# FEASIBILITY FOR IN-LINE MEASUREMENT OF TAPIOCA STARCH MOISTURE CONTENT USING NEAR INFRARED SPECTROSCOPY

KITTISAK PHETPAN

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENT FOR THE DEGREE OF
MASTER OF ENGINEERING IN AGRICULTURAL ENGINEERING
FACULTY OF ENGINEERING
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG
2015

KMITL-2015-EN-M-100-042

# FEASIBILITY FOR IN-LINE MEASUREMENT OF TAPIOCA STARCH MOISTURE CONTENT USING NEAR INFRARED SPECTROSCOPY


KITTISAK PHETPAN



A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENT FOR THE DEGREE OF
MASTER OF ENGINEERING IN AGRICULTURAL ENGINEERING
FACULTY OF ENGINEERING
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG
2015

KMITL-2015-EN-M-100-042

| | |
|---|---|
| **หัวข้อวิทยานิพนธ์** | ความเป็นไปได้ในการวัดปริมาณความชื้นของแป้งมันสำปะหลังในสายการผลิตโดยใช้เนียร์อินฟราเรดสเปกโทรสโกปี |
| **นักศึกษา** | ว่าที่ ร.ต. กิตติศักดิ์ เพ็ชรพันธ์ |
| **รหัสประจำตัว** | 56601115 |
| **ปริญญา** | วิศวกรรมศาตรมหาบัณฑิต |
| **สาขาวิชา** | วิศวกรรมเกษตร |
| **พ.ศ.** | 2558 |
| **อาจารย์ที่ปรึกษาวิทยานิพนธ์** | รศ.ดร. ปานมนัส ศิริสมบูรณ์ |
| **อาจารย์ที่ปรึกษาวิทยานิพนธ์ (ร่วม)** | ดร. วสุ อุดมเพทายกุล |

## บทคัดย่อ


วัตถุประสงค์ของงานวิจัยนี้เพื่อศึกษาความเป็นได้ในการวัดปริมาณความชื้นของแป้งมันสำปะหลังในสายการผลิต (จุดสิ้นสุดของกระบวนการอบแห้ง) โดยใช้เนียร์อินฟราเรดสเปกโทรสโกปี การทดลองถูกแบ่งออกเป็น 3 ส่วน ประกอบด้วย การทดลองเบื้องต้น, การทดลองที่สายการผลิต (At-line) และการทดลองในสายการผลิต (In-line)

สำหรับตัวอย่าง-ของการทดลองเบื้องต้น ใช้แป้งปริมาณ 70 กรัมต่อตัวอย่างมาปรับระดับความชื้นที่แตกต่างกันโดยการเติมน้ำ โดยมีระดับความชื้นคือ 12.5, 20.63, 28.75 และ 36.88% wb การทดลองนี้ใช้เครื่อง diode array NIR spectrometer (DA7200, Perten, Sweden) ที่ช่วงความยาวคลื่นเท่ากับ 950-1650 นาโนเมตร ในการสแกนตัวอย่างแป้ง โดยที่ปริมาณความชื้นจริงของตัวอย่างถูกวัดโดยใช้เครื่อง infrared moisture analyzer และวิธีนี้ถูกใช้เป็นวิธีการวัดความชื้นอ้างอิงในทุกการทดลอง สมการความสัมพันธ์ระหว่างข้อมูลเชิงแสงกับค่าความชื้นถูกสร้างด้วยวิธี partial least squares regression แบบจำลองที่ดีที่สุดสำหรับตัวอย่างแป้งได้ค่าสัมประสิทธิ์การตัดสินใจ ($R^2$), ค่าความผิดพลาดมาตรฐานของการพิสูจน์แบบไขว้ (SECV), bias และค่า residual prediction deviation (RPD) เท่ากับ 0.997, 0.52%, -0.001% และ 16.8, ตามลำดับ

ตัวอย่างสำหรับการทดลองที่สายการผลิต (At-line) ถูกเก็บใน 2 รูปแบบ คือ ตัวอย่างแป้งหมาดที่จุดเริ่มต้นกระบวนการอบแห้ง และตัวอย่างแป้งแห้งที่จุดสิ้นสุดกระบวนการอบแห้ง การทดลองนี้ใช้เครื่อง diode array NIR spectrometer (DA7200, Perten, Sweden) ที่ช่วงความยาวคลื่นเท่ากับ 950-1650 นาโนเมตร ในการสแกนตัวอย่างแป้ง ข้อมูลเชิงแสงที่ใช้สำหรับการสร้างสมการประกอบไปด้วย 3 กลุ่มคือ ข้อมูลเชิงแสงที่ได้จากการสแกนแป้งหมาด, ข้อมูลเชิงแสงที่ได้จากการสแกนแป้งแห้ง และข้อมูลเชิงแสงที่รวมกันของแป้งหมาดและแป้งแห้ง แบบจำลองที่ดีที่สุดได้ค่า $R^2$, ค่าความผิดพลาดมาตรฐานของการพิสูจน์แบบภายนอก (SEP), bias และค่า residual prediction deviation (RPD) เท่ากับ 0.943, 0.41%, 0.069% และ 4.00, ตามลำดับ, สำหรับตัวอย่างแป้งหมาด; 0.928, 0.15%, -0.062% และ 4.00, ตามลำดับ, สำหรับตัวอย่างแป้งแห้ง;และ

0.996, 0.65%, 0.043% and 15.88, ตามลำดับ, สำหรับตัวอย่างแป้งที่รวมกันของแป้งหมาดและแป้งแห้ง

สำหรับการทดลองในสายการผลิต (In-line) ตัวอย่างแป้งมันสำปะหลังถูกเก็บทันทีหลังจากการสแกนด้วยเครื่อง diode array in-line NIR spectrometer (DA7300, Perten, Sweden) ที่ช่วงความยาวคลื่นเท่ากับ 950-1650 นาโนเมตร ซึ่งทำการติดตั้งในสายการผลิตที่จุดสิ้นสุดกระบวนการอบแห้ง การทดลองนี้ทำการสร้างสมการด้วยวิธี partial least squares regression ทั้งหมด 4 แบบคือ 1) การสร้างสมการโดยใช้เฉพาะข้อมูลการทดลองในสายการผลิต (In-line) และทำการพิสูจน์สมการแบบไขว้ (Cross-validation) ผลการสร้างสมการที่ดีที่สุดให้ค่า $R^2$, SECV, a bias และ RPD เท่ากับ 0.806, 0.54%, -0.014% and 2.00, ตามลำดับ 2) การสร้างสมการโดยใช้เฉพาะข้อมูลการทดลองในสายการผลิต (In-line) และทำการพิสูจน์สมการแบบภายนอก (External-validation) ผลการสร้างสมการที่ดีที่สุดให้ค่า $R^2$, SEP, bias และ RPD เท่ากับ 0.641, 0.62%, -0.069% and 1.68, ตามลำดับ 3) การสร้างสมการโดยใช้เฉพาะข้อมูลการทดลองที่สายการผลิต (At-line) และใช้ข้อมูลการทดลองในสายการผลิต (In-line) เป็นชุดพิสูจน์สมการ แบบจำลองที่ดีที่สุดให้ค่า $R^2$, SEP, bias and RPD เท่ากับ 0.667, 0.62%, -0.086% และ 1.74, ตามลำดับ 4) การสร้างสมการโดยใช้ข้อมูลทั้งหมดของการทดลองที่สายการผลิต (100% ของ at-line) และข้อมูลครึ่งหนึ่งของการทดลองในสายการผลิต (50% ของ in-line) โดยใช้ข้อมูลการทดลองในสายการผลิตที่เหลืออีกครึ่งหนึ่งเป็นชุดพิสูจน์สมการ ผลการสร้างสมการที่ดีที่สุดให้ค่า $R^2$, SEP, bias และ RPD เท่ากับ 0.658, 0.61%, 0.001% และ 1.70, ตามลำดับ

โดยสรุป สมการที่ดีที่สุดที่ได้ใช้ข้อมูลการทดลองที่สายการผลิต (At-line) และใช้ข้อมูลจากตัวอย่างแป้งแห้งสามารถใช้เป็นทางเลือกสำหรับการประเมินปริมาณความชื้นของแป้งมันสำปะหลังอย่างรวดเร็วในห้องปฏิบัติการควบคุมคุณภาพของโรงงาน อย่างไรก็ตาม สมการที่ใช้ข้อมูลการทดลองในสายการผลิต (In-line) เพื่อทำนายปริมาณความชื้นแป้งมันสำปะหลังสามารถใช้ได้ดีสำหรับการตรวจคัดแยกแบบหยาบ (Rough screening) สิ่งที่จำเป็นสำหรับงานวิจัยต่อไป คือ การปรับปรุงวิธีการสแกนตัวอย่าง โดยเฉพาะอย่างยิ่งการเลือกตำแหน่งการติดตั้งสเปกโตรมิเตอร์ที่สามารถสแกนตัวอย่างได้โดยมีระยะการเคลื่อนที่ของแสง (Path range) ของการสแกนแต่ละครั้งใกล้เคียงกัน

| | |
|---|---|
| **Thesis** | Feasibility for in-line measurement of tapioca starch moisture content using near infrared spectroscopy |
| **Student** | Acting Sub Lt. Kittisak Phetpan |
| **Student ID.** | 56601115 |
| **Degree** | Master of Engineering |
| **Program** | Agricultural Engineering |
| **Year** | 2015 |
| **Thesis Advisor** | Assoc. Prof. Dr. Panmanas Sirisomboon |
| **Thesis Co-Advisor** | Dr. Vasu Udompetaikul |


## ABSTRACT

The objective of this research was to study the feasibility for in-line measurement of tapioca starch moisture content at the end of the drying process by near infrared (NIR) spectroscopy technique. The experiments were divided into three parts including preliminary experiment, at-line experiment and in-line experiment.

In the preliminary experiment, the samples (70 g each) were adjusted for different levels of moisture content (12.5, 20.63, 28.75 and 36.88% wb) by mixing with distilled water and scanned using diode array NIR spectrometer (DA7200, Perten, Sweden) in wavelength range of 950-1650 nm. The actual moisture content was measured by an infrared moisture analyzer and this method was used as reference method for all experiments. The partial least squares (PLS) regression was used for the NIR spectroscopic models establishment for predicting the moisture content tapioca starch. The optimum models for starch samples provided the coefficient of determination ($R^2$), standard error of cross-validation (SECV), a bias and residual prediction deviation (RPD) of 0.997, 0.52%, -0.001% and 16.8, respectively.

In the at-line experiment, the tapioca starch samples were collected (tapioca starch cake at the inlet of the drying process and dried tapioca starch at the end) and also scanned using a diode array NIR spectrometer (DA7200, Perten, Sweden) in wavelength range of 950-1650 nm. Three groups of sample spectra were used in model development: those of tapioca starch cake samples, dried tapioca starch samples and combined samples (cake and dried samples were pooled). The optimum PLS regression models provided an $R^2$ value, standard error of prediction (SEP), bias and RPD of 0.943, 0.41%, 0.069% and 4.00, respectively, for starch cake

samples; 0.928, 0.15%, -0.062% and 4.00, respectively, for dried tapioca starch samples; 0.996, 0.65%, 0.043% and 15.88, respectively, for combined samples.

In the in-line experiment, the NIR scanned tapioca starch samples were collected immediately after scanning at the end of the drying process. It was scanned using a diode array in-line NIR spectrometer (DA7300, Perten, Sweden) in wavelength range of 950-1650 nm. The PLS regression models of four different types were as follows: 1) the NIR model establishment using in-line data with internal validation (Cross-validation) provided the optimum model with $R^2$ value, SECV, a bias and RPD of 0.806, 0.54%, -0.014% and 2.00, respectively; 2) the NIR model establishment using in-line data with external validation (Test set validation) provided the optimum model with $R^2$ value, SEP, bias and RPD of 0.641, 0.62%, -0.069% and 1.68, respectively; 3) the NIR model establishment using at-line data as calibration set and in-line data as test set provided the optimum model with $R^2$ value, SEP, bias and RPD of 0.667, 0.62%, -0.086% and 1.74, respectively; 4) the NIR model establishment using the calibration set consisted of 100% of at-line data and 50% of in-line data and test set consisted of another 50% of the in-line data provided the optimum model with $R^2$ value, SEP, bias and RPD of 0.658, 0.61%, 0.001% and 1.70, respectively.

In conclusion, the best NIR model used at-line data and developed from dried starch samples could be used as a rapid alternative to evaluate the moisture content of tapioca starch in factory quality control laboratory. However, the NIR models used in-line data for predicting the tapioca starch moisture content were good for rough screening. Therefore the further research for the improvement of the method is needed.

# Acknowledgements

I wish to express my appreciation and gratitude to many individuals for their assistance and contributions towards the success of this research. First of all, I would like to express my thanks to my advisers, Assoc. Prof. Dr. Panmanas Sirisomboon and Dr. Vasu Udompetaikul for their valuable time, advices as well as their encouragement throughout the study period which enabled the completion of this thesis.

I would like to thank Sangpetch Tapioca Flour Co., Ltd. in Nongbua-rawe district, Chaiyaphum, Thailand, especially to Mr. Siripong Soontreerat (คุณ ศิริพงษ์ สุนทรีรัตน์) and his son, Mr. Siwa Soontreerat (คุณ ศิวะ สุนทรีรัตน์), for providing samples and the experiment station, and thank Charpa Techcenter Co., Ltd. Thailand and Perten Insruments, Sweden, for supporting the NIR spectrometers.

I am very grateful to all my friends in curriculum of Agricultural Engineering at King Mongkut's Institute of Technology Ladkrabang for their supports and encouragement. Moreover, I would like to thank Faculty of Engineering, KMITL for supporting the tuition fee and thank "KMITL Educational Fund: Master Scholarship Academic Year 2013" for supporting monthly living expenses as well as transportation costs for the research paper presentation outside the university.

Finally, I would like to dedicate this thesis to my family and teachers. I have been extremely fortunate in my life to have family and teachers, they have shown me unconditional love and support and have taught me a great deal about the aging process and growing old gracefully. I would like to show appreciation of kindness received from them. Without their support and encouragement, I could not have completed this thesis.

Kittisak Phetpan

# Table of contents

# Table of contents (continued)

# Table of contents (continued)

# Table of contents (continued)

# List of tables

# List of tables (continued)

# List of figures

# List of figures (continued)

# List of figures (continued)

**Figure** **Page**

# List of figures (continued)

**Figure** **Page**

# Chapter 1
# Introduction

## 1.1 Introduction to the research problems and its significance

Thailand has been one of the top tapioca product exporters. The tapioca starch industry is an economically important industry in Thailand. From the total export value of all tapioca products, about one-third is from tapioca starch [1]. In 2014, Thailand exported approximately 3,012 million tons of tapioca starch, corresponding to about 1,259 million USD or 41,053 million THB [2].

One of the important processes in tapioca starch production is the drying process. Pneumatic conveying dryers have been widely used in tapioca starch factories. In the drying process [3], water in the tapioca starch cake was evaporated using hot air and the dried tapioca starch was conveyed to a cyclone separator, where the dried tapioca starch and saturated air were separated. Moisture content of the finished product (dried starch) can be controlled by varying the feed rate of the starch cake at the inlet of the system. At present, most tapioca starch factories use air temperature at the outlet of the drying process as the parameter to control the starch cake feed rate. When the outlet air temperature is high, indicating that the cake feed rate is too low, the controller will increase the speed of the feeder. The maximum allowable temperature is 60°C. If the outlet temperature reaches 60°C, the controller will stop the feeder. In addition, moisture content of the dried tapioca starch is regularly checked during the drying process using an infrared moisture analyzer, which requires approximately 10 min per sample. Whenever the moisture content is found exceeding the acceptable standard value (13% wb), the dried starch from that lot must be sent to the drying process again. Thus, the starch production time is approximately 30 min (20 min of drying tube retention and 10 min of moisture content measurement). For the factory with production capacity of 200 tons per 12 hr, typically, 8 tons of those must be repeated in the drying process. This is very expensive, with costs including the drying energy, product unpacking and repacking, material handling, and labor costs. This does not include the time required to re-adjust the system to meet the proper conditions, which must be performed by

experienced workers. Based on discussions with the factory production engineers, the use of the moisture content of the dried starch as a direct control parameter of the starch cake feeder would solve this problem. However, there is no existing sensors that can be installed to measure directly the final starch moisture content in tapioca starch factories.

Near-infrared (NIR) spectroscopy is a rapid method for chemical component analysis. Its advantages include its high precision and accuracy, non-destructive measurements, chemical-free procedure and environmental-friendliness as well as no or minimal sample preparation. However, this technique needs a calibration model to be developed for prediction, which is calibrated with a standard reference method. If the validation of its accuracy is proved, analysis takes only 2-3 s.

There were some researches that used NIR for measuring the moisture content of agricultural products. Vesela et al. [4] reported a calibration model developed from the NIR absorbance spectra (1100-2500 nm) of cocoa powder samples, which was used for the determination of the moisture content. The results showed that the relative root mean square error of cross-validation (RMSECV) was 5.2% and the determination correlation ($R^2$) was 0.94. Camps et al. [5] developed a NIR spectroscopy method to determine the moisture in flour of dry *Artemisia annua* leaves, and the model accurately predicted moisture with $R^2$, RMSECV and root mean square error of prediction (RMSEP) of 0.99, 0.8% and 1.4%, respectively. Recently, Phoonphatthanachai and Sirisomboon [6] reported the feasibility study of using FT-NIR spectrometer, micro-NIR with Linear Variable Filter (LVF), and portable VIS-NIR spectrometer in determination of moisture content of tapioca starch cake and the best model from FT-NIR spectrometer provided $R^2$ and RMSECV of 0.99 and 0.755%, respectively.

These confirm that the NIR technique could be used to precisely measure the agricultural products moisture content. From reviewed studies, there are still no reports on the determination of tapioca starch moisture content using in-line NIR spectroscopy especially the in-line NIR spectrometer that is installed at the end of the drying process to measure or monitor directly the tapioca starch moisture content. Therefore, from the problems as mentioned above, the in-line NIR spectroscopy should be studied as a rapid and powerful alternative to evaluate the

tapioca starch moisture content at the end of the drying process for tapioca starch industries.

## 1.2 Objective

The objectives of this study were: 1) to conduct the preliminary study for the evaluation of moisture content of tapioca starch using NIR spectroscopy; 2) to evaluate the moisture content of tapioca starch using NIR spectroscopy; and 3) to study the feasibility for the evaluation of moisture content of tapioca starch using in-line NIR spectroscopy.

The final goal of the research was to study the feasibility for in-line measurement of tapioca starch moisture content at the end of the drying process using near infrared spectroscopy at wavelength range of 950-1650 nm.

## 1.3 Scope of research

1.3.1 The tapioca starch samples were only collected at the Sangpetch Tapioca Flour Co., Ltd. factory in Nongbua-rawe district, Chaiyaphum, Thailand.

1.3.2 The diode array NIR spectrometer (DA7200, Perten, Sweden) and diode array in-line NIR spectrometer (DA7300, Perten, Sweden) in reflection mode of 950-1650 nm were used for scanning the samples.

1.3.3 The infrared moisture analyzer (HB43-S Halogen, Mettler Toledo, Switzerland) was used as the reference method for measuring the tapioca starch moisture content.

## 1.4 Expected results

1.4.1 The NIR models using at-line and/or in-line spectra data for accurately predicting the tapioca starch moisture content are obtained.

1.4.2 Knowledge of feasibility for in-line measurement of tapioca starch moisture content using NIR spectroscopy are obtained to distribute to the tapioca starch factories.

## 1.5 Experimental plan

The experiments in this study were divided into three parts including preliminary experiment, at-line experiment and in-line experiment and were briefly summarized in figure 1.1, 1.2 and 1.3, respectively.

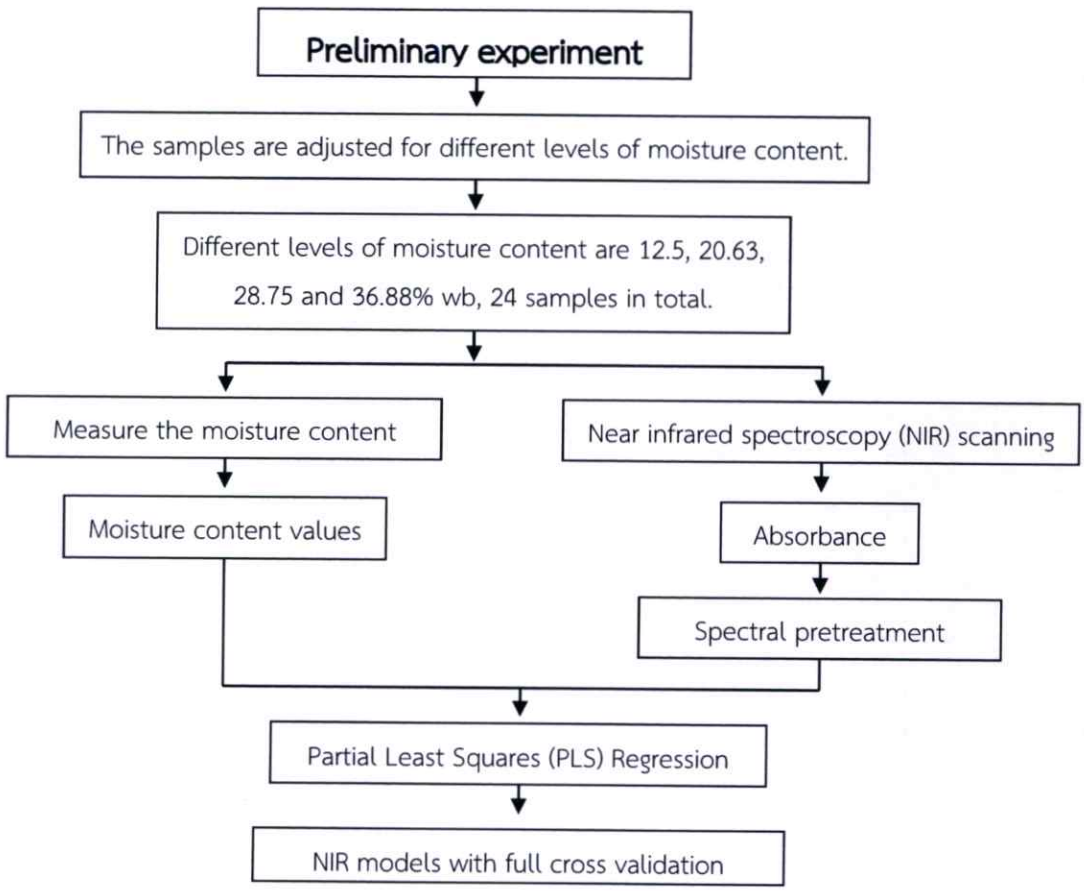



**Figure 1.1** Preliminary experiment plan

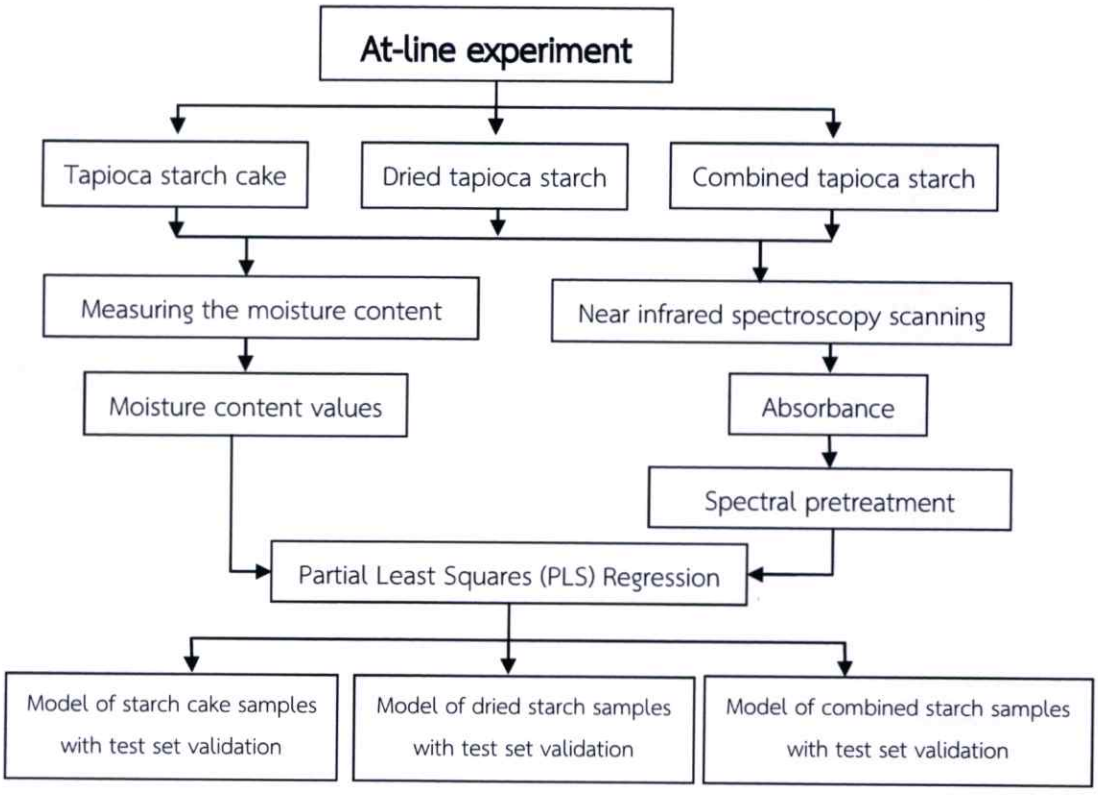
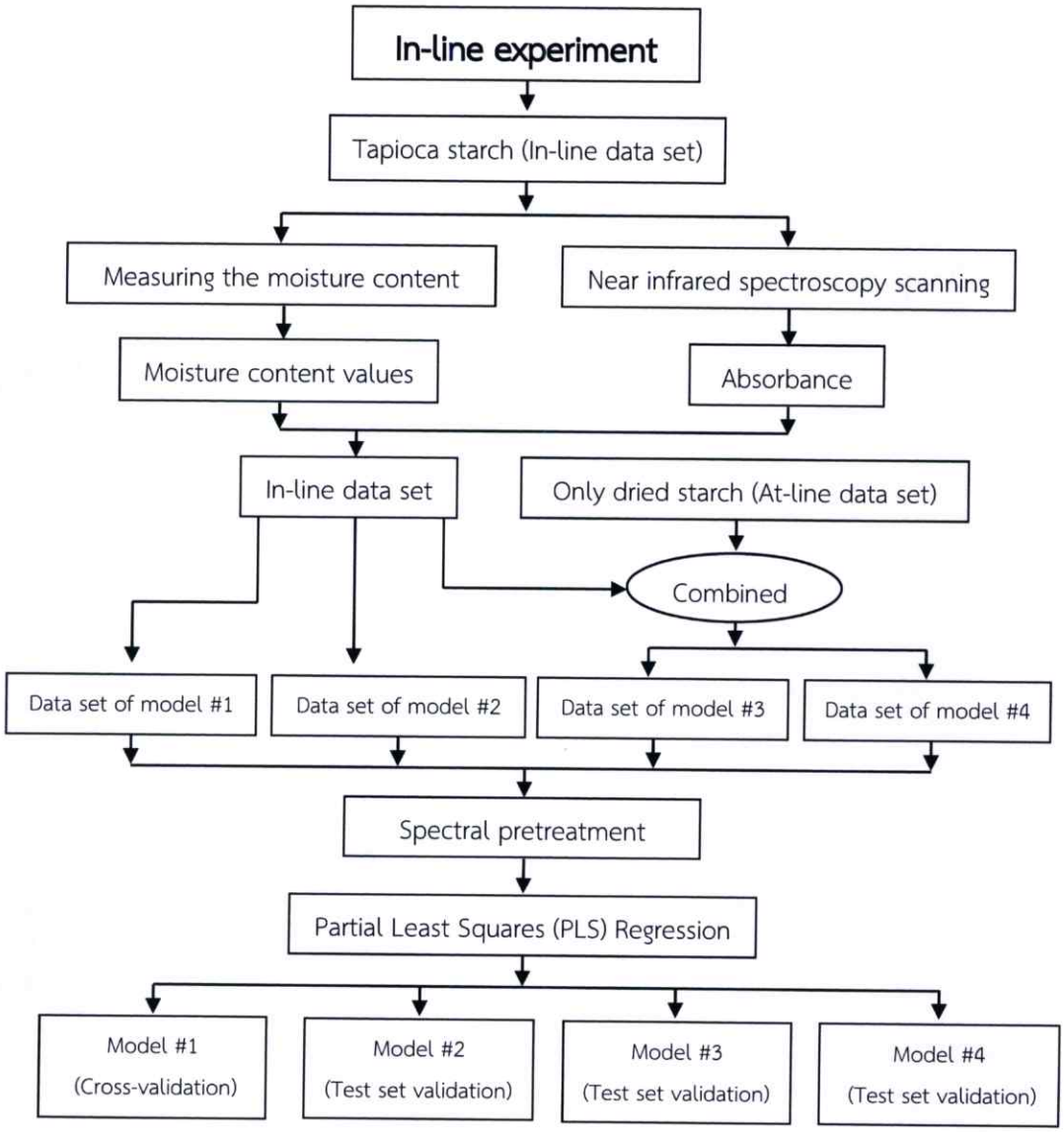



**Figure 1.2** At-line experiment plan




Note: Model #1 used in-line data and internal validation (Cross-validation).

Model #2 used in-line data and external validation (Test set validation).

Model #3 used at-line data as calibration set and in-line data as test set.

Model #4 used calibration set consisted of 100% of at-line data and 50% of in-line data and test set consisted of another 50% of the inline data.

**Figure 1.3** In-line experiment plan

# Chapter 2

# Theory and literature review

## 2.1 Thai tapioca

The tapioca is an important economical crop in Thailand. In 2014, Thailand has an approximately 1.4 million hectare (8.8 million rai) for tapioca plantation in 2014/2015 (Table 2.1) [7] and exported approximately 9.8 million tons of tapioca products (Table 2.2) which was 2,787 million USD (Table 2.3) [2]. Specially, the exported product, which is tapioca starch, was 3.0 million tons valued at 1,260 million USD.

**Table 2.1** Tapioca plantation areas (rai) in Thailand [7]

| Region/Province | Tapioca starch plantation areas (rai) | | |
|---|---|---|---|
| | 2012/2013 | 2013/2014 | 2014/2015 |
| Northern | 1,549,938 | 1,876,311 | 1,894,534 |
| Northeast | 4,366,997 | 4,493,264 | 4,627,719 |
| Central | 1,988,121 | 2,287,367 | 2,313,326 |
| **Total All Region** | **7,905,056** | **8,656,942** | **8,835,576** |

**Table 2.2** Export statistics of tapioca products by quantity in 2010-1014 [2]

| Category | Quantity (tons) | | | | |
|---|---|---|---|---|---|
| | 2010 | 2011 | 2012 | 2013 | 2014 |
| Tapioca starch | 1,740,806 | 1,891,343 | 2,235,574 | 2,445,612 | 3,012,111 |
| Starch chips | 4,116,726 | 3,693,514 | 4,611,976 | 5,755,376 | 6,777,097 |
| Starch pellets | 156,069 | 36,694 | 84,215 | 59,082 | 23,054 |
| Sogo | 25,006 | 30,893 | 23,540 | 27,005 | 28,061 |
| **Total** | **6,038,607** | **5,652,444** | **6,955,305** | **8,287,075** | **9,840,323** |

**Table 2.3** Export statistics of tapioca products by value in 2010-2014 [2]

| Category | Value (USD) | | | | |
|---|---|---|---|---|---|
| | 2010 | 2011 | 2012 | 2013 | 2014 |
| Tapioca starch | 753,727,600 | 866,861,108 | 945,399,463 | 1,070,744,478 | 1,260,271,814 |
| Starch chips | 773,369,062 | 897,991,429 | 1,020,389,442 | 1,213,059,620 | 1,500,317,995 |
| Starch pellets | 24,101,579 | 8,715,157 | 17,725,195 | 12,781,555 | 4,805,001 |
| Sogo | 17,218,061 | 20,745,874 | 17,830,524 | 19,527,075 | 21,959,589 |
| **Total** | **1,568,416,302** | **1,794,313,568** | **2,001,344,624** | **2,316,112,728** | **2,787,354,399** |

## 2.2 Tapioca starch industry

Compared to other tapioca producers of the world e.g. Brazil, Nigeria and Indonesia, Thailand is the only country that uses the most of the tapioca in tapioca starch industry [8]. Thus, Thailand has become the largest tapioca starch manufacturers in the world (see section 1.1).

### 2.2.1 Flour and starch

Many people have been confused about the meaning of flour and starch. There are several meanings defined in dictionaries. For example, flour is the powder made from grain, especially wheat, used for making bread, cakes, pasta, pastry, etc whereas starch is a white substance that exists in large amounts in potatoes and particular grains such as rice [9]. Similarly, flour is the powder obtained by grinding grain, typically wheat, and used to make bread, cakes, and pastry, whereas starch is an odourless, tasteless white substance occurring widely in plant tissue and obtained chiefly from cereals and potatoes, and is also a polysaccharide which functions as a carbohydrate store [10].

However, the difference between flour and starch have been more clearly described in Food focus Thailand [11] as following ways; flour is the products which are made from raw agricultural tuber plants or cereals. It is cleaned and ground. Thus, it contains the original constituents from raw materials such as carbohydrate, protein, fat and etc, whereas starch is products which are also made from agricultural tuber plants but carbohydrate is extracted by the processes, then it is dried and extracted other constituents. Therefore, starch contains most of carbohydrate.

### 2.2.2 Standards of tapioca starch

However, the standards of tapioca starch in Thailand were defined in Ministry of Commerce notification "Prescribing tapioca starch as a standardized commodity and the standards of tapioca starch, dated 29th September, 2006" which has given the involved details as follows [12]:

2.2.2.1 Definitions:

1. "Tapioca starch" means the starch, which is the processed product of tapioca roots, appears in white and pale yellow.

2. "Tapioca modified starch" means the starch, which is the product of tapioca starch, has undergone a modification of its physical and/or chemical characteristic by applying heat and/or enzymes and/or other chemical substances in order to prepare the starch for appropriate uses.

2.2.2.2 The standards of tapioca starch shall be divided into 3 classes including 1) Premium grade tapioca starch, 2) First grade tapioca starch and 3) Second grade tapioca starch. The details could be seen in Table 2.4.

**Table 2.4** Required features of tapioca starch [12]

| Features | Premium grade | First grade | Second grade |
|---|---|---|---|
| Moisture (% by weight) not exceeding | 13 | 14 | 14 |
| Starch (% by weight) not less than | 85 | 83 | 80 |
| Ashes (% by weight) not exceeding | 0.2 | 0.3 | 0.5 |
| Fiber (Cubic centimeters per 50 g starch) not exceeding | 0.2 | 0.5 | 1.0 |
| pH | 4.5 to 7.0 | 4.5 to 7.0 | 4.5 to 7.0 |
| Starch (% by weight) shall be able to pass through 150 micro-meter sieve at not less than | 95 | 95 | 95 |

### 2.2.3 Usage of tapioca starch

Tapioca can be grouped as the economical crop in Thailand (see Table 2.2 and 2.3). In this case the tapioca starch, which is the product produced from tapioca roots, was mentioned about the various usages. According to report [8], tapioca starch is not currently produced only for human consumption such as in food and beverage industry, sweetener industry but It is also used as a raw material for products in other industries such as, textile industry, paper industry, glue industry, plywood industry, medical industry and so on.

### 2.2.4 The process of tapioca starch production

In Thailand small-scale factories have been replaced by the large-scale tapioca starch production that includes advanced technologies, according to report [13], the processes of production were divided into two ways i.e. traditional and modern process. Their details and procedures are as follows:

2.2.4.1 The traditional process

The traditional process is usually found in the small-scale factories. The tapioca starch was produced by this process is low quality. The procedure in this process has 9 stages as follows:

1. Weight the fresh tapioca for determination of starch percentage.
2. Chop out the roots and stems connected along with tapioca bulbs before putting in the peeling devices.
3. Remove the peel and put them in the cleansing well.
4. Put the clean ones in the chopper and grinder. Grind until they becomes fine particles and sediment them in the water. In this step, they appear as the slurry mix of starch, water, and fiber.
5. Filter the slurry to separate starch slurry from the fiber, however, the fiber still has some starch content that can be used as animal food.
6. Silt up the starch in the wooden pail.
7. Dry the moist starch to the sun light on the yard.
8. Use the grindstone to grind the completely dried starch.
9. Pack in the bag.

2.2.4.2 The modern process

The modern process is found in the large-and medium-scale factories. This process requires only short period of time and provides the higher quality of starch. The process has just 8 stages as follows:

1. Weight the fresh tapioca for determination of the starch percentage.
2. Remove sand and impurity in the rotary screener.
3. Put it into the peeling and cleansing device.
4. Put the clean tapioca into the grindstone device and then the separating device to extract the protein out.
5. Pass the slurry through the screener to remove fiber.
6. Separate the fine fiber and impurity by "Centrifuge". The starch cake is obtained.
7. Dry out the starch cake by passing it through the hot-aired dryer.
8. Pass the starch through the sifter and pack the fine powder into the sacks for sale.

Since, the tapioca starch factory in this research used the modern process. Thus, the flow chart of tapioca starch production by modern process (Figure 2.1) is shown for better understanding.

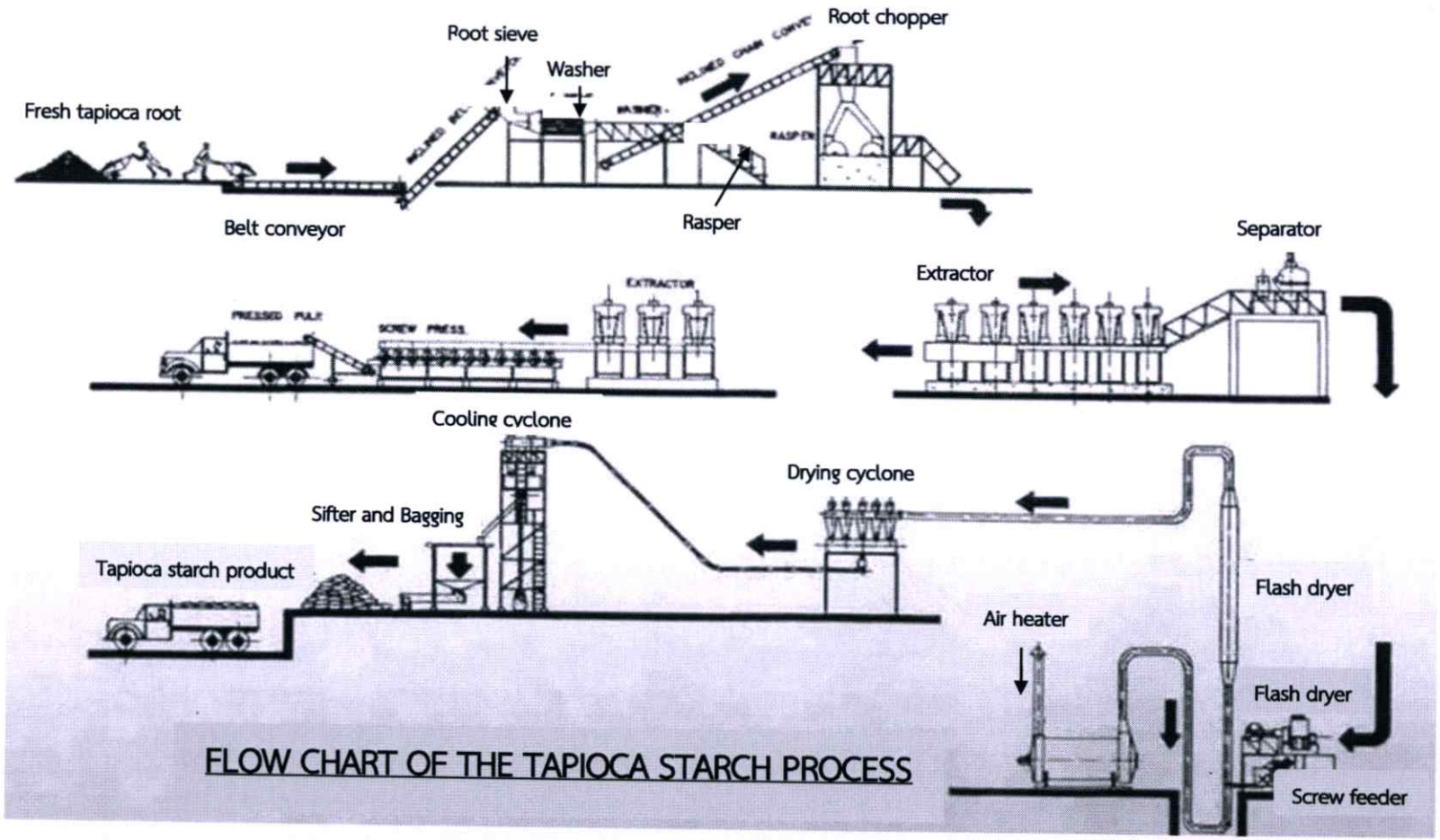



**Figure 2.1** Flow chart of tapioca starch production by modern process [13]

## 2.3 Determination of moisture content

Since, tapioca starch moisture content is one parameter used to evaluate the standard of tapioca starch class (see section 2.2.2.2). Of course, the standard method is acceptable but it spends a lot of times for evaluating the moisture content. Nowadays, most of tapioca starch factories rely on the infrared moisture analyzer as the reference method.

### 2.3.1 Standard method of moisture content determination

The standard method of tapioca starch moisture content determination was defined in Thai Industrial Standards "Tapioca starch" that Thai Industrial Standards Institute [14] has given the details as follows:

2.3.1.1 Method

1. Dry the empty aluminium dish and lid in the oven at 105 to 107 °C for 15 minute and transfer to desiccator to cool. Then, weight the empty dish and lid.
2. Weigh about 5 g of sample into the dish.
3. Place the dish with sample in the oven. Dry for 5 h at 105 to 107 °C.
4. After drying, transfer the dish with lid to the desiccator to cool. Reweigh the dish with its dried sample.
5. Repeat drying for every 30 minutes until the least difference of initial and final weight equal to 2 mg.

2.3.1.2 Calculation

$$\text{Moisture (\%)} = \frac{100(W_1 - W_2)}{W_1 - W} \tag{2.1}$$

Where the $W$ is the empty aluminum dish with lid weight (g), $W_1$ is the aluminum dish with initial sample weight (g), and $W_2$ is the aluminum dish with final sample weight (g).

### 2.3.2 Infrared moisture analyzer

The infrared moisture analyzer works on the thermogravimetric principle [15]. At the start of the measurement the "infrared moisture analyzer" determines the

weight of the sample. It is then quickly heated by the integrated halogen heating module to vaporize the moisture. During the drying process the weight of the sample is continually measured by the instrument and displays the reduction in moisture. Once the drying process has been completed, the moisture content of the sample is displayed as the final result.

## 2.4 Basic principles of near-infrared (NIR)

Infrared energy is the electromagnetic energy of molecular vibration that energy band defined for convenience as the NIR is 780 to 2,500 nm, 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 [16]. However, the NIR region of the electromagnetic spectrum were 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 $cm^{-1}$, it related to overtones and combinations of fundamental vibrations of C–H, N–H, O–H (and S–H) functional groups as well [17; 18; 19]. Moreover, according to reviews the principle of vibrational spectroscopy has interestingly been mentioned by Burns and Ciurczak [18] that it is based on the concept that atom-to-atom bonds within molecules vibrate with frequencies. When these molecular vibrators absorb light of particular frequency, they are excited to a higher energy level. At ambient temperature, most molecules are at their rest or zero energy level.

### 2.4.1 Physics of the interaction of radiation with matter

The interaction of radiation with matter depends on various factors such as sample surface, particle size within the sample and so on. However, Osborne et al. [20] explained the incident radiation using the law of conservation of energy that when monochromatic radiation interacts with a sample it may be absorbed, transmitted or reflected (Figure 2.2). Hence, the total radiant intensity incident on the sample $(I_0)$ must equal the sum of intensity absorbed $(I_A)$, transmitted $(I_T)$ and reflected $(I_R)$.

$$I_0 = I_A + I_T + I_R \tag{2.2}$$

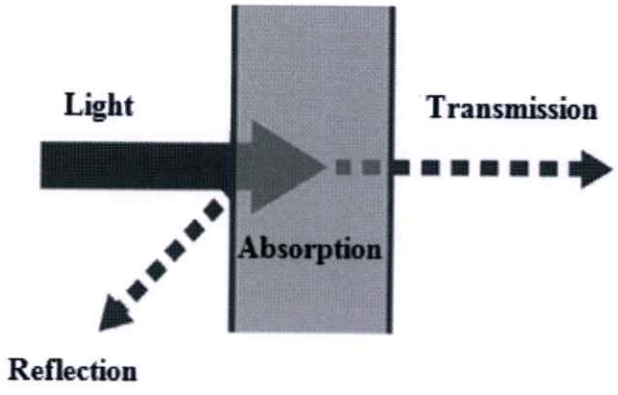



**Figure 2.2** Interaction of radiation with matter

### 2.4.2 Absorption of radiation

If consider the equation 2.2 by focusing only absorbed intensity in order to measure either the transmitted or reflected intensity. Thus this means that absorption of radiation by sample involves the loss of energy by the radiation. However Osborne et al. [20] mentioned the Beer-Lambert law that it relates 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 is proportional to the number of molecules $dn$ in that thickness, or

$$-dI/I = kdn \tag{2.3}$$

Take integrate into equation 2.3

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

$$-\ln(I_T - I_0) = k(n-0)$$

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

$$\ln(I_0 / I_T) = kn \tag{2.4}$$

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

$$\ln \frac{1}{T} = kn \tag{2.5}$$

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

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

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

$$A = \log \frac{1}{T} = kn \tag{2.6}$$

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

$$A = \log \frac{1}{T} = \log(I_0 / I_T) = abc \tag{2.7}$$

and equation 2.7 becomes

$$\mathrm{A = abc} \tag{2.8}$$

### 2.4.3 Reflection of radiation

The reflective behaviors of the incident radiation on the samples depend on the various factors such as the matter surface, different particles within the sample. Anyway the figure 2.3 shows the reflective behavior of both smooth and rough surface. Osborne et al. [20] reported that sample which is opaque and non-absorbing will reflect in a manner which is similar to the law of mirror whereas, if the sample surface is rough, the beam of light will reflect and diffuse in many different directions.

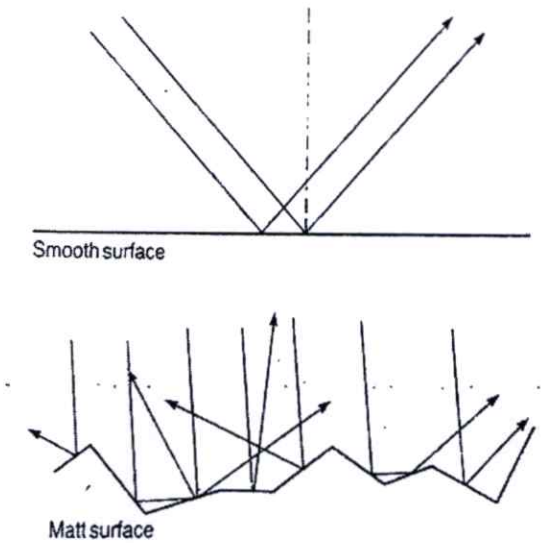



**Figure 2.3** Surface effect-specular reflectance [20]

Moreover there are some cases that the radiation can be transmitted through the sample surface and then it is absorbed according to Beer Lambert law, especially, it can be found in solid sample such as powder, rice and so on. These behaviors were also explained by Osborne et al. [20] that the radiation faces the discrete particles within the sample and radiation disseminates in all directions; this phenomenon was discovered by Tyndall in 1869 and is known as scattering (Figure 2.4).

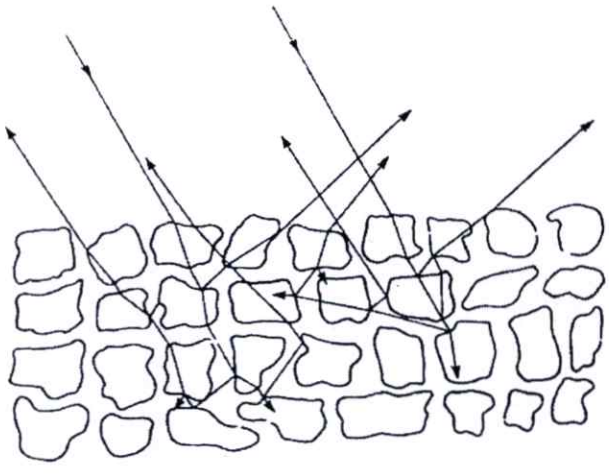


**Figure 2.4** Diffuse (body) reflectance [20]

Since the scattering effect is difficult to describe mathematically for particulate samples. However, Kubelka and Munk [20; 21; 22] founded that the absorbance of layer of infinite thickness, which is completely opaque, depend on its scattering and absorption as the following equation:

$$\frac{(1-R\infty)^2}{2R\infty} = \frac{k}{s} \tag{2.9}$$

where $R\infty$ is the reflectance of the infinitely thick layer, $s$ is the scattering constant and $k$ is the absorption constant. According to the Beer-Lambert law [20; 23; 24], the absorption coefficient $k$ in equation 2.7 is actually equal to concentration ($c$) multiplied by the absorptivity ($a$) as below:

$$\frac{(1-R\infty)^2}{2R\infty} = \frac{ac}{s} \tag{2.10}$$

In practical, Norris [20; 25] 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:

$$\log(R'/R) = \log 1/R + \log R' \sim ac/s \tag{2.11}$$

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

$$\log 1/R \sim \frac{ac}{s} \tag{2.12}$$

### 2.4.4 Near-infrared (NIR) instrumentation

A NIR spectrometer is generally composed of a light source, a monochromator, a sample holder or a sample presentation interface, and a detector, allowing for transmittance or reflectance measurements (Figure 2.5) [17].

The light source is usually a tungsten halogen lamp, while detector types include silicon, lead sulfide (PbS) and indium gallium arsenide (InGaAs) [17; 26]. However, the special properties of these types have interestingly been reported by Reich [17] as follows: Silicon detectors are fast, low noise, small and highly sensitive from the visible region to 1100 nm; PbS detectors are slower, but very popular since they are sensitive from 1100 to 2500 nm and provide good signal-to-noise properties. The advantages of both types, which are speed and size characteristics of silicon detector and wavelength range of PbS detector, were combined to become the InGaAs detector.

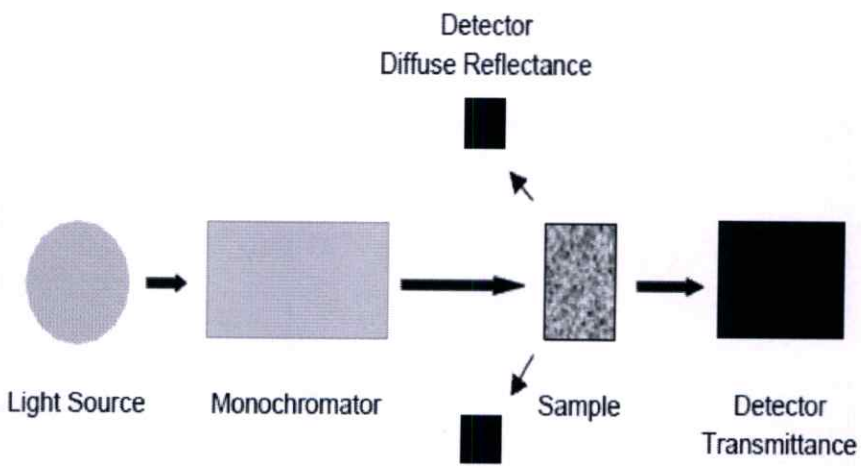



**Figure 2.5** Basic NIR spectrometer configurations [17]

Currently, various technologies are used to separate the polychromatic NIR light into monochromatic light for both qualitative and quantitative purposes, including Filter photometer, diffraction grating, interferometer, diode-array and acousto-optic tunable filter (AOTF) [27, 28]. The basic of each technique were described as following ways:

2.4.4.1 Filter spectrophotometer

The concept of the filter-type NIR spectrophotometers has clearly been described by Workman and Burns [29], that it utilized interference filter wedge by allowing incident angle of the light pass through itself is the transmitted intensity and band-passes. However, the transmitted intensity at varying wavelength still depends on incident angle.

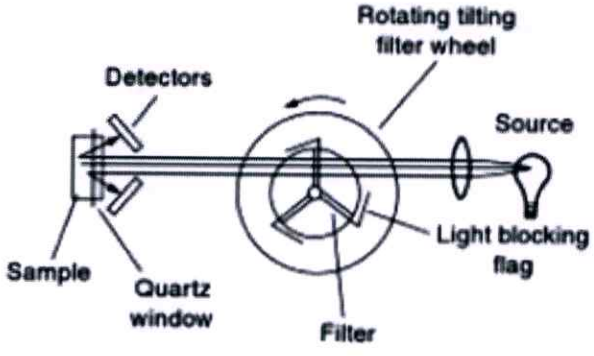



**Figure 2.6** Two types of tilting filter instrument design [29]

2.4.4.2 Diffraction grating

Grating is device that used to separate the light into the various wavelengths prior to striking the detector [29]. It is covered the parallel grooves at very close intervals on the top [20]. However, Loewen and Popov [30] described the diffraction grating concepts as following ways: when light is incident on a grating it is diffracted from the grooves that becomes a small source of reflected and/or transmitted light (Figure 2.7). Thus only difference between angle $\theta_i$ and $\theta_d$ equals the wavelength of light, the light will be phase.

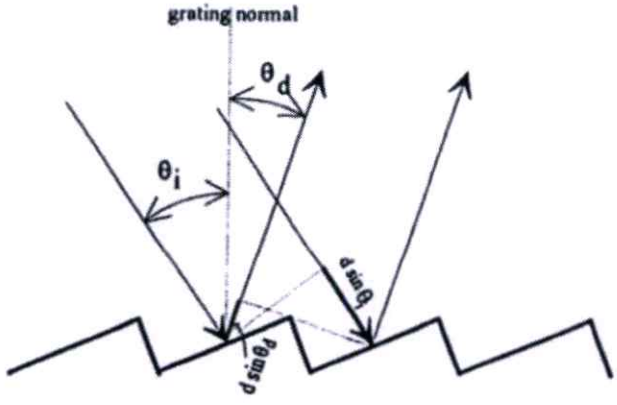



**Figure 2.7** Diagram for phase relation between the rays diffracted from adjacent grooves [30]

2.4.4.3 Interferometer or Fourier transform (FT) NIR spectrophotometer

Interferometer or FT-NIR spectrometer operates by applying Fourier transform concept to interferograms that obtained from a Michelson interferometer with a movable mirror [27]. McCarthy and Kemeny [29] mentioned about the Fourier transform spectrometers that two mutually perpendicular plane mirrors is the simplest form of an interferometer (Figure 2.8). Moreover, they understandably explained the principle of FT-NIR spectrometer as follows: "The beam-splitter partially reflects the light source to the fixed mirror and reflects the remaining light to the moving mirror. The beams reflected from the mirror are recombined by the beam-splitter and directed to the detector. The optical retardation is equal to twice the moving mirror displacement. When the fixed and moving mirror is equidistant, the retardation is zero for all frequencies and two beams interfere constructively. Therefore, all light source reaches the detector at this point. The variation in intensity as the moving mirror is displaced contains the spectral information that is retrieved by a FT".

2.4.4.4 Diode-array spectrometers

Polychromatic light is sent to a monochromator that comprises mirrors, slits and a grating (Figure 2.9) [31]. Moreover, Choi [32] described the diode array that polychromatic beam passed through the sample is irradiated onto the inlet slit of the polychromator. Hence, the polychromator disperses the narrow band of the spectrum onto the diode array. Then the photodiode will convert light into electrical

signals and temporarily stores them. These signals are then read out as time-series signals via the output line.

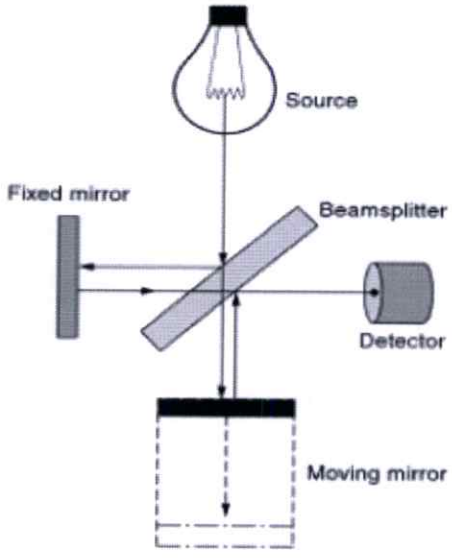



**Figure 2.8** A Michelson interferometer [29]

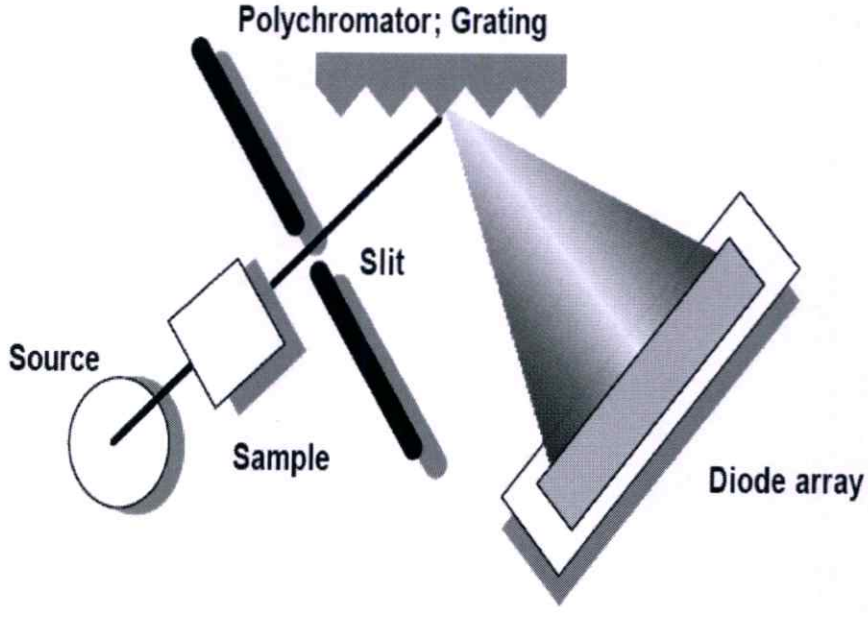



**Figure 2.9** Schematic of photodiode array spectrophotometer [32]

2.4.4.5 Acousto-optic tunable filter (AOTF) spectrometer

An acousto-optic tunable filter (AOTF) is a device that operates on the basis of the interaction between optical radiation and a compression or shear wave traveling within an anisotropic material at a near-acoustic velocity. Hence, their interaction allows light in a very narrow wavelength range to be transmitted. Therefore, the wavelength output is dictated by the applied frequency (Figure 2.10) [33].

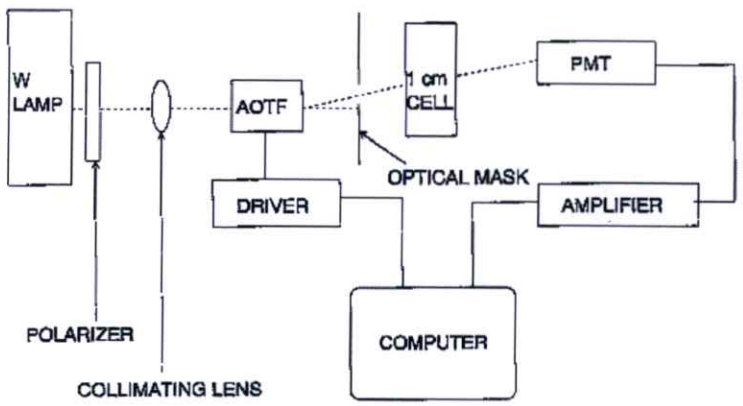



**Figure 2.10** Acousto-optic tunable filter spectrophotometer [33]

### 2.4.5 Principle of selected NIR spectrometer used for this study

Diode array NIR spectrometers (DA7200 and DA7300, Perten, Sweden) were selected to be used for scanning the tapioca starch in this research. According to the experimental plan (see section 1.5), it was needed to scan the tapioca starch in both at-line process and in-line process. Thus, the NIR spectrometer used in each part was different. However, both spectrometer principles are described as follows:

2.4.5.1 Principle of diode array at-line spectrometer (DA7200 Perten, Sweden)

The DA7200 spectrometer consists of a stationary grating for wavelength dispersion and 256 pixel Indium-Gallium-Arsenide (InGaAs) detector operating in the wavelength range 950-1650 nm for energy detection (Figure 2.11) [34]. More details of diode array principle can be found above (see section 2.4.4.4).

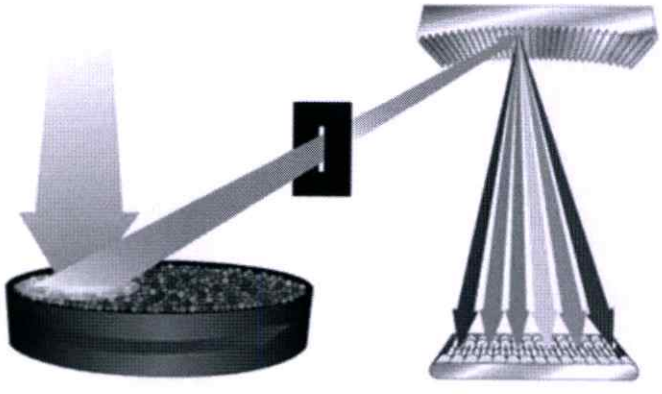


**Figure 2.11** Principle of operation of DA7200 [34]

2.4.5.2 Principle of diode array in-line spectrometer (DA7300 Perten, Sweden)

A solid-state optics design based on a stationary grating for wavelength separation and a diode array for simultaneous collection of light at all wavelengths are used in the DA 7300 NIR spectrometer [35]. Moreover, the concepts of DA7300 NIR instrument have been described by FF instrument [36] as follows: "a portion of the light is absorbed and also the rest is mirrored. However, the light mirrored to a stationary grating that separates the light by wavelength. Therefore, each wavelength is measured by a dedicated detector" (Figure 2.12). More details of diode array principle can also be found above (see section 2.4.4.4).

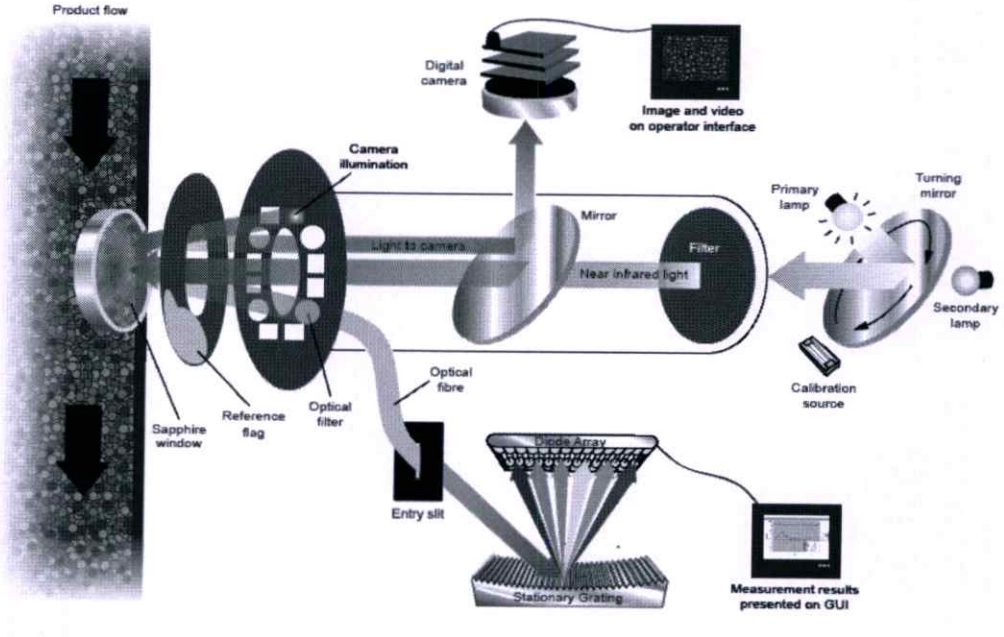



**Figure 2.12** Principle of operation of DA7300 [35]

## 2.5 NIR analysis strategies

There are several strategies for installing the NIR instrument. However, Metrohm [28] mentioned that it can be installed in the laboratory, at-line or directly into the processes. Hence, the location of the NIR analyzer depends on the optical properties of the sample, the required analyze selectivity, the duration of the process run, and the monitoring and control requirements (Figure 2.13).

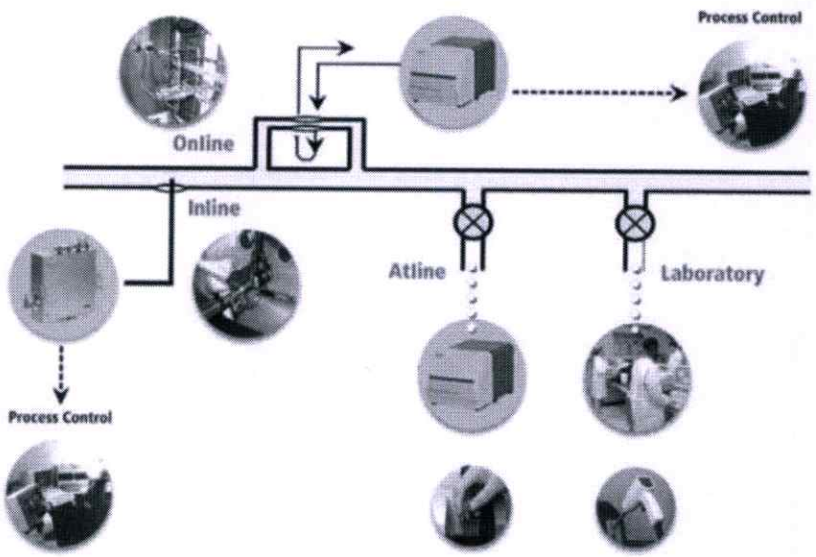



**Figure 2.13** NIR analysis strategies, the process monitoring and control requirements and measurement performance determine ideal strategy [28]

### 2.5.1 Laboratory analysis

Manual sampling and turnaround time for results are the operation that appropriate to the laboratory analysis. Hence the pilot plant facilities, or measurement of long duration runs where process information is required infrequently, are very useful for using this analysis strategy [28].

### 2.5.2 At-line analysis

In this analysis strategy, Metrohm [28] identified that the location of NIR analyzer closes to a process stream. However, the manual sampling is required for at-line analysis. Moreover, the turnaround time for results can also be less than 5 minutes.

Therefore, this strategy is appropriate in process monitoring and control, and in manufacturing operations.

### 2.5.3 In-line analysis

NIR analyzer can be interfaced directly to the process using fiber optics and the probe. Inline analysis strategies operate unattended near real-time (results in < 10 s) analysis on specific media. Although, in-line analysis requires the minimal supporting hardware but maintenance cannot be performed unless the process is shut down. Hence, this strategy is appropriate in closed-loop monitoring and control for scale-up and manufacturing operations [28].

### 2.5.4 On-line analysis

An on-line NIR analysis is interfaced to the process using a sample-loop. NIR spectral are measured on a continuous flow of sample as it passes through a flow-cell. Moreover, slide-streams is the convenient means for performing maintenance and checking the samples to be analyzed, while the process is operating [28].

## 2.6 Data pretreatment

Interfering spectral parameters, such as light scattering, path length variations are resulted from variable physical sample properties or instrumental effects. Hence, data pretreatments are needed prior to multivariate modeling in order to reduce, eliminate their impact on the spectra [17]. Since, there are several means that to be the data pretreatments. Therefore, the principles of these pretreatments were briefly described as follows:

### 2.6.1 Normalization

The normalization, which is used in this study, contains three methods i.e. mean, maximum, and range normalization. Each method was briefly showed as follows [37]:

2.6.1.1 Mean normalization

It consists in dividing each row (each observation; $X_{i,k}$) of a data matrix by its average ($\overline{X}_{i,\bullet}$) as following equation [37]:

$$X_{i,k} = \frac{X_{i,k}}{\left|\overline{X}_{i,\bullet}\right|} \tag{2.13}$$

2.6.1.2 Maximum normalization [37]

This is an alternative to classical normalization which divides each row ($X_{i,k}$) by its maximum absolute value ($\max(|X_{i,\bullet}|)$) as following equation:

$$X_{i,k} = \frac{X_{i,k}}{\max(|X_{i,\bullet}|)} \tag{2.14}$$

2.6.1.3 Range normalization [37]

Each row ($X_{i,k}$) is divided by its range, i.e. "max value – min value" as following equation:

$$X_{i,k} = \frac{X_{i,k}}{\max(X_{i,\bullet}) - \min(X_{i,\bullet})} \tag{2.15}$$

**2.6.2 Derivatives**

Derivatives are typically applied to correct for baseline effects in spectra for the purpose of removing nonchemical effects and creating robust calibration models. Moreover, it may also aid in resolving overlapped bands [37]. First and second derivatives are widely used in practice [38]. Therefore, their concepts are summarized as following ways:

2.6.2.1 First derivative

The first derivative of spectrum is simply a measure of the slope of the spectral curve at every point. Its slope is not affected by purely additive baseline

offsets in the spectrum. Hence, the first derivative is a very effective method for removing the offsets [37].

2.6.2.2 Second derivative

The second derivative is used to measure the change in the slope of the curve and to remove pure additive offset. It is not affected by any linear "tilt" that may exist in the data. Hence, it is a very effective method for removing both the baseline offset and slope from a spectrum, especially the peaks in raw spectra change sign and turn to negative peaks with lobes on either side in the second derivative [37].

### 2.6.3 Baseline offset

Baseline offset is used to adjust the spectral offset by adjusting the data to the minimum point in the data. Hence, the formula for the offset correction can be written as follow [37]:

$$f(x) = x - \min X \tag{2.16}$$

where $x$ is a variable and $X$ denotes all selected variables for this sample.

### 2.6.4 Standard Normal Variate (SNV)

SNV is usually applied to remove scatter effects by centering and scaling each individual spectrum. Moreover, it is also used in combination with de-trending to reduce multicollinearity, baseline shift and curvature in spectroscopic data [37]. Each value $x_k$ in a row of data $X$ is transformed according to the formula [39]:

$$SNV = \frac{x_k - \overline{X}}{SDev(X)} \tag{2.17}$$

### 2.6.5 De-trending

According to report in The Unscrambler Appendices: Method References [37], de-trending concept is described as follows: "it is used to remove nonlinear trends in spectroscopic data. Conceptually, baseline function is calculated as a least square fit

of a polynomial to the sample spectrum. As the polynomial order of the de-trend increases, additional baseline effects are removed".

### 2.6.6 Multiplicative Scatter Correction (MSC)

The multiplicative scatter correction concept has clearly been explained by Buddenbaum and Steffens [39] as follows: it is also pretreatment for baseline correction in spectra. It assumes that mean spectrum $\bar{x}$ is this ideal spectrum. This spectrum represents the mean scattering and offset. Each spectrum $x_i$ is then fit to the mean spectrum using a least squares method:

$$x_i = a_i + b_i\bar{x} + e_i \tag{2.18}$$

However, they continuously described that, $e_i$ contains the chemical information, because scattering and offset are represented by the coefficients $a_i$ and $b_i$. Hence, the MSC spectrum is calculated by determining the coefficients for each spectrum and then transforming the spectrum as follows:

$$MSC_i = \frac{x_i - a_i}{b_i} \tag{2.19}$$

## 2.7 Partial least squares (PLS) regression

PLS regression is a technique that combines features from principal component analysis (PCA) and multiple linear regression (MLR). Its goal is to predict a set of dependent variables $Y$ (predicted matrix) from a set of independent variables $X$ (predictor matrix). Hence, the prediction is achieved by extracting from $X$ as a set of orthogonal components called latent variables (or principal components) [40].

However, Abdi [41] described about PLS regression that two set of weights denoted $w$ and $c$ in order to create (respectively) a linear combination of the columns of $X$ and $Y$ such that these two linear combinations have maximum covariance. Specifically, the goal is to obtain a first pair of vectors:

$$t = Xw \text{ and } u = Yc \tag{2.20}$$

with the constraints that $w^T w = 1, t^T t = 1$ and $t^T u$ is maximal. When the first latent vector is found, it is subtracted from both $X$ and $Y$ and the procedure is re-iterated until $X$ becomes a null matrix [41].

NIPALS: PLS Algorithm, Abdi [41] also emphasized that the first step is to create two matrices: $E = X$ and $F = Y$. These matrices are then column centered and normalized (i.e., transformed into $Z$-scores). In addition, he described about before starting the iteration process that the vector $u$ is initialized with random values. Therefore, the NIPALS algorithm then performs the following steps (in what follows the symbol " $\propto$ " means "to normalize the result of the operation"):

Step 1; $w \propto E^T u$ (estimate $X$ weights).
Step 2; $t \propto Ew$ (estimate $X$ factor scores).
Step 3; $c \propto F^T t$ (estimate $Y$ weights).
Step 4; $u = Fc$ (estimate $Y$ scores).

However, he also described the detailed process as follows: "if $t$ has not converged, then go to Step 1, if $t$ has converged, then compute the value of $b$ which is used to predict $Y$ from $t$ as $b = t^T u$, and compute the factor loading for $X$ as $p = E^T t$. Now subtract (i.e., partial out) the effect of $t$ from both $E$ and $F$ as follows $E = E - tp^T$ and $F = F - btc^T$. This subtraction is called a deflation of the matrices $E$ and $F$. The vectors $t, u, w, c$ and $p$ are then stored in the corresponding matrices, and the scalar $b$ is stored as a diagonal element of $B$. If $E$ is a null matrix, then the whole set of latent vectors has been found, otherwise the procedure can be re-iteration from Step 1 on."

In conclusion, the dependent variables are predicted using the multivariate regression formula as [41]:

$$\hat{Y} = TBC^T \tag{2.21}$$

## 2.8 Model validation

The objective of the model validation is to demonstrate that the developed model is suitable for its intended purpose. Hence, the validation is performed after the model is developed [42]. Two types of validation are internal validation (Cross-validation) and external validation (Test-set validation) [43]. Test-set validation is best suited to large population, and cross-validation is more suitable than test-set validation for small population [44].

### 2.8.1 Internal validation (Cross validation)

The principle of an internal validation, individual samples are taken from the calibration set. The chemometric model is then established by the remaining samples and used to analyze the previously extracted samples. Therefore, the whole process of cross validation is clearly illustrated in the following steps [43]:

2.8.1.1 Remove a sample from the calibration data set.

2.8.1.2 Develop the model with the remaining samples.

2.8.1.3 Analyze the removed test sample; calculate the error of analysis for this sample: $\mathbf{Y}_1^{meas} - \mathbf{Y}_1^{pred}$.

2.8.1.4 Return the removed sample to the data set and remove a new sample. Calculate new model and predict new sample: $\mathbf{Y}_2^{meas} - \mathbf{Y}_2^{pred}$.

2.8.1.5 Repeat step 4 until all samples of the calibration data set have been analyzed once; calculate the mean error of prediction RMSECV.

### 2.8.2 External validation (Test set validation)

In external validation, no samples are excluded from the calibration set. Therefore, the data set is divided into calibration set samples and test-set samples. Since only the samples of test-set are analyzed. So steps of a test-set validation are shown as below [43]:

2.8.2.1 Develop the model, using all calibration set samples.

2.8.2.2 Evaluate the model, using a separate data set of test-set samples with known concentrations; calculate the mean error of prediction RMSEP from the respective analysis errors.

## 2.9 Literature review

Xiao et al. [45] determined the moisture content in Chinese fried bread sticks (FBS) samples. Rapid Fourier transform near-infrared (FT-NIR) methods were developed for determining moisture content in FBS collected from 123 different vendors in Shanghai, China with interval moisture (17.39%-32.65%) content. The moisture content in FBS predicted by FT-NIR methods had very good correlation with their values determined via traditional methods with $R^2$ of 0.983, which clearly indicated that FT-NIR methods could be used as an effective tool for rapid determination of moisture content in FBS.

Haase [46] predicted the potato processing quality of ground raw tubers by near infrared (NIR) reflectance spectroscopy with scanning 850-2500 nm. Calibration equations were developed for dry matter, starch, of dehydrated potatoes. The best coefficients of determination ($R^2$) within independent validation sets were about 0.99 for dry matter, 0.96 for starch. The residual prediction deviation (RPD) statistic was up to 8.5 and 5.4 (dry matter and starch content, respectively).

Haase [47] also estimated the dry matter and starch concentration in potatoes by determination of under-water weight and NIR spectroscopy. The results showed that the determination coefficients ($R^2$) were 0.92 and 0.83 (starch concentration between 13 and 23%), and 0.94 and 0.88 (starch concentration ≥13%) for dry matter and starch concentration, respectively. NIR spectroscopy models for both constituents were then calculated ($R^2$ of validation set was 0.98 and 0.96 for dry matter and starch concentration, respectively).

Ait Kaddour and Cuq [48] studied to investigate the ability of the NIR spectroscopy method to describe the physical and chemical changes occurring during wet agglomeration of wheat flour. The NIR spectra were analyzed as raw spectra and after second derivative treatment by using principal components analysis

(PCA). The results confirm the ability of NIR spectroscopy to identify the wet agglomeration time of wheat flour and the possibility to propose physical and chemical analysis of NIR spectra.

Sudar et al. [49] studied the application of near infrared transmission (NIT) for the determination of ash in wheat flour. The wheat flour samples with ash content of 0.367-0.964% were used for this study. Simple linear regression analyses showed high significant correlation between the NIT and the reference method (r=0.953). However, it still showed that the NIT results were not accurate enough for all flour types when compared to the results obtained by the reference method.

Manley et al. [50] used the FT-NIR spectroscopy to determine the moisture content of a single sample set of whole wheat flour. The calibrations were derived by performing partial least square (PLS) regression on multiplicative scatter corrected (MSC) and baseline corrected data. The results showed standard error of prediction (SEP), root mean standard error of prediction (RMSEP) and correlation coefficient (r) values of 1.16%, 0.15% and 0.85, respectively.

Dong and Sun [51] used the NIR spectroscopy to determine moisture of wheat flour nondestructively combined with characteristic bands selection methods. Pearson product-moment correlation coefficients and interval partial least squares (iPLS) were performed comparatively to choose characteristic bands associated with moisture distributions. The characteristic bands of 4000–4896, 5504–6704 $cm^{-1}$ were chosen by iPLS which were related to moisture. The R and RMSEP of the best models are 0.990 and 0.088%.

Van Zyl et al. [52] studied using the different sample holders in determining moisture content in whole wheat flour by FT-NIR spectroscopy. The conventional sample cup with a sapphire-glass base (provided with the spectrophotometer), borosilicate-glass vials and soda-glass vials were used. Calibrations were derived by performing PLS regression on multiplicative scatter corrected (MSC) spectra and tested using independent validation procedures. Best results were obtained with the sample set analyzed in the borosilicate-glass vials with the standard error of prediction (SEP), root mean standard error of prediction (RMSEP) and the correlation coefficient (R) of 0.15%, 0.38% and 0.94, respectively.

Peiris et al. [53] developed the calibrations for estimating moisture content (MC) of wheat grain samples using a Perkin Elmer Spectrum 400 FTIR/FTNIR spectrometer. Grains in a glass petri-dish were scanned in 10000-4000 $cm^{-1}$ (1000-2500 nm) range. The MC was determined by oven drying method. Mean centered absorbance spectra with MC values were used to develop calibration models using the PLS regression technique. The MC calibration predicted MC in grain samples with $R^2$ of 0.95, SEP of 0.2%. The result suggests that MC of a grain sample can be non-destructively estimated in about 30 seconds per sample using FTNIR spectroscopy.

Hermida et al. [54] determined the moisture and starch in common beans (*Phaseolus vulgaris* L.) by NIR spectroscopy. A set of 96 samples was used to calibrate the instrument by modified PLS regression. The following statistical results were achieved: standard error of calibration (SEC) of 0.31 and square correlation coefficient ($R^2$) of 0.96 for moisture; SEC of 0.76 and $R^2$ of 0.92 for starch. To validate the calibration, a set of 25 bean samples was used. Standard errors of prediction were 0.39 and 0.90 for moisture and starch, respectively, and $R^2$ for the regression of measurements by the reference method versus NIR analysis were 0.94 and 0.88 for moisture and starch, respectively.

Hong at al. [55] studied the feasibility of NIR diffuse reflectance spectroscopy scanning from 1100 to 2500 nm to analyze amounts of moisture, starch in buckwheat flours. The multiple regression equations (MREs) established between the second derivative NIR spectra data and the reference data. The best MREs for the contents of moisture and starch gave multiple correlation coefficients (MCCs) of 0.94 and 0.93, and feasible SEPs of 0.24 and 1.15%. The correlation coefficients (R) between the predicted and the reference data were 0.93 and 0.84 for the contents of moisture and starch, respectively.

Fernandez et al. [56] used the FT-NIR spectrometer in reflectance mode with 4,000-10,000 $cm^{-1}$ wavenumber region for determination of dry matter content of mashed potatoes. The calibration model showed the $R^2$ of 0.92 and SECV of 4.5. Moreover, they also reported that the preliminary results may be possible to use NIR models developed at-line for the prediction of samples analyzed in-line.

Hartmann and Büning-Pfau [57] reported the NIR model developed from the NIR absorbance spectra (1100-2500 nm) to measure the dry matter content of mashed potatoes. The model provided the $R^2$ and SEP of 0.97 and 0.19%, respectively.

From reviewed studies, these confirm that the NIR technique could be used to precisely measure the parameters in agricultural products such as moisture, starch, dry matter and ash content. However, the performance of NIR technique can summarily be distinguished as follows: The ability of NIR technique for measuring the dry matter, starch and ash content in tuber crops provided the high $R^2$. Specially, the moisture content of wheat flour can precisely be measured using NIR technique. Nevertheless, there is no report of tapioca starch moisture content.

# Chapter 3

# Methodology

## 3.1 Samples and sampling

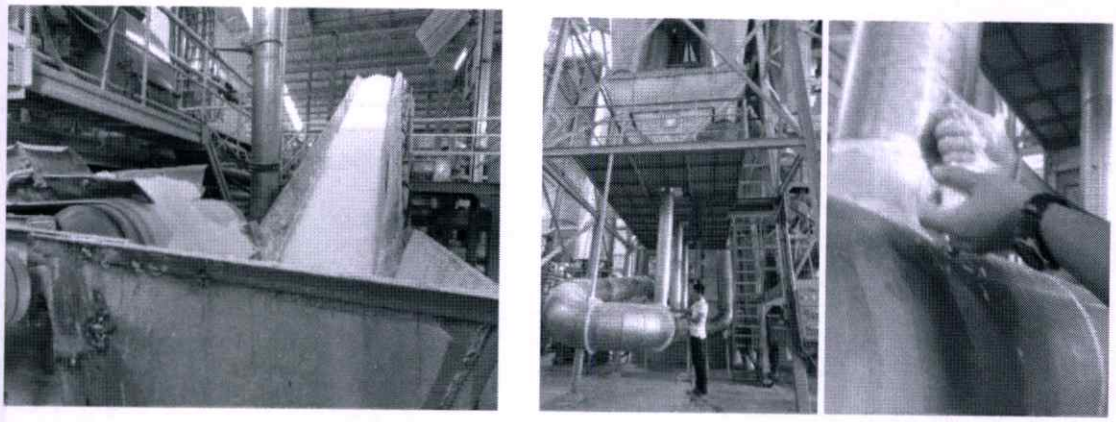
The experiments in this study were divided into three parts including preliminary study for the evaluation of moisture content of tapioca starch using near-infrared (NIR) spectroscopy (Preliminary experiment), the evaluation of the moisture content of tapioca starch using at-line near-infrared spectroscopy (At-line experiment) and the feasibility study of using in-line near-infrared spectroscopic model for predicting the tapioca starch moisture content at the end of the drying process (In-line experiment).

### 3.1.1 Samples and sampling for the preliminary experiment

The tapioca starch samples of different production dates (23, 24 and 25 August 2013) were collected at Sangpetch tapioca flour Co., Ltd. factory in Nongbua-rawe district, Chaiyaphum, Thailand. The samples (70 g each) were adjusted for different levels of moisture content (12.5, 20.63, 28.75 and 36.88% wb) by mixing with distilled water of which the amount were pre-calculated and the sample was called starch cake. The starch cake samples were left in the refrigerator for 24 h before near infrared scanning. There were 24 samples in total. The experiment was done in duplicate.

### 3.1.2 Samples and sampling for the at-line experiment

The tapioca starch samples were collected at the Sangpetch Tapioca Flour Co., Ltd. factory in Nongbua-rawe district, Chaiyaphum, Thailand, in two forms: tapioca starch cake at the inlet of the drying process (Figure 3.1a) and dried tapioca starch at the end (Figure 3.1b). The 210 samples (105 for each form) were kept in plastic cups with covers and immediately brought to the factory laboratory, where the room temperature was 25±1°C. The samples were collected over two periods: 64 samples on 24-26 August 2013 and 146 samples on 19-23 December 2013.



(a) (b)

**Figure 3.1** The sampling position, tapioca starch cake at the inlet of the drying process (a), dried tapioca starch at the end of the drying process (b)

### 3.1.3 Samples and sampling for the in-line experiment

The tapioca starch samples were collected at Sangpetch tapioca flour Co., Ltd. factory in Nongbua-rawe district, Chaiyaphum, Thailand. Each sample was collected while being scanned to obtain near-infrared (NIR) spectrum at the end of the drying process and kept in plastic cups with covers (Figure 3.2). It was immediately brought to the factory laboratory for measuring the moisture content (see section 3.3). The samples were collected about every 1/2 h. There were 93 samples in total.

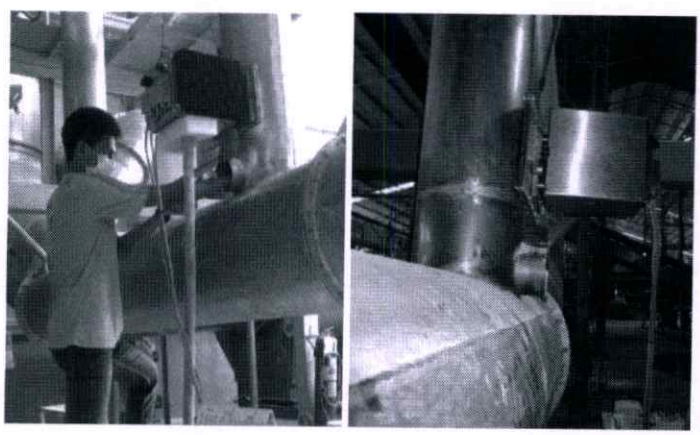


**Figure 3.2** The sampling at the end of drying process

## 3.2 Near-infrared spectroscopy experiment

In this study, the researchers focused on developing the near-infrared (NIR) spectroscopic model for predicting the moisture content of tapioca starch at the end of the drying process. Since, wavelength range of the in-line NIR spectrometer (DA7300, Perten, Sweden) installed at the end of the drying process was 950-1650 nm, the experimental designs for evaluation of the tapioca starch moisture content in the range were important.

### 3.2.1 Preliminary near-infrared spectroscopy experiment

This part was done to study the feasibility for evaluation of the moisture content of tapioca starch in wavelength range of 950-1650 nm, which was the range of the in-line NIR spectrometer.

3.2.1.1 Materials and equipments

1. Diode array NIR spectrometer (DA7200, Perten, Sweden)
2. Sample dish with 75 mm diameter (72.05.04, Perten, Sweden)
3. Plastic cups with covers
4. Samples for the preliminary experiment (see section 3.1.1)

3.2.1.2 Methods

The diode array NIR spectrometer (Figure 3.3) in reflection mode of 950-1650 nm with resolution of 2 nm was used for samples scanning. The NIR exposure time was 0.164 s per scan. The sample was presented to the spectrometer in 75 mm diameter sample dish. The scanning was done while the sample dish was rotated consecutively for 3 rounds and the reference material (Ceramic) was automatically scanned before each sample scan. Therefore there were 72 spectra (24 samples and 3 spectra per sample) in total.

### 3.2.2 At-line near-infrared spectroscopy experiment

This part was done to evaluate the moisture content of tapioca starch in wavelength range of 950-1650 nm, which was the range of the in-line NIR spectrometer.

3.2.2.1 Materials and equipments

1. Diode array NIR spectrometer (DA7200, Perten, Sweden)
2. Sample dish with 75 mm of diameter (72.05.04, Perten, Sweden)

3. Plastic cups with covers
4. Samples for at-line experiment (as mentioned in section 3.1.2)

3.2.2.2 Methods

The spectra of tapioca starch samples were measured with a diode-array NIR spectrometer (Figure 3.3) in reflection mode at 950-1650 nm with a resolution of 2 nm. The NIR exposure time was 0.164 s per scan. The sample was presented to the spectrometer in a 75-mm-diameter sample dish. The NIR illumination was over the dish area. The scanning was done while the sample dish was rotated consecutively for 3 rounds and a reference material (Ceramic) was automatically scanned before each sample scan. Therefore, there were 3 spectra for each sample and they were averaged, 210 spectra in total.

**Figure 3.3** The scanning using a diode-array NIR spectrometer (DA7200, Perten, Sweden)

### 3.2.3 In-line near-infrared spectroscopy experiment

Feasibility study in this part was done for developing the calibration model for predicting the moisture content of tapioca starch at the end of drying process. Therefore, the in-line NIR spectrometer was installed for scanning the tapioca starch stream line at the end of drying process.

3.2.3.1 Materials and equipments

1. Diode array in-line NIR spectrometer (DA7300, Perten, Sweden)
2. Plastic cups with covers
3. Samples for in-line experiment (see section 3.1.2)

3.2.3.2 Methods

A diode-array in-line NIR spectrometer (DA7300, Perten, Sweden) was installed at the end of drying process about 30 cm above the sampling point (Figure 3.4). The NIR spectra were scanned at the tapioca starch stream line through the quartz window where the samples behind moved downward. The spectra were recorded using the reflection mode from 950-1650 nm with a resolution of 2 nm. The sample collected while NIR scanning, was put in a plastic cup with cover and immediately brought to the factory laboratory for measuring the moisture content (see section 3.3).

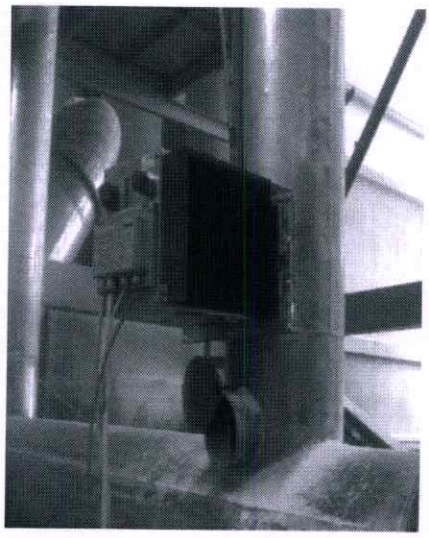


**Figure 3.4** Installation of a diode-array in-line NIR spectrometer (DA7300, Perten, Sweden)

## 3.3 Moisture measurement

### 3.3.1 Comparison of the reference method and standard method

Since, the factory has used the infrared moisture analyzer as the reference method for measuring the moisture content of the tapioca starch. However, it is not the standard method. Hence, the comparison of the reference method with the standard method, which is the hot air oven method (Figure 3.5), was performed. The tapioca starch samples (25 g each) were adjusted for different levels of moisture content (15, 16, 18, 19 and 25% wb) by mixing thoroughly with distilled water of which the amount were pre-calculated. Then it was left in the refrigerator for 24 h

before measuring the moisture content. The samples were left to reach the room temperature before the experiment. There were 3 samples in each level. Therefore, it was 15 samples in total. The 5-g of each sample was used for measuring by reference method, and also 5-g for measuring by standard method. Both methods' procedures were described in chapter 2 (see section 2.3). Then, the data sets measured were statistically compared based on correlation and residual to prove whether there were no significant difference between the tests.

The correlation coefficient ($\mathbf{r}$) was calculated to assess the linear correlation of both methods. For high linear correlation, $\mathbf{r}$ is near to +1 or -1. It can be calculated by the following formula [58].

$$r = \frac{\sum(X-\overline{X})(Y-\overline{Y})}{\sqrt{\sum(X-\overline{X})^2\sum(Y-\overline{Y})^2}} \tag{3.1}$$

In addition, in order to investigate the different measures between standard method and reference method, the residual ($\mathbf{e}$) was evaluated, and was defined [58] as follows:

$$e = X - Y \tag{3.2}$$

where $\mathbf{X}$ is the measured value of standard method and $\mathbf{Y}$ is the corresponded measured value of reference method. Then, residuals were plotted against the measured values by standard method.

### 3.3.2 Moisture measurement using the infrared moisture analyzer

After scanning, the moisture content of a 5-g tapioca starch sample was immediately analyzed by the factory reference method using an infrared moisture analyzer (HB43-S Halogen, Mettler Toledo, Switzerland) at 130°C shown in figure 3.6. Duplicate tests were run for each sample. Therefore, the repeatability, which was the standard deviation of the difference between duplicates, was calculated to indicate the precision of the test.

**Figure 3.5** The tapioca starch moisture content measurement using standard method (Hot air oven)

**Figure 3.6** The tapioca starch moisture content measurement using an infrared moisture analyzer

## 3.4 Spectrum pre-treatment and NIR spectroscopy model establishment

Before model establishment, the NIR spectra used for model development were either not pre-treated or pre-treated mathematically. The following pretreatments were used: mean normalization, maximum normalization, range normalization, first derivatives (5, 11 and 21 points), second derivatives (5, 11 and 21 points), baseline offset, standard normal variate (SNV), detrending, SNV+detrending or multiplicative scatter correction (MSC). There were three parts in this study. The partial least squares regression (PLS) was used for the NIR spectroscopic models establishment for predicting the moisture content of starch using The Unscrambler v. 9.8 (Camo, Norway) in part of preliminary experiment and using The Unscrambler X

10.3 (Camo, Norway) in part of at-line and in-line experiment. Therefore, the model establishments were also divided as follows:

### 3.4.1 NIR spectroscopy model establishment for preliminary experiment

The 24 samples were not enough for external validation. Thus, the internal validation was used in this part. The models were then validated using full cross-validation. The optimum model was selected by coefficient of determination ($R^2$), standard error of cross-validation (SECV), a bias and residual prediction deviation (RPD).

### 3.4.2 The NIR spectroscopy model establishment for at-line experiment

Three groups of sample spectra were used for model development: those of tapioca starch cake samples, dried tapioca starch samples and combined samples (cake and dried samples). The data set of each group was divided into a calibration set (two-thirds of the data set) and a test set (one-third of the data set), after the reference data were arranged in ascending order. The calibration set was used for model development. The model accuracy was determined using full cross-validation, as indicated by $R^2$ and SECV. In external validation using the test set, the optimum model was selected based on the high $R^2$, low SEP, low prediction bias and high RPD.

### 3.4.3 The NIR spectroscopy model establishment for in-line experiment

In order to study for feasibility of using at-line and/or in-line near-infrared spectroscopic model for predicting the tapioca starch moisture content at the end of the drying process, the model establishments were divided into four parts as follows:

3.4.3.1 The NIR model establishment using in-line data (Model #1) and internal validation (Cross-validation)

Since, the physical properties of samples between the in-line and at-line experiment were different (e.g. density, sample temperature, sample feature). Therefore, this part was performed using in-line data in order to evaluate the tapioca starch moisture content. The models were then validated using full cross validation. The optimum model was selected by $R^2$, SECV, a bias and RPD.

3.4.3.2 The NIR model establishment using in-line data (Model #2) and external validation (Test set validation)

The internal validation (Cross-validation) that was used in previous section (see section 3.4.3.1) could not completely confirm the accuracy of the models. So that, the external validation (Test set) was performed on in-line data. This method which is better than the internal validation can better indicate the model accuracy. The in-line data set (93 samples) was arranged in ascending order. It was then divided into a calibration set (47 samples) and a test set (46 samples). This was done by assigning the first one sample to be in the calibration set and second one to be in test set consecutively until all samples had been allocated.

3.4.3.3 The NIR model establishment using at-line data as calibration set and in-line data as test set (Model #3)

The at-line samples group was used as calibration set (105 samples) for model development. Therefore, the test set was in-line samples group (93 samples). The model accuracy was determined using full cross-validation, as indicated by $R^2$ and SECV. In external validation using the test set, the optimum model was selected based on the high $R^2$, low SEP, low prediction bias and high RPD.

3.4.3.4 The NIR model establishment using the calibration set consisted of 100% of at-line data and 50% of in-line data and test set consisted of another 50% of the inline data (Model #4)

There were two groups of sample data set, i.e. at-line data and in-line data. The in-line data set (93 samples) was arranged in ascending order. It was then divided into a calibration set (47 samples) and a test set (46 samples). This was done by assigning the first one sample to be in the calibration set and second one sample to be in test set until all samples had been allocated. All of at-line data (105 samples) and 50% of in-line data (47 samples) were used as calibration set. Another 50% of in-line data (46 samples) was used as test set. The model accuracy was determined using full cross-validation, as indicated by $R^2$ and SECV. In external validation using the test set, the optimum model was selected based on the high $R^2$, low SEP, low prediction bias and RPD.

# Chapter 4

# Results and discussion

## 4.1 Accuracy of the reference method

According to the standard method, the measurement of tapioca starch moisture content must be evaluated by using hot air oven. Since, reference method used in this study, which was acceptable in most of tapioca starch factories, was the infrared moisture analyzer, the comparison of standard method with reference method was performed in order to analyze the difference of measured values. Table 4.1 shows the moisture content measured by the standard and reference methods.

**Table 4.1** The moisture content of tapioca starch samples measured by hot air oven (Standard method) and infrared moisture analyzer (Reference method)

| Samples | Infrared moisture analyzer (% wb) (Reference method) | Hot air oven (% wb) (Standard method) |
|---|---|---|
| 1 | 15.65 | 15.51 |
| 2 | 15.61 | 15.60 |
| 3 | 15.62 | 15.64 |
| 4 | 16.04 | 16.16 |
| 5 | 16.07 | 16.23 |
| 6 | 16.20 | 16.18 |
| 7 | 18.14 | 18.43 |
| 8 | 18.00 | 18.05 |
| 9 | 18.08 | 18.08 |
| 10 | 19.32 | 19.28 |
| 11 | 19.47 | 19.68 |
| 12 | 20.06 | 19.72 |
| 13 | 25.45 | 25.93 |
| 14 | 26.83 | 26.41 |
| 15 | 26.71 | 26.55 |

The correlation between standard method and reference method, which was calculated by equation (3.1), was high with the correlation coefficient of 0.998. In addition, the difference of both methods was showed in residual plot (Figure 4.1). It shows that the points in a residual plot are randomly dispersed around the horizontal axis and also shows that the residuals of high moisture content samples are somewhat higher.

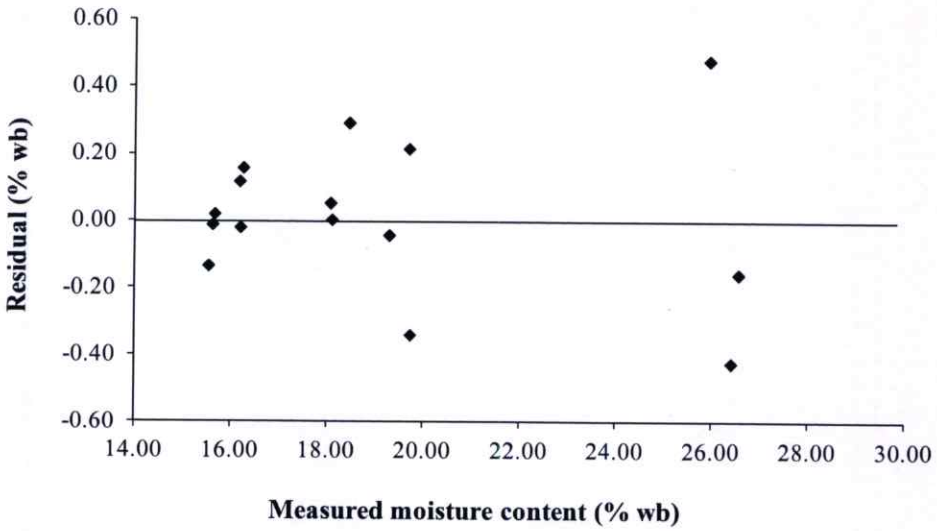



**Figure 4.1** Residual plot of measured moisture content (Standard method) in X axis with error values in Y axis

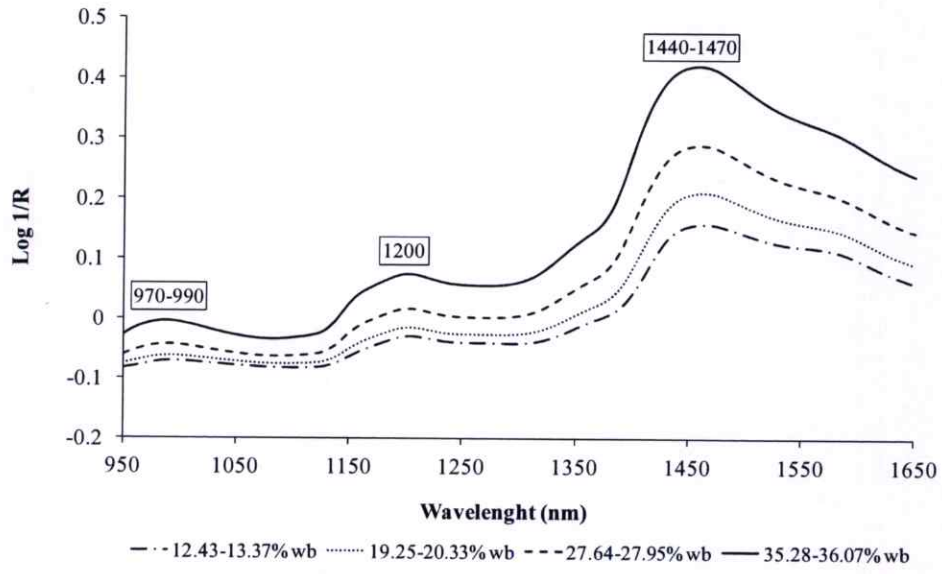
## 4.2 Preliminary study for the evaluation of moisture content of tapioca starch using near-infrared (NIR) spectroscopy (Preliminary experiment)

### 4.2.1 Spectral Analysis of Absorption Features

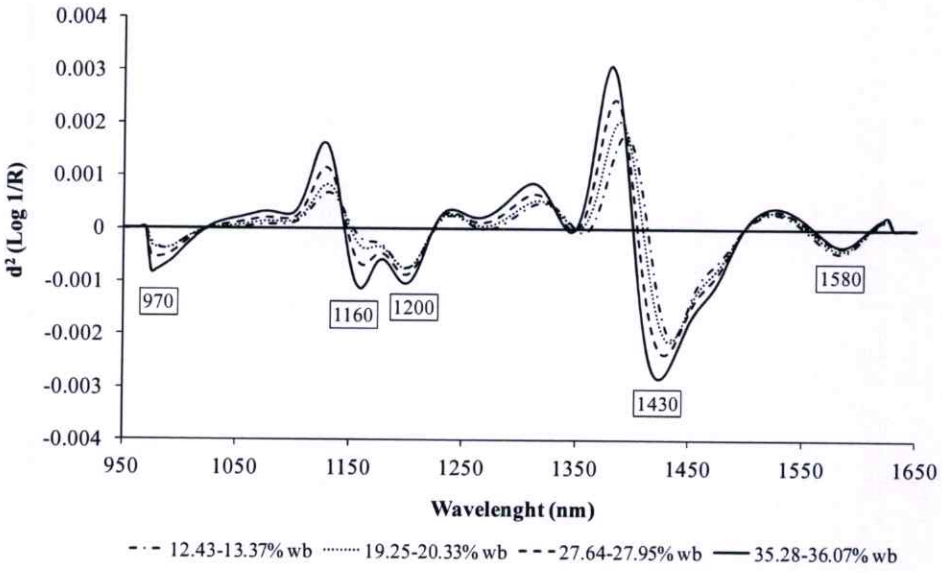
There were 4 levels of the moisture content in all starch samples including 12.43-13.37, 19.25-20.33, 27.64-27.95 and 35.28-36.07% wb. Therefore, the NIR spectra were averaged into 4 spectra of the 4 levels of moisture content. The average raw and average second-derivative spectra of tapioca starch samples are shown in figure 4.2 (a) and (b), respectively.

There were 3 dominant absorption regions in the raw spectra (Fig. 4.2a) i.e. 970-990, 1200 and 1440-1470 nm. In the measured range 950-1650 nm, water absorption due to the vibration of O-H bonds can be found at approximately 970, 1200 and

1450 nm [59, 60]. However, according to Williams, the peak in the 1400 nm region was associated with the glucose molecules in the starch constituents [61].




(a)




(b)

**Figure 4.2** Averaged NIR spectra of tapioca starch at different levels of moisture content, (a) Average raw absorbance spectra, (b) Average second-derivative spectra (11 points)

At 1400 nm region, the second-derivative spectra (Fig. 4.2b) shows that the tapioca starch at highest moisture content (35.28-36.07% wb) has higher absorption than that of the others which indicated that the peak in 1400 nm region is more due to the absorption of water in the samples. Bands at 970, 1200, 1430 and 1580 nm were also observed in all spectra, corresponding to the second overtone associated with the O-H stretching of water (970 nm), the starch band (1202 nm), the first overtone associated with the O-H stretching of water and starch (1450 nm), and the first overtone associated with the O-H stretching of starch (1580 nm) [20]. Moreover, there was a peak at 1160 nm (Fig. 4.2b) which was a shifted peak of 1190 nm of absorption band of water [62] where the absorbance intensity of the highest moisture content range samples were prominent.

### 4.2.2 Tapioca starch moisture content prediction using Partial Least Squares Regression

The repeatability of the reference moisture content evaluation method was 0.42% for starch samples. The value was accepted by the factory. The maximum (Max), minimum (Min), mean and standard deviation (SD) of the tapioca starch moisture content are shown in Table 4.2.

The results of Partial Least Squares (PLS) regression models for predicting the tapioca starch moisture content are shown in Table 4.3. The optimum models for starch samples were developed from the range normalization. One PLS factor was used for predicting the tapioca starch moisture content.

The scatter plot of the near-infrared spectroscopic model for predicting the tapioca starch moisture content is shown in figure 4.3. The model showed that the coefficient of determination ($R^2$), standard error of cross-validation (SECV), a bias and residual prediction deviation (RPD) of 0.997, 0.52%, -0.001% and 16.8, respectively. According to Williams [44], an $R^2$ of 0.98 or more and the RPD of 8.1 or more implies that a model can be used excellently for any application.

**Table 4.2** Moisture content (% wb) of tapioca starch (Preliminary experiment data set) measured by the reference method

| Number of samples | Maximum | Minimum | Mean | Standard deviation |
|---|---|---|---|---|
| 24 | 36.07 | 12.43 | 23.99 | 8.78 |

**Table 4.3** Results of the PLS calibration models (Preliminary experiment data set)

| Spectrum pretreatment | Factor | Calibration model | | | RPD |
|---|---|---|---|---|---|
| | | $R^2$ | SECV (%) | Bias (%) | |
| No-pretreatment | 2 | 0.982 | 1.17 | -0.002 | 7.49 |
| Normalize | | | | | |
| - Mean | 4 | 0.972 | 1.48 | -0.002 | 5.93 |
| - Max | 1 | 0.988 | 0.95 | -0.004 | 9.21 |
| - **Range*** | **1** | **0.997** | **0.52** | **-0.001** | **16.81** |
| Derivative | | | | | |
| - 1st derivative (5 points) | 2 | 0.995 | 0.62 | 0.002 | 14.06 |
| - 1st derivative (11 points) | 2 | 0.994 | 0.71 | 0.004 | 12.41 |
| - 1st derivative (21 points) | 2 | 0.990 | 0.89 | 0.005 | 9.90 |
| - 2nd derivative (5 points) | 1 | 0.990 | 1.06 | 0.010 | 8.28 |
| - 2nd derivative (11 points) | 2 | 0.996 | 0.58 | -0.000 | 15.28 |
| - 2nd derivative (21 points) | 2 | 0.994 | 0.70 | 0.003 | 12.59 |
| Baseline-offset | 2 | 0.980 | 0.91 | 0.003 | 9.60 |
| SNV | 1 | 0.993 | 0.74 | -0.001 | 11.85 |
| SNV+De-trending | 1 | 0.993 | 0.71 | -0.002 | 12.29 |
| De-trending | 2 | 0.991 | 0.82 | 0.005 | 10.73 |
| MSC | 1 | 0.993 | 0.74 | -0.001 | 11.89 |

* Selected optimum model, Coefficient of determination ($R^2$), Standard error of cross-validation (SECV), Standard normal variate (SNV), Multiplicative scatter correction (MSC) and Residual prediction deviation (RPD)

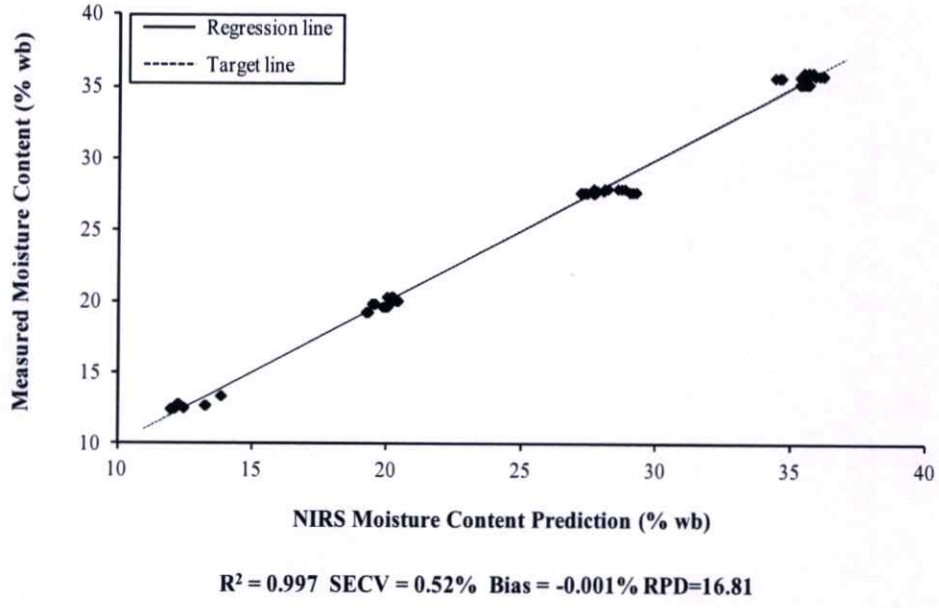



**Figure 4.3** Scatter plot of predicted moisture content with measured moisture content in tapioca starch samples

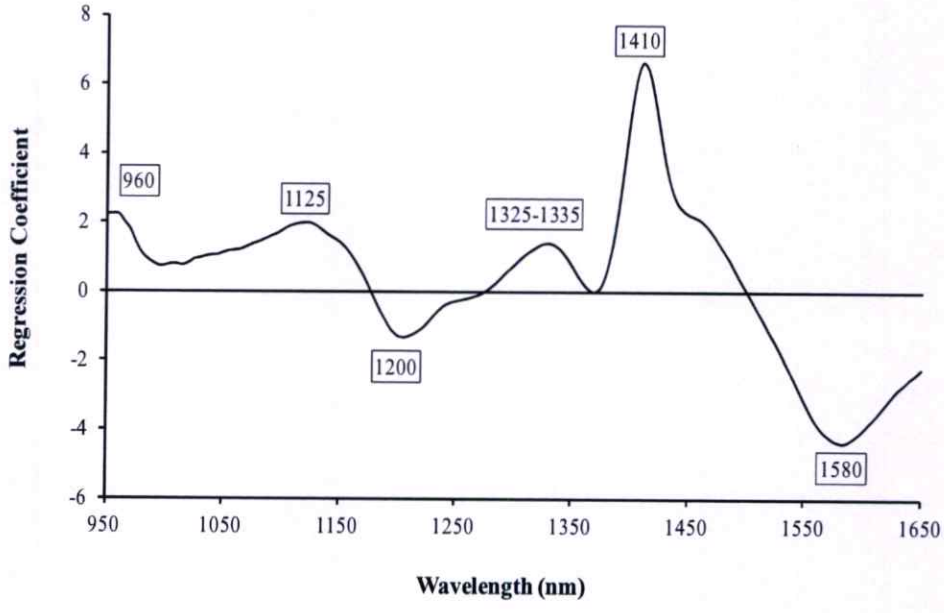



**Figure 4.4** Regression coefficient plot of optimum model for the moisture content in tapioca starch samples

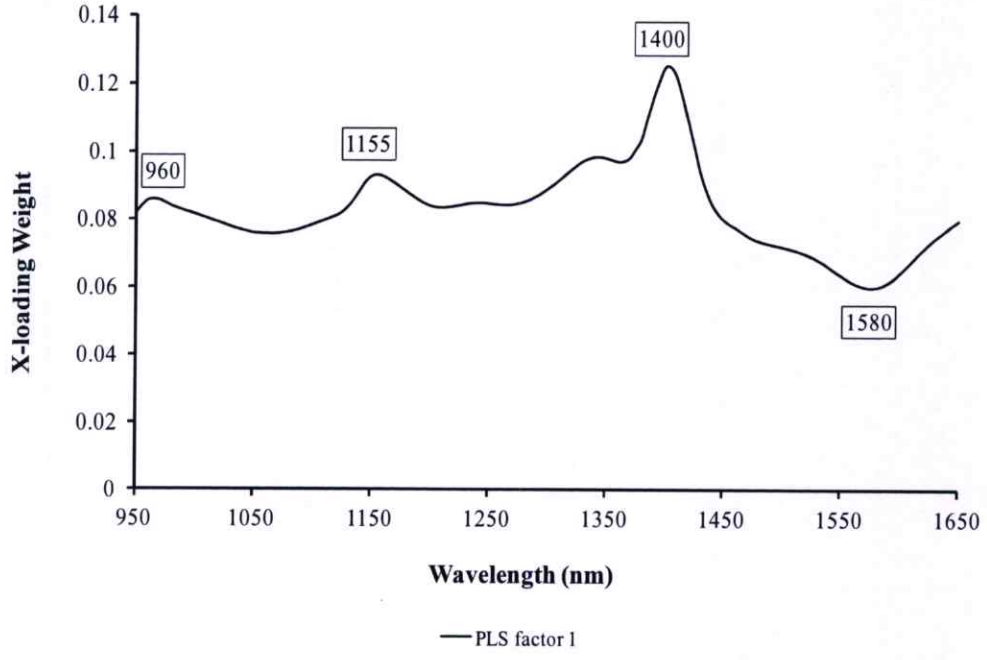



**Figure 4.5** X-loading weight plot of optimum model for the moisture content in tapioca starch samples

Figure 4.4 shows the regression coefficient plot of the optimum model for predicting the moisture content of starch. There were bands of water and starch appeared on the plot. The peak of the starch spectrum at 1410 nm might be the strong absorption band of water in the samples [44] or was possibly associated with the glucose molecules that formed the starch constituents [61]. There were the peaks at 960, 1200 and 1580 nm which were the second overtone associated with O-H stretching of water (970 nm) and that of the bands of 1200 and 1580 nm were due to the starch band and absorption band of the first overtone associated with O-H stretching of starch, respectively [20].

The X-loading weight for predicting the moisture content of starch model is shown in figure 4.5. The appropriate number of PLS factors were selected based on the minimum value of SECV. Only one PLS latent variable was used in the optimum model of starch. The highest proportion of the total variance of the system is normally accounted by the first PLS factor [44]. The explained variance in X variables was 100% (NIR spectra) and in Y variable (Moisture content) was 99%. The peaks at 960, 1155, 1400 and 1580 nm were observed which were the second overtone associated with O-H stretching of water [20], the shifted peak of 1190 nm of

absorption band of water [62], the strong absorption of water [44] or the glucose molecules that form the starch constituents [61], and absorption band of the first overtone associated with O-H stretching of starch [20], respectively. These confirmed that absorption bands of starch and water influenced the starch moisture content prediction. Therefore, the absorption bands that are involved with prediction of tapioca starch moisture content, which appeared in figure 4.2, 4.4 and 4.5, are clearly shown in table 4.4.

**Table 4.4** Vibration bands of some peaks at wavelength appeared on average second-derivative spectra of tapioca starch, regression coefficient plot and X-loading weight plot in the preliminary experiment

| Appeared wavelength (nm) | Cited wavelength (nm) | Vibration band | Structure | Source |
|---|---|---|---|---|
| 960, 970 | 970 [20] | O-H str. second overtone | $H_2O$ | 2AS, RC, F1 |
| 1160 | 1190 [62] | Absorption band | $H_2O$ | 2AS, F1 |
| 1200 | 1202 [20] | Absorption band | starch | 2AS, RC |
| 1400, 1410 | 1400 [61], [44] | Absorption band | water, glucose | RC, F1 |
| 1430 | 1450 [20] | O-H str. first overtone | water, starch | 2AS |
| 1580 | 1580 [20] | O-H str. first overtone | starch | 2AS, RC, F1 |

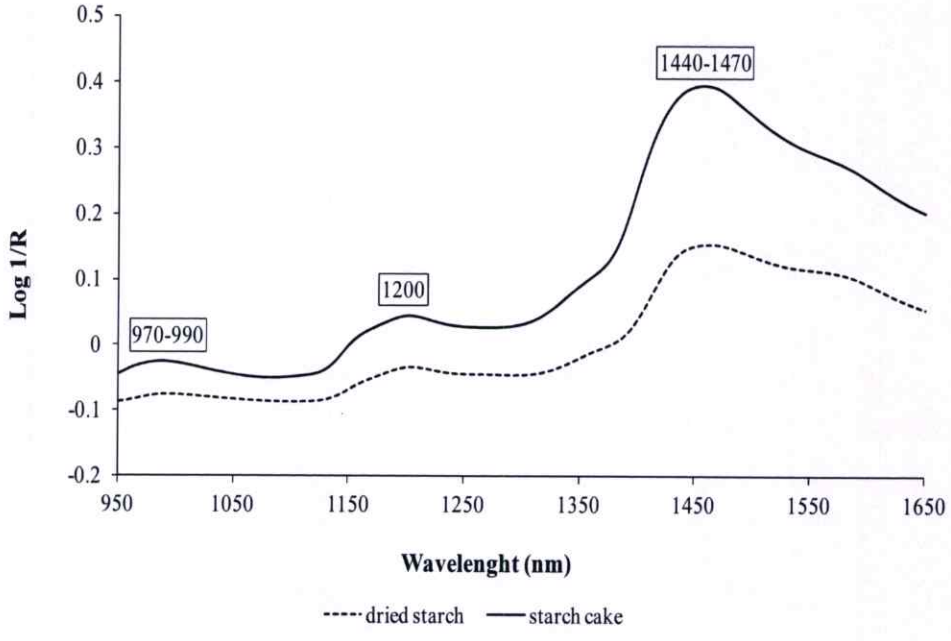
F1, F2 and F3 are PLS factor 1, 2 and 3 in X-loading plot, respectively. RC is regression coefficient plot and, 2AS is average second-derivative spectra of tapioca starch.

## 4.3 The evaluation of the moisture content of tapioca starch using at-line near-infrared (NIR) spectroscopy (At-line experiment)

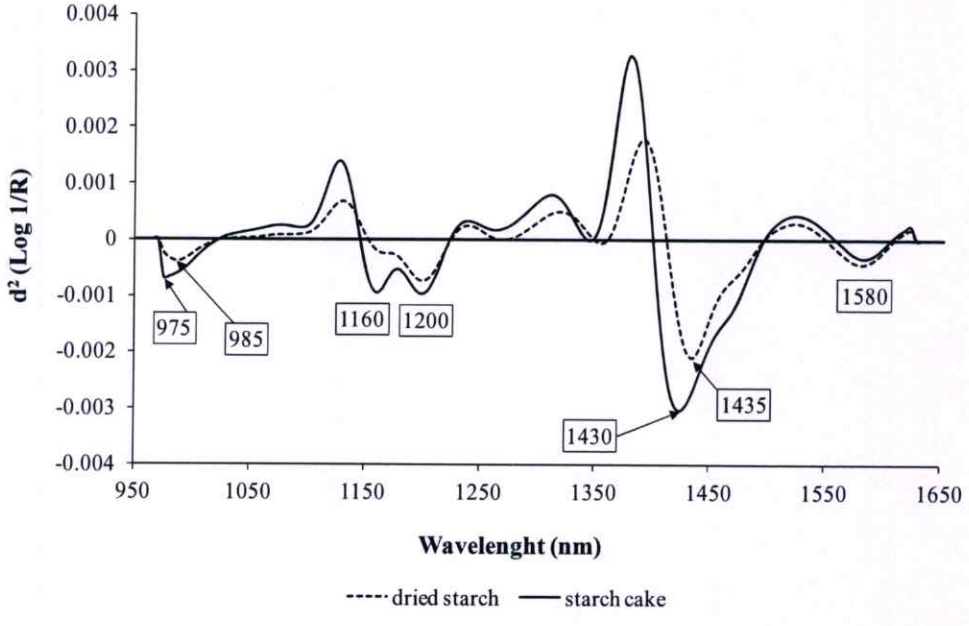
Three groups of sample spectra were used for model development i.e. cake samples, dried tapioca starch samples and combined samples (Cake and dried samples).

### 4.3.1 Spectral Analysis of Absorption Features

In order to analyze absorption features, the NIR spectra were thus averaged into 2 spectra of cake and dried tapioca starch samples.




(a)




(b)

**Figure 4.6** Averaged NIR spectra of tapioca starch cake and dried tapioca starch, (a) Average raw absorbance spectra, (b) Average second-derivative spectra (11 points)

Figure 4.6 (a) and (b) show the average absorbance of the raw and second-derivative spectra of cake and dried samples, respectively. There were 3 dominant absorption regions in the raw spectra (Fig. 4.6a) i.e. 970-990, 1200 and 1440-1470 nm. It is similar to average absorbance of raw and second-derivative spectra as mentioned in section 4.2.1 (see Fig. 4.2). In the measured range 950-1650 nm, water absorption due to the vibration of O-H bonds can be found at approximately 970, 1200 and 1450 nm [59, 60]. Although Williams [61] indicated that the peak in the 1400 nm region was associated with the glucose molecules in the starch constituents, in this case the second-derivative spectra (Fig. 4.6b) shows that the tapioca starch cake has higher absorption than that of the dried tapioca starch which indicated that the peak in 1400 nm region is more due to the absorption of water in the samples. Bands at 1200 and 1580 nm were also observed in both the cake and the dried starch spectra, corresponding to the starch band (1202 nm) and the first overtone associated with the O-H stretching of starch (1580 nm) [20].

From second-derivative spectra (Fig. 4.6b), there was a peak at 1160 nm which was the shifted peak of 1190 nm of absorption band of water [62] where the absorbance intensity of starch cake is higher than that of dried starch. However, a band at 975 nm in the cake spectrum and 985 nm in dried starch spectrum could be observed, which were the second overtone associated with the O-H stretching of water (970 nm) and the second overtone associated with starch (990 nm) [20]. The shift in the position of the band at 975 nm in the cake spectrum to 985 nm in the dried starch spectrum may have been a result of the removal of water during drying.

### 4.3.2 Tapioca starch moisture content prediction using Partial Least Squares Regression

The repeatability of the reference moisture content evaluation method was 0.45% and 0.15% for starch cake samples and dried starch samples, respectively. The maximum (Max), minimum (Min), mean and standard deviation (SD) of the moisture content in tapioca starch of different groups are shown in Table 4.5.

**Table 4.5** Moisture content (% wb) of tapioca starch (At-line experiment data sets) measured by the reference method used to develop the prediction model and validate the test set

| Group | Calibration | | | | | Test | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | N | Max | Min | Mean | SD | N | Max | Min | Mean | SD |
| Starch cake | 71 | 41.13 | 29.32 | 33.52 | 2.04 | 34 | 37.49 | 29.93 | 33.38 | 1.64 |
| Dried starch | 71 | 18.82 | 10.91 | 12.90 | 0.97 | 34 | 14.01 | 11.39 | 12.81 | 0.60 |
| Combined-starch | 141 | 41.13 | 10.91 | 23.24 | 10.48 | 69 | 37.49 | 11.39 | 23.03 | 10.32 |

**Table 4.6 (a)** Results of the PLS calibration models (At-line experiment data sets) for tapioca starch cake samples prediction

| Spectrum pretreatment | Factor | Calibration | | Test | | | RPD |
|---|---|---|---|---|---|---|---|
| | | $R^2$ | SECV (%) | $R^2$ | SEP (%) | Bias (%) | |
| No pretreatment | 4 | 0.951 | 0.48 | 0.887 | 0.55 | 0.040 | 2.98 |
| Normalized | | | | | | | |
| - Mean | 4 | 0.961 | 0.42 | 0.938 | 0.41 | 0.047 | 4.00 |
| - Max | 4 | 0.973 | 0.35 | 0.940 | 0.42 | 0.083 | 3.90 |
| **- Range*** | **4** | **0.973** | **0.35** | **0.943** | **0.41** | **0.069** | **4.00** |
| Derivatives | | | | | | | |
| - 1st derivative (5 points) | 4 | 0.956 | 0.45 | 0.930 | 0.43 | 0.005 | 3.81 |
| - 1st derivative (11 points) | 4 | 0.961 | 0.42 | 0.933 | 0.42 | 0.015 | 3.90 |
| - 1st derivative (21 points) | 4 | 0.957 | 0.44 | 0.926 | 0.44 | 0.008 | 3.73 |
| - 2nd derivative (5 points) | 3 | 0.938 | 0.54 | 0.906 | 0.50 | -0.004 | 3.28 |
| - 2nd derivative (11 points) | 5 | 0.965 | 0.41 | 0.933 | 0.42 | 0.009 | 3.90 |
| - 2nd derivative (21 points) | 4 | 0.960 | 0.42 | 0.933 | 0.42 | 0.010 | 3.90 |
| Baseline-offset | 5 | 0.963 | 0.41 | 0.931 | 0.43 | 0.033 | 3.81 |
| SNV | 3 | 0.968 | 0.37 | 0.919 | 0.49 | 0.094 | 3.35 |
| SNV + Detrending | 3 | 0.972 | 0.35 | 0.925 | 0.47 | 0.092 | 3.49 |
| Detrending | 4 | 0.961 | 0.42 | 0.929 | 0.44 | 0.018 | 3.73 |
| MSC | 3 | 0.968 | 0.37 | 0.919 | 0.49 | 0.095 | 3.35 |

* Selected optimum model, coefficient of determination ($R^2$), standard error of cross-validation (SECV), standard error of prediction (SEP), standard normal variate (SNV), multiplicative scatter correction (MSC) and residual prediction deviation (RPD)

**Table 4.6 (b)** Results of the PLS calibration models (At-line experiment data sets) for dried tapioca starch samples prediction

| Spectrum pretreatment | Factor | Calibration | | Test | | | RPD |
|---|---|---|---|---|---|---|---|
| | | $R^2$ | SECV (%) | $R^2$ | SEP (%) | Bias (%) | |
| **No pretreatment*** | **4** | **0.972** | **0.17** | **0.928** | **0.15** | **-0.062** | **4.00** |
| Normalized | | | | | | | |
| - Mean | 4 | 0.219 | 0.89 | 0.015 | 0.70 | 0.212 | 0.86 |
| - Max | 3 | 0.965 | 0.19 | 0.881 | 0.19 | -0.083 | 3.16 |
| - Range | 4 | 0.972 | 0.17 | 0.873 | 0.19 | -0.096 | 3.16 |
| Derivatives | | | | | | | |
| - 1st derivative (5 points) | 3 | 0.968 | 0.18 | 0.918 | 0.16 | -0.058 | 3.75 |
| - 1st derivative (11 points) | 3 | 0.970 | 0.17 | 0.909 | 0.17 | -0.064 | 3.53 |
| - 1st derivative (21 points) | 3 | 0.974 | 0.16 | 0.884 | 0.18 | -0.094 | 3.33 |
| - 2nd derivative (5 points) | 4 | 0.967 | 0.20 | 0.920 | 0.16 | -0.047 | 3.75 |
| - 2nd derivative (11 points) | 3 | 0.968 | 0.18 | 0.926 | 0.16 | -0.051 | 3.75 |
| - 2nd derivative (21 points) | 3 | 0.970 | 0.17 | 0.897 | 0.18 | -0.068 | 3.33 |
| Baseline-offset | 4 | 0.969 | 0.18 | 0.932 | 0.15 | -0.051 | 4.00 |
| SNV | 3 | 0.966 | 0.19 | 0.915 | 0.16 | -0.065 | 3.75 |
| SNV + Detrending | 3 | 0.970 | 0.17 | 0.894 | 0.18 | -0.072 | 3.33 |
| Detrending | 3 | 0.970 | 0.17 | 0.885 | 0.19 | -0.079 | 3.16 |
| MSC | 3 | 0.966 | 0.19 | 0.915 | 0.16 | -0.065 | 3.75 |

* Selected optimum model, coefficient of determination ($R^2$), standard error of cross-validation (SECV), standard error of prediction (SEP), standard normal variate (SNV), multiplicative scatter correction (MSC) and residual prediction deviation (RPD)

**Table 4.6 (c)** Results of the PLS calibration models (At-line experiment data sets) for combined tapioca starch samples prediction

| Spectrum pretreatment | Factor | Calibration | | Test | | | RPD |
|---|---|---|---|---|---|---|---|
| | | $R^2$ | SECV (%) | $R^2$ | SEP (%) | Bias (%) | |
| No pretreatment | 3 | 0.986 | 1.25 | 0.980 | 1.47 | 0.129 | 7.02 |
| Normalized | | | | | | | |
| - Mean | 1 | 0.033 | 12.03 | 0.014 | 10.28 | 0.461 | 1.00 |
| - Max | 2 | 0.991 | 1.01 | 0.993 | 0.90 | 0.037 | 11.47 |
| - Range | 2 | 0.993 | 0.86 | 0.994 | 0.82 | 0.059 | 12.59 |
| Derivatives | | | | | | | |
| - 1st derivative (5 points) | 2 | 0.992 | 0.98 | 0.987 | 1.20 | 0.216 | 8.60 |
| - 1st derivative (11 points) | 2 | 0.988 | 1.14 | 0.982 | 1.44 | 0.252 | 7.17 |
| - 1st derivative (21 points) | 1 | 0.975 | 1.67 | 0.966 | 1.93 | 0.146 | 5.35 |
| **- 2nd derivative (5 points)*** | **2** | **0.997** | **0.61** | **0.996** | **0.65** | **0.043** | **15.88** |
| - 2nd derivative (11 points) | 2 | 0.995 | 0.74 | 0.994 | 0.82 | 0.135 | 12.59 |
| - 2nd derivative (21 points) | 2 | 0.988 | 1.12 | 0.983 | 1.42 | 0.269 | 7.27 |
| Baseline-offset | 1 | 0.972 | 1.76 | 0.958 | 2.15 | 0.256 | 4.80 |
| SNV | 1 | 0.992 | 0.95 | 0.992 | 0.95 | 0.161 | 10.86 |
| SNV + Detrending | 1 | 0.992 | 0.92 | 0.992 | 0.95 | 0.161 | 10.86 |
| Detrending | 1 | 0.975 | 1.66 | 0.966 | 1.90 | 0.131 | 5.43 |
| MSC | 1 | 0.992 | 0.95 | 0.992 | 0.95 | 0.161 | 10.86 |

* Selected optimum model, coefficient of determination ($R^2$), standard error of cross-validation (SECV), standard error of prediction (SEP), standard normal variate (SNV), multiplicative scatter correction (MSC) and residual prediction deviation (RPD)

The results of PLS regression models for predicting the moisture content of tapioca starch of starch cake, dried starch and combined sample groups are shown in Table 4.6 (a), (b) and (c), respectively. The optimum models for the three different groups were developed from the range-normalization, no pretreatment and second-derivative (5 points) spectra, respectively, using 950-1650 nm wavelength range. The appropriate number of PLS factors were selected based on the minimum value of SECV. The optimum models for starch cake, dried starch and combined starch samples were developed from four, four and two PLS factors, respectively. The first PLS factor usually accounts for the highest proportion of the total variance (i.e., the combined spectra and reference data) of the system [44]. The explained variance in the X variables (NIR spectra) and Y variable (Moisture content) by the PLS factors are shown in Table 4.7.

**Table 4.7** X (NIR spectra) and Y (Moisture) explained variance for tapioca starch moisture content prediction models (At-line experiment data sets)

| Model | PLS factor | X-explained variance (%) | Y-explained variance (%) |
|---|---|---|---|
| Tapioca starch cake samples | 1 | 97 | 15 |
| (Range normalize) | 2 | 3 | 62 |
| | 3 | 0 | 7 |
| | 4 | 0 | 13 |
| Dried tapioca starch samples | 1 | 95 | 20 |
| (No pretreatment) | 2 | 4 | 60 |
| | 3 | 0 | 23 |
| | 4 | 0 | 7 |
| Combined samples | 1 | 99 | 99 |
| (Second derivative, 5 points) | 2 | 0 | 1 |

The optimum models performances are explained as follows:

4.3.2.1 The NIR model for predicting the moisture content of tapioca starch cake samples

Figure 4.7 shows the scatter plot of the near-infrared spectroscopic model for predicting the moisture content of tapioca starch cake. It provided an $R^2$ of 0.943, standard error of prediction (SEP) of 0.41%, bias of 0.069% and RPD of 4.00.

The regression coefficient plot of the optimum model for the moisture content of starch cake is shown in figure 4.8. The common bands relevant to the moisture content prediction of the starch cake model were at approximately 990, 1200, 1365, 1395 and 1580 nm. Moreover, there was also an obvious band of the highest regression coefficient at 1430 nm. Therefore, the absorption bands that highly influenced the prediction of tapioca starch moisture content are clearly shown in table 4.8.

The X-loading weight plot for the prediction model of the starch cake moisture content is shown in Figure 4.9. The X-loading spectra of various PLS factors show that vibration bands relevant to the moisture prediction of tapioca starch cake model were located at 990, 1200, 1365, 1395, 1430 and 1575-1585 nm. These bands were associated with starch and water [20, 44, 59, 60].

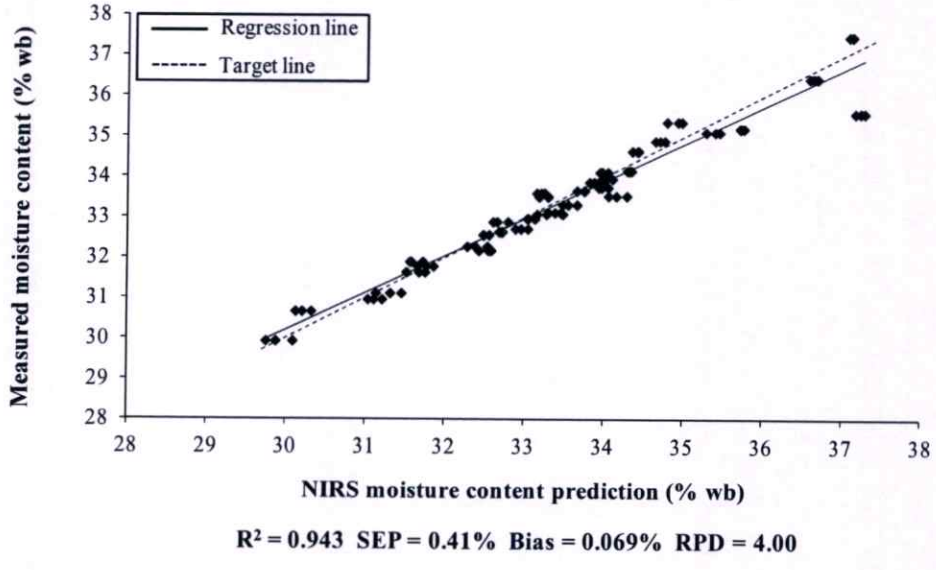



**Figure 4.7** Scatter plot of predicted moisture content with measured moisture content in tapioca starch cake samples

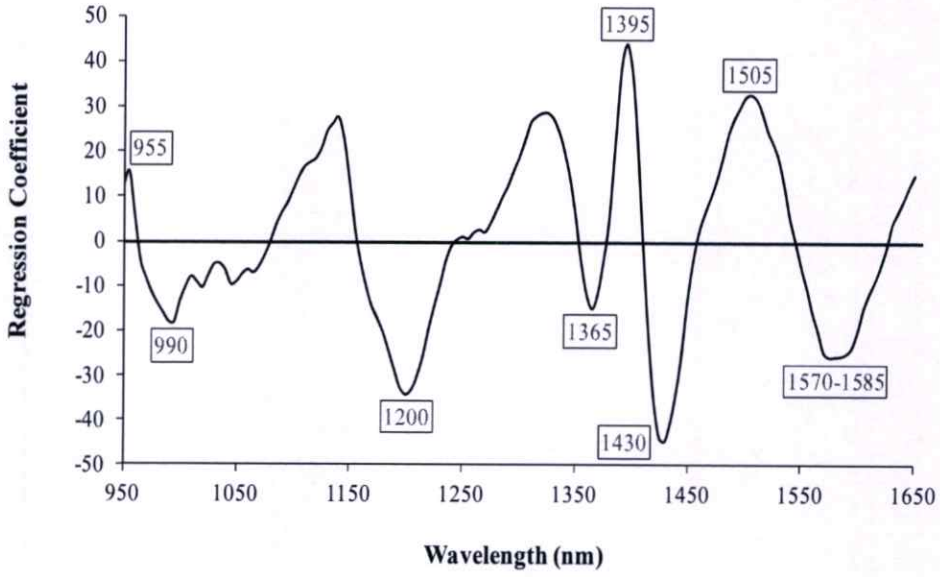



**Figure 4.8** Regression coefficient plot of optimum model for the moisture content in tapioca starch cake samples

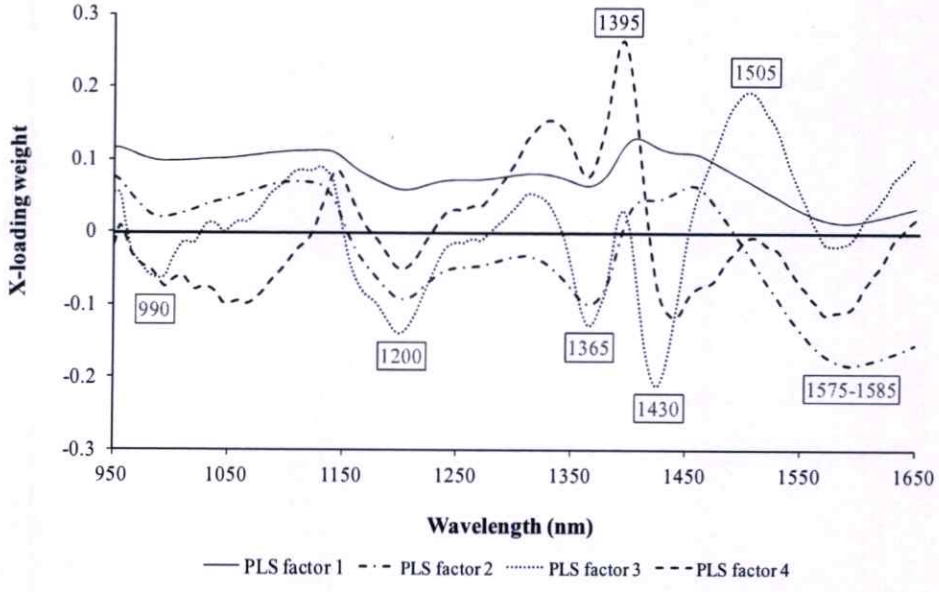



**Figure 4.9** X-loading weight plot of optimum model for the moisture content in tapioca starch cake samples

4.3.2.2 The NIR model for predicting the moisture content of dried tapioca starch samples

Figure 4.10 shows the scatter plot of the near-infrared spectroscopic model for predicting the moisture content of dried tapioca starch. It provided an $R^2$ of 0.928, SEP of 0.15%, bias of -0.062% and RPD of 4.00.

The regression coefficient plot of the optimum model for the moisture content of dried starch is shown in figure 4.11. The common bands relevant to the moisture content prediction of the dried starch model were at approximately 985, 1205, 1405 and 1575 nm. Moreover, there was also an obvious band of the highest regression coefficient at 1435 nm. Therefore, these bands that highly affect the moisture content prediction of tapioca starch are also shown in table 4.8.

The X-loading weight plot for the prediction model of the moisture content of dried starch is shown in Figure 4.12. The X-loading spectra of various PLS factors show that vibration bands relevant to the moisture prediction of dried tapioca starch model were located at 985, 1205, 1410, 1435 and 1580 nm. These bands were also associated with starch and water [20, 44, 59, 60].

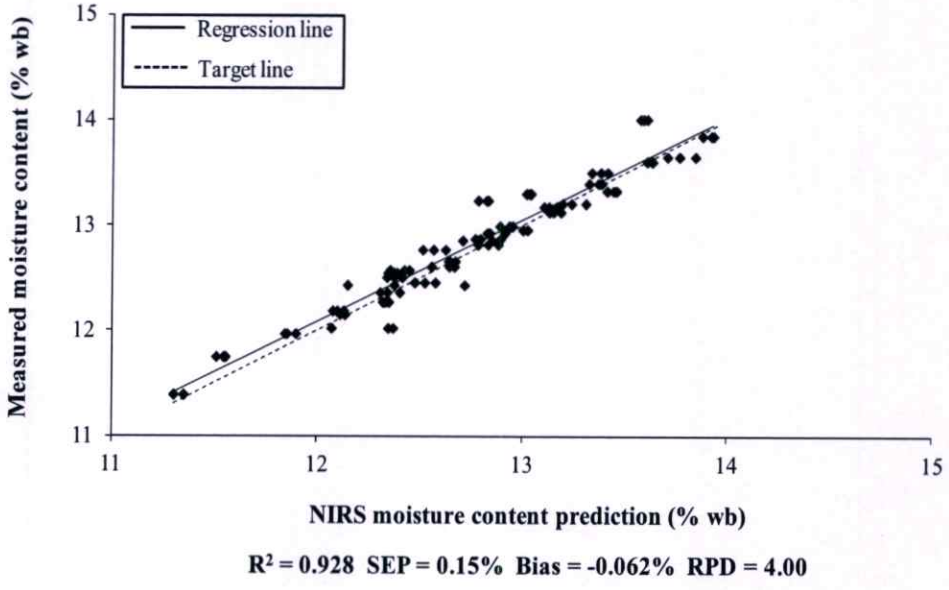



**Figure 4.10** Scatter plot of predicted moisture content with measured moisture content in dried tapioca starch samples

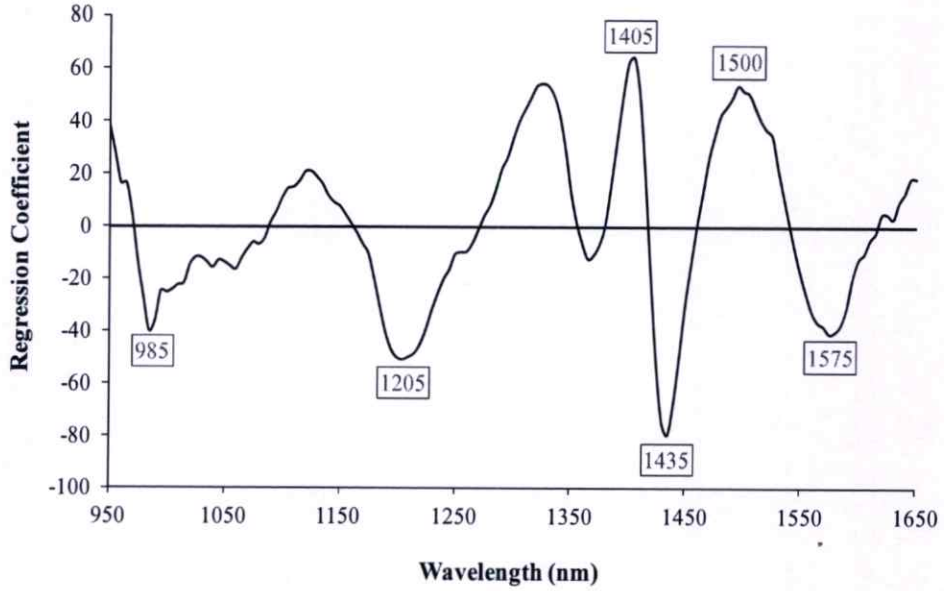



**Figure 4.11** Regression coefficient plot of optimum model for moisture content in dried tapioca starch samples

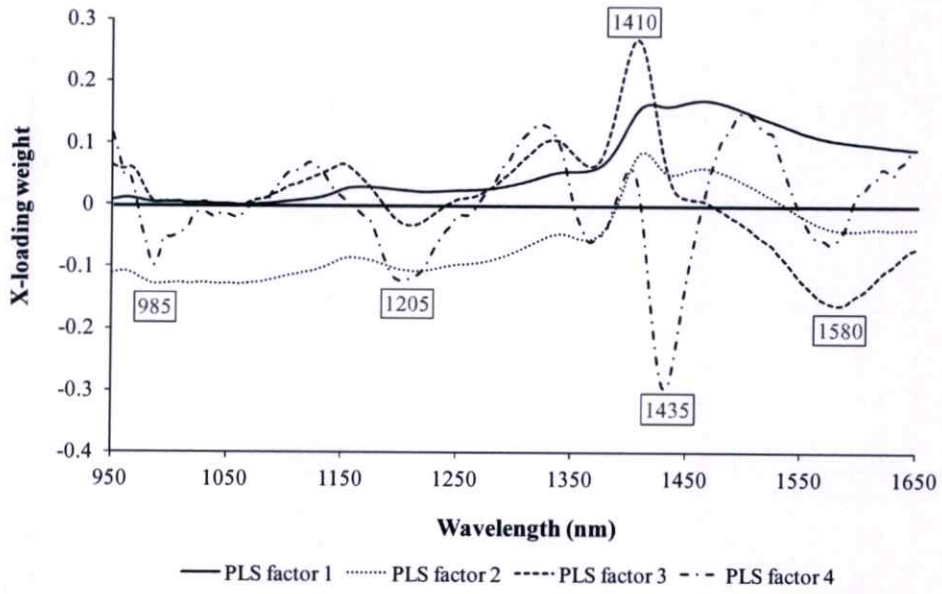



**Figure 4.12** X-loading weight plot of optimum model for moisture content in dried tapioca starch samples

4.3.2.3 The NIR model for predicting the moisture content of combined tapioca starch samples

Figure 4.13 shows the scatter plot of the near-infrared spectroscopic model for predicting the moisture content of combined tapioca starch. It provided an $R^2$ of 0.996, SEP of 0.65%, bias of 0.043% and RPD of 15.88.

The regression coefficient plot of the optimum model for the moisture content of combined starch is shown in figure 4.14. The bands relevant to the moisture content prediction of the combined starch model were at approximately 985, 1200, 1385, 1410, 1435 and 1580 nm. Details of these vibration bands are summarized in table 4.8.

The X-loading weight plot for the prediction model of the moisture content of combined starch is shown in Figure 4.15. The X-loading spectra of various PLS factors show that vibration bands relevant to the moisture prediction of combined tapioca starch model were located at 985, 1200, 1380, 1410, 1435 and 1580 nm. These bands were also associated with starch and water [20, 44, 59, 60].

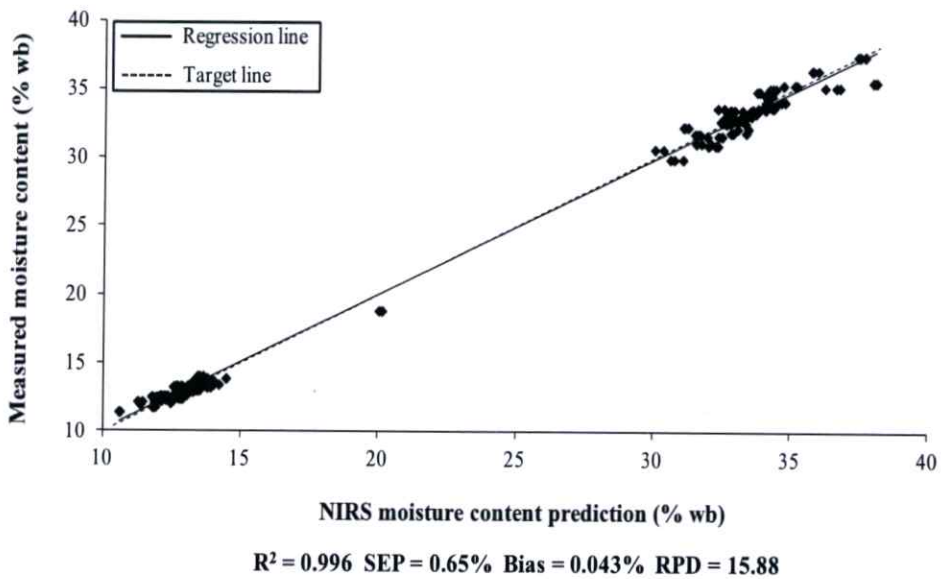



**Figure 4.13** Scatter plot of predicted moisture content with measured moisture content in combined tapioca starch samples

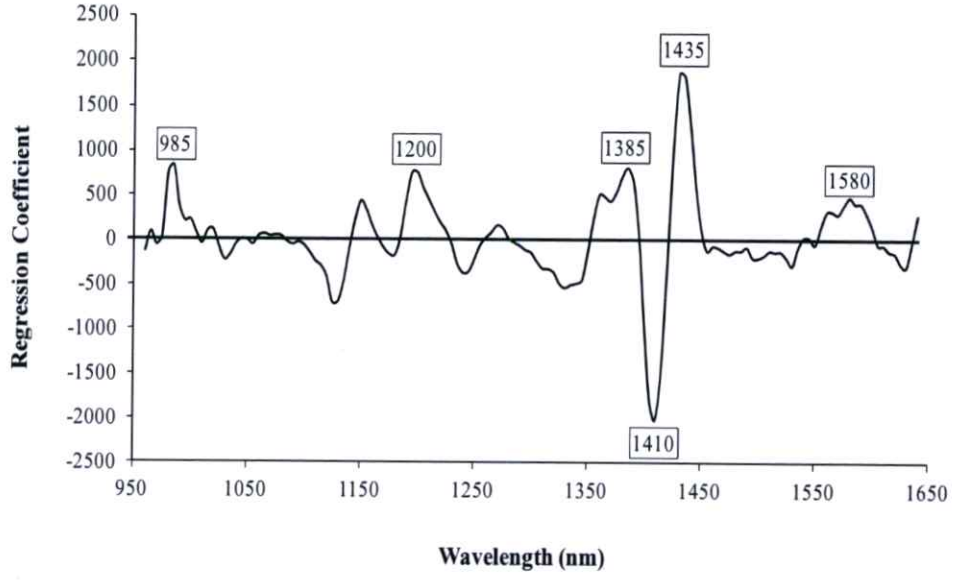



**Figure 4.14** Regression coefficient plot of optimum model for moisture content in combined tapioca starch samples

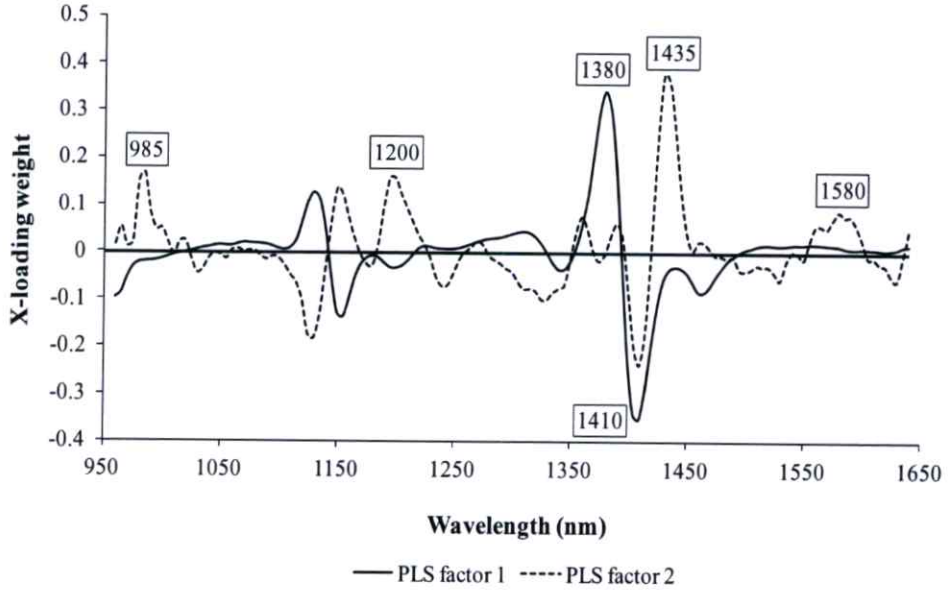



**Figure 4.15** X-loading weight plot of optimum model for moisture content in combined tapioca starch samples

**Table 4.8** Vibration bands of some peaks at wavelength appeared on average second-derivative spectra of tapioca starch samples, regression coefficient plot and X-loading weight plot in the at-line experiment

| Appeared wavelength (nm) | Cited wavelength (nm) | Vibration band | Structure | Source |
|---|---|---|---|---|
| 975 | 970 [20] | O-H str. second overtone | $H_2O$ | 2AS |
| 985, 990 | 990 [20] | O-H str. second overtone | starch | RC, F2, F3, F4 |
| 1160 | 1190 [62] | Absorption band | $H_2O$ | 2AS |
| 1200, 1205 | 1202 [20] | Absorption band | starch | 2AS, RC, F1, F2, F3, F4 |
| 1365, 1380, 1385 | 1360 [20] | Absorption band | starch | RC, F1, F3 |
| 1395, 1405, 1410 | 1400 [44], [61] | Absorption band | glucose, water | RC, F1, F2, F3, F4 |
| 1430, 1435 | 1450 [20] | O-H str. first overtone | water, starch | 2AS, RC, F2, F3, F4 |
| 1575, 1580, 1585 | 1580 [20] | O-H str. first overtone | starch | 2AS, RC, F2, F3, F4 |

F1, F2, F3 and F4 are PLS factor 1, 2, 3 and 4 in X-loading plot, respectively. RC is regression coefficient plot and, 2AS is average second-derivative spectra of tapioca starch.

From the results presented in the at-line experiment, it can clearly be summarized that the optimum near-infrared spectroscopic models for the moisture content of tapioca starch cake, dried starch and combined samples provided a coefficient of determination ($R^2$) of 0.928-0.996, standard error of prediction (SEP) of 0.15-0.65%, bias of (0.069)-(-0.062)% and residual prediction deviation (RPD) of 4.00-15.88. The model of dried tapioca starch samples using no pretreated spectra (Fig. 4.10) had the highest prediction ability, i.e., the lowest SEP (0.15%), while the $R^2$, bias and RPD were 0.928, -0.062% and 4.00, respectively. Williams [44] has indicated that an $R^2$ of between 0.66-0.81, 0.83-0.90 and 0.92-0.96 implies that a model can be used for screening, usable with caution for most applications, including research, and in most applications, including quality assurance, respectively. In addition, this result was comparable to the results of Fourier transform near-infrared (FT-NIR) spectroscopy for the prediction of moisture content of whole-wheat flour, where SEP, root mean

standard error of prediction (RMSEP) and correlation coefficient (r) values of 0.15%, 0.38% and 0.85, respectively, were obtained [50]. Moreover, the highest peak (1430-1435 nm region), which appeared in both regression coefficient and X-loading weight plot, showed the high influence of absorption of water in prediction of moisture content. Other peaks also confirmed that absorption bands of starch and water influenced the tapioca starch moisture content prediction.

## 4.4 Feasibility study for the evaluation of the moisture content of tapioca starch using in-line near-infrared (NIR) spectroscopy (In-line experiment)

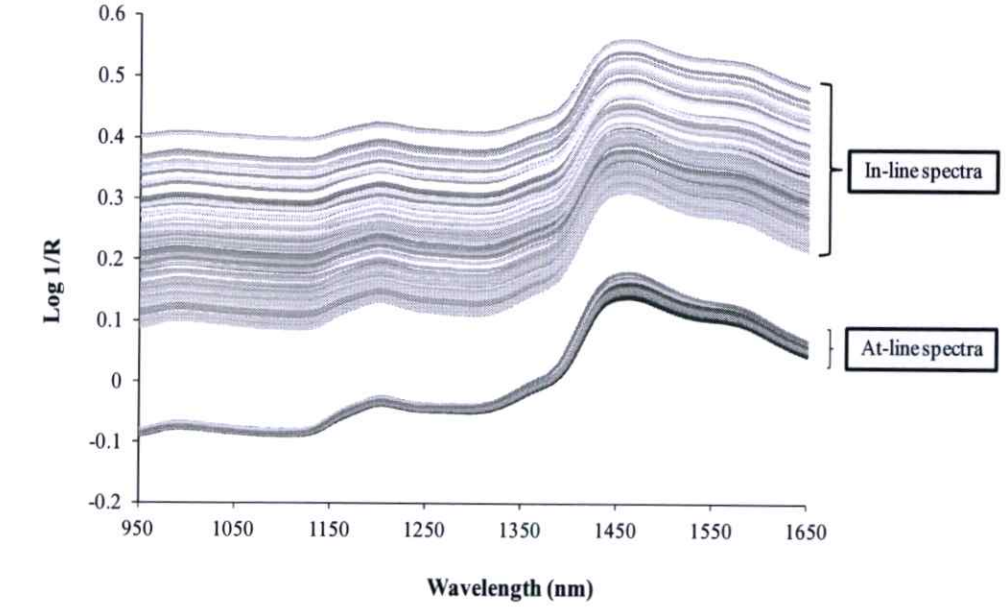
### 4.4.1 Spectral Analysis of Absorption Features

Since the physical characteristic of samples scanned in-line and at-line (Only dried starch) were different, appeared NIR spectra were distinctly separated into two groups as shown in figure 4.16a. So that, the spectral analysis was needed in order to consider the absorption features.

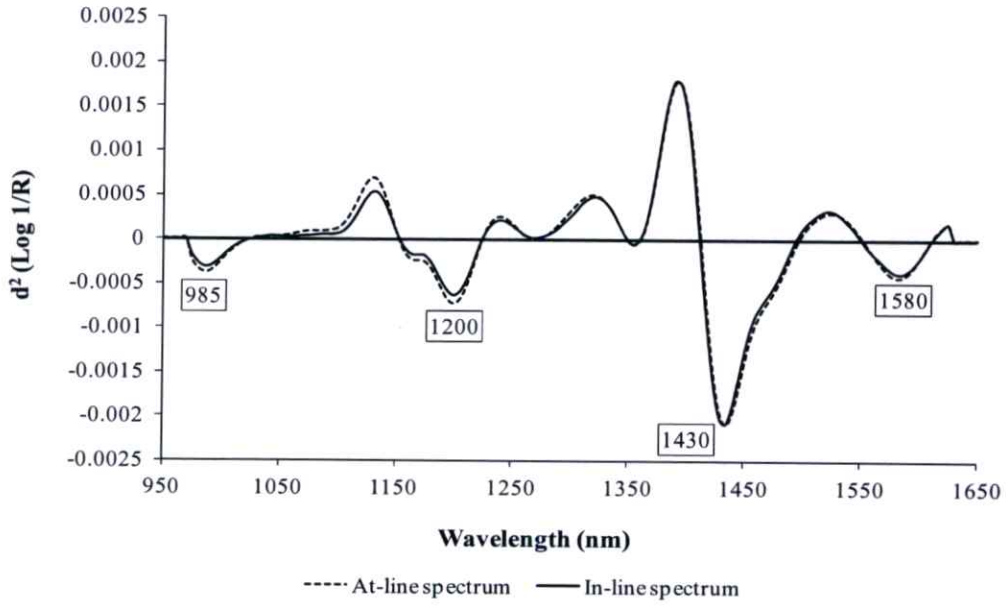
Figure 4.16 (a) and (b) show the absorbance of the raw and average second-derivative spectra of at-line and in-line samples, respectively. There were 4 dominant absorption regions in the average second-derivative spectra (Fig. 4.16b) i.e. 985, 1200, 1430 and 1580 nm. In addition, figure 4.16b shows that there is no difference in 1430 region between at-line and in-line spectrum. The vibration bands observed in both of spectra at 985, 1200, 1430 and 1580 nm, were clearly shown in table 4.16.

### 4.4.2 Tapioca starch moisture content prediction using Partial Least Squares Regression

The repeatability of the reference moisture content evaluation method was 0.15% and 0.12% for at-line samples (Only dried starch) and in-line samples, respectively. Since there were four parts for model establishment in the in-line experiment (see section 3.4.3), the maximum (Max), minimum (Min), mean and standard deviation (SD) of the tapioca starch moisture content in all parts are shown in Table 4.9.




(a)




(b)

**Figure 4.16** NIR spectra of tapioca starch between at-line and in-line experiment, (a) Average raw absorbance spectra, (b) Average second-derivative spectra (11 points)

**Table 4.9** Moisture content (% wb) of tapioca starch (In-line experiment data sets) measured by the reference method used for model development and validation

| Group | Calibration | | | | | Test | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | N | Max | Min | Mean | SD | N | Max | Min | Mean | SD |
| **Model #1** | 93 | 16.11 | 10.46 | 12.79 | 1.08 | - | - | - | - | - |
| **Model #2** | 47 | 16.11 | 10.46 | 12.79 | 1.13 | 46 | 15.75 | 10.47 | 12.78 | 1.04 |
| **Model #3** | 105 | 18.82 | 10.91 | 12.87 | 0.87 | 93 | 16.11 | 10.46 | 12.79 | 1.08 |
| **Model #4** | 152 | 18.82 | 10.46 | 12.86 | 0.90 | 46 | 15.75 | 10.47 | 12.78 | 1.04 |

**Note:** Details of all models were illustrated in Chapter 3 (see section 3.4.3).

**Table 4.10** Results of the PLS calibration models for the in-line data using internal validation (Cross-validation) (Model #1)

| Spectrum pretreatment | Factor | Calibration model | | | RPD |
|---|---|---|---|---|---|
| | | $R^2$ | SECV (%) | Bias (%) | |
| No-pretreatment | 7 | 0.818 | 0.53 | -0.003 | 2.04 |
| Normalize | | | | | |
| - Mean | 5 | 0.740 | 0.60 | -0.004 | 1.80 |
| - Max | 6 | 0.793 | 0.56 | -0.001 | 1.93 |
| - Range | 6 | 0.805 | 0.53 | -0.003 | 2.04 |
| Derivative | | | | | |
| - 1$^{st}$derivative (5 points) | 3 | 0.800 | 0.52 | -0.004 | 2.08 |
| - 1$^{st}$derivative (11 points) | 3 | 0.784 | 0.53 | -0.002 | 2.04 |
| - 1$^{st}$derivative (21 points) | 5 | 0.801 | 0.54 | -0.000 | 2.00 |
| **- 2$^{nd}$derivative (5 points)*** | **2** | **0.806** | **0.54** | **-0.014** | **2.00** |
| - 2$^{nd}$derivative (11 points) | 3 | 0.798 | 0.52 | -0.003 | 2.08 |
| - 2$^{nd}$derivative (21 points) | 4 | 0.798 | 0.52 | -0.001 | 2.08 |
| Baseline-offset | 5 | 0.798 | 0.53 | -0.004 | 2.04 |
| SNV | 5 | 0.804 | 0.53 | -0.006 | 2.04 |
| SNV+De-trending | 4 | 0.800 | 0.53 | -0.006 | 2.04 |
| De-trending | 4 | 0.796 | 0.52 | -0.003 | 2.08 |
| MSC | 5 | 0.803 | 0.53 | -0.006 | 2.04 |

* Selected optimum model, coefficient of determination ($R^2$), standard error of cross-validation (SECV), standard error of prediction (SEP), standard normal variate (SNV), multiplicative scatter correction (MSC) and residual prediction deviation (RPD)

**Table 4.11** Results of the PLS calibration models for the in-line data using external validation (Test set) (Model #2)

| Spectrum pretreatment | Factor | Calibration | | Test | | | RPD |
|---|---|---|---|---|---|---|---|
| | | $R^2$ | SECV (%) | $R^2$ | SEP (%) | Bias (%) | |
| No pretreatment | 6 | 0.897 | 0.47 | 0.583 | 0.66 | -0.100 | 1.58 |
| Normalized | | | | | | | |
| - Mean | 5 | 0.813 | 0.60 | 0.515 | 0.72 | -0.081 | 1.44 |
| - Max | 5 | 0.851 | 0.53 | 0.526 | 0.71 | -0.074 | 1.46 |
| - Range | 5 | 0.884 | 0.47 | 0.490 | 0.74 | -0.050 | 1.41 |
| Derivatives | | | | | | | |
| - 1st derivative (5 points) | 3 | 0.885 | 0.45 | 0.641 | 0.63 | -0.033 | 1.65 |
| - 1st derivative (11 points) | 5 | 0.895 | 0.36 | 0.596 | 0.66 | -0.081 | 1.58 |
| - 1st derivative (21 points) | 5 | 0.887 | 0.47 | 0.509 | 0.72 | -0.113 | 1.44 |
| **- 2nd derivative (5 points)*** | **3** | **0.900** | **0.43** | **0.641** | **0.62** | **-0.069** | **1.68** |
| - 2nd derivative (11 points) | 3 | 0.884 | 0.45 | 0.647 | 0.61 | -0.054 | 1.70 |
| - 2nd derivative (21 points) | 3 | 0.843 | 0.50 | 0.623 | 0.64 | -0.006 | 1.63 |
| Baseline-offset | 5 | 0.890 | 0.45 | 0.585 | 0.66 | -0.102 | 1.58 |
| SNV | 3 | 0.885 | 0.47 | 0.511 | 0.73 | -0.082 | 1.42 |
| SNV + Detrending | 5 | 0.907 | 0.45 | 0.623 | 0.63 | -0.089 | 1.65 |
| Detrending | 5 | 0.906 | 0.44 | 0.633 | 0.62 | -0.099 | 1.68 |
| MSC | 3 | 0.885 | 0.47 | 0.509 | 0.73 | -0.084 | 1.42 |

* Selected optimum model, coefficient of determination ($R^2$), standard error of cross-validation (SECV), standard error of prediction (SEP), standard normal variate (SNV), multiplicative scatter correction (MSC) and residual prediction deviation (RPD)

**Table 4.12** Results of the PLS calibration models using at-line data as calibration set and in-line data as test set (Model #3)

| Spectrum pretreatment | Factor | Calibration | | Test | | | RPD |
|---|---|---|---|---|---|---|---|
| | | $R^2$ | SECV (%) | $R^2$ | SEP (%) | Bias (%) | |
| No pretreatment | 4 | 0.968 | 0.16 | NA | 2.78 | -15.646 | 0.39 |
| Normalized | | | | | | | |
| - Mean | 4 | 0.200 | 0.81 | NA | 1.12 | -0.136 | 0.96 |
| - Max | 3 | 0.954 | 0.19 | NA | 0.86 | -1.547 | 1.26 |
| - Range | 3 | 0.952 | 0.20 | NA | 0.32 | -9.897 | 3.38 |
| Derivatives | | | | | | | |
| - 1st derivative (5 points) | 3 | 0.961 | 0.17 | NA | 0.83 | -0.794 | 1.30 |
| - 1st derivative (11 points) | 3 | 0.962 | 0.17 | NA | 1.05 | -1.257 | 1.03 |
| - 1st derivative (21 points) | 3 | 0.962 | 0.17 | NA | 1.58 | -3.690 | 0.68 |
| **- 2nd derivative (5 points)*** | **4** | **0.958** | **0.19** | **0.667** | **0.62** | **-0.086** | **1.74** |
| - 2nd derivative (11 points) | 3 | 0.963 | 0.17 | 0.543 | 0.68 | -0.277 | 1.59 |
| - 2nd derivative (21 points) | 3 | 0.961 | 0.17 | NA | 1.96 | 0.756 | 0.55 |
| Baseline-offset | 4 | 0.965 | 0.17 | NA | 1.12 | -14.600 | 0.96 |
| SNV | 3 | 0.959 | 0.18 | NA | 0.98 | -7.590 | 1.10 |
| SNV + Detrending | 3 | 0.962 | 0.18 | NA | 0.82 | -7.414 | 1.32 |
| Detrending | 3 | 0.959 | 0.18 | NA | 1.53 | -1.527 | 0.71 |
| MSC | 4 | 0.959 | 0.18 | NA | 0.99 | -7.664 | 1.09 |

* Selected optimum model, coefficient of determination ($R^2$), standard error of cross-validation (SECV), standard error of prediction (SEP), standard normal variate (SNV), multiplicative scatter correction (MSC) and residual prediction deviation (RPD), Not applicable (NA)

**Table 4.13** Results of the PLS calibration models using the calibration set consisted of 100% of at-line data and 50% of in-line data and test set consisted of another 50% of the in-line data (Model #4)

| Spectrum pretreatment | Factor | Calibration | | Test | | | RPD |
|---|---|---|---|---|---|---|---|
| | | $R^2$ | SECV (%) | $R^2$ | SEP (%) | Bias (%) | |
| No pretreatment | 8 | 0.927 | 0.27 | 0.547 | 0.73 | 0.100 | 1.42 |
| Normalized | | | | | | | |
| - Mean | 3 | 0.111 | 0.86 | NA | 1.09 | 0.151 | 0.95 |
| - Max | 5 | 0.841 | 0.38 | 0.497 | 0.74 | 0.036 | 1.41 |
| - Range | 6 | 0.915 | 0.29 | 0.562 | 0.69 | 0.001 | 1.51 |
| Derivatives | | | | | | | |
| - 1st derivative (5 points) | 5 | 0.915 | 0.28 | 0.650 | 0.62 | -0.018 | 1.68 |
| - 1st derivative (11 points) | 5 | 0.902 | 0.30 | 0.647 | 0.62 | 0.004 | 1.68 |
| - 1st derivative (21 points) | 7 | 0.913 | 0.29 | 0.562 | 0.69 | -0.007 | 1.51 |
| - 2nd derivative (5 points) | 5 | 0.930 | 0.27 | 0.581 | 0.67 | -0.048 | 1.55 |
| **- 2nd derivative (11 points)*** | **4** | **0.912** | **0.28** | **0.658** | **0.61** | **0.001** | **1.70** |
| - 2nd derivative (21 points) | 3 | 0.881 | 0.32 | 0.642 | 0.63 | -0.013 | 1.65 |
| Baseline-offset | 6 | 0.902 | 0.31 | 0.641 | 0.63 | 0.011 | 1.65 |
| SNV | 7 | 0.943 | 0.25 | 0.636 | 0.63 | 0.024 | 1.65 |
| SNV + Detrending | 6 | 0.938 | 0.25 | 0.623 | 0.64 | 0.036 | 1.67 |
| Detrending | 7 | 0.933 | 0.27 | 0.614 | 0.65 | -0.017 | 1.60 |
| MSC | 7 | 0.943 | 0.25 | 0.639 | 0.63 | 0.016 | 1.65 |

* Selected optimum model, coefficient of determination ($R^2$), standard error of cross-validation (SECV), standard error of prediction (SEP), standard normal variate (SNV), multiplicative scatter correction (MSC) and residual prediction deviation (RPD), Not applicable (NA)

Feasibility of using at-line and/or in-line near-infrared spectroscopic models for predicting the tapioca starch moisture content at the end of the drying process was done in four different ways (see section 3.4.3). The corresponded results for predicting the moisture content of tapioca starch of PLS regression models (Model #1, #2, #3, and #4) are shown in table 4.10, 4.11, 4.12, and 4.13, respectively and the performance of the optimum models were summarized in table 4.14.

**Table 4.14** Summary of the optimum models for the four parts

| Model | | Model #1 | Model #2 | Model #3 | Model #4 |
|---|---|---|---|---|---|
| Spectrum pretreatment | | $d^2$ (5 points) | $d^2$ (5 points) | $d^2$ (5 points) | $d^2$ (11 points) |
| Factor | | 2 | 3 | 4 | 4 |
| Validation | Cross-validation | | | | |
| | $R^2$ | 0.806 | | | |
| | SECV | 0.54 | | | |
| | Bias | -0.014 | | | |
| | External-validation | | | | |
| | $R^2$ | | 0.641 | 0.667 | 0.658 |
| | SEP | | 0.62 | 0.62 | 0.61 |
| | Bias | | -0.069 | -0.086 | 0.001 |
| RPD | | 2.00 | 1.68 | 1.74 | 1.70 |

**Note:** Model #1, the in-line data using internal validation (Cross-validation)

Model #2, in-line data using external validation (Test set)

Model #3, the at-line data as calibration set and in-line data as test set

Model #4, the calibration set consisted of 100% of at-line data and 50% of in-line data and test set consisted of another 50% of the in-line data

Details of all models were illustrated in Chapter 3 (see section 3.4.3).

Second-derivative ($d^2$), coefficient of determination ($R^2$), standard error of cross-validation (SECV), standard error of prediction (SEP), and residual prediction deviation (RPD)

**Table 4.15** X (NIR spectra) and Y (Moisture) explained variance for tapioca starch moisture content models: Case of in-line experiment

| Model | PLS factor | X-explained variance (%) | Y-explained variance (%) |
|---|---|---|---|
| Model #1, $d^2$ (5 points) | 1 | 16% | 68% |
| | 2 | 49% | 12% |
| Model #2, $d^2$ (5 points) | 1 | 8% | 86% |
| | 2 | 58% | 2% |
| Model #3, $d^2$ (5 points) | 1 | 29% | 88% |
| | 2 | 10% | 4% |
| Model #4, $d^2$ (11 points) | 1 | 33% | 69% |
| | 2 | 23% | 16% |

Note: Second-derivative ($d^2$)

The appropriate number of PLS factors were selected based on the minimum value of SECV. Two, three, four and four PLS factors were used for developing the model #1, #2, #3, and #4, respectively. The first PLS factor usually accounts for the highest proportion of the total variance (i.e., the combined spectra and reference data) of the system [44]. The explained variance in the X variables (NIR spectra) and Y variable (Moisture content) by the PLS factors are shown in Table 4.15. The performance of optimum models were explained as follows:

4.4.2.1 The NIR model developed by using the in-line data (Model #1) for predicting the moisture content of tapioca starch and validated by internal validation (Cross-validation)

Figure 4.17 shows the scatter plot of the near-infrared spectroscopic model in using in-line data (Internal validation) for predicting the tapioca starch moisture content. It provided an $R^2$ of 0.806, SECV of 0.54%, bias of -0.014% and RPD of 2.00. Although Williams [44] has indicated that an $R^2$ between 0.66-0.81 implies that a model can be used for screening, cross-validation is theoretically satisfactory for the evaluation of any NIRS calibration. Because the cross-validation is carried out on the same overall sample set. Thus, this is basically the confidence's assessable for the prediction of the tapioca starch moisture content using in-line data.

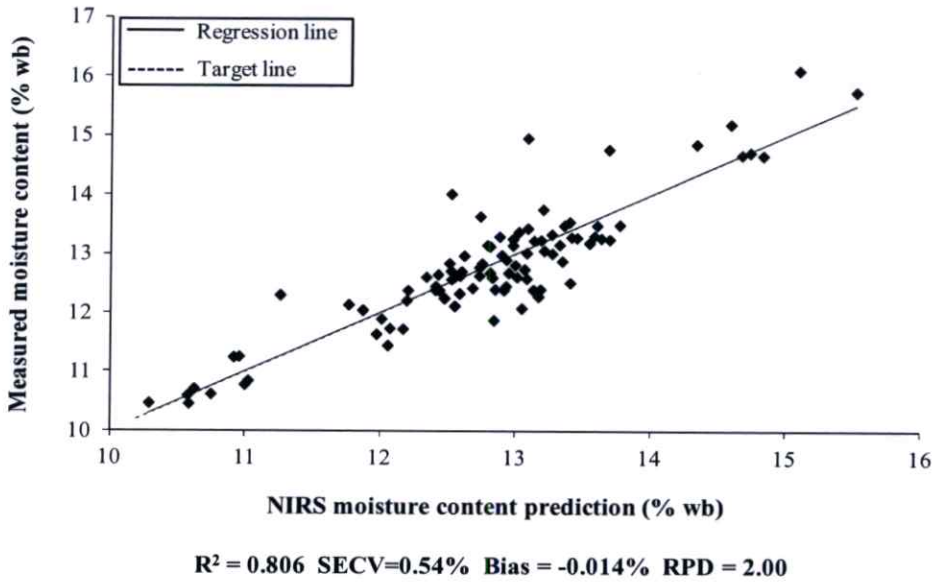



**Figure 4.17** Scatter plot of predicted moisture content with measured moisture content in tapioca starch using in-line data

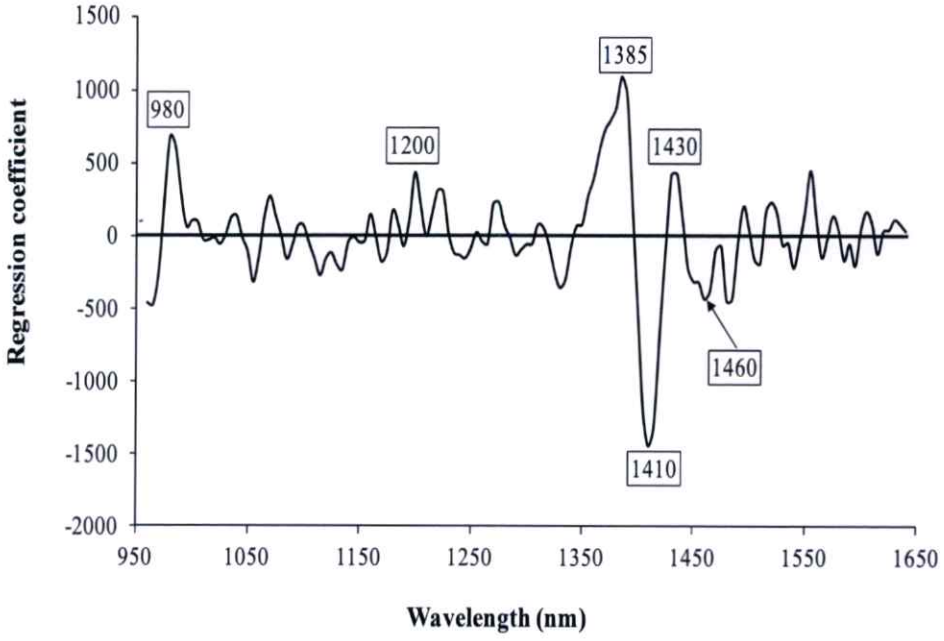



**Figure 4.18** Regression coefficient plot of optimum model in using in-line data for predicting the moisture content in tapioca starch

The regression coefficient plot of the optimum model in using in-line data (Internal validation) for the moisture content prediction of tapioca starch is shown in figure 4.18. The common bands relevant to the moisture content prediction were at approximately 980, 1200, 1385, 1430 and 1460 nm. Moreover, there was also an obvious band of the highest regression coefficient at 1410 nm. Therefore, all of them that highly influenced moisture content prediction were clearly summarized in table 4.16.

Figure 4.19 shows the X-loading weight plot of the model used in-line data (Internal validation) for the moisture content prediction of tapioca starch. The X-loading spectra of various PLS factors show that vibration bands relevant to the moisture prediction of tapioca starch model were located at 980, 1200-1210, 1385, 1410, 1430 and 1460 nm. These bands were associated with starch and water [20, 44, 59, 60].

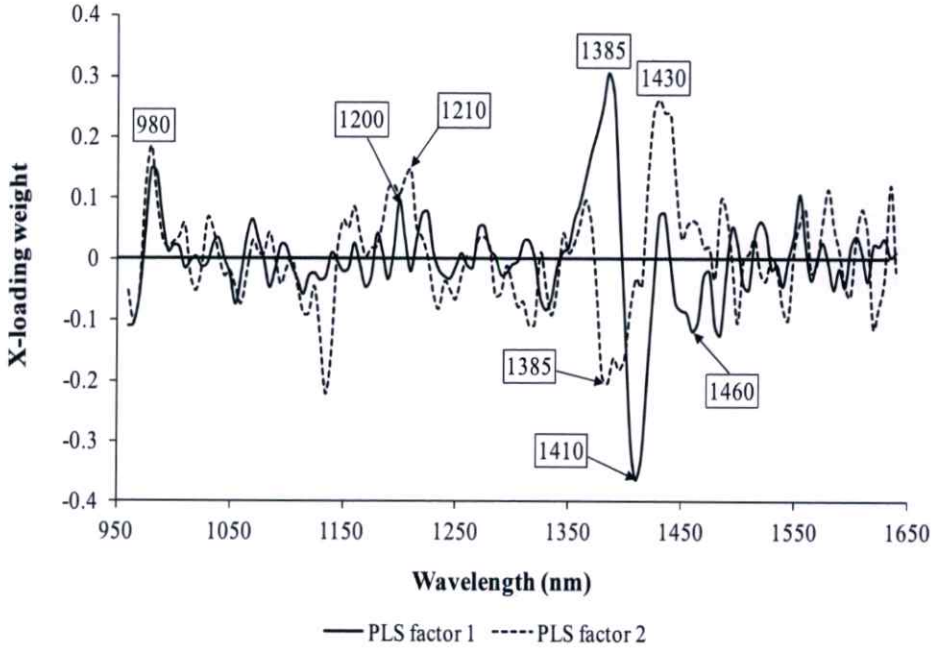



**Figure 4.19** X-loading weight plot of optimum model in using in-line data for predicting the moisture content in tapioca starch

4.4.2.2 The NIR model establishment using in-line data (Model #2) and using external validation (Test set)

The scatter plot of the near-infrared spectroscopic model in using in-line data (External validation) for predicting the tapioca starch moisture content is shown in figure 4.20. It provided an $R^2$ of 0.641, SEP of 0.62%, bias of -0.069% and RPD of 1.68. Williams [44] has indicated that an $R^2$ of between 0.50-0.64 implies that a model can be used for rough screening. Of course, high values of the RPD (Ideally 5 or more, but at least 3, [44]) indicated the efficient NIRS predictions. In this case, the samples used in the test set were the dried tapioca starch scanned in-line, which were obtained from the end of the drying process. The range of moisture content of the samples was narrow. Therefore, the standard deviation (SD) is very low (see Table 4.9).

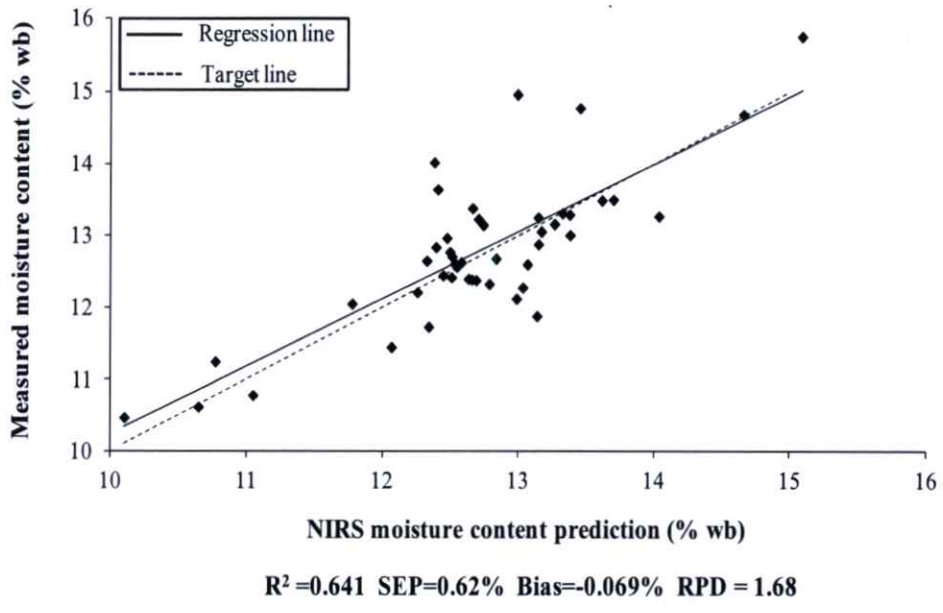



**Figure 4.20** Scatter plot of predicted moisture content with measured moisture content in tapioca starch using in-line data (external validation)

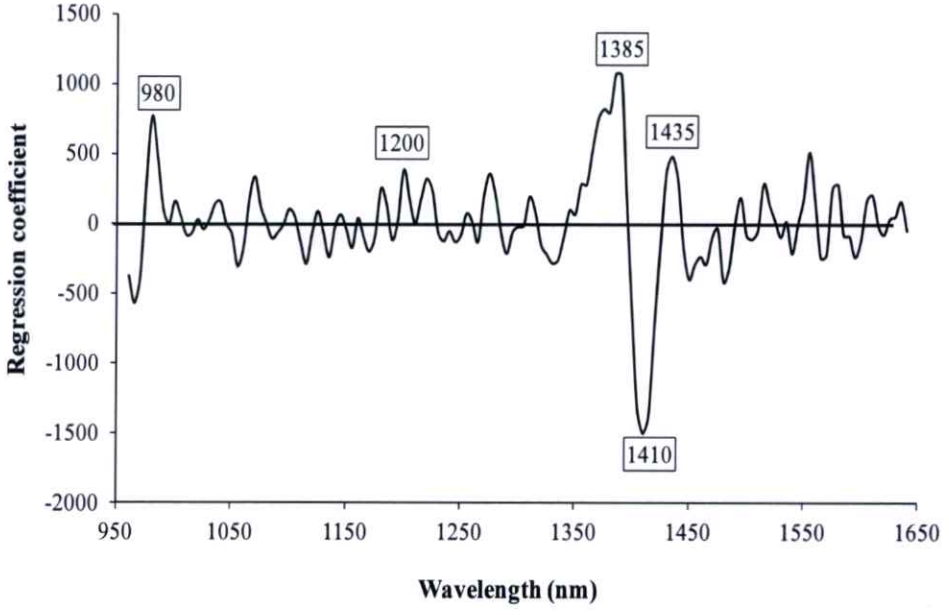



**Figure 4.21** Regression coefficient plot of optimum model in using in-line data for predicting the moisture content in tapioca starch (external validation)

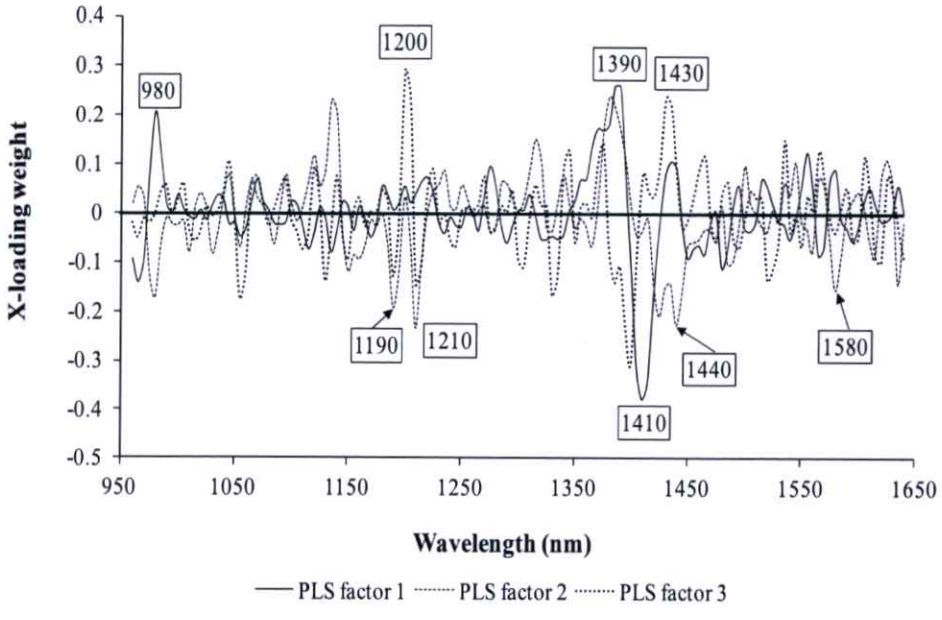



**Figure 4.22** X-loading weight plot of optimum model in using in-line data for predicting the moisture content in tapioca starch (external validation)

Figure 4.21 shows the regression coefficient plot of the optimum model in using in-line data (External validation) for the moisture content prediction of tapioca starch. The vibration bands relevant to the moisture content prediction were at approximately 980, 1200, 1385 and 1435 nm, which were similar to the bands as mentioned before (see section 4.4.2.1). Moreover, there was also an obvious band of the highest regression coefficient at 1410 nm, which might also be both the strong absorption of water in the samples [44] or is possibly associated with the glucose molecules that form the starch constituents [61].

The X-loading weight plot of the model used in-line data (External validation) for the moisture content prediction of tapioca starch is shown in figure 4.22. The X-loading spectra of various PLS factors show that vibration bands relevant to the moisture prediction of tapioca starch model, which used in-line data and was then validated using external validation (Test set), were located at 980, 1190, 1200-1210, 1390, 1410, 1430-1440 and 1580 nm. These bands were associated with starch and water [20, 44, 59, 60].

4.4.2.3 The NIR model establishment using at-line data as calibration set and in-line data as test set (Model #3)

Figure 4.23 shows the scatter plot of the near-infrared spectroscopic model in using at-line data as calibration set and in-line data as test set for predicting the tapioca starch moisture content. It provided an $R^2$ of 0.667, SEP of 0.62%, bias of -0.086% and RPD of 1.74. Williams [44] has also indicated that an $R^2$ of between 0.66-0.81 implies that a model can be used for screening. For RPD, the samples used in the test set were also the dried tapioca starch scanned in-line, which were obtained from the end of the drying process. It is the same type of situation as mentioned in section 4.4.2.2. The standard deviation (SD) was very low (see Table 4.9).

The regression coefficient plot of the optimum model in using at-line data as calibration set and in-line data as test set for the moisture content prediction of tapioca starch is shown in figure 4.24. There were absorption bands of starch and water appeared in the plot. The band at approximately 975 nm was the lowest among them. The prominent bands appeared at around 1385, 1405, 1430-1450 nm. Details of the vibration bands were summarized in table 4.16.

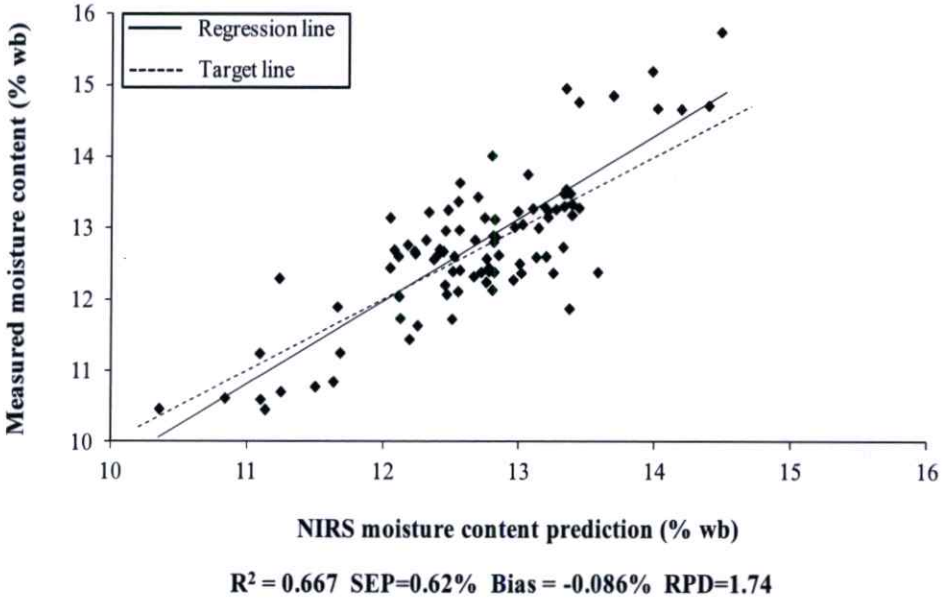



**Figure 4.23** Scatter plot of predicted moisture content with measured moisture content in tapioca starch using at-line data as calibration set and in-line data as test set

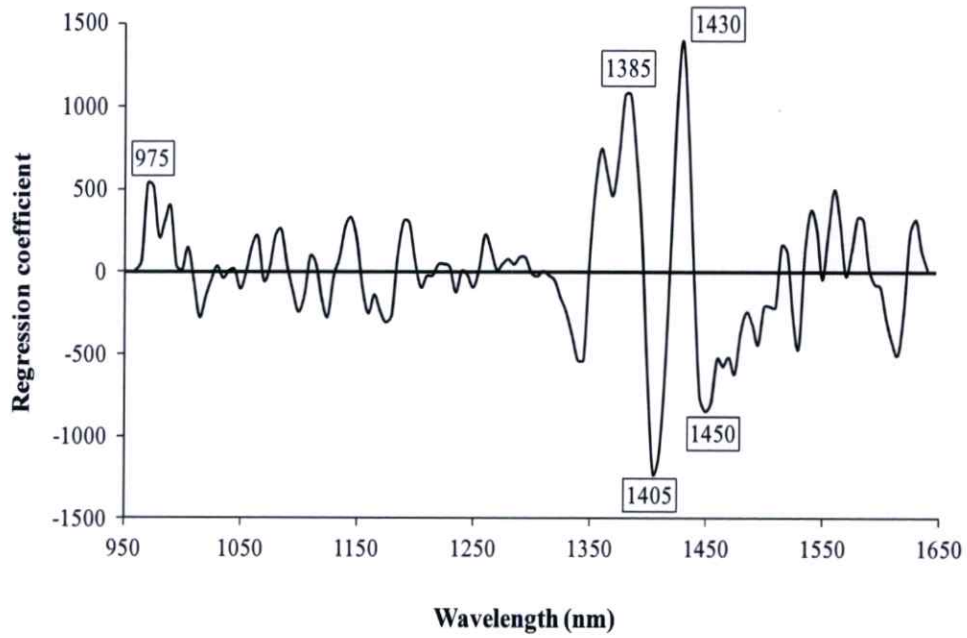



**Figure 4.24** Regression coefficient plot of optimum model in using at-line data as calibration set and in-line data as test set for predicting the moisture content in tapioca starch

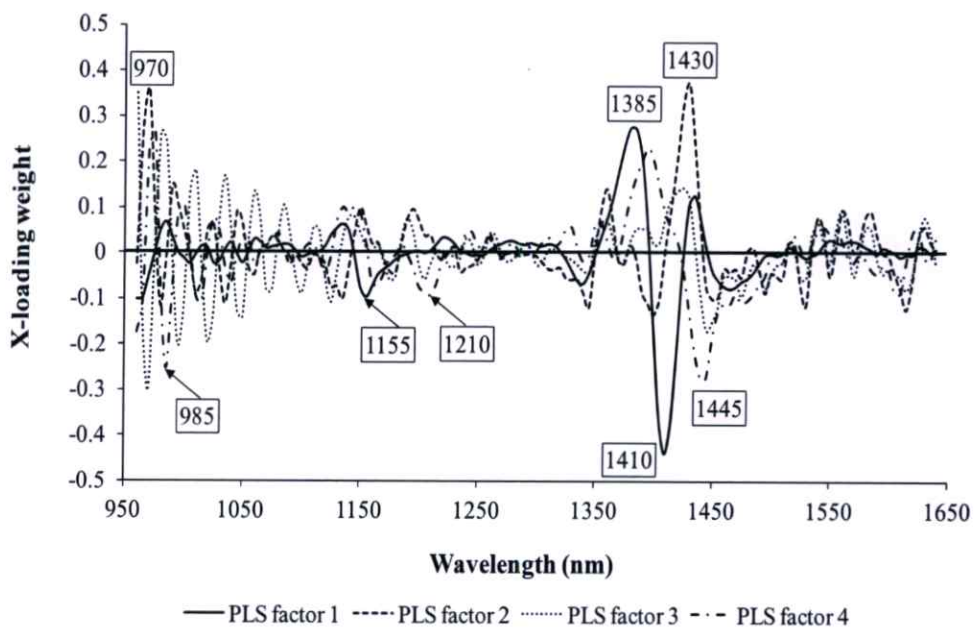



**Figure 4.25** X-loading weight plot of optimum model in using at-line data as calibration set and in-line data as test set for predicting the moisture content in tapioca starch

Figure 4.25 shows the X-loading weight plot of the model used at-line data as calibration set and in-line data as test set for predicting the moisture content in tapioca starch. The X-loading spectra of various PLS factors show that vibration bands relevant to the moisture prediction of tapioca starch model were located at 970, 985, 1155, 1210, 1385, 1410, 1430-1445 nm. These bands were associated with starch and water [20, 44, 59, 60].

4.4.2.4 The NIR model establishment using the calibration set consisted of 100% of at-line data and 50% of in-line data and test set consisted of another 50% of the in-line data (Model #4)

Figure 4.26 shows the scatter plot of the near-infrared spectroscopic model using the calibration set consisted of 100% of at-line data and 50% of in-line data and the test set consisted of another 50% of the in-line data for predicting the tapioca starch moisture content. It provided an $R^2$ of 0.658, SEP of 0.61%, bias of 0.001% and RPD of 1.70. Williams [44] has also indicated that an $R^2$ of between 0.66-0.81 implies that a model can be used for screening. In values of RPD, It is also the same type of situation as mentioned in section 4.4.2.2 and 4.4.2.3.

The regression coefficient plot of the optimum model in using the calibration set consisted of 100% of at-line data and 50% of in-line data and the test set consisted of another 50% of the in-line data for the moisture content prediction of tapioca starch is shown in figure 4.27. The vibration bands relevant to the moisture content prediction were at approximately 985, 1170, 1205, 1375, 1405 and 1430-1460 nm (see table 4.16).

Figure 4.28 shows the X-loading weight plot of the model used the calibration set consisted of 100% of at-line data and 50% of in-line data and test set consisted of another 50% of the in-line data for predicting the moisture content in tapioca starch. The X-loading spectra of various PLS factors show that vibration bands relevant to the moisture prediction of tapioca starch model were located at 980, 1200, 1380, 1395-1410, 1430-1450 and 1580 nm. These bands were associated with starch and water [20, 44, 59, 60].

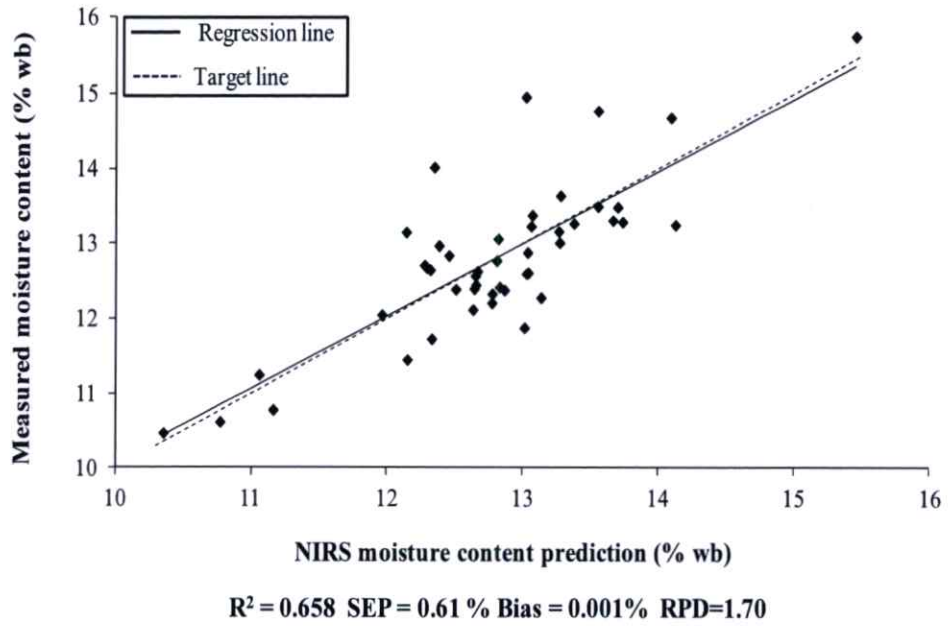



**Figure 4.26** Scatter plot of predicted moisture content with measured moisture content in tapioca starch using the calibration set consisted of 100% of at-line data and 50% of in-line data and test set consisted of another 50% of the in-line data

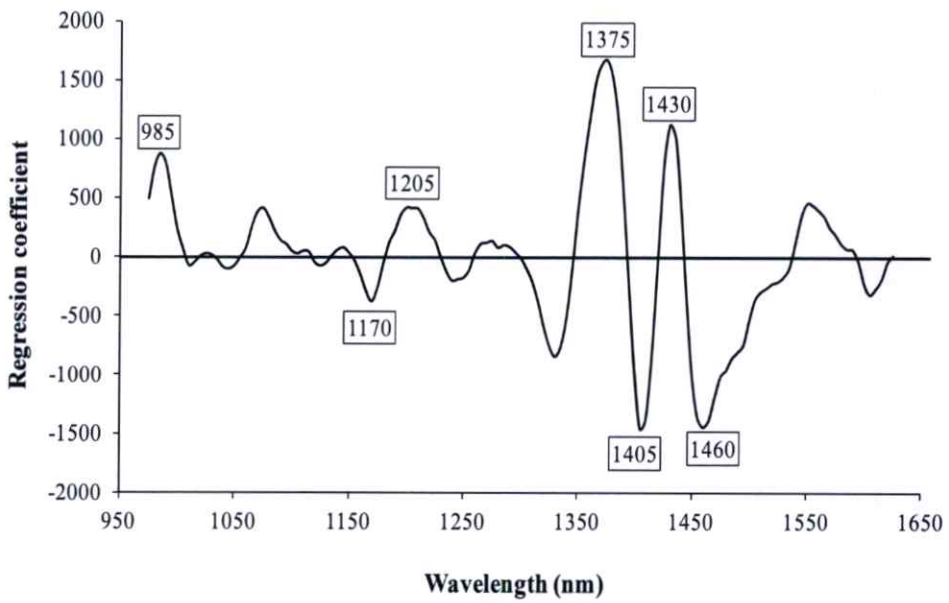



**Figure 4.27** Regression coefficient plot of optimum model in using the calibration set consisted of 100% of at-line data and 50% of in-line data and the test set consisted of another 50% of the in-line data for predicting the moisture content in tapioca starch

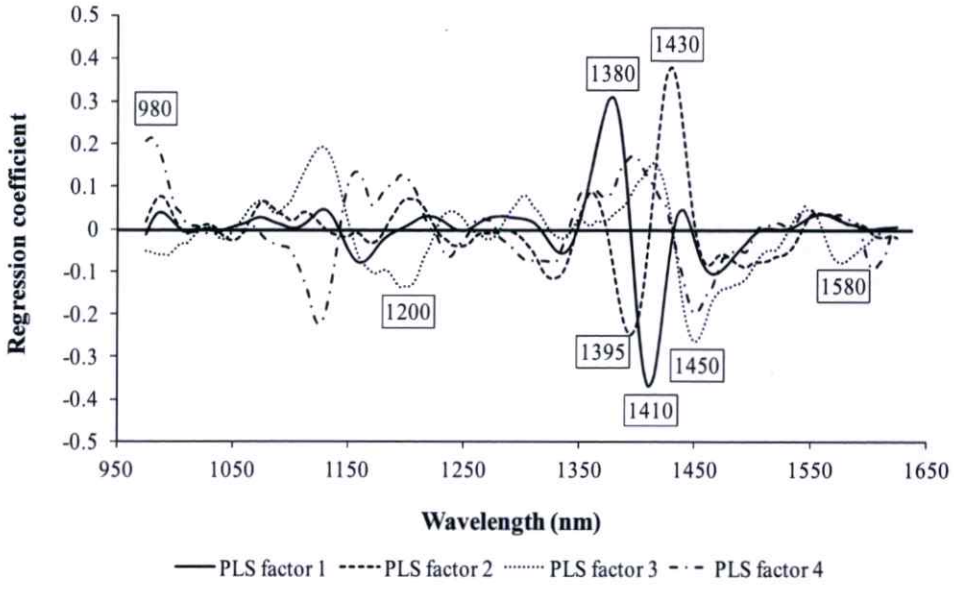



**Figure 4.28** X-loading weight plot of optimum model in using the calibration set consisted of 100% of at-line data and 50% of in-line data and the test set consisted of another 50% of the in-line data for predicting the moisture content in tapioca starch

**Table 4.16** Vibration bands of some peaks at wavelength appeared on average second-derivative spectra of tapioca starch samples, regression coefficient plot and X-loading weight plot of all models for in-line experiment

| Appeared wavelength (nm) | Cited wavelength (nm) | Vibration band | Structure | Source |
|---|---|---|---|---|
| 970, 975 | 970 [20] | O-H str. second overtone | $H_2O$ | RC, F2 |
| 980, 985 | 990 [20] | O-H str. second overtone | starch | 2AS, RC, F1, F2, F4 |
| 1155, 1170, 1190 | 1190 [62] | Absorption band | $H_2O$ | RC, F1, F2 |
| 1200, 1205, 1210 | 1202 [20] | Absorption band | starch | 2AS, RC, F1, F2, F3, F4 |
| 1375, 1380, 1385, 1390 | 1360 [20] | Absorption band | starch | RC, F1, F2, |
| 1395, 1400, 1410 | 1400 [44], [61] | Absorption band | glucose, water | RC, F1, F2, |
| 1430, 1435, 1440, 1445, 1450, 1460 | 1450 [20] | O-H str. first overtone | water, starch | 2AS, RC, F1, F2, F3, F4 |
| 1580 | 1580 [20] | O-H str. first overtone | starch | 2AS, F3 |

F1, F2, F3 and F4 are PLS factor 1, 2, 3 and 4 in X-loading plot, respectively. RC is regression coefficient plot and, 2AS is average second-derivative spectra of tapioca starch.

Moreover, the highest peak (1405-1410 nm region), which appeared in both regression coefficient and X-loading weight plot of all models, showed the high influence of absorption of water in prediction of moisture content. Other peaks also confirmed that absorption bands of starch and water influenced the tapioca starch moisture content prediction.

# Chapter 5

# Conclusions

## 5.1 Near infrared (NIR) spectroscopy model establishment for tapioca starch moisture content prediction

There were three main parts in this study including preliminary experiment, at-line experiment and in-line experiment (see section 3.4). All of them, the diode array NIR spectrometer with wavelength range of 950-1650 nm was used for NIR scanning. The diode array NIR spectrometer (DA7200, Perten, Sweden) was used in part of preliminary experiment and at-line experiment. A diode-array in-line NIR spectrometer (DA7300, Perten, Sweden) was used in part of in-line experiment. Moreover, the infrared moisture analyzer (HB43-S Halogen, Mettler Toledo, Switzerland) was used as the reference method for measuring the tapioca starch moisture content. Since, the results on the statistically comparison of the reference method and standard method (Hot air oven method) confirmed that the correlation between standard method and reference method was high with the correlation of 0.998. Therefore, calibration model developments for evaluating the moisture content of tapioca starch using the NIR spectral data in conjunction with PLS regression were summarized as follows:

### 5.1.1 Conclusion of preliminary study for the evaluation of moisture content of tapioca starch using near-infrared (NIR) spectroscopy (Preliminary experiment)

The optimum NIR model for predicting the tapioca starch moisture content showed an $R^2$, SECV, a bias and RPD of 0.997, 0.52%, -0.001% and 16.81, respectively. According to Williams [44], an $R^2$ of 0.98 or more and the RPD of 8.1 or more implies that a model can be used excellent for any application. Of course, this part is just preliminary experiment and the samples were obtained by adjusting the moisture content levels. The models were then validated using full cross-validation in order to assess the feasibility of using NIR technique for tapioca starch moisture content prediction. The results showed that the NIR spectroscopy is feasible for analysis of moisture content in the tapioca starch.

### 5.1.2 Conclusion of the evaluation of the moisture content of tapioca starch using at-line near-infrared (NIR) spectroscopy (At-line experiment)

Three groups of sample spectra were used for model development, i.e. tapioca starch cake samples, dried tapioca starch samples and combined samples (cake and dried samples). The optimum NIR models for the moisture content of tapioca starch cake, dried starch and combined samples provided an $R^2$ of 0.928-0.996, SEP of 0.15-0.65%, bias of (0.069)-(-0.062)% and RPD of 4.00-15.88. The model of dried tapioca starch samples using no pretreatment spectra had the highest prediction ability, i.e., the lowest SEP (0.15%), while the $R^2$, bias and RPD were 0.928, -0.062% and 4.00, respectively. Williams [44] has indicated that an $R^2$ of 0.92-0.96 implies that a model can be used for in most applications, including quality assurance. The best NIR model developed from dried starch samples could be used as a rapid alternative to evaluate the moisture content of tapioca starch in factory quality control laboratory and leads to the possibility of using the moisture content prediction of the starch at the end of the drying process for the process control purpose.

### 5.1.3 Conclusion of feasibility study for the evaluation of the moisture content of tapioca starch using in-line near-infrared (NIR) spectroscopy (In-line experiment)

There were four parts for model establishment in the in-line experiment (see section 3.4.3). The tapioca starch moisture content prediction at the end of the drying process was done in four different ways as follows:

5.1.3.1 The NIR model developed by using the in-line data (Model #1) for predicting the moisture content of tapioca starch and validated by internal validation (Cross-validation), provided an $R^2$ of 0.806, SECV of 0.54%, bias of -0.014% and RPD of 2.00.

5.1.3.2 The NIR model establishment using in-line data (Model #2) and using external validation (Test set), provided an $R^2$ of 0.641, SEP of 0.62%, bias of -0.069% and RPD of 1.68.

5.1.3.3 The NIR model establishment using at-line data as calibration set and in-line data as prediction set (Model #3), provided an $R^2$ of 0.667, SEP of 0.62%, bias of -0.086% and RPD of 1.74.

5.1.3.4 The NIR model establishment using the calibration set consisted of 100% of at-line data and 50% of in-line data and prediction set consisted of another

50% of the inline data (Model #4), provided an $R^2$ of 0.658, SEP of 0.61%, bias of 0.001% and RPD of 1.70.

In conclusion, the at-line spectrum can also be used for developing the calibration model for predicting the moisture content of the samples. Among these models, the model #4 shows its better performance in prediction than the others due to its very low bias. In addition, the in-line NIR protocol developed in this study could be used for rough screening of the tapioca starch moisture content with $R^2$ of 0.658.

## 5.2 Suggestion

Since a diode-array in-line NIR spectrometer (DA7300, Perten, Sweden) was installed at the end of drying process where there was a tapioca starch stream line, the installation was at the pneumatic conveying circular tube through the flat quartz window (see Fig. 3.4). This caused the occasionally un-flow starch stuck on the window. Another reason for inaccurate in-line model might be because of the uncertainty of NIR scattering on the mixing of tapioca starch and air at the end of drying process. The mixture tends to have variation in density with time. These reasons might be the problem affecting spectra data used in the NIR model development. Therefore, the more suitable place for installing the spectrometer should be at the position where stream flow is steady.

# References


[1] National Food Institute. "**Thailand Food Industry Profiles: Tapioca Industry and related product (in Thai).**" [Online]. Available : www.google.co.th/url?sa=t&rct=j&q=&esrc=s&frm=1&source=web&cd=2&ved=0CDQQFjAB&url=http%3A%2F%2Ffic.nfi.or.th%2Ffood%2Fupload%2Fdoc%2F13_135.docx&ei=6cFBUpWEHYK4rAem5YGwBg&usg=AFQjCNE9F0_teVJl30fHXoGiZsxf7gwr0Q&sig2=qYpUkYjSe0vfYUpccQGM3w. Accessed on April 7, 2014.

[2] Office of Agricultural Economics. **"Export statistics."** [Online]. Available : http://www.oae.go.th/oae_report/export_import/export_result.php. Accessed on February 1, 2015.

[3] Thailand tapioca starch newsletter. **"Analysis of energy consumption in starch drying unit (Part 3): Industrial development guideline (in Thai)."** [Online]. Available : www.thailandtapiocastarch.net/newsletters/newsletters-12.pdf. Accessed on April 7, 2013.

[4] Vesela A., Barros A.S., Synytsya A., Delgadillo I., Copikova J. and Coimbra M.A. "Infrared spectroscopy and outer product analysis for quantification of fat, nitrogen, and moisture of cocoa powder." **Analytica Chimica Acta.,** vol. 601, no. 1, 2007. Pp. 77-86

[5] Camps C., Toussirot M., Quennoz M. and Simonnet X. "Determination of *artemisinin* and moisture content of *Artemisia annua* L. dry powder using a hand-held near infrared spectroscopy device." **Journal of Near Infrared Spectroscopy.,** vol. 19, 2011. Pp. 191-198

[6] Poonpatanachai W. and Sirisomboon P. "Feasibility study for the evaluation of moisture content in tapioca starch cake by near Infrared spectroscopy." **The 3rd International Conference on Engineering, Applied Sciences, and Technology (ICEAST 2013),** The Sukosol, Bangkok, Thailand, August 21-24, 2013

[7] Thai Tapioca Starch Association. **"Crop Survey."** [Online]. Available : http://www.thaitapiocastarch.org/crop.asp?xyear=14-15. Accessed on February 24, 2015.

[8] Thai Tapioca Starch Association **"24th Anniversary The Thai Tapioca Flour Industries Trade Association."** 2000.

[9] Cambridge dictionaries online. **"Flour and starch."** [Online]. Available : http://dictionary.cambridge.org/dictionary/british. Accessed on March 5, 2015.

[10] Oxford dictionaries. **"Flour and starch."** [Online]. Available : http://www.oxforddictionaries.com/definition/english/flour. Accessed on March 5, 2015.

[11] Food focus Thailand. **"Industry-focused magazine for F&B professionals."** Oct 2013. Pp 46-47

[12] Ministry of Commerce. **"Prescribing tapioca starch as a standardized commodity and the standards of tapioca starch."** dated 29th September, 2006.

[13] Thai Tapioca Starch Association. **"The Process of Tapioca Starch Production."** [Online]. Available : http://www.thaitapiocastarch.org/products.asp. Accessed on March 7, 2015.

[14] Thai Industrial Standards Institute. **"Thai Industrial Standards, Tapioca starch."** TIS (274-1978).

[15] Mettler-Toledo. **Operating Instructions of Moisture Analyzer.** Switzerland : Mttler-Toledo AG Laboratory & Weighing Technologies

[16] Workman J. **"An Introduction to Near Infrared Spectroscopy."** [Online]. Available : http://www.spectroscopynow.com/details/education/sepspec1881education/An-Introduction-to-Near-Infrared-Spectroscopy.html?&tzcheck=1. Accessed on February 21, 2015.

[17] Reich G. "Near-infrared spectroscopy and imaging: Basic principles and pharmaceutical applications." **Advanced Drug Delivery Reviews.**, vol. 57, 2005. Pp. 1109–1143

[18] Burns D.A., Ciurczak E.W, editors. "Principles of near-infrared spectroscopy." Pp. 7-18. **Handbook of Near-Infrared Analysis**. 2nd ED. New York : Marcel Dekker Inc. 2001.

[19] Siesler H.W., Ozaki Y., Kawata S. and Heise H.M., editors. "Origin of near-infrared absorption bands." Pp. 11–41. **Near-Infrared Spectroscopy: Principles, Instruments, Applications**. Wiley-VCH Verlag GmbH, Weinheim. 2002.

[20] Osborne B.G., Fearn T. and Hindle P.H. **Practical NIR spectroscopy with applications in food and beverage analysis.** 2nd ED. UK : Longman Science & Technical. 1993.

[21] Kubelka P., Munk F. "Ein Beitrag zur Optik der Farbanstriche, Zeitshrift fur technische Physik." vol. 12, 1931. Pp. 593-604

[22] Kubelka P. "New contributions to the optics of intensely light-scattering materials Part 1." **Journal of the Optical Society of America**., vol. 38, no. 5, 1948. Pp. 448-448

[23] Kortum G., Braun W. and Herzog G. "Principles and thchniques of diffuse-reflectance spectroscopy." **Angewandke Chemie International Edition Spectroscopy**., vol. 2, no. 7, 1963. Pp. 333-341

[24] Birth G.S., Zachariah G.L. "Spectrophotometry of agricultural products, in Quality Detection in Foods Gaffney." **American Society of Agricultural Engineering**., vol. 1, no. 76, 1976. Pp. 6-11

[25] Shemilt L. W., editor. "Multivariate analysis of raw materials." Pp. 521-535. **Chemistry and World Food Supplies: the New Frontiers, Chemrawn II**. British : Pergamon Press. 1983.

[26] Siesler H.W., Ozaki Y., Kawata S. and Heise H.M., editors. "New techniques in near-infrared spectroscopy." Pp. 75–84. **Near Infrared Spectroscopy: Principles, Instruments, Applications**. Wiley-VCH Verlag GmbH, Weinheim. 2002.

[27] Sun D.W. **Infrared Spectroscopy for Food Quality Analysis and Control.** Academic Press, 2009.

[28] Metrohm A.G. **"A guide to near-infrared spectroscopic analysis of industrial manufacturing processes."** [Online]. Available : http://www.metrohm.com/com/Search/index.html?identifier=81085026EN&language=en Accessed on February 22, 2015

[29] Burns D.A., Ciurczak E.W., editors. **Handbook of Near-Infrared Analysis.** 3nd ED. New York : Taylor & Francis Group, LLC. 2008.

[30] Loewen E. G., Popov E. **Diffraction Gratings and Applications.** USA : CRC Press. 1997.

[31] Fong G. W. **HPLC in the Pharmaceutical Industry.** CRC Press, 1991.

[32] Choi H. **"Advantages of Photodiode Array."** [Online]. Available : www.hwe.oita-u.ac.jp/kiki/ronnbunn/paper_choi.pdf. Accessed on February 23, 2015.

[33] Bucher E. G., Carnahan J. W. "Characterization of an Acousto-optic Tunable Filter and Use in Visible Spectrophotometry." **Society for Applied Spectroscopy.,** vol. 53, no. 5, 1999. Pp. 603-611

[34] Perten instrument. **"DA 7200 Technology."** [Online]. Available : http://www.perten.com/Products/DA-7200-NIR-Analysis-System/Technology. Accessed on February 23, 2015.

[35] Perten instrument. **"Diode array 7300 In-line NIR measurement."** [Online]. Available : http://www.perten.com/Products/DA-7300-NIR-On-Line/Brochures. Accessed on February 23, 2015.

[36] FF Instrument. **"DA7300 NIR In-Line."** [Online]. Available : http://www.ffinstrumentation.co.nz/fooddairy/constituent-analysis/da7300-line-nir. Accessed on February 23, 2015.

[37] CAMO. **"The Unscrambler Appendices: Method References."** [Online]. Available : http://www.camo.com/TheUnscrambler/Appendices. Accessed on February 23, 2015.

[38] Huang J., Romero-Torres S. and Moshgbar M. "Practical Considerations in Data Pre-treatment for NIR and Raman Spectroscopy." **American Pharmaceutical Review**. [Online]. Available : http://www.americanpharmaceuticalreview.com/Featured-Articles/116330-Practical-Considerations-in-Data-Pre-treatment-for-NIR-and-Raman-Spectroscopy. Accessed on February 23, 2015.

[39] Buddenbaum H., Steffens M. "The Effects of Spectral Pretreatments on Chemometric Analyses of Soil Profiles Using Laboratory Imaging Spectroscopy." **Applied and Environmental Soil Science.**, 2012. doi:10.1155/2012/274903.

[40] Ruiz M, Mujica L. E, Berjaga X. and Rodellar J. "Partial least square/projection to latent structures (PLS) regression to estimate impact localization in structures" **Smart Materials and Structures.,** 2013. doi:10.1088/0964-1726/22/2/025028.

[41] Abdi H. **Partial least squares regression and projection on latent structure regression (PLS Regression)** John Wiley & Sons, Inc. doi: 10.1002/wics.051. 2010.

[42] Kurtyka B. "Validation of Multivariate Models with a Focus on Chemometric NIR Models" **IFPAC 2014 Annual Meeting Arlington, VA.,** January 23, 2014.

[43] Conzen J.P. **Multivariate Calibration: A practical guide for developing methods in the quantitative analytical chemistry.** 2nd ED. Bruker Optik GmbH. 2006.

[44] Williams P. **Near-infrared Technology–Getting the best out of light, A Short Course in the Practical Implementation of Near-infrared Spectroscopy for the User**. 5th ED. Canada : PDK Grain, Nanaimo. 2007.

[45] Xiao Z., Lai K., Du R., Shen Y., Sun X., Pan Y., Rasco B.A. and Huang Y. "Fat and moisture content in Chinese fried bread sticks: Assessment and rapid near-infrared spectroscopic method development." **Journal of Spectroscopy.,** vol. 1, no. 1, 2013.

[46] Haase N.U. "Prediction of potato processing quality by near infrared reflectance spectroscopy of ground raw tubers." **Journal of Near Infrared Spectroscopy.**, vol. 19, no. 1, 2011. Pp. 37-45

[47] Haase N.U. "Estimation of dry matter and starch concentration in potatoes by determination of under-water weight and near infrared spectroscopy." **Potato Research.**, vol. 46, no. 3-4, 2003. Pp. 117-127

[48] Ait-Kaddour A., Cuq B. "In line monitoring of wet agglomeration of wheat flour using near infrared spectroscopy." **Powder Technology.**, vol. 190, no. 1-2, 2009. Pp. 10-18

[49] Sudar R., Jurković Z., Galonja M., Turk I. and Arambašić M. "Application of near infrared transmission for the determination of ash in wheat flour." **Agriculturae Conspectus Scientificus.**, vol. 72, no. 3, 2007. Pp. 233-238

[50] Manley M., Van Zyl L. and Osborne B.G. "Using Fourier transform near infrared spectroscopy in determining kernel hardness, protein and moisture content of whole wheat flour." **Journal of Near Infrared Spectroscopy.**, vol. 10, no. 1, 2002. Pp. 71-76

[51] Dong X., Sun X. "A case study of characteristic bands selection in near-infrared spectroscopy: Nondestructive detection of ash and moisture in wheat flour." **Journal of Food Measurement and Characterization.**, vol. 7, no. 3, 2013. Pp. 141-148

[52] Van Zyl L., Manley M. and Osborne B.G. "Using different sample holders in determining protein and moisture content in whole wheat flour by means of Fourier transform near infrared (FT-NIR) spectroscopy." **South African Journal of Plant and Soil.**, vol. 18, no. 2, 2001. Pp. 50-55

[53] Peiris K.H.S., Dong Y., Bockus W.W. and Dowell F.E. "Estimation of bulk deoxynivalenol and moisture content of wheat grain samples by FT-NIR spectroscopy." **American Society of Agricultural and Biological Engineers Annual International Meeting 2013.,** vol. 2, 2013. Pp. 1244-1251

[54] Hermida M., Rodriguez N. and Rodriguez-Otero J.L. "Determination of moisture, starch, protein, and fat in common beans (*Phaseolus vulgaris L.*) by near infrared spectroscopy." **Journal of AOAC International.**, vol. 89,no. 4, 2006. Pp. 1039-1041

[55] Hong J. H., Ikeda K., Kreft I. and Yasumoto K. "Near-infrared diffuse reflectance spectroscopic analysis of the amounts of moisture, protein, starch, amylase and tannin in Buckwheat flours." **Journal of nutritional science and vitaminology.**, vol. 42, 1996. Pp. 359-366

[56] Fernàndez-Ahumada E., Garrido-Varo A., Guerrero-Ginel J.E., Wubbels A., Van der Sluis C. and Van der Mer J. "Understanding factors affecting near infrared analysis of potato constituents." **Journal of Near Infrared Spectroscopy.**, vol. 14, 2006. Pp. 27-35

[57] Hartmann R., Büning-Pfaue H. "NIR determination of potato constituents." **Potato Research.**, vol. 41, 1998. Pp. 327-334

[58] Steel R.G.D., Torrie J.H. **Principles and procedures of statistics.** 2nd ED. Singapore : McGraw-Hill. 1980.

[59] Clevers J.G.P.W., Kooistra L. "Using spectral information at the NIR water absorption features to estimate canopy water content and biomass." **Commission VII, WG VII/1 and VII/3**. [Online]. Available : http://www.isprs.org/proceedings/XXXVI/part7/PDF/124.pdf. Accessed on October 10, 2014.

[60] Curran P.J. "Estimating foliar chemical concentrations with the airborne visible/infrared imaging spectroscopy (AVIRIS)." **ISPRS Commission VII**. [Online]. Available : http://www.isprs.org/proceedings/XXIX/congress/part7/705_XXIX-part7.pdf. Accessed on October 10, 2014.

[61] Williams P. "Influence of water on prediction of composition and quality factors: the Aquaphotomics of low moisture agricultural materials." **Journal of Near Infrared Spectroscopy.**, vol. 17, 2009. Pp. 315-328

[62] Workman J., Weyer L. **Practical Guide to Interpretive Near-Infrared Spectroscopy.** USA : Taylor & Francis Group. 2008.

# Appendix

## Published papers

Journal of Innovative Optical Health Sciences
Vol. 8, No. 2 (2015) 1550014 (12 pages)

DOI: 10.1142/S1793545815500145

World Scientific
www.worldscientific.com



# Evaluation of the moisture content of tapioca starch using near-infrared spectroscopy

Kittisak Phetpan and Panmanas Sirisomboon*
*Curriculum of Agricultural Engineering*
*Department of Mechanical Engineering, Faculty of Engineering*
*King Mongkut's Institute of Technology Ladkrabang*
*Bangkok 10520, Thailand*
**kspanman@kmitl.ac.th*

Received 18 September 2014
Accepted 13 October 2014
Published 14 November 2014

The purpose of this study was to develop a calibration model to evaluate the moisture content of tapioca starch using the near-infrared (NIR) spectral data in conjunction with partial least square (PLS) regression. The prediction ability was assessed using a separate prediction data set. Three groups of tapioca starch samples were used in this study: tapioca starch cake, dried tapioca starch and combined tapioca starch. The optimum model obtained from the baseline-offset spectra of dried tapioca starch samples at the outlet of the factory drying process provided a coefficient of determination ($R^2$), standard error of prediction (SEP), bias and residual prediction deviation (RPD) of 0.974, 0.16%, −0.092% and 7.4, respectively. The NIR spectroscopy protocol developed in this study could be a rapid method for evaluation of the moisture content of the tapioca starch in factory laboratories. It indicated the possibility of real-time online monitoring and control of the tapioca starch cake feeder in the drying process. In addition, it was determined that there was a stronger influence of the NIR absorption of both water and starch on the prediction of moisture content of the model.

*Keywords*: Moisture content; tapioca starch; near-infrared spectroscopy.

## 1. Introduction

The tapioca starch industry is an economically important industry in Thailand. Thailand has been the top country in tapioca product export. Moreover, in 2011, 33.91% of the export value of the total tapioca products of Thailand was from tapioca starch.[1] The total tapioca starch export from Thailand in 2013 was 2,445,612.18 tons, corresponding to 1,162,652,368.50 US$ in value.[2]

The Thailand Tapioca Starch Newsletter[3] described the principle of pneumatic drying for the tapioca starch drying process, wherein the water in the tapioca starch cake was evaporated using hot

*Corresponding author.

air and the dried tapioca starch was conveyed to a cyclone separator, where the dried tapioca starch and saturated air were separated. At present, most tapioca starch factories use the outlet air temperature in the drying process, measured by a thermocouple, as a parameter for controlling the starch cake feed rate. When the outlet air temperature is high, indicating that the cake feed rate is too low, the controller will increase the speed of the feeder. The set point is 60°C. If the outlet temperature reaches 60°C, the controller will stop the feeder. However, the moisture content of dried tapioca starch is checked regularly during the drying process using an infrared moisture analyzer, which requires approximately 10 min per sample. When the moisture content is found to exceed the acceptable standard value (13%), the dried starch lot must be returned to the process. Thus, the starch production time is approximately 30 min (20 min of drying tube retention and 10 min of moisture content measurement). For the production capacity factory of 200 tons per 12 h, 8 tons of unacceptable dried starch must be returned to the process to be dried again. This is very expensive, with costs including the drying energy, product unpacking and repacking, material handling, and labor costs. This does not include the time required to readjust the system to the proper conditions, which must be performed by experienced workers. Based on discussions with the factory production engineers, the use of the moisture content of the dried starch as a direct control parameter of the starch cake feeder would solve this problem.

Near-infrared (NIR) spectroscopy is a rapid method for chemical component analysis. Its advantages include its high precision and accuracy, nondestructive measurements, chemical-free procedure and environmental-friendliness as well as no or minimal sample preparation. However, this technique needs a calibration model to be developed for prediction, which is calibrated with a standard reference method. If the validation of its accuracy is proved, analysis takes only 2–3 s.

Some studies use NIR spectroscopy to measure the moisture content of powders of agricultural products. Vesela *et al.*[4] reported a calibration model developed from the NIR absorbance spectra (1100–2500 nm) of cocoa powder samples, which was used for the determination of the moisture content. The results showed that the relative root mean square error of cross-validation (RMSECV) was 5.2% and the determination correlation ($R^2$) was 0.94. Camps *et al.*[5] developed a NIR spectroscopy method to determine the moisture in flour of dry *Artemisia annua* leaves, and the model accurately predicted moisture with $R^2$, RMSECV and root mean square error of prediction (RMSEP) of 0.99, 0.8% and 1.4%, respectively. Gillon *et al.*[6] found close relationships between the foliage moisture content and ground foliage absorbance in the 400–2500 nm spectral range ($R^2 = 0.93$–$0.99$, standard error of cross-validation (SECV) = 2–7%). Ren and Chen[7] used NIR reflectance spectroscopy to measure the moisture content in hot-air-dried Asian ginseng (*Panax ginseng*) roots, freeze-dried Asian ginseng roots, red Asian ginseng roots and hot-air-dried American ginseng (*Panax quinquefolium*) roots. The calibration equation showed a high correlation between the NIR and reference analytical methods ($R^2 = 0.998$), with a standard error of prediction (SEP) of 0.14% and a bias of only 0.12%. The NIR spectroscopy in diffuse reflectance mode has been used for determination of starch content, total sugar, sucrose, cellulose, total nitrogen and ash of tuber crop flour, e.g., cassava,[8] taro,[8,9] yam[8,10] and sweet potato.[8,11] However, there are few papers that are using NIR spectroscopy for determination of moisture content or dry matter of flour or starch of tuber crop. Fernandez *et al.*[12] used the Fourier transformation near-infrared (FT-NIR) spectrometer in reflectance mode with 4000–10,000 $cm^{-1}$ wavenumber region for determination of dry matter content of mashed potatoes. The calibration model showed the $R^2$ of 0.92 and SECV of 4.5. Hartmann and Büning-Pfau[13] reported the NIR model developed from the NIR absorbance spectra (1100–2500 nm) to measure the dry matter content of mashed potatoes. The model provided the $R^2$ and SEP of 0.97 and 0.19%, respectively. Based on these previous studies, NIR spectroscopy is promising for use in tapioca starch factories for the evaluation of the moisture content of the starch. In factories, the inconsistent moisture contents of starch cake and dried starch affect the process control to ensure acceptable products. Measurement using NIR spectroscopy may facilitate rapid measurement of the starch moisture content and response to the process control. It is worth researching the performance of NIR spectroscopic models using samples obtained from the real conditions of the factory production process. Therefore, this research aimed to evaluate the moisture content of both tapioca starch

cake and dried starch in the drying process of a tapioca starch factory using NIR spectroscopy. This information is useful for quality assurance and process control.

## 2. Materials and Methods

### 2.1. *Samples*

The tapioca starch samples were collected at the Sangpetch Tapioca Flour Co., Ltd., factory in Nongbua-rawe district at Chaiyaphum, Thailand, in two forms: tapioca starch cake at the inlet of the drying process and dried tapioca starch at the outlet. The 210 samples (105 for each form) were kept in plastic cups with covers and immediately brought to the factory laboratory, where the room temperature was $25 \pm 1$°C. The samples were collected over two periods: 64 samples on 24–26 August 2013 and 146 samples on 19–23 December 2013.

### 2.2. *Near-infrared spectroscopy experiment*

The spectra of tapioca starch samples were measured with a diode-array NIR spectrometer (DA7200, Perten, Sweden) in reflection mode at 950–1650 nm with a resolution of 2 nm. The NIR exposure time was 0.164 s per scan. The sample was presented to the spectrometer in a 75-mm-diameter sample dish (72.05.04, Perten, Sweden). The NIR illumination was over the dish area. The scanning was done while the sample dish was rotated consecutively for 3 rounds and a reference material (polystyrene) was automatically scanned before each sample scan. Therefore, there were 3 spectra for each sample and they were averaged, 210 spectra in total.

### 2.3. *Moisture measurement*

After scanning, the moisture content of a 5 g tapioca starch sample was immediately analyzed by the factory reference method using an infrared moisture analyzer (HB43-S Halogen, Mettler Toledo, Switzerland) at 130°C. Duplicate tests were run for each sample. Therefore, the repeatability, which was the standard deviation of the difference between duplicates, was calculated to indicate the precision of the test.

### 2.4. *Spectrum pre-treatment and NIR spectroscopy model establishment*

The NIR spectroscopic models for predicting the moisture content of tapioca starch were established using partial least squares (PLS) regression using the Unscrambler X 10.3 (Camo, Norway). Three groups of sample spectra were used for model development: those of tapioca starch cake samples, dried tapioca starch samples and combined samples (cake and dried samples). The data set of the first two groups was divided into a calibration data set (two-thirds of the data set, 70 samples) and a prediction data set (one-third of the data set, 35 samples) after the data were arranged in ascending order. Similarly, the data for the third group were 140 samples for the calibration set and 70 samples for the prediction set. This was done by assigning the first two samples to be in the calibration set and the third one to be in prediction set until all samples had been allocated. The calibration set was used for model development. The NIR spectra used for model development were either not pre-treated or pre-treated. The following pretreatments were used: mean normalization, maximum normalization, range normalization, first derivatives (5, 11 and 21 points), second derivatives (5, 11 and 21 points), baseline offset, standard normal variate (SNV), detrending, SNV + detrending or multiplicative scatter correction (MSC). The model accuracy was determined using full cross-validation, as indicated by the coefficient of determination ($R^2$) and SECV.

In external validation using the prediction set, the optimum model was selected based on the high coefficient of determination ($R^2$), low SEP and low prediction bias.

## 3. Results and Discussion

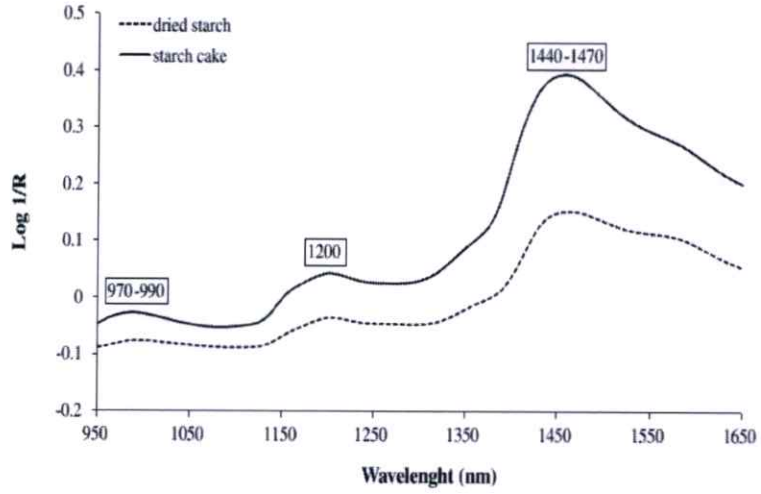
The repeatability of the traditional moisture content evaluation method was 0.45% and 0.15% for starch cake samples and dried starch samples, respectively. These values were accepted by the factory. The maximum (Max), minimum (Min), mean and standard deviation (SD) of the moisture content in tapioca starch of different groups are shown in Table 1.

Figures 1(a) and 1(b) show the average absorbance of the raw and second-derivative spectra of cake and dried tapioca starch samples, respectively. There were 3 dominant absorption regions in the

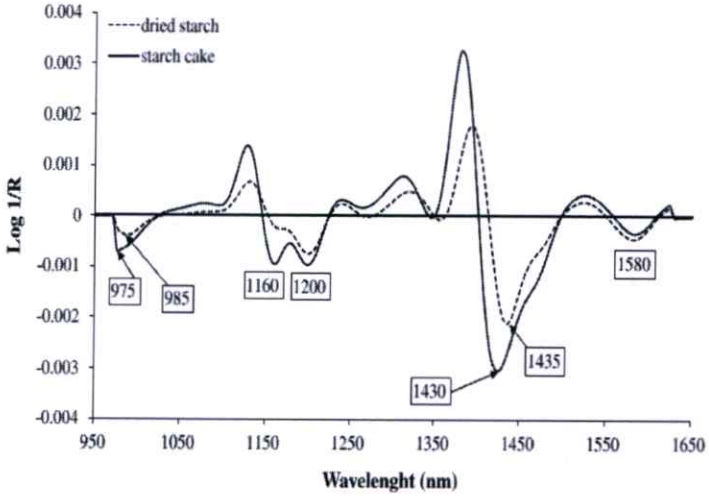
Table 1. Moisture content (% wet basis) of tapioca starch measured by the standard method used to develop the prediction model and validate the test set.

| | Calibration | | | | | Prediction | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Group | *N* | Max | Min | Mean | SD | *N* | Max | Min | Mean | SD |
| Tapioca starch cake | 70 | 39.42 | 29.32 | 33.41 | 1.84 | 35 | 41.13 | 29.93 | 33.60 | 2.09 |
| Dried tapioca starch | 70 | 14.53 | 10.91 | 12.81 | 0.66 | 35 | 18.82 | 11.39 | 12.98 | 1.18 |
| Combined tapioca starch | 140 | 39.42 | 10.91 | 23.12 | 10.43 | 70 | 41.13 | 11.39 | 23.29 | 10.52 |






(a)

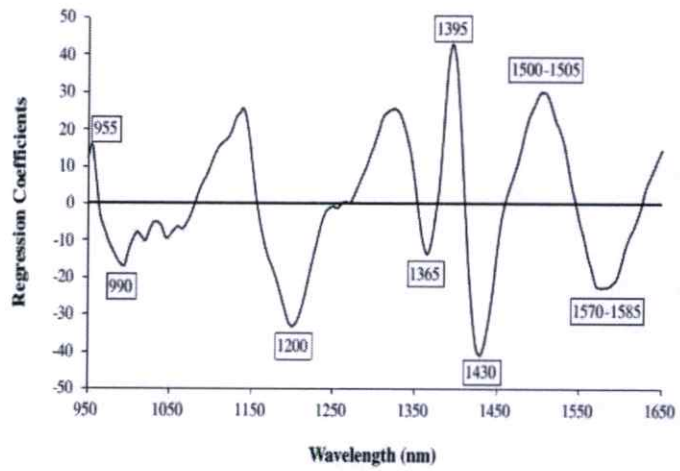



(b)

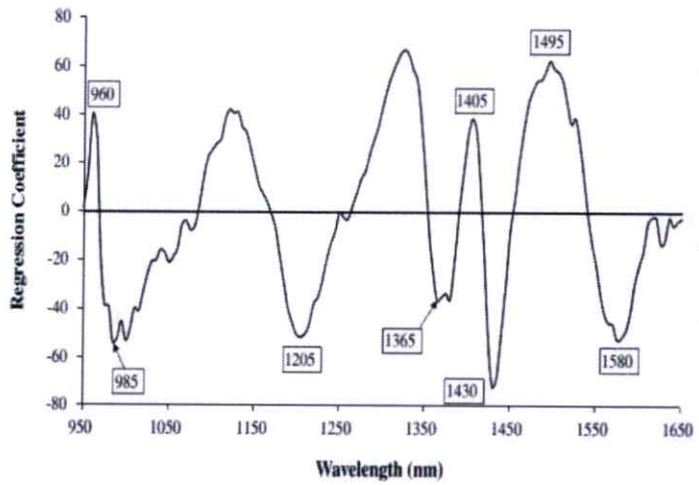
Fig. 1. Averaged NIR spectra of tapioca starch cake and dried tapioca starch. (a) Average raw absorbance spectra. (b) Average second derivative spectra (11 points).

Table 2. X (NIR spectra) and Y (Moisture) explained variance for tapioca starch moisture content models.

| Model | PLS factor | X-explained variance (%) | Y-explained variance (%) |
|---|---|---|---|
| Tapioca starch cake samples | 1 | 88 | 17 |
| (Max normalize) | 2 | 11 | 55 |
| | 3 | 0 | 9 |
| | 4 | 0 | 15 |
| Dried tapioca starch samples | 1 | 99 | 37 |
| (Baseline) | 2 | 1 | 32 |
| | 3 | 0 | 23 |
| Combined samples | 1 | 99 | 99 |
| (Second derivative, 5 points) | 2 | 0 | 1 |

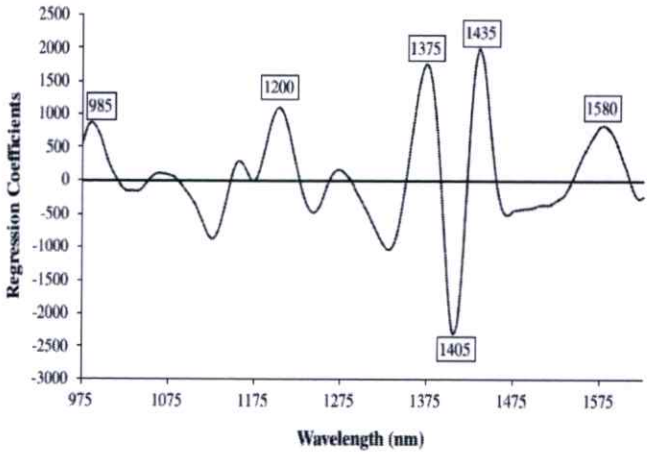



(a)




(b)

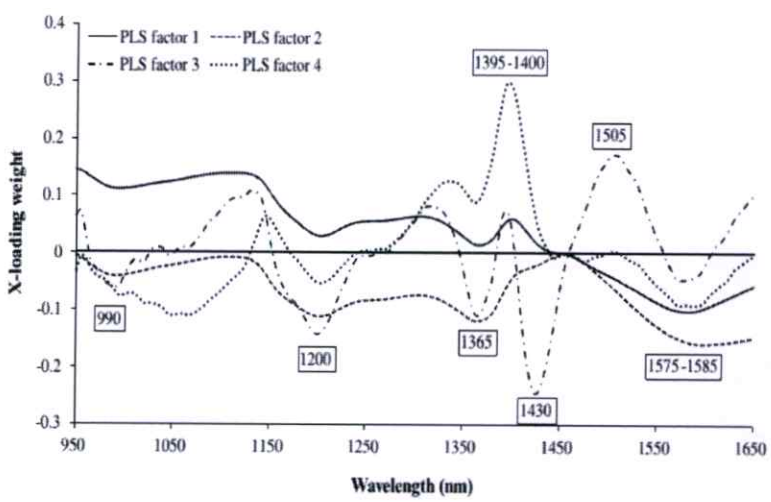
Fig. 2. Regression coefficient plots of optimum models for moisture content in tapioca starch. (a) Tapioca starch cake samples. (b) Dried tapioca starch samples. (c) Combined tapioca starch samples.




(c)

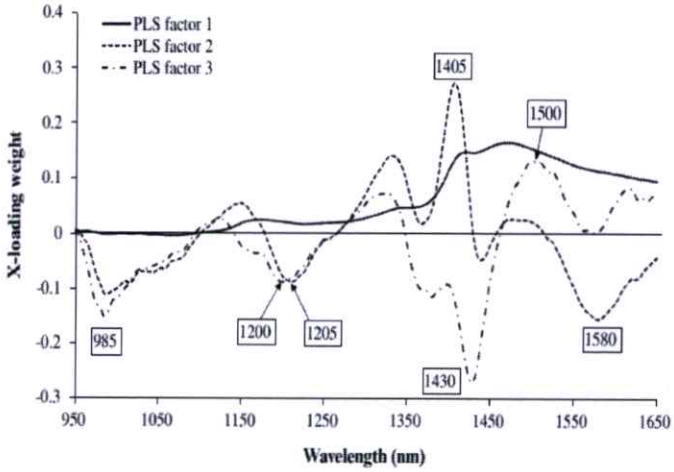
Fig. 2. (*Continued*)

raw spectra (Fig. 1(a)) i.e., 970–990, 1200 and 1440–1470 nm. In the measured range 950–1650 nm, water absorption due to the vibration of O–H bonds can be found at approximately 970, 1200 and 1450 nm.[14,15] However, according to Williams, the peak in the 1400 nm region was associated with the glucose molecules in the starch constituents.[16] In our case, the second-derivative spectra (Fig. 1(b)) shows that the tapioca starch cake has higher absorption than that of the dried tapioca starch which indicated that the peak in 1400 nm region is more due to the absorption of water in the samples. Bands at 1200 and 1580 nm were also observed in both the cake and the dried starch spectra, corresponding to the starch band (1202 nm) and the first overtone associated with the O–H stretching of starch (1580 nm).[17] This was confirmed by Fig. 1(b) where there was very slightly different of absorption intensity between starch cake and dried starch indicated pure starch absorption. From second-derivative spectra (Fig. 1(b)), there is a peak at 1160 nm which is the shifted peak of 1190 nm of

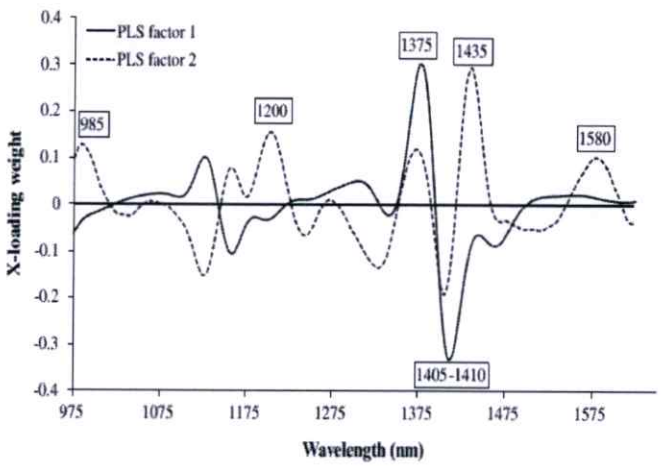



(a)

Fig. 3. X-loading weight plots of optimum models for moisture content in tapioca starch. (a) Tapioca starch cake samples. (b) Dried tapioca starch samples. (c) Combined tapioca starch samples.




(b)



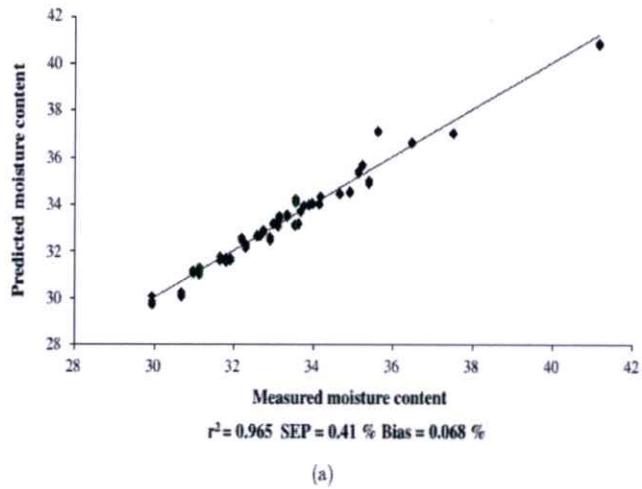

(c)

Fig. 3. (*Continued*)

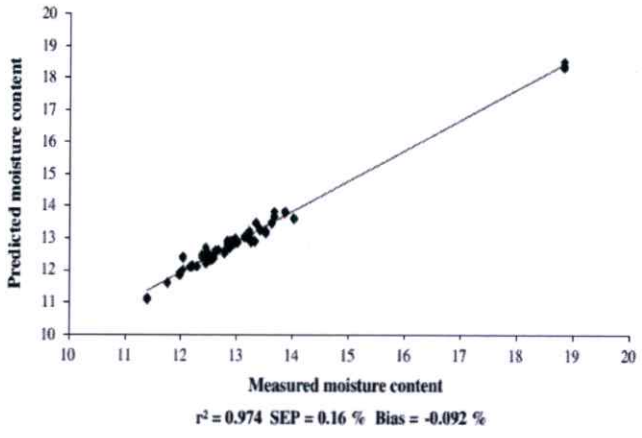
absorption band of water[18] where the absorbance intensity of starch cake is higher than that of dried starch. However, a band at 975 nm in the cake spectrum and 985 nm in dried starch spectrum could be observed, which were the second overtone associated with the O–H stretching of water (970 nm) and the second overtone associated with starch (990 nm).[17] The shift in the position of the band at 975 nm in the cake spectrum to 985 nm in the dried starch spectrum may have been a result of the removal of water during drying.

The optimum models for the three different groups were developed from the maximum-normalization, baseline-offset and second-derivative (11 points) spectra, respectively, using 950–1650 nm wavelength range. The appropriate number of PLS factors were selected based on the minimum value of SECV. The optimum models for starch cake, dried starch and combined starch samples were developed from four, three and two PLS factors, respectively. The first PLS factor usually accounts for the highest proportion of the total variance (i.e., the combined spectra and reference data) of the system.[19] The explained variance in the X variables (NIR spectra) and Y variable (moisture content) by the PLS factors are shown in Table 2.

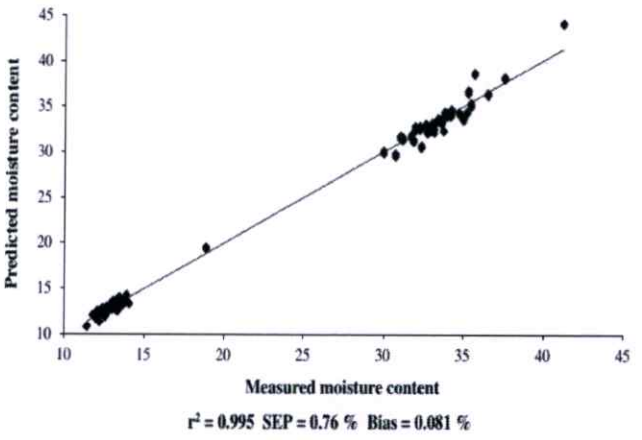
Figures 2(a)–2(c) show the regression coefficient plots of the optimum models for the moisture




(a)




(b)




(c)

Fig. 4. Validation plots of measured moisture content (X) with predicted moisture content (Y) in tapioca starch samples. (a) Tapioca starch cake samples. (b) Dried tapioca starch samples. (c) Combined tapioca starch samples.

content of starch cake, dried starch and the combined samples, respectively. The common bands relevant to the moisture prediction of the three different groups of starch models were those at approximately 985–990, 1200–1205, 1365–1375 and 1580 nm, which were, respectively, the second overtone associated with the O–H stretching of starch (990 nm), the starch bands (1202 nm) and (1360 nm) and the first overtone associated with the O–H stretching of starch (1580 nm), as indicated by Osborne *et al.*[17] Moreover, there was also an obvious band of the highest regression coefficient between 1430–1435 nm, where it was confirmed by second-derivative spectra (Fig. 1(b)) that they were the band of water. Therefore, these confirmed the high influence of O–H bands of water on the prediction of moisture content of tapioca starch.

The X-loading weights for the prediction models of the moisture content of tapioca starch are shown in Fig. 3. The X-loading spectra of various PLS factors show that vibration bands relevant to the moisture prediction of the three different groups starch models were located at 985–990, 1200–1205, 1395–1405, 1430–1435 and 1575–1585 nm. These bands were associated with starch and water.[14,15,17] It was again observed that the highest peak was at the 1430–1435 nm region showed the high influence of absorption of water in prediction of moisture content. These confirmed that O–H bands of starch and the O–H bands of water influenced the prediction of the moisture content of tapioca starch.

Figures 4(a)–4(c) show the validation plots of the NIR spectroscopic models for the moisture content of tapioca starch cake, dried starch and combined samples, respectively, which are useful for illustrating the precision of each calibration model. The optimum models provided a coefficient of determination ($R^2$) of 0.965–0.995, SEP of 0.16–0.76%, bias of (0.068)–(−0.092)% and residual prediction deviation (RPD) of 5.1–13.8. The model of dried tapioca starch samples using baseline-offset spectra (Fig. 4(b)) had the highest prediction ability, i.e., the lowest SEP (0.16%), while the $R^2$, bias and RPD were 0.974, −0.092% and 7.4, respectively. Williams[19] has indicated that an $R^2$ of between 0.66–0.81, 0.83–0.90 and 0.92–0.96 implies that a model can be used for screening, usable with caution for most applications, including research, and in most applications, including quality assurance, respectively. In addition, he also indicated that an RPD of 3.1 to 4.9, 5.0 to 6.4 and 6.5 to 8.0 implies that the model can be used for screening, for quality control, and for process control, respectively. This confirm that both values are higher two level than screening application. The results of PLS regression models for predicting the moisture content of tapioca

Table 3. Results of the PLS calibration models.

| | Calibration | | Prediction | | | |
|---|---|---|---|---|---|---|
| Spectrum pretreatment | $R^2$ | SECV | $r^2$ | SEP | Bias | RPD |
| For tapioca starch cake samples prediction | | | | | | |
| No pretreatment | 0.973 | 0.32 | 0.902 | 0.71 | 0.164 | 2.9 |
| Normalized | | | | | | |
| Mean | 0.955 | 0.41 | 0.958 | 0.42 | −0.007 | 5.0 |
| Max* | 0.967 | 0.35 | 0.965 | 0.41 | 0.068 | 5.1 |
| Range | 0.967 | 0.35 | 0.965 | 0.41 | 0.074 | 5.1 |
| Derivatives | | | | | | |
| 1st derivative (5 points) | 0.962 | 0.37 | 0.910 | 0.66 | 0.125 | 3.2 |
| 1st derivative (11 points) | 0.967 | 0.34 | 0.911 | 0.66 | 0.130 | 3.2 |
| 1st derivative (21 points) | 0.973 | 0.31 | 0.903 | 0.71 | 0.167 | 2.9 |
| 2nd derivative (5 points) | 0.937 | 0.49 | 0.909 | 0.65 | 0.103 | 3.2 |
| 2nd derivative (11 points) | 0.968 | 0.35 | 0.918 | 0.62 | 0.105 | 3.4 |
| 2nd derivative (21 points) | 0.965 | 0.35 | 0.917 | 0.63 | 0.119 | 3.3 |
| Baseline | 0.971 | 0.33 | 0.905 | 0.69 | 0.154 | 3.0 |
| SNV | 0.964 | 0.36 | 0.952 | 0.49 | 0.108 | 4.3 |
| SNV + Detrending | 0.968 | 0.34 | 0.954 | 0.48 | 0.107 | 4.4 |
| Detrending | 0.967 | 0.35 | 0.910 | 0.66 | 0.126 | 3.2 |
| MSC | 0.964 | 0.36 | 0.952 | 0.49 | 0.108 | 4.3 |

Table 3. (*Continued*)

| | Calibration | | Prediction | | | |
|---|---|---|---|---|---|---|
| Spectrum pretreatment | $R^2$ | SECV | $r^2$ | SEP | Bias | RPD |
| For dried tapioca starch samples prediction | | | | | | |
| No pretreatment | 0.942 | 0.16 | 0.963 | 0.21 | −0.086 | 5.6 |
| Normalized | | | | | | |
| Mean | 0.384 | 0.54 | 0.047 | 0.14 | −0.050 | 1.0 |
| Max | 0.947 | 0.16 | 0.914 | 0.32 | −0.123 | 3.7 |
| Range | 0.949 | 0.15 | 0.940 | 0.26 | −0.122 | 4.5 |
| Derivatives | | | | | | |
| 1st derivative (5 points) | 0.939 | 0.17 | 0.961 | 0.21 | −0.086 | 5.6 |
| 1st derivative (11 points) | 0.941 | 0.16 | 0.961 | 0.21 | −0.089 | 5.6 |
| 1st derivative (21 points) | 0.947 | 0.16 | 0.968 | 0.18 | −0.102 | 6.6 |
| 2nd derivative (5 points) | 0.939 | 0.19 | 0.959 | 0.22 | −0.076 | 5.4 |
| 2nd derivative (11 points) | 0.939 | 0.17 | 0.967 | 0.20 | −0.075 | 5.9 |
| 2nd derivative (21 points) | 0.941 | 0.16 | 0.963 | 0.21 | −0.088 | 5.6 |
| Baseline* | 0.949 | 0.16 | 0.974 | 0.16 | −0.092 | 7.4 |
| SNV | 0.941 | 0.16 | 0.931 | 0.29 | −0.105 | 4.1 |
| SNV + Detrending | 0.944 | 0.16 | 0.949 | 0.24 | −0.098 | 4.9 |
| Detrending | 0.939 | 0.17 | 0.964 | 0.20 | −0.095 | 5.9 |
| MSC | 0.941 | 0.16 | 0.931 | 0.29 | −0.105 | 4.1 |
| For combined tapioca starch samples (starch cake and dried starch) prediction | | | | | | |
| No pretreatment | 0.982 | 1.42 | 0.955 | 2.28 | 0.293 | 4.6 |
| Normalized | | | | | | |
| Mean | 0.033 | 11.95 | 0.015 | 10.39 | 0.076 | 1.0 |
| Max | 0.992 | 0.96 | 0.991 | 0.97 | −0.032 | 10.8 |
| Range | 0.994 | 0.85 | 0.994 | 0.83 | 0.023 | 12.7 |
| Derivatives | | | | | | |
| 1st derivative (5 points) | 0.995 | 0.78 | 0.992 | 0.96 | 0.127 | 11 |
| 1st derivative (11 points) | 0.993 | 0.92 | 0.988 | 1.17 | 0.157 | 9 |
| 1st derivative (21 points) | 0.981 | 1.45 | 0.952 | 2.36 | 0.356 | 4.5 |
| 2nd derivative (5 points) | 0.988 | 1.17 | 0.972 | 1.85 | 0.321 | 5.7 |
| 2nd derivative (11 points)* | 0.997 | 0.63 | 0.995 | 0.76 | 0.082 | 13.8 |
| 2nd derivative (21 points) | 0.995 | 0.79 | 0.992 | 0.98 | 0.155 | 10.7 |
| Baseline | 0.977 | 1.58 | 0.947 | 2.55 | 0.456 | 4.1 |
| SNV | 0.992 | 0.92 | 0.992 | 0.97 | 0.102 | 10.8 |
| SNV + Detrending | 0.993 | 0.90 | 0.992 | 0.96 | 0.105 | 11 |
| Detrending | 0.981 | 1.44 | 0.953 | 2.34 | 0.343 | 4.5 |
| MSC | 0.992 | 0.93 | 0.992 | 0.97 | 0.103 | 10.8 |

*Selected optimum model, coefficient of determination for calibration model ($R^2$), coefficient of determination for prediction ($r^2$), SECV, SEP, SNV, MSC and RPD.

starch of starch cake, dried starch and combined sample groups are shown in Table 3. It could be used as guidance for process control and further application in a tapioca starch factory. This result was comparable to the results of FT-NIR spectroscopy for the prediction of moisture content of whole-wheat flour, where SEP, root mean standard error of prediction (RMSEP) and correlation coefficient ($r$) values of 0.15%, 0.38% and 0.85, respectively, were obtained.[20]

## 4. Conclusion

From the results presented in this study, NIR spectroscopy could be used as a rapid and powerful alternative to evaluate the moisture content of tapioca starch in factory quality control laboratories. The model developed from dried tapioca starch samples using baseline-offset spectra performed best. The effectiveness of the model leads to the possibility of using the moisture content of the starch at the end of the drying process as an alternative control

parameter for the starch cake feeder instead of the outlet temperature of the process. The NIR-based model, which was obtained from NIR off-line spectrometer (DA7200, Perten, Sweden), developed in the study can be installed into the NIR on-line spectrometer (DA7300, Perten, Sweden) for the on-line measurement of the moisture content of tapioca starch at the end of drying process in a tapioca starch factory. In addition, it was found that the influence of the NIR absorption of starch was stronger than that of water in the prediction of the moisture content of the model.

## Acknowledgments

The authors are grateful to Sangpetch Tapioca Flour Co., Ltd., in Nongbua-rawe district in Chaiyaphum, Thailand, for providing samples and the experiment station.

## References

1. National Food Institute, "Thailand Food Industry Profiles: Tapioca Industry and related product," (in Thai) Available at http://www.google.co.th/url?sa=t&rct=j&q=&esrc=s&frm=1&source=web&cd=2&ved=0CDQQFjAB&url=http%3A%2F%2Ffic.nfi.or.th%2Ffood%2Fupload%2Fdoc%2F13_135.docx&ei=6cFBUpWEHYK4rAem5YGwBg&usg=AFQjCNE9F0_teVJl30fHXoGiZsxf7gwr0Q&sig2=qYpUkYjSe0vfYUpccQGM3w, Accessed on 7 April 2014.
2. Office of Agricultural Economics, Export statistics, Available at http://www.oae.go.th/oae_report/export_import/export_result.php, Accessed on 26 February 2014.
3. Thailand tapioca starch newsletter, "Analysis of energy consumption in starch drying unit (Part 3), Industrial development guideline (in Thai)," Available at www.thailandtapiocastarch.net/newsletters/newsletters-12.pdf, Accessed on 7 April 2013.
4. A. Vesela, A. S. Barros, A. Synytsya, I. Delgadillo, J. Copikova, M. A. Coimbra, "Infrared spectroscopy and outer product analysis for quantification of fat, nitrogen, and moisture of cocoa powder," *Anal. Chem. Acta.* **601**(1), 77–86 (2007).
5. C. Camps, M. Toussirot, M. Quennoz, X. Simonnet, "Determination of artemisinin and moisture content of Artemisia annua L. dry powder using a hand-held near infrared spectroscopy device," *J. Near Infrared Spectrosc.* **19**, 191–198 (2011).
6. D. Gillon, F. Dauriac, M. Deshayes, J. C. Valette, C. Moro, "Estimation of foliage moisture content using near infrared reflectance spectroscopy," *Agric. Forest Meteorol.* **124**, 51–62 (2004).
7. G. Ren, F. Chen, "Determination of moisture content of ginseng by near infrared reflectance spectroscopy," *Food Chem.* **60**, 433–436 (1997).
8. V. Lebot, A. Champagne, R. Malapa, D. Shiley, "NIR determination of major constituents in tropical root and tuber crop flours," *J. Agric. Food Chem.* **57**, 10539–10547 (2009).
9. V. Lebot, R. Malapa, M. Bourrieau, "Rapid estimation of taro quality by near infrared spectroscopy," *J. Food Agric. Chem.* **59**(17), 9327–9334 (2011).
10. V. Lebot, R. Malapa, "Application of near infrared reflectance spectroscopy to the evaluation of yam (*Dioscorea alata*) germplasm and breeding lines," *J. Sci. Food Agric.* **293**, 1788–1797 (2013).
11. G. Lu, H. Huang, D. Zhang, "Prediction of sweetpotato starch physio-chemical quality and pasting properties using near-infrared reflectance," *Food Chem.* **94**, 632–639 (2006).
12. E. Fernàndez-Ahumada, A. Garrido-Varo, J. E. Guerrero-Ginel, A. Wubbels, C. Van der Sluis, J. Van der Mer, "Understanding factors affecting near infrared analysis of potato constituents," *J. Near Infrared Spectrosc.* **14**, 27–35 (2006).
13. R. Hartmann, H. Büning-Pfaue, "NIR determination of potato constituents," *Potato Res.* **41**, 327–334 (1998).
14. J. G. P. W. Clevers, L. Kooistra, "Using spectral information at the NIR water absorption features to estimate canopy water content and biomass," Commission VII, WG VII/1 and VII/3, Available at http://www.isprs.org/proceedings/XXXVI/part7/PDF/124.pdf, Accessed on 10 October 2014.
15. P. J. Curran, "Estimating foliar chemical concentrations with the airborne visible/infrared imaging spectroscopy (AVIRIS)," ISPRS Commission VII, Available at http://www.isprs.org/proceedings/XXIX/congress/part7/705_XXIX-part7.pdf, Accessed on 10 October 2014.
16. P. Williams, "Influence of water on prediction of composition and quality factors: the Aquaphotomics of low moisture agricultural materials," *J. Near Infrared Spectrosc.* **17**, 315–328 (2009).
17. B. G. Osborne, T. Fearn, P. H. Hindle, *Practical NIR Spectroscopy with Applications in Food and Beverage Analysis*, 2nd Edition, Longman Science & Technical, UK (1993).
18. J. Workman, L. Weyer, *Practical Guide to Interpretive Near-Infrared Spectroscopy*, Taylor & Francis Group, USA (2008).

19. P. Williams, *Near-Infrared Technology–Getting the Best Out of Light, A Short Course in the Practical Implementation of Near-Infrared Spectroscopy for the User*, 5th Edition, PDK Grain, Nanaimo, Canada (2007).

20. M. Manley, L. V. Zyl, B. G. Osborne, "Using Fourier transform near infrared spectroscopy in determining kernel hardness, protein and moisture content of whole wheat flour," *J. Near Infrared Spectrosc.* **10**(1), 71–76 (2002).

7th International Conference on Agricultural Engineering (TSAE2014)
2nd April – 4th April 2014, Krungsri river Hotel Pranakhornsriayothanya

FPT-03

# FEASIBILITY STUDY FOR THE EVALUATION OF MOISTURE CONTENT IN TAPIOCA STARCH CAKE BY NEAR INFRARED REFLECTANCE SPECTROSCOPY

*Kittisak PHETPAN and Panmanas SIRISOMBOON
*Curriculum of Agricultural Engineering, Department of Mechanical Engineering, Faculty of Engineering, King Mongkut's Institute of Technology Ladkrabang, Bangkok 10520, Thailand.*
**Corresponding author: Kittisak PHETPAN. E-mail: kphetpan@gmail.com**

**Keywords:** Moisture content; Tapioca starch; Near Infrared Spectroscopy

## ABSTRACT

A preliminary study of rapid predictive method based on near-infrared reflectance (NIR) spectroscopy was developed to measure tapioca starch cake moisture content. The starch cake samples were scanned by diode array NIR spectrometer in reflection mode of 950-1650 nm and analyzed for moisture content by infrared moisture analyzer. Result of statistical modeling indicated that the NIR spectroscopy was reasonably accurate in predicting moisture content. The optimum model obtained from range normalize spectra with the coefficient of determination ($R^2$), root mean square error of cross validation (RMSECV), a bias and residual prediction deviation (RPD) of 0.997, 0.52 %, -0.00137 % and 16.8, respectively. The NIR-based protocol developed in this study can be used as the guidance for further application in model development to be used in the tapioca starch factory.

## INTRODUCTION

In Thailand, tapioca starch industry is one of the economical industries. There are 69 tapioca starch factories in Thailand [1]. The moisture content of the tapioca starch cake is one of the drying process control parameter. It would be excellent in practice if there is a rapid method for moisture determination which could be applied online in a drying system. Near infrared (NIR) spectroscopy technique offers a number of important advantages over traditional chemical methods. Its advantages include high precision and accuracy, non-destructive, no chemical used, environmental friendly and no or minimal sample preparation. Near infrared spectroscopy has been used for determination of moisture content of agricultural product. Colette et al. [2] reported prediction of moisture content of dedicated bio-energy crops i.e. Miscanthus (*Miscanthus* x *giganteus*), Tora (*Salix schwerinii* x *Salix viminalis*), and Karin ((*Salix schwerinii* x *Salix viminalis*) x *Salix burjatica*) using near-infrared spectroscopy which provided a root mean square error of cross validation of 0.90% (coefficient of determination, $R^2$ = 0.99). Carles et al. [3] reported feasibility of near-infrared (NIR) spectroscopy for predicting moisture parameter using a FT-NIR spectrometer with the wavelength between 12,000 and 4000 $cm^{-1}$ (833–2500 nm) in reflectance mode. Partial least squares (PLS) regression was used to process spectra and develop calibrations. Predictive models showed coefficient of determination in prediction ($R_p^2$) of 0.997 and root mean square errors of prediction (RMSEP) of 0.675. Carmen et al. [4] developed the model using NIR reflectance spectra (400–2498 nm) for prediction of moisture content of processed cheese. Calibrations to predict moisture content (37.7–54.8% w/w) developed by a partial least squares (PLS) regression procedure. Moisture content, the preferred prediction was obtained using the wavelength range between 1100 and 2498 nm (Standard error of cross validation, SECV = 0.50 and correlation coefficient, R = 0.99) using four loadings. Manley et al. [5] used Fourier transform near infrared spectroscopy in determining moisture content of whole wheat flour. Good calibration and prediction results were obtained with standard error of prediction (SEP), root mean standard error of prediction (RMSEP) and correlation coefficient (R) values of 0.15%, 0.38% and 0.85, respectively. The results confirmed that NIR spectroscopy is a useful technique for predicting moisture content.

The aim of this work is to study a preliminary of a rapid predictive method based on near-

infrared reflectance (NIR) spectroscopy for measuring tapioca starch cake moisture content. This information is very useful in predictive of moisture content of starch for the tapioca flour production factory.

## MATERIALS AND METHODS

### Samples

The tapioca starch samples of different production dates (23, 24 and 25 August 2013) were collected at Sangpetch tapioca flour Co., Ltd factory in Nongbua-rawe district at Chaiyaphum, Thailand. The samples (70 g each) were adjusted for different levels of moisture content (12.5, 20.63, 28.75 and 36.88 % wet basis) by mixing with distilled water which the amount were calculated by Equation (1), which is called starch cake. There were 24 samples in total. The experiment was done in duplicate.

$$q = [w_i(m_f - m_i)/(100 - m_f)] \quad (1)$$

Where q is the mass of distilled water to be added in grams, $w_i$ is the initial mass of the sample in grams, $m_i$ is the initial moisture content in % wet basis, and $m_f$ is the final moisture content in % wet basis.

### Near infrared spectroscopy experiment

The diode array NIR spectrometer (DA7200, Perten, Sweden) in reflection mode of 950-1650 nm with resolution of 2 nm was used for starch cake samples scanning. The sample was presented to the spectrometer in 75 mm diameter sample dish (72.05.04, Perten, Sweden). The reference material (Polystyrene) automatically was scanned before every sample scanning. Each sample was measured for 3 scans. There were 72 spectra in total.

### Moisture measurement

The moisture content of 5 grams starch cake sample was analyzed by infrared moisture analyzer (HB43-S Halogen, Mettler Toledo, Switzerland) at the temperature of 130°C.

### Chemometric

The partial least squares regression (PLS) was used for the NIR spectroscopic models establishment for predicting the moisture content of starch cake using The Unscrambler v. 9.8 (Camo, Norway) software. The NIR spectra used for model development were no pre-treatment or pre-treated in the following way, mean normalization, maximum normalization, range normalization, first derivatives (5, 11 and 21 points), second derivatives (5, 11 and 21 points), baseline offset, standard normal variate (SNV), detrending, SNV+detrending and multiplicative scatter correction (MSC). After model development there was no outlier identified by the software. The models were then validated using full cross validation. The optimum model was selected by coefficients of determination ($R^2$), root mean squared error of cross validation (RMSECV), a bias and residual prediction deviation (RPD).

## RESULTS AND DISCUSSION

Table 1 shows the statistic of the moisture content (% wet basis) of starch cake measured by the reference method used in developing prediction model. Table 2 shows the result of partial least square (PLS) regression models for predicting the moisture content of starch cake. The optimum models for starch cake samples were developed from the range normalization.

Figure 1 shows the scatter plot of calibration model. The model showed that the coefficient of determination ($R^2$), root mean square error of cross-validation (RMSECV), a bias and residual prediction deviation (RPD) of 0.997, 0.52 %, -

0.00137 % and 16.8, respectively. According to Williams [6], an $R^2$ of 0.98 or more and the RPD of 8.1 or more implies that a model can be used excellent for any application.

The regression coefficient plot of optimum model for moisture content in starch cake is shown in Figure 2. There were the bands of water and starch appeared on the plot. The peak of the starch cake spectrum at 1410 nm is possibly associated with the glucose molecules that form the starch constituents [7]. There were the peaks at 960, 1200 and 1580 nm which are the second overtone associated with O-H stretching of water (970 nm) and that of the bands of 1200 and 1580 nm were due to the starch band and absorption band of the first overtone associated with O-H stretching of starch, respectively [8].

Only one PLS latent variable was used in the optimum model of starch cake. The highest proportion of the total variance of the system is normally accounted by the first PLS factor [6]. The explained variance in X variables was 100 % (NIR spectra) and in Y variable (moisture content) was 99 %. The X-loading weight for prediction model of moisture content of starch cake is shown in Figure 3. The positive peaks at 960 and 1580 nm were observed which was the second overtone associated with O-H stretching of water and absorption band of the first overtone associated with O-H stretching of starch, respectively [8].

Table 1. The statistic of the moisture content (% wet basis) of tapioca starch measured by the reference method.

| Number of samples | Maximum | Minimum | Mean | SD |
|---|---|---|---|---|
| 24 | 36.07 | 12.43 | 23.99 | 8.78 |

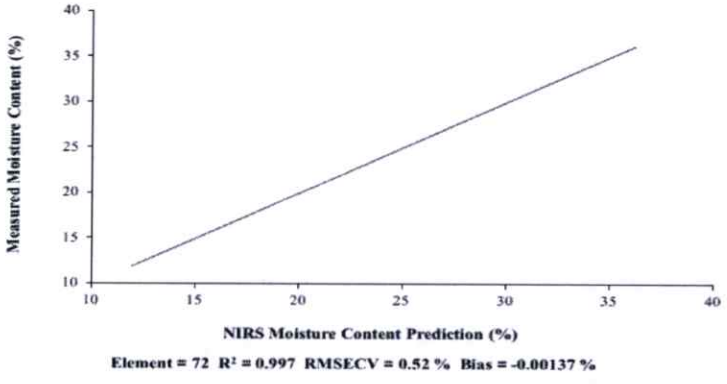



Figure 1. The scatter plot of prediction data moisture content (X) with reference data (Y) in tapioca starch cake

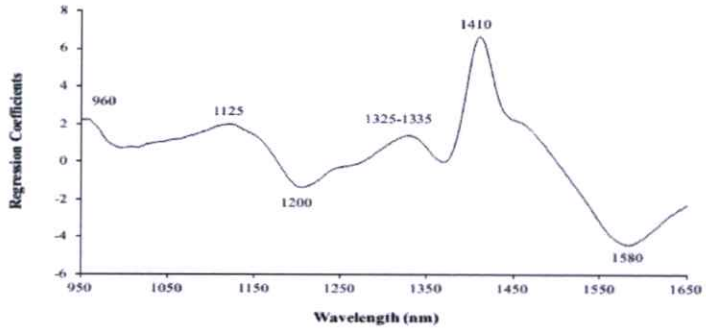



Figure 2. Regression coefficient plot of optimum model for moisture content in tapioca starch cake.

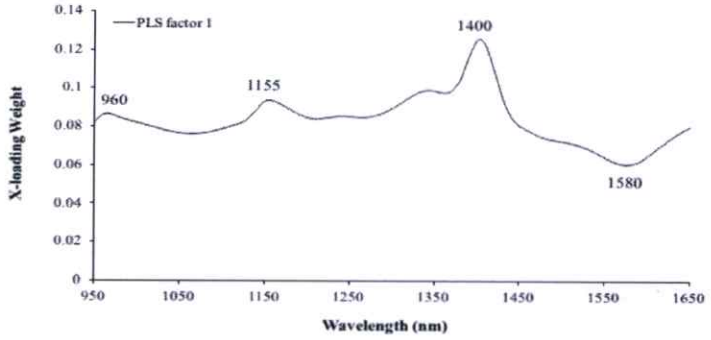



Figure 3. X-loading weight plot of optimum model for moisture content in tapioca starch cake.

Table 2. The results of calibration model of tapioca starch prediction.

| Spectrum pretreatment | Calibration model | | | | RPD |
|---|---|---|---|---|---|
| | $R^2$ | RMSECV | SECV | Bias | |
| No-pretreatment | 0.982 | 1.16 | 1.17 | -0.0021 | 7.5 |
| Normalize | | | | | |
| - Mean | 0.972 | 1.47 | 1.48 | -0.0021 | 5.9 |
| - Max | 0.988 | 0.95 | 0.95 | -0.0044 | 9.2 |
| - Range* | 0.997 | 0.52 | 0.52 | -0.0014 | 16.8 |
| Derivative | | | | | |
| - 1st derivative (5 points) | 0.995 | 0.62 | 0.62 | 0.0018 | 14.1 |
| - 1st derivative (11 points) | 0.994 | 0.70 | 0.71 | 0.0038 | 12.4 |
| - 1st derivative (21 points) | 0.990 | 0.88 | 0.89 | 0.0047 | 9.9 |
| - 2nd derivative (5 points) | 0.990 | 1.05 | 1.06 | 0.0095 | 8.3 |

| | | | | | |
|---|---|---|---|---|---|
| - 2nd derivative (11 points) | 0.996 | 0.57 | 0.58 | -0.0002 | 15.3 |
| - 2nd derivative (21 points) | 0.994 | 0.69 | 0.70 | 0.0029 | 12.6 |
| Baseline | 0.980 | 0.91 | 0.91 | 0.0028 | 9.6 |
| SNV | 0.993 | 0.74 | 0.74 | -0.0011 | 11.9 |
| SNV+De-trending | 0.993 | 0.71 | 0.71 | -0.0016 | 12.3 |
| De-trending | 0.991 | 0.81 | 0.82 | 0.0052 | 10.7 |
| MSC | 0.993 | 0.73 | 0.74 | -0.0011 | 11.9 |

* Selected optimum model, Coefficient of determination ($R^2$), Root mean square error of cross-validation (RMSECV), Standard error of cross-validation (SECV), Standard normal variate (SNV), Multiplicative scatter correction (MSC) and Residual prediction deviation (RPD)

## CONCLUSIONS

From the results presented in this study, near infrared (NIR) spectroscopy is feasible for analysis of moisture content in the tapioca starch. This information is very useful for the tapioca starch factory for further robust model development for applying the NIR spectroscopy technique in the process control of the factory.

## ACKNOWLEDGMENT

The authors are grateful for Sangpetch tapioca flour Co., Ltd factory in Nongbua-rawe district at Chaiyaphum, Thailand, for providing samples.

## REFERENCES

1. Department of foreign trade, Ministry of Commerce. Available online, Accessed on 29 January 2013, http://www.dft.go.th/Default.aspx?TabId=168&ctl=DetailUserContent&mid=685&contentID=393

2. Colette, C.F., Colm, D.E. and Kevin, M. Prediction of moisture, calorific value, ash and carbon content of two dedicated bioenergy crops using near-infrared spectroscopy. Bioresource Technology. 102: 5200–5206 (2011)

3. Carles, C., Pere, G., Pierre, P., Jacint, A. and Josep, C. Feasibility of near-infrared spectroscopy to predict $a_w$ and moisture and NaCl contents of fermented pork sausages. Meat Science. 85: 325–330 (2010)

4. Carmen, B., Gerard, D., Colm, O.D., Donal, O.C. and Vincent, H. Prediction of moisture, fat and inorganic salts in processed cheese by near infrared reflectance spectroscopy and multivariate data analysis. Journal of Near Infrared Spectroscopy. 12(3), 149–158 (2004) doi: 10.1255/jnirs.420

5. Manley, M., Van Zyla, L. and Osborneb, B.G. Using Fourier transform near infrared spectroscopy in determining kernel hardness, protein and moisture content of whole wheat flour. Journal of Near Infrared Spectroscopy. 10(1), 71–76 (2002) doi: 10.1255/jnirs.323

6. Williams, P. Near-infrared Technology–Getting the best out of light. A Short Course in the Practical Implementation of Near-infrared Spectroscopy for the




User, Edn 5.0. PDK Grain, Nanaimo, Canada (2007)

7. Williams, P. Influence of water on prediction of composition and quality factor: the Aquaphotomics of low moisture agricultural materials. Journal of Near Infrared Spectroscopy. 17(6), 315–328 (2009) doi: 10.1255/jnirs.862

8. Osborne, B.G., Fearn, T. and Hindle, P.H. Practical NIR spectroscopy with applications in food and beverage analysis, 2nd Ed., Longman Science & Technical, UK (1993)

# Author biography

| | |
|---|---|
| Name-Surname | Acting Sub Lt. Kittisak Phetpan |
| Date of birth | 11 December 1989 |
| Current address | 113/1 102 room Kakeingam soi 1 Khunhyingliam Rd. Ladkrabang, Bangkok, 10520 |
| Contact address | 72/1 M.9 Khuankalong, Khuankalong district, Satun 91130 |
| Email address | kphetpan@gmail.com |
| Contact number | 099-3242127, 081-4785862 |
| Education background | 2012 Bachelor of Engineering (Agricultural Engineering) (1st class) King Mongkut's Institute of Technology Ladkrabang, Chumphon Campus,Thailand<br>2014 Master of Engineering (Agricultural Engineering) King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand |
| Area of expertise | Near infrared spectroscopy technique, Machinery design and development |
| Publications | K. Phetpan and P. Sirisomboon, "Evaluation of the moisture content of tapioca starch using near-infrared spectroscopy", Journal of Innovative Optical Health Sciences, (2015).<br>K. Phetpan, V. Udompetaikul and P. Sirisomboon, "Study of moisture content prediction in tapioca starch drying process using in-line near-infrared spectroscopy", (Under manuscript writing).<br>P. Sirisomboon, K. Kaewsorn, K. Satthawat and K. Phetpan, "Some texture characteristics and stress relaxation behavior of cooked white, brown and germinated brown Thai Jasmine rice under large strain", (Under manuscript writing). |

Conferences

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).