

<sup>1</sup>TRƯỜNG ĐẠI HỌC TÀI NGUYÊN VÀ MÔI TRƯỜNG TP. HCM  
<sup>2</sup>VIỆN MÔI TRƯỜNG VÀ TÀI NGUYÊN, ĐẠI HỌC QUỐC GIA TP. HCM  
Phan Nguyễn Hồng Ngọc<sup>1</sup>, Hoàng Thị Thanh Thủy<sup>1</sup>, Nguyễn Hải Âu<sup>2</sup>

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

Người trình bày: Phan Nguyễn Hồng Ngọc

TP Cần Thơ, tháng 11/2017

# NỘI DUNG TRÌNH BÀY

---

1

- GIỚI THIỆU CHUNG

2

- TÀI LIỆU VÀ PHƯƠNG PHÁP NGHIÊN CỨU

3

- KẾT QUẢ VÀ THẢO LUẬN

4

- KẾT LUẬN

# GIỚI THIỆU CHUNG

# GIỚI THIỆU CHUNG

Phân tích thống kê đa biến:

- Phân tích cụm  
(CA - Cluster Analysis)
- Phân tích biệt số  
(DA - Discriminant Analysis - DA)

1

Khả năng phân tích dữ liệu lớn và phức tạp, có nhiều biến và nhiều đơn vị thí nghiệm.

2

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

- 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.

- 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.

# GIỚI THIỆU CHUNG

## Các nghiên cứu trong lĩnh vực môi trường ở Việt Nam:

- 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. 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 **18.608.430 m<sup>3</sup>/năm** được khai thác chủ yếu trong tầng chứa nước lỗ hổng các trầm tích Pleistocen

# **TÀI LIỆU VÀ PHƯƠNG PHÁP NGHIÊN CỨU**

# TÀI LIỆU VÀ PHƯƠNG PHÁP 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).
- Nền nhiệt độ trung bình hàng năm cao (26,3°C)
- Lượng mưa phân bố đều trong năm (trung bình 1356.5 mm/năm)

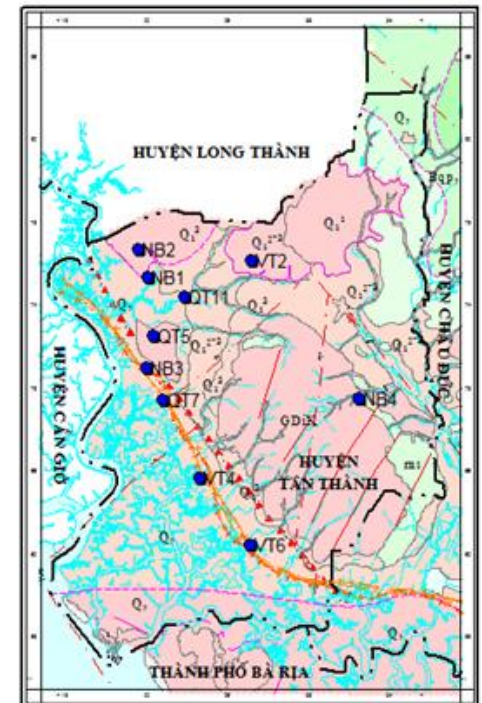
**15** thông số chất lượng nước từ **18** giếng quan trắc chất lượng nước dưới đất trên địa bàn huyện Tân Thành được Sở Tài nguyên và Môi trường tỉnh Bà Rịa – Vũng Tàu thực hiện vào mùa khô năm 2012 được lựa chọn xử lý và đánh giá



## CHÚ THÍCH

- Q<sub>1</sub>: Trầm tích Pleistocen
- Q<sub>2</sub>: Trầm tích Holocen
- ms: Trầm tích lục nguyên tuổi Mesozoi
- GDiK: Phức hệ: Cù Mông, Phan Rang, Ankreot, Đèo Cả, Định Quán
- : Giếng quan trắc
- : Ranh mặn
- - -: Ranh giới huyện
- - -: Đứt gãy
- : Sông, biển

0 5 10  
kilometers



# TÀI LIỆU VÀ PHƯƠNG PHÁP NGHIÊN CỨU

## CA Cluster Analysis

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

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

Khi  $d(x_i,x_j)$  là khoảng cách Euclid bình phương trong công

## DA Discriminant Analysis

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 là 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

$$D = b_0 + b_1X_1 + b_2X_2 + \dots + b_kX_k$$

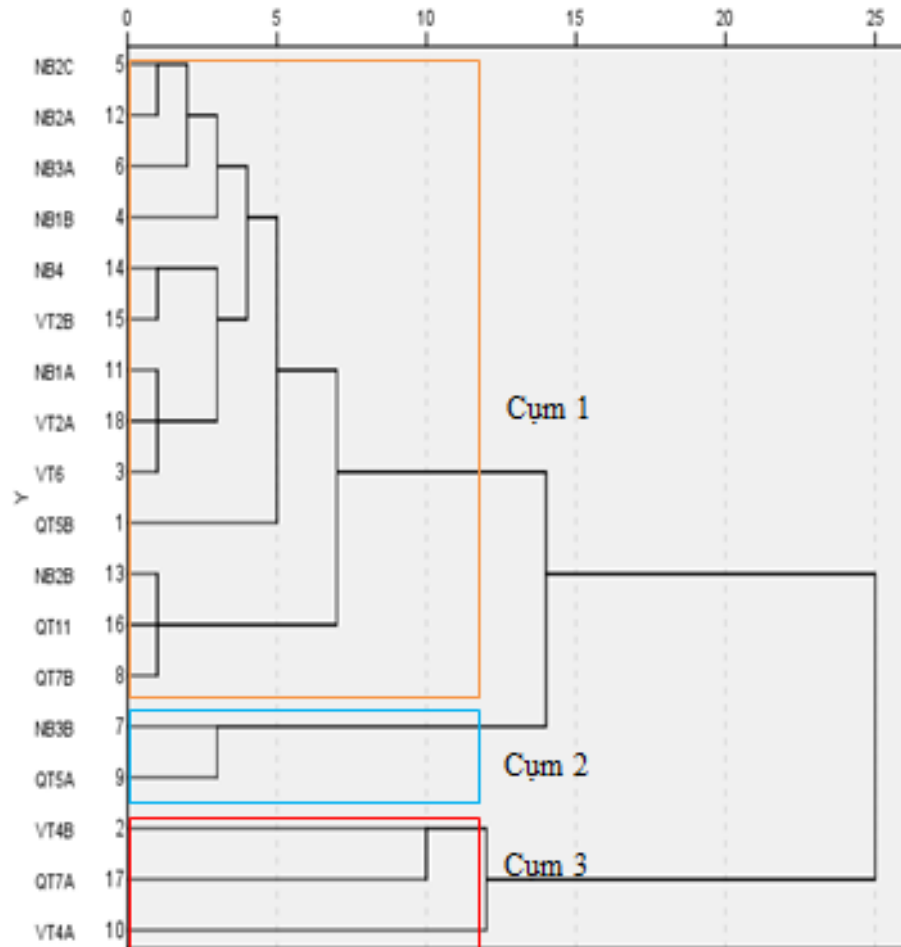
Trong đó  $D$  là biệt số,  $b$  là hệ số hay trọng số phân biệt và  $X$  là biến độc lập



# KẾT QUẢ VÀ THẢO LUẬN

# KẾT QUẢ VÀ THẢO LUẬN

Sự biến thiên theo không gian các cụm giếng



# KẾT QUẢ VÀ THẢO LUẬN

Sự biến thiên theo không gian các cụm giếng

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 <sup>2+</sup>	mg/l	17,3	14,7	<b>179,358</b>
Mg <sup>2+</sup>	mg/l	9,9	2,250	<b>89,072</b>
Na <sup>+</sup>	mg/l	69,6	7,340	<b>681,878</b>
K <sup>+</sup>	mg/l	5,9	0,935	<b>53,183</b>
HCO <sub>3</sub> <sup>-</sup>	mg/l	54,17	42,715	<b>96,615</b>
pH	mg/l	6,25	4,76	5,66
Độ cứng	mg/l	96,21	13,75	<b>407,83</b>
TDS	mg/l	216	102	<b>1041</b>
Cl <sup>-</sup>	mg/l	90,24	27,47	<b>608,55</b>
F <sup>-</sup>	mg/l	0,275	0,105	0,59
NO <sub>3</sub> <sup>-</sup>	mg/l	<b>0,827</b>	0,566	0,363
SO <sub>4</sub> <sup>2-</sup>	mg/l	48,28	36,98	<b>72,84</b>
Cr <sup>6+</sup>	mg/l	0,009	<b>0,047</b>	0,010
Cu <sup>2+</sup>	mg/l	0,027	<b>0,102</b>	0,044
Fe <sup>2+</sup>	mg/l	7,229	1,390	<b>29,517</b>

# KẾT QUẢ VÀ THẢO LUẬN

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

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

# KẾT QUẢ VÀ THẢO LUẬN

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

	CỤM		
	Nông nghiệp	Công nghiệp	Nhiễm mặn
Ca <sup>2+</sup>	0,852	0,900	-1,060
Mg <sup>2+</sup>	-4,664	-2,355	54,613
Na <sup>+</sup>	0,140	0,003	-6,703
K <sup>+</sup>	-0,735	-2,663	16,933
HCO <sub>3</sub> <sup>-</sup>	-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 <sup>-</sup>	0,293	0,218	2,626
F <sup>-</sup>	75,844	55,220	273,652
SO <sub>4</sub> <sup>2-</sup>	0,157	0,111	0,524
(Constant)	-201,036	-127,917	-1093,672

# KẾT QUẢ VÀ THẢO LUẬN

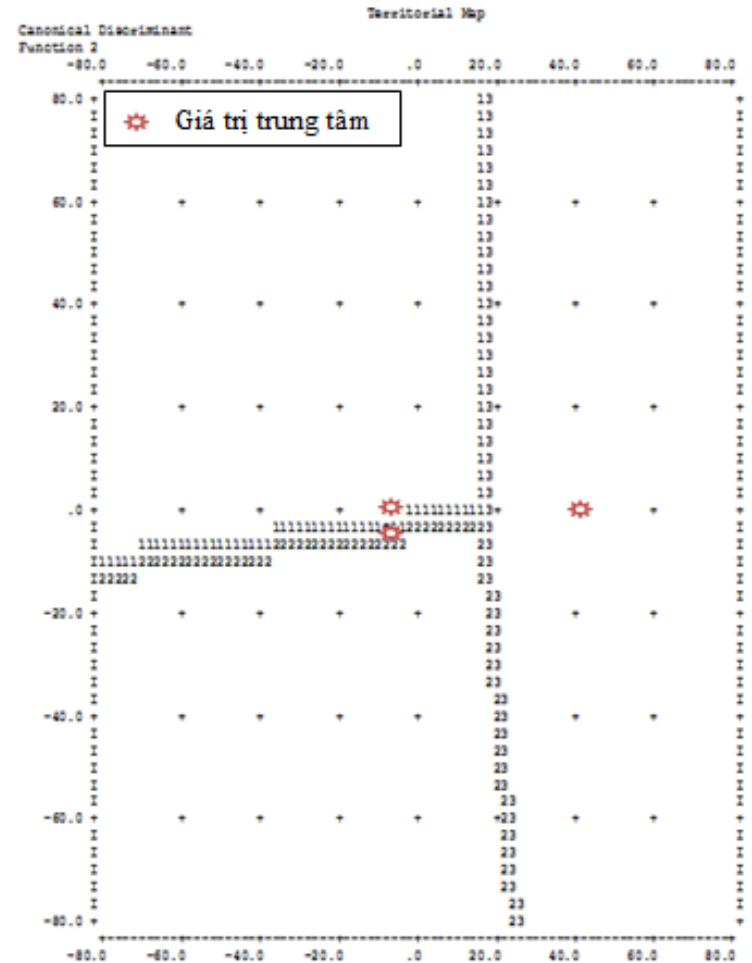
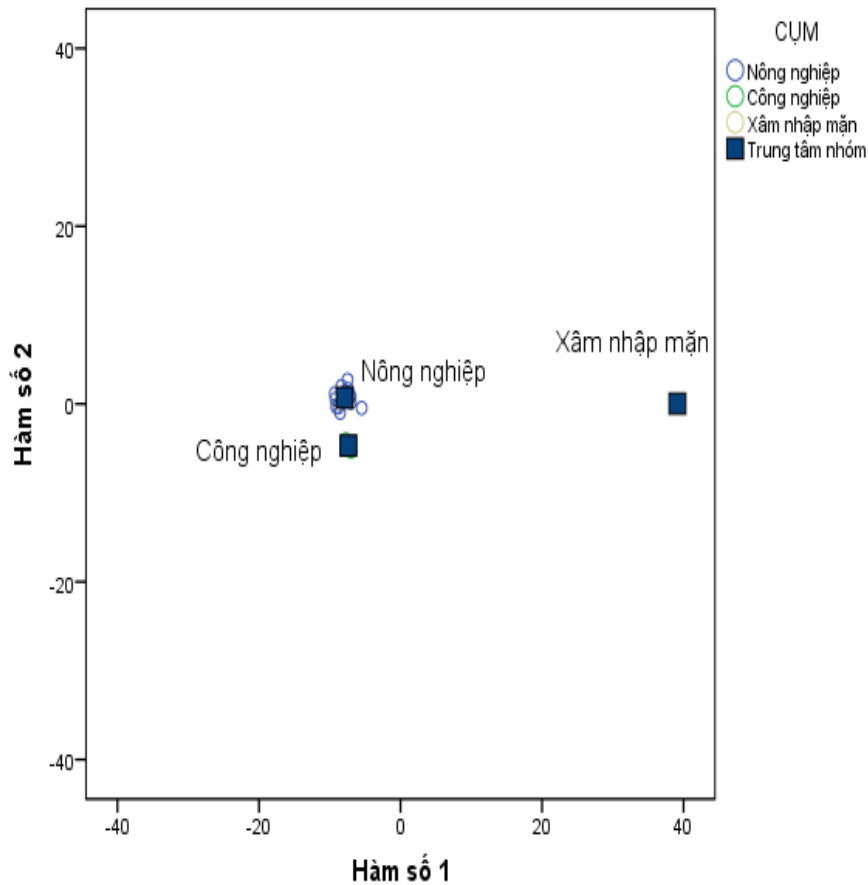
Xác định các biến phân biệt có ý nghĩa giữa các cụm

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

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

# KẾT QUẢ VÀ THẢO LUẬN

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



# KẾT QUẢ VÀ THẢO LUẬN

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

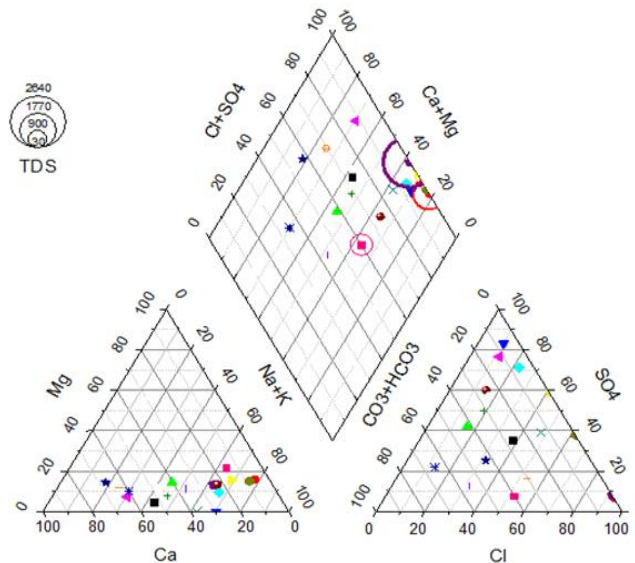
Sự nhiễm mặn thể hiện ở hàm lượng cao của TDS cũng như xu hướng tập trung cao các ion  $\text{Cl}^-$ ,  $\text{SO}_4^{2-}$ ,  $\text{Na}^+$ ,  $\text{K}^+$ ,  $\text{Mg}^{2+}$  và một số muối sắt hiện hữu. Bên cạnh đó, vào mùa khô, lượng nước bổ cập cho tầng chứa nước là rất ít, hầu như không xảy ra 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.



# KẾT QUẢ VÀ THẢO LUẬN

## Sự phân bố các ion chính

Hầu hết các giếng quan trắc đều phân bố ở khu vực thuộc loại nước  $\text{Ca}^{2+}$ -  $\text{Mg}^{2+}$ ,  $\text{Na}^{+}$ -  $\text{K}^{+}$  và loại nước  $\text{Cl}^{-}$  -  $\text{SO}_4^{2-}$ . Hơn thế nữa, các giếng quan trắc còn có xu hướng trội hơn nghiêng về loại nước  $\text{Cl}^{-}$  -  $\text{SO}_4^{2-}$ , 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 .



- QT5B
- VT4B
- ▲ VT6
- ▼ NB1B
- ◆ NB2C
- ◆ NB3A
- ◆ NB3B
- QT7B
- ★ QT5A
- ◆ VT4A
- NB1A
- NB2A
- NB2B
- NB4
- VT2B
- QT11
- QT7A
- VT2A

# KẾT LUẬN

# 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.

**XIN CHÂN THÀNH CẢM  
ƠN QUÝ HỘI ĐỒNG!**