VNU Journal of Science: Earth and Environmental Sciences
Journal homepage: https://js.vnu.edu.vn/EES

Original Article

Application of Machine Learning Algorithms for Forecasting Sea Surface Temperature at Hon Dau Station

Do Thi Thu Ha[1], Nguyen Quoc Trinh[2,*], Nguyen The Duc[1], Nguyen Van Tung[1], Nguyen Thieu Hoa[1]

[1]*Institute of Mechanics, Vietnam Academy of Science and Technology (VAST), 264 Doi Can, Ngoc Ha, Hanoi, Vietnam*

[2]*Graduate University of Science and Technology, VAST, 18 Hoang Quoc Viet, Nghia Do, Hanoi, Vietnam*

Received 18th December 2025
Revised 19th July 2026; Accepted 01st October 2026

Abstract: Sea Surface Temperature (SST) plays a crucial role in climate studies, tropical cyclone prediction, and marine resource management. This study analyzes SST observations at Hon Dau station from 2000–2022 and applies ARIMAX, SARIMAX, Prophet, and LSTM models for SST forecasting. Time series decomposition was conducted to separate trend, seasonal, and residual components to guide model selection. Results indicate that ARIMAX, SARIMAX, and Prophet achieved high accuracy with mean absolute errors (MAE) of 0.82–0.92 °C, while LSTM performed less effectively due to data limitations. For long-term projection, a linear regression model with Fourier harmonics suggests an increase of approximately +1.7 °C in mean annual SST by 2100 relative to 2021, consistent with moderate climate scenarios. These findings highlight the potential of machine learning methods for short-term SST forecasting and emphasize the need for integrating multi-source data with physics-informed deep learning models to improve the reliability of climate and marine applications.

Keywords: Sea Surface Temperature (SST); Machine learning; Short-term forecasting; Climate trend.

* Corresponding author.
E-mail address: nqtrinh@ies.vast.vn

https://doi.org/10.25073/2588-1094/vnuees.5463

Ứng dụng thuật toán học máy cho dự báo nhiệt độ nước biển bề mặt trạm Hòn Dấu

Đỗ Thị Thu Hà[1], Nguyễn Quốc Trinh[2, *], Nguyễn Thế Đức[1], Nguyễn Văn Tùng[1], Nguyễn Thiếu Hoa[1]

[1]*Viện Cơ học, Viện Hàn lâm Khoa học và Công nghệ Việt Nam, 264 Đội Cấn, Ngọc Hà, Hà Nội, Việt Nam*

[2]*Học viện Khoa học và Công nghệ, Viện Hàn lâm Khoa học và Công nghệ Việt Nam, 18 Hoàng Quốc Việt, Nghĩa Đô, Hà Nội, Việt Nam*

Nhận ngày 18 tháng 12 năm 2025
Chỉnh sửa ngày 19 tháng 7 năm 2026; Chấp nhận đăng ngày 01 tháng 10 năm 2026

Tóm tắt: Nhiệt độ nước biển bề mặt (SST) là yếu tố quan trọng trong nghiên cứu khí hậu, dự báo bão và quản lý tài nguyên biển. Nghiên cứu này sử dụng chuỗi số liệu quan trắc SST tại trạm Hòn Dấu giai đoạn 2000–2022 để thử nghiệm dự báo SST bằng một số mô hình học máy gồm ARIMAX, SARIMAX, Prophet và LSTM. Dữ liệu được phân tích thành các thành phần xu thế, mùa vụ và nhiễu để lựa chọn mô hình phù hợp. Kết quả cho thấy các mô hình ARIMAX, SARIMAX và Prophet dự báo khá chính xác với sai số tuyệt đối trung bình (MAE) khoảng 0,82–0,92 °C, trong khi LSTM kém hiệu quả hơn do hạn chế về dữ liệu huấn luyện. Với dự báo dài hạn, mô hình hồi quy tuyến tính kết hợp Fourier cho thấy SST trung bình năm tại Hòn Dấu có thể tăng khoảng +1,7 °C vào năm 2100 so với năm 2021, phù hợp với kịch bản khí hậu trung bình. Kết quả khẳng định tiềm năng của các phương pháp học máy trong dự báo ngắn hạn SST, đồng thời gợi mở hướng kết hợp dữ liệu đa nguồn và mô hình học sâu có ràng buộc vật lý nhằm nâng cao độ tin cậy cho các ứng dụng dự báo và quản lý biển.

Từ khóa: Nhiệt độ nước biển bề mặt; Học máy; Dự báo ngắn hạn; Xu thế khí hậu.

1. Mở đầu

Nhiệt độ nước biển bề mặt (Sea Surface Temperature – SST) là một thông số hải dương học quan trọng, đóng vai trò then chốt trong trao đổi năng lượng, ẩm và các tương tác khác giữa khí quyển và đại dương. Biến động của SST ảnh hưởng trực tiếp đến các quá trình quy mô toàn cầu như dao động ENSO, Dipole Ấn Độ Dương, gió mùa, hoạt động xoáy thuận nhiệt đới và các hiện tượng sóng nhiệt biển [1-3]. SST cũng tương quan mật thiết đến cân bằng năng lượng của Trái đất và biến đổi khí hậu [4, 5]. Do đó, dự báo SST chính xác phục vụ cảnh báo thiên tai, quản lý nguồn lợi hải sản mà cũng đóng góp vào nghiên cứu biến đổi khí hậu toàn cầu [6, 7].

Các phương pháp dự báo SST dựa vào mô hình động lực đại dương – khí quyển [8, 9] và viễn thám [10]. Tuy nhiên, các mô hình này yêu cầu tài nguyên tính toán lớn và xử lý dữ liệu đầu vào phức tạp. Phương pháp thống kê đã được ứng dụng [11] nhưng khả năng mô tả tính phi tuyến còn hạn chế.

Với sự phát triển của trí tuệ nhân tạo và các mô hình học máy (RF - Random Forest, CNN -

* Tác giả liên hệ.
Địa chỉ email: nqtrinh@ies.vast.vn

https://doi.org/10.25073/2588-1094/vnuees.5463

Convolutional Neural Network, LSTM - Long Short-Term Memory, GRU- Gated Recurrent Unit) đã chứng minh tính ưu việt trong dự báo chuỗi thời gian SST [12-14]. LSTM là có khả năng xử lý biến động phi tuyến và tính mùa [15]. Ứng dụng phương pháp hiện đại (CNN, LSTM) đối với SST cho Biển Đông trong dự báo ngắn hạn cũng được thể hiện [16, 17].

Các công trình nghiên cứu nhấn mạnh vai trò của vật lý - trí tuệ nhân tạo (PINNs) và học sâu - vật lý (PGTransNet) đã được trình bày [18-20]. Bên cạnh đó, tiềm năng của AI và học sâu trong dự báo hải dương cũng được thể hiện [21, 22].

Đối với khu vực Biển Đông cũng có đặc trưng hải dương phức tạp của gió mùa, ENSO và các hệ thống thống hoàn lưu [23]. SST của khu vực này biến động mạnh như mùa đông 24–26°C, mùa hè 29–31°C [24]. Hơn nữa, khu vực này cũng có hiện tượng nước trồi điển hình [25], ENSO tác động rõ rệt [26]. Do đó, xu thế tăng SST nhiều thập kỷ qua đã được xác nhận với mức tăng 0,2–0,3°C mỗi thập kỷ [27, 28]. Tuy nhiên, hạn chế lớn vẫn là thiếu dữ liệu quan trắc tại chỗ và sự gián đoạn dữ liệu vệ tinh [29].

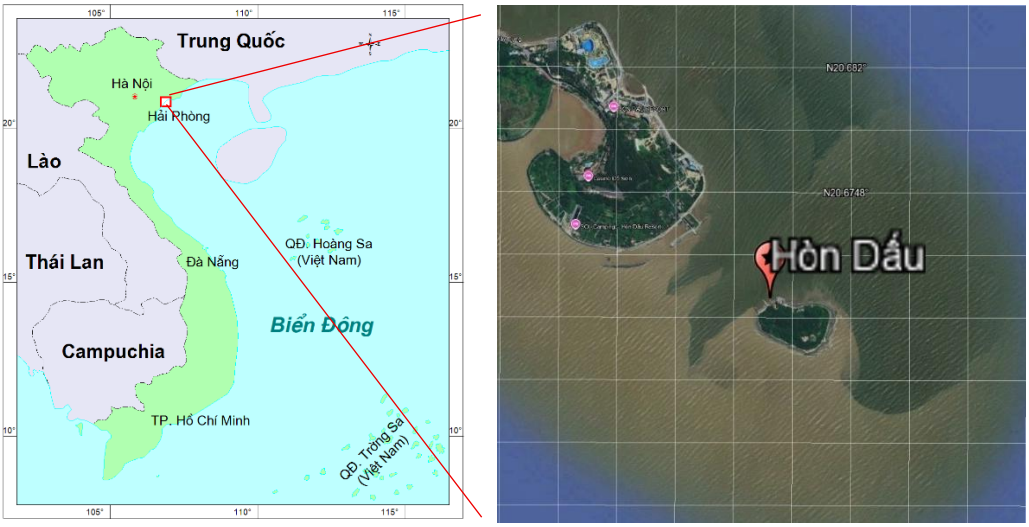
Tuy nhiên, nghiên cứu tại Việt Nam vẫn chủ yếu dựa vào phương pháp truyền thống, trong khi ứng dụng học máy - đặc biệt ở quy mô trạm - còn hạn chế. Đây chính là khoảng trống nghiên cứu cần được làm rõ.

Nghiên cứu này áp dụng các mô hình ARIMAX (AutoRegressive Integrated Moving Average with Exogenous variables), SARIMAX (Seasonal Autoregressive Integrated Moving Average with eXogenous regressors), Prophet và LSTM (Long Short-Term Memory) để dự báo SST tại trạm Hòn Dấu giai đoạn 2000–2022, nhằm i) Đánh giá khả năng dự báo ngắn hạn (7–30 ngày); ii) Phân tích xu thế dài hạn; và iii) Kiểm chứng tiềm năng của học máy trong điều kiện dữ liệu thực tế tại Việt Nam. Kết quả kỳ vọng cung cấp bằng chứng khoa học cho việc ứng dụng AI trong dự báo hải dương nghiệp vụ.

2. Dữ liệu và phương pháp nghiên cứu

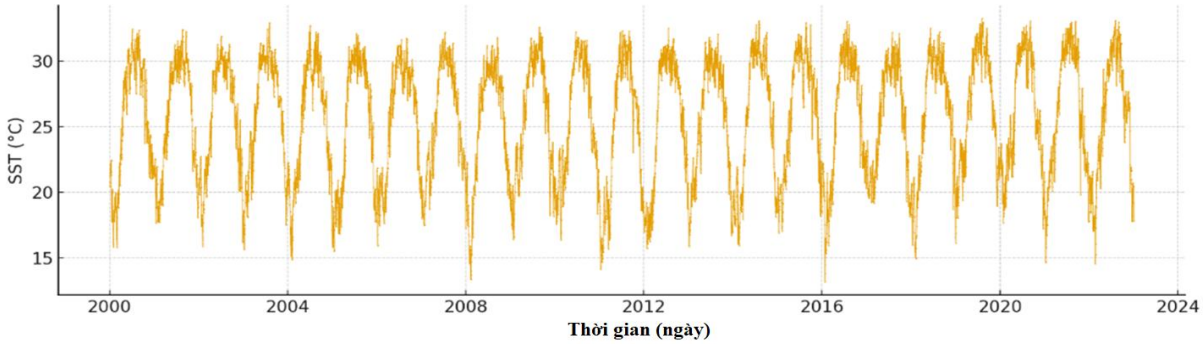
2.1. Dữ liệu sử dụng

Nghiên cứu sử dụng chuỗi số liệu nhiệt độ nước biển bề mặt thu thập được tại trạm hải văn Hòn Dấu giai đoạn 2000–2022 (Hình 1). Trạm hải văn Hòn Dấu có vị trí $20^0$40'N-$106^0$48'E tại đảo Hòn Dấu, phường Đồ Sơn, thành phố Hải Phòng. Trạm được thành lập năm 1928, đảm nhận quan trắc các yếu tố khí tượng và hải văn (mực nước, sóng, độ mặn, SST, …). Dữ liệu SST từ trạm Hòn Dấu được thu thập theo tần suất ngày, được kiểm định và hiệu chỉnh theo quy trình chất lượng của ngành khí tượng thủy văn [28].



Hình 1. Khu vực nghiên cứu.

Chuỗi số liệu SST được xử lý và phân tích theo mô hình dự báo với hạn 7 ngày, 15 ngày, 30 ngày và xu thế nhiều năm. Định dạng chuỗi số liệu của SST là thời gian (năm, tháng, ngày, giờ) và giá trị theo đơn vị ºC. Tần suất số liệu là theo 04 obs/ngày (04 lần đo/ngày) vào thời điểm theo quy định của tổ chức khí tượng thế giới tương ứng giờ của Việt Nam (lúc 01, 07, 13, 19 giờ). Khoảng thời gian thu thập số liệu từ năm 2000 đến năm 2022 tương đương với chuỗi dài hơn 20 năm. Sau khi thu thập và xử lý, SST được chuyển về dạng số liệu trung bình ngày (Hình 2).



Hình 2. Biến trình chuỗi số liệu SST (ºC) trung bình ngày giai đoạn 2000–2022.

2.2. Phương pháp nghiên cứu

Để bảo đảm tính khách quan và khả năng so sánh, tất cả các mô hình được xây dựng trên cùng bộ dữ liệu đầu vào, áp dụng thống nhất quy trình tiền xử lý, phương pháp nghiên cứu và hệ thống chỉ số đánh giá. Sự khác biệt giữa các mô hình chỉ nằm ở phương pháp xây dựng và quy trình lựa chọn cấu hình tối ưu, nhờ đó kết quả so sánh phản ánh đúng năng lực dự báo của từng mô hình.

i) Tiền xử lý dữ liệu: Dữ liệu được kiểm soát chặt chẽ, loại bỏ giá trị bất thường và nội suy tuyến tính với các khoảng thiếu nhỏ hơn 3 ngày. Xu thế dài hạn được xác định bằng phương pháp hồi quy tuyến tính [28].

ii) Xây dựng mô hình dự báo: Bốn mô hình được lựa chọn như sau:

Mô hình ARIMAX là một dạng kết hợp các thành phần tự hồi quy, tích hợp và trung bình trượt với các biến SST để dự báo chuỗi thời gian [11];

Mô hình SARIMAX là mô hình dự báo chuỗi thời gian mạnh dựa vào tích hợp các thành phần mùa và SST [11];

Mô hình Prophet là mô hình dự báo theo chuỗi thời gian, được phát triển bởi Facebook để dự đoán xu hướng trong tương lai có tính mùa và xu hướng rõ ràng;

Mô hình LSTM là mô hình học sâu cho chuỗi thời gian, có khả năng mô tả quan hệ phi tuyến và tính chất chu kỳ [13, 14].

iii) Quy trình nghiên cứu: gồm bảy bước: a) Thu thập và kiểm tra chất lượng chuỗi số liệu nhiệt độ nước biển (SST) tại trạm Hòn Dấu; b) Tiền xử lý và xây dựng chuỗi SST trung bình ngày; c) Chia dữ liệu thành các tập huấn luyện, kiểm định và kiểm tra; d) Xây dựng năm mô hình dự báo gồm Persistence, ARIMAX, SARIMAX, Prophet và LSTM; e) Lựa chọn cấu hình tối ưu cho từng mô hình; f) Thực hiện dự báo trên tập kiểm tra; và g) đánh giá, so sánh hiệu năng dự báo bằng cùng một hệ thống chỉ số thống kê.

Đối với các mô hình ARIMAX và SARIMAX, các tham số được xác định theo quy trình dò tìm tự động kết hợp với phân tích (hàm tương quan, hàm tương quan riêng phần và các tiêu chí). Trong quá trình hiệu chỉnh, nhiều cấu hình mô hình khác nhau được thử nghiệm tự động trong các khoảng tham số xác định trước nhằm lựa chọn mô hình có hiệu năng tốt nhất.

Đối với mô hình Prophet, tham số xu thế và mùa tự điều chỉnh thông qua quá trình thử nghiệm lặp đã được thử nghiệm với nhiều giá trị khác nhau.

Đối với mô hình LSTM, các tham số (số neuron của lớp LSTM, Kịch thước, số epoch và tốc độ học) của mô hình được thử nghiệm với nhiều cấu hình khác nhau. Dữ liệu đầu vào cho mô hình LSTM được xây dựng theo thời gian gồm n bước trước đó để dự báo giá trị SST kế tiếp.

iv) Tiêu chí lựa chọn các mô hình: Mỗi mô hình được lựa chọn theo tiêu chí tối ưu phù hợp với đặc điểm của từng phương pháp, đồng thời sử dụng thống nhất tập kiểm định để bảo đảm tính khách quan trước khi đánh giá trên tập kiểm tra. Đối với ARIMAX và SARIMAX, cấu hình tối ưu được xác định theo giá trị AIC (Akaike Information Criterion) nhỏ nhất. Đối với Prophet và LSTM, cấu hình tối ưu được lựa chọn dựa trên kết quả đánh giá trên tập kiểm định, ưu tiên mô hình có sai số dự báo thấp và khả năng tổng quát hóa tốt. Mô hình Persistence không yêu cầu hiệu chỉnh tham số và được sử dụng làm mô hình chuẩn để đánh giá mức độ cải thiện của các mô hình còn lại.

v) Đánh giá hiệu năng các mô hình: Các mô hình trên được đánh giá theo chuỗi số liệu SST về hiêu chỉnh và kiểm định. Sai số được đánh giá bằng phương pháp sai số bình phương trung bình (RMSE) và sai số tuyệt đối trung bình (MAE) [17].

vi) Dự báo dài hạn: Xu thế SST đến năm 2100 được ngoại suy bằng các mô hình, đồng thời so sánh với kịch bản biến đổi khí hậu toàn cầu (CMIP6) để kiểm chứng tính hợp lý [30].

Phương pháp tiếp cận này cho phép đánh giá đồng thời hiệu quả dự báo ngắn hạn của học máy và xu thế dài hạn của SST trong bối cảnh biến đổi khí hậu. Mô hình dự báo SST học máy kết hợp của bốn mô hình nêu trên được thể hiện trong công thức sau:

$$SST_t = A_0 + A_1 t + \sum_{k=1}^{K} \left(a_k \sin\left(\frac{2\pi kt}{365}\right) + b_k \cos\left(\frac{2\pi kt}{365}\right)\right) + \varepsilon_t \quad (1)$$

Trong đó: A_0 là giá trị trung bình; A_1là tham số biến đổi tuyến tính theo thời gian; a_k và b_k là các biên độ dao động theo theo thời gian; t là thời gian (ngày), k=1, 2, …, K (thường chọn K = 3 là đủ để bắt dạng sóng năm).

Hạn dự báo thường được thể hiện ở các bản tin dự báo biển là ngắn hạn (1-3 ngày, 7-10 ngày, 15-30 ngày) và dài hạn (lớn hơn 30 ngày, mùa, năm, nhiều năm).

Chuỗi số liệu SST được chia thành ba tập dữ liệu gồm: tập huấn luyện, tập kiểm định, và tập kiểm tra. Trong nghiên cứu này, dữ liệu giai đoạn 2000–2020 được sử dụng để huấn luyện mô hình. Khi dữ liệu năm 2021 được sử dụng cho kiểm định và đánh giá khả năng dự báo. Việc áp dụng cùng một cách chia dữ liệu cho tất cả các mô hình nhằm bảo đảm tính nhất quán và khách quan trong quá trình so sánh hiệu năng dự báo.

3. Kết quả và thảo luận

3.1. Biến động SST quan trắc tại Hòn Dấu giai đoạn 2000–2022

Kết quả phân tích chuỗi số liệu tại trạm Hòn Dấu cho thấy SST có dao động mùa rõ rệt: thấp nhất vào tháng 1–2 (trung bình 16–18°C), cao nhất vào tháng 7–8 (trung bình 28–30°C). Xu thế tăng nhiệt độ trung bình năm khoảng 0,25°C/thập kỷ, phù hợp với xu thế chung ở Biển Đông [31] và toàn cầu [5]. So sánh với các nghiên cứu cho thấy tốc độ tăng SST tại Hòn Dấu tương đồng với nhiều khu vực ven bờ Việt Nam [27, 28].

Ngoài ra, các dị thường SST có liên quan chặt chẽ đến ENSO. Trong các năm El Niño mạnh (2009–2010, 2015–2016), SST cao hơn trung bình 0,8–1,2 °C; trong khi các năm La Niña (2007–2008, 2010–2011) SST thấp hơn trung bình 0,6–0,9°C. Kết quả này phù hợp với nhận định về ảnh hưởng của ENSO đến khu vực Nam Trung Bộ và Bắc Biển Đông [23, 25].

3.2. Đánh giá hiệu quả mô hình dự báo ngắn hạn

Mô hình dự báo SST đã chạy trên chuỗi số liệu mẫu từ ngày 01/01/2020 đến 31/12/2021 và kiểm tra trên chuỗi 01/01/2022 đến 31/12/2022). Chuỗi số liệu mẫu này có tổng số 8401 mẫu. Trong đó, số liệu được sử dụng 8036 mẫu, số liệu kiểm tra 365 mẫu và số liệu mẫu lỗi bị loại bỏ. Kết quả đã được đưa ra đánh giá sai số MAE và

RMSE khi dự báo cho các chuỗi 7, 15, 30 ngày (Bảng 1) và biến trình SST được so sánh gồm 3 đường (tương ứng hạn dự báo cho 7, 15, 30 ngày) hiển thị chuỗi quan trắc trong tập kiểm tra và các điểm dự báo được đánh dấu trên ngày dự báo thực tế (Hình 3).

Bảng 1. Đánh giá sai số dự báo bằng mô hình dự báo SST trên toàn tập kiểm tra

Hạn dự báo	MAE (°C)	RMSE (°C)
7 ngày	1,54	1,96
15 ngày	1,77	2,27
30 ngày	2,69	3,35

Tiếp theo, kết quả dự báo SST theo các mô hình dự báo (ARIMAX, SARIMAX, Prophet, LSTM) và so sánh với số liệu quan trắc tại trạm Hòn Dấu cho hạn dự báo 7 ngày. Kết quả đánh giá sai số (MAE, RMSE) để xác định mối quan hệ giữa mô hình tính toán và thực đo (Bảng 2).

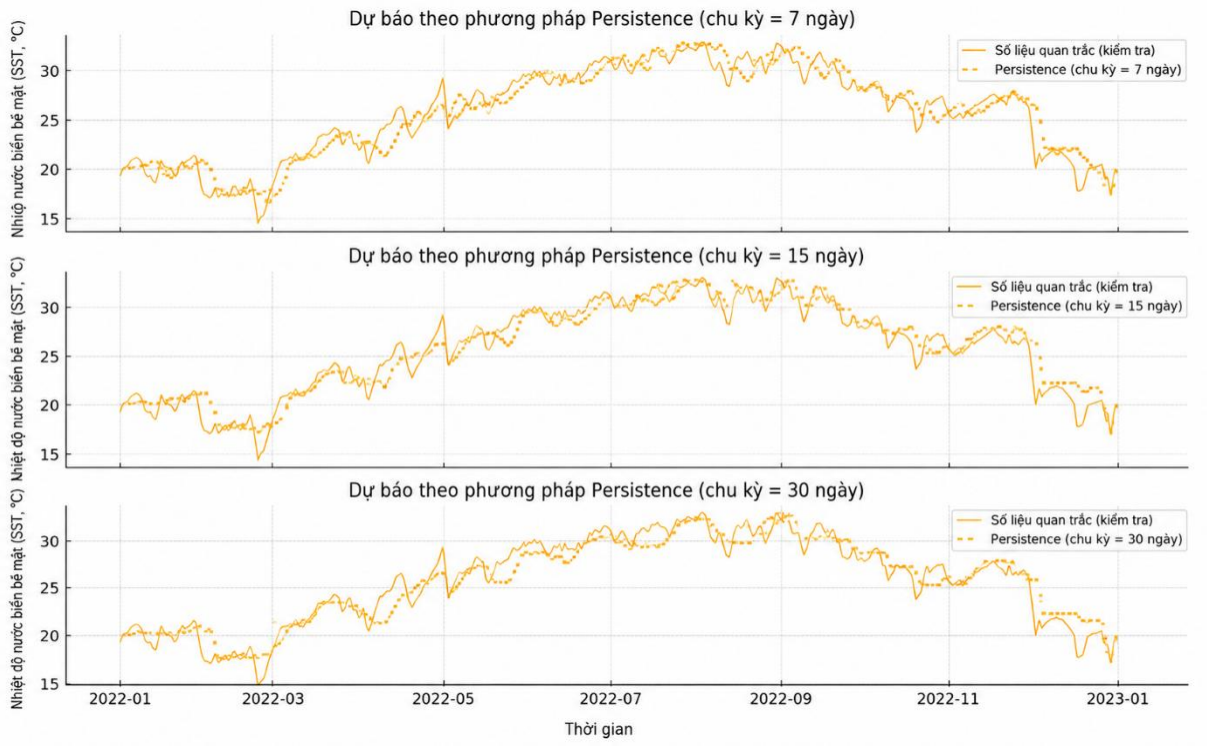
Bảng 2. Đánh giá sai số giữa các mô hình và thực đo

Mô hình	MAE (°C)	RMSE (°C)
ARIMAX	0,824	1,056
SARIMAX	0,874	1,064
Prophet	0,920	1,161
LSTM	2,323	2,342

Kết quả so sánh các mô hình cho thấy:

- Các mô hình ARIMAX và SARIMAX dự báo có độ chính xác đảm bảo, sai số MAE khoảng 0,824 - 0,874°C và sai số RMSE khoảng 1,056 - 1,064°C. Các mô hình này thể hiện tính mùa khá tốt.

- Mô hình Prophet thể hiện ưu thế trong việc tách xu thế – mùa vụ, sai số MAE 0,920°C và sai số RMSE 1,161°C, song vẫn chưa phản ánh chính xác các dao động ENSO. Mô hình này phản ánh độ chính xác chấp nhận được.



Hình 3. Biến trình SST (°C) so sánh giữa số liệu quan trắc và kết quả mô hình dự báo (Persistence) cho hạn dự báo (a. 7 ngày (trên); b. 15 ngày (giữa); c. 30 ngày (dưới)).

- Mô hình LSTM cho thấy sai lệch đáng kể với sai số RMSE 2,342 °C và MAE 2,323 °C. Mô hình này định hướng dài hạn dao động mùa và ENSO. Kết quả này phù hợp với các nghiên cứu quốc tế về ưu thế của LSTM trong dự báo SST [11, 13, 14].

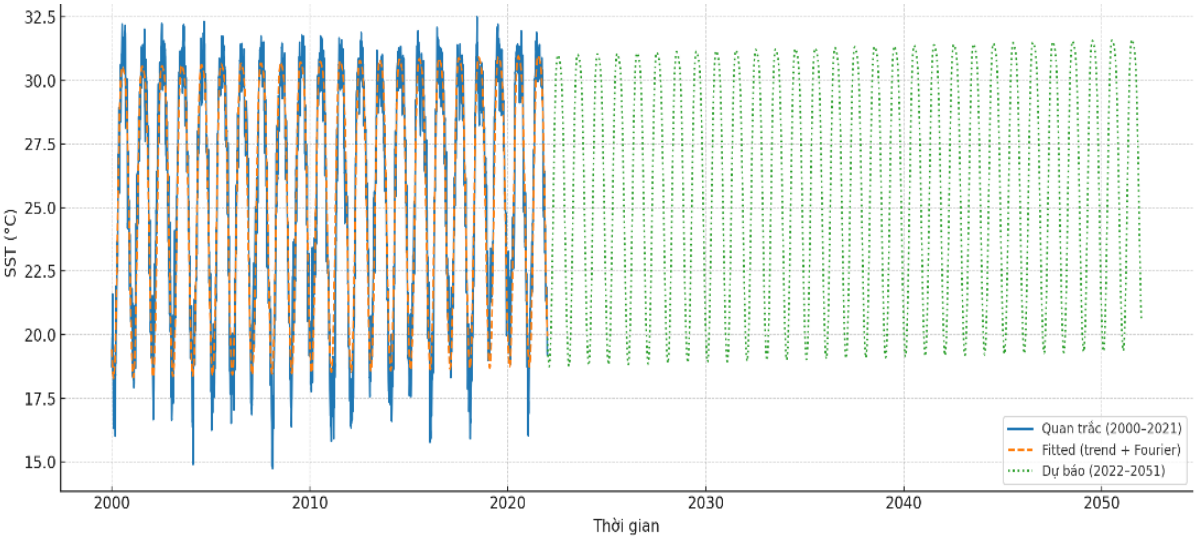
Khi áp dụng LSTM cho Biển Đông cũng ghi nhận sai số dự báo ngắn hạn khoảng 0,6–0,9 °C, gần với kết quả tại Hòn Dấu [16, 17]. Điều này khẳng định tiềm năng của học sâu trong cải thiện dự báo SST khu vực.

3.3. Dự báo xu thế SST dài hạn đến năm 2100

Mô hình dự báo xu thế đã được thực hiện ứng với K=3 (chu kỳ năm) sử dụng chuỗi số liệu SST tại trạm Hòn Dấu giai đoạn 2000–2021 để phân tích, sau đó tái phân tích và dự báo giai 2000-2050 (Hình 4). Kết quả đưa ra tính mức thay đổi nhiệt độ so với năm 2021 tại các mốc 2050, 2080 và 2100 (Bảng 3).

Bảng 3. Kết quả dự báo SST tại trạm Hòn Dấu

Năm	SST trung bình (°C)	ΔSST so với năm 2021 (°C)
2021	25,75	0,00
2050	26,38	+0,63
2080	26,97	+1,22
2100	27,41	+1,66



Hình 4. Đồ thị biến thiên SST, xu thế và dự báo SST đến 2050.

Bảng 3 cho thấy kết quả dự báo SST so với mốc năm 2021 đến năm 2050 (sau khoảng 30 năm), 2080 và 2100 (cuối thế kỷ) xu thế tăng lên tương ứng 0,63, 1,22 và 1,66°C. Trong khi đó, một số công trình công bố có thể tăng lên khoảng 1,8–2,3°C vào cuối thế kỷ XXI so với trung bình 2000–2020 [24, 29, 30]. Kịch bản này phù hợp với xu thế tăng SST toàn cầu theo CMIP6 [30] và kết quả nghiên cứu ở khu vực biển ven bờ Việt Nam [24, 29].

Sự gia tăng SST đã được thể hiện qua nhiều hiện tượng khác nhau như: i) Tăng nguy cơ sóng nhiệt biển đã được ghi nhận ở Địa Trung Hải [4] và Nam Thái Bình Dương [6]; ii) Anh hưởng đến nghề cá và hệ sinh thái biển ven bờ Việt Nam, vốn đã nhạy cảm với biến đổi khí hậu [31]; iii) Làm thay đổi các đặc trưng khí tượng – hải dương cực đoan, đặc biệt là bão nhiệt đới [32].

3.4. Thảo luận

Kết quả nghiên cứu cho thấy hiệu năng dự báo SST có sự khác biệt rõ rệt giữa các mô hình. ARIMAX và SARIMAX đạt độ chính xác cao nhất với MAE = 0,824–0,874 °C và RMSE =

1,056–1,064 °C. Prophet cho kết quả ở mức chấp nhận được (MAE = 0,920 °C; RMSE = 1,161 °C), trong khi LSTM có sai số lớn nhất (MAE = 2,323 °C; RMSE = 2,342 °C), cao gấp khoảng 2 lần so với ARIMAX và SARIMAX.

Kết quả cho thấy đối với chuỗi SST đơn trạm có đặc trưng mùa vụ và tự tương quan mạnh, các mô hình thống kê như ARIMAX và SARIMAX phù hợp hơn mô hình học sâu LSTM. Mặc dù LSTM có khả năng mô hình hóa quan hệ phi tuyến, hiệu quả của mô hình phụ thuộc nhiều vào quy mô và chất lượng dữ liệu huấn luyện. Với bộ dữ liệu của nghiên cứu này, LSTM chưa phát huy được ưu thế và không vượt trội so với các mô hình thống kê.

Mặc dù chênh lệch sai số giữa ARIMAX, SARIMAX và Prophet không lớn, việc giảm sai số dự báo vẫn có ý nghĩa thực tiễn đối với giám sát môi trường biển, cảnh báo sớm và hỗ trợ quản lý vùng ven bờ. Kết quả nghiên cứu khẳng định ARIMAX và SARIMAX là hai mô hình phù hợp nhất để dự báo SST tại trạm Hòn Dấu trong điều kiện dữ liệu hiện có.

Dữ liệu SST tại Hòn Dấu thể hiện biến động mùa rõ rệt, dao động trung bình từ khoảng 16–18°C vào mùa đông đến 28–30°C vào mùa hè. Xu thế tăng dài hạn khoảng 0,25 °C/thập kỷ tương đồng với xu thế ghi nhận ở Biển Đông [27, 28] và trên phạm vi toàn cầu [5]. ENSO cũng cho thấy ảnh hưởng rõ rệt, với các pha El Niño làm SST tăng cao và La Niña gây giảm SST bất thường [23, 25]. Tuy nhiên, dữ liệu quan trắc thực đo còn hạn chế về chiều dài chuỗi và tính liên tục. Vì vậy, việc tích hợp dữ liệu vệ tinh, tái phân tích và dữ liệu mô hình là cần thiết nhằm mở rộng tập dữ liệu phục vụ các phương pháp học máy và học sâu [8, 10, 24].

Mô hình xu thế cho thấy SST tại Hòn Dấu có thể tăng khoảng +1,7°C vào cuối thế kỷ XXI, tương đồng với các kịch bản khí hậu CMIP6 [30]. Xu thế này có thể làm gia tăng nguy cơ sóng nhiệt biển, ảnh hưởng đến hệ sinh thái biển, nguồn lợi thủy sản ven bờ và các hoạt động kinh tế biển, đồng thời góp phần thay đổi đặc trưng khí hậu cực đoan tại khu vực [4, 6]. Trong tương lai, việc phát triển các mô hình học sâu có tích hợp chặt chẽ với vật lý tự nhiên [18, 20] có thể nâng cao độ tin cậy, khả năng giải thích và giảm hạn chế của mô hình "hộp đen". Đây được xem là hướng nghiên cứu tiềm năng cần tiếp tục mở rộng tại Việt Nam.

4. Kết luận

Dự báo nhiệt độ mặt biển (SST) ở Biển Đông đã đạt được nhiều tiến bộ nhờ sự phát triển của dữ liệu vệ tinh và các phương pháp học máy, tuy nhiên vẫn còn nhiều thách thức liên quan đến chất lượng dữ liệu, cấu trúc mô hình và khả năng ứng dụng thực tiễn. Để nâng cao độ chính xác và giá trị ứng dụng của sản phẩm dự báo SST, cần phát triển đồng bộ trên ba phương diện chính gồm dữ liệu, mô hình và triển khai ứng dụng.

Kết quả nghiên cứu cho thấy các mô hình ARIMAX, SARIMAX và Prophet đạt hiệu quả dự báo tương đối cao với sai số tuyệt đối trung bình (MAE) khoảng 0,82–0,92°C, đồng thời cải thiện rõ rệt so với persistence model. Trong khi đó, mô hình LSTM cho hiệu quả thấp hơn do hạn chế về dữ liệu huấn luyện và chiều dài chuỗi thời gian. Kết quả này cho thấy các mô hình học sâu không phải lúc nào cũng vượt trội hơn các phương pháp thống kê truyền thống đối với chuỗi số liệu SST đơn trạm có kích thước hạn chế.

Đối với dự báo dài hạn, mô hình hồi quy tuyến tính kết hợp Fourier cho thấy SST trung bình năm tại Hòn Dấu có thể tăng khoảng +1,7°C vào năm 2100 so với năm 2021, phù hợp với các kịch bản biến đổi khí hậu trung bình. Kết quả nghiên cứu khẳng định tiềm năng của các phương pháp học máy trong dự báo ngắn hạn SST, đồng thời gợi mở hướng kết hợp dữ liệu đa nguồn và các mô hình học sâu có ràng buộc vật lý nhằm nâng cao độ tin cậy cho các ứng dụng dự báo và quản lý biển trong tương lai.

Lời cảm ơn

Nhóm tác giả xin được gửi lời cảm ơn Cục Khí tượng Thủy văn đã hỗ trợ, cung cấp cơ sở dữ liệu để thực hiện bài báo này.

Tài liệu tham khảo

[1] N. H. Saji et al., A Dipole Mode in the Tropical Indian Ocean, Nature, Vol. 401, 1999, pp. 360-363, https://doi.org/10.1038/43854.

[2] R. A. Betts et al., El Niño and a Record CO_2 Rise, Nature Climate Change, Vol. 6, No. 9, 2016, pp. 806-810, https://doi.org/10.1038/nclimate3063.

[3] C. Deser et al., Sea Surface Temperature Variability: Patterns and Mechanisms, Annual Review of Marine Science, Vol. 2, 2010, pp. 115-143, https://doi.org/10.1146/annurev-marine-120408-151453.

[4] D. Baird et al., Ecosystem Response to Increasing Ambient Water Temperatures due to Climate Warming in the Sylt- Rømø Bight, Northern Wadden Sea, Germany, Estuarine, Coastal and Shelf Science, Vol. 228, 2019, pp. 106-322, https://doi.org/10.1016/j.ecss.2019.106322.

[5] P. M. Forster et al., Indicators of Global Climate Change 2023: Annual Update of Key Indicators of the State of the Climate System and Human Influence, 2024, https://doi.org/10.5194/essd-16-2625-2024.

[6] G. Bonino et al., Machine Learning Methods to Predict Sea Surface Temperature and Marine Heatwave Occurrence: A Case Study of the Mediterranean Sea, Ocean Science, Vol. 20, No. 2, 2024, pp. 417-432, https://doi.org/10.5194/os-20-417-2024.

[7] P. Heimbach et al., Crafting the Future: Machine Learning for Ocean Forecasting, 2024, https://doi.org/10.5194/sp-2024-18.

[8] J. Wallcraft et al., Software Design Description for the HYbrid Coordinate Ocean Model (HYCOM), Version 2.2, 2009.

[9] V. Alari et al., Surface Wave Effects on Water Temperature in the Baltic Sea: Simulations with the Coupled NEMO-WAM Model, Ocean Dynamics, Vol. 66, No. 8, 2016, pp. 917-930, https://doi.org/10.1007/s10236-016-0963-x.

[10] B. Huang et al., Improvements of the Daily Optimum Interpolation Sea Surface Temperature (DOISST) version 2.1, Journal of Climate, Vol. 34, No. 8, 2021, pp. 2923-2939, https://doi.org/10.1175/JCLI-D-20-0166.1.

[11] S. Wolff et al., Statistical and Machine Learning Ensemble Modelling to Forecast Sea Surface Temperature, Journal of Marine Systems, Vol. 208, 2020, 103347, https://doi.org/10.1016/j.jmarsys.2020.103347.

[12] L. Breiman, Random Forests, Machine Learning, Vol. 45, 2001, pp. 5-32, https://doi.org/10.1023/A:1010933404324.

[13] Q. Zhang et al., Prediction of Sea Surface Temperature Using Long Short-Term Memory, IEEE Geoscience and Remote Sensing Letters, Vol. 14, No. 10, 2017, pp. 1745-1749.

[14] Y. Fu et al., Prediction and Analysis of Sea Surface Temperature Based on LSTM-Transformer Model, Regional Studies in Marine Science, Vol. 78, 2004, 103726, https://doi.org/10.1016/j.rsma.2024.103726.

[15] G. E. Hinton, Deep Learning - A Technology with the Potential to Transform AI, Communications of the ACM, Vol. 61, No. 6, 2018, pp. 53-59, https://doi.org/10.1145/3211562.

[16] P. Hao et al., Prediction of Sea Surface Temperature in the South China Sea Based on Deep Learning, Remote Sensing, Vol. 15, No. 6, 2023, https://doi.org/10.3390/rs15061656.

[17] Y. Miao et al., A Multivariable Convolutional Neural Network for Forecasting Synoptic-Scale Sea Surface Temperature Anomalies in the South China Sea, Weather and Forecasting, Vol. 38, No. 6, 2023, pp. 849-863, https://doi.org/10.1175/WAF-D-22-0094.1.

[18] S. Cuomo et al., Scientific Machine Learning through Physics-Informed Neural Networks: Where We Are and What's Next, 2022, http://arxiv.org/abs/2201.05624.

[19] T. Yuan et al., A Space-Time Partial Differential Equation Based Physics-Guided Neural Network for Sea Surface Temperature Prediction, Remote Sensing, Vol. 15, No. 14, 2023, https://doi.org/10.3390/rs15143498.

[20] S. Wu et al., PGTransNet: A Physics-Guided Transformer Network for 3D Ocean Temperature and Salinity Predicting in Tropical Pacific, Frontiers in Marine Science, Vol. 11, 2024, https://doi.org/10.3389/fmars.2024.1477710.

[21] Y. Wang et al., Impact of Physical and Attention Mechanisms on U-Net for SST Forecasting, Intelligent Marine Technology and Systems, Vol. 2, No. 1, 2024, https://doi.org/10.1007/s44295-024-00025-4.

[22] X. Zhao et al., Deep Learning for Ocean Temperature Forecasting: A Survey, Intelligent Marine Technology and Systems, Vol. 2, No. 28, 2024, https://doi.org/10.1007/s44295-024-00042-3.

[23] B. Wang et al., Multi - Scale Climate Variability of the South China Sea Monsoon: A Review, Dynamics of Atmospheres and Oceans, Vol. 47,

No. 1-3, 2009, pp. 15-37, https://doi.org/10.1016/j.dynatmoce.2008.09.004.

[24] N. T. Hai et al., Analysis of Sea Surface Temperature Using MODIS Imagery, Journal of Hydro - Meteorology, Vol. 625, 2013, pp. 32-37 (in Vietnamese).

[25] V. V. Tac et al., Sea Surface Temperature Anomaly in South Central Vietnam Waters Related to ENSO Phenomenon, Vietnam Journal of Marine Science and Technology, Vol. 17, No. 2, 2017, pp. 111-120, https://doi.org/10.15625/1859-3097/17/2/10153.

[26] N. T. Thanh et al., The Impact of Satellite Sea Surface Temperature on Intensity and Track of Tropical Cyclones over the Vietnam East Sea by Using WRF Model, Journal of Hydro - Meteorology, Vol. 676, 2017, pp. 47-55 (in Vietnamese).

[27] N. X. Hien et al., Trends in Sea Surface Temperature Changes in the East Vietnam Sea, Journal of Hydro - Meteorology, Vol. 639, 2014, pp. 5-9 (in Vietnamese).

[28] N. Q. Trinh et al., Assessment of Sea surface Temperature in the Vietnam Sea for the Period 2000-2022 According to Data from the HYCOM Model, Journal of Hydro - Meteorology, Vol. 761, 2024, pp. 70-81 (in Vietnamese).

[29] N. N. Tuan et al., Research of Rain and Air Temperature in the Gulf of Tonkin from Satellite Data, Journal of Hydro - Meteorology, Vol. 755, 2023, pp. 19-28 (in Vietnamese).

[30] K. E. Trenberth et al., Earth's Energy Imbalance, Journal of Climate, Vol. 27, No. 9, 2014, pp. 3129-3144, https://doi.org/10.1175/JCLI-D-13-00294.1.

[31] N. V. Huong et al., Study on Sea Surface Temperateture front Distribution in Viet Nam Waters for Forcast of Fishing Ground, Journal of Hydro - Meteorology, Vol. 719, 2020, pp. 66-75 (in Vietnamese).

[32] R. A. Betts et al., El Niño and a Record CO_2 Rise, Nature Climate Change, Vol. 6, No. 9, 2016, pp. 806-810, https://doi.org/10.1038/nclimate3063.