

Ứng dụng mô hình hồi quy thành lập bản đồ chất lượng không khí tỉnh Tiền Giang

Applying regression model to create air quality map in Tien Giang province

> TRẦN NGỌC HUYỀN TRANG*, LÊ THUY LINH, TRẦN PHẠM ANH PHI

Trường Đại học Tài nguyên và Môi trường TP. HCM; *Email: tnhttrang@hcmunre.edu.vn

TÓM TẮT

Ô nhiễm không khí có thể gây ra nhiều vấn đề sức khỏe, như hen suyễn, ung thư phổi, bệnh tim và đột quỵ. Các nguyên nhân dẫn đến ô nhiễm không khí bao gồm các chất gây ô nhiễm trong không khí, có thể là các hạt rắn, giọt chất lỏng, hoặc khí. Chất gây ô nhiễm thường có nguồn gốc tự nhiên hoặc do con người tạo ra. Do đó việc đánh giá chất lượng môi trường nói chung và môi trường không khí nói riêng là một vấn đề cấp thiết cần được quan tâm và có những giải pháp xử lý kịp thời. Phương pháp tiếp cận trong bài báo là sử dụng kết hợp viễn thám và GIS tính hệ số tương quan, từ đó lập phương trình hồi quy giữa chỉ số chất lượng không khí tại các trạm đo và giá trị bức xạ của các kênh phổ trên ảnh vệ tinh Landsat 8. Kết quả là tìm được phương trình hồi quy bậc 3 đơn biến với hệ số $R^2=0.511137922$. Kiểm định phương trình hồi quy bằng 4 trạm quan trắc, kết quả đạt độ chính xác với $RMSE=2.5$.

Từ khóa: Tương quan - Hồi quy; ô nhiễm không khí; chỉ số chất lượng không khí (AQI); GIS, Landsat

ABSTRACT

Air pollution can cause many health problems, such as asthma, lung cancer, heart disease and stroke. Causes of air pollution include airborne contaminants, which can be solid particles, liquid droplets, or gases. Pollutants are often of natural or man-made origin. Therefore, assessing the quality of the environment in general and the air environment in particular is an urgent issue that needs attention and timely solutions. The approach in this article is to use a combination of remote sensing and GIS to calculate the correlation coefficient, thereby establishing a regression equation between the air quality index at measuring stations and the radiation value of the spectrum channels on Landsat 8 satellite images. The result is to find a univariate third-order linear regression equation with the coefficient $R^2=0.511137922$. Testing the regression equation with 4 monitoring stations, the results achieved accuracy with $RMSE=2.5$.

Key words: Correlation – Regression; air pollution; Air Quality Index (AQI); GIS; Landsat.

1. GIỚI THIỆU

Ô nhiễm không khí là trạng thái của bầu không khí bị bẩn hoặc chứa các chất gây hại như các hạt bụi, khí ô nhiễm, thành phần hóa học độc hại, hay khí thải từ các hoạt động công nghiệp, giao thông, nông nghiệp, và sinh hoạt người dân. Nó gây ra tác động xấu đến sức khỏe con người, động vật, cây cối và môi trường, gây ra các vấn đề về hô hấp, viêm phổi, bệnh tim mạch và ung thư. Ô nhiễm không khí trở thành một trong những vấn đề nghiêm trọng tại tỉnh Tiền Giang năm 2019. Tình hình này đã ảnh hưởng đến đời sống và sức khỏe của người dân trong khu vực. Từ tháng 10/2019 đến nay, mức độ ô nhiễm không khí trong tỉnh Tiền Giang đã tăng lên mức báo động, trong đó chủ yếu là do sự gia tăng các hoạt động công nghiệp và giao thông. Theo số liệu thống kê của Sở Tài nguyên và Môi trường, chỉ số chất lượng không khí trong khu vực đã vượt quá mức cho phép và ảnh hưởng đến sức khỏe của người dân.

Ứng dụng ảnh vệ tinh trong nghiên cứu chất lượng không khí đã được thực hiện ở các khu vực với quy mô, đặc trưng khác nhau và đã thu được một số kết quả nhất định trong thời gian gần và chi phí thấp. Các nghiên cứu như: Ứng dụng mô hình hồi quy tuyến tính trong nghiên cứu phân bố nồng độ bụi PM_{10} trên địa bàn các quận nội thành Hà Nội (Doãn Hoài Trang et al, 2020); Đánh giá sự phân bố

nồng độ $PM_{2.5}$ tại khu vực TP.HCM bằng công nghệ viễn thám (Trần Quang Trà et al, 2021); phân tán ô nhiễm không khí dựa vào viễn thám và dữ liệu mặt đất (Abdul Hameed Saleh và Ghada Hasan, 2014).

Bảng 1. Các mức chất lượng không khí (Tổng cục Môi trường, 2019)

Khoảng giá trị AQI	Chất lượng không khí	Ảnh hưởng sức khỏe	Màu
0-50	Tốt	Không	Xanh
51-100	Trung bình	Nhóm nhạy cảm, đôi khi nên giới hạn thời gian ở bên ngoài.	Vàng
101-200	Kém	Nhóm nhạy cảm nên hạn chế thời gian ở ngoài.	Da cam
201-300	Xấu	Nhóm nhạy cảm tránh ra ngoài. Những người khác nên hạn chế thời gian ở ngoài.	Đỏ
Trên 300	Nguy hại	Mọi người nên ở trong nhà.	Nâu

Chỉ số chất lượng không khí (viết tắt là AQI) là chỉ số được tính toán từ các thông số quan trắc các chất ô nhiễm trong không khí nhằm cho biết tình trạng chất lượng không khí và mức độ ảnh hưởng đến sức khỏe con người, được biểu diễn qua một thang điểm

(Tổng cục môi trường, 2019). Hiện nay có 5 mức AQI do tổng cục môi trường ban hành để thể hiện tình trạng chất lượng không khí như bảng sau:

Hình ảnh của Landsat 8 có độ phân giải cao, cho phép tách biệt chi tiết nhỏ và cung cấp thông tin chính xác về sự khác biệt của các vùng đất. Điều này rất hữu ích cho các nghiên cứu về bảo vệ môi trường, theo dõi những thay đổi đất đai và các tác động của nhân loại lên môi trường tự nhiên.

Kỹ thuật Tương quan - Hồi quy là phương pháp phân tích dựa trên mối liên hệ phụ thuộc của một biến kết quả (biến phụ thuộc) vào một hay nhiều biến nguyên nhân (biến độc lập). Tương quan là mối liên hệ giữa hai biến, khi một biến thay đổi thì biến còn lại cũng thay đổi theo một cách nào đó. Tương quan có thể được đo bằng hệ số tương quan, thường được ký hiệu bằng "r". Hệ số tương quan có giá trị từ -1 đến 1 và cho biết mức độ tương quan giữa hai biến, với giá trị 0 cho biết không có tương quan. Hồi quy là một kỹ thuật thống kê để đo và dự đoán mối quan hệ giữa hai biến số. Hồi quy được sử dụng để tìm ra mối liên hệ giữa một biến phụ thuộc và một hoặc nhiều biến độc lập. Một khi mô hình hồi quy được xác định, nó có thể được sử dụng để dự đoán giá trị của biến phụ thuộc dựa trên giá trị của biến độc lập.

2. KHU VỰC NGHIÊN CỨU

Tiền Giang thuộc ĐBSCL có diện tích tự nhiên là 2.484,2 km², nằm trong tọa độ 105°50' - 106°45' độ kinh Đông và 10°35' - 10°12' độ vĩ Bắc. Phía Bắc và Đông Bắc giáp Long An và TP.HCM, phía Tây giáp Đồng Tháp, phía Nam giáp Bến Tre và Vĩnh Long, phía Đông giáp biển Đông. Tiền Giang nằm trải dọc trên bờ Bắc sông Tiền (một nhánh của sông Mê Kông) với chiều dài 120km. Theo báo cáo của Sở Tài nguyên và Môi trường Tiền Giang, trong năm 2019, chất lượng không khí ở Tiền Giang chịu ảnh hưởng bởi nhiều nguyên nhân khác nhau, đặc biệt là từ sự phát triển kinh tế và công nghiệp. Các thành phần ô nhiễm khí thải như NO_x, SO₂, hạt bụi PM10 và PM2.5 đều vượt quá mức cho phép theo quy định. Mức độ ô nhiễm nặng nhất xảy ra tại các khu vực tập trung giao thông và công nghiệp, đặc biệt là trong các ngày nắng nóng dẫn đến tình trạng bức xạ nhiệt và tăng cường sự phát tán của khí thải. Việc áp dụng

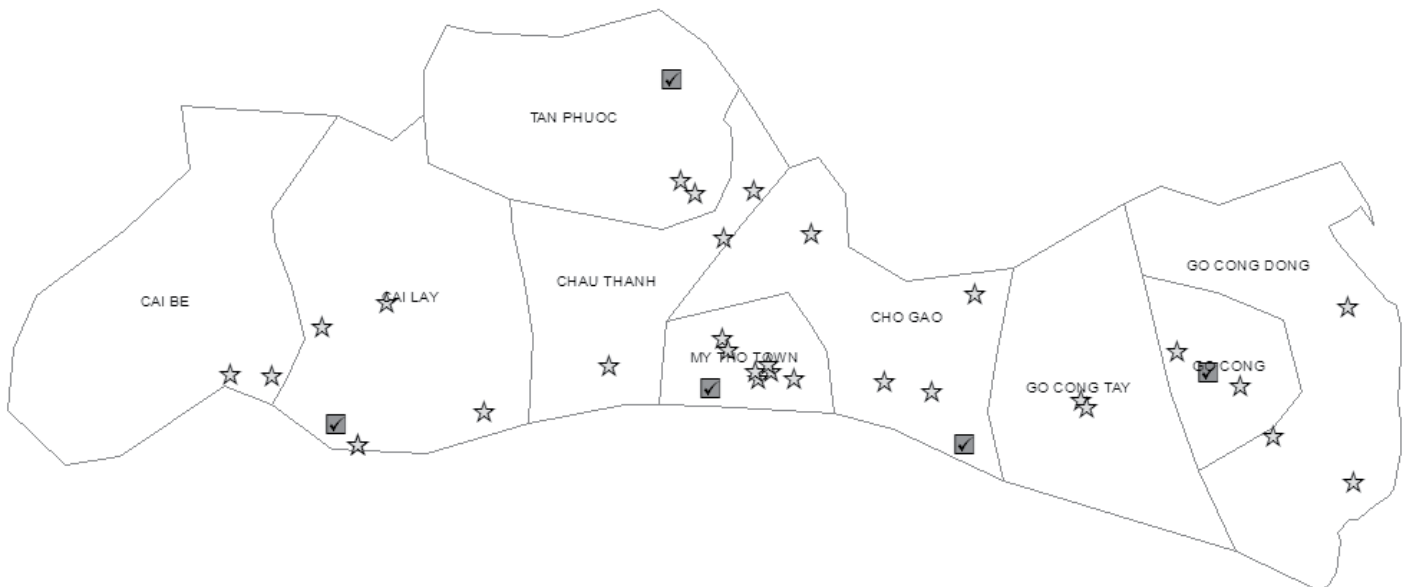
các biện pháp đối phó và cải thiện chất lượng không khí đang được chính quyền địa phương quan tâm và triển khai.



Hình 1. Vị trí tỉnh Tiền Giang ở khu vực ĐBSCL

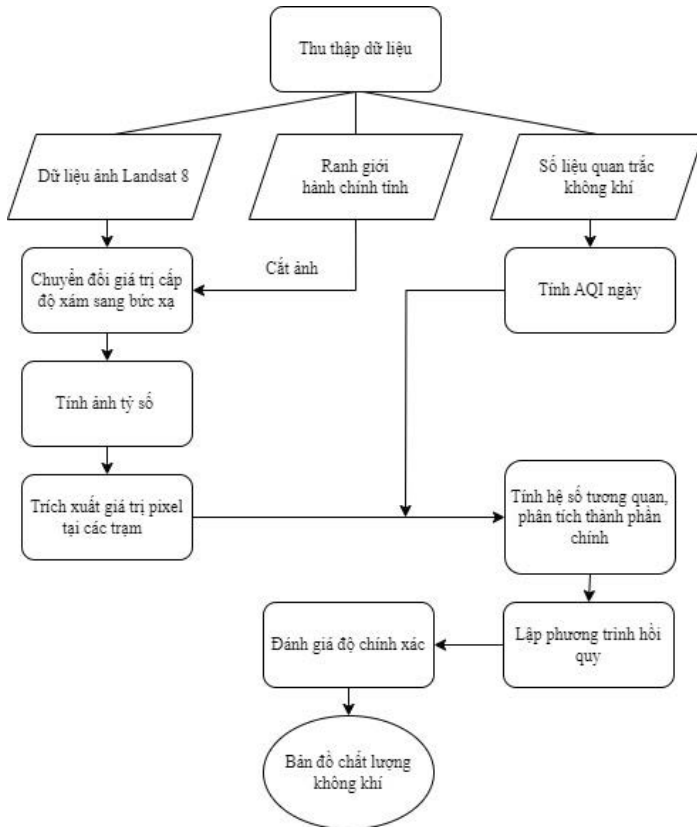
3. PHƯƠNG PHÁP NGHIÊN CỨU

Tư liệu sử dụng là ảnh vệ tinh Landsat 8 được chụp vào ngày 21/12/2019. Số liệu quan trắc của tỉnh Tiền Giang năm 2019, trong đó có tổng cộng 34 trạm đo, các trạm đo các thông số sau: Ô₃, Bụi lơ lửng, SO₂, NO₂, CO, O₃, PM₁₀, PM_{2.5}. Các trạm đều thuộc kiểu quan trắc tác động, loại quan trắc là thủ công. Để tính hệ số tương quan, tác giả sử dụng 30 trạm đo, còn lại 4 trạm đo dùng để đánh giá độ chính xác. Các trạm phân bố không đều trong khu vực nghiên cứu, cụ thể ở TP Mỹ Tho có 7 trạm đo và huyện Cai Lậy cũng có 7 trạm đo, huyện Chợ Gạo có 5 trạm, huyện Châu Thành, Gò Công Tây và thị xã Gò Công có 3 trạm, huyện Gò Công Đông, Cái Bè và huyện Tân Phước có 2 trạm. Khi phân tích vị trí các điểm quan trắc, tính toán chỉ số NNI thể hiện sự sắp xếp trong không gian được kết quả 0.8 với mật độ 30 điểm/2.484,2 km², NNI=0.8 cho thấy các điểm phân bố tương đối tập trung, và nhìn bản đồ cũng có thể thấy rằng các huyện vùng ven rất ít điểm quan trắc (Cái Bè, Tân Phước). Tuy nhiên sự phân bố này cũng khá ngẫu nhiên (gần 1).



Hình 2. Phân bố của các trạm quan trắc

Trình tự các bước được thực hiện theo hình sau đây:



Hình 3. Quy trình thực hiện

Phân tích vị trí các điểm kiểm tra đánh giá độ chính xác, với 4 điểm có chỉ số Nearest Neighbor Index NNI=0.93 cho thấy sự phân bố các điểm kiểm tra này là ngẫu nhiên (gần 1).

Với dữ liệu ảnh Landsat, công cụ Raster Calculator trong ArcGIS để chuyển đổi giá trị số nguyên (DN) của các kênh ảnh sang giá trị thực của bức xạ điện từ qua công thức:

$$L\lambda = MLQ_{cal} + AL \quad (1)$$

Trong đó:

• $L\lambda$: Giá trị bức xạ điện từ tại cảm biến (kênh ảnh gốc được tải xuống từ USGS);

• ML : Giá trị $RADIANCE_MULT_BAND_x$

• Q_{cal} : Giá trị số trên ảnh (Digital Number)

• AL : Giá trị $RADIANCE_ADD_BAND_x$

Để mở rộng tìm kiếm các kênh ảnh tương quan, tác giả tính toán các kênh ảnh tỷ số với công cụ Raster Calculator trên phần mềm ArcGIS. Ví dụ muốn tạo kênh ảnh tỷ số Band1/Band2 thì cú pháp nhập $Float(Band1) / Float(Band2)$.

Tính toán chỉ số chất lượng không khí AQI (Air Quality Index) từ dữ liệu thu thập tại các trạm quan trắc bao gồm: SO_2 , NO_2 , CO , O_3 , PM_{10} , $PM_{2.5}$ bằng công thức sau:

$$AQI_x = \frac{I_{i+1} - I_i}{BP_{i+1} - BP_i} (C_x - BP_i) + I_i \quad (2)$$

Trong đó:

•Bảng giá trị BP_i và I_i lấy trong Bảng 1

• AQI_x : Giá trị AQI_d thông số của thông số x

• BP_i : Nồng độ giới hạn dưới của giá trị thông số quan trắc được quy định trong Bảng 1 tương ứng với mức i

• BP_{i+1} : Nồng độ giới hạn trên của giá trị thông số quan trắc được quy định trong Bảng 1 tương ứng với mức $i+1$

• I_i : Giá trị AQI ở mức i đã cho trong Bảng 1 tương ứng với giá trị BP_i

• I_{i+1} : Giá trị AQI ở mức $i+1$ cho trong Bảng 1 tương ứng với giá trị BP_{i+1}

C_x : được quy định cụ thể như sau:

- Đối với thông số $PM_{2.5}$ và PM_{10} : C_x là giá trị trung bình 24 giờ
- Đối với thông số O_3 : C_x là giá trị lớn nhất trong giá trị trung bình 1 giờ lớn nhất trong ngày và giá trị trung bình 8 giờ lớn nhất trong ngày.

Lưu ý: Không tính toán AQI thông số O_3 khi giá trị trung bình 8 giờ lớn nhất trong ngày cao hơn $400 \mu g/m^3$ (lúc này chỉ tính toán AQI đối với trung bình 1 giờ lớn nhất trong ngày).

- Đối với thông số SO_2 , NO_2 và CO : C_x giá trị trung bình 1 giờ lớn nhất trong ngày.

Bảng 2. Các giá trị BP_i đối với các thông số (Tổng cục môi trường, 2019)

i	I_i	Giá trị BP _i quy định đối với từng thông số (Đơn vị: $\mu g/m^3$)						
		$O_3(1h)$	$O_3(8h)$	CO	SO_2	NO_2	PM_{10}	$PM_{2.5}$
1	0	0	0	0	0	0	0	0
2	50	160	100	10000	125	100	50	25
3	100	200	120	30000	350	200	150	50
4	150	300	170	45000	550	700	250	80
5	200	400	210	60000	800	1200	350	150
6	300	800	400	90000	1600	2350	420	250
7	400	1.000	-	120000	120000	3100	500	350
8	500	≥ 1200	-	≥ 150000	≥ 2630	≥ 3850	≥ 600	≥ 500

Sau khi đã có số liệu giá trị bức xạ của 99 kênh ảnh bao gồm các kênh ảnh đơn lẻ cùng các kênh ảnh tỷ số và số liệu AQI_{max} của 30 trạm đo thì ta có thể tiến hành tính tương quan. Hệ số tương quan bằng công cụ Correlation trong Excel giữa giá trị AQI_{max} với 9 kênh ảnh gốc đã chuyển từ giá trị cấp độ xám sang giá trị bức xạ và các kênh ảnh tỷ số. Sau đó, lựa chọn các kênh có hệ số tương quan lớn hơn hoặc bằng 0.5 (giá trị 0.5 được các nhà nghiên cứu: Francis và White (2002); Kim và Stoele (2004) đã công bố trong các nghiên cứu của mình, đây là ngưỡng mà những nhà nghiên cứu này đã đưa ra ở mức tốt). Kết quả cho thấy các kênh tỷ số với kênh 1 cho giá trị tương quan cao nhất.

Bảng 3. Kết quả các kênh ảnh có giá trị tương quan cao với chỉ số AQI

Kênh ảnh	Hệ số tương quan	Kênh ảnh	Hệ số tương quan
B2/B1	0.586880317	B7/B1	0.583429622
B3/B1	0.586367855	B8/B1	0.583989467
B4/B1	0.586117796	B9/B1	0.585041597
B5/B1	0.584708645	B10/B1	0.585017365
B6/B1	0.58724833	B11/B1	0.584737508

Phân tích thành phần chính để loại bỏ các kênh tương quan với nhau thì thấy rằng các kênh này đều có sự tương quan cao, chính vì thế chỉ chọn kênh B6/B1 có giá trị tương quan cao nhất để thành lập phương trình hồi quy.

Sau nhiều lần khảo sát hồi quy đơn biến theo từng bậc để lựa chọn hàm hồi quy có giá trị R^2 lớn nhất của kênh B6/B1.

Bảng 4. Hệ số R^2 ứng với các bậc của phương trình hồi quy

Bậc phương trình	Hệ số R^2
1	0.344860601
2	0.510594140
3	0.511137922
4	0.342085141
5	0.342085140

Hệ số R^2 tăng dần từ bậc 1 đến bậc 3 sau đó bắt đầu giảm, chính vì vậy lựa chọn phương trình bậc 3 có giá trị R^2 cao nhất để xây dựng hàm hồi quy.

Phương trình hồi quy đơn biến có dạng như sau:

$$y = b_0 + b_1 \cdot x + b_2 \cdot x^2 + \dots + b_n \cdot x^n \quad (3)$$

Kết quả của phương trình hồi quy bậc 3 được xây dựng với công cụ Excel như sau:

$$AQI = 28.62752715 + 2.716556599 \times B6/B1 - 0.054903125 \times B6/B1^2 + 0.0000368084 \times B6/B1^3$$

4. KẾT QUẢ VÀ THẢO LUẬN

Kết quả hồi quy được đánh giá độ chính xác theo sai số Root Mean Square Error (RMSE).

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}} \quad (4)$$

Trong đó:

$\hat{y}_i$ là giá trị AQI dự đoán (từ phương trình hồi quy)

y_i là giá trị AQI đo

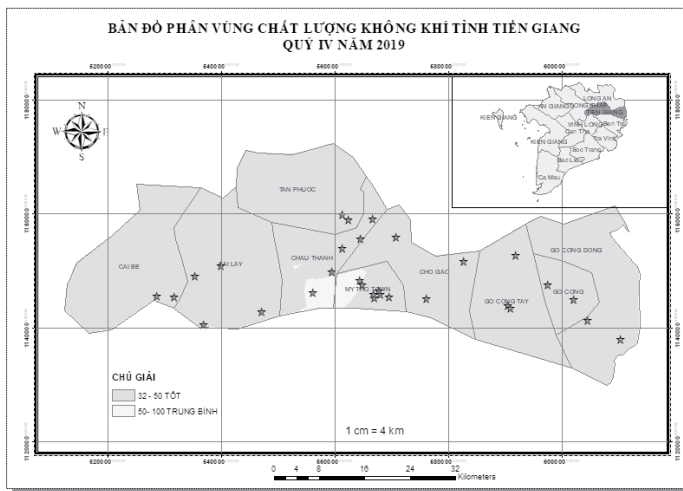
n là số trạm kiểm tra

Sai số giới hạn được tính dựa theo văn bản kỹ thuật đo lường Việt Nam ĐLVN 333: 2016: Phương tiện đo nồng độ SO_2 , CO, NO, NO_2 của trạm quan trắc không khí tự động, liên tục - quy trình kiểm định. Độ lệch trung bình được phép bằng 5% giá trị đọc của các thông số đã được tính toán chuẩn hóa thành AQI nên lấy sai số giới hạn cho AQI bằng 5% AQI_{max}.

Bảng 5. Đánh giá độ chính xác kết quả hồi quy

Mã trạm	AQI _{Hồi quy}	AQI _{Đo}	Độ lệch	RMSE
K8	28.9	32.9	-4.0	2.5
K14	29.0	30.1	-1.1	
K21	29.1	27.4	1.7	
K33	28.9	31.3	-2.4	

Theo tiêu chuẩn ĐLVN 333:2016, sai số giới hạn của số liệu đo trực tiếp là 5% giá trị đọc (ở đây sử dụng giá trị đọc lớn nhất sau khi quy đổi thành AQI là 55), sai số cho phép là 2.75. Kết quả cho thấy rằng, sai số khi sử dụng phương trình hồi quy ảnh vệ tinh là 2.5 nhỏ hơn sai số cho phép là 2.75 nên có thể sử dụng phương pháp này để thay thế cho phép đo trực tiếp với mức độ chính xác tương đương.



Hình 4. Bản đồ phân vùng chất lượng không khí

Bản đồ phân vùng chất lượng không khí được tính toán từ phương trình hồi quy, sau đó tái chia mẫu và biên tập với tỷ lệ 1/500,000 gồm hệ thống khung, lưới, chú giải, thước tỷ lệ, bản đồ phụ,... Kết quả cho thấy chất lượng không khí của tỉnh Tiền Giang vào cuối năm 2019 (tháng 12 – quý IV) chia thành 2 mức: 0-50 tương ứng với mức độ tốt, mức này chiếm phần lớn diện tích của tỉnh; 50-100 tương ứng với mức trung bình phân bố ở khu vực Thành phố Mỹ Tho và lân cận, nguyên nhân là do khu vực này tập trung nhiều

khu công nghiệp, các nhà máy và dân cư đông đúc. Đây cũng chính là lý do gây ô nhiễm môi trường khu vực.

IV. KẾT LUẬN

Giải pháp kết hợp viễn thám và GIS đã và đang trở thành công cụ hỗ trợ hiệu quả trong quản lý môi trường nói chung và môi trường không khí nói riêng. Ảnh Landsat 8 là nguồn tư liệu ảnh miễn phí, độ phân giải cao, chu kỳ tương đối ngắn phù hợp cho mục đích giám sát chất lượng không khí bề mặt.

Ứng dụng mô hình tương quan - hồi quy từ ảnh vệ tinh áp dụng cho chỉ số chất lượng không khí đạt độ tin cậy cao (RMSE=2.5), thời gian thực hiện nhanh chóng, có thể trở thành công cụ trợ giúp hiệu quả cho chính quyền địa phương trong việc quản lý tình trạng ô nhiễm không khí trên khu vực.

TÀI LIỆU THAM KHẢO

[1] Tổng cục Môi trường. *Hướng dẫn kỹ thuật tính toán và công bố chỉ số chất lượng không khí Việt Nam (VN_AQI)*. Số: 1459/QĐ-TCMT, 2019.

[2] Tổng cục Tiêu chuẩn Đo lường chất lượng. *Phương tiện đo nồng độ SO_2 , CO, NO, NO_2 của trạm quan trắc không khí tự động, liên tục - Quy trình kiểm định*. Văn bản kỹ thuật đo lường Việt Nam, 2016.

[3] Doãn Hoài Trang, Đặng Vũ Khắc, Nguyễn Khắc Anh, Phạm Thanh Hải. *Ứng dụng mô hình hồi quy tuyến tính trong nghiên cứu phân bố nồng độ bụi PM_{10} trên địa bàn các quận nội thành Hà Nội, 2020*.

[4] Salah Abdul Hameed Saleh và Ghada Hasan. Estimation of PM_{10} Concentration using Ground Measurements and Landsat 8 OLI Satellite Image. *Journal of Remote Sensing & GIS*. Vol. 3, Số. 120, 2014.

[5] Trần Quang Trà, Nguyễn Phúc Hiếu, Đào Nguyên Khôi. *Đánh giá sự phân bố nồng độ $PM_{2.5}$ tại khu vực TP.HCM bằng công nghệ viễn thám - một số kết quả ban đầu, 2021*.