

ỨNG DỤNG GIS VÀ THUẬT TOÁN TỐI ƯU TRONG XÂY DỰNG MẠNG LƯỚI QUAN TRẮC KHÔNG KHÍ TỈNH CÀ MAU

HUỖNH ANH HOÀNG^{1*}, VĂN TUẤN VŨ²

¹ Trường Đại học Tài nguyên và Môi trường TP. Hồ Chí Minh

² Trung tâm Tài nguyên Nước và Môi trường - Chi nhánh Công ty TNHH MTV Tài nguyên và Môi trường miền Nam

Tóm tắt

Chất lượng môi trường không khí (MTKK) đóng vai trò quan trọng trong bảo vệ sức khỏe cộng đồng và phát triển bền vững. Vì vậy, nghiên cứu được triển khai với mục đích là tạo lập mạng lưới quan trắc MTKK tối ưu, phù hợp với không gian phân bố dân số, giao thông và công nghiệp hiện nay của tỉnh Cà Mau. Nghiên cứu đã áp dụng phương pháp GIS, phân tích thống kê phân vị, kết hợp các thuật toán tối ưu Heuristic và Genetic Algorithm (GA) để xây dựng mạng lưới quan trắc không khí. Trên cơ sở chồng ghép các lớp dữ liệu dân số, giao thông và khu công nghiệp, 40 điểm quan trắc hiện hữu được sàng lọc, giữ lại 33 trạm có mức độ ưu tiên cao. Kết quả tối ưu cho thấy phương pháp Heuristic và GA đều đề xuất tập trung 3–5 trạm online tại TP. Cà Mau, trục QL1A và các khu công nghiệp (KCN) lớn, vừa đảm bảo tính đại diện, vừa tiết kiệm chi phí. Mạng lưới đề xuất là nền tảng cho việc tái thiết kế mạng lưới sau khi sáp nhập đơn vị hành chính cấp tỉnh.

Từ khóa: Cà Mau, Quan trắc không khí, GIS, Genetic Algorithm, Heuristic.

Ngày nhận bài: 9/9/2025; **Ngày sửa chữa:** 15/9/2025. **Ngày duyệt đăng:** 25/9/2025.

Application of GIS and optimization algorithms to design the air quality monitoring network in Ca Mau province

Abstract

Air quality (AQ) plays an important role in protecting public health and promoting sustainable development. Therefore, this study was conducted with the aim of establishing an optimal AQ monitoring network that aligns with the current spatial distribution of population, traffic, and industry in Cà Mau province. The research applied GIS methods, percentile statistical analysis, and combined optimization algorithms, including Heuristic and Genetic Algorithm (GA), to design the air monitoring network. By overlaying data layers of population, traffic, and industrial zones, 40 existing monitoring points were screened, retaining 33 stations with high priority levels. The optimization results showed that both Heuristic and GA methods recommended focusing 3–5 online stations in Cà Mau City, along National Highway 1A, and in major industrial zones, ensuring both representativeness and cost-effectiveness. The proposed network provides a foundation for redesigning the system after the provincial-level administrative merger.

Keywords: Air quality monitoring, GIS, Genetic Algorithm, Heuristic, Ca Mau.

JEL Classifications: Q 51, Q52, Q53, Q54.

1. GIỚI THIỆU

ÔNKK là một trong những thách thức lớn hiện nay, ảnh hưởng đến sức khỏe cộng đồng và hệ sinh thái. Theo WHO, việc xây dựng mạng lưới quan trắc với mật độ và tính đại diện đầy đủ là điều kiện tiên quyết để đánh giá phơi nhiễm của cộng đồng và đưa ra chính sách bảo vệ sức khỏe (World Health Organization, 2021). Đối với tỉnh Cà Mau là cực Nam của Tổ quốc, áp lực ÔNKK xuất phát từ nhiều nguồn khác nhau, bao gồm hoạt động công nghiệp, giao thông, chế biến thủy sản và năng lượng. Trong khi đó, hệ thống quan trắc MTKK hiện nay chủ yếu dựa vào các điểm đo thủ công, phân bố chưa đồng bộ, chưa đáp ứng yêu cầu giám sát và cảnh báo ô nhiễm (Vietnam Environment Administration (VEA), 2021). Một số nghiên cứu

trong nước gần đây cũng đã bước đầu ứng dụng GIS kết hợp các phương pháp đa tiêu chí để hỗ trợ quy hoạch mạng lưới quan trắc không khí (Trần Thị Hồng, 2020).

Vấn đề đặt ra là tỉnh Cà Mau hiện chưa có trạm quan trắc tự động, dữ liệu thu thập còn rời rạc, khó phản ánh xu thế ô nhiễm dài hạn và khó phục vụ dự báo, cảnh báo. Thực tiễn này đòi hỏi một giải pháp khoa học nhằm sàng lọc lại các trạm quan trắc hiện hữu, loại bỏ các trạm kém hiệu quả và xây dựng mạng lưới quan trắc tối ưu, bảo đảm tính đại diện và hiệu quả chi phí. Điểm mới của nghiên cứu là ứng dụng công nghệ GIS kết hợp với các thuật toán tối ưu (Heuristic và Genetic Algorithm – GA) để thiết kế mạng lưới quan trắc không khí (Chen, Y., Li, X., Zheng, Y., & Guan, Y.,



2019) (Lee, M., Kim, H., & Park, J., 2020), một hướng tiếp cận hiện đại chưa được áp dụng phổ biến ở khu vực đồng bằng sông Cửu Long.

Ngoài ra, trong bối cảnh Việt Nam đang tiến hành sắp xếp lại đơn vị hành chính cấp tỉnh, việc nghiên cứu xây dựng mạng lưới quan trắc cho Cà Mau không chỉ mang ý nghĩa địa phương, mà còn đóng vai trò nền tảng để mở rộng, tái thiết kế mạng lưới quan trắc không khí cho các tỉnh mới hình thành sau khi sáp nhập. Điều này làm tăng tính cấp thiết và giá trị thực tiễn của nghiên cứu.

Mục tiêu của nghiên cứu là tạo lập mạng lưới quan trắc MTKK tối ưu, phù hợp với không gian phân bố dân số, giao thông và công nghiệp hiện nay của tỉnh Cà Mau, đồng thời định hướng cho việc xây dựng mạng lưới quan trắc hiệu quả trong tương lai.

2. ĐỐI TƯỢNG VÀ PHƯƠNG PHÁP NGHIÊN CỨU

2.1. Đối tượng

Mạng lưới quan trắc MTKK tỉnh Cà Mau 2024-2025, bao gồm 40 điểm quan trắc thủ công do địa phương quản lý triển khai.

2.2. Phương pháp nghiên cứu

2.2.1. Phương pháp GIS

Nghiên cứu kế thừa các nguyên lý của lý thuyết tối ưu hóa mạng lưới quan trắc môi trường, trong đó GIS được sử dụng để tích hợp và phân tích dữ liệu không gian như mật độ dân số, hệ thống giao thông, phân bố khu công nghiệp.

- Dữ liệu và lớp thông tin không gian:

Khu vực nghiên cứu được chia thành 253 ô với kích thước mỗi ô lưới 5x5 km bao phủ toàn tỉnh Cà Mau (cũ). Mỗi ô lưới được gán ba yếu tố chính ảnh hưởng đến chất lượng không khí: Dân số (Population density), Giao thông (Traffic intensity), Khu công nghiệp (Industrial zones – IZ)

Các lớp dữ liệu này được chuẩn hóa về cùng một thang giá trị [0;1] nhằm loại bỏ sự khác biệt đơn vị đo (Nguyễn, 2020).

- Hàm mục tiêu:

Chỉ số tổng hợp S cho mỗi ô lưới được tính theo công thức (Malczewski, J., 1999):

$$S_i = \sum_{j=1}^m w_j \cdot x_{ij}, \quad \sum_{j=1}^m w_j = 1 \quad (1)$$

w_j: Trọng số của các yếu tố x_{ij}

x_{ij}: Các yếu tố ảnh hưởng đến chỉ số tổng hợp S

Phương trình (1) được triển khai theo 3 yếu tố chính (dân số, giao thông, KCN) có ảnh hưởng lớn đến vị trí đặt trạm quan trắc như sau:

$$S = w_1 * Norm_{Pop} + w_2 * Norm_{Traffic} + w_3 * Norm_{IZ} \quad (2)$$

Trong đó:

Norm-Pop: giá trị chuẩn hóa của mật độ dân số

Norm-Traffic: giá trị chuẩn hóa của mức độ giao thông

Norm_IZ: giá trị chuẩn hóa của phân bố khu công nghiệp

Trọng số w₁, w₂, w₃ được lựa chọn dựa trên kinh nghiệm và theo chuyên gia.

Chỉ số S vừa là cơ sở cho bước sàng lọc, vừa được dùng làm hàm mục tiêu (Fitness) trong quá trình tối ưu hóa bằng Heuristic và GA.

2.2.2. Phân tích thống kê phân vị

Nghiên cứu sử dụng phương pháp thống kê phân vị để phân loại mức độ ưu tiên hoặc sàng lọc, giữ lại các điểm có giá trị (Nguyen, H.

T., & Le, T. P., 2021) (Helsel, D. R., & Hirsch, R. M., 2002). Trên toàn bộ 253 ô lưới, các phân vị của chỉ số S được tính toán là Q1, Q2 (Median), Q3. Kết quả phân vị này được dùng để đối chiếu với 40 điểm quan trắc thủ công hiện hữu. Nguyên tắc sàng lọc:

Giữ lại các điểm (trạm thủ công) thuộc ô lưới có S ≥ Q2, từ trung vị trở lên;

Loại bỏ các điểm có S < Q2 và đồng thời cả ba yếu tố (Norm_Pop, Norm_Traffic, Norm_IZ) đều thấp hơn trung vị;

Xem xét đặc biệt: nếu trạm nằm trong ô có chỉ số tổng hợp S thấp nhưng một yếu tố vượt trội (≥ P90), trạm đó vẫn được cân nhắc giữ lại.

Kết quả phân vị giúp loại bỏ các điểm ít ý nghĩa trong giám sát, đồng thời rút gọn tập ứng viên đầu vào cho các bước tối ưu hóa bằng Heuristic và GA.

2.2.3. Phương pháp tối ưu Heuristic và GA

Sau khi sàng lọc, các điểm còn lại được đưa vào hai phương pháp tối ưu:

(i) Heuristic: Lựa chọn tuần tự các điểm có giá trị S cao, đảm bảo khoảng cách tối thiểu giữa các trạm, đến khi đạt số lượng trạm cần thiết (Lee, M., Kim, H., & Park, J., 2020); (ii) Genetic Algorithm: Mã hóa mỗi tập hợp vị trí trạm thành một cá thể; áp dụng các bước chọn lọc, lai ghép và đột biến để tối đa hóa hàm mục tiêu S cho toàn mạng lưới, đồng thời duy trì phân bố không gian hợp lý. Đây là thuật toán tiến hóa kinh điển (Holland, J. H., 1975) và đã được ứng dụng rộng rãi trong tối ưu hóa mạng lưới quan trắc môi trường (Chen, Y., Li, X., Zheng, Y., & Guan, Y., 2019).

Heuristic là nhóm thuật toán tối ưu dựa trên các quy tắc tìm kiếm đơn giản, nhằm chọn lựa phương án khả thi tốt nhất ở mỗi bước lặp. Phương pháp này thường cho kết quả nhanh chóng, phù hợp với các bài toán có dữ liệu lớn nhưng không yêu cầu độ chính xác tuyệt đối. Trong nghiên cứu này, thuật toán Heuristic (Lee, M., Kim, H., & Park, J., 2020) (Hình 1a) được sử dụng để lựa chọn lần lượt các vị trí trạm quan trắc không khí có chỉ

số phù hợp tổng hợp cao nhất, đảm bảo khoảng cách tối thiểu giữa các trạm.

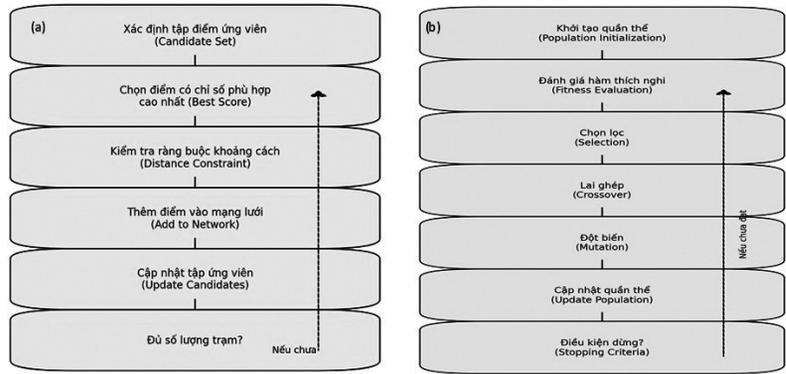
Genetic Algorithm (GA) là một thuật toán metaheuristic, mô phỏng quá trình tiến hóa tự nhiên thông qua các phép toán di truyền như chọn lọc, lai ghép và đột biến (Chen, Y., Li, X., Zheng, Y., & Guan, Y., 2019) (Randy L. Haupt, Sue Ellen Haupt, 2004) (Hình 1b). GA có khả năng tìm ra lời giải tối ưu gần toàn cục cho các bài toán tối ưu phức tạp, trong đó không gian tìm kiếm rất lớn và khó giải bằng các phương pháp truyền thống. Trong nghiên cứu này, GA được dùng để kiểm chứng kết quả một vài tổ hợp vị trí trạm quan trắc tối ưu của Heuristic, với hàm mục tiêu là tối đa hóa độ bao phủ dân số, giao thông và khu công nghiệp.

3. KẾT QUẢ VÀ THẢO LUẬN

3.1. Hiện trạng mạng lưới quan trắc không khí tỉnh Cà Mau

Dữ liệu đầu vào của nghiên cứu bao gồm thông tin về dân số, giao thông, khu công nghiệp và mạng lưới trạm quan trắc hiện hữu. Tất cả được chuẩn hóa và chồng ghép trên hệ lưới 5x5 km để phục vụ phân tích (Bảng 1).

Kết quả tổng hợp cho thấy toàn tỉnh Cà Mau có diện tích tự nhiên 5.274,51 km² được chia thành 253 ô lưới 5x5 km, trong đó có 40 ô chứa trạm quan trắc không khí thủ công (Hình 2). Các trạm này phân bố không đồng đều: tập trung nhiều ở khu vực thành phố Cà Mau và các trục giao thông chính, trong khi vùng ven biển và khu vực nông thôn thưa thớt trạm.



Hình 1. Sơ đồ khung lý thuyết thuật toán Heuristic (a) và Genetic Algorithm (b)

Do chỉ dựa vào phương pháp thủ công, số liệu quan trắc hiện có còn rời rạc, thiếu tính liên tục theo thời gian và chưa phản ánh đầy đủ sự biến động của chất lượng không khí. Bên cạnh đó, sự phân bố không đồng đều của các trạm khiến một số khu vực chịu áp lực phát thải lớn chưa được giám sát hiệu quả.

Những hạn chế trên cho thấy cần thiết phải phân tích không gian, sàng lọc và tái cấu trúc lại mạng lưới quan trắc hiện hữu, tiến tới thiết lập hệ thống trạm tự động tại các vị trí ưu tiên. Đây là cơ sở để thực hiện các bước phân tích phân vị và tối ưu hóa trong nghiên cứu này.

3.2. Phân tích không gian và tính toán chỉ số tổng hợp

Toàn tỉnh được chia thành 253 ô lưới 5x5 km. Trên mỗi ô i tính ba chỉ số thành phần, sau đó chuẩn hóa min-max về [0,1] và tích hợp thành chỉ số tổng hợp Si.

a. Dân số

Dữ liệu dân số được khai thác từ WorldPop (2023) là cơ sở dữ liệu dân số toàn cầu có độ phân giải 100 m kết hợp với Niên giám thống kê tỉnh Cà Mau (2023) với số dân là 1.207.400 người. Giá trị dân số tại từng ô lưới 5x5 km được xác định bằng phương pháp phân bố theo diện tích chồng lấn (areal interpolation) giữa raster WorldPop và ranh giới hành chính.

- Dân số của ô i được phân bố theo diện tích chồng lấn hành chính, sau đó quy về mật độ (người/km²): Popi.

-mChuẩn hóa min-max (Randy L. Haupt, Sue Ellen Haupt,

Bảng 1. Dữ liệu đầu vào phục vụ phân tích mạng lưới quan trắc không khí

Nhóm dữ liệu	Chỉ tiêu	Mô tả/Đơn vị
Dân số	Dân số raster (100 m)	Phân bố dân số trung bình theo ô lưới 5x5 km, tính mật độ (người/km ²) (WorldPop, 2023)
Giao thông	Số tuyến đường, tổng chiều dài đường	Tính chỉ số: Số đường + (Chiều dài/1000), đơn vị km (OpenStreetMap contributors, 2023)
Công nghiệp	Khoảng cách đến KCN	Tính khoảng cách trung bình từ tâm ô đến các KCN, nghịch đảo để chuẩn hóa
Lưới nghiên cứu	253 ô lưới 5x5 km	Toàn tỉnh Cà Mau được chia thành 253 ô, trong đó có 40 ô chứa trạm quan trắc hiện hữu
Mạng lưới hiện hữu	40 trạm quan trắc thủ công	Vị trí các trạm quan trắc không khí truyền thống, phân bố trong lưới nghiên cứu (Sở Nông nghiệp và Môi trường tỉnh Cà Mau, 2025)

2004) (Niharika, Venkatadri M., 2014) (Jiawei Han, Micheline Kamber, Jian Pei., 2021):

$$Norm_{Pop}(i) = \frac{Pop_i - \min(Pop)}{\max(Pop) - \min(Pop)} \quad (3)$$

b. Giao thông

Dữ liệu đường giao thông được khai thác từ OpenStreetMap (OpenStreetMap contributors, 2023). Tại mỗi ô lưới 5x5 km, tham khảo cách tiếp cận của Kanaroglou và cộng sự (Kanaroglou, P. S., Jerrett, M., Morrison, J., Beckerman, B., Arain, M. A., Gilbert, N. L., & Brook, J. R., 2005), chỉ số giao thông được tính dựa trên tổng số tuyến đường cắt qua ô kết hợp với tổng chiều dài đường trong ô (m), quy đổi về đơn vị km:

$$Traffic_i = Nroad_i + \frac{L_i}{1000} \quad (4)$$

Với Nroadi là số tuyến đường cắt qua ô; Li là tổng chiều dài đường trong ô (m). Việc chia 1000 đưa chiều dài về đơn vị km để cộng cùng số tuyến (không thứ nguyên).

- Chuẩn hóa min-max:

$$Norm_{Traffic}(i) = \frac{Traffic_i - \min(Traffic)}{\max(Traffic) - \min(Traffic)} \quad (5)$$

c. Khu công nghiệp

Trong nghiên cứu này, dữ liệu về bốn khu công nghiệp chính của tỉnh Cà Mau gồm Khánh An, Hòa Trung, Năm Căn và Sông Đốc được sử dụng. Mỗi KCN được chuẩn hóa về hệ tọa độ VN2000 và quy đổi thành điểm đại diện để thuận tiện cho tính toán.

- Tính khoảng cách trung bình từ tâm ô i đến các KCN (Liu, H., Shen, Z., & Chen, T., 2017) trong tỉnh:

$$\bar{d}_i = \frac{1}{M} \sum_{j=1}^M d(i, KCN_j) \quad (6)$$

với M là số KCN, d (i, KCNj) là khoảng cách tâm ô i đến từng KCN.

- Nghịch đảo để ra chỉ số thô gần KCN thì có giá trị cao:

$$IZ_i = \frac{1}{\bar{d}_i} \quad (7)$$

- Chuẩn hóa min-max:

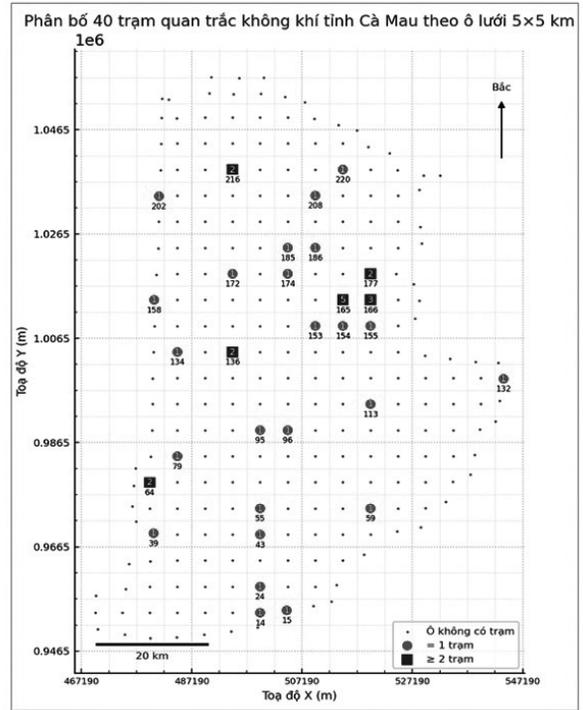
$$Norm_{IZ}(i) = \frac{IZ_i - \min(IZ)}{\max(IZ) - \min(IZ)} \quad (8)$$

(d) Chỉ số tổng hợp (fitness)

$$S_i = 0,5 * Norm_{Pop}(i) + 0,3 * Norm_{Traffic}(i) + 0,2 * Norm_{IZ}(i) \quad (9)$$

Bảng 2. Phân bố chỉ số S của 253 ô lưới theo phân vị

Giá trị S theo phân vị	Số ô lưới	Tỷ lệ (%)	Ghi chú
< Q1 (S < 0,065)	64	25,3	Ven biển, thưa dân
Q1-Q2 (S: 0,065-0,125)	63	24,9	Vùng đệm
Q2-Q3 (S: 0,125-0,216)	63	24,9	Đô thị phụ cận
≥ Q3 (S ≥ 0,216)	63	24,9	TP. Cà Mau, QL1A, gần KCN



Hình 2. Bản đồ phân bố 40 điểm quan trắc thủ công

Trọng số (0,5; 0,3; 0,2) được lựa chọn dựa trên kinh nghiệm và thực tế phản ánh tầm quan trọng ưu tiên của các yếu tố trong thực tiễn quản lý MTKK tại Cà Mau. Chỉ số Si phản ánh mức độ ưu tiên của từng ô và là đầu vào cho bước phân tích phân vị và sàng lọc các trạm hiện hữu.

Kết quả tính toán chỉ số tổng hợp S cho 253 ô lưới cho thấy sự phân hóa rõ rệt. Giá trị S dao động từ 0,006 đến 0,938, trung bình 0,150. Các ô có giá trị cao tập trung tại TP. Cà Mau, dọc QL1A và gần các KCN lớn (Khánh An, Hòa Trung); trong khi các ô giá trị thấp (<0,1) phân bố chủ yếu ở ven biển thưa dân. Bảng 2, thống kê số ô lưới có giá trị S theo các ngưỡng phân vị Q1, Q2, Q3:

Bảng 2 cho thấy phân bố các ô lưới theo phân vị được chia đều thành bốn nhóm, mỗi nhóm chiếm khoảng 25% tổng số ô. Không gian phân bố: nhóm <Q1 tập trung chủ yếu ở ven biển và nông thôn thưa dân; nhóm ≥Q3 chủ yếu phân bố ở TP. Cà Mau, dọc QL1A và các KCN lớn. Kết quả này tạo cơ sở để bước tiếp theo đối chiếu với 40 trạm quan trắc hiện

hữu nhằm thực hiện sàng lọc.

3.3. Phân tích thống kê phân vị và sàng lọc trạm hiện hữu

Chỉ số tổng hợp Si của 253 ô lưới được phân chia theo phân vị: Q1: S = 0,065; Q2 (trung vị): S = 0,125; Q3: S = 0,216

Khi đối chiếu 40 trạm hiện hữu vào lưới, kết quả phân bố: <Q1 có 1

Bảng 3. Kết quả phân vị và sàng lọc trạm quan trắc hiện hữu tại Cà Mau

Nhóm	Số trạm	Tỷ lệ (%)
Q1 ($S < 0,065$)	1	2,5
Q1-Q2 ($0,065 - 0,125$)	6	15,0
Q2-Q3 ($0,125 - 0,216$)	5	12,5
$\geq Q3$ ($S \geq 0,216$)	28	70,0
Tổng	40	100

trạm, Q1-Q2 có 6 trạm, Q2-Q3 có 5 trạm, $\geq Q3$ có 28 trạm. Theo nguyên tắc sàng lọc (giữ các trạm có $S \geq Q2$):

Loại bỏ: tổng cộng 7 trạm (1 trạm $< Q1$ và 6 trạm Q1-Q2).

Giữ lại: 33 trạm (5 trạm Q2-Q3 và 28 trạm $\geq Q3$).

Việc loại bỏ 7 trạm thấp giúp tập trung nguồn lực, đồng thời duy trì được 80% trạm ở mức ưu tiên cao ($\geq Q2$).

Đáng chú ý, trong nhóm giữ lại có 6 ô chứa nhiều trạm (từ 2–5), cụ thể: ô 165 (5 trạm), ô 166 (3 trạm), ô 64, 136, 177, 216 (mỗi ô 2 trạm). Các trạm này được giữ toàn bộ và sẽ được tinh gọn bằng thuật toán tối ưu ở bước tiếp theo (Bảng 3).

3.4. Tối ưu hóa vị trí đặt trạm online

Sau khi sàng lọc bằng phân vị, tập ứng viên còn lại gồm 33 trạm (23 ô lưới). Nghiên cứu tiếp tục tiến hành tối ưu hóa vị trí các trạm quan trắc không khí online bằng hai phương pháp: Heuristic và Thuật toán di truyền (GA).

(i) Phương pháp Heuristic:

Nguyên tắc lựa chọn là tuân tự chọn các ô lưới có giá trị chỉ số tổng hợp S cao nhất, đồng thời bảo

đảm khoảng cách tối thiểu 15km (khoảng cách này có thể thay đổi theo yêu cầu của địa phương) giữa các trạm để tránh chồng lặp. Quá trình được thực hiện theo kịch bản giả định: trong tương lai, nếu có điều kiện đầu tư từ 1 đến 10 trạm online, thì lần lượt chọn các ô ưu tiên từ trên xuống (Bảng 4).

(ii) Kiểm chứng bằng GA:

GA được áp dụng cho kịch bản 3 trạm online, với tham số: quần thể 100 cá thể, số thế hệ 100, tỷ lệ đột biến 10%, ràng buộc khoảng cách tối thiểu 15 km. Kết quả GA chọn ba ô: 95, 136, 165 với Fitness (S) trung bình đạt 0.58, hoàn toàn trùng khớp với kết quả Heuristic (Hình 3). Sự đồng nhất này khẳng định tính hợp lý và độ tin cậy của phương pháp Heuristic khi áp dụng cho bộ dữ liệu nghiên cứu.

Danh sách các ô ưu tiên từ Heuristic cho phép xây dựng lộ trình đầu tư theo giai đoạn: khi kinh phí hạn chế có thể đầu tư trước 1–3 trạm, sau đó mở rộng dần lên 5–10 trạm. Kết quả cũng cho thấy cả hai phương pháp đều nhấn mạnh các vị trí ưu tiên tập trung ở TP Cà Mau, trục QL1A và các KCN lớn (Khánh An, Hòa Trung), trong khi các vùng ven biển ít dân cư ít được lựa chọn.

Hình 4, phân bố 33 vị trí trạm quan trắc không khí sau sàng lọc và 10 trạm online tối ưu được lựa chọn bằng thuật toán Heuristic.

Thảo luận:

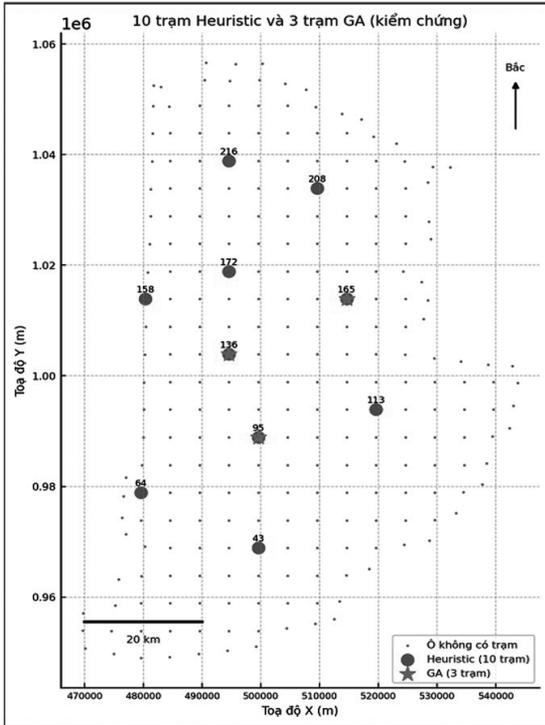
Kết quả nghiên cứu đã chỉ ra sự cần thiết phải sàng lọc và tối ưu lại mạng lưới quan trắc không khí tỉnh Cà Mau. Việc áp dụng thống kê phân vị giúp loại bỏ 7 trạm ít ý nghĩa, giữ lại 33 trạm có mức độ ưu tiên cao hơn. Các nghiên cứu (Nguyen, H. T., & Le, T. P., 2021) (Helsel, D.

R., & Hirsch, R. M., 2002) trước đây tại Việt Nam và quốc tế cũng khẳng định cách tiếp cận phân vị là phù hợp khi số liệu phân bố không đồng đều.

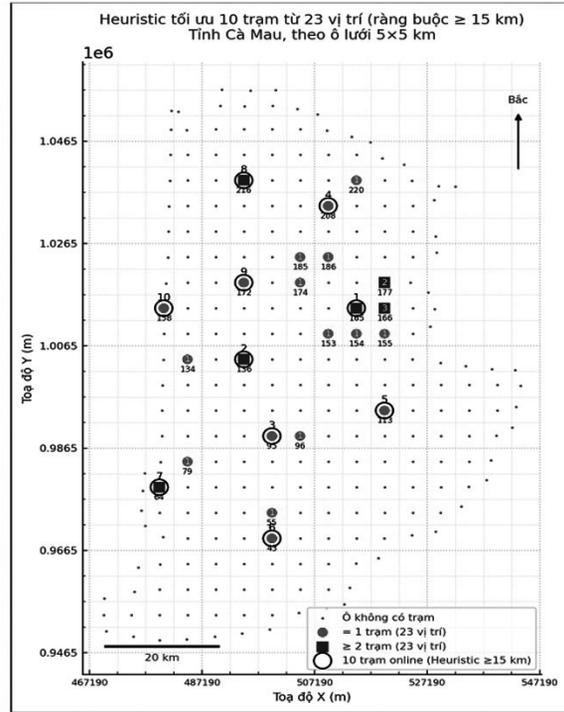
Phương pháp Heuristic cho kết quả nhanh, dễ áp dụng và có thể xây dựng lộ trình đầu tư theo giai đoạn (1–10 trạm). Đặc biệt, việc kiểm chứng bằng GA cho kịch bản 3 trạm đã trùng khớp hoàn toàn, củng cố độ tin cậy của phương pháp. Điều này phù hợp với xu hướng các nghiên cứu quốc tế như Chen & cộng sự (2019) (Chen, Y., Li, X., Zheng, Y., & Guan, Y., 2019), Lee & cộng sự (2020) (Lee, M., Kim, H., & Park, J., 2020), khi GA thường dùng để kiểm chứng tính tối ưu của các thuật toán đơn giản hơn.

Bảng 4. Kết quả lựa chọn vị trí trạm online bằng Hueristic

Số trạm online	Ô lưới được chọn (ID)	$S_{\text{Trung bình}}$
1	165	0.94
2	165, 136	0.68
3	165, 136, 95	0.58
4	165, 136, 95, 208	0.52
5	165, 136, 95, 208, 113	0.48
6	165, 136, 95, 208, 113, 43	0.44
7	165, 136, 95, 208, 113, 43, 64	0.41
8	165, 136, 95, 208, 113, 43, 64, 216	0.38
9	165, 136, 95, 208, 113, 43, 64, 216, 172	0.36
10	165, 136, 95, 208, 113, 43, 64, 216, 172, 158	0.34



Hình 3. Vị trí 10 trạm online theo Heuristic và 3 trạm trù theo GA



Hình 4. Sàng lọc 33 trạm thủ công và chùng ghép 10 trạm online tối ưu

Từ biểu đồ Hình 5 cho thấy giá trị trung bình S giảm nhanh khi số trạm tăng. Với 1-5 trạm, S trung bình vẫn duy trì ở mức cao (~0.5), phản ánh các vị trí được chọn đều thuộc nhóm ưu tiên. Khi mở rộng đến 10 trạm, S trung bình chỉ còn 0.34, thấp hơn nhiều so với trung vị 0.46 của các ô được chọn. Điều này chứng tỏ một số ô có giá trị S rất thấp đã kéo hạ giá trị trung bình, làm cho mạng lưới kém đồng đều và giảm hiệu quả. Theo Kanaroglou và cộng sự (Kanaroglou, P. S., Jerrett, M., Morrison, J., Beckerman, B., Arain, M. A., Gilbert, N. L., & Brook, J. R., 2005), khi số trạm vượt quá mức hợp lý, hiệu quả giảm do các trạm bổ sung thường rơi vào vùng ít dân cư hoặc xa nguồn thải. Do đó, quy mô 3-5 trạm được xem là phù hợp nhất cho tỉnh Cà Mau, vừa đảm bảo hiệu quả giám sát, vừa tránh đầu tư dàn trải vào các khu vực ít ưu tiên.

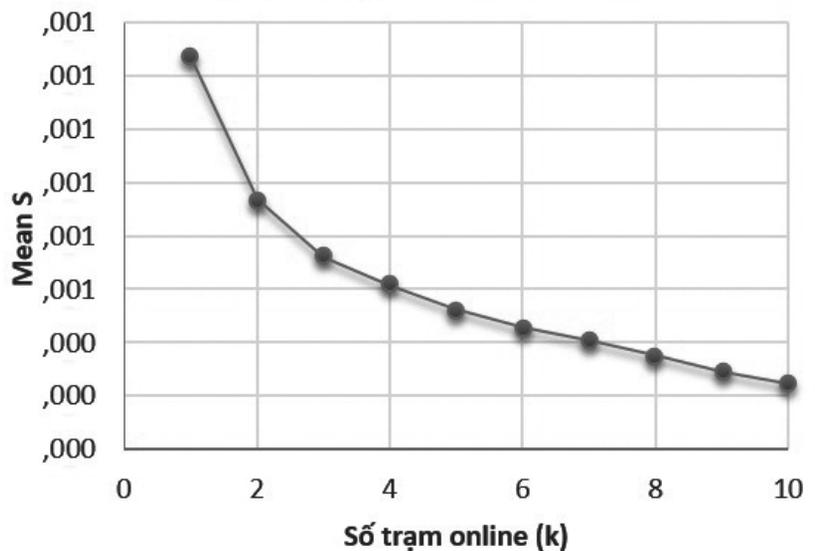
Điều này cho thấy kết quả của nghiên cứu tại Cà Mau phù hợp với xu thế quốc tế nhưng đồng thời có tính đặc thù địa phương.

KẾT LUẬN

Nghiên cứu đã đề xuất cách tiếp cận kết hợp phân tích thống kê kê phân vị, phương pháp Heuristic và thuật toán GA để sàng lọc và tối ưu mạng lưới quan trắc không khí tại tỉnh Cà Mau.

Trên cơ sở dữ liệu dân số, giao thông và khu công nghiệp, chỉ số tổng hợp S được xây dựng cho 253 ô lưới. Kết quả sàng lọc loại bỏ 8 trạm có giá trị thấp (Q1 và Q2), giữ lại 33 trạm có mức độ ưu tiên cao hơn.

Heuristic: Mean S theo k



Hình 5. Giá trị trung bình S theo số trạm online



Phương pháp Heuristic cho phép xác định thứ tự ưu tiên đầu tư 1–10 trạm online; trong đó 3-5 trạm online là phù hợp đặt tại các vị trí quan trọng nhất tập trung ở TP. Cà Mau, trục QL1A và các KCN lớn.

Thuật toán GA kiểm chứng cho kịch bản 3 trạm online đã cho kết quả trùng khớp hoàn toàn với Heuristic, chứng minh tính hợp lý và tin cậy của cách tiếp cận.

Kết quả nghiên cứu có ý nghĩa thực tiễn trong việc xây dựng lộ trình đầu tư mạng lưới trạm quan trắc online theo giai đoạn, đồng thời cung cấp cơ sở khoa học để điều chỉnh khi Cà Mau và các tỉnh lân cận được sáp nhập trong tương lai.

Tuy nhiên, nghiên cứu vẫn còn một số hạn chế. Mạng lưới quan trắc được tối ưu hóa mới chỉ dựa trên ba yếu tố chính gồm dân số, giao thông và khu công nghiệp. Trong thực tế, chất lượng không khí còn chịu ảnh hưởng bởi nhiều yếu tố khác như khí tượng, địa hình, hiện trạng sử dụng đất, cũng như các nguồn thải phân tán từ sinh hoạt và nông nghiệp. Do thiếu dữ liệu đầy đủ và đồng bộ, các yếu tố này chưa được tích hợp trong nghiên cứu này. Các nghiên cứu tiếp theo cần mở rộng phạm vi tiêu chí, kết hợp dữ liệu quan trắc tự động, mô hình lan truyền và phân tích chi phí – lợi ích để nâng cao độ chính xác hơn.

Lời cảm ơn: Nhóm nghiên cứu xin chân thành cảm ơn Trường Đại học Tài nguyên và Môi trường Thành phố Hồ Chí Minh đã hỗ trợ và tạo điều kiện thuận lợi cho quá trình thực hiện và hoàn thành nghiên cứu này ■

TÀI LIỆU THAM KHẢO

- Chen, Y., Li, X., Zheng, Y., & Guan, Y. (2019). Optimization of air quality monitoring station network based on genetic algorithm and GIS. *Atmospheric Environment*, 214(116854).
- Helsel, D. R., & Hirsch, R. M. (2002). *Statistical methods in water resources*. U.S. Geological Survey.
- Holland, J. H. (1975). *Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence*. Ann Arbor: University of Michigan Press.
- Jiawei Han, Micheline Kamber, Jian Pei. (2021). *Data mining: concepts and techniques*. Waltham, MA 02451, USA: Morgan Kaufmann, Elsevier.
- Kanaroglou, P. S., Jerrett, M., Morrison, J., Beckerman, B., Arain, M. A., Gilbert, N. L., & Brook, J. R. (2005). Establishing an air pollution monitoring network for intra-urban population exposure assessment: A location–allocation approach. *Atmospheric Environment*, 39(2), 239–251.
- Lee, M., Kim, H., & Park, J. (2020). Heuristic approaches for optimizing air quality monitoring networks: A case study in urban areas. *Environmental Monitoring and Assessment*, 192(4)(256).
- Liu, H., Shen, Z., & Chen, T. (2017). Optimization of industrial site selection based on GIS and multi-criteria decision analysis. *Sustainability*, 9(11).
- Malczewski, J. (1999). *GIS and Multicriteria Decision Analysis*. New York: John Wiley & Sons.
- Nguyen, H. T., & Le, T. P. (2021). Application of quantile statistics in environmental data analysis. *Vietnam Journal of Science and Technology*, 59(5), 620–632.
- Nguyễn, T. T. (2020). Phương pháp chuẩn hóa và tích hợp chỉ số trong đánh giá chất lượng môi trường. *Tạp chí Khoa học ĐHQGHN: Các Khoa học Trái đất và Môi trường*, 36(2), 87–95.
- Niharika, Venkatadri M. (2014). A Novel Air Quality Prediction Model Using Artificial Neural Networks. *International Journal of Engineering Research & Technology (IJERT)*, 3(2), 2973–2977.
- OpenStreetMap contributors. (2023). Vietnam road network data. Geofabrik GmbH. Được truy lục từ <https://download.geofabrik.de/asia/vietnam.html>
- Randy L. Haupt, Sue Ellen Haupt. (2004). *Practical Genetic Algorithms*. Second Edition, Wiley-InterScience.
- Sở Nông nghiệp và Môi trường tỉnh Cà Mau. (2025). Quyết định số E2500271010_2506181701, ngày 18 tháng 6 năm 2025 về việc phê duyệt E-HSMT gói thầu số 01: Quan trắc chất lượng môi trường định kì năm 2025 trên địa bàn tỉnh Cà Mau.
- Trần Thị Hồng, & N. (2020). Ứng dụng GIS và phương pháp đa tiêu chí trong quy hoạch mạng lưới quan trắc không khí đô thị. *Tạp chí Khoa học ĐHQGHN: Các Khoa học Trái đất và Môi trường*, 36(3), 45–56.
- Vietnam Environment Administration (VEA). (2021). *National report on environmental monitoring network*. Hanoi: VEA.
- World Health Organization. (2021). *WHO global air quality guidelines: Particulate matter (PM_{2.5} and PM₁₀), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide*. Geneva: World Health Organization.
- WorldPop. (2023). *WorldPop Global High Resolution Population Data (100m resolution)*. University of Southampton. Được truy lục từ <https://www.worldpop.org>.