

Bài báo khoa học

Nghiên cứu kết hợp mô hình thủy lực và mô hình trí tuệ nhân tạo mô phỏng chất lượng nước sông Nhuệ – Đáy

Hoàng Thị An^{1*}, Trần Thùy Nhung¹, Nguyễn Thanh Thủy², Trần Anh Phương¹

¹ Phòng Công nghệ và Kỹ thuật tài nguyên nước, Viện Khoa học tài nguyên nước;
anht510@wru.vn; tranthuynhung1990@gmail.com; phuongtran.monre@gmail.com

² Trường Đại học Thủy Lợi; thanhthuy_rt@tlu.edu.vn

*Tác giả liên hệ: anht510@wru.vn; Tel.: +84–963952484

Ban Biên tập nhận bài: 11/6/2022; Ngày phản biện xong: 18/07/2022; Ngày đăng bài: 25/07/2022

Tóm tắt: Lưu vực sông Nhuệ–Đáy được biết đến như một điểm nóng về chất lượng nước, gây ảnh hưởng lớn đến sức khỏe người dân cũng như cảnh quan, môi trường lưu vực sông trong những năm gần đây. Nghiên cứu xây dựng mô hình trí tuệ nhân tạo (AI) mô phỏng chất lượng nước sông Nhuệ–Đáy dựa trên kết quả mô phỏng thủy động lực và chất lượng nước của mô hình thủy lực MIKE 11. Dữ liệu đầu vào bao gồm: địa hình, số liệu thủy văn, dữ liệu nguồn thải tại các vị trí quan trắc để mô phỏng chất lượng nước trong mô hình MIKE 11. Mô hình AI sử dụng thuật toán MLP–ANN xây dựng quan hệ hồi qui giữa lưu lượng nước sông và nồng độ các nguồn thải với các chỉ số chất lượng nước để dự báo chất lượng nước tại các vị trí kiểm soát theo các kịch bản kiểm soát và quản lý nguồn thải khác nhau một cách nhanh chóng. Kết quả từ mô hình AI gần sát với kết quả từ mô hình MIKE 11 với chỉ số R^2 đạt từ 0,98 trở lên với thời gian mô phỏng nhanh hơn hàng nghìn lần so với mô hình thủy lực, là một công cụ hữu hiệu, cho phép đưa ra dự báo chất lượng nước nhanh. Nghiên cứu cung cấp công cụ quản lý cho các nhà quản lý trong đánh giá nhanh tác động của các giải pháp qui hoạch, quản lý và kiểm soát ô nhiễm nguồn nước trên lưu vực sông Nhuệ–Đáy.

Từ khóa: Chất lượng nước; Lưu vực sông Nhuệ–Đáy; Mô hình MIKE 11; Mô hình trí tuệ nhân tạo AI; MLP–ANN.

1. Đặt vấn đề

Chất lượng nước luôn là vấn đề cấp thiết được quan tâm hơn cả do tác động của nó đến sức khỏe của người dân cũng như đến môi trường và hệ sinh thái. Lưu vực sông Nhuệ–Đáy có vai trò quan trọng không chỉ trong phát triển kinh tế của các tỉnh thành phố ven sông mà còn cả trong sự phát triển của vùng đồng bằng sông Hồng [1]. Tuy nhiên, nguồn nước sông đang bị ô nhiễm nghiêm trọng và là một trong những điểm nóng về ô nhiễm môi trường nước. Theo Tổng cục Môi trường, Bộ Tài nguyên và Môi trường, trong tổng số 15/185 điểm quan trắc bị ô nhiễm nặng nhất trên 5 lưu vực sông khu vực phía Bắc vào năm 2020 thì có đến 13 điểm thuộc lưu vực sông Nhuệ–Đáy và có tới 62% số điểm quan trắc cho kết quả của chỉ số chất lượng nước WQI < 50 ở mức xấu đến rất xấu. Chất lượng nước kém đã ảnh hưởng rất lớn tới cuộc sống, kinh tế, môi trường và xã hội. Do đó việc kiểm soát ô nhiễm trên lưu vực sông Nhuệ–Đáy thuộc địa bàn thành phố Hà Nội đóng vai trò quan trọng không chỉ với Hà Nội mà còn đối với toàn lưu vực.

Việc tính toán mô phỏng biến đổi theo không gian và thời gian của các chất ô nhiễm đóng vai trò quan trọng trong công tác đánh giá và kiểm soát chất lượng nước. Với sự phát triển của công nghệ thông tin cũng như mô hình hóa chất lượng nước, ngày càng nhiều các

mô hình mô phỏng chất lượng nước ra đời và phát triển. Một số mô hình toán thông dụng đã được phát triển bao gồm QUAL, WASP, QUASAR, MIKE Ecolab, các mô hình này thuộc nhóm mô hình số trị, là mô hình hóa quá trình lan truyền nước trên hệ thống sông, kênh mương. Ưu điểm của nhóm mô hình này là có thể cung cấp thông tin đánh giá chất lượng nước theo không gian và thời gian. Tuy nhiên, việc thiết lập mô hình tương đối phức tạp và mất nhiều thời gian, yêu cầu người sử dụng mô hình cần có kiến thức chuyên môn hiệu chỉnh, kiểm định và đánh giá kết quả mô hình. Đặc biệt, trong điều kiện ở các nước đang phát triển như Việt Nam, do tính phức tạp trong việc thiết lập mô hình, sẽ rất khó khăn trong việc chuyển giao mô hình cho các cơ quan quản lý địa phương sử dụng. Do đó việc phát triển các mô hình đơn giản dễ tiếp cận, dễ sử dụng là một yêu cầu cấp bách. Thời gian gần đây, các mô hình trí tuệ nhân tạo phục vụ mô phỏng dự báo chất lượng nước đã và đang được ứng dụng rộng rãi do khả năng tính toán nhanh với độ tin cậy và hiệu quả cao. Bên cạnh đó, các quá trình thủy văn nhìn chung là các quá trình phi tuyến tính trong tự nhiên, thay đổi theo không gian và thời gian như dòng chảy, chất lượng nước [2]. Do đó, việc mô tả các quá trình và yếu tố trên cần được phân tích phi tuyến tính. Khi các dữ liệu đủ lớn, cùng với thuật toán và hệ số của mô hình trí tuệ nhân tạo (AI) được thiết lập một cách phù hợp, mô hình có khả năng mô phỏng các yếu tố với hiệu suất và độ chính xác cao, có thể thay thế cho mô hình thủy lực số trị.

Thực tế cho thấy AI có ưu điểm vượt bậc và phù hợp trong việc quản lý, đánh giá và mô phỏng, dự báo chất lượng nước hiệu quả tại một số quốc gia trên thế giới. Tuy vậy, cách tiếp cận sử dụng công nghệ trí tuệ nhân tạo mô phỏng chất lượng nước vẫn là một trong những hướng nghiên cứu tương đối mới trên thế giới hiện nay. Ở Việt Nam, cho đến thời điểm hiện tại vẫn chưa có nhiều nghiên cứu đánh giá mô phỏng chất lượng nước ứng dụng mạng trí tuệ nhân tạo. Một trong những lý do mô hình trí tuệ nhân tạo chưa được sử dụng rộng rãi là do thiếu dữ liệu để huấn luyện các mô hình trí tuệ nhân tạo. Mặc dù số liệu quan trắc chất lượng nước trên lưu vực sông Nhuệ–Đáy ngày càng nhiều, chuỗi số liệu này vẫn chưa đủ để xây dựng một mô hình trí tuệ nhân tạo đủ tin cậy. Năm 2006, [3] đã xem xét kết hợp mô hình trí tuệ nhân tạo và mô hình chất lượng nước và đưa ra các lý do cần tích hợp hai loại mô hình và đánh giá hiệu quả. Trong đề xuất này nhóm thực hiện sẽ xây dựng công cụ mô phỏng chất lượng nước dựa vào công nghệ trí tuệ nhân tạo. Theo đó, nghiên cứu sẽ sử dụng công nghệ này xây dựng quan hệ tương quan giữa đầu vào (tải lượng các nguồn xả thải, chỉ số chất lượng nước sông thời điểm hiện tại và lưu lượng nước sông) và đầu ra (chỉ số chất lượng nước tại một số vị trí có số liệu đo trên sông Nhuệ–Đáy). Quan hệ này sau khi xây dựng sẽ được sử dụng để mô phỏng chất lượng nước sông góp phần cho dự báo và phục vụ công tác kiểm soát ô nhiễm sau này. So với các mô hình số trị, công nghệ trí tuệ nhân tạo không yêu cầu người sử dụng cần có kiến thức chuyên môn sâu vì các công việc phân tích, xử lý số liệu đã được ẩn phía sau. Do đó, công nghệ rất thuận lợi trong việc chuyển giao cho thành phố Hà Nội. Công nghệ cũng cho kết quả tính toán một cách nhanh chóng, rất phù hợp phục vụ công tác mô phỏng, dự báo chất lượng nước cũng như hỗ trợ ra quyết định trong việc kiểm soát ô nhiễm. Vì vậy, nghiên cứu sẽ kết hợp cả mô hình thủy lực và mô hình trí tuệ nhân tạo nhằm tăng độ tin cậy cho mô hình trí tuệ nhân tạo trong đó mô hình thủy lực được sử dụng để tạo lập cơ sở dữ liệu theo các kịch bản khác nhau cùng với dữ liệu đo đạc có sẵn làm đầu vào cho mô hình trí tuệ nhân tạo.

Bài báo này trình bày một số kết quả đạt được của cách tiếp cận mới này áp dụng cho lưu vực sông Nhuệ–Đáy. Kết quả nghiên cứu trình bày trong bài báo có thể hỗ trợ các nhà quản lý trong việc đánh giá hiện trạng và dự báo chất lượng nước, phục vụ công tác kiểm soát và quản lý chất lượng nước trên lưu vực sông Nhuệ Đáy.

2. Phương pháp nghiên cứu và tài liệu thu thập

2.1. Giới thiệu về khu vực nghiên cứu

Lưu vực sông Nhuệ–Đáy nằm ở hữu ngạn sông Hồng tọa độ từ 20° đến 21°20' vĩ độ Bắc, và từ 105° đến 105°30' kinh độ Đông, với tổng diện tích tự nhiên là 7665 km² [1] (hình 1). Sông Đáy là một phân lưu của sông Hồng bắt nguồn từ hệ thống phân chứa lũ Vân Cốc thuộc huyện Phúc Thọ chảy qua các tỉnh Hà Nội, Hà Nam, Ninh Bình, Nam Định trước khi đổ ra biển ở cửa Đáy. Sông Nhuệ lấy nước từ sông Hồng qua cống Liêm Mạc ở Từ Liêm, Hà Nội để cấp nước tưới và đồng thời làm chức năng tiêu nước cho các quận/huyện Cầu Giấy, Hà Đông, Từ Liêm, Thanh Trì, Hoài Đức, Thường Tín, Thanh Oai, Phú Xuyên của TP. Hà Nội rồi đổ vào sông Đáy ở Phú Lý [4].

BẢN ĐỒ VỊ TRÍ SÔNG NHUỆ, ĐÁY

Tỷ lệ: 1:400,000



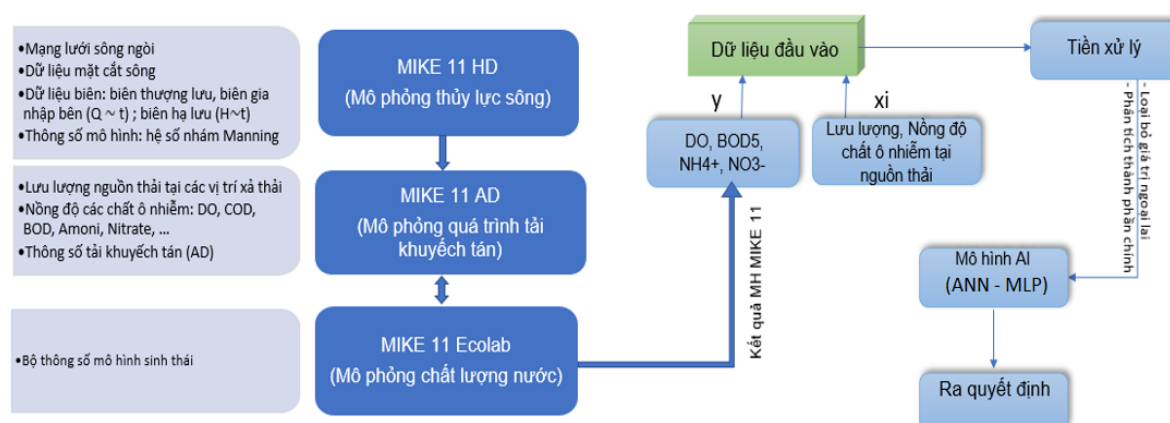
Hình 1. Sông Nhuệ Đáy thuộc địa bàn thành phố Hà Nội.

Khu vực nằm trong vùng có khí hậu nhiệt đới ẩm gió mùa, có một mùa đông lạnh và khô; một mùa hè nóng ẩm và mưa nhiều [4]. Mùa mưa thường kéo dài 5 tháng, từ tháng V đến tháng X; ba tháng liên tục có lượng mưa cao nhất là tháng VII, VIII và IX [5]. Chế độ thủy văn trên lưu vực sông Đáy phụ thuộc vào dòng chảy của nước sông Hồng và các sông nhánh [5]. Do đó mà dòng chảy trên lưu vực sông phân bố không đều theo không gian và thời gian. Theo không gian, dòng chảy lớn nhất ở núi Ba Vì, phần hữu ngạn lưu vực có dòng chảy lớn hơn phần tả ngạn. Sự phân bố theo thời gian theo phân phối dòng chảy năm phụ thuộc vào sự phân phối theo mùa. Trên lưu vực sông Nhuệ–Đáy, các nguồn xả thải chủ yếu từ các nguồn sinh hoạt và đô thị, từ các làng nghề, tiểu thủ công nghiệp; bệnh viện; một phần từ nông nghiệp, thủy sản và các nguồn khác do đó chất lượng nước đang bị ô nhiễm, nhiều vị trí cục bộ ô nhiễm nghiêm trọng. Chất lượng nước lưu vực sông Nhuệ Đáy những năm gần

đây có sự ô nhiễm cục bộ theo thời gian. Vào mùa khô, nước ít, một số vị trí nước có màu đen, mùi hôi, đặc biệt một số vùng mặt nước bị thu hẹp, dòng chảy thấp: đoạn Phúc La–Đồng Quan, cầu Mai Lĩnh. Mùa mưa lượng nước được tăng cường nhiều hơn nhưng lưu vực sông Nhuệ–Đáy lúc này là khu vực xả lũ cho thượng lưu (Hà Nội) nên ngoài lưu lượng nước thường kèm theo rác thải gây ô nhiễm cho lưu vực sông.

2.2. Phương pháp nghiên cứu

Nghiên cứu sử dụng kết hợp các mô hình thủy lực và mô hình trí tuệ nhân tạo trong mô phỏng tính toán, mô phỏng chất lượng nước lưu vực sông Nhuệ – Đáy. Trong đó, phương pháp mô hình toán sử dụng mô hình MIKE HD, AD và Ecolab một chiều để mô phỏng diễn biến chất lượng nước và tạo cơ sở dữ liệu đầu vào cho mô hình trí tuệ nhân tạo. Nghiên cứu sử dụng mô hình MIKE mô phỏng thủy lực và chất lượng nước lưu vực sông với chuỗi dữ liệu nằm trong giai đoạn 2016 – 2020 bước thời đoạn ngày để hiệu chỉnh và kiểm định mô hình, sau đó kết quả mô phỏng chất lượng nước được sử dụng làm dữ liệu đầu vào cho mô hình AI. Mô hình AI sử dụng thuật toán trí tuệ nhân tạo để xây dựng quan hệ hồi quy giữa chất lượng nước tại một số vị trí kiểm soát và các biến ảnh hưởng đến chất lượng tại các vị trí kiểm soát này (lưu lượng nước sông, nồng độ các chất ô nhiễm và lưu lượng tại các nguồn thải). Hình 2 trình bày sơ đồ kết nối giữa mô hình thủy lực, mô hình chất lượng nước và mô hình trí tuệ nhân tạo sử dụng trong nghiên cứu này. Chi tiết về hai mô hình này được trình bày dưới đây.



Hình 2. Phương pháp nghiên cứu.

2.2.1. Mô hình MIKE 11

✓ Mô hình thủy động lực MIKE 11

Mô hình MIKE 11 là bộ mô hình một chiều được phát triển bởi Viện thủy lực Đan Mạch (DHI) từ mô hình gốc đầu tiên ra đời năm 1972 dùng để mô phỏng thủy lực, chất lượng nước và vận chuyển nước trong sông [6–9]. Mô hình giúp mô phỏng diễn biến dòng chảy cũng như mực nước trong hệ thống sông. Mô đun thủy động lực học được xây dựng trên cơ sở lý thuyết của hệ phương trình Saint–Venant, có dạng:

$$\frac{\partial Q}{\partial x} + b \frac{\partial h}{\partial t} = q \quad (1)$$

$$\frac{\partial Q}{\partial t} + \frac{\partial \left(\alpha \frac{Q^2}{A} \right)}{\partial x} + gA \frac{\partial h}{\partial x} + \frac{gQ|Q|}{C^2 AR} = 0 \quad (2)$$

Trong đó Q là lưu lượng dòng chảy, h là mực nước; t là thời gian; x là biến khoảng cách; g là gia tốc trọng trường; A, b là diện tích và chiều rộng mặt cắt ướt; g là gia tốc trọng trường; α là hệ số hiệu chỉnh lưu tốc; C là hệ số Chezy; R là bán kính thủy lực.

✓ Mô đun chất lượng nước

Mô đun truyền tải khuếch tán AD: quá trình truyền tải khuếch tán được mô tả bằng phương trình [8]:

$$\frac{\partial AC}{\partial t} + \frac{\partial QC}{\partial x} - \frac{\partial}{\partial x} \left(AD \frac{\partial C}{\partial x} \right) = -AKC + C_2 q \quad (3)$$

Trong đó C là nồng độ (Kg/m^3); D là hệ số khuếch tán; q là lưu lượng nhập lưu đơn vị (m^2/s); K là hệ số phân hủy sinh học, K chỉ được dùng khi các hiện tượng hay quá trình xem xét có liên quan đến các phản ứng sinh hóa.

Phương trình tải khuếch tán phản ánh hai cơ chế vận chuyển: (1) Quá trình vận chuyển chất do dòng chảy (*advection*); (2) Quá trình khuếch tán các chất do dòng chảy rối (*turbulent diffusion*).

Mô đun Ecolab: Mô đun sinh thái trong mô hình MIKE 11 mô phỏng các quá trình biến đổi sinh-hóa của các biến chất lượng nước trong Template 4 với các quá trình được thể hiện trong phương trình: trao đổi nhiệt, biến đổi chất hữu cơ, biến đổi nitơ, biến đổi photpho và biến đổi oxy trong sông bằng các phương trình:

$$\frac{dT}{dt} = \text{insolation} - \text{radiation} \quad \text{khi } t \in [t_{up}; t_{down}]; \quad (4)$$

$$- \text{radiation} \quad \text{khi } t \notin [t_{up}; t_{down}]$$

$$\frac{dBOD}{dt} = -BOD_{decay} \quad (5)$$

$$\frac{dNH_4_N}{dt} = N_BOD_{decay} - \text{Nitrification} \quad (6)$$

$$\frac{dNO_3_N}{dt} = \text{Nitrification} - \text{Denitrification} \quad (7)$$

$$\frac{dOP}{dt} = \text{PPdecay} - \text{PPformation} + \text{OPreleaseFromBOD} - \text{OpplantUptake} \quad (8)$$

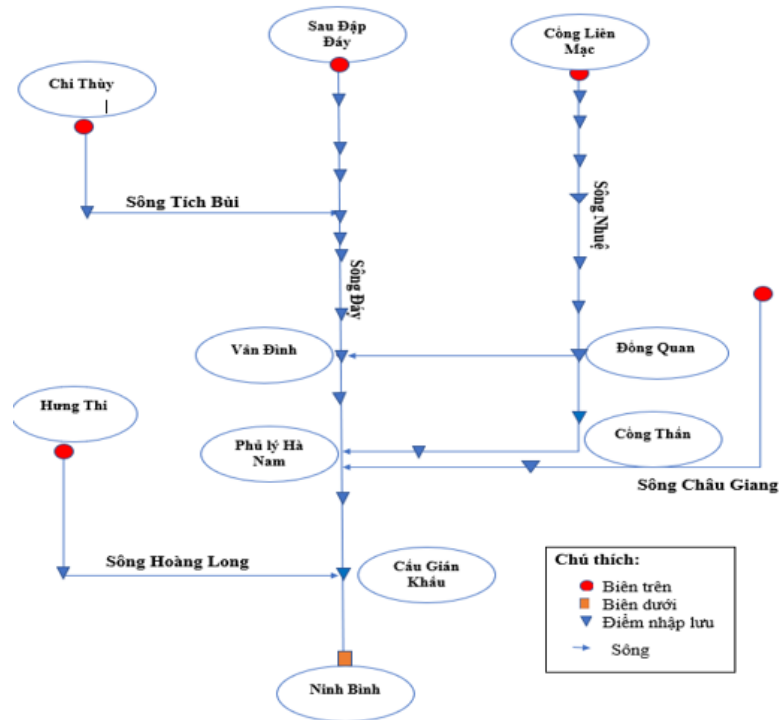
$$\frac{dPP}{dt} = - \text{PPdecay} + \text{PPformation} - \text{PPsedimentation} + \text{PPresuspension}$$

$$\frac{dDO}{dt} = \text{Reaera} + \text{phtsyn} - \text{respT} - BOD_{decay} - SOD - \text{Nitrification} \quad (9)$$

Trong đó T: nhiệt độ của nước ($^{\circ}\text{C}$); insolation, radiation: Bức xạ mặt trời vào và ra BOD_{decay} : BOD của chất hữu cơ hòa tan (mg/l); K_{d3} là tốc độ phân hủy của chất hữu cơ hòa tan ($1/\text{day}$); DO: nồng độ oxy hòa tan (mg/l); NH_4_N : Amoni tính theo Nitơ; NO_3_N : Nitrate tính theo Nitơ, PP: Photpho dạng mảnh; OP: photpho hòa tan.

Sơ đồ mạng sông mô phỏng diễn biến chất lượng nước trên hệ thống sông Nhuệ – Đáy được trình bày ở Hình 3 dưới đây. Đầu vào cho các mô đun thủy lực HD, tải khuếch tán AD và Ecolab của mô hình số trị MIKE 11 bao gồm: mạng lưới sông, dữ liệu mặt cắt sông, dữ liệu biên, thông số hệ số nhám (*MIKE 11 HD*); lưu lượng, nồng độ các chất ô nhiễm và vị trí nguồn thải, bộ thông số của các mô đun tải khuếch tán (*MIKE 11 AD*) và sinh thái (*MIKE 11 Ecolab*). Hệ thống sông bao gồm các sông Đáy, Nhuệ, Tích Bù, Hoàng Long, Châu Giang với tổng chiều dài là 434 km và số mặt cắt là 179 mặt cắt (Bảng 1). Biên trên của mô hình là lưu lượng tại các vị trí Chi Thủy, đập Đáy, cống Liên Mạc, Hưng Thi và Châu Giang. Biên

dưới của hệ thống là mực nước tại trạm thủy văn Ninh Bình. Các yếu tố chất lượng nước được mô phỏng bao gồm BOD₅, NH₄-N, NO₃-N, DO và nhiệt độ.



Hình 3. Sơ đồ thủy lực lưu vực sông Nhuệ-Đáy.

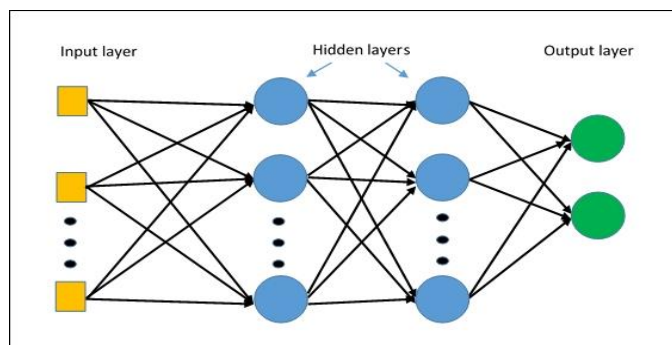
Bảng 1. Mạng lưới sông, số lượng mặt cắt trên các sông.

STT	Tên sông	Chiều dài (m)	Số mặt cắt 1
1	Sông Đáy	168.858	90
2	Sông Nhuệ	73.755	43
3	Sông Tích Bù	101.888	20
4	Sông Châu Giang	27.688	11
5	Sông Hoàng Long	61.589	15

Nguồn: ¹Viện Khoa học Khí tượng thủy văn và Biến đổi khí hậu.

2.2.2. Mô hình trí tuệ nhân tạo

Mô hình AI xây dựng quan hệ hồi qui $y=f(\mathbf{p},x_i)$ trong đó x_i là biến dự báo (lưu lượng, nồng độ các chất ô nhiễm tại các nguồn thải, lưu lượng nước sông tại biên trên), y là biến phụ thuộc (nồng độ BOD₅, DO, Amoniac, Nitrate tại một số điểm kiểm soát trên sông Nhuệ-Đáy), $\mathbf{p}$ là tham số của phương trình hồi qui sẽ được xác định trong quá trình huấn luyện (*training*) sử dụng các thuật toán trí tuệ nhân tạo. Trong nghiên cứu này, thuật toán mạng hồi quy Multilayer Perceptron (*ANN-MLP*) đã được sử dụng [10]. *ANN-MLP* là một thuật toán học máy có giám sát (*Machine Learning*) thuộc lớp Mạng nơ-ron nhân tạo, là tập hợp của các perceptron chia làm nhiều nhóm, mỗi nhóm tương ứng với một layer [10–12]. Các layer này được kết nối hoàn toàn với nhau: Mỗi nút ở từng layer đều kết nối với tất cả các nút của layer kế tiếp. Một neuron nhân tạo (còn được gọi là perceptron) là một hàm biến đổi toán học nhận một hoặc nhiều đầu vào đã được nhân với các giá trị gọi là “*weights*”, cộng các giá trị đó lại với nhau thành một giá trị duy nhất. Sau đó giá trị này được đưa vào một hàm phi tuyến (được gọi là *activation function*) và kết quả của hàm này chính là đầu ra của neuron [10–11].



Hình 4. Minh họa mạng hồi quy Multilayer Perceptron (ANN-MLP): Input layer (lớp đầu vào), Output layer (lớp đầu ra) và Hidden layers (các lớp ẩn).

Các mô hình được xây dựng từ các thuật toán trí tuệ nhân tạo bởi việc huấn luyện dữ liệu. Quá trình huấn luyện dữ liệu trong mô hình nhằm xác định các tham số, có mối liên hệ chặt chẽ với số liệu được đưa vào huấn luyện. Trong khi đó, quá trình kiểm định sẽ đánh giá mức độ tin cậy của mô hình. Các thuật toán được lựa chọn trong từng mô hình tương ứng nhằm tìm ra mô hình phù hợp với nghiên cứu. Lựa chọn phương pháp tối ưu phù hợp trong mô hình với mục đích tăng độ chính xác của mô hình. Các thông số trong mô hình sẽ được xác định bằng cách tính toán và thử sai. Hiệu chỉnh các thông số của mô hình đối với khu vực sao cho kết quả tính toán phù hợp với số liệu thực đo. Sử dụng hệ số tương quan (R^2) để đánh giá hiệu quả mô hình trí tuệ nhân tạo so với kết quả từ mô hình MIKE 11. Chỉ số này phản ánh mức độ giải thích của các biến độc lập đối với biến phụ thuộc trong mô hình hồi quy [13].

Số liệu mô hình bao gồm dữ liệu huấn luyện và kiểm tra, được phân chia với tỷ lệ 90% dữ liệu sử dụng cho huấn luyện, 10% dữ liệu sử dụng cho kiểm tra.

Trước khi đưa vào mô AI để huấn luyện, tập dữ liệu sẽ được phân tích thành phần chính PCA để tìm ra được các thành phần có tương quan tốt với mô hình. Đối với mỗi thành phần chất lượng nước tại mỗi điểm, tìm ra được một bộ dữ liệu huấn luyện riêng sau khi đã loại bỏ các thông số có tương quan thấp. Việc loại bỏ giá trị ngoại lai được thực hiện bởi phương pháp Standard Scaler dựa trên phương sai đơn vị.

Các tham số được hiệu chỉnh của mạng nơ-ron hồi quy nhiều lớp MLP bao gồm: `hidden_layer_sizes` (số lớp ẩn), `activation` (hàm kích hoạt), `solver` (bộ tối ưu hóa), `learning_rate` (tốc độ học). Sử dụng phương pháp thử sai, nghiên cứu đã xác định được các tham số này như sau: số lớp ẩn được cài đặt là 100, hàm kích hoạt là hàm *relu*, bộ tối ưu hóa là *adam*, tốc độ học là *constant* [12, 14–16].

2.2.3. Xây dựng tập dữ liệu cho mô hình trí tuệ nhân tạo từ mô hình MIKE 11

Mô hình AI yêu cầu một lượng lớn dữ liệu để huấn luyện và kiểm định mô hình trong khi chuỗi số liệu đo đạc trên hệ thống sông Nhuệ–Đáy tương đối ngắn, không đủ để xây dựng một mô hình AI đủ tin cậy. Do đó nghiên cứu sẽ sử dụng mô hình MIKE 11 để chạy mô hình với các kịch bản khác nhau để tạo tập dữ liệu cho mô hình AI. Trong mô hình AI, các biến đầu vào bao gồm: lưu lượng, BOD₅, DO, Amoniac, Nitrate tại cống Liên Mạc, Hà Đông, Cầu Tó, Đồng Quan (lưu vực sông Nhuệ) và đập Đáy, Song Phương, cầu Mai Lĩnh, cầu Ba Thá (lưu vực sông Đáy). Các biến-mô phỏng bao gồm: nồng độ BOD₅, DO, Amoniac, Nitrate tại các vị trí kiểm soát Phúc La, cầu Chiếu, Đồng Quan (lưu vực sông Nhuệ) và cầu Mai Lĩnh, cầu Ba Thá, cầu Tế Tiêu (lưu vực sông Đáy). Nghiên cứu đã tạo một chuỗi các số liệu đầu vào cho các biến mô phỏng theo nguyên tắc sau:

- Các tham số chất lượng nước BOD₅, Amoniac, Nitrate được tạo ra ngẫu nhiên trong khoảng từ giá trị cận dưới là các chỉ tiêu chất lượng cho cột B1 của QNVN MT–08/2015–

BTNMT (đủ tiêu chuẩn cho mục đích tưới tiêu, thủy lợi) và cận trên là giá trị gấp đôi hàm lượng chất ô nhiễm hiện tại.

- Tham số DO dao động ngẫu nhiên trong khoảng cận trên là giá trị tại cột B1 của QNVN MT-08/2015-BTNMT và cận dưới là nửa giá trị DO tại thời điểm hiện tại.

- Nhiệt độ tại tất cả các nguồn thải được tạo ngẫu nhiên trong khoảng 15–35°C.

- Lưu lượng nước sông, nghiên cứu xem xét biến đổi của lưu lượng nước tại các biên đập Đáy trên sông Đáy và cống Liên Mạc trên sông Nhuệ. Lưu lượng được tạo ngẫu nhiên dao động trong khoảng từ 3–7 m³/s tại đập Đáy, trong khi đó giá trị này dao động trong khoảng từ 5–60 m³/s tại cống Liên Mạc. Các yếu tố chất lượng nước tại 2 biên này được giữ cố định như điều kiện hiện trạng.

Để có được cơ sở dữ liệu đủ lớn, nghiên cứu đã tạo ngẫu nhiên chuỗi số liệu gồm 6872 dữ liệu (dữ liệu thủy văn, các biến chất lượng nước: nhiệt độ, nồng độ BOD₅, Amoniac, Nitrate, DO) với bước thời gian là 1 ngày. Ngoại trừ các vị trí điều chỉnh này, các vị trí khác không được xem xét trong nghiên cứu này.

3. Kết quả nghiên cứu

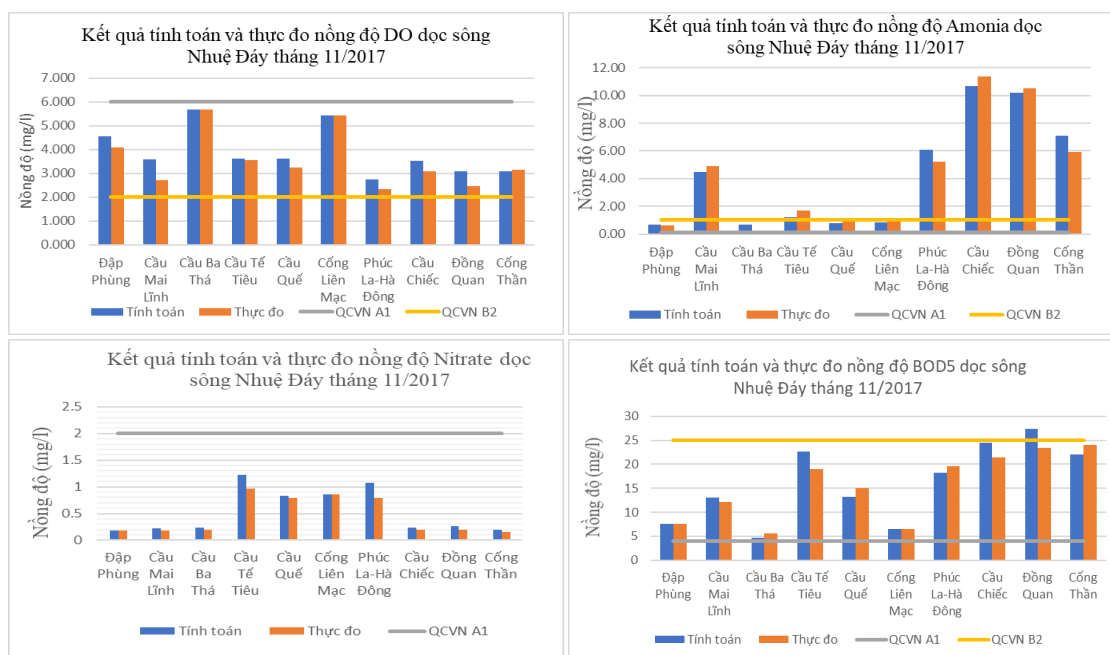
3.1. Kết quả mô phỏng chất lượng nước bằng mô hình MIKE 11

Theo không gian: trên sông Nhuệ nước sông bị ảnh hưởng rất lớn bởi nước thải sinh hoạt, công nghiệp, nông nghiệp của thành phố Hà Nội. Về mùa cạn chất lượng nước phụ thuộc vào chế độ vận hành cống Liên Mạc, chế độ xả nước đập Thanh Liệt và chế độ lấy nước tưới của hệ thống thủy nông. Về mùa lũ, cống Liên Mạc thường đóng, nước sông Nhuệ chủ yếu là nước thải từ thành phố, nước mưa, nước tiêu nông nghiệp, nguồn nước bị ô nhiễm nhưng được bơm thoát ra sông Đáy. Chất lượng nước sông Đáy thay đổi thất thường và phụ thuộc chủ yếu vào chất lượng nước thải từ sinh hoạt, công nghiệp, nông nghiệp, làng nghề xuống các kênh mương, sông nội địa, sau đó đổ dồn vào sông Đáy dọc theo sông.

Với nguồn số liệu thực đo nồng độ các chất tại các vị trí dọc sông, việc hiệu chỉnh mô hình dựa trên việc thay đổi hệ số triết giảm cho từng đoạn sông ứng với mỗi vị trí xả thải. Quá trình hiệu chỉnh thông số mô hình dựa trên sự phù hợp giữa tính toán và thực đo tại các trạm kiểm tra, cụ thể là sự phù hợp về giá trị nồng độ chất lượng nước.

Việc hiệu chỉnh thông số mô hình chất lượng nước Ecolab cũng cần dựa trên sự tương đồng giữa số liệu tính toán và thực đo tại các vị trí quan trắc; với việc hiệu chỉnh mô hình được thực hiện chủ yếu thông qua việc thay đổi các biến trạng thái, hệ số khuếch tán và các chỉ số sinh thái Ecolab của các chất ô nhiễm.

Từ kết quả hiệu chỉnh mô hình (Hình 5) cho kết quả mô phỏng thực đo và tính toán tại Đập Phùng cho thấy nồng độ các chất giữa thực đo và tính toán chênh lệch không nhiều. Nồng độ Amoni, và BOD₅ sai số không đáng kể, nồng độ DO và Nitrate mô phỏng cao hơn thực đo với sai số tương đối 11–26%. Có thể thấy kết quả mô phỏng của mô hình chất lượng nước là tương đối phù hợp khi tính toán lan truyền đối với DO, BOD₅, Amonia (NH₄⁺) và Nitrate (NO₃). Nhìn chung có nồng độ 3 thông số DO, Amoni và BOD₅ so với quy chuẩn A₁ đều vượt ngưỡng quy định, chỉ có nồng độ Nitrate không vượt nên không được sử dụng cho mục đích cấp nước sinh hoạt. Với mục đích sử dụng cho mục đích tưới tiêu và giao thông thủy được đáp ứng tại một số vị trí cho nồng độ các chất khác nhau.



Hình 5. Nồng độ DO, Amoni, Nitrate, BOD₅ thực đo và tính toán giai đoạn hiệu chỉnh tháng 11/2017 tại các vị trí dọc sông Nhuệ Đáy.



Hình 6. Nồng độ DO, Amoni, Nitrate, BOD₅ thực đo và tính toán giai đoạn kiểm định tháng 09/2019 và tháng 09/2020 tại các vị trí dọc sông Nhuệ Đáy.

Hình 6 biểu diễn kết quả mô phỏng cho thấy theo các vị trí quan trắc, giá trị nồng độ DO sông Nhuệ Đáy đều nằm trong giới hạn B₂ QCVN 08:2015/BTNMT [20]. Biến đổi theo không gian cho thấy nồng độ DO tại sông Đáy tốt hơn so với sông Nhuệ. Xét theo biến đổi theo thời gian, nồng độ DO tháng 9/2020 tăng cao hơn tháng 9/2019 nhưng vẫn chỉ nằm ở mức B₂, một số vị trí trên sông Đáy đạt mức A₂ (Đập Phụng) và B₁. Giai đoạn tháng 9/2020, nồng độ DO sông Nhuệ tăng cao tại cầu Chiềng (70%), những vị trí khác tăng từ 4–56%, đều nằm trong mức B₁.

Nồng độ amonia tại sông Đáy phần lớn nằm trong mức B₂, nhưng tại sông Nhuệ vượt ngoài mức B₂ rất nhiều lần tại Phúc La–Hà Đông và cầu Chiềc. So sánh nồng độ amonia trong 2 thời đoạn tháng 9/2019 và tháng 9/2020 trên các vị trí dọc sông Nhuệ Đáy cho thấy sự chênh lệch giữa 2 thời đoạn cao nhất tại cầu Chiềc trên sông Nhuệ với mức tăng 12,3 mg/l, sau đó là cống Liên Mạc, tăng 9,1 mg/l. Tuy nhiên, xét sự sai khác giữa tính toán và thực tế giữa 2 thời đoạn cho thấy mô phỏng chất lượng nước trong tháng 9/2020 cho kết quả tốt với sai số tương đối giữa thực đo và tính toán cao nhất chỉ 24% tại Đồng Quan, còn lại tại các vị trí khác đều cho kết quả sai số dưới 20%.

Trong tháng 9/2020 tại cầu Mai Lĩnh, nồng độ thải nitrate cao gấp 3,3 lần so với cùng kỳ năm trước, điều này cho thấy tại đây xuất hiện lượng xả thải cao đánh kể gây ô nhiễm nguồn nước nghiêm trọng. Ngược lại, tại Đồng Quan, nồng độ thải nitrate cùng kỳ tháng 9 nhưng năm 2020 giảm đi rất nhiều, đưa chất lượng nước sông đang từ vượt mức A₂ về dưới mức A₁, cho thấy đã có những biện pháp can thiệp giúp bảo vệ nguồn nước khu vực này. Đánh giá sai số tương đối giữa nồng độ nitrat tính toán và quan trắc nhận thấy mô hình mô phỏng cho kết quả tương đối tốt. Sai số tương đối giữa tính toán và thực đo kỳ tháng 9/2019 cho sai số dưới 20%; trong khi đó tháng 9/2020 cao nhất cũng chỉ 21%. Với kết quả này, nghiên cứu đã lựa chọn được bộ thông số phù hợp để mô phỏng cho các thời đoạn khác trong lưu vực sông.

Trên các vị trí sông Nhuệ, tại Phúc La–Hà Đông và Cầu Chiềc, nồng độ BOD₅ cao gấp nhiều lần hạng B₂, do vậy cần có các biện pháp cải tạo nguồn nước. Tại cầu Chiềc, nồng độ BOD₅ thời đoạn tháng 9/2020 giảm hơn 50% so với cùng kỳ năm 2019, cho thấy nên có những cơ chế hạn chế nguồn thải và làm sạch nguồn nước sông sẽ giúp cho chất lượng nước sông được cải thiện đáng kể. Xét sai số tương đối giữa kết quả mô phỏng và thực đo nhận thấy mô hình mô phỏng thời đoạn tháng 9/2020 (sai số tương đối (SSTD) max = 28% tại cầu Mai Lĩnh) tốt hơn tháng 9/2019 (SSTD = 22% tại Đồng Quan). Tuy nhiên kết quả mô phỏng này tương đối hợp lý và có thể sử dụng mô phỏng cho các vị trí trong các thời đoạn khác trên lưu vực sông Nhuệ Đáy để đánh giá chất lượng nước sông, qua đó tìm các giải pháp giúp cải thiện nguồn nước sông tốt hơn.

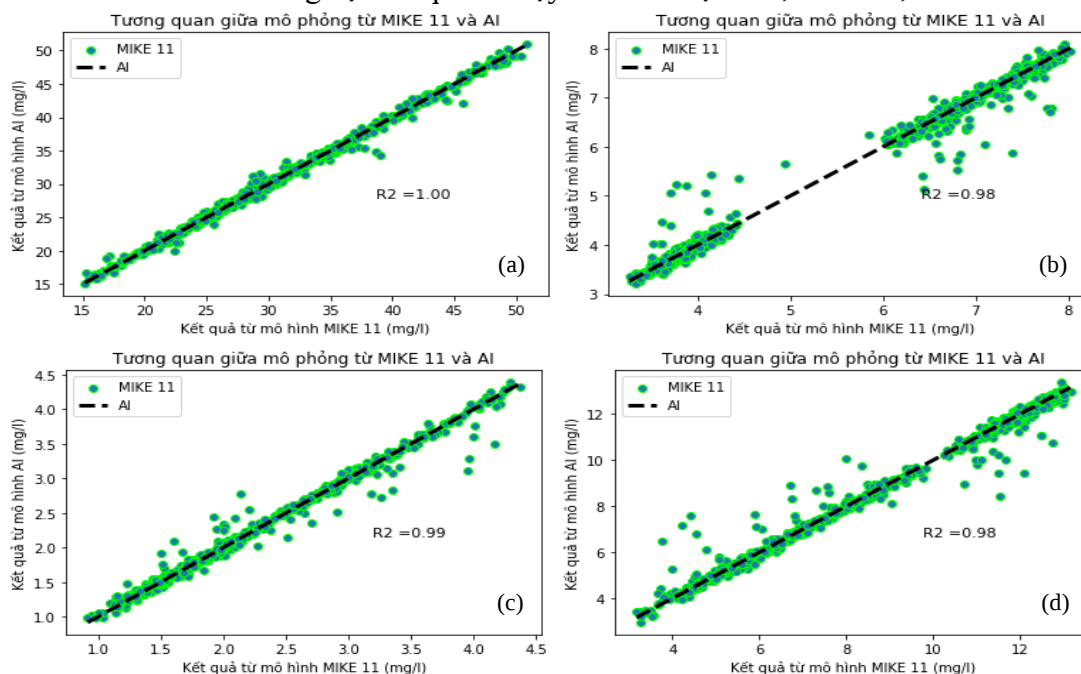
3.2. Kết quả mô hình trí tuệ nhân tạo AI

Quá trình huấn luyện được đánh giá bởi các hệ số tương quan R² thể hiện mức độ tin cậy của kết quả huấn luyện từ mô hình AI trong nghiên cứu (ANN–MLP). Bên cạnh đó, mức độ tin cậy cũng được đánh giá thông qua chỉ số Nash (NSE). Mô hình mô phỏng tính toán tại các điểm: Phúc La, cầu Chiềc, Đồng Quan (lưu vực sông Nhuệ) và cầu Mai Lĩnh, cầu Ba Thá, cầu Tế Tiêu (lưu vực sông Đáy). Bảng 1, Hình 7 và 8 trình bày kết quả kiểm định mô hình AI tương ứng với các tham số này.

Bảng 1. Hệ số tương quan R², chỉ số NSE so sánh kết quả từ mô hình MIKE 11 và Mô hình AI.

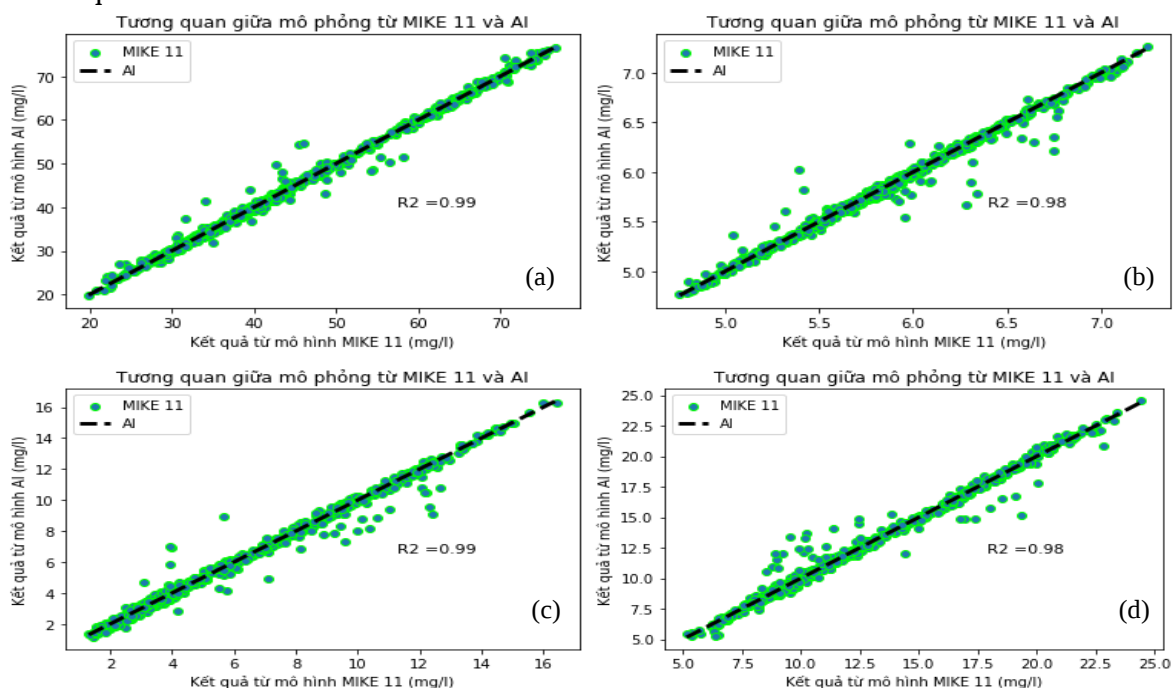
Vị trí	BOD ₅		DO		NH ₄ ⁺		NO ₃ ⁻	
	R ²	NSE	R ²	NSE	R ²	NSE	R ²	NSE
Cầu Mai Lĩnh	~1,00	0,98	0,98	0,98	0,98	0,98	0,98	0,96
Cầu Ba Thá	0,99	0,99	0,98	0,98	0,97	0,97	0,98	0,98
Cầu Tế Tiêu	0,98	0,97	0,99	0,99	0,98	0,99	0,98	0,99
Phúc La	~1,00	0,98	0,98	0,98	0,99	0,98	0,98	0,97
Cầu Chiềc	~1,00	0,98	0,99	0,99	0,99	0,96	0,99	0,99
Đồng Quan	0,99	0,99	0,98	0,98	0,98	0,98	0,98	0,98
Cầu Cống Thần	0,99	0,98	0,98	0,99	0,99	0,98	0,98	0,99

Nếu phần biến thiên do phần dư càng nhỏ, nghĩa là khoảng cách từ các điểm quan sát đến đường ước lượng hồi quy càng nhỏ thì phần biến thiên do hồi quy sẽ càng cao, khi đó giá trị R^2 sẽ càng cao. Kết quả thu được từ quá trình huấn luyện ở tất cả các trạm đạt giá trị R^2 từ 0,98 đến ~1 cho thấy khả năng dự báo của mô hình AI gần sát với kết quả từ mô hình MIKE 11. Chỉ số NSE cũng đạt kết quả tin cậy cao lần lượt từ 0,96 đến 0,99.



Hình 7. Tương quan giữa kết quả mô phỏng từ mô hình MIKE 11 và AI tại điểm Cầu Mai Lĩnh – sông Đáy: (a) Nồng độ BOD₅; (b) Nồng độ DO; (c) Nồng độ NH₄⁺; (d) Nồng độ NO₃⁻.

Hình 7 và Hình 8 so sánh kết quả chất lượng nước từ mô hình MIKE 11 và mô hình AI tại hai trạm Cầu Mai Lĩnh và Cầu Cống Thần. Kết quả cho thấy việc dự báo từ AI gần như trùng lặp với kết quả từ mô hình MIKE 11. Trong đó, thông số BOD₅ có độ hồi quy tốt nhất, có kết quả tốt nhất.



Hình 8. Tương quan giữa kết quả mô phỏng từ mô hình MIKE 11 và AI tại điểm Cầu Cống Thần – sông Nhuệ: (a) Nồng độ BOD₅; (b) Nồng độ DO; (c) Nồng độ NH₄⁺; (d) Nồng độ NO₃⁻.

Với kết quả được cho là tin cậy, mô hình AI cho phép dự báo chất lượng nước mặt lưu vực sông Nhuệ–Đáy thay cho mô hình MIKE 11 nhằm thực hiện dự báo chất lượng nước mặt cho lưu vực sông cho các kịch bản trong tương lai. Đây là công cụ thực hiện dự báo với tốc độ nhanh chóng, thuận lợi cho thành phố Hà Nội trong việc quản lý cũng như đưa ra các chính sách phù hợp.

4. Kết luận

Bài báo đã thu được các kết quả chính và nổi bật như sau:

- Sử dụng phương pháp sử dụng mô hình số trị MIKE 11 kết hợp với mô hình trí tuệ nhân tạo mô phỏng chất lượng nước sông tương đối tốt.

- Mô hình trí tuệ nhân tạo mạng hồi quy Multilayer Perceptron (ANN–MLP) đã được ứng dụng để mô phỏng nồng độ BOD₅, DO, Amoniac, Nitrate trên lưu vực sông Nhuệ–Đáy. Kết quả thu được từ quá trình huấn luyện ở tất cả các trạm đạt chỉ số R² từ 0.98 đến ~1 cho thấy khả năng dự báo của mô hình AI gần sát với kết quả từ mô hình MIKE 11. Với kết quả được cho là tin cậy, mô hình AI cho phép mô phỏng chất lượng nước mặt lưu vực sông Nhuệ–Đáy thay cho mô hình MIKE 11 đồng thời là cơ sở để xây dựng một mô hình dự báo chất lượng nước mặt cho lưu vực sông cho các kịch bản trong tương lai. Đây là công cụ thực hiện dự báo với tốc độ nhanh chóng, thuận lợi cho thành phố Hà Nội trong việc quản lý cũng như đưa ra các chính sách phù hợp.

- Mô phỏng chất lượng nước trên lưu vực sông đã đánh giá được tình trạng chất lượng nước như sau: chất lượng nước lưu vực sông vẫn chưa được cải thiện rõ rệt, xuất hiện tình trạng gia tăng ô nhiễm cục bộ; chất lượng nước sông Nhuệ đoạn chảy qua đoạn Phúc La– Hà Đông đến hạ nguồn sông phân xuyên ở mức kém, tại một số thời điểm còn xảy ra tình trạng ô nhiễm nghiêm trọng và ít chuyển biến giữa các mùa trong năm; chất lượng nước sông Đáy nhìn chung ở mức trung bình, ô nhiễm cục bộ tại khu vực cầu Mai Lĩnh. Nước sông Nhuệ chủ yếu chỉ đáp ứng được cho mục giao thông thủy và các mục đích tương đương khác, trên sông Đáy tại một số vị trí được sử dụng thêm cho mục đích tưới tiêu. Do đó, cần có các giải pháp khắc phục tình trạng ô nhiễm nguồn nước lưu vực sông.

Ngoài những kết quả đã đạt được, do hạn chế về thời gian nên việc mô phỏng các thông số lượng nước khác: Photpho, Coliform, Ecoli, chưa được tiến hành để đánh giá tổng quan hơn về chất lượng nước lưu vực sông. Đồng thời, nhóm tác giả định hướng nghiên cứu dự báo và kiểm soát nguồn ô nhiễm chất lượng nước cho nguồn nước lưu vực sông Nhuệ–Đáy nói riêng và hệ thống sông nói chung sẽ được thực hiện trong các nghiên cứu tiếp theo.

Đóng góp của tác giả: Xây dựng ý tưởng nghiên cứu: T.A.P.; Lựa chọn phương pháp nghiên cứu: T.A.P., N.T.T.; Xử lý số liệu: T.A.P., H.T.A., T.T.N.; Viết bản thảo bài báo: T.A.P., H.T.A., T.T.N.; Chỉnh sửa bài báo: T.A.P., N.T.T., H.T.A., T.T.N.

Lời cảm ơn: Nghiên cứu được thực hiện bởi sự tài trợ của đề tài nghiên cứu khoa học thuộc chương trình nghiên cứu ứng dụng và phát triển công nghệ cấp thành phố, mã số đề tài 01C–09/01–2020–3 và đề tài nghiên cứu ứng dụng và phát triển công nghệ cấp Quốc gia, mã số ĐTĐL.CN.18/21.

Lời cam đoan: Tập thể tác giả cam đoan bài báo này là công trình nghiên cứu của tập thể tác giả chưa được công bố ở đâu, không được sao chép từ những nghiên cứu trước đây; không có sự tranh chấp lợi ích trong nhóm tác giả.

Tài liệu tham khảo

1. Trung tâm quan trắc Môi trường, Tổng cục Môi trường (<http://www.cem.gov.vn/tin-tuc-moi-truong/dien-bien-chat-luong-nuoc-luu-vuc-song-nhue-day>).
2. Chapra, S.C. Surface water–quality modeling. *Waveland Press*, 2008.
3. Chau, K.W. A review on integration of artificial intelligence into water quality modelling. *Mar. Pollut. Bull.* 2006, 52(7), 726–733.

4. Xuân, T.T. Tài nguyên nước các hệ thống sông chính ở Việt Nam. NXB Khoa học và Kỹ thuật, Hà Nội, **2012**, tr. 298.
5. Hằng, V.T. Đánh giá tác động của biến đổi khí hậu đến biến động tài nguyên nước lưu vực sông Nhuệ–Đáy thuộc thành phố Hà Nội (Doctoral dissertation, Luận văn thạc sỹ khoa học, Trường Đại học KHTN Hà Nội, 2010).
6. Tayfur, G.; Singh, V.P. Predicting mean and bankfull discharge from channel cross-sectional area by expert and regression methods. *Water Resour.* **2011**, 25(5), 1253–1267.
7. Techow, V.; Maidment, D.R.; Mays, L.W. Thủy văn ứng dụng – Đỗ Hữu Thành dịch – NXB Giáo dục, 1994.
8. DHI Water and Environment (DHI). MIKE 11 User Manual. DHI, Demark, 2009.
9. DHI. MIKE 11, A modelling system for rivers and channel, user guide. http://manuals.mikepoweredbydhi.help/2017/Water_Resources/MIKE11_UserManual.pdf, 2017.
10. Pham, Q.B.; Abba, S.I.; Usman, A.G.; Linh, N.T.T.; Gupta, V.; Malik, A.; Costache, R.; Vo, N.D.; Tri, D.Q. Potential of hybrid data–intelligence algorithms for multi-station modelling of rainfall. *Water Res. Manage.* **2019**, 33(15), 5067–5087.
11. Chena, Y.H.; Chang, F.J. Evolutionary artificial neural networks for hydrological systems forecasting. *J. Hydrol.* **2009**, 367(1–2), 125–137.
12. Haghiabi, A.H.; Nasrolahi, A.H.; Parsaie, A. Water quality prediction using machine learning methods. *Water Quality Res. J.* **2018**, 53(1), 3–13.
13. Hải, C.H.; Phương, T.A.; Như, T.Q.; Cường, T.M. Áp dụng mô hình trí tuệ nhân tạo vào dự báo lưu lượng đến hồ lưu vực sông Ba. *Tạp chí Khí tượng Thủy văn* **2019**, 705, 22–33.
14. Philippe, B. The WGNE survey of verification methods for numerical prediction of weather elements and severe weather events. Meteo – France, Toulouse, 2003.
15. Chen, Yingyi. A review of the artificial neural network models for water quality prediction. *Applied Sci.* **2020**, 10(17), 5776.
16. Ighalo, J.O.; Adewale, G.A.; Gonçalo, M. Artificial intelligence for surface water quality monitoring and assessment: a systematic literature analysis. *Model. Earth Syst. Environ.* **2021**, 7(2), 669–681.
17. Han, H.G.; Qiao, J.F.; Chen, Q.L. Model predictive control of dissolved oxygen concentration based on a self–organizing RBF neural network. *Control Eng. Pract.* **2012**, 20, 465–476.
18. Rajae, T.; Boroumand, A. Forecasting of chlorophyll–a concentrations in South San Francisco Bay using five different models. *Appl. Ocean Res.* **2015**, 53, 208–217.
19. Augspurger, T.; Keller, A.E.; Black, M.C.; Cope, W.G.; Dwyer, F.J. Water quality guidance for protection of freshwater mussels (Unionidae) from ammonia exposure. *Environ. Toxicol. Chem.: Int. J.* **2003**, 22(11), 2569–2575.
20. Bộ Tài nguyên và Môi trường. QCVN 08:MT 2015/BTNMT – Quy chuẩn kỹ thuật quốc gia về chất lượng nước mặt, Hà Nội, 2015.

Research on combining hydraulic model and artificial intelligence model to simulate water quality of Nhue – Day river in Ha Noi City

Hoang Thi An^{1*}, Tran Thuy Nhung¹, Nguyen Thanh Thuy², Tran Anh Phuong¹

¹ Department of Water Resources Engineering and Technology, Water Resources Institute; anht510@wru.vn; tranthuynhung1990@gmail.com; phuongtran.monre@gmail.com

² ThuyLoi University; thanhthuy_rt@tlu.edu.vn

Abstract: The Nhue–Day river basin is well-known by its critical water quality, causing negative impacts on locals' health as well as landscape and environmental storage in recent years. The study develops artificial intelligence (AI) model to simulate the water quality of Nhue–Day river, based on the results of hydrodynamic simulation and water quality in the hydraulic model MIKE11. The input data include: topography, hydrological, and discharge data at monitoring locations to simulate river water quality. The AI model with MLP – ANN algorithm builds a regression relationship between river water flow and concentration of discharge sources (predictor variables) with water quality indicators (dependent variable) to forecast river water quality at control locations under different control and waste management scenarios quickly. The results from the AI model are close to the results from the MIKE 11 model with an R^2 index of 0.98 or higher, with simulation time being thousands of times faster than hydraulic model, which is an effective tool, allowing to get water quality forecast quickly. This study provides management tools for managers to quickly assess the impact of solutions to planning, managing, and controlling water pollution in the Nhue – Day river basin.

Keywords: Water quality; Nhue–Day river basin; MIKE 11 model; Artificial Intelligence model; MLP–ANN.