

ผลของรูปแบบและอัตราการสูญหายของข้อมูลต่อความแม่นยำของแบบจำลอง ARIMA สำหรับการพยากรณ์การใช้พลังงานไฟฟ้า Effects of Missing Data Patterns and Rates on the Accuracy of ARIMA Model for Electricity Consumption Forecasting

นัสสมล บุตรวิเศษ* และ อุปริษฐา อินทรสาด

Nassamon Bootwisas* and Uparitttha Intarasat

สาขาวิชาวิทยาการคำนวณ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ

จ.พระนครศรีอยุธยา ประเทศไทย

Department of Computation Science, Faculty of Science and Technology,

Rajamangala University of Technology Suvarnabhumi, Phra Nakhon Si Ayutthaya, Thailand

วันที่ส่งบทความ : 20 พฤศจิกายน 2568 วันที่แก้ไขบทความ : 14 พฤษภาคม 2569 วันที่ตอบรับบทความ : 17 พฤษภาคม 2569

Received: 20 November 2025, Revised: 14 May 2026, Accepted: 17 May 2026

บทคัดย่อ

การวิจัยนี้มุ่งศึกษาผลของรูปแบบและอัตราการสูญหายของข้อมูลที่มีต่อความแม่นยำของแบบจำลอง ARIMA ในเชิงระเบียบวิธี โดยอาศัยข้อมูลการใช้พลังงานไฟฟ้ารายเดือนของประเทศไทยช่วงปี ค.ศ. 2002–2025 งานวิจัยนี้ใช้การจำลองแบบแฟกทอเรียลเต็ม (Full factorial design) เพื่อจำลองสถานการณ์ข้อมูลสูญหายตามรูปแบบ Missing Completely at Random (MCAR) และ Missing at Random (MAR) ที่อัตราร้อยละ 10, 20 และ 30 โดยทำการทดลองซ้ำเงื่อนไขละ 50 ครั้ง และดำเนินการจัดการข้อมูลด้วยที่สูญหายวิธีการทดแทนข้อมูล 3 วิธี ที่ใช้กันอย่างแพร่หลายในงานพยากรณ์อนุกรมเวลา ได้แก่ วิธีแทนค่าด้วยค่าก่อนหน้า (Last Observation Carried Forward: LOCF) การประมาณค่าเชิงเส้น (Linear interpolation) และตัวกรองคาลมาน (Kalman filtering) ผลการวิจัยพบว่า วิธีการทดแทนข้อมูลเป็นปัจจัยที่มีอิทธิพลต่อความแม่นยำของการพยากรณ์มากที่สุด รองลงมาคืออัตราการสูญหายของข้อมูล โดยวิธี Kalman filtering ให้ค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Squared Error: RMSE) ต่ำที่สุดในภาพรวมและมีความเสถียรสูง ในขณะที่วิธี Linear interpolation ให้ค่า RMSE สูงที่สุดอย่างสม่ำเสมอ นอกจากนี้ยังพบว่า ประสิทธิภาพของบางวิธี โดยเฉพาะ LOCF มีความแปรผันตามสัดส่วนข้อมูลที่สูญหายอย่างชัดเจน ผลลัพธ์ดังกล่าวชี้ให้เห็นว่า Kalman filtering เป็นทางเลือกที่เหมาะสมและมีความทนทานสูง สำหรับการจัดการข้อมูลอนุกรมเวลาด้านพลังงานที่ไม่สมบูรณ์ ซึ่งสามารถสนับสนุนการวางแผนการใช้พลังงานระยะยาวภายใต้ความไม่แน่นอนของข้อมูลได้อย่างมีประสิทธิภาพ

คำสำคัญ : การทดแทนข้อมูล แบบจำลอง ARIMA Kalman filtering รูปแบบการสูญหายของข้อมูล การพยากรณ์การใช้พลังงานไฟฟ้า

*ที่อยู่ติดต่อ E-mail address: nassamon.b@rmutsb.ac.th

<https://doi.org/10.55003/scikmitl.2026.269931>

Abstract

This study aims to investigate the effects of missing data patterns and missing rates on the forecasting accuracy of the ARIMA model from a methodological perspective using monthly electricity consumption data in Thailand from 2002 to 2025. A full factorial design was employed to simulate missing data under the Missing Completely at Random (MCAR) and Missing at Random (MAR) mechanisms at missing rates of 10%, 20%, and 30%, with 50 replications conducted for each experimental condition. Three widely used imputation methods in time series forecasting were applied, namely Last Observation Carried Forward (LOCF), Linear interpolation, and Kalman filtering. The results reveal that the imputation method is the most influential factor affecting forecasting accuracy, followed by the missing data rate. Kalman filtering consistently produced the lowest Root Mean Squared Error (RMSE) and demonstrated high stability across experimental conditions, whereas Linear interpolation consistently yielded the highest RMSE values. In addition, the performance of certain methods, particularly LOCF, varied substantially according to the proportion of missing data. These findings suggest that Kalman filtering is a robust and appropriate approach for handling incomplete energy time series data and can effectively support long-term energy consumption planning under data uncertainty.

Keyword: Data imputation, ARIMA Model, Kalman filtering, Missing data patterns, Electricity consumption forecasting

1. บทนำ

การพยากรณ์ข้อมูลอนุกรมเวลามีบทบาทสำคัญอย่างยิ่งต่อการวางแผนเชิงยุทธศาสตร์ในหลายภาคส่วน ทั้งด้านเศรษฐกิจ พลังงาน และสาธารณสุข โดยเฉพาะอย่างยิ่งในบริบทของ “พลังงานไฟฟ้า” ซึ่งข้อมูลการใช้ไฟฟ้ามีความจำเป็นต่อการบริหารจัดการโครงข่าย การจัดสรรกำลังการผลิต และการประเมินความต้องการในอนาคต งานวิจัยจำนวนมากยืนยันว่าความแม่นยำของการพยากรณ์ส่งผลโดยตรงต่อความมั่นคงของระบบและต้นทุนการดำเนินงาน ดังที่ปรากฏในงานศึกษาของ Hussain et al. (2025), Duarte et al. (2024) และ Wongoutong (2021) ซึ่งชี้ให้เห็นว่าความคลาดเคลื่อนเพียงเล็กน้อยในแบบจำลองอาจนำไปสู่ต้นทุนด้านกำลังการผลิตสำรองที่เพิ่มขึ้นอย่างมีนัยสำคัญ

ในกลุ่มเครื่องมือทางสถิติสำหรับการพยากรณ์ แบบจำลอง ARIMA ได้รับความนิยมอย่างแพร่หลายเนื่องจากความสามารถในการถอดรหัสโครงสร้างข้อมูลตามลำดับเวลาได้อย่างมีประสิทธิภาพ ทั้งในด้านแนวโน้มและความผันผวน (Hussain et al., 2025) อย่างไรก็ตาม ประสิทธิภาพของ AutoRegressive Integrated Moving Average (ARIMA) ขึ้นอยู่กับความสมบูรณ์ของข้อมูลเป็นสำคัญ ซึ่งเป็นความท้าทายหลักในข้อมูลพลังงานที่มักประสบปัญหาข้อมูลสูญหาย (Missing data) อันเนื่องมาจากความขัดข้องของ

อุปกรณ์ตรวจวัดหรือความไม่ต่อเนื่องของการรายงานข้อมูล (Chhabra, 2023; Wongoutong, 2021) ปัญหาข้อมูลสูญหายมีความซับซ้อนขึ้นเมื่อพิจารณาถึงรูปแบบการสูญหาย (Missing pattern) โดยเฉพาะรูปแบบการสูญหายแบบสุ่มสมบูรณ์ (Missing Completely at Random: MCAR) และการสูญหายแบบสุ่ม (Missing at Random: MAR)

ผลกระทบของรูปแบบการสูญหายมีความสำคัญอย่างยิ่งต่อขั้นตอนการเตรียมข้อมูล (Data preparation) การเลือกใช้วิธีการทดแทนข้อมูล (Imputation) ที่ไม่สอดคล้องกับกลไกการสูญหาย อาจก่อให้เกิดความเอนเอียงและลดทอนความแม่นยำของการพยากรณ์ได้ (Chhabra, 2023; Moritz & Bartz-Beielstein, 2017) ในทางปฏิบัติมีแนวทางการทดแทนข้อมูลที่หลากหลาย ตั้งแต่วิธีพื้นฐาน เช่น การแทนค่าด้วยค่าก่อนหน้า (Last Observation Carried Forward: LOCF) และการประมาณค่าเชิงเส้น (Linear interpolation) ไปจนถึงวิธีที่มีความซับซ้อนอย่างตัวกรองคาลมาน (Kalman filtering) ซึ่งมีจุดเด่นในการรักษาโครงสร้างพลวัตของข้อมูล โดยงานวิจัยของ Fung (2006) และ Moritz & Bartz-Beielstein (2017) และพบว่า Kalman filtering มีความเหมาะสมกับข้อมูลที่มีฤดูกาล ในขณะที่วิธีเชิงเส้นมักมีข้อจำกัดเมื่อข้อมูลมีความผันผวนสูง

จากการทบทวนวรรณกรรมที่เกี่ยวข้อง พบว่า งานวิจัยส่วนใหญ่ที่ศึกษาผลกระทบของข้อมูลสูญหายยังขาดการเปรียบเทียบเชิงระบบที่ครอบคลุมปัจจัยร่วมต่าง ๆ งานวิจัยจำนวนหนึ่งมักประเมินผลภายใต้ระดับอัตราการสูญหาย (Missing rate) เพียงระดับเดียว หรือเปรียบเทียบวิธีการทดแทนโดยไม่ได้อแยกผลกระทบของรูปแบบการสูญหาย (MCAR และ MAR) อย่างชัดเจน ทำให้ข้อสรุปที่ได้อาจไม่ครอบคลุมเงื่อนไขที่ซับซ้อนในสถานการณ์จริง นอกจากนี้ ในบริบทของประเทศไทย งานวิจัยด้านการพยากรณ์พลังงานส่วนใหญ่ มักมุ่งเน้นที่การเปรียบเทียบประสิทธิภาพของตัวแบบพยากรณ์ แต่ยังขาดงานวิจัยที่ประเมินผลกระทบของกระบวนการเตรียมข้อมูลต่อความแม่นยำของแบบจำลอง ARIMA โดยตรง โดยเฉพาะการใช้ข้อมูลอนุกรมเวลารายเดือนที่มีความยาวต่อเนื่องมากกว่า 20 ปี ซึ่งมีความสำคัญอย่างยิ่งต่อการจับสัญญาณวัฏจักรการดำเนินงานในระยะยาว (Chhabra, 2023; Moritz & Bartz-Beielstein, 2017; Wongoutong, 2021)

แม้ว่าการพยากรณ์ข้อมูลอนุกรมเวลาที่มีฤดูกาลสามารถประยุกต์ใช้แบบจำลองเชิงฤดูกาล เช่น SARIMA ได้ อย่างไรก็ตาม งานวิจัยนี้เลือกใช้แบบจำลอง ARIMA เป็นกรณีศึกษา เนื่องจากมีวัตถุประสงค์หลักเพื่อมุ่งเน้นการประเมินผลกระทบของกระบวนการจัดการข้อมูลสูญหาย (Missing data handling) ต่อความแม่นยำของการพยากรณ์ มากกว่าการเปรียบเทียบประสิทธิภาพระหว่างตัวแบบพยากรณ์หลายประเภท

ในบริบทนี้แบบจำลอง ARIMA ถูกนำมาใช้ในฐานะตัวแบบอ้างอิงมาตรฐาน (Baseline forecasting model) ที่ได้รับการยอมรับอย่างกว้างขวางในงานพยากรณ์พลังงานและมีโครงสร้างเชิงสถิติที่ชัดเจน ซึ่งเอื้อต่อการตรวจสอบผลกระทบของการทดแทนข้อมูลต่อขั้นตอนการประมาณค่าและการพยากรณ์ได้อย่างเป็นระบบ การใช้ ARIMA ช่วยลดความซับซ้อนขององค์ประกอบของตัวแบบ ทำให้สามารถแยกอิทธิพลของรูปแบบการสูญหาย อัตราการสูญหาย และวิธีการทดแทนข้อมูลออกจากผลของโครงสร้างตัวแบบได้อย่างชัดเจน

การเลือกใช้ ARIMA ยังสอดคล้องกับวัตถุประสงค์เชิงระเบียบวิธี (Methodological contribution) ของงานวิจัยนี้ ที่มุ่งเน้นนำเสนอข้อค้นพบเชิงโครงสร้างเกี่ยวกับอคติและความเสถียรของวิธีการทดแทนข้อมูลภายใต้เงื่อนไขข้อมูลสูญหายที่ต่างกัน โดยเฉพาะในกรณีข้อมูลรายเดือนที่มีความยาวต่อเนื่องระยะยาว การศึกษานี้จึงมีได้มุ่งสรุปว่า ARIMA เป็นตัวแบบที่ดีที่สุดสำหรับการพยากรณ์พลังงานไฟฟ้า หากแต่ใช้เป็นกรอบมาตรฐานเพื่อสร้างข้อสรุปที่สามารถนำไปขยายผลสู่ตัวแบบพยากรณ์อื่น ๆ ได้ในอนาคต รวมถึงไปใช้เป็นแนวทางในการออกแบบกระบวนการเตรียมข้อมูลสำหรับการพยากรณ์พลังงานในระดับนโยบายและการวางแผนระบบไฟฟ้าในระยะยาว

เพื่อเพิ่มเติมช่องว่างทางวิชาการดังกล่าว งานวิจัยนี้จึงมุ่งศึกษาผลกระทบร่วมของ 3 ปัจจัยหลัก ได้แก่ รูปแบบการสูญหาย (MCAR และ MAR) อัตราการสูญหาย (ร้อยละ 10, 20 และ 30) และวิธีการทดแทนข้อมูล (LOCF, Linear Interpolation และ Kalman Filtering) ที่มีต่อความแม่นยำของแบบจำลอง ARIMA การศึกษานี้ใช้ข้อมูลปริมาณการใช้พลังงานไฟฟ้ารายเดือนของประเทศไทยในช่วงปี ค.ศ. 2002–2025 ผ่านการออกแบบการทดลองแบบแฟกทอเรียลเต็ม (Full factorial design) และวิเคราะห์ผลด้วยสถิติวิเคราะห์ความแปรปรวนแบบสามทาง (Three-way ANOVA) เพื่อให้ได้ข้อสรุปเชิงประจักษ์สำหรับการเลือกวิธีการจัดการข้อมูลสูญหายที่เหมาะสมที่สุดสำหรับข้อมูลอนุกรมเวลาด้านพลังงาน

2. วิธีดำเนินการวิจัย

การวิจัยนี้มุ่งศึกษาผลกระทบของปัจจัยด้านรูปแบบการสูญหาย อัตราการสูญหาย และวิธีการทดแทนข้อมูลที่มีต่อความแม่นยำของแบบจำลอง ARIMA โดยมีกรอบแนวคิดและขั้นตอนการดำเนินงานดังนี้

2.1 ทฤษฎีและแนวคิดที่เกี่ยวข้อง

1) ทฤษฎีอนุกรมเวลา (Time series theory) อนุกรมเวลาเป็นข้อมูลที่ถูกบันทึกตามลำดับเวลาอย่างต่อเนื่อง และมักประกอบด้วยองค์ประกอบสำคัญ ได้แก่ แนวโน้ม (Trend) ฤดูกาล (Seasonality) และส่วนสุ่ม (Random component) (Box et al., 2015; Hyndman & Athanasopoulos, 2021) การพยากรณ์อนุกรมเวลาที่มีประสิทธิภาพจำเป็นต้องสร้างแบบจำลองที่สามารถอธิบายและแยกแยะองค์ประกอบเหล่านี้ได้อย่างเหมาะสม นอกจากนี้ การวิเคราะห์ด้วยแบบจำลองเชิงสถิติส่วนใหญ่รวมถึง ARIMA จำเป็นต้องมีการตรวจสอบคุณสมบัติความนิ่ง (Stationarity) ของข้อมูล กล่าวคือ ข้อมูลต้องมีค่าเฉลี่ยและความแปรปรวนคงที่ตามเวลา (Box et al., 2015)

2) แบบจำลอง ARIMA(p, d, q) เป็นแบบจำลองที่อาศัยความสัมพันธ์ของข้อมูลในอดีตเพื่อพยากรณ์ค่าในอนาคต ตามแนวคิดของ Box-Jenkins (Box et al., 2015) ประกอบด้วย 3 ส่วนหลัก คือ (1) Autoregressive (p) ซึ่งเป็นส่วนที่อธิบายความสัมพันธ์ระหว่างค่าปัจจุบันกับค่าในอดีต (2) Integrated (d) หรือลำดับการหาผลต่างเพื่อปรับข้อมูลให้มีความนิ่ง และ (3) Moving Average (q) ซึ่งอธิบายความสัมพันธ์ของค่าความคลาดเคลื่อน ความสัมพันธ์ดังกล่าวสามารถเขียนในรูปสมการคณิตศาสตร์ได้ดังสมการที่ (1)

$$\phi_p(B)(1-B)^d Y_t = c + \theta_q(B)\varepsilon_t \quad (1)$$

โดยที่	Y_t	คือ ค่าสังเกต ณ เวลา t
	ε_t	คือ ค่าความคลาดเคลื่อนแบบสุ่ม (White noise)
	B	คือ ตัวดำเนินการย้อนหลัง (Backshift operator)
	$\phi_p(B)$	คือ พหุนามของส่วน Autoregressive อันดับที่ p
	$\theta_q(B)$	คือ พหุนามของส่วน Moving average อันดับที่ q
	d	คือ ลำดับของการหาผลต่าง (Differencing)
	c	คือ ค่าคงที่

การพัฒนาแบบจำลองตามกรอบแนวคิดนี้ประกอบด้วยขั้นตอนการระบุรูปแบบ (Identification) การประมาณค่าพารามิเตอร์ (Estimation) และการตรวจสอบความเหมาะสม (Diagnostic checking) โดยพิจารณาจากเศษเหลือ (Residual) ที่ต้องมีคุณสมบัติเป็น White noise เพื่อยืนยันว่าแบบจำลองสามารถจับข้อมูลข่าวสารในอนุกรมเวลาได้ครบถ้วน

3) ปัญหาข้อมูลสูญหายในอนุกรมเวลา (Missing data in time series) ข้อมูลสูญหายเป็นปัญหาที่พบได้บ่อยในข้อมูลอนุกรมเวลา ซึ่งอาจเกิดจากความขัดข้องของระบบบันทึกข้อมูล ความผิดพลาดของอุปกรณ์ตรวจวัด หรือข้อจำกัดในการจัดเก็บและส่งผ่านข้อมูล ปัญหาดังกล่าวสามารถส่งผลกระทบต่อความถูกต้องของการวิเคราะห์และความแม่นยำของการพยากรณ์ หากไม่ได้รับการจัดการอย่างเหมาะสม

ตามกรอบแนวคิดทางสถิติของ Little & Rubin (2019) ข้อมูลสูญหายสามารถจำแนกตามกลไกการเกิดการสูญหายได้หลายประเภท โดยงานวิจัยนี้มุ่งพิจารณา 2 รูปแบบหลัก ได้แก่ ข้อมูลสูญหายแบบสุ่มสมบูรณ์ ซึ่งการสูญหายเกิดโดยไม่ขึ้นกับค่าของข้อมูลใด ๆ และข้อมูลสูญหายแบบสุ่มตามเงื่อนไข ซึ่งความน่าจะเป็นของการสูญหายขึ้นอยู่กับข้อมูลที่สามารถสังเกตได้ก่อนหน้านี้

นอกจากรูปแบบการสูญหายแล้ว อัตราการสูญหายของข้อมูล ยังเป็นอีกปัจจัยสำคัญที่มีผลต่อประสิทธิภาพของกระบวนการวิเคราะห์และการพยากรณ์ โดยอัตราการสูญหายที่สูงขึ้นอาจนำไปสู่ความเอนเอียงของค่าประมาณและการสูญเสียสารสนเทศในข้อมูลอนุกรมเวลา ซึ่งส่งผลกระทบต่อความแม่นยำของแบบจำลองพยากรณ์ในภาพรวม (Schafer & Graham, 2002)

4) วิธีการทดแทนข้อมูลข้อมูล (Imputation techniques) เพื่อจัดการกับปัญหาข้อมูลสูญหายในอนุกรมเวลา ได้มีการพัฒนาวิธีการแทนข้อมูลหลากหลายแนวทาง ตั้งแต่วิธีการพื้นฐานไปจนถึงวิธีที่อาศัยแบบจำลองทางสถิติที่มีความซับซ้อน งานวิจัยนี้เลือกใช้วิธีการแทนข้อมูล 3 วิธี ได้แก่ วิธี LOCF วิธี Linear interpolation และวิธี Kalman filtering ซึ่งสะท้อนระดับความซับซ้อนของวิธีการแทนข้อมูลที่แตกต่างกัน และเป็นวิธีที่นิยมใช้ในงานวิเคราะห์และพยากรณ์อนุกรมเวลา (Fung, 2006; Moritz & Bartz-Beielstein, 2017)

วิธี LOCF เป็นวิธีการแทนข้อมูลแบบพื้นฐาน โดยใช้ค่าที่สังเกตได้ล่าสุดก่อนเกิดการสูญหายมาแทนค่าที่หายไป วิธีนี้เหมาะกับข้อมูลที่มีการเปลี่ยนแปลงแบบค่อยเป็นค่อยไป แต่มีข้อจำกัดในกรณีที่ข้อมูลมีแนวโน้มหรือความผันผวนสูง ในขณะที่วิธี Linear interpolation เป็นการประมาณค่าข้อมูลสูญหายโดยใช้เส้นตรงเชื่อมระหว่างค่าข้อมูลก่อนหน้าและถัดไป ซึ่งเหมาะกับข้อมูลที่มีการเปลี่ยนแปลงอย่างต่อเนื่องในช่วงเวลาสั้น และสำหรับวิธี Kalman filtering เป็นวิธีการแทนข้อมูลที่อาศัยกรอบแบบจำลองเชิงพลวัตในรูปแบบ State-space ซึ่งสามารถคำนึงถึงโครงสร้างของข้อมูล ความไม่แน่นอน และความแปรปรวนของระบบตามเวลาได้อย่างเป็นระบบ วิธีนี้จึงมีประสิทธิภาพสูงในการจัดการข้อมูลอนุกรมเวลาที่มีความซับซ้อนและถูกนำมาใช้อย่างแพร่หลายในงานพยากรณ์และการประมาณค่าข้อมูลสูญหาย (Durbin & Koopman, 2012)

2.2 งานวิจัยที่เกี่ยวข้อง

จากการทบทวนวรรณกรรมด้านการพยากรณ์พลังงาน พบว่า ความแม่นยำของแบบจำลองส่งผลโดยตรงต่อความเสถียรของระบบและการจัดการต้นทุน (Hussain et al., 2025; Duarte et al., 2024) อย่างไรก็ตาม งานวิจัยส่วนใหญ่มักใช้ชุดข้อมูลที่สมบูรณ์และไม่ได้คำนึงถึงผลกระทบจากปัญหาข้อมูลสูญหายที่มักเกิดขึ้นในสถานการณ์จริง สำหรับแนวทางการจัดการข้อมูลสูญหายนั้น การศึกษาของ Chhabra (2023) ได้เปรียบเทียบประสิทธิภาพของวิธีการทดแทนข้อมูลและพบว่า วิธีการพื้นฐานอย่าง LOCF มีประสิทธิภาพดีเพียงพอสำหรับข้อมูลที่สูญหายในช่วงสั้น ๆ หรือมีอัตราการสูญหายต่ำ ในขณะที่ Fung (2006) พบว่า ข้อมูลอนุกรมเวลาที่มีโครงสร้างซับซ้อน เช่น มีแนวโน้มและฤดูกาล วิธีการแบบ State-space เช่น วิธี Kalman filtering จะให้ผลลัพธ์ที่แม่นยำกว่าวิธีการประมาณค่าเชิงเส้น

ในบริบทของประเทศไทย งานวิจัยของ Wongoutong (2021) ได้ประยุกต์ใช้ ARIMA ในการพยากรณ์ แต่ยังขาดการศึกษาเชิงระบบที่พิจารณาอิทธิพลร่วมระหว่างรูปแบบการสูญหาย อัตราการสูญหาย และวิธีการทดแทนข้อมูล ด้วยการวิเคราะห์ความแปรปรวน (Analysis of Variance: ANOVA) ซึ่งเป็นช่องว่างทางวิชาการที่งานวิจัยนี้มุ่งแก้ไข

2.3 ข้อมูลและแหล่งข้อมูล

ข้อมูลที่ใช้ในการศึกษานี้เป็นข้อมูลปริมาณการใช้พลังงานไฟฟ้ารายเดือนของประเทศไทยครอบคลุมระยะเวลาตั้งแต่เดือนมกราคม พ.ศ. 2545 ถึงเดือนสิงหาคม พ.ศ. 2568 รวมทั้งสิ้น 284 เดือน ข้อมูลดังกล่าวรวบรวมจากฐานข้อมูลสถิติพลังงาน สำนักงานนโยบายและแผนพลังงาน (EPPO) กระทรวงพลังงาน โดยผู้วิจัยได้ตรวจสอบความสมบูรณ์ของข้อมูลเบื้องต้นก่อนนำเข้าสู่กระบวนการวิเคราะห์

2.4 การเตรียมข้อมูล

ข้อมูลปริมาณการใช้พลังงานไฟฟ้ารายเดือนถูกนำมาจัดรูปแบบเป็นอนุกรมเวลา โดยกำหนดความถี่เป็นรายเดือน ก่อนเข้าสู่กระบวนการสร้างแบบจำลองได้มีการตรวจสอบคุณสมบัติความนิ่งของข้อมูล ด้วยการทดสอบ Augmented Dickey-Fuller (ADF test) ภายใต้สมมติฐานว่า ข้อมูลมีค่าคงที่ (Intercept) แต่ไม่มีแนวโน้มเชิงเส้น เพื่อประเมินความนิ่งเชิงสถิติของอนุกรมเวลา

ผลการทดสอบ ADF พบว่า สามารถปฏิเสธสมมติฐานศูนย์ของการมีหน่วยราก (Unit root) ได้ที่ระดับนัยสำคัญทางสถิติที่กำหนด แสดงให้เห็นว่าข้อมูลมีคุณสมบัติความนิ่งเชิงสถิติ (Statistical stationarity) จึงกำหนดลำดับของการหาผลต่างเป็น $d = 0$ ในการสร้างแบบจำลอง ARIMA ทั้งนี้ แม้ว่าลักษณะการพลของข้อมูลจะปรากฏแนวโน้มเพิ่มขึ้นในระยะยาว แต่ความแตกต่างดังกล่าวสะท้อนถึงความไม่เป็นนิ่งในเชิงปฏิบัติ (Practical non-stationarity) มากกว่าความไม่เป็นนิ่งในเชิงสถิติ ซึ่งเป็นประเด็นที่พบได้ทั่วไปในข้อมูลเศรษฐกิจและพลังงานระยะยาว ทั้งนี้ผลการทดสอบ ADF ได้แก่ ค่า test statistic, Critical values และ p-value ได้ถูกรายงานเพื่อตรวจสอบการวิเคราะห์ทางสถิติ

นอกจากนี้ เพื่อให้สอดคล้องกับกระบวนการวิเคราะห์ภายหลังการจัดการข้อมูลสูญหาย ผู้วิจัยได้ทำการตรวจสอบความนิ่งของข้อมูลอีกครั้งหลังการทดแทนข้อมูลด้วยวิธี LOCF, Linear interpolation และ Kalman filtering โดยผลการทดสอบพบว่า ข้อมูลที่ผ่านกระบวนการทดแทนแล้วยังคงมีคุณสมบัติความนิ่งเชิงสถิติในทุกกรณี จึงสามารถใช้พารามิเตอร์ $d = 0$ ได้อย่างสอดคล้องตลอดกระบวนการวิเคราะห์

การดำเนินการดังกล่าวช่วยลดความเสี่ยงของการบิดเบือนโครงสร้างข้อมูลจากกระบวนการทดแทน และสนับสนุนความถูกต้องของการเปรียบเทียบผลกระทบของรูปแบบการสูญหาย อัตราการสูญหาย และวิธีการทดแทนข้อมูลต่อความแม่นยำของแบบจำลอง ARIMA ได้อย่างเป็นระบบ

2.5 การจำลองข้อมูลสูญหาย (Missing data simulation)

ข้อมูลสมมติถูกใช้พื้นฐานในการสร้างชุดข้อมูลจำลองที่มีการสูญหายภายใต้การออกแบบการทดลองแบบแฟกทอเรียลเต็ม โดยกำหนดรูปแบบการสูญหาย 2 รูปแบบ ได้แก่ MCAR และ MAR และกำหนดอัตราการสูญหาย 3 ระดับ คือ 10% 20% และ 30% โดยทำการจำลองซ้ำเงื่อนไขละ 50 ครั้ง เพื่อให้ผลการเปรียบเทียบมีความเสถียรในเชิงการทดลองแบบมอนติคาร์โล

อัตราการสูญหายของข้อมูลที่กำหนดไว้ที่ระดับ 10% 20% และ 30% ในงานวิจัยนี้ ถูกเลือกเพื่อให้ครอบคลุมช่วงของการสูญหายตั้งแต่ระดับต่ำ ระดับปานกลาง ไปจนถึงระดับค่อนข้างสูง ซึ่งสอดคล้องกับแนวปฏิบัติที่พบในงานวิจัยด้านข้อมูลพลังงานและระบบสมาร์ตมิเตอร์ ที่รายงานว่าการสูญหายของข้อมูลอาจเกิดขึ้นจากความขัดข้องของอุปกรณ์ตรวจวัด ปัญหาการสื่อสารข้อมูล และกระบวนการจัดเก็บข้อมูลในระดับที่แตกต่างกัน ทั้งนี้ การกำหนดอัตราการสูญหายในลักษณะดังกล่าวมีวัตถุประสงค์เพื่อวิเคราะห์ความไว (Sensitivity) ของวิธีการทดแทนข้อมูลต่อระดับความรุนแรงของข้อมูลสูญหาย และมีได้มุ่งสะท้อนอัตราการสูญหายที่แน่นอนของระบบพลังงานใดระบบหนึ่งโดยเฉพาะ

1) รูปแบบการสูญหายแบบสุ่มสมบูรณ์ (MCAR)

สำหรับ MCAR กำหนดให้โอกาสการสูญหายเป็นอิสระจากค่าของข้อมูลทั้งในอดีตและปัจจุบัน กล่าวคือ สำหรับแต่ละเวลา t จะสุ่มให้ข้อมูลสูญหายด้วยความน่าจะเป็นคงที่ r (โดย $r \in \{0.10, 0.20, 0.30\}$) ดังสมการที่ (2)

$$R_t \sim \text{Bernoulli}(r) \quad (2)$$

เมื่อ $R_t = 1$ หมายถึง ค่าสังเกต ณ เวลา t ถูกกำหนดให้สูญหาย
 $R_t = 0$ หมายถึง ยังคงสังเกตได้

2) รูปแบบการสูญหายแบบสุ่มตามข้อมูลที่สังเกตได้ (MAR)

สำหรับ MAR งานวิจัยนี้กำหนดกลไกการสูญหายให้ขึ้นอยู่กับข้อมูลที่สังเกตได้ก่อนหน้านี้ (lag-dependence) เพื่อสะท้อนสถานการณ์จริงของระบบตรวจวัดที่มีโอกาสขัดข้องมากขึ้น เมื่อค่าการใช้ไฟฟ้าในช่วงก่อนหน้านี้สูงหรือผิดปกติ โดยกำหนดความน่าจะเป็นของการสูญหาย ณ เวลา t ให้เป็นฟังก์ชันของค่าที่สังเกตได้ที่เวลา $t-1$ ผ่านแบบจำลองลอจิสติก ดังสมการที่ (3)

$$\Pr(R_t = 1 \mid Y_{t-1}) = \text{logit}^{-1}(\alpha + \beta Z_{t-1}) \quad (3)$$

โดยที่ R_t คือ ตัวบ่งชี้การสูญหาย ณ เวลา t
 Y_{t-1} คือ ค่าสังเกตที่เวลา $t-1$ (เป็นข้อมูลที่สังเกตได้)
 Z_{t-1} คือ ค่าที่ทำให้เป็นมาตรฐานของ Y_{t-1} เช่น $Z_{t-1} = (Y_{t-1} - \bar{Y}) / s_Y$
เพื่อให้พารามิเตอร์มีสเกลที่เหมาะสม
 α, β คือ พารามิเตอร์กำหนดระดับความโน้มเอียงของการสูญหายให้ขึ้นกับค่าที่สังเกตได้ก่อนหน้านี้

จากนั้นทำการสุ่มตัวแปรการสูญหายตามสมการที่ (4)

$$R_t \sim \text{Bernoulli}(\Pr(R_t = 1 \mid Y_{t-1})) \quad (4)$$

เพื่อให้สัดส่วนการสูญหายโดยรวมตรงตามอัตราที่กำหนด (10% 20% และ 30%) ผู้วิจัยปรับค่าพารามิเตอร์ α (โดยคงทิศทางของ β) ให้ได้อัตราการสูญหายเฉลี่ยใกล้เคียง r ในแต่ละการจำลอง และทำการตรวจสอบสัดส่วนการสูญหายที่เกิดขึ้นจริงก่อนเข้าสู่ขั้นตอนการทดแทนข้อมูล ทั้งนี้ กลไกดังกล่าวทำให้การสูญหายขึ้นอยู่กับข้อมูลที่สังเกตได้ (Lagged observed value) จึงสอดคล้องกับนิยามของ MAR

2.6 การทดแทนข้อมูล

ดำเนินการทดแทนข้อมูลที่สูญหายในแต่ละชุดข้อมูลจำลองด้วย 3 วิธี ได้แก่ LOCF, Linear interpolation และ Kalman filtering โดยใช้ฟังก์ชันมาตรฐานจากแพ็คเกจ imputeTS ในโปรแกรม R ซึ่งเป็นเครื่องมือเฉพาะทางสำหรับการจัดการค่าสูญหายในอนุกรมเวลา

2.7 การสร้างแบบจำลอง ARIMA และการตรวจสอบความเหมาะสม (ARIMA modeling and diagnostic checking)

การสร้างแบบจำลอง ARIMA ในงานวิจัยนี้ดำเนินการตามกรอบแนวคิดของ Box-Jenkins ซึ่งประกอบด้วย 3 ขั้นตอนหลัก ได้แก่ การระบุรูปแบบของแบบจำลอง การประมาณค่าพารามิเตอร์ และการตรวจสอบความเหมาะสมของแบบจำลอง เพื่อให้มั่นใจว่าแบบจำลองสามารถอธิบายโครงสร้างของข้อมูลอนุกรมเวลาได้อย่างเหมาะสม

1) การระบุรูปแบบของแบบจำลอง

หลังจากข้อมูลผ่านกระบวนการทดแทนข้อมูลและได้รับการยืนยันคุณสมบัติความนิ่งเชิงสถิติแล้ว ผู้วิจัยใช้การพิจารณารูปกราฟฟังก์ชันสหสัมพันธ์ตนเอง (Autocorrelation Function: ACF) และฟังก์ชันสหสัมพันธ์ตนเองบางส่วน (Partial Autocorrelation Function: PACF) ร่วมกับการเลือกแบบจำลองอัตโนมัติด้วยฟังก์ชัน `auto.arima` เพื่อระบุค่าพารามิเตอร์ p และ q ที่เหมาะสม ภายใต้เงื่อนไข $d = 0$ ตามผลการทดสอบความนิ่งของข้อมูล

การใช้ `auto.arima` ช่วยลดอคติจากการเลือกแบบจำลองด้วยดุลยพินิจของผู้วิจัย และทำให้การคัดเลือกแบบจำลองอยู่บนเกณฑ์ข้อมูล (Data-driven) โดยใช้ค่า Akaike Information Criterion (AIC) เป็นเกณฑ์หลักในการตัดสินใจ

2) การประมาณค่าพารามิเตอร์

เมื่อได้รูปแบบของแบบจำลองที่เหมาะสมแล้ว ทำการประมาณค่าพารามิเตอร์ของแบบจำลอง ARIMA ด้วยวิธี Maximum Likelihood Estimation (MLE) บนชุดข้อมูลฝึก (Training set) ซึ่งกำหนดสัดส่วนเป็นร้อยละ 80 ของข้อมูลทั้งหมด การประมาณค่าพารามิเตอร์ดำเนินการแยกตามแต่ละชุดข้อมูลจำลองที่เกิดจากเงื่อนไขของรูปแบบการสูญหาย อัตราการสูญหาย และวิธีการทดแทนข้อมูล เพื่อให้ผลการพยากรณ์สะท้อนผลกระทบของกระบวนการจัดการข้อมูลสูญหายอย่างแท้จริง

3) การตรวจสอบความเหมาะสมของแบบจำลอง

ความเหมาะสมของแบบจำลอง ARIMA ได้รับการตรวจสอบผ่านการวิเคราะห์เศษเหลือ (Residuals) โดยพิจารณาว่าเศษเหลือควรมีลักษณะเป็นสัญญาณรบกวนแบบสุ่ม ผู้วิจัยตรวจสอบกราฟ ACF ของเศษเหลือ และทดสอบความเป็นอิสระของเศษเหลือด้วยการทดสอบ Ljung-Box ซึ่งผลการทดสอบพบว่า ไม่สามารถปฏิเสธสมมติฐานศูนย์ของความเป็นอิสระของเศษเหลือได้ในระดับนัยสำคัญที่กำหนด แสดงให้เห็นว่าแบบจำลอง ARIMA ที่ได้มีความเหมาะสมในการอธิบายโครงสร้างของข้อมูลอนุกรมเวลา

การตรวจสอบดังกล่าวช่วยยืนยันว่าแบบจำลองไม่มีโครงสร้างของข้อมูลตกค้างอยู่ในเศษเหลือ และสามารถนำไปใช้ในการพยากรณ์บนชุดข้อมูลทดสอบ (Test set) ได้อย่างเหมาะสม

2.8 การประเมินความแม่นยำและการวิเคราะห์สถิติ

สมมติฐานทางสถิติสำหรับการวิเคราะห์ความแปรปรวนสามทาง เพื่อให้การวิเคราะห์ความแปรปรวนแบบสามปัจจัย มีความชัดเจนเชิงระเบียบวิธี งานวิจัยนี้กำหนดสมมติฐานเชิงสถิติสำหรับผลหลัก (Main effects) และผลปฏิสัมพันธ์ (Interaction effects) ของปัจจัยทั้งสาม ได้แก่ รูปแบบการสูญหาย

ของข้อมูล (Missing pattern: P และระดับ MCAR/MAR) อัตราการสูญหายของข้อมูล (Missing rate: R และระดับ 10% 20% 30%) และวิธีการทดแทนข้อมูล (Imputation method: M และระดับ LOCF, Linear interpolation และ Kalman filtering) ต่อค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Squared Error: RMSE) ดังนี้

1) สมมติฐานสำหรับผลหลัก

- ผลของวิธีการทดแทนข้อมูลสูญหาย (M)
 H_{0M} : ค่าเฉลี่ยของค่า RMSE จากวิธี LOCF วิธี Linear Interpolation และวิธี Kalman Filtering เท่ากัน
 H_{1M} : มีอย่างน้อยหนึ่งวิธีที่ให้ค่าเฉลี่ย RMSE แตกต่างจากวิธีอื่น
- ผลของรูปแบบการสูญหายของข้อมูล (P)
 H_{0P} : ค่าเฉลี่ยของค่า RMSE ภายใต้ MCAR และ MAR เท่ากัน
 H_{1P} : ค่าเฉลี่ยของค่า RMSE ภายใต้ MCAR และ MAR แตกต่างกัน
- ผลของอัตราการสูญหายของข้อมูล (R)
 H_{0R} : ค่าเฉลี่ยของค่า RMSE ภายใต้อัตราการสูญหาย 10%, 20% และ 30% เท่ากัน
 H_{1R} : มีอย่างน้อยหนึ่งระดับอัตราการสูญหายที่ให้ค่าเฉลี่ย RMSE แตกต่างจากระดับอื่น

2) สมมติฐานสำหรับผลปฏิสัมพันธ์

- ปฏิสัมพันธ์ระหว่างวิธีทดแทนและรูปแบบการสูญหาย ($M \times P$)
 H_{0MP} : ผลของวิธีทดแทนต่อค่าเฉลี่ย RMSE ไม่ขึ้นกับรูปแบบการสูญหาย (ไม่มีปฏิสัมพันธ์)
 H_{1MP} : ผลของวิธีทดแทนต่อค่าเฉลี่ย RMSE ขึ้นกับรูปแบบการสูญหาย (มีปฏิสัมพันธ์)
- ปฏิสัมพันธ์ระหว่างวิธีทดแทนและอัตราการสูญหาย ($M \times R$)
 H_{0MR} : ผลของวิธีทดแทนต่อค่าเฉลี่ย RMSE ไม่ขึ้นกับรูปแบบการสูญหาย (ไม่มีปฏิสัมพันธ์)
 H_{1MR} : ผลของวิธีทดแทนต่อค่าเฉลี่ย RMSE ขึ้นกับรูปแบบการสูญหาย (มีปฏิสัมพันธ์)
- ปฏิสัมพันธ์ระหว่างรูปแบบและอัตราการสูญหาย ($P \times R$)
 H_{0PR} : ผลของรูปแบบการสูญหายต่อค่าเฉลี่ย RMSE ไม่ขึ้นกับอัตราการสูญหาย (ไม่มีปฏิสัมพันธ์)
 H_{1PR} : ผลของรูปแบบการสูญหายต่อค่าเฉลี่ย RMSE ขึ้นกับอัตราการสูญหาย (มีปฏิสัมพันธ์)
- ปฏิสัมพันธ์ร่วมแบบสามทาง ($M \times P \times R$)
 H_{0MPR} : ไม่มีปฏิสัมพันธ์ร่วมแบบสามทางของปัจจัยทั้งสามต่อค่าเฉลี่ย RMSE
 H_{1MPR} : มีปฏิสัมพันธ์ร่วมแบบสามทางของปัจจัยทั้งสามต่อค่าเฉลี่ย RMSE

จากสมมติฐานดังกล่าว ผู้วิจัยทำการทดสอบด้วย Three-way ANOVA ที่ระดับนัยสำคัญ $\alpha = 0.05$

ประสิทธิภาพของแบบจำลองประเมินด้วยค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) ซึ่งคำนวณจากผลต่างระหว่างค่าจริง (y_t) และค่าพยากรณ์ ($\hat{y}_t$) ของชุดข้อมูลทดสอบ จำนวน n ตัวอย่าง ดังแสดงในสมการที่ (5)

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2} \quad (5)$$

จากนั้นทำการวิเคราะห์ความแปรปรวนแบบสามปัจจัย เพื่อระบุอิทธิพลของ รูปแบบการสูญหายของข้อมูล ระดับอัตราการสูญหาย และ วิธีการทดแทนข้อมูล และปฏิสัมพันธ์ของปัจจัยเหล่านี้ต่อความแม่นยำของแบบจำลอง ARIMA โดยเบื้องต้นผู้วิจัยได้ตรวจสอบข้อตกลงเบื้องต้น (Assumptions) ของการวิเคราะห์ความแปรปรวน ได้แก่ การแจกแจงปกติของความคลาดเคลื่อน (Normality of residuals) และความแปรปรวนเท่ากัน (Homogeneity of variance) พบว่า ข้อมูลอยู่ในเกณฑ์ที่ยอมรับได้สำหรับการวิเคราะห์ทางสถิติ

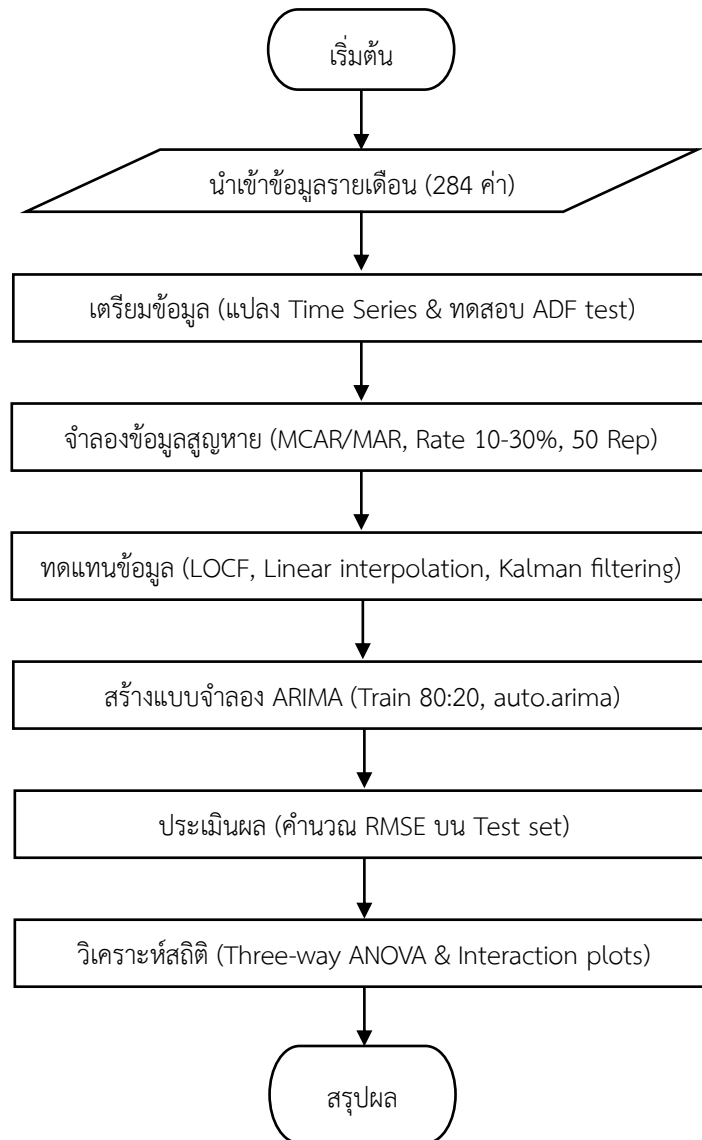
เพื่อประเมินความเพียงพอของจำนวนการจำลองซ้ำ (Replications) ในการออกแบบการทดลองแบบมอนติคาร์โล งานวิจัยนี้กำหนดจำนวนการจำลองซ้ำเท่ากับ 50 ครั้งต่อหนึ่งเงื่อนไขการทดลอง ซึ่งสอดคล้องกับแนวปฏิบัติทั่วไปในงานจำลองเชิงสถิติที่มุ่งเปรียบเทียบผลของปัจจัยหลายระดับ ภายใต้ข้อจำกัดด้านภาระการคำนวณ จำนวนการจำลองซ้ำดังกล่าวช่วยลดความคลาดเคลื่อนจากการสุ่ม ให้อยู่ในระดับที่ยอมรับได้สำหรับการเปรียบเทียบค่าเฉลี่ย RMSE ระหว่างกลุ่ม และทำให้ผลการวิเคราะห์ความแปรปรวนมีความเสถียรเพียงพอสำหรับการสรุปเชิงระเบียบวิธี ทั้งนี้ งานวิจัยนี้ได้มุ่งประมาณค่าพารามิเตอร์เชิงความน่าจะเป็นอย่างแม่นยำ หากแต่มุ่งวิเคราะห์แนวโน้มเชิงเปรียบเทียบของผลกระทบจากรูปแบบการสูญหาย อัตราการสูญหาย และวิธีการทดแทนข้อมูล ซึ่งจำนวนการจำลองซ้ำ 50 ครั้ง ถือว่ามีความเหมาะสมกับวัตถุประสงค์ของการศึกษาในบริบทนี้

ในงานวิจัยนี้เลือกใช้ค่า RMSE เป็นตัวชี้วัดหลักในการประเมินความแม่นยำของการพยากรณ์ เนื่องจาก RMSE ให้ค่าน้ำหนักกับความคลาดเคลื่อนขนาดใหญ่ในระดับกำลังสอง ซึ่งเหมาะสมกับบริบทของข้อมูลการใช้พลังงานไฟฟ้า ที่ความผิดพลาดขนาดใหญ่สามารถส่งผลกระทบต่อการทำงานของระบบและการตัดสินใจเชิงนโยบายได้อย่างมีนัยสำคัญ เมื่อเทียบกับตัวชี้วัดอื่น เช่น MAE หรือ MAPE ซึ่งให้ค่าน้ำหนักความคลาดเคลื่อนอย่างเท่าเทียม หรืออาจมีข้อจำกัดในกรณีที่ข้อมูลมีค่าใกล้ศูนย์ ดังนั้น RMSE จึงเหมาะสมสำหรับการประเมินประสิทธิภาพของแบบจำลองพยากรณ์ในบริบทของการศึกษานี้

2.9 กรอบแนวคิดการวิจัย

กรอบแนวคิดการวิจัยเริ่มจากการเตรียมข้อมูลและจำลองข้อมูลสูญหายภายใต้กลไก MCAR และ MAR จากนั้นทดแทนข้อมูลสูญหายด้วยวิธี LOCF วิธี Linear interpolation และวิธี Kalman filtering

แล้วนำข้อมูลที่ได้ไปสร้างแบบจำลอง ARIMA และประเมินประสิทธิภาพด้วยค่า RMSE ก่อนวิเคราะห์ผลด้วยการวิเคราะห์ความแปรปรวนสามทางและแผนภาพปฏิสัมพันธ์ ดังแสดงในรูปที่ 1

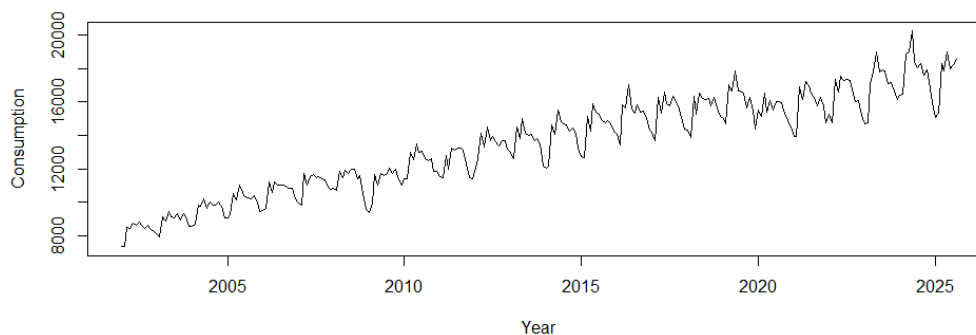


รูปที่ 1. กรอบแนวคิดและขั้นตอนการดำเนินงานวิจัย

3. ผลการวิจัยและอภิปรายผล

3.1 ลักษณะข้อมูลอนุกรมเวลาเบื้องต้น

จากการพิจารณาลักษณะข้อมูลอนุกรมเวลาปริมาณการใช้พลังงานไฟฟ้าของประเทศไทย (รายเดือน) ในช่วงปี ค.ศ. 2002-2025 ดังแสดงในรูปที่ 2 พบว่า ข้อมูลมีลักษณะโครงสร้างที่ประกอบด้วย แนวโน้ม ที่เพิ่มสูงขึ้นอย่างต่อเนื่องในระยะยาว และปรากฏวัฏจักรฤดูกาล ที่ชัดเจนในรอบ 12 เดือน ซึ่งสะท้อนพฤติกรรมการใช้ไฟฟ้าที่มีรูปแบบซ้ำเดิมในแต่ละช่วงเดือนของปี



รูปที่ 2. อนุกรมเวลาปริมาณการใช้พลังงานไฟฟ้า (รายเดือน) แสดงแนวโน้มขาขึ้นและวัฏจักรฤดูกาล

3.2 ค่าสถิติเชิงพรรณนาของ RMSE

จากการวิเคราะห์ภาพรวมประสิทธิภาพของแบบจำลอง พบว่า ระดับความคลาดเคลื่อนของการพยากรณ์มีความผันแปรตามเงื่อนไขของข้อมูลอย่างชัดเจน โดยผลการวิเคราะห์ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของค่าความคลาดเคลื่อนรากที่สองเฉลี่ย (RMSE) ดังปรากฏในตารางที่ 1 พบว่า ในภาพรวมรูปแบบการสูญหายแบบ MCAR มีค่าเฉลี่ย RMSE ต่ำกว่ารูปแบบ MAR เล็กน้อยในเกือบทุกระดับอัตราการสูญหาย ตัวอย่างเช่น อัตราการสูญหายร้อยละ 10 รูปแบบ MCAR มีค่าเฉลี่ย RMSE เท่ากับ 3,594.40 ในขณะที่ MAR อยู่ที่ 3,602.63 เมื่อพิจารณาผลของอัตราการสูญหาย พบว่า เมื่อข้อมูลสูญหายมากขึ้น (จาก 10% เป็น 30%) ค่าเฉลี่ย RMSE มีแนวโน้มลดลงเล็กน้อยในทั้งสองรูปแบบการสูญหาย ซึ่งอาจเป็นผลมาจากวิธีการทดแทนข้อมูลบางวิธีที่ลดทอนความแปรปรวนของข้อมูลลง (Variance reduction) ตามที่ Chhabra (2023) ได้ตั้งข้อสังเกตไว้ในการศึกษาลักษณะผลกระทบของการทดแทนข้อมูลแบบดั้งเดิม

ตารางที่ 1. ค่าสถิติเชิงพรรณนาของค่า RMSE จำแนกตามรูปแบบการสูญหายและอัตราการสูญหายของข้อมูล

รูปแบบการสูญหาย	อัตราการสูญหาย (%)	ค่าเฉลี่ย RMSE	ส่วนเบี่ยงเบนมาตรฐาน	จำนวน
MAR	10	3,602.63	389.66	150
MAR	20	3,592.82	331.91	150
MAR	30	3,515.15	352.37	150
MCAR	10	3,594.40	348.49	150

ตารางที่ 1. (ต่อ) ค่าสถิติเชิงพรรณนาของค่า RMSE จำแนกตามรูปแบบการสูญหายและอัตราการสูญหายของข้อมูล

รูปแบบการสูญหาย	อัตราการสูญหาย (%)	ค่าเฉลี่ย RMSE	ส่วนเบี่ยงเบนมาตรฐาน	จำนวน
MCAR	20	3,533.11	317.39	150
MCAR	30	3,490.36	317.85	150

3.3 ความแตกต่างของความแม่นยำระหว่างวิธีการทดแทนข้อมูล

เมื่อพิจารณาเปรียบเทียบประสิทธิภาพรายวิธีพบความแตกต่างของความแม่นยำอย่างมีนัยสำคัญจากการเปรียบเทียบประสิทธิภาพของวิธีการทดแทนข้อมูลทั้ง 3 วิธีในภาพรวม แสดงดังตารางที่ 2 พบว่าวิธี Kalman filtering มีประสิทธิภาพสูงที่สุด โดยให้ค่าเฉลี่ย RMSE ต่ำที่สุดที่ 3,459.91 รองลงมาคือวิธี LOCF (3,565.22) และวิธี Linear interpolation ให้ค่าความคลาดเคลื่อนสูงที่สุด (3,639.10) ผลลัพธ์นี้ยืนยันว่าวิธีการแบบ State-space เหมาะสมกับข้อมูลรายเดือนที่มีความผันผวนและฤดูกาลมากกว่าวิธีการประมาณค่าเชิงเส้น

ตารางที่ 2. การเปรียบเทียบค่าเฉลี่ย RMSE ระหว่างวิธีการทดแทนข้อมูลทั้ง 3 วิธี

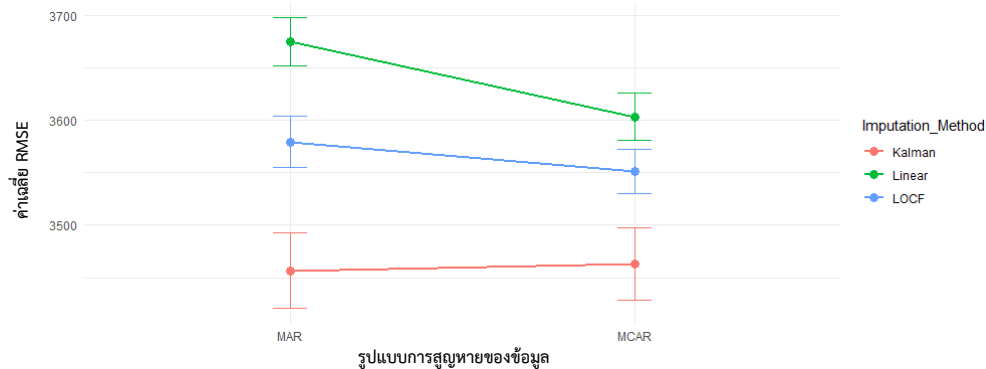
วิธีการทดแทนข้อมูล	ค่าเฉลี่ย RMSE	ส่วนเบี่ยงเบนมาตรฐาน	อันดับ
Kalman filtering	3,459.91	430.80	1
LOCF	3,565.22	278.86	2
Linear interpolation	3,639.10	282.20	3

3.4 ผลของรูปแบบการสูญหาย (MCAR/MAR) ต่อความแม่นยำ

ผลการวิเคราะห์ปฏิสัมพันธ์ระหว่างวิธีการทดแทนข้อมูลและรูปแบบการสูญหาย แสดงในตารางที่ 3 และดังรูปที่ 3 ซึ่งชี้ให้เห็นว่า ความเหมาะสมของวิธีการทดแทนขึ้นอยู่กับกลไกการเกิดข้อมูลสูญหาย โดยในกรณี MAR วิธี Kalman filtering มีความโดดเด่นอย่างมาก (RMSE = 3,456.48) ซึ่งต่ำกว่าวิธี Linear interpolation (3,674.90) อย่างชัดเจน ส่วนในกรณี MCAR ความแตกต่างระหว่างวิธี Kalman filtering (3,463.34) และ LOCF (3,551.21) มีช่องว่างที่แคบลง แต่ Kalman filtering ยังคงให้ผลลัพธ์ที่ดีกว่าในภาพรวม

ตารางที่ 3. ค่าเฉลี่ย RMSE จำแนกตามปฏิสัมพันธ์ระหว่างรูปแบบการสูญหายและวิธีการทดแทนข้อมูล

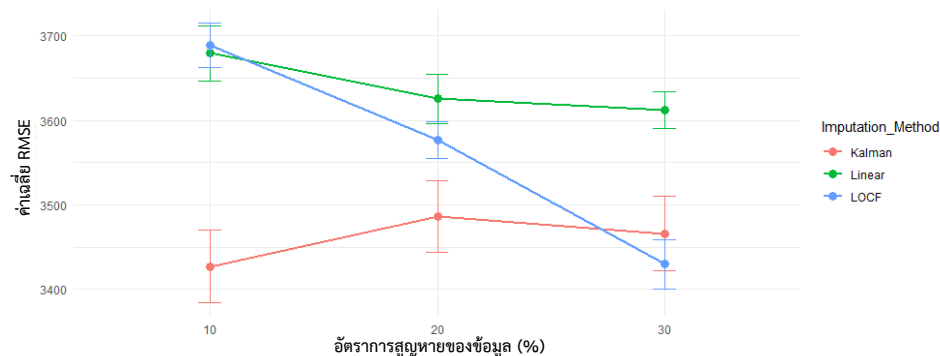
รูปแบบการสูญหาย	วิธีการทดแทนข้อมูล		
	Kalman filtering	LOCF	Linear interpolation
MAR	3,456.48	3,579.22	3,674.90
MCAR	3,463.34	3,551.21	3,603.31



รูปที่ 3. กราฟอิทธิพลร่วมระหว่างวิธีการทดแทนข้อมูลและรูปแบบการสูญหายที่มีต่อค่า RMSE

3.5 ผลของอัตราการสูญหาย

การพิจารณาความสัมพันธ์ระหว่างวิธีการทดแทนและอัตราการสูญหาย จากรูปที่ 4 พบพฤติกรรมที่น่าสนใจ คือ วิธี LOCF แสดงแนวโน้มค่า RMSE ที่ลดลงอย่างมากเมื่ออัตราการสูญหายเพิ่มขึ้น (จาก 3,688.96 ที่ 10% เหลือ 3,429.83 ที่ 30%) ซึ่งสันนิษฐานได้ว่าเกิดจากผลการเกลี่ยเรียบของข้อมูล (Smoothing effect) ที่ลดความผันผวนของข้อมูลลง ทำให้แบบจำลอง ARIMA จับทิศทางได้ง่ายขึ้นในเชิงตัวเลข ในขณะที่ Kalman filtering แสดงความเสถียรในทุกช่วงอัตราการสูญหาย (RMSE อยู่ระหว่าง 3,427 - 3,486) ซึ่งสะท้อนถึงความน่าเชื่อถือในการรักษาโครงสร้างข้อมูลจริงได้ดีกว่า



รูปที่ 4. กราฟอิทธิพลร่วมระหว่างวิธีการทดแทนข้อมูลและอัตราการสูญหายที่มีต่อค่า RMSE

3.6 วิเคราะห์ความแปรปรวนแบบสามทาง

ผลการวิเคราะห์ความแปรปรวนดังแสดงในตารางที่ 4 ยืนยันว่า วิธีการทดแทนข้อมูล ($F=22.24$, $p < 0.001$) และอัตราการสูญหาย ($F=6.42$, $p=0.002$) เป็นปัจจัยหลักที่ส่งผลต่อค่า RMSE อย่างมีนัยสำคัญทางสถิติ ในขณะที่รูปแบบการสูญหาย (Pattern) เพียงปัจจัยเดียวไม่ส่งผลอย่างมีนัยสำคัญ ($p = 0.161$)

อย่างไรก็ตาม เมื่อพิจารณาอิทธิพลร่วม (Interaction effects) พบประเด็นที่น่าสนใจคือ ปฏิสัมพันธ์ระหว่างวิธีการทดแทนข้อมูลและรูปแบบการสูญหาย ไม่พบนัยสำคัญทางสถิติ ($p=0.347$) แต่

กลับพบปฏิสัมพันธ์ที่มีนัยสำคัญสูงระหว่าง วิธีการทดแทนข้อมูลและอัตราการสูญหาย ($F=5.49$ และ $p < 0.001$) ผลลัพธ์นี้สอดคล้องกับรูปที่ 3 ซึ่งแสดงให้เห็นว่าประสิทธิภาพของวิธี LOCF แปรผันอย่างมากตามระดับอัตราการสูญหายที่เพิ่มขึ้น ในขณะที่วิธี Kalman filtering มีความเสถียรมากกว่า นอกจากนี้ยังพบปฏิสัมพันธ์ร่วมแบบสามทาง อย่างมีนัยสำคัญ ($p=0.004$) ซึ่งชี้ให้เห็นว่า ความแม่นยำของการพยากรณ์ขึ้นอยู่กับความซับซ้อนเชิงโครงสร้างของทั้งสามปัจจัยร่วมกัน

ตารางที่ 4. ผลการวิเคราะห์ความแปรปรวนแบบสามทางของปัจจัยที่มีผลต่อค่า RMSE

แหล่งความแปรปรวน	df	ผลบวกกำลังสอง	กำลังสองเฉลี่ย	ค่า F	ค่า p-value
Imputation Method	2	4865872	2432936	22.24	<0.001***
Pattern	1	215004	215004	1.97	0.161
Rate	2	1405831	702916	6.42	0.002**
Method × Pattern	2	231698	115849	1.06	0.347
Method × Rate	4	2404490	601123	5.49	<0.001***
Pattern × Rate	2	103614	51807	0.47	0.623
Method × Pattern × Rate	4	1689801	422450	3.86	0.004

หมายเหตุ: *** $p < 0.001$, ** $p < 0.01$

3.7 อภิปรายผล (Discussion)

จากการศึกษาผลของรูปแบบและอัตราการสูญหายของข้อมูลต่อความแม่นยำของแบบจำลอง ARIMA สามารถอภิปรายประเด็นสำคัญร่วมกับวรรณกรรมที่เกี่ยวข้องได้ดังนี้

1) ประสิทธิภาพและความเสถียรของวิธี Kalman filtering ผลการวิจัยยืนยันว่าวิธี Kalman filtering เป็นวิธีการทดแทนข้อมูลที่มีประสิทธิภาพสูงสุด (RMSE ต่ำที่สุด) และมีความเสถียร (Robustness) มากที่สุดเมื่อเทียบกับวิธีอื่น โดยสามารถรักษาความแม่นยำได้ดีแม้ในสถานการณ์ที่อัตราการสูญหายเปลี่ยนแปลงไป ผลลัพธ์นี้สอดคล้องกับงานวิจัยของ Fung (2006) และ De Bézenac et al. (2020) ที่ระบุว่าวิธี State-space เหมาะสมอย่างยิ่งกับการกู้คืนโครงสร้างพลวัตของข้อมูลอนุกรมเวลาที่มีแนวโน้มและฤดูกาลชัดเจน เช่น ข้อมูลการใช้พลังงานไฟฟ้า เป็นต้น

2) ผลกระทบของอัตราการสูญหายและปรากฏการณ์ Smoothing effect การค้นพบที่ตีพิมพ์ร่วมกันระหว่างวิธีการทดแทนและอัตราการสูญหาย อย่างมีนัยสำคัญทางสถิติ เป็นประเด็นที่น่าสนใจ โดยพบว่าวิธี LOCF มีความคลาดเคลื่อนลดลงอย่างมากเมื่อข้อมูลสูญหายเพิ่มขึ้น (30%) ปรากฏการณ์นี้สอดคล้องกับข้อสังเกตของ Chhabra (2023) เกี่ยวกับผลของการเกลี่ยเรียบ (Smoothing effect) การแทนค่าด้วยข้อมูลเดิมจำนวนมากติดต่อกันทำให้ข้อมูลสูญเสียความผันผวนตามธรรมชาติ ส่งผลให้แบบจำลอง ARIMA พยากรณ์ค่ากลางได้ง่ายขึ้นในเชิงตัวเลข แต่อาจลดทอนความเป็นตัวแทนของพฤติกรรมข้อมูลจริงลง ในขณะที่วิธี Kalman filtering ไม่ได้รับผลกระทบจากปรากฏการณ์นี้มากนัก แม้ว่าวิธีการทดแทนข้อมูลที่มีความซับซ้อนสูงขึ้น เช่น วิธีที่อาศัยการเรียนรู้ของเครื่องหรือโครงสร้างเชิงลึก จะมีศักยภาพในการจัดการข้อมูลสูญหายในบางบริบท แต่อาศัยข้อมูลจำนวนมากและสมมติฐานเพิ่มเติม ซึ่งอยู่นอกขอบเขตของ

การศึกษานี้ที่มุ่งเน้นการเปรียบเทียบวิธีการทดแทนข้อมูลที่สามารถนำไปใช้ได้ง่ายและแพร่หลายในงานพยากรณ์อนุกรมเวลา

3) บทบาทของรูปแบบการสูญหาย (Role of missing pattern) แม้ผลการวิเคราะห์ความแปรปรวนจะไม่พบอิทธิพลร่วมระหว่างวิธีการทดแทนและรูปแบบการสูญหาย อย่างมีนัยสำคัญทางสถิติ ซึ่งแตกต่างจากการศึกษาบางส่วนในอดีต แต่เมื่อพิจารณาอิทธิพลร่วมแบบสามทาง กลับพบนัยสำคัญที่ระดับ 0.01

จากผลการวิเคราะห์ดังกล่าว แม้จะพบว่า อิทธิพลร่วมระหว่างวิธีการทดแทนและรูปแบบการสูญหาย (Method $\times$ Pattern) ไม่มีนัยสำคัญทางสถิติในหลายกรณี อย่างไรก็ตาม ข้อค้นพบนี้ควรถูกตีความภายใต้ขอบเขตของกรอบการศึกษานี้เท่านั้น กล่าวคือ ผลลัพธ์ดังกล่าวเป็นผลที่ได้จากการประยุกต์ใช้แบบจำลอง ARIMA กับข้อมูลรายเดือน ภายใต้กลไกการสูญหายแบบ MCAR และ MAR และการทดแทนข้อมูลด้วยวิธีที่พิจารณาในงานวิจัยนี้

ในบริบทอื่น เช่น การใช้ตัวแบบพยากรณ์ที่มีโครงสร้างไม่เชิงเส้น การใช้ข้อมูลที่มีความถี่สูงกว่าหรือการพิจารณาผลกระทบการสูญหายที่ซับซ้อนยิ่งขึ้น ข้อสรุปเกี่ยวกับอิทธิพลของรูปแบบการสูญหายของข้อมูลอาจแตกต่างออกไป ดังนั้น ผลการศึกษานี้จึงมิได้มุ่งสรุปเชิงทั่วไปว่า รูปแบบการสูญหายของข้อมูลไม่มีความสำคัญโดยสมบูรณ์ หากแต่เป็นข้อค้นพบเชิงบริบทที่สะท้อนพฤติกรรมของข้อมูลและตัวแบบภายใต้เงื่อนไขการทดลองที่กำหนดไว้ในงานวิจัยนี้

ผลลัพธ์นี้บ่งชี้ว่ารูปแบบการสูญหาย (MCAR หรือ MAR) อาจไม่ได้เป็นปัจจัยเดียวที่กำหนดประสิทธิภาพของวิธีทดแทนข้อมูลโดยตรงในข้อมูลลักษณะรายเดือน แต่ผลกระทบของรูปแบบจะปรากฏชัดเจนขึ้นเมื่อทำงานร่วมกับระดับอัตราการสูญหายที่สูงขึ้น ข้อค้นพบนี้สะท้อนให้เห็นว่าในบริบทของการพยากรณ์ระยะยาวปริมาณข้อมูลที่สูญหาย (Missing rate) เป็นปัจจัยวิกฤตที่มีน้ำหนักต่อการตัดสินใจเลือกวิธีทดแทนมากกว่ากลไกการสูญหายเพียงอย่างเดียว

4) ข้อจำกัดของการใช้แบบจำลอง auto.arima ทั้งนี้ การใช้ฟังก์ชัน auto.arima ในงานวิจัยนี้มีข้อจำกัด เนื่องจากการเลือกแบบจำลองอาศัยเกณฑ์สารสนเทศและสมมติฐานพื้นฐานของแบบจำลอง ARIMA ซึ่งอาจไม่สามารถสะท้อนโครงสร้างข้อมูลที่มีความไม่เชิงเส้นหรือการเปลี่ยนแปลงเชิงโครงสร้างได้อย่างครบถ้วน อย่างไรก็ตาม งานวิจัยนี้มุ่งใช้ ARIMA ในฐานะแบบจำลองอ้างอิง (Baseline model) เพื่อศึกษาผลกระทบของข้อมูลสูญหายและวิธีการทดแทนข้อมูลต่อความแม่นยำของการพยากรณ์ มากกว่าการเปรียบเทียบศักยภาพของตัวแบบพยากรณ์ขั้นสูง

5) ความสำคัญในบริบทงานวิจัยไทย งานวิจัยนี้ได้ทดแทนช่องว่างทางวิชาการ (Research Gap) ในบริบทของประเทศไทย โดยขยายผลจากการศึกษาเดิมของ Rothjanawan (2021) และ Sutthison & Pienkhawsook (2024) ที่มุ่งเน้นการเปรียบเทียบตัวแบบพยากรณ์เป็นหลัก มาสู่การตรวจสอบผลกระทบของกระบวนการเตรียมข้อมูลอย่างเป็นระบบ

โดยเฉพาะอย่างยิ่ง ผลการค้นพบเรื่องความเสี่ยงของการใช้วิธี LOCF ในข้อมูลที่มีอัตราการสูญหายสูงเป็นองค์ความรู้ใหม่เชิงประจักษ์ที่ไม่เคยมีการกล่าวถึงในงานวิจัยข้อมูลพลังงานไทยมาก่อน ซึ่งจะเป็นประโยชน์อย่างยิ่งต่อการวางแผนจัดการข้อมูลของหน่วยงานภาครัฐเพื่อเพิ่มความแม่นยำในการพยากรณ์ระยะยาว

ในเชิงปฏิบัติ ผลการศึกษานี้สามารถนำไปใช้เป็นแนวทางในการจัดการข้อมูลด้านพลังงานของหน่วยงานภาครัฐ เช่น สำนักงานนโยบายและแผนพลังงาน (EPPO) โดยเฉพาะในกรณีที่ข้อมูลการใช้พลังงานมีความไม่สมบูรณ์หรือเกิดการสูญหายเป็นช่วง ๆ ผลลัพธ์ชี้ให้เห็นว่าการเลือกวิธีทดแทนข้อมูลที่เหมาะสมมีความสำคัญอย่างยิ่งต่อความแม่นยำของการพยากรณ์ ซึ่งส่งผลโดยตรงต่อการวางแผนกำลังการผลิตและการกำหนดนโยบายพลังงานในระยะยาว ทั้งนี้ การหลีกเลี่ยงวิธีการทดแทนที่อาจก่อให้เกิดความเอนเอียง เช่น LOCF ในกรณีที่มียอดการสูญหายสูง และการเลือกใช้วิธีที่มีความเสถียรสูงกว่า เช่น วิธี Kalman filtering จะช่วยลดความเสี่ยงในการตัดสินใจเชิงนโยบายจากข้อมูลที่บิดเบือน

ในระดับเชิงนโยบาย ผลการศึกษานี้สามารถสนับสนุนการพัฒนากระบวนการสร้างพื้นฐานด้านข้อมูลพลังงานของภาครัฐในหลายมิติ ประการแรก ผลลัพธ์ชี้ให้เห็นความจำเป็นของการออกแบบและพัฒนาระบบตรวจวัดอัจฉริยะ (Smart meter) ที่สามารถรองรับข้อมูลสูญหายได้อย่างเหมาะสม โดยควรคำนึงถึงกระบวนการจัดการข้อมูลสูญหายตั้งแต่ขั้นตอนการออกแบบระบบ ประการที่สอง การออกแบบระบบตรวจวัดและจัดเก็บข้อมูลควรมีมาตรการลดความเสี่ยงของการสูญหายของข้อมูล เช่น การสำรองข้อมูลและการตรวจสอบความสมบูรณ์ของข้อมูลอย่างสม่ำเสมอ เพื่อลดการพึ่งพาวิธีการทดแทนข้อมูลที่อาจก่อให้เกิดความเอนเอียงในภายหลัง ประการสุดท้าย ผลการศึกษานี้สะท้อนให้เห็นถึงความสำคัญของการกำหนดกรอบการกำกับดูแลข้อมูล (Data governance) ในภาครัฐด้านพลังงาน โดยเฉพาะการกำหนดแนวปฏิบัติที่ชัดเจนเกี่ยวกับการจัดการข้อมูลที่ไม่สมบูรณ์ เพื่อสนับสนุนการวิเคราะห์และการตัดสินใจเชิงนโยบายที่มีความน่าเชื่อถือในระยะยาว

4. สรุปผลการวิจัย

การวิจัยนี้มุ่งศึกษาอิทธิพลของรูปแบบและอัตราการสูญหายของข้อมูลที่มีต่อความแม่นยำของแบบจำลอง ARIMA ผลการศึกษาพบว่า วิธีการทดแทนข้อมูลเป็นปัจจัยวิกฤตที่สุดต่อความแม่นยำของการพยากรณ์ โดยวิธี Kalman filtering มีประสิทธิภาพโดดเด่นและมีความเสถียรสูงสุดในทุกเงื่อนไขการทดลอง ผลการวิเคราะห์ทางสถิติชี้ให้เห็นว่า ประสิทธิภาพของวิธีการทดแทนข้อมูลขึ้นอยู่กับ อัตราการสูญหายอย่างมีนัยสำคัญมากกว่ารูปแบบการสูญหาย โดยเฉพาะอย่างยิ่งวิธี LOCF ที่แสดงพฤติกรรมความคลาดเคลื่อนลดลงเมื่อข้อมูลสูญหายมากขึ้นจากการลดทอนความแปรปรวนของข้อมูล โดยสรุป งานวิจัยนี้เสนอแนะให้เลือกใช้ Kalman filtering เป็นวิธีหลักในการเตรียมข้อมูลสำหรับแบบจำลอง ARIMA ในการพยากรณ์พลังงานไฟฟ้า เนื่องจากเป็นวิธีที่ให้ความแม่นยำสูงและมีความทนทานต่อการเปลี่ยนแปลงของสัดส่วนข้อมูลสูญหายได้ดีที่สุด ข้อค้นพบนี้ช่วยสนับสนุนแนวทางการเลือกกระบวนการเตรียมข้อมูลสำหรับการพยากรณ์พลังงานระยะยาวของหน่วยงานภาครัฐและภาคอุตสาหกรรมให้มีประสิทธิภาพยิ่งขึ้น

ข้อเสนอแนะเพื่อนำไปใช้ประโยชน์

1) การเลือกวิธีทดแทนข้อมูล ผลการวิจัยบ่งชี้ว่า สำหรับข้อมูลอนุกรมเวลาที่มีวัฏจักรฤดูกาลชัดเจน วิธี Kalman filtering เป็นทางเลือกที่มีความเหมาะสมสูงสุด เนื่องจากให้ความแม่นยำและมีความทนทานต่อการเปลี่ยนแปลงของอัตราการสูญหาย

2) ข้อควรระวังในการใช้วิธี LOCF ควรระมัดระวังการตีความผลลัพธ์จากวิธี LOCF ในกรณีที่ข้อมูลสูญหายในสัดส่วนสูง เนื่องจากค่าความคลาดเคลื่อนที่ต่ำลงเกิดจากการลดทอนความแปรปรวน ซึ่งอาจไม่สะท้อนพฤติกรรมความผันผวนที่แท้จริงของระบบไฟฟ้า

ข้อเสนอแนะสำหรับการวิจัยในอนาคตและข้อจำกัด

1) การศึกษาเปรียบเทียบกับแบบจำลองอื่น ควรมีการขยายผลการศึกษาโดยเปรียบเทียบผลกระทบของการทดแทนข้อมูลต่อแบบจำลองการพยากรณ์สมัยใหม่อื่น ๆ เช่น Long Short-Term Memory (LSTM) หรือ Hybrid models เพื่อตรวจสอบว่าปรากฏการณ์ Smoothing effect ของวิธี LOCF จะส่งผลกระทบต่อลักษณะเดียวกันหรือไม่

2) ข้อจำกัดด้านความละเอียดของข้อมูล งานวิจัยนี้มุ่งเน้นที่ข้อมูลรายเดือน ซึ่งเหมาะสมกับการวางแผนระยะกลางถึงยาว การศึกษาในอนาคตอาจพิจารณาใช้ข้อมูลที่มีความละเอียดสูงขึ้น (รายวันหรือรายชั่วโมง) เพื่อตรวจสอบความอ่อนไหวของวิธีการทดแทนต่อความผันผวนระยะสั้น (High-frequency volatility)

3) การศึกษาในสถานการณ์การสูญหายของข้อมูลแบบไม่สุ่ม (Missing Not At Random: MNAR) ซึ่งเป็นสถานการณ์ที่ซับซ้อนและอาจเกิดขึ้นได้จริงในระบบตรวจวัดอัจฉริยะ

เอกสารอ้างอิง (References)

- Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: Forecasting and control* (5th ed.). John Wiley & Sons.
- Chhabra, G. (2023). Comparison of imputation methods for univariate time series. *International Journal on Recent and Innovation Trends in Computing and Communication*, 11(2s), 286-292. <https://doi.org/10.17762/ijritcc.v11i2s.6148>
- De Bézenac, E., Rangapuram, S. S., Benidis, K., Bohlke-Schneider, M., Kurl, R., Stella, L., Hilmkil, H., Januschowski, T., & Gasthaus, J. (2020). Normalizing Kalman filters for multivariate time series analysis. *Proceedings of the 34th International Conference on Neural Information Processing Systems* (pp. 2995-3007). Curran Associates Inc.
- Duarte, O., Duarte, J. E., & Rosero-Garcia, J. (2024). Data imputation in electricity consumption profiles through shape modeling with autoencoders. *Mathematics*, 12(19), Article 3004. <https://doi.org/10.3390/math12193004>
- Durbin, J., & Koopman, S. J. (2012). *Time series analysis by state space methods* (2nd ed.). Oxford University Press.
- Fung, D. S. (2006). *Methods for the estimation of missing values in time series* [Master's thesis, Edith Cowan University]. <https://ro.ecu.edu.au/theses/63>
- Hussain, A., Giangrande, P., Franchini, G., Fenili, L., & Messi, S. (2025). Analyzing the effect of error estimation on random missing data patterns in mid-term electrical forecasting. *Electronics*, 14(7), Article 1383. <https://doi.org/10.3390/>

electronics14071383

- Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and practice* (3rd ed.). OTexts.
- Little, R. J. A., & Rubin, D. B. (2019). *Statistical analysis with missing data* (3rd ed.). John Wiley & Sons.
- Moritz, S., & Bartz-Beielstein, T. (2017). imputeTS: Time series missing value imputation in R. *The R Journal*, 9(1), 207-218. <https://doi.org/10.32614/RJ-2017-009>
- Rothjanawan, K., & Phetjirachotkul, W. (2021). Feature selection methods for imputation missing values of time series data using data mining. *Princess of Naradhiwas University Journal*, 13(2), 326-341. (in Thai)
- Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, 7(2), 147-177. <https://doi.org/10.1037/1082-989X.7.2.147>
- Sutthison, T., & Pienkhawsook, T. (2024). Hybrid model for forecasting time series data of monthly household electrical distribution units in Thailand. *The Journal of King Mongkut's University of Technology North Bangkok*, 34(2), 1-18. (in Thai)
- Wongoutong, C. (2021). Imputation methods in time series with a trend and a consecutive missing value pattern. *Thailand Statistician*, 19(4), 866-879.