

Bài báo khoa học

## Dự báo khu vực lan truyền nhạy cảm Covid-19 dựa vào phương pháp hồi quy

Bùi Hồng Sơn<sup>1\*</sup>, Dương Thị Thúy Nga<sup>2\*</sup>, Lê Trung Chơn<sup>3</sup>, Đồng Thị Bích Phương<sup>4</sup>,  
Phạm Minh Khan<sup>2</sup>

<sup>1</sup> Trung tâm Công nghệ thông tin – Sở TN&MT TPHCM; son.ciren@gmail.com

<sup>2</sup> Trường Đại học Tài nguyên và Môi trường TPHCM; dttnnga\_cntt@hcmunre.edu.vn;  
pmkhan@hcmunre.edu.vn

<sup>3</sup> Trường Đại học Bách Khoa TPHCM; ltchon@hcmut.edu.vn

<sup>4</sup> Viện Khoa học đo đạc và bản đồ; dbphuong@yahoo.com

\*Tác giả liên hệ: son.ciren@gmail.com; dttnnga\_cntt@hcmunre.edu.vn; Tel.: +84–  
903746638; +84–938589589

Ban Biên tập nhận bài: 5/2/2022; Ngày phản biện xong: 16/3/2022; Ngày đăng bài:  
25/4/2022

**Tóm tắt:** Các diễn biến phức tạp và sự lây lan nhanh chóng của Covid-19 đang diễn ra tại Thành phố Hồ Chí Minh đang là một thách thức cho chính quyền và người dân thành phố trong công tác phòng và dập dịch. Để kiểm soát được dịch bệnh, thành phố cần một giải pháp có thể dự báo và phân vùng được các vùng dịch để có thể đưa ra những giải pháp chính xác và nhanh nhất. Trong nghiên cứu này, chúng tôi đề xuất một giải pháp là áp dụng các mô hình hồi quy địa lý (Geographically Weighted Regression gọi tắt là GWR) cục bộ, không phải là toàn cục (global) như các mô hình hồi quy khác như Ordinary Least Squares (OLS). Ưu điểm của giải pháp này là GWR sẽ loại trừ bớt sai số do thiếu các yếu tố xã hội mà mô hình chưa có dữ liệu để sử dụng. Ưu điểm của giải pháp này là nó sẽ loại trừ bớt sai số do thiếu các yếu tố xã hội mà mô hình chưa có dữ liệu để sử dụng. Giải pháp của chúng tôi đề xuất có thể dự báo và đánh giá tính hình dịch trên toàn địa bàn thành phố, phạm vi dự báo là các vùng nhạy cảm cao và rất cao, nơi xảy ra hiện tượng nhiễm bệnh nhiều và rất nhiều. Với các kịch bản dự báo, chúng tôi thử nghiệm tại Thành phố Hồ Chí Minh và kết quả khá sát với dữ liệu thực trong dự báo ngắn hạn.

**Từ khóa:** Covid-19; Dịch bệnh; Dữ liệu; Hồi quy; Nhạy cảm.

### 1. Giới thiệu

COVID-19 là một đại dịch mới do hội chứng coronavirus hô hấp cấp tính 2 (SARS-CoV-2) gây ra và lây lan nhanh chóng từ người này sang người khác. Đại dịch này là cuộc khủng hoảng thế giới đáng chú ý nhất, bùng phát bắt nguồn từ Vũ Hán, Trung Quốc, vào cuối năm 2019. Tổ chức Y tế Thế giới (WHO) cho biết tên gọi chính thức của bệnh viêm đường hô hấp cấp do chủng mới của vi-rút corona (nCoV) gây ra là COVID-19. Tên gọi mới này gọi tắt của coronavirus disease 2019. Tháng 2/2020, Ủy ban quốc tế về phân loại Virus – International Committee on Taxonomy of Viruses (ICTV) chính thức đặt tên cho chủng mới của vi-rút corona là SARS-CoV-2. Đây là tên gọi khác với tên virus Corona mới (2019-nCov) mà WHO đã chỉ định trước đó. Các nhà khoa học đã tiến hành nghiên cứu và phân lập được một chủng corona virus mới, có trình tự gen giống với SARS-CoV trước đây, với mức tương đồng lên tới 79,5%.

Đây là dạng virus mới nên con người chưa từng có miễn dịch, kể cả miễn dịch chéo trước đó và đã gây ra một cuộc khủng hoảng kinh tế toàn cầu. Việt Nam trong đợt bùng dịch thứ 4

này đã chịu rất nhiều tổn thất về kinh tế, đặc biệt là con người, tính đến ngày 02 tháng 10 năm 2021 cả nước có 803,149 người nhiễm bệnh, 19,601 người mất. Riêng tại thành phố Hồ Chí Minh có 395,052 người nhiễm bệnh.

Gần đây, mô hình thống kê và chuỗi thời gian cũng đã được giới thiệu để mô hình hóa và dự đoán mức độ phổ biến của đại dịch này. Nhưng nghiên cứu này nhằm dự đoán sự lây nhiễm của COVID-19 ở Thành phố Hồ Chí Minh, Việt Nam chúng tôi sử dụng phương pháp hồi quy OLS – Ordinary Least Square và hồi quy có trọng số không gian (*Geographically Weighted Regression*).

## 2. Phương pháp

### 2.1. Phương pháp hồi quy OLS – Ordinary Least Square

Ước lượng các tham số của mô hình:

$$\text{Giá trị thực tế } Y_i = \beta_1 + \beta_2 X_i + e_i$$

$$\text{Giá trị ước lượng } \hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$$

$$\text{Sai số } e_i = Y_i - \hat{Y}_i = Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i$$

Tìm  $\hat{\beta}_1, \hat{\beta}_2$  sao cho tổng bình phương sai số là nhỏ nhất

$$\text{Tức là } \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)^2 \rightarrow \min$$

Giải bài toán cực trị hàm hai biến, ta được:

$$\hat{\beta}_2 = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2} = \frac{\sum_{i=1}^n Y_i X_i - n \bar{X} \bar{Y}}{\sum_{i=1}^n X_i^2 - n (\bar{X})^2} = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}$$

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}$$

Trong đó

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} \text{ là giá trị trung bình của X và } x_i = X_i - \bar{X}$$

$$\bar{Y} = \frac{\sum_{i=1}^n Y_i}{n} \text{ là giá trị trung bình của Y và } y_i = Y_i - \bar{Y}$$

Các giả thiết của OLS:

Giả thiết 1: Quan hệ giữa Y và X là tuyến tính Các giá trị  $X_i$  cho trước và không ngẫu nhiên.

Giả thiết 2: Các sai số  $U_i$  là đại lượng ngẫu nhiên có giá trị trung bình bằng 0.

$$E(U_i | X_i)$$

Giả thiết 3: Các sai số  $U_i$  là đại lượng ngẫu nhiên có phương sai không thay đổi.

$$\text{Var}(U_i | X_i) = \sigma^2 = \text{const}$$

Giả thiết 4: Không có sự tương quan giữa các  $U_i$ .

$$\text{Cov}(U_i, U_j | X_i, X_j) = 0, i \neq j$$

Giả thiết 5: Không có sự tương quan giữa  $U_i$  và  $X_i$ .

$$\text{Cov}(U_i, X_i) = 0$$

Định lý Gauss–Markov: Khi các giả thiết này được đảm bảo thì các ước lượng tính được bằng phương pháp OLS là các ước lượng tuyến tính không chệch, hiệu quả nhất của hàm hồi quy tổng thể.

Giả thiết 6: các sai số  $U_i$  có phân phối chuẩn  $U_i \approx N(0, \sigma^2)$

Hệ số xác định của mô hình:

Tổng bình phương toàn phần TSS (*Total Sum of Squares*):

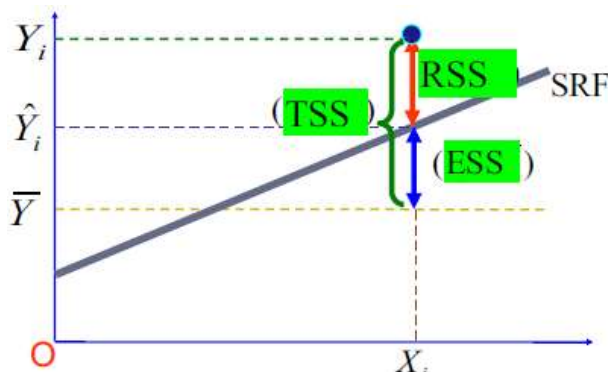
$$TSS = \sum_{i=1}^n (Y_i - \bar{Y})^2 = \sum_{i=1}^n Y_i^2 - n(\bar{Y})^2$$

Tổng bình phương hồi quy ESS (*Explained Sum of Squares*):

$$ESS = \sum (\hat{Y}_i - \bar{Y})^2 = \hat{\beta}_2^2 (\sum X_i^2 - n\bar{X}^2)$$

Tổng bình phương phần dư RSS (*Residual Sum of Squares*):

$$RSS = \sum (Y_i - \hat{Y}_i)^2 = \sum e_i^2$$



Ta có:  $TSS = ESS + RSS$

Hệ số xác định:  $R^2 = 1 - \frac{RSS}{TSS} = \frac{ESS}{TSS}$

Lưu ý:  $0 \leq R^2 \leq 1$ ;  $R^2 = 1$ : Mô hình phù hợp hoàn toàn với mẫu nghiên cứu;  $R^2 = 0$ : Mô hình hoàn toàn không phù hợp với mẫu nghiên cứu.

## 2.2. Phương pháp hồi quy địa lý (GWR)

Phương pháp hồi quy địa lý (GWR) mô hình hóa dữ liệu bằng các phương pháp hồi quy nhưng có thêm phần tính toán dựa trên dữ liệu địa lý. Nó mô tả thêm mối liên quan về vị trí địa lý của các giá trị. Mô hình hồi quy có trọng số địa lý sẽ có dạng [8]:

$$y_i = \beta_0(u_i, v_i) + \sum_k \beta_k(u_i, v_i)x_{ik} + \varepsilon_i$$

So sánh với công thức mô hình hồi quy tuyến tính:

$$y_i = \beta_0 + \sum_k \beta_k x_{ik} + \varepsilon_i$$

Trong đó  $(u_i, v_i)$  biểu thị tọa độ của điểm thứ  $i$  trong không gian và  $\beta_k(u_i, v_i)$  thể hiện giá trị của hàm  $\beta$  tại điểm  $i$ . Dễ nhận thấy công thức mô hình hồi quy tuyến tính là giá trị đặc biệt của công thức mô hình hồi quy địa lý, khi các tham số không gian  $\beta_k(u_i, v_i)$  là bất biến bằng một. Tương tự như hồi quy tuyến tính ta cần tìm các tham số hồi quy  $\beta_k(u_i, v_i)$ . Nhưng trong GWR các điểm ở gần có ảnh hưởng lớn hơn các điểm ở xa. Vậy nên sẽ có trọng số địa lý để thể hiện mức độ ảnh hưởng khác nhau [8].

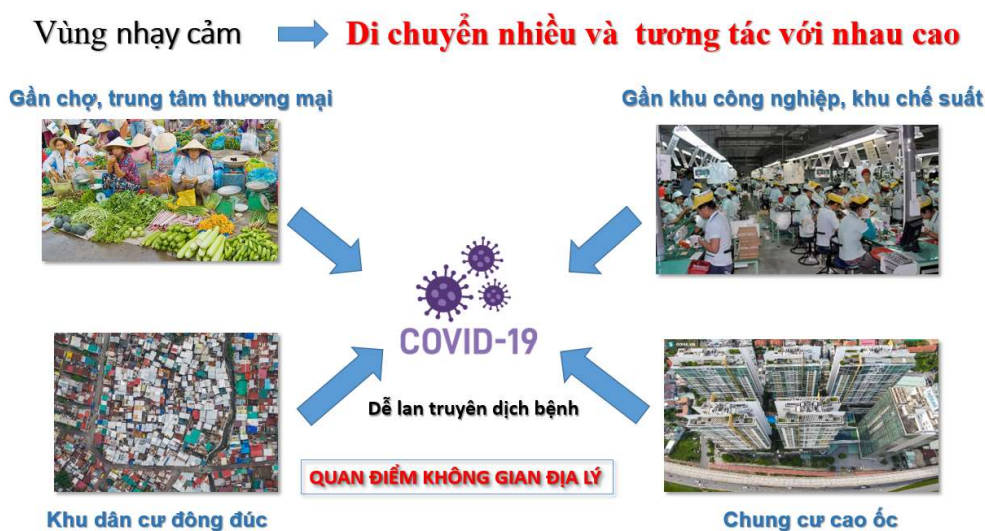
$$Y = \beta X + \varepsilon$$

$$\hat{\beta}(u_i, v_i) = (X^T W(u_i, v_i) X)^{-1} X^T W(u_i, v_i) y$$

Trong đó  $\beta$  là ma trận trọng số địa lý của  $x$  tương ứng;  $\hat{\beta}(u_i, v_i)$  là ma trận các tham số ước tính của  $\beta_k(u_i, v_i)$ .  $X$  là ma trận các giá trị  $x$ .  $Y$  là ma trận các giá trị  $y$ .  $X^T$  là ma trận chuyển vị của  $X$ .  $X^{-1}$  là ma trận nghịch đảo của  $X$ .  $W(u_i, v_i)$  là một ma trận đường chéo thể hiện trọng số địa lý của dữ liệu cho điểm hồi quy  $i$ .

## 3. Phân tích mức độ nhạy cảm và phân vùng nhạy cảm

Phân tích xác định mức độ nhạy cảm lan truyền dịch dựa trên quan điểm các vùng có tiếp xúc cao là có nguy cơ cao. Đó là các khu vực đông dân cư, các khu chung cư, cao ốc, chợ, siêu thị, trung tâm thương mại, khu công nghiệp, khu chế xuất, các trung tâm hành chính, dịch vụ...



Hình 1. Các đối tượng địa lý có nguy cơ lan truyền dịch bệnh.

Dữ liệu sử dụng trong thử nghiệm bao gồm:

- Dữ liệu dân số của World Population năm 2020 độ phân giải 90 x 90m.
- Dữ liệu các đối tượng hạ tầng kinh tế xã hội của Vietbando bao gồm chợ (226 điểm), siêu thị (255 điểm), chung cư, cao ốc (1076 điểm), khu công nghiệp, khu chế xuất, nhà máy (41 điểm).

Phần mềm mã mở QGIS được sử dụng trong phân tích các quan hệ không gian và tính toán các chỉ số nguy cơ.

Dữ liệu đối chứng là Heatmap dựa theo phân bố F0 khu vực TP.HCM đợt dịch từ đầu tháng 5 đến tháng 7 năm 2021.

Quy trình xây dựng bản đồ phân vùng nhạy cảm lan truyền bệnh:

1. Lượng hóa các giá trị đặc trưng cho mức độ nhạy cảm của từng ô (cell) kích thước 100x100m (tương đương 01hectar) bao gồm tổng dân số trong bán kính 500 mét và khoảng cách từ từng cell đến các đối tượng địa lý nhạy cảm, để trở thành tâm lan truyền dịch như chợ, siêu thị, chung cư, khu công nghiệp

2. Xác định các chỉ số nhạy cảm cho từng cell theo công thức sau:

Đối với dân số xác định chỉ số PopIndex:

$$\text{PopIndex}_i = \frac{\text{Pop2020}_i}{\text{MaxOfPop2020}} \quad (1)$$

Trong đó  $\text{Pop2020}_i$  là dân số trong vòng 500 mét đối với từng cell và  $\text{MaxOfPop2020}$  là giá trị Max của  $\text{Pop2020}$ .

- Chỉ số nhạy cảm đối với chợ

$$\text{ChoIndex}_i = \frac{(1000 - \text{Dist2cho}_i)}{1000} \quad (2)$$

Trong đó  $\text{Dist2cho}_i$  là khoảng cách theo đường chim bay đến chợ của từng cell.

- Chỉ số nhạy cảm đối với chung cư, cao ốc:

$$\text{ChungcuIndex}_i = \frac{(1000 - \text{Dist2cc}_i)}{1000} \quad (3)$$

Trong đó  $\text{Dist2cc}_i$  là khoảng cách từng cell đến chung cư.

- Chỉ số nhạy cảm đối với siêu thị, trung tâm thương mại:

$$\text{SiethiIndex}_i = \frac{(1000 - \text{Dist2st}_i)}{1000} \quad (4)$$



Trong đó  $Dist2st_i$  là khoảng cách từng cell đến siêu thị.  
 - Chỉ số nhạy cảm đối với khu công nghiệp, khu chế xuất:

$$KhuCNIndex_i = \frac{(1000 - Dist2c_i)}{1000} \quad (5)$$

Trong đó  $Dist2cn_i$  là khoảng cách từng cell đến khu công nghiệp (tính theo đường thẳng, không tính theo lộ trình).

Chỉ số nhạy cảm tổng hợp xác định theo công thức sau:

$$SensIndex_i = k_p \cdot PopIndex_i + k_{ch} \cdot ChoIndex_i + k_{cc} \cdot ChungcuIndex_i + k_{st} \cdot SieuthiIndex_i + k_{cn} \cdot KhuCNIndex_i \quad (6)$$

Trong đó  $k_p, k_{ch}, k_{cc}, k_{st}, k_{cn}$  là các trọng số các yếu tố nhạy cảm dân số, chợ, chung cư, siêu thị, khu công nghiệp ...

Do chuỗi số liệu thu thập được chưa đủ lớn, trong thử nghiệm dưới đây  $k_p, k_{ch}, k_{cc}, k_{st}, k_{cn}$  nhận giá trị bằng 1.

Để thuận lợi trong khai thác thực quan, sử dụng giá trị chỉ số nhạy cảm tổng hợp tương đối:

$$CovidSensIndex_i = 100 \cdot \frac{SensIndex_i}{MaxOfSensIndex} \quad (7)$$

Trong đó  $MaxOfSensIndex$  là giá trị lớn nhất trong các giá trị  $SensIndex$  của tất cả các cell.

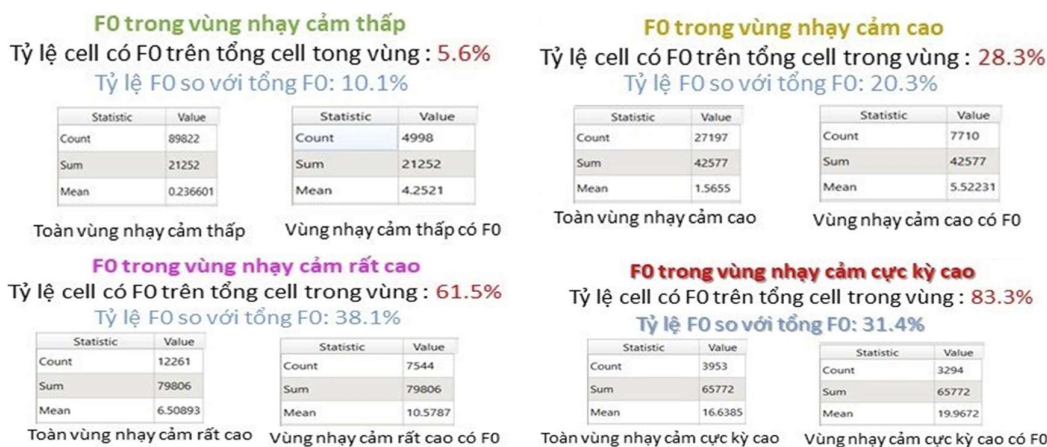
#### 4. Kết quả và thảo luận

Vùng nhạy cảm càng cao thì càng dễ bị lây nhiễm ( $5.6\% \rightarrow 28.3\% \rightarrow 61.5\% \rightarrow 83.3$ ), vùng nhạy cảm rất cao và cực kỳ cao chiếm gần 70% số ca F0 trên toàn thành phố và vùng nhạy cảm càng cao thì mật độ F0/ha càng lớn khi bị nhiễm (Số F0 trung bình = 20F0/ha).



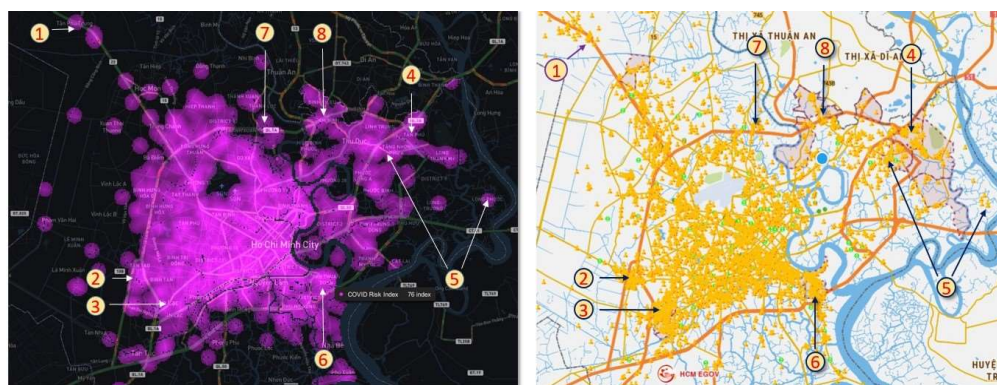
**Hình 2.** Số liệu thống kê theo không gian địa lý thì dữ liệu F0 từ ngày 27/04/2021 đến ngày 07/09/2021.

Trong đó tổng F0: 209,407, tổng cell bị F0: 23,546, tổng cell: 133,233 biết rằng 1 cell = 1 hectare.

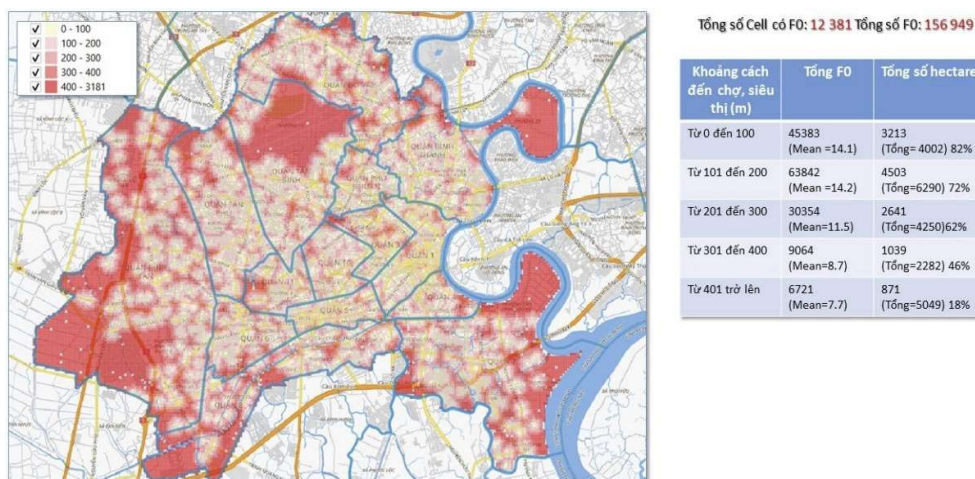


Đa số các khu vực có chỉ số nhạy cảm thấp có dân cư thưa thớt và ở ngoại thành, khu có chỉ số nhạy cảm cao nằm ở quận vùng ven và khu vực có chỉ số nhạy cảm rất cao nằm nội thành. Một số khu vực có sự đan xen các vùng nhạy cảm do đặc thù cảnh quan đô thị.

Kết quả so sánh dữ liệu xuất hiện các ca dương tính trên địa bàn thành phố trên bản đồ COVID-19 trực tuyến của thành phố ngày 18/07/2021 (<https://bando.tphcm.gov.vn/ogis>) có nhiều điểm trùng khớp, tương đồng, đặc biệt tại những vùng có sự đan xen giữa các mức độ nhạy cảm khác nhau.

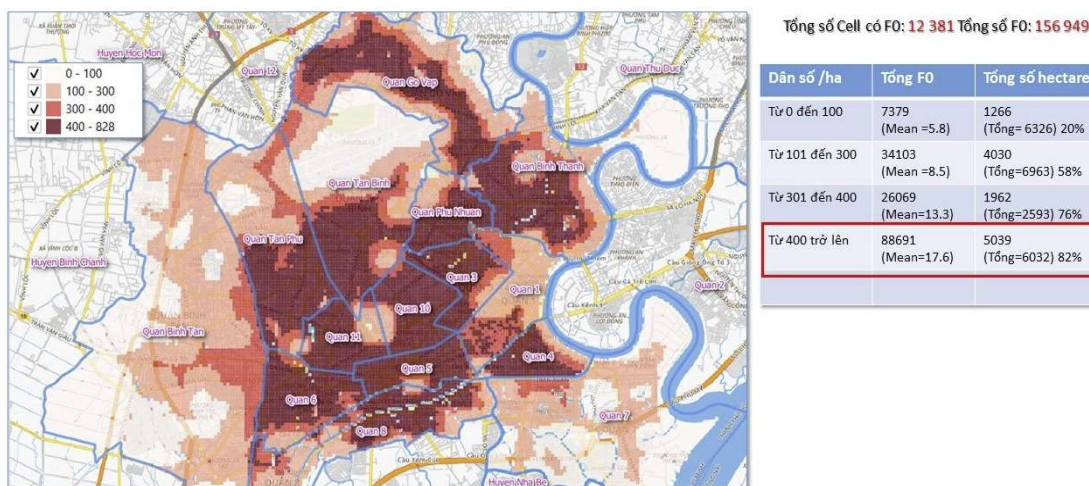


**Hình 3.** Các vị trí đặc biệt trên bản đồ vùng nhạy cảm cao đến rất cao (trái) và vị trí các ca dương tính đến ngày 18/07/2021 (phải).

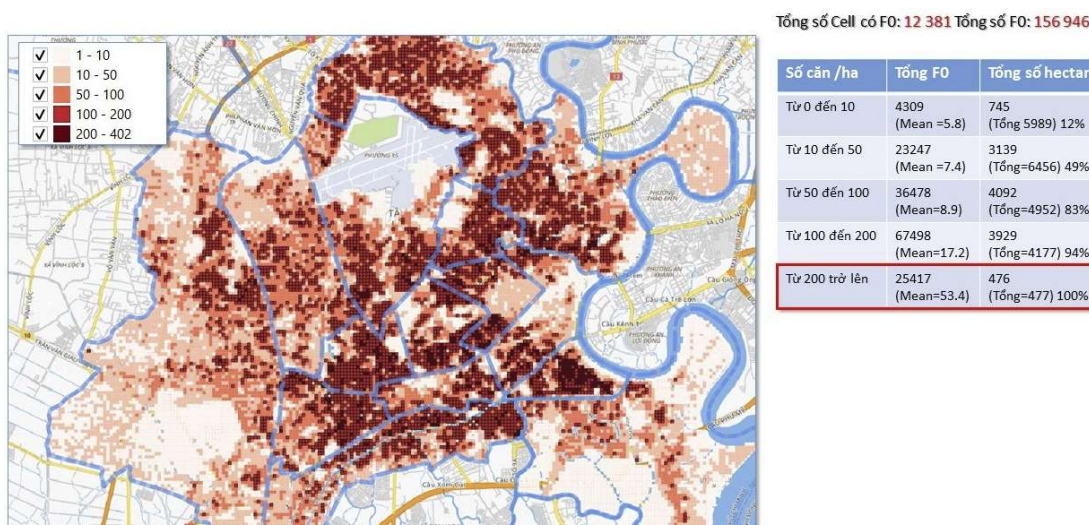


**Hình 4.** Số liệu thống kê F0 theo khoảng cách đến chợ, siêu thị (mét), dữ liệu F0 từ ngày 27/04/2021 đến ngày 07/09/2021.





**Hình 5.** Số liệu thống kê F0 theo mật độ dân số trên một cell (1 cell = 1 hectare), dữ liệu F0 từ ngày 27/04/2021 đến ngày 07/09/2021.



**Hình 6.** Số liệu thống kê F0 theo mật độ căn nhà (không tính chung cư) trên một cell (1 cell = 1 hectare), dữ liệu F0 từ ngày 27/04/2021 đến ngày 07/09/2021.

Như vậy các ca dương tính xuất hiện đều tập trung vào các vùng nhạy cảm cao và rất cao cho dù phạm vi không gian của các vùng này là rất nhỏ so với vùng nhạy cảm thấp. Các số liệu và sự phân bố không gian địa lý cũng nói lên rằng mối quan hệ không gian địa lý giữa các đối tượng nguy cơ trên địa bàn Thành phố Hồ Chí Minh có thể làm cơ sở để phân vùng nhạy cảm lan truyền dịch bệnh qua tiếp xúc, cụ thể đối với dịch COVID-19 đang hoành hành tại Thành phố Hồ Chí Minh.

Kết quả nghiên cứu cho thấy việc phân vùng nhạy cảm lan truyền dịch bệnh qua tiếp xúc dựa trên mối quan hệ không gian địa lý là cần thiết và mang nhiều ý nghĩa thực tiễn:

Khẳng định vùng nhạy cảm và vùng nguy cơ lan truyền dịch không phụ thuộc vào địa giới hành chính, do vậy các biện pháp phòng ngừa và ứng phó cần phải được xác định trên cơ sở khoa học và thực tiễn.

Công cụ xác định vùng nhạy cảm và vùng nguy cơ các mức độ khác nhau trong thời gian ngắn để có biện pháp phòng ngừa và khống chế hợp lý với nguồn lực hiện có.

Xây dựng bản đồ phân vùng nhạy cảm và phân vùng nguy cơ lan truyền dịch qua tiếp xúc sẽ là cơ sở hoạch định cơ sở hạ tầng, tổ chức triển khai phòng chống dịch đúng nơi, đúng chỗ và có trọng tâm, trọng điểm tránh bị thụ động.

Phân vùng nhạy cảm lan truyền dịch là cơ sở để xác định, phân vùng nguy cơ, theo đó, song song với biện pháp truy vết, cách ly, xét nghiệm có thể có các biện pháp chủ động phòng thủ nơi nguy cơ cao và rất cao như xét nghiệm tầm soát, tăng cường ý thức chống dịch, kiểm tra, giám sát thực hiện quy định chống dịch.

#### 4. Kết luận

Kết quả nghiên cứu, thử nghiệm cho thấy quan hệ không gian giữa các đối tượng địa lý có thể là cơ sở phân vùng nhạy cảm lan truyền dịch bệnh, hỗ trợ các cơ quan chức năng trong việc ra quyết định, thực hiện các biện pháp phòng chống dịch hiệu quả.

Mô hình dự báo GWR giúp cho việc dự báo và phân loại các vùng địa lý để phân loại các nguy cơ từ thấp đến cao, từ đó giúp cho việc khoanh vùng phong tỏa và điều trị được chuẩn xác hơn.

**Đóng góp của tác giả:** Xây dựng ý tưởng nghiên cứu: L.H.T.; Lựa chọn phương pháp nghiên cứu: L.H.T., D.N.D.P.; Xử lý số liệu: D.N.D.P., P.T.H.; Chạy mô hình: L.H.T., D.N.D.P., P.T.H.; Phân tích và đánh giá kết quả: L.H.T., N.K.L.; Viết bản thảo bài báo: L.H.T., D.N.D.P., P.T.H., N.K.L., N.T.H.; Chỉnh sửa bài báo: L.H.T., D.N.D.P., P.T.H., N.T.H., N.K.L.

**Lời cảm ơn:** Để hoàn thành nghiên cứu này, chúng tôi trân trọng gửi lời cảm ơn đến Sở Khoa học và Công nghệ tỉnh Gia Lai đã cấp kinh phí thông qua đề tài: “Xác định tập đoàn giống cây trồng nông nghiệp chủ lực của tỉnh Gia Lai”, Mã số: KHGL-09-18.

**Lời cam đoan:** Tập thể tác giả cam đoan bài báo này là công trình nghiên cứu của tập thể tác giả, chưa được công bố ở đâu, không được sao chép từ những nghiên cứu trước đây; không có sự tranh chấp lợi ích trong nhóm tác giả.

#### Tài liệu tham khảo

1. Hiền, L.T.T. Ảnh hưởng của kích thước ô lưới tới kết quả tính toán thủy lực dòng chảy lũ. *Tạp chí Khoa học và Công nghệ Thủy lợi* **2017**, *40*, 107–117.
2. Thảo, N.Đ.P. và cs. Ứng dụng GIS và phương pháp phân tích đa chỉ tiêu xác định vị trí bãi chôn lấp chất thải rắn cho quận Thủ Đức, thành phố Hồ Chí Minh. Hội thảo ứng dụng GIS toàn quốc, 2011.
3. Việt, N.Q. và cs. Ứng dụng GIS và đánh giá đa tiêu chí xác định khu vực thích hợp cho phát triển khu dân cư trong bối cảnh biến đổi khí hậu ở huyện Phú Vang, tỉnh Thừa Thiên Huế. Kỷ yếu hội nghị Khoa học Địa lý toàn quốc lần thứ XI, 2019.
4. Richards, M.; Ghanem, M.; Osmond, M.; Guo, Y.; Hasard, J. Grid-based analysis of air pollution data. *Ecol. Modell.* **2006**, *194*, 274–286.
5. Wesseloo, J.; Woodward, K.; Pereira, J. Grid-based analysis of seismic data. *J. S. Afr. Inst. Min. Metall.* **2014**, *114(10)*, 815–822.
6. Bagan, H.; Yamagata, Y. Landsat analysis of urban growth: How Tokyo became the world’s largest megacity during the last 40 years. *Remote Sens. Environ.* **2012**, *127*, 210–222.
7. Su, M.D.; Kang, J.L.; Chang, L.F.; Chen, A.S. A grid-based GIS approach to regional flood damage assessment. *J. Mar. Sci. Technol.* **2005**, *13(3)*, 184–192.
8. Fotheringham, A.S.; Brunson, C.; Charlton, M. Geographically Weighted Regression: The Analysis of Spatially Varying Relationships, Wiley, 2002.



## **Forecasting Covid–19 sensitive transmission areas based on regression method**

**Bui Hong Son<sup>1\*</sup>, Duong Thi Thuy Nga<sup>2\*</sup>, Le Trung Chon<sup>3</sup>, Dong Thi Bich Phuong<sup>4</sup>,  
Pham Minh Khan<sup>2</sup>**

<sup>1</sup> Ho Chi Minh City Natural Resources and Environment IT Center; son.ciren@gmail.com

<sup>2</sup> Ho Chi Minh City University of Natural Resources and Environment;

dttnnga\_cntt@hcmunre.edu.vn; pmkhan@hcmunre.edu.vn

<sup>3</sup> HCMC University of Technology; ltchon@hcmut.edu.vn

<sup>4</sup> Institute of Geodesy and Cartography; dbphuong@yahoo.com

**Abstract:** Complicated developments and the rapid spread of Covid–19 taking place in Ho Chi Minh City is a challenge for the city's authorities and people in epidemic prevention and control. To control the disease, the city needs a solution that can forecast and partition epidemic areas to be able to provide the most accurate and fastest solutions. In this research, we propose a solution that is applying local Geographically Weighted Regression (GWR), not global like other regression models such as Ordinary Least Squares (OLS). The advantage of this solution is that it will eliminate errors due to the lack of social factors that the model does not have data to use. Our proposed solution can forecast and evaluate the epidemic pattern in the whole city, the forecast range is high and very sensitive areas, where many and many infections occur. With forecast scenarios, we test in Ho Chi Minh City and the result is rather closed to real data in short term forecasting.

**Keywords:** Covid–19; Data; Epidemic; Regression; Sensitive.