

# PHƯƠNG PHÁP LOẠI BỎ NHIỀU DỮ LIỆU ĐÁM MÂY ĐIỂM LIDAR SỬ DỤNG KHOẢNG CÁCH DANH NGHĨA (NOMINAL POINT SPACING)

Nguyễn Thị Hữu Phương<sup>1</sup>, Đặng Văn Đức<sup>2</sup>, Nguyễn Trường Xuân<sup>1</sup>

<sup>1</sup> Khoa Công nghệ thông tin, Trường Đại học Mở - Địa chất

<sup>2</sup> Viện Công nghệ thông tin, Viện Hàn lâm Khoa học và Công nghệ Việt Nam

nguyenthihuuphuong@humg.edu.vn, dvduc@ioit.ac.vn, nguyentruongxuan@humg.edu.vn

**TÓM TẮT**— Với khả năng thu thập được dữ liệu trên không gian rộng lớn trong khoảng thời gian ngắn và không phụ thuộc thời tiết công nghệ LiDAR ngày càng được sử dụng rộng rãi trong lĩnh vực Khoa học Trái đất. Dữ liệu thu được từ dữ liệu LiDAR thường có độ chính xác cao, mỗi điểm được đặc trưng bởi tọa độ, độ cao và độ chính xác của nó. Tuy nhiên, với số lượng lớn điểm được thu thập trong mỗi lần đo, nhiều trong dữ liệu đám mây điểm LiDAR ảnh hưởng nhiều đến kết quả xử lý cho các ứng dụng sau này. Do đó, cần phải loại bỏ nhiều trước khi thực hiện các bước xử lý dữ liệu tiếp theo với dữ liệu LiDAR. Trong bài báo này, các tác giả đề xuất phương pháp loại bỏ nhiều dữ liệu LiDAR sử dụng khoảng cách danh nghĩa (NPS) trong quá trình tiền xử lý. Phương pháp đã được thử nghiệm với đám mây điểm LiDAR được thu nhận tại Bắc Ninh cho độ chính xác 93.6%.

**Từ khóa**— loại bỏ nhiều, đám mây điểm LiDAR, khoảng cách danh nghĩa, LiDAR, NPS.

## I. GIỚI THIỆU

Nhiều trong dữ liệu là những dữ liệu bị ngắt, bị hỏng hoặc những dữ liệu bị thêm vào không mong muốn. Nhiều thường là những dữ liệu không có giá trị và ảnh hưởng trực tiếp đến độ chính xác đến kết quả của bất kỳ phân tích dữ liệu nào và làm sai kết quả dự đoán của bất kỳ thông tin có ý nghĩa nào nếu không được xử lý đúng cách [1]. Đã có nhiều nghiên cứu thực nghiệm chỉ ra rằng nhiều trong tập dữ liệu ảnh hưởng đáng kể dẫn đến làm giảm độ chính xác của kết quả của bài toán phân loại và kết quả dự đoán, dự báo kém [2]. Do đó, xác định và loại bỏ nhiều trong các bài toán của khai phá dữ liệu là cần thiết và phải được tiến hành trước khi thực hiện các bước xử lý tiếp theo. Nếu việc loại bỏ nhiều được thực hiện với độ chính xác sẽ giúp tiết kiệm thời gian và nâng cao độ chính xác của quá trình xử lý dữ liệu.

Với khả năng thu nhận dữ liệu trên một khu vực rộng lớn, bộ dữ liệu lớn có thể lên đến hàng trăm triệu điểm, LiDAR thu thập thông tin về đối tượng bằng cách phát tia laser tới mục tiêu và thu nhận tia phản xạ. LiDAR thực hiện quét bề mặt Trái đất với mật độ điểm dày đặc, nhiều trong quá trình thu thập dữ liệu là không thể tránh khỏi. Nhiều trước hết thường do độ nhạy cao của máy quét khi quét các địa hình địa vật, các đám mây điểm hiển thị rất nhiều “điểm trong không khí”. Thứ hai, phạm vi tự nó có xu hướng bị nhiễu phạm vi cao hơn đáng kể so với dạng sóng tuyến tính LiDAR [3]. Nhiều của dữ liệu LiDAR thường là Gaussian, ngoại biên và ngoại lệ [4].

- Nhiễu Gaussian thường được tạo ra do sai số của bộ quét laser trong quá trình thu nhận về đối tượng. Nhiễu này thường ảnh hưởng đến tất cả các điểm trong đám mây điểm và có thể được loại bỏ trong quá trình xuất dữ liệu và được mô hình hóa bởi độ lệch chuẩn.

- Nhiễu ngoại biên thường được gây ra do máy quét không xác định được biên của khu vực quét. Sai số này có thể loại bỏ bằng ước lượng mật độ điểm dựa trên độ cao,

- Nhiễu ngoại lệ thường được tạo ra trong quá trình thu nhận đối tượng khi có những điểm được thêm vào như các hạt trong không khí, các đối tượng không mong muốn.

Với mỗi loại nhiễu sẽ có phương pháp loại bỏ nhiễu và đã được đưa ra trong các nghiên cứu của nhiều nhà khoa học trên thế giới. Thông thường, phương pháp loại bỏ nhiễu LiDAR sẽ được thực hiện theo hai hướng sau [5]:

- Khử nhiễu miền tần số: trước hết chuyển đổi tín hiệu vào miền tần số trước khi loại bỏ nhiễu và biến đổi tín hiệu trở lại miền không gian bằng cách nghịch đảo biến đổi sau khi khử nhiễu, chẳng hạn như biến đổi Fourier và biến đổi wavelet.

- Khử nhiễu miền không gian: đây là phương pháp áp dụng trực tiếp phép biến đổi không gian cho dữ liệu. Thường dựa trên phương pháp lân cận, có nghĩa là giá trị của một điểm lấy mẫu trong các dữ liệu đầu ra được tính toán bởi một thuật toán nhất định bằng cách sử dụng giá trị của điểm lân cận từ các tín hiệu đầu vào. Các thuật toán có thể là lọc trung bình, lọc Gaussian, v.v.

Nhiều trong dữ liệu LiDAR từ nhiều nguồn khác nhau và có sự phân bố phức tạp. Để lựa chọn phương pháp loại bỏ nhiễu phù hợp và áp dụng được cho mọi bộ dữ liệu là rất khó thực hiện. Trong bài báo này, nhóm tác giả tập trung nghiên cứu phương pháp khử nhiễu miền không gian. Khử nhiễu miền không gian là phương pháp có thể khử được nhiều loại nhiễu như nhiễu ngoại biên, nhiễu ngoại lệ.

Trong nghiên cứu [5], tác giả đã cải tiến thuật toán  $\lambda/\mu$  của tác giả Taubin dựa trên phương pháp giá trị trung bình để loại bỏ nhiễu dữ liệu LiDAR. Kết quả thử nghiệm đã chỉ ra rằng quá trình loại bỏ nhiễu giúp tăng độ chính xác

của bài toán ứng dụng. Các tác giả trong nghiên cứu [6] đã loại bỏ nhiễu dựa trên các phương pháp thống kê gợi ý chỉ số trước khi xử lý dữ liệu và so sánh với dữ liệu chưa loại nhiễu để đánh giá độ chính xác.

Trong khi đó, nghiên cứu [7] đã đưa ra phương pháp loại bỏ biên sử dụng sự sai khác đột ngột của độ cao của điểm so với điểm lân cận. Sử dụng cách so sánh này, nhóm tác giả loại những điểm cao hay thấp đột biến. Bài báo [8] đánh giá một phương pháp miền thời gian để làm mịn và giảm mức độ nhiễu trong dữ liệu đó. Phương pháp Savitzky-Golay (S-G) xấp xỉ và làm mịn dữ liệu bằng cách tận dụng lợi thế của việc điều chỉnh một đa thức bậc d, sử dụng bình phương nhỏ nhất cục bộ. Là kết quả của việc tích hợp phương pháp này với Phân tích giá trị số ít (SVD) và áp dụng bộ lọc này trên các vectơ riêng của SVD, có thể thu được các kết quả làm giảm hiệu quả thỏa đáng. Kết quả của phương pháp tiếp cận S-G dựa trên SVD này đã được đánh giá bằng cách sử dụng hai bộ dữ liệu LiDAR khác nhau và cũng được so sánh với các bộ dữ liệu khác các phương pháp phổ biến về mức độ bảo toàn các mômen của tín hiệu và độ gần của tín hiệu nhiễu. Kết quả chỉ ra rằng phương pháp S-G dựa trên SVD có hiệu suất vượt trội trong việc khử dữ liệu LiDAR.

Thông qua điểm lưới không gian 3D, dữ liệu đám mây bên trong mỗi các điểm lưới khối có thuộc tính của chỉ số không gian, theo không gian mối quan hệ lân cận giữa các phán đoán lưới 3D trong điểm lưới dưới dạng nhiễu điểm là phương pháp được đề xuất trong [9]. Kết quả thử nghiệm cho thấy thuật toán này có thể lọc địa hình một cách hiệu quả đám mây điểm gồm các điểm nhiễu rời rạc và các cụm nhiễu nhỏ, bằng cách sử dụng hệ số tương quan của khoảng cách trung bình giữa các cạnh của lưới 3D có thể giảm đáng kể sai số xác định nhiễu. Với nghiên cứu [10] các tác giả sử dụng một phương pháp lọc đám mây điểm 3D cho lá dựa trên khoảng cách đa tạp và ước lượng bình thường được đề xuất. Đầu tiên, lá được chiết từ ngọn cây đám mây và phân cụm ban đầu được thực hiện như bước tiền xử lý. Thứ hai, các cụm ngoại lệ lọc và lọc điểm ngoại lệ đã được thực hiện liên tiếp bằng cách sử dụng khoảng cách đa tạp và phương pháp cắt ngắn. Thứ ba, các điểm nhiễu trong mỗi cụm được lọc dựa trên bề mặt cục bộ ước lượng bình thường. Kết quả tái tạo 3D của lá sau khi áp dụng bộ lọc được đề xuất phương pháp chứng minh rằng phương pháp này tốt hơn các phương pháp lọc cổ điển khác.

Với những nghiên cứu đã được đề cập, có thể thấy phương pháp loại nhiễu dựa trên đặc trưng điểm hay điểm lân cận được sử dụng nhiều và cho độ chính xác. Chính vì thế, nhóm tác giả đã sử dụng phương pháp loại bỏ nhiễu miền không gian dựa trên khái niệm điểm lân cận cho nghiên cứu của mình. Bài báo sử dụng khái niệm NPS – đặc trưng của điểm dữ liệu LiDAR kết hợp với thuật toán kNN để loại bỏ nhiễu cho đám mây điểm trước khi thực hiện phân lớp thành lập DEM.

## II. KHOẢNG CÁCH DANH NGHĨA TRONG DỮ LIỆU LIDAR

Trong công nghệ LiDAR, NPS được nêu trong danh mục của USGS là thước đo phổ biến về mật độ của tập dữ liệu LiDAR. NPS là khoảng cách giữa các điểm trong bộ dữ liệu thu thập được trong quá trình quét, thường được biểu thị bằng mét và được tính đơn giản nhất là căn bậc hai của diện tích trung bình trên mỗi điểm trả về đầu tiên. Đây là khoảng cách chỉ được sử dụng trong công nghệ LiDAR. NPS là một giá trị thường được dự đoán bởi phần mềm lập kế hoạch chuyến bay của nhà sản xuất hệ thống để chỉ ra khoảng cách lưới của các điểm được thu thập dựa trên các tham số kế hoạch chuyến bay đầu vào. NPS cũng có thể được ước tính bằng cách xác định căn bậc hai nghịch đảo của mật độ tia phản xạ đầu tiên trả về, không có điểm trung bình, trên một  $\text{km}^2$  hoặc diện tích lớn hơn [11].

NPS được thu nhận như sau [12]:

- Tạo bản sao của bộ dữ liệu
- Lọc các tín hiệu phản xạ
- Lấy thông tin về các tín hiệu phản xạ đầu tiên
- Tạo một tập hợp đa giác theo đường bay, tính toán diện tích của các đa giác này
- Tính tổng những điểm phản xạ đầu tiên trên mỗi ô đa giác chia cho diện tích của đa giác sẽ ra được mật độ điểm

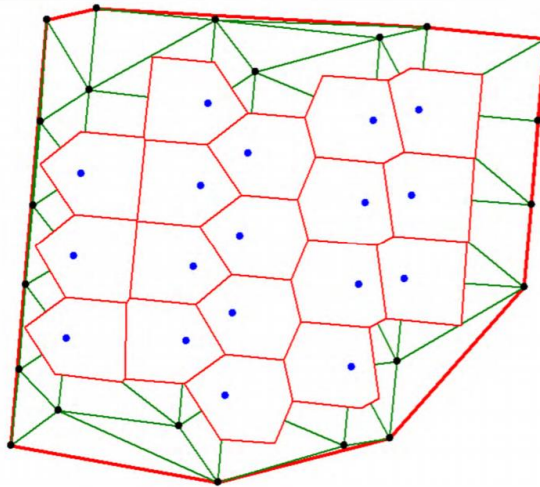
Giả sử một đám mây điểm phân bố đều đặn, khoảng cách điểm là căn bậc hai của diện tích trung bình trên một điểm (diện tích của đa giác chia cho số điểm mà nó chứa). Theo ngữ cảnh của mật độ, NPS có thể được tính theo công thức:

$$NPS = \frac{1}{\sqrt{NPD}}$$

Trong đó, NPD (Nominal Point Denstiy) là mật độ điểm danh nghĩa. Mật độ điểm của đám mây điểm LiDAR đặc trưng cho số điểm trong một khu vực nhất định. Thường mật độ điểm được tính toán cho mỗi  $\text{m}^2$  và sử dụng đơn vị  $\text{pt}/\text{m}^2$ .

NPS có sự liên quan mật thiết đến mật độ điểm, các nhóm điểm càng gần nhau thì mật độ điểm càng cao. Tuy nhiên, khoảng cách giữa các điểm trong đám mây điểm thực tế có thể có sự khác biệt, do đó trong [3] tác giả đã đề xuất khái niệm NPS lấy giá trị cạnh lớn nhất thay vì trung bình. Điều này sẽ đảm bảo rằng các khoảng trống trong việc lấy

mẫu được tính toán và số liệu không bị giảm do các phép đo dày đặc hoặc chùng chéo trong một dòng quét có khoảng cách rộng. Một yếu tố khác ảnh hưởng đến sự phân bố điểm là kiểu quét được tạo ra bởi các cảm biến khác nhau. Theo USGS khi tính toán theo các vùng dạng đa giác hay tam giác sẽ dễ dàng tính toán được sự phân bố này.

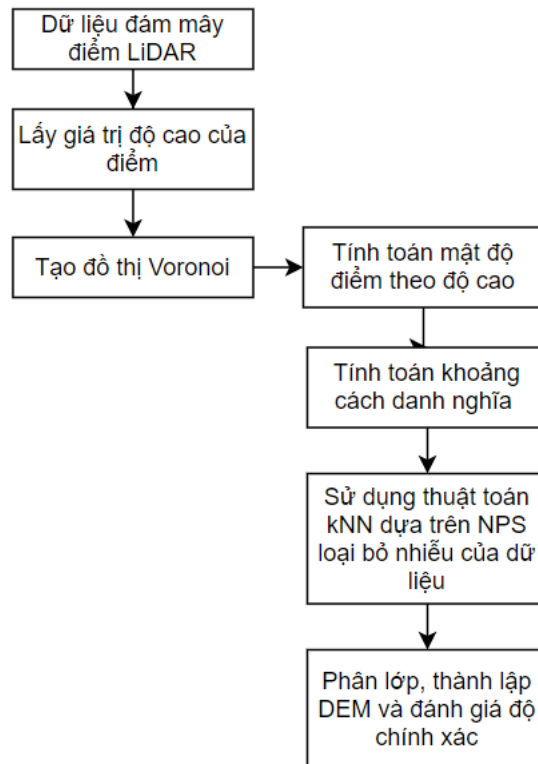


**Hình 1.** Giá trị NPS được tính toán dựa trên đồ thị Voronoi

Yêu cầu về mật độ điểm và khoảng cách danh nghĩa của dữ liệu sẽ là yếu tố chính giúp tính toán các tham số cho LiDAR. Mặc dù có thể có nhiều cân nhắc khác nhau để chọn một mật độ điểm và khoảng cách nhất định, nhưng mức độ chi tiết cần hiển thị trong tập dữ liệu sẽ ít nhất giới hạn mật độ điểm và khoảng cách tối thiểu. Mật độ điểm cao hơn sẽ có khoảng cách điểm thấp hơn và do đó tiết lộ nhiều đặc trưng hơn trong đám mây điểm so với các loại dữ liệu khác [13]. Dựa trên giá trị NPS ta có thể nhận biết nhiều một cách đáng kể trong dữ liệu LiDAR.

### III. PHƯƠNG PHÁP LOẠI BỎ NHIỀU DỮ LIỆU ĐÁM MÂY ĐIỂM LIDAR DỰA TRÊN GIÁ TRỊ NPS

Phương pháp loại bỏ nhiều dữ liệu đám mây điểm LiDAR dựa trên giá trị NPS được nhóm tác giả đề xuất như trong hình 2.



**Hình 2.** Phương pháp đề xuất

Chi tiết quy trình như sau:

- Đám mây điểm LiDAR được thu nhận từ quá trình sẽ được xuất ra file dưới định dạng .las
- Trích xuất giá trị độ cao (Z) của điểm, giá trị độ cao của điểm trong đám mây điểm LiDAR là giá trị đặc trưng của dữ liệu LiDAR. Dựa trên giá trị độ cao ta có thể thực hiện các bài toán phân lớp, xây dựng các mô hình, tạo

mô phỏng, .... Trong bài báo này, nhóm tác giả sử dụng giá trị độ cao của điểm LiDAR để thực hiện tính toán phân bố theo độ cao để từ đó tính được khoảng cách danh nghĩa giữa các điểm theo phân bố độ cao. Theo đó, bước trích xuất giá trị độ cao của điểm từ file .las sẽ được thực hiện trước các bước tính toán tiếp theo.

c) Tạo đồ thị Voronoi:

Đồ thị Voronoi là khái niệm cơ bản chỉ cấu trúc dữ liệu được nghiên cứu trong hình học. Một đồ thị Voronoi được định nghĩa như một đồ thị tối thiểu của một tập hữu hạn các hàm liên tục [14]. Một đồ thị Voronoi thường chia tập dữ liệu thành các vùng, mỗi vùng bao gồm các điểm gần với điểm nhất định hơn so với các điểm khác, và như vậy ta có thể nói mỗi điểm trong đám mây điểm sẽ có khoảng cách nhất định với một điểm nào đó [15]. Do đó, sử dụng đồ thị Voronoi rất dễ để tìm ra vùng đa giác theo đường bay, từ đó tính toán được NPS. Để tìm được đồ thị Voronoi của đám mây điểm nhóm tác giả dựa trên thuật toán dòng quét Fortune (từ dưới lên trên), thuật toán được thực hiện theo các bước sau:

Ta có tập điểm  $P = \{p_1, p_2, \dots, p_n\}$

- Bước 1: Xác định vị trí các điểm dựa trên tọa độ điểm  $(x, y)$  – kí hiệu  $s$
- Bước 2: Xác định một phép biến đổi  $*(P)$  với  $*(P) = (px, py + d(P))$ , xác định đường quét  $T$ ,  $R_s$  là vùng bao quanh vị trí  $s$ ,  $B_{ij}$  là vùng biên giữa hai vị trí  $s_i, s_j$ . Coi các  $s_m$  là các vị trí với giá trị tối thiểu của tọa độ  $y$  theo  $x$ . Tập  $Q = P - \{s_m\}$ , tạo các đường biên dọc ký hiệu  $B_{0s_1, s_2}, B_{0s_2, s_3}, \dots, B_{0s_{m-1}, s_m}$ . Đường quét  $T$  là tập hợp bao gồm các giá trị  $\{*(R_{s_1}), B_{0s_1, s_2}, *(R_{s_2}), B_{0s_2, s_3}, \dots, *(R_{s_{m-1}}), B_{0s_{m-1}, s_m}\}$
- Bước 3: nếu  $s_i$  là vị trí trong  $*(P)$  sử dụng phép biến đổi  $*P$  để tìm tần suất xảy ra của vùng  $R_s$  trong  $T$ . Nếu  $s_i$  thuộc vào  $T$  tạo một đường biên  $B$  mới và loại bỏ bất kỳ điểm nào nằm ở vùng ranh giới, và thêm bất kỳ điểm nào nằm ở vùng giao của vùng trên và dưới của đường quét.
- Bước 4: Nếu  $s_i$  là đỉnh của Voronoi, xét  $s_i$  theo vị trí bên trái hay bên phải của đường quét, tiến hành tạo mới đường biên  $B$  mới và loại bỏ hay thêm vào các điểm ở vùng biên theo vị trí của điểm
- Bước 5: Ghi nhận lại các vị trí  $s_i$  này vào tạo đường phân vùng vị trí
- Bước 6: Lưu lại các đường biên này trong đường quét để tạo thành đồ thị Voronoi

d) Sau khi tìm được đồ thị Voronoi của đám mây điểm LiDAR, NPS được tính theo mật độ điểm và sự phân bố điểm trong không gian theo các bước sau:

- Sắp xếp giá trị khoảng cách và mật độ từ thấp đến cao

Trong đó, khoảng cách (Point Spacing – PS) và mật độ điểm (Point Density – PD) được tính theo các công thức sau:

$$PD = \frac{1}{\delta_{Voronoi}}$$

$$PS = MAX(\Delta p_{ki} \in P_k)$$

Trong đó,  $\Delta p_{ki}$  – khoảng cách giữa hai điểm,  $P_k$  – tập điểm lân cận của  $p_k$ .

Tính toán phân vị của giá trị PS và PD, gán giá trị phân vị này = NPS

e) Sử dụng thuật toán kNN dựa trên NPS loại bỏ nhiễu đám mây điểm LiDAR

Dựa trên mật độ điểm và NPS, sử dụng kNN tìm kiếm điểm lân cận gần nhất theo NPS, sử dụng giá trị ngưỡng là một số nguyên  $k$  thường được lấy theo độ lớn của dữ liệu, và giá trị NPS cho mọi điểm được kí hiệu là  $x$ . Khoảng cách giữa các điểm mẫu và  $x$  được sắp xếp theo:

$$D_1(x) < \dots < D_{k-1}(x) < D_k(x) < \dots < D_N(x)$$

Tính toán xác suất theo công thức:

$$\mathcal{P}(x) = N^{-1} D_k(x) \sum_{i=1}^N K[D_k(x)(x - X_i)]$$

Trong đó,  $D_k(x)$  là khoảng cách từ  $x$  đến  $k$  láng giềng của nó,  $K(\dots)$  là hàm Gaussian.

Sau khi có giá trị xác suất, loại bỏ biên bằng cách so sánh độ cao của các điểm với điểm lân cận của nó, nhằm tìm kiếm những điểm loại bỏ. Điểm bị loại bỏ là điểm có giá trị độ cao thay đổi quá lớn trước và sau khi so sánh nó với  $k$  láng giềng của nó và giá trị mật độ của nó là thấp.

Để so sánh được độ cao của các điểm với điểm lân cận, sử dụng công thức:

$$\left\{ \begin{array}{l} Z_{th} = 3 * \mathcal{P}(x) * Z_{std} \\ Z_{std} = \sqrt{\frac{1}{k} \sum_{k=1}^k (Z_k - Z_{mean})^2} \\ Z_{mean} = \frac{1}{k} \sum_{k=1}^k Z_k \end{array} \right.$$

Với,  $Z_{th}$  là ngưỡng độ cao để phát hiện điểm nhiễu,  $Z_{std}$  là phân bố chuẩn của điểm lân cận,  $\mathcal{P}(x)$  là hàm xác suất,  $Z_{mean}$  là trung bình độ cao của điểm lân cận,  $Z_k$  là độ cao của điểm đang xét nhiễu. Nếu giá trị  $Z_k > Z_{th}$ , đó là nhiễu, điểm sẽ bị loại bỏ.

f) Phân lớp đám mây điểm sau loại bỏ nhiễu, thành lập DEM và đánh giá độ chính xác

Đám mây điểm sau khi loại bỏ nhiễu thực hiện phân lớp với thuật toán cực đại hóa kỳ vọng được công bố tại Hội thảo Khoa học Quốc gia “Một số vấn đề chọn lọc của công nghệ thông tin và truyền thông” lần thứ XXII, tháng 6/2019 tại Thái Bình. Kết quả phân lớp sẽ được so sánh với quá trình phân lớp đám mây điểm không được loại bỏ với nhiễu với các độ đo Precision, Recall và F1.

**IV. THỬ NGHIỆM**

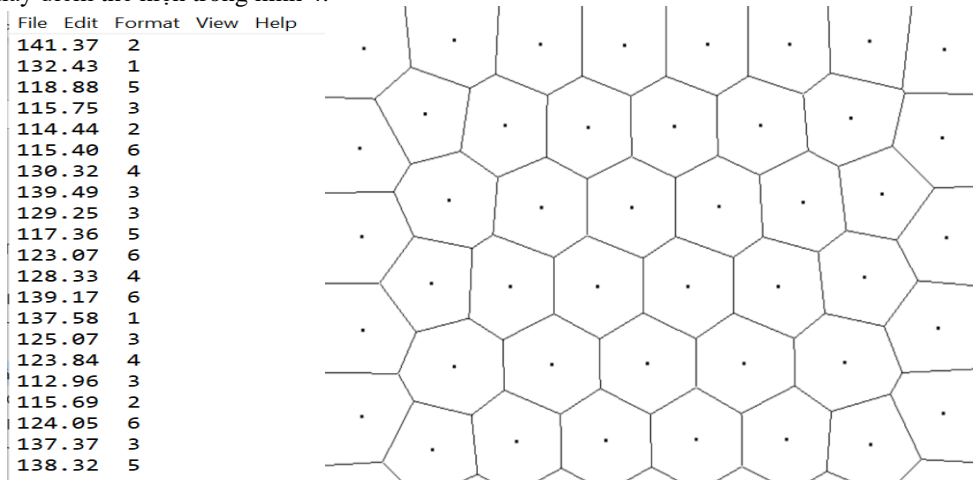
Bộ dữ liệu được sử dụng trong thực nghiệm là khu vực dân cư của Thành phố Bắc Ninh. Đây là khu vực có mật độ dân cư bình thường, có diện tích khoảng 6km<sup>2</sup>, mật độ dân cư cao, có đường giao thông và nhiều cây. Đám mây điểm với 606.249 điểm, giá trị tọa độ Xmin = 442160.213633; Ymin = 3940060.013977; Xmax = 442377.312573; Ymax = 3940171.064913. Mật độ điểm là 33.13 pts/m<sup>2</sup>, sử dụng công nghệ ALS (Airborne Laser Scanning).



**Hình 3.** Ảnh vệ tinh của khu vực thu nhận dữ liệu

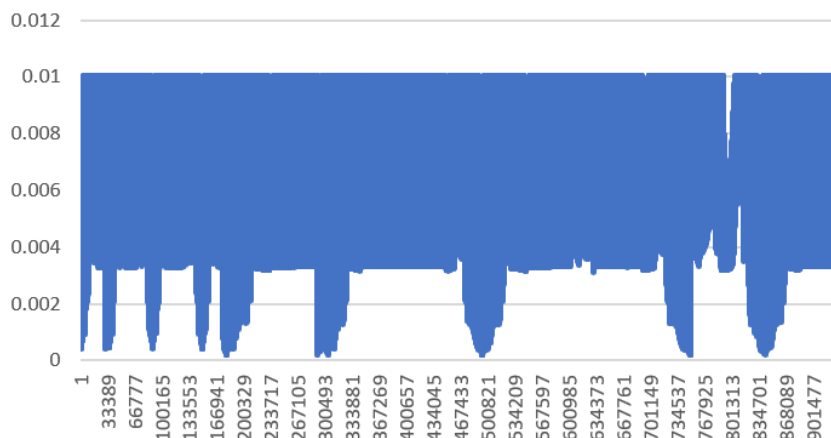
Thực hiện loại bỏ nhiễu đám mây điểm sử dụng NPS theo phương pháp đề xuất thu được kết quả như sau:

- Trích xuất dữ liệu độ cao điểm của đám mây điểm và một phân đồ thị voronoi được xây dựng trên tập hợp độ cao đám mây điểm thể hiện trong hình 4.



**Hình 4.** Giá trị độ cao của điểm và một phân đồ thị Voronoi được tạo

- Dựa trên đồ thị Voronoi, tính toán sự phân bố mật độ độ cao điểm dữ liệu, và tính toán giá trị PD thể hiện trong hình 5, 6



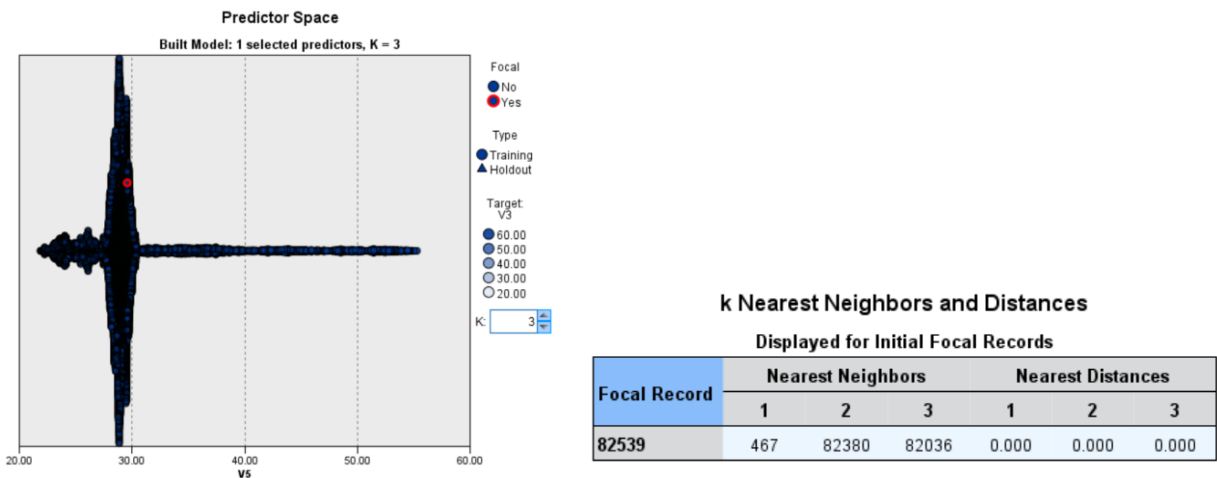
**Hình 5.** Phân bố mật độ điểm

X	Y	Z	Z	
420996.5	4467690	140.3	140.3	0.005153
420999.6	4467690	141.8	141.8	0.00493
420999.1	4467689	143.17	143.17	0.004729
420995.9	4467690	143.73	143.73	0.004648
420995.8	4467689	139.49	139.49	0.005274
420993	4467690	148.41	148.41	0.003989
420986.4	4467690	140.9	140.9	0.005063
420983.2	4467690	139.12	139.12	0.005329
420980.2	4467690	142	142	0.0049
420949.9	4467690	128.33	128.33	0.006967
420981.9	4467690	139.18	139.18	0.00532
420985.2	4467690	140.73	140.73	0.005088
420991.8	4467690	151.53	151.53	0.003576
420994.7	4467689	143.98	143.98	0.004611
420997.8	4467689	142.86	142.86	0.004774
420998.2	4467688	142.29	142.29	0.004858
420995.2	4467689	143.84	143.84	0.004632
420995	4467688	139.17	139.17	0.005322
420992.2	4467689	148.57	148.57	0.003968
420992	4467689	141.66	141.66	0.00495

**Hình 6.** Giá trị PD của bộ dữ liệu

- Sắp xếp giá trị PD từ lớn đến nhỏ lấy phân vị của nó ta có giá trị của NPS = 3.02
- Sử dụng kNN tìm điểm lân cận, thực hiện theo bước f của phương pháp, mô hình cho ta kết quả trong hình 7,

8, 9



**Hình 7.** Mô hình sau khi chạy, điểm dữ liệu và khoảng cách

So sánh điểm với ngưỡng loại bỏ nhiễu, kết quả tỉ lệ nhiễu được thể hiện trong hình 8

**Error Summary**

Partition	Sum-of-Squares Error
Training	0.269

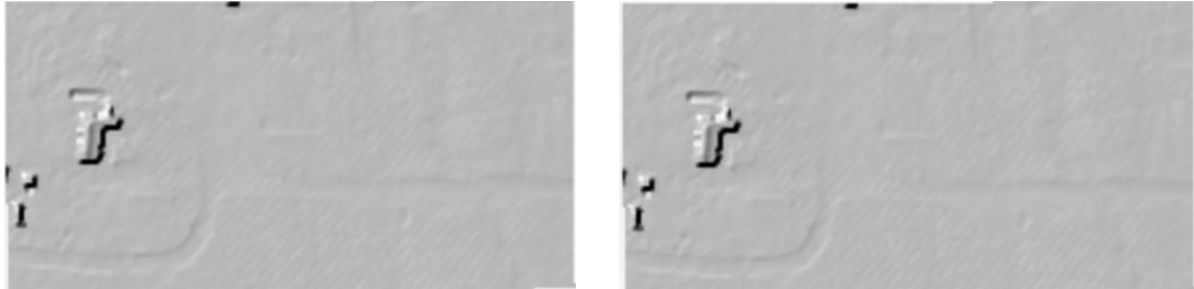
**Hình 8.** Tỉ lệ nhiễu trong dữ liệu

Dữ liệu sau khi loại bỏ nhiễu được thực hiện phân lớp thành lớp mặt đất và không mặt đất, với lớp mặt đất được sử dụng để thành lập DEM. Sau khi thực hiện, kết quả được so sánh với bộ dữ liệu chưa loại bỏ nhiễu trước khi thực hiện phân lớp và thành lập DEM dựa trên các độ đo Precision, Recall, F1.

Kết quả so sánh độ chính xác trên độ đo được thể hiện trong bảng 2, mô hình DEM được thể hiện trong hình 9.

Bảng 1. So sánh kết quả phân lớp trên các độ đo

	Không loại bỏ nhiễu	Có loại bỏ nhiễu
Precision	89.20%	93.60%
Recall	89.50%	92.70%
F1	89.35%	93.15%



(a) DEM trước khi loại nhiễu

(b) DEM sau khi loại nhiễu

**Hình 9.** Mô hình DEM được thành lập với bộ dữ liệu trước và sau khi loại bỏ nhiễu

Qua đó, có thể thấy, việc loại bỏ nhiễu trước khi sử dụng dữ liệu đám mây điểm LiDAR vào các bài toán ứng dụng là hoàn toàn cần thiết. Qua bảng đánh giá độ chính xác của kết quả phân lớp chỉ ra rằng, nhiễu ảnh hưởng nhiều đến bài toán này. Nếu thực hiện được bài toán loại bỏ nhiễu trước khi xử lý dữ liệu LiDAR sẽ giúp kết quả của bài toán ứng dụng chính xác hơn.

## V. KẾT LUẬN

Hiện nay, LiDAR ngày càng được sử dụng rộng rãi trong các lĩnh vực của đời sống xã hội nhờ khả năng ứng dụng cao của mình. Loại bỏ nhiễu trong quá trình xử lý là hoàn toàn cần thiết để giúp tăng độ chính xác cho bài toán xử lý dữ liệu LiDAR sau này. Với NPS là giá trị đặc trưng của bộ dữ liệu, được tính toán dựa trên giá trị độ cao của điểm, được sử dụng trong quá trình loại bỏ nhiễu là hoàn toàn đúng đắn, khi có thể sử dụng và loại bỏ nhiễu hiệu quả. Trong bài báo, nhóm tác giả đã sử dụng khái niệm NPS trong thuật toán kNN tìm điểm lân cận và tính hàm phân bố mật độ, sau đó dựa trên ngưỡng để thực hiện loại bỏ nhiễu trong dữ liệu LiDAR. Kết quả thực nghiệm chỉ ra rằng việc thực hiện phương pháp giúp tăng độ chính xác của bài toán phân lớp thành lập DEM. Tuy nhiên, cần thử nghiệm trên đa dạng bộ dữ liệu để đánh giá tốt hơn phương pháp đề xuất.

## VI. TÀI LIỆU THAM KHẢO

- [1] X. Zhu, X. Wu, "Class Noise vs. Attribute Noise: A Quantitative Study," *Artificial Intelligence Review*, 2004.
- [2] Shivani Gupta, Atul Gupta, "Dealing with noise problem in machine learning data-sets: a systematic review," *Procedia Computer Science*, tập 161, pp. 466-474, 2019.
- [3] A Ullrich, M.Plennigbauer, "Noisy LiDAR point clouds: impact on information extraction in high - precision LiDAR surveying," trong *Laser radar technology and application*, Florida, 2018.
- [4] Z. Afrose, "Adaptive methods for point cloud and mesh processing," Old Dominion University, 2018.
- [5] Xudong Lai, Min Zheng, "A denoising method for LiDAR full waveform data," *Mathematical Problems in Engineering*, pp. 1-8, 2015.
- [6] Adrian Ghimbasan, "PRE-PROCESSING AND POST-PRECESSING LIDAR DATA," *Vasile Goldiș Western University of Ara*, tập 10, số 1, pp. 7-13, 2015.
- [7] Zhenyang Hui, Dajun Li, Shuanggen Jin, "Automatic DTM extraction from airborne LiDAR based on expectation - maximization," *Optics and Laser Technology*, tập 112, pp. 43-55, 2019.
- [8] Mohsen Azadbakht a,b, \*, Clive S. Fraser a,b, Chunsun Zhang a,c, Joseph Leach b, "A SIGNAL DENOISING METHOD FOR FULL-WAVEFORM LIDAR DATA," trong *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Antalya, Turkey, 2013.
- [9] Shan Yong-hua1, Zhang Xu-qing1, "Denoising Algorithm of Airborne LIDAR Point Cloud Based on 3D Grid," *International Journal of Signal Processing, Image Processing and Pattern Recognition*, tập 10, số 3, pp. 85-92, 2017.
- [10] Chunhua Hu 1,†Zhou Pan, Pingping Li, "A 3D Point Cloud Filtering Method for Leaves Based on Manifold Distance and Normal Estimation," *Remote Sensing MDPI*, tập 11, 2019.
- [11] Micheal Ty Naus, Furgo Horizon, "LiDAR density and spacing speci," *ASPRS*, 2010.

- [12] MOA, *LiDAR Quality Inspection*, 2015.
- [13] FELIX, “Point Density and Point Spacing,” 2015. [Trực tuyến]. Available: <https://felix.rohrba.ch/en/2015/point-density-and-point-spacing/>.
- [14] B. Kalantari, “The state of the art of Voronoi diagram research,” *Transactions on Computational Science*, tập XX, 2013.
- [15] Wojciech Polojski, Paulina Polojska, “Voronoi diagrams - Inventor, method, applications,” *Polish Cartographical Review*, tập 50, số 3, pp. 141-150, 2018.

## LIDAR POINT CLOUD DENOISING METHOD USING NOMINAL POINT SPACING (NPS)

**Nguyen Thi Huu Phuong, Dang Van Duc, Nguyen Truong Xuan**

**ABSTRACT**— *With the ability to collect data over a large space in a short period of time and regardless of weather, LiDAR technology is increasingly widely used in the field of Earth Science. The data obtained from LiDAR data is usually highly accurate, each point is characterized by its coordinates, altitude, and precision. However, with the large number of points collected in each measurement, the noise in the LiDAR point cloud data greatly affects the processing results for later applications. Therefore, it is necessary to remove noise before performing further data processing with LiDAR data. In this paper, the authors propose a method to remove noise from LiDAR data using nominal point spacing (NPS). The method has been tested with the LiDAR point cloud acquired in Bac Ninh for an accuracy of 93.6%.*