

propose

January 2026

1 Proposed Study

1.1 Bối cảnh và vấn đề nghiên cứu

Do đó, nghiên cứu này đề xuất một khung phương pháp tích hợp theo hướng:

- Xây dựng **bản đồ mức độ phơi nhiễm mặn** (salinity exposure) như một lớp **tính** phản ánh điều kiện nền địa chất–môi trường;
- Mô hình hóa **động lực xâm nhập mặn theo thời gian** (salinity state dynamics) bằng mô hình học sâu chuỗi thời gian, có điều kiện bởi lớp phơi nhiễm nền;
- Phân tích liên hệ giữa **phơi nhiễm mặn** và **chất lượng rừng ven biển**, đồng thời xác định các vùng điểm nóng (hotspots) phục vụ quản lý.

1.2 Khu vực nghiên cứu

Khu vực nghiên cứu là dải ven biển dài trên 50 km thuộc 08 xã ven biển của tỉnh Hưng Yên, gồm Đông Thụy Anh, Thái Thụy, Thái Ninh, Đông Thái Ninh, Đông Tiền Hải, Đông Châu, Nam Cường và Hưng Phú. Khu vực này nằm trong vùng cửa sông ven biển của hệ thống sông Hồng với năm cửa sông chính: Thái Bình, Diêm Hộ, Trà Lý, Lân và Ba Lạt. Trong đó, sông Trà Lý chảy qua trung tâm vùng ven biển, đóng vai trò ranh giới tự nhiên giữa hai tiểu vùng: nhóm xã phía Đông Bắc (Đông Thụy Anh, Thái Thụy, Thái Ninh, Đông Thái Ninh) và nhóm xã phía Đông Nam (Đông Tiền Hải, Đông Châu, Nam Cường và Hưng Phú).

1.3 Mục tiêu và câu hỏi nghiên cứu

1. Xây dựng bản đồ phơi nhiễm mặn ở độ phân giải 30 m dựa trên các yếu tố môi trường nền (địa hình, kết nối thủy văn, thổ nhưỡng, khí hậu nền).
2. Xây dựng bản đồ chất lượng rừng ven biển ở độ phân giải 30 m dựa trên bản đồ phân bố rừng và các chỉ số thực vật từ ảnh viễn thám.
3. Mô hình hóa động lực không gian–thời gian của trạng thái mặn và dự báo bản đồ mặn cho các mốc thời gian tương lai bằng mô hình ConvLSTM.

1.4 Phương pháp: tách lớp tĩnh và lớp động

Nghiên cứu tách biệt rõ:

- **Phơi nhiễm mạn (Exposure):** lớp nền không gian *quasi-static*, đại diện cho khả năng/tiềm năng chịu tác động mạn mang tính cấu trúc; không gắn với một năm cụ thể.
- **Trạng thái mạn (State):** lớp biến động theo thời gian, phản ánh mức độ mạn tại các thời điểm/năm khác nhau; chịu chi phối bởi các điều kiện khí tượng-thủy văn theo năm và các quá trình động lực.

Theo đó, mô hình học máy (RF/XGBoost) được sử dụng để lập bản đồ phơi nhiễm mạn tĩnh; còn mô hình ConvLSTM được sử dụng để học động lực trạng thái mạn theo thời gian, với bản đồ phơi nhiễm như một kênh nền tĩnh.

1.5 Dữ liệu và tiền xử lý

1.5.1 Nguyên tắc chuẩn hóa dữ liệu

Tất cả các lớp dữ liệu được:

- Chuẩn hóa về cùng hệ tọa độ và phạm vi vùng nghiên cứu;
- Tái lấy mẫu về độ phân giải không gian 30 m (biến liên tục: song tuyến tính; biến phân loại: lân cận gần nhất);
- Cắt theo mặt nạ vùng nghiên cứu và kiểm soát giá trị ngoại lai/khuyết dữ liệu.

1.5.2 Dữ liệu cho lớp phơi nhiễm mạn (tĩnh)

Các biến đầu vào được cố định và lựa chọn theo cơ chế địa chất–môi trường:

- **Địa hình:** DEM, độ dốc (từ DEM).
- **Kết nối thủy văn:** khoảng cách tới đường bờ biển; khoảng cách tới các cửa sông; khoảng cách tới sông chính; mật độ sông/kênh; chỉ số tích tụ dòng chảy (flow accumulation) hoặc chỉ số ẩm địa hình (TWI).
- **Thổ nhưỡng:** loại đất/thành phần cơ giới; chỉ thị đất mặn/đất phèn tiềm tàng (nếu có).
- **Khí hậu nền:** lượng mưa trung bình năm và nhiệt độ trung bình năm (hoặc chỉ số khô hạn trung bình).

Lưu ý về “năm” của đầu vào tĩnh. Các biến khí hậu trong mô hình phơi nhiễm không sử dụng theo từng năm, mà được biểu diễn dưới dạng **trung bình nhiều năm** nhằm đại diện cho bối cảnh khí hậu nền. Các biến địa hình, kết nối thủy văn và thổ nhưỡng được xem là tĩnh trong giai đoạn nghiên cứu.

1.5.3 Dữ liệu cho lớp chất lượng rừng (tính)

Chỉ số chất lượng rừng được phân loại thành năm cấp: rất kém, kém, trung bình, tốt, rất tốt.

1.5.4 Dữ liệu chuỗi thời gian (động)

Chuỗi dữ liệu động phục vụ mô hình ConvLSTM bao gồm:

- **Chuỗi bản đồ mặn** theo năm (salinity/soil salinity hoặc proxy từ NDWI/salinity index), giai đoạn 2000–2023.
- **Chuỗi biến khí tượng–thủy văn theo năm** (ví dụ lượng mưa năm, nhiệt độ năm, các chỉ tiêu liên quan) ở dạng raster hoặc trường không gian tương ứng.

1.6 Hợp phần 1: Lập bản đồ mức độ phơi nhiễm mặn bằng học máy

1.6.1 Xây dựng biến mục tiêu (nhân) cho phơi nhiễm mặn

Biến mục tiêu của mô hình phơi nhiễm mặn không phải là trạng thái mặn tại từng năm, mà là **chỉ số phơi nhiễm mặn dài hạn** được xây dựng từ thống kê chuỗi dữ liệu mặn trong giai đoạn 2000–2023. Một lựa chọn phù hợp là sử dụng **tần suất vượt ngưỡng mặn** tại mỗi pixel:

$$E(i) = \frac{1}{N} \sum_{t=1}^N I(S_t(i) > \tau),$$

trong đó $E(i)$ là chỉ số phơi nhiễm tại pixel i , $S_t(i)$ là giá trị mặn tại năm t , τ là ngưỡng mặn lựa chọn, và $I(\cdot)$ là hàm chỉ thị.

Chỉ số $E(i)$ được chuẩn hóa về khoảng $[0, 1]$ và được xem là nhân giám sát để huấn luyện mô hình học máy. Cách định nghĩa này đảm bảo bản chất “tính theo nghĩa thống kê” của phơi nhiễm mặn và tránh rò rỉ thông tin thời gian.

1.6.2 Mô hình huấn luyện

Mô hình Random Forest (hoặc XGBoost) được sử dụng để học mối quan hệ phi tuyến giữa các yếu tố môi trường nền và chỉ số phơi nhiễm mặn:

$$\hat{E} = f(DEM, Slope, Dist_{sea}, Dist_{estuary}, Dist_{river}, RiverDensity, TWI, Soil, P_{mean}, T_{mean}).$$

Do dữ liệu raster có tự tương quan không gian, nghiên cứu sử dụng **kiểm định chéo theo khối không gian** (spatial blocked cross-validation) để giảm nguy cơ overfitting. Sau huấn luyện, mô hình được áp dụng để suy rộng chỉ số phơi nhiễm mặn cho toàn bộ vùng nghiên cứu ở độ phân giải 30 m.

1.6.3 Sản phẩm bản đồ phơi nhiễm mặn

Đầu ra trực tiếp của mô hình là bản đồ phơi nhiễm mặn liên tục $\hat{E} \in [0, 1]$. Bản đồ này được phân loại thành 05 cấp độ (rất thấp, thấp, trung bình, cao, rất cao) để phục vụ trình bày và phân tích theo vùng ưu tiên. Quy tắc phân loại có thể dựa trên phân vị (quantile) hoặc các ngưỡng sinh thái/môi trường phù hợp.

1.7 Hợp phần 2: Mô hình hóa và dự báo trạng thái xâm nhập mặn bằng ConvLSTM

1.7.1 Định nghĩa bài toán dự báo

Bài toán dự báo được định nghĩa là dự báo bản đồ trạng thái mặn ở thời điểm $t+1$ (hoặc $t+k$) từ chuỗi quan sát trong T năm trước đó. ConvLSTM coi chuỗi bản đồ mặn như một chuỗi không gian–thời gian (tương tự một “video” môi trường), trong đó trạng thái mặn tại mỗi pixel chịu ảnh hưởng của quá khứ và tương tác lân cận theo không gian.

1.7.2 Đầu vào mô hình ConvLSTM

Tại mỗi thời điểm t , đầu vào được biểu diễn dưới dạng tensor đa kênh:

$$X_t \in R^{H \times W \times C},$$

trong đó $H \times W$ là kích thước không gian và C là số kênh đầu vào. Các kênh bao gồm:

- S_t : bản đồ trạng thái mặn tại thời điểm t ;
- D_t : các biến khí tượng–thủy văn tại thời điểm t (ví dụ mưa năm, nhiệt độ năm);
- \hat{E} : bản đồ phơi nhiễm mặn (tĩnh), được lặp lại như một kênh nền ở mọi thời điểm.

Chuỗi đầu vào có dạng:

$$\{X_{t-T+1}, X_{t-T+2}, \dots, X_t\}.$$

1.7.3 Đầu ra mô hình ConvLSTM

Đầu ra của mô hình là bản đồ dự báo trạng thái mặn:

$$\hat{S}_{t+1} = g(X_{t-T+1:t}),$$

hoặc dự báo nhiều bước:

$$\{\hat{S}_{t+1}, \hat{S}_{t+2}, \dots, \hat{S}_{t+k}\}.$$

Như vậy, ConvLSTM dự báo **trạng thái mặn tương lai** (động), trong khi bản đồ phơi nhiễm \hat{E} được dùng như **kênh nền tĩnh** để điều tiết động lực.

1.8 Sản phẩm dự kiến

- Bản đồ mức độ phơi nhiễm mặn (30 m), dạng liên tục và 05 cấp độ.
- Bộ bản đồ trạng thái mặn dự báo theo các mốc thời gian tương lai (30 m).
- Bản đồ chất lượng rừng ven biển (30 m) và phân vùng chất lượng (05 cấp).