



## การพัฒนาตัวแบบผสมเพื่อพยากรณ์ข้อมูลอนุกรมเวลาปริมาณหน่วยจำหน่ายไฟฟ้าบ้านอยู่อาศัยรายเดือนของประชาชนในเขตภาคตะวันออกเฉียงเหนือของประเทศไทย

ธนกร สุทธิสินธ์\*

สาขาวิชาสถิติประยุกต์ คณะวิทยาศาสตร์ มหาวิทยาลัยราชภัฏอุบลราชธานี

สุรัตน์ หารวย

สาขาวิชาสาธารณสุขชุมชน คณะสาธารณสุขศาสตร์ มหาวิทยาลัยราชภัฏอุบลราชธานี

\* ผู้นิพนธ์ประสานงาน โทรศัพท์ 0 4535 2000 ต่อ 1430 อีเมล: thanakon.suthison@gmail.com DOI: 10.14416/j.kmutnb.2020.04.009

รับเมื่อ 6 พฤศจิกายน 2562 แก้ไขเมื่อ 5 มกราคม 2563 ตอบรับเมื่อ 24 มกราคม 2563 เผยแพร่ออนไลน์ 20 เมษายน 2563

© 2020 King Mongkut's University of Technology North Bangkok. All Rights Reserved.

### บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อพัฒนาตัวแบบที่เหมาะสมในการพยากรณ์ข้อมูลอนุกรมเวลาปริมาณหน่วยจำหน่ายไฟฟ้าบ้านอยู่อาศัยรายเดือนของประชาชนในเขตภาคตะวันออกเฉียงเหนือ เก็บรวบรวมข้อมูลจากการไฟฟ้าส่วนภูมิภาค ตั้งแต่เดือนมกราคม 2552 ถึงเดือนพฤษภาคม 2562 จำนวน 127 ค่า จำนวน 3 ชุดข้อมูล คือ กฟผ.1 กฟผ.2 และ กฟผ.3 การพัฒนาตัวแบบผู้วิจัยประยุกต์ใช้เทคนิคการวิเคราะห์ข้อมูลเชิงประจักษ์เพื่อลดสัญญาณแกว่งไกว (Fast Oscillation) ของข้อมูลก่อนนำไปพยากรณ์ด้วยตัวแบบอนุกรมเวลาเชิงเดี่ยว 2 วิธี คือ วิธีปรับให้เรียบเอ็กซ์โพเนนเชียลแบบโฮลท์-วินเทอร์ (Holt -Winters Exponential Smoothing Method) และตัวแบบออริมาแบบมีฤดูกาล (Seasonal Auto - Regressive Integrated Moving Average) จากนั้นนำมาสร้างตัวแบบผสม (Hybrid Forecasting Model) ด้วยเทคนิคการพยากรณ์ร่วมและเปรียบเทียบประสิทธิภาพการพยากรณ์ด้วยเกณฑ์การวัดประสิทธิภาพ 5 เกณฑ์ ได้แก่ 1) ความคลาดเคลื่อนสัมบูรณ์เฉลี่ย 2) รากของค่าคลาดเคลื่อนกำลังสองเฉลี่ย 3) ร้อยละของความคลาดเคลื่อนสัมบูรณ์เฉลี่ย 4) เพอร์เซ็นต์ของมัธยฐานความคลาดเคลื่อนสัมบูรณ์ และ 5) เพอร์เซ็นต์ของความคลาดเคลื่อนสัมบูรณ์เฉลี่ยแบบมาตรฐาน ผลการวิจัยบ่งชี้ว่าตัวแบบผสมที่พัฒนาขึ้น มีประสิทธิภาพการพยากรณ์ดีกว่าตัวแบบอนุกรมเวลาเชิงเดี่ยว 2 ตัวแบบ ทุกเกณฑ์การวัดประสิทธิภาพการพยากรณ์ ดังนั้น สรุปได้ว่าตัวแบบผสมเป็นตัวแบบที่เหมาะสมสำหรับการพยากรณ์ข้อมูลอนุกรมเวลาปริมาณหน่วยจำหน่ายไฟฟ้าบ้านอยู่อาศัยรายเดือนของประชาชนในเขตภาคตะวันออกเฉียงเหนือของประเทศไทย

**คำสำคัญ:** ปริมาณหน่วยจำหน่ายไฟฟ้ารายเดือน การวิเคราะห์ข้อมูลเชิงประจักษ์ ตัวแบบผสม ตัวแบบอนุกรมเวลาเชิงเดี่ยว

การอ้างอิงบทความ: ธนกร สุทธิสินธ์ และ สุรัตน์ หารวย, “การพัฒนาตัวแบบผสมเพื่อพยากรณ์ข้อมูลอนุกรมเวลาปริมาณหน่วยจำหน่ายไฟฟ้าบ้านอยู่อาศัยรายเดือนของประชาชนในเขตภาคตะวันออกเฉียงเหนือของประเทศไทย,” *วารสารวิชาการพระจอมเกล้าพระนครเหนือ*, ปีที่ 30, ฉบับที่ 4, หน้า 630-644, ต.ค.-ธ.ค. 2563.



## The Development of a Hybrid Model for Forecasting Time Series Data of Monthly Household Electrical Distribution Units of People in the Northeast, Thailand

Thanakon Sutthison\*

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

Surat Haruay

Community Health Program, Faculty of Public Health, Ubon Ratchathani Rajabhat University, Ubon Ratchathani, Thailand

\*Corresponding Author, Tel. 0 4535 2000 Ext. 1430, E-mail: thanakon.sutthison@gmail.com DOI: 10.14416/j.kmutnb.2020.04.009

Received 6 November 2019; Revised 10 January 2020; Accepted 24 January 2020; Published online: 20 April 2020

© 2020 King Mongkut's University of Technology North Bangkok. All Rights Reserved.

### Abstract

The objective of this research was to develop a suitable model for forecasting time series data of monthly household electrical distribution units of people in the Northeast, Thailand. Data were collected from the Provincial Electricity Authority from January 2009 – May 2019, a total of 127 values, consisting of 3 data sets, namely EGAT 1, EGAT 2, and EGAT 3. For the development of the model, the researcher applied empirical mode decomposition to reduce Fast Oscillation of the data before being forecasted by two univariate time series models, including Holt –Winters Exponential Smoothing Method and Seasonal Auto-regressive integrated moving average. After that, a hybrid forecasting model was created using combined forecasting method. In addition, the forecasting efficiency was compared by 5 performance measurement criteria, 1) namely Mean Absolute Error, 2) Root Mean Square Error, 3) Mean Absolute Percentage Error, 4) Median Absolute Percentage Error and 5) Symmetric Mean Absolute Percentage Error. The findings indicated that the forecasting efficiency of the developed model was better than two univariate time series models in all criteria. Therefore, it can be concluded that the hybrid model is the suitable model for forecasting time series data of monthly household electrical distribution units of people in the Northeast, Thailand.

**Keywords:** Monthly Electrical Distribution Units, Empirical Mode Decomposition, Hybrid Model, Univariate Time Series Model

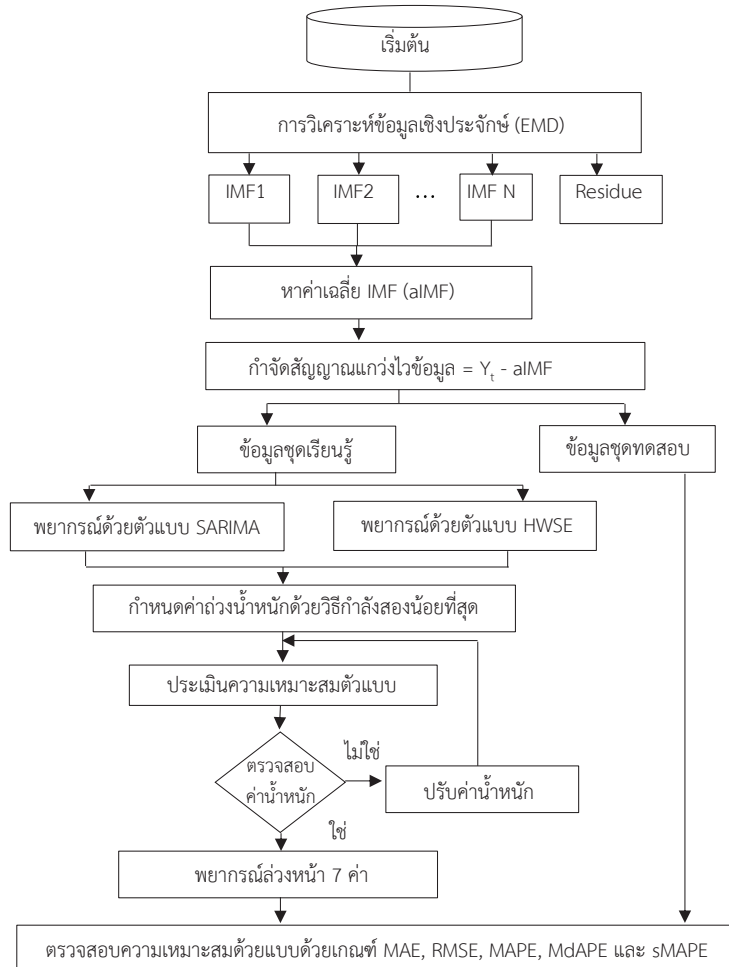
Please cite this article as: T. Sutthison and S. Haruay, "The development of a hybrid model for forecasting time series data of monthly household electrical distribution units of people in the Northeast, Thailand," *The Journal of KMUTNB*, vol. 30, no. 4, pp. 630–644, Oct.–Dec. 2020 (in Thai).

## 1. บทนำ

ปัจจุบันมีการนำความรู้เกี่ยวกับการพยากรณ์อนุกรมเวลาไปประยุกต์ใช้อย่างแพร่หลาย เช่น การพยากรณ์ความเร็วลม [1] การพยากรณ์การใช้พลังงานไฟฟ้า [2], [3] การพยากรณ์ค่าความเข้มข้นของฝุ่นละอองขนาดเล็กไม่เกิน 10 ไมครอน [4] การพยากรณ์ค่าความเข้มข้นของฝุ่นละอองขนาดเล็กไม่เกิน 2.5 ไมครอน [5] เป็นต้น จากการศึกษารายงานการวิจัยดังกล่าวพบว่า นักวิจัยแต่ละท่านพยายามพัฒนาตัวแบบการพยากรณ์ที่เหมาะสมกับข้อมูลอนุกรมเวลาที่นำมาพยากรณ์ เพื่อให้ได้ค่าพยากรณ์ที่มีความแม่นยำมากที่สุด การสร้างตัวแบบของนักวิจัยแต่ละท่านยังคงอาศัยจุดเด่นของการพยากรณ์เชิงเดี่ยวแบบดั้งเดิมเป็นฐาน เช่น ตัวแบบการปรับให้เรียบเอ็กซ์โพเนนเชียลแบบโฮลท์วินเทอร์ (Holt-Winters Exponential Smoothing Method; HWES) ตัวแบบอาร์มาแบบมีฤดูกาล (Seasonal Auto-Regressive Integrated Moving Average; SARIMA) เป็นต้น หากข้อมูลอนุกรมเวลาที่นำมาพยากรณ์เหมาะสมกับวิธีการพยากรณ์เชิงเดี่ยวตั้งแต่ 2 วิธีขึ้นไป จึงมีแนวคิดจะรวมวิธีการพยากรณ์เชิงเดี่ยวเข้าด้วยกัน โดยกำหนดน้ำหนักที่เหมาะสมให้กับค่าพยากรณ์อนุกรมเวลาเชิงเดี่ยวเพื่อพยากรณ์อนุกรมเวลาชุดนั้น จะทำให้ได้ค่าพยากรณ์มีความแม่นยำมากขึ้น [6] โดยเรียกวิธีการดังกล่าวว่าตัวแบบผสมมีการนำตัวแบบตามแนวคิดดังกล่าวไปใช้ในการพยากรณ์พบว่า มีความแม่นยำมากกว่าตัวแบบอนุกรมเวลาเชิงเดี่ยว [7] ถึงแม้ว่าตัวแบบผสมจะมีประสิทธิภาพในการพยากรณ์มากกว่าตัวแบบอนุกรมเวลาเชิงเดี่ยวก็ตาม แต่หากข้อมูลอนุกรมเวลาที่นำมาพยากรณ์มีความถี่หรือมีความกว้าง (Fast Oscillation) ของข้อมูลมาก ตัวแบบผสมที่นำมาใช้พยากรณ์อาจมีความคลาดเคลื่อนสูง การกรองข้อมูลเพื่อลดความถี่ หรือสัญญาณแกว่งไกวของข้อมูลก่อนนำไปพยากรณ์จะทำให้ค่าพยากรณ์มีความถูกต้องแม่นยำมากขึ้น วิธีการแยกองค์ประกอบข้อมูลเชิงประจักษ์ (Empirical Mode Decomposition; EMD) เป็นวิธีที่มีการประยุกต์ใช้เพื่อลดความถี่ของข้อมูลอนุกรมเวลาอย่างหลากหลาย เช่น ประยุกต์ใช้ในการพยากรณ์ความเร็วลม [8], [9] การพยากรณ์ปริมาณการ

ใช้พลังงานไฟฟ้า [10] ค่าพยากรณ์ที่ได้จากการประยุกต์ใช้วิธีการแยกองค์ประกอบข้อมูลเชิงประจักษ์ ทำให้ค่าพยากรณ์มีความแม่นยำมากขึ้น

จากเหตุผลข้างต้นผู้วิจัยจึงนำแนวคิดดังกล่าวมาสร้างตัวแบบผสมขึ้นเพื่อพยากรณ์อนุกรมเวลาในการศึกษาครั้งนี้ โดยนำเอาเทคนิคการแยกองค์ประกอบด้วยข้อมูลเชิงประจักษ์มาประยุกต์ใช้เพื่อการลดสัญญาณแกว่งไกวของข้อมูลอนุกรมเวลาก่อนนำไปพยากรณ์ด้วยตัวแบบอนุกรมเวลาเชิงเดี่ยว 2 วิธีคือ ตัวแบบ HWES และตัวแบบ SARIMA เนื่องจากตัวแบบทั้ง 2 มีประสิทธิภาพในการพยากรณ์ได้ดีและนิยมใช้ในปัจจุบัน [11] จากนั้นนำมาสร้างตัวแบบผสม โดยรวมตัวแบบ HWES และตัวแบบ SARIMA เข้าด้วยกัน และให้น้ำหนักที่เหมาะสม จากนั้นนำไปทดลองใช้กับข้อมูลอนุกรมเวลาหน่วยจำหน่ายไฟฟ้าประเภทบ้านอยู่อาศัยภาคตะวันออกเฉียงเหนือของประเทศไทย และนำไปเปรียบเทียบกับประสิทธิภาพการพยากรณ์กับอนุกรมเวลาเชิงเดี่ยว ด้วยเกณฑ์วัดประสิทธิภาพ 5 เกณฑ์ ได้แก่ 1) ความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Error; MAE) 2) รากของความคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Square Error; RMSE) 3) ร้อยละความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error; MAPE) 4) เปอร์เซ็นต์ของมัธยฐานความคลาดเคลื่อนสัมบูรณ์ (Median Absolute Percentage Error; MdAPE) และ 5) เปอร์เซ็นต์ของความคลาดเคลื่อนสัมบูรณ์แบบสมมาตร (Symmetric Mean Absolute Percentage Error; sMAPE) เพื่อค้นหาตัวแบบที่เหมาะสมที่สุดสำหรับการพยากรณ์ ผลการวิจัยที่ได้จะเป็นประโยชน์แก่หน่วยงานที่เกี่ยวข้อง ได้แก่ การไฟฟ้าส่วนภูมิภาคสามารถนำตัวแบบที่ผู้วิจัยพัฒนาขึ้นไปใช้ในการพยากรณ์ปริมาณหน่วยจำหน่ายไฟฟ้าบ้านอยู่อาศัยรายเดือนของประชาชนในเขตภาคตะวันออกเฉียงเหนือล่วงหน้า เพื่อประกอบการวางแผนและบริหารจัดการพลังงานไฟฟ้าให้เพียงพอต่อความต้องการของประชาชนในอนาคต และงานวิจัยชิ้นนี้ยังใช้เป็นแนวทางเพื่อต่อยอดสำหรับการเพิ่มประสิทธิภาพการพยากรณ์อนุกรมเวลาได้



รูปที่ 1 ขั้นตอนการสร้างตัวแบบพยากรณ์แบบผสม

## 2. วิธีการวิจัย

### 2.1 การจัดเตรียมข้อมูล

2.1.1 ข้อมูลที่ใช้ในการทดลองครั้งนี้เป็นข้อมูลอนุกรมเวลาปริมาณหน่วยจำหน่ายไฟฟ้าบ้านอยู่อาศัยของการไฟฟ้าส่วนภูมิภาค [12] เป็นข้อมูลทุติยภูมิจากรายงานสถานการณ์การจำหน่ายไฟฟ้าประจำเดือน ภาคตะวันออกเฉียงเหนือ จำนวน 3 ชุดข้อมูล ชุดละ 127 ค่า ตั้งแต่เดือนมกราคม 2552 ถึงเดือนกรกฎาคม 2562 ตามเขตน้อย คือ กพล.1 อุตรธานี รับผิดชอบการบริการในพื้นที่ 8 จังหวัด ได้แก่ อุตรธานี ขอนแก่น หนองบัวลำภู เลย หนองคาย บึงกาฬ นครพนม และสกลนคร กพล.2 อุบลราชธานี รับผิดชอบการบริการใน

พื้นที่ 8 จังหวัด ได้แก่ อุบลราชธานี อำนาจเจริญ มุกดาหาร กาฬสินธุ์ มหาสารคาม ร้อยเอ็ด ยโสธร และศรีสะเกษ กพล.3 นครราชสีมา รับผิดชอบการบริการในพื้นที่ 4 จังหวัด ได้แก่ นครราชสีมา ชัยภูมิ บุรีรัมย์ และสุรินทร์ ดังแสดงในรูปที่ 1 ผู้วิจัยได้แบ่งข้อมูลออกเป็น 2 ชุด คือ ข้อมูลชุดที่ 1 ตั้งแต่เดือนมกราคม 2552 ถึงเดือนธันวาคม 2561 จำนวน 120 ค่า ใช้สำหรับสร้างตัวแบบ และข้อมูลชุดที่ 2 ตั้งแต่เดือนมกราคม 2562 ถึงเดือนกรกฎาคม 2562 จำนวน 7 ค่า ใช้สำหรับการทดสอบตัวแบบ ซึ่งเกณฑ์ในการแบ่งข้อมูลเพื่อวัดประสิทธิภาพการพยากรณ์ (Cross-Validation) ข้างต้นใช้เกณฑ์ร้อยละ 95 และ 5 ตามลำดับ

2.1.2 การศึกษาความเคลื่อนไหวของข้อมูลอนุกรมเวลา การทดสอบความเคลื่อนไหวของข้อมูล เพื่อเป็นการพิจารณาว่าข้อมูลอนุกรมเวลาที่นำมาศึกษานั้นมีลักษณะแบบใด โดยการทดสอบแนวโน้มและความผันแปรของฤดูกาล โดยพิจารณาจากกราฟแสดงความสัมพันธ์ระหว่างค่าสัมประสิทธิ์สหสัมพันธ์แบบอัตโนมัติ ( $r_k$ ) และช่วงเวลาห่าง  $k$  การทดสอบว่าอนุกรมเวลามีลักษณะการเคลื่อนไหวแบบใดแบบหนึ่ง หรือการทดสอบ  $H_0 : \rho_k = 0$  มีตัวทดสอบสถิติเป็นค่าสัมประสิทธิ์สหสัมพันธ์แบบอัตโนมัติที่ช่วงเวลาห่างกันดังสมการที่ (1) ต่อไปนี้

$$r_k = \frac{\sum_{i=k+1}^n (y_i - \bar{y})(y_{i-k} - \bar{y})}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (1)$$

เมื่อ  $y_i$  แทนค่าที่แท้จริงในช่วงเวลา  $t$ ;  $t = 0, 1, 2, 3, \dots, n$   
 $k = 0, 1, 2, 3, \dots, \bar{y}$  แทน ค่าเฉลี่ย

## 2.2 รูปแบบการพยากรณ์

2.2.1 วิธีการวิเคราะห์ข้อมูลอนุกรมเวลาเชิงประจักษ์ (EMD) เป็นวิธีการแยกความถี่ของข้อมูลอนุกรมเวลาที่ไม่เป็นเชิงเส้นตรง และมีลักษณะเคลื่อนไหวไม่คงที่ (Non - Stationary) ออกจากกัน นำเสนอโดย Huang และคณะ [13] เป็นการลดสัญญาณแกว่งไกวของข้อมูล โดยแยกออกเป็นฟังก์ชันภายใน (Intrinsic Mode Function; IMF) มีเงื่อนไข ดังนี้ 1) จำนวนจุดปลายสุด (Extrema Points) ซึ่งจะประกอบด้วย จุดสูงสุดและต่ำสุด และจำนวนจุดตัดแกนศูนย์จะต้องมีค่าเท่ากันหรือมีค่าต่างกันไม่เกินหนึ่ง 2) ณ จุดใดๆ ของสัญญาณค่าเฉลี่ยของขอบบนและขอบล่างจะมีค่าเท่ากับศูนย์ กระบวนการทำงานเพื่อสกัดชุดฟังก์ชันภายใน (IMFs) ของการแยกองค์ประกอบด้วยวิธีเชิงประจักษ์ (EMD) ประกอบด้วย 2 ขั้นตอน คือ

1) การตรวจจับจุดปลายสุดของสัญญาณ (Extrema Detection) โดยการหาจุดสูงสุดและจุดต่ำสุดของแต่ละช่วงสัญญาณย่อย

2) การประมาณค่าขอบเขตพื้นผิว (Surface Interpolation) ของสัญญาณโดยการหาค่าประมาณของสัญญาณขอบบน

และสัญญาณขอบล่าง แล้วคำนวณหาสัญญาณเฉลี่ยตามสมการที่ (2) จากนั้นนำสัญญาณต้นฉบับลบออกด้วยสัญญาณเฉลี่ยจะได้สัญญาณผลลัพธ์ ขั้นตอนต่อไปตรวจสอบว่าสัญญาณผลลัพธ์นั้นเป็นฟังก์ชันภายใน (IMF) หรือไม่ ถ้าผลปรากฏว่าสัญญาณผลลัพธ์ที่ได้ไม่เป็นฟังก์ชันภายใน (IMF) ให้แทนที่สัญญาณต้นฉบับด้วยสัญญาณผลลัพธ์ แล้วทำซ้ำกระบวนการดังกล่าวข้างต้น หากสัญญาณผลลัพธ์ที่ได้เป็นฟังก์ชันภายใน (IMF) ให้จัดเก็บสัญญาณผลลัพธ์นั้น จากนั้นทำการสกัดฟังก์ชันภายใน (IMF) ลำดับถัดไป จากสัญญาณต้นฉบับใหม่ที่ได้จากการนำสัญญาณต้นฉบับเดิมลบด้วยฟังก์ชันภายใน (IMF) แล้วทำกระบวนการข้างต้น ซ้ำจนสัญญาณต้นฉบับเป็นสัญญาณฟังก์ชันเดียวดังสมการที่ (2)

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

เมื่อ  $m(t)$  แทน สัญญาณเฉลี่ย  $e_{\max}$ ,  $e_{\min}$  แทนสัญญาณขอบบนและสัญญาณขอบล่างตามลำดับ และสามารถเขียนให้อยู่ในรูปสมการที่ (3) ได้ดังนี้

$$x(t) = \sum_{i=1}^n h_i(t) + r_N(t) \quad (3)$$

เมื่อ  $x(t)$  แทน สัญญาณต้นฉบับ  $h_i(t)$ ;  $i = 0, 1, 2, 3, \dots, N$  เป็นชุดฟังก์ชันภายใน (IMFs) และ  $r_N(t)$  เป็นส่วนเหลือ (Residual)

### 2.2.2 การลดสัญญาณแกว่งไกวของข้อมูล

จากการวิเคราะห์ห่องค์ประกอบเชิงประจักษ์ (EMD) ในสมการที่ (2) จะได้ฟังก์ชันภายใน (IMF<sub>*i*</sub>) ของข้อมูล  $h_i(t)$  จากนั้นนำ  $h_i(t)$  มาหาค่าเฉลี่ยโดยไม่รวมส่วนเหลือ ( $r_N(t)$ ) จากสมการที่ (4)

$$aIMF = \frac{\sum_{i=1}^n h_i(t)}{n} \quad (4)$$

นำ  $aIMF$  ลบสัญญาณต้นฉบับดังสมการที่ (5)

$$Y_n(t) = x(t) - aIMF \quad (5)$$

เมื่อ  $y_n(t)$  คือ ชุดข้อมูลใหม่ที่มีการลดข้อมูลสัญญาณกว้างไว้

### 2.2.3 ตัวแบบอาร์มา

ตัวแบบอาร์มาเป็นตัวแบบการพยากรณ์ที่มีความแม่นยำสูงพัฒนาขึ้นโดยบอกซ์เจนกินส์ [14] สามารถเขียนสมการทางคณิตศาสตร์ของตัวแบบอาร์มาแบบมีฤดูกาล ดังสมการที่ (6)–(11)

$$\phi_p(B)\Phi_p(B^s)(1-B)^d(1-B^s)^D Z_t = \delta + \theta_q(B)\Theta_q(B^s)\varepsilon_t \quad (6)$$

$$\delta = \mu\phi_p(B)\Phi_p(B^s) \quad (7)$$

$$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \quad (8)$$

$$\Phi_p(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \dots - \Phi_p B^{ps} \quad (9)$$

$$\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q \quad (10)$$

$$\Theta_q(B^s) = 1 - \Theta_1 B^s - \Theta_2 B^{2s} - \dots - \Theta_q B^{qs} \quad (11)$$

เมื่อ  $Z_t$  แทน อนุกรมเวลา ณ เวลา  $t$

$\varepsilon_t$  แทน ความคลาดเคลื่อนที่มีการแจกแจงแบบปกติ มีค่าเฉลี่ยเท่ากับศูนย์มีความแปรปรวนคงที่และเป็นอิสระกัน ( $\varepsilon_t \sim NID(0, \sigma^2)$ )

$t$  แทน คาบเวลาซึ่งมีค่าตั้งแต่ 1 ถึง  $n$

$s$  แทน ความยาวของคาบฤดูกาลใน 1 ฤดูกาล

$d$  และ  $D$  แทน ลำดับที่ของการหาผลต่างภายในฤดูกาลและผลต่างระหว่างฤดูกาล

$B$  แทน ตัวดำเนินการถอยหลัง (Backward Operator) โดยที่  $B^s Z_t = Z_{t-s}$  และ  $BZ_t = Z_{t-1}$

$\phi_p(B)$  แทน ตัวดำเนินการสหสัมพันธ์ในตนเองแบบไม่มีฤดูกาลอันดับที่  $p$

$\Phi_p(B^s)$  แทน ตัวดำเนินการสหสัมพันธ์ในตัวเองแบบมีฤดูกาลอันดับที่  $P$

$\theta_q(B)$  แทน ตัวดำเนินการเฉลี่ยเคลื่อนที่แบบภายใน

ฤดูกาลอันดับที่  $q$

$\Theta_q(B^s)$  แทน ตัวดำเนินการเฉลี่ยเคลื่อนที่ระหว่างฤดูกาลอันดับที่  $Q$

การนำข้อมูลอนุกรมเวลาไปสร้างตัวแบบผู้วิจัยได้ทำการลดสัญญาณกว้างไวของข้อมูลโดยใช้เทคนิคการแยกองค์ประกอบเชิงประจักษ์ และนำไปสร้างตัวแบบ SARIMA โดยกำหนดรูปแบบ SARIMA(p,d,q)(P,D,Q) การดำเนินการผู้วิจัยเขียนคำสั่งโดยใช้โปรแกรมภาษาอาร์มีฟังก์ชัน `auto.arima()` ซึ่งอยู่ใน Package “forecast” [15] ที่ใช้ในการระบุรูปแบบ ARIMA(p,d,q)(P,D,Q) ที่เหมาะสม ขั้นตอนการสร้างตัวแบบ 4 ขั้นตอน ดังนี้คือ

**ขั้นตอนที่ 1** พิจารณาอนุกรมเวลาว่าคงที่หรือไม่โดยพิจารณาจากกราฟของอนุกรมเวลาชุดนั้น หากกราฟอนุกรมเวลาชุดนั้นมีแนวโน้มเป็นส่วนประกอบกล่าวคือ ข้อมูลลักษณะเพิ่มขึ้นหรือลดลงอย่างเป็นสัดส่วนกับเวลา และมีการกระจายไม่สม่ำเสมอมีลักษณะบานออกเป็นปากแตร แสดงว่าข้อมูลอนุกรมเวลาชุดนั้นมีสถานะไม่คงที่นั่นคือ มีค่าเฉลี่ยไม่คงที่และความแปรปรวนไม่คงที่ งานวิจัยนี้จะทำการทดสอบด้วย Augmented Dickey – Fuller (ADF) หากผลการทดสอบไม่มีนัยสำคัญทางสถิติแสดงว่า ข้อมูลอนุกรมเวลาชุดนั้นมีลักษณะไม่คงที่ เมื่อทราบว่าข้อมูลมีลักษณะไม่คงที่ที่จะต้องทำการแปลงข้อมูลให้มีลักษณะคงที่ โดยการหาลอการิทึมของข้อมูลก่อน และจึงหาผลต่างของข้อมูลชุดดังกล่าว เพื่อให้ข้อมูลอนุกรมเวลามีลักษณะคงที่ก่อนที่จะนำไปกำหนดตัวแบบ

**ขั้นตอนที่ 2** กำหนดตัวแบบพยากรณ์

**ขั้นตอนที่ 3** ประเมินค่าพารามิเตอร์และตรวจสอบความเหมาะสมของตัวแบบ (Diagnostics Checking) มีทั้งหมด 4 ข้อ คือ 1) ทดสอบความคลาดเคลื่อนมีค่าเฉลี่ยเท่ากับศูนย์ โดยใช้สถิติทดสอบ t-test 2) ทดสอบความคลาดเคลื่อนมีการแจกแจงแบบปกติโดยใช้สถิติทดสอบ Kolmogorov – Smirnov 3) ทดสอบความเป็นอิสระกันของความคลาดเคลื่อนโดยใช้สถิติทดสอบ Ljung – Box Q- statistics 4) ความคลาดเคลื่อนมีความแปรปรวนคงที่โดยพิจารณาจากการพล็อตกราฟดูลักษณะการกระจายระหว่างความคลาดเคลื่อนกับ

ค่าพยากรณ์หากกราฟมีการกระจายไม่มีรูปแบบแสดงว่าความคลาดเคลื่อนมีความแปรปรวนคงที่ ถ้าตัวแบบผ่านการตรวจสอบตัวแบบครบทั้ง 4 ข้อ จึงสามารถนำตัวแบบไปใช้ในการพยากรณ์ในขั้นตอนที่ 4 ต่อไป

**ขั้นตอนที่ 4** นำตัวแบบที่เหมาะสมไปใช้ในการพยากรณ์ล่วงหน้า

2.2.4 การพยากรณ์ด้วยวิธีการปรับให้เรียบเอ็กซ์โพเนนเชียลแบบโฮลท์ – วินเทอร์

วิธีการนี้เป็นการพยากรณ์ข้อมูลอนุกรมเวลาที่มีองค์ประกอบของแนวโน้มและฤดูกาลแบ่งออกเป็น 2 กรณีคือตัวแบบการพยากรณ์แบบบวกและตัวแบบการพยากรณ์แบบคูณ ในการวิจัยครั้งนี้ใช้ตัวแบบคูณ มีจำนวนพารามิเตอร์อยู่ทั้งหมด 3 ตัว คือ  $\alpha$ ,  $\beta$  และ  $\gamma$  ค่าพารามิเตอร์จะมีค่าอยู่ระหว่าง 0-1 มีตัวแบบการพยากรณ์ดังสมการที่ (12)

$$L_t = \alpha \frac{Y_t}{S_{t-s}} + (1-\alpha)(L_{t-1} + T_{t-1}) \quad (12)$$

ประมาณค่าแนวโน้ม  $T_t = \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1}$

การประมาณค่า  $S_t = \gamma\left(\frac{Y_t}{L_t}\right) + (1-\gamma)S_{t-s}$

สมการพยากรณ์  $\hat{Y}_{t+p} = (L_t + pT_t)S_{t-s+p}$

เมื่อ  $L_t$  แทน ค่าที่ปรับให้เรียบหรือการประมาณระดับปัจจุบัน

$\alpha$  แทน ค่าคงที่ใช้ในการปรับให้เรียบสำหรับระดับ

$Y_t$  แทน ค่าจริงในช่วงเวลา

$\beta$  แทน ค่าคงที่ในการปรับให้เรียบของแนวโน้ม

$T_t$  แทน การประมาณค่าแนวโน้ม

$\gamma$  แทน การปรับให้เรียบตามฤดูกาล

$S_t$  แทน การประมาณฤดูกาล

$p$  แทน ระยะเวลาที่จะพยากรณ์ในอนาคต

$S$  แทน จำนวนฤดูกาล

$\hat{Y}_{t+p}$  แทน การพยากรณ์สำหรับระยะเวลา  $p$  ในอนาคต การนำข้อมูลอนุกรมเวลาไปสร้างตัวแบบผู้วิจัยได้

ทำการลดสัญญาณแกว่งไกวของข้อมูลโดยใช้เทคนิคการแยกองค์ประกอบเชิงประจักษ์ ก่อนนำไปสร้างตัวแบบ จากนั้นจะดำเนินการตรวจสอบคุณลักษณะของความคลาดเคลื่อนจากการพยากรณ์ตามขั้นตอนที่ 3 ของการสร้างตัวแบบการอาร์มาแบบมีฤดูกาล

## 2.2.5 ตัวแบบผสม

การพัฒนาตัวแบบผสมที่นำเสนอในงานวิจัยนี้ จะใช้เทคนิคการพยากรณ์ร่วม ซึ่งมีแนวคิดที่ว่าหากข้อมูลอนุกรมเวลาที่นำมาพยากรณ์เหมาะกับการพยากรณ์เชิงเดี่ยว ตั้งแต่ 2 วิธีขึ้นไป โดยกำหนดค่าน้ำหนักที่เหมาะสมกับค่าพยากรณ์เชิงเดี่ยว จะทำให้ได้ค่าพยากรณ์ที่มีความคลาดเคลื่อนน้อยลง [6] แต่ตัวแบบผสมที่ผู้วิจัยเสนอมีความแตกต่างกับเทคนิคการพยากรณ์ร่วมคือ ก่อนที่จะนำไปพยากรณ์เชิงเดี่ยว ผู้วิจัยจะทำการกรองข้อมูลด้วยวิธีการวิเคราะห์อนุกรมเวลาเชิงประจักษ์ก่อนนำไปพยากรณ์ เพื่อเป็นการลดสัญญาณแกว่งไกวของข้อมูลซึ่งคาดว่าจะทำให้ค่าพยากรณ์มีความแม่นยำมากขึ้น วิธีการพยากรณ์เชิงเดี่ยวที่พิจารณาในงานวิจัยนี้ มี 2 วิธีคือ วิธีปรับให้เรียบเอ็กซ์โพเนนเชียลแบบโฮลท์ – วินเทอร์ และวิธีการพยากรณ์ด้วยตัวแบบอาร์มาแบบมีฤดูกาล โดยมีตัวแบบผสมดังสมการที่ (13) ต่อไปนี้

$$\hat{Y}_t = \beta_1 F_{1t} + \beta_2 F_{2t} + \dots + \beta_i F_{it} \quad (13)$$

เมื่อ  $B_i$  แทน ค่าถ่วงน้ำหนักของแต่ละเทคนิคการพยากรณ์ที่  $i$  ด้วยวิธีกำลังสองน้อยที่สุด

$F_{it}$  แทน ค่าพยากรณ์สำหรับเทคนิคการพยากรณ์ที่  $i$  ในช่วงเวลา  $t$  เมื่อมีการกรองข้อมูลด้วยวิธีการวิเคราะห์ข้อมูลเชิงประจักษ์ แสดงดังรูปที่ 1

## 2.2.5 เกณฑ์การประเมินประสิทธิภาพของตัวแบบ

1) ความคลาดเคลื่อนสัมบูรณ์เฉลี่ยดังสมการที่ (14)

$$MAE = \sum_{t=1}^n |Y_t - \hat{Y}_t| / n \quad (14)$$

2) รากของความคลาดเคลื่อนกำลังสองเฉลี่ย ดังสมการที่ (15)

$$RMSE = \sqrt{\sum_{t=1}^n (Y_t - \hat{Y}_t)^2 / n} \quad (15)$$

3) ร้อยละความคลาดเคลื่อนสัมบูรณ์เฉลี่ยดังสมการที่ (16)

$$MAPE = 100 \times \sum_{t=1}^n |1 - \hat{Y}_t / Y_t| / n \quad (16)$$

4) เปอร์เซ็นต์ของมัธยฐานความคลาดเคลื่อนสัมบูรณ์ดังสมการที่ (17)

$$MdAPE = \text{median} \sum_{t=1}^n \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right| \times 100 \quad (17)$$

5) เปอร์เซ็นต์ของความคลาดเคลื่อนสัมบูรณ์เฉลี่ยแบบสมมาตรดังสมการที่ (18)

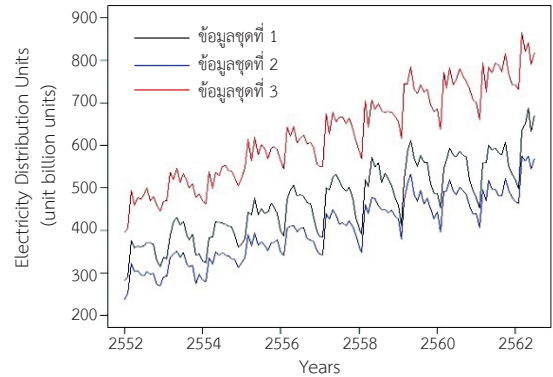
$$sMAPE = \sum_{t=1}^n \frac{2 \times |Y_t - \hat{Y}_t| / |Y_t + \hat{Y}_t|}{n} \times 100 \quad (18)$$

โดยที่ค่า MAE, RMSE, MAPE, MdAPE และ sMAPE ยิ่งน้อยก็ยิ่งแสดงว่าตัวแบบมีประสิทธิภาพสูง

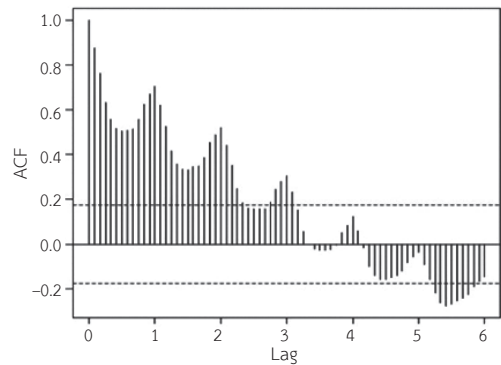
### 3.ผลการวิจัย

#### 3.1 ผลการวิเคราะห์ลักษณะการเคลื่อนไหวของอนุกรมเวลา

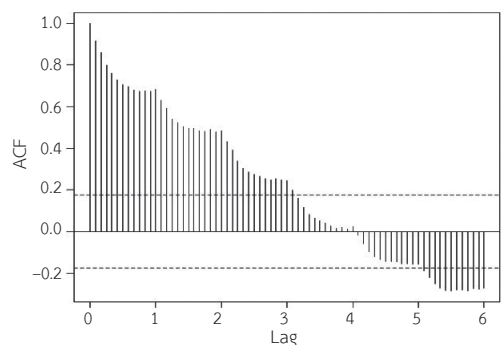
จากการพิจารณาลักษณะการเคลื่อนไหวของอนุกรมเวลาปริมาณหน่วยจำหน่ายไฟฟ้าประเภทบ้านอยู่อาศัยทั้ง 3 เขต คือ กฟผ.1 กฟผ.2 และ กฟผ.3 ย้อนหลัง 10 ปี 7 เดือน คือ ตั้งแต่เดือนมกราคม 2552 ถึงเดือนกรกฎาคม 2562 จำนวน 127 ค่า ดังรูปที่ 2 พบว่า อนุกรมเวลาทั้ง 3 ชุด มีแนวโน้มและความผันแปรตามฤดูกาลเป็นส่วนประกอบ และผลจากการทดสอบองค์ประกอบแนวโน้มและความผันแปรตามฤดูกาล โดยพิจารณาจากกราฟแสดงความสัมพันธ์ระหว่างค่าสัมประสิทธิ์สหสัมพันธ์แบบอัตโนมัติและค่าระยะห่าง จากรูปที่ 3-5 พบว่า ข้อมูลอนุกรมเวลาทั้ง 3 ชุด ข้อมูลมีองค์ประกอบของแนวโน้มและฤดูกาลที่ชัดเจน เนื่องจาก  $r_k$  มีค่ามากและเป็นบวก  $k$  มีค่าน้อย และลดลงอย่างช้าๆ เมื่อ  $k$  มีค่าเพิ่มขึ้น



รูปที่ 2 ลักษณะการเคลื่อนไหวของอนุกรมเวลาปริมาณหน่วยจำหน่ายไฟฟ้าประเภทบ้านอยู่อาศัยทั้ง 3 เขต

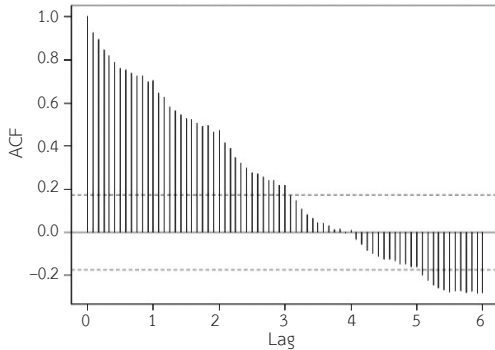


รูปที่ 3 ความสัมพันธ์ระหว่างค่าสัมประสิทธิ์สหสัมพันธ์แบบอัตโนมัติและค่าของปริมาณหน่วยจำหน่ายไฟฟ้าประเภทบ้านอยู่อาศัยของ กฟผ.1



รูปที่ 4 ความสัมพันธ์ระหว่างค่าสัมประสิทธิ์สหสัมพันธ์แบบอัตโนมัติและค่าของปริมาณหน่วยจำหน่ายไฟฟ้าประเภทบ้านอยู่อาศัยของ กฟผ.2





รูปที่ 5 ความสัมพันธ์ระหว่างค่าสัมประสิทธิ์สหสัมพันธ์แบบบอโตและค่าของปริมาณหน่วยจำหน่ายไฟฟ้าประเภทบ้านอยู่อาศัยของ กฟผ.3

### 3.2 ผลการวิเคราะห์ข้อมูลค่าสถิติพื้นฐาน

ผลการวิเคราะห์ค่าสถิติพื้นฐานพบว่า ปริมาณหน่วยจำหน่ายไฟฟ้าประเภทบ้านอยู่อาศัยภาคตะวันออกเฉียงเหนือ กฟผ.1 เฉลี่ยต่อเดือนเท่ากับ 470.65 ล้านหน่วยต่อเดือน ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 88.74 ล้านหน่วยต่อเดือน มีหน่วยจำหน่ายไฟฟ้าสูงสุดเท่ากับ 687.58 ล้านหน่วยต่อเดือน ต่ำสุด 281.92 ล้านหน่วยต่อเดือน กฟผ.2 มีปริมาณหน่วยจำหน่ายไฟฟ้าเฉลี่ยต่อเดือนเท่ากับ 402.47 ล้านหน่วยต่อเดือน ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 78.61 ล้านหน่วยต่อเดือน มี

หน่วยจำหน่ายไฟฟ้าสูงสุดเท่ากับ 574.94 ล้านหน่วยต่อเดือน ต่ำสุด 237.29 ล้านหน่วยต่อเดือน กฟผ.3 มีปริมาณหน่วยจำหน่ายไฟฟ้าเฉลี่ยต่อเดือนเท่ากับ 628.43 ล้านหน่วยต่อเดือน ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 109.95 ล้านหน่วยต่อเดือน มีหน่วยจำหน่ายไฟฟ้าสูงสุดเท่ากับ 864.92 ล้านหน่วยต่อเดือน ต่ำสุด 394.42 ล้านหน่วยต่อเดือน รายละเอียดดังตารางที่ 1

### 3.3 ผลพยากรณ์โดยใช้ตัวแบบอาร์มาแบบมีฤดูกาล

3.3.1 จากการพิจารณาข้อมูลปริมาณหน่วยจำหน่ายไฟฟ้าประเภทบ้านอยู่อาศัยภาคตะวันออกเฉียงเหนือทั้ง 3 เขต (ดังรูปที่ 2) พบว่า กราฟมีแนวโน้มเป็นส่วนประกอบโดยมีลักษณะการกระจายไม่สม่ำเสมอลักษณะของกราฟบานออกเป็นปากแตรแสดงว่า ข้อมูลมีลักษณะไม่คงที่ สอดคล้องกับผลการทดสอบความสัมพันธ์ระหว่างค่าสัมประสิทธิ์สหสัมพันธ์แบบบอโต (รูปที่ 3-5) ซึ่งพบว่า ค่า  $r_k$  มีค่ามากและเป็นบวก  $k$  มีค่าน้อย และลดลงอย่างช้าๆ เมื่อ  $k$  มีค่าเพิ่มขึ้น และผลการทดสอบความคงที่ของข้อมูลด้วย Augmented Dickey Fuller (ADF) (ตารางที่ 2) พบว่า ค่า ADF ของข้อมูลทั้ง 3 เขต ไม่นับสำคัญทางสถิติ ( $p$ -value > 0.05) แสดงว่า ข้อมูลทั้ง 3 เขต มีแนวโน้มและฤดูกาลเป็นส่วนประกอบ เมื่อทราบว่า ข้อมูลอนุกรมเวลาทั้ง 3 เขต มีลักษณะไม่คงที่ ผู้วิจัยจึงทำการ

ตารางที่ 1 ค่าสถิติพื้นฐานปริมาณหน่วยจำหน่ายไฟฟ้าประเภทบ้านอยู่อาศัยภาคตะวันออกเฉียงเหนือทั้ง 3 เขตข้อมูล (หน่วย : ล้านหน่วย)

ชุดข้อมูล	ข้อมูล	จำนวน	ค่าเฉลี่ย	ค่าสูงสุด	ค่ามัธยฐาน	ค่าต่ำสุด	ส่วนเบี่ยงเบนมาตรฐาน
กฟผ.1	All sample	127	470.65	687.58	475.41	281.92	88.74
	Training	120	462.32	611.06	463.16	281.92	82.41
	Testing	7	613.42	687.58	634.03	498.88	73.80
กฟผ.2	All sample	127	402.47	574.94	405	237.29	78.61
	Training	120	394.65	531.89	394.89	237.29	72.80
	Testing	7	536.63	574.94	563.02	463.45	49.42
กฟผ.3	All sample	127	628.43	864.92	623.53	394.42	109.95
	Training	120	618.34	815.16	616.54	394.42	104.00
	Testing	7	801.30	864.92	817.56	731.77	49.31

แปลงข้อมูลโดยการหาลอการิทึมก่อน แล้วจึงหาผลต่างของชุดข้อมูลก่อนนำไปสร้างตัวแบบ SARIMA ผลการวิเคราะห์ข้อมูล ADF (ดังตารางที่ 2) เมื่อมีการแปลงข้อมูลพบว่าหลังแปลงข้อมูลอนุกรมเวลาทั้ง 3 ชุด ข้อมูลมีลักษณะคงที่ โดยมีค่า ADF มีนัยสำคัญทางสถิติทั้ง 3 ชุด ( $p\text{-value} < 0.05$ ) เมื่อทราบว่าข้อมูลมีลักษณะคงที่ผู้วิจัยจะนำไปใช้ในการกำหนดตัวแบบในลำดับต่อไป

**ตารางที่ 2** ค่า ADF ปริมาณหน่วยจำหน่ายไฟฟ้าประเภทบ้านอยู่อาศัยภาคตะวันออกเฉียงเหนือทั้ง 3 เขต

ชุดข้อมูล	ก่อนแปลงข้อมูล		หลังแปลงข้อมูล	
	ADF	p-value	ADF	p-value
กพฉ.1	0.9056	0.9009	-7.3681	<0.01
กพฉ.2	1.2871	0.9496	-8.6256	<0.01
กพฉ.3	1.6598	0.9758	-8.8500	<0.01

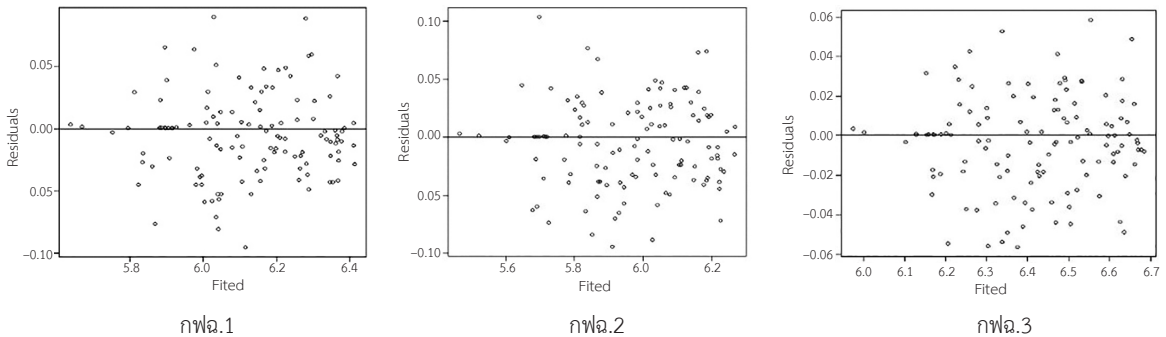
3.3.2 กำหนดรูปแบบของตัวแบบ SARIMA ผลการวิเคราะห์ข้อมูลดังปรากฏตารางที่ 3

3.3.3 การตรวจสอบความเหมาะสมของตัวแบบ โดย

พิจารณาตามข้อตกลงเบื้องต้นของตัวแบบ SARIMA ซึ่งมีทั้งหมด 4 ข้อ พบว่า ค่าความคลาดเคลื่อนมีค่าเฉลี่ยเท่ากับศูนย์ (ข้อมูลชุดที่ 1 มีค่า  $t = -1.609$ ,  $p\text{-value} = 0.110$  ข้อมูลชุดที่ 2 มีค่า  $t = -1.704$ ,  $p\text{-value} = 0.091$  และข้อมูลชุดที่ 3 มีค่า  $t = -1.856$ ,  $p\text{-value} = 0.065$ ) ค่าความคลาดเคลื่อนมีการแจกแจงแบบปกติ (ข้อมูลชุดที่ 1 มีค่า  $KS = 0.081$ ,  $p\text{-value} = 0.413$  ข้อมูลชุดที่ 2 มีค่า  $KS = 0.040$ ,  $p\text{-value} = 0.991$  และข้อมูลชุดที่ 3 มีค่า  $KS = 0.068$ ,  $p\text{-value} = 0.634$ ) มีค่าความคลาดเคลื่อนเป็นอิสระกัน (ข้อมูลชุดที่ 1 มีค่า  $Q\text{-statistics} = 19.555$ ,  $p\text{-value} = 0.549$  ข้อมูลชุดที่ 2 มีค่า  $Q\text{-statistics} = 20.206$ ,  $p\text{-value} = 0.508$  และข้อมูลชุดที่ 3 มีค่า  $Q\text{-statistics} = 14.712$ ,  $p\text{-value} = 0.837$ ) ความคลาดเคลื่อนมีความแปรปรวนคงที่โดยพิจารณาจากรูปที่ 6 พบว่า ค่าความคลาดเคลื่อนมีการกระจายไม่มีรูปแบบซึ่งสามารถกล่าวได้ว่า ค่าความคลาดเคลื่อนมีลักษณะคงที่ ผลการทดสอบแสดงดังตารางที่ 4 และรูปที่ 6 จากการตรวจสอบตัวแบบ SARIMA สำหรับพยากรณ์ปริมาณหน่วยจำหน่ายไฟฟ้าประเภทบ้านอยู่อาศัยภาคตะวันออกเฉียงเหนือทั้ง 3 เขต ผ่านข้อตกลงเบื้องต้นทุกข้อ จึงสรุปได้ว่าตัวแบบมีความเหมาะสมที่จะนำไปพยากรณ์ต่อไป

**ตารางที่ 3** การประมาณของค่าตัวแบบ SARIMA ทั้ง 3 ชุดข้อมูล

ชุดข้อมูล		Estimate	Std. Error	z-value	p-value
กพฉ.1 SARIMA(0,1,2)(0,1,1) <sub>12</sub>	MA1	-0.327	0.094	-3.492	<0.001
	MA2	-0.208	0.091	-2.290	0.022
	SMA1	-0.870	0.174	-4.995	<0.001
MAE = 26.509, RMSE = 29.431, MAPE = 4.141, MdAPE = 4.628, sMAPE = 4.245					
กพฉ.2 SARIMA(0,1,2)(0,1,1) <sub>12</sub>	MA1	-0.337	0.099	-3.378	<0.001
	MA2	-0.239	0.105	-2.275	0.023
	SMA1	-0.698	0.096	-7.271	<0.001
MAE = 28.246, RMSE = 31.001, MAPE = 5.149, MdAPE = 6.093, sMAPE = 5.312					
กพฉ.3 SARIMA(0,1,2)(0,1,1) <sub>12</sub>	MA1	-0.1983	0.0960	-2.0668	0.038
	MA2	-0.2401	0.0935	-2.5689	0.010
	SMA1	-0.7434	0.0972	-7.6466	<0.001
MAE = 11.827, RMSE = 14.339, MAPE = 1.483, MdAPE = 1.588, sMAPE = 1.495					



รูปที่ 6 แผนภาพการกระจายระหว่างความคลาดเคลื่อนกับค่าพยากรณ์จากตัวแบบ SARIMA

ตารางที่ 4 ตรวจสอบความคลาดเคลื่อนของตัวแบบ SARIMA ทั้ง 3 ชุดข้อมูล

ชุดข้อมูล	ค่าเฉลี่ยเท่ากับ 0		การแจกแจงแบบปกติ		ความเป็นอิสระกัน		
	t-value	p-value	KS -value	p-value	Ljung – Box Q-statistics		
					Q-statistics	df	p-value
กพฉ.1 SARIMA(0,1,2)(0,1,1) <sub>12</sub>	-1.609	0.110	0.081	0.413	19.555	21	0.549
กพฉ.2 SARIMA(0,1,2)(0,1,1) <sub>12</sub>	-1.704	0.091	0.040	0.991	20.206	21	0.508
กพฉ.3 SARIMA(0,1,2)(0,1,1) <sub>12</sub>	-1.856	0.065	0.068	0.634	14.712	21	0.837

### 3.4 การพยากรณ์ด้วยวิธีการปรับให้เรียบเอ็กซ์โพเนนเชียลแบบไฮลท์ – วินเทอร์

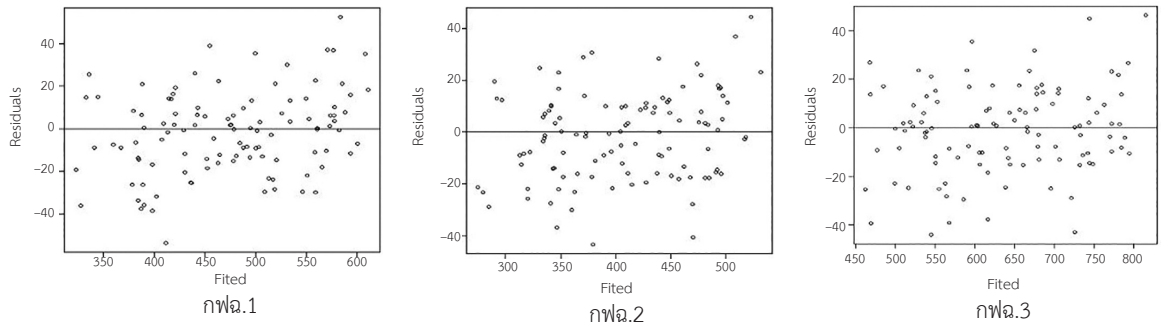
กำหนดรูปแบบของตัวแบบ HWSE ผู้วิจัยประยุกต์ใช้โปรแกรมภาษาอาร์ ซึ่งมีฟังก์ชัน HoltWinters() ซึ่งอยู่ใน Package “forecast” สำหรับกำหนดตัวแบบและกำหนดค่าพารามิเตอร์ที่เหมาะสม ผลการวิเคราะห์ข้อมูลดังตารางที่ 5

การตรวจสอบความเหมาะสมของตัวแบบ โดยพิจารณาตามข้อตกลงเบื้องต้นของตัวแบบ HWSE ซึ่งมีทั้งหมด 4 พบว่า ค่าความคลาดเคลื่อนมีค่าเฉลี่ยเท่ากับศูนย์ (ข้อมูลชุดที่ 1 มีค่า  $t = -0.063$ ,  $p\text{-value} = 0.950$  ข้อมูลชุดที่ 2 มีค่า  $t = -0.089$ ,  $p\text{-value} = 0.929$  และข้อมูลชุดที่ 3 มีค่า  $t = 0.167$ ,  $p\text{-value} = 0.867$ ) ค่าความคลาดเคลื่อนมีการแจกแจงแบบปกติ (ข้อมูลชุดที่ 1 มีค่า  $KS = 0.083$ ,  $p\text{-value} = 0.446$  ข้อมูลชุดที่ 2 มีค่า  $KS = 0.069$ ,  $p\text{-value} = 0.668$  ข้อมูลชุดที่ 3 มีค่า  $KS = 0.077$ ,  $p\text{-value} = 0.544$ ) มีค่าความคลาดเคลื่อนเป็นอิสระกัน (ข้อมูลชุดที่ 1 มีค่า  $Q\text{-statistics} = 30.939$ ,  $p\text{-value} = 0.075$  ข้อมูลชุดที่ 2 มีค่า  $Q\text{-statistics} = 26.513$ ,

$p\text{-value} = 0.1876$  และข้อมูลชุดที่ 3 มีค่า  $Q\text{-statistics} = 18.133$ ,  $p\text{-value} = 0.641$ ) ความคลาดเคลื่อนมีความแปรปรวนคงที่โดยพิจารณาจากรูปที่ 7 พบว่า ค่าความคลาดเคลื่อนมีการกระจายไม่มีรูปแบบซึ่งสามารถกล่าวได้ว่า ค่าความคลาดเคลื่อนมีลักษณะคงที่ จากการตรวจสอบตัวแบบ HWSE ข้างต้นพบว่า ตัวแบบผ่านข้อตกลงเบื้องต้นทุกข้อ แสดงว่า ตัวแบบมีความเหมาะสมที่จะนำไปพยากรณ์ต่อไป ผลการทดสอบแสดงดังตารางที่ 6 และรูปที่ 7

### 3.5 การพยากรณ์ด้วยตัวแบบผสม

เมื่อได้ค่าพยากรณ์ข้อมูลอนุกรมเวลาเชิงเดี่ยวด้วยวิธี HWES และ SARIMA โดยนำค่าทั้ง 2 มาใส่ในน้ำหนักที่เหมาะสมโดยกำหนดค่าถ่วงน้ำหนักด้วยวิธีกำลังสองน้อยที่สุด (ตามขั้นตอนการสร้างตัวแบบในขั้นตอนที่ 2) ในการให้น้ำหนักที่เหมาะสมให้กับค่าพยากรณ์ ผู้วิจัยใช้โปรแกรมภาษาอาร์ซึ่งมี Package “ForecastComb” ซึ่งเป็นฟังก์ชันที่กำหนดน้ำหนักที่เหมาะสมให้กับค่าพยากรณ์ผลการวิเคราะห์ดังตารางที่ 7



รูปที่ 7 แผนภาพการกระจายระหว่างความคลาดเคลื่อนกับค่าพยากรณ์ จากตัวแบบ HESW

ตารางที่ 5 การประมาณของค่าตัวแบบ HWSE กพล. ทั้ง 3 ชุดข้อมูล

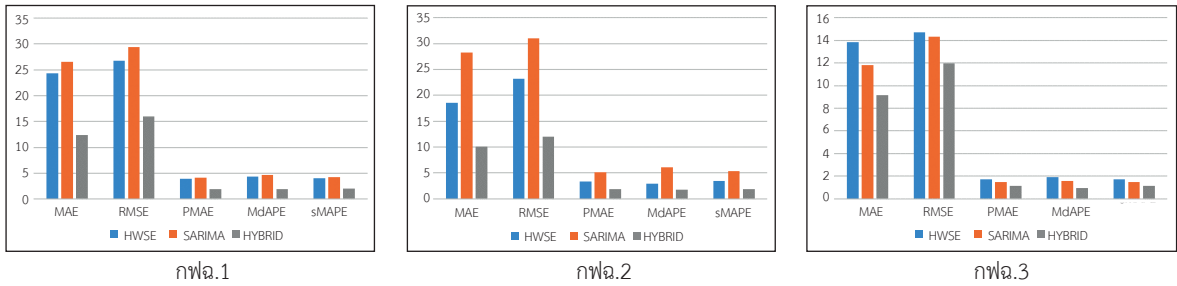
ชุดข้อมูล	สมการ
กพล.1	$\hat{Y}_i(p) = (591.184 + p(2.837)) S_i; i = 1, 2, 3, \dots, 12 \quad p = 1, 2, 3, \dots$ $\alpha = 0.502, \beta = 0.007, \gamma = 0.608 \quad \hat{s}_1 = -57.304, \hat{s}_2 = -73.889, \hat{s}_3 = 8.011, \hat{s}_4 = 19.676, \hat{s}_5 = 45.715, \hat{s}_6 = 27.809,$ $\hat{s}_7 = 21.809, \hat{s}_8 = 14.235, \hat{s}_9 = 5.847, \hat{s}_{10} = 8.495, \hat{s}_{11} = -30.296, \hat{s}_{12} = -46.728$
MAE = 24.380, RMSE = 26.756, MAPE = 3.952, MdAPE = 4.368, sMAPE = 3.997	
กพล.2	$\hat{Y}_i(p) = (511.774 + p(2.388)) S_i; i = 1, 2, 3, \dots, 12 \quad p = 1, 2, 3, \dots$ $\alpha = 0.413, \beta = 0.00, \gamma = 0.541 \quad \hat{s}_1 = -31.974, \hat{s}_2 = -54.677, \hat{s}_3 = 19.645, \hat{s}_4 = 12.658, \hat{s}_5 = 37.226, \hat{s}_6 = 16.263,$ $\hat{s}_7 = 8.256, \hat{s}_8 = 6.639, \hat{s}_9 = -1.280, \hat{s}_{10} = 6.069, \hat{s}_{11} = -14.652, \hat{s}_{12} = -29.421$
MAE = 18.599, RMSE = 23.208, MAPE = 3.349, MdAPE = 2.930, sMAPE = 3.424	
กพล.3	$\hat{Y}_i(p) = (787.248 + p(3.155)) S_i; i = 1, 2, 3, \dots, 12 \quad p = 1, 2, 3, \dots$ $\alpha = 0.489, \beta = 0.00, \gamma = 0.698 \quad \hat{s}_1 = -38.573, \hat{s}_2 = -64.916, \hat{s}_3 = 35.733, \hat{s}_4 = 0.935, \hat{s}_5 = 40.627, \hat{s}_6 = 15.852$ $\hat{s}_7 = 18.918, \hat{s}_8 = 18.422, \hat{s}_9 = 2.471, \hat{s}_{10} = 10.145, \hat{s}_{11} = -18.172, \hat{s}_{12} = -36.526$
MAE = 13.832, RMSE = 14.740, MAPE = 1.714, MdAPE = 1.898, sMAPE = 1.705	

ตารางที่ 6 ตรวจสอบความคลาดเคลื่อนของตัวแบบ HWSE กพล. ทั้ง 3 ชุดข้อมูล

ชุดข้อมูล	ค่าเฉลี่ยเท่ากับ 0		การแจกแจงแบบปกติ		ความเป็นอิสระกัน		
	t-value	p-value	KS -value	p-value	Ljung - Box Q-statistics		
					Q-statistics	df	p-value
กพล.1	-0.063	0.950	0.083	0.446	30.939	21	0.075
กพล.2	-0.089	0.929	0.069	0.668	26.513	21	0.1876
กพล.3	0.167	0.867	0.077	0.544	18.133	21	0.641

ตารางที่ 7 ค่า MAE, RMSE, MAPE, MdAPE และ sMAPE จากการพยากรณ์โดยใช้ตัวแบบผสม

ชุดข้อมูล	MAE	RMSE	MAPE	MdAPE	sMAPE
กพล.1	12.393	15.935	1.958	1.923	1.984
กพล.2	10.058	11.974	1.848	1.770	1.868
กพล.3	9.185	11.942	1.138	0.970	1.132



รูปที่ 8 เปรียบเทียบเกณฑ์วัดประสิทธิภาพการพยากรณ์ระหว่างวิธีการการพยากรณ์อนุกรมเวลาเชิงเดี่ยวและวิธีการพยากรณ์แบบผสม

3.6 การเปรียบเทียบความคลาดเคลื่อนของค่าพยากรณ์ระหว่างตัวแบบ HWSE, SARIMA และตัวแบบผสมที่พัฒนาขึ้น

จากผลการทดลองใช้ตัวแบบการพยากรณ์ทั้ง 3 ตัวแบบ จากตารางที่ 8 พบว่า ตัวแบบอนุกรมเวลาเชิงเดี่ยว 2 ตัวแบบ ได้แก่ ตัวแบบ HWSE ตัวแบบ SARIMA และตัวแบบผสมต่างก็มีประสิทธิภาพที่ดีในแต่ละเกณฑ์การวัดประสิทธิภาพ การพยากรณ์แตกต่างกัน ชุดข้อมูลที่ 1 และ 2 ตัวแบบ HWSE มีประสิทธิภาพการพยากรณ์ดีกว่าตัวแบบ SARIMA ในทุกเกณฑ์การวัดประสิทธิภาพ ส่วนข้อมูล ชุดที่ 3 พบว่า ตัวแบบ SARIMA มีประสิทธิภาพดีกว่าตัวแบบ HWSE ในทุกเกณฑ์การวัดประสิทธิภาพ เมื่อพิจารณาตัวแบบผสมที่พัฒนาขึ้นพบว่า มีประสิทธิภาพดีกว่าตัวแบบอนุกรมเวลาเชิงเดี่ยวทั้ง 2 ตัวแบบ ในทุกเกณฑ์การวัด แสดงการเปรียบเทียบในรูปที่ 8 และรูปที่ 9 ตามลำดับ

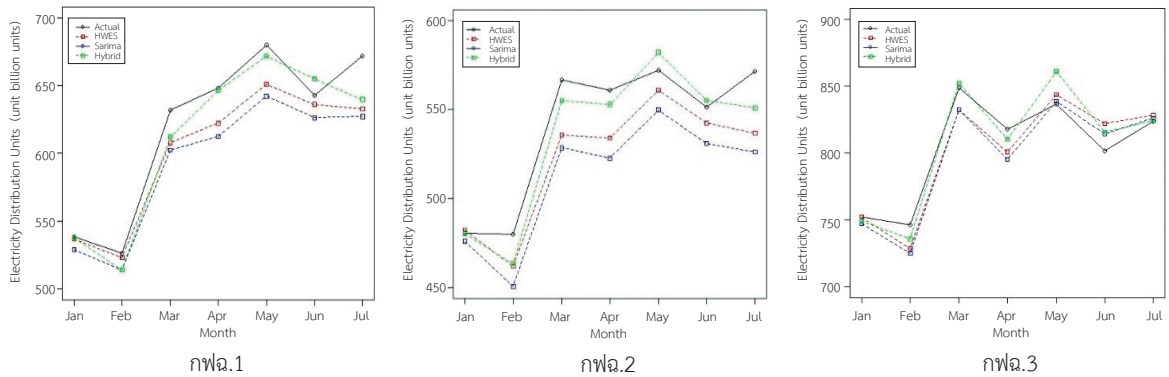
4. อภิปรายผลและสรุป

จากการนำเสนอวิธีการสร้างตัวแบบผสมเพื่อลดความคลาดเคลื่อนของค่าพยากรณ์ เมื่อเปรียบเทียบกับตัวแบบ HWSE และตัวแบบ SARIMA จากตารางที่ 8 รูปที่ 8 และรูปที่ 9 พบว่า ตัวแบบผสมจะให้ค่าความคลาดเคลื่อนต่ำกว่าตัวแบบอนุกรมเวลาเชิงเดี่ยวทั้ง 2 ตัวแบบ ประมาณ 2 เท่า โดยเฉพาะข้อมูลชุดที่ 1 กพฉ.1 และข้อมูลชุดที่ 2 กพฉ.2 ข้อค้นพบจากการวิจัยนี้บ่งชี้ว่า ตัวแบบผสมที่ผู้วิจัยพัฒนาขึ้น มีประสิทธิภาพและมีความแม่นยำในการพยากรณ์มากกว่าตัวแบบอนุกรมเวลาเชิงเดี่ยวทั้ง 2 ตัวแบบ ซึ่งสอดคล้องกับผลงานวิจัยที่อ้างอิง อย่างไรก็ตาม แม้ว่าตัวแบบอนุกรมเวลาเชิงเดี่ยวที่นำมาใช้ในการศึกษาเปรียบเทียบในครั้งนี้ ให้ค่าความคลาดเคลื่อนสูงกว่าตัวแบบที่ผู้วิจัยพัฒนาขึ้น หากจะนำตัวแบบอนุกรมเวลาเชิงเดี่ยวไปใช้ประโยชน์เพราะเป็นตัวแบบที่ใช้ง่าย และมีความสะดวกให้ค่าความคลาดเคลื่อนที่ยอมรับได้

ตารางที่ 8 เปรียบเทียบค่าเกณฑ์การวัดประสิทธิภาพของตัวแบบอนุกรมเวลาเชิงเดี่ยวและตัวแบบผสม

เกณฑ์การวัดประสิทธิภาพ	ข้อมูลชุดที่ 1 กพฉ.1			ข้อมูลชุดที่ 2 กพฉ.2			ข้อมูลชุดที่ 3 กพฉ.3		
	HWSE	SARIMA	HYBIRD	HWSE	SARIMA	HYBIRD	HWSE	SARIMA	HYBIRD
MAE	24.380	26.509	12.393	18.599	28.246	10.058	13.832	11.827	9.185
RMSE	26.756	29.431	15.935	23.208	31.001	11.974	14.740	14.339	11.942
MAPE	3.952	4.141	1.958	3.349	5.149	1.848	1.714	1.483	1.138
MdAPE	4.368	4.628	1.923	2.930	6.093	1.770	1.898	1.588	0.970
sMAPE	3.997	4.245	1.984	3.424	5.312	1.868	1.705	1.495	1.132

ธนกร สุทธิสนธิ์ และ สุรัตน์ หารวย, “การพัฒนาตัวแบบผสมเพื่อพยากรณ์ข้อมูลอนุกรมเวลาปริมาณหน่วยจำหน่ายไฟฟ้าบ้านอยู่อาศัยรายเดือนของประชาชนในเขตภาคตะวันออกเฉียงเหนือของประเทศไทย.”



รูปที่ 9 เปรียบเทียบค่าจริงกับค่าพยากรณ์ระหว่างวิธีการพยากรณ์อนุกรมเวลาเชิงเดี่ยวและวิธีการพยากรณ์แบบผสม

แต่ควรมีการกรองข้อมูลเพื่อลดสัญญาณแกว่งไกวของข้อมูล ก่อนนำไปพยากรณ์ด้วยตัวแบบอนุกรมเชิงเดี่ยว จะทำให้มีความแม่นยำของค่าพยากรณ์มากขึ้น และหากต้องการความแม่นยำของค่าพยากรณ์สูง แนะนำให้เลือกใช้ตัวแบบผสมที่ผู้วิจัยพัฒนาขึ้น ทั้งนี้ เป็นเพราะทำการกรองข้อมูลก่อนนำไปพยากรณ์เพื่อกำจัดสัญญาณแกว่งไกวของข้อมูลที่ไม่สามารถทำนายได้ และรวมค่าพยากรณ์อนุกรมเวลาเชิงเดี่ยว 2 ตัวแบบเข้าด้วยกันจึงทำให้ค่าพยากรณ์มีความแม่นยำสูง ตัวแบบการพยากรณ์ที่พัฒนาขึ้นในงานวิจัยนี้สามารถนำไปประยุกต์ใช้กับข้อมูลอนุกรมเวลาที่มีลักษณะการเคลื่อนไหวคล้ายกันเพื่อเป็นการเพิ่มประสิทธิภาพในการพยากรณ์ได้ ข้อเสนอแนะในการวิจัยครั้งต่อไปอาจเลือกใช้วิธีการกรองข้อมูลที่หลากหลายวิธี เช่น วิธีเวฟเลต (Wavelet Transform) วิธี SSA (Singular Spectrum Analysis) อาจทำให้ได้ค่าพยากรณ์ที่มีความแม่นยำมากขึ้นเป็นต้น

## 5. กิตติกรรมประกาศ

คณะผู้วิจัยขอขอบคุณการไฟฟ้าส่วนภูมิภาคที่อนุเคราะห์ข้อมูลปริมาณหน่วยจำหน่ายไฟฟ้าประเภทบ้านอยู่อาศัยภาคตะวันออกเฉียงเหนือในการศึกษาครั้งนี้ และขอขอบคุณสาขาวิชาสถิติประยุกต์ คณะวิทยาศาสตร์ มหาวิทยาลัยราชภัฏอุบลราชธานี ที่อนุเคราะห์สถานที่สำหรับการทำวิจัยในครั้งนี้

## เอกสารอ้างอิง

- [1] P. Du, J. Wang, W. Yang, and T. Niu, "A novel hybrid model for short-term wind power forecasting," *Applied Soft Computing*, no. 80, pp. 93–106, 2019.
- [2] J. Zhang, Y.M. Wei, D. Li, Z. Tan, and J. Zhou, "Short term electricity load forecasting using a hybrid model," *Energy*, no. 158, pp. 774–781, 2018.
- [3] W. Yang, J. Wang, T. Niu, and P. Du, "A hybrid forecasting system based on a dual decomposition strategy and multi-objective optimization for electricity price forecasting," *Applied Energy*, vol. 235, pp. 1205–1225, 2019.
- [4] H. Luo, D. Wang, C. Yue, Y. Liu, and H. Guo. "Research and application of a novel hybrid decomposition-ensemble learning paradigm with error correction for daily PM 10 forecasting," *Atmospheric Research*, vol. 201, pp. 34–45, 2018.
- [5] Y. Cheng, H. Zhang, Z. Liu, L. Chen, and P. Wang, "Hybrid algorithm for short-term forecasting



- of  $PM_{2.5}$  in China,” *Atmospheric Environment*, vol. 200, pp. 264–279, 2019.
- [6] M. Manmin, *Time Series and Forecasting*. Bangkok : Four Printing, 2006 (in Thai).
- [7] N. Papukdee, N. Senawong, and P. Busababodhin, “A comparative forecasting model of monthly rainfall in Northeast of Thailand,” *The Journal of KMUTNB*, vol. 29, no. 2, pp. 302–313, 2019 (in Thai).
- [8] Z. Yang and J.Wang, “A hybrid forecasting approach applied in wind speed forecasting based on a data processing strategy and an optimized artificial intelligence algorithm,” *Energy*, vol. 160, pp. 87–100, 2018.
- [9] M. Santhosh, C. Venkaiah, and D. M. Vinod Kumar, “Ensemble empirical mode decomposition based adaptive wavelet neural network method for wind speed prediction,” *Energy Conversion and Management*, vol. 168, pp. 482–493, 2018.
- [10] Y. Yaslan and B. Bican, “Empirical mode decomposition based denoising method with support vector regression for time series prediction: A case study for electricity load forecasting,” *Measurement*, vol. 103, pp. 52–61, 2017.
- [11] J. F. Rendon-Sanchez and L. M. de Menezes, “Structural combination of seasonal exponential smoothing forecasts applied to load forecasting,” *European Journal of Operational Research*, vol. 275, no. 3, pp. 916–924, 2019.
- [12] Power Economics Division, “Electricity Sales Report” Provincial Electricity Authority, Bangkok, Thailand. Accessed: March, 2019. [Online]. Available: <http://peaoc.pea.co.th/ped/>
- [13] N. Huang, Z. Shen, S. Long, M. Wu, H. Shih, Q. Zheng, N. Yen, C. Tung, and H. Liu, “The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis,” in *Proceedings of The Royal Society a Mathematical Physical and Engineering Sciences*, 1998, no. 454, pp. 903–995, 1998.
- [14] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time Series Analysis: Forecasting and Control*, 4th ed., New Jersey: John Wiley and Sons, Inc., 2008.
- [15] R Core Team. forecast: Forecasting Functions for Time Series and Linear Models. R Foundation for Statistical Computing, Vienna, Austria. Accessed: March 2019. [Online]. Available: <https://cran.r-project.org/web/packages/forecast/>