurement system, a) side view and b) top view

In order to obtain four representative optical spectral sets of each clump, two replications with two repeated measurements were obtained. One set of scans was performed for each repeated measurement. After finishing the first measurement, the slat and chain were retracted to the starting position before proceeding with the second measurement. This was performed without reshuffling the cane billets. To date, two sets of spectra (19 spectra each) were obtained for first replication. The cane billets were then collected for ºBrix (SSC) determination (detailed in the next section). With the first replication, two spectral sets and a corresponding SSC value were acquired. After the first replication, the remaining billets were removed from

the elevator and reshuffled before reloading to the elevator for the second replication (using the same procedure). Therefore, four sets of spectra (2×19 spectra for the 1st replication + 2×19 spectra for the 2nd replication) and two SSC values (for the 1st and 2nd replications) were obtained for each billet clump. Two hundred spectral sets, along with 100 SSC values were obtained in total.

4.2.3 Soluble solids (°Brix) determination

Reference SSC values for the sugarcane billets those lied on the scanning path (Figure 4.2 b) were obtained by squeezing the billet in a small hydraulic press. The juice from each billet was stirred, screened with filter paper, and immediately poured onto a digital refractometer (Pocket PAL-1, ATAGO, Japan) to enable °Brix measurement. The SSC values obtained from each replication were averaged.

4.2.4 Spectral filtration

Spectral bands outside the range of 450-900 nm were first removed due to excessive noise. This is because low signal-to-noise ratio at both ends of the detector. However, the absorption bands relating sugar can still be found at 450-900 nm. All 200 spectral sets obtained from the previously described experiments were performed using this process. Each of the spectral sets contained spectra from 3 different sources - cane, slats and floor – as showed in Figure 4.3. Undesirable spectra from the slats and floor must be removed before further analysis. Slat spectra could be easily filtered out if the difference in percent reflection value at the wavelength of 550 and 620 nm was higher than 0.6. For floor spectra, the reflection value at 800 nm was used as the marking point to eliminate this group of spectra from the set. The eliminated spectra were those with the reflection at the marking point less than 20%. In addition, spectra from partial scans of sugarcane with weak signal were also eliminated.

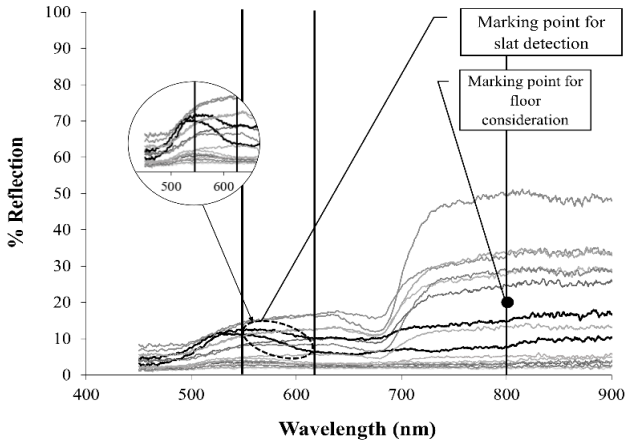


Figure 4.3 An example of NIR spectra obtained from a group of the sugarcane billets on the elevator

4.2.5 NIR modeling

After the filtration step above, 150 sets of spectra remained for modeling against their SSC values. Note that 50 sets of spectra were eliminated. Partial least squares (PLS) regression is regarded as one of the most popular linear statistical methods for modeling the linear relationships between the variable matrix X or the spectra and the variable matrix Y or the properties of interest (Jie et al., 2014). It was applied in order to obtain the linear model correlating the online spectra of sugarcane billets on the elevator with their SSC. In this study, the software used for multivariate analysis (Unscrambler X 10.3, Camo, Norway) was used in both spectral pretreatment and model development.

To obtain the model with more reliable performance, this necessitated the formation of two independent datasets, which were obtained by randomly splitting the 100 sets of spectra with their corresponding SSC from 150 sets mentioned above to be calibration set. The other set (50 sets) was used for external testing set. However, from observation, signal noises and offset were spectral characteristics that have a negative impact on model development. Spectral pretreatment is a key step to improve the model accuracy. Several techniques including smoothing, multiplicative scatter correction (MSC), standard normal variate (SNV) scaling, mean normalization and baseline offset were applied in the calibration and the external testing sets separately to overcome these characteristics. In this paper, moving average (MA) smoothing with segment size of 21 points was first applied to minimize

spectral noises and then the others were individually applied to diminish the offset effect in spectra. After the pretreatments, each spectral set in both datasets were averaged. Thus, 100 samples and 50 samples with matching averaged spectra and their SSC were obtained and ready for PLS modeling and the external testing, respectively.

To initially seek an optimum model based on the data obtained from different pretreatments, the PLS models were developed and validated using the leave-one-out cross-validation. The optimum one was selected based on high coefficient of determination (R^2) and low root mean square error of cross-validation (RMSECV). Then, this selected model was used to test its true performance by predicting the external testing set. This procedure is a more realistic measure of model performance as the samples used in the analysis were not used to develop the initial model. The model performance is assessed statistically using the coefficient of determination (R^2) of prediction, root mean square error of prediction (RMSEP) and residual predictive deviation (the standard deviation to standard error of prediction, RPD).

4.3 Results and discussion

4.3.1 Spectral description

Four different spectral pre-processing methods were considered in this study including moving average (MA) + baseline offset, MA + mean normalization, MA + MSC and MA + SNV. Figure 4.4 shows the spectra after the filtration process, while Figure 4.5 shows the spectra in the calibration dataset after the pre-processing operations. The normalized descriptors were used for modeling against the average SSC values. The effect of noise in the original spectra was minimized by utilizing the smoothing technique (MA). The baseline offset, one of mathematical techniques well known in the correction of offset existing in group of spectra, was applied in this study to diminish some effects that occur during the spectral detection. There is still a scattering effect in the spectra, which is tolerable if eliminated by the other pretreatments, particularly the MSC and SNV operations.

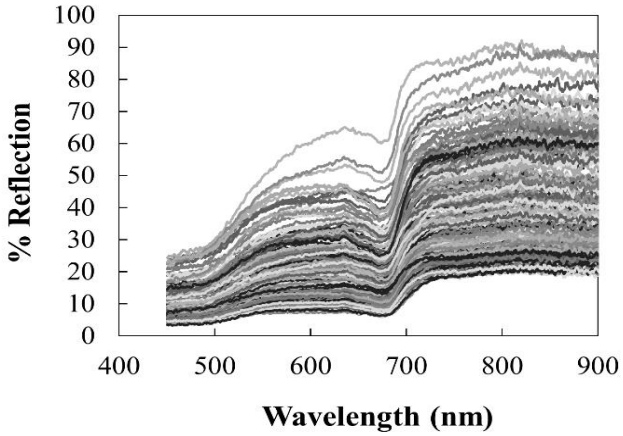


Figure 4.4 The sugarcane spectra remained from the filtration of all sample groups

The absorption region at around 670-680 nm, which could be seen clearly in all pretreated spectra, appears to be related to the chlorophyll content at the peel (Guo et al., 2003) of cane billets. The prominent peak does not appear in the region around 700-900 nm, instead, this region is the location of the bands typically related to the third or fourth overtone C-H and O-H stretching of both sugar and water groups (Osborne et al., 1993; Golic, et al., 2003; Guo et al., 2003; Ecarnot et al., 2013).

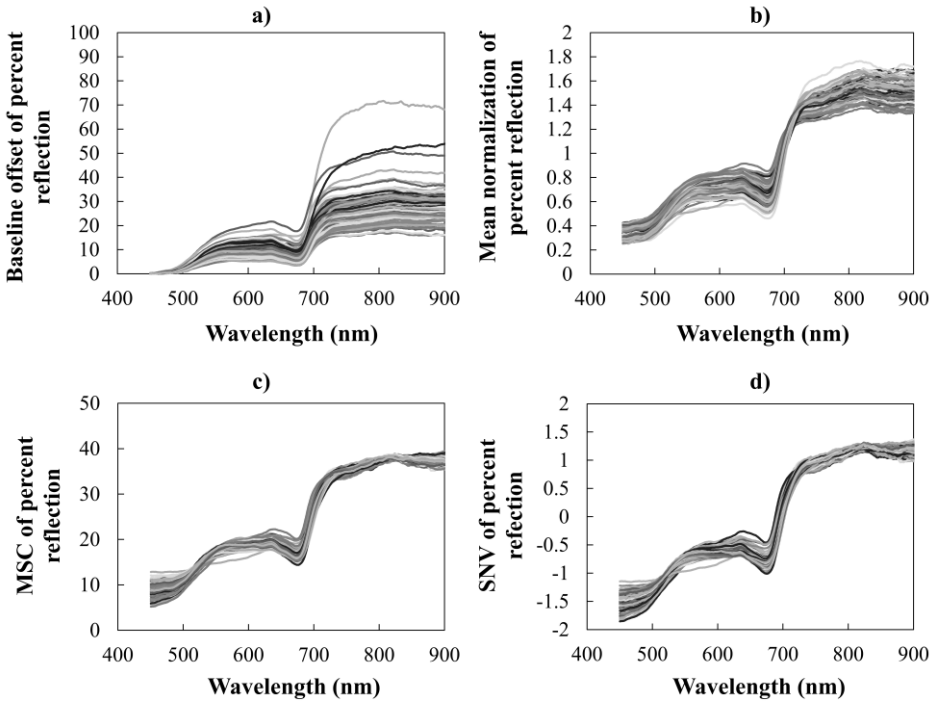


Figure 4.5 Average NIR spectra pretreated by different pretreatments, a) MA + baseline offset, b) MA + mean normalization, c) MA + MSC and d) MA + SNV

4.3.2 Near-infrared spectroscopy models for SSC prediction

Figure 4.6 shows the distribution of the SSC values used for PLS modeling and external testing. A summary of their statistical characteristics for the two sets are shown in Table 4.1. The 100 average spectra after pretreatments were used for modeling against the average SSC values and evaluated with full cross-validation to initially seek an optimum model. The regression and validation results for SSC prediction with different pretreatments are shown in Table 4.2 and corresponding regression coefficient plots are displayed in Fig 4.7. The models constructed from the spectra pretreated by the smoothing (MA) and combined with SNV, MSC and normalization show good performances. They provide R^2 values approximately 0.7-0.8 and RMSECV values of around 0.3 ºBrix. Among the best models, the optimum one was that obtained using MA + SNV, displaying an R^2 and RMSECV of 0.807 and 0.33 ºBrix, respectively. On the other hand, using the spectra obtained from the MA + Baseline offset operation provides the lowest R^2 and highest RMSECV values of 0.656 and 0.45 ºBrix, respectively. For the regression coefficient plots, their spectral patterns look similar. The two dominant peaks at 755 and 890 nm stand out.

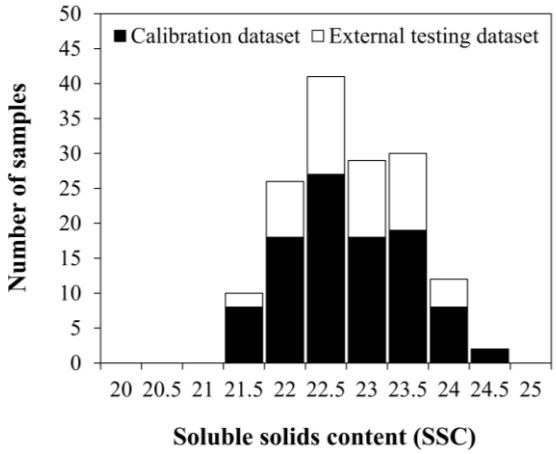


Figure 4.6 Frequency histograms of the soluble solids content values used for PLS model development (100 samples) and external testing (50 samples)

Table 4.1 Statistical SSC values of sugarcane billets used in developing and testing the PLS model

Dataset	N	Max (°Brix)	Min (°Brix)	Mean (°Brix)	SD
Calibration set	100	24.5	21.2	22.6	0.76
External testing set	50	23.9	21.4	22.6	0.66

N is the number of samples. Max is maximum. Min is minimum. SD is standard deviation.

Table 4.2 Regression and validation results for SSC prediction with different pretreatments

Pre-Processing	Calibration			Validation	
	LVs	R^2	RMSEC (°Brix)	R^2	RMSECV (°Brix)
MA + Baseline offset	4	0.695	0.42	0.656	0.45
MA + Mean Normalization	4	0.821	0.32	0.767	0.37
MA + MSC	4	0.844	0.30	0.805	0.34
MA + SNV	4	0.846	0.30	0.807	0.33

Note: LVs is Latent variables, MA is moving average method, MSC is Multiplicative scatter correction, SNV is Standard normal variate.

To confirm the independent prediction performance of the PLS model from MA + SNV operation, the external validation was performed. This selected model presents the performance for predicting a series of 50 samples not used during model generation by explaining the 78.5% of the variation existing in this dataset. RMSEP and RPD values of 0.30 °Brix and 2.16 were obtained, respectively. The corresponding scatter plots are displayed in Figure 4.8.

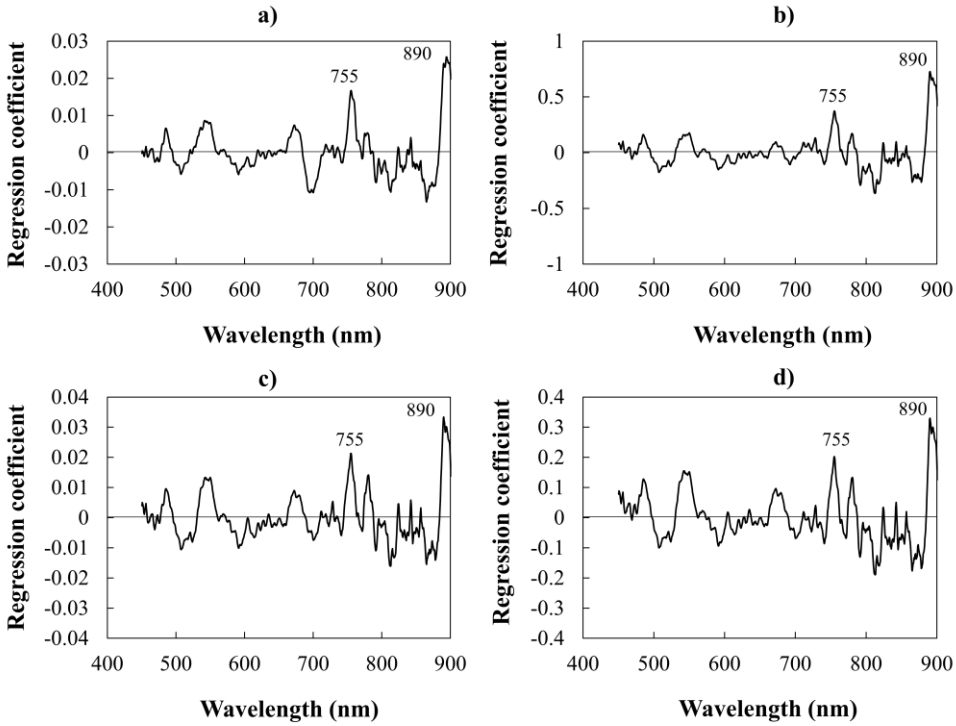


Figure 4.7 Regression coefficient plots of the models constructed from the different pretreated spectra, a) MA + Baseline offset, b) MA + Mean normalization, c) MA + MSC and d) MA + SNV

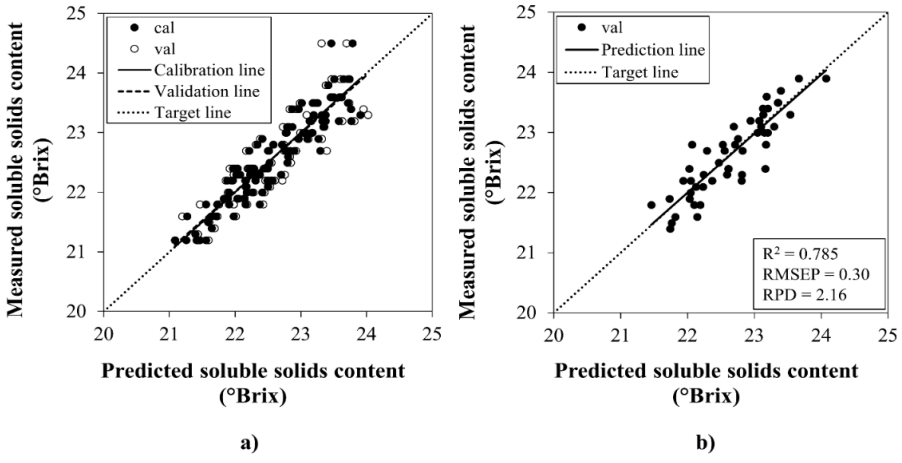


Figure 4.8 PLS model with external validation constructed from the spectra pretreated with MA combined with SNV, comparison of SSC predicted by PLS model and measured by the standard reference for a) the calibration set and b) the external testing set

Based on the results of the spectral pre-processing methods applied in this study, we found that they improve the linear relationship obtained between the spectral signals and the SSC values which were obtained from the sugarcane billets moving on the elevator. Good predictability was obtained using three pre-processing techniques after spectral smoothing including mean normalization, MSC and SNV. They are not different because these techniques have the same concept for solving the scattering effect existing in the spectra. Among the three spectral treatment methods, the MSC and SNV techniques provided the best results (R^2 and RMSECV). The SNV technique is more suitable for practical application, especially for our proposed system, due to there is no need for a reference spectrum. Self-correction is achieved by using only the standard deviation and mean of the scan in question.

To build robust PLS models, it is important to have a large sample size with good sample variability. Low variability in the SSC values used for modeling in this study was caused by collecting only one sugarcane variety with a narrow range of maturity stages. Although expanding the variability could be done by starting the sampling before 11 months after cane planting, it is not necessarily useful for real world application since it is not desirable to harvest such immature sugarcane. Given the emphasis on the primary evaluation, the proposed system for online SSC measurement of the sugarcane, it is adequate for this study. However, to obtain more robust models samples of several cane cultivars with different ranges of maturity (ºBrix) could be added in the future application.

With the spectral range of 450-900 nm used for the PLS modeling, the coefficient plots show that the models contained two sugar related peaks at 755 nm (the 4th overtone of C-H stretching of sugar at 762 nm (Osborne et al., 1993) or the 3rd overtone of O-H stretching of sucrose in water at 740 nm (Golic et al., 2003)) and 890 nm (the 3rd overtone of C-H stretching of sucrose in water at 910 nm (Golic et al., 2003)). Based on those, the absorption bands relating sugar are mainly weighted when predicting a particular SSC value of sugarcane billets moving on the elevator.

Typically, a few latent variables (LVs) are required to describe most of the data variance, however, the first LV accounts for the greatest amount of variance (Nawi et al., 2013b). Additionally, it is desirable to use only a limited number of LVs

in the model to avoid the inclusion of signal noise (Xiaobo et al., 2007). In this study only 4 LVs were used to explain the data variance in the SSC values.

This laboratory scale analysis is an important first step in the development of an on-the-go sensing system for the assessment of cane yield and quality during harvesting. The results presented here show the potential of online Vis/ SW-NIR spectroscopic techniques for the determination of soluble solids content in sugarcane billets moving on an elevator. From the findings present here, it should be possible to adapt this technique as an all-encompassing tool for PA already had the ability in yield assessment. This combination would allow the production of spatial variability maps describing the yield and quality within sugarcane farmlands. This variability map could help farmers reach their yields and quality responses, customize crop input and maximize farm profits. This could also allow the establishment of a fairer payment system for growers and allow mills to optimize their production processes.

4.4 Conclusions

In this work a lab-scale prototype of an online detection system has been designed and developed for the real-time SSC assessment of the sugarcane billets on an elevator conveyor. The system detects the spectra of the cane billets which are conveyed at the speed of 2 m/s (typical speed of a sugarcane harvester elevator) using an integration time of 14 ms. PLS modeling was used to correlate the obtained spectra with the SSC values. The results showed that the system is certainly feasible for the online SSC measurement of the sugarcane billets on the elevator, with an R^2, RMSEP, and RPD values of 78.5% , 0.30 ºBrix and 2.16 for the prediction set, respectively. Nevertheless, it is acknowledged that modeling with a dataset consisting of a greater number of sugarcane cultivars is necessary for a production of on-the-go SSC sensing system. This additional validation would help to improve the robustness of method for online SSC measurements in real world sugarcane fields.

This on-the-go sensing system would benefit agriculturists in that it would minimize yield and quality variations across sugarcane farmland. A further side effect would be the establishment of a fairer payment system for the growers reflecting the quality of their product and optimized production processes within mills.

4.5 References

Bramley R. 2009. “Lessons from nearly 20 years of precision agricultural research, development, and adoption as a guide to its appropriate application.” **Crop and Pasture Science.** 60, 197-217.

Cookson C. 2012. “A tank of sugar: how Brazil runs on biofuel.” **FT Magazine**, United Kingdom.

Dixon P.B. and Johnson D.T. 1988. “Pricing of Queensland sugarcane: appraisal of the present formula and a suggestion for reform.” **Review of Marketing and Agricultural Economics**. 56, 27-35.

Ecarnot M., Compan F. and Roumet P. 2013. “Assessing leaf nitrogen content and leaf mass per unit area of wheat in the field throughout plant cycle with a portable spectrometer.” **Field Crops Research.** 140, 44-50.

Golic M., Walsh K. and Lawson P. 2003. “Short-wavelength near infrared spectra of sucrose, glucose, and fructose with respect to sugar concentration and temperature.” **Applied Spectroscopy.** 57, 139-145.

Guo Z., Huang W., Peng Y., Chen Q. and Ouyang Q. 2003. “Color compensation and comparison of shortwave near infrared and long wave near infrared spectroscopy for determination of soluble solids content of Fuji apple.” **Postharvest Biology and Technology**. 27, 197-211.

International Trade Centre. “List of exporters for the selected product in 2018 – Product : 1701 Cane or beet sugar and chemically pure sucrose, in solid form” http://www.trademap.org/Country_SelProduct.aspx?nvpm=1|||||1701|||4|1|1|2|1||2|1|1. Accessed on 4 October 2017.

Jie D., Xie L., Rao X. and Ying Y. 2014. “Using visible and near infrared diffuse transmittance technique to predict soluble solids content of watermelon in an on-line detection system.” **Postharvest Biology and Technology.** 90, 1-6.

Klute U. 2007. “Microwave measuring technology for the sugar industry.” **International Sugar Journal**. 109, 1-6.

Lawes R.A., Wegener M.K., Basford K.E. and Lawn R.J. 2000. “Commercial cane sugar trends in the Tully sugar district.” **Australian Journal of Experimental Agriculture.** 40, 969-973.

Mccarthy S. and Billingsley J. 2002. “A sensor for the sugar cane harvester topper.” **Sensor Review**. 22, 242-246.

Naderi-Boldaji M., Ghasemi-Varnamkhasti M., Fazeliyan-Dehkordi M., Mireei S.A. and Hosseinzadeh B. 2016. “A non-destructive sensing technique for sugarcane quality in term of commercial cane sugar (CCS).” **Proceedings of the third Iranian International NDT Conference**. Tehran, Iran.

Nawi N.M., Chen G. and Jensen T. 2013a. “Visible and shortwave near infrared spectroscopy for predicting sugar content of sugarcane based on a cross-sectional scanning method.” **Journal of Near Infrared Spectroscopy.** 21, 289-297.

Nawi N.M., Chen G. and Jensen T. 2014. “In-field measurement and sampling technologies for monitoring quality in the sugarcane industry: a review.” **Precision Agriculture.** 15, 684-703.

Nawi N.M., Chen G., Jensen T. And Mehdizadeh S. 2013b. “Prediction and classification of sugar content of sugarcane based on skin scanning using visible and shortwave near infrared.” **Biosystems Engineering.** 115, 154-161.

Nelson S.O. 1987. “Potential agricultural applications for RF and microwave energy.” **Transactions of the American Society of Agricultural Engineers.** 30, 818-831.

Osborne B., Fearn T. and Hindle P. 1993. **Practical NIR spectroscopy with applications in food and beverage analysis.** UK: Longman Scientific & Technical.

Shah S. and Joshi M. 2010. “Modeling microwave drying kinetics of sugarcane bagasse.” **International Journal of Electronics Engineering**. 2, 159-163.

Sorol N., Arancibia E., Bortolato S.A. and Olivieri, A.C. 2010. “Visible/near infrared-partial least-squares analysis of Brix in sugar cane juice A test field for variable selection methods.” **Chemometrics and Intelligent Laboratory Systems**. 102, 100-109.

Walsh K.B., Guthrie J.A. and Burney J. 2000. “Application of commercially available, low cost, miniaturised NIR spectrometers to the assessment of the sugar content of intact fruit.” **Australian Journal of Plant Physiology.** 27, 1175-1186.

Xiaobo Z., Jiewen Z., Xingyi H. and Yanxiao L. 2007. “Use of FTNIR NIR spectrometry in non-invasive measurements of soluble solid contents SSC of “Fuji” apple based on different PLS models.” **Chemometrics and Intelligent Laboratory Systems**. 87, 43-51.

Chapter 5

Development of the partial least-squares model to determine the soluble solids content of sugarcane billets on an elevator conveyor

Developing a spectroscopic sensing system for bio-material applications always involves multivariate analysis requiring strong validation for further field use. As the Chapter 4 proposed the online spectroscopic sensing system and showed the possibility in measuring the sweetness (SSC) of sugarcane billets being transferred on the elevator. So, developing the multivariate model suitable for real use is very important. Partial least squares (PLS) regression was used for constructing the model. The exploration of an appropriate spectral range for modeling and the assessment of the influence of different levels of cane billets delivery on the sugar content predictive ability are the important issues presented in this chapter. Besides, the predictions of the samples, collected from the same and different growing seasons as the samples for the modeling, were also performed. Overview of PLS modeling is shown in Figure 5.1.

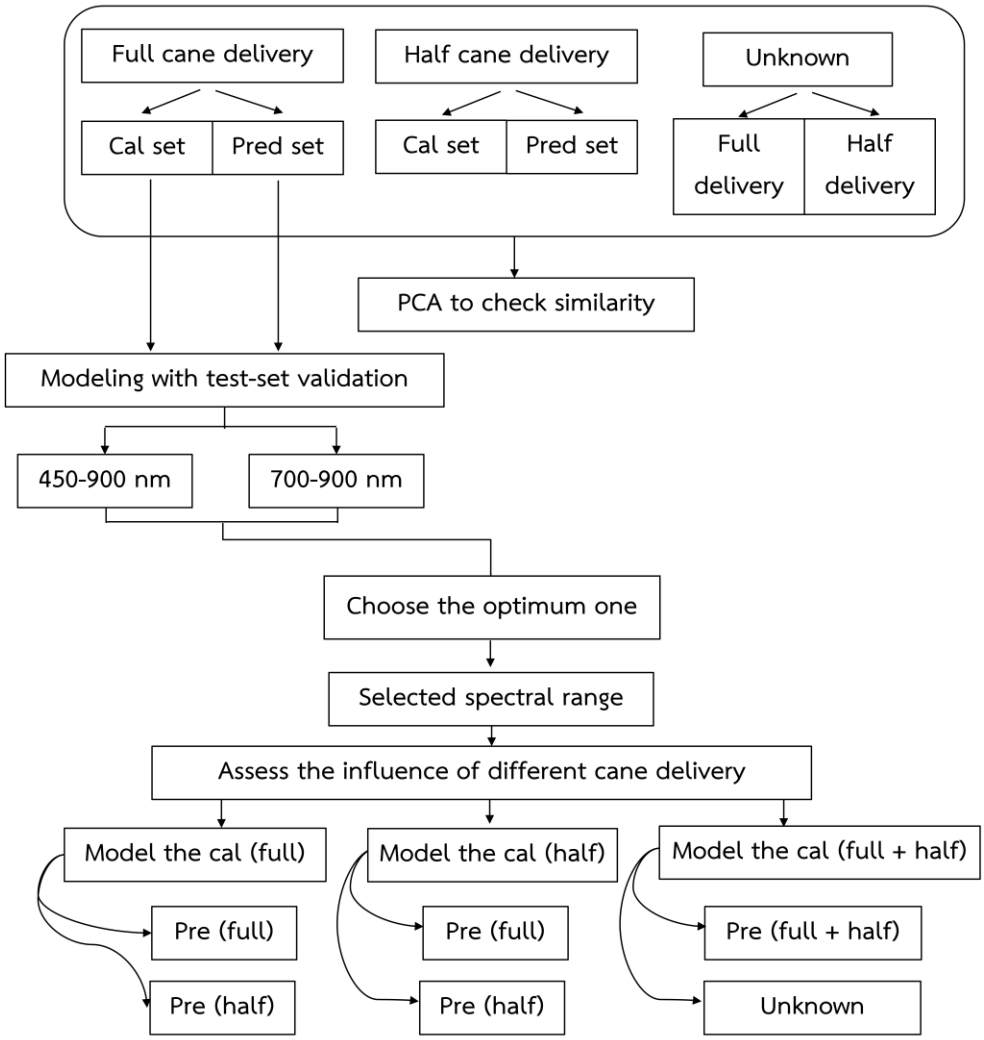


Figure 5.1 Overview of development of PLS model for measuring the soluble solids content of sugarcane billets on an elevator conveyor

5.1 Introduction

A chief aim in commercial crop production is to gain the highest yield and quality in the least productive areas. One of the many barriers to this aim is the variation in the yield and quality within these areas, especially in sugarcane production, in which the spatial variation of commercial cane sugar (CCS), i.e., the primary industrial sugarcane quality index, can reach up to 6.5 units within a single 8.8 ha sugarcane block (Kingston and Hyde, 1995), whereas yield variation can reach 30 – 45% in terms of the coefficient of variation (Bramley and Quabba, 2001). These variations can be caused by many factors, including weather, soil quality, plant density, age, and nutrients. Surely, cane growers want to manage their farms based

on uniformity by increasing their yield and quality. Therefore, the best option to increase yield and quality involves the site-specific management of the field input and activities, thus helping to minimise these variations. However, it is not easy to access and overcome field variations without exact site-specific information (Nawi 2014).

Precision agriculture (PA) is a technology used to respond and to recognise the spatial variation over a field to improve the management of agricultural production processes. The application of PA in the sugarcane industry is not an arduous process, requiring the monitoring of sugarcane to construct maps that show variability in yield. The monitoring of the cane quality still faces challenges (Nawi et al. 2014). In the future, if the information of both spatial variations across the field can be accessed, farm management practices will take action against this information, and a fair payment system to the growers will be implemented (Nawi et al. 2013a). In addition, the information could be beneficial to the sugarcane industry in dealing with individual small-holding farming and cooperative farming on a large scale.

Hence, several techniques have been studied as an alternative approach for the development of a cane quality measuring system in the field such as electronic refractometer (Mccarthy and Billingsley, 2002), microwave (Klute 2007; Nelson 1987; Shah and Joshi 2010) and spectroscopic techniques (Nawi et al. 2013a; Nawi et al. 2013b; Phetpan et al. 2018). From these techniques, the microwave and spectroscopic approaches have been proposed, as they have more potential in field applications. Nawi et al. (2013a) conducted research using the spectroscopic technique in a spectral region from 400-1000 nm for sugar content measurements in cane billets based on surface scanning. A non-contact sample spectral measurement was performed in their study. Based on the study, they claimed that this technique looked promising because it showed good results with a coefficient of determination (R^2) and root means square error of predictions (RMSEP) of 0.91 and 0.721 °Brix, respectively.

The spectroscopic technique involves the utilisation of the electromagnetic spectrum covering the wavelength range from approximately 350-2500 nm (Nawi et al. 2013b). The first half of this range, 350-1100 nm, is defined as an area of visible

and shortwave near-infrared (Vis/SW-NIR), ascribed to the third and fourth overtones of O-H and C-H stretching modes (Walsh et al. 2000). It has been claimed that it was suitable to develop as a sensor for monitoring sugarcane quality during harvesting (Nawi et al. 2014). However, some researchers who conducted the spectroscopic technique in the prediction of soluble solids content (SSC) mentioned that the presence of a chlorophyll absorption peak at ~ 680 nm, related to the peel of the fruits (Guo et al. 2003), can compromise model accuracy (Travers et al. 2014). Therefore, the use of the spectroscopic technique in the cane quality monitoring system during harvesting in terms of SSC may be better if the developed model contains only useful information. This use will lead to faster processing of the system due to fewer numbers of variables.

According to Nawi et al. (2014)'s review, it could be concluded that the optimum location to install the SSC measuring system was on the elevator. It was an open section in which the cane billets could be scanned directly with approximately 90% of trash, already blown out by the primary extractor. In addition, the system could be placed at any point along the elevator. Phetpan et al. (2018) has developed a laboratory scale prototype of the online detection system for real-time SSC assessment of the sugarcane billets on an elevator conveyor. Based on this prototype, the system detected the spectra of sugarcane billets which are conveyed with the typical speed of a harvester elevator and showed that the online SSC prediction of sugarcane being transferred on the elevator could certainly be feasible. However, with the spatial variation of yield across the field, it directly and differently affected the levels of sugarcane billets in the conveyor. This leads to differences in pathlength which is the key factor affecting the S/N ratio (Inagaki et al. 2017) as well as offset and scattering behaviours in spectra and did not cover the developed system. Hence, the influence of this factor on SSC prediction needs to be studied.

The objectives of this paper were to enhance the prediction of the system developed by Phetpan et al. (2018) in two main issues; 1) to assess whether it was better for predicting the SSC of sugarcane billets moving on the elevator if the model was developed without the visible range (450-700 nm) and 2) to determine whether it was possible to develop a calibration model for measuring SSC in cane billets by neglecting the influence of different cane levels on the elevator. In

addition, this paper also presented the developed model's validation with the sugarcane as collected from a different growing season.

5.2 Material and methods

5.2.1 Sugarcane samples

Sugarcane samples from the Khonkaen 3 cultivar were randomly collected from commercial fields in Song-Phi-Nong District, Suphan Buri, Thailand. They were collected at approximately 11 and 12 months after planting to cover the commercial harvesting stage for this variety.

Fifty clumps of sugarcane collected in February 2017 were used for constructing and validating the models in this study. The cane stalks in each clump were chopped into billets with an approximate length of 20 centimetres, which is the same size as that chopped by the cane harvester. Therefore, the 50 clumps were chopped and grouped into 200 cane billet groups.

To present the model's ability in unknown sugarcane prediction, thirty clumps of the sugarcane were collected from the fields covering approximately 30 square kilometers in the same district in February 2018. The sugarcane of each clump was chopped with the same procedure above to obtain a cane billet group. Thus, 30 groups were obtained and used for the unknown prediction.

5.2.2 On-line spectroscopic data collection

This work was performed at the Near Infrared Spectroscopy Research Center for Agricultural Product and Food, Bangkok, Thailand, where a simulated elevator conveyor in a sugarcane harvester (John Deere 3520) was built as shown by the scheme in Figure 5.2. On the conveyor, a chamber was built for the installation of spectroscopic devices and for the protection of interference from environmental light. Inside the chamber, four halogen lamps (Aluline Halogen, 12 V/50 W, R111, Royal Philips, Holland) were used as the light source, oriented at 45° to the elevator floor. Reflected light was received by a fibre optics probe (FC-IR1000, Avantes BV, Netherlands) with a field-of-view (FOV) of 25°, mounted perpendicular to the elevator floor. A Vis/NIR spectrometer (AvaSpec-2048-USB2, Avantes BV, Netherlands) fitted to the probe was used to acquire reflectance spectra in the wavelength range

between 350 and 1100 nm, at a resolution of 2.4 nm. Spectra were acquired with an integration time of 14 ms, and the optimum sensitivity based on our chamber environment and yielding was approximately 90% at full-scale measured by the Analog-to-Digital Converter (ADC) of the reference material reflectance. Each on-line spectroscopic response covers the compartment between 2 slats (a distance of 0.52 m). As a result of this integration time, 19 spectra (a spectral set), arising from transferring cane billets on the conveyor with a speed of 2 m/s, were obtained for each time scanning.

5.2.2.1 Data collection for modeling and model validation

From interviewing the operator and observing the harvester, the presentation of cane billets being transferred on the conveyor to a truck or a haul-out vehicle did not provide full delivery all the time. Hence, the influence of cane levels on the conveyor needs to be studied.

To collect the data for the study of this influence on the prediction of soluble solids content of sugarcane by using the spectroscopic technique, two conditions were designed and conducted to simulate the cane billets delivery on the conveyor. In this collection, 200 cane billet groups mentioned above were used. In the first condition, 100 groups were used. Each group was fully loaded on the conveyor compartment and was transported with a constant speed at 2 m/s, simulated for the full delivery (Figure 5.3a), to detect spectral data with two repetitions (2 scans). For each scan, the slat transferred the cane billets along the conveyor and stopped at the ending part of the conveyor. After finishing each scan, it was retracted to the starting position before proceeding to the next scan. Each delivery of the canes throughout this step was done without reshuffling. Two on-line spectroscopic scanning spectra (2 × 19 spectra) were acquired per cane billets group. The SSC (ºBrix) of the scanned cane billets was then determined (detailed in the next section). Therefore, two hundred spectral scanning spectra (200 × 19 spectra) along with 100 SSC values were obtained in the first condition. For the second condition, the other 100 groups were used. Each group was half loaded on the conveyor compartment and was also transferred with a constant speed at 2 m s^{-1}, simulated for the half delivery (Figure 5.3b). Data acquisition for the second condition also

followed the same procedure as the first condition. In total, four hundred spectral sets (Full and half delivery) and 200 SSC values were obtained.

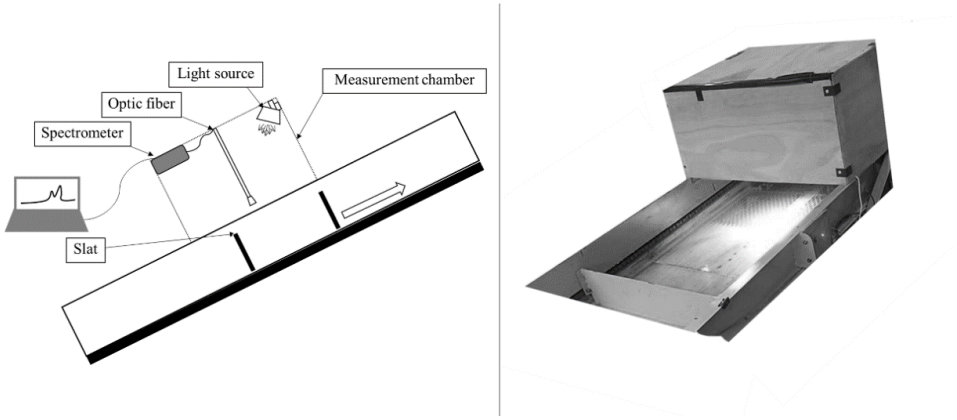


Figure 5.2 Overall components of spectral scanning system

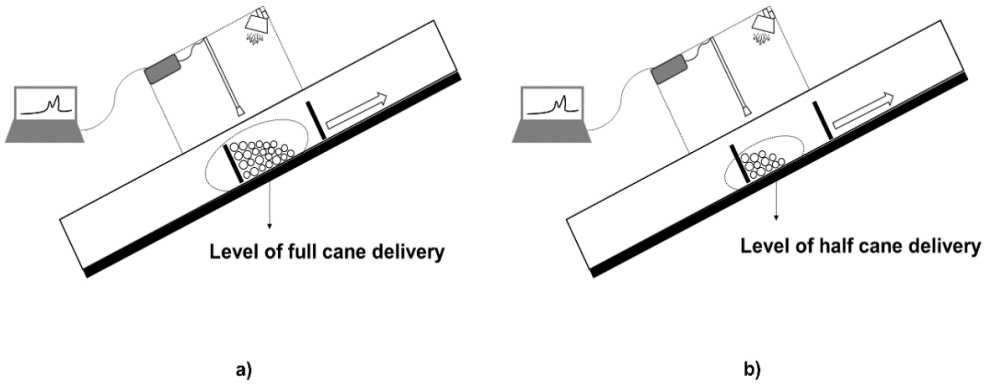


Figure 5.3 Descriptive scheme of two conditions, designed for simulating two levels of cane billets delivery on the conveyor of harvester, a) full delivery and b) half delivery

5.2.2.2 Data collection for unknown sugarcane prediction

To obtain data for unknown prediction, the 30 clumps of sugarcane (formed into 30 billet groups) collected in the different growing season were used. The data for one-half of the 30 groups were collected under the full delivery and that of the other half were collected under the half delivery. For each group, the spectral data were obtained with 1 scan (19 spectra). For the SSC determination, the procedure was below. Therefore, thirty scans (30 × 19 spectra) and 30 SSC values were implemented for the unknown prediction.

5.2.3 Soluble solids (°Brix) determination

Scanned sugarcane billets were collected and then squeezed to obtain juice. The juice from each billet was filtered using a filter paper and immediately poured onto a digital refractometer (Pocket PAL-1, ATAGO, Japan), which used distilled water as the calibration substance, to obtain the SSC value. The SSC values from the billets were averaged for each group. Although, the commercial cane sugar (CCS) is the standard sugarcane quality index which is derived from SSC (°Brix), Pol and fibre content, only the SSC was used in this initial stage of the development because it is the easiest quality parameter to be measured. Surely, significant relationship between near-infrared spectra and CCS value could identify that it could also be developed as an online system if the monitoring the SSC in sugarcane billets on the elevator is perfect.

5.2.4 Chemometric analysis

The absorbance at both ends of the range was first eliminated for all online spectra collected in this work because of the revealed noise caused by the low signal-to-noise ratio at the ends of the detector. This elimination resulted in the spectra between 450 and 900 nm, which were used as a full range for further analysis. In addition, the filtration process proposed by Phetpan et al. (2018) was used to filter out the undesired spectral line (slat and floor detections) in each spectral scanning. Out of the 430 spectral sets, 325 spectral sets (150 for full delivery, 145 for half delivery and 30 for unknown dataset) passed the filtration process. Note that approximately 75% of the overall data obtained from this online spectral detection system remained, and this number was adequate to study whether the different levels of cane delivery on the conveyor affected the developed models in the SSC, i.e., sweetness, regardless of prediction.

The quantitative calibrations using partial least squares (PLS) regression were developed with leave-one-out cross-validation to predict the sweetness (SSC) of sugarcane moving on the conveyor. Before PLS modeling, smoothing of the spectra by the moving average (MA) method with a segment size of 21 points was applied to minimise noises. The standard normal variate (SNV), the best one compared with the others, was then applied to remove the large amount of variability likely caused by

the scattering effect. SNV is one of the pre-processing techniques that deal with individual spectrum independently. The chemometric and multivariate analyses were performed by the Unscrambler software (Unscrambler X 10.3, Camo, Norway). The performance of the developed PLS models was analysed based on the statistical terms, i.e., the coefficient of determination (R^2), root mean square error of cross-validation (RMSECV), and prediction (RMSEP). In addition, the number of latent variables (LVs) used in the modeling was also considered.

Each of the 325 spectral sets obtained after the filtering process was subjected to the pretreatments above and then averaged. Therefore, there were 325 samples (150, 145, and 30 samples for full delivery, half delivery, and the unknown data, respectively) matching the averaged spectra and their SSC. In this paper, two main interesting issues need to be conducted as follows below.

5.2.4.1 Assessment of the PLS models accompanied with and without the visible range

The construction and assessment of the PLS models accompanied with and without the visible range were developed. In this step, the study was conducted to assess whether it was better for predicting the SSC of sugarcane billets moving on the elevator if the model was developed without the visible range (450-700 nm). The dataset from full delivery was selected, which is the ideal practical pattern of sugarcane being transferred on the elevator. The dataset (150 samples) was randomly split into a calibration set (two-thirds of the dataset, 105 samples) and a testing set (one-third of the dataset, 45 samples). Two different PLS models were developed based on 2 different spectral ranges including 450-900 and 700-900 nm. To evaluate the performance of both developed models, the prediction of the external dataset (45 samples) was performed. Their results were compared to each other for selecting the optimum one.

5.2.4.2 Influence of different cane levels on the conveyor on the ability of the SSC prediction

Before the assessment, the influence was studied using principal component analysis (PCA). This algorithm is an explorative data technique that can be used to find patterns in data and leads to the interpretation of sample grouping, including similarities or differences. PCA was applied to check the distribution patterns in the

dataset (325 samples) from full delivery, half delivery (collected in 2017) and unknown sample groups (collected in 2018).

To assess the influence of different cane levels on the ability of the developed models in the SSC prediction, the datasets of both full delivery and half delivery were randomly split into calibration (2/3 of each dataset) and testing (1/3 of each dataset) sets. Three different strategies were performed by developing and validating the PLS models using different calibration and testing sets:

- Strategy 1: Constructed the model with the calibration set of full delivery (105 samples) and validated it with itself (full delivery) and with half delivery testing sets (45 samples each).
- Strategy 2: Constructed the model with the calibration set of half delivery (100 samples) and validated it with itself (half delivery) and with full delivery testing sets (45 samples each).
- Strategy 3: Constructed the global model with the combination of the calibration set in full delivery and half delivery (105 + 100 samples) and validated it with the testing set (45 + 45 samples) and also with the unknown dataset (30 samples).

5.3 Results and discussion

5.3.1 Reference data description

Figure 5.4 shows the distribution of the SSC values used for constructing and validating the PLS models, whereas their summaries of statistical characteristics are shown in Table 1. As seen from the table, all datasets come out within the narrow range caused by one cane cultivar sampling in this study. Although the wide sample variability is necessary to establish a robust model for practical uses, it is adequate for the study to emphasise the possibility of spectroscopic use for monitoring the SSC values of cane moving on the conveyor based on different cane level factors that would be confronted in real situations.

Table 5.1 Statistical soluble solids content values of sugarcane billets used in model development

Dataset	N	Min (°Brix)	Max (°Brix)	Mean (°Brix)	SD
Full delivery					
- Calibration set	105	21.20	24.50	22.64	0.73
- Testing set	45	21.20	23.90	22.49	0.71
Half delivery					
- Calibration set	100	20.80	24.90	22.63	0.89
- Testing set	45	21.20	24.40	22.75	0.81
Data combination (Full + Half delivery)	205	20.80	24.90	22.63	0.81
- Calibration set	90	21.20	24.40	22.62	0.77
- Testing set					
Unknown set	30	21.7	24.33	22.99	0.64

5.3.2 The PLS models for SSC prediction based on 450-900 and 700-900 nm

The comparison of the PLS model results constructed from 2 different ranges is shown in Table 5.2. The model established from the wavelength range of 450-900 nm, covering the visible range and shortwave near infrared range, slightly gave better results than that of 700-900 nm. As an overall result, it seems necessary for the modeling in this work to include the spectral range of the visible region, i.e., 450-700 nm, although the regression coefficient plots (Figure 5.5) of both models are the same with two sugar related peaks at 755 nm (the 4^{th} overtone of C-H stretching of sugar at 762 nm (Osborne et al. 1993) or the 3^{rd} overtone of O-H stretching of sucrose in water at 740 nm) and 890 nm (the 3^{rd} overtone of C-H stretching of sucrose in water at 910 nm (Golic et al. 2003)). However, a few differences in the predictability, the model (700-900 nm) developed without including the visible region employed lower LVs in accounting spectral variance related to the variation in SSC than the model of full wavelength (450-900 nm). This result could be seen obviously

in the x-loading weight plots (Figure 5.6) that could be used to interpret in detail how the information in each X-variable related to the variation in Y. With this concept, the variables with large loading weight values were important for the prediction of SSC. Each spectral line appearing in the plot expresses each LV for data variance explanation. The first 1-2 LVs account for the most variance, especially the first LV accounting for the greatest amount of variance. Hence, the loading weight plots of the models developed from the ranges of 450-900 (Figure 5.6a) and 700-900 (Figure 5.6b) nm were compared. Constructing the model with the wavelength range of 450-900 nm results in the first LV in explaining 70% and 62% of the variance in on-line spectral data and SSC data, respectively. This LV exhibited prominent appearances of the band at approximately 670-700 nm, as related to the chlorophyll content (Guo et al. 2003), and the amplitude of the sugar-related bands was at 755 and 890 nm (Golic et al. 2003; Osborne et al. 1993) and was in the visible range. Similarly, the first LV of the modeling accompanied with the range of 700-900 nm explained 66% and 65% of the variance in the spectral data and SSC, respectively, and showed an obvious amplitude of the sugar bands at 755 and 890 nm. With the ability in explaining data variance of first LV for the both models, it is very attractive in developing SSC prediction model without including the visible range. For the overall variance explanation, the PLS model that was developed from the spectral range of 450-900 nm employed 4 LVs to explain 96% and 81%, whereas that of 700-900 nm desired only 3 LVs for explaining 93% and 80% of the variance in the spectral data and SSC data, respectively. Based on this outcome, the optimum spectral range, suitable for the modeling in this research, was 700-900 nm.

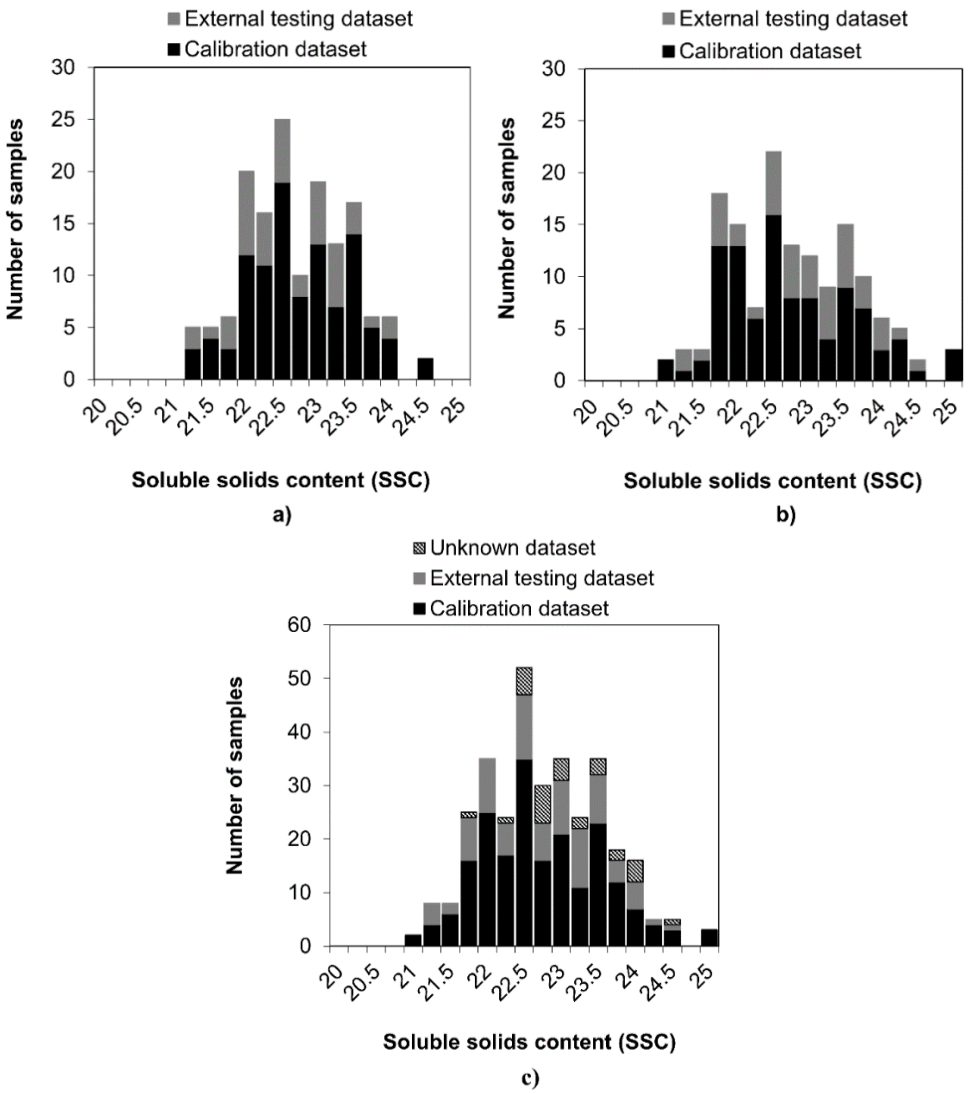


Figure 5.4 Histograms of the soluble solids content values; a) for the calibration and external validation sets of full cane delivery, b) for that of half cane delivery and c) for that of the combination of full and half delivery and for the unknown dataset

Table 5.2 Results of the partial least squares regression models for soluble solids content prediction based on different wavelength ranges

Spectral range (nm)	LVs	Calibration model			External testing	
		R^2	RMSEC (°Brix)	RMSECV (°Brix)	R^2	RMSEP (°Brix)
450-900	4	0.817	0.31	0.34	0.832	0.29
700-900	3	0.799	0.32	0.35	0.805	0.31

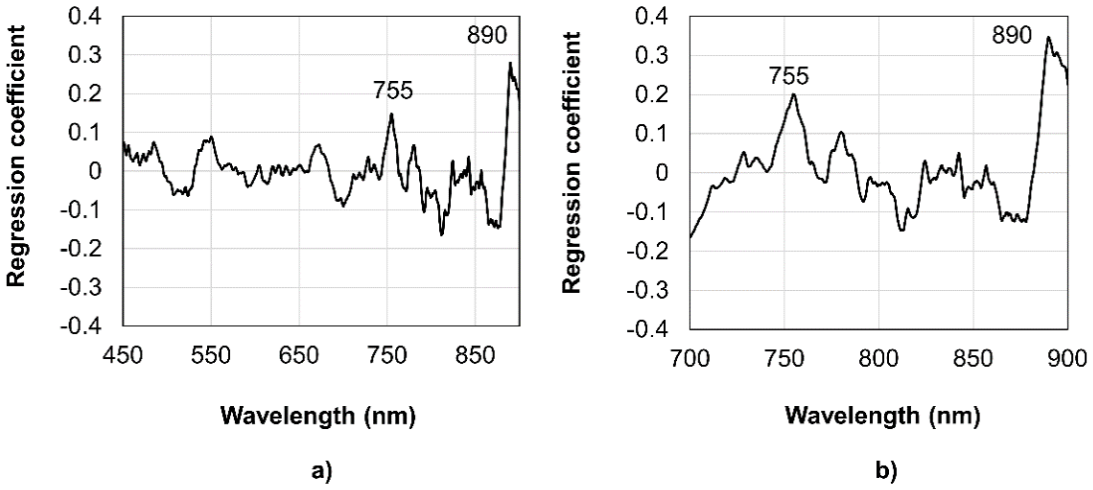


Figure 5.5 Regression coefficient plots of the PLS models constructed from a) the wavelength range with covering the visible range (450-900 nm) and b) the wavelength range without covering the visible range (700-900 nm)

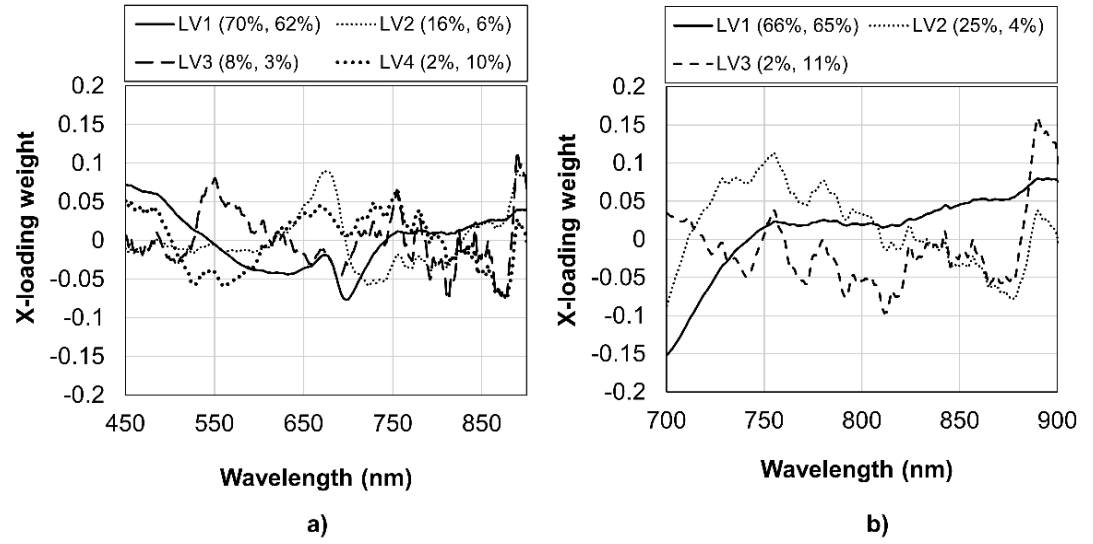


Figure 5.6 X-loading weight plots of the PLS models constructed from a) the wavelength range with covering the visible range (450-900 nm) and b) the wavelength range without covering the visible range (700-900 nm)

5.3.3 The assessment regarding the influence of the cane levels on the SSC prediction

The spectral range of 700-900 nm, i.e., the optimum one, was used for analysis throughout this part. The PCA plot (Figure 5.7) presents the distribution patterns of full delivery, half delivery, and the unknown sample groups, where the first two PCs explained 91% of the total variation in the spectral data. The patterns identified the similarity of all sample groups.

For each strategy studied, the results of constructing and validating the PLS models are summarised in Table 5.3. The analysis of the PLS models based on level-specific cane delivery for the SSC prediction was performed in strategies 1 and 2, in which the models showing the regression coefficient plots in Figure 5.8 were validated by the samples originating from the same and different conditions. As seen in the table, the SSC can be directly predicted if the samples in the PLS models and the prediction sets are the same condition. This outcome shows that the R^2 and RMSEP are approximately 0.69-0.81 and 0.31-0.45 °Brix, respectively, which are also visualised by the plot (Figure 5.9a) scattering the NIR-predicted data versus the reference data (SSC) for the full delivery and half delivery testing sets. In contrast, the assessment of the predictive ability of the models to validate the datasets from the different condition presents results that are not as accurate as those from the same condition. Some fluctuations in the prediction are observed as shown in Figure 5.9b. Based on this outcome, if the PLS models developed based on the both strategies are considered, it is found that an error term (RMSEP) obtained from modeling the dataset of the half cane delivery had higher value than that of full cane delivery. According to Inagaki et al. (2017) and Molla et al. (2016), the higher the pathlength in the spectral scans the low S/N ratio they are which tends to increase the error of prediction. This result corresponds to the behavior of the regression coefficient plot of the model for half cane delivery (Figure 5.8b) that presents several peaks not related to sugar but have the effect on the SSC prediction. Therefore, it could be concluded that the levels of different cane delivery on the elevator substantially affected the predictive accuracy. This outcome led to the analysis in strategy 3 that constructed the PLS model by the samples containing different effects in the calibration set and validated it by the external testing set that had

already been split out. In this paper, the number of samples and the range of SSC values used for constructing the model were 205 samples and 20.8-24.9 °Brix, respectively. This model employed 3 LVs, accounting for 92% and 75% in the on-line spectral and SSC variations, respectively. Its predictive performance, validated by 90 samples in the testing set, could be explained by the error term (RMSEP) of 0.42 °Brix. In addition, this model was tested for the performance by predicting 30 unknown samples collected from different planting seasons and locations. Its performance looked satisfying with an error term (RMSEP) of 0.42 °Brix. Figure 5.10 shows the overall results of the combination model analysis. It consists of the scatter plot of the model for calibration and its internal validation (top left), the scatter plot presenting the predictive ability of the model for the external testing set and unknown set (top right), and regression coefficient plot (bottom).

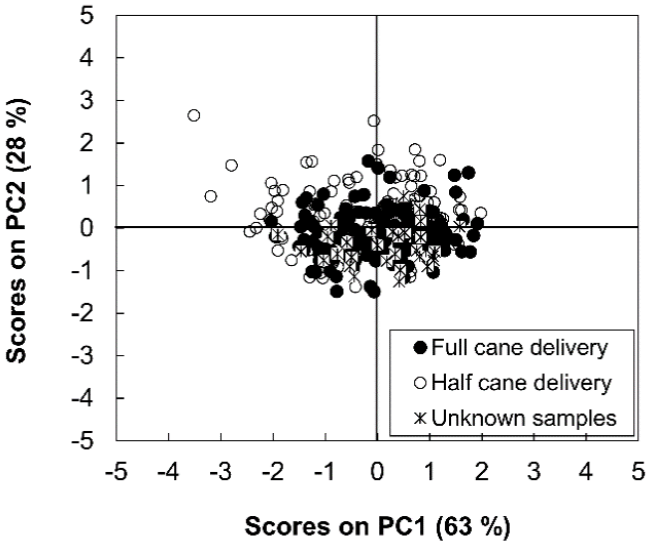


Figure 5.7 Principal component analysis (PC1 and PC2) for the samples from full cane level, half cane level and unknown sample groups obtained after the filtration and the spectral pre-processing processes

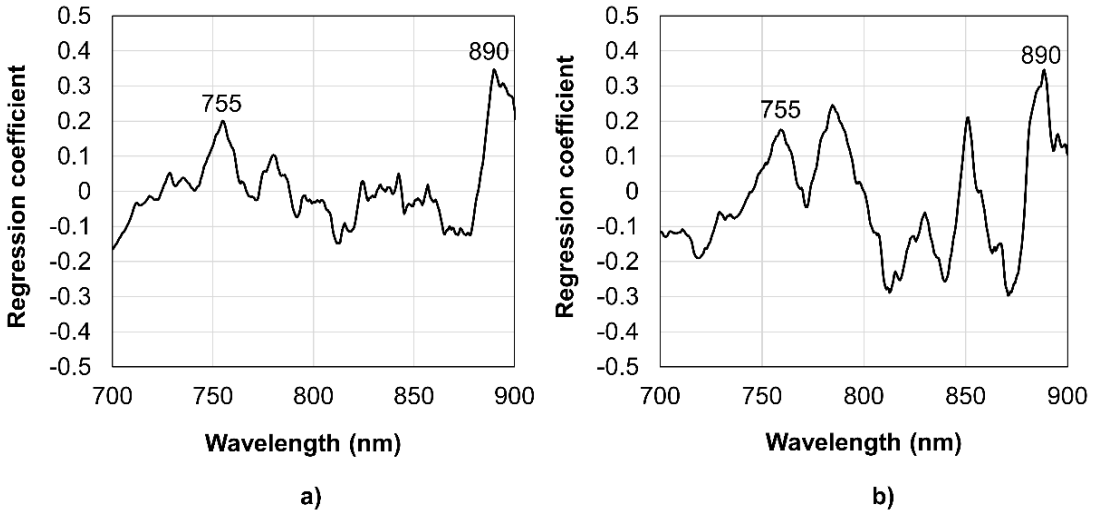


Figure 5.8 Regression coefficient plots of the PLS models constructed from the dataset of a) full cane delivery and b) half cane delivery

Compared to the cane level-specific model, the combined model yielded a satisfying result. It reduced some effects in the SSC prediction caused by the levels of different cane delivery on the elevator, especially if the scattering patterns were compared in the plots shown in Figure 5.9b and top-right of Figure 5.10. Based on this result, the combined model tended to overcome or compensate the interference from different cane levels during spectral detection.

These results could be sent as an initial message to the worldwide sugarcane industry that the desire to recognise or to access the spatial variability of the sugarcane quality is now in a promising place. Nevertheless, the modeling with wide sample variabilities based on sugarcane cultivars and agro-climatic conditions (different growing zones) is necessary to obtain a robust model for practical uses. This step would be done in a further work based on the body of knowledge obtained from this paper, i.e., the combined or global modeling accompanied without the spectra in the visible range.

Table 5.3 Results of soluble solids content prediction for assessing the influence of the different levels of cane delivery on the elevator

	Calibration model				External testing		
	Sample set	R^2	RMSEC (°Brix)	RMSECV (°Brix)	Sample set	R^2	RMSEP (°Brix)
Strategy	Full delivery	0.799	0.32	0.35	Full delivery	0.805	0.31
					Half delivery	0.521	0.55
Strategy	Half delivery	0.794	0.45	0.45	Half delivery	0.686	0.45
					Full delivery	0.529	0.48
Strategy	(Full + Half delivery)	0.752	0.40	0.42	(Full + Half)	0.692	0.42
					Unknown	0.556	0.42

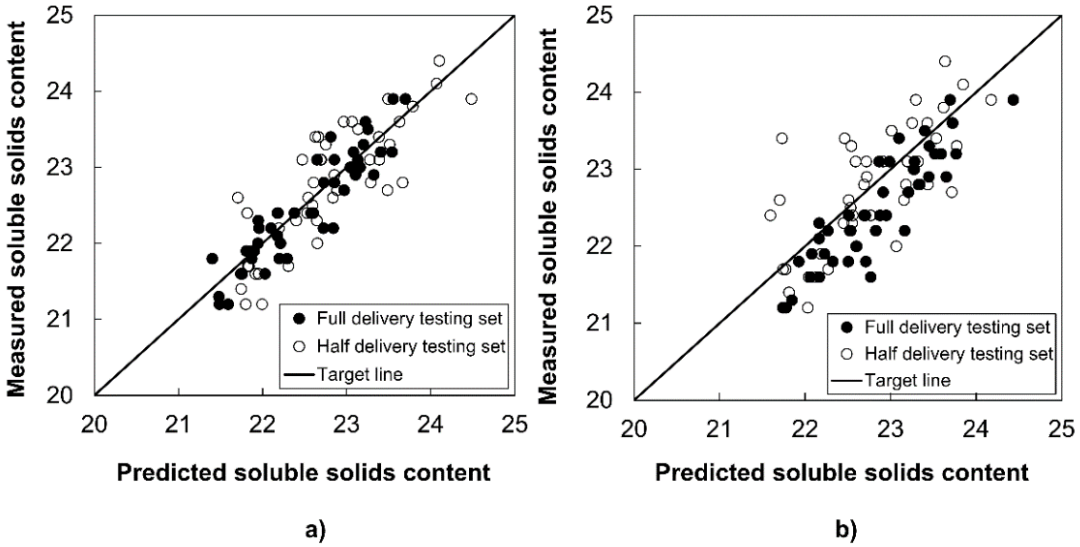


Figure 5.9 The external validation of the specific PLS models developed by either full or half cane delivery; a) with datasets contain the same condition and b) with datasets contain the different condition

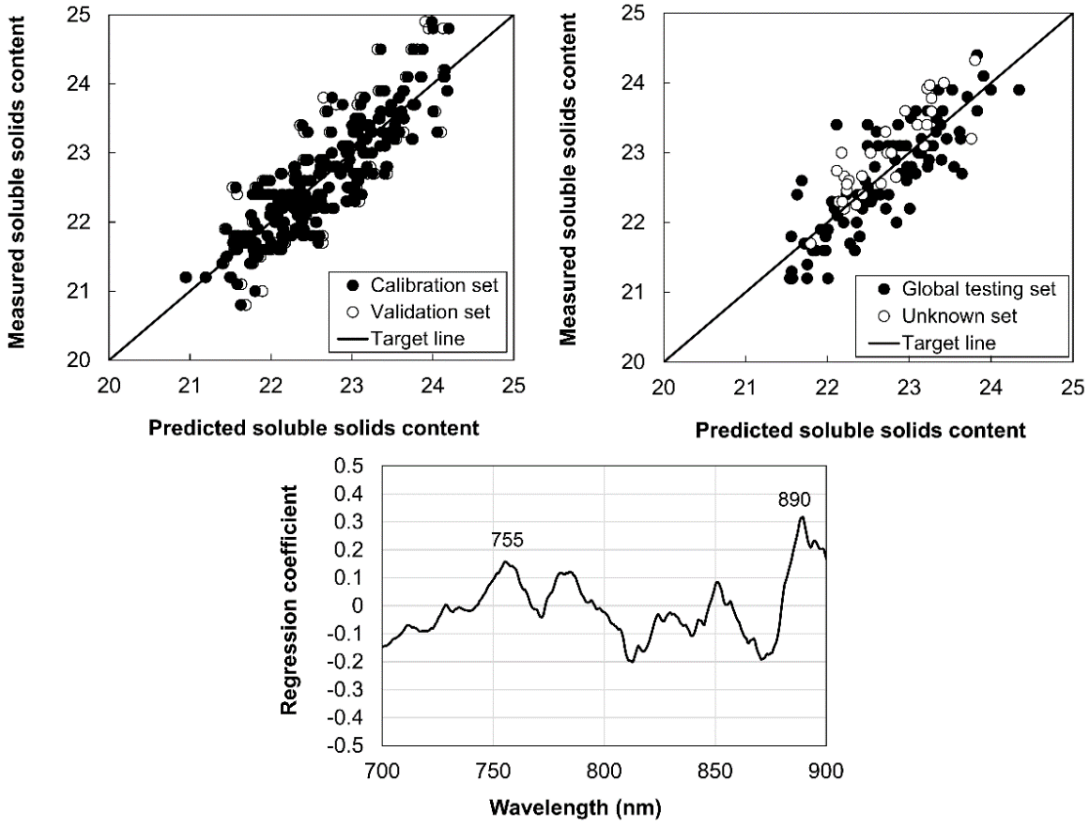


Figure 5.10 The results of the combined model analysis; comparison of the predicted and measured SSC for the PLS modeling (top-left) and its internal validation (top-right), the regression coefficient plot (bottom)

5.4 Conclusion

This work was conducted to explore and present the PLS model which was a partner of the spectroscopic monitoring system for real-time measuring SSC of sugarcane on the elevator. The studies consisted of 2 main parts. They started with the comparative assessment for modeling based on two different spectral ranges. This part was conducted to investigate whether the ability of the PLS model in predicting the SSC of sugarcane billets moving on the elevator was better if the model was developed without the visible range. As a result, it could be concluded that the modeling for the sweetness prediction of sugarcane in this work no needed to include the visible region. Therefore, the model constructed from the spectral range of 700-900 nm is the best, showing an R^2 of approximately 0.80 and error terms obtained from internal (RMSECV) and external (RMSEP) validations of 0.35 and 0.31 °Brix, respectively. The study in the second part, regarding the influence of different

levels of cane delivery on the conveyor, was conducted to assess whether it is possible to develop the PLS model by neglecting this influence. From the study, it can be concluded that the variations that originated from different spectral scanning distances (from the probe to the top layer of cane billets) should not be ignored. This factor can occur in a real situation and is difficult to control. Therefore, a combined or global model that contained data with different effects was developed. This model could account for 92% and 75% of the variation in the spectral data and SSC, respectively, using only 3 LVs. The performance was tested by validating two external testing sets, 90 samples from the same growing season and 30 samples from a different growing season, presenting error terms (RMSEP) for 0.42 and 0.42 °Brix, respectively.

The findings presented here can lead to the development of a sweetness monitoring system of sugarcane during harvesting. Surely, the system would access the site-specific quality of sugarcane in the fields and would allow the production of spatial variability maps describing the quality within the farmlands. Farmers could obtain information corresponding to their crop quality and customise crop input to maximise farm profits. Furthermore, the establishment of a fairer payment system for growers and optimisation of production processes in the mills could happen from having this system.

5.5 References

Bramley R.G.V. and Quabba R.P. 2001. “Opportunities for improving the management of sugarcane production through the adoption of precision agriculture - An Australian perspective.” **Proceedings of the 24th Congress of the International Society of Sugar Cane Technologists.** 38-46.

Golic M., Walsh K. and Lawson P. 2003. “Short-wavelength near-infrared spectra of sucrose, glucose, and fructose with respect to sugar concentration and temperature.” **Applied spectroscopy.** 57(2), 139-145.

Guo Z., Huang W., Peng Y., Chen Q. and Ouyang Q. 2003. “Color compensation and comparison of shortwave near infrared and long wave near infrared spectroscopy for determination of soluble solids content of Fuji apple.” **Postharvest Biology and Technology.** 27, 197-211.

Inagaki T., Watanabe T. and Tsuchikawa S. 2017. “The effect of path length, light intensity and co-added time on the detection limit associated with NIR spectroscopy of potassium hydrogen phthalate in aqueous solution.” **PLOS One**. 12(5), e0176920.

Kingston G. and Hyde R.E. 1995. “Intra-field variation of commercial cane sugar (CCS) values.” **Proceedings of the Australian Society of Sugar Cane Technologists.** 17, 30-38.

Klute U. 2007. “Microwave measuring technology for the sugar industry.” **International Sugar Journal.** 109(1308): 1-6.

Mccarthy S. and Billingsley J. 2002. “A sensor for the sugar cane harvester topper.” **Sensor Review.** 22(3), 242-246.

Molla N., Bakardzhiyski I., Manolova Y., Bambalov V., Cozzolino D. and Antonov L. 2016. “The Effect of Path Length on the Measurement Accuracies of Wine Chemical Parameters by UV, Visible, and Near-Infrared Spectroscopy.” **Food Analytical Methods.** doi: 10.1007/s12161-016-0735-8

Nawi N.M., Chen G. and Jensen T. 2014. “In-field measurement and sampling technologies for monitoring quality in the sugarcane industry: a review.” **Precision Agriculture.** 15, 684–703.

Nawi N.M. 2014. “Development of new measurement methods to determine sugarcane quality from stalk samples.” Ph.D. Thesis, University of Southern Queensland.

Nawi N.M., Chen G. and Jensen T. 2013b. “Visible and shortwave near infrared spectroscopy for predicting sugar content of sugarcane based on a cross-sectional scanning method.” **Journal of Near Infrared Spectroscopy.** 21, 289–297.

Nawi N.M., Chen G., Jensen T. and Mehdizadeh S.A. 2013a. “Prediction and classification of sugar content of sugarcane based on skin scanning using visible and shortwave near infrared.” **Biosystems Engineering**. 115(2), 154-161.

Nelson S.O. 1987. “Potential agricultural applications for RF and microwave energy.” **Transactions of the American Society of Agricultural Engineers**. 30(3), 818-831.

Osborne B.G., Fearn T. and Hindle P.H. 1993. **Practical NIR spectroscopy with applications in food and beverage analysis.** UK: Longman Scientific & Technical.

Phetpan K., Udompetaikul V. and Sirisomboon P. 2018. "An online visible and near-infrared spectroscopic technique for the real-time evaluation of the soluble solids content of sugarcane billets on an elevator conveyor." **Computers and Electronics in Agriculture.** 154, 460-466.

Shah S. and Joshi M. 2010. "Modeling microwave drying kinetics of sugarcane bagasse." **International Journal of Electronics Engineering.** 2(1), 159–163.

Travers S., Bertelsen M.G., Petersen K.K. and Kucheryavskiy S.V. 2014. "Predicting pear (cv. Clara Frijs) dry matter and soluble solids content with near infrared spectroscopy." **LWT - Food Science and Technology.** 59(2), 1107-1113.

Walsh K.B., Guthrie J.A. and Burney J.W. 2000. "Applicaion of commercially available, low cost, miniaturused NIR spectrometers to the assessment of the sugar content of intact fruit." **Australian Journal of Plant Physiology.** 27(12), 1175-1186.

Chapter 6

Informative selection of spectra obtained from an online sugar content prediction system of sugarcane by using statistical index

Main ideas of Chapter 4 and 5 are to present the possibility in real-time sweetness prediction of sugarcane on the elevator by the proposed system and to explore the optimal features for modeling. One of important processes of the system is the spectral filtration which is the first step after the spectral acquisition. However, the process used in the system previously presented is just the basic screening by eliminating all low signal reflections despite being the sugarcane responses. So, the need to explore the informative selection of spectra obtained from the online soluble solids content (SSC) measuring system of sugarcane was focused. This chapter presents the evaluation of strategies in classifying between sugarcane and non-sugarcane (slat and floor) spectral detection. At the end of this chapter, spectral outputs obtained from this new filtration process were modeled against their SSC values in order to compare with the results obtained from the old process. Overview of this chapter could be seen in Figure 6.1.

* This chapter constituted the publication article: Phetpan K, Udompetiakul V, Sirisomboon P. “Informative selection of spectra obtained from an online sugar content prediction system of sugarcane by using statistical index.” **Thai Society of Agricultural Engineering Journal,** 25(2) (2019) pp.19-27.

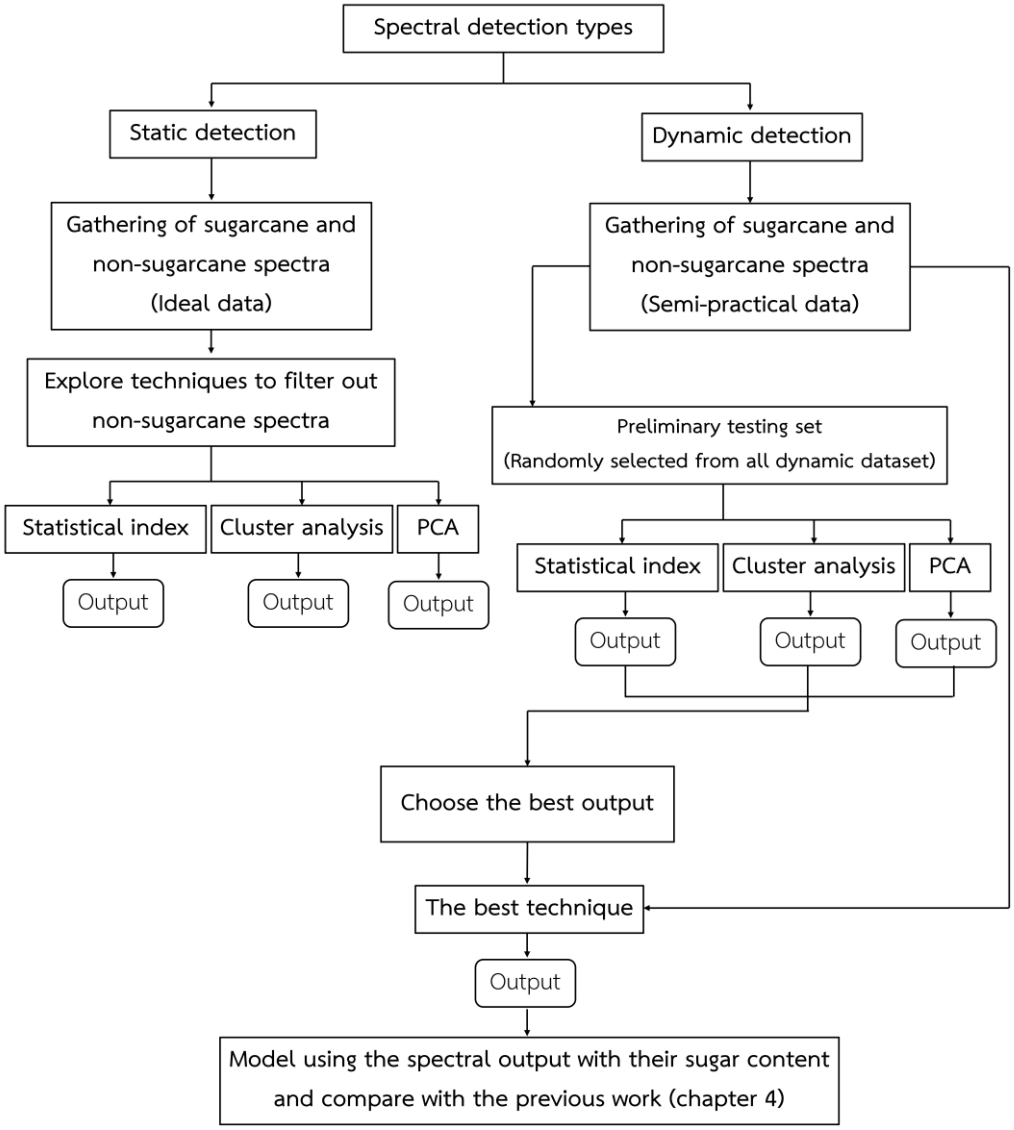


Figure 6.1 Overview for the informative selection of spectra.

6.1 Introduction

Agricultural products are one of the necessary resources for all living things around the world. Sugarcane is one of the main agricultural products, used in sugar production worldwide and used to produce alternative fuels in the form of ethanol in some countries as well (Cookson, 2012). With the increasing world population, it is expected to reach 8.6, 9.8 and 11.2 billion in 2030, 2050 and 2100, respectively (United Nations, 2017). So, to increase the yield and maintain the quality in the production of sugarcane without enlargement of growing areas should be focused to support the increasing population.

However, the world of sugarcane production is still confronted with the problem of variation in the yield and especially quality within the fields (Kingston and Hyde 1995; Bramley and Quabba 2001). This has prompted efforts in accessing the variation within the fields to achieve improvement in the production such as field input.

Several means such as electronic refractometers (Mccarthy and Billingsley 2002), microwaves (Klute 2007; Nelson 1987; Shah and Joshi 2010) and spectroscopic techniques (Nawi et al. 2013a; Nawi et al. 2013b; Phetpan et al. 2018) have been studied to develop a rapid sugarcane quality measurement. Interestingly, Phetpan et al. (2018) developed the non-destructive, visible and near-infrared spectroscopic technique for online measuring soluble solids content (SSC) of sugarcane billets on an elevated conveyor. With their results reporting a root mean squares error of prediction (RMSEP) of 0.30 °Brix, it shows the possibility in measuring the SSC of moving sugarcane billets using the spectroscopic technique. For the continuous online spectral measurement, a spectral screening process is needed to filter out unwanted spectra because many sources are detected and recorded. However, the spectral filtration proposed in Phetpan et al. (2018) study was just the basic screening process by eliminating all low signal reflections despite being the sugarcane responses.

Pattern recognition and classification are typically divided into two main groups, i.e. unsupervised and supervised methods. Grouping of any objects with no supervisor in the sense of known membership is called unsupervised manner (Otto, 2017). Principal component analysis (PCA) and cluster analysis are some of the unsupervised learning methods, widely used in the chemometric analysis. If the membership of objects to classify is known, the methods of the supervision called pattern recognition can be used (Otto, 2017). Linear learning machine (LLM), discriminant analysis, the soft independent modeling of class analogies (SIMCA), and support vector machines (SVMs) are some of this one for the application in recognizing patterns (Otto, 2017). In the real world, not only these were two methods applied for the object classification, but also statistical index such as Normalized Difference Vegetation Index (NDVI) and Green-Red Vegetation Index (GRVI) were widely used in the remote sensing area. These are the well-known indexes

used for tracking phenological changes which are probably used in distinguishing between the green vegetation and the other types of the ground covers by measuring the difference between two bands, i.e. near-infrared and red light for the NDVI and green and red light for the GRVI (Motohka et al. 2010).

Based on spectral data set in this study gathering sugarcane and non-sugarcane detecting signals, adopting some statistical procedures in order to obtain the required spectra of sugarcane during an online scanning is necessary. The unsupervised and statistical index methods are suitable for this data set because of their unknown memberships of the spectra during classification. Therefore, the objectives of this research are threefold: 1) to initially assess the application of principal component analysis (PCA), cluster analysis and statistical index in filtering out non-sugarcane spectra for static data set, 2) to select the best one for classifying the sugarcane and non-sugarcane spectra in dynamic data set and 3) to establish the partial least squares (PLS) model by using the output spectra from the best filtration method against their SSC values.

6.2 Materials and methods

6.2.1 Samples

Fifty clumps of sugarcane were used in this study. In each clump, the sugarcane was chopped into billets with an approximate length of 20 cm (see in Phetpan et al., 2018 for more details). So, there were 50 groups of sugarcane billets in total.

6.2.2 Spectral detection

An online measurement system used in this study consisted of two main parts, a cane billet elevator and a spectral acquisition system. A Vis/SW-NIR spectrometer (AvaSpec-2048-USB2, Avantes BV, Netherlands) was installed for the spectral acquisition system, operating in the spectral range of 350-1100 nm with the spectral resolution of 2.4 nm. The integration time was set for 14 ms, yielding approximately 90% full-scale Analog-to-Digital Converter (ADC) of the reference material reflectance (see in Phetpan et al., 2018 for more details). Static and dynamic spectral detections were performed.

6.2.3 Data analysis

Grouping of analytical data according to their similar elemental pattern is possible by either means of clustering methods or by projecting the high dimensional data onto lower dimensional space (Otto, 2017). These methods are performed with no supervisor in the sense of known membership of objects to classify. So, principal component analysis (PCA), one of the projection methods, and cluster analysis were used for classifying sugarcane and non-sugarcane samples.

To decompose the original matrix (X) by a product of the score (T) and loading (P) matrices, is the key idea of PCA which can be formulated as follows:

$$X = TP^T \tag{6.1}$$

With this concept, to project matrix (X) or spectral matrix onto the lower dimensional space (T), the equation (6.1) is to be converted by:

$$T = XP \tag{6.2}$$

To interpret the results of PCA for the discrimination between the sample groups, visualizing by plotting the elements of two scored vectors of the matrix (T), especially the first two vectors explaining the most variance in the original matrix, could be done.

For the cluster analysis, the similarity of the objects is decided based on the distance measures. The shorter the distance between objects the more similar they are (Otto, 2017). In this study, the Euclidean distance, widely used, was applied according to:

$$d_{ij} = \left[\sum_{k=1}^{K} \left|x_{ik} - x_{jk}\right|^2\right]^{1/2} \tag{6.3}$$

where K is the number of variables; i, j are the indices for object i and j.

In addition, the Normalized Difference Vegetation Index (NDVI) and the Green-Red Vegetation Index (GRVI), which are a normalized ratio of near-infrared and red reflectance and that of green and red reflectance, respectively, were initially tested

to assess the possibilities for classifying the different spectra in this work. No such approaches were possible.

With the behavior of spectra scanned under the static state, it seems that the spectra had their own pattern according to their sources (Figure 6.3). The own pattern of sugarcane responses presents the low and high variation in the range from 560 to 640 nm and from 680 to 750 nm, respectively. High variation in both ranges is the characteristic of slat detection, whereas low variation throughout the spectral range is that of floor detection. With very low %reflection and variation of floor spectra, these original floor spectra can be easily cut off from other raw spectra. The standard deviation (SD) throughout the spectral line for 1 was used as a threshold for removing the floor spectra.

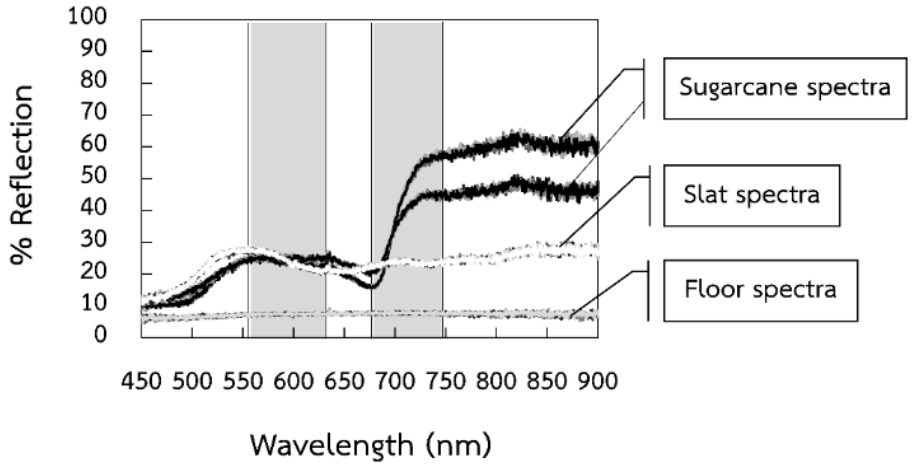


Figure 6.3 Spectral characteristics based on the detection of the different objects.

Based on these two leftover patterns, the concept of the normalized ratio mentioned above was modified to be an alternative channel for this work. So, a statistical index relying on a normalized ratio of SD of reflection value in the range from 560 to 640 nm and 680 to 750 nm was proposed in this study. The index is defined as follows:

$$\text{Statistical index} = \frac{SD_{560\ to\ 640} - SD_{680\ to\ 750}}{SD_{560\ to\ 640} + SD_{680\ to\ 750}} \tag{6.4}$$

For the data analysis, the spectral range of 450-900 nm, defined as a full range, and that of 560-640 and 680-750 nm were used for the application of PCA and cluster analysis. PCA and cluster analysis were firstly applied to the static data set after the floor spectra removal in order to test the filtering out slat spectra based on both the full range and the 560-640 and 680-750 nm. In the case of PCA, estimation of the loading matrices P^T from the scaled-matrices (X) based on two spectral ranges and computation of new score matrices (T) according to the equation (6.2) were performed, respectively. With this computation, two matrices (T) based on these two ranges were the output of the PCA and were used for the discrimination between the sugarcane and slat spectra. For the classification with cluster analysis, the Euclidean distance was applied based on two spectral ranges resulting in different K according to the equation (6.3). Based on these two means, four outputs of the classification were obtained, two outputs (450-900 nm and 560-640, 680-750 nm) from the PCA and the other two outputs from the cluster analysis. For the statistical index, it was also applied to the data of static data set. The specific index value was then obtained to be an indicator to keep up the sugarcane responses. In order to evaluate the performance of these three methods in filtering the slat spectra out, they were applied to dynamic spectral data set which was already cut the floor spectra off. Ten sample groups (74 spectra) out of one-hundred and seventy-six groups were randomized and used as a testing set for the evaluation. The results showing the percentage of correct classification of those compared among the three methods were presented. Based on these, the best one was applied to screen non-sugarcane (slat) spectra out from all the dynamic data set. The spectra survived from this process were mathematically pretreated by moving average (MA) smoothing with segment size of 21 points and standard normal variate (SNV) in order to minimize spectral noises and to diminish the offset effect in the spectra, respectively. At the end of this study, these spectra were modeled against their SSC values with partial least square (PLS) regression in order to compare with the results obtained from previous work (Phetpan et al., 2018). So, the PLS model, resulting standard error of cross-validation ($SECV$) was compared using a Fisher's test (F value) (Molla et al., 2016) to see a significant difference in statistics. The F value was calculated as

$$F = \frac{SECV2}{SECV1}, \text{ where } SECV1 < SECV2 \quad (6.5)$$

The calculated F value was compared with the confidence limit F critical (1 – α, n_1 – 1, n_2 – 2), obtained from the distribution F table, where α is the test significance level (α = 0.05), n_1 for the sample number modeled at the previous work, and n_2 for that in this work (n_1 and n_2 = 100 for this test) (Molla et al., 2016). The differences between these two models are significant when $F > F$ limit.

In this study, the statistical index computations were conducted by R software (R Core Team, 2018), using the dplyr package (Wickham et al., 2017), whereas the PCA, cluster analysis, mathematical pretreatments as well as the PLS modeling were conducted by the software for multivariate analysis (Unscrambler X 10.3, Camo, Norway).

6.3 Results and discussion

The application of PCA and statistical index employed with the static data set to filter out slat spectra are shown in Figure 6.4 and 6.5, respectively, whereas that of cluster analysis is shown in Table 6.1. The PCA outcome with the first principle component (PC) explained 99% of the total variation in the spectra identifies the tendency in classifying between sugarcane and slat spectra. The results of PCA based on the two spectral ranges show the same distribution patterns. With this result, the spectral range of 560-640 and 680-750 was used for further analysis, whereas zero was used as a threshold for the classification.

For the statistical index test, the scatter plot clearly groups between two different objects (cane and slat). Sugarcane detection has a very high value in negative, whereas slat detection has very high in positive. Therefore, the value of -0.6 was selected as a threshold to distinguish between sugarcane and slat detections. With this threshold, the spectra could be filtered out if the value was higher than -0.6. In the case of cluster analysis, the results shown in the Table 6.1 identify that this method probably works well in the future for filtering out the slat detecting signals. However, these tests were done based on ideal spectral data set (static spectral detection). So with the very good results they are, these applications were also tested with the dynamic data set.

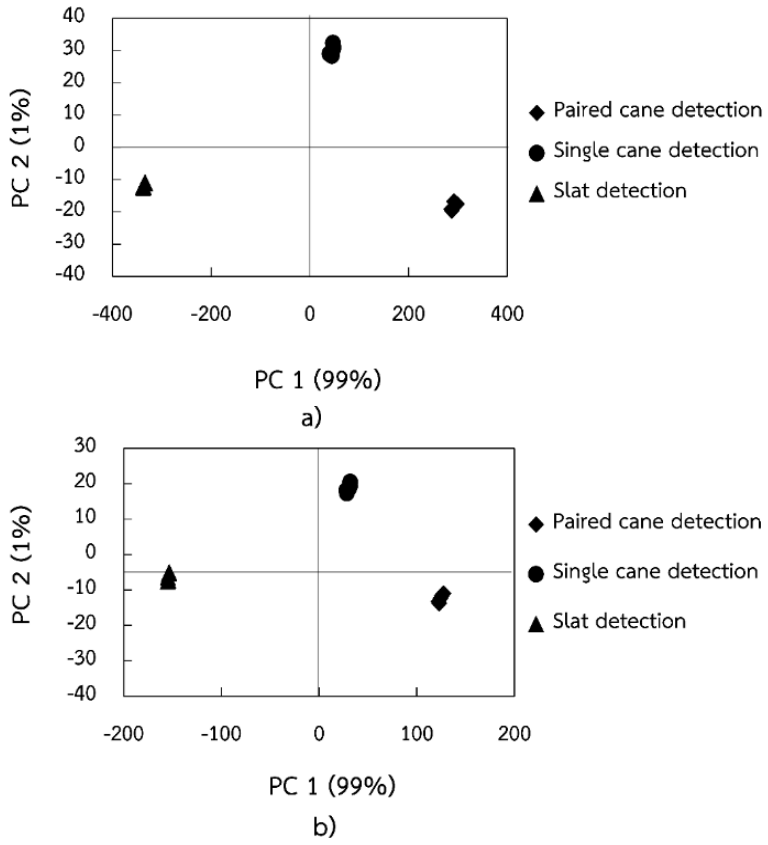


Figure 6.4 Spectral filtration output of the PCA applied to static data set a) outcome based on full spectral range (450-900 nm) and b) outcome based on 560-640 and 680-750 nm

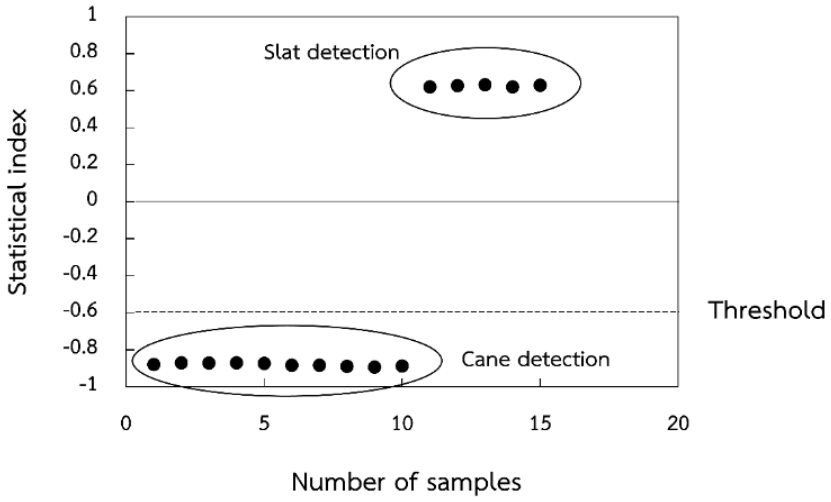


Figure 6.5 Spectral filtration output of statistical index applied to the static data set

Table 6.1 Application of cluster analysis in filtering out slat spectra for data set of static spectral detection

Objects	450-900 (nm)	560-640 & 680-750 (nm)
Single cane detection	0	0
	0	0
	0	0
	0	0
	0	0
Paired cane detection	0	0
	0	0
	0	0
	0	0
	0	0
Slat detection	1	1
	1	1
	1	1
	1	1
	1	1
Overall accuracy	100%	100%

Note: 0 represents sugarcane identification and 1 represents slat identification.

Table 6.2 shows the evaluation results of the performance of the three methods in classifying between sugarcane and slat spectra from the testing data set (74 spectra previously mentioned). Percentage of overall accuracy obtained from PCA and cluster analysis are between 48.65-56.76%. There is a bit difference of overall accuracy in PCA employed based on two different spectral ranges, whereas the same result is obtained from the cluster analysis. This identifies that the different range of spectra has no influence for the PCA and clustering applications. Based on the results, these two methods could not be used to filter out slat spectra from the dynamic data set. The best result for the evaluation in filtering out non-sugarcane (slat) spectra from the dynamic data set is from the statistical index application, 89.19% for overall accuracy. With this accuracy, eight sugarcane detections were

identified as slat spectra whereas no slat detection was identified as the cane. This is what we expected is that a few cane spectral detections with low reflectance could be removed out as non-sugarcane spectra but keeping up slat detection as sugarcane spectra are not allowed. This statistical index was modified from the NDVI and GRVI, which are the normalized ratio relying on measuring the difference between two bands. The application of this index for identifying the difference between spectra of sugarcane and slat that characterizes their own pattern based on different standard deviation is satisfied. So, this index was applied to screen non-sugarcane (slat) spectra out from all the dynamic data set. The filtration result is shown in Figure 6.6, displaying the minimum signal reflection in the region around 750-900 nm of around 3%. This number is quite low compared to that (20%) of filtration proposed by Phetpan et al. (2018). However, this indicates the performance of this index to keep up only the sugarcane detecting signals. The result could also be confirmed by the pretreated spectra in Figure 5b with no presentation of the slat signal except for the sugarcane spectra. With this result, all of 176 spectral sets (100%) survived for this study whereas only 150 out of the 176 sets (85%) survived for the filtration process in Phetpan et al. (2018). Based on this, the statistical index proposed in this study works very well in classifying the sugarcane and non-sugarcane detection especially keeping up the sugarcane signal despite having a low signal reflection.

A PLS model was developed using the spectra obtained from the filtration by statistical index to compare with the PLS model presented in Phetpan et al. (2018). So, the number of samples and pretreatment used for modeling were the same as the previous work. Figure 6 shows the result of PLS modeling employed using the spectra kept up by the statistical index. The model employed 4 latent variables (LVs) to account 92% and 77% in dynamic spectral and SSC variations, respectively. Two sugar related peaks at 755 nm (the 4th overtone of C-H stretching of sugar at 762 nm (Osborne et al. 1993) or the 3rd overtone of O-H stretching of sucrose in water at 740 nm) and 890 nm (the 3rd overtone of C-H stretching of sucrose in water at 910 nm (Golic et al. 2003)) were still found in the regression coefficient plot. The predictive performance of this model, validated by 50 samples in the prediction set, could be explained by root mean square error of prediction (RMSEP) of 0.44 °Brix.

Comparison the SECV between the two PLS models found that the modeling using the data set obtained from the statistical index in this work tends to increase the standard error of cross-validation, SECV of 0.43, whereas the model presented in previous work had 0.34 in term of SECV. However, there is no statistically significant difference between these two PLS models when F (0.05, 99, 99) < F limit (1.26 < 1.39).

Table 6.2 The applications of PCA, cluster analysis and statistical index for filtering out the slat spectra from the dynamic data set

Methods for classification	Identified cane as slat	Identified slat as cane	% Overall accuracy
PCA *	38 (74)	0 (74)	48.65
PCA **	36 (74)	0 (74)	51.35
Cluster analysis *	31 (74)	1 (74)	56.76
Cluster analysis **	31 (74)	1 (74)	56.76
Statistical index	8 (74)	0 (74)	89.19

Note: * employed with the spectral range of 450-900 nm, whereas ** employed with that of 560-640 and 680-750 nm.

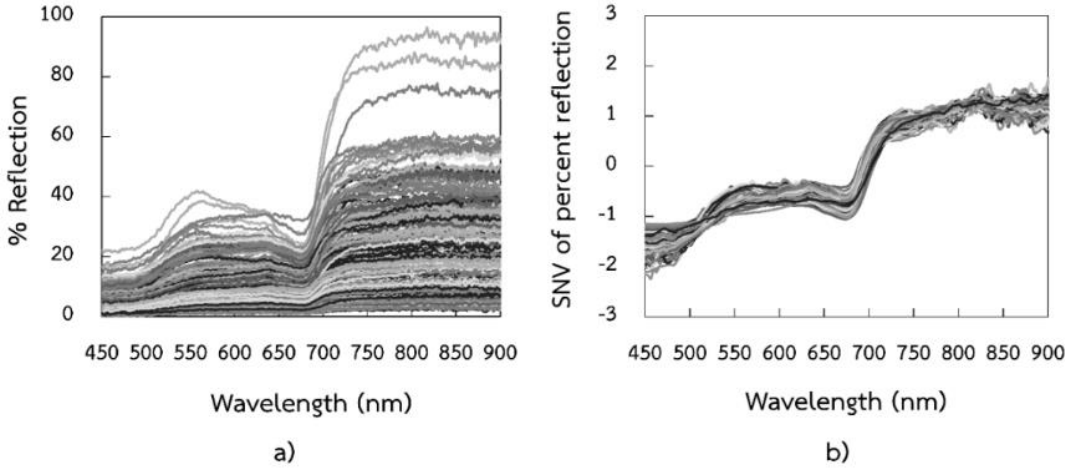


Figure 6.6 Spectral output remained from classifying by the statistical index, a) raw output and b) pretreated output

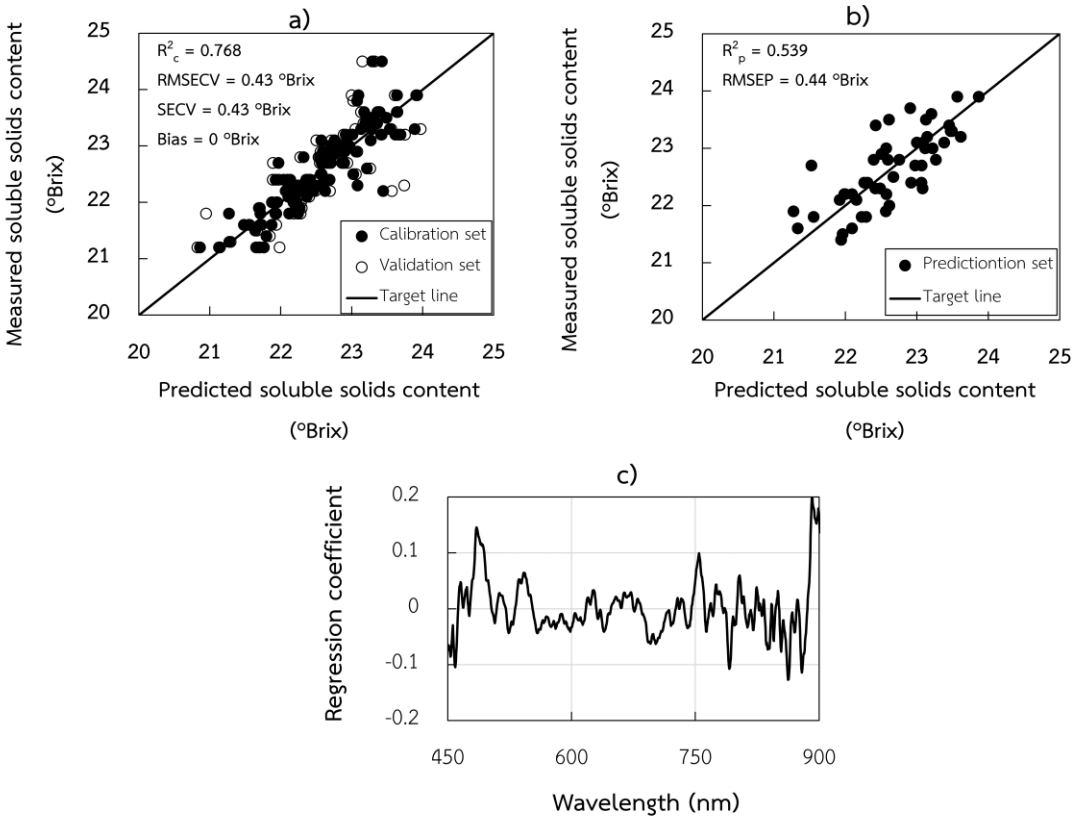


Figure 6.7 The result of PLS model constructed using spectra kept up by the statistical index; a) scatter plot the PLS modeling and its internal validation, b) scatter plot of external validation and c) the regression coefficient plot

6.4 Conclusion

The application of principal component analysis (PCA), cluster analysis and statistical index showed a good tendency in screening out non-sugarcane spectra from the static data set (ideal case) with clear grouping between the different objects.

The statistical index was the best for removing the slat detection signals from the testing data set (randomly selected 74 spectra from the dynamic data set), showing 89.19% for overall accuracy, compared to the PCA and cluster analysis reflecting the accuracy between 48.65-56.76%.

The PLS modeling that employed the spectra obtained from the filtration using the statistical index against their SSC values displayed a coefficient of determination of calibration (R2) of 0.768. It was internally validated and showed a

root mean square error of cross-validation (RMSECV) of 0.43 °Brix, whereas the RMSECV of 0.33 °Brix was shown in previous work. However, no statistically significant difference was observed ($p > 0.05$) from the error increased. So, this index would be applied as the filtration process for further work in developing the on-line soluble solid content measuring system of sugarcane on the conveyor.

6.5 References

Bramley R.G.V. and Quabba R.P. 2001. "Opportunities for improving the management of sugarcane production through the adoption of precision agriculture - An Australian perspective." **Proceedings of the 24th Congress of the International Society of Sugar Cane Technologists**. 38-46.

Cookson C. 2012. "A tank of sugar: how Brazil runs on biofuel." **FT Magazine.** United Kingdom.

Golic M., Walsh K. and Lawson P. 2003. "Short-wavelength near infrared spectra of sucrose, glucose, and fructose with respect to sugar concentration and temperature." **Applied Spectroscopy.** 57, 139-145.

Kingston G. and Hyde R.E. 1995. "Intra-field variation of commercial cane sugar (CCS) values." **Proceedings of the Australian Society of Sugar Cane Technologists.** 17, 30-38.

Klute U. 2007. "Microwave measuring technology for the sugar industry." **International Sugar Journal**. 109(1308), 1-6.

Mccarthy S. and Billingsley J. 2002. "A sensor for the sugar cane harvester topper." **Sensor Review**. 22(3), 242-246.

Molla, N., Bakardzhiyski, I., Manolova, Y., Bambalov, V., Cozzolino, D., Antonov, L. 2016. "The Effect of Path Length on the Measurement Accuracies of Wine Chemical Parameters by UV, Visible, and Near-Infrared Spectroscopy." **Food Analytical Methods**. 10(5), 1156–1163.

Motohka T., Nasahara K.N., Oguma H., Tsuchida S. 2010. "Applicability of Green-Red Vegetation Index for remote sensing of vegetation phenology." **Remote Sensing**. 2, 2369-2387.

Nawi N.M., Chen G. and Jensen T. 2013b. "Visible and shortwave near infrared spectroscopy for predicting sugar content of sugarcane based on a cross-

sectional scanning method." **Journal of Near Infrared Spectroscopy**. 21, 289–297.

Nawi N.M., Chen G., Jensen T. and Mehdizadeh S.A. 2013a. "Prediction and classification of sugar content of sugarcane based on skin scanning using visible and shortwave near infrared." **Biosystems Engineering.** 115(2), 154-161.

Nelson S.O. 1987. "Potential agricultural applications for RF and microwave energy." **Transactions of the American Society of Agricultural Engineers.** 30(3), 818-831.

Osborne B., Fearn T. and Hindle P. 1993. **Practical NIR spectroscopy with applications in food and beverage analysis.** UK: Longman Scientific & Technical.

Otto M. 1999. **Chemometrics: Statistics and Computer Application in Analytical Chemistry.** New York: Wiley-VCH.

Phetpan K., Udompetaikul V. and Sirisomboon P. 2018. "An online visible and near-infrared spectroscopic technique for the real-time evaluation of the soluble solids content of sugarcane billets on an elevator conveyor." **Computers and Electronics in Agriculture**. 154, 460-466.

R Core Team. 2018. "**R: A language and environment for statistical computing. R Foundation for Statistical Computing.**" Vienna, Austria. URL https://www.R-project.org/.

Shah S. and Joshi M. 2010. "Modeling microwave drying kinetics of sugarcane bagasse." **International Journal of Electronics Engineering** 2(1): 159–163.

United Nations. 2017. "**News: World population projected to reach 9.8 billion in 2050, and 11.2 billion in 2100.**" Available at: https://www.un.org/development/desa/en/news/population/world-population-prospects-2017.html. Accessed on 4 January 2019.

Wickham H., Francois R., Henry L. and Müller K. 2017. "**dplyr: A Grammar of Data Manipulation.**" R package version 0.7.4. https://CRAN.R-project.org/package=dplyr.

Chapter 7
Conclusions and recommendations

The need to access site-specific information of sugarcane quality leads to attempts in searching for sensors for the use as real-time sensing devices. Near-infrared (NIR) spectroscopic technique has been studied for a period and claimed that it is suitable to develop as a sensor for monitoring sugarcane quality during harvesting. Just an idea for real-time measurement of sugar content in sugarcane using the NIR technique is found. Based on this, outcome of this thesis can confirm if the NIR sensor is possible to develop as an online sugar content monitoring system of sugarcane being transferred on the elevator conveyor of sugarcane harvester.

7.1 Summary of findings

The preliminary study of this thesis was conducted regarding the influence of sugarcane variety on sugar content prediction using near-infrared spectroscopy presented in Chapter 3. The study demonstrated that the sugarcane cultivar affects the sugar content prediction based on the results presented that the specific variety spectroscopic models could not be able to predict sugar content of the other variety. For field use, the global calibration model was suggested as it compensated the source of errors by gathering different sources of variance in the model.

The proposed prototype online sugar content measuring system of sugarcane was designed to be installed on an elevator conveyor, exhibited in Chapter 4. The system can detect the spectra of the cane billets being conveyed at the speed of 2 m/s (typical speed of a sugarcane harvester elevator) using an integration time of 14 ms. Based on the results obtained from previous study (Chapter 3), only one commercial sugarcane cultivar was used for assessing the performance of the calibration model for the real-time sugar content prediction. PLS modeling showed that the system is certainly feasible for the online SSC measurement of the sugarcane billets on the elevator.

Determining the calibration model suitable for real use in the future, exhibited in Chapter 5, was constructed based on two different spectral ranges. One covers the visible and shortwave near-infrared (Vis/SW-NIR) spectral range (450-900 nm), whereas another one includes only the SW-NIR range (700-900 nm). The study was conducted to assess whether the sugar content predictive ability of the model was better if it was developed without the visible range. As a result, it could be concluded that the modeling for the sugar content prediction of sugarcane in this work was no needed to include the visible region. For a study on influence of different levels of cane delivery on the predictive ability of the model, the results identified that we should not develop the calibration model by neglecting this influence. Therefore, a global model gathering data with different effects is one of the solutions proposed in this study.

Exploration and evaluation of spectral classification methods for classifying between sugarcane and non-sugarcane detecting signals, presented in Chapter 6, could be concluded that filtering non-sugarcane spectra out using the statistical index method was the best, showing 89.19% for overall accuracy. Modeling by employing the sugarcane spectra classified by the statistical index against their SSC values showed no statistically significant difference compared to the model constructed from the previous work.

7.2 Conclusion

Based on the findings, this thesis can confirm that the spectroscopic technique can be used for developing an online sugar content monitoring system of sugarcane to be installed on the elevator conveyor of sugarcane harvester.

Modeling the calibration for sugar content prediction of sugarcane by excluding the visible region seemed to be better and more suitable. Using SW-NIR spectral range for modeling, 700-900 nm, results in neglecting the effects of different pigment of sugarcane cultivars to be modeled and predicted for field use. Gathering the data variances that originated from different sources of spectral scanning distances, sugarcane varieties, etc for modeling, called global model, is the best for field use. Using the statistical index in screening non-sugarcane out is rather more

suitable compared to the basic screening process, eliminating all low signal reflections despite being the sugarcane responses.

7.3 Recommendations

This thesis has successfully achieved all objectives defined in Chapter 1. It is the first research aimed to prove that near-infrared spectroscopic technique could be adopted as the main heart in developing an online sugar content monitoring system of sugarcane to be installed on the elevator conveyor of sugarcane harvester. Main experiments appeared in Chapter 4 to 6 were performed by using the proposed system that limited some issues for studying. Hence, such aspects considered based on the experience gained from this thesis are listed as the recommendations for further state.

- From the observation of the cane billets movement on the harvester's conveyor, not only the horizontal style is the pattern of cane billets movement, but also the vertical style. With this behavior, some billets stood and appeared over the top edge of the slat. However, the overhead part of the slat could not be over the top edge of the elevator in which the bars are mounted. The distance between the top edge of the slat and elevator was standard at 9 cm for cane harvester (John Deere 3520). Hence, the optic fiber installation inside the proposed system at 9 cm over the top edge of the elevator was chosen to prevent the harm caused by the billets movement. Based on our finding, the difference in spectral scanning distances affects the sugar content prediction of the model. Of course, it would be interesting to see the results obtained with the fiber fixed in different positions. So, determining the rather suitable position for the optic installation that provides no statistically significant difference of predictive ability of the model is recommended for further study.
- Based on above recommendation, the fact that the closer the objects (sugarcane billets) of the optic installation, the smaller the field of view (FOV) for spectral detection is. So, if the other more suitable positions of the optic

installation are found, the modeling developed for one-by-one billets prediction will also be interesting.

- The system studied in this thesis was conducted under the laboratory circumstance simulating by sensing the sugarcane billets being transferred on the elevator conveyor with regardless of leafy trash or even soil particles mixed with the billets. These are the real situation factors beyond the control that interfere in the NIR spectral measurement if the system is installed on the elevator of sugarcane harvester. So, training the spectral filtration process of the system by recognizing these environmental factors in order to screen the undesired spectra out prior the prediction process is needed for further study. Also, dealing with temperature variation in the field is another unavoidable factor. So, global-temperature modeling for sugar content prediction of sugarcane being harvested is needed for field application.

Author biography

Name-Surname Acting Sub Lt. Kittisak Phetpan

Date of birth 11 December 1989

Contact address 72/1 M.9 Khuankalong, Khuankalong district, Satun 91130

Email address kphetpan@gmail.com

Contact number 099-3242127

Education background

D.Eng. (Agricultural Engineering) from King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand (2015-2019, GPA: 4.00)

M.Eng. (Agricultural Engineering) from King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand (2013-2014, GPA: 3.88)

B.Eng. (Agricultural Engineering, 1st class honor) from King Mongkut's Institute of Technology Ladkrabang, Chumphon Campus, Thailand (2010-2012, GPA: 3.76)

Dip. (Farm Mechanics) from Rajamangala University of technology Srivijaya, Thailand (2008-2009, GPA: 3.93)

Training attended:

- Visiting researcher in Biological and Agricultural Engineering, University of California, Davis, USA (4 months, 2019)
- Summer training program at Tokai University, Japan (1 month, 2012)

Scholarships:

- Royal Golden Jubilee scholarship (RGJ) PhD program from the Thailand Research Fund (TRF) (PHD/0102/2558)
- Full KMITL Educational Fund: Master Scholarship from King Mongkut's Institute of Technology Ladkrabang, Thailand.

Awards:

- Outstanding oral presentation award "Real-time evaluation system of the sugarcane soluble solids content on an elevator conveyor using visible and shortwave near-infrared spectroscopy" in RGJ-Ph.D. Congress 19 Innovation Challenges toward Thailand 4.0 organized by The Thailand Research Fund (2018)

- Best presentation award (Silver Medal) “Development of multi–fertilizer applicator prototype to achieve good nutrient management for oil palm plantation: Testing of centrifugal disc” in Agricultural Engineering projects competition organized by Thai Society of Agricultural Engineering (2012)

Oral presentation:

- **K. Phetpan**, V. Udompetiakul and P. Sirisomboon, “Real-time evaluation system of the sugarcane soluble solids content on an elevator conveyor using visible and shortwave near-infrared spectroscopy” RGJ-Ph.D. Congress 19, Jomtien Palm Beach Hotel & Resort, Pattaya, Thailand (2018).
- **K. Phetpan**, V. Udompetiakul and P. Sirisomboon, “A study of the influence of sugarcane variety on sugar content prediction using shortwave near-infrared spectroscopy” 9th Thai Society of Agricultural Engineering International Conference (TSAE2016), Thailand (2016).
- **K. Phetpan** and P. Sirisomboon, “Feasibility study for the evaluation of moisture content in tapioca starch cake by near infrared reflectance spectroscopy”, 7th Thai Society of Agricultural Engineering International Conference (TSAE2014), Thailand (2014).
- N. Ruttanadech, C. Yenphayep, N. Boontan, P. Nontarit and **K. Phetpan**, “Evaluation of multi- fertilizer applicator prototype attached to a tractor for oil palm plantation”, 7th Thai Society of Agricultural Engineering International Conference (TSAE2014), Thailand (2014).
- **K. Phetpan** and P. Sirisomboon, “Application of near infrared spectroscopy for detection of steroids adulteration in Traditional Thai Medicines”, 2nd International Conference on Agricultural and Food Engineering, CAFEi2014, Malaysia (2014).

International Published papers:

- **K. Phetpan**, V. Udompetiakul and P. Sirisomboon “Informative selection of spectra obtained from an online sugar content prediction system of sugarcane by using statistical index.” Thai Society of Agricultural Engineering Journal, 25(2), 19-27 (2019).

- **K. Phetpan**, V. Udompetaikul and P. Sirisomboon, “In-line near infrared spectroscopy for prediction of moisture content in tapioca starch drying process”, Powder Technology, 345, 608-615 (2019).
- **K. Phetpan**, V. Udompetaikul and P. Sirisomboon, “An online visible and near-infrared spectroscopic technique for the real-time evaluation of the soluble solids content of sugarcane billets on an elevator conveyor”, Computers and Electronics in Agriculture, 154, 460-466 (2018).
- P. Sirisomboon, K. Kaewsorn, S. Thanimkarn and **K. Phetpan**, “Non-linear viscoelastic behavior of cooked white, brown and germinated brown Thai jasmine rice by large deformation relaxation test”, International Journal of Food Properties, 20(7), (2017).
- **K. Phetpan** and P. Sirisomboon, “Evaluation of the moisture content of tapioca starch using near-infrared spectroscopy”, Journal of Innovative Optical Health, 8(2), 1550014 (2015).

Areas of expertise and interest:

- Applications in near infrared spectroscopy
- Non-destructive testing techniques for agricultural products and food
- Future of agriculture technology & Biosensor in food and agricultural sectors

Reference persons:

- Assoc. Prof. Dr. Panmanas Sirisomboon, Department of Agricultural Engineering, Faculty of Engineering, King Mongkut's Institute of Technology Ladkrabang, Thailand Email: panmanas.si@kmitl.ac.th
- Dr. Vasu Udompetaikul, Department of Agricultural Engineering, Faculty of Engineering, King Mongkut's Institute of Technology Ladkrabang, Thailand
  Email: vasu.ud@kmitl.ac.th