VNU Journal of Science: Natural Sciences and Technology

Journal homepage: https://js.vnu.edu.vn/NST

Original Article

Assessment of Groundwater Quality in Ha Tinh Province Using Water Quality Index

Truong Trung Kien, Nguyen Dang Giang Chau,
Nguyen Van Hop*, Truong Quy Tung

University of Sciences, Hue University, 77 Nguyen Hue, Hue, Thua Thien Hue, Vietnam

Received 08th October 2024
Revised 16th March 2025; Accepted 04th April 2025

Abstract: The Groundwater Quality Index (GWQI) was developed and applied to assessment of groundwater quality in Ha Tinh province. This study utilized water quality monitoring data in 2018 - 2020 period at 58 locations with a frequency of 4 times/year in 13 districts, towns and city in the province. Ten (10) parameters were selected (i = 1 - 10) to establish the GWQI, including: pH, total dissolved solids (TDS), hardness (HARD); sulfate (SO4), chemical oxygen demand (COD), ammonium as nitrogen (N-NH4), nitrate as nitrogen (N-NO3), total dissolved iron (Fe), total dissolved manganese (Mn) and total coliform (TC). The weight (w_i) of each parameter was based on principal component analysis (PCA) using monitoring data in 2018 and 2019. The sub-index (q_i) of each parameter was determined by a linear function, based on the limit value specified in the National Technical Regulation on Domestic Water Quality (QCVN 01-1:2018/BYT). The GWQI as a function of w_i and q_i was calculated by the multiplicate formula. The results showed that in 2020, in the province 21.1% of GWQI values classified as VERY GOOD or GOOD, 31.9% as AVERAGE, 33.2% as POOR and 13.8% as VERY POOR.

Keywords: Groundwater, GWQI, PCA, Ha Tinh province.

* Corresponding author.
E-mail address: ngvanhopkh@gmail.com

https://doi.org/10.25073/2588-1140/vnunst.5845

Nghiên cứu đánh giá chất lượng nước dưới đất ở tỉnh Hà Tĩnh sử dụng chỉ số chất lượng nước

Trương Trung Kiên, Nguyễn Đặng Giang Châu,
Nguyễn Văn Hợp*, Trương Quý Tùng

Trường Đại học Khoa học, Đại học Huế, 77 Nguyễn Huệ, Thành phố Huế, Thừa Thiên Huế, Việt Nam

Nhận ngày 08 tháng 10 năm 2024
Chỉnh sửa ngày 16 tháng 3 năm 2025; Chấp nhận đăng ngày 04 tháng 4 năm 2025

Tóm tắt: Chỉ số chất lượng nước dưới đất (GWQI) đã được xây dựng và áp dụng để đánh giá chất lượng nước dưới đất ở tỉnh Hà Tĩnh. Nghiên cứu này sử dụng dữ liệu quan trắc chất lượng nước (CLN) giai đoạn 2018-2020 tại 58 vị trí với tần suất 4 đợt/năm ở 13 huyện, thị xã và thành phố ở tỉnh. Mười (10) thông số được lựa chọn (i = 1 – 10) để xây dựng GWQI gồm: pH, tổng chất rắn hòa tan (TDS), độ cứng (HARD); sunfat (SO4), COD, amoni tính theo nitơ (N-NH4), nitrat tính theo nitơ (N-NO3), tổng sắt tan (Fe), tổng mangan tan (Mn) và tổng coliform (TC). Trọng số *(w_i)* của các thông số CLN được xác định dựa vào phương pháp phân tích thành phần chính (PCA) trên tập dữ liệu năm 2018 và 2019. Chỉ số thông số (q_i) của các thông số được xác định theo hàm tuyến tính và chúng được thiết lập dựa vào các giá trị giới hạn được quy định trong Quy chuẩn kỹ thuật quốc gia về chất lượng nước sử dụng cho mục đích sinh hoạt (QCVN 01-1:2018/BYT). Chỉ số GWQI là hàm của w_i và q_i, được tính theo công thức dạng tích. Kết quả cho thấy, năm 2020, có 21,1% giá trị GWQI ứng với CLN loại RẤT TỐT/TỐT; 31,9% - TRUNG BÌNH; 33,2% - KÉM và 13,8% - RẤT KÉM.

Từ khóa: Nước dưới đất, GWQI, PCA, tỉnh Hà Tĩnh.

1. Mở đầu

Phương pháp đánh giá CLN phổ biến hiện nay là dựa vào so sánh kết quả quan trắc thông số CLN với giá trị giới hạn được quy định trong quy chuẩn kỹ thuật quốc gia về CLN tương ứng, chẳng hạn dựa vào QCVN 08:2023/BTNMT (về CLN mặt), QCVN 09:2023/BTNMT (về CLN dưới đất) [1, 2]. Phương pháp này khó hiểu đối với các nhà quản lý và cộng đồng, do họ ít hiểu thông số CLN riêng lẻ như COD, N-NH4, độ cứng,... Một phương pháp dễ hiểu hơn là dựa vào một chỉ số và thang đo duy nhất, đó là Chỉ số chất lượng nước (WQI/water quality index) hoặc Chỉ số chất lượng nước dưới đất (GWQI/groundwater quality index) [3, 4]. Chỉ số WQI hoặc GWQI là một số nguyên, thường có giá trị 0 - 100, càng gần 100 CLN càng tốt và ngược lại [3], [5]. Chỉ số WQI hoặc GWQI được tính toán từ *chỉ số thông số* q_i (thể hiện chất lượng của thông số CLN thứ *i*, không có đơn vị, nhận giá trị 0 - 100, q_i càng gần 100, thông số có chất lượng càng tốt và ngược lại) và *trọng số* w_i (thể hiện tầm quan trọng tương đối của thông số *i*, tổng trọng số w_i bằng 1, hay tổng tầm quan trọng của 10 thông số lựa chọn là 100%, thông số nào có *w* lớn hơn, thì quan trọng hơn và ngược lại) [3]. Hầu hết các nghiên cứu đều cho rằng, việc xây dựng chỉ số WQI hoặc GWQI gồm 4 bước: i) Lựa chọn các thông số CLN; ii) Tính chỉ số thông số q_i thông qua việc chuyển nồng độ quan trắc hay phân tích được của mỗi thông số thành giá trị nằm trong khoảng 0 - 100; iii) Xác định trọng số w_i của mỗi thông số; và iv) Tổ hợp các giá trị q_i và w_i thành WQI hoặc GWQI và đánh giá CLN dựa vào thang điểm phân loại WQI hoặc GWQI

* Tác giả liên hệ.
Địa chỉ email: ngvanhopkh@gmail.com

https://doi.org/10.25073/2588-1140/vnunst.5845

[3]. Để xác định w_i có 2 phương pháp thường dùng: i) Dựa vào ý kiến của các chuyên gia; ii) Áp dụng phân tích thống kê đa biến như phương pháp phân tích thành phần chính (PCA/principal component analysis) và một dạng tương tự là phân tích yếu tố (FA/factor analysis). Nhiều tác giả cho rằng, phương pháp PCA cho phép xác định w_i khách quan hơn (vì nó dựa vào phân tích dữ liệu quan trắc CLN), khắc phục được tính chủ quan của phương pháp lấy ý kiến chuyên gia. Nhiều tác giả đã áp dụng phương pháp PCA hoặc FA để xác định w_i trong phát triển chỉ số GWQI [6-8].

Ở Việt Nam, Bộ Tài nguyên và Môi trường (BTNMT) đã ban hành chỉ số WQI để áp dụng đánh giá chất lượng nước của các nguồn nước mặt (từ năm 2011 và điều chỉnh năm 2019) [9]. Trong đó, chỉ số WQI không tính đến w_i, mà chỉ tính đến q_i (hay coi các thông số CLN lựa chọn có trọng số bằng nhau, tức là các giá trị w_i đều bằng 1). Đến nay, BTNMT chưa ban hành chỉ số GWQI để đánh giá CLN dưới đất. Nguyễn Hải Âu và cộng sự [10] đã áp dụng chỉ số GWQI do N. Subba Rao [11] đề xuất năm 2019 để đánh giá CLN dưới đất tầng Pleistocen tại xã Phú Mỹ, tỉnh Bà Rịa Vũng Tàu: lựa chọn 10 thông số CLN, xác định w_i theo phương pháp entropy. Nói chung, các nghiên cứu về GWQI ở Việt Nam còn rất hạn chế.

Quy trình tổng quát xây dựng chỉ số WQI công bố năm 2022 [12] đã được sử dụng để thiết lập chỉ số GWQI và áp dụng trong đánh giá CLN dưới đất ở tỉnh Hà Tĩnh.

2. Phương pháp nghiên cứu

2.1. Phạm vi nghiên cứu và dữ liệu chất lượng nước dưới đất

Hà Tĩnh là một tỉnh ở dải đất miền Trung, Việt Nam, phía Bắc giáp tỉnh Nghệ An, phía Nam giáp tỉnh Quảng Bình, phía Tây giáp Nước Cộng hòa Dân chủ Nhân dân Lào, phía Đông giáp biển Đông. Địa hình đa dạng, có đủ các vùng đồi núi, trung du, đồng bằng và ven biển. Hà Tĩnh có tới 17 con sông lớn nhỏ và nhiều hồ nước. Tỉnh Hà Tĩnh nằm trong khu vực nhiệt đới, gió mùa nóng ẩm, mưa nhiều. Diện tích đất tự nhiên 5,994,45 km^2. Vị trí Hà Tĩnh nằm trên trục kinh tế Bắc - Nam, là nơi giao thương kinh tế, giao lưu văn hóa Đông - Tây, là cửa ngõ lớn nối Việt Nam với thế giới qua đường bộ và đường biển [13]. Dân số của Tỉnh Hà Tĩnh là 1,323,746 người (2023) [14]. Người dân ở Hà Tĩnh ngoài khai thác nước mặt, còn một bộ phận dân cư ở khu vực nông thôn (khoảng 77,7 %) sử dụng nước dưới đất cho nhiều hoạt động khác nhau: sinh hoạt, tưới cây, cấp nước cho công nghiệp,… [13] (Hình 1).

Hình 1. Vị trí quan trắc CLN dưới đất tỉnh Hà Tĩnh.

Hàng năm, Trung tâm Quan trắc Tài nguyên và Môi trường (QT-TNMT) tỉnh Hà Tĩnh tiến hành quan trắc CLN dưới đất tại 58 điểm với tần suất 04 đợt/năm ở 13 huyện, thị xã và thành phố thuộc tỉnh (số điểm quan trắc ghi trong ngoặc đơn), gồm: thành phố Hà Tĩnh (3); các thị xã Kỳ Anh (16) và Hồng Lĩnh (3); các huyện Đức Thọ (3), Hương Sơn (8), Hương Khê (1), Cẩm Xuyên (5), Lộc Hà (4), Nghi Xuân (6), Vũ Quang (1), Thạch Hà (5), Can Lộc (1) và Kỳ Anh (2).

Dữ liệu gốc về CLN dưới đất của tỉnh Hà Tĩnh (giai đoạn 2018-2020), do Trung tâm QT-TNMT tỉnh Hà Tĩnh cung cấp, được xử lý để tạo ra tập dữ liệu phù hợp cho phân tích theo PCA. Tập dữ liệu này được chia thành 2: *tập dữ liệu 1* (năm 2018 và 2019, với 464 mẫu hay N = 464) được dùng để xây dựng chỉ số GWQI; *tập dữ liệu 2* (năm 2020 với N = 232) được dùng để kiểm chứng chỉ số GWQI xây dựng được và đánh giá CLN dưới đất ở tỉnh Hà Tĩnh. Do dung lượng 2 tập dữ liệu khá lớn nên số liệu chi tiết của tập dữ liệu 1 và 2 không đưa ra ở đây.

Mười (10) thông số được lựa chọn để xây dựng chỉ số GWQI gồm: pH, tổng chất rắn hòa tan (TDS), độ cứng (HARD), nhu cầu oxy hóa học (COD) hay chỉ số pemanganat, sunfat (viết tắt là SO4), amoni tính theo nitơ (N-NH4), nitrat tính theo nitơ (N-NO3), tổng sắt tan (Fe), tổng mangan tan (Mn) và tổng coliform (TC).

2.2. Lựa chọn thông số

Để thuận lợi cho đánh giá CLN dưới đất dựa vào GWQI, các thông số được lựa chọn là các thông số cơ bản và được quan trắc thường xuyên trong chương trình quan trắc môi trường của các địa phương (tỉnh/thành). Số thông số lựa chọn (n) thường khoảng 9 - 13 [3, 15]. Trong nghiên cứu này, lựa chọn 10 thông số như nêu ở mục 2.1. Trong đó, các thông số pH, TDS, HARD, SO4 thể hiện đặc điểm hóa lý của nguồn nước; thông số COD thể hiện mức ô nhiễm các chất hữu cơ; thông số N-NH4 và N-NO3 thể hiện mức ô nhiễm các chất dinh dưỡng; Fe và Mn thể hiện mức nhiễm kim loại nặng; TC thể hiện mức nhiễm các vi khuẩn có nguồn gốc từ phân người và động vật. Các thông số này cũng được quy định trong QCVN 09-MT:2023/BTNMT về CLN dưới đất cho đa mục đích sử dụng và QCVN 01-1:2018/BYT về CLN sạch sử dụng cho mục đích sinh hoạt. Trong dữ liệu quan trắc, các kim loại nặng và độc (Hg, Cd, Pb, As, Ni, Cr, Cu, Zn) cũng như các hóa chất bảo vệ thực vật nhóm clo (nhóm DDT, HCH) đều có hàm lượng nhỏ hơn giới hạn phát hiện của phương pháp phân tích hoặc nhỏ hơn nhiều so với giới hạn cho phép trong các quy chuẩn kỹ thuật liên quan, nên chúng không được đưa vào để xây dựng chỉ số GWQI.

2.3. Phương pháp xác định chỉ số thông số

Chỉ số thông số q_i thể hiện chất lượng của thông số CLN. Có nhiều cách để chuyển kết quả quan trắc mỗi thông số (x) thành chỉ số thông số q. Một số tác giả xác định giá trị q theo công thức (1) [3, 11, 16]:

$$q_i = \frac{C_i}{C_{oi}} \times 100 \qquad (1)$$

Trong đó, C_i và C_{oi} tương ứng là nồng độ của mỗi thông số trong mẫu (mg/L) và giới hạn được quy định trong quy chuẩn kỹ thuật CLN của quốc gia hoặc WHO đối với mỗi thông số. Theo công thức (1), đối với mỗi thông số, q_i là hàm tuyến tính của C_i; q_i càng tăng, chất lượng của thông số càng giảm và ngược lại; khi $C_i > C_{oi}$, giá trị của $q_i > 100$.

Có một cách thuận tiện hơn là dựa vào các mức được quy định trong quy chuẩn quốc gia để xây dựng hàm tuyến tính dạng y = q = a + bx [17, 18]. Hiện nay ở nước ta người ta thường sử dụng nước giếng khoan, giếng đào cho sinh hoạt và do vậy, trong nghiên cứu này, sử dụng QCVN 01-1:2018/BYT (viết tắt là QCVN 01-1) quy định về CLN sạch sử dụng cho mục đích sinh hoạt [19]. Hệ số a và b trong phương trình tuyến tính được suy ra từ hai phương trình tuyến tính (2) và (3):

100 = a + b × (giới hạn cho phép của QCVN 01-1 đối với thông số – 3×S) (2)

Trong đó, y = q = 100 ứng với chất lượng tốt của thông số; Khi kết quả quan trắc thông số (x) < (giới hạn cho phép của QCVN 01-1) – 3×S, giá trị q của nó = 100; Khi x = QCVN 01-1, thì q = 50.

1 = a + b × (giới hạn cho phép của QCVN 01-1 đối với thông số + 3×S) (3)

Trong đó, y = q = 1 ứng với chất lượng kém của thông số; Khi kết quả quan trắc của thông số (x) ≥ (giới hạn cho phép) + 3×S, giá trị q của nó = 1. Điều này được hiểu là: giới hạn cho phép đối với mỗi thông số là trung bình của một phân bố chuẩn (hay một quần thể); chỉ những giá trị x nằm ngoài khoảng tin cậy 99,87% của giá trị trung bình đó mới được xem là phân bố khác (hay quần thể khác), tức là lớn hơn giới hạn cho phép; số 3 là giá trị của biến chuẩn hóa (z) ứng với xác suất tin cậy P = 0,9987; S là độ lệch chuẩn của quần thể [20].

Độ lệch chuẩn S có thể tính toán từ độ lệch chuẩn tương đối (RSD, %):

$$S = \frac{\text{RSD} \times \text{Trung bình số học}}{100} \quad (4)$$

Trong đó, Trung bình số học chính là giới hạn cho phép của thông số CLN trong QCVN 01-1; RSD được xác định dựa vào phương trình Horwitz [21]:

$$\text{RSD} = 2^{1-0,5\lg C} \quad (5)$$

Với C là nồng độ được biểu diễn bằng số thập phân; chẳng hạn khi C = 2 ppm (hay 2 mg/L), thì thay $C = 2.10^{-6}$ vào công thức (5).

2.4. Phương pháp xác định trọng số

Trọng số (w_i) thể hiện tầm quan trọng tương đối của thông số CLN và chúng có thể bằng hoặc khác nhau. Thông thường, tổng các trọng số bằng 1 [3, 5, 15].

Nhiều chỉ số GWQI sử dụng trọng số w_i khác nhau và được tính theo công thức (6) và (7) [3, 7, 11, 16]:

$$w_{i*} = \frac{K}{C_{oi}} \quad (6)$$

Trong đó, hệ số $K = 1/\sum(1/C_{oi})$; C_{oi} là giới hạn của thông số i được quy định trong quy chuẩn kỹ thuật CLN của quốc gia hoặc WHO; w_{i*} là trọng số của mỗi thông số CLN được xác định theo ý kiến chuyên gia [3]. Từ các giá trị w_{i*}, tính toán trọng số w_i theo công thức (7):

$$w_i = \frac{w_{i*}}{\sum_{i=1}^{n} W_{i*}} \quad (7)$$

Trong nghiên cứu này, phương pháp PCA được sử dụng để xác định trọng số của các thông số CLN [22, 23]. Phương pháp PCA là phương pháp giảm chiều, cho phép giảm dữ liệu từ không gian *m* chiều (mỗi chiều ứng với một biến/thông số CLN) thành không gian ít chiều hơn: *k* chiều (*k* < *m*); mỗi chiều được gọi là một thành phần chính (PC/principal component) - là tổ hợp tuyến tính của các biến gốc (x_{ij}) và các PC không tương quan với nhau, mà trực giao nhau.

Thông thường, chỉ cần giữ lại một vài PC đầu tiên, chẳng hạn PC1, PC2 vì chúng giải thích được đa số biến động hay phương sai của tập dữ liệu gốc. Những PC được giữ lại là những PC có *Giá trị riêng* lớn hơn 1 (giá trị riêng là lượng phương sai của các biến gốc được giải thích bởi PC tương ứng) và phương sai tích lũy của chúng lớn hơn 60% [22]. Mặt khác, từ kết quả tính toán theo PCA, sẽ xác định được *tải lượng bình phương* (squared loading) của mỗi biến trên mỗi PC. Tải lượng bình phương của mỗi biến chính là hệ số xác định (R^2) của mỗi biến trên PC tương ứng: Tải lượng bình phương càng lớn, biến càng đóng góp quan trọng vào PC. Từ *tổng tải lượng bình phương* của mỗi biến trên các PC giữ lại (communality) và Tổng tải lượng bình phương của tất cả các biến trên các PC giữ lại sẽ tính được trọng số (hay tầm quan trọng tương đối) của mỗi biến theo công thức (8). Để thực hiện các tính toán theo phương pháp PCA, nghiên cứu này sử dụng phần mềm R - phần mềm được áp dụng phổ biến trên thế giới: R version 4.0.3/64-bit (10-10-2020) với module R-Studio, dùng gói (package) Factoextra (version 1.0.7).

$$\text{Trọng số } (w_i) = \frac{\text{Tổng tải lượng bình phương của thông số i trên các PC giữ lại}}{\text{Tổng tải lượng bình phương của tất cả các thông số trên các PC giữ lại}} \quad (8)$$

2.5. Phương pháp tính chỉ số GWQI và thang điểm phân loại chất lượng nước dưới đất

Nhiều tác giả đề xuất thang điểm ngược để phân loại CLN dưới đất, tức là giá trị GWQI càng lớn, CLN càng giảm và ngược lại [6, 7], [16]. Để thuận lợi cho người dùng chỉ số GWQI, nghiên cứu này sử dụng thang điểm thuận để phân loại CLN dưới đất tương tự như thang điểm phân loại của Tổng cục Môi trường quy định [9]. GWQI càng lớn, CLN dưới đất càng tốt và ngược lại (Bảng 1).

Để chọn công thức tính phù hợp, tiến hành tính chỉ số GWQI cuối cùng cho Tập dữ liệu 1 theo phương pháp tích như công thức (9):

$$GWQI = \prod_{i=1}^{n} q_i^{w_i} \quad (9)$$

Mặt khác, để so sánh, ở đây cũng tính toán giá trị GWQI cuối cùng theo một phương pháp khác đã được một số tác giả đã đề xuất [6, 7], [16] như công thức (10):

$$GWQI = \sum_{i=1}^{n} q_i \times \frac{w_{i*}}{\sum_{i=1}^{n} w_{i*}} \quad (10)$$

Trong đó, trọng số w_i được tính theo (6) và (7), Chỉ số thông số q_i được tính theo (11):

$$q_i = \frac{100 \times (C_i - C_o)}{S_i - C_o} \quad (11)$$

Ở đây, C_i và C_o tương ứng là nồng độ của thông số CLN trong mẫu và nồng độ lý tưởng của thông số trong nước tinh khiết = 0 (ngoại trừ pH có C_o = 7,0).

Theo công thức (11), GWQI được phân chia thành 5 mức CLN: < 50 (RẤT TỐT); 51 - 100 (TỐT); 101 - 200 (KÉM); 201 - 300 (RẤT KÉM) và > 300 (KHÔNG PHÙ HỢP CHO ĂN UỐNG). GWQI càng bé, CLN dưới đất càng tốt và ngược lại.

Bảng 1. Phân loại chất lượng nước dưới đất theo chỉ số GWQI

Loại/ Mức	Chỉ số GWQI	Đánh giá chất lượng
I	81 – 100	RẤT TỐT (RT)
II	61 – 80	TỐT (T)
III	41 – 60	TRUNG BÌNH (TB)
IV	21 – 40	KÉM (K)
V	≤ 20	RẤT KÉM (RK/ không phù hợp cho ăn uống)

3. Kết quả và thảo luận

3.1. Chỉ số thông số của các thông số chất lượng nước

Các hàm tuyến tính (phương trình) xác định Chỉ số thông số q_i đối với thông số CLN i (i = 1-10), gồm:

(1) Khi 5,5 ≤ pH < 6: y = 99x – 445,5; 8,5 < pH ≤ 9,5: y = -99x + 951,5; 6 ≤ pH ≤ 8,5: y = 100; pH <5,5 và pH > 9,5: y = 1.

(2) Khi nồng độ TDS ≤ 1000 mg/L: y = 100; 1000 mg/L < TDS ≤ 1170 mg/L: y = -0,5834x + 683,36; TDS > 1170 mg/L: y = 1.

(3) Khi nồng độ HARD (tính theo $CaCO_3$) ≤ 300 mg/L: y = 100; 300 mg/L < HARD ≤ 361 mg /L: y = -1,6222x + 586,67; HARD > 361 mg /L: y = 1.

(4) Khi nồng độ SO4 ≤ 250 mg/L: y = 100; 250 mg/L < SO4 ≤ 302 mg/L: y = -1,894x + 573,5; SO4 > 302 mg/L: y = 1.

(5) Khi nồng độ COD ≤ 2,0 mg/L: y = 100; 2,0 mg/L < COD ≤ 2,9 mg/L: y = -114,47x + 328,92; COD > 2,9 mg/L: y = 1.

(6) Khi nồng độ N-NH4 ≤ 0,3 mg/L: y = 100; 0,3 mg/L < N-NH4 ≤ 0,47 mg/L: y = -573,55x + 272,07; N-NH4 > 0,47 mg/L: y = 1.

(7) Khi nồng độ N-NO3 ≤ 2,0 mg/L: y = 100; 2,0 mg/L < N-NO3 ≤ 2,86 mg/L: y = -573,55x + 272,07; N-NO3 > 2,86 mg/L: y = 1.

(8) Khi nồng độ Fe ≤ 0,3 mg/L: y = 100; 0,3 mg/L < Fe ≤ 0,47 mg/L: y = -573,55x + 272,07; Fe > 0,47 mg/L: y = 1.

(9) Khi nồng độ Mn ≤ 0,1 mg/L: y = 100; 0,1 mg/L < Mn ≤ 0,17 mg/L: y = -1458,41x + 245,84; Mn > 0,17 mg/L: y = 1

(10) Khi nồng độ TC ≤ 3,0 MPN/100 mL: y = 100; 3,0 MPN/100 mL < TC ≤ 4,2 MPN/100 mL: y = -81,11 + 343,34; TC > 4,2 MPN/100 mL: y = 1.

Số liệu thống kê mô tả tập dữ liệu 1, được tính toán bằng Microsoft-Excel với công cụ Real Statistics (Bảng 2). Quy chuẩn QCVN 01-1 [17] cũng được đưa vào Bảng 2 để chỉ ra các giới hạn cho phép của các thông số được sử dụng để thiết lập các hàm tuyến tính xác định chỉ số thông số q_i.

Bảng 2. Tổng hợp các đại lượng thống kê mô tả cho 10 thông số được chọn của tập dữ liệu 1 (n=464)

Đại lượng thống kê (*)	pH (mg/L)	TDS (mg/L)	HARD (mg/L)	SO4 (mg/L)	COD (mg/L)	N-NH4 (mg/L)	N-NO3 (mg/L)	Fe (mg/L)	Mn (mg/L)	TC (MPN/ 100mL)
Min	3,8	7	10	5	0,2	0,1	0,0	0,01	0,10	1
Max	8,2	4800	4590	2624	18,0	19,0	36,0	90,00	4,10	600
Mean	6,1	236	222	62	2,7	0,4	3,5	2,01	0,30	43
±Stdev	0,9	456	460	226	2,1	1,3	4,8	7,73	0,46	76
Median	6,1	125	111	18	2,0	0,1	1,4	0,17	0,14	11
MAD	0,6	2	3	4	4,6	5,6	6,6	7,64	8,64	10
Quartile 1	5,5	51	47	10	2,0	0,1	0,4	0,10	0,10	3
Quartile 3	6,8	223	230	45	2,0	0,1	4,5	0,60	0,32	36
CV (%)	14	193	208	362	75,6	328	139	384	155	180
CL 95%	0,1	73	74	36	0,3	0,2	0,8	1,24	0,07	12
QCVN 01-1	6-8,5	1000	300	250	2,0	0,3	2,0	0,30	0,10	3
$RSD_{Horwitz}$	-	6	7	7	14,4	19,2	14,4	19,2	22,6	14
(+/-) S	-	65	20	17	0,3	0,1	0,3	0,06	0,02	1

(*) Min và Max: giá trị nhỏ nhất và giá trị lớn nhất; Mean: trung bình số học;
±Stdev: Độ lệch chuẩn; Median: trung vị; MAD: độ lệch tuyệt đối trung bình; CV(%): hệ số biến động;
CL95%: giới hạn tin cậy 95%; QCVN 01-1: viết tắt của QCVN 01-1:2018/BYT - Quy chuẩn kỹ thuật quốc gia về chất lượng nước dùng cho mục đích sinh hoạt; $RSD_{Horwitz}$ và ± S: tương ứng là độ lệch chuẩn tương đối và độ lệch chuẩn được tính từ phương trình Horwitz với C_o được biểu diễn bằng phân số
(C_o là mức giới hạn của thông số được quy định trong QCVN 01-1).

3.2. Trọng số của các thông số chất lượng nước từ tính toán theo PCA

Trên tập dữ liệu 1, thực hiện chuyển dạng các biến gốc x_{ij} (10 thông số CLN) thành biến chuẩn hóa z_{ij}: z_{ij} = (x_{ij} – Độ lệch chuẩn)/Trung bình số học; biến z_{ij} có giá trị trung bình quần thể bằng 0 và phương sai bằng 1 [22, 23]. Thực hiện phương pháp PCA trên ma trận tương quan của các biến z_{ij}, thu được các giá trị riêng, phần phương sai và phương sai tích lũy của các thành phần chính (PC) ở Bảng 3. Các kết quả cho thấy: Chỉ cần giữ lại 04 thành phần chính (PC1 – PC4) vì chúng đều có giá trị riêng ≥ 1 và giải thích được 68,7% phương sai của tập dữ liệu gốc; Các thành phần chính còn lại (PC5 – PC10) đều có giá riêng < 1 và chỉ giải thích 31,3 % phương sai, được xem là biến động nền hay "nhiễu nền" của tập dữ liệu gốc.

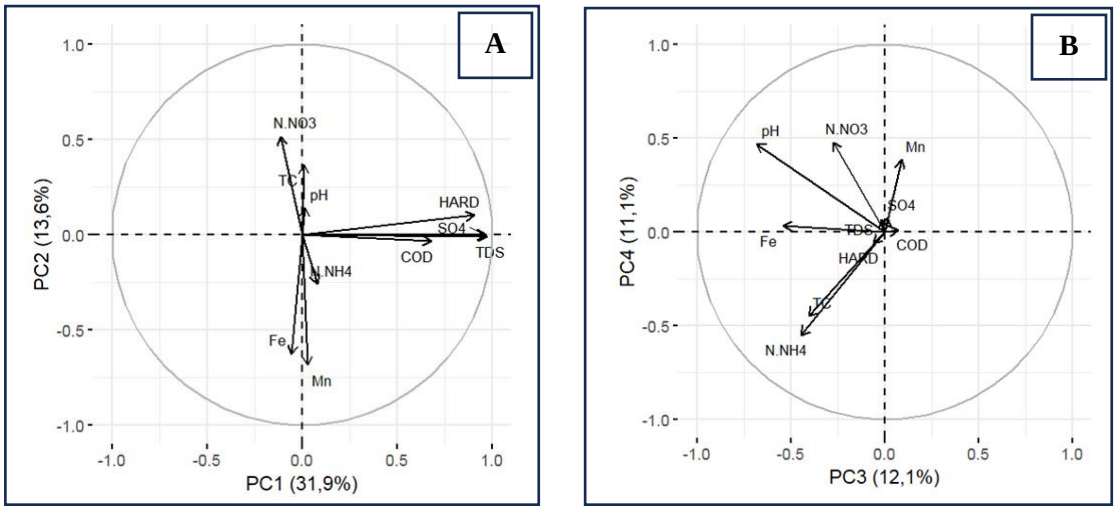
Bảng 3. Giá trị riêng, phần phương sai và phương sai tích lũy của 04 thành phần chính đầu tiên

Thành phần chính	Giá trị riêng	Phần phương sai (%)	Phương sai tích lũy (%)
PC1	3,2	31,9	31,9
PC2	1,3	13,6	45,5
PC3	1,2	12,1	57,6
PC4	1,1	11,1	68,7

Bảng 4. Trọng số w của thông số chất lượng nước(*)

Thông số	pH	TDS	HARD	SO4	COD	N-NH4	N-NO3	Fe	Mn	TC
Tổng tích lũy trên 4 PC (1-4)	0,701	0,950	0,839	0,941	0,468	0,575	0,581	0,689	0,626	0,500
Trọng số (w)	**0,10**	**0,14**	**0,12**	**0,14**	**0,07**	**0,08**	**0,08**	**0,10**	**0,09**	**0,08**

(*) Ngoại trừ các giá trị w, số con số sau dấu phẩy của các số liệu trong bảng được giữ lại chỉ để phục vụ tính toán. Tổng tích lũy của 10 biến trên 4 PC giữ lại bằng tổng giá riêng của chúng bằng 6,8.



Hình 2. Tải lượng yếu tố A. PC1 × PC2 và B. PC3 × PC4.

Biểu diễn tải lượng yếu tố của các biến (factor loading plot) trong không gian các thành phần chính PC1×PC2 và PC3×PC4, thu được Hình 2. Hình 2 cho thấy: PC1 giải thích tốt các biến TDS, HARD, SO4 và COD; PC2 giải thích tốt biến Mn, Fe, N-NO3; PC3 chủ yếu giải thích biến pH và một phần Fe, N-NH4; còn PC4 chủ yếu giải thích biến N-NH4 và TC (tải lượng mỗi biến trên mỗi PC là hình chiếu vuông góc của nó trên PC đó).

Từ tải lượng của các biến ở Hình 2 tính được *tải lượng bình phương* của mỗi biến và *tổng tải lượng bình phương* của nó trên các PC giữ lại. Từ đó tính được *tổng tải lượng bình phương* của 10 biến trên các PC giữ lại và trọng số của mỗi biến theo công thức (8). Trọng số w_i của 10 thông số CLN được nêu trong Bảng 4.

3.3. Kiểm tra sự phù hợp của chỉ số GWQI xây dựng được

Để kiểm tra sự phù hợp của công thức tính GWQI khi phản ánh CLN thực tế ở vùng khảo sát, tiến hành so sánh kết quả tính toán và phân loại CLN theo 02 (hai) công thức khác nhau là công thức tích (9) và công thức (10) trên tập dữ liệu 1, được ký hiệu tương ứng là GWQI và $GWQI_{Ref}$.

Bảng 5. So sánh kết quả tính các chỉ số GWQI, $GWQI_{Ref}$, và phân loại CLN đối với một số trường hợp của tập dữ liệu 1 (*)

Mẫu	pH	HARD (mg/L)	N-NO3 (mg/L)	TC (MPN/ 100mL)	GWQI - phân loại	$GWQI_{Ref}$ - phân loại
TK18-3	4,9 (*)	65	0,1	3 (*)	39 - RK	100 - T
CN18-1	6,5	428 (*)	8,8 (*)	7 (*)	27 - RK	84 - T
CN18-2	7,5	363 (*)	9,7 (*)	3 (*)	39 - RK	83 - T
KN18-2	8,0	590 (*)	0,7	17 (*)	39 - RK	80 - T
KNa18-4	6,0	439 (*)	3,3 (*)	3 (*)	26 - RK	100 - T
KP18-1	4,5 (*)	32	3,6 (*)	26 (*)	30 - RK	90 - T
KT18-4	4,6 (*)	123	3,9 (*)	6 (*)	30 - RK	84 - T
KLg18-1	5,2 (*)	106	7,3 (*)	18 (*)	40 - RK	90 - T
KgL18-4	4,8 (*)	33	8,9 (*)	13 (*)	30 - RK	92 - T
KL18-4	4,8 (*)	47	2,5 (*)	5 (*)	40 - RK	83 - T
KB18-1	5,2 (*)	71	6,1 (*)	16 (*)	40 - RK	97 - T

(*): không đạt yêu cầu so với quy định của QCVN 01-1.

Các kết quả ở Bảng 5 cho thấy, các chỉ số GWQI và $GWQI_{Ref}$ phản ánh CLN khác nhau. Đối với các trường hợp đó, mặc dù có một vài thông số không đạt yêu cầu (so với quy định trong QCVN 01-1), tuy nhiên chỉ số $GWQI_{Ref}$ vẫn phản ánh CLN loại TỐT là không phù hợp. Nói cách khác, trong những trường hợp đó, chỉ số $GWQI_{Ref}$ bị mắc tính "mờ" hoặc "che khuất" (eclipsing) trong phản ánh CLN [3]. Trong khi đó, chỉ số GWQI phản ánh CLN phù hợp với thực tế hơn. Rõ ràng, chỉ số GWQI phù hợp hơn trong phản ánh CLN dưới đất cho mục đích ăn uống ở vùng khảo sát.

3.4. Áp dụng chỉ số GWQI để đánh giá CLN dưới đất của tỉnh Hà Tĩnh trong năm 2020

Áp dụng công thức GWQI - công thức (9) cho tập dữ liệu 2 để tính các giá trị GWQI - là tích của các chỉ số thông số q_i được lũy thừa các giá trị trọng số w_i tương ứng. Đồng thời, so sánh dữ liệu trong tập dữ liệu 2 với các giá trị giới hạn trong QCVN 01-1.

Kết quả GWQI cho thấy có 21,1% trường hợp đạt loại RẤT TỐT và TỐT (49/232 trường hợp), 31,9% trường hợp thuộc loại TRUNG BÌNH (74/232 trường hợp), 33,2% trường hợp thuộc loại KÉM (77/232 trường hợp) và 13,8% trường hợp thuộc loại RẤT KÉM (32/232 trường hợp).

Kết quả trên cũng phù hợp với kết quả đánh giá CLN dưới đất dựa vào các thông số riêng biệt, tức là so sánh với giới hạn của QCVN 01-1: tổng số giá trị GWQI loại KÉM/RẤT KÉM là 47,0% tương đương với số thông số quan trắc vượt giới hạn của QCVN 01-1 là 46,1%.

4. Kết luận

Trên cơ sở áp dụng phương pháp PCA đã xác định được trọng số (w_i) của các thông số CLN một cách khách quan dựa vào đặc điểm thực tế của CLN dưới đất ở khu vực nghiên cứu. Trọng số của các thông số (ghi trong ngoặc đơn) là pH (0,10); TDS (0,14); HARD (0,12); SO4 (0,14); COD (0,07); N-NH4 (0,08); N-NO3 (0,08); Fe (0,10); Mn (0,09) và TC (0,08).

Chỉ số thông số (q_i) của các thông số CLN được xác định dựa trên tương quan tuyến tính giữa q_i và nồng độ/hàm lượng của thông số CLN, nên thuận tiện cho người dùng GWQI. Chỉ số GWQI xây dựng được phản ánh phù hợp CLN dưới đất ở tỉnh Hà Tĩnh. Có thể áp dụng nhân rộng chỉ số GWQI để đánh giá CLN khu vực nghiên cứu và các khu vực khác có đặc điểm thủy văn tương tự.

Lời cảm ơn

Nhóm tác giả chân thành cảm ơn Trung tâm Quan trắc Tài nguyên và Môi trường, Sở Tài nguyên và Môi trường tỉnh Hà Tĩnh đã cung cấp dữ liệu quan trắc nước dưới đất cho nghiên cứu này.

Tài liệu tham khảo

[1] MONRE, QCVN 08:2023/BTNMT - National Technical Regulation on Surface Water Quality, 2023, https://datafiles.chinhphu.vn/cpp/files/vbpq/2023/3/01-btnmt-qc08.pdf, in Vietnamese/, 2023 (accessed on: Jan 14th, 2024).

[2] MONRE, QCVN 09:2023/BTNMT - National Technical Regulation on Ground Water Quality, 2023, https://datafiles.chinhphu.vn/cpp/files/vbpq/2023/3/01-btnmt-qc09.pdf, in Vietnamese/, 2023 (accessed on: Jan 14th, 2024).

[3] T. Abbasi, S. A. Abbasi, Water Quality Indices, Elsevier, 2012.

[4] R. Noori, R. Berndtsson, M. Hosseinzadeh, J. F. Adamowski, M. R. Abyaneh, A Critical Review on the Application of the National Sanitation Foundation Water Quality Index, Environ, Pollut., Vol. 244, 2019, pp. 575-587, https://doi.org/10.1016/j.envpol.2018.10.076.

[5] A. A. Bordalo, R. Teixeira, W. J. Wiebe, A Water Quality Index Applied to an International Shared River Basin: The Case of the Douro River, Environ, Manage., Vol. 38, No. 6, 2006, pp. 910-920, https://doi.org/10.1007/s00267-004-0037-6.

[6] N. Adimalla, H. Qian, Groundwater Quality Evaluation using Water Quality Index (WQI) for Drinking Purposes and Human Health Risk (HHR) Assessment in an Agricultural Region of Nanganur, South India, Ecotoxicol, Environ. Saf., Vol. 176, 2019, pp. 153-161, https://doi.org/10.1016/j.ecoenv.2019.03.066.

[7] N. R. Ekere, V. E. Agbazue, B. U. Ngang, J. N. Ihedioha, Hydrochemistry and Water Quality Index of Groundwater Resources in Enugu North District, Enugu, Nigeria, Environ, Monit, Assess., Vol. 191, No. 3, 2019, https://doi.org/10.1007/s10661-019-7271-0.

[8] J. Barbosa Filho, I. B. de Oliveira, Development of a Groundwater Quality Index: GWQI, for the Aquifers of the State of Bahia, Brazil using Multivariable Analyses, Sci. Rep., Vol. 11, No. 1, 2021, https://doi.org/10.1038/s41598-021-95912-9.

[9] Vietnam Environment Administration, 1460/QD-TCMT, Decision on Promulgating Technical

Guidelines for Calculating and Publicizing Vietnam Water Quality Index (VN_WQI), Hanoi, November 12, 2019 (in Vietnamese).

A. N. Hai, B. T. Minh, N. P. T. Tuyet, V. T. H. Minh, H. T. Tan, H. T. Ngoc, L. L. Khanh, H. L. T. Hai. Application of Groundwater Quality Index (GWQI) and GIS in Groundwater Quality Zoing in Pleistocene Aquifer in Phu My Town, Ba Ria - Vung Tau Province, Science and Technology Development Journal - Science of The Earth and Environment, Vol. 4, No. 1, 2020, pp. 149-161 (in Vietnamese),

https://doi.org/10.32508/stdjsee.v4i1.525.

[10] N. S. Rao, B. Sunitha, N. Adimalla, M. Chaudhary, Quality Criteria for Groundwater use from a Rural Part of Wanaparthy District, Telangana State, India, through Ionic Spatial Distribution (ISD), Entropy Water Quality Index (EWQI) and Principal Component Analysis (PCA), Environ, Geochem, Health, Vol. 42, No. 2, 2020, pp. 579-599,

https://doi.org/10.1007/s10653-019-00393-5.

[11] H. N. Van, H. N. Viet, K. T. Trung, P. N. Hai, C. N. D. Giang, A Comprehensive Procedure to Develop Water Quality Index: A Case Study to the Huong River in Thua Thien Hue Province, Central Vietnam, PLoS One, Vol. 17, No. 9, 2022, pp. e0274673,

https://doi.org/10.1371/journal.pone.0274673.

[12] Prime Minister of Vietnam, 1163/QD-TTg, Decision Approving the Planning of Ha Tinh Province for the Period 2021-2030, with a Vision to 2050, Hanoi, November 8, 2022.

[13] Ha Tinh Provincial Statistics Agency, Ha Tinh Provincial Statistical Yearbook 2023, Ha Tinh, June, 2023.

[14] M. E. Soltan, Evaluation of Ground Water Quality in Dakhla Oasis (Egyptian Western Desert), Environ, Monit. Assess., Vol. 57, No. 2, 1999, pp. 157-168,

https://doi.org/10.1023/A:1005948930316.

[15] M. Prasad, V. Sunitha, Y. S. Reddy, B. Suvarna, B. M. Reddy, M. R. Reddy, Data on Water Quality Index Development for Groundwater Quality Assessment from Obulavaripalli Mandal, YSR district, A. P India, Data Br., Vol. 24, Jun. 2019, https://doi.org/10.1016/j.dib.2019.103846.

[16] M. Tripathi, S. K. Singal, Allocation of Weights using Factor Analysis for Development of a Novel Water Quality Index, Ecotoxicol, Environ. Saf., Vol. 183, No. August, 2019, pp. 109510, https://doi.org/10.1016/j.ecoenv.2019.109510.

[17] C. Coletti, R. Testezlaf, T. A. P. Ribeiro, R. T. G. D. Souza, D. D. A. Pereira, Water Quality Index using Multivariate Factorial Analysis, Rev. Bras. Eng. Agric, Ambient., Vol. 14, No. 5, 2010, pp. 517-522,

https://doi.org/10.1590/S1415-43662010000500009.

[18] MOH, QCVN 01-1:2018/BYT - National Technical Regulation on Domestic Water Quality, 2018 (in Vietnamese).

[19] J. N. Miller, J. C. Miller, The Quality of Analytical Measurements, 2010, http://www.tandfonline.com/doi/abs/10.1198/tech.2004.s248/, 2010 (accessed on: July 14th, 2021).

[20] C. Rivera, R. Rodriguez, Horwitz Equation as Quality Benchmark in ISO-IEC 17025 Horwitz Ratio (HorRat), IIE Annu. Conf., Vol. 5, 2010, pp. 1-6.

[21] D. J. Denis, Univariate, bivariate, and multivariate statistics using R: Quantitative tools for data analysis and data science. 2020, https://doi.org/10.1002/9781119549963.

[22] A. Afifi, S. May, R. A. Donatello, V. A. Clark, Practical Multivariate Analysis, 2019, https://doi.org/10.1201/9781315203737.