

การวิเคราะห์อนุกรมเวลาของข้อมูลหน่วยจำหน่ายไฟฟ้า เพื่อค้นหาตัวแบบการพยากรณ์ที่เหมาะสม ด้วยภาษาอาร์

Time Series Analysis of Electrical Distribution Units to Find a Suitable Forecasting Model with R Language

รณชัย ชื่นธวัช^{1*} ชนินทร เรืองอุดมสกุล¹ นิตยา เกิดประสพ¹ และกิตติศักดิ์ เกิดประสพ¹

Received: November, 2015; Accepted: May, 2016

บทคัดย่อ

การวิจัยนี้มีวัตถุประสงค์เพื่อวิเคราะห์อนุกรมเวลาของข้อมูลหน่วยจำหน่ายไฟฟ้า และค้นหาตัวแบบการพยากรณ์ที่เหมาะสม โดยใช้การเขียนโปรแกรมด้วยภาษาอาร์ เป็นเครื่องมือดำเนินการวิจัย ซึ่งอนุกรมเวลาที่นำมาวิเคราะห์ เป็นข้อมูลหน่วยจำหน่ายไฟฟ้าของการไฟฟ้านครหลวง (กฟน.) เก็บรวบรวมข้อมูลจากรายงานสถานการณ์การจำหน่ายไฟฟ้าของการไฟฟ้านครหลวง ตั้งแต่เดือนมกราคม ปี พ.ศ. 2553 ถึงเดือนพฤษภาคม ปี พ.ศ. 2558 ซึ่งข้อมูลที่สนใจมีขอบเขตเฉพาะจำนวนหน่วยจำหน่ายไฟฟ้าให้กับผู้บริโภคประเภทบ้านอยู่อาศัย โดยนำข้อมูลตั้งแต่เดือนมกราคม ปี พ.ศ. 2553 ถึงเดือนธันวาคม ปี พ.ศ. 2557 ไปสร้างตัวแบบการพยากรณ์ที่แตกต่างกันทั้งสิ้น 3 ตัวแบบ ด้วยภาษาอาร์ ได้แก่ ตัวแบบ glm, HoltWinters และ ARIMA จากนั้นเปรียบเทียบตัวแบบทั้ง 3 ด้วยการวัดค่าคลาดเคลื่อน ได้แก่ ค่า RMSE และค่า MAPE จากข้อมูลที่แต่ละตัวแบบพยากรณ์ข้อมูลหน่วยจำหน่ายไฟฟ้าของเดือนมกราคม ถึงเดือนพฤษภาคม ปี พ.ศ. 2558 เปรียบเทียบกับข้อมูลจริงที่ได้จากรายงานสถานการณ์การจำหน่ายไฟฟ้าของการไฟฟ้านครหลวง ในเดือนมกราคมถึงเดือนพฤษภาคม ปี พ.ศ. 2558 ซึ่งผลการทดลอง สรุปได้ว่าตัวแบบ ARIMA เป็นตัวแบบที่ให้ค่า RMSE และค่า MAPE ต่ำที่สุด ดังนั้น ตัวแบบ ARIMA เป็นตัวแบบที่เหมาะสมที่สุดสำหรับนำไปสร้างเป็นตัวแบบการพยากรณ์ในรูปของสมการคณิตศาสตร์ เพื่อพยากรณ์ข้อมูลหน่วยจำหน่ายไฟฟ้าของการไฟฟ้านครหลวงในส่วนของผู้บริโภคประเภทบ้านอยู่อาศัย

คำสำคัญ : ข้อมูลอนุกรมเวลา; หน่วยจำหน่ายไฟฟ้า; การพยากรณ์; ภาษาอาร์

¹ สำนักวิชาวิศวกรรมศาสตร์ มหาวิทยาลัยเทคโนโลยีสุรนารี นครราชสีมา

* Corresponding Author E - mail Address: c_ronnachai@hotmail.com

Abstract

This research aimed to analyze time series of electrical distribution unit data and to find a suitable forecasting model by using R language programming as a tool to carry out this research. The analyzed time series were the electrical distribution unit data of Metropolitan Electricity Authority (MEA) which had been collected from report of electrical distribution since January 2010 to May 2015. Residential consumers were the interested specific data. The electrical distribution units as the training data set were the data from January 2010 to December 2014. This data set were modeled by 3 different forecasting models with R language programming, i.e. glm model, HoltWinters model and ARIMA model. Then the comparison among the 3 models was done by comparing the model forecasting against the actual data of MEA during January - May, 2015. From the comparative results, it was concluded that the ARIMA model is the most suitable one for creating a forecasting model in term of mathematic equation to predict the electrical distribution units of MEA for the section of residential consumers.

Keywords: Time Series Data; Electrical Distribution Units; Forecasting; R Language

บทนำ

การวิเคราะห์อนุกรมเวลาเป็นวิธีการพยากรณ์เชิงปริมาณ หมายถึง การใช้ข้อมูลเชิงปริมาณ ที่มีการเก็บรวบรวมอย่างเป็นระบบ และต่อเนื่องในอดีต มาวิเคราะห์ด้วยวิธีการเชิงสถิติ เพื่อพยากรณ์ข้อมูลที่เกิดขึ้นในอนาคต ส่วนข้อมูลอนุกรมเวลา หมายถึง ข้อมูลที่เก็บบันทึกตามลำดับเวลา และการบันทึกข้อมูลอาจกระทำเป็นรายวัน รายสัปดาห์ รายเดือน รายไตรมาส หรือรายปี ขึ้นอยู่กับความเหมาะสมของการนำข้อมูลไปใช้ โดยข้อมูลอนุกรมเวลาประกอบด้วย 4 ส่วนประกอบ คือ แนวโน้ม ฤดูกาล รอบเวลา และความผิดปกติ ซึ่งการวิเคราะห์ข้อมูลอนุกรมเวลาโดยใช้เทคนิคการพยากรณ์จะทำการกำหนดตัวแบบการพยากรณ์จากข้อมูลในอดีตด้วย สมมุติฐานว่าเหตุการณ์ที่เกิดขึ้นในอนาคต มีรูปแบบเหมือนกับเหตุการณ์ที่เกิดขึ้นในอดีต และงานวิจัยนี้ข้อมูลที่น่าสนใจมาพยากรณ์เป็นข้อมูลการจำหน่ายไฟฟ้าของการไฟฟ้านครหลวงที่จำหน่ายให้กับผู้บริโภคที่อยู่ในเขตพื้นที่กรุงเทพมหานคร สมุทรปราการ และนนทบุรี ประเภทบ้านอยู่อาศัย โดยอาศัยรายงานสถานการณ์การจำหน่ายไฟฟ้าประจำเดือนตั้งแต่เดือนมกราคม ปี พ.ศ. 2553 ถึงเดือนพฤษภาคม ปี พ.ศ. 2558 [1] และข้อมูลที่เกี่ยวข้อง คือ หน่วยจำหน่ายไฟฟ้า โดย 1 หน่วยจำหน่าย หมายถึง การใช้ไฟฟ้า 1,000 วัตต์ต่อชั่วโมง ซึ่งข้อมูลหน่วยจำหน่ายไฟฟ้า เป็นข้อมูลที่สะท้อนถึงความต้องการใช้ไฟฟ้าของผู้บริโภค มีลักษณะเป็นอนุกรมเวลาที่บันทึกข้อมูลเป็นรายเดือนตามลำดับเวลา และการพยากรณ์ค่าความต้องการไฟฟ้าของผู้บริโภคช่วยวางแผนด้านทรัพยากร จัดสรรงบประมาณ และลดค่าใช้จ่ายในการดำเนินการลงได้ สอดคล้องกับงานวิจัยของ Bunn, D.W. and

Farmer, E.D. [2] ได้ศึกษาการพยากรณ์ความต้องการไฟฟ้า พบว่า ค่าใช้จ่ายการดำเนินการจะเพิ่มขึ้น 10 ล้านปอนด์ทุกปี เมื่อมีการพยากรณ์ผิดพลาดไปทุก 1% กล่าวได้ว่า การพยากรณ์ความต้องการของหน่วยจำหน่ายไฟฟ้า มีความสำคัญอย่างยิ่งต่อการวางแผนการผลิตกระแสไฟฟ้า และทำให้การดำเนินการผลิตกระแสไฟฟ้า มีประสิทธิภาพมากยิ่งขึ้น

สำหรับวิธีการวิเคราะห์อนุกรมเวลามีหลากหลายวิธี แต่ละวิธีจะดำเนินการสร้างตัวแบบการพยากรณ์ที่ให้ค่าความแม่นยำของการพยากรณ์แตกต่างกัน โดยสามารถวัดค่าความแม่นยำด้วยการวัดค่าความคลาดเคลื่อนทางสถิติ เช่น ค่า SSE (Sum Square Error) ค่า RMSE (Root Mean Square Error) ค่า MAE (Mean Absolute Error) หรือค่า MAPE (Mean Absolute Percentage Error) ซึ่งค่าสถิติดังกล่าว เป็นการวัดค่าความแม่นยำด้วยการเปรียบเทียบค่าพยากรณ์กับค่าสังเกตจริงในช่วงเวลาเดียวกัน ว่ามีค่าแตกต่างกันมากน้อยเพียงใด ดังนั้น ค่าสถิติดังกล่าว สามารถนำมาใช้เปรียบเทียบตัวแบบการพยากรณ์ที่ถูกสร้างด้วยวิธีการพยากรณ์ที่แตกต่างกันได้ และวิธีการวิเคราะห์อนุกรมเวลาเพื่อสร้างตัวแบบการพยากรณ์ที่สำคัญ มีดังนี้

1. วิธีการวิเคราะห์การถดถอย (Regression Analysis) เป็นการศึกษาความสัมพันธ์ระหว่างปัจจัยหรือตัวแปรตั้งแต่ 2 ตัวแปรขึ้นไป โดยเป็นความสัมพันธ์ของตัวแปรอิสระ ที่มีค่าของตัวแปรเกิดขึ้นอย่างอิสระ และตัวแปรอิสระมีตั้งแต่ 1 ตัวแปรขึ้นไป กับตัวแปรตาม 1 ตัวแปร ซึ่งค่าของตัวแปรตามขึ้นอยู่กับรูปแบบความสัมพันธ์กับตัวแปรอิสระ และวัตถุประสงค์ของการวิเคราะห์การถดถอย คือ เพื่อให้ทราบถึงลักษณะของความสัมพันธ์ ขนาดของความสัมพันธ์ และทิศทางของความสัมพันธ์ระหว่างตัวแปรอิสระกับตัวแปรตาม

2. วิธีการปรับให้เรียบแบบเอ็กซ์โปเนนเชียล (Exponential Smoothing) เป็นวิธีการพยากรณ์ที่มีประสิทธิภาพ เมื่อองค์ประกอบของอนุกรมเวลา ได้แก่ แนวโน้ม และฤดูกาล ที่มีการเปลี่ยนแปลงตลอดช่วงเวลาที่น่าสนใจ สำหรับวิธีการนี้เป็นวิธีการให้น้ำหนักกับค่าสังเกตของอนุกรมเวลาที่นำมาใช้ในการพยากรณ์ไม่เท่ากัน โดยให้น้ำหนักค่าสังเกตที่เกิดขึ้นล่าสุดมากที่สุด ส่วนน้ำหนักของค่าสังเกตก่อนหน้ามีค่าลดลงตามลำดับเวลา หมายถึง ถ้าค่าสังเกตล่าสุดมาก ยิ่งมีน้ำหนักน้อย และน้ำหนักของค่าสังเกต จะลดลงแบบเอ็กซ์โปเนนเชียลตามลำดับเวลา ซึ่งใช้ค่าคงที่ 1 หรือมากกว่า 1 ตัว กำหนดน้ำหนักให้กับค่าสังเกตแต่ละค่า และเรียกค่าคงที่ว่า ค่าคงที่ปรับให้เรียบ (Smoothing Constant)

3. วิธีของบ็อกซ์และเจนกินส์ (Box and Jenkins Method) เป็นวิธีการวิเคราะห์อนุกรมเวลาที่อยู่ภายใต้กระบวนการเชิงสุ่ม (Stochastic Process) โดยถือว่าข้อมูลที่เกิดขึ้นตามเวลาที่เปลี่ยนแปลงไป มีลักษณะการเกิดค่าของข้อมูลที่เป็นไปตามกฎความน่าจะเป็น ซึ่งวิธีของบ็อกซ์และเจนกินส์ จะพยากรณ์ข้อมูลจากรูปแบบหลัก 2 รูปแบบ ได้แก่ รูปแบบ Autoregressive คือ รูปแบบที่กำหนดว่าค่าพยากรณ์ที่เวลาใด ๆ ขึ้นอยู่กับค่าสังเกตก่อนหน้า และรูปแบบ Moving Average คือ รูปแบบที่กำหนดว่าค่าพยากรณ์ที่เวลาใด ๆ ขึ้นอยู่กับค่าคลาดเคลื่อนก่อนหน้า

การศึกษาวิธีการสร้างตัวแบบการพยากรณ์ตามวิธีการวิเคราะห์อนุกรมเวลาแต่ละวิธี จะใช้โปรแกรมคอมพิวเตอร์เป็นเครื่องมือช่วยในการสร้างตัวแบบการพยากรณ์ และพยากรณ์ข้อมูลล่วงหน้าจากตัวแบบที่สร้างขึ้น โดยงานวิจัยนี้นำเสนอการเขียนโปรแกรมเพื่อสร้างตัวแบบการพยากรณ์และพยากรณ์ข้อมูล

ล่วงหน้าด้วยภาษาอาร์ (R Language) ซึ่งเป็นภาษาคอมพิวเตอร์ที่สร้างขึ้นโดย Ross Ihaka และ Robert Gentleman แห่งมหาวิทยาลัยออกแลนด์ (The University of Auckland) โดยภาษาอาร์เป็นภาษาคอมพิวเตอร์ที่ใช้สำหรับการวิเคราะห์เชิงสถิติ ซึ่งมีฟังก์ชันทางสถิติหลากหลาย สามารถนำมาเขียนโปรแกรมและพัฒนาวิธีการทางสถิติใหม่ ๆ ที่แสดงผลในรูปของกราฟได้ ทำให้ผู้ใช้นำมาวิเคราะห์และจัดการข้อมูลได้อย่างสะดวก รวดเร็วมากขึ้น ซึ่งต่างจากโปรแกรมสำเร็จรูปทางสถิติที่ใช้โดยทั่วไป และการใช้ภาษาอาร์สำหรับการวิเคราะห์เชิงสถิติ สามารถจำแนกได้เป็น 2 ลักษณะ ได้แก่ แบบโต้ตอบ (Interactive Mode) และแบบชุดคำสั่งในลักษณะฟังก์ชันและวัตถุ (Object and Functional Programming) ดังนั้น ภาษาอาร์ ซึ่งเป็นภาษาคอมพิวเตอร์ในรูปแบบของโอเพนซอร์ส (Open Source) จึงเป็นภาษาที่มีผู้พัฒนาฟังก์ชันที่หลากหลายเพื่อใช้สำหรับการวิเคราะห์ทางสถิติ และสามารถใช้ภาษาอาร์สนับสนุนการเขียนโปรแกรมแบบ Procedural Programming หรือการเขียนโปรแกรมแบบ Object Oriented Programming โดยมีโครงสร้างข้อมูลพื้นฐาน ได้แก่ เวกเตอร์ (Vectors) เมตริกซ์ (Metrics) อาร์เรย์ (Arrays) และคาตาเฟรม (Data Frames)

จากการทบทวนวรรณกรรมและงานวิจัยที่เกี่ยวข้องกับการวิเคราะห์อนุกรมเวลาที่เป็นข้อมูลแสดงปริมาณการใช้ไฟฟ้า เพื่อสร้างตัวแบบการพยากรณ์ข้อมูล สามารถอธิบายได้ดังนี้

1. Maçaira, P.M. et al. [3] ศึกษาการกำหนดตัวแบบการพยากรณ์ และพยากรณ์ปริมาณการใช้ไฟฟ้ารายปีของผู้บริโภคประเภทบ้านอยู่อาศัย ในประเทศบราซิล จนถึงปี ค.ศ. 2050 ด้วยเทคนิคการพยากรณ์แบบ Pegels Exponential Smoothing โดยปรับค่าพารามิเตอร์ เพื่อพยากรณ์ข้อมูลร่วมกับค่าประมาณการของบริษัท ERC (The Energy Research Company) ที่กำหนดไว้ในรายงาน PDE (The Ten Year Energy Planning) และรายงาน PNE (The Nation Energy Planning) ผลจากการศึกษาทำให้ทราบตัวแบบการพยากรณ์ที่ให้ค่าพยากรณ์ใกล้เคียงกับค่าประมาณการปริมาณการใช้ไฟฟ้าในรายงาน PDE และ PNE โดยตัวแบบการพยากรณ์สามารถพยากรณ์ปริมาณการใช้ไฟฟ้ารายปีได้ถึงปี ค.ศ. 2050

2. Pereira, C.M. et al. [4] ศึกษาการเปรียบเทียบตัวแบบการพยากรณ์ข้อมูลอนุกรมเวลา 2 ตัวแบบ โดยใช้ข้อมูลการจ่ายกระแสไฟฟ้าของสถานีจ่ายไฟฟ้า Coelba ในประเทศบราซิล ซึ่งมีลักษณะเป็นข้อมูลอนุกรมเวลา มาสร้างตัวแบบการพยากรณ์ 2 ตัวแบบ คือ ตัวแบบ SARIMAX และตัวแบบ FIC (Fuzzy Inference System) โดยรวมตัวแปรภายนอก 3 ตัวแปร ได้แก่ จำนวนลูกค้า อุณหภูมิ และปริมาณน้ำฝน เข้าไปในตัวแบบทั้ง 2 จากนั้นเปรียบเทียบค่าพยากรณ์ที่ได้จากตัวแบบการพยากรณ์ด้วยค่า MAPE (Mean Average Percentage Error) และค่า SD (Standard Deviation) ผลจากการศึกษาสรุปได้ว่า ตัวแบบ FIC ให้ค่าคลาดเคลื่อนต่ำกว่าตัวแบบ SARIMAX

3. Keka, I. and Hamiti, M. [5] ศึกษาการหาตัวแบบทางคณิตศาสตร์ที่แสดงความสัมพันธ์ระหว่างพลังงานไฟฟ้าและเวลา ด้วยหลักการวิเคราะห์ความถดถอยแบบเชิงเส้น (Linear Regression) ซึ่งข้อมูลที่ใช้หาตัวแบบทางคณิตศาสตร์ได้ใช้ข้อมูลการจ่ายกระแสไฟฟ้าของสถานีจ่ายไฟฟ้าย่อยที่มีการวัดปริมาณการจ่ายกระแสไฟฟ้าทุก 15 นาที แล้วนำมาสร้างเป็นข้อมูลอนุกรมเวลาด้วยภาษาอาร์ โดยแบ่งคาบเวลาเป็นรายวัน รายสัปดาห์ และรายเดือน จากนั้นคำนวณค่าพารามิเตอร์ทางสถิติ และนำเสนอตัวแบบเชิงเส้นทางคณิตศาสตร์ที่เป็นสมการแสดงความสัมพันธ์ระหว่างพลังงานไฟฟ้าและเวลา

4. Chujai, P. et al. [6] ได้ศึกษาการวิเคราะห์ข้อมูลอนุกรมเวลาของปริมาณการใช้กระแสไฟฟ้าภายในบ้าน ซึ่งชุดข้อมูลที่ใช้เป็นชุดข้อมูลมาตรฐาน UCI โดยนำมาสร้างเป็นข้อมูลอนุกรมเวลาที่แบ่งคาบเวลาเป็น 4 ประเภท คือ รายวัน รายสัปดาห์ รายเดือน และรายไตรมาส แล้วนำข้อมูลมาสร้างตัวแบบการพยากรณ์ 2 ตัวแบบ คือ ตัวแบบ ARIMA และตัวแบบ ARMA ด้วยภาษาอาร์และโปรแกรม Rstudio จากนั้นเปรียบเทียบความถูกต้องของค่าพยากรณ์จากตัวแบบทั้งสอง โดยวัดค่าคลาดเคลื่อนด้วยค่า AIC (Akaike Information Criterion) และค่า RMSE (Root Mean Square Error) ซึ่งสรุปได้ว่าตัวแบบ ARIMA เป็นตัวแบบที่เหมาะสมสำหรับการพยากรณ์คาบเวลารายเดือนและรายไตรมาส ส่วนตัวแบบ ARMA เป็นตัวแบบที่เหมาะสมสำหรับการพยากรณ์คาบเวลารายวันและรายสัปดาห์

5. Wang, Y. et al. [7] ได้ศึกษาการปรับปรุงความแม่นยำของค่าพยากรณ์ความต้องการไฟฟ้าในเขตไฟฟ้าภาคตะวันออกเฉียงเหนือของประเทศไทย ที่ได้จากตัวแบบ Seasonal ARIMA โดยทำการปรับปรุงความแม่นยำของค่าพยากรณ์ด้วยการใช้วิธีการปรับปรุงค่าคลาดเคลื่อน (Residual) ที่เกิดจากการพยากรณ์ด้วยตัวแบบ Seasonal ARIMA ซึ่งใช้วิธีการที่แตกต่างกัน 3 วิธี ได้แก่ 1) การปรับปรุงค่าคลาดเคลื่อนด้วยวิธี PSO Optimal Fourier 2) วิธี Seasonal ARIMA และ 3) วิธีแบบผสมผสานระหว่างวิธี PSO Optimal Fourier และวิธี Seasonal ARIMA ผลการศึกษาสรุปได้ว่า การใช้วิธีการปรับปรุงค่าคลาดเคลื่อนทั้ง 3 วิธี ร่วมกับการพยากรณ์ด้วยตัวแบบ Seasonal ARIMA ให้ความถูกต้องของค่าพยากรณ์สูงกว่าการพยากรณ์ด้วยตัวแบบ Seasonal ARIMA เพียงอย่างเดียว นอกจากนี้วิธีการปรับปรุงค่าคลาดเคลื่อนที่ให้ผลลัพธ์ดีที่สุด คือ วิธีการปรับปรุงค่าคลาดเคลื่อนแบบผสมผสาน

งานวิจัยนี้ผู้วิจัยต้องการนำเสนอวิธีการวิเคราะห์อนุกรมเวลา ด้วยการสร้างตัวแบบการพยากรณ์จากวิธีการที่แตกต่างกัน 3 วิธี ได้แก่ วิธีการวิเคราะห์การถดถอย วิธีการปรับให้เรียบแบบเอ็กซ์โปเนนเชียลของ Holt และ Winters และวิธีของบ็อกซ์และเจนกินส์ จากนั้นประเมินผลความแม่นยำของแต่ละตัวแบบด้วยการวัดค่าคลาดเคลื่อน RMSE และ MAPE ซึ่งใช้การเขียนโปรแกรมด้วยภาษาอาร์ เป็นเครื่องมือในการดำเนินการวิจัย ดังนั้น วัตถุประสงค์ของงานวิจัยสามารถแบ่งเป็นประเด็นหลัก ได้ดังนี้

1. เพื่อศึกษาการวิเคราะห์อนุกรมเวลา ด้วยการสร้างตัวแบบการพยากรณ์ด้วยภาษาอาร์และโปรแกรม Rstudio จากข้อมูลหน่วยจำหน่ายไฟฟ้า ตั้งแต่เดือนมกราคม ปี พ.ศ. 2553 ถึงเดือนธันวาคม ปี พ.ศ. 2557 ทั้งหมด 3 ตัวแบบ ได้แก่ 1) ตัวแบบ Generalized Linear Models (glm) 2) ตัวแบบ Holt-Winters (HoltWinters) และ 3) ตัวแบบ Autoregressive Integrated Moving Average (ARIMA)

2. เพื่อเปรียบเทียบตัวแบบทั้ง 3 ด้วยการทดสอบความถูกต้องของค่าพยากรณ์ โดยใช้ภาษาอาร์คำนวณค่า RMSE และ MAPE ของแต่ละตัวแบบ ซึ่งตัวแบบที่เหมาะสมสำหรับการพยากรณ์ข้อมูลหน่วยจำหน่ายไฟฟ้า คือตัวแบบที่ให้ค่า RMSE และ MAPE ค่าที่ต่ำที่สุด

3. เพื่อสร้างตัวแบบการพยากรณ์ที่เหมาะสมที่สุด ในรูปของสมการคณิตศาสตร์ ที่สามารถนำไปพัฒนาเป็นโปรแกรมคอมพิวเตอร์ สำหรับพยากรณ์ข้อมูลหน่วยจำหน่ายไฟฟ้าประเภทบ้านอยู่อาศัยของการไฟฟ้านครหลวง

วิธีดำเนินการวิจัย

สำหรับงานวิจัยนี้ผู้วิจัยต้องการศึกษาข้อมูลการจำหน่ายไฟฟ้าของการไฟฟ้านครหลวง ที่จำหน่ายให้กับผู้บริโภคประเภทบ้านอยู่อาศัย โดยแบ่งคาบเวลาเป็นรายเดือน ซึ่งแบ่งข้อมูลเป็นสองกลุ่ม กลุ่มที่ 1 เป็นกลุ่มที่ใช้สำหรับการสร้างตัวแบบการพยากรณ์ โดยใช้ข้อมูลหน่วยจำหน่ายไฟฟ้า ตั้งแต่เดือนมกราคม ปี พ.ศ. 2553 ถึงเดือนธันวาคม ปี พ.ศ. 2557 เป็นข้อมูลฝึกสอน ส่วนกลุ่มที่ 2 เป็นกลุ่มที่ใช้สำหรับทดสอบความถูกต้องของค่าพยากรณ์สำหรับแต่ละตัวแบบการพยากรณ์ที่ได้สร้างขึ้น โดยใช้ข้อมูลหน่วยจำหน่ายไฟฟ้า ตั้งแต่เดือนมกราคมถึงเดือนพฤษภาคม ปี พ.ศ. 2558 เป็นข้อมูลทดสอบ ซึ่งวิธีดำเนินการวิจัยจะดำเนินการสร้างตัวแบบการพยากรณ์ จากวิธีการพยากรณ์ที่แตกต่างกัน 3 วิธี ได้แก่ วิธีการวิเคราะห์ความถดถอย วิธีการปรับให้เรียบแบบเอ็กซ์โปเนนเชียล และวิธีของบอกรีและเจนกินส์ ได้เป็นตัวแบบการพยากรณ์ 3 ตัวแบบ จากนั้น นำตัวแบบทั้ง 3 มาทำนายค่าพยากรณ์ล่วงหน้า 5 คาบเวลา แล้วนำค่าพยากรณ์ของแต่ละตัวแบบไปเปรียบเทียบกับค่าสังเกตจริง ที่แบ่งไว้เป็นชุดข้อมูลทดสอบ เพื่อวัดค่าคลาดเคลื่อนทางสถิติ ได้แก่ ค่า RMSE และค่า MAPE และใช้ค่าคลาดเคลื่อนดังกล่าว เลือกตัวแบบการพยากรณ์ที่เหมาะสมที่สุด แล้วนำไปสร้างสมการพยากรณ์ของตัวแบบที่เหมาะสมที่สุด ซึ่งตัวแบบการพยากรณ์ที่สร้างจากวิธีการพยากรณ์ที่แตกต่างกันทั้ง 3 วิธี มีดังนี้

1. ตัวแบบ glm เป็นตัวแบบที่สร้างจากวิธีการวิเคราะห์ความถดถอย โดยมีรูปแบบการถดถอยแบบเชิงเส้น ดังนี้ [8]

$$Y = \beta_0 + \beta_1 X + \varepsilon \quad (1)$$

โดยที่

X	คือ	คาบเวลามีหน่วยเป็นเดือน ซึ่งเป็นตัวแปรอิสระ
Y	คือ	ค่าสังเกตที่สนใจ ในที่นี้คือหน่วยจำหน่ายไฟฟ้ารายเดือน
ε	คือ	ค่าคลาดเคลื่อน

2. ตัวแบบ HoltWinters เป็นตัวแบบที่สร้างจากวิธีการปรับให้เรียบแบบทริเปิลเอ็กซ์โปเนนเชียลสามพหุติ (Triple Exponential Smoothing) ซึ่งเป็นวิธีการปรับให้เรียบที่นำมาใช้สร้างตัวแบบการพยากรณ์กับข้อมูลที่มีส่วนประกอบของแนวโน้ม และการเปลี่ยนแปลงตามฤดูกาล โดยมีรูปแบบ ดังนี้ [9]

$$\begin{aligned} \text{Level} : L_t &= \alpha(Y_t - S_{t-s}) + (1 - \alpha)(L_{t-1} + b_{t-1}) \\ \text{Trend} : b_t &= \beta(L_t - L_{t-1}) + (1 - \beta)b_{t-1} \\ \text{Seasonal} : S_t &= \gamma(Y_{t-1} - L_t) + (1 - \gamma)S_{t-s} \end{aligned} \quad (2)$$

และสมการพยากรณ์ ที่ได้จากส่วนประกอบทั้ง 3 ส่วน แสดงได้ดังนี้

$$\text{Forecast: } F_{t+m} = L_t + b_{tm} + S_{t-s+m} \quad (3)$$

โดยที่

α	คือ	ค่าคงที่ปรับให้เรียบของส่วนประกอบระดับ มีค่าระหว่าง 0 กับ 1
β	คือ	ค่าคงที่ปรับให้เรียบของส่วนประกอบแนวโน้ม มีค่าระหว่าง 0 กับ 1
γ	คือ	ค่าคงที่ปรับให้เรียบของส่วนประกอบฤดูกาล มีค่าระหว่าง 0 กับ 1
L_t	คือ	ค่าประมาณระดับของอนุกรมเวลา ณ เวลา t
Y_t	คือ	ค่าสังเกต ณ เวลา t
b_t	คือ	ค่าประมาณความชันของอนุกรมเวลา ณ เวลา t
S_t	คือ	ส่วนประกอบฤดูกาล
s	คือ	ความยาวของแต่ละฤดูกาล เช่น ฤดูกาลละ 12 เดือน ($s = 12$)
m	คือ	จำนวนคาบเวลาล่วงหน้าที่ต้องการพยากรณ์
F_{t+m}	คือ	ค่าพยากรณ์ล่วงหน้า m คาบเวลา

3. ตัวแบบ ARIMA เป็นตัวแบบที่สร้างจากวิธีของบอซซ์และเจนกินส์ โดยมีรูปแบบทั่วไปที่เขียนในรูปของตัวดำเนินการย้อนกลับ (Backward Shift Operator: B) ในรูปสมการดังนี้ [7]

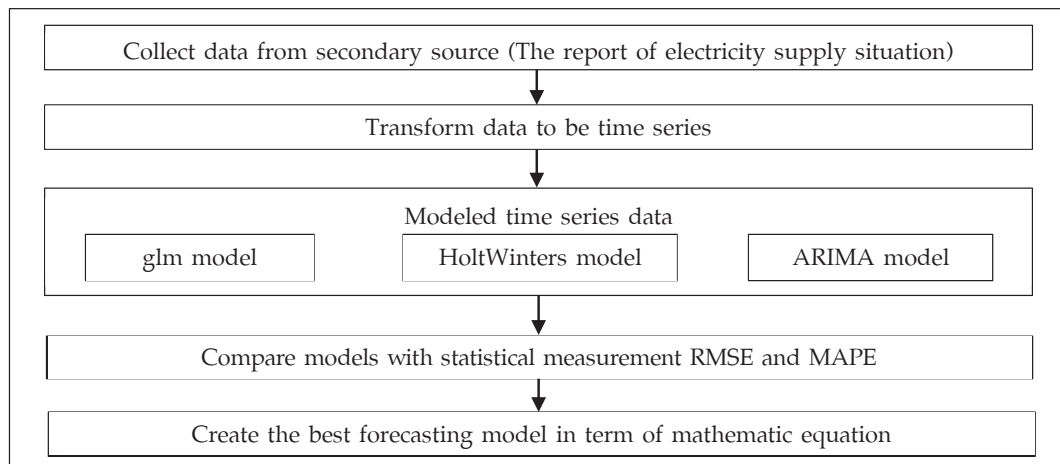
$$\theta_p(B)\Theta_p(B^s)(1-B)^d(1-B^s)^D Y_t = w_q(B)W_Q(B^s)a_t \quad (4)$$

เมื่อ $BY_t = Y_{t-1}$ และ $B^k Y_t = Y_{t-k}$

โดยที่

$$\begin{aligned} w_q(B) &= 1 - w_1 B - w_2 B^2 - \dots - w_q B^p \\ \theta_p(B) &= 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_p B^p \\ \Theta_p(B^s) &= 1 - \Theta_1 B^s - \Theta_2 B^{2s} - \dots - \Theta_p B^{ps} \\ W_Q(B^s) &= 1 - W_1 B^s - W_2 B^{2s} - \dots - W_Q B^{Qs} \end{aligned} \quad (5)$$

เมื่อได้ 3 ตัวแบบ จึงนำตัวแบบมาเปรียบเทียบ ด้วยการวัดค่าสถิติ RMSE และ MAPE ซึ่งตัวแบบที่เหมาะสมสำหรับการพยากรณ์ข้อมูลอนุกรมเวลาของข้อมูลหน่วยจำหน่ายไฟฟ้า คือ ตัวแบบที่ให้ค่า RMSE และค่า MAPE ต่ำที่สุด โดยสรุปวิธีการวิเคราะห์ข้อมูลอนุกรมเวลา เพื่อหาตัวแบบที่เหมาะสมสำหรับการพยากรณ์ข้อมูล ดังรูปที่ 1



รูปที่ 1 กรอบแนวคิดแสดงขั้นตอนการเลือกตัวแบบที่เหมาะสมที่สุดสำหรับการพยากรณ์ข้อมูลอนุกรมเวลา

จากรูปที่ 1 อธิบายแต่ละขั้นตอนได้ดังนี้

1. Collect Data from Secondary Source คือ ขั้นตอนรวบรวมข้อมูล จากข้อมูลทุติยภูมิ ที่ได้จากรายงานสถานการณ์การจำหน่ายไฟฟ้าประจำเดือน ของการไฟฟ้านครหลวง ตั้งแต่เดือนมกราคม ปี พ.ศ. 2553 ถึงเดือนพฤษภาคม ปี พ.ศ. 2558 กับผู้บริโภคนประเภทบ้านอยู่อาศัย ซึ่งตัวอย่างของข้อมูล ในรายงาน ที่นำมาสร้างเป็นข้อมูลอนุกรมเวลา แสดงได้ดังรูปที่ 2

รายการ	ก เดือน พ.ย. 2553	ข เดือน ธ.ค. 2552	ค เดือน ธ.ค. 2553
1. ข้อมูลการใช้ไฟฟ้า			
1.1 ข้อมูลหน่วยซื้อ(หน่วย : ล้านหน่วย)	3,679.03	3,474.84	3,586.96
1.1.1 ข้อมูลหน่วยซื้อจากกฟผ.(หน่วย : ล้านหน่วย)	3,678.68	3,473.74	3,586.95
1.1.1.1 แยกตามแรงดัน			
1.1.1.1.1 69 เควี	1,677.30	1,626.20	1,715.35
1.1.1.1.2 115 เควี	1,546.66	1,408.64	1,501.67
1.1.1.1.3 230 เควี	454.72	438.9	369.93
1.1.1.2 จำนวนสถานีต้นทาง แยกตามแรงดัน			
1.1.1.2.1 69 เควี	11	11	11
1.1.1.2.2 115 เควี	12	12	12
1.1.1.2.3 230 เควี	3	3	3
1.1.2 ข้อมูลหน่วยซื้อจาก VSPP(หน่วย : ล้านหน่วย)	0.35	1.1	0.01
1.2 หน่วยจำหน่าย(หน่วย : ล้านหน่วย)	3,529.06	3,318.09	3,477.63
1.2.1 แยกตามแรงดัน			
1.2.1.1 220 โวลท์	1,660.78	1,650.83	1,637.19
1.2.1.2 12 - 24 เควี	1,589.66	1,407.74	1,567.07
1.2.1.3 69 - 115 เควี	304.5	285.25	300.17
1.2.2 ตามประเภทผู้ใช้ไฟฟ้า			
1.2.2.1 บ้านอยู่อาศัย	787.32	722.16	776.67

รูปที่ 2 ตัวอย่างข้อมูลหน่วยจำหน่ายไฟฟ้าประจำเดือนธันวาคม ปี พ.ศ. 2553

2. Transform Data to be Time Series คือ ขั้นตอนที่น่าข้อมูลทุกข้อมูลที่รวบรวมได้ มาสร้างเป็นข้อมูลอนุกรมเวลา เรียงตามลำดับเวลา โดยมีคาบเวลาเป็นรายเดือน ตั้งแต่เดือนมกราคม ปี พ.ศ. 2553 ถึงเดือนพฤษภาคม ปี พ.ศ. 2558 แสดงได้ดังรูปที่ 3 และแบ่งข้อมูลเป็น 2 ชุด ได้แก่ 1) ชุดข้อมูลฝึกสอน (Training Set) ซึ่งใช้ข้อมูลตั้งแต่เดือนมกราคม ปี พ.ศ. 2553 ถึงเดือนธันวาคม ปี พ.ศ. 2557 สำหรับสร้างตัวแบบพยากรณ์ และ 2) ชุดข้อมูลทดสอบ (Test Set) ซึ่งใช้ข้อมูลตั้งแต่เดือนมกราคมถึงเดือนพฤษภาคม ปี พ.ศ. 2558 สำหรับทดสอบตัวแบบพยากรณ์ เพื่อวัดค่าคลาดเคลื่อน RMSE และ MAPE

	A	B	C	D	E
1	Years	Months	Electric Units	No_Consumers	Mean_Units
2	2553	1	776.32	2420767	320.69
3	2553	2	836.16	2423046	345.09
4	2553	3	925.55	2433022	380.41
5	2553	4	1026.5	2439689	420.75
6	2553	5	1082.18	2448223	442.03
7	2553	6	1003.41	2456309	408.5
8	2553	7	930.91	2463398	377.9
9	2553	8	876.83	2473742	354.45

รูปที่ 3 ข้อมูลอนุกรมเวลาที่สร้างจากข้อมูลทุกข้อมูล

3. Modeled Time Series Data คือ ขั้นตอนการสร้างตัวแบบที่แตกต่างกัน 3 ตัวแบบ จากข้อมูลอนุกรมเวลาที่แบ่งไว้เป็นชุดข้อมูลฝึกสอน โดยการสร้างตัวแบบทั้ง 3 ใช้ชุดข้อมูลฝึกสอนชุดเดียวกัน แต่ละตัวแบบมีคำสั่งสร้างตัวแบบด้วยภาษาอาร์ สรุปลได้ดังรูปที่ 4

4. Compare Models with Statistical Measurement RMSE and MAPE คือ ขั้นตอนการเปรียบเทียบตัวแบบพยากรณ์ 3 ตัวแบบ ด้วยการวัดค่าคลาดเคลื่อนทางสถิติ คือ ค่า RMSE (Root Mean Square Error) และ MAPE (Mean Absolute Percentage Error) [10] ซึ่งมีสมการดังนี้

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2} \quad (6)$$

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| 100 \frac{y_t - \hat{y}_t}{y_t} \right| \quad (7)$$

เมื่อ y_t = ค่าสังเกตที่เวลา t , $\hat{y}_t$ = ค่าพยากรณ์ที่เวลา t และ n = จำนวนคาบเวลาที่พยากรณ์

5. Create the Best Forecasting Model in Term of Mathematic Equation คือ ขั้นตอนการสร้างตัวแบบการพยากรณ์ที่เหมาะสมที่สุด ในรูปของสมการคณิตศาสตร์ ที่สามารถนำไปพัฒนาเป็นโปรแกรมคอมพิวเตอร์ สำหรับพยากรณ์ข้อมูลหน่วยจำหน่ายไฟฟ้าประเภทบ้านอยู่อาศัยของการไฟฟ้านครหลวง ซึ่งตัวแบบการพยากรณ์ที่เหมาะสมที่สุด คือ ตัวแบบที่ให้ค่า RMSE และ MAPE ต่ำที่สุด

```
myDFlr <- read.csv(file="d:/R/EuseLRmodel.csv", header=TRUE, sep=",")
glm.D9 <- glm(myDFlr[,4] ~ myDFlr[,3])
```

(ก) คำสั่งสร้างตัวแบบ glm (Generalized Linear Models)

```
library("forecast", lib.loc=~ /R/win-library/3.2")
myDF53_57 <- read.csv(file="d:/R/Euse53_57.csv", header=TRUE, sep=",")
myts53_57 <- ts(myDF53_57[,3], start=c(2553,1), end=c(2557,12), frequency=12)
HWfit <- HoltWinters(myts53_57)
```

(ข) คำสั่งสร้างตัวแบบ HoltWinters (Holt-Winters)

```
library("forecast", lib.loc=~ /R/win-library/3.2")
myDF53_57 <- read.csv(file="d:/R/Euse53_57.csv", header=TRUE, sep=",")
myts53_57 <- ts(myDF53_57[,3], start=c(2553,1), end=c(2557,12), frequency=12)
auto.arima(myts53_57)
fitM53_57 <- arima(myts53_57, order=c(1,0,0), list(order=c(1,0,0), period=12))
```

(ค) คำสั่งสร้างตัวแบบ ARIMA (Autoregressive Integrated Moving Average)

รูปที่ 4 ชุดคำสั่งภาษาอาร์ สำหรับสร้างตัวแบบอนุกรมเวลาที่แตกต่างกัน 3 ตัวแบบ

ผลการวิจัย

ผู้วิจัยนำตัวแบบพยากรณ์ทั้ง 3 ตัวแบบ มาพยากรณ์ข้อมูลหน่วยจำหน่ายไฟฟ้า 5 คาบเวลา ตั้งแต่เดือนมกราคมถึงเดือนพฤษภาคม ปี พ.ศ. 2558 เพื่อนำค่าพยากรณ์ไปวัดค่าคลาดเคลื่อน RMSE และ MAPE โดยเปรียบเทียบกับค่าสังเกตจริงที่แบ่งไว้เป็นชุดข้อมูลทดสอบ ซึ่งผลการวิจัยพบว่า

1. การพยากรณ์ข้อมูลด้วยตัวแบบพยากรณ์

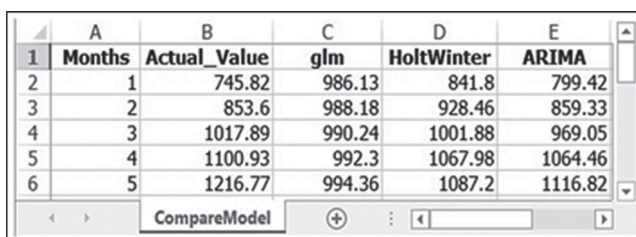
คำสั่งในภาษาอาร์ที่ใช้สำหรับพยากรณ์ข้อมูลหน่วยจำหน่ายไฟฟ้าจากตัวแบบพยากรณ์ และค่าพยากรณ์ 5 คาบเวลาของแต่ละตัวแบบ เป็นดังนี้

1. ตัวแบบ glm มีคำสั่งพยากรณ์ข้อมูลและค่าพยากรณ์ เป็นดังรูปที่ 5



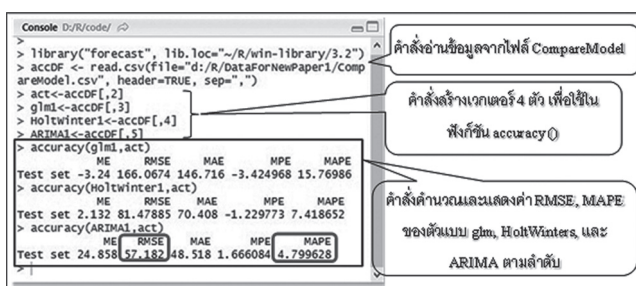
ค่าพยากรณ์กับค่าสังเกตจริง โดยตัดเครื่องหมายบวกและลบออก เพื่อมิให้ค่าคลาดเคลื่อนที่เป็นบวกและลบหักล้างกันเอง ดังแสดงในสมการที่ (1) และ (2) ดังนั้น ค่า RMSE และค่า MAPE จึงมีความเหมาะสมในการนำมาใช้วัดความแม่นยำของการพยากรณ์ได้ดี ซึ่งตัวแบบที่เหมาะสมมากที่สุด คือ ตัวแบบที่มีความแม่นยำในการพยากรณ์มากที่สุด หรือให้ค่า RMSE และค่า MAPE ต่ำที่สุด

คำสั่งในภาษาอาร์สำหรับคำนวณค่า RMSE และค่า MAPE คือ การเรียกใช้ฟังก์ชัน `accuracy()` ซึ่งเป็นฟังก์ชันที่อยู่ในแพ็คเกจ `forecast` โดยมีการติดตั้งแพ็คเกจไว้ในขั้นตอนการสร้างตัวแบบ `HoltWinters` และตัวแบบ `ARIMA` เรียบร้อยแล้ว สำหรับการเรียกใช้ฟังก์ชัน `accuracy()` ต้องส่งค่าพยากรณ์และค่าจริงในรูปแบบของเวกเตอร์ให้กับฟังก์ชัน ดังนั้นจึงนำค่าจริงและค่าพยากรณ์ที่ได้จากตัวแบบทั้ง 3 มาสร้างเป็นไฟล์ชนิด CSV จากนั้นใช้คำสั่งอ่านข้อมูลจากไฟล์ แล้วใช้คำสั่งสร้างเวกเตอร์ขึ้นมา 4 ตัว เพื่อเก็บค่าจริงและค่าพยากรณ์ที่ได้จากตัวแบบทั้ง 3 โดยตั้งชื่อไฟล์ชนิด CSV ด้วยชื่อ “CompareModel” ซึ่งมีข้อมูลเป็นดังรูปที่ 8 และชุดคำสั่งในภาษาอาร์ สำหรับคำนวณค่า RMSE และ MAPE ของตัวแบบทั้ง 3 แสดงได้ดังรูปที่ 9



	A	B	C	D	E
1	Months	Actual_Value	glm	HoltWinter	ARIMA
2	1	745.82	986.13	841.8	799.42
3	2	853.6	988.18	928.46	859.33
4	3	1017.89	990.24	1001.88	969.05
5	4	1100.93	992.3	1067.98	1064.46
6	5	1216.77	994.36	1087.2	1116.82

รูปที่ 8 ไฟล์ชนิด CSV ชื่อ “CompareModel”



```

> library("forecast", lib.loc="/R/win-library/3.2")
> accDF <- read.csv(file="d:/DataForNewPaper1/compareModel.csv", header=TRUE, sep=",")
> act<-accDF[,2]
> glm1<-accDF[,3]
> Holtwinter1<-accDF[,4]
> ARIMA1<-accDF[,5]
> accuracy(glm1,act)
      ME      RMSE      MAE      MPE      MAPE
Test set -3.24 166.0674 146.716 -3.424968 15.76986
> accuracy(Holtwinter1,act)
      ME      RMSE      MAE      MPE      MAPE
Test set 2.132 81.47885 70.408 -1.229773 7.418652
> accuracy(ARIMA1,act)
      ME      RMSE      MAE      MPE      MAPE
Test set 24.858 57.182 48.518 1.666084 4.799628
  
```

คำสั่งอ่านข้อมูลจากไฟล์ CompareModel

คำสั่งสร้างเวกเตอร์ 4 ตัว เพื่อใช้ฟังก์ชัน `accuracy()`

คำสั่งคำนวณและแสดงค่า RMSE, MAPE ของตัวแบบ `glm`, `HoltWinters`, และ `ARIMA` ตามลำดับ

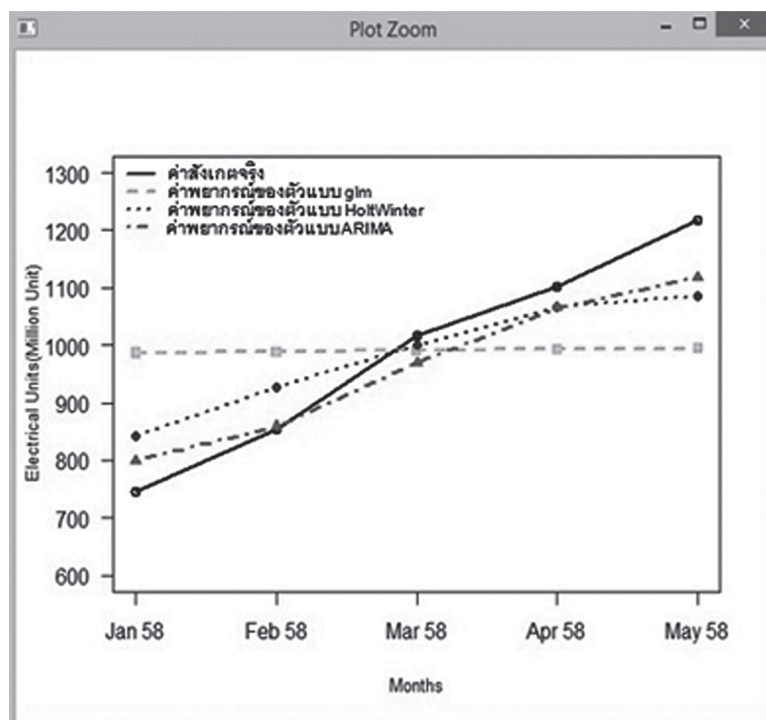
รูปที่ 9 ชุดคำสั่งในภาษาอาร์ สำหรับคำนวณค่า RMSE และ MAPE

จากรูปที่ 9 ตัวแบบ `ARIMA` มีค่า RMSE และค่า MAPE ต่ำที่สุด ซึ่งมีค่าเท่ากับ 57.18 และ 4.80 ตามลำดับ ดังนั้น ตัวแบบ `ARIMA` เป็นตัวแบบที่มีความเหมาะสมที่สุดสำหรับการพยากรณ์ข้อมูลอนุกรมเวลา โดยสามารถแสดงค่าพยากรณ์ 5 คาบเวลา ค่า RMSE และค่า MAPE ของแต่ละตัวแบบได้ดังตารางที่ 1 นอกจากนี้ เมื่อนำค่าสังเกตจริงและค่าพยากรณ์ที่ได้จากตัวแบบทั้ง 3 มาสร้างเป็นกราฟเส้น ดังรูปที่ 10 ทำให้ทราบว่า ตัวแบบ `glm` ที่สร้างจากวิธีการวิเคราะห์ความถดถอยเชิงเส้นไม่เหมาะสำหรับนำมาใช้พยากรณ์อนุกรมเวลา เพราะเมื่อนำอนุกรมเวลามาสร้างกราฟอนุกรมเวลา

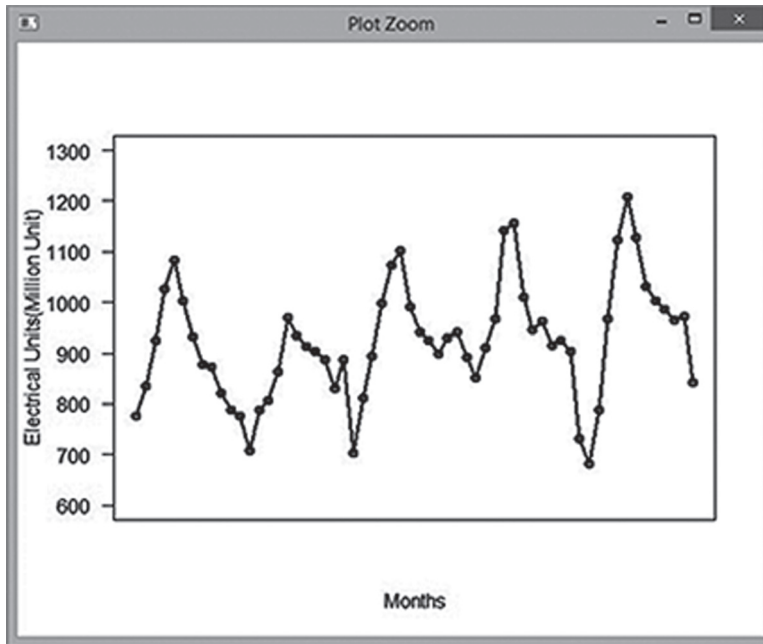
ดังรูปที่ 11 จากลักษณะของกราฟอนุกรมเวลา แสดงให้เห็นว่า อนุกรมเวลามีแนวโน้มเป็นเส้นตรงเชิงบวก และมีส่วนประกอบฤดูกาล เพราะมีรูปแบบการเกิดซ้ำตามรอบฤดูกาล ซึ่งตัวแบบ glm ที่สร้างจากวิธีการวิเคราะห์ความถดถอยเชิงเส้น ไม่สามารถตรวจจับความเคลื่อนไหวเนื่องจากแนวโน้มและฤดูกาลได้ โดยสอดคล้องกับค่า RMSE และค่า MAPE ของตัวแบบ glm ที่มีค่ามากที่สุด ส่วนตัวแบบ HoltWinters และตัวแบบ ARIMA สามารถตรวจจับความเคลื่อนไหวเนื่องจากแนวโน้มและฤดูกาลได้ แต่สำหรับอนุกรมเวลาชุดนี้ ตัวแบบ ARIMA มีความเหมาะสมมากกว่าตัวแบบ HoltWinters เนื่องจากให้ค่า RMSE และค่า MAPE ต่ำที่สุด

ตารางที่ 1 ค่าพยากรณ์ค่า RMSE และค่า MAPE ของแต่ละตัวแบบ

หน่วยจำหน่ายไฟฟ้า	ม.ค. 2558	ก.พ. 2558	มี.ค. 2558	เม.ย. 2558	พ.ค. 2558	RMSE	MAPE
ค่าสังเกตจริง	745.82	853.60	1017.89	1100.93	1216.77	-	-
ค่าพยากรณ์ของตัวแบบ glm	986.13	988.18	990.24	992.30	994.36	166.07	15.77
ค่าพยากรณ์ของตัวแบบ HoltWinters	841.80	928.46	1001.88	1067.98	1087.20	81.48	7.42
ค่าพยากรณ์ของตัวแบบ ARIMA	799.24	859.33	969.05	1064.46	1116.82	57.18	4.80



รูปที่ 10 กราฟเปรียบเทียบระหว่างค่าสังเกตจริงและค่าพยากรณ์ 5 คาบเวลา



รูปที่ 11 กราฟอนุกรมเวลาสำหรับสร้างตัวแบบการพยากรณ์

3. การสร้างสมการพยากรณ์ของตัวแบบ ARIMA

เมื่อได้ตัวแบบ ARIMA เป็นตัวแบบการพยากรณ์ที่เหมาะสมที่สุด สามารถสร้างสมการพยากรณ์หน่วยจำหน่ายไฟฟ้าประเภทบ้านอยู่อาศัยของการไฟฟ้านครหลวง ที่ได้จากการแปลงรูปแบบทั่วไป (General Form) ของตัวแบบ ARIMA ด้วยการกำหนดรูปแบบที่เหมาะสมกับอนุกรมเวลา และประมาณค่าพารามิเตอร์ที่ไม่ทราบค่า โดยใช้ฟังก์ชัน `auto.arima()` ดังรูปที่ 12 ซึ่งทำให้ทราบว่า ตัวแบบ ARIMA สำหรับอนุกรมเวลาชุดนี้ ประกอบด้วย Autoregressive(AR) อันดับ $p=1$ และ Seasonal Autoregressive(SAR) อันดับ $P=1$ ซึ่งประมาณค่าสัมประสิทธิ์ $AR_1 (\hat{\theta}_1)$ และ $SAR_1 (\hat{\Theta}_1)$ ได้เท่ากับ 0.7301 และ 0.6403 โดยที่อันดับ $d=D=0$ และอันดับ $q=Q=0$ ดังนั้น สามารถสร้างสมการพยากรณ์หน่วยจำหน่ายไฟฟ้าประเภทบ้านอยู่อาศัยของการไฟฟ้านครหลวง ได้ดังนี้

$$(1 - \hat{\theta}_1 B)(1 - \hat{\Theta}_1 B^{12})Y_t = \theta_0 \quad (8)$$

เมื่อ $\theta_0 =$ ค่าคงที่

$$\begin{aligned} (1 - \hat{\theta}_1 B - \hat{\Theta}_1 B^{12} + \hat{\theta}_1 \hat{\Theta}_1 B^{13})Y_t &= \theta_0 \\ Y_t - \hat{\theta}_1 Y_{t-1} - \hat{\Theta}_1 Y_{t-12} + \hat{\theta}_1 \hat{\Theta}_1 Y_{t-13} &= \theta_0 \\ Y_t &= \theta_0 + \hat{\theta}_1 Y_{t-1} + \hat{\Theta}_1 Y_{t-12} - \hat{\theta}_1 \hat{\Theta}_1 Y_{t-13} \\ Y_t &= \theta_0 + 0.73Y_{t-1} + 0.64Y_{t-12} - 0.47Y_{t-13} \end{aligned} \quad (9)$$

จากค่าสังเกตของอนุกรมเวลา ทำให้ทราบว่า ค่าคงที่ $\theta_0 = 92.42$ ดังนั้น ได้สมการพยากรณ์เป็นดังสมการที่ (10)

$$Y_t = 92.42 + 0.73Y_{t-1} + 0.64Y_{t-12} - 0.47Y_{t-13} \quad (10)$$

```

> auto.arima(myts53_57)
Series: myts53_57
ARIMA(1,0,0)(1,0,0)[12] with non-zero mean
Coefficients:
ar1      sar1      intercept
0.7301    0.6403    929.9639
s.e.      0.0835    0.0984    57.2129

sigma^2 estimated as 3320: log likelihood=-331.93
AIC=671.85  AICC=672.58  BIC=680.23
>

```

รูปที่ 12 ฟังก์ชัน auto.arima() สำหรับกำหนดรูปแบบของตัวแบบ ARIMA

เมื่อสามารถสร้างตัวแบบ ARIMA ซึ่งเป็นตัวแบบการพยากรณ์ที่เหมาะสมที่สุด ในรูปของสมการคณิตศาสตร์ ได้ดังสมการที่ (10) โดยอยู่ในรูปของสมการถดถอยเชิงเส้น ที่มีตัวแปรอิสระ 3 ตัว คือ Y_{t-1} , Y_{t-12} และ Y_{t-13} หมายถึงค่าสังเกตที่ล่าหลังไป 1, 12 และ 13 คาบเวลา ตามลำดับ และตัวแปรตามคือ Y_t หมายถึง ค่าสังเกต ณ คาบเวลา t ใด ๆ ดังนั้น สามารถนำสมการที่ (10) ไปใช้สร้างโปรแกรมสำหรับพยากรณ์หน่วยจำหน่ายไฟฟ้ารายเดือน ประเภทบ้านพักอาศัยของการไฟฟ้านครหลวงได้ดังรูปที่ 13

รูปที่ 13 ตัวอย่างโปรแกรมพยากรณ์ของตัวแบบ ARIMA

การอภิปรายผล

จากผลการวิจัยพบว่า อนุกรมเวลาที่ศึกษามีการเคลื่อนไหวเนื่องจากแนวโน้มและฤดูกาล เพราะลักษณะของกราฟอนุกรมเวลามีแนวโน้มเป็นเส้นตรงเชิงบวก และมีรูปแบบการเกิดซ้ำตามรอบฤดูกาล เมื่อนำอนุกรมเวลามาสร้างตัวแบบการพยากรณ์ด้วยวิธีการวิเคราะห์ความถดถอยเชิงเส้น วิธีการปรับให้เรียบ

แบบเอ็กซ์โปเนนเชียล และวิธีของบ็อกซ์และเจนกินส์ แล้วทำการเปรียบเทียบตัวแบบการพยากรณ์ด้วยการวัดค่า RMSE และค่า MAPE พบว่า ตัวแบบ ARIMA มีค่า RMSE และค่า MAPE ต่ำที่สุด โดยมีค่า RMSE เท่ากับ 57.18 และค่า MAPE เท่ากับ 4.80 ซึ่งค่า MAPE มีค่าอยู่ในเกณฑ์ที่ต่ำหมายความว่าตัวแบบการพยากรณ์มีความแม่นยำในการพยากรณ์สูง ดังนั้น ตัวแบบ ARIMA จึงเป็นตัวแบบที่เหมาะสมสำหรับการพยากรณ์ข้อมูลอนุกรมเวลาชุดดังกล่าวนี้ ซึ่งสอดคล้องกับงานวิจัยในอดีตของ Chujai, P. et al. [6] และงานวิจัยของ Wang, Y. et al. [7] ที่พบว่า ตัวแบบ ARIMA มีความเหมาะสมสำหรับการนำมาใช้พยากรณ์อนุกรมเวลาที่มีส่วนประกอบของแนวโน้มหรือฤดูกาล และจากการใช้ฟังก์ชัน `auto.arima()` เพื่อกำหนดรูปแบบของตัวแบบ ARIMA พบว่า ตัวแบบ ARIMA มีส่วนประกอบของ Autoregressive ที่อันดับ $p=1$ และส่วนประกอบของ Seasonal Autoregressive ที่อันดับ $P=1$ ซึ่งหมายความว่า ค่าสังเกต ณ เวลาใด ขึ้นอยู่กับค่าสังเกตก่อนหน้า 1 คาบเวลา และขึ้นอยู่กับค่าสังเกตก่อนหน้า 1 ฤดูกาล แต่ไม่มีส่วนประกอบของ Moving Average และ Seasonal Moving Average เพราะอันดับ $q=Q=0$ หมายความว่า ค่าสังเกต ณ เวลาใด ๆ ไม่ขึ้นอยู่กับค่าคลาดเคลื่อนของการพยากรณ์และอนุกรมเวลามีลักษณะนิ่ง (Stationary) เนื่องจากอันดับ $d=D=0$ ทั้งนี้ จากรูปแบบของตัวแบบ ARIMA สามารถสร้างสมการพยากรณ์ของตัวแบบ ARIMA เพื่อใช้พยากรณ์ข้อมูลล่วงหน้าได้ ดังสมการที่ (10) และนำมาใช้พยากรณ์หน่วยจำหน่ายไฟฟ้ารายเดือนของผู้บริโภคประเภทบ้านพักอาศัย ประจำปี พ.ศ. 2558 ได้ดังตารางที่ 2

ตารางที่ 2 ค่าพยากรณ์หน่วยจำหน่ายไฟฟ้ารายเดือน ประจำปีพ.ศ. 2558 (หน่วยเป็น ล้านบาท)

ม.ค.	ก.พ.	มี.ค.	เม.ย.	พ.ค.	มิ.ย.	ก.ค.	ส.ค.	ก.ย.	ต.ค.	พ.ย.	ธ.ค.
799.42	859.33	969.05	1064.46	1116.82	1062.40	998.62	980.24	968.71	954.07	957.64	874.13

บทสรุป

จากวัตถุประสงค์การวิจัย เพื่อศึกษาผลการวิเคราะห์ข้อมูลอนุกรมเวลาในการค้นหาตัวแบบที่เหมาะสมที่สุด โดยการเขียนโปรแกรมด้วยภาษาอาร์ พบว่า ข้อมูลหน่วยจำหน่ายไฟฟ้าของการไฟฟ้านครหลวงสำหรับผู้บริโภคประเภทบ้านพักอาศัย เป็นอนุกรมเวลาที่มีส่วนประกอบของแนวโน้มและฤดูกาล ซึ่งตัวแบบ ARIMA เป็นตัวแบบที่มีความเหมาะสมมากที่สุด สำหรับนำมาใช้ในการพยากรณ์ข้อมูล เพราะสามารถตรวจจับการเคลื่อนไหวของอนุกรมเวลาเนื่องจากแนวโน้มหรือฤดูกาลได้ดี และจากรูปที่ 10 เมื่อพิจารณาลักษณะของกราฟค่าพยากรณ์ของตัวแบบ `glm` สามารถสรุปเพิ่มเติมได้ว่า ตัวแบบ `glm` ที่สร้างจากวิธีการวิเคราะห์ความถดถอยเชิงเส้น ไม่เหมาะสมสำหรับนำมาใช้ในการพยากรณ์อนุกรมเวลาที่มีการเคลื่อนไหวเนื่องจากแนวโน้มหรือฤดูกาล เพราะไม่สามารถตรวจจับการเคลื่อนไหวของอนุกรมเวลาเนื่องจากแนวโน้มหรือฤดูกาลได้ นอกจากนี้ การวิจัยนี้ แสดงให้เห็นถึงแนวทางการวิเคราะห์อนุกรมเวลาโดยการเขียนโปรแกรมด้วยภาษาอาร์ ซึ่งช่วยให้นักวิจัยสามารถวิเคราะห์อนุกรมเวลาได้ง่าย สะดวก และรวดเร็วขึ้น โดยลดขั้นตอนการวิเคราะห์ข้อมูลบางขั้นตอน เช่น ขั้นตอนการสร้างตัวแบบ ARIMA ต้องมีการตรวจสอบสถานะนิ่ง และกำหนดรูปแบบของตัวแบบ ARIMA แต่ในภาษาอาร์ มีฟังก์ชัน

auto.arima() ที่ใช้กำหนดรูปแบบของตัวแบบ ARIMA โดยอัตโนมัติ ทำให้ผู้วิจัยไม่ต้องวิเคราะห์อนุกรมเวลา เพื่อกำหนดรูปแบบของตัวแบบ ARIMA อีกทั้งภาษาอาร์มีคำสั่งสร้างกราฟ ทำให้สามารถวิเคราะห์ข้อมูลจากกราฟได้ง่ายขึ้น

References

- [1] Metropolitan Electricity Authority. (2015). Report of Electrical Distribution. Access (10 August 2015). Available (<http://www.meo.or.th/download/index.php#>)
- [2] Bunn, D.W., and Farmer, E.D. (1985). Comparative Models for Electrical Load Forecasting. New York: John Wiley and Sons
- [3] Maçaira, P.M., Souza, R.C. and Oliveira, F.L.C. (2015). Modelling and Forecasting the Residential Electricity Consumption in Brazil with Pegels Exponential Smoothing Techniques. *Procedia Computer Science*. Vol. 55. p. 328-335
- [4] Pereira, C.M., Almeida de, N.N. and Velloso, M.L.F. (2015). Fuzzy Modeling to Forecast an Electric Load Times Series. *Procedia Computer Science*. Vol. 55. pp. 395-404
- [5] Keka, I., and Hamiti, M. (2013). Load profile analyses using R language. *Proceedings of the ITI 2013 35th International Conference on Information Technology Interfaces (ITI)*. Croatia : Cavtat. pp. 245-250
- [6] Chujai, P., Kerdprasop, N. and Kerdprasop, K. (2013). Time Series Analysis of Household Electric Consumption with ARIMA and ARMA Models. *Proceedings of the International MultiConference of Engineers and Computer Scientists*. China : Hong Kong. pp. 295-300
- [7] Wang, Y., Wang, J., Zhao, G. and Dong, Y. (2012). Application of Residual Modification Approach in Seasonal ARIMA for Electricity Demand Forecasting: A Case Study of China. *Energy Policy*. Vol. 48. pp. 284-294
- [8] Boldina, I., and Beninger, P.G. (2016). Strengthening Statistical Usage in Marine Ecology: Linear Regression. *Journal of Experimental Marine Biology and Ecology*. Vol. 474. pp. 81-91
- [9] Prema, V. and Rao, K.U. (2015). Development of Statistical Time Series Models for Solar Power Prediction. *Renewable Energy*. Vol. 83. pp. 100-109
- [10] Bergmeir, C., and Benítez, J.M. (2012). On the Use of Cross-Validation for Time Series Predictor Evaluation. *Information Sciences*. Vol. 191. pp. 192-213