



รายงานวิจัยฉบับสมบูรณ์

โครงการ การติดตามคุณภาพของน้ำมันประกอบอาหารโดยใช้
ระบบปัญญาประดิษฐ์

โดย ดร. ศีลา กิตติวัชนะ และคณะ

เมษายน 2559

รายงานวิจัยฉบับสมบูรณ์

โครงการ การติดตามคุณภาพของน้ำมันประกอบอาหาร โดยใช้ระบบปัญญาประดิษฐ์

ผู้วิจัย

สังกัด

- | | |
|--------------------------|-------------------------------------|
| 1. ดร. ศิลา กิตติวัชนะ | คณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่ |
| 2. รศ. ดร. จุฑา จักร์มณี | คณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่ |

สนับสนุนโดย สำนักงานกองทุนสนับสนุนการวิจัย และมหาวิทยาลัยเชียงใหม่

(ความเห็นในรายงานนี้เป็นของผู้วิจัย สกว. ไม่จำเป็นต้องเห็นด้วยเสมอไป)

สารบัญ

	หน้า
สารบัญ	i
สารบัญรูป	iii
สารบัญตาราง	v
1. บทคัดย่อภาษาไทย และภาษาอังกฤษ	3
2. ที่มาและความสำคัญของปัญหา-ทบทวนวรรณกรรม	5
3. วัตถุประสงค์	11
4. วิธีทดลอง	12
4.1 ขั้นตอนและวิธีการทดลอง	12
4.1.1 การเตรียมอุปกรณ์ และการเก็บตัวอย่างน้ำมัน	12
4.1.2 วิธีการวิเคราะห์ตัวอย่างน้ำมันด้วยเทคนิค NIR	13
4.2 การวิเคราะห์ด้วยเทคนิคเคโมเมทริกซ์	16
4.2.1 การวิเคราะห์องค์ประกอบหลัก (Principal component analysis, PCA)	16
4.2.2 เทคนิคพาร์เชียลลีสแควร์ (partial least squares, PLS)	18
4.2.3 การวิเคราะห์ข้อมูลด้วยเทคนิค self organizing maps (SOM)	19
5. ผลการทดลอง	23
5.1 ผลที่ได้จากการวิเคราะห์ด้วยเทคนิค NIR spectroscopy	23
5.2 ผลที่ได้จากการทำ data preprocessing ได้แก่ normalization, log scaling, standardization และ mean centring โดยวิเคราะห์ด้วยเทคนิค PCA	26
5.2.1 ผลของตัวอย่างน้ำมันที่ใช้อากาศเป็นตัวอ้างอิง	26
5.2.2 ผลของตัวอย่างน้ำมันที่ใช้ไขมันใหม่เป็นตัวอ้างอิง	29
5.2.3 ผลจากการวิเคราะห์ด้วยเทคนิค PCA และ log scaling	31
5.3 ผลจากการวิเคราะห์ด้วยเทคนิค PLS	31
5.3.1 ผลที่ได้จากข้อมูลที่ใช้อากาศเป็นตัวอ้างอิง	32
5.3.2 ผลที่ได้จากข้อมูลที่ใช้ไขมันใหม่เป็นตัวอ้างอิง	34
5.4 การสร้างแบบจำลองเพื่อติดตามคุณภาพน้ำมันด้วยเทคนิค SOM	37
5.4.1 การสร้างแบบจำลองเพื่อติดตามคุณภาพน้ำมันด้วยเทคนิค multiple SOMs	38

สารบัญ (ต่อ)

	หน้า
6. สรุปและวิจารณ์ผลการทดลอง และข้อเสนอแนะสำหรับงานวิจัยในอนาคต	43
6.1 การวิเคราะห์ด้วยเทคนิค PCA และ log scaling	43
6.2 การวิเคราะห์ด้วยเทคนิค PLS	43
7. ภาคผนวก	45
7.1 เคมีเมทริกซ์ (Chemometrics)	45
7.2 Univariate data analysis และ Multivariate data analysis	47
7.3 Principal Component Analysis (PCA)	50
8. Output จากโครงการวิจัยที่ได้รับทุนจาก สกว.	59

สารบัญรูป

	หน้า
รูป 2.1 การเปลี่ยนแปลงพันธะของโมเลกุลในรูปแบบต่างๆ	7
รูป 2.2 ตัวอย่าง กราฟ NIR spectra ของตัวอย่างน้ำมัน ช่วงความยาวคลื่น 860-1490 nm	8
รูป 2.3 กระบวนการวิเคราะห์ข้อมูลด้วย PLS	9
รูป 4.1 ขั้นตอนการเตรียมน้ำมันตัวอย่าง	12
รูป 4.2 เครื่อง UV-vis spectroscopy ที่สามารถวัดในช่วง Near infrared (NIR) ได้	13
รูป 4.3 ตัวอย่างการ project ข้อมูลลงบนแกน PC1 และค่า variation	16
รูป 4.4 ผลที่ได้จากการวิเคราะห์ข้อมูล (data) ด้วย PCA ประกอบด้วย score, loading และ residue	18
รูป 4.5 องค์ประกอบของ SOM ที่มีขนาด $P \times Q$ และใช้จำนวนพารามิเตอร์เท่ากับ J	20
รูป 4.6 ขั้นตอนการเรียนรู้ของ SOM ที่กำหนดให้ map unit มีรูปทรงหกเหลี่ยม	21
รูป 4.7 การประยุกต์ใช้ multiple SOMs สำหรับข้อมูลที่มี 3 กลุ่ม (class membership)	22
รูป 5.1 สเปกตรัมของข้อมูลน้ำมันที่ใช้ประกอบอาหารทั้งหมดที่ได้จากการวิเคราะห์ด้วย NIR spectroscopy (ก) ใช้อากาศเป็นตัวอ้างอิง (ข) ใช้น้ำมันใหม่เป็นตัวอ้างอิง	23
รูป 5.2 สเปกตรัมของข้อมูลตัวอย่างน้ำมันที่วิเคราะห์ด้วย NIR spectroscopy ช่วงความยาวคลื่น 557 – 1500 nm (ก) ใช้อากาศเป็นตัวอ้างอิง (ข) ใช้น้ำมันใหม่เป็นตัวอ้างอิง	24
รูป 5.3 การทำ moving average 3, 5 และ 7 ของตัวอย่างน้ำมันตัวที่ 10	24
รูป 5.4 การทำ moving average (MA = 5) ของข้อมูลตัวอย่างน้ำมันทั้งหมด (ก) ใช้อากาศเป็นตัวอ้างอิง (ข) ใช้น้ำมันใหม่เป็นตัวอ้างอิง	25
รูป 5.5 การทำ baseline correction ของข้อมูลตัวอย่างน้ำมันทั้งหมดในช่วง 557 – 1500 nm (ก) ใช้อากาศเป็นตัวอ้างอิง (ข) ใช้น้ำมันใหม่เป็นตัวอ้างอิง	26
รูป 5.6 การทำ data preprocessing (ก) ข้อมูลที่ไม่มีการทำ data preprocessing (ข) การทำ normalization (ค) การทำ log scaling (ง) การทำ standardization และ (จ) การทำ mean centring	28
รูป 5.7 การทำ data preprocessing (ก) ข้อมูลที่ไม่มีการทำ data preprocessing (ข) การทำ normalization (ค) การทำ log scaling (ง) การทำ standardization และ (จ) การทำ mean centring	29

สารบัญรูป (ต่อ)

	หน้า
รูป 5.8 ผลการทำ PCA และ log scaling ของข้อมูลตัวอย่างน้ำมันทั้งหมดกำหนดให้	31
รูป 5.9 (ก) score plot ของข้อมูล NIR ที่ผ่านการทำ mean centring (ข) แบบจำลอง SOM ที่แสดงด้วยเทคนิค supervised color shading	37
รูป 5.10 (ก) แบบจำลอง SOM ที่แสดงด้วยเทคนิค u-matrix โดยตัวเลขแสดงจำนวนวันของการใช้น้ำมันซ้ำ และ (ข) แสดง component plane ของค่าการดูดกลืนแสงในช่วงความยาวคลื่นเท่ากับ 1100 nm	38
รูป 5.11 (ก)-(ค) แสดง u-matrices ของข้อมูลน้ำมัน โดย แต่ละ map ตัวอย่างที่เก็บได้แต่ละวัน จำนวน 3 วัน และ (ง)-(ช) แสดง response planes สำหรับระยะเวลาของน้ำมัน ที่ได้จากการทำนายของแต่ละ map unit ในรูป (ก)-(ค) ตามลำดับ	40
รูป 5.12 control chart ที่สร้างโดยใช้ (ก) single SOM และ (ข) multiple SOMs	41
รูป 7.1 แบบจำลองการทำงานของเคโมเมตริกซ์	46
รูป 7.2 ตัวอย่างการประยุกต์ใช้เคโมเมตริกซ์ในการแปรผลการทดลอง	46
รูป 7.3 ตัวอย่างข้อต่อที่เกิดจากการวิเคราะห์โดยใช้เพียงตัวแปรเดียว	48
รูป 7.2 UV-Vis สเปกตรัมของสารตัวอย่างจำนวน 39 ตัวอย่าง	50
รูป 7.3 ตัวอย่างการประยุกต์ใช้เคโมเมตริกซ์ในการแปรผลการทดลอง	51
รูป 7.4 การประยุกต์ใช้เคโมเมตริกซ์สำหรับงานวิจัยในแขนงต่างๆ	52
รูป 7.5 เมตริกซ์ที่บรรจุค่าการดูดกลืนแสงของ 39 ตัวอย่าง โดยมีจำนวนความยาวคลื่นที่ถูกใช้ในการตรวจวัดทั้งหมด 24 ความยาวคลื่น	53
รูป 7.6 องค์ประกอบของข้อมูลที่อยู่ในเมตริกซ์	54
รูป 7.7 ข้อมูลแสดงอัตราสารกินน้ำมันของรถยนต์	55
รูป 7.8 ข้อมูลแสดงอัตราสารกินน้ำมันของรถยนต์และแกน PC1	56
รูป 7.11 ข้อมูลแสดงอัตราสารกินน้ำมันของรถยนต์และแกน PC1 และ PC2	57
รูป 7.9 แสดงความสัมพันธ์ระหว่าง score และ loading ในสมการการคูณเมตริกซ์	58
รูป 7.10 Eigen value สำหรับข้อมูลสเปกตรัมจาก 39 ตัวอย่าง	59

สารบัญตาราง

	หน้า
ตาราง 5.1 ผลที่ได้จากการใช้ตัวอย่างที่เป็นเลขคู่เป็นตัวสร้างโมเดล และตัวอย่างที่เป็นเลขคี่เป็นตัวทดสอบ จากตัวอย่างทั้งหมด 59 ตัวอย่าง	32
ตาราง 5.2 ผลที่ได้จากการใช้ตัวอย่างตัวที่ 1, 5, 10, 15, ..., 59 เป็นตัวสร้างโมเดล และตัวอย่างตัวที่เหลือทั้งหมดเป็นตัวทดสอบ จากตัวอย่างทั้งหมด 59 ตัว	32
ตาราง 5.3 ค่า Q^2 และ R^2 ของการใช้ตัวอย่างที่เป็นเลขคู่เป็นตัวสร้างโมเดล และตัวอย่างที่เป็นเลขคี่เป็นตัวทดสอบ จากตัวอย่างทั้งหมด 59 ตัวอย่าง	32
ตาราง 5.4 ค่า Q^2 และ R^2 ของการใช้ตัวอย่างตัวที่ 1, 5, 10, 15, ..., 59 เป็นตัวสร้างโมเดล และตัวอย่างตัวที่เหลือทั้งหมดเป็นตัวทดสอบ จากตัวอย่างทั้งหมด 59 ตัว	34
ตาราง 5.5 ผลที่ได้จากการใช้ตัวอย่างที่เป็นเลขคู่เป็นตัวสร้างโมเดล และตัวอย่างที่เป็นเลขคี่เป็นตัวทดสอบ จากตัวอย่างทั้งหมด 68 ตัวอย่าง	34
ตาราง 5.6 ผลที่ได้จากการใช้ตัวอย่างที่เป็นเลข 1, 5, 10, 15, ..., 68 เป็นตัวสร้างโมเดล และตัวอย่างตัวที่เหลือทั้งหมดเป็นตัวทดสอบ จากตัวอย่างทั้งหมด 68 ตัวอย่าง	35
ตาราง 5.7 ค่า Q^2 และ R^2 ของการใช้ตัวอย่างที่เป็นเลขคู่เป็นตัวสร้างโมเดล และตัวอย่างที่เป็นเลขคี่เป็นตัวทดสอบ จากตัวอย่างทั้งหมด 68 ตัวอย่าง	36
ตาราง 5.8 ค่า Q^2 และ R^2 ของการใช้ตัวอย่างที่เป็นเลข 1, 5, 10, 15, ..., 68 เป็นตัวสร้างโมเดล และตัวอย่างตัวที่เหลือทั้งหมดเป็นตัวทดสอบ จากตัวอย่างทั้งหมด 68 ตัวอย่าง	36

บทคัดย่อ

รหัสโครงการ: TRG5780175
ชื่อโครงการ: การติดตามคุณภาพของน้ำมันประกอบอาหารโดยใช้ระบบ
ปัญญาประดิษฐ์
ชื่อนักวิจัย และสถาบัน
ดร. ศิลา กิตติวิชนะ
ภาควิชาเคมี คณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่
อีเมล: silacmu@gmail.com
ระยะเวลาโครงการ: 2 ปี

บทคัดย่อ:

งานวิจัยนี้เป็นการประยุกต์ใช้เทคนิคเคโมเมทริกซ์โดยใช้เทคนิคปริ้นซิเปิลคอมโพเนนท์อะนาไลซิส (พีซีเอ) และเทคนิคพาร์เชียลลีสท์สแควร์ (พีเอลเอส) ช่วยในการวิเคราะห์ข้อมูลเนียร์อินฟราเรดสเปกตรัม จากน้ำมันสำหรับประกอบอาหารที่ถูกใช้ซ้ำ เพื่อศึกษาการเปลี่ยนแปลงและสร้างโมเดลสำหรับทำนายจำนวนครั้งของน้ำมันพีซีสำหรับประกอบอาหารหลังจากถูกใช้ทอดซ้ำๆ เป็นระยะเวลา นาน โดยตัวอย่างน้ำมันถูกเก็บทันทีหลังจากการทอดมันฝรั่งแต่ละครั้ง ครั้งละประมาณ 5 นาที นาน 2 ชั่วโมง เป็นเวลาติดต่อกัน 3 วัน และวิเคราะห์ด้วยเทคนิคเนียร์อินฟราเรดสเปกโตรสโกปี โดยข้อมูลค่า การดูดกลืนแสงที่บันทึกได้นั้นถูกนำไปวิเคราะห์คุณลักษณะพื้นฐานของข้อมูลด้วยเทคนิคพีซีเอ พบว่า การเตรียมข้อมูลด้วยวิธีลือกสเกลลิ่งสามารถช่วยแสดงความแตกต่างของตัวอย่างน้ำมันที่ผ่านการทอด ในแต่ละครั้งได้อย่างชัดเจน และจากการสร้างโมเดลสำหรับการทำนายจำนวนซ้ำที่น้ำมันถูกใช้ด้วย เทคนิคพีเอลเอส พบว่าการเตรียมข้อมูลด้วยวิธีมินเซนตริงให้ผลการทำนายถูกต้องมากที่สุด โดยมีค่า อาร์เอ็มเอสอี, ค่าอาร์สแควร์ และค่าคิวิสแควร์ เท่ากับ 3.12, 0.9960 และ 0.9733 ตามลำดับ จากนั้น เทคนิคมัลติเปิลเชลฟ์ออกานิชิงค์แมพซึ่งจัดเป็นเทคนิคเคโมเมทริกซ์แบบปัญญาประดิษฐ์ที่สามารถ สร้างโมเดลที่มีลักษณะการทำงานคล้ายแผนที่ โดยแผนที่สำหรับการทำนายที่สร้างขึ้นใหม่ดังกล่าวจะถูก ใช้เพื่อวิเคราะห์ข้อมูลที่ได้จากระบบหลังจากที่มีการเปลี่ยนแปลงเกิดขึ้น ทั้งนี้แผนที่ที่สร้างขึ้นใหม่นี้ จะถูกใช้ร่วมกับแผนที่การทำนายหลักที่ถูกใช้ก่อนหน้านี้ในการทำนายคุณภาพของน้ำมันพีซีสำหรับปรุง อาหารในกระบวนการผลิต โดยโมเดลที่พัฒนาขึ้นมีจุดเด่นคือ สามารถทำงานได้อย่างต่อเนื่องและให้ผล การวิเคราะห์ของคุณภาพน้ำมันทันที ถึงแม้ว่าจะมีการเปลี่ยนแปลงขึ้นในระบบของกระบวนการผลิต

คำหลัก: เคโมเมทริกซ์, ปริ้นซิเปิลคอมโพเนนท์อะนาไลซิส, พาร์เชียลลีสท์สแควร์, เนียร์อินฟราเรดสเปกโตรสโกปี

Abstract

Project Code : TRG5780175
Project Title : Monitoring of cooking oil quality using supervised self organizing maps (SOMs)
Investigator : Dr. Sila Kittiwachana / Department of Chemistry / Faculty of Science / Chiang Mai University
E-mail Address : silacmu@gmail.com
Project Period : 2 years

Abstract:

This research applied chemometric techniques called principle component analysis (PCA) and partial least squares (PLS) for investigating any change in deep-frying cooking oil after repeatedly and long-term used. The vegetable cooking oil used for French-fried potatoes cooking was sampled immediately after each cook, approximately every 5 minutes. The cooking process was continued for 2 hours on consecutively 3 days. The oil samples were then analyzed using near infrared (NIR) spectroscopy. To exploratory analyze the NIR data, PCA was performed and it could be seen that using log scaling could effectively reveal the pattern among the oil samples. PLS, a multivariate calibration model, was established for predicting the period of times the cooking oil samples were used. It was found that the PLS model could provide the best predictive ability where RMSE, R^2 and Q^2 values of the established model were 3.12, 0.9960 and 0.9733, respectively. After that a chemometric model based on the use of multiple self organizing maps (mSOM) was developed for continuous process monitoring. The developed method was aimed to solve the problem when a manufacturing process has to be restarted or there was a disconnection in the process, for example, background changes or drifts of the instrumental detections. Using mSOM, an additional dimension/layer of the supervised SOM model was established for the newly update data from the refreshed process. Therefore, instead of using only the original SOM model to estimate the process status, when the process was restarted, new samples from the process were collected and used for establishing another level of the SOM model. The extension of the modified SOM model was regarded as optional and was used as the same time with the main/original SOM model. The competitive learning within and between maps provided accurate health status of the ongoing process especially the insight change of the process after refreshed.

Key words: Process monitoring, self organizing Maps (SOMs), artificial neural networks (ANNs), deep-frying vegetable oil, near infrared (NIR) spectroscopy

1. บทคัดย่อภาษาไทย และภาษาอังกฤษ

ชื่อโครงการ การติดตามคุณภาพของน้ำมันประกอบอาหารโดยใช้ระบบ
ปัญญาประดิษฐ์

บทคัดย่อ

งานวิจัยนี้เป็นการประยุกต์ใช้เทคนิคเคโมเมทริกซ์โดยใช้เทคนิคปรินซิเปิลคอมโพเนนท์อะนาไลซิส (พีซีเอ) และเทคนิคพาร์เชียลลีสทิสแควร์ (พีแอลเอส) ช่วยในการวิเคราะห์ข้อมูลเนียร์อินฟราเรดสเปกตรัม จากน้ำมันสำหรับประกอบอาหารที่ถูกใช้ซ้ำ เพื่อศึกษาการเปลี่ยนแปลงและสร้างโมเดลสำหรับทำนายจำนวนครั้งของน้ำมันพีซีสำหรับประกอบอาหารหลังจากถูกใช้ทอดซ้ำๆ เป็นระยะเวลา นาน โดยตัวอย่างน้ำมันถูกเก็บทันทีหลังจากการทอดมันฝรั่งแต่ละครั้ง ครั้งละประมาณ 5 นาที นาน 2 ชั่วโมง เป็นเวลาติดต่อกัน 3 วัน และวิเคราะห์ด้วยเทคนิคเนียร์อินฟราเรดสเปกโตรสโกปี โดยข้อมูลค่า การดูดกลืนแสงที่บันทึกได้นั้นถูกนำไปวิเคราะห์คุณลักษณะพื้นฐานของข้อมูลด้วยเทคนิคพีซีเอ พบว่า การเตรียมข้อมูลด้วยวิธีล็อกสเกลลิ่งสามารถช่วยแสดงความแตกต่างของตัวอย่างน้ำมันที่ผ่านการทอด ในแต่ละครั้งได้อย่างชัดเจน และจากการสร้างโมเดลสำหรับการทำนายจำนวนซ้ำที่น้ำมันถูกใช้ด้วย เทคนิคพีแอลเอส พบว่าการเตรียมข้อมูลด้วยวิธีมีนเซนตริงให้ผลการทำนายถูกต้องมากที่สุด โดยมีค่า อาร์เอ็มเอสอี, ค่าอาร์สแควร์ และค่าคิวสแควร์ เท่ากับ 3.12, 0.9960 และ 0.9733 ตามลำดับ จากนั้น เทคนิคมัลติเปิลเชลฟ์ออกานิชิงค์แมฟซึ่งเป็นเทคนิคเคโมเมทริกซ์แบบปัญญาประดิษฐ์ที่สามารถสร้าง โมเดลที่มีลักษณะการทำงานคล้ายแผนที่ โดยแผนที่สำหรับการทำนายที่สร้างขึ้นใหม่ดังกล่าวจะถูกใช้ เพื่อวิเคราะห์ข้อมูลที่ได้จากระบบหลังจากที่มีการเปลี่ยนแปลงเกิดขึ้น ทั้งนี้แผนที่ที่สร้างขึ้นใหม่นี้จะ ถูกใช้ร่วมกับแผนที่การทำนายหลักที่ถูกใช้ก่อนหน้านี้ในการทำนายคุณภาพของน้ำมันพีซีสำหรับปรุง อาหารในกระบวนการผลิต โดยโมเดลที่พัฒนาขึ้นมีจุดเด่นคือ สามารถทำงานได้อย่างต่อเนื่องและให้ผล การวิเคราะห์ของคุณภาพน้ำมันทันที ถึงแม้ว่าจะมีการเปลี่ยนแปลงขึ้นในระบบของกระบวนการผลิต

คำหลัก: เคโมเมทริกซ์, ปรินซิเปิลคอมโพเนนท์อะนาไลซิส, พาร์เชียลลีสทิสแควร์, เนียร์อินฟราเรดสเปกโตรสโกปี

Project Title Monitoring of cooking oil quality using supervised self organizing maps (SOMs)

Abstract

This research applied chemometric techniques called principle component analysis (PCA) and partial least squares (PLS) for investigating any change in deep-frying cooking oil after repeatedly and long-term used. The vegetable cooking oil used for French-fried potatoes cooking was sampled immediately after each cook, approximately every 5 minutes. The cooking process was continued for 2 hours on consecutively 3 days. The oil samples were then analyzed using near infrared (NIR) spectroscopy. To exploratory analyze the NIR data, PCA was performed and it could be seen that using log scaling could effectively reveal the pattern among the oil samples. PLS, a multivariate calibration model, was established for predicting the period of times the cooking oil samples were used. It was found that the PLS model could provide the best predictive ability where RMSE, R^2 and Q^2 values of the established model were 3.12, 0.9960 and 0.9733, respectively. After that a chemometric model based on the use of multiple self organizing maps (mSOM) was developed for continuous process monitoring. The developed method was aimed to solve the problem when a manufacturing process has to be restarted or there was a disconnection in the process, for example, background changes or drifts of the instrumental detections. Using mSOM, an additional dimension/layer of the supervised SOM model was established for the newly update data from the refreshed process. Therefore, instead of using only the original SOM model to estimate the process status, when the process was restarted, new samples from the process were collected and used for establishing another level of the SOM model. The extension of the modified SOM model was regarded as optional and was used as the same time with the main/original SOM model. The competitive learning within and between maps provided accurate health status of the ongoing process especially the insight change of the process after refreshed.

Key words: Process monitoring, self organizing Maps (SOMs), artificial neural networks (ANNs), deep-frying vegetable oil, near infrared (NIR) spectroscopy

2. ที่มาและความสำคัญของปัญหา-ทบทวนวรรณกรรม

น้ำมันสำหรับประกอบอาหารที่เราใช้อยู่ในชีวิตประจำวัน แบ่งออกเป็น 2 ชนิด ได้แก่ น้ำมันจากไขมันสัตว์ เช่น น้ำมันหมู และน้ำมันวัว เป็นต้น และน้ำมันจากพืช เช่น น้ำมันปาล์ม น้ำมันถั่วเหลือง น้ำมันมะพร้าว เป็นต้น นอกจากการอุปโภคและบริโภคน้ำมันในครัวเรือน น้ำมันจัดได้ว่าเป็นองค์ประกอบหลักที่สำคัญในการประกอบอาหารมากมายหลายชนิดในโรงงานอุตสาหกรรมอาหาร เช่น การใช้เพื่อถนอมอาหาร การใช้เป็นส่วนประกอบของอาหาร และการใช้เป็นสื่อกลางสำหรับส่งผ่านความร้อนเพื่อให้อาหารสุก เช่น กระบวนการผลิตมันฝรั่งทอด บะหมี่กึ่งสำเร็จรูป และ ขนมขบเคี้ยว เป็นต้น อย่างไรก็ตาม เมื่อมีการใช้น้ำมันไปนานๆ น้ำมันจะเกิดการเสื่อมสภาพเกิดขึ้น คือ มีลักษณะเหนียวข้น ผิดปกติ มีสีดำ เกิดฟองมาก มีกลิ่นเหม็นไหม้ เกิดควันมากขณะทอด ซึ่งลักษณะเหล่านี้จะเกิดขึ้นเมื่อน้ำมันถูกความร้อนสูง มีความชื้น และทำปฏิกิริยากับออกซิเจน เกิดเป็นสารมีพิษ เช่น สารประกอบที่มีขั้ว (polar compounds) และ สารก่อมะเร็งเกิดขึ้นในน้ำมันทอดซ้ำ ได้แก่ cyclic fatty acids, aldehydic triglycerides, triglyceride hydroperoxides, aldehydes และ polycyclic aromatic hydrocarbon (PAH) เป็นต้น¹ โดยสารประกอบเหล่านี้จัดเป็นสารอันตราย เมื่อเกิดการสะสมในร่างกายปริมาณมาก จะก่อให้เกิดโรคไขมันอุดตันในเส้นเลือด ความดันสูง หัวใจขาดเลือด หัวใจตีบ รวมถึงความเสี่ยงในการเกิดมะเร็ง² ได้มีการทดลองเพื่อหาเปอร์เซ็นต์การดูดซับของน้ำมันที่ถูกใช้ทอดมันฝรั่ง พบว่า มันฝรั่งจะดูดซับน้ำมันโดยเฉลี่ยประมาณ 10% ซึ่งแสดงให้เห็นว่าร่างกายได้รับน้ำมันที่ใช้ประกอบอาหารเป็นจำนวนมากจากการบริโภคอาหารที่ถูกทอด เมื่อพิจารณาถึงปริมาณสารโพลาร์ (total polar material, TPM) พบว่าการใช้น้ำมันทอดซ้ำ 2 - 3 ครั้ง ยังอยู่ในเกณฑ์ค่อนข้างปลอดภัย เนื่องจากปริมาณสารโพลาร์ที่ร่างกายจะได้รับเข้าไปไม่เกินขีดจำกัดที่มีอันตรายต่อสุขภาพ (25 - 27 %) ส่วนการบริโภคในร้านอาหารและอาหารจานด่วน พบว่าค่อนข้างอันตรายต่อสุขภาพ เนื่องจากพบปริมาณสารโพลาร์ มากกว่า 25 % ในตัวอย่างอาหารค่อนข้างมาก จากน้ำมันที่ถูกใช้ซ้ำๆ กันเป็นระยะเวลานาน

การวิเคราะห์คุณภาพของน้ำมันที่ใช้ในการประกอบอาหารที่ผ่านการใช้แล้ว สามารถทำได้หลายวิธี เช่น การวิเคราะห์ปริมาณกรดไขมันอิสระ (free fatty acid, %FFA) การวิเคราะห์ปริมาณสารโพลาร์ (total polar material, TPM), peroxide value (PV) และ TOTOX value³ รวมถึงการตรวจสอบสมบัติทางกายภาพ (physical property) เช่น การดูสี ความชื้น หรือความหนืด เป็นต้น ถึงแม้ว่าการ

¹ Christian Gertz, " Chemical and physical parameters as Quality indicators of used frying fats " Eur. J. Lipid Sci. Technol. 102 (2000) 566-572

² สำนักส่งเสริมและสนับสนุนอาหารปลอดภัย (ออนไลน์). Bureau of Food Safety Extension and Support Ministry of Public Health. แหล่งที่มา http://www.fda.moph.go.th/food_safety/frontend/theme_1/index.php. 25 พฤศจิกายน 2556

³ F. Pedreshi, V.H. Segtnan, S.H. Knutsen, Food Chem. 121 (2010) 616-620.

ประเมินคุณภาพของน้ำมันด้วยการวิเคราะห์ทางเคมีดังกล่าวมีความถูกต้องและแม่นยำค่อนข้างสูงในการประเมินคุณภาพของน้ำมัน แต่ยังมีขั้นตอนในการดำเนินการที่ซับซ้อนและใช้เวลาค่อนข้างนาน ทำให้มีข้อจำกัดในกรณีที่มีจำนวนตัวอย่างที่ต้องการวิเคราะห์เยอะ โดยเฉพาะในโรงงานที่มีกระบวนการผลิตจำนวนมากและความจำเป็นต้องวิเคราะห์ตัวอย่างในลักษณะต่อเนื่อง "online detection" และต้องการผลการวิเคราะห์แบบทันทีหรือ "real time analysis"⁴ ดังนั้นการประยุกต์นำเทคนิคการวิเคราะห์ทางเคมีที่สามารถทำได้ด้วยความรวดเร็ว ไม่ทำลายตัวอย่าง และสามารถติดตั้งเข้าร่วมกับกระบวนการผลิตเพื่อทำการตรวจวัดในลักษณะ online จึงมีความจำเป็น⁵

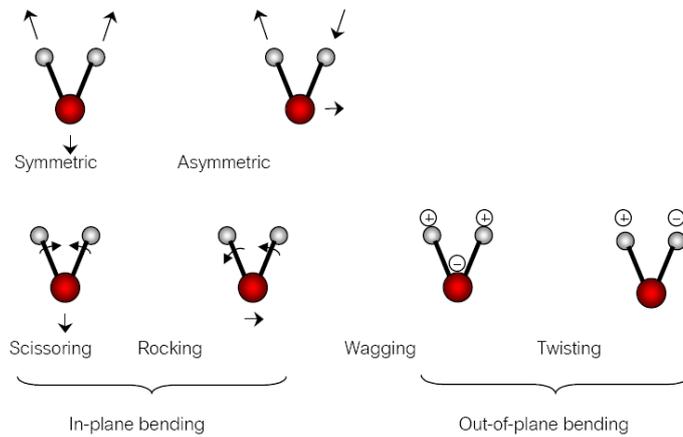
เทคนิคเนียร์อินฟราเรดสเปกโตรสโกปี (near-infrared spectroscopy)⁶ เป็นหนึ่งในกลุ่มเทคนิควิเคราะห์เชิงแสงที่ได้รับความนิยมอย่างมากในโรงงานอุตสาหกรรม เนื่องจากสามารถตรวจวัดสารตัวอย่างได้อย่างสะดวก รวดเร็ว และสามารถติดตั้งรวมเข้าไปในระบบการผลิตสำหรับการตรวจวัดแบบ online detection โดยไม่ทำลายสารตัวอย่าง วิธีการวิเคราะห์ด้วยเทคนิค NIR จะคล้ายกับการวิเคราะห์ด้วยเทคนิค UV-Vis คือ อาศัยหลักการดูดกลืนแสงของสารตัวอย่าง โดยคลื่นแสง NIR มีช่วงความยาวคลื่นตั้งแต่ 800-2500 นาโนเมตร สามารถแบ่งช่วงความยาวคลื่น ออกเป็น 2 ช่วง คือ ช่วงคลื่นสั้นที่มีความยาวคลื่น 800-1100 นาโนเมตร และช่วงคลื่นยาวที่มีความยาวคลื่น 1100-2500 นาโนเมตร พลังงานของคลื่นแสง NIR จะอยู่ในช่วงที่สอดคล้องกับการสั่น (vibration) ของพันธะภายในโมเลกุล หากการสั่นของพันธะใดเกิดที่ความถี่ที่ตรงกับความถี่ของคลื่นแสง NIR ก็จะทำให้เกิดการดูดกลืนขึ้น การสั่นของโมเลกุลในกลุ่มฟังก์ชันต่างๆ สามารถแบ่งออกได้เป็น 2 ลักษณะหลักๆ ลักษณะที่หนึ่ง คือ การยืดหด (stretching) ของพันธะเป็นการสั่นที่ทำให้เกิดการเปลี่ยนแปลงความยาวระหว่างอะตอมที่สร้างพันธะกันซึ่งมีสองรูปแบบย่อยคือ แบบสมมาตร (symmetric) และแบบอสมมาตร (asymmetric) และลักษณะที่สองคือ การงอ (bending) หรือการผิดรูป (deformation) เป็นการสั่นที่ทำให้เกิดการเปลี่ยนแปลงมุมพันธะ ที่มีรูปแบบปลีกย่อยอีกได้แก่ scissoring, rocking, wagging และ twisting ซึ่งทั้งหมดนี้อาจเกิดการเคลื่อนที่อยู่ในระนาบเดียวกัน (in-plane) หรือต่างระนาบกัน (out-of-plane) ก็ได้

ดัง รูป 2.1

⁴ S. Kittiwachana, D.L.S. Ferreira, G.R. Lloyd, L.A. Fido, D.R. Thompson, R.E.A. Escott, R.G. Brereton, Journal of Chemometrics, 24 (2010) 96-110.

⁵ S. Kittiwachana, D.L.S. Ferreira, L.A. Fido, D.R. Thompson, R.E.A. Escott, R.G. Brereton, Analytical Chemistry, 82 (2010) 5972-5982.

⁶ G.R. Lloyd, R.G. Brereton, J.C. Duncan, Analyst, 133 (2008) 1046-1059.

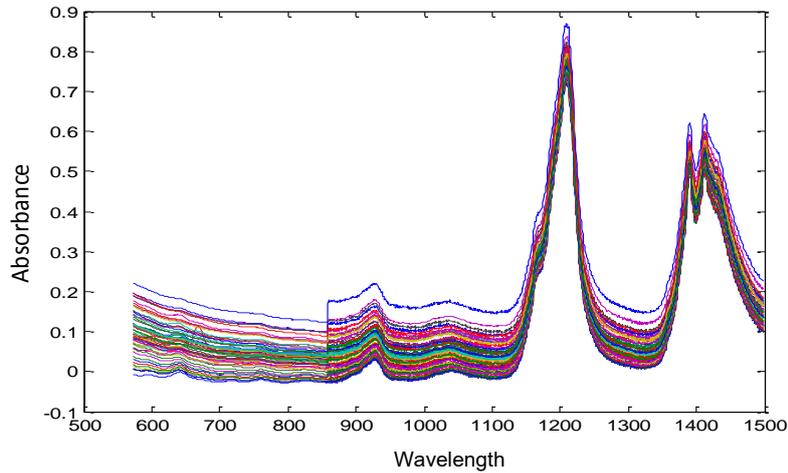


รูป 2.1 การเปลี่ยนแปลงพันธะของโมเลกุลในรูปแบบต่างๆ⁷

การสั่นของพันธะในลักษณะต่างๆ จะมีการดูดกลืนพลังงานที่เป็นเฉพาะของมัน เนื่องจากโมเลกุลหนึ่งๆ มีพันธะได้หลายแบบ ซึ่งแต่ละพันธะก็มีรูปแบบการสั่นได้อีกหลายรูปแบบทำให้โมเลกุลหนึ่งๆ จะแสดงการดูดกลืนแสง NIR ได้หลายช่วงคลื่นพร้อมกัน ลักษณะการดูดกลืนแสงจะเกิดเป็นแถบ (band) หรือ พีค (peak) แสดงถึงพลังงานคลื่นแสง NIR ที่ถูกดูดกลืน ซึ่งสามารถวัดได้ใน 2 รูปแบบ คือ รูปแบบความเข้มแสงที่ส่องผ่าน (transmittance) และรูปแบบความเข้มแสงสะท้อนออกมา (reflectance) แล้วนำค่าความเข้มแสงที่ได้ในแต่ละความยาวคลื่น มาเขียนกราฟโดยให้แกนนอน (แกน x) เป็นค่าความยาวคลื่น แกนตั้ง (แกน y) เป็นค่าการดูดกลืนแสง จะได้กราฟการดูดกลืนแสงของตัวอย่างนั้นๆ กราฟที่ได้เรียกว่า เนียร์อินฟราเรดสเปกตรัม (NIR spectrum) โมเลกุลของสารแต่ละชนิดมีความสามารถในการดูดกลืนแสงที่ความยาวคลื่นต่างกัน เมื่อวัดค่าการดูดกลืนแสงของสารแต่ละชนิดจึงได้สเปกตรัมที่มีลักษณะต่างกัน ตัวอย่างกราฟ NIR แสดงดังรูป 2.2

ทั้งนี้ข้อมูลการดูดกลืนแสงเนียร์อินฟราเรดมักมีความซับซ้อนจากสัญญาณพีคที่อาจเหลื่อมกัน (overlap) เนื่องจากความจำเพาะเจาะจงของการดูดกลืนแสงในช่วงเนียร์อินฟราเรดของสารเคมีที่มีหมู่ฟังก์ชันเป็นองค์ประกอบมักจะต่ำ ดังนั้น การนำเทคนิคการวิเคราะห์ข้อมูลที่มีลักษณะหลายตัวแปร (multivariate analysis) เข้ามาประยุกต์ใช้จึงมีความจำเป็น

⁷ D. Harvey, Modern Analytical Chemistry. The McGraw-Hill Companies, U.S.A., 2000.



รูป 2.2 ตัวอย่าง กราฟ NIR spectra ของตัวอย่างน้ำมัน ช่วงความยาวคลื่น 860-1490 nm

เคโมเมทริกซ์⁸ คือ กระบวนการวิเคราะห์แบบ multivariate analysis ที่เริ่มเป็นที่นิยมในการนำมาใช้ เพื่อวิเคราะห์ข้อมูลมากขึ้นในปัจจุบัน โดยเคโมเมทริกซ์ถือเป็นศาสตร์แขนงหนึ่งของเคมี ที่มีจุดมุ่งหมายเพื่อวิเคราะห์ผลการทดลองปริมาณมากและมีความซับซ้อน เพื่อกลับกรองข้อมูลที่สำคัญออกมาโดยใช้ความรู้พื้นฐานทางสถิติ คณิตศาสตร์ และคอมพิวเตอร์ มาช่วยในการวิเคราะห์ ตัวอย่างเช่น การวิเคราะห์อิทธิพลของข้อมูลหลายตัวแปร และวิธีการตรวจสอบรูปแบบของข้อมูล เป็นต้น นอกจากนี้ยังอาศัยวิทยาการคอมพิวเตอร์ในการพัฒนาโปรแกรม และเคมีวิเคราะห์ในการตรวจวัดผลการทดลอง เช่น ในทางอุตสาหกรรมยา จะประยุกต์ใช้เคโมเมทริกซ์ในการจำแนกกลุ่มยาและติดตามกระบวนการผลิต เป็นต้น นอกจากนี้เคโมเมทริกซ์ยังมีความสัมพันธ์กับวิทยาศาสตร์แขนงอื่นอีกหลายด้าน จนก่อให้เกิดวิทยาศาสตร์แขนงใหม่เช่น เมแทบอลอมิกส์ (metabolomics) โปรตีโอมิกส์ (proteomics) และจีโนมิกส์ (genomics) เป็นต้น

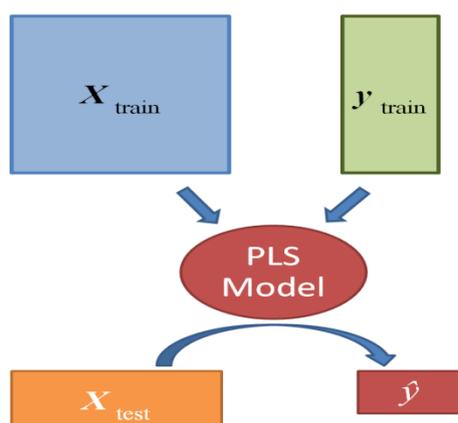
การวิเคราะห์ข้อมูลโดยใช้เคโมเมทริกซ์นั้นมีอยู่หลายเทคนิค การนำแต่ละเทคนิคไปใช้ต้องคำนึงถึงวัตถุประสงค์ด้วยว่าต้องการวิเคราะห์ข้อมูลแบบไหน และลักษณะของข้อมูลเป็นอย่างไร โดยกลุ่มเทคนิคหลักๆ อาจแบ่งได้เป็น 5 กลุ่มเทคนิค ได้แก่ 1. เทคนิคการวิเคราะห์เพื่อสำรวจข้อมูลเบื้องต้น (data exploratory) 2. เทคนิคการแบ่งกลุ่ม (pattern recognition) 3. เทคนิคการติดตามกระบวนการผลิต (process monitoring) 4. เทคนิคการสอบเทียบข้อมูลที่มีหลายตัวแปร (multivariate calibration) และ 5. เทคนิคการออกแบบการทดลองอย่างเป็นระบบ (design of experiments) ซึ่งในการวิจัยครั้งนี้ได้เลือกใช้เทคนิค partial least squares (PLS)⁹ ที่อยู่ในประเภท multivariate calibration

⁸ Brereton, R. G., *Introduction to Multivariate Calibration in Analytical Chemistry*. *Analyst* **2006**, 125 (11), 2125-2154.

⁹ Geladi, P.; Dowalski, B. R., *Partial Least-Squares Regression – a Tutorial*. *Anal. Chim. Acta* **1986**, 185, 1-17.

เพื่อสร้างแบบจำลองสำหรับการทำนายจำนวนครั้งในการใช้ซ้ำของน้ำมันสำหรับประกอบอาหาร รวมถึงการประยุกต์ใช้เทคนิค self organizing maps¹⁰ สำหรับการติดตามคุณภาพของกระบวนการผลิต

เทคนิค PLS จัดเป็น multivariate calibration ที่ศึกษาความสัมพันธ์ระหว่างตัวแปรต้น (x) และตัวแปรตาม (y) ของตัวอย่างที่ใช้สำหรับสร้างโมเดล (training sample) จากนั้นความสัมพันธ์ดังกล่าวสามารถนำมาประยุกต์ใช้ทำนาย การสอบเทียบข้อมูลที่มีหลายตัวแปร (multivariate calibration) โดยเป็นส่วนขยายจาก regression analysis ใช้เพื่อทำนายหาผลลัพธ์ (responses) ของข้อมูลที่มีหลายตัวแปร (multivariate-data) ตัวอย่างเช่น ข้อมูลของตัวแปรต้น ได้แก่ ข้อมูลค่าการดูดกลืนแสงที่ได้จากการวิเคราะห์ด้วยเทคนิค NIR ข้อมูลของอุณหภูมิ ความเข้มข้น หรือค่า pH เป็นต้น โดยตัวแปรตามหรือผลลัพธ์ที่ทำนายได้อาจจะเป็น อายุของการใช้งาน ปริมาณผลผลิตที่ได้ หรือจำนวนครั้งของการใช้งาน เป็นต้น



รูป 2.3 กระบวนการวิเคราะห์ข้อมูลด้วย PLS

อย่างไรก็ตาม การสร้างโมเดลสำหรับการวิเคราะห์หรือทำนายโดยใช้เทคนิคทางเคโมเมทริกซ์ยังคงมีจุดอ่อน กล่าวคือ ประสิทธิภาพและความสามารถของโมเดลที่สร้างนั้นจะขึ้นอยู่กับทางเลือกข้อมูลตัวอย่างหรือต้นแบบ (training sample) ที่นำมาใช้ในการสร้างโมเดล ยกตัวอย่างเช่น โมเดลที่สร้างโดยใช้ข้อมูลจากตัวอย่างน้ำมันยี่ห้อ ก. อาจจะไม่สามารถใช้ในการทำนายคุณภาพของตัวอย่างน้ำมันยี่ห้อ ข. ได้อย่างถูกต้องและแม่นยำ เนื่องจากคุณสมบัติที่สำคัญบางประการที่พบได้ในน้ำมันยี่ห้อ ข. อาจมีความแตกต่างหรือไม่สามารถพบได้ในน้ำมันยี่ห้อ ก. เป็นต้น ดังนั้นโมเดลทางเคโมเมทริกซ์จึงมีความจำเป็นต้องถูกปรับปรุงหรือสร้างใหม่อย่างสม่ำเสมอเพื่อยืนยันความถูกต้องและประสิทธิภาพในการทำนายเมื่อเกิดความเปลี่ยนแปลงขึ้นในกระบวนการผลิต เช่น การปิดหรือหยุดระบบเพื่อเปลี่ยนน้ำมัน

¹⁰ Kittiwachana, S.; Diana L.; S, Ferreira.; Louise,A.; Fido, Duncan R.; Thompson, Richard,E.; A, Escott.; Richard G, Brereton. *Self-Organizing Map Quality Control Index. Anal. chem.* 2010 (82), 5972-5982.

พืชใหม่สำหรับการดูแลรักษากระบวนการผลิตตามปกติ หรือการปิดระบบเพื่อเปลี่ยนแปลงหรือซ่อมแซมอุปกรณ์ในการผลิตในกรณีที่มีความจำเป็น ปัญหาสำคัญที่เกิดขึ้นระหว่างการปรับปรุงหรือสร้างโมเดลใหม่นอกจากค่าใช้จ่ายและเวลา คือการสูญเสียความสามารถในการทำนายในช่วงเวลาดังกล่าว

เทคนิค self organizing maps หรือ SOMs¹¹ จัดเป็นเทคนิคการวิเคราะห์แบบปัญญาประดิษฐ์ (artificial neural networks) กล่าวคือ แบบจำลองที่สร้างขึ้นเกิดจากการเลียนแบบลักษณะโครงสร้างข้อมูลของตัวอย่างที่ถูกกำหนดให้เป็นตัวอย่างต้นแบบ (training samples) การเรียนรู้ที่เกิดขึ้นนั้นมีลักษณะพิเศษคือ การจำลองที่เกิดขึ้นโดยไม่อ้างอิงสมการเชิงคณิตศาสตร์ ส่งผลให้แบบจำลองที่สร้างขึ้นมีความยืดหยุ่น สามารถปรับตัวได้ขึ้นอยู่กับลักษณะของข้อมูลที่ต้องการทำนาย หรืออาจกล่าวได้ว่าเทคนิค SOMs เป็นเทคนิคการทำนายแบบไม่เป็นเส้นตรง (nonlinear method) จุดเด่นของเทคนิค SOMs ความสามารถในการแสดงผลในลักษณะแผนที่ (map) ที่เกิดจากการรวมกันของ map unit ย่อยๆ ซึ่งทำให้ง่ายต่อการแสดงผลของข้อมูลที่อาจมีความซับซ้อน¹² ทั้งนี้ map ที่สร้างขึ้นอาจถูกแบ่งย่อยให้มีขนาดเล็กลงตามลักษณะของกลุ่มข้อมูลที่สร้างขึ้น เกิดเป็น multiple SOMs โดยมีจุดเด่นคือ แบ่งจำลองที่สร้างขึ้นอาจถูกปรับแต่งได้ หากมีการเปลี่ยนแปลงของ เช่น การเพิ่มจำนวนของ training samples ที่อาจจะมีพฤติกรรมที่เปลี่ยนแปลงจาก training samples ที่ถูกใช้ในการสร้างแบบจำลองดั้งเดิม เป็นต้น

จากความเสียงต่างๆข้างต้นนี้ งานวิจัยนี้จึงได้เสนอวิธีที่สามารถตรวจสอบคุณภาพของน้ำมันซึ่งในการศึกษาในครั้งนี้ได้ใช้เทคนิค near infrared (NIR) spectroscopy มาช่วยในการวิเคราะห์คุณภาพของน้ำมันที่ใช้ประกอบอาหารเป็นเวลานาน โดยเทคนิค NIR spectroscopy เป็นเทคนิคที่นิยมใช้กันอย่างแพร่หลาย เนื่องจากเป็นเทคนิคที่ง่าย สะดวกรวดเร็ว และไม่ทำลายสารตัวอย่าง แต่สเปกตรัมที่ได้จะมีความซับซ้อนมาก ดังนั้นจึงมีการนำความรู้ทางเคโมเมทริกซ์ (chemometrics) มาช่วยในการวิเคราะห์ข้อมูลของน้ำมันที่ได้จาก NIR spectroscopy ได้แก่เทคนิค principal component analysis (PCA) และ partial least squares (PLS) รวมถึงการพัฒนาโมเดลทางสถิติด้วยเทคนิคเคโมเมทริกซ์แบบปัญญาประดิษฐ์ที่เรียกว่า self organizing map (SOM) ซึ่งสามารถสร้างโมเดลที่มีลักษณะการทำงานคล้ายแผนที่ โดยแผนที่สำหรับการทำนายที่สร้างขึ้นใหม่ดังกล่าวจะถูกใช้เพื่อวิเคราะห์ข้อมูลที่ได้จากระบบหลังจากที่มีการเปลี่ยนแปลงเกิดขึ้น ทั้งนี้แผนที่ที่สร้างขึ้นมาใหม่นี้จะถูกใช้ร่วมกับแผนที่การทำนายหลักที่ถูกใช้ก่อนหน้านี้ในการทำนายคุณภาพของน้ำมันพืชสำหรับปรุงอาหาร

¹¹ R.G. Brereton, Chemometrics for Pattern Recognition. Wiley: Chichester, U.K., 2009.

¹² G.R. Lloyd, S. Ahmad, M. Wasim, R.G. Brereton, Analytica Chimica Acta 649(2009) 33-42.

ในกระบวนการผลิต โดยโมเดลที่พัฒนาขึ้นมีจุดเด่นคือ สามารถทำงานได้อย่างต่อเนื่องและให้ผลการวิเคราะห์ของคุณภาพน้ำมันทันที ถึงแม้ว่าจะมีการเปลี่ยนแปลงขึ้นในระบบของกระบวนการผลิต

3. วัตถุประสงค์

- 3.1 เพื่อเพิ่มประสิทธิภาพการใช้เทคนิคเนียร์อินฟราเรดสเปกโตรสโคปีสำหรับการติดตามคุณภาพของน้ำมันพืชสำหรับปรุงอาหารโดยการใช้เทคนิคเคโมเมทริกซ์
- 3.2 เพื่อสร้างโมเดลทางสถิติด้วยเทคนิคเคโมเมทริกซ์แบบปัญญาประดิษฐ์ที่สามารถติดตามคุณภาพของน้ำมันพืชสำหรับปรุงอาหารในกระบวนการผลิตที่ดำเนินการอย่างต่อเนื่อง โดยสามารถให้ผลการวิเคราะห์อย่างทันทีทันใด (on-line monitoring of continuous process)
- 3.3 เพื่อนำแบบจำลองปัญญาประดิษฐ์ที่พัฒนาขึ้นไปใช้สำหรับการแบ่งกลุ่มของข้อมูลเคมีที่มีความซับซ้อน เช่น ข้อมูลองค์ประกอบของดินที่ได้จากแปลงนาในสำหรับปลูกข้าวหอมมะลิในภาคเหนือและภาคตะวันออกเฉียงเหนือของประเทศไทย

4. วิธีทดลอง

4.1 ขั้นตอนและวิธีการทดลอง

4.1.1 การเตรียมอุปกรณ์ และการเก็บตัวอย่างน้ำมัน

ในขั้นตอนการเตรียมตัวอย่างน้ำมันที่จะใช้สำหรับการวิเคราะห์ในงานวิจัยนี้ เริ่มต้นโดย ตั้งกระทะไฟฟ้าที่จะใช้ในการทอด และตั้งเทอร์โมมิเตอร์เพื่อใช้วัดอุณหภูมิในขณะการทอดโดยอุณหภูมิระหว่างการทอดอยู่ที่ประมาณ 178 °C จากนั้นนำมันฝรั่งที่เตรียมไว้ลงทอดในปริมาณที่พอเหมาะจนกระทั่งมันฝรั่งสุกซึ่งใช้เวลาประมาณ 5 นาที จึงตักมันฝรั่งขึ้นจากกระทะไฟฟ้า เก็บตัวอย่างน้ำมันใส่ในขวดเก็บตัวอย่างขนาด 10 ml และสังเกตดูอุณหภูมิจากเทอร์โมมิเตอร์ โดยจะพยายามควบคุมให้อุณหภูมิใกล้เคียงกันตลอดการดำเนินการอยู่ที่ประมาณ 178 °C ทำการทดลองอย่างต่อเนื่องเป็นเวลา 3 ชั่วโมง โดยน้ำมันที่ใช้แล้วจะถูกนำมาใช้ซ้ำเป็นเวลาติดต่อกัน 3 วัน ขั้นตอนการเตรียมน้ำมันตัวอย่างแสดงดังรูป 4.1



(ก)



(ข)



(ค)



(ง)

รูป 4.1 ขั้นตอนการเตรียมน้ำมันตัวอย่าง (ก) ตั้งอุปกรณ์ (ข) นำมันฝรั่งลงทอด (ค) เก็บน้ำมันตัวอย่าง (ง) ตัวอย่างน้ำมันในขวดเก็บตัวอย่าง

4.1.2 วิธีการวิเคราะห์ผลด้วย NIR

น้ำมันตัวอย่างจากขั้นตอนที่ 4.2.1 จะถูกนำไปวิเคราะห์ด้วยเทคนิค near infrared (NIR) spectroscopy ด้วยเครื่อง UV/VIS/NIR, Spectrometer Lambda 19, บริษัท PERKIN ELMER ซึ่งแสดงดังรูป 4.2



รูป 4.2 เครื่อง UV-Vis spectroscopy ที่สามารถวัดในช่วง Near infrared (NIR) ได้

4.1.3 การวิเคราะห์ปริมาณกรดไขมันอิสระ (free fatty acid, %FFA)¹³

การวิเคราะห์ปริมาณกรดไขมันอิสระ (free fatty acid, %FFA) มีขั้นตอน ดังนี้

1. ชั่งตัวอย่างน้ำมันให้ได้น้ำหนักแน่นอน 1-10 กรัม ในขวดรูปชมพู่ขนาด 250 มิลลิลิตร
2. เตรียมสารละลายแอลกอฮอล์ (ปรับสภาพให้เป็นกลาง) โดยเติมฟีนอล์ฟทาลีน 5 หยด จากนั้นหยดโซเดียมไฮดรอกไซด์ 0.1 นอร์มอล ที่ละหยดพร้อมทั้งเขย่าจนได้สีชมพูถาวร
3. เติมแอลกอฮอล์ที่มีสภาพเป็นกลาง 50.00 มิลลิลิตร ลงในตัวอย่าง เขย่าอย่างแรงให้ตัวอย่างละลายในแอลกอฮอล์
4. ไตเตรทสารละลายตัวอย่างด้วยโซเดียมไฮดรอกไซด์ 0.1 นอร์มอล จนกระทั่งได้สีชมพูถาวร
5. คำนวณปริมาณกรดไขมันอิสระจากสูตร

$$\% \text{ FFA} = (\text{ปริมาตรต่างที่ใช้ (มิลลิลิตร)} \times \text{ความเข้มข้นต่าง (นอร์มอล)} \times 25.6) / \text{น้ำหนักตัวอย่าง (g)}$$

หมายเหตุ สูตรคำนวณนี้ใช้สำหรับคำนวณปริมาณกรดไขมันอิสระในน้ำมันที่มีกรดปาล์มมิติกเป็นองค์ประกอบหลัก เช่น น้ำมันปาล์ม

4.1.4 การวิเคราะห์ปริมาณสารโพลาร์ (total polar material, TPM)¹³

การวิเคราะห์ปริมาณสารโพลาร์ (total polar material, TPM) มีขั้นตอน ดังนี้

1. เตรียมคอลัมน์แก้ว
2. ชั่งซิลิกาเจล 60 ปริมาณ 25 กรัมใส่ปิกเกอร์ จากนั้นใส่สารละลายผสมระหว่างปิโตรเลียมอีเทอร์และไดเอทิลอีเทอร์ (87:13 v/v) 80 มิลลิลิตร คนด้วยแท่งแก้วจนหมดฟองอากาศ
3. นำสำลีปริมาณเล็กน้อยอุดในคอลัมน์แก้ว ป้องกันการไหลออกของซิลิกาเจล 60 จากนั้นเปิดวาล์วคอลัมน์แล้วจึงคนซิลิกาเจล 60 ที่เตรียมไว้ ค่อย ๆ รินลงในคอลัมน์จนหมด ปล่อยให้สารละลายในคอลัมน์ ไหลออกจนเหลือเพียงเล็กน้อยบนผิวของซิลิกาเจล 60 แล้วจึงปิดวาล์ว
4. เตรียมตัวอย่างโดยการชั่งน้ำมัน 2.5 ± 1 กรัม ลงในขวดวัดปริมาตรขนาด 50.00 มิลลิลิตร ปรับปริมาตรด้วยสารละลายผสมระหว่างปิโตรเลียมอีเทอร์ และไดเอทิลอีเทอร์ (87:13 v/v)
5. บีบอัดสารละลายตัวอย่างน้ำมัน 20 มิลลิลิตร ลงในคอลัมน์ นำขวดก้นกลมมารองรับสารละลายจากคอลัมน์ ส่วนนี้จะเป็นส่วนของ nonpolar fraction เตรียมสารละลายผสมระหว่างปิโตรเลียมอีเทอร์ และไดเอทิลอีเทอร์ (87:13 v/v) 150 มิลลิลิตรไว้ จากนั้นเปิดก๊อก ปล่อยให้สารละลายตัวอย่างน้ำมันจากคอลัมน์ไหลช้า ๆ จนเหลือบนผิวซิลิกาเจล 60 ประมาณ 1 เซนติเมตรค่อย ๆ รินสารละลายผสมที่เตรียมไว้ลงในคอลัมน์ (ระวังผิวหน้าของซิลิกาเจลในคอลัมน์) จนหมดและเหลือสารละลายผสมบนผิวหน้าซิลิกาเจลไว้พอประมาณ ปรับการไหลของคอลัมน์ให้สารละลายไหลหมดภายใน 60-70 นาที
6. เปลี่ยนขวดก้นกลมใหม่มารองรับสารละลายจากคอลัมน์ ซึ่งส่วนนี้จะเป็น polar fraction เตรียมสารละลายไดเอทิลอีเทอร์ไว้ 150 มิลลิลิตร จากนั้นจึงค่อย ๆ รินลงไปคอลัมน์จนหมด ปรับการไหลของคอลัมน์ให้สารละลายไหลหมดภายใน 60-70 นาที
7. นำขวดก้นกลมทั้งสองระเหยสารละลายออก อบและชั่งน้ำหนักสารที่ได้ออก และนำไปคำนวณ ดังนี้

$$TPM = \frac{(E - A)}{E} \times 100$$

โดย A = น้ำหนักของส่วน nonpolar fraction (กรัม)

E = น้ำหนักของสารละลายน้ำมันใน 10 มิลลิลิตร (กรัม)

8. ตรวจสอบประสิทธิภาพของการแยกด้วยคอลัมน์โดยใช้ TLC ทำโดยละลายส่วนที่เป็น polar และ nonpolar fraction ด้วยคลอโรฟอร์มปริมาณเล็กน้อย เตรียม spot สารละลายโดยใช้แท่งแก้วคาปิลารี ลงบนแผ่น TLC ที่ตัดมา ใช้สารละลายผสมปิโตรเลียมอีเทอร์และไดเอทิลอีเทอร์ และกรดแอซิติค (70:30:2 v/v) เป็นตัวทำละลายเคลื่อนที่ หลังจากตัวทำละลายเคลื่อนที่ถึงขอบแผ่น TLC แล้ว นำออกจากตัวทำละลายเคลื่อนที่ ให้สารละลายระเหยหมด จากนั้นนำไปสเปรย์ด้วยกรดโมลิบโดฟอสฟอริก 10 % ที่ละลายในแอลกอฮอล์ แล้วนำไป

ระเหยด้วยการให้ความร้อนที่อุณหภูมิประมาณ 120-130 °C จนปรากฏจุดบนแผ่น TLC ชัดเจน โดย spot ที่เป็นส่วน nonpolar fraction จะเป็นจุดกลมไม่มีหาง ในขณะที่ส่วนที่เป็น polar fraction มีลักษณะเป็นหางยาว และอยู่ต่ำกว่าตำแหน่งของส่วน nonpolar fraction บนแผ่น TLC

4.1.5 การวิเคราะห์ปริมาณเปอร์ออกไซด์ (peroxide value, PV)¹³

การวิเคราะห์ปริมาณเปอร์ออกไซด์ (peroxide value, PV) มีขั้นตอน ดังนี้

1. นำขวดรูปชมพู่ ขนาด 250 มิลลิลิตร ชั่งบนเครื่องชั่งหยาบ เติมน้ำมันตัวอย่างลงไปให้ได้ 5.000 กรัม
2. เติมกรดแอสติติกเข้มข้น ต่อ คลอโรฟอร์ม (อัตราส่วน 3 ต่อ 2) ปริมาตร 0.5 มิลลิลิตร เขย่าเบา ๆ ให้เข้ากัน
3. เติมสารละลายโพแทสเซียมไอโอไดด์ ความเข้มข้น 0.05 นอร์มอล ปริมาตร 0.5 มิลลิลิตร เขย่าเบา ๆ ให้เข้ากัน
4. เติมน้ำกลั่นปริมาตร 30.00 มิลลิลิตร และเติมอินดิเคอร์น้ำแป้งสุก ปริมาตร 2 มิลลิลิตร เขย่าเบา ๆ ให้เข้ากัน นำไปเก็บไว้ในที่มืดนาน 5-10 นาที จากนั้นนำไปไตเตรตด้วยสารละลายโซเดียมไธโอซัลเฟตความเข้มข้น 0.05 นอร์มอล จนกระทั่งสีน้ำเงินจางหายไป จดปริมาตรของสารละลายโซเดียมไธโอซัลเฟตที่ใช้ไป เพื่อนำไปใช้คำนวณต่อไป
5. ทำการทดลองของสารละลายแบลงค์ โดยเติมกรดแอสติติกเข้มข้น ต่อ คลอโรฟอร์ม (อัตราส่วน 3 ต่อ 2) ปริมาตร 30.00 มิลลิลิตร เติมน้ำกลั่นปริมาตร 0.5 มิลลิลิตร เขย่าเบา ๆ ให้เข้ากัน เติมน้ำกลั่น ปริมาตร 30.00 มิลลิลิตร เขย่าเบา ๆ ให้เข้ากัน นำไปไตเตรตด้วยสารละลายโซเดียมไธโอซัลเฟตความเข้มข้น 0.05 นอร์มอล จนกระทั่งสีน้ำเงินจางหายไป จดปริมาตรของสารละลายโซเดียมไธโอซัลเฟตที่ใช้ เพื่อนำไปใช้คำนวณจากสมการ

$$PV = \frac{(a - b) \times N \times 1000}{W}$$

โดย a = ปริมาตร (มิลลิลิตร) ของสารละลายโซเดียมไธโอซัลเฟตที่ใช้ไตเตรตกับตัวอย่าง

b = ปริมาตร (มิลลิลิตร) ของสารละลายโซเดียมไธโอซัลเฟตที่ใช้ไตเตรตกับแบลงค์

N = ความเข้มข้นของโซเดียมไธโอซัลเฟต (นอร์มอล)

W = น้ำหนักตัวอย่าง (กรัม)

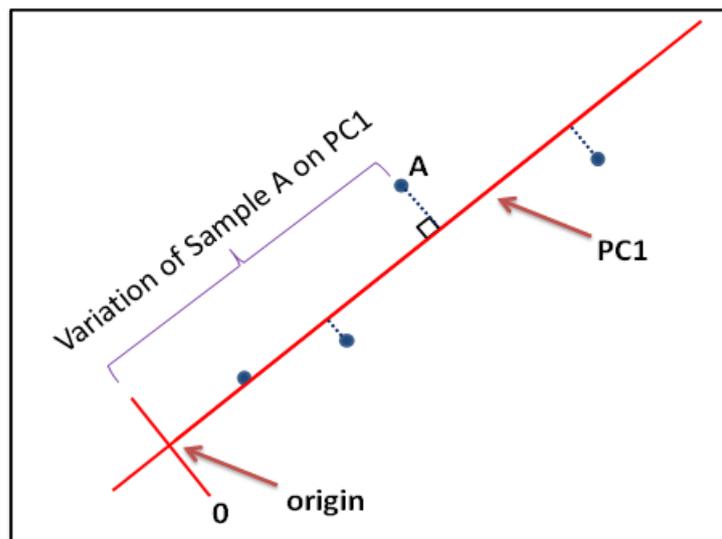
¹³

สุนิสา วิชชุเขต, ปัจจัยบางประการที่มีผลต่อคุณภาพน้ำมันทอดไก่, สาขาวิชาวิทยาศาสตร์เทคโนโลยีอาหาร มหาวิทยาลัยสงขลานครินทร์, 2553

4.2 การวิเคราะห์ด้วยเทคนิคเคโมเมตริกซ์

4.2.1 การวิเคราะห์องค์ประกอบหลัก (Principal component analysis, PCA)

เทคนิค PCA เป็นเทคนิคหนึ่งในเคโมเมตริกซ์ที่ใช้วิเคราะห์คุณลักษณะพื้นฐานของข้อมูลที่ได้จากการบันทึกผลการทดลองซึ่งเป็นข้อมูลที่มีหลายตัวแปร (multivariate data analysis) เทคนิคนี้ใช้ลดจำนวนข้อมูลที่มีจำนวนมาก และมีความซับซ้อนให้อยู่ในรูปที่สะดวกต่อการวิเคราะห์ หรืออาจกล่าวได้ว่าเป็นวิธีหนึ่งที่สามารถช่วยจัดกลุ่มของข้อมูลที่มีความสัมพันธ์กันให้อยู่ใกล้เคียงกัน โดยในการนำข้อมูลมาวิเคราะห์เพื่อหาว่าข้อมูลนั้น ๆ มีความสัมพันธ์กันอย่างไร สามารถแสดงผลของข้อมูลได้ในรูปของกราฟความสัมพันธ์ระหว่างแกนเส้นตรงที่กำหนดขึ้นมาใหม่ (สร้างตัวแปรใหม่จากการรวมตัวแปรที่มีอยู่เดิมเข้าด้วยกัน) เพื่อรวมข้อมูลที่มีจำนวนมากให้น้อยลงโดยแกนที่สร้างขึ้นมาใหม่ยังคงสามารถเก็บข้อมูล (variation) ที่มีความสำคัญไว้ได้เกือบทั้งหมด และสามารถใช้ตัวแปรที่สร้างขึ้นมานี้อธิบายข้อมูลชุดเดียวกันได้ เรียกเป็น principal component หรือ แกน PC เนื่องจากแกน PC ที่กำหนดขึ้นมาเป็นแกนเส้นตรง จึงสามารถทำให้ข้อมูลทั้งหมดมารวมกันอยู่ในแกนเส้นตรงเดียวกัน โดยการใช้การกระจัดของเวกเตอร์ที่ได้จากการ project จุดของตัวอย่างแต่ละตัวลงบนแกนเส้นตรง ดังรูป 4.3 โดยระยะกระจัดบนแกนที่สร้างขึ้นมาใหม่ จากจุด 0 ถึงตำแหน่งของข้อมูลแต่ละตัว จะได้ค่าที่เรียกว่า “variation” หรือ “score”



รูป 4.3 ตัวอย่างการ project ข้อมูลลงบนแกน PC1 และค่า variation

การใช้เทคนิคนี้จะทำให้สามารถสังเกตพฤติกรรมของข้อมูลได้สะดวกและรวดเร็ว ทำให้เห็นแนวโน้มของการจัดกลุ่ม สามารถใช้ได้กับข้อมูลที่ไม่ซับซ้อนมากไปจนถึงข้อมูลที่มีความซับซ้อนในระดับหนึ่ง

ความหมายของ scores และ loadings

- Scores

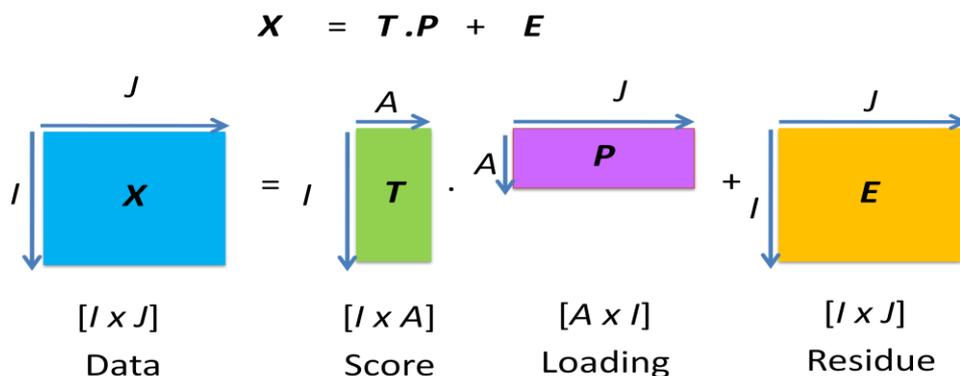
score คือ ค่า variation ของแต่ละตัวอย่าง โดยจำนวน score ของแต่ละ PC จะเท่ากับจำนวนตัวอย่าง ค่า score ที่คำนวณได้จะเป็นค่าที่แสดงถึงความสัมพันธ์ (relationship) ของตัวอย่างแต่ละตัวบนแกน PC นั้นๆ โดยตัวอย่างที่มีค่า score ใกล้เคียงกันจะมีลักษณะบางประการที่คล้ายคลึงกัน และตัวอย่างที่มีค่า score ต่างกันก็จะมีลักษณะบางประการที่แตกต่างกัน ความสัมพันธ์ดังกล่าวนี้จะสังเกตได้โดยการพลอตค่า score ของแต่ละ PC ลงบนกราฟ และกราฟที่ถูกพลอตขึ้นนี้จะเรียกว่า “score plot” ซึ่งหากตัวอย่างมีลักษณะใกล้เคียงกัน ตำแหน่งของตัวอย่างดังกล่าวเหล่านี้จะอยู่บริเวณใกล้เคียงกัน และในทางกลับกัน หากตัวอย่างมีลักษณะที่แตกต่างกัน ก็จะมีตำแหน่งบนกราฟ score plot ที่แตกต่างกัน

- Loadings

loading เป็นค่าที่แสดงความสัมพันธ์ของตัวแปรแต่ละตัวแปร หาก loading ของตัวแปรใดๆ คล้ายกัน ตัวแปรนั้นก็จะมีลักษณะการเปลี่ยนแปลงที่คล้ายคลึงกัน นั่นคือถ้าตัวหนึ่งเพิ่มตัวแปรอีกตัวที่มี loading ที่ใกล้เคียงกันก็ควรจะมีแนวโน้มเพิ่มขึ้น หรือถ้า loading ต่างกันก็จะมีลักษณะการเปลี่ยนแปลงที่ต่างกัน คือหากตัวแปรตัวหนึ่งมีค่าเพิ่ม ตัวแปรอีกตัวหนึ่งที่มี loading ต่างกันอาจจะลดลง ในการวิเคราะห์ค่า loading จะถูกคำนวณออกมาพร้อมกับค่า score เสมอ

- ความสัมพันธ์ของ score และ loading

เมื่อเข้าสู่กระบวนการทำ PCA ค่า score และ loading จะถูกคำนวณออกมาโดยโปรแกรมทางคอมพิวเตอร์ในรูปแบบของเมทริกซ์ จากรูป 4.4 หากมีเรามีข้อมูลดิบ (raw data) ในรูปเมทริกซ์ X ที่มีขนาดเท่ากับ $[I \times J]$ หมายถึง ข้อมูลทั้งหมดมีจำนวนตัวอย่างเท่ากับ I และมีจำนวนตัวแปรเท่ากับ J เมื่อเข้าสู่กระบวนการ PCA จะได้เป็นเมทริกซ์ T และ P ออกมา โดยเมทริกซ์ T คือ score หรือความสัมพันธ์ของข้อมูลแต่ละตัว มีขนาดเท่ากับ $[I \times A]$ และ P คือ loading หรือ ความสัมพันธ์ของตัวแปรแต่ละตัว มีขนาดเท่ากับ $[A \times J]$ โดยที่ค่า A คือจำนวน PCs ที่ใช้ในการวิเคราะห์ข้อมูล นอกจากนี้ยังได้เมทริกซ์ E หรือค่า residue ออกมาด้วย โดยจะมีขนาดเท่ากับเมทริกซ์ data



รูป 4.4 ผลที่ได้จากการวิเคราะห์ข้อมูล (data) ด้วย PCA ประกอบด้วย score, loading และ residue

4.2.2 เทคนิคพาร์เชียลลีสแควร์ (partial least squares, PLS)

เทคนิค PLS จัดเป็น multivariate calibration ที่ศึกษาความสัมพันธ์ระหว่างตัวแปรต้น (x) และตัวแปรตาม (y) ของตัวอย่างที่ใช้สำหรับสร้างโมเดล (training sample) จากนั้นความสัมพันธ์ดังกล่าวสามารถนำมาประยุกต์ใช้ทำนาย การสอบเทียบข้อมูลที่มีหลายตัวแปร (multivariate calibration) โดยเป็นส่วนขยายจาก regression analysis ใช้เพื่อทำนายหาผลลัพธ์ (responses) ของข้อมูลที่มีหลายตัวแปร (multivariate-data) ตัวอย่างเช่น ข้อมูลของตัวแปรต้น ได้แก่ ข้อมูลค่าการดูดกลืนแสงที่ได้จากการวิเคราะห์ด้วยเทคนิค NIR ข้อมูลของอุณหภูมิ ความเข้มข้น หรือค่า pH เป็นต้น โดยตัวแปรตามหรือผลลัพธ์ที่ทำนายได้อาจจะเป็น อายุของการใช้งาน ปริมาณผลผลิตที่ได้ หรือจำนวนครั้งของการใช้งาน เป็นต้น

การวิเคราะห์ข้อมูลด้วย PLS จะมีลักษณะคล้ายกับการวิเคราะห์ด้วย Multiple Linear Regression (MLR) และ Principal Component Regression (PCR) แต่การทำนายด้วยเทคนิค PLS มักจะได้ผลการคำนวณที่มีความถูกต้องมากกว่า เนื่องจากข้อมูลที่ต้องการทำการวิเคราะห์จะผ่านกระบวนการลดจำนวนตัวแปรลงด้วยเทคนิคที่เรียกว่า Non-linear Iterative Partial Least Square (NIPALS) ซึ่งจะช่วยลดปัญหาความคลาดเคลื่อนจากการวิเคราะห์ที่มีหลายตัวแปร ทำให้ได้ผลการวิเคราะห์ที่ดีกว่า โดยจะช่วยลดจำนวนตัวแปรลงให้เป็นจำนวน PC ที่ใช้ ซึ่งจะมีค่าน้อยกว่าหรือเท่ากับจำนวนของตัวอย่าง เนื่องจากการที่มีจำนวนตัวแปรมากกว่าจำนวนตัวอย่าง ทำให้เกิดข้อจำกัดที่ทำให้ไม่สามารถคำนวณสมการเพื่อสร้างโมเดล MLR ได้ อีกทั้งเทคนิค NIPALS ที่ใช้จะไปปรับให้ค่าความสัมพันธ์ (correlation) ของค่า variation ระหว่าง X และ y มีค่าสูงที่สุดเท่าที่จะเป็นไปได้ ดังนั้น PLS จึงเป็นเทคนิคที่ให้ค่าการทำนายที่ถูกต้อง และแม่นยำมากกว่า MLR และ PCR

การนำข้อมูลที่เป็น training set ได้แก่ ข้อมูลที่มีหลายตัวแปร ($\mathbf{X}_{\text{train}}$) และข้อมูลของผลลัพธ์ ($\mathbf{y}_{\text{train}}$) มาสร้างเป็น PLS Model ด้วยเทคนิคที่เรียกว่า NIPALS เราจะได้โมเดลมาหนึ่งโมเดล จากข้อมูลที่เราเลือกเป็น training set ขั้นตอนต่อไปคือขั้นตอนการทำนายผล โดยนำข้อมูลหลายตัวแปรที่เราต้องการทราบผลลัพธ์ (\mathbf{X}_{test}) ผ่านเข้าไปใน PLS Model ก็จะได้ค่าผลลัพธ์ที่ทำนายได้ออกมา เรียกว่า $\mathbf{y}_{\text{prediction}}$ (\hat{y}) ดังรูป 2.3

กระบวนการที่กล่าวมาข้างต้นอาจดูเหมือนไม่ซับซ้อน แต่ในความเป็นจริง ค่าที่เราทำนายได้มักมีความคลาดเคลื่อนด้วยเสมอ ความคลาดเคลื่อนนี้จะแสดงในค่า รุทมีนสแควร์เออเรอร์ (root mean square error, RMSE) ค่า RMSE เป็นรากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย ซึ่งก็คือค่าความคลาดเคลื่อนของการทำนายผลจาก PLS Model เป็นวิธีที่ใช้ประเมินความถูกต้องของ PLS Model ที่เราสร้างขึ้น โดยคำนวณดังสมการ

$$\text{RMSE} = \sqrt{\frac{\sum (y_{\text{test}} - \hat{y})^2}{N}}$$

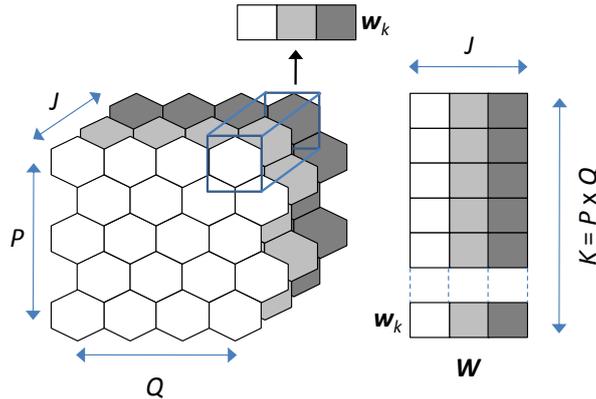
เมื่อ y_{test} = ผลลัพธ์ที่ได้จากการทดลอง
 \hat{y} = ผลลัพธ์ที่ได้จากการทำนาย
 N = จำนวนของผลลัพธ์ทั้งหมด

โดยหากค่า RMSE ที่ได้มีค่าน้อย แสดงว่าโมเดลที่เราสร้างขึ้นนั้น มีความถูกต้องในการทำนายมาก ในทางกลับกันหากค่า RMSE ที่ได้มีค่ามาก ก็หมายถึงโมเดลที่สร้างขึ้นมีความถูกต้องน้อย ในการสร้าง PLS โมเดล ค่า RMSE ที่ได้จะมากหรือน้อยต้องคำนึงถึงการใช้ data preprocessing ในการเตรียมข้อมูลที่เหมาะสม และการเลือกใช้จำนวน PC ในการสร้างด้วย ซึ่งเทคนิคที่ใช้เลือกจำนวน PC ที่เหมาะสมคือ เทคนิค cross validation¹⁰

4.2.3 การวิเคราะห์ข้อมูลด้วยเทคนิค self organizing maps (SOM)

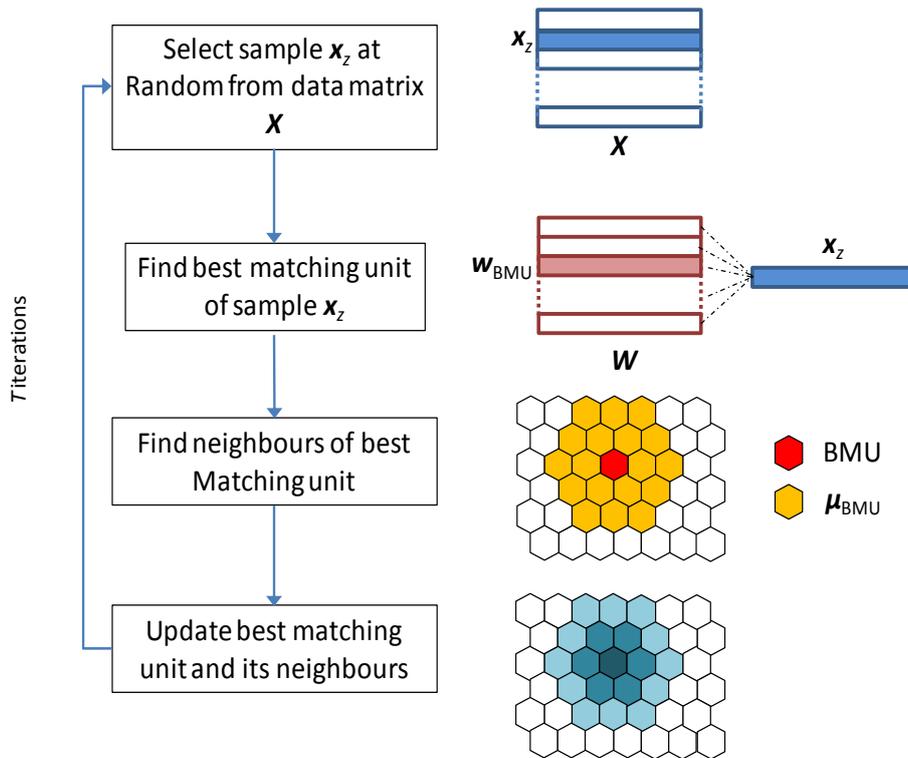
เทคนิค SOM จัดเป็นแบบจำลองแบบปัญญาประดิษฐ์ (artificial neural network) กล่าวคือแบบจำลองที่สร้างขึ้นจะพยายามเลียนแบบโครงสร้างของข้อมูลต้นแบบที่นำมาใช้ในการสร้างแบบจำลองนั้น โดยการเรียนรู้จะมีลักษณะแบบซ้ำๆ (iterative learning) และการปรับตัวของแบบจำลองขึ้นอยู่กับค่าการสุ่มเรียกของข้อมูลต้นแบบ ในที่สุดแบบจำลอง SOM ที่สร้างขึ้นจะสามารถนำไปใช้ในการแสดงลักษณะของข้อมูลโดยการจัดวางตัวอย่างลงบนตำแหน่งที่เหมาะสมบนโครงสร้างของแบบจำลอง SOM ที่สร้างขึ้นนั่นเอง ซึ่งโดยปกติแล้ว โครงสร้างของ SOM มักถูกแสดงในรูปแบบของแผนที่ (map)

ที่มีลักษณะเป็น 2 มิติ ที่ประกอบด้วยส่วนประกอบเล็ก (map unit) ที่มีลักษณะรูปทรงสี่เหลี่ยมหรือหกเหลี่ยมเรียงต่อกันอย่างเป็นระเบียบ ดังรูป 4.5



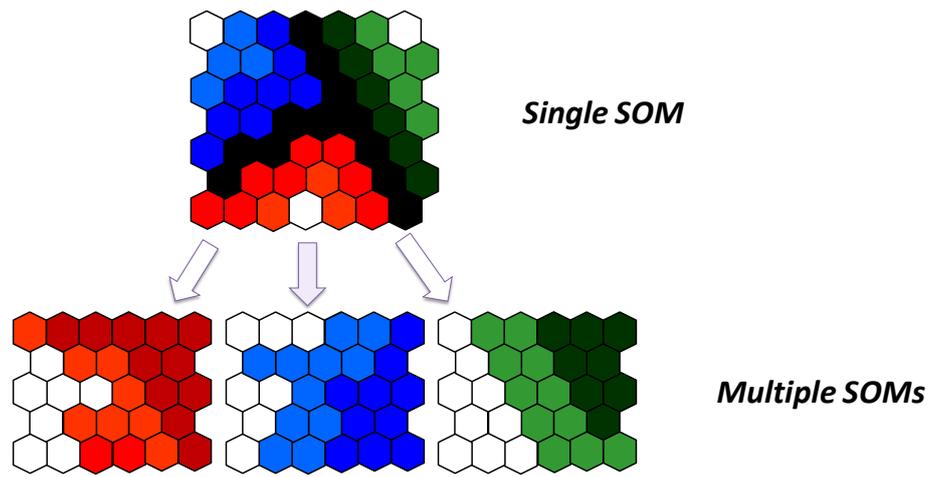
รูป 4.5 องค์ประกอบของ SOM ที่มีขนาด $P \times Q$ และใช้จำนวนพารามิเตอร์เท่ากับ J

รูปที่ 4.5 แสดงแผนที่ SOM ที่ใช้ map unit ที่มีลักษณะรูปทรงสี่เหลี่ยมจัตุรัส โดยแผนที่ที่มีขนาด $[P \times Q]$ หน่วย ส่วนจำนวนชั้นหรือค่า J จะเท่ากับจำนวนของพารามิเตอร์ที่ใช้ในการสร้างแบบจำลอง และเมื่อเรานำ map unit แต่ละหน่วยมาต่อกันเป็นแถวยาว จะทำให้เกิด matrix ซึ่งเรียกว่า weight matrix โดยมีจำนวนสมาชิกเท่ากับจำนวนของ map unit ที่ใช้ ($K = P \times Q$) และจำนวนคอลัมน์ของ weight matrix เท่ากับ J พารามิเตอร์หรือตัวแปร หลังจาก that weight matrix ได้ถูกกำหนดขึ้นในขั้นตอนแรก ขั้นตอนถัดมาคือการเรียนรู้ของแบบจำลอง โดยเริ่มจากการเลือกตัวอย่างจาก training samples มา 1 ตัวอย่างโดยใช้วิธีเลือกแบบ random กล่าวคือ ทุกๆตัวอย่างที่ถูกใช้เป็น training samples มีความน่าจะเป็นเท่าเทียมกันที่จะถูกเลือก จากนั้นตัวอย่างที่ถูกเลือกนี้ (\mathbf{x}_2) จะถูกนำไปเปรียบเทียบกับ weight matrix เพื่อพิจารณาหา map unit ที่มีความคล้ายคลึงกันมากที่สุด โดยในกรณีนี้ความคล้ายคลึงประเมินโดยการคำนวณค่า Euclidean distance กล่าวคือ map unit ที่มีลักษณะคล้ายคลึงกับตัวอย่าง \mathbf{x}_2 มากที่สุดจะมีค่า Euclidean distance ต่ำที่สุดและถูกกำหนดให้เป็น "best matching unit" หรือ BMU ของตัวอย่างที่ถูกเลือกมา หลังจากนั้นเมื่อ map unit ที่อยู่ล้อมรอบ BMU จะถูกกำหนดและเรียกเป็น "neighbor" ของ BMU เพื่อใช้ในการกำหนดปริมาณการเรียนรู้ โดย BMU จะถูกปรับให้มีความคล้ายคลึงกับตัวอย่าง \mathbf{x}_2 มากที่สุด ตามด้วย neighbor ที่อยู่ใกล้ๆ ส่วน neighbor ที่อยู่ไกลออกไปจะถูกปรับให้มีความคล้ายคลึงน้อยที่สุดตามลำดับความใกล้ชิด กระบวนการเรียนรู้นี้ดังกล่าวนี้จะถูกทำซ้ำอย่างต่อเนื่องจนกว่าจะถึงจำนวนรอบของการเรียนรู้ (number of iterations) ที่กำหนดไว้ รูปที่ 4.6 แสดงลักษณะการเรียนรู้ของ SOM ที่ใช้ในการศึกษานี้



รูปที่ 4.6 ขั้นตอนการเรียนรู้ของ SOM ที่กำหนดให้ map unit มีรูปทรงหกเหลี่ยม

โดยปกติขนาดของ map สามารถปรับได้ตามลักษณะของข้อมูล เช่น ข้อมูลที่มีความซับซ้อนน้อยหรือตัวอย่างแต่ละตัวมีความแตกต่างกันไม่มากหรือมีจำนวนกลุ่ม (class membership) ไม่มาก จำนวน map unit ที่ใกล้เคียงกับจำนวนของตัวอย่างก็เพียงพอที่จะสามารถอธิบายสมบัติของข้อมูลดังกล่าวได้ แต่สำหรับในกรณีที่ข้อมูลมีความแตกต่างกันมาก ๆ หรือจำนวนกลุ่มของข้อมูลมีขนาดใหญ่แบบจำลอง SOM อาจต้องการจำนวนของ map unit เพิ่มขึ้นเพื่อใช้ในการอธิบายความแตกต่างของกลุ่มข้อมูล รวมถึงการมีจำนวน map unit ที่เพียงพอต่อการแสดงจุดเชื่อมต่อระหว่างข้อมูลที่มาจากรวมถึง class membership ที่แตกต่างกัน นอกจากการเพิ่มจำนวนของ map unit ในบางกรณีจำนวนของ map unit ที่ใช้อาจถูกเพิ่มขึ้น เพื่อแสดงข้อมูลที่มีจำนวนมาก หรือข้อมูลที่มีลักษณะบางประการที่แตกต่างกัน การเพิ่มจำนวนของ map นอกจากจะเพิ่มจำนวนของ map unit แล้วยังทำให้แบบจำลองที่ได้มีความยืดหยุ่นในการจัดแบ่งข้อมูล เช่น การกำหนดให้แต่ละ map ที่สร้างขึ้นแสดงพฤติกรรมของตัวอย่างที่มาจากกลุ่มที่แตกต่างกันหรือมี class membership ที่แตกต่างกัน เป็นต้น ลักษณะการเพิ่มจำนวนของ map ที่ใช้ในการสร้างแบบจำลองดังกล่าวอาจเรียกได้ว่าเป็น multiple SOMs ซึ่งแสดงในรูปที่ 4.7 สำหรับการใช่มultiple SOM ในการจำลองพฤติกรรมของ training samples ที่มีจำนวนกลุ่มที่แตกต่างกันอยู่ 3 กลุ่ม

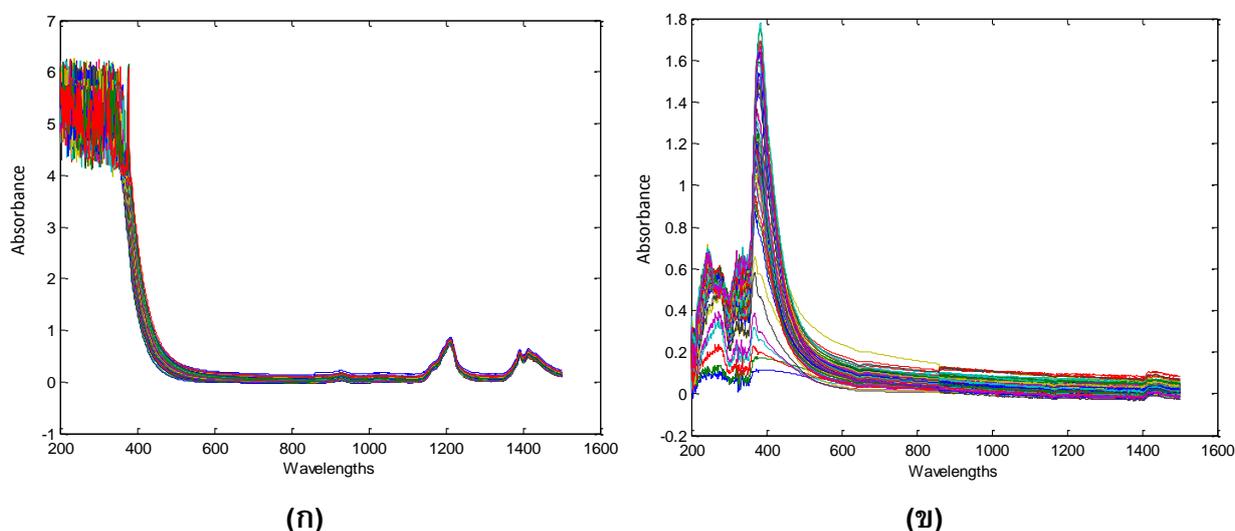


รูปที่ 4.7 การประยุกต์ใช้ multiple SOMs สำหรับข้อมูลที่มี 3 กลุ่ม (class membership)

5. ผลการทดลอง

5.1 ผลที่ได้จากการวิเคราะห์ด้วยเทคนิค NIR spectroscopy

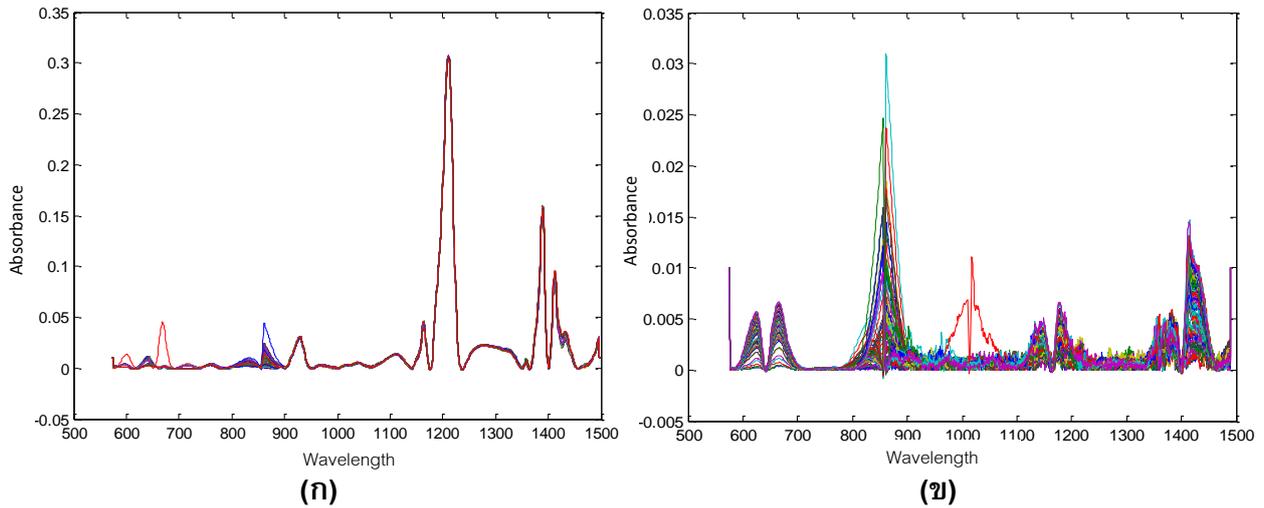
ในการวิเคราะห์น้ำมันที่ใช้ประกอบอาหารที่ใช้ไปแล้วทั้งหมดประมาณ 9 ชั่วโมง เก็บตัวอย่าง ทุกๆครั้งที่ทอด ครั้งละประมาณ 5 นาที จากนั้นนำไปวิเคราะห์ผลโดยเทคนิค NIR spectroscopy ได้ สเปกตรัม ดังรูป 5.1



รูป 5.1 สเปกตรัมของข้อมูลน้ำมันที่ใช้ประกอบอาหารทั้งหมดที่ได้จากการวิเคราะห์ด้วย NIR spectroscopy (ก) ใช้อากาศเป็นตัวอ้างอิง (ข) ใช้น้ำมันใหม่เป็นตัวอ้างอิง

รูป 5.1 แสดงเส้นสเปกตรัมของข้อมูลตัวอย่างน้ำมันทั้งหมด 68 ตัวอย่าง วัดค่าการดูดกลืนแสงที่ ความยาวคลื่นช่วง 200 – 1500 nm ซึ่งอยู่ในช่วงทั้ง UV-Vis และ NIR เห็นได้ว่าการเปลี่ยนแปลงของ สเปกตรัมเกิดขึ้น แต่สเปกตรัมที่ได้มีความซับซ้อนมากเกินไป ทำให้สังเกตถึงความเปลี่ยนแปลงนั้นได้ ไม่ชัดเจน ดังนั้นจึงนำเทคนิค PCA เข้ามาช่วยเพื่อให้การวิเคราะห์ข้อมูลนั้นง่ายขึ้น

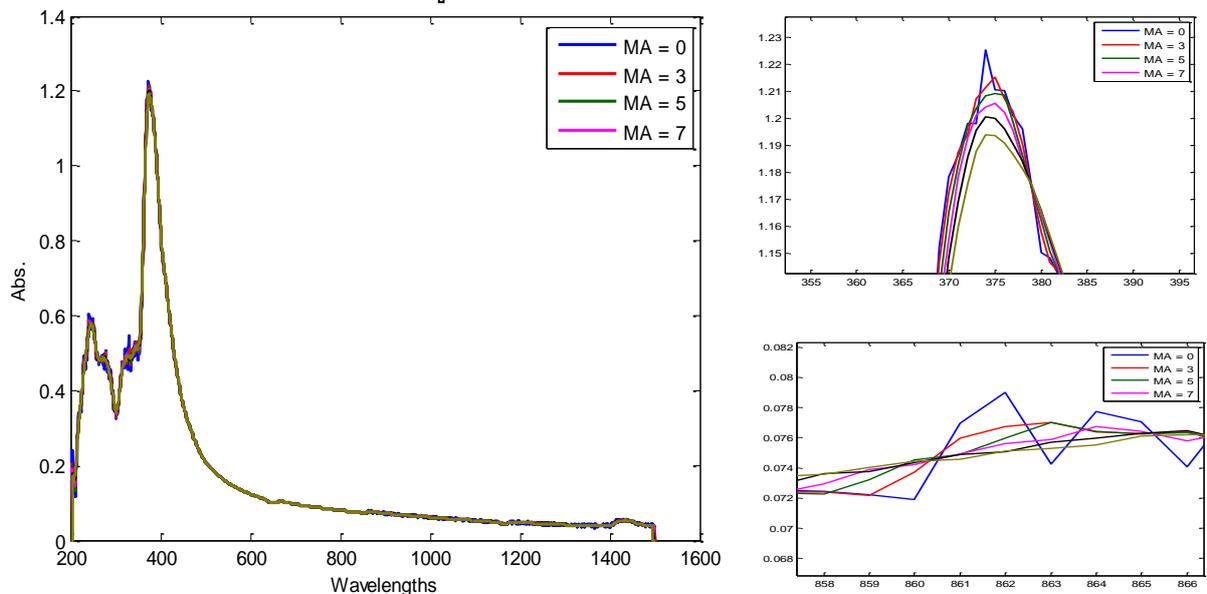
จากสเปกตรัมที่ได้นั้น สังเกตเห็นว่าในช่วงความยาวคลื่น 200 – 550 nm มีสัญญาณการ เปลี่ยนแปลงของสเปกตรัมที่ค่อนข้างสูง โดยในช่วงความยาวคลื่นนี้จะเป็นช่วงของ UV-Vis ซึ่งเป็นช่วง ของการวิเคราะห์การเปลี่ยนแปลงสีของน้ำมัน และเนื่องจาก spectrum ที่ปรากฏในช่วงความยาวคลื่น 200-550 nm สัญญาณที่สูง ทำให้การสังเกตการเปลี่ยนแปลงเส้นสเปกตรัมในช่วงความยาวคลื่น 557– 1500 nm ซึ่งอยู่ในช่วงของ UV-Vis และ NIR นั้นทำได้ยาก ดังนั้นในการวิเคราะห์ครั้งนี้เราจึงเลือกใช้ ความยาวคลื่นในช่วงตั้งแต่ 557 – 1500 nm โดยจะได้สเปกตรัมดังรูป 5.2



รูป 5.2 สเปกตรัมของข้อมูลตัวอย่างน้ำมันที่วิเคราะห์ด้วย NIR spectroscopy ช่วงความยาวคลื่น 557 – 1500 nm (ก) ใช้อากาศเป็นตัวอ้างอิง (ข) ใช้น้ำมันใหม่เป็นตัวอ้างอิง

5.1.1 ผลที่ได้จากการทำ data preprocessing ได้แก่ moving average และ baseline correction

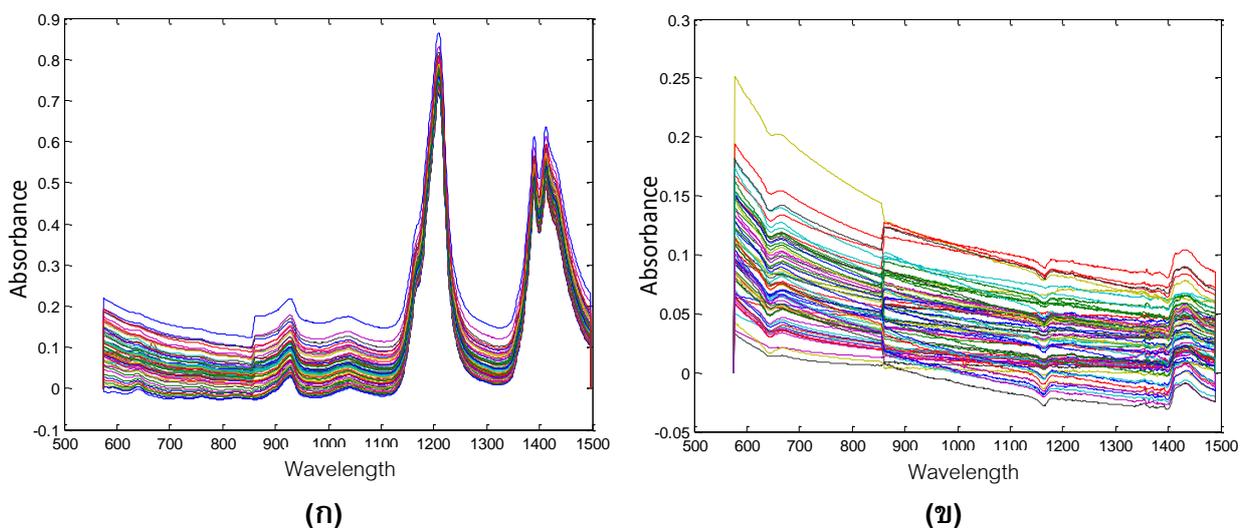
ก่อนที่จะใช้ PCA นั้น จะต้องมีเตรียมข้อมูลให้เหมาะสมกับการวิเคราะห์ก่อน โดยการทำ moving average (MA) และ baseline correction (BC) เพื่อลดปัญหาสัญญาณรบกวนจากเครื่องมือหรือผลของ baseline ที่ไม่เท่ากัน ดังรูป 5.3



รูป 5.3 การทำ moving average 3, 5 และ 7 ของตัวอย่างน้ำมันตัวที่ 10 (รูปเล็กทางด้านขวา แสดงส่วนขยายของจุดยอดพีคและหางพีค)

รูป 5.3 เป็นการทำ moving average 3, 5 และ 7 ของตัวอย่างน้ำมันตัวที่ 10 (สุ่มเลือกมาเพียงตัวอย่างเดียว) เพื่อหาว่าควรใช้ช่วงของ MA เท่ากับเท่าไรถึงจะเหมาะสมกับข้อมูลมากที่สุด โดยที่ MA = 0 (สีน้ำเงิน) คือเส้นสเปกตรัมที่ยังไม่ได้ทำ moving average จะเห็นว่ามึลักษณะคล้ายฟันปลา ที่ MA = 3 (สีแดง) กราฟจะเรียบและมีความเป็นระเบียบมากกว่าที่ MA = 0 แต่ก็ยังเห็นได้ชัดว่าเส้นสัญญาณยังไม่นิ่งพอ ที่ MA = 7 (สีชมพู) จะมีลักษณะที่มีความเรียบเป็นระเบียบมาก แต่อาจทำให้สัญญาณบางสัญญาณซึ่งอาจมีผลต่อการเปลี่ยนแปลงของข้อมูลหายไป ดังนั้นจึงเลือกใช้ MA = 5 (สีเขียว) เพราะเส้นสเปกตรัมมีความ smooth ที่เหมาะสม และยังคงใกล้เคียงกับค่าเดิมก่อนการทำ moving average

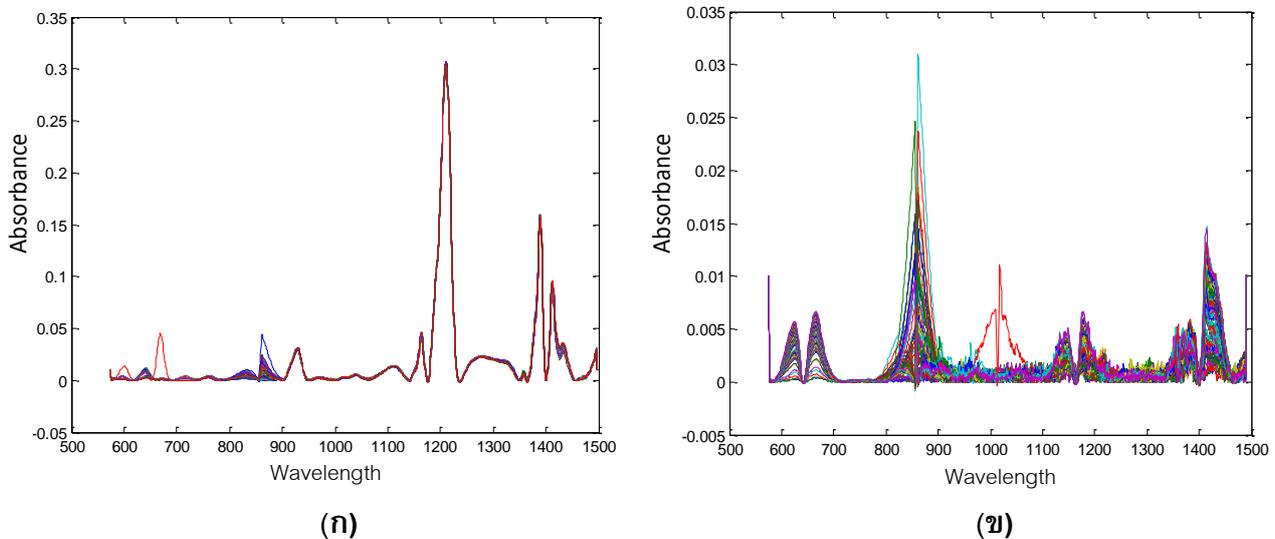
เมื่อเราได้ค่า MA ที่เหมาะสมกับข้อมูลแล้ว จึงนำเอาข้อมูลทั้งหมดไปทำ moving average โดยใช้ค่า MA = 5 จะได้ผลดังรูป 5.4



รูป 5.4 การทำ moving average (MA = 5) ของข้อมูลตัวอย่างน้ำมันทั้งหมด (ก) ใช้อากาศเป็นตัวอ้างอิง (ข) ใช้น้ำมันใหม่เป็นตัวอ้างอิง

จากรูป 5.4 พบว่า เมื่อทำ moving average แล้ว เส้นสเปกตรัมที่ได้มีความเป็นระเบียบมากขึ้น มีการซ้อนทับของสัญญาณตัวอย่างแต่ละตัวน้อยลง ทำให้สังเกตเห็นความเปลี่ยนแปลงของตัวอย่างน้ำมันแต่ละตัวได้ง่ายขึ้น แต่จะเห็นว่าข้อมูลแต่ละตัวจะมี baseline ที่แตกต่างกัน โดยอาจเกิดอุณหภูมิ ความร้อนจากเครื่องมือ หรือความชื้นที่เกิดขึ้นในการทดลอง ดังนั้นก่อนการนำข้อมูลที่ไปวิเคราะห์ จึงต้องมีการทำ baseline correction ก่อนเพื่อลดปัญหาที่จะเกิดขึ้นจาก baseline

การทำ baseline correction เป็นการช่วยให้ข้อมูลทุกตัวมี baseline เท่ากัน เพื่อลดปัญหาที่เกิดจากความร้อน ความชื้น ของเครื่องมือขณะที่ทำการวิเคราะห์ จากข้อมูลที่ผ่านมาการทำ moving average ข้างต้น เราก็นำมาทำ baseline correction ได้ดังรูป 5.5



รูป 5.5 การทำ baseline correction ของข้อมูลตัวอย่างน้ำมันทั้งหมดในช่วง 557 – 1500 nm (ก) ใช้อากาศเป็นตัวอ้างอิง (ข) ใช้น้ำมันใหม่เป็นตัวอ้างอิง

เมื่อได้ข้อมูลที่เหมาะสมแล้ว จึงนำไปสู่กระบวนการวิเคราะห์ในขั้นตอนต่อไป ได้แก่ PCA และ PLS ตามลำดับ

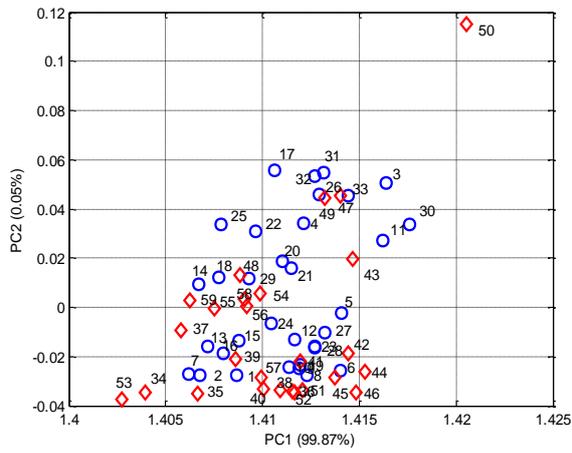
5.2 ผลที่ได้จากการทำ data preprocessing ได้แก่ normalization, log scaling, standardization และ mean centring โดยวิเคราะห์ด้วยเทคนิค PCA

เมื่อเราได้ข้อมูลที่เหมาะสมซึ่งผ่านการทำ moving average และ baseline correction แล้วต่อไปจึงนำมาทำ normalization, log scaling, standardization และ mean centring เพื่อพิจารณาว่า data preprocessing แบบไหนที่เหมาะสม และสามารถวิเคราะห์ข้อมูลได้ดีที่สุดสำหรับการวิเคราะห์ผลด้วย PCA ดังต่อไปนี้

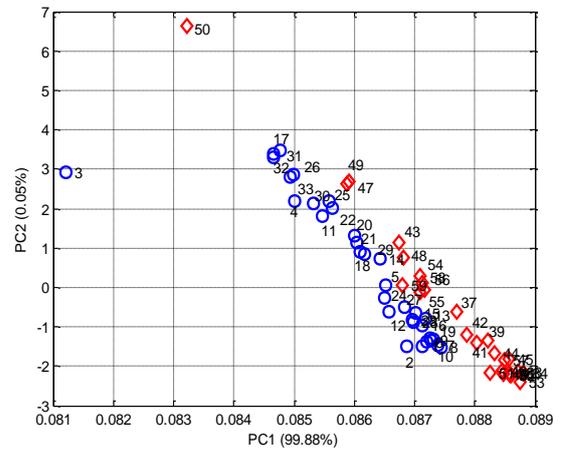
5.2.1 ผลของตัวอย่างน้ำมันที่ใช้อากาศเป็นตัวอ้างอิง

รูป 5.6 เป็นการทำ PCA ของข้อมูลน้ำมันที่ใช้อากาศเป็นตัวอ้างอิง โดยวงกลมสีฟ้าแทนตัวอย่างน้ำมันที่ทอดในวันที่หนึ่ง และสี่เหลี่ยมสีแดงแทนตัวอย่างน้ำมันที่ทอดในวันที่สอง โดยที่มีข้อมูล

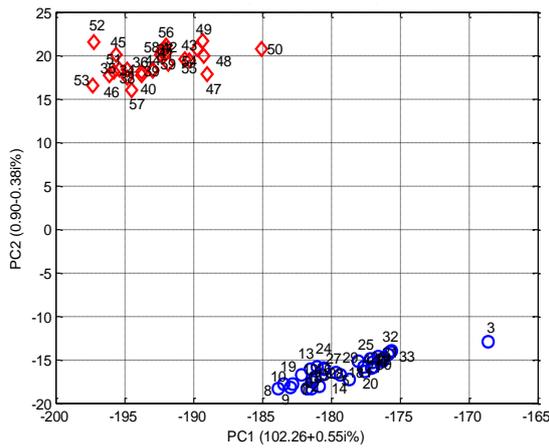
เพียงสองวันเนื่องจากวันที่สามน้ำมันตัวอย่างเริ่มเป็นไข ทำให้ไม่สามารถนำไปวิเคราะห์ข้อมูลได้ รูป 5.6 ก เป็นข้อมูลที่ไม่ทำ preprocessing จะเห็นว่าลักษณะข้อมูลของทั้งสองวันซ้อนทับกันทำให้เห็นการเปลี่ยนแปลงได้ไม่ชัดเจน จึงได้มีการทำ data preprocessing ได้แก่ normalization, log scaling, standardization และ mean centring เป็นต้น รูป 5.6 ข เป็นการทำ normalization หรือการทำให้ข้อมูลทุกตัวมีความสำคัญเท่ากัน เริ่มสังเกตเห็นความเปลี่ยนแปลงได้เล็กน้อย แต่ก็ยังไม่ชัดเจน รูป 5.6 ค รูป 5.6 ง และ รูป 5.6 จ เป็นการทำ log scaling, standardization และ mean centring ตามลำดับ ลักษณะข้อมูลที่ได้จะเห็นว่าข้อมูลของน้ำมันแยกกลุ่มกันอย่างชัดเจน ในทางทฤษฎีอาจเป็นผลการทดลองที่ดี เนื่องจากในการทดลองผู้ทดลองสามารถแยกน้ำมันที่ใช้ไปในแต่ละวันได้ชัดเจน แต่ในความเป็นจริง มีทั้งปัจจัยต่างๆ มากมายที่ทำให้ผลการทดลองนี้ได้มาในลักษณะรูปแบบของกลุ่มข้อมูลแตกต่างกัน เนื่องจากการใช้อากาศเป็นตัวอ้างอิง อาจมีผลของความชื้น หรืออุณหภูมิบริเวณห้องทดลองในแต่ละวันที่ต่างกันมาด้วย ไม่สามารถระบุได้แน่ชัดว่า การที่กลุ่มข้อมูลแยกจากกันอย่างชัดเจนนี้เกิดจากองค์ประกอบภายใน หรือปัจจัยต่างๆ ภายนอก ดังนั้นเพื่อให้ได้ผลการทดลองที่แน่นอน ควรใช้น้ำมันใหม่เป็น blank หรือตัวอ้างอิงในการทำการทดลอง ซึ่งจะได้ลักษณะของผลการทดลองเป็นดังต่อไปนี้



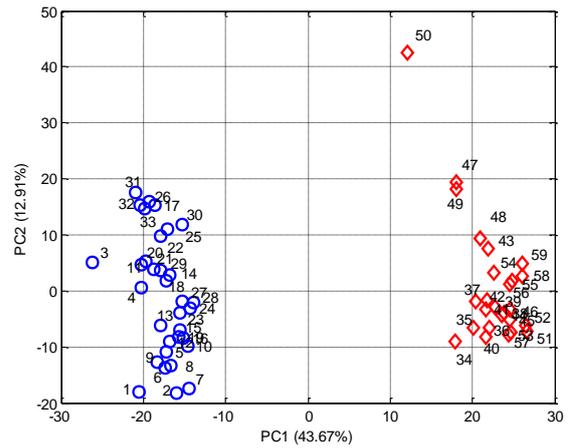
(ก)



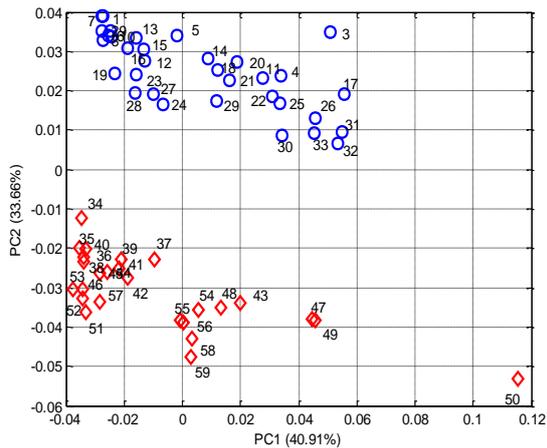
(ข)



(ค)



(ง)



(จ)

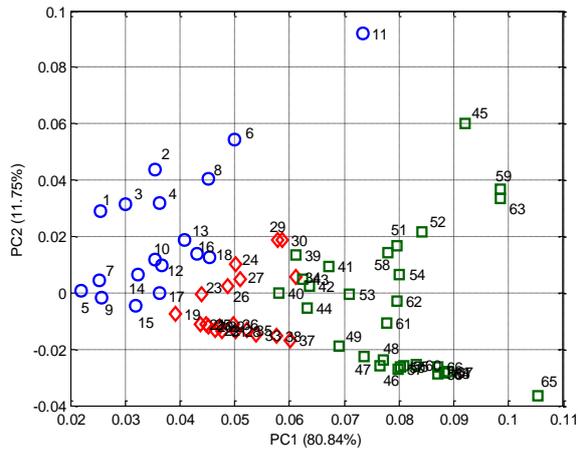
○ → ตัวอย่างน้ำมันที่ทอดในวันที่ 1
 ◇ → ตัวอย่างน้ำมันที่ทอดในวันที่ 2

รูป 5.6 การทำ data preprocessing (ก) ข้อมูลที่ไม่มีการทำ data preprocessing (ข) การทำ normalization (ค) การทำ log scaling (ง) การทำ standardization และ (จ) การทำ mean centring

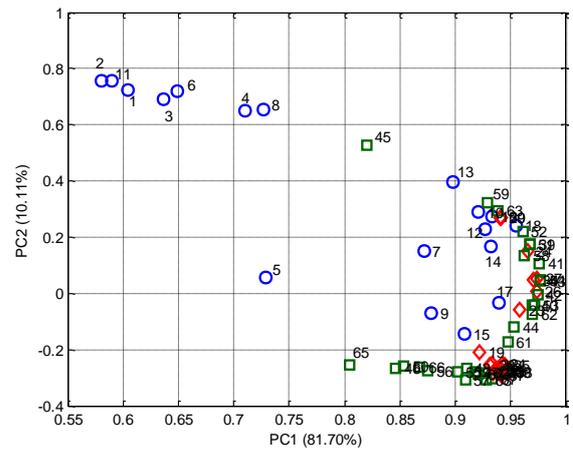
5.2.2 ผลของตัวอย่างน้ำมันที่ใช้น้ำมันใหม่เป็นตัวอ้างอิง

รูป 5.7 แสดงผลการทำ data processing ของตัวอย่างน้ำมันที่ใช้น้ำมันใหม่เป็นตัวอ้างอิง จากรูปกำหนดให้วงกลมสีฟ้าแทนตัวอย่างน้ำมันที่ทอดในวันที่หนึ่ง สีเหลี่ยมสีแดงแทนตัวอย่างน้ำมันที่ทอดในวันที่สอง และสีเหลี่ยมสีเขียวแทนตัวอย่างน้ำมันที่ทอดในวันที่สาม จากรูป 5.7 ก จะเห็นว่าลักษณะของข้อมูลทั้งหมดจะมีการเปลี่ยนแปลงไปอย่างมีรูปแบบ (pattern) จากวันที่หนึ่งค่อยๆ เลื่อนไปจนถึงวันที่สาม แต่ยังเห็นข้อมูลบางส่วนซ้อนทับกันอยู่ ทำให้พิจารณาได้ไม่ชัดเจน ดังนั้นจึงทำ data preprocessing เพื่อช่วยให้สังเกตข้อมูลได้ง่ายขึ้น ได้แก่ normalization, log scaling, standardization และ mean centring จากรูป 5.7 ข เป็นการทำ normalization จะเห็นว่ากลุ่มข้อมูลในวันที่หนึ่งจะกระจายตัวมากขึ้น แต่ข้อมูลของวันที่สอง และสามซ้อนทับกัน จนแทบสังเกตความแตกต่างไม่ได้ ส่วนรูป 5.7 ค เป็นการทำ log scaling หรือการทำให้ช่วงของข้อมูลแต่ละความยาวคลื่นมีความสำคัญใกล้เคียงกัน คือความเปลี่ยนแปลงของแต่ละพีคในสเปกตรัมจะถูกพิจารณาให้เท่าเทียมกัน จากรูป จะสังเกตเห็นว่าข้อมูลที่ได้อาจมีการเปลี่ยนแปลงรูปแบบของกลุ่มข้อมูลที่ชัดเจนมากขึ้นตั้งแต่วันที่หนึ่งถึงวันที่สาม ส่วนในรูป 5.7 ง เป็นการทำ standardization คือการทำให้ข้อมูลแต่ละความยาวคลื่นมีความสำคัญเท่ากัน จะเห็นว่าการกระจายตัวของข้อมูลมีมากขึ้น แต่ไม่ค่อยเรียงตัวกัน ทำให้อาจพิจารณาถึงการเปลี่ยนแปลงได้ยาก เพราะกลุ่มข้อมูลของวันที่หนึ่ง และวันที่สามแทบจะคล้ายกัน และในรูป 5.7 จ ที่เป็นการทำ mean centring เป็นการทำให้จุดศูนย์กลางข้อมูลทั้งหมดอยู่ที่จุด origin ได้ ลักษณะข้อมูลที่คล้ายกับ รูป 5.7 ง แต่เป็นแบบกลับหัว ซึ่งก็มีการกระจายที่ไม่ค่อยดี มีการซ้อนทับกันของกลุ่มข้อมูลแต่ละวัน ทำให้ดูลักษณะการเปลี่ยนแปลงได้ไม่ค่อยชัดเจน

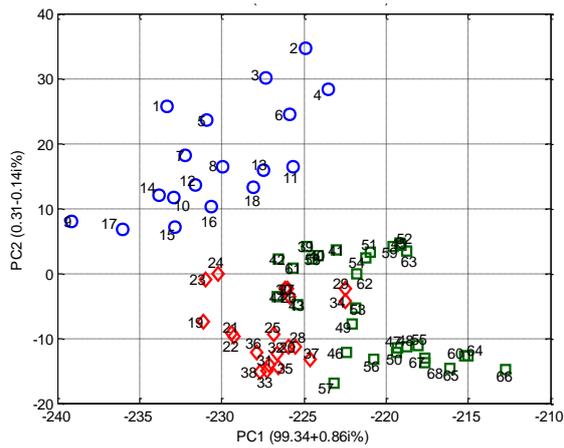
จากรูป 5.7 นั้น เมื่อพิจารณาการทำ data preprocessing ทั้ง 4 แบบ (normalization, log scaling, standardization และ mean centring) พบว่าแบบที่เหมาะสมสำหรับข้อมูลตัวอย่างน้ำมันมากที่สุดคือการทำ log scaling เนื่องจากสังเกตเห็นการเปลี่ยนแปลงได้ชัดเจนมากที่สุด ส่วน normalization ที่ไม่เลือกเพราะจากที่กล่าวไว้ในข้างต้นว่า การทำ normalization คล้ายกับการทำ baseline correction คือการลดปัญหา baseline ไม่เท่ากัน ดังนั้นจึงไม่เลือกใช้ normalization เนื่องจากในตอนแรกได้ทำ baseline correction ไปแล้ว และการเปลี่ยนแปลงของข้อมูลก็เห็นได้ไม่ชัดเจน สำหรับ standardization ซึ่งเป็นการทำให้ข้อมูลของทุกๆ ความยาวคลื่นมีความสำคัญเท่ากัน ที่ไม่เลือกใช้เพราะการทำ standardization นี้อาจไปทำให้สัญญาณที่เป็นของตัวรบกวนอื่นๆ เข้ามามีผล ซึ่งจะทำการวิเคราะห์ผลคลาดเคลื่อนไปด้วย และสำหรับ mean centring ลักษณะของข้อมูลที่ได้นั้นไม่ค่อยชัดเจน ดังนั้นการทำ log scaling จึงถือว่าเหมาะสมสำหรับการวิเคราะห์ข้อมูลตัวอย่างน้ำมันมากที่สุด



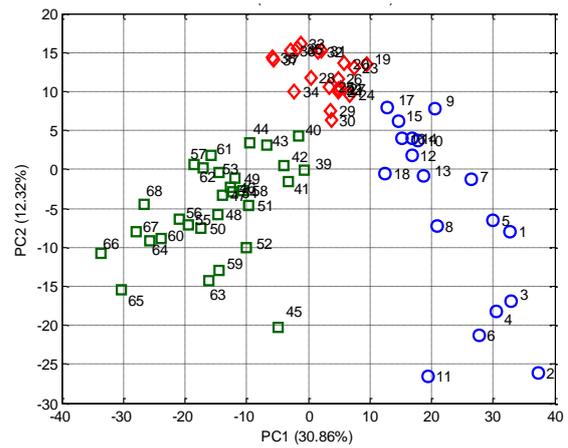
(ก)



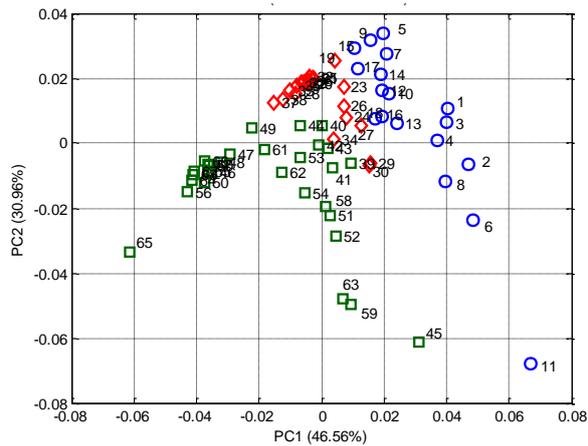
(ข)



(ค)



(ง)



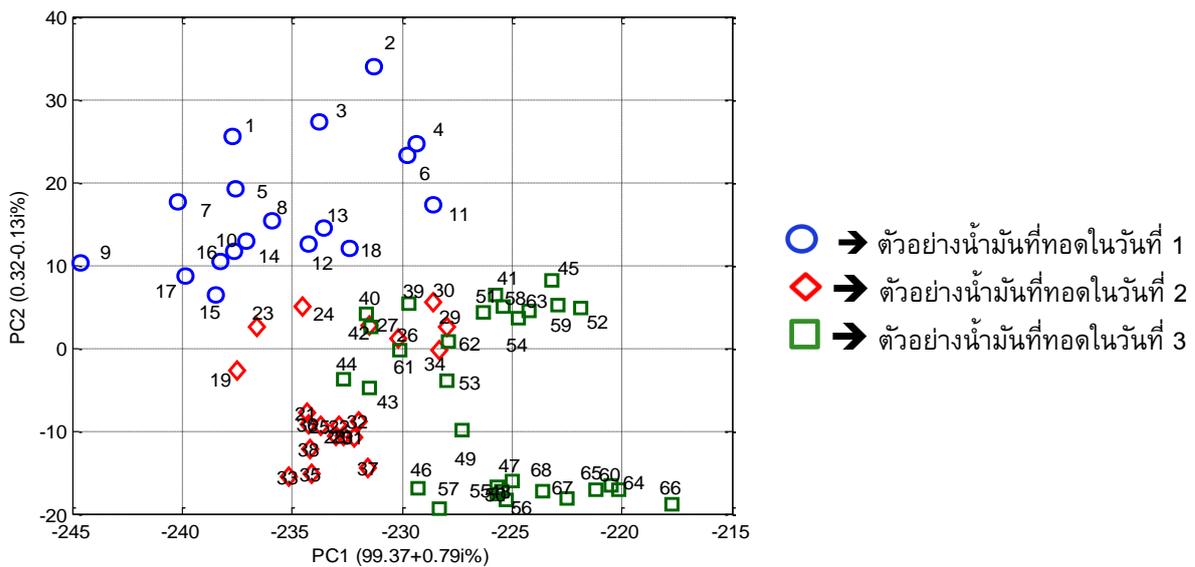
(จ)

- → ตัวอย่างน้ำมันที่ทอดในวันที่ 1
- ◇ → ตัวอย่างน้ำมันที่ทอดในวันที่ 2
- → ตัวอย่างน้ำมันที่ทอดในวันที่ 3

รูป 5.7 การทำ data preprocessing (ก) ข้อมูลที่ไม่มีการทำ data preprocessing (ข) การทำ normalization (ค) การทำ log scaling (ง) การทำ standardization และ (จ) การทำ mean centring

5.2.3 ผลจากการวิเคราะห์ด้วยเทคนิค PCA และ log scaling

รูป 5.8 แสดงผลการทำ PCA และ log scaling ของข้อมูลตัวอย่างน้ำมันทั้งหมดกำหนดให้ จากผลการวิเคราะห์ที่ได้สามารถเห็นการเปลี่ยนแปลงของข้อมูลตัวอย่างน้ำมันที่ผ่านการทอดซ้ำเป็นเวลาสามวันได้อย่างชัดเจน โดยจะแยกเป็นสามกลุ่ม กลุ่มแรกคือ วันที่หนึ่งจะมีการกระจายตัวมาก อาจเกิดจากปฏิกิริยาที่เกิดขึ้นในระหว่างการทอด ทั้งความชื้น และอุณหภูมิ จะทำให้องค์ประกอบภายในของน้ำมันเช่น free fatty acid หรือสารโพลาร์เปลี่ยนแปลงไปอย่างรวดเร็ว ส่วนในกลุ่มที่สองและสาม ข้อมูลตัวอย่างน้ำมันจะค่อยๆ เปลี่ยนแปลง ไม่มากเท่ากับวันแรก นั่นคือการเปลี่ยนแปลงองค์ประกอบภายในเริ่มคงที่และค่อยๆ เปลี่ยนไปซ้ำๆ



รูป 5.8 ผลการทำ PCA และ log scaling ของข้อมูลตัวอย่างน้ำมันทั้งหมดกำหนดให้

5.3 ผลจากการวิเคราะห์ด้วยเทคนิค PLS

การวิเคราะห์ผลของ PLS เป็นการทำนายจำนวนครั้งของน้ำมันที่ผ่านการทอดซ้ำ โดยใช้ข้อมูลค่าการดูดกลืนแสงจาก NIR สร้างโมเดล ค่า RMSE เป็นค่าที่บอกถึงความคลาดเคลื่อนของค่าที่ทำนายได้เทียบกับค่าจริง ดังนั้นค่าที่ได้ต้องนำไปตรวจสอบเข้ากับค่าจำนวนครั้งจากการทำนาย ส่วนค่า R^2 และ Q^2 จะเป็นตัวบ่งบอกถึงประสิทธิภาพของโมเดลที่สร้างขึ้น โดยค่า R^2 และ Q^2 ต้องมีค่ามากที่สุด และใกล้เคียงกันมากที่สุด ถึงจะถือว่าโมเดล PLS ที่สร้างขึ้นมีประสิทธิภาพดี

การวิเคราะห์ผลด้วย PLS จะวิเคราะห์ทั้งหมด 5 แบบ

สัญลักษณ์ย่อ	Data preprocessing
non	ไม่ทำ preprocessing
mean centring	mean centring
log	log scaling
std	Standardization (auto scaling)
norm	Normalization (sum = 1)

โดยข้อมูลที่ใช้เป็นตัวสร้างโมเดลจะแบ่งออกเป็น 2 แบบ คือ การใช้ตัวอย่างที่เป็นเลขคู่เป็นตัวสร้างโมเดล แล้วตัวอย่างที่เป็นเลขคี่เป็นตัวทดสอบ เพื่อต้องการดูว่าข้อมูลที่วิเคราะห์ได้ สามารถนำมาสร้างเป็นโมเดลสำหรับทำนายจำนวนครั้งได้หรือไม่ และการใช้ตัวอย่างตัวที่ 1, 5, 10, 15,....., n (n คือ จำนวนตัวอย่าง) เป็นตัวสร้างโมเดล และตัวอย่างตัวที่เหลือทั้งหมดเป็นตัวทดสอบ เพื่อดูว่าหากเราใช้ข้อมูลจำนวนครั้งที่ห่างกันมาสร้างเป็นโมเดล โมเดลที่เราสร้างได้นั้นยังสามารถให้ผลการทำนายที่ดีอยู่หรือไม่ ได้ผลดังต่อไปนี้

5.3.1 ผลที่ได้จากข้อมูลที่ใช้อากาศเป็นตัวอ้างอิง

ตาราง 5.1 ผลที่ได้จากการใช้ตัวอย่างที่เป็นเลขคู่เป็นตัวสร้างโมเดล แล้วตัวอย่างที่เป็นเลขคี่เป็นตัวทดสอบ จากตัวอย่างทั้งหมด 59 ตัวอย่าง

Data preprocessing	200-1500 nm		575-1500 nm		860-1500 nm		200-575 nm	
	PCs	RMSE	PCs	RMSE	PCs	RMSE	PCs	RMSE
Non	7	3.34	4	4.29	7	2.67*	4	6.55
Log	3	3.70	7	2.82	4	7.35	3	4.40
Norm	6	3.42	4	4.43	6	3.14	4	6.28
Centring	6	3.77	3	4.29	5	3.06	3	6.13
Std	5	2.82	1	6.12	7	3.33	2	3.37

*ค่า RMSE ที่น้อยที่สุด

ตาราง 5.2 ผลที่ได้จากการใช้ตัวอย่างตัวที่ 1, 5, 10, 15,...., 59 เป็นตัวสร้างโมเดล และตัวอย่างตัวที่เหลือทั้งหมดเป็นตัวทดสอบ จากตัวอย่างทั้งหมด 59 ตัว

Data preprocessing	200-1500 nm		575-1500 nm		860-1500 nm		200-575 nm	
	PCs	RMSE	PCs	RMSE	PCs	RMSE	PCs	RMSE
Non	3	8.72	3	6.21	7	5.13	3	8.59
Log	4	4.39*	6	5.27	3	9.30	3	5.36
Norm	3	8.42	5	4.60	7	5.32	3	8.37
Centring	2	9.30	2	6.22	8	4.78	2	9.21
Std	5	10.77	1	6.25	1	7.09	3	5.17

*ค่า RMSE ที่น้อยที่สุด

จากตาราง 5.1 ช่วงที่เหมาะสมสำหรับการวิเคราะห์ด้วยเทคนิค PLS คือ ช่วงความยาวคลื่น 860 – 1500 nm โดยไม่ทำ preprocessing ใช้ PC = 7 และมีค่า RMSE = 2.67 และจากตาราง 5.2 ช่วงที่เหมาะสมสำหรับการวิเคราะห์ด้วยเทคนิค PLS คือ ช่วงความยาวคลื่น 200 – 1500 nm โดยใช้ preprocessing คือ log scaling ใช้ PC = 4 และมีค่า RMSE = 4.39 ซึ่งสังเกตเห็นว่าค่า RMSE ของโมเดลที่สร้างจากจำนวนครั้งเป็นเลขคู่จะมีค่าน้อยกว่า เนื่องจากจำนวนข้อมูลที่นำมาสร้างเป็นโมเดลนั้นค่อนข้างมากทำให้การทำนายจำนวนครั้งมีความผิดพลาดน้อย เมื่อเทียบกับโมเดลที่สร้างจากจำนวนครั้งที่เป็นเลข 1,5,10....,59 เนื่องจากข้อมูลค่อนข้างห่างกัน ทำให้การทำนายจำนวนครั้งมีโอกาสผิดพลาดได้มากกว่า

ตาราง 5.3 ค่า Q^2 และ R^2 ของการใช้ตัวอย่างที่เป็นเลขคู่เป็นตัวสร้างโมเดล แล้วตัวอย่างที่เป็นเลขคี่เป็นตัวทดสอบ จากตัวอย่างทั้งหมด 59 ตัวอย่าง

Data preprocessing	200-1500 nm		575-1500 nm		860-1500 nm		200-575 nm	
	Q^2	R^2	Q^2	R^2	Q^2	R^2	Q^2	R^2
Non	0.9600	0.9967	0.9343	0.9479	0.9745*	0.9883*	0.8469	0.9538
Log	0.9511	0.9687	0.9716	0.9993	0.8071	0.9223	0.9308	0.9741
Norm	0.9583	0.9870	0.9298	0.9530	0.9649	0.9770	0.8590	0.9711
Centring	0.9493	0.9937	0.9342	0.9482	0.9667	0.9797	0.8656	0.9596
Std	0.9715	0.9964	0.8661	0.8539	0.9603	0.9978	0.9595	0.9841

*ค่า R^2 และ Q^2 ที่สูงที่สุด และเหมาะสมที่สุด

ตาราง 5.4 ค่า Q^2 และ R^2 ของการใช้ตัวอย่างตัวที่ 1, 5, 10, 15, ..., 59 เป็นตัวสร้างโมเดล และตัวอย่างตัวที่เหลือทั้งหมดเป็นตัวทดสอบ จากตัวอย่างทั้งหมด 59 ตัว

Data preprocessing	200-1500 nm		575-1500 nm		860-1500 nm		200-575 nm	
	Q2	R2	Q2	R2	Q2	R2	Q2	R2
Non	0.7241	0.9437	0.8599	0.9554	0.9045	0.9876	0.7322	0.945
Log	0.9300*	0.9966*	0.8992	0.9662	0.6861	0.9334	0.8958	0.9742
Norm	0.7426	0.9546	0.9231	0.9748	0.8972	0.9665	0.7461	0.9571
Centring	0.6860	0.9502	0.8596	0.9553	0.9169	1.0000	0.6925	0.9514
Std	0.5789	0.9599	0.8582	0.9487	0.8176	0.9541	0.9031	0.9979

*ค่า R^2 และ Q^2 ที่สูงที่สุด และเหมาะสมที่สุด

จากตาราง 5.3 ค่า R^2 และ Q^2 ที่มากที่สุด และเหมาะสมที่สุด อยู่ในช่วงความยาวคลื่น 860–1500 nm โดยมีค่าเท่ากับ 0.9883 และ 0.9745 ตามลำดับ และจากตาราง 5.4 ค่า R^2 และ Q^2 ที่มากที่สุด และเหมาะสมที่สุด อยู่ในช่วงความยาวคลื่น 200–1500 nm โดยมีค่าเท่ากับ 0.9966 และ 0.9300 ตามลำดับ data preprocessing ที่ใช้คือ log scaling หากเปรียบเทียบทั้งสองตารางนี้จะเห็นว่า ตารางที่ 5.3 จะได้ค่า R^2 และ Q^2 ที่ดีกว่าเพราะ มีค่าสูงและใกล้เคียงกัน ส่วนในตารางที่ 5.4 ค่า R^2 และ Q^2 ที่มากที่สุดจะอยู่ในช่วงความยาวคลื่น 200-1500 nm ซึ่งในช่วงนี้อาจมีผลจากสัญญาณรบกวนในช่วงของ UV-Vis มาเกี่ยวข้องด้วย นอกจากนี้ยังเป็นโมเดลที่สร้างจากจำนวนครั้งที่ห่างกันมากกว่า ความผิดพลาดในการทำนายจึงมากกว่า ดังนั้นค่า R^2 และ Q^2 ที่เหมาะสมที่สุดสำหรับข้อมูลน้ำมันที่ใช้อากาศเป็นตัวอ้างอิงจึงเป็นข้อมูลที่ตาราง 5.3

5.3.2 ผลที่ได้จากข้อมูลที่ใช้น้ำมันใหม่เป็นตัวอ้างอิง

ตาราง 5.5 ผลที่ได้จากการใช้ตัวอย่างที่เป็นเลขคู่เป็นตัวสร้างโมเดล แล้วตัวอย่างที่เป็นเลขคี่เป็นตัวทดสอบ จากตัวอย่างทั้งหมด 68 ตัวอย่าง

Data preprocessing	200-1500 nm		575-1500 nm		860-1500 nm		200-575 nm	
	PCs	RMSE	PCs	RMSE	PCs	RMSE	PCs	RMSE
Non	6	3.55	10	3.34	6	3.37	7	3.17
Log	7	4.84	7	7.19	7	7.24	3	3.67
Norm	7	4.30	7	8.44	7	11.55	6	3.26
Centring	5	3.68	8	2.02*	3	2.95	5	3.51
Std	4	2.10	7	2.40	4	2.99	5	2.59

*ค่า RMSE ที่น้อยที่สุด

ตาราง 5.6 ผลที่ได้จากการใช้ตัวอย่างที่เป็นเลข 1, 5, 10, 15,..., 68 เป็นตัวสร้างโมเดล แล้ว ตัวอย่างตัวที่เหลือทั้งหมดเป็นตัวทดสอบ จากตัวอย่างทั้งหมด 68 ตัวอย่าง

Data preprocessing	200-1500 nm		575-1500 nm		860-1500 nm		200-575 nm	
	PCs	RMSE	PCs	RMSE	PCs	RMSE	PCs	RMSE
Non	2	4.30	6	3.76	4	3.42	2	4.39
Log	3	6.88	3	11.49	3	10.31	3	3.90
Norm	2	6.11	6	11.12	5	12.70	2	6.33
Centring	2	4.02	6	3.12	3	3.12*	2	4.20
Std	7	2.91*	8	4.32	6	5.70	1	5.60

*ค่า RMSE ที่น้อยที่สุด

จากตาราง 5.5 ช่วงที่เหมาะสมสำหรับการวิเคราะห์ด้วยเทคนิค PLS คือ ช่วงความยาวคลื่น 575 – 1500 nm โดยใช้ preprocessing คือ mean centring ใช้ PC = 8 และมีค่า RMSE = 2.02 และ จากตาราง 5.6 ช่วงที่เหมาะสมสำหรับการวิเคราะห์ด้วยเทคนิค PLS คือ ช่วงความยาวคลื่น 200 – 1500 nm โดยใช้ preprocessing คือ standardization ใช้ PC = 7 และมีค่า RMSE = 2.91

ตาราง 5.7 ค่า Q^2 และ R^2 ของการใช้ตัวอย่างที่เป็นเลขคู่เป็นตัวสร้างโมเดล แล้วตัวอย่างที่เป็น เลขคี่เป็นตัวทดสอบ จากตัวอย่างทั้งหมด 68 ตัวอย่าง

Data preprocessing	200-1500 nm		575-1500 nm		860-1500 nm		200-575 nm	
	Q^2	R^2	Q^2	R^2	Q^2	R^2	Q^2	R^2
Non	0.9652	0.9946	0.9692	0.9995	0.9687	0.9974	0.9723	0.9969
Log	0.9353	0.9994	0.8573	0.9993	0.8553	0.9989	0.9629	0.9595
Norm	0.9491	0.9949	0.8035	0.9502	0.6323	0.9562	0.9707	0.9897
Centring	0.9627	0.9943	0.9887*	0.9994*	0.9761	0.9906	0.9660	0.9948
Std	0.9878	0.9984	0.9841	0.9996	0.9754	0.9954	0.9815	0.9971

*ค่า Q^2 และ R^2 ที่มากที่สุด และเหมาะสมที่สุด

จากตาราง 5.7 ค่า R^2 และ Q^2 ที่มากที่สุด และเหมาะสมที่สุด อยู่ในช่วงความยาวคลื่น 575–1500 nm โดยมีค่าเท่ากับ 0.9994 และ 0.9887 ตามลำดับ data preprocessing ที่ใช้คือ mean centring จากตาราง 5.8 ค่า R^2 และ Q^2 ที่มากที่สุด และเหมาะสมที่สุด อยู่ในช่วงความยาวคลื่น 860–1500 nm โดยมีค่าเท่ากับ 0.9999 และ 0.9768 ตามลำดับ data preprocessing ที่ใช้คือ standardization

ตาราง 5.8 ค่า Q^2 และ R^2 ของการใช้ตัวอย่างที่เป็นเลข 1, 5, 10, 15,..., 68 เป็นตัวสร้างโมเดล แล้วตัวอย่างตัวที่เหลือทั้งหมดเป็นตัวทดสอบ จากตัวอย่างทั้งหมด 68 ตัวอย่าง

Data preprocessing	200-1500 nm		575-1500 nm		860-1500 nm		200-575 nm	
	Q^2	R^2	Q^2	R^2	Q^2	R^2	Q^2	R^2
Non	0.9495	0.9667	0.9614	0.9967	0.9681	0.9930	0.9473	0.9676
Log	0.8706	0.9740	0.6392	0.9465	0.7096	0.9626	0.9585	0.9678
Norm	0.8979	0.8826	0.6624	0.9784	0.5592	0.9557	0.8904	0.8788
Centring	0.9559	0.9701	0.9734	0.9994	0.9733*	0.9960*	0.9519	0.9700
Std	0.9768*	0.9999*	0.9490	1.0000	0.9111	1.0000	0.9143	0.9578

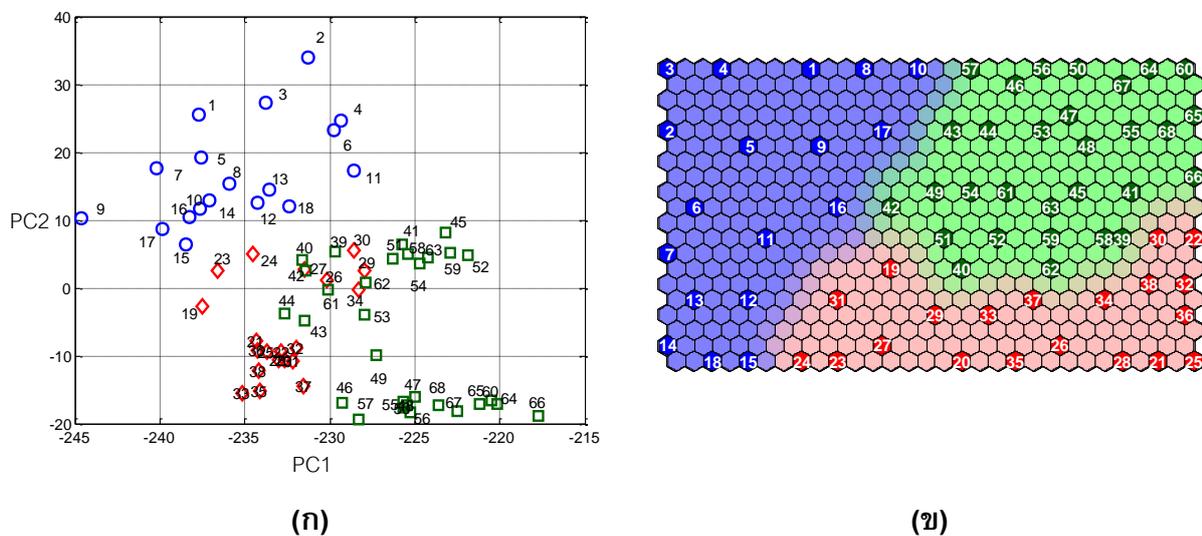
*ค่า Q^2 และ R^2 ที่มากที่สุด และเหมาะสมที่สุด

จากผลการสร้างโมเดล PLS เพื่อทำนายจำนวนครั้งของน้ำมันที่ถูกใช้ซ้ำเป็นเวลานานๆ จะเห็นว่า จำนวน PC, ค่า RMSE, R^2 และ Q^2 นั้นจะแตกต่างกันไปตามการเลือกข้อมูลในการทำโมเดล หากเลือกข้อมูลจำนวนหลายๆ ข้อมูลดังเช่นการเลือกตัวอย่างที่เป็นจำนวนคู่มาทำเป็นโมเดล จะพบว่า ให้ผลการทำนายที่ดีกว่าเลือกข้อมูลจำนวนน้อยๆ แต่ในทางปฏิบัติ การเลือกข้อมูลจำนวนมากๆ มาสร้างเป็นโมเดลนั้นอาจทำให้เสียเวลาเพราะในโรงงานอุตสาหกรรมมีการทำงานแบบต่อเนื่อง ตัวอย่างจะถูกตรวจวัดอยู่ตลอดเวลา ทำให้มีจำนวนตัวอย่างมากมายมหาศาล อีกทั้งยังทำให้เกิดความคลาดเคลื่อนได้ง่าย ดังนั้นในการสร้างโมเดลเราจึงจำเป็นต้องเลือกจำนวนตัวอย่างมาให้เหมาะสม ไม่มากและไม่น้อยจนเกินไป และจำนวน PC จะต้องมิต่ำน้อยๆ เพราะยังมีจำนวน PC มาก โอกาสที่ค่าตัวรบกวน (noise) ถูกวิเคราะห์ด้วยจะมีมาก ส่งผลให้การทำนายคลาดเคลื่อน ดังนั้นในงานวิจัยครั้งนี้จึงเลือกผลที่ได้จากการใช้ตัวอย่างจำนวนครั้งที่ 1, 5, 10, 15,..., 68 เป็นตัวสร้างโมเดล ส่วนตัวอย่างตัวที่เหลือทั้งหมดเป็นตัวทดสอบจากตัวอย่างทั้งหมด 68 ตัวอย่างของตัวอย่างน้ำมันที่ใช้ น้ำมันใหม่เป็นตัวอ้างอิงมาวิเคราะห์ผลที่ได้คือ ค่า RMSE ที่น้อยที่สุดเท่ากับ 2.91 (ตาราง 5.6) และใช้ data preprocessing คือ standardization แต่เนื่องจากอยู่ในช่วงความยาวคลื่น 200-1500 nm เป็นช่วงของ UV-Vis ซึ่งมีสัญญาณของตัวรบกวนสูง อีกทั้งการทำ standardization อาจไปทำให้พีคของตัวรบกวนมีผลต่อการวิเคราะห์ด้วย ดังนั้นจึงเลือกใช้ช่วงที่มีค่า RMSE ที่น้อยรองลงมาเท่ากับ 3.12 เท่ากันอยู่สองช่วง คือ ช่วง 575-1500 nm และ 860-1500 nm ซึ่งใช้จำนวน PC เท่ากับ 6 และ 3 ตามลำดับ จากที่กล่าวไว้ข้างต้นแล้วว่า ไม่ควรใช้จำนวน PC มากเกินไป เพราะจะทำให้ตัวรบกวนเข้ามามีผลมาก ดังนั้นจึงเลือกใช้ในช่วง 860-1500 nm ซึ่งอยู่ในช่วงของ NIR มีค่า RMSE = 3.12 มีจำนวน PC = 3 ค่า R^2 และ

Q^2 เท่ากับ 0.9960 และ 0.9733 ตามลำดับ (ตาราง 5.8) ใช้ data preprocessing คือ mean centring มาสร้างโมเดล PLS สำหรับทำนายจำนวนทอดซ้ำของน้ำมันที่ใช้ทอดมันฝรั่ง

5.4 การสร้างแบบจำลองเพื่อติดตามคุณภาพน้ำมันด้วยเทคนิค SOM

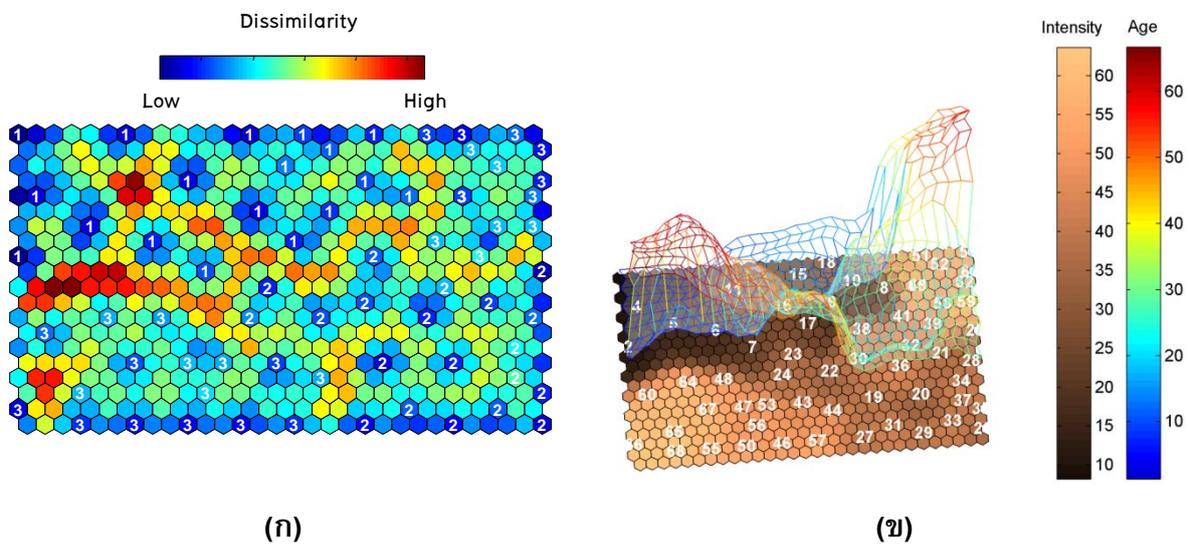
เมื่อสร้างแบบจำลองโดยใช้ข้อมูล NIR ที่ผ่านการทำ mean centring โดยใช้ map unit ที่มีขนาด 20×30 และแสดงผลด้วยเทคนิค supervised color shading ซึ่งแสดงในรูปที่ 5.9 (หมายเหตุ supervised color shading คือการแสดงผลของ map โดยระบุกลุ่มหรือ class membership ให้กับทุก map unit ขึ้นอยู่กับความคล้ายคลึงที่สุดของแต่ละ map unit กับ class membership ที่ใช้ในการสร้างแบบจำลอง) โดยรูปที่ 5.9 (ก) เป็นผลการทำ PCA ซึ่งแสดงให้เห็นว่าข้อมูลสามารถแบ่งได้เป็น 3 กลุ่มตามลำดับวันที่ของการใช้น้ำมัน ถึงแม้ว่าวันที่ 2 และ 3 ตัวอย่างน้ำมันมีลักษณะค่อนข้างหลวมซ้อนกัน แสดงถึงความคล้ายคลึงกันระหว่างตัวอย่างน้ำมันวันที่ 2 และ 3 ซึ่งเหมือนกันมากกว่าน้ำมันใหม่ที่ถูกใช้ในครั้งแรก



รูป 5.9 (ก) score plot ของข้อมูล NIR ที่ผ่านการทำ mean centring (ข) แบบจำลอง SOM ที่แสดงด้วยเทคนิค supervised color shading โดยสัญลักษณ์ที่ใช้เท่าเดียวกับสัญลักษณ์ที่แสดงในรูป 5.8

อย่างไรก็ตาม เมื่อพิจารณาผลการทำ supervised color shading ในรูป 5.9 (ข) พบว่า map ที่ใช้แสดงผลสามารถแบ่งได้เป็น 3 ส่วนอย่างชัดเจนตามตัวอย่างน้ำมันที่เก็บได้แต่ละวัน แสดงให้เห็นถึงประสิทธิภาพของเทคนิค SOM เมื่อประยุกต์ใช้กับข้อมูล NIR ของน้ำมันประกอบอาหาร ทั้งนี้จะ

สังเกตเห็นว่าตัวอย่างน้ำมันที่มีอายุการใช้งานน้อยจะอยู่ตำแหน่งมุมบนซ้ายของ map และแนวโน้มของอายุน้ำมันเพิ่มขึ้นบริเวณตรง map ส่วนน้ำมันที่มีอายุมากจะรวมกันอยู่การอยู่บริเวณด้านบนซ้าย โดยลักษณะการเปลี่ยนแปลงดังกล่าวสามารถพิจารณาได้จาก response value ซึ่งเป็นค่าแสดงอายุของน้ำมันที่ได้จากการทำนายของแต่ละ map unit ที่ถูกแสดงด้วยกราฟ contour บนผิวของกราฟ component plane ในรูปที่ 5.10 (ข)

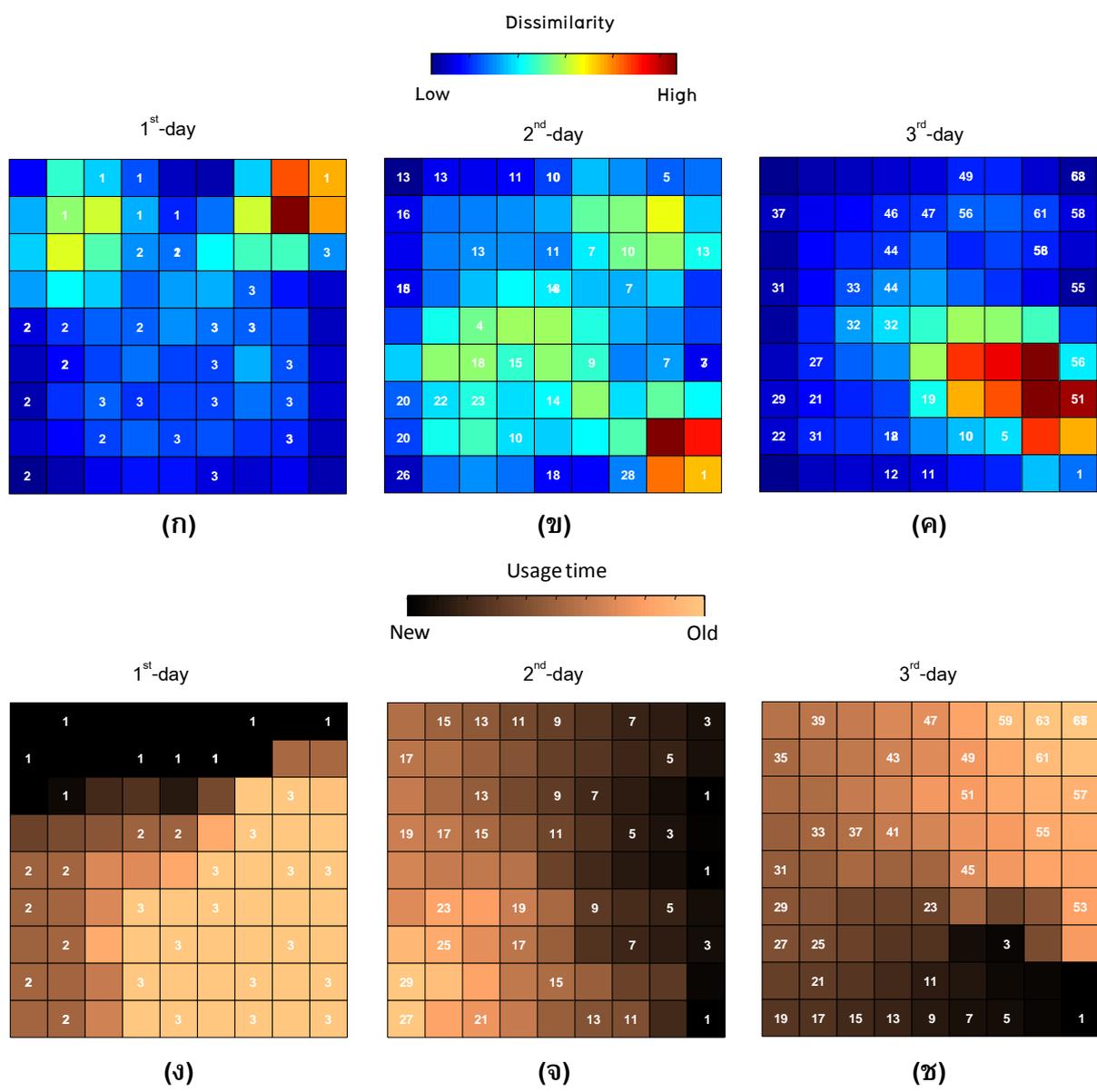


รูป 5.10 (ก) แบบจำลอง SOM ที่แสดงด้วยเทคนิค u-matrix โดยตัวเลขแสดงจำนวนวันของการใช้น้ำมันซ้ำ และ (ข) แสดง component plane ของค่าการดูดกลืนแสงในช่วงความยาวคลื่นเท่ากับ 1100 nm ส่วน response value สำหรับระยะเวลาของน้ำมันที่ได้จากการทำนายของแต่ละ map unit ถูกแสดงด้วยกราฟ contour บนผิวของกราฟ component plane

5.4.1 การสร้างแบบจำลองเพื่อติดตามคุณภาพน้ำมันด้วยเทคนิค multiple SOMs

การสร้างแบบจำลอง SOM ข้างต้นที่กล่าวมาในหัวข้อ 5.4 เป็นการสร้างแบบจำลองโดยใช้ map เพียงแผ่นเดียวในการแสดงข้อมูลน้ำมันทั้งหมด อย่างไรก็ตามมีความเป็นไปได้ในการแบ่งข้อมูลน้ำมันออกเป็น 3 กลุ่มย่อย ตามจำนวนวันที่เก็บบันทึกข้อมูล ดังนั้นจำนวน map ที่เกิดขึ้นจึงเท่ากับ 3 map สำหรับขั้นตอนการสร้างแบบจำลอง multiple SOMs การกำหนดขนาดของ map และวิธีการทำนายผลได้ถูกอธิบายใน manuscript ที่ได้ส่งตีพิมพ์ในวารสารนานาชาติชื่อวารสาร Journal of Chemometrics ซึ่งอ้างถึงในรายงานฉบับนี้

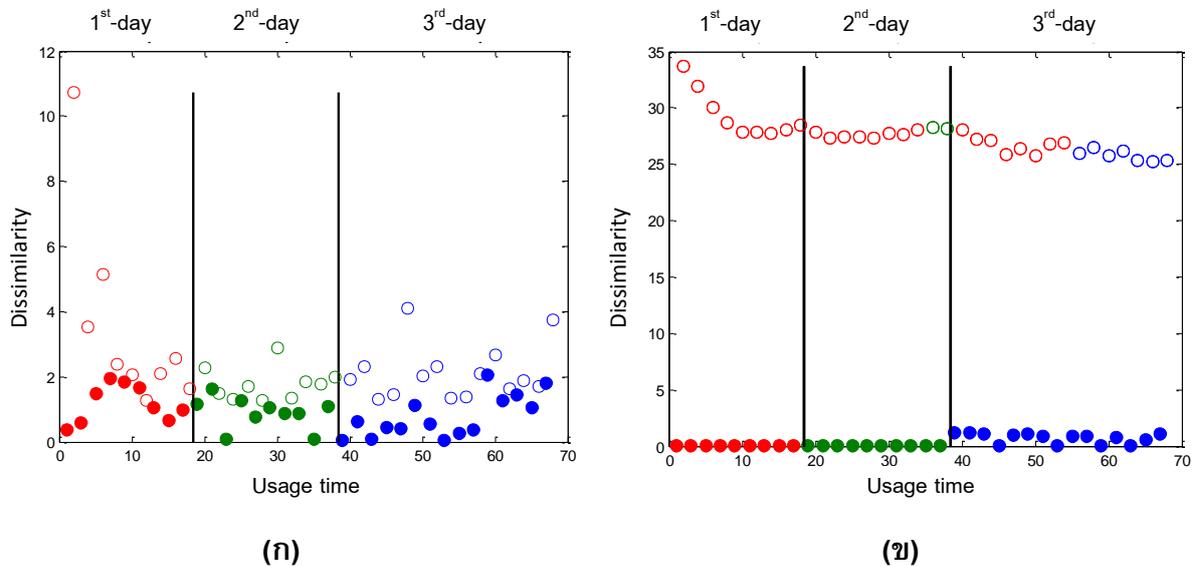
จุดเด่นของการประยุกต์ใช้ multiple SOMs ในกรณีนี้คือ แบบจำลองที่สร้างขึ้นสามารถใช้งานได้อย่างต่อเนื่อง ในกรณีในวันที่สองของกระบวนการผลิต ข้อมูล NIR ของตัวอย่างน้ำมันในวันที่หนึ่ง ยังสามารถใช้ในการติดตามคุณภาพกระบวนการผลิตในขั้นต้น หลังจากที่จำนวนตัวอย่างน้ำมันในวันที่สองมีจำนวนมากพอแบบจำลอง SOM สามารถสร้างขึ้น เพื่อใช้ในการทำนายร่วมกับ map ที่สร้างขึ้นไว้ก่อนหน้านี้ รูป 5.11 แสดงแบบจำลอง multiple SOMs ที่สร้างขึ้นจากข้อมูล NIR ของตัวอย่างน้ำมันทั้ง 3 วัน เกิดเป็น map ที่มีขนาดเล็กๆ 3 map โดย รูป 5.11 (ก) – (ค) แสดง u-matrices ของข้อมูลน้ำมัน โดย แต่ละ map ตัวอย่างที่เก็บได้แต่ละวัน ไม่แสดงถึง cluster หรือกลุ่มของข้อมูลอย่างชัดเจน เนื่องจากตัวอย่างของน้ำมันในแต่ละวันไม่ควรมีความแตกต่างกันมาก อย่างไรก็ตามเมื่อพิจารณา รูป 5.11 (ง) – (ข) ซึ่งแสดง response planes สำหรับระบุอายุของน้ำมันที่ได้จากการทำนายของแต่ละ map unit ในรูป (ก) - (ค) ตามลำดับ พบแนวโน้มการเปลี่ยนแปลงของตัวอย่างน้ำมันแต่ละวัน เช่น รูป 5.11 (จ) ตัวอย่างน้ำมันที่มีอายุการใช้งานช้าน้อยจะจัดเรียงตัวทางด้านขวา และน้ำมันที่มีอายุการใช้งานช้านานจะมีแนวโน้มเคลื่อนที่มาทางซ้ายของพื้นที่ map เป็นต้น



รูป 5.11 (ก)-(ค) แสดง u-matrices ของข้อมูลน้ำมัน โดย แต่ละ map ตัวอย่างที่เก็บได้แต่ละวัน จำนวน 3 วัน และ (ง)-(ช) แสดง response planes สำหรับระยะเวลาของน้ำมันที่ได้จากการทำนายของแต่ละ map unit ในรูป (ก)-(ค) ตามลำดับ

เมื่อแบ่งข้อมูล NIR ของตัวอย่างน้ำมันออกเป็น 2 กลุ่ม เพื่อสร้างเป็น training set และ test set จากนั้นสร้างแบบจำลองโดยใช้ training samples และจากแบบจำลอง SOM ที่สร้างขึ้นทดลองคำนวณค่า dissimilarity ของ test samples แต่ละตัว เมื่อนำผลการคำนวณที่ได้มาสร้างกราฟโดยกำหนดให้แกนนอนแสดงอายุ (จำนวนครั้งของน้ำมันที่ถูกใช้) และแกนตั้งแสดง dissimilarity ของตัวอย่างแต่ละตัว จะ

ได้กราฟที่เรียกว่า "control chart" ซึ่งสามารถใช้สังเกตพฤติกรรมของกระบวนการผลิตแสดงในรูป 5.12 รูป 5.12 แสดง control chart ที่สร้างโดยใช้ (ก) single SOM และ (ข) multiple SOMs โดยกำหนดให้สัญลักษณ์สีที่แตกต่างแทนจำนวนวันที่น้ำมันถูกใช้ วงกลมที่บัพแทน training samples ส่วนวงกลมเปิดแทน test samples



รูปที่ 5.12 control chart ที่สร้างโดยใช้ (ก) single SOM และ (ข) multiple SOMs โดยสัญลักษณ์สีที่ใช้ดังแสดงในรูปที่ 5.4 และวงกลมที่บัพแทน training samples ส่วนวงกลมเปิดแทน test samples

จาก control chart ที่ได้พบว่า ค่า dissimilarity ที่ได้จากการคำนวณด้วยเทคนิค multiple SOMs มีค่ามากกว่าผลการคำนวณด้วยเทคนิค single SOM อย่างไรก็ตาม control chart ที่ได้แสดงความแตกต่างระหว่าง training และ test samples อย่างชัดเจน และเมื่อพิจารณาผลการทำนายตามจำนวนวันที่ที่ผ่านการใช้งาน พบว่าการใช้ multiple SOMs ให้ผลการทำนายของตัวอย่างในวันที่สองมีคุณภาพคล้ายคลึงกับน้ำมันวันที่หนึ่ง ถึงแม้ว่าจะมีสองตัวอย่างสุดท้ายของวันที่สองถูกระบุว่าเป็นน้ำมันจากวันที่สอง ส่วนน้ำมันในวันที่สามในช่วงครึ่งแรกจะถูกทำนายให้อยู่ในกลุ่มที่มีคุณภาพใกล้เคียงกับน้ำมันวันแรก และหลังจากนั้นจึงถูกทำนายว่าน้ำมันมีสภาพที่แตกต่างจากวันที่หนึ่งอย่างชัดเจน ถึงแม้ว่าค่า dissimilarity ที่ได้มีความแตกต่างไม่ชัดเจนมากนัก

ดังนั้น การใช้ multiple SOMs อาจมีปัญหาเกี่ยวกับ overfitting คือ แต่ละ map จะถูกเรียนรู้จากตัวอย่างกลุ่มใดกลุ่มหนึ่งเพียงเท่านั้น ทำให้ค่า dissimilarity ที่คำนวณได้สูงกว่าการรวมตัวอย่างทุกตัวลงบน map เพียงอันเดียว (single SOM) และน้ำมันจะถูกจัดกลุ่มตาม class membership หรือจำนวน

วันที่กำหนดให้ไว้เท่านั้น แต่ถ้าพิจารณาถึงการทำนายเชิงคุณภาพ ในกรณีนี้คือการระบุความสดของน้ำมัน และการบ่งบอกจุดที่น้ำมันควรถูกเปลี่ยน การใช้ multiple SOM จะให้ผลการทำนายที่ถูกต้องมากกว่า

6. สรุปและวิจารณ์ผลการทดลอง และข้อเสนอแนะสำหรับงานวิจัยในอนาคต

สรุปผลการวิจัย

NIR มักถูกใช้อย่างแพร่หลายในงานวิจัยการศึกษาและวิเคราะห์คุณภาพของน้ำมันหรืออาหาร โดยทั่วไป เนื่องจากเป็นวิธีที่วิเคราะห์ได้ง่าย สะดวก รวดเร็ว อีกทั้งยังไม่ทำลายสารตัวอย่าง ทำให้สามารถนำกลับมาวิเคราะห์ใหม่ได้อีกครั้ง แต่เนื่องจากการวิเคราะห์ด้วย NIR ช่วงแสงของ near infrared จะตรวจสอบกับหมู่ฟังก์ชันของสารเคมีที่เป็นองค์ประกอบภายในของสารตัวอย่างได้หลากหลายชนิดในความยาวคลื่นเดียว ทำให้สเปกตรัมที่ได้มีความซับซ้อนมาก จึงมีการนำเอาเทคนิคทางเคโมเมทริกซ์มาประยุกต์ใช้โดยใช้เทคนิค principal component analysis (PCA) เพื่อช่วยในการวิเคราะห์ข้อมูลที่มีความซับซ้อนมาก ทำให้ได้ผลที่ง่ายต่อการวิเคราะห์มากขึ้น และเทคนิค partial least squares (PLS) เพื่อช่วยในการทำนายจำนวนครั้งของน้ำมันที่ถูกใช้ซ้ำเป็นเวลานาน จากผลการดำเนินการสามารถสรุปได้ดังต่อไปนี้

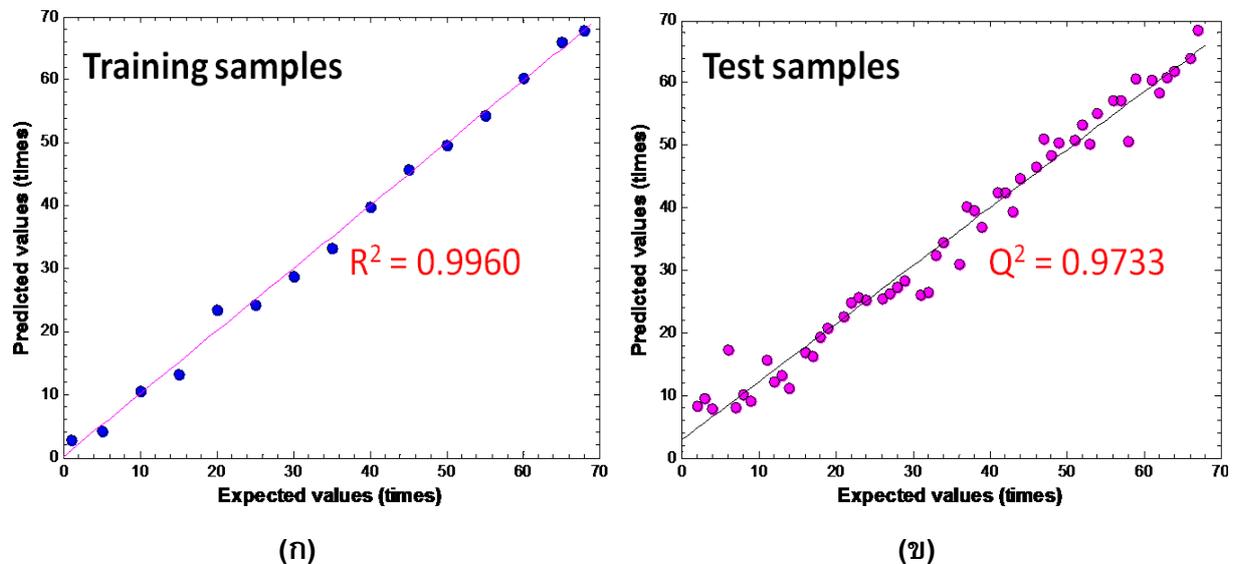
6.1 การวิเคราะห์ด้วยเทคนิค PCA และ log scaling

จากการศึกษาการเปลี่ยนแปลงของลักษณะข้อมูลน้ำมันที่ใช้ในการประกอบอาหารที่ได้จากการวิเคราะห์ด้วยเทคนิค NIR spectroscopy โดยใช้เทคนิค PCA เข้ามาช่วย และการศึกษาการสร้างโมเดล PLS เพื่อใช้ทำนายจำนวนครั้งของการใช้น้ำมันทอดซ้ำด้วยเทคนิค PLS จากผล PCA ของตัวอย่างน้ำมันที่ผ่านการใช้งานแล้ว เมื่อพิจารณาจากการทำ log scaling จะสังเกตเห็นว่า น้ำมันเมื่อผ่านการใช้งานไปเรื่อยๆ สารเคมีที่เป็นองค์ประกอบในน้ำมัน เช่น ปริมาณความชื้น ปริมาณสารโพลาร์ หรือ ปริมาณ FFA อาจจะมีการเปลี่ยนแปลงไป ส่งผลให้รูปแบบ (pattern) ของข้อมูลมีการเปลี่ยนแปลงไปเรื่อยๆ ข้อมูลน้ำมันที่มีองค์ประกอบภายในคล้ายคลึงกัน หรือไม่ต่างกันมากจะแสดงค่า score ที่ใกล้เคียงกัน ส่วนข้อมูลน้ำมันที่มีองค์ประกอบภายในแตกต่างกันมาก ก็จะมีแสดงค่า score ที่อยู่ไกลกัน ดังแสดงในรูป 5.8

6.2 การวิเคราะห์ด้วยเทคนิค PLS

ในส่วนของการวิเคราะห์ด้วย PLS ผลที่ได้จากการสร้างโมเดลพบว่า โมเดล PLS ที่มีประสิทธิภาพการทำนายที่ดีที่สุด สร้างการใช้ตัวอย่างจำนวนครั้งที่ 1, 5, 10, 15, ..., 68 เป็นตัวสร้างโมเดล แล้วตัวอย่างตัวที่เหลือทั้งหมดเป็นตัวทดสอบ จากตัวอย่างทั้งหมด 68 ตัวอย่าง ได้ค่า R^2 และ Q^2 ที่มากที่สุด และเหมาะสมที่สุด อยู่ในช่วงความยาวคลื่น 860–1500 nm โดยมีค่าเท่ากับ 0.9960 และ

0.9733 ตามลำดับ data preprocessing ที่ใช้คือ mean centring ซึ่งสามารถแสดงกราฟ correlation ได้ ดังรูป 6.2



รูปที่ 6.2 กราฟ correlation ระหว่างค่า ที่ได้จากการทดลอง และค่าที่ได้จากการทำนาย (ก) กราฟ correlation ของข้อมูล training samples (ข) กราฟ correlation ของข้อมูล test samples

6.3 การสร้างแบบจำลองด้วยเทคนิค multiple SOMs

การพัฒนาโมเดลทางสถิติด้วยเทคนิคเคโมเมตริกซ์แบบปัญญาประดิษฐ์ที่เรียกว่า multiple self organizing map (SOM) ซึ่งสามารถสร้างโมเดลที่มีลักษณะการทำงานคล้ายแผนที่ โดยแผนที่สำหรับการทำนายที่สร้างขึ้นใหม่ดังกล่าวจะถูกใช้เพื่อวิเคราะห์ข้อมูลที่ได้จากระบบหลังจากที่มีการเปลี่ยนแปลงเกิดขึ้น ทั้งนี้แผนที่ที่สร้างขึ้นใหม่นี้จะถูกใช้ร่วมกับแผนที่การทำนายหลักที่ถูกใช้ก่อนหน้านี้ในการทำนายคุณภาพของน้ำมันพืชสำหรับปรุงอาหารในกระบวนการผลิต โดยโมเดลที่พัฒนาขึ้นมีจุดเด่นคือสามารถทำงานได้อย่างต่อเนื่องและให้ผลการวิเคราะห์ของคุณภาพน้ำมันทันที ถึงแม้ว่าจะมีการเปลี่ยนแปลงขึ้นในระบบของกระบวนการผลิต

7. ภาคผนวก

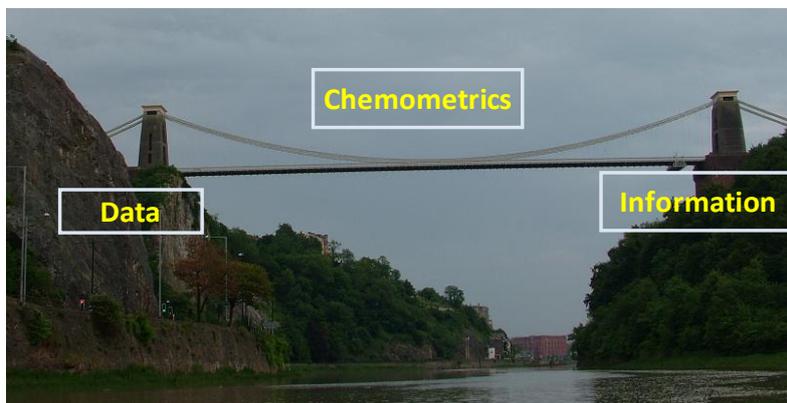
เอกสารและงานวิจัยที่เกี่ยวข้อง

7.1 เคโมเมทริกซ์ (Chemometrics)

คำว่า เคโมเมทริกซ์ (Chemometrics) อาจเป็นคำศัพท์ใหม่สำหรับนักวิจัยหลายๆ ท่าน แต่แท้ที่จริงแล้วคำว่าเคโมเมทริกซ์ถูกบัญญัติและใช้มาตั้งแต่ในช่วงต้นศตวรรษที่ 80 แต่อย่างไรก็ตามการประยุกต์ใช้ของเคโมเมทริกซ์สำหรับงานวิจัยในอดีตนั้นยังไม่เป็นที่แพร่หลายนัก อาจเป็นเพราะว่างานวิจัยในอดีตมักเกี่ยวข้องกับข้อมูลที่ไม่ค่อยมีความซับซ้อน เนื่องจากเทคโนโลยีของอุปกรณ์ที่ใช้ในการทำวิจัยในอดีตนั้นยังไม่ก้าวหน้าเมื่อเปรียบเทียบกับในปัจจุบัน หรืออาจเป็นเพราะนักวิจัยส่วนใหญ่เข้าใจว่าเคโมเมทริกซ์เป็นสิ่งที่ซับซ้อน น่ากลัว ยากต่อการเข้าใจ และยากที่นำมาประยุกต์ใช้ ผู้เขียนยอมรับว่ามีบางส่วนหรือบางแง่มุมของเคโมเมทริกซ์ที่มีความซับซ้อนและอาจต้องใช้เวลาบ้างในการทำความเข้าใจ แต่ในความเป็นจริงแล้วมีหลายสิ่งหลายอย่างที่เรอาจไม่จำเป็นต้องรู้หรือเข้าใจในหลักการบางอย่างละเอียดเพื่อที่จะใช้ประโยชน์จากมัน เช่น ในการขับรถยนต์ เรอาจไม่จำเป็นต้องทราบว่าการเครื่องยนต์และเบรกมีหลักการทำงานอย่างไร เครื่องยนต์ของรถของเรามีจำนวนกระบอกสูบกี่กระบอกและมีปริมาตรรวมเป็นกี่ซีซี เป็นกระบอกสูบแบบนอนหรือเรียงขวาง แต่สิ่งที่เราจำเป็นต้องทราบคือ การเข้าใจว่ารถยนต์นั้นเอาไว้ใช้ทำอะไร รถกระบะจะสามารถบรรทุกของได้มากกว่ารถเก๋ง แต่รถเก๋งจะนั่งสบายกว่ารถกระบะ ถ้าจะชนต้องเบรก ถ้าจะเร่งก็ต้องเหยียบคันเร่ง เวลาเลี้ยวก็ต้องหมุนพวงมาลัย เป็นต้น องค์ความรู้พื้นฐานเหล่านี้ก็เพียงพอแล้วสำหรับการขับรถและใช้ประโยชน์จากรถยนต์อย่างปลอดภัย (ทั้งนี้ก็ต้องยอมรับว่าองค์ความรู้ในขั้นสูงขึ้นไปก็มีประโยชน์ แต่มันอาจจะไม่มีความจำเป็นในขั้นต้น) เช่นเดียวกันสำหรับการประยุกต์ใช้เคโมเมทริกซ์ในงานวิจัย ขอให้ผู้อ่านพึงระลึกตัวอย่างที่กล่าวไว้ข้างต้นเพื่อความสบายใจ

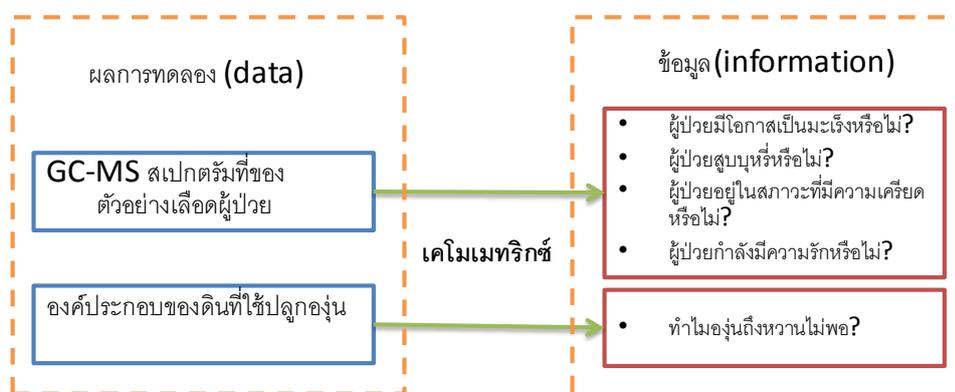
เคโมเมทริกซ์เป็นการประยุกต์ใช้ความรู้ทางคณิตศาสตร์ สถิติ และคอมพิวเตอร์ เพื่อช่วยประมวลผลที่ได้จากการวิเคราะห์ หรือหาข้อสรุปจากผลการทดลองทางเคมี ในความเป็นจริงแล้วการประยุกต์ใช้ของเคโมเมทริกซ์นั้นไม่ได้ถูกจำกัดเฉพาะในทางเคมีเท่านั้น เทคนิคต่างๆ ในทางเคโมเมทริกซ์ยังสามารถประยุกต์ใช้งานได้ในศาสตร์แขนงต่างๆ และในบางครั้งถูกเรียกโดยใช้ชื่อที่แตกต่างออกไป เช่น Econometrics สำหรับการประยุกต์ใช้ในสาขาเศรษฐศาสตร์, Machine Learning สำหรับการประยุกต์ในกลุ่มทางวิศวกรรมศาสตร์, Bioinformatics สำหรับการประยุกต์ใช้ในกลุ่มทางสิ่งมีชีวิต เป็นต้น ทั้งนี้ถึงแม้ว่าตัวเทคนิคต่างๆ ที่ถูกประยุกต์ใช้อย่างหลากหลายในศาสตร์ที่แตกต่างกัน แต่นักวิจัยที่ศึกษาเคโมเมทริกซ์จะสามารถเข้าใจหลักการทำงานของเทคนิคที่ถูกประยุกต์ใช้ดังกล่าวซึ่งมี

ความคล้ายคลึงกันได้โดยไม่ยากเย็นนัก ทั้งนี้เนื่องจากความแตกต่างนั้นอาจเป็นเพียงชนิดและที่มาของข้อมูลที่ถูกใช้ในการวิเคราะห์รวมไปถึงมุมมองของนักวิจัยที่มองถึงตัวปัญหา



รูป 7.1 แบบจำลองการทำงานของเคโมเมทริกซ์

เคโมเมทริกซ์เปรียบได้ดังสะพานที่ใช้เชื่อมระหว่างผลการทดลอง (data) และข้อมูลที่เราต้องการหรือคาดว่าจะได้รับจากการทดลอง (information) ทั้งนี้ความใกล้ชิดกัน (ความสัมพันธ์) ระหว่าง data กับ information นั้นขึ้นอยู่กับข้อมูลและโจทย์ที่นักวิจัยกำลังศึกษา ในกรณีที่โจทย์และข้อมูลไม่ซับซ้อนเราอาจใช้เพียงสะพานเชือกสั้นๆ ในการเชื่อมระหว่าง 2 สิ่งนี้ หรืออาจจะใช้วิธีการโดดข้ามแทนการใช้สะพานก็ได้ ในกรณีนี้เคโมเมทริกซ์อาจไม่มีความจำเป็นมากนัก แต่หากโจทย์นั้นเกี่ยวข้องกับ data ที่มีความซับซ้อน ต้องใช้สะพานที่มีขนาดใหญ่และแข็งแรง การนำเคโมเมทริกซ์มาประยุกต์ใช้จึงมีความจำเป็น



รูปที่ 7.2 ตัวอย่างการประยุกต์ใช้เคโมเมทริกซ์ในการแปรผลการทดลอง

รูปที่ 7.2 แสดงตัวอย่างของผลของการทดลองและข้อมูลที่ได้จากการประมวลผลที่สามารถเชื่อมต่อกันได้โดยใช้เคโมเมทริกซ์ ในกรณีแรกตัวอย่างเลือดของผู้ป่วยถูกใช้สำหรับการวิเคราะห์เพื่อ

วินิจฉัยข้อมูลบางประการเกี่ยวกับตัวผู้ป่วย เช่น ความเสี่ยงในการเป็นมะเร็ง, การสูบบุหรี่ และ ความเครียดของผู้ป่วย เป็นต้น ในกรณีนี้ข้อมูลเลือดที่บันทึกได้จัดเป็นข้อมูลผลการทดลอง (data) และ ข้อมูลที่ได้จากการวินิจฉัยจัดเป็นข้อมูลที่เราต้องการได้รับจากการทดลอง (information) สำหรับในกรณี ที่ 2 เป็นการรวบรวมข้อมูลองค์ประกอบของดินที่ใช้ในการปลูกองุ่น เช่น ปริมาณของธาตุ N, P, K, Cu, Zn และอินทรีย์วัตถุในดิน เพื่อใช้ในการวิเคราะห์ว่าอะไรเป็นปัจจัยที่ทำให้องุ่นมีความหวานน้อย จาก ตัวอย่างนี้ องค์ประกอบแต่ละชนิดของดินที่บันทึกได้จะถูกใช้เป็นข้อมูลผลการทดลองหรือ data และ หลังจากการประมวลผลพบว่า ดินมีปริมาณธาตุ Cu มากเกินไปและมีปริมาณธาตุ Zn ที่น้อยเกินไป คือ สาเหตุที่ทำให้องุ่นหวานไม่พอ ซึ่งองค์ประกอบหรือปัจจัยที่เป็นเหตุให้องุ่นมีความหวานน้อยนี้เองคือ ข้อมูลหรือ information ที่เราต้องการทราบ ดังนั้นการบรรลุผลในการประยุกต์ใช้ของเคโมเมทริกซ์นั้น นอกจากจะต้องรู้จักการเลือกประยุกต์ใช้เทคนิคเคโมเมทริกซ์ที่เหมาะสมกับข้อมูลแล้ว นักวิจัยยังต้อง อาศัยความร่วมมือจากนักวิจัยผู้มีประสบการณ์ในสาขานั้นๆ เพื่อเพิ่มศักยภาพในการประมวลผลด้วย

7.2 Univariate data analysis และ Multivariate data analysis

เคมีวิเคราะห์ (analytical chemistry) อาจจำแนกได้เป็น 2 ประเภท คือ การวิเคราะห์เชิง คุณภาพ (qualitative analysis) และการวิเคราะห์เชิงปริมาณ (quantitative analysis) การวิเคราะห์เชิง คุณภาพเป็นการวิเคราะห์เพื่อหาคำตอบว่า “มีสารอะไรบ้างในตัวอย่างที่เราต้องการวิเคราะห์” เช่น การ วิเคราะห์ว่ามียาฆ่าแมลงชนิดใดตกค้างในผลิตภัณฑ์ทางการเกษตรบ้าง หรือการวิเคราะห์ว่ามีโลหะหนัก ชนิดใดเจือปนในเนื้อหอยแครงบ้าง ส่วนการวิเคราะห์เชิงปริมาณเป็นการวิเคราะห์เพื่อหาคำตอบว่า “สารที่ เราต้องการวิเคราะห์มีปริมาณเท่าไร” เช่น การวิเคราะห์ว่ามีปริมาณของยาฆ่าแมลงในกลุ่มฟอสเฟต ตกค้างอยู่ในผลผลิตทางการเกษตรปริมาณเท่าไร หรือการวิเคราะห์ว่ามีโลหะปรอท (Hg) เจือปนใน เนื้อหอยแครงปริมาณเท่าไร เป็นต้น สำหรับเคโมเมทริกซ์นั้น สามารถประยุกต์ใช้ได้ทั้งการวิเคราะห์ เชิงคุณภาพและเชิงปริมาณ

ลักษณะการประยุกต์ใช้ของเคโมเมทริกซ์ อาศัยลักษณะของข้อมูลที่ถูกใช้ในการวิเคราะห์ โดย แบ่งเป็น การวิเคราะห์ข้อมูลที่มีเพียง 1 ตัวแปร (univariate data analysis) และการวิเคราะห์ข้อมูลที่มี มากกว่า 1 ตัวแปร (multivariate data analysis) คำว่า “ตัวแปร” ในที่นี้หมายถึง ลักษณะของข้อมูลที่เรา บันทึกได้จากตัวอย่างในระหว่างการทดลอง ยกตัวอย่างเช่น การวิเคราะห์ความอ้วนของชายไทย ใน กรณีนี้ตัวแปรที่ถูกบันทึกและจะถูกนำไปใช้ในการวิเคราะห์ คือ น้ำหนักของชายไทยแต่ละคนที่เข้าร่วม ในการทดลอง จากตัวอย่างนี้ ข้อมูลผลการทดลอง (data) คือ น้ำหนักที่บันทึกได้ ส่วนข้อมูลที่เรา ต้องการทราบจากการวิเคราะห์ (information) คือ คำตอบว่าชายคนนั้นอ้วนหรือไม่ การวิเคราะห์ข้อมูล ในเบื้องต้นอาจทำได้โดยกำหนดเกณฑ์ว่า “ชายไทยคนใดมีน้ำหนักมากกว่า 80 กิโลกรัม แสดงว่า

ชายไทยคนนั้นอ้วน” อย่างไรก็ตาม การวิเคราะห์ข้อมูลโดยใช้เกณฑ์ข้างต้นเพียงอย่างเดียวอาจไม่เหมาะสมนัก เช่น ในกรณีที่มีชายไทยที่ตัวเตี้ยมากๆ แต่หนักเพียง 70 กิโลกรัมซึ่งไม่ถึง 80 กิโลกรัม ในกรณีนี้ถ้าพิจารณาจากลักษณะทางกายภาพภายนอก จะสรุปได้ทันทีเลยว่าเขาอ้วน (พิจารณาโดยใช้รูป 3 ประกอบ) แต่เนื่องจากเกณฑ์ที่เราได้กำหนดไว้ในตอนแรก เขาจะไม่ถูกจัดอยู่ในกลุ่มคนอ้วน ดังนั้นจากตัวอย่างนี้จะพบว่าการวิเคราะห์ข้อมูลโดยใช้ตัวแปรน้ำหนักเพียงตัวแปรเดียวอาจไม่เหมาะสมที่จะสรุปว่าเขาอ้วนหรือไม่ สิ่งนี้จุดด้อยของการวิเคราะห์โดยใช้เพียงตัวแปรเดียว ถึงแม้การวิเคราะห์โดยใช้เพียงตัวแปรเดียวนั้นจะไม่ซับซ้อน ง่ายต่อการเข้าใจและการประยุกต์ใช้งาน แต่เคโมเมทริกซ์ อาจจะไม่ค่อยมีบทบาทมากนักสำหรับการวิเคราะห์ในลักษณะนี้



รูป 7.3 ตัวอย่างข้อด้อยที่เกิดจากการวิเคราะห์โดยใช้เพียงตัวแปรเดียว

จากตัวอย่างข้างต้น เราพบว่าการใช้ข้อมูลที่มีเพียงตัวแปรเดียวอาจให้ข้อสรุปที่ไม่ถูกต้องนัก แนวทางหนึ่งที่เป็นไปได้ในการแก้ปัญหาคือ การบันทึกข้อมูลเพิ่มหรือเพิ่มจำนวนตัวแปรที่ใช้ในการวิเคราะห์ เช่น การวัดส่วนสูงและนำมาใช้ประกอบสำหรับการพิจารณา โดยเราอาจใช้สัดส่วนของส่วนสูง ต่อน้ำหนักในการวิเคราะห์ แทนการใช้เพียงข้อมูลน้ำหนัก เช่นระบุว่า “ชายไทยคนใดมีสัดส่วนของความสูงต่อน้ำหนักต่ำกว่า 2.66 ชายไทยคนนั้นจะถูกจัดอยู่ในกลุ่มคนอ้วน” นี่คือนัยของการวิเคราะห์ที่มีการใช้ตัวแปรมากกว่า 1 หรือการวิเคราะห์ multivariate data analysis ที่มี 2 ตัวแปร ซึ่งโดยปรกติแล้ว การวัดความอ้วนเพื่อประเมินหาไขมันส่วนเกินในร่างกาย เพื่อคำนวณความเสี่ยงในการเป็นโรค อาจคำนวณได้โดยใช้ค่า Body Mass Index (BMI) หรือ ดัชนีมวลกาย ซึ่งเท่ากับ น้ำหนักตัว / ความสูง ยกกำลังสอง

Multivariate data analysis มีบทบาทในชีวิตประจำวันของเราอย่างมากโดยที่เรามักจะไม่รู้ตัว แต่โดยมากแล้วมักจะเกี่ยวข้องกับการวิเคราะห์ที่ใช้ข้อมูล 2-3 ตัวแปรพร้อมๆ กัน เช่น เวลาเราจะเลือกซื้อมะม่วงเราอาจใช้เงื่อนไข ดังนี้ 1. ลูกมีขนาดใหญ่ 2. ผิวไม่ขำ และ/หรือ 3. ไม่สุกเกินไป เป็นต้น ทั้งนี้

เนื่องจากมนุษย์ส่วนใหญ่เคยชินกับการรับรู้และวิเคราะห์กับข้อมูลได้มากที่สุดเพียง 3 มิติ บางท่านอาจพิจารณาได้มากกว่า 3 มิติพร้อมๆ กันแต่ก็เป็นส่วนน้อย แต่ในความเป็นจริงแล้วโจทย์ของงานวิจัยหลายๆ อย่างมีความจำเป็นที่จะต้องใช้อ้อมมูลจำนวนมากเพื่อช่วยประกอบในการตัดสินใจ เช่น ในกรณีที่ต้องการวิเคราะห์หาสาเหตุที่ทำให้ผลผลิตของขนาดความหวาน ปัจจัยหรือตัวแปรใช้ในการศึกษาอาจมีถึงมากถึง 32 ตัวแปร ดังตาราง 7.1

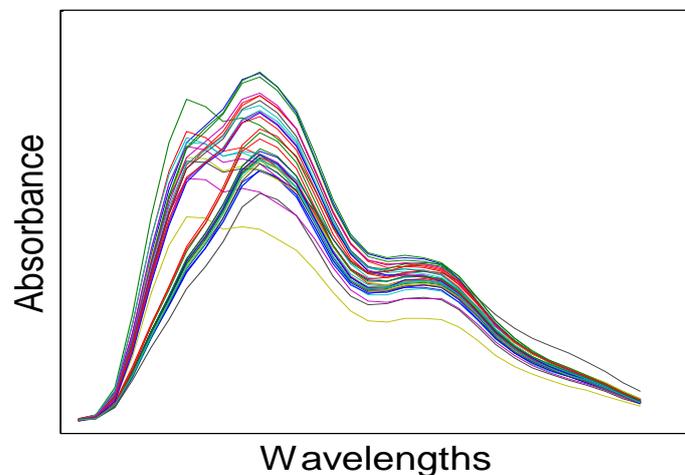
ตาราง 7.1 ข้อมูลที่ได้จากการวิเคราะห์องค์ประกอบจำนวน 32 ตัวแปรของตัวอย่างดินจำนวน 10 ตัวอย่าง

NO.	ตัวแปร (parameters)	หน่วย	ตัวแปร										
			Sp1	Sp2	Sp3	Sp4	Sp5	Sp6	Sp7	Sp8	Sp9	Sp10	Sp11
1	pH-1:1	water	5.4	5.8	5.0	6.2	6.2	6.3	6.3	5.9	6.0	6.3	5.8
2		KCl	4.6	4.5	4.3	5.0	5.3	5.3	4.6	4.4	5.0	4.6	4.5
3	CEC	(cmol/kg)	8.47	5.64	7.84	6.59	12.23	11.92	9.72	10.35	13.49	9.72	12.54
4	Organic Matter	(%)	3.18	4.59	4.70	2.97	5.55	2.81	7.51	5.71	5.66	3.99	5.82
5	Extractable Phosphorus, P	(mg/kg)	220.0	10.0	594.0	20.0	125.0	155.0	12.0	27.0	594.0	120.0	9.0
6	Extractable Potassium, K	(mg/kg)	52	52	325	135	210	726	265	96	462	59	89
7	Extractable Manganese, Mn	(mg/kg)	89	7	71	115	11	27	32	11	14	8	20
8	Extractable Copper, Cu	(mg/kg)	1.24	0.39	1.6	0.68	1.49	0.6	0.16	0.99	5.98	0.48	0.27
9	Extractable Zinc, Zn	(mg/kg)	2.02	0.23	0.99	0.6	1.25	1.77	0.32	0.15	4.53	0.3	0.39
10	Extractable Iron, Fe	(mg/kg)	41	42	37	25	54	33	27	25	95	53	50
11	Fe ion	(%)	1.74	2.84	2.54	3.31	2.12	1.78	1.10	1.88	0.97	0.20	2.03
12	Al ion	(%)	1.05	0.89	1.10	0.00	0.00	0.00	0.00	1.53	0.00	0.00	0.75
13	Exchangeable Potassium, K	(cmol/kg)	0.15	0.18	0.78	0.32	0.42	1.48	0.63	0.26	0.10	0.20	0.28
14	Exchangeable Calcium, Ca	(cmol/kg)	2.87	1.26	1.11	1.46	3.56	6.39	1.75	5.96	1.32	3.56	3.60
15	Exchangeable Magnesium, Mg	(cmol/kg)	0.25	0.16	0.27	0.53	0.47	1.00	0.49	0.31	1.22	0.36	0.35
16	Exchangeable Sodium, Na	(cmol/kg)	0.1	0.09	0.05	0.08	0.05	0.07	0.03	0.03	0.14	0.07	0.04
17	Extractable Boron, B	(mg/kg)	0.13	0.09	0.18	0.07	0.58	0.71	0.53	0.17	1.7	0.44	0.28
18	Particle size distribution	% sand	45.86	20.13	42.50	61.92	40.23	56.01	45.69	45.86	61.15	70.86	32.91
19		% silt	19.27	22.36	12.78	7.56	22.86	33.39	25.56	16.76	20.23	19.47	25.56
20		% clay	34.87	57.50	44.73	30.52	36.91	10.60	28.75	37.38	18.62	9.67	41.53
21	Bulk Density, Db	(g/cm ³)	1.03	1.24	1.18	0	1.14	1.45	1.13	1.06	1.02	1.35	0.97
22	Particle Density, Ds	(g/cm ³)	2.40	2.48	2.45	2.45	2.49	2.49	2.43	2.48	2.47	2.50	2.49
23	Porosity, E	(%)	57.2	50.0	51.6	0.0	54.3	41.5	53.7	57.3	58.7	45.9	60.9
24	Soil moisture (% by weight)	FC	33.6	33.9	29.9	34.1	30.7	25.8	33.4	26.4	25.4	20.0	31.4
25		PWP	18.4	18.1	16.5	14.8	18.8	11.9	12.8	12.3	10.1	6.3	17.4
26		AWCA	15.2	15.8	13.4	19.3	11.9	13.9	20.6	14.1	15.4	13.8	14.0
27	Hydraulic Conductivity, K sat	(m/day)	0.232	2.208	10.122	0.000	4.899	9.299	0.323	4.416	2.123	0.004	4.050
28	Water stable aggregates (%)	Macro Agg.	80.41	84.59	76.35	53.75	89.43	82.53	83.53	82.06	79.96	67.62	90.49
29		Micro Agg.	19.59	15.41	23.65	46.25	10.57	17.47	16.47	17.94	20.04	32.38	9.51
30	Soil aggregate stability (mm)	MWD (dry)	2.15	3.09	2.13	3.03	2.99	2.72	2.09	1.64	3.28	4.30	2.50
31		MWD(wet)	1.23	1.47	0.94	0.68	1.90	1.10	1.72	1.46	1.55	0.79	2.43
32		CMWD	0.91	1.61	1.19	2.35	1.09	1.62	0.37	0.18	1.74	3.51	0.07

ทั้งนี้อาจนำปัจจัยอื่นๆ ที่คาดว่าจะเกี่ยวข้อง มาใช้ในการวิเคราะห์ด้วย เช่น ค่าความเป็นกรด เบส ค่าความชื้นในอากาศ ปริมาณน้ำฝนและปริมาณแสงอาทิตย์ ดังนั้นเราจะพบว่าเทคนิคทางเคโมเมทริกซ์สามารถช่วยให้การวิเคราะห์นั้นทำได้ง่ายขึ้น รวมไปถึงการหาเหตุผลมาช่วยยืนยันผลการวิเคราะห์

ตัวอย่างการประยุกต์ใช้เคโมเมทริกซ์ในการวิจัย

รูป 7.4 เป็นกราฟแสดงค่าการดูดกลืนแสงของสารตัวอย่างจำนวน 39 ตัวอย่าง หากเราพิจารณาโดยใช้เฉพาะข้อมูลที่ให้มา เราจะไม่สามารถทราบว่าคุณค่าการดูดกลืนแสงเหล่านี้มาจากตัวอย่างอะไรบ้าง อย่างไรก็ตามสิ่งที่เราสังเกตเห็นได้ คือ บางสเปกตรัมมีลักษณะที่คล้ายคลึงกัน ในขณะที่บางสเปกตรัมมีลักษณะที่แตกต่างกัน ทั้งนี้ความเหมือนและความแตกต่างของข้อมูลจะแสดงถึงรูปแบบหรือ pattern ของข้อมูลนั่นเอง คำถามต่อมาที่น่าสนใจ คือ จากตัวอย่างทั้งหมด 39 ตัวอย่างนี้จะถูกจำแนกได้เป็นกี่กลุ่ม ในกรณีนี้คำว่ากลุ่ม หมายถึงเซตของตัวอย่างที่มีลักษณะคล้ายคลึงกัน แต่จากภาพที่นำมาแสดง การพิจารณาโดยใช้ข้อมูลดิบหรือสเปกตรัมดังกล่าวอาจเป็นการยากที่จะตอบคำถามนี้

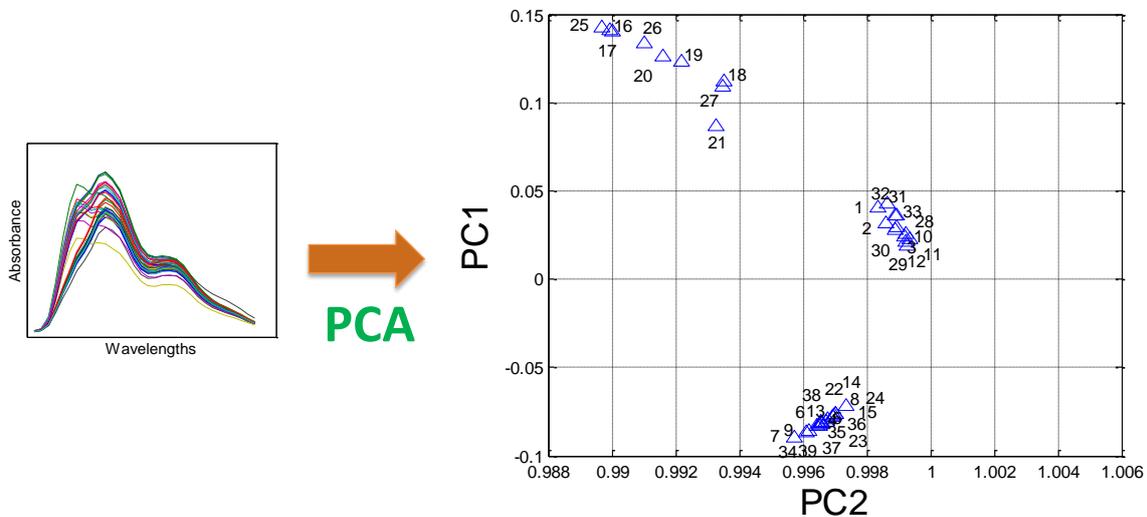


รูปที่ 7.2 UV-Vis สเปกตรัมของสารตัวอย่างจำนวน 39 ตัวอย่าง

ในกรณีนี้ เราจะใช้เทคนิคเคโมเมทริกซ์ที่เรียกว่า Principal Component Analysis (PCA) มาช่วยแก้ปัญหา เทคนิค PCA เป็นการวิเคราะห์เชิงพหุตัวแปรที่สามารถใช้ในการลดจำนวนตัวแปรสำหรับการวิเคราะห์ โดยมีวัตถุประสงค์เพื่อทำให้ข้อมูลที่มีจำนวนมากและมีความซับซ้อนนั้นอยู่ในรูปที่สะดวกต่อการวิเคราะห์ โดยตัวแปรที่สร้างขึ้นมานั้น ยังคงสามารถใช้เก็บข้อมูล หรือ variation ที่มีความสำคัญไว้ได้เกือบทั้งหมด ยกตัวอย่างในกรณีข้างต้น ถ้ามีค่าการดูดกลืนแสงที่ถูกใช้ในการวิเคราะห์ทั้งหมด 24 ความยาวคลื่น หลังจากการทำ PCA พบว่าตัวแปรที่สร้างขึ้นใหม่จำนวน 2 ตัวแปร (ในที่นี้ให้เป็น PC1 และ PC2) สามารถใช้ในการเก็บข้อมูลหลักจากข้อมูลทั้งหมด 24 ความยาวคลื่น (หรือตัว

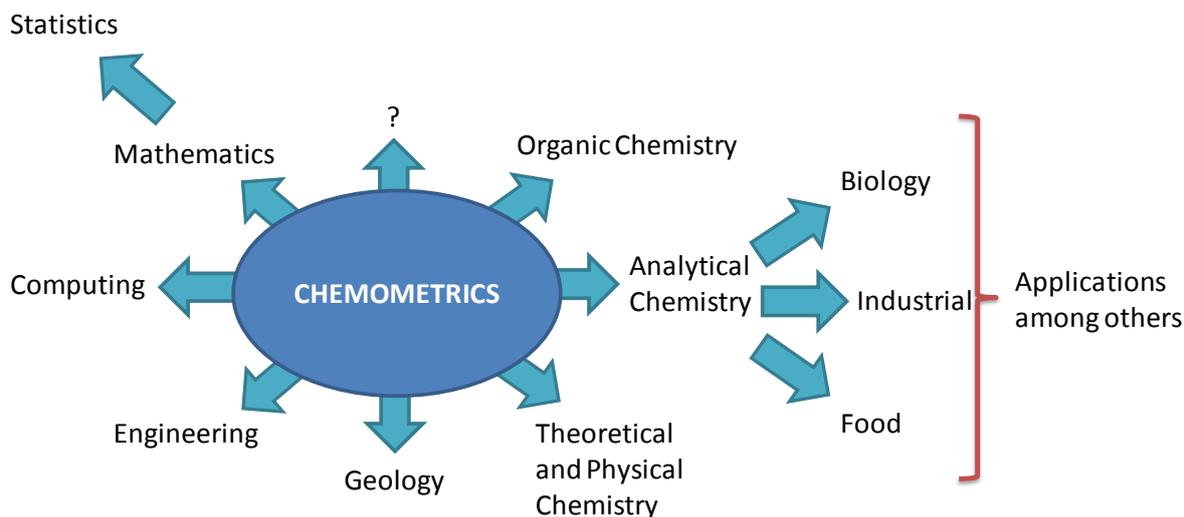
แปร) เมื่อนำตัวแปรทั้ง 2 ที่สร้างขึ้นใหม่มาพล็อตเพื่อแสดงผลก็จะได้เป็นกราฟดังรูป 5 และจากผลดังกล่าว เราจะสามารถสังเกตและสรุปในขั้นต้นได้ว่า ข้อมูลสเปกตรัมนี้สามารถจัดรวมตัวเป็นได้เป็น 3 กลุ่มที่แตกต่างกัน

PCA มักจะใช้ในการสำรวจข้อมูลพื้นฐาน (data exploratory) แต่อย่างไรก็ตาม PCA ก็ยังสามารถถูกประยุกต์เข้าใช้กับงานอื่นได้อย่างหลากหลาย ซึ่งจะกล่าวถึงในหัวข้อถัดไป



รูปที่ 7.3 ตัวอย่างการประยุกต์ใช้เคโมเมตริกซ์ในการแปรผลการทดลอง

นอกจากการใช้เคโมเมตริกซ์สำหรับการใช้สำรวจข้อมูลขั้นต้น (data exploratory) ยังมีการประยุกต์ต่างๆ อีกมากมายเช่น experimental designs, pattern recognition/classification, multivariate calibration, process monitoring และ resolution เป็นต้น ดังนั้นเราจะเห็นได้ว่าการประยุกต์ใช้งานของเคโมเมตริกซ์นั้นมีความหลากหลาย ทั้งนี้ขึ้นอยู่กับข้อมูลที่น่าสนใจ รวมทั้งมุมมองของนักวิจัยที่มีต่อข้อมูลนั้น การเชื่อมต่อระหว่างเคโมเมตริกซ์และงานวิจัยแขนงต่างๆ อาจแสดงได้ดังรูป 6 อาจกล่าวได้ว่าการประยุกต์ใช้งานของเคโมเมตริกซ์นั้นแฝงอยู่ในวิทยาศาสตร์แขนงต่างๆ ที่มีส่วนเกี่ยวข้องกับข้อมูลและการประมวลผลเพื่อตอบคำถามของโจทย์ที่ได้ตั้งไว้นั่นเอง



รูปที่ 7.4 การประยุกต์ใช้เคโมเมทริกซ์สำหรับงานวิจัยในแขนงต่างๆ

7.3 Principal Component Analysis (PCA)

จากที่ได้ยกตัวอย่างการใช้งานของเคโมเมทริกซ์ในหัวข้อที่ผ่านมา PCA ถูกใช้ในการสำรวจข้อมูลขั้นต้น (data exploratory) แต่จริงๆ แล้ว PCA ยังสามารถถูกประยุกต์ใช้เพื่อวัตถุประสงค์อื่นได้อีกด้วย เช่น การปรับข้อมูลให้อยู่ในสถานะที่เหมาะสมก่อนที่จะใช้เทคนิคอื่นๆ หรือถูกนำมาใช้ในการพิจารณาตัวแปรหรือพารามิเตอร์ที่มีความสำคัญ และ PCA ยังแฝงอยู่ในขั้นตอนที่สำคัญสำหรับเทคนิคเคโมเมทริกซ์อื่นๆ เช่น principal component regression (PCR), partial least square regression (PLS), D and Q statistics เป็นต้น ในหัวข้อนี้จะกล่าวถึง PCA ในส่วนของรายละเอียดที่มีความจำเป็น ได้แก่ 1. ลักษณะทางโครงสร้าง 2. การประยุกต์ใช้ score และ 3. การประยุกต์ใช้ loading

	1	2	3	4	Wavelengths	21	22	23	24													
1	385	498	1128	2826	5098	6878	8235	8714	9179	10017	10188	9792	9074	7832	6500	5565	5063	4938	4962	4930	4779	4364	3710	3099
2	385	476	1033	2633	4869	6656	8037	8552	9018	9897	10099	9704	9016	7809	6443	5512	5010	4876	4909	4864	4742	4328	3696	3078
3	381	458	947	2376	4478	6180	7495	8047	8433	9292	9573	9224	8601	7543	6266	5349	4856	4699	4736	4679	4569	4197	3628	3036
4	353	411	723	1553	2700	3637	4585	5262	5982	7013	7472	7241	6743	5867	4876	4212	3919	3909	4061	4084	4009	3700	3189	2659
5	356	424	771	1671	2880	3851	4873	5579	6367	7468	7916	7684	7143	6181	5138	4447	4138	4155	4307	4315	4244	3911	3339	2793
6	354	417	757	1656	2858	3850	4890	5578	6350	7427	7881	7649	7119	6196	5138	4448	4133	4142	4269	4294	4225	3887	3344	2787
7	348	409	716	1480	2444	3217	4075	4703	5438	6386	6786	6607	6150	5335	4441	3863	3622	3667	3800	3839	3772	3485	2995	2498
8	350	428	792	1672	2712	3599	4550	5237	6086	7092	7426	7224	6696	5771	4786	4191	3940	4002	4133	4179	4130	3769	3223	2689
9	356	434	826	1799	3019	4053	5133	5933	6876	8060	8497	8267	7666	6612	5508	4808	4520	4568	4720	4786	4703	4324	3707	3076
10	378	472	1022	2548	4632	6304	7607	8093	8539	9359	9578	9226	8607	7490	6252	5375	4903	4775	4802	4728	4616	4219	3624	3027
11	377	464	988	2422	4414	5982	7204	7700	8092	8893	9143	8822	8232	7192	5984	5149	4676	4560	4587	4526	4412	4046	3485	2908
12	377	459	936	2330	4333	5934	7166	7674	8054	8847	9098	8745	8159	7135	5924	5056	4599	4462	4500	4457	4336	3971	3427	2864
13	355	421	777	1687	2878	3841	4851	5568	6366	7452	7904	7643	7081	6145	5105	4435	4125	4149	4304	4314	4239	3905	3357	2801
14	354	419	763	1640	2814	3790	4794	5505	6308	7387	7822	7595	7059	6131	5095	4403	4097	4122	4276	4293	4223	3891	3343	2781
15	354	423	785	1706	2904	3890	4917	5653	6473	7569	7996	7809	7295	6363	5332	4639	4327	4312	4438	4445	4356	3989	3433	2856
16	396	522	1290	3368	6167	8223	9456	9262	8830	8933	8725	8346	7852	6932	5858	5093	4679	4650	4789	4839	4803	4427	3801	3175
17	387	501	1176	3053	5565	7430	8545	8377	7997	8081	7881	7521	7075	6264	5287	4600	4230	4186	4316	4362	4308	3992	3425	2862
18	384	472	1034	2721	5158	7059	8211	8199	7832	7978	7905	7579	7191	6451	5498	4791	4392	4311	4408	4445	4376	4068	3547	2984
19	375	457	972	2459	4609	6227	7214	7192	6837	6932	6824	6509	6160	5521	4679	4051	3732	3693	3801	3821	3796	3528	3066	2573
20	364	435	894	2172	3981	5305	6139	6077	5779	5849	5757	5493	5187	4660	3968	3441	3163	3136	3242	3240	3218	2998	2610	2199
21	380	469	1044	2861	4929	6659	7713	7674	7371	7506	7458	7197	6910	6316	5476	4861	4540	4453	4547	4564	4538	4295	3823	3348
22	351	413	740	1581	2704	3625	4571	5250	5997	7015	7464	7259	6738	5886	4909	4252	3967	3975	4122	4140	4054	3719	3213	2686
23	354	428	793	1692	2845	3789	4773	5485	6288	7349	7757	7508	6979	6024	5011	4346	4067	4093	4243	4269	4186	3842	3303	2750
24	354	442	849	1863	3094	4136	5197	5934	6849	7977	8331	8085	7480	6433	5329	4641	4363	4400	4554	4606	4518	4140	3537	2940
25	384	506	1209	3113	5546	7324	8385	8178	7848	7930	7721	7365	6960	6143	5189	4520	4179	4152	4254	4281	4240	3919	3357	2809
26	380	479	1098	2832	5240	7053	8138	8030	7661	7772	7621	7259	6847	6076	5147	4439	4088	4046	4170	4202	4147	3845	3319	2780
27	379	468	1034	2654	4963	6730	7807	7758	7419	7551	7459	7171	6839	6149	5274	4605	4252	4178	4288	4302	4269	3974	3480	2942
28	380	464	991	2462	4552	6214	7481	7959	8350	9172	9404	9054	8416	7335	6078	5187	4710	4576	4641	4577	4468	4102	3521	2917
29	377	451	910	2254	4249	5862	7117	7637	7977	8788	9061	8720	8134	7105	5916	5055	4591	4463	4488	4433	4313	3964	3415	2860
30	384	475	1021	2603	4870	6712	8085	8657	9077	9969	10212	9814	9135	7969	6805	5637	5116	4981	5026	4968	4820	4429	3800	3159
31	375	464	994	2453	4456	6064	7250	7682	8036	8781	8964	8625	8013	6960	5790	4953	4511	4407	4459	4419	4329	3966	3397	2834
32	379	490	1095	2700	4770	6393	7618	8021	8419	9156	9295	8934	8313	7183	5971	5128	4698	4597	4643	4612	4490	4101	3509	2933
33	379	484	1077	2679	4824	6526	7821	8248	8680	9455	9642	9269	8623	7489	6227	5334	4880	4763	4820	4799	4676	4275	3659	3049
34	352	415	749	1617	2770	3719	4709	5409	6192	7261	7692	7492	6949	6032	4996	4335	4034	4059	4210	4241	4160	3840	3310	2746
35	354	414	742	1602	2760	3684	4654	5350	6090	7156	7618	7398	6868	5993	4974	4302	4016	4021	4166	4196	4135	3800	3283	2722
36	353	428	802	1726	2901	3856	4853	5558	6373	7441	7865	7625	7072	6134	5108	4418	4132	4139	4322	4336	4250	3904	3348	2792
37	357	424	777	1685	2892	3894	4920	5651	6465	7578	8050	7827	7284	6365	5312	4597	4294	4299	4425	4441	4357	4013	3462	2891
38	358	430	796	1774	3073	4157	5269	6063	6962	8141	8624	8397	7791	6784	5647	4878	4548	4569	4720	4765	4666	4290	3682	3062
39	353	421	773	1647	2765	3688	4669	5391	6231	7315	7733	7514	6965	6023	5003	4343	4075	4120	4280	4330	4254	3907	3349	2800

รูป 7.5 เมทริกซ์ที่บรรจุค่าการดูดกลืนแสงของ 39 ตัวอย่าง โดยมีจำนวนความยาวคลื่นที่ใช้ในการตรวจวัดทั้งหมด 24 ความยาวคลื่น

จากข้อมูลสเปกตรัมของตัวอย่าง 39 ตัวอย่างในรูป 7.7 มีความยาวคลื่นที่ถูกตรวจวัดค่าการดูดกลืนแสงทั้งหมด 24 ความยาวคลื่น เมื่อเรานำข้อมูลดังกล่าวมาจัดเรียงในเมทริกซ์ (matrix) จะได้ดังรูป 7.7 (โดยปรกติแล้วการจัดเรียงข้อมูลลงในเมทริกซ์สามารถทำได้ 2 ลักษณะคือ 1. row wise system และ 2. column wise system) ในกรณีนี้จะขอกล่าวถึงการจัดเรียงแบบ row wise system โดยการจัดเรียงข้อมูลในลักษณะนี้ ในแต่ละแถว (row) จะอ้างอิงถึงข้อมูลของแต่ละตัวอย่าง เช่น ในแถวที่ 1 จะบรรจุข้อมูลค่าการดูดกลืนแสง (absorbance) ของตัวอย่างหมายเลข 1 สำหรับในแต่ละคอลัมน์ (column) จะอ้างอิงถึงข้อมูลของแต่ละค่าการดูดกลืนแสงของแต่ละความยาวคลื่น เช่น ในคอลัมน์ที่ 24 จะบรรจุค่าการดูดกลืนแสงของค่าความยาวคลื่นหมายเลข 24 เป็นต้น ดังนั้นถ้าเราตั้งชื่อเมทริกซ์นี้ว่าเมทริกซ์ X การระบุสมาชิก $x_{2,24}$ จะหมายถึงค่าการดูดกลืนแสงของตัวอย่างหมายเลข 2 ที่ค่าความยาวคลื่นหมายเลข 24 เป็นต้น เราจะเห็นว่าการนำข้อมูลมาจัดเรียงในเมทริกซ์นั้นทำให้มีข้อได้เปรียบคือข้อมูลจะถูกจัดเรียงอย่างเป็นระเบียบซึ่งจะทำให้ง่ายต่อการอ้างอิงข้อมูล รวมทั้งช่วยให้สะดวกในการประมวลผลโดยใช้การ matrix operation เช่น การบวก คูณ หาร รวมไปถึงการหาค่าอินเวอร์ส (inverse) เป็นต้น การคำนวณโดยใช้ matrix operation เข้ามาช่วยนั้นนอกจากจะมีความสะดวกแล้วยัง

ช่วยเพิ่มความรวดเร็วในการคำนวณและความคล่องตัวในการออกแบบโปรแกรมเนื่องจากข้อมูลจำนวนมากสามารถถูกจัดเรียงในเมทริกซ์ ทั้งนี้ขนาดของข้อมูลหรือ dimension ของเมทริกซ์อาจถูกควบคุมด้วยประสิทธิภาพของคอมพิวเตอร์ที่ใช้

โดยปกติแล้วสัญญาณที่บันทึกได้มักจะไม่ได้อะไรเฉพาะข้อมูลที่เรากำลังต้องการ แต่มักจะมีสัญญาณที่ไม่เกี่ยวข้องในการวิเคราะห์รวมอยู่ด้วย เช่น ค่าความคลาดเคลื่อนที่ได้จากการวัดโดยใช้เครื่องชั่งน้ำหนัก ไม้บรรทัด หรือกระบอกตวง เป็นต้น รวมไปถึงสัญญาณรบกวน (noise) ที่แฝงอยู่ในสเปกตรัมและสัญญาณที่เกิดจากกระแสรบกวนจากไฟฟ้า ทั้งนี้อาจกล่าวได้ว่าเมทริกซ์ที่บรรจุข้อมูลที่ได้จากการทดลองเป็นเมทริกซ์ที่เกิดจากการรวมกันของเมทริกซ์ย่อยๆ ที่บรรจุข้อมูลที่แตกต่างกัน โดยในกรณีนี้จะแบ่งเป็น 1. เมทริกซ์ที่บรรจุข้อมูลที่เราต้องการใช้ในการวิเคราะห์ และ 2. เมทริกซ์ที่บรรจุข้อมูลค่าความคลาดเคลื่อนหรือ noise ซึ่งแสดงได้ดังรูป 7.8

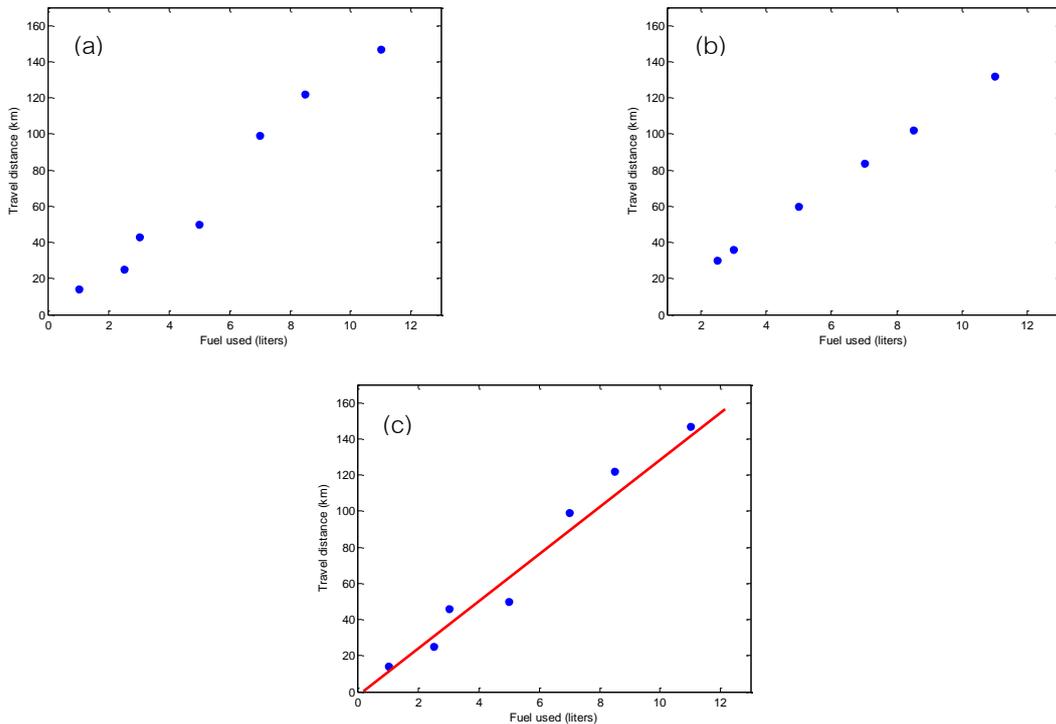
$$\boxed{\begin{array}{c} \text{สัญญาณที่ได้} \\ \text{จากการทดลอง} \end{array}} = \boxed{\begin{array}{c} \text{ข้อมูลที่ใช้ในการ} \\ \text{วิเคราะห์} \end{array}} + \boxed{\begin{array}{c} \text{ค่าความคลาด} \\ \text{เคลื่อน(noise)} \end{array}}$$

รูป 7.6 องค์ประกอบของข้อมูลที่อยู่ในเมทริกซ์

เพื่อแสดงให้เห็นถึงการทำงานของ PCA ที่ชัดเจนยิ่งขึ้น ขอยกตัวอย่างการใช้ข้อมูลที่มีเพียง 2 ตัวแปรเพื่อความสะดวกต่อการเข้าใจ รูป 7.9 เป็นผลการทดลองที่ได้จากการบันทึกปริมาณน้ำมัน (ลิตร) ที่ใช้กับระยะทาง (กิโลเมตร) ที่รถยนต์วิ่งไปได้ ในทางทฤษฎีถ้ารถยนต์คันนี้มีอัตราการใช้เชื้อเพลิงคือ 12 กิโลเมตรต่อลิตร ผลการทดลองที่ได้ควรจะเป็นเส้นตรงเรียงอยู่บนแนวเดียวกันดังรูป 9(b) แต่ในความเป็นจริงแล้วผลการทดลองจริงที่ได้ (รูป 7.9(a)) มีการเบี่ยงเบนไปจากค่าทางทฤษฎี ทั้งนี้อาจเนื่องมาจากปัจจัยต่างๆ เช่น ความขรุขระของผิวถนน กระแสลมในขณะการทดลอง หรือความร้อนสะสมที่เกิดจากเครื่องยนต์ เป็นต้น อย่างไรก็ตาม ปัจจัยที่ทำให้ผลการทดลองคลาดเคลื่อนจะยังไม่ถูกนำมาพิจารณาในขั้นตอนนี้

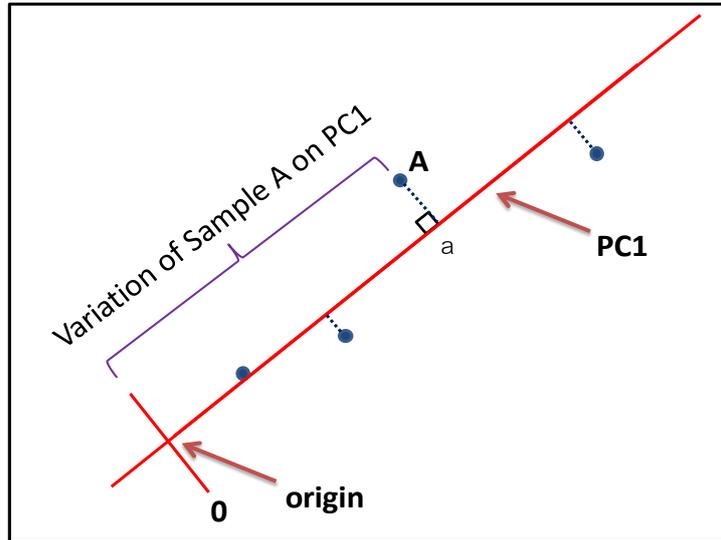
ถ้าเรากลับมาพิจารณาผลการทดลองที่ได้ เราจะพบว่าการอธิบายผลการทดลองในกราฟนั้นจะต้องใช้ 2 ตัวแปรประกอบกัน เช่น ที่ตำแหน่งจุด A จะพบว่าตัวแปรน้ำมันที่ใช้ไปมีค่า 1 ลิตร และตัวแปรระยะทางที่วิ่งได้มีค่า 14 กิโลเมตร เป็นต้น การพิจารณาข้อมูลในลักษณะนี้สามารถทำได้เช่นเดียวกันสำหรับข้อมูลที่มี 3 ตัวแปรหรือมากกว่า แต่สำหรับข้อมูลที่มีมากกว่า 3 ตัวแปรนั้นการแสดงผลอาจทำได้ลำบากเนื่องจากการรับรู้ของมนุษย์ยังสามารถทำได้เพียง 3 มิติเท่านั้น (ไม่มีกราฟที่มี 4 มิติ) เมื่อเป็นเช่นนี้เพื่อให้สะดวกต่อการพิจารณาข้อมูล แทนที่เราจะใช้ข้อมูลทั้งหมดทุกตัวแปรที่มีอยู่

เราจะใช้ PCA เพื่อสร้างแกนหรือตัวแปรขึ้นมาใหม่ โดยอาศัยหลักการคือ แกนหรือตัวแปรที่สร้างขึ้นมานั้นจะต้องสามารถใช้ในการบรรจุข้อมูลที่มีอยู่เดิมเอาไว้ได้มากที่สุด ในกรณีนี้เมื่อพิจารณารูป 7.9(c) จะพบว่าแกนที่เป็นไปได้คือ เส้นสีแดงที่ลากผ่านตำแหน่งใกล้เคียงของข้อมูลเกือบทั้งหมด



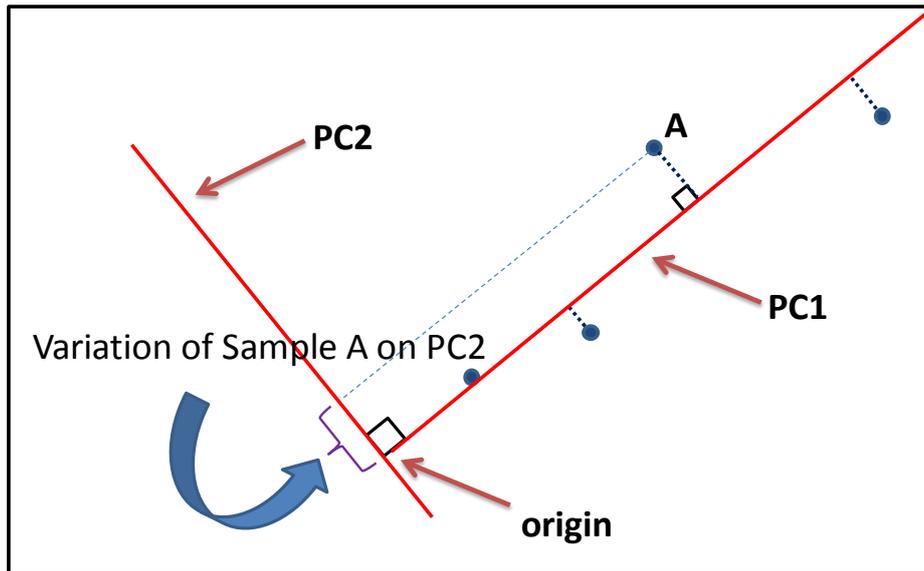
รูป 7.7 ข้อมูลแสดงอัตราการใช้แก๊สของรถยนต์

เมื่อพิจารณาแกนใหม่ที่สร้างขึ้น ถ้าเรา project ข้อมูลของเราลงบนแกนที่สร้างขึ้นใหม่ ดังรูป 10 คราวนี้แทนที่เราจะพิจารณาข้อมูลโดยใช้ 2 ตัวแปรพร้อมกัน เราสามารถใช้เพียง 1 ตัวแปรที่สร้างขึ้นใหม่สำหรับการอธิบายข้อมูลชุดเดียวกันได้ โดยใช้ค่าการกระจัดบนแกนที่ได้ออกมาจากการ project ข้อมูลลงไปบนแกนนั้นๆ เช่น จุด A ในรูป 10 การกระจัดบนแกนที่สร้างขึ้นใหม่ก็คือระยะทางจากจุด 0 ถึงจุด a ในทางเคโม่เมตริกซ์การกระจัดของข้อมูล A บนแกนที่สร้างขึ้นใหม่มักจะถูกเรียกว่า variation และแกนหรือตัวแปรที่สร้างขึ้นใหม่มักจะถูกเรียกว่า component โดยปกติแล้ว PC แรกที่ถูกกำหนดขึ้นจะเป็น PC ที่บรรจุข้อมูลเอาไว้มากที่สุด กล่าวคือ ผลรวมของ variation ของตัวอย่างทุกตัวบนแกนนี้จะมีค่ามากที่สุดและแกนนี้จะถูกเรียกว่า principal component number 1 (PC1) ซึ่งเป็นที่มาของคำว่า principal component analysis (PCA)



รูป 7.8 ข้อมูลแสดงอัตราสารกนน้ำมันของรถยนต์และแกน PC1

คุณสมบัติที่สำคัญอย่างหนึ่งของ PCA คือความเป็น orthogonal กล่าวคือแต่ละ component นั้นจะตั้งฉากกัน นั่นคือ correlation ของแต่ละคู่ PC นั้นเป็นศูนย์ หมายความว่า PC1 จะตั้งฉากกับ PC2 และ PC2 นั้นจะตั้งฉากกับทั้ง PC1 และ PC3 เป็นต้น เมื่อพิจารณารูป 11 จะเห็นว่า PC2 ตั้งฉากกับ PC1 และพบว่า variation ของตัวอย่างแต่ละตัวที่อยู่บน PC2 จะน้อยกว่าเมื่อเทียบกับ PC1 อย่างเห็นได้ชัด ทั้งนี้พบว่าข้อมูลที่อยู่บนแกน PC2 เป็นข้อมูลที่ไม่มีความสอดคล้องกับตัวแปรที่ใช้ในการศึกษาทั้ง 2 ตัว (พิจารณาโดยใช้รูป 7.9 ประกอบ) เพราะเมื่อพิจารณาอย่างละเอียดโดยใช้ความรู้พื้นฐานพบว่าความสัมพันธ์ของทั้ง 2 ตัวแปรควรเป็นแบบแปรผันตรง (ยกตัวอย่างความสัมพันธ์แบบแปรผันตรงเช่น เมื่อเติมน้ำมันเพิ่มขึ้นรถยนต์ควรจะวิ่งไปได้ไกลมากขึ้น) ดังนั้นเราจะพบแนวโน้มการกระจัดของ variation ของตัวอย่างที่เป็นบวกบน PC1 เท่านั้น ซึ่งตรงกันข้ามกับแนวโน้มการกระจัดของ variation บน PC2 ที่มีทั้งในด้านบวกและลบซึ่งไม่สอดคล้องกับหลักความเป็นจริง โดยปรกติเราจะพบว่า สำหรับ PC แกนถัดๆ ไป (PC2, PC3, PC4,... หมายถึง จำนวนของ component ที่มากที่สุดคือเท่ากับจำนวนตัวแปรที่ใช้ในการวิเคราะห์) จะแสดงค่าผลรวมของ variation ที่ลดลง ดังนั้นหากข้อมูลไม่ซับซ้อนมากนักการใช้เพียง 2 หรือ 3 PCs ก็เพียงพอแล้วสำหรับการแสดง variation ที่สำคัญ ของข้อมูล (สำหรับข้อมูลการกนน้ำมันของรถยนต์ข้างต้น การใช้เพียง PC1 นั้นก็เพียงพอแล้วในการอธิบายข้อมูลชุดนี้) การใช้ PC มากเกินไปเช่นการรวมเอา PC2 เข้ามาร่วมใช้ในการคำนวณ อาจจะเป็นการเอาข้อมูลที่ไม่เกี่ยวข้องกับการทดลองหรือ noise เข้ามาร่วมพิจารณาด้วยในโมเดล

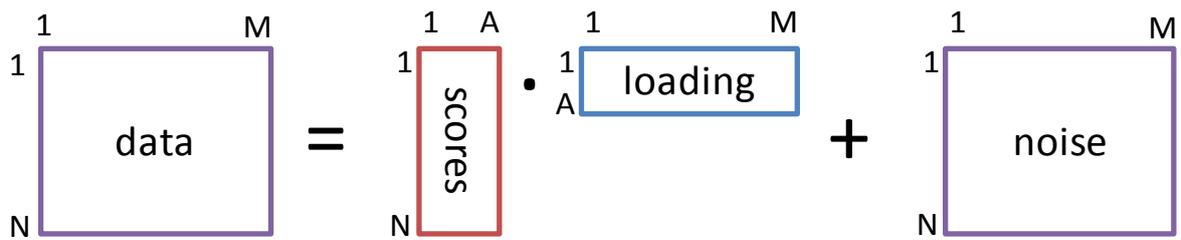


รูป 7.11 ข้อมูลแสดงอัตราการกินน้ำมันของรถยนต์และแกน PC1 และ PC2

- **Scores และ loadings**

จากที่ได้กล่าวไปแล้วว่า PCA สามารถใช้ในการสร้างตัวแปรใหม่ (PC) เพื่อใช้ในการแสดง variation ของข้อมูลแทนการใช้ตัวแปรเดิมที่มีอยู่ คำว่า variation เป็นคำที่มักใช้ในทางคณิตศาสตร์และมีความหมายจากพจนานุกรมว่า “การเปลี่ยนแปลง, การผันแปร, การแปรปรวน, จำนวนที่เปลี่ยนแปลง, รูปแบบที่เปลี่ยนแปลง, ความคลาดเคลื่อน” เพื่อลดความสับสนในการใช้งาน นักเคโมเมทริกซ์จะใช้คำว่า score แทนคำว่า variation ดังนั้นจำนวน score ของแต่ละ PC จะเท่ากับจำนวนตัวอย่าง และการแปรความจากค่า score ที่คำนวณได้อาจพิจารณาอย่างง่าย ๆ ว่าเป็นค่าที่แสดงถึงความสัมพันธ์ (relationship) ของแต่ละตัวอย่างบนแกนนั้นๆ ยกตัวอย่างเช่น ตัวอย่างที่มีค่า score ใกล้เคียงกันจะมีลักษณะบางประการที่คล้ายกัน (พิจารณารูป 7.5 ประกอบ) ความสัมพันธ์พื้นฐานนี้อาจสังเกตได้โดยการ plot ค่า score ของ PC ที่บรรจุ variation ไว้เป็นจำนวนมาก กราฟที่ได้จากการ plot ดังกล่าว จะถูกเรียกว่า score plot ในทางตรงกันข้าม ตัวอย่างที่มีค่า score แตกต่างกันอย่างมีลักษณะบางประการที่แตกต่างกัน ทำให้ถูกจัดวางในตำแหน่งที่แตกต่างกันบน score plot เป็นต้น

ในการคำนวณ PCA นอกจากจะได้ค่า score ของแต่ละตัวอย่างบนแกน PC แล้วยังมีอีกค่าหนึ่งซึ่งมีประโยชน์และความสำคัญในการวิเคราะห์เช่นกัน และจะถูกคำนวณควบคู่ออกมาพร้อมกันด้วยเสมอ คำนี้เรียกว่า loading จากที่ได้กล่าวไปแล้วว่า score เป็นค่าที่แสดงความสัมพันธ์ของแต่ละตัวอย่าง ในทำนองเดียวกัน loading จะเป็นค่าที่แสดงถึงความสัมพันธ์ของแต่ละตัวแปร



รูป 7.9 แสดงความสัมพันธ์ระหว่าง score และ loading ในสมการการคูณเมทริกซ์

ในรูป 7.12 แสดงความสัมพันธ์ระหว่าง score และ loading ในสมการการคูณเมทริกซ์ ในกรณีนี้สมมุติว่า data เมทริกซ์ ซึ่งมีขนาด $[N \times M]$ หรือมี N ตัวอย่างและมี M ตัวแปรที่ใช้ในการศึกษา หลังจากการทำ PCA พบว่ามีจำนวน PC ที่จะใช้ในการศึกษาอยู่ A PCs (ในขั้นตอนนี้จะยังไม่ขอกล่าวถึงวิธีการคำนวณจำนวน PC ที่เหมาะสมสำหรับการวิจัย) ดังนั้นจะพบว่าเมทริกซ์ score ของเรามีขนาด $[N \times A]$ และเมทริกซ์ loading มีขนาด $[A \times M]$ และผลของการคูณกันระหว่างเมทริกซ์ score และเมทริกซ์ loading จะได้เมทริกซ์ผลลัพธ์ที่มีขนาด $[N \times M]$ (สำหรับรายละเอียดวิธีการคูณกันของเมทริกซ์สามารถศึกษาได้จากตำราคณิตศาสตร์ทั่วไป) สามารถแสดงสัญลักษณ์ทางเมทริกซ์ได้ดังสมการ

$$X = T.P + E$$

โดยกำหนดให้ X แทนเมทริกซ์ที่บรรจุข้อมูล, E แทนเมทริกซ์ที่บรรจุ noise และ T แทนเมทริกซ์ score ส่วน P แทนเมทริกซ์ loading ในทำนองเดียวกันถ้าเราพิจารณาถึงสมาชิกแต่ละตัวในเมทริกซ์ก็จะสามารถแสดงได้ดังสมการ

$$x_{ij} = \sum_{r=1}^R t_{ir} p_{rj} + e_{ij}$$

ในกรณีนี้ x_{ij} แทนสมาชิกของเมทริกซ์ X ของตัวอย่างที่ i และ ตัวแปร j เช่นเดียวกับ e_{ij} แทนสมาชิกของเมทริกซ์ E ของตัวอย่างที่ i และ ตัวแปร j ในขณะที่ t_{ir} แทน score ของตัวอย่าง i บน r^{th} PC และ p_{rj} แทน loading ของตัวแปร j บน r^{th} PC จากสมการข้างต้น ในกรณีที่เราทราบจำนวน component ที่แน่นอน เราจะสามารถคำนวณหา residual เมทริกซ์ E โดย

$$X - TP = E$$

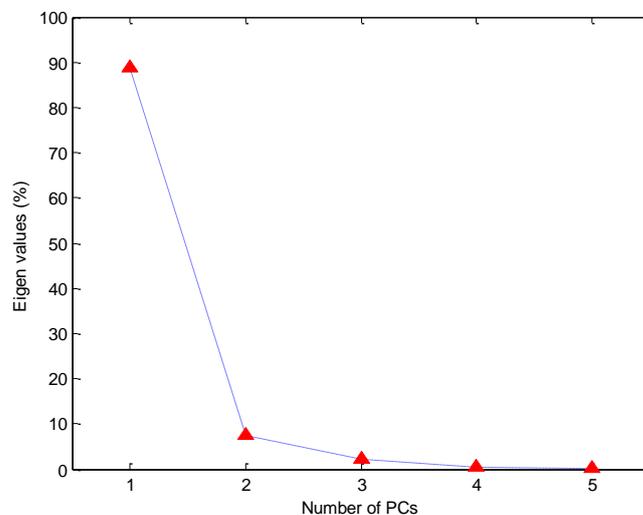
หรือ ในกรณีที่พิจารณาแต่ละสมาชิก

$$x_{ij} - \sum_{r=1}^R t_{ir} p_{rj} = e_{ij}$$

ดังนั้นเราจะพบว่าการกำหนดจำนวนของ component ที่ใช้ในการวิเคราะห์นั้นเป็นเรื่องที่สำคัญ ในกรณีที่มีการใช้จำนวน component มากเกินไป ข้อมูลที่ไม่เกี่ยวข้องกับปัจจัยที่เราต้องการศึกษาหรือ noise อาจจะถูกนำมาพิจารณาในโมเดลด้วย ยกตัวอย่างเช่นการใช้เพียง 1 component ก็เพียงพอแล้วในการ

แสดงถึง variation ที่จำเป็นสำหรับข้อมูลการกินน้ำมันของรถยนต์ เนื่องจากมีความเป็นไปได้สูงที่ variation on component ที่ 2 นั้นแสดงข้อมูลที่เป็นความคลาดเคลื่อนที่เกิดขึ้นในระหว่างการทดลอง ในทางตรงกันข้ามการใช้เพียง 1 component นั้นอาจจะไม่เพียงพอสำหรับข้อมูลสเปกตรัมที่ยกตัวอย่าง ในรูป 4 เป็นต้น วิธีที่ง่ายที่สุดในการพิจารณาจำนวน component ที่เหมาะสมสำหรับแต่ละข้อมูลคือการพิจารณาค่า eigen value ของแต่ละ component (eigen value เท่ากับ ผลรวมของ variation บน component นั้นๆ ยกกำลังสอง) component ที่ควรจะถูกนำมาใช้ในการพิจารณาค่า eigen value ที่สูงและในทางตรงกันข้าม component ที่แสดงแต่ variation ของ noise มักมีค่า eigen value ต่ำ

รูป 7.13 แสดงค่า eigen value ของแต่ละ component สำหรับข้อมูลสเปกตรัม จากกราฟที่ได้ เราจะพบว่าหลังจาก component ที่ 2 ค่า eigen value มีการลดลงอย่างเห็นได้ชัด ดังนั้นการใช้เพียง 2 component นั้นก็เพียงพอแล้วสำหรับนำมาใช้พิจารณาข้อมูลชุดนี้ ทั้งนี้การพิจารณาค่า eigen value นั้น เป็นเพียงการพิจารณาอย่างผิวเผิน ซึ่งมีเทคนิคอื่นๆ ที่ได้รับความนิยมสำหรับการพิจารณาจำนวน component ที่เหมาะสม ยกตัวอย่างเช่น cross validation และ boot strap เป็นต้น



รูป 7.10 eigen value สำหรับข้อมูลสเปกตรัมจาก 39 ตัวอย่างที่แสดงในรูป 7.5

8. Output จากโครงการวิจัยที่ได้รับทุนจาก สกว.

1. ผลงานตีพิมพ์ในวารสารวิชาการนานาชาติ (ระบุชื่อผู้แต่ง ชื่อเรื่อง ชื่อวารสาร ปี เล่มที่ เลขที่ และหน้า) หรือผลงานตามที่คาดไว้ในสัญญาโครงการ

โครงการวิจัยภายใต้รหัสสัญญา TRG5780175 ได้ดำเนินการจนเสร็จแล้ว และแบบจำลองสร้างขึ้นโดยเทคนิคเคโมเมทริกซ์แบบปัญญาประดิษฐ์นี้ได้ถูกนำมาประยุกต์ใช้ในงานวิจัยที่จำเป็นต้องวิเคราะห์ข้อมูลทางเคมีที่ซับซ้อน ซึ่งผลลัพธ์ที่ได้จากงานวิจัยนั้น ผู้วิจัยได้จัดทำบทความ (manuscript) เพื่อส่งตีพิมพ์ในวารสารวิชาการนานาชาติ จำนวน 2 ฉบับ ดังนี้

1. วารสาร Chemometrics and Intelligent Laboratory Systems (impact factor = 2.321)

ซึ่งทำการ submit วันที่ 21 พฤศจิกายน 2558 ขณะนี้อยู่ในสถานะ Under revision (กำหนดให้ส่งกลับ revised manuscript คือ 21 พฤษภาคม 2559)

2. วารสาร Journal of Chemometrics (impact factor = 1.5)

ซึ่งทำการ submit วันที่ 23 มีนาคม 2559 ขณะนี้อยู่ในสถานะ Under revision

2. การนำผลงานวิจัยไปใช้ประโยชน์

- **เชิงสาธารณะ**

- แบบจำลองที่ผู้วิจัยได้พัฒนาขึ้นนี้ ก่อให้เกิดการสร้างเครือข่ายความร่วมมือในการวิเคราะห์ข้อมูลโดยใช้เทคนิคเคโมเมทริกซ์ รวมถึงแก้ไขปัญหาในงานวิจัยในโครงการศึกษาปัจจัยที่มีผลกระทบต่อผลผลิตและคุณภาพของข้าวหอมมะลิในภาคตะวันออก เฉียงเหนือ โดยสร้างเครือข่ายความร่วมมือกับศูนย์วิจัยข้าวอุบลราชธานี อำเภอเมือง จังหวัดอุบลราชธานี

- **เชิงวิชาการ**

- ส่วนหนึ่งของงานวิจัยถึงนำไปใช้ในการเรียนการสอนกระบวนวิชา 203458 "การประยุกต์ใช้เคโมเมทริกซ์สำหรับการติดตามคุณภาพของกระบวนการผลิตในโรงงานเคมี"
- โครงการวิจัยนี้มีส่วนในการสร้างนักศึกษาปริญญาโท จำนวน 2 คน คือ น.ส. สุกุลนา วงศ์สายปັນ รหัสนักศึกษา 570531034 และ น.ส. สุจิตรา ฝั้นสีบ รหัสนักศึกษา 560531044

3. อื่นๆ

- การเสนอผลงานแบบบรรยายในการประชุมวิชาการ "The 9th conference on science and technology for youths" วันที่ 30 พ.ค. 2557 - 1 มิ.ย. 2557 ณ ศูนย์นิทรรศการและการประชุมไบเทค บางนา ในหัวข้อ " Chemometric analysis of near-infrared (NIR) spectra from reused deep-frying cooking oil samples"
- การเสนอผลงานแบบโปสเตอร์ในการประชุมประจำปี "นักวิจัยรุ่นใหม่...พบ...เมธีวิจัยอาวุโส สกว." ครั้งที่ 15 ระหว่างวันที่ 6 – 8 มกราคม พ.ศ. 2559 ณ โรงแรมเดอะรีเจ้นท์ ชะอำ บีช รีสอร์ท จังหวัดเพชรบุรี ในหัวข้อ " Application of multiple self organizing maps (mSOM) for continuous process monitoring"
- การเสนอผลงานแบบโปสเตอร์ในการประชุมวิชาการนานาชาติ "XVI CHEMOMETRICS IN ANALYTICAL CHEMISTRY (CAC 2016)" ระหว่างวันที่ 6 - 10 มิถุนายน 2559 ณ เมือง Barcelona ประเทศสเปน ในหัวข้อ "Development of Multiple Self Organizing Maps for Classification Problems" และหัวข้อ "Outlier Detection using Bootstrap Estimates"