



Tạp chí Khoa học Trường Đại học Cần Thơ  
Phần A: Khoa học Tự nhiên, Công nghệ và Môi trường

website: [sj.ctu.edu.vn](http://sj.ctu.edu.vn)



DOI:10.22144/ctu.jsi.2017.038

## SO SÁNH KHẢ NĂNG DỰ ĐOÁN CHẤT LƯỢNG NƯỚC SÔNG ĐỒNG NAI BẰNG MẠNG NƠ RON NHÂN TẠO VÀ LÝ THUYẾT XÁM

Nguyễn Hiền Thân<sup>1</sup>, Chế Đình Lý<sup>1</sup> và Phạm Văn Tất<sup>2</sup>

<sup>1</sup>Khoa Tài nguyên Môi trường, Đại học Thủ Dầu Một, tỉnh Bình Dương

<sup>2</sup>Khoa Khoa học Công nghệ, Đại học Hoa Sen, Thành phố Hồ Chí Minh

### Thông tin chung:

Ngày nhận bài: 28/07/2017

Ngày nhận bài sửa: 09/09/2017

Ngày duyệt đăng: 26/10/2017

### Title:

Comparison of artificial neural network and grey theory in predicting water quality of Dong Nai River

### Từ khóa:

Chất lượng nước, dự đoán, lý thuyết xám, mạng nơ ron nhân tạo, so sánh

### Keywords:

Artificial neural network, comparing, forecast, grey theory, water quality

### ABSTRACT

Water pollution has been increasing quickly and complexly for recent years. Water quality forecast to provide prompt and timely information on water pollution is very necessary. In this study, the aim of the study was to compare capability of water quality forecast by the multilayer perceptron neural network method and the grey theory method that emphasized rapid predictability and accuracy, contributing to improving the efficiency of water quality forecast. The data were collected from 23 monitoring stations of Dong Nai river from 2010 to 2014 in Dong Nai and Binh Duong Department of Natural Resources and Environment for forecasting the water quality index. The results showed that both methods had good performance of water quality forecast. The water quality index forecasted by the multilayer perceptron neural network model showed higher accuracy ( $RMSE = 2.88$ ,  $R^2 = 0.987$  and  $P = 0$ ) than that forecasted by the grey theory model ( $RMSE = 7.84$ ,  $R^2 = 0.879$  and  $P = 0$ ).

### TÓM TẮT

Ô nhiễm môi trường nước đang gia tăng nhanh chóng và phức tạp trong những năm gần đây. Dự đoán chất lượng nước nhằm cung cấp thông tin nhanh chóng và kịp thời về tình trạng ô nhiễm nguồn nước là rất cần thiết. Trong nghiên cứu này, khả năng dự đoán chất lượng nước được so sánh bằng phương pháp mạng nơ ron perceptron nhiều lớp và phương pháp lý thuyết xám tập trung khả năng dự đoán nhanh và độ chính xác góp phần nâng cao hiệu quả công tác dự đoán chất lượng nước. Dữ liệu nghiên cứu được thu thập tại 23 điểm quan trắc chất lượng nước Sông Đồng Nai từ 2010 – 2014 tại Sở Tài nguyên và Môi trường tỉnh Đồng Nai và Bình Dương dùng để dự đoán chỉ số chất lượng nước (WQI). Kết quả nghiên cứu cho thấy cả hai phương pháp cho kết quả dự đoán tốt chất lượng nước. Chỉ số chất lượng nước được dự đoán bằng mạng nơ ron có độ chính xác cao hơn ( $RMSE = 2,88$ ,  $R^2 = 0,987$  và  $P = 0$ ) so với phương pháp dự đoán bằng lý thuyết xám ( $RMSE = 7,84$ ,  $R^2 = 0,879$  và  $P = 0$ ).

Trích dẫn: Nguyễn Hiền Thân, Chế Đình Lý và Phạm Văn Tất, 2017. So sánh khả năng dự đoán chất lượng nước sông Đồng Nai bằng mạng nơ ron nhân tạo và lý thuyết xám. Tạp chí Khoa học Trường Đại học Cần Thơ. Số chuyên đề: Môi trường và Biến đổi khí hậu (1): 120-127.

## 1 GIỚI THIỆU

Quản lý chất lượng nước là một trong những phương diện quan trọng của quản lý môi trường. Nhu cầu nước ngày một gia tăng cùng với tăng dân số và hoạt động của con người dẫn đến chất lượng nước ngày càng xấu đi. Dự đoán sự thay đổi chất lượng nước là một thách thức mà các nhà quản lý đang phải đối mặt (Diamantopoulou *et al.*, 2005). Đánh giá và dự đoán chất lượng nước là một trong những nhiệm vụ quan trọng của quản lý môi trường. Người quản lý không những dễ dàng hoạch định và đưa ra quyết định bảo vệ môi trường mà còn cung cấp nhiều thông tin chất lượng môi trường cho cộng đồng (Juan D *et al.*, 2012).

Phương pháp mạng trí tuệ nhân tạo (Artificial Neuron Network - ANN) là công cụ đánh giá mới trong thời kỳ phát triển của công nghệ thông tin, có thể kết hợp thuận lợi giữa phương pháp truyền thống và trí tuệ nhân tạo (Zulin & Benlin, 2012). Khái niệm nơ ron nhân tạo được giới thiệu lần đầu tiên vào năm 1943 (Sarani *et al.*, 2012) và được xem như một công cụ tiềm năng hữu ích cho hệ thống phi tuyến (Barzegar *et al.*, 2016; Patki *et al.*, 2013). Nghiên cứu mạng trí tuệ nhân tạo được ứng dụng thành công trong nhiều lĩnh vực từ thập niên 80 - 90, đặc biệt là lĩnh vực năng lượng máy tính. Trải qua các thập kỷ, mạng trí tuệ nhân tạo cải tiến không chỉ là một công cụ nghiên cứu mà còn ứng dụng để giải nhiều vấn đề trong thế giới thực. Kết quả là ANN ngày càng ứng dụng trong nhiều lĩnh vực, trong đó có quản lý tài nguyên nước (Lobbrecht *et al.*, 2002).

Trong những năm gần đây, nhiều nghiên cứu ứng dụng ANN để dự đoán và đánh giá chất lượng nước (Barzegar *et al.*, 2016). Stewart (2002) sử dụng ANN để dự đoán hàm lượng oxy hòa tan trong nước sông. Các biến tham gia vào mô hình dự đoán bao gồm lưu lượng chảy, bức xạ mặt trời, nhiệt độ và lượng mưa. Trong nghiên cứu này, tác giả sử dụng mạng lan truyền ngược để thực hiện huấn luyện mạng. Alireza và Jaber (2013) đã ứng dụng ANN ba lớp để dự đoán nồng độ đồng trong nước uống. Bốn biến đầu vào độ tinh điện, chất rắn lơ lửng, nhiệt độ và pH được sử dụng. Kết quả đánh giá cho thấy nồng độ đồng dự đoán có độ chính xác cao so với nồng độ đồng thực tế. Việc ứng dụng ANN để dự đoán chất lượng nước đối với các thông số đơn lẻ được nhiều nghiên cứu quan tâm, nhưng việc nghiên cứu dự đoán chỉ chất lượng nước sông từ các thông số đo đạc nhanh ngoài hiện trường và số liệu khí tượng ít được đề cập đến.

Một phương pháp khác được phát triển sau ANN là lý thuyết xám. Lý thuyết xám được phát triển bởi Deng (1982), nó thật sự là lý thuyết di truyền và đa ngành để giải quyết các hệ thống có đặc điểm thông tin ít và thiếu (Hsu & Chen, 2003; Ali & Zade, 2011). Khái niệm “Grey” được hiểu là ít (nghèo), không hoàn toàn và không chắc chắn (Slavek & Jović, 2012). Các lĩnh vực ứng dụng lý thuyết xám bao gồm: phân tích hệ thống, xử lý số liệu, mô hình, dự đoán và kiểm soát. Điểm mạnh của lý thuyết xám là phân tích hệ thống. Mô hình dự đoán được sử dụng trong nhiều lĩnh vực (Dounis *et al.*, 2006). Thuận lợi của lý thuyết hệ thống xám là thực hiện nghiên cứu không chắc chắn. Đặc biệt, nó có thể sử dụng với số lượng mẫu ít hoặc không chắc chắn dữ liệu nào là đại diện (Slavek & Jović, 2012). Hiện nay, lý thuyết xám được ứng dụng nhiều trong dự đoán các vấn đề môi trường như: Dự đoán nhiệt độ (Dounis *et al.*, 2006), dự đoán lưu lượng dòng chảy (Vishnu & Syamala, 2012), dự đoán chất lượng môi trường (Haiyan, 2002), dự đoán bụi lơ lửng (Pai *et al.*, 2011) và dự đoán lượng mưa (Lin *et al.*, 2013).

Trong bài báo này, khả năng dự đoán của hai mô hình ANN và lý thuyết xám được so sánh nhằm đánh giá khả năng dự đoán của hai mô hình để ứng dụng vào công tác dự đoán chất lượng nước. Kết quả nghiên cứu sẽ góp phần ứng dụng các phương pháp dự đoán chất lượng nước mới vào lĩnh vực môi trường. Đồng thời, kết quả nghiên cứu sẽ mở ra hướng ứng dụng các số liệu khí tượng hiện có và thông số chất lượng nước đo đạc nhanh để dự đoán chất lượng nước.

## 2 PHƯƠNG PHÁP NGHIÊN CỨU

### 2.1 Dữ liệu

Dữ liệu quan trắc được thu thập từ năm 2010 đến năm 2014 gồm 23 điểm trên sông Đồng Nai thuộc hai tỉnh Bình Dương và Đồng Nai từ Trung Tâm quan trắc Môi trường. Trong 23 điểm quan trắc có 4 điểm thuộc địa phận tỉnh Bình Dương bao gồm: ĐN1, ĐN2, ĐN3, ĐN4 và tỉnh Đồng Nai bao gồm 19 điểm: SW-DN-01, SW-DN-02, SW-DN-03, SW-DN-04, SW-DN-05, SW-DN-06, SW-DN-07, SW-DN-08, SW-DN-09, SW-DN-10, SW-DN-11, SW-DN-12, SW-DN-13, SW-DN-14, SW-DN-15, SW-DN-16, SW-DN-17, SW-DN-18 và SW-DN-19.

Dữ liệu sau khi thu thập được kiểm tra trị bất thường (outliers) và dữ liệu khuyết (missing data). Đối với dữ liệu bất thường được loại bỏ được thêm vào thông qua phương pháp phân tích dữ liệu nhớt (input missing data) bằng SPSS 18.0.

## 2.2 Phương pháp nghiên cứu

### 2.2.1 Mạng nơ ron perceptron nhiều lớp (Multi-layer perceptron neural network – MLPNN)

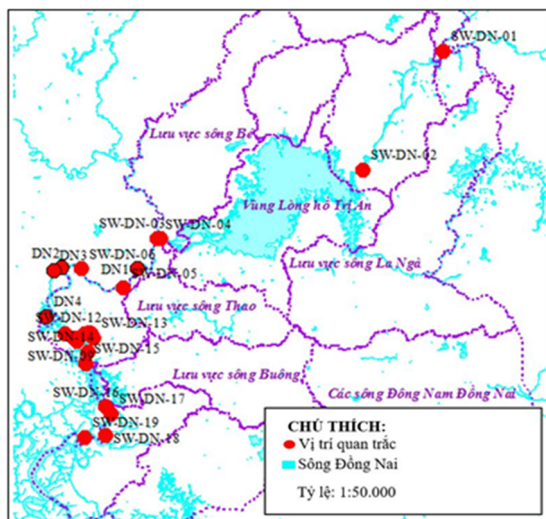
Mạng nơ ron MLPNN gồm một lớp input và một lớp output với các lớp ẩn giữa mỗi lớp chứa một nơ ron nhân tạo. Một nơ ron nhân tạo trong một loại cấu trúc mạng nơ ron nhân tạo nhận một tập hợp tín hiệu đầu vào ( $x$ ) với trọng số ( $w$ ), tính toán trung bình trọng số của chúng ( $y$ ), sử dụng hàm tổng và sau đó sử dụng hàm chuyển đổi để tạo

giá trị output ( $o = f(y)$ ), trong đó:  $y = \sum_{i=1}^n w_i x_i$

(1) (Mekanik *et al.*, 2012).

Bước đầu tiên của quá trình huấn luyện, dữ liệu được chuẩn hóa theo phương pháp min-max. Cả hai biến đầu vào và đầu ra cho mô hình được chuẩn hóa để tránh lỗi trong tính toán (Thair *et al.*, 2014) vì các thông số đều không đồng nhất đơn vị tính toán theo công thức sau:

$$X_n = (X - X_{\min}) / (X_{\max} - X_{\min}) \quad (2)$$



Hình 1: Vị trí quan trắc sông Đồng Nai

Trong đó,  $X_n$  và  $X$  thể hiện giá trị chuẩn hóa và dữ liệu gốc,  $X_{\max}$  và  $X_{\min}$  là giá trị lớn nhất và nhỏ nhất của từng thông số trong tập dữ liệu.

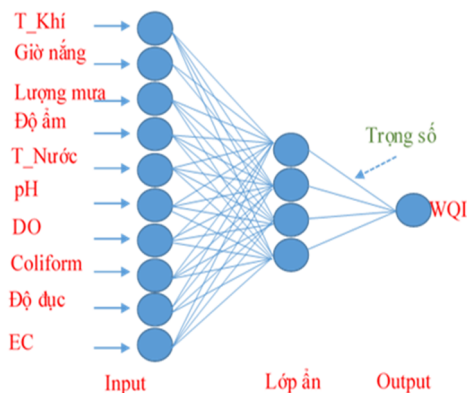
Biến dự đoán sẽ được xác định thông qua phương pháp phân tích nhóm và phân tích thành phần chính. Biến đầu vào là thông số chất lượng nước có thể thu thập nhanh tại hiện trường và dữ liệu khí tượng trong niên giám thống kê nhằm giảm chi phí và thời gian tính toán nhưng vẫn cho kết quả đánh giá nhanh. Các biến được lựa chọn từ kết quả phân tích hệ số tương quan, phân tích thành phần chính và phân tích gộp nhóm. Các biến lựa chọn là các biến có mối tương quan cao với chỉ số

chất lượng nước. Biến đầu ra là chỉ số chất lượng nước (WQI) được tính toán theo quyết định số 879/2011 của Tổng cục Môi trường (Tổng cục Môi trường, 2011). Các thông số tính toán bao gồm 09 thông số: BOD<sub>5</sub>, COD, N-NH<sub>3</sub>, Độ đục, TSS, Coliform, DO% bão hòa, pH và nhiệt độ nước. Công thức tính toán chỉ số chất lượng nước cụ thể như sau:

$$WQI = \frac{WQI_{pH}}{100} \left[ \frac{1}{4} \sum_{a=1}^4 WQI_a \times \frac{1}{2} \sum_{b=1}^2 WQI_b \times WQI_c \right]^{1/3} \quad (3)$$

Trong đó,  $WQI_a$ : Giá trị WQI đã tính toán đối với 04 thông số: DO, BOD<sub>5</sub>, COD, N-NH<sub>4</sub>;  $WQI_b$ : Giá trị WQI đã tính toán đối với 02 thông số: TSS, độ đục;  $WQI_c$ : Giá trị WQI đã tính toán đối với thông số Tổng Coliform;  $WQI_{pH}$ : Giá trị WQI đã tính toán đối với thông số pH. Ngưỡng giá trị của chỉ số WQI từ 0-100.

Tập dữ liệu tham gia dự đoán chất lượng nước từ năm 2010 – 2013 dùng để xây dựng và kiểm tra mô hình huấn luyện và dữ liệu 2014 dùng để đánh giá mô hình. Mỗi điểm quan trắc được thu thập 1 tháng/lần, và lấy 3 mẫu tại mỗi điểm: mẫu giữa dòng, mẫu hai bên bờ trái và phải. Theo nhiều nghiên cứu, tập dữ liệu kiểm tra mô hình (test) nên xấp xỉ 10 - 40% (Palani *et al.*, 2008). Do đó, tác giả thiết lập tỷ lệ dữ liệu huấn luyện và kiểm tra mô hình là 70:30, trong tổng 2.276 mẫu tương ứng 1.593 mẫu huấn luyện và 683 mẫu kiểm tra mô hình.



Hình 2: Cấu trúc mạng MLPNN

### 2.2.2 Phân tích nhóm (cluster analysis)

Phân tích nhóm là một trong những kỹ thuật thống kê đa biến với mục đích chính phân chia các nhóm riêng biệt hoặc các nhóm theo các thông số nhóm có sự tương đồng lớn nhất và các nhóm có sự tương đồng ít (Sayadi *et al.*, 2014). Phương pháp phân bậc dùng để lựa chọn biến là phương pháp liên kết đơn giản (single linkage method),

khoảng cách hoặc sự tương tự giữa hai nhóm A và B được xác định như khoảng cách nhỏ nhất giữa điểm A và điểm B (khoảng cách Euclid):

$$D(A,B)=\min\{d(x_i,x_j), x_i \text{ trong } A, x_j \text{ trong } B\} \quad (4)$$

Trong đó,  $d(x_i, x_j)$  là khoảng cách Euclid (là khoảng cách thường giữa hai điểm có thể đo được bằng cây thước, được tính bằng công thức Pytago) (Alkarkhi *et al.*, 2009). Ở mỗi bước khoảng cách là được tìm thấy cho mỗi cặp nhóm và hai nhóm với khoảng cách nhỏ nhất (sự tương đồng lớn nhất) được gộp lại. Sau 2 nhóm được gộp thì quá trình được lặp lại cho bước tiếp theo: Khoảng cách giữa các cặp được tính lại lần nữa và cặp có khoảng cách nhỏ nhất được gộp thành nhóm đơn lẻ. Kết quả được thể hiện trong sơ đồ nhánh (dendrogram). Phương pháp này kết hợp với phân tích thành phần chính được sử dụng để xác định các biến dự đoán cho mô hình dự đoán chất lượng nước.

### 2.2.3 Phương pháp kiểm định thống kê

Mô hình dự đoán được đánh giá bằng cách so sánh giá trị WQI ước tính thu được so với giá trị thực tế. Tác giả sử dụng bình phương sai số gốc (Root Mean Squared Error - RMSE) để đánh giá sai số mô hình. Mô hình có sai số càng nhỏ, thì mô hình đó càng tốt.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (WQI_{act} - WQI_{pre})_i^2}{n}} \quad (5)$$

Trong đó,  $WQI_{act}$  là giá trị thực của chỉ số chất lượng nước,  $WQI_{pre}$  là giá trị dự đoán của chỉ số chất lượng nước,  $n$  là số lượng mẫu dữ liệu giám sát (Banejad & Olyae, 2011).

### 2.2.4 Phương pháp dự đoán lý thuyết xám

Phương pháp dự đoán lý thuyết xám là phương pháp dự đoán nội suy dựa trên diễn biến tập dữ liệu chất lượng nước. Qui trình tính toán được khái quát qua ba bước cơ bản: Tính tổng lũy tiến (accumulated generating operator – AGO), Tích lũy toán tử chuyển vị (inverse accumulating operator – IAGO) và Mô hình xám (GM) (Kazemi *et al.*, 2011), cụ thể như sau:

Bước 1: Dữ liệu gốc sử dụng ước lượng trong mô hình như sau:

$$\{X^{(0)}\} = \{X^{(0)}, X^{(0)}, X^{(0)}, \dots, X^{(0)}\} \quad (m \geq 4) \quad (6)$$

Trong đó,  $x$  là số năm dữ liệu thu thập, số lượng phải lớn hơn bằng 4 năm.

Dữ liệu dự đoán chất lượng nước là giá trị trung bình chỉ số WQI các điểm quan trắc trên sông Đồng Nai từ năm 2010 – 2013 ( $\geq 4$  năm). Để tiện theo dõi, dữ liệu điểm ĐN1 được sử dụng để minh họa cho kết quả tính toán. Dữ liệu chỉ số WQI điểm ĐN1 từ 2010-2013 tương ứng là  $\{83,4; 80,0; 86,6; 89,9\}$ . Dữ liệu môi trường thường có biến động lớn, do đó để có kết quả dự đoán chính xác cần làm trơn dữ liệu như sau:

$$X_{(1)}^{(0)'} = \frac{3}{4}X_{(1)}^{(0)} + \frac{1}{4}X_{(2)}^{(0)} \quad (7)$$

$$X_{(i)}^{(0)'} = (X_{(i-1)}^{(0)} + 2X_{(i)}^{(0)} + X_{(i+1)}^{(0)})/4 \quad (i = 2, 3, \dots, m-1) \quad (8)$$

$$X_{(m)}^{(0)'} = \frac{1}{4}X_{(m-1)}^{(0)} + \frac{3}{4}X_{(m)}^{(0)} \quad (\text{Haiyan, 2002}) \quad (9)$$

Trong đó,  $X_{(1)}^{(0)'}$  là dữ liệu năm đầu tiên của dãy dữ liệu,  $X_{(i)}^{(0)'}$  là dữ liệu của các năm tiếp theo,  $X_{(m)}^{(0)'}$  là dữ liệu năm cuối cùng.

Dựa theo dữ liệu điểm sông ĐN1 và công thức 7, 8, 9 ta thu được dữ liệu được làm trơn ĐN1  $\{82,6; 82,5; 85,8; 89,1\}$ . Tiếp theo tích tổng lũy tiến dãy số liệu tương ứng:  $\{X_{(i)}^{(1)}\} = \{X_{(1)}^{(1)}, X_{(2)}^{(1)}, X_{(3)}^{(1)}, \dots, X_{(m)}^{(1)}\}$ .

Trong đó:

$X_{(i)}^{(1)} = \sum_{k=1}^i X_{(k)}^{(0)'}$ . Rõ ràng  $X_{(i)}^{(0)}$  có thể dễ dàng phục hồi từ  $X_{(i)}^{(1)}$  như sau:

$X_{(i)}^{(0)} = X_{(i)}^{(1)} - X_{(i-1)}^{(1)}$ , trong đó:  $X_{(1)}^{(0)} = X_{(1)}^{(1)}$ . Bước này được gọi là IAGO (Kazemi *et al.*, 2011)

Kết quả tính tổng lũy tiến dữ liệu điểm ĐN1 =  $\{82,6; 165,1; 250,8; 339,9\}$

Bước 2: Hình thành mô hình xám bằng cách thiết lập phương trình vi phân xám:

$$\frac{dx^1}{dt} + ax^1 = b \quad \text{Hay} \quad X_{(i)}^{(0)} + az_{(i)}^{(1)} = b \quad (10)$$

Trong đó:  $Z_{(k)}^{(1)} = 0,5 X_{(i)}^{(1)} + (1-\alpha) X_{(i-1)}^{(1)}$ , ( $i=2, 3, 4, \dots, n$ )

$a$  được gọi là hệ số phát triển và  $b$  được gọi là hệ số động lực (hay đầu vào xám) (Kazemi *et al.*, 2011; Pai *et al.*, 2011)

$$X_{(2)}^{(0)} + az_{(2)}^{(1)} = b$$

$$X_{(3)}^{(0)} + az_{(3)}^{(1)} = b$$

$$\vdots \quad \quad \quad \vdots$$

$$X_{(n)}^{(0)} + az_{(n)}^{(1)} = b$$

(11) Chuyển đổi công thức (11) vào ma trận ta



được: 
$$\begin{bmatrix} X_{(2)}^{(0)} \\ X_{(3)}^{(0)} \\ \vdots \\ X_{(n)}^{(0)} \end{bmatrix} = \begin{bmatrix} -Z_{(2)}^{(1)} & 1 \\ -Z_{(3)}^{(1)} & 1 \\ \vdots & \vdots \\ -Z_{(n)}^{(1)} & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix}$$
 Sau đó, hệ số có thể

được ước lượng bằng ma trận,  $p = \begin{bmatrix} a \\ b \end{bmatrix} = (B^T B)^{-1} B^T Y$  (12)

Trong đó: Y là vector dữ liệu  $Y = \begin{bmatrix} X_{(2)}^{(0)} \\ X_{(3)}^{(0)} \\ \vdots \\ X_{(n)}^{(0)} \end{bmatrix}$ , B

được gọi là ma trận dữ liệu  $B = \begin{bmatrix} X_{(2)}^{(0)} \\ X_{(3)}^{(0)} \\ \vdots \\ X_{(n)}^{(0)} \end{bmatrix}$ , T là ký hiệu

của ma trận chuyển vị. Ta có:  $Y = \begin{bmatrix} 82,5 \\ 85,8 \\ 89,1 \end{bmatrix}$ ,

$B = \begin{bmatrix} -123,8 & 1 \\ -207,9 & 1 \\ -295,4 & 1 \end{bmatrix}$ ,  $p = \begin{bmatrix} a \\ b \end{bmatrix} = (B^T B)^{-1} B^T Y = \begin{bmatrix} -0,04 \\ 77,78 \end{bmatrix}$

Theo công thức (6),  $X^{(1)}$  ở thời gian t:

Thông số	Thành phần (pr)				
	1	2	3	4	5
Nhiệt độ không khí TB tháng	-0,19	0,01	<b>0,84</b>	0,02	0,03
Số giờ nắng TB tháng	<b>-0,68</b>	-0,05	0,54	0,15	0,06
Lượng mưa TB tháng	<b>0,82</b>	0,04	-0,02	-0,18	-0,12
Độ ẩm TB tháng	<b>0,82</b>	0,05	-0,25	-0,17	-0,07
Nhiệt độ nước	-0,14	0,22	0,36	0,09	<b>-0,52</b>
pH	-0,08	0,06	0,12	0,01	<b>0,79</b>
DO	0,26	<b>-0,60</b>	0,07	-0,39	0,19
COD	0,37	0,45	0,02	<b>0,50</b>	-0,06
BOD <sub>5</sub>	0,43	<b>0,59</b>	0,10	0,23	0,00
TSS	<b>0,63</b>	-0,09	-0,01	0,39	0,16
NH <sub>4</sub> <sup>+</sup>	0,08	<b>0,70</b>	0,00	-0,26	0,13
Độ đục	<b>0,80</b>	-0,06	-0,08	0,22	0,11
Coliform	-0,03	0,17	-0,28	-0,04	0,19
EC	-0,09	-0,01	-0,01	<b>0,79</b>	-0,03
NO <sub>3</sub> <sup>-</sup>	0,22	0,07	<b>-0,65</b>	0,45	-0,07
NO <sub>2</sub> <sup>-</sup>	-0,23	<b>0,63</b>	-0,38	0,07	-0,10

Hình 3: Hệ số thành phần chính

Theo Hình 3 và Hình 4, nhóm thứ 1 có 2 nhóm phụ là DO và nhóm nhiệt độ không khí trung bình tháng (T\_Air), nhiệt độ nước (T\_Water), độ ẩm (humidity), số giờ nắng (shunshine). Với mục đích nghiên cứu là khảo sát các biến dự đoán dễ thu thập tại hiện trường và dữ liệu thống kê có sẵn, do đó các biến trong nhóm thứ 1 được lựa chọn.

$$\hat{X}^{(1)}(t+1) = (X^{(0)}(1) - \frac{b}{a})e^{-at} + \frac{b}{a}, (t = 1, 2, 3, \dots), \hat{X}^{(1)}(1) = X_{(1)}^{(1)} \quad (13)$$

Trong đó, dấu “^” biểu thị cho giá trị dự đoán Grey và  $\hat{X}^{(1)}(t+1)$  là giá trị ước đoán của  $X^{(1)}(t+1)$ . Vì vậy, dữ liệu gốc có thể tính theo công thức sau:

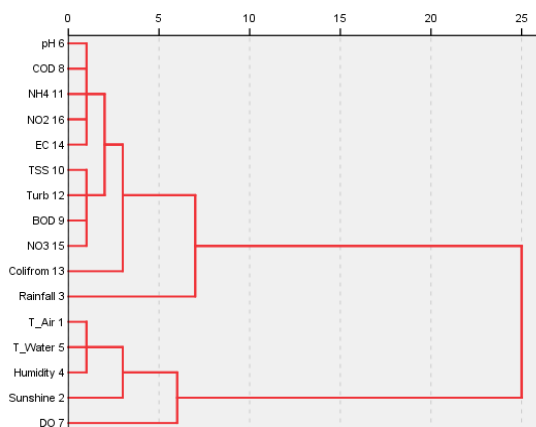
$$\hat{X}^{(0)}(t) = \hat{X}^{(1)}(t+1) - \hat{X}^{(1)}(t) = (X^{(0)}(1) - \frac{b}{a})(1-e^{-a})e^{-a(t-1)}, \hat{X}^{(0)}(1) = X^{(0)}(t=2, 3, 4) \quad (14)$$

Ta có thời gian t là 4 năm, thế các giá trị vào công thức ta thu được  $\hat{X}^{(1)} = \{82,6; 165,1; 250,8; 339,9; 432,4\}$ . Giá trị WQI cho điểm ĐN1 năm 2014 là 92,5. Tương tự, tác giả tiến hành tính toán cho các giá trị WQI dự đoán của các điểm khác.

### 3 KẾT QUẢ VÀ THẢO LUẬN

#### 3.1 Lựa chọn biến cho mô hình mạng nơ ron MLP

Các biến đầu vào ban đầu gồm 16 biến có liên hệ với chỉ số chất lượng nước được xác định thông qua phân tích hệ số tương quan pearson. Trong đó, có 4 biến dữ liệu khí tượng thủy văn và 12 biến là thông số quan trắc chất lượng nước. Sau đó, dữ liệu được xác định thành phần đại diện cho tập dữ liệu. Kết quả phân tích có 16 thành phần đại diện cho dữ liệu chất lượng nước trong đó có 5 thành phần có giá trị riêng liên hệ cao > 1. 5 thành phần này giải thích 61,92% phương sai giải thích của tập dữ liệu gốc. Sau đó, tiến hành xoay trục của các thành phần chính nhằm thu được hệ số thành phần tốt nhất.



Hình 4: Phân nhóm chất lượng nước

Trong nhóm hai bao gồm 2 nhóm phụ: nhóm thứ nhất chỉ có lượng mưa trung bình tháng và nhóm thứ hai được chia làm 3 nhóm nhỏ gồm: nhóm thứ nhất có coliform, nhóm thứ 2 gồm TSS, độ đục (Turb), BOD<sub>5</sub>, NO<sub>3</sub><sup>-</sup>, NO<sub>2</sub><sup>-</sup> và nhóm thứ 3 có độ tinh điện (EC), NH<sub>4</sub><sup>+</sup>, COD và pH. Căn cứ

vào sự tương đồng của các nhóm, mục đích nghiên cứu và hệ số thành phần chính (Hình 3) ta có:

– Lượng mưa trung bình tháng (rainfall) là nhóm phụ riêng lẻ do đó được lựa chọn

– Nhóm thứ 2 gồm TSS, độ đục, BOD<sub>5</sub>, NO<sub>3</sub><sup>-</sup>, NO<sub>2</sub><sup>-</sup> có hệ số thành phần chính tương ứng là 0,63 (pr1), 0,80 (pr1), 0,59 (pr2), -0,65 (pr3), 0,63 (pr2). Thông số độ đục là thông số có hệ số nhân tố lớn nhất. Do đó, độ đục là thông số đại diện tốt nhất cho nhóm.

Nhóm thứ 3 có độ tinh điện (EC), NH<sub>4</sub><sup>+</sup>, COD và pH có hệ số thành phần tương ứng là 0,79 (pr4), 0,7 (pr2), 0,5 (pr4) và 0,79 (pr5). Độ tinh điện (EC) và pH là thông số có hệ số cao nhất trong nhóm và thỏa mãn mục tiêu lựa chọn biến, vì vậy đây là hai thông số tốt nhất đại diện cho nhóm.

Như vậy, các biến được lựa chọn từ phân tích gộp nhóm gồm 10 thông số:  $x_1$  = Nhiệt độ không khí trung bình,  $x_2$  = Số giờ nắng trung bình,  $x_3$  =

Lượng mưa trung bình tháng,  $x_4$  = Độ ẩm,  $x_5$  = Nhiệt độ nước,  $x_6$  = pH,  $x_7$  = DO,  $x_8$  = Độ đục,  $x_9$  = Coliform và  $x_{10}$  = Độ dẫn điện là các biến độc lập (input) và WQI (Y) là biến phụ thuộc cho các mô hình mạng ANN (output). Trong tập hợp các biến dự đoán, chỉ duy nhất thông số coliform là thông số được xác định ở phòng thí nghiệm, ngược lại các thông số còn lại là các thông số có thể đo đạc ngay tại sông Đồng Nai. Dữ liệu khí tượng sử dụng trong nghiên cứu có thể thu thập dễ dàng và được cập nhật liên tục trong sách Niên giám Thống kê tỉnh Đồng Nai.

## 3.2 Kết quả xây dựng mô hình dự đoán

### 3.2.1 Mô hình mạng nơ ron nhân tạo

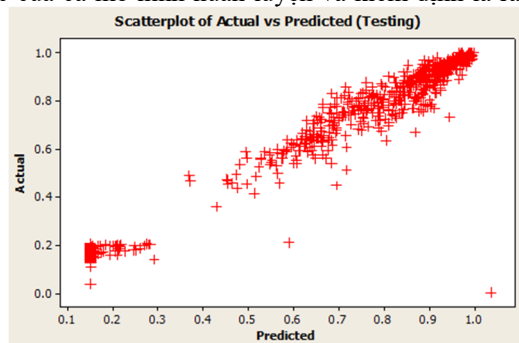
Kết quả chạy mô hình nơ ron MLPNN tìm được loại mạng tối ưu là 4 nơ ẩn và sai số kiểm định mô hình RMSE = 0,0492 thấp nhất. Đây là cấu trúc mạng tối ưu cho mô hình dự đoán chất lượng nước sông Đồng Nai.

**Bảng 1: Tìm mạng nơ ron perceptron nhiều lớp tốt nhất**

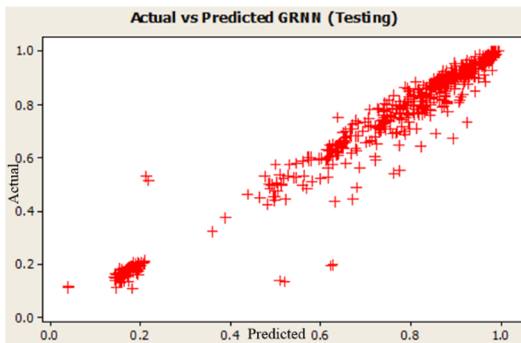
Tim mạng tốt nhất	RMSE kiểm định	Thời gian huấn luyện	Điểm dừng
MLFN 3 nơ	0,0540	0:12:00	Auto-Stopped
MLFN 4 nơ	0,0492	0:12:00	Auto-Stopped
MLFN 5 nơ	0,0531	0:12:00	Auto-Stopped
MLFN 6 nơ	0,0536	0:12:00	Auto-Stopped

Theo Hình 5 và Hình 6, mức độ phân tán các điểm mẫu phân phối theo đường chuẩn, số điểm rời rạc của cả mô hình huấn luyện và kiểm định là rất

ít. Điều này cho thấy, mô hình dự đoán chất lượng nước bằng mạng nơ ron perceptron đa lớp là rất tốt.



**Hình 5: Đồ thị phân tán các biến mô hình huấn luyện**



**Hình 6: Đồ thị phân tán các biến mô hình huấn luyện**

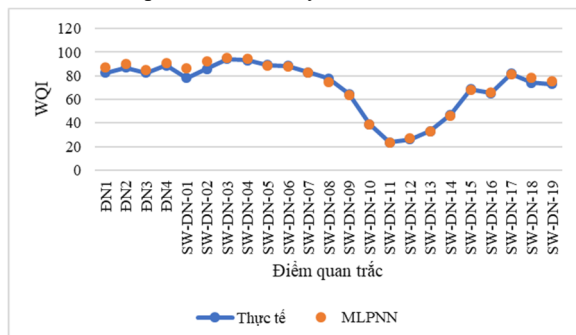
### 3.2.2 So sánh kết quả dự đoán chất lượng nước bằng mạng MLPNN và mô hình lý thuyết xám

Kết dự đoán chất lượng nước sông Đồng Nai năm 2014 bằng mạng nơ ron MLP cho sai số mô hình RSME = 2,8,  $R^2 = 0,987$  và p mức ý nghĩa với giá trị WQI thực tế bằng 0. Điều này cho thấy kết quả dự đoán chất lượng nước bằng mạng MLP cho kết quả rất tốt và có độ chính xác rất cao. Hơn thế nữa, kết quả dự đoán chất lượng nước từ các thông số khí tượng và thông số chất lượng nước dễ thu

thập tại hiện trường là rất tốt. Kết quả này cho thấy sử dụng các thông số đo đạc nhanh tại hiện trường và dữ liệu khí tượng hoàn toàn có thể dự đoán được chất lượng nước và cho kết quả đánh giá là rất tốt. Mô hình dự đoán ANN hoạt động tốt đối với dữ liệu dài và khả năng mô phỏng kết quả có độ chính xác cao.

Đối với mô hình lý thuyết xám, cho kết quả sai số WQI so với giá trị WQI thực tế là khá thấp 7,84,  $R^2 = 0,879$  và p mức ý nghĩa với giá trị WQI thực

tế bằng 0. Điều này cho thấy phương pháp dự đoán chất lượng nước bằng lý thuyết xám là khá tốt. So với phương pháp mạng nơ ron nhân tạo thì phương pháp dự đoán chất lượng nước bằng lý thuyết xám cho kết quả không tốt bằng. Tuy nhiên, mô hình lý thuyết xám có ưu điểm là có thể dự đoán nội suy dựa trên dữ liệu thực tế với chuỗi dữ liệu ít nhưng vẫn cho kết quả có độ tin cậy tốt.

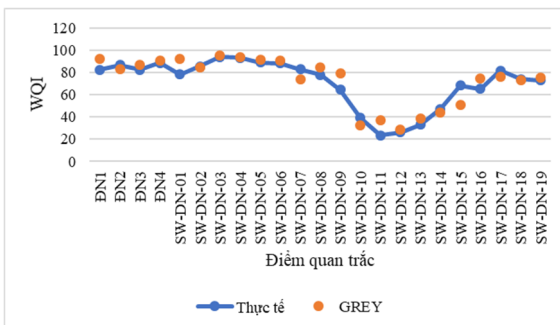


**Hình 7: WQI dự đoán bằng mạng MLPNN và giá trị WQI tính toán thực tế**

Phương pháp dự đoán Grey cho kết quả đánh giá tốt khi dữ liệu dự đoán không có nhiều biến động (Xuepeng *et al.*, 2006). Mức độ chính xác càng cao khi dữ liệu tăng hoặc giảm liên tục qua các năm theo một chiều hướng diễn biến nhất định. Phương pháp dự đoán Grey phù hợp với nguồn dữ liệu ít, trong khi mô hình ANN phù hợp với dữ liệu lớn.

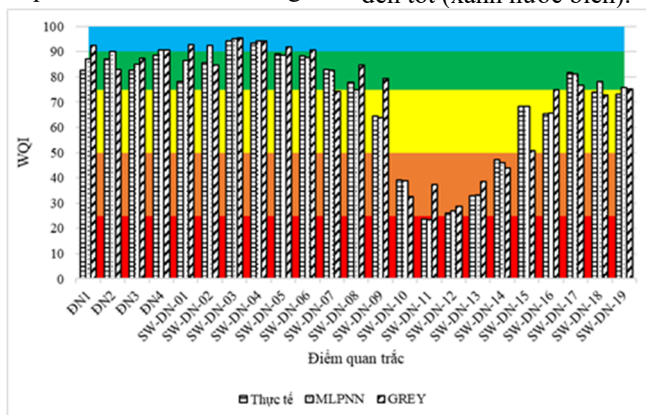
Kết quả chất lượng nước năm 2014 của sông Đồng Nai được tính toán từ 3 phương pháp WQI, dự đoán ANN và mô hình lý thuyết xám được thể hiện trong Hình 9. Kết quả dự đoán chất lượng

Hình 7 và Hình 8 cho thấy mức độ phân tán của chỉ số WQI giữa hai mô hình mạng ANN và lý thuyết xám so với giá trị WQI thực tế. Qua đó, có thể thấy mức độ chính xác của hai mô hình dự đoán chất lượng nước này. Rõ ràng ANN cho kết quả dự đoán chất lượng nước tốt hơn mô hình lý thuyết xám.



**Hình 8: WQI dự đoán bằng mô hình lý thuyết xám và giá trị WQI tính toán thực tế**

nước cho thấy WQI thực tế và ANN không có sự chênh lệch đáng kể về mức độ ô nhiễm của chất lượng nước. Ngược lại một số điểm trong kết quả dự đoán bằng mô hình lý thuyết xám có sự chênh lệch cao hơn như điểm: SW-DN-08, SW-DN-09, SW-DN-11. Kết quả chất lượng nước năm 2014 sông Đồng Nai đã bị ô nhiễm, chỉ có điểm ô nhiễm nặng SW-ĐN-11 (vùng màu đỏ), 4 điểm ô nhiễm nhẹ SW-ĐN-10, SW-ĐN-12, SW-ĐN-13, SW-ĐN-14 (vùng màu cam). Các điểm này đều nằm trong đoạn sông chảy qua thành phố Biên Hòa. Các điểm còn lại có chất lượng nước từ ôn hòa (màu cam) đến tốt (xanh nước biển).



**Hình 9: Chỉ số WQI thực tế, MLPNN và lý thuyết xám năm 2014**

#### 4 KẾT LUẬN

Hai mô hình dự đoán chất lượng nước cho kết quả có độ chính xác sát với giá trị thực tế. Mô hình mạng nơ ron perceptron nhiều lớp thể hiện khả năng dự đoán vượt trội trong điều kiện biến dữ liệu

đầu vào là các biến có thể thu thập nhanh tại hiện trường nhưng vẫn có độ chính xác rất cao và hoạt động tốt đối với tập dữ liệu lớn (nhiều điểm và nhiều năm). Mô hình lý thuyết xám cho kết quả dự đoán cũng khá tốt và phù hợp với điều kiện thống kê hạn chế (dữ ít).

Cả hai mô hình là 2 công cụ dự đoán có thể ứng dụng tốt trong lĩnh vực môi trường. Tùy vào điều kiện dữ liệu hiện có để có thể áp dụng phương pháp dự đoán nào là phù hợp. Trong đó, mô hình ANN là kỹ thuật dự đoán có thể đáp ứng tốt về mặt không gian và thời gian, dữ liệu thống kê lớn và không bắt buộc theo một cấu trúc nhất định.

## TÀI LIỆU THAM KHẢO

- Ali Mohammadi, & Zade, S. Z., 2011. Applying grey forecasting method to forecast the portfolio's rate of return in stock market of Iran. *Australian Journal of Business and Management Research*. 1: 1-16.
- Alireza Shakeri Abdolmaleki, A. G. A., Jaber Soltani, 2013. Artificial neural network (ann) approach for predicting Cu concentration in drinking water of Chahnimeh1 reservoir in Sistan-Balochistan, Iran. *Health Scope*. 2: 31-38.
- Alkarkhi, A. F. M., Ahmad, A., & Easa, A. M., 2009. Assessment of surface water quality of selected estuaries of Malaysia: multivariate statistical techniques. *The Environmentalist*. 29: 255-262.
- Vishnu B and P, Syamala, 2012. Grey model for stream flow prediction. *Aceh International Journal of Science and Technology*. 1:14-19.
- Banejad, H., & Olyaie, E., 2011. Application of an artificial neural network model to rivers water quality indexes prediction – a case study. *Journal of American Science*. 7: 60-65.
- Barzegar, R., Adamowski, J., & Moghaddam, A. A., 2016. Application of wavelet-artificial intelligence hybrid models for water quality prediction: a case study in Aji-Chay River, Iran. *Stochastic Environmental Research and Risk Assessment*. 1-23.
- Diamantopoulou, M. J., Antonopoulos, V. Z., & Papamichail, D. M., 2005. The use of a neural network technique for the prediction of water quality parameters of Axios River in Northern Greece. *European Water*. 11(12): 55-62.
- Dounis, A. I., Tiropanis, P., Tseles, D., Nikolaou, G., & Syrcos, G. P., 2006. A comparison of grey model and fuzzy predictive model for time series. *International Journal of Information and Mathematical Sciences*. 2: 176-181.
- Haiyan, W., 2002. Assessment and prediction of overall environmental quality of Zhuzhou City, Hunan Province, China. *Environmental Management*. 66: 329-340.
- Hsu, C.-C., & Chen, C.-Y., 2003. Applications of improved grey prediction model for power demand forecasting. *Energy Conversion and Management*. 44: 2241-2249.
- Kazemi, A., Modarres, M., & al, e., 2011. A markov chain grey forecasting model: a case study of energy demand of industry sector in Iran. 2011 3rd International Conference on Information and Financial Engineering. 12: 13-18.
- Lin, Y.-H., Chiu, C.-C., Lin, Y.-J., & Lee, P.-C., 2013. Rainfall prediction using innovative grey model with the dynamic index. *Journal of Marine Science and Technology*. 21: 63-75.
- Lobrecht, A. H., Dibike, Y. B., & Solomatine, D. P., 2002. Applications of neural networks and fuzzy logic to integrated water management. 5: 50-58
- Mekanik, F., Lee, T. S., & Imteaz, M. A., 2011. Rainfall modeling using artificial neural network for a mountainous region in West Iran. Paper presented at the Proceedings of the 19th International Congress on Modelling and Simulation, Perth, Australia.
- Pai, T.-Y., Ho, C.-L., & et al., 2011. Using seven types of GM (1, 1) model to forecast hourly particulate matter concentration in Banciao City of Taiwan. *Water, Air, & Soil Pollution*. 217: 25-33.
- Palani, S., Liong, S.-Y., & Tkalich, P., 2008. An ANN application for water quality forecasting. *Marine Pollution Bulletin*. 56: 1586 - 1597.
- Patki, V. K., Shrihari, S., & Manu, B., 2013. Water quality prediction in distribution system using cascade feed forward neural network. *International Journal of Advanced Technology in Civil Engineering*. 2(1): 84-91.
- Sarani, N., Soltani, J., Sarani, S., & Moasheri, A., 2012. Comparison of artificial neural network and multivariate linear regression model to predict sodium adsorption ratio (SAR) (case study: Sistan River, Iran). *International Journal of Biological, Ecological and Environmental Sciences*. 1: 2277 – 4394.
- Sayadi, M., Rezaei, A., Rezaei, M., & Nourozi, K., 2014. Multivariate statistical analysis of surface water chemistry: A case study of Gharasoo River, Iran Paper presented at the Proceedings of the International Academy of Ecology and Environmental Sciences.
- Slavek, N., & Jovi'c, A., 2012. Application of grey system theory to software projects ranking. 53:284-293.
- Thair S.K, Abdul Hameed M. J, & Ayad S. M., 2014. Prediction of water quality of Euphrates River by using artificial neural network model (spatial and temporal study). *International Research Journal of Natural Sciences*. 2(3): 25-38.
- Tổng cục Môi trường, 2011. Quyết định số 879/QĐ-TCMT về việc ban hành sổ tay hướng dẫn tính toán chỉ số chất lượng nước. Hà Nội.
- Xuepeng, J., Zhisheng, X., Jianming, W., & Xiangbing1, L., 2006. Fire prediction based on reverse sequence grey model. Paper presented at the 2006 International Symposium on Safety Science and Technology, Beijing.
- Zulin, H., & Benlin, D., 2012. Water quality health assessment of offshore area in North Branch of the Yangtze River Estuary. *Dyna*. 79: 50-58.