

การเปรียบเทียบวิธีการพยากรณ์ สำหรับอนุกรมเวลาที่มีลักษณะไม่เป็นเชิงเส้นและไม่คงที่

Comparison of Forecasting Methods for Non-Linear and Non-Stationary Time Series

ณัชพล อธิประยูร*, อำไพ ทองธีรภาพ และเสาวภา ชัยพิทักษ์

ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตบางเขน

ถนนงามวงศ์วาน แขวงลาดยาว เขตจตุจักร กรุงเทพมหานคร 10900

Natchapol Athipayoon*, Ampai Thongteeraparp and Saowapa Chaipitak

Department of Statistics, Faculty of Science, Kasetsart University, Bangkhen Campus,

Ngamwongwan Road, Lad Yao District, Chatuchak, Bangkok 10900

บทคัดย่อ

งานวิจัยนี้ศึกษาการพยากรณ์อนุกรมเวลาที่มีลักษณะไม่เป็นเชิงเส้นและไม่คงที่ด้วยวิธีการพยากรณ์ 3 วิธี คือ วิธีบ็อกซ์และเจนกินส์ (Box-Jenkins แบ่งอนุกรมเวลาที่จำลองขึ้นมาออกเป็น 2 ส่วน ส่วนที่ 1 มีจำนวน 150 ค่า ใช้สำหรับสร้างตัวแบบพยากรณ์ และส่วนที่ 2 มีจำนวน 20 ค่า ใช้สำหรับเปรียบเทียบประสิทธิภาพในการพยากรณ์ โดยใช้เกณฑ์เปรียบเทียบประสิทธิภาพในการพยากรณ์ คือ ค่า MSE และค่า MAPE ผลการศึกษาพบว่าวิธีบ็อกซ์และเจนกินส์ด้วยตัวแบบ ARIMA มีประสิทธิภาพในการพยากรณ์ดีที่สุดในการพยากรณ์อนุกรมเวลาที่มีแนวโน้มเอ็กซ์โพเนนเชียลที่มีเลขชี้กำลังบวกและโลจิสติก ในขณะที่วิธี EMD-ARIMA มีประสิทธิภาพในการพยากรณ์ดีที่สุดในการพยากรณ์อนุกรมเวลาที่มีแนวโน้มเอ็กซ์โพเนนเชียล โดยวิธี EEMD-ARIMA ด้วยตัวแบบ ARIMA วิธี empirical mode decomposition ด้วยตัวแบบ ARIMA ของวิธีบ็อกซ์และเจนกินส์ (EMD-ARIMA) และวิธี ensemble empirical mode decomposition ด้วยตัวแบบ ARIMA ของวิธีบ็อกซ์และเจนกินส์ (EEMD-ARIMA) โดยการจำลองข้อมูลอนุกรมเวลาจำนวน 170 ค่า ด้วยโปรแกรม R จำนวน 6 รูปแบบ คือ อนุกรมเวลาที่มีแนวโน้มโพลีโนเมียลอันดับสองและอันดับสาม อนุกรมเวลาที่มีแนวโน้มเอ็กซ์โพเนนเชียลที่มีเลขชี้กำลังบวกและกำลังลบ และอนุกรมเวลาที่มีแนวโน้มโลจิสติกที่มีเลขชี้กำลังบวกและกำลังลบ แททิภาพในการพยากรณ์ดีที่สุดในการพยากรณ์อนุกรมเวลาที่มีแนวโน้มโพลีโนเมียล เอ็กซ์โพเนนเชียลที่มีเลขชี้กำลังลบและโลจิสติก

คำสำคัญ : การจำลอง; วิธีการแยกส่วนประกอบ; วิธีบ็อกซ์และเจนกินส์

Abstract

This research studied non-linear and non-stationary time series for three forecasting methods that were Box-Jenkins method with ARIMA model, empirical mode decomposition with ARIMA model of Box-Jenkins method (EMD-ARIMA) and ensemble empirical mode decomposition with ARIMA model of Box-Jenkins method (EEMD-ARIMA). Time series data with 170 values were simulated by program R in six models consisting of second and third-order polynomial trend time series, exponential trend time series with positive and negative exponents and logistic trend time series with positive and negative exponents. The data were divided into two parts. The first part with 150 values, was used to create forecasting model and the second part with 20 values, was used for comparing efficiency of forecasting methods. Efficiency criteria were MSE and MAPE. The results showed that the Box-Jenkins method with ARIMA model was the most efficient method for forecasting exponential trend time series with positive exponent and logistic trend time series, whereas the EMD-ARIMA method was the most efficient method for exponential trend time series. In addition, the EEMD-ARIMA method was the most efficient method for forecasting polynomial, exponential trend time series with positive exponent and logistic trend time series.

Keywords: simulation; decomposition method; Box-Jenkins method

1. บทนำ

การพยากรณ์ในอนุกรมเวลา คือ การทำนายค่าของอนุกรมเวลาในอนาคตล่วงหน้าโดยใช้อนุกรมเวลาในอดีตและปัจจุบันเพื่อช่วยในการพยากรณ์ ซึ่งการพยากรณ์ในอนุกรมเวลานั้นมีความสำคัญต่อผู้คนในปัจจุบันอย่างมาก เนื่องจากถ้าสามารถพยากรณ์อนุกรมเวลาชนิดต่าง ๆ เช่น อนุกรมเวลาราคาหุ้น อนุกรมเวลาปริมาณน้ำฝน อนุกรมเวลาการใช้พลังงานไฟฟ้าในแต่ละปี จะทำให้ทราบว่าจะเกิดเหตุการณ์ในอนาคตเป็นอย่างไร หรือมีแนวโน้มเป็นอย่างไร ซึ่งจะช่วยในการรับมือกับเหตุการณ์ในอนาคตได้ อย่างไรก็ตาม การพยากรณ์อนุกรมเวลาที่ดีจะต้องมีความแม่นยำสูงและพยากรณ์ได้ใกล้เคียงกับค่าจริงมากที่สุด [1]

การพยากรณ์อนุกรมเวลาในปัจจุบันมีหลากหลายวิธี เช่น การพยากรณ์โดยใช้ค่าเฉลี่ยเคลื่อนที่

การแยกส่วนประกอบของอนุกรมเวลา การพยากรณ์โดยใช้ตัวแบบบ็อกซ์และเจนกินส์ ซึ่งงานวิจัยนี้ได้เลือกใช้วิธีการพยากรณ์โดยใช้วิธีบ็อกซ์และเจนกินส์ด้วยตัวแบบ ARIMA ซึ่งเป็นวิธีการที่มีการใช้อย่างแพร่หลายวิธีหนึ่ง วิธีบ็อกซ์และเจนกินส์ด้วยตัวแบบ ARIMA มีหลักการ คือ พิจารณาสัมประสิทธิ์สหสัมพันธ์ในตนเองและสัมประสิทธิ์สหสัมพันธ์บางส่วนในตนเองของอนุกรมเวลา แล้วกำหนดตัวแบบที่เหมาะสมให้กับอนุกรมเวลานั้น อย่างไรก็ตาม วิธีการพยากรณ์อนุกรมเวลาด้วยตัวแบบบ็อกซ์และเจนกินส์ยังพยากรณ์อนุกรมเวลาที่มีลักษณะไม่เป็นเชิงเส้น (non-linear time series) ได้ไม่ดีนัก [2] จึงมีผู้เสนอวิธีการผสมที่มีการใช้ตัวแบบบ็อกซ์และเจนกินส์ ซึ่งสามารถช่วยในการพยากรณ์อนุกรมเวลาที่มีลักษณะไม่เป็นเชิงเส้นดีขึ้น

ในปี ค.ศ. 2014 Abadan และ Shabri [3] ศึกษาเปรียบเทียบประสิทธิภาพการพยากรณ์ราคาข้าวระหว่างวิธีบ็อกซ์และเจนกินส์ด้วยตัวแบบ ARIMA และวิธี EMD-ARIMA ซึ่งข้อสรุปที่ได้ คือ วิธี EMD-ARIMA มีประสิทธิภาพในการพยากรณ์ดีกว่าวิธีตัวแบบบ็อกซ์และเจนกินส์ และในปี ค.ศ. 2016 Zhou และ Huang [4] ศึกษาการพยากรณ์อายุของแบตเตอรี่ลิเทียมไอออนด้วยวิธีการ 4 วิธี ได้แก่ EMD-ARIMA, ARIMA, relevance vector machine (RVM) และ ensemble echo state networks (EESN) เมื่อทดสอบประสิทธิภาพด้วยค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (mean absolute error, MAE) พบว่าวิธี EMD-ARIMA มีประสิทธิภาพสูงที่สุด

ในปี ค.ศ. 2012 Bao และคณะ [2] ศึกษาการพยากรณ์ความเชื่อมั่นของซอฟต์แวร์โดยใช้วิธี EEMD-ARIMA พบว่าวิธีการดังกล่าวมีประสิทธิภาพสูงสุดเมื่อเปรียบเทียบกับวิธี neural network และ support vector machine และวิธี EEMD-ARIMA โดยข้อดีของ EEMD คือ ช่วยทำให้อนุกรมเวลามีลักษณะเป็นเชิงเส้นและคงที่มากขึ้น และมีความยุ่งยากในการใช้งานน้อยกว่าวิธี neural network และ support vector machine

ในปี ค.ศ. 2017 Min และคณะ ศึกษาการพยากรณ์ความเร็วลมซึ่งมีลักษณะของอนุกรมเวลาที่ไม่เป็นเชิงเส้นและไม่คงที่ด้วยวิธี EEMD-ARIMA และเปรียบเทียบกับวิธีบ็อกซ์และเจนกินส์และวิธี EMD-ARIMA โดยใช้เกณฑ์เปรียบเทียบ คือ ค่าเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (mean absolute percentage error, MAPE) ค่าเฉลี่ยความคลาดเคลื่อนสัมบูรณ์ (mean absolute error, MAE) และค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (root mean square error, RMSE) ผลที่ได้ คือ วิธี EEMD-ARIMA มีประสิทธิภาพในการพยากรณ์สูงกว่า

อีกสองวิธี

ผลสรุปของงานวิจัยที่กล่าวมาข้างต้นจะเห็นว่าวิธี EMD-ARIMA และวิธี EEMD-ARIMA มีประสิทธิภาพในการพยากรณ์สูงกว่าวิธีพยากรณ์โดยใช้ตัวแบบบ็อกซ์และเจนกินส์ด้วยตัวแบบ ARIMA กับข้อมูลอนุกรมเวลาที่มีอยู่จริง จึงทำให้มีสนใจการพยากรณ์ด้วยวิธี EMD-ARIMA และวิธี EEMD-ARIMA และได้เปรียบเทียบประสิทธิภาพของวิธีการพยากรณ์ทั้ง 2 วิธีกับวิธีการพยากรณ์ด้วยตัวแบบบ็อกซ์และเจนกินส์ โดยการจำลองอนุกรมเวลาที่มีลักษณะไม่เป็นเชิงเส้นและไม่คงที่ 6 รูปแบบ ซึ่งอนุกรมเวลาที่จำลองทั้ง 6 รูปแบบ คือ อนุกรมเวลาที่มีแนวโน้มโพลีโนเมียลอันดับสองและอันดับสาม อนุกรมเวลาที่มีแนวโน้มเอ็กซ์โปเนนเชียลที่มีเลขชี้กำลังบวกและกำลังลบ และอนุกรมเวลาที่มีแนวโน้มโลจิสติกที่มีเลขชี้กำลังบวกและกำลังลบ และเปรียบเทียบประสิทธิภาพการพยากรณ์ของวิธีการพยากรณ์ทั้ง 3 วิธี ด้วยค่าเฉลี่ยความคลาดเคลื่อนกำลังสองและค่าเฉลี่ยของเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์

2. วิธีการดำเนินการวิจัย

2.1 การจำลองอนุกรมเวลา

งานวิจัยนี้ศึกษาประสิทธิภาพการพยากรณ์ของวิธีบ็อกซ์และเจนกินส์ด้วยตัวแบบ ARIMA วิธี EMD-ARIMA และวิธี EEMD-ARIMA ในการพยากรณ์อนุกรมเวลาที่มีลักษณะไม่เป็นเชิงเส้นและไม่คงที่ โดยจำลองอนุกรมเวลา (y_t) ที่มีลักษณะไม่เป็นเชิงเส้นและไม่คงที่ 6 รูปแบบดังรูปที่ 1 [6-8] โดยใช้โปรแกรม R เวอร์ชัน 3.4.1 กำหนดให้ข้อมูลอนุกรมเวลา (n) เท่ากับ 170 ซึ่งรายละเอียดของอนุกรมเวลาทั้ง 6 รูปแบบ มีดังนี้

2.1.1 อนุกรมเวลารูปแบบที่ 1 คือ อนุกรมเวลาที่มีแนวโน้มโพลีโนเมียลอันดับสอง มีตัวแบบ คือ

$y_t = at^2 + bt + c + \omega_t$ กำหนดให้ $a=1, b=-1, c=-2$ และ ω_t แทนตัวแปรปรบกวนโดยสุ่ม (noise) ที่เวลา t มีการแจกแจงปกติ มีค่าเฉลี่ยเท่ากับ 0 และมีความแปรปรวนไม่คงที่เท่ากับ σ_t^2 โดยที่ σ_t^2 มีการแจกแจงปกติ มีค่าเฉลี่ยเท่ากับ 0 และมีความแปรปรวนเท่ากับ 3,000

2.1.2 อนุกรมเวลารูปแบบที่ 2 คือ อนุกรมเวลาที่มีแนวโน้มโพลิโนเมียลอันดับสาม มีตัวแบบ คือ $y_t = at^3 + bt^2 + ct + d + \omega_t$ กำหนดให้ $a=1, b=1, c=-1, d=-2$ และ ω_t แทนตัวแปรปรบกวนโดยสุ่มที่เวลา t มีการแจกแจงปกติ มีค่าเฉลี่ยเท่ากับ 0 และมีความแปรปรวนไม่คงที่เท่ากับ σ_t^2 โดยที่ σ_t^2 มีการแจกแจงปกติ มีค่าเฉลี่ยเท่ากับ 0 และมีความแปรปรวนเท่ากับ 200,000

2.1.3 อนุกรมเวลารูปแบบที่ 3 คือ อนุกรมเวลาที่มีแนวโน้มเอ็กซ์โปเนนเชียลที่มีเลขชี้กำลังบวก มีตัวแบบ คือ $y_t = ab^t + \omega_t$ กำหนดให้ $a=1, b=1.01$ และ ω_t แทนตัวแปรปรบกวนโดยสุ่มที่เวลา t มีการแจกแจงปกติ มีค่าเฉลี่ยเท่ากับ 0 และมีความแปรปรวนไม่คงที่เท่ากับ σ_t^2 โดยที่ σ_t^2 มีการแจกแจงปกติ มีค่าเฉลี่ยเท่ากับ 0 และมีความแปรปรวนเท่ากับ 0.5

2.1.4 อนุกรมเวลารูปแบบที่ 4 คือ อนุกรมเวลาที่มีแนวโน้มเอ็กซ์โปเนนเชียลที่มีเลขชี้กำลังติดลบ มีตัวแบบ คือ $y_t = ab^{-t} + \omega_t$ กำหนดให้ $a=1, b=1.01$ และ ω_t แทนตัวแปรปรบกวนโดยสุ่มที่เวลา t มีการแจกแจงปกติ มีค่าเฉลี่ยเท่ากับ 0 และมีความแปรปรวนไม่คงที่เท่ากับ σ_t^2 โดยที่ σ_t^2 มีการแจกแจงปกติ มีค่าเฉลี่ยเท่ากับ 0 และมีความแปรปรวนเท่ากับ 0.1

2.1.5 อนุกรมเวลารูปแบบที่ 5 คือ อนุกรมเวลาที่มีแนวโน้มโลจิสติกที่มีเลขชี้กำลังบวก มีตัวแบบ

คือ $y_t = (1+ab^t)^{-1} + \omega_t$ กำหนดให้ $a=1, b=1.1$ และ ω_t แทนตัวแปรปรบกวนโดยสุ่มที่เวลา t มีการแจกแจงปกติ มีค่าเฉลี่ยเท่ากับ 0 และมีความแปรปรวนไม่คงที่เท่ากับ σ_t^2 โดยที่ σ_t^2 คือ มีการแจกแจงปกติ มีค่าเฉลี่ยเท่ากับ 0 และมีความแปรปรวนเท่ากับ 0.05

2.1.6 อนุกรมเวลารูปแบบที่ 6 คือ อนุกรมเวลาที่มีแนวโน้มโลจิสติกที่มีเลขชี้กำลังลบ มีตัวแบบ คือ $y_t = (1+ab^{-t})^{-1} + \omega_t$ กำหนดให้ $a=1, b=1.1$ และ ω_t แทนตัวแปรปรบกวนโดยสุ่มที่เวลา t มีการแจกแจงปกติ มีค่าเฉลี่ยเท่ากับ 0 และมีความแปรปรวนไม่คงที่เท่ากับ σ_t^2 โดยที่ σ_t^2 มีการแจกแจงปกติ มีค่าเฉลี่ยเท่ากับ 0 และมีความแปรปรวนเท่ากับ 0.05

หลังจากนั้นแบ่งอนุกรมเวลาที่จำลองขึ้นทั้งหมด 170 ค่า เป็น 2 ส่วน โดยอนุกรมเวลาส่วนที่ 1 มีจำนวน 150 ค่าแรกใช้สำหรับสร้างตัวแบบพยากรณ์ และอนุกรมเวลาส่วนที่ 2 มีจำนวน 20 ค่าใช้สำหรับตรวจสอบค่าพยากรณ์ที่ได้จากตัวแบบพยากรณ์จากอนุกรมเวลาส่วนที่ 1

2.2 วิธีการพยากรณ์อนุกรมเวลา

2.2.1 วิธีบ็อกซ์และเจนกินส์ด้วยตัวแบบ ARIMA เป็นวิธีการพยากรณ์ที่มีหลักการ คือ กำหนดตัวแบบพยากรณ์โดยการพิจารณาฟังก์ชันสหสัมพันธ์ในตนเอง (autocorrelation function, ACF) และฟังก์ชันสหสัมพันธ์ในตนเองบางส่วน (partial autocorrelation function, PACF) ของอนุกรมเวลาที่มีลักษณะคงที่ (stationary time series) หรืออนุกรมเวลาที่มีค่าเฉลี่ยและความแปรปรวนคงที่ [9] ดังนั้นก่อนที่จะกำหนดตัวแบบ autoregressive integrated moving average (ARIMA) ให้กับอนุกรมเวลา จะต้องมีการตรวจสอบว่าอนุกรมเวลามีลักษณะ

คงที่หรือไม่ ถ้าอนุกรมเวลายังมีค่าเฉลี่ยไม่คงที่สามารถหาผลต่างของอนุกรมเวลา [1] และถ้าหากอนุกรมเวลาที่มีความแปรปรวนไม่คงที่ สามารถแปลงอนุกรมเวลาให้มีความแปรปรวนคงที่มากขึ้นด้วยการใช้ลอการิทึมธรรมชาติ (natural logarithm) หรือรากที่สอง (square root) ของ y_t นั่นคือ $Z_t = \ln(y_t)$ หรือ $Z_t = \sqrt{y_t}$ [1]

หลังจากที่อนุกรมเวลาคงที่แล้วจะกำหนดตัวแบบให้กับอนุกรมเวลาเพื่อนำไปใช้ในการพยากรณ์ต่อไป ซึ่งมีขั้นตอนดังนี้ [1]

(1) กำหนดตัวแบบพยากรณ์ โดย

พิจารณา ACF และ PACF ของอนุกรมเวลาที่มีลักษณะคงที่ ซึ่งตัวแบบที่กำหนดให้ คือ ตัวแบบ ARIMA(p,d,q) ดังสมการที่ (2.1) [9]

$$\phi_p(B)(1-B)^d z_t = \delta + \theta_q(B)\varepsilon_t \quad (2.1)$$

โดยที่ Z_t แทนอนุกรมเวลาที่มีลักษณะคงที่; $\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ แทนการถดถอยในตัวเองแบบไม่มีฤดูกาลอันดับ p [non-seasonal autoregressive order p, AR(p)]; $(1-B)^d$ แทนอันดับของผลต่างแบบไม่มีฤดูกาล (non-seasonal difference) โดยที่ d คือ จำนวนครั้งของผลต่างแบบไม่มีฤดูกาล; $\delta = \phi_p(B)\mu$ แทนค่าคงที่ โดยที่ μ แทนค่าเฉลี่ยของอนุกรมเวลาที่มีลักษณะคงที่; $\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$ แทนค่าเฉลี่ยเคลื่อนที่แบบไม่มีฤดูกาลอันดับ q (non-seasonal moving average order q, MA(q)); B แทนตัวดำเนินการถอยหลัง (backward shift operator) โดยที่ $B^k y_t = y_{t-k}$ และ ε_t แทนค่าคลาดเคลื่อนสุ่มที่มีการแจกแจงปกติ มีค่าเฉลี่ยเท่ากับ 0 และค่าความแปรปรวนคงที่

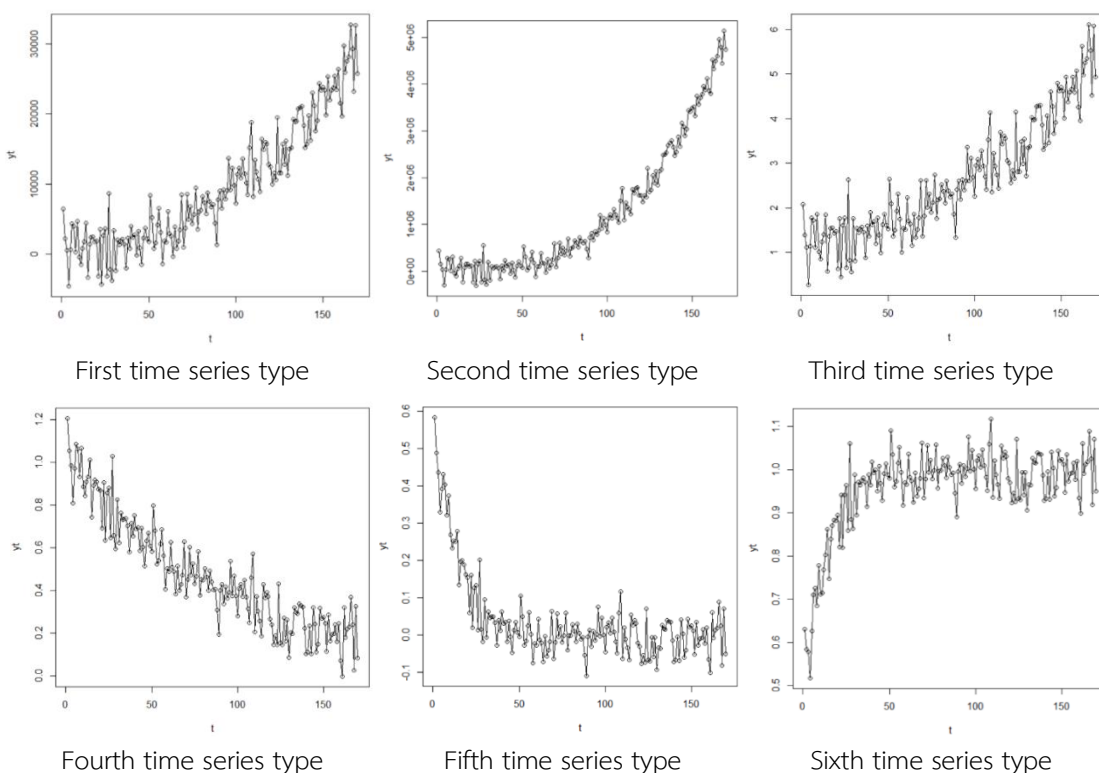


Figure 1 Six time series types plots

(2) ประมาณค่าพารามิเตอร์ของตัวแบบพยากรณ์ด้วยวิธีกำลังสองน้อยสุด และตรวจสอบความมีนัยสำคัญของพารามิเตอร์ที่ประมาณได้ด้วยการทดสอบ t-test

(3) ตรวจสอบความถูกต้องของตัวแบบพยากรณ์ ได้แก่ การตรวจสอบความเป็นอิสระกันของค่าความคลาดเคลื่อนของการพยากรณ์โดยพิจารณาจากการทดสอบสหสัมพันธ์ในตนเองของ Box และ Ljung และการตรวจสอบการแจกแจงปกติของค่าความคลาดเคลื่อนของการพยากรณ์ว่ามีค่าเฉลี่ยเท่ากับศูนย์และมีความแปรปรวนคงที่หรือไม่ โดยพิจารณาจากแผนภาพความน่าจะเป็นแบบปกติ (normal probability plot) ฮิสโทแกรมและแผนภาพการกระจายระหว่างค่าพยากรณ์ (fitted value) กับค่าความคลาดเคลื่อนของการพยากรณ์ หากตัวแบบที่ได้ไม่ผ่านการตรวจสอบความถูกต้องจะย้อนกลับไปที่ขั้นตอนที่ (1) ใหม่จนกระทั่งได้ตัวแบบพยากรณ์ที่เหมาะสม [9]

(4) นำตัวแบบที่ผ่านการตรวจสอบไปใช้ในการพยากรณ์ ซึ่งผู้วิจัยจะพยากรณ์ทั้งหมด 20 ค่าแล้วนำไปเปรียบเทียบกับข้อมูลในส่วนที่สอง โดยหาค่า MSE และ MAPE เพื่อใช้ในการเปรียบเทียบวิธีการพยากรณ์ต่อไป

2.2.2 วิธี empirical mode decomposition ด้วยตัวแบบ ARIMA (EMD-ARIMA) หลักการของวิธีการนี้ คือ แยกส่วนประกอบของอนุกรมเวลา y_t ออกมาเป็นฟังก์ชัน intrinsic mode functions (IMFs) จำนวนหนึ่งและค่าเศษเหลือ (residue) จากนั้นใช้วิธีบ็อกซ์และเจนกินส์พยากรณ์ IMFs และ residue แล้วนำค่าพยากรณ์ IMFs และ residue ที่ได้จากวิธีบ็อกซ์และเจนกินส์มารวมกันจะได้เป็นค่าพยากรณ์ $\hat{y}_t$ ซึ่งมีขั้นตอนดังนี้ [3]

(1) คำนวณค่า $d_t = y_t - m_t$ โดยที่ $m_t = (e_{\max(t)} + e_{\min(t)}) / 2$, $e_{\max(t)}$ แทน upper

envelope และ $e_{\min(t)}$ แทน lower envelope

(2) ตรวจสอบว่าค่า d_t ที่ได้ มีเงื่อนไขของการเป็น IMF ครบทั้ง 2 ข้อ หรือไม่ ซึ่งเงื่อนไข คือ d_t จะต้องมีความถี่ของค่า local extrema และจำนวนของ zero-crossing เท่ากันหรือต่างกันไม่เกิน 1 และที่จุดใด ๆ ก็ตามของ d_t จะต้องมีความถี่ของ envelope ที่ได้จากค่า local maxima กับ local minima เท่ากับ 0 ถ้า d_t มีเงื่อนไขครบ จะให้ d_t คือ IMF_t และหาค่า residue (r_t) ได้จาก $r_t = y_t - d_t$ แต่ในกรณีที่ d_t มีเงื่อนไขไม่ครบ จะถือว่า d_t ไม่เป็น IMF แล้วกำหนดให้ y_t เท่ากับ d_t แล้วย้อนกลับไปทำขั้นตอนที่ (1)

(3) ทำตามขั้นตอนที่ (1) และ (2) จนค่า r_t จะเป็นค่าคงที่ หรือ IMF_t ที่ได้มีค่าสูงสุดต่ำสุดเพียงค่าเดียวและไม่สามารถหา IMF ได้อีก

(4) เมื่อทำครบทุกขั้นตอนแล้วจะสามารถเขียนสมการของอนุกรมเวลา y_t ในรูปของผลบวกของฟังก์ชัน IMF และค่า residue ได้ดังสมการ

$$y_t = \sum_{k=1}^N IMF_{kt} + r_t \quad (2.2)$$

โดยที่ IMF_{kt} แทน IMF ลำดับที่ k ณ เวลา t และ N แทนจำนวนของฟังก์ชัน IMF

(5) กำหนดตัวแบบ ARIMA ในการพยากรณ์ให้ค่า IMF_{kt} และค่า r_t เพื่อหาค่าพยากรณ์ ณ เวลา t โดยที่ $t = 1-170$ ของฟังก์ชัน IMF_{kt} ทั้งหมด ($\hat{IMF}_{kt}$) และค่าพยากรณ์ ณ เวลา t ของค่า residue ($\hat{r}_t$) หลังจากนั้นนำค่า $\hat{IMF}_{kt}$ และค่า $\hat{r}_t$ มารวมกันจะได้เป็นค่าพยากรณ์อนุกรมเวลาที่เวลา t ตั้งแต่ 1-170 หรือ $\hat{y}_t$ ดังสมการ

$$\hat{y}_t = \sum_{k=1}^N \hat{IMF}_{kt} + \hat{r}_t \quad (2.3)$$

โดยสามารถสรุปขั้นตอนของวิธี EMD-ARIMA เป็นแผนภาพดังรูปที่ 2

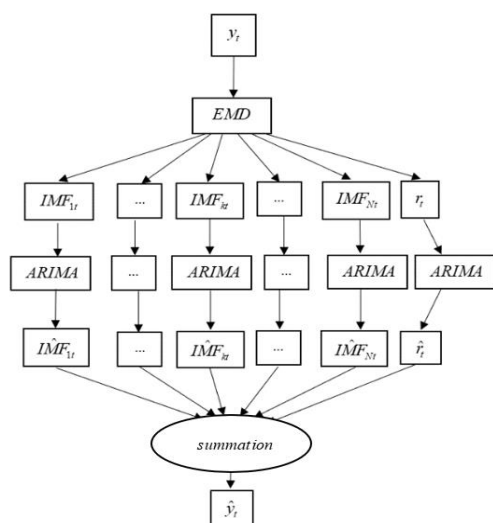


Figure 2 Diagram of EMD- ARIMA method procedures [5]

2.2.3 วิธี ensemble empirical mode decomposition ด้วยตัวแบบ ARIMA (EEMD-ARIMA) หลักการของวิธีการนี้ คือ แยกส่วนประกอบของอนุกรมเวลา y_t ออกมาเป็นฟังก์ชัน IMF จำนวนหนึ่งและค่า residue คล้ายกับวิธีการ EMD แต่วิธีการนี้ จะมีการใช้ gaussian white noise เข้ามาช่วยแก้ปัญหา mode mixing ซึ่งเป็นปัญหาที่เกิดขึ้นในวิธีการ EMD ซึ่งหลักการที่เลือกจะมีความคล้ายกับวิธี EMD-ARIMA โดยมีขั้นตอนดังนี้ [11]

(1) สร้างอนุกรมเวลา y_{it} จากสมการดังต่อไปนี้ $y_{it} = y_t + \varepsilon_{it}; i = 1, 2, \dots, e$ โดยที่ e แทนจำนวนครั้งทั้งหมดในการบวกอนุกรมเวลาด้วย ε_{it} ซึ่งในงานวิจัยนี้กำหนด $e = 250$ และ ε_{it} แทน gaussian white noise ณ เวลา t ในการบวกครั้งที่ i ซึ่งมีการแจกแจงปกติ มีค่าเฉลี่ยเท่ากับ 0 และมีความแปรปรวนเท่ากับ 0.2 นั่นคือ $\varepsilon_{it} \sim N(0, 0.2)$

(2) แยกส่วนประกอบของอนุกรมเวลาใหม่ (y_{it}) ออกมาเป็น intrinsic mode functions และค่า residue ด้วยวิธีการ EMD

(3) ทำซ้ำขั้นตอน (1) กับ (2) แต่เปลี่ยน

ชุดของ gaussian white noise

(4) หลังจากที่ทำขั้นตอน (3) จะได้ฟังก์ชัน IMF ทั้งหมด 250 ชุด และค่า residue ทั้งหมด 250 ชุด แล้วคำนวณค่าเฉลี่ยดังสมการ

$$IMF_{kt} = \frac{\sum_{i=1}^e IMF_{kit}}{e} \quad (2.4)$$

$$r_t = \frac{\sum_{i=1}^e r_{it}}{e} \quad (2.5)$$

โดยที่ IMF_{kit} คือ ค่าของ IMF ลำดับที่ k ชุดที่ i ณ เวลา t และ r_{it} คือ ค่า residue ชุดที่ i ณ เวลา t

สุดท้ายจะได้ผลลัพธ์เป็นฟังก์ชัน IMF 1 ชุด และค่า residue 1 ชุด ซึ่งสามารถเขียนเป็นสมการของอนุกรมเวลา y_t ในรูปของผลบวกของฟังก์ชัน IMF และค่า residue ได้ดังสมการ (2.2)

(5) กำหนดตัวแบบ ARIMA ในการพยากรณ์ให้แก่ค่า IMF_{kt} และค่า r_t เพื่อทำการพยากรณ์ ณ เวลา t ของฟังก์ชัน IMF_{kt} ทั้งหมด ($\hat{IMF}_{kt}$) และค่าพยากรณ์ ณ เวลา t ของค่า residue ($\hat{r}_t$) แล้วนำค่า $\hat{IMF}_{kt}$ และค่า $\hat{r}_t$ มารวมกันจะได้เป็นค่าพยากรณ์อนุกรมเวลาที่เวลา t หรือ $\hat{y}_t$ ดังสมการ (2.3)

โดยสามารถสรุปขั้นตอนของวิธี EEMD-ARIMA เป็นแผนภาพได้ดังรูปที่ 3

2.3 เกณฑ์การตรวจสอบประสิทธิภาพในการพยากรณ์

งานวิจัยนี้ใช้เกณฑ์ตรวจสอบประสิทธิภาพ 2 เกณฑ์ ได้แก่ ค่า MSE และ MAPE ซึ่งมีสูตรในการคำนวณดังนี้ [10]

$$MSE = \frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{n} \quad (2.6)$$

$$MAPE = \frac{100}{n} \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right| \quad (2.7)$$

โดยที่ n แทนจำนวนข้อมูลของอนุกรมเวลา

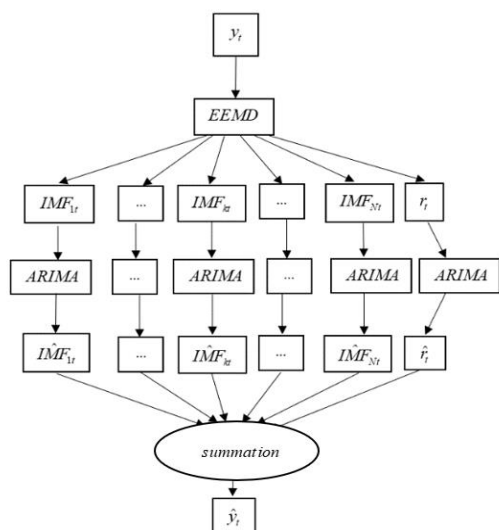


Figure 3 Diagram of EEMD-ARIMA method procedures [5]

3. ผลการวิจัย

อนุกรมเวลา 6 รูปแบบ ได้แก่ อนุกรมเวลาที่มีแนวโน้มโพลีโนเมียล 2 รูปแบบ อนุกรมเวลาที่มีแนวโน้มเอ็กซ์โพเนนเชียล 2 รูปแบบ และอนุกรมเวลาที่มีแนวโน้มโลจิสติก 2 รูปแบบ และกำหนดตัวแบบที่ใช้ในการพยากรณ์อนุกรมเวลาด้วยวิธีการพยากรณ์ 3 วิธี ได้แก่ วิธีบ็อกซ์และเจนกินส์ด้วยตัวแบบ ARIMA วิธี EMD-ARIMA และวิธี EEMD-ARIMA ผลของค่า MSE และ MAPE แสดงดังตารางที่ 1

ตารางที่ 1 เมื่อพิจารณาจากค่า MSE พบว่าสำหรับการพยากรณ์อนุกรมเวลารูปแบบที่ 1, 2 และ 4 ซึ่งเป็นอนุกรมเวลาที่มีแนวโน้มโพลีโนเมียลอันดับสองและสาม และเอ็กซ์โพเนนเชียลที่มีเลขชี้กำลังลบ วิธี EEMD-ARIMA ให้ค่า MSE ต่ำที่สุดเท่ากับ 8.8×10^6 , 4.7×10^{10} และ 0.0099 ตามลำดับ ขณะที่การพยากรณ์อนุกรมเวลารูปแบบที่ 3 ซึ่งเป็นอนุกรมเวลาที่มีแนวโน้มเอ็กซ์โพเนนเชียลที่มีเลขชี้กำลังบวก วิธีบ็อกซ์และเจนกินส์ด้วยตัวแบบ ARIMA และวิธี EMD-ARIMA

ให้ค่า MSE ต่ำที่สุดและมีค่าใกล้เคียงกันซึ่งเท่ากับ 0.3392 และ 0.3459 ตามลำดับ และสำหรับการพยากรณ์อนุกรมเวลารูปแบบที่ 5 และ 6 ซึ่งเป็นอนุกรมเวลาที่มีแนวโน้มโลจิสติกที่มีเลขชี้กำลังบวกและลบ ตามลำดับ วิธีบ็อกซ์และเจนกินส์ด้วยตัวแบบ ARIMA และวิธี EEMD-ARIMA ให้ค่า MSE ต่ำที่สุดและมีค่าใกล้เคียงกันซึ่งเท่ากับ 0.0024 และ 0.0026 ตามลำดับ สำหรับการพยากรณ์อนุกรมเวลารูปแบบที่ 5 และ MSE เท่ากับ 0.0031 และ 0.0028 ตามลำดับ สำหรับการพยากรณ์อนุกรมเวลารูปแบบที่ 6 และเมื่อพิจารณาจากค่า MAPE พบว่าสำหรับการพยากรณ์อนุกรมเวลารูปแบบที่ 1 และรูปแบบที่ 2 ซึ่งเป็นอนุกรมเวลาที่มีแนวโน้มโพลีโนเมียลอันดับสองและสาม วิธี EEMD-ARIMA ให้ค่า MAPE ต่ำที่สุดเท่ากับ 9.6223 และ 3.9734 ตามลำดับ ในขณะที่การพยากรณ์อนุกรมเวลารูปแบบที่ 3 ซึ่งเป็นอนุกรมเวลาที่มีแนวโน้มเอ็กซ์โพเนนเชียลที่มีเลขชี้กำลังบวก วิธีบ็อกซ์และเจนกินส์ด้วยตัวแบบ ARIMA และวิธี EMD-ARIMA ให้ค่า MAPE ต่ำที่สุดและมีค่าใกล้เคียงกันซึ่งเท่ากับ 9.0091 และ 9.8261 ตามลำดับ สำหรับการพยากรณ์อนุกรมเวลารูปแบบที่ 4 ซึ่งเป็นอนุกรมเวลาที่มีแนวโน้มเอ็กซ์โพเนนเชียลที่มีเลขชี้กำลังลบ วิธี EMD-ARIMA ให้ค่า MAPE ต่ำที่สุดซึ่งเท่ากับ 356.3533 และสำหรับการพยากรณ์อนุกรมเวลารูปแบบที่ 5 และรูปแบบที่ 6 ซึ่งเป็นอนุกรมเวลาที่มีแนวโน้มโลจิสติกที่มีเลขชี้กำลังบวกและลบ ตามลำดับ วิธีบ็อกซ์และเจนกินส์ด้วยตัวแบบ ARIMA และวิธี EEMD-ARIMA ให้ค่า MAPE ต่ำที่สุดและมีค่าใกล้เคียงกันซึ่งเท่ากับ 100.6131 และ 100.2875 ตามลำดับ สำหรับการพยากรณ์อนุกรมเวลารูปแบบที่ 5 และเท่ากับ 4.4402 และ 4.3056 ตามลำดับ สำหรับการพยากรณ์อนุกรมเวลารูปแบบที่ 6

Table 1 MSE and MAPE of ARIMA, EMD-ARIMA and EEMD-ARIMA methods for six time series types

Time series type	ARIMA method for y_t		ARIMA method for $Z_t = \sqrt{y_t}$		EMD-ARIMA method for y_t		EEMD-ARIMA method for y_t	
	MSE	MAPE	MSE	MAPE	MSE	MAPE	MSE	MAPE
First type	1.0×10^7	10.0101	2.6×10^7	14.5330	1.3×10^7	11.7029	$8.8 \times 10^{6*}$	9.6223**
Second type	9.3×10^{10}	5.2262	1.0×10^{12}	20.2491	1.1×10^{11}	5.9245	$4.7 \times 10^{10*}$	3.9734**
Third type	0.4356	9.7783	0.3392	9.0091**	0.3459	9.8261	0.8935	16.8173
Fourth type	0.0141	1,702.5667	0.0139	1,690.9879	0.0331	356.3533**	0.0099*	1,254.9433
Fifth type	0.0030	135.2633	0.0024*	100.6131	0.0061	282.5146	0.0026	100.2875**
Sixth type	0.0035	4.8702	0.0031	4.4402	0.0514	18.1653	0.0028*	4.3056**

*Means the least MSE for each time series type; **Means the least MAPE for each time series type

ดังนั้นพบว่าวิธี EEMD-ARIMA มีประสิทธิภาพในการพยากรณ์สูงที่สุดสำหรับอนุกรมเวลาที่มีแนวโน้มโพลีโนเมียล และมีประสิทธิภาพในการพยากรณ์อนุกรมเวลาที่มีแนวโน้มโลจิสติกใกล้เคียงกับวิธีบ็อกซ์และเจนกินส์ด้วยตัวแบบ ARIMA และเมื่อพิจารณาเฉพาะค่า MSE เพียงอย่างเดียว วิธี EEMD-ARIMA จะมีประสิทธิภาพสูงที่สุดในการพยากรณ์อนุกรมเวลาที่มีแนวโน้มเอ็กซ์โปเนนเชียลที่มีเลขชี้กำลังลบ ในขณะที่วิธี EMD-ARIMA มีประสิทธิภาพในการพยากรณ์อนุกรมเวลาที่มีแนวโน้มเอ็กซ์โปเนนเชียลที่มีเลขชี้กำลังบวกใกล้เคียงกับวิธีบ็อกซ์และเจนกินส์ด้วยตัวแบบ ARIMA และเมื่อพิจารณาเฉพาะค่า MAPE เพียงอย่างเดียว วิธี EMD-ARIMA จะมีประสิทธิภาพสูงที่สุดในการพยากรณ์อนุกรมเวลาที่มีแนวโน้มเอ็กซ์โปเนนเชียลที่มีเลขชี้กำลังลบ

4. สรุปผล อภิปรายผล และข้อเสนอแนะ

งานวิจัยนี้ได้ศึกษาวิธีการพยากรณ์ 3 วิธี คือ วิธีบ็อกซ์และเจนกินส์ด้วยตัวแบบ ARIMA วิธี EMD-ARIMA และวิธี EEMD-ARIMA ในการพยากรณ์อนุกรมเวลาที่มีลักษณะไม่เป็นเชิงเส้นและไม่คงที่ 6 รูปแบบ ที่จำลองขึ้นโดยใช้โปรแกรม R ผลการวิจัยพบว่าวิธี

EEMD-ARIMA มีประสิทธิภาพสูงที่สุดในการพยากรณ์ในการพยากรณ์อนุกรมเวลาที่มีแนวโน้มโพลีโนเมียล และมีประสิทธิภาพในการพยากรณ์ใกล้เคียงกับวิธีบ็อกซ์และเจนกินส์ด้วยตัวแบบ ARIMA ในการพยากรณ์อนุกรมเวลาที่มีแนวโน้มโลจิสติก ขณะที่วิธี EMD-ARIMA มีประสิทธิภาพในการพยากรณ์ใกล้เคียงกับวิธีบ็อกซ์และเจนกินส์ด้วยตัวแบบ ARIMA ในการพยากรณ์อนุกรมเวลาที่มีแนวโน้มเอ็กซ์โปเนนเชียลที่มีเลขชี้กำลังบวก ส่วนวิธี EMD-ARIMA และวิธี EEMD-ARIMA มีประสิทธิภาพในการพยากรณ์อนุกรมเวลาที่มีแนวโน้มเอ็กซ์โปเนนเชียลที่มีเลขชี้กำลังลบที่แตกต่างกัน โดยเมื่อพิจารณาเฉพาะค่า MSE วิธี EEMD-ARIMA จะมีประสิทธิภาพในการพยากรณ์อนุกรมเวลาที่มีแนวโน้มเอ็กซ์โปเนนเชียลที่มีเลขชี้กำลังลบได้ดีที่สุด แต่เมื่อพิจารณาเฉพาะค่า MAPE วิธี EMD-ARIMA จะมีประสิทธิภาพในการพยากรณ์อนุกรมเวลาที่มีแนวโน้มเอ็กซ์โปเนนเชียลที่มีเลขชี้กำลังลบได้ดีที่สุด

ยิ่งกว่านั้น ผลการวิจัยพบว่าวิธี EEMD-ARIMA มีประสิทธิภาพสูงที่สุดในการพยากรณ์อนุกรมเวลาที่มีแนวโน้มโพลีโนเมียล และเมื่อพิจารณาจากค่า MSE วิธี EEMD-ARIMA ยังมีประสิทธิภาพสูงที่สุดในการพยากรณ์อนุกรมเวลาที่มีแนวโน้มเอ็กซ์โปเนนเชียลที่มี

เลขชี้กำลังลบ ซึ่งผลที่ได้มีความสอดคล้องกับงานวิจัยของ Min และคณะ [5] ขณะที่เมื่อพิจารณาจากค่า MAPE วิธี EMD-ARIMA จะมีประสิทธิภาพในการพยากรณ์สูงที่สุดในการพยากรณ์อนุกรมเวลาที่มีแนวโน้มเอ็กซ์โปเนนเชียลที่มีเลขชี้กำลังลบ ซึ่งผลที่ได้มีความสอดคล้องกับงานวิจัยของ Abadan และ Shabri [3] กับ Zhou และ Huang [4]

ข้อสังเกตจากวิธี EMD-ARIMA และ EEMD-ARIMA คือ วิธีการพยากรณ์ทั้งสองวิธีอาจให้ค่า MSE และ MAPE ที่มากผิดปกติได้ เช่น ในการพยากรณ์อนุกรมเวลาที่มีแนวโน้มเอ็กซ์โปเนนเชียลที่มีเลขชี้กำลังบวก และในการพยากรณ์อนุกรมเวลาที่มีแนวโน้มโลจิสติกที่มีเลขชี้กำลังลบ ซึ่งเกิดขึ้นได้อาจเนื่องมาจากวิธีทั้งสองมีการพยากรณ์ IMF หลายชุด ทำให้ความคลาดเคลื่อนของการพยากรณ์สะสมเพิ่มขึ้น ส่งผลให้ค่า MSE และ MAPE มากผิดปกติ

งานวิจัยครั้งต่อไปที่มีการใช้วิธี EMD-ARIMA และ EEMD-ARIMA ขอเสนอแนะให้จำลองอนุกรมเวลาที่มีส่วนประกอบของอนุกรมเวลาที่ประกอบด้วยฤดูกาล วัฏจักร หรือเหตุการณ์ที่ผิดปกติ หรือจำลองอนุกรมเวลาที่มีแนวโน้มอื่นนอกเหนือจากแนวโน้มโพลิโนเมียล เอ็กซ์โปเนนเชียล และโลจิสติก เช่น อนุกรมเวลาที่มีแนวโน้มแบบตรีโกณมิติ และนอกจากการแยกส่วนประกอบของอนุกรมเวลาด้วยวิธี EMD และ EEMD แล้ว ยังมีวิธีอื่นที่สามารถแยกส่วนประกอบของอนุกรมเวลาได้ เช่น วิธี complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) และยังมีวิธีการแปลงอนุกรมเวลารูปแบบอื่นที่สามารถใช้ร่วมกับวิธีบ็อกซ์และเจนกินส์ด้วยตัวแบบ ARIMA เช่น Fourier transformation และ Wavelet transformation

5. References

- [1] Taesombut, S., 2006, Quantitative Forecasting, Kasetsart Publisher, Bangkok. (in Thai)
- [2] Bao, Y. , Yi, D. and Chen, J. , 2012, Forecasting software reliability using Arima with ensemble empirical mode decomposition, Int. J. Model. Simul. 32: 104-110.
- [3] Abadan, S. and Shabri, A. , 2014, Hybrid empirical mode decomposition-ARIMA for forecasting price of rice, Appl. Math. Sci. 8: 3133-3143.
- [4] Zhou, Y. and Huang, M., 2016, Lithium-ion batteries remaining useful life prediction based on a mixture of empirical mode decomposition and ARIMA model, Microelect. Reliab. 65: 265-273.
- [5] Min, Y., Bin, Y., Liang-li, Z. and Xi, C., 2015, Wind Speed Forecasting Based on EEMD and ARIMA, Chinese Automation Congress, Wuhan.
- [6] Polynomial, Available Source: <https://en.wikipedia.org/wiki/Polynomial>, June 5, 2018.
- [7] Exponential Function, Available Source: https://en.wikipedia.org/wiki/Exponential_function, June 5, 2018.
- [8] Logistic Function, Available Source: https://en.wikipedia.org/wiki/Logistic_function, June 5, 2018.
- [9] Choopradit, B. and Chaipitak, S. , 2018, Forecasting model for mango export volumes of Thailand, Naresuan Univ. J. Sci. Technol. 26(2): 74-85. (in Thai)

-
- [10] Mantmin, M. , 2006, Time Series Forecasting, Praguypreak Publisher, Bangkok. (in Thai)
- [11] Huang, N. and Wu, Z., 2009, Ensemble empirical mode decomposition: A Noise assisted data analysis method, Adv. Adapt. Data Anal. 1: 1-12.