



รายงานการวิจัย

การตรวจจับสิ่งผิดปกติในกระบวนการผลิตปุ๋ยอินทรีย์ชีวภาพ
ด้วยเครือข่ายเซ็นเซอร์ไร้สาย

(Anomaly Detection in a Bioorganic Fertilizer Production Process
using Wireless Sensor Networks)

มหาวิทยาลัยเทคโนโลยีสุรนารี

ได้รับทุนอุดหนุนการวิจัยจาก
มหาวิทยาลัยเทคโนโลยีสุรนารี

ผลงานวิจัยเป็นความรับผิดชอบของหัวหน้าโครงการวิจัยแต่เพียงผู้เดียว



รายงานการวิจัย

การตรวจจับสิ่งผิดปกติในกระบวนการผลิตปุ๋ยอินทรีย์ชีวภาพ

ด้วยเครือข่ายเซ็นเซอร์ไร้สาย

(Anomaly Detection in a Bioorganic Fertilizer Production Process
using Wireless Sensor Networks)

คณะผู้วิจัย

หัวหน้าโครงการ

ผศ. ดร. วิภาวี หัตถกรรม

สาขาวิชาวิศวกรรมโทรคมนาคม

สำนักวิชาวิศวกรรมศาสตร์

มหาวิทยาลัยเทคโนโลยีสุรนารี

ผู้ร่วมวิจัย

ผศ. ดร. สุธชต วุ่นประเสริฐ

สาขาวิชาเทคโนโลยีการผลิตพืช

สำนักวิชาเทคโนโลยีการเกษตร

มหาวิทยาลัยเทคโนโลยีสุรนารี

ได้รับทุนอุดหนุนการวิจัยจากมหาวิทยาลัยเทคโนโลยีสุรนารี ปีงบประมาณ 2554-2555

ผลงานวิจัยเป็นความรับผิดชอบของหัวหน้าโครงการวิจัยแต่เพียงผู้เดียว

มีนาคม 2560

กิตติกรรมประกาศ

การวิจัยครั้งนี้ได้รับทุนอุดหนุนการวิจัยจากมหาวิทยาลัยเทคโนโลยีสุรนารี ปีงบประมาณ 2554-2555 ทั้งนี้ ผู้วิจัยขอขอบคุณผู้ร่วมวิจัย ผศ.ดร. สุธชด วุ่นประเสริฐ ผู้ให้ความอนุเคราะห์ความรู้เกี่ยวกับลักษณะและคุณสมบัติต่างๆของดิน ศ.ดร. หนึ่ง เตียอำรุง ที่ปรึกษาโครงการวิจัย ผู้ให้ความอนุเคราะห์ความรู้ด้านการทำงานของจุลินทรีย์ย่อยสลายในปุ๋ยชีวภาพ นอกจากนี้ ผู้วิจัยขอขอบคุณนักศึกษาบัณฑิตศึกษา ที่ได้ทำงานวิจัยด้วยความวิริยะอุตสาหะ ภายใต้หัวข้อการวิจัยนี้ ได้แก่ นางสาวเสาวลักษณ์ ตะเคียนงาม ที่ศึกษาการตรวจจับสิ่งผิดปกติในเครือข่ายเซนเซอร์ไร้สาย โดยใช้การแปลงลิฟต์ตั้งเวฟเล็ต (Lifting Wavelet Transform) และซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine) นายศุภกิตต์ ศิริพนาดร ที่ศึกษาการตรวจจับสิ่งผิดปกติในเครือข่ายเซนเซอร์ไร้สาย โดยใช้การแปลงดิสครีตเวฟเล็ต (Discrete Wavelet Transform) และเซลล์ฟอว์แมป (Self-Organizing Map) นายอรรถชัย ภูพานิล และนางสาวศจี สิงห์สง่า ที่แปลและเรียบเรียงจากวิทยานิพนธ์ดั้งเดิมซึ่งเป็นภาษาอังกฤษ

นอกจากนี้ ผู้วิจัยขอขอบพระคุณสาขาวิชาวิศวกรรมโทรคมนาคม มหาวิทยาลัยเทคโนโลยีสุรนารี ที่ให้โอกาสและสนับสนุนในการทำวิจัยครั้งนี้ รวมทั้งผู้ที่มีส่วนเกี่ยวข้องในการทำวิจัยครั้งนี้ทุกท่าน ที่ได้กรุณาให้ความช่วยเหลือ สนับสนุนมาโดยตลอด ผู้วิจัยหวังเป็นอย่างยิ่งว่าผลการวิจัยนี้จะเป็นประโยชน์ในการสร้างองค์ความรู้ ด้านการหาเส้นทางส่งข้อมูลในเครือข่ายเซ็นเซอร์ไร้สาย เพื่อการตรวจจับสิ่งผิดปกติด้านการเกษตรต่อไป

วิภาวี หัตถกรรม

มีนาคม 2560

บทคัดย่อ

ปัจจุบันมีการประยุกต์ใช้งานเทคโนโลยีการสื่อสารไร้สายเพื่อการเกษตรเป็นจำนวนมาก เช่น การใช้เครือข่ายเซ็นเซอร์ไร้สายเพื่อเฝ้าระวังทางการเกษตร โดยอาศัยการเก็บข้อมูลทางกายภาพภายในแปลงเกษตรจากโหนดเซ็นเซอร์ไร้สายซึ่งถูกวางในตำแหน่งต่างๆกันในไร่หรือแปลงเกษตร และส่งข้อมูลกลับมาที่สถานีฐานผ่านการสื่อสารวิทยุระยะสั้น ในกรณีที่แปลงเกษตรมีขนาดใหญ่ ข้อมูลจากโหนดเซ็นเซอร์ไร้สายจะถูกส่งผ่านโหนดเซ็นเซอร์ข้างเคียง ซึ่งจะทำการส่งต่อข้อมูลภายในเครือข่ายในลักษณะนี้จนข้อมูลถูกนำส่งถึงสถานีฐาน เครือข่ายเซ็นเซอร์ไร้สายประกอบด้วยโหนดเซ็นเซอร์หลายโหนดซึ่งมีขนาดเล็กและราคาถูก เครือข่ายตัวตรวจรู้ไร้สายมีข้อจำกัดหลายอย่าง เช่น หน่วยความจำ, ความกว้างแถบความถี่ (แบนด์วิธ), อัตราการส่งข้อมูลต่ำ, แหล่งพลังงานและการใช้พลังงาน, และความสามารถเชิงประมวลผล เป็นต้น ข้อจำกัดของอุปกรณ์เหล่านี้ส่งผลกระทบต่อความสามารถในการตรวจจับ ความผิดปกติของตัวตรวจรู้ และสามารถก่อให้เกิดความเสียหายต่อผลผลิตได้ นอกจากนี้ข้อจำกัดด้านแหล่งพลังงานในเครือข่ายตัวตรวจรู้ไร้สาย ต้องการใช้พลังงานให้น้อยที่สุด เนื่องจากการส่งข้อมูลในเครือข่ายตัวตรวจรู้ไร้สายนั้น ใช้พลังงานมากกว่ากระบวนการในการประมวลผล และ การคำนวณข้อมูลที่มีขนาดเล็กจะใช้พลังงานน้อยกว่าการคำนวณข้อมูลขนาดใหญ่ด้วย

ดังนั้น การวิจัยนี้จึงมุ่งเน้นที่จะผสมผสานการทำงานระหว่าง การบีบอัดข้อมูลด้วย Discrete Wavelet Transform (DWT) และ Lifting Wavelet Transform (LWT) ร่วมกับ การตรวจจับความผิดปกติของข้อมูลโดยใช้ One-Class Support Vector Machine (OCSVM)

วิธีการแรกที่น่าเสนอในการวิจัยนี้ คือ OCSVM + DWT เมื่อนำไปเปรียบเทียบกับวิธีการก่อนหน้านี้ที่ถูกนำเสนอมาแล้ว เช่น Self-Organizing Map (SOM) + DWT พบว่า OCSVM + DWT สามารถเพิ่มประสิทธิภาพในการตรวจจับความผิดปกติได้ สำหรับการทดลองกับข้อมูลสังเคราะห์ พบว่า OCSVM + DWT ที่เลือกใช้ค่าสัมประสิทธิ์ความถี่ต่ำ มีอัตราความถูกต้องในการตรวจจับความผิดปกติถึง 100% ในขณะที่อัตราความผิดพลาดในการตรวจจับข้อมูลเพิ่มขึ้นเพียงเล็กน้อย เมื่อเปรียบเทียบกับวิธีการอื่น ๆ และในการทดลองกับชุดข้อมูลจริงพบว่า OCSVM + DWT ที่เลือกใช้ค่าสัมประสิทธิ์ความถี่ต่ำ ทำงานได้ดีที่สุด โดยมีอัตราความถูกต้องในการตรวจจับความผิดปกติสูงถึงเกือบ 100% แม้ว่าการทดลองกับข้อมูลที่มีความผิดปกติแบบ Short หรือ Noise จะให้อัตราการ

ตรวจจับข้อมูลที่มีความผิดปกติสูงกว่าวิธีการอื่น ๆ ก็ตาม จากการทดลองจะเห็นว่า OCSVM + DWT ที่เลือกใช้ค่าสัมประสิทธิ์ความถี่ต่ำ เหมาะกับการตรวจจับข้อมูลที่มีความผิดปกติแบบ Short หรือ Noise เป็นองค์ประกอบ ในขณะที่ SOM + DWT ที่เลือกใช้ค่าสัมประสิทธิ์ความถี่ต่ำ เหมาะกับการตรวจจับข้อมูลที่มีความผิดปกติแบบ Constant เป็นองค์ประกอบ

อีกวิธีการหนึ่งที่นำเสนอในงานวิจัยนี้ คือ OCSVM + LWT ซึ่งจะถูกนำไปเปรียบเทียบกับประสิทธิภาพการทำงานกับวิธีการอื่น ซึ่งได้แก่ OCSVM + DWT และ OCSVM + Principal Component Analysis (PCA) สำหรับการทดลองกับข้อมูลสังเคราะห์และข้อมูลจริงที่มีความผิดปกติแบบ Short เป็นองค์ประกอบ พบว่า OCSVM + LWT มีประสิทธิภาพการทำงานใกล้เคียงกับ OCSVM, OCSVM + DWT และ OCSVM + PCA สำหรับการทดลองกับข้อมูลสังเคราะห์และข้อมูลจริงที่มีความผิดปกติแบบ Noise และ Constant เป็นองค์ประกอบ พบว่า OCSVM + LWT และ OCSVM + DWT ที่เลือกใช้สัมประสิทธิ์ความถี่ต่ำ มีประสิทธิภาพการทำงานที่ดีกว่า OCSVM และ OCSVM + PCA ในทางกลับกัน OCSVM + LWT และ OCSVM + DWT ที่เลือกใช้สัมประสิทธิ์ความถี่สูง มีประสิทธิภาพการทำงานที่แย่ที่สุด ซึ่งแสดงให้เห็นว่า LWT มีความต้องการที่น้อยกว่า DWT ในแง่ของหน่วยความจำที่ใช้งาน และเวลาในการคำนวณ และจากผลการทดลองแสดงให้เห็นว่า OCSVM + LWT เหมาะที่จะนำไปติดตั้งเครือข่ายตัวตรวจรู้ไร้สายมากกว่าวิธีการอื่น ๆ ที่ได้กล่าวมาแล้ว

งานวิจัยนี้ยังได้ทำการประยุกต์ใช้ในการตรวจจับความผิดปกติของชุดข้อมูลจริงจากโรงปุ๋ยชีวภาพ (bioorganic fertilizer plant หรือ BOF) มหาวิทยาลัยเทคโนโลยีสุรนารี โดยได้มีการพัฒนาชุดอุปกรณ์ต้นแบบเซ็นเซอร์ โหนดไร้สายและมีการส่งค่าความชื้นและอุณหภูมิของดินไปยังสถานีฐานอย่างต่อเนื่อง และมีการนำข้อมูลผ่านการประมวลผลแบบ DWT และใช้ SOM ในการตรวจจับความผิดปกติในข้อมูล โดยพบว่าอัลกอริทึมที่ได้นำเสนอที่ประกอบด้วยเวฟเลตแม่แบบฮาร์ (Haar) ที่มีการใช้สัมประสิทธิ์ต่ำผ่านนั้นมีค่าการแจ้งเตือนจริง 95% โดยที่อัลกอริทึมที่นำเสนอนั้นมีประสิทธิภาพดีเทียบเท่าอัลกอริทึม SOM และมีประสิทธิภาพที่ดีกว่าอัลกอริทึม DWT ถึง 75%

ผลการทดลองแสดงให้เห็นว่าอัลกอริทึมที่ได้นำเสนอสามารถรักษาค่าการตรวจจับความผิดปกติจริงที่ยอมรับได้ ในขณะที่ใช้ข้อมูลป้อนเข้าเพียงครั้งเดียว (ด้วยการใช้ DWT level 1)

Abstract

Nowadays wireless communication technologies have been applied to enhance yield in agriculture. One such application is the use of wireless sensor networks (WSNs) for agriculture monitoring which collects physical measurements from the environment through wireless sensor nodes embedded at various locations in the field. The collected data are then transmitted back to the base station through short wave communication. In scenarios of large crop fields, such data are relayed hop-by-hop within the network to the base station. A WSN is a network composed of several small lowcost sensor nodes. Wireless sensor networks (WSNs) have many limitations such as memory, bandwidth, low-rate radio communication, energy supply and consumption, and computational capabilities. These limitations can affect the sensor node ability to detect anomalies and can damage produce. Furthermore, the battery supply limitations in WSNs require minimal energy consumption. Since radio communication in WSNs consume more energy than processing and computing, computation with small datasets is likely to consume less energy than a large dataset.

Therefore, this research is focused on incorporating the discrete wavelet transform (DWT) and lifting wavelet transform (LWT) data compression schemes with one-class support vector machine (OCSVM) anomaly detection.

Our first proposed algorithm (OCSVM + DWT) was compared with a previous algorithm i.e., self-organizing map (SOM) + DWT. We found the OCSVM + DWT can increase the efficiency of anomaly detection. For synthetic data, the OCSVM + DWT with low-pass coefficients (LP) achieved 100% detection rate (DR) with marginal increase in false positive rate (FPR) when compared with all other algorithms. For real world datasets, the OCSVM + DWT with LP coefficients performed best by achieving nearly 100% DR although with slightly higher FPR for datasets containing short and noise faults. These results suggest that OCSVM + DWT (LP) algorithm is suited for short and noise faults whereas SOM + DWT (LP) is suited for short and constant faults.

Our second proposed algorithm (OCSVM + LWT) was compared with other variants of integration such as OCSVM + DWT and OCSVM + principal component analysis (PCA) and OCSVM alone (with uncompressed data). For synthetic data and real world datasets with short faults, the OCSVM + LWT performed equally well as the OCSVM alone, OCSVM + DWT and OCSVM + PCA. For synthetic data and real world datasets with noise and constant faults, the OCSVM + LWT [LP] and the OCSVM + DWT [LP] gave better performance than the OCSVM alone and OCSVM + PCA. On the contrary the OCSVM + LWT [HP] with high-pass coefficients and the OCSVM + DWT [HP] gave the worst performance. It was also demonstrated that LWT was less demanding in terms of memory requirement and computation time than DWT. Our results therefore suggest that OCSVM + LWT was more suitable for implementation in WSNs.

This research has also implemented a prototype for datasets collected from a bioorganic fertilizer (BOF) plant in Suranaree University of Technology. The measured parameters include the soil moisture and temperature which were continuously measured and transmitted to the base station. The data collected were processed with DWT and SOM anomaly detection was used. The results showed that the proposed algorithm with Haar as a mother wavelet using low pass coefficients can attain 95% of true alarm rates. Our proposed algorithm also performed as equally well as the SOM algorithm and outperformed the DWT algorithm by up to 75%.

The results showed that our proposed algorithm can maintain acceptable anomaly detection accuracy while using just half of the input data (using level 1 DWT).

สารบัญ

	หน้า
กิตติกรรมประกาศ.....	ก
บทคัดย่อ (ภาษาไทย).....	ข
บทคัดย่อ (ภาษาอังกฤษ).....	ง
สารบัญ.....	ฉ
สารบัญรูป.....	ญ
สารบัญตาราง.....	ฐ
บทที่	
1 บทนำ	
1.1 ความเป็นมาของปัญหา.....	1
1.1.1 เทคนิคการตรวจจับความผิดปกติ.....	2
1.1.1.1 เทคนิคการตรวจจับความผิดปกติเชิงสถิติแบบพารามетริก.....	2
1.1.1.2 เทคนิคการตรวจจับความผิดปกติแบบที่ไม่ใช้พารามิเตอร์.....	3
1.1.2 เทคนิคการบีบอัดข้อมูล.....	8
1.2 จุดประสงค์งานวิจัย.....	9
1.3 ขอบเขตของโครงการวิจัย.....	10
1.4 ประโยชน์ที่คาดว่าจะได้รับ.....	10
2 วิธีการแปลงเวฟเล็ตแบบเต็มหน่วยและเทคนิคซัพพอร์ตเวกเตอร์แมชชีนคลาสหนึ่งสำหรับ	
การตรวจจับความผิดปกติในเครือข่ายเซ็นเซอร์ไร้สาย	
2.1 บทนำ.....	12
2.2 การตรวจจับความผิดปกติ.....	14
2.2.1 One-Class Support Vector Machines (OCSVM).....	14
2.2.2 Self-Organizing Map (SOM).....	17
2.3 การบีบอัดข้อมูล.....	19
2.3.1 การแปลงเวฟเล็ตแบบเต็มหน่วย.....	19

สารบัญ (ต่อ)

	หน้า
2.4 ชุดข้อมูลสำหรับการทดลอง.....	20
2.4.1 ข้อมูลที่มาจากการสังเคราะห์.....	22
2.4.2 ชุดข้อมูล INTEL.....	22
2.4.3 ชุดข้อมูล SensorScope Station no.39 dataset (SS39).....	22
2.4.4 ชุดข้อมูล SensorScope pdg2008-metro-1 dataset (pdg2008).....	23
2.4.5 ชุดข้อมูล NAMOS.....	23
2.5 ผลการทดลอง.....	23
2.5.1 การประเมินผลของ DWT ด้วยอัลกอริทึม OCSVM.....	23
2.5.2 เปรียบเทียบประสิทธิภาพกับงานที่ผ่านมา.....	29
2.6 บทสรุป.....	36
3 วิธีการแปลงเวฟเล็ตแบบลิฟติงและเทคนิคซัพพอร์ตเวกเตอร์แมชชีนคลาสหนึ่งสำหรับการ ตรวจจับความผิดปกติในเครือข่ายเซ็นเซอร์ไร้สาย	
3.1 บทนำ.....	37
3.2 การตรวจจับความผิดปกติ.....	38
3.2.1 One-Class Support Vector Machines (OCSVM).....	39
3.3 การบีบอัดข้อมูล.....	41
3.3.1 Principal Component Analysis (PCA).....	41
3.3.2 การแปลงเวฟเล็ตแบบเต็มหน่วย.....	43
3.3.3 Lifting Wavelet Transform (LWT).....	45
3.4 ผลการทดลอง.....	45
3.4.1 ชุดข้อมูลสำหรับการทดลอง.....	45
3.4.1.1 ข้อมูลสังเคราะห์.....	46
3.4.1.2 ชุดข้อมูลที่ได้จากการวัดจริง.....	50
3.4.2 การประเมินประสิทธิภาพ.....	57

สารบัญ (ต่อ)

	หน้า
3.4.3 ขยายผลสู่ชุดข้อมูล 3 KPI.....	61
3.4.4 การประเมินการคำนวณเชิงเวลา.....	66
3.5 สรุป.....	67
4 การประเมินผลการตรวจจับความผิดปกติของข้อมูลจากเครือข่ายเซ็นเซอร์ไร้สายสำหรับ เฟ้าระวังโรงผลิตปุ๋ยอินทรีย์ชีวภาพ	
4.1 บทนำ.....	68
4.2 อุปกรณ์ต้นแบบสำหรับเฟ้าระวังโรงผลิตปุ๋ยอินทรีย์ชีวภาพ.....	69
4.2.1 สถานีฐาน (Base Station).....	70
4.2.2 เซ็นเซอร์โหนด (Sensor mote).....	70
4.2.3 บอร์ดรวมข้อมูล (Data acquisition board).....	71
4.2.4 อุปกรณ์เซ็นเซอร์ภายนอกสำหรับตรวจวัดอุณหภูมิและความชื้นในดิน.....	72
4.3 ผลการทดลอง.....	73
4.4 สรุป.....	77
5 บทสรุป	
5.1 กระบวนการตรวจจับความผิดปกติโดยใช้สัมประสิทธิ์ DWT.....	78
5.2 กระบวนการตรวจจับความผิดปกติโดยใช้สัมประสิทธิ์ LWT.....	79
5.3 กระบวนการตรวจจับความผิดปกติในเครือข่ายเซ็นเซอร์ไร้สายสำหรับเฟ้าระวัง โรงผลิตปุ๋ยอินทรีย์ชีวภาพ.....	79
บรรณานุกรม	81
ภาคผนวก	
ภาคผนวก ก.....	86
ภาคผนวก ข.....	104
ภาคผนวก ค.....	147



สารบัญรูป

รูปที่	หน้า
2.1 เรขาคณิตของอัลกอริทึม quarter-sphere OCSVM.....	16
2.2 ตัวอย่างของวิธีการ SOM.....	17
2.3 ความผิดปกติในการอ่านข้อมูลของเซ็นเซอร์.....	21
2.4 ROC สำหรับข้อมูลสังเคราะห์ที่เพิ่มความผิดปกติ 1/80 เข้าไป.....	24
2.5 ROC สำหรับข้อมูลสังเคราะห์ที่เพิ่มความผิดปกติ 5/16 เข้าไป.....	25
2.6 ROC สำหรับข้อมูลสังเคราะห์ที่เพิ่มความผิดปกติ 10/8 เข้าไป.....	25
2.7 ROC สำหรับข้อมูลสังเคราะห์ที่เพิ่มความผิดปกติ 20/4 เข้าไป.....	26
2.8 ROC สำหรับชุดข้อมูลINTEL.....	26
2.9 ROC สำหรับชุดข้อมูลSS39.....	27
2.10 ROC สำหรับชุดข้อมูล pdg2008-metro-1.....	27
2.11 ROC สำหรับชุดข้อมูลNAMOS.....	28
2.12 อัตราการตรวจจับความผิดปกติด้วยอัลกอริทึมที่แตกต่างกัน สำหรับข้อมูลสังเคราะห์ ที่มีความผิดปกติ 1/80.....	31
2.13 อัตราการตรวจจับความผิดปกติด้วยอัลกอริทึมที่แตกต่างกัน สำหรับข้อมูลสังเคราะห์ ที่มีความผิดปกติ 5/16.....	32
2.14 อัตราการตรวจจับความผิดปกติด้วยอัลกอริทึมที่แตกต่างกัน สำหรับข้อมูลสังเคราะห์ ที่มีความผิดปกติ 10/8.....	32
2.15 อัตราการตรวจจับความผิดปกติด้วยอัลกอริทึมที่แตกต่างกัน สำหรับข้อมูลสังเคราะห์ ที่มีความผิดปกติ 20/4.....	33
2.16 อัตราการตรวจจับความผิดปกติด้วยอัลกอริทึมที่แตกต่างกันสำหรับชุดข้อมูล INTEL.....	33
2.17 อัตราการตรวจจับความผิดปกติด้วยอัลกอริทึมที่แตกต่างกันสำหรับชุดข้อมูล SS39.....	34
2.18 อัตราการตรวจจับความผิดปกติด้วยอัลกอริทึมที่แตกต่างกันสำหรับชุดข้อมูล pdg2008.....	34
2.19 อัตราการตรวจจับความผิดปกติด้วยอัลกอริทึมที่แตกต่างกันสำหรับชุดข้อมูล NAMOS.....	35

สารบัญรูป (ต่อ)

รูปที่	หน้า
3.1 เรขาคณิตของอัลกอริทึม Quarter-sphere OCSVM	39
3.2 รูปแบบการทำงานของหลักการวิเคราะห์องค์ประกอบหลัก (PCA).....	42
3.3 รูปแบบการทำงานของ Lifting Wavelet Transform (LWT).....	44
3.4 ความผิดปกติในการอ่านข้อมูลของเซ็นเซอร์	46
3.5 ข้อมูลสังเคราะห์ 2KPI กับความผิดพลาด 1/80.....	47
3.6 ข้อมูลสังเคราะห์ 3KPI กับความผิดพลาด 1/80.....	48
3.7 ข้อมูลสังเคราะห์ 2KPI กับความผิดพลาด 20/4	48
3.8 ข้อมูลสังเคราะห์ 3KPI กับความผิดพลาด 20/4.....	49
3.9 ข้อมูลสังเคราะห์ 2KPI กับความผิดพลาด 80/1.....	49
3.10 ข้อมูลสังเคราะห์ 3KPI กับความผิดพลาด 80/1.....	50
3.11 ชุดข้อมูล INTEL (อุณหภูมิที่อ่านได้).....	51
3.12 ชุดข้อมูล INTEL (ความชื้นที่อ่านได้).....	51
3.13 ชุดข้อมูล INTEL (แรงดันที่อ่านได้).....	52
3.14 ชุดข้อมูล NAMOS (แสงฟลูออเรสเซนซ์ที่อ่านได้).....	53
3.15 ชุดข้อมูล NAMOS (อุณหภูมิที่อ่านได้จากเซ็นเซอร์ 1).....	54
3.16 ชุดข้อมูล NAMOS (อุณหภูมิที่อ่านได้จากเซ็นเซอร์ 2).....	54
3.17 ชุดข้อมูล pdg2008 (อุณหภูมิโดยรอบที่อ่านได้).....	55
3.18 ชุดข้อมูล pdg2008 (อุณหภูมิพื้นผิวที่อ่านได้).....	56
3.19 ชุดข้อมูล pdg2008 (รังสีดวงอาทิตย์ที่อ่านได้).....	56
3.20 ROC ของข้อมูลสังเคราะห์ 2KPI เพิ่มด้วยความผิดปกติ 1/80.....	58
3.21 ROC ของชุดข้อมูล INTEL 1KPI (ประกอบด้วยความผิดปกติระยะสั้น).....	59
3.22 ROC ของข้อมูลสังเคราะห์ 2KPI เพิ่มด้วยความผิดปกติ 20/4.....	59
3.23 ROC ของชุดข้อมูล pdg2008 2KPI (ประกอบด้วยความผิดปกติจากสัญญาณรบกวน).....	60

สารบัญรูป (ต่อ)

รูปที่	หน้า
3.24 ROC ของชุดข้อมูลสังเคราะห์ 2KPI เพิ่มด้วยความผิดปกติ 80/1.....	60
3.25 ROC ของชุดข้อมูล NAMOS 1KPI (ประกอบด้วยความผิดปกติคงที่).....	61
3.26 ROC ของข้อมูลสังเคราะห์ 3KPI เพิ่มด้วยความผิดปกติ 1/80.....	63
3.27 ROC ของชุดข้อมูล INTEL 3KPI (ประกอบด้วยความผิดปกติระยะสั้น).....	64
3.28 ROC ของข้อมูลสังเคราะห์ 3KPI เพิ่มด้วยความผิดปกติ 20/4.....	64
3.29 ROC ของชุดข้อมูล pdg2008 3KPI (ประกอบด้วยความผิดปกติจากสัญญาณรบกวน).....	65
3.30 ROC ของชุดข้อมูลสังเคราะห์ 3KPI เพิ่มด้วยความผิดปกติ 80/1.....	65
3.31 ROC ของชุดข้อมูล NAMOS 3KPI (ประกอบด้วยความผิดปกติคงที่).....	66
3.32 เวลาในการคำนวณของแต่ละเทคนิคการบีบอัดข้อมูล.....	67
4.1 ชุดอุปกรณ์ต้นแบบ.....	68
4.2 โรงเรือนผู้ป่วยอินทรีชีวาภาพ มหาวิทยาลัยเทคโนโลยีสุรนารี.....	69
4.3 บอร์ด MIB 520CB (USB interface).....	70
4.4 โหนดเซ็นเซอร์ MPR2400 หรือ MicaZ.....	71
4.5 บอร์ดรวมข้อมูล MDA300.....	72
4.6 เซ็นเซอร์วัดอุณหภูมิ หรือ เทอร์โมคัปเปิล.....	72
4.7 เซ็นเซอร์วัดความชื้นในดิน (EC-5).....	73
4.8 ชุดข้อมูลจากโรงผู้ป่วยอินทรีชีวาภาพจำนวน.....	74
4.9 อัตราการตรวจจับสำหรับชุดข้อมูลจากโรงผู้ป่วยอินทรีชีวาภาพด้วยการสอน รูปแบบต่างๆ 50 ครั้ง.....	75
4.10 อัตราการตรวจจับสำหรับชุดข้อมูลจากโรงผู้ป่วยอินทรีชีวาภาพด้วยการสอน รูปแบบต่างๆ 10 ครั้ง.....	75
4.11 อัตราการตรวจจับสำหรับชุดข้อมูลจากโรงผู้ป่วยอินทรีชีวาภาพด้วยขนาดของ เครือข่ายประสาทเทียม 10x10.....	76

สารบัญตาราง

ตารางที่	หน้า
1.1 ความสามารถในการตรวจจับความผิดปกติในงานวิจัยที่ผ่านมา.....	7



คำอธิบายสัญลักษณ์และคำย่อ

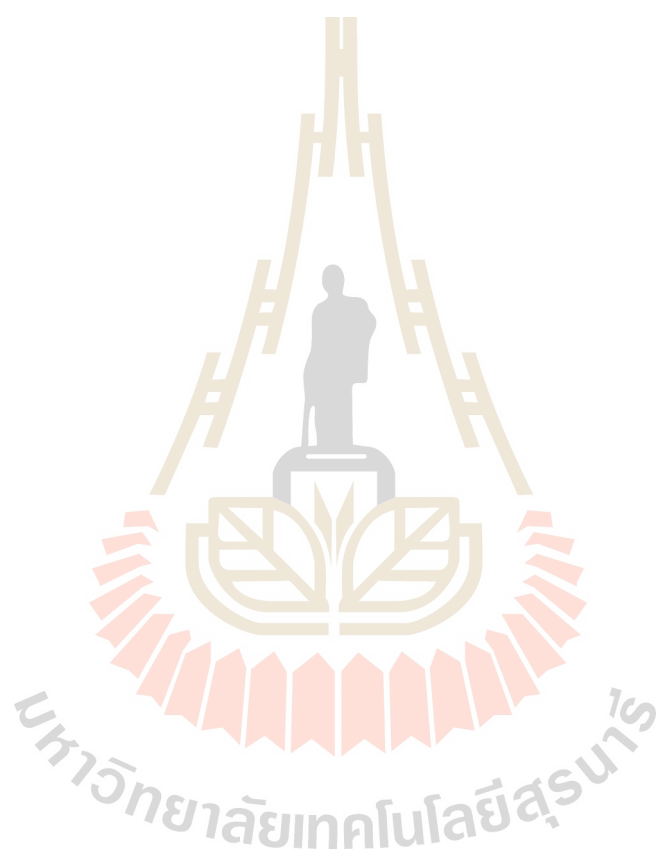
WSN	=	เครือข่ายเซ็นเซอร์ไร้สายทาง
DWT	=	การแปลงเวฟเล็ตแบบคิสคริต
LWT	=	การแปลงเวฟเล็ตแบบลิปท์ดิง
PCA	=	การวิเคราะห์องค์ประกอบหลัก
PC	=	องค์ประกอบหลัก
SOM	=	แผนผังจัดระเบียบตนเอง
OCSVM	=	ซัพพอร์ตเวกเตอร์แมชชีนแบบคลาสเดียว
KPI	=	ดัชนีชี้วัด
μ, ϖ	=	ดัชนีค่าที่สังเกต
n	=	จำนวนเวกเตอร์ข้อมูลในชุดข้อมูล
p, q	=	จำนวนชนิดของพารามิเตอร์ หรือ ดัชนีชี้วัด
X	=	ชุดข้อมูลป้อนเข้า
x^μ, x^ϖ	=	เวกเตอร์ของข้อมูลป้อนเข้า
x^{new}	=	เวกเตอร์ของข้อมูลป้อนเข้าชุดใหม่
BMU	=	หน่วยที่คล้ายคลึงกันที่สุด
i	=	นิวรอน
t	=	ดัชนีการวนรอบ
m_i	=	เวกเตอร์น้ำหนักของนิวรอน i
m_c	=	เวกเตอร์น้ำหนักของนิวรอน BMU
$\ \ $	=	ระยะ Euclidean
η_t	=	อัตราการเรียนรู้ที่ดัชนีเวลา t
$h_c(i, t)$	=	ฟังก์ชันเพื่อนบ้านที่นิวรอน i และเวลา t
$r_i(t)$	=	ฟังก์ชันตำแหน่งของนิวรอน i ที่เวลา t
$r_c(t)$	=	ฟังก์ชันตำแหน่งของนิวรอน BMU ที่เวลา t
σ	=	รัศมีของฟังก์ชันเพื่อนบ้าน

คำอธิบายสัญลักษณ์และคำย่อ (ต่อ)

$\phi(\cdot)$	=	ฟังก์ชันไม่เป็นเชิงเส้น
$X\phi$	=	เวกเตอร์อิมเมจ
$\phi(x^\mu)$	=	ฟังก์ชันไม่เป็นเชิงเส้นของเวกเตอร์อิมเมจ
R	=	รัศมีทรงกลม
δ_μ	=	ตัวแปรสเกลิก
v	=	พารามิเตอร์ควบคุม
$k(x^\mu, x^\omega)$	=	เคอร์เนลฟังก์ชัน
α_μ	=	ตัวคูณลากรางจ์
K	=	เมทริกซ์ของเคอร์เนล
K_c	=	เมทริกซ์ของเคอร์เนลกึ่งกลาง
\bar{X}	=	ค่าเฉลี่ยเลขคณิตของชุดข้อมูล
j	=	ระดับของเวฟเลต
h_0	=	ฟังก์ชันเวฟเลต
g_0	=	ฟังก์ชันสเกลลิ่ง
m	=	ดัชนีสเกลเชิงเวลา
f	=	ดัชนีสเกลเชิงเวลา
L	=	ความยาวของฟังก์ชันเวฟเลต หรือฟังก์ชันสเกลลิ่ง
a_j^{DWT}	=	สัมประสิทธิ์ DWT แบบหยาบในระดับ j
a_j^{LWT}	=	สัมประสิทธิ์ LWT แบบหยาบในระดับ j
d_j^{DWT}	=	สัมประสิทธิ์ DWT แบบละเอียดในระดับ j
d_j^{LWT}	=	สัมประสิทธิ์ LWT แบบละเอียดในระดับ j
a	=	จำนวนความผิดพลาดต่อชุดข้อมูล
s	=	จำนวนชุดข้อมูล
FPR	=	อัตราผลบวกคลวง

คำอธิบายสัญลักษณ์และคำย่อ (ต่อ)

DR	=	อัตราผลบวกจริง
LP	=	สัมประสิทธิ์ความถี่ผ่านต่ำ
HP	=	สัมประสิทธิ์ความถี่ผ่านสูง



บทที่ 1

บทนำ

1.1 ความเป็นมาของปัญหา

ปัจจุบันมีการประยุกต์ใช้งานเทคโนโลยีการสื่อสารไร้สายเพื่อการเกษตรเป็นจำนวนมาก เช่น การใช้เครือข่ายเซ็นเซอร์ไร้สายเพื่อเฝ้าระวังทางการเกษตร โดยอาศัยการเก็บข้อมูลทางกายภาพภายในแปลงเกษตรจากโหนดเซ็นเซอร์ไร้สายซึ่งถูกวางในตำแหน่งต่างๆกันในไร่หรือแปลงเกษตร และส่งข้อมูลกลับมายังสถานีฐานผ่านการสื่อสารวิทยุระยะสั้น ในกรณีที่แปลงเกษตรมีขนาดใหญ่ ข้อมูลจากโหนดเซ็นเซอร์ไร้สายจะถูกส่งผ่านโหนดเซ็นเซอร์ข้างเคียง ซึ่งจะทำการส่งต่อข้อมูลภายในเครือข่ายในลักษณะนี้จนข้อมูลถูกนำส่งถึงสถานีฐาน เครือข่ายเซ็นเซอร์ไร้สายประกอบด้วยโหนดเซ็นเซอร์หลายโหนดซึ่งมีขนาดเล็กและราคาถูก มีหน่วยประมวลผล (CPU) และหน่วยความจำขนาดเล็ก ทั้งยังมีแหล่งจ่ายพลังงานที่จำกัดอีกด้วย ข้อจำกัดดังกล่าวส่งผลทำให้เกิดความผิดพลาดระหว่างการส่งข้อมูลและก่อให้เกิดความผิดปกติในชุดข้อมูลที่วัดได้ด้วย (Goh, et al., 2007)

ความผิดปกติของข้อมูล (anomaly) ในชุดข้อมูล โดยนิยามคือ ค่าของข้อมูลที่สังเกตได้โดยมีลักษณะแตกต่างจากชุดข้อมูลอื่นๆ (Rajasegarar, et al., 2008) ความผิดปกติในข้อมูลนั้นอาจเกิดจากปรากฏการณ์ซึ่งผิดปกติในพื้นที่เฝ้าระวังนั้นๆ อาจส่งผลกระทบต่อผลผลิตทางการเกษตรนั้นได้ เมื่อเกิดความผิดปกติในพื้นที่เฝ้าระวัง พืชผลทางการเกษตรบางชนิด เช่น แปลงผักไฮโดรโปนิกส์ (hydroponics farm) ซึ่งจำเป็นต้องอาศัยความเป็นกรดต่าง (pH) ที่คงที่ของสารละลายอาหาร หรือกระทั่งการเพาะเลี้ยงสัตว์น้ำ ซึ่งอาศัยการวัดค่าปริมาณออกซิเจนที่ละลายในน้ำ (dissolved oxygen) ล้วนแล้วแต่ได้รับผลกระทบทั้งสิ้น หากสถานการณ์หรือปัญหาไม่ได้รับการแก้ไขอย่างทันท่วงที โดยเฉพาะอย่างยิ่งพืชผลทางการเกษตรแม่นยำ (precision agricultural) ซึ่งต้องเจริญเติบโตในสภาพแวดล้อมที่แม่นยำ และคุณภาพผลผลิตแปรเปลี่ยนได้ง่าย หากสภาพพื้นที่เพาะปลูกเปลี่ยนแปลงแม้เพียงเล็กน้อย

1.1.1 เทคนิคการตรวจจับความผิดปกติ (outlier)

โดยทั่วไปกลไกการตรวจจับความผิดปกติมักจะแบ่งออกเป็น 3 กลุ่มตามลักษณะชนิดของข้อมูลที่สนใจ คือ กลุ่มที่ 1 วิธีการหา outlier โดยไม่ใช้ความรู้เบื้องต้น (prior knowledge) ของข้อมูล อันได้แก่กลุ่มที่ใช้เทคนิคการตรวจจับความผิดปกติเชิงสถิติแบบพารามेटริก (parametric statistical anomaly detection) กลุ่มที่ 2 เป็นกลุ่มที่ใช้การจำแนกซึ่งจะถูกสอนให้รู้จักรูปแบบข้อมูลที่ผิดปกติ กลุ่มที่ 3 เป็นกลุ่มที่เป็นกลไกการตรวจจับแบบกึ่งสอนจำ (semi-supervised recognition) (Rajasegarar, et al., 2008) โดยในกลุ่มที่ 2 และ 3 นั้นจัดเป็นเทคนิคการตรวจจับความผิดปกติประเภทที่ไม่ใช้พารามิเตอร์ (nonparametric anomaly detection techniques)

1.1.1.1 เทคนิคการตรวจจับความผิดปกติเชิงสถิติแบบพารามेटริก

เทคนิคการตรวจจับความผิดปกติเชิงสถิติแบบพารามेटริก (parametric statistical anomaly detection techniques) นั้นมักตั้งสมมติฐานว่าข้อมูลซึ่งเป็นข้อมูลปกติมักถูกสังเคราะห์มาจากการแจกแจงแบบพารามेटริก (Rajasegarar, et al., 2008) ดังนั้นความหนาแน่นความน่าจะเป็นของข้อมูลจึงมักทราบล่วงหน้า โดยทราบค่าพารามิเตอร์ทางสถิติของข้อมูล ดังนั้นค่าเฉลี่ย (mean) ค่าความแปรปรวน (variance) ค่าฟังก์ชันความหนาแน่นความน่าจะเป็นจึงถูกคำนวณมาจากชุดข้อมูลที่ปกติดังกล่าว ส่วนความผิดปกติของข้อมูล (outlier) นั้นถูกตรวจจับโดยอาศัยการตรวจจับจากจุดข้อมูลซึ่งมีความเป็นไปได้ต่ำภายใต้สมมติฐานการแจกแจงของข้อมูลนั้น เช่น การทดสอบเชิงสถิติแบบไคแอสควร์ (Chi-Square Test Statistical Method) ถูกประยุกต์ใช้ในการตรวจจับการโจมตีของแหล่งรวมข้อมูล (sinkhole attacks) ในเครือข่ายเซ็นเซอร์ไร้สาย (Rajasegarar, et al., 2008) วิธีการใช้แบบจำลองเกาส์เซียน (Gaussian Model Based Method) ถูกประยุกต์ใช้ในการทดสอบชุดข้อมูลซึ่งมีลักษณะการแจกแจงแบบเกาส์เซียน โดยการใช้ประมาณค่าเฉลี่ยในการคำนวณค่าแอมพลิจูดความผิดปกติ (Chandola, et al., 2009) วิธีการประมาณค่าแบบเชิงเส้น (Linear Least – Squares Estimation Method, LLSE) นำค่าเฉลี่ยและค่าความแปรปรวนที่คำนวณได้จากชุดข้อมูลจากเซ็นเซอร์รวมทั้งค่าความแปรปรวนร่วม (covariance) ระหว่างข้อมูลเซ็นเซอร์ต่างๆมาพิจารณาในการตรวจจับความผิดปกติในชุดข้อมูล (Sharma, Golubchik, et al., 2010)

อย่างไรก็ดี เทคนิคการตรวจจับความผิดปกติเชิงสถิติเหล่านี้มักใช้งานได้ดีเมื่อค่าพารามิเตอร์ทางสถิติของข้อมูลที่วัดได้มีความถูกต้องหรือข้อมูลมีปริมาณมาก

เพียงพอ ดังนั้นประสิทธิภาพการตรวจจับด้วยวิธีการนี้จึงขึ้นอยู่กับข้อมูลและการประยุกต์ใช้งานเป็นหลัก หากเซ็นเซอร์วัดข้อมูลที่มีความผันผวนค่อนข้างสูง หรือมีการเปลี่ยนแปลงเชิงเวลาด้วยเทคนิคการตรวจจับวิธีการนี้จึงไม่เหมาะสมในการใช้งานในเครือข่ายเซ็นเซอร์ไร้สาย (Rajasegarar, et al., 2008; Chandola, et al., 2009)

1.1.1.2 เทคนิคการตรวจจับความผิดปกติแบบที่ไม่ใช้พารามิเตอร์

เทคนิคการตรวจจับความผิดปกติแบบที่ไม่ใช้พารามิเตอร์ (Nonparametric anomaly detection techniques) ไม่ได้อาศัยค่าทางสถิติจากชุดข้อมูลที่วัดได้จากเซ็นเซอร์ ดังนั้นเทคนิคเหล่านี้จึงเหมาะแก่การประยุกต์ใช้งานในเครือข่ายเซ็นเซอร์ไร้สายที่มีความจำกัดด้านทรัพยากร และมักถูกติดตั้งในสภาพแวดล้อมที่มีการเปลี่ยนแปลงของการแจกแจงข้อมูลอยู่ด้วย (Rajasegarar, et al., 2008; Chandola, et al., 2009) เช่น วิธีการใช้ฮิสโตแกรม (Histogram method) เป็นวิธีการแบบที่ไม่ใช้พารามิเตอร์ที่ง่ายที่สุด ซึ่งทำการแบ่งข้อมูลเชิงเวลาจากเซ็นเซอร์ออกเป็นช่วงๆ N ช่วงตามค่าของข้อมูลเพื่อที่จะหาขอบแบ่ง (threshold) ของความผิดปกติของชุดข้อมูลสำหรับประสิทธิภาพของวิธีการนี้มักขึ้นอยู่กับค่าของ N ซึ่งใช้แบ่งข้อมูลแบบปกติมากำหนดค่าขอบแบ่งตามการเปลี่ยนแปลงในสภาพแวดล้อม และขอบเขตความเหมาะสมต่อการใช้งานตรวจจับความผิดปกติ (Rajasegarar, et al., 2008)

วิธีการโดยใช้ความหนาแน่น (Density Based Methods) ในการตรวจจับความผิดปกติของข้อมูลแบบกระจาย (distributed anomaly detection) วิธีการนี้มีประเด็นเปิดคือ การระบุจำนวนที่เหมาะสมของโหนดย่อย (children node) ต่อหนึ่งโหนดหลัก (parent node) โดยพิจารณาปริมาณการคำนวณและการสื่อสาร (computational and communication load) ที่เกิดขึ้นที่โหนดหลักต่ออายุของเครือข่าย (Rajasegarar, et al., 2008)

วิธีการโดยใช้การรวมกลุ่มคลัสเตอร์ข้อมูล (Data Clustering Based Methods) เช่น K-Nearest Neighbor (KNN) ซึ่งถูกนำเสนอใน (Chandola, et al., 2009; Kiziloren, et al., 2009; Rajasegarar, et al., 2006; Yao, et al., 2010) วิธีการนี้ทำการรวมกลุ่มของข้อมูลที่มีความคล้ายคลึงกันเข้าในคลัสเตอร์ (cluster) เดียวกันที่มีขนาดเท่าๆกัน และหาค่าเฉลี่ยของระยะระหว่างคลัสเตอร์ โดยประสิทธิภาพของวิธีการนี้ขึ้นอยู่กับขนาดของคลัสเตอร์อย่างยิ่ง และสามารถระบุความผิดปกติของข้อมูลในระดับชุดข้อมูล (data set) ได้เท่านั้น แต่ไม่สามารถระบุตำแหน่งหรือค่าข้อมูลที่ผิดปกติภายในชุดข้อมูลได้ ดังนั้นวิธีการนี้จึงไม่เหมาะแก่การตรวจจับความผิดปกติในชุด

ข้อมูลที่มีความผิดปกติในระยะยาว (long term anomaly) เนื่องจากไม่ได้พิจารณาสหสัมพันธ์เชิงเวลา (temporal correlation) ของชุดข้อมูล (Yao, et al., 2010)

วิธีการแบบผลรวมสะสม (Cumulative Summation หรือ CUSUM) สามารถตรวจจับการเปลี่ยนแปลงของค่าเฉลี่ย (mean) ค่าความแปรปรวน (variance) และค่าความแปรปรวนร่วม (covariance) ได้โดยไม่ต้องกำหนดสมมติฐานเกี่ยวกับการแจกแจงข้อมูล (Yao, et al., 2010) วิธีการนี้มีการคำนวณที่ง่ายแต่ต้องอาศัยระยะเวลาสอนจำ (training phase) ที่ยาว เนื่องจากวิธีการนี้พิจารณาคุณลักษณะ (feature) ของข้อมูลแยกกัน (Rajasegarar, et al., 2008)

งานวิจัยที่ผ่านมาทำการตรวจจับความผิดปกติโดยการอาศัยการเรียนรู้คุณลักษณะของความผิดปกติของข้อมูลล่วงหน้า (offline training) ยังมีงานวิจัยอีกกลุ่มหนึ่งที่ศึกษาวิธีการตรวจจับความผิดปกติของข้อมูลแบบเวลาจริง (real time) ซึ่งเป็นวิธีการวิเคราะห์คุณลักษณะเชิงเวลา (Time series analysis based methods) เช่น Autoregressive Integrated Moving Average (ARIMA) (Chandola, et al., 2009; Sharma, et al., 2010; Yao, et al., 2010) ซึ่งเป็นวิธีการมาตรฐานในการจำลองและคาดคะเนค่าของข้อมูลที่มีความเป็นคาบ (time series data with periodicity) ค่าสหสัมพันธ์เชิงเวลาของข้อมูลมาจากเซ็นเซอร์จะถูกใช้ในการสร้างแบบจำลอง ARIMA ในการตรวจจับความผิดปกติของข้อมูลนั้น ค่าของข้อมูลชุดใหม่จากเซ็นเซอร์จะถูกเปรียบเทียบกับค่าที่คาดคะเนจากแบบจำลอง หากเกิดความคลาดเคลื่อนเกินกว่าขนาดที่กำหนดไว้ ข้อมูลใหม่ที่ได้จะถูกจำแนกเป็นข้อมูลที่ผิดปกติ อย่างไรก็ตาม ARIMA มีประสิทธิภาพต่ำเมื่อใช้ในการตรวจจับความผิดปกติที่มีระยะเวลายาว อีกวิธีการหนึ่งคือ Segment Sequence Analysis (SSA) ซึ่งเหมาะสำหรับข้อมูลเซ็นเซอร์ที่เป็นคาบและมีสหสัมพันธ์เชิงเวลา โดยวิธีการ SSA นี้สามารถตรวจจับความผิดปกติระยะยาวได้ดี โดยเฉพาะเมื่อมีการนำ SSA ร่วมกับวิธีการตั้งค่า rule-based method แล้วพบว่าความสามารถตรวจจับความผิดปกติทั้งที่มีระยะสั้นและยาวได้ (Short and long term anomalies) ที่ยังต้องการปริมาณการคำนวณและหน่วยความจำต่ำอีกด้วย จึงเหมาะแก่การประยุกต์ใช้งานในเครือข่ายเซ็นเซอร์ไร้สาย (Yao et al., 2010)

วิธีการแบบเคอร์เนล (Kernel-based methods) เช่น Support Vector Machine (SVM) เป็นวิธีการหนึ่งที่ถูกใช้ในการจำแนกข้อมูล (data classification) และตรวจจับความผิดปกติ (Hsu, et al., 2003), (Du, et al., 2010), (Lutsa, et al., 2010) โดยหลักการของวิธีการนี้คือการเชื่อมโยงข้อมูลไปยังปริภูมิหลายมิติ (high dimensional space) เรียกว่า feature โดยใช้ kernel

function (Lutsa, et al., 2010) ดังนั้น kernel function จึงเป็นส่วนสำคัญยิ่งในการตรวจจับความผิดปกติด้วยในกระบวนการ SVM งานวิจัยที่ผ่านมามีการประยุกต์ใช้ kernel function หลายประเภท เช่น เชิงเส้น (linear), โพลีโนเมียล (polynomial), เกาส์เซียน (Gaussian), Radial Based function (RBF), Sigmoid shaped function, (Du, et al., 2010) สำหรับจำแนกข้อมูล และมีงานวิจัยที่ประยุกต์ใช้วิธีการ SVM เรียกว่า One-class Support Vector Machine (OSVM) ในการตรวจจับความผิดปกติในข้อมูลด้วย (F. Wang, et al., 2010) นอกจากนี้งานวิจัยหลายงานยังใช้วิธีการ OSVM ในการตรวจจับความผิดปกติในชุดข้อมูลจริงจากเครือข่ายเซ็นเซอร์ไร้สายอีกด้วย (Laskov, et al., 2004), (Rajasegarar et al., 2007), (Meratnia, et al, 2009) โดยเชื่อมโยงข้อมูลที่ปกติจากเซ็นเซอร์เพื่อแปลงไปสู่ quarter sphere ของปริภูมิคุณลักษณะ (feature space) ซึ่งถูกปรับเปลี่ยนตามสภาพแวดล้อมได้ อย่างไรก็ตามอัลกอริทึม OSVM มี communication overhead เล็กน้อยและเหมาะสำหรับการใช้งานในเครือข่ายเซ็นเซอร์ที่มีการติดตั้งในสภาพแวดล้อมเดียวกัน (homogenous environment) ซึ่งทำให้การแจกแจงของข้อมูลที่แต่ละโหนดเซ็นเซอร์มีลักษณะเหมือนกันและเหมาะสมต่อการประยุกต์การใช้งานแบบเวลาจริง (online application) นอกจากนี้อัลกอริทึม OSVM ยังมีความยืดหยุ่นต่อการเปลี่ยนแปลงของชุดข้อมูลที่มีความปกติจากการพิจารณาคุณลักษณะอีกด้วย อย่างไรก็ตามเมื่อเปรียบเทียบความซับซ้อนของการคำนวณแล้วพบว่าอัลกอริทึม OSVM นั้นใช้การคำนวณที่ซับซ้อนกว่าวิธีการ Rule-based และอัลกอริทึม CUSUM

วิธีการ learning-based เป็นวิธีการกลุ่มสุดท้ายที่จัดอยู่ในกลุ่มเทคนิคการตรวจจับความผิดปกติของข้อมูลแบบที่ไม่ใช่พารามิเตอร์ วิธีการนี้อาศัยการเรียนรู้พารามิเตอร์ของแบบจำลองเช่น Hidden Markov Model (HMM) จากการสุ่มตัวอย่างของข้อมูลจากเซ็นเซอร์ นอกจาก HMM แล้วยังมีอีกวิธีการหนึ่งซึ่งให้ผลตรวจจับที่ดีกว่า HMM เรียกว่า Self-Organizing Map (SOM) (Min, et al., 2009) SOM เป็นแบบจำลองนิเวศ (neural network) ซึ่งใช้ในการวิเคราะห์และมองภาพข้อมูลจำนวนหลายมิติเพื่อแปลงเป็นสองมิติ (two-dimensional lattices) (F. Wang, et al., 2010) และมักถูกใช้ในการจำแนกข้อมูล (Doshi, et al., 2007) โดยในงานวิจัย (F. Wang, et al., 2010) อัลกอริทึม SOM ถูกใช้งานร่วมกับอัลกอริทึม OSVM เพื่อตรวจจับความผิดปกติของข้อมูล ในงานวิจัย (Siripanadorn, et al., 2010) อัลกอริทึม SOM ถูกใช้เพื่อตรวจจับความผิดปกติของข้อมูลการวัดเซ็นเซอร์ซึ่งถูกบีบอัดโดยการแปลงเวฟเลตแบบเต็มหน่วย (Discrete Wavelet Transform, DWT) อย่างไรก็ตามอัลกอริทึม SOM จะใช้เวลาในการสอนให้รู้จักรูปแบบของข้อมูลแปรผันตามปริมาณของ

ข้อมูลที่ใช้ในการสอน ในงาน (Kiziloren, et al., 2009) อัลกอริทึม SOM ถูกใช้ในการตรวจจับความผิดปกติในกราฟฟิคเครือข่าย ซึ่งถูกบีบอัดโดยหลักการวิเคราะห์องค์ประกอบหลัก (principal component analysis, PCA)

ดังนั้นงานวิจัยนี้จึงมุ่งเน้นที่จะตรวจจับความผิดปกติในข้อมูลซึ่งถูกบีบอัดเพื่อให้ได้ประสิทธิภาพสูงยิ่งขึ้น ทั้งยังหาวิธีบีบอัดข้อมูลในเครือข่ายเซ็นเซอร์ไร้สายเพื่อหาวิธีการบีบอัดข้อมูลและตรวจจับความผิดปกติที่เหมาะสมสำหรับเครือข่ายเซ็นเซอร์ไร้สาย โดยเฉพาะอย่างยิ่งสำหรับการประยุกต์ใช้งานด้านเกษตร



ตารางที่ 1.1 ความสามารถของการตรวจจับความผิดปกติในงานวิจัยที่ผ่านมา

Previous anomaly detection works		Ability to detect faults			Adaptive to dynamic	Used with data compression
		Short	Noise	Constant		
Parametric Statistical Anomaly Detection Technique						
1	Chi-Square Test Statistic Method	-	-	-	x	x
2	Gaussian Model Based Method	-	-	-	x	x
3	Linear Least – Squares Estimation (LLSE)	✓	✓	x	x	x
Nonparametric Anomaly Detection Techniques						
4	Rule-Based method • Histogram Method	✓	✓	✓	x	x
5	Density Based Method	x	✓	✓	✓	x
6	Data Clustering Based Method • K-Nearest Neighbor (KNN)	✓	x	x	✓	x
7	Cumulative Summation (CUSUM)	x	✓	x	✓	x
8	Time Series Analysis Based Method • Autoregressive Integrated Moving Average (ARIMA) • Segment Sequence Analysis (SSA)	✓	x	x	✓	x
9	Kernel-based method • Support Vector Machine (SVM) • One-Class Support Vector Machine (OCSVM)	-	-	-	x	x
10	Learning Based Method • Hidden Markov Model (HMM) • Self-Organizing Map (SOM)	✓	✓	x	✓	x
		✓	x	✓	✓	✓

1.1.2 เทคนิคการบีบอัดข้อมูล

ในเครือข่ายเซ็นเซอร์ไร้สายซึ่งมีความจำกัดทางด้านพลังงาน, ความสามารถในการประมวลผลและแบนด์วิธ (bandwidth) นั้น การบีบอัดข้อมูลจะสามารถทำให้ลดปริมาณการส่งข้อมูลในเครือข่ายลงได้ เพื่อให้ผู้ใช้ทรัพยากรที่มีอยู่อย่างจำกัดในเครือข่ายให้มีประสิทธิภาพยิ่งขึ้น อีกทั้งยังดึงคุณลักษณะเด่นของข้อมูลให้เด่นยิ่งขึ้นเพื่อให้ตรวจจับความผิดปกติของข้อมูลได้ดียิ่งขึ้น

งานวิจัยแรกเริ่มที่ศึกษาการบีบอัดข้อมูลได้ทำการเปรียบเทียบอัลกอริทึมบีบอัดข้อมูล 4 วิธีการด้วยกันได้แก่ การแปลงโคไซน์แบบเต็มหน่วย (Discrete Cosine Transform, DCT), การแปลงฟูเรียร์แบบเต็มหน่วย (Discrete Fourier Transform, DFT), การแปลงเวฟเล็ตแบบเต็มหน่วย และการประมาณค่าเวกเตอร์ (Vector Quantization, VQ) (Watson, et al., 1995) โดยพบว่าอัลกอริทึม VQ สามารถบีบอัดได้ดีกว่าวิธีการอื่นๆ ต่อมาอัลกอริทึม VQ ถูกพัฒนาเป็น Adaptive Learning Vector Quantization (ALVQ) และถูกประยุกต์ใช้ในงานที่แบนด์วิธมีการเปลี่ยนแปลง (dynamic bandwidth application) (S. Lin et al., 2005) อย่างไรก็ตามทั้งอัลกอริทึม VQ และอัลกอริทึม ALVQ มักถูกใช้ในการบีบอัดข้อมูลที่เป็นภาพมากกว่าที่จะลดปริมาณข้อมูลทุกๆ ไป

ต่อมางานวิจัย (Marcelloni, et al., 2008) ได้นำเสนอวิธีการบีบอัดข้อมูลแบบการไม่มีสูญเสีย (lossless) ในเครือข่ายเซ็นเซอร์ไร้สายเรียกว่า Sensor Lempel Ziv Welch (S-LZW) และงานวิจัย (Capo-Chichi, et al., 2009) นำเสนอวิธีการบีบอัดข้อมูลที่เรียกว่า K-Run Length Encoding (K-RLE) ซึ่งให้อัตราการบีบอัดข้อมูลดีกว่าอัลกอริทึม S-LZW แต่ใช้พลังงานสูงกว่าอัลกอริทึม S-LZW

ต่อมางานวิจัย (Sharma, et al., 2010) นำเสนอวิธีการบีบอัดข้อมูลที่เรียกว่า หลักการวิเคราะห์องค์ประกอบหลัก (principal component analysis, PCA) ซึ่งเป็นวิธีการเชิงสถิติสำหรับแปลงคุณลักษณะ (attributes) ของชุดข้อมูลให้เป็นคุณลักษณะที่ไม่สัมพันธ์ (uncorrelated attributes) เรียกว่า principle components (PCs)

นอกจาก PCA แล้วยังมีวิธีการบีบอัดข้อมูลโดยอาศัยหลักการแปลงข้อมูลเป็นเวฟเล็ตเช่น Data Aggregation based on Wavelet Entropy (DAWE) (Bruce, et al., 2002) และการแปลงเวฟเล็ตแบบเต็มหน่วย (X. L. Li, et al., 2009) และ (Xu, et al., 2010) เนื่องจาก DWT นอกจากจะสามารถบีบอัดข้อมูลแล้วยังสามารถดึงคุณลักษณะที่สำคัญของข้อมูลออกมาจากชุดข้อมูลเพื่อใช้ในการจำแนกข้อมูลและตรวจจับความผิดปกติของข้อมูลได้อีกด้วย ในงานวิจัย (Siripanadorn, et al.,

2010a, 2010b) ใช้ DWT บีบอัดข้อมูลและใช้อัลกอริทึม SOM เพื่อตรวจจับความผิดปกติของข้อมูล ซึ่งเก็บจากเครือข่ายเซ็นเซอร์ไร้สาย นอกจากนี้ยังมีงานวิจัยที่นำเสนอ Lifting Wavelet Transform (LWT) มาใช้ในการบีบอัดข้อมูล และพบว่าให้ประสิทธิภาพดีในแง่ของการประหยัดพลังงานและบีบอัดข้อมูลด้วย (Ciancio, et al., 2005), (Manjunath, et al., 2010) อย่างไรก็ตามการนำ LWT มาใช้ร่วมกับเทคนิคตรวจจับความผิดปกติของข้อมูลยังไม่ได้มีการวิจัยเปรียบเทียบประสิทธิภาพการตรวจจับข้อมูล

ดังนั้นงานวิจัยนี้ จึงมีจุดประสงค์ที่จะศึกษาและเปรียบเทียบวิธีการบีบอัดข้อมูล ร่วมกับการตรวจจับความผิดปกติของข้อมูล ซึ่งเหมาะแก่การใช้งานในการเฝ้าระวังสภาพแวดล้อมด้วยเครือข่ายเซ็นเซอร์ไร้สายซึ่งมีทรัพยากรพลังงานที่จำกัด โดยมีสมมติฐานว่าวิธีการบีบอัดข้อมูลที่สามารถดึงคุณลักษณะสำคัญของชุดข้อมูลจากเซ็นเซอร์ออกมาได้ จะทำให้สามารถตรวจจับความผิดปกติของข้อมูลได้ดียิ่งขึ้น โดยพิจารณาวิธีการบีบอัดข้อมูลแบบ LWT, DWT และ PCA ร่วมกับวิธีการบีบอัดข้อมูลแบบ OCSVM และ SOM เพื่อหากระบวนการบีบอัดข้อมูลและตรวจจับความผิดปกติที่เหมาะสมสำหรับการใช้งานในเครือข่ายเซ็นเซอร์ไร้สาย โดยเฉพาะการเฝ้าระวังด้านการเกษตร

1.2 จุดประสงค์งานวิจัย

เพื่อสร้างรูปแบบโรงผลิตปุ๋ยอินทรีย์ชีวภาพที่มีคุณภาพได้มาตรฐาน สำหรับธุรกิจขนาดกลางและขนาดใหญ่ หรือธุรกิจการส่งออก เพื่อสร้างรายได้ให้แก่ประเทศและตอบสนองนโยบายรัฐบาลในการผลิตพืชเกษตรอินทรีย์โดยมีวัตถุประสงค์จำเพาะดังนี้

- 1.2.1 เพื่อศึกษาประสิทธิภาพในการตรวจจับความผิดปกติในข้อมูลและการบีบอัดข้อมูลที่เหมาะสมในเครือข่ายเซ็นเซอร์ไร้สาย
- 1.2.2 เพื่อค้นหารูปแบบผสมของการทำงานที่เหมาะสมระหว่างวิธีการบีบอัดข้อมูลและการตรวจจับความผิดปกติในเครือข่ายเซ็นเซอร์ไร้สาย
- 1.2.3 เพื่อพัฒนาอุปกรณ์ระบบต้นแบบสำหรับบีบอัดข้อมูล และตรวจจับความผิดปกติของข้อมูลเซ็นเซอร์ในเครือข่ายเซ็นเซอร์ไร้สายเพื่อเฝ้าระวังด้านการเกษตร
- 1.2.4 เพื่อให้ได้กระบวนการตรวจจับความผิดปกติของข้อมูลจากเครือข่ายเซ็นเซอร์ไร้สายสำหรับเฝ้าระวังโรงผลิตปุ๋ยอินทรีย์ชีวภาพ (bioorganic fertilizer production plant monitoring)

1.3 ขอบเขตของโครงการวิจัย

ขอบเขตของการวิจัยจะครอบคลุมถึงการสร้างระบบเครือข่ายเซ็นเซอร์ไร้สายต้นแบบที่ใช้ได้จริงสำหรับการเฝ้าระวังโรงผลิตปุ๋ยอินทรีย์ชีวภาพ เพื่อควบคุมคุณภาพในการผลิตให้ได้มาตรฐาน และสร้างกระบวนการตรวจจับความผิดปกติของข้อมูลจากเครือข่ายเซ็นเซอร์ไร้สาย ซึ่งจะนำไปสู่แนวทางในการพัฒนาเทคโนโลยีการควบคุมคุณภาพในการผลิตปุ๋ยอินทรีย์ชีวภาพ ที่สามารถควบคุมคุณภาพของผลผลิตได้อย่างมีประสิทธิภาพ

1.4 ประโยชน์ที่คาดว่าจะได้รับ

1.4.1 แก้ปัญหาในการดำเนินงานของหน่วยงานที่ทำการวิจัย:

ระบบเครือข่ายเซ็นเซอร์ไร้สาย (wireless sensor network) เพื่อเฝ้าระวังโรงผลิตปุ๋ยอินทรีย์ชีวภาพจะเป็นเครือข่ายต้นแบบในการพัฒนาระบบตรวจวัด และควบคุม (detection and control) ด้านการเกษตรอื่นๆ ที่มหาวิทยาลัยเทคโนโลยีสุรนารี โดยทำให้นักวิจัยทราบถึงการเปลี่ยนแปลงที่เกิดขึ้นได้รวดเร็ว และมีการเก็บข้อมูลเพื่อการวิเคราะห์ภายหลัง ซึ่งจะช่วยให้การวิจัยและพัฒนากระบวนการผลิต และวิเคราะห์แก้ปัญหาได้สะดวกและรวดเร็วขึ้น

1.4.2 เป็นองค์ความรู้ที่เป็นประโยชน์ในการวิจัยต่อไป:

กลไกการทำงานภายในเครือข่ายเซ็นเซอร์ไร้สายที่สามารถตรวจจับความผิดปกติของข้อมูลได้อย่างมีประสิทธิภาพ สามารถเผยแพร่และพัฒนาต่อเพื่อเชื่อมโยงสถานีฐานเข้ากับเครือข่ายโทรศัพท์เคลื่อนที่เพื่อส่งข้อมูลที่ผิดปกติมายังผู้ควบคุมดูแลหรือเกษตรกร ทำให้ทราบถึงความผิดปกติของข้อมูลได้อย่างรวดเร็วโดยไม่ต้องเฝ้าประจำอยู่ที่สถานีฐาน นอกจากนี้ ยังสามารถขยายผลต่อไปเพื่อการรวบรวมข้อมูลที่ต้องการ สำหรับช่วยคาดการณ์คุณภาพของผลผลิตได้แม่นยำมากขึ้น

1.4.3 บริการความรู้แก่ภาคธุรกิจ:

ระบบเครือข่ายเซ็นเซอร์ไร้สายที่มหาวิทยาลัยเทคโนโลยีสุรนารี จะเป็นเครือข่ายต้นแบบ ซึ่งคาดว่าจะนำไปสู่การเป็นสถานีตัวอย่าง หรือ ศูนย์วิจัยพิเศษเฉพาะด้านที่มีมูลค่าสูง ช่วยให้นักวิชาการเกษตรสามารถวิจัยและพัฒนาได้สะดวกและรวดเร็วยิ่งขึ้น ส่งผลให้เกิดการพัฒนากระบวนการผลิตและวิเคราะห์แก้ปัญหาด้านคุณภาพของผลผลิตทางการเกษตร ซึ่งเป็นประโยชน์ต่อเกษตรกรและภาคธุรกิจทั้งทางตรงและทางอ้อม

1.4.4 เพิ่มคุณภาพในการผลิตและได้มาตรฐาน:

การพัฒนาระบบเครือข่ายเซ็นเซอร์ไร้สายสำหรับเฟ้าระวังโรงผลิตปุ๋ยอินทรีย์ชีวภาพ จะช่วยให้สามารถตรวจจับความผิดปกติในกระบวนการผลิตและแจ้งเตือนได้รวดเร็วขึ้น ส่งผลให้สามารถเพิ่มคุณภาพในการผลิต อีกทั้งยังสามารถควบคุมคุณภาพให้คงที่และได้มาตรฐานอีกด้วย

1.4.5 เป็นประโยชน์ต่อประชากรกลุ่มเป้าหมาย:

- นักวิชาการเกษตร ระบบเครือข่ายเซ็นเซอร์ไร้สายที่สามารถตรวจจับความผิดปกติได้อย่างมีประสิทธิภาพ จะเป็นประโยชน์ต่อการสามารถวิจัยและพัฒนาวิเคราะห์ปัญหาและพัฒนากระบวนการควบคุมคุณภาพการผลิตที่มีประสิทธิภาพ และได้มาตรฐาน
- ผู้ประกอบอาชีพเกษตรกร จะได้รับประโยชน์ทั้งทางตรงจากการใช้ระบบได้จริง และทางอ้อม อันได้แก่จากองค์ความรู้ที่เกิดจากการวิจัยและพัฒนาจากทางนักวิชาการเกษตรที่ใช้ระบบนี้พัฒนากระบวนการควบคุมคุณภาพการผลิตปุ๋ย
- หน่วยงานภาครัฐที่จะนำผลการวิจัยไปใช้ประโยชน์จากองค์ความรู้เพิ่มเติมจากการพัฒนาเครือข่ายเซ็นเซอร์ไร้สายที่สามารถตรวจจับความผิดปกติได้อย่างมีประสิทธิภาพ ได้แก่ NECTEC ซึ่งเป็นผู้ให้การสนับสนุนการวิจัยภายใต้ชุดโครงการระบบสมองฝังตัว (Cluster 05 Embedded Systems)

บทที่ 2

วิธีการแปลงเวฟเล็ตแบบเต็มหน่วยและเทคนิคซัพพอร์ตเวกเตอร์แมชชีนคลาสหนึ่ง สำหรับการตรวจจับความผิดปกติในเครือข่ายเซ็นเซอร์ไร้สาย

การอ่านข้อมูลจากเครือข่ายเซ็นเซอร์ไร้สายอาจเกิดความผิดปกติได้เนื่องมาจากการตรวจพบปรากฏการณ์ที่ผิดปกติ พลังงานจากแบตเตอรี่ที่มีอยู่อย่างจำกัด การทำงานผิดพลาดของอุปกรณ์เซ็นเซอร์ หรือสัญญาณรบกวนจากช่องสัญญาณการสื่อสาร ดังนั้นในการตรวจจับความผิดปกติของข้อมูลที่มีอยู่ในเครือข่ายเซ็นเซอร์ไร้สายจึงเป็นเรื่องสำคัญอย่างยิ่งเพื่อหาแนวทางในการปฏิบัติที่เหมาะสม เนื้อหาในบทนี้นำเสนอการบีบอัดของข้อมูลและอัลกอริทึมในการตรวจจับความผิดปกติในเครือข่ายเซ็นเซอร์ไร้สายซึ่งสามารถตรวจจับความผิดปกติได้อย่างแม่นยำโดยการใช้งานเพียงครั้งหนึ่งของข้อมูลที่วัดได้จากเซ็นเซอร์แทนการใช้ทั้งหมด เนื้อหาในบทนี้จะมุ่งไปที่การบีบอัดข้อมูลโดยใช้การแปลงเวฟเล็ตแบบเต็มหน่วย (Discrete Wavelet Transform หรือ DWT) เพื่อป้องกันการตรวจจับความผิดปกติที่ใช้อัลกอริทึม One-Class Support Vector Machine (OCSVM) การวิจัยนี้ได้ทำการทดสอบอัลกอริทึมด้วยข้อมูลสังเคราะห์แบบต่างๆและชุดข้อมูลจริง หลังจากนั้นทำการเปรียบเทียบประสิทธิภาพของอัลกอริทึมกับเทคนิคself-organizing map (SOM) ที่ใช้ร่วมกับ DWT พบว่าอัลกอริทึมที่นำเสนอมีประสิทธิภาพที่ดีกว่า โดยสามารถตรวจจับได้ใกล้เคียง 100% และทำให้ค่า False Positive Rates (FPR) เพิ่มขึ้นเล็กน้อยในส่วนของการผิดพลาดระยะสั้น (short faults) และความผิดพลาดจากสัญญาณรบกวน (noise faults)

2.1 บทนำ

เครือข่ายเซ็นเซอร์ไร้สายประกอบด้วยโหนดเซ็นเซอร์ไร้สายที่ถูกนำไปติดตั้งในบริเวณต่างๆของพื้นที่ที่สนใจ ข้อมูลที่ตรวจวัดได้จะถูกส่งต่อไปยังสถานีฐาน เครือข่ายเซ็นเซอร์ไร้สายถูกสร้างขึ้นด้วยการใช้โหนดเซ็นเซอร์หลายตัวที่มีข้อจำกัดหลายอย่าง เช่น หน่วยความจำ แบนด์วิดท์ (bandwidth) การใช้พลังงาน และความสามารถทางด้านการคำนวณต่างๆ (Goh, et al., 2007) โดย

ข้อจำกัดเหล่านี้ส่งผลให้การสื่อสารไม่น่าเชื่อถือ ซึ่งเป็นส่วนที่สามารถทำให้เกิดความผิดปกติของข้อมูลที่วัดได้

ความผิดปกติหรือค่าที่ผิดปกติในชุดข้อมูลที่วัดได้จาก โหนดเซ็นเซอร์จะส่งผลให้เกิดความไม่สอดคล้องกันกับข้อมูลที่เหลือ (Rajasegarar, et al., 2008) ความผิดปกติที่เกิดขึ้นจากข้อมูลการตรวจวัดเหล่านี้สามารถสร้างความเสียหายให้กับผลผลิตทางการเกษตรได้ ตัวอย่างการประยุกต์ใช้งาน เช่น ฟาร์มไฮโดรโปนิกส์นั้นต้องการความแม่นยำในการวัดค่า pH เพื่อใช้ในการควบคุมผลผลิตหรือในโรงงานปุ๋ยอินทรีย์ชีวภาพที่ต้องการควบคุมอุณหภูมิในกระบวนการหมักปุ๋ย ดังนั้นการตรวจจับความผิดปกติในชุดข้อมูลที่วัดได้อย่างรวดเร็วก็จะหาวิธีรับมือต่อความผิดพลาดได้อย่างทันท่วงที

อย่างไรก็ตาม เนื่องจากข้อจำกัดทางด้านฮาร์ดแวร์ของเครือข่ายเซ็นเซอร์ไร้สายโดยมีความต้องการลดการใช้พลังงาน เนื่องจากการสื่อสารผ่านสัญญาณวิทยุ (Radio communication) ในเครือข่ายเซ็นเซอร์ไร้สายมีการใช้พลังงานมากกว่าการคำนวณและการประมวลผลข้อมูล (Rajasegarar, et al., 2008) บทความวิจัย Siripanadorn, et al., 2010a และ Siripanadorn, et al., 2010b ใช้การบีบอัดข้อมูลโดยวิธีการแปลงเวฟเลตแบบเต็มหน่วย (Discrete Wavelet Transform หรือ DWT) เพื่อป้อนข้อมูลให้กับอัลกอริทึมการตรวจจับความผิดปกติ (Anomaly Detection Algorithm) ซึ่งวิธีการดังกล่าวสามารถเพิ่มประสิทธิภาพในการตรวจจับความผิดปกติ การวิจัยนี้ได้ทำการต่อยอดผลการศึกษาของบทความวิจัยดังกล่าว โดยการรวมการบีบอัดข้อมูล DWT กับเทคนิคการตรวจจับความผิดพลาดด้วยวิธีการอื่น

เทคนิคซัพพอร์ตเวกเตอร์แมชชีนคลาสหนึ่ง (One-Class Support Vector หรือ OCSVM) คือวิธีการตรวจจับความผิดปกติที่มีประโยชน์และเป็นที่ยอมรับที่ไม่ต้องใช้ความรู้เดิม (prior knowledge) เกี่ยวกับการแจกแจงข้อมูล และพบว่ามีความเหมาะสมกับทรัพยากรที่มีอยู่อย่างจำกัดในเครือข่ายเซ็นเซอร์ไร้สาย (Wang, et al., 2010) อัลกอริทึม OCSVM สามารถอัพเดทแบบจำลองพฤติกรรมปกติของข้อมูลในการตรวจวัดข้อมูลแบบออนไลน์ได้ บทความวิจัย Laskov, et al., 2004, Rajasegarar et al. 2007; 2008; 2010, และ Zhang, et al., 2009 ประสบความสำเร็จในการใช้อัลกอริทึม OCSVM เพื่อตรวจจับความผิดปกติในเครือข่ายเซ็นเซอร์ไร้สายด้วยชุดข้อมูลจริงซึ่งเป็นข้อมูลปกติที่สามารถเปลี่ยนแปลงได้ในสภาพแวดล้อมที่เปลี่ยนแปลงตลอดเวลา (dynamic environment) อย่างไรก็ตาม

การวิจัยนี้พบว่า การรวมกันของอัลกอริทึม OCSVM และวิธี DWT ยังไม่ได้ถูกนำเสนอ ดังนั้นวัตถุประสงค์ของบทนี้คือ การศึกษาถึงผลกระทบและประสิทธิภาพของการบีบอัดข้อมูลแบบ DWT บนเทคนิคการตรวจจับความผิดปกติด้วยอัลกอริทึม OCSVM และประเมินความเหมาะสมต่อการนำไปใช้ในเครือข่ายเซ็นเซอร์ไร้สายที่มีทรัพยากรจำกัด

2.2 การตรวจจับความผิดปกติ

ขั้นตอนแรกของการตรวจจับความผิดปกติคือการเลือกตัวแปรข้อมูลที่ตรวจวัดได้ และจัดกลุ่มข้อมูลเหล่านั้นในรูปของเวกเตอร์ $x^\mu \in \mathcal{R}^p, \mu = 1, 2, \dots, n$

$$x^\mu = \begin{bmatrix} x_1^\mu \\ x_2^\mu \\ x_3^\mu \\ \vdots \\ x_p^\mu \end{bmatrix} = \begin{bmatrix} KPI_1^\mu \\ KPI_2^\mu \\ KPI_3^\mu \\ \vdots \\ KPI_p^\mu \end{bmatrix} \quad (2.1)$$

เมื่อ μ คือดัชนีการสำรวจ p คือจำนวนของประเภทตัวแปร หรือ ดัชนีชี้วัดผลงานหรือความสำเร็จของงาน (Key Performance Indices, KPIs) ที่ถูกเลือกเพื่อใช้ตรวจวัดสภาพแวดล้อมที่สนใจ

2.2.1 One-Class Support Vector Machines (OCSVM)

บทความวิจัย (Tax, et al., 2004) ได้นำเสนอวิธี OCSVM สำหรับการตรวจจับค่าผิดปกติ บทความวิจัย (Laskov, et al., 2004) ได้ขยายขอบเขตของวิธีการนี้เข้าไปในวิธีการ SVM แบบพิเศษ ที่เรียกว่า Quarter-Sphere OCSVM โดยแนวความคิดของอัลกอริทึมนี้คือ การล้อมข้อมูลด้วย hypersphere anchor ที่จุดศูนย์กลางของข้อมูลในพื้นที่ ดังนั้นการวิจัยนี้จะนำเสนอสมการทางคณิตศาสตร์ของวิธีการ one-class quarter-sphere SVM

พิจารณาชุดข้อมูลที่ป้อนเข้า (input dataset) $X = \{x^\mu: \mu = 1, 2, 3, \dots, n\}$ ของเวกเตอร์ p ที่แตกต่างกัน $x^\mu = [x_1^\mu, x_2^\mu, x_3^\mu, \dots, x_p^\mu]$ สำหรับข้อมูลนำเข้า (input space) \mathcal{R}^p เมื่อจำนวนของเวกเตอร์ข้อมูลในชุดข้อมูล X คือ n กล่าวคือ X จะถูกรวมเข้าไปในปริภูมิคุณลักษณะ (feature space) \mathcal{R}^q ด้วยฟังก์ชันไม่เชิงเส้น (nonlinear function) ผลลัพธ์ที่ได้จะอยู่ในรูปของเวกเตอร์อิมเมจ (image vectors) $x_\phi = \{\phi(x^\mu): \mu = 1, 2, 3, \dots, n\}$ เมื่อแถวของเวกเตอร์อิมเมจคือ $\phi(x^\mu) = [\phi(x_1^\mu), \phi(x_2^\mu), \phi(x_3^\mu), \dots, \phi(x_n^\mu)]$ วัตถุประสงค์เพื่อ hypersphere ที่เหมาะสมต่อ

ปริภูมิคุณลักษณะด้วยรัศมีที่มีประสิทธิภาพขั้นต่ำ $R > 0$ ที่จุดศูนย์กลาง ซึ่งครอบคลุมเวกเตอร์อิมเมจ X_ϕ ซึ่งกระบวนการดังกล่าวสามารถแก้ปัญหาได้อย่างมีประสิทธิภาพ (Laskov, et al., 2004)

$$\min_{\alpha \in \mathbb{R}^+, \xi \in \mathbb{R}} R^2 + \frac{1}{vn} \sum_{\mu=1}^n \xi_\mu$$

$$\text{ภายใต้เงื่อนไข: } k(x^\mu, x^\mu) \leq R^2 + \xi_\mu \quad (2.2)$$

$$\xi_\mu \geq 0$$

เมื่อ ξ_μ คือตัวแปรช่วยที่ยอมให้บางส่วนของเวกเตอร์อิมเมจอยู่ข้างนอกทรงกลมได้ ตัวแปร $\nu = \epsilon \in (0, 1)$ คือค่าคงที่สำหรับควบคุมค่าคลาดเคลื่อน (Regularization Parameter) ซึ่งคอยควบคุมส่วนของเวกเตอร์อิมเมจที่อยู่ด้านนอกทรงกลม กล่าวคือส่วนของเวกเตอร์อิมเมจสามารถผิดพลาดได้ สังเกตว่า $k(x^\mu, x^\mu) = \phi(x^\mu) \cdot \phi(x^\mu)^T$ สำหรับเมอร์เซอร์เคอร์เนล (Mercer kernel) และ $k(x^\mu, x^\mu)$ คือฟังก์ชันเคอร์เนล (kernel function) ที่ใช้ในการคำนวณความเหมือนของ 2 เวกเตอร์ใดๆ ในปริภูมิคุณลักษณะ (feature space) สมการที่ 2.3 คือสมการคู่ (dual formulation) ของปัญหาเริ่มต้นในสมการที่ 3.2 (Laskov, et al., 2004) สามารถเขียนได้ดังนี้

$$\min_{\alpha \in \mathbb{R}} - \sum_{\mu=1}^n \alpha_\mu k(x^\mu, x^\mu)$$

$$\text{ภายใต้เงื่อนไข } \sum_{\mu=1}^n \alpha_\mu = 1 \quad (2.3)$$

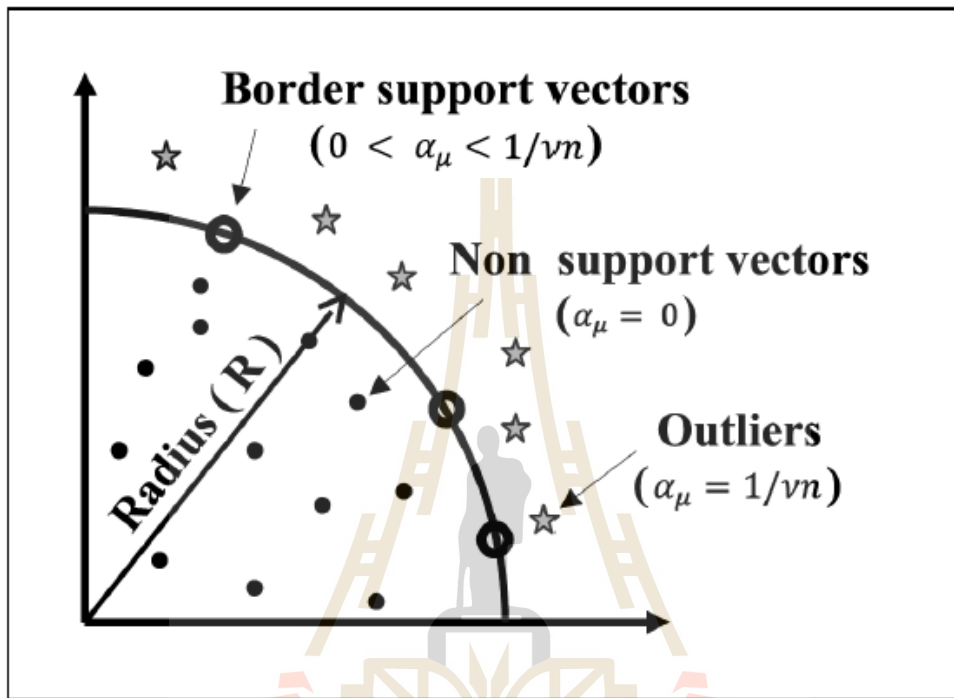
$$0 \leq \alpha_\mu \leq \frac{1}{vn}$$

เมื่อ $\alpha_\mu \geq 0$ คือตัวคูณลากรางจ์ (Lagrange Multipliers) และ $\mu = 1, 2, \dots, n$ ปัญหาควบคู่ (Dual Problem) ในสมการที่ (2.3) นี้คือปัญหาการหาค่าที่เหมาะสมเชิงเส้น (linear optimization problem) เพื่อที่จะลดปัญหาเช่นนี้ เวกเตอร์อิมเมจในปริภูมิคุณลักษณะจะต้องเป็นจุดศูนย์กลางของพื้นที่ โดยใช้เมตริกซ์เคอร์เนล (kernel matrix) (Laskov, et al., 2004) ดังนี้

$$K_c = K - 1_n K - K 1_n + 1_n K 1_n \quad (2.4)$$

เมื่อ K คือเมตริกซ์เคอร์เนลขนาด $n \times n$ ที่ประกอบด้วย $k(x^\mu, x^{\bar{\omega}})$ โดยที่ $\mu, \bar{\omega} = 1, 2, 3, \dots, n$ ถ้า $\mu = \bar{\omega}$ จะได้ว่า $k(x^\mu, x^{\bar{\omega}}) = k(x^\mu, x^\mu)$ ดังนั้นจะได้ $k(x^\mu, x^{\bar{\omega}})$ จากการหาเกณฑ์ปกติ (norms)

ของเวกเตอร์อิมเมจ ในทางตรงข้าม $k(x^\mu, x^{\bar{\omega}})$ สามารถหาได้จากฟังก์ชันเคอร์เนล เช่น เชิงเส้น โพลิโนเมียล เคอร์เนล RBF นอกจากนี้ 1_n คือเมตริกซ์ขนาด $n \times n$ มีค่าทุกค่าเท่ากับ $1/n$ เมื่อเวกเตอร์อิมเมจคือศูนย์กลาง การหาเกณฑ์ปกติของเคอร์เนลก็จะไม่เท่ากัน ดังนั้นจึงสามารถแก้ปัญหาคว่ำคู่ (Dual Problem) ในสมการที่ (2.7) ได้ดังนี้

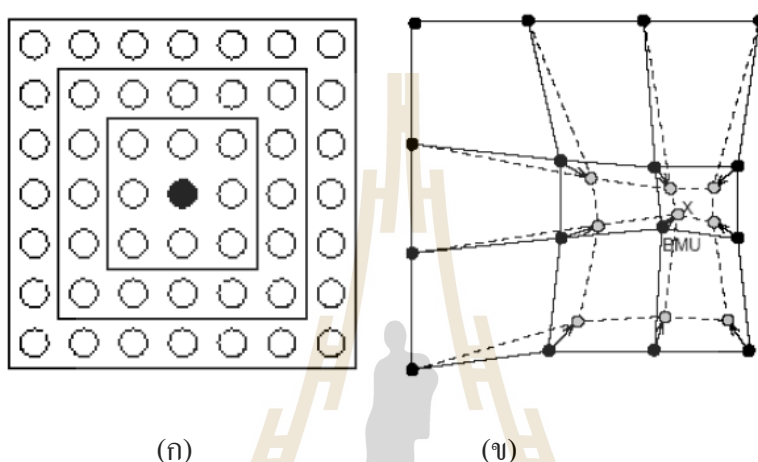


รูปที่ 2.1 เรขาคณิตของอัลกอริทึม quarter-sphere OCSVM

เซต $\{\alpha_\mu\}$ หาได้จากเทคนิคการหาค่าที่เหมาะสมเชิงเส้น (linear optimization techniques) ที่ใช้อย่างแพร่หลาย เวกเตอร์อิมเมจสามารถจัดให้อยู่ในรูปที่ 2.2 เวกเตอร์อิมเมจที่ประกอบด้วย $\alpha_\mu = 0$ จะอยู่ในพื้นที่ของทรงกลม เวกเตอร์อิมเมจที่ประกอบด้วย $\alpha_\mu > 0$ จะถูกเรียกว่า เวกเตอร์ซัพพอร์ต (support vectors) โดยซัพพอร์ตเวกเตอร์ประกอบด้วย $\alpha_\mu = 1/vn$ จะเป็นข้อมูลที่ผิดปกติ (outliers) ซึ่งมีค่าผิดปกติอยู่นอกทรงกลม ด้วยเหตุนี้จึงถูกเรียกว่าขอบเขตของเวกเตอร์ซัพพอร์ต (border support vectors) นอกจากนี้รัศมีของทรงกลม R หาได้จาก $R^2 = k(x^\mu, x^\mu)$ สำหรับ ขอบเขตของเวกเตอร์ซัพพอร์ตของ x^μ (Rajasegarar, et al., 2007)

2.2.2 Self-Organizing Map (SOM)

รูปแบบประสาทเทียมที่มีการแข่งขัน (Competitive neural model) เช่น แผนผังการจัดระบบตัวเอง (Self-Organizing Map, SOM) สามารถดึงแบบแผนทางสถิติจากเวกเตอร์ข้อมูลที่ป้อนเข้า และเข้ารหัสโดยปราศจากการควบคุมดูแล โดยจะทำการส่งผ่านข้อมูลจากมิติระดับสูงไปยังระดับที่ต่ำกว่า โดยทั่วไปจะใช้เป็น 2 มิติ ในแบบกริด (Siripanadorn, et al., 2010a; 2010b)



รูปที่ 2.2 ตัวอย่างของวิธีการ SOM (ก) โครงสร้างเพื่อนบ้านแบบตั้งฉาก ซึ่งอยู่ลึกที่สุดในประสาทเทียม (จุดสีดำ) ที่สอดคล้องกับเพื่อนบ้านที่ 1, 2 และ 3 (ข) SOM ทำการอัปเดต BMU ด้วยเพื่อนบ้านใกล้เคียง 1 ตัว

พื้นฐานของวิธีการ SOM ประกอบด้วยแผนผังกริด หรือประสาทเทียม แสดงในรูปที่ 3.2 (ก) แต่ละเซลล์ประสาท i (จุดสีดำ) มีชุดของชั้นเซลล์ประสาทใกล้เคียง (จุดสีขาว) ซึ่งแสดงในรูปที่ 3.2 (ก) (Siripanadorn, et al., 2010a; 2010b)

เซลล์ประสาท i จะรักษาเวกเตอร์ถ่วงน้ำหนัก m_i เพื่อที่จะติดตามคุณลักษณะของข้อมูลที่ป้อนเข้า เวกเตอร์ดังกล่าวจะถูกอัปเดตในระหว่างกระบวนการฝึกฝน (training process) ตัวอย่างเช่น รูปที่ 2.2 (ข) แสดง SOM ในรูปแบบประสาทเทียมแบบกริด 2 มิติ ขนาด 4×4 โดยขนาดของแต่ละเวกเตอร์มีค่าเท่ากับขนาดของข้อมูลที่ป้อนเข้า ในรูปแสดงให้เห็นว่าเวกเตอร์ของข้อมูลที่ป้อนเข้า (เครื่องหมาย x) ถูกฝึกฝนด้วยเวกเตอร์ถ่วงน้ำหนัก SOM (จุดสีดำ) เซลล์ประสาทที่ชนะ

(BMU) และ เซลล์ประสาทที่ใกล้เคียง (1-neighborhood neurons) จะปรับเปลี่ยนเวกเตอร์ให้เป็นค่าใหม่ (จุดสีเทา)

อัลกอริทึม SOM จะถูกฝึกเข้าไปซ้ำมา ในแต่ละขั้นตอนของการฝึกฝน เวกเตอร์ตัวอย่าง $X' = \{x^\mu: \mu = 1, 2, 3, \dots, s\}$ จากชุดข้อมูล $X = \{x^\mu: \mu = 1, 2, 3, \dots, n\}$ จะถูกเลือก เมื่อจำนวนของเวกเตอร์ตัวอย่าง X' เท่ากับ s และจำนวนของเวกเตอร์ข้อมูลในชุดข้อมูล X คือ n ระยะระหว่างข้อมูลตัวอย่างและเวกเตอร์ถ่วงน้ำหนักทั้งหมดใน SOM จะถูกคำนวณด้วยการใช้การวัดระยะทางบางช่วง สมมติว่าช่วงเวลา t เซลล์ประสาท i ที่มีเวกเตอร์ถ่วงน้ำหนัก $m_i(t)$ อยู่ใกล้กับเวกเตอร์ที่ป้อนเข้า $x^\mu(t)$ ที่สุด ดังนั้นเวกเตอร์ถ่วงน้ำหนัก $m_c(t)$ ซึ่งอ้างอิงถึง Best-Matching Unit (BMU) (Siripanadorn, et al., 2010a; 2010b)

$$\|x^\mu(t) - m_c(t)\| = \arg \min_i \|x^\mu(t) - m_i(t)\| \quad (2.5)$$

เมื่อ $\|\cdot\|$ คือระยะทางแบบยูคลิด (Euclidian distance)

สมมติให้เซลล์ประสาท i ถูกอัปเดต โดยหลักการอัปเดตของ SOM สำหรับเวกเตอร์ถ่วงน้ำหนักของเซลล์ประสาท i (Siripanadorn, et al., 2010a; 2010b)

$$m_i(t+1) = m_i(t) + \eta_t h_c(i, t) [x^\mu(t) - m_i(t)] \quad (2.6)$$

เมื่อ t คือดัชนีการซ้ำ $x^\mu(t)$ คือเวกเตอร์ป้อนเข้า η_t คืออัตราการเรียนรู้ $h_c(i, t)$ คือฟังก์ชันเพื่อนบ้านของอัลกอริทึม ฟังก์ชันเกาส์เซียนเพื่อนบ้าน (Gaussian neighborhood function) ที่อาจถูกนำมาใช้ (Siripanadorn, et al., 2010a; 2010b) ได้แก่

$$h_c(i, t) = \exp\left[-\frac{\|r_c(t) - r_i(t)\|^2}{2\sigma^2(t)}\right] \quad (2.7)$$

เมื่อ $r_i(t)$ และ $r_c(t)$ คือตำแหน่งของเซลล์ประสาท i และค่า BMU c ตามลำดับ และ σ^2 คือรัศมีของฟังก์ชันเพื่อนบ้านที่เวลา t โดยที่ $h_c(i, t)$ นิยามถึงความกว้างของเพื่อนบ้าน โดยมีความสำคัญอย่างยิ่งที่ $\lim_{t \rightarrow \infty} h_c(i, t) = 0$ และ $\lim_{t \rightarrow \infty} \eta_c(i, t) = 0$ สำหรับอัลกอริทึม convergence (Siripanadorn, et al., 2010a; 2010b)

2.3 การบีบอัดข้อมูล

2.3.1 การแปลงเวฟเลตแบบเต็มหน่วย

การแปลงเวฟเลตแบบเต็มหน่วย (Discrete Wavelet Transform หรือ DWT) การแปลงทางคณิตศาสตร์ที่แยกสัญญาณข้อมูลเป็นข้อมูลสเกลที่ดี (fine-scale information) ที่รู้จักในชื่อสัมประสิทธิ์รายละเอียด (detail coefficients) และข้อมูลสเกลหยาบ (rough-scale information) ที่รู้จักในชื่อค่าสัมประสิทธิ์การประมาณ (approximate coefficients) ข้อดีของหลักการนี้คือ การนำเสนอด้วยความละเอียดที่หลากหลาย (multi-resolution representation) และคุณสมบัติการจำกัดวงทางความถี่และเวลาของสัญญาณ โดยปกติการร่างชุดเวลาตั้งต้นสามารถกลับไปตั้งต้นใหม่ได้ได้ด้วยการใช้ค่าสัมประสิทธิ์การจำแนกต่ำผ่าน (low-pass-cut off decomposition coefficients) สามารถจำลองจากค่าสัมประสิทธิ์การจำแนกระดับกลาง (middle-level decomposition coefficients) ส่วนที่เหลือจะถูกพิจารณาเป็นสัญญาณรบกวนหรือความผิดปกติ สมการต่อไปนี้อธิบายการคำนวณในกระบวนการแจกแจงของ DWT (Siripanadorn, et al., 2010a; 2010b)

$$a_{j+1}^{DWT}(f) = \sum_n h_0(n - 2f) a_j^{DWT}(f) \quad (2.8)$$

$$d_{j+1}^{DWT}(f) = \sum_n g_0(n - 2f) a_j^{DWT}(f) \quad (2.9)$$

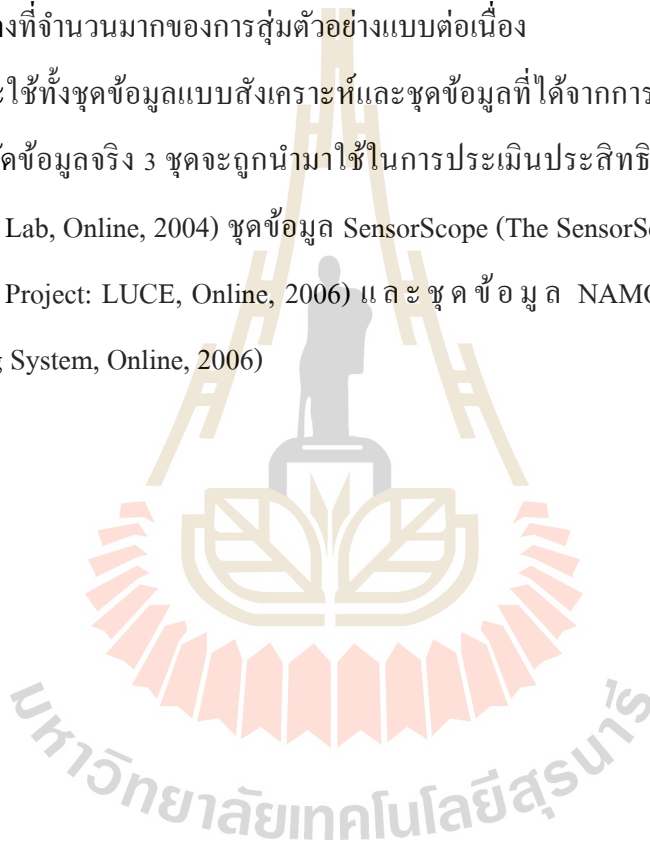
เมื่อค่าสัมประสิทธิ์สเกลหยาบ (หรือ ค่าประมาณ) a_j^{DWT} ถูกรวมเข้าด้วยกัน ด้วย h_0 และ g_0 ซึ่งเป็นฟังก์ชันเวฟเลตและฟังก์ชันสเกล (scaling function) ตามลำดับ n คือดัชนีสเกลเวลา (time scaling index) f คือดัชนีการแปลงค่าทางความถี่ (frequency translation index) สำหรับเวฟเลตระดับ j ค่าสัมประสิทธิ์ลัพธ์จะถูกลดขนาดลงด้วย 2 กระบวนการนี้จะแยก a_{j+1}^{DWT} ครึ่งหนึ่งอย่างหยาบ โดยแยกให้อยู่ในเซตของสเกลที่ดี หรือสัมประสิทธิ์แบบละเอียด d_{j+1}^{DWT} และเซตสเกลหยาบของสัมประสิทธิ์แบบค่าประมาณ a_{j+1}^{DWT} (Siripanadorn, et al., 2010a; 2010b)

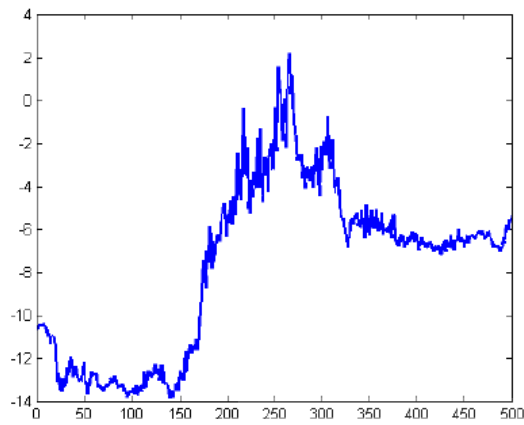
DWT มีความสามารถในการเข้ารหัสความละเอียดปลีกย่อยของชุดเวลาตั้งต้นกับค่าสัมประสิทธิ์ลำดับชั้น นอกจากนี้ DWT ยังสามารถคำนวณแบบเวลาเชิงเส้นได้อย่างมีประสิทธิภาพ ซึ่งมีความสำคัญอย่างยิ่งในการรับมือกับชุดข้อมูลขนาดใหญ่

2.4 ชุดข้อมูลสำหรับการทดลอง

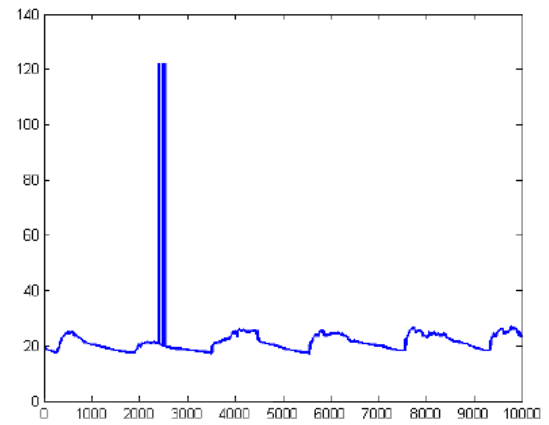
การวิจัยนี้จำแนกความผิดพลาดออกเป็น 3 ประเภทแสดงในรูปที่ 2.3 กล่าวคือ ความผิดปกติจากสัญญาณรบกวน (noise faults) ความผิดปกติระยะสั้น (short faults) และความผิดปกติคงที่ (constant faults) (Sharma, et al., 2010) ความผิดปกติจากสัญญาณรบกวนคือความผิดปกติที่เกิดขึ้นเมื่อความแปรปรวนของการอ่านข้อมูลของเซ็นเซอร์ที่สูงขึ้นและส่งผลกระทบต่อจำนวนตัวอย่างต่อเนื่อง (successive samples) ความผิดปกติระยะสั้นคือการเปลี่ยนแปลงแบบพุ่งสูงขึ้นของค่าที่วัดได้ระหว่างข้อมูลที่ต่อเนื่องและส่งผลกระทบต่อตัวอย่างที่เวลาหนึ่ง ความผิดปกติคงที่คือความผิดปกติที่เกิดจากค่าคงที่จำนวนมากของการสุ่มตัวอย่างแบบต่อเนื่อง

การวิจัยนี้จะใช้ทั้งชุดข้อมูลแบบสังเคราะห์และชุดข้อมูลที่ได้จากการวัดข้อมูลจริง โดยชุดข้อมูลที่ได้จากการวัดข้อมูลจริง 3 ชุดจะถูกนำมาใช้ในการประเมินประสิทธิภาพ ได้แก่ ชุดข้อมูล INTEL (The INTEL Lab, Online, 2004) ชุดข้อมูล SensorScope (The SensorScope Lausanne Urban Canopy Experiment Project: LUCE, Online, 2006) และชุดข้อมูล NAMOS (Network Aquatic Microbial Observing System, Online, 2006)

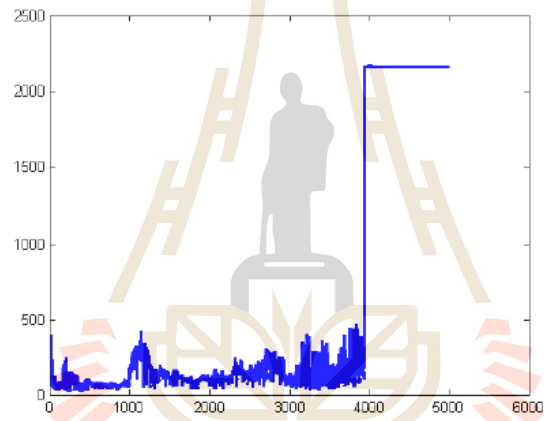




(ก)



(ข)



(ค)

รูปที่ 2.3 ความผิดปกติในการอ่านข้อมูลของเซ็นเซอร์

2.4.1 ข้อมูลที่มาจากการสังเคราะห์

ข้อมูลที่มาจากการสังเคราะห์ (synthetic data) ถูกสร้างขึ้นจากการผสมกันระหว่างการแจกแจงแบบเกาส์เซียน (Gaussian distributions) ด้วยค่ากลางที่สุ่มจาก (0.3, 0.35, 0.45) และค่าเบี่ยงเบนมาตรฐาน 0.03 โดยใช้โปรแกรม MATLAB ข้อมูลจะถูกสร้างขึ้นให้กับโหนดเซ็นเซอร์ 15 ตัว และคุณสมบัติทั้งสองของเวกเตอร์ข้อมูล 106ตัวต่อหนึ่งโหนดเซ็นเซอร์ ข้อมูลที่รวบรวมได้จึงประกอบด้วยเวกเตอร์ข้อมูล 1590 ข้อมูล หลังจากนั้นจำนวนของความผิดปกติที่มีการแจกแจงแบบปกติอยู่ในช่วงระหว่าง [0.5, 1] จะถูกใส่เข้าไปในคุณสมบัติของแต่ละข้อมูล (Rajasegarar, et al., 2007) จำนวนของความผิดปกติแทนด้วย a/s เมื่อ a คือจำนวนความผิดปกติต่อชุดข้อมูล และ s คือจำนวนชุดข้อมูลที่มีความผิดปกติ ผลรวมทั้งหมดของข้อมูลผิดปกติจึงเป็น $a \times s$ ความผิดปกติที่สร้างขึ้นนั้นจะถูกป้อนเข้าไปในข้อมูลอินพุตซึ่งมาจากความผิดปกติจากสัญญาณรบกวนในช่วง 20/4 10/8 5/16 และความผิดปกติระยะสั้น 1/80 ความผิดปกติทั้งหมดที่ใช้จึงเท่ากับ 80 ข้อมูล ชุดข้อมูลทั้งหมดจะถูกทำให้อยู่ในรูปแบบมาตรฐานในช่วง [0, 1] ตำแหน่งที่แน่นอนของความผิดพลาดถูกใส่เข้าไปในข้อมูลอินพุตที่กำหนดไว้เพื่อใช้ในการตรวจจับการแจ้งเตือนความผิดปกติที่ถูกและผิด

2.4.2 ชุดข้อมูล INTEL

เซ็นเซอร์ตรวจจับความชื้น แสงสว่าง และอุณหภูมิ ชนิด Mica2Dot จำนวน 54 ตัว ถูกใช้ในห้องแลปวิจัยของ INTEL ระหว่างวันที่ 28 กุมภาพันธ์ ถึง 5 เมษายน 2547 (The INTEL Lab, Online, 2004) การวิจัยนี้ได้แสดงผลของการตรวจจับความผิดปกติในการอ่านข้อมูลอุณหภูมิ และเลือกค่าเทรชโฮลด์ (threshold) 16 และ 30 เพื่อใช้เป็นขอบบนและขอบล่างของพื้นที่ที่มีข้อมูลปกติ ค่าเหล่านี้ได้มาจากการทำฮิสโตแกรม (Histogram) โดยพิจารณาจากตำแหน่งของข้อมูลที่ผิดปกติ และพบว่าชุดข้อมูลนี้เป็นข้อมูลที่มีความผิดพลาดระยะสั้น

2.4.3 ชุดข้อมูล SensorScope Station no.39 dataset (SS39)

ในการทดลองนี้ได้นำเสนอผลของการตรวจจับความผิดปกติใน KPI หนึ่งของ SensorScope ซึ่งถูกรวบรวมมาจากสถานีเลขที่ 39 (SS39) โดยใช้การตรวจสอบภาพเสมือนจริงและการใช้วิธีการทางฮิสโตแกรม ค่าเทรชโฮลด์ที่เป็นขอบบนและขอบล่างที่ใช้สำหรับการตรวจจับความผิดปกติมีค่าเป็น 1.5 และ 9 เมื่อพิจารณาถึงตำแหน่งของข้อมูลที่ผิดปกติพบว่า ชุดข้อมูลนี้เป็นข้อมูลที่มีความผิดปกติแบบสั้น

2.4.4 ชุดข้อมูล SensorScope pdg2008-metro-1 dataset (pdg2008)

ในการทดลองนี้ใช้ KPIs 2 ชนิดจากข้อมูลในชุดข้อมูล pdg2008-metro-1 ในการตรวจจับความผิดปกติ กล่าวคือ เป็นข้อมูลอุณหภูมิพื้นผิวและอุณหภูมิโดยรอบ โดยใช้การตรวจสอบภาพเสมือนจริงและการใช้วิธีการทางฮิสโตแกรม ค่าเทรสโฮลด์ที่เป็นขอบบนและขอบล่างที่ใช้สำหรับการตรวจจับความผิดปกติมีค่าเป็น -14 และ 4 สำหรับอุณหภูมิพื้นผิว และค่า 12 และ 4 สำหรับอุณหภูมิโดยรอบ เมื่อพิจารณาถึงตำแหน่งของข้อมูลที่ผิดปกติพบว่า ชุดข้อมูลนี้เป็นข้อมูลที่มีความผิดปกติแบบคงที่

2.4.5 ชุดข้อมูล NAMOS

ข้อมูลในชุดนี้ ใช้หุ่นลอยน้ำ 9 ตัวกับเซ็นเซอร์ตรวจจับความเข้มข้นของคลอโรฟิลล์และอุณหภูมิในทะเลสาป Fulmor ตลอด 24 ชั่วโมงในเดือนสิงหาคมปีพ.ศ. 2549 (Network Aquatic Microbial Observing System, Online, 2006) เมื่อวิเคราะห์ผลการตรวจวัดจากเซ็นเซอร์วัดคลอโรฟิลล์บนหุ่นลอยน้ำหมายเลข 13 จำนวน 10,000 ตัวอย่าง โดยในการทดลองได้ใช้วิธีการทางฮิสโตแกรมสำหรับจำแนกความผิดปกติต่างๆของชุดข้อมูล NAMOS โดยทำการเลือกค่าเทรสโฮลด์ 0 และ 500 เป็นขอบบนและขอบล่างของพื้นที่ปกติตามลำดับ เมื่อพิจารณาถึงตำแหน่งของข้อมูลที่ผิดปกติพบว่า ชุดข้อมูลนี้เป็นข้อมูลที่มีความผิดปกติแบบคงที่

2.5 ผลการทดลอง

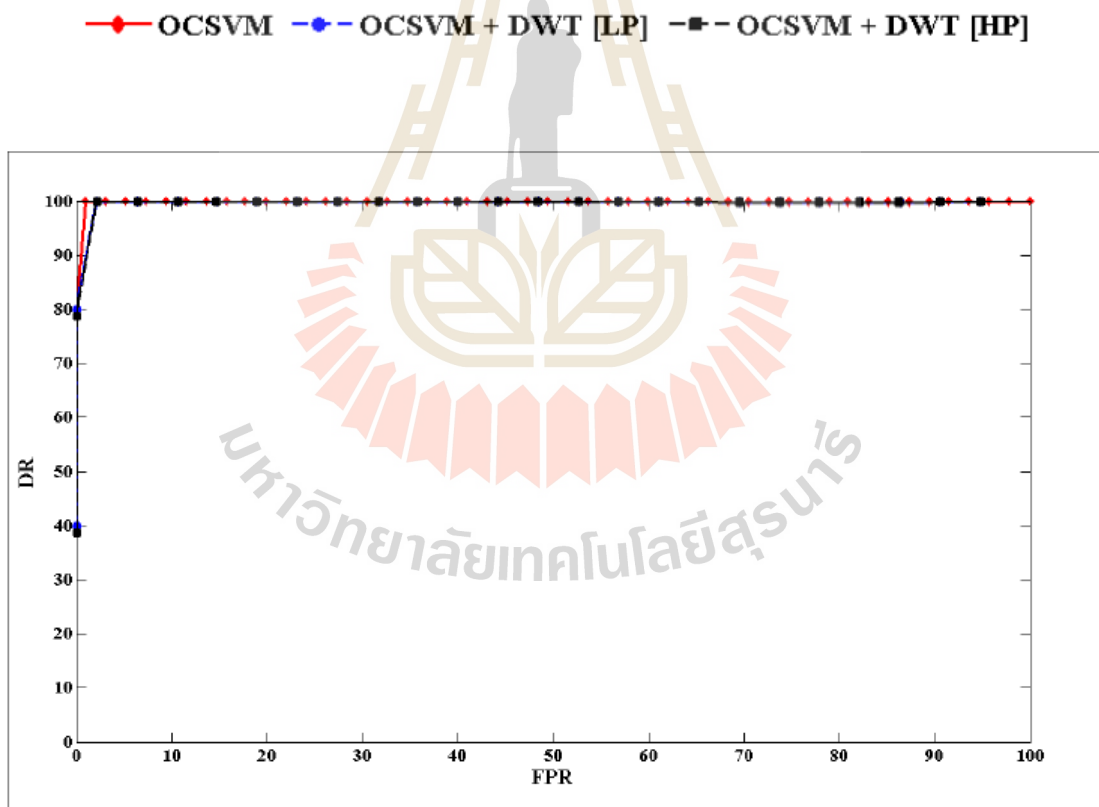
ในหัวข้อนี้ประกอบไปด้วย 2 ส่วน กล่าวคือ ส่วนที่หนึ่งเป็นการประเมินการทำงานของอัลกอริทึมที่รวม DWT และ OCSVM โดยการตรวจจับความผิดปกติในชุดข้อมูลสังเคราะห์และชุดข้อมูลที่ได้จากการวัดจริง จากนั้นจึงประเมินผลอัลกอริทึมที่นำเสนอเทียบกับเทคนิคก่อนหน้าที่ใช้ DWT และ SOM (Siripanadorn, et al., 2010a; 2010b)

2.5.1 การประเมินผลของ DWT ด้วยอัลกอริทึม OCSVM

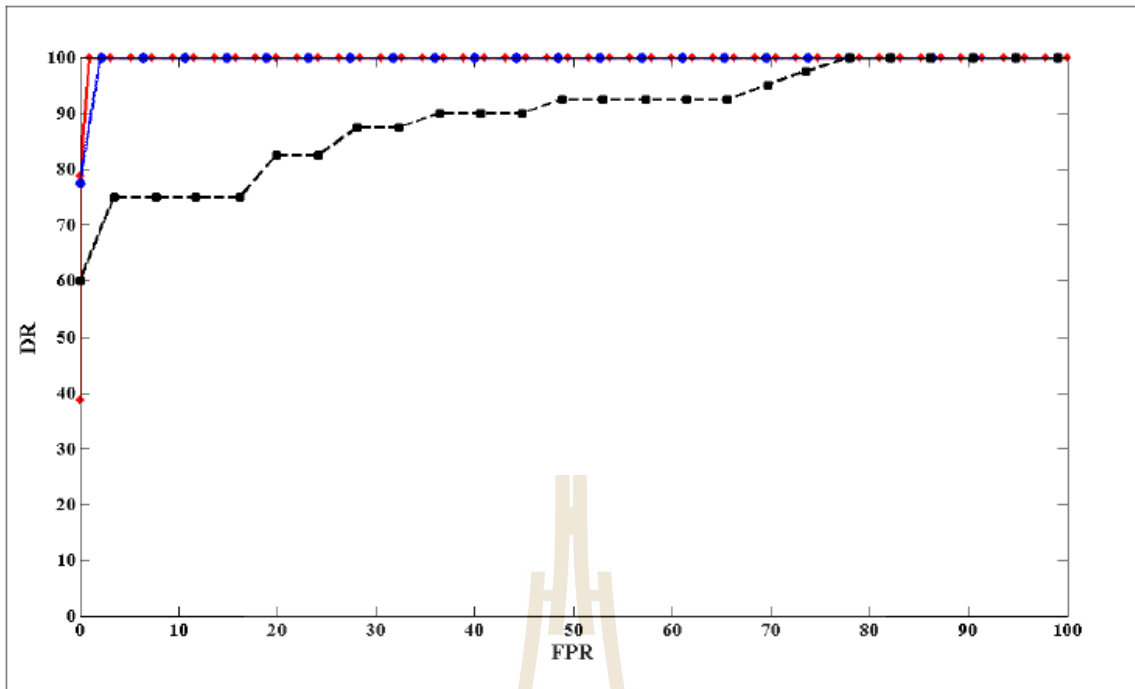
ในบทนี้จะใช้เคอร์เนลเชิงเส้น (linear kernel) เป็นค่าระยะทางเคอร์เนล โดยฟังก์ชันเคอร์เนลเชิงเส้นสำหรับเวกเตอร์ข้อมูล x^μ และ $x^{\bar{w}}$ ถูกกำหนดโดย $k_{\text{linear}}(x^\mu, x^{\bar{w}}) = \phi(x^\mu) \cdot \phi(x^{\bar{w}})$

ในแต่ละการจำลอง (simulation) ได้ทำการบันทึกค่า False positives ซึ่งเกิดขึ้นเมื่อการวัดแบบปกติถูกระบุเป็นความผิดปกติจากการตรวจจับ และ True positive ซึ่งเกิดขึ้นจากการวัดความผิดปกติที่เกิดขึ้นจริง ซึ่งเป็นการตรวจจับที่ระบุได้อย่างถูกต้อง false positive rate (FPR) ถูกคำนวณจากอัตราส่วนระหว่างผลบวกลวงและการวัดค่าปกติที่วัดได้จริง และอัตราการตรวจจับ (detection rate, DR) ถูกคำนวณจากอัตราส่วนระหว่างผลบวกจริงและการวัดค่าปกติที่วัดได้จริง

อัลกอริทึมที่ได้นำเสนอในบทความนี้ได้ทำการเพิ่มส่วนของDWTด้วยOCSVM (OCSVM+DWT) เพื่อปรับปรุงประสิทธิภาพการทำงานในส่วนของ OCSVM โดยการแทนที่ข้อมูลอินพุตชุดเดิมด้วยค่าสัมประสิทธิ์DWTสูงผ่านหรือต่ำผ่านด้วยการใช้เวฟเล็ตที่เรียกว่า Haar mother wavelet ค่าสัมประสิทธิ์ต่ำผ่าน(หรือสูงผ่าน)DWTหาได้จากข้อมูลต่ำผ่าน(หรือสูงผ่าน)โดยที่เป็นเพียงข้อมูลเพียงครั้งเดียวจากขนาดของข้อมูลต้นฉบับ ในขณะที่ข้อมูลต้นฉบับนั้นเป็นข้อมูลที่ยังไม่ได้ผ่านการบีบอัดข้อมูล

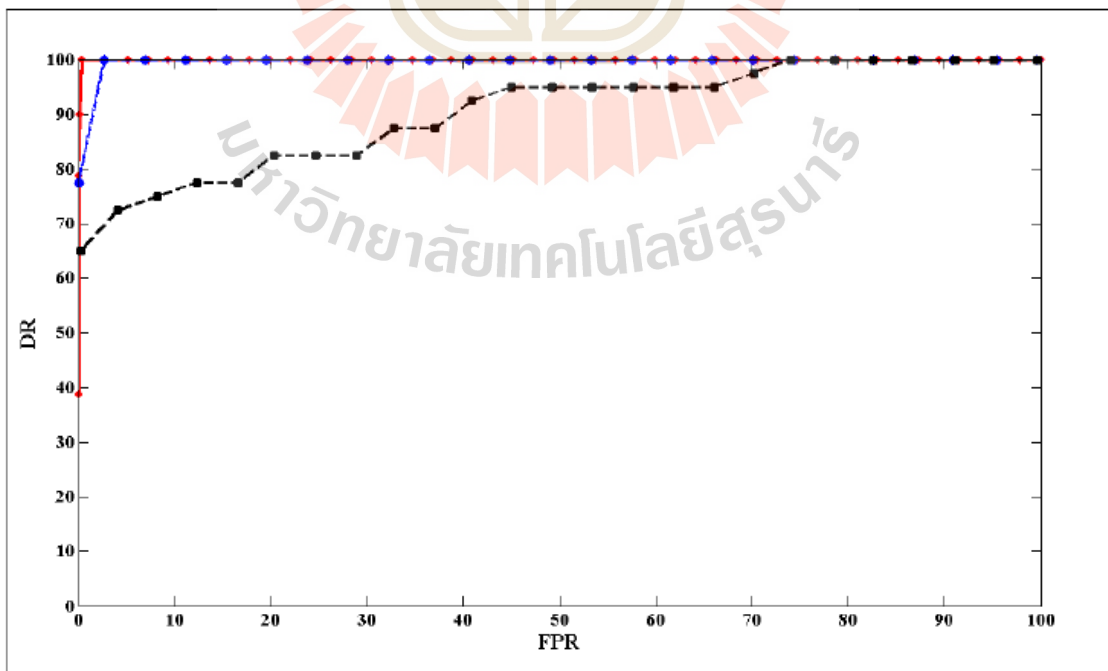


รูปที่ 2.4 ROC สำหรับข้อมูลสังเคราะห์ที่เพิ่มความผิดปกติ 1/80 เข้าไป

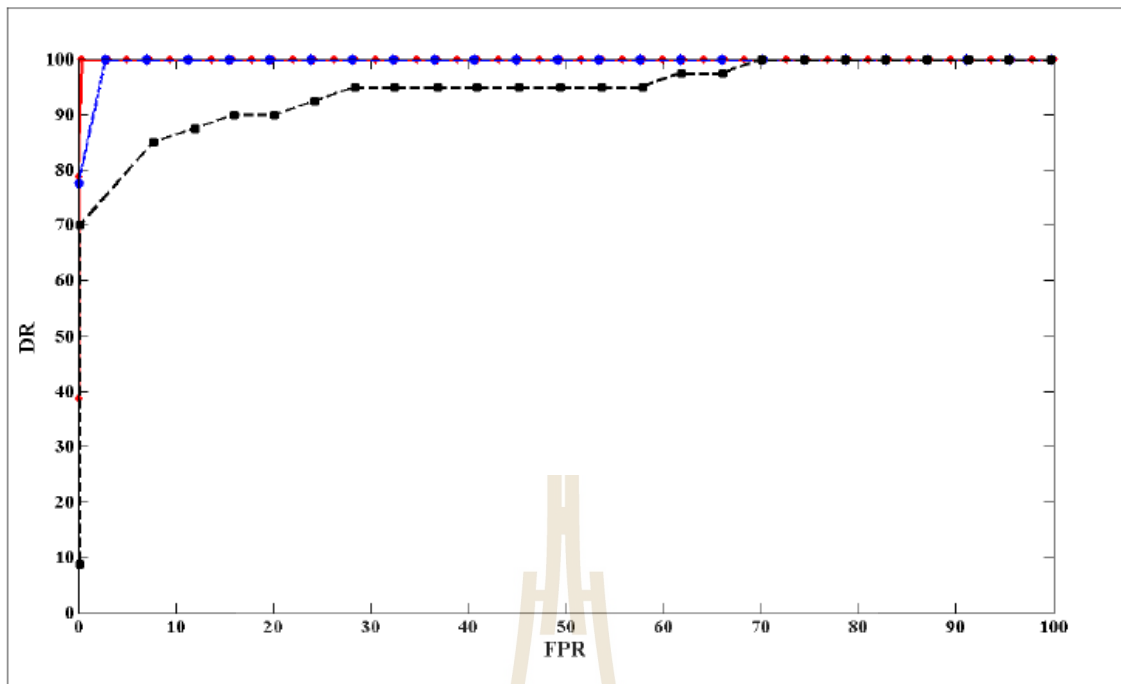


รูปที่ 2.5 ROC สำหรับข้อมูลสังเคราะห์ที่เพิ่มความผิดปกติ 5/16 เข้าไป

—◆— OCSVM -●- OCSVM + DWT [LP] -■- OCSVM + DWT [HP]

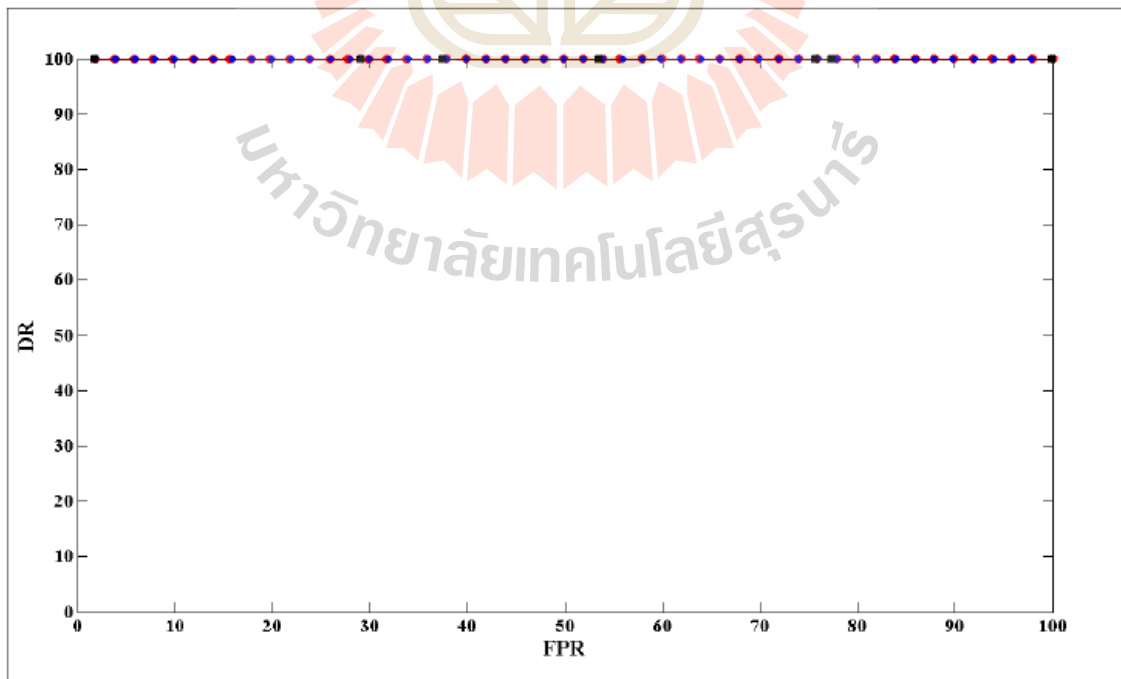


รูปที่ 2.6 ROC สำหรับข้อมูลสังเคราะห์ที่เพิ่มความผิดปกติ 10/8 เข้าไป

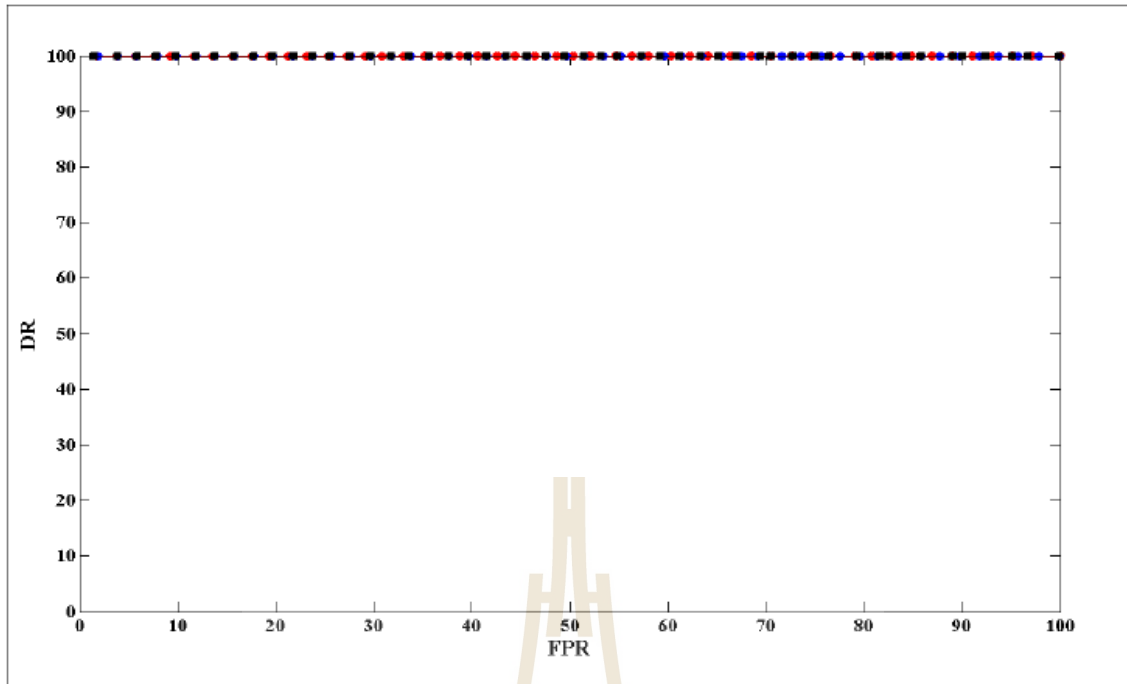


รูปที่ 2.7 ROC สำหรับข้อมูลสังเคราะห์ที่เพิ่มความผิดปกติ 20/4 เข้าไป

—◆— OCSVM —●— OCSVM + DWT [LP] -■- OCSVM + DWT [HP]

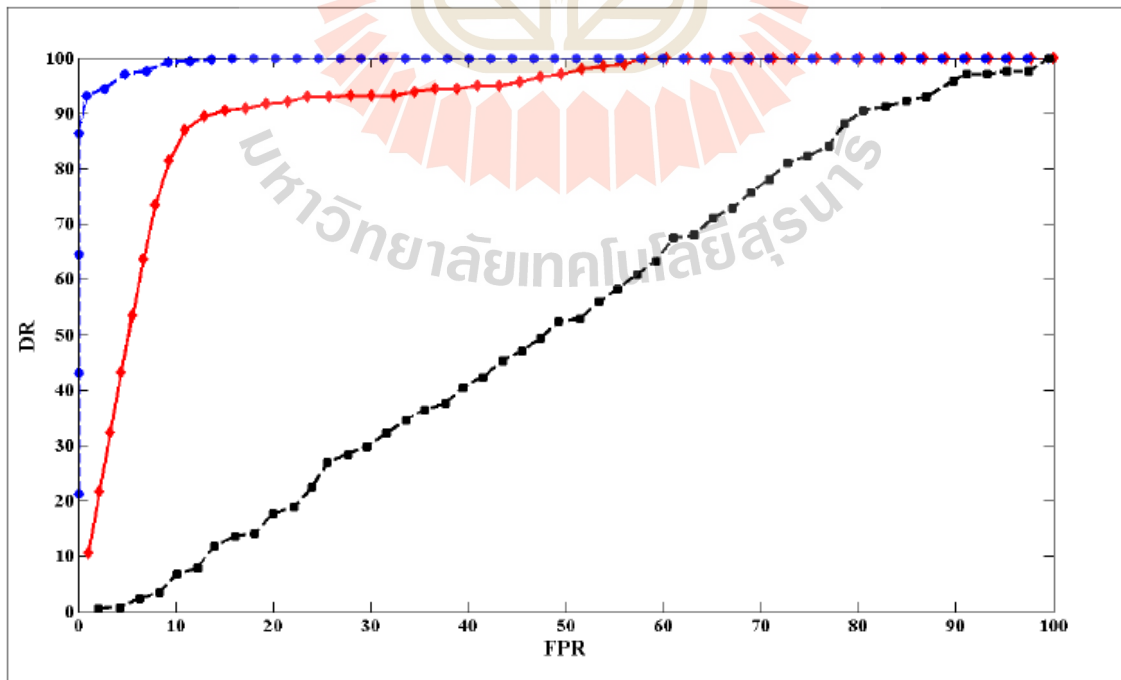


รูปที่ 2.8 ROC สำหรับชุดข้อมูล INTEL

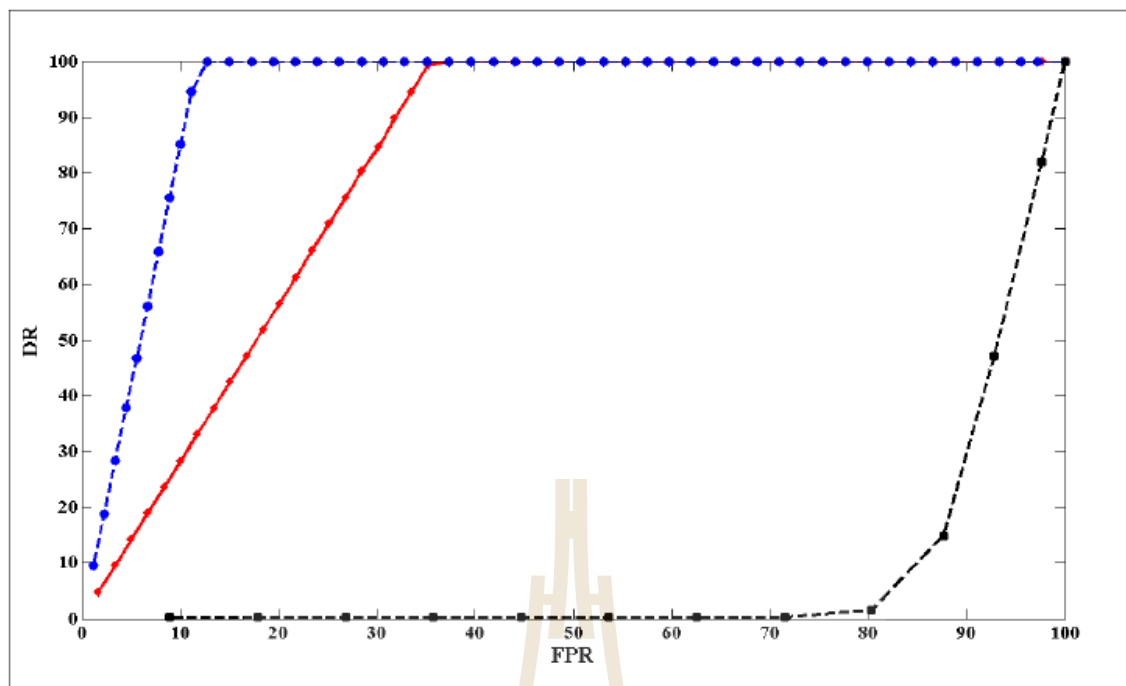


รูปที่ 2.9 ROC สำหรับชุดข้อมูล SS39

—●— OCSVM —■— OCSVM + DWT [LP] —■— OCSVM + DWT [HP]



รูปที่ 2.10 ROC สำหรับชุดข้อมูล pdg2008-metro-1



รูปที่ 2.11 ROC สำหรับชุดข้อมูล NAMOS

รูปที่ 2.4-2.11 แสดงกราฟคุณลักษณะของการดำเนินการที่ภากรับ (receiver operating characteristics, ROC) ของวิธีการแบบ OCSVM+DWT สำหรับชุดข้อมูลต่างๆด้วยการเปลี่ยนแปลงค่าพารามิเตอร์ ν จาก 0.02 ถึง 1 โดยเพิ่มขึ้นทีละ 0.02 ผลที่ได้แสดงให้เห็นว่า ν มีผลกระทบต่อ DR และ FPR ในวิธีการแบบ OCSVM อย่างยิ่ง โดยที่ค่าพารามิเตอร์ ν คือสัดส่วนของความผิดพลาดที่ตรวจจับได้ (Zhang, et al., 2009) ซึ่งเป็นสัดส่วนโดยตรงต่อค่ารัศมี R (รูปที่ 2.1) ดังนั้นเมื่อค่า ν มีค่าสูงขึ้นจึงแสดงถึงการตรวจจับความผิดพลาดที่สูงขึ้น (ส่งผลให้ค่า DR และค่า FPR สูงขึ้นเช่นกัน)

รูปที่ 2.4 แสดงกราฟของชุดข้อมูลสังเคราะห์ที่เพิ่มความผิดพลาดระยะสั้น (short faults) เข้าไป พบว่าทุกๆอัลกอริทึมสามารถทำงานได้ดีเท่าๆกัน รูปที่ 2.5 2.6 และ 2.7 แสดงผลของข้อมูลสังเคราะห์ที่เพิ่มความผิดพลาดจากสัญญาณรบกวน (noise faults) เข้าไป โดยพบว่าอัลกอริทึม OCSVM+DWT (LP) นั้นสามารถทำงานได้ดีในเทอมของ DR เช่นเดียวกับอัลกอริทึม OCSVM สำหรับข้อมูลที่ไม่มีการบีบอัดข้อมูล อย่างไรก็ตามอัลกอริทึม OCSVM+DWT (HP) ให้ประสิทธิภาพการทำงานที่แย่ที่สุด ซึ่งมาสาเหตุมาจากค่าสัมประสิทธิ์ HP นั้นเหมาะสมต่อความผิดพลาดระยะสั้น

ในขณะที่ค่าสัมประสิทธิ์ LP นั้นเหมาะสมต่อความผิดปกติที่มีการเปลี่ยนแปลงอย่างช้าๆเหมือนความผิดปกติจากสัญญาณรบกวน

รูปที่ 2.8-2.11 แสดงผลของชุดข้อมูลที่ได้จากการวัดจริง โดยที่รูปที่ 2.8 และ 2.9 แสดงผลของชุดข้อมูล INTEL และชุดข้อมูล SS39 ตามลำดับ ซึ่งจะเห็นว่าทุกอัลกอริทึมให้ค่า DR 100% ที่เป็นเช่นนี้เพราะว่าชุดข้อมูล INTEL และชุดข้อมูล SS39 ประกอบด้วยความผิดปกติระยะสั้น ซึ่งมีแอมพลิจูดสูงซึ่งง่ายต่อการตรวจจับ รูปที่ 2.10 และ 2.11 แสดงผลของชุดข้อมูล pdg2008-metro-1 และชุดข้อมูล NAMOS ตามลำดับ โดยพบว่าอัลกอริทึม OCSVM+DWT (LP) ให้ค่า DR ที่สูงกว่าและค่า FPR ที่ต่ำกว่า ซึ่งดีกว่าอัลกอริทึม OCSVM สำหรับข้อมูลที่ไม่มีการบีบอัดและอัลกอริทึม OCSVM+DWT (HP) ทั้งนี้เป็นเพราะว่าชุดข้อมูล pdg2008-metro-1 ประกอบด้วยความผิดปกติจากสัญญาณรบกวน และชุดข้อมูล NAMOS ประกอบด้วยความผิดปกติคงที่ (constant faults) โดยความผิดปกติทั้ง 2 ชนิดนี้มีแนวโน้มที่จะเปลี่ยนแปลง ดังนั้นจึงง่ายต่อการตรวจจับด้วยค่าสัมประสิทธิ์ต่ำผ่านได้มากกว่าค่าสัมประสิทธิ์สูงผ่าน โดยจะเห็นว่าอัลกอริทึม OCSVM+DWT (HP) มีการทำงานที่แย่ที่สุด

จะสังเกตได้ว่าผลที่ได้จากทั้งข้อมูลสังเคราะห์และข้อมูลที่ได้จากการวัดจริงแสดงให้เห็นว่าอัลกอริทึม OCSVM+DWT (LP) ทำงานได้ดีในเทอมของเปอร์เซ็นต์ DR เนื่องจากค่าสัมประสิทธิ์ต่ำผ่านจะให้แอมพลิจูดของความผิดปกติที่สูงกว่าชุดข้อมูลต้นฉบับ จึงทำให้ง่ายต่อการตรวจจับความผิดปกติ (แสดงในภาคผนวก ก)

2.5.2 เปรียบเทียบประสิทธิภาพกับงานที่ผ่านมา

ในส่วนนี้จะแสดงให้เห็นว่าผลที่ได้จากอัลกอริทึม OCSVM+DWT ในแต่ละชุดข้อมูลที่ถูกลีเลือกมาเปรียบเทียบกับอัลกอริทึม OCSVM อัลกอริทึม SOM และอัลกอริทึม SOM+DWT (Siripanadorn, et al., 2010a; 2010b) ภายใตชุดข้อมูลเดียวกันดังที่อธิบายไว้ในหัวข้อก่อนหน้า โดยในแต่ละชุดข้อมูลจะทำการเลือกค่าพารามิเตอร์ v เพื่อนำมาใช้ในอัลกอริทึมที่น่าเสนอ ซึ่งจะเป็นค่าที่ทำให้เกิดประสิทธิภาพการทำงานที่ดีที่สุด (DR สูงสุด, FPR ต่ำสุด) ดังนี้

- $v = 0.06$ สำหรับชุดข้อมูลสังเคราะห์ทั้งหมด
- $v = 0.02$ สำหรับชุดข้อมูล INTEL และชุดข้อมูล SS39
- $v = 0.14$ สำหรับชุดข้อมูล pdg2008-metro-1

- $v = 0.22$ สำหรับชุดข้อมูล NAMOS

อัลกอริทึม SOM และอัลกอริทึม SOM+DWT ถูกฝึกซ้ำ 50 ครั้งเพื่อป้องกันการเกิดการฝึกไม่เพียงพอ โดยใช้ประสาทเทียมแบบกริดขนาด 50×50 จำนวนตัวอย่างที่ใช้ในการฝึกอัลกอริทึม SOM และอัลกอริทึม SOM+DWT ถูกเลือกจากเขตพื้นที่ที่ไม่มีคามผิดปกติ โดยเลือกมา 2,000 ตัวอย่างสำหรับชุดข้อมูล INTEL และชุดข้อมูล pdg2008-metro-1 3,000 ตัวอย่างสำหรับชุดข้อมูล NAMOS และ 3,200 ตัวอย่างสำหรับชุดข้อมูล SS39

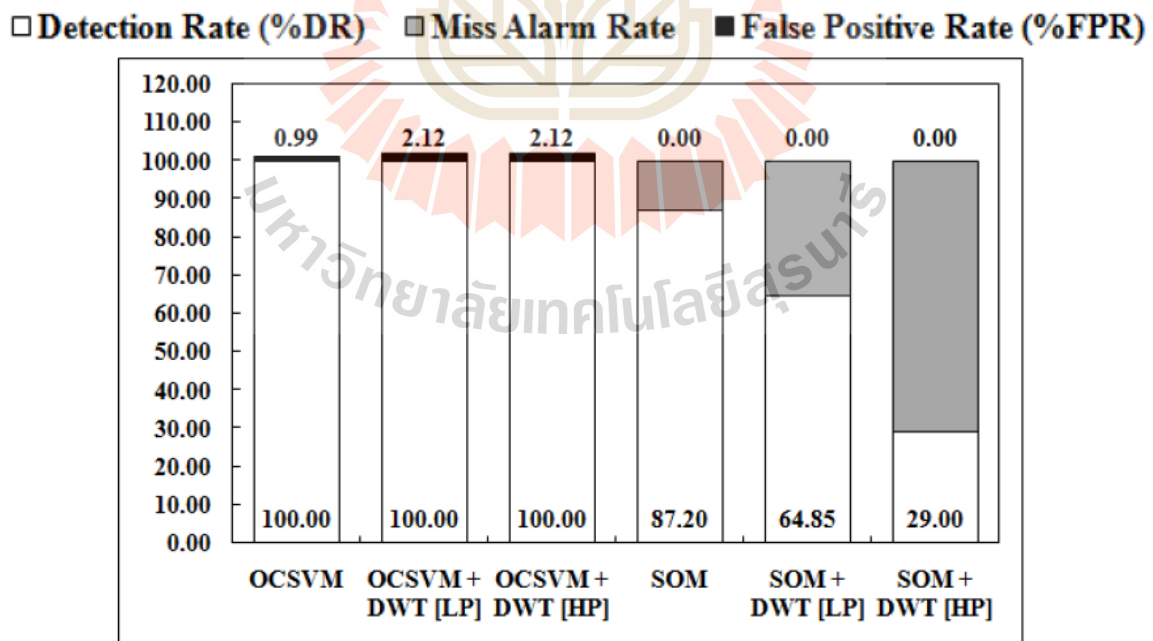
รูปที่ 2.12 แสดงผลการใช้อัลกอริทึม OCSVM และอัลกอริทึม OCSVM+DWT ด้วยข้อมูลสังเคราะห์ที่เพิ่มความผิดปกติระยะสั้นเข้าไป $1/80$ โดยให้ค่า DR 100% แม้ว่าอัลกอริทึม OCSVM+DWT จะให้ค่า FPR 2.12% และอัลกอริทึม OCSVM ให้ค่า FPR 0.99% สำหรับอัลกอริทึม SOM และอัลกอริทึม SOM+DWT นั้นให้ผลของค่า DR ที่น้อยกว่าและไม่มีค่า FPR เลย เมื่อความผิดปกติสูงขึ้นอย่างรวดเร็ว (ความผิดปกติจากสัญญาณรบกวนมากขึ้น) แสดงในรูปที่ 2.13 2.14 และ 2.15 อัลกอริทึม OCSVM และอัลกอริทึม OCSVM+DWT (LP) มีค่า DR 100% ในขณะที่อัลกอริทึม OCSVM+DWT (HP) มีค่า DR เพียง 75-85% อย่างไรก็ตามค่า FPR กลับเพิ่มสูงขึ้นเป็น 0.13-0.99% 6.36-7.02%, และ 7.68-8.21% สำหรับอัลกอริทึม OCSVM อัลกอริทึม OCSVM+DWT (LP) และอัลกอริทึม OCSVM+DWT (HP) ตามลำดับ โดยจะเห็นว่าอัลกอริทึม SOM ที่ประกอบด้วยข้อมูลที่ยังไม่ได้ถูกบีบอัดนั้นมีค่า DR อยู่ในช่วง 86-90% และไม่ทำให้เกิดค่า FPR เลย โดยใช้ข้อมูลอินพุตเพียงครั้งเดียวเท่านั้น

สำหรับชุดข้อมูลที่ได้จากการวัดจริง โดยที่ชุดข้อมูล INTEL และชุดข้อมูล SS39 จะมีเพียงความผิดปกติระยะสั้นเท่านั้น ซึ่งง่ายต่อการตรวจจับความผิดปกติ ผลการทดลองแสดงในรูปที่ 2.16 และ 2.17 ซึ่งสอดคล้องกับรูปที่ 2.12 ภายใต้การใช้อัลกอริทึม OCSVM โดยมีค่า DR 100% แต่ค่า FPR กลับลดต่ำลงโดยเหลือเพียง 1.09% การเพิ่มประสิทธิภาพการทำงานของทุกอัลกอริทึมสามารถทำได้เนื่องจากความจริงที่ว่า ความผิดปกติระยะสั้นที่เกิดในชุดข้อมูล INTEL และชุดข้อมูล SS39 นั้นสามารถตรวจจับได้ง่ายกว่าชุดข้อมูลสังเคราะห์

รูปที่ 2.18 แสดงผลที่ได้จากชุดข้อมูล pdg2008 ที่ประกอบด้วยความผิดปกติจากสัญญาณรบกวน พบว่าความผิดปกติจากสัญญาณรบกวนนี้ส่งผลให้อัลกอริทึม OCSVM มีค่า FPR ที่

มากกว่าอัลกอริทึม SOM ซึ่งสอดคล้องกับผลของชุดข้อมูลสังเคราะห์ที่แสดงในรูปที่ 2.13 2.14 และ 2.15 อย่างไรก็ตามอัลกอริทึม OCSVM+DWT (LP) นั้นให้ผลลัพธ์ที่ดีที่สุด โดยมีค่า DR 97.04% และค่า FPR 4.81% ซึ่งดีกว่าอัลกอริทึม OCSVM

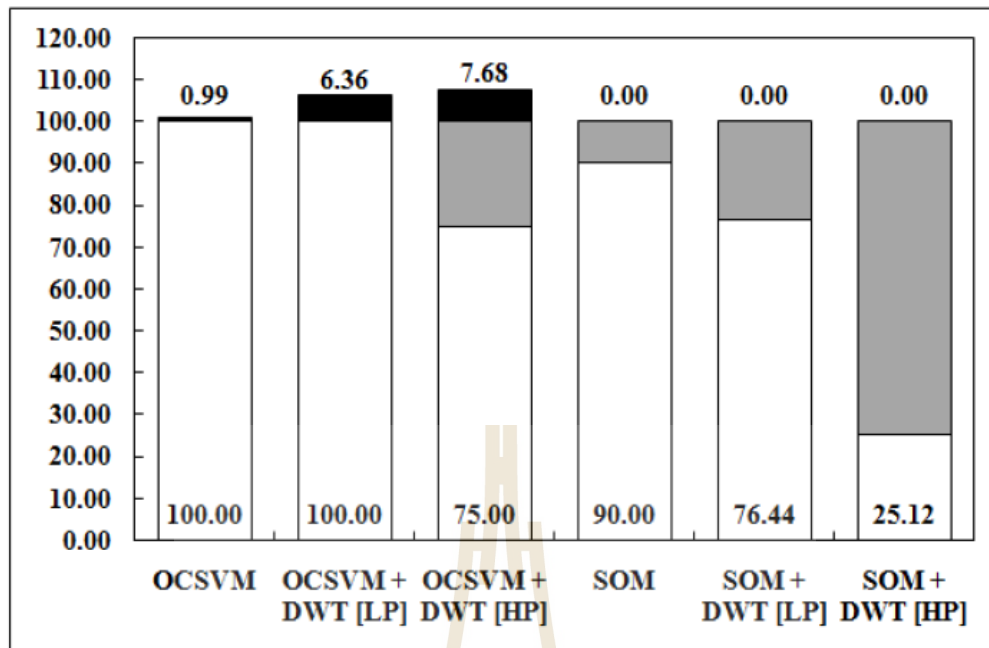
รูปที่ 2.19 แสดงผลที่ได้จากชุดข้อมูล NAMOS ที่ประกอบด้วยความคิดปกติกงที่ซึ่งความคิดปกตินี้ยากต่อการตรวจจับเนื่องจากข้อมูลมีลักษณะเหมือนกับข้อมูลปกติทั่วไป แม้ว่าจะใช้อัลกอริทึม SOM และอัลกอริทึม SOM+DWT (LP) ก็ไม่สามารถตรวจจับความคิดปกตินี้ได้ถ้าไม่ได้รับการฝึกมาอย่างเพียงพอ (Siripanadorn, et al., 2010a; 2010b) จากผลการทดลองสามารถสังเกตได้ว่าอัลกอริทึม SOM และอัลกอริทึม SOM+DWT (LP) นั้นมีค่า DR เกือบจะ 100% ในขณะที่อัลกอริทึม OCSVM+DWT (LP) นั้นมีค่า DR 100% โดยมีค่า FPR เป็น 12.77% โดยที่อัลกอริทึม SOM และอัลกอริทึม SOM+DWT (LP) มีค่า FPR เพียงเล็กน้อยเท่านั้น จากผลการทดลองนี้ชี้ให้เห็นว่าการบีบอัดข้อมูลและการใช้ข้อมูลอินพุตเพียงครั้งเดียวนั้น การใช้อัลกอริทึม OCSVM+DWT (LP) จะเหมาะสมสำหรับข้อมูลที่ความผิดปกติระยะสั้นและความผิดปกติจากสัญญาณรบกวน ในขณะที่อัลกอริทึม SOM+DWT (LP) เหมาะสมสำหรับข้อมูลที่ความผิดปกติระยะสั้นและข้อมูลที่ความผิดปกติ



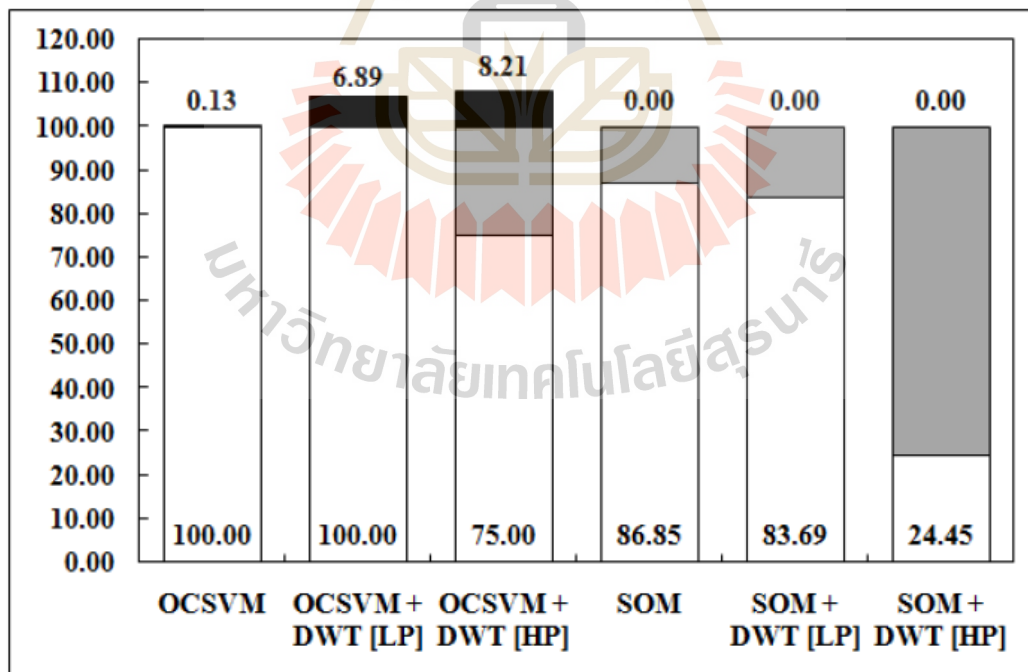
รูปที่ 2.12 อัตราการตรวจจับความคิดปกติด้วยอัลกอริทึมที่แตกต่างกัน สำหรับข้อมูลสังเคราะห์ที่มี

ความผิดปกติ 1/80

□ Detection Rate (%DR) ■ Miss Alarm Rate ■ False Positive Rate (%FPR)

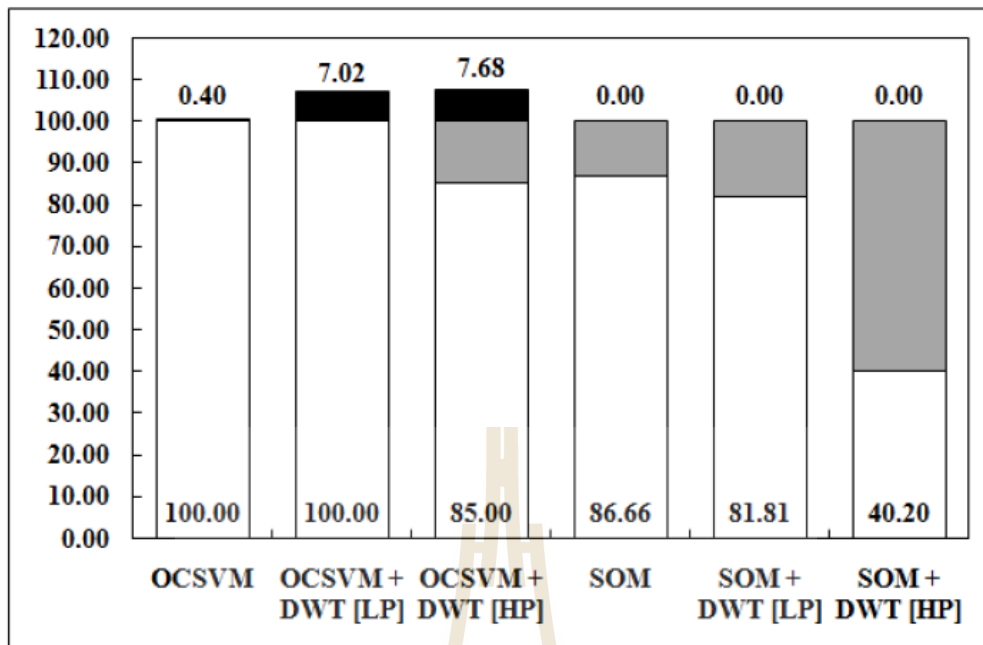


รูปที่ 2.13 อัตราการตรวจจับความผิดปกติด้วยอัลกอริทึมที่แตกต่างกัน สำหรับข้อมูลสังเคราะห์ที่มีความผิดปกติ 5/16



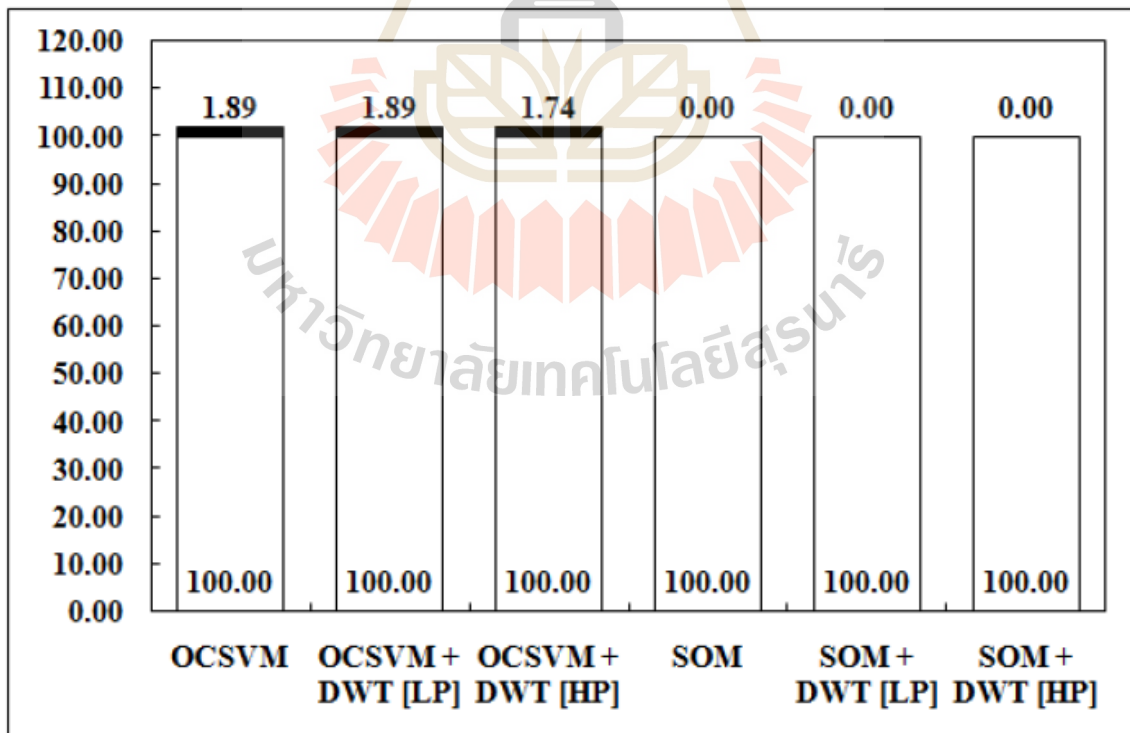
รูปที่ 2.14 อัตราการตรวจจับความผิดปกติด้วยอัลกอริทึมที่แตกต่างกัน สำหรับข้อมูลสังเคราะห์ที่มีความผิดปกติ 10/8

□ Detection Rate (%DR) ■ Miss Alarm Rate ■ False Positive Rate (%FPR)



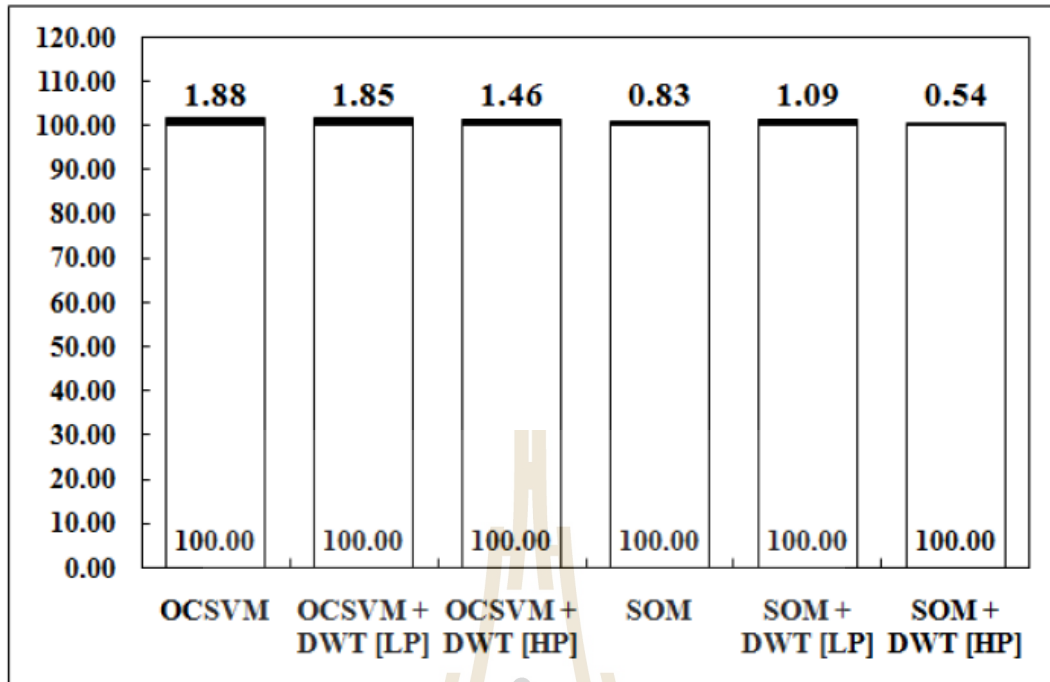
รูปที่ 2.15 อัตราการตรวจจับความผิดปกติด้วยอัลกอริทึมที่แตกต่างกัน สำหรับข้อมูลสังเคราะห์ที่มี

ความผิดปกติ 20/4

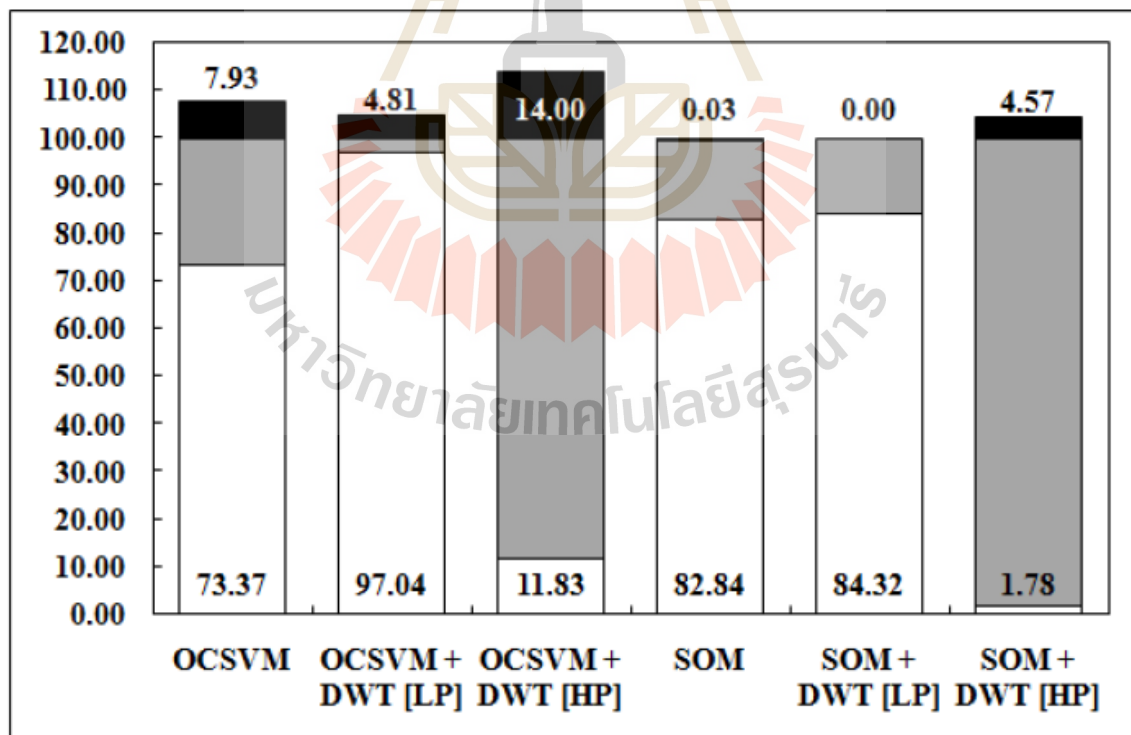


รูปที่ 2.16 อัตราการตรวจจับความผิดปกติด้วยอัลกอริทึมที่แตกต่างกันสำหรับชุดข้อมูล INTEL

□ Detection Rate (%DR) ■ Miss Alarm Rate ■ False Positive Rate (%FPR)

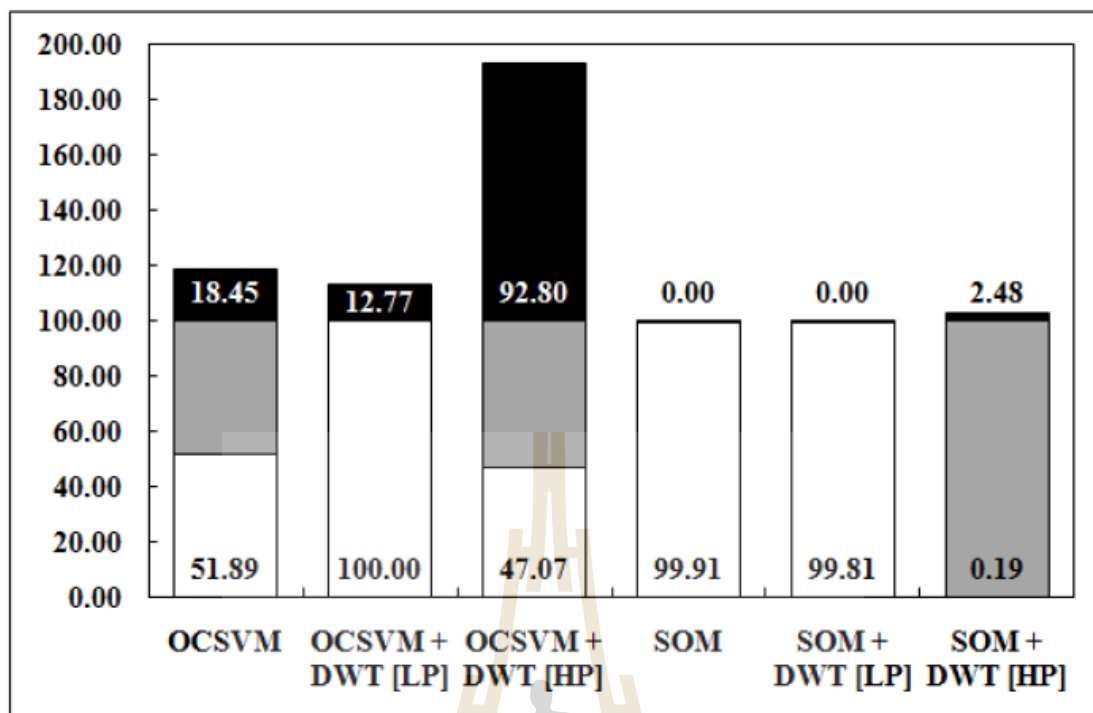


รูปที่ 2.17 อัตราการตรวจจับความผิดปกติด้วยอัลกอริทึมที่แตกต่างกันสำหรับชุดข้อมูล SS39



รูปที่ 2.18 อัตราการตรวจจับความผิดปกติด้วยอัลกอริทึมที่แตกต่างกันสำหรับชุดข้อมูล pdg2008

□ Detection Rate (%DR) ■ Miss Alarm Rate ■ False Positive Rate (%FPR)



รูปที่ 2.19 อัตราการตรวจจับความผิดปกติด้วยอัลกอริทึมที่ต่างกันสำหรับชุดข้อมูล NAMOS

2.6 บทสรุป

ในบทนี้ได้นำเสนออัลกอริทึม OCSVM ที่รวมกับวิธี DWT (OCSVM+DWT) สำหรับการตรวจจับความผิดปกติในเครื่องข่ายเช่นเซอร์ไวร์สาย โดยทำการประเมินประสิทธิภาพของอัลกอริทึมโดยใช้โปรแกรม MATLAB และทดสอบอัลกอริทึมด้วยข้อมูลสังเคราะห์และข้อมูลที่ได้จากการวัดจริง สำหรับชุดข้อมูลสังเคราะห์ พบว่าอัลกอริทึมที่นำเสนอด้วยค่าสัมประสิทธิ์ค่าผ่านสามารถให้ผลของ DR ที่ 100% โดยค่า FPR จะสูงขึ้นเล็กน้อยเมื่อเทียบกับอัลกอริทึมอื่น สำหรับชุดข้อมูลที่ได้จากการวัดจริง พบว่าอัลกอริทึมที่นำเสนอสามารถทำงานได้ดีที่สุด โดยมีค่า DR ที่ใกล้ 100% แม้ว่าค่า FPR จะสูงขึ้นเล็กน้อยสำหรับชุดข้อมูลที่ประกอบด้วยความผิดปกติระยะสั้นและความผิดปกติจากสัญญาณรบกวน ผลการทดลองนี้ชี้ให้เห็นว่าการบีบอัดข้อมูลและการใช้ข้อมูลอินพุตเพียงครั้งเดียวนั้น การใช้อัลกอริทึม OCSVM+DWT (LP) จะเหมาะสำหรับข้อมูลที่มีความผิดปกติระยะสั้นและความผิดปกติจากสัญญาณรบกวน ในขณะที่อัลกอริทึม SOM+DWT (LP) เหมาะสำหรับข้อมูลที่มีความผิดปกติระยะสั้นและข้อมูลที่มีความผิดปกติคงที่

สำหรับการตรวจจับความผิดปกติโดยใช้อัลกอริทึม SOM และอัลกอริทึม OCSVM ด้วยค่าสัมประสิทธิ์ต่ำผ่านของวิธี DWT หรือ OCSVM+DWT (LP) นั้นให้ประสิทธิภาพการทำงานที่ดีกว่าอัลกอริทึม SOM และอัลกอริทึมที่ใช้ OCSVM เพียงอย่างเดียว เนื่องจากค่าสัมประสิทธิ์ต่ำผ่านจะให้แอมพลิจูดของความผิดปกติสูงกว่าชุดข้อมูลต้นฉบับ ซึ่งทำให้ง่ายต่อการตรวจจับ (แสดงใน ภาคผนวก ก) นอกจากนี้อัลกอริทึม OCSVM ยังทนทานต่อผลกระทบจากการปรับให้อยู่ในรูปมาตรฐาน (normalization effects) (แสดงใน ภาคผนวก ค) ซึ่งทำให้เกิดแรงจูงใจในการใช้งานอัลกอริทึมนี้ในบทความต่อไป



บทที่ 3

วิธีการแปลงเวฟเล็ตแบบลิฟติงและเทคนิคซัพพอร์ตเวกเตอร์แมชชีนคลาสหนึ่ง สำหรับการตรวจจับความผิดปกติในเครือข่ายเซ็นเซอร์ไร้สาย

บทนี้จะนำเสนอการรวมกันระหว่างอัลกอริทึมในการตรวจจับความผิดปกติ และการบีบอัดข้อมูล โดยมุ่งไปที่การบีบอัดข้อมูลโดยใช้วิธีการแปลงเวฟเล็ตแบบลิฟติง (lifting wavelet transform หรือ LWT) จากนั้นเพื่อนำมาใช้ในการตรวจจับความผิดปกติ ที่เรียกว่า เทคนิคซัพพอร์ตเวกเตอร์แมชชีนคลาสหนึ่ง (one-class support vector machine หรือ OCSVM) เราทดสอบอัลกอริทึมนี้โดยข้อมูลจริงละข้อมูลสังเคราะห์จำนวนมากและทำการเปรียบเทียบกับอัลกอริทึมการบีบอัดแบบอื่นที่มีอยู่เช่น หลักการวิเคราะห์องค์ประกอบหลัก (principal component analysis หรือ PCA) และ การแปลงเวฟเล็ตแบบเต็มหน่วย (Discrete Wavelet Transform หรือ DWT)

3.1 บทนำ

จากบทที่ 2 ซึ่งแสดงถึงการรวมกันของการตรวจจับความผิดปกติด้วย OSVM และการบีบอัดข้อมูลด้วย DWT นั้นให้ประสิทธิภาพที่ดีกว่าวิธีการที่มีอยู่ นอกจากนี้ (Kiziloren, et al., 2009) ได้เสนอการรวมกันระหว่างการตรวจจับความผิดปกติและการบีบอัดข้อมูลโดยการใช้วิธีการ PCA โดยเป็นเครื่องมือที่มีความเหมาะสมในการลดมิติของข้อมูลก่อนที่จะป้อนให้กับอัลกอริทึมการตรวจจับความผิดปกติ ซึ่งวิธีการนี้สามารถเพิ่มประสิทธิภาพการตรวจจับความผิดปกติได้ แรงจูงใจจากการค้นพบนี้ทำให้เราขยายงานจากบทที่ 2 เพื่อรวมการตรวจจับความผิดปกติ OCSVM ด้วยการบีบอัดข้อมูลอย่างอื่น

PCA ลดมิติของชุดข้อมูลโดยการลดตัวแปรหรือจำนวนของประเภทพารามิเตอร์ที่วัดได้ อย่างไรก็ตามมีความเป็นไปได้ว่าผลลัพธ์จะมีคุณสมบัติบางอย่างของชุดข้อมูลหายไป เพราะว่า PCA สามารถละเลยบางองค์ประกอบที่มีความสำคัญน้อย (Smith, 2002) ซึ่งเหมือนว่าจะทำให้ประสิทธิภาพของการตรวจจับลดลง

ในทางกลับกัน DWT มีความสามารถในการเข้ารหัส รายละเอียดปลีกย่อยของชุดเวลาดั้งเดิม ด้วยค่าสัมประสิทธิ์ความเป็นลำดับชั้นของมัน DWT ลดเฉพาะขนาดของเวกเตอร์ข้อมูลเท่านั้น ไม่ลดตัวแปรพารามิเตอร์ นอกจากนี้ DWT ยังสามารถถูกคำนวณในรูปแบบเวลาเชิงเส้นได้อย่างมีประสิทธิภาพอีกด้วย ซึ่งวิธีนี้มีความสำคัญเมื่อจะจัดการกับข้อมูลขนาดใหญ่ (Siripanadorn, et al., 2010) อย่างไรก็ตาม (Acharya, et.al , 2006), (Manjunath, et al., 2010) เสนอว่า DWT ที่ถูกสร้างจากการคอนโวลูชัน ซึ่งต้องการหน่วยความจำขนาดใหญ่ การคำนวณทางคณิตศาสตร์จำนวนมากและเวลาในการคำนวณที่ยาวนาน มากกว่าการแปลงเวฟเลตอันใหม่ที่เรียกว่า lifting wavelet transforms (LWT) (Sweldens, 1998) จากข้อดีที่เราทราบว่าการรวมกันระหว่าง OCSVM และ LWT นั้นยังไม่ถูกนำเสนอ เพราะฉะนั้นจุดมุ่งหมายของบทนี้จะนำไปเพื่อการศึกษาเกี่ยวกับผลกระทบและประสิทธิภาพของการบีบอัดข้อมูลโดยใช้ LWT บนเทคนิคการตรวจจับข้อมูลแบบ OSCVM จากนั้นเราจะประเมินความเหมาะสมของวิธีการนี้ในการนำไปใช้ในเงื่อนไขข้อจำกัดทางทรัพยากรของเครือข่ายเซ็นเซอร์ไร้สาย และเปรียบเทียบกับวิธีการการรวมกันของการบีบอัดข้อมูลกับการตรวจจับความผิดปกติอื่นที่มีอยู่

3.2 การตรวจจับความผิดปกติ

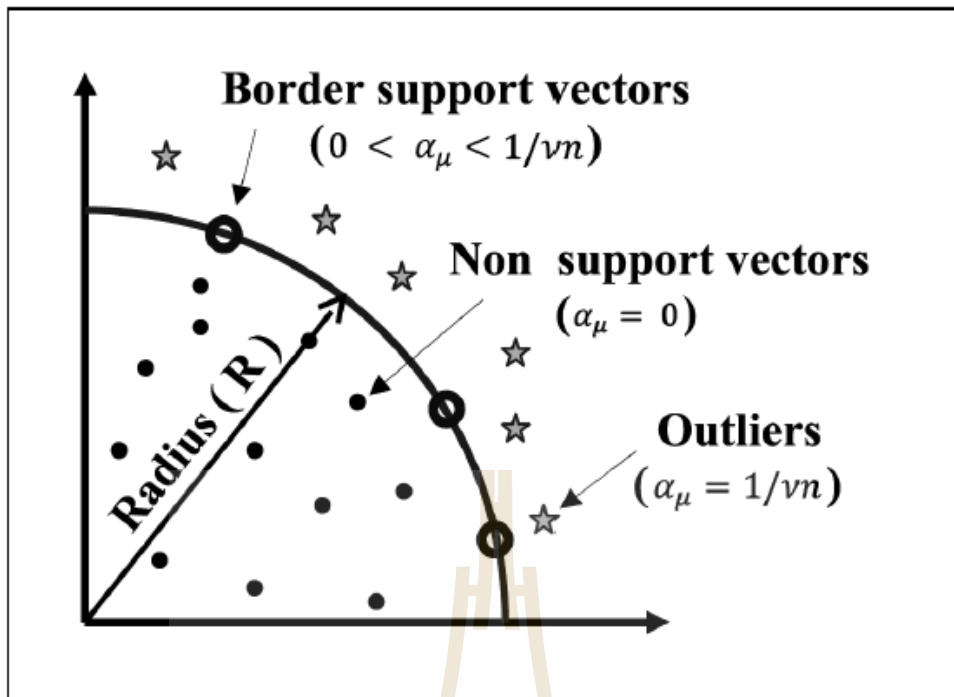
ขั้นตอนแรกของการตรวจจับความผิดปกติเกี่ยวข้องกับทางเลือกพารามิเตอร์ข้อมูลที่จะทำการสังเกตการณ์และรวมกลุ่มไว้ด้วยกันให้อยู่ในเวกเตอร์รูปแบบ $x^\mu \in \mathcal{R}$

$$x^\mu = \begin{bmatrix} x_1^\mu \\ x_2^\mu \\ x_3^\mu \\ \vdots \\ x_p^\mu \end{bmatrix} = \begin{bmatrix} KPI_1^\mu \\ KPI_2^\mu \\ KPI_3^\mu \\ \vdots \\ KPI_p^\mu \end{bmatrix} \quad (3.1)$$

เมื่อ $\mu = 1, 2, 3, \dots, n$ คือ ดัชนีการสังเกต

n คือ จำนวนเวกเตอร์ข้อมูล ในชุดข้อมูล

p คือ จำนวนประเภทพารามิเตอร์หรือตัวชี้วัดคุณภาพ (KPIs) ที่ถูกเลือกเพื่อสังเกตการณ์
เงื่อนไขทางสิ่งแวดล้อม



รูปที่ 3.1 เรขาคณิตของอัลกอริทึม quarter-sphere OCSVM (Rajasegarar, et al., 2007)

3.2.1 One-Class Support Vector Machines (OCSVM)

(Tax, et al., 2004) ได้นำเสนอหลักเกณฑ์ของ OCSVM สำหรับตรวจจับค่าผิดปกติ ต่อมา (Laskov, et al., 2004) ได้ขยาย OCSVM ไปสู่ข้อมูลแบบ one-side non-negative ที่มีความต้องการให้จุดศูนย์กลางของทรงกลมคงที่อยู่จุดกำเนิดเรียกวิธีการนี้ว่า Quarter-Sphere OCSVM รูปทรงเรขาคณิตของวิธีการนี้แสดงในรูปที่ 3.1

พิจารณาชุดข้อมูลอินพุต $X = \{x^\mu: \mu = 1, 2, 3, \dots, n\}$ ของเวกเตอร์ตัวแปร p $x^\mu = [x_1^\mu, x_2^\mu, \dots, x_p^\mu]$ ในปริภูมิอินพุต \mathcal{R}^p ที่จำนวนเวกเตอร์ข้อมูลในชุดข้อมูล X คือ n โดยหลักการ X ถูกแปลงให้เข้าไปอยู่ในปริภูมิคุณลักษณะ \mathcal{R}^q ผ่านฟังก์ชันที่ไม่เป็นเชิงเส้น $\phi(\cdot)$ ให้ผลลัพธ์อยู่ในชุดของเวกเตอร์อิมเมจ $x_\phi = \{\phi(x^\mu): \mu = 1, 2, 3, \dots, n\}$

โดยที่ $\phi(x^\mu) = \{\phi(x_1^\mu), \phi(x_2^\mu), \phi(x_3^\mu), \dots, \phi(x_q^\mu)\}$ คือเวกเตอร์แถวของเวกเตอร์อิมเมจ ข้อมูลปกติสามารถอธิบายโดยหนึ่งส่วนของรูปทรงกลมในรูปที่ 3.1 จุดมุ่งหมายคือการจัดไฮเปอร์สเฟียร์ (hypersphere) ให้พอดีในปริภูมิคุณลักษณะด้วยรัศมีที่มีประสิทธิภาพขั้นต่ำ $R > 0$ จุดศูนย์กลางอยู่ที่จุดกำเนิดครอบคลุมส่วนใหญ่ของอิมเมจเวกเตอร์ x_ϕ ความผิดปกติที่แสดงในชุดข้อมูลแก้ไขได้โดยการ

ใช้ตัวแปรช่วย ξ_μ ในทางคณิตศาสตร์ปัญหาการจัดควอเตอร์สเฟียร์ (Quarter-sphere) ให้เข้ากับข้อมูล ถูกอธิบายด้วยสมการ (Laskov, et al., 2004):

$$\min_{R \in \mathbb{R}^+, \xi \in \mathbb{R}} R^2 + \frac{1}{vn} \sum_{\mu=1}^n \xi_\mu$$

$$\text{ภายใต้เงื่อนไข: } k(x^\mu, x^\mu) \leq R^2 + \xi_\mu, \quad (3.2)$$

เมื่อ ξ_μ คือตัวแปรช่วยที่ยอมให้บางส่วนของเวกเตอร์อิมเมจอยู่ข้างนอกทรงกลม พารามิเตอร์ $v \in (0,1)$ คือพารามิเตอร์ควบคุมที่จะคอยควบคุมเศษของเวกเตอร์อิมเมจที่อยู่ข้างนอกทรงกลมเช่น เศษของเวกเตอร์อิมเมจที่สามารถเป็นความผิดปกติได้

สังเกตว่า $\|\phi(x^\mu)\| = \phi(x^\mu) \cdot \phi(x^\mu)^T = k(x^\mu, x^\mu)$ สำหรับเมอร์เซอร์เคอร์เนล (Mercer kernel) $k(x^\mu, x^\mu)$ คือฟังก์ชันเคอร์เนล (kernel function) ที่ใช้เพื่อคำนวณความเหมือนกันของสองเวกเตอร์ใดๆในปริภูมิคุณลักษณะ โดยใช้ชุดคุณลักษณะดั้งเดิม จากสมการที่ 3.2 ไม่สามารถจะแก้ได้โดยตรงเพราะฉะนั้นจึงใช้วิธีการแก้ปัญหาแบบปัญหาควบคู่ (dual problem) ตามสมการ (Laskov, et al., 2004), (Rajasegarar, et al., 2007):

$$\min_{\alpha \in \mathbb{R}} - \sum_{\mu=1}^n \alpha_\mu k(x^\mu, x^\mu)$$

$$\text{ภายใต้เงื่อนไข: } \sum_{\mu=1}^n \alpha_\mu = 1 \quad (3.3)$$

$$0 \leq \alpha_\mu \leq \frac{1}{vn}$$

ปัญหาควบคู่ในสมการที่ 3.3 คือการแก้ปัญหาค่าเหมาะสมเชิงเส้น (linear optimization problem) เพราะฉะนั้นตัวคูณลากรองซ์ (Lagrangian multiplier) $\{\alpha_\mu\}$ สามารถหาได้จากการใช้เทคนิคการหาค่าที่เหมาะสมเชิงเส้นที่ใช้กันอย่างแพร่หลาย ตัวแปรนี้ถูกใช้เพื่อจัดแบ่งประเภทของเวกเตอร์อิมเมจ $\{\phi(x^\mu)\}$ ดังต่อไปนี้ (แสดงในรูปที่ 1) เวกเตอร์อิมเมจที่ค่า $\alpha_\mu = 0$ จะอยู่ในพื้นที่ของทรงกลม เวกเตอร์อิมเมจที่ $0 < \alpha_\mu < \frac{1}{vn}$ จะอยู่บนพื้นผิวของทรงกลม ด้วยเหตุนี้จึงถูกเรียกว่าขอบเขตของเวกเตอร์ซัพพอร์ต (border support vectors) เวกเตอร์สนับสนุนที่ $\alpha_\mu = \frac{1}{vn}$ จะ

เป็นค่าที่ผิดปกติซึ่งตกอยู่นอกของทรงกลม มากไปกว่านั้นรัศมีของทรงกลม R สามารถหาได้จากการใช้สมการ $R^2 = k(x^\mu, x^\mu)$ สำหรับขอบเขตของเวกเตอร์ซัพพอร์ตใดๆ (Rajasegarar, et al., 2007)

นอกจากนี้วิธีการแก้ปัญหของปัญหาควบคู่ในสมการที่ (3.3) ยังได้รับผลกระทบจากการหาค่าปกติ (norms) ของการแปลงข้อมูลโดยใช้เคอร์เนล $k(x^\mu, x^\mu)$ ที่ไม่เป็นเชิงเส้น ด้วยเหตุนี้จะสร้างปัญหาสำหรับการประยุกต์วิธีการแก้ปัญหานี้กับเคอร์เนลที่อ้างอิงด้วยระยะทาง (distance-based kernel) ซึ่งค่าปกติของเคอร์เนลจะมีค่าเท่ากันสำหรับเวกเตอร์ข้อมูลทุกๆตัว ซึ่งเป็นวิธีการแก้ปัญหที่ไม่มี ความหมายในปัญหาควบคู่ เพื่อที่จะแก้ปัญหานี้เวกเตอร์อิมเมจในปริภูมิคุณลักษณะจะเป็นศูนย์กลางในปริภูมิโดยการลบค่าเฉลี่ยออกจากเวกเตอร์อิมเมจทุกตัว (Laskov, et al., 2004), (Rajasegarar, et al., 2007)

$$\tilde{\phi}(x^\mu) = \phi(x^\mu) - \frac{1}{n} \sum_{\mu=1}^n \phi(x^\mu) \quad (3.4)$$

ผลคูณเชิงสเกลาร์ของเวกเตอร์อิมเมจที่อยู่ศูนย์กลาง $\tilde{\phi}(x^\mu)$ สามารถคำนวณอย่างง่ายโดยใช้ เมตริกซ์ศูนย์กลางเคอร์เนล (center kernel matrix) K_c ดังต่อไปนี้

$$K_c = K - \mathbf{1}_n K - K \mathbf{1}_n + \mathbf{1}_n K \mathbf{1}_n \quad (3.5)$$

เมื่อ $\mathbf{1}_n$ คือเมตริกซ์ $n \times n$ ที่มีค่าทั้งหมดเท่ากับ $\frac{1}{n}$, K คือเคอร์เนลเมตริกซ์ขนาด $n \times n$ ที่ประกอบไปด้วย $k(x^\mu, x^{\bar{\omega}})$ เมื่อ $\mu, \bar{\omega} = 1, 2, 3, \dots, n$ ถ้า $\mu = \bar{\omega}$ จะได้ว่า $k(x^\mu, x^{\bar{\omega}}) = k(x^\mu, x^\mu)$ และดังนั้นจะได้ค่าปกติของเวกเตอร์อิมเมจ $\tilde{\phi}(x^\mu)$ มิฉะนั้น $k(x^\mu, x^{\bar{\omega}})$ สามารถหาได้จาก ฟังก์ชันเคอร์เนลเช่น เคอร์เนลเชิงเส้น เคอร์เนลพหุนาม หรือเคอร์เนล RBF (RBF kernel) เมื่อเวกเตอร์อิมเมจเป็นจุดศูนย์กลาง ค่าปกติของเคอร์เนลจะไม่มีเท่ากันอีกต่อไป ด้วยเหตุนี้ปัญหาควบคู่ในสมการ (3.3) จะสามารถแก้ปัญหได้อย่างง่ายดาย

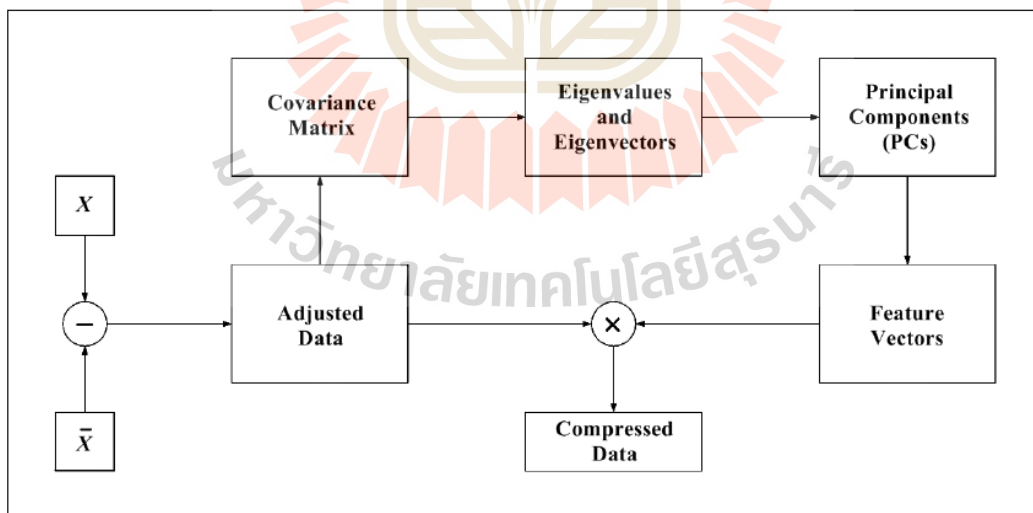
3.3 การบีบอัดข้อมูล

3.3.1 Principal Component Analysis (PCA)

หลักการวิเคราะห์องค์ประกอบหลัก (PCA) ถูกนำเสนอเพื่อเครื่องมือในการลดมิติของชุดข้อมูล (Kiziloren, et al., 2009) PCA คือหลักการทางสถิติแบบดั้งเดิมซึ่งสามารถย้อนกลับ

ได้อย่างสมบูรณ์ (ข้อมูลต้นฉบับอาจถูกกู้คืนได้โดยตรงจาก PCs) เป็นเครื่องมืออเนกประสงค์ มีประโยชน์สำหรับลดปริมาณข้อมูล กำจัดสัญญาณรบกวน การสร้างภาพและการบีบอัดข้อมูล (Dwinnell, Online, 2010) อย่างไรก็ตาม PCA ขอมรับที่จะไม่สนใจบางองค์ประกอบของข้อมูลที่มีความสำคัญน้อยโดยการพิจารณาค่าลักษณะเฉพาะ (eigenvalue) ของเมทริกซ์ความแปรปรวนของข้อมูล ถ้าค่าลักษณะเฉพาะของข้อมูลมีค่าน้อยข้อมูลนั้นจะถูกทิ้งซึ่งจะทำให้ข้อมูลบางส่วนหายไป (Smith, 2002)

รูปแบบการทำงานในรูปที่ 3.2 สามารถอธิบายได้ดังนี้ ขั้นตอนแรกจะทำการค้นหาค่าเฉลี่ยของชุดข้อมูลอินพุต X และนำค่าเฉลี่ยที่ได้ลบออกจากข้อมูลอินพุต X และให้ผลลัพธ์เป็นข้อมูลที่ถูกปรับเปลี่ยนแล้ว หลังจากนั้นจะทำการหาเมทริกซ์ความแปรปรวนของข้อมูลที่ถูกปรับเปลี่ยนแล้ว เพื่อที่จะทำการหาค่าเวกเตอร์ลักษณะเฉพาะและค่าลักษณะเฉพาะของเมทริกซ์ความแปรปรวนตามลำดับ หลังจากนั้นจะทำการเลือกค่าลักษณะเฉพาะที่มีค่ามากที่สุดเป็น PCs ตัวแรก PCs ลำดับถัดไปจะถูกเลือกจากค่าลักษณะเฉพาะที่มีค่าสูงสุดลำดับถัดไป จากนั้นนำ PCs ที่ทำการสลับเปลี่ยนแล้ว (transpose) เรียกว่า เวกเตอร์คุณลักษณะ (feature vector) มาคูณกับข้อมูลที่ถูกปรับเปลี่ยนแล้ว สุดท้ายแล้วจะได้ข้อมูลที่ถูกบีบอัดที่มีจำนวนเท่ากับข้อมูลแต่ KPIs น้อยกว่าชุดข้อมูลต้นฉบับ X



รูปที่ 3.2 รูปแบบการทำงานของหลักการวิเคราะห์องค์ประกอบหลัก (PCA)

3.3.2 การแปลงเวฟเล็ตแบบเต็มหน่วย

การแปลงเวฟเล็ตแบบเต็มหน่วย (Discrete Wavelet Transform หรือ DWT) คือการแปลงทางคณิตศาสตร์ที่แยกสัญญาณข้อมูลเป็นการแปลงเวฟเล็ตแบบเต็มหน่วยที่รู้จักกันว่าสัมประสิทธิ์รายละเอียด (detail coefficients) และข้อมูลที่มีมาตราส่วนแบบหยาบ (rough-scale information) ที่รู้จักในชื่อ ค่าสัมประสิทธิ์ประมาณ (approximate coefficients) โดยใช้วิธีการคอนโวลูชัน (convolution) ข้อได้เปรียบหลักของ DWT คือการนำเสนอด้วยความละเอียดที่หลากหลาย (multi-resolution representation) และคุณสมบัติการจำกัดวงทางความถี่และเวลาของสัญญาณ (Siripanadorn, et al., 2010) โดยปกติการร่างชุดเวลาดั้งเดิมสามารถกลับไปตั้งต้นใหม่ได้ได้ด้วยการใช้ค่าสัมประสิทธิ์การจำแนกต่ำผ่าน (low-pass-cut off decomposition coefficients) สามารถจำลองจากค่าสัมประสิทธิ์การจำแนกระดับกลาง (middle-level decomposition coefficients) ส่วนที่เหลือจะถูกพิจารณาเป็นสัญญาณรบกวนหรือความผิดปกติ สมการต่อไปนี้อธิบายการคำนวณในกระบวนการแจกแจงของ DWT (Siripanadorn, et al., 2010)

$$a_{j+1}^{DWT}(f) = \sum_n h_0(n - 2f) \cdot a_j^{DWT}(f) \quad (3.6)$$

$$d_{j+1}^{DWT}(f) = \sum_n g_0(n - 2f) \cdot a_j^{DWT}(f) \quad (3.7)$$

เมื่อค่าสัมประสิทธิ์สเกลหยาบ (หรือ ค่าประมาณ) a_j^{DWT} ถูกรวมเข้าด้วยกัน ด้วย h_0 และ g_0 ซึ่งเป็นฟังก์ชันเวฟเล็ตและฟังก์ชันสเกล (scaling function) ตามลำดับ n คือดัชนีสเกลเวลา (time scaling index) f คือดัชนีการแปลงค่าทางความถี่ (frequency translation index) สำหรับเวฟเล็ตระดับ j ค่าสัมประสิทธิ์ผลลัพธ์จะถูกลดขนาดลงด้วย 2 กระบวนการนี้จะแยก a_{j+1}^{DWT} ครั้งหนึ่งอย่างหยาบ โดยแยกให้อยู่ในเซตของสเกลที่ดี หรือสัมประสิทธิ์แบบละเอียด d_{j+1}^{DWT} และเซตสเกลหยาบของสัมประสิทธิ์แบบค่าประมาณ a_{j+1}^{DWT}

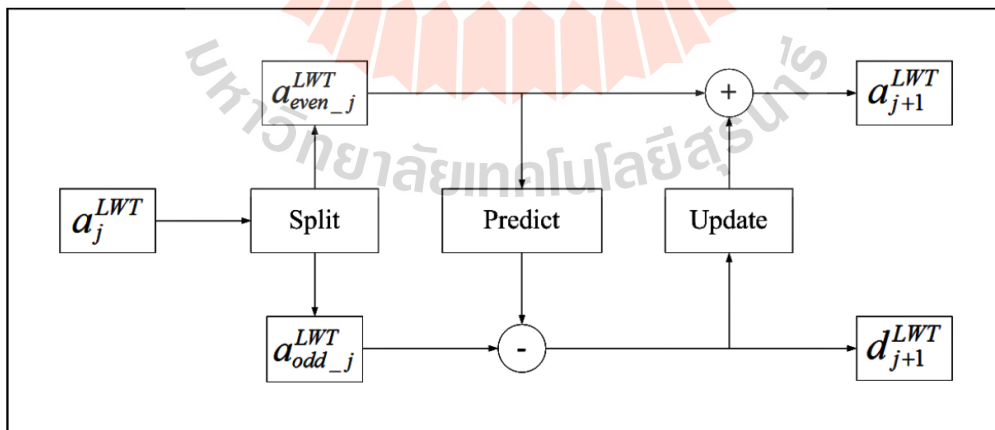
สำหรับการบีบอัดข้อมูล DWT จะลดขนาดของข้อมูลลงครึ่งหนึ่ง ในขณะที่ยังคงขนาดของ KPIs ไว้ขนาดเท่าเดิม DWT มีความสามารถในการเข้ารหัสความละเอียดปลีกย่อยของชุดเวลาดั้งเดิมกับค่าสัมประสิทธิ์ลำดับชั้น นอกจากนี้ DWT ยังมีความสามารถคำนวณแบบเวลาเชิงเส้นได้อย่างมีประสิทธิภาพ ซึ่งมีความสำคัญอย่างยิ่งในการรับมือกับชุดข้อมูลขนาดใหญ่ อย่างไรก็ตาม DWT ที่ใช้วิธีการคอนโวลูชันต้องการการคำนวณทางคณิตศาสตร์หน่วยการเก็บข้อมูลที่มากกว่า LWT (Achaya, et al., 2006)

3.3.3 Lifting Wavelet Transform (LWT)

LWT ถูกนำเสนอเป็นยุคที่สองของเวฟเลต (Sweldens, 1998) LWT สืบทอดคุณสมบัติหลายลำดับความละเอียดมาจากยุคแรกของเวฟเลตและคุณลักษณะหลักของ LWT การแปลงการดำเนินการคอนโวลูชันของ DWT ไปเป็นการคำนวณเมทริกซ์แถบ (band matrix) คุณด้วย in-place เพราะฉะนั้น LWT ต้องการพื้นที่หน่วยความจำและการคำนวณที่น้อยกว่า DWT (Sweldens, 1998) (Achaya, et al., 2006), (X. L. Li, et al., 2009)

จากรูปที่ 3.3 อัลกอริทึม LWT แบ่งชุดข้อมูลอินพุต $X = \{x^\mu : \mu = 1, 2, 3, \dots, n\}$ หรือ ค่าสัมประสิทธิ์กระแส rough-scale (หรือการประมาณ) a_j^{LWT} เป็น 3 ขั้นตอนดังต่อไปนี้:

- 1) แยก: ชุดข้อมูลอินพุตหรือค่าสัมประสิทธิ์กระแส rough-scale (หรือการประมาณ) a_j^{LWT} ถูกแยกเป็นครึ่งคู่ $a_{even_j}^{LWT}$ และครึ่งคี่ $a_{odd_j}^{LWT}$
- 2) ทำนาย: ครึ่งคี่คือการทำนายโดยการลบการรวมกันแบบเชิงเส้นกับครึ่งคู่ ผลลัพธ์อยู่ในรูปของการทำนายความผิดพลาดหรือค่าสัมประสิทธิ์ผ่านสูงหรือค่าสัมประสิทธิ์รายละเอียด d_{j+1}^{DWT}
- 3) ปรับปรุง: ครึ่งคู่ถูกปรับปรุงโดยการเพิ่มตัวของมันเองเข้าไปในการรวมกันแบบเชิงเส้นกับการทำนายความผิดพลาด ผลลัพธ์อยู่ในรูปของค่าสัมประสิทธิ์ผ่านต่ำหรือค่าสัมประสิทธิ์การประมาณ a_{j+1}^{DWT}



รูปที่ 3.3 รูปแบบการทำงานของ Lifting Wavelet Transform (LWT)

สังเกตว่า LWT ใช้การรวมกันแบบเชิงเส้นเพื่อหาค่าสัมประสิทธิ์ของเวฟเลต ขณะที่ DWT ใช้การคอนโวลูชัน แสดงให้เห็นว่า LWT มีความละเอียดน้อยกว่า DWT (Manjunath, et al. 2010)

3.4 ผลการทดลอง

ในส่วนนี้เราใช้โปรแกรม MATLAB ในการประเมินคุณภาพเชิงตัวเลขของ 4 อัลกอริทึมการบีบอัดข้อมูล ได้แก่ อัลกอริทึม OCSVM+LWT ซึ่งเป็นอัลกอริทึมที่น่าเสนอ, เปรียบเทียบกับอัลกอริทึม OSCVM เพียงอย่างเดียว (ข้อมูลที่ไม่บีบอัด), อัลกอริทึม OCSVM+PCA (Kiziloren, et al., 2009) และอัลกอริทึม OSCVM+DWT (Takianggam, et al., 2011) โดยใช้ตรวจจับความผิดปกติในชุดข้อมูลที่ได้จากการวัดจริงและข้อมูลสังเคราะห์

ในงานวิจัยนี้ใช้เคอร์เนลเชิงเส้นที่เป็นเคอร์เนลที่อ้างอิงด้วยระยะทาง (distance-based kernel) ฟังก์ชันเคอร์เนลเชิงเส้น สำหรับเวกเตอร์ข้อมูล x^μ และ $x^{\bar{\omega}}$ ถูกกำหนดโดย $k_{linear}(x^\mu, x^{\bar{\omega}}) = \phi(x^\mu) \cdot \phi(x^{\bar{\omega}})$

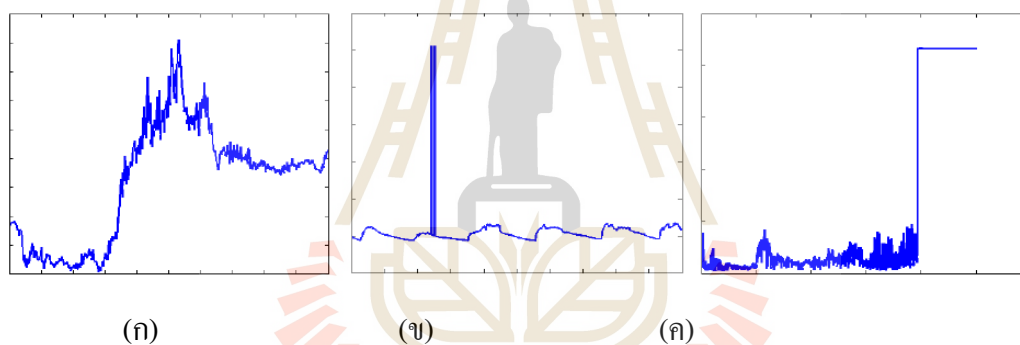
ในแต่ละการจำลอง ได้ทำการเปลี่ยนแปลงพารามิเตอร์ ν ในสมการที่ (3.3) ตั้งแต่ 0.02 ถึง 1 โดยเพิ่มทีละ 0.02 หลังจากนั้นทำการบันทึกค่า false positives ซึ่งเกิดขึ้นเมื่อการวัดปกติถูกระบุว่าเป็นความผิดปกติที่ถูกตรวจจับ และ true positives ซึ่งเกิดขึ้นเมื่อการวัดความผิดปกติถูกระบุว่าเป็นความผิดปกติที่ถูกตรวจจับ อัตรา false positive rate (FRT) จะถูกคำนวณ โดยเปอร์เซ็นต์ที่เป็นอัตราส่วนระหว่างความเท็จเชิงบวกและการวัดปกติที่เกิดขึ้นจริง และอัตราการตรวจจับ (detection, DR) คำนวณจากเปอร์เซ็นต์ที่เป็นอัตราส่วนระหว่างความจริงเชิงบวกกับการวัดปกติที่เกิดขึ้นจริง สุดท้ายทำการแสดงผลการทดลองในรูปแบบเส้นโค้ง Receiver Operating Characteristics (ROC)

3.4.1 ชุดข้อมูลสำหรับการทดลอง

การวิจัยนี้จำแนกความผิดพลาดออกเป็น 3 ประเภทแสดงในรูปที่ 3.3 กล่าวคือ ความผิดปกติจากสัญญาณรบกวน (noise faults) ความผิดปกติระยะสั้น (short faults) และความผิดปกติคงที่ (constant faults) (Sharma, et al., 2010) ความผิดปกติจากสัญญาณรบกวนคือความผิดปกติที่เกิดขึ้นเมื่อความแปรปรวนของการอ่านข้อมูลของเซ็นเซอร์ที่สูงขึ้นและส่งผลกระทบต่อจำนวนตัวอย่างต่อเนื่อง (successive samples) ความผิดปกติระยะสั้นคือการเปลี่ยนแปลงแบบพุ่งสูงขึ้นของ

ค่าที่วัดได้ระหว่างข้อมูลที่ต่อเนื่องและส่งผลกระทบต่อตัวอย่างที่เวลาหนึ่ง ความผิดปกติก็คือ ความผิดปกติที่เกิดจากค่าคงที่จำนวนมากของการสุ่มตัวอย่างแบบต่อเนื่อง

การวิจัยนี้จะใช้ทั้งชุดข้อมูลแบบสังเคราะห์และชุดข้อมูลที่ได้จากการวัดข้อมูลจริง โดยชุดข้อมูลที่ได้จากการวัดข้อมูลจริง 3 ชุดจะถูกนำมาใช้ในการประเมินประสิทธิภาพ ได้แก่ ชุดข้อมูล INTEL (The Intel Lab, Online, 2004) ชุดข้อมูล SensorScope (The SensorScope Lausanne Urban Canopy Experiment Project (LUCE), Online, 2006) และ ชุดข้อมูล NAMOS (Network Aquatic Microbial Observing System, Online, 2006)

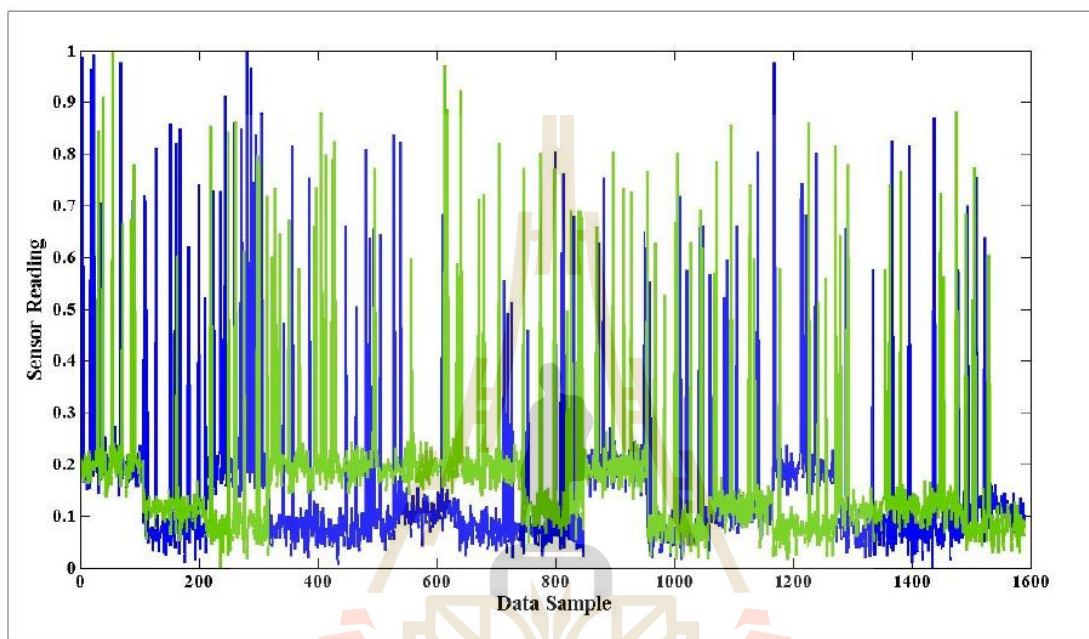


รูปที่ 3.4 ความผิดปกติในการอ่านข้อมูลของเซ็นเซอร์

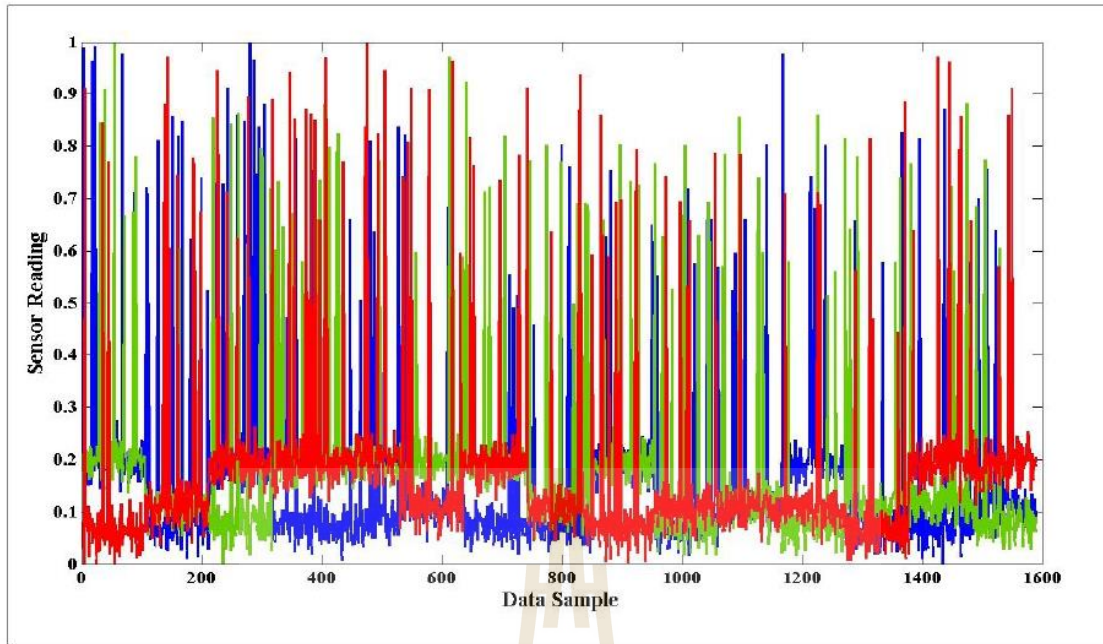
3.4.1.1 ข้อมูลสังเคราะห์

ข้อมูลสังเคราะห์ในรูปที่ 4.5-4.10 ถูกสร้างจากการรวมกันของการแจกแจงแบบเกาส์เซียน(Gaussian distributions) กับค่าเฉลี่ยที่สุ่มเลือกจาก (0.3, 0.35, 0.45) และค่าเบี่ยงเบนมาตรฐานคือ 0.03 โดยใช้โปรแกรม MATLAB ในการจำลอง ข้อมูลถูกสร้างสำหรับ 15 โหนดเซ็นเซอร์และมี 2 ข้อมูล KPIs สำหรับเวกเตอร์ข้อมูล 160 ตัวต่อหนึ่งโหนดเซ็นเซอร์ รวมแล้วเป็นเวกเตอร์ข้อมูลทั้งหมด 1590 ชุด หลังจากนั้นนำเข้าตัวเลขความผิดปกติที่มีการแจกแจงแบบปกติอยู่ในช่วงระหว่าง [0.5, 1] สำหรับแต่ละ KPIs ของข้อมูล จำนวนของความผิดปกติถูกแทนด้วยสัญลักษณ์ a/s เมื่อ a จำนวนความผิดปกติต่อชุดข้อมูล และ s คือจำนวนชุดข้อมูลที่มีความผิดปกติ ผลรวมของความผิดปกติจะเป็น $a \times s$ ความผิดปกติที่ถูกสร้างขึ้นจะถูกป้อนเข้าไปในข้อมูลอินพุต

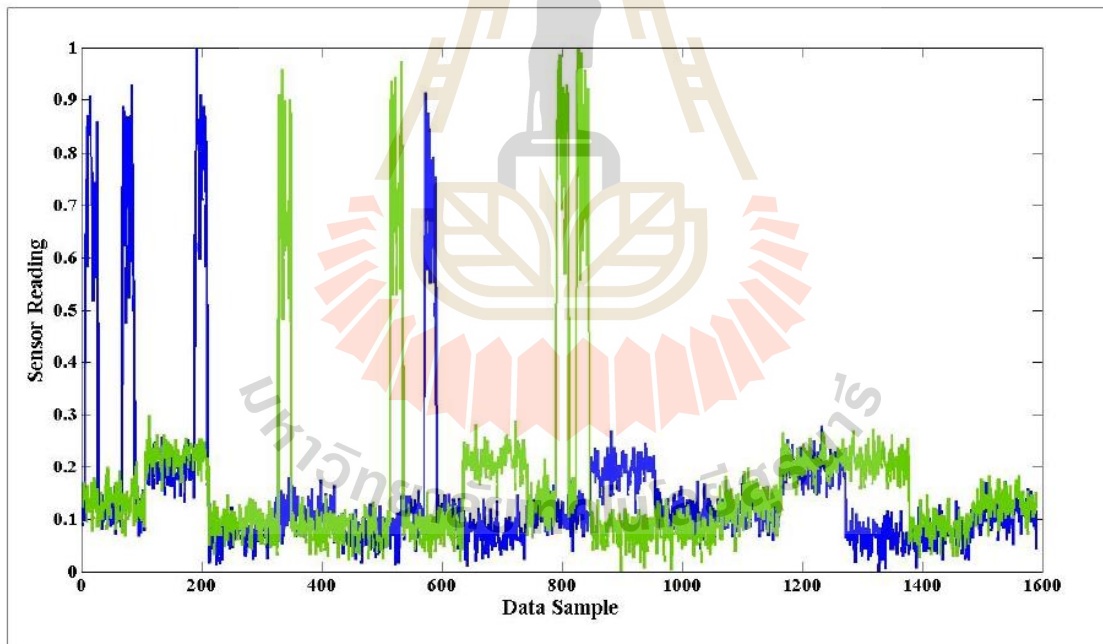
จะประกอบไปด้วยความผิดพลาดที่ $(80/1)$, จากนั้นความผิดพลาดจากสัญญาณรบกวน $(20/4)$, และสุดท้ายความผิดพลาดระยะสั้น $(1/80)$ โดยความผิดพลาดทุกชนิดรวมกันมีค่าความผิดพลาดทั้งหมด 80 ความผิดพลาด ชุดข้อมูลทั้งหมดจะถูกทำให้อยู่ในรูปแบบมาตรฐานให้อยู่ในช่วง $[0, 1]$ ตำแหน่งที่แน่นอนของความผิดพลาดถูกใส่เข้าไปในข้อมูลอินพุตที่กำหนดไว้เพื่อใช้ในการตรวจจับการแจ้งเตือนความผิดพลาดที่ถูกต้องและผิด



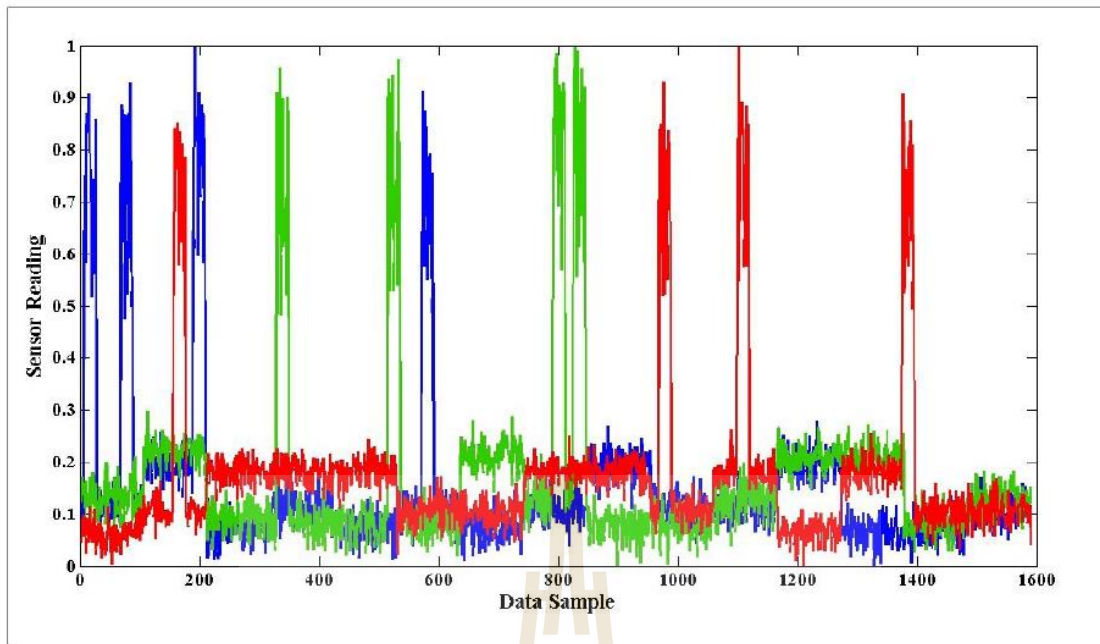
รูปที่ 3.5 ข้อมูลสังเคราะห์ 2KPI กับความผิดพลาด 1/80



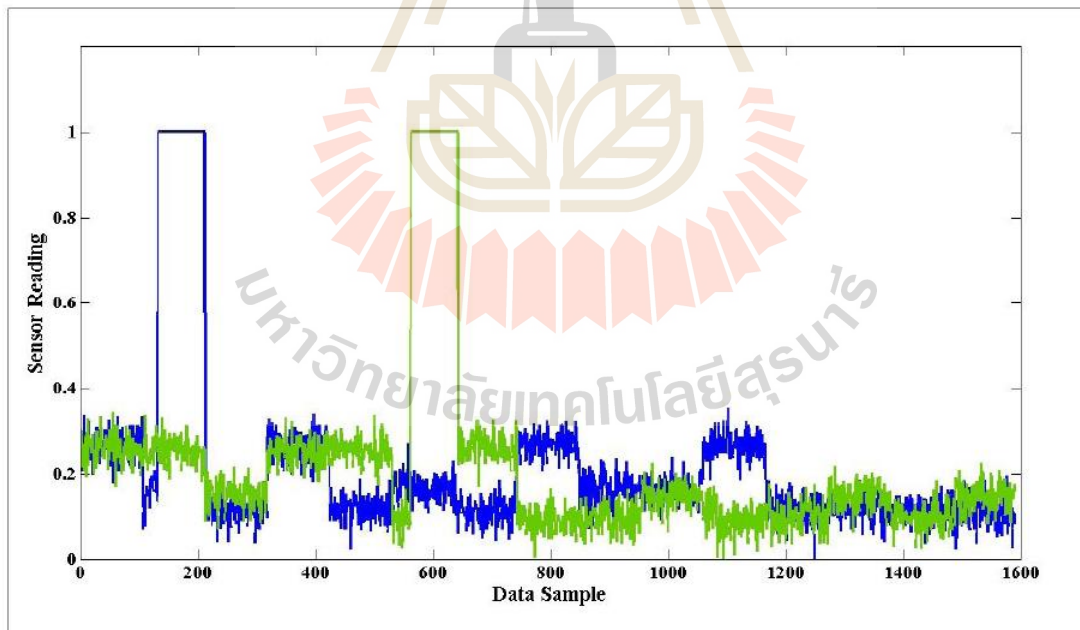
รูปที่ 3.6 ข้อมูลสังเคราะห์ 3KPI กับความผิดพลาด 1/80



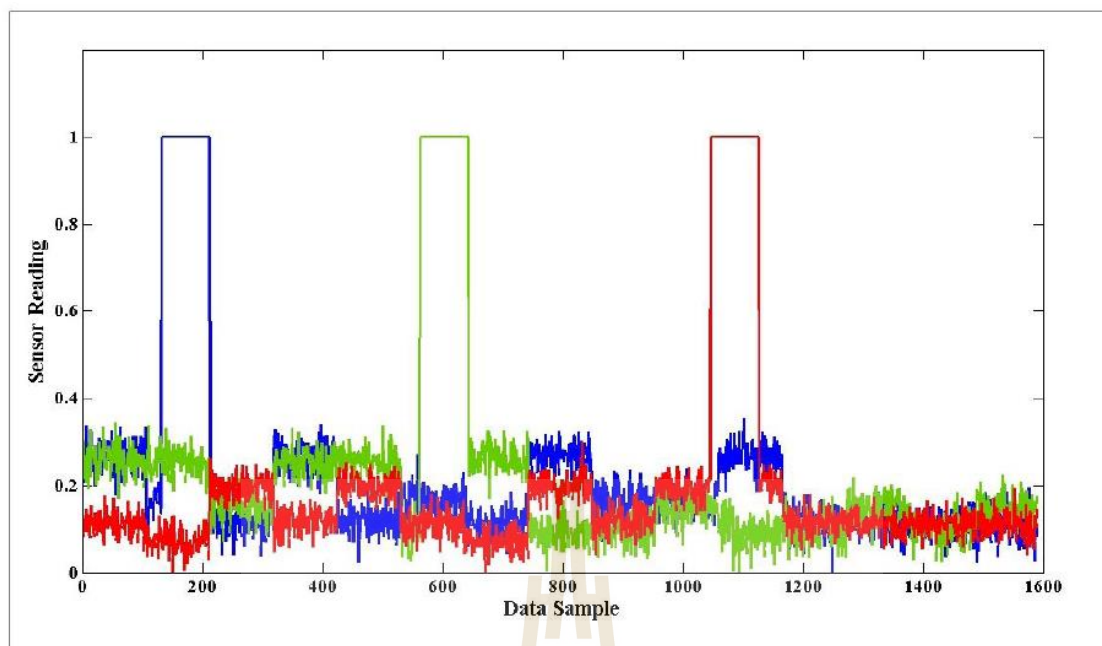
รูปที่ 3.7 ข้อมูลสังเคราะห์ 2KPI กับความผิดพลาด 20/4



รูปที่ 3.8 ข้อมูลสังเคราะห์ 3KPI กับความผิดพลาด 20/4



รูปที่ 3.9 ข้อมูลสังเคราะห์ 2KPI กับความผิดพลาด 80/1

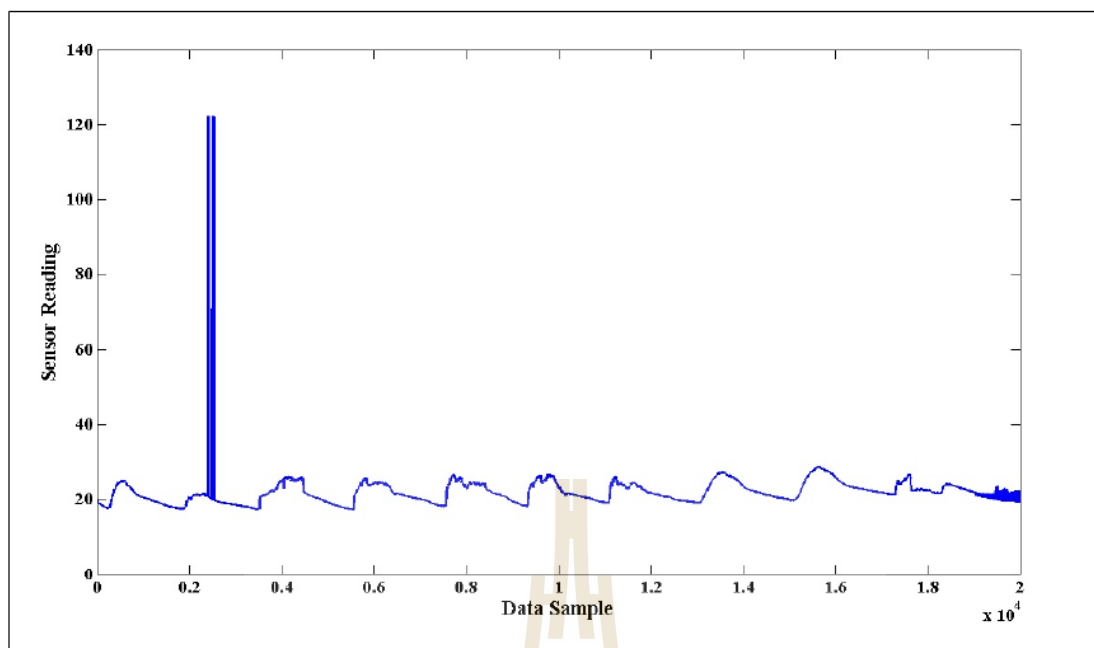


รูปที่ 3.10 ข้อมูลสังเคราะห์ 3KPI กับความผิดพลาด 80/1

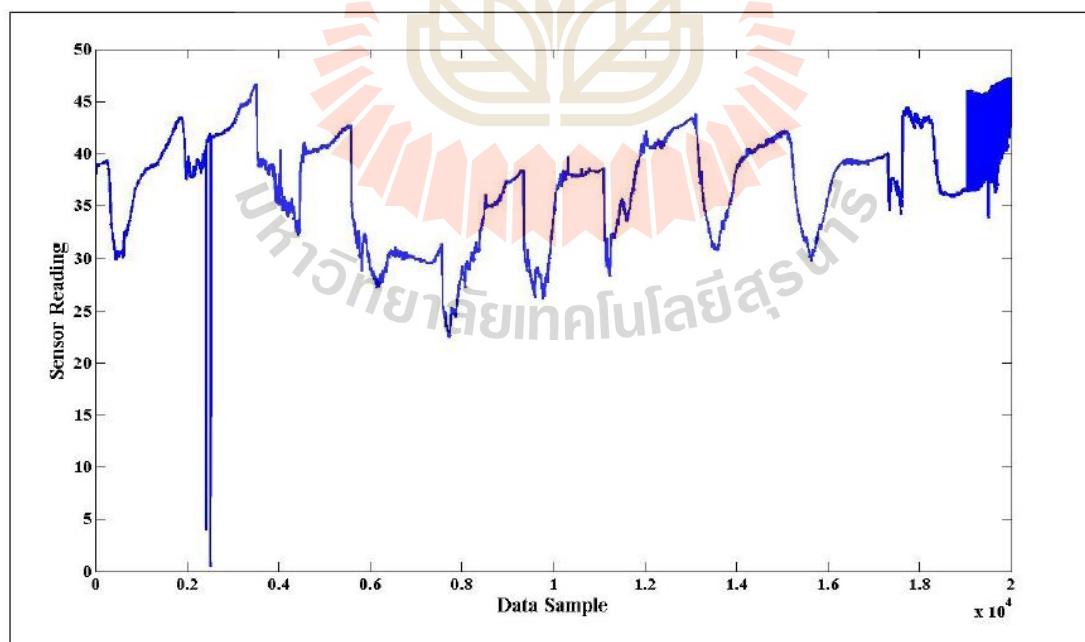
3.4.1.2 ชุดข้อมูลที่ได้จากการวัดจริง

สำหรับข้อมูลที่ได้จากการวัดจริง เนื่องจากไม่มีข้อมูลภาคพื้นจริงเกี่ยวกับความผิดปกติในแต่ละชุดข้อมูล ดังนั้นจึงใช้หลักการสวิตช์โปรแกรมเพื่อแยกข้อมูลปกติออกจากข้อมูลที่ไม่ปกติ

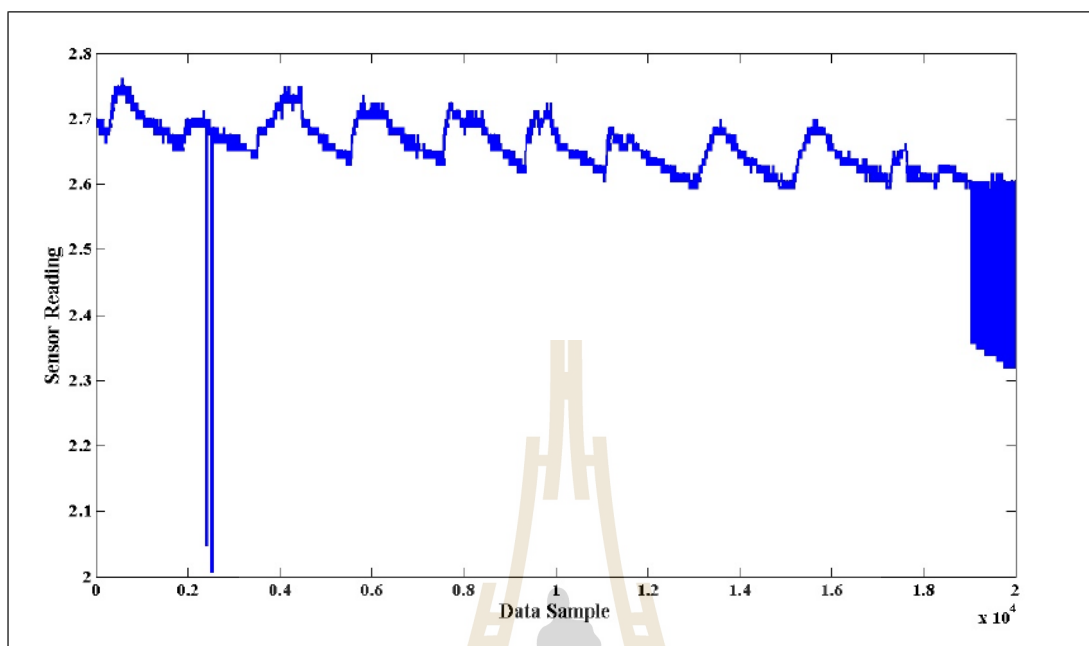
1) ชุดข้อมูล INTEL: Mica2Dot จำนวน 54 ตัวที่มีการวัดอุณหภูมิ, ความชื้น, แสงสว่างและแรงดันถูกใช้ใน ห้องปฏิบัติการวิจัย Intel Berkeley ระหว่าง วันที่ 28 กุมภาพันธ์และวันที่ 5 เมษายน ปี 2004 (The Intel Lab, Online, 2004) ในการทดลอง ทำการเลือก 20,000 ตัวอย่างจาก อุณหภูมิ, ความชื้นและแรงดันที่อ่านค่าได้ ในส่วนแรกของการทดลองได้นำเสนอผลลัพธ์ของการตรวจจับความผิดปกติในการอ่านค่าอุณหภูมิโดยเลือกค่าเทรชโฮลด์ (threshold) โดยที่ 16 เป็นขอบบนและ 30 เป็นขอบล่างของขอบเขตข้อมูลปกติ ส่วนที่สอง นำเสนอค่าความชื้นและแรงดันที่อ่านค่าได้ โดยเลือกค่าเทรชโฮลด์ โดยที่ 24 และ 47 เป็นขอบบนและขอบล่างสำหรับขอบเขตความชื้นที่อ่านค่าได้ และ 2.5 และ 2.8 สำหรับค่าแรงดันที่อ่านได้ จากการใช้หลักการสวิตช์โปรแกรม พบว่าข้อมูลชุดนี้เป็นความผิดปกติระยะสั้น



รูปที่ 3.11 ชุดข้อมูล INTEL (อุณหภูมิที่อ่านได้)

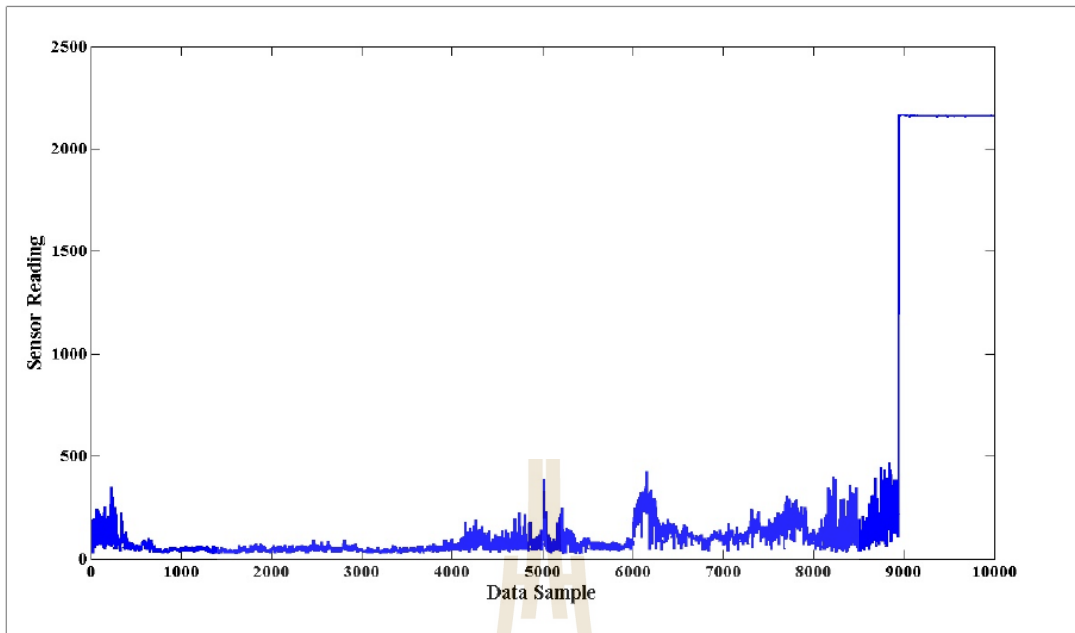


รูปที่ 3.12 ชุดข้อมูล INTEL (ความชื้นที่อ่านได้)



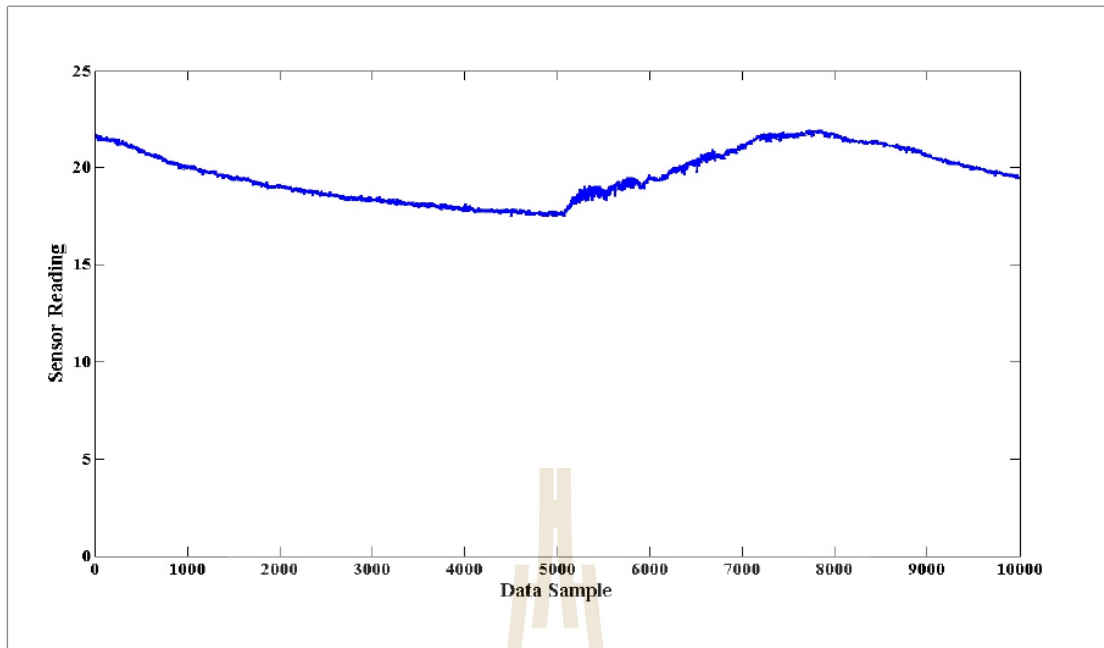
รูปที่ 3.13 ชุดข้อมูล INTEL (แรงดันที่อ่านได้)

2) ชุดข้อมูล NAMOS: ข้อมูลชุดนี้ ได้มาจากทุ่นลอยน้ำ 9 ตัวกับ เซ็นเซอร์วัดอุณหภูมิและเซ็นเซอร์วัดความเข้มของคลอโรฟิลล์ (เครื่องวัดแสงฟลูออเรสเซนซ์) ถูกใช้ในทะเลสาบ Fulmor ตลอด 24 ชั่วโมงในเดือนสิงหาคมปี 2006 (Network Aquatic Microbial Observing System, Online, 2006) ทำการวิเคราะห์ 10,000 ตัวอย่างที่วัดแสงฟลูออเรสเซนซ์บนทุ่นหมายเลข 103 เลือกค่าเทรสโฮลด์ 0 และ 500 เป็นขอบล่างและขอบบนของขอบเขตปกติตามลำดับ จากการพิจารณาตำแหน่งของข้อมูลที่ผิดปกติพบว่าข้อมูลชุดนี้เป็นความผิดปกติครั้งที่ ในช่วงที่สองทำการเลือกค่าที่อ่านได้จาก 2 เซ็นเซอร์วัดอุณหภูมิและให้ข้อมูลนั้นเป็นข้อมูลปกติเนื่องจากหลักการอิตโตแกรมไม่สามารถแยกขอบเขตความปกติและความผิดปกติได้อย่างชัดเจน

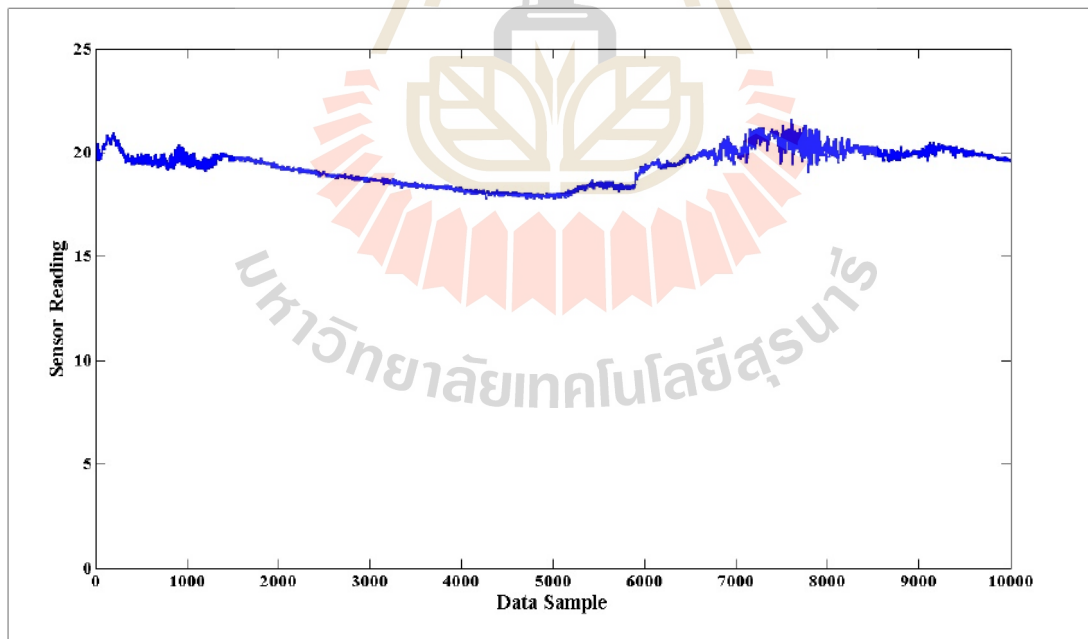


รูปที่ 3.14 ชุดข้อมูล NAMOS (แสดงฟลูออเรสเซนซ์ที่อ่านได้)



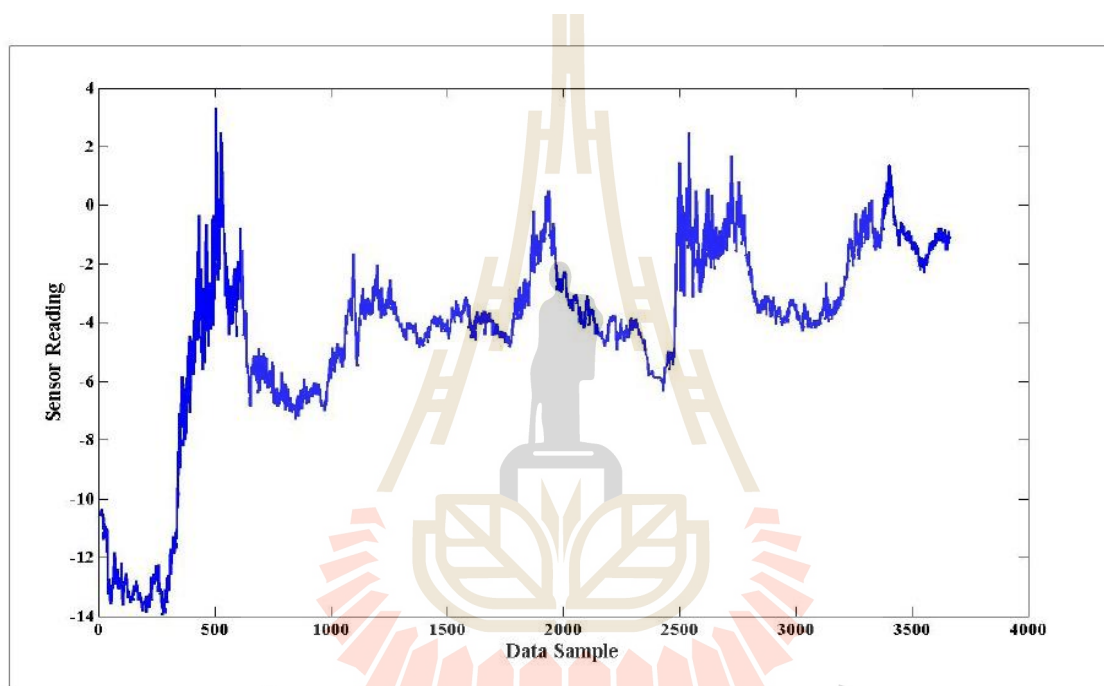


รูปที่ 3.15 ชุดข้อมูล NAMOS (อุณหภูมิที่อ่านได้จากเซ็นเซอร์ 1)

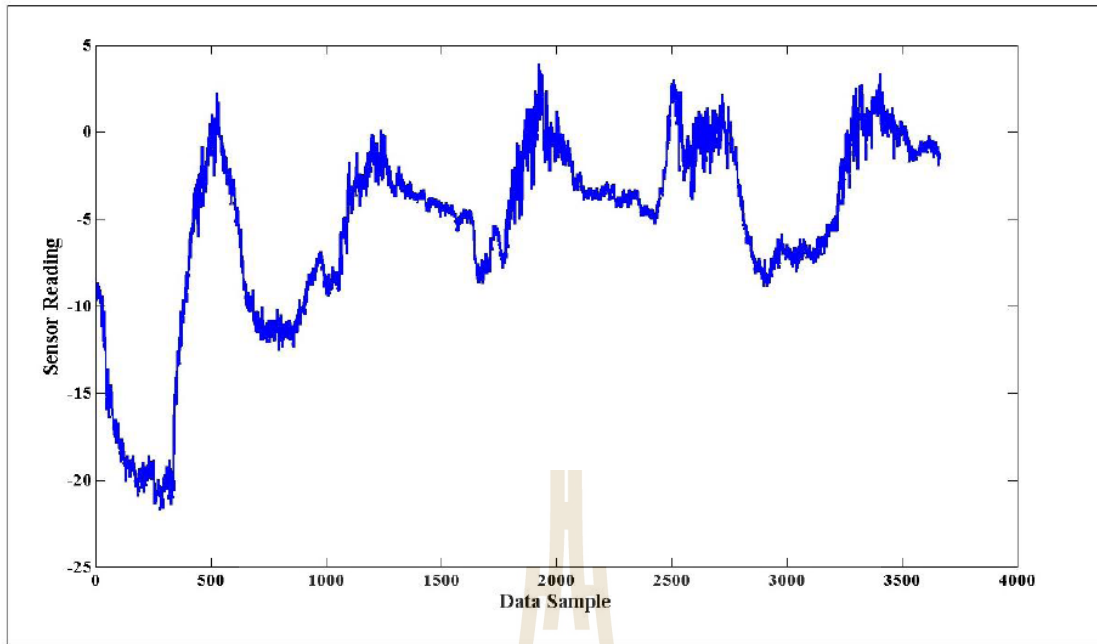


รูปที่ 3.16 ชุดข้อมูล NAMOS (อุณหภูมิที่อ่านได้จากเซ็นเซอร์ 2)

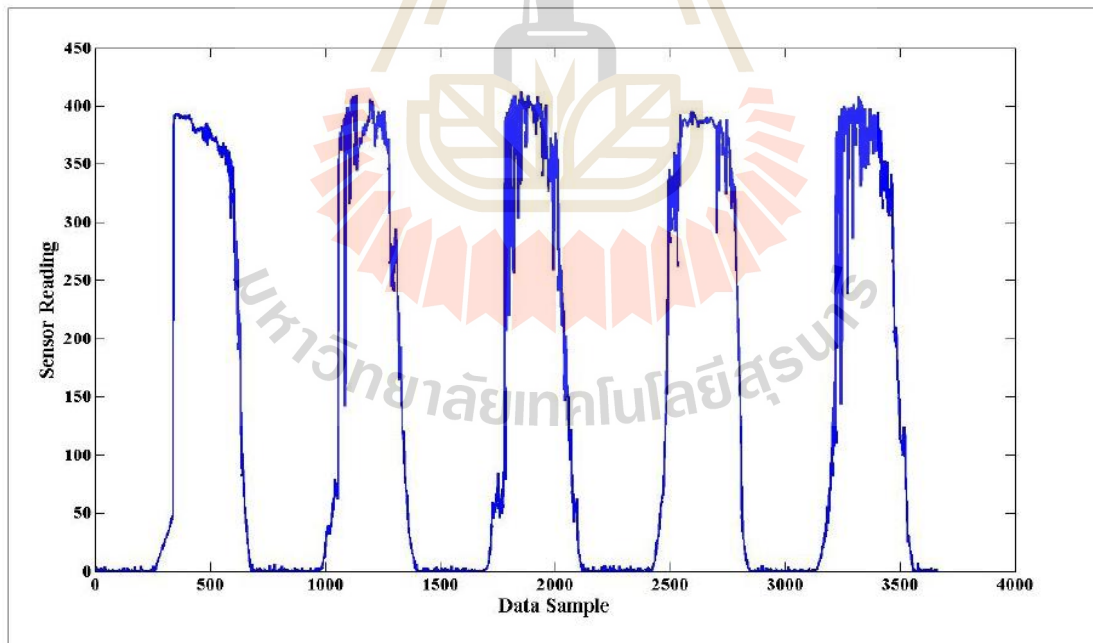
3) ชุดข้อมูล SensorScope pdg2008-metro-1 (pdg2008): ในส่วนแรกของการทดลอง ใช้ KPIs 2 ชนิดในข้อมูล pdg2008 สำหรับการตรวจจับความผิดปกติ เช่น การอ่านอุณหภูมิพื้นผิวและอุณหภูมิโดยรอบ ใช้ -12 และ 4 เป็นขอบบนและขอบล่างของค่าเทอร์สโพลด์สำหรับอุณหภูมิโดยรอบ และ -14 และ 4 สำหรับอุณหภูมิพื้นผิว จากการพิจารณาดำแหน่งของข้อมูลที่ผิดปกติ พบว่าข้อมูลชุดนี้ประกอบไปด้วยความผิดปกติจากสัญญาณรบกวน ในส่วนที่สอง ใช้ข้อมูลที่อ่านได้จากการแผ่รังสีดวงอาทิตย์เป็นข้อมูลปกติ เนื่องจากวิธีการทางฮิโดแกรมนั้นไม่สามารถแยกขอบเขตที่ผิดปกติและไม่ปกติได้อย่างชัดเจน



รูปที่ 3.17 ชุดข้อมูล pdg2008 (อุณหภูมิโดยรอบที่อ่านได้)



รูปที่ 3.18 ชุดข้อมูล pdg2008 (อุณหภูมิพื้นผิวที่อ่านได้)



รูปที่ 3.19 ชุดข้อมูล pdg2008 (รังสีดวงอาทิตย์ที่อ่านได้)

3.4.2 การประเมินประสิทธิภาพ

การวิจัยนี้นำเสนอการรวมกันระหว่างการตรวจจับความผิดปกติแบบ OCSVM และการบีบอัดข้อมูลโดยใช้ LWT และเปรียบเทียบกับอัลกอริทึม OCSVM+DWT และอัลกอริทึม OSCVM+ PCA สำหรับอัลกอริทึม OCSVM+LWT และอัลกอริทึม OSCVM+DWT ทำการแทนชุดข้อมูลต้นฉบับของข้อมูลอินพุตด้วยสัมประสิทธิ์ผ่านต่ำหรือผ่านสูงของ LWT โดยใช้วิธีการ Haar mother wavelet ค่าสัมประสิทธิ์เวฟเลตผ่านต่ำ (สูง) ได้ถูกเรียกว่า “ข้อมูลผ่านต่ำ (สูง)” ด้วยจำนวน KPIs ที่มีจำนวนเท่ากันและมีความยาว (นั่นคือความยาวที่สังเกต) เพียงครึ่งหนึ่งของเวกเตอร์ข้อมูลต้นฉบับ โดยข้อมูลต้นฉบับคือข้อมูลที่ยังไม่ถูกบีบอัด

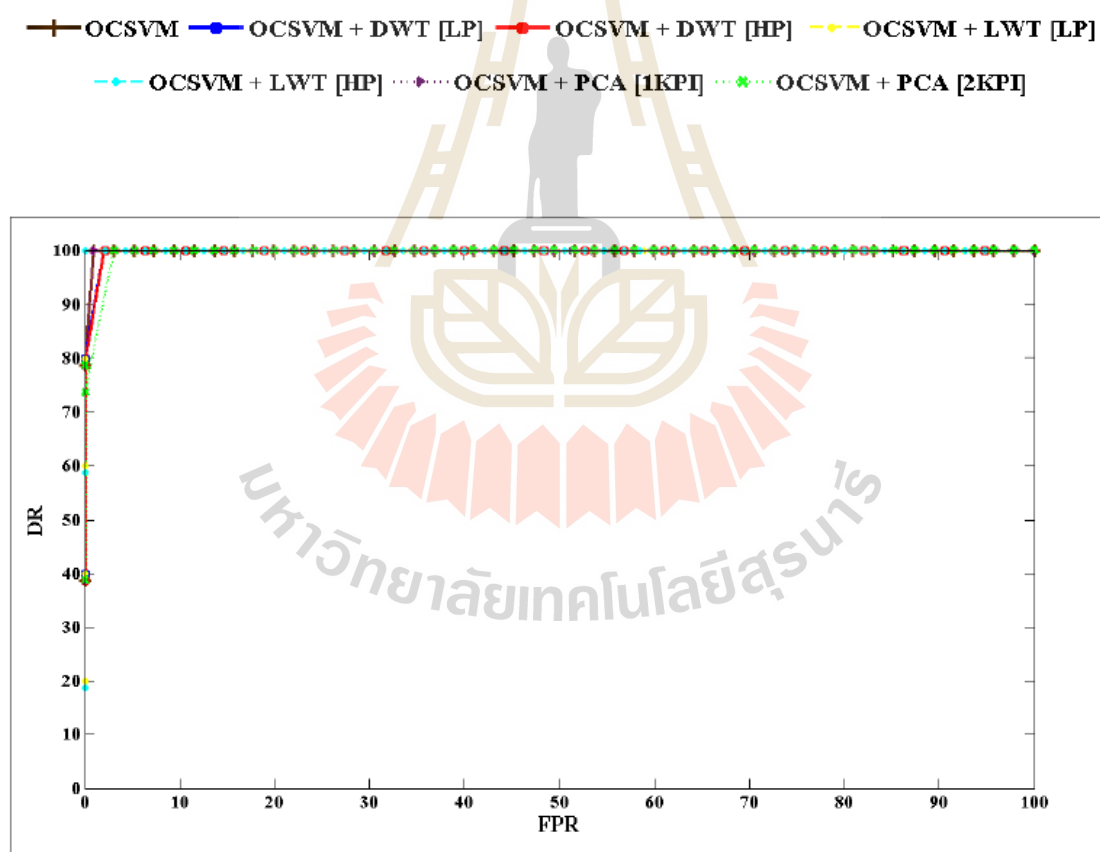
สำหรับอัลกอริทึม OSCVM กับ PCA ถูกแทนด้วยชุดข้อมูลต้นฉบับของข้อมูลอินพุตกับชุดข้อมูล PCA จำนวน 1 KPI (2KPI, 3KPI) ซึ่งมี 1 KPI (2KPI, 3KPI) กับความยาวที่สังเกตเท่ากับกับเวกเตอร์ข้อมูลต้นฉบับ โดยเรียกวิธีการว่าข้อมูล PCA 1 KPI (2KPI, 3KPI)

รูปที่ 3.20 แสดงชุดข้อมูลสังเคราะห์ 2KPI ถูกใส่ความผิดปกติระยะสั้นเข้าไป รูปที่ 3.21 แสดงผลลัพธ์ของชุดข้อมูล INTEL 1KPI ที่ประกอบด้วยความผิดปกติระยะสั้น ซึ่งสอดคล้องกับผลลัพธ์สำหรับข้อมูลสังเคราะห์ที่ใส่ความผิดปกติระยะสั้น 1/80 เข้าไป โดยที่อัลกอริทึมทั้งหมดทำงานได้ดีเท่าๆกันเพราะความผิดปกติระยะสั้นมีแอมพลิจูดที่สูงต่อการตรวจจับ

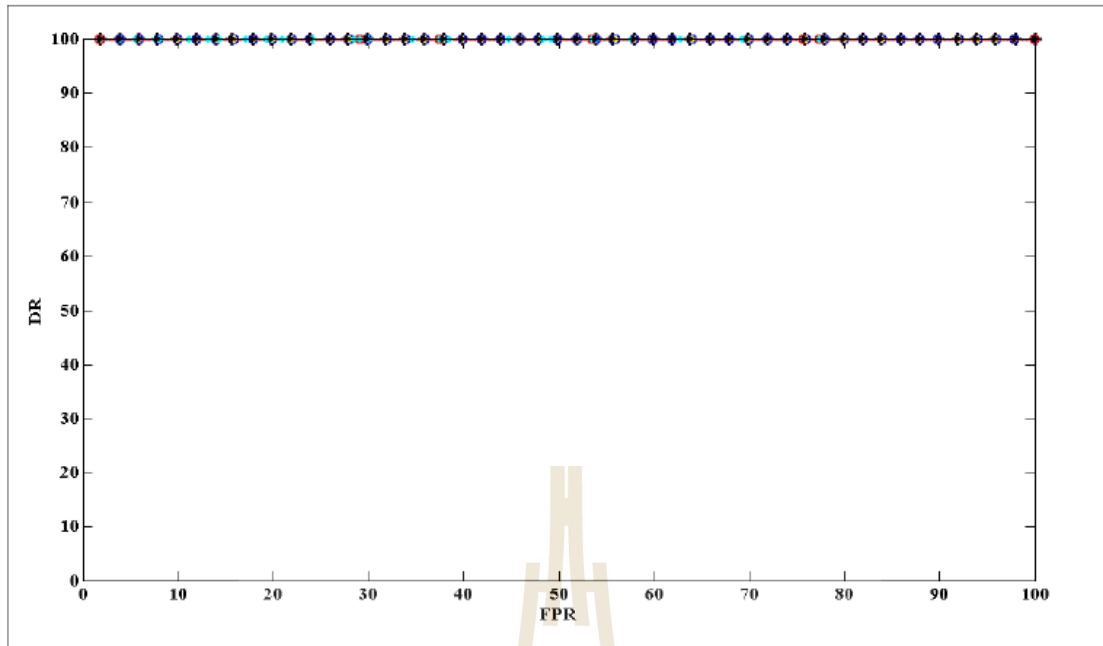
รูปที่ 3.22 คือผลลัพธ์ของชุดข้อมูลสังเคราะห์ที่ใส่ความผิดปกติจากสัญญาณรบกวนเข้าไป ผลลัพธ์แสดงให้เห็นว่าทุกอัลกอริทึมทำงานได้ดียกเว้นอัลกอริทึม OSCVM+DWT (HP) และ OSCVM + LWT (HP) ที่ให้ประสิทธิภาพแย่มากที่สุด

รูปที่ 3.23 แสดงผลลัพธ์ของชุดข้อมูล pdg2008 ซึ่งความผิดปกติในข้อมูลชุดนี้มีแอมพลิจูดที่เล็กทำให้ยากต่อการตรวจจับอัลกอริทึม OSCVM เพียงอย่างเดียวและอัลกอริทึม OSCVM+PCA ทำงานได้ไม่ดีตลอดจน OSCVM+DWT (LP) และ OSCVM+LWT (LP) ในทางตรงกันข้ามอัลกอริทึม OCSVM+DWT (HP) และอัลกอริทึม OCSVM+LWT (HP) ให้ประสิทธิภาพที่แย่มากซึ่งสอดคล้องกับผลลัพธ์ของข้อมูลสังเคราะห์ ซึ่งเป็นเพราะว่าสัมประสิทธิ์ HP สะท้อนถึงอัตราการเปลี่ยนแปลงระหว่างสองตัวอย่างที่ต่อเนื่องกัน เพราะฉะนั้นค่าสัมประสิทธิ์ HP จึงเหมาะสมกับความผิดปกติระยะสั้นมากกว่า ในขณะที่ค่าสัมประสิทธิ์ LP เหมาะสมกับความผิดปกติที่เปลี่ยนแปลงอย่างช้าๆเช่นความผิดปกติจากสัญญาณรบกวน

รูปที่ 3.24 และ 3.25 แสดงข้อมูลสัณเคราะห์ที่ถูกใส่ความผิดพลาดที่เข้าไปและผลลัพธ์ของชุดข้อมูล NAMOS ตามลำดับ ผลลัพธ์ของชุดข้อมูล NAMOS สอดคล้องกับผลลัพธ์ของข้อมูลสัณเคราะห์ที่ความผิดพลาดที่เข้าไป อัลกอริทึม OSCVM+LWT (LP) และอัลกอริทึม OSCVM+DWT (LP) ทำงานได้ดีกว่าอัลกอริทึม OSCVM+PCA และอัลกอริทึม OSCVM เพียงอย่างเดียว ในขณะที่อัลกอริทึม OSCVM+LWT (HP) และ OSCVM+DWT (HP) ให้ประสิทธิภาพที่แย่มากที่สุด สังเกตว่าผลลัพธ์จากความผิดพลาดที่นั่นคล้ายกับผลลัพธ์จากความผิดพลาดจากสัญญาณรบกวน เป็นเพราะว่าความผิดพลาดทั้งสองชนิดมีการเปลี่ยนแปลงที่มีแนวโน้มเหมือนกัน ดังนั้นความผิดพลาดทั้งสองจึงมีคุณสมบัติที่มันสำคัญมากขึ้นเมื่อถูกจับด้วยสัมประสิทธิ์ LP มากกว่าสัมประสิทธิ์ HP

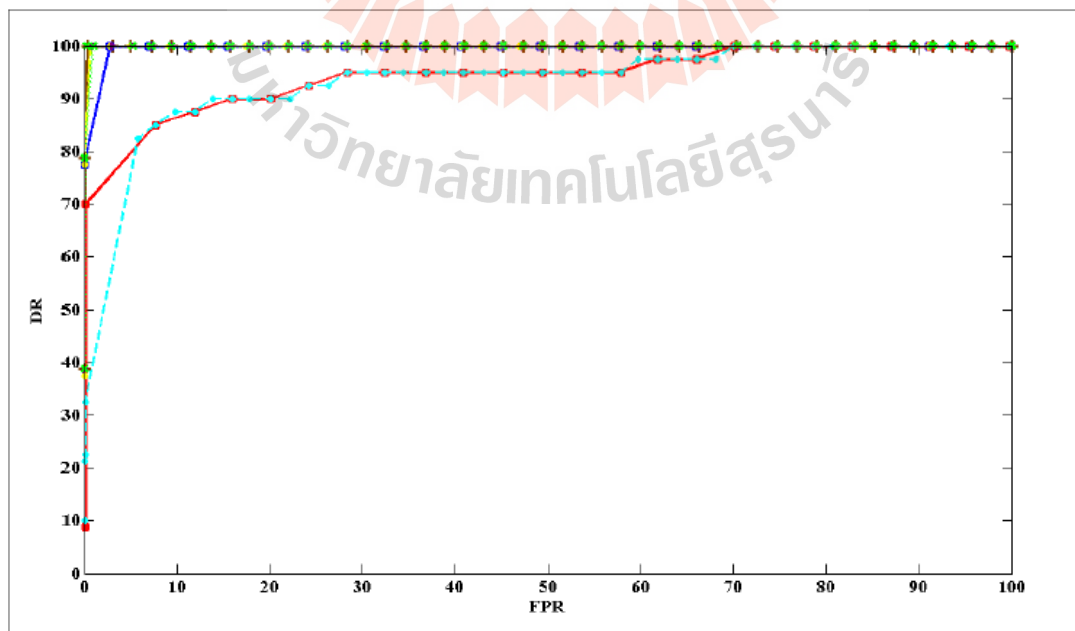


รูปที่ 3.20 ROC ของข้อมูลสัณเคราะห์ 2KPI เพิ่มด้วยความผิดพลาด 1/80

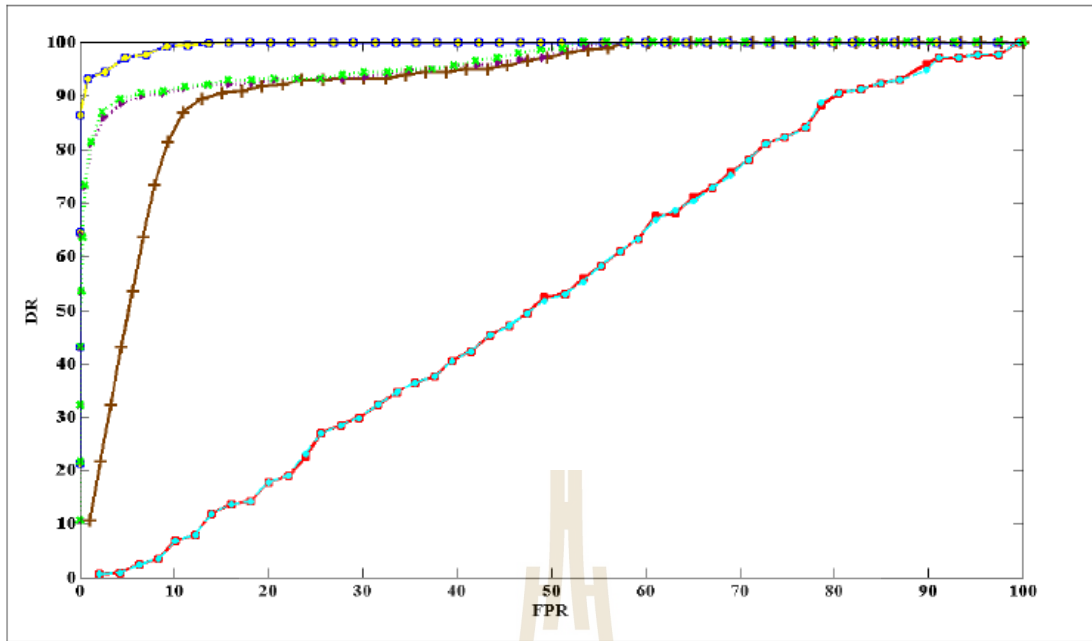


รูปที่ 3.21 ROC ของชุดข้อมูล INTEL 1KPI (ประกอบด้วยความผิดปกติระยะสั้น)

— OCSVM — OCSVM + DWT [LP] — OCSVM + DWT [HP] — OCSVM + LWT [LP]
 — OCSVM + LWT [HP] — OCSVM + PCA [1KPI] — OCSVM + PCA [2KPI]

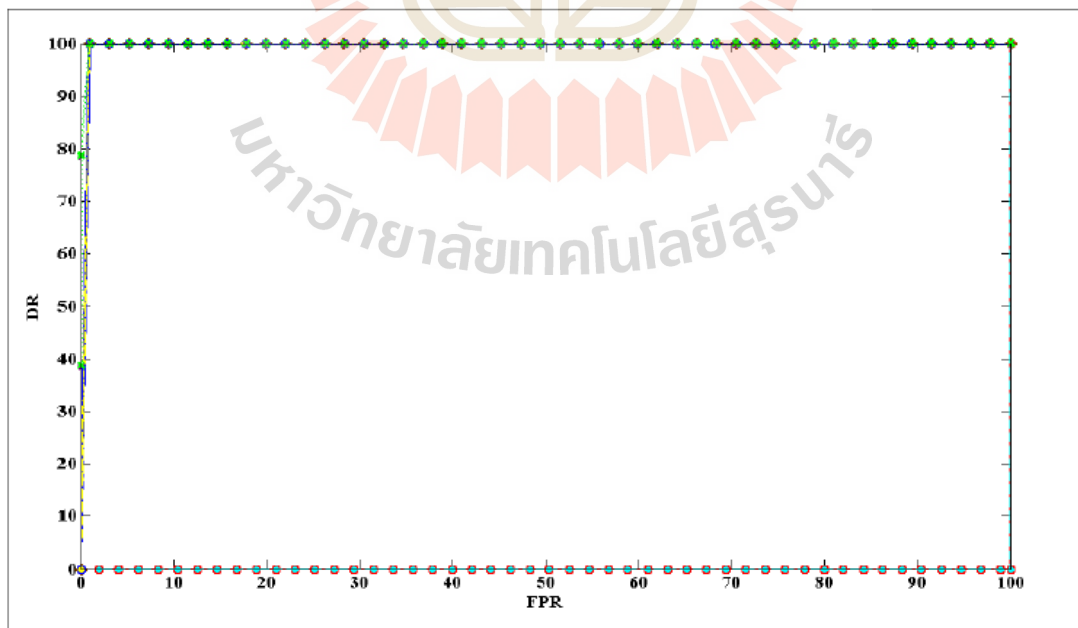


รูปที่ 3.22 ROC ของข้อมูลถังกระดาษ 2KPI เพิ่มด้วยความผิดปกติ 20/4

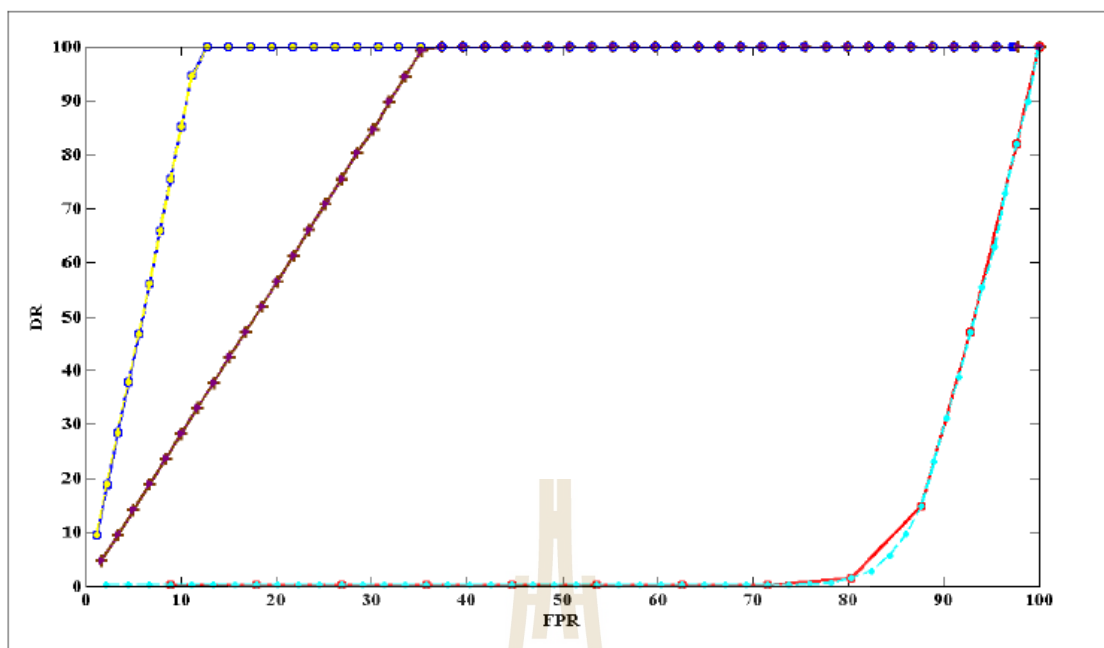


รูปที่ 3.23 ROC ของชุดข้อมูล pdg2008 2KPI
(ประกอบด้วยความผิดปกติจากสัญญาณรบกวน)

—+— OCSVM —■— OCSVM + DWT [LP] —■— OCSVM + DWT [HP] - - - ■ - - - OCSVM + LWT [LP]
- - - ■ - - - OCSVM + LWT [HP] ···· ◆ ···· OCSVM + PCA [1KPI] ···· ◆ ···· OCSVM + PCA [2KPI]



รูปที่ 3.24 ROC ของชุดข้อมูลสังเคราะห์ 2KPI เพิ่มด้วยความผิดปกติ 80/1



รูปที่ 3.25 ROC ของชุดข้อมูล NAMOS 1KPI (ประกอบด้วยความผิดปกติคงที่)

3.4.3 ขยายผลสู่ชุดข้อมูล 3 KPI

รูปที่ 3.26 แสดงชุดข้อมูลสังเคราะห์ 3KPI ที่ถูกใส่ความผิดปกติระยะสั้นเข้าไป ผลการทดลองพบว่าทุกอัลกอริทึมสามารถทำงานได้ดี เพราะความผิดปกติระยะสั้นมีแอมพลิจูดที่สูงง่ายต่อการตรวจจับ ยกเว้นอัลกอริทึม OSCVM+PCA [1 KPI] ที่ให้ประสิทธิภาพแย่มากที่สุด เป็นเพราะว่าข้อมูลสังเคราะห์ที่สร้างขึ้นโดยการใส่ความผิดปกติเข้าไปที่ตำแหน่งแตกต่างกัน ทำให้ข้อมูลในแต่ละ KPI แตกต่างกันด้วย (แสดงในรูปที่ 3.6) การบีบอัดข้อมูลโดย PCA กับเพียง 1KPI นั้นละทิ้งบางองค์ประกอบของข้อมูล ดังนั้นข้อมูลที่ปรากฏจึงมีความผิดเพี้ยนไป เพราะตำแหน่งของความผิดปกติที่บิดเบี้ยวไปทำให้ยากต่อการตรวจจับ

รูปที่ 3.27 แสดงผลของชุดข้อมูล INTEL 3KPI ที่ประกอบด้วยความผิดปกติระยะสั้นใน 3KPI ผลการทดลองแสดงให้เห็นว่าทุกอัลกอริทึมทำงานได้ดีเท่ากัน ซึ่งเป็นเพราะว่าค่าสัมประสิทธิ์ HP สะท้อนถึงอัตราการเปลี่ยนแปลงระหว่างสองตัวอย่างที่ต่อเนื่องกันและความผิดปกติระยะสั้นมีแอมพลิจูดสูง เพราะฉะนั้นความผิดปกติระยะสั้นจึงง่ายต่อการตรวจจับ นอกจากนี้แต่ละ KPI ของชุดข้อมูล INTEL ประกอบด้วยความผิดปกติในตำแหน่งเดียวกัน (แสดงในรูปที่ 3.11-3.13) การบีบอัดข้อมูลโดย PCA กับ 1KPI นั้นข้อมูลบางองค์ประกอบของข้อมูลดังนั้นจึงเกิดความ

ผิดเพิ่มขึ้น อย่างไรก็ตามตำแหน่งของความผิดปกติยังคงปรากฏให้เห็นชัดเจน และง่ายต่อการตรวจจับแม้ว่าจะบีบอัดข้อมูลด้วย PCA 1KPI

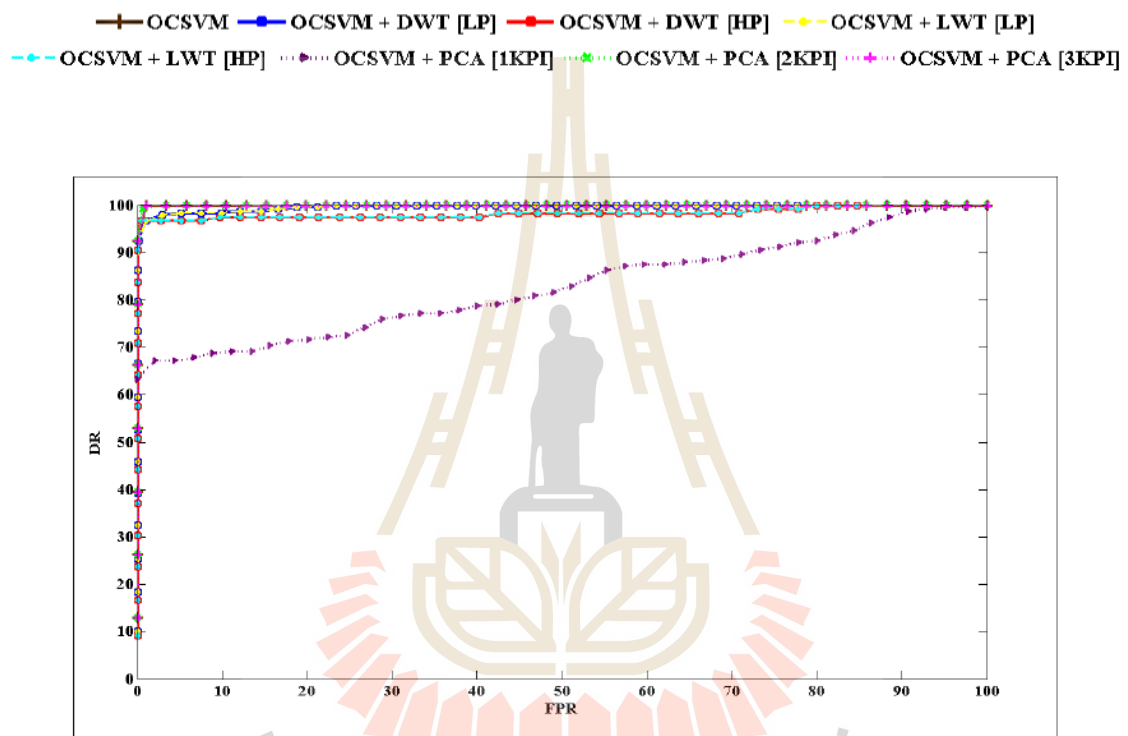
รูปที่ 3.28 แสดงผลข้อมูลสังเคราะห์ 3KPI ที่เพิ่มความผิดปกติจากสัญญาณรบกวน ผลการทดลองแสดงให้เห็นว่าอัลกอริทึม OSCVM+LWT (LP) และ OSCVM+DWT (LP) ทำงานได้ดีเท่ากับอัลกอริทึม OSCVM เพียงอย่างเดียว และอัลกอริทึม OSCVM+PCA [2KPI] และ [3KPI] อัลกอริทึม OSCVM+LWT (HP) และ OSCVM+DWT (HP) ให้ประสิทธิภาพที่แยกว่าอัลกอริทึมเหล่านั้น เพราะค่าสัมประสิทธิ์ HP สะท้อนถึงอัตราการเปลี่ยนแปลงระหว่างสองตัวอย่างที่ต่อเนื่องกัน เพราะฉะนั้นค่าสัมประสิทธิ์ HP จึงมีความเหมาะสมสำหรับความผิดปกติระยะสั้น ในขณะที่ค่าสัมประสิทธิ์ LP เหมาะสมกับความผิดปกติที่เปลี่ยนแปลงอย่างช้าๆ เช่น ความผิดปกติจากสัญญาณรบกวน นอกจากนี้อัลกอริทึม OSCVM+PCA [1KPI] ให้ประสิทธิภาพที่แย่ที่สุดในทุกๆ อัลกอริทึม เนื่องจากข้อมูลถูกสร้างขึ้นด้วยตำแหน่งความผิดปกติที่แตกต่างกัน ข้อมูลที่ปรากฏในแต่ละ KPI จึงแตกต่างกัน (แสดงในรูปที่ 3.8) ข้อมูลที่ถูกบีบอัดด้วย PCA [1KPI] จะเกิดความผิดเพี้ยน เพราะฉะนั้นตำแหน่งของความผิดปกติที่ผิดเพี้ยนจึงทำให้ยากต่อการตรวจจับ

รูปที่ 3.29 แสดงผลชุดข้อมูล pdg2008 3KPI ซึ่งประกอบด้วยความผิดปกติที่เกิดจากสัญญาณรบกวน พบว่าอัลกอริทึม OSCVM+LWT (LP) และ OSCVM+DWT (LP) ทำงานได้ดีกว่าอัลกอริทึม OSCVM เพียงอย่างเดียว และอัลกอริทึม OSCVM+PCA อัลกอริทึม OSCVM+LWT (HP) และ OSCVM+DWT (HP) ให้ประสิทธิภาพที่แย่ที่สุด ในขณะที่ผลของข้อมูลสังเคราะห์ในรูปที่ 3.28 แสดงให้เห็นว่า OSCVM+PCA [1KPI] ให้ประสิทธิภาพที่แย่ที่สุด เพราะค่าสัมประสิทธิ์ HP สะท้อนถึงอัตราการเปลี่ยนแปลงระหว่างสองตัวอย่างที่ต่อเนื่องกัน มากไปกว่านั้นชุดข้อมูล pdg2008 (แสดงในรูป 3.17-3.18) มีแอมพลิจูดของสัญญาณรบกวนที่ต่ำกว่าข้อมูลสังเคราะห์ (แสดงในรูป 3.8) เพราะฉะนั้นผลของชุดข้อมูล pdg2008 3KPI แสดงให้เห็นว่า OSCVM+LWT (HP) และ OSCVM+DWT (HP) ให้ประสิทธิภาพที่แย่ที่สุด

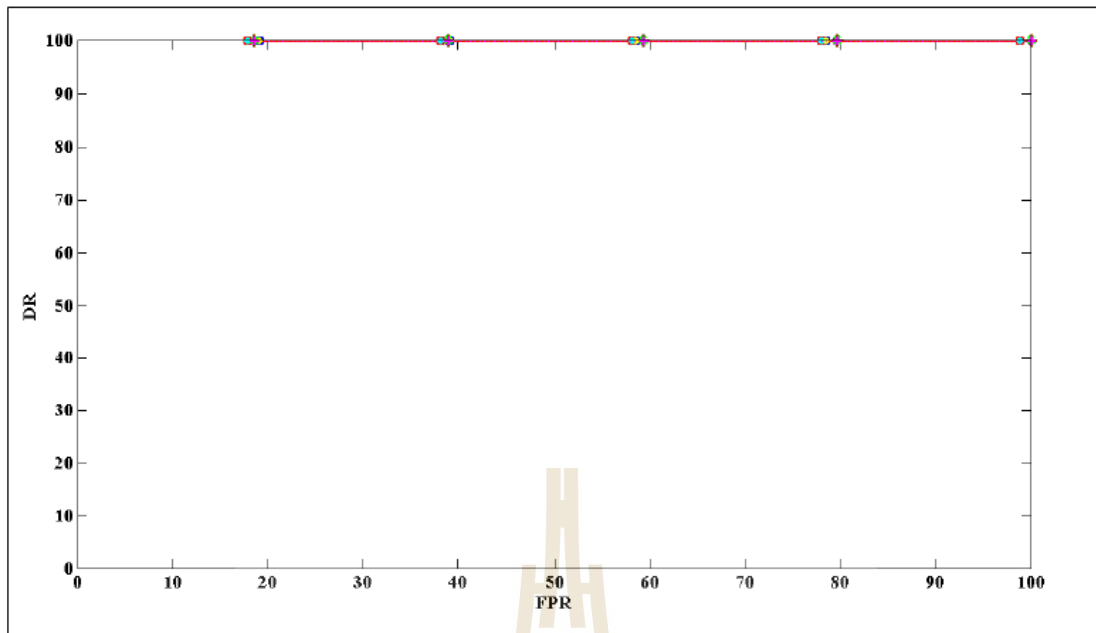
รูปที่ 3.30 แสดงผลชุดข้อมูลสังเคราะห์ 3KPI ที่เพิ่มความผิดปกติคงที่ ทุกอัลกอริทึมแสดงการทำงานได้ดียกเว้นสำหรับอัลกอริทึม OSCVM+LWT (HP) และ OSCVM+DWT (HP) ที่ให้ประสิทธิภาพที่แย่ที่สุด

รูปที่ 3.31 แสดงผลของชุดข้อมูล NAMOS 3KPI ซึ่งให้ผลสอดคล้องกับข้อมูลสังเคราะห์ในรูปที่ 3.30 อัลกอริทึม OSCVM+LWT (LP) และ OSCVM+DWT (LP) ทำงานได้ดีกว่า

อัลกอริทึม OSCVM เพียงอย่างเดียว และอัลกอริทึม OSCVM+PCA อัลกอริทึม OSCVM+LWT (HP) และ OSCVM+DWT (HP) ให้ประสิทธิภาพที่แย่ที่สุด สังเกตว่าผลการทดลองของความผิดพลาดที่คล้ายกับผลการทดลองของความผิดพลาดจากสัญญาณรบกวน เพราะว่าความผิดพลาดทั้งสองชนิดมีการเปลี่ยนแปลงที่มีแนวโน้มเหมือนกัน ดังนั้นความผิดพลาดทั้งสองจึงมีคุณสมบัติที่มีนัยสำคัญมากขึ้นเมื่อถูกจับด้วยสัมประสิทธิ์ LP มากกว่าสัมประสิทธิ์ HP

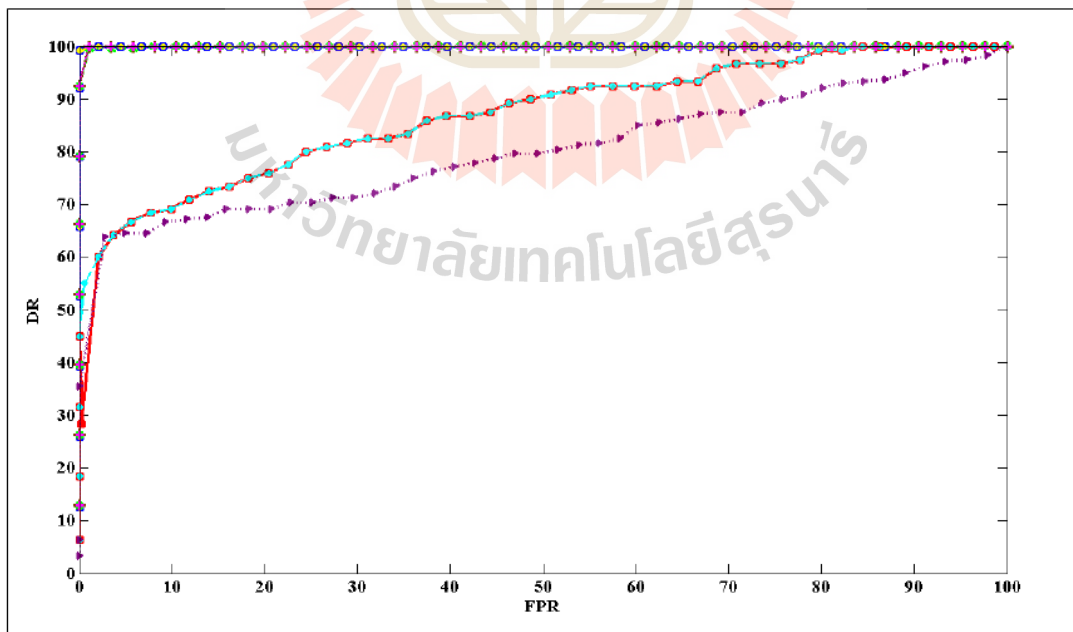


รูปที่ 3.26 ROC ของข้อมูลตั้งเกราะห้ 3KPI เพิ่มด้วยความผิดพลาด 1/80

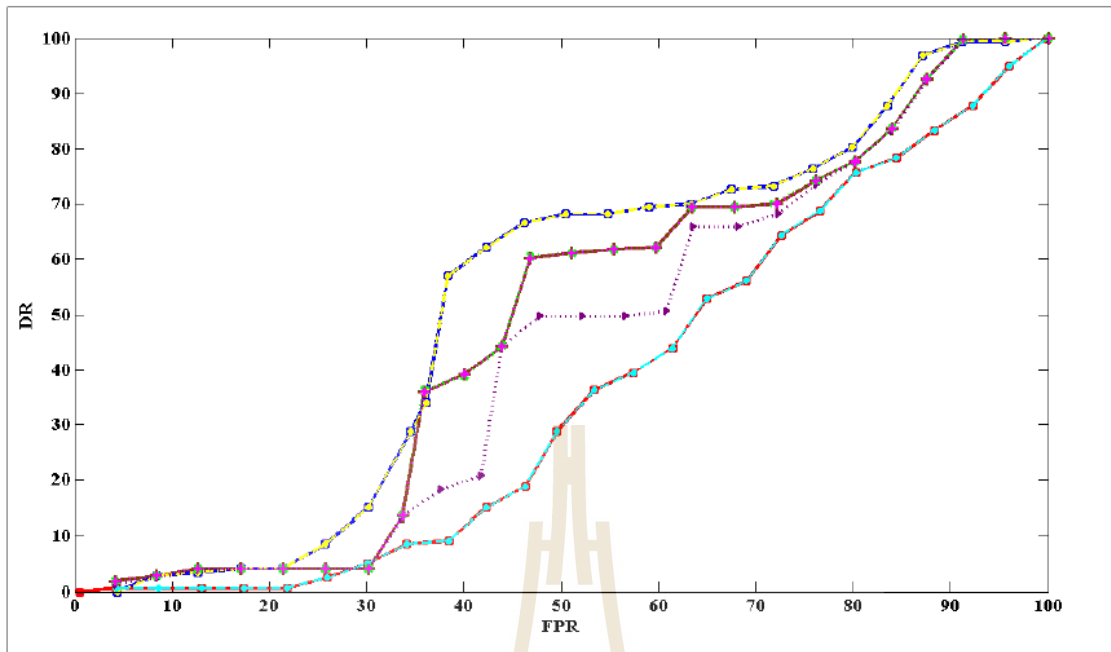


รูปที่ 3.27 ROC ของชุดข้อมูล INTEL 3KPI (ประกอบด้วยความผิดปกติระยะสั้น)

— OCSVM — OCSVM + DWT [LP] — OCSVM + DWT [HP] — OCSVM + LWT [LP]
 — OCSVM + LWT [HP] — OCSVM + PCA [1KPI] — OCSVM + PCA [2KPI] — OCSVM + PCA [3KPI]

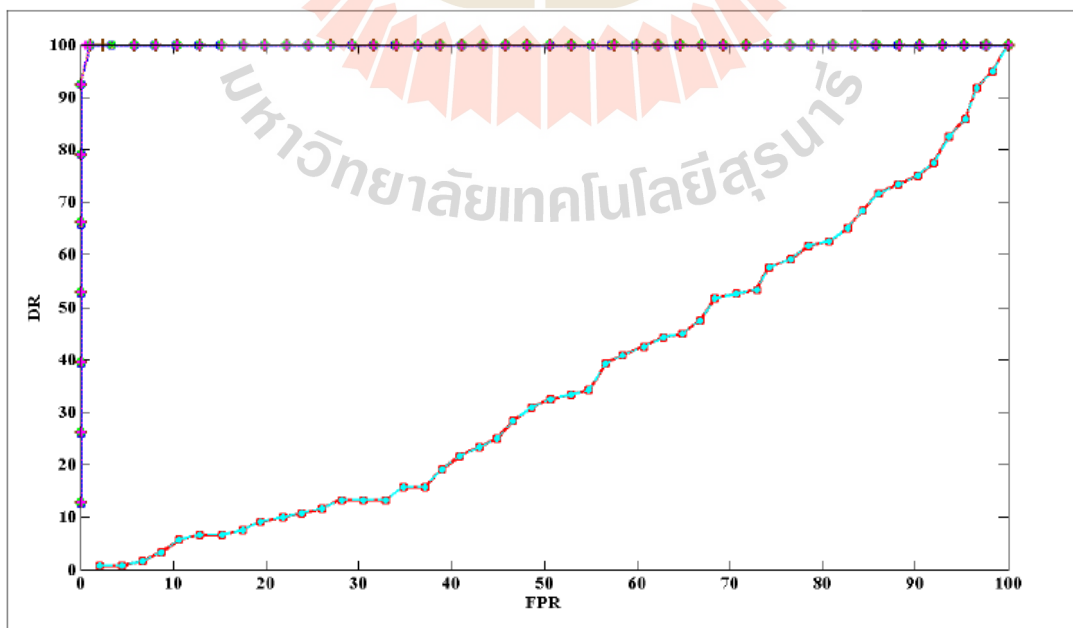


รูปที่ 3.28 ROC ของข้อมูลสังเคราะห์ 3KPI เพิ่มด้วยความผิดปกติ 20/4

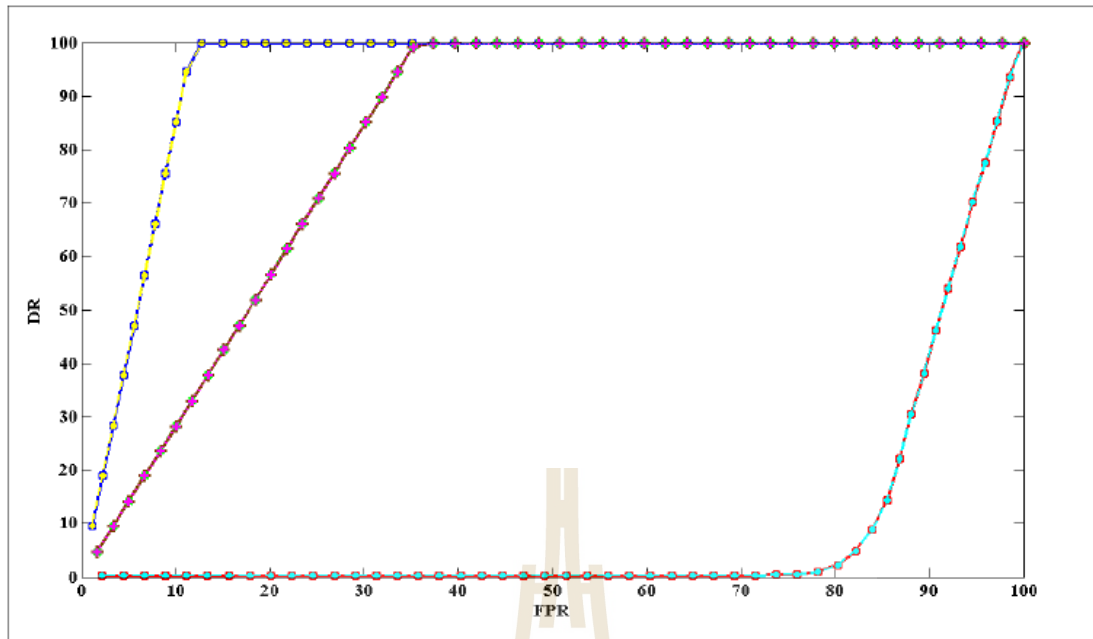


รูปที่ 3.29 ROC ของชุดข้อมูล pdg2008 3KPI
(ประกอบด้วยความผิดปกติจากสัญญาณรบกวน)

—+— OCSVM —■— OCSVM + DWT [LP] —■— OCSVM + DWT [HP] —◆— OCSVM + LWT [LP]
—■— OCSVM + LWT [HP] ...◆... OCSVM + PCA [1KPI] ...◆... OCSVM + PCA [2KPI] ...◆... OCSVM + PCA [3KPI]



รูปที่ 3.30 ROC ของชุดข้อมูลตั้งเกาะห้ 3KPI เพิ่มด้วยความผิดปกติ 80/1



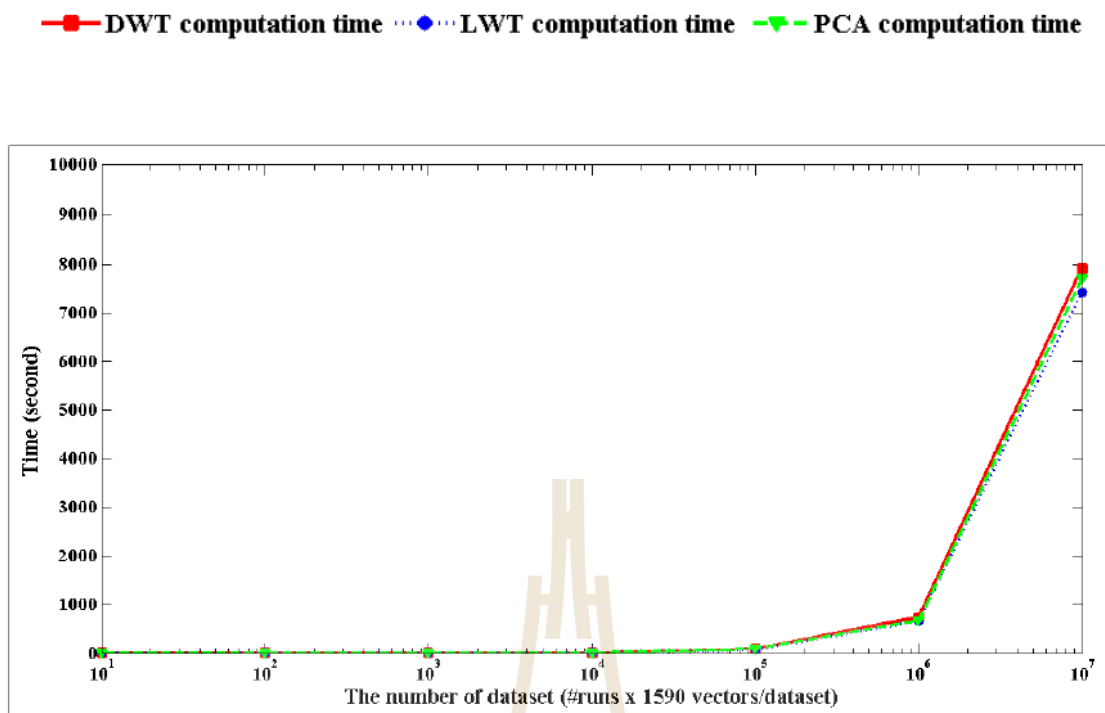
รูปที่ 3.31 ROC ของชุดข้อมูล NAMOS 3KPI (ประกอบด้วยความผิดปกติทั้ง)

3.4.4 การประเมินการคำนวณเชิงเวลา

มีหลายงานวิจัยที่กล่าวถึงเทคนิคการบีบอัดข้อมูลโดยใช้ DWT ได้รับความนิยมนำมาใช้ในการคอนโวลูชันซึ่งต้องการเวลาในการดำเนินการ, พื้นที่หน่วยความจำและการคำนวณทางคณิตศาสตร์จำนวนมากกว่า LWT (Sweldens, 1998), (Achaya, et al., 2006), (X. L. Liet al., 2009)

เพราะฉะนั้นการทดลองจึงดำเนินการป้อนข้อมูลเวกเตอร์ขององค์ประกอบข้อมูล 1590 ตัวเข้าไปในการบีบอัดข้อมูลแต่ละหลักการเพื่อสำหรับวัดเวลาที่ใช้ในการคำนวณของแต่ละหลักการ

รูปที่ 3.32 แสดงเวลาในการคำนวณของแต่ละหลักการที่ใช้สำหรับการประมวลผลที่เพิ่มขึ้นในแต่ละขนาดของชุดข้อมูลอินพุต สังเกตว่า LWT ใช้เวลาในการคำนวณน้อยกว่า DWT และ PCA



รูปที่ 3.32 เวลาในการคำนวณของแต่ละเทคนิคการบีบอัดข้อมูล

3.5 สรุป

งานวิจัยได้นำเสนอการรวมกันของอัลกอริทึม OSCVM กับ LWT เพื่อการตรวจจับความผิดปกติในเครือข่ายเซ็นเซอร์ไร้สาย โดยใช้การประเมินเชิงตัวเลขกับอัลกอริทึมโดยใช้โปรแกรม MATLAB และทดสอบอัลกอริทึมด้วยข้อมูลสังเคราะห์และข้อมูลที่ได้จากการวัดจริง สำหรับข้อมูลสังเคราะห์และข้อมูลที่ได้จากการวัดจริงที่ประกอบด้วยความคิดปกติระยะสั้น อัลกอริทึม OSCVM กับ LWT ทำงานได้ดีเท่ากับอัลกอริทึม OSCVM เพียงอย่างเดียว, อัลกอริทึม OSCVM กับ DWT และอัลกอริทึม OSCVM กับ LWT สำหรับข้อมูลสังเคราะห์และข้อมูลที่ได้จากการวัดจริงที่ประกอบด้วยความคิดปกติจากสัญญาณรบกวนและความคิดปกติ อัลกอริทึม OSCVM กับ LWT (LP) และอัลกอริทึม OSCVM กับ DWT (LP) ให้ประสิทธิภาพที่ดีกว่าอัลกอริทึม OSCVM เพียงอย่างเดียวและอัลกอริทึม OSCVM กับ PCA ในขณะที่อัลกอริทึม OSCVM กับ LWT (HP) และอัลกอริทึม OSCVM กับ DWT (HP) ให้ประสิทธิภาพที่แย่ที่สุด

อย่างไรก็ตาม LWT ที่การคำนวณที่ง่ายและใช้เวลาในการคำนวณน้อยกว่า DWT เพราะฉะนั้นอัลกอริทึม OSCVM กับ LWT จึงเหมาะสมกับเครือข่ายเซ็นเซอร์ไร้สายมากกว่า

บทที่ 4

การประเมินผลการตรวจจับความผิดปกติของข้อมูลจากเครือข่ายเซ็นเซอร์ไร้สายสำหรับ เฟ้าระวังโรงผลิตปุ๋ยอินทรีย์ชีวภาพ

4.1 บทนำ

ในการวิจัยนี้ได้ทำการประยุกต์ใช้อัลกอริทึมในการตรวจจับความผิดปกติของข้อมูลจากเครือข่ายเซ็นเซอร์ไร้สายสำหรับเฟ้าระวังโรงผลิตปุ๋ยอินทรีย์ชีวภาพ (bioorganic fertilizer production plant monitoring) โดยข้อมูลที่เก็บได้มาจากชุดอุปกรณ์ต้นแบบเครือข่ายเซ็นเซอร์ไร้สายที่ได้ติดตั้งอยู่ที่โรงปุ๋ยอินทรีย์ชีวภาพ มหาวิทยาลัยเทคโนโลยีสุรนารี

ในการเก็บข้อมูลที่ได้จากชุดอุปกรณ์ต้นแบบนั้นได้ทำการออกแบบระบบให้มีความเหมาะสมต่อเงื่อนไขในกระบวนการหมักปุ๋ยอินทรีย์ชีวภาพ การออกแบบระบบเบื้องต้นสำหรับชุดอุปกรณ์ต้นแบบประกอบด้วย สถานีฐาน โหนดเซ็นเซอร์ โพรบเซ็นเซอร์ (sensor probe) สำหรับวัดอุณหภูมิและความชื้นในดิน ซึ่งถูกติดตั้งไว้บนสุดของ โหนดเซ็นเซอร์ โดยอุปกรณ์ต้นแบบนี้ได้ถูกติดตั้งไว้ 2 จุดภายในโรงปุ๋ยอินทรีย์ชีวภาพ โหนดเซ็นเซอร์จะทำการตรวจวัดและส่งข้อมูลที่วัดได้ไปยังสถานีฐานอย่างต่อเนื่อง



รูปที่ 4.1 ชุดอุปกรณ์ต้นแบบ



(ก)

(ข)



(ค)

รูปที่ 4.2 โรงเรือนปุ๋ยอินทรีย์ชีวภาพ มหาวิทยาลัยเทคโนโลยีสุรนารี (ก) บริเวณภายใน โรงเรือน (ข) ตำแหน่งที่ติดตั้งอุปกรณ์ต้นแบบ (ค) ตำแหน่งของสถานีฐาน

4.2 อุปกรณ์ต้นแบบสำหรับเฝ้าระวังโรงผลิตปุ๋ยอินทรีย์ชีวภาพ

ผู้วิจัยได้ทำการสร้างอุปกรณ์ต้นแบบสำหรับการตรวจวัดข้อมูลของปุ๋ยอินทรีย์ชีวภาพด้วย เครื่องข่ายเซ็นเซอร์ไร้สาย ซึ่งถูกติดตั้งที่โรงปุ๋ยอินทรีย์ชีวภาพ มหาวิทยาลัยเทคโนโลยีสุรนารี ข้อมูลที่ตรวจวัดได้จากโรงเรือนจะถูกส่งไปยังสถานีฐานด้วยการรับ-ส่งข้อมูลแบบไร้สาย

โดยอุปกรณ์สำหรับนำไปใช้สร้างชุดต้นแบบสำหรับการตรวจวัดข้อมูลของปุ๋ยอินทรีย์ชีวภาพด้วยเครือข่ายเซ็นเซอร์ไร้สาย มีองค์ประกอบดังนี้

4.2.1 สถานีฐาน (Base Station)

อุปกรณ์ที่ใช้เป็นสถานีฐานได้แก่ บอร์ดไมโครคอนโทรลเลอร์รุ่น MIB 520CB เป็นผลิตภัณฑ์จากบริษัท Crossbow ประกอบด้วย USB สำหรับการเชื่อมต่อกับคอมพิวเตอร์ อุปกรณ์ MIB 520CB แบ่งออกเป็น 2 พอร์ตทำหน้าที่แตกต่างกัน โดยพอร์ตแรกทำหน้าที่เป็นช่องทางสำหรับการโปรแกรมข้อมูลให้แก่เซ็นเซอร์โหนด พอร์ตที่สองทำหน้าที่ในการรับข้อมูลที่ตรวจวัดได้จากเครือข่ายเซ็นเซอร์ไร้สาย โดยปกติแล้วอุปกรณ์ MIB 520CB จะทำหน้าที่เชื่อมต่อกับคอมพิวเตอร์ทางพอร์ต USB และทำหน้าที่เป็นสถานีฐานด้วยการใช้เซ็นเซอร์โหนดชนิด IRIS/MICAZ/Mica2 มาต่อเข้ากับบอร์ดอินเตอร์เฟซ (interface board) ของ MIB 520CB และถูกโปรแกรมให้ทำหน้าที่เป็นสถานีฐานเรียบร้อยแล้ว



รูปที่ 4.3 บอร์ด MIB 520CB (USB interface)

4.2.2 เซ็นเซอร์โหนด (Sensor mote)

สำหรับอุปกรณ์ที่ใช้เป็นโหนดเซ็นเซอร์ ได้แก่ เซ็นเซอร์โหนดรุ่น MPR2400 หรือ MicaZ ผลิตภัณฑ์จากบริษัท Crossbow โดย MPR2400 ประกอบด้วยไมโครคอนโทรลเลอร์ Atmel ATMEGA 128 ซึ่งเป็นไมโครคอนโทรลเลอร์ที่มีการใช้พลังงานต่ำที่ประมวลผลจากแฟลชเมมโมรี่ภายนอก มีพิน 51 ขาเพื่อรองรับสัญญาณขาเข้าแบบอนาล็อก ดิจิตอล I2C SPI และอินเตอร์เฟซ UART โดยอินเตอร์เฟซนี้จะทำให้ง่ายต่อการเชื่อมต่อกับอุปกรณ์ต่อพ่วงภายนอก



รูปที่ 4.4 โหนดเซ็นเซอร์ MPR2400 หรือ MicaZ

4.2.3 บอร์ดรวมข้อมูล (Data acquisition board)

อุปกรณ์สำหรับใช้เป็น บอร์ดรวมข้อมูล (Data acquisition board) ได้แก่ บอร์ด MDA300 ของบริษัท Crossbow บอร์ดรวมข้อมูล (Data Acquisition board หรือ DAQ) ถูกนำมาใช้เพื่อรวบรวมข้อมูลที่ตรวจวัดได้จากแหล่งข้อมูลหลายแหล่งที่แตกต่างกัน โดยแหล่งข้อมูลภายนอกเหล่านี้มาจากอุปกรณ์เซ็นเซอร์ภายนอก ซอร์ฟแวร์ของ DAQ จะมีอินเตอร์เฟซของแหล่งข้อมูลไปยังเอาต์พุตที่แตกต่างกัน จากคุณสมบัติดังกล่าวจึงทำให้ MDA300 เป็นอุปกรณ์ที่ง่ายต่อการเข้าถึงข้อมูลเหมาะสำหรับการประยุกต์ใช้งานที่หลากหลาย เช่น สามารถเชื่อมต่อกับอุปกรณ์เซ็นเซอร์ภายนอก ไม่ว่าจะเป็นเซ็นเซอร์วัดความเป็นกรด-ด่างของสารละลาย เซ็นเซอร์วัดความชื้น หรือเซ็นเซอร์วัดอุณหภูมิ เป็นต้น นอกจากนี้ยังมีการใช้พลังงานต่ำ การเก็บและแสดงผลข้อมูลถูกสนับสนุนภายใต้ อุปกรณ์เชื่อมต่อของ Crossbow ด้วยกัน ข้อมูลที่ตรวจวัดได้จะถูกรวบรวมไปยังสถานีฐานซึ่งมีการเชื่อมต่อกับเครื่องคอมพิวเตอร์โดยอัตโนมัติเพื่อบันทึกค่าลงฐานข้อมูล

4.2.4 อุปกรณ์เซ็นเซอร์ภายนอกสำหรับตรวจวัดอุณหภูมิและความชื้นในดิน

ในงานวิจัยนี้ได้ใช้อุปกรณ์เซ็นเซอร์ 2 ชนิด ได้แก่ เซ็นเซอร์วัดอุณหภูมิ และ เซ็นเซอร์วัดความชื้นในดิน เทอร์โมคัปเปิล (thermocouple) คืออุปกรณ์ที่ใช้เซ็นเซอร์วัดอุณหภูมิ สำหรับวัดอุณหภูมิ ซึ่งสามารถเปลี่ยนอุณหภูมิหรือความร้อนเป็นแรงเคลื่อนไฟฟ้า โดยอุปกรณ์นี้จะมี ราคาค่อนข้างถูกและปรับเปลี่ยนได้ง่าย มีหัวต่อมาตรฐาน และสามารถนำไปใช้วัดอุณหภูมิได้ใน ช่วงกว้าง สำหรับเซ็นเซอร์วัดความชื้นในดินในงานวิจัยนี้ใช้ EC-5 เนื่องจากอุปกรณ์ชนิดนี้มีความ ทนทานต่ออุณหภูมิสูงซึ่งเป็นอุณหภูมิภายในกระบวนการหมักปุ๋ย



รูปที่ 4.5 บอร์ดรวมข้อมูล MDA300



รูปที่ 4.6 เซ็นเซอร์วัดอุณหภูมิ หรือ เทอร์โมคัปเปิล



รูปที่ 4.7 เซ็นเซอร์วัดความชื้นในดิน (EC-5)

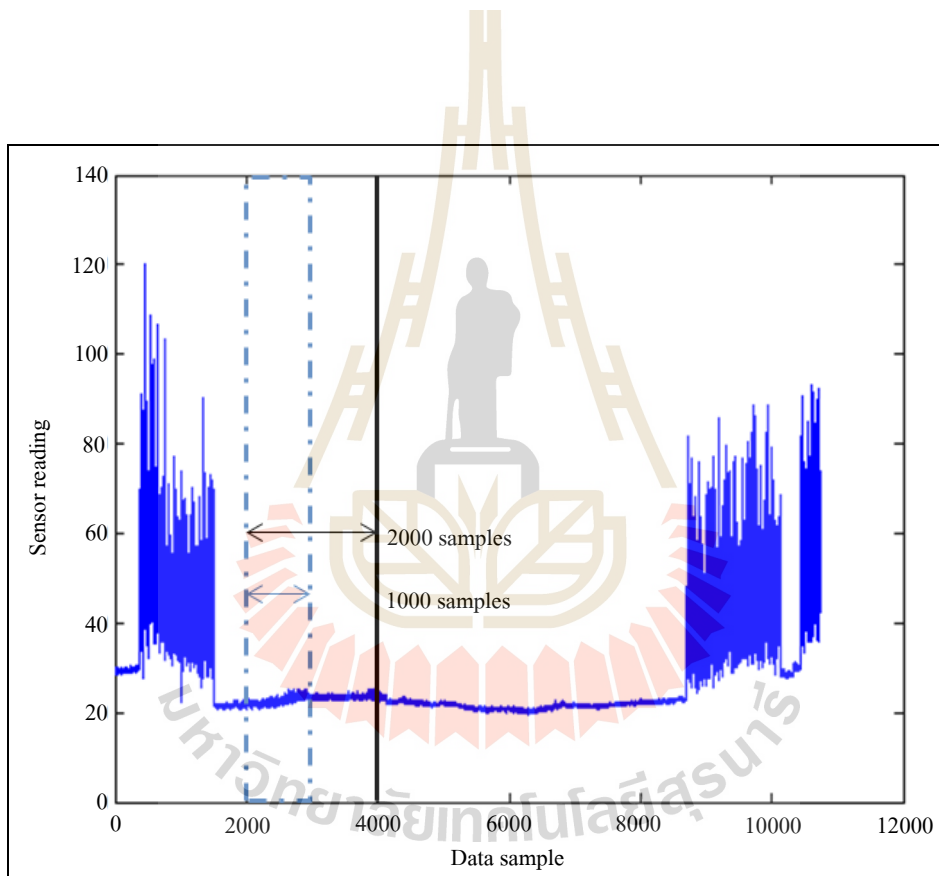
4.3 ผลการทดลอง

การวิจัยนี้ทำการตรวจจับความผิดปกติของข้อมูลโดยใช้ตัวชี้วัดสมรรถภาพ (KPI) ชนิดหนึ่งเพื่อวัดสมรรถภาพของชุดข้อมูล BOF นั่นคือ การอ่านข้อมูลอุณหภูมิซึ่งถูกอ่านค่าทุกๆ 5 นาทีต่อวัน ด้วยการใช้การตรวจจับข้อมูลจริงและวิธีการทางฮิสโตแกรม ทำให้ค่าขอบแบ่งบนและขอบแบ่งล่างถูกนำมาใช้ในการตรวจจับความผิดปกติของชุดข้อมูล BOF มีค่าเป็น 0 และ 30 ขนาดของการสอนให้จำรูปแบบ คือ 1000 ตัวอย่าง และ 2000 ตัวอย่างดังที่แสดงในรูปที่ 4.8

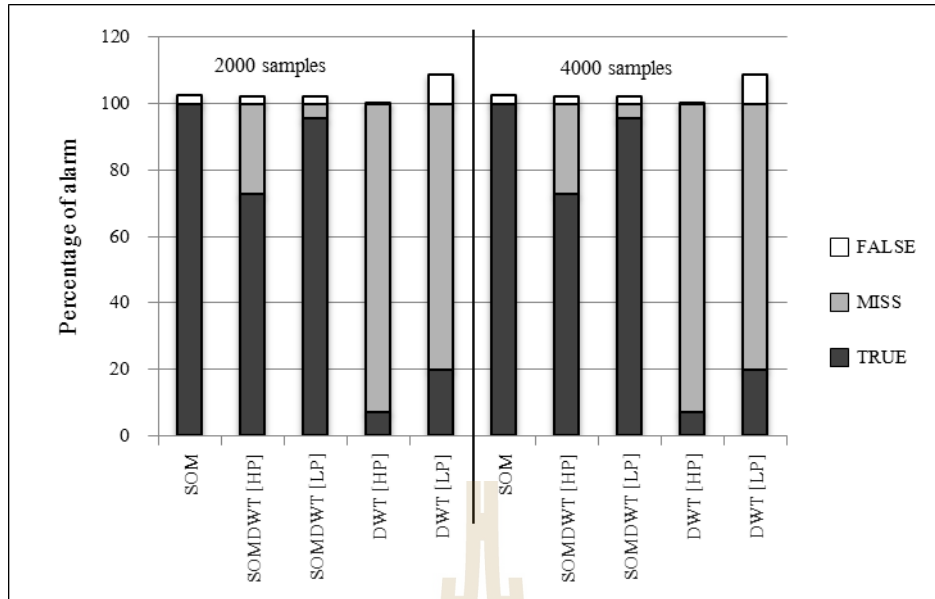
รูปที่ 4.9 แสดงเปอร์เซ็นต์ของอัตราส่วนของการเตือนที่ตรวจพบ (detected alarm rates) สำหรับการเตือนจริง (true alarm) การเตือนตกหล่น (miss alarm) การเตือนที่ผิดพลาด (false alarm) ในการ train ตัวอย่างข้อมูล สังเกตได้ว่าอัลกอริทึมที่นำเสนอที่มีการใช้สัมประสิทธิ์ต่ำผ่านมีค่า true alarm rate 5% ซึ่งน้อยกว่าอัลกอริทึม SOM ในขณะที่ค่า false alarm rate มีค่าต่ำกว่า 3% อัลกอริทึมที่นำเสนอมีการใช้สัมประสิทธิ์ต่ำผ่านนั้นสามารถบรรลุค่า true alarm rate ได้สูงถึง 75% ซึ่งมากกว่าการใช้อัลกอริทึม DWT เพียงอย่างเดียว สำหรับอัลกอริทึมที่รวมระหว่าง SOM กับ DWT และอัลกอริทึม DWT เพียงอย่างเดียวที่ได้มีการใช้สัมประสิทธิ์สูงผ่านนั้นให้ประสิทธิภาพในการทำงานที่ต่ำกว่า ทั้งนี้เป็นเพราะว่าสัมประสิทธิ์สูงผ่านเหมาะกับความผิดปกติแบบช่วงสั้น (short duration faults) เช่น ความผิดปกติแบบ spike หรือ sparse ในขณะที่ข้อมูลในรูปที่ 4.8 เป็นข้อมูลที่มีความ

ผิดปกติแบบมีสัญญาณรบกวน (noise faults) ซึ่งส่งผลให้จำนวนของตัวอย่างที่สำเร็จสูงขึ้นเมื่อค่าความแปรปรวนเพิ่มขึ้น

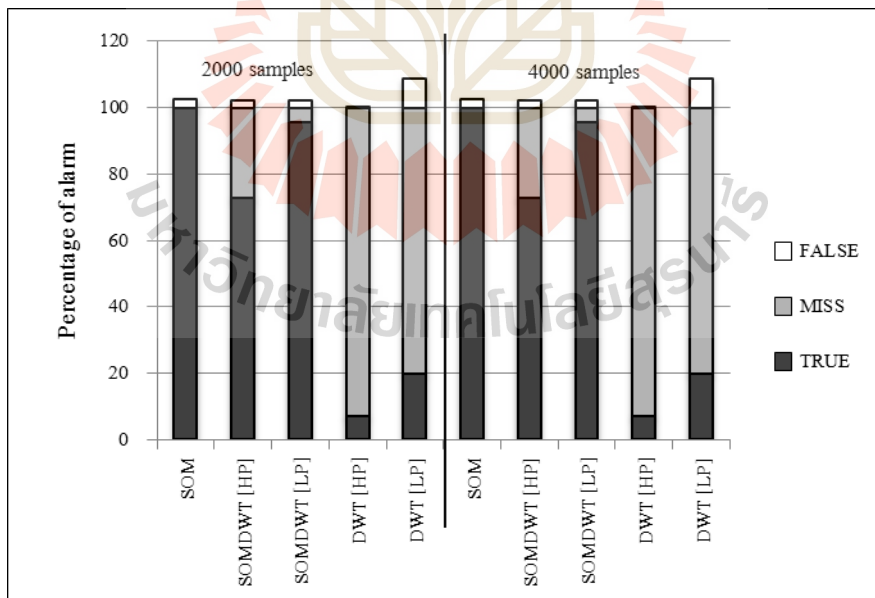
ผลกระทบจากการลดจำนวนการtrainลง ซึ่งแสดงในรูปที่ 4.10 จากผลดังกล่าวพบว่าการลดจำนวนการtrainลงไม่ได้ส่งผลต่อประสิทธิภาพของอัลกอริทึม SOM และ อัลกอริทึม SOM+DWT แต่อย่างใด เช่นเดียวกับรูปที่ 4.11 จะเห็นว่าไม่มีการเปลี่ยนแปลงที่ส่งผลสำคัญต่อเปอร์เซ็นต์ของอัตราส่วนของการเตือนที่ตรวจพบสำหรับ true alarm, miss alarm และ false alarm เมื่อขนาดของเครือข่ายประสาทเทียมลดลง



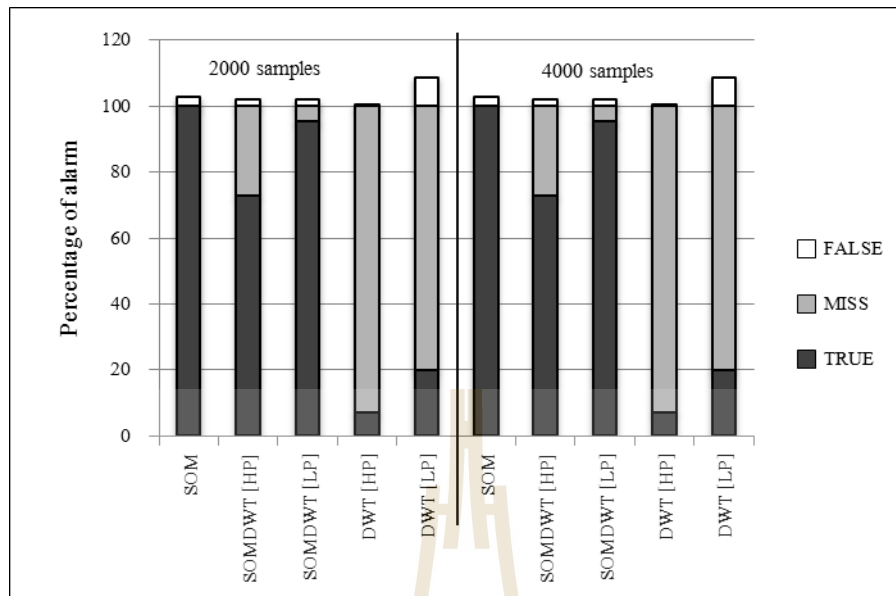
รูปที่ 4.8 ชุดข้อมูลจากโรงปุ๋ยอินทรีย์ชีวภาพจำนวน 10,740 ตัวอย่าง



รูปที่ 4.9 อัตราการตรวจจับสำหรับชุดข้อมูลจากโรงพยาบาลด้วยการสอน
รูปแบบต่างๆ 50 ครั้ง



รูปที่ 4.10 อัตราการตรวจจับสำหรับชุดข้อมูลจากโรงพยาบาลด้วยการสอน
รูปแบบต่างๆ 10 ครั้ง



รูปที่ 4.11 อัตราการตรวจจับสำหรับชุดข้อมูลจาก โรงปู้ยอินทรีย์ชีวภาพด้วยขนาดของเครือข่ายประสาทเทียม 10x10

ผลที่ได้จากชุดข้อมูลจริงและชุดข้อมูลจากโรงปู้ยอินทรีย์ชีวภาพแสดงให้เห็นว่า อัลกอริทึมที่นำเสนอ นั่นคืออัลกอริทึม SOM+DWT มีประสิทธิภาพการทำงานที่ดีเทียบเท่ากับอัลกอริทึม SOM ในขณะที่ใช้ข้อมูลอินพุตเพียงครั้งเดียว (ใช้ level 1 ของอัลกอริทึม DWT) ทั้งนี้เนื่องมาจากอัลกอริทึม DWT สามารถแยกคุณลักษณะของข้อมูลที่เกี่ยวข้องเนื่องกันโดยไม่สูญเสียรายละเอียดที่สำคัญของข้อมูลไป ทำให้ลดการใช้พลังงานที่เสียเปล่า (wasted energy) จากการส่งข้อมูลที่ตรวจวัดได้ทั้งหมดไปยังสถานีฐาน ดังนั้นผลที่ได้จึงสรุปได้ว่า การประยุกต์ใช้ DWT ในโหนดเซ็นเซอร์ไร้สาย เพื่อให้การประมวลผลข้อมูลภายในเครือข่ายและขนาดของข้อมูลในการส่งลดลงซึ่งยังคงรักษาความสามารถในการตรวจจับข้อมูลที่ผิดปกติได้เป็นอย่างดี

อย่างไรก็ตามด้วยคุณลักษณะที่หลากหลายของข้อมูลสามารถส่งผลกระทบต่อประสิทธิภาพในการตรวจจับความผิดปกติของข้อมูลของอัลกอริทึม SOM+DWT ได้ดังที่แสดงในผลจากชุดข้อมูลแบบ NAMOS ดังนั้นการตั้งค่าที่เหมาะสมของอัลกอริทึม เช่น ขนาดของการ train ข้อมูล จะต้องถูกพิจารณาอย่างระมัดระวัง นอกจากนี้ในเทอมของขนาดของเครือข่ายประสาทเทียมพบว่า ยิ่งใช้เครือข่ายประสาทเทียมมากขึ้น ยิ่งสามารถจำแนกประเภทของ SOM ได้ดี อย่างไรก็ตามผลจากชุด

ข้อมูลจริง (real-world datasets) และชุดข้อมูล BOF พบว่าไม่มีการเปลี่ยนแปลงที่ส่งผลสำคัญต่อประสิทธิภาพในการตรวจจับ ในเทอมของการเลือกสัมประสิทธิ์ wavelet นั้น ค่าสัมประสิทธิ์สูงผ่านจะต้องเหมาะสมต่อการตรวจพบจุดของข้อมูลที่มีการเปลี่ยนแปลง ในขณะที่ค่าสัมประสิทธิ์ต่ำผ่านจะต้องเหมาะสมต่อการตรวจพบแนวโน้มของข้อมูลที่มีการเปลี่ยนแปลง ซึ่งการตั้งค่าเหล่านี้สามารถคาดคะเนล่วงหน้าได้จากการพิจารณาธรรมชาติของเซ็นเซอร์ที่ถูกหว่านลงไป ตัวอย่างเช่น ความผิดพลาดที่มาจาก การปรับเทียบ (calibration errors) ของเซ็นเซอร์มีสาเหตุมาจากการชดเชยที่ผิดพลาด (offset faults) โดยที่ค่าจากการวัดให้ผลที่แตกต่างจากค่าจริง นอกจากนี้แรงดันแบตเตอรี่ที่ต่ำ (low battery voltage) ซึ่งมีสาเหตุมาจากการรวมกันของความผิดปกติจากสัญญาณรบกวนและความผิดปกติที่ ในขณะที่ความผิดปกติระยะสั้นมีสาเหตุมาจากความผิดพลาดของซอฟต์แวร์ระหว่างการติดต่อสื่อสารข้อมูล (Sharma, et al., 2010)

4.4 สรุป

ในบทนี้ได้มีการนำข้อมูลที่ได้จากโรงปุ๋ยอินทรีย์ชีวภาพ (BOF dataset) มาใช้ในการทดสอบประสิทธิภาพของอัลกอริทึม โดยพบว่าอัลกอริทึมที่ได้นำเสนอที่ประกอบด้วยเวฟเลทแม่แบบฮาร์ (Haar) ที่มีการใช้สัมประสิทธิ์ต่ำผ่านนั้นมีค่าการแจ้งเตือนจริง 95% โดยที่อัลกอริทึมที่นำเสนอนั้นมีประสิทธิภาพดีเทียบเท่าอัลกอริทึม SOM และมีประสิทธิภาพที่ดีกว่าอัลกอริทึม DWT ถึง 75%

ผลการทดลองแสดงให้เห็นว่าอัลกอริทึมที่ได้นำเสนอสามารถรักษาค่าการตรวจจับความผิดปกติจริงที่ยอมรับได้ ในขณะที่ใช้ข้อมูลป้อนเข้าเพียงครั้งเดียว (ด้วยการใช้ DWT ระดับ 1)

บทที่ 5

บทสรุป

งานวิจัยฉบับนี้นำเสนอกระบวนการตรวจจับความผิดปกติที่เรียกว่าเทคนิคซัพพอร์ตเวกเตอร์แมชชีนคลาสหนึ่ง (one-class support vector machines หรือ OCSVM) และเทคนิคการบีบอัดข้อมูลที่เรียกว่า การแปลงเวฟเลตแบบเต็มหน่วย (Discrete Wavelet Transform หรือ DWT) และวิธีการแปลงเวฟเลตแบบลิฟติง (lifting wavelet transform หรือ LWT) เพื่อบีบอัดข้อมูลและตรวจจับข้อมูลที่ผิดปกติในเครือข่ายเซ็นเซอร์ไร้สาย อัลกอริทึมที่ได้นำเสนอถูกออกแบบเพื่อการส่งข้อมูลบีบอัดที่มีความผิดปกติไปยังสถานีฐานเพื่อที่จะช่วยลดการใช้พลังงาน โดยเปล่าประโยชน์อันเนื่องมาจากการส่งข้อมูลที่ตรวจวัดทั้งหมดไปยังสถานีฐาน งานวิจัยนี้ทำการประเมินประสิทธิภาพของอัลกอริทึมโดยใช้โปรแกรม MATLAB และทดสอบด้วยชุดข้อมูลสังเคราะห์และชุดข้อมูลจริงซึ่งประกอบไปด้วยข้อมูลที่มีความผิดปกติ 3 ชนิด ได้แก่ ความผิดปกติระยะสั้น (short faults) ความผิดปกติจากสัญญาณรบกวน (noise faults) และความผิดปกติคงที่ (constant faults) โดยการทดลองถูกแบ่งเป็น 2 ส่วน ซึ่งสามารถกล่าวโดยสรุปได้ดังนี้

5.1 กระบวนการตรวจจับความผิดปกติโดยใช้สัมประสิทธิ์ DWT

อัลกอริทึม OCSVM ได้ถูกนำมารวมกับวิธี DWT เพื่อศึกษาถึงผลกระทบของการบีบอัดข้อมูลในกระบวนการตรวจจับความผิดปกติ โดยพบว่าการผสมรวมกันของทั้งสองอัลกอริทึมนี้สามารถเพิ่มประสิทธิภาพในการตรวจจับความผิดปกติได้โดยมีค่าอัตราการตรวจจับที่ถูกต้อง (detection rate หรือ DR) ในขณะที่ใช้ขนาดของข้อมูลในการส่งเพียงครึ่งเดียวจากข้อมูลต้นฉบับ สำหรับชุดข้อมูลแบบสังเคราะห์พบว่า อัลกอริทึม OCSVM+DWT ด้วยการใช้ค่าสัมประสิทธิ์ต่ำผ่าน (OCSVM+DWT with LP) มีค่าอัตราการตรวจจับ 100% โดยมีการเพิ่มขึ้นของค่า false positive rate (FPR) เมื่อเปรียบเทียบกับอัลกอริทึมอื่น ซึ่งประกอบด้วย อัลกอริทึม OCSVM alone, OCSVM+DWT ด้วยการใช้ค่าสัมประสิทธิ์สูงผ่าน (OCSVM+DWT with HP) และอัลกอริทึม self-organizing map (SOM) สำหรับชุดข้อมูลจริงพบว่า อัลกอริทึม OCSVM+DWT ด้วยการใช้ค่า

สัมประสิทธิ์ค่าผ่านมีประสิทธิภาพสูงสุด โดยมีอัตราการตรวจจับใกล้เคียง 100% แม้ว่าจะมีค่า FPR สูงขึ้นอันเนื่องมาจากข้อมูลที่มีความผิดปกติระยะสั้นและความผิดปกติจากสัญญาณรบกวน จากผลการทดลองนี้ชี้ให้เห็นว่า ด้วยการบีบอัดข้อมูลและการใช้ข้อมูลอินพุตเพียงครั้งเดียวนั้น อัลกอริทึม OCSVM+DWT (LP) เหมาะสำหรับข้อมูลที่มีความผิดปกติระยะสั้นและความผิดปกติจากสัญญาณรบกวน ในขณะที่อัลกอริทึม SOM+DWT (LP) นั้นเหมาะกับข้อมูลที่มีความผิดปกติระยะสั้นและความผิดปกติคงที่

5.2 กระบวนการตรวจจับความผิดปกติโดยใช้สัมประสิทธิ์ LWT

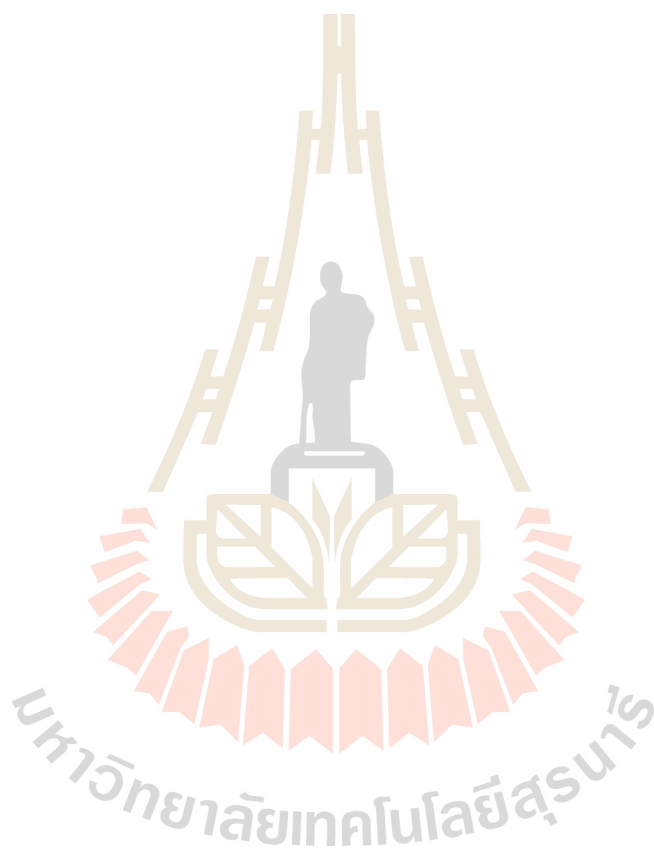
จากการศึกษาพบว่าเทคนิคการบีบอัดข้อมูลที่เรียกว่าวิธีการแปลงเวฟเลตแบบลิฟติง (lifting wavelet transform หรือ LWT) มีความซับซ้อนในการคำนวณน้อย ใช้ memory space น้อย และใช้เวลาในการคำนวณผลต่ำกว่า DWT งานวิจัยนี้จึงได้รวมวิธี OCSVM เข้ากับ LWT (OCSVM + LWT) และทดสอบประสิทธิภาพด้วยชุดข้อมูลสังเคราะห์และชุดข้อมูลจริง สำหรับข้อมูลสังเคราะห์และข้อมูลจริงที่ประกอบด้วยความผิดปกติระยะสั้น อัลกอริทึม OCSVM + LWT มีประสิทธิภาพดีเทียบเท่าอัลกอริทึม OCSVM อัลกอริทึม OCSVM+DWT และอัลกอริทึม OCSVM+PCA สำหรับข้อมูลสังเคราะห์และข้อมูลจริงที่ประกอบด้วยความผิดปกติจากสัญญาณรบกวนและความผิดปกติคงที่พบว่าอัลกอริทึม OCSVM + LWT (LP) และอัลกอริทึม OCSVM + DWT (LP) มีประสิทธิภาพที่ดีกว่าอัลกอริทึม OCSVM และอัลกอริทึม OCSVM+PCA ในทางตรงข้าม อัลกอริทึม OCSVM + LWT (HP) และอัลกอริทึม OCSVM + DWT (HP) มีประสิทธิภาพที่แย่ที่สุด จากผลการทดลองนี้พิสูจน์ให้เห็นว่าวิธี LWT มีความต้องการในเทอมของการคำนวณและเวลาที่ใช้ในการคำนวณต่ำกว่าวิธี DWT ผลการทดลองชี้ให้เห็นว่าอัลกอริทึม OCSVM + LWT นั้นเหมาะสมที่จะนำไปใช้งานจริงในเครือข่ายเซ็นเซอร์ไร้สาย

5.3 กระบวนการตรวจจับความผิดปกติในเครือข่ายเซ็นเซอร์ไร้สายสำหรับเฝ้าระวังโรงผลิตปุ๋ยอินทรีย์ชีวภาพ

งานวิจัยฉบับนี้ได้มีการนำข้อมูลที่ได้จากโรงปุ๋ยอินทรีย์ชีวภาพ (bioorganic fertilizer production plant หรือ BOF dataset) มาใช้ในการทดสอบประสิทธิภาพของอัลกอริทึม โดยพบว่า

อัลกอริทึมที่ได้นำเสนอซึ่งประกอบด้วยเวฟเลตแม่แบบฮาร์ (Haar) ที่มีการใช้สัมประสิทธิ์ต่ำผ่านนั้น มีอัตราการแจ้งเตือนจริง (true alarm rates) 95% โดยที่อัลกอริทึมที่นำเสนอนั้นมีประสิทธิภาพ เทียบเท่าอัลกอริทึม SOM และมีประสิทธิภาพที่ดีกว่าอัลกอริทึม DWT ถึง 75%

ผลการทดลองแสดงให้เห็นว่าอัลกอริทึมที่ได้นำเสนอสามารถรักษาค่าการตรวจจับความผิดปกติจริงที่ยอมรับได้ ในขณะที่ใช้ข้อมูลป้อนเข้าเพียงครั้งเดียว (ด้วยการใช้ DWT level 1)



บรรณานุกรม

- Acharya T. and Chakrabarti C. (2006). A Survey on Lifting-Based Discrete Wavelet Transform Architectures. In **Proceedings of the Journal of VLSI Signal Processing**. 42: pp. 321-339.
- Burrus, C. S., Gopinath, R. A. and Guo, H. (1998) **Introduction to Wavelets and Wavelet Transforms**. New Jersey: Prentice-Hall
- Bruce, L. M., Koger, C. H., and Li, J. (2002). Dimensionality Reduction of Hyperspectral Data Using Discrete Wavelet Transform Feature Extraction. In **Proceedings of the IEEE Transactions International Geosciences and Remote Sensing**. 40 (10): pp. 2331-2338.
- Capo-Chichi, E. P., Guyennet, H., and Friedt, J. M. (2009). K-RLE : A new Data Compression Algorithm for Wireless Sensor Network. In **Proceedings of the Third International Conference on Sensor Technologies and Applications**. pp. 502-507.
- Chandola, V., Banerjee, A., and Kumar, V. (2009). Anomaly Detection: A Survey. In **Proceedings of the ACM Computing Surveys**. 41 (3): pp. 1 - 72.
- Ciancio, A., and Ortega, A. (2005). A distributed wavelet compression algorithm for wireless multihop sensor networks using lifting. In **Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing 2005**. 4: pp. IV-825-IV-828.
- Du, P., Tan, K., and Xing, X. (2010). Wavelet SVM in Reproducing Kernel Hilbert Space for hyper spectral remote sensing image classification. In **Proceedings of the Optics Communications**. 283 (24): pp. 4978-4984.
- Dwinnell, W. (2010). **Principal Components Analysis**. [Online]. Available: <http://matlabdatamining.blogspot.com/2010/02/principal-components-analysis.html>

- Goh, H.G., et al. (2007). Agriculture monitoring. In N. P. Mahalik (ed.). **Sensor Networks and Configuration**. New York: Springer. pp. 439-462.
- Haykin, S. (1999). **Neural Networks**, 2nd edition, New Jersey: Prentice-Hall, Inc
- Hsu, C. W., Chang, C. C., and Lin, C. J. (2003). **A Practical Guide to Support Vector Classification**. [Online]. Available: <http://www.csie.ntu.edu.tw/~cjlin>
- Kiziloren, T., and Germen, E. (2009). Anomaly detection with Self-Organizing Maps and effects of Principal Component Analysis on feature vectors. In **Proceedings of the 5th International Conference on Natural Computation**. 6: pp. 509-513.
- Kohonen, T. (1990). Self-Organizing Maps. In **Proceedings of IEEE**, vol. 78, no. 9, pp. 1464-1480
- Laskov, P., Schafer, C., and Kotenko, I. (2004). Intrusion detection in unlabeled data with quarter sphere support vector machines. In **Proceedings of the Detection of Intrusions and Malware & Vulnerability Assessment**. 27 (4): pp. 228-236.
- Li, X. L., Zhang, J. W., and Fang, W. H. (2009). The research of data compression algorithm based on lifting wavelet transform for wireless sensor network. In **Proceedings of the International Conference on Apperceiving Computing and Intelligence Analysis**. pp. 228-233.
- Lin, S., et al. (2005). A Data Compression Technique for Sensor Networks with Dynamic Bandwidth Allocation. In **Proceedings of the 12th International Symposium on Temporal Representation and Reasoning**. pp. 186-188.
- Lutsa, J., et al. (2010). A tutorial on support vector machine-based methods for classification problems in chemometrics. In **Proceedings of the Analytica Chimica Acta**. 665 (2): pp. 129-145.

- Manjunath, A., and Ravikumar, H. M. (2010). Comparison of discrete wavelet transform (DWT) lifting wavelet transform (LWT) stationary wavelet transform (SWT) and S-Transform in power quality analysis. In **Proceedings of the European Journal of Scientific Research**. 39 (4): pp. 569-576.
- Marcelloni, F., and Vecchio, M. (2008). A Simple Algorithm for Data Compression in Wireless Sensor Networks. In **Proceedings of the IEEE Communications Letters**. 12 (6): pp. 411-413.
- Min, L., and Dongliang, W. (2009). Anomaly Intrusion Detection Based on SOM. In **Proceedings of the WASE International Conference on Information Engineering**, pp. 40-43.
- Network Aquatic Microbial Observing System, Online, 2006
- Rajasegarar, S., et al. (2006). Distributed Anomaly Detection in Wireless Sensor Networks. In **Proceedings of the 10th IEEE Singapore International Conference on Communication systems**, pp. 1-5.
- Rajasegarar, S., et al. (2007). Quarter sphere based distributed anomaly detection in wireless sensor networks. In **Proceedings of the IEEE International Conference on Communications**. New Jersey, USA: IEEE Operations Center. pp. 3864-3869.
- Rajasegarar, S., Leckie, C., and Palaniswami, M. (2008). Anomaly detection in wireless sensor networks. In **Proceedings of the IEEE Wireless Communications**. 15 (4): pp. 34-40.
- Sharma, A.B., Golubchik, L., and Govindan, R. (2010). Sensor faults: detection methods and prevalence in real-world datasets. In **Proceedings of the ACM Transactions on Sensor Networks**. 6 (3): pp. 1-34.
- Siripanadorn, S., Hattagam, W., and Teaumroong, N. (2010) Anomaly detection in wireless sensor networks using Self-Organizing Map and Wavelets. In **Proceedings of the International Journal of Communications**. 4 (3): pp. 74-83.

- Siripanadorn, S., Hattagam, W., and Teaumroong, N. (2010a) Anomaly detection in wireless sensor networks using Self-Organizing Map and Wavelets. In **Proceedings of the International Journal of Communications**. 4 (3): pp. 74-83.
- Siripanadorn, S., Hattagam, W., and Teaumroong, N. (2010b) Anomaly detection using self-organizing map and wavelets in wireless sensor networks. In **Proceedings of the International Journal of Communications**. 4 (3): pp. 74-83.
- Smith, L. I. (2002). **A tutorial on Principal Components Analysis**. [Online]. Available: http://www.cs.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf
- Sweldens, W. (1998). The lifting scheme: a construction of second generation wavelets. In **Proceedings of the Siam Journal on Mathematical Analysis**. 29: pp. 511-546.
- Takiangam, S., and Hattagam, W., (2011). Discrete Wavelet Transform and One-Class Support Vector Machines for anomaly detection in wireless sensor networks, **Proceedings of International Symposium on Intelligent Signal Processing and Communications Systems**.
- Tax, D. M. J., and Duin, R. P. W. (2004). Support vector data description. . In **Proceedings of the Machine Learning**. 54 (1): pp. 45-66.
- The INTEL Lab, Online, 2004
- The SensorScope Lausanne Urban Canopy Experiment Project: LUCE, Online, 2006
- Wang, F., et al. (2010). A Model Based on Hybrid Support Vector Machine and Self-Organizing Map for Anomaly Detection. In **Proceedings of the 2010 International Conference on Communications and Mobile Computing**. 1: pp. 97-101.
- Watson, M. J., et al. (1995). Wavelet techniques in the compression of process data. In **Proceedings of the American Control Conference**. 2: pp. 1265-1269.

- Xu, Y., and Chow, T. W.S. (2010). Efficient Self-Organizing Map Learning Scheme Using Data Reduction Preprocessing. In **Proceedings of the World Congress on Engineering 2010**. 2183 (1): pp. 273-276.
- Yao, Y., et al. (2010). Online Anomaly Detection for Sensor Systems: a Simple and Efficient Approach. In **Proceedings of the Journal Performance Evaluation**. 67 (11): pp. 1-24.
- Zhang, Y., Meratnia, N., and Havinga, P. (2009). Adaptive and Online One-Class Support Vector Machine-based Outlier Detection Techniques for Wireless Sensor Networks. In **Proceedings of the 2009 International Conference on Advanced Information Networking and Applications Workshops**. pp. 990-995.



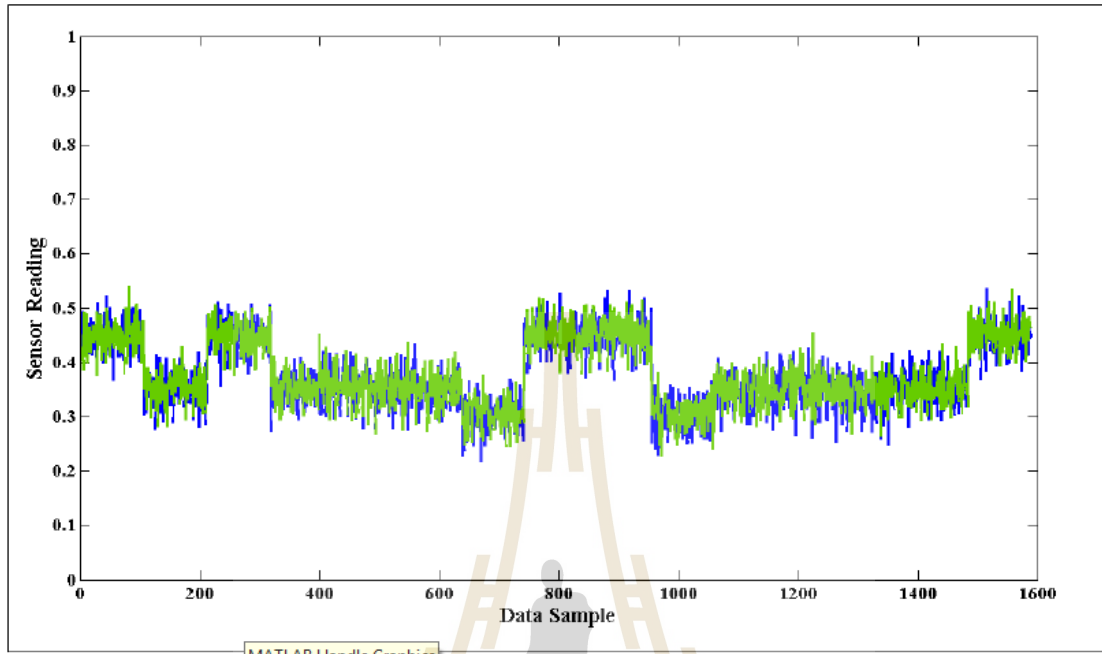
ภาคผนวก ก

ชุดข้อมูลสำหรับการทดลองบทที่ 3

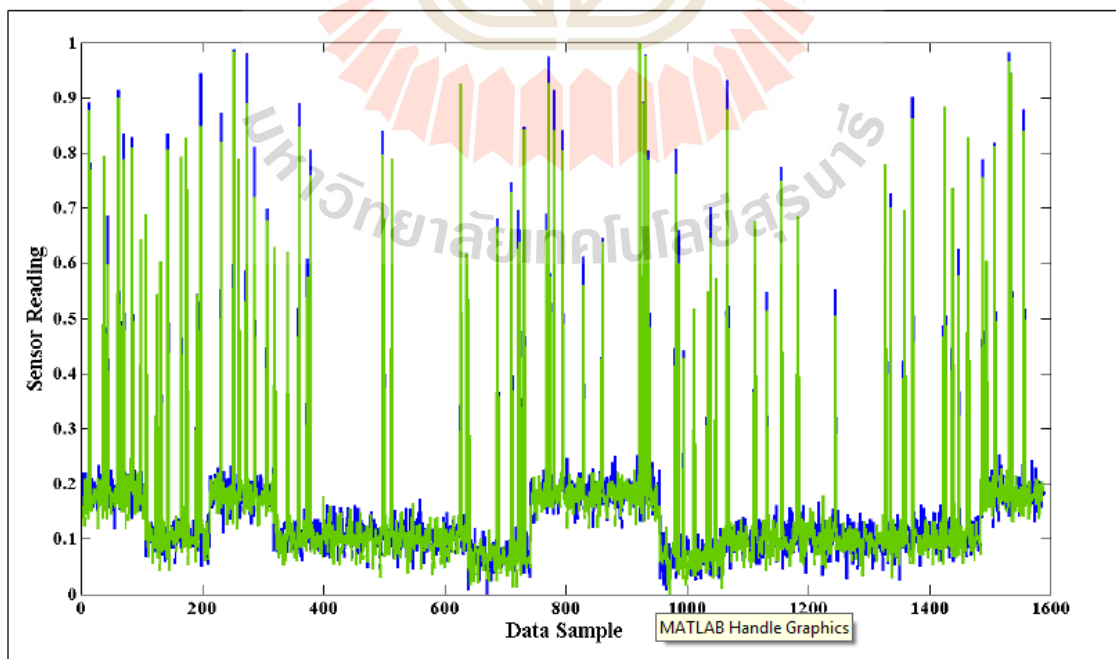
มหาวิทยาลัยเทคโนโลยีสุรนารี

ชุดข้อมูลสำหรับบทที่ 3

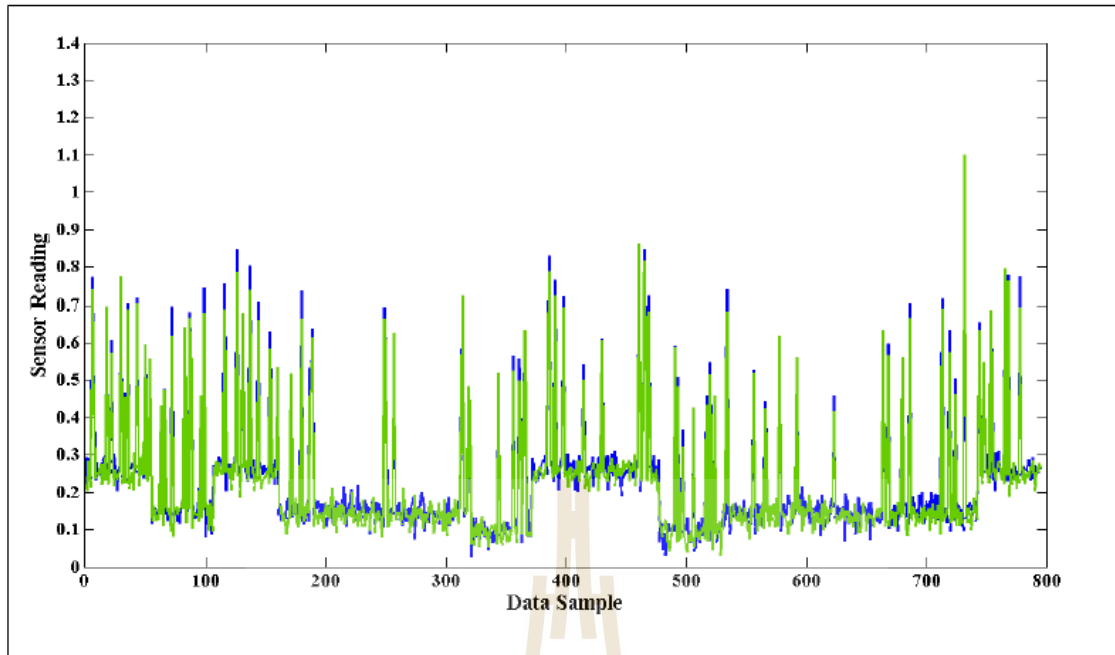
1. ข้อมูลสังเคราะห์



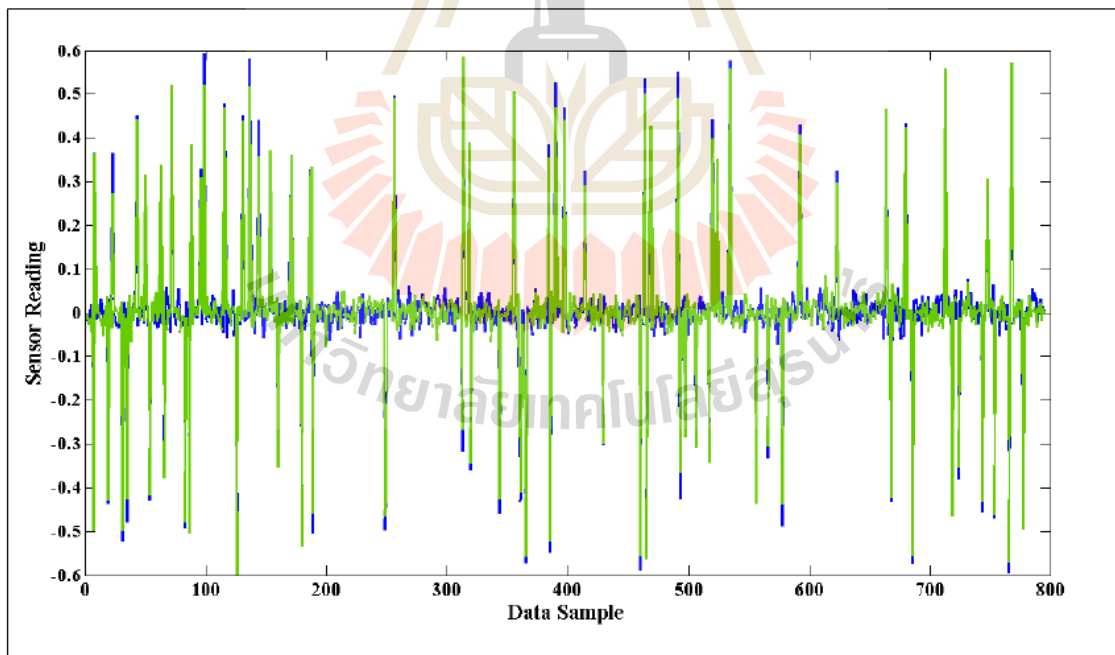
รูปที่ ก.1 ข้อมูลสังเคราะห์เมื่อปราศจากความผิดพลาดสำหรับการฝึก SOM อัลกอริทึม



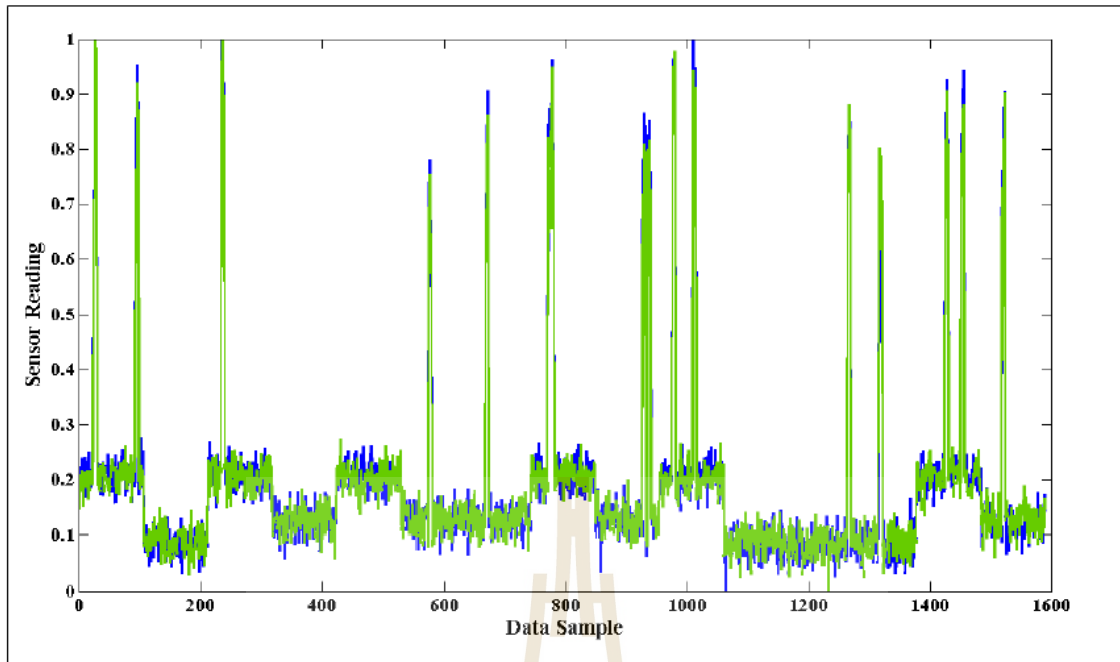
รูปที่ ก.2 ข้อมูลสังเคราะห์ที่ประกอบด้วยความผิดพลาด 1/80



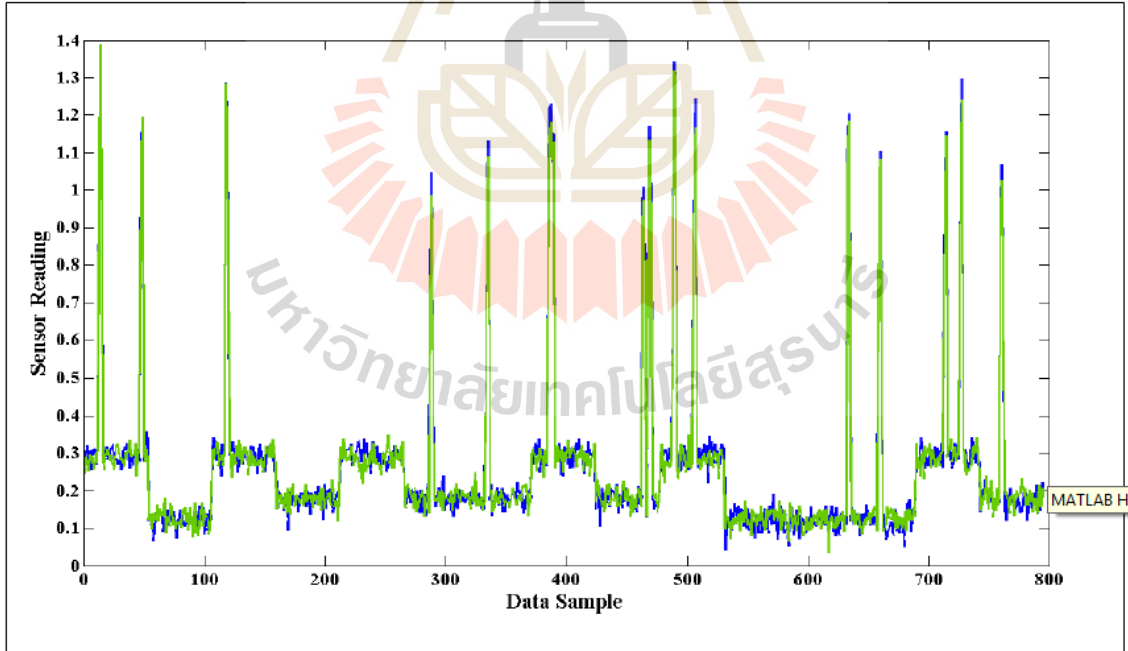
รูปที่ 3.3 DWT ค่าสัมประสิทธิ์ต่ำผ่านของข้อมูลสังเคราะห์ที่ประกอบด้วยความผิดพลาด 1/80



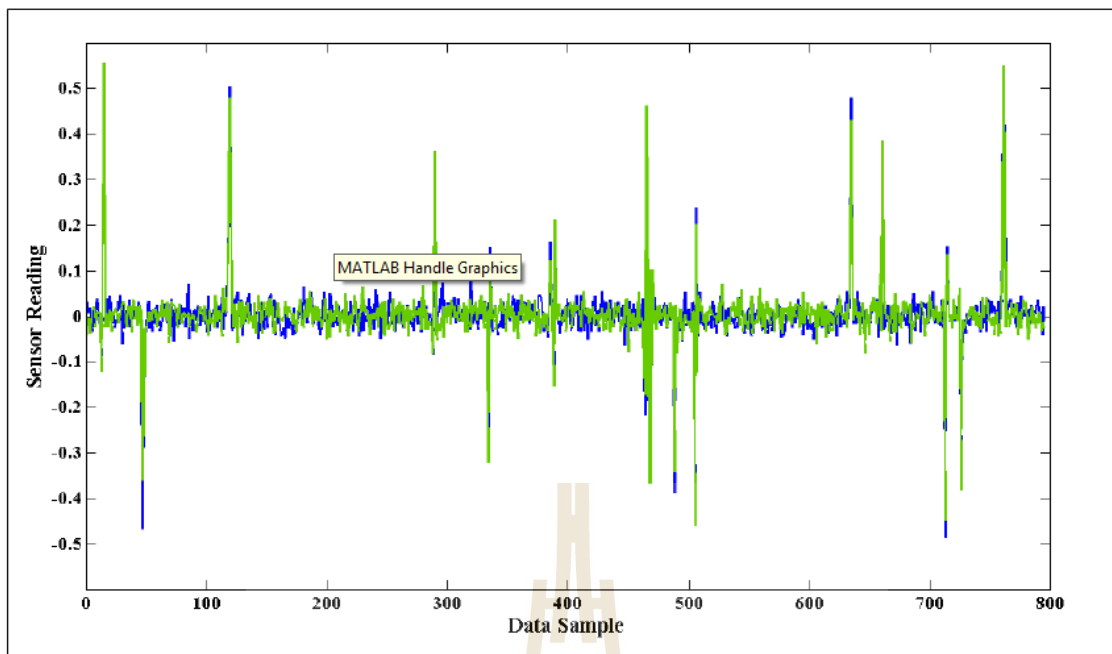
รูปที่ 3.4 DWT ค่าสัมประสิทธิ์สูงผ่านของข้อมูลสังเคราะห์ที่ประกอบด้วยความผิดพลาด 1/80



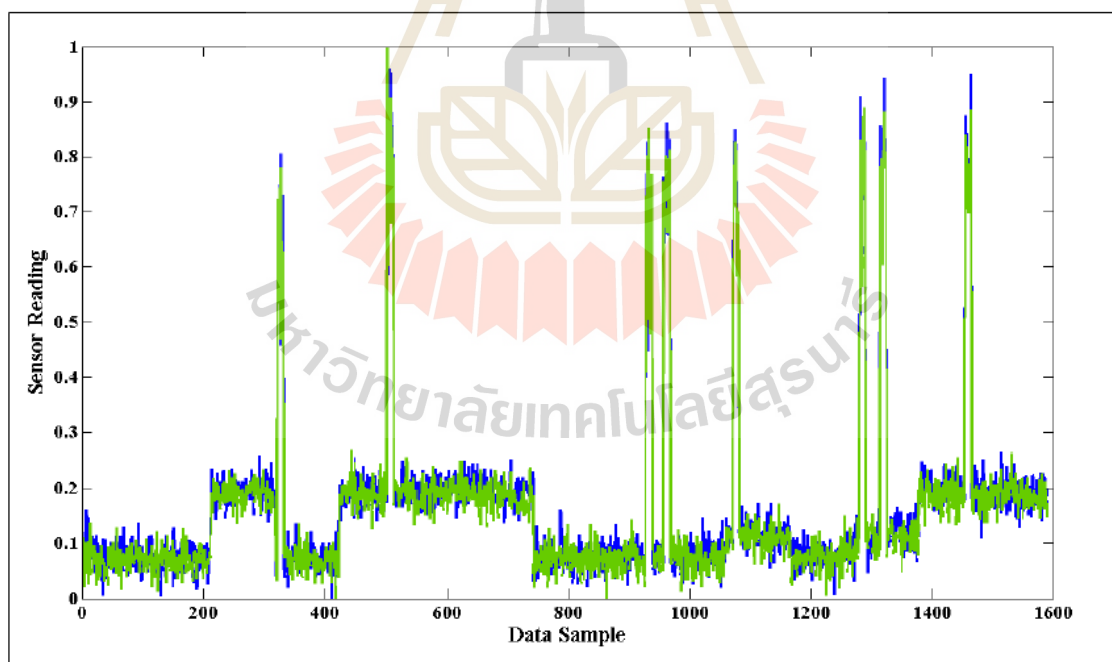
รูปที่ ก.5 ข้อมูลสังเคราะห์ที่ประกอบด้วยความผิดพลาด 5/16



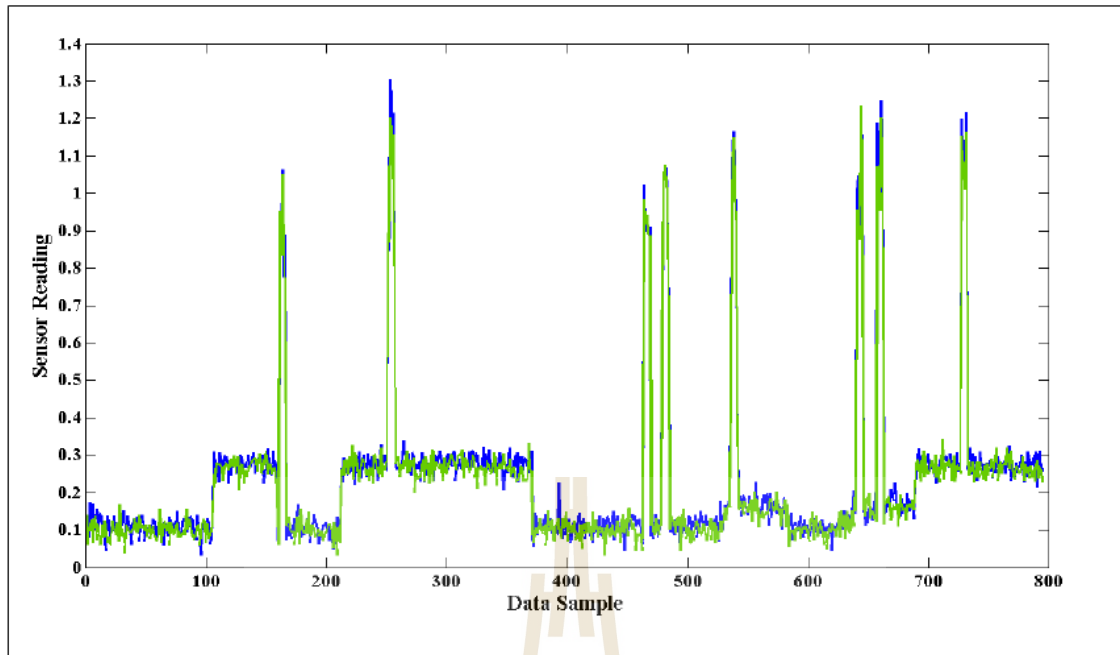
รูปที่ ก.6 DWT ค่าสัมประสิทธิ์ต่ำผ่านของข้อมูลสังเคราะห์ที่ประกอบด้วยความผิดพลาด 5/16



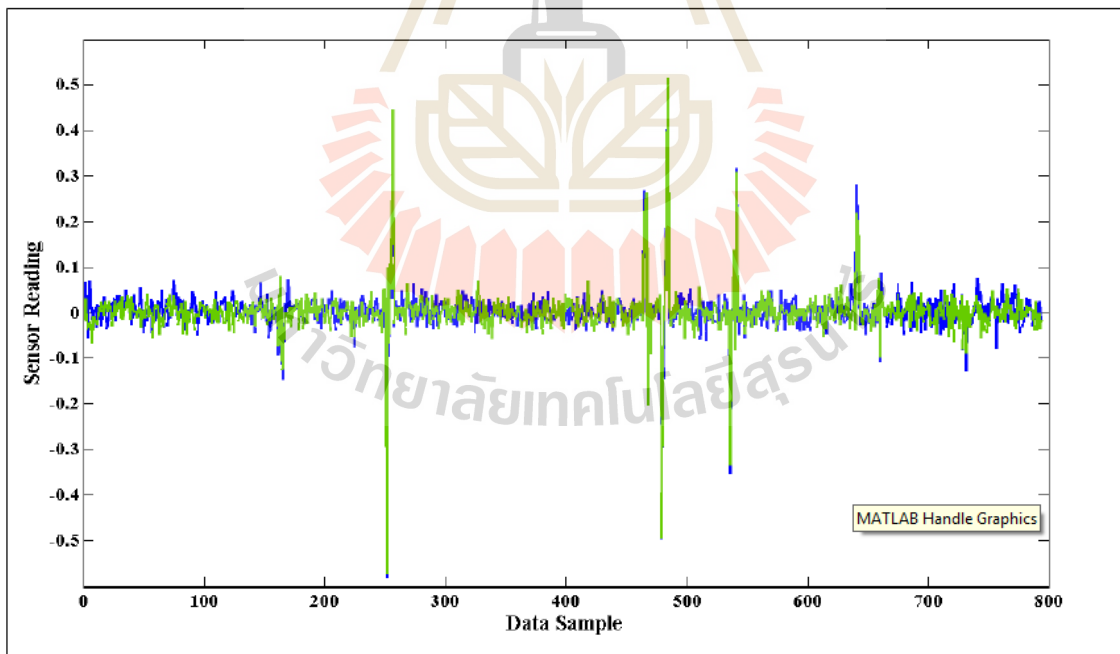
รูปที่ ก.7 DWT ค่าสัมประสิทธิ์สูงผ่านของข้อมูลสังเคราะห์ที่ประกอบด้วยความผิดพลาด 5/16



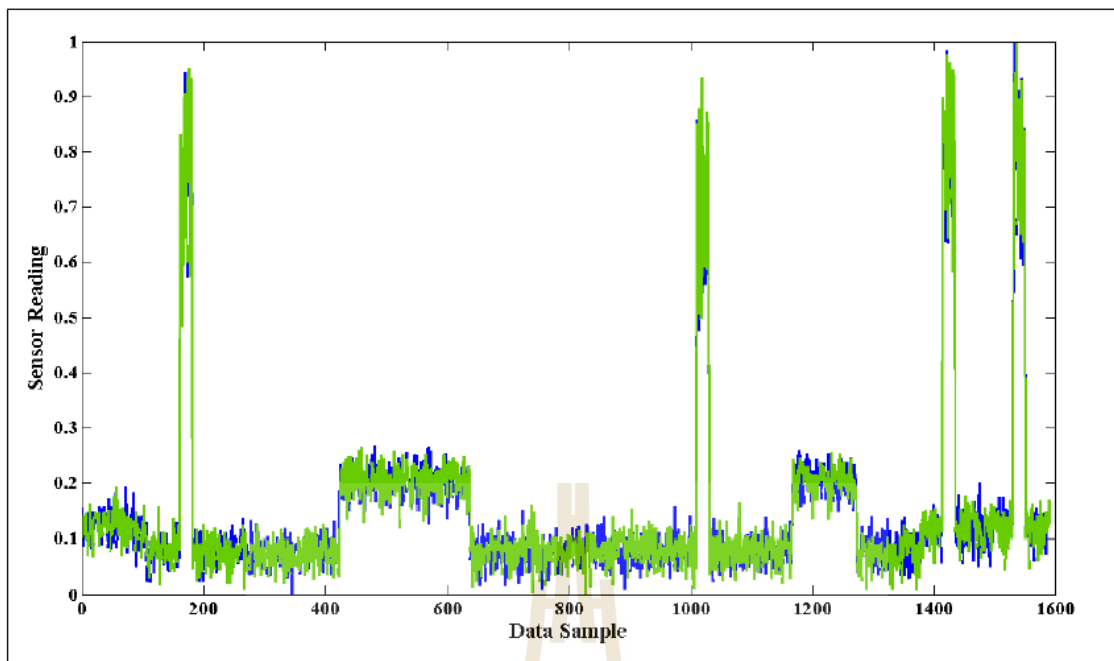
รูปที่ ก.8 ข้อมูลสังเคราะห์ที่ประกอบด้วยความผิดพลาด 10/8



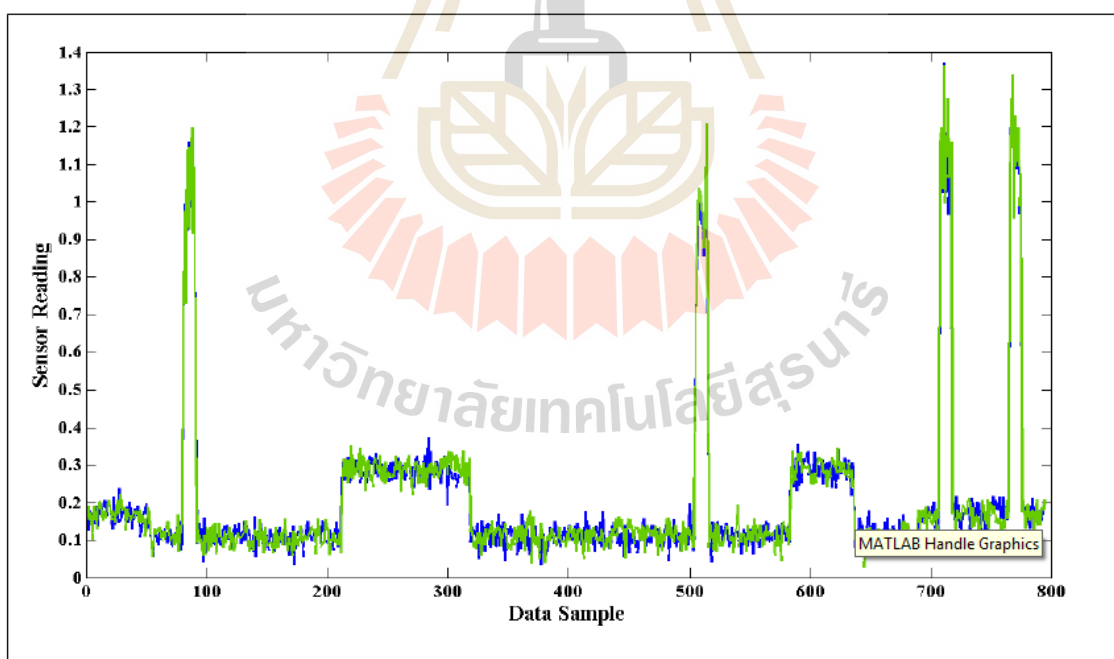
รูปที่ ๙.๙ DWT ค่าสัมประสิทธิ์ต่ำผ่านของข้อมูลสังเคราะห์ที่ประกอบด้วยความผิดพลาด 10/8



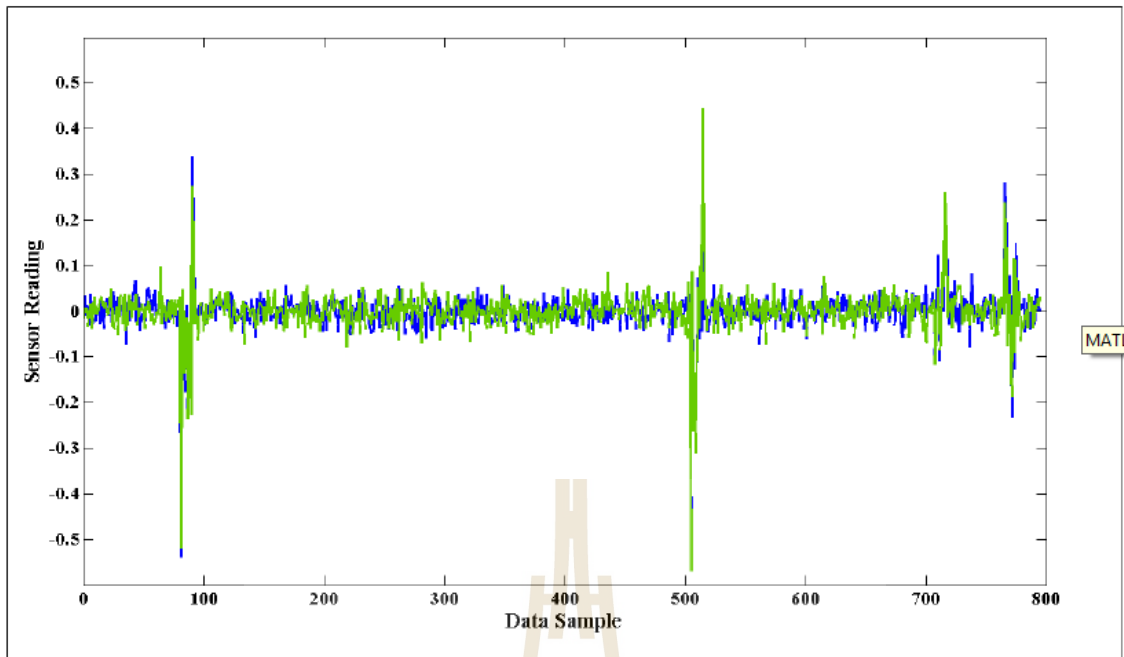
รูปที่ ๙.๑๐ DWT ค่าสัมประสิทธิ์สูงผ่านของข้อมูลสังเคราะห์ที่ประกอบด้วยความผิดพลาด 10/8



รูปที่ ก.11 ข้อมูลสังเคราะห์ที่ประกอบด้วยความผิดพลาด 20/4

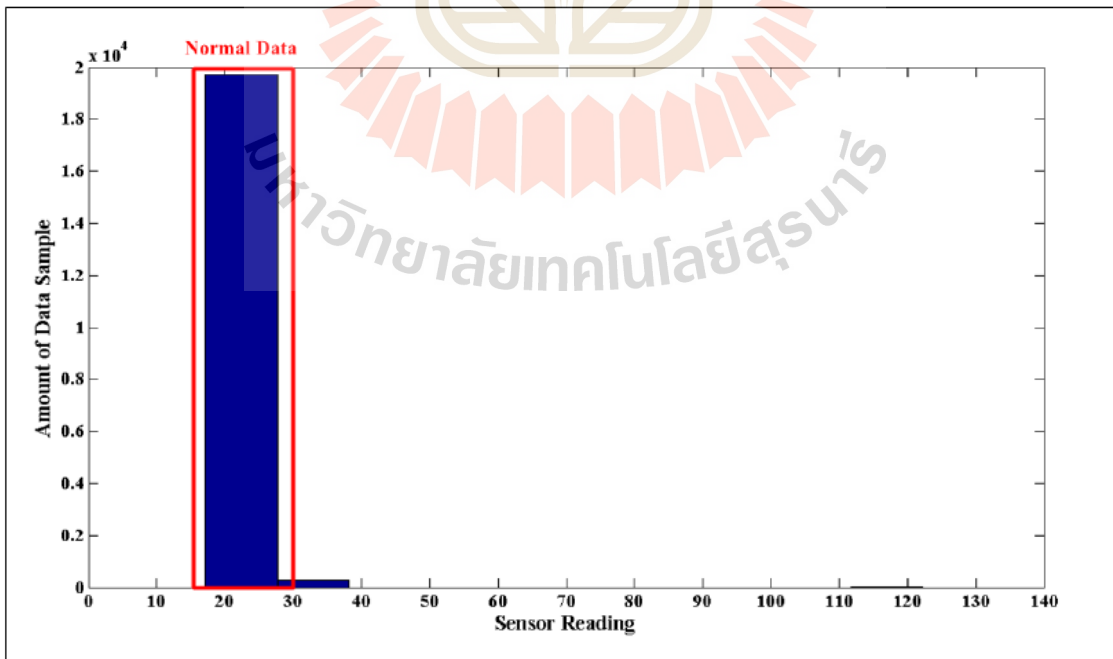


รูปที่ ก.12 DWT ค่าสัมประสิทธิ์ต่ำผ่านของข้อมูลสังเคราะห์ที่ประกอบด้วยความผิดพลาด 20/4

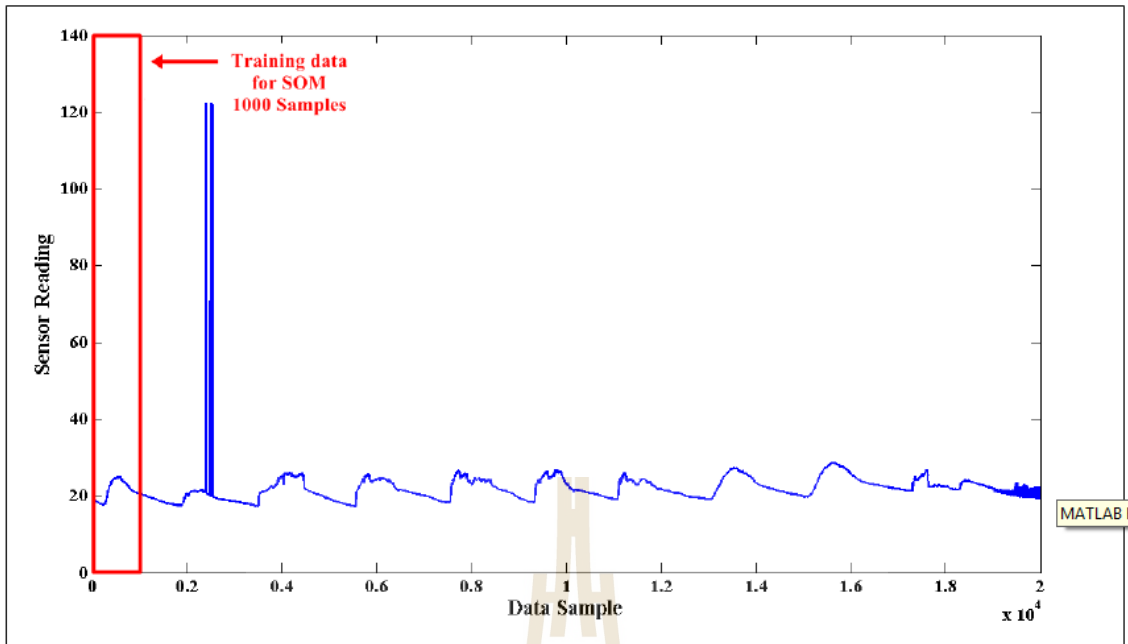


รูปที่ ก.13 DWT ค่าสัมประสิทธิ์สูงผ่านของข้อมูลสังเคราะห์ที่ประกอบด้วยความผิดพลาด 20/4

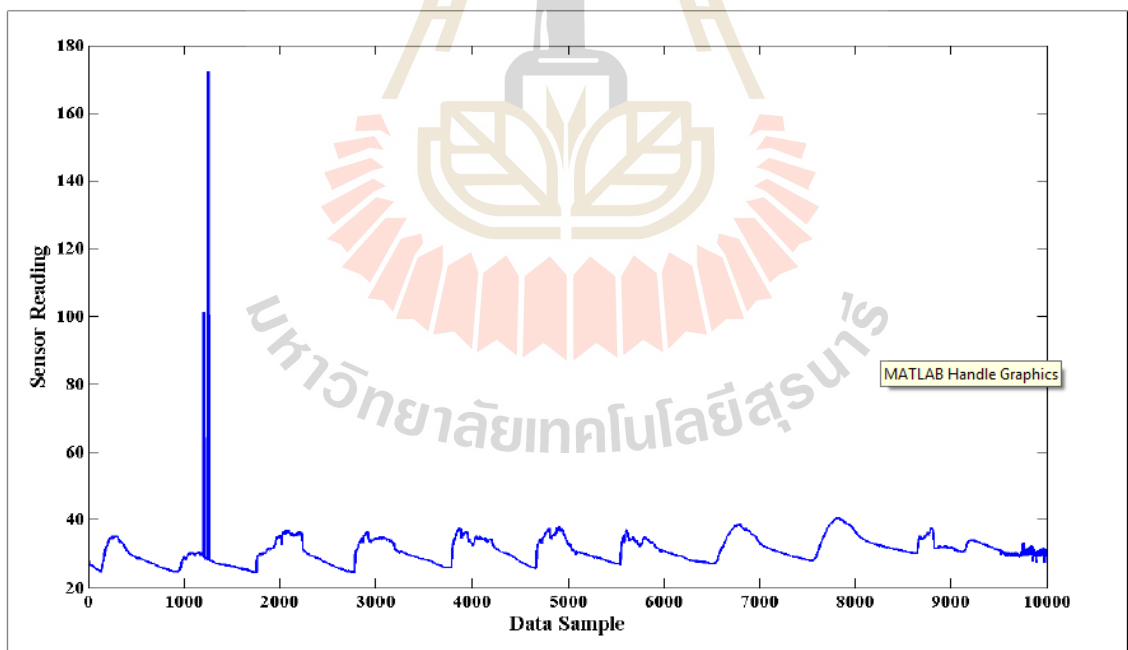
2. ชุดข้อมูล INTEL



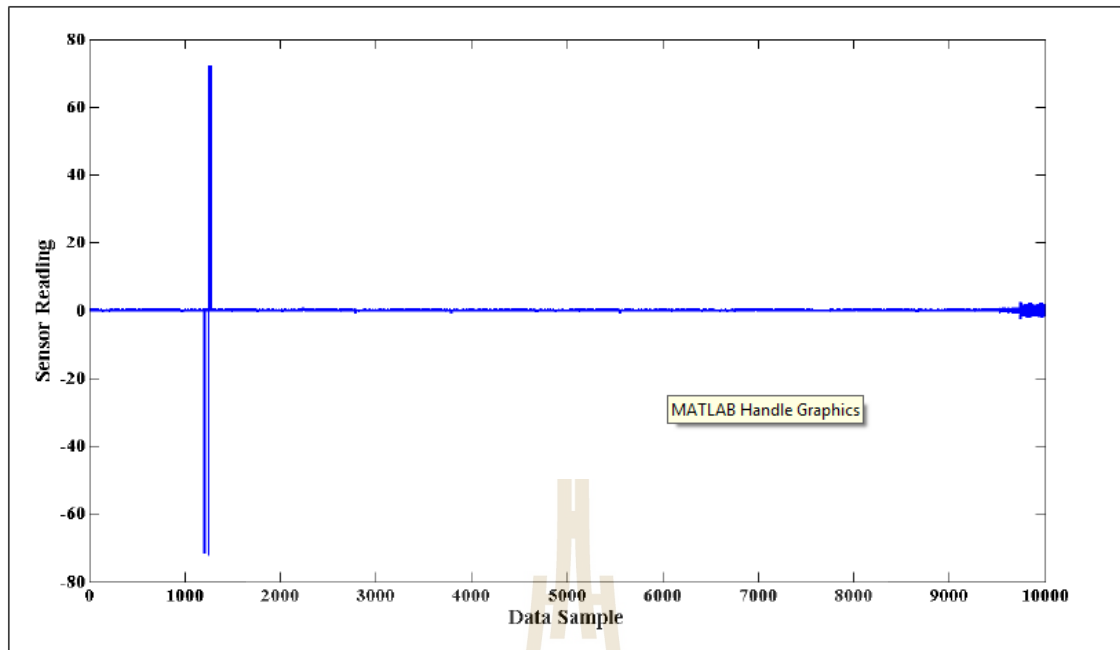
รูปที่ ก.14 ฮิสโตแกรมของชุดข้อมูล INTEL (อุณหภูมิที่อ่านได้)



รูปที่ ก.15 ชุดข้อมูล INTEL (อนุกรมที่อ่านได้)

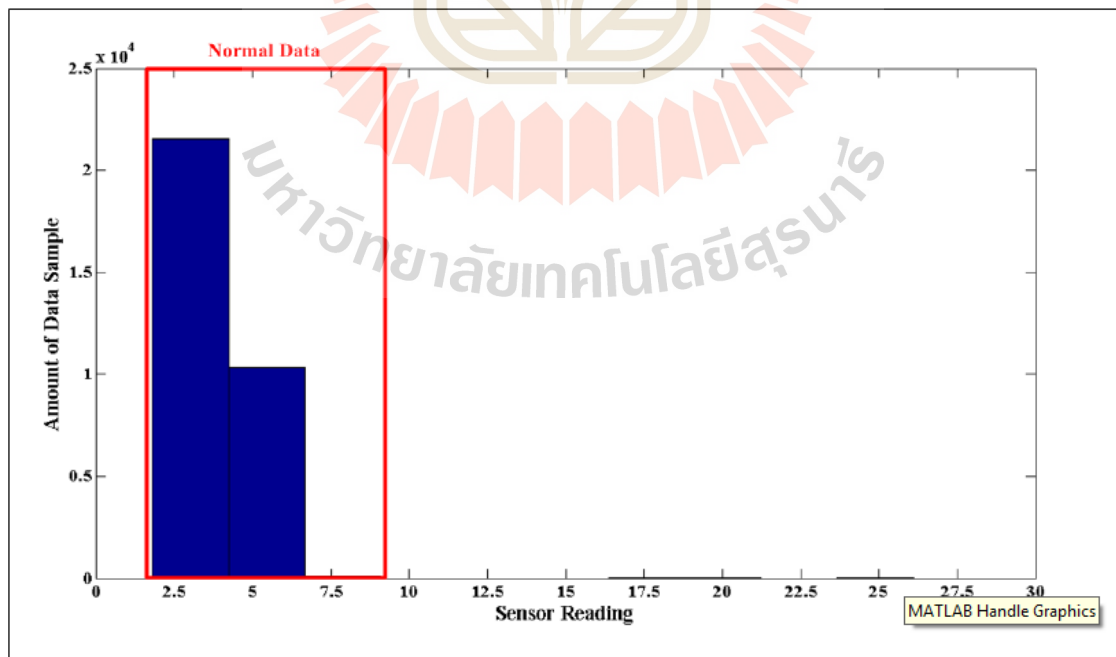


รูปที่ ก.16 DWT ค่าสัมประสิทธิ์ที่คำนวณของชุดข้อมูล INTEL (อนุกรมที่อ่านได้)

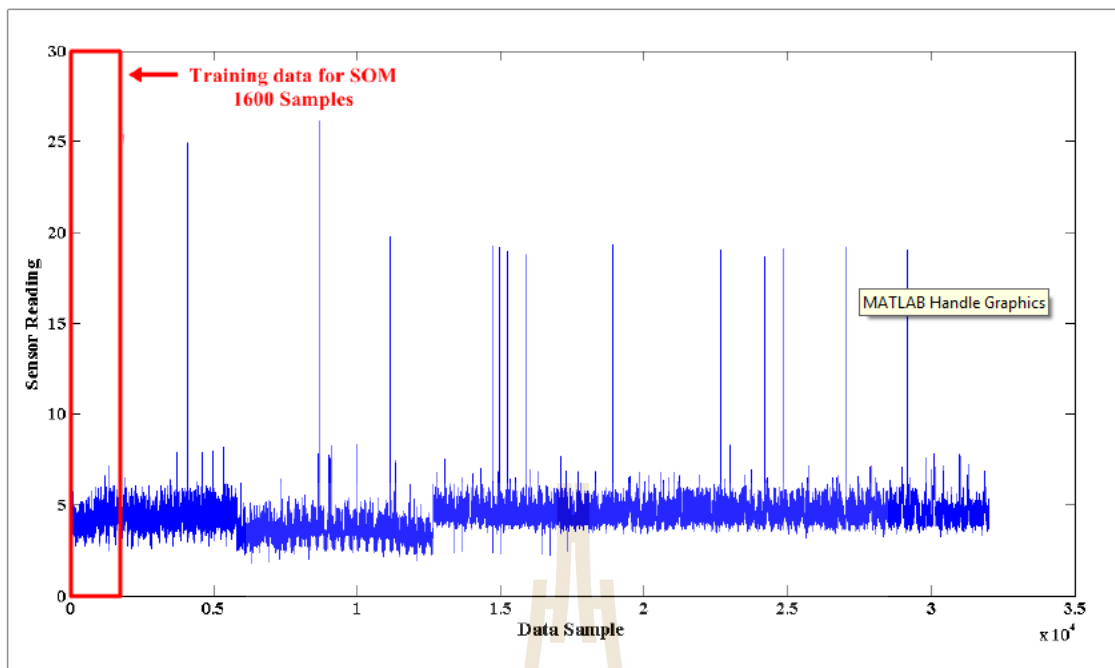


รูปที่ ก.17 DWT ค่าสัมประสิทธิ์สูงผ่านของชุดข้อมูล INTEL (อุณหภูมิที่อ่านได้)

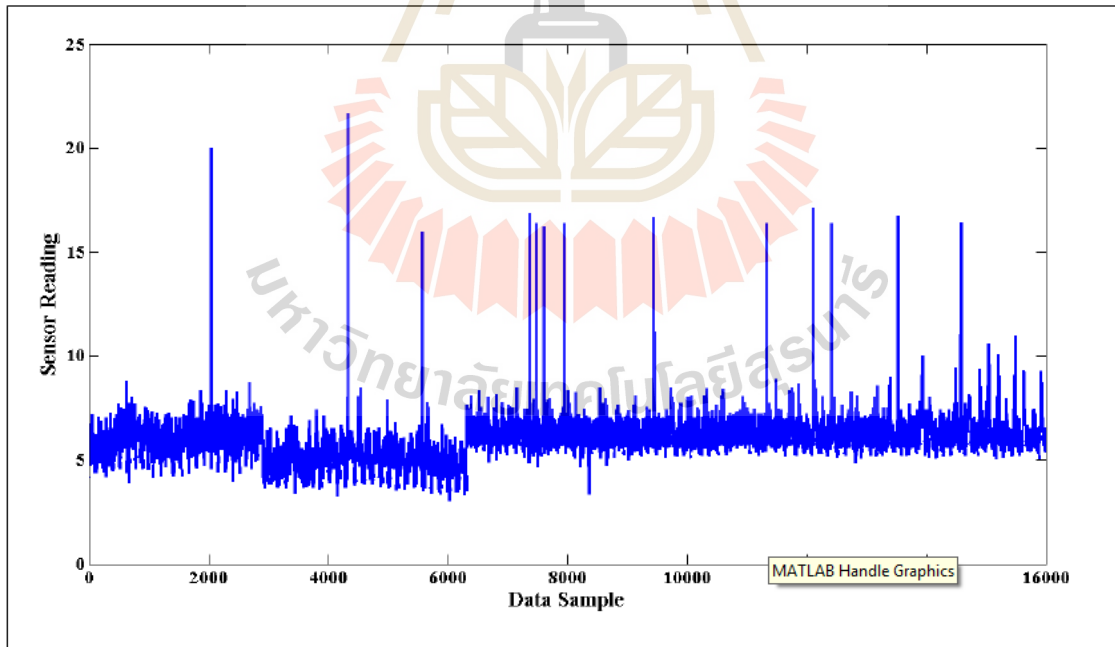
3. ชุดข้อมูลสถานีเซ็นเซอร์สโครปที่ 39 (SS39)



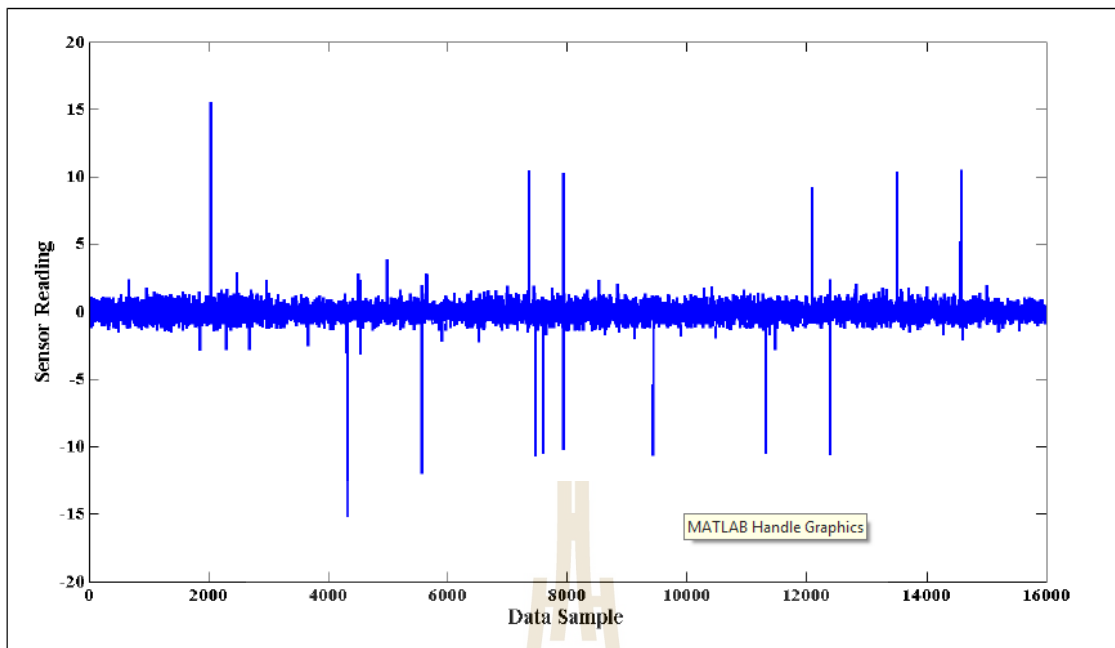
รูปที่ ก.18 ฮิสโตแกรมของชุดข้อมูลสถานีเซ็นเซอร์สโครปที่ 39 (SS39)



รูปที่ ก.19 ชุดข้อมูลสถานีเซ็นเซอร์สโครปที่ 39 (SS39), (ค่าglobalที่อ่านได้ในขณะนั้น)

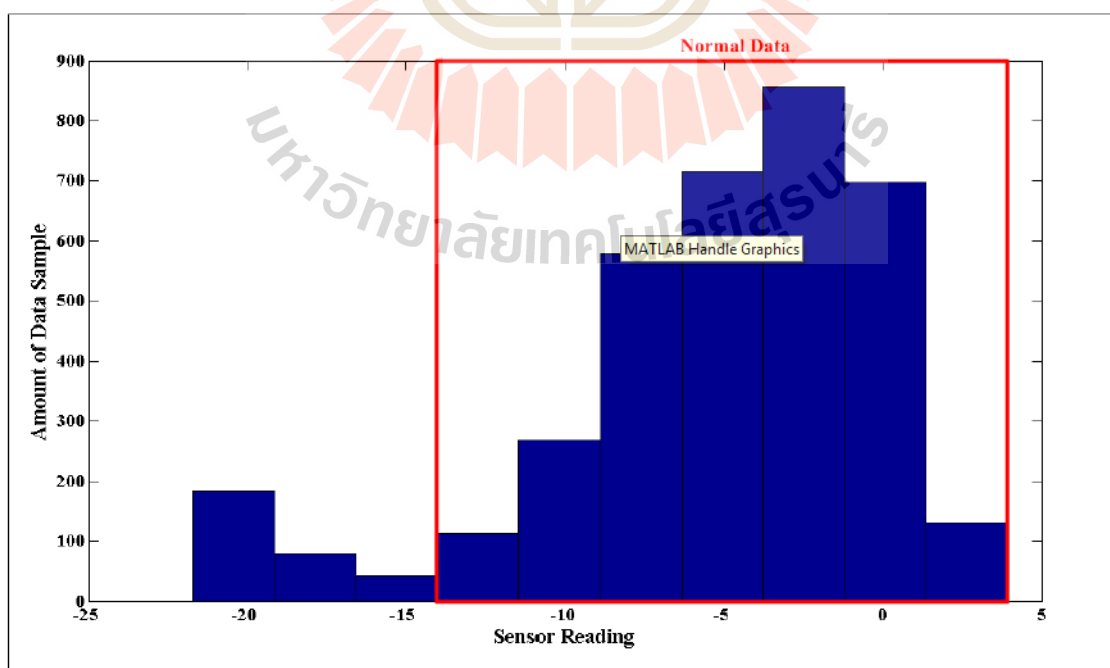


รูปที่ ก.20 DWT ค่าสัมประสิทธิ์ที่ต่ำผ่านของชุดข้อมูลสถานีเซ็นเซอร์สโครปที่ 39 (SS39),
(ค่าglobalที่อ่านได้ในขณะนั้น)

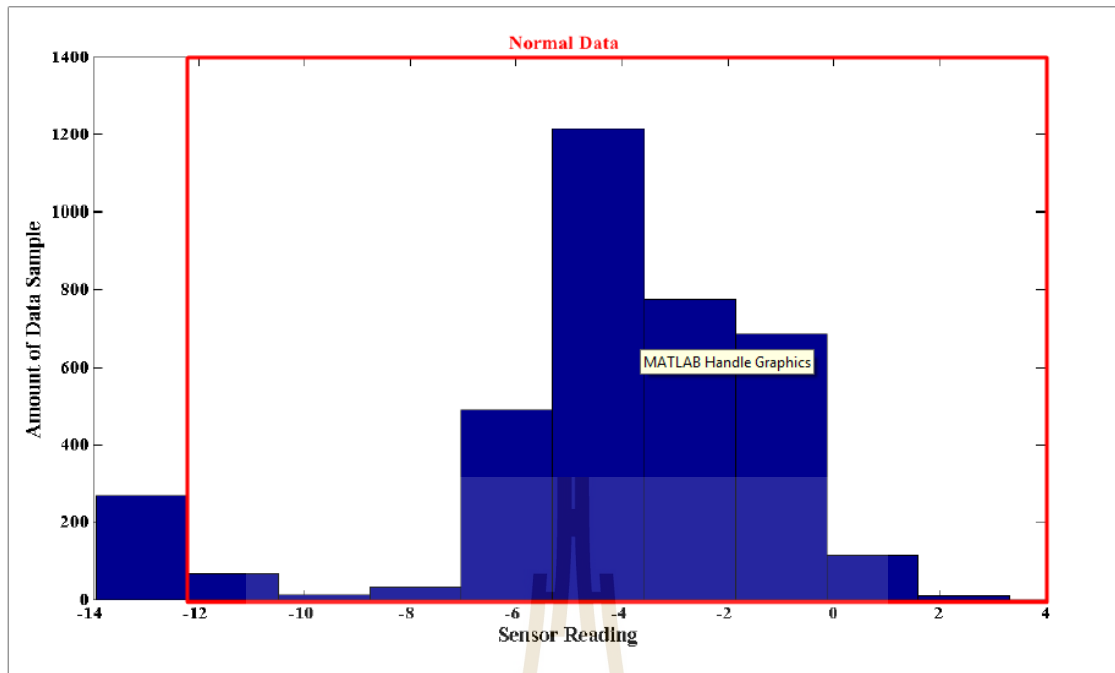


รูปที่ ก.21 DWT ค่าสัมประสิทธิ์ที่สูงผ่านของชุดข้อมูลสถานีเซ็นเซอร์สโครปที่ 39 (SS39),
(ค่าglobalที่อ่านได้ในขณะนั้น)

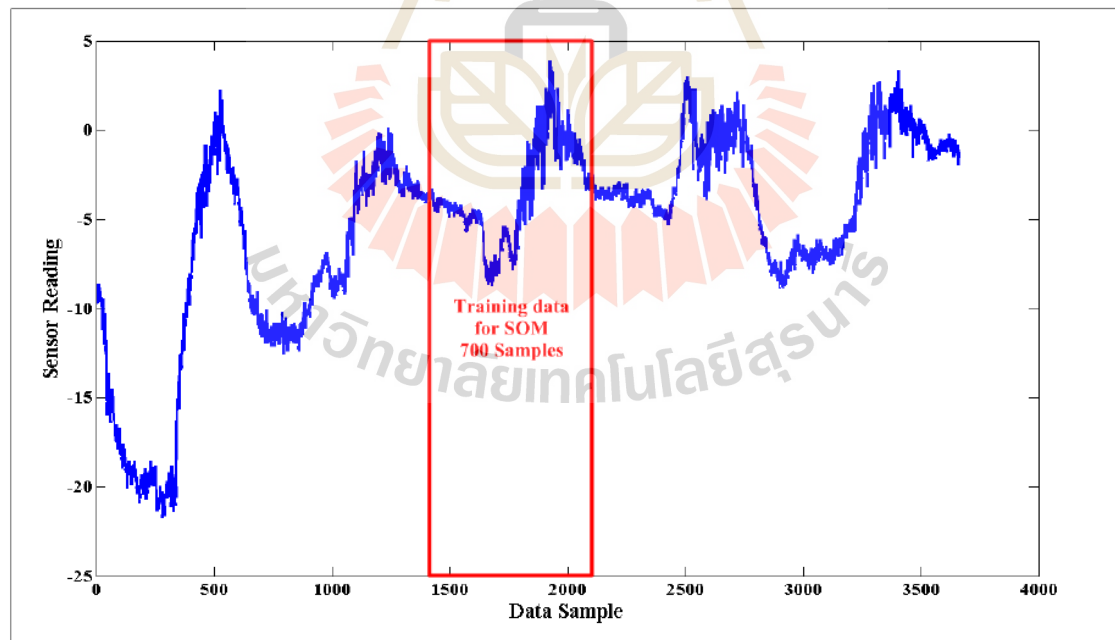
4. ชุดข้อมูลสถานีเซ็นเซอร์สโครป pdg2008-metro-1 (pdg2008)



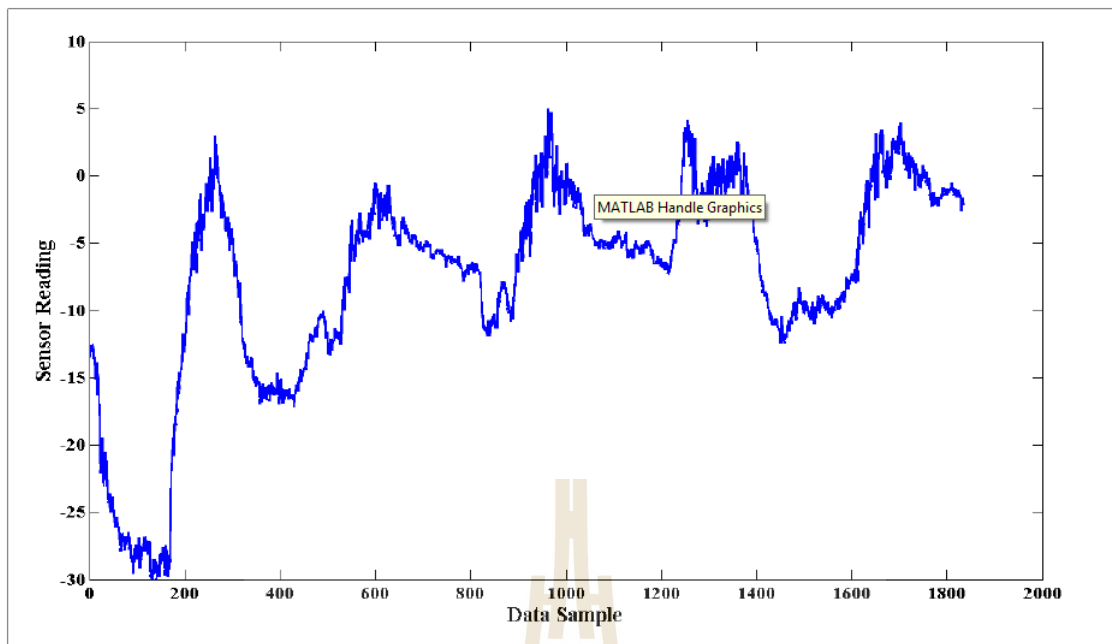
รูปที่ ก.22 ฮิสโตแกรมของชุดข้อมูล pdg2008 (ค่าอุณหภูมิพื้นผิวที่อ่านได้)



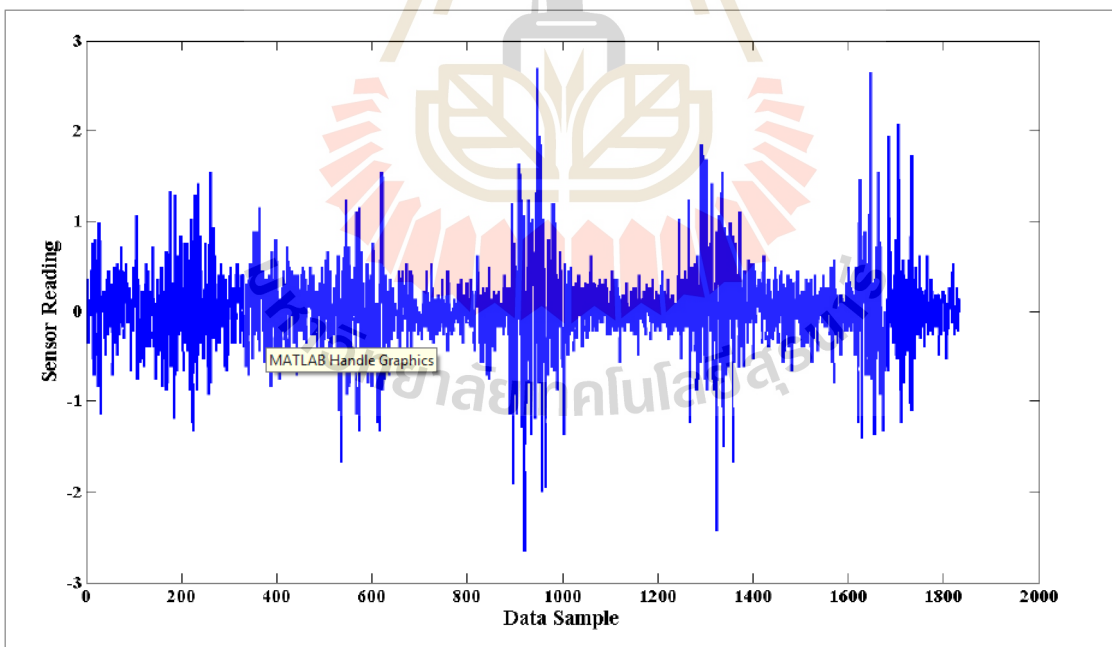
รูปที่ ก.23 ฮิสโตแกรมของชุดข้อมูล pdg2008 (ค่าอุณหภูมิโดยรอบที่อ่านได้)



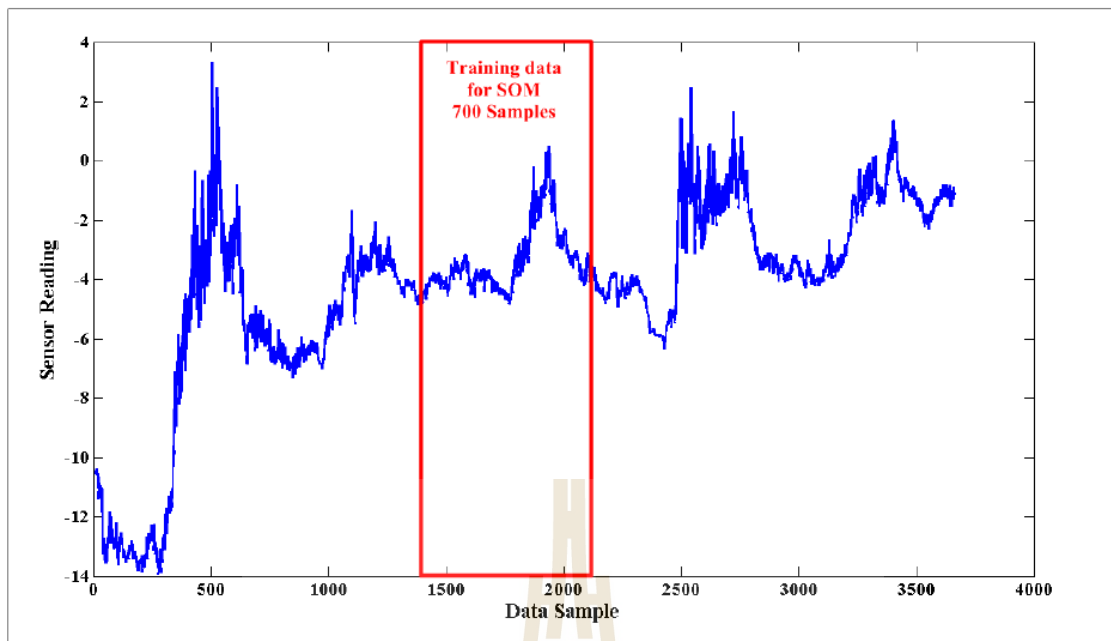
รูปที่ ก.24 ชุดข้อมูล pdg2008 (ค่าอุณหภูมิพื้นผิวที่อ่านได้)



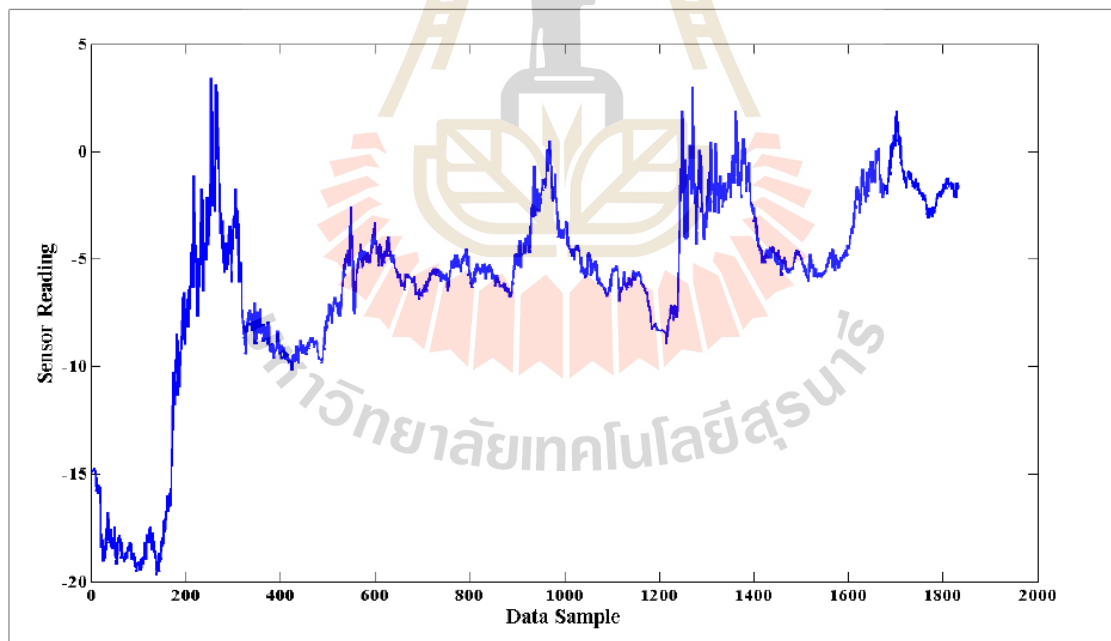
รูปที่ ก.25 DWT ค่าสัมประสิทธิ์ต่ำผ่านของชุดข้อมูล pdg2008 (ค่าอนุกรมมีพื้นผิวที่อ่านได้)



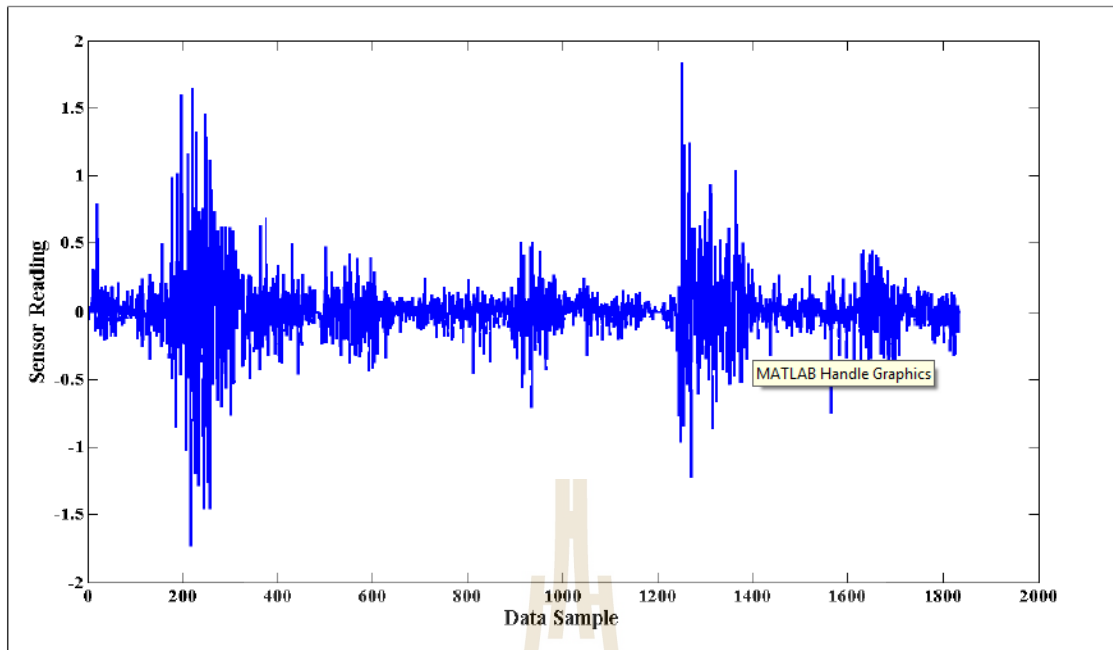
รูปที่ ก.26 DWT ค่าสัมประสิทธิ์สูงผ่านของชุดข้อมูล pdg2008 (ค่าอนุกรมมีพื้นผิวที่อ่านได้)



รูปที่ ก.27 ชุดข้อมูล pdg2008 (ค่าอนุหภูมิโดยรอบที่อ่านได้)

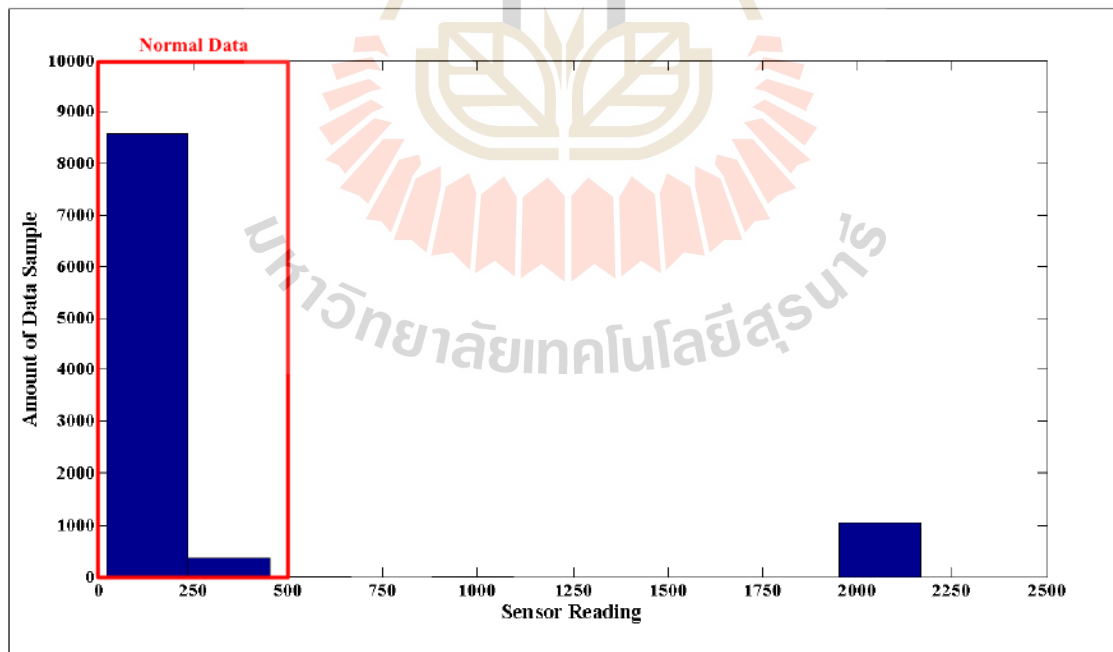


รูปที่ ก.25 DWT ค่าสัมประสิทธิ์ต่ำผ่านของชุดข้อมูล pdg2008 (ค่าอนุหภูมิโดยรอบที่อ่านได้)

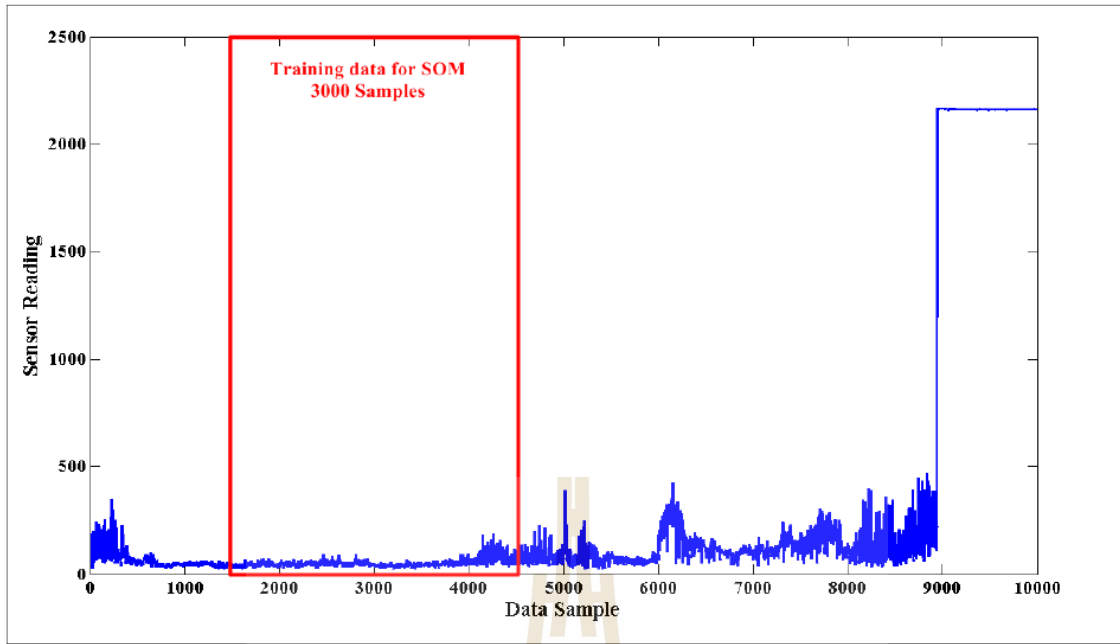


รูปที่ ก.26 DWT ค่าสัมประสิทธิ์สูงผ่านของชุดข้อมูล pdg2008 (ค่าอุณหภูมิโดยรอบที่อ่านได้)

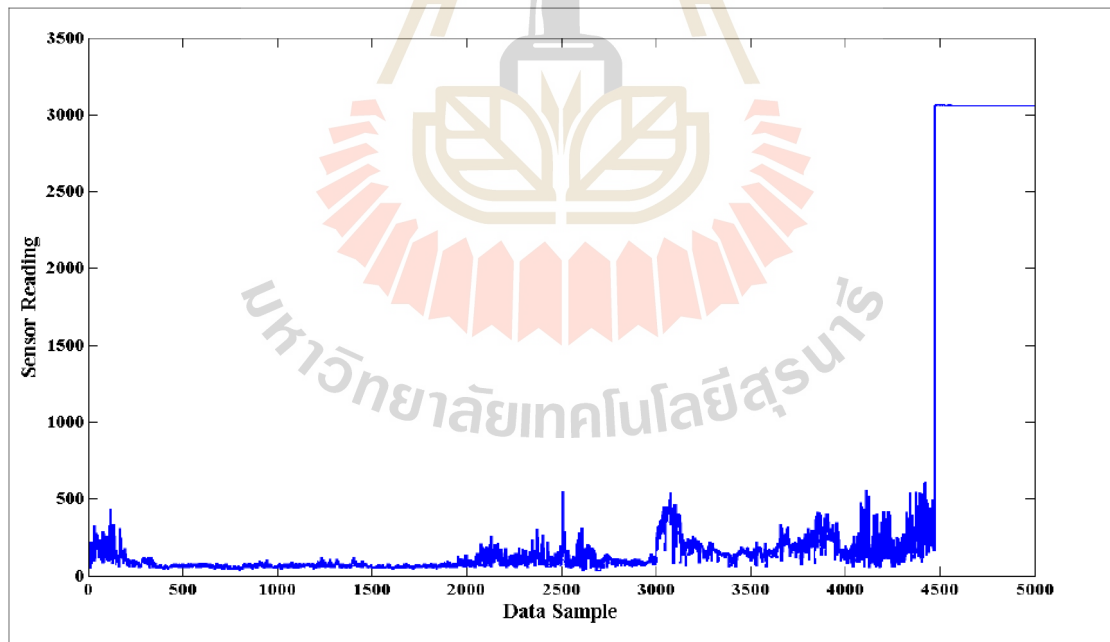
5. ชุดข้อมูล NAMOS



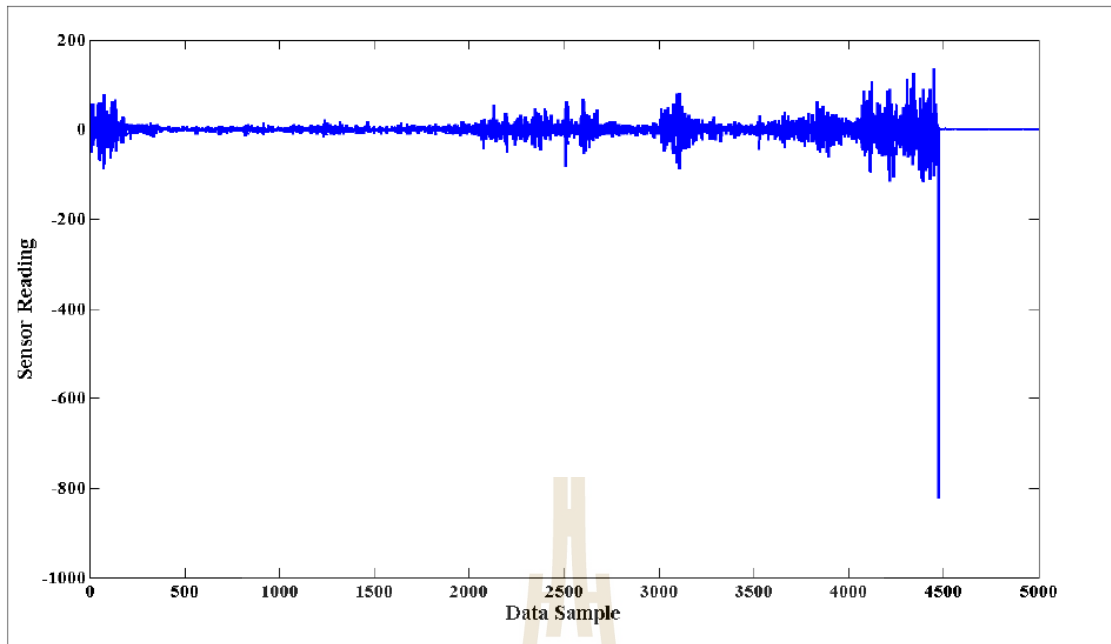
รูปที่ ก.30 ฮิสโตแกรมของชุดข้อมูล NAMOS (ค่าที่อ่านได้จากฟลูออริมิเตอร์)



รูปที่ ก.31 ชุดข้อมูล NAMOS (ค่าที่อ่านได้จากฟลูออริมิเตอร์)



รูปที่ ก.32 DWT ค่าสัมประสิทธิ์ต่ำผ่านของชุดข้อมูล NAMOS (ค่าที่อ่านได้จากฟลูออริมิเตอร์)



รูปที่ ก.33 DWT ค่าสัมประสิทธิ์สูงผ่านของชุดข้อมูล NAMOS (ค่าที่อ่านได้จากฟลูออริมิเตอร์)



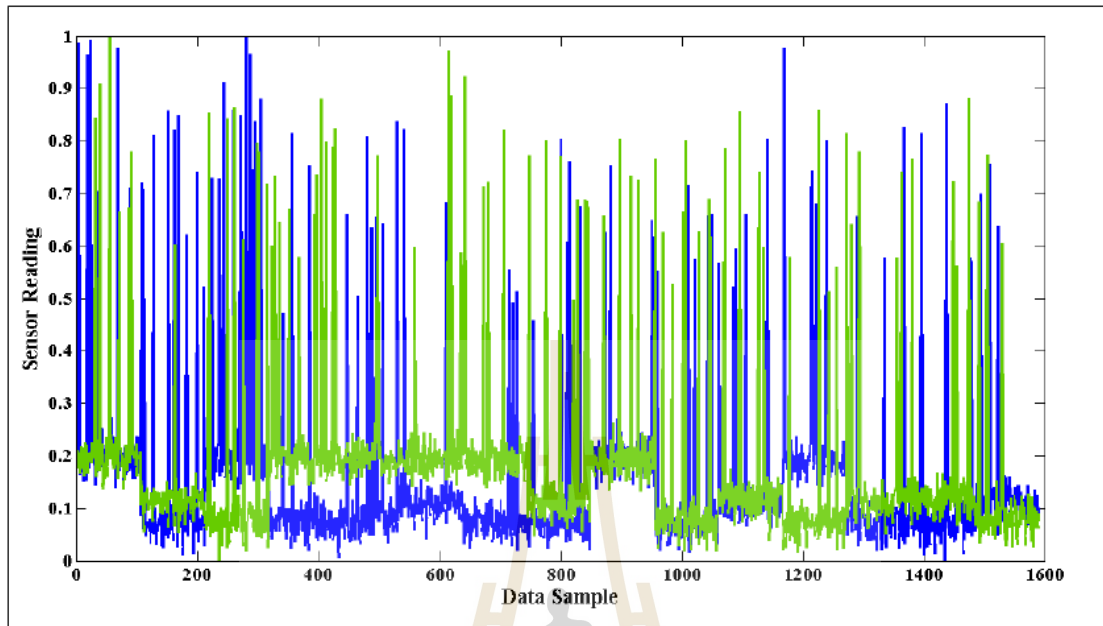
ภาคผนวก ข

ชุดข้อมูลสำหรับการทดลองบทที่ 4

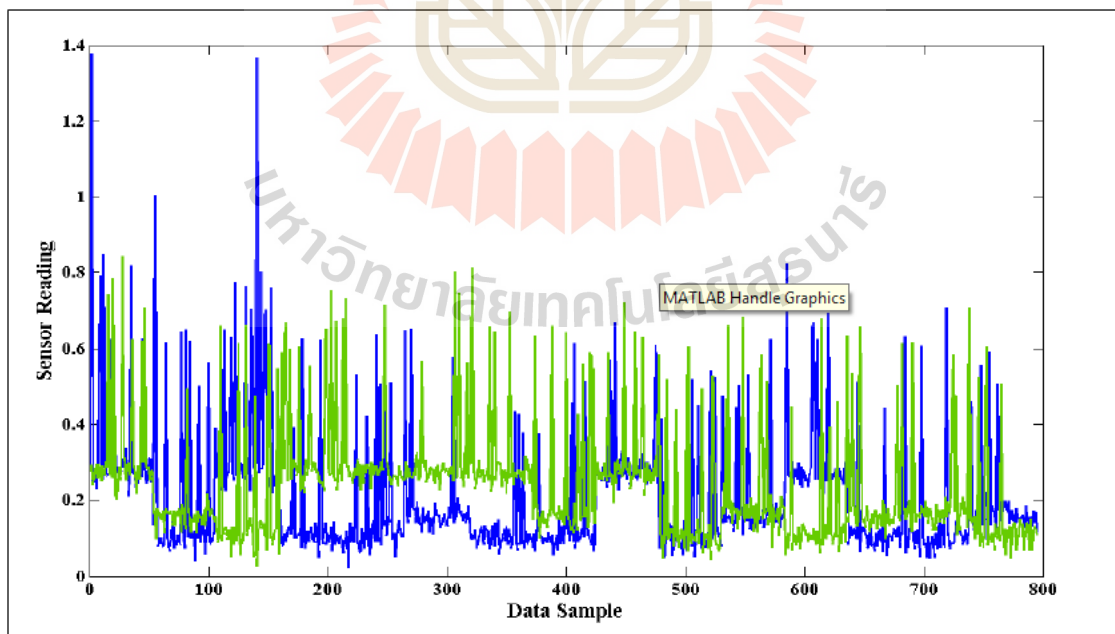


ชุดข้อมูลสำหรับบทที่ 4

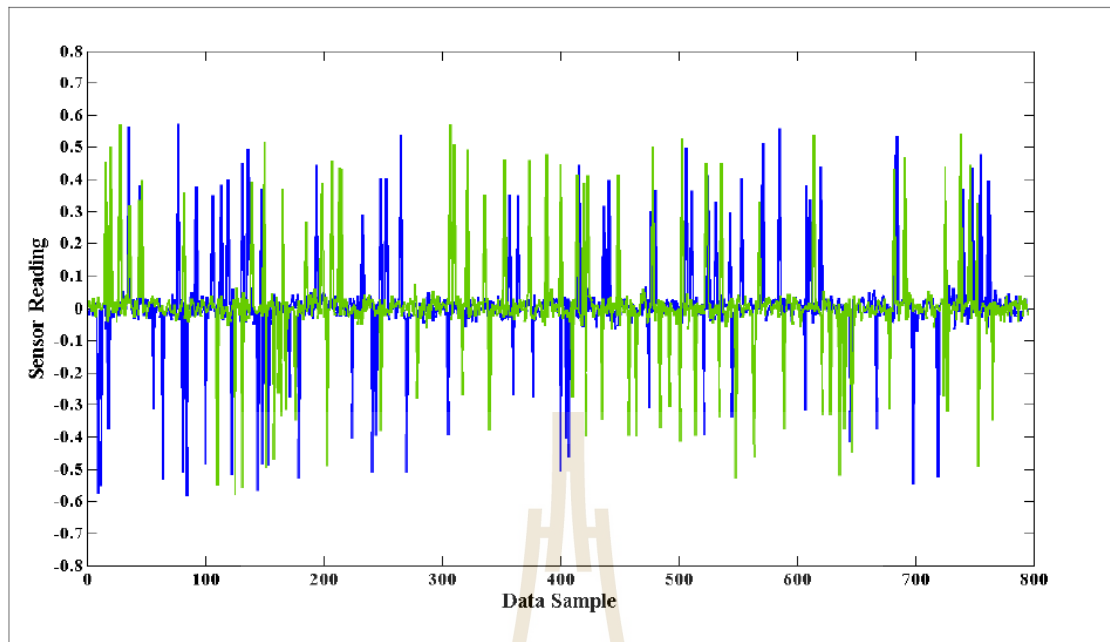
1. ข้อมูลสังเคราะห์



รูปที่ ข.1 ข้อมูลสังเคราะห์ 2KPI ที่ประกอบด้วยความผิดพลาด 1/80

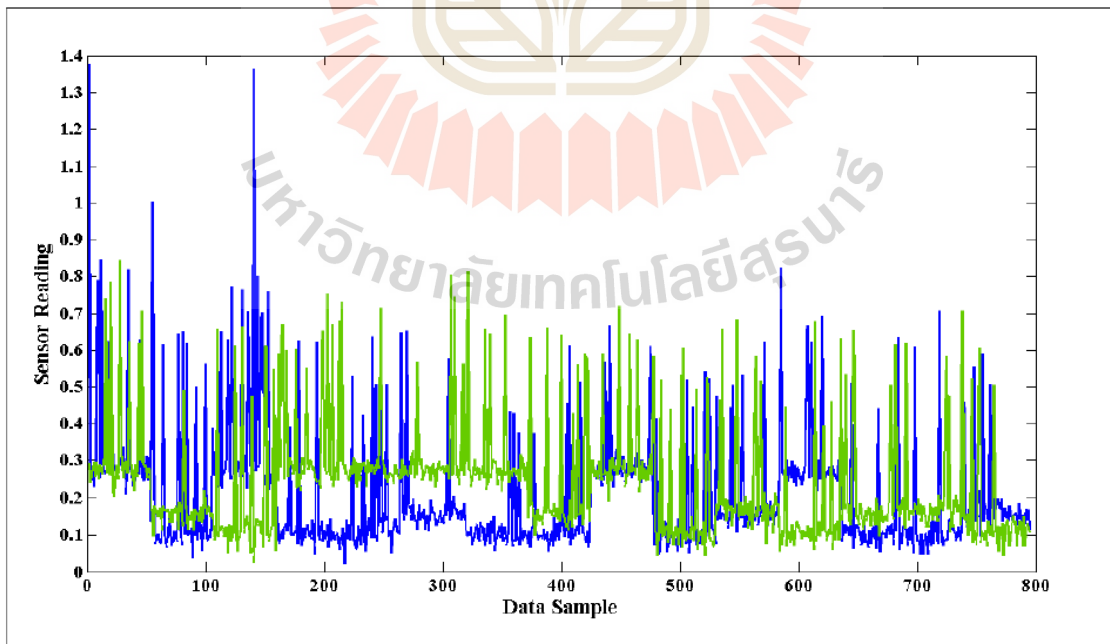


รูปที่ ข.2 DWT ค่าสัมประสิทธิ์ต่ำผ่านของชุดข้อมูลสังเคราะห์ 2KPI ที่ประกอบด้วยความผิดพลาด



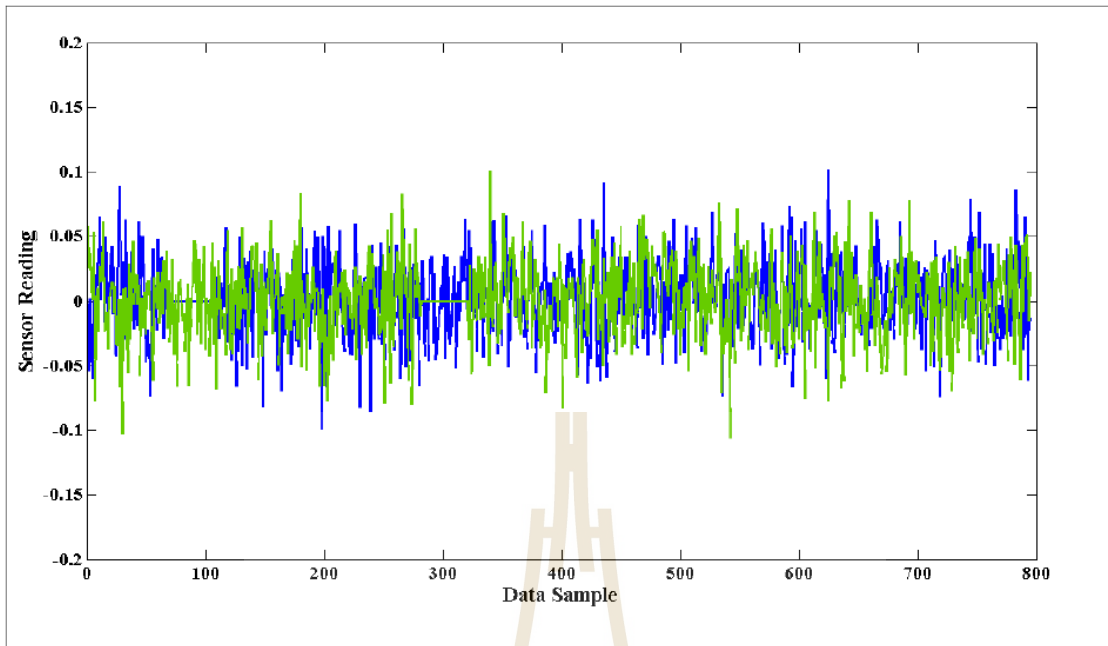
รูปที่ ข.3 DWT ค่าสัมประสิทธิ์สูงผ่านของชุดข้อมูลสังเคราะห์ 2KPI ที่ประกอบด้วยความผิดพลาด

1/80



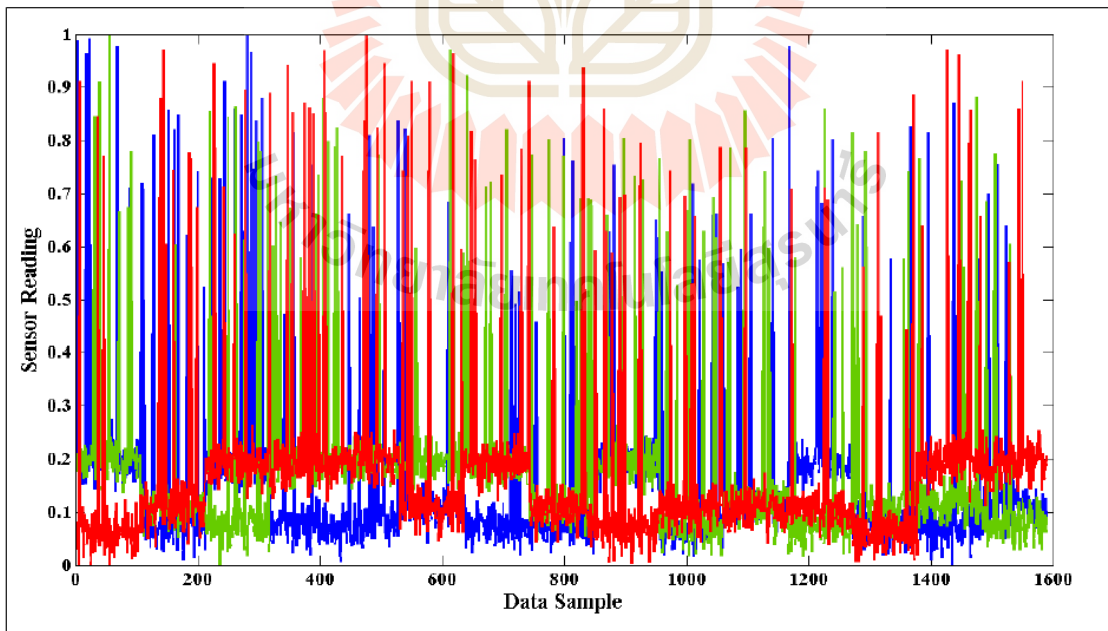
รูปที่ ข.4 LWT ค่าสัมประสิทธิ์ต่ำผ่านของชุดข้อมูลสังเคราะห์ 2KPI ที่ประกอบด้วยความผิดพลาด

1/80

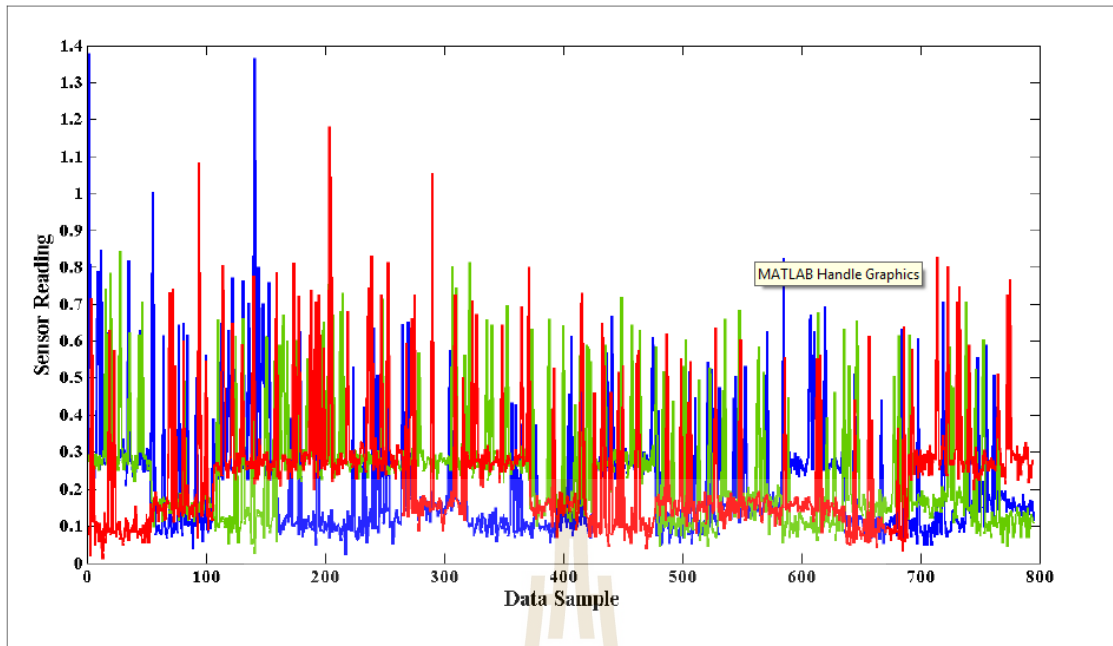


รูปที่ ข.5 LWT ค่าสัมประสิทธิ์สูงผ่านของชุดข้อมูลสังเคราะห์ 2KPI ที่ประกอบด้วยความผิดพลาด

1/80

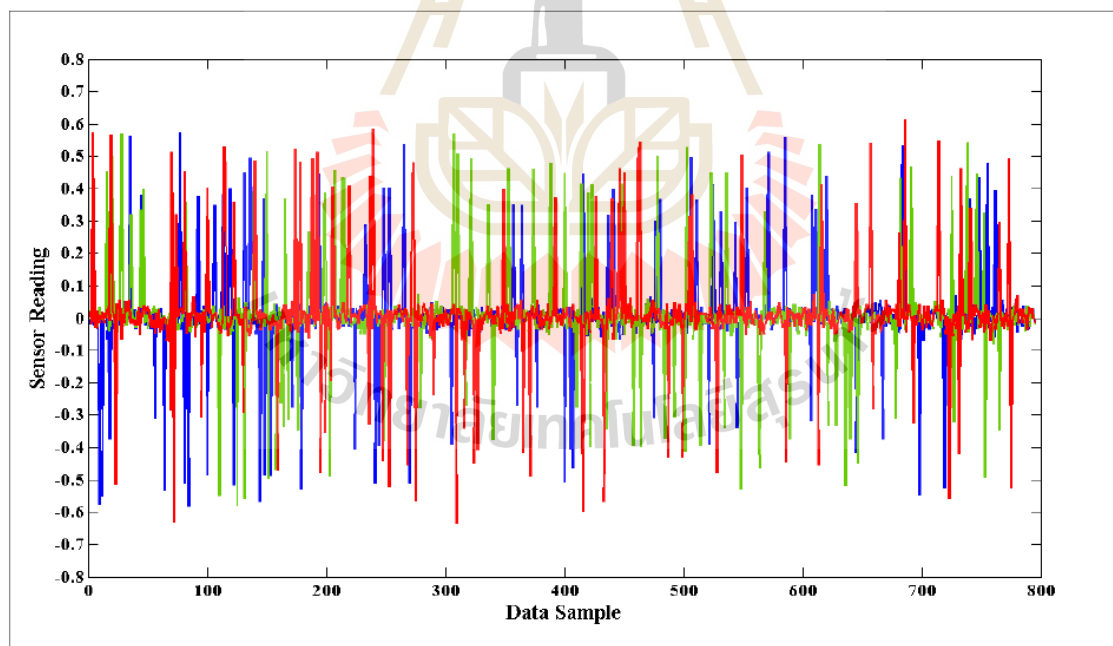


รูปที่ ข.6 ข้อมูลสังเคราะห์ 3KPI ที่ประกอบด้วยความผิดพลาด 1/80



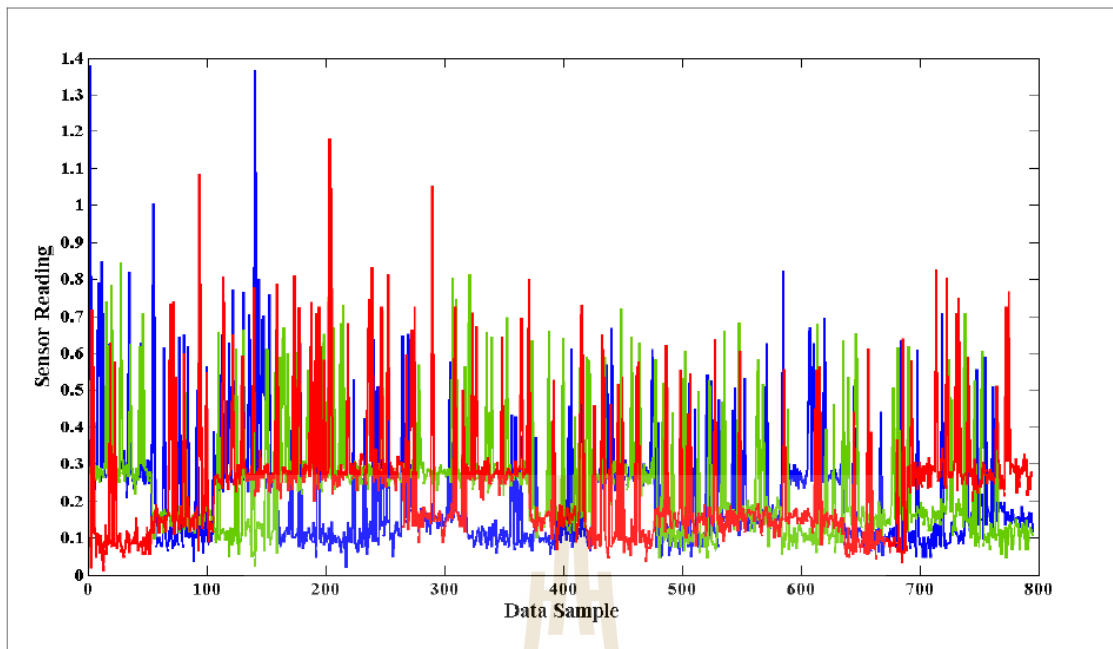
รูปที่ ข.7 DWT ค่าสัมประสิทธิ์ต่ำผ่านของชุดข้อมูลสังเคราะห์ 3KPI ที่ประกอบด้วยความผิดพลาด

1/80



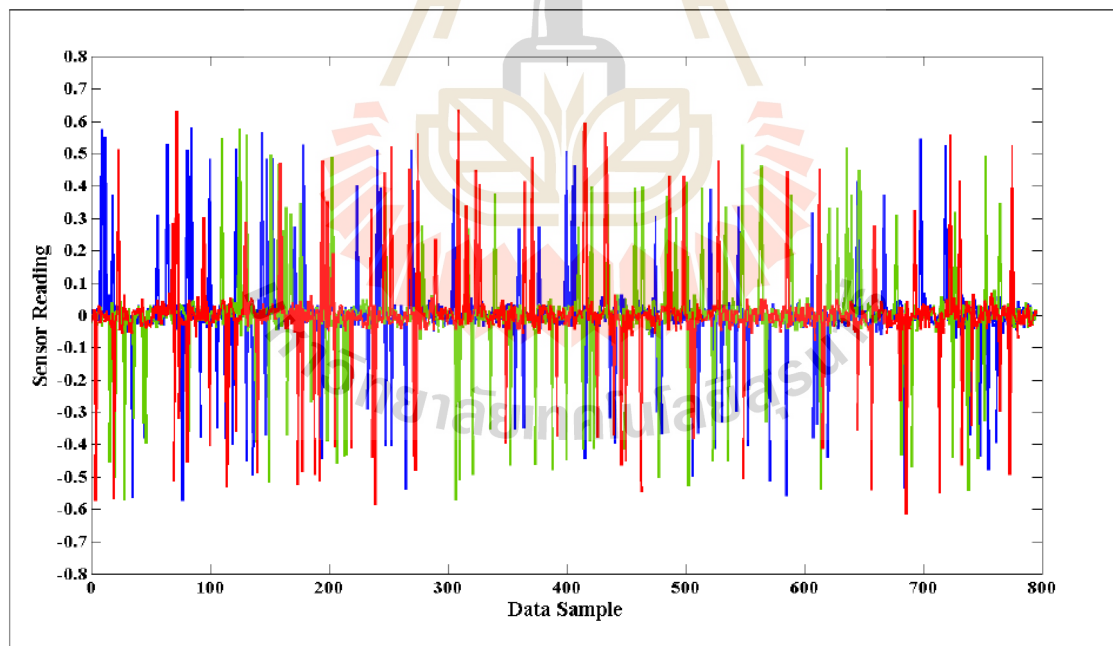
รูปที่ ข.8 DWT ค่าสัมประสิทธิ์สูงผ่านของชุดข้อมูลสังเคราะห์ 3KPI ที่ประกอบด้วยความผิดพลาด

1/80



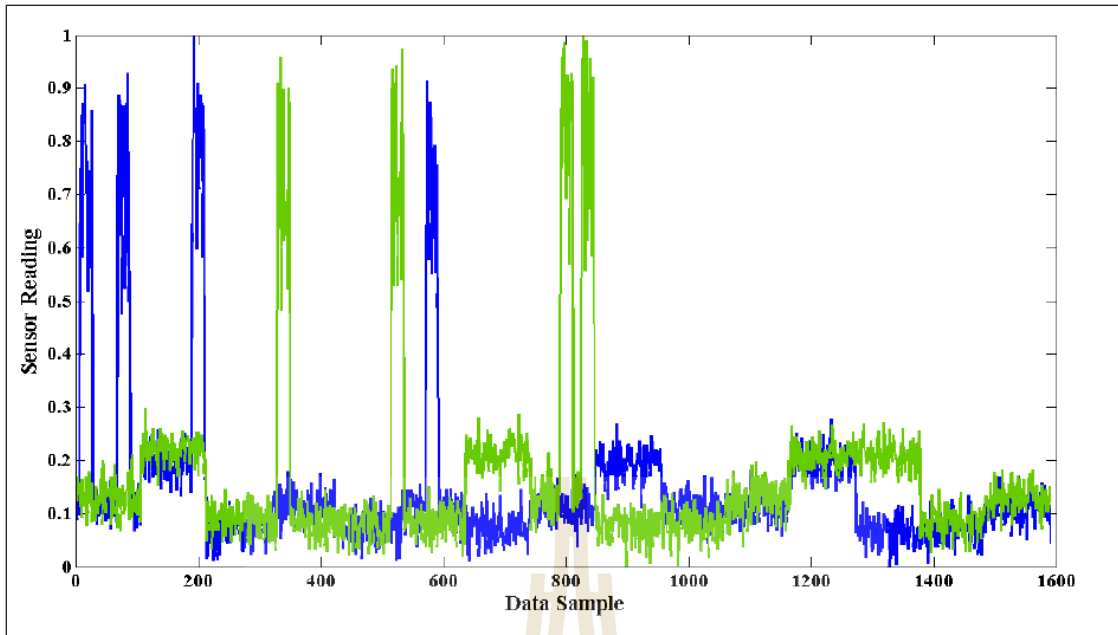
รูปที่ ข.9 LWT ค่าสัมประสิทธิ์ต่ำผ่านของชุดข้อมูลสังเคราะห์ 3KPI ที่ประกอบด้วยความผิดพลาด

1/80

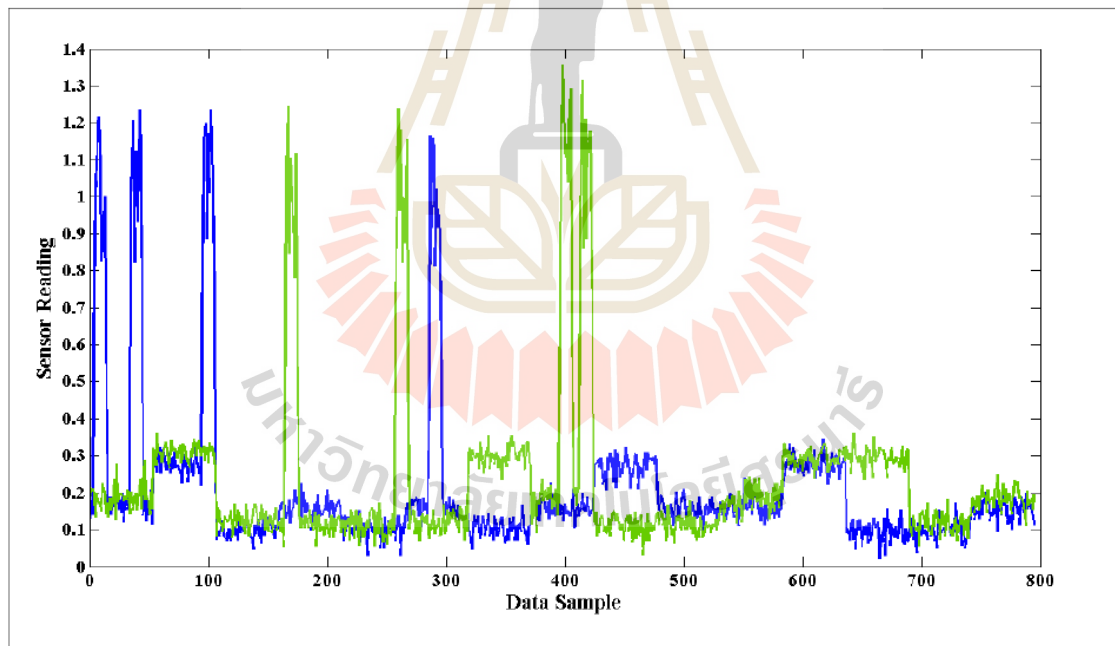


รูปที่ ข.10 LWT ค่าสัมประสิทธิ์สูงผ่านของชุดข้อมูลสังเคราะห์ 3KPI ที่ประกอบด้วยความผิดพลาด

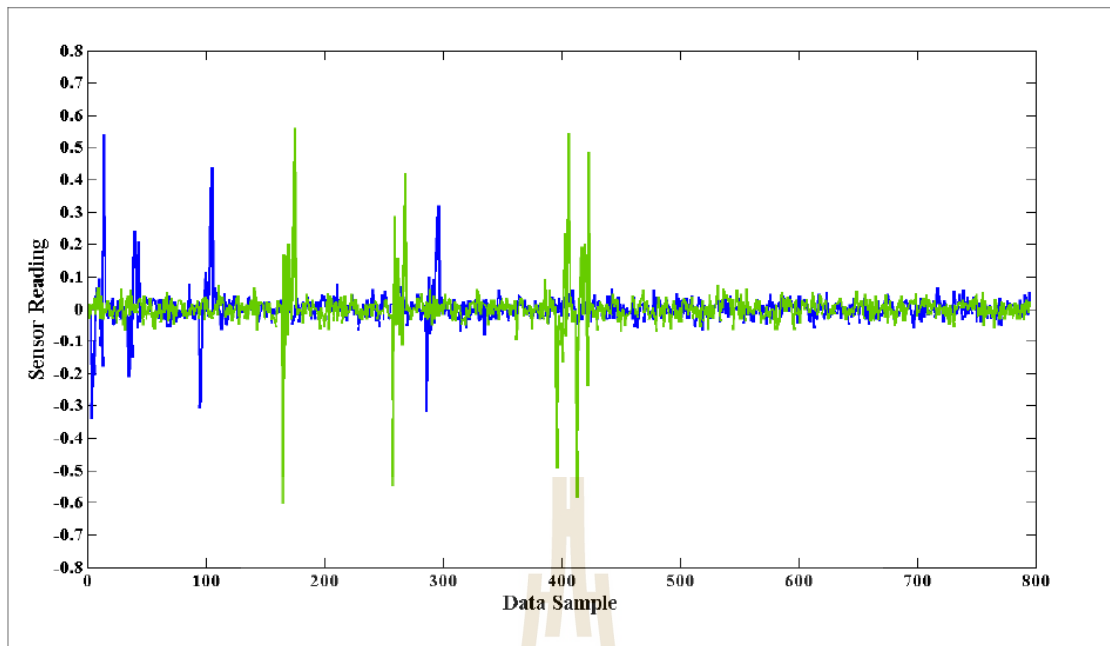
1/80



รูปที่ ข.11 ข้อมูลสังเคราะห์ 2KPI ที่ประกอบด้วยความผิดพลาด 20/4

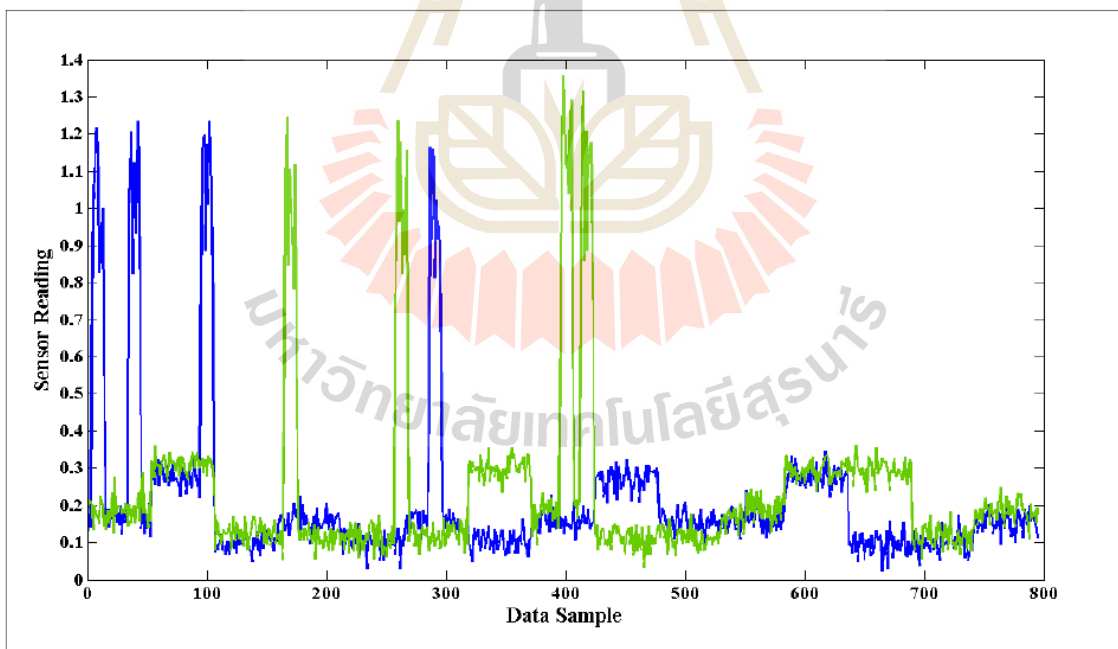


รูปที่ ข.12 DWT ค่าสัมประสิทธิ์ต่ำผ่านของชุดข้อมูลสังเคราะห์ 2KPI ที่ประกอบด้วยความผิดพลาด



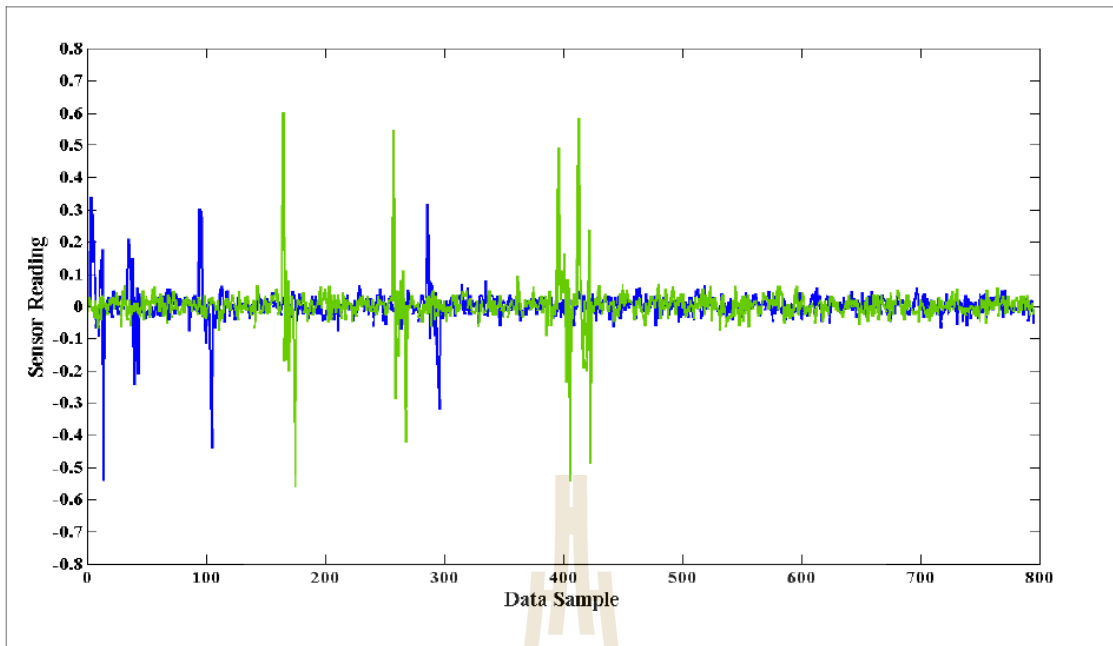
รูปที่ ข.13 DWT ค่าสัมประสิทธิ์สูงผ่านของชุดข้อมูลสังเคราะห์ 2KPI ที่ประกอบด้วยความผิดพลาด

20/4



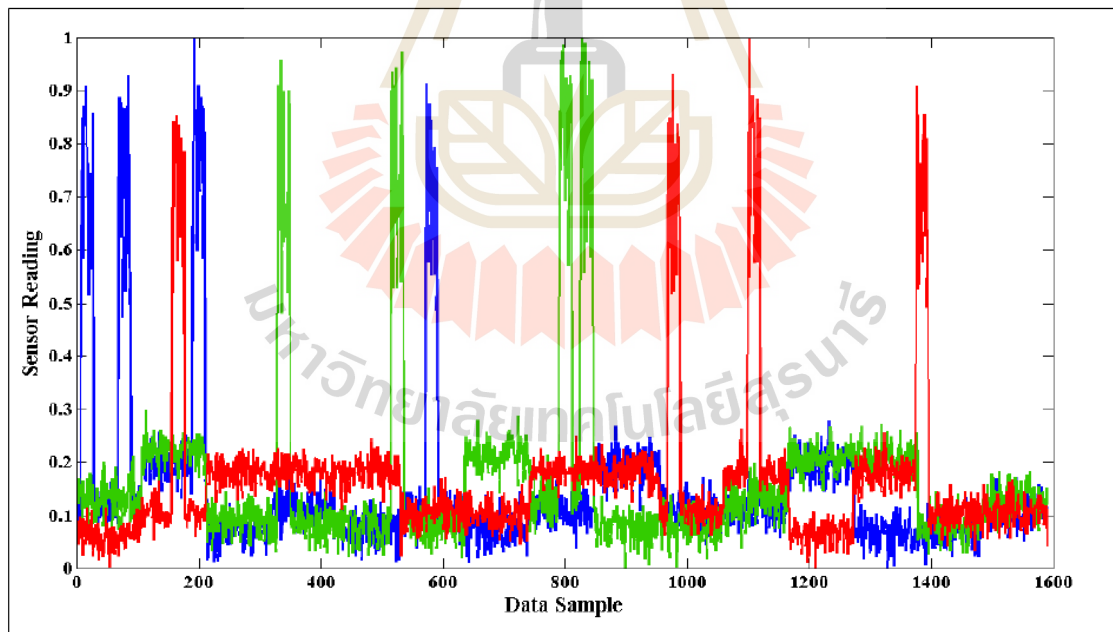
รูปที่ ข.14 LWT ค่าสัมประสิทธิ์ต่ำผ่านของชุดข้อมูลสังเคราะห์ 2KPI ที่ประกอบด้วยความผิดพลาด

20/4

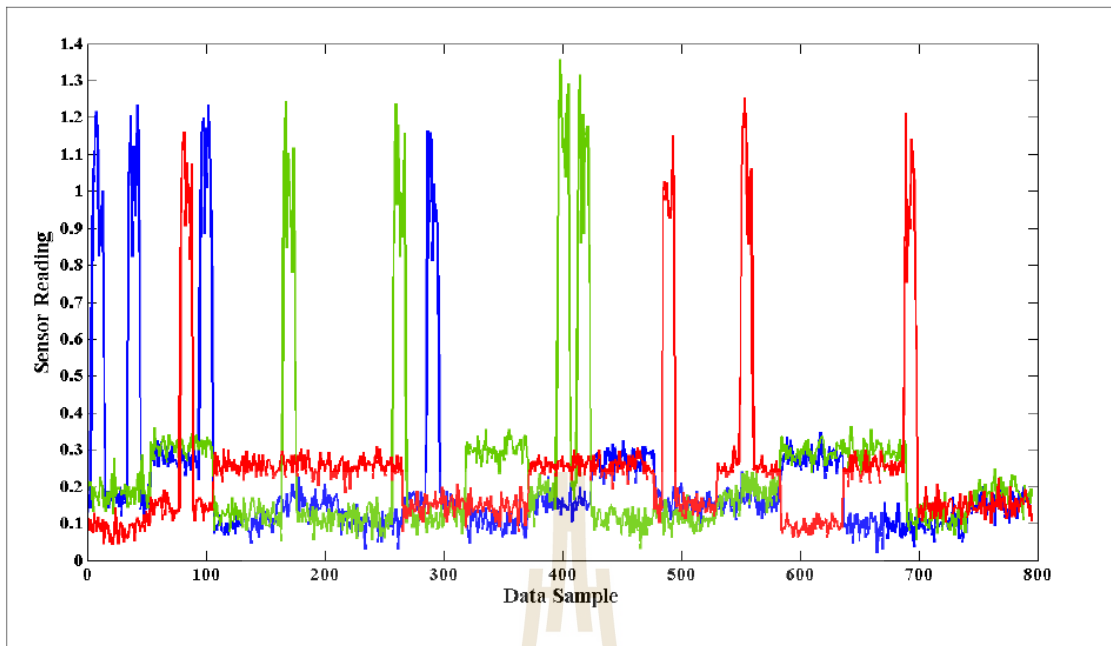


รูปที่ ข.15 LWT ค่าสัมประสิทธิ์ที่สูงผ่านของชุดข้อมูลสังเคราะห์ 2KPI ที่ประกอบด้วยความผิดพลาด

20/4

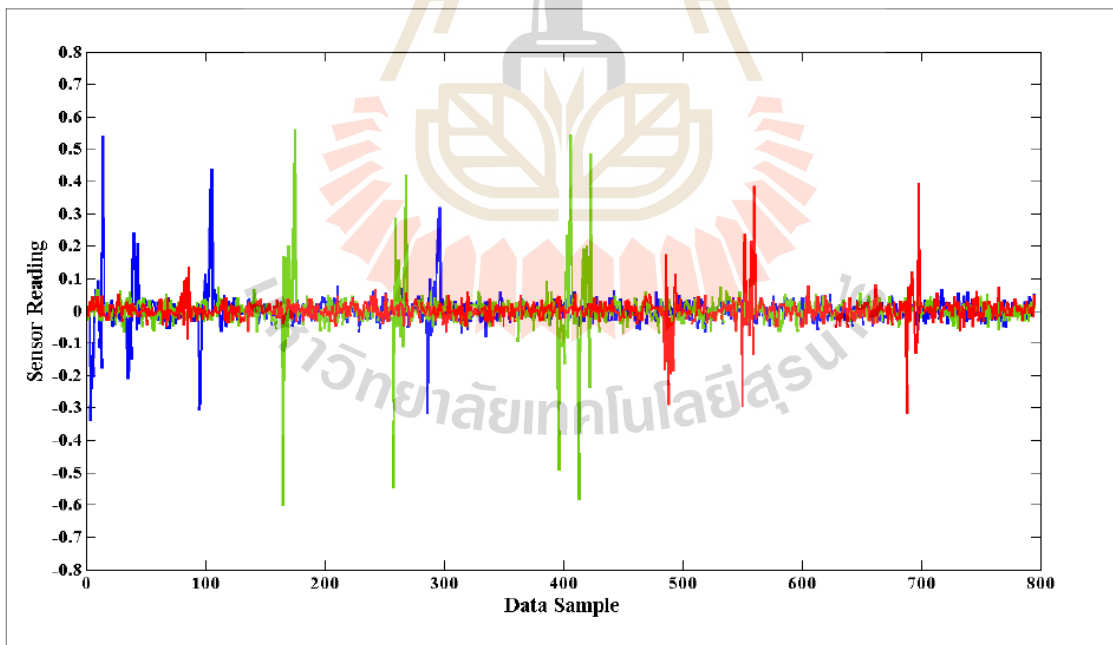


รูปที่ ข.16 ข้อมูลสังเคราะห์ 3KPI ที่ประกอบด้วยความผิดพลาด 20/4



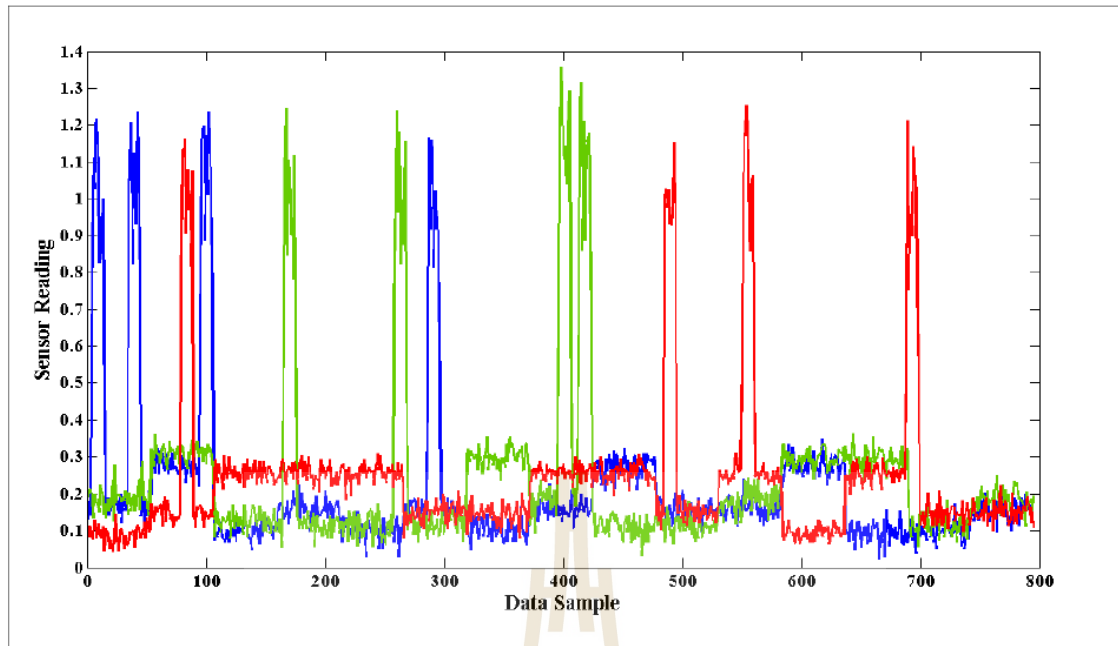
รูปที่ ข.17 DWT ค่าสัมประสิทธิ์ต่ำผ่านของชุดข้อมูลสังเคราะห์ 3KPI ที่ประกอบด้วยความผิดพลาด

20/4



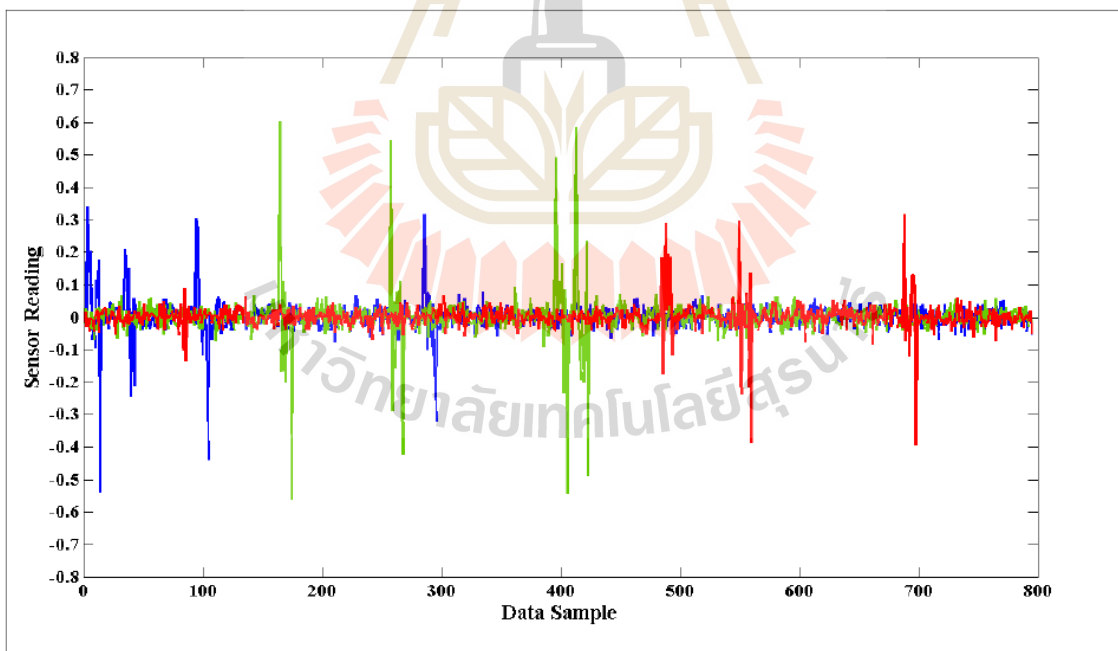
รูปที่ ข.18 DWT ค่าสัมประสิทธิ์สูงผ่านของชุดข้อมูลสังเคราะห์ 3KPI ที่ประกอบด้วยความผิดพลาด

20/4



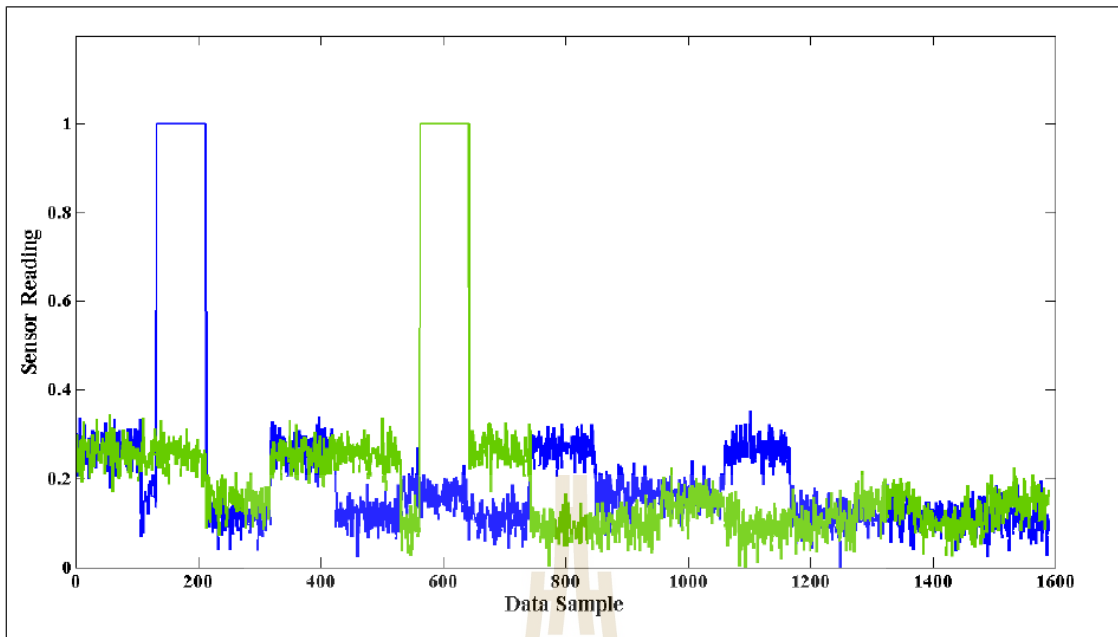
รูปที่ ข.19 LWT ค่าสัมประสิทธิ์ต่ำผ่านของชุดข้อมูลสังเคราะห์ 3KPI ที่ประกอบด้วยความผิดพลาด

20/4

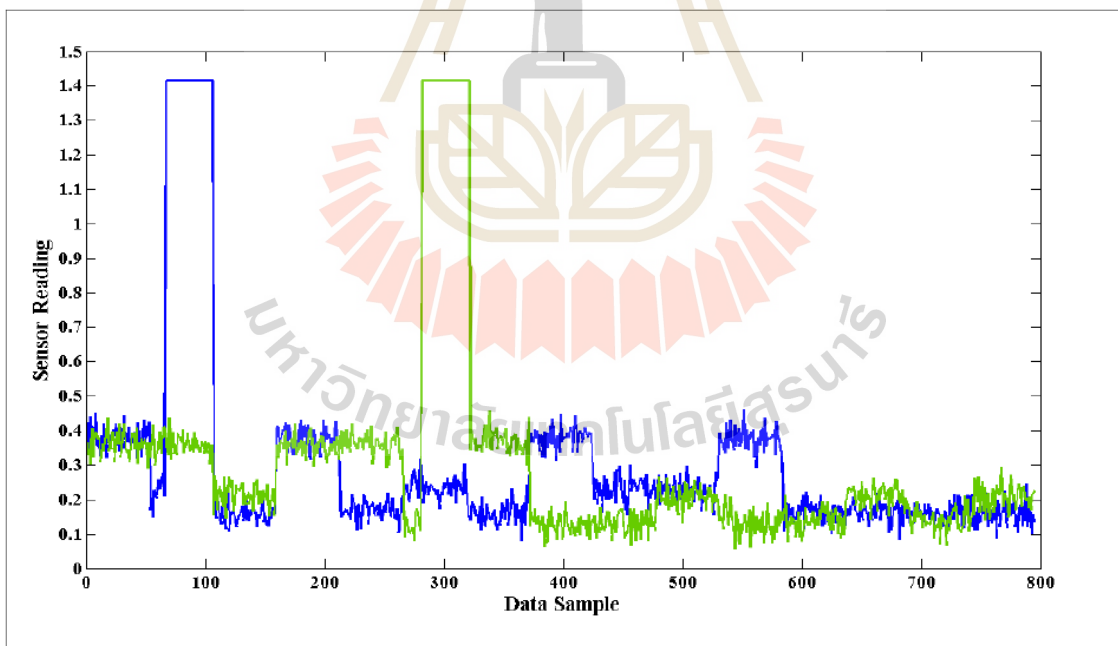


รูปที่ ข.20 LWT ค่าสัมประสิทธิ์สูงผ่านของชุดข้อมูลสังเคราะห์ 3KPI ที่ประกอบด้วยความผิดพลาด

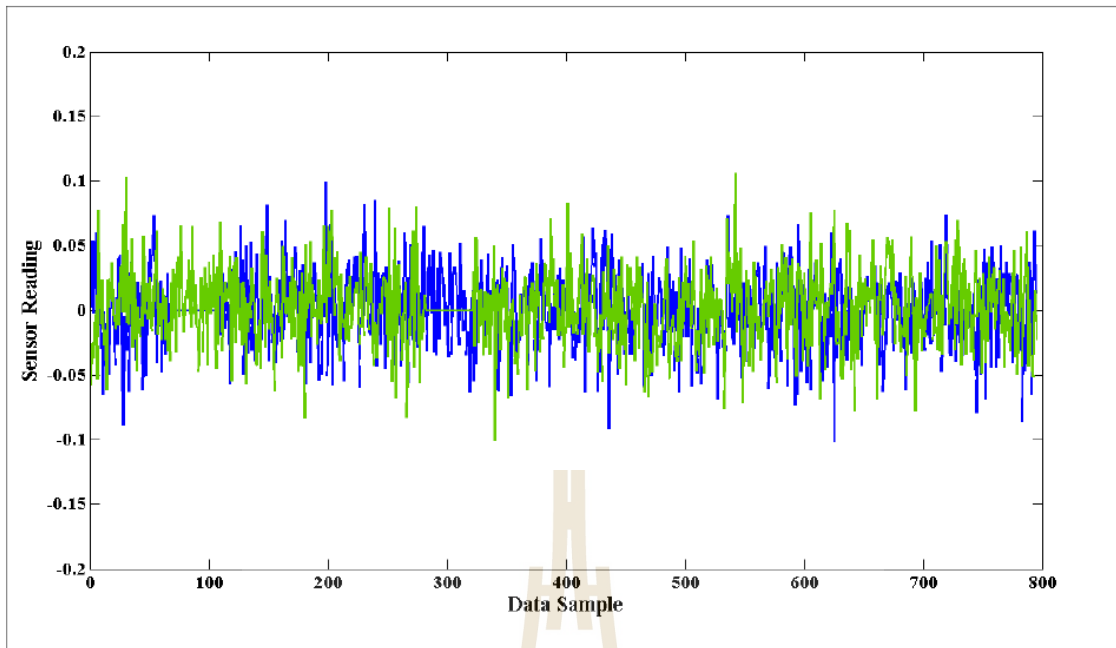
20/4



รูปที่ ข.21 ข้อมูลสังเคราะห์ 2KPI ที่ประกอบด้วยความผิดพลาด 80/1

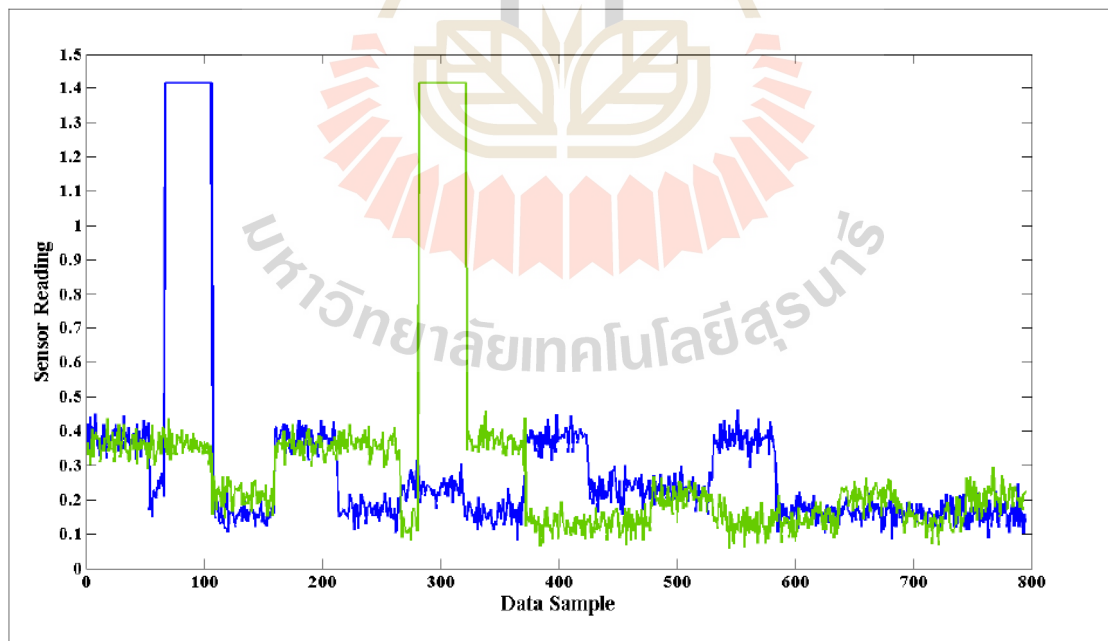


รูปที่ ข.22 DWT ค่าสัมประสิทธิ์ต่ำผ่านของข้อมูลสังเคราะห์ 2KPI ที่ประกอบด้วยความผิดพลาด



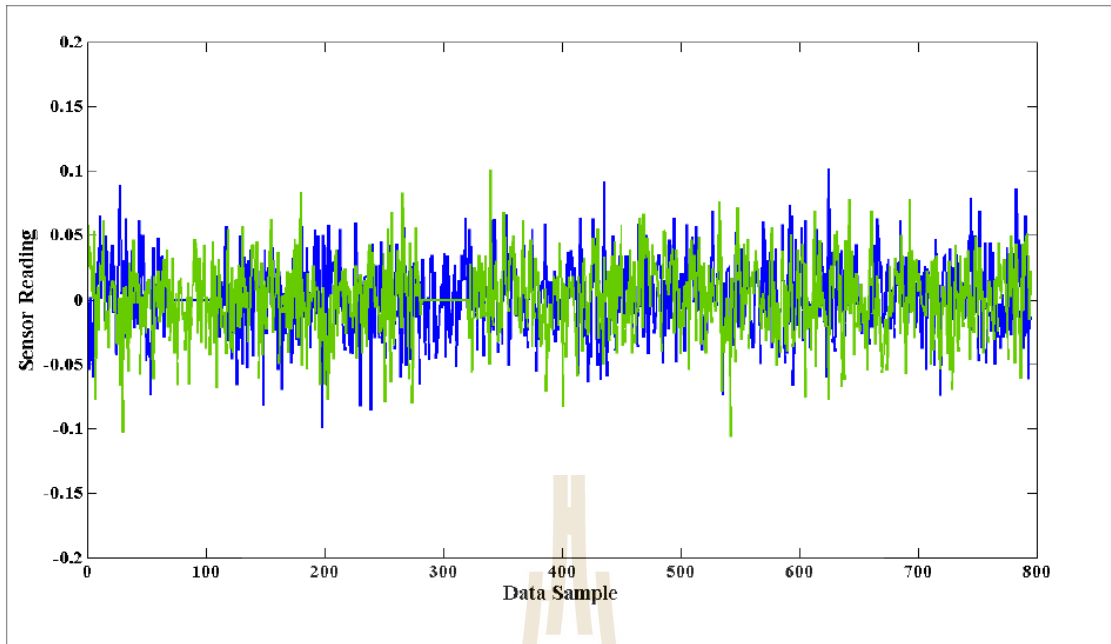
รูปที่ ข.23 DWT ค่าสัมประสิทธิ์สูงผ่านของข้อมูลสังเคราะห์ 2KPI ที่ประกอบด้วยความผิดพลาด

80/1



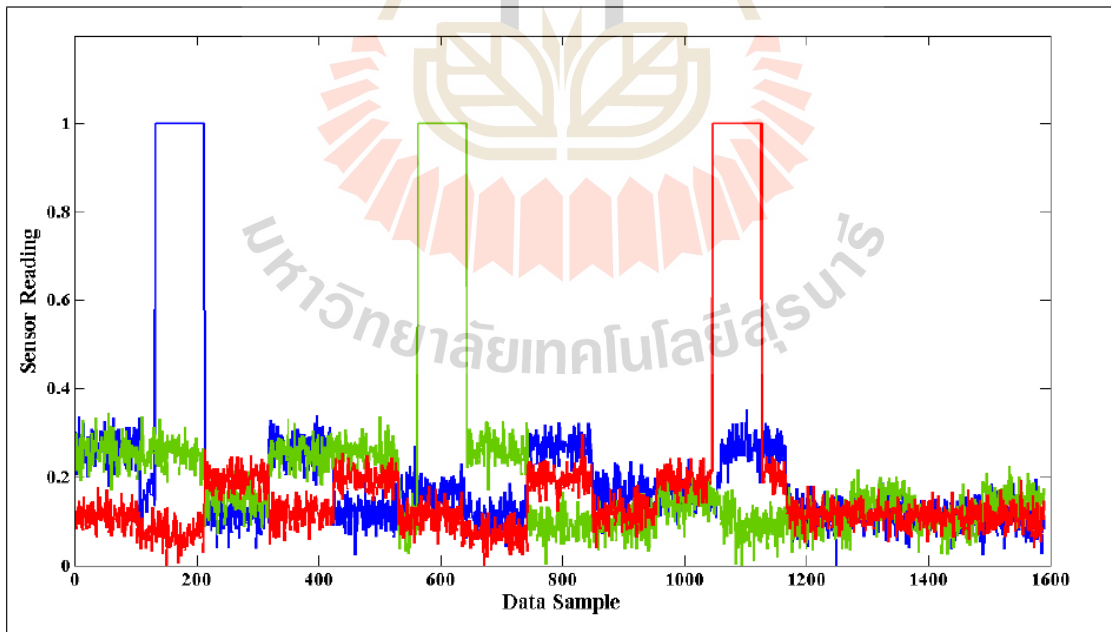
รูปที่ ข.24 LWT ค่าสัมประสิทธิ์ต่ำผ่านของข้อมูลสังเคราะห์ 2KPI ที่ประกอบด้วยความผิดพลาด

80/1

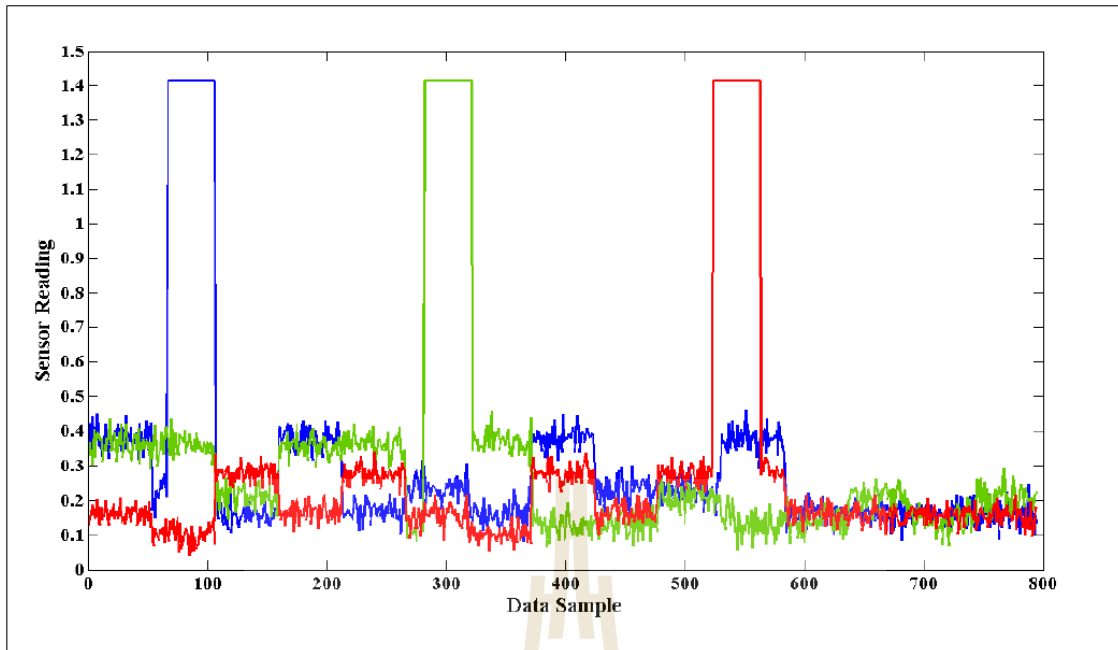


รูปที่ ข.25 LWT ค่าสัมประสิทธิ์สูงผ่านของข้อมูลสังเคราะห์ 2KPI ที่ประกอบด้วยความผิดพลาด

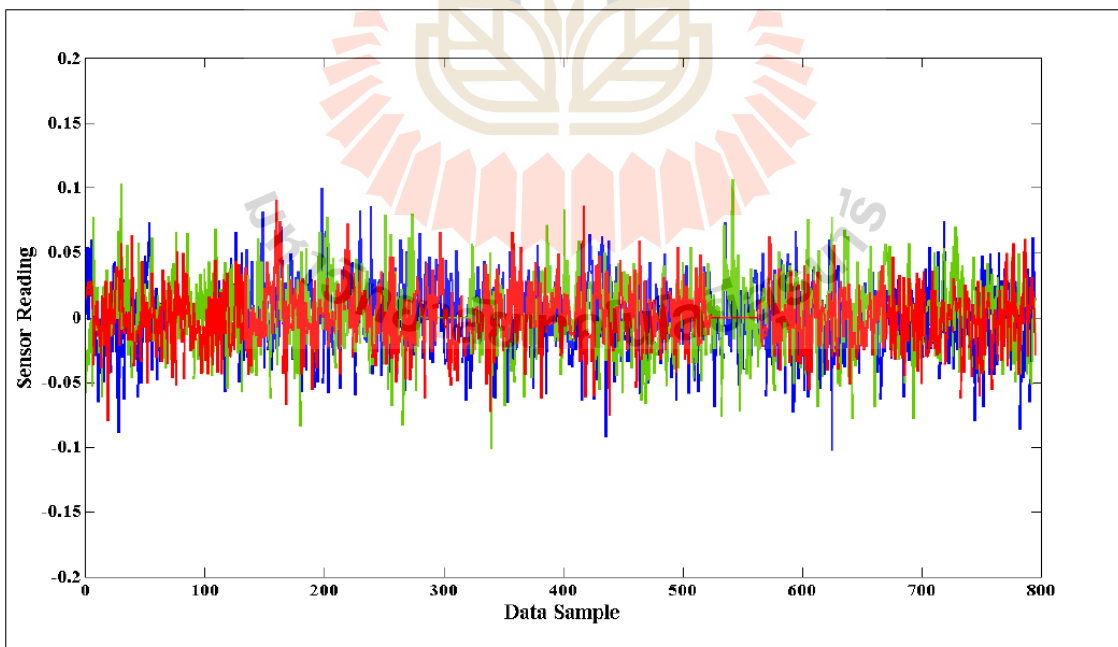
80/1



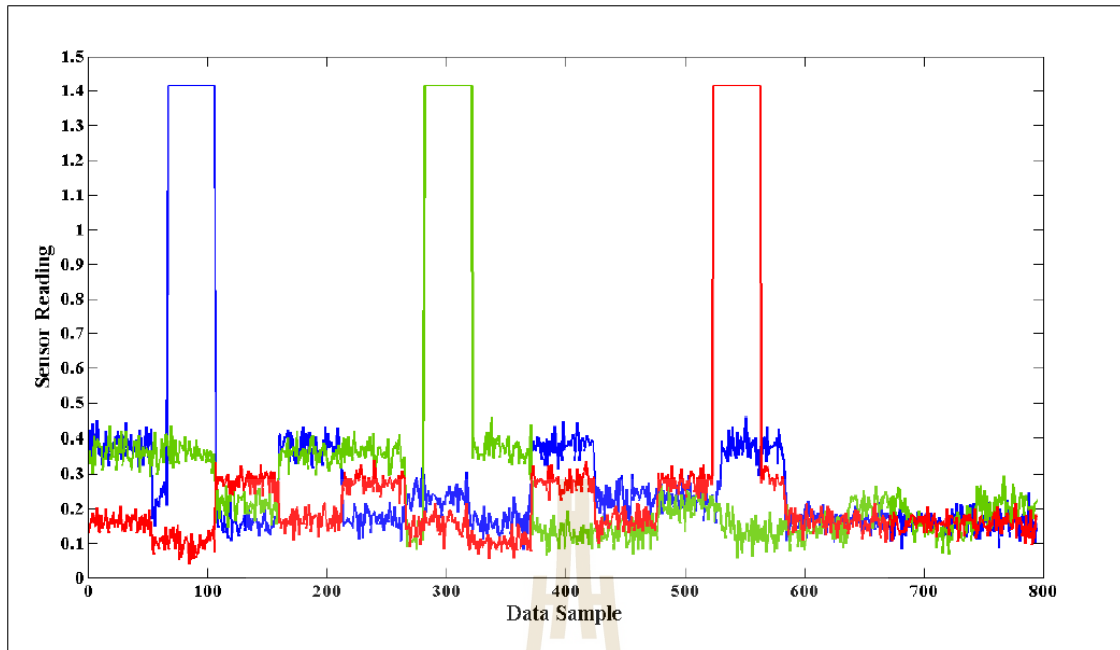
รูปที่ ข.26 ข้อมูลสังเคราะห์ 3KPI ที่ประกอบด้วยความผิดพลาด 80/1



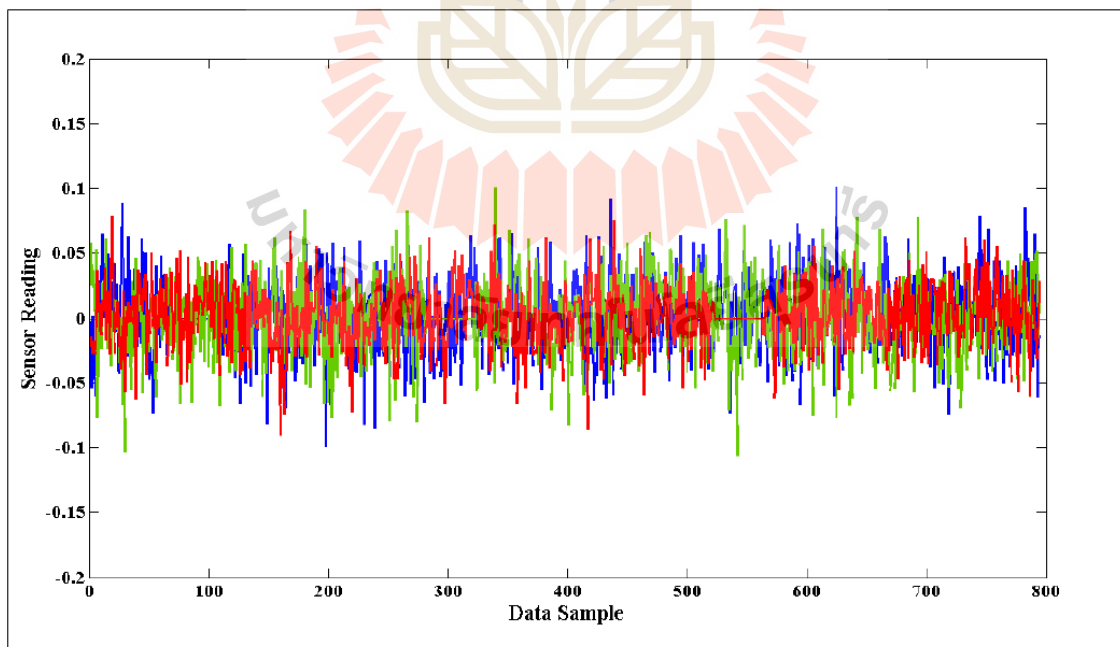
รูปที่ ข.27 DWT สัมประสิทธิ์ผ่านต่ำของข้อมูลสังเคราะห์ 3KPI กับความผิดพลาด 80/1



รูปที่ ข.28 DWT สัมประสิทธิ์ผ่านสูงของข้อมูลสังเคราะห์ 3KPI กับความผิดพลาด 80/1

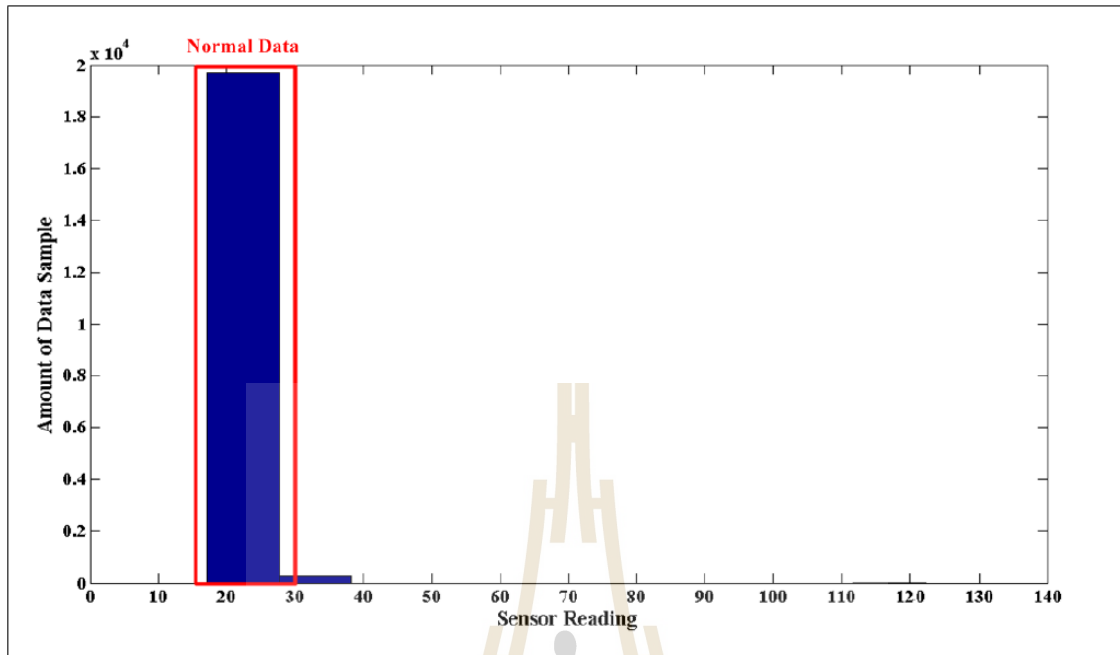


รูปที่ ข.29 LWT สัมประสิทธิ์ผ่านต่ำของข้อมูลสังเคราะห์ 3KPI กับความผิดพลาด 80/1

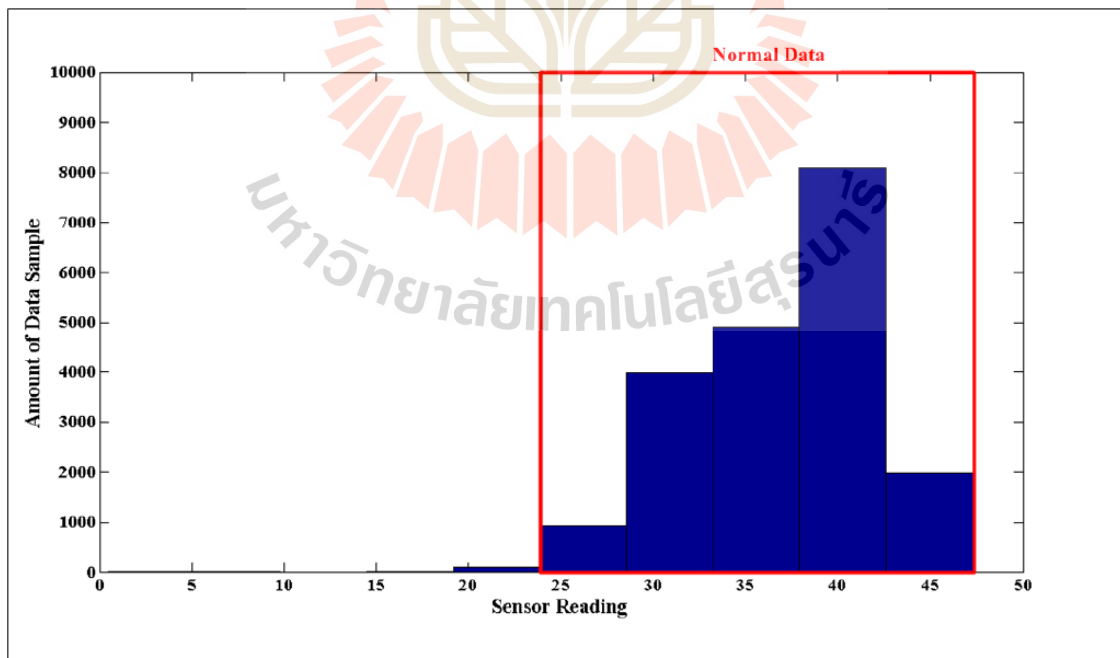


รูปที่ ข.30 LWT สัมประสิทธิ์ผ่านสูงของข้อมูลสังเคราะห์ 3KPI กับความผิดพลาด 80/1

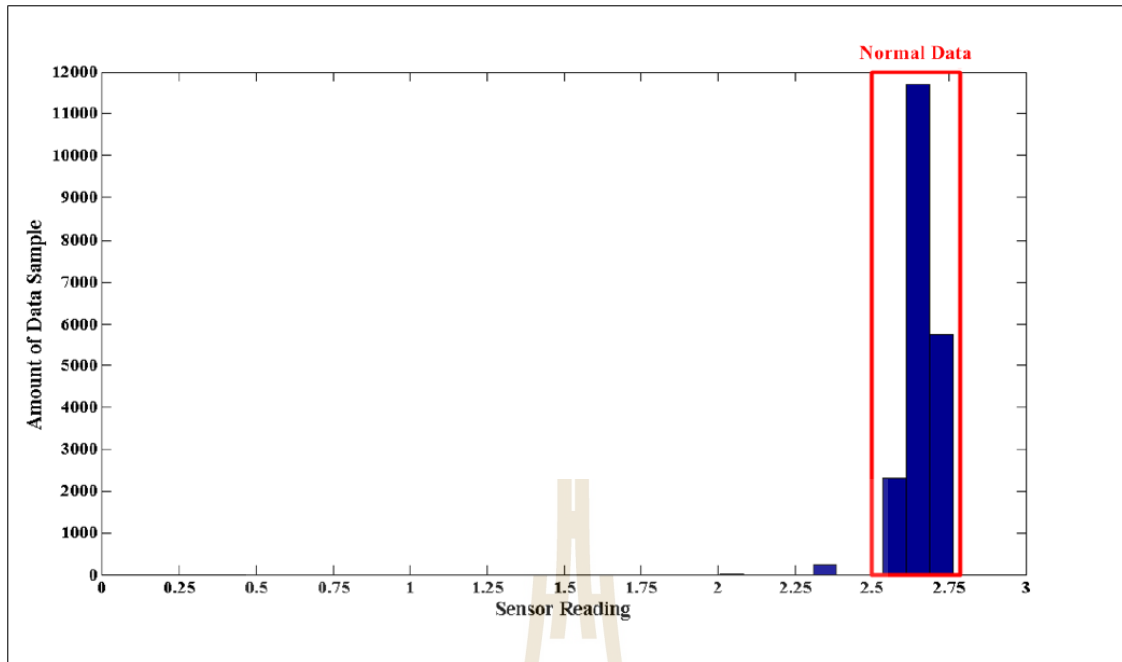
2. ชุดข้อมูล INTEL



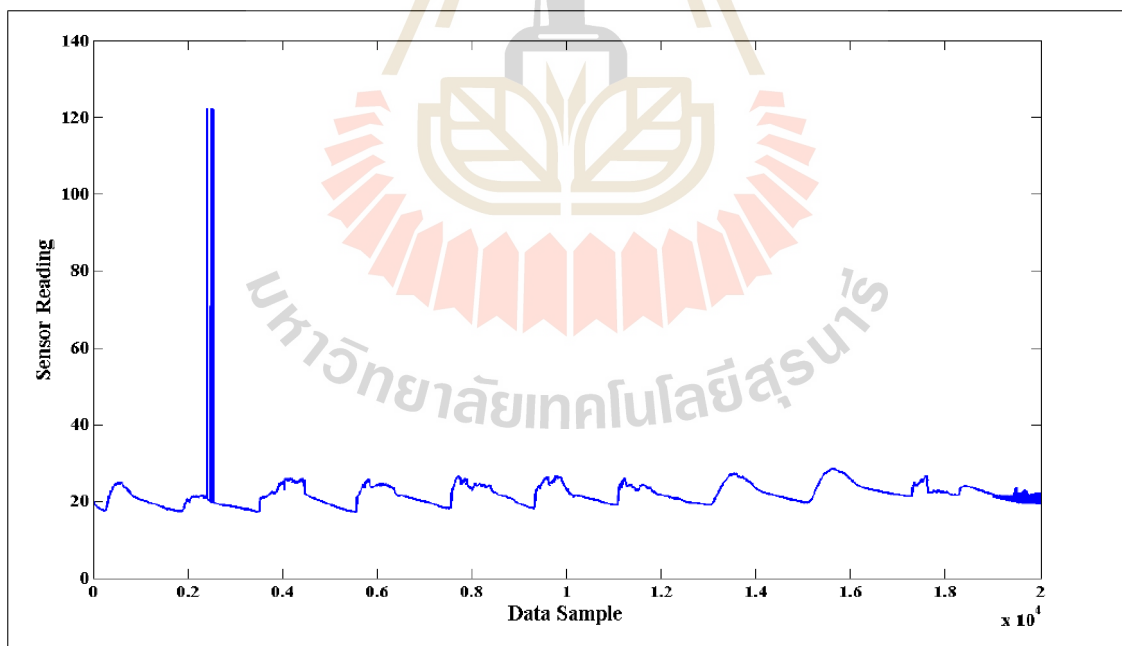
รูปที่ ข.31 ฮิตโตแกรมแสดงค่าของชุดข้อมูล INTEL (อุณหภูมิที่อ่านได้)



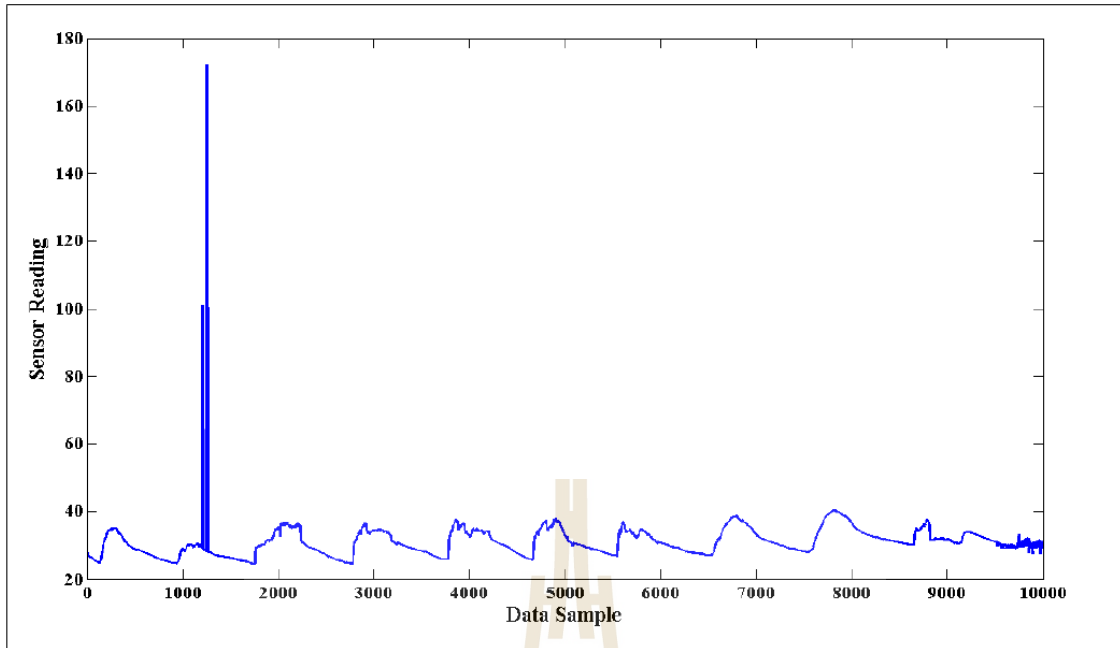
รูปที่ ข.32 ฮิตโตแกรมแสดงค่าของชุดข้อมูล INTEL (ความชื้นที่อ่านได้)



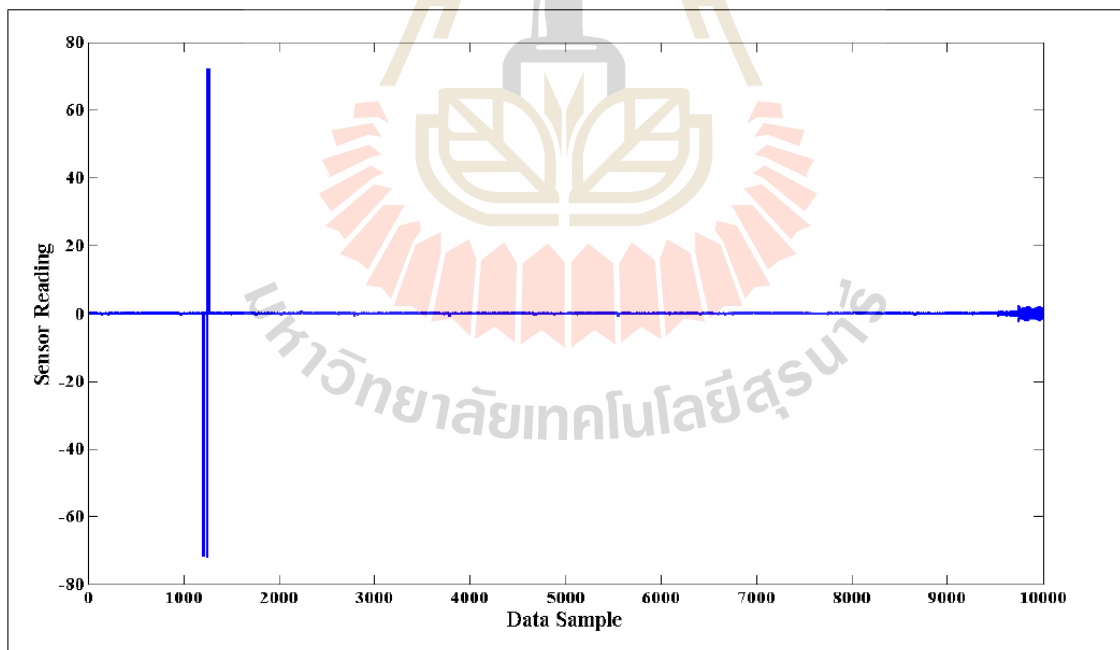
รูปที่ ข.33 ฮิตโตแกรมแสดงค่าของชุดข้อมูล INTEL (แรงดันไฟฟ้าที่อ่านได้)



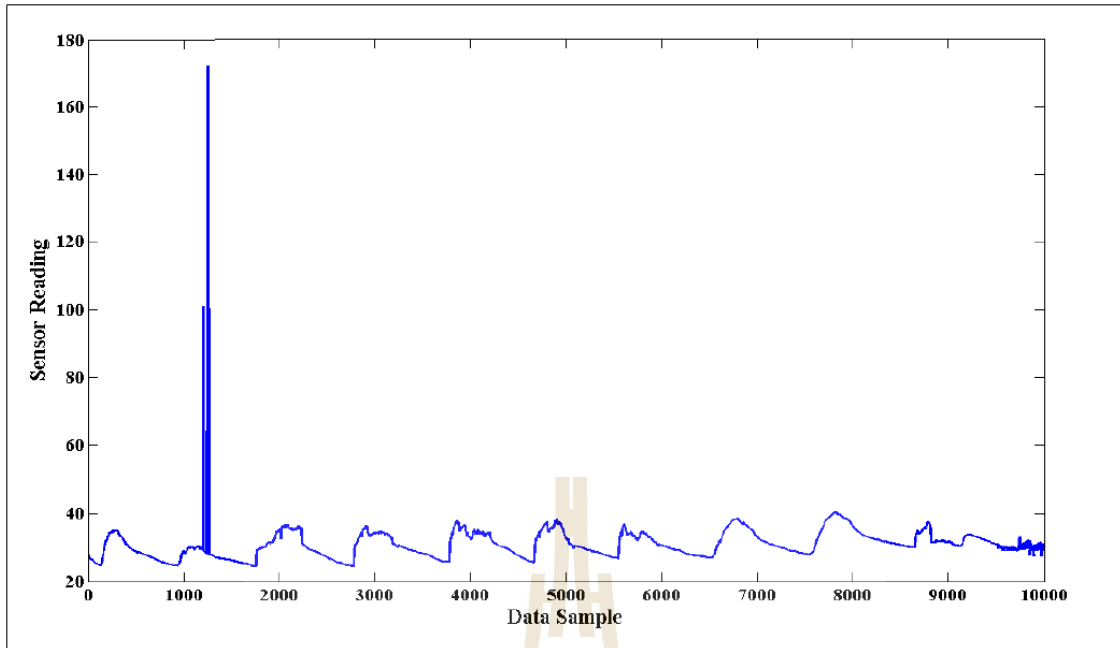
รูปที่ ข.34 ชุดข้อมูล INTEL (อุณหภูมิที่อ่านได้)



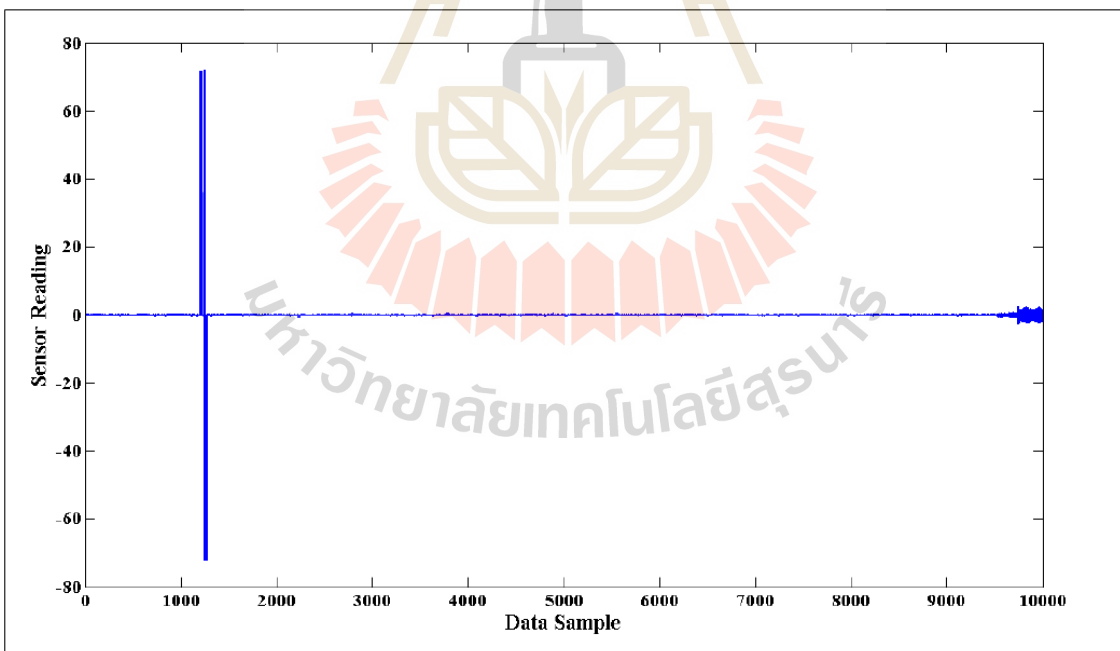
รูปที่ ข.35 DWT สัมประสิทธิ์ผ่านต่ำของชุดข้อมูลสังเคราะห์ INTEL (อนุกรมที่อ่านได้)



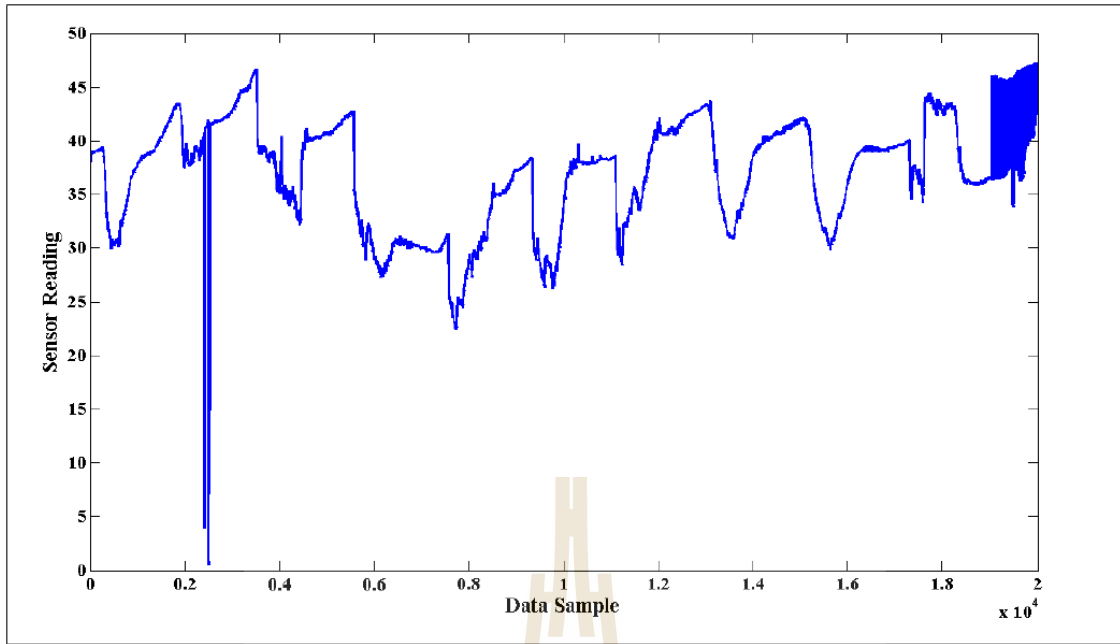
รูปที่ ข.36 DWT สัมประสิทธิ์ผ่านสูงของชุดข้อมูลสังเคราะห์ INTEL (อนุกรมที่อ่านได้)



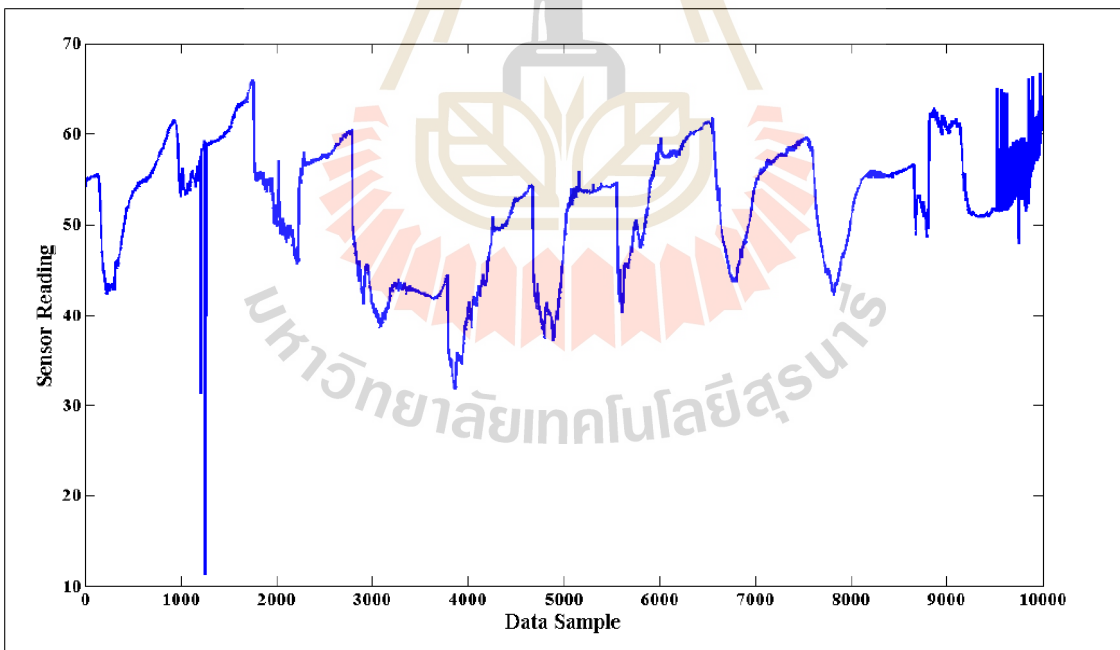
รูปที่ ข.37 LWT สัมประสิทธิ์ผ่านต่ำของชุดข้อมูลสังเคราะห์ INTEL (อุณหภูมิที่อ่านได้)



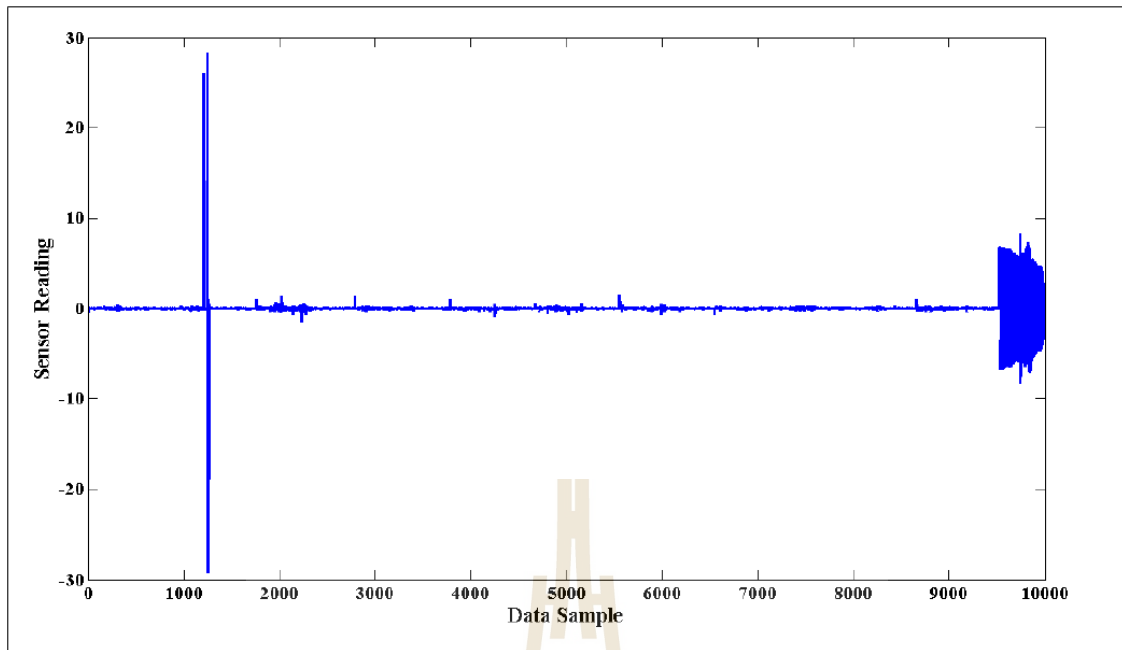
รูปที่ ข.38 LWT สัมประสิทธิ์ผ่านสูงของชุดข้อมูลสังเคราะห์ INTEL (อุณหภูมิที่อ่านได้)



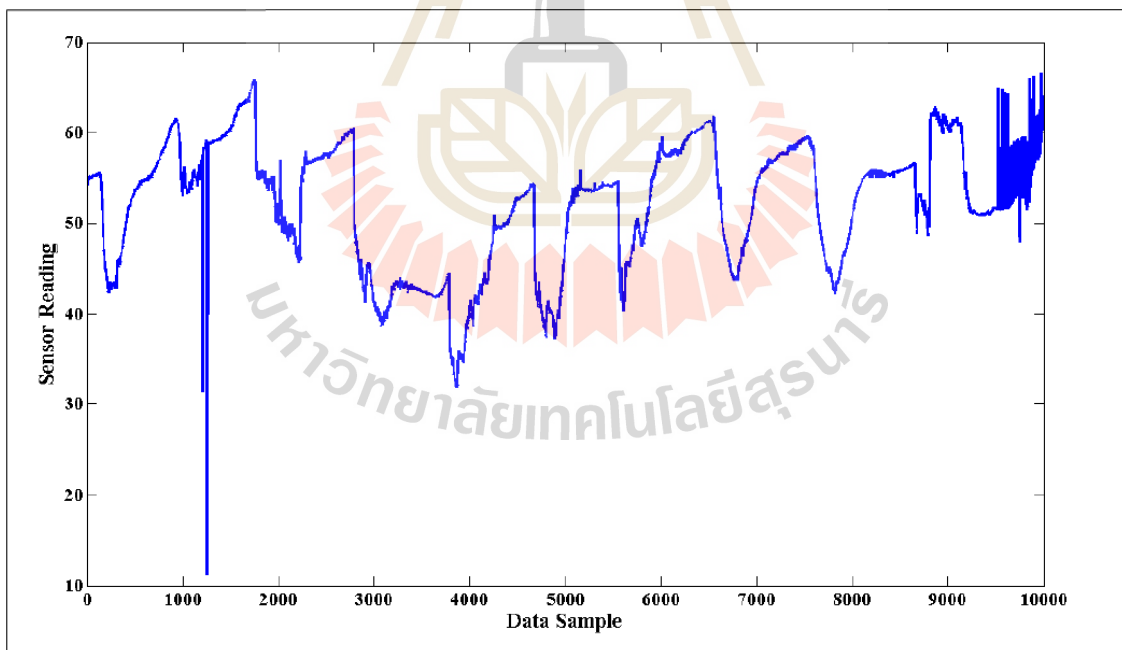
รูปที่ ข.39 ชุดข้อมูล INTEL (ความชื้นที่อ่านได้)



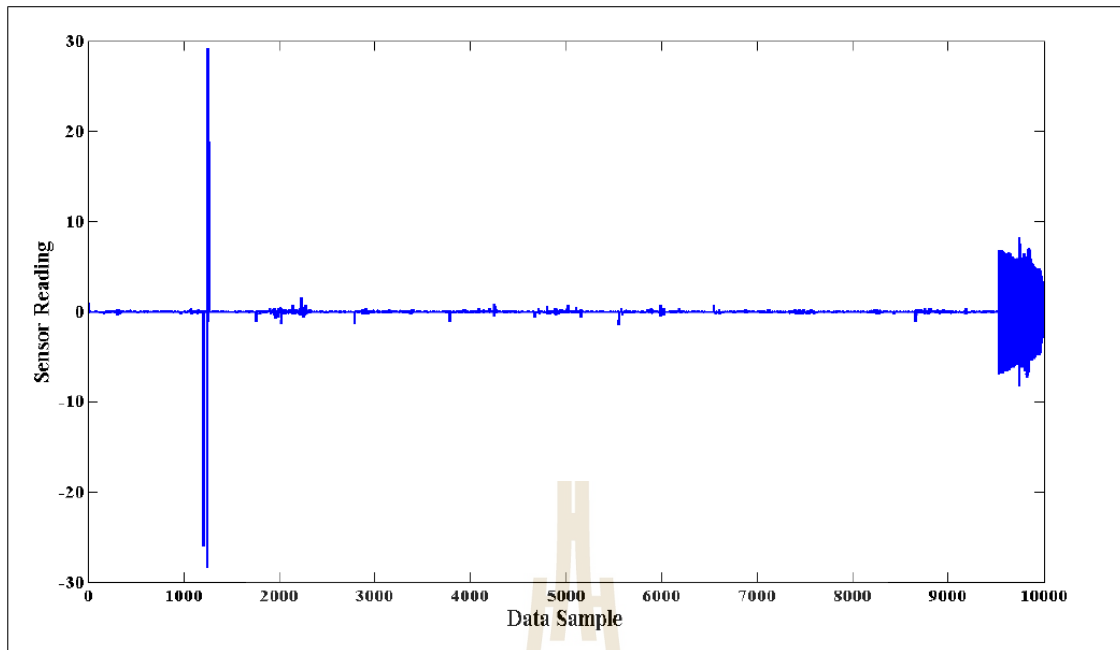
รูปที่ ข.40 DWT สัมประสิทธิ์ผ่านต่ำของชุดข้อมูลสังเคราะห์ INTEL (ความชื้นที่อ่านได้)



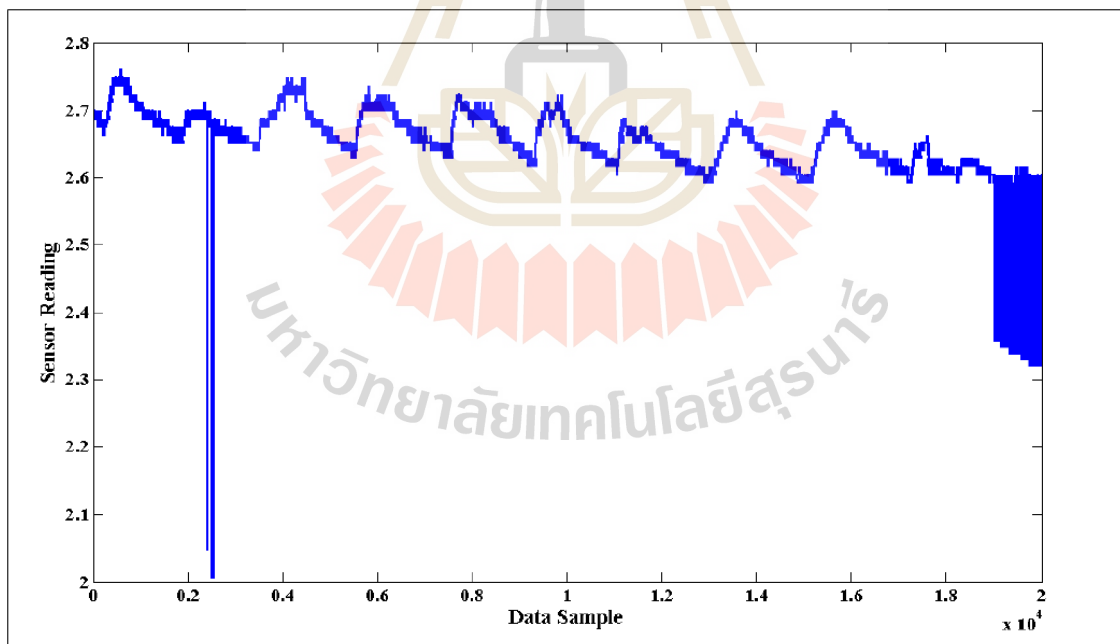
รูปที่ ข.41 DWT สัมประสิทธิ์ผ่านสูงของชุดข้อมูลสังเคราะห์ INTEL (ความถี่ที่อ่านได้)



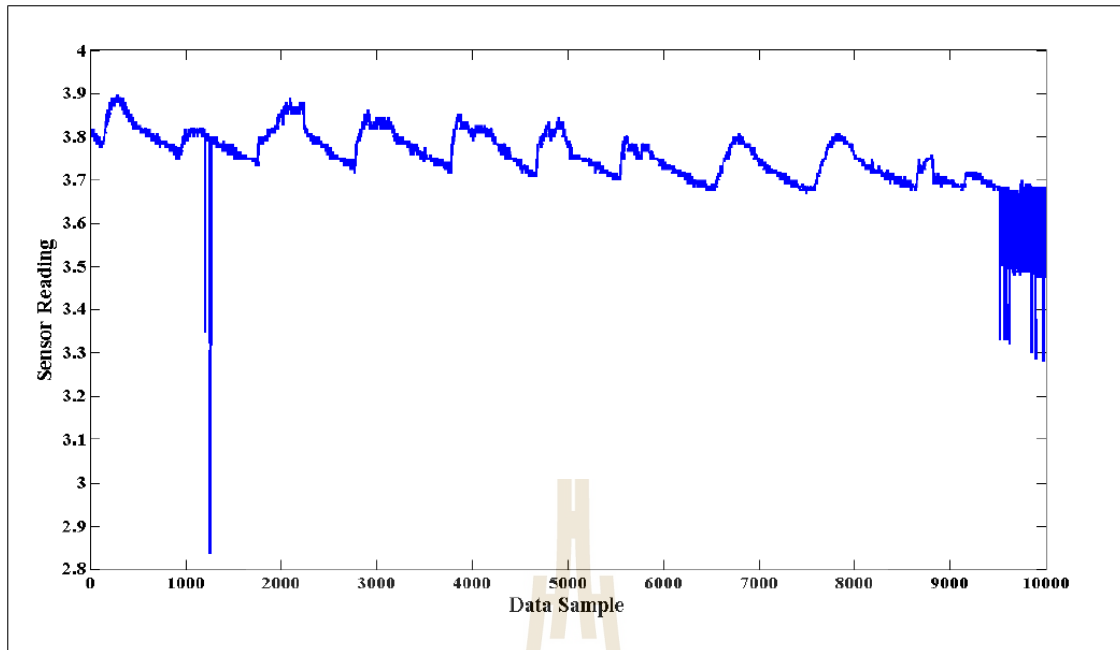
รูปที่ ข.42 LWT สัมประสิทธิ์ผ่านต่ำของชุดข้อมูลสังเคราะห์ INTEL (ความถี่ที่อ่านได้)



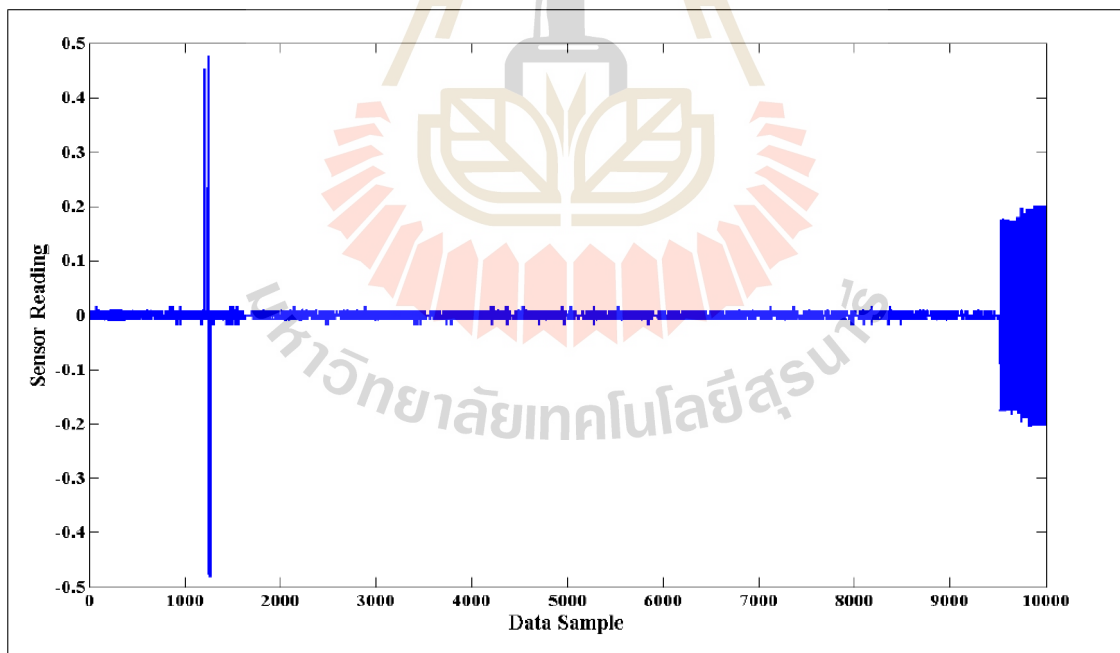
รูปที่ ข.43 LWT สัมประสิทธิ์ผ่านสูงของชุดข้อมูลสังเคราะห์ INTEL (ความถี่ที่อ่านได้)



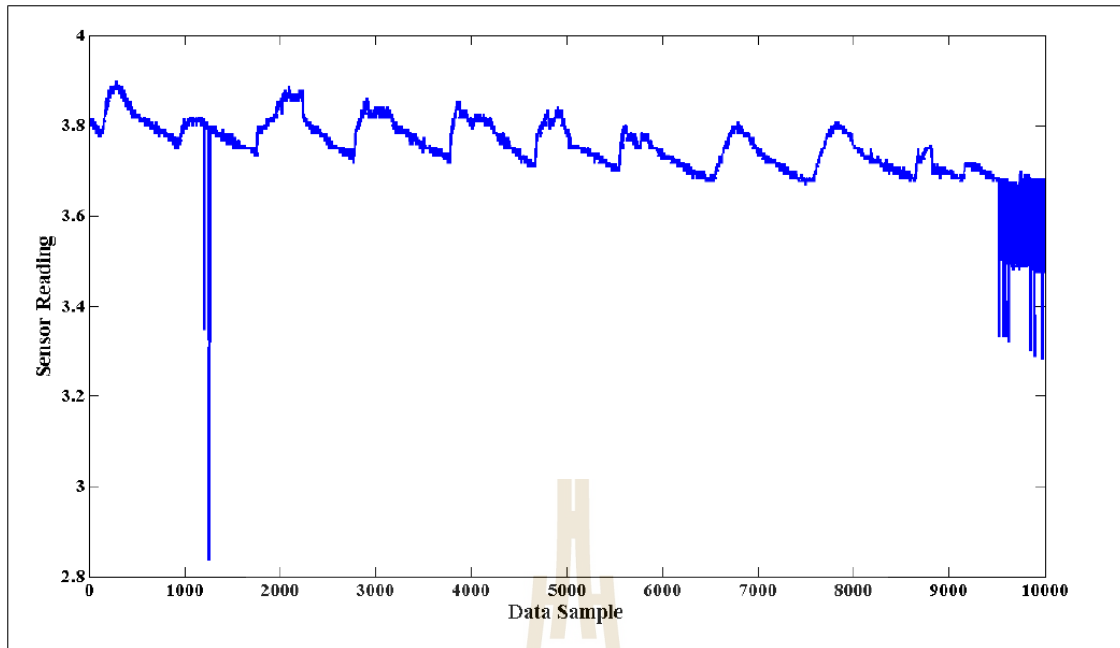
รูปที่ ข.44 ชุดข้อมูล INTEL (แรงดันไฟฟ้าอ่านได้)



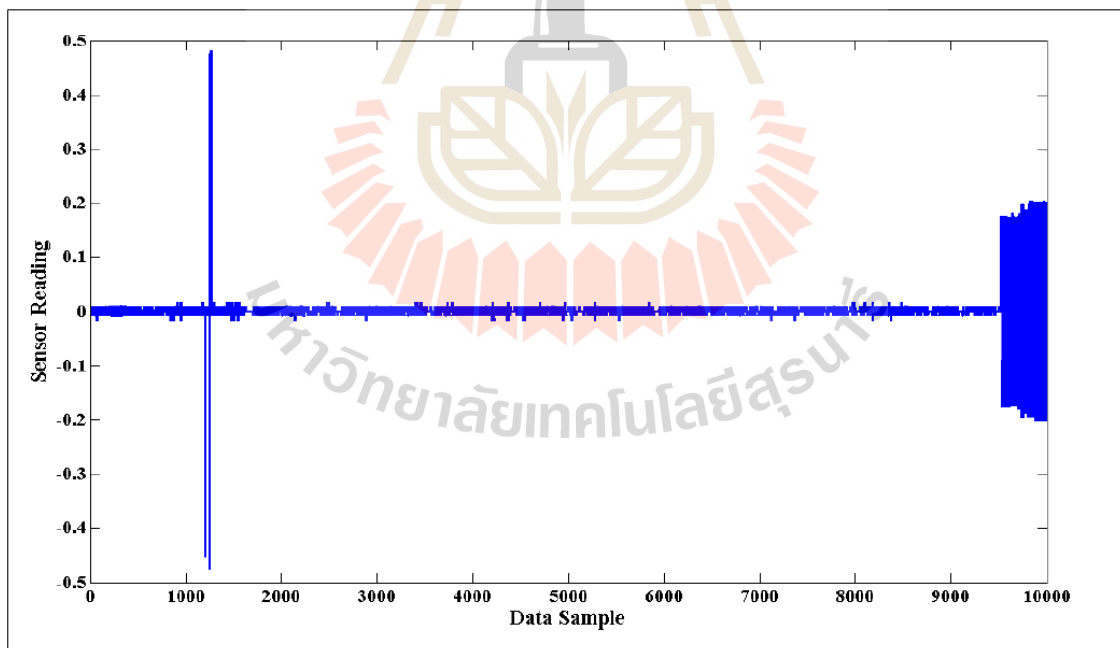
รูปที่ ข.45 DWT สัมประสิทธิ์ผ่านต่ำของชุดข้อมูลสังเคราะห์ INTEL (แรงดันไฟฟ้าอ่านได้)



รูปที่ ข.46 DWT สัมประสิทธิ์ผ่านสูงของชุดข้อมูลสังเคราะห์ INTEL (แรงดันไฟฟ้าที่อ่านได้)

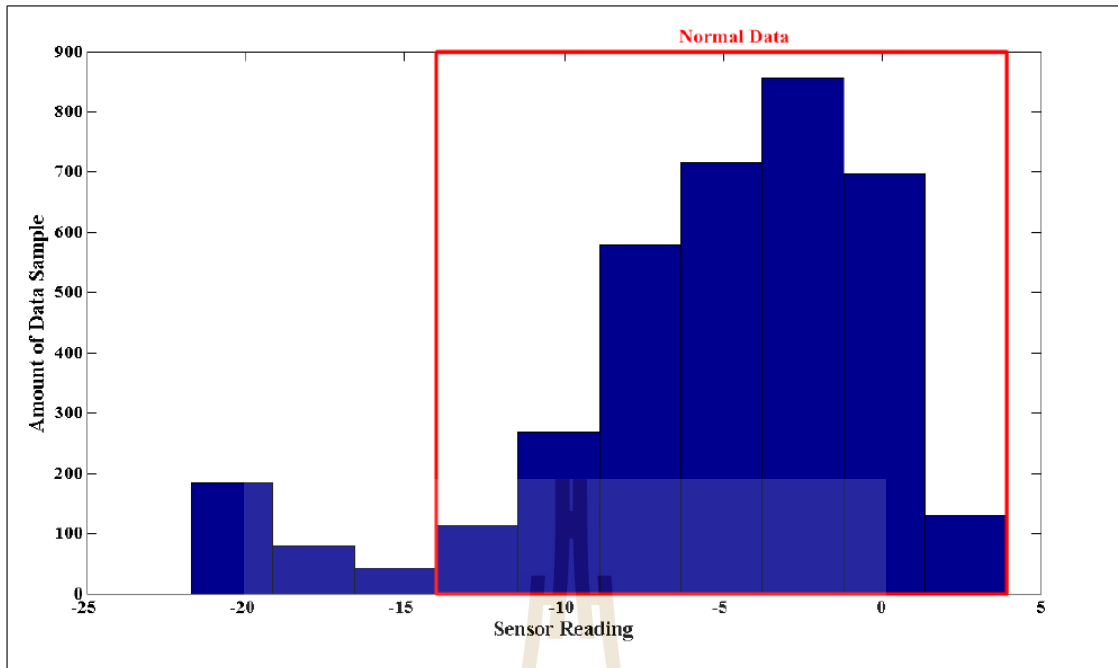


รูปที่ ข.47 LWT สัมประสิทธิ์ผ่านต่ำของชุดข้อมูลสังเคราะห์ INTEL (แรงดันไฟฟ้าที่อ่านได้)

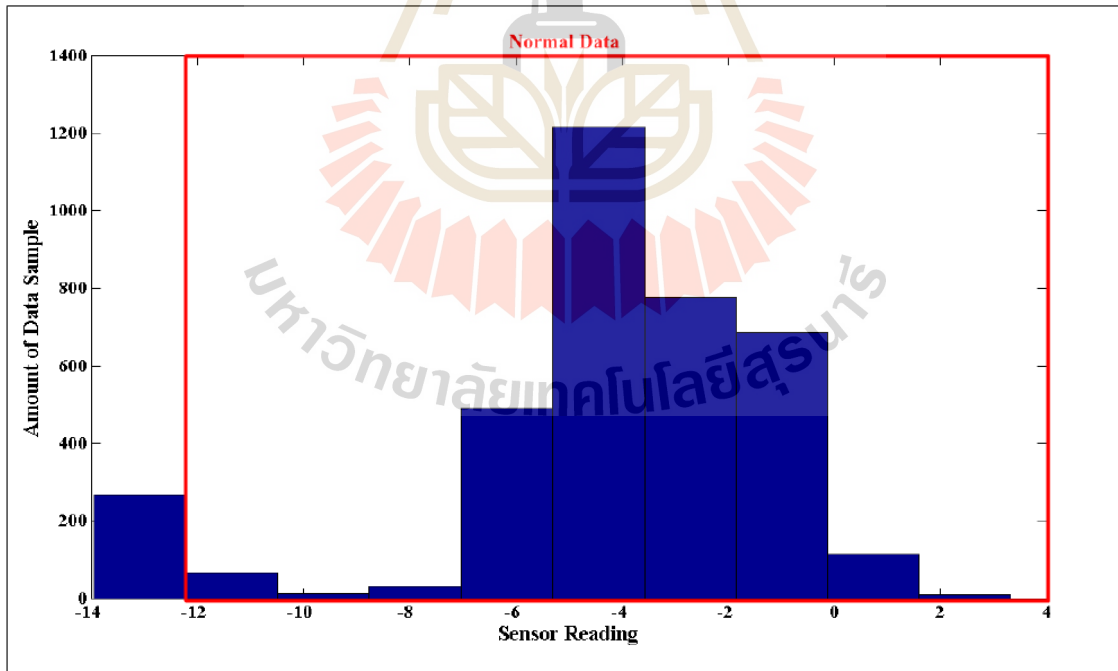


รูปที่ ข.48 LWT สัมประสิทธิ์ผ่านสูงของชุดข้อมูลสังเคราะห์ INTEL (แรงดันไฟฟ้าที่อ่านได้)

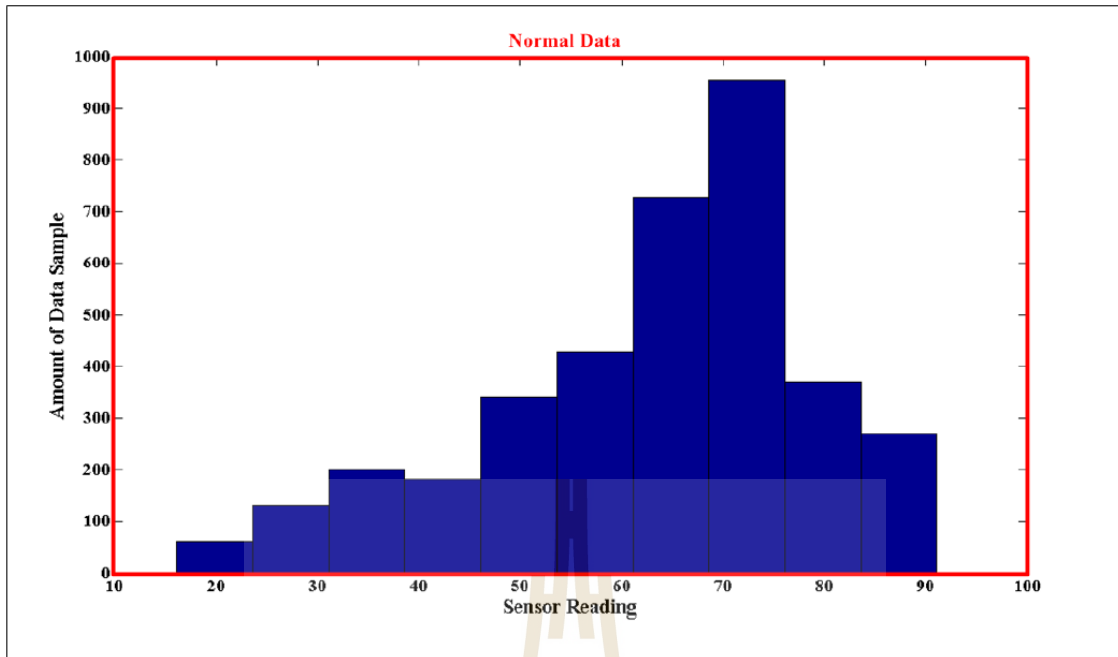
3. เครื่องมือตรวจรู้ pdg2008-metro-1 dataset (pdg2008)



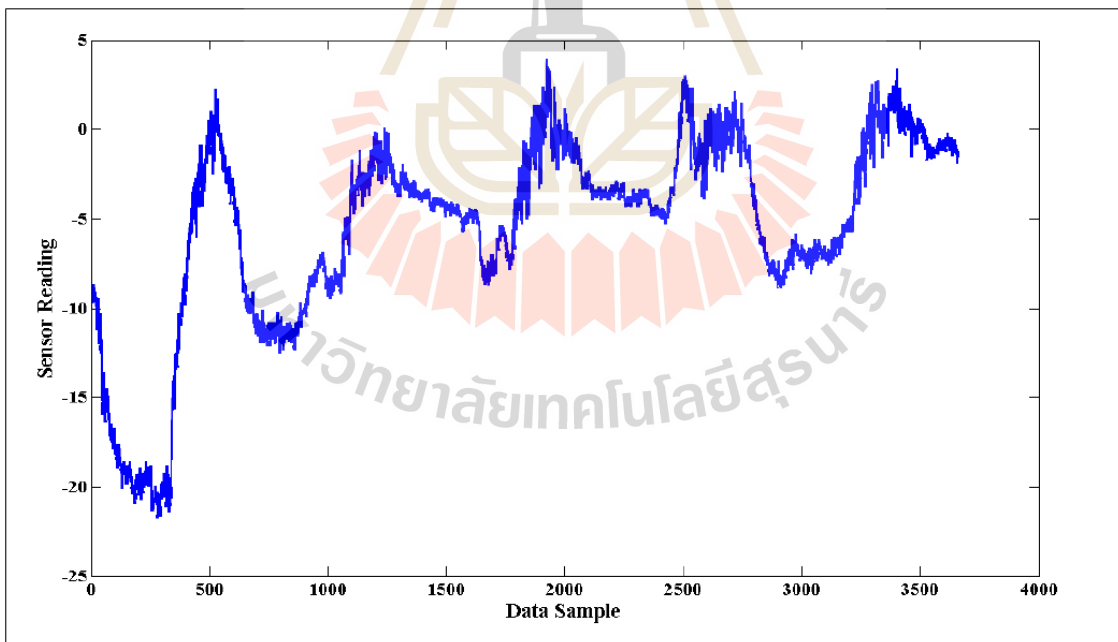
รูปที่ ข.49 ฮิตโตแกรมแสดงค่าของชุดข้อมูล pdg2008 (อุณหภูมิพื้นผิวที่อ่านได้)



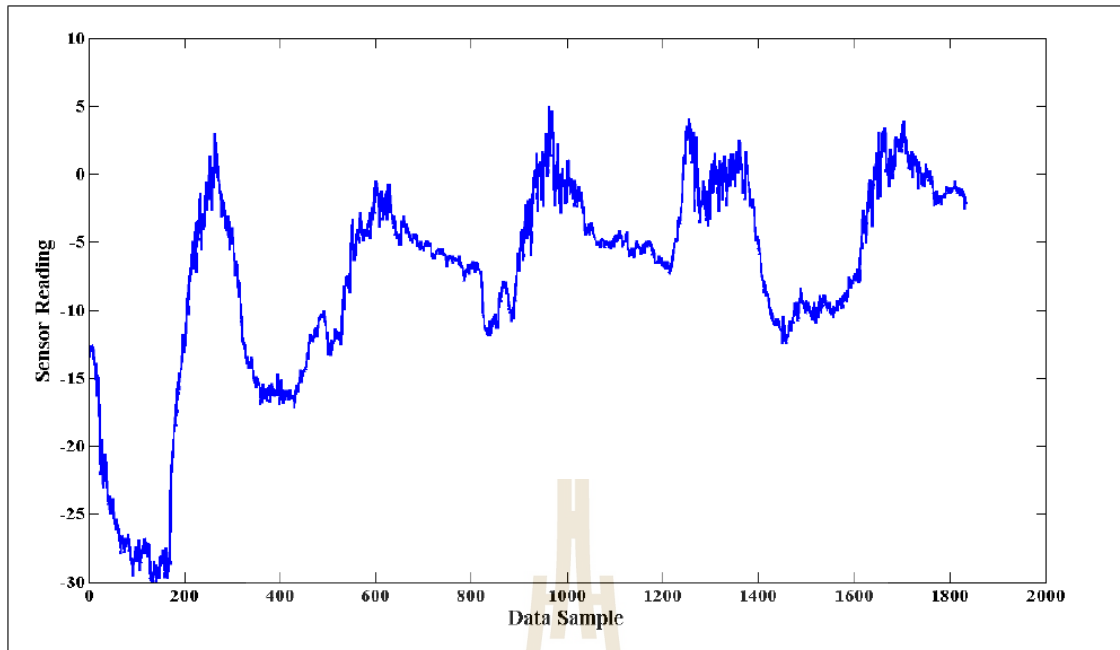
รูปที่ ข.50 ฮิตโตแกรมแสดงค่าของชุดข้อมูล pdg2008 (อุณหภูมิโดยรอบที่อ่านได้)



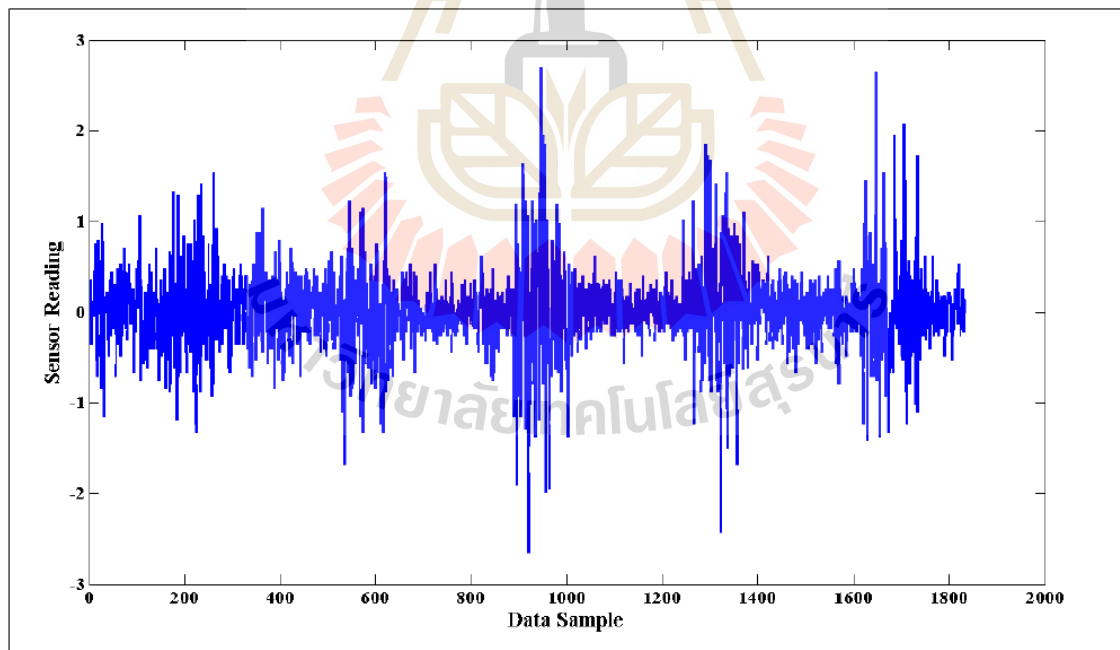
รูปที่ ข.51 ฮิตโตแกรมแสดงค่าของชุดข้อมูล pdg2008 (รังสีดวงอาทิตย์ที่อ่านได้)



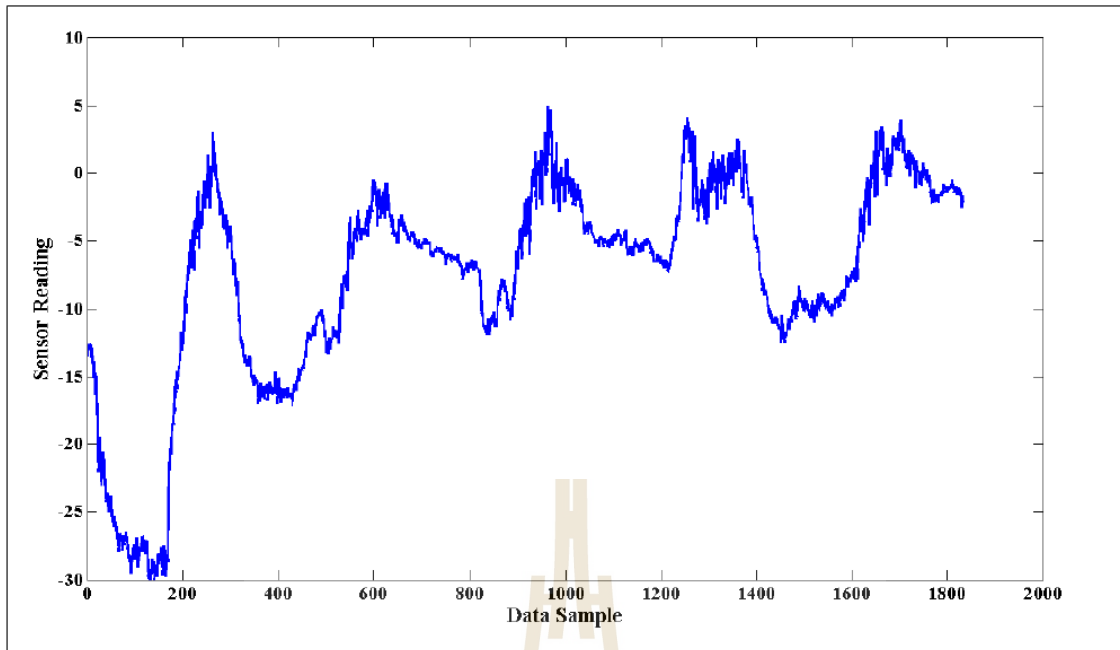
รูปที่ ข.52 ชุดข้อมูล pdg2008 (อุณหภูมิพื้นผิวที่อ่านได้)



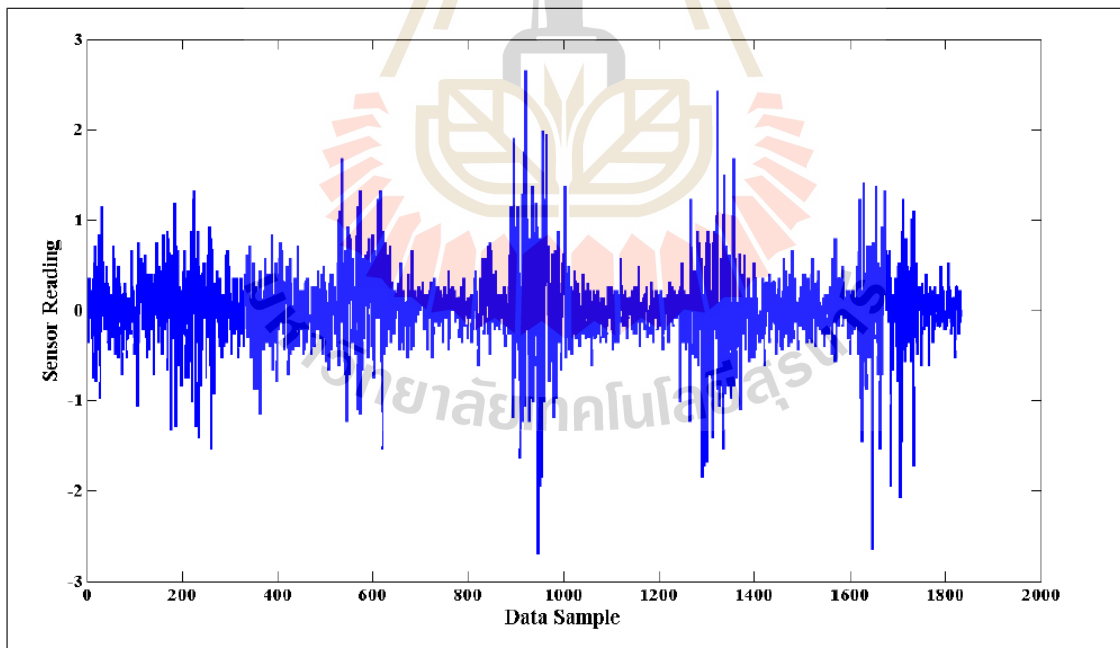
รูปที่ ข.53 DWT สัมประสิทธิ์ผ่านต่ำของชุดข้อมูล pdg2008 (อุณหภูมิพื้นผิวที่อ่านได้)



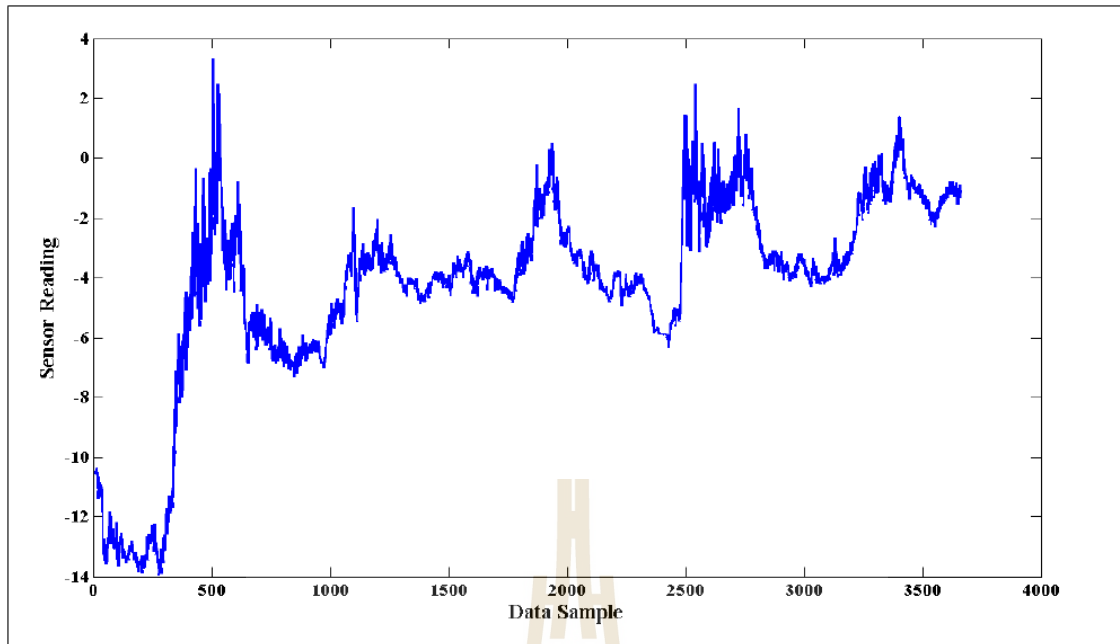
รูปที่ ข.54 DWT สัมประสิทธิ์ผ่านสูงของชุดข้อมูล pdg2008 (อุณหภูมิพื้นผิวที่อ่านได้)



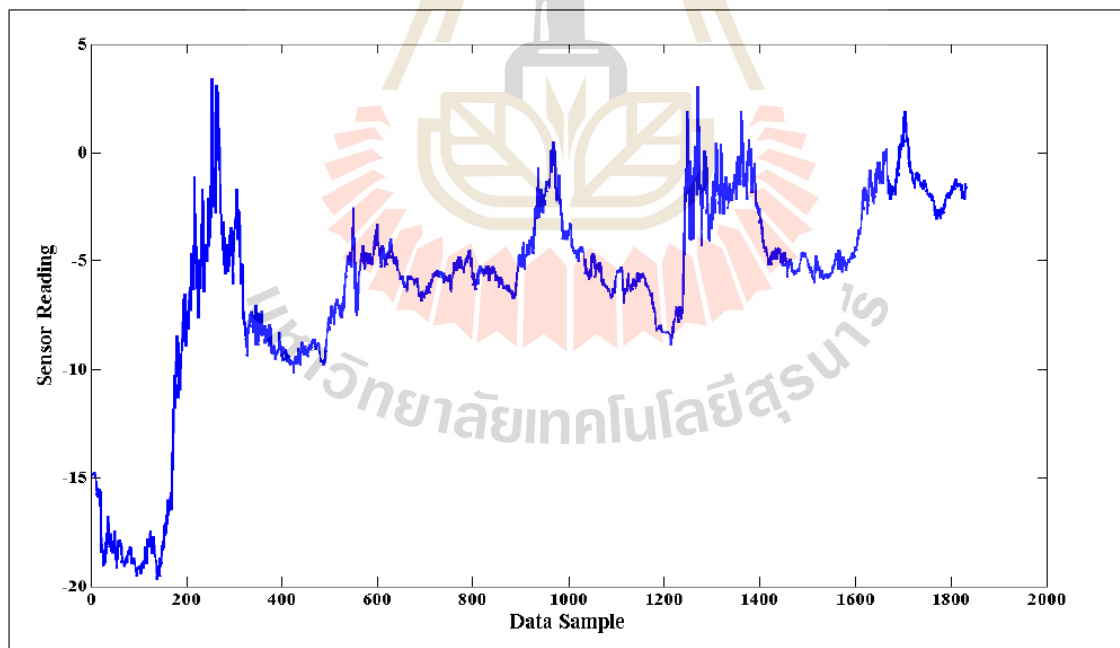
รูปที่ ข.55 LWT สัมประสิทธิ์ผ่านต่ำของชุดข้อมูล pdg2008 (อนุกรมพื้นที่อ่านได้)



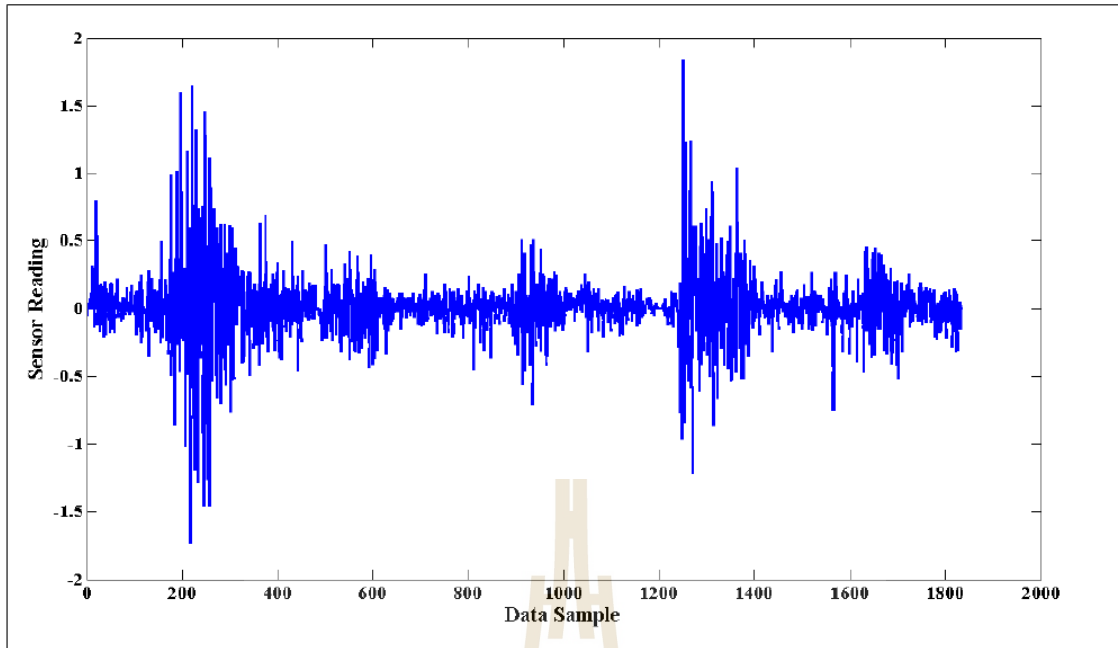
รูปที่ ข.56 LWT สัมประสิทธิ์ผ่านสูงของชุดข้อมูล pdg2008 (อนุกรมพื้นที่อ่านได้)



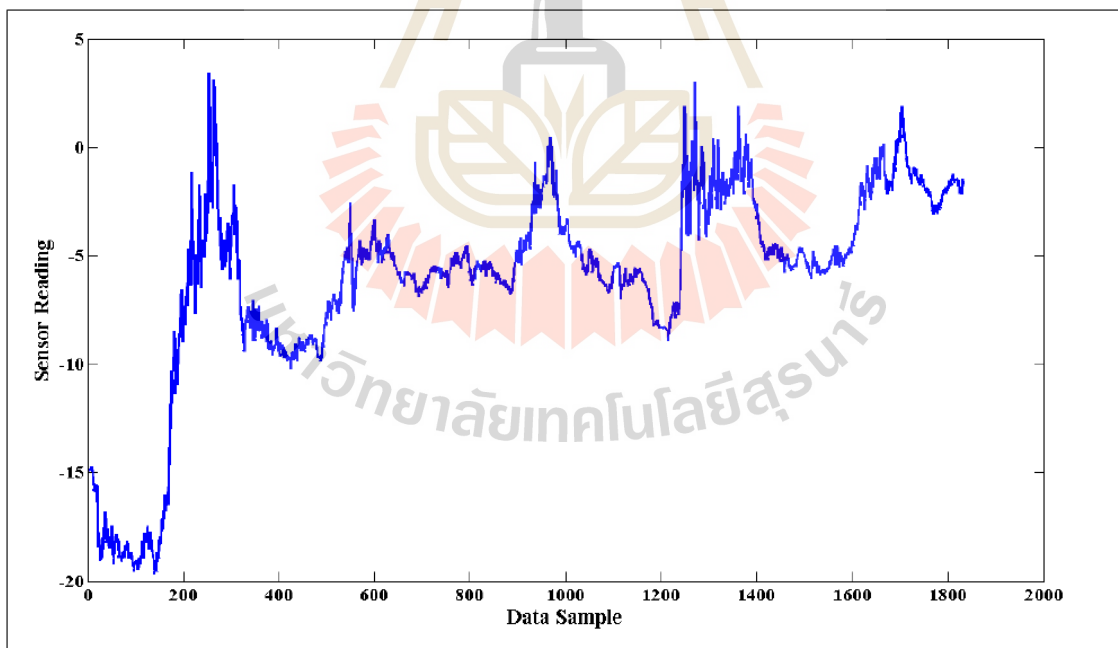
รูปที่ ข.57 ชุดข้อมูล pdg2008 (อุณหภูมิโดยรอบที่อ่านได้)



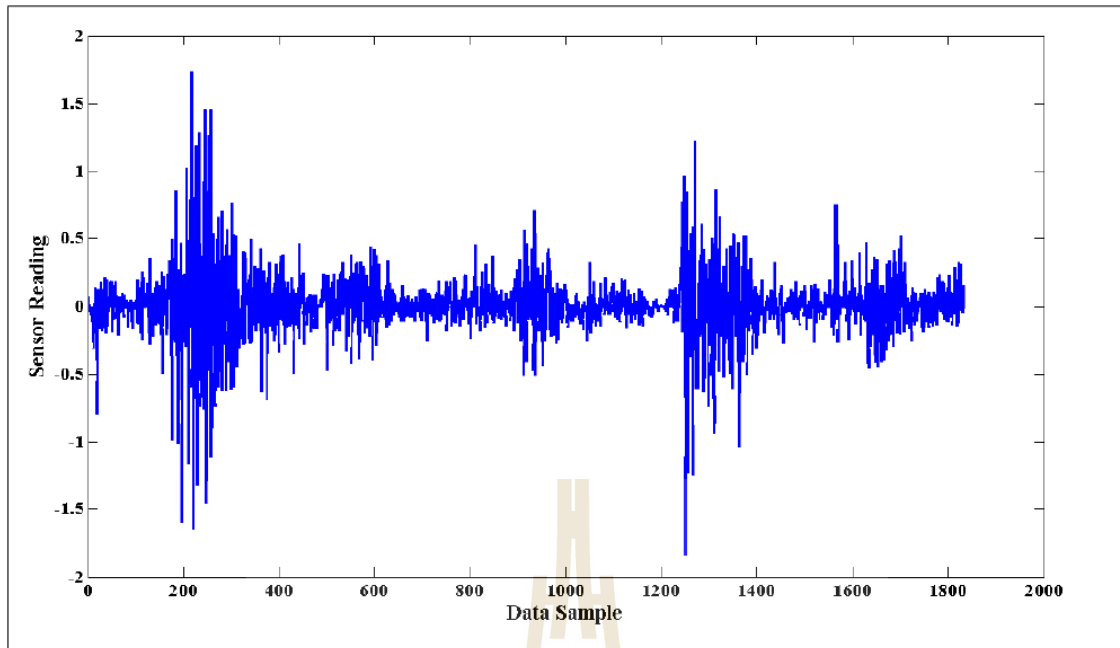
รูปที่ ข.58 DWT สัมประสิทธิ์ผ่านต่ำของชุดข้อมูล pdg2008 (อุณหภูมิโดยรอบที่อ่านได้)



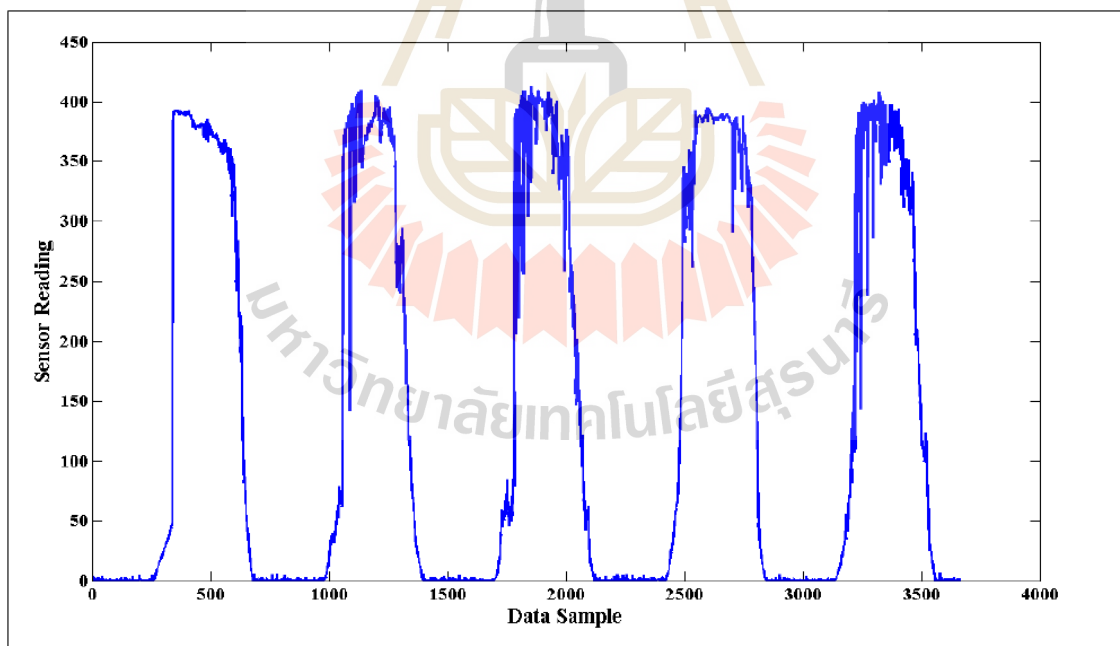
รูปที่ ข.59 DWT สัมประสิทธิ์ผ่านสูงของชุดข้อมูล pdg2008 (อุณหภูมิโดยรอบที่อ่านได้)



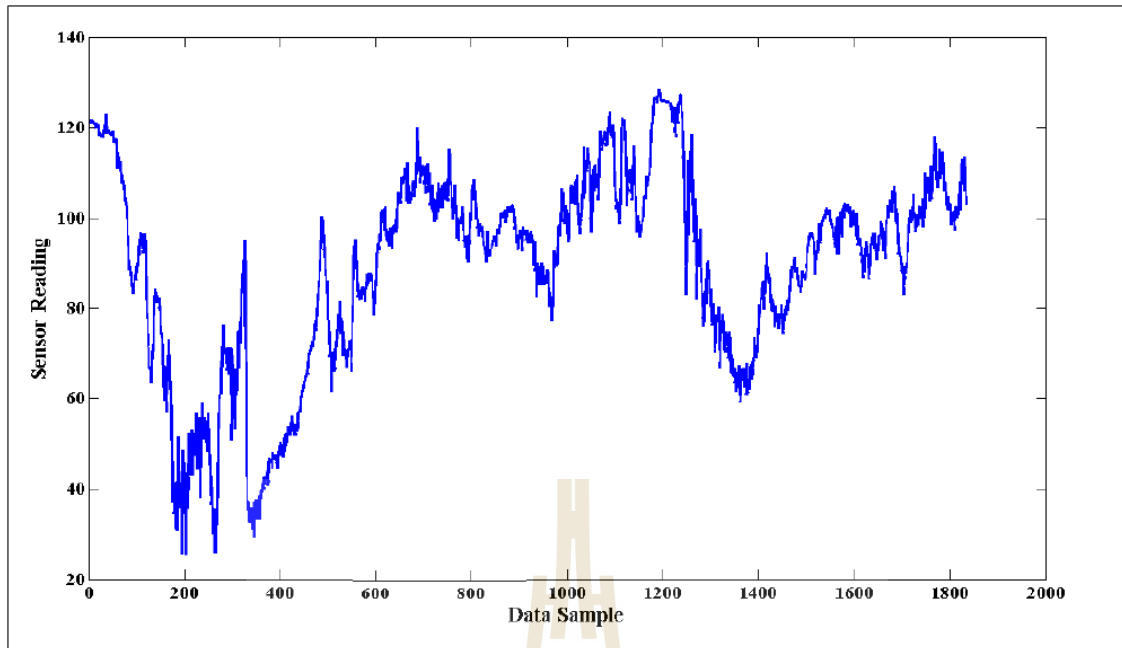
รูปที่ ข.60 LWT สัมประสิทธิ์ผ่านต่ำของชุดข้อมูล pdg2008 (อุณหภูมิโดยรอบที่อ่านได้)



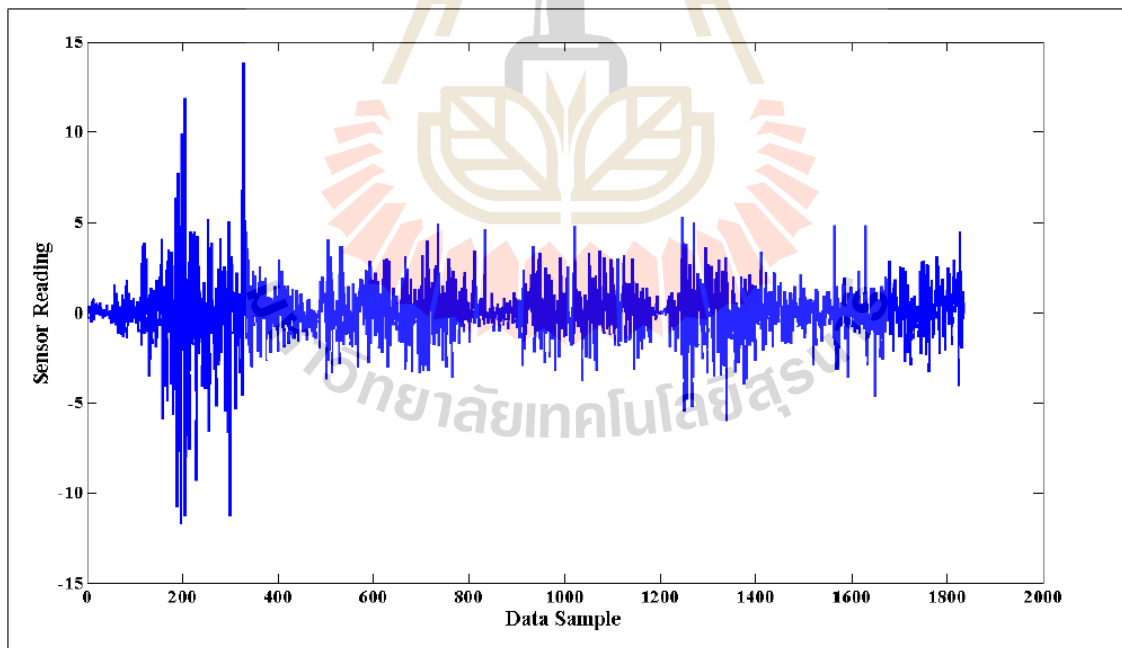
รูปที่ ข.61 LWT สัมประสิทธิ์ผ่านสูงของชุดข้อมูล pdg2008 (อุณหภูมิโดยรอบที่อ่านได้)



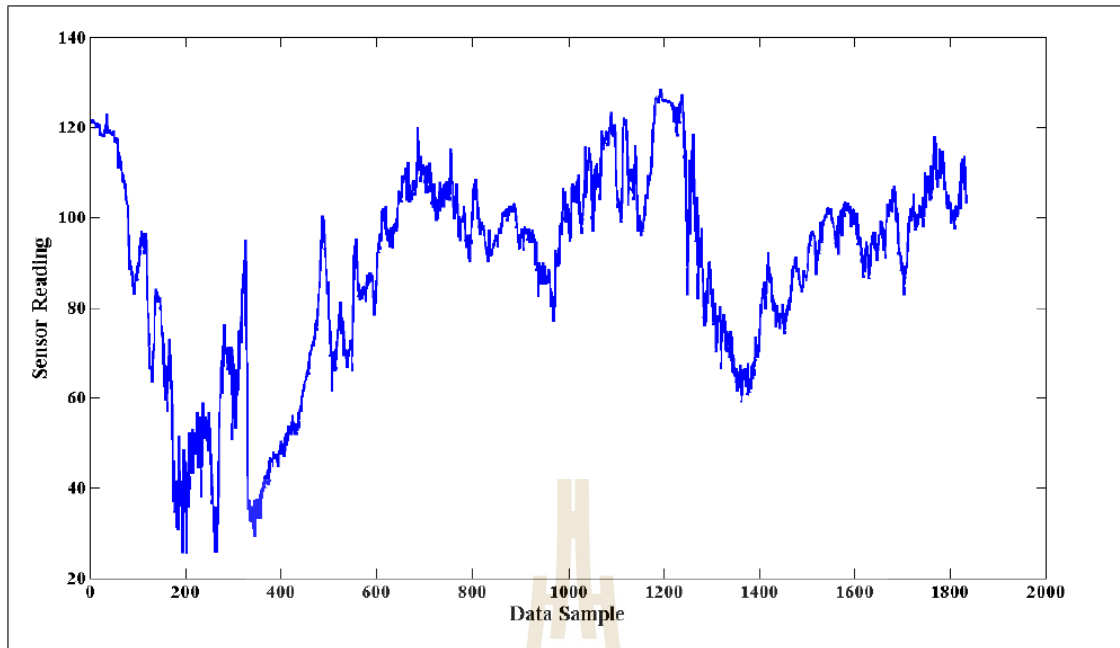
รูปที่ ข.62 ชุดข้อมูล pdg2008 (รังสีดวงอาทิตย์ที่อ่านได้)



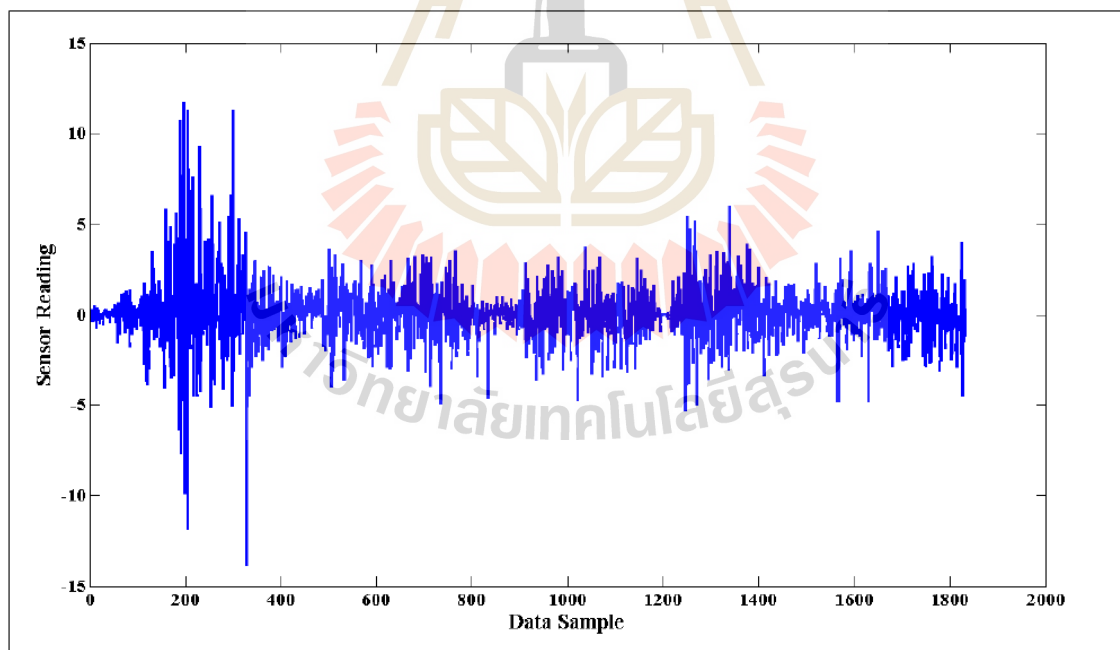
รูปที่ ข.63 DWT สัมประสิทธิ์ผ่านต่ำของชุดข้อมูล pdg2008 (รังสีดวงอาทิตย์ที่อ่านได้)



รูปที่ ข.64 DWT สัมประสิทธิ์ผ่านสูงของชุดข้อมูล pdg2008 (รังสีดวงอาทิตย์ที่อ่านได้)

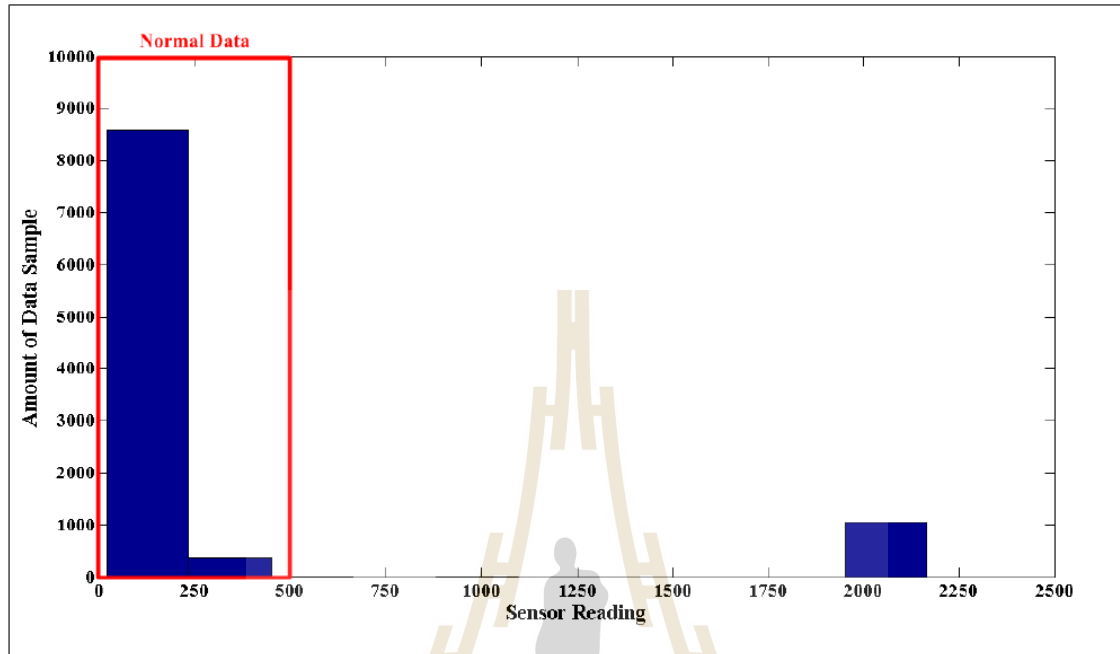


รูปที่ ข.65 LWT สัมประสิทธิ์ผ่านต่ำของชุดข้อมูล pdg2008 (รังสีดวงอาทิตย์ที่อ่านได้)

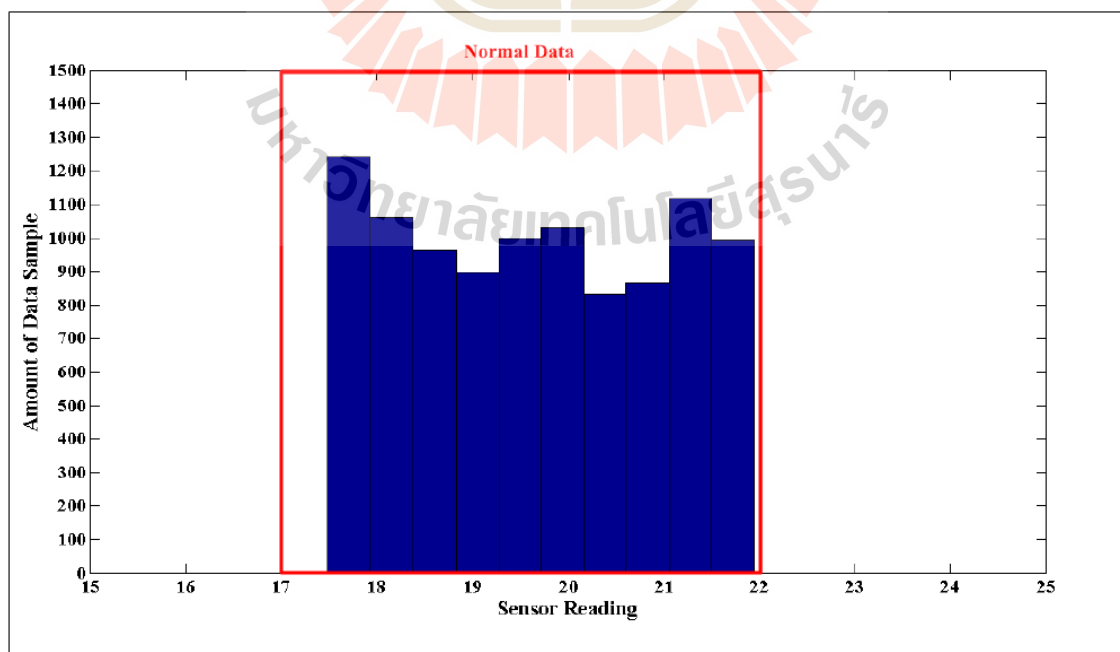


รูปที่ ข.66 LWT สัมประสิทธิ์ผ่านสูงของชุดข้อมูล pdg2008 (รังสีดวงอาทิตย์ที่อ่านได้)

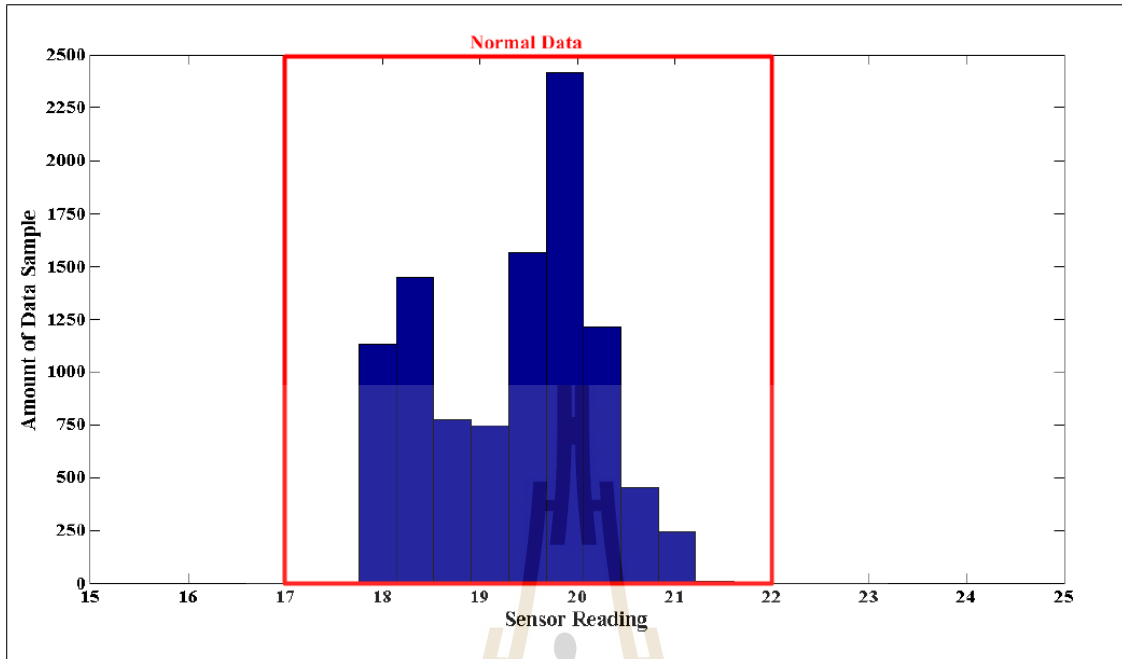
4. ชุดข้อมูล NAMOS



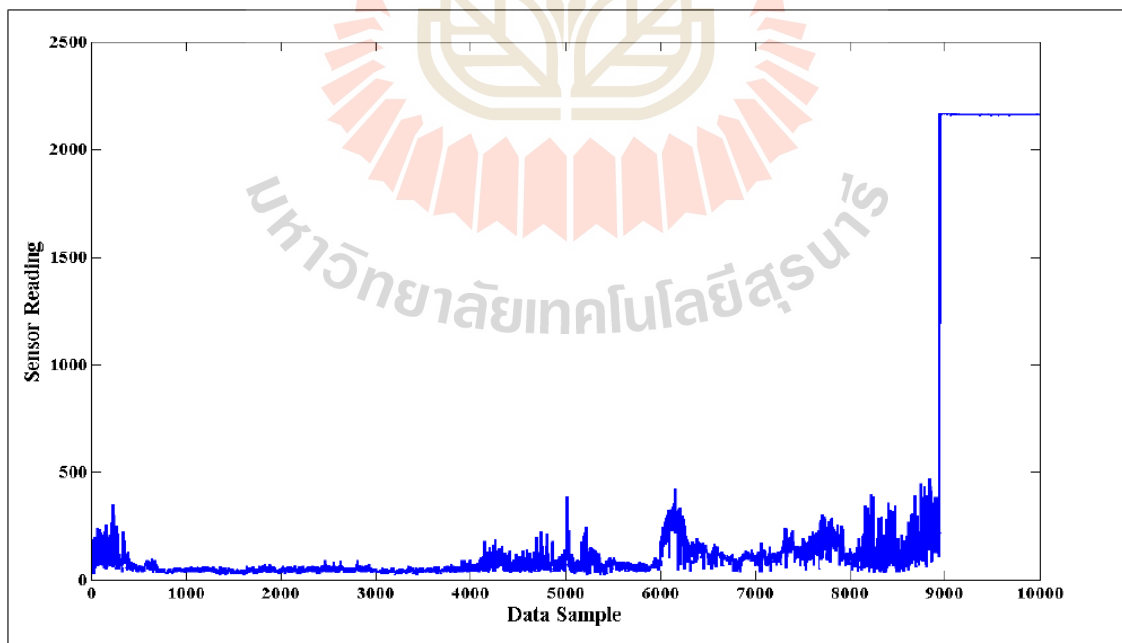
รูปที่ ข.67 ฮิตโตแกรมแสดงค่าของชุดข้อมูล NAMOS (แสงฟลูออเรสเซนซ์ที่อ่านได้)



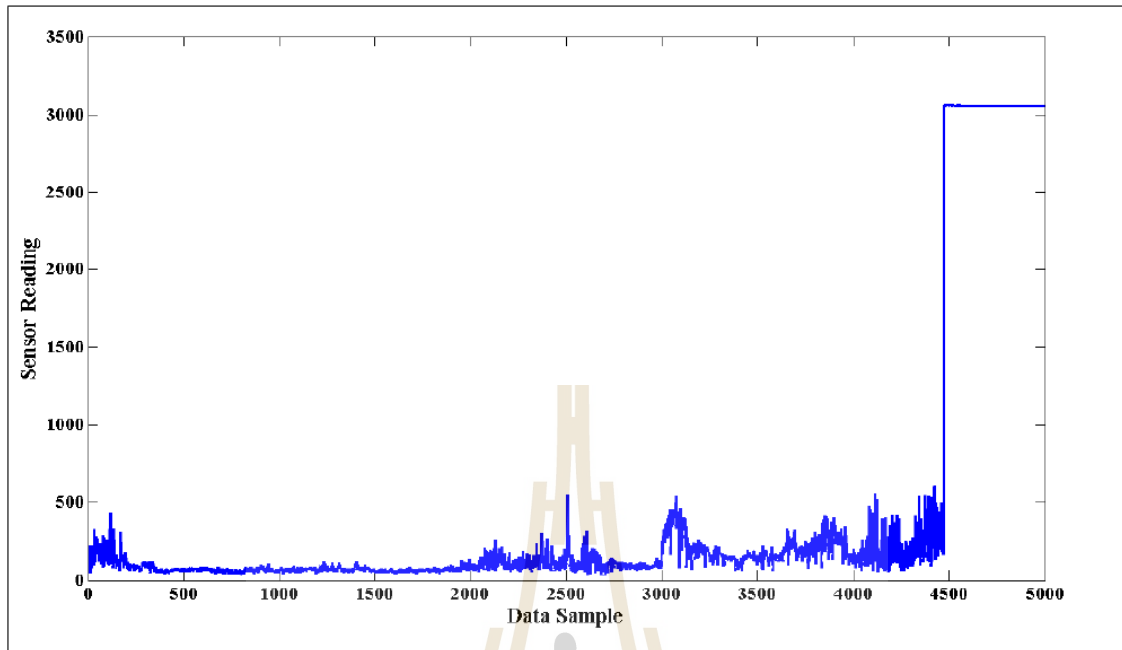
รูปที่ ข.68 ฮิตโตแกรมแสดงค่าของชุดข้อมูล NAMOS (อุณหภูมิที่อ่านได้จากเซ็นเซอร์ 1)



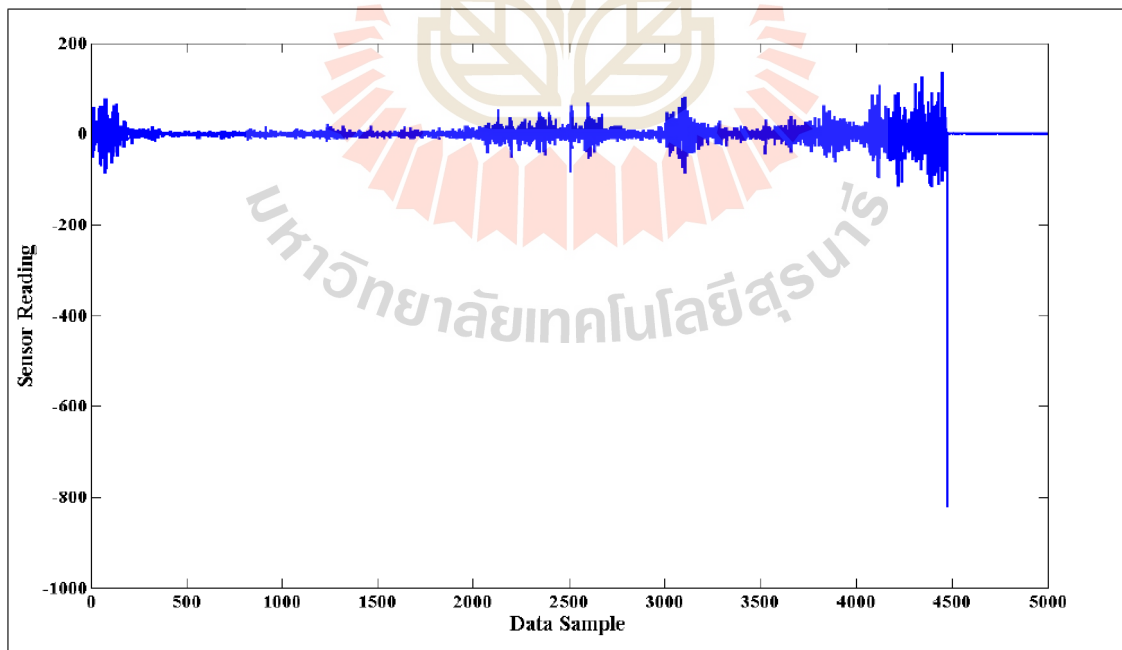
รูปที่ ข.69 ฮิสโตแกรมแสดงค่าของชุดข้อมูล NAMOS (อุณหภูมิที่อ่านได้จากเซ็นเซอร์ 2)



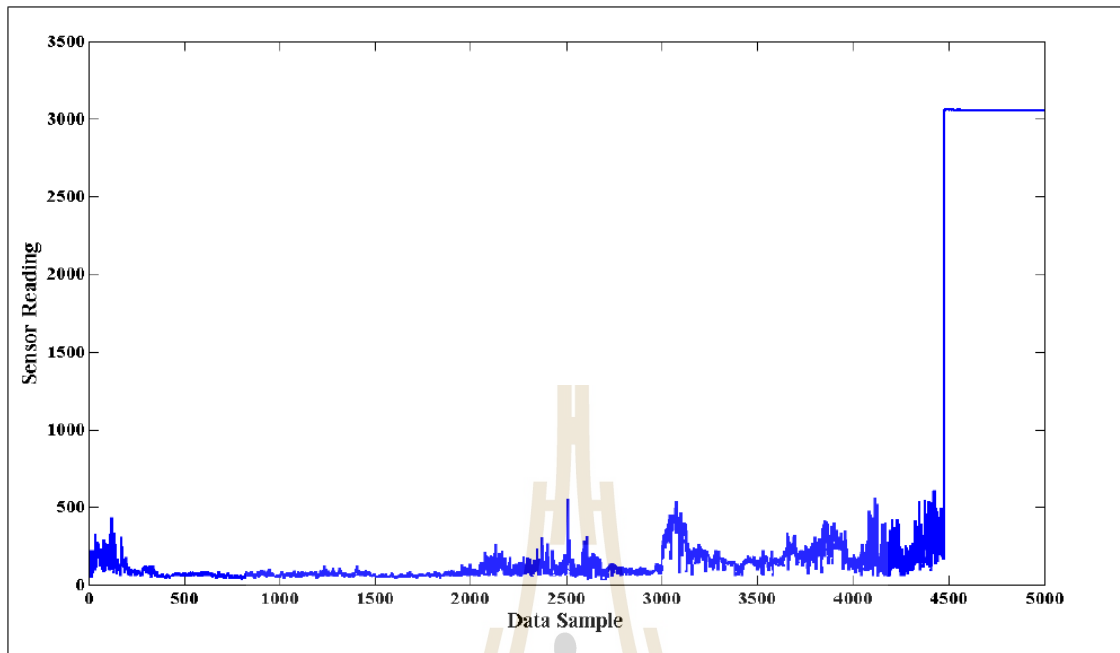
รูปที่ ข.70 ชุดข้อมูล NAMOS (แสงฟลูออเรสเซนซ์ที่อ่านได้)



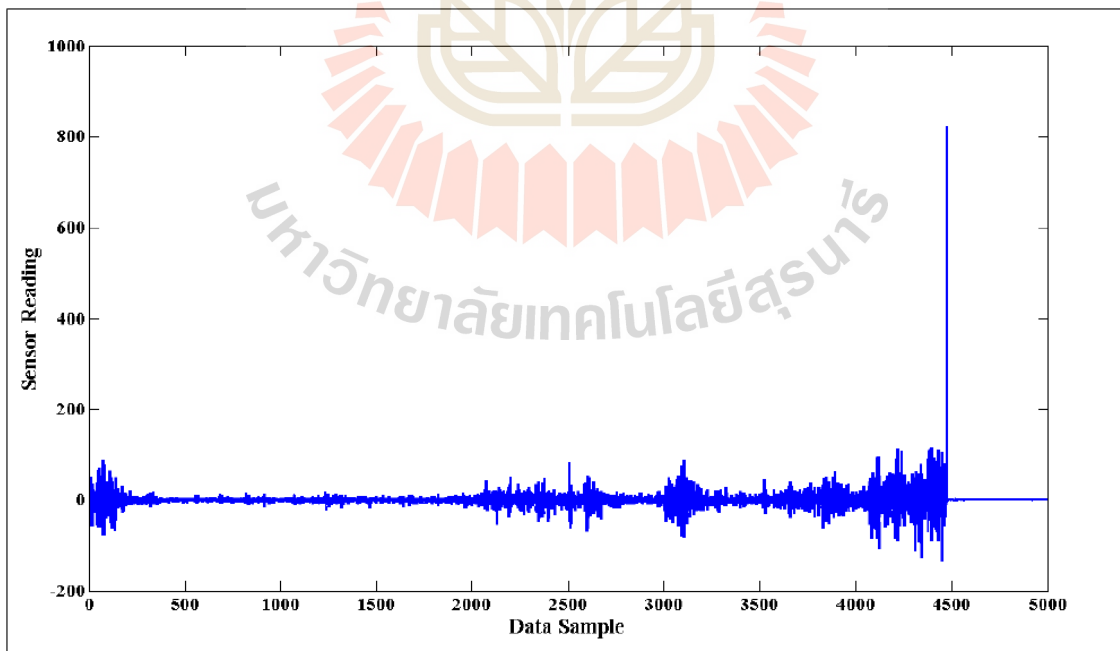
รูปที่ ข.71 DWT สัมประสิทธิ์ผ่านต่ำของชุดข้อมูล NAMOS (แสดงพลูออเรสเซนซ์ที่อ่านได้)



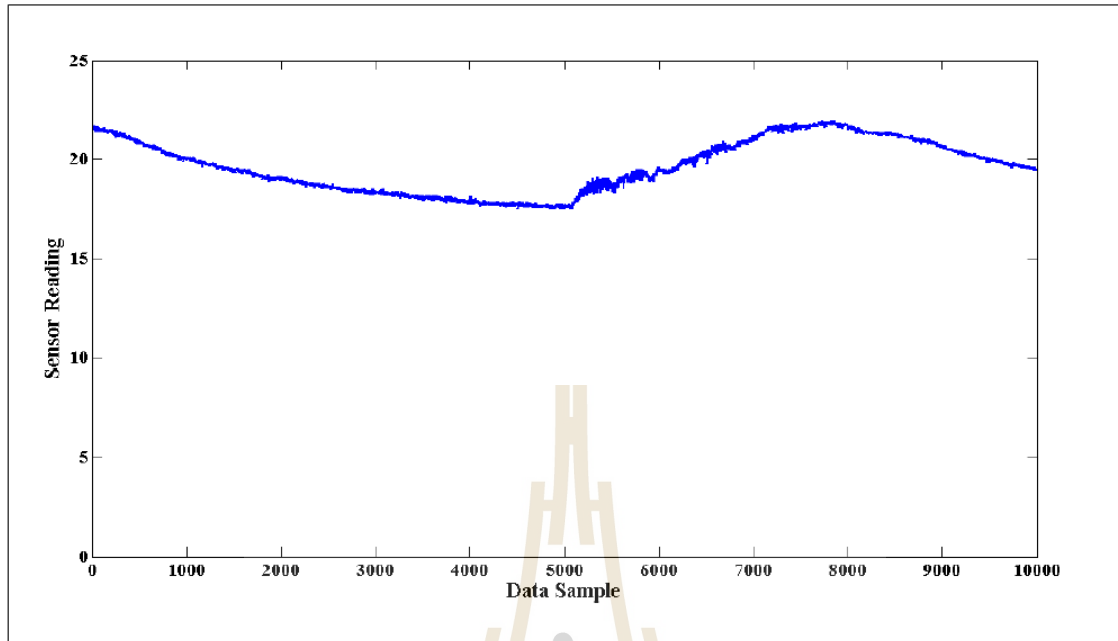
รูปที่ ข.72 DWT สัมประสิทธิ์ผ่านสูงของชุดข้อมูล NAMOS (แสดงพลูออเรสเซนซ์ที่อ่านได้)



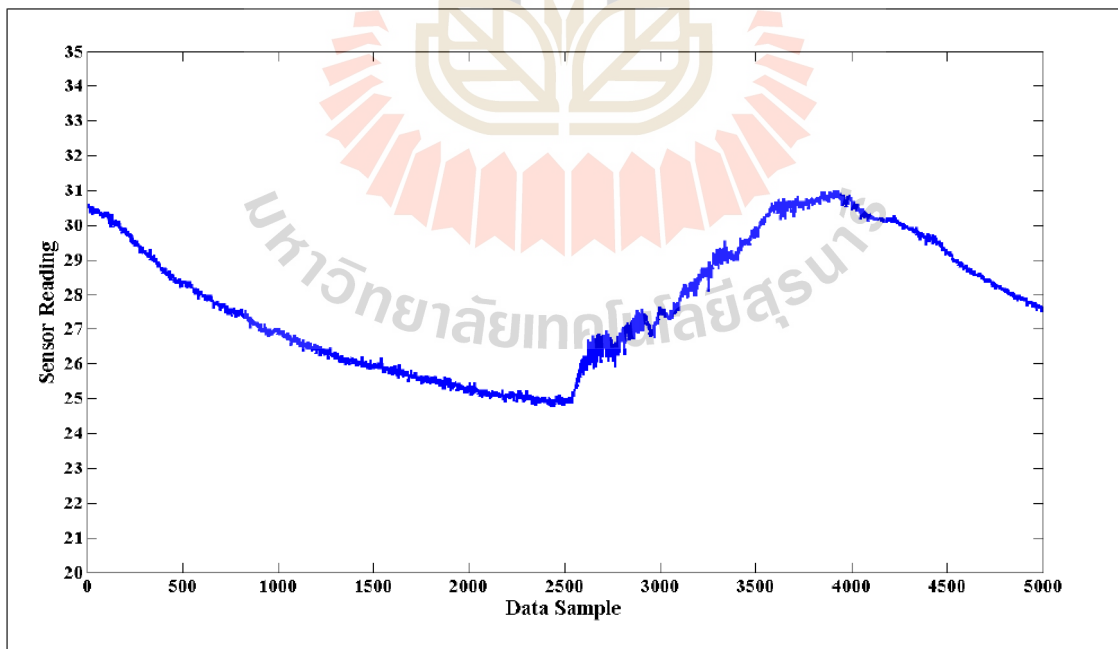
รูปที่ ข.73 LWT สัมประสิทธิ์ผ่านต่ำของชุดข้อมูล NAMOS (แสงฟลูออเรสเซนซ์ที่อ่านได้)



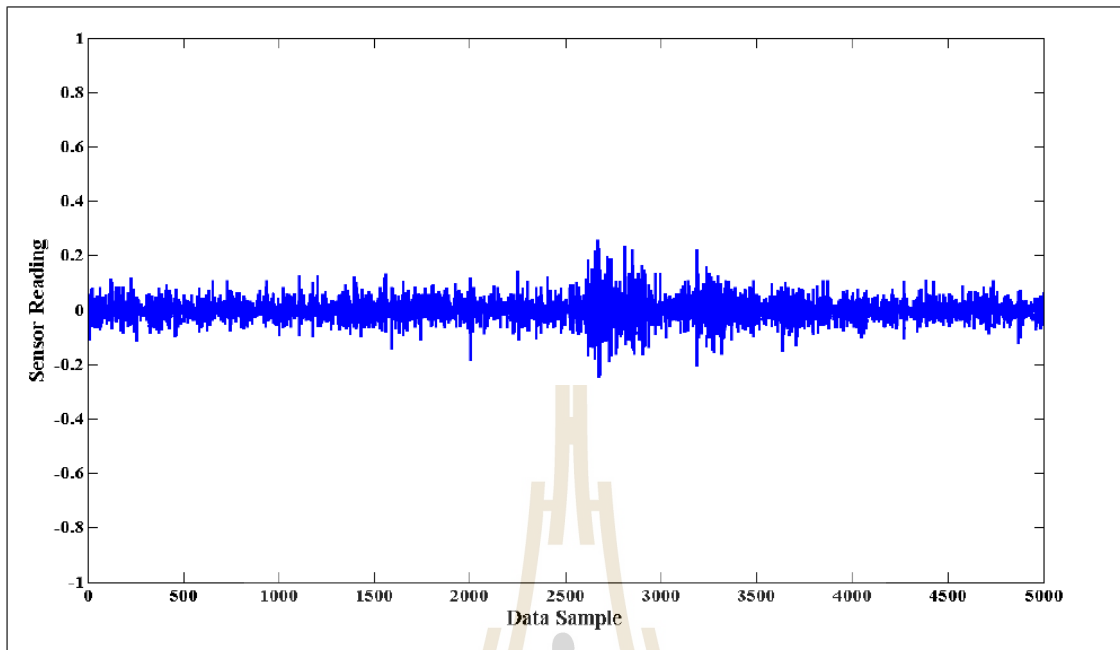
รูปที่ ข.74 LWT สัมประสิทธิ์ผ่านสูงของชุดข้อมูล NAMOS (แสงฟลูออเรสเซนซ์ที่อ่านได้)



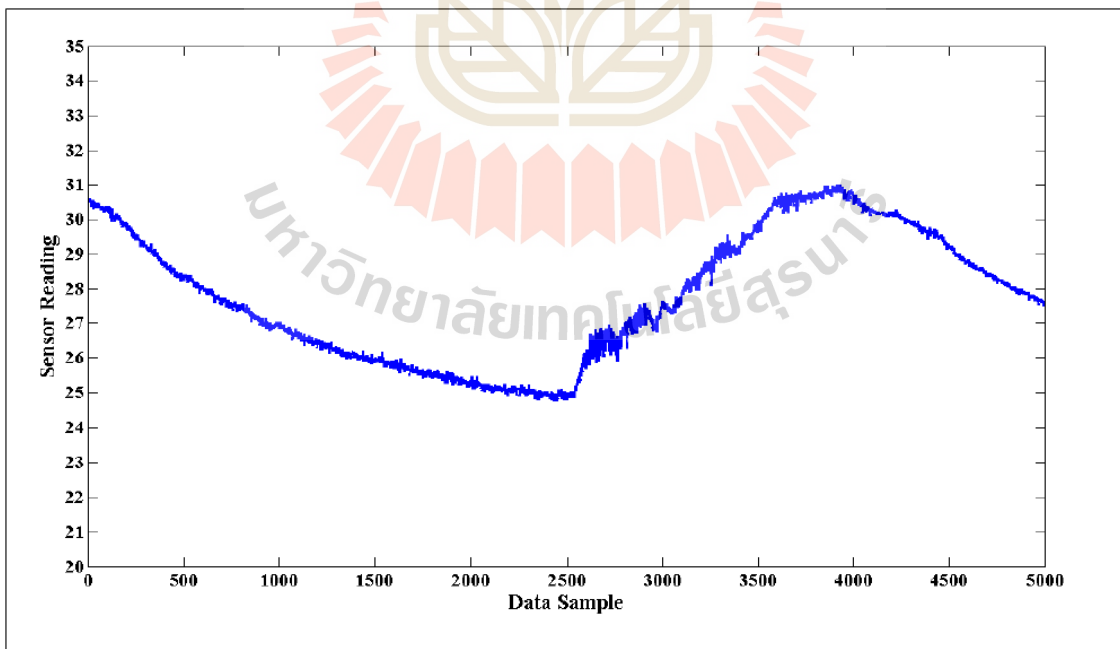
รูปที่ ข.75 ชุดข้อมูล NAMOS (อุณหภูมิที่อ่านได้จากเซ็นเซอร์ 1)



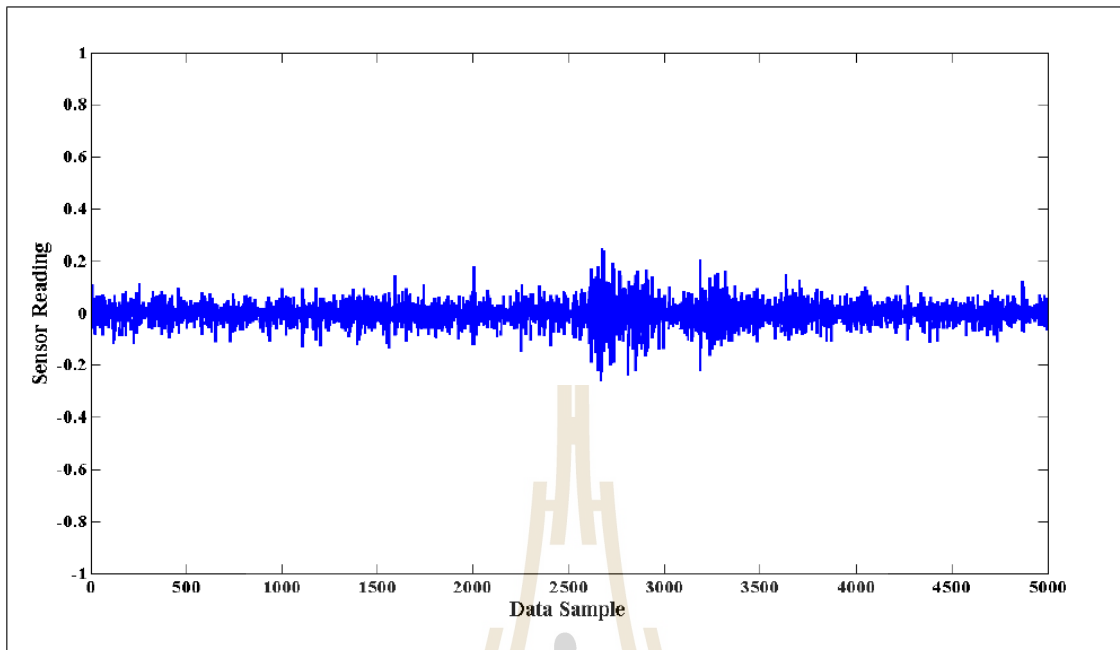
รูปที่ ข.76 DWT สัมประสิทธิ์ผ่านต่ำของชุดข้อมูล NAMOS (อุณหภูมิที่อ่านได้จากเซ็นเซอร์ 1)



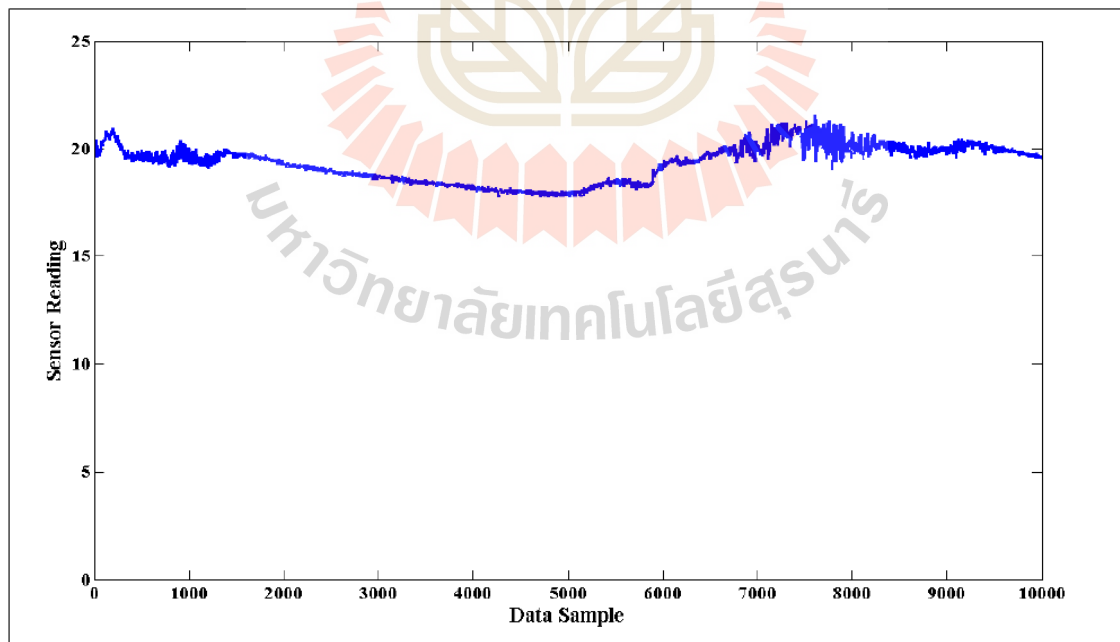
รูปที่ ข.77 DWT สัมประสิทธิ์ผ่านสูงของชุดข้อมูล NAMOS (อุณหภูมิที่อ่านได้จากเซ็นเซอร์ 1)



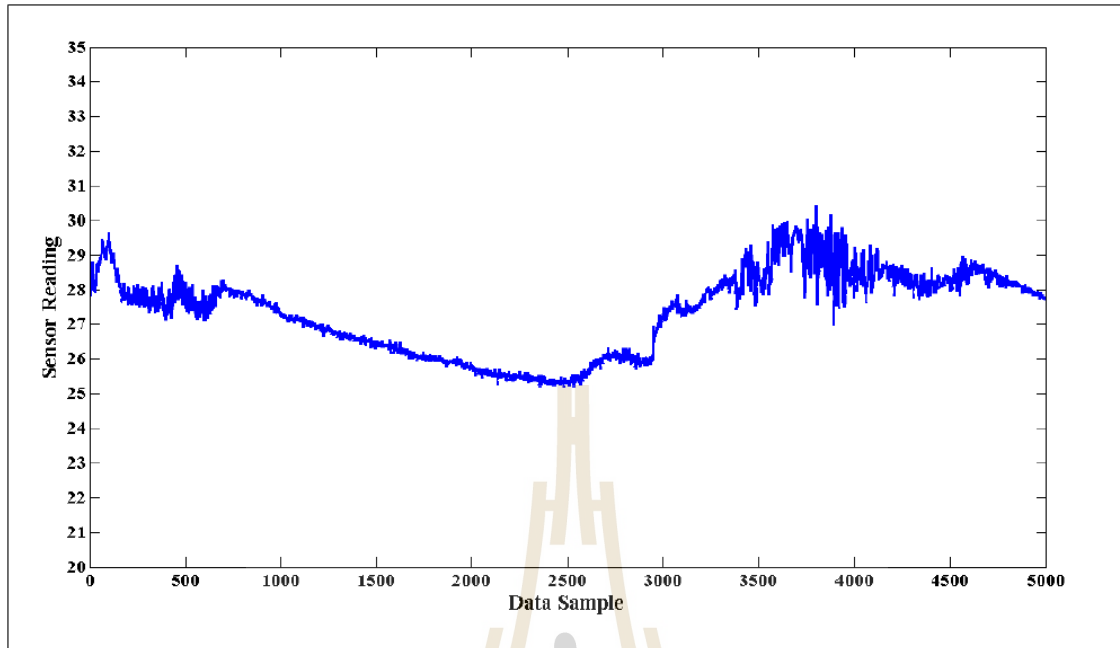
รูปที่ ข.78 LWT สัมประสิทธิ์ผ่านต่ำของชุดข้อมูล NAMOS (อุณหภูมิที่อ่านได้จากเซ็นเซอร์ 1)



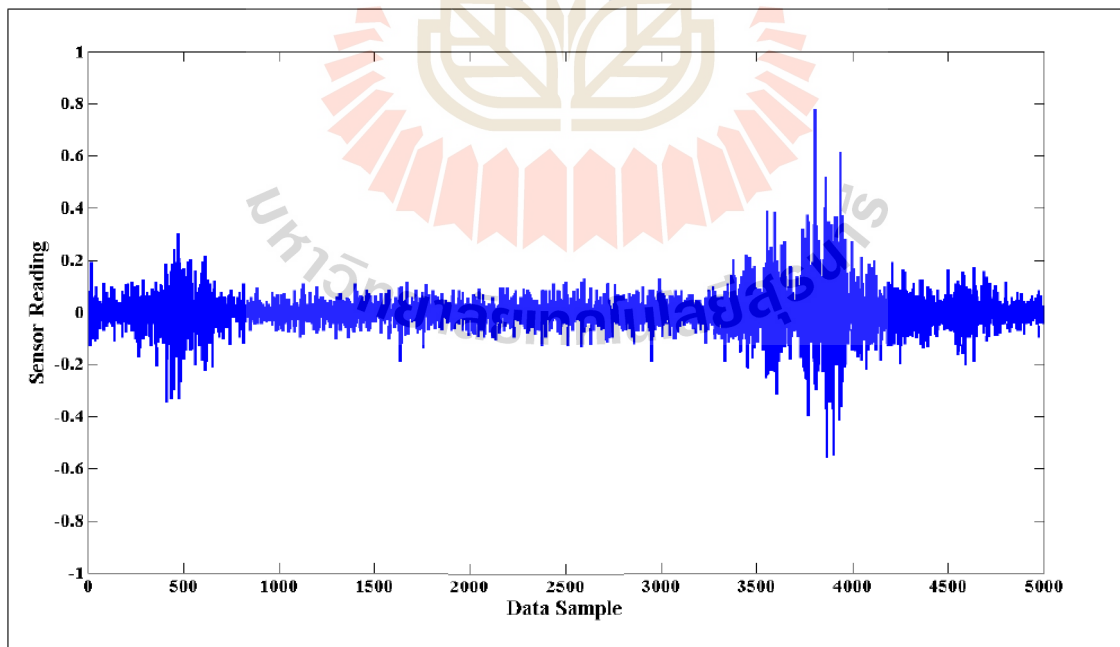
รูปที่ ข.79 LWT สัมประสิทธิ์ผ่านสูงของชุดข้อมูล NAMOS (อุณหภูมิที่อ่านได้จากเซ็นเซอร์ 1)



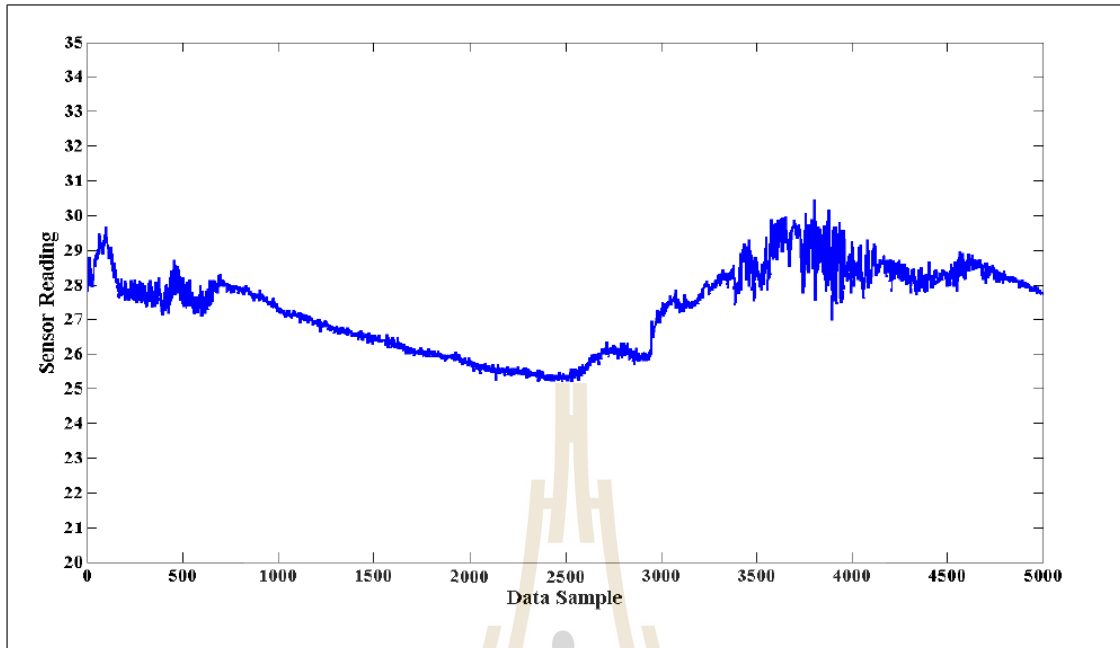
รูปที่ ข.80 ชุดข้อมูล NAMOS (อุณหภูมิที่อ่านได้จากเซ็นเซอร์ 2)



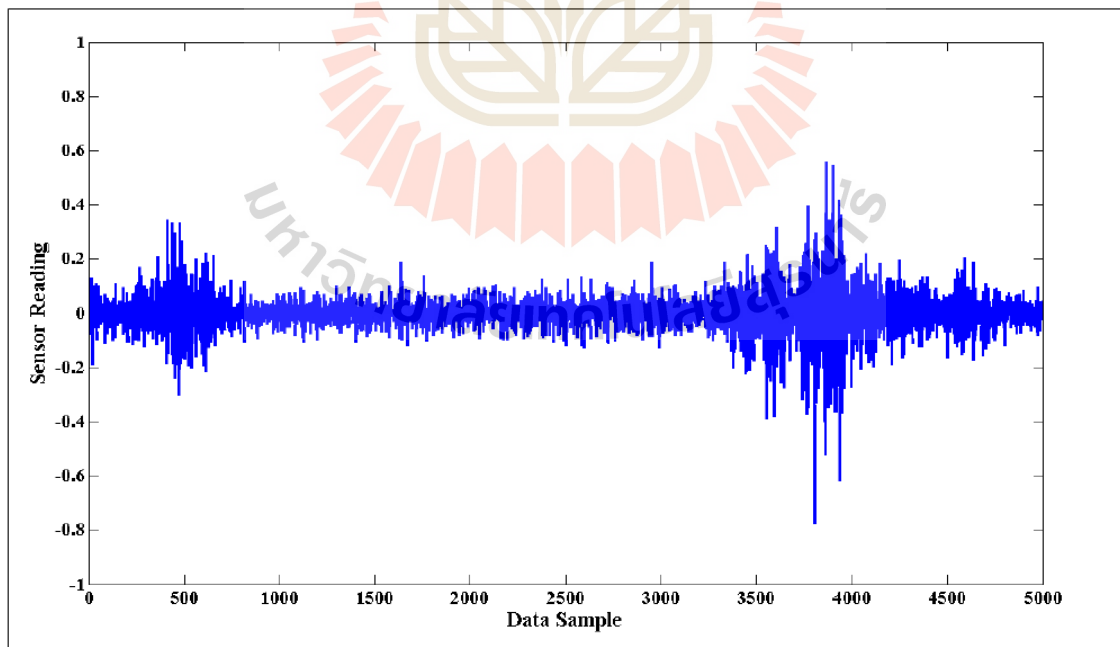
รูปที่ ข.81 DWT สัมประสิทธิ์ผ่านต่ำของชุดข้อมูล NAMOS (อุณหภูมิที่อ่านได้จากเซ็นเซอร์ 2)



รูปที่ ข.82 DWT สัมประสิทธิ์ผ่านสูงของชุดข้อมูล NAMOS (อุณหภูมิที่อ่านได้จากเซ็นเซอร์ 2)



รูปที่ ข.83 LWT สัมประสิทธิ์ผ่านต่ำของชุดข้อมูล NAMOS (อุณหภูมิที่อ่านได้จากเซ็นเซอร์ 2)



รูปที่ ข.84 LWT สัมประสิทธิ์ผ่านสูงของชุดข้อมูล NAMOS (อุณหภูมิที่อ่านได้จากเซ็นเซอร์ 2)

ภาคผนวก ค

ผลกระทบจากการทำให้อยู่ในรูปแบบมาตรฐาน



ผลกระทบจากการนอร์มัลไลเซชัน

ข้อมูลสังเคราะห์ในบทความวิจัยของ (Siripanadorn, et al., 2010a; 2010b) ถูกทำการนอร์มัลไลเซชัน (Normalization) หรือ กระบวนการปรับปรุงโครงสร้างข้อมูลของฐานข้อมูลที่มีความซ้ำซ้อนให้อยู่ในรูปแบบที่เป็นบรรทัดฐานเดียวกัน โดยใช้สมการ (ค.1) ข้อมูลสังเคราะห์ถูกทำให้อยู่ในรูปแบบการแจกแจงแบบปกติ ประกอบด้วยค่ากลางเท่ากับ 0 และค่าเบี่ยงเบนมาตรฐานเท่ากับ 1

$$\frac{x_k - \text{mean}(x_k)}{\sqrt{\text{var}(x_k)}} \quad (\text{ค.1})$$

เมื่อ k คือค่าดัชนีของ KPI และ x_k คือเวกเตอร์คอลัมของข้อมูลใน k th KPI

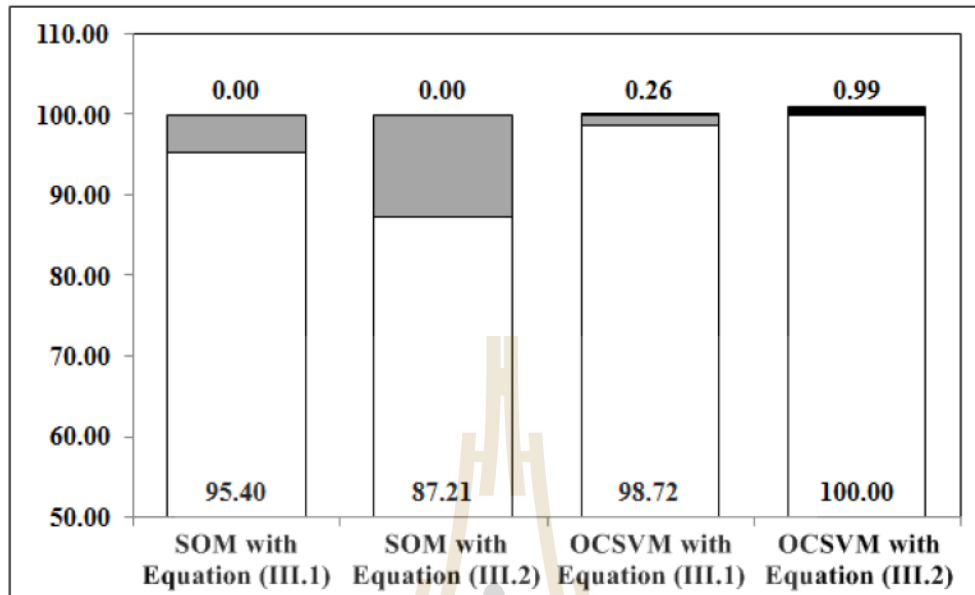
ข้อมูลสังเคราะห์ในบทความวิจัยนี้สร้างขึ้นจากการผสมกันของการแจกแจงแบบเกาส์เซียนด้วยค่ากลางที่สุ่มจาก (0.3, 0.35, 0.45) และค่าเบี่ยงเบนมาตรฐาน 0.03 และถูกทำให้อยู่ในรูปแบบมาตรฐานในช่วง [0,1] โดยใช้สมการ (ค.2)

$$\frac{x_k - \min(x_k)}{\max(x_k) - \min(x_k)} \quad (\text{ค.2})$$

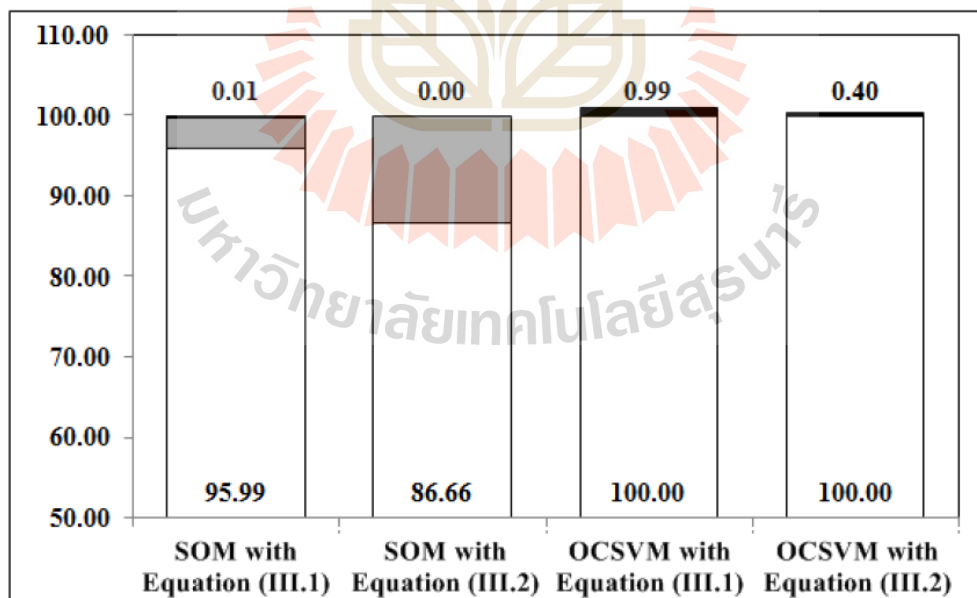
เพื่อที่จะศึกษาถึงผลกระทบจากการนอร์มัลไลเซชันในกระบวนการตรวจจับความผิดปกติของข้อมูล ข้อมูลสังเคราะห์จะถูกสร้างขึ้นจากการผสมกันของการแจกแจงแบบเกาส์เซียนด้วยค่ากลางที่สุ่มจากค่า (0.3, 0.35, 0.45) และค่าเบี่ยงเบนมาตรฐาน 0.03 และถูกทำให้อยู่ในรูปนอร์มัลไลซ์ด้วยสมการที่กล่าวมาข้างต้น ซึ่งถูกเพิ่มเข้าไปในกระบวนการตรวจจับความผิดปกติของข้อมูล

รูปที่ ค.1, ค.2 แสดงผลกระทบการนอร์มัลไลเซชัน ของการตรวจสอบความผิดปกติโดยใช้ชุดข้อมูลสังเคราะห์ที่มีความผิดพลาด 1/80 และ 20/4 ตามลำดับ สำหรับอัลกอริทึม SOM ทำการนอร์มัลไลเซชัน ด้วยค่าการแจกแจงปกติ $N(0,1)$ โดยใช้สมการ (ค.1) ซึ่งให้ค่า DR ที่ดีว่าการนอร์มัลไลเซชัน ในช่วง [0,1] โคนโดยใช้สมการ (ค.2) ในทางตรงกันข้าม OCSVM แสดงออกถึงค่าที่ดีเท่ากันทั้งสองแบบของการนอร์มัลไลซ์ นอกจากนี้การนอร์มัลไลเซชัน ข้อมูลสังเคราะห์ SOM ด้วย $N(0,1)$ ยังแสดงสมรรถนะที่ดีเท่ากับ OCSVM อีกด้วย เพราะฉะนั้นอัลกอริทึม SOM จึงเหมาะสมสำหรับชุดข้อมูลที่มีการกระจายตัวแบบปกติ อย่างไรก็ตามชุดข้อมูลอาจจะไม่มีการแจกแจงแบบปกติซึ่งจะส่งผลกระทบต่อประสิทธิภาพการทำงานของ SOM เพราะฉะนั้นในงานวิจัยนี้หลักการนอร์มัลไลเซชัน ไม่ได้คำนึงถึงข้อมูลที่กระจายตัว เช่นในสมการ (ค.2) ได้ถูกเลือกให้ประมวลผลข้อมูลก่อนที่จะมีการตรวจสอบความผิดปกติ

□ Detection Rate (%DR) □ Miss Alarm Rate ■ False Positive Rate (%FPR)



รูปที่ ค.1 ผลกระทบการนอร์มัลไลเซชันในชุดข้อมูลสังเคราะห์กับความผิดพลาด 1/80



รูปที่ ค.2 ผลกระทบการนอร์มัลไลเซชันในชุดข้อมูลสังเคราะห์กับความผิดพลาด 20/4

ประวัตินักวิจัย

วิภาวี หัตถกรรม จบการศึกษาระดับปริญญาตรี ด้วย B.Eng (Hons) ด้านวิศวกรรมไฟฟ้า (สื่อสาร) จากสถาบันเทคโนโลยีนานาชาติสิรินธร มหาวิทยาลัยธรรมศาสตร์ ระดับปริญญาโท M.Sc. และ ปริญญาเอก Ph.D. ด้าน Communication and Signal Processing จาก Department of Electrical and Electronic Engineering, Imperial College London ประเทศอังกฤษ ปัจจุบัน ดำรงตำแหน่งเป็น ผู้ช่วยศาสตราจารย์ ประจำสาขาวิชาวิศวกรรมโทรคมนาคม สำนักวิชาวิศวกรรมศาสตร์ มหาวิทยาลัยเทคโนโลยีสุรนารี 111 ถนนมหาวิทยาลัย อำเภอเมือง จังหวัดนครราชสีมา 30000 โทร. (044)224392 โทรสาร (044) 224603

