



ỨNG DỤNG PHƯƠNG PHÁP PHÂN TÍCH CỤM VÀ PHÂN TÍCH BIỆT SỐ ĐÁNH GIÁ NHIỄM MẶN TẦNG CHỨA NƯỚC PLEISTOCEN Ở HUYỆN TÂN THÀNH, TỈNH BÀ RỊA - VŨNG TÀU

Phan Nguyễn Hồng Ngọc¹, Hoàng Thị Thanh Thủy¹ và Nguyễn Hải Âu²

¹Trường Đại học Tài nguyên và Môi trường Thành phố Hồ Chí Minh

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

Thông tin chung:

Ngày nhận bài: 28/07/2017

Ngày nhận bài sửa: 19/09/2017

Ngày duyệt đăng: 26/10/2017

Title:

Application of cluster analysis and discriminant analysis assess salinity intrusion in Pleistocene aquifer of Tan Thanh district, Ba Ria - Vung Tau province, Vietnam

Từ khóa:

Nước dưới đất, phân tích cụm, phân tích biệt số, xâm nhập mặn

Keywords:

Cluster analysis, discriminant analysis, groundwater, salinity intrusion

ABSTRACT

Cluster Analysis (CA) and Discriminant Analysis (DA) were applied to assess salinity intrusion of groundwater in the Pleistocene aquifer in Tan Thanh district, Ba Ria – Vung Tau province. Groundwater samples were collected from 18 monitoring wells in April 2012. The water quality parameters selected for statistical analysis include pH, hardness, TDS, Cl⁻, F⁻, NO₃⁻, SO₄²⁻, Cr⁶⁺, Cu²⁺, Ca²⁺, Mg²⁺, Na⁺, K⁺, HCO₃⁻ and Fe²⁺.

CA groups the wells into three distinct clusters related to agricultural activities, industry, and salinization. DA revealed the existence of significant differences between these clusters, built a function discriminant analysis and determined group which causes difference among clusters in the data set. The parameters representing salinity intrusion (TDS, Mg²⁺, Cl⁻, Hardness, Na⁺, K⁺, Ca²⁺, and SO₄²⁻) accounted for 99.8% of the variance of dependent variables explained by the DA model.

The results showed that cluster analysis and discriminant analysis are the effective statistical methods in the clustering of salt intrusion.

TÓM TẮT

Phương pháp phân tích cụm (Cluster Analysis - CA) và phân tích biệt số (Discriminant Analysis - DA) đã được sử dụng nhằm đánh giá nhiễm mặn nước dưới đất ở khu vực huyện Tân Thành, tỉnh Bà Rịa – Vũng Tàu. Các mẫu nước dưới đất được thu thập từ 18 giếng quan trắc vào tháng 4 năm 2012. Các thông số chất lượng nước được lựa chọn trong kỹ thuật phân tích thống kê trên bao gồm: pH, độ cứng, TDS, Cl⁻, F⁻, NO₃⁻, SO₄²⁻, Cr⁶⁺, Cu²⁺, Ca²⁺, Mg²⁺, Na⁺, K⁺, HCO₃⁻ và Fe²⁺.

Kết quả phân tích CA nhóm bộ dữ liệu quan trắc thành ba cụm giếng có liên quan đến hoạt động nông nghiệp, hoạt động công nghiệp và sự nhiễm mặn. Phân tích biệt số được thực hiện nhằm nghiên cứu sự tồn tại khác biệt có ý nghĩa giữa các cụm, xây dựng hàm phân tích phân biệt và xác định cụm gây ra sự khác biệt giữa các nhóm cụm. Các thông số đại diện cho yếu tố xâm nhập mặn (TDS, Mg²⁺, Cl⁻, độ cứng, Na⁺, K⁺, Ca²⁺, SO₄²⁻) chiếm 99,8% phương sai các biến phụ thuộc cụm được giải thích bởi mô hình DA.

Kết quả nghiên cứu đã cho thấy phân tích cụm và biệt số là những phương pháp thống kê hiệu quả trong phân vùng xâm nhập mặn.

Trích dẫn: Phan Nguyễn Hồng Ngọc, Hoàng Thị Thanh Thủy và Nguyễn Hải Âu, 2017. Ứng dụng phương pháp phân tích cụm và phân tích biệt số đánh giá nhiễm mặn tầng chứa nước pleistocen ở huyện Tân Thành, tỉnh Bà Rịa - Vũng Tàu. Tạp chí Khoa học Trường Đại học Cần Thơ. Số chuyên đề: Môi trường và Biến đổi khí hậu (2): 129-136.

1 GIỚI THIỆU

Phân tích thống kê đa biến (MSA - Multivariate Statistics Analysis) bao gồm các kỹ thuật thống kê khác nhau, chẳng hạn như phân tích cụm (CA - Cluster Analysis) và phân tích biệt số (DA - Discriminant Analysis - DA), phân tích nhân tố (FA - Factor Analysis), phân tích thành phần chính (PCA - Principal Component Analysis),... Các phương pháp này giúp giải thích các ma trận dữ liệu phức tạp để hiểu rõ hơn về chất lượng nước, cho phép xác định các nhân tố hoặc nguồn ảnh hưởng đến chất lượng nước và cung cấp một công cụ có giá trị để quản lý và khai thác sử dụng tài nguyên nước.

Một trong những thuận lợi chính của những kỹ thuật này là khả năng phân tích dữ liệu lớn và phức tạp, có nhiều biến và nhiều đơn vị thí nghiệm. Những phương pháp này đôi khi tạo ra các biến mới bằng cách giảm số lượng của các biến ban đầu trong việc so sánh và giải thích các dữ liệu. Phân tích cụm (CA) là phương pháp phân loại các đối tượng hay các biến sao cho các đối tượng trong cùng một cụm xét theo các đặc tính được chọn để phân tích. Phân tích phân biệt (DA) là phương pháp phân tích dữ liệu khi biến phụ thuộc là biến phân loại và biến độc lập là biến định lượng. Phân tích phân biệt đòi hỏi phải biết trước số nhóm và các đối tượng trong mỗi nhóm trước khi tiến hành phân tích, xây dựng hàm phân tích phân biệt, nghiên cứu sự tồn tại khác biệt có ý nghĩa giữa các nhóm, xác định biến độc lập là nguyên nhân lớn nhất gây ra sự khác biệt giữa các nhóm và phân loại các quan sát vào một nhóm nhất định dựa vào giá trị của các biến độc lập.

Trong những năm gần đây, các phương pháp phân tích cụm và phân tích phân biệt đã được sử dụng rộng rãi trong các ứng dụng môi trường, bao gồm các đánh giá quan trắc diễn biến chất lượng nước ngầm, nước mặt, kiểm tra kết quả các mô hình mô phỏng chất lượng nước theo không gian và thời gian, xác định các yếu tố hóa học liên quan đến các điều kiện thủy văn, và đánh giá các chỉ thị chất lượng môi trường. Ở Mỹ và các nước châu Âu như Pháp, Thổ Nhĩ Kỳ (Varol *et al.*, 2012) và các quốc gia ở châu Á như Nhật Bản (Shrestha *et al.*, 2007), Malaysia (Kura *et al.*, 2013), Ấn Độ (Mavukkandy *et al.*, 2014), Trung Quốc (Yang *et al.*, 2015), các nghiên cứu này đã ứng dụng nhiều phương pháp thống kê đa biến đánh giá chất lượng nước mặt, nước dưới đất ở các lưu vực sông dựa vào mối quan hệ giữa các thông số đo đạc với các đặc điểm nguồn thải, đề xuất được các thông số đặc trưng chất lượng nước của lưu vực để giám sát và quản lý hiệu quả.

Ở Việt Nam, các kỹ thuật thống kê đa biến cũng được sử dụng rộng rãi trong nhiều lĩnh vực khác nhau như tâm lý, kinh tế, xã hội, kỹ thuật trong đó có lĩnh vực môi trường. Các nghiên cứu trong lĩnh vực môi trường chủ yếu sử dụng phương pháp phân tích hồi quy tuyến tính để xử lý các số liệu phân tích thí nghiệm, kiểm định thông số mô hình mô phỏng. Phân tích đánh giá chất lượng nước chỉ dựa vào phương pháp so sánh với qui chuẩn cho phép, phân vùng dựa vào chỉ số chất lượng nước (WQI) và mô hình toán, chưa có nghiên cứu nào sử dụng các kỹ thuật phân tích thống kê đa biến đánh giá chất lượng nước một cách đầy đủ, riêng biệt.

Huyện Tân Thành đã và đang trở thành một trong ba địa phương có nền kinh tế phát triển bậc nhất tỉnh Bà Rịa – Vũng Tàu, là nơi tập trung nhiều khu công nghiệp nhất của tỉnh. Nhu cầu sử dụng nước trên địa bàn huyện là khá lớn để phục vụ phát triển kinh tế, trong khi đó các nguồn khai thác nước mặt từ sông hồ không đáp ứng nhu cầu đó. Theo số liệu điều tra năm 2012 từ Sở Tài Nguyên và Môi trường tỉnh Bà Rịa - Vũng Tàu thì tổng lưu lượng khai thác nước dưới đất trung bình là 13.174.310m³/năm, trong đó cấp nước sinh hoạt là 5.434.120 m³/năm, nước được khai thác chủ yếu trong tầng chứa nước lỗ hổng các trầm tích Pleistocen. Tầng chứa nước lỗ hổng Pleistocen được tạo thành từ đất đá hạt thô nằm dưới hệ tầng Cù Chi, hệ tầng Thủ Đức và hệ tầng Trảng Bom với các khoáng vật chính: Fluorit-apatit, feldpat, thạch cao, turmalin, montmorilonit, ilmenit và một số tạp chất khác. Hai nguyên nhân chính ảnh hưởng đến chất lượng nước dưới đất ở khu vực nghiên cứu là do phía Tây huyện Tân Thành nằm dọc theo sông Thị Vải và hoạt động khai thác nước dưới đất từ các trạm cấp nước lớn (Phú Mỹ - Mỹ Xuân, Tóc Tiên) và nhỏ lẻ thuộc địa bàn; một số giếng có độ tổng khoáng hóa tăng cao cho thấy chất lượng nước dưới đất có khả năng chịu sự tác động của quá trình biên mặn lấn sâu vào đất liền và hoạt động phát triển kinh tế - xã hội trên lưu vực.

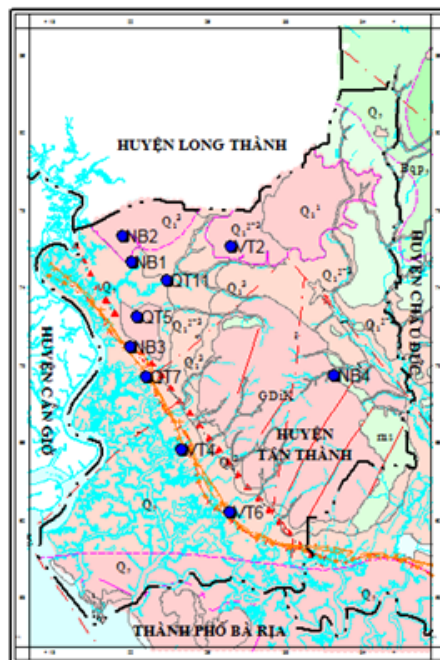
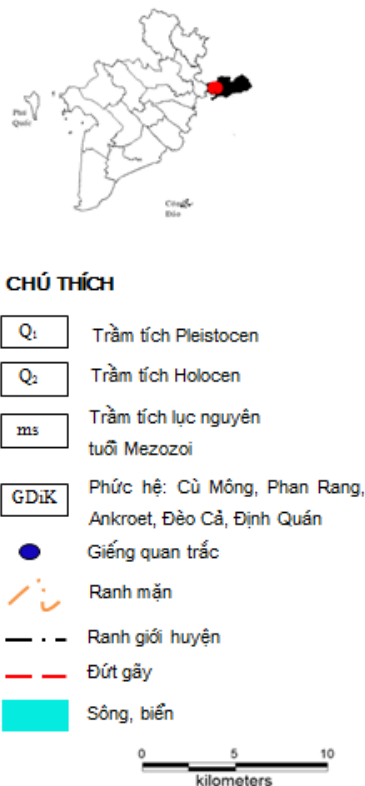
Trong nghiên cứu này, các thông số lý hóa của chất lượng nước dưới đất trong tầng chứa nước Pleistocen khu vực huyện Tân Thành, tỉnh Bà Rịa – Vũng Tàu được tập trung phân tích bằng các phương pháp thống kê đa biến (CA và DA). Kết quả phân tích cho phép đánh giá thông tin về sự tương đồng trong các trạm quan trắc khác nhau, từ đó xác định các thông số chất lượng nước đặc trưng theo không gian, tác động của các nguồn ô nhiễm trên lưu vực đến chất lượng nước.

2 TÀI LIỆU VÀ PHƯƠNG PHÁP NGHIÊN CỨU

2.1 Mô tả vùng nghiên cứu

Huyện Tân Thành thuộc tỉnh Bà Rịa – Vũng Tàu, phía Đông giáp huyện Châu Đức, phía Tây giáp huyện Cần Giờ và thành phố Vũng Tàu, phía Nam giáp thành phố Bà Rịa và phía Bắc giáp

huyện Long Thành, tỉnh Đồng Nai. Diện tích tự nhiên (33.825 ha), dân số trung bình (137.334 người). Mùa khô kéo dài từ tháng 12 đến tháng 4 năm sau, có nền nhiệt độ trung bình hàng năm cao (26,3°C) và hầu như không thay đổi nhiều trong năm. Mùa mưa bắt đầu từ tháng 5 đến tháng 11, lượng mưa phân bố đều trong năm (trung bình 1356.5 mm/năm).



Hình 1: Sơ đồ vị trí quan trắc huyện Tân Thành, tỉnh Bà Rịa – Vũng Tàu

2.2 Tài liệu nghiên cứu

Trong bài báo này, 15 thông số chất lượng nước (Ca^{2+} , Mg^{2+} , Na^+ , K^+ , HCO_3^- , pH, độ cứng, TDS, Cl^- , F^- , NO_3^- , SO_4^{2-} , Cr^{6+} , Cu^{2+} và Fe^{2+}) từ 18 giếng quan trắc chất lượng nước dưới đất trên địa bàn huyện Tân Thành được Sở Tài nguyên và Môi trường tỉnh Bà Rịa – Vũng Tàu thực hiện vào mùa khô năm 2012 (NB2C, NB2A, NB3A, NB1B, NB4, VT2B, NB1A, VT2A, VT6, QT5B, NB2B, QT11, QT7B, NB3B, QT5A, VT4B, QT7A, VT4A) được lựa chọn xử lý và đánh giá. Vị trí các giếng quan trắc được trình bày trong sơ đồ vị trí quan trắc ở Hình 1.

2.3 Phương pháp phân tích thống kê đa biến và xử lý số liệu

Tất cả các tính toán toán học và thống kê được thực hiện bằng cách sử dụng phần mềm Excel 2013 (Microsoft Office). Các phân tích CA,

DA được xử lý bằng phần mềm SPSS (PAS Statistics 21).

2.3.1 Phân tích cụm (Clustering Analysis - CA)

Trong nghiên cứu này, phương pháp phân tích CA được lựa chọn là phương pháp phân tích cụm tích tụ dựa vào phương sai là “thủ tục Ward” trong loại thủ tục phân cụm thứ bậc (Hierarchical clustering). Theo thủ tục Ward thì ta sẽ tính giá trị trung bình tất cả các biến cho từng cụm một; sau đó, tính khoảng cách Euclid bình phương (Squared Euclidean distance) giữa các phần tử trong cụm với giá trị trung bình của cụm, rồi lấy tổng tất cả các khoảng cách bình phương trong nội bộ cụm nếu kết hợp với nhau là nhỏ nhất sẽ được kết hợp. Cụ thể hơn, trong phương pháp này khoảng cách hoặc sự

giống nhau giữa 2 nhóm A và B được xem là khoảng cách nhỏ nhất giữa 2 điểm A và B:

$$D(A,B) = \text{Range}\{d(x_i,x_j), \text{với } x_i \in A \text{ và } x_j \in B\} \quad (1)$$

Khi $d(x_i,x_j)$ là khoảng cách Euclid bình phương trong công thức (1). Tại mỗi bước khoảng cách được tìm thấy từng cặp nhóm và 2 nhóm có khoảng cách nhỏ nhất (giống nhau nhiều nhất) được gộp lại. Sau khi 2 nhóm được gộp lại, tiếp tục lặp lại các bước tiếp theo: khoảng cách giữa tất cả các cặp nhóm được tính lại lần nữa, và cặp có khoảng cách ngắn nhất được gộp vào nhóm đơn. Kết quả của việc phân nhóm cấu trúc được biểu diễn bằng đồ thị - biểu đồ hình cây (Nguyễn Hải Âu và *ctv.*, 2014).

2.3.2 Phân tích biệt số (Discriminant Analysis - DA)

Phương pháp phân tích phân biệt (DA) là kỹ thuật phân tích dữ liệu khi biến phụ thuộc là biến phân loại và biến độc lập là biến định lượng. Mục tiêu của phân tích biệt số là việc xây dựng các hàm phân tích phân biệt hay một hàm tuyến tính kết hợp các biến độc lập sao cho phân biệt rõ các biểu hiện của biến phụ thuộc, nghiên cứu sự tồn tại khác biệt có ý nghĩa giữa các nhóm xét theo biến độc lập, xác định biến độc lập nào là nguyên nhân chính gây ra sự khác biệt giữa các nhóm và phân loại các quan sát này vào trong một nhóm nhất định dựa vào các giá trị của biến độc lập. Mô hình phân tích biệt số có dạng tuyến tính như sau:

$$D = b_0 + b_1X_1 + b_2X_2 + \dots + b_kX_k \quad (2)$$

Trong đó, D là biệt số, b là hệ số hay trọng số phân biệt và X là biến độc lập. Các hệ số hay trọng số được tính toán sao cho các nhóm có các giá trị của hàm phân biệt (biệt số D) khác nhau càng nhiều càng tốt. Điều này xảy ra khi tỉ lệ của tổng các độ lệch bình phương của biệt số giữa các nhóm so với tổng các độ lệch bình phương của biệt số trong nội bộ nhóm đạt cực đại (Hoàng Trọng và *ctv.*, 2008).

3 KẾT QUẢ VÀ THẢO LUẬN

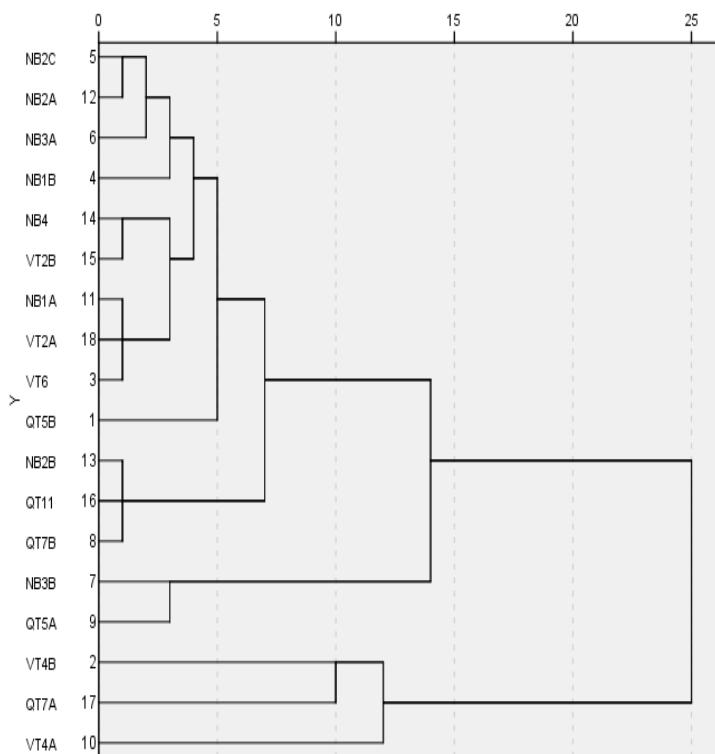
3.1 Sự biến thiên theo không gian các cụm giếng

Dựa vào biểu đồ phân tích cụm (Hình 2) và kết quả phân tích phân bố các cụm giếng quan trắc (Bảng 2) cho thấy chất lượng nước của 18 giếng quan trắc được phân thành 3 cụm như sau:

– Cụm 1 gồm 13 giếng với giá trị trung bình NO_3^- vượt trội hơn so với hai cụm còn lại, hiện tượng này có thể kết luận rằng các giếng thuộc cụm 1 chịu ảnh hưởng khá lớn bởi hoạt động nông nghiệp diễn ra xung quanh khu vực quan trắc. Các giếng NB2C, NB2A, NB1B, VT2B, VT2A, NB1A và QT11 nằm ở vị trí gần khu vực trồng cây công nghiệp; giếng VT6 tọa lạc ở khu vực nuôi trồng thủy sản, chịu ảnh hưởng bởi các chất hữu cơ từ lượng thức ăn tàn dư trong quá trình nuôi trồng; giếng NB4 nằm ngay khu vực bãi rác Tóc Tiên với lượng hữu cơ phân hủy lớn. Tuy nhiên, giá trị trung bình NO_3^- đạt 0,827 mg/l vẫn nằm trong quy chuẩn cho phép (<15 mg/l) do đó chất lượng nước thuộc các giếng ở cụm 1 vẫn còn khá tốt (Yang, *et al.*, 2015).

– Cụm 2 gồm 2 giếng có giá trị Cr^{6+} và Cu^{2+} cao hơn so với hai cụm còn lại. Khả năng khá lớn khi nhận định rằng cụm 2 bị ảnh hưởng bởi quá trình xả thải của các khu công nghiệp bởi vị trí hiện diện của hai giếng (NB3B, QT5A) thuộc cụm 2 là rất gần với các khu công nghiệp nặng (Mỹ Xuân B1).

Cụm 3 gồm 3 giếng mang giá trị Mg^{2+} , Na^+ , K^+ , độ cứng, TDS, Cl^- , F^- , SO_4^{2-} và Fe^{2+} vượt trội hơn rất nhiều so với hai cụm đầu tiên. Đặc biệt, giá trị Cl^- và Fe^{2+} còn vượt cả quy chuẩn cho phép với Cl^- vượt gấp 2,4 lần và Fe^{2+} vượt 5,9 lần cho thấy rằng cụm 3 có khả năng chịu ảnh hưởng của quá trình xâm nhập mặn lên các giếng có vị trí gần bờ biển cũng như gần các ranh mặn.



Hình 2: Biểu đồ phân tích cụm

Bảng 2: Bảng phân bố các cụm giếng quan trắc

Thông số quan trắc	Đơn vị	Cụm 1	Cụm 2	Cụm 3
		NB2C, NB2A, NB3A, NB1B, NB4, VT2B, NB1A, VT2A, VT6, QT5B, NB2B, QT11, QT7B	NB3B, QT5A	VT4B, QT7A, VT4A
		Giá trị trung bình		
Ca ²⁺	mg/l	17,3	14,7	179,358
Mg ²⁺	mg/l	9,9	2,250	89,072
Na ⁺	mg/l	69,6	7,340	681,878
K ⁺	mg/l	5,9	0,935	53,183
HCO ₃ ⁻	mg/l	54,17	42,715	96,615
pH	mg/l	6,25	4,76	5,66
Độ cứng	mg/l	96,21	13,75	407,83
TDS	mg/l	216	102	1041
Cl ⁻	mg/l	90,24	27,47	608,55
F ⁻	mg/l	0,275	0,105	0,59
NO ₃ ⁻	mg/l	0,827	0,566	0,363
SO ₄ ²⁻	mg/l	48,28	36,98	72,84
Cr ⁶⁺	mg/l	0,009	0,047	0,010
Cu ²⁺	mg/l	0,027	0,102	0,044
Fe ²⁺	mg/l	7,229	1,390	29,517

3.2 Xác định các biến phân biệt có ý nghĩa giữa các cụm

Từ kết quả phân tích DA cho thấy các biến thuộc yếu tố nhiễm mặn (TDS, Mg²⁺, Cl⁻, độ cứng, Na⁺, K⁺, Ca²⁺, SO₄²⁻) chiếm ưu thế hệ số chuẩn hóa, hệ số chuẩn hóa càng lớn thì đóng góp nhiều vào khả năng phân biệt của hàm, các biến này chủ

yếu có liên hệ chặt chẽ với hàm thứ nhất, cụ thể là yếu tố nhiễm mặn (Bảng 3 và 4).

Giá trị Eigenvalue tương ứng của hàm 1 lớn và chiếm 99,1% phương sai giải thích kết quả. Hệ số Canonical tương ứng là 0,999, bình phương của hệ số này là (0,999)² = 0,998 cho thấy 99,8% phương sai của biến phụ thuộc Cụm được giải thích bởi mô hình DA (Bảng 5).

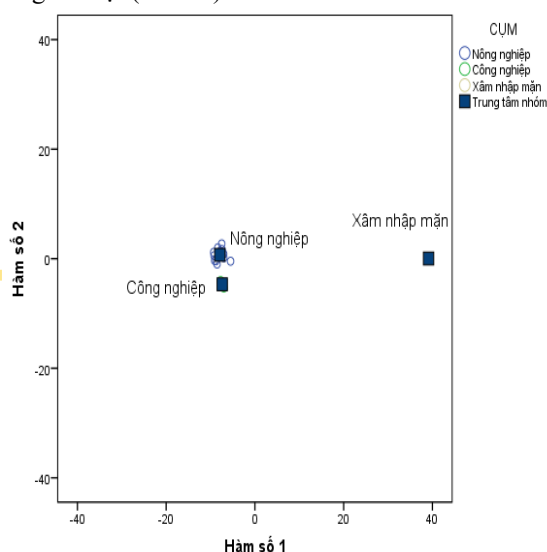
Bảng 3: Ma trận cấu trúc trong DA

	Hàm số	
	1	2
NO ₃ ⁻	0,241*	0,129
TDS	0,150*	0,021
Cu ²⁺	0,138*	-0,113
Mg ²⁺	0,094*	0,018
Cl ⁻	0,091*	0,004
Độ cứng	0,089*	0,083
Na ⁺	0,077*	0,028
K ⁺	0,064*	0,033
Ca ²⁺	0,044*	-0,001
HCO ₃ ⁻	0,015*	0,001
pH	-0,017	0,435*
Cr ⁶⁺	-0,008	-0,088*
SO ₄ ²⁻	0,025	0,081*
F ⁻	0,023	0,060*
Fe ²⁺	0,007	0,052*

Bảng 5: Bảng Hệ số Eigenvalue trong DA

Hàm	Eigenvalue	% Phương sai	% Phương sai trích	Tương quan Canonical
1	367.556	99,1	99,1	0,999
2	3.350	0,9	100	0,878

Biểu đồ phân tán của các nhóm ảnh hưởng đến chất lượng nước dưới đất (Nông nghiệp, Công nghiệp và Xâm nhập mặn) được diễn tả theo hàm thứ nhất và hàm thứ hai của biểu đồ. Có thể thấy rõ rằng nhóm 3 (Xâm nhập mặn) có trị số cao nhất theo hàm thứ nhất so với hai nhóm còn lại, vì hàm thứ nhất chủ yếu gắn với các biến (TDS, Mg²⁺, Cl⁻, độ cứng, Na⁺, K⁺, Ca²⁺, SO₄²⁻) nên có thể thấy rằng nhóm 3 được phân biệt một cách rõ ràng theo các biến này. Những giếng mang giá trị cao ở hàm thứ hai có nguy cơ bị nhiễm mặn nhiều hơn so với các giếng còn lại (Hình 3).

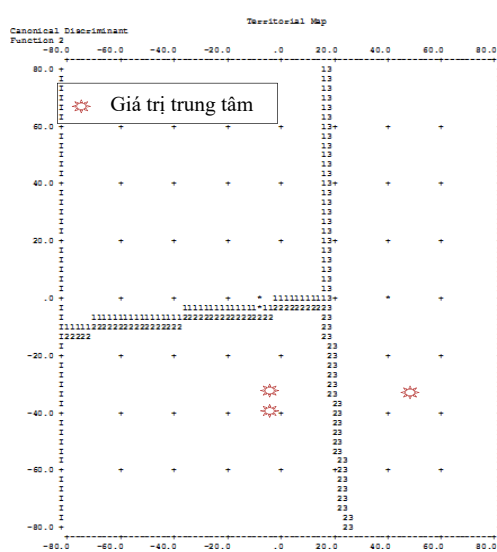


Hình 3: Sự phân bố các cụm phân biệt

Bảng 4: Hệ số hàm phân biệt

	CỤM		
	Nông nghiệp	Công nghiệp	Nhiễm mặn
Ca ²⁺	0,852	0,900	-1,060
Mg ²⁺	-4,664	-2,355	54,613
Na ⁺	0,140	0,003	-6,703
K ⁺	-0,735	-2,663	16,933
HCO ₃ ⁻	-0,251	-0,303	-1,659
pH	62,914	50,306	16,675
Độ cứng	-0,671	-0,677	1,466
TDS	0,095	0,156	-1,108
Cl ⁻	0,293	0,218	2,626
F ⁻	75,844	55,220	273,652
SO ₄ ²⁻	0,157	0,111	0,524
(Constant)	-201,036	-127,917	-1093,672

Biểu đồ vị trí Territorial Map (Hình 4) biểu diễn giá trị trung tâm (centroid) của mỗi nhóm. Ranh giới của các nhóm được thể hiện bằng các con số 1, 2, 3 tương ứng với cụm Nông nghiệp, Công nghiệp và Xâm nhập mặn. Có thể thấy rằng trung tâm của nhóm 3 (Xâm nhập mặn) nằm tại vị trí ổn định và ít bị lẫn so với hai nhóm còn lại, dấu hiệu này cho thấy nhóm 3 mang tính đại diện cao nhất cho bộ dữ liệu quan trắc với ba giếng đặc trưng cho yếu tố nhiễm mặn (VT4A, VT4B và QT7A) của chất lượng nước dưới đất năm 2012.



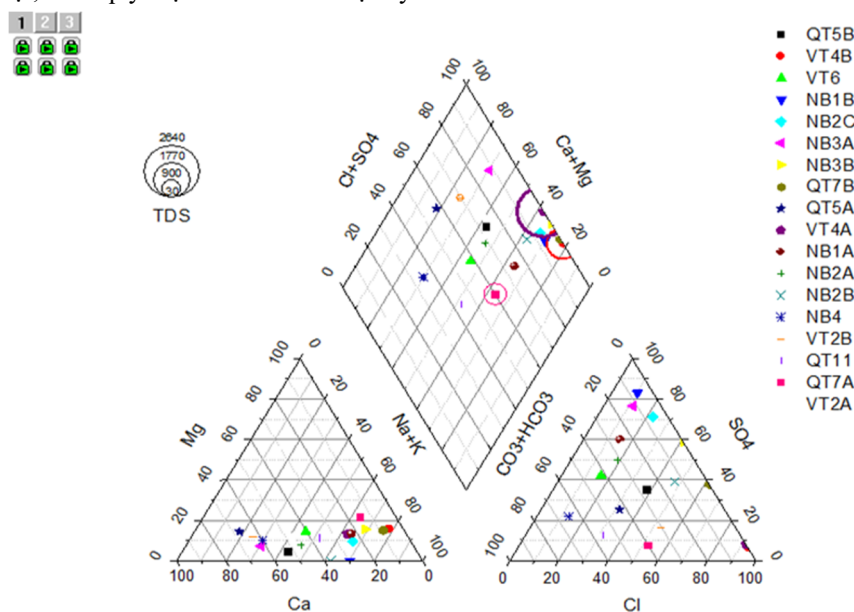
Hình 4: Biểu đồ thể hiện giá trị trung tâm của mỗi nhóm

Sự nhiễm mặn thể hiện ở hàm lượng cao của TDS cũng như xu hướng tập trung cao các ion Cl^- , SO_4^{2-} , Na^+ , K^+ , Mg^{2+} và một số muối sắt hiện hữu. Bên cạnh đó, vào mùa khô, lượng nước bổ cấp cho tầng chứa nước là rất ít, hầu như không xảy ra mạnh mẽ như mùa mưa. Một nguyên nhân nữa là do mực nước các vùng sông hồ giảm gây khó khăn trong việc khai thác, do đó tần suất khai thác nước dưới đất phục vụ cho sinh hoạt cũng như các hoạt động nông nghiệp, công nghiệp diễn ra với cường độ lớn, góp phần gây suy giảm chất lượng và cạn kiệt nước dưới đất. VT4A, VT4B, QT7A là các giếng quan trắc có quan hệ thủy lực với nước sông, có đặc điểm là tầng chứa nước nằm nông, lớp cách nước có nguồn gốc sông - biển và nằm ngay ranh giới mặn - nhạt, theo quy luật thì mùa mưa bị đẩy

ra biển và mùa khô xâm nhập mặn sẽ lại tiến sâu vào đất liền gây ảnh hưởng đến chất lượng nước dưới đất khu vực nghiên cứu.

3.3 Sự phân bố các ion chính

Hầu hết các giếng quan trắc đều phân bố ở khu vực thuộc loại nước Ca^{2+} - Mg^{2+} , Na^+ - K^+ và loại nước Cl^- - SO_4^{2-} (Hình 5). Hơn thế nữa, các giếng quan trắc còn có xu hướng trội hơn nghiêng về loại nước Cl^- - SO_4^{2-} , đặc biệt một số giếng quan trắc còn mang hàm lượng TDS rất lớn (VT4A, VT4B). Từ đó có thể giả định được rằng kết quả trên có thể là do sự hạ thấp mực nước dưới đất vào mùa khô và quá trình nhiễm mặn đã ảnh hưởng nhất định đến các giếng quan trắc phân bố ở khu vực gần bờ biển.



Hình 5: Biểu đồ phân bố các ion chính

4 KẾT LUẬN

Kỹ thuật thống kê đa biến được ứng dụng trong nghiên cứu này như một công cụ hữu hiệu trong việc phân tích và giải thích chất lượng nước, giúp các nhà quản lý hiểu rõ hơn về sự biến đổi theo không gian của chất lượng nước dưới đất, từ đó đưa ra được các giải pháp nhằm quản lý bền vững nguồn tài nguyên nước. Kết quả phân tích cụm (CA) đã nhóm 18 giếng quan trắc thành 3 cụm đặc trưng cho chất lượng nước của các giếng quan trắc (hoạt động nông nghiệp, hoạt động công nghiệp và nhiễm mặn) dựa vào đặc tính tương đồng của bộ dữ liệu. Phân tích biệt số (DA) tiếp tục kiểm định kết quả phân tích cụm bằng cách xác định các biến phân biệt có ý nghĩa giữa các cụm, giải thích được

99,8% phương sai các biến và chỉ ra nhóm 3 (VT4A, VT4B, QT7A) mang tính đại diện cao nhất cho bộ dữ liệu ban đầu.

LỜI CẢM ƠN

Nghiên cứu này được thực hiện bởi sự hỗ trợ dữ liệu từ Sở Tài nguyên và Môi trường tỉnh Bà Rịa – Vũng Tàu.

TÀI LIỆU THAM KHẢO

- Hoàng Trọng và Chu Nguyễn Mộng Ngọc., 2008. Phân tích dữ liệu nghiên cứu với SPSS. Nhà xuất bản Hồng Đức, TP.HCM, 179 trang.
- Kura, N. U., Ramli, M. F., Sulaiman, W. N. A., Ibrahim, S., Aris, A. Z., and Mustapha, A., 2013. Evaluation of factors influencing the groundwater

- chemistry in a small tropical island of Malaysia. International journal of environmental research and public health. 10: 1861-1881.
- Mavukkandy, M. O., Karmakar, S., and Harikumar, P., 2014. Assessment and rationalization of water quality monitoring network: a multivariate statistical approach to the Kabbini River (India). Environmental Science and Pollution Research. 21: 10045-10066.
- Nguyễn Hải Âu và Vũ Văn Nghi, 2014. Bước đầu áp dụng kỹ thuật phân tích thống kê đa biến phân tích số liệu chất lượng nước lưu vực sông Thị Tính, tỉnh Bình Dương. Tạp chí Phát triển Khoa học và Công nghệ của Viện Hàn Lâm Khoa học và Công nghệ Việt Nam. 52: 190-199.
- Shrestha, S. and Kazama, F., 2007. Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji river basin, Japan. Environmental Modelling & Software. 22: 464-475.
- Sở Tài nguyên và Môi trường tỉnh Bà Rịa – Vũng Tàu, 2013. Báo cáo “Vận hành mạng quan trắc nước dưới đất tỉnh Bà Rịa – Vũng Tàu giai đoạn 2012 – 2013.
- Varol, M. and Şen, B., 2012. Assessment of nutrient and heavy metal contamination in surface water and sediments of the upper Tigris River, Turkey. Catena. 92: 1-10.
- Yang, Q., Zhang, J., Wang, Y., Fang, Y., and Martín, J. D., 2015. Multivariate statistical analysis of hydrochemical data for shallow ground water quality factor identification in a coastal aquifer. Polish Journal of Environmental Studies. 24: 102-11