

ĐÁNH GIÁ HIỆN TRẠNG Ô NHIỄM KIM LOẠI NẶNG VÀ ĐỀ XUẤT MÔ HÌNH TIỀN LƯỢNG TẠI HẠ LƯU SÔNG SÀI GÒN, TỈNH BÌNH DƯƠNG

EVALUATION OF HEAVY METALS POLLUTION AND PROPOSING PREDICTIVE MODELS ON DOWNSTREAM OF SAI GON RIVER SUBBASIN, BINH DUONG PROVINCE

Nguyễn Ngọc Huy¹, Lê Đức Trung², Chế Đình Lý², Lê Thị Xuân Thùy³

¹Công ty TNHH Môi trường xanh SUSTECH; nguyemt@gmail.com

²Viện Môi trường và Tài nguyên - Đại học Quốc gia Thành phố Hồ Chí Minh

³Trường Đại học Bách khoa - Đại học Đà Nẵng

Tóm tắt - Bài báo đánh giá hiện trạng ô nhiễm kim loại nặng trong nước, trầm tích tại tiểu lưu vực hạ lưu sông Sài Gòn và đề xuất mô hình tiên lượng nồng độ kim loại nặng trong nước bằng phương pháp thống kê và mạng trí tuệ nhân tạo. Kết quả cho thấy tại khu vực nghiên cứu, nồng độ Fe vượt quá quy chuẩn kỹ thuật quốc gia về chất lượng nước mặt, nồng độ các kim loại nặng khác nằm trong quy chuẩn cho phép. Bên cạnh đó, hàm lượng các kim loại nặng trong trầm tích tại khu vực cuối tuyến khảo sát cao hơn khu vực đầu tuyến và tương quan thuận với nồng độ Cu, Zn và Cr, tương quan nghịch với nồng độ Ni và Cd trong nước. Mô hình tiên lượng kim loại nặng được xây dựng nhằm tiên đoán nồng độ các kim loại nặng trong nước theo phương pháp hồi quy tuyến tính đa biến và mạng trí tuệ nhân tạo đã tìm thấy 3 mô hình tiên lượng có xác suất xuất hiện hệ số tương quan $r \geq 0,9$ đạt khoảng 9 – 15%.

Từ khóa - sông Sài Gòn; ô nhiễm kim loại nặng; mô hình tiên lượng; hồi quy tuyến tính đa biến; mạng trí tuệ nhân tạo

1. Đặt vấn đề

Hiện nay, ô nhiễm kim loại nặng (KLN) trong nước và trầm tích tại các kênh, rạch, sông, suối đang diễn ra phổ biến trên thế giới. Các biện pháp loại bỏ ion KLN trong nước tại các nhà máy xử lý nước hiện nay còn hạn chế. Ngoài ra, do tính khả thi về mặt kinh tế và khó khăn trong việc kiểm soát thành phần KLN trong sản phẩm sau xử lý nên các công nghệ xử lý bùn thải hiện nay chưa đáp ứng được yêu cầu xử lý triệt để. Khi thải bỏ tùy tiện chất thải vào môi trường, KLN sẽ dễ dàng phát tán sang môi trường đất, nước mặt và nước ngầm. Đây là mối nguy hiểm tiềm tàng đối với sức khỏe của con người cũng như hệ sinh thái tự nhiên.

Tiểu lưu vực (TLV) hạ lưu sông Sài Gòn, tỉnh Bình Dương hiện nay đang phải tiếp nhận nước thải từ 5 khu công nghiệp, 2 cụm công nghiệp và 2 nhà máy xử lý nước thải sinh hoạt. Trong số những ngành nghề phát triển tại địa phương thì dệt may; hóa chất, nhựa, cao su; thiết bị điện, điện tử; cơ khí, sản phẩm kim loại; vật liệu xây dựng, gỗ, giấy; xử lý chất thải là những ngành sản xuất có khả năng phát thải KLN trong nước và bùn thải [1].

Hiện nay trên thế giới, nhiều nghiên cứu đã ứng dụng các mô hình diễn toán nhằm mô tả các quá trình ô nhiễm, lan truyền KLN trong môi trường nước và trầm tích [2 - 3]. Dữ liệu đầu vào cho các mô hình này cần độ phong phú, đủ lớn và có độ tin cậy. Thêm vào đó, sự hiểu biết về các quá trình lan truyền của KLN trong môi trường là yêu cầu bắt buộc để xây dựng một mô hình tính toán lan truyền ô nhiễm. Tại TLV hạ lưu sông Sài Gòn, Bình Dương, công tác quan trắc KLN trong nước và trầm tích ở một số địa phương được tiến hành từ năm 2012 với tần suất quan trắc

Abstract - This research evaluates the current concentration of heavy metals in the downstream of Sai Gon River subbasin and proposes the predictive models of heavy metals in water by statistical and artificial neural network methods. The results have shown that the concentration of Fe exceeds National Technical Regulation on surface water quality while the other heavy metals remain in the standard. Moreover, the concentration of heavy metals in sediments on downstream is higher than that upstream, and it has positive correlation with that of Cu, Zn and negative correlation with that of Ni and Cd in water. The predictive models of heavy metals are built to predict the concentration of heavy metals in water by applying multiple linear regression and artificial neural network methods and have found that 3 models have the existing probability of correlation coefficient $\geq 0,9$ of about 9 – 15%.

Key words - Sai Gon river; heavy metal pollution; predictive model; multiple linear regression; artificial neural network

6 lần/năm. Mức độ ô nhiễm một số KLN trong nước như Fe, Zn, Cr, Ni có dấu hiệu tăng trong 05 năm qua (2012 – 2016), vượt quy chuẩn cho phép theo QCVN 08-MT:2015/BTNMT – Quy chuẩn kỹ thuật quốc gia về chất lượng nước mặt. Ngoài ra, hàm lượng KLN trong bùn lắng, trầm tích cũng cho thấy dấu hiệu tăng dần về phía hạ lưu, một số thời điểm vượt ngưỡng quy chuẩn cho phép theo QCVN 43:2012/BTNMT – Quy chuẩn kỹ thuật quốc gia về chất lượng trầm tích.

Trước sự gia tăng mức độ ô nhiễm KLN tại TLV hạ lưu sông Sài Gòn, một số nghiên cứu đã được tiến hành nhằm đánh giá hiện trạng ô nhiễm KLN trong nước và trầm tích [4]. Tuy vậy, chưa có nghiên cứu nào đề xuất mô hình tiên lượng nồng độ KLN. Bài báo ứng dụng phương pháp phân tích thống kê và mạng trí tuệ nhân tạo để đánh giá hiện trạng ô nhiễm KLN tại TLV hạ lưu sông Sài Gòn và đề xuất mô hình tiên lượng nồng độ một số KLN trong nước.

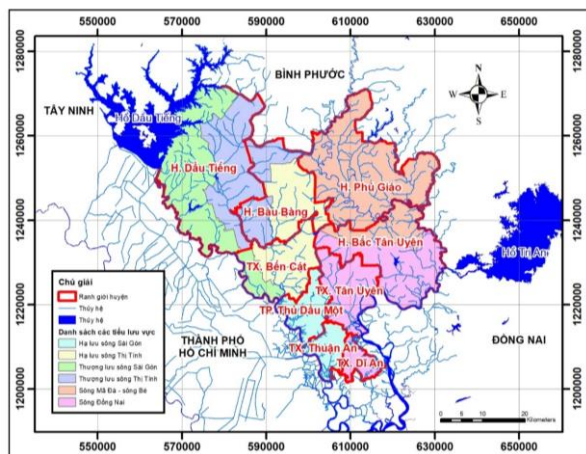
2. Nội dung nghiên cứu

2.1. Phạm vi nghiên cứu

Tiểu lưu vực hạ lưu sông Sài Gòn thuộc phạm vi nghiên cứu có giới hạn theo hành chính bao gồm toàn bộ thành phố Thủ Dầu Một và toàn bộ thị xã Thuận An. Theo Quyết định 3613/QĐ-UBND 26/12/2016 của Ủy ban nhân dân (UBND) tỉnh Bình Dương, việc bảo vệ tài nguyên nước mặt được quy định đối với chất lượng nước sông Sài Gòn đạt cột A2 theo QCVN 08-MT:2015/BTNMT và đạt cột B1 cho các sông, suối còn lại.

Khu vực mang nhiều đặc tính của vùng khí hậu nhiệt đới gió mùa nóng ẩm như nắng nhiều, nhiệt độ cao đều quanh năm, lượng mưa trung bình tháng khá cao, mùa mưa

thường bắt đầu từ tháng 05 đến tháng 10 và mùa khô bắt đầu từ tháng 11 đến tháng 04 năm sau. Địa chất khu vực gồm 3 nhóm đất chính là đất phèn, đất xám và đất mặt nước (kênh, rạch, suối). Trong đó đất phèn phân bố chủ yếu dọc sông Sài Gòn từ thượng lưu về hạ lưu.



Hình 1. Bản đồ các tiểu lưu vực quy hoạch tài nguyên nước tỉnh Bình Dương

2.2. Phương pháp nghiên cứu

2.2.1. Thu thập tài liệu

Dữ liệu chất lượng nước mặt và trầm tích được thu thập từ các báo cáo quan trắc nước mặt và trầm tích định kỳ hàng năm của Sở Tài nguyên và Môi trường tỉnh Bình Dương; dữ liệu lượng mưa và lưu lượng dòng chảy được thu thập từ báo cáo liên quan [5].

Tần suất quan trắc 1 tháng/lần đối với nước mặt và 2 tháng/lần đối với trầm tích. Phương pháp lấy mẫu và bảo quản mẫu, phương pháp phân tích mẫu theo các quyết định 1153/QĐ-BTNMT và 1613/QĐ-BTNMT của Bộ Tài nguyên và Môi trường. Các chỉ tiêu KLN trong nước và trầm tích được nghiên cứu bao gồm Fe, Cu, Zn, Cd, Ni, Cr. Một số chỉ tiêu KLN trong nước như Hg, Pb, As có nồng độ quan trắc luôn nằm dưới ngưỡng phát hiện của thiết bị phân tích.

2.2.2. Xử lý dữ liệu

Việc lưu trữ, phân tích và xử lý dữ liệu được thực hiện nhờ phần mềm R kết hợp với Microsoft Excel. Các kỹ thuật thống kê được sử dụng gồm thống kê mô tả, thống kê suy diễn, phân tích tương quan [6 - 7]. Mục đích sử dụng phân tích thống kê là tính toán các thông số như cỡ mẫu (n), trung bình (mean), sd (độ lệch chuẩn), median (trung vị), min (giá trị tối thiểu), max (giá trị tối đa) và xây dựng mô hình hồi quy tuyến tính đa biến.

2.2.3. Mô hình tiên lượng nồng độ kim loại nặng

a) Mô hình hồi quy tuyến tính đa biến

Mô hình hồi quy tuyến tính đa biến (multiple linear regression model – MLR) là mô hình được diễn đạt bằng phương trình:

$$y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \dots + \beta_k x_{ki} + \varepsilon_i$$

Trong phương trình trên, biến cần tiên lượng y_i là biến phụ thuộc, các biến $x_1, x_2, \dots$ đến x_k là các biến độc lập. Mỗi biến độc lập có một thông số β_j ($j = 1, 2, \dots, k$) cần phải ước tính. Phương pháp ước tính β_j dựa vào phương pháp bình phương nhỏ nhất. Gọi $\hat{y} = \hat{\alpha} + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \hat{\beta}_3 x_{3i} +$

$\dots + \hat{\beta}_k x_{ki}$ là ước tính của y_i , phương pháp bình phương nhỏ nhất tìm giá trị $\hat{\alpha}, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \dots, \hat{\beta}_k$ sao cho $\sum_{i=1}^n (y_i - \hat{y}_i)^2$ nhỏ nhất, trong đó $\hat{\alpha}$ và $\hat{\beta}$ (với dấu mũ phía trên) là 2 ước số (estimates) của α và β và $(y_i - \hat{y}_i)$ gọi là phần dư (residual).

Hệ số xác định R^2 và hệ số xác định hiệu chỉnh R^2 adj được sử dụng để đo lường mối liên hệ giữa kết quả và giá trị thực trong thống kê. Ngoài ra, hệ số tương quan r (coefficient of correlation) được sử dụng để đánh giá tính tương quan giữa 2 biến số và đánh giá hiệu quả trong các mô hình tiên lượng. Mô hình MLR đã được ứng dụng trong các nghiên cứu liên quan trên thế giới [8 - 9]. Trong nghiên cứu này, mô hình hồi quy tuyến tính đa biến được ứng dụng để tiên lượng nồng độ của một KLN dựa trên một phương trình quan hệ mà trong đó dữ liệu chất lượng nước của các thông số khác là những biến độc lập. Mô hình sau đó được tối ưu để loại bỏ những biến không cần thiết trong phương trình theo tiêu chuẩn Bayesian [6].

b) Mô hình mạng trí tuệ nhân tạo

Mô hình mạng trí tuệ nhân tạo (hay còn gọi là mạng neuron) (Artificial Neural Network – ANN) là một mô hình toán học được sử dụng nhằm xây dựng phương trình tiên lượng biến y (biến phụ thuộc) dựa vào sự tương quan với các biến x (biến độc lập). Cũng giống như mô hình hồi quy tuyến tính đa biến, mô hình ANN là một mô hình hộp đen (blackbox model) mà trong đó dữ liệu đầu vào và dữ liệu đầu ra được sử dụng để chạy mô hình [10 - 11].

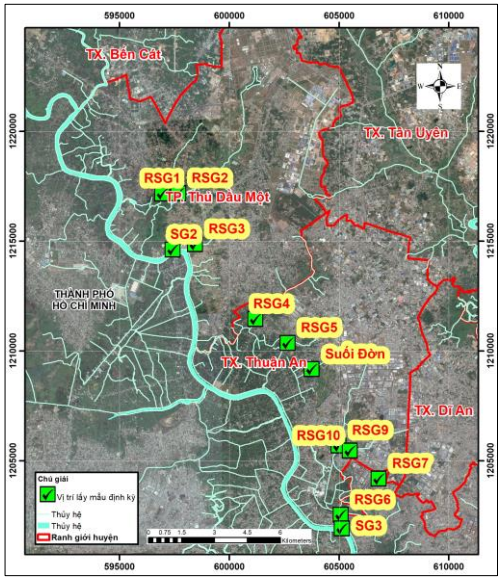
Công cụ NNetool (Neural Fitting Tool) trong phần mềm Matlab được sử dụng để xây dựng mô hình tiên lượng nồng độ KLN trong nước. Nghiên cứu xây dựng một mô hình ANN với 10 lớp ẩn (hidden layer) và thuật toán huấn luyện Levenberg Marquardt để tiên lượng nồng độ của một KLN trong nước. Kết quả mô hình được xuất ra hàm lưu trữ trong file .m của Matlab và có thể được sử dụng cho mục đích tiên lượng trong tương lai.

3. Kết quả nghiên cứu và thảo luận

3.1. Đánh giá hiện trạng ô nhiễm kim loại nặng trong trầm tích

Hiện trạng ô nhiễm trầm tích trên tuyến khảo sát tại TLV hạ lưu sông Sài Gòn được đánh giá theo 2 khu vực đầu tuyến và cuối tuyến đặc trưng cho tính chất nguồn thải, cụ thể: đầu tuyến (vị trí RSG3, SG2) – tiếp nhận nước thải sinh hoạt và cuối tuyến (vị trí RSG6, SG3) – tiếp nhận dòng nước thải sinh hoạt và công nghiệp. Kết quả thống kê dữ liệu cho thấy hàm lượng trung bình các KLN trong trầm tích trong 5 năm (2012 – 2016) đều thấp hơn quy chuẩn cho phép theo QCVN 43:2012/BTNMT – cột trầm tích nước ngọt, ngoài ra hàm lượng trung bình các KLN tại khu vực cuối tuyến khảo sát cao hơn so với đầu tuyến.

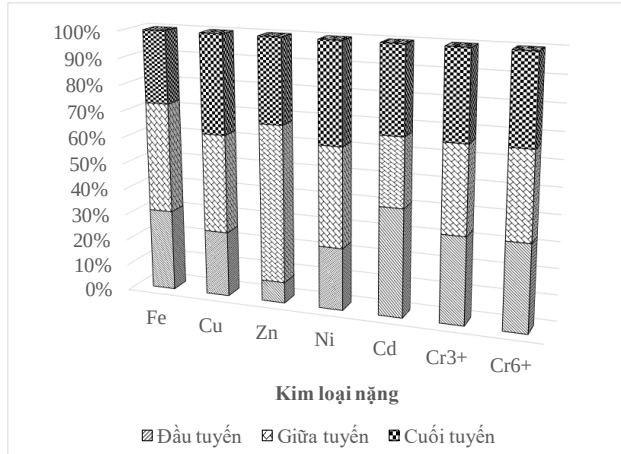
Trong khi đó, giá trị tối đa của hàm lượng Zn trong trầm tích trong 05 năm trên tuyến khảo sát vượt quy chuẩn cho phép theo QCVN 43:2012/BTNMT - cột trầm tích nước ngọt từ 3,8 đến 5,5 lần. Hàm lượng Ni tối đa vượt tiêu chuẩn tham khảo từ 1,7 đến 5,8 lần. Hàm lượng Cr tổng cuối tuyến tối đa vượt quy chuẩn 1,7 lần. Các KLN còn lại đều nằm trong quy chuẩn cho phép. Tính trong năm 2016, các chỉ tiêu KLN trong trầm tích hầu hết đều đạt quy chuẩn ngoại trừ Ni tại SG2 (đợt 1) và RSG6 (đợt 2).



Hình 2. Bản đồ vị trí lấy mẫu định kỳ

3.2. Đánh giá hiện trạng ô nhiễm kim loại nặng trong nước mặt

Hiện trạng ô nhiễm KLN trong nước mặt tại khu vực nghiên cứu cũng được đánh giá trên 3 khu vực đầu tuyến, giữa tuyến và cuối tuyến đặc trưng cho tính chất nguồn thải: đầu tuyến (vị trí RSG1, RSG2, RSG3, SG2) – tiếp nhận nước thải sinh hoạt; giữa tuyến (vị trí RSG4, RSG5, Suối Đồn) và cuối tuyến (vị trí RSG9, RSG10, RSG7, RSG6, SG3) – tiếp nhận nước thải công nghiệp.



Hình 3. Phần trăm nồng độ trung bình của kim loại nặng trong nước mặt trên tuyến khảo sát trong 05 năm 2012 – 2016

Kết quả thống kê dữ liệu nồng độ trung bình các KLN trong nước trên tuyến khảo sát trong 5 năm (2012 – 2016) cho thấy nồng độ Fe và Zn tại khu vực giữa tuyến cao hơn 2 khu vực còn lại. Trong khi đó nồng độ Cu và Ni cao tại khu vực giữa tuyến và cuối tuyến, nồng độ Cr³⁺ và Cr⁶⁺ không thay đổi trên tuyến khảo sát (Hình 3).

Trong khi đó, tại các kênh, rạch, suối là nguồn tiếp nhận trực tiếp của nước thải, nồng độ trung bình các KLN Fe, Zn, Ni, Cd cao hơn trên sông Sài Gòn. Trong đó, nồng độ trung bình của Fe gấp 1,6 – 1,9 lần, Zn gấp 3,8 – 4,7 lần, Ni gấp 1,7 – 2,3 lần và Cd gấp 1,5 lần, tuy nhiên nhìn chung vẫn thấp hơn quy chuẩn cho phép theo QCVN 08-MT:2015/BTNMT, cột B1.

3.3. Đánh giá mối tương quan giữa nồng độ kim loại nặng trong nước và trầm tích

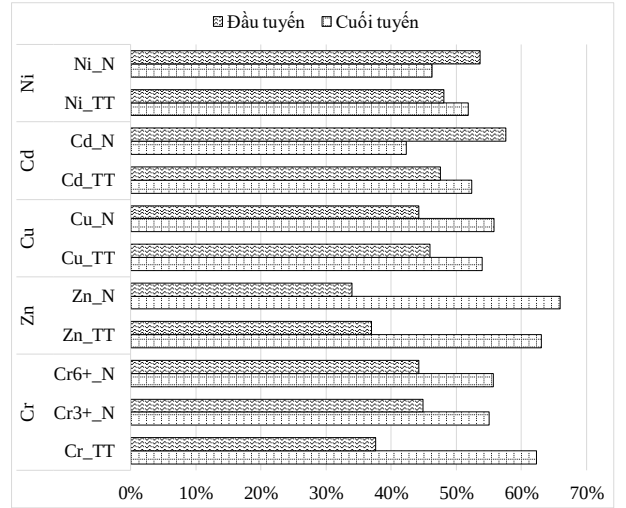
3.3.1. Tại mỗi vị trí lấy mẫu

Trong môi trường nước, KLN tồn tại dưới dạng ion hoặc hấp phụ lên các chất rắn lơ lửng, bùn đáy và tồn tại trong 5 dạng khác nhau [12]. Tại khu vực nghiên cứu, kết quả đánh giá tương quan giữa KLN gồm Cu, Zn, Ni, Cd và Cr trong nước và trầm tích tại mỗi vị trí lấy mẫu trong 5 năm cho thấy sự tương quan yếu với hệ số tương quan Kendall $|\tau| < 0,5$ (Bảng 1). Kết quả này là phù hợp bởi vì dưới tác động của dòng chảy, các chất ô nhiễm sẽ khuếch tán hay di chuyển. Do đó quá trình hấp phụ KLN lên trầm tích hoặc giải phóng KLN từ trầm tích sẽ không tạo nên sự tương quan thuận.

Bảng 1. Hệ số tương quan giữa nồng độ kim loại nặng trong nước và trầm tích tại mỗi vị trí lấy mẫu

Kim loại	Hệ số tương quan Kendall τ					
	SG2 (Cường)	SG2 (Kiệt)	SG3 (Cường)	SG3 (Kiệt)	RSG3	RSG6
Cu	0,16	0,42	0,30	0,34	0,18	-0,24
Zn	0,31	0,20	0,31	0,04	-0,04	0,38
Ni	0,20	-0,28	0,02	-0,07	0,09	0,22
Cd	-0,53	0,00	-0,13	0,13	0,05	-0,02
Cr	0,32	-0,13	0,07	-0,24	0,34	0,16
As	-0,07	-0,18	-0,49	-0,40	-0,05	0,23

3.3.2. Giữa khu vực đầu tuyến và cuối tuyến



Hình 4. Phần trăm nồng độ KLN trong nước và trầm tích trên tuyến khảo sát trong 05 năm 2012 – 2016

Kết quả phân tích tương quan giữa KLN trong nước và trầm tích giữa khu vực đầu tuyến và cuối tuyến trong 05 năm thể hiện qua Hình 4.

Đọc theo tuyến khảo sát, hàm lượng các KLN trong trầm tích gồm Cu, Zn, Ni, Cd, Cr_{tổng} tăng dần từ khu vực đầu tuyến đến cuối tuyến. Trong khi đó, nồng độ KLN trong nước cũng có sự thay đổi, cụ thể nồng độ Cu, Zn, Cr⁶⁺ và Cr³⁺ tăng dần từ khu vực đầu tuyến đến cuối tuyến, tương quan thuận với hàm lượng KLN trong trầm tích. Trong khi đó, nồng độ Ni, Cd trong nước giảm dần từ khu vực đầu tuyến đến cuối tuyến, tương quan nghịch với hàm lượng KLN trong trầm tích.

Kết quả đánh giá cho thấy sự tương quan nghịch của Ni và Cd trong nước và trầm tích cho thấy sự cân bằng vật chất dưới tác động của dòng chảy trên sông Sài Gòn, hay nói cách khác, ảnh hưởng từ nguồn thải KLN trong môi trường là rất thấp, không đáng kể. Trong khi đó, Cu, Zn và Cr, cho thấy sự tương quan thuận, tức là có sự tác động đáng kể của nguồn thải đến chất lượng nước ở khu vực cuối tuyến, đặc biệt KLN.

3.4. Xây dựng mô hình tiên lượng nồng độ kim loại nặng trong nước

3.4.1. Mô hình hồi quy tuyến tính đa biến

Trong nhiều nghiên cứu trước đây [8 - 9], nồng độ một số KLN trong nước có thể được tiên lượng dựa trên một phương trình hồi quy mà các biến là những chỉ tiêu chất lượng nước như nhiệt độ, pH, ... Nghiên cứu này ứng dụng mô hình hồi quy tuyến tính đa biến để xây dựng phương trình tiên lượng nồng độ một KLN tại SG2 và SG3 dựa trên các chỉ tiêu chất lượng nước, chỉ số WQI, lượng mưa (mm) và lưu lượng dòng chảy (m³/s). Tất cả dữ liệu trong phương trình tiên lượng đều được chuẩn hóa bằng hàm log().

Kết quả xây dựng phương trình tiên lượng và tối ưu mô hình theo tiêu chuẩn Bayesian chọn được 10 mô hình hồi quy tuyến tính đa biến tiên lượng KLN có hệ số xác định điều chỉnh $R^2 \text{ adj} \geq 0,7$ thể hiện trong Bảng 2 gồm: Mô hình Cu, Zn, Cd, Cr³⁺, Cr⁶⁺ tại SG2 (Cường); Mô hình Zn tại SG2 (Kiệt); Mô hình Cr⁶⁺ tại SG3 (Cường) và mô hình Fe, Ni, Cr⁶⁺ tại SG3 (Kiệt).

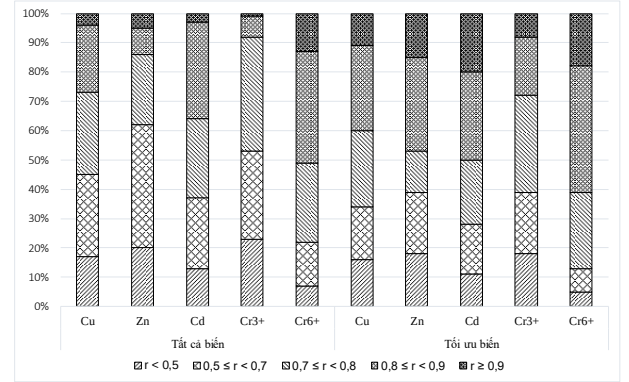
Bảng 2. Hệ số xác định của mô hình hồi quy tuyến tính đa biến sau khi loại bỏ biến KLN

Mô hình tối ưu	SG2 (Cường)				
	Cu	Zn	Cd	Cr ³⁺	Cr ⁶⁺
R ²	0,782	0,831	0,841	0,877	0,810
R ² adj	0,691	0,761	0,754	0,818	0,769
Mô hình tối ưu	SG2 (Kiệt)	SG3 (Cường)	SG3 (Kiệt)		
	Zn	Cr ⁶⁺	Fe	Ni	Cr ⁶⁺
R ²	0,879	0,803	0,819	0,79	0,935
R ² adj	0,804	0,735	0,725	0,69	0,897

3.4.2. Mô hình mạng trí tuệ nhân tạo

Nhằm mục đích so sánh và đánh giá hiệu quả của mô hình hồi quy tuyến tính đa biến, mô hình mạng trí tuệ nhân tạo được ứng dụng để xây dựng phương trình phi tuyến tiên lượng nồng độ KLN trong nước.

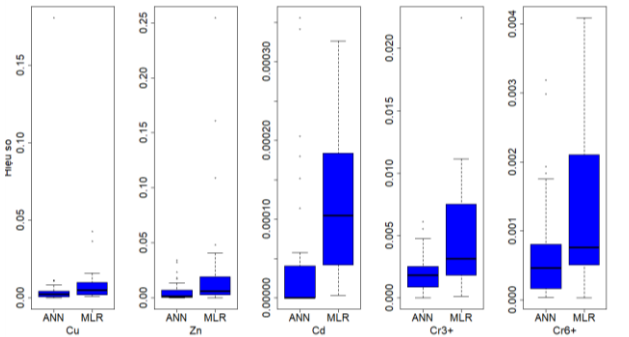
Dữ liệu sử dụng để chạy mô hình ANN là dữ liệu sử dụng để chạy mô hình MLR. Tuy nhiên phương pháp xử lý dữ liệu của mô hình ANN có tính ngẫu nhiên trong việc lựa chọn dữ liệu để huấn luyện, hiệu chỉnh và kiểm định, do đó nghiên cứu thực hiện việc huấn luyện mô hình trong 100 lần để tìm ra một mô hình có hệ số tương quan tổng cao nhất. Bên cạnh đó, kết quả tối ưu biến cho thấy hiệu quả tiên lượng của mô hình ANN tăng lên sau khi loại bỏ những biến không cần thiết theo tiêu chuẩn Bayesian (tiêu chuẩn đã được áp dụng trong mô hình MLR). Hình 5 minh họa cho các mô hình tiên lượng KLN tại SG2 (Cường) trước và sau khi tối ưu biến. Tại SG2 (Kiệt), SG3 (Cường & Kiệt) cũng cho kết quả tương tự.



Hình 5. Hệ số tương quan r trong mô hình ANN tiên lượng nồng độ KLN trong nước tại SG2 (Cường) với tất cả các biến và tối ưu biến theo tiêu chuẩn Bayesian

3.4.3. Kiểm định và đánh giá hai mô hình tiên lượng

Từ kết quả tiên lượng trong mô hình hồi quy tuyến tính MLR và mô hình mạng trí tuệ nhân tạo ANN, tiến hành kiểm định và đánh giá 2 mô hình. Hình 6 là biểu đồ box-plot minh họa hiệu quả tiên lượng nồng độ KLN trong nước tại SG2 (Cường) bằng 2 mô hình ANN và MLR. Theo đó, độ lệch giữa giá trị tiên lượng và nồng độ đo đạc thực tế tại các đợt quan trắc (2012 – 2016) ở mô hình ANN thấp hơn và có phạm vi nhỏ hơn so với mô hình MLR. Kết quả tương tự tại SG2 (Kiệt), SG3 (Cường & Kiệt).



Hình 6. Thống kê độ lệch giữa giá trị tiên lượng và nồng độ quan trắc KLN tại SG2 (Cường) theo 2 mô hình MLR và ANN

Như vậy, mô hình ANN cho thấy khả năng tiên lượng tốt hơn với cùng số biến đầu vào so với mô hình hồi quy tuyến tính. Kết quả này cũng phù hợp với một số nghiên cứu trên thế giới [13].

3.5. Đề xuất mô hình tiên lượng nồng độ kim loại nặng

Trên lý thuyết, các mô hình tiên lượng nồng độ KLN trong nước đã tìm được đều là những mô hình đáp ứng đủ cơ sở khoa học. Tuy nhiên vì những lý do liên quan đến chi phí phân tích các chỉ tiêu KLN trong nước, một mô hình tiên lượng KLN tốt hơn hết không nên có sự phụ thuộc vào giá trị của các KLN khác.

Việc loại bỏ hoàn toàn các biến độc lập là giá trị nồng độ của KLN cần phải được xem xét bởi vì hiệu quả tiên lượng của các mô hình này có thể giảm do một số KLN trong nước có mối tương quan chặt chẽ với nhau. Kết quả loại bỏ các biến độc lập là KLN từ 10 mô hình MLR được lựa chọn cho thấy chỉ có 2 mô hình cho hệ số xác định $R^2 \geq 0,7$ gồm mô hình tiên lượng Zn tại SG2 (Cường) và mô hình tiên lượng Cr⁶⁺ tại SG3 (Kiệt).

Phương trình tiên lượng cho Zn tại SG2 (Cường):

$$[Zn] = -22,9974 + 8,8327.[pH] + 1,1222.[TDS] + 0,6664.[\text{Độ đục}] - 1,6135.[DO] - 0,6475.[COD] - 0,2504.[NO_2-N] + 0,6327.[NH_3-N] - 0,7045.[EC] - 0,2432.[F^-] \quad (1)$$

Và phương trình tiên lượng cho Cr^{6+} tại SG3 (Kiệt) là:

$$[Cr^{6+}] = -12,3760 - 0,2779.[SS] + 0,0577.[TDS] + 0,2576.[\text{Độ đục}] - 0,9904.[DO] - 0,1434.[NH_3-N] + 0,0115.[EC] + 0,1555.[NaCl] - 0,0663.[Cl^-] - 0,0486.[Coliform] - 0,6146.[WQI] + 1,6818.[Q] \quad (2)$$

Các mô hình ANN cũng cho thấy kết quả tương tự khi loại bỏ các biến KLN, xác suất xuất hiện các mô hình tiên lượng có hệ số tương quan $r \geq 0,9$ khá thấp. Trong số các mô hình, có 3 mô hình cho xác suất xuất hiện cao hơn các mô hình còn lại (hệ số r thu được từ 9 – 15%) trong 100 lần huấn luyện (Bảng 3) gồm mô hình tiên lượng Zn tại SG2 (Cường), Zn tại SG2 (Kiệt) và Cr^{6+} tại SG3 (Cường). Đây cũng là 3 mô hình được đề xuất để tiên lượng nồng độ KLN trong nước (được lưu trong các file .m của Matlab).

4. Kết luận

Kết quả đánh giá hiện trạng ô nhiễm KLN trong nước và trầm tích trên các kênh, rạch, sông, suối tiêu lưu vực hạ lưu sông Sài Gòn cho thấy hầu hết các KLN vẫn nằm trong giới hạn cho phép theo QCVN 08-MT:2015/BTNMT, tuy nhiên một số KLN trong nước có dấu hiệu tăng mạnh tại cuối tuyến khảo sát, nhất là tại RSG10 – nơi tiếp nhận nước thải từ Khu Công nghiệp (KCN) Đồng An. Ngoài ra, kết quả phân tích thống kê dữ liệu cũng cho thấy tại mỗi vị trí lấy mẫu, nồng độ kim loại nặng trong nước tương quan yếu với hàm lượng kim loại nặng trong trầm tích. Đồng thời, xét trên tuyến thí nghiệm gia tăng nồng độ Cu, Zn và Cr trong nước được chứng minh là do ảnh hưởng của nguồn thải tại cuối tuyến khảo sát.

Bảng 3. Hệ số tương quan của mô hình ANN sau khi loại bỏ biến KLN

Kim loại	$r < 0,5$	$0,5 \leq r < 0,7$	$0,7 \leq r < 0,8$	$0,8 \leq r < 0,9$	$r \geq 0,9$
I. SG2 (Cường)					
Cu	34%	29%	19%	17%	1%
Zn	25%	16%	27%	20%	12%
Cd	14%	35%	27%	22%	2%
Cr^{3+}	32%	31%	20%	14%	3%
Cr^{6+}	39%	56%	5%	0%	0%
II. SG2 (Kiệt)					
Zn	36%	33%	12%	10%	9%
III. SG3 (Cường)					
Cr^{6+}	19%	29%	32%	18%	2%
IV. SG3 (Kiệt)					
Fe	40%	27%	21%	11%	1%
Ni	34%	39%	14%	12%	1%
Cr^{6+}	8%	21%	16%	40%	15%

Kết quả xây dựng mô hình hồi quy tuyến tính đa biến với tiêu chuẩn Bayesian đã tìm ra được mô hình tiên lượng Zn

tại SG2 (Cường) và mô hình tiên lượng Cr^{6+} tại SG3 (Kiệt) cho hệ số $R^2 \geq 0,7$. Đồng thời, mô hình ANN cũng được thử nghiệm nhằm đánh giá, so sánh hiệu quả 2 mô hình tiên lượng. Kết quả đã chỉ ra mô hình ANN có hiệu quả tiên lượng tốt hơn so với mô hình hồi quy tuyến tính có cùng số biến lựa chọn. Ngoài ra, mô hình ANN còn tìm được thêm 1 mô hình tiên lượng Zn tại SG3 (Cường) với xác suất xuất hiện mô hình có hệ số tương quan $r \geq 0,9$ đạt khoảng 9%.

Vì sự thiếu sót dữ liệu chất lượng nguồn thải công nghiệp và sinh hoạt, các mô hình tiên lượng đưa ra vẫn còn hạn chế, tính hiệu quả thực tiễn của mô hình chưa cao. Vì vậy, nghiên cứu nhằm mục đích bước đầu đánh giá và xây dựng mô hình tiên lượng nồng độ KLN trong môi trường dựa trên các phương pháp mới và phổ biến, làm cơ sở tham khảo cho các nghiên cứu tiếp theo trong tương lai.

TÀI LIỆU THAM KHẢO

- [1] Bộ Tài nguyên và Môi trường, *Thông tư số 06/2013/TT-BTNMT ngày 07/05/2013 của Bộ Tài nguyên và Môi trường ban hành Danh mục lĩnh vực, ngành sản xuất, chế biến có nước thải chứa kim loại nặng phục vụ tính phí bảo vệ môi trường đối với nước thải*, Hà Nội, 2013.
- [2] C. Garneau, S. Sauvage, A. Probst, and J. M. Sánchez-Pérez, "Modelling of Trace Metal Transfer in A Large River Under Different Hydrological Conditions (The Garonne River in Southwest France)", *Ecological Modelling*, Vol. 306, 2015, pp. 195-204.
- [3] M. L. Sámano, A. García, J. A. Revilla, and C. Álvarez, "Modeling Heavy Metal Concentration Distributions in Estuarine Waters: An Application to Suances Estuary (Northern Spain)", *Environmental Earth Sciences*, Vol. 72, No. 8, 2014, pp. 2931-2945.
- [4] Đặng Vũ Bích Hạnh, "Tóm tắt kết quả khảo sát kim loại vết trong nước và bùn trầm tích lưu vực sông Sài Gòn xác định mối tương quan giữa nước và bùn trầm", *Tạp chí Tài nguyên và Môi trường*, Số 17, 2015, trang 25-26.
- [5] Sở Khoa học và Công nghệ Thành phố Hồ Chí Minh, *Đánh giá mức độ khan hiếm tài nguyên nước ngọt cho TP. HCM bằng chỉ số áp lực về nước WSI theo các kịch bản quy hoạch phát triển đến năm 2030 trong điều kiện biến đổi khí hậu khi nước biển dâng và đề xuất các giải pháp tổng thể giảm thiểu*, Tp. Hồ Chí Minh, 2016.
- [6] Nguyễn Văn Tuấn, *Phân tích dữ liệu với R*, NXB Tổng hợp Thành phố Hồ Chí Minh, 2016.
- [7] Chế Đình Lý, *Thống kê và xử lý dữ liệu môi trường*, NXB Đại học Quốc gia Thành phố Hồ Chí Minh, 2012.
- [8] M. C. Faier, G. A. Dumitrel, and D. Perju, "Experimental Modeling of Heavy Metals Concentration Distribution in Rivers", *Chemical Bulletin of "Politehnica" University of Timișoara, România*, Vol. 56(70), No. 1, 2011, pp. 38-42.
- [9] T. M. Akpomie, E. O. Ekanem, M. M. Adamu, and J. O. Akpomie, "Computer Modelling of the Concentration of Heavy Metals in Artificial Borings", *World Journal of Analytical Chemistry*, Vol. 4, No. 1, 2016, pp. 6-10.
- [10] A. S. Abdolmaleki, A. G. Ahangar, and J. Soltani, "Artificial Neural Network (ANN) Approach for Predicting Cu Concentration in Drinking Water of Chahnimah1 Reservoir in Sistan-Balochistan, Iran", *Health Scope*, Vol. 2, No. 1, 2013, pp. 31-38.
- [11] A. Sarkar and P. Pandey, "River Water Quality Modelling Using Artificial Neural Network Technique", *Aquatic Procedia*, Vol. 4, 2015, pp. 1070-1077.
- [12] A. Tessier, P. G. C. Campbell, and M. Bisson, "Sequential Extraction Procedure for The Speciation of Particulate Trace Metals", *Analytical Chemistry*, Vol. 51, No. 7, 1979, pp. 844-851.
- [13] A. Abdallaoui and H. E. Badaoui, "Comparative Study of Two Stochastic Models Using The Physicochemical Characteristics of River Sediment to Predict The Concentration of Toxic Metals", *Journal of Materials and Environmental Science*, Vol. 6, No. 2, 2015, pp. 445-454.