

การพยากรณ์ปริมาณการปล่อยคาร์บอนไดออกไซด์จากกลุ่มพลังงานในประเทศไทย ด้วยตัวแบบ SARIMA-ANN-REG

PREDICTION OF CARBON DIOXIDE EMISSION FROM ENERGY CONSUMPTION IN THAILAND WITH SARIMA-ANN-REG MODEL

สุธีมนต์ กาสีวงศ์, กัลยา บุญหล้า*

*Suteemon Karseewong, Kanlaya Boonlha**

สาขาวิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยนเรศวร

Department of Mathematics, Faculty of Science, Naresuan University.

*Corresponding author, e-mail: kanlayab@nu.ac.th

Received: 14 November 2023; **Revised:** 26 March 2024; **Accepted:** 2 April 2024

บทคัดย่อ

การศึกษานี้มีจุดประสงค์เพื่อศึกษาการพยากรณ์การปล่อยก๊าซ CO₂ ในประเทศไทยโดยใช้ข้อมูลจากกระทรวงพลังงาน ด้วยตัวแบบผสม SARIMA-ANN-REG โดยใช้สมการการถดถอยเชิงเส้นตรงในการสร้างตัวแบบพยากรณ์ผสมระหว่างตัวแบบ SARIMA กับตัวแบบโครงข่ายประสาทเทียม (ANN) ข้อมูลการปล่อยปริมาณก๊าซ CO₂ รายเดือนตั้งแต่เดือนมกราคม พ.ศ. 2558 ถึงธันวาคม พ.ศ. 2565 (96 เดือน) ใช้เป็นชุดข้อมูลสำหรับการเรียนรู้ ในขณะที่ข้อมูลตั้งแต่เดือนมกราคม ถึงกรกฎาคม พ.ศ. 2566 (7 เดือน) ถูกนำมาใช้ในพยากรณ์ พิจารณาเลือกตัวแบบด้วยรากของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) ที่ต่ำที่สุด ผลการศึกษาพบว่า ตัวแบบ SARIMA-ANN-REG เมื่อ SARIMA(1,0,2)(1,1,1)₁₂ และกำหนดจำนวนโหนดในชั้นซ่อนเท่ากับ 8 ในตัวแบบ ANN ทำให้ค่า RMSE ต่ำที่สุดเท่ากับ 0.4992 สำหรับชุดข้อมูลเรียนรู้ เมื่อนำไปพยากรณ์ข้อมูลค่า RMSE เท่ากับ 1.2106 โดยมีสมการพยากรณ์ คือ

$$\hat{Y}_t = 1.0003\hat{F}_t^{SARIMA} + 1.0379\hat{F}_t^{ANN}$$

เมื่อ $\hat{Y}_t$, $\hat{F}_t^{SARIMA}$, $\hat{F}_t^{ANN}$ คือ ค่าพยากรณ์ ค่าพยากรณ์ด้วยตัวแบบ SARIMA และค่าพยากรณ์ด้วยตัวแบบ ANN ณ เวลา t ตามลำดับ

คำสำคัญ: การพยากรณ์; คาร์บอนไดออกไซด์; วิธีบ็อกซ์-เจนกินส์; โครงข่ายประสาทเทียม; วิธีตัวแบบผสม

Abstract

This study aims to forecast monthly CO₂ emissions in Thailand using data from the Ministry of Energy using the model SARIMA-ANN-REG. Using linear regression model combined both SARIMA and Artificial Neural Network (ANN) approaches. Monthly CO₂ emission data from January 2015 to December 2022 (96 months) as the training dataset, while data from January to July 2023 (7 months) was used for forecasting. The model selection was based on minimizing the root mean square error (RMSE). The result found that the SARIMA-ANN-REG model with SARIMA(1,0,2)(1,1,1)₁₂ and the 8 hidden nodes in the ANN component, achieving an RMSE of 0.4992 on the training set. Applied to the forecasting period, the model yielded an RMSE of 1.2106 with the forecast model is

$$\hat{Y}_t = 1.0003\hat{F}_t^{SARIMA} + 1.0379\hat{F}_t^{ANN}$$

where $\hat{Y}_t$, $\hat{F}_t^{SARIMA}$, $\hat{F}_t^{ANN}$ are the forecast value, forecast value using SARIMA mode and forecast value with ANN model at time, respectively.

Keywords: Forecasting; Carbon Dioxide; Box-Jenkins; Artificial Neural Networks; Hybrid Model

บทนำ

ปัญหาภาวะโลกร้อนเป็นปรากฏการณ์ที่โลกมีอุณหภูมิเฉลี่ยสูงขึ้น เนื่องจากการรวมตัวกันอย่างหนาแน่นของก๊าซเรือนกระจกที่ดูดความร้อนไว้ไม่ให้สะท้อนออกไป ก๊าซเรือนกระจก เช่น คาร์บอนไดออกไซด์ มีเทน ไนตรัส ออกไซด์ ก๊าซไฮโดรฟลูออโรคาร์บอน ก๊าซเพอร์ฟลูออโรคาร์บอน ก๊าซซัลเฟอร์เฮกซะฟลูออไรด์ ก๊าซไนโตรเจน ไตรฟลูออไรด์ เป็นต้น เมื่อลอยสู่บรรยากาศจะดูดซับความร้อนไว้หากปริมาณมากเกินไป จะส่งผลให้ชั้นบรรยากาศมีการกักเก็บรังสีความร้อนไว้มากขึ้น ทำให้อุณหภูมิเฉลี่ยของชั้นบรรยากาศเพิ่มขึ้น และทำลายชั้นโอโซนที่ทำหน้าที่กรองรังสียูวีให้บางลง จากข้อมูลของกระทรวงพลังงาน พบว่าก๊าซคาร์บอนไดออกไซด์ (CO₂) เป็นก๊าซเรือนกระจกที่สะสมในชั้นบรรยากาศมากที่สุด [1] หากปริมาณเกินความต้องการจะผลให้อุณหภูมิเฉลี่ยในโลกสูงขึ้น และเป็นสาเหตุของการเกิดภาวะโลกร้อน ซึ่งก๊าซ CO₂ ที่ถูกปล่อยสู่ชั้นบรรยากาศโลกส่วนใหญ่เกิดจากการปล่อยออกมาจากการเผาไหม้เชื้อเพลิง ปัจจุบันประเทศไทยมีอุตสาหกรรมเพิ่มมากยิ่งขึ้น มีการจัดตั้งโรงงานอุตสาหกรรม มีจำนวนรถที่เพิ่มมากขึ้นจึงมีการใช้พลังงานเพิ่มมากขึ้น ก่อให้เกิดการเผาไหม้ การปล่อยก๊าซ CO₂ จากโรงงานอุตสาหกรรมในปริมาณมหาศาลมาก ส่งผลกระทบให้เกิดความแปรปรวนทางสภาพอากาศ ในปี พ.ศ. 2565 พบว่าประเทศไทยได้ปล่อยก๊าซคาร์บอนไดออกไซด์อยู่ที่ระดับ 247.7 ล้านตัน เพิ่มขึ้นร้อยละ 1.5 เมื่อเทียบกับปี พ.ศ. 2564 โดยภาคการขนส่งมีส่วนการปล่อยก๊าซ CO₂ เพิ่มขึ้นมากที่สุด [2] หากทราบแนวโน้มของปริมาณการปล่อยก๊าซ CO₂ ได้ล่วงหน้า จะทำให้หน่วยงานที่เกี่ยวข้องวางแผนเพื่อลดปริมาณปล่อยก๊าซ CO₂ หรือหามาตรการณรงค์ลดการปล่อยก๊าซ CO₂ ในอนาคตได้และนำไปสู่การพัฒนาด้านพลังงานของประเทศต่อไปในอนาคต

การพยากรณ์ตัวแบบอนุกรมเวลา (Time Series Model) เป็นการจำลองทางสถิติอย่างหนึ่งที่ใช้ข้อมูลในอดีตของตัวแปรที่ต้องการพยากรณ์ เพื่อพยากรณ์ค่าของตัวแปรในอนาคต วิธีบอกซ์-เจนกินส์ (Box-Jenkins Method) เป็นวิธีการพยากรณ์ที่นิยมใช้กันเป็นส่วนใหญ่ มีความแม่นยำในการพยากรณ์ค่อนข้างสูง เหมาะสำหรับใช้พยากรณ์ข้อมูลอนุกรมเวลาแบบคงที่เป็นฟังก์ชันเชิงเส้นตรงของค่าพารามิเตอร์โดยตรวจสอบฟังก์ชันสหสัมพันธ์ในตัวเอง และฟังก์ชันสหสัมพันธ์ในตัวเองบางส่วน ในปัจจุบันมีการพัฒนาตัวแบบที่สามารถใช้ในการศึกษาความสัมพันธ์ที่ไม่ใช่

เชิงเส้นตรงของอนุกรมเวลา เช่น ตัวแบบเครือข่ายประสาทเทียม (Artificial Neural Network: ANN) เพื่อจำลองการทำงานของเครือข่ายประสาทในสมองมนุษย์ ด้วยวัตถุประสงค์ที่จะสร้างเครื่องมือซึ่งมีความสามารถในการเรียนรู้และการจดจำรูปแบบ (Pattern Recognition) และการสร้างความรู้ใหม่ประกอบด้วย 3 ชั้นตอน คือ ชั้นนำเข้า (Input Layer) ชั้นซ่อน (Hidden Layer) และชั้นผลลัพธ์ (Output Layer) [3]

การศึกษาการพยากรณ์การปล่อยก๊าซ CO_2 มีการศึกษาหลายกรณี เช่น การทำนายการปล่อยก๊าซ CO_2 ในประเทศอินเดีย [4] ตั้งแต่ปี ค.ศ. 1960 ถึง ค.ศ. 2017 จำนวน 58 ค่า ผลการศึกษาพบว่า ตัวแบบที่เหมาะสมคือ ตัวแบบ ARIMA (2,2,0) มีค่ารากของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) น้อยที่สุด การพยากรณ์ก๊าซที่เกิดจากผลิตภัณฑ์ชีวภาพในจังหวัดสุพรรณบุรี [5] โดยใช้ข้อมูลรายเดือนตั้งแต่ปี พ.ศ. 2553 ถึงปี พ.ศ. 2560 ด้วยวิธีบอซ-เจนกินส์ วิธีปรับเรียบด้วยเลขชี้กำลังของโฮลต์ (Holt's Exponential Smoothing Method) วิธีปรับเรียบด้วยเลขชี้กำลังของบราวน์ (Brown's Exponential Smoothing Method: Brown) และวิธีปรับเรียบด้วยเลขชี้กำลังแบบแนวโน้มของแดม (Damped Trend Exponential Smoothing Method Damped) ผลการศึกษาพบว่า ตัวแบบการพยากรณ์ก๊าซ CO_2 เมื่อเปรียบเทียบค่ารากที่สองค่าคลาดเคลื่อนกำลังสองพบว่า การพยากรณ์กับตัวแบบ ARIMA (0,1,1) ให้ค่าใกล้เคียงกับค่าจริงมากที่สุด การพยากรณ์ปริมาณการปล่อยก๊าซ CO_2 ในประเทศไทยในภาคส่วนอุตสาหกรรม การขนส่งและภาคเศรษฐกิจ [6] ข้อมูลรายเดือนตั้งแต่ปี พ.ศ. 2552 ถึงปี พ.ศ. 2561 โดยวิธีการวิเคราะห์การถดถอยและวิธีโครงข่ายประสาทเทียมแบบหลายชั้น เมื่อพิจารณาค่าเบี่ยงเบนสัมบูรณ์เฉลี่ย ผลการศึกษาพบว่า วิธีโครงข่ายประสาทเทียมสามารถพยากรณ์ปริมาณก๊าซ CO_2 ได้อย่างมีประสิทธิภาพและแม่นยำกว่าวิธีการวิเคราะห์การถดถอย

อีกทั้งยังมีการประยุกต์ใช้ตัวแบบการพยากรณ์แบบผสม (Hybrid Forecasting) ซึ่งรูปแบบการพยากรณ์แบบผสม เพื่อให้การพยากรณ์มีความแม่นยำมาก เช่น การพยากรณ์ราคาขายปลีกมะนาวเบอร์ 1-2 ซึ่งเป็นข้อมูลจากกรมการค้าภายในกระทรวงพาณิชย์ [7] โดยศึกษาเปรียบเทียบความแม่นยำของตัวแบบ ARIMA ที่มีฤดูกาล ตัวแบบผสมระหว่างตัวแบบ ARIMA ที่มีฤดูกาลกับตัวแบบโครงข่ายประสาทเทียม และตัวแบบผสมระหว่างตัวแบบ ARIMA ที่มีฤดูกาลกับตัวแบบซัพพอร์ตเวกเตอร์แมชชีน ซึ่งพบว่าวิธีการผสมระหว่างตัวแบบ ARIMA ที่มีฤดูกาลกับตัวแบบโครงข่ายประสาทเทียม เป็นตัวแบบที่มีค่ารากของค่าคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุดเมื่อเปรียบเทียบกับ การพยากรณ์อื่น ๆ

ในการศึกษาวิจัยครั้งนี้ผู้วิจัยจึงมีความสนใจศึกษาการพยากรณ์ปริมาณการปล่อยก๊าซ CO_2 ในประเทศไทย จากกลุ่มพลังงานซึ่งประกอบด้วยภาคการผลิตไฟฟ้า ภาคการขนส่ง ภาคอุตสาหกรรม และภาคเศรษฐกิจอื่น ๆ ซึ่งเป็นอนุกรมเวลาที่มีทั้งการเคลื่อนไหวจากแนวโน้มและอิทธิพลของฤดูกาลด้วยวิธีบอซ-เจนกินส์ ที่มีฤดูกาลเป็นตัวแบบรวมการถดถอยในตัวกับการเฉลี่ยเคลื่อนที่เชิงฤดูกาล หรือ SARIMA ร่วมกับตัวแบบโครงข่ายประสาทเทียม (ANN) โดยใช้ตัวแบบผสมด้วยวิธีวิเคราะห์การถดถอย หรือ SARIMA-ANN-REG และทำการเปรียบเทียบประสิทธิภาพการพยากรณ์ของข้อมูลโดยใช้เกณฑ์รากของค่าคลาดเคลื่อนกำลังสองเฉลี่ย

วัตถุประสงค์ของการวิจัย

ศึกษาตัวแบบผสมระหว่างวิธีบอซ-เจนกินส์ ที่มีฤดูกาลเป็นตัวแบบรวมการถดถอยในตัวกับการเฉลี่ยเคลื่อนที่เชิงฤดูกาล (SARIMA) กับโครงข่ายประสาทเทียม (ANN) โดยวิธีวิเคราะห์การถดถอยเชิงเส้น (SARIMA-ANN-REG) สำหรับพยากรณ์ปริมาณก๊าซ CO_2 โดยใช้เกณฑ์ค่ารากของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Square Error: RMSE) ที่ต่ำที่สุด

วิธีดำเนินการวิจัย

การศึกษาครั้งนี้จะดำเนินการสร้างตัวแบบพยากรณ์ปริมาณการปล่อยก๊าซ CO₂ จากสำนักงานนโยบายและแผนพลังงาน กระทรวงพลังงาน [3] ข้อมูลรายเดือน (หน่วย : ล้านตัน CO₂) ตั้งแต่เดือนมกราคม พ.ศ. 2558 ถึงเดือนมิถุนายน พ.ศ. 2566 จำนวน 102 เดือน โดยแบ่งข้อมูลออกเป็น 2 ชุด ได้แก่ ข้อมูลเดือนมกราคม พ.ศ. 2558 ถึงเดือนธันวาคม พ.ศ. 2565 เป็นข้อมูลสำหรับการสร้างตัวแบบการพยากรณ์ และข้อมูลเดือนมกราคม ถึงเดือนกรกฎาคม พ.ศ. 2566 สำหรับการเปรียบเทียบประสิทธิภาพของตัวแบบโดยพิจารณาจากค่ารากของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) ต่ำที่สุด

วิธีตัวแบบผสม SARIMA-ANN โดยวิธีวิเคราะห์การถดถอย (Hybrid Model, SARIMA-ANN-REG)

วิธีตัวแบบผสม SARIMA-ANN เป็นการพยากรณ์ค่าโดยผสมระหว่างตัวแบบ SARIMA กับตัวแบบ ANN ด้วยการใช้วิธีการวิเคราะห์การถดถอย (Regression Analysis) เพื่อรวมค่าพยากรณ์จากตัวแบบ SARIMA และค่าพยากรณ์จากตัวแบบ ANN เข้าด้วยกัน ผู้วิจัยใช้แนวคิดของ Kulvanich, N. (2020) [7] โดยมีสมการต่อไปนี้

$$Y_t = w_1 F_t^{SARIMA} + w_2 F_t^{ANN} + \varepsilon_t$$

โดย	Y_t	แทน	ค่าสังเกตของข้อมูลอนุกรมเวลา ณ เวลา t
	F_t^{SARIMA}	แทน	ค่าพยากรณ์ด้วยตัวแบบ SARIMA
	F_t^{ANN}	แทน	ค่าพยากรณ์ด้วยตัวแบบ ANN
	w_1	แทน	ค่าถ่วงน้ำหนักข้อมูลส่วนที่เป็นฟังก์ชันเชิงเส้นตรงคำนวณได้จากการวิเคราะห์การถดถอย
	w_2	แทน	ค่าถ่วงน้ำหนักข้อมูลส่วนที่ไม่เป็นฟังก์ชันเชิงเส้นตรงคำนวณได้จากการวิเคราะห์การถดถอย
	ε_t	แทน	ค่าคลาดเคลื่อน ณ เวลา t

ผู้วิจัยทำการวิเคราะห์ตัวแบบอนุกรมเวลาแบบผสม SARIMA-ANN-REG โดยแนวคิดของ Kulvanich, N. (2020) [7] โดยมีขั้นตอนดังนี้

ขั้นตอนที่ 1 วิเคราะห์ข้อมูลอนุกรมเวลาข้อมูลเดือนมกราคม พ.ศ. 2558 ถึงเดือนธันวาคม พ.ศ. 2565 ด้วยตัวแบบบอซ-เจนกินส์ เพื่อพยากรณ์ข้อมูลส่วนที่เป็นฟังก์ชันเชิงเส้นตรงในเทอมของค่าพยากรณ์ คือ $\hat{F}_t^{SARIMA}$ โดยทำการทดสอบกระบวนการคงที่ (Stationary Process) พิจารณาจากกราฟฟังก์ชันสัมประสิทธิ์สหสัมพันธ์ในตัวเอง (ACF) และกราฟฟังก์ชันสัมประสิทธิ์สหสัมพันธ์ในตัวเองบางส่วน (PACF) กำหนดตัวแบบพยากรณ์จากกราฟ ACF และกราฟ PACF พิจารณาข้อมูลอนุกรมเวลาในหน่วยเวลาเริ่มแรกของคาบฤดูกาลที่แตกต่างไปจากศูนย์อย่างมีนัยสำคัญทางสถิติ โดยในการศึกษาครั้งนี้ผู้วิจัยกำหนดระดับนัยสำคัญทางสถิติเท่ากับ 0.05 พิจารณากำหนดตัวแบบที่ใช้สำหรับอนุกรมเวลาแล้ว ทำการประมาณค่าพารามิเตอร์ให้มีความเหมาะสมกับอนุกรมเวลา ตรวจสอบความเหมาะสมของตัวแบบโดยพิจารณาจากการประมาณค่าพารามิเตอร์ของตัวแบบ SARIMA โดยค่าพารามิเตอร์ทุกตัวจะต้องมีค่าแตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 ตรวจสอบค่าความคลาดเคลื่อนมีการแจกแจงปรกติ โดยค่าสถิติทดสอบของโคลโมโกรอฟ-สไมร์นอฟ (Kolmogorov-Smirnov Test) ตรวจสอบความแปรปรวนของความคลาดเคลื่อนโดยค่าสถิติทดสอบเลวิน (Levene test) และทดสอบภาพรวมของความคลาดเคลื่อนด้วยสถิติของ Ljung-Box โดยทุก Lag จะต้องไม่แตกต่างไปจากศูนย์ อย่างมีนัยสำคัญทาง

สถิติที่ระดับ 0.05 เมื่อได้ตัวแบบที่มีความเหมาะสมสำหรับการพยากรณ์นำตัวแบบไปพยากรณ์ข้อมูลเดือนมกราคม ถึงเดือนกรกฎาคม พ.ศ. 2566

ขั้นตอนที่ 2 วิเคราะห์ข้อมูลอนุกรมเวลาข้อมูลเดือนมกราคม พ.ศ. 2558 ถึงเดือนธันวาคม พ.ศ. 2565 ด้วยตัวแบบเครือข่ายประสาทเทียม โดยใช้ library(nnfor) เพื่อนำมาพยากรณ์ข้อมูลส่วนที่ไม่เป็นฟังก์ชันเชิงเส้นตรง ในเทอมของค่าพยากรณ์ คือ $\hat{F}_t^{ANN}$ กำหนดโหนดในชั้นนำเข้า (Input Node) แทนตัวแปร ได้แก่ ปริมาณการปล่อยก๊าซ CO₂ ในประเทศไทย จากกลุ่มพลังงานฟังก์ชันกระตุ้น (Activation Function) ภายใต้ตัวแบบโครงข่ายประสาทเทียมปรากฏอยู่ 2 ที่ คือ โหนดชั้นซ่อน (Hidden Node) ถูกกำหนดเป็นฟังก์ชันซิกมอยด์ และโหนดในชั้นผลลัพธ์ (Output Node) ถูกกำหนดเป็นฟังก์ชันเชิงเส้น ค่าน้ำหนัก (Weight) โปรแกรมจะทำการสุ่มค่าน้ำหนักเริ่มต้นให้กับโครงข่าย กำหนดจำนวนรอบการทำงานซ้ำ (Iteration) เท่ากับ 20 รอบ กำหนดจำนวนโหนดในชั้นซ่อน ตั้งแต่ 1-10 อัตราการเรียนรู้ เท่ากับ 0.01 นำข้อมูลนำเข้า (Input Data) เข้าสู่โครงข่ายแบบ Multi-Layer Perceptron [8] เมื่อมีการเรียนรู้ครบ 20 รอบ จะปรากฏค่าส่วนเหลือ จากนั้นจะทำการคำนวณ RMSE ของแต่ละตัวแบบจำแนกตามจำนวนโหนดในชั้นซ่อน และทำการเลือกตัวแบบที่เหมาะสมที่สุดสำหรับการพยากรณ์ โดยพิจารณาค่า RMSE ต่ำที่สุด

ขั้นตอนที่ 3 นำค่าพยากรณ์ด้วยตัวแบบ SARIMA ในขั้นตอนที่ 1 และค่าพยากรณ์ด้วยตัวแบบ ANN ในขั้นตอนที่ 2 มาเป็นตัวแปรต้นในสมการถดถอย และให้ค่าสังเกตข้อมูล Y_t เป็นตัวแปรตาม

ขั้นตอนที่ 4 ใช้ค่าสัมประสิทธิ์การถดถอยเป็นค่าถ่วงน้ำหนัก คือ w_1 และ w_2 จากนั้นคำนวณค่าพยากรณ์รวม (Total Forecasting) จากสมการ

$$\hat{Y}_t = w_1 \hat{F}_t^{SARIMA} + w_2 \hat{F}_t^{ANN}$$

เมื่อ $\hat{Y}_t, \hat{F}_t^{SARIMA}, \hat{F}_t^{ANN}$ คือ ค่าพยากรณ์ ค่าพยากรณ์ด้วยตัวแบบ SARIMA และค่าพยากรณ์ด้วยตัวแบบ ANN ณ เวลา t ตามลำดับ

เมื่อได้ตัวแบบที่มีความเหมาะสมสำหรับการพยากรณ์นำตัวแบบไปพยากรณ์ข้อมูลเดือนมกราคม ถึงเดือนกรกฎาคม พ.ศ. 2566

ขั้นตอนที่ 5 คำนวณค่า การเปรียบเทียบตัวแบบพยากรณ์ค่ารากของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Square Error; RMSE) ของแต่ละวิธีจาก

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

เมื่อ	Y_t	แทน	อนุกรมเวลา ณ เวลา
	$\hat{Y}_t$	แทน	พยากรณ์ข้อมูล ณ เวลา t
	n	แทน	จำนวนข้อมูลทั้งหมด

พิจารณาโดยใช้โดยเกณฑ์ค่า RMSE เป็นเกณฑ์ในการเปรียบเทียบ โดยตัวแบบที่เหมาะสมที่สุดจะให้ค่า RMSE ต่ำที่สุด

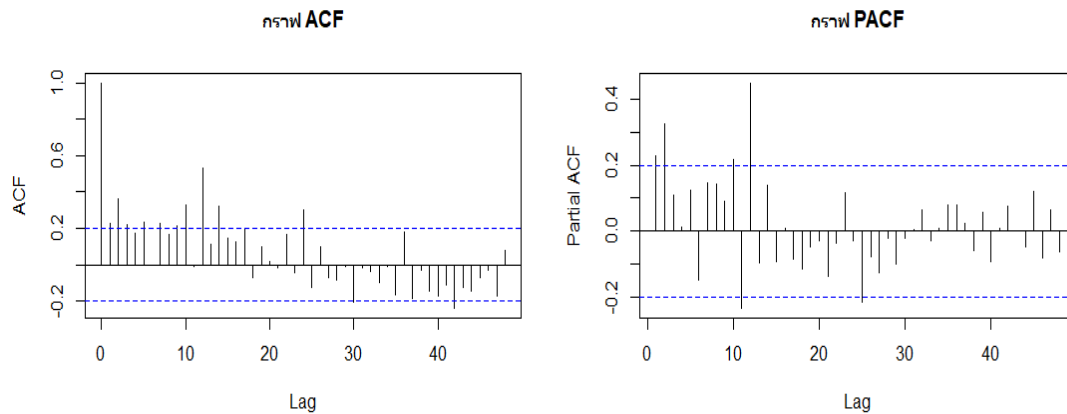
ผลการวิจัย

การพยากรณ์ข้อมูลข้อมูลปริมาณการปล่อย CO₂ ตั้งแต่เดือนมกราคมปี พ.ศ. 2558 ถึงเดือนมิถุนายนปี พ.ศ. 2566 พบว่าข้อมูลปริมาณการปล่อย CO₂ ในปี พ.ศ. 2558-2566 มีค่าใกล้เคียงกัน โดยในปี พ.ศ. 2561 มีค่าเฉลี่ยต่อเดือนสูงสุดเท่ากับ 21.707 และในปี พ.ศ. 2564 มีค่าเฉลี่ยต่อเดือนต่ำสุดเท่ากับ 20.336 ล้านตัน CO₂ เมื่อพิจารณาแนวโน้มของปริมาณการปล่อย CO₂ มีแนวโน้มลดลงในปี พ.ศ. 2564 ซึ่งเป็นช่วงการแพร่ระบาดของ Covid-19 ดังตารางที่ 1

ตารางที่ 1 ค่าต่ำสุด ค่าสูงสุด ค่าเฉลี่ย และค่าส่วนเบี่ยงเบนมาตรฐานของปริมาณการปล่อยก๊าซคาร์บอนไดออกไซด์ในประเทศไทย จากกลุ่มพลังงานตั้งแต่เดือนมกราคม พ.ศ. 2558 ถึงเดือนมิถุนายน พ.ศ. 2566

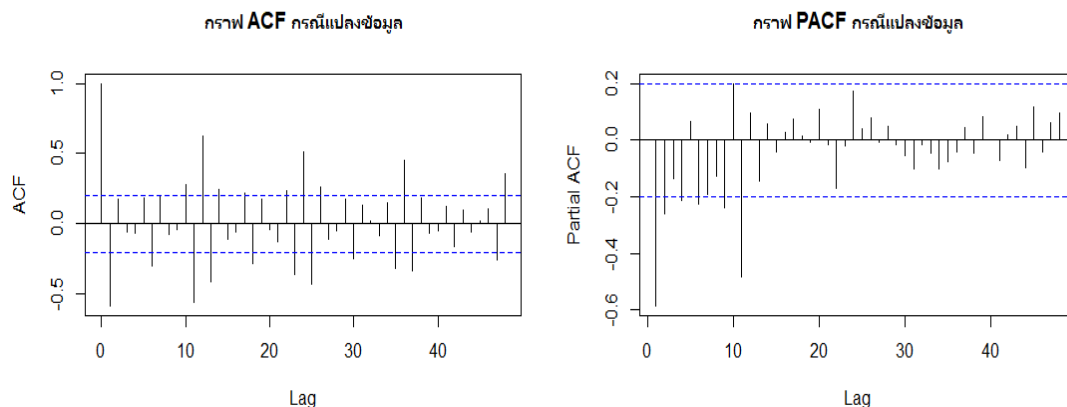
ปี	ค่าต่ำสุด	ค่าสูงสุด	$\bar{X}$	<i>S. D.</i>
2558	20.165	22.286	21.203	0.602
2559	20.979	22.776	21.558	0.639
2560	20.326	23.070	21.513	0.710
2561	20.405	23.677	21.707	0.852
2562	19.916	22.991	21.454	0.925
2563	19.093	21.755	20.649	0.682
2564	18.731	22.466	20.336	1.081
2565	19.378	21.864	20.638	0.683
2566	19.609	22.082	20.983	0.968
ภาพรวม	18.731	23.677	21.123	0.900

การพยากรณ์การวิเคราะห์ข้อมูลอนุกรมเวลาเดือนมกราคม พ.ศ. 2558 ถึงเดือนธันวาคม พ.ศ. 2565 ด้วยตัวแบบ SARIMA ผู้วิจัยตรวจสอบความคงที่ (Stationary) ของอนุกรมเวลาดังภาพที่ 1 พบว่าฟังก์ชัน ACF ในภาพลดลงอย่างช้า ๆ เมื่อ lag เพิ่มขึ้น แสดงว่าข้อมูลอนุกรมเวลาไม่คงที่ และมีส่วนประกอบของฤดูกาล และกราฟ PACF พบว่ามีบางค่าต่างจากศูนย์ แสดงว่าข้อมูลอนุกรมเวลามีความแปรปรวน (ADF Test = -3.444, p-value = 0.0518) จึงต้องหาผลต่างฤดูกาล 1 ครั้ง ได้กราฟฟังก์ชัน ACF และกราฟ PACF ใหม่ (ADF Test = -5.411, p-value = 0.01) ดังภาพที่ 2



ภาพที่ 1 ความสัมพันธ์ของ ACF และ PACF ของปริมาณการปล่อยก๊าซคาร์บอนไดออกไซด์

ก่อนการแปลงข้อมูล



ภาพที่ 2 ความสัมพันธ์ของ ACF และ PACF ของปริมาณการปล่อยก๊าซคาร์บอนไดออกไซด์

หลังการแปลงข้อมูล

จากการพิจารณารูปภาพ ACF และ PACF ของข้อมูลเดือนมกราคม พ.ศ. 2558 ถึงเดือนธันวาคม พ.ศ. 2565 ผู้วิจัยสามารถกำหนดตัวแบบที่เป็นไปได้ จำนวน 3 ตัวแบบ ได้แก่ SARIMA (1,0,0)(2,1,0)₁₂ SARIMA(1,01)(2,1,0)₁₂ SARIMA(0,0,1)(2,1,0)₁₂ และ SARIMA(1,0,2)(1,1,1)₁₂ โดยมีค่า AIC เท่ากับ 180.88, 180.27, 184.23 และ 167.16 ตามลำดับ เมื่อพิจารณาจากค่า AIC ที่ต่ำที่สุด จึงกำหนดตัวแบบที่เป็นไปได้ คือ SARIMA(1,0,2)(1,1,1)₁₂ ให้ค่า AIC เท่ากับ 167.16 ทำการประมาณของพารามิเตอร์และสถิติทดสอบความเหมาะสมของตัวแบบ SARIMA(1,0,2)(1,1,1)₁₂ พบว่าค่าทดสอบ Box-Ljung test มีค่าสถิติทดสอบ Q เท่ากับ 0.4451 มีค่า p-value เท่ากับ 0.9786 ซึ่งมากกว่าระดับนัยสำคัญ 0.05 ที่ผู้วิจัยกำหนด นั่นคือตัวแบบมีความเหมาะสมกับการพยากรณ์ ที่ระดับนัยสำคัญ 0.05 โดยตัวแบบที่ได้มีค่า RMSE เท่ากับ 0.4963 ดังตารางที่ 2

ตารางที่ 2 ค่าประมาณของพารามิเตอร์และสถิติทดสอบด้วยตัวแบบ SARIMA(p,d,q)(P,D,Q)₁₂

ค่าประมาณพารามิเตอร์		ตัวแบบ SARIMA(p,d,q)(P,D,Q) ₁₂			
		SARIMA (1,0,0)(2,1,0) ₁₂	SARIMA (1,01)(2,1,0) ₁₂	SARIMA (0,0,1)(2,1,0) ₁₂	SARIMA (1,0,2)(1,1,1) ₁₂
AR(1)	ค่าประมาณ	0.5268	0.9159		0.9869
	p-value	<0.0001	<0.0001		<0.0001
MA(1)	ค่าประมาณ		-0.6175	0.4941	-0.4794
	p-value		0.0010	<0.0001	<0.0001
MA(2)	ค่าประมาณ				-0.3081
	p-value				0.0069
SAR(1)	ค่าประมาณ	-0.5312	-0.6877	-0.4338	0.0257
	p-value	<0.0001	<0.0001	0.0001	0.8338
SAR(1)	ค่าประมาณ	-0.2871	-0.3505	-0.2118	
	p-value	0.0191	0.0038	0.0808	
SMR(1)	ค่าประมาณ				-1.0000
	p-value				<0.0001
AIC		180.88	180.27	184.23	167.16
ค่าสถิติทดสอบ Q		3.6290	3.4830	6.5753	0.4451
p-value		0.4585	0.4805	0.1601	0.9786
RMSE		0.6162	0.5986	0.6347	0.4963

การวิเคราะห์ข้อมูลเดือนมกราคม พ.ศ. 2558 ถึงเดือนธันวาคม พ.ศ. 2565 ด้วยวิธี SARIMA-ANN-REG โดยนำค่าพยากรณ์จากตัวแบบ SARIMA(1,0,2)(1,1,1)₁₂ และนำค่าส่วนเหลือพยากรณ์ด้วยตัวแบบ ANN แบบ Multi-Layer Perceptron (MLP) โดยเปลี่ยนจำนวนโหนดในชั้นซ่อน ตั้งแต่ 1-10 โดยใช้อัตราการเรียนรู้ 0.01 พบว่าโครงข่ายประสาทเทียมที่มีจำนวนโหนดในชั้นซ่อน เท่ากับ 8 มีค่า RMSE ต่ำที่สุด เท่ากับ 0.4993 เมื่อนำค่าพยากรณ์จากตัวแบบ ANN ไปสร้างตัวแบบการพยากรณ์ปริมาณการปล่อยก๊าซ CO₂ จากกลุ่มพลังงานในประเทศไทยร่วมกับค่าพยากรณ์จากตัวแบบ SARIMA(1,0,2)(1,1,1)₁₂ โดยวิธีการวิเคราะห์การถดถอย พบว่าเมื่อจำนวนโหนดในชั้นซ่อน เท่ากับ 8 มีค่า RMSE ต่ำที่สุด เท่ากับ 0.4992 ดังตารางที่ 3 โดยมีสมการการพยากรณ์เป็น

$$\hat{Y}_t = 1.0003\hat{F}_t^{SARIMA} + 1.0379\hat{F}_t^{ANN}$$

เมื่อนำตัวแบบ SARIMA(1,0,2)(1,1,1)₁₂ และตัวแบบ SARIMA-ANN-REG ไปพยากรณ์ปริมาณการปล่อยก๊าซ CO₂ ในเดือนมกราคม ถึงเดือนธันวาคม พ.ศ. 2566 จะได้ค่าพยากรณ์ดังตารางที่ 4 และพบว่าตัวแบบ SARIMA(1,0,2)(1,1,1)₁₂ มีค่า RMSE ของการพยากรณ์เท่ากับ 1.2307 ตัวแบบ SARIMA-ANN-REG เมื่อจำนวนโหนดในชั้นซ่อน เท่ากับ 8 มีค่า RMSE ของการพยากรณ์เท่ากับ 1.2106 และเมื่อนำตัวแบบ SARIMA-ANN-REG ไปพยากรณ์ปริมาณการปล่อยก๊าซ CO₂ ล่วงหน้าในเดือนสิงหาคม ถึงเดือนธันวาคม พ.ศ. 2566 ได้ค่าพยากรณ์ล่วงหน้าดังตารางที่ 5

ตารางที่ 3 ค่า RMSE ในการพยากรณ์ชุดข้อมูลด้วยวิธี ANN และ SARIMA-ANN-REG จำแนกตามจำนวนโหนดในชั้นซ่อน

จำนวนโหนด ในชั้นซ่อน	1	2	3	4	5	6	7	8	9	10
ANN	0.5020	0.5016	0.5016	0.5016	0.5017	0.5015	0.5015	0.4993	0.5013	0.5016
SARIMA-ANN- REG	0.5020	0.5016	0.5016	0.5016	0.5017	0.5014	0.5014	0.4992	0.5013	0.5016

ตารางที่ 4 ค่าพยากรณ์ปริมาณการปล่อย CO₂ ในเดือนมกราคม ถึงเดือนกรกฎาคม พ.ศ. 2566 โดยตัวแบบ SARIMA (1,0,2)(1,1,1)₁₂ และตัวแบบ SARIMA-ANN-REG

ช่วงเวลา	ค่าจริง	ค่าพยากรณ์	
		SARIMA (1,0,2)(1,1,1) ₁₂	SARIMA-ANN-REG โหนดซ่อน = 8
มกราคม 2566	20.3846	20.4083	20.3385
กุมภาพันธ์ 2566	20.4253	19.5937	19.5399
มีนาคม 2566	19.6091	22.0600	21.8919
เมษายน 2566	22.0822	20.5990	20.5073
พฤษภาคม 2566	21.0440	21.3328	21.2891
มิถุนายน 2566	22.0679	20.7949	20.7674
กรกฎาคม 2566	20.6670	20.6541	20.4964
RMSE		1.2307	1.2106

ตารางที่ 5 ค่าพยากรณ์ปริมาณการปล่อย CO₂ ล่วงหน้า 5 เดือนด้วยตัวแบบ SARIMA-ANN-REG โหนดซ่อน = 8

ช่วงเวลา	ค่าพยากรณ์
สิงหาคม 2566	20.6101
กันยายน 2566	19.9285
ตุลาคม 2566	20.2133
พฤศจิกายน 2566	20.5518
ธันวาคม 2566	20.5805

สรุปและอภิปรายผล

การศึกษาตัวแบบสำหรับการพยากรณ์ปริมาณการปล่อยก๊าซ CO₂ จากกลุ่มพลังงานในประเทศไทย ตั้งแต่เดือนมกราคม พ.ศ. 2558 ถึงเดือนมิถุนายน พ.ศ. 2566 พบว่าข้อมูลที่ใช้ในการศึกษามีค่าใกล้เคียงกัน เมื่อพิจารณาแนวโน้มของปริมาณการปล่อยก๊าซ CO₂ มีแนวโน้มลดลงหลังจากปี พ.ศ. 2563 ซึ่งเป็นช่วงสถานการณ์การแพร่ระบาดของ Covid-19 ทำให้ประชาชนไม่ออกจากบ้านจึงมีการใช้รถยนต์เป็นจำนวนน้อยลง จึงทำให้ปริมาณการปล่อยก๊าซ CO₂ จำนวนลดลง เมื่อทำการพยากรณ์ข้อมูลเดือนมกราคม พ.ศ. 2558 ถึงเดือนธันวาคม พ.ศ. 2565 พบว่าตัวแบบ SARIMA(1,0,2)(1,1,1)₁₂ ให้ค่า RMSE เท่ากับ 0.4963 และเมื่อนำไปพยากรณ์ข้อมูลเดือนมกราคม ถึงเดือนกรกฎาคม พ.ศ. 2566 ให้ค่า RMSE สำหรับการพยากรณ์เท่ากับ 1.2307 ต่อมานำค่าส่วนเหลือจากการพยากรณ์ด้วยตัวแบบ SARIMA(1,0,2)(1,1,1)₁₂ ไปทำการวิเคราะห์ตัวแบบพยากรณ์ด้วยวิธี ANN กำหนดจำนวนโหนดในชั้นซ่อน ตั้งแต่ 1-10 โดยใช้อัตราการเรียนรู้ 0.01 พบว่าจำนวนโหนดในชั้นซ่อน เท่ากับ 8 มีค่า RMSE ต่ำที่สุด ทำให้ตัวแบบ SARIMA-ANN-REG สำหรับการพยากรณ์ข้อมูลเดือนมกราคม พ.ศ. 2558 ถึงเดือนธันวาคม พ.ศ. 2565 มีค่า RMSE เท่ากับ 0.4992 และเมื่อนำไปพยากรณ์ข้อมูลเดือนมกราคม ถึงเดือนกรกฎาคม พ.ศ. 2566 ให้ค่า RMSE สำหรับการพยากรณ์เท่ากับ 1.2106 ผลการศึกษาค้นพบว่า การวิเคราะห์อนุกรมเวลาโดยการสร้างสมการการถอยจากตัวแบบ SARIMA และตัวแบบ ANN หรือตัวแบบ SARIMA-ANN-REG เป็นตัวแบบที่มีความแม่นยำในการพยากรณ์ปริมาณการปล่อยก๊าซ CO₂ จากกลุ่มพลังงานในประเทศไทย ซึ่งให้ค่า RMSE ต่ำกว่า ตัวแบบ SARIMA เพียงอย่างเดียว นั่นคือการวิเคราะห์ข้อมูลแบบผสม ระหว่างตัวแบบ SARIMA และตัวแบบ ANN ด้วยสมการการถอยทำให้การพยากรณ์ข้อมูลให้มีความแม่นยำมากยิ่งขึ้น โดยมีค่า RMSE ของตัวแบบเท่ากับ 0.4992 และค่า RMSE สำหรับการพยากรณ์ข้อมูลเดือนมกราคม ถึงเดือนกรกฎาคม พ.ศ. 2566 เท่ากับ 1.2106 ซึ่งพบว่าการนำตัวแบบ SARIMA ค่าส่วนเหลือจากการพยากรณ์ด้วยตัวแบบ SARIMA ไปทำการวิเคราะห์ตัวแบบพยากรณ์ด้วยวิธี ANN และนำสร้างตัวแบบผสมโดยการสร้างสมการการถอย ให้ผลลัพธ์สอดคล้องกับการศึกษาการพยากรณ์ข้อมูลราคาไข่ ข้อมูลราคาหมู ข้อมูลราคาบะหมี่กึ่งสำเร็จรูป ข้อมูลราคาหุ้น [7] ซึ่งพบว่าการพยากรณ์ข้อมูลด้วยตัวแบบ ARIMA-ANN-REG มีค่า RMSE ต่ำที่สุด และเมื่อนำผลจากการพยากรณ์ปริมาณการปล่อยก๊าซ CO₂ จากกลุ่มพลังงานในประเทศไทย ด้วยตัวแบบ SARIMA-ANN-REG นำไปพยากรณ์ล่วงหน้าเดือนสิงหาคม ถึงธันวาคม พ.ศ. 2566 ค่าพยากรณ์ที่ได้สามารถนำไปเป็นแนวทางให้หน่วยงานที่เกี่ยวข้องวางแผนเพื่อการลดการปล่อยก๊าซ CO₂ จากการศึกษาพบว่า ค่า RMSE ของตัวแบบ SARIMA เพียงอย่างเดียวมีค่า RMSE สูงกว่าการใช้ตัวแบบผสม SARIMA-ANN-REG เล็กน้อย จึงควรนำตัวแบบผสม SARIMA-ANN-REG ไปใช้ในการพยากรณ์ปริมาณการปล่อยก๊าซ CO₂ เนื่องจากสามารถลดความผิดพลาดในการพยากรณ์ได้ดีกว่า แต่การพยากรณ์ด้วยตัวแบบผสม SARIMA-ANN-REG จะใช้เวลาในการพยากรณ์มากกว่าการใช้ตัวแบบ SARIMA เพียงอย่างเดียว เนื่องจากมีขั้นตอนในการวิเคราะห์ที่มีมากกว่า ในการศึกษาครั้งนี้เป็นการพยากรณ์ในระยะสั้น (Short-term Forecasting) ไม่เกิน 1 ปี ในการศึกษาครั้งต่อไปอาจนำการพยากรณ์ในระยะยาว (36 เดือนขึ้นไป) มาเปรียบเทียบกันเพื่อหาตัวแบบที่มีประสิทธิภาพตามระยะเวลาการพยากรณ์บางวิธีให้ค่าความแม่นยำในการพยากรณ์ระยะสั้นแต่บางวิธีเหมาะสมกับการพยากรณ์ระยะยาว และในการศึกษาค้นคว้าวิจัยสร้างตัวแบบโครงข่ายประสาทเทียม (ANN) ด้วยเทคนิค Multi-Layer Perceptron (MLP) [8] ซึ่งผู้วิจัยอื่นที่สนใจศึกษาอาจเลือกใช้เทคนิคโครงข่ายแบบอื่น เช่น Recurrent Network, Boltzmann Machine [9] รวมถึงการศึกษาตัวแบบผสมตัวแบบ SARIMA กับ ANN ในลักษณะอื่น เช่น ตัวแบบ SARIMA-ANN [10] และตัวแบบ SARIMA-SVM [11] เป็นต้น

เอกสารอ้างอิง

- [1] Pitoonpong, S. (2021). *Global warming and labor*. Matichon Online. Retrieved August 1, 2023, from https://www.matichon.co.th/columnists/news_3037715
- [2] The Energy Policy and Planning Office. (2022). *Energy statistics*. Retrieved August 1, 2023, from <https://www.eppo.go.th/index.php/en/en-energystatistics/co2-statistic>
- [3] Boonmana, C., and Kulvanich, N. (2017). A Comparative prediction accuracy of hybrid time series models. *Thai Science and Technology Journal*, 25(2), 177-190. <https://li01.tci-thaijo.org/index.php/tstj/article/view/70308/57090>
- [4] Nyoni, T., and Bonga, W. G. (2019). Prediction of co₂ emissions in India using Arima models. *DRJ-Journal of Economics & Finance*, 4(2), 01-10. <https://ssrn.com/abstract=3346378>
- [5] Amphanthong, P. (2022). Comparison of methods for analysis of the air quality: A case study of alternative energy for biogas in suphanburi province. *Journal of Applied Research on Science and Technology (JARST)*, 21(1), 44-54. <https://doi.org/10.14456/jarst.2022.5>
- [6] Sala-ngam, S. (2022). Towards a predictor for CO₂ emission using regression analysis and an artificial neural network. *The Journal of Industrial Technology: Suan Sunandha Rajabhat University*, 10(2). 54-65. <https://ph01.tci-thaijo.org/index.php/fit-ssru/article/view/248813>
- [7] Kulvanich, N. (2020). Improving Prediction Accuracy of Time Series Data Using ARIMA-ANN Hybrid Model and Regression Analysis. *Thai Science and Technology Journal*, 28(12), 2101-2112. <https://li01.tci-thaijo.org/index.php/tstj/article/view/210424/170099>
- [8] Devadoss, A. V., and Ligori, T. A. A. (2013). Forecasting of stock prices using multi-layer perceptron. *International journal of computing algorithm*, 2(1), 440-449.
- [9] Li, W., Han, M., and Wang, J. (2020). Recurrent restricted Boltzmann machine for chaotic time-series prediction. In *2020 12th International Conference on Advanced Computational Intelligence (ICACI)* (pp. 439-445). The Institute of Electrical and Electronics Engineers (IEEE).
- [10] Krishnan, G., Mehta, K. B. V., and Yadav, R. S. (2022). Assessment of future pattern of rainfall in different zones of Kerala using incorporation of SARIMA, ANN, and hybrid SARIMA-ANN models. *Economic Affairs (New Delhi)*, 67(5), 823-832.
- [11] Abdullah, A. S., Ruchjana, B. N., and Jaya, I. G. N. M. (2021). Comparison of SARIMA and SVM model for rainfall forecasting in Bogor city, Indonesia. *Journal of Physics: Conference Series*, 1722(1), Article number 012061.