

## Application of correlative equations to “filling up” the water quality monitoring data for assessment of the quality and self-cleaning capacity of water

Hung V. Bui<sup>1\*</sup>, & Diep N. Nguyen<sup>2</sup>

<sup>1</sup>Faculty of Environment, University of Science, Vietnam National University Ho Chi Minh City, Vietnam

<sup>2</sup>University of Labour and Social Affairs (Campus II), Ho Chi Minh City, Vietnam

### ARTICLE INFO

#### Research Paper

Received: August 24, 2022

Revised: December 05, 2022

Accepted: February 15, 2023

#### Keywords

Highest order polynomial empirical curve equation (HoCEq)

Nhieu Loc Thi Nghe canal

Self-cleaning capacity

The multivariable regressive correlative equation (MREq)

Water quality

#### \*Corresponding author

Bui Viet Hung

Email: bvhung@hcmus.edu.vn

### ABSTRACT

In environmental management, continuous water quality monitoring is essential to provide information on the status, distribution, and trends of water quality. However, monitoring work may not be conducted continuously for various reasons, resulting in a lack of data or discontinuity in data sets. The lack of data and the discontinuity of the monitoring data set can result in the heterogeneity or weak representativeness of the analysis/assessment results regarding the level quality or self-cleaning capacity of water. High-order polynomial empirical curve equation (HoCEq) and multivariable regression correlative equation (MREq) are commonly used interpolation/simulation methods because they are integrated in the office analysis tools like Excel or SPSS and give suitable results. In the study, the assessment of water quality and self-cleaning capacity of Nhieu Loc Thi Nghe canal in Ho Chi Minh City, the HoCEq and MREq were applied to “fill up” monitoring data sets for the period 2012 - 2021. This approach helps to increase efficiency in the analysis/assessment and increases the representativeness of research results with an appropriate square correlation coefficient ( $R^2$  larger than 0.5) and corresponding degree of close correlation.

**Cited as:** Bui, H. V., & Nguyen, D. N. (2023). Application of correlative equations to “filling up” the water quality monitoring data for assessment of the quality and self-cleaning capacity of water. *The Journal of Agriculture and Development* 22(1), 41-53.

## Ứng dụng phương trình tương quan "lấp đầy" dữ liệu quan trắc chất lượng nước phục vụ đánh giá chất lượng và khả năng tự làm sạch của nước

Bùi Việt Hưng<sup>1\*</sup> & Nguyễn Ngọc Diệp<sup>2</sup>

<sup>1</sup>Khoa Môi Trường, Trường Đại Học Khoa Học Tự Nhiên, Đại Học Quốc Gia TP. Hồ Chí Minh, TP. Hồ Chí Minh

<sup>2</sup>Trường Đại Học Lao Động - Xã Hội (Cơ sở II), TP. Hồ Chí Minh

### THÔNG TIN BÀI BÁO

#### Bài báo khoa học

Ngày nhận: 24/08/2022

Ngày chỉnh sửa: 05/12/2022

Ngày chấp nhận: 15/02/2023

#### Từ khóa

Chất lượng nước

Kênh Nhiêu Lộc Thị Nghè (NLTN)

Khả năng tự làm sạch của nước

Phương trình đường cong thực nghiệm đa thức bậc cao (HoCEq)

Phương trình tương quan hồi quy đa biến (MREp)

#### \*Tác giả liên hệ

Bùi Việt Hưng

Email: bvhung@hcmus.edu.vn

### TÓM TẮT

Trong công tác quản lý môi trường, quan trắc chất lượng nước được thực hiện liên tục nhằm cung cấp thông tin về hiện trạng chất lượng nước, phân bố và xu hướng diễn ra. Tuy nhiên, do nhiều nguyên nhân khác nhau, công tác quan trắc không diễn ra liên tục, gây gián đoạn / thiếu số liệu cho bộ dữ liệu quan trắc. Việc thiếu số liệu và không liên tục của bộ dữ liệu quan trắc gây ra sự không đồng nhất hoặc làm yếu tính đại diện của các kết quả phân tích / đánh giá về mức độ chất lượng hoặc khả năng tự làm sạch của nước. Phương trình đường cong thực nghiệm đa thức bậc cao (HoCEq) và phương trình hồi quy đa biến (MREq) là một số những phương pháp nội suy / mô phỏng thường được sử dụng do được tích hợp trong bộ công cụ phân tích văn phòng như excel hay SPSS (Statistical Product and Services Solutions) và chúng cho kết quả phù hợp. Trong nghiên cứu đánh giá chất lượng nước kênh Nhiêu Lộc Thị Nghè (NLTN), TP. Hồ Chí Minh (TP.HCM), HoCEq và MREq được áp dụng cho "lấp đầy" cho bộ dữ liệu quan trắc chất lượng nước giai đoạn 2012 - 2021 (nội suy giữa các vị trí và giữa các thời điểm quan trắc). Điều này đã giúp tăng hiệu quả việc phân tích / đánh giá và tăng tính đại diện của kết quả nghiên cứu với hệ số tương quan thích hợp ( $R^2$  trên 0,5), tương ứng mức tương quan chặt.

### 1. Đặt Vấn Đề

Công tác quan trắc môi trường được coi là một công tác quan trọng trong quản lý môi trường, công tác này đưa ra hiện trạng chất lượng môi trường với các chỉ tiêu chất lượng cụ thể. Bộ số liệu, thông tin về chất lượng môi trường, trong đó có môi trường nước, là cơ sở dữ liệu đầu vào quan trọng cho hầu hết các nghiên cứu đánh giá, quy hoạch hay các giải pháp ứng phó, khắc phục và cải thiện môi trường. Quan trắc chất lượng môi trường thường đòi hỏi chi phí và nhân lực lớn. Bộ số liệu quan trắc thường không liên tục hoặc bị "hổng" do nhiều nguyên nhân thực tiễn khác nhau (Le, 2019). Hậu quả của những vấn đề trên

gây ra cho bộ số liệu quan trắc chất lượng môi trường, đặc biệt là môi trường nước, là sự không đồng bộ và thiếu hụt số liệu quan trắc cả về số lượng chỉ tiêu lẫn vị trí quan trắc... (Le, 2019).

Việc mô phỏng các thông số chất lượng nước chủ yếu dựa vào các mô hình toán và có thể chia các mô hình này thành hai nhóm (Le, 2019): (1) Mô hình kết hợp toán - vật lý - hóa học mô phỏng theo tiến trình (process - based models) là loại mô hình mô phỏng các quy luật thủy văn, thủy lực, lan truyền chất và phản ứng hóa học, hóa sinh đã được mô tả và ước tính chính xác theo các phương trình toán hóa các quy luật vật lý, hóa học và sinh học (nhóm mô hình toán hóa). Ưu điểm của các nhóm mô hình toán hóa là tính chính xác rất

cao trong việc thể hiện xu thế, diễn tiến các quy luật; tuy nhiên, nhóm này thường đòi hỏi phải có các số liệu nền lớn, liên tục (số liệu đo đạc) với khoảng thời gian đủ lớn. Do đó, số lượng dữ liệu đầu vào (biên và điều chỉnh) quyết định đến mức độ chính xác (Lin, 2019; Ross, 2019). (2) Mô hình nội suy theo tương quan thường được gọi là “mô hình trí tuệ nhân tạo”. Dạng mô hình này sử dụng các thuật toán học nội suy khác nhau nhằm tìm/thiết lập tương quan phù hợp giữa các trường thông tin, ví dụ K - Nearest Neighbors, tương quan đa biến phi tuyến, hồi quy phi tuyến, Decision Tree, Random Forest, Support Vector Machine, Artificial Neural Network, và Long Short - Term Memory (LSTM). Dạng mô hình này có ưu điểm là dễ áp dụng, độ chính xác cao, không đòi hỏi số lượng dữ liệu lớn liên tục. Ngoài ra, nhóm mô hình này thường chuẩn hóa nhiều thông tin khác nhau, trong đó có các thông tin liên quan đến kinh tế, xã hội và nhận thức của cộng đồng khu vực. Tuy nhiên, do các mô hình dạng này thuộc nhóm “black - box” nên đôi khi rất khó để giải thích kết quả (Palani, 2008; Hunter, 2018; Ross, 2019) dù bản chất của các mô hình đều dựa trên cơ sở lựa chọn một hay nhiều hàm hay phương trình phân bố của một hay nhiều biến số có mối tương quan khác nhau. Việc lựa chọn nhóm mô hình nào phụ thuộc nhiều yếu tố như từ đầu vào (khối lượng thông tin, dữ liệu - số liệu cho các điều kiện biên đến hiệu chỉnh mô hình, phần mềm. . .) đến mục tiêu đạt được về mức độ chính xác. Theo quan điểm của Harvey (1990), các tiêu chuẩn để đánh giá một mô hình tốt là:

- Tính tiết kiệm (parsimony): mô hình càng đơn giản (nhưng phải chứa biến chính ảnh hưởng đến biến phụ thuộc) càng tốt.

- Tính đồng nhất (identifiability): với mỗi tập dữ liệu đã cho thì các tham số ước lượng được phải có giá trị thống nhất.

- Tính thích hợp (goodness of fit): Mục đích của phân tích hồi quy là giải thích sự biến động của biến phụ thuộc bằng các biến giải thích của mô hình. Mô hình càng thích hợp nếu các biến giải thích càng giải thích được nhiều sự thay đổi của biến phụ thuộc, tức là hệ số  $R^2$  hoặc  $\bar{R}^2$  càng lớn càng tốt.

- Tính vững về mặt lý thuyết (theoretical consistency): mô hình phải phù hợp với cơ sở lý thuyết nền tảng của lĩnh vực đang xét. Nếu có hệ số xác định cao nhưng dấu của hệ số hồi quy sai thì mô hình không thể được đánh giá là tốt.

- Khả năng dự báo tốt (predictive power): mô

hình có khả năng dự báo càng chính xác, càng phù hợp với thực tế càng tốt.

Với các tiêu chuẩn lựa chọn mô hình toán nhóm 1 và nhóm 2, hầu hết các mô hình toán đang được áp dụng các nghiên cứu hiện nay đều đạt được một hay nhiều tiêu chuẩn trên (Hunter, 2018; Le, 2019). Điều có nghĩa là các mô hình toán hiện nay đều là những công cụ hữu ích phân tích diễn biến các yếu tố vật lý, hóa học, sinh học. Tuy nhiên, trong các trường hợp bộ số liệu hạn chế về mặt khối lượng (phạm vi không gian và thời gian), các mô hình nhóm (2) thường được sử dụng do loại mô hình này không đòi hỏi bộ số liệu lớn.

Như phân tích ở trên nhóm (2), các mô hình dựa trên nền tảng các hàm phân tích thống kê theo hướng nội suy, tối ưu. Hàm hồi quy đa biến (MREq) cũng như hàm phân bố theo phương trình đường cong thực nghiệm đa thức bậc cao (HoCEq) là những dạng thường được lựa chọn dùng trong các bài toán phân tích thống kê, nội suy. Hai trong phương trình hồi quy này thường được dùng do chúng được tích hợp trong nhiều công cụ tổng hợp, lưu trữ và phân tích rất phổ biến như Microsoft Excel hay SPSS. Do vậy, việc sử dụng chúng khá giản đơn và có cơ sở lý thuyết về thực hiện và kiểm tra mức độ phù hợp khá cụ thể và rõ ràng. Trong nghiên cứu, chúng tôi đã sử dụng hai phương trình (HoCEq) và (MREq) cho việc “nội suy” nhằm “lấp” các khoảng trống trong bộ số liệu quan trắc chất lượng nước kênh rạch nội đô - kênh Nhiều Lộc Thị Nghè tại thành phố Hồ Chí Minh giai đoạn 2012 - 2021.

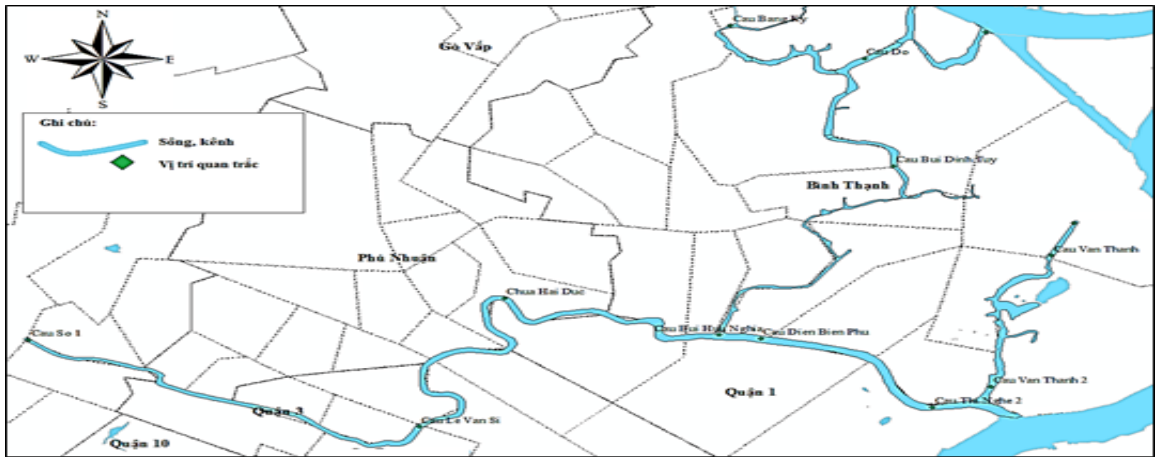
## 2. Vật Liệu và Phương Pháp Nghiên Cứu

### 2.1. Khu vực nghiên cứu

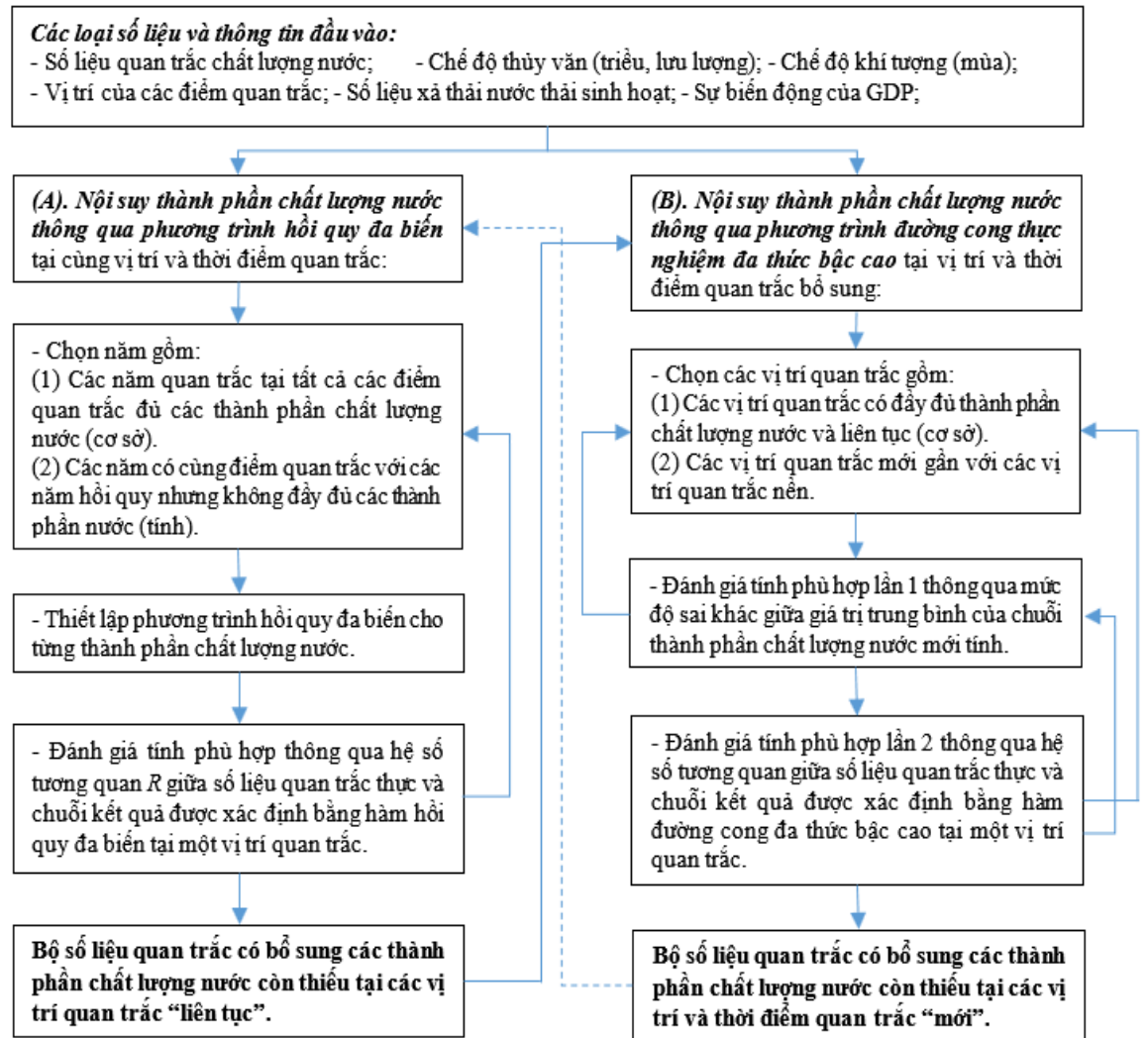
Khu vực nghiên cứu là kênh nội đô Như Lộc Thị Nghè (NLTN) trên địa bàn thành phố Hồ Chí Minh. Kênh NLTN và các điểm quan trắc xem Hình 1. Bộ số liệu sử dụng trong nghiên cứu là kết quả quan trắc chất lượng nước kênh Nhiều Lộc Thị Nghè của Trung tâm Quan trắc Môi trường thuộc sở Tài nguyên và Môi trường TP.HCM giai đoạn 2012 - 2021. Số liệu quan trắc của bộ số liệu này - đối tượng nghiên cứu - không “đồng bộ” và không “liên tục” giữa các năm và giữa các vị trí.

### 2.2. Tiếp cận nghiên cứu

Tiếp cận nghiên cứu “lấp đầy” số liệu các thành phần quan trắc chất lượng nước thiếu hoặc không thực hiện của nguồn nước kênh rạch nội đô thành



Hình 1. Kênh Nhiều Lọc Thị Nghè và các vị trí quan trắc chất lượng nước kênh.



Hình 2. Tiếp cận trong nghiên cứu “lấp đầy” bộ số liệu quan trắc chất lượng nước.

phố Hồ Chí Minh giai đoạn 2012 - 2021, như Hình 2.

Nhằm đa dạng thông tin và dữ liệu có thể ảnh hưởng / tác động đến chất lượng môi trường nước, nghiên cứu đã xác định một số yếu tố chính đưa vào phân tích như:

- Bộ số liệu quan trắc của các thành phần nước tại các điểm quan trắc trên hệ thống kênh nội đô thành phố được thực hiện bởi Trung tâm Quan trắc Môi trường thành phố Hồ Chí Minh giai đoạn 2012 - 2021

- Thông tin về dân số, lưu lượng xả thải các khu vực phường xã trên địa bàn thành phố Hồ Chí Minh nằm trên các vùng thoát nước kế cận tuyến kênh

- Thông tin về phát triển kinh tế của khu vực quận huyện có tuyến kênh chảy qua (tuy nhiên, trong nghiên cứu chỉ sử dụng số liệu phát triển kinh tế (GDP) hàng năm của thành phố).

- Kết quả mô phỏng chế độ dòng chảy (Q - lưu lượng) trong hệ thống kênh rạch trong giai đoạn 2012 - 2021.

- Các thông tin về phát triển kinh tế, xã hội, chế độ thủy văn, khí hậu được chuẩn hóa trước khi đưa vào phân tích.

Dựa trên cách tiếp cận nghiên cứu, các bước chính thực hiện đề tài mô tả bên dưới.

### 2.3. Lựa chọn các năm và vị trí quan trắc cơ sở

Xác định các vị trí cơ sở và cần tính toán bổ sung thành phần chất lượng nước và những năm đầy đủ điểm đo và thành phần chất lượng nước kênh NLTN.

Căn cứ vào chuỗi số liệu quan trắc chất lượng nước tại các điểm khảo sát trên kênh NLTN của Trung tâm Quan trắc môi trường TP.HCM giai đoạn 2012-2021, năm thực hiện quan trắc tại các điểm khảo sát được tổng hợp trong Bảng 1.

Các vị trí được đo đạc liên tục nhiều nhất trong giai đoạn 2012 - 2021 gồm cầu Số 1, cầu Lê Văn Sĩ, chùa Hải Đức, cầu Điện Biên Phủ và cầu Thị Nghè 2 với các năm quan trắc từ 2014 - 2019 & 2021.

Xác định năm quan trắc cần bổ sung giá trị thành phần nước theo các thời điểm quan trắc. Ví dụ như tại vị trí quan trắc cầu Điện Biên Phủ, các điểm quan trắc được hiện đo đạc theo các thời điểm trong năm (Bảng 2).

**Bảng 1.** Thống kê các vị trí được quan trắc trong giai đoạn 2012 - 2021

Năm	Cầu Số 1		Cầu Lê Văn Sĩ		Chùa Hải Đức		Bùi Hữu Nghĩa		Điện Biên Phủ		Thị Nghè 2		Văn Thánh		Cầu Thánh 2		Cầu Đỗ	
	Vị trí	Số 1	Lê Văn Sĩ	Hải Đức	Chùa	Bùi Hữu Nghĩa	Điện Biên Phủ	Thị Nghè 2	Văn Thánh	Cầu Thánh 2	Cầu Đỗ							
2012		x					x											
2013		x					x											
2014		x	x				x											
2015		x	x	x			x											
2016		x	x	x			x											
2017		x	x	x			x											
2018		x	x	x			x											x
2019		x	x	x			x											x
2020																		
2021		x		x			x											x

**Bảng 2.** Thống kê thời điểm quan trắc tại vị trí cầu Điện Biên Phủ giai đoạn 2012 - 2021

Năm	Tháng											
	1	2	3	4	5	6	7	8	9	10	11	12
2012		x		x					x		x	
2013		x		x					x		x	
2014	x	x	x		x		x		x		x	
2015	x	x	x	x	x	x	x	x	x	x	x	x
2016				x	x	x	x	x	x	x	x	x
2017	x	x	x	x	x	x	x	x	x	x	x	x
2018	x	x	x	x	x				x	x	x	x
2019							x	x	x	x	x	x
2020	x	x	x	x	x							
2021	x	x	x	x	x							

Nguồn: Trung tâm Quan trắc Môi trường, Sở Tài nguyên và Môi trường TP.HCM.

Như vậy, từ các kết quả tổng hợp trong Bảng 1 và Bảng 2 trên, nghiên cứu đã chọn các vị trí quan trắc và năm cơ sở cho kênh NLTN như sau:

- Các vị trí quan trắc cơ sở: cầu Lê Văn Sĩ; cầu Điện Biên Phủ

- Các vị trí cần bổ sung theo nội suy hàm tương quan hồi quy đa biến: (1) dòng chính: cầu số 1, chùa Hải Đức, cầu Thị Nghè 2; (2) dòng nhánh: cầu Bùi Hữu Nghĩa, cầu Văn Thánh, cầu Văn Thánh 2, cầu Sơn và các cửa rạch

- Các năm quan trắc cơ sở và các năm quan trắc cần bổ sung thông qua nội suy tương quan hồi quy đa biến: 2014 - 2019

- Các năm quan trắc cơ sở cho lập phương trình đường cong thực nghiệm đa thức bậc cao: 2015 & 2017. Các năm quan trắc còn lại được bổ sung thông qua tính toán bằng phương trình đường cong thực nghiệm đa thức bậc cao trên.

#### 2.4. Thiết lập các phương trình (MREq) và (HoCEq)

- Phương trình tương quan hồi quy đa biến áp dụng trong nghiên cứu:

$$Y_j = f\left(\sum_{i=1}^n y_i, \text{mùa, triều, } Q_{\text{kênh}}, Q_{\text{xã thái}}, TT_{\text{kinh tế}}, TT_{\text{xã hội}}\right)$$

Trong đó:

$Y_j$  – giá trị thành phần chất lượng nước cần tính toán tại 1 vị trí quan trắc và tại 1 thời điểm xác định trong năm thứ (j) – đơn vị tương ứng với thành phần chất lượng nước như (mg/L) hoặc (MPN/100 mL)

$y_i$  – giá trị thành phần chất lượng nước đã được quan trắc trong năm thứ (i) tại cùng vị trí và thời điểm của năm (j)

Mùa – là thông tin mùa (1-mùa khô, 2-mùa mưa)

Triều – là thông tin về chế độ triều (1-lớn, 2-ròng)

$Q_{\text{kênh}}$  – là lưu lượng trung bình tại vị trí quan trắc ( $\text{m}^3/\text{s}$ )

$Q_{\text{xã thái}}$  – là lưu lượng nước thải xả ra gần vị trí quan trắc ( $\text{m}^3/\text{ngày đêm}$ )

$TT_{\text{kinh tế}}$  – thông tin kinh tế khu vực (sử dụng giá trị GDP khu vực)

$TT_{\text{xã hội}}$  – thông tin xã hội khu vực (xung quanh vị trí quan trắc) thông qua biến động dân số khu vực

Để đánh giá sự phù hợp, các thành phần trong phương trình hồi quy được xem xét thông qua giá trị hệ số của thành phần đó. Nếu giá trị hệ số hồi quy thành phần nhỏ hơn 1% thì được xem xét không tính đến, do mức độ “ảnh hưởng” của thành phần đó đến giá trị thành phần chất lượng nước cần tính nhỏ.

Nghiên cứu sử dụng 1 phần chuỗi đầy đủ các thành phần chất lượng nước quan trắc cho phân tích tương quan với kết quả tính “nội suy lại” thông qua phương trình hồi quy trên tại các vị trí cơ sở. Phương trình hồi quy đa biến cho nội suy được sử dụng khi kết quả hệ số tương quan bình phương (R<sup>2</sup>) giữa chuỗi số liệu quan trắc và “tính lại” phù hợp (trên 0,5).

- Phương trình đường cong thực nghiệm đa thức bậc cao áp dụng trong nghiên cứu:

$$Y_{i,y} = A_{i,o} + \sum_{k=1}^9 A_{i,k} T^k$$

Trong đó:

$Y_{i,j}$  – giá trị thành phần nước cần tính tại 1 vị trí quan trắc và tại thời điểm (j) trong năm thứ (i); T – thời điểm quan trắc trong năm;  $A_{i,o}$ ,  $A_{i,k}$  – là các hệ số của hàm đường cong thực nghiệm đa thức bậc cao cho năm thứ (i). Các hệ số này được tính theo phương pháp nội suy IDW (Inverse distance weight) từ các hệ số tương ứng được xác định thông qua chuỗi số liệu của chính loại thành phần chất lượng nước cần tính được quan trắc tại cùng 1 vị trí quan trắc của các năm thứ (i-1; i+1...).

Đánh giá sự phù hợp của các giá trị thành phần chất lượng thông qua tính toán phương trình đường cong thực nghiệm đa thức bậc cao. Nghiên cứu thực hiện 2 bước với hướng bước 2 là tính ngược lại bước 1 cho năm cơ sở.

Phương trình đường cong thực nghiệm đa thức bậc cao được cho là phù hợp khi kết quả tính giá trị quan trắc và tính “lại” thành phần chất nước trong năm cơ sở có (1) mức độ sai khác giữa tiêu chí “trung bình” dưới 10% so với giá trị trung bình của chuỗi quan trắc; (2) hệ số tương quan bình phương ( $R^2$ ) giữa chúng cao (trên 0,5).

**2.5. Đánh giá khả năng tự làm sạch**

Đánh giá khả năng tự làm sạch (KNTLS) của nước kênh NLTN giai đoạn 2012 - 2021 là một nhiệm vụ của đề tài nghiên cứu và sử dụng bộ số liệu sau khi “lấp đầy”. Phương trình Streeter-Phelpse được áp dụng cho đánh giá khả năng tự làm sạch trong nghiên cứu:

$$D_t = \frac{k_1 L_0}{k_2 - k_1} (e^{-k_1 t} - e^{-k_2 t}) + \frac{k_2 N_0}{k_3 - k_1} (e^{-k_3 t} - e^{-k_2 t}) + \frac{SOD}{k_2 H} (1 - e^{-k_2 t}) + D_0 e^{-k_2 t}$$

Trong đó:

$k_1$  – hệ số quá trình phân hủy do oxy hóa chất hữu cơ và được tính

$k_2$  – hệ số quá trình thẩm thấu oxy vào nước và được tính

$k_3$  – hệ số quá trình phân hủy do oxy hóa  $NH_4$  được tính

U và H là vận tốc và chiều sâu dòng chảy

$D_t$  – mức độ suy giảm nồng độ oxy trong dòng nước;  $L_0$ ,  $N_0$ ,  $D_0$ , SOD giá trị ban đầu thành phần  $BOD_5$ ;  $NH_4$ ; Oxy và nhu cầu oxy của bùn cát.

Công thức đánh giá KNTLS được nghiên cứu áp dụng cho kênh NLTN như sau (Le, 2004):

$$f = \frac{k_2}{k_1}$$

Trong đó:

$f$  - tỷ lệ giữa các hệ số quá trình suy giảm và bổ sung thành phần oxy. Mức độ KNTLS của nguồn nước được phân loại như sau: (i) kém ( $f < 2$ ), (ii) trung bình ( $2 < f < 4$ ), (iii) khá ( $4 < f < 10$ ) and (iv) tốt ( $f > 10$ ) (Le, 2004).

**2.6. Thiết lập bản đồ phân vùng khả năng tự làm sạch**

Thiết lập bản đồ phân vùng khả năng tự làm sạch bằng công cụ GIS với phương pháp nội suy IDW. Phương pháp nội suy IDW của công cụ sẽ hỗ trợ xác định giá trị tại một số vị trí điểm giao giữa các nhánh kênh rạch.

**3. Kết Quả và Thảo Luận**

Từ bộ số liệu quan trắc chất lượng nước thực hiện trong năm 2012 và 2015 tại cầu Điện Biên Phủ trên kênh NLTN (Bảng 3) cho thấy một số tháng trong và thông số DO, độ đục không được thực hiện trong năm 2012, nhưng được thực hiện trong năm 2015.

**3.1. Đánh giá mức độ phù hợp của các phương trình nội suy**

Theo phương pháp tiếp cận trong “lấp đầy” số liệu quan trắc chất lượng nước (CLN) kênh NLTN giai đoạn 2012 - 2021, việc tính toán nội suy theo hướng (A) hay (B) đều có kiểm định mức độ phù hợp của tương quan giữa kết quả tính toán và số liệu quan trắc các năm. Qua tính toán, mức độ tương quan bình phương ( $R^2$ ) đa biến của từng thành phần nước được tổng hợp Bảng 4.

Trong Bảng 4 cho thấy hệ số tương quan bình phương  $R^2$  giữa chuỗi số liệu chất lượng nước quan trắc và chuỗi kết quả “tính ngược lại” chất lượng nước tại hai vị trí cơ sở là cầu Lê Văn Sĩ và cầu Điện Biên Phủ trên kênh Nhiêu Lộc Thị Nghè phần lớn có giá trị lớn hơn 0,70 tương ứng với mức độ quan hệ chặt.

Giá trị hệ số tương quan bình phương ( $R^2$ ) của hàm hồi quy đa biến tính nội suy thành phần nước cho các vị trí quan trắc còn lại trên kênh Nhiêu Lộc - Thị Nghè phần lớn có giá trị trên 0,55 tương đương với mức độ quan hệ khá chặt.

**3.2. Nội suy hàm hồi quy đa biến (MREQ)**

Qua phân tích tương quan hồi quy đa biến, phương trình hồi quy đa biến (1) có các biến có

**Bảng 4.** Giá trị hệ số R2 đa biến của các chuỗi kết quả tại các vị trí quan trắc

Thành phần	Vị trí					
	Cầu Số 1	Chùa Hải Đức	Cầu Thị Nghè 2	Cầu Lê Văn Sĩ	Cầu Điện Biên Phủ	
pH	0,60	0,94		0,91		0,94
TSS	0,66	0,87		0,86		0,82
Độ đục	0,80	0,85		0,87		0,85
NH <sub>4</sub>	0,58	0,75		0,72		0,78
PO <sub>4</sub>	0,55	0,27		0,40		0,48
DO	0,64	0,93		0,89		0,91
COD	0,55	0,83		0,80		0,72
BOD <sub>5</sub>	0,49	0,76		0,72		0,68
Coliforms	0,87	0,34		0,29		0,73
<i>E. coli</i>	0,94	0,71		0,70		0,97

pH: chỉ số ion hydro; TSS: tổng thành phần lơ lửng; NH<sub>4</sub>: amoniac; PO<sub>4</sub>: thành phần Phốt pho; DO: lượng oxy hòa tan trong nước; COD: như cầu oxy hóa học; BOD<sub>5</sub>: như cầu oxy sinh hóa; *E. Coli*: *Escherichia coli*.

**Bảng 3.** Số liệu quan trắc chất lượng nước kênh Nhiêu Lộc Thị Nghè tại cầu Điện Biên Phủ

Năm	Thời điểm		T	pH	Độ đục	NH <sub>4</sub>	DO	COD	BOD <sub>5</sub>	Cd	Pb	Cr	Cu	MPN/100ml	Coliform
	Tháng	Triều													
2012	01	Ròng	26,4	6,94	3,28	3,65	5,33	9,97	4,8	0,002	0,01	0,001	0,003	0,004	4600000
	02	Ròng	26,8	6,94		13,14		193,3	100	0,001	0,007	0,003	0,004	4600000	
	04	Ròng	29,9	7,03	8,909	8,909	3,1	14,68	5,81	0,002	0,003	0,006	0,01	4600000	
	09	Ròng	28,6	6,76	10,95	10,95	3,35	9,81	5,6	0,001	0,002	0,002	0,002	240000	
	11	Ròng	29,1	7,04	9,074	9,074	3,8	6,3	2,9	0,001	0,005	0,002	0,009	23000	
	01	Ròng	26,4	6,94	3,28	3,65	5,33	9,97	4,8	0,002	0,01	0,001	0,003	900	
	02	Ròng	26,8	7,02	17,29	3,45	3,1	14,68	5,81	0,004	0,007	0,003	0,015	43000	
	03	Ròng	29	7	3,5	0,52	3,35	9,81	5,6	0,001	0,002	0,002	0,002	240000	
	04	Ròng	29,9	6,98	9,05	1,60	3,8	6,3	2,9	0,001	0,005	0,002	0,009	23000	
	05	Ròng	30,7	7,02	8,95	1,03	4,92	7,6	2,5		0,002			46000	
	2015	06	Ròng	29,2	7,37	8,11	0,89	3,19	13,17	4,24		0,001	0,001	0,001	43000
07		Ròng	28,9	6,84	10,06	0,56	3,1	8,75	6,04	0,002	0,001	0,001	0,001	460000	
08		Ròng	29	6,57	7,53	0,78	3,68	23	15	0,002	0,005	0,03	0,05	150000	
09		Ròng	28,6	6,12	12,33	0,04	1,81	37,76	24,95	0,002	0,005	0,03	0,05	93000	
10		Ròng	28,7	6,79	14,16	3,64	3,97	19	12	0,002	0,005	0,03	0,05	460000	
11		Ròng	29,1	5,83	9,31	2,12	3,79	22	15	0,002	0,005	0,03	0,05	43000	
12		Ròng	28,6	6,75	8,21	0,08	4,41	25	17	0,002	0,005	0,03	0,05	110000	

NTU: đơn vị đo độ đục; MPN: chỉ số sử dụng để ước tính nồng độ vi sinh vật khả thi; T: nhiệt độ; pH: chỉ số ion hydro; NH<sub>4</sub>: amoniac; PO<sub>4</sub>: thành phần Phốt pho; DO: lượng oxy hòa tan trong nước; COD: như cầu oxy hóa học; BOD<sub>5</sub>: như cầu oxy sinh hóa; Cd: cadmium; Pb: chì; Cr: crom; Cu: crom; *E. Coli*: *Escherichia coli*.  
 Nguồn: Trung tâm Quan trắc Môi trường, Sở Tài nguyên và Môi trường TP.HCM.



**Bảng 5.** Tổng hợp các hệ số hàm hồi quy đa biến cho nội suy thành phần chất lượng nước tại vị trí quan trắc chùa Hải Đức

Thành phần CLN	A <sub>triều</sub>	A <sub>mùa</sub>	A <sub>Lê Văn Sĩ</sub>	A <sub>Điện Biên Phủ</sub>	A <sub>o</sub>
pH	-0,01	0,06	0,56	0,42	0,04
TSS	-0,87	-4,86	0,38	0,58	9,01
Độ Đục	2,18	2,94	0,60	0,38	-6,98
NH <sub>4</sub>	-0,59	-0,99	0,61	0,39	2,58
PO <sub>4</sub>	-0,01	0,06	0,56	-0,03	-0,02
DO	0,005	0,06	0,51	0,42	0,10
COD	0,41	-0,71	0,53	0,58	-2,74
BOD <sub>5</sub>	0,44	-0,92	0,54	0,55	-0,79
Coliform	-7095,39	55824,10	0,14	0,12	-7058,90
<i>E. coli</i>	7364,43	-3922,20	0,18	0,47	-4378,61

CLN: chất lượng nước, pH: chỉ số ion hydro, TSS: tổng thành phần lơ lửng, NH<sub>4</sub>: amoniac, PO<sub>4</sub>: thành phần Phốt pho, DO: lượng oxy hòa tan trong nước, COD: nhu cầu oxy hóa học, BOD<sub>5</sub>: nhu cầu oxy sinh hóa, *E. Coli*: *Escherichia coli*.

nghĩa (hệ số của các biến lớn hơn 0,01) gồm thành phần chất lượng nước (CLN) tại các vị trí cơ sở (cầu Lê Văn Sĩ và cầu Điện Biên Phủ), thông tin về triều (triều lớn, triều ròng) và mùa (mưa, khô).

Hàm hồi quy đa biến cho nội suy thành phần CLN có dạng như sau:

$$Y_j = A_{\text{triều}} \text{Tr} + A_{\text{mùa}} \text{Mùa} + A_{\text{Lê Văn Sĩ}} Y_{\text{Lê Văn Sĩ}} + A_{\text{Điện Biên Phủ}} Y_{\text{Điện Biên Phủ}} + A_o$$

Tổng hợp các hệ số của hàm hồi quy đa biến dung nội suy các thành phần CLN tại các vị trí quan trắc cơ sở được thể hiện trong bảng dưới đây (ví dụ cho vị trí chùa Hải Đức).

Bảng 5 cho thấy một số yếu tố trong phương trình hồi quy đa biến không xuất hiện. Lý do chính là giá trị của các hệ số (mức độ ảnh hưởng của chúng đến giá trị thành phần nước) tính toán nhỏ nên có thể bỏ qua như yếu tố về biến động dân khu vực, GDP hay thành phần lưu lượng dòng chảy, lưu lượng xả thải.

**3.3. Nội suy đường cong thực nghiệm đa thức bậc cao (HoCEq)**

Phương trình HoCEq (9 bậc) được áp dụng cho xác định / nội suy thành phần CLN tại 1 vị trí quan trắc trong năm quan trắc có dạng như sau:

$$Y_{i,j} = A_{o,j} + A_{1,j} \text{Th} + A_{2,j} \text{Th}^2 + A_{3,j} \text{Th}^3 + A_{4,j} \text{Th}^4 + A_{5,j} \text{Th}^5 + A_{6,j} \text{Th}^6 + A_{7,j} \text{Th}^7 + A_{8,j} \text{Th}^8 + A_{9,j} \text{Th}^9$$

Trong đó: Th – là thời điểm quan trắc trong năm (tháng); Phương trình đường cong thực

nghiệm đa thức bậc cao với bộ hệ số của năm cơ sở là 2015 và 2017 được thiết lập thông qua bộ số liệu quan trắc tại các điểm được tổng hợp trong Bảng 6 (ví dụ cho BOD<sub>5</sub> và COD tại vị trí quan trắc cầu Lê Văn Sĩ).

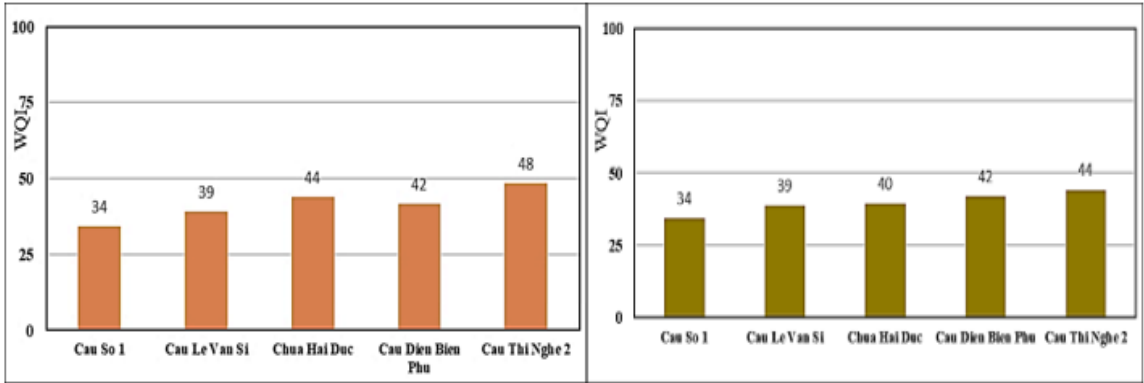
Trên cơ sở bộ giá trị hệ số phương trình HoCEq của các thành phần CLN tại các vị trí cơ sở (cầu Lê Văn Sĩ, cầu Điện Biên Phủ) các năm cơ sở (2015 & 2017), bộ giá trị các hệ số trên cho các vị trí quan trắc và năm tính toán còn lại trong giai đoạn 2012 - 2021 được xác định thông qua nội suy IDW.

Các thành phần CLN nước còn thiếu trong bộ số liệu quan trắc giai đoạn 2012 - 2021 được tính theo phương trình HoCEq. Ví dụ số liệu thành phần CLN tại cầu Lê Văn Sĩ năm 2012 sau bổ sung như Bảng 7.

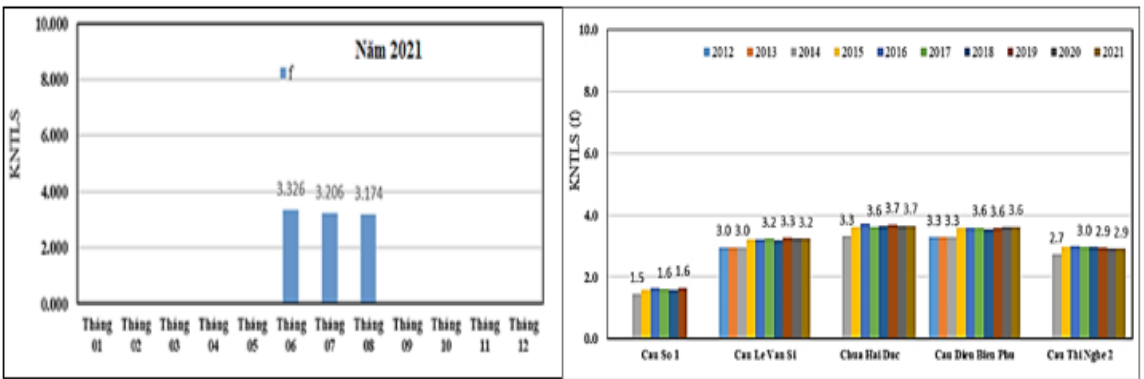
**3.4. Đánh giá sự thay đổi kết quả đánh giá sử dụng bộ số liệu chất lượng nước trước và sau “lấp đầy”**

Sau khi bổ sung giá trị thành phần CLN cho các vị trí và năm quan trắc còn trống giai đoạn 2012 - 2021, việc phân tích và đánh giá chất lượng nước theo chỉ số chất lượng nước của Việt Nam (Vietnam water quality index, VN-WQI) và khả năng tự làm sạch của nguồn nước kênh NLTN thuận lợi và thể hiện được các xu hướng biến động theo thời gian và không gian.

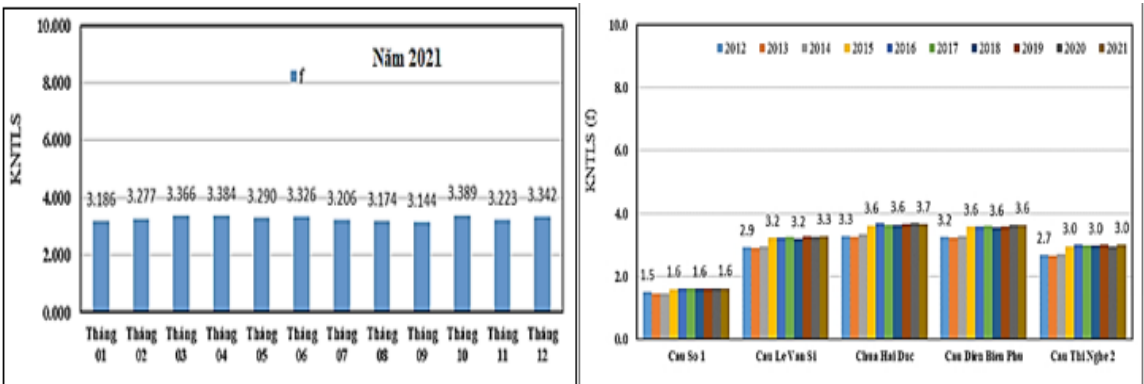
Kết quả cho thấy giá trị VN - WQI tăng dần từ cuối kênh ra đến đầu kênh (cửa ra sông Sài Gòn) đối với bộ số liệu đã được “lấp đầy” so với việc sử dụng bộ số liệu thiếu thành phần nước và không liên tục (Hình 3).



**Hình 3.** Phân bố giá trị chất lượng nước WQI (water quality index) của Việt Nam dọc kênh Nhiêu Lộc Thị Nghè trường hợp sử dụng bộ số liệu “thiếu” (bên trái) và bộ số liệu đã được “lấp đầy” (bên phải).



**Hình 4.** Giá trị  $f$  - đánh giá khả năng tự làm sạch (KNTLS) của nước tại vị trí cầu Lê Văn Sĩ năm 2021 và toàn kênh Nhiêu Lộc Thị Nghè (trước bổ sung).

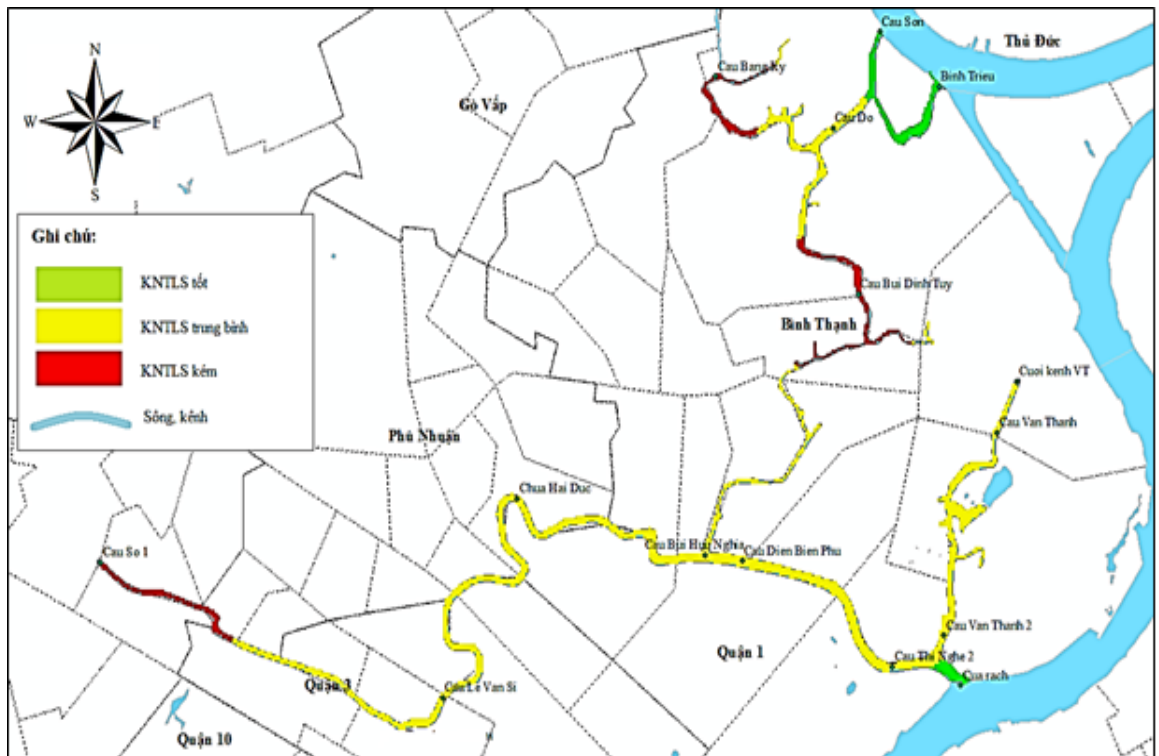


**Hình 5.** Giá trị  $f$  - đánh giá khả năng tự làm sạch của nước tại vị trí cầu Lê Văn Sĩ năm 2021 và toàn kênh Nhiêu Lộc Thị Nghè (sau bổ sung).

Đánh giá KNTLS thông qua tỷ số  $f$  - giữa hệ số hai quá trình phân hủy và bổ sung oxy cũng cho thấy sự phù hợp của bộ số liệu sau “lấp đầy”. Hình 4 cho thấy năm 2021, tại vị trí quan trắc

- cầu Lê Văn Sĩ chỉ quan trắc 3 tháng/12 tháng, giá trị trung bình đạt (3,2).

Hình 5 cho thấy, giá trị  $f$  phân bố tất cả các tháng trong năm và thể hiện sự biến động chất



**Hình 6.** Bản đồ phân vùng khả năng tự làm sạch (KNTLS) nước kênh Nhiêu Lộc Thị Nghè giai đoạn 2012 - 2021.

chất lượng nước của kênh (thường đầu mùa khô và mùa mưa chất lượng nước kém) và giá trị đại diện năm của vị trí cao hơn (3,3).

Dựa kết quả phân tích đánh giá KNTLS của nước kênh NLTN giai đoạn 2012 - 2021, sử dụng công cụ phần mềm ARGIS lập bản đồ phân vùng mức độ KNTLS của kênh (Hình 6).

**4. Kết Luận**

Với việc sử dụng hàm MREq và phương trình HoCEq luân phiên trong việc nội suy thành phần CLN kênh Nhiêu Lộc Thị Nghè tại các vị trí quan trắc còn trống trong giai đoạn 2012 - 2021 cho thấy phù hợp với hệ số tương quan bình phương R<sup>2</sup> khá cao (trên 0,5) tương ứng với mức quan hệ khá chặt. Bên cạnh đó, việc sử dụng của hàm và phương trình trên khá dễ dàng và thuận tiện trong quá trình tính toán vì không có “hộp đen” như trong nhiều mô hình AI đang được sử dụng trong nhiều nghiên cứu. Điều này thông qua việc nghiên cứu đã sử dụng bộ số liệu quan trắc tại các vị trí cơ sở và năm cơ sở để đánh giá mức độ phù hợp khi sử dụng hàm (MREq) và phương trình (HoCEp) vừa lập tính “ngược lại” thành

phần CLN.

Sau khi bộ số liệu quan trắc thành phần CLN được bổ sung và lấp đầy các khoảng trống đã cho thấy tính liên tục, tính đại diện và cho thấy tính xu thế “biến đổi” chất lượng cũng như khả năng tự làm sạch của nguồn nước kênh Nhiêu Lộc Thị Nghè theo không gian và thời gian những năm quan trắc và cả giai đoạn quan trắc.

Tuy nhiên, trong tính toán cho thấy một số thành phần CLN có hệ số tương quan bình phương R<sup>2</sup> còn thấp nên cần có thêm nghiên cứu và điều chỉnh (Ví dụ thành phần Coliform, PO4 trong Bảng 4). Ngoài ra, trong hàm hồi quy đa biến, một số yếu tố tác động đến chất lượng nước kênh bị “bỏ qua” do thiếu thông tin và số liệu hoặc lấy chung toàn Thành phố như số lượng dân cư khu vực, GDP các năm. Do vậy, cần thu thập thêm thông tin như dân số khu vực xả thải ra kênh cùng tình hình kinh tế (mức độ biến động %) hàng năm nhằm làm giàu thêm bộ số liệu sử dụng cho “lấp đầy” bộ số liệu quan trắc khi thiếu hoặc không được đo đạc.

**Bảng 7.** Số liệu thành phần CLN kênh NLTN tại Cầu Lê Văn Sĩ năm 2012 sau bổ sung

Năm	Thời điểm	T	pH	TSS	NH <sub>4</sub>	PO <sub>4</sub>	DO	COD	BOD <sub>5</sub>	Coliform	<i>E. coli</i>	
												Thăng
2012	01	Ròng	26,4	7,71	137,3	4,05	0,02	3,36	13,4	8,5	13003	1483
	02	Ròng	26,8	6,94	106	13,14	0,189	1,49	193,3	100,0	4600000	430000
	03	Ròng	29,0	7,56	104,3	0,23	0,05	3,31	23,7	35,4	66801	63276
	04	Ròng	29,9	7,03	86	8,91	0,832	3,20	91,6	15,1	4600000	930000
	05	Ròng	30,7	8,30	112,5	2,25	0,06	2,73	26,2	29,7	12863	0
	06	Ròng	29,2	8,42	102,3	3,74	0,05	3,01	23,1	28,1	98177	205230
	07	Ròng	28,9	6,84	28,00	6,14	0,06	4,95	14,6	16,3	187187	449018
	08	Ròng	29,0	3,89	2,65	13,27	0,00	5,52	27,7	10,5	168194	272217
	09	Ròng	28,6	6,76	55	10,95	1,168	1,44	73,4	37,0	930000	40000
	10	Ròng	28,7	6,80	16,25	0,64	0,18	3,96	19,5	3,8	100395	452077
	11	Ròng	29,1	7,04	48	9,07	0,899	5,84	60,4	29,9	430000	90000
	12	Ròng	28,6	6,25	26,12	20,92	0,20	5,29	36,3	15,9	68980	26560

Chi chú: số nghiêng là giá trị bổ sung.  
 MPN: chỉ số sử dụng để ước tính nồng độ vi sinh vật khả thi, T: nhiệt độ, pH: chỉ số ion hydro, TSS: tổng thành phần lơ lửng, NH<sub>4</sub>: amoniac, PO<sub>4</sub>: thành phần Phot pho, DO: lượng oxy hòa tan trong nước, COD: nhu cầu oxy hóa học, BOD<sub>5</sub>: nhu cầu oxy sinh hóa, *E. Coli*: *Escherichia coli*.

**Bảng 6.** Tổng hợp các hệ số phương trình HoCeq cho BOD<sub>5</sub> và COD tại cầu Lê Văn Sĩ

CLN	COD						BOD <sub>5</sub>					
	Triều rông			Triều lớn			Triều rông			Triều lớn		
Năm	2015	2017	2015	2017	2015	2017	2015	2017	2015	2017	2015	2017
A9	-8,82E-05	-7,91E-04	-1,01E-04	2,48E-04	-1,14E-04	-4,86E-04	-1,62E-04	3,85E-04				
A8	4,79E-03	4,61E-02	5,41E-03	-1,33E-02	6,47E-03	2,85E-02	9,22E-03	-2,14E-02				
A7	-1,10E-01	-1,15E+00	-1,22E-01	3,04E-01	-1,57E-01	-7,11E-01	-2,24E-01	5,07E-01				
A6	1,39E+00	1,59E+01	1,52E+00	-3,85E+00	2,14E+00	9,88E+00	3,02E+00	-6,66E+00				
A5	-1,07E+01	-1,34E+02	-1,15E+01	2,97E+01	-1,78E+01	-8,37E+01	-2,48E+01	5,33E+01				
A4	5,19E+01	7,08E+02	5,40E+01	-1,44E+02	9,32E+01	4,44E+02	1,28E+02	-2,67E+02				
A3	-1,56E+02	-2,32E+03	-1,58E+02	4,40E+02	-3,06E+02	-1,46E+03	-4,07E+02	8,33E+02				
A2	2,75E+02	4,48E+03	2,71E+02	-8,05E+02	5,96E+02	2,82E+03	7,61E+02	-1,54E+03				
A1	-2,56E+02	-4,52E+03	-2,43E+02	7,96E+02	-6,09E+02	-2,86E+03	-7,43E+02	1,51E+03				
A0	1,06E+02	1,80E+03	9,70E+01	-3,04E+02	2,47E+02	1,14E+03	2,89E+02	-5,77E+02				

CLN: Chất lượng nước, COD: nhu cầu oxy hóa học, BOD<sub>5</sub>: nhu cầu oxy sinh hóa.

## Lời Cam Đoan

Bài báo được sự đồng thuận của tất cả các tác giả.

## Lời Cảm Ơn

Bài báo được thực hiện với kết quả nghiên cứu “Khảo sát, đánh giá chất lượng nước và khả năng tự làm sạch, xây dựng hệ thống dự báo chất lượng nước tự động và đề xuất các giải pháp kiểm soát quản lý chất lượng nước”.

## Tài Liệu Tham Khảo (References)

- Harvey, A. C. (1990). *The econometric analysis of time series* (2<sup>nd</sup> ed.). Massachusetts, USA: The MIT Press.
- Hunter, J. M., Maier, H. R., Gibbs, M. S., Foale, E. R., Grosvenor, N. A., Harders, N. P., & Kikuchi-Miller, T. C. (2018). Framework for developing hybrid process-driven, artificial neural network and regression models for salinity prediction in river systems. *Hydrology and Earth System Sciences* 22(5), 2987-3006. <https://doi.org/10.5194/hess-22-2987-2018>.
- Le, T., & Le, H. Q. (2004). *Environment in Dong Nai - Sai Gon River Basin*. Ha Noi, Vietnam: Science and Technics Publishing House.
- Le, X. H., Ho, H. V., Lee, G., & Jung, S. (2019). Application of long short-term memory (LSTM) neural network for flood forecasting. *Water* 11(7), 1387. <https://doi.org/10.3390/w11071387>.
- Lin, K., Lu, P., Xu, C. Y., Yu, X., Lan, T., & Chen, X. (2019). Modeling saltwater intrusion using an integrated bayesian model averaging method in the Pearl River Delta. *Journal of Hydroinformatics* 21(6), 1147-1162. <https://doi.org/10.2166/hydro.2019.073>.
- Palani, S., Liang, S. Y., & Tkalich, P. (2008). An ANN application for water quality forecasting. *Marine Pollution Bulletin* 56(9), 1586-1597. <https://doi.org/10.1016/j.marpolbul.2008.05.021>.
- Ross, A. C., & Stock, C. A. (2019). An assessment of the predictability of column minimum dissolved oxygen concentrations in Chesapeake Bay using a machine learning model. *Estuarine, Coastal and Shelf Science* 221, 53-65. <https://doi.org/10.1016/j.ecss.2019.03.007>.