

**การประยุกต์ใช้เทคนิคการแยกข้อมูลเชิงประจักษ์ร่วมกับวิธีบ็อก-เจนกินส์และ
วิธีซัพพอร์ตเวกเตอร์รีเกรสชันในการพยากรณ์ข้อมูลอนุกรมเวลา**
**Application of Empirical Mode Decomposition with Box – Jenkins and
Support Vector Regression for Time Series Forecasting**

ธนกร สุทธิสนธิ* และสมพร เทพฉิม

Thanakorn Sutthison* and Somporn Thepchim

สาขาวิชาสถิติประยุกต์ คณะวิทยาศาสตร์ มหาวิทยาลัยราชภัฏอุบลราชธานี

Program of Applied Statistics, Faculty of Science, Ubon Ratchathani Rajabhat University

*Email: thanakon.sutthison@gmail.com

Received: 03 Oct 19

Revised: 13 Nov 19

Accepted: 20 Jan 20

บทคัดย่อ

งานวิจัยนี้ ผู้วิจัยได้พัฒนาตัวแบบผสมระหว่างตัวแบบ SARIMA และตัวแบบซัพพอร์ตเวกเตอร์รีเกรสชันขึ้นเพื่อใช้พยากรณ์ปริมาณหน่วยจำหน่ายไฟฟ้ารายเดือนประเภทบ้านอยู่อาศัยของประชาชน โดยเก็บรวบรวมข้อมูลจากการไฟฟ้าส่วนภูมิภาคตั้งแต่เดือนมกราคม 2551 ถึงเดือนพฤษภาคม 2562 จำนวน 137 ค่า ผู้วิจัยประยุกต์ใช้เทคนิคการแยกข้อมูลเชิงประจักษ์เพื่อลดสัญญาณแกว่งไว (Fast Oscillation) ของข้อมูลก่อนนำไปสร้างตัวแบบ โดยการสร้างตัวแบบผสมจะอาศัยจุดเด่นของตัวแบบแต่ละตัว คือ ตัวแบบ SARIMA มีจุดเด่นในการอธิบายข้อมูลอนุกรมเวลาที่มีความสัมพันธ์เชิงเส้นตรงได้ดี ส่วนตัวแบบซัพพอร์ตเวกเตอร์รีเกรสชันมีจุดเด่นในการสร้างฟังก์ชันการทำนายที่ซับซ้อนและอธิบายข้อมูลอนุกรมเวลาที่มีความสัมพันธ์ไม่เป็นเชิงเส้นตรงได้ดี จากนั้นทำการเปรียบเทียบประสิทธิภาพการพยากรณ์กับตัวแบบ 3 ตัวแบบ คือ ตัวแบบอย่างง่าย ตัวแบบ SARIMA และตัวแบบซัพพอร์ตเวกเตอร์รีเกรสชัน ด้วยเกณฑ์การวัดประสิทธิภาพการพยากรณ์ 5 เกณฑ์ ได้แก่ ความคลาดเคลื่อนสัมบูรณ์เฉลี่ย รากของค่าคลาดเคลื่อนกำลังสองเฉลี่ย ร้อยละของความคลาดเคลื่อนสัมบูรณ์เฉลี่ย เปอร์เซนต์ของมีฐานความคลาดเคลื่อนสัมบูรณ์ และเปอร์เซนต์ของความคลาดเคลื่อนสัมบูรณ์เฉลี่ยแบบมาตรฐาน ผลการวิจัยบ่งชี้ว่า ตัวแบบผสมให้ประสิทธิภาพการพยากรณ์ดีกว่าตัวแบบอนุกรมเวลาเชิงเดี่ยว 3 ตัวแบบ ในทุกเกณฑ์การวัดประสิทธิภาพการพยากรณ์ ดังนั้นตัวแบบผสมสามารถใช้เป็นเครื่องมือในการพยากรณ์ข้อมูลอนุกรมเวลาหน่วยจำหน่ายไฟฟ้ารายเดือนประเภทบ้านอยู่อาศัยของประชาชนได้อย่างเหมาะสมและสามารถใช้ประกอบการตัดสินใจสำหรับการวางแผนการผลิตหรือจำหน่ายไฟฟ้าให้เพียงพอต่อความต้องการของประชาชนในอนาคตได้

คำสำคัญ: ตัวแบบผสม ตัวแบบอย่างง่าย ตัวแบบ SARIMA ตัวแบบซัพพอร์ตเวกเตอร์รีเกรสชัน การแยกข้อมูลเชิงประจักษ์

Abstract

In this research, the researcher developed the hybrid model of the SARIMA model and the support vector regression model in order to forecast the amount of monthly electricity distribution units for residential houses. The data of 137 values had been collected from the Provincial Electricity Authority from January 2008 to May 2019. The researcher applied the empirical mode decomposition to reduce the fast oscillation of the data. The creation of the hybrid model was based on the strengths of each model. The SARIMA model had a strong point in describing time series data with a linear relationship, and the support vector regression model had a strong point in creating complex predictive functions and describing time series data with a non-linear relationship. After that, the forecasting efficiency was compared in 3 models: the naïve model, the SARIMA model, and the support vector regression model using 5 forecasting performance criteria, namely Mean Absolute Error, Root Mean Square Error, Mean Absolute Percentage Error, Median Absolute Percentage Error and Symmetric Mean Absolute Percentage Error. The results indicated the hybrid model provided better forecasting performance than the 3 Univariate Time Series models in all forecasting performance criteria. Therefore, the hybrid model can be used as a suitable tool for forecasting the

monthly electricity distribution units of the residential houses time series data. It can be used for decision in electricity generation and supply to served the consumer's demand in the future.

Keywords: hybrid model, naïve model, SARIMA model, support vector regression model, Empirical mode decomposition

1. บทนำ

ปัจจุบันนักวิจัยจำนวนมากให้ความสำคัญเกี่ยวกับการพยากรณ์โดยมีการประยุกต์ใช้ในการแก้ปัญหาต่าง ๆ เช่น การพยากรณ์การใช้พลังงานไฟฟ้า [1], [2], [3], [4] การพยากรณ์ความเร็วลม [5],[6] การพยากรณ์ความต้องการด้านการท่องเที่ยว [7] การพยากรณ์ความต้องการด้านอุตสาหกรรมการบิน [8] การพยากรณ์คุณภาพของอากาศ [9] เป็นต้น โดยมีการแบ่งการพยากรณ์อนุกรมเวลาออกเป็น 2 กลุ่มใหญ่ คือ (1) วิธีการพยากรณ์แบบเชิงเส้นตรง (Linear approach) วิธีการพยากรณ์ที่นิยม ได้แก่ วิธีของบ็อก-เจนกินส์ตัวแบบที่ใช้คือ ตัวแบบ ARIMA และ SARIMA เป็นวิธีการพยากรณ์ที่มีประสิทธิภาพใช้ได้ในการพยากรณ์อนุกรมเวลาที่เป็นเชิงเส้นตรงข้อเสีย คือ ไม่สามารถอธิบายอนุกรมเวลาที่ซับซ้อนหรือมีความสัมพันธ์ไม่เป็นเชิงเส้นตรงได้ ทำให้ค่าพยากรณ์มีความคลาดเคลื่อนค่อนข้างมาก และ (2) วิธีการพยากรณ์แบบไม่เป็นเชิงเส้นตรง (Non -Linear approach) วิธีที่นิยมใช้ ได้แก่ วิธีโครงข่ายประสาทเทียม (Artificial Neural Network) และ วิธีซัพพอร์ตเวกเตอร์รีเกรสชัน (Support vector regression model: SVR) เป็นวิธีการที่มีความยืดหยุ่นสูงใช้ได้กับข้อมูลอนุกรมเวลาที่ไม่เป็นเชิงเส้นตรง แต่มีข้อเสีย คือ ไม่สามารถวิเคราะห์ข้อมูลอนุกรมเวลาที่เป็นเชิงเส้นตรงได้ โดยเฉพาะข้อมูลที่มีลักษณะการเปลี่ยนแปลงตามฤดูกาลและแนวโน้ม [10] จากวิธีการพยากรณ์ทั้ง 2 กลุ่มข้างต้น หากผู้พยากรณ์ประยุกต์ใช้วิธีการพยากรณ์อนุกรมเวลาเชิงเดียวด้วยวิธีการใดวิธีการหนึ่ง อาจไม่เพียงพอและไม่สามารถอธิบายความสัมพันธ์ของข้อมูลอนุกรมเวลาที่ซับซ้อนได้ครบถ้วน จึงมีแนวคิดในการพัฒนาตัวแบบผสมขึ้นโดยอาศัยลักษณะเด่นของแต่ละตัวแบบเพื่ออธิบายลักษณะความสัมพันธ์ของข้อมูลที่ต้องการศึกษาให้สามารถเรียนรู้และอธิบายคุณลักษณะของข้อมูลอนุกรมเวลาได้ชัดเจนยิ่งขึ้น รวมทั้งลดความเสี่ยงของการใช้ตัวแบบที่ไม่เหมาะสมกับข้อมูลที่ต้องการศึกษาได้ [11] จากการทบทวนวรรณกรรมที่เกี่ยวข้อง พบว่า ระหว่างการพยากรณ์อนุกรมเวลาเชิงเดียวและการพยากรณ์ด้วยตัวแบบผสม [12] การพยากรณ์ด้วยตัวแบบผสมจะให้ความแม่นยำมากกว่าการพยากรณ์ด้วยตัวแบบอนุกรมเวลาเชิงเดียว ด้วยเหตุที่อนุกรมเวลาในปัจจุบันมีความซับซ้อนไม่สามารถระบุความสัมพันธ์ที่ชัดเจนได้ จึงมีนักวิจัยหลายท่านพยายามปรับโมเดลที่มีประสิทธิภาพมากขึ้นโดยการแยกองค์ประกอบด้วยข้อมูลเชิงประจักษ์ (Empirical mode decomposition:

EMD) เพื่อการลดความถี่หรือสัญญาณแกว่งไวของข้อมูลอนุกรมเวลาที่ไม่เป็นเชิงเส้นและมีลักษณะการเคลื่อนไหวไม่คงที่ (Non-stationary) ก่อนนำไปพยากรณ์ด้วยเทคนิคการพยากรณ์ต่าง ๆ ทั้งอนุกรมเวลาเชิงเดียวและตัวแบบผสม ซึ่งพบว่า อนุกรมเวลาที่มีการแยกองค์ประกอบเพื่อลดความถี่หรือสัญญาณแกว่งไวของข้อมูลก่อนนำไปพยากรณ์ให้ค่าความแม่นยำสูงกว่าข้อมูลอนุกรมเวลาที่ไม่มีการแยกองค์ประกอบ [13], [14]

จากเหตุผลดังกล่าวข้างต้นเพื่อเพิ่มความแม่นยำในการพยากรณ์ ผู้วิจัยจึงพัฒนาวิธีการพยากรณ์แบบผสมขึ้น โดยจะประยุกต์ใช้เทคนิคการแยกองค์ประกอบด้วยข้อมูลเชิงประจักษ์ เพื่อการลดความถี่หรือสัญญาณแกว่งไวของข้อมูลอนุกรมเวลาที่ไม่เป็นเชิงเส้นและมีลักษณะการเคลื่อนไหวไม่คงที่ก่อนนำไปพยากรณ์ด้วยตัวแบบผสมระหว่างตัวแบบการพยากรณ์ตามวิธีของบ็อก-เจนกินส์ ด้วยตัวแบบ SARIMA เนื่องจากตัวแบบดังกล่าวมีประสิทธิภาพในการพยากรณ์อนุกรมเวลาที่มีความสัมพันธ์เชิงเส้นตรงได้ดีและมีการประยุกต์ใช้อย่างหลากหลายในปัจจุบัน [15], [16], [17], [18] แต่มีข้อเสียคือ มีข้อกำหนดในการสร้างตัวแบบที่เหมาะสมจำนวนมากประกอบกับข้อมูลอนุกรมเวลาในปัจจุบันมีความสัมพันธ์ที่ซับซ้อนจึงทำให้ค่าพยากรณ์ที่ได้มีความคลาดเคลื่อนสูง และวิธีซัพพอร์ตเวกเตอร์รีเกรสชันเป็นวิธีการพยากรณ์ที่มีความแม่นยำและมีประสิทธิภาพสูง โดยมีจุดเด่นในการสร้างฟังก์ชันการทำนายจากลักษณะข้อมูลที่มีความสัมพันธ์ไม่เป็นเชิงเส้นตรงได้ดีและมีข้อกำหนดไม่มากและเป็นที่ยอมรับในปัจจุบัน [8],[13], [19], [20] ซึ่งเหมาะที่จะนำมาใช้สำหรับการสร้างความสัมพันธ์เพื่ออธิบายลักษณะข้อมูลอนุกรมเวลาที่มีความสัมพันธ์ไม่เป็นเส้นตรงของส่วนเหลือ (residuals) จากตัวแบบ SARIMA [21] โดยตัวแบบผสมระหว่างตัวแบบ SARIMA กับตัวแบบซัพพอร์ตเวกเตอร์รีเกรสชันจะถูกนำมาทดลองใช้กับข้อมูลอนุกรมเวลาหน่วยจำหน่ายไฟฟ้าประเภทบ้านอยู่อาศัยของประเทศไทย และนำไปเปรียบเทียบประสิทธิภาพการพยากรณ์กับตัวแบบการพยากรณ์แบบอย่างง่าย (Naïve model) ตัวแบบ SARIMA ตัวแบบซัพพอร์ตเวกเตอร์รีเกรสชัน และตัวแบบผสมระหว่างตัวแบบ SARIMA ตัวแบบซัพพอร์ตเวกเตอร์รีเกรสชัน ด้วยเกณฑ์วัดประสิทธิภาพ 5 เกณฑ์ ได้แก่ ความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Error : MAE) รากของความคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Square Error :

RMSE) ร้อยละความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error : MAPE) เปอร์เซ็นต์ของมัธยฐานความคลาดเคลื่อนสัมบูรณ์ (Median Absolute Percentage Error: MdAPE) และเปอร์เซ็นต์ของความคลาดเคลื่อนสัมบูรณ์แบบสมมาตร (Symmetric Mean Absolute Percentage Error: sMAPE) เพื่อค้นหาตัวแบบที่เหมาะสมที่สุดสำหรับการพยากรณ์ และผลการวิจัยที่ได้จะเป็นประโยชน์และเป็นแนวทางสำหรับการเพิ่มประสิทธิภาพการพยากรณ์ต่อไป

2. วัสดุ อุปกรณ์ และวิธีการดำเนินการวิจัย/ทดลอง

งานวิจัยนี้มีลำดับขั้นตอนการวิจัยดังนี้

2.1 ข้อมูลที่ใช้ในการทดลอง

ข้อมูลที่ใช้ในการทดลองครั้งนี้เป็นข้อมูลอนุกรมเวลาหน่วยจำหน่ายไฟฟ้าบ้านอยู่อาศัย: ประเภทอัตราค่าไฟฟ้าของการไฟฟ้าส่วนภูมิภาค ซึ่งเป็นข้อมูลทุติยภูมิจากรายงานสถานการณ์การจำหน่ายไฟฟ้าประจำเดือน ตั้งแต่เดือนมกราคม 2551 ถึงเดือนพฤษภาคม 2562 ที่เผยแพร่ในเว็บไซต์ของการไฟฟ้าส่วนภูมิภาค [22] จำนวน 137 ค่า ดังแสดงในรูปที่ 1 ผู้วิจัยได้แบ่งข้อมูลออกเป็น 2 ชุด คือ ข้อมูลชุดที่ 1 ตั้งแต่เดือนมกราคม 2551 ถึงเดือนธันวาคม 2559 จำนวน 108 ค่า ใช้สำหรับสร้างตัวแบบและข้อมูลชุดที่ 2 ตั้งแต่เดือนมกราคม 2560 ถึงเดือนพฤษภาคม 2562 จำนวน 29 ค่า ใช้สำหรับการทดสอบตัวแบบเกณฑ์ในการแบ่งข้อมูลเพื่อวัดประสิทธิภาพการพยากรณ์ (Cross - validation) ข้างต้นใช้เกณฑ์ร้อยละ 80 และ 20 ตามลำดับ

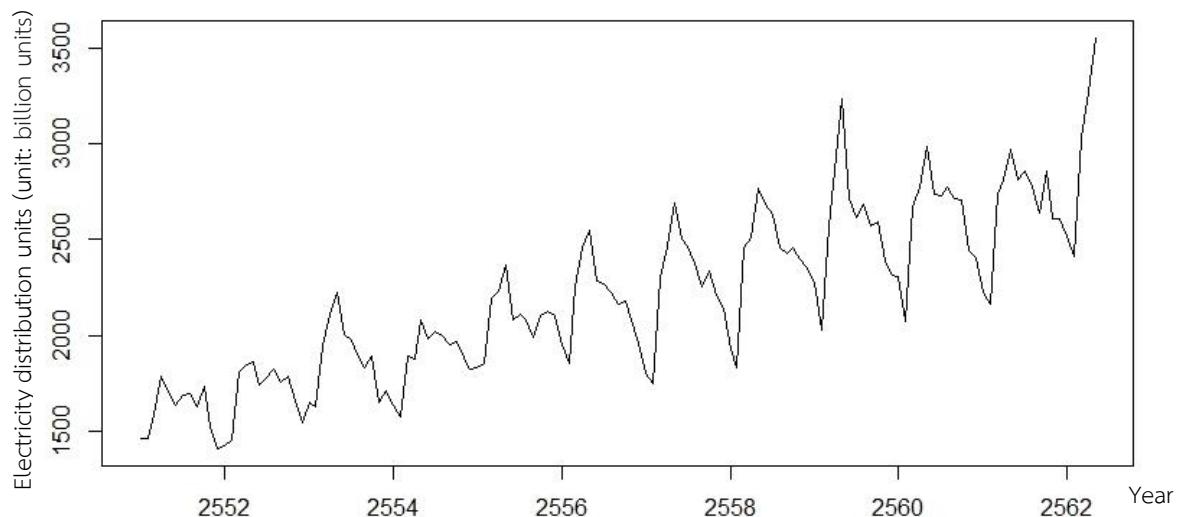


Figure 1 The experimental electricity distribution units for residents' houses time series

2.2 วิธีการวิเคราะห์ข้อมูลอนุกรมเวลาเชิงประจักษ์

วิธีการเชิงประจักษ์ (Empirical mode decomposition : EMD) เป็นการประยุกต์ใช้เพื่อวิเคราะห์ความถี่หรือสัญญาณแกว่งไกวของข้อมูลอนุกรมเวลาที่ไม่เป็นเชิงเส้นและมีลักษณะการเคลื่อนไหวไม่คงที่ (Non-stationary) นำเสนอโดย Huang และคนอื่น ๆ [23] โดยมีสมมติฐานว่าข้อมูลอนุกรมเวลาหนึ่งจะประกอบด้วยสัญญาณที่มีการแกว่งไกว (Fast Oscillation) และสัญญาณที่มีการแกว่งช้า (Slow Oscillation) ดังนั้น เมื่อต้องการแยกองค์ประกอบสัญญาณต้นฉบับให้อยู่ในรูปของชุดฟังก์ชันภายใน (Intrinsic Mode Functions : IMFs) โดยไม่จำเป็นต้องใช้ฟังก์ชันก่อนหน้าใด ๆ (Prior Function) แต่จะใช้การคำนวณเพื่อหาค่าเฉลี่ยของสัญญาณจากข้อมูลจริงของสัญญาณนำเข้า เพื่อกำหนดเป็นเงื่อนไขหรือเกณฑ์ในการแยกองค์ประกอบของสัญญาณ วิธี EMD จึงจัด เป็นการแยก

องค์ประกอบของสัญญาณโดยการขับเคลื่อนด้วยข้อมูล (Data Driven) Huang และคณะตั้งชื่อขั้นตอน และวิธีสำหรับสกัดชุดฟังก์ชันภายใน (IMFs) จากสัญญาณต้นฉบับนี้ว่า Sifting Process และให้คำจำกัดความของชุดข้อมูลฟังก์ชันภายใน (IMFs) ไว้ว่าฟังก์ชันใด ๆ จะจัดเป็นฟังก์ชันภายใน (IMF) ได้ก็ต่อเมื่อสัญญาณนั้น ๆ เป็นไปตามเงื่อนไขดังนี้ 1) จำนวนจุดปลายสุด (Extrema Points) ซึ่งจะประกอบด้วยจุดสูงสุดและต่ำสุดและจำนวนจุดตัดแกนศูนย์จะต้องมีค่าเท่ากันหรือมีค่าต่างกันไม่เกินหนึ่ง 2) ณ จุดใด ๆ ของสัญญาณค่าเฉลี่ยของขอบบนและขอบล่างจะมีค่าเท่ากับศูนย์กระบวนการทำงานเพื่อสกัดชุดฟังก์ชันภายใน (IMFs) ของการแยกองค์ประกอบด้วยวิธีเชิงประจักษ์ (EMD) แสดงดังรูปที่ 2 ประกอบด้วย 2 ขั้นตอนคือ

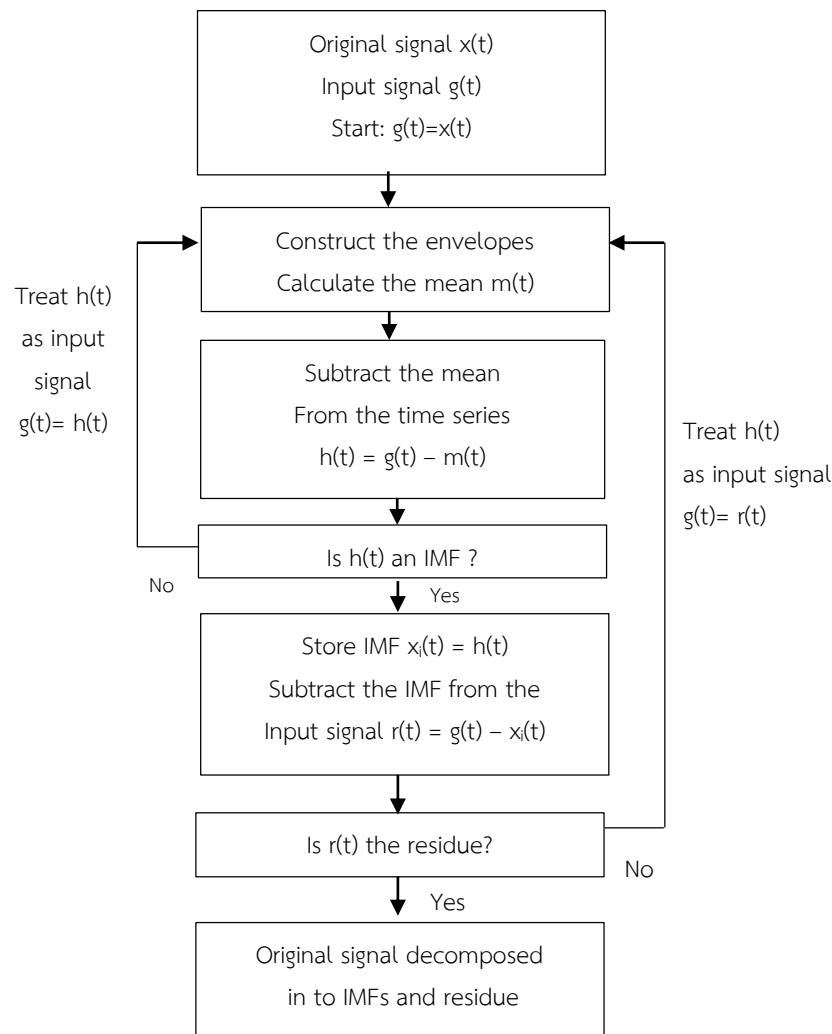


Figure 2 Flow chart of Empirical mode decomposition.

1) การตรวจจับจุดปลายสุดของสัญญาณ (Extrema Detection) โดยการหาจุดสูงสุดและจุดต่ำสุดของแต่ละช่วงสัญญาณย่อย

2) การประมาณค่าขอบแบบพื้นผิว (Surface Interpolation) ของสัญญาณ โดยการหาค่าประมาณของสัญญาณขอบบนและสัญญาณขอบล่าง แล้วคำนวณหาค่าสัญญาณเฉลี่ยตามสมการที่ (1) จากนั้นนำสัญญาณต้นฉบับลบออกจากสัญญาณเฉลี่ยจะได้สัญญาณผลลัพธ์ ขั้นตอนต่อไปตรวจสอบว่า สัญญาณผลลัพธ์นั้นเป็นฟังก์ชันภายใน (IMF) หรือไม่ ถ้าผลปรากฏว่าสัญญาณผลลัพธ์ที่ได้ไม่เป็นฟังก์ชันภายใน (IMF) ให้แทนที่สัญญาณต้นฉบับด้วยสัญญาณผลลัพธ์ แล้วทำซ้ำกระบวนการดังกล่าวข้างต้น หากสัญญาณผลลัพธ์ที่ได้เป็นฟังก์ชันภายใน (IMF) ให้จัดเก็บสัญญาณผลลัพธ์นั้น จากนั้นทำการสกัดฟังก์ชันภายใน (IMF) ลำดับถัดไป จากสัญญาณต้นฉบับใหม่ที่ได้จากการนำสัญญาณต้นฉบับเดิมลบด้วย

ฟังก์ชันภายใน (IMF) แล้วทำกระบวนการข้างต้น ซ้ำจนสัญญาณต้นฉบับเป็นสัญญาณฟังก์ชันเดียว

$$m(t) = \frac{e_{\max} + e_{\min}}{2} \quad (1)$$

เมื่อ $m(t)$ แทน สัญญาณเฉลี่ย $e_{\max}$, $e_{\min}$ สัญญาณขอบบน และสัญญาณขอบล่าง ตามลำดับ และสามารถเขียนให้อยู่ในรูปสมการได้ดังนี้

$$x(t) = \sum_{i=1}^n h_i(t) + r_N(t) \quad (2)$$

เมื่อ $x(t)$ แทนสัญญาณต้นฉบับ $h_i(t)$,
 $i=1,2,3,...,N$ เป็นชุดฟังก์ชันภายใน (IMFs) และ
 $r_N(t)$ เป็นส่วนเหลือ (residual)

2.3 ตัวแบบอย่างง่าย

ตัวแบบอย่างง่ายเป็นแบบที่มีฟังก์ชันความสัมพันธ์ไม่ยุ่งยากและง่ายต่อการประยุกต์ใช้ โดยอาศัยข้อมูลที่เราพบในปัจจุบันเป็นค่าพยากรณ์ของช่วงเวลาถัดไป ซึ่งสามารถด้วยตัวแบบทางคณิตศาสตร์ดังสมการที่ (3)

$$Y_{t+1} = Y_t + \varepsilon_{t+1} \quad (3)$$

เมื่อ Y_{t+1} คือ ค่าพยากรณ์ของช่วงเวลาถัดไปหนึ่งหน่วย
 เวลา ณ เวลาที่ t+1

Y_t คือ ค่าข้อมูล ณ เวลาที่ t

ε_{t+1} คือ ค่าความคลาดเคลื่อน ณ เวลาที่ t+1

2.4 ตัวแบบการพยากรณ์โดยวิธีของบ็อก-เจนกินส์

ตัวแบบการพยากรณ์ตามวิธีของบ็อก-เจนกินส์ เป็นวิธีการพยากรณ์หนึ่งที่มีความถูกต้องแม่นยำสูง เนื่องจากมีการพิจารณาลักษณะของอนุกรมเวลาว่ามีสหสัมพันธ์ในตนเอง (Autocorrelation Function: ACF) และอัตสหสัมพันธ์ในตนเองบางส่วน (Partial Autocorrelation Function: PACF) เป็นอย่างไร เพื่อใช้ในการกำหนดตัวแบบการพยากรณ์ เช่น ตัวแบบ AR(1) ค่าสหสัมพันธ์ในตนเอง (ACF) มีลักษณะที่มีค่าลดลงอย่างรวดเร็วแบบเอ็กโปเนนเชียลโดยมีเครื่องหมายสลับกัน ($\phi < 0$) แต่ฟังก์ชันสหสัมพันธ์ในตนเองบางส่วน (PACF) จะมีค่าไม่เท่ากับ 0 เพียงค่าเดียว คือ ช่วงเวลาห่างกัน 1 หน่วยเวลานั้น นอกจากค่านี้แล้วสหสัมพันธ์ในตนเองส่วนย่อยจะมีค่าเป็น 0 เสมอ เมื่อ $k > 0$ [24] ดังรูปที่ 3 จะเห็นได้ว่าการพยากรณ์ข้อมูลอนุกรมเวลาตามวิธีของบ็อก-เจนกินส์ จะพิจารณาจากรูปแบบหลัก 2 รูปแบบ ได้แก่ Autoregressive (AR) คือ รูปแบบที่กำหนดว่าค่าพยากรณ์ที่เวลาใด ๆ ขึ้นอยู่กับค่าสังเกตก่อนหน้าและรูปแบบ Moving average (MA) คือ รูปแบบที่กำหนดว่าค่าพยากรณ์ที่เวลาใด ๆ ขึ้นอยู่กับความคลาดเคลื่อนก่อนหน้าและเงื่อนไขสำคัญของการใช้ตัวแบบบ็อก-เจนกินส์ คือ อนุกรมเวลาที่นำมาพยากรณ์จะต้องมีสถานะคงที่ (Stationary) กล่าวคือ ค่าเฉลี่ยคงที่และค่าความแปรปรวนคงที่ อย่างไรก็ตามหากข้อมูลอนุกรมเวลาที่นำมาวิเคราะห์มีฤดูกาลเป็นส่วนประกอบเราจะต้องคำนึงถึงความสัมพันธ์ตามฤดูกาลซึ่งเป็นส่วนประกอบที่สำคัญของอนุกรมเวลาร่วมด้วยโดยมีตัวแบบทั่วไป คือ SARIMA (p,d,q) (P,D,Q)s [25] แสดงดังสมการที่ (4) – (9)

$$\phi_p(B)\Phi_p(B^s)(1-B)^d(1-B^s)^D Y_t = \delta + \theta_q(B)\Theta_q(B^s)\varepsilon_t \quad (4)$$

$$\delta = \mu\phi_p(B)\Phi_p(B^s) \quad (5)$$

$$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \quad (6)$$

$$\Phi_p(B^s) = 1 - \phi_1 B^s - \phi_2 B^{2s} - \dots - \phi_p B^{ps} \quad (7)$$

$$\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q \quad (8)$$

$$\Theta_q(B^s) = 1 - \theta_1 B^s - \theta_2 B^{2s} - \dots - \theta_q B^{qs} \quad (9)$$

เมื่อ Y_t แทน อนุกรมเวลา ณ เวลา t และ ε_t คือ ความคลาดเคลื่อน ซึ่งมีการแจกแบบปกติ มีค่าเฉลี่ยเท่ากับ ศูนย์ มีความแปรปรวนคงที่และเป็นอิสระกัน ($\varepsilon_t \sim NID(0, \sigma^2)$) และ t แทน คาบเวลา ซึ่งมีค่าตั้งแต่ 1 ถึง n, s แทนความยาวของคาบฤดูกาลใน 1 ฤดูกาล, d และ D แทน ลำดับที่ของการหาผลต่างภายในฤดูกาลและผลต่างระหว่างฤดูกาล, B แทน ตัวดำเนินการถอยหลัง (Backward operator) โดยที่ $B^s Y_t = Y_{t-s}$ และ $BY_t = Y_{t-1}$, $\phi_p(B)$ แทนตัวดำเนินการสหสัมพันธ์ในตนเองแบบไม่มีฤดูกาลอันดับที่ p, $\Phi_p(B^s)$ แทน ตัวดำเนินการสหสัมพันธ์ในตัวเองแบบมีฤดูกาลอันดับที่ P, $\theta_q(B)$ แทน ตัวดำเนินการเฉลี่ยเคลื่อนที่แบบภายในฤดูกาลอันดับที่ q, $\Theta_q(B^s)$ แทน ตัวดำเนินการเฉลี่ยเคลื่อนที่ระหว่างฤดูกาลอันดับที่ Q

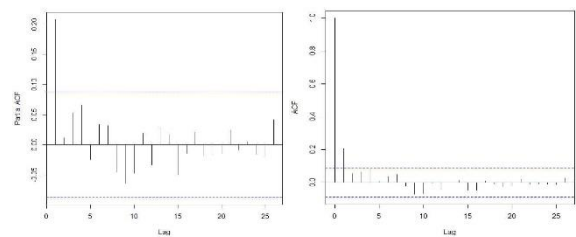


Figure 3 Graph of autocorrelation function (ACF) and partial autocorrelation function (PACF) for AR(1) model

การสร้างตัวแบบ SARIMA ต้องวิเคราะห์อนุกรมเวลาเพื่อกำหนดรูปแบบ SARIMA (p,d,q) (P,D,Q), การดำเนินการผู้วิจัยเขียนคำสั่งโดยใช้โปรแกรมภาษาอาร์มีฟังก์ชัน auto.arima() ซึ่งอยู่ใน Package “forecast” [26] ที่ใช้ในการระบุรูปแบบ ARIMA (p,d,q) (P,D,Q), ที่เหมาะสม ขั้นตอนการสร้างตัวแบบคือ 1) พิจารณาอนุกรมเวลาว่าคงที่หรือไม่ โดยพิจารณาจากกราฟของอนุกรมเวลาชุดนั้น หากกราฟอนุกรมเวลาชุดนั้นมี

แนวโน้มเป็นส่วนประกอบกล่าวคือ ข้อมูลลักษณะเพิ่มขึ้นหรือลดลงอย่างเป็นสัดส่วนกับเวลาและมีการกระจายไม่สม่ำเสมอ มีลักษณะบานออกเป็นปากแตร แสดงว่าข้อมูลอนุกรมเวลาชุดนั้นมีสถานะไม่คงที่นั่นคือมีค่าเฉลี่ยไม่คงที่และความแปรปรวนไม่คงที่งานวิจัยนี้จะทำการทดสอบด้วย Augmented Dickey – Fuller (ADF) หากผลการทดสอบไม่มีนัยสำคัญทางสถิติแสดงว่าข้อมูลอนุกรมเวลาชุดนั้นมีลักษณะไม่คงที่ เมื่อทราบว่าข้อมูลมีลักษณะไม่คงที่จะต้องทำการแปลงข้อมูลให้มีลักษณะคงที่ โดยการหาลอการิทึมของข้อมูลก่อนแล้วจึงหาผลต่างของข้อมูลชุดดังกล่าว เพื่อให้ข้อมูลอนุกรมเวลามีลักษณะคงที่ก่อนที่จะนำไปกำหนดตัวแบบ 2) กำหนดตัวแบบพยากรณ์เนื่องจากข้อมูลอนุกรมเวลาหน่วยจำหน่ายไฟฟ้ามีแนวโน้มและฤดูกาลเป็นส่วนประกอบ ดังนั้นตัวแบบที่เหมาะสมในการพยากรณ์คือ ตัวแบบ SARIMA การใช้ตัวแบบจะต้องระบุอันดับ (p,d,q) และอันดับ (P,D,Q)_s ในงานวิจัยนี้ผู้วิจัยใช้โปรแกรมภาษาอาร์ซึ่งมีฟังก์ชัน auto.arima () ซึ่งอยู่ใน Package “forecast” สำหรับกำหนดอันดับ (p,d,q) และอันดับ (P,D,Q)_s ที่เหมาะสมต่อไป 3) ประมาณค่าพารามิเตอร์และตรวจสอบความเหมาะสมของตัวแบบ เมื่อได้ตัวแบบที่เหมาะสมก่อนที่จะนำตัวแบบไปใช้ในการพยากรณ์จะต้องมีการวินิจฉัยตัวแบบ (Diagnostics Checking) ซึ่งมีทั้งหมด 4 ข้อ คือ 3.1) ทดสอบความคลาดเคลื่อนมีค่าเฉลี่ยเท่ากับศูนย์ โดยใช้สถิติทดสอบ t-test 3.2) ทดสอบความคลาดเคลื่อนมีการแจกแจงแบบปกติ โดยใช้สถิติทดสอบ Kolmogorov – Smirnov 3.3) ทดสอบความเป็นอิสระกันของความคลาดเคลื่อนโดยใช้สถิติทดสอบ Ljung – Box Q- statistics 3.4) ความคลาดเคลื่อนมีความแปรปรวนคงที่โดยพิจารณาจากการพล็อตกราฟดูลักษณะการกระจายระหว่างความคลาดเคลื่อนกับค่าพยากรณ์หากกราฟมีการกระจายไม่มีรูปแบบแสดงว่าความคลาดเคลื่อนมีความแปรปรวนคงที่ 4) นำตัวแบบที่เหมาะสมไปใช้ในการพยากรณ์ล่วงหน้าต่อไป

2.5. ซัพพอร์ตเวกเตอร์รีเกรสชัน

ซัพพอร์ตเวกเตอร์รีเกรสชัน (Support vector Regression : SVR) เป็นวิธีการพยากรณ์วิธีหนึ่งที่มีความแม่นยำสูงและเป็นที่ยอมรับอย่างมากและมีการใช้อย่างแพร่หลาย [8], [13], [19] เสนอโดย Vapnik ในปี ค.ศ. 1995 ซึ่งเป็นการดัดแปลงมาจากวิธีซัพพอร์ตเวกเตอร์แมชชีน โดยใช้สมการไฮเปอร์เพลนเป็นสมการพยากรณ์ข้อมูลที่เป็นเลขจำนวนจริง

หลักการคือจะลดความเสี่ยงเชิงโครงสร้างให้มีค่าต่ำที่สุด ด้วยกระบวนการปรับรูปแบบของข้อมูลที่มีมิติต่ำ (Low dimension data set) บนพื้นที่ข้อมูลนำเข้า (Input space) ให้อยู่ในรูปแบบของข้อมูลที่มีมิติสูง (High dimension data set) บนพื้นที่ข้อมูลคุณลักษณะ (Feature space) โดยใช้ฟังก์ชันในการปรับรูปแบบข้อมูลเรียกว่า ฟังก์ชันเคอร์เนล (kernel function) โดยมีสมการพยากรณ์ดังสมการที่ (10)

$$f(x) = w^T \phi(x) + b \quad (10)$$

เมื่อ W น้ำหนักของเวกเตอร์ และ b คือ ความคลาดเคลื่อนของเส้นถดถอยการกำหนดค่าของ W และ b ทำได้โดยการหาค่าต่ำสุดจากสมการที่ (11)

$$R(C) = \frac{1}{2} \|w\|^2 + C \cdot \frac{1}{n} \sum_{i=1}^N |y_i - f(x)|_{\epsilon} \quad (11)$$

การใช้เทคนิค SVR ทำนายค่าเอาต์พุตจากอินพุตเวกเตอร์ จะมีการสร้างท่อเอปซิลอน (Epsilon tube) โดยใช้ฟังก์ชันสูญเสีย (Loss function) ในหลายรูปแบบ เช่น Quadratic, Laplacian, Huber และ $\mathcal{E}$ -insensitive แต่ที่นิยมใช้คือ $\mathcal{E}$ -insensitive ที่นำเสนอโดย Vapnik [27] เนื่องจากเป็นฟังก์ชันสูญเสียที่นิยมใช้ในงานวิจัยอย่างแพร่หลาย เช่น งานวิจัยของ Yashlan & Bican [13] ได้ประยุกต์ใช้ฟังก์ชันการสูญเสียแบบ $\mathcal{E}$ -insensitive เพื่อใช้ในการพยากรณ์การใช้พลังงานไฟฟ้าของประเทศจีน เป็นต้น ฟังก์ชันการสูญเสียแบบ $\mathcal{E}$ -insensitive สามารถแสดงได้ดังสมการที่ (12)

$$|y_i - f(x)| = \begin{cases} 0 & \text{if } |y_i - f(x)| \leq \epsilon \\ |y_i - f(x)|_{\epsilon} & \text{otherwise} \end{cases} \quad (12)$$

กำหนดให้ ξ_i และ ξ_i^* แทนตัวแปรที่ทำให้เกิดค่าคลาดเคลื่อน (Slack variable) และอยู่ห่างจากระยะขอบของเอปซิลอน ($\mathcal{E}$ -tube) ดังรูปที่ 4

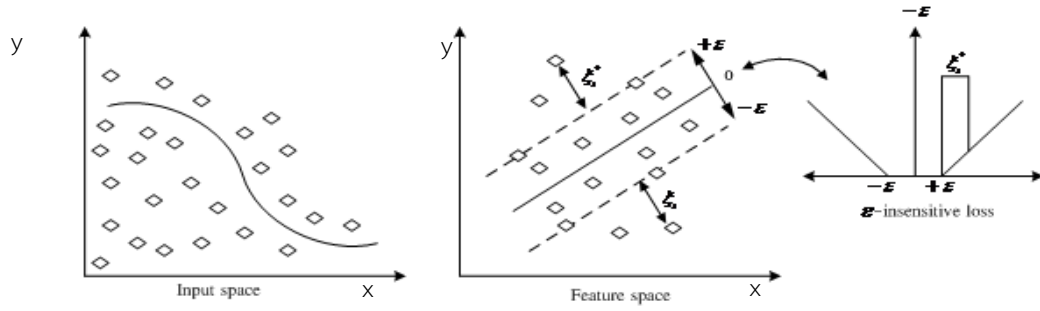


Figure 4 Boundary of Support Vector Regression. [13]

การเพิ่มตัวแปร ξ_i และ ξ_i^* ซึ่งเป็นการวัดส่วนเบี่ยงเบนทั้งหมดของจุดฝึกสอนที่อยู่นอกขอบระนาบหรือคลาดเคลื่อนของข้อมูลจากขอบระนาบบนและล่าง ดังนั้นการเพิ่ม ξ_i และ ξ_i^* จะทำให้สมการที่ (11) ถูกแปลงให้อยู่ในเงื่อนไขบังคับ (Constrained) ดังสมการที่ (13)

$$\text{Minimize} = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N (\xi_i + \xi_i^*) \quad (13)$$

$$\text{Subject to} \begin{cases} y_i - w\phi_i - b \leq \xi_i + \varepsilon \\ -y_i + w\phi_i + b \leq \xi_i^* + \varepsilon \\ \xi_i, \xi_i^* \geq 0 \end{cases} \quad (14)$$

การแก้ปัญหาของสมการที่ 13 โดยมีเงื่อนไขตามสมการที่ 14 ปัญหานี้สามารถแก้ได้โดยแปลงเป็นปัญหาควคู่ (Dual Problem) ด้วยตัวคูณลากรองจ์ (Lagrange multipliers) แสดงดังสมการที่ 15 และ เงื่อนไขข้อบังคับดังสมการที่ 16

$$\text{Maximize } (\alpha_i, \alpha_i^*) = \sum_{i=1}^N y_i (\alpha_i - \alpha_i^*) - \varepsilon \sum_{i=1}^N (\alpha_i + \alpha_i^*) \quad (15)$$

$$- \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) k(x_i, x_j)$$

$$\text{Subject to} \begin{cases} \sum_{i=1}^N (\alpha_i + \alpha_i^*) = 0 \\ 0 \leq \alpha_i \leq C, i = 1, \dots, N \\ 0 \leq \alpha_i^* \leq C, i = 1, \dots, N \end{cases} \quad (16)$$

เมื่อ X_i, X_j คือ ข้อมูลอินพุต α_i, α_i^* คือ ตัวคูณลากรองจ์ C คือ ค่าคงที่ N คือ จำนวนซัพพอร์ตเวกเตอร์ซึ่งมีอินพุตเป็นซัพพอร์ตเวกเตอร์ จะมี $\alpha_i, \alpha_i^* > 0$ ส่วนอินพุตเวกเตอร์ที่ไม่ใช่ซัพพอร์ตเวกเตอร์ จะมี $\alpha_i, \alpha_i^* = 0$ และหลังจากที่คำนวณค่า α_i และ α_i^* จากชุดข้อมูลชุดที่ 1 ดังนั้นเราสามารถสร้างสมการ SVR เพื่อใช้ทำนายค่าเอาต์พุตจากอินพุตเวกเตอร์ได้ดังสมการที่ 17

$$f(x) = \sum_{i=1}^N (\alpha_i - \alpha_i^*) k(x_i, x) + b \quad (17)$$

เมื่อ α_i และ α_i^* คือ ตัวคูณลากรองจ์, $k(x_i, x)$ คือ เคอร์เนลฟังก์ชันที่ใช้ในการแปลงข้อมูลให้มีมิติที่สูงขึ้น (High dimensional feature space) โดยเคอร์เนลฟังก์ชันที่นิยมใช้ในวิธีซัพพอร์ตเวกเตอร์รีเกรสชัน เช่น

1) ลีเนียร์เคอร์เนล (Linear Kernel)

$$k(x_i, x) = x_i^T x$$

2) โพลีโนเมียลเคอร์เนล (Polynomial kernel)

$$k(x_i, x) = (1 + x_i \cdot x)^d$$

3) เกาส์เซียนเคอร์เนล (Gaussian (RBF) kernel)

$$k(x_i, x) = \exp\left(-\frac{\|x_i - x\|^2}{2\sigma^2}\right)$$

และ b คือ ความคลาดเคลื่อน (bias) ตัวแบบซัพพอร์ตเวกเตอร์รีเกรสชัน ที่ใช้ในงานวิจัยนี้เขียนขึ้นจากโปรแกรมภาษาอาร์ ด้วยฟังก์ชัน svm() ในแพ็คเกจชื่อ e1071 ซึ่งได้รับการพัฒนาและถูกประยุกต์ใช้ในงานวิจัยอย่างแพร่หลาย โดยการจัดเรียงข้อมูลอนุกรมเวลาให้อยู่ในรูปแบบสมการถดถอยจำนวน m คอลัมน์ดังรูปที่ 5 ซึ่งเป็นข้อมูลย้อนหลังตั้งแต่คอลัมน์ที่ 1 ถึงคอลัมน์ที่ $m-1$ เป็นข้อมูลป้อนเข้า (Input data) และคอลัมน์ที่ m เป็นคอลัมน์เป้าหมาย (Target data) เพื่อให้ตัวแบบซัพพอร์ตเวกเตอร์รีเกรสชันเรียนรู้และสร้างฟังก์ชันการทำนายจากคุณลักษณะของข้อมูลที่ใช้ในการเรียนรู้ของตัวแบบซัพพอร์ตเวกเตอร์รีเกรสชัน

$$y_1, y_2, y_3, \dots, y_{t-3}, y_{t-2}, y_{t-1}, y_t$$

$$\downarrow$$

$$\begin{bmatrix} y_1 & y_2 & y_3 & \cdots & y_{m-2} & y_{m-1} & y_m \\ y_2 & y_3 & y_4 & \cdots & y_{m-1} & y_m & y_{m+1} \\ y_3 & y_4 & y_5 & \cdots & y_m & y_{m+1} & y_{m+2} \\ \vdots & & & \vdots & & & \vdots \\ y_{t-m+1} & y_{t-m+2} & y_{t-m+3} & \cdots & y_{t-2} & y_{t-1} & y_t \end{bmatrix}$$

Figure 5 Data formatting for support vector regression model

เพื่อให้การอธิบายผลการวิจัยมีความกระชับจึงใช้สัญลักษณ์ $SVR(m)$ แทน ตัวแบบซัพพอร์ตเวกเตอร์รีเกรสชันที่ถูกจัดเรียงในรูปแบบใหม่จำนวน m คอลัมน์ โดยงานวิจัยนี้เลือกใช้จำนวนคอลัมน์ตั้งแต่ 2 ถึง 12 คอลัมน์ ซึ่งเหตุผลที่เลือกใช้จำนวนไม่เกิน 12 คอลัมน์ เนื่องจากผู้วิจัยเก็บข้อมูลเป็นรายเดือนจำนวนคาบเวลาใน 1 ฤดูกาลเท่ากับ 12 คาบเวลาตามรอบปฏิทินและเพื่อไม่ให้ใช้ข้อมูลย้อนหลังหลายช่วงเวลามากจนเกินไปในการพยากรณ์ค่าในอนาคตเพียงค่าเดียวสำหรับการสร้างและพิจารณาตัวแบบซัพพอร์ตเวกเตอร์รีเกรสชันที่เหมาะสม [21] นอกจากนี้ การปรับค่าพารามิเตอร์ของตัวแบบจะอาศัยการค้นหาแบบกริด (Grid search) ด้วยการเพิ่มค่าขึ้นทีละหน่วย เพื่อให้ได้ตัวแบบที่เหมาะสมที่สุด

2.6 ตัวแบบผสมด้วยวิธีเชิงประจักษ์ร่วมกับวิธีบ็อก-เจนกินส์และวิธีซัพพอร์ตเวกเตอร์รีเกรสชัน

จากวิธีการพยากรณ์โดยใช้ตัวแบบ SARIMA และตัวแบบ SVR ข้างต้นจะเห็นได้ว่าเป็นวิธีการพยากรณ์ที่เหมาะสมกับการพยากรณ์ข้อมูลอนุกรมเวลาต่างกักล่าวคือตัวแบบ

SARIMA เหมาะสำหรับการวิเคราะห์อนุกรมเวลาที่เป็นเชิงเส้นตรง ส่วนตัวแบบ SVR เป็นตัวแบบที่ใช้ได้ดีกับข้อมูลอนุกรมเวลาที่ไม่เป็นเชิงเส้นตรง ด้วยเหตุนี้การศึกษาวิจัยครั้งนี้ผู้วิจัยได้นำเสนอตัวแบบผสมด้วยวิธีการเชิงประจักษ์ร่วมกับตัวแบบ SARIMA และวิธีซัพพอร์ตเวกเตอร์รีเกรสชัน ดังรูปที่ 6 จากรูปที่ 6 สามารถอธิบายรายละเอียดตัวแบบผสม EMD – SARIMA – SVR ได้ดังนี้

ขั้นตอนที่ 1 วิเคราะห์ข้อมูลโดยใช้เทคนิคอนุกรมเวลาเชิงประจักษ์เพื่อแยกองค์ประกอบของอนุกรมเวลาด้านฉบับให้อยู่ในรูปของชุดฟังก์ชันภายใน (Intrinsic Mode Functions : IMFs) และค่าความคลาดเคลื่อน

ขั้นตอนที่ 2 หาค่าเฉลี่ยของข้อมูลอนุกรมเวลาของชุดฟังก์ชันภายใน โดยไม่รวมค่าความคลาดเคลื่อน

ขั้นตอนที่ 3 นำข้อมูลอนุกรมเวลา Y_t ลบด้วยค่าเฉลี่ยของข้อมูลอนุกรมเวลาของชุดฟังก์ชันภายในเพื่อเป็นการกำจัดความถี่ของข้อมูลอนุกรมเวลาที่ไม่เป็นเชิงเส้นและมีลักษณะการเคลื่อนไหวไม่คงที่

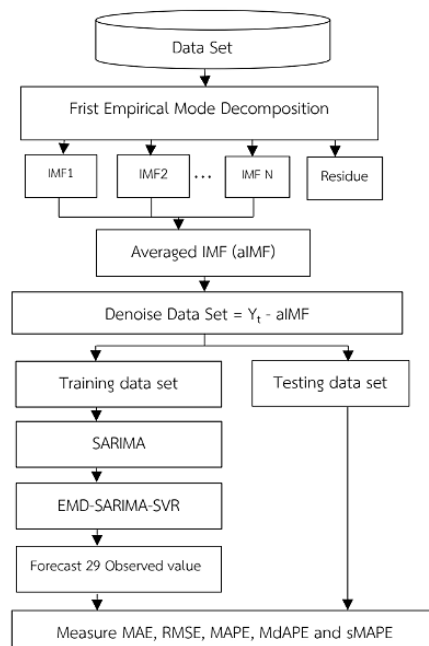


Figure 6 The Flowchart of the proposed hybrid EMD – SARIMA – SVR forecasting system.

ขั้นตอนที่ 4 นำข้อมูลที่กำจัดความถี่ไปสร้างสมการพยากรณ์โดยใช้ ตัวแบบ SARIMA ดังสมการที่ 4-9 เพื่ออธิบายลักษณะข้อมูลแบบไม่เป็นเชิงเส้นตรง เมื่อได้ตัวแบบที่เหมาะสมแล้วจากนั้นพยากรณ์ล่วงหน้าหนึ่งช่วงเวลา ดังสมการที่ (18)

$$\hat{L}_{t+1} = f(Y_t, Y_{t-1}, Y_{t-2}, \dots, Y_1) \quad (18)$$

เมื่อ $\hat{L}_{t+1}$ คือ ค่าพยากรณ์ข้อมูลที่มีลักษณะความสัมพันธ์แบบเชิงเส้นตรง ณ เวลาที่ t+1

ขั้นตอนที่ 5 หาค่าส่วนเหลือจากสมการที่ 5 โดยการนำค่าจริง (Y_t) ลบด้วยค่าพยากรณ์ ($\hat{L}_{t+1}$) เพื่อสร้างฟังก์ชันความสัมพันธ์ไม่เป็นเส้นตรงสำหรับอธิบายลักษณะความสัมพันธ์แบบไม่เป็นเส้นตรงของส่วนเหลือที่ได้จากตัวแบบ SARIMA ที่เหมาะสม ซึ่งการสร้างสมการพยากรณ์ ดังกล่าวผู้วิจัยจะใช้ตัวแบบซัพพอร์ตเวกเตอร์รีเกรสชัน โดยแปลงส่วนเหลือให้เป็นชุดข้อมูล $D = \{(x_i, y_i)\}_{i=1}^n$ ตามหัวข้อที่ 5 จากนั้นพยากรณ์ค่าข้อมูลที่มีความสัมพันธ์ไม่เป็นเชิงเส้นตรงหนึ่งช่วงเวลาดังสมการที่ (19)

$$\hat{N}_{t+1} = f((Y_t - \hat{Y}_t), (Y_{t-1} - \hat{Y}_{t-1}), (Y_{t-2} - \hat{Y}_{t-2}), \dots, (Y_1 - \hat{Y}_1)) \quad (19)$$

เมื่อ $\hat{N}_{t+1}$ คือ ค่าพยากรณ์ข้อมูลที่มีลักษณะความสัมพันธ์แบบไม่เชิงเส้นตรง ณ เวลาที่ t+1

ขั้นตอนที่ 6 นำผลการพยากรณ์ค่าข้อมูลแบบเชิงเส้นตรงและไม่เป็นเส้นตรงที่ได้จากสมการที่ (18) และ (19) รวมกันแบบความสัมพันธ์เชิงบวก เพื่อหาค่าพยากรณ์ของข้อมูลอนุกรมเวลาดังสมการที่ (20)

$$\hat{Y}_{t+1} = \hat{L}_{t+1} + \hat{N}_{t+1} + \mathcal{E}_{t+1} \quad (20)$$

$\hat{Y}_{t+1}$ คือ ค่าพยากรณ์ข้อมูลอนุกรมเวลา ณ เวลาที่ t+1

$\hat{L}_{t+1}$ คือ ค่าพยากรณ์ข้อมูลที่มีลักษณะความสัมพันธ์แบบเชิงเส้นตรง ณ เวลาที่ t+1

$\hat{N}_{t+1}$ คือ ค่าพยากรณ์ข้อมูลที่มีลักษณะความสัมพันธ์แบบไม่เชิงเส้นตรง ณ เวลาที่ t+1

$\mathcal{E}_{t+1}$ คือ ความคลาดเคลื่อน ณ เวลาที่ t+1

2.7 เกณฑ์การประเมินประสิทธิภาพของตัวแบบ

การประเมินประสิทธิภาพของตัวแบบเพื่อวัดความผิดพลาดเชิงปริมาณ เพื่ออธิบายความสามารถในการให้ค่าพยากรณ์ที่ถูกต้องและเหมาะสมในการอธิบายข้อมูลอนุกรมเวลา โดยใช้เกณฑ์ที่ใช้สำหรับการวัดประสิทธิภาพของตัวแบบการพยากรณ์ในงานวิจัยนี้มีด้วยกันทั้งหมด 5 เกณฑ์ ดังนี้

7.1 ความคลาดเคลื่อนสัมบูรณ์เฉลี่ย

$$MAE = \sum_{i=1}^n |Y_i - \hat{Y}_i| / n \quad (21)$$

7.2 รากของความคลาดเคลื่อนกำลังสองเฉลี่ย

$$RMSE = \sqrt{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2 / n} \quad (22)$$

7.3 ร้อยละความคลาดเคลื่อนสัมบูรณ์เฉลี่ย

$$MAPE = 100 \times \sum_{i=1}^n |1 - \hat{Y}_i / Y_i| / n \quad (23)$$

7.4 เปอร์เซ็นต์ของมัธยฐานความคลาดเคลื่อนสัมบูรณ์

$$MdAPE = median \sum_{i=1}^n \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right| \times 100 \quad (24)$$

7.5 เปอร์เซ็นต์ของความคลาดเคลื่อนสัมบูรณ์เฉลี่ยแบบสมมาตร

$$sMAPE = \sum_{i=1}^n \frac{2 \times |Y_i - \hat{Y}_i| / |Y_i + \hat{Y}_i|}{n} \times 100 \quad (25)$$

โดยที่ค่า MAE, RMSE, MAPE, MdAPE และ sMAPE ยิ่งน้อยก็ยิ่ง แสดงว่า ตัวแบบมีประสิทธิภาพสูง

3. ผลการวิจัย

3.1 ผลการวิเคราะห์ข้อมูลค่าสถิติพื้นฐาน

ผลการวิเคราะห์ค่าสถิติพื้นฐาน พบว่า ปริมาณหน่วยจำหน่ายไฟฟ้าประเภทบ้านอยู่อาศัยโดยเฉลี่ยต่อเดือน เท่ากับ 2201.319 ล้านหน่วยต่อเดือน ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 440.973 ล้านหน่วยต่อเดือน โดยมีหน่วยจำหน่ายไฟฟ้าสูงสุด เท่ากับ 3556.2 ล้านหน่วยต่อเดือน และ ต่ำสุด 1406.59 ล้านหน่วยต่อเดือน รายละเอียดดังตารางที่ 1

Table 1 Basic statistics of the electricity distribution units for residents' houses

data	Number	Mean	Std.	Max.	Min.	Median
All sample	137	2201.319	440.973	3556.2	1406.59	2165.5
Training	108	2068.427	370.0976	3233.3	1406.59	2012.37
Testing	29	2696.226	315.0829	3556.2	2069.01	2733.45

3.2 ผลการวิเคราะห์ตัวแบบพยากรณ์อย่างง่าย

เนื่องจากตัวแบบอย่างง่ายเป็นแบบที่มีฟังก์ชันความสัมพันธ์ที่ไม่ยุ่งยากและง่ายต่อการประยุกต์ใช้ โดยอาศัยข้อมูลที่ทราบในปัจจุบันเป็นค่าพยากรณ์ของช่วงเวลาถัดไป โดยใช้สมการที่ (1) ผู้วิจัยประยุกต์ใช้โปรแกรมภาษาอาร์มีฟังก์ชัน `snaiive ()` ซึ่งอยู่ใน Package “forecast” ผลการวิเคราะห์ข้อมูลแสดงดังตารางที่ 4

3.3 ผลพยากรณ์โดยใช้วิธีบ็อก-เจนกินส์

3.3.1 จากการพิจารณาข้อมูลอนุกรมเวลาหน่วยจำหน่ายไฟฟ้าประเภทบ้านอยู่อาศัยมีลักษณะไม่คงที่ โดยการพิจารณาจากกราฟ (รูปที่1) ซึ่งพบว่ากราฟมีแนวโน้มเป็นส่วนประกอบมีลักษณะการกระจายไม่สม่ำเสมอ

ของกราฟบานออกเป็นปากแตรแสดงว่าข้อมูลมีลักษณะไม่คงที่ สอดคล้องกับผลการทดสอบความคงที่ของข้อมูลด้วย Augmented Dickey – Fuller ($ADF = -0.0792$, $p\text{-value} = 0.588$) จึงสรุปได้ว่าข้อมูลชุดนี้มีลักษณะไม่คงที่ นั่นคือมีค่าเฉลี่ยไม่คงที่และความแปรปรวนไม่คงที่ เมื่อทราบว่าข้อมูลมีลักษณะไม่คงที่ ผู้วิจัยจึงทำการแปลงข้อมูลโดยการหาลอการิทึมก่อนแล้วจึงหาผลต่างของชุดข้อมูลก่อนนำไปสร้างตัวแบบ SARIMA และทำการทดสอบความคงที่ของข้อมูลด้วย Augmented Dickey – Fuller (-7.0434 , $p\text{-value} = 0.01$) พบว่า ข้อมูลมีลักษณะคงที่ แสดงดังรูปที่ 7 ซึ่งเป็นข้อมูลที่มีลักษณะคงที่แล้วจากนั้นนำไปใช้ในการกำหนดตัวแบบในลำดับต่อไป

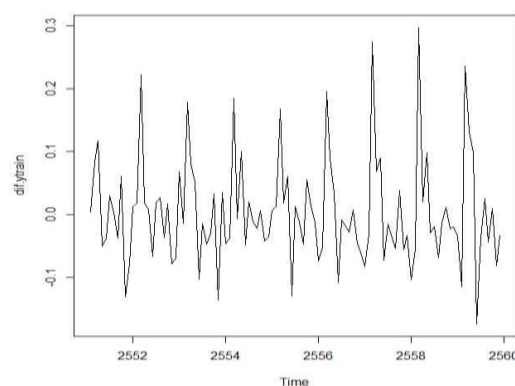


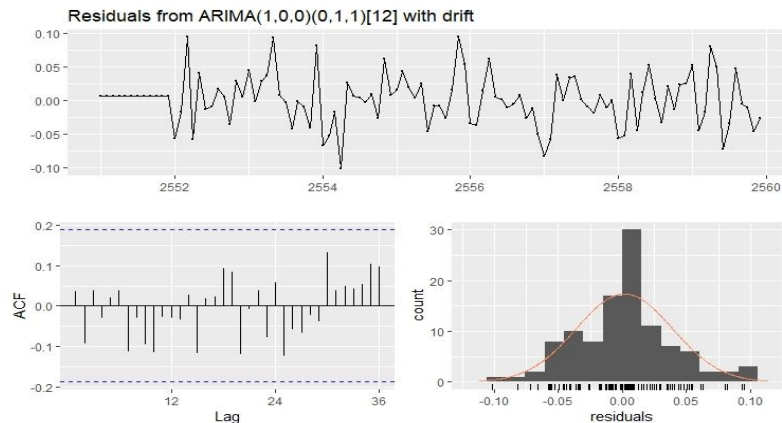
Figure 7 Difference log transform data to make data stationary on both mean and variance.

3.3.2 กำหนดรูปแบบของตัวแบบ SARIMA ต้องระบุอันดับ (p,d,q) และอันดับ (P,D,Q)_s ซึ่งโปรแกรมภาษาอาร์มีฟังก์ชัน `auto.arima()` ซึ่งอยู่ใน Package “forecast” สำหรับกำหนดอันดับ (p,d,q) และอันดับ (P,D,Q)_s ที่เหมาะสม ผลจากการใช้ฟังก์ชัน `auto.arima()` กำหนดรูปแบบของตัวแบบของ SARIMA ผลการวิเคราะห์ข้อมูล พบว่า ตัวแบบที่เหมาะสมสำหรับการพยากรณ์หน่วยจำหน่ายไฟฟ้าประเภทบ้านอยู่อาศัยคือ SARIMA(1,0,0)(0,1,1)₁₂ ไม่มีพจน์ค่าคงที่ โดยมีคาบเวลา 1 ฤดูกาลเท่ากับ 12 ($S=12$) สามารถประมาณค่าสัมประสิทธิ์ $AR_1(\hat{\phi}_1)$ เท่ากับ 0.571 ($Z = 6.714^{**}$, $p < 0.00$) และ $SAR_1(\hat{\Phi}_1)$ เท่ากับ -0.589 ($Z = -5.915^{**}$, $p < 0.00$) ตามลำดับ

3.3.3 การตรวจสอบความเหมาะสมของตัวแบบ โดยพิจารณาตามข้อตกลงเบื้องต้นของตัวแบบ SARIMA ซึ่งมีทั้งหมด 4 ข้อ ได้แก่ 1) ความคลาดเคลื่อนมีค่าเฉลี่ยเท่ากับศูนย์ ผลการตรวจสอบ พบว่า ค่าเฉลี่ยความคลาดเคลื่อนมีค่าเท่ากับ 0 ($t\text{-test} = 0.1656$, $p\text{-value} = 0.6783$) 2) ความคลาดเคลื่อนมีการแจกแจงแบบปกติ ผลการตรวจสอบ พบว่า ความคลาดเคลื่อนมีการแจกแจงแบบปกติ ($KS\text{-value} = 0.1148$, $p\text{-value} = 0.1157$) 3) ความคลาดเคลื่อนเป็นอิสระต่อกัน โดยใช้ Ljung – Box Q-statistics พบว่า ความคลาดเคลื่อนเป็นอิสระต่อกัน ($Q\text{-statistics} = 12.16$, $p\text{-value} = 0.8638$) (ดังตารางที่ 2) และ 4) ความคลาดเคลื่อนมีค่าความแปรปรวนคงที่ (ดังรูปที่ 8) ซึ่งพบว่า ความคลาดเคลื่อนมีค่าความแปรปรวนคงที่ (พิจารณาจากกราฟ ACF) เนื่องจากความคลาดเคลื่อนมีการกระจายคงที่รอบค่าใดค่าหนึ่ง ดังนั้นสามารถสรุปได้ว่าตัวแบบที่ได้มีความเหมาะสมที่จะนำไปพยากรณ์

Table 2 Examining the basic assumption of the SARIMA model

model	The mean error value is 0		The error value is normal distribution		The errors are dependent	
	t-value	p-value	KS -value	p-value	Ljung – Box Q-statistics	p-value
SARIMA(1,0,0)(0,1,1) ₁₂	0.1656	0.6783	0.1148	0.1157	12.16	0.8638

**Figure 8** Residuals from the fitted SARIMA (1,0,0) (0,1,1)₁₂ model for the electricity distribution units for resident s' houses data.

3.4 ผลการพยากรณ์โดยใช้วิธีซัพพอร์ตเวกเตอร์รีเกรสชัน

ขั้น

การสร้างตัวแบบซัพพอร์ตเวกเตอร์รีเกรสชันผู้วิจัยดำเนินการเขียนโปรแกรมโดยใช้โปรแกรมภาษาอาร์ซึ่งมีฟังก์ชัน svm () ในแพ็คเกจชื่อ e1071 ที่เผยแพร่โดย Meyer และคนอื่น ๆ [28] ก่อนที่จะนำ SVR ไปใช้กับการพยากรณ์อนุกรมเวลาผู้วิจัยได้ปรับชุดข้อมูลให้อยู่ในรูปแบบของอินพุตเวกเตอร์และทาร์เก็ต ($D = \{(x_i, y_i)\}_{i=1}^n$) ตามตัวแบบที่นำเสนอในขั้นตอนที่ 2.3 เนื่องจากอนุกรมเวลาที่ใช้ในการศึกษามีคาบเวลาเป็นรายเดือนและจำนวนคาบเวลาใน 1 ฤดูกาลเท่ากับ 12 คาบเวลาตามรอบปฏิทิน ดังนั้นผู้วิจัยจึงเลือกอินพุตเวกเตอร์ เป็นค่าสังเกตย้อนหลัง 12 คาบเวลาจากค่าพยากรณ์

การสร้างตัวแบบ SVR จะต้องมีการกำหนดเคอร์เนลฟังก์ชันและกำหนดค่าพารามิเตอร์ที่เหมาะสมงานวิจัยนี้ใช้ฟังก์ชัน tune () ค้นหาเคอร์เนลฟังก์ชันและกำหนดค่าพารามิเตอร์ที่เหมาะสม โดยฟังก์ชัน tune () จะค้นหาค่าที่เหมาะสมแบบกริดเสิร์จ (Grid search) ตลอดช่วงของพารามิเตอร์ที่กำหนดไว้งานวิจัยนี้กำหนดค่าไว้ที่ 0 -1 และให้เพิ่มค่อย ๆ เพิ่มค่าทีละ 0.001 และกำหนดค่า Cost ไว้ที่

1-16 และประเมินความเหมาะสมด้วยการวัดค่า MSE (Mean Square Error) ผลลัพธ์การค้นหาเคอร์เนลฟังก์ชันที่เหมาะสมจากฟังก์ชัน tune () พบว่า เคอร์เนลแบบเรเดียลเบซิส (Radial Basis Function: RBF) เป็นเคอร์เนลฟังก์ชันที่เหมาะสมในการนำไปใช้พยากรณ์หน่วยจำหน่ายไฟฟ้าประเภทบ้านอยู่อาศัยของประชาชน การใช้เคอร์เนลแบบเรเดียลเบซิสจะต้องมีการกำหนดค่าพารามิเตอร์ Cost, ϵ และ γ ผลจากการใช้ฟังก์ชัน tune () ได้ค่าที่เหมาะสม คือ Cost = 1, ϵ = 0.147 และ γ = 0.0833 ตามลำดับ เมื่อได้เคอร์เนลฟังก์ชันและค่าพารามิเตอร์ที่เหมาะสม ผู้วิจัยได้ดำเนินการนำค่าดังกล่าวไปใช้ในการพยากรณ์เพื่อเปรียบเทียบตัวแบบและเลือกตัวแบบที่เหมาะสมในการพยากรณ์หน่วยจำหน่ายไฟฟ้าประเภทบ้านอยู่อาศัยของประชาชนต่อไป (ดังตารางที่ 4)

3.5 ตัวแบบผสมด้วยวิธีเชิงประจักษ์ร่วมกับวิธีบ็อก-เจนกินส์และวิธีซัพพอร์ตเวกเตอร์รีเกรสชัน

3.5.1 ทำการวิเคราะห์ข้อมูลโดยใช้เทคนิคอนุกรมเวลาเชิงประจักษ์เพื่อแยกองค์ประกอบของอนุกรมเวลาด้านฉบับให้อยู่ในรูปของชุดฟังก์ชันภายใน (Intrinsic Mode Functions: IMFs) และค่าความคลาดเคลื่อน ผลการวิเคราะห์ดังตารางที่ 3 และรูปที่ 9

Table 3 Comparison of EMD and original data values

Data	Mean	Std.	Min	Max
dataset	2201.32	440.97	1406.59	3556.20
Denoise dataset	2188.17	413.17	1434.47	3525.73
IMF1	6.64	91.55	-306.64	245.03
IMF2	-2.20	111.80	-230.24	216.62
IMF3	0.97	168.48	-424.34	383.74
IMF4	5.17	100.98	-218.87	232.76
IMF5	-23.25	254.38	-540.52	306.58
IMF6	25.18	74.37	-77.45	182.97
IMF7	62.72	179.43	-198.02	370.58
IMF8	29.95	227.38	-335.84	424.17
Residue	2096.13	362.85	1275.54	3312.45
aIMF	13.15	47.65	-92.25	169.61

3.5.2 ผลการวิเคราะห์โดยใช้วิธีบ็อก-เจนกินส์ เมื่อการจัดความถี่ของข้อมูลรวมเวลาที่ไม่เป็นเชิงเส้นและมีลักษณะการเคลื่อนไหวไม่คงที่ออกเรียบร้อยแล้วก่อนที่จะไปหาตัวแบบ SARIMA ที่เหมาะสม ผลการวิเคราะห์ข้อมูล พบว่า ตัวแบบที่เหมาะสม คือตัวแบบ SARIMA(1,0,1)(1,1,2)₁₂ ไม่มีเทอมค่าคงที่ ผู้วิจัยนำตัวแบบดังกล่าวมาพยากรณ์เพื่อหาค่าพยากรณ์ข้อมูลที่มีลักษณะความสัมพันธ์แบบเชิงเส้นตรง ณ เวลาที่ $t+1$ แล้วนำค่าจริงมาลบค่าพยากรณ์ข้อมูลที่มีลักษณะความสัมพันธ์แบบเชิงเส้นตรงจะได้ส่วนเหลือ จากนั้นนำค่าส่วนเหลือไปพยากรณ์โดยใช้ตัวแบบซัพพอร์ตเวกเตอร์รีเกรสชันตามขั้นตอนของการสร้างตัวแบบผสม (ขั้นตอนที่ 6) การพยากรณ์ค่าส่วนเหลือผู้วิจัยจะทำการแปลงข้อมูลให้อยู่ในรูปแบบของอินพุตเวกเตอร์และทาร์เก็ต ($D = \{(x_i, y_i)\}_{i=1}^n$) ตามตัวแบบที่นำเสนอในขั้นตอนที่ 2.3 เนื่องจากอนุกรมเวลาที่ใช้ในการศึกษามีคาบเวลาเป็นรายเดือนและจำนวนคาบเวลาใน 1 ฤดูกาลเท่ากับ 12 คาบเวลาตามรอบปฏิทิน จากนั้นใช้

ฟังก์ชัน tune () ค้นหาเคอร์เนลฟังก์ชันและกำหนดค่าพารามิเตอร์ที่เหมาะสม งานวิจัยนี้กำหนดค่าไว้ที่ 0 -1 และให้เพิ่มค่อย ๆ เพิ่มค่าทีละ 0.001 และกำหนดค่า Cost ไว้ที่ 1-16 และประเมินความเหมาะสมด้วยการวัดค่า MSE (Mean Square Error) ผลลัพธ์การค้นหาเคอร์เนลฟังก์ชันที่เหมาะสมจากฟังก์ชัน tune () พบว่า เคอร์เนลแบบเรเดียลเบซิส (Radial Basis Function: RBF) เป็นเคอร์เนลฟังก์ชันที่เหมาะสมในการนำไปใช้พยากรณ์หน่วยจำหน่ายไฟฟ้าประเภทบ้านอยู่อาศัยของประชาชน การใช้เคอร์เนลแบบเรเดียลเบซิสจะต้องมีการกำหนดค่าพารามิเตอร์ Cost, ϵ และ γ ผลจากการใช้ฟังก์ชัน tune () ได้ค่าที่เหมาะสม คือ Cost = 1, ϵ = 0.0833 และ γ = 0.892 ตามลำดับ เมื่อได้เคอร์เนลฟังก์ชันและค่าพารามิเตอร์ที่เหมาะสม ผู้วิจัยได้ดำเนินการนำค่าดังกล่าวไปใช้ในการพยากรณ์เพื่อเปรียบเทียบตัวแบบและเลือกตัวแบบที่เหมาะสมในการพยากรณ์หน่วยจำหน่ายไฟฟ้าประเภทบ้านอยู่อาศัยของประชาชนต่อไป ผลแสดงในตารางที่ 4

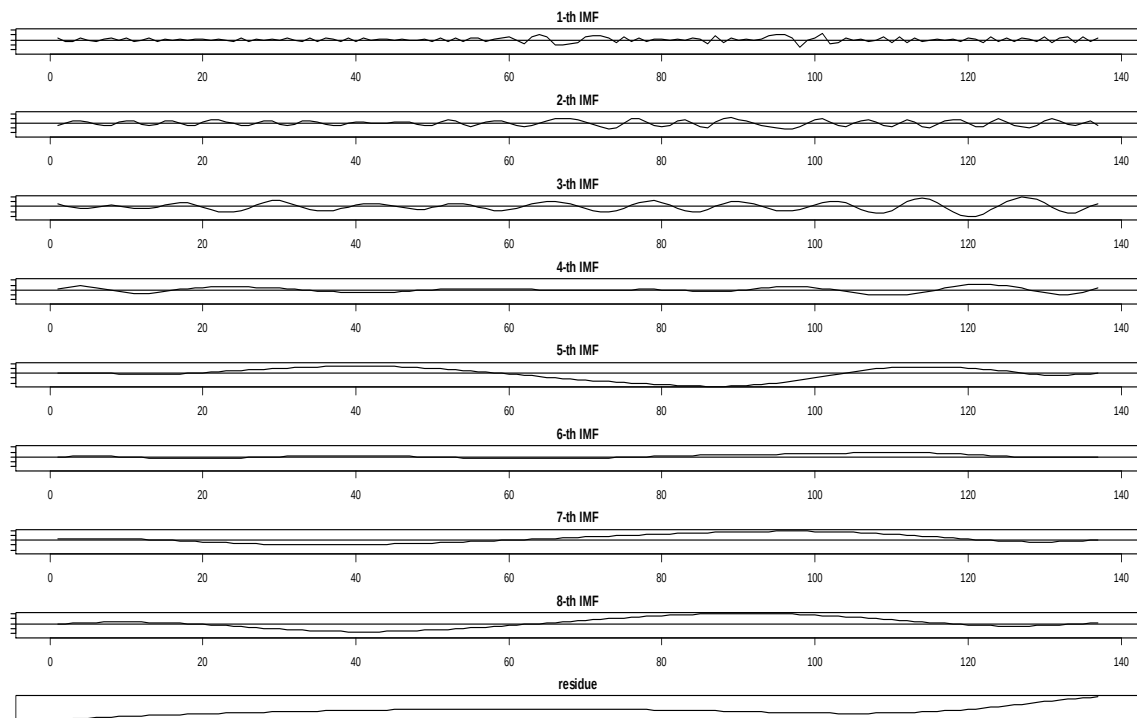


Figure 9 EMD components of electricity distribution units for residents' houses dataset of experiment.

Table 4 Comparison between different actual and forecasted of hybrid model and univariate model

Months / Year	electricity distribution units	Naïve	SARIMA	SVR(12)	EMD-SARIMA- SVR(12)
Jan 2560	2304.37	2276.29	2235.74	2321.288	2166.98
Feb 2560	2069.01	2030.88	2113.55	2281.77	2052.18
Mar 2560	2672.77	2571.12	2720.85	2517.425	2677.53
Apr 2560	2777.48	2930.48	2957.86	2602.47	2875.21
May 2560	2985	3233.3	3232.34	2679.595	3160.69
Jun 2560	2737.79	2716.09	2902.34	2712.624	2875.94
Jul 2560	2733.45	2619.76	2844.33	2645.746	2806.95
Aug 2560	2774.81	2685.85	2804.08	2683.271	2720.37
Sep 2560	2715.63	2569.31	2706.45	2646.193	2629.05
Oct 2560	2710.59	2593.34	2753.56	2604.595	2672.59
Nov 2560	2445.83	2389.31	2602.84	2505.662	2589.40
Dec 2560	2402.64	2311.74	2524.50	2429.432	2511.96
Jan 2561	2222.01	2276.29	2408.65	2419.024	2310.21
Feb 2561	2165.5	2030.88	2259.46	2324.303	2133.19
Mar 2561	2740.76	2571.12	2895.85	2548.303	2792.82
Apr 2561	2811.77	2930.48	3140.17	2626.356	3034.73
May 2561	2970.28	3233.3	3426.61	2675.539	3345.23
Jun 2561	2812.79	2716.09	3074.24	2620.227	3025.90
Jul 2561	2857.32	2619.76	3011.38	2621.032	2936.02
Aug 2561	2775.3	2685.85	2967.97	2635.667	2860.14
Sep 2561	2640.44	2569.31	2864.19	2618.986	2763.95

Table 4 Comparison between different actual and forecasted of hybrid model and univariate model (Continue)

Months / Year	electricity distribution units	Naïve	SARIMA	SVR(12)	EMD-SARIMA- SVR(12)
Oct 2561	2858.07	2593.34	2913.79	2582.397	2802.06
Nov 2561	2610.08	2389.31	2754.17	2504.196	2691.40
Dec 2561	2606.71	2311.74	2671.19	2452.368	2607.85
Jan 2562	2519.72	2276.29	2548.57	2408.592	2409.36
Feb 2562	2410.78	2030.88	2390.69	2391.025	2209.73
Mar 2562	3031.67	2571.12	3064.03	2520.025	2903.77
Apr 2562	3271.79	2930.48	3322.52	2546.561	3182.06
May 2562	3556.2	3233.3	3625.60	2498.065	3512.46
	MAE	171.316	129.061	203.864	105.443
	RMSE	205.090	164.543	299.252	130.400
	MAPE	6.187%	4.768%	7.075%	3.899%
	MdAPE	5.389%	4.553%	5.921%	3.188%
	sMAPE	6.398	4.602%	7.491%	3.850%

4. สรุปและข้อเสนอแนะ

จากผลการทดลองใช้ตัวแบบการพยากรณ์ทั้ง 4 ตัวแบบ จากตารางที่ 4 พบว่า ตัวแบบอนุกรมเวลาเชิงเดี่ยวทั้ง 3 ตัวแบบ ได้แก่ ตัวแบบอย่างง่าย ตัวแบบ SARIMA และตัวแบบซัพพอร์ตเวกเตอร์รีเกรสชัน ต่างก็มีประสิทธิภาพที่ดีในแต่ละเกณฑ์การวัดประสิทธิภาพการพยากรณ์แตกต่างกัน ตัวแบบ SARIMA มีประสิทธิภาพดีกว่าตัวแบบอย่างง่าย และตัวแบบซัพพอร์ตเวกเตอร์รีเกรสชันตามเกณฑ์การวัดประสิทธิภาพซึ่งบ่งชี้ว่า การใช้ตัวแบบ SARIMA สำหรับการพยากรณ์ข้อมูลอนุกรมเวลาที่เป็นเชิงเส้นตรงจะมีประสิทธิภาพดีกว่าตัวแบบอย่างง่าย และตัวแบบซัพพอร์ตเวกเตอร์รีเกรสชันในการทำงานหากพิจารณาตัวแบบอย่างง่ายเทียบกับตัวแบบซัพพอร์ตเวกเตอร์รีเกรสชัน พบว่า ตัวแบบอย่างง่ายให้ประสิทธิภาพดีกว่าตัวแบบซัพพอร์ตเวกเตอร์รีเกรสชันทุกเกณฑ์การวัดเช่นเดียวกัน แต่อย่างไรก็ตามตัวแบบผสมระหว่างตัวแบบ SARIMA และตัวแบบซัพพอร์ตเวกเตอร์รีเกรสชันโดยมีการกรองข้อมูลโดยใช้วิธีการเชิงประจักษ์เพื่อลดความถี่ของข้อมูลก่อนนำไปใช้วิธีการพยากรณ์ พบว่า ตัวแบบผสมสามารถให้ประสิทธิภาพการพยากรณ์ดีทุกเกณฑ์การวัดประสิทธิภาพซึ่งบ่งชี้ว่า การประยุกต์ใช้จุดเด่นและประสิทธิภาพในการอธิบายลักษณะความสัมพันธ์ของข้อมูลแบบเชิงเส้นตรงของตัวแบบ SARIMA และจุดเด่นของตัวแบบซัพพอร์ตเวกเตอร์รีเกรสชันสำหรับพยากรณ์ข้อมูลที่มีลักษณะสัมพันธ์ที่ไม่เป็นเชิงเส้นตรงของส่วนเหลือที่ได้จากตัวแบบ SARIMA สามารถทำให้ค่าพยากรณ์ที่ใกล้เคียงกับค่าจริงมากที่สุดเมื่อเปรียบเทียบ

กับตัวแบบที่เหลือ ซึ่งผลการวิจัยที่ได้สอดคล้องกับงานวิจัยที่อ้างอิง โดยบ่งชี้ว่าตัวแบบผสมมีประสิทธิภาพดีกว่าตัวแบบอนุกรมเวลาเชิงเดี่ยว จึงสามารถกล่าวได้ว่าตัวแบบผสมที่พัฒนาขึ้นโดยใช้การกรองข้อมูลด้วยวิธีการเชิงประจักษ์ก่อนนำไปพยากรณ์ด้วยตัวแบบ SARIMA และตัวแบบซัพพอร์ตเวกเตอร์รีเกรสชันเป็นตัวแบบที่เหมาะสมและสามารถนำไปประยุกต์ใช้การพยากรณ์ข้อมูลอนุกรมเวลาต่าง ๆ เพื่อช่วยในการตัดสินใจได้อย่างมีประสิทธิภาพตามลำดับ

ข้อเสนอแนะในการเลือกใช้ตัวแบบสำหรับการพยากรณ์ จากการประเมินประสิทธิภาพของตัวแบบอนุกรมเวลาเชิงเดี่ยวทั้ง 3 ตัวแบบ ตัวแบบ SARIMA เป็นตัวแบบการพยากรณ์ที่เหมาะสมมากที่สุด รองลงมาคือ ตัวแบบอย่างง่าย และตัวแบบตัวแบบซัพพอร์ตเวกเตอร์รีเกรสชัน ดังนั้นเมื่อผู้พยากรณ์ต้องการใช้ตัวแบบการพยากรณ์อนุกรมเวลาเชิงเดี่ยวเพื่อสะดวกในการใช้งานเบื้องต้น ควรเลือกใช้ตัวแบบ SARIMA และหากต้องการได้ผลการพยากรณ์ที่มีความถูกต้องเพิ่มมากขึ้น ตัวแบบการพยากรณ์แบบผสมระหว่างตัวแบบ SARIMA และตัวแบบซัพพอร์ตเวกเตอร์รีเกรสชันโดยมีการกรองข้อมูลโดยใช้วิธีการเชิงประจักษ์เพื่อลดความถี่ของข้อมูลก่อนนำไปใช้วิธีการพยากรณ์เป็นตัวแบบการพยากรณ์ที่มีประสิทธิภาพมากที่สุด ดังนั้นการนำข้อมูลอนุกรมเวลามาพยากรณ์เพื่อช่วยในการตัดสินใจผู้พยากรณ์ควรเลือกใช้ตัวแบบผสมเพื่อให้ได้ค่าพยากรณ์ใกล้เคียงกับค่าจริงมากที่สุด ทั้งนี้เนื่องจากหากค่าพยากรณ์มีค่าใกล้เคียงกับค่าจริงมากเท่าใดจะเป็นสารสนเทศที่เป็นประโยชน์ประกอบการวางแผนในการ

ตัดสินใจของผู้พยากรณ์เองหรือผู้ที่มีส่วนเกี่ยวข้องที่จะนำค่าพยากรณ์ไปใช้ประโยชน์ต่อไป ข้อควรระวังในการนำค่าพยากรณ์ไปใช้ประโยชน์ ถึงแม้ตัวแบบผสมจะเป็นตัวแบบที่มีประสิทธิภาพและความแม่นยำมากที่สุดเมื่อเทียบกับตัวแบบอนุกรมเวลาเชิงเดี่ยวก็ตาม การใช้ประโยชน์ควรพิจารณาให้รอบด้านทั้งนี้เพราะว่าข้อมูลอนุกรมเวลาชุดใด ๆ ที่ผู้พยากรณ์นำมาใช้ในการพยากรณ์อาจจะมีปัจจัยอื่น ๆ ที่เข้ามามีผลกระทบทำให้ค่าพยากรณ์มีความผิดพลาดได้เป็นต้น

5. References

- [1] Cheng, H. and et al. 2019. A hybrid electricity price forecasting model with Bayesian optimization for German energy exchange. **International Journal of Electrical Power & Energy Systems**. 110, 653–666.
- [2] de Marcos, R. A., Bello, A. and Reneses, J. 2019. Electricity price forecasting in the short term hybridising fundamental and econometric modelling. **Electric Power Systems Research**. 167, 240–251.
- [3] Maluf de Lima, L. and Piedade Bacchi, M. R. 2019. Assessing the impact of Brazilian economic growth on demand for electricity. **Energy**. 172, 861–873.
- [4] AL-Musaylh, M. S. and et al. 2018. Two-phase particle swarm optimized- support vector regression hybrid model integrated with improved empirical mode decomposition with adaptive noise for multiple- horizon electricity demand forecasting. **Applied Energy**. 217, 422–439.
- [5] Yang, Z. and Wang, J. 2018. A hybrid forecasting approach applied in wind speed forecasting based on a data processing strategy and an optimized artificial intelligence algorithm. **Energy**. 160, 87–100.
- [6] Mi, X., Liu, H. and Li, Y. 2019. Wind speed prediction model using singular spectrum analysis, empirical mode decomposition and convolutional support vector machine. **Energy Conversion and Management**. 180, 196–205.
- [7] Silva, E. S. and et al. 2019. Forecasting tourism demand with denoised neural networks. **Annals of Tourism Research**. 74, 134–154.
- [8] Xu, S., Chan, H. K. and Zhang, T. 2019. Forecasting the demand of the aviation industry using hybrid time series SARIMA- SVR approach. **Transportation Research Part E: Logistics and Transportation Review**. 122, 169–180.
- [9] Li, H. and et al. 2019. Novel analysis–forecast system based on multi-objective optimization for air quality index. **Journal of Cleaner Production**. 208, 1365–1383.
- [10] Junsagoon, S. 2015. Time Series Forecasting: Linear Approach, Non – Linear Approach and Hybrid Models. **EAU HERITAGE JOURNAL Science and Technology**. 9(2) : 50 – 63. (in Thai)
- [11] Zhang, G. P. 2003. Time series forecasting using a hybrid ARIMA and neural network model. **Neurocomputing**. 50, 159 – 175.
- [12] Du, P., Wang, J., Yang, W. and Niu, T. 2019. A novel hybrid model for short- term wind power forecasting. **Applied Soft Computing**. 80, 93–106.
- [13] Yaslan, Y. and Bican, B. 2017. Empirical mode decomposition based denoising method with support vector regression for time series prediction: A case study for electricity load forecasting. **Measurement**. 103, 52–61.
- [14] Wang, J. and Wang, J. 2017. Forecasting stochastic neural network based on financial empirical mode decomposition. **Neural Networks**. 90, 8–20.
- [15] Sutthison, T. 2018. A Comparison of the Forecasting Methods of the Electric Consumption of Ubon Ratchathani Rajabhat University. **Journal of industrial Ubon Ratchathani Rajabhat University**. 7(1): 58 – 74. (in Thai)
- [16] Mao, Q. and et al. 2018. Forecasting the incidence of tuberculosis in China using the seasonal auto- regressive integrated moving average (SARIMA) model. **Journal of Infection and Public Health**. 11(5), 707–712.
- [17] Arumugam, P. and Saranya, R. 2018. Outlier Detection and Missing Value in Seasonal ARIMA Model Using Rainfall Data. **Materials Today: Proceedings**. 5(1), 1791–1799.

- [18] Curceac, S. and et al. 2019. Short-term air temperature forecasting using Nonparametric Functional Data Analysis and SARMA models. **Environmental Modelling & Software**. 111, 394–408.
- [19] Chen, Y. and et al. 2017. Short-term electrical load forecasting using the Support Vector Regression (SVR) model to calculate the demand response baseline for office buildings. **Applied Energy**. 195, 659–670.
- [20] Maldonado, S., González, A. and Crone, S. 2019. Automatic time series analysis for electric load forecasting via support vector regression. **Applied Soft Computing**. 83, 2-9.
- [21] Sujjaviriyasup, T. 2019. Hybrid Model of Linear and Nonlinear Model for Forecasting Annual Molumes of International Rice Exports. **Burapha Science Journal**. 24(2): 517 – 531. (in Thai)
- [22] PROVINCIAL ELECTRICITY AUTHORITY. **Electricity distribution units Situation**. [Internet]. 2019. [cited 2019 March 20] available form: <http://peaoc.pea.co.th/ped/>
- [23] Huang NE. and et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and nonstationary time series analysis. **Proc R Soc**. 454 (1971): 903–95.
- [24] Ungpansattawong, S. **Statistical Forecasting Techniques**. Khon Kaen: Khon Kaen University Press; 2012. (in Thai)
- [25] Box, G.E.P. and et al. 2008. **Time Series Analysis: Forecasting and Control**, 4th ed. John Wiley and Sons, Inc., New Jersey.
- [26] R Core Team. 2019. forecast: Forecasting Functions for Time Series and Linear Models. R Foundation for Statistical Computing, Vienna, Austria.URL <https://cran.rproject.org/web/packages/forecast/>
- [27] Vapnik, V. 1998. **The Nature of Statistical Learning Theory**. Springer-Verlag.
- [28] Meyer, D. and et al. 2015. **Misc Functions of the Department of Statistics**, Probability Theory Group.<http://cran.r-project.org/web/packages/e1071/index.html>,5 April 2018.