

การพยากรณ์ปริมาณฝุ่นละออง PM2.5 ในจังหวัดนครราชสีมา
ด้วยตัวแบบอนุกรมเวลาและการเรียนรู้ของเครื่องจักร
Forecasting PM2.5 Quantity in Nakhon Ratchasima Province
Using Time Series and Machine Learning

พียดา นาภูวงศ์¹ ชันัญกาญจน์ แสงประสาน^{1*} สมบูรณ์ ชาวชายโขง¹ และ ธนวรรณ วงศ์จิรัญธิ²

Piyada Napuvong¹ Chanankarn Seangprasan^{1*} Somboon Chaochaikong¹ and Thanawan Wongjiratthiti²

¹คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏสกลนคร สกลนคร 47000 ประเทศไทย

²โรงเรียนสกลราชวิทยานุกูล สกลนคร 47000 ประเทศไทย

¹Faculty of Science and Technology, Sakon Nakhon Rajabhat University, Sakon Nakhon 47000, Thailand

²Sakolrajwittayanukul School, Sakon Nakhon 47000, Thailand

*E-mail: chanankarn@snru.ac.th

Received: Jun 17, 2024

Revised: Jul 16, 2024

Accepted: Jul 23, 2024

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อสร้างตัวแบบพยากรณ์ปริมาณฝุ่นละออง PM2.5 รายเดือนในจังหวัดนครราชสีมา โดยใช้ 3 วิธีการคือวิธีบ็อกซ์-เจนกินส์ การถดถอยเชิงเส้นพหุคูณ และโครงข่ายประสาทเทียม และเปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์ดังกล่าวด้วยค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) สำหรับการสร้างตัวแบบการพยากรณ์ ใช้ข้อมูลปริมาณ PM2.5 ชุดที่ 1 ที่เก็บจาก 42 คาบเวลา ตั้งแต่เดือนกรกฎาคม พ.ศ. 2562 ถึงเดือนธันวาคม พ.ศ. 2565 ส่วนการเปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์ ใช้ข้อมูลปริมาณ PM2.5 ชุดที่ 2 ที่เก็บจาก 12 คาบเวลา ตั้งแต่เดือนมกราคม พ.ศ. 2566 ถึงเดือนธันวาคม พ.ศ. 2566 ผลการวิจัย พบว่า ตัวแบบพยากรณ์ปริมาณ PM2.5 รายเดือนในจังหวัดนครราชสีมาที่เหมาะสม คือ ตัวแบบโครงข่ายประสาทเทียม ซึ่งใช้ตัวแปรอิสระทั้ง 16 ตัวแปร มีชั้นซ่อน 1 ชั้น และจำนวนโหนดในชั้นซ่อน เท่ากับ 7 โหนด มีค่าอัตราการเรียนรู้ (Learning rate) เท่ากับ 0.1 ค่าโมเมนตัม (Momentum) เท่ากับ 0.1 และจำนวนการเรียนรู้ (Epoch) 500 รอบ นอกจากนี้ยังมีค่า RMSE ต่ำสุดเท่ากับ 2.560 และค่าสัมประสิทธิ์การตัดสินใจ (R^2) เท่ากับ 0.988

คำสำคัญ: ฝุ่นละออง PM2.5 วิธีบ็อกซ์-เจนกินส์ การถดถอยเชิงเส้นพหุคูณ โครงข่ายประสาทเทียม

Abstract

This research aimed to create models to forecast monthly PM2.5 quantity in Nakhon Ratchasima province by using 3 methods including Box-Jenkins method, Multiple linear regression and Artificial neural networks and to compare the performance of the forecasting models using Root mean square error (RMSE). For creating forecasting models, the data set 1 of PM2.5 quantity collected from 42 time periods from July 2019 to December 2022 was used while for the performance comparison among the forecasting models, the data set 2 of PM2.5 quantity collected from 12 time periods from January 2023 to December 2023 was used. The results demonstrated that the Artificial neural network was the suitable model for forecasting monthly PM2.5 quantity in Nakhon Ratchasima province. This model consisted of 14 inputs, 1 hidden layer with 7 nodes, a learning rate of 0.1, a momentum of 0.1, and 500 epochs. Furthermore, it had the lowest RMSE of 2.560 and a coefficient of determination (R^2) of 0.988.

Keywords: PM2.5, Box-Jenkins method, Multiple linear regression, Artificial neural networks

1. บทนำ

มลพิษทางอากาศเป็นปัญหาที่ทั่วโลกต้องเผชิญอยู่ ซึ่งในปัจจุบันมลพิษทางอากาศเป็นปัญหาที่สำคัญที่มีผลต่อสุขภาพของมนุษย์ อย่างไรก็ตามการกระทำของมนุษย์ส่งผลต่อการสร้างมลพิษทางอากาศแทบทั้งหมดทั้งภาคอุตสาหกรรม ด้านคมนาคมการขนส่ง และด้านการเกษตร โดยการเผาพื้นที่ทำการเกษตรหรือการลักลอบเผาป่า ดังนั้นปัญหามลพิษทางอากาศจึงเป็นหนึ่งในความท้าทายเกือบทุกประเทศที่กำลังประสบปัญหาอยู่ ณ ขณะนี้ เพราะมีผลกระทบต่อสุขภาพและเป็นสาเหตุสำคัญของการเสียชีวิต [1] จังหวัดนครราชสีมาเป็นพื้นที่ที่มีประชากรหนาแน่น มีปริมาณโรงงานอุตสาหกรรมมากเป็นอันดับต้น ๆ ของภาคตะวันออกเฉียงเหนือ และเป็นศูนย์กลางการคมนาคมไปสู่ภาคตะวันออกเฉียงเหนือ จึงประสบปัญหาค่าฝุ่นละอองขนาดเล็กไม่เกิน 2.5 ไมครอน (PM2.5) เกินค่ามาตรฐาน ซึ่งมักจะเกิดอยู่ในช่วงรอยต่อฤดูหนาวเข้าสู่ฤดูร้อน ตั้งแต่เดือนพฤศจิกายนจนถึงเดือนมีนาคม [2] แต่ในปัจจุบันปัญหาปริมาณฝุ่น PM2.5 มีแนวโน้มเพิ่มสูงขึ้นเรื่อย ๆ ดังที่สำนักงานสิ่งแวดล้อมและควบคุมมลพิษที่ 11 (นครราชสีมา) รายงานว่า ไฟป่าจากอุทยานแห่งชาติเป็นต้นเหตุก่อฝุ่นควันอันดับหนึ่งในพื้นที่ สภาพอากาศค่อนข้างมีเมฆมาก ส่งผลให้ท้องฟ้าปิด ประกอบกับการเผาป่า และเผาในพื้นที่การเกษตรมากถึงวันละประมาณ 45 จุด เป็นต้นเหตุทำให้เกิดฝุ่นละออง PM2.5 สูงขึ้นเกินค่ามาตรฐาน [3] จากสถานการณ์ที่กล่าวมา จะเห็นได้ว่า ปริมาณฝุ่น PM2.5 ส่วนใหญ่เกิดจากการเผาวัสดุทางการเกษตรและพื้นที่ป่า การคมนาคม รวมทั้งโรงงานอุตสาหกรรมและประเทศเพื่อนบ้าน ด้วยเหตุนี้ปริมาณฝุ่น PM2.5 ก่อให้เกิดปัญหาด้านสุขภาพ สิ่งแวดล้อมและเศรษฐกิจเป็นอย่างยิ่ง ถ้าทราบข้อมูลปริมาณ PM2.5 ล่วงหน้าจะสามารถประเมินความรุนแรงของปริมาณฝุ่น PM2.5 ตลอดจนการพิจารณาปัจจัยที่เกี่ยวข้องที่จะส่งผลต่อปริมาณฝุ่น PM2.5 สามารถนำไปวางแผนเพื่อป้องกันปัญหาฝุ่น PM2.5 ได้ สำหรับการสร้างตัวแบบ (Model) การพยากรณ์ปริมาณฝุ่น PM2.5 ในพื้นที่ต่าง ๆ ในประเทศไทยมีหลากหลายวิธี เช่น การพยากรณ์ปริมาณฝุ่น PM2.5 รายชั่วโมงของสถานีตรวจวัดอากาศที่สถานีสูบน้ำประตูปลแสน อำเภอเมือง จังหวัดนครราชสีมา ใช้ข้อมูล ณ พ.ศ. 2562-2563 ด้วยการพยากรณ์แบบอนุกรมเวลา (Time series forecasting) โดยวิธีการปรับเรียบ (Smoothing forecasting method) และวิธีบ็อกซ์-เจนกินส์ (Box-Jenkins method) พบว่า วิธีบ็อกซ์-เจนกินส์ พยากรณ์ปริมาณ PM2.5 รายชั่วโมงได้ความแม่นยำ [4] การพยากรณ์

แบบโครงข่ายประสาทเทียม (Artificial neural networks) ในการพยากรณ์ปริมาณความเข้มข้นของฝุ่น PM2.5 ในเขตพื้นที่ของสถานีโรงไฟฟ้าพระนครใต้ ตำบลบางโปร่ง อำเภอมือง จังหวัดสมุทรปราการ ด้วยการใช้วิธีผสมของโครงข่ายประสาทเทียมเพอร์เซพตรอนแบบหลายชั้น (Multi-layer perceptron) โดยมีการเรียนรู้แบบแพร่ย้อนกลับและห่วงโซ่มาร์คอฟ (Markov chain) พบว่าให้ผลการพยากรณ์ที่มีความแม่นยำ [5] และการพยากรณ์ปริมาณฝุ่น PM2.5 เขตเมืองในภาคเหนือของประเทศไทยด้วยวิเคราะห์การถดถอยเชิงเส้นพหุคูณหลายตัวแปร (Multivariate linear regression) และใช้ข้อมูลจากสถานีวัดในจังหวัดน่าน เชียงใหม่ และลำปาง ซึ่งพบว่าการเพิ่มตัวแปรก๊าซ SO_2 , NO_2 , CO และ O_3 เป็นตัวแปรอิสระเพิ่มเติมจากตัวแปรทางอุตุนิยมวิทยาภาคพื้นดิน ช่วยให้ตัวแบบมีความแม่นยำขึ้น [6] นอกจากนี้ยังมีการพยากรณ์ปริมาณฝุ่น PM2.5 ในกรุงเทพฯ จากข้อมูลดาวเทียมร่วมกับตัวแปรทางอุตุนิยมวิทยา [7] การพยากรณ์ปริมาณฝุ่น PM2.5 ในสถานีวัดมหาวิทยาลัยเกษตรศาสตร์ กรุงเทพฯ เปรียบเทียบประสิทธิภาพการพยากรณ์ด้วยชุดอนุกรมเวลาปริมาณฝุ่น พ.ศ. 2563 พบว่าวิธีการวิเคราะห์ถดถอยมีความแม่นยำมากกว่าวิธีโครงข่ายประสาทเทียม [8] ในต่างประเทศ มีการพยากรณ์ปริมาณฝุ่น PM2.5 ในเมืองเดลีประเทศอินเดีย ด้วยวิธีโครงข่ายประสาทเทียมมีความแม่นยำกว่าวิธีซัพพอร์ตเวกเตอร์แมชชีน (Support vector machine) [9] เช่นเดียวกับการพยากรณ์ปริมาณฝุ่น PM2.5 จาก 7 สถานีวัดในประเทศมาเลเซีย พบว่า วิธีการวิเคราะห์ถดถอย และวิธีโครงข่ายประสาทเทียมมีเหมาะสมในการพยากรณ์ชุดข้อมูล แต่วิธีโครงข่ายประสาทเทียมมีความแม่นยำกว่า [10]

จากการศึกษาข้างต้นจะเห็นได้ว่า วิธีการพยากรณ์ปริมาณฝุ่น PM2.5 มีการประยุกต์ใช้เทคนิคทางด้านอนุกรมเวลา คือ วิธีบ็อกซ์-เจนกินส์ และเทคนิคด้านการเรียนรู้ของเครื่องจักร (Machine learning) คือ การวิเคราะห์ถดถอยและวิธีโครงข่ายประสาทเทียม ซึ่งสามารถใช้พยากรณ์ข้อมูลได้อย่างมีประสิทธิภาพ แตกต่างกันไปตามพื้นที่ที่ทำการศึกษา สำหรับการพยากรณ์ปริมาณฝุ่น PM2.5 ของจังหวัดนครราชสีมาในอดีต [4] ใช้เทคนิคด้านอนุกรมเวลา แต่การศึกษาหลายพื้นที่ [5], [8]-[10] มีการใช้ตัวแปรอิสระร่วมพยากรณ์ปริมาณฝุ่น PM2.5 ด้วยเทคนิคการเรียนรู้ของเครื่องจักร ช่วยการพยากรณ์มีประสิทธิภาพมากขึ้น

ดังนั้นผู้วิจัยจึงศึกษาหาตัวแบบการพยากรณ์ปริมาณมลพิษทางอากาศ PM2.5 รายเดือนในจังหวัดนครราชสีมา

โดยใช้วิธีตัวแบบวิธีบ็อกซ์-เจนกินส์ (Box-Jenkins method) การถดถอยเชิงเส้นพหุคูณ (Multiple linear regression: MLR) และโครงข่ายประสาทเทียม (Artificial neural networks: ANN) และเปรียบเทียบประสิทธิภาพตัวแบบการพยากรณ์ปริมาณมลพิษทางอากาศ PM2.5 ช่วงหน้าในการพยากรณ์ระยะสั้น ด้วยเกณฑ์รากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (Root mean square error: RMSE) ต่ำสุด เพื่อให้ได้ตัวแบบสำหรับการคาดการณ์ปริมาณมลพิษทางอากาศ PM2.5 ในอนาคตให้มีความแม่นยำ และสามารถนำไปใช้วางแผนงานที่เกี่ยวข้องในอนาคตได้อย่างมีประสิทธิภาพยิ่งขึ้น

2. วัสดุ อุปกรณ์และวิธีการวิจัย

2.1. เทคนิคการพยากรณ์

เทคนิคการพยากรณ์ที่ใช้ในการวิจัยนี้มี 3 วิธี ดังนี้

2.1.1. วิธีบ็อกซ์-เจนกินส์ (Box-Jenkins)

การพยากรณ์โดยวิธีบ็อกซ์-เจนกินส์เป็นวิธีการพยากรณ์ข้อมูลอนุกรมเวลาที่มีส่วนประกอบฤดูกาล มักใช้ตัวแบบเชิงคูณ คือ $ARIMA(p,d,q)(P,D,Q)_s$ [11] ประกอบด้วยกระบวนการ Autoregressive (AR), I (Integrated) และ Moving average (MA) รูปแบบทั่วไปดังสมการ (1)

$$\phi_p(B)\Phi_p(B^s)(1-B)^d(1-B^s)^D Y_t = \theta_q(B)\Theta_q(B^s)a_t \quad (1)$$

เมื่อ

Y_t คือ ข้อมูลอนุกรมเวลา ณ คาบเวลา t

a_t คือ ค่าคลาดเคลื่อน ณ คาบเวลา t ซึ่งมีการแจกแจงปกติ มีค่าเฉลี่ย 0 ความแปรปรวน σ^2

S คือ จำนวนคาบฤดูกาล

d, D คือ ลำดับครั้งที่หาผลต่างแบบไม่มีฤดูกาล d และแบบมีฤดูกาล D

$\phi_p(B)$ คือ ตัวดำเนินการสหสัมพันธ์ในตัวเองแบบไม่มีฤดูกาล อันดับ p ซึ่งเท่ากับ $1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$

$\Phi_p(B^s)$ คือ ตัวดำเนินการสหสัมพันธ์ในตัวเองแบบมีฤดูกาล อันดับ P ซึ่งเท่ากับ $1 - \phi_1 B^{1s} - \phi_2 B^{2s} - \dots - \phi_p B^{Ps}$

$\theta_q(B)$ คือ ตัวดำเนินการเฉลี่ยเคลื่อนที่แบบไม่มีฤดูกาลอันดับ q ซึ่งเท่ากับ $1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$

$\Theta_q(B^s)$ คือ ตัวดำเนินการเฉลี่ยเคลื่อนที่แบบมีฤดูกาลอันดับ Q ซึ่งเท่ากับ $1 - \theta_1 B^{1s} - \theta_2 B^{2s} - \dots - \theta_q B^{Qs}$

B คือ ตัวดำเนินการแบบถอยหลัง (Backward shift operator) นั่นคือ $(B)Y_t = Y_{t-1}$

การสร้างตัวแบบพยากรณ์ด้วยวิธีบ็อกซ์-เจนกินส์ มีขั้นตอนที่สำคัญ 4 ขั้นตอน ดังนี้ 1) การทดสอบกระบวนการคงที่ (Stationary time series) เป็นอนุกรมเวลามีค่าเฉลี่ยและความแปรปรวนคงที่ โดยใช้สถิติทดสอบ Augmented Dickey-Fuller (ADF) 2) การกำหนดตัวแบบ (Identification) เป็นการกำหนดตัวแบบที่คาดว่าจะเป็นไปได้ของตัวแบบ $ARIMA(p,d,q)(P,D,Q)_s$ โดยพิจารณาจากกราฟฟังก์ชันสหสัมพันธ์ในตัวเอง (Autocorrelation function: ACF) และกราฟฟังก์ชันสหสัมพันธ์ในตัวเองบางส่วน (Partial autocorrelation function: PACF) เพื่อกำหนดตัวแบบที่คาดว่าจะเป็นไปได้ 3) ทำการประมาณค่าพารามิเตอร์ (Parameter estimation) ที่คาดว่าจะเป็นไปได้ และตรวจสอบความเหมาะสมของตัวแบบ จากนั้นทำการทดสอบนัยสำคัญของค่าประมาณพารามิเตอร์ด้วยสถิติทดสอบ Z-test 4) นำตัวแบบที่ได้จากขั้นตอนที่ 3 มาตรวจสอบความเป็นอิสระต่อกัน พิจารณาจากสถิติการทดสอบของสถิติทดสอบ Ljung-Box อย่างมีนัยสำคัญทางสถิติที่ระดับนัยสำคัญ 0.05 และต้องมีค่า Akaike information criterion (AIC) ต่ำสุด ในกรณีตัวแบบที่มีค่า AIC ต่ำสุดใกล้เคียงกัน จะต้องทำการวัดประสิทธิภาพของตัวแบบ เพื่อให้ได้ตัวแบบที่เหมาะสมในการพยากรณ์

2.1.2. การถดถอยเชิงเส้นพหุคูณ (Multiple linear regression)

วิธีการวิเคราะห์ทางสถิติที่เกี่ยวข้องกับการสร้างตัวแบบทางคณิตศาสตร์ในระดับประชากรเพื่อแสดงความสัมพันธ์ระหว่างตัวแปรตามแทนด้วย Y_i และตัวแปรอิสระแทนด้วย X_i ซึ่งเป็นตัวแปรที่ใช้ในการพยากรณ์ อาจมีตัวแปรอิสระมากกว่า 1 ตัวแปร [12] สามารถเขียนความสัมพันธ์ในรูปแบบเมทริกซ์ของสมการการถดถอยเชิงเส้นพหุคูณของข้อมูลกลุ่มตัวอย่าง ดังสมการ (2)

$$\hat{y}_j = \hat{\beta}_0 + \hat{\beta}_1 x_{1j} + \hat{\beta}_2 x_{2j} + \dots + \hat{\beta}_p x_{pj} \quad (2)$$

เมื่อ

$\hat{y}_j$ คือ ค่าพยากรณ์ตัวแปรตามตัวที่ j

x_{ij} คือ ค่าตัวแปรอิสระที่ i ของข้อมูลตัวที่ j

เมื่อ $i = 1, 2, 3, \dots, p$ และ

$j = 1, 2, 3, \dots, n$

p คือ จำนวนตัวแปรอิสระในสมการถดถอย

$\hat{\beta}_0$ คือ ค่าประมาณค่าคงที่ หรือจุดตัดแกน Y

$\hat{\beta}_i$ คือ ค่าประมาณสัมประสิทธิ์การถดถอย (Regression coefficient) ของตัวแปรอิสระที่ i

เมื่อ $i = 1, 2, 3, \dots, p$

n คือ จำนวนข้อมูล

สามารถเขียนอยู่ในรูปแบบของเมทริกซ์ ได้ดังสมการ (3)

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_p \end{bmatrix} \quad (3)$$

หรือ $\hat{y} = x\hat{\beta}$ ซึ่งค่า $\hat{\beta}$ คำนวณได้จาก $\hat{\beta} = (x'x)^{-1}x'y$

2.1.3. โครงข่ายประสาทเทียม (Artificial neural networks)

โครงข่ายประสาทเทียมเป็นการเรียนรู้วิธีการสร้างตัวแบบที่มีลักษณะการทำงานคล้ายกับสมองมนุษย์ ซึ่งประกอบด้วยเซลล์ประสาท (Neuron) ที่มีเส้นประสาทแตกแขนงออกมาเรียกว่า เดนไดรต์ (Dendrites) มีหน้าที่ส่งสัญญาณเชื่อมโยงระหว่างเส้นประสาทอื่น ๆ ณ ตำแหน่งแอกซอน (Axon) ทำหน้าที่รับส่งสัญญาณไปยังเซลล์ต่าง ๆ ในรูปแบบของสัญญาณกระตุ้นไฟฟ้า องค์ประกอบที่สำคัญของโครงข่ายประสาทเทียม ประกอบด้วยชั้นอินพุต (Input layer) ชั้นซ่อน (Hidden layer) ซึ่งภายในชั้นซ่อนมีกระบวนการ

ประมวลผล ประกอบไปด้วย 2 ฟังก์ชัน ได้แก่ ฟังก์ชันรวม (Summation function) และฟังก์ชันกระตุ้น (Activation function) ชั้นซ่อนนี้สามารถมีได้มากกว่า 1 ชั้นซ่อน และชั้นสุดท้าย คือ ชั้นเอาต์พุต (Output layer) โครงข่ายประสาทเทียมมี 2 ลักษณะสำคัญ ได้แก่ โครงข่ายแบบชั้นเดียว (Single layer neural network) และโครงข่ายแบบหลายชั้น (Multi-layer neural network) [13] ซึ่งโครงข่ายแบบหลายชั้นมีกระบวนการทำงานมี 2 ลักษณะ คือ โครงข่ายประสาทเทียมแบบแพร่ไปข้างหน้า (Feed forward neural networks) และการแพร่ย้อนกลับ (Backpropagation) ดัง Figure 1

การแพร่ย้อนกลับมีพารามิเตอร์ที่สำคัญในการวิเคราะห์ ได้แก่ อัตราการเรียนรู้ (Learning rate: η) เป็นการปรับค่าน้ำหนัก (Weight) ของตัวแบบเพื่อลดค่าผิดพลาด (Error) เกณฑ์การหยุด (Termination criteria) เป็นการปรับค่าน้ำหนักให้คงที่ และลดความคลาดเคลื่อนของการพยากรณ์โมเมนตัม (Momentum: β) เป็นการปรับค่าน้ำหนักเข้าใกล้ค่าที่เหมาะสมที่สุด จำนวนรอบการเรียนรู้ (Epochs) เป็นพารามิเตอร์ที่มีผลต่อการเรียนรู้ของตัวแบบ ในการกำหนดจำนวนรอบการเรียนรู้มากเกินไปจะก่อให้เกิดการเรียนรู้มากเกินไป ซึ่งเรียกว่า Overfitting และการกำหนดจำนวนโหนดชั้นซ่อน จะต้องไม่เกินโหนดของชั้นอินพุต ซึ่งจำนวนเซลล์ประสาทในแต่ละชั้น ควรมีประมาณ 2 ใน 3 ของจำนวนข้อมูลอินพุต [14]

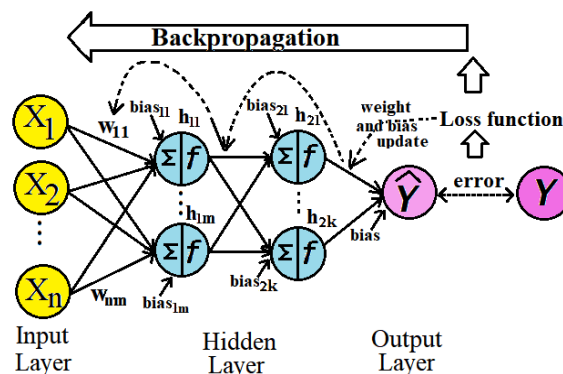


Figure 1 Backpropagation process

2.2. การเตรียมข้อมูลและการวิเคราะห์ข้อมูล

การศึกษาในครั้งนี้ใช้ข้อมูลอนุกรมเวลาของปริมาณมลพิษทางอากาศ PM2.5 รายเดือนในจังหวัดนครราชสีมา และข้อมูลที่เกี่ยวข้องตั้งแต่เดือนกรกฎาคม พ.ศ. 2562 ถึงเดือนธันวาคม พ.ศ. 2566 จำนวน 54 คาบเวลา ซึ่งเก็บข้อมูล

โดยกรมควบคุมมลพิษ [15] และศูนย์อุตุนิยมวิทยาภาคตะวันออกเฉียงเหนือตอนล่าง [16] การวิจัยนี้พยากรณ์ตัวแปรตาม คือ ปริมาณมลพิษทางอากาศ PM2.5 เฉลี่ยรายเดือน ในจังหวัดนครราชสีมา (PM25_mean : หน่วย $\mu\text{g}/\text{m}^3$) โดยใช้ตัวแปรอิสระ 16 ตัว ดังนี้

- 1) ปริมาณ PM2.5 เฉลี่ยรายเดือนของเดือนก่อนหน้า (PM2.5_meanbefore) หน่วย $\mu\text{g}/\text{m}^3$
- 2) ปริมาณ PM2.5 ค่าสูงสุดของเดือนก่อนหน้า (PM2.5_maxbefore) หน่วย $\mu\text{g}/\text{m}^3$
- 3) ปริมาณฝุ่นละอองขนาดไม่เกิน 10 ไมครอน ค่าสูงสุดของเดือน (PM10_max) หน่วย $\mu\text{g}/\text{m}^3$
- 4) ปริมาณฝุ่นละอองขนาดไม่เกิน 10 ไมครอน เฉลี่ยรายเดือน (PM10_mean) หน่วย $\mu\text{g}/\text{m}^3$
- 5) ปริมาณก๊าซซัลเฟอร์ไดออกไซด์ ค่าสูงสุดของเดือน (SO2_max) หน่วย ppb
- 6) ปริมาณก๊าซซัลเฟอร์ไดออกไซด์เฉลี่ยรายเดือน (SO2_mean) หน่วย ppb
- 7) ปริมาณก๊าซไนโตรเจนไดออกไซด์ค่าสูงสุดของเดือน (NO2_max) หน่วย ppb
- 8) ปริมาณก๊าซไนโตรเจนไดออกไซด์เฉลี่ยรายเดือน (NO2_mean) หน่วย ppb
- 9) ปริมาณก๊าซคาร์บอนมอนนอกไซด์ค่าสูงสุดของเดือน (CO_max) หน่วย ppm

- 10) ปริมาณก๊าซคาร์บอนมอนนอกไซด์เฉลี่ยรายเดือน (CO_mean) หน่วย ppm
 - 11) ปริมาณก๊าซโอโซนค่าสูงสุดของเดือน (O3_max) หน่วย ppb
 - 12) ปริมาณก๊าซโอโซนเฉลี่ยรายเดือน (O3_mean) หน่วย ppb
 - 13) อุณหภูมิต่ำสุดของเดือน (Temp_min) หน่วย $^{\circ}\text{C}$
 - 14) ความชื้นสัมพัทธ์เฉลี่ยรายเดือน (RH) หน่วย %
 - 15) ความเร็วลมเฉลี่ยรายเดือน (Wind_speed) หน่วย m/s
 - 16) ปริมาณน้ำฝนเฉลี่ยรายเดือน (Rainfall) หน่วย mm
- ผู้วิจัยคัดเลือกตัวแปรอิสระหรือคัดเลือกคุณลักษณะ (Feature selection) จำนวน 16 ตัวแปรที่มีความสัมพันธ์อย่างมีนัยสำคัญทางสถิติกับตัวแปรตาม ด้วยการทดสอบนัยสำคัญของสัมประสิทธิ์สหสัมพันธ์เพียร์สัน (Pearson correlation coefficient) ของตัวแปรรายคู่ ที่ระดับนัยสำคัญ 0.05 ค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สันเฉพาะคู่ตัวแปรที่มีนัยสำคัญทางสถิติ ดัง Figure 2

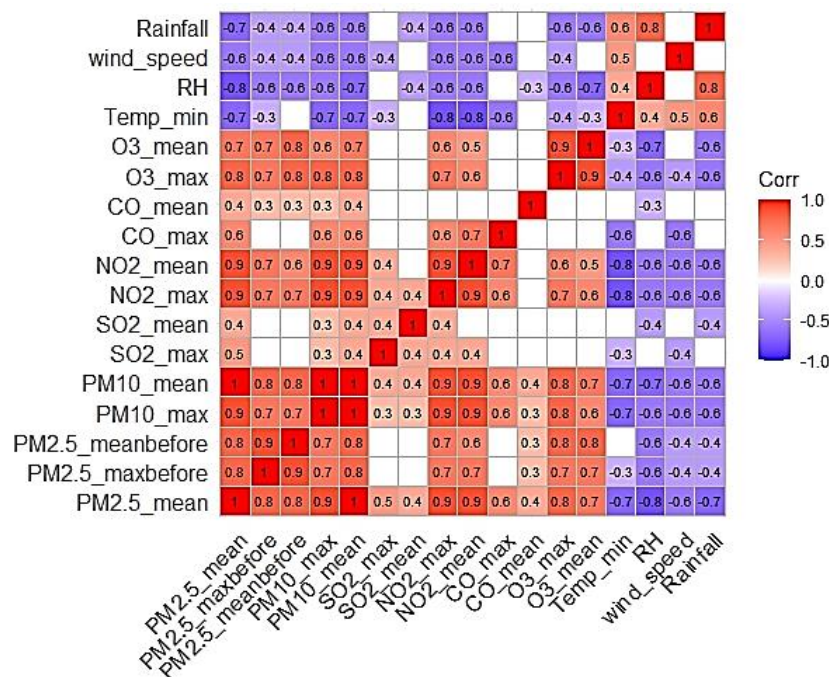


Figure 2 A correlation matrix

จาก Figure 2 พบว่า ตัวแปรอุณหภูมิต่ำสุดของเดือน ความชื้นสัมพัทธ์เฉลี่ยรายเดือน ความเร็วลมเฉลี่ยรายเดือน และปริมาณน้ำฝนเฉลี่ยรายเดือน มีความสัมพันธ์เชิงลบอย่างมีนัยสำคัญกับปริมาณมลพิษทางอากาศเฉลี่ยรายเดือนในจังหวัดนครราชสีมา ส่วนตัวแปรอิสระอื่น ๆ จำนวน 12 ตัวแปร มี

ความสัมพันธ์เชิงบวกอย่างมีนัยสำคัญกับปริมาณมลพิษทางอากาศเฉลี่ยรายเดือนในจังหวัดนครราชสีมา

ผู้วิจัยได้ทำการแบ่งข้อมูลออกเป็น 2 ชุด คือ ชุดที่ 1 ตั้งแต่เดือนกรกฎาคม พ.ศ. 2562 ถึงเดือนธันวาคม พ.ศ. 2565 จำนวน 42 คาบเวลา สร้างตัวแบบการพยากรณ์ด้วยวิธี

บอซ-เจนกินส์ด้วยภาษา R คำสั่ง `adf.test` ใน `library(tseries)` คำสั่ง `Arima()` ใน `library(forecast)` คำสั่ง `coef.test` ใน `library(lmtest)` สร้างตัวแบบการถดถอยเชิงเส้นพหุคูณด้วยภาษา R คำสั่ง `lm()` ใน `library(stats)` และสร้างตัวแบบโครงข่ายประสาทเทียมด้วยโปรแกรม Rapidminer ข้อมูลชุดที่ 2 ตั้งแต่เดือนมกราคม พ.ศ. 2566 ถึงเดือนธันวาคม พ.ศ. 2566 จำนวน 12 คาบเวลา ทั้งนี้เพื่อให้ข้อมูลชุดทดสอบสะท้อนอิทธิพลของฤดูกาล ทำการทดสอบประสิทธิภาพของตัวแบบการพยากรณ์ทั้ง 3 วิธี โดยใช้เกณฑ์ค่า Root mean squared error (RMSE) ต่ำสุด

2.3. การเปรียบเทียบประสิทธิภาพของตัวแบบ

การเปรียบเทียบประสิทธิภาพของตัวแบบการพยากรณ์ทั้ง 3 วิธี ใช้เกณฑ์ค่า RMSE ต่ำสุด ดังสมการ (4)

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n}} \quad (4)$$

เมื่อ

Y_i คือ ค่าสังเกตข้อมูลที i เมื่อ $i = 1, 2, 3, \dots, n$

$\hat{Y}_i$ คือ ค่าพยากรณ์ของข้อมูลที i

n คือ จำนวนข้อมูล

3. ผลการวิจัย

3.1. ผลการสร้างตัวแบบบอซ-เจนกินส์

ปริมาณมลพิษทางอากาศ PM2.5 โดยเฉลี่ยรายเดือนของจังหวัดนครราชสีมา ตั้งแต่เดือนกรกฎาคม พ.ศ. 2562 ถึงเดือนธันวาคม พ.ศ. 2565 จำนวน 42 คาบเวลา มีความเคลื่อนไหวของข้อมูลดัง Figure 3

จาก Figure 3 พิจารณาการเคลื่อนไหวข้อมูลชุดที่ 1 ของปริมาณมลพิษทางอากาศ PM2.5 โดยเฉลี่ยรายเดือนของจังหวัดนครราชสีมา ทำการทดสอบกระบวนการคงที่ของ

ข้อมูลด้วยสถิติทดสอบ ADF พบว่า ค่าสถิติ Dickey-Fuller เท่ากับ -4.9718 และค่า p-value < 0.01 ซึ่งน้อยกว่าระดับนัยสำคัญ 0.05 แสดงว่าข้อมูลมีกระบวนการคงที่ แต่จาก Figure 3 จะเห็นว่าปริมาณมลพิษทางอากาศ PM2.5 มีการเคลื่อนไหวคล้ายคลื่นไซน์ (Sine wave) อาจมีอิทธิพลของฤดูกาล ผู้วิจัยจึงสร้างทางเลือกในการสร้างตัวแบบเพิ่มมากขึ้น โดยทดลองแปลงค่าข้อมูลด้วยการหาผลต่างฤดูกาลอันดับที่ 1 และการหาผลต่างอันดับที่ 1 ทำการทดสอบกระบวนการคงที่พบว่าแปลงค่าข้อมูลด้วยการหาผลต่างฤดูกาลอันดับที่ 1 ไม่มีกระบวนการคงที่ แต่ผลการทดสอบกระบวนการคงที่ของข้อมูลที่แปลงค่าด้วยผลต่างอันดับที่ 1 มีค่าสถิติ Dickey-Fuller เท่ากับ -3.7317 และค่า p-value เท่ากับ 0.03503 ซึ่งน้อยกว่าระดับนัยสำคัญ 0.05 แสดงว่าข้อมูลปริมาณมลพิษทางอากาศ PM2.5 เมื่อทำการแปลงข้อมูลด้วยผลต่างอันดับที่ 1 แล้วมีกระบวนการคงที่เช่นเดียวกัน จากนั้นทำการระบุตัวแบบที่คาดว่าจะเป็นไปได้จากกราฟ ACF และ PACF ของข้อมูลเดิมดัง Figure 4 และกราฟ ACF และ PACF ของข้อมูลที่แปลงค่าแล้ว (d=1) ดัง Figure 5

จาก Figure 4 และ Figure 5 เมื่อกำหนดตัวแบบที่คาดว่าจะเป็นไปได้ แล้วประมาณพารามิเตอร์ และทดสอบนัยสำคัญด้วยสถิติ Z เพื่อคัดเลือกตัวแบบที่เหมาะสม ได้ 7 ตัวแบบ ดังนี้ ARIMA(1,0,0)(1,0,0)₁₂, ARIMA(0,0,0)(1,0,0)₁₂, ARIMA(2,0,1), ARIMA(2,0,0), ARIMA(3,0,2)(1,0,0)₁₂, ARIMA(3,0,2), ARIMA(2,1,2) จากนั้นทำการทดสอบความเป็นอิสระกันของคลาตเคลื่อนด้วยสถิติ Ljung-Box และคัดเลือกตัวแบบที่เหมาะสมที่สุดด้วยเกณฑ์ค่า AIC ต่ำสุด ได้ตัวแบบที่เหมาะสมที่สุดดัง Table 1

จาก Table 1 ตัวแบบบอซ-เจนกินส์ที่เหมาะสมในการพยากรณ์ปริมาณมลพิษทางอากาศ PM2.5 โดยเฉลี่ยรายเดือนของจังหวัดนครราชสีมา คือ ตัวแบบ ARIMA(3,0,2)

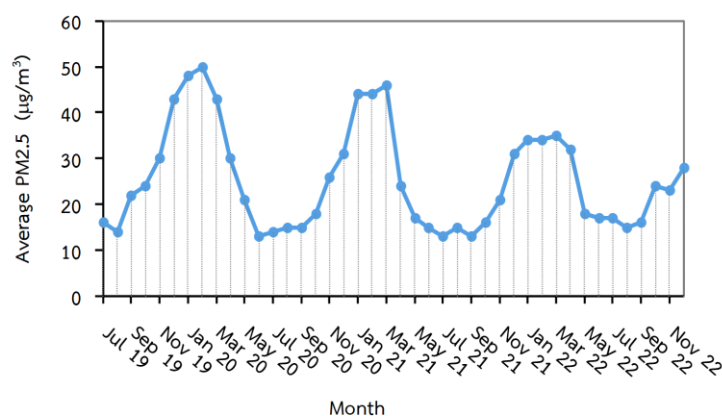


Figure 3 The monthly average PM2.5 air pollution levels in Nakhon Ratchasima province

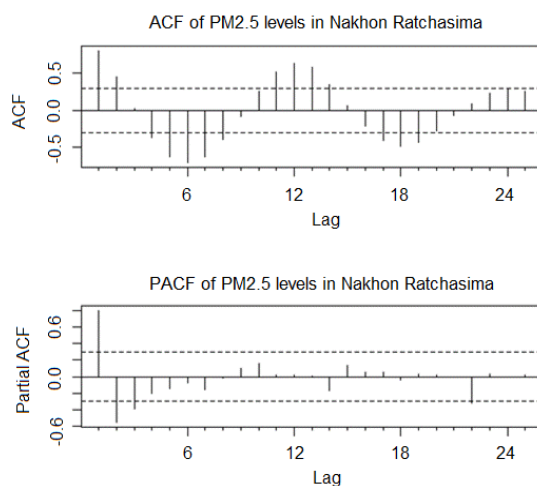


Figure 4 ACF and PACF of the raw data

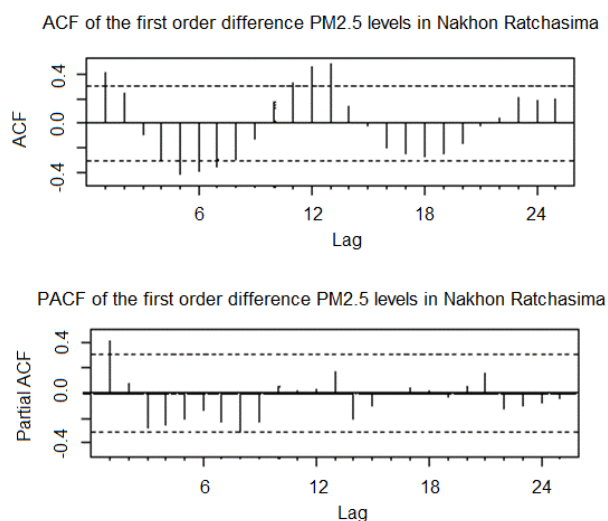


Figure 5 ACF and PACF of the first-order difference data

Table 1 The significance testing of ARIMA(3,0,2)

Parameter	Estimate	Z-test (p-value)	Ljung-box (p-value)	AIC
AR(1)	2.59008	23.7941 ($<0.001^{**}$)	6.0811 (0.9119 ^{ns})	257.568
AR(2)	-2.45801	-12.8519 ($<0.001^{**}$)		
AR(3)	0.83688	7.4721 ($<0.001^{**}$)		
MA(1)	-1.90603	-16.1859 ($<0.001^{**}$)		
MA(2)	0.99998	8.5650 ($<0.001^{**}$)		

AIC = Akaike information criterion

** = Significantly different at $p < 0.01$; ns = Not significant

3.2. ผลการสร้างตัวแบบการถดถอยเชิงเส้นพหุคูณ

ตัวแบบการถดถอยเชิงเส้นพหุคูณใช้ตัวแปรอิสระจาก Table 1 เข้าสู่สมการถดถอย แต่เพื่อป้องกันตัวแปรอิสระที่มีความสัมพันธ์กับตัวแปรอิสระในสมการถดถอย จึงคัดกรองตัวแปรอิสระที่มีความสัมพันธ์กันสูงจนทำให้เกิดปัญหา Multicollinearity เพื่อคัดออกจากสมการด้วยวิธีขั้นบันได (Stepwise regression) โดยผู้วิจัยใช้เกณฑ์ค่า Variance inflation factor (VIF) ไม่เกิน 10 [12] ซึ่งผลการประมาณค่าสัมประสิทธิ์การถดถอย และผลการทดสอบนัยสำคัญทางสถิติของสมการถดถอยที่ได้ ดัง Table 2

จาก Table 2 จะได้สมการการถดถอยดังสมการ (5)

$$\hat{Y} = 19.489 + 0.512PM10_mean - 0.228RH \quad (5)$$

เมื่อ

$\hat{Y}$ คือ ค่าพยากรณ์ปริมาณฝุ่น PM_{2.5} เฉลี่ยรายเดือน

PM10_mean คือ ปริมาณฝุ่น PM₁₀ เฉลี่ยรายเดือน

RH คือ ความชื้นสัมพัทธ์เฉลี่ยรายเดือน

สมการถดถอยพหุคูณที่ได้ดังสมการ (5) แสดงให้เห็นว่าตัวแปรปริมาณฝุ่นละอองขนาดไม่เกิน 10 ไมครอนเฉลี่ยรายเดือน และตัวแปรความชื้นสัมพัทธ์เฉลี่ยรายเดือน สามารถอธิบายความผันแปรของปริมาณ PM_{2.5} โดยเฉลี่ยรายเดือนของจังหวัดนครราชสีมาได้ร้อยละ 96.9

Table 2 The results of the stepwise regression model

Variables	Regression coefficient		t-test (p-value)	VIF
	b	Std. Error		
constant	19.489	4.714	4.134 (0.000**)	
PM10_mean	0.512	0.023	22.326 (0.000**)	1.912
RH	-0.228	0.054	-4.246 (0.000**)	1.912
R ²	0.969			

VIF = Variance inflation factor; R² = Coefficient of determination

** = Significantly different at p<0.01

3.3. ผลการสร้างตัวแบบโครงข่ายประสาทเทียม

ผลการสร้างตัวแบบโครงข่ายประสาทเทียม เมื่อผู้วิจัยได้กำหนดโครงข่ายไม่เกิน 2 ชั้น เนื่องจากข้อมูลมีจำกัด และส่งผลต่อการพยากรณ์ แบ่งเป็นสองกรณี คือ 1) กรณีการสร้างตัวแบบด้วยตัวแปรอิสระที่มีความสัมพันธ์กับตัวแปรตามทั้ง 16 ตัวแปร เป็นตัวแปรนำเข้า กำหนดชั้นซ่อน 1 ชั้นซ่อน และจำนวนโหนดในชั้นซ่อน เท่ากับ 3, 7 และ 11 โหนด และ 2) กรณีการสร้างตัวแบบด้วยตัวแปรอิสระ 2 ตัวแปรที่ได้จากสมการการถดถอย (5) กำหนดชั้นซ่อน 1 ชั้นซ่อน และจำนวนโหนดในชั้นซ่อน เท่ากับ 2 โหนด

การกำหนดค่าพารามิเตอร์ในการวิเคราะห์โครงข่ายประสาทเทียมด้วยกระบวนการแพร่ย้อนกลับ ทั้ง 2 กรณี คือ ค่าอัตราการเรียนรู้ (Learning Rate: η) และค่าโมเมนตัม

(Momentum: β) เท่ากับ 0.1 - 0.9 เพิ่มค่าทีละ 0.1 จำนวนการเรียนรู้ เท่ากับ 500, 1,000 และ 1,500 รอบ ซึ่งคัดเลือกตัวแบบโครงข่ายประสาทเทียม 16-3-1, 16-7-1, 16-11-1 และ 2-2-1 ที่เหมาะสมด้วยเกณฑ์ค่า RMSE ต่ำสุด นำมาเปรียบเทียบประสิทธิภาพของตัวแบบโครงข่ายประสาทเทียมได้ดัง Table 3

จาก Table 3 พบว่า ตัวแบบโครงข่ายประสาทเทียมใช้ตัวแปรอิสระทั้ง 16 ตัวแปร มีชั้นซ่อน 1 ชั้น และจำนวนโหนดในชั้นซ่อน เท่ากับ 7 โหนด มีค่าอัตราการเรียนรู้ เท่ากับ 0.1 ค่าโมเมนตัม เท่ากับ 0.1 และจำนวนการเรียนรู้ 500 รอบ มีค่า RMSE ต่ำสุดเท่ากับ 2.560 และค่าสัมประสิทธิ์การตัดสินใจ (Coefficient of determination: R²) เท่ากับ 0.988 สามารถแสดงโครงข่ายประสาทเทียมได้ดัง Figure 6

Table 3 The results of backpropagation in artificial neural networks

Model	Learning rate	Momentum	Epoch	RMSE	R ²
16-3-1	0.1	0.1	500	3.157	0.990
			1,000	3.364	0.993
			1,500	2.955	0.995
16-7-1	0.1	0.1	500	2.560	0.988
			1,000	2.961	0.993
			1,500	3.204	0.995
16-11-1	0.1	0.2	500	3.370	0.991
			1,000	3.747	0.996
			1,500	3.876	0.997
2-2-1	0.3	0.3	500	3.782	0.977
			1,000	3.800	0.977
			1,500	3.803	0.977

RMSE = Root mean square error; R² = Coefficient of determination

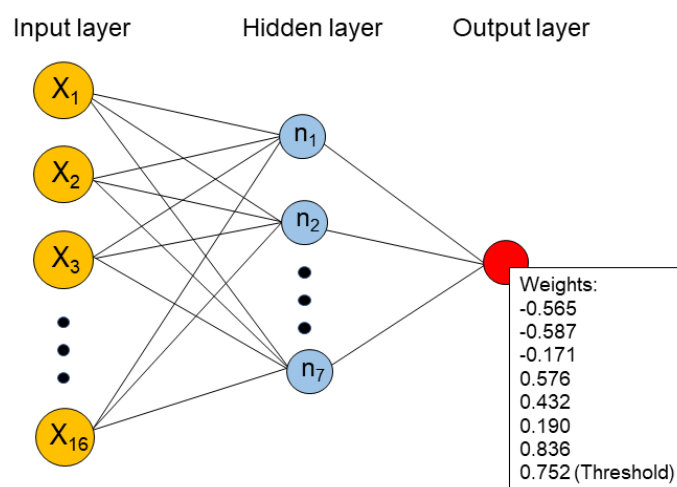


Figure 6 The 16-7-1 artificial neural network (ANN) model

3.4. ผลการเปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์

การเปรียบเทียบประสิทธิภาพของตัวแบบด้วยวิธีบอกซ์-เจนกินส์ การถดถอยเชิงเส้นพหุคูณ และโครงข่ายประสาทเทียม โดยใช้ข้อมูลชุดทดสอบตั้งแต่เดือนมกราคม พ.ศ. 2566 ถึง ธันวาคม พ.ศ. 2566 จำนวน 12 คาบเวลา ผลการพยากรณ์ข้อมูลชุดทดสอบดัง Table 4

จาก Table 4 พบว่า ตัวแบบที่เหมาะสมในพยากรณ์ปริมาณมลพิษทางอากาศ PM2.5 รายเดือนของจังหวัดนครราชสีมา คือ ตัวโครงข่ายประสาทเทียม 16-7-1 ที่มีค่า RMSE ต่ำสุด Figure 7 แสดงการเปรียบเทียบประสิทธิภาพของการพยากรณ์ด้วยตัวแบบด้วยวิธีบอกซ์-เจนกินส์ การถดถอยเชิงเส้นพหุคูณ และโครงข่ายประสาทเทียม

Table 4 The performance comparison of the three forecasting models

Month in Year 2023	Actual value (Y_t)	Forecasting value		
		Box-Jenkins	MLR	ANN
January	32	30.4699	29.9850	36.3676
February	41	31.4234	37.3250	45.2399
March	48	30.7436	43.6410	46.6054
April	39	28.7063	35.4490	40.5725
May	28	25.8985	27.7090	28.6858
June	18	23.0646	18.6050	19.6103
July	17	20.9215	16.3850	17.2577
August	19	19.9864	19.4010	17.0441
September	14	20.4608	14.6730	16.1959
October	18	22.1941	18.9410	14.7728
November	25	24.7353	18.7450	21.2430
December	28	27.4534	21.4650	27.0602
RMSE		7.0941	3.3334	2.560

For Box-Jenkins, the values were predicted from ARIMA(3,0,2).

For multiple linear regression (MLR), the values were predicted from $\hat{Y} = 19.489 + 0.512PM10_mean - 0.228RH$.

For artificial neural networks (ANN), the values were predicted from 16-7-1 ANN model.

RMSE = Root mean square error

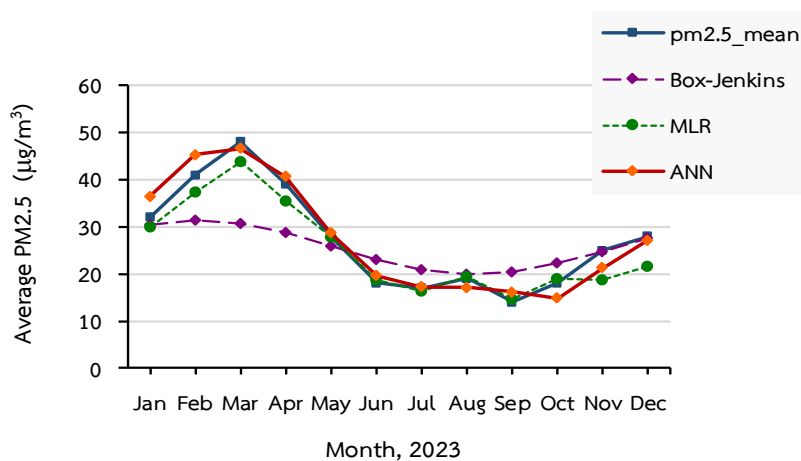


Figure 7 The performance comparison of the three forecasting models

4. อภิปรายผลการวิจัย

การสร้างตัวแบบพยากรณ์ปริมาณมลพิษทางอากาศ PM2.5 รายเดือนของจังหวัดนครราชสีมา ด้วยวิธีบ็อกซ์-เจนกินส์ เริ่มจากการศึกษาการเคลื่อนไหวของข้อมูล พบว่า ปริมาณฝุ่น PM2.5 มีปริมาณลดต่ำลงในช่วงเดือนมิถุนายนถึงเดือนกันยายน เนื่องจากอยู่ในช่วงฤดูฝน และเริ่มมีค่าเพิ่มมากขึ้นในช่วงปลาย

เดือนพฤศจิกายนจนถึงต้นเดือนเมษายน เนื่องจากอยู่ในช่วงฤดูแล้งเข้าสู่ฤดูร้อน จึงมีปริมาณเพิ่มสูงขึ้นเปรียบเสมือนปลายฤดูหนาว ซึ่งผลการทดสอบกระบวนการคงที่ของข้อมูลเดิม และข้อมูลที่แปลงค่าด้วยผลต่างอันดับ 1 ล้วนเป็นข้อมูลที่มีกระบวนการคงที่ สอดคล้องกับงานวิจัยของ Sukkasem [4] ที่พบว่า ข้อมูลปริมาณ PM2.5 รายชั่วโมงในจังหวัดนครราชสีมา

จากปี พ.ศ. 2562 ถึง พ.ศ. 2563 เมื่อแปลงข้อมูลผลต่างอันดับที่ 1 จะมีลักษณะกระบวนการคงที่ ทั้งนี้เนื่องจากปริมาณฝุ่น PM2.5 ได้รับอิทธิพลจากฤดูกาล โดยเฉพาะฤดูฝนที่มีความชื้นสัมพัทธ์สูง จะช่วยลดปริมาณ PM2.5 ซึ่งข้อเท็จจริงนี้ก็สอดคล้องกับผลพยากรณ์ด้วยสมการการวิเคราะห์การถดถอยในครั้งนี้ จากการคัดเลือกตัวแปรอิสระ หรือคัดเลือกคุณลักษณะที่ส่งผลต่อปริมาณมลพิษทางอากาศ PM2.5 ได้แก่ ปริมาณฝุ่น PM10 เฉลี่ยรายเดือน และความชื้นสัมพัทธ์โดยเฉลี่ย ซึ่งสอดคล้องกับรายงานของกรมควบคุมมลพิษ [17] ที่พบว่า การเกิดฝุ่น PM2.5 และฝุ่น PM10 มีสาเหตุจากการเผาไหม้ทั้งจากยานพาหนะ การเผาวัสดุการเกษตร ไฟป่า และกระบวนการอุตสาหกรรม รวมทั้งช่วงที่สภาพอากาศนิ่ง ลมสงบ ทั้งหมดล้วนก่อให้เกิดมลพิษทางอากาศทั้งสิ้น ดังนั้นช่วงเวลาที่ปริมาณฝุ่น PM10 เพิ่มขึ้น และความชื้นสัมพัทธ์โดยเฉลี่ยลดลง จะส่งผลต่อปริมาณฝุ่น PM2.5 เพิ่มขึ้น ซึ่งจะเกิดขึ้นในช่วงฤดูหนาวจนถึงฤดูร้อน หรือปริมาณฝุ่น PM10 ลดลง และความชื้นสัมพัทธ์โดยเฉลี่ยเพิ่มขึ้น จะส่งผลต่อปริมาณฝุ่น PM2.5 ลดลง ซึ่งสถานการณ์จะอยู่ในระหว่างฤดูฝน

ผลการวิจัย พบว่า ตัวแบบพยากรณ์ปริมาณฝุ่น PM2.5 รายเดือนของจังหวัดนครราชสีมาที่เหมาะสม คือ ตัวแบบโครงข่ายประสาทเทียม ใช้ตัวแปรอิสระทั้ง 16 ตัวแปร มีชั้นซ่อน 1 ชั้น และจำนวนโหนดในชั้นซ่อน เท่ากับ 7 โหนด มีค่าอัตราการเรียนรู้เท่ากับ 0.1 ค่าโมเมนตัมเท่ากับ 0.1 และจำนวนการเรียนรู้ 500 รอบ นอกจากนี้มีค่า RMSE เท่ากับ 2.560 และค่าสัมประสิทธิ์การตัดสินใจ (R^2) เท่ากับ 0.988 ซึ่งสอดคล้องกับงานวิจัยของ Sobri et al. [10] ที่พบว่า โครงข่ายประสาทเทียมมีความเหมาะสมในการพยากรณ์ปริมาณฝุ่น PM2.5 ได้ดีกว่าการถดถอยเชิงเส้นพหุคูณ ทั้งนี้เนื่องจากตัวแปรนำเข้ามีความสัมพันธ์อย่างมีนัยสำคัญกับปริมาณฝุ่น PM2.5 และโครงข่ายประสาทเทียมแบบแพร่ย้อนกลับมีกระบวนการปรับค่าน้ำหนักและค่าความเอนเอียงของตัวแบบเพื่อลดค่าผิดพลาดในการพยากรณ์

5. บทสรุป

จากการศึกษานี้สามารถสรุปได้ว่าตัวแบบโครงข่ายประสาทเทียมมีความเหมาะสมในการพยากรณ์ปริมาณฝุ่น PM2.5 รายเดือนของจังหวัดนครราชสีมา ในการศึกษาครั้งต่อไปอาจนำตัวแบบโครงข่ายประสาทเทียมไปประยุกต์เพื่อพยากรณ์ระดับปริมาณฝุ่น PM2.5 ของจังหวัดนครราชสีมาแบบรายวันหรือรายชั่วโมง หรืออาจเพิ่มมิติตัวแปรที่เกี่ยวกับ

จำนวนผู้ป่วยที่เป็นโรคสัมพัทธ์กับปริมาณฝุ่น PM2.5 เข้าร่วมในการศึกษารั้งต่อไป เพื่อให้เป็นตัวแบบที่มีประสิทธิภาพมากยิ่งขึ้น สามารถนำไปประยุกต์ในการพัฒนาแอปพลิเคชันเพื่อการแจ้งเตือนระดับปริมาณฝุ่น PM2.5 ของจังหวัดนครราชสีมา หรือพยากรณ์โอกาสการเกิดโรคอื่นเนื่องจากการได้รับมลพิษทางอากาศต่อไปได้

6. กิตติกรรมประกาศ

ผู้วิจัยขอขอบคุณสถาบันวิจัยและพัฒนา มหาวิทยาลัยราชภัฏสุรินทร์ที่ให้ทุนสนับสนุนการวิจัย และขอขอบคุณกรมควบคุมมลพิษ และศูนย์อุตุนิยมวิทยาภาคตะวันออกเฉียงเหนือตอนล่างที่ให้ความอนุเคราะห์ข้อมูลสำหรับการศึกษาในครั้งนี้

7. References

- [1] Pochanart, P. 2016. The present state of urban air pollution problems in Thailand's large cities: cases of Bangkok, Chiang mai, and Rayong. *Journal of Environmental Management*. 12(1):114-33. (in Thai)
- [2] Department of Disease Control. 2022. *Disease and Health Hazard Weekly Forecast No. 9/2022*. https://ddc.moph.go.th/odpc9/forecast_detail.php?publish=12452. Accessed 15 January 2023. (in Thai)
- [3] Thairathonline. 2023. *Korat Sky Clears, PM2.5 Dust Levels Decrease, Mainly Caused by Forest Fires in the National Park*. <https://www.thairath.co.th/news/local/northeast/2627065>. Accessed 20 February 2023. (in Thai)
- [4] Sukkasem, I. 2020. Forecasting the amount of particulate matter having a diameter of less than 2.5 micrometers (PM2.5) in Nakhon Ratchasima province. *The Office of Disease Prevention and Control 9 Nakhon Rathasima Journal*. 27(1): 16-25. (in Thai)
- [5] Sangsri, P. and Chaiwuttisak, P. 2021. Hybrid of neural network and Markov chain method for predicting PM2.5 concentrations. *Thai Science and Technology Journal*. 29(5): 752-765. (in Thai)

- [6] Amnuaylojaroen, T. 2022. Prediction of PM2.5 in an urban area of northern Thailand using multivariate linear regression model. **Advances in Meteorology**. 2022: 3190484.
- [7] Kumharn, W. and et al. 2022. Improved hourly and long-term PM2.5 prediction modeling based on MODIS in Bangkok. **Remote Sensing Applications: Society and Environment**. 28: 100864.
- [8] Saiohai, J. and et al. 2023. Statistical PM2.5 prediction in an urban area using vertical meteorological factors. **Atmosphere**. 14(3): 589.
- [9] Masood, A. and Ahmad, K. 2020. A model for particulate matter (PM2. 5) prediction for Delhi based on machine learning approaches. **Procedia Computer Science**. 167: 2101-2110.
- [10] Sobri, M.N. and et al. 2021. Predicting particulate matter (PM2.5) in Malaysia using multiple linear regression and artificial neural network. **Journal of Physics: Conference Series**. 2084: 012010.
- [11] Hyndman, R.J. and Athanasopoulos, G. 2021. **Forecasting: Principles and Practice**. 3rd edition. Melbourne: OTexts.
- [12] Hair, J.F. and et al. 2009. **Multivariate Data Analysis**. 7th edition. New York: Pearson.
- [13] Han, J., Kamber, M. and Pei, J. 2012. **Data Mining: Concepts and Techniques**. Amsterdam: Elsevier.
- [14] Pasilatetsang, B. 2021. **Creating Learning for AI with Python Machine Learning**. Bangkok: V. Print (1991) Co., Ltd. (*in Thai*)
- [15] Pollution Control Department. 2022. **Historical Air Quality Data**. <http://air4thai.pcd.go.th/webV3/#/History>. Accessed 1 October 2022. (*in Thai*)
- [16] Northeastern Meteorological Center (Lower Part). 2022. **Request Online Data**. <http://www.ubonmet.tmd.go.th/>. Accessed 1 October 2022. (*in Thai*)
- [17] Pollution Control Department. 2023. **Air Quality Index**. <http://air4thai.pcd.go.th/webV3/#/AQIInfo>. Accessed 15 October 2023. (*in Thai*)