บทที่ 4

ผลการศึกษา และการอภิปรายผล

ในบทนี้จะแสดงถึงผลการศึกษาเกี่ยวกับความสัมพันธ์ของพารามิเตอร์ต่างๆกับปริมาณออกซิเจนในน้ำ และการแสดงแบบจำลองโครงข่ายใยประสาทประดิษฐ์ที่เหมาะสมในการทำนายปริมาณออกซิเจนที่มีอยู่ในน้ำซึ่งเป็นพารามิเตอร์ที่มีความสำคัญในการบอกถึงคุณภาพน้ำ

4.1 ผลการศึกษาสัมประสิทธิ์สหสัมพันธ์

ผลจากการศึกษาสัมประสิทธิ์สหสัมพันธ์ระหว่างปริมาณค่าออกซิเจนในน้ำ (DO) ในเดือนถัดไปกับข้อมูลคุณภาพน้ำ ได้แก่ ค่าความเป็นกรด-ด่าง (pH) ปริมาณจุลินทรีย์ที่ใช้ในการสลายสารอินทรีย์ในน้ำ (BOD) ปริมาณออกซิเจนที่สารเคมีใช้ในการย่อยสลายอินทรีย์ในน้ำ (COD) ปริมาณของแข็งในน้ำ (SS) ปริมาณไนโตรเจนที่มีอยู่ในน้ำเสีย (TKN) ปริมาณแอมโมเนีย-ไนโตรเจน (NH_3N), ปริมาณไนไตร-ไนโตรเจน (NO_2N), ปริมาณไนเตรท-ไนโตรเจน (NO_3N), ปริมาณฟอสฟอรัสทั้งหมด (T-P) and ปริมาณแบคทีเรียกลุ่มโคลิฟอร์มทั้งหมด (total coliform) ในวันปัจจุบัน ย้อนหลัง 1 เดือน ในพื้นที่เขตดุสิต กรุงเทพมหานคร ดังแสดงในตารางที่ 4.1 โดยพบว่า ความสัมพันธ์ของปริมาณค่าออกซิเจนในน้ำ (DO) ในเดือนถัดมากับตัวแปรต่างๆที่ใช้บ่งบอกคุณภาพน้ำ ของทั้งพื้นที่เขตดุสิตนั้น มีค่าสัมประสิทธิ์สหสัมพันธ์ใกล้เคียงกัน โดยตัวแปร DO ในเดือนถัดมามีความสัมพันธ์กับตัวแปร ปริมาณจุลินทรีย์ที่ใช้ในการสลายสารอินทรีย์ในน้ำ (BOD) และ ปริมาณฟอสฟอรัสทั้งหมด (T-P) มากที่สุด ส่วนความสัมพันธ์ในระดับรองลงมาคือ ตัวแปร TKN, NH_3N, COD, total coliform, pH, NO_2N, NO_3N, และ SS ตามลำดับ ซึ่งตัวแปรที่มีทิศทางความสัมพันธ์ไปในทิศทางเดียวกันกับ DO คือ pH, NO_2N, NO_3N, และ SS ส่วนตัวแปรที่มีความสัมพันธ์กันแบบผกผันกับ DO คือ BOD, T-P, TKN, NH_3N, COD และ total coliform

ตารางที่ 4.1 สัมประสิทธิ์สหสัมพันธ์ระหว่าง ปริมาณค่าออกซิเจนในน้ำ (DO) ในเดือนถัดไป และตัวแปรต่างๆที่ใช้บ่งบอกคุณภาพน้ำ ณ 11 จุดเก็บในพื้นที่เขตดุสิต กรุงเทพมหานคร

ตัวแปรอิสระ / ตัวแปรตาม	pH	BOD	COD	SS	TKN	NH3-N	NO2-N	NO3-N	T-P	T.Coliform
DO	0.24*	-0.64*	-0.27*	0.08*	-0.43*	-0.40*	0.24*	0.20*	- 0.65 *	- 0.27*

หมายเหตุ : * หมายถึง มีความสัมพันธ์กันอย่างมีนัยสำคัญที่ระดับความเชื่อมั่นร้อยละ 95

4.2 ผลการศึกษาแบบจำลองโครงข่ายใยประสาทประดิษฐ์

จากผลการศึกษาสัมประสิทธิ์สหสัมพันธ์ เพื่อนำมาสร้างแบบจำลองโครงข่ายใยประสาทประดิษฐ์ โดยมีปริมาณค่าออกซิเจนในน้ำ (DO) ในเดือนถัดไปเป็นตัวแปรตาม ซึ่งได้แสดงผลใน 11 จุดเก็บในพื้นที่เขตดุสิต กรุงเทพมหานคร ดังที่กล่าวไว้แล้วในหัวข้อที่ 4.1

4.2.1 การสร้างแบบจำลอง

ในที่นี้จะอธิบายแบบจำลองสำหรับทำนาย ปริมาณค่าออกซิเจนในน้ำ (DO) ในเดือนถัดไป ณ 11 จุดเก็บในพื้นที่เขตดุสิต กรุงเทพมหานคร โดยมีขั้นตอนดังนี้

4.2.1.1 การเตรียมข้อมูล การออกแบบ และการสอนโครงข่าย

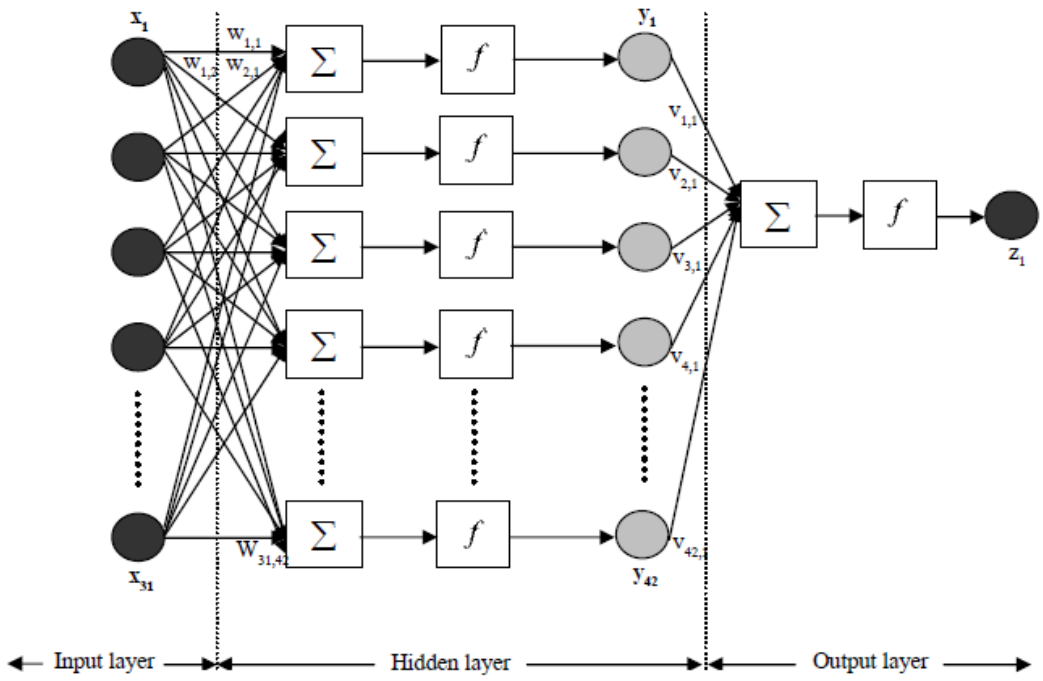
1. เนื่องจากข้อมูลที่นำมาสร้างแบบจำลองมีความแตกต่างเกี่ยวกับลักษณะของโครงสร้างและปัจจัยภายนอกของแต่ละจุดเก็บที่ต่างกัน ดังนั้นจึงต้องทำการปรับค่าของข้อมูล (Normalize) ให้อยู่ในช่วงของข้อมูลที่เหมาะสม ซึ่งในงานวิจัยนี้กำหนดให้ช่วงของข้อมูลอยู่ในช่วงระหว่าง 0.1 และ 0.9 ซึ่งเป็นช่วงข้อมูลที่เหมาะสมในการปรับข้อมูลทางคุณภาพน้ำ (E.Dogan et al, 2008) โดยอธิบายได้จากสมการ (4.1)

$$x_{new} = 0.8\frac{x - x_{\min}}{x_{\max} - x_{\min}} + 0.1 \qquad (4.1)$$

โดยที่ x_{new} คือ ข้อมูลที่ได้รับการ Normalize แล้ว
X_n คือ ข้อมูลที่ต้องการ Normalize
X_{min} คือ ข้อมูลที่มีค่าต่ำที่สุด
X_{max} คือ ข้อมูลที่มีค่าสูงที่สุด

2. แบ่งชุดข้อมูลเป็น 2 ส่วน คือ ชุดสอนโครงข่าย 70% (training set) ชุดทดสอบ 30% (test set)

3. ในการออกแบบขั้นต้นกำหนดให้โครงข่ายมี 3 ชั้น และมีจำนวนโหนดในชั้นซ่อน 30 โหนด โดยที่มีจำนวนอินพุต (ตัวแปรอิสระ) 10 โหนด และจำนวนเอาต์พุต (ตัวแปรตาม)1 โหนด จะได้โครงข่ายที่ออกแบบ คือ 10-30-1 ดังแสดงในรูปที่ 4.1



รูปที่ 4.1 โครงข่ายแบบ 3 ชั้นของ 11 จุดเก็บตัวอย่างในเขตดุสิต

4. กำหนดค่าพารามิเตอร์ที่จำเป็นในโครงข่าย คือ อัตราการเรียนรู้ (η) โมเมนตัม (α) และค่าถ่วงน้ำหนักเริ่มต้น (w) โดยมีค่า 0.05, 0.1 และ 0.2 ตามลำดับ

5. เริ่มทำการสอนโครงข่ายโดยนำชุดสอนโครงข่ายที่ได้แบ่งไว้ 70% (269 ชุดข้อมูล โดยเลือกมา 1 ชุดข้อมูลต่อการสอน 1 รอบ) คำนวณหาค่าเนทของแต่ละโหนดในชั้นซ่อนที่ 1จากการใช้ฟังก์ชันผลรวม (Summation function = Σ) และหลังจากนั้นทำการแปลงค่าโดยใช้ฟังก์ชันซิกมอยด์ (sigmoid function = f) ซึ่งจะทำให้ได้ค่า y1, y2,..., y42 ของชั้นซ่อนที่ 1 (ตัวอย่างการคำนวณหาค่า y1 จะแสดงให้เห็นในสมการที่ (4.2) และ (4.3))

$$(net_1)_h = \sum_{i=1}^{31} w_{ij} x_i = \sum_h \qquad (4.2)$$

$$y_1 = \frac{1}{1+\exp((-net_1)_h)} = f_h \qquad (4.3)$$

6. คำนวณหาค่าเนทของโหนดในชั้นเอาต์พุต และหลังจากนั้นทำการแปลงค่าโดยใช้ฟังก์ชันซิกมอยด์ ซึ่งจะทำให้ได้ค่า z_1 ของชั้นเอาต์พุต (ตัวอย่างการคำนวณหาค่า z_1 จะแสดงให้เห็นในสมการที่ (4.4) และ (4.5))

$$(net_1)_o = \sum_{i=1}^{42} v_{jk} y_i = \sum_o \qquad (4.4)$$

$$z_1 = \frac{1}{1+\exp((-net_1)_o)} = f_o \qquad (4.5)$$

7. จะทำให้ได้โครงข่ายที่มีการเรียนรู้ 1 รอบ จากนั้นนำชุดทดสอบที่ได้แบ่งไว้ 30% (115 ชุดข้อมูล) มาทดสอบความคลาดเคลื่อน (E_{tot}) ดังสมการที่ (4.6) ซึ่งหากค่าคลาดเคลื่อนที่คำนวณได้ยังไม่น้อยกว่าหรือเท่ากับ 0.01 จึงคำนวณเทอมผิดพลาดในชั้นเอาต์พุต (δ_0) เพื่อนำกลับไปปรับแก้ค่าถ่วงน้ำหนักใหม่ (โดยใช้ชุดข้อมูลเดิมที่ได้จากการสอนในรอบนั้น ๆ มาคำนวณเทอมผิดพลาด) ดังสมการที่ (4.7)

$$E_{tot} = \frac{1}{115}\sum_{p=1}^{243} E^p \quad \text{โดยที่} \quad E^p = \frac{1}{2}(t_i - z_i)^2 \qquad (4.6)$$

$$\delta_o = (z_i - t_i) f_o'((net_i)_o) \qquad (4.7)$$

8. หลังจากนั้นคำนวณเทอมผิดพลาดในชั้นซ่อน (δ_h) โดยเริ่มจากชั้นซ่อนชั้นสุดท้าย ดังสมการที่ (4.8)

$$\delta_h = f_h'((net_i)_h)\sum_{i=1}^{30} \delta_o v_{jk} \qquad (4.8)$$

9. จากนั้นเริ่มทำการปรับแก้ค่าถ่วงน้ำหนักก่อนที่จะเริ่มทำการสอนในรอบที่ 2 โดยเริ่มปรับแก้จากชั้นเอาต์พุตก่อน ดังสมการที่ (4.9)

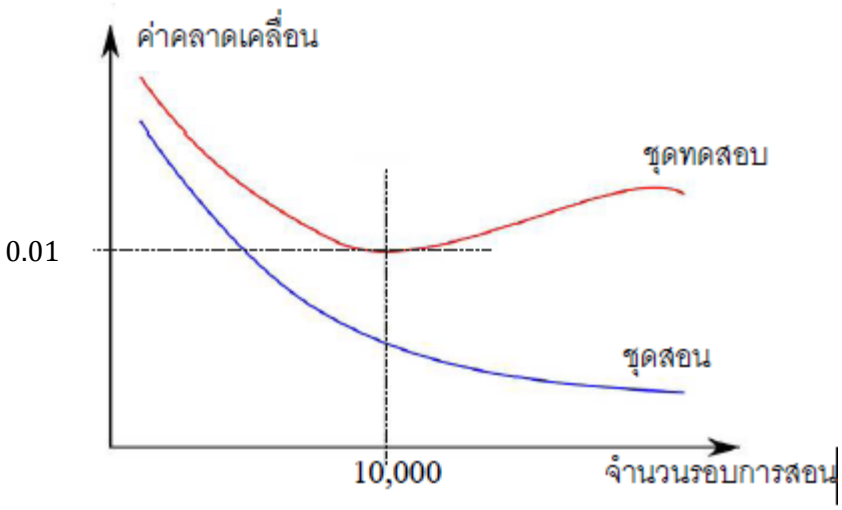
$$v_{(t+1)} = \alpha v_{(t)} + \eta \delta_o y_i \qquad (4.9)$$

โดยที่ t คือ จำนวนรอบการสอนโครงข่าย (t = 1, 2, 3,....,n)

10. หลังจากนั้นจึงปรับแก้ค่าถ่วงน้ำหนักในชั้นซ่อน ดังสมการที่ (4.10)

$$w_{(t+1)} = \alpha w_{(t)} + \eta \delta_h x_i \qquad (4.10)$$

11. หลังจากทำการปรับแก้ค่าถ่วงน้ำหนักในชั้นเอาต์พุต และชั้นซ่อนจึงกลับไปเริ่มสอนโครงข่ายในรอบต่อไปตามขั้นตอนที่ 5 ดังที่กล่าวมาแล้วข้างต้น จนกระทั่งค่าความคลาดเคลื่อนจากการนำชุดทดสอบที่ได้แบ่งไว้ 30% (115 ชุดข้อมูล) ที่นำมาทดสอบในแต่ละรอบของการสอนโครงข่ายมีค่าคลาดเคลื่อนน้อยกว่า 0.01 หรือสอนโครงข่ายได้ 10,000 รอบ (กรณีใดกรณีหนึ่ง) ดังรูปที่ 4.2 จึงจะหยุดสอนโครงข่าย ซึ่งจะได้โครงข่ายของทั้ง 11 จุดเก็บตัวอย่างของคลองในเขตดุสิต สำหรับการปรับค่า 1 ครั้ง



รูปที่ 4.2 จุดสิ้นสุดการสอน

หลังจากนั้นทำการสร้างโครงข่ายอีกครั้งโดยใช้ค่ากำหนดที่เปลี่ยนไป ดังนี้

• กำหนดให้โครงข่ายหนึ่ง ๆ มีการกำหนดค่าพารามิเตอร์ที่จำเป็นในการสอนโครงข่าย คือ อัตราการเรียนรู้ โมเมนตัม และค่าถ่วงน้ำหนักเริ่มต้น หลังจากสิ้นสุดการสอนโครงข่ายแล้ว ให้ปรับเปลี่ยนค่าพารามิเตอร์ทั้ง 3 เป็นดังนี้ อัตราการเรียนรู้ (ปรับค่า 3 ครั้ง) โมเมนตัม (ปรับค่า 4 ครั้ง) และค่าถ่วงน้ำหนักเริ่มต้น (ปรับค่า 4 ครั้ง)

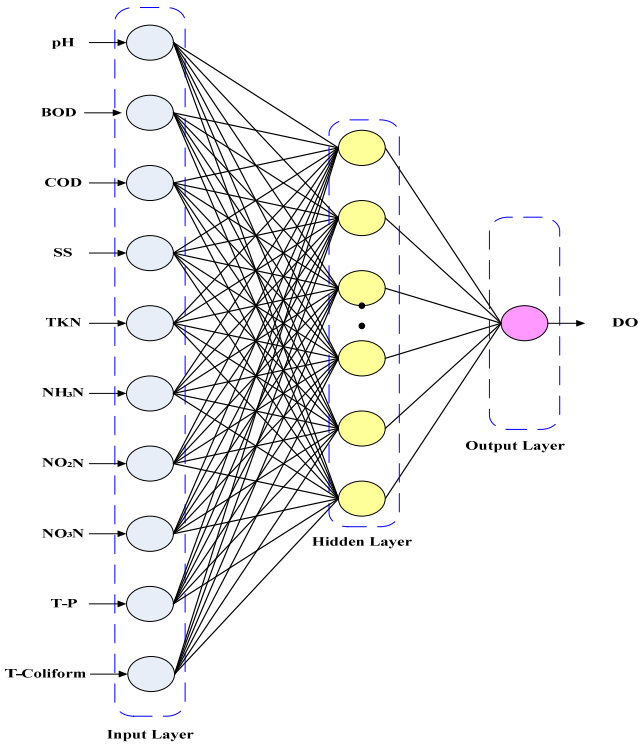
• กำหนดให้โครงข่ายหนึ่ง ๆ มีจำนวนโหนดในชั้นซ่อนเริ่มต้นค่าหนึ่ง หลังจากสิ้นสุดการสอนโครงข่ายแล้ว ให้ปรับเปลี่ยนจำนวนโหนดในชั้นซ่อนอีก 2 ครั้ง คือ เพิ่มและลดจำนวนโหนดจากค่าเริ่มต้น 10 โหนด และ 20 โหนด

4.2.2 แบบจำลองโครงข่ายใยประสาทประดิษฐ์ที่เหมาะสม

จากการสร้างแบบจำลองโครงข่ายใยประสาทประดิษฐ์ดังที่กล่าวในหัวข้อที่ 4.2.1 ได้ผลเป็นแบบจำลองโครงข่ายใยประสาทประดิษฐ์ที่เหมาะสม สำหรับการทำนายปริมาณค่าออกซิเจนในน้ำ (DO) ล่วงหน้า 1 เดือน ในพื้นที่เขตดุสิต กรุงเทพมหานคร 11 จุดเก็บตัวอย่าง ดังตารางที่ 4.2 และรูปที่ 4.3 ซึ่งจะเห็นว่าจำนวนโหนดในชั้นซ่อนที่เหมาะสมกับตัวแปรอิสระทั้ง 10 ตัว (pH) (BOD) (COD) (SS) (TKN) (NH_3N) (NO_2N) (NO_3N) (T-P) และ (total coliform) กับตัวแปรตาม 1 ตัว (DO) คือ 16 โหนด สรุปผลจากตารางแบบจำลองโครงข่ายใยประสาทประดิษฐ์ที่เหมาะสม พบว่าแบบจำลองส่วนใหญ่ต้องมีการแปลงค่าของตัวแปรอิสระให้อยู่ในรูปของลอการิทึมธรรมชาติ ซึ่งสอดคล้องกับสมมติฐานที่ว่าการกระจายตัวของตัวแปรทางสิ่งแวดล้อมส่วนใหญ่จะมีการกระจายตัวอยู่ในรูปแบบ log – normal และพบว่าข้อมูลทางสิ่งแวดล้อมมีความสัมพันธ์กันอย่างซับซ้อนและสอดคล้องกับการศึกษาของ Perez et al., (2000) จากการปรับแก้ค่าพารามิเตอร์ของโครงข่าย พบว่าต้องปรับค่าอัตราการเรียนรู้ (η) ให้สูงกว่าค่าเริ่มต้นที่ 0.05 ซึ่งเป็นผลมาจากโครงข่ายทำการสอนจนครบรอบที่กำหนดก่อนที่จะลู่เข้าหาค่าผิดพลาดที่น้อยที่สุดได้ทำให้ผลทำนายส่วนใหญ่ที่อัตราการเรียนรู้ที่ 0.05 ยังมีค่าคลาดเคลื่อนที่สูง ส่วนค่าถ่วงน้ำหนัก (w) และค่าโมเมนตัม (α) ที่เหมาะสมในโครงข่าย จะมีค่าต่าง ๆ กันขึ้นอยู่กับลักษณะ หรือความสัมพันธ์ของชุดข้อมูลเป็นกรณี ๆ ไป (Gardner and Dorling, 1998; Dimopoulos et al., 1999; Khare and Nagendra, 2006)

ตารางที่ 4.2 แบบจำลองโครงข่ายใยประสาทประดิษฐ์ที่เหมาะสม

โครงสร้าง	ชั้นซ่อน	โหนดในชั้นซ่อน	ค่าสถิติ	
			R	MSE
10-3-1	1	3	0.8	0.8
10-4-1	1	4	0.82	0.76
10-5-1	1	5	0.81	0.76
10-6-1	1	6	0.83	0.75
10-7-1	1	7	0.8	0.9
10-8-1	1	8	0.82	0.77
10-9-1	1	9	0.8	0.8
10-10-1	1	10	0.83	0.7
10-11-1	1	11	0.75	0.88
10-12-1	1	12	0.78	0.83
10-13-1	1	13	0.77	0.86
10-14-1	1	14	0.79	0.79
10-15-1	1	15	0.81	0.79
10-16-1	**1**	**16**	**0.84**	**0.78**
10-17-1	1	17	0.79	0.78
10-18-1	1	18	0.81	0.76
10-19-1	1	19	0.82	0.77
10-20-1	1	20	0.81	0.77
10-21-1	1	21	0.79	0.82
10-22-1	1	22	0.82	0.79
10-23-1	1	23	0.73	0.9
10-24-1	1	24	0.83	0.74
10-25-1	1	25	0.79	0.78
10-26-1	1	26	0.81	0.75
10-27-1	1	27	0.83	0.74
10-28-1	1	28	0.79	0.82
10-29-1	1	29	0.8	0.78
10-30-1	1	30	0.82	0.78



รูปที่ 4.3 โครงสร้างโครงข่ายใยประสาทประดิษฐ์ที่เหมาะสม **10-16-1**

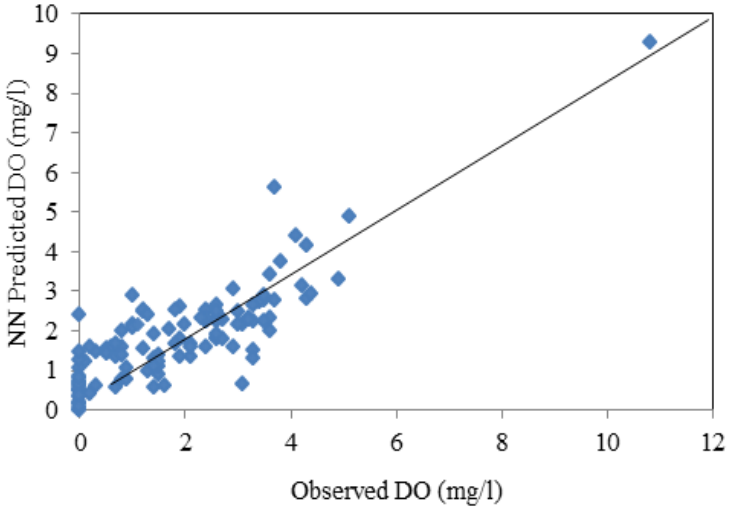
4.3 การศึกษาประสิทธิภาพของแบบจำลอง

ในการวิจัยนี้ได้นำข้อมูลคุณภาพน้ำจากจุดเก็บตัวอย่าง 11 จุดเก็บ ในพื้นที่เขตดุสิต กรุงเทพมหานคร ในช่วงปี 2549 – 2551 มาทำการสร้างแบบจำลองโครงข่ายใยประสาทประดิษฐ์ เพื่อทำนายปริมาณค่าออกซิเจนในน้ำ (DO) ในเดือนถัดไป ดังแสดงไว้ในหัวข้อที่ 4.2 ซึ่งในขั้นต่อไปจะทำการศึกษาประสิทธิภาพของแบบจำลอง โดยทำการสุ่มข้อมูลชุดใหม่ในช่วงปี 2552 – 2553 จำนวน 350 ชุดข้อมูล มาประเมินและวัดประสิทธิภาพของแบบจำลองโดยใช้วิธีทางสถิติดังนี้

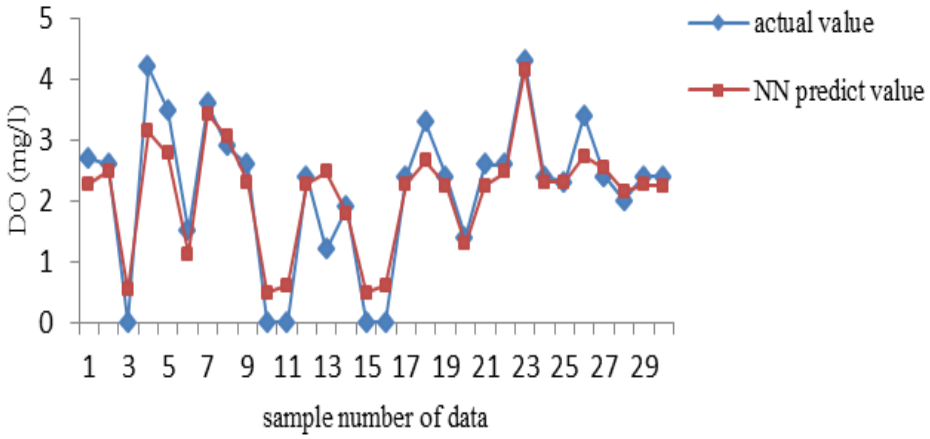
- Mean square error (MSE)
- Correlation (R)

4.3.1 ผลการวัดประสิทธิภาพของแบบจำลอง

จากการวัดประสิทธิภาพแบบจำลองโดยใช้วิธีทางสถิติ ซึ่งมีการนำข้อมูลปริมาณค่าออกซิเจนในน้ำที่ได้จากการเก็บตัวอย่าง (Observed DO) ของแต่ละจุดเก็บในพื้นที่เขตดุสิตมาเปรียบเทียบกับข้อมูลที่ได้จากการทำนายปริมาณค่าออกซิเจนในน้ำโดยโครงข่ายใยประสาทประดิษฐ์ (Neural Network Prediction DO) โดยมีค่าความผิดพลาดกำลังสองเฉลี่ย (MSE) เท่ากับ 0.78 ค่าร้อยละความผิดพลาดเฉลี่ย (MAE) เท่ากับ 0.7 และมีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ 0.84 ดังแสดงในรูปที่ 4.4 และ 4.5



รูปที่ 4.4 กราฟแสดงความสัมพันธ์ระหว่างข้อมูล DO ที่ได้จากการจัดเก็บกับ ข้อมูล DO ที่ได้จากการทำนายโดยใช้โครงข่ายใยประสาทประดิษฐ์



รูปที่ 4.5 กราฟแสดงการเปรียบเทียบกลุ่มตัวอย่างของชุดข้อมูล DO ที่ได้ค่าจากการจัดเก็บกับ ข้อมูล DO ที่ค่าได้จากการทำนายโดยใช้โครงข่ายใยประสาทประดิษฐ์

สรุปผลการวัดประสิทธิภาพของแบบจำลองโดยการทดสอบด้วยตัวชี้วัดทางสถิติพบว่าแบบจำลองโครงข่ายใยประสาทประดิษฐ์สามารถใช้เป็นเครื่องมือในการทำนายปริมาณค่าออกซิเจนในน้ำ ในพื้นที่เขตดุสิต กรุงเทพมหานครได้ ซึ่งจะเห็นได้จากผลของค่าสัมประสิทธิ์สหสัมพันธ์ที่อยู่ในระดับสูง และค่าความคลาดเคลื่อนของข้อมูลที่ไม่มากนัก อย่างไรก็ตามการใช้งานแบบจำลองโครงข่ายใยประสาทประดิษฐ์ที่ถูกสร้างมาจากความสัมพันธ์ของข้อมูลคุณภาพน้ำทั้งทางกายภาพและทางเคมีทั้ง10 พารามิเตอร์ (pH) (BOD) (COD) (SS) (TKN) (NH_3N) (NO_2N) (NO_3N) (T-P) และ (total coliform) เป็นข้อมูลเฉพาะในพื้นที่เขตดุสิต กรุงเทพมหานครเท่านั้น ซึ่งมีลักษณะของความสัมพันธ์เฉพาะที่ จึงมีข้อจำกัดในการทำนาย คือ สามารถทำนายปริมาณค่าออกซิเจนในน้ำ (DO) ในเดือนถัดไปได้เฉพาะพื้นที่นี้เท่านั้น