



## Journal of Mining and Earth Sciences

Website: <https://jmes.humg.edu.vn>



# Optimizing the Prophet model using Bayesian Optimization for land subsidence prediction in central wards and communes of Ca Mau



Khien Trung Ha <sup>1</sup>, Anh Van Tran <sup>2,\*</sup>, Nhan Quy Pham <sup>3</sup>, Dung Ngoc Luong <sup>1</sup>, Chieu Dinh Vu <sup>1</sup>, Hue Dieu Dang <sup>1</sup>, Huy Dinh Nguyen <sup>1</sup>

<sup>1</sup> Hanoi University of Civil Engineering, Hanoi, Vietnam

<sup>2</sup> Hanoi University of Mining and Geology, Hanoi, Vietnam

<sup>3</sup> Hanoi University of Natural Resources and Environment, Hanoi, Vietnam

### ARTICLE INFO

#### Article history:

Received 09<sup>th</sup> July 2025

Revised 03<sup>rd</sup> Sept. 2025

Accepted 08<sup>th</sup> Sept. 2025

#### Keywords:

Bayesian Optimization,  
Forecasting,  
Land subsidence,  
Prophet,  
PS-InSAR.

### ABSTRACT

Land subsidence in the Ca Mau region has become increasingly complex, posing a significant threat to the sustainable development of this coastal delta area. In this context, the temporal monitoring and forecasting of land subsidence are essential for early warning systems and for supporting disaster response and spatial planning. This study proposes the application of an advanced machine learning model-Prophet-integrated with Bayesian Optimization (BO) to improve the accuracy of temporal subsidence prediction. The dataset comprises time series of ground deformation at 1817 points, extracted using Persistent Scatterer Interferometric Synthetic Aperture Radar (PS-InSAR) technology. These data were collected continuously from January 2015 to January 2019, with a total of 178 temporal acquisitions. Experimental results demonstrate that the Prophet model combined with BO achieves high predictive performance, with an average Root Mean Square Error (RMSE) of 3,4 mm and a Mean Absolute Error (MAE) of 2,6 mm. Notably, at the reference date of January 31, 2019, the predicted values exhibited a strong correlation with PS-InSAR observations ( $R^2 = 0,96$ ). Given this level of accuracy, the proposed model shows great potential for long-term subsidence trend monitoring and risk mapping, particularly in areas with slow and stable subsidence such as the Ca Mau coastal plain.

Copyright © 2025 Hanoi University of Mining and Geology. All rights reserved.

\*Corresponding author

E - mail: [tranvananh@humg.edu.vn](mailto:tranvananh@humg.edu.vn)

DOI: 10.46326/JMES.2025.66(5).04



## Tạp chí Khoa học Kỹ thuật Mỏ - Địa chất

Trang điện tử: <https://tapchi.humg.edu.vn>



# Tối ưu hóa mô hình Prophet bằng Bayesian Optimization trong dự báo lún đất tại một số phường, xã trung tâm của Cà Mau

Hà Trung Khiên<sup>1</sup>, Trần Văn Anh<sup>2,\*</sup>, Phạm Quý Nhân<sup>3</sup>, Lương Ngọc Dũng<sup>1</sup>, Vũ Đình Chiều<sup>1</sup>, Đặng Diệu Huệ<sup>1</sup>, Nguyễn Đình Huy<sup>1</sup>

<sup>1</sup> Trường Đại học Xây dựng Hà Nội, Hà Nội, Việt Nam.

<sup>2</sup> Trường Đại học Mỏ - Địa chất, Hà Nội, Việt Nam.

<sup>3</sup> Trường Đại học Tài nguyên - Môi trường, Hà Nội, Việt Nam

### THÔNG TIN BÀI BÁO

#### Quá trình:

Nhận bài 09/7/2025

Sửa xong 03/9/2025

Chấp nhận đăng 08/9/2025

#### Từ khóa:

Bayesian Optimization,

Dự báo,

Lún đất,

Prophet,

PS-InSAR.

### TÓM TẮT

Hiện tượng lún đất tại Cà Mau ngày càng phức tạp, đe dọa sự phát triển bền vững của vùng đồng bằng ven biển này. Trong bối cảnh đó, việc giám sát và dự báo lún đất theo thời gian là cần thiết để cảnh báo sớm và hỗ trợ quy hoạch ứng phó thiên tai. Nghiên cứu này đề xuất ứng dụng mô hình học máy tiên tiến Prophet kết hợp với phương pháp tối ưu hóa Bayesian (Bayesian Optimization-BO) nhằm nâng cao độ chính xác trong dự báo lún đất theo thời gian. Dữ liệu sử dụng trong nghiên cứu bao gồm chuỗi thời gian lún đất tại 1817 điểm, được trích xuất từ công nghệ radar giao thoa tán xạ cố định (Persistent Scatterer InSAR-PS-InSAR). Các dữ liệu này được thu thập liên tục từ tháng 01/2015 đến tháng 01/2019 với tổng cộng 178 mốc thời gian. Kết quả thử nghiệm cho thấy mô hình Prophet kết hợp BO đạt hiệu suất dự báo cao, với sai số trung bình RMSE là 3,4 mm và MAE là 2,6 mm. Đặc biệt, tại thời điểm 31/01/2019, kết quả dự báo có độ tương quan rất cao với dữ liệu thực đo từ PS-InSAR ( $R^2 = 0,96$ ). Với độ chính xác này, mô hình có thể được ứng dụng hiệu quả trong giám sát xu hướng lún đất dài hạn và xây dựng bản đồ nguy cơ tại các khu vực có tốc độ lún chậm và ổn định như đồng bằng ven biển Cà Mau.

© 2025 Trường Đại học Mỏ - Địa chất. Tất cả các quyền được bảo đảm.

\*Tác giả liên hệ

E - mail: tranvananh@humg.edu.vn

DOI: 10.46326/JMES.2025.66(5).04

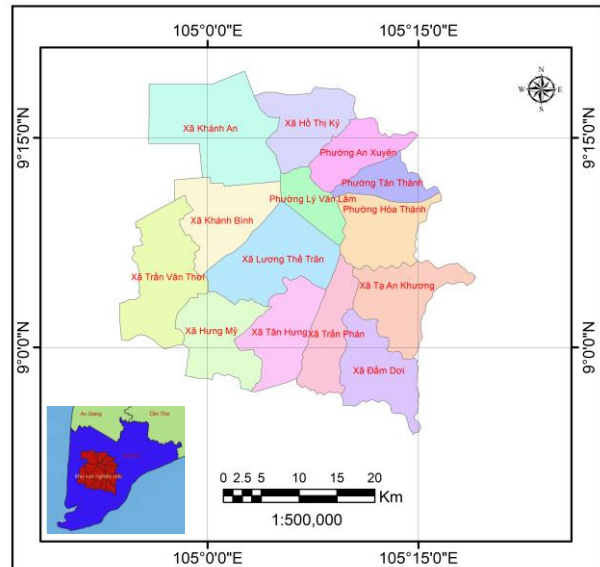
## 1. Mở đầu

Lún đất là hiện tượng biến dạng lún xuống của mặt đất do các yếu tố tự nhiên hoặc nhân tạo như khai thác nước ngầm, tải trọng công trình hoặc thay đổi địa chất. Hiện tượng lún đất có thể gây ra những hệ quả nghiêm trọng liên quan đến địa chất, thủy văn, môi trường và phát triển kinh tế. Mặc dù khó có thể tránh được trong ngành công nghiệp khai thác nhưng lún đất vẫn có thể được kiểm soát thông qua quy định pháp luật, giám sát khai thác và quy hoạch hợp lý khi có những bản đồ cảnh báo khả năng lún (Shi và nnk., 2008).

Trên thế giới cũng như tại Việt Nam, đã có nhiều nghiên cứu tập trung vào dự báo lún đất, với các phương pháp từ mô hình truyền thống đến các mô hình hiện đại dựa trên trí tuệ nhân tạo và học máy. Một số mô hình dựa trên cơ học đất như mô hình Terzaghi hoặc Biot được sử dụng rộng rãi để tính toán lún do cố kết của tầng đất yếu dưới tác động của tải trọng và thay đổi áp suất nước lỗ rỗng (Chunlin, 2014). Ngoài ra, các mô hình hồi quy tuyến tính hoặc ARIMA cũng được áp dụng nhằm thiết lập mối quan hệ giữa độ lún và các yếu tố đầu vào như mực nước ngầm, loại đất, tải trọng công trình (Đình và nnk., 2018). Tuy nhiên, những mô hình này thường giả định quan hệ tuyến tính, ít linh hoạt trong các điều kiện phức tạp.

Để giải quyết những hạn chế đó, những năm gần đây, các phương pháp học máy (Machine Learning) đã được nghiên cứu và ứng dụng mạnh mẽ. Nhiều nghiên cứu trên thế giới đã áp dụng

thành công các thuật toán học máy để dự báo lún đất. Các mô hình hồi quy Logistic (Logistic Regression - LR), rừng ngẫu nhiên (Random Forest - RF), cây hồi quy tăng cường (Boosted Regression Trees - BRT), máy vector hỗ trợ (Support Vector Machine - SVM) được Sekkeravani và nnk. (2022) ứng dụng xây dựng bản đồ nguy cơ lún đất và RF cho hiệu suất tốt nhất ( $AUC = 0,953$ ). Wang và nnk. (2023) cho thấy mạng nơ-ron truyền ngược (Backpropagation Neural Network - BPNN) hoạt động tốt hơn mô hình rừng ngẫu nhiên (RF) tại Sơn Đông, Trung Quốc, trong khi Liu và nnk. (2023) kết hợp dữ liệu Sentinel-1 và Landsat để dự báo lún đất tại đồng bằng sông Tây Giang với mô hình hồi quy rừng ngẫu nhiên (Random Forest Regressor - RFR) cho  $R^2 = 0,631$ . Eghrari và nnk. (2023) đã chứng minh XGBoost ( $R^2 = 0,932$ ) hiệu quả hơn rừng ngẫu nhiên (RF) ( $R^2 = 0,8355$ ) tại đồng bằng Kashan. Nghiên cứu của Ku và Liu (2023) sử dụng mạng nơ-ron nhân tạo (ANN) để dự báo lún đất tại Yunlin (Đài Loan) với kết quả phù hợp thực địa. Gần đây, Liu và nnk. (2024) đã so sánh mô hình hồi quy sử dụng thuật toán XGBoost (XGBR) và mô hình Long Short-Term Memory (LSTM), kết quả cho thấy XGBR đạt  $RMSE = 0,37$ , tốt hơn LSTM. Tại Việt Nam, nghiên cứu học máy trong dự báo lún đất còn hạn chế. Các nghiên cứu chủ yếu tập trung vào dự báo nguy cơ sụt lún, lún lút và lớp phủ bề mặt (Doãn và Trần, 2023; Phạm và Vũ, 2019; Vân Anh và nnk, 2023).



Hình 1. Khu vực nghiên cứu.

Đối với khu vực đồng bằng sông Cửu Long nói chung và Cà Mau nói riêng việc ứng dụng các thuật toán học máy để dự báo lún đất cũng đã được làm trong những năm gần đây, tuy nhiên đa số những nghiên cứu này chỉ tập trung vào dự báo nhạy cảm hoặc nguy cơ lún đất bằng một số thuật toán khác nhau như Random Forest, Classification and Regression Trees (CAST) (Anh và nnk., 2023) hoặc Gradient Boosting (GB) và Support Vector Regression (SVR) dựa vào dữ liệu được xử lý từ ảnh Sentinel-1 và các yếu tố địa môi trường (Anh và nnk., 2024). Một nghiên cứu khác sử dụng chuỗi thời gian GNSS để dự báo lún đất nhưng số lượng điểm dữ liệu rất hạn chế (Do và nnk., 2024).

Trong nghiên cứu này, mô hình Prophet kết hợp với phương pháp tối ưu Bayesian được sử dụng để phân tích và dự báo lún đất theo chuỗi thời gian trên cơ sở dữ liệu PS-InSAR đa thời gian. Prophet là một mô hình dự báo chuỗi thời gian do Facebook phát triển, được thiết kế để xử lý dữ liệu có xu hướng rõ rệt và chu kỳ. Mô hình hoạt động theo cấu trúc cộng tuyến, dễ sử dụng và cho phép dự báo linh hoạt trong tương lai ngay cả khi dữ liệu có thiếu hụt hoặc không đều. Prophet đặc biệt phù hợp với các bài toán như dự báo lún đất, nơi xu hướng dài hạn đóng vai trò quan trọng (Taylor và Letham, 2018). Mô hình Prophet đã được ứng dụng thành công trong dự báo lún đất ở một số khu vực trên thế giới cho kết quả có độ chính xác đáng tin cậy (Hung và nnk., 2024; Dongwei và nnk., 2024).

## 2. Khu vực nghiên cứu

Khu vực nghiên cứu bao gồm toàn bộ diện tích của 14 phường và xã thuộc trung tâm tỉnh Cà Mau, bao gồm các phường: Lý Văn Lâm, Hòa Thành, Tân Thành, An Xuyên và các xã: Hồ Thị Kỷ, Khánh An, Khánh Bình, Trần Văn Thới, Lương Thế Trân, Hưng Mỹ, Tân Hưng, Trần Phán, Đầm Dơi, Tạ An Khương (Hình 1). Địa hình khu vực chủ yếu là vùng đồng bằng thấp có độ cao trung bình từ 0÷1,5 m so với mực nước biển (Dao và Nguyen, 2018).

Hiện tượng lún đất tại khu vực Đồng bằng sông Cửu Long nói chung và tỉnh Cà Mau nói riêng đang trở thành vấn đề đáng báo động. Theo nhiều nghiên cứu, nguyên nhân chính gây ra lún đất tại khu vực này là do khai thác nước ngầm quá mức (Erban và nnk., 2014; Karlsrud và nnk., 2017). Bên cạnh đó, các yếu tố như tải trọng công trình và đặc

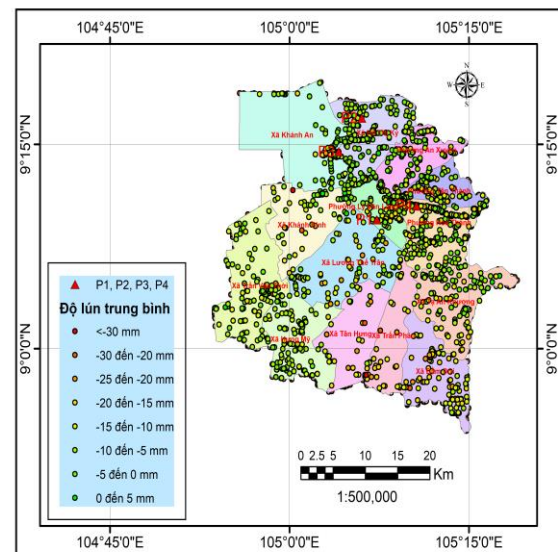
điểm nền đất yếu cũng góp phần đáng kể vào quá trình sụt lún (Tran và nnk., 2024).

## 3. Dữ liệu và phương pháp nghiên cứu

### 3.1. Dữ liệu sử dụng

Dữ liệu sử dụng là 1817 điểm lún với 178 mốc thời gian, được lấy trong tập điểm giám sát lún tại Đồng bằng sông Cửu Long của ESA thu thập từ 01/2015 đến 01/2019. Các điểm lún này được xác định bằng kỹ thuật Radar giao thoa tán xạ cố định (Persistent Scatterer Interferometry - PSI) từ tập hợp ảnh radar Sentinel-1 thuộc chương trình Copernicus. Các điểm sụt lún đất này đã được chứng minh có độ chính xác thỏa đáng, có thể sử dụng sau khi kiểm định với các điểm đo thực địa

(EMSN062, 2019) và được kiểm chứng bởi các nghiên cứu khác (Minderhoud, 2021; Tran và nnk., 2024). Vị trí các điểm quan trắc lún được thể hiện tại Hình 2.



Hình 2. Các điểm lún PS-InSAR sử dụng trong nghiên cứu.

### 3.2. Mô hình Prophet và phương pháp tối ưu hóa Bayesian

Prophet là một công cụ dự báo chuỗi thời gian mã nguồn mở được Facebook phát triển và công bố vào năm 2017. Mô hình này dựa trên khung hồi quy cộng tính (additive regression framework) (Taylor và Letham, 2018):

$$y_{(t)} = g_{(t)} + s_{(t)} + h_{(t)} + \varepsilon_t \quad (1)$$

Trong đó tín hiệu đầu ra được biểu diễn như tổng của ba thành phần chính:

- Thành phần xu thế ( $g_{(t)}$ ): Thành phần này mô tả sự biến động dài hạn không mang tính chu kỳ, được mô hình hóa bằng hồi quy tuyến tính từng phần hoặc hồi quy logistic, đặc biệt phù hợp với các chuỗi có xu hướng tăng trưởng hoặc bão hòa (Chang và nnk., 2025). Đối với xu hướng logistic, hàm được định nghĩa như sau (Taylor và Letham, 2018):

$$g_{(t)} = \frac{C_{(t)}}{1 + \exp\{[-(k+a(t)^T\delta)], [t-(m+a(t)^T\gamma)]\}} \quad (2)$$

Trong đó:  $C_{(t)}$  là khả năng bão hòa (saturation capacity);  $k(t)$  là tốc độ tăng trưởng (growth rate);  $m(t)$  là thời điểm chuyển đổi (offset parameter).

- Thành phần thời vụ ( $s_{(t)}$ ): Thành phần này phản ánh các chu kỳ định kỳ như hàng tuần hoặc hàng năm. Prophet sử dụng khai triển Fourier để mô hình hóa đặc trưng mùa vụ (Taylor và Letham, 2018):

$$s_{(t)} = \sum_{n=1}^N \left[ a_n \cos\left(\frac{2\pi nt}{P}\right) + b_n \sin\left(\frac{2\pi nt}{P}\right) \right] = x_{(t)} \beta \quad (3)$$

Trong đó:  $a_n, b_n \sim N(0, \sigma^2)$  - các hệ số Fourier;  $\sigma$  - tham số điều chỉnh biên độ ảnh hưởng của mùa vụ.

- Thành phần ảnh hưởng của sự kiện đặc biệt ( $h_{(t)}$ ): Thành phần này cho phép mô hình tính đến các sự kiện đặc biệt như ngày lễ, kỳ nghỉ, hoặc các sự kiện mang tính chu kỳ không đều.

-  $\varepsilon_t$  - nhiễu trắng (sai số ngẫu nhiên không hệ thống).

Tối ưu hóa Bayesian là một phương pháp hiệu quả để tìm cực trị toàn cục của các hàm mục tiêu phức tạp và tốn kém khi đánh giá (Snoek và nnk., 2012). Phương pháp này đặc biệt phù hợp cho việc điều chỉnh siêu tham số - những tham số không được học trực tiếp từ dữ liệu mà phải được thiết lập trước quá trình huấn luyện (Bernd và nnk., 2023).

Tối ưu hóa Bayesian (Bayesian Optimization - BO) là phương pháp xây dựng một mô hình xác suất, thường là Gaussian Process (GP), để mô phỏng hành vi của hàm mục tiêu dựa trên các đánh giá đã có. GP không chỉ dự đoán giá trị trung

bình mà còn thể hiện độ không chắc chắn tại từng điểm trong không gian tìm kiếm. Dựa trên mô hình này, BO sử dụng một hàm thu nhận (acquisition function) như Expected Improvement hoặc Upper Confidence Bound để chọn điểm tiếp theo cần đánh giá, với mục tiêu cân bằng giữa khai thác các vùng có giá trị tốt (exploitation) và khám phá những vùng chưa biết rõ (exploration). Bayesian Optimization đặc biệt hữu ích trong các bài toán tối ưu phức tạp, hàm mục tiêu không rõ ràng và có nhiều điểm cực trị khiến việc tìm lời giải tối ưu trở nên khó khăn.

BO-Prophet là sự kết hợp giữa mô hình Prophet và thuật toán tối ưu hóa Bayesian nhằm tự động hóa quá trình điều chỉnh siêu tham số cho Prophet. Quy trình hoạt động bao gồm các bước sau (Hình 3).

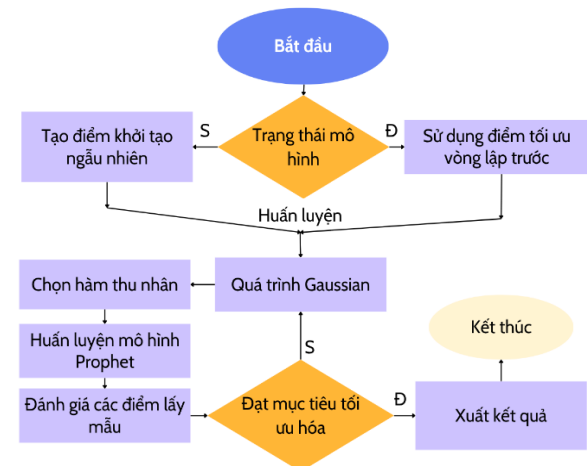
(1) Lấy mẫu ban đầu: Một số tổ hợp siêu tham số được chọn ngẫu nhiên để khởi tạo quá trình.

(2) Huấn luyện và đánh giá: Prophet được huấn luyện với từng cấu hình, và hiệu suất được đo lường bằng các chỉ số như MAE, RMSE hoặc MAPE.

(3) Xây dựng mô hình Gaussian Process: Một Gaussian Process được huấn luyện dựa trên dữ liệu thu thập được (siêu tham số và hiệu suất tương ứng) nhằm mô hình hóa hàm mục tiêu.

(4) Tối ưu hàm thu nhận: Hàm thu nhận (ví dụ: Expected Improvement, Upper Confidence Bound) được tối ưu để chọn ra tổ hợp siêu tham số kế tiếp có kỳ vọng mang lại cải thiện tốt nhất.

(5) Lặp lại quá trình: Các bước trên được lặp lại cho đến khi đạt điều kiện dừng (số vòng lặp nhất định hoặc hội tụ).



Hình 3. Sơ đồ thuật toán mô hình dự báo Prophet dựa trên tối ưu hóa Bayesian.

Khung BO-Prophet giúp cải thiện đáng kể hiệu suất dự báo của Prophet, giảm sự phụ thuộc vào chuyên môn thủ công trong việc lựa chọn siêu tham số và tăng khả năng tổng quát hóa mô hình với các chuỗi thời gian khác nhau.

#### 4. Kết quả và thảo luận

Nghiên cứu đã sử dụng 30 điểm lún có độ tương quan cao trên khu vực nghiên cứu được chọn để tối ưu hóa mô hình Prophet bằng phương pháp BO. Mỗi điểm có chuỗi giá trị lún ở 178 mốc thời gian từ 01/2015 đến 01/2019 trong đó giá trị lún ở 142 mốc thời gian đầu tiên được chọn làm tập huấn luyện và giá trị lún ở 36 mốc thời gian còn lại được chọn làm tập kiểm tra. Hiệu quả của mô hình được đánh giá thông qua sai số tuyệt đối trung bình (MAE). Kết quả các tham số tối ưu của mô hình được trình bày trong Bảng 1.

*Bảng 1. Các tham số mô hình được chuẩn hóa bằng BO.*

| Tham số                                                                         | Giá trị   |
|---------------------------------------------------------------------------------|-----------|
| Hệ số điều chỉnh độ linh hoạt của thành phần mùa vụ                             | 0,00049   |
| Chu kỳ theo năm                                                                 | Không     |
| Chu kỳ theo tuần                                                                | Không     |
| Chu kỳ theo ngày                                                                | Có        |
| Chế độ kết hợp thành phần mùa vụ vào mô hình                                    | Cộng thêm |
| Hệ số điều chỉnh độ nhạy của mô hình đối với các điểm thay đổi xu hướng         | 0,100393  |
| Tỷ lệ phần đầu của chuỗi thời gian mà Prophet sẽ tìm các điểm thay đổi xu hướng | 0,716926  |

Sau khi hoàn tất bước tối ưu, mô hình Prophet được huấn luyện và dự báo độ lún tại 1787 điểm được xác định bằng công nghệ PS-InSAR, với mỗi điểm có 178 mốc thời gian. Trong quá trình huấn luyện, độ lún của 142 mốc thời gian đầu tiên được dùng làm tập huấn luyện (Bảng 2) và mô hình được huấn luyện để dự báo độ lún cho 36 mốc thời gian còn lại.

*Bảng 2. Số lượng mốc thời gian trên huấn luyện và tập kiểm tra.*

| Tập dữ liệu    | Số lượng | Mốc thời gian                                                        |
|----------------|----------|----------------------------------------------------------------------|
| Tập huấn luyện | 142      | 10/1/2015; 3/2/2015; 15/2/2015; ...; 24/5/2018; 30/5/2018; 5/6/2018. |

|              |    |                                                                        |
|--------------|----|------------------------------------------------------------------------|
| Tập kiểm tra | 36 | 11/6/2018; 17/6/2018; 23/6/2018; ...; 19/1/2019; 25/1/2019; 31/1/2019. |
|--------------|----|------------------------------------------------------------------------|

Để đánh giá độ chính xác của các mô hình, hai chỉ số được sử dụng là sai số tuyệt đối trung bình (MAE) và căn bậc hai của sai số bình phương trung bình (RMSE).

Các công thức tính toán MAE và RMSE như sau (Willmott & Matsuura, 2005):

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

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (5)$$

Trong đó:  $y_i$  - giá trị quan sát thực tế;  $\hat{y}_i$  - giá trị dự báo từ mô hình. Việc áp dụng chuẩn hóa đối xứng trong các chỉ số MAE và RMSE nhằm hạn chế tác động của các giá trị ngoại lai đến quá trình đánh giá sai số.

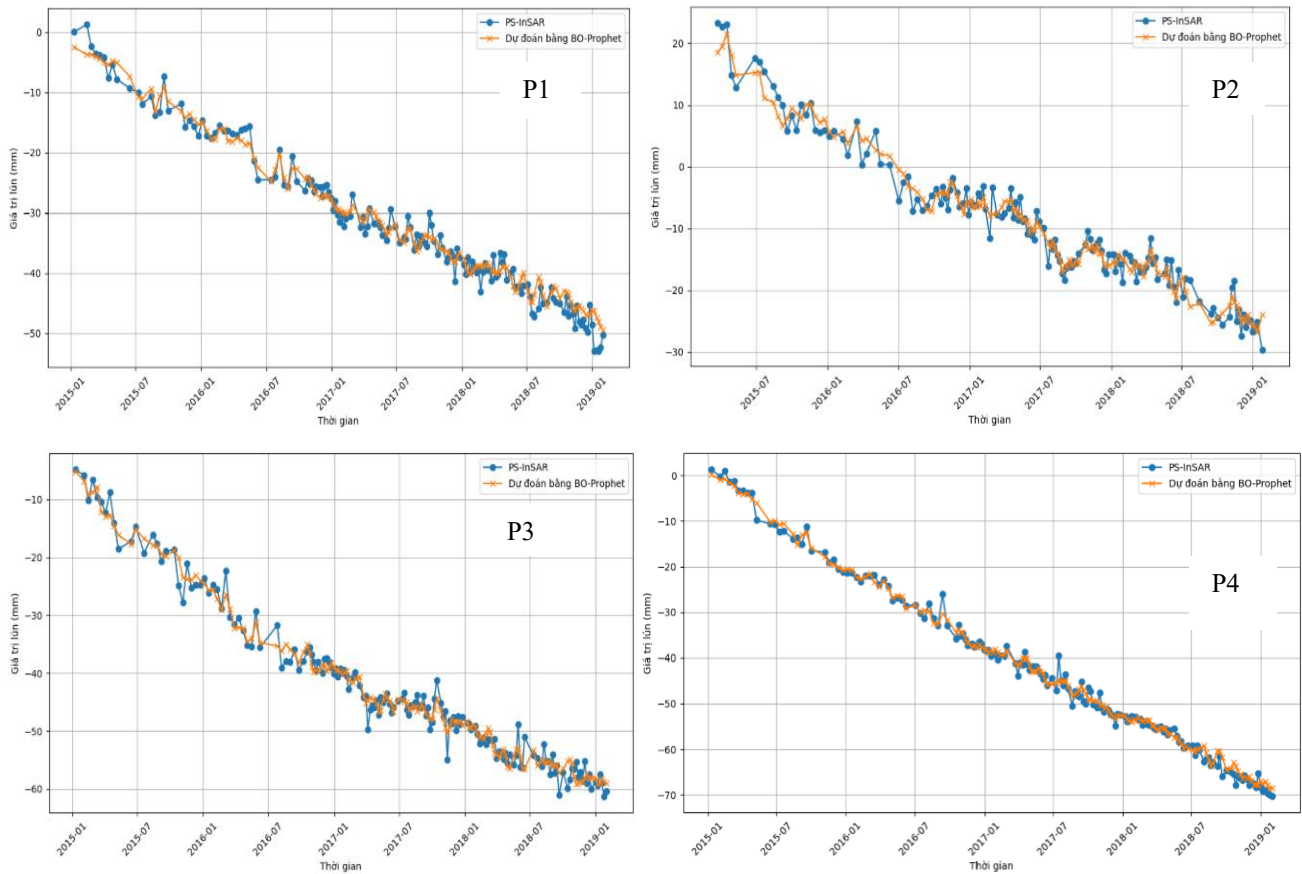
Hình 4 thể hiện biểu đồ lún của 4 điểm P1, P2, P3, P4 dự báo bằng mô hình BO-Prophet so với giá trị lún xác định bằng PS-InSAR (P1, P2, P3, P4 là các điểm lún điển hình được lựa chọn trên khu vực khảo sát, vị trí 4 điểm này được thể hiện ở Hình 2).

Biểu đồ tương quan giữa giá trị RMSE và MAE được thể hiện trong Hình 5. Kết quả cho thấy, giá trị RMSE và MSE trung bình của các chuỗi có giá trị từ 0÷4 mm.

Bảng 3 thể hiện các giá trị MAE và RMSE trên tập huấn luyện và tập kiểm tra của 10 chuỗi đầu tiên và 10 chuỗi cuối cùng.

*Bảng 3. Giá trị độ chính xác của 10 chuỗi đầu tiên và 10 chuỗi cuối cùng.*

| TT | Tập huấn luyện (mm) |      | Tập kiểm tra (mm) |      |
|----|---------------------|------|-------------------|------|
|    | MAE                 | RMSE | MAE               | RMSE |
| 0  | 2,5                 | 3,3  | 3,1               | 4,0  |
| 1  | 1,7                 | 2,4  | 2,6               | 3,1  |
| 2  | 1,6                 | 2,4  | 2,0               | 2,8  |
| 3  | 1,5                 | 2,2  | 3,0               | 3,6  |
| 4  | 1,9                 | 3,0  | 2,4               | 2,9  |
| 5  | 2,2                 | 3,5  | 2,4               | 3,0  |
| 6  | 2,5                 | 3,4  | 2,9               | 3,8  |
| 7  | 1,6                 | 2,5  | 1,7               | 2,5  |

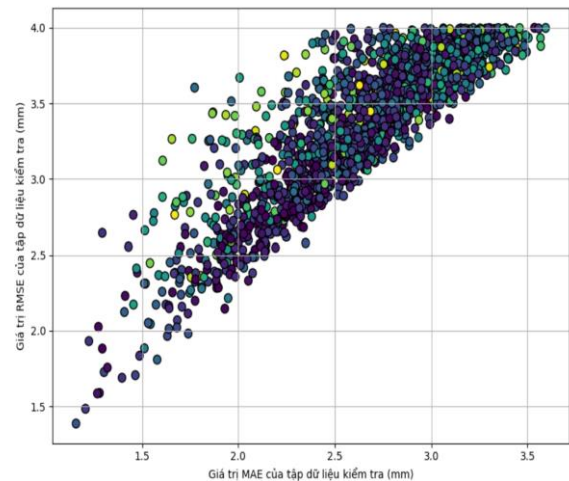


Hình 4. Biểu đồ lún dự báo bằng BO-Prophet so với lún thực tế PS-InSAR tại 4 điểm P1, P2, P3, P4.

|      |     |     |     |     |
|------|-----|-----|-----|-----|
| 8    | 2,3 | 3,8 | 2,5 | 3,1 |
| 9    | 1,4 | 2,0 | 2,9 | 3,5 |
| 10   | 1,2 | 2,1 | 2,4 | 3,1 |
| ...  | ... | ... | ... | ... |
| 1777 | 1,5 | 2,3 | 2,6 | 3,1 |
| 1778 | 1,5 | 2,1 | 3,3 | 3,7 |
| 1779 | 1,4 | 2,3 | 2,3 | 3,2 |
| 1780 | 2,3 | 3,1 | 3,3 | 3,9 |
| 1781 | 1,9 | 2,7 | 1,9 | 2,4 |
| 1782 | 1,2 | 1,9 | 1,5 | 2,1 |
| 1783 | 1,1 | 1,9 | 2,4 | 3,0 |
| 1784 | 2,5 | 3,4 | 2,6 | 3,3 |
| 1785 | 1,5 | 2,8 | 2,4 | 2,9 |
| 1786 | 2,0 | 2,6 | 3,5 | 4,0 |

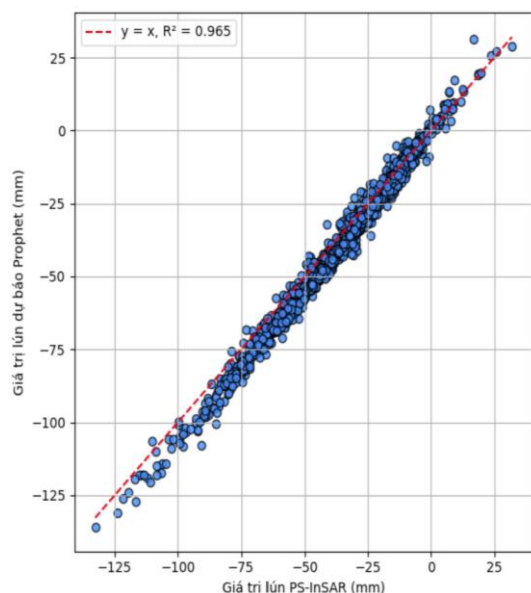
Xét trên toàn tập dữ liệu, đối với tập huấn luyện có giá trị RMSE trung bình là 3,2 mm và MAE trung bình là 2,1 mm; trên tập kiểm tra, RMSE trung bình là 3,4 mm và MAE trung bình là 2,6 mm.

Sau khi huấn luyện, mô hình Prophet được sử dụng để dự báo giá trị lún cho 1787 điểm lún tại

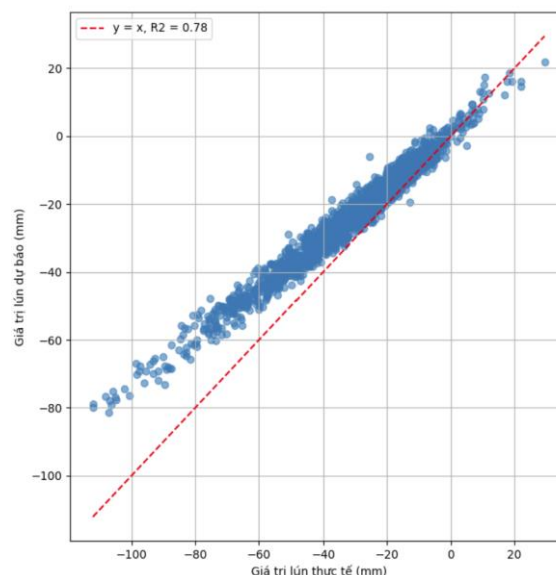


Hình 5. Biểu đồ tương quan giữa giá trị RMSE và MAE trên tập kiểm tra của 1787 điểm.

thời điểm cuối cùng (31/01/2019), các giá trị dự báo này được so sánh với các giá trị lún xác định bằng PS-InSAR ở cùng thời điểm thông qua mô hình hồi quy tuyến tính để phân tích tương quan, cho hệ số  $R^2 = 0,96$  (Hình 6).



Hình 6. Phân tích độ tương quan giữa độ lún dự báo bằng Prophet và độ lún xác định bằng PS-InSAR của 1787 điểm, tại thời điểm 31/01/2019.



Hình 7. Phân tích độ tương quan giữa độ lún dự báo bằng SVR và độ lún xác định bằng PS-InSAR của 1787 điểm, tại thời điểm 31/01/2019.

Để đánh giá hiệu quả của mô hình Prophet, nghiên cứu đã tiến hành so sánh với mô hình Support Vector Regression (SVR). SVR là một biến thể của thuật toán máy vector hỗ trợ (Support Vector Machine - SVM), được thiết kế để giải quyết các bài toán dự báo trong học máy (Drucker và nnk., 1997). Trong nghiên cứu (Anh và nnk., 2024) đã chứng minh cơ sở lý thuyết cũng như hiệu quả ứng dụng của SVR trong dự báo lún với bài toán phân loại.

Để dự báo lún bằng mô hình SVR, dữ liệu được sử dụng và chia tương tự như khi huấn luyện mô hình Prophet, trong đó 142 mốc thời gian đầu được sử dụng để huấn luyện, còn 36 mốc thời gian còn lại dùng để đánh giá khả năng dự báo của mô hình. Kết quả đánh giá mô hình SVR trên tập kiểm tra cho thấy RMSE trung bình là 6,1 mm và MAE trung bình là 4,8 mm. Sau khi huấn luyện, mô hình SVR cũng được sử dụng để dự báo giá trị lún tại 1787 điểm ở thời điểm cuối cùng (31/01/2019) và so sánh với dữ liệu lún từ PS-InSAR, cho hệ số tương quan  $R^2 = 0,78$  (Hình 7). So sánh các thông số này với kết quả đánh giá mô hình Prophet cho thấy Prophet có độ chính xác cao hơn.

## 5. Kết luận

Nghiên cứu đã phân tích khả năng của mô hình Prophet kết hợp với phương pháp tối ưu hóa

mô hình BO để dự báo lún theo chuỗi thời gian tại khu vực Cà Mau trên cơ sở dữ liệu đa thời gian PS-InSAR. Kết quả kiểm định mô hình trên tập dữ liệu kiểm tra cho thấy mô hình Prophet kết hợp với BO đạt hiệu suất dự báo cao, với sai số trung bình (RMSE) là 3,4 mm và sai số tuyệt đối trung bình (MAE) là 2,6 mm. Kết quả độ lún của 1787 điểm dự đoán bằng mô hình Prophet tại thời điểm 31/01/2019 so sánh với giá trị độ lún xác định bằng công nghệ PS-InSAR ở cùng thời điểm cho độ tương quan cao ( $R^2 = 0,96$ ).

Với độ chính xác trên, mô hình có thể được ứng dụng hiệu quả trong các bài toán giám sát xu hướng lún đất dài hạn và xây dựng bản đồ nguy cơ, đặc biệt tại các khu vực có xu hướng lún chậm và ổn định như đồng bằng ven biển Cà Mau. Tuy nhiên, mô hình Prophet giả định cấu trúc có tính chu kỳ trong dữ liệu, điều này có thể không phản ánh đầy đủ tính chất phi tuyến hoặc các biến động đột ngột của quá trình lún đất. Đồng thời, phương pháp tối ưu hóa BO cũng tiêu tốn tài nguyên tính toán, nhất là với tập dữ liệu có số lượng điểm lớn.

## Lời cảm ơn

Bài báo được sự hỗ trợ dữ liệu và kinh phí từ đề tài Khoa học và Công nghệ cấp trường Đại học Xây dựng Hà Nội, mã số 05-2024/KHXD-TĐ.

## Đóng góp của tác giả

Hà Trung Khiên - nghiên cứu phương pháp; thu thập, phân tích và xử lý dữ liệu; soạn thảo bài báo; viết bản thảo cuối cùng; Trần Văn Anh - nghiên cứu phương pháp; phân tích và xử lý dữ liệu; kiểm tra, biên tập bài báo; viết bản thảo cuối cùng; Phạm Quý Nhân - nghiên cứu phương pháp, kiểm tra bài báo; Lương Ngọc Dũng - nghiên cứu phương pháp; phân tích và xử lý dữ liệu; Vũ Đình Chiêu, Đặng Diệu Huệ, Nguyễn Đình Huy - đóng góp ý kiến, đánh giá, chỉnh sửa bài báo.

## Tài liệu tham khảo

- Anh, T. V., Dong, K. T., Khiên H. T., Hanh, T. H., Nghi, L. T., Hoa, P. T. T., Dung, N., Anh, L. H., Dinh, N. Q. (2023). Land Subsidence Susceptibility Mapping Using Machine Learning in the Google Earth Engine Platform. Trong N.N. Dao, T. N. Thinh, N. T. Nguyen (Ed.), *Intelligence of Things: Technologies and Applications* (pp. 55 - 64). Springer.
- Anh, T. V., Khiên H.T.; Đông, K.T.; Nghi, L.T.; Hạnh, T.H.; & Phong, D.H. (2024). Nghiên cứu khả năng của mô hình học máy GB và SVR trong thành lập bản đồ nguy cơ lún đất khu vực bán đảo Cà Mau, Việt Nam. *Tạp chí Khí tượng Thủy văn*, số 757, trang 60-73.
- Bernd, B., Martin, B., Michel, L., Tobias, P., Jakob, R., Stefan, C., Janek, T., Theresa, U., Marc, B., Anne-Laure, B., Difan, D., & Marius, L. (2023). Hyperparameter optimization: Foundations, algorithms, best practices, and open challenges. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 13(4), 1474.
- Chang, H., Yang, Z., & Zhang, Y. (2025). Forecasting daily freight flow in cold regions of China using the hybrid Prophet model considering the importance of festivals and epidemic prevention policy. *Research in Transportation Business & Management*, 59, 101294.
- Chunlin, L. (2014). A simplified method for prediction of embankment settlement in clays. *J. R. M. Geo. China*, 6, 61-66.
- Dao, H. H., Nguyen, D. T. (2018). Assessment of the impacts of climate change on groundwater resources in Ca Mau peninsula. *VNUHCM Journal of Engineering and Technology*, 1, 43-52.
- Đinh, X. V., Lê, T. N., & Nguyễn, V. Q. (2018). Xác định biến dạng của công trình xây dựng theo mô hình tham số ứng dụng lọc Kalman và mô hình phi tham số ứng dụng chuỗi thời gian. *Tạp chí Khoa học ĐHQGHN: Các Khoa học Trái đất và Môi trường*, số 34(3), trang 86-99.
- Do, T. M., Nguyen, H. D. Q., Pham, Q. N., Le, D. T., Nguyen, L. Q., Tran, V. A., Gia, T. N. (2024). GNSS Time-Series Prediction Using Moving Average Filter and Multi-Layer Perceptron Neuron Network. *Journal of the Polish Mineral Engineering Society*, 121-131.
- Doãn, H. P., Trần, H. T. (2023). Ứng dụng trí tuệ nhân tạo trong xây dựng bản đồ nguy cơ ngập do nước biển dâng cho thành phố Hồ Chí Minh. *Tạp chí Khí tượng Thủy văn*, số 2, trang 96-107.
- Dongwei, Q., Yuci, T., Yuzheng, W., Keliang, D., Tiancheng, L., & Shanshan, W. (2024). Land subsidence analysis along high-speed railway based on EEMD-Prophet method. *Scientific Reports*, 14(1).
- Drucker, H., Burges, C. J. C., Kaufman, L.; Smola, A; & Vapnik, V. (1997). Support Vector Regression Machines. *Conference: Advances in Neural Inform&on Processing Systems* 9, 155-161.
- Eghrari, Z., Delavar, M. R., Zare, M., Beitollahi, A., & Nazari, B. (2023). Land subsidence susceptibility mapping using machine learning algorithms. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, X-4/W1-2022, 129-136.

- EMSN062 (2019). Assessing changes in ground subsidence rates, Mekong Delta, Vietnam. <https://mapping.emergency.copernicus.eu/activations/EMSN062/>.
- Erban, L., Gorelick, S., & Zebker, H. (2014). Groundwater extraction, land subsidence, and sea-level rise in the Mekong Delta, Vietnam. *Environmental Research Letters*, 9, 084010.
- Hung, W. C., Hwang, C., Tosi, L., Lin, G.-Z., Lin, S.-H., & Chen, Y.-A. . (2024). Near real-time subsidence monitoring and AI forecasting with multi-depth extensometers. *SSRN*.
- Karlsrud, K., Vangelsten, B. V., & Frauenfelder, R. (2017). Subsidence and Shoreline Retreat in the Ca Mau Province - Vietnam. Causes, Consequences and Mitigation Options. *Geotechnical Engineering Journal of the SEAGS & AGSSEA*, 48, 26-32.
- Ku, C.-Y., & Liu, C.-Y. (2023). Modeling of land subsidence using GIS-based artificial neural network in Yunlin County, Taiwan. *Scientific Reports*, 13, 4090.
- Liu, J., Liu, W., Allechy, F., Zheng, Z., Liu, R., & Kouadio, L. (2024). Machine learning-based techniques for land subsidence simulation in an urban area. *Journal of environmental management*, 352, 120078.
- Liu, Z., Ng, A. H.-M., Wang, H., Chen, J., Du, Z., & Ge, L.. (2023). Land subsidence modeling and assessment in the West Pearl River Delta from combined InSAR time series, land use and geological data. *International Journal of Applied Earth Observation and Geoinformation*, 118, 103228.
- Minderhoud, P. (2021). Modelling Mekong Delta Subsidence, Challenges and How to Improve Quantifications. *Conference: 4th Asia Pacific Meeting on Near Surface Geoscience & Engineering*. Nov 2021, Ho Chi Minh, Vietnam, 1 - 7.
- Phạm, M. H., Vũ K. L. (2019). Nghiên cứu ứng dụng phương pháp học máy (machine learning) trong phân loại rừng ngập mặn trên ảnh viễn thám spot6 với khu vực thử nghiệm tại tỉnh Cà Mau. *Tạp chí Khoa học Đo đạc và Bản đồ*, số 40-6/2019, trang 17-21.
- Sekkeravani, A. M., Bazrafshan, O., Pourghasemi, H. R., & Holisaz, A. (2022). Spatial modeling of land subsidence using machine learning models and statistical methods. *Environmental Science and Pollution Research*, 29, 28866-28883.
- Shi, X. W., J., Ye, S., Zhang, Y., Xue, Y., Wei, Z., & Yu, J. (2008). Regional land subsidence simulation in Su-xi-Chang area and Shanghai City, China. *Engineering Geology*, 100(12), 27-42.
- Willmott, C.J. & Matsuura, K. (2005). Advantages of the Mean Absolute Error (MAE) over the Root Mean Square Error (RMSE) in assessing average model performance. *Climate Research*, 30, 79-82.
- Snoek, J., Larochelle, H., & Adams, R. P. (2012). Practical bayesian optimization of machine learning algorithms. *Proceedings of the 26th International Conference on Neural Information Processing Systems*. December 03-06, Lake Tahoe, NV, United States, 2951-2959.
- Taylor, S. J., & Letham, B. (2018). Forecasting at scale. *The American Statistician*, 72, 37-45.
- Tran, A. V., Brovelli, M. A., Ha, K. T., Khuc, D. T., Tran, D. N., Tran, H. H., & Le, N. T. . (2024). Land subsidence susceptibility mapping in Ca Mau Province, Vietnam, using boosting models. *SPRS International Journal of Geo-Information*, 13(5), 161.
- Tran, Q. C., Vu, T. T., Batelaan, O., Pham, Q.-N. . (2024). Factors controlling land subsidence in the Southern Hau River Region, Vietnam.

*Groundwater for Sustainable Development*, 27, 101383.

Vân Anh, T., Hạnh, T. H., Hòa, P. T. T., Anh, T. Q., Vân Anh, N. T., Thu, P. T., & Thọ, T. Q. (2023). Nghiên cứu một số phương pháp học máy trong thành lập bản đồ lớp phủ bề mặt tỉnh Cà Mau trên nền tảng Google Earth Engine. *Tạp chí Khoa học Đo đạc và Bản đồ*, Số 55, trang 18-26.

Wang, H., Jia, C., Ding, P., Feng, K., Yang, X., & Zhu, X. (2023). Analysis and Prediction of Regional Land Subsidence with InSAR Technology and Machine Learning Algorithm. *KSCE Journal of Civil Engineering*, 27, 782-793.