

<http://journal.rmutp.ac.th/>

การพัฒนาแบบจำลองสำหรับคาดการณ์ปริมาณน้ำท่าบริเวณลุ่มน้ำมูล

รติพร จันทร์กลั่น^{1*} เกตุกาญจน์ ไชยจันทร์² กิตติศักดิ์ เกิดประสพ¹ และ นิตยา เกิดประสพ¹

¹ สำนักวิศวกรรมศาสตร์ วิศวกรรมคอมพิวเตอร์ มหาวิทยาลัยเทคโนโลยีสุรนารี

² คณะวิศวกรรมศาสตร์และสถาปัตยกรรมศาสตร์ วิศวกรรมคอมพิวเตอร์ มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน

¹ 111 ถนนมหาวิทยาลัย อำเภอเมือง จังหวัดนครราชสีมา 30000

² 744 ถนนสุรนารายณ์ อำเภอเมือง จังหวัดนครราชสีมา 30000

รับบทความ 4 กุมภาพันธ์ 2017; ตอรับบทความ 1 สิงหาคม 2017

บทคัดย่อ

บทความนี้นำเสนอการใช้การสำรวจระยะไกลด้วยดัชนีผลต่างพืชพรรณจากดาวเทียม NOAA ค่าการจัดกลุ่มจาก k-means อุณหภูมิ ปริมาณน้ำฝน จำนวนวันที่ฝนตก และปริมาณน้ำท่า เพื่อสร้างโมเดลคาดการณ์ค่าปริมาณน้ำท่ารายเดือนด้วยโครงข่ายประสาทเทียม ประเมินประสิทธิภาพโดยใช้ค่า R^2 และ RMSE เปรียบเทียบประสิทธิภาพด้วยสมการถดถอย ผลการทดลองสรุปว่าเมื่อเพิ่มค่าการจัดกลุ่มร่วมกับพารามิเตอร์อื่น ๆ เพื่อนำมาสร้างโมเดลสามารถเพิ่มประสิทธิภาพการคาดการณ์ได้ เมื่อใช้ค่าดัชนีผลต่างพืชพรรณร่วมกับอุณหภูมิ ที่เวลาย้อนหลัง 1 และ 2 เดือน และค่าการจัดกลุ่ม ที่เวลาย้อนหลัง 1 เดือน สร้างโมเดลให้มีประสิทธิภาพดีที่สุด โดยให้ผล RMSE=0.09 และ $R^2=0.743$ ผลการทดลองแสดงให้เห็นว่าการใช้ข้อมูลการสำรวจระยะไกล และค่าการจัดกลุ่ม สามารถใช้ในการคาดการณ์ค่าปริมาณน้ำท่าได้อย่างมีประสิทธิภาพ

คำสำคัญ: การสำรวจระยะไกล; น้ำท่า; โครงข่ายประสาทเทียม; สมการถดถอยเชิงเส้น

* ผู้นิพนธ์ประสานงาน โทร: +669 9469 6164, ไปรษณีย์อิเล็กทรอนิกส์: arc_angle@hotmail.com

<http://journal.rmutp.ac.th/>

Development Model for Predict Runoff in Mun Basin

Ratiporn Chanklan^{1*} Kedkarn Chaiyakhan² Kittisak Kerdprasop¹ and Nittaya Kerdprasop¹

¹ Institute of Engineering, Computer Engineering, Suranaree University of Technology

² Faculty of Engineering and Architecture, Computer Engineering, Rajamangala University of Technology Isan

¹ 111 University Avenue, Muang, Nakhon Ratchasima 30000

² 744 Suranarai Road, Muang, Nakhon Ratchasima 30000

Received 4 February 2017; accepted 1 August 2017

Abstract

In this paper proposed remote sensing using Normalized Difference Vegetation Index from NOAA STAR, cluster value from k-means, temperature, rainfall, number of rainy days and runoff to create runoff prediction model using Artificial Neural Network (ANN) and evaluated runoff models with the R^2 and RMSE. The results show that the using of cluster value with other parameters to create predictive models can enhance forecasting results. When using Normalized Difference Vegetation Index with temperature value at lag time 1-2 month and cluster value at lag time 1 month to create model with ANN, we have got the best performance which are RMSE=0.09 and $R^2=0.743$. The experimental results shows that remote sensing data and cluster value from k-means can be used to predictive the runoff effectively.

Keywords: Remote Sensing; Runoff; Artificial Neural Network; Linear Regression

1. บทนำ

การคาดการณ์ปริมาณน้ำท่าหรือปริมาณน้ำในแม่น้ำที่เกิดขึ้นจากฝนเป็นการวิเคราะห์ค่อนข้างยาก เพราะมีกระบวนการเกิดที่ซับซ้อนและมีความสัมพันธ์ที่ซับซ้อนไม่เป็นเชิงเส้น การคาดการณ์ปริมาณน้ำท่าที่ได้จะช่วยในการตัดสินใจการวางแผนและจัดการการใช้ทรัพยากรน้ำ กล่าวคือถ้าเราสามารถรู้ปริมาณน้ำท่าหรือปริมาณน้ำในแม่น้ำได้ล่วงหน้า ทำให้สามารถรู้ว่าจะเกิดปัญหาจากน้ำได้แก่ น้ำท่วม และการขาดแคลนน้ำ ในอนาคตหรือไม่ ดังนั้นการสร้างโมเดลเพื่อคาดการณ์น้ำท่าจะช่วยให้นักสิ่งแวดล้อมที่อาจจะเกิดขึ้นเนื่องจากน้ำท่วมและภัยแล้ง โครงข่ายประสาทเทียม (ANN) เป็นเครื่องมือที่ได้ถูกนำมาใช้ในอุทกวิทยา และสร้างแบบจำลองคาดการณ์ปริมาณน้ำท่า เพราะมีความสามารถในการจำลองทั้งเชิงเส้นและไม่เชิงเส้น โดยไม่จำเป็นต้องตั้งสมมติฐานใด ๆ ตามวิธีการทางสถิติแบบดั้งเดิม มีงานวิจัยหลายชิ้นที่ประสบความสำเร็จในการสร้างแบบจำลองน้ำท่าโดยใช้ ANN

S. Riad et al. [1] ศึกษา ANN เพื่อยืนยันว่าเครือข่ายประสาทเทียมเหมาะที่จะคาดการณ์น้ำท่ามากกว่าวิธีการพื้นฐานด้วยการวิเคราะห์การถดถอยโดยใช้ตัวแปรอินพุต 14 ตัวคือ ปริมาณน้ำฝนและปริมาณน้ำท่าย้อนหลัง 1-7 วัน เพื่อทำนายน้ำท่าที่เวลา t โดยใช้มาตรวัด Squared of Error (ASE), ค่าสัมประสิทธิ์การตัดสินใจ (R^2) ผลการทดลองแสดงให้เห็นว่าวิธี ANN ช่วยให้การคาดการณ์ที่ดีกว่าวิธีการวิเคราะห์การถดถอย

A.R. Ghumman et al. [2] พัฒนา ANN เพื่อให้เหมาะกับชุดข้อมูลที่เก็บรวบรวมที่เป็นระยะสั้น เพื่อทำการคาดการณ์น้ำท่าเดือนถัดไป ค่าพารามิเตอร์ที่ใช้ได้แก่ ค่าน้ำท่าเดือนปัจจุบัน, ค่าน้ำท่าก่อนหน้า, ค่าน้ำฝนเดือนปัจจุบัน, ค่าน้ำฝนก่อนหน้า โดยใช้ค่าจำนวนโหนดในชั้นซ่อนตั้งแต่ 2-15 โดยเลือกค่าจำนวนโหนดในชั้นซ่อนที่ให้ค่าเฉลี่ยความผิดพลาดยกกำลังสอง (MSE) น้อยที่สุด ทดสอบประสิทธิภาพ

การคาดการณ์น้ำท่าโดยใช้ ค่าสัมประสิทธิ์ของความมีประสิทธิภาพ (CE) และค่าเฉลี่ยกำลังสองข้อผิดพลาด (RMSE) ผลการทดลองยืนยันว่า ANN ให้ประสิทธิภาพดี

R. Modarres et al. [3] ใช้ค่าทางสถิติ 17 ค่าในการทดสอบการคาดการณ์ปริมาณน้ำท่า ใช้พารามิเตอร์ 6 ตัวในการสร้างแม่แบบด้วย ANN ได้แก่ ปริมาณน้ำฝนประจำวันของสถานีที่ 1 ก่อนหน้า 1 และ 2 วัน, น้ำฝนประจำวันของสถานีที่ 2 ก่อนหน้า 1 วัน, สถานีที่ 3 ก่อนหน้า 2 วัน, น้ำท่ารายวันในเวลาก่อนหน้า 1 และ 2 วัน เปรียบเทียบกับการวิเคราะห์การถดถอย โดยใช้ข้อมูลน้ำฝนและน้ำท่ารายวันตั้งแต่ ค.ศ. 1978-2000 ผลการทดลองแสดงให้เห็นว่า ANN ให้ผลที่ดีกว่าการวิเคราะห์การถดถอย

Gunwant Sharma et al. [4] ใช้ข้อมูลปริมาณน้ำฝนและน้ำท่ารายปี ค.ศ. 1987-2012 จาก 6 สถานี เพื่อวิเคราะห์น้ำท่าโดยใช้สมการถดถอยแบบ Linear กับ Parabolic โดยใช้ค่าปริมาณน้ำฝนปีก่อนหน้าเป็นอินพุตเพื่อคาดการณ์ปริมาณน้ำท่า ทำการเปรียบเทียบประสิทธิภาพโดยใช้ค่า R^2 จะเห็นว่าเมื่อใช้ Parabolic มีประสิทธิภาพการทำนายน้ำท่าดีกว่า Linear และแสดงให้เห็นว่าความสัมพันธ์ระหว่างปริมาณน้ำฝนปีก่อนหน้าและน้ำท่าในปัจจุบันสอดคล้องกัน

F. Machado et al. [5] สร้างแบบจำลองการทำนายปริมาณน้ำท่าด้วย ANN ด้วยข้อมูลรายเดือนที่มีความแม่นยำ แบ่งข้อมูลออกเป็น 3 ชุด กำหนดค่าโหนดใน Hidden Layer 3, 5, 8, 10 ทดสอบข้อมูลด้วยวิธีการ Nash-Sutcliffe Efficiency Coefficient (NS), R-square Value (R^2) ผลการทดลองที่ข้อมูลชุดที่หนึ่งใช้ อินพุต $P(t)$, $EVT(t)$ และ $Q(t-1)$ ดีที่สุด ชุดที่สองและสามใช้อินพุต $P(t-1)$, $P(t)$, $EVT(t-1)$, $EVT(t)$, $Q(t-1)$ ประสิทธิภาพดีที่สุด

จากที่กล่าวมีข้างต้นจะเห็นว่างานวิจัยส่วนใหญ่ นิยมใช้ค่าปริมาณน้ำฝนและปริมาณน้ำท่าย้อนหลังมาใช้ในการสร้างโมเดลคาดการณ์ปริมาณน้ำท่า และ

พยายามปรับปรุงค่าอินพุตที่ใช้กับ ANN เพื่อเพิ่มประสิทธิภาพการคาดการณ์ปริมาณน้ำท่า

ในงานวิจัยนี้เสนอการใช้ข้อมูลรายเดือนได้แก่ ค่าจากการสำรวจระยะไกลคือดัชนีผลต่างพีชพรรณ (NDVI) อุณหภูมิ ค่าคลัสเตอร์หรือค่าการจัดกลุ่ม น้ำฝน น้ำท่า จำนวนวันที่ฝนตก จาก K-means เพื่อสร้างโมเดลในการคาดการณ์น้ำท่าที่สถานี M.145 ลำพระเพลิง อ.ปากช่อง จ.นครราชสีมา และเปรียบเทียบผลการคาดการณ์น้ำท่าด้วยการใช้การวิเคราะห์การถดถอย โดยทำการประเมินประสิทธิภาพการคาดการณ์โดยใช้ค่า R^2 และ RMSE

2. วิธีการศึกษา

2.1 ทฤษฎีที่เกี่ยวข้อง

2.1.1 การวิเคราะห์การถดถอยเชิงเส้นอย่างง่าย (Simple Linear Regression Analysis)

การวิเคราะห์การถดถอยเชิงเส้นอย่างง่าย เป็นการศึกษาความสัมพันธ์ระหว่างสองตัวแปร กำหนดให้ตัวแปร X เป็นตัวแปรอิสระเป็นตัวแปรที่ทราบค่า ตัวแปร Y เป็นตัวแปรตามเป็นตัวแปรที่ไม่ทราบค่า สองตัวแปรนี้มีความสัมพันธ์กันในลักษณะเส้นตรง (Linear) เมื่อนำค่า X และ Y หลาย ๆ ค่ามาพล็อตบนแกน X และ Y แล้วลากเส้นตรงผ่านจุดที่พล็อตลงในกราฟ เส้นตรงที่ลากนั้นจะแสดงความสัมพันธ์ระหว่างค่าเฉลี่ยของตัวแปร X และตัวแปร Y ซึ่งเส้นตรงนี้เรียกว่า เส้นกราฟถดถอย (Regression Line) ค่าของตัวแปร Y จะเปลี่ยนแปลงไปตามตัวแปร X ค่าของ X หนึ่งค่าจะมีค่า Y ที่เป็นคู่ของค่า X เมื่อนำข้อมูลจากตัวแปรมาวิเคราะห์หาความสัมพันธ์ซึ่งสามารถบอกแนวโน้มของความสัมพันธ์โดยใช้แผนภาพเส้นตรง และจะทำการหาเส้นตรงที่ดีที่สุดเพื่อเป็นตัวแทนของรูปแบบความสัมพันธ์ของตัวแปรที่ศึกษา เส้นตรงที่ดีที่สุดจะมีเพียงเส้นเดียว

2.1.2 การวิเคราะห์การถดถอยเชิงเส้นพหุ

(Multiple Linear Regression Analysis)

ในการหาความสัมพันธ์ของตัวแปรอิสระกับตัวแปรตามบางครั้งที่มีจำนวนตัวแปรอิสระที่สนใจในการศึกษามีมากกว่าหนึ่งตัว ความสัมพันธ์นี้ไม่สามารถใช้การถดถอยเชิงเส้นอย่างง่ายในการวิเคราะห์ได้ สำหรับกรณีที่มีตัวแปรอิสระ 2 ตัว (X_1 และ X_2) ที่มีความสัมพันธ์เชิงเส้นกับตัวแปรตาม สมการถดถอยสามารถเขียนในรูปสมการคือ $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \epsilon_i$ เมื่อ Y_i เป็นตัวแปรตามทีในการเก็บข้อมูลครั้งที่ i ตัวแปรอิสระตัวที่ 1 และ 2 ในการเก็บข้อมูลครั้งที่ i แทนด้วย X_{i1} และ X_{i2} ตามลำดับ β_0 คือค่าเฉลี่ยของตัวแปรตามที่ตัวแปรอิสระทั้งสองมีค่าเป็นศูนย์ β_1 และ β_2 ค่าสัมประสิทธิ์การถดถอยของ X_1 และ X_2 ตามลำดับ และ ϵ_i เป็นค่าความคลาดเคลื่อนในการเก็บข้อมูลครั้งที่ i

การประมาณค่าพารามิเตอร์ใช้วิธีการเช่นเดียวกับสมการถดถอยเชิงเส้นอย่างง่ายคือการใช้วิธีกำลังสองน้อยที่สุด ซึ่งสมการถดถอยเชิงเส้นพหุของกลุ่มตัวอย่างคือ $\hat{Y} = b_0 + b_1 X_1 + b_2 X_2$ ในที่นี้ b_0 คือค่าโดยประมาณ โดยสามารถเขียนในรูปของเมตริกซ์ซึ่งสามารถหาเวกเตอร์ b ได้ดังนี้ $b = (X'X)^{-1} \cdot X'Y$ โดย $X'X$ ต้องสามารถหาเมตริกซ์ผกผันได้และเป็นเมตริกซ์สมมาตรที่มีขนาด $p \times p$ โดย p คือจำนวนตัวแปรอิสระบวกหนึ่งและค่าของสมาชิกในแนวเฉียงเป็นผลรวมกำลังสองของค่าในแต่ละหลัก

2.1.3 ดัชนีผลต่างพีชพรรณ (Normalized

Differential Vegetation Indices: NDVI)

ดัชนีผลต่างพีชพรรณแบบนอร์มัลไลซ์เป็นค่าที่ใช้แสดงการสะท้อนของคลื่นแม่เหล็ก (Electromagnetic Spectrum) ของช่วงคลื่นใกล้อินฟราเรด (Near Infrared Reflectance) กับช่วงคลื่นตามองเห็นสีแดง (Visible Red Reflectance) โดยสามารถนำมาใช้ ในการ

วิเคราะห์การวัดระยะไกล (Remote Sensing Analysis) โดย NDVI เป็นการคำนวณหาสัดส่วนของช่วงคลื่นที่เกี่ยวข้องกับพืชพรรณ โดยนำค่าความแตกต่างของการสะท้อนของพื้นผิว ระหว่างช่วงคลื่นใกล้อินฟราเรด กับช่วงคลื่นตามองเห็นสีแดง มาทำสัดส่วนกับค่าผลรวมของทั้งสองช่วงคลื่น เพื่อปรับให้เป็นลักษณะการกระจายแบบปกติ เพื่อให้การแปรผลทำได้ง่ายขึ้น NDVI นิยมใช้ในการตรวจวัดความสมบูรณ์ของป่าไม้ หรือพืชพรรณ เพื่อประเมินว่าพื้นที่ที่ทำการวิเคราะห์มีพืชพรรณสีเขียวหนาแน่นมากน้อยเพียงใด [6] NDVI มีค่าอยู่ระหว่าง -1.0 ถึง +1.0 บริเวณที่ค่า NDVI อยู่ในช่วงค่าลบพื้นที่จะเป็นผืนน้ำ หรือทะเล ในบริเวณที่มีค่า NDVI เข้าใกล้ค่า 0 แสดงถึงพื้นที่ที่มีพืชพรรณสีเขียวน้อย และหากค่า NDVI มีค่าเข้าใกล้ +1.0 แสดงถึงพื้นที่ที่พืชสีเขียวปกคลุมมาก

2.1.4 โครงข่ายประสาทเทียม

(Artificial Neural Network: ANN)

โครงข่ายประสาทเทียมเป็นแนวความคิดที่ต้องการให้คอมพิวเตอร์มีความสามารถในการเรียนรู้เหมือนมนุษย์ โครงข่ายประสาทเทียมประกอบไปด้วย เซตของโหนดและเส้นเชื่อมระหว่างโหนด สามารถแบ่งโหนดเป็น 3 ระดับได้แก่ โหนดในชั้นอินพุต (Input Layer) มีจำนวนเท่ากับจำนวนคุณสมบัติ (Attribute) ที่ใช้อธิบายข้อมูลแต่ละตัว โหนดในชั้นเอาต์พุต (Output Layer) มีจำนวนเท่ากับจำนวนกลุ่มหรือจำนวนประเภทของข้อมูลที่ต้องการจำแนก โหนดในชั้นซ่อน (Hidden Layer) ในแต่ละชั้นซ่อนอาจจะมีได้มากกว่า 1 ชั้น จำนวนชั้นและจำนวนโหนดจะขึ้นอยู่กับผู้ออกแบบ โดยต้องทดลองหลายๆ แบบแล้วพิจารณาว่าแบบใดให้ประสิทธิภาพที่ดีที่สุด

ในโครงข่ายจะมีเส้นเชื่อมจากทุกโหนดในชั้นอินพุตไปยังทุกโหนดในชั้นซ่อนและมีเส้นเชื่อมจากทุกโหนดในชั้นซ่อนไปยังทุกโหนดในชั้นเอาต์พุต ในแต่ละโหนดจะมีค่า Bias (b) เส้นเชื่อมแต่ละเส้นมีค่าน้ำหนัก (Weight) การทำงานของแต่ละโหนดเทียบได้กับ

เซลล์ประสาทในสมองมนุษย์ 1 เซลล์ อินพุตที่เข้าสู่โหนดจะเป็นเวกเตอร์ของคุณสมบัติของข้อมูลตัวอย่าง มีค่า $p = [p_1, p_2, \dots, p_R]$ มีจำนวน R องค์ประกอบ (คอลัมน์) และเวกเตอร์น้ำหนัก $W = [w_1, w_2, \dots, w_R]$ นำอินพุตมาคูณกับน้ำหนักของแต่ละเส้นเชื่อม ผลที่ได้จากอินพุตทุกๆ เส้นเชื่อมของโหนดจะเอามารวมกันแล้วส่งต่อไปยังฟังก์ชันถ่ายโอน (Transfer Function) ซึ่งเกิดเป็นค่าเอาต์พุต a ในที่นี้ f เป็นฟังก์ชันถ่ายโอนที่รับค่าอินพุต n เพื่อเปลี่ยนเป็นค่าเอาต์พุต a ค่าเอาต์พุต a สามารถคำนวณค่าเอาต์พุต a [7] ได้จาก $a = f(Wp + b)$ เมื่อ $n = w_{1,1}p_1 + w_{1,2}p_2 + \dots + w_{1,R}p_R + b$ ดังนั้น $n = Wp + b$

การค้นหาค่าน้ำหนักของเส้นเชื่อมแต่ละเส้นที่เหมาะสมที่ทำให้สามารถจำแนกประเภทของข้อมูลตัวอย่างที่ใช้สอน (Training Data) ได้ถูกต้องมากที่สุด เป็นการสอนโครงข่ายประสาทเทียมให้เรียนรู้ ค่าน้ำหนักจะทำการปรับจนกว่าค่าความผิดพลาดจะน้อยลงหรืออยู่ในเกณฑ์ที่ยอมรับ

สิ่งสำคัญคือต้องทราบค่า Weight สำหรับสิ่งที่ต้องการให้โครงข่ายเรียนรู้ ซึ่งเป็นค่าที่ไม่แน่นอน แต่สามารถกำหนดให้โครงข่ายปรับค่าเหล่านั้นได้โดยการสอนให้รู้จักรูปแบบ (Pattern) ของสิ่งที่ต้องการให้รู้จัก เรียกว่า “Back Propagation” ถ้าโครงข่ายประสาทเทียมให้ค่าเอาต์พุตผิด

2.1.5 การแบ่งกลุ่มข้อมูลแบบเคมีน

(K-means Clustering)

เป็นการจัดให้ข้อมูลที่มีลักษณะคล้ายกันมาไว้ด้วยกัน ใช้สำหรับการแบ่งข้อมูลจำนวน n เป็น k กลุ่ม [8] ชั้นแรกจะรับค่าพารามิเตอร์ k ซึ่งค่านี้อาจเป็นจำนวน Cluster ที่ต้องการค้นหา จากนั้นกำหนดจุดศูนย์กลาง (Centroid) ชั้นตอนต่อไปคือสร้างกลุ่มข้อมูลและความสัมพันธ์กับจุดศูนย์กลางที่ใกล้กันมากที่สุด ซึ่งจุดแต่ละจุดแทนด้วยข้อมูลหนึ่งตัว โดยแต่ละจุดจะถูกกำหนดกลุ่มไปยังจุดศูนย์กลางที่ใกล้ที่สุดจนครบหมดทุกจุด และคำนวณจุดศูนย์กลางใหม่ โดยการหา

ค่าเฉลี่ยทุกจุดที่อยู่ในกลุ่ม หากจุดศูนย์กลางในแต่ละกลุ่มถูกเปลี่ยนตำแหน่ง จะได้จุดที่มีความสัมพันธ์กับกลุ่มใหม่และใกล้กับจุดศูนย์กลางใหม่ ทำซ้ำไปเรื่อย ๆ จนกระทั่งจุดศูนย์กลางไม่มีการเปลี่ยนแปลง

2.1.6 ค่าสัมประสิทธิ์การตัดสินใจ

(Coefficient of Determination: R^2)

ค่าสัมประสิทธิ์การตัดสินใจ เป็นค่าที่บอกถึงความสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรอิสระว่ามีความสัมพันธ์กันมากน้อยเพียงใด ค่าสัมประสิทธิ์การตัดสินใจมีค่าตั้งแต่ 0-1 ถ้าค่า R^2 มีค่าเข้าใกล้ 1 แสดงว่าตัวแปรอิสระและตัวแปรตามมีความสัมพันธ์กันมากในเชิงเส้นตรง คำนวณได้จากสูตร

$$R^2 = \frac{(N \sum XY - (\sum X)(\sum Y))^2}{(N \sum X^2 - (\sum X)^2)(N \sum Y^2 - (\sum Y)^2)} \quad (1)$$

โดยที่

X, Y = ตัวแปรที่พิจารณา

N = จำนวนข้อมูลทั้งหมด

2.1.7 ค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Squared Error: RMSE)

ค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ยใช้ในการประเมินความผิดพลาดในการคาดการณ์ปริมาณน้ำท่า ถ้าค่า RMSE ยิ่งน้อยหมายถึงการคาดการณ์ยิ่งแม่นยำ สามารถคำนวณได้ดังสมการต่อไปนี้

$$RMSE = \sqrt{\frac{\sum (T_i - O_i)^2}{N}} \quad (2)$$

โดยที่

T_i = ค่าของข้อมูลน้ำท่าจริง

O_i = ค่าของข้อมูลน้ำท่าที่คาดการณ์

N = จำนวนข้อมูลทั้งหมด

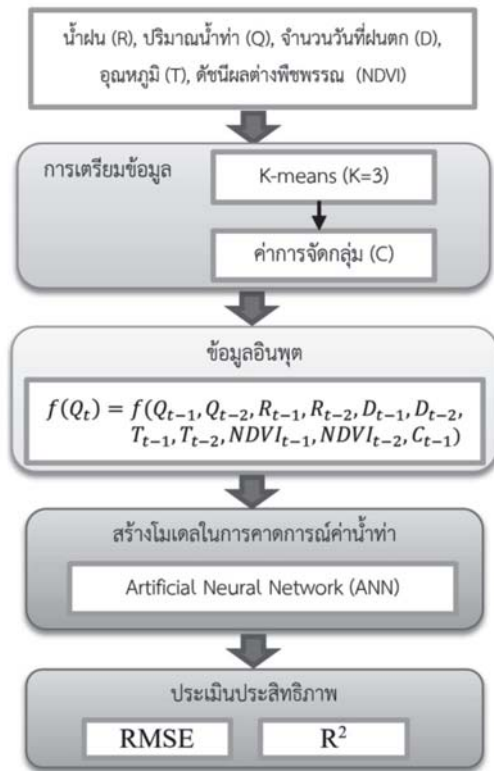
2.2 ขั้นตอนการดำเนินงาน

งานวิจัยนี้ใช้ข้อมูลอุทกภูมิ ปริมาณน้ำฝน และ

ปริมาณน้ำท่าเฉลี่ย ปริมาณน้ำฝน จำนวนวันที่ฝนตก และอุณหภูมิตามเดือน โดยเป็นข้อมูลที่เกี่ยวข้องกับสถานีน้ำท่า M.145 ที่ตั้งอยู่ในจังหวัดนครราชสีมา ซึ่งเป็นบริเวณลุ่มน้ำมูล จากศูนย์อุทกวิทยาชลประทานภาคตะวันออกเฉียงเหนือตอนล่าง (ดาวินโหลดข้อมูลได้จาก www.hydro-4.com) และใช้ข้อมูลดัชนีผลต่างพืชพรรณ จากดาวเทียม NOAA STAR - Global Vegetation Health Products (ดาวินโหลดข้อมูลจาก www.star.nesdis.noaa.gov) โดยจะใช้ข้อมูลที่เก็บรวบรวม 15 ปีตั้งแต่ปี พ.ศ. 2543 ถึง พ.ศ. 2557 โดยในขั้นตอนแรกจะนำข้อมูลทั้งหมดได้แก่ ปริมาณน้ำฝน จำนวนวันที่ฝนตก อุณหภูมิ ปริมาณน้ำท่า ดัชนีผลต่างพืชพรรณ ซึ่งเป็นข้อมูลที่ผ่านการลดความซ้ำซ้อนของข้อมูลด้วยวิธีการนอร์มัลไลเซชัน มาทำการจัดกลุ่มด้วยอัลกอริทึม k-means โดยอยู่ในสมมติฐานที่ว่าข้อมูลส่วนมากจะแบ่งเป็นข้อมูลที่มีค่าต่ำปานกลาง และสูง จึงจัดกลุ่มเป็น $k=3$ หลังจากนั้นจะนำค่าดัชนีผลต่างพืชพรรณ ปริมาณน้ำฝน ปริมาณน้ำท่า จำนวนวันที่ฝนตก และอุณหภูมิต่อมาจัดกลุ่มหนึ่งและสองเดือนร่วมกับค่าการจับกลุ่มที่เวลาย้อนหลังหนึ่งเดือน แล้วนำไปสร้างโมเดลการคาดการณ์ด้วยโครงข่ายประสาทเทียม โดยจะสลับการใช้ข้อมูลในการสร้างโมเดลเพื่อพิจารณาว่าควรใช้ข้อมูลใดในการสร้างโมเดลในการทำนายปริมาณน้ำท่าให้ประสิทธิภาพดีที่สุด กล่าวคือ ถ้าต้องการคาดการณ์ค่าน้ำท่าเดือนมีนาคม จะทำการสร้างโมเดลการคาดการณ์โดยใช้ค่าดัชนีผลต่างพืชพรรณ ปริมาณน้ำฝน ปริมาณน้ำท่า จำนวนวันที่ฝนตก และอุณหภูมิต่อมากรกฎาคมและเดือนกุมภาพันธ์ ส่วนค่าการจับกลุ่มจะใช้ที่เดือนกุมภาพันธ์เพื่อนำไปใช้ในการสร้างโมเดลการคาดการณ์ด้วยโครงข่ายประสาทเทียม

ในการวิจัยนี้จะใช้โครงข่ายประสาทเทียมแบบการแพร่ย้อนกลับ (Back-propagation Algorithm) ใช้จำนวนชั้นในชั้นซ่อน 1 ชั้น ทำการเลือกจำนวนโหนดภายในชั้นซ่อนตั้งแต่ 1-10 แล้วเลือกใช้จำนวนโหนด

ที่ให้ค่าเฉลี่ยความผิดพลาดยกกำลังสอง (RMSE) ที่น้อยที่สุดกับการคาดการณ์ค่าน้ำท่าในชุดข้อมูลฝึกสอน



รูปที่ 1 กรอบแนวคิดของงานวิจัย

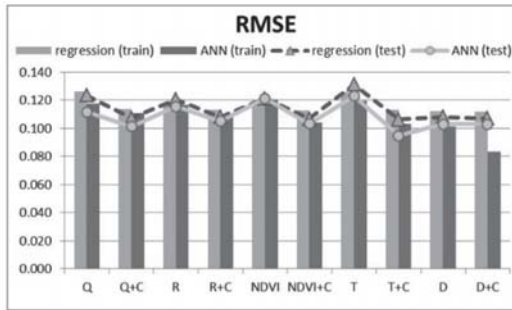
นอกจากนี้เปรียบเทียบประสิทธิภาพด้วยการวิเคราะห์การถดถอยเชิงเส้น และใช้พารามิเตอร์ตัวอื่น ๆ ในการสร้างโมเดลการคาดการณ์ปริมาณน้ำท่า ได้แก่ ปริมาณน้ำฝนรายเดือน (R) จำนวนวันที่ฝนตกในแต่ละเดือน (D) และปริมาณน้ำท่ารายเดือน (Q) ที่เวลาย้อนหลัง 1 เดือนและ 2 เดือน (t-1, t-2) การทดลองนี้ใช้มาตรวัดทางสถิติคือ ค่าสัมประสิทธิ์การตัดสินใจ (R^2) ซึ่งใช้แสดงความสัมพันธ์ระหว่างค่าน้ำท่าจริงกับค่าที่คาดการณ์ และใช้ค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) ซึ่งเป็นค่าที่แสดงความผิดพลาดของการคาดการณ์ โดยกรอบแนวคิดของงานวิจัยนี้สามารถแสดงดังรูปที่ 1

3. ผลการศึกษาและอภิปรายผล

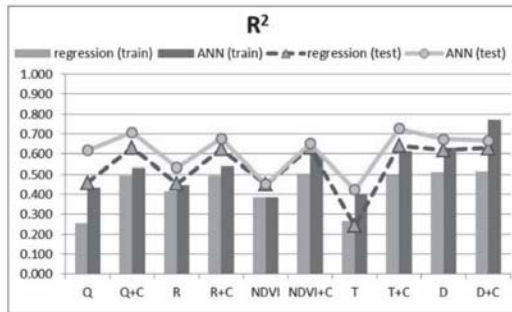
ในขั้นตอนแรกจะใช้อินพุตเป็นพารามิเตอร์ 1 ชนิด เปรียบเทียบกับการใช้ค่าการจับกลุ่มร่วมด้วย เพื่อจะแสดงการทดสอบว่าค่าการจับกลุ่มมีผลต่อประสิทธิภาพการคาดการณ์น้ำท่า จากนั้นสร้างแบบจำลองการคาดการณ์น้ำท่าด้วยโครงข่ายประสาทเทียม (ANN) และ การวิเคราะห์การถดถอยเชิงเส้น (Linear Regression: LR) เปรียบเทียบประสิทธิภาพการคาดการณ์ด้วยมาตรวัดทางสถิติด้วยการหาค่า R^2 และ RMSE ผลการทดลองแสดงในตารางที่ 1 และกราฟแสดงในรูปที่ 2

ตารางที่ 1 ใช้พารามิเตอร์ 1 ชนิดและผลการเพิ่มค่าการจับกลุ่มเพื่อใช้คาดการณ์น้ำท่า (ข้อมูลฝึกสอน)

Parameters	Hidden Nodes	Test Data			
		RMSE		R^2	
		LR	ANN	LR	ANN
Q	9	0.123	0.111	0.459	0.618
Q+C	9	0.107	0.101	0.631	0.707
R	6	0.12	0.115	0.453	0.533
R+C	6	0.108	0.105	0.623	0.677
NDVI	7	0.121	0.121	0.451	0.451
NDVI+C	7	0.107	0.103	0.631	0.651
T	7	0.131	0.123	0.242	0.423
T+C	7	0.106	<u>0.095</u>	0.642	<u>0.727</u>
D	6	0.108	0.103	0.620	0.673
D+C	6	0.107	0.103	0.630	0.667



(ก) ค่า RMSE



(ข) ค่า R²

รูปที่ 2 กราฟการทดลองใช้พารามิเตอร์ 1 ชนิด และการเพิ่มค่าการจัดกลุ่มของข้อมูลฝึกสอน (กราฟแท่ง) และข้อมูลทดสอบ (กราฟเส้น) ในรูป (ก) แสดงค่า RMSE (ข) แสดงค่า R²

ตารางที่ 2 ผลการใช้พารามิเตอร์ 2 ชนิดร่วมกับค่าการจัดกลุ่มเพื่อใช้พยากรณ์ (ข้อมูลฝึกสอน)

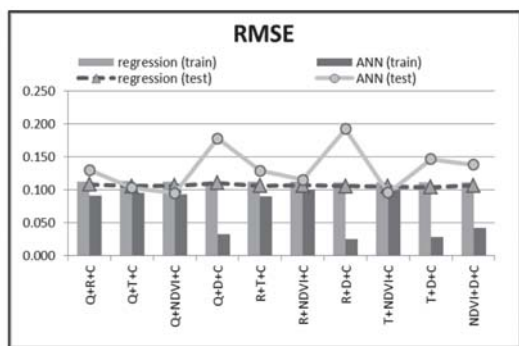
Parameters	Hidden Nodes	Test Data			
		RMSE		R ²	
		LR	ANN	LR	ANN
Q+R+C	8	0.109	0.13	0.615	0.409
Q+T+C	9	0.106	0.103	0.644	0.653
Q+NDVI+C	2	0.107	0.095	0.632	0.711
Q+D+C	10	0.111	0.178	0.595	0.498
R+T+C	9	0.107	0.129	0.638	0.519
R+NDVI+C	6	0.107	0.116	0.629	0.545
R+D+C	10	0.106	0.193	0.642	0.419
T+NDVI+C	3	0.105	<u>0.096</u>	0.66	<u>0.743</u>
T+D+C	10	0.104	0.147	0.655	0.532
NDVI+D+C	9	0.107	0.138	0.629	0.458

จากตารางที่ 1 เมื่อใช้พารามิเตอร์ 1 ชนิดที่เวลาย้อนหลัง 1 เดือนและ 2 เดือน ร่วมกับค่าการจัดกลุ่มที่เวลาย้อนหลัง 1 เดือน ให้ค่า RMSE ลดลง และค่า R² เพิ่มขึ้นในทุกพารามิเตอร์ ซึ่งแสดงให้เห็นว่าหากใช้ค่าการจัดกลุ่มร่วมกับพารามิเตอร์อื่น สามารถเพิ่มประสิทธิภาพการพยากรณ์น้ำทำได้ใน ANN และ Regression จากผลการทดลองให้ประสิทธิภาพดีที่สุดเมื่อใช้อุณหภูมิร่วมกับการจัดกลุ่ม (T+C) โดยที่ ANN ให้ค่า RMSE=0.095 และค่า R²= 0.727

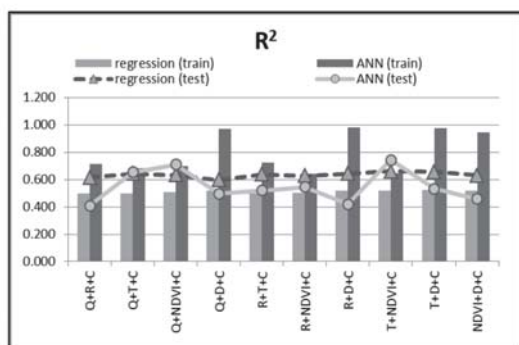
รูปที่ 2 แสดงให้เห็นว่าในข้อมูลฝึกสอนและข้อมูลทดสอบมีค่า RMSE และค่า R² ไปในทิศทางเดียวกัน เนื่องจากการทดลองตารางที่ 1 แสดงให้เห็นว่าเมื่อใช้ค่าการจัดกลุ่มสามารถเพิ่มประสิทธิภาพการพยากรณ์ได้ ดังนั้นจึงทดลองการใช้พารามิเตอร์ 2 ชนิดร่วมกับค่าการจัดกลุ่ม ซึ่งแสดงในตารางที่ 2

จากตารางที่ 2 เมื่อใช้ค่าอุณหภูมิ และดัชนีผลต่างพีชพรรณ ร่วมกับค่าการจัดกลุ่ม (T+NDVI+C) ที่ ANN มีประสิทธิภาพดีที่สุดในข้อมูลทดสอบโดยให้ค่า RMSE = 0.096 และ R² =0.743 Regression ให้ประสิทธิภาพการพยากรณ์โดยพิจารณาจาก RMSE ไม่แตกต่างจากการใช้พารามิเตอร์ 1 ชนิดแต่ใน R² มีค่าสูงตั้งแต่ 0.6 ขึ้นไปในทุกกรณี

เมื่อนำผลการทดลองมาแสดงเป็นกราฟในรูปที่ 3 จะเห็นว่ารูปที่ 3 (ก) ค่า RMSE และรูปที่ 4 (ข) ค่า R² ที่ ANN เกิดปัญหาการ Overfitting ชัดเจนในกรณี Q+D+C, R+D+C และ NDVI+D+C กล่าวคือการที่โมเดลที่ได้จากการใช้ชุดข้อมูลฝึกสอนให้ประสิทธิภาพการพยากรณ์สูง แต่เมื่อนำโมเดลไปใช้กับชุดข้อมูลทดสอบให้ประสิทธิภาพการพยากรณ์ต่ำ จะสังเกตเห็นว่าในทุกกรณีเมื่อใช้ D+C จะทำให้เกิดปัญหา Overfitting ดังนั้นผลการทดลองต่อไปจะใช้พารามิเตอร์ Q, T, NDVI, R และ C เป็นพารามิเตอร์ในการทดลองต่อไป เพราะไม่มีแนวโน้มทำให้เกิดปัญหา Overfitting และจะไม่ใช่ D+C ผลการทดลองใช้พารามิเตอร์ 3 ชนิดและ 4 ชนิดร่วมกับค่าการจัดกลุ่มแสดงในตารางที่ 3



(ก) ค่า RMSE



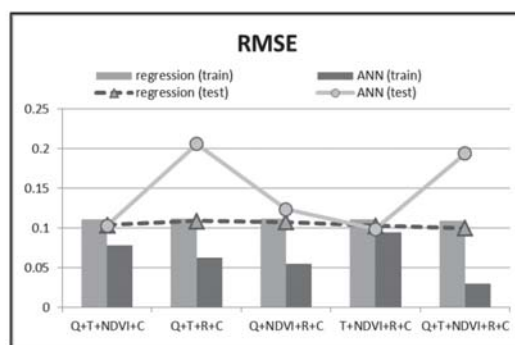
(ข) ค่า R²

รูปที่ 3 กราฟการใช้พารามิเตอร์ 2 ชนิด ร่วมกับค่าการจับกลุ่มของข้อมูลฝึกสอน (กราฟแท่ง) และข้อมูลทดสอบ (กราฟเส้น) ในรูป (ก) แสดงค่า RMSE (ข) แสดงค่า R²

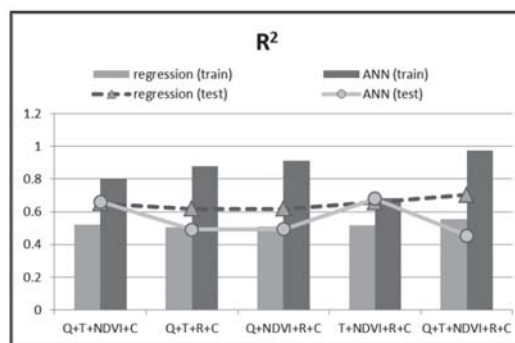
ตารางที่ 3 การใช้พารามิเตอร์ 3 และ 4 ชนิดร่วมกับค่าการจับกลุ่มเพื่อใช้พยากรณ์ (ข้อมูลฝึกสอน)

Parameters	Hidden Nodes	Test Data			
		RMSE		R ²	
		LR	ANN	LR	ANN
Q+T+NDVI+C	7	0.104	0.103	0.652	0.661
Q+T+R+C	10	0.109	0.206	0.617	0.491
Q+NDVI+R+C	10	0.108	0.124	0.617	0.494
T+NDVI+R+C	8	0.103	0.099	0.658	0.681
Q+T+NDVI+R+C	10	0.100	0.194	0.702	0.453

จากตารางที่ 3 เมื่อใช้ T+NDVI+R+C มีประสิทธิภาพดีที่สุดที่ ANN ในข้อมูลทดสอบให้ค่า RMSE=0.099 และ Q+T+NDVI+R+C ให้ค่า R²=0.702 มีประสิทธิภาพดีที่สุดที่ ANN และจากรูปที่ 4 จะเห็นว่าเกิดปัญหาการ Overfitting ที่ ANN ยกเว้นกรณี T+NDVI+R+C และประสิทธิภาพการพยากรณ์น้ำท่าไม่ดีกว่าการใช้พารามิเตอร์ 2 ชนิดร่วมกับค่าการจับกลุ่ม (T+NDVI+C) ที่แสดงในตารางที่ 2



(ก) ค่า RMSE



(ข) ค่า R²

รูปที่ 4 กราฟการใช้พารามิเตอร์ 3 และ 4 ชนิด ร่วมกับค่าการจับกลุ่มของข้อมูลฝึกสอน (กราฟแท่ง) และข้อมูลทดสอบ (กราฟเส้น) ในรูป (ก) แสดงค่า RMSE (ข) แสดงค่า R²

จากการทดลองทั้งหมดแสดงให้เห็นว่าเมื่อใช้ ANN ในการสร้างแบบจำลองมีประสิทธิภาพดีที่สุด โดยใช้อินพุตเป็น $T+NDVI+C$ ที่ชุดข้อมูลทดสอบให้ค่า $RMSE = 0.096$ และ $R^2 = 0.743$ (แสดงในตารางที่ 2) และไม่เกิดปัญหาการ Overfitting ส่วนในการสร้างแบบจำลองด้วย Regression การคาดการณ์น้ำทำให้ประสิทธิภาพดีที่สุดเมื่อใช้ $Q+T+NDVI+R+C$ ในชุดข้อมูลทดสอบโดยให้ค่า $RMSE = 0.100$ และ $R^2 = 0.702$ (แสดงในตารางที่ 3) และไม่เกิดการ Overfitting จะเห็นว่าเมื่อใช้ ANN ให้ประสิทธิภาพดีที่สุดในการทดลองนี้และมีประสิทธิภาพการคาดการณ์น้ำทำดีกว่า Regression นอกจากนี้ยังแสดงให้เห็นว่าการเลือกใช้พารามิเตอร์ในการสร้างโมเดลในการทำนายมีผลต่อความถูกต้องของการทำนาย

4. สรุป

การคาดการณ์น้ำทำถือเป็นข้อมูลสำคัญที่สามารถช่วยให้สามารถป้องกันปัญหาที่เกิดจากน้ำได้แก่ น้ำท่วม และการขาดแคลนน้ำ กล่าวคือถ้าเราสามารถรู้ปริมาณน้ำทำหรือปริมาณน้ำในแม่น้ำได้ล่วงหน้า ทำให้สามารถรู้ว่าจะเกิดปัญหาเกี่ยวกับน้ำขึ้นในอนาคตหรือไม่ ทำให้มีการวางแผนรับมือกับปัญหาที่จะเกิดขึ้นในอนาคตได้ งานวิจัยนี้เสนอการสร้างโมเดลคาดการณ์ค่าปริมาณน้ำทำด้วยโครงข่ายประสาทเทียมโดยการทดลองใช้ค่าอุณหภูมิ (T) และดัชนีผลต่างพืชพรรณ (NDVI) ปริมาณน้ำฝน (R) จำนวนวันที่ฝนตก (D) และปริมาณน้ำทำ (Q) ที่เวลาย้อนหลัง 1 และ 2 เดือน (t-1 และ t-2) ร่วมกับ ค่าการจัดกลุ่ม (C) ที่เวลาย้อนหลัง 1 เดือน (t-1) เพื่อแสดงให้เห็นว่าการเลือกใช้พารามิเตอร์ในการสร้างโมเดลมีผลต่อการทำนาย การใช้ค่าการจัดกลุ่มเพื่อเพิ่มประสิทธิภาพของการทำนายค่า น้ำทำ และค่าจากการสำรวจระยะไกลสามารถใช้ในการทำนายน้ำทำได้ การประเมินประสิทธิภาพโดยใช้วิธีการทางสถิติ 2 ชนิดคือค่า $RMSE$ ที่แสดงความผิดพลาดของการคาดการณ์ปริมาณน้ำทำและ R^2 เพื่อ

ดูความสัมพันธ์ระหว่างค่า น้ำทำจริงกับค่า น้ำทำที่ได้จากการคาดการณ์ นอกจากนี้ยังเปรียบเทียบประสิทธิภาพกับการสร้างแบบจำลองด้วยสมการถดถอย

ผลการทดลองสรุปได้ว่าเมื่อใช้การจัดกลุ่มร่วมกับพารามิเตอร์อื่น ๆ เพื่อสร้างโมเดลคาดการณ์ปริมาณน้ำทำสามารถเพิ่มประสิทธิภาพการคาดการณ์ได้ และเมื่อใช้ค่าดัชนีผลต่างพืชพรรณ ร่วมกับอุณหภูมิ และค่าการจัดกลุ่ม ($NDVI+T+C$) สร้างโมเดลด้วยโครงข่ายประสาทเทียม ให้ประสิทธิภาพดีที่สุดโดยให้ค่า $RMSE=0.09$ และ $R^2=0.743$ (แสดงในตารางที่ 2) จะเห็นว่า การใช้ข้อมูลการสำรวจระยะไกลด้วยดัชนีผลต่างพืชพรรณ และค่าการจัดกลุ่มจาก k-means สามารถนำมาใช้ในการสร้างโมเดลคาดการณ์ค่าปริมาณน้ำทำได้ และการสร้างโมเดลด้วยโครงข่ายประสาทเทียมเป็นวิธีการที่มีการคำนวณที่ซับซ้อน การเลือกใช้พารามิเตอร์ในการสร้างโมเดลที่ใช้ในการคาดการณ์ค่าปริมาณน้ำทำมีประสิทธิภาพ และสามารถป้องกันเกิดปัญหาการ Overfitting จำนวนพารามิเตอร์และการเลือกใช้พารามิเตอร์ไม่เหมาะสม และพารามิเตอร์ที่เลือกใช้จะต้องมีความสัมพันธ์กับค่าที่คาดการณ์ ถ้าเราเลือกใช้พารามิเตอร์ไม่เหมาะสมจะทำให้โมเดลการคาดการณ์น้ำทำจากโครงข่ายประสาทเทียมคาดการณ์ปริมาณน้ำทำผิดพลาด

นอกจากนี้ข้อมูลปริมาณน้ำฝนและน้ำทำที่ใช้ทำการทดลองในงานวิจัยนี้ใช้แค่หนึ่งสถานี หากทำการทดลองเพิ่มโดยนำข้อมูลปริมาณน้ำฝนและน้ำทำจากสถานีข้างเคียงมาพิจารณา หรือต้องการวิธีการเลือกพารามิเตอร์ที่เหมาะสมในการสร้างโมเดลคาดการณ์ปริมาณน้ำทำอาจจะช่วยให้เพิ่มประสิทธิภาพการคาดการณ์ปริมาณน้ำทำได้ การทดลองต่อไปในอนาคตจะต้องทดลองกับข้อมูลน้ำทำสถานีอื่นและลุ่มน้ำบริเวณอื่น เพื่อแสดงให้เห็นประสิทธิภาพของวิธีการที่นำเสนอ

5. เอกสารอ้างอิง

- [1] S. Riad, J. Mania, L. Bouchaou and Y. Najjar, "Rainfall-runoff model using an artificial neural network approach," *Mathematical and Computer Modelling*, vol. 40, no. 7, pp. 839-846, 2004.
- [2] A. R. Ghumman, Y. M. Ghazaw, A. R. Sohail and K. Watanabe, "Runoff forecasting by artificial neural network and conventional mode," *Alexandria Engineering Journal*, vol. 50, no. 4, pp. 345-350, 2011.
- [3] R. Modarres, "Multi-criteria validation of artificial neural network rainfall-runoff modeling," *Hydrology and Earth System Sciences*, vol. 13, no.3, pp. 411-421, 2009.
- [4] G. Sharm, Y.P. Mathur, S. K. Vyas, & P.K. Navin, "Rainfall-Runoff Regression Model for Meja Catchment," *International Journal of Inventions in Reasearch, Engineering Science and Technology (IJIREST)*, vol. 1, no. 1, pp. 58-62, 2014.
- [5] F. Machado, M. Mine, E. Kaviski and H. Fill, "Monthly rainfall-runoff modelling using artificial neural networks," *Hydrological Sciences Journal-Journal des Sciences Hydrologiques*, vol. 56, no.3, pp. 349-361, 2011.
- [6] F.N. Kogan, "Vegetation index for a real analysis of crop conditions," in *Proceedings of the 18th Conference on Agricultural and Forest Meteorology*, AMS, W. Lafayette, Indiana, 15-18 September 1987, Indiana, USA, pp. 103-106.
- [7] Mark Hudson Beale, Martin T. Hagan and Howard B. Demuth, *Neural Network Toolbox™ User's Guide*: MathWorks Inc, 2015.
- [8] Alsabti Khaled, Sanjay Ranka and Vineet Singh, "An efficient k-means clustering algorithm," *Electrical Engineering and Computer Science*, 1997.