



## ANALYSIS OF URBAN PORT TRAFFIC MOBILITY BEHAVIOR IN HAI PHONG CITY USING GNSS DATA AND THE DBSCAN ALGORITHM

Do Van Manh\*

University of Transport and Communications, No 3 Cau Giay Street, Hanoi, Vietnam

### ARTICLE INFO

TYPE: Research Article

Received: 05/08/2026

Revised: 16/09/2026

Accepted: 07/10/2026

Published online: 15/10/2026

<https://doi.org/10.47869/tcsj.77.8.7>

\* *Corresponding author*

Email: manhdv@utc.edu.vn; Tel: +84 972573673

**Abstract.** Traffic in port cities is shaped by logistics operations, container truck flows and the links between seaports and the urban road network, which produces spatiotemporal mobility patterns unlike those of ordinary cities. This paper examines these patterns in Hai Phong City using GNSS data from an in-vehicle tracking system. After preprocessing, 131,498 vehicle stop points were clustered with the DBSCAN algorithm (50 m radius, minimum of 5 points) for three time windows: 6:30-8:30, 11:00-13:00 and 17:00-19:00. Clustering quality was checked with the Calinski-Harabasz, Davies-Bouldin and Silhouette scores and set against K-means and KDE results. DBSCAN identified 3,377 clusters with a noise ratio of 7.95%. The most stable clusters lie around the port area, container routes and corridors to industrial zones, indicating that logistics activity is tightly bound to the urban traffic structure. Density in the city centre varies strongly with peak hours, whereas the port area remains high in all three windows. The results provide an empirical basis for traffic analysis and for intelligent transport solutions in port cities.

**Keywords:** GNSS, DBSCAN, port city, mobility behavior, ITS, Hai Phong.

@ 2026 University of Transport and Communications



# PHÂN TÍCH ĐẶC TRƯNG HÀNH VI DI CHUYỂN GIAO THÔNG ĐÔ THỊ CẢNG TẠI THÀNH PHỐ HẢI PHÒNG TỪ DỮ LIỆU GNSS SỬ DỤNG THUẬT TOÁN DBSCAN

Đỗ Văn Mạnh\*

Trường Đại học Giao thông vận tải, Số 3 Cầu Giấy, Hà Nội, Việt Nam

## THÔNG TIN BÀI BÁO

CHUYÊN MỤC: Công trình khoa học

Ngày nhận bài: 05/08/2026

Ngày nhận bài sửa: 16/09/2026

Ngày chấp nhận đăng: 07/10/2026

Ngày xuất bản Online: 15/10/2026

<https://doi.org/10.47869/tcsj.77.8.7>

\* Tác giả liên hệ

Email: manhdv@utc.edu.vn; Tel: +84 972573673

**Tóm tắt.** Ở đô thị cảng, vận tải logistics, dòng xe container và liên kết giữa cảng biển với đường đô thị chi phối hoạt động giao thông, tạo nên quy luật không gian - thời gian khác đô thị thông thường. Bài báo phân tích các quy luật này tại Hải Phòng từ dữ liệu GNSS của hệ thống giám sát hành trình. Sau tiền xử lý, 131.498 điểm dừng được phân cụm bằng thuật toán DBSCAN (bán kính 50 m, tối thiểu 5 điểm) theo ba khung giờ 6h30-8h30, 11h-13h và 17h-19h; độ tin cậy của kết quả được kiểm tra qua ba chỉ số Silhouette, Davies-Bouldin, Calinski-Harabasz và đối chiếu với K-means, KDE. Thuật toán xác định 3.377 cụm với tỷ lệ nhiễu 7,95%. Các cụm ổn định nhất nằm quanh khu cảng, tuyến vận tải container và hành lang nổi khu công nghiệp, cho thấy hoạt động logistics gắn chặt với cấu trúc giao thông đô thị; mật độ tại khu trung tâm dao động mạnh theo giờ cao điểm, trong khi khu cảng giữ mức cao ở cả ba khung giờ. Kết quả là cơ sở thực nghiệm cho phân tích giao thông và cho việc lựa chọn giải pháp điều hành tại đô thị cảng.

**Từ khóa:** GNSS, DBSCAN, Đô thị cảng, Hành vi di chuyển, Giao thông thông minh, Hải Phòng.

@ 2026 Trường Đại học Giao thông vận tải

## 1. ĐẶT VẤN ĐỀ

Logistics, vận tải container và các khu công nghiệp mở rộng nhanh, tạo sức ép ngày một lớn cho giao thông ở các đô thị cảng Việt Nam [1, 2]. Tại Hải Phòng, đầu mối cảng biển và logistics chủ lực của miền Bắc, mạng lưới đường phải đồng thời phục vụ đi lại nội đô và vận chuyển hàng hóa liên vùng, nên mật độ phương tiện và mức biến động giao thông khác rõ so với đô thị thông thường.

Thiết bị giám sát hành trình cho chuỗi tọa độ, tốc độ và hướng chạy của từng xe trên phạm vi rộng, nên nguồn số liệu GNSS này ngày càng được dùng trong nghiên cứu điều hành giao thông [3, 4]. So với khảo sát thủ công hay đếm xe, cách thu thập này tự động, phủ rộng và cho phép phân tích gần thời gian thực [5].

Nhiều công trình đã kết hợp GNSS với GIS để phát hiện điểm nóng ùn tắc hoặc lập bản đồ mật độ giao thông [6, 7]. Tuy vậy, các công trình đó mới trực quan hóa ùn tắc hoặc mật độ; đặc trưng hành vi di chuyển của phương tiện ở đô thị cảng, nơi luồng xe container và hành lang logistics có phân bố không gian phức tạp, chưa được phân tích đầy đủ.

Với điểm GNSS rải không đều và lẫn nhiễu, DBSCAN là lựa chọn thích hợp [8, 9] vì phát hiện được các cụm di chuyển mà không phải khai báo trước số cụm.

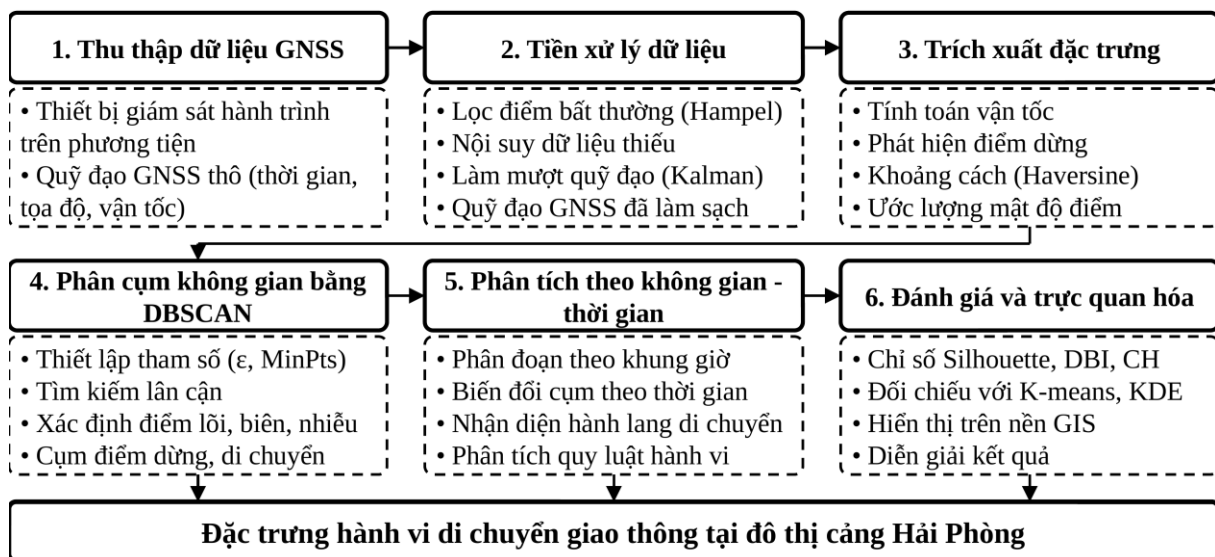
Bài báo vì vậy dùng dữ liệu GNSS và thuật toán DBSCAN để phân tích hành vi di chuyển giao thông đô thị cảng tại Hải Phòng, với ba mục tiêu: nhận diện các cụm di chuyển, xác định hành lang vận tải chủ đạo và làm rõ sự biến đổi hành vi giao thông theo thời gian trong ngày.

## 2. PHƯƠNG PHÁP PHÂN TÍCH HÀNH VI DI CHUYỂN TỪ DỮ LIỆU GNSS

Quy trình gồm bốn bước nối tiếp (Hình 1): thu nhận số liệu, làm sạch, rút đặc trưng di chuyển và phân cụm không gian bằng DBSCAN; các cụm thu được sau đó được xem xét theo vị trí và theo khung giờ.

### 2.1. Dữ liệu nghiên cứu và tiền xử lý GNSS

Nguồn số liệu là hệ thống giám sát hành trình gắn trên xe kinh doanh vận tải chạy tại Hải Phòng, cùng nguồn với nghiên cứu [5]. Số liệu được ghi từ 6h00 đến 22h00 hằng ngày, từ 01/02/2020 đến 26/02/2020 (26 ngày); mỗi phương tiện ứng với một tệp hành trình, chủ yếu là xe tải, xe đầu kéo container và phương tiện logistics quanh khu vực cảng, khu công nghiệp. Mỗi bản ghi gồm mốc thời gian, tọa độ (hệ WGS84), vận tốc tức thời và mã hiệu phương tiện. Số liệu thô được xử lý theo quy trình tác giả đã công bố [10] (khử nhiễu, bù giá trị khuyết, làm trơn quỹ đạo); từ tập quỹ đạo sạch, 131.498 điểm dừng được trích xuất làm đầu vào phân cụm. Điểm dừng là đơn vị phân tích và cũng là thước đo quy mô dữ liệu của bài.



Hình 1. Sơ đồ các bước của nghiên cứu.

## 2.2. Thuật toán phân cụm DBSCAN

DBSCAN gom cụm theo mật độ điểm: thuật toán tự xác định số cụm và tách riêng các điểm nhiễu, nên thích hợp khi điểm GNSS rải không đều.

Tập lân cận của điểm  $p$  và điều kiện để  $p$  là điểm lõi được xác định như sau [11, 12]:

$$N_{\varepsilon}(p) = \{q \in D \mid \text{dist}(p, q) \leq \varepsilon\} \quad (1)$$

$$|N_{\varepsilon}(p)| \geq \text{MinPts} \quad (2)$$

Trong đó:  $D$  - tập điểm GNSS;  $\text{dist}(p, q)$  - khoảng cách từ  $p$  tới  $q$ ;  $\varepsilon$  - bán kính xét lân cận; MinPts - ngưỡng số điểm để tạo cụm.

## 2.3. Đánh giá và trực quan hóa kết quả

Kết quả được kiểm định bằng ba thước đo nội tại: Silhouette, Davies-Bouldin (DBI) và Calinski-Harabasz (CH) [13, 14]. Với điểm  $i$ , hệ số Silhouette là:

$$S(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (3)$$

với  $a(i)$  - trung bình khoảng cách từ  $i$  tới các điểm chung cụm;  $b(i)$  - giá trị nhỏ nhất của trung bình khoảng cách từ  $i$  tới các điểm của từng cụm còn lại.

$S$  trung bình nằm giữa -1 và 1; theo [15],  $S > 0,70$ : cấu trúc cụm mạnh; 0,51-0,70: hợp lý; 0,26-0,50: yếu;  $\leq 0,25$ : chưa có cấu trúc đáng kể. DBI [16] và CH [17] được tính như sau:

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{j \neq i} \left( \frac{\sigma_i + \sigma_j}{d(c_i, c_j)} \right) \quad (4)$$

$$CH = \frac{\text{tr}(B_k)}{\text{tr}(W_k)} \times \frac{n-k}{k-1} \quad (5)$$

Ở đây  $k, n$  - số cụm và số điểm;  $\sigma_i$  - độ phân tán của cụm  $i$  (trung bình khoảng cách tới tâm  $c_i$ );  $d(c_i, c_j)$  - khoảng cách hai tâm;  $\text{tr}(B_k)$ ,  $\text{tr}(W_k)$  - tổng phương sai liên cụm và nội cụm. DBI và CH không có ngưỡng tuyệt đối; chúng chỉ cho biết phương án nào tốt hơn khi cùng áp dụng cho một tập điểm [13].

Kết quả phân cụm được hiển thị trên nền GIS để nhận diện hành lang giao thông và khu vực logistics tập trung.

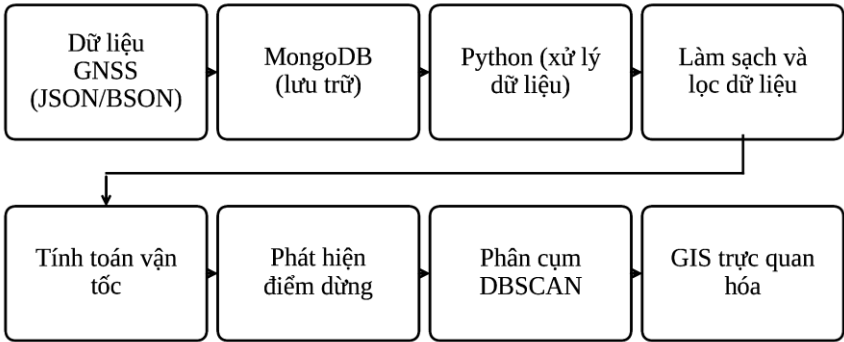
## 2.4. Phương pháp đối chứng: K-means và KDE

K-means [18] phân tập điểm vào  $k$  nhóm sao cho tổng bình phương độ lệch so với tâm nhóm là nhỏ nhất; phải chọn trước  $k$  và mọi điểm đều bị gán nhóm nên nhiễu không được tách. KDE (Kernel Density Estimation) [19, 20] không gán nhãn cụm mà ước lượng mật độ liên tục bằng tổng các hàm nhân đặt tại từng điểm với băng thông  $h$ , cho kết quả dạng bản đồ nhiệt. Cả hai là phương pháp đối chứng cho DBSCAN.

## 3. KẾT QUẢ THỰC NGHIỆM VÀ PHÂN TÍCH HÀNH VI DI CHUYỂN

### 3.1. Chuẩn bị dữ liệu

Trước khi phân tích, dữ liệu được đọc theo từng phần, chuẩn hóa thời gian, kiểm tra tọa độ, loại bản ghi sai, tính vận tốc và phát hiện điểm dừng (Hình 2, Bảng 1).



Hình 2. Luồng xử lý bằng Python và MongoDB.

Tiếp đó, tập điểm sạch được đưa vào GIS và tách lớp theo khung giờ để so sánh cao điểm sáng, giữa ngày và cao điểm chiều.

Bảng 1. Trình tự xử lý trước khi phân cụm.

| Bước xử lý          | Nội dung thực hiện                                          | Mục đích                                |
|---------------------|-------------------------------------------------------------|-----------------------------------------|
| Đọc dữ liệu lớn     | Đọc dữ liệu theo từng phần hoặc theo truy vấn MongoDB       | Tránh quá tải bộ nhớ                    |
| Chuẩn hóa mốc giờ   | Đưa các mốc giờ về một định dạng                            | Phân tích theo hành trình và khung giờ  |
| Lọc tọa độ          | Loại bỏ điểm ngoài vùng nghiên cứu hoặc tọa độ không hợp lệ | Giảm nhiễu không gian                   |
| Lọc vận tốc         | Loại bỏ vận tốc bất thường                                  | Tăng độ tin cậy khi phát hiện điểm dừng |
| Phân tách khung giờ | Tách dữ liệu theo 6h30-8h30, 11h-13h, 17h-19h               | So sánh giữa các khung giờ              |
| Phân cụm DBSCAN     | Phát hiện vùng mật độ cao                                   | Nhận diện khu vực hoạt động tập trung   |
| Hiển thị GIS        | Vẽ các cụm lên bản đồ nền                                   | Diễn giải theo địa bàn thực tế          |

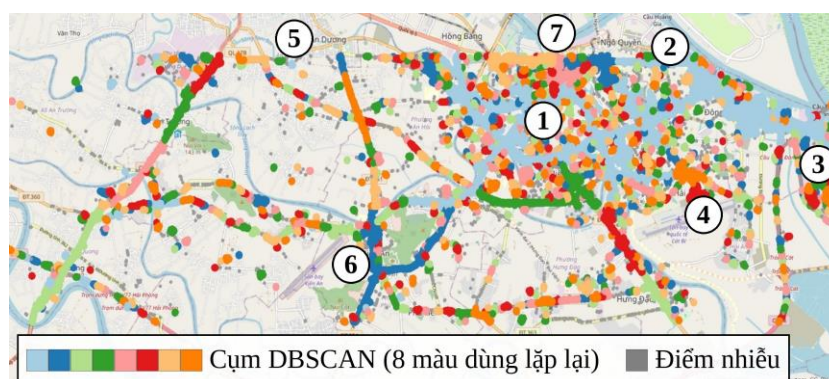
3.2. Kết quả phân tích không gian - thời gian theo khung giờ

Nhằm làm rõ mức độ biến động của hành vi di chuyển trong ngày, dữ liệu nghiên cứu được tách thành ba khung giờ đại diện: 6h30-8h30, 11h-13h và 17h-19h, tương ứng lần lượt với giờ cao điểm sáng, giờ giữa ngày và giờ cao điểm chiều - ba mốc thời gian có ý nghĩa giao thông rõ rệt nhất trong ngày. Kết quả theo từng khung giờ được thể hiện tại Hình 3, Hình 4, Hình 5 và tổng hợp ở Bảng 2. Riêng khung giờ 19h-21h (sau cao điểm chiều) chỉ được dùng bổ sung ở phần đối chiếu bằng KDE (Hình 8) nhằm quan sát sự suy giảm mật độ sau giờ cao điểm, không đưa vào phân tích chính.

• Khung giờ 6h30-8h30:

Kết quả cho khung giờ 6h30-8h30 cho thấy mật độ giao thông tập trung cao nhất tại khu vực trung tâm, bao gồm Ngô Quyền và Lê Chân, đồng thời hình thành một dải mật độ kéo dài dọc trục An Dương - nội đô trên nền GIS. Quy luật này phản ánh dòng xe từ vùng ven dòn vào nội đô đầu ngày, phù hợp với hành vi đi làm, đi học và khởi động các hoạt động đô thị.

Ngoài khu vực trung tâm, mật độ cao còn được ghi nhận tại Sân bay Cát Bi và Cảng Hải Phòng, cho thấy các hoạt động vận tải, dịch vụ hàng không và logistics đã bắt đầu gia tăng ngay từ buổi sáng, dù mức độ tập trung ở khung giờ này chưa đạt đỉnh so với buổi chiều.

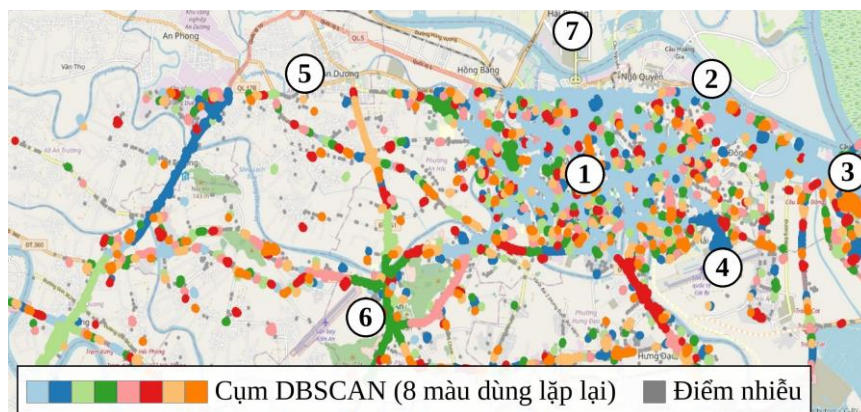


Hình 3. Phân bố cụm mật độ phương tiện trong khung giờ 6h30-8h30 trên nền GIS. Các số chỉ vị trí gần đúng của khu vực: (1) trung tâm Ngõ Quyền - Lê Chân; (2) khu cảng dọc sông Cấm; (3) Đình Vũ; (4) sân bay Cát Bi; (5) An Dương; (6) Kiến An; (7) hướng đi Thủy Nguyên.

- **Khung giờ 11h-13h:**

Sang khung giờ 11h-13h, mật độ giao thông giảm rõ rệt và có xu hướng phân bố đồng đều hơn trên toàn khu vực khảo sát; các điểm nóng còn lại nằm chủ yếu tại Ngõ Quyền - Lê Chân, Sân bay Cát Bi và khu cảng Hải Phòng. Điều đó cho thấy giao thông dân sinh giảm vào giữa ngày, còn các hoạt động đặc thù (chở hàng, logistics, hàng không) vẫn duy trì.

Trái lại, các khu vực ngoại vi như Kiến An và An Dương trở nên thông thoáng rõ rệt, gần như không ghi nhận ùn tắc đáng kể. Điều này khẳng định rằng đặc điểm giao thông theo khung giờ tại Hải Phòng có sự phân hóa rõ nét giữa ba nhóm không gian: khu vực trung tâm, vùng cảng và khu vực ngoại vi.

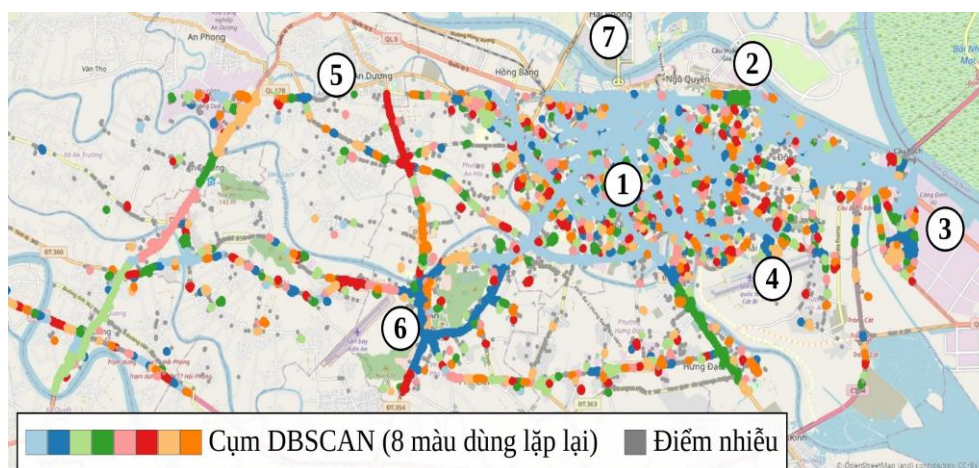


Hình 4. Phân bố cụm mật độ phương tiện trong khung giờ 11h-13h trên nền GIS (ký hiệu khu vực như Hình 3).

- **Khung giờ 17h-19h:**

Đến khung giờ 17h-19h, mật độ giao thông tăng trở lại đáng kể và mở rộng trên toàn mạng lưới nghiên cứu. Các điểm nóng không còn giới hạn ở trung tâm Ngõ Quyền - Lê Chân mà lan tỏa theo nhiều hướng ra An Dương, Kiến An và khu vực qua cầu Thủy Nguyên - dấu hiệu đặc trưng của dòng phương tiện rời nội đô sau giờ làm việc.

Đáng lưu ý, khu vực Cảng Hải Phòng - đặc biệt là vùng Đình Vũ và các tuyến gần với vận tải hàng hóa - trở thành điểm ùn tắc nghiêm trọng nhất trong khung giờ này, nhiều khả năng do sự cộng hưởng giữa dòng xe container, xe tải và giao thông đô thị vào giờ cao điểm chiều. Sân bay Cát Bi cũng duy trì mật độ cao, phản ánh tính chất hỗn hợp của giao thông tại đây giữa nhu cầu dân sinh, dịch vụ hàng không và vận tải. Kết quả tại Đình Vũ phù hợp với tình trạng ùn tắc xe container đã được ghi nhận trong thực tế ở khu vực này (Hình 6).



Hình 5. Phân bố cụm mật độ phương tiện trong khung giờ 17h-19h trên nền GIS (ký hiệu khu vực như Hình 3).

#### • So sánh ba khung giờ:

Nhìn chung qua cả ba khung giờ, khu vực cảng và cảng cá luôn duy trì mật độ giao thông cao ở mức tương đối ổn định. Điều này khác biệt với khu vực trung tâm - nơi mật độ biến động mạnh theo nhịp giờ cao điểm - và cho thấy hoạt động tại khu cảng ít phụ thuộc vào chu kỳ sinh hoạt đô thị hơn, do chịu chi phối chủ yếu bởi vận tải hàng hóa và logistics.



Hình 6. Tắc đường nghiêm trọng xảy ra tại cảng Đình Vũ (ảnh tư liệu minh họa).

Như vậy, ngay trong một đô thị, mỗi khu vực chức năng vận hành theo một nhịp giao thông riêng.

Bảng 2. So sánh đặc điểm mật độ giao thông theo khung giờ.

| Khung giờ | Khu vực nổi bật                                                    | Đặc điểm mật độ                              | Diễn giải hành vi                                                |
|-----------|--------------------------------------------------------------------|----------------------------------------------|------------------------------------------------------------------|
| 6h30-8h30 | Ngô Quyền, Lê Chân, trục An Dương - nội đô, Cát Bi, Cảng Hải Phòng | Mật độ tăng tại trung tâm và trục vào nội đô | Dòng phương tiện vào trung tâm; hoạt động logistics bắt đầu tăng |
| 11h-13h   | Ngô Quyền - Lê Chân, Cát Bi, Cảng Hải Phòng                        | Mật độ giảm và phân bố đều hơn               | Giao thông dân sinh giảm; vận tải và hàng không vẫn duy trì      |
| 17h-19h   | Trung tâm, An Dương, Kiến An, Thủy Nguyên, Đình Vũ, Cát Bi         | Mật độ tăng mạnh và lan rộng                 | Dòng phương tiện rời nội đô; cảng Đình Vũ chịu áp lực lớn        |
| Tổng hợp  | Khu cảng/cảng cá                                                   | Mật độ cao ổn định                           | Phản ánh hoạt động logistics, vận tải và dịch vụ cảng            |

### 3.3. Phân cụm DBSCAN và cách hiển thị trên GIS

Trong thực nghiệm, DBSCAN được cấu hình với bán kính lân cận  $\epsilon = 50$  m và số điểm tối thiểu MinPts = 5, nghĩa là một vùng được xem là có mật độ cao khi trong bán kính 50 m tồn tại ít nhất 5 điểm lân cận. Giá trị  $\epsilon = 50$  m tương ứng với quy mô không gian của một vị trí dừng, chờ diễn hình trong đô thị cảng (nút giao, công cảng, bãi tập kết); giá trị này lớn hơn sai số định vị GNSS thường gặp trong môi trường đô thị do che khuất và đa đường [7, 9] để các điểm dừng tại cùng một vị trí không bị tách cụm bởi nhiễu định vị, đồng thời đủ nhỏ để không gộp các tuyến đường song song lân cận. Giá trị MinPts = 5 dựa trên các giá trị kinh nghiệm cho dữ liệu hai chiều (MinPts = 4 theo [11] và hai lần số chiều theo [12, 21]), được lấy cao hơn một đơn vị do tập dữ liệu lớn và nhiễu nhiễu [12]. Những vùng mật độ cao không liên kết về mặt không gian sẽ tự động được thuật toán tách thành các cụm độc lập.

Áp dụng vào toàn bộ 131.498 điểm dừng, thuật toán xác định được 3.377 cụm và 10.455 điểm nhiễu (Bảng 3). Do số lượng cụm lớn, trên bản đồ GIS (Hình 3, Hình 4, Hình 5) các cụm được tô bằng một bảng 8 màu định tính dùng lặp lại theo mã cụm; vì vậy màu sắc chỉ nhằm phân biệt các cụm nằm gần nhau, hai cụm cùng màu ở cách xa nhau là hai cụm khác nhau, còn các điểm nhiễu được hiển thị bằng màu xám.

### 3.4. Kiểm định chất lượng phân cụm

Xét trên tập điểm dừng đưa vào phân cụm, DBSCAN xử lý tổng cộng 131.498 điểm, trong đó 121.043 điểm được gán vào các cụm và 10.455 điểm còn lại được xác định là nhiễu, tương ứng tỷ lệ nhiễu khoảng 7,95% (Bảng 3). Tỷ lệ này cho thấy phần lớn điểm dừng thu thập được có thể tổ chức thành các cụm không gian mang ý nghĩa giao thông.

Bảng 3. Các chỉ số kiểm định kết quả DBSCAN.

| Chỉ số            | Giá trị | Diễn giải                          | Nhận xét               |
|-------------------|---------|------------------------------------|------------------------|
| Số điểm đầu vào   | 131.498 | Tổng số điểm dừng đưa vào phân cụm | Quy mô dữ liệu đủ lớn  |
| Số điểm thuộc cụm | 121.043 | Số điểm được DBSCAN gán vào cụm    | Chiếm phần lớn dữ liệu |
| Số điểm nhiễu     | 10.455  | Điểm không thuộc cụm nào           | Chấp nhận được         |

| Chỉ số                 | Giá trị  | Diễn giải                                                                    | Nhận xét                                             |
|------------------------|----------|------------------------------------------------------------------------------|------------------------------------------------------|
| Noise Ratio            | 7,95%    | Tỷ lệ nhiễu của DBSCAN                                                       | Hợp lý với dữ liệu GNSS thực tế                      |
| Số cụm phát hiện       | 3.377    | Tổng số cụm DBSCAN                                                           | Phản ánh dữ liệu phân bố chi tiết                    |
| Silhouette             | 0,271    | Mức độ chặt chẽ và tách biệt của cụm; thuộc $[-1; 1]$ , càng gần 1 càng tốt  | Dương, thuộc mức cấu trúc cụm yếu (0,26-0,50)        |
| Davies-Bouldin (DBI)   | 0,552    | Độ giống nhau giữa các cụm; không âm, càng nhỏ càng tốt                      | Thấp hơn K-means (Bảng 4), các cụm tách biệt khá tốt |
| Calinski-Harabasz (CH) | 2.116,24 | Mức độ phân tách giữa cụm và trong cụm; không có cận trên, càng lớn càng tốt | Chỉ có ý nghĩa so sánh tương đối (Bảng 4)            |

Ghi chú: Các chỉ số Silhouette, Davies-Bouldin và Calinski-Harabasz được tính trên mẫu đánh giá gồm 3.000 điểm không nhiễu được lấy ngẫu nhiên từ tập điểm dừng sau DBSCAN. Sau khi loại các nhãn cụm xuất hiện đơn lẻ trong mẫu, còn 2.486 điểm thuộc 282 cụm được sử dụng để đánh giá.

Với giá trị 0,271, chỉ số Silhouette tuy không cao nhưng vẫn dương và thuộc khoảng cấu trúc cụm yếu (0,26-0,50) theo thang diễn giải ở mục 2.3, cho thấy các cụm có tách nhau ở mức nhất định. Với dữ liệu giao thông đô thị, mức này chấp nhận được, do các điểm GNSS thường trải dọc theo mạng lưới đường thay vì hình thành các cụm tròn, tách biệt hoàn toàn như trong các bộ dữ liệu lý tưởng.

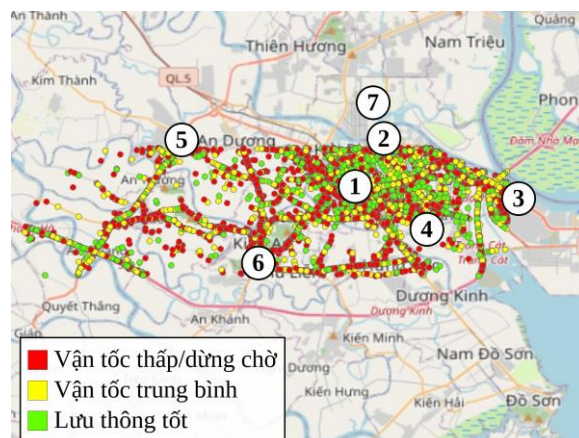
Chỉ số Davies-Bouldin 0,552 cho thấy các cụm ít chồng lấn, tức khả năng phân tách tương đối tốt; trong khi đó, chỉ số Calinski-Harabasz đạt 2.116,24 phản ánh cấu trúc cụm rõ ràng trong dữ liệu. Kết hợp các chỉ số này với kết quả trực quan trên nền GIS - nhất là các cụm ổn định quanh khu vực cảng, cảng cá và các trục giao thông chính - có cơ sở để xem DBSCAN là phương pháp phù hợp cho bài toán phát hiện vùng tập trung điểm dừng từ dữ liệu GNSS tại Hải Phòng.

Bảng 4. So sánh nhanh DBSCAN và K-means trên mẫu điểm dừng.

| Phương pháp | Số cụm | Silhouette | Davies-Bouldin | Calinski-Harabasz | Nhận xét                                          |
|-------------|--------|------------|----------------|-------------------|---------------------------------------------------|
| DBSCAN      | 3.377  | 0,271      | 0,552          | 2.116,24          | Phát hiện cụm chi tiết, xử lý nhiễu tốt           |
| K-means     | 8      | 0,378      | 0,851          | 4.159,97          | Chỉ số Silhouette cao hơn nhưng không xử lý nhiễu |
| K-means     | 20     | 0,458      | 0,665          | 5.792,22          | Cụm hình học rõ hơn nhưng cần chọn trước số cụm   |

Mặc dù chỉ số Silhouette và Calinski-Harabasz của K-means cao hơn trên cùng mẫu dữ liệu (Bảng 4), điều đó không đồng nghĩa K-means là lựa chọn phù hợp hơn cho bài toán đang xét. Do có xu hướng tạo ra các cụm hình học đều và buộc mọi điểm phải được gán vào một cụm nào đó, K-means tỏ ra kém tương thích với dữ liệu GNSS thực tế - vốn chứa nhiều điểm nhiễu và các cụm kéo dài theo hình dạng tuyến đường. Vì vậy, DBSCAN vẫn là lựa chọn phù hợp hơn cho mục tiêu phát hiện vùng điểm dừng tập trung đồng thời loại bỏ nhiễu trong dữ liệu

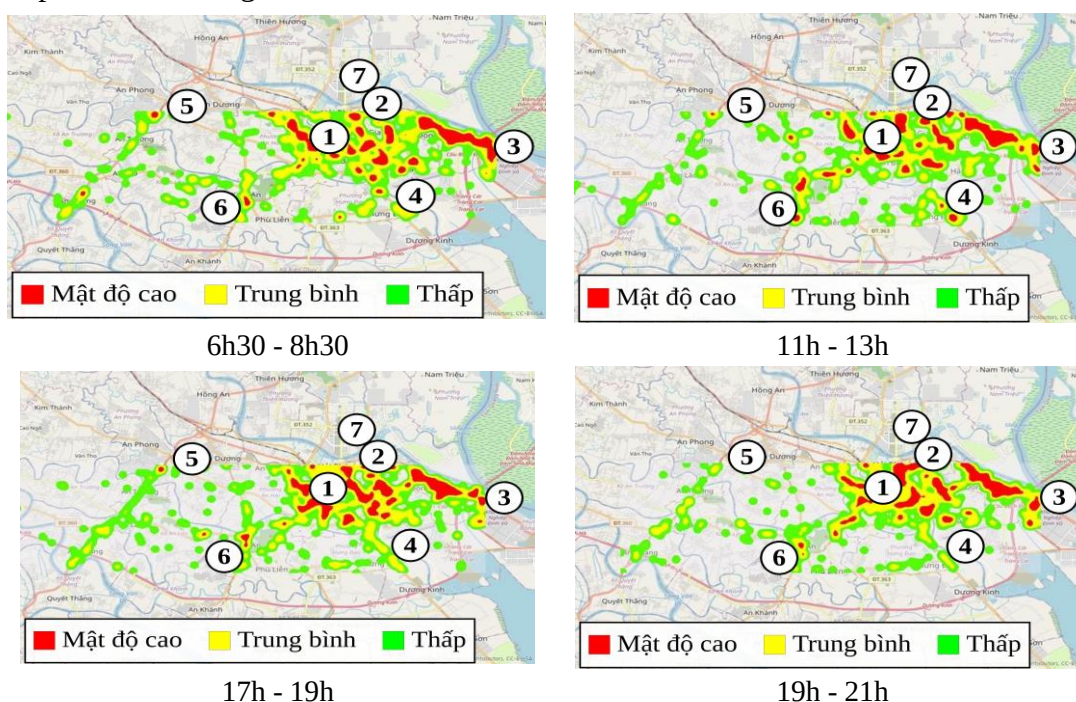
GNSS đô thị. Hình 7 minh họa kết quả K-means theo ba nhóm trạng thái giao thông (vận tốc thấp/dừng chờ, vận tốc trung bình và lưu thông tốt) trên nền GIS.



Hình 7. Phân cụm và hiển thị bằng K-Means và GIS (ký hiệu khu vực như Hình 3).

Bên cạnh K-means, nghiên cứu tiếp tục đối chiếu DBSCAN với phương pháp KDE. Các bản đồ nhiệt theo từng khung giờ (Hình 8) cho thấy mật độ phương tiện dồn về trung tâm Hải Phòng, khu cảng và các trục chính, đồng thời cho phép quan sát trực quan xu hướng lan truyền mật độ cũng như sự biến đổi mật độ theo thời gian trong ngày.

Ngược lại, DBSCAN thiên về phát hiện từng cụm điểm dừng cụ thể, kể cả cụm kéo dài theo tuyến đường, đồng thời bỏ được điểm lẻ, nên hợp hơn cho bài toán suy luận hành vi di chuyển. Như vậy, KDE phản ánh tốt xu thế phân bố giao thông ở quy mô tổng thể, còn DBSCAN cho phép xác định cụ thể các vùng hoạt động và điểm dừng có ý nghĩa giao thông; dùng kết hợp hai phương pháp cho cái nhìn đầy đủ hơn về không gian - thời gian ở một đô thị phức tạp như Hải Phòng.



Hình 8. Bản đồ nhiệt mật độ giao thông theo khung giờ bằng phương pháp KDE tại Hải Phòng (ký hiệu khu vực như Hình 3).

### 3.5. So sánh kết quả theo chức năng đô thị

Dựa trên kết quả hiển thị GIS, các cụm mật độ cao tại Hải Phòng có thể được quy về một số nhóm chức năng không gian khác nhau (Bảng 5). Khu vực trung tâm, tiêu biểu là Ngô Quyền - Lê Chân, phản ánh giao thông dân sinh gắn với nhu cầu đi lại thường nhật. Cụm tại Cảng Hải Phòng, Đình Vũ và khu vực cảng cá thể hiện rõ hoạt động vận tải, logistics và dịch vụ cảng. Khu vực Sân bay Cát Bi mang đặc thù giao thông hỗn hợp giữa hành khách, dịch vụ và phương tiện kết nối. Trong khi đó, các trục An Dương, Kiến An và Thủy Nguyên phản ánh dòng phương tiện ra/vào nội đô theo nhịp sinh hoạt và làm việc.

Đối chiếu với thực tế, vị trí các cụm ổn định trùng với những đầu mối giao thông hiện hữu trên bản đồ nền (Hình 3, Hình 4, Hình 5), gồm khu cảng dọc sông Cấm, khu vực Đình Vũ, sân bay Cát Bi cùng các cửa ngõ An Dương, Kiến An; riêng tại Đình Vũ, kết quả phù hợp với tình trạng ùn tắc xe container trong thực tế (Hình 6). Việc đối chứng mới dừng ở mức định tính; đối chứng định lượng bằng số liệu đếm xe hoặc camera giao thông sẽ được thực hiện trong nghiên cứu tiếp theo.

Bảng 5. Diễn giải các vùng cụm theo chức năng đô thị.

| Nhóm khu vực                  | Biểu hiện trên GIS          | Đặc điểm theo khung giờ           | Ý nghĩa hành vi suy luận                             |
|-------------------------------|-----------------------------|-----------------------------------|------------------------------------------------------|
| Trung tâm Ngô Quyền - Lê Chân | Mật độ cao, biến động mạnh  | Cao sáng và chiều, giảm giữa ngày | Giao thông đi làm, đi học, dừng chờ và ùn tắc đô thị |
| Trục An Dương - nội đô        | Dải mật độ vào trung tâm    | Rõ trong 6h30-8h30                | Dòng phương tiện vào nội đô đầu ngày                 |
| Đình Vũ - Cảng Hải Phòng      | Cụm mạnh, ổn định           | Cao cả sáng, trưa và chiều        | Logistics, container, vận tải hàng hóa               |
| Cảng cá/khu cảng              | Cụm đồng bộ nhiều khung giờ | Duy trì mật độ cao                | Hoạt động dịch vụ cảng, tập trung phương tiện        |
| Sân bay Cát Bi                | Mật độ cao cục bộ           | Duy trì nhiều khung giờ           | Giao thông hỗn hợp, dịch vụ sân bay                  |
| Kiến An, Thủy Nguyên          | Mở rộng rõ vào chiều        | Cao trong 17h-19h                 | Dòng rời nội đô sau giờ làm                          |

## 4. KẾT LUẬN

Nghiên cứu đã làm rõ cách phương tiện di chuyển ở đô thị cảng Hải Phòng dựa trên số liệu hành trình GNSS và thuật toán DBSCAN. Trên 131.498 điểm dừng, thuật toán xác định 3.377 cụm với tỷ lệ nhiễu 7,95%; các cụm rõ nhất nằm tại khu cảng biển, hành lang vận tải container và các tuyến nối khu công nghiệp.

Hành vi di chuyển khác nhau rõ giữa các khu vực và các khung giờ: mật độ ở trung tâm dao động theo giờ cao điểm, còn khu cảng giữ mức cao trong cả ba khung giờ, cho thấy logistics và hoạt động chở hàng chi phối mạnh cấu trúc giao thông của thành phố.

Vẫn còn những giới hạn: số liệu giám sát hành trình có sai số vị trí, có khoảng trống và phân bố phương tiện không đều giữa các khu vực; phân tích mới dựa trên GNSS, chưa tích hợp camera giao thông, thời tiết hay tín hiệu đèn. Bài báo cũng chưa tách ngày làm việc và ngày cuối tuần, chưa phân tích các khung giờ thấp điểm, và việc đối chứng với thực tế mới ở mức định tính.

Hướng tiếp theo: tích hợp đa nguồn dữ liệu; dùng học máy, học sâu trên chuỗi GNSS để dự đoán luồng di chuyển và nhận biết bất thường; mở rộng phân tích theo loại ngày, theo khung giờ thấp điểm; kiểm chứng định lượng bằng số liệu quan trắc hiện trường.

## TÀI LIỆU THAM KHẢO

- [1]. P. N. Duy, L. Chapman, M. Tight, Resilient transport systems to reduce urban vulnerability to floods in emerging-coastal cities: A case study of Ho Chi Minh City, Vietnam, *Travel Behaviour and Society*, 15 (2019) 28-43. <https://doi.org/10.1016/j.tbs.2018.11.001>
- [2]. H. Nguyen, T. T. Phan, D. T. Binh, Unlocking logistics efficiency in transitional economies: the impact of economic, social and institutional factors on provincial logistics costs in Vietnam, *Economic Change and Restructuring*, 59 (2026) 44. <https://doi.org/10.1007/s10644-026-09995-3>
- [3]. H. Jing, Y. Gao, S. Shahbeigi, M. Dianati, Integrity monitoring of GNSS/INS based positioning systems for autonomous vehicles: State-of-the-art and open challenges, *IEEE Transactions on Intelligent Transportation Systems*, 23 (2022) 14166-14187. <https://doi.org/10.1109/TITS.2022.3149373>
- [4]. Đỗ Văn Mạnh, Đinh Tuấn Hải, Development of a Sustainable National Traffic Information Notification System: A GNSS-Based with Enhanced-LSTM for Urban Road Traffic Speed Forecasting, *International Journal of Intelligent Transportation Systems Research*, 23 (2025) 489-502. <https://doi.org/10.1007/s13177-025-00463-2>
- [5]. Đỗ Văn Mạnh, Trần Quang Học, Lê Khánh Giang, Vương Xuân Cẩn, Vũ Văn Trường, Enhanced Deep Neural Networks for Traffic Speed Forecasting Regarding Sustainable Traffic Management Using Probe Data from Registered Transport Vehicles on Multilane Roads, *Sustainability*, 16 (2024) 2453. <https://doi.org/10.3390/su16062453>
- [6]. B. Konakoglu, S. S. Kurt Konakoglu, K. T. Celik, T. Ustun Topal, A Bibliometric Analysis of GPS/GNSS Contributions to GIS, *Journal of Geovisualization and Spatial Analysis*, 10 (2026) 9. <https://doi.org/10.1007/s41651-026-00252-9>
- [7]. Y. Lee, P. Wang, B. Park, Nonlinear regression-based GNSS multipath dynamic map construction and its application in deep urban areas, *IEEE Transactions on Intelligent Transportation Systems*, 24 (2023) 5082-5093. <https://doi.org/10.1109/TITS.2023.3246493>
- [8]. T. Z. Abdulhameed, S. A. Yousif, V. W. Samawi, H. I. Al-Shaikhli, SS-DBSCAN: semi-supervised density-based spatial clustering of applications with noise for meaningful clustering in diverse density data, *IEEE Access*, 12 (2024) 131507-131520. <https://doi.org/10.1109/ACCESS.2024.3457587>
- [9]. H. J. Kim, Y. H. Kim, J. H. Lee, S. J. Park, B. S. Ko, J. W. Song, Improving the accuracy of vehicle position in an urban environment using the outlier mitigation algorithm based on GNSS multi-position clustering, *Remote Sensing*, 15 (2023) 3791. <https://doi.org/10.3390/rs15153791>
- [10]. Đỗ Văn Mạnh, Tiền xử lý nâng cao dữ liệu GNSS cho phân tích giao thông bằng Hampel filter, mô hình nội suy kết hợp và Kalman filter, *Tạp chí Khoa học Giao thông vận tải*, 77 (2026) 328-341. <https://doi.org/10.47869/tcsj.77.3.8>
- [11]. M. Ester, H.-P. Kriegel, J. Sander, X. Xu, A density-based algorithm for discovering clusters in large spatial databases with noise, in: *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96)*, 1996, AAAI Press, Portland, pp. 226-231.

- [12]. E. Schubert, J. Sander, M. Ester, H.-P. Kriegel, X. Xu, DBSCAN revisited, revisited: why and how you should (still) use DBSCAN, *ACM Transactions on Database Systems*, 42 (2017) 1-21. <https://doi.org/10.1145/3068335>
- [13]. O. Arbelaitz, I. Gurrutxaga, J. Muguerza, J. M. Pérez, I. Perona, An extensive comparative study of cluster validity indices, *Pattern Recognition*, 46 (2013) 243-256. <https://doi.org/10.1016/j.patcog.2012.07.021>
- [14]. P. J. Rousseeuw, Silhouettes: A graphical aid to the interpretation and validation of cluster analysis, *Journal of Computational and Applied Mathematics*, 20 (1987) 53-65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)
- [15]. L. Kaufman, P. J. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*, John Wiley & Sons, New York, 1990. <https://doi.org/10.1002/9780470316801>
- [16]. D. L. Davies, D. W. Bouldin, A cluster separation measure, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1 (1979) 224-227. <https://doi.org/10.1109/TPAMI.1979.4766909>
- [17]. T. Caliński, J. Harabasz, A dendrite method for cluster analysis, *Communications in Statistics*, 3 (1974) 1-27. <https://doi.org/10.1080/03610927408827101>
- [18]. S. Lloyd, Least squares quantization in PCM, *IEEE Transactions on Information Theory*, 28 (1982) 129-137. <https://doi.org/10.1109/TIT.1982.1056489>
- [19]. B. W. Silverman, *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London, 1986.
- [20]. Z. Xie, J. Yan, Kernel Density Estimation of traffic accidents in a network space, *Computers, Environment and Urban Systems*, 32 (2008) 396-406. <https://doi.org/10.1016/j.compenvurbsys.2008.05.001>
- [21]. J. Sander, M. Ester, H.-P. Kriegel, X. Xu, Density-based clustering in spatial databases: The algorithm GDBSCAN and its applications, *Data Mining and Knowledge Discovery*, 2 (1998) 169-194. <https://doi.org/10.1023/A:1009745219419>