

JTST - JOURNAL OF TRANSPORTATION SCIENCE & TECHNOLOGY

MANUSCRIPT ID: JIST-2022-0027

CÁC MÔ HÌNH THỐNG KÊ NGHIÊN CỨU Ô NHIỄM KHÔNG KHÍ TỪ HOẠT ĐỘNG GIAO THÔNG ĐƯỜNG BỘ ĐẾN SỨC KHỎE - TRƯỜNG HỢP THÀNH PHỐ HỒ CHÍ MINH

Manuscript ID	JIST-2022-0027
Full title	Các Mô Hình Thống Kê Nghiên Cứu Ô Nhiễm Không Khí Từ Hoạt Động Giao Thông Đường Bộ Đến Sức Khỏe - Trường Hợp Thành Phố Hồ Chí Minh
Summary	<p>Ô nhiễm môi trường là vấn nạn đáng báo động hiện nay trên toàn cầu. Trong đó ô nhiễm không khí do khí thải từ các phương tiện giao thông là một trong những nguyên nhân chính góp phần tạo nên vấn nạn đó. Bài báo sử dụng các mô hình toán học bằng công cụ thống kê để nghiên cứu tình trạng ô nhiễm không khí từ hoạt động xả thải của các phương tiện giao thông đường bộ. Bằng việc sử dụng các mô hình hồi qui, mô hình tự hồi quy tuyến tính (AR), tích hợp (I) và thay đổi (dịch chuyển) trung bình (MA) có xét với yếu tố mùa vụ, bài báo phân tích và dự báo hiện trạng ô nhiễm không khí tại Thành phố Hồ Chí Minh với các dữ liệu cập nhật. Các tính toán mô phỏng trong bài báo được trình bày trên ngôn ngữ lập trình R.</p>
Research Area	Môi trường
Research Topic	Môi trường
Keywords	Mô hình thống kê; ô nhiễm không khí; AQI; giao thông đô thị; sức khỏe con người
Authors	<p>Nguyễn Thị Thanh Hà, Email: , Insitution: Trường Đại học Giao thông vận tải Thành phố Hồ Chí Minh, Department: . Trần Đình Tường, Email: tuong.tran@ut.edu.vn, Insitution: Trường Đại học Giao thông vận tải Thành phố Hồ Chí Minh, Department: . Nguyễn Thị Thu Trang, Email: , Insitution: Trường Đại học Giao thông vận tải Thành phố Hồ Chí Minh, Department: . Mai Thị Hải Hà, Email: , Insitution: Trường Đại học Giao thông vận tải Thành phố Hồ Chí Minh, Department: . Trần Văn Thành, Email: , Insitution: Trường Đại học Giao thông vận tải Thành phố Hồ Chí Minh, Department: .</p>

CÁC MÔ HÌNH THỐNG KÊ NGHIÊN CỨU Ô NHIỄM KHÔNG KHÍ TỪ HOẠT ĐỘNG GIAO THÔNG ĐƯỜNG BỘ ĐẾN SỨC KHỎE - TRƯỜNG HỢP THÀNH PHỐ HỒ CHÍ MINH

STATISTICAL MODELS FOR AIR POLLUTION CAUSED BY GROUND TRANSPORT: A CASE STUDY IN HCM

¹Nguyễn Thị Thanh Hà, ²Trần Đình Tường, ³Nguyễn Thị Thu Trang, ⁴Mai Thị Hải Hà, ⁵Trần Văn Thành, ⁶Nguyễn Điệp, ⁷Trương Thị Dung, ⁸Hà Nguyễn Thùy Linh
^{1,2,7,8}Khoa Cơ bản, ³Viện đào tạo và hợp tác quốc tế (IEC), ^{4,5}Viện nghiên cứu Môi trường và Giao thông Trường Đại học Giao thông vận tải Thành phố Hồ Chí Minh.
⁶Khoa Cơ bản, Trường Cao đẳng Công thương Thành phố Hồ Chí Minh.

Tóm tắt. Ô nhiễm môi trường là vấn nạn đáng báo động hiện nay trên toàn cầu. Trong đó ô nhiễm không khí do khí thải từ các phương tiện giao thông là một trong những nguyên nhân chính góp phần tạo nên vấn nạn đó. Bài báo sử dụng các mô hình toán học bằng công cụ thống kê để nghiên cứu tình trạng ô nhiễm không khí từ hoạt động xả thải của các phương tiện giao thông đường bộ. Bằng việc sử dụng các mô hình hồi qui, mô hình tự hồi quy tuyến tính (AR), tích hợp (I) và thay đổi (dịch chuyển) trung bình (MA) có xét với yếu tố mùa vụ, bài báo phân tích và dự báo hiện trạng ô nhiễm không khí tại Thành phố Hồ Chí Minh với các dữ liệu cập nhật. Các tính toán mô phỏng trong bài báo được trình bày trên ngôn ngữ lập trình R.

Từ khóa. Mô hình thống kê; ô nhiễm không khí; AQI; giao thông đô thị; sức khỏe con người.

Mã phân loại: 6.2

Abstract. Environmental pollution is an alarming issue globally. One of the main sources contributing to that problem is the air pollution induced by emission from transportation vehicles. This paper uses mathematical models to study the air pollution from the exhaust operation of ground transport. By using regression models, ARIMA (Autoregression – Integration – Moving Average) models with seasonal variations of the AQI, we analyze and forecast the air pollution in HCM city. The computations were simulated in programming language R.

Keywords. Statistical models; air pollution; AQI; ground transport; health.

Classification code: 6.2

1. Giới thiệu

Giao thông vận tải (GTVT) là một trong những yếu tố chính góp phần gây ra tình trạng ô nhiễm không khí tại Việt Nam. GTVT thải ra môi trường các chất gây ô nhiễm không khí và các hạt bụi (Particulate Matter - PM) với những thành phần và kích cỡ khác nhau. Theo báo cáo của tổ chức WHO, khoảng 30% lượng bụi mịn (PM_{2.5} - các hạt có đường kính nhỏ hơn 2.5 μm) thành phần chính xây dựng chỉ số chất lượng không khí AQI (Air Quality Index) được thải ra từ các ống xả khí của các loại xe có động cơ [1]. Các hạt bụi có đường kính lớn hơn (2.5 μm - 10 μm) đến từ các hoạt động khác của giao thông đường bộ như bụi đường, thắng xe và cấu tạo lốp của lốp xe. Ngoài ra, giao thông đường bộ cũng là một trong những nguồn chính thải ra nitrogen dioxide và benzen.

Ô nhiễm không khí đến từ hoạt động GTVT làm tăng nguy cơ tử vong, cụ thể là từ những

bệnh liên quan tới tim và phổi. Nghiên cứu thực nghiệm chỉ ra rằng các tác động này có liên quan đến những thay đổi trong việc hình thành các gốc oxy hóa tự do (reactive oxygen species), sự thay đổi khả năng chống oxy hóa và tăng phản ứng viêm, do đó gây ra các tổn thương sức khỏe. Ô nhiễm không khí đến từ GTVT có thể làm tăng nguy cơ dị ứng và làm trầm trọng thêm các triệu chứng, đặc biệt với nhóm người nhạy cảm. Một số nghiên cứu khác báo cáo sự gia tăng đáng kể nguy cơ nhồi máu cơ tim sau khi tiếp xúc với nguồn không khí ô nhiễm đến từ các hoạt động GTVT. Việc tiếp xúc với không khí ô nhiễm cũng có thể dẫn đến những thay đổi trong cơ chế điều hòa hệ thần kinh thực vật và tăng phản ứng viêm. Một số nghiên cứu khác cũng cho thấy tỷ lệ mắc ung thư phổi gia tăng ở những người tiếp xúc lâu dài với không khí ô nhiễm đến từ hoạt động GTVT chẳng hạn như [6]-[8] và các tài liệu được trích dẫn trong đó.

Tại Thành phố Hồ Chí Minh (TP. HCM) có khoảng 30 trạm quan trắc với tần suất 10 ngày/tháng thường đo vào 2 thời điểm (7 giờ 30 phút đến 8 giờ 30 phút và 15 giờ đến 16 giờ), chủ yếu đặt tại các giao lộ, vòng xoay, các tuyến đường có mật độ giao thông cao như khu vực vòng xoay An Sương (Quận 12), vòng xoay Mỹ Thủy (Quận 2), ngã tư Huỳnh Tấn Phát - Nguyễn Văn Linh (Quận 7), ngã tư Bình Phước (Quận Thủ Đức)... Công tác lấy mẫu được thực hiện bán thủ công, dữ liệu không được quan trắc 24/24 giờ dẫn đến dữ liệu được cập nhật chậm, không đầy đủ gây khó khăn cho công tác cảnh báo ô nhiễm và các biện pháp kiểm soát ô nhiễm. Bên cạnh đó còn có các tổ chức xã hội, đa phần là nước ngoài quan trắc dữ liệu liên tục với những trạm quan trắc cố định như Tổng lãnh sự quán Hoa Kỳ quan trắc thông số $PM_{2.5}$.

Với tác động nguy hại đến sức khỏe người dân thành phố, đòi hỏi có những nghiên cứu tin cậy với số liệu cập nhật và các thông số ô nhiễm của không khí là hệ quả trực tiếp từ hoạt động xả thải của các phương tiện giao thông đường bộ như: TSP (bụi tổng lơ lửng), SO_2 , NO_2 , CO, benzen, hơi xăng dầu, bụi chì v.v... tại những nơi có mật độ giao thông cao theo đề án qui hoạch tổng thể mạng lưới quan trắc tại TP. HCM được Ủy ban Nhân dân Thành phố phê duyệt (2015), đồng thời sử dụng chỉ số AQI với số liệu đủ lớn để phân tích, dự đoán tình hình ô nhiễm tại Thành phố trong thời gian tới.

Việc sử dụng các mô hình toán học trong nghiên cứu ô nhiễm không khí đã thu hút nhiều nhóm nghiên cứu từ trong nước cũng như trên thế giới [3]-[5], [11], v.v... Bài báo xây dựng các mô hình thống kê để giải quyết vấn đề trên như xây dựng các mô hình hồi qui nghiên cứu mối tương quan thống kê giữa các thành phần ô nhiễm không khí chính từ các hoạt động giao thông đường bộ với số liệu cập nhật. So với [5], kết quả này đã xây dựng các mô hình hồi qui tuyến tính từ đơn giản đến mô hình hồi qui tuyến tính đa biến trong sự phân tích ý nghĩa từ các tham số của mô hình. Ngoài ra bài báo còn sử dụng mô hình Box-Jenkins [2], và kế thừa kỹ hiệu và kết quả [4] cho việc sử dụng chuỗi thời gian, một khái niệm quan trọng trong thống kê, để phân tích và dự đoán chỉ số AQI trong những năm gần đây tại TP.HCM.

Cấu trúc bài báo được trình bày như sau, trong mục 1, nhóm tác giả giới thiệu tình hình thời sự và động lực của vấn đề nghiên cứu. Mục 2 trình bày nội dung chính của bài báo. Phần đầu tiên của mục 2 nhóm trình bày các thông tin cơ bản của dữ liệu quan sát và phân tích lựa chọn mô hình. Ngoài ra mục này còn phân tích mô hình tuyến tính chuỗi thời gian từ đó đưa ra các dự đoán cho tình hình ô nhiễm không khí trong thời gian sắp đến tại TP. HCM. Ở mục cuối, bài báo tóm tắt các kết quả nghiên cứu và đề xuất các giải pháp để kiểm soát, hạn chế ô nhiễm không khí do các phương tiện giao thông đường bộ tại TP. HCM.

2. Kết quả

2.1. Các mô hình hồi qui tuyến tính

Kết quả quan trắc không khí với bốn thông số: TSP, SO_2 , NO_2 , CO, đơn vị tính $\mu g/m^3$, mỗi mẫu lấy trung bình trong một giờ, chỉ tiêu được VIMCERTS công nhận, thời điểm tháng 11 năm 2020 tại năm địa điểm: Vòng xoay An Lạc 218, Quốc lộ 1A, Phường Tân Tạo A, Bình Tân; Vòng xoay An Sương: 6C – 10Bis Trường Chinh, Quận 12; ngã tư Bảy Hiền – ngã tư Lý Thường Kiệt – Cách mạng tháng tám; 300 Điện Biên Phủ, Quận 3; ngã tư Hàng Xanh – Điện Biên Phủ giao với Xô Viết Nghệ Tĩnh, trong điều kiện trời nắng, các hoạt động bình thường được thể hiện trong bảng 1.

Ta tham số bốn thông số TSP, SO_2 , NO_2 , CO lần lượt bằng các biến X_2, X_3, X_4, X_5 . Dữ liệu về các biến quan sát được thể hiện hình 1. Dựa trên biểu đồ của hình 1 trên, dữ liệu thu thập của các biến quan sát khá tốt, không có thể hiện phần ngoại lai (outliers). Bảng thống kê mô tả các thông số cho ta thấy, ngoài thành phần CO, các thông số còn lại đều vượt ngưỡng so với QCVN 05:2013/BTNMT và đặc biệt NO_2 có biên độ dao động khá mạnh so với trung bình (bảng 2). Biểu đồ tán xạ và hệ số tương quan các thông số quan trắc thể hiện hình 2.

Dựa vào kết quả phân tích mối tương quan (dương) giữa bụi tổng lơ lửng TSP với các thông số còn lại bao gồm SO_2 , NO_2 , CO khá cao đặc biệt là hai thông số SO_2 , NO_2 được thể hiện ở mức tin cậy cao (hình 2).

Dựa vào kết quả quan sát, với độ tin cậy 95%, ta có thể ước lượng lượng bụi tổng lơ lửng

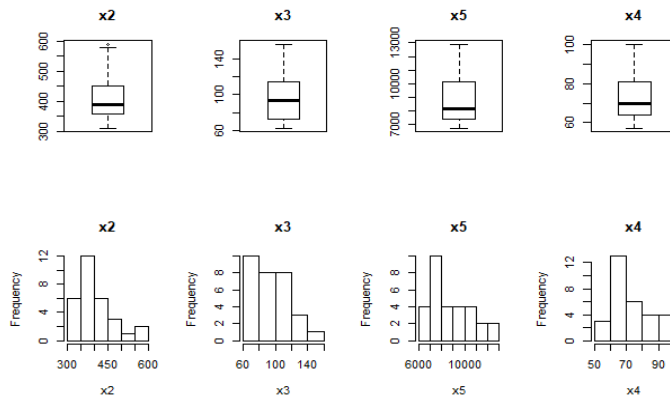
của TP.HCM vào khoảng (381,07; 434,26) $\mu\text{g}/\text{m}^3$ đều vượt khá cao so với QCVN 05:2013/BTNM. Ta xây dựng các phương trình hồi qui, mô tả sự phụ thuộc của các thông số SO_2 , NO_2 , CO lần lượt vào thông số TSP với các kết quả lần lượt như sau (bảng 3). Trong đó giá trị của thống kê F trong từng phương trình hồi qui đều có ý nghĩa thống kê.

Dựa vào bảng 3, ta thấy mô hình $\hat{Y}_3 \approx 18.78 + 0.3X_3 + 5.98X_4 - 0.01X_5$ có hệ số xác định hiệu chỉnh là 0.71, cao nhất trong 3 mô hình, trong khi giá trị thống kê $F = 12.9$ có ý nghĩa thống kê. Do đó ta lựa chọn mô hình này. Đường hồi qui thực nghiệm và đường hồi qui tính toán thể hiện ở hình 3.

Bảng 1. Kết quả quan trắc.

STT	TSP	SO_2	NO_2	CO
1	330	105	68	7500
2	360	96	64	7900
3	380	108	73	8100
4	410	104	68	7300
5	370	118	75	7800
6	430	129	79	9300
7	480	120	90	9200
8	540	114	93	9000
9	580	119	98	10400
10	450	125	94	9350
11	460	137	85	9800
12	590	156	100	11200
13	310	62	62	6900
14	330	66	68	7400
15	360	72	70	7800
16	350	64	60	7100
17	380	66	64	6700
18	400	73	74	7700
19	390	70	57	6800
20	410	75	60	7000
21	430	73	64	8200
22	380	68	67	7600
23	360	83	64	7200
24	480	90	78	8500
25	390	91	63	10100
26	370	95	67	10200
27	350	84	70	10400
28	320	86	74	11800
29	390	98	81	12100
30	450	104	87	12900

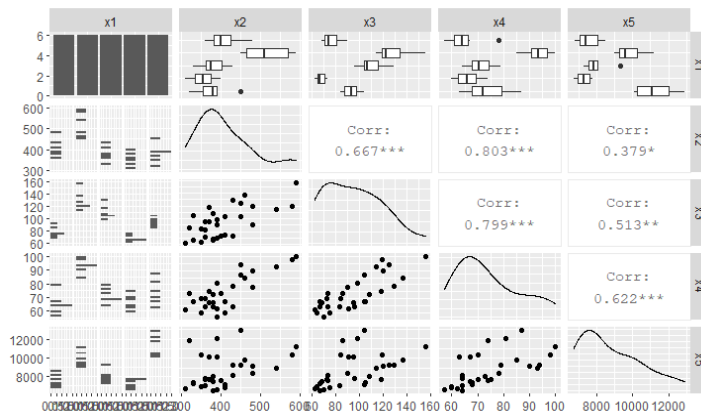
Nguồn. Nhóm tác giả.



Hình 1. Biểu đồ dữ liệu quan sát.
Nguồn. Nhóm tác giả **Bảng 2.** Thống kê mô tả kết quả quan trắc.

	X ₂	X ₃	X ₄	X ₅
Mean	407.67	95.03	73.90	8,775.00
SE	13.00	4.48	2.21	314.36
Med	390.00	93.00	70.00	8,150.00
Min	310.00	62.00	57.00	6,700.00
Max	590.00	156.00	100.00	12,900.00
Sum	12,230.00	2,851.00	2,217.00	263,250.00
Count	30.00	30.00	30.00	30.00

Nguồn. Nhóm tác giả.



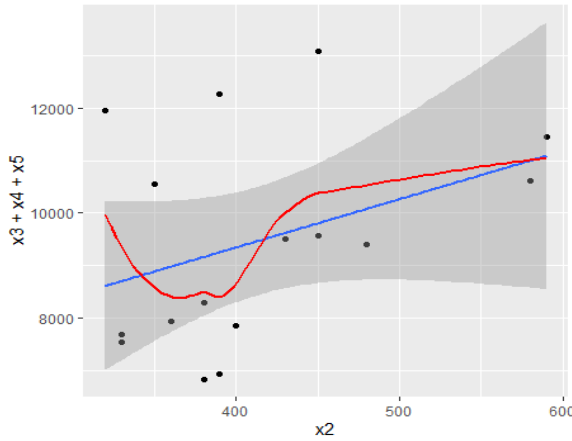
Hình 2. Phân tích tương quan và hàm mật độ các biến quan sát.

Nguồn. Nhóm tác giả.

Bảng 3. Các phương trình hồi qui.

STT	Phương trình hồi qui	R ² hiệu chỉnh	Thống kê F
1	$\hat{Y}_1 \approx 188.65 + 2.27X_2$	0.52	17.25
2	$\hat{Y}_2 \approx 4.98 + 0.42X_3 + 4.7X_4$	0.68	16.65
3	$\hat{Y}_3 \approx 18.78 + 0.3X_3 + 5.98X_4 - 0.01X_5$	0.71	12.9

Nguồn. Nhóm tác giả.



Hình 3. Đường hồi qui thực nghiệm (màu đỏ) và đường hồi qui tuyến tính (màu xanh) của phương trình $\hat{Y}_3 \approx 18.78 + 0.3X_3 + 5.98X_4 - 0.01X_5$.

Nguồn. Nhóm tác giả.

2.2. Các mô hình chuỗi thời gian

Để đánh giá ô nhiễm không khí thông qua chỉ số AQI với dữ liệu đủ lớn trong thời gian dài, ta sử dụng dữ liệu từ quan trắc của Tổng Lãnh sự quán Hoa Kỳ tại TP. HCM từ tháng 1 năm 2017 đến tháng 8 năm 2020 từ website <https://www.airnow.gov> [1]. Bài báo sử dụng mô hình Box-Jenkins [2] để phân tích mô hình tuyến tính chuỗi thời gian. Phương pháp Box-Jenkins bao gồm ba yếu tố: Tự hồi quy tuyến tính (Autoregression-AR), Tích hợp (Integration-I) và thay đổi trung bình (Moving Average-MA). Mô hình được viết tắt: ARIMA.

Mô hình tự hồi quy cấp p , viết tắt ($AR(p)$) có dạng

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \epsilon_t \quad (1)$$

Trong phương trình (1), giá trị hiện tại của chuỗi Y_t là tổ hợp tuyến tính của p giá trị gần nhất và nhiều. Trong mô hình thay đổi trung bình (MA), giá trị của Y_t phụ thuộc vào những giá trị của nhiều ở một khoảng thời gian thay vì phụ thuộc vào những giá trị trong quá khứ của chuỗi. Phương trình của mô hình dịch chuyển trung bình cấp q có dạng:

$$Y_t = \mu + \epsilon_t - \theta_1 \epsilon_{t-1} - \dots - \theta_q \epsilon_{t-q}$$

Mô hình tổng hợp của tự hồi quy và thay đổi trung bình, kí hiệu bởi ($ARMA(p, q)$), trong đó p và q lần lượt là cấp của phần tự hồi quy và dịch chuyển trung bình. Phương trình có dạng:

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \epsilon_t + \mu + \epsilon_t - \theta_1 \epsilon_{t-1} - \dots - \theta_q \epsilon_{t-q} \quad (2)$$

Mô hình (2) này thích hợp để mô tả những chuỗi thời gian dừng. Do bao gồm cả hai phần tự hồi quy và trung bình dịch chuyển, việc dự đoán của mô hình này phụ thuộc vào hiện tại và quá khứ của chuỗi thời gian cũng như hiện tại và quá khứ của thặng dư. Khi chuỗi thời gian không dừng, tức là dữ liệu không dao động quanh một giá trị trung bình và một xu hướng thay đổi nào có thể nhận ra, có một cách để làm dữ liệu trở thành chuỗi dừng là xem xét sai phân của chuỗi thời gian ban đầu. Định nghĩa toán tử dịch chuyển lùi $B: BY_t = Y_{t-1}$. Sai phân cấp 1 được định nghĩa bởi: $\Delta Y_t := Y_t - Y_{t-1} = (1 - B)Y_t$. Sai phân các cấp cao hơn được định nghĩa bởi $(1 - B)^d Y_t$. Mô hình tích hợp tự hồi quy và trung bình dịch chuyển có dạng:

$$Y'_t = c + \phi_1 Y'_{t-1} + \dots + \phi_p Y'_{t-p} + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q} + \epsilon_t \quad (3)$$

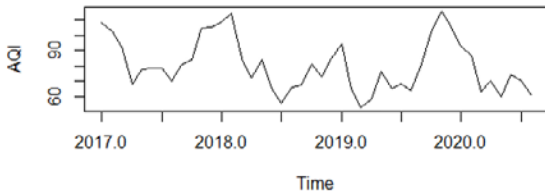
Trong đó Y'_t là chuỗi đã được chuyển hóa qua một toán tử hiệu nào đó. Mô hình (3) được viết dưới dạng $ARIMA(p, d, q)$ trong đó d là số lần lấy sai phân (differencing), trong thực tế người ta thường lấy $d \leq 2$. Phương trình có thể viết dưới dạng:

$$(1 - \phi_1 B - \dots - \phi_p B^p)(1 - B)^d Y_t = c + (1 + \theta_1 B + \dots + \theta_q B^q) \epsilon_t \quad (4)$$

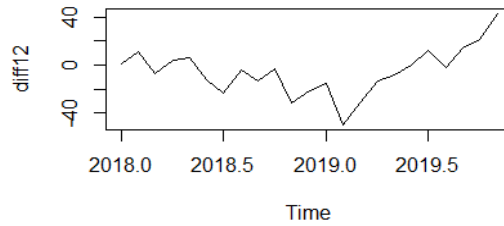
Để xác định sự chính xác của việc dự đoán dựa trên một mô hình nào đó, ta có thể xét các loại sai số sau:

- Sai số tuyệt đối trung bình: $MAE = \frac{1}{n} \sum_{t=1}^n |Y_t - \hat{Y}_t|$;
- Sai số phần trăm tuyệt đối trung bình $MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right| \times 10$;
- Sai số bình phương trung bình $MSE = \frac{1}{n} \sum_{t=1}^n (Y_t - \hat{Y}_t)^2$;
- Căn bậc hai sai số bình phương trung bình $RMSE = \sqrt{MSE}$.

Trong \hat{Y}_t là giá trị dự đoán, n là cỡ mẫu được dự đoán. Mô hình với các sai số nhỏ sẽ là mô hình tốt. Kết quả tính toán và mô phỏng được thực hiện dựa trên phần mềm R như sau:

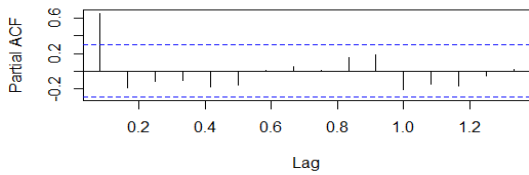


Hình 4. Dữ liệu AQI.
Nguồn. Nhóm tác giả.

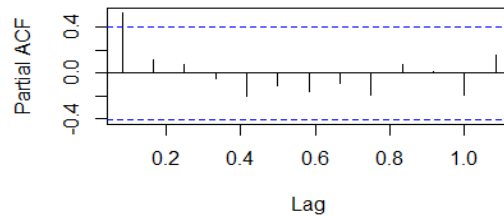


Series AQI

Series diff12



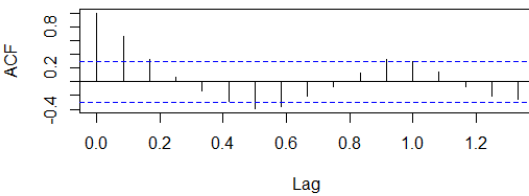
(a)



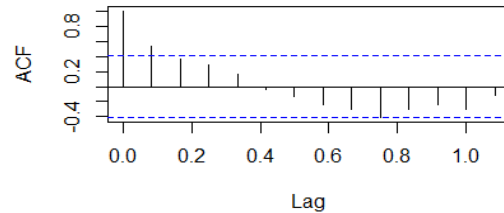
Lag

Series AQI

Series diff12



(b)



Lag

Hình 5. Đồ thị hệ số tự tương quan hồi quy ACF (a) và hệ số tự tương quan hồi quy thành phần PACF (b).
Nguồn. Nhóm tác giả.

Hình 6. Đồ thị hệ số sự khác biệt, hệ số tự tương quan hồi quy ACF và hệ số tự tương quan hồi quy thành phần và PACF có yếu tố mùa vụ.
Nguồn. Nhóm tác giả.

Đồ thị hệ số tương quan hồi quy và hệ số tương quan hồi quy thành phần ACF và PACF được vẽ để xác định giá trị thích hợp của p và q . Giá trị PACF không đáng kể đối với trễ, gợi ý $p = 0$. Giá trị ACF bé dần sau trễ hai gợi ý $q = 1$ hoặc 2 . Chuỗi thời gian không có xu hướng rõ ràng gợi ý $d = 0$.

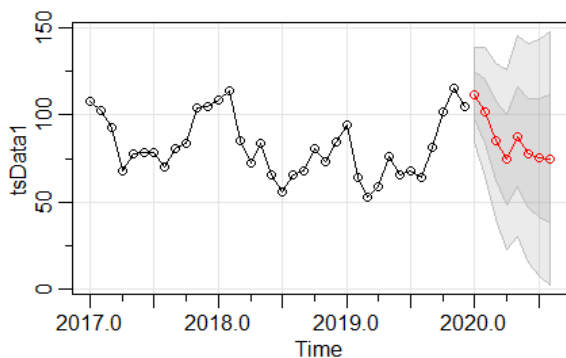
Để xác định ảnh hưởng mùa vụ, ta sử dụng mô hình ARIMA có ảnh hưởng đến yếu tố mùa vụ, gọi tắt là mô hình SARIMA (Seasonal ARIMA). Ta ký hiệu SARIMA(p, d, q)(P, D, Q)^S trong đó chỉ số S là yếu tố mùa vụ (xét $S = 12$, tương ứng với 12 tháng trong nghiên cứu này); bộ chỉ số (P, D, Q) biểu thị cho phần yếu tố mùa vụ (seasonal part). Từ đó, ta vẽ đồ thị và tính ACF, PACF của hiệu $Y_t - Y_{t-12}$ dưới đây:

Từ đồ thị ACF và PACF của $Y_t - Y_{t-12}$, ta có thể dự đoán $P = 0, Q = 1$ hoặc $Q = 2$.

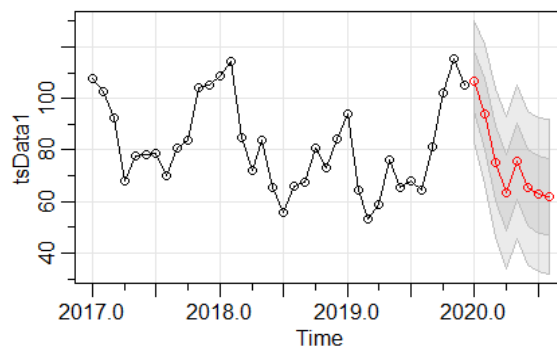
Với những dự đoán trên, nhóm chạy mô phỏng 10 mô hình với giá trị (p, d, q) và (P, D, Q) khác nhau. Ở đây nhóm chỉ trình bày một vài mô hình tiêu biểu:

- Mô hình 1: SARIMA(0, 1, 0)(0, 1, 1)¹²
- Mô hình 2: SARIMA(0, 1, 2)(0, 1, 1)¹²
- Mô hình 3: SARIMA(0, 1, 1)(0, 1, 1)¹²
- Mô hình 4: SARIMA(1, 1, 0)(0, 1, 1)¹²
- Mô hình 5: SARIMA(0, 0, 2)(0, 1, 1)¹²

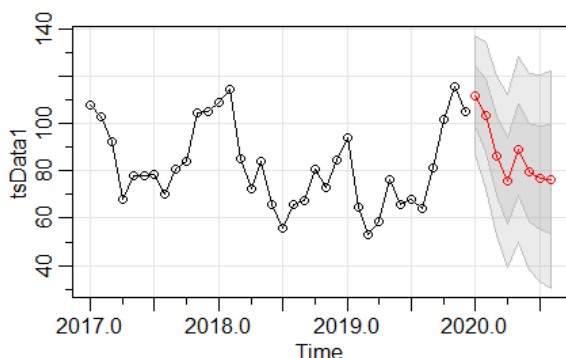
Số liệu được sử dụng từ tháng 1 năm 2017 đến tháng 12 năm 2019 để ước lượng các tham số của từng mô hình. Sau đó số liệu của tám tháng đầu năm 2020 được dùng để tính toán sai số dự đoán.



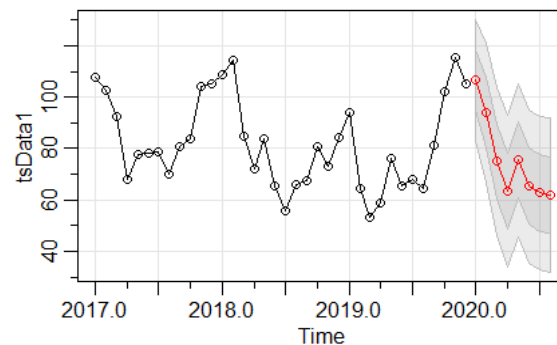
Hình 7. Đồ thị mô phỏng dự đoán cho mô hình 1.
Nguồn. Nhóm tác giả.



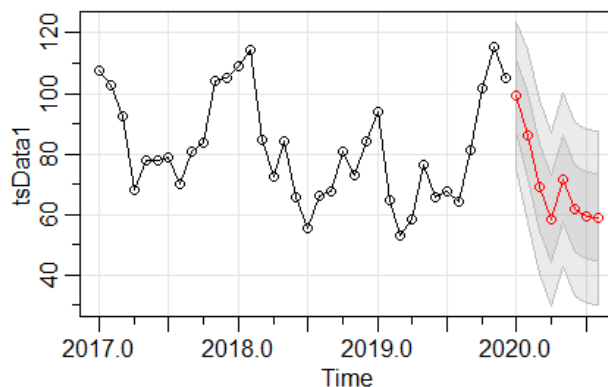
Hình 9. Đồ thị mô phỏng dự đoán cho mô hình 3.
Nguồn. Nhóm tác giả.



Hình 8. Đồ thị mô phỏng dự đoán cho mô hình 2.
Nguồn. Nhóm tác giả.



Hình 10. Đồ thị mô phỏng dự đoán cho mô hình 4.
Nguồn. Nhóm tác giả.



Hình 11. Đồ thị mô phỏng dự đoán cho mô hình 5.
Nguồn. Nhóm tác giả.

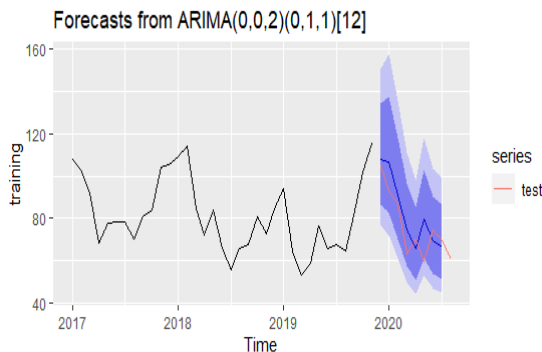
Các sai số dự đoán được đưa ra trong bảng sau:

Bảng 4. Sai số của các mô hình.

	ME	RMSE	MSE	MPE	MAPE
Mô hình 1	-13.885	16.148	13.884	-20.069	20.069
Mô hình 2	-15.092	17.150	15.093	-21.842	21.842
Mô hình 3	-17.262	19.130	17.262	-24.848	24.848
Mô hình 4	-3.2	9.984	8.972	-4.545	12.604
Mô hình 5	1.668	8.753	7.619	2.093	10.968

Nguồn. Nhóm tác giả.

Dựa vào bảng sai số của các mô hình (Bảng 4), ta thấy sai số của mô hình 5 là bé nhất do vậy ta chọn mô hình này trong công tác dự báo.



Hình 12. Dự báo vùng chỉ số AQI.
Nguồn. Nhóm tác giả.

Dựa trên đồ thị hình 12 ta có thể thấy vùng dự báo cho khoảng thời gian tiếp theo, (chẳng hạn sáu tháng đầu năm 2021) AQI dao động khoảng (45; 155), so sánh kết quả đường cong thực nghiệm tính toán khá gần với dữ liệu quan sát.

3. Kết luận và khuyến nghị

Bài báo đã sử dụng các mô hình hồi qui, mô hình tự hồi quy tuyến tính tích hợp và thay đổi trung bình (ARIMA) và mô hình tự hồi quy tuyến tính tích hợp và thay đổi trung bình có ảnh hưởng của yếu tố mùa vụ (SARIMA) để phân tích hiện trạng ô nhiễm và dự đoán tình trạng ô nhiễm từ các hoạt động xả thải của các phương tiện giao thông đường bộ tại TP. HCM và dự báo sự ô nhiễm này trong thời gian tới. Dưới tác động nghiêm trọng đến sức khỏe người dân, thiết nghĩ cả hệ thống chính quyền và mỗi người dân cần nâng cao ý thức trong việc giảm thiểu khí thải từ các phương tiện giao thông hàng ngày. Tuy nhiên, để đánh giá chính xác hơn về ảnh hưởng và những thay đổi về sức khỏe con người trong và sau khi tiếp xúc với không khí ô nhiễm, các cơ sở dữ liệu nghiên cứu vẫn cần được cải thiện. Đặc biệt, cần có nhiều nghiên cứu hơn về các mô hình và tác động xấu đến sức khỏe của việc tiếp xúc với cộng đồng và vai trò của các thành phần khác nhau trong hỗn hợp ô nhiễm. Cần có thêm nhiều nghiên cứu đánh giá lợi ích sức khỏe cộng đồng của các biện pháp cải thiện chất lượng không khí khác nhau, đặc biệt là thông qua các can thiệp nhằm giải quyết ô nhiễm không khí liên quan đến giao thông, từ đó có những hỗ trợ thích hợp cho các chính sách □

Tài liệu tham khảo

- [1] [https://www.airnow.gov/international/us-embassies-and-consulates/?fbclid=IwAR0jT3BUu2AgNjPrOLTXKW9CkY2jOckJ9IL2T533wrS1gOIy_13tIU68qxxg#Vietnam\\$Ho_Chi_Minh_City](https://www.airnow.gov/international/us-embassies-and-consulates/?fbclid=IwAR0jT3BUu2AgNjPrOLTXKW9CkY2jOckJ9IL2T533wrS1gOIy_13tIU68qxxg#Vietnam$Ho_Chi_Minh_City), Ngày truy cập: 20/09/2020;
- [2] Box G.E.P., Jenkins G.M. (1976), *Time Series Analysis: Forecasting and Control*, Revised Edition, San Francisco: Holden Day;
- [3] Bùi Tá Long, Nguyễn Hoàng Phong, Nguyễn Châu Mỹ Duyên (2019), *Nghiên cứu ứng dụng mô hình hoá tính toán ô nhiễm không khí cho nguồn thải đường và thể tích - trường hợp áp dụng tại mỏ khai thác đá tỉnh Bình Dương*, Tạp chí khí tượng thủy văn, số tháng 7;
- [4] Lee M.H., Rahman N., Suhartono S., Latif M.T., Nor M., Kamisan N. (2012), *Seasonal ARIMA for forecasting air pollution index: A case study*. Am. J. Appl. Sci., 9, 570 – 578;
- [5] Monika Z., Katarzyna G. (2020), *Regression Linear Model of Air Pollution Emission on the Example of a Waste Incineration Plant*, MDPI, 51, 32;
- [6] M. Jerrett, et al. (2005), *A review and evaluation of intraurban air pollution exposure models*, J Expo Anal Environ Epidemiol, 15, 185 – 204;
- [7] M. Jerrett, et al. (2005), *Spatial analysis of air pollution and mortality in Los Angeles*, Epidemiology 16, 727 – 736;
- [8] Xi C., et al. (2016), *Long-term exposure to urban air pollution and lung cancer mortality: A 12-year cohort study in Northern China*, Science of The Total Environment, 571, 855 – 861;
- [9] Tran Dinh Tuong (2020), *Epidemic SIS model in air-polluted environment*, J. Appl. Math. Comput., 64, p.53 - 69.

Lời cảm ơn: Bài báo này được tài trợ một phần từ đề tài “Xây dựng các mô hình toán học trong việc nghiên cứu tác động ô nhiễm môi trường từ các hoạt động giao thông đô thị đối với sức khỏe con người” với mã số DT203006. Nhóm tác giả xin bày tỏ lòng biết ơn đến Bộ Giao thông vận tải, Trường Đại học Giao thông vận tải Thành phố Hồ Chí Minh đã hỗ trợ trong quá trình nghiên cứu. Cuối cùng, nhóm nghiên cứu xin chân thành cảm ơn đến các đơn vị: Viện nghiên cứu Môi trường và Giao thông, Viện Đào tạo và Hợp tác Quốc tế (IEC), Khoa Cơ Bản - Trường Đại học Giao thông vận tải Thành phố Hồ Chí Minh; Khoa Cơ Bản, trường Cao đẳng Công thương Tp. HCM, Trung tâm Quan trắc tài nguyên môi trường (CEM) - Sở Tài nguyên và Môi trường Thành phố Hồ Chí Minh đã tạo điều kiện thuận lợi để bài báo được hoàn thành.

Ngày nhận bài: 25/12/2020

Ngày chuyển phản biện: 29/12/2020

Ngày hoàn thành sửa bài: 19/01/2021

Ngày chấp nhận đăng: 26/01/2021