

แบบจำลองโครงข่ายประสาทเทียมสำหรับคลอรีนอิสระคงเหลือ โดยใช้ข้อมูลของการประปาบางเขนเป็นกรณีศึกษา

Free Residual Chlorination Models Using Artificial Neural Network: Bangkhen Water Treatment Plant Case Study

กิงกาญจน์ พิทยกิจติวงศ์¹ วรณศักดิ์ เลี่ยมแหลม¹ เสฎฐา ศาสนนันท์^{2*}

¹ภาควิชาวิศวกรรมโยธา, คณะวิศวกรรมศาสตร์ มหาวิทยาลัยธรรมศาสตร์ ตำบลคลองหนึ่ง
อำเภอคลองหลวง จังหวัดปทุมธานี 12120

²ภาควิชาวิศวกรรมโยธา, คณะวิศวกรรมศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ ตำบลองครักษ์
อำเภอองครักษ์ จังหวัดนครนายก 12110

*Corresponding author: E-mail: e41stomos@gmail.com

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์ในการประยุกต์ใช้โครงข่ายประสาทเทียม (Artificial Neural Networks) ในการสร้างแบบจำลองทางคณิตศาสตร์ทำนายค่าคลอรีนอิสระคงเหลือ (Free Residual Chlorine) โดยจุดประสงค์หลักของแบบจำลองจะนำไปใช้ในการฝึกผู้ปฏิบัติงานที่ไม่มีประสบการณ์ในการควบคุมระบบจ่ายคลอรีน ในการสร้างแบบจำลองทางคณิตศาสตร์นี้จะใช้ข้อมูลอนุกรมเวลาในการเดินระบบจ่ายคลอรีนของโรงผลิตน้ำประปาบางเขน การประปานครหลวง ข้อมูลในการเดินระบบนี้เป็นข้อมูลรายวันโดยเก็บเป็นระยะเวลา 1 ปี ตั้งแต่วันที่ 1 มกราคม พ.ศ. 2553 ถึง วันที่ 31 ธันวาคม พ.ศ. 2553 ในการสร้างแบบจำลองโครงข่ายประสาทเทียมได้ทำการแบ่งข้อมูลออกเป็น 3 ส่วน คือ Training Set, Test Set และ Validation Set เพื่อเป็นข้อมูลในการสอน การทดสอบแบบจำลองเพื่อเปรียบเทียบประสิทธิภาพและลดการเกิด Over Fitting ของแบบจำลอง โดยสอนแบบจำลองให้เกิดการเรียนรู้ ด้วยเทคนิคการแพร่กระจายย้อนหลัง (Back Propagation Learning) การวัดประสิทธิภาพของแบบจำลองทำได้โดยการเปรียบเทียบค่า Coefficient of Determination และค่า Mean Absolute Error ของชุดข้อมูล Test Set

จากผลการวิจัยพบว่าแบบจำลองโครงข่ายประสาทเทียมที่มีประสิทธิภาพการทำนายที่ดีที่สุด เป็นแบบจำลองแบบสองชั้นซ่อน โดยมีจำนวนโหนดเท่ากับ 50 และ 90 โหนด ในชั้นซ่อนที่หนึ่งและสองตามลำดับแบบจำลองนี้ให้ผลการทำนายที่ดีมากโดยมีค่า Coefficient of Determination เท่ากับ 0.99 และค่าความผิดพลาด Mean Absolute Error เท่ากับ 0.02 มิลลิกรัมต่อลิตร

คำสำคัญ: โครงข่ายประสาทเทียม คลอรีน อนุกรมเวลา

ABSTRACT

The objective of this research is to apply Artificial Neural Network Constructing for mathematical model to predict Free Residual Chlorine. The main purpose of the model is to train unexperienced operators in the control of chlorination system. The mathematical model was constructed by using Time Series operation data collected from the Metropolitan Waterworks Authority. The data was collected daily from 1st January 2010 to 31th December 2010. In constructing the Artificial Neural Network Model, the input data was divided into 3 sets: Training Set, Test Set, and Validation Set, in order to be used for teaching, testing the model efficiency, and reducing over fitting. Back Propagation learning was applied in training the model while model performance was evaluated by using the Coefficient of Determination and the Mean Absolute Error of the data in the Test Set.

The results reveal that the Artificial Neural Network Model with two hidden layers obviously shows the best performance, with 50 nodes in the first hidden layer and 90 nodes in the second hidden layer. The model gives excellent prediction as reflected by a Coefficient of Determination of 0.99 and a Mean Absolute Error of 0.02 milligrams per liter.

Keyword: Artificial neural networks, Chlorine, Time series

1. บทนำ

กระบวนการที่สำคัญที่สุดกระบวนการหนึ่งในการผลิตน้ำประปาคือกระบวนการกำจัดเชื้อโรค (Disinfection) ในประเทศไทยนิยมใช้การเติมคลอรีนเพื่อให้เชื้อโรคและจุลินทรีย์ที่อาจเล็ดลอดจากระบบกรองถูกทำลายจนอยู่ในระดับที่ยอมรับได้ ในปัจจุบันการประมาณปริมาณคลอรีน ผู้ปฏิบัติงาน (Operator) จะใช้ความชำนาญในการกำหนดผู้ปฏิบัติงานเป็นผลให้ไม่สามารถเดินระบบได้อย่างเที่ยงตรงและแม่นยำส่งผลให้ไม่สามารถนำค่ามาปรับใช้ได้อย่างทันทีในระบบผลิต ซึ่งการนำเทคโนโลยีการใช้โครงข่ายประสาทเทียม มาช่วยในการทำนายค่าปริมาณคลอรีนอิสระคงเหลือ จะสามารถลดระยะเวลา และเพิ่มความมั่นใจให้กับผู้ปฏิบัติงานได้

แบบจำลองโครงข่ายประสาทเทียมเป็นแบบจำลองแบบ Black Box [11] แบบจำลองนี้เป็นแบบจำลองที่อาศัยข้อมูลจำนวนมากในการสร้างความสัมพันธ์ระหว่างชั้นขาเข้าและชั้นขาออกของแบบจำลอง โดยไม่ต้องคำนึงถึงความหมายทางกายภาพ (Physical Meaning) ของกระบวนการสร้างแบบจำลอง โดยการสร้างแบบจำลองโครงข่ายประสาท

เทียม นี้จะมีการกำหนดตัวแปรชั้นขาเข้าของแบบจำลองโดยการพิจารณาค่าสัมประสิทธิ์ความถดถอย (Regression Coefficient) [15] ซึ่งแสดงความสัมพันธ์ระหว่างตัวแปรชั้นขาเข้ากับตัวแปรชั้นขาออก

2. วิธีดำเนินการ

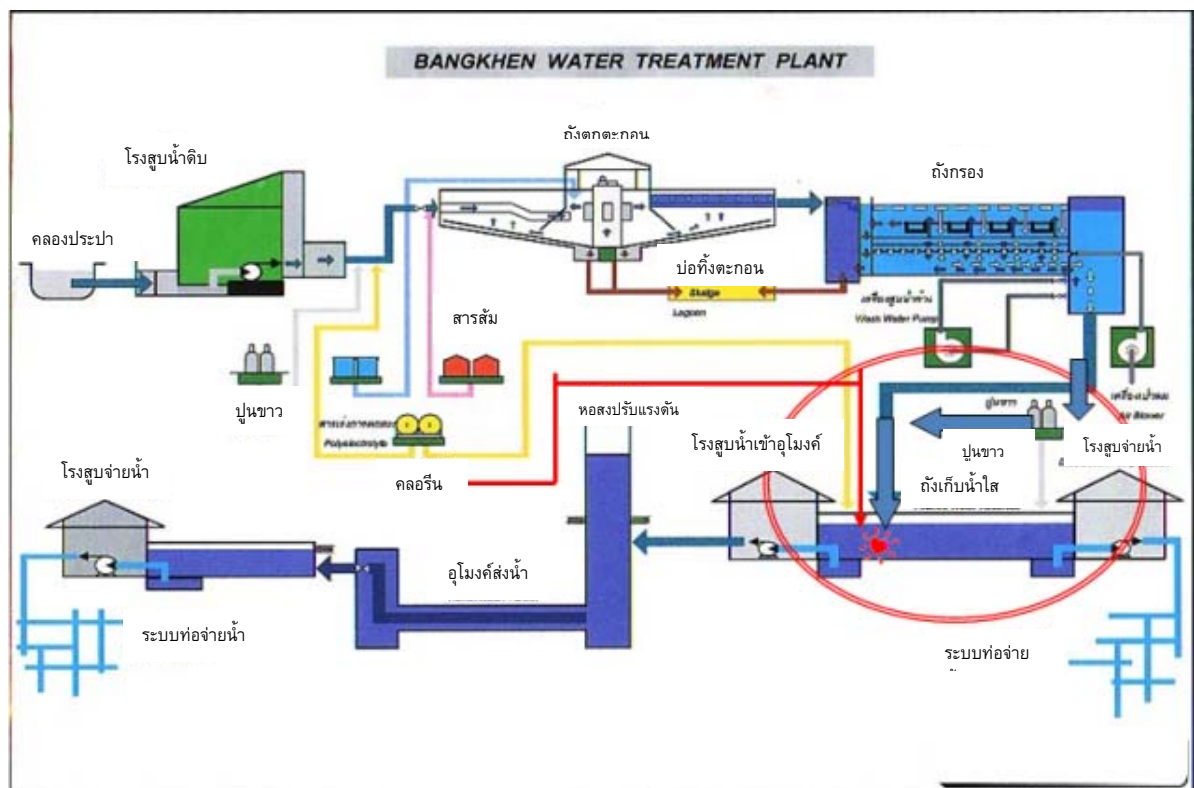
การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อประยุกต์ใช้แบบจำลองโครงข่ายประสาทเทียม (Artificial Neural Network Model, ANN) [6] แบบ Back Propagation Learning [16] ในการทำนายค่าปริมาณคลอรีนอิสระคงเหลือ (Free Residual Chlorine) ในการสร้างแบบจำลองโครงข่ายประสาทเทียม ประกอบด้วย 3 ขั้นตอน คือ การเก็บข้อมูล (Data Collection) การเตรียมข้อมูล (Pre Process) และกระบวนการของแบบจำลอง (Modeling Process) โดยใช้ข้อมูลโรงผลิตน้ำประปาบางเขน การประปานครหลวง

กระบวนการผลิตน้ำประปาของโรงผลิตน้ำประปาบางเขน [13] มีการเติมปูนขาว (Lime) ในท่อลำเลียงน้ำดิบ เพื่อปรับสภาพให้มีความเป็นด่างในน้ำดิบ ช่วยให้สารส้ม (Alum) ทำปฏิกิริยาได้ดี ในกรณีที่

น้ำดิบมีสีและตะไคร่น้ำสูงเกินค่ามาตรฐาน จะมีการเติมคลอรีนเพื่อกำจัดสิ่งปนเปื้อนดังกล่าวในน้ำดิบ การเติมคลอรีนดังกล่าวนี้เรียกว่า Pre - Chlorination เมื่อน้ำดิบไหลเข้าสู่ถังตกตะกอน สารเคมีจะถูกกวนให้สัมผัสและทำปฏิกิริยากับตะกอนที่อยู่ในน้ำ จับกันเป็นก้อนเล็กๆแล้วค่อยๆมีขนาดใหญ่ขึ้นและตกลงสู่ก้นถังเหลือเพียงน้ำใสไหลไปยังบ่อกรอง (Filter) ในฤดูกาลที่น้ำดิบมีความขุ่นสูงหรือต่ำผิดปกติ อาจมีการเติมสารช่วยในการตกตะกอน (Poly-electrolyte) ในถังตกตะกอน เพื่อช่วยให้เกิดการตกตะกอนได้ดีขึ้น น้ำที่ผ่านการกรองแล้วจะมีความขุ่นต่ำกว่า 1-2 NTU [13] ซึ่งเป็นสภาพที่เหมาะสมสำหรับการฆ่าเชื้อโรคโดยใช้คลอรีน กระบวนการนี้ว่า Post Chlorination การเติมคลอรีนลงในน้ำจะต้องมีปริมาณที่เหมาะสมและเวลาที่เพียงพอ นอกจากนี้แล้วจะต้องเติมคลอรีนมากพอจนมีปริมาณคลอรีนอิสระคงเหลือ อยู่อย่างน้อย 0.7 ppm [13]

2.1 แบบจำลองโครงข่ายประสาทเทียม (Artificial neural network model, ANN)

เป็นแบบจำลองทางคณิตศาสตร์ [6] สำหรับประมวลผลสารสนเทศด้วยการคำนวณแบบคอนเนกชันนิสต์ (Connectionist) เทคนิคนี้ได้มาจากการศึกษาข่ายงานไฟฟ้าชีวภาพในสมองซึ่ง ประกอบด้วย เซลล์ประสาท หรือ โหนด และ ข่ายงานประสาทเกิดจากการเชื่อมต่อระหว่างเซลล์ประสาท จนเป็นเครือข่ายที่ทำงานร่วมกัน ซึ่งรูปแบบการคำนวณนั้นค่อนข้างซับซ้อน โดยใช้ข้อมูลต่างๆ ที่มีอยู่ให้ออกมาในรูปแบบของโครงสร้างวิธีการคำนวณ เพื่อที่จะนำรูปแบบที่ได้จากชุดข้อมูลที่เลือกนี้ไปใช้ในการวิเคราะห์จดจำความหมายของข้อมูล (Pattern) ที่อยู่ในลักษณะเดียวกันหรือในบางโอกาสนักวิจัยบางกลุ่มเรียกกระบวนการนี้ว่ากระบวนการจัดรูปแบบ



รูปที่ 1 กระบวนการผลิตน้ำประปาของโรงงานผลิตน้ำประปาบางเขน ที่มา : โรงผลิตน้ำบางเขน

และนำไปใช้ (Pattern Recognition) ซึ่งวิธีการดังกล่าวจะเป็นการเลียนแบบวิธีการทำงานของสมองมนุษย์หรืออาจจะกล่าวได้ว่าวิธีการดังกล่าวเป็นการจำลองการทำงานของสมองและความรู้ที่ได้เกิดขึ้นได้จากกระบวนการเรียนรู้

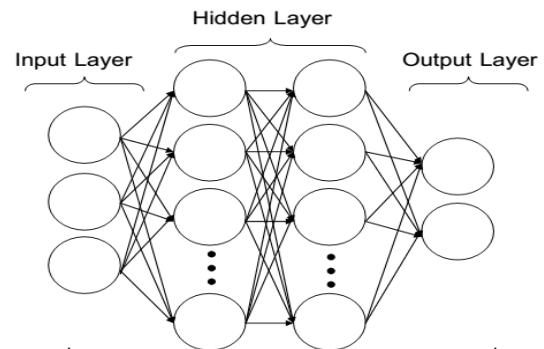
แบบจำลองจะประกอบด้วย ชั้นขาเข้า (Input Layer) ชั้นขาออก (Output Layer) และ ชั้นซ่อน (Hidden Layer) [5] ดังรูปที่ 2

ชั้นขาเข้าเป็นชั้นแรกในโครงสร้างของแบบจำลองโครงข่ายประสาทเทียม จะประกอบด้วย โหนดรับข้อมูล (Input Nodes) ทำหน้าที่รับข้อมูลเข้าสู่แบบจำลองโครงข่ายประสาทเทียม

ชั้นขาออก เป็นชั้นสุดท้ายในโครงสร้างของแบบจำลองโครงข่ายประสาทเทียม จะประกอบด้วย โหนดแสดงผลข้อมูล (Output Nodes) ทำหน้าที่แสดงผลลัพธ์ (Output) ของแบบจำลองโครงข่ายประสาทเทียม

ชั้นซ่อน (Hidden Layer) จะอยู่ระหว่างชั้นขาเข้าและชั้นขาออก ประกอบด้วยโหนด ในชั้นนี้ทำหน้าที่รวมข้อมูลจากชั้นขาเข้าและทำหน้าที่ปรับน้ำหนัก (Weight) เพื่อแปลงให้เป็นผลลัพธ์ (Output) และส่งข้อมูลไปยังชั้นขาออกเพื่อปรับแก้ ซึ่งชั้นซ่อนและจำนวนโหนดนั้นสามารถมีได้หลายชั้นและหลายโหนดตามความเหมาะสมของการวิเคราะห์หรือการ

สร้างแบบจำลอง ในหลายงานวิจัยการหาจำนวนชั้นซ่อนและจำนวนโหนดที่เหมาะสมสามารถหาได้จากกระบวนการลองผิดลองถูก (Trial-Error) [10,12]



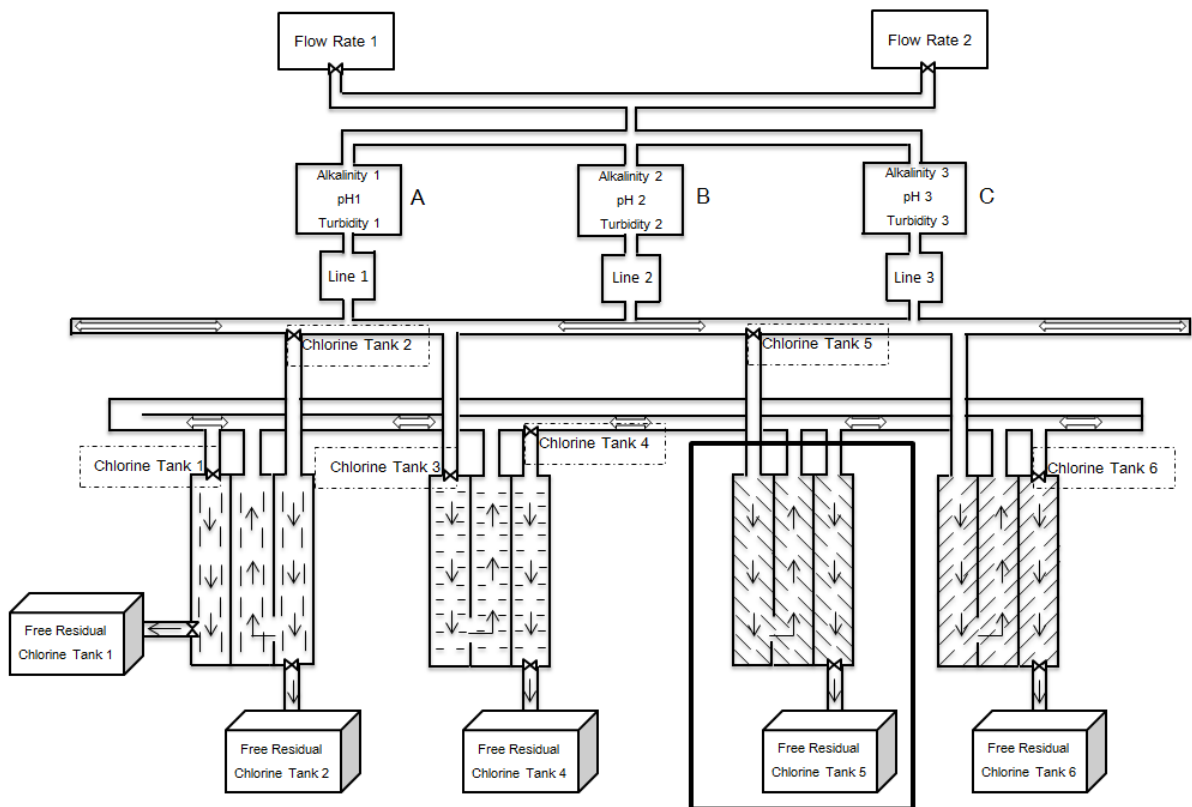
รูปที่ 2 แสดงถึงระบบโครงข่ายประสาทเทียมที่มีจำนวนชั้นทั้งหมด 3 ชั้น

2.2 การเก็บข้อมูล

ข้อมูลที่ใช้ในการทดลองเป็นข้อมูลในการเดินระบบของโรงงานผลิตน้ำประปาบางเขน การประมาณคร่าวๆเป็นข้อมูลรายวัน โดยเก็บเป็นระยะเวลา 1 ปี ตั้งแต่วันที่ 1 มกราคม พ.ศ. 2553 ถึงวันที่ 31 ธันวาคม พ.ศ. 2553 โรงผลิตน้ำประปาบางเขนมีการเก็บข้อมูล 2 แบบ คือการเก็บข้อมูลโดยผู้ปฏิบัติการ (Laboratory) และช่างผู้ชำนาญ (Operator)

ตารางที่ 1 ชนิดของข้อมูล และความถี่ในการเก็บข้อมูลพร้อมตารางที่ 1 ชนิดของข้อมูล และความถี่ในการเก็บข้อมูลพร้อมตำแหน่งที่เก็บข้อมูล

ข้อมูล	ความถี่ (ชั่วโมง)	การเก็บข้อมูล	ตำแหน่งที่เก็บข้อมูล
ความเป็นด่าง (Alkalinity) (mg/l CaCO ₃)	4	Operator	A,B,C
ความเป็นกรด-เบส (pH) (pH units)	4	Operator	A,B,C
ความขุ่น (Turbidity) (NTU)	4	Operator	A,B,C
คลอรีน (Chlorine) (ppm)	2	Laboratory	Chlorine Tank 2 Chlorine Tank 5
คลอรีนอิสระคงเหลือ (Free Residual Chlorine) (ppm)	2	Laboratory	Free Residual Chlorine Tank 5



รูปที่ 3 กระบวนการเติมคลอรีนของโรงงานผลิตน้ำประปาบางเขน

2.3 การจัดข้อมูลและการวิเคราะห์ข้อมูลเบื้องต้น

การจัดการข้อมูลเป็นสิ่งสำคัญในกระบวนการวิเคราะห์ข้อมูลเนื่องจากการสร้างแบบจำลองโครงข่ายประสาทเทียมใช้หลักการการเรียนรู้ความสัมพันธ์ของตัวแปรต่างๆในการสอนแบบจำลองงานวิจัยนี้จึงได้ทำการกรองข้อมูลโดยกำหนดค่ามากที่สุด (Maximum) และค่าน้อยที่สุด (Minimum) เป็นเกณฑ์การคัดกรองของแต่ละตัวแปรในการสร้างแบบจำลองโครงข่ายประสาทเทียมจำเป็นต้องใช้ข้อมูลอนุกรมเวลาที่มีความถี่ของเวลาที่เท่ากันดังนั้นในงานวิจัยนี้จึงได้ทำการปรับข้อมูลที่มีความถี่ในการเก็บข้อมูล จาก 4 ชั่วโมง เป็น 2 ชั่วโมง เพื่อให้มีความสัมพันธ์กับค่าคลอรีนที่นำมาทำนายในแบบจำลอง

การวิเคราะห์ข้อมูลทางสถิติทำได้โดยวิเคราะห์ตัวแปรทางสถิติพื้นฐาน เช่น ค่าเฉลี่ย (Mean) และช่วงของข้อมูล (Range) เพื่อให้เกิดความ

เข้าใจในตัวแปรนั้นๆและสามารถทำการวิเคราะห์ความสัมพันธ์เบื้องต้นระหว่างตัวแปรขาเข้า (Input) และตัวแปรขาออก (Output) โดยการใช้ค่าการถดถอยแบบเชิงเส้น (Linear Regression Analysis) [7] ในการวิเคราะห์ความสัมพันธ์ดังกล่าว จะใช้ค่าดังกล่าวเป็นแนวทางในการเลือกตัวแปรขาเข้าที่มีความสัมพันธ์อย่างมีนัยยะกับตัวแปรขาออก และเป็นการป้องกันการนำข้อมูลที่มีสิ่งปลอมปน (Noisy Data) มาใช้

2.4 การเตรียมข้อมูลในการสร้างแบบจำลองโครงข่ายประสาทเทียม

2.4.1 การจัดเตรียมข้อมูล

การจัดเตรียมข้อมูลในการสร้างแบบจำลองโครงข่ายประสาทเทียมประกอบด้วย 3 ขั้นตอน คือ การจัดแบ่งข้อมูล การเลือกข้อมูลขาเข้าของ

แบบจำลอง และการลดขนาดของข้อมูลด้วยการ Normalization [4]

2.4.2 การจัดแบ่งข้อมูล

การจัดแบ่งข้อมูลเนื่องจากแบบจำลองโครงข่ายประสาทเทียมเป็นแบบจำลองที่ต้องอาศัยข้อมูลเป็นจำนวนมาก เพื่อการสอนให้แบบจำลองโครงข่ายประสาทเทียมจดจำรูปแบบและนำไปใช้ (Pattern Recognition) ในการสร้างแบบจำลองโครงข่ายประสาทเทียมจะต้องทำการแบ่งข้อมูลออกเป็น 3 ส่วนคือ Training Set, Test Set และ Validation Set เพื่อเป็นการลดการผิดพลาด และเพิ่มประสิทธิภาพของแบบจำลองการแบ่งชุดข้อมูลควรให้ชุดข้อมูลทั้ง 3 ชุดเป็นตัวแทนซึ่งกันและกันโดยมีค่าเฉลี่ยใกล้เคียงกัน นอกจากนี้แล้วข้อมูลของ Training Set จะต้องครอบคลุมข้อมูลของ Test Set และ Validation Set ในงานวิจัยนี้จะแบ่งข้อมูลโดยมีอัตราส่วนของจำนวนข้อมูลใน Training Set: Test Set: Validation Set เท่ากับ 4:1:1 [1]

การเลือกตัวแปรขาเข้าสำหรับแบบจำลองที่เกี่ยวข้องกับข้อมูลอนุกรมเวลา จะต้องคำนึงถึงปัจจัยสำคัญ 2 ปัจจัย คือ ชนิดของตัวแปรขาเข้า และเวลาย้อนหลัง (Time Lags) ในการเลือกชนิดของตัวแปรจะพิจารณาโดยใช้วิธีการวิเคราะห์ความถดถอยแบบเชิงเส้น (Linear Regression Analysis) [7] โดยพิจารณาค่าสัมประสิทธิ์ความถดถอย (Regression Coefficient) ซึ่งแสดงความสัมพันธ์ระหว่างตัวแปรขาเข้ากับตัวแปรขาออก ในกรณีเลือกช่วงเวลาย้อนหลังที่เหมาะสม จะพิจารณาโดยใช้ค่า Autocorrelation (ACF) [2]

การพิจารณาการเลือกตัวแปรขาเข้าสำหรับแบบจำลองโครงข่ายประสาทเทียมเพื่อทำนายค่าคลอรีนอิสระคงเหลือ (Free Residual Chlorine) จะใช้ค่าสัมประสิทธิ์ความถดถอย (Linear Regression Analysis) [7] ระหว่างตัวแปรขาเข้ากับตัวแปรขาออก ดังตารางที่ 2 ถ้าค่าสัมประสิทธิ์ความถดถอย (R) ที่มีค่าน้อยกว่า 0.5 จะถือว่าตัวแปรดังกล่าวส่งผลน้อยกับตัวแปรขาออก และไม่ควรรนำไปใช้เป็นตัวแปรขาเข้าของแบบจำลองค่าสัมประสิทธิ์ความถดถอย มีค่าอยู่ในช่วง 0.0519 – 0.1291 แสดงถึงความไม่มีนัย

สัมพันธ์ระหว่างตัวแปรขาเข้าและตัวแปรขาออก เป็นเพราะว่าความสัมพันธ์ของตัวแปรทั้งหมดเป็นความสัมพันธ์แบบไม่เชิงเส้น (Non Linear) เป็นผลให้ค่าสัมประสิทธิ์ความถดถอยมีค่าน้อยมากในกรณีของการมีความสัมพันธ์แบบไม่เชิงเส้นนี้แบบจำลองโครงข่ายประสาทเทียมจึงเป็นหนึ่งในตัวเลือกที่เหมาะสมที่สุดที่จะนำมาเป็นเครื่องมือในการสร้างแบบจำลองของความสัมพัทธ์แบบไม่เชิงเส้น โดยใช้ตัวแปรขาเข้าทั้งหมดเป็นตัวแปรขาเข้าของแบบจำลอง [8, 9, 14]

ตารางที่ 2 ความสัมพันธ์ระหว่างค่าคลอรีนอิสระคงเหลือกับตัวแปรต่างๆที่ใช้ในแบบจำลอง

ตัวแปร	R
ความเป็นด่าง (mg/l CaCO ₃)	-0.13
ความเป็นกรด-เบส (pH units)	0.05
ความขุ่น (NTU)	-0.02
อัตราการไหล(Q,m ³ /d)	0.10
คลอรีนถึงที่ 2 (ppm)	-0.16
คลอรีนถึงที่ 5 (ppm)	-0.16

2.4.3 การกำหนดตัวแปร

การกำหนดตัวแปรขาเข้าของแบบจำลองโดยใช้ค่า Autocorrelation Function (ACF) [2] ค่า ACF เป็นค่าที่ใช้แสดงผลข้อมูลในอดีตของข้อมูลอนุกรมเวลาที่มีผลกับค่าในเวลาปัจจุบัน ค่า ACF สามารถหาได้จากความสัมพันธ์ดังต่อไปนี้

ค่าสัมประสิทธิ์ของ Autocorrelation ที่ Lag k คือ

$$C_k = \frac{1}{N-k} \sum_{t=1}^{N-k} (x_t - \bar{x})(x_{t+k} - \bar{x}) \quad (1)$$

โดยที่ k คือ lag , $k = 0, 1, \dots, N$ ถ้าถูกตั้งค่าเป็น 0 หมายถึงค่าปัจจุบัน และ \bar{x} คือค่าเฉลี่ยทั้งหมดของชุด ถูกนิยามเป็น

$$\bar{x} = \frac{1}{N} \sum_{t=0}^N x_t \quad (2)$$

Autocorrelation ที่ lag k หาได้จาก

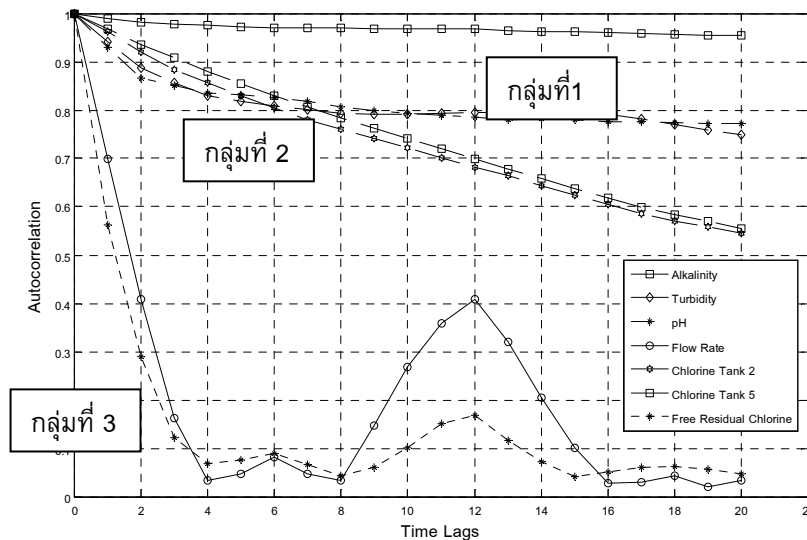
$$ACF(k) = \frac{C_k}{C_0} \quad (3)$$

เป็นค่าที่แสดงผลของข้อมูลในอดีตของข้อมูลอนุกรมเวลา (Time Series) ที่มีผลกับค่าในปัจจุบัน ในงานวิจัยครั้งนี้ได้ทำการหาค่า ACF ของอนุกรมเวลาของตัวแปรขาเข้าทั้งหมด โดยทำการแปรค่าเวลาย้อนหลังจากเวลาปัจจุบัน (Lag ที่ 0) จนถึงเวลาย้อนหลัง 40 ชั่วโมง (Lags ที่ 20) โดยแสดงผลอยู่ในรูปที่ 4 โดยภาพรวมจะแยกการพิจารณาในกลุ่มของข้อมูลอนุกรมเวลาออกเป็น 3 กลุ่ม ตามลักษณะการลดลงของค่า ACF เมื่อเพิ่มช่วงเวลาย้อนหลังถ้าค่า ACF ที่ช่วงเวลาย้อนหลังใดๆของอนุกรมเวลามีค่าน้อยกว่า ± 0.5 ถือว่าเป็นข้อมูลที่มีสิ่งปลอมปน (Noise data) และจะไม่นำไปใช้เป็นข้อมูลตัวแปรขาเข้าของแบบจำลองโครงข่ายประสาทเทียม [10]

กลุ่มที่ 1 (อนุกรมเวลาของค่าความเป็นด่าง (Alkalinity)) ค่า ACF ของอนุกรมเวลาของค่าความเป็นด่างมีค่าลดลงน้อยมาก เมื่อเวลาย้อนหลังเพิ่มขึ้นนั้นหมายความว่า ข้อมูลในอดีตของค่าความเป็นด่างมีผลกับข้อมูลในปัจจุบันอย่างมาก

กลุ่มที่ 2 (อนุกรมเวลาของค่าความขุ่น (Turbidity) พีเอช (pH) และค่าความเข้มข้นของคลอรีน (Chlorine) ถึงที่ 2 และ ถึงที่ 5) เมื่อมาพิจารณาตัวแปรในกลุ่มของความขุ่น พีเอช ความเข้มข้นของคลอรีนถึงที่ 2 และ ถึงที่ 5 พบว่าค่า ACF แปรอยู่ระหว่าง 1.0 - 0.6 เมื่อเวลาย้อนหลังเพิ่มขึ้นถึง 40 ดังนั้นข้อมูลทั้งหมดในช่วงเวลาย้อนหลัง 40 ชั่วโมง จึงเป็นข้อมูลที่มีประโยชน์ในการใช้สร้างแบบจำลอง ANN

กลุ่มที่ 3 (อนุกรมเวลาของคลอรีนอิสระคงเหลือ (Free Residual Chlorine) และค่าอัตราการไหล (Flow Rate)) เมื่อมาพิจารณาข้อมูลของกลุ่มที่มีค่า ACF ลดลงอย่างมากในช่วงแรก (ข้อมูลอนุกรมเวลา คลอรีนอิสระคงเหลือและค่าอัตราการไหล) พบว่าข้อมูลย้อนหลังเกินกว่า 4 ชั่วโมง ไม่ควรนำมาใช้เป็นข้อมูลตัวแปรขาเข้าเนื่องจากเป็นข้อมูลที่มีสิ่งปลอมปน (Noise Data)



รูปที่ 4 Autocorrelation (ACF) ของตัวแปรขาเข้าและตัวแปรขาออกของค่าปริมาณคลอรีนอิสระคงเหลือที่ใช้ในการเลือกเวลาย้อนหลัง (Time Lags)

ดังนั้นจากการพิจารณาค่า ACF ของข้อมูลทั้งสามกลุ่มแล้วจึงไม่ควรนำข้อมูลย้อนหลังไม่เกิน 4 ชั่วโมงมาใช้เป็นตัวแปรขาเข้า อย่างไรก็ตามในการแปรค่าเวลาย้อนหลังจะทำการแปรค่าจากเวลาปัจจุบัน (Lag ที่ 0) จนถึงเวลาย้อนหลัง 12 ชั่วโมง (Lags ที่ 6) เพื่อให้สามารถพิจารณาประสิทธิภาพการทำนายของแบบจำลองเมื่อมีค่าเวลาย้อนหลังมากและน้อยกว่าช่วงเวลาย้อนหลังที่เหมาะสม

2.4.4 Normalization

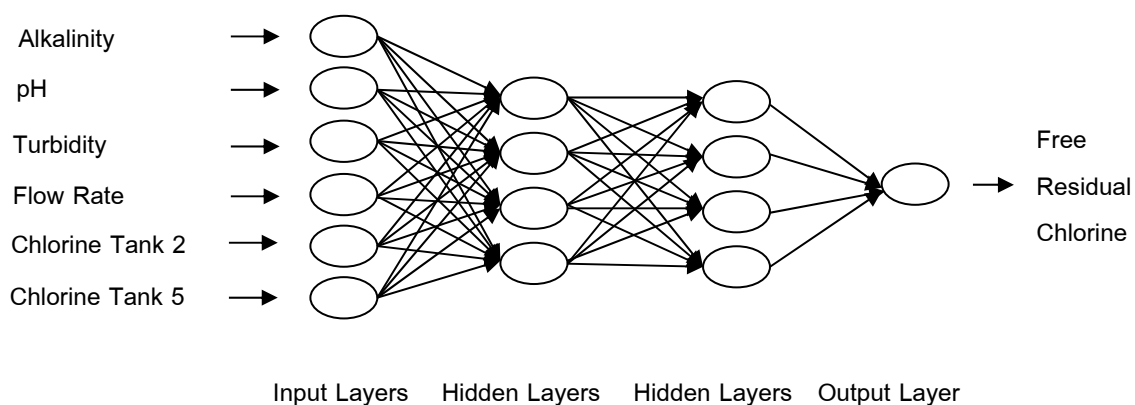
การทำ Normalization [4] เป็นวิธีการลดขนาดของช่วงข้อมูลที่ต่างกัน โดยเป็นการทำการจัดค่าของข้อมูลให้อยู่ในช่วงเดียวกัน การจัดข้อมูลให้เหมาะสมกับการสอน และการเรียนรู้แบบจำลองโครงข่ายประสาทเทียม โดยเป็นการนำข้อมูลขาเข้าและขาออกที่ได้ทำการ Normalization [4] ตามสมการ (4) ไปใช้เพื่อที่จะได้ข้อมูลที่มีช่วงข้อมูลขนาดเดียวกันเพื่อที่จะทำให้ตัวแปรทั้งหมดได้รับความสนใจในการ Training ที่เท่ากัน ถ้าตัวแปรมีขนาดช่วงข้อมูลที่ต่างกัน จะทำให้ตัวแปรบางตัวมีความสำคัญมากกว่าความเป็นจริง ดังนั้นจึงมีความจำเป็นจะต้องปรับปรุงข้อมูลให้มีขนาดข้อมูลเดียวกัน เป็นผลให้ไม่มีตัวแปรใดมีความสำคัญมากกว่าความเป็นจริง

$$x_{nor} = \frac{(x - x_{min})(0.75 - (-0.75))}{(x_{max} - x_{min})} + (-0.75) \quad (4)$$

- เมื่อ
- x = ค่าของข้อมูล
 - x_{max} = ค่าที่มากที่สุดของข้อมูล
 - x_{min} = ค่าที่น้อยที่สุดของข้อมูล

2.5 การบวนการสร้างแบบจำลอง

รูปแบบโครงสร้างของแบบจำลองโครงข่ายประสาทเทียม ที่ใช้ในงานวิจัยนี้ ประกอบด้วยชั้นขาเข้าที่มีข้อมูลจำนวน 7 ตัวแปร ดังรูปที่ 5 ที่เวลาปัจจุบัน (t) และเวลาย้อนหลังต่างๆ (t-1, t-2, t-3, t-4, t-5 และ t-6) และชั้นขาออกที่ต้องการทำนาย 1 ตัวแปร (ค่าคลอรีนอิสระคงเหลือ Free Residual Chlorine, t+1) โดยการทดลองหาสถาปัตยกรรมที่เหมาะสม (ขนาดหรือจำนวนของโหนดในชั้นซ่อนที่เหมาะสมกับแบบจำลอง) โดยจะทำการทดลอง 2 กรณี คือ กรณีแรกจะใช้ชั้นซ่อนจำนวน 1 ชั้นซ่อน และกรณีที่สองจะใช้ชั้นซ่อนจำนวน 2 ชั้นซ่อน โดยที่โหนดในแต่ละชั้นซ่อนจะเริ่มตั้งแต่ 5 โหนดเพิ่มขึ้นทีละ 5 โหนดจนถึง 100 โหนด (5, 10, 15, ..., 100)



รูปที่ 5 แบบจำลองโครงข่ายประสาทเทียม

2.5.1 ประสิทธิภาพของแบบจำลอง

ประสิทธิภาพของแบบจำลองโครงข่ายประสาทเทียมสามารถทำการประเมินโดยวัดจากความผิดพลาดของการทำนายบน Test Set ความผิดพลาดในการทำนายสามารถวัดได้โดยใช้ตัวแปรทางสถิติเช่น r^2 หรือ Mean Absolute Error (MAE) [3] ค่า r^2 และค่า MAE สามารถหาได้จากสมการดังต่อไปนี้

$$r^2 = 1 - \frac{SS_r}{SS_T} \quad (5)$$

โดยที่ SS_r คือ ค่าผลรวมกำลังสองของ residual
 SS_T คือ ค่าผลรวมกำลังสองทั้งหมด

$$SS_r = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (6)$$

$$SS_T = \sum_{i=1}^n (y_i - \bar{y})^2 \quad (4)$$

โดยที่ y_i คือ ค่าจากการสังเกต
 \hat{y}_i คือ ค่าจากการทำนายสำหรับ y_i
 \bar{y} คือ ค่าเฉลี่ยของการสังเกต

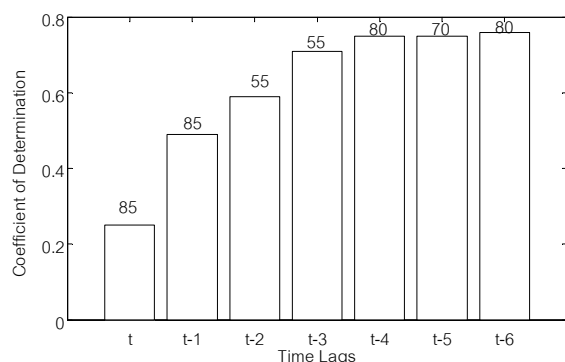
$$MAE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i) \quad (7)$$

โดยที่ y_i คือ ค่าจากการสังเกต
 \hat{y}_i คือ ค่าจากการทำนาย y_i

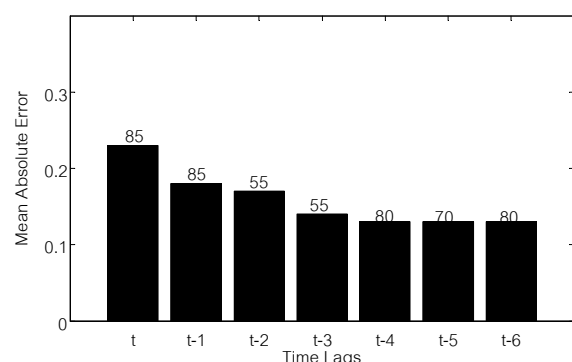
3. ผลการทดลอง

การแบ่งข้อมูลของแบบจำลองโครงข่ายประสาทเทียมแบ่งข้อมูลออกเป็น 3 ชุดด้วยกันคือ Training Set, Test Set และ Validate Set ในอัตราส่วน 4:1:1 ตามลำดับ ค่าที่เหมาะสมของสถาปัตยกรรมโครงข่ายประสาทเทียม (ANN) ที่มีชั้นซ่อน (Hidden layer) 1 ชั้น และ 2 ชั้น

สถาปัตยกรรมแบบชั้นซ่อน 1 ชั้น ประสิทธิภาพการทำนายของโครงข่ายประสาทเทียมที่มี 1 ชั้นซ่อนแสดงอยู่ในรูปที่ 6 และ 7 สรุปอยู่ในตารางที่ 3 โดยในรูปที่ 6 แสดง r^2 ที่มากที่สุด ประสิทธิภาพการทำนายสูงสุดของแต่ละเวลาย้อนหลัง และในรูปที่ 7 แสดงค่า MAE ที่น้อยที่สุด (ประสิทธิภาพการทำนายสูงสุด) ของแต่ละเวลาย้อนหลังจากผลการทดลองดังกล่าว พบว่าสำหรับแบบจำลองโครงข่ายประสาทเทียมที่มีชั้นซ่อน 1 ชั้น เวลาย้อนหลังที่เหมาะสมคือ 12 ชั่วโมง และจำนวนโหนดเท่ากับ 80 โหนด โดยแบบจำลองนี้จะให้ค่า r^2 เท่ากับ 0.76 และค่า MAE เท่ากับ 0.13 mg/l 0.76 และค่า MAE เท่ากับ 0.13 mg/l



รูปที่ 6 การเปรียบเทียบค่า Coefficient of Determination



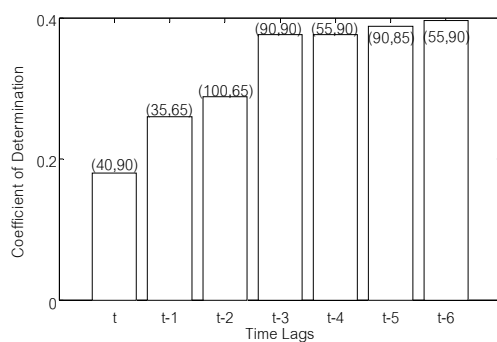
รูปที่ 7 การเปรียบเทียบค่า Mean Absolute Error

ตารางที่ 3 ประสิทธิภาพของแบบจำลองโครงข่ายประสาทเทียมที่มี 1 ชั้นซ่อน

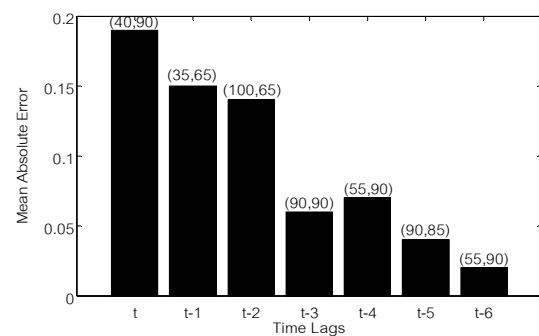
เวลาย้อนหลัง (Time Lags) (hours)	จำนวนโหนด	Coefficient of Determination (r^2)	Mean Absolute Error (MAE) (mg/l)
ปัจจุบัน	85	0.25	0.23
2	85	0.49	0.18
4	55	0.59	0.17
6	55	0.71	0.14
8	80	0.75	0.13
10	70	0.75	0.13
<u>12</u>	<u>80</u>	<u>0.76</u>	<u>0.13</u>

ประสิทธิภาพการทำนายของโครงข่ายประสาทเทียมที่มี 2 ชั้นซ่อนแสดงผลการเปรียบเทียบ Coefficient of Determination และ Mean Absolute Error ที่ดีที่สุดของสถาปัตยกรรมโครงข่ายประสาทเทียมที่เวลาย้อนหลังต่างๆในรูปที่ 7, 8 และในตารางที่

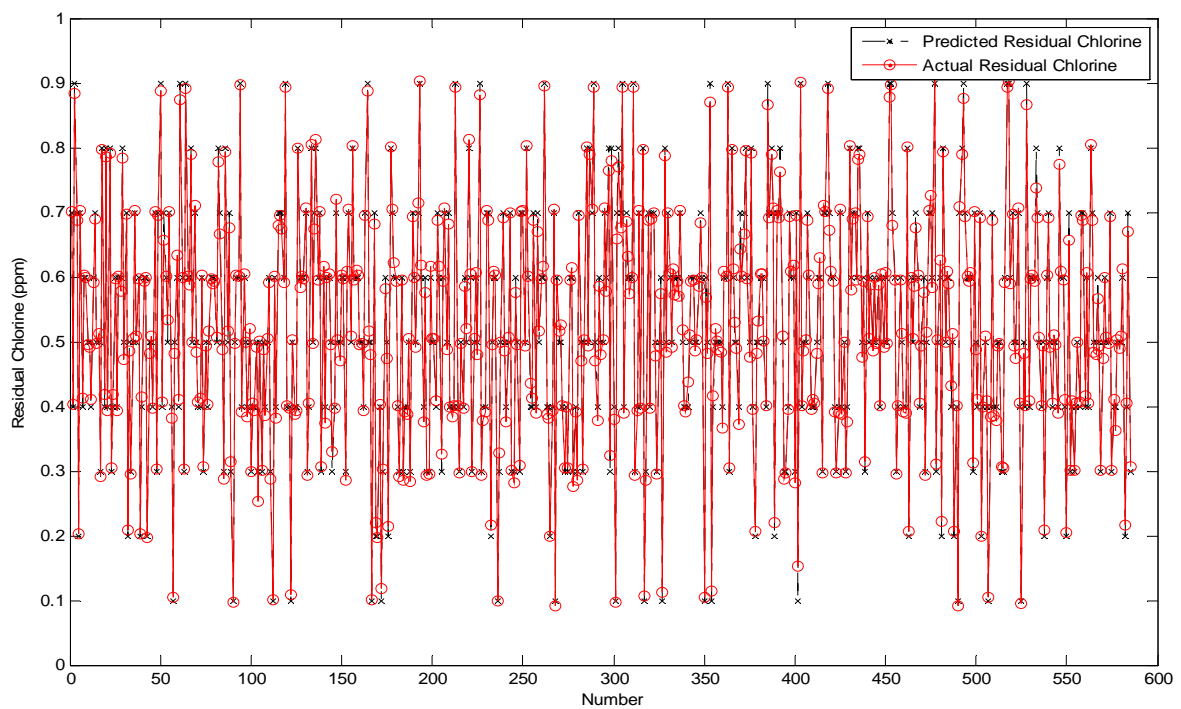
4 พบว่าสำหรับแบบจำลองโครงข่ายประสาทเทียมที่มี 2 ชั้นซ่อนเวลาย้อนหลังที่เหมาะสมคือ 12 ชั่วโมง และจำนวนโหนดจะเท่ากับ (55,90) โหนดโดยแบบจำลองนี้จะให้ค่า Coefficient of Determination เท่ากับ 0.99 และค่า Mean Absolute Error เท่ากับ 0.02 mg/l



รูปที่ 8 การเปรียบเทียบค่า Coefficient of Determination



รูปที่ 9 การเปรียบเทียบค่า Mean Absolute Error



รูปที่ 10 แสดงค่าคลอรีนอิสระที่ทำนายโดยแบบจำลองของสิ่งที่ดีกว่าเปรียบเทียบกับค่าคลอรีนอิสระของ Test Set

ตารางที่ 4 ประสิทธิภาพของแบบจำลองโครงข่ายประสาทที่มี 2 ชั้นซ่อน

เวลา ย้อยหลัง (Time Lags)	จำนวนชั้นซ่อน	จำนวนชั้นซ่อน	Coefficient of Determination (r^2)	Mean Absolute Error (MAE) (mg/l)
(hours)	1 st hidden	2 nd hidden		
ปัจจุบัน	40	90	0.45	0.19
2	35	65	0.65	0.15
4	100	65	0.72	0.14
6	90	90	0.94	0.06
8	55	90	0.94	0.07
10	90	85	0.97	0.04
<u>12</u>	<u>55</u>	<u>90</u>	<u>0.99</u>	<u>0.02</u>

เมื่อเปรียบเทียบกับความคลาดเคลื่อน (Accuracy) ของเครื่องวัดคลอรีนซึ่งมีค่าเท่ากับ $\pm 5\%$ และค่าคลอรีนอิสระคงเหลือที่ออกแบบเท่ากับ 0.7 mg/l ดังนั้นค่าความคลาดเคลื่อนของการวัดจะมีค่าเท่ากับ ± 0.035 mg/l เมื่อนำมาเปรียบเทียบกับแบบจำลองแบบ 1 ชั้นซ้อนพบว่าค่า MAE มีค่ามากกว่า (0.13 mg/l) ดังนั้นประสิทธิภาพการทำนายของแบบจำลองโครงข่ายประสาทเทียมที่มีชั้นซ้อน 1 ชั้นไม่สามารถยอมรับได้

แบบจำลองที่มี 2 ชั้นซ้อน เวลาย้อนหลังที่เหมาะสมคือ 12 ชั่วโมง และจำนวนโหนดเท่ากับ (55, 90) โหนด โดยแบบจำลองนี้จะให้ค่า r^2 เท่ากับ 0.99 และค่า MAE เท่ากับ 0.02 mg/l เมื่อนำมาเปรียบเทียบกับความคลาดเคลื่อนของเครื่องมือวัดพบว่า ค่า MAE มีค่าน้อยกว่า (0.02 mg/l) ดังนั้นประสิทธิภาพการทำนายของแบบจำลองโครงข่ายประสาทเทียมที่มี 2 ชั้นซ้อนเป็นที่ยอมรับได้ดังรูปที่ 10

4. วิเคราะห์ผลการทดลอง

ผลการทดลองของการทำนายแบบจำลองที่มี 1 ชั้นซ้อน ไม่สามารถยอมรับได้ เนื่องจากค่าความคลาดเคลื่อนของแบบจำลอง (0.13 มิลลิกรัมต่อลิตร) มีค่ามากกว่าค่าความคลาดเคลื่อนของเครื่องมือวัดคลอรีน (± 0.035 มิลลิกรัมต่อลิตร) และแบบจำลองที่มี 2 ชั้นซ้อนเป็นที่ยอมรับได้เนื่องจากค่า MAE ของแบบจำลองมีค่าต่ำกว่าค่าความคลาดเคลื่อนของเครื่องมือวัดคลอรีน (± 0.035 มิลลิกรัมต่อลิตร) โดยปกติแล้วการประสานครหลวงกำหนดให้ค่าความคลอรีนอิสระคงเหลือมีค่าเท่ากับ 0.7 มิลลิกรัมต่อลิตร

นอกจากนี้ยังพบว่าเมื่อพิจารณาค่าเวลากักเก็บ (HRT) ซึ่งในการออกแบบถังเติมคลอรีนใช้เวลากักเก็บ (Hydraulic Retention Time, HRT) ประมาณ 20-45 นาที แต่แบบจำลองโครงข่ายประสาทเทียมต้องใช้ข้อมูลเวลาย้อนหลังถึง 12 ชั่วโมง เพื่อให้ได้ผลดีที่สุด แสดงว่าผลของการเดินระบบ (Operate)

และคุณภาพน้ำที่เข้าสู่ถังคลอรีนตั้งแต่เวลาปัจจุบันถึงเวลาย้อนหลัง 12 ชั่วโมงมีผลกับค่าความเข้มข้นของคลอรีนอิสระคงเหลือ ในกรณีนี้สามารถอธิบายได้ว่าอนุภาคหรือสิ่งปนเปื้อนซึ่งมีผลต่อปริมาณคลอรีนที่เหลืออยู่ ใช้เวลาอยู่ในถังเติมคลอรีนมากกว่าเวลากักเก็บ (HRT) เนื่องจากรูปร่างของถังตกตะกอนที่มีลักษณะคอคเคียว ทำให้คุณสมบัติของน้ำเข้าระบบที่เวลามากกว่าเวลากักเก็บ จะส่งผลมาถึงความเข้มข้นของคลอรีนที่เหลืออยู่ นอกจากนี้แล้ว อัตราการไหลและปริมาณคลอรีนที่เติมเป็นตัวแปรในการเดินระบบที่กำหนดโดยผู้ปฏิบัติการ (Operator)

5. สรุป

จากผลการศึกษาและทดลองพบว่าแบบจำลองโครงข่ายประสาทเทียมที่ใช้ในการทำนายค่าคลอรีนอิสระคงเหลือ (Free Residual Chlorine) โดยใช้ข้อมูลตัวแปรทั้งหมดจำนวน 7 ตัวแปร ซึ่งตัวแปรที่ใช้ประกอบด้วยข้อมูลในเวลาปัจจุบันและข้อมูลเวลาย้อนหลัง 2, 4, 6, 8, 10 และ 12 ชั่วโมง โดยมีจำนวนชั้นซ้อน 1 และ 2 ชั้นซ้อน พบว่าแบบจำลองโครงข่ายประสาทเทียมที่มี 1 ชั้นซ้อนเวลาย้อนหลังที่เหมาะสมคือ 12 ชั่วโมง และจำนวนโหนดเท่ากับ 80 โหนด ค่า r^2 เท่ากับ 0.76 และค่า MAE เท่ากับ 0.13 มิลลิกรัมต่อลิตรสำหรับแบบจำลองโครงข่ายประสาทเทียมที่มี 2 ชั้นซ้อนให้ผลดีกว่า เวลาย้อนหลังที่เหมาะสมคือ 12 ชั่วโมง และจำนวนโหนดเท่ากับ (55,90) โหนดโดยแบบจำลองนี้จะให้ค่า r^2 เท่ากับ 0.99 และค่า MAE เท่ากับ 0.02 มิลลิกรัมต่อลิตร

ประสิทธิภาพของการทำนายของแบบจำลองที่มี 2 ชั้นซ้อนเป็นที่ยอมรับได้ เนื่องจากให้ค่าความผิดพลาดน้อยกว่าความคลาดเคลื่อนของเครื่องมือวัด อย่างไรก็ตามแบบจำลองที่มี 2 ชั้นซ้อนเป็นแบบจำลองที่มีศักยภาพสูงกว่าเนื่องจากค่าความคลาดเคลื่อนน้อยกว่าค่าความคลาดเคลื่อนของเครื่องมือวัด

6. กิตติกรรมประกาศ

ขอขอบคุณ การประสานครหลวง Metropolitan Waterworks Authority (Thailand). www.mwa.co.th สำหรับข้อมูลที่ใช้ในงานวิจัยนี้

7. เอกสารอ้างอิง

- [1] Bowden, G. J., Maier, H. R., and Dandy, G. C., 2002. Optimal division of data for neural network models in water resources applications." Water Resour. Res., Vol. 38, No. 2, pp 1010.
- [2] Box, G.E.P., G.M. Jenkins and G.C. Reinsel. 1994. Time Series Analysis Forecasting and Control. 3rd ed. Prentice Hall, Inc., New jersey. 598 p.
- [3] Demuth, H., and Beale, M. (2000). "Neural network toolbox."The MathWorksInc.USA.
- [4] D.R. Baughman , Y. A. Liu 1995. Neural Networks in Bioprocessing and Chemical Engineering,Page 53
- [5] Fausett, L. 1994. Fundamentals of Neural Networks. Prentice Hall, Inc., New jersey. 461 p.
- [6] Gurney K. 1997. An introduction to neural networks.UCL Press. London, UK
- [7] Haan, C.T. 1977. Statistical Methods in Hydrology. The Iowa State University Press-Ames, Iowa. 378 p.
- [8] Haykin, S. S., 1999: Neural networks : a comprehensive foundation. Prentice Hall, 842 pp.
- [9] Ikonen E., Najim K., 2002. Advanced Process Identification and Control, Marcel Dekker Inc., New York, U.S.A., 310 p.
- [10] Maier, H.R. and Dandy, G.C., 2000, Neural Networks for the Prediction and Forecasting of Water Resources Variables: a Review of Modeling Issues and Applications, Environmental Modeling & Software, Vol.15, Page 101-124
- [11] Mario Bunge .1963. "A general black-box theory". Philosophy of Science. Vol. 30.No. 4, pp. 346-358.
- [12] Masters, T. 1993. Practical Neural Network Recipes in C++. California: Academic Press, Inc.
- [13] Metropolitan Waterworks Authority (Thailand). *มาตรฐานคุณภาพน้ำประปาของการประปาครหลวง* [ระบบออนไลน์], แหล่งที่มา: www.mwa.co.th
- [14] Silva Filho, AC et al. Using the Lag of Autocorrelation Function in Order to Identify the Anaerobic Threshold During Dynamic Physical Exercise. Computers in Cardiology 2006; 33: 625-628.
- [15] Theil, H. (1950). "A rank-invariant method of linear and polynomial regression analysis. I, II, III". Nederl. Akad. Wetensch., Proc. 53: 386–392, 521–525, 1397–1412.
- [16] Witbrock, Michael and Zagha, Marco October,1994."Back-Propagation Learning on the IBM GF11," Chapter in Przytula, K.W., and Prasanna Kumar