การพยากรณ์ปริมาณฝนโดยใช้การแยกองค์ประกอบข้อมูลร่วมกับการเรียนรู้เชิงลึก

Rainfall Forecasting Using Data Decomposed Technique and Deep Learning

พงษ์บวร รัตนพันธ์[1,*] และ วรรณดี ไทยสยาม[1]

[1] ภาควิชาวิศวกรรมทรัพยากรน้ำ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเกษตรศาสตร์ กรุงเทพมหานคร

**Corresponding author; E-mail address: pongbavorn.ra@ku.th*

บทคัดย่อ

ข้อมูลปริมาณน้ำฝนเป็นข้อมูลพื้นฐานสำคัญสำหรับการคาดการณ์ วางแผน และบริหารจัดการน้ำ ปัจจุบันมีการใช้แบบจำลองพยากรณ์ปริมาณน้ำฝนหลากหลายรูปแบบ โดยเฉพาะแบบจำลองอากาศเชิงตัวเลข ซึ่งต้องอาศัยข้อมูลและทรัพยากรจำนวนมาก งานวิจัยนี้มีวัตถุประสงค์เพื่อคาดการณ์ปริมาณฝนในพื้นที่รับน้ำเหนือเขื่อนลำตะคอง เพื่อใช้ในการพยากรณ์น้ำไหลเข้าอ่างเก็บน้ำลำตะคองต่อไป ลักษณะพื้นที่ศึกษาเป็นการไหลตามธรรมชาติ (free flow) ไม่มีสิ่งก่อสร้างควบคุมการไหล น้ำที่ไหลเข้าสู่อ่างเก็บน้ำจึงสะท้อนปริมาณน้ำที่เกิดขึ้นจากธรรมชาติอย่างแท้จริง ทำให้เหมาะสมสำหรับการศึกษาพฤติกรรมของฝนและน้ำไหล ในการศึกษานี้ได้ตั้งสมมติฐานว่า การแยกองค์ประกอบข้อมูล (Data Decomposition) ด้วย Variational Mode Decomposition (VMD) ก่อนนำไปพยากรณ์ด้วยแบบจำลองจะช่วยให้แบบจำลองสามารถเรียนรู้ลักษณะเฉพาะของข้อมูลได้ง่ายขึ้น โดยลดความซับซ้อนของสัญญาณต้นฉบับและแยกแยะลักษณะรูปแบบต่าง ๆ ของข้อมูลที่มีความแปรปรวนสูง ซึ่งคาดว่าจะช่วยเพิ่มความแม่นยำในการพยากรณ์ได้อย่างมีประสิทธิภาพ โดย VMD เป็นเทคนิคการแยกสัญญาณที่สามารถแยกข้อมูลออกเป็นองค์ประกอบย่อยตามช่วงความถี่ที่แตกต่างกัน เพื่อลดความซับซ้อนของข้อมูล ขณะที่ Stack Long Short-Term Memory (Stack LSTM) เป็นโครงข่ายประสาทเทียมลำดับเวลาที่ซ้อนหลายชั้น เพื่อเพิ่มความสามารถในการจับลักษณะเชิงลึกของข้อมูล ดังนั้นจึงแบ่งการศึกษานี้ออกเป็น 2 กรณีเพื่อเปรียบเทียบประสิทธิภาพ กรณีที่ 1 ใช้ Stack LSTM เพียงอย่างเดียว และกรณีที่ 2 ใช้ VMD ร่วมกับ Stack LSTM ผลการศึกษาพบว่า การใช้ VMD ร่วมกับ Stack LSTM ช่วยเพิ่มความแม่นยำในการพยากรณ์ได้อย่างมีนัยสำคัญ โดยการพยากรณ์ฝนเฉลี่ยล่วงหน้า 1 วันมีค่า R เท่ากับ 0.928, NSE เท่ากับ 0.861, RMSE เท่ากับ 2.858 มิลลิเมตร และ MAE เท่ากับ 1.409 มิลลิเมตร สูงกว่าการใช้ค่าฝนเฉลี่ยรายวันโดยตรง ดังนั้นการแยกองค์ประกอบข้อมูลก่อนพยากรณ์จึงช่วยลดข้อผิดพลาดและเพิ่มความแม่นยำของแบบจำลองได้อย่างมีประสิทธิภาพ

คำสำคัญ: การพยากรณ์ปริมาณฝน, โครงข่ายความจำระยะยาวระยะสั้นแบบซับซ้อน, เทคนิคการแยกองค์ประกอบข้อมูล, อ่างเก็บน้ำลำตะคอง

Abstract

Rainfall data is essential for forecasting, planning, and water resource management. Various rainfall forecasting models are currently used, particularly numerical weather prediction models, which require large amounts of data and computational resources. This study aims to forecast rainfall in the upstream catchment of Lam Ta Khong Dam to support reservoir inflow prediction. The study area is characterized by natural river flow (free flow) without hydraulic structures controlling the flow, so the inflow into the reservoir directly reflects natural runoff, making it highly suitable for studying rainfall and runoff behavior. This study hypothesizes that applying Variational Mode Decomposition (VMD) before forecasting helps the model learn key data characteristics by reducing signal complexity and distinguishing patterns within highly variable data, thereby improving forecasting accuracy. VMD is a signal decomposition technique that separates data into modes across different frequency ranges to reduce data complexity, while Stack Long Short-Term Memory (Stack LSTM) is a deep sequential neural network designed to enhance the model's ability to capture complex temporal features. Therefore, the study is divided into two cases for performance comparison: Case 1 applies Stack LSTM alone, and Case 2 applies VMD combined with Stack LSTM. The results show that using VMD combined with Stack LSTM significantly improves forecasting accuracy. The one-day-ahead rainfall forecasting achieved an R value of 0.928, NSE of 0.861, RMSE of 2.858 millimeters, and MAE of 1.409 millimeters, which are better than using daily average rainfall directly. Therefore, applying data decomposition before forecasting effectively reduces errors and improves the model's accuracy.

Keywords: Rainfall Forecasting, Stack LSTM, Data Decomposition, Lam Takhong Reservoir

1. คำนำ

การบริหารจัดการทรัพยากรน้ำในเขื่อนหรืออ่างเก็บน้ำขนาดใหญ่จำเป็นต้องอาศัยข้อมูลอุทกวิทยาที่แม่นยำ โดยเฉพาะข้อมูลปริมาณน้ำฝนซึ่งส่งผลโดยตรงต่อปริมาณน้ำท่าและระดับน้ำในอ่าง ความแม่นยำในการคาดการณ์ฝนจึงมีบทบาทสำคัญต่อการวางแผนด้านเกษตรกรรม การป้องกันอุทกภัย และการกระจายน้ำเพื่ออุปโภคบริโภคในเขตเมือง ทั้งนี้ การพยากรณ์ปริมาณน้ำฝนมีความสำคัญอย่างยิ่ง เนื่องจากสามารถนำผลที่ได้ไปใช้ในการคาดการณ์ปริมาณน้ำไหลเข้าอ่างเก็บน้ำ ซึ่งเป็นข้อมูลพื้นฐานในการบริหารจัดการน้ำอย่างมีประสิทธิภาพ อย่างไรก็ตาม ลักษณะของข้อมูลฝนมักมีความแปรปรวนสูงและได้รับอิทธิพลจากฤดูกาล จึงจำเป็นต้องเลือกวิธีที่สามารถจัดการกับความซับซ้อนของข้อมูลประเภทนี้ได้อย่างเหมาะสม

ปัจจุบันมีการประยุกต์ใช้เทคนิคการเรียนรู้ของเครื่อง (Machine Learning) ในการพยากรณ์ปริมาณฝนล่วงหน้าอย่างแพร่หลาย [1, 2] โดยเฉพาะแบบจำลอง Long Short-Term Memory (LSTM) ซึ่งเป็นโครงข่ายประสาทเทียมแบบวนซ้ำ (Recurrent Neural Network: RNN) ที่สามารถจัดเก็บและประมวลผลข้อมูลลำดับเวลาได้อย่างมีประสิทธิภาพ [3, 4] จุดเด่นของ LSTM คือการใช้กลไก “ประตู” (Gate) เพื่อคัดกรองข้อมูลสำคัญและลดปัญหา vanishing gradient จึงเหมาะกับข้อมูลที่ซับซ้อนและเปลี่ยนแปลงตามเวลา

นอกจากนี้ ยังมีการประยุกต์ใช้เทคนิคการแยกองค์ประกอบข้อมูล (Data Decomposition) เพื่อเพิ่มประสิทธิภาพในการวิเคราะห์สัญญาณและลดสัญญาณรบกวน ก่อนป้อนข้อมูลเข้าสู่แบบจำลองการเรียนรู้ของเครื่องและแบบจำลองเชิงลึก โดยเฉพาะเทคนิค Variational Mode Decomposition (VMD) ซึ่งสามารถแยกสัญญาณออกเป็นโหมดความถี่ย่อยที่ไม่ทับซ้อนกันในเชิงความถี่ ช่วยให้สามารถแยกโครงสร้างของข้อมูลออกจากความผันผวนหรือ noise ได้อย่างมีประสิทธิภาพ [5]

จากการศึกษางานวิจัยที่ผ่านมา พบว่ามีการประยุกต์ใช้เทคนิค Variational Mode Decomposition (VMD) ร่วมกับแบบจำลองการเรียนรู้ของเครื่องหลายรูปแบบ เช่น Long Short-Term Memory (LSTM), Recurrent Neural Network (RNN), Deep Neural Computation (DNC), Artificial Neural Network (ANN) และ Least Squares Support Vector Machine (LSSVM) ในการพยากรณ์ปริมาณน้ำฝนหรือปริมาณการไหลของน้ำในแม่น้ำและอ่างเก็บน้ำ ข้อมูลที่ผ่านการเตรียมด้วย VMD ช่วยลดความไม่เสถียรของสัญญาณ และแยกรูปแบบเชิงเวลาออกเป็นหลายช่วงความถี่ ส่งผลให้โมเดลสามารถเรียนรู้พฤติกรรมลำดับเวลาได้แม่นยำยิ่งขึ้น ทั้งนี้ แบบจำลองแบบผสม เช่น VMD-LSTM, VMD-ANN และ VMD-LSSVM แสดงให้เห็นถึงประสิทธิภาพที่เหนือกว่าแบบจำลองเดี่ยว โดยเฉพาะในกรณีการพยากรณ์ระยะยาว หรือช่วงเหตุการณ์ที่มีความผันผวนสูงของข้อมูล [6-9]

ในการศึกษานี้ เลือกพื้นที่ลุ่มน้ำบริเวณเหนือเขื่อนลำตะคอง ซึ่งมีความสำคัญต่อการจัดสรรน้ำในภาคตะวันออกเฉียงเหนือตอนล่าง ที่มีบทบาทหลักในการสนับสนุนการอุปโภคบริโภค การเกษตรกรรม และการผลิตไฟฟ้า อย่างไรก็ตาม พื้นที่ดังกล่าวประสบปัญหาน้ำท่วมและภัยแล้งสลับกันเป็นระยะ เนื่องมาจากความแปรปรวนของปริมาณน้ำฝนและปริมาณน้ำไหลเข้าอ่างในแต่ละปี ความจำเป็นในการพยากรณ์ปริมาณน้ำฝนล่วงหน้าจึงมีความสำคัญอย่างยิ่ง โดยเฉพาะอย่างยิ่งในการจัดหาข้อมูลฝนล่วงหน้า สำหรับแบบจำลอง rainfall–runoff model เพื่อใช้ในการประเมินและบริหารจัดการปริมาณน้ำไหลเข้าอ่างเก็บน้ำอย่างมีประสิทธิภาพ

ในการศึกษานี้มีวัตถุประสงค์เพื่อพัฒนาแบบจำลองการพยากรณ์ปริมาณฝนโดยประยุกต์ใช้เทคนิค VMD เพื่อแยกองค์ประกอบข้อมูลก่อนนำเข้าสู่แบบจำลอง LSTM เพื่อให้แบบจำลองสามารถเรียนรู้ลักษณะเชิงเวลาได้อย่างมีประสิทธิภาพยิ่งขึ้น ลดผลกระทบจากความซับซ้อนและความแปรปรวนของข้อมูลต้นฉบับ และเพิ่มความแม่นยำในการพยากรณ์ปริมาณน้ำฝน ซึ่งคาดว่าจะนำไปสู่การสนับสนุนการวางแผนบริหารจัดการน้ำในอ่างเก็บน้ำลำตะคองได้อย่างมีประสิทธิภาพมากยิ่งขึ้น

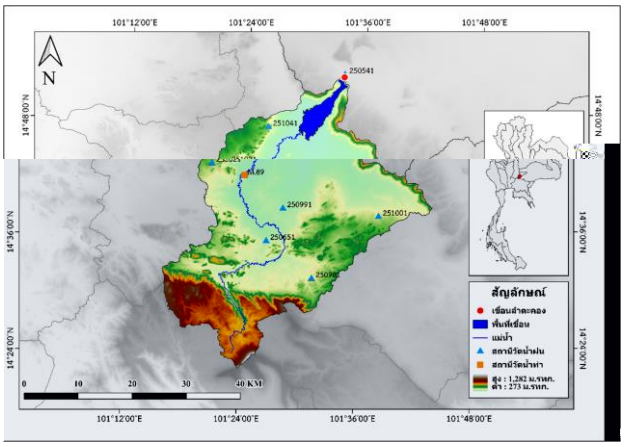
2. อุปกรณ์และวิธีการ

2.1 พื้นที่ศึกษา

พื้นที่ศึกษาอยู่บริเวณ ลุ่มน้ำลำตะคองบริเวณเหนือเขื่อนลำตะคอง ครอบคลุมพื้นที่รวม 8 ตำบล อำเภอปากช่อง จังหวัดนครราชสีมา โดยมีรายละเอียดที่สำคัญดังนี้

2.1.1 สภาพภูมิประเทศ

ลุ่มน้ำลำตะคอง มีขนาดพื้นที่ลุ่มน้ำ 3,352.64 ตารางกิโลเมตร โดยพื้นที่ศึกษาคือลุ่มน้ำลำตะคองบริเวณเหนืออ่างเก็บน้ำลำตะคอง มีพื้นที่ทั้งหมด ประมาณ 1,299.11 ตารางกิโลเมตร คิดเป็นร้อยละ 38.75 ของพื้นที่ลุ่มน้ำลำตะคองทั้งหมด โดยมีพื้นที่ส่วนใหญ่อยู่ในจังหวัดนครราชสีมา สภาพภูมิประเทศส่วนใหญ่เป็นที่ราบแคบบริเวณริมลำน้ำ และ ลักษณะเป็นภูเขาสูงชันทางตอนใต้ของลุ่มน้ำ ดังแสดงในรูปที่ 1 มีแม่น้ำสายหลักของลุ่มน้ำ คือ ลำน้ำลำตะคอง มีความยาวประมาณ 135 กิโลเมตร ซึ่งต้นน้ำของลำน้ำลำตะคองอยู่บริเวณเทือกเขาดงพญาเย็น ไหลขึ้นจากทางทิศตะวันตกเฉียงใต้ ไปสู่ทางทิศตะวันออกเฉียงเหนือ รูปตัดตามยาวของลำน้ำลำตะคอง



รูปที่ 1 สภาพภูมิประเทศบริเวณพื้นที่ศึกษา

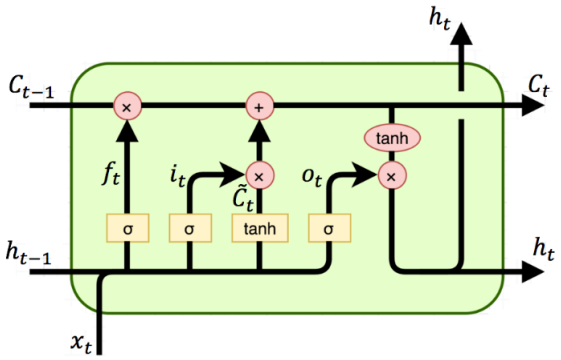
2.1.2 ข้อมูลทั่วไปของอ่างเก็บน้ำลำตะคอง

เขื่อนลำตะคอง เริ่มก่อสร้างเมื่อ พ.ศ. 2512 สร้างแล้วเสร็จเมื่อ พ.ศ. 2517 ตั้งอยู่ที่ ตำบลคลองไผ่ อำเภอสีคิ้ว จังหวัดนครราชสีมา ตัวเขื่อนเป็นเขื่อนดินสูง 40.30 เมตร สันเขื่อนยาว 527 เมตร กว้าง 10 เมตร อ่างเก็บน้ำเหนือเขื่อนมีความยาวตลอดลำน้ำ 19 กิโลเมตร มีพื้นที่ 277,000 ไร่ มีปริมาตรเก็บกักสูงสุด 445 ล้าน ลบม. ปริมาตรเก็บกักปกติ 314.49 ล้าน ลบม. ปริมาตรเก็บกักต่ำสุด 22.70 ล้าน ลบม.

2.2 ทฤษฎีที่เกี่ยวข้อง

2.2.1 แบบจำลองหน่วยความจำระยะยาวแบบสั้น (LSTM)

Long Short-Term Memory (LSTM) เป็นโครงข่ายประสาทเทียมแบบวนซ้ำ (Recurrent Neural Network: RNN) ที่พัฒนาขึ้นเพื่อแก้ปัญหา vanishing gradient ในการเรียนรู้ข้อมูลลำดับระยะยาว โดยใช้กลไกที่เรียกว่า ประตู (Gate) เพื่อควบคุมการไหลของข้อมูลภายในหน่วยความจำ ทำให้สามารถรักษาข้อมูลสำคัญไว้ได้ในช่วงเวลาที่เหมาะสม LSTM ประกอบด้วยสามประตูหลัก ได้แก่ Forget Gate, Input Gate และ Output Gate [10] ซึ่งสามารถอธิบายรายละเอียดการทำงานได้ ตามสมการที่ (1)–(6) และ มีโครงสร้างแบบจำลอง ตามรูปที่ 2



รูปที่ 2 โครงสร้างแบบจำลอง LSTM

ที่มา: Colah, 2015

ในกระบวนการทำงานของ LSTM นั้น หน่วยความจำในแต่ละช่วงเวลาจะถูกควบคุมโดยกลไกที่เรียกว่า ประตู (Gate) ซึ่งมีทั้งหมด 3 ประตูหลัก ได้แก่ Forget Gate, Input Gate และ Output Gate โดยจะมีบทบาทเฉพาะในการควบคุมข้อมูลที่ไหลเข้าและออกจากหน่วยความจำ เริ่มจาก Forget Gate ซึ่งมีหน้าที่ตัดสินว่าข้อมูลจากช่วงเวลาก่อนหน้าควรถูกลบหรือเก็บไว้ในหน่วยความจำ สำหรับการคำนวณค่าของ Forget Gate จะใช้ข้อมูลนำเข้าในช่วงเวลาปัจจุบัน x_t ร่วมกับค่าของ hidden state จากช่วงเวลาก่อนหน้า h_{t-1} ผ่านฟังก์ชัน sigmoid ดังแสดงในสมการที่ (1)

$$f_t = \sigma\left(W_f[h_{t-1}, x_t] + b_f\right) \quad (1)$$

ค่าที่ได้จาก f_t จะอยู่ในช่วงระหว่าง 0 ถึง 1 เพื่อแสดงระดับความสำคัญของข้อมูลเดิมในหน่วยความจำ ต่อมา Input Gate มีหน้าที่ควบคุมการเพิ่มข้อมูลใหม่เข้าสู่หน่วยความจำ โดยกระบวนการจะประกอบด้วยสองส่วน ได้แก่ การตัดสินว่าจะรับข้อมูลใหม่มากน้อยเพียงใด และการสร้างค่าข้อมูลใหม่ที่เสนอเข้าสู่หน่วยความจำ โดยสมการของ Input Gate ดังแสดงในสมการที่ (2)

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \quad (2)$$

และการคำนวณข้อมูลใหม่ที่เสนอเข้าสู่หน่วยความจำจะแสดงในสมการที่ (3) เมื่อได้ผลลัพธ์จากทั้งสองส่วนแล้ว ระบบจะนำมาผสมกับข้อมูลเดิมเพื่อลงทะเบียนเป็นสถานะหน่วยความจำใหม่ C_t ดังแสดงในสมการที่ (4)

$$\tilde{C}_t = \tanh(W_C[h_{t-1}, x_t] + b_C) \quad (3)$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \quad (4)$$

จากนั้นจึงเข้าสู่ขั้นตอนของ Output Gate ซึ่งจะกำหนดว่าข้อมูลใดจากหน่วยความจำควรถูกส่งออกไปใช้งาน โดยมีการคำนวณค่าของ output gate ตามสมการที่ (5)

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad (5)$$

และสุดท้ายจะคำนวณค่าของ hidden state ที่เป็นผลลัพธ์ของ LSTM ในช่วงเวลานั้น ดังแสดงในสมการที่ (6)

$$h_t = o_t \cdot \tanh(C_t) \quad (6)$$

2.2.2 การแยกองค์ประกอบข้อมูลด้วยวิธี (VMD)

การแยกองค์ประกอบข้อมูล (Data Decomposition) เป็นกระบวนการหนึ่งที่ช่วยแยกข้อมูลต้นฉบับออกเป็นส่วนย่อย ๆ ซึ่งสะท้อนพฤติกรรมในแต่ละความถี่หรือรูปแบบเฉพาะตัวของข้อมูล การแยกข้อมูลก่อนนำไปวิเคราะห์หรือพยากรณ์จึงมีประโยชน์อย่างยิ่ง โดยเฉพาะในระบบที่มีความไม่แน่นอนสูง เช่น ข้อมูลเวลา (Time Series) ที่มีความซับซ้อน

Variational Mode Decomposition (VMD) เป็นเทคนิคการแยกสัญญาณที่สามารถแบ่งข้อมูลออกเป็นหลายโหมดย่อย (sub-signals) ที่ไม่ทับซ้อนกันในเชิงความถี่ โดยโหมดที่ได้จะมีพฤติกรรมความถี่ที่ไม่ซ้อนทับกัน [5]

2.2.3 การประเมินประสิทธิภาพแบบจำลอง

การประเมินประสิทธิภาพของแบบจำลองในงานวิจัยนี้ใช้ตัวชี้วัดทางสถิติ 4 รายการ ได้แก่ สัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient: R) ค่าประสิทธิภาพของแบบจำลอง (Nash–Sutcliffe Efficiency: NSE) ค่าความคลาดเคลื่อนเฉลี่ยรากที่สอง (Root Mean Square Error: RMSE) และค่าเฉลี่ยสัมบูรณ์ของความคลาดเคลื่อน (Mean Absolute Error: MAE) เพื่อวัดความแม่นยำของผลการพยากรณ์เมื่อเปรียบเทียบกับค่าจริง ตามสมการที่ (7) ถึง (10)

$$R = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}\sqrt{\sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}})^2}} \quad (7)$$

$$\mathrm{NSE} = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (8)$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (9)$$

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (10)$$

โดยที่ y_i แทนค่าจริงของข้อมูลในช่วงเวลา i , $\hat{y}_i$ แทนค่าที่แบบจำลองพยากรณ์ได้ในช่วงเวลาเดียวกัน $\bar{y}$ และ$\bar{\hat{y}}$ แทนค่าเฉลี่ยของค่าจริงและค่าพยากรณ์ตามลำดับ ส่วน n คือจำนวนข้อมูลทั้งหมดที่ใช้ในการคำนวณ โดย R ใช้วัดความสัมพันธ์เชิงเส้นระหว่างค่าที่พยากรณ์กับค่าจริง, NSE ใช้วัดความแม่นยำของแนวโน้มผลการพยากรณ์, RMSE และ MAE ใช้วัดขนาดของความคลาดเคลื่อนระหว่างค่าพยากรณ์กับค่าจริง โดย RMSE ให้ความสำคัญกับข้อผิดพลาดขนาดใหญ่ ส่วน MAE เป็นการวัดข้อผิดพลาดเฉลี่ยในเชิงสัมบูรณ์ ทั้งนี้ ในการประเมินประสิทธิภาพของแบบจำลองในงานวิจัยนี้ ใช้ค่าประสิทธิภาพของแบบจำลอง (Nash-Sutcliffe Efficiency; NSE) เป็นตัวชี้วัดหลัก โดยเกณฑ์การประเมินอ้างอิงตามแนวทางของ Moriasi et al. (2007) ซึ่งแบ่งระดับประสิทธิภาพตามช่วงค่าของ NSE ดังแสดงในตารางที่ 1

ตารางที่ 1 เกณฑ์การประเมินประสิทธิภาพของแบบจำลองด้วยค่า NSE

เกณฑ์	คุณภาพการพยากรณ์
0.75 < NSE ≤ 1.00	ดีมาก (Very Good)
0.65 < NSE ≤ 0.75	ดี (Good)
0.50 < NSE ≤ 0.65	พอใช้ (Satisfactory)
NSE ≤ 0.50	แย่ (Unsatisfactory)

2.3 ขั้นตอนการดำเนินงาน

สำหรับการศึกษานี้ การพยากรณ์ปริมาณน้ำฝนเฉลี่ยของลุ่มน้ำบริเวณเหนือเขื่อนลำตะคอง โดยใช้แบบจำลองหน่วยความจำระยะยาวแบบสั้น ซึ่งมีการใช้ ข้อมูลปริมาณน้ำฝนรายวัน ข้อมูลสภาพอากาศรายวัน(ข้อมูลความชื้นสัมพัทธ์รายวัน ข้อมูลอุณหภูมิเฉลี่ยรายวัน และ ข้อมูลน้ำระเหยถาดรายวัน) ซึ่งสามารถแบ่งกรณีศึกษาได้ 2 กรณี ดังแสดงตามรูปที่ 3



รูปที่ 3 แผนภาพแสดงขั้นตอนการดำเนินงานวิจัย

2.4 รวบรวมข้อมูลที่ใช้ในการศึกษา

รวบรวมข้อมูลปริมาณน้ำฝนรายวัน จากสถานีฝนของกรมชลประทานในลุ่มน้ำที่อยู่บริเวณเหนืออ่างเก็บน้ำลำตะคอง และ ข้อมูลสภาพอากาศรายวัน (ข้อมูลความชื้นสัมพัทธ์รายวัน ข้อมูลอุณหภูมิเฉลี่ยรายวัน และข้อมูลน้ำระเหยถาดรายวัน) จากสถานีอุตุนิยมวิทยา นครราชสีมา ปากช่อง (สถานีเกษตร) ดังแสดงใน ตารางที่ 2

ตารางที่ 2 รายละเอียดข้อมูลในพื้นที่ศึกษา

ลำดับ	รหัสสถานี	ชนิดข้อมูล (รายวัน)	พิกัด		สถิติข้อมูล
			ละติจูด (น.)	ลองจิจูด (อ.)	
1	250541	ปริมาณน้ำฝน	14.8701	101.5610	2550-2564
2	250651	ปริมาณน้ำฝน	14.5845	101.4519	2550-2564
3	250981	ปริมาณน้ำฝน	14.5197	101.5297	2550-2564
4	250991	ปริมาณน้ำฝน	14.6403	101.4808	2550-2564
5	251001	ปริมาณน้ำฝน	14.6258	101.6446	2550-2564
6	251021	ปริมาณน้ำฝน	14.7181	101.3592	2550-2564
7	251041	ปริมาณน้ำฝน	14.7800	101.4300	2550-2564
8	431301	ความชื้นสัมพัทธ์	14.6979	101.4152	2550-2564
9	431301	อุณหภูมิเฉลี่ย	14.6979	101.4152	2550-2564
10	431301	น้ำระเหยถาด	14.6979	101.4152	2550-2564

2.5 การเตรียมข้อมูลเพื่อนำเข้าแบบจำลอง

ข้อมูลปริมาณน้ำฝนรายวันจากสถานีตรวจวัดในพื้นที่ศึกษา ก่อนนำไปใช้ในกระบวนการสร้างแบบจำลอง จำเป็นต้องผ่านขั้นตอนการตรวจสอบความถูกต้องและการปรับปรุงข้อมูล เพื่อให้มั่นใจว่าเป็นข้อมูลที่มีความน่าเชื่อถือ และเหมาะสมสำหรับใช้ในการฝึกและทดสอบแบบจำลอง โดยในงานวิจัยนี้ได้ดำเนินการใน 3 ขั้นตอนหลัก ได้แก่ การตรวจสอบความถูกต้องของข้อมูล การปรับปรุงข้อมูล และการแบ่งชุดข้อมูลสำหรับการเรียนรู้ของแบบจำลอง ดังนี้

2.5.1 ตรวจสอบความถูกต้องของข้อมูล

ข้อมูลปริมาณน้ำฝนรายวันจากสถานีวัดฝนจะได้รับการตรวจสอบความสอดคล้องและความน่าเชื่อถือเบื้องต้น โดยใช้วิธี Double Mass Curve (DMC) ซึ่งเป็นเทคนิคที่นิยมใช้ในการตรวจสอบความต่อเนื่องและความคงที่ของความสัมพันธ์ระหว่างข้อมูลจากสถานีวัดที่ต้องการตรวจสอบ กับสถานีวัดฝนโดยรอบในบริเวณใกล้เคียง โดยจะทำการพล็อตกราฟของผลรวมสะสมของปริมาณน้ำฝนจากสถานีที่สนใจ เปรียบเทียบกับผลรวมสะสมของสถานีเปรียบเทียบ หากกราฟที่ได้มีลักษณะเป็นเส้นตรงหรือมีแนวโน้มต่อเนื่อง แสดงว่าข้อมูลมีความน่าเชื่อถือ แต่หากกราฟแสดงแนวโน้มที่ผิดปกติหรือมีจุดเปลี่ยนแปลงชัดเจน อาจหมายถึงข้อมูลช่วงนั้นมีความคลาดเคลื่อนหรือเกิดความผิดพลาดจากการวัด ซึ่งจะต้องพิจารณาเพื่อทำการปรับแก้ไขในขั้นตอนถัดไป

2.5.2 การปรับปรุงข้อมูล

ในกรณีที่ตรวจสอบพบข้อมูลปริมาณน้ำฝนที่ขาดหายหรือมีค่าผิดปกติ จะทำการปรับปรุงข้อมูลโดยใช้วิธี Inverse Distance Weighting (IDW) ซึ่งเป็นเทคนิคการประมาณค่าที่อาศัยหลักการถ่วงน้ำหนักตามระยะทาง โดยจะใช้ข้อมูลจากสถานีรอบข้างที่อยู่ในบริเวณใกล้เคียงกันและมีข้อมูลในวันเดียวกันกับข้อมูลที่ขาดหาย เพื่อนำมาประมาณค่าปริมาณน้ำฝนของสถานีเป้าหมาย โดยสถานีที่อยู่ใกล้จะมีอิทธิพลมากกว่าสถานีที่อยู่ไกล สูตรการคำนวณ IDW จะพิจารณาน้ำหนักกลับตามกำลังสองของระยะทาง เพื่อให้ได้ค่าที่สอดคล้องกับสภาพพื้นที่และความสัมพันธ์เชิงพื้นที่ของข้อมูล

2.5.3 การแบ่งชุดข้อมูล

หลังจากผ่านกระบวนการตรวจสอบและปรับปรุงข้อมูลเรียบร้อยแล้ว จะทำการแบ่งชุดข้อมูลสำหรับการเรียนรู้ของแบบจำลอง LSTM โดยข้อมูลปริมาณน้ำฝนที่ใช้ครอบคลุมระยะเวลา 14 ปี ตั้งแต่ปี พ.ศ. 2550 ถึง พ.ศ. 2564 โดยพิจารณาถึงความต่อเนื่องของข้อมูลเชิงเวลา เพื่อหลีกเลี่ยงการตัดข้อมูลกลางฤดูซึ่งอาจส่งผลกระทบต่อการเรียนรู้ของแบบจำลอง ข้อมูลถูกแบ่งตามลักษณะปีน้ำ (Hydrological Year) โดยกำหนดช่วงเวลาเริ่มตั้งแต่วันที่ 1 เมษายนของปี ถึงวันที่ 31 มีนาคมของปีถัดไป แบ่งออกเป็น 3 ส่วน ได้แก่

2.7 การฝึกแบบจำลอง

ในการฝึกแบบจำลอง LSTM สำหรับการพยากรณ์ปริมาณฝนเฉลี่ยรายวัน ได้มีการกำหนดค่าพารามิเตอร์หลัก (hyperparameters) ที่มีผลต่อประสิทธิภาพของการเรียนรู้ ได้แก่ จำนวนหน่วยในแต่ละชั้นของ LSTM (node1, node2), ขนาดชุดข้อมูลย่อย (batch size), ขนาดบัฟเฟอร์ (buffer size), จำนวนรอบการฝึก (epochs), จำนวนขั้นต่อรอบ (steps per epoch) และอัตราการเรียนรู้ (learning rate) โดยใช้วิธีการปรับจูนค่าพารามิเตอร์ด้วยตนเอง (manual tuning) ผ่านการทดลองค่าหลายชุด แล้วเลือกค่าที่ให้ผลลัพธ์ดีที่สุดบนชุดข้อมูลตรวจสอบ (validation set) ช่วงค่าที่พิจารณาในการทดลองสรุปไว้ในตารางที่ 2

ตารางที่ 2 รายละเอียดช่วงในการปรับค่าพารามิเตอร์ของแบบจำลอง

Hyperparameters	Range of Values
node1	[16, 32, 64, 128, 256]
node2	[16, 32, 64, 128, 256]
batch size	[16, 32, 64, 128, 256]
buffer size	[1000, 3000, 5000]
epochs	[100, 150]
steps per epoch	[100, 150]
learning rate	[0.0001, 0.001, 0.01, 0.1]

สำหรับการปรับน้ำหนักของแบบจำลอง ใช้ RMSprop เป็นตัวปรับค่าพารามิเตอร์ (optimizer) ซึ่งเหมาะสมกับข้อมูลลำดับเวลา และใช้ค่าเฉลี่ยสัมบูรณ์ของความคลาดเคลื่อน (Mean Absolute Error: MAE) เป็นฟังก์ชันความสูญเสีย (loss function) ในการเรียนรู้ของแบบจำลองในแต่ละรอบ ทั้งนี้ จากผลการฝึกแบบจำลองพบว่า ค่าความสูญเสียลดลงอย่างต่อเนื่องในช่วงต้นของการฝึก และมีแนวโน้มคงที่ในช่วงท้ายของการเรียนรู้ แสดงให้เห็นว่าแบบจำลองสามารถเรียนรู้ได้อย่างมีเสถียรภาพ ไม่เกิดปัญหา overfitting หรือ underfitting อย่างชัดเจน และเหมาะสมสำหรับการนำไปใช้ในการพยากรณ์ข้อมูลลำดับเวลา

2.8 ประเมินประสิทธิภาพแบบจำลอง

การประเมินประสิทธิภาพของแบบจำลองดำเนินการโดยใช้ชุดข้อมูลทดสอบ (testing set) ซึ่งไม่ถูกนำไปใช้ในการฝึกแบบจำลอง เพื่อทดสอบความสามารถของแบบจำลองในการพยากรณ์ปริมาณฝนล่วงหน้า โดยใช้ตัวชี้วัดทางสถิติ 4 รายการ ได้แก่ สัมประสิทธิ์สหสัมพันธ์ (R), ค่าประสิทธิภาพของแบบจำลอง (Nash–Sutcliffe Efficiency: NSE), ค่าความคลาดเคลื่อนเฉลี่ยรากที่สอง (Root Mean Square Error: RMSE) และค่าเฉลี่ยสัมบูรณ์ของความคลาดเคลื่อน (Mean Absolute Error: MAE) ซึ่งใช้ในการเปรียบเทียบผลการพยากรณ์กับค่าจริง เพื่อประเมินประสิทธิภาพของแบบจำลองทั้งในระดับวันเดียวล่วงหน้า (T+1) และหลายวันล่วงหน้า (T+2 ถึง T+7)

3. ผลและการวิจารณ์ผลการทดลอง

3.1 ผลการประเมินแบบจำลองกรณีที่ 1

การประเมินประสิทธิภาพของแบบจำลองในกรณีที่ 1 ซึ่งใช้ LSTM เพียงอย่างเดียวโดยไม่มีการเตรียมข้อมูลล่วงหน้าด้วยเทคนิคใด ๆ ได้ดำเนินการด้วยชุดข้อมูลทดสอบ (testing set) เช่นเดียวกับกรณีอื่น เพื่อประเมินความสามารถในการพยากรณ์ปริมาณฝนเฉลี่ยรายวันล่วงหน้าในช่วงเวลา 7 วัน (T+1 ถึง T+7) ผลการคำนวณตัวชี้วัดทางสถิติ ได้แก่ ค่าสัมประสิทธิ์สหสัมพันธ์ (R), ค่าประสิทธิภาพของแบบจำลอง (NSE), ค่าความคลาดเคลื่อนเฉลี่ยรากที่สอง (RMSE) และค่าเฉลี่ยสัมบูรณ์ของความคลาดเคลื่อน (MAE) สรุปใน ตารางที่ 3 โดยจะเห็นว่า ค่าตัวชี้วัดในภาพรวมอยู่ในระดับต่ำ โดยเฉพาะค่า R และ NSE ที่ไม่แสดงความสัมพันธ์ที่ชัดเจนระหว่างค่าพยากรณ์กับค่าจริง เช่น ในช่วง T+1 ได้ค่า R เพียง 0.258 และ NSE มีค่าใกล้ศูนย์ (−0.009) แสดงให้เห็นว่าแบบจำลองไม่สามารถอธิบายพฤติกรรมของข้อมูลได้ดีนักซึ่งอยู่ในเกณฑ์แย่ ตามตารางที่1 ขณะที่ค่า RMSE และ MAE อยู่ในระดับสูงถึง 7.69 และ 2.64 มิลลิเมตร ตามลำดับ ซึ่งแสดงถึงความคลาดเคลื่อนของแบบจำลองในการพยากรณ์ข้อมูลรายวัน

ตารางที่ 3 ผลการประเมินแบบจำลองรายวัน กรณีที่1 (T+1 ถึง T+7)

Day	R	NSE	RMSE (mm)	MAE (mm)
1	0.258	-0.009	7.687	2.640
2	0.167	-0.066	7.899	2.766
3	0.146	-0.106	8.048	2.776
4	0.147	-0.103	8.037	2.791
5	0.164	-0.100	8.026	2.788
6	0.111	-0.124	8.115	2.797
7	0.179	-0.122	8.104	2.799

3.2 ผลการประเมินแบบจำลองกรณีที่ 2

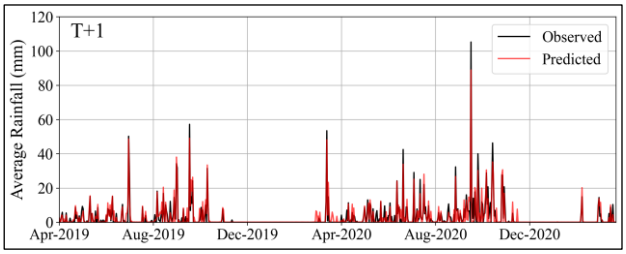
การประเมินประสิทธิภาพของแบบจำลองในกรณีที่ 2 ซึ่งใช้ LSTM ร่วมกับการเตรียมข้อมูลล่วงหน้าด้วยเทคนิคการแยกองค์ประกอบข้อมูลแบบ Variational Mode Decomposition (VMD) เพื่อช่วยแยกสัญญาณข้อมูลฝนออกเป็นหลายโหมดก่อนนำไปใช้ฝึกแบบจำลอง เพื่อช่วยให้แบบจำลองสามารถเรียนรู้ความสัมพันธ์ของข้อมูลเชิงเวลาได้ง่ายและมีประสิทธิภาพยิ่งขึ้น การประเมินประสิทธิภาพของแบบจำลองดำเนินการโดยใช้ชุดข้อมูลทดสอบ (testing set) ซึ่งเป็นข้อมูลที่ไม่เคยถูกใช้ในการฝึกหรือปรับพารามิเตอร์ของแบบจำลองมาก่อน โดยประเมินผลลัพธ์ของการพยากรณ์ในช่วงเวลา 7 วันล่วงหน้า (T+1 ถึง T+7) ด้วยตัวชี้วัดทางสถิติ 4 รายการ ได้แก่ สัมประสิทธิ์สหสัมพันธ์ (R), ค่าประสิทธิภาพของแบบจำลอง (Nash–Sutcliffe Efficiency: NSE), ค่าความคลาดเคลื่อนเฉลี่ยรากที่สอง (Root Mean Square Error: RMSE) และค่าเฉลี่ยสัมบูรณ์ของความคลาดเคลื่อน (Mean Absolute Error: MAE) ผลการประเมินประสิทธิภาพของแบบจำลองในกรณีที่ 2 สรุปในตารางที่ 4 โดยจะเห็นว่าแบบจำลองให้ผลการพยากรณ์ที่แม่นยำที่สุดในวันที่ 1 (T+1) โดยมีค่า R เท่ากับ 0.928 และค่า NSE เท่ากับ 0.861 ซึ่งอยู่ในเกณฑ์ที่ดีมาก ตามตารางที่1 ขณะที่ค่า

RMSE และ MAE อยู่ในระดับต่ำ คือ 2.858 มิลลิเมตร และ 1.409 มิลลิเมตร ตามลำดับ อย่างไรก็ตาม เมื่อจำนวนวันในการพยากรณ์ล่วงหน้าเพิ่มขึ้น ความแม่นยำของแบบจำลองมีแนวโน้มลดลง โดยในวันที่ 7 (T+7) ค่า R ลดลงเหลือ 0.631 และค่า NSE เหลือ 0.308 ขณะที่ RMSE และ MAE เพิ่มขึ้นเป็น 6.364 มิลลิเมตร และ 3.074 มิลลิเมตร ตามลำดับ ซึ่งสะท้อนถึงความไม่แน่นอนที่เพิ่มขึ้นเมื่อต้องพยากรณ์ในช่วงระยะยาว

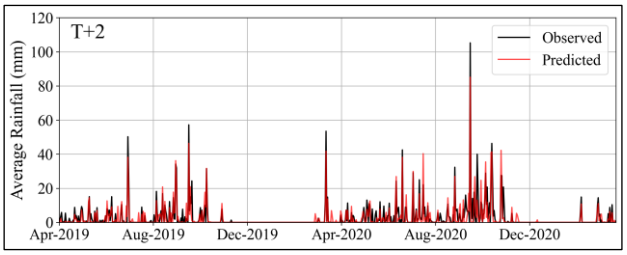
ตารางที่ 4 ผลการประเมินแบบจำลองรายวัน กรณีที่2 (T+1 ถึง T+7)

Day	R	NSE	RMSE (mm)	MAE (mm)
1	0.928	0.861	2.858	1.409
2	0.875	0.762	3.730	1.700
3	0.868	0.737	3.924	2.127
4	0.832	0.671	4.387	2.276
5	0.786	0.610	4.782	2.149
6	0.732	0.508	5.368	2.299
7	0.631	0.308	6.364	3.074

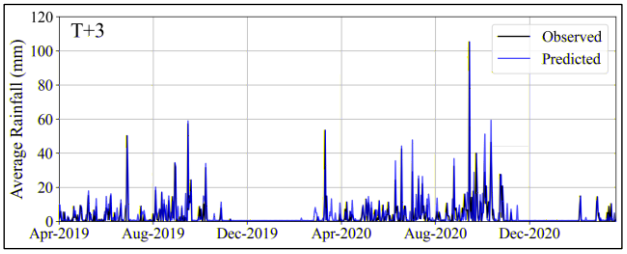
จากการแสดงผลใน รูปที่ 4 ถึง 6 ซึ่งแสดงกราฟเปรียบเทียบค่าฝนเฉลี่ยรายวันระหว่างค่าที่สังเกตจริง (Observed) และค่าที่แบบจำลองพยากรณ์ได้ (Predicted) สำหรับกรณี T+1 ถึง T+3 จะเห็นว่าแบบจำลองสามารถติดตามแนวโน้มของข้อมูลจริงได้ค่อนข้างดี โดยเฉพาะในช่วงฤดูฝนที่มีปริมาณฝนเพิ่มสูงขึ้น อย่างไรก็ตาม เมื่อระยะเวลาการพยากรณ์เพิ่มขึ้นจาก 1 วัน ไปสู่ 2 และ 3 วัน พบว่าความแม่นยำของผลพยากรณ์มีแนวโน้มลดลง โดยมีลักษณะเบี่ยงเบนจากค่าจริงมากขึ้นในบางช่วง โดยเฉพาะช่วงฝนตกหนักเฉียบพลัน ซึ่งแสดงถึงความไม่แน่นอนของข้อมูลที่มีความแปรปรวนสูงและคาดการณ์ได้ยาก นอกจากนี้ยังพบว่า เมื่อค่าการพยากรณ์ระยะเวลาเพิ่มขึ้นลักษณะของกราฟจุดสูงสุด (peak) มีแนวโน้มลดลงเมื่อเทียบกับค่าจริง แสดงถึงข้อจำกัดของแบบจำลองในการพยากรณ์เหตุการณ์ฝนที่รุนแรงในช่วงระยะยาว อย่างไรก็ตาม ในบางช่วงเวลา แบบจำลองมีการพยากรณ์ค่าฝนที่สูงกว่าค่าจริง ซึ่งอาจสะท้อนถึงการตอบสนองต่อสัญญาณบางประเภทที่ขยายจากโหมดความถี่ของข้อมูลเดิมที่ผ่านการแยกองค์ประกอบด้วย VMD



รูปที่ 4 กราฟเปรียบเทียบค่าจริงและค่าที่พยากรณ์ สำหรับ T+1

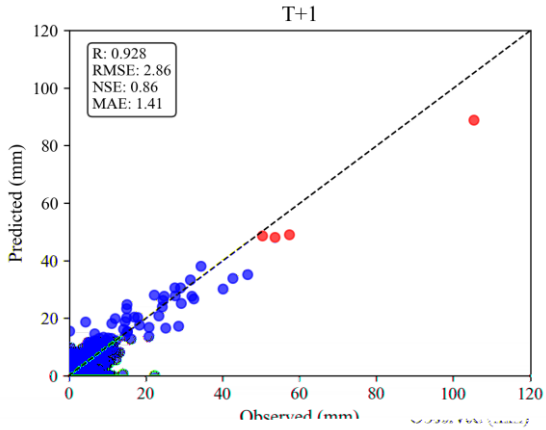


รูปที่ 5 กราฟเปรียบเทียบค่าจริงและค่าที่พยากรณ์ สำหรับ T+2

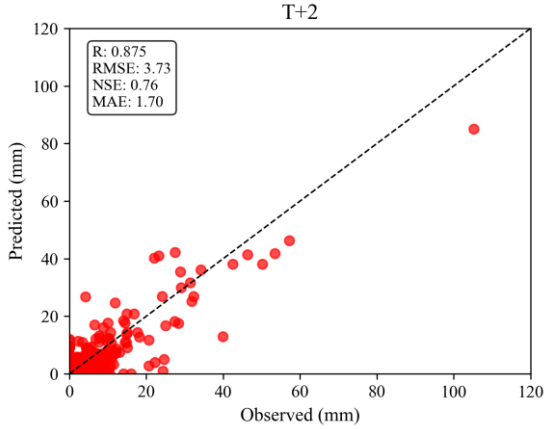


รูปที่ 6 กราฟเปรียบเทียบค่าจริงและค่าที่พยากรณ์ สำหรับ T+3

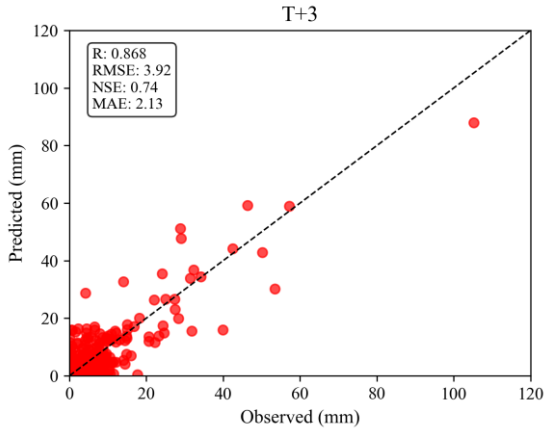
ขณะเดียวกัน รูปที่ 7 ถึง 9 แสดงความสัมพันธ์ระหว่างค่าที่สังเกตจริงและค่าที่แบบจำลองพยากรณ์ได้ สำหรับช่วงการพยากรณ์ล่วงหน้า 1 ถึง 3 วัน โดยจุดข้อมูลส่วนใหญ่ในทั้งสามกรณีปรากฏอยู่ใกล้กับเส้นทแยงมุม ซึ่งแสดงให้เห็นถึงความสามารถของแบบจำลองในการประมาณค่าฝนเฉลี่ยได้ใกล้เคียงกับค่าจริง โดยเฉพาะในช่วง T+1 ซึ่งมีค่าตัวชี้วัดทางสถิติที่อยู่ในระดับดีมาก อย่างไรก็ตาม เมื่อระยะเวลาการพยากรณ์เพิ่มขึ้น ค่า R และ NSE มีแนวโน้มลดลง ขณะที่ RMSE และ MAE เพิ่มขึ้น ซึ่งสอดคล้องกับลักษณะของการกระจายจุดที่แสดงถึงความแม่นยำที่ลดลง โดยเฉพาะในกรณี T+3 ที่เริ่มปรากฏจุดเบี่ยงเบนจากเส้นทแยงมุมมากขึ้น



รูปที่ 7 Scatter plot ค่าจริงเทียบกับค่าที่พยากรณ์ สำหรับ T+1



รูปที่ 8 Scatter plot ค่าจริงเทียบกับค่าที่พยากรณ์ สำหรับ T+2



รูปที่ 9 Scatter plot ค่าจริงเทียบกับค่าที่พยากรณ์ สำหรับ T+3

3.3 การเปรียบเทียบผลการทดลอง

จากผลการทดลองทั้งสองกรณี พบว่าแบบจำลองที่ผ่านกระบวนการแยกองค์ประกอบข้อมูลด้วยเทคนิค Variational Mode Decomposition (VMD) ก่อนนำเข้าสู่แบบจำลอง LSTM (กรณีที่ 2) ให้ผลการพยากรณ์ที่ดีกว่าแบบจำลองที่ใช้ LSTM เพียงอย่างเดียว (กรณีที่ 1) อย่างชัดเจนในทุกตัวชี้วัดทางสถิติ โดยเฉพาะในช่วงพยากรณ์ล่วงหน้า 1 วัน (T+1) ซึ่งกรณีที่ 2 มีค่า R เท่ากับ 0.928 และค่า NSE เท่ากับ 0.861 ขณะที่กรณีที่ 1 มีค่า R เท่ากับ 0.258 และ NSE ติดลบ ซึ่งสะท้อนถึงความแม่นยำที่แตกต่างกันอย่างมีนัยสำคัญ นอกจากนี้ RMSE และ MAE ของกรณีที่ 2 ยังมีค่าต่ำกว่ากรณีที่ 1 อย่างต่อเนื่องตลอดช่วงเวลา T+1 ถึง T+7 ซึ่งแสดงถึงความคลาดเคลื่อนที่ลดลงและศักยภาพของการแยกสัญญาณย่อยในการช่วยให้แบบจำลองเรียนรู้คุณลักษณะเชิงเวลาได้แม่นยำยิ่งขึ้น จากกราฟแสดงผลช่วง T+1 ถึง T+3 จะเห็นว่าเส้นกราฟของค่าพยากรณ์ในกรณีที่ 2 สามารถติดตามแนวโน้มของข้อมูลจริงได้ใกล้เคียงกัน โดยเฉพาะในช่วงฤดูฝนที่มีปริมาณฝนสูงซึ่งมักเป็นช่วงเวลาที่แบบจำลองทั่วไปคาดการณ์ได้ยาก อย่างไรก็ตาม แม้แบบจำลองกรณีที่ 2 จะมีแนวโน้มที่ค่าพยากรณ์ลดลงเมื่อระยะเวลาพยากรณ์ยาวขึ้น แต่ยังสามารถสะท้อนทิศทางข้อมูลได้ดี และแม้บางช่วงของเหตุการณ์ฝนตกหนักเฉียบพลันจะมีค่าพยากรณ์ที่สูงกว่าค่าจริงเล็กน้อย แต่โดยรวมยังแสดงให้เห็นว่าแบบจำลองที่ผสาน VMD ช่วยเพิ่มความสามารถในการพยากรณ์ได้อย่างมีประสิทธิภาพมากกว่าแบบจำลอง LSTM เพียงอย่างเดียว

4. บทสรุป

งานวิจัยนี้มีที่มาจากความสำคัญของการพยากรณ์ปริมาณฝนเฉลี่ยรายวัน ซึ่งเป็นปัจจัยสำคัญในการบริหารจัดการน้ำในระดับลุ่มน้ำ การคาดการณ์ที่แม่นยำช่วยเพิ่มประสิทธิภาพในการจัดเก็บและระบายน้ำ รวมถึงการเตรียมรับมือกับน้ำท่วม วัตถุประสงค์ของงานวิจัย คือ การพัฒนาแบบจำลองพยากรณ์ปริมาณฝนเฉลี่ยรายวัน โดยประยุกต์ใช้เทคนิค VMD ร่วมกับแบบจำลอง LSTM เพื่อเพิ่มความแม่นยำในการพยากรณ์ โดยมีการเปรียบเทียบกับแบบจำลอง LSTM พื้นฐาน ที่ไม่ได้ใช้เทคนิค VMD โดยหลักการสำคัญที่ใช้ในงานวิจัยนี้ คือ การลดความซับซ้อนของข้อมูลด้วยการแยกสัญญาณข้อมูลออกเป็นโหมดย่อย ๆ ผ่านเทคนิค VMD เพื่อให้แบบจำลอง LSTM ที่มีความสามารถในเรียนรู้ลักษณะข้อมูลลำดับเวลา ให้สามารถเรียนรู้ได้อย่างมีประสิทธิภาพมากขึ้น

กระบวนการวิจัยเริ่มจากการรวบรวมข้อมูลปริมาณฝนเฉลี่ยรายวันและตัวแปรสภาพอากาศในพื้นที่ศึกษา จากนั้นดำเนินการตรวจสอบความถูกต้องของข้อมูลด้วยวิธี Double Mass Curve และเติมข้อมูลขาดหายด้วยเทคนิค IDW และ Linear Interpolation ต่อมาแบ่งชุดข้อมูลออกเป็นชุดฝึก ชุดตรวจสอบ และชุดทดสอบตามปีน้ำ พร้อมทั้งทำการปรับสเกลข้อมูลให้อยู่ในช่วง [-1,1] ในกรณีที่ประยุกต์ใช้ VMD จะทำการแยกองค์ประกอบข้อมูลก่อนการป้อนเข้าสู่แบบจำลอง LSTM ซึ่งใช้โครงสร้างสำหรับการพยากรณ์หลายช่วงเวลา (Multi-step Forecasting) ทำนายค่าปริมาณฝนเฉลี่ยรายวันล่วงหน้าในช่วงเวลา 7 วัน (T+1 ถึง T+7) ผลการพยากรณ์ประเมินด้วยตัวชี้วัดทางสถิติ ได้แก่ ค่าสัมประสิทธิ์สหสัมพันธ์ (R), ค่าประสิทธิภาพของแบบจำลอง (NSE), ค่าความคลาดเคลื่อนเฉลี่ยรากที่สอง (RMSE) และค่าความคลาดเคลื่อนเฉลี่ยสัมบูรณ์ (MAE)

จากการเปรียบเทียบประสิทธิภาพของแบบจำลอง พบว่าแบบจำลองที่ใช้ข้อมูลจากการแยกองค์ประกอบร่วมกับ LSTM ให้ผลการพยากรณ์ที่ดีกว่าในทุกตัวชี้วัด ได้แก่ ค่าสัมประสิทธิ์สหสัมพันธ์ (R), ค่าประสิทธิภาพของแบบจำลอง (NSE), ค่าความคลาดเคลื่อนเฉลี่ยรากที่สอง (RMSE) และค่าความคลาดเคลื่อนเฉลี่ยสัมบูรณ์ (MAE) โดยเฉพาะในช่วงพยากรณ์ล่วงหน้า 1 วัน (T+1) กรณีที่ 2 มีค่า R เพิ่มขึ้นจาก 0.258 เป็น 0.928 และค่า NSE เพิ่มขึ้นจาก -1.361 เป็น 0.861 เมื่อเทียบกับกรณีที่ 1 ซึ่งสะท้อนถึงการปรับปรุงด้านความแม่นยำอย่างมีนัยสำคัญ ขณะที่ค่า RMSE ลดลงจาก 5.487 เป็น 2.858 มิลลิเมตร และค่า MAE ลดลงจาก 3.667 เป็น 1.409 มิลลิเมตร ผลการทดลองแสดงให้เห็นว่าการเตรียมข้อมูลด้วยการแยกองค์ประกอบสามารถช่วยเสริมความสามารถในการเรียนรู้ของแบบจำลองหน่วยความจำระยะยาวแบบสั้น โดยเฉพาะในบริบทของข้อมูลปริมาณฝนที่มีความแปรปรวนสูง จึงเป็นแนวทางที่เหมาะสมสำหรับการประยุกต์ใช้ในงานด้านการบริหารจัดการน้ำ เช่น การวางแผนการเก็บกักหรือระบายน้ำจากอ่างเก็บน้ำ เพื่อให้สามารถรับมือกับสถานการณ์น้ำได้อย่างแม่นยำและมีประสิทธิภาพมากยิ่งขึ้น

สำหรับแนวทางการพัฒนาต่อยอดงานวิจัยในอนาคต สามารถประยุกต์ใช้เทคนิคการแยกองค์ประกอบข้อมูล (Data Decomposition) ร่วมกับแบบจำลองโครงข่ายประสาทเทียมประเภทคอนโวลูชัน (Convolutional Neural Network: CNN) เพื่อนำไปใช้กับข้อมูลปริมาณฝนในรูปแบบกริด (Grid-based Rainfall Data) ที่มีความละเอียดเชิงพื้นที่สูงกว่าข้อมูลปริมาณฝนเฉลี่ยของลุ่มน้ำ นอกจากนี้ยังสามารถต่อยอดโดยการประยุกต์ใช้เทคนิคการแยกองค์ประกอบหลายชั้น (Multi-layer VMD) เพื่อเพิ่มระดับความละเอียดในการแยกสัญญาณย่อย อย่างไรก็ตาม การเพิ่มความละเอียดดังกล่าวย่อมแลกมากับความต้องการทรัพยากรในการคำนวณที่สูงขึ้น

กิตติกรรมประกาศ

ขอขอบคุณมหาวิทยาลัยเกษตรศาสตร์ คณะวิศวกรรมศาสตร์ ภาควิชาทรัพยากรน้ำ สำหรับการสนับสนุนทุนการศึกษาและทรัพยากรที่ใช้ในการดำเนินงานวิจัยครั้งนี้ รวมถึงการให้คำแนะนำและสนับสนุนด้านวิชาการอย่างต่อเนื่องจากอาจารย์ที่ปรึกษาและอาจารย์ที่ปรึกษาร่วมตลอดระยะเวลาศึกษา ทั้งนี้ งานวิจัยได้ดำเนินการในพื้นที่เขื่อนลำตะคอง โดยได้รับความร่วมมือด้านข้อมูลจากกรมชลประทานและหน่วยงานที่เกี่ยวข้อง ซึ่งมีส่วนสำคัญอย่างยิ่งต่อความสำเร็จของงานวิจัยนี้

เอกสารอ้างอิง

[1] W. M. Ridwan, M. Sapitang, A. Aziz, K. F. Kushiar, A. N. Ahmed,A. El-Shafie, Rainfall forecasting model using machine learning methods: Case study Terengganu, Malaysia, Ain Shams Engineering Journal 12 (2021) 1651-1663, https://doi.org/10.1016/j.asej.2020.09.011.

[2] Y. Dash, S. K. Mishra,B. K. Panigrahi, Rainfall prediction for the Kerala state of India using artificial intelligence approaches, Computers & Electrical Engineering 70 (2018) 66-73, https://doi.org/10.1016/j.compeleceng.2018.06.004.

[3] C. Z. Basha, N. Bhavana, P. Bhavya,V. S, Rainfall Prediction using Machine Learning & Deep Learning Techniques, in 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), 2-4 July 2020 2020, pp. 92-97, https://doi.org/10.1109/ICESC48915.2020.9155896.

[4] S. Aswin, P. Geetha,R. Vinayakumar, Deep Learning Models for the Prediction of Rainfall, in 2018 International Conference on Communication and Signal Processing (ICCSP), 3-5 April 2018 2018, pp. 0657-0661, https://doi.org/10.1109/ICCSP.2018.8523829.

[5] K. Dragomiretskiy,D. Zosso, Variational Mode Decomposition, IEEE Transactions on Signal Processing 62 (2014) 531-544, https://doi.org/10.1109/TSP.2013.2288675.

[6] H. Zhou, D. Schertzer, and I. Tchiguirinskaia, "Combining Recurrent Neural Networks with Variational Mode Decomposition and Multifractals to Predict Rainfall Time Series," *EGUsphere,* vol. 2024, pp. 1-36, 2024, doi: https://doi.org/10.5194/egusphere-2023-2710.

[7] X. Zhang, Q. Yin, F. Liu, H. Li, and Y. Qi, "Comparative study of rainfall prediction based on different decomposition methods of VMD," *Scientific Reports,* vol. 13, no. 1, p. 20127, 2023/11/17 2023, doi: https://doi.org/10.1038/s41598-023-47416-x.

[8] X. Jiao and Z. He, "A novel coupled rainfall prediction model based on stepwise decomposition technique," *Scientific Reports,* vol. 14, no. 1, p. 10853, 2024/05/13 2024, doi: https://doi.org/10.1038/s41598-024-61855-0.

[9] S. Guo, S. Sun, X. Zhang, H. Chen, and H. Li, "Monthly precipitation prediction based on the EMD–VMD–LSTM coupled model," *Water Supply,* vol. 23, no. 11, pp. 4742-4758, 2023, doi: https://doi.org/10.2166/ws.2023.275.

[10] S. Hochreiter,J. Schmidhuber, Long Short-Term Memory, Neural Computation 9 (1997) 1735-1780, https://doi.org/10.1162/neco.1997.9.8.1735.