Journal of Environmental Management 401 (2026) 128811

Contents lists available at ScienceDirect

Journal of Environmental Management

journal homepage: www.elsevier.com/locate/jenvman

Research article

Machine learning prediction of saltwater intrusion to support coastal water resource management

Jian Shen [a,*], Xun Cai [b], Qubin Qin [c]

[a] Virginia Institute of Marine Science, William & Mary, Gloucester Point, VA, 23062, USA
[b] Yale School of the Environment, Yale University, New Haven, CT, 06511, USA
[c] Department of Coastal Studies, East Carolina University, Wanchese, NC, 27981, USA

ARTICLE INFO

Keywords:
Saltwater intrusion
Machine learning model
Water resource
Backward prediction
Chesapeake Bay
Climate changes

ABSTRACT

Saltwater intrusion (SWI), exacerbated by climate change, poses a significant threat to estuarine ecosystems, impacting drinking water supplies, agriculture, aquaculture, biodiversity, and habitat stability. Improving our ability to understand and predict the SWI is crucial to support decision-making for estuarine and coastal resource management. We developed machine learning (ML) models that estimate daily SWI length (*Ls*) in the Chesapeake Bay main stem and its eight major tributaries, supporting both historical reconstructions and future projections. The ML models were trained on two decades (2001–2020) of simulations from a validated, high-resolution three-dimensional hydrodynamic model. The ML model effectively reproduced variability of SWI, providing a simplified and computationally efficient alternative to traditional numerical modeling. Model skills were consistently high, with correlation coefficients ranging from 0.88 to 0.95 and root-mean-square errors (RMSE) between 1.53 and 5.03 km. In addition, the ML model demonstrated predictive capability for 7-day and 14-day forecasts using the preceding 90 days of discharge data. Scenario experiments simulating increased or reduced river discharge confirmed the model's utility for management applications. A key advantage of the ML approach is its ability to reconstruct historical or future SWI under limited forcing data conditions, providing valuable insights into long-term hydrological changes and the effects of climate variability on SWI in the Chesapeake Bay.

Key points

- A machine learning model trained on 20 years of 3D simulations accurately predicts saltwater intrusion in the Chesapeake Bay and major tributaries.
- The model provides daily, 7-day and 14-day forecasts, and scenario-based analyses to support coastal water resource management.
- The model reconstructs historical and future saltwater intrusion using limited data, offering insights into climate change impacts.

1. Introduction

An estuary is a transitional zone connecting freshwater from rivers and saline coastal seawater. An increase in saltwater intrusion (SWI) due to climate change poses a significant global threat to estuaries (Lee et al., 2024). The position of the salt-freshwater front is determined by complex estuarine dynamics driven by estuary geometry (length and depth) and external forcings such as river discharge, tides, and salinity at coastal boundary (MacCready and Geyer, 2010; Ralston and Geyer, 2019). Interactions among these forcings cause SWI to vary over spring-neap, seasonal, and interannual timescales. Beyond these natural variations, SWI is exacerbated by climate change. Extreme SWI events driven by drought and sea-level rise (SLR) have been documented in estuaries and delta worldwide, such as Rhine-Meuse Delta (Wegman et al., 2024), Bengal Delta (Sherin et al., 2020), Changjiang Estuary (Dai et al., 2011), Pearl River Estuary (Hong et al., 2020), Mekong Delta (Eslami et al., 2021), and Chesapeake Bay (Najjar et al., 2010; Hong and Shen, 2012). Many climate studies show that droughts are projected to increase in frequency and intensity by the 21st century (Naumann et al., 2018; Ji et al., 2023). Such extremes, combined with sea-level rise, enhanced the likelihood of severe SWI events, placing estuarine ecosystem at increasing risks. As these pressures intensify,

* Corresponding author.
E-mail address: shen@vims.edu (J. Shen).

https://doi.org/10.1016/j.jenvman.2026.128811
Received 18 October 2025; Received in revised form 9 January 2026; Accepted 30 January 2026

understanding and predicting SWI becomes increasingly critical, drawing greater attention from both governments and society.

Among U.S. estuaries, Chesapeake Bay presents a clear case of the challenges posed by SWI. Observation records and climate research projections have provided abundant evidence that freshwater resources are vulnerable and have the potential to be strongly affected by SLR (IPCC, 2007; Najjar et al., 2010; Hilton et al., 2008; Hong and Shen, 2012). Changes in SWI has far-reaching ecological and socio-economic consequences, impacting marsh and wetland ecosystems (Sutter et al., 2014, 2015), fish and shellfish habitats (Joseph et al., 2008), drinking water supplies and human infrastructure (Rice et al., 2012), and agriculture (Tully et al., 2019; Weissman and Tully, 2020). Although direct observations of groundwater and biogeochemical alterations are limited in Chesapeake Bay, studies in other estuaries demonstrate that SWI affects the groundwater (Gao et al.) and chemical reactions (Zhu et al., 2022). With a growing regional population, sustaining freshwater supplies in the face of SWI is crucial for long-term resilience of systems like Chesapeake Bay. Therefore, Chesapeake Bay provides a valuable case study for this research.

Mitigating the impacts of SWI requires urgent development of predictive tools that can forecast intrusion patterns and guide water resource managers in implementing effective strategies. A variety of simplified and empirical models have been developed over the past decades to address this need, each offering different approaches to predicting SWI length (*Ls*). Savenije (1986) developed a simplified model based on theoretical analysis to evaluate the *Ls* in an estuary. Gay and O'Donnell (2007) developed a one-dimensional model for a linearly tapered segment of an estuary, which allows for a simple analytic solution of *Ls*. The model has been successfully applied to compare SWI in three estuaries (Gay and O'Donnell, 2008). MacCready and Geyer (2010) demonstrated that changes in *Ls* can be expressed as a power function of freshwater (Q) as $Ls = aQ^b$, where a and b are constants. The empirical equations have shown a good prediction skill for SWI over seasonal to interannual variations (Ralston and Geyer, 2019; Rice et al., 2012).

While simplified and empirical models perform well for average conditions or long-term seasonal salinity variations, they are often inadequate for accurate predictions in estuaries with complex geometry and dynamics. It is widely recognized that three-dimensional (3D) numerical models are highly effective for simulating and predicting salinity variations and SWI in estuaries (Blumberg and Mellor, 1987; Wu et al., 2006; Werner et al., 2009; Ralston and Geyer, 2019; Gong and Shen, 2011; Irby et al., 2016; Ye et al., 2016; Jiang and Xia, 2016). However, the use of 3D numerical models requires high-resolution mode grids and extensive input data, including freshwater discharge, surface heat fluxes, wind forcing, and salinity and tidal conditions at the open boundary. Since these forcing data are not always available during historical and future periods, evaluating past SWI events and projecting future changes can be challenging.

Recently, machine learning (ML) has emerged as a powerful alternative to traditional modeling approaches (Brunton et al., 2019). Weng et al. (2024) applied tree-based classification techniques to develop an early warning system for SWI. Wang and Ge (2025) tested various ML methods, including neural networks (NNs), and combined ML outputs with a 3D model to improve SWI predictions. Hutton et al. (2016) developed a data-driven model to predict historical SWI in San Francisco Bay and the Delta, and they demonstrated that integrating numerical and ML models can enhance predictive skill (Hunter et al., 2017). More recently, Gorski et al. (2024) developed a Long Short-Term Memory (LSTM) SWI model that successfully predicted 7-day mean *Ls* in the Delaware River with a root-mean-square error (RMSE) of less than 2.52 miles (4.06 km). However, although ML models have shown strong skill in simulating SWI, few studies have demonstrated their ability to capture system responses to changes in discharge or to support scenario-based applications for water resource management.

This study aims to apply ML modeling techniques to develop a predictive tool capable of simulating and forecasting *Ls* in the Chesapeake Bay and its tributaries based on freshwater discharge. The ML-based model is designed to serve as practical tools for water resource management, enabling scenario analyses that address challenges from climate change and increased freshwater demand. Our framework supports both forward and backward predictions, allowing reconstruction of historical trends as well as diagnosis of future SWI in response to changes in freshwater discharge. The ML model achieved predictive skill comparable to high-resolution 3D hydrodynamic model while offering substantially greater computational efficiency. Importantly, it can provide reliable estimates even under limited data conditions, offering insights into climate-driven salinization and supporting sustainable management strategies in Chesapeake Bay and other estuarine systems facing similar challenges.

2. Materials and methods

2.1. Study area

The Chesapeake Bay is the largest estuary in the United States and one of the most productive estuaries in the world. The length of the Bay extends about 322 km from its northernmost point at the Susquehanna River to its mouth at the Atlantic Ocean. The major tributaries of the Bay include the Potomac River, James River, Rappahannock River, and York River on the western shore, and Nanticoke, Choptank, and Chester Rivers on the eastern shore (Fig. 1b). Chesapeake Bay is a partially mixed estuary (Pritchard, 1952). The mean water depth is about 6.46 m, and the mean tidal range at the mouth is 0.78 m. The Susquehanna and Potomac Rivers account for 50% and 19% discharges, respectively. The mean residence time is about 180 days (Du and Shen, 2016).

2.2. Numerical model

A seamless unstructured grid model for the North American Atlantic Coast (NAAC, Cai et al., 2025) has been developed based on the Semi-implicit Cross-scale Hydroscience Integrated System (SCHISM, Zhang et al., 2015). The model domain covers the US East Coast with a high-resolution (down to 100 m at tributaries) coverage in the Chesapeake Bay (Fig. 1a). The grid structure for the Bay's main stem is consistent with the validated Chesapeake Bay model developed in previous studies (Ye et al., 2016; Cai et al., 2022). The bathymetric configuration (Fig. 1b) is based on data from the NOAA Continuously Updated Digital Elevation Model (CUDEM; https://www.ncei.noaa.gov/metadata/geoportal/rest/metadata/item/gov.noaa.ngdc.mgg.dem:999919/html) and the USGS Coastal National Elevation Database (CoNED; https://www.usgs.gov/coastal-changes-and-impacts/coned). River discharges for the U.S. East Coast are sourced from the National Water Model (NWM; https://water.noaa.gov/about/nwm). Oceanic boundary conditions are derived from the Hybrid Coordinate Ocean Model (HYCOM; https://www.hycom.org) combined with eight spatially and temporally varying tidal constituents, including O1, K1, Q1, P1, M2, S2, K2, and N2, and atmospheric forcings are obtained from the North American Regional Reanalysis (NARR; https://www.ncei.noaa.gov/products/weather-climate-models/north-american-regional). The model simulation was conducted from 1/1/2001-12/31/2020, capturing a wide range of hydrological conditions, including the record-breaking dry year of 2002 and other wet years, therefore providing a comprehensive dataset for ML model training. Over this two-decade period, gradual sea-level rise signals are also embedded in the simulation, enhancing the model's ability to represent long-term salinity dynamics.

The three-dimensional model was validated using salinity observations at all available stations across the major bays along the US East Coast, demonstrating consistent model skill. However, direct validation of explicit saltwater intrusion length is only available for the Delaware Bay (Cai et al., 2025). Given the comparable salinity simulation

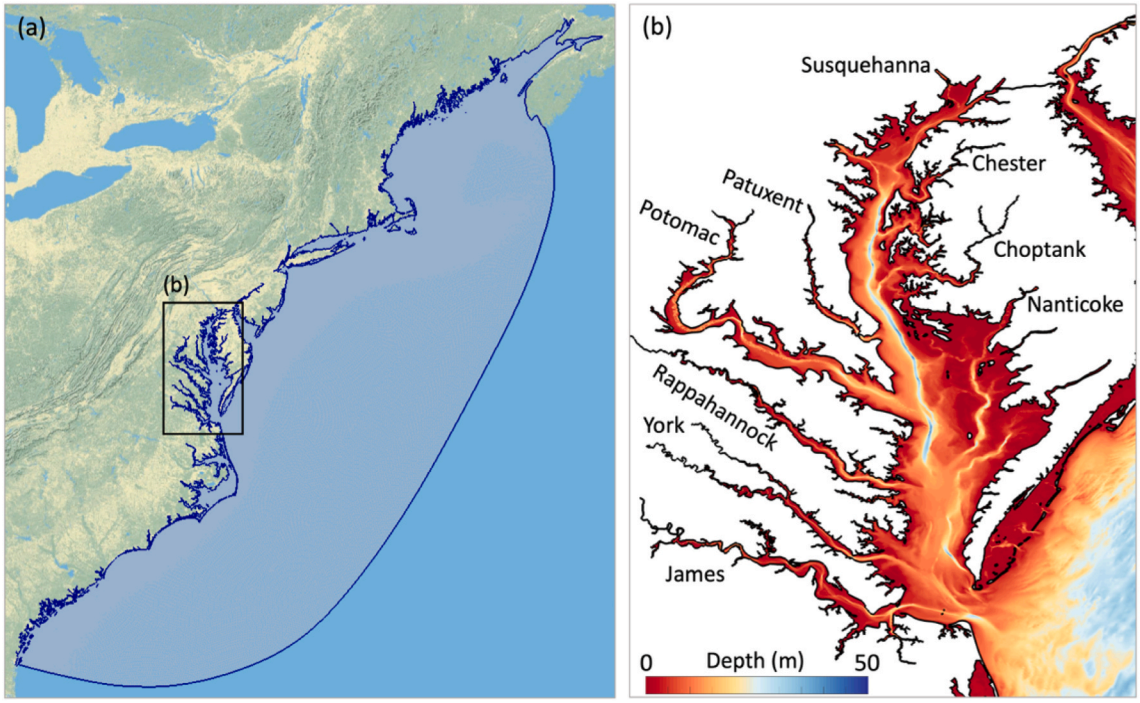

Fig. 1. (a) Numerical model domain and (b) a zoom-in at the Chesapeake Bay region.

performance achieved across stations in both the Delaware Bay and the Chesapeake Bay, we consider the simulated saltwater intrusion to be reliable across the broader model domain (Cai et al., 2025).

2.3. Spectral and correlation analysis

The daily SWI depends not only on the current day discharge, but also on the existing condition and preceding cumulative discharges. We applied spectral analysis to understand low and high-frequency fluctuations that drive SWI and how changes in *Ls* depend on cumulative discharge. Low-frequency seasonal variations in discharge highly contribute to the *Ls* variations (See Section 3.1). MacCready and Geyer (2010) show that the *Ls* has a scale of

$$Ls \approx aQ^{-\gamma}, \tag{1}$$

where γ varies from 1/7 to 1. Previous studies have shown that Eq. (1) provides a good estimate of *Ls* for monthly or annually averaged variations (Ralston and Geyer, 2018; Rice et al., 2012).

When using Eq. (1), *Ls* responds to the salinity in a time lag of Δt, which can be accounted for by parameter transformation as:

$$Ls(t) = \frac{a}{\Delta t}\int_{t-\Delta t}^{t} \widetilde{Ls}(\tau)d\tau + Er(t) \approx \frac{a}{\Delta t}\sum_{t-\Delta t}^{t} Q(t)^{-\gamma} + Er(t), \tag{2}$$

where $\frac{1}{\Delta t}\sum_{t-\Delta t}^{t} Q(t)^{-\gamma}$ is preceding cumulative mean transformed discharge, and Δt varies with time and tributaries. Besides, *Ls* can also be driven by the tide, salinity at the Bay mouth, and discharges discharged from multiple tributaries. To understand the correlations between *Ls* and discharge and other external forcings, we computed the correlations between *Ls* and cumulative discharge, salinity, and tide, which will provide essential information to select external forcings to develop ML models.

2.4. Machine learning approach

From a modeling perspective, the salinity intrusion length (*Ls*) can be expressed as:

$$Ls(t) = \widetilde{Ls}(t,p) + Er(t), \tag{3}$$

Where *Ls* is the saltwater intrusion length measured from the estuary or tributary mouth to the location where vertical mean salinity equals 0.5. $\widetilde{Ls}(t)$ is an empirical or dynamic model, t is time, p is model parameters (e.g., discharge, tide, wind, salinity, etc.), and $Er(t)$ is the deviations between the model and *Ls*,

$$Er(t) = Ls(t) - \widetilde{Ls}(t,p). \tag{4}$$

We can train a ML model to simulate $\widetilde{Ls}(t,p)$ to compute *Ls*. Because there are no available long-term observations of *Ls*, it has been a common practice to compute *Ls* using a 3D model, which has proven to have higher skill than empirical models (Ralston and Geyer, 2019; Irby et al., 2016; Gorski et al., 2024). We applied well calibrated SCHISM model (Cai et al., 2025) to estimate *Ls*, which are used as "observations" (hereafter referred to as observations) of *Ls* for training ML models.

Regarding the choice of machine-learning techniques, Random Forest (RF) and LSTM have often been used for time series prediction. For example, Gorski et al. (2024) used the past 360 days' discharge and an LSTM model to estimate 7-day mean *Ls* in Delaware Bay. More sophisticated methods, such as combined LTSM and Transform with attention, have also been applied to time series analysis (Shi et al., 2024). However, the Convolutional Neural Network (CNN) outperformed the other methods in sensitivity tests, consistently delivering the most accurate simulations of daily *Ls* across the Bay and its tributaries by effectively incorporating both discharge accumulation and temporal fluctuations into its input features. The network architecture, implemented in MATLAB, consisted of an input layer accepting sequences of length 90 to 210 input features, including discharges, tide, and salinity, followed by a convolutional layer with 64 filters of size 4×1, batch normalization, a ReLU activation, and a 2×1 max pooling layer (stride 1×1). A second convolutional layer with 24 filters of size 2×1, batch normalization, and ReLU activation was applied before a fully connected layer and regression output layer.

The first 70% of the data (1/1/2001-12/31/2012) were used for training, and the rest 30% data for testing the model. The test data are

not within the training period to ensure that the model has high predictive skill. To simulate Bay SWI, we tested models trained with different lengths of cumulative discharge data to evaluate how the duration of accumulation influences performance. We found that model performance decreases when using longer than 90-day preceding discharge data from the Susquehannock River. The final model used preceding 90-day discharge data from the Susquehanna River and Potomac River, plus preceding 30-day daily mean tide and salinity at the Bay mouth (not at 3D model open boundary).

For tributary configurations, incorporating the preceding 90-day discharges from the Susquehanna River and Potomac River did not improve the model performance. Instead of using the preceding 90-day data, we applied Eq. (2) to calculate cumulative discharges over 1, 3, 5, 10, 15, 20, 30, 40, 50, 60, and 80 days for each tributary as well as the Susquehanna and Potomac Rivers. We also incorporated tide and salinity data at 1, 3, 7, 15, and 30-day for the tributaries, which provided stronger model performance.

When conducting simulations using CNN model, variability in model outcomes can arise due to random initialization of filter and fully connected layer weights, as well as the stochastic order of data presentation during each training epoch (caused by training data shuffling). To enhance prediction accuracy and account for this variability, we implemented an ensemble training approach (Dietterich, 2000; Yu et al. 2020). Following the principles of multi-model inference and the Akaike Information Criterion (AIC) as described by Burnham and Anderson (2002), we evaluated each model in the ensemble (50 trained models) using scores of the RMSE, which equivalent to maximum likelihood under a Gaussian noise model of error. These scores were then used to compute model weights for model selection, which informed the selection of the best-performing model. The 95% confidence interval and standard deviations from ensemble test results can be used to quantify the model uncertainty.

3. Results

3.1. Frequency and correlation analysis

The spectral analysis results show the power distribution from 20 years of 3D model simulation of daily *Ls* in the main Bay (Fig, 2). The analysis reveals that semiannual and annual periods exhibit high spectral power, while spring-neap tidal cycles and the 29-day lunar period show relatively low power. This indicates that seasonal and interannual

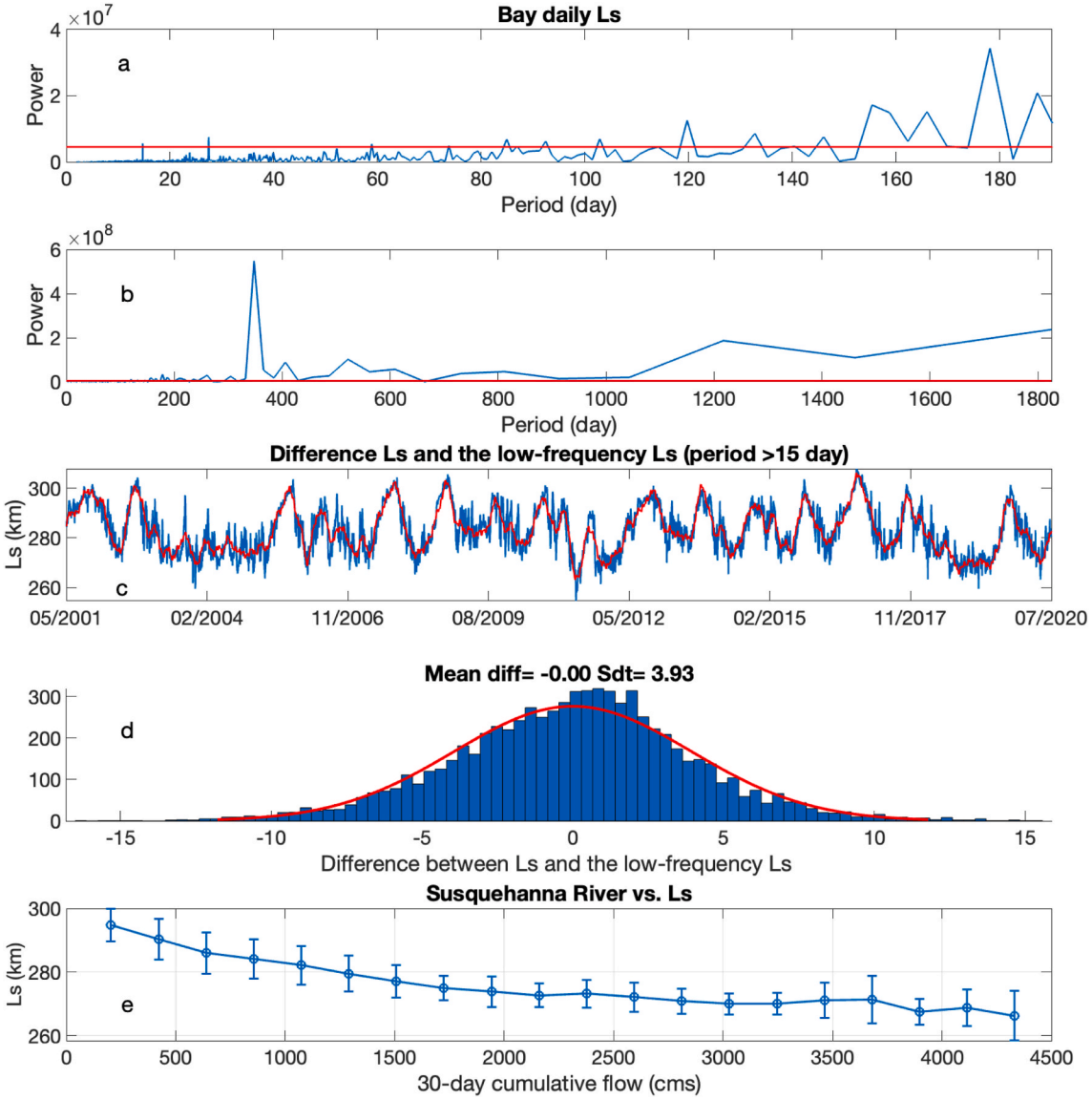

Fig. 2. Spectral analysis results of 20-year simulations of daily *Ls* in the Bay.

variations of discharge predominantly influence *Ls* dynamics. After filtering out low-frequency components with periods longer than 15 days, the high-frequency *Ls* follows a normal distribution, with a standard deviation of 3.9 km (Fig. 2d). This suggests that *Ls* is strongly influenced by low-frequency fluctuations. Therefore, both the cumulative discharge effects and the delayed response of *Ls* should be considered. In general, *Ls* increases as discharge decreases (Fig. 2e). *Ls* increases faster as discharge decreases. For example, when the discharge decreases from 1000 to 500 m^3s^{-1}, a 50% decrease, *Ls* increases by about 10 km, while reducing discharge by 500 m^3s^{-1} when the discharge is higher than 3000 m^3s^{-1}, *Ls* experiences less increase.

The correlation analysis shows the coefficients (CC) between daily *Ls* and transformed cumulative daily mean discharge ($Q^{-1/3}$), as well as daily tide and salinity (Fig. 3). The highest correlations are observed at 76 and 36 days for average windows of discharge from the Susquehanna and Potomac Rivers, with CC values of 0.83 and 0.76, respectively. The strongest correlation with tide occurs at day 31 (CC = 0.27), while the correlation with salinity is less than 0.5. The continuous increase in CC as the averaging period increases suggests that salinity exhibits a prolonged cumulative effect on *Ls*. Fig. 3 also presents the delayed correlation of *Ls* with all these variables. The discharge lags are 16 and 11 days for the Susquehanna and Potomac Rivers, respectively, while the lags are 1 and 61 days for tide and salinity, respectively. The corresponding CC values are 0.11 and 0.35, respectively. High correlation between *Ls* and tide also occurs at days 7 and 31. The results indicate that discharge is the dominant forcing to control *Ls* with high correlations, while high-frequency variation of *Ls* has a delayed response to these variables with weak correlations.

3.2. Machine learning model skills

ML model results are evaluated by CC, RMSE, error distribution, and ensemble error distribution (Fig. 4). This model demonstrated good predictive capability, with CC = 0.94 and 0.91, and RMSE = 3.08 and 4.26 km, respectively, for *train and test* runs. The similar performance between training and testing models indicates that the model is not overfitting. The error distribution followed a normal distribution with a mean of 0.12 km and standard deviation of 4.25 km, indicating relatively consistent performance across different discharge regimes. The errors are comparable to other ML or numerical models (Gorski et al., 2024; Gong and Shen, 2011). However, the results showed the discrepancies in short-term variations. Interestingly, the high-frequency fluctuations of *Ls* follow a normal distribution with a standard deviation of 3.93 km (Fig. 2d), comparable to the error of the ML model. This highlights the difficulty for a CNN model to fully capture high-frequency variability. This is expected, as high-frequency variations often depend on additional factors such as lateral discharge, precipitation, and wind, which were not included among the input variables. The weak correlation with tide and salinity indicates that *Ls* does not respond strongly to their high-frequency variability (Fig. 3), while the ML model tends to smooth such short-term fluctuations, as it has been observed in other ML predictions (Gorski et al., 2024). As our focus is on management-oriented applications, the ML model was not tested against additional variables that are often difficult to obtain. We assessed model uncertainty through the ensemble runs (Fig. 4, bottom panel), showing a mean daily standard deviation of about 2.09 km.

We also evaluated the model against observed *Ls* for each tributary (Fig. 5). The results indicate strong performances across all systems, with CC ranging from 0.88 to 0.95 and RMSE values between 1.53 and 5.03 km. Overall skill is comparable to that of the main Bay assessment.

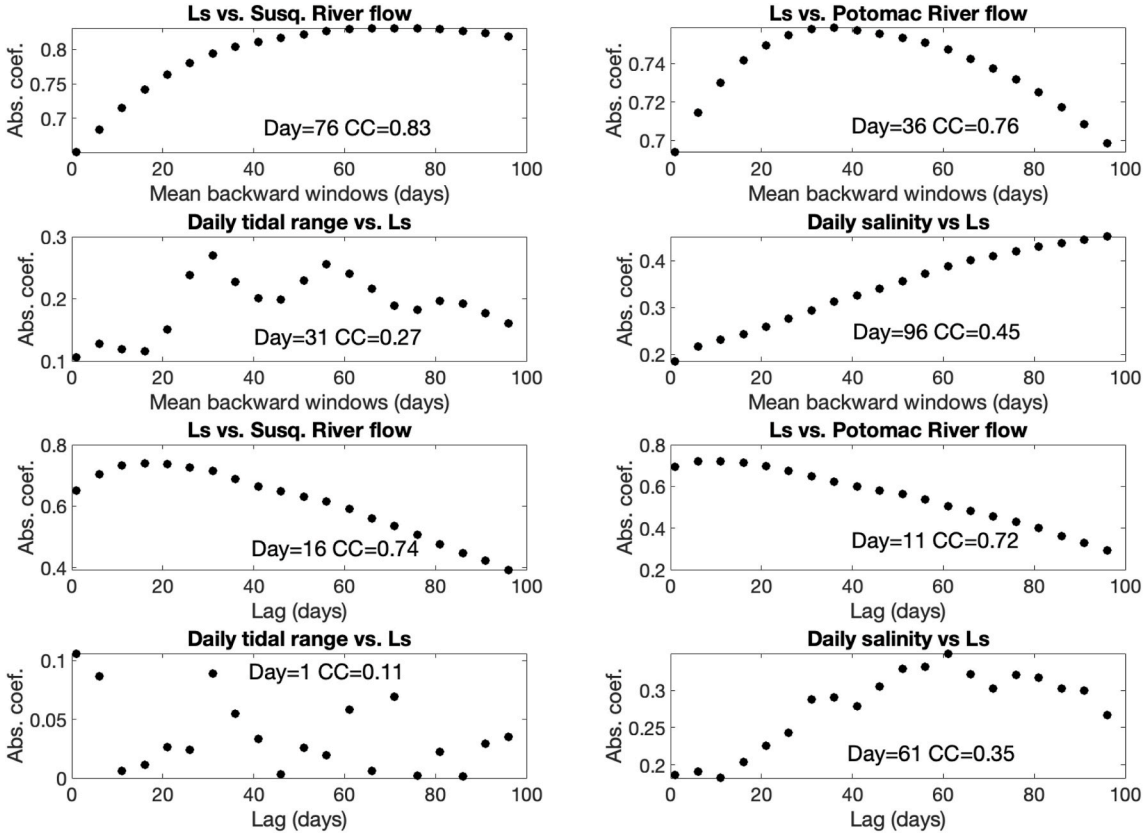

Fig. 3. Correlation coefficients (CC) between *Ls* and the mean cumulative discharge $Q^{-1/3}$ from the Susquehanna and Potomac Rivers, as well as daily tide and salinity at the Bay mouth. The figure also shows delayed correlations (the numbers indicate the time lag or averaging window in days at which the maximum correlation occurs for each variable).

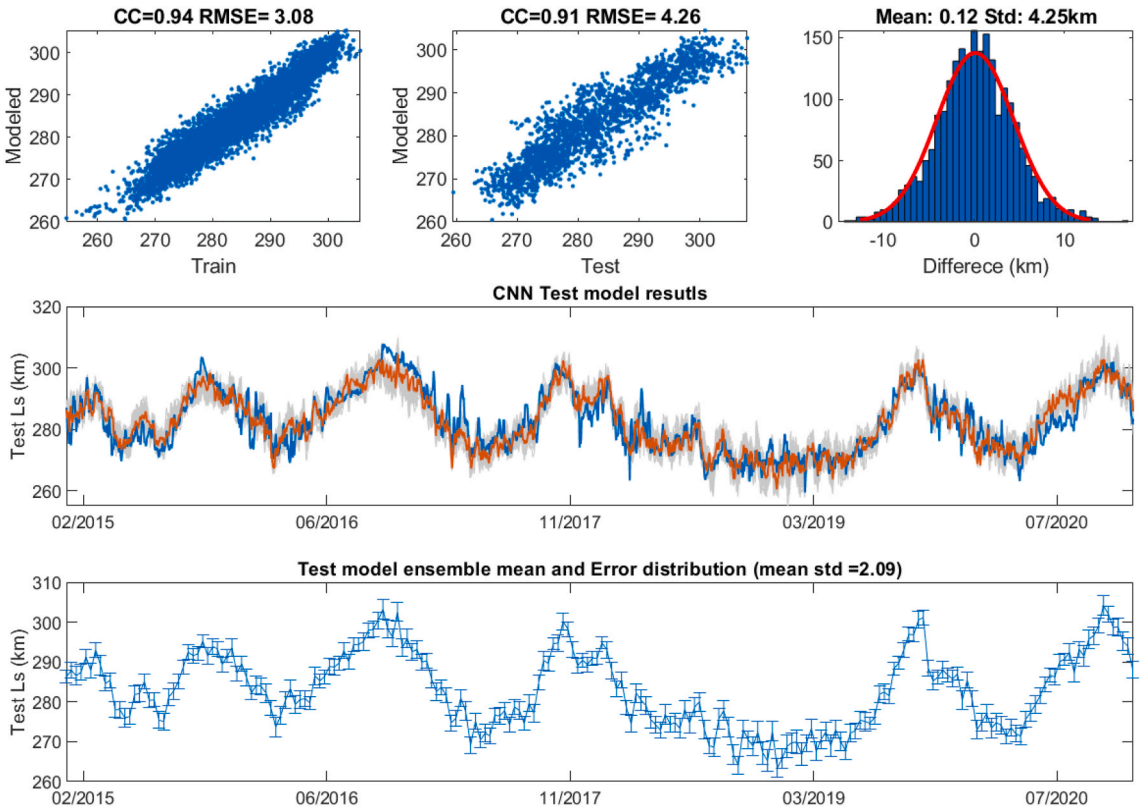

Fig. 4. Comparison of the selected best model against the observations (the top panel shows the correlation of training and testing model results against observations, and error distribution; the middle panel shows the selected test model (brown), observation (blue), and ensemble results (gray); the bottom panel shows variability of test models of ensemble run. The model uncertainty is quantified by the error bar of mean standard deviation from multiple models.

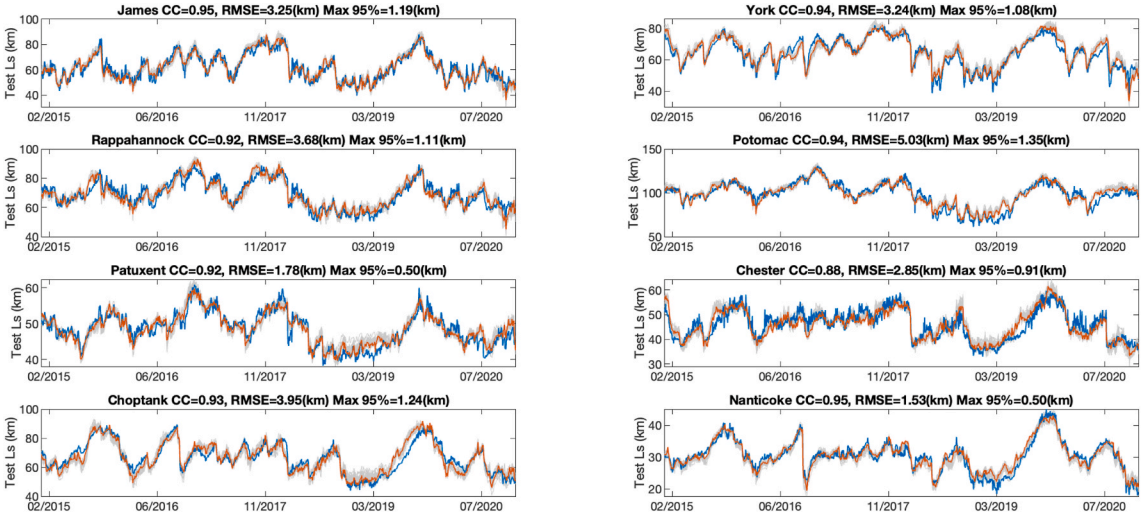

Fig. 5. Comparison of model prediction of *Ls* against observations for eight tributaries (blue lines are observations, brown lines are the weighted best model, gray lines are ensemble runs, Max 95% is estimated maximum 95% confidence interval for each day for ML model uncertainty based on results of ensemble run).

Among the tributaries, the Potomac exhibits the largest RMSE, while the Chester shows the lowest CC. Despite of these differences, all tributaries capture SWI dynamics well. One limitation is that during high discharge events, the ML model tends to over-predict SWI, particularly in small tributaries. These findings highlight that SWI dynamics in tributaries are highly sensitive to variations in upstream discharges.

3.3. *Forward prediction*

A key forecasting question is whether preceding discharge conditions can be used to predict future SWI and, if so, how accurate these forecasting will be. We compared the weighted best models from multiple ensemble simulations against the observations (Fig. 6). Firstly, when only using historical 90-day discharge data to forecast the future 7 days and 14-day *Ls*, the results are within acceptable range with CC = 0.87, 0.85 and RMSE = 4.92, 5.61, respectively (Fig. 6ac). The major

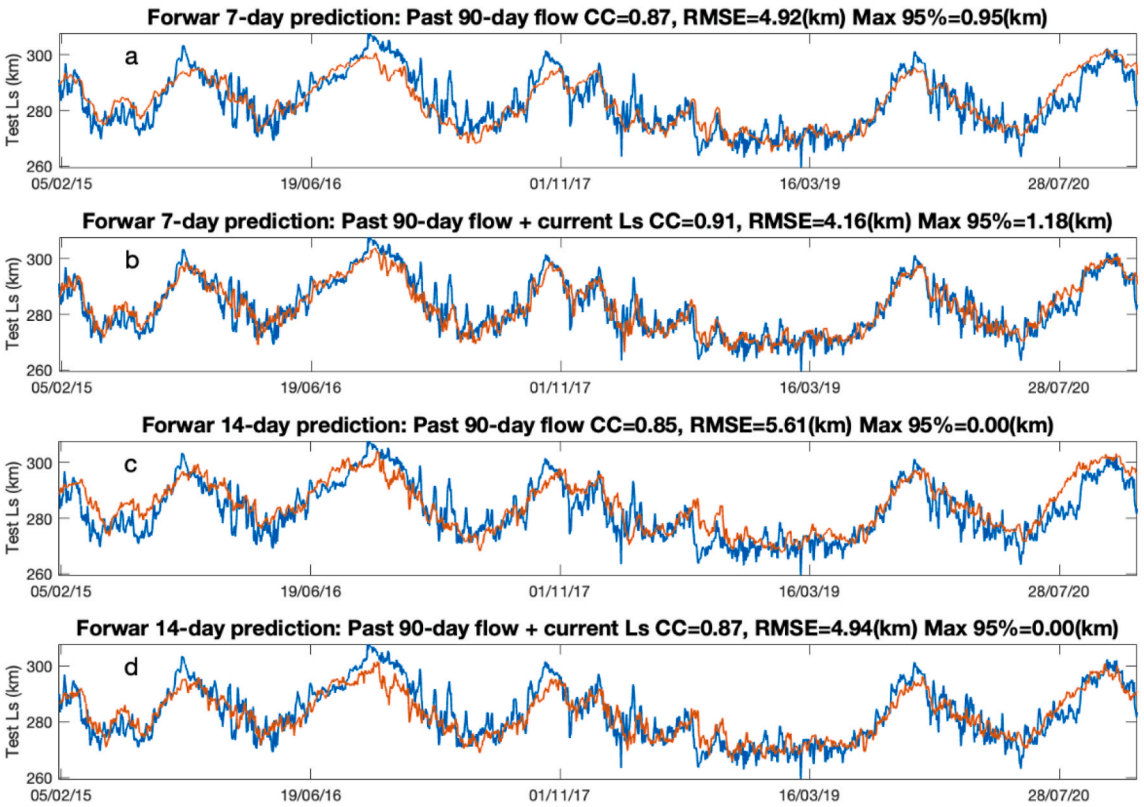

Fig. 6. Forward prediction on days 7th and 14th (brown lines are weighted ensemble simulations, and blue lines are observations, Max 95% are maximum 95 percentile confidence interval of ensemble model).

discrepancy is during the summer of 2016 when discharge was low. The error in both forward predictions is on the same order as daily prediction with slightly lag response when discharge changes rapidly. The maximum 95% percentile interval of trained ML model simulations is about 1.0 km.

Previous work has shown that current condition of SWI is important for forward prediction (Gong and Shen, 2011). Building on this, we included current day *Ls* along with preceding 90-day discharge as inputs for the ML model. The results demonstrate strong predictive skill (Fig. 6bd). For the 7-day forward prediction (Fig. 6b), the CC value and RMSE are 0.91 and 4.16 km, respectively. For the 14-day forward prediction, the CC value and RMSE are 0.87 and 4.94 km, respectively (Fig. 6d). Although the 14-day forecast is not as accurate as the 7-day forecast, the results are still acceptable. The results suggest that using past discharge data to forecast future increase time lag in response to rapid discharge changes, but the RMSE is within the same order of daily prediction.

4. Discussion

4.1. Bay model with reduced independent variables

We trained the Bay model using discharges from the Susquehanna and Potomac Rivers, along with tidal range and salinity at the Bay mouth. We found using input data longer than preceding 90 days reduces the model performance. This can be understood as correlation between *Ls* and discharge is unchanged to even decreased when cumulative discharge is longer than 90 days (Fig. 3). To assess the impact of tide and salinity on *Ls*, we trained the model without including tidal range and salinity. The results showed only minor differences between the models with and without tide and salinity, as both variables were only weakly correlated with *Ls* (Fig. 3).

We also evaluated the impact of Potomac River discharge on *Ls* by training the model using only Susquehanna River discharge as input. This configuration yielded a CC of 0.87 and a RMSE of 5.04 km during the test period. The error distributions were of the same order (Fig. 7a). Compared to the model trained with all variables (Fig. 4), prediction accuracy decreased slightly (Fig. 7b). However, the cumulative distributions between ML model prediction and obsecration are very close (Fig. 7c). These results suggest that Susquehanna River discharge is the dominant factor controlling SWI in the Bay. Interestingly, the ML model appears to reduce the influence of Potomac River discharge, likely due to the correlation between the two rivers. However, since Potomac River discharge also affects the phase of the SWI, an increase in error is expected. Given the minor increase in RMSE, the model driven solely by Susquehanna River discharge remains acceptable.

We applied the same approach to individual tributaries. When the model was trained using discharge data from a single tributary, similar performance was observed: RMSE increased slightly, and CC decreased slightly. Overall, the differences were not significant. The advantage of using only the dominant discharge variable is that it simplifies model application for resource management, especially when studying historical discharge impacts or projecting future climate scenarios where input data are limited.

4.2. Sensitivity of SWI response to change in discharge

The test results indicate that the trained ML model can successfully simulate SWI across the Bay and its tributaries. A key remaining question is whether this model can also serve as a tool for water resource management, specifically in evaluating how changes in river discharge alter SWI dynamics. To investigate this, we conducted two experiments in which Susquehanna River discharge was increased and decreased by 30%, respectively. Significant daily variation in *Ls* was observed with changes in discharge, as *Ls* depends on cumulative discharge over time (Fig. 8). The model responded appropriately to these changes, with

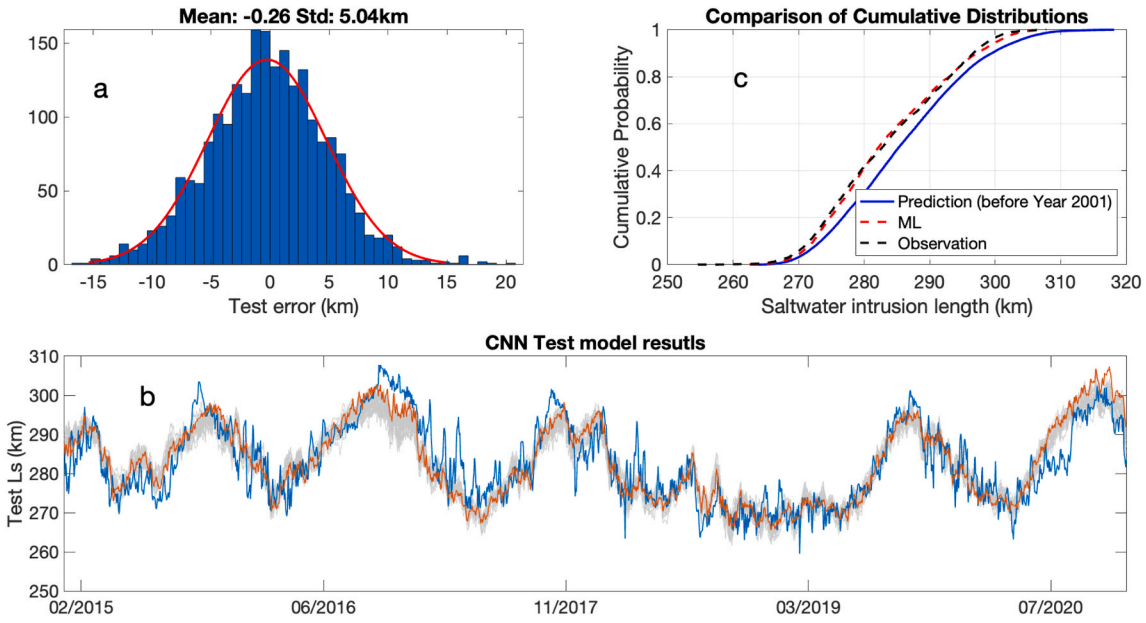

Fig. 7. Model simulations of *Ls* with 90-day backward mean discharge from Susquehanna River (gray lines are ensemble runs, brown line is the best trained model, and blue line are observation).

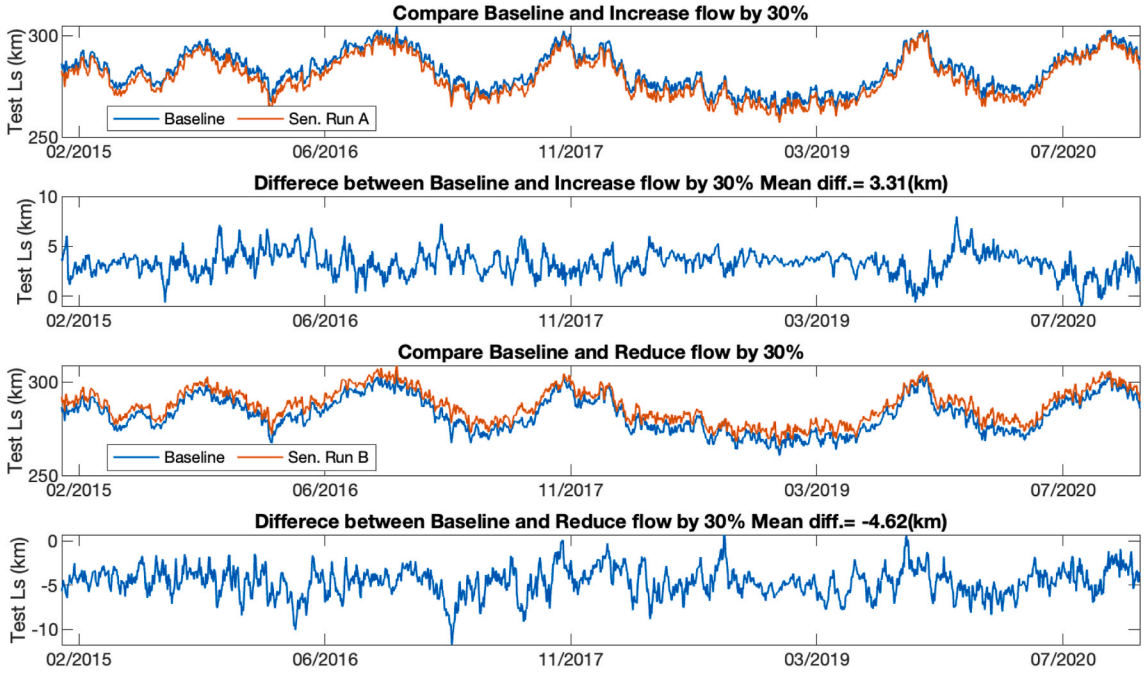

Fig. 8. Comparison of model simulations with increasing or decreasing discharge by 30%.

variations in *Ls* less than8 km comparing to the baseline condition. Most changes are about 5 km. The magnitude of change in *Ls* was consistent with the change in discharge (Fig. 2e). On average, *Ls* decreased by approximately 3.3 km when discharge increased 30% and increased by about 4.6 km when discharge decreased 30%. The impact of reduced discharge on *Ls* was greater than that of increased discharge, which is consistent with observations (Fig. 2e), suggesting the physical processes driven *Ls* is correctly simulated.

4.3. Reconstruction of historical saltwater intrusion

To understand the impact of climate-driven changes in river discharge on SWI, historical variability in SWI provides critical evidence (Hilton et al., 2008). However, there is insufficient observational data to directly assess historical changes in SWI in the Chesapeake Bay. Although three-dimensional numerical models (e.g., Hong and Shen, 2012) can simulate past SWI conditions, the lack of boundary condition data to force these hydrodynamic models, such as wind, tide, and salinity prior to 1984 in Chesapeake Bay limits their applicability.

In contrast, this trained ML model offer a promising alternative for reconstructing historical SWI. Based on our model validation, the standard deviation of prediction error is approximately 5 km, which is comparable to the deviations observed in numerical models and other ML models (Rice et al., 2012; Gorski et al., 2024). Since this ML model

only requires historical river discharge data, it enables reconstruction of SWI in the Chesapeake Bay from 1890 to the present.

We applied the trained ML model to estimate historical *Ls* in the Bay 1891–2000 and compared the cumulative distribution of observed and ML-predicted *Ls* from 2001 to 2020 with the reconstructed values from the historical period (Fig. 7c). The cumulative curves for observations and ML test period match well, while backward prediction of *Ls* values were generally higher during 1891–2000, suggesting a decreasing trend over time (Fig. 7c). Fig. 9 presents the monthly mean *Ls* over a 134-year period, capturing both seasonal and interannual variability. The results indicate an overall decreasing trend in *Ls*, likely driven by prolonged dry periods from 1958 to 1971 and 1980–1981 (Summer, 1981). This pattern aligns with recent studies reporting increasing stream discharge variability in the Bay watershed, particularly during extreme wet events, consistent with observed precipitation trends (Groisman et al., 2001, 2004). The ML-predicted *Ls* values accurately reflect hydrological changes in the Bay watershed; however, seasonal trends in SWI diverge, with opposite trends observed between spring and winter. Overall, this information provides valuable insight into the effects of climate change on SWI in the Chesapeake Bay.

4.4. Limitations and future directions

Due to the lack of long-term observations of SWI in the Chesapeake Bay and its tributaries, we used a high-resolution numerical model to estimate SWI. 3D model simulation of SWI was only verified in the Delaware Bay (Cai et al., 2025). As the model simulates entire East Coast of the USA and model has the same salinity simulation skill of Delaware Bay and Chesapeake, the simulation of saltwater intrusion is reliable. Although the model has been well-calibrated for a 20-year salinity simulation (Cai et al., 2025), and the accuracy of the SCHISM model is comparable to, or better than, other models used for estimating salinity in the Bay (Irby et al., 2016), uncertainty remains due to inherent assumptions and simplifications involved in numerical modeling approaches.

Our ML model performed well in capturing seasonal and interannual variations. However, it remains challenging to simulate short-term variations with periods less than 7 days. This limitation warrants further investigation if high-frequency predictions are desired.

To reduce uncertainty in ML model development, we applied an ensemble modeling approach. The resulting RMSE and standard deviation of the error distribution were both less than 5 km, which can be used to quantify prediction uncertainty and set margin of safety. To enhance applicability for climate change studies and water resource management, the long-term prediction model was trained using only discharge data from the Susquehanna River for the Bay and upstream discharge to individual tributaries for tributaries. This simplification slightly increased RMSE by less than 1 km, but the error remains within an acceptable range.

In reconstructing historical SWI, SLR was not explicitly included as a model input. Boon (2005) reported a rate of sea-level rise of 4.25 mm/year based on long-term tidal data from Hampton Roads (James River mouth), predicting an increase of 29.8 cm over the past 70 years. A sea-level rise of 50 cm in the Bay could result in a 6 km increase in SWI (Hong and Shen, 2012), which could contribute to reduce the trend due to increase of stream discharge. While the current ML model is trained on 20 years of 3D hydrodynamic simulations that already embed some SLR signals, explicitly incorporating sea-level adjustments in future versions would improve accuracy. Expanding the training dataset to include longer-term simulations and a wider range of sea-level conditions could further improve model performance and reliability.

5. Conclusions

In this study, we developed ML models to predict the SWI length in the main stem of the Chesapeake Bay and its eight major tributaries,

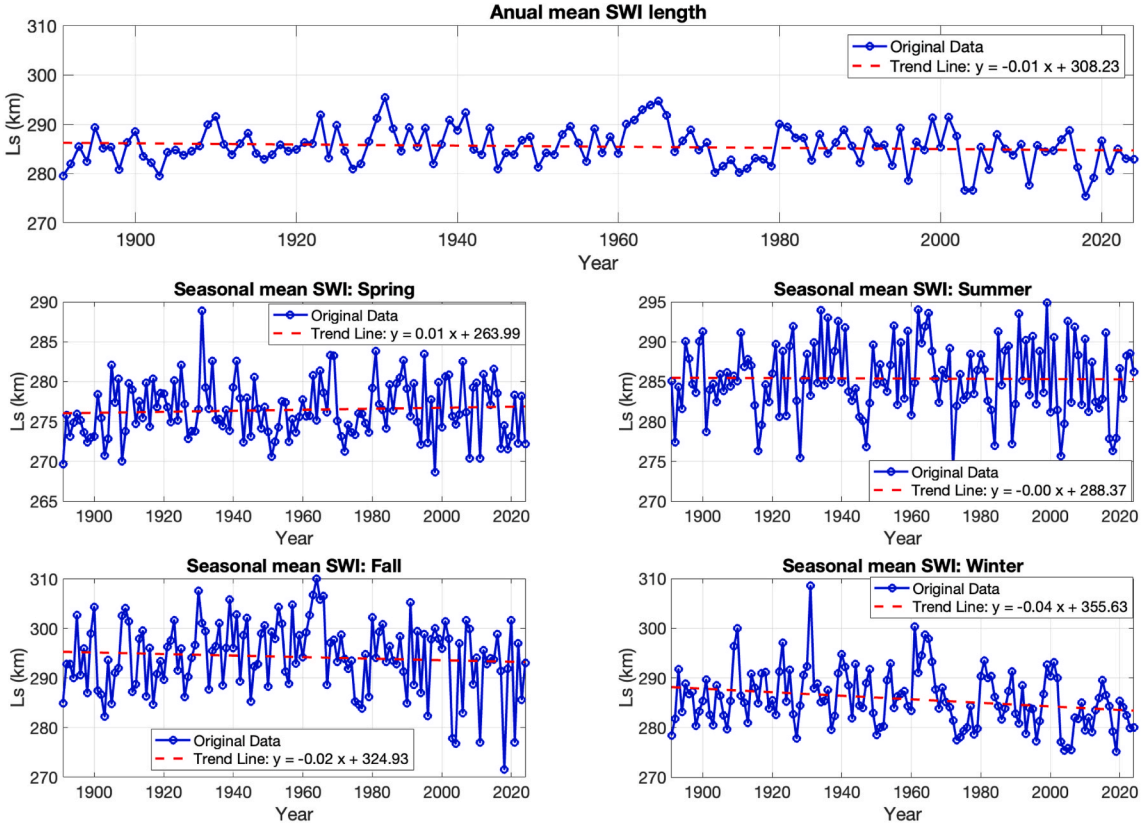

Fig. 9. Inter annual varying monthly mean and seasonal variations of SWI length (*Ls*) predicted by the ML model in the Chesapeake Bay.

trained on two decades (2001–2020) of high-resolution simulations from a regional 3D hydrodynamic model, which serve as a valuable substitute for long-term observational records. Following a comparative evaluation of multiple ML architectures, we selected CNN model as the optimal framework for simulating both the Bay and its tributaries, using discharge inputs from the Susquehanna and Potomac Rivers along with individual tributary discharges. To minimize training uncertainty, we employed ensemble runs to identify the best-performing configurations and quantify prediction variability.

Overall, the trained ML models demonstrated strong predictive skills, with correlation coefficients ranging from 0.88 to 0.95 and RMSE values between 1.53 and 5.03 km. The model is capable of forward prediction for 7-day and 14-day periods using only the preceding 90 days of cumulative upstream discharge. Prediction accuracy improves when current-day *Ls* is included as an input. A key advantage of this ML approach is its ability to reconstruct historical SWI under limited data conditions, offering valuable insights into how climate change has influenced SWI in the Chesapeake Bay.

Given the lack of long-term observational data for *Ls*, this study demonstrates that integrating high-resolution 3D model outputs with ML models offers a robust and effective alternative for estimating and predicting *Ls*. Nonetheless, some limitations remain, particularly in capturing high-frequency, short-term variations. Addressing these gaps will require further refinement of the model framework and incorporation of additional hydrodynamic and meteorological drivers.

CRediT authorship contribution statement

Jian Shen: Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Conceptualization. **Xun Cai:** Writing – review & editing, Validation, Methodology, Investigation, Conceptualization. **Qubin Qin:** Writing – review & editing, Methodology, Investigation, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Xun Cai reports financial support was provided by US NSF. Xun Cai reports a relationship with US NSF that includes: funding grants. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

XC is supported by an NSF OCE-PRF fellowship (grant no. 2403359). Simulations presented in this paper were conducted using Sciclone at William & Mary.

Data availability

All discharge data are publicly available from USGS (https://dashboard.waterdata.usgs.gov/app/nwd/en/). Saltwater intrusion data will be available upon request.

References

Blumberg, A.F., Mellor, G.M., 1987. A description of a three-dimensional coastal ocean circulation model. In: Heaps, N.S. (Ed.), Three-Dimensional Coastal Ocean Models. Coastal and Estuarine Science, 4. American Geophysical Union, p. 1e19.

Boon, J., 2005. Isabel's silent partners: seasonal and secular sea level change. In: Sellner, K.G. (Ed.), Hurricane Isabel in Perspective.

Brunton, S.L., Noack, B.R., Koumoutsakos, P., 2019. Machine learning for fluid mechanics. Annu. Rev. Fluid Mech. https://doi.org/10.1146/annurev-fluid-010719-060214\.

Burnham P., K., Anderson, D.R., 2002. Model Selection and Multimodal Inference: a Practical information-theoretic Approach, second ed. Springer-Verlag, New York, Inc.

Cai, X., Zhang, Y., Shen, J., Wang, H., Wang, Z., Qin, Q., Ye, F., 2022. A numerical study of hypoxia in Chesapeake Bay using an unstructured grid model: validation and sensitivity to bathymetry representation. JAWRA J. Am Water Res. Assoc. 58 (6), 898–921. https://doi.org/10.1111/1752-1688.12887.

Cai, X., Qin, Q., Cui, L., Yang, X., Zhang, Y.J., Shen, J., 2025. NAAC (v1.0): a seamless two-decade cross-scale simulation from the north American Atlantic Coast to tidal wetlands using the 3D unstructured-grid model SCHISM. v5.11.0. https://doi.org/10.5194/egusphere-2025-593.

Dai, Z., Chu, A., Stive, M., Zhang, X., Yan, H., 2011. Unusual salinity conditions in the Yangtze Estuary in 2006: impacts of an extreme drought or of the three Gorges Dam? Ambio 40 (5), 496–505. EGUsphere, 2025, 1-23. doi: Egusphere-2025-593.

Dietterich, T.G., 2000. An experimental comparison of three methods for constructing ensembles of decision trees: bagging, boosting, and randomization. Mach. Learn. 40, 139–157, 2000.

Du, J., Shen, J., 2016. Water residence time in Chesapeake Bay for 1980–2012. J. Mar. Syst. https://doi.org/10.1016/j.jmarsys.2016.08.011.

Eslami, S., Hoekstra, P., Kernkamp, H.W.J., Nguyen Trung, N., Do Duc, D., Nguyen Nghia, H., et al., 2021. Dynamics of salt intrusion in the Mekong delta: results of field observations and integrated coastal–inland modelling. Earth Surf. Dyn. 9 (4), 953–976. https://doi.org/10.5194/esurf-9-953-2021.

Gao, S., Zheng, T., Zheng, X., Luo, J., Impact of river-groundwater interactions on residual saltwater pollution in estuarine groundwater reservoirs, Water Res., (279), 123474, https://doi.org/10.1016/j.watres.2025.123474.

Gay, P.S., O'Donnell, J., 2007. A simple advection-dispersion model for the salt distribution in linearly tapered estuaries. J. Geophys. Res. 112, C07021. https://doi.org/10.1029/2006JC003840.

Gay, P.S., O'Donnell, J., 2008. Comparison of the salinity structure of the Chesapeake Bay, the Delaware Bay, and Long Island Sound using a linearly tapered advection-dispersion model. Estuaries Coasts. https://doi.org/10.1007/s12237-008-9101-4.

Gong, W., Shen, J., 2011. The response of salt intrusion to changes in river discharge and tidal mixing during the dry season in the Modaomen Estuary, China. Cont. Self Res. 31 (7-8), 2484–2502. https://doi.org/10.1016/j.csr.2011.01.011.

Gorski, G., Cook, S., Snyder, A., Appling, A.P., Thompson, T., Smith, J.D., Warner, J.C., Topp, S.N., 2024. Deep learning of estuary salinity dynamics is physically accurate at a fraction of hydrodynamic model computational cost. Limnol. Oceanogr. 69, 1070–1085. https://doi.org/10.1002/lno.12549.

Groisman, P.Y., Knight, R.W., Karl, T.R., 2001. Heavy precipitation and high streamflow in the contiguous United States: trends in the twentieth century. Bull. Am. Meteorol. Soc. 82, 219–246.

Groisman, P.Y., Knight, R.W., Karl, T.R., Easterling, D.R., Sun, B.M., Lawrimore, J.H., 2004. Contemporary changes of the hydrological cycle over the contiguous United States: trends derived from in situ observations. J. Hydrometeorol. 5, 64–85.

Hong, B., Shen, J., 2012. Responses of estuarine salinity and transport processes to potential future sea-level in the Chesapeake Bay. Estuar. Coast Shelf Sci. 104–105, 33–45. https://doi.org/10.1016/j.ecss.2012.03.014.

Hong, B., Liu, Z., Shen, J., Wu, H., Gong, W., Xu, H., Wang, D., 2020. Potential physical impacts of sea-level rise on the Pearl River Estuary, China. J. Mar. Syst. 201, 103245. https://doi.org/10.1016/j.jmarsys.2019.103245.

Hilton, T.W., Najjar, R.G., Zhong, L., Li, M., 2008. Is there a signal of sea-level rise in Chesapeake Bay salinity? J. Geophys. Res. 113, C09002. https://doi.org/10.1029/2007JC004247.

Hunter, J.M., Maier, H.R., Gibbs, M.S., Foale, E.R., Grosvenor, N.A., Harders, N.P., Kikuchi-Miller, T.C., 2017. Modelling salinity in river systems using hybrid process and data-driven models. Hydrol. Earth Syst. Sci. Discuss. https://doi.org/10.5194/hess-2017-571.

Hutton, P.H., Rath, J.S., Chen, L., Ungs, M.J., Roy, S.B., 2016. Nine decades of salinity observations in the San Francisco Bay and Delta: modeling and trend evaluations. J. Water Resour. Plann. Manag. 142, 04015069.

Intergovernmental Panel on Climate Change, 2007. Climate change 2007dthe physical science basis. In: Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K.B., Tignor, M., Miller, H.L. (Eds.), Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom. Available on-line at: http://www.ipcc.ch/ipccreports/ar4-wg1.htm.

Irby, I.D., Friedrichs, M.A.M., Friedrichs, C.T., Bever, A.J., Hood, R.R., Lanerolle, L.W.J., Li, M., Linker, L., Scully, M.E., Sellner, K., Shen, J., Testa, J., Wang, H., Wang, P., Xia, M., 2016. Challenges associated with modeling low-oxygen waters in Chesapeake Bay: a multiple model comparison. Biogeosciences 13, 2011–2028. https://doi.org/10.5194/bg-13-2011-2016.

Ji, Y., Fu, J., Lu, Y., Liu, B., 2023. Three-dimensional-based global drought projection under global warming tendency. Atmos. Res. 291, 106812.

Jiang, L., Xia, M., 2016. Dynamics of the Chesapeake Bay outflow plume: realistic plume simulation and its seasonal and interannual variability. J. Geophys. Res., Oceans 121 (2), 1424–1445. https://doi.org/10.1002/2015JC011191.

Joseph, W. Love, Gill, John, Joshua, J.N., 2008. Saltwater intrusion impacts fish diversity and distribution in the blackwater river drainage (Chesapeake Bay watershed). Wetlands 28 (4), 967–974. https://doi.org/10.1672/07-238.1.

Lee, J., Biemond, B., Keulen, D.V., Huismans, Y., Westen, R.M., Swart, H.E., Dijkstra, H. A., Hranenburg, W.M., 2024. Global increases of salt intrusion in estuaries under future environmental condition. Nat. Commun. 16, 3444. https://doi.org/10.1038/s41467-025-58783-6.

MacCready, P., Geyer, W.R., 2010. Advances in estuarine physics. Ann. Rev. Mar. Sci 2 (1), 35–58. https://doi.org/10.1146/annurev-marine-120308-081015.

Najjar, R.G., Pyke, C.R., Adams, M.B., Breitburg, D., Hershner, C., Kemp, M., Howarth, R., Mulholland, M.R., Paolisso, M., Secor, D., Sellner, K., Wardrop, D., Wood, R., 2010. Potential climate-change impacts on the Chesapeake Bay. Estuar. Coast Shelf Sci. 86, 1e20.

Naumann, G. L. Alfieri, Wyser, K., Mentaschi, L., Betts, R.A., Carrao, H., Spinoni, J., Vogt, J., Feyen, L., 2018. Global changes in drought conditions under different levels of warming. Geophys. Res. Lett. 45, 3285–3296.

Pritchard, D.W., 1952. Salinity distribution and circulation in the Chesapeake Bay estuaries system. J. Mar. Res. 11, 106–123.

Ralston, D.K., Geyer, W.R., 2018. Sediment transport time scales and trapping efficiency in a tidal river. J. Geophys. Res.: Earth Surf. 122 (11), 2042–2063. https://doi.org/10.1002/2017JF004337.

Ralston, D.K., Geyer, W.R., 2019. Response to channel deepening of the salinity intrusion, estuarine circulation, and stratification in an urbanized estuary. J. Geophys. Res., Oceans 124. https://doi.org/10.1029/2019JC015006.

Rice, K.C., Hong, B., Shen, J., 2012. ssessmet f salnity intrusin in the ames and Chickahominy Riers as a result of simulated sea-level rise in Chesapeake Bay, ast Coast, USA. J. Enviro. Management. 111, 61–69.

Savenije, H.H.G., 1986. A one-dimensional model of salinity intrusion in alluvial estuaries. J. Hydrol. 85, 87–109.

Sherin, V., Durand, F., Papa, F., Islam, A.S., Gopalakrishna, V., Khaki, M., Suneel, V., 2020. Recent salinity intrusion in the Bengal Delta: observations and possible causes. Cont. Shelf Res. 202, 104142. https://doi.org/10.1016/j.csr.2020.104142.

Shi, J., Wang, S., Qu, P., Shao, J., 2024. Time series prediction model using LSTM-Transformer neural network for mine water inflow. Sci. Rep. 14. https://doi.org/10.1038/s41598-024-69418-z. Article 11306369.

Summer, S.G., 1981. Contributing factors to the 1980-1981 water supply drought. Northeast U.S. NOAA Technical memorandum NSW Er-66.

Sutter, L.S., Perry, J.E., Chambers, R.M., 2014. Tidal freshwater Marsh plant responses to low level salinity increases. Wetlands 34, 167–175.

Sutter, L.S., Chambers, R.M., Perry, J.E., 2015. Seawater intrusion mediates species transition in low salinity, tidal marsh vegetation. Aquat. Bot. 122, 32–39.

Tully, K., Weissman, D., Wyner, W.J., Miller, J., Jordan, T., 2019. Soils in transition: saltwater intrusion alters soil chemistry in agricultural fields. Biogeochemistry 142, 339–356.

Wang, N., Ge, J., 2025. Predictions of saltwater intrusion in the Changjiang Estuary: integrating machine learning methods with FVCOM. J. Hydrol., 132739. Available online 30 January 2025.

Wegman, T.M., Pietrzak, J.D., Horner-Devine, A.R., Dijkstra, H.A., Ralston, D.K., 2024. Observations of estuarine salt intrusion dynamics during a prolonged drought event in the Rhine Meuse Delta. J. Geophys. Res., Oceans 129. https://doi.org/10.1029/2024JC021655 e2024JC021655.

Weissman, D., Tully, K., 2020. Saltwater intrusion alters nutrient cycling in coastal ecosystems. Ecosphere 11 (2), e03041.

Weng, P., Tian, Y., Zhou, H., Zheng, Y., Jiang, Y., 2024. Saltwater intrusion early warning in Pearl River Delta based on the temporal clustering method. J. Environ. Manag. 349, 119443.

Werner, A.D., Jakovovic, D., Simmons, C.T., 2009. Experimental observations of saltwater up-coning. J. Hydrol. 373, 230–241.

Wu, H., Zhu, J., Chen, B., Chen, Y., 2006. Quantitative relationship of runoff and tide to saltwater spilling over from the north branch in the Changjiang estuary: a numerical study. Estuar. Coast Shelf Sci. 69 (1), 125–132.

Ye, F., Zhang, Y.J., Friedrichs, M.A.M., Wang, H.V., Irby, I.D., Shen, J., Wang, Z., 2016. A 3D, cross-scale, baroclinic model with implicit vertical transport for the Upper Chesapeake Bay and its tributaries. Ocean Model.

Yu, X., Shen, J., Du, J., 2020. A machine-learning-based model for water quality in coastal waters, taking dissolved oxygen and hypoxia in Chesapeake Bay as an example. Water Resour. Res. https://doi.org/10.1029/2020WR027227.

Zhang, Y.J., Ateljevich, E., Yu, H.C., Wu, C.H., Jason, C.S., 2015. A new vertical coordinate system for a 3D unstructured-grid model. Ocean Model. 85, 16–31. https://doi.org/10.1016/j.ocemod.2014.10.003.

Zhu, L., Zhang, X., Zhang, J., Liu, T., Qiu, Y., 2022. Saltwater intrusion weakens Fe-(oxyhydr)oxide-mediated (im)mobilization of Ni and Zn in redox-fluctuating soil–groundwater system. Water Res. 221, 118799. https://doi.org/10.1016/j.watres.2022.118799.