Applied Ocean Research 175 (2026) 105233

Contents lists available at ScienceDirect

Applied Ocean Research

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

Research paper

A data-driven channel infilling rate forecasting framework with quantified uncertainty for Southwest Pass and Houston Ship Channel

Yichao Zeng [a], Mayank Chadha [b], Dania Ammar [c], Dingbao Wang [d], Magdalena Asborno [e,g], Sarah Miele [c], Charles J. McKnight [e], Natalie P. Memarsadeghi [e,f], Michael A. Hartman [e], Kenneth N. Mitchell [e], Guga Gugaratshan [c], Michael D. Todd [b], Zhen Hu [a,*]

[a] Department of Industrial and Manufacturing Systems Engineering, University of Michigan-Dearborn, Dearborn, 48128, MI, USA
[b] Department of Structural Engineering, University of California San Diego, La Jolla, 92093, CA, USA
[c] Hottinger Bruel & Kjaer Solutions LLC, Southfield, 48076, MI, USA
[d] Department of Civil, Environmental and Construction Engineering, University of Central Florida, Orlando, 32816, FL, USA
[e] Coastal and Hydraulics Laboratory, US Army Corps of Engineers, Vicksburg, 39180, MS, USA
[f] Earth System Science Interdisciplinary Center, University of Maryland, College Park, 20742, MD, USA
[g] Applied Research Associates, Inc., Vicksburg, 39180, MS, USA

ARTICLE INFO

Keywords:
Dredging
Channel infilling rate
Shoaling forecasting
Uncertainty
Machine learning

ABSTRACT

Dredging operations in navigational waterways are essential to maintain channel depth to ensure the safe passage of vessels. Existing shoaling forecasting tools often rely on heuristics or direct modeling of cumulative sedimentation, limiting lead time and predictive reliability. In this work, we propose an end-to-end data-driven forecasting framework that models the rate of change in sedimentation volume, referred to as Channel Infilling Rate (CIR), 30 days ahead, which is then integrated to recover cumulative forecasts. Our approach starts with the calculation of CIR using the Corps Shoaling Analysis Tool, developed by the U.S. Army Corps of Engineers. Following that, we employ variance decomposition-based global sensitivity analysis to identify the most influential upstream streamflow gauges from a large set of candidate gauges, which serve as inputs for CIR prediction. To model the complex relationship between upstream river discharge and the CIR of a region of interest, we explore four data-driven modeling approaches including multivariate linear regression, the long short-term memory (LSTM) networks, a hybrid model architecture that combines Convolutional Neural Network (CNN) with LSTM network (CNN-LSTM), and a Temporal Convolutional Network (TCN) model combined with LSTM network (TCN-LSTM). Finally, the trained models are employed to predict the CIR with uncertainty quantified to provide confidence intervals of the predictions. Unlike previous studies that model aggregate sedimentation volume across entire systems, we perform reach-level modeling to better capture spatial heterogeneity and hydrodynamic variability. Demonstrated on the Houston Ship Channel (HSC) and Southwest Pass (SWP), this framework represents a step toward data-adaptive shoaling forecasts that can support proactive sediment management in complex riverine environments. By providing rolling short-term forecasts with uncertainty bounds, the proposed framework can support risk-aware sediment management and proactive dredging decision-making. The comparison between HSC and SWP further demonstrates that model selection should be adapted to local survey-data availability, with simpler models favored under sparse survey conditions and LSTM-based models favored when dense survey records preserve nonlinear sedimentation dynamics.

1. Introduction

Shoaling in navigational waterways reduces channel depth and threatens the safe passage of deep-draft vessels. This disruption imposes significant logistical and economic burdens on maritime trade, including increased shipping costs and delays. To maintain navigability, the U.S. Army Corps of Engineers (USACE) maintains an estimated 40,200 km of channels across both coastal and riverine areas (Asborno et al., 2024), investing heavily in maintenance dredging, with nearly

* Corresponding author.
E-mail addresses: zyichao@umich.edu (Y. Zeng), machadha@eng.ucsd.edu (M. Chadha), dania.ammar@hbkeso.com (D. Ammar), Dingbao.Wang@ucf.edu (D. Wang), Magdalena.Asborno@usace.army.mil (M. Asborno), Sarah.Miele@hbkeso.com (S. Miele), Charles.J.Mcknight@usace.army.mil (C.J. McKnight), Natalie.P.Memarsadeghi@usace.army.mil (N.P. Memarsadeghi), Michael.A.Hartman@usace.army.mil (M.A. Hartman), Kenneth.N.Mitchell@usace.army.mil (K.N. Mitchell), guga.gugaratshan@hbkworld.com (G. Gugaratshan), mdtodd@ucsd.edu (M.D. Todd), zhennhu@umich.edu (Z. Hu).

https://doi.org/10.1016/j.apor.2026.105233
Received 12 March 2026; Received in revised form 16 August 2026; Accepted 20 August 2026
Available online 2 September 2026
0141-1187/© 2026 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

one and a half billion dollars contracted annually across the U.S. (Yeates et al., 2020). Ports and channels require periodic maintenance dredging to ensure minimum draft for safe vessel access (Mateo-Pérez et al., 2021). However, shoaling forecasts and dredging decisions often depend on non-automated and site specific rule-of-thumb practices, short-term stage trend extrapolations, or institutional knowledge gained from years of experience to estimate dredging patterns offering limited lead time (typically 10–14 days) and restricting proactive planning (Yeates et al., 2020; Asborno et al., 2024).

The accurate forecasting of shoaling remains challenging, due to highly non-linear, noisy, and non-stationary channel infilling processes, influenced by upstream flow, weather, geomorphology, and anthropogenic activities (Afan et al., 2016; Bhattacharya and Solomatine, 2006; Chou and Lin, 2020; Asborno et al., 2023). Traditional models often fail to capture this complexity, leading to inefficiencies in dredging operations (Afan et al., 2016). In this context, developing robust, data-informed forecasting models for shoaling is not only a scientific priority but also a practical necessity. Over the last two decades, Artificial Intelligence (AI) and Machine Learning (ML) techniques have become increasingly popular alternatives to traditional empirical or physics-based models. A comprehensive review by Afan et al. outlines the evolution of ML techniques in this domain, emphasizing the superior performance of hybrid and data-driven models (Afan et al., 2016). Early contributions include (Jain, 2001), who developed a sediment rating curve using feedforward neural networks, and Nagy et al. (2002), who demonstrated that multilayer perceptrons (MLP) based on backpropagation could outperform classical formulas as described by Engelund et al. (1967). Subsequent studies in Refs. Cigizoglu and Alp (2006), Alp and Cigizoglu (2007) and Jothiprakash and Garg (2009) confirmed the efficacy of artificial neural network (ANN) models in predicting suspended sediment loads across diverse environments. Melesse et al. (2011) and Kisi (2008) further highlighted ANN's superiority over multiple linear regression (MLR), autoregressive models, and conventional sediment rating curves, particularly in complex river systems. Beyond ANN, researchers have explored alternative techniques, including Adaptive Neuro-Fuzzy Inference Systems (ANFIS), Support Vector Machines (SVM), and ensemble methods. ANFIS, introduced by Ref. Jang (1993), has been shown to more effectively handle uncertainty and non-linearity in hydrological processes (Kisi, 2005; Lohani et al., 2007; Firat et al., 2009; Firat and Güngör, 2010). SVM and Least Squares SVM (LSSVM) have demonstrated competitive accuracy in sediment estimation and offered advantages in regularization and generalization (Kisi, 2012; Mateo-Pérez et al., 2021). Evolutionary techniques such as Linear Genetic Programming (LGP) and Gene Expression Programming (GEP) have been used by Refs. Guven and Kişi (2011) and Kisi and Shiri (2012), outperforming traditional methods in several sediment forecasting scenarios. Preprocessing techniques like wavelet transforms have been widely integrated with ML models to address non-stationarity in time-series data, yielding Wavelet-ANN, Wavelet-ANFIS, and Wavelet-LSSVM hybrids that capture high-frequency variability and improve peak prediction accuracy (Partal and Cigizoglu, 2008; Nourani et al., 2014; Rajaee, 2010).

In parallel to the aforementioned efforts, a growing body of work has focused specifically on dredging-related applications and shoaling prediction. Bhattacharya and Solomatine (2006) used MLPs and M5 model trees to predict sedimentation in the Port of Rotterdam by incorporating discharge, wave energy, and sediment potential. Chou and Lin (2020) developed a novel risk-informed prediction model for dredging project duration in Taiwan using support vector regression (SVR) and Monte Carlo simulations, making it the first such effort in that region. In ports and coastal systems, Mateo-Pérez et al. (2021) used SVM and Random Forests to derive satellite-based bathymetry in highly anthropized environments, providing rapid and cost-effective alternatives to sonar surveys. Graw et al. (2021) applied Random Forest Regressors for the first high-resolution global prediction of seafloor sediment density, incorporating uncertainty estimates to support real-world use. Dunkin et al. (2015) showed how combining AIS-based vessel tracking with hydrographic surveys can reveal shoaling "hot spots" and limiting depths, improving navigational safety and planning. Yeates et al. (2020) linked inland streamflow to dredging requirements using the USACE's Streamflow Prediction Tool (SPT), and found that high flows could sometimes reduce shoaling by scouring sediment, suggesting that upstream hydrology is a critical, but often underused, predictor of dredging needs. Building on these advances, Asborno et al. (2024) proposed a multivariate, multistep ML framework for shoaling prediction in the Southwest Pass (SWP) of the Mississippi River. Their use of Long Short-Term Memory (LSTM) networks, combined with physics-informed variables and Extreme Gradient Boosting (XGB) for feature selection, achieved significantly better 45-day forecasts than autoregressive integrated moving average (ARIMA) and univariate baselines. Zeng et al. (2026) recently enhanced the LSTM model for shoaling forecasting by incorporating a recursive forecasting scheme and shifted delays for input variables.

Although ML has shown promise for dredging-related applications, several important research gaps remain in existing studies. First, many existing data-driven approaches directly predict cumulative sedimentation volume or shoaling trend, which can be sensitive to datum shifts, dredging interventions, and abrupt disturbances. Only recently (Zeng et al., 2026) explicitly treated channel infilling rate (CIR) as a time-varying forecasting target in a data-driven framework and demonstrated that forecasting CIR can improve accuracy of sedimentation predictions relative to directly predicting cumulative sedimentation. Second, existing studies have not systematically examined how model choice should depend on bathymetric survey availability, even though sparse and dense survey records contain substantially different levels of temporal information. Third, even though the work by Zeng et al. (2026) represents an advance in existing studies, that study mainly focused on improving a basic LSTM model through a recursive forecasting framework with shifted delays based on aggregated sedimentation data from a data-rich region. It did not evaluate more advanced architectures that combine LSTM networks with complementary modeling components, nor did it comprehensively assess model performance under data-sparse conditions. Moreover, that study did not quantify predictive uncertainty, limiting its applicability to risk-informed dredging decisions. Finally, it relied on precomputed CIR values without fully characterizing how CIR is derived from and related to outputs from the U.S. Army Corps of Engineers' Corps Shoaling Analysis Tool (CSAT).

Motivated by the limitations of existing data-driven approaches to shoaling forecasting and built upon the work by Zeng et al. (2026), we propose a comprehensive end-to-end CIR forecasting framework that transforms raw CSAT outputs into actionable shoaling forecasts through integrated data preprocessing, model development and selection, and uncertainty quantification. We first establish an explicit and reproducible connection between CSAT-derived bathymetric information and CIR forecasting. After that, we investigate modeling strategies tailored to different levels of shoaling-survey availability. We develop and compare multiple data-driven CIR models ranging from simple linear regression models to advanced machine learning methods to evaluate performance across scenarios with limited versus abundant survey data. Because CIR predictions are subject to uncertainty arising from measurement noise and limited training data, we incorporate uncertainty quantification to provide confidence intervals for shoaling forecasts. The framework is demonstrated for two major U.S. navigation channels with contrasting data availability: Southwest Pass (SWP), which has dense bathymetric survey coverage, and the Houston Ship Channel (HSC), where bathymetric surveys are comparatively sparse. The main contributions of this paper can be summarized as follows:

- An end-to-end CIR-based shoaling forecasting framework with quantified uncertainty to predict the rate of sediment-volume change, with the predicted CIR subsequently integrated to recover cumulative sedimentation volume.

- A preprocessing procedure for CSAT/eHydro survey data is developed to construct reach-level sedimentation volume and CIR time series from historical bathymetric surveys, enabling consistent model-ready datasets to be generated across channels with different survey intervals.
- A systematic comparison of linear regression and advanced hybrid LSTM-based models for sedimentation forecasting under different conditions and two different modeling strategies (i.e., aggregated and disaggregated). The hybrid models integrate LSTM networks with convolutional neural networks (CNNs) or temporal convolutional networks (TCNs). We evaluate these models under contrasting survey-density conditions to determine how model suitability varies with data availability.
- Reach-level and aggregate forecasting strategies are evaluated to quantify the effect of spatial resolution on sediment-volume prediction accuracy.
- Predictive uncertainty in shoaling forecasts is quantified using MC-dropout across LSTM-based model variants, providing calibrated uncertainty estimates to support risk-aware dredging decisions.
- The proposed framework is demonstrated on two operationally important U.S. navigation channels, HSC and SWP, representing sparse- and dense-survey conditions, respectively, illustrating its applicability across differing data regimes.

The remainder of the paper is organized as follows: Section 2 presents the background on shoaling, the two study areas of interest (i.e., HSC and SWP), and available data sources for shoaling forecasting. Section 3 defines the problem statement, processes the CSAT data to obtain CIR, explains the input and output variables and the data processing techniques used, and outlines the solution methodology. Section 4 discuss the results and key findings in respectively the HSC region and the SWP region, while Section 5 concludes the paper with a summary of limitations and potential directions for future work.

2. Background

2.1. The process of shoaling

In the context of river channels and hydrology, shoaling is the process by which sediment, such as sand, silt, and gravel, accumulates on a riverbed, resulting in a decrease in water depth and the formation of submerged or exposed landforms called shoals. Navigable channels often necessitate dredging to remove shoals and maintain navigable depths. The term shoaling encompasses both vertical and lateral sedimentation processes (Rosati and Kraus, 2009). *Bed infilling*, or deposition, occurs when sediments settle at the bottom of a channel, directly decreasing the available navigable depth. *Lateral sediment encroachment*, by contrast, involves features such as spits, shoals, or sand waves migrating from the banks into the channel, thereby reducing its effective width. Fig. 1 schematically illustrates the effect on the channel's cross-section due to shoaling.

Shoaling is driven by a complex interplay of sediment transport dynamics, hydro-meteorological forces, sediment properties, geological features, and human activities. At its core, it involves the nonlinear interaction between water flow and sediment particles, from detachment to deposition. Density currents associated with freshwater discharges can shift transport sediments downstream, while salt-wedge intrusions in estuaries can transport them upstream. Seasonal stratification can further promote sediment retention in deeper channels (Bhattacharya and Solomatine, 2006; Asborno et al., 2023). Wind, waves, and currents further influence sediment movement, with reduced river discharge and colder temperatures favoring deposition. Sediment traits, like size, density, and concentration, affect settling behavior, while biological activity can enhance or stabilize deposits. Geographical features and anthropogenic factors such as land use changes and dredging practices also shape sediment delivery and accumulation.

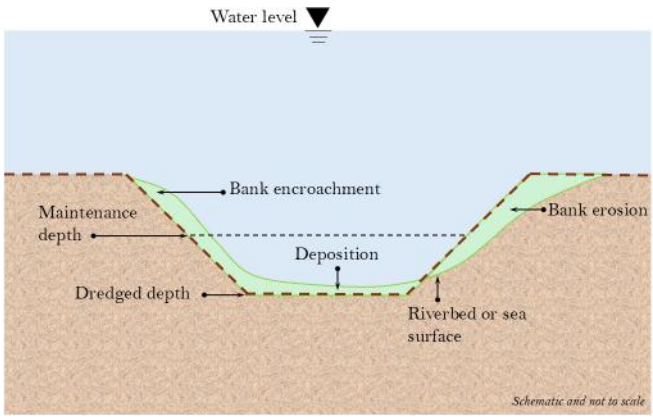


Fig. 1. Schematic diagram showing progressive morphological changes in channel geometry.

As discussed above, shoaling is a complex process influenced by numerous factors. However, consistent observations and reliable predictive models are not available for all of these factors. Previous studies have demonstrated that upstream river discharge contains important information about shoaling in downstream reaches, and streamflow forecasting has been extensively studied in the hydrologic community (Yeates et al., 2020). Therefore, this study uses upstream discharge as the primary predictor of shoaling. We acknowledge that excluding factors such as tides, waves, and salinity is a limitation of the proposed framework. These factors can be incorporated in the future as reliable observational data or suitable predictive models become available. We expect that including those additional important factors could further improve the performance of the presented framework.

2.2. Study sites: HSC and SWP

2.2.1. The Houston Ship Channel (HSC)

The HSC, as shown in Fig. 2, is a navigable waterway located in the state of Texas and the USACE Galveston District (SWG). It extends 53 miles upstream from Galveston Bay to the Turning Basin in Houston, facilitating the movement of vessels from the Gulf of Mexico to the city. In 2005, the SWG undertook an expansion project for the HSC, resulting in varying authorized channel dimensions along its length: (1) a depth of 45 ft and a width of 530 ft for the segment extending from Bolivar Roads to Boggy Bayou, (2) a depth of 40 ft and a width of 300 ft for the segment extending between Boggy Bayou and Sims Bayou, and (3) a depth of 36 ft and a width of 300 ft for the third segment extending between Sims Bayou and the Main Turning Basin (Kabir et al., 2022). The upstream gauge stations and their corresponding IDs are also shown in this figure. These stations serve as the input variables for the GSA described in Section 3.3.

Based on USACE National Channel Framework (NCF) (U.S. Army Corps of Engineers, Mobile District, 2025), the HSC is divided into 12 main sections, also known as reaches. In addition to being a deep-draft navigation channel, the HSC is home to the Port of Houston—one of the busiest ports in the United States in terms of total tonnage. It also connects the nation's largest petrochemical complex to global markets. These attributes make it the country's leading waterway by waterborne tonnage, supporting 3.37 million jobs nationwide and generating $906 billion in economic impact across the United States (Port Houston, 2024).

Maintaining the HSC's safety, efficiency, and navigability, especially given the high vessel traffic, is therefore critical. This is achieved through ongoing improvement and maintenance strategies, including new work dredging for channel expansions and maintenance dredging to address shoaling hotspots. The HSC remains an active expansion

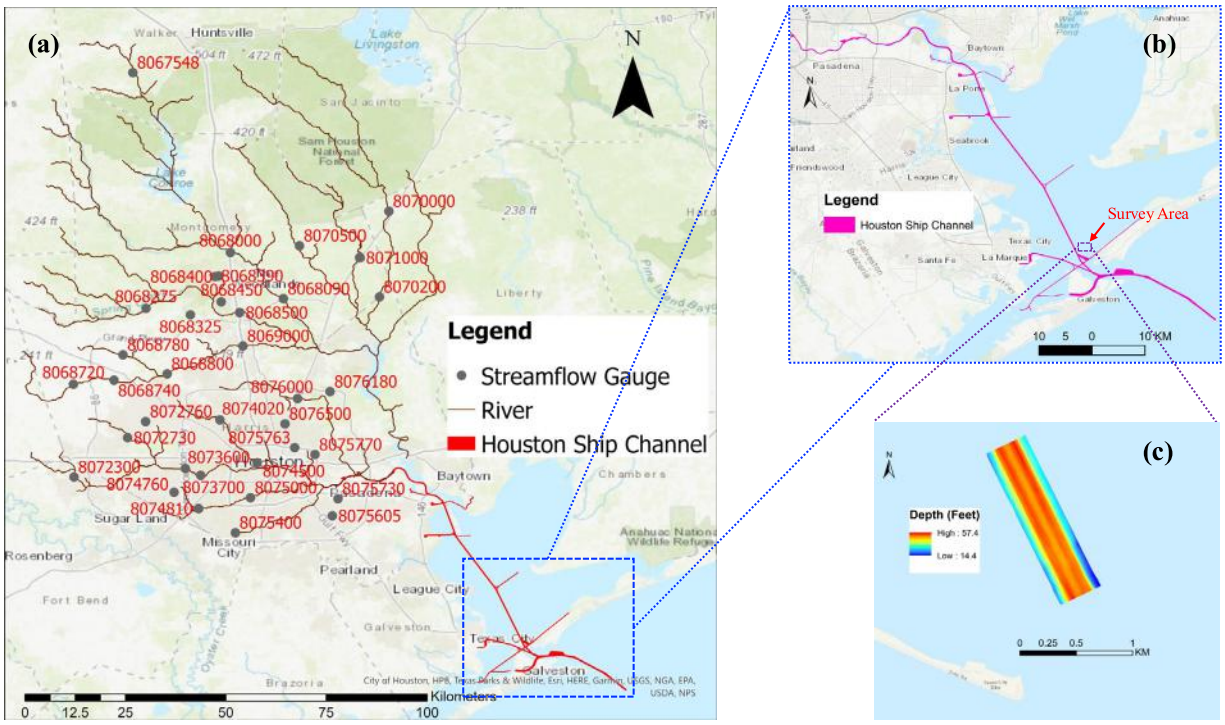


Fig. 2. Houston Shipping Channel (HSC). (a) Buffalo Bayou and San Jacinto River Basins showing data collection stations considered for sensitivity analysis to select inputs for building the forecasting model for the HSC region. (b) Zoomed-in view of the HSC. (c) Bathymetric survey for a portion of channel in HSC.

project, which aims to widen certain segments from 530 to 700 ft and deepen upstream segments to 46.5 ft (U.S. Army Corps of Engineers, Galveston District, 2025; Port of Houston, 2025).

2.2.2. The Southwest Pass (SWP)

Located at the USACE New Orleans District (MVN), the SWP, as shown in Fig. 3, is a critical navigable channel in the United States, located in the state of Louisiana. It extends from the mouth of the Mississippi River to the Gulf, serving as the primary gateway for maritime traffic traveling between the Gulf and the interior states (Ayres et al., 2014), particularly where the Port of New Orleans and other key ports are located. Similar to the HSC region, the upstream gauge stations and their corresponding IDs are also shown in this figure. These stations serve as the input variables for the GSA described in Section 3.3.

The SWP currently maintains a navigable depth of 50 ft, with a recommended draft limit of 45 ft (U.S. Department of Commerce et al., 2025). Its width ranges from 600 to 750 ft along a stretch of approximately 35 miles (Hartman et al., 2022), divided into 13 NCF reaches. These characteristics make the SWP the nation's premier waterway for deep-draft vessels, with over 250 million tons of cargo passing through annually—transported by an average of 13 such vessels per day (Hartman et al., 2022). Given its critical role, any disruption to the route of these oceangoing vessels can result in significant economic losses. As such, the SWP is a top priority for the USACE and MVN, particularly in terms of maintaining full navigability through timely dredging operations.

2.3. Data sources and existing tools

2.3.1. Bathymetric surveys and ehydro tool

Bathymetric surveys, which measure channel depth to support navigation, engineering, and sediment management, and is hosted on a centralized data platform known as *eHydro* (Kress, 2021; U.S. Army Corps of Engineers, Mobile District, 2025). *eHydro* is an enterprise-level tool developed by USACE to process, store, and standardize hydrographic survey data nationwide. It organizes bathymetric data by individual channel reaches as defined by the National Channel Framework (NCF)—an enterprise Geographic Information System (eGIS) database that documents congressionally authorized navigation channels (USACE, 2022). *eHydro* compiles the survey data into structured files with uniform spatial grids, maintaining consistent X and Y coordinates while capturing temporal changes in bathymetry.

Bathymetric surveys (Irish and White, 1998; Byrnes et al., 2002) capture underwater terrain by measuring the horizontal position (X-Y) and the bottom elevation using mechanical or acoustic methods, with modern systems offering high precision. Depths are often measured using single-beam or multi-beam echo sounders, such as the Navisound 210, which provide centimeter-level vertical precision across a range of acoustic frequencies (Mateo-Pérez et al., 2021). These measurements are georeferenced using Differential GPS and corrected for tidal variations to ensure accuracy (Mateo-Pérez et al., 2021; Byrnes et al., 2002). The resulting point data are interpolated into digital terrain models of the channel bed, typically referenced to standard projection systems like UTM/WGS84 (Mateo-Pérez et al., 2021). Figs. 2(c) and 3(c) illustrate respectively an instance of a Bathymetric survey for a portion of the channel in the HSC and SWP region. Survey accuracy is influenced by sea state, instrument calibration, and vessel motion, with errors potentially arising from random, systematic, or human sources. These are managed through rigorous calibration and quality control to limit such errors. The data are typically processed into Triangulated Irregular Networks (TINs) or gridded models, with TINs preferred when data are sparse. Interpolation and terrain irregularities introduce uncertainty, which is minimized through repeated measurements and surface comparisons.

2.3.2. Corps Shoaling Analysis Tool (CSAT)

The CSAT calculates channel infilling rates and shoaling volumes using historical channel surveys (available through *eHydro* (U.S. Army Corps of Engineers, Mobile District, 2025)) and provides a first-order prediction of future dredging volumes (U.S. Army Corps of Engineers, 2024). The tool consists of two parts. The first is the *CSAT input*, which contains pre-processed and cleaned bathymetric data from eHydro, stored in NetCDF format and includes details from historical surveys.

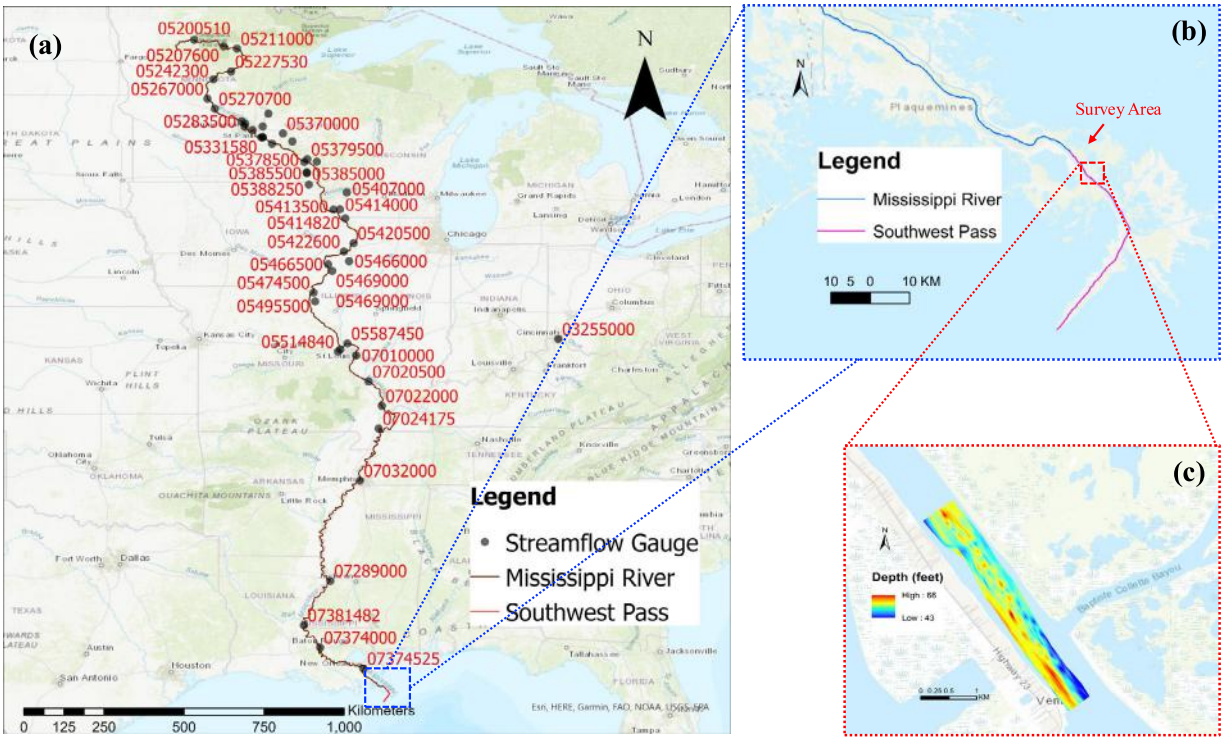


Fig. 3. The Southwest Pass (SWP). (a) Mississippi River and its tributaries showing discharge and the locations of data collection stations considered for sensitivity analysis to select inputs for building the forecasting model for the SWP region. (b) Zoomed-in view of SWP. (c) Bathymetric survey for a portion of channel in SWP.

Since these surveys span wide geographic areas, the region is divided into grids referred to as bathymetric survey points (typically 10-by-10 ft), with each point representing the center of a grid cell and identified by its *latitude* and *longitude*. Each survey, identified by a unique *survey ID*, provides elevation data measured as positive downward. In essence, the elevation value is treated as a function of the *survey ID*, *survey time*, and the *latitude* and *longitude* of each bathymetric survey point (see Figure 10 in Bain et al. (2025)).

The second part of the tool uses the provided Python code (U.S. Army Corps of Engineers, 2024) to process the *CSAT input* and generate various *CSAT output* files, including: (a) a table of average, minimum, and maximum CIR for each bathymetric survey point; (b) a table of volume differences between surveys by reach; and (c) a table of forecast dredging volumes by reach, which includes projected volumes for various cut depths at 6-month intervals, among others as listed in Section 4 of Ref. Bain et al. (2025). For a fixed cut depth, the forecast is provided at 6-month intervals, covering periods from 6 to 60 months. However, these forecasts are first-order, piecewise linear projections informed solely by past sedimentation volume change data. They are intended to indicate linear sedimentation trends but lack more informed short-term projections (e.g., 30 to 45 days ahead, which this paper aims to provide), which are more useful for immediate decision-making.

The *CSAT output* in its current form (U.S. Army Corps of Engineers, 2024) does not provide historical time series of sedimentation volume. Therefore, it is not the *CSAT output* that interests us, but rather the *CSAT input*, which contains data from historical bathymetric surveys derived from *eHydro*. The *CSAT input* (in the ascending time order) is utilized to derive the historical sedimentation volume time series at the reach level.

2.3.3. Survey frequencies in HSC and SWP

When comparing the number of conducted hydrographic surveys between the HSC and the SWP — based on a portion of the *CSAT input* database, which compiles all hydrographic surveys for each reach into a single file (Dunkin et al., 2018) — the SWP shows a significantly higher number of surveys per reach across different years. From 2012 to 2024, the SWP recorded 5.65 times more surveys than the HSC, with a total of 21,208 survey entries compared to just 3751 for HSC. While district-specific survey processing practices (MVN for SWP and SWG for HSC) could introduce procedural differences, a comparison of representative surveys from both regions showed that the spatial coverage of individual surveys evaluated using reach length, shape area, and shape length, does not systematically vary with survey frequency. In the SWP, reach lengths are relatively uniform (approximately 2.7–3 miles), whereas in the HSC they range more widely (0.3–11.3 miles). Although HSC reaches vary widely in size, as reflected in their corresponding survey shape area and length, this variability does not correlate with the number of surveys conducted. Moreover, even for smaller HSC reaches, whose survey shape area and length are less than those of SWP reaches, the number of conducted surveys remains significantly lower than in the SWP. Thus, the disparity primarily reflects the SWP's inherently dynamic morphology and rapid shoaling behavior, which compel the MVN to perform frequent hydrographic surveys to monitor and maintain navigable conditions (Hartman et al., 2022). Fig. 4 illustrates the number of annual hydrographic survey entries for the SWP and the HSC from 2012 to 2024. The bar chart (left y-axis) compares the total number of surveys conducted in each region per year, while the overlaid line plot (right y-axis) depicts the ratio of SWP to HSC survey counts. This visualization highlights the consistently higher frequency of surveys in the SWP, underscoring the channel's more aggressive shoaling behavior and the increased monitoring efforts required to maintain navigability.

2.3.4. Streamflow gauges near HSC and SWP

Previous studies by Yeates et al. (2020) and Asborno et al. (2024) highlight the critical role of upstream streamflow measurements in influencing shoaling. In the United States, streamflow measurements are primarily sourced from the National Water Information System (NWIS) maintained by the U.S. Geological Survey (USGS) (https://waterdata.usgs.gov/nwis). Supplementary, shorter-duration data are

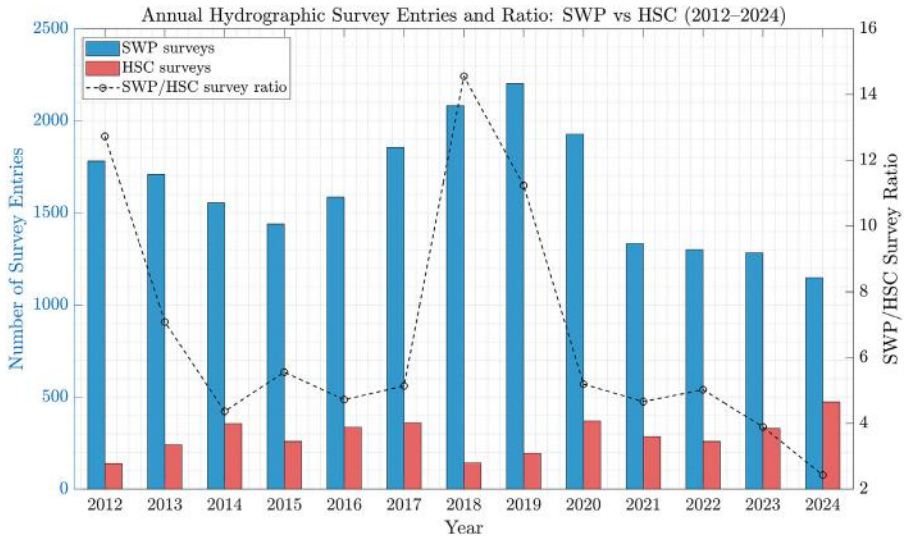


Fig. 4. Annual hydrographic survey entries for the Southwest Pass (SWP) and Houston Ship Channel (HSC) from 2012 to 2024.

sometimes provided by local state or municipal agencies, though these are generally considered less reliable than the NWIS records (Ghimire et al., 2023; Zhao et al., 2024). We considered 36 gauge stations in the HSC (Buffalo Bayou and San Jacinto River) basin and 43 gauge stations in the SWP (Mississippi River) basin for sensitivity analysis. Figs. 2(a) and 3(a) illustrate respectively the gauge stations along with the corresponding station IDs for the HSC and SWP regions.

It is worth noting that, although other predictors, such as turbidity, have been used to forecast shoaling in prior work (Asborno et al., 2024) (see Section 1), we focus primarily on streamflow as the main driver of shoaling in this study. This choice is motivated by both physical relevance and operational practicality. Upstream streamflow is routinely monitored, broadly available, and can be forecast using hydrologic or physics-based simulations; river-discharge prediction also remains an active area of research in hydrology. These characteristics make streamflow a suitable predictor for dredging decision support. In contrast, variables such as turbidity, tides, winds, wave climate, salinity, rainfall, and storm-related indicators may also influence channel infilling processes, but they are not always consistently available across navigation channels or directly predictable with the same operational reliability. Their exclusion should therefore not be interpreted as evidence that they are unimportant features. Rather, the present study focuses on establishing an operationally feasible CIR-based forecasting framework using routinely available and forecastable predictors. Future work will extend the proposed framework by incorporating additional hydro-meteorological and sediment-related variables to evaluate their marginal contribution to CIR prediction under different channel conditions.

2.4. Challenges and needs of data-driven models for shoaling forecasting

Shoaling forecasting remains difficult to implement in practice for several reasons, summarized below.

- First, physics-based sediment transport models, although theoretically capable of simulating sediment dynamics, are impractical for large-scale and long-term operational forecasting in complex systems, such as the HSC and SWP shown in Figs. 2 and 3. Process-based numerical models have also been widely used to investigate sediment transport mechanisms and system-scale sediment dynamics in complex coastal and estuarine environments (e.g., Uphues et al., 2025; Pourzangbar and Brocchini, 2022; Dodd et al., 2008). These models require extensive bathymetric detail, boundary conditions, and computational resources, making them unsuitable for routine forecasting at the spatial and temporal scales required by navigation authorities. To clarify the practical role of the proposed approach, Table 1 summarizes the distinction between process-based sediment transport models and the proposed CIR-based forecasting framework.
- Second, while data-driven models offer a promising alternative for shoaling forecasting, training them has proven challenging because the data are affected by various sources of noise. For example, dredging interventions introduce abrupt, extreme jumps in the CSAT data, obscuring the underlying natural shoaling process. Consequently, using the raw CSAT data directly for model training can lead to substantial prediction errors.
- Third, bathymetric survey frequency varies by region. In areas such as the SWP, surveys are conducted frequently, producing a densely sampled sedimentation-volume time series that captures finer temporal variability. In contrast, the HSC has a much sparser survey record, yielding a sedimentation-volume curve that is effectively a coarse, piecewise-linear approximation. These differences require distinct data-driven modeling strategies; no universal approach is suitable for all scenarios.
- Fourth, although upstream river discharge is a critical predictor for shoaling forecasting (Figs. 2(a) and 3(a)), discharge data are available from a large number of gauging stations. This high-dimensional input space complicates training of data-driven shoaling models. Moreover, more features are not necessarily better: if some predictors are weakly related to the quantity of interest, including them can degrade model performance rather than improve it. Therefore, it is important to identify the most relevant stations among the many upstream discharge gauges.
- Fifth, while CSAT provides valuable insights into past shoaling patterns, it does not offer nonlinear predictive capabilities. Since shoaling behavior is nonlinear in nature, nonlinear predictive capabilities can better support forward-looking decision-making for navigation management than just linear patterns (Asborno et al., 2024; Yeates et al., 2020).

In the following section, we propose a systematic shoaling forecasting framework that complements existing survey-based tools and enables data-driven prediction of shoaling volume. Unlike approaches that focus only on trend, the proposed framework forecasts the actual shoaling-volume values.

Table 1
Practical comparison between process-based sediment transport models and the proposed CIR-based forecasting framework.

Aspect	Process-based models	Proposed CIR-based framework
Purpose	Process-level simulation of sediment transport	Forecasting of CIR and sedimentation volume
Data	Bathymetry, boundary conditions, sediment properties, hydrodynamic forcing, and calibration data	CSAT/eHydro sedimentation records and forecastable upstream streamflow data
Setup	High setup and computational burden; site-specific calibration is required	Moderate setup and low prediction cost after training
Uncertainty	Ensemble or parameter-uncertainty simulations	MC-dropout for uncertainty quantification
Limitation	Difficult to deploy routinely without detailed physical inputs	Sensitive to survey density and predictor availability

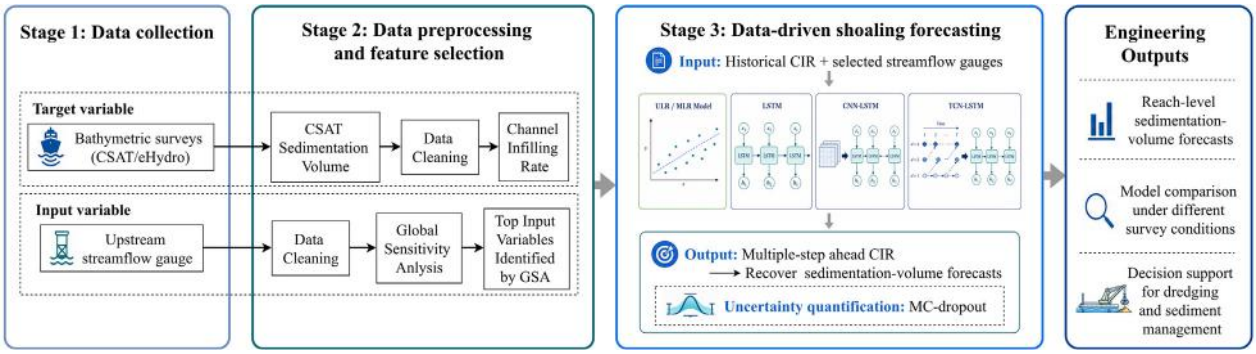


Fig. 5. Proposed framework of the solution approaches.

3. Methodology

3.1. Overview

Fig. 5 provides an overview of the proposed framework. As shown in this figure, it consists of three main stages.

- The first stage focuses on data collection for both target variable (i.e., raw NetCDF files from the eHyrdo database (Kress, 2021; U.S. Army Corps of Engineers, Mobile District, 2025)) and the explanatory inputs (i.e., streamflow data from the upstream of the region of interest).
- The second stage centers on data preprocessing, which converts the raw datasets obtained from the first stage into physically meaningful and model-ready inputs. For the target variable, sedimentation volume is first computed using the CSAT tool, followed by data cleaning and transformation into the CIR. On the input side, gauge data go through cleaning and quality control before being analyzed to select the most important features. This step identifies the most influential predictors, allowing the modeling effort to focus on a reduced and physically interpretable set of input variables.
- The final stage develops data-driven shoaling forecast models using the processed CIR data and selected input features. Modeling strategies are tailored to the survey frequency in the region of interest. Prediction uncertainty is also quantified to produce confidence intervals for the shoaling forecasts.

3.2. Data preprocessing

Before analysis, the *CSAT input* data must be preprocessed. As noted above, this paper forecasts the CIR rather than sedimentation volume directly, because CIR captures the temporal dynamics of sedimentation and helps distinguish gradual, natural shoaling from abrupt shifts induced by external interventions. Next, we describe the preprocessing steps for the CSAT input and the computation of the CIR after that.

3.2.1. Computation of channel infilling rate

There are several data issues that need to be addressed before the computation of the CIR. (a) For a given reach, bathymetric surveys typically focus on the channel; grid cells outside the channel are often unsurveyed and are therefore recorded as either a very large value (e.g., $\sim 10^{38}$) or *NaN*. (b) Multiple surveys (distinguished by survey ID) may be recorded on the same day. (c) Surveys are irregularly spaced in time and can be sparse. Notably, CSAT applies region-specific, preprogrammed datum corrections.

These issues present both a *spatial* (item *(a)*) and a *temporal* (items *(b)* and *(c)*) challenge. To address the spatial issue, we first replace all extremely large values (e.g., $\sim 10^{38}$) with *nan*. For the temporal issue, we apply linear interpolation over the *nan* values to estimate reasonable elevation values for the grids that are part of the channel but were not surveyed in certain instances. Grids that have *nan* values across all survey IDs are assumed to be outside the channel and are excluded from volume calculations. When there are multiple surveys for the same day (item *(c)*), we combine them by computing the mean elevation value (excluding *nan*) for each grid. Finally, the elevation values are *linearly* interpolated over time, yielding continuous elevation data at a daily frequency (which can be adjusted to weekly or monthly if desired) for each grid. This process is schematically illustrated in the flowchart below (Fig. 6).

The sedimentation volume of the reach at time t is obtained by first computing the volume associated with each grid and then summing the volumes across all grids within the reach to obtain the total volume. Let $v_i(t)$ denote the volume of the grid i at time t, and let $d_e(t)$ represent the elevation of the sea surface (obtained from *CSAT input*) and d_m be the maintained depth. Both quantities are assumed to be positive in the downward direction, measured in feet from a fixed datum. For a grid size of 10-by-10 ft, the volume $v_i(t)$ is given by

$$v_i(t) = \begin{cases} 0, & \text{if } d_e(t) - d_m > 0 \\ \frac{(d_e(t)-d_m)\times 10\times 10}{27}\text{yard}^3, & \text{if } d_e(t) - d_m \le 0. \end{cases} \quad (1)$$

The two scenarios presented in Eq. (1), one corresponding to the channel bed being below the maintenance depth (i.e., $d_e(t) - d_m > 0$),

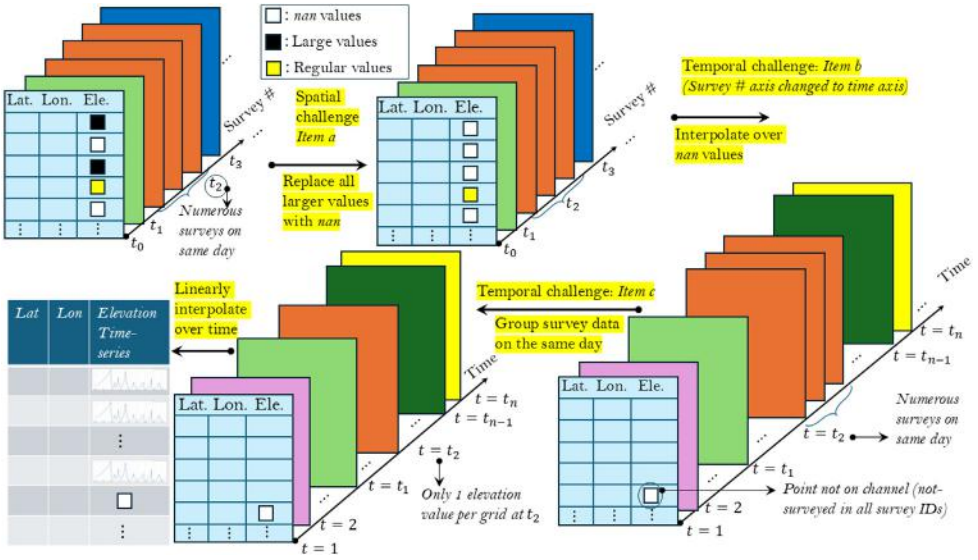


Fig. 6. Pre-processing of historic bathymetry survey data extracted from *CSAT input*.

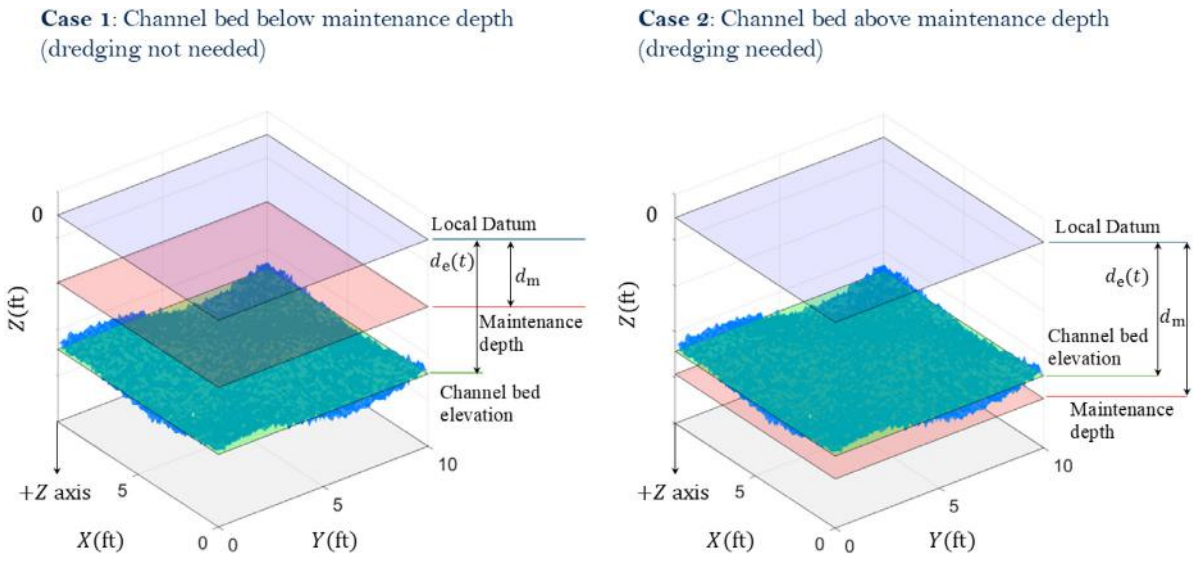


Fig. 7. Calculating the sediment volume for a fixed survey grid and time instance.

and the other where the channel bed is above the maintenance depth (i.e., $d_e(t) - d_m \leq 0$), are schematically illustrated in Fig. 7.

The volume at the reach level, denoted by reach r, at time t, represented as $V_{raw_r}(t)$, is computed by summing the contributions from all N_{grids} individual grids, as defined in Eq. (2):

$$V_{raw_r}(t) = \sum_{i=1}^{N_{grids}} v_i(t). \tag{2}$$

The raw volume $V_{raw_r}(t)$ is often contaminated by short-term, high-frequency fluctuations resulting from measurement noise and interpolation artifacts in survey data. To mitigate this, we applied the *Savitzky-Golay filter* (SGF) with a 30-day window size to obtain a smoothed volume time series at the reach level. Let $V_r(t)$ denote the smoothed sedimentation volume for reach r. Mathematically, this can be expressed as:

$$V_r(t) = \mathrm{SGF}(V_{raw_r}(t), 30). \tag{3}$$

The *SGF* is a digital smoothing technique that fits a low-degree polynomial (typically of degree 2 or 3) to a moving window of consecutive data points using least squares. Fig. 8 shows a comparison of the raw sediment volumes with their SGF-processed counterparts for a HSC reach and a SWP reach respectively. As shown in this figure, unlike simple moving averages, SGF preserves the shape, local trends, and features, such as peaks and inflection points, of the data while reducing noise. This makes it particularly suitable for our application, where preserving the underlying trend in sedimentation volume is crucial for reliable CIR computation and downstream modeling. The 30-day window was chosen to strike a balance between removing short-term noise and retaining meaningful temporal variations in sedimentation dynamics.

Occasionally, it becomes important to consider combinations of reaches. Let C denote the set of reaches to be combined or aggregated, consisting of $N_{reaches}$ reaches. The total sedimentation volume for the combined set can be obtained by summing the volumes of the

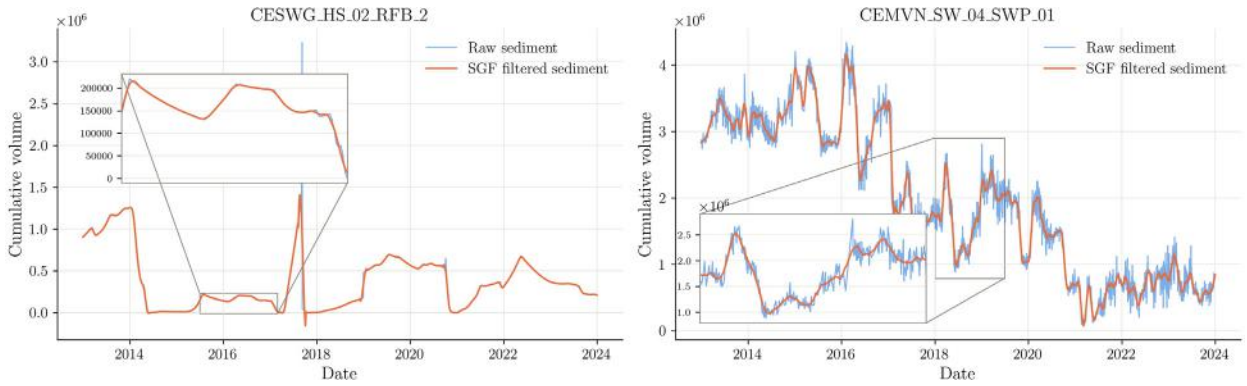


Fig. 8. Comparison of raw sediment volume data and sediment data after using SGF for a HSC reach (i.e., CESWG_HS_02_RFB_2) and a SWP reach (i.e., CEMVN_SW_04_SWP_01), respectively.

individual reaches.

$$V_C(t) = \sum_{r=1}^{N_{\text{reaches}}} V_r(t). \tag{4}$$

The CIR, computed using a lag of Δ days at the reach level (denoted by $R_{\Delta,r}(t)$) and at the aggregate level (denoted by $R_{\Delta,C}(t)$), is given by:

$$R_{\Delta,r}(t) = \frac{V_r(t) - V_r(t-\Delta)}{\Delta}, \tag{5a}$$

$$R_{\Delta,C}(t) = \frac{V_C(t) - V_C(t-\Delta)}{\Delta}. \tag{5b}$$

In summary, the CIR calculation involves four main steps. First, at the grid level, extreme values in the survey data are removed, and missing values in the time series are estimated using linear interpolation. Second, at the reach level, sedimentation volume is calculated by summing the volumes across all individual grid cells. Third, the reach-level volume data are smoothed using a Savitzky–Golay filter to reduce noise. Finally, the reach-level volumes are aggregated to obtain regional volumes, and the CIR is calculated using a finite-difference method. Alternatively, the CIR can be computed at reach level based on the reach-level sediment volume data.

3.2.2. Identification of abrupt CIR anomalies caused by non-natural human activities

We used an *Interquartile Range* (IQR)–based approach to filter abrupt shifts induced by external non-natural interventions, since the goal of this paper is to model the natural shoaling process. The procedure consists of the following four steps:

1. *Compute the quartiles:* Obtain the 25th and 75th percentile values of the CIR time series (either $R_{\Delta,r}(t)$ or $R_{\Delta,C}(t)$), denoted as Q_{25} and Q_{75}, respectively.
2. *Compute the interquartile range (IQR):* The IQR is defined as the difference between the 75th and 25th percentiles:

   $$\text{IQR} = Q_{75} - Q_{25} \tag{6}$$

3. *Determine the outlier bounds:* Compute the lower and upper cutoff thresholds, denoted as Q_{lower} and Q_{upper}, using the following expressions:

   $$Q_{\text{lower}} = Q_{25} - q\,\text{IQR} \tag{7a}$$

   $$Q_{\text{upper}} = Q_{75} + q\,\text{IQR} \tag{7b}$$

   Here, q is the outlier threshold chosen by the modeler. A smaller value of q results in a more sensitive outlier filter.
4. *Identify the extreme events:* Values of the CIR time series that are either above Q_{upper} (e.g., due to sudden volume addition from extreme rainfall) or below Q_{lower} (e.g., due to dredging) are identified as extreme values, and the corresponding time instances are omitted for the purpose of training. Let N_{time} denote the total number of time instances in the CIR data after removing (i.e., not counting) the outliers.

Figs. 9 and 10 illustrate the IQR-based filter applied to the aggregated CIR, denoted by $R_{\Delta,C}$, using $\Delta = 1$ days. We set $q = 2$ in this case study. This outlier threshold value was chosen because it provides a good balance between sensitivity and robustness in filtering abnormal data points. Several threshold values were tested (e.g., $q = 1, 1.5, 2, 2.5, \ldots, 10$), and it was found that $q = 2$ effectively eliminated most abnormal trends, such as sudden artificial drops due to dredging or spikes from extreme events, while preserving the meaningful variability in sediment dynamics. It should be noted that this filtering step is designed to reduce the influence of abrupt discontinuities in the model-training target, rather than to provide definitive attribution of each flagged point to dredging, measurement error, or natural extreme sedimentation processes. For reproducibility purposes, the corresponding quartile statistics are reported as follows:

- **HSC**: $Q_{\text{lower}} = -28,854.70$ yd^3/day, $Q_{\text{upper}} = 30,946.55$ yd^3/day for $q = 2$.
- **SWP**: $Q_{\text{lower}} = -152,820.23$ yd^3/day, $Q_{\text{upper}} = 159,095.30$ yd^3/day for $q = 2$.

Since the survey data are linearly interpolated, the ability of the sedimentation volume time series to capture the true volume evolution for a given reach depends heavily on the density of available surveys. When surveys are sparse, the resulting sedimentation volume curve reduces to an approximate piecewise linear fit. While this may capture the overall trend at the system level, it serves only as a first-order approximation of short-term variations.

3.3. Feature selection

Before we present the proposed solution strategy for forecasting CIR in Section 3.4, we first outline the approach employed to select input features used in the modeling framework. As mentioned in Section 2.4, we use the upstream streamflow gauge data as inputs (i.e., predictors) for CIR forecasting. However, the large number of available streamflow gauges (see Figs. 2(a) and 3(a)) makes the problem high-dimensional, which poses challenges to the data-driven CIR modeling. In order to identify the most informative streamflow gauges for CIR forecasting, the variance decomposition-based Global Sensitivity Analysis (GSA) is employed to select the N_{sg} gauges with the highest sensitivity to the CIR time series.

Let $G_i(t)$ (or, equivalently, its abbreviated form G_i) denote the streamflow gauge time series for the ith upstream gauge, where $1 \le i \le N_{\text{g}}$ and N_{g} denote the total number of upstream gauges considered for the sensitivity analysis. To ensure compatibility with the CIR data, the

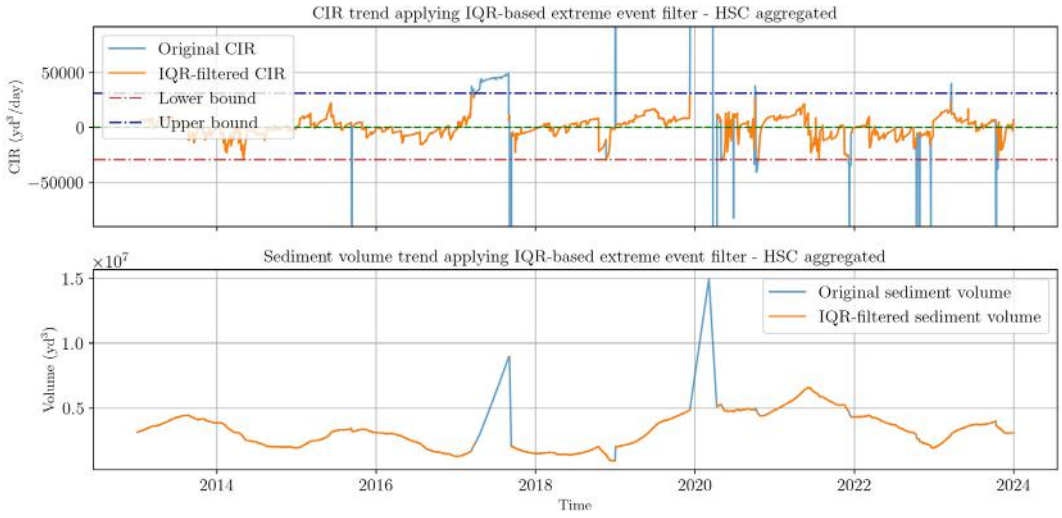


Fig. 9. IQR-based extreme event filter for HSC region considering aggregated sedimentation data.

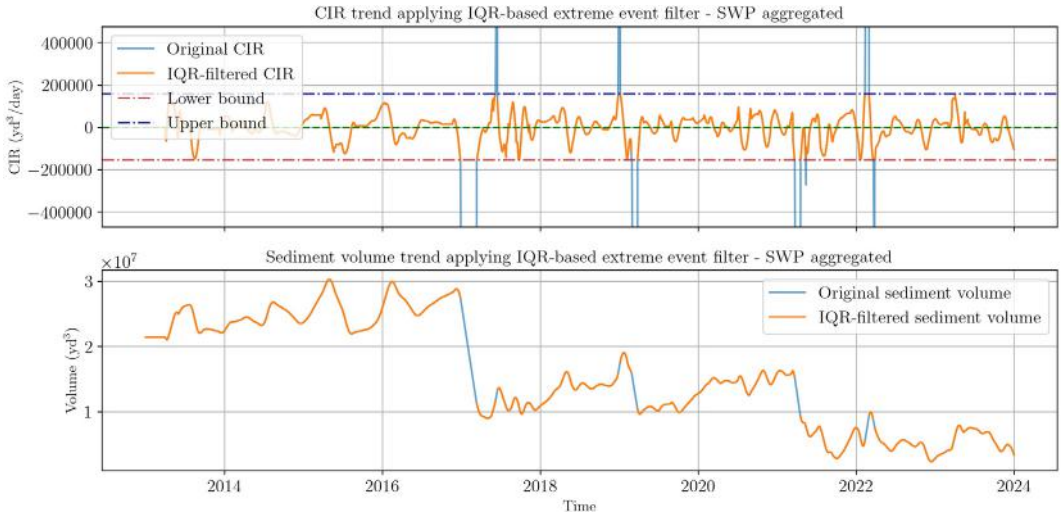


Fig. 10. IQR-based extreme event filter for SWP region considering aggregated sedimentation data.

time instances corresponding to extreme event outliers were removed, resulting in each $G_i(t)$ vector containing N_{time} elements. These time series were then arranged as columns to form a matrix of streamflow gauge data, denoted by $\mathbf{G} \in \mathbb{R}^{N_{\text{time}} \times N_g}$. The objective of GSA is to identify which columns of $\mathbf{G}$ (i.e., which upstream gauges) are most sensitive to the CIR data, which serves as the output variable in the forecasting models (more details on modeling in Section 3.4). For description, we use reach-level CIR data, denoted by $R_{\Delta,r} \in \mathbb{R}^{N_{\text{time}}}$.

In variance-based GSA (most commonly known as Sobol' sensitivity analysis), the variance of the reach-level CIR data $R_{\Delta,r}$ is decomposed into contributions from individual inputs $G_i, \forall i = 1, \ldots, N_g$ and their interactions as follows (Li and Mahadevan, 2016).

$$\text{Var}(R_{\Delta,r}) = \sum_{i=1}^{N_g} V_i + \sum_{1 \le i < j \le N_g} V_{ij} + \sum_{1 \le i < j < k \le N_g} V_{ijk} + \cdots + V_{1 \ldots N_g}, \tag{8}$$

where V_u in which $u \in \{1, \ldots, N_g\}$ is the variance contribution form the interaction involving exactly those input factors in u. For instance, $V_i, i = 1, \ldots, N_g$ are the variance contributions involve only the ith streamflow gauge G_i.

Based on the above variance decomposition, two types of Sobol' indices are mostly widely used, namely the *First-Order Sobol' Index* and the *Total Sobol' Index*. The First-Order Sobol' Index quantifies the main effect of an individual input vector $G_i \in \mathbb{R}^{N_{\text{time}}}$ on the output vector $R_{\Delta,r}$, representing the fraction of the output variance attributable to G_i alone. It is defined as:

$$S_i = \frac{\text{Var}_{G_i}\left[\mathbb{E}\left(R_{\Delta,r} \mid G_i\right)\right]}{\text{Var}(R_{\Delta,r})}, \tag{9}$$

where $\text{Var}_{G_i}\left[\mathbb{E}\left(R_{\Delta,r} \mid G_i\right)\right]$ is the variance of the conditional expectation of $R_{\Delta,r}$ given G_i, and $\text{Var}(R_{\Delta,r})$ is the total variance of the output.

A high value of S_i indicates that the corresponding streamflow gauge G_i has a strong influence on the output variability and is therefore a strong candidate for inclusion in the forecasting model. Although the *Total Sobol' Index* captures both the main effects and interaction effects among variables, it is computationally more expensive to estimate. In our case, the primary goal was to identify the most influential individual gauges with minimal computational burden. We adopted the approach proposed by Ref. Li and Mahadevan (2016) to estimate the *First-Order Sobol' Index* S_i. This is a model-free method that directly computes sensitivity indices from the input and output data without requiring an explicit surrogate or simulation model linking them.

For the proposed model, additional computation was necessary because it operates on sequences of input data segments and produces

corresponding output sequences. Since the model's predictions and sensitivity may vary across different positions within these sequences, it is important to aggregate the GSA indices over all time steps to obtain a comprehensive measure of input importance. Firstly, the local Sobol' index for each input variable $i \in \{1, \ldots, N_g\}$ at each time step $m \in \{1, \ldots, W_{\text{past}}\}$ within the input sequence can be defined as follows:

$$S_{i,m,n} = \text{Sobol}\left(G_{i,t-w+m} \rightarrow R_{\Delta,r,t+n}\right), \tag{10}$$

where $G_{i,t-w+m}$ denotes the value of the ith input variable at the mth time step in the input sequence and $R_{\Delta,r,t+n}$ represents the nth time step in the output sequence.

Building on the calculation of the local Sobol' index, the aggregated GSA indices are calculated using the following equation:

$$S_i^{(\text{agg})} = \sum_{m=1}^{W_{\text{past}}} \sum_{n=1}^{W_{\text{pred}}} S_{i,m,n}, \tag{11}$$

We then selected the top N_{sg} most sensitive streamflow gauges, based on each variable's aggregated Sobol' index $S_i^{(\text{agg})}$, and denote the kth most sensitive gauge as $\tilde{G}_k$, where $1 \le k \le N_{\text{sg}}$. In this paper, we tried different N_{sg} values and finally considered two input configurations, top 3 and top 8 gauges, to balance parsimony and coverage: the top 3 represent the most dominant controls, while the top 8 incorporate additional gauges to capture broader spatial variability. It is worth mentioning that the GSA-selected gauges are not intended to be universal or permanently fixed predictors. Rather, they represent the most influential streamflow stations for the specific study region, historical period, preprocessing procedure, and forecast horizon considered in this study. To evaluate the robustness of feature selection, the GSA procedure can be repeated over different training horizons or rolling calibration windows. The stability of the selected gauges can then be quantified using top-k overlap or rank-correlation measures between the gauge rankings obtained from different windows. For example, a high overlap among the top-ranked gauges indicates that the dominant upstream hydrological controls are stable over the considered period, whereas large changes in the rankings suggest nonstationary hydrological controls or changing channel-response behavior.

3.4. CIR forecasting problem description and solution strategies

3.4.1. Modeling strategies

In this paper, our objective is to predict sedimentation volume using two different modeling strategies.

1. **Strategy 1 (Disaggregated Strategy)**: This strategy treats each reach independently and forecasts the sedimentation volume at the individual reach level. The forecasted values are then aggregated to obtain the cumulative volume across multiple reaches.
2. **Strategy 2 (Aggregated Strategy)**: This strategy begins with the cumulative volume and the respective CIR obtained for the desired set of reaches C and directly forecasts the aggregate volume.

For *Strategy 1*, which involves modeling at the reach level, $R_{\Delta,r}(t)$ is used, while for *Strategy 2*, which involves modeling at the aggregate level, $R_{\Delta,C}(t)$ is used. These two quantities are defined in Eq. (5) and are preprocessed to remove extreme outliers, as discussed in Section 3.2.2. For both of these strategies, we consider different data-driven forecasting approaches. The only differences among these data-driven approaches are the CIR data used to train the forecasting models and the inclusion of upstream streamflow gauge data.

3.4.2. Simple linear regression

(a) Univariate Linear Regression (ULR). The first approach used the CIR data only for training. For the sake of describing this approach, we use the CIR of the reach level, $R_{\Delta,r}(t)$. A past window of W_{past} days of CIR data was used to fit a best-fit linear curve (via univariate linear regression, denoted by $\mathcal{G}_1(\cdot)$ in Eq. (12)), which was then used to predict the CIR over the next W_{pred} days. Specifically, for the current time t, the model was trained on the past W_{past} days of CIR data (including time t), denoted by $R_{\Delta,r}([t - W_{\text{past}} + 1 : t])$. The predicted CIR at future time $t < t + h \le t + W_{\text{pred}}$ is functionally defined as:

$$\hat{R}_{\Delta,r}(t+h) = \mathcal{G}_1\left(t+h; a_{\text{uni}}, b_{\text{uni}}\right), \forall h = 1, \ldots, W_{\text{pred}}. \tag{12}$$

Here, a_{uni} and b_{uni} are model parameters for univariate linear regression, as discussed later in Section 3.5.1.

(b) Multivariate Linear Regression (MLR). The second approach performs a $(N_{\text{sg}} + 1)$-dimensional multivariate linear regression, denoted by $\mathcal{G}_2(\cdot)$, to regress the output variable CIR on a multi-dimensional input vector consisting of time and the most sensitive streamflow gauge data (i.e., identified in Section 3.3). Since the most sensitive gauges were selected based on the CIR data, these had to be obtained at the individual reach level for *Strategy 1*, or at the aggregate level for *Strategy 2*.

For the sake of describing MLR, we use the reach-level CIR, denoted by $R_{\Delta,r}(t)$. At the current time t, the model was trained using the past W_{past} days of CIR data as output, denoted by $R_{\Delta,r}([t - W_{\text{past}} + 1 : t])$, which was regressed against streamflow data from N_{sg} selected gauges, denoted by $\tilde{G}_i([t - W_{\text{past}} + 1 : t])$, and the corresponding time vector $[t - W_{\text{past}} + 1 : t]$ as inputs. The predicted CIR at a future time $[t < t + h \le t + W_{\text{pred}}]$ is functionally defined as:

$$\hat{R}_{\Delta,r}(t+h) = \mathcal{G}_2\left(t+h, \tilde{G}_1(t+h), \ldots, \tilde{G}_{N_{\text{sg}}}(t+h); \mathbf{a}_{\text{multi}}, b_{\text{multi}}\right), \tag{13}$$

$\forall h = 1, \ldots, W_{\text{pred}}$.

Here, $\mathbf{a}_{\text{multi}}$ and b_{multi} are model parameters for multivariate linear regression, as discussed later in Section 3.5.1.

3.4.3. Advanced machine learning (ML) forecasting models

(a) Long Short-Term Memory (LSTM) model. This scenario utilizes historical CIR and streamflow gauge data as input, with the future CIR time series serving as the output. The LSTM algorithm, denoted by $\mathcal{G}_3(\cdot)$, was employed to construct the predictive model. Unlike the linear regression models used in *Approaches 1* and *2*, which are trained independently at each time instance t using only the most recent W_{past} days of data (i.e., $[t - W_{\text{past}} + 1 : t]$), the LSTM model is trained *once* on the entire available training history. Once trained, the same LSTM model is used to predict the future output for any given time step.

In this regard, the linear regression approach can be viewed as a *localized fitting technique*, capturing short-term or local effects that may vary from one time step to another. In contrast, the LSTM approach captures *global temporal patterns* by learning from the complete time history, allowing it to generalize and detect long-term dependencies across the data sequence.

The proposed LSTM neural network was specifically designed for the prediction of the sequence-to-sequence of future CIR. The core of the proposed architecture is the LSTM cell, a type of recurrent neural network (RNN) unit specifically designed to capture long-term dependencies in sequential data (Hochreiter and Schmidhuber, 1997). Unlike traditional RNN cells, which can suffer from vanishing or exploding gradients, the LSTM cell incorporates a memory mechanism through the use of gating functions that regulate the flow of information. By selectively retaining or discarding information via its gating mechanisms, the LSTM can model complex dependencies across widely varying timescales, improving robustness to noise and irregularities in the data. This makes the LSTM particularly well-suited for forecasting

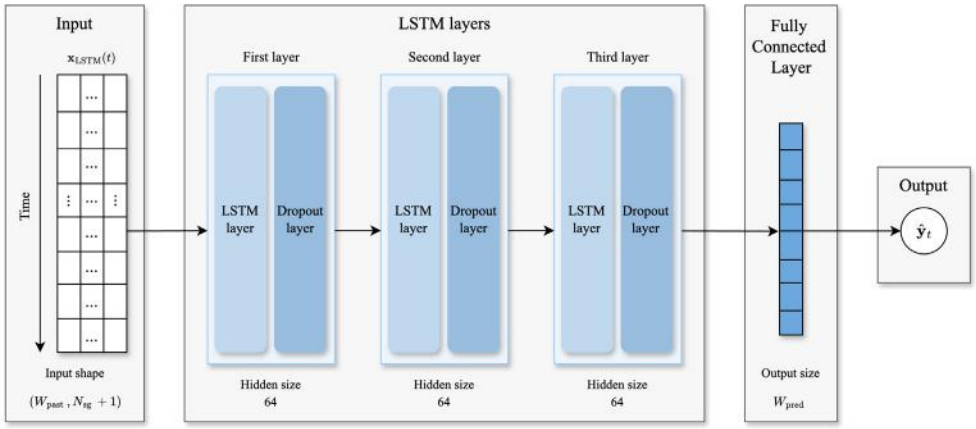


Fig. 11. The overall structure of the LSTM model.

the CIR, where the underlying physical processes may involve delayed responses and long-range correlations.

The architecture of the LSTM model is illustrated in Fig. 11. There are three stacked LSTM layers in the model, each followed by a dropout layer to reduce overfitting by randomly zeroing a fraction of the activations during training. The final hidden state of the last LSTM layer is passed to a fully connected output layer, which maps the learned temporal representation (i.e., streamflow discharge and past CIR values) to the desired output sequence length W_{pred} (i.e., future CIR).

Defining the functional form of $\mathcal{G}_3(\cdot)$ requires further clarification. The input data was structured as a two-dimensional matrix, denoted by $\mathbf{X}_{\text{train}}$, where each row corresponds to a time instance. Each row contains a total of $W_{\text{past}}\cdot(N_{\text{sg}}+1)$ columns, representing the input features: specifically, W_{past} days of past CIR data and streamflow data from N_{sg} gauges.

The number of available days of data, denoted by N_{time}, was divided into N_{train} training days and N_{test} testing days, such that $N_{\text{time}} = N_{\text{train}} + N_{\text{test}}$. Because each training input requires W_{past} days of historical data, the total number of training samples (i.e., the number of rows in $\mathbf{X}_{\text{train}}$) is given by $(N_{\text{train}} - W_{\text{past}} + 1)$.

The output used for training the LSTM model is denoted by $\mathbf{Y}_{\text{train}}$, which contains the same number of rows as $\mathbf{X}_{\text{train}}$, corresponding to each training time instance. $\mathbf{Y}_{\text{train}}$ consists of W_{pred} columns, representing the next W_{pred} days of future CIR data.

$$\mathbf{X}_{\text{train}} \in \mathbb{R}^{(N_{\text{train}}-W_{\text{past}}+1)\times\left(W_{\text{past}}\cdot(N_{\text{sg}}+1)\right)}; \tag{14a}$$

$$\mathbf{Y}_{\text{train}} \in \mathbb{R}^{(N_{\text{train}}-W_{\text{past}}+1)\times W_{\text{pred}}}. \tag{14b}$$

The LSTM model $\mathcal{G}_3(\cdot)$ learns the mapping from the input $\mathbf{X}_{\text{train}}$ to the output $\mathbf{Y}_{\text{train}}$. For the sake of describing the functional form of $\mathcal{G}_3(\cdot)$, we use the reach-level CIR, denoted by $R_{\Delta,r}(t)$. At time t, let $\mathbf{x}_{\text{LSTM}}(t) \in \mathbb{R}^{W_{\text{past}}(N_{\text{sg}}+1)}$ denote the feature vector at time t within a given input window. The predicted CIR, denoted by $\hat{R}_{\Delta,r}([t+1:t+W_{\text{pred}}])$, is given by:

$$\begin{aligned}&\hat{R}_{\Delta,r}([t+1:t+W_{\text{pred}}])\\&=\mathcal{G}_3\left(\mathbf{x}_{\text{LSTM}}(t);\mathbf{W}_{\text{LSTM}},\mathbf{b}_{\text{LSTM}}\right).\end{aligned} \tag{15}$$

Here, $\mathbf{W}_{\text{LSTM}}$ and $\mathbf{b}_{\text{LSTM}}$ are model parameters obtained by training the LSTM model, as discussed later in Section 3.5.2.

(b) Convolutional neural network–long short-term memory (CNN–LSTM) model. To improve the representation of temporal dynamics in CIR forecasting, we also employ a hybrid convolutional neural network–long short-term memory (CNN–LSTM) architecture. This structure combines the strengths of convolutional layers for local temporal feature extraction with recurrent layers for modeling long-term temporal dependencies. The CNN–LSTM structure was proposed in the context of deep learning for speech recognition by Sainath et al. (2015) and has since been successfully applied to a range of temporal prediction tasks.

Unlike the standalone LSTM model, which directly processes the raw input sequence, the CNN–LSTM architecture first applies one-dimensional convolutional filters along the temporal dimension to learn localized temporal features and interactions among the input variables. These learned features are then passed to stacked LSTM layers to model longer-range temporal dependencies. The architecture of CNN-LSTM is illustrated in Fig. 12.

The input and output data structures are similar to those of the LSTM model. The difference is that there are CNN layers between the input and the LSTM layers. The convolutional stage applies one-dimensional convolutional filters along the temporal axis. Two convolutional layers with kernel size k are used, each followed by a ReLU activation function. Dropout is applied after the first convolutional layer to reduce overfitting (dropout rate = 0.2). These convolutional layers learn local temporal patterns and short-term variations in the CIR and streamflow signals within the input window. After the convolutional feature extraction stage, the feature representation maintains the same temporal length while expanding the feature dimension to 64 channels. Let $C(\cdot)$ denote the nonlinear transformation implemented by the convolutional feature extractor. The output of the CNN block is given by

$$\mathbf{Z}(t) = C\left(\mathbf{x}_{\text{CNN-LSTM}}(t)\right), \tag{16}$$

where $\mathbf{Z}(t)$ is a sequence of learned temporal features with reduced noise and enhanced local structure, $\mathbf{x}_{\text{CNN-LSTM}}(t) \in \mathbb{R}^{W_{\text{past}}\times(N_{\text{sg}}+1)}$ denotes the input feature vector at each prediction time t.

The extracted feature sequence $\mathbf{Z}(t)$ is then passed to stacked LSTM layers. Similar to the standalone LSTM model, the LSTM component is trained once using the entire available training history and is subsequently applied to all prediction time steps. The LSTM layers capture global temporal dependencies and long-range correlations across the filtered feature sequence. The final hidden state of the last LSTM layer is passed through a ReLU activation function and a dropout layer, and then mapped to the output space via a fully connected layer.

The CNN–LSTM model, denoted by $\mathcal{G}_4(\cdot)$, learns the mapping from the input sequence to the future CIR sequence as

$$\begin{aligned}&\hat{R}_{\Delta,r}\left(\left[t+1:t+W_{\text{pred}}\right]\right)\\&=\mathcal{G}_4\left(\mathbf{x}_{\text{CNN-LSTM}}(t);\mathbf{W}_{\text{CNN-LSTM}},\mathbf{b}_{\text{CNN-LSTM}}\right),\end{aligned} \tag{17}$$

where $\mathbf{W}_{\text{CNN-LSTM}}$ and $\mathbf{b}_{\text{CNN-LSTM}}$ represent the trained parameters of both the convolutional and LSTM layers.

(c) Temporal Convolutional Network–Long Short-Term Memory (TCN–LSTM) model. This scenario extends the LSTM-based forecasting framework by incorporating a Temporal Convolutional Network (TCN) as

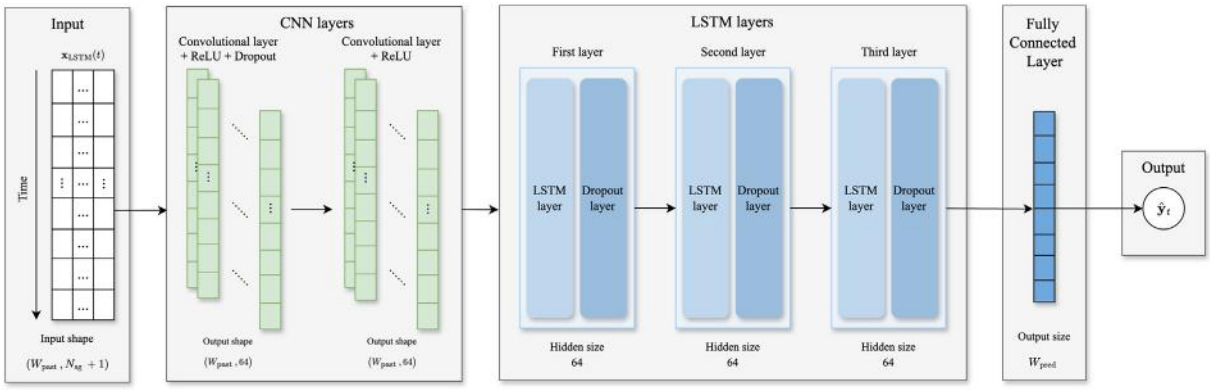


Fig. 12. The overall structure of the CNN-LSTM model for CIR forecasting.

a front-end feature extractor, followed by stacked LSTM layers for sequence-to-sequence prediction. Unlike standard convolutional neural networks with fixed receptive fields, the TCN employs causal and dilated one-dimensional convolutions, enabling the model to capture temporal dependencies over multiple timescales while preserving the temporal ordering of the input sequence (Bai, 2018). By stacking convolutional layers with increasing dilation factors, the TCN effectively expands its temporal receptive field without increasing model depth. This design enables the TCN to encode long-range temporal patterns directly within the convolutional feature maps prior to recurrent processing.

The overall structure of the TCN–LSTM model is illustrated in Fig. 13. The input data are first processed by a sequence of dilated causal convolutional layers, forming the TCN component. It applies a series of dilated one-dimensional convolutional layers to capture temporal patterns at multiple scales. In this study, three dilated convolution layers are used with dilation factors $d = 1, d = 2,$ and $d = 4$, respectively. Each convolutional layer contains 32 filters and is followed by a ReLU activation function and a dropout layer (dropout rate = 0.2) to enhance nonlinearity and reduce overfitting. After the TCN module, the extracted features maintain the same temporal length while expanding the feature dimension to 32 channels, resulting in an output tensor of shape $(W_{\text{past}}, 32)$. This output feature sequence is then passed to a stack of LSTM layers, followed by a fully connected output layer that maps the learned temporal representation to the desired prediction horizon W_{pred}. At a given time t, let $\mathbf{x}_{\text{TCN–LSTM}}(t) \in \mathbb{R}^{W_{\text{past}}(N_{\text{sg}}+1)}$ denote the input feature vector within the sliding window. The predicted CIR sequence is expressed as

$$\begin{aligned}&\hat{R}_{\Delta,r}\left(\left[t+1:t+W_{\text{pred}}\right]\right)\\&=\mathcal{G}_5\left(\mathbf{x}_{\text{TCN–LSTM}}(t);\mathbf{W}_{\text{TCN–LSTM}},\mathbf{b}_{\text{TCN–LSTM}}\right),\end{aligned}\tag{18}$$

where $\mathbf{W}_{\text{TCN–LSTM}}$ and $\mathbf{b}_{\text{TCN–LSTM}}$ denote the learned parameters of the TCN, LSTM, and fully connected layers.

Remark 1. Recall from Fig. 4 that the HSC region has sparse survey data, whereas the SWP region has relatively abundant survey data. Therefore, the advanced ML model-based approaches, such as variants of LSTM-based models adopted in this paper, which provide a flexible global fitting capability, are more suitable for capturing the complex dynamics of the SWP region. This modeling approach can effectively learn temporal depth and non-linear patterns from rich data. For the SWP region modeled using LSTM, CNN-LSTM, and TCN-LSTM models, we set both W_{past} and W_{pred} to 30 days. The 30-day input window was selected as it provided the optimal balance between model accuracy and computational efficiency, while also representing the typical hydrologic memory of discharge and sediment processes at the monthly scale. On the other hand, the sparse survey data in the HSC region produces a sedimentation volume curve that approximates a coarse, piecewise linear form. In such cases, the data structure aligns more naturally with simpler linear models (as suggested by ULR and MLR), which are better suited to learning local behavior. The sparsity of data limits the ability to capture long-term patterns. Additionally, such localized linear fitting techniques help avoid overfitting and align well with the first-order characteristics introduced by interpolation. For the HSC region modeled using ULR and MLR, we set relatively lower values of W_{past} to 5 days.

3.4.4. Obtaining the volume from CIR data

For all three approaches and both scenarios, the CIR data used for training were obtained using consecutive days of volume data with $\Delta = 1$.

Consider *Strategy 1* and a current time instance t, such that the past CIR and streamflow gauge data are available. Once $\hat{R}_{\Delta=1,r}([t+1:t+W_{\text{pred}}])$ is obtained (using any of the three approaches), and since $\Delta = 1$, the forecasted volume of the reach, denoted by $\hat{V}_r$, and the combined volume, denoted by $\hat{V}_C$, at time $(t+1)$ for *Strategy 1* is computed as:

$$\hat{V}_r(t+i)\tag{19a}$$

$$=\begin{cases}V_r(t)+\hat{R}_{\Delta=1,r}(t+i), & \text{if } i=1\\ \hat{V}_r(t+i-1)+\hat{R}_{\Delta=1,r}(t+i), & \text{if } i>1\end{cases};\tag{19b}$$

$$\hat{V}_C(t+1)=\sum_{i=1}^{N_{\text{reaches}}}\hat{V}_r(t).\tag{19c}$$

Similarly, for *Strategy 2*, the predicted sedimentation volume, denoted by $\hat{V}_C$, is computed as:

$$\hat{V}_C(t+i)=\begin{cases}V_C(t)+\hat{R}_{\Delta=1,C}(t+i), & \text{if } i=1\\ \hat{V}_C(t+i-1)+\hat{R}_{\Delta=1,C}(t+i), & \text{if } i>1\end{cases}\tag{20}$$

As discussed in Remark 1, since HSC has sparse data that are linearly interpolated to obtain the volume time series, it serves as a natural choice to demonstrate ULR and MLR. On the other hand, SWP, being rich in survey data, is more suitable for illustrating the workings of LSTM-based models (i.e., LSTM, CNN-LSTM, and TCN-LSTM). All methods will be discussed in more detail in Section 3.5.

3.5. Forecasting framework

The remainder of this section presents a comprehensive overview of the linear regression and LSTM-based ML methods utilized in this study. To describe the regression techniques used across the three approaches, we focus on the reach-level CIR (i.e., *Scenario 1*). The regression methods remain the same for *Scenario 2*, except that CIR data is considered at the aggregate level, with all other aspects held constant.

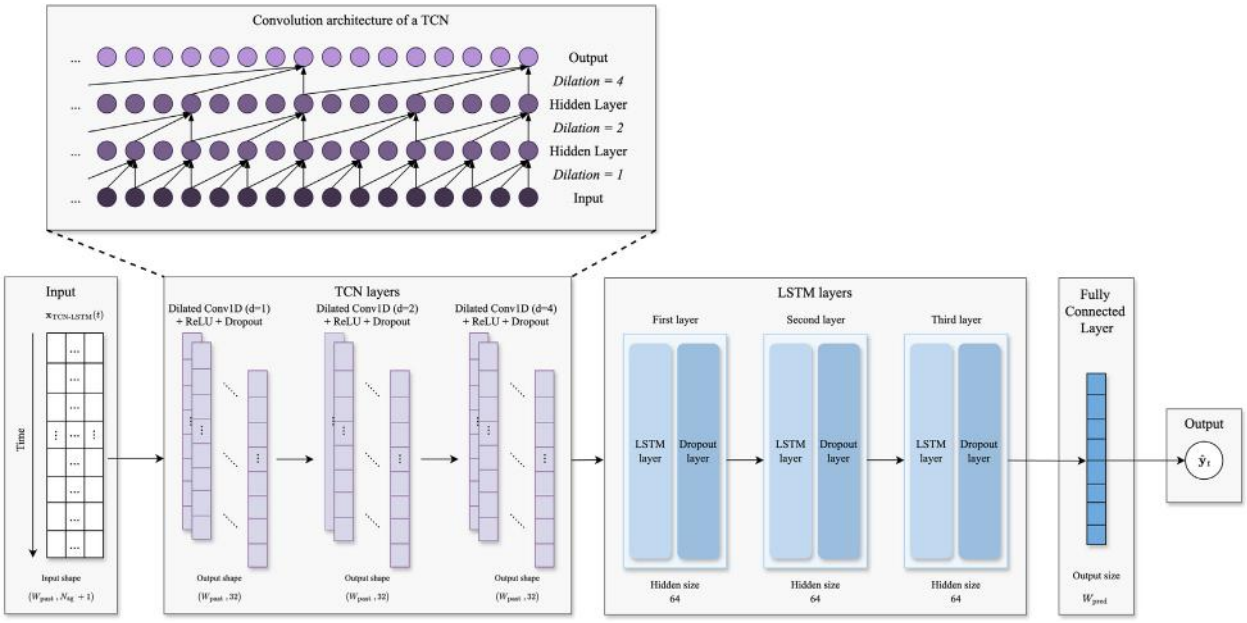


Fig. 13. The overall structure of the TCN-LSTM model for CIR forecasting.

3.5.1. Linear regression based model

Solution ULR and MLR are two different linear regression strategies: a univariate approach using only the recent history of the target variable, and a multivariate approach that incorporates additional gauge data.

1. In the ULR case, predictions rely exclusively on past observations of the target variable, $R_{\Delta,r}([t-W_{\text{past}}+1:t])$. For each prediction step, a simple linear regression model is fitted using the CIR data over the past W_{past} days as the output and time as the input. The model estimates a linear trend — characterized by a single slope and intercept — across these time points. This linear model is then extrapolated to predict sediment volumes for the subsequent W_{pred} days.
   The fitted linear function is expressed as

$$R_{\Delta,r}(t) = a_{\text{uni}} \cdot t + b_{\text{uni}}. \tag{21}$$

   Here, t is the time index, a_{uni} is the estimated slope, and b_{uni} is the intercept—both obtained by least-squares fitting of the training input $[t-W_{\text{past}}+1:t]$ to the output $R_{\Delta,r}([t-W_{\text{past}}+1:t])$.
   Prediction for the next W_{pred} days ($h = 1,\ldots,W_{\text{pred}}$) are given by:

$$\hat{R}_{\Delta,r}(t+h) = a_{\text{uni}} \cdot (t+h) + b_{\text{uni}}. \tag{22}$$

   where the same slope and intercept are applied for all forecast steps $1 \le h \le W_{\text{pred}}$.
2. Similar to ULR, in MLR, a local model is trained at each time step using a multivariate linear regression. The regression models the CIR over the past W_{past} days as the output variable, using the corresponding time steps and the most sensitive gauge data over the same period as input features. The fitted linear model is expressed as:

$$R_{\Delta,r}(t) = \mathbf{x}_{\text{multi}}(t) \cdot \mathbf{a}_{\text{multi}} + b_{\text{multi}}. \tag{23}$$

   Here, $\mathbf{x}_{\text{multi}}(t)$ denotes the input vector defined as:

$$\mathbf{x}_{\text{multi}}(t) = [t, \tilde{G}_1(t), \ldots, \tilde{G}_{N_{\text{sg}}}(t)]. \tag{24}$$

   The parameters $\mathbf{a}_{\text{multi}}$ and b_{multi} are obtained by least-squares fitting of the training input $\mathbf{x}_{\text{multi}}(t_{\text{train}})$ to the output $R_{\Delta,r}(t_{\text{train}})$ for all training time steps $t_{\text{train}} \in [t-W_{\text{past}}+1:t]$. The term $\mathbf{x}_{\text{multi}}(t) \cdot \mathbf{a}_{\text{multi}}$ denotes a dot product between input vector $\mathbf{x}_{\text{multi}}(t)$ and the parameter vector $\mathbf{a}_{\text{multi}}$.
   Prediction for the next W_{pred} days ($h = 1,\ldots,W_{\text{pred}}$) are given by:

$$\hat{R}_{\Delta,r}(t+h) = \mathbf{x}_{\text{multi}}(t+h) \cdot \mathbf{a}_{\text{multi}} + b_{\text{multi}}. \tag{25}$$

   where the same slope and intercept are applied for all forecast steps $1 \le h \le W_{\text{pred}}$.

3.5.2. ML model-based based CIR forecasting

Let $\mathbf{x}_{\text{ML-based}}(t) \in \mathbb{R}^{W_{\text{past}}(N_{\text{sg}}+1)}$ denotes the feature vector at time t within a given input window, where $\mathbf{x}_{\text{ML-based}}(t)$ is $\mathbf{x}_{\text{LSTM}}(t)$ for the LSTM model, is $\mathbf{x}_{\text{CNN-LSTM}}(t)$ after passing through the CNN layers for the CNN-LSTM model, and is $\mathbf{x}_{\text{TCN-LSTM}}(t)$ after passing through the TCN layers for the TCN-LSTM model. For each LSTM layer $l = 1,2,3$, the hidden state and cell state are recursively updated according to:

$$\left[\mathbf{h}_t^{(l)}, \mathbf{c}_t^{(l)}\right] = \text{LSTM}^{(l)}\left(\mathbf{h}_{t-1}^{(l)}, \mathbf{c}_{t-1}^{(l)}, \mathbf{z}_t^{(l)}\right), \tag{26}$$

where $\mathbf{z}_t^{(l)} = \mathbf{x}_{\text{ML-based}}(t)$ for the first layer and $\mathbf{z}_t^{(l)} = \mathbf{h}_t^{(l-1)}$ for subsequent layers.

After passing through the stacked LSTM layers (i.e., CNN and LSTM layers for the CNN-LSTM model, TCN and LSTM layers for the TCN-LSTM model) and dropout regularization, the final hidden state sequence is processed by a dense output layer:

$$\hat{\mathbf{y}}_t = \mathbf{W}_{\text{LSTM}}\mathbf{h}_t^{(3)} + \mathbf{b}_{\text{LSTM}}. \tag{27}$$

Here, $\mathbf{W}_{\text{LSTM}}$ and $\mathbf{b}_{\text{LSTM}}$ are the weights and bias of the output layer in the LSTM module, and $\hat{\mathbf{y}} \in \mathbb{R}^{W_{\text{pred}}}$ is the vector of predicted values for the forecast window. For *Scenario 1*, $\hat{\mathbf{y}}_t = \hat{R}_{\Delta,r}([t+1:t+W_{\text{pred}}])$; and for *Scenario 2*, $\hat{\mathbf{y}}_t = \hat{R}_{\Delta,C}([t+1:t+W_{\text{pred}}])$.

3.6. Uncertainty quantification of CIR forecasting

In addition to point predictions, uncertainty quantification (UQ) is incorporated into all LSTM-based models considered in this study, including the standard LSTM, CNN–LSTM, and TCN–LSTM architectures. Given the cumulative nature of sediment volume prediction, where small errors in predicted CIR can propagate over time, the predictive uncertainty is critical for reliable forecasting and risk-informed decision

making. To achieve this, Monte Carlo (MC) dropout is implemented as a computationally efficient, distribution-free approach for estimating predictive uncertainty in deep neural networks. MC-dropout provides an approximation to Bayesian inference by interpreting dropout as a variational approximation over the model parameters (Gal and Ghahramani, 2016).

Let $\mathcal{G}_{\mathrm{ML}}(\cdot)$ denote a trained LSTM-based model or a hybrid model that combines LSTM model with a CNN or TCN model (i.e., LSTM, CNN–LSTM, or TCN–LSTM), parameterized by weights W. During standard training, dropout layers are used to randomly deactivate a subset of neurons, reducing overfitting. In MC-dropout, this stochastic behavior is retained during inference, enabling multiple forward passes through the network with different realizations of the dropout mask. For a given input sequence $\mathbf{x}(t)$, N stochastic forward passes are performed:

$$\hat{\mathbf{y}}^{(i)}(t) = \mathcal{G}_{\mathrm{ML}}\left(\mathbf{x}(t); \mathbf{W}^{(i)}\right), \quad i = 1, \ldots, N, \tag{28}$$

where $\mathbf{W}^{(i)}$ represents an implicit sample from the approximate posterior induced by dropout. Each forward pass produces a realization of the predicted CIR sequence over the forecast horizon W_{pred}. Each realization of the predicted CIR sequence is integrated over time to obtain the corresponding cumulative sedimentation-volume trajectory. The uncertainty in the predicted volume is then quantified using the ensemble distribution of these stochastic realizations.

The predictive mean is estimated as the ensemble average of the MC samples:

$$\bar{\mathbf{y}}(t) \approx \frac{1}{N}\sum_{i=1}^{N} \hat{\mathbf{y}}^{(i)}(t), \tag{29}$$

while the predictive uncertainty is quantified by the sample variance:

$$\sigma^2(t) \approx \frac{1}{N-1}\sum_{i=1}^{N}\left(\hat{\mathbf{y}}^{(i)}(t) - \bar{\mathbf{y}}(t)\right)^2. \tag{30}$$

Assuming approximate normality of the predictive distribution, a two-sided confidence interval can be constructed at each forecast horizon step. In this study, we select 95% confidence interval which can be find in the case study section. These uncertainty bounds are propagated through the sediment volume integration process, enabling uncertainty estimates not only for CIR predictions but also for the corresponding cumulative sediment volume forecasts.

3.7. Error metrics

The efficacy of predicting the sedimentation volume, obtained by integrating the predicted CIR, is quantified using three different error metrics. For each time instance t in the test segment, the output consists of a predicted time series of W_{pred} future time steps. Let the vector $\hat{\mathbf{V}}_{t+1:t+W\,\mathrm{pred}}$ denote the predicted volume time series and $\mathbf{V}_{t+1:t+W\,\mathrm{pred}}$ denote the corresponding observed true values for test time instance t, with the total number of test samples denoted by N_{test}. Let $\hat{V}_{t+k}$ and V_{t+k} denote the predicted and true values at time $t+k$ (i.e., the kth elements of their respective vectors).

The first error metric used is the Root Mean Squared Error (RMSE) defined as:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N_{\mathrm{test}} \cdot W_{\mathrm{pred}}}\sum_{t=1}^{N_{\mathrm{test}}}\sum_{k=1}^{W_{\mathrm{pred}}}\left(V_{t+k} - \hat{V}_{t+k}\right)^2}. \tag{31}$$

Since observed sediment volumes vary substantially across test windows, RMSE alone may not provide a fair performance comparison. This issue is addressed by normalizing RMSE by the mean observed value within each window, which facilitates consistent comparison across periods and scales. The average normalized RMSE, denoted as nRMSE, is defined as:

$$\mathrm{nRMSE} = \frac{1}{N_{\mathrm{test}}}\left(\frac{\sqrt{\frac{1}{W_{\mathrm{pred}}}\sum_{k=1}^{W_{\mathrm{pred}}}\left(V_{t+k} - \hat{V}_{t+k}\right)^2}}{\frac{1}{W_{\mathrm{pred}}}\sum_{k=1}^{W_{\mathrm{pred}}} V_{t+k}}\right). \tag{32}$$

The third error metric, used to assess the relative accuracy of predictions across periods with varying sediment volume scales, is the Mean Absolute Percentage Error (MAPE). The average MAPE is computed as:

$$\mathrm{MAPE} = \frac{1}{N_{\mathrm{test}}}\left(\frac{100}{W_{\mathrm{pred}}}\sum_{k=1}^{W_{\mathrm{pred}}}\left|\frac{V_{t+k} - \hat{V}_{t+k}}{V_{t+k}}\right|\right). \tag{33}$$

3.8. Implementation details and reproducibility

To improve reproducibility, this subsection summarizes the main implementation details used in the case studies. The preprocessing of the CSAT/eHydro-derived bathymetric survey data consisted of the following steps. First, extremely large invalid values in the bathymetric grids were replaced with missing values. Grid cells that remained missing across all survey IDs were treated as outside the surveyed channel region and excluded from volume calculations. For grid cells with intermittent missing values, linear interpolation was applied over time to construct a continuous daily elevation record. When multiple surveys were available on the same day, the elevation values were aggregated by taking the mean value for each grid cell after excluding missing values. The resulting daily bathymetric time series was then used to compute reach-level sedimentation volume, followed by smoothing and CIR calculation as described in Section 3.2.

For model inputs, the selected upstream streamflow gauges and the CIR series were aligned to the same daily temporal resolution. Normalization parameters were estimated using the training set only and then applied to the validation and testing sets to avoid information leakage. For the LSTM-based models, each training sample was constructed using a sliding-window structure. Specifically, the input at prediction time t consisted of the past W_{past} days of CIR and the selected streamflow gauge data, while the output consisted of the future W_{pred} days of CIR. The predicted CIR sequence was then integrated to recover the cumulative sedimentation-volume forecast. In this study, W_{pred} was set to 30 days to support short-term dredging decision-making. For LSTM-based models in the data-rich SWP region, W_{past} was set to 30 days. For the linear models applied to the data-sparse HSC region, a shorter local window was used.

The data were split chronologically into training, validation, and testing periods. This chronological split was used to avoid temporal leakage and to reflect the operational forecasting setting in which future observations are unavailable during model development. In the case studies, the training set is from 2013 to 2021, the validation set from 2021 to 2022, and the test period covers 2022–2023. Rolling 30-day forecasts were generated throughout the test period.

Hyperparameters for the LSTM-based models were selected based on validation-set performance. The main hyperparameters include the input window length W_{past}, prediction horizon W_{pred}, number of LSTM layers, hidden dimension, dropout rate, learning rate, batch size, number of training epochs, and the number of MC-dropout samples used for UQ. The final model was then evaluated on the independent test set using RMSE, nRMSE, and MAPE. For operational deployment in a new region, the same procedure shown above can be followed.

4. Case studies

The methodology described above is evaluated using two navigation channels with contrasting survey-data conditions: the Houston Ship Channel (HSC), which has relatively sparse bathymetric survey records, and the Southwest Pass (SWP), which has substantially denser

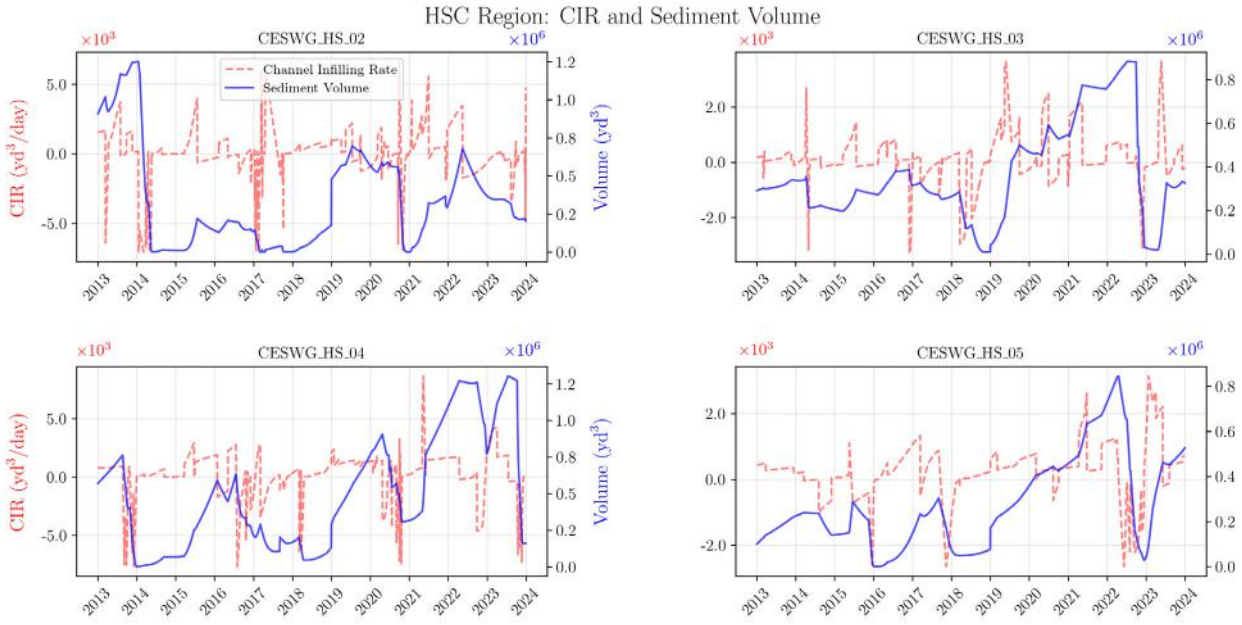


Fig. 14. Sedimentation volume and channel infilling rate (CIR) for representative reaches in the HSC region (The legend for all subplots is identical to that of the top-left figure).

survey coverage. This contrast allows the proposed framework to be evaluated under two representative data-availability regimes. The case-study results are organized to first examine reach-level and aggregate forecasting strategies, then compare different forecasting models, and finally evaluate uncertainty quantification for LSTM-based models.

In this section, the two study sites of interest (i.e., HSC and SWP) as described in Section 2.2 are employed to demonstrate and validate the proposed framework. As mentioned previously, these two sites have different survey frequencies, which represent two different representative scenarios. Through this case study, we aim to identify the best modeling approach for each scenario.

4.1. Houston Ship Channel: Forecasting under sparse survey conditions

Fig. 14 illustrates historic CIR and sedimentation volume of four representative reaches in the HSC region. These figures show that the HSC has relatively fewer measurements and bathymetric surveys compared to the SWP region (also reflected in Figs. 4 and 21). In this case study, data spanning January 1, 2013, to the end of 2023 were used. The dataset was divided into three subsets: the training set (2013–early 2021), the validation set (2021–2022), and the test set (2022–end of 2023). The model was trained to produce rolling 30-day forecasts for each timestep within the test period.

Two strategies are evaluated and compared in the case study. The first is a disaggregated strategy, in which GSA is performed, and models are trained separately for each reach. After selecting an identical testing period for all reaches, the predicted sediment volumes from each reach are aggregated to obtain the total sediment volume trend for the entire target region. In contrast, the aggregated strategy directly applies GSA and trains the model using the total sediment volume aggregated across all reaches. Although the aggregated approach can retain a larger number of timesteps within the same testing period, the testing dataset is constrained to match that of the disaggregated strategy to ensure a fair comparison. The remainder of this section presents the results obtained for this region.

4.1.1. Disaggregated strategy

The prediction results obtained using the disaggregated strategy are presented in Fig. 15 and Table 2. Fig. 15 illustrates the predicted sedimentation volume and CIR at the reach level for a representative reach (CESWG_HS_02_RFB _2) in the HSC region at four different time instances. Table 2 reports the error metrics corresponding to these predictions. The full test set consists of 686 filtered time steps, covering the period from January 1, 2022, to December 31, 2023. From the results, it can be observed that:

- In the HSC region, linear regression (ULR and MLR) proves better suited for the HSC region than LSTM, which yields poor predictive performance with significantly larger errors. This weakness arises from the HSC's sparse survey record, which produces a sedimentation volume curve that is essentially a coarse, piecewise linear approximation, as shown in Fig. 14. In such cases, the data structure aligns more naturally with simpler linear models, which not only avoid overfitting but also capture the first-order characteristics introduced by the interpolation process.
- The results of ULR and MLR are not very different. Based on the nRMSE and MAPE error metrics, MLR performs marginally better than ULR. Therefore, although MLR is objectively the best, the minimal difference between ULR and MLR arises from the sparse nature of the HSC data, where adding information from streamflow gauges provides little additional value, and the correlation between the CIR data and the gauge data is weak.

The predicted sedimentation volume using MLR at the reach level is evaluated under two predictor configurations based on the most sensitive upstream gauges measuring streamflow discharge (ranking obtained from the GSA analysis). The first configuration uses the three most sensitive gauges, representing a parsimonious set that captures the dominant upstream hydrological signals influencing CIR dynamics. The second configuration expands the predictor set to the top eight gauges to examine whether incorporating additional upstream observations provides additional predictive information. This comparison allows assessment of the trade-off between model simplicity and potential information gain from a larger set of hydrological predictors. Table 3 reports the error metrics corresponding to these predictions. The full test set consists of 686 time steps, covering the period from January 1, 2022, to December 31, 2023. Several prediction instances show

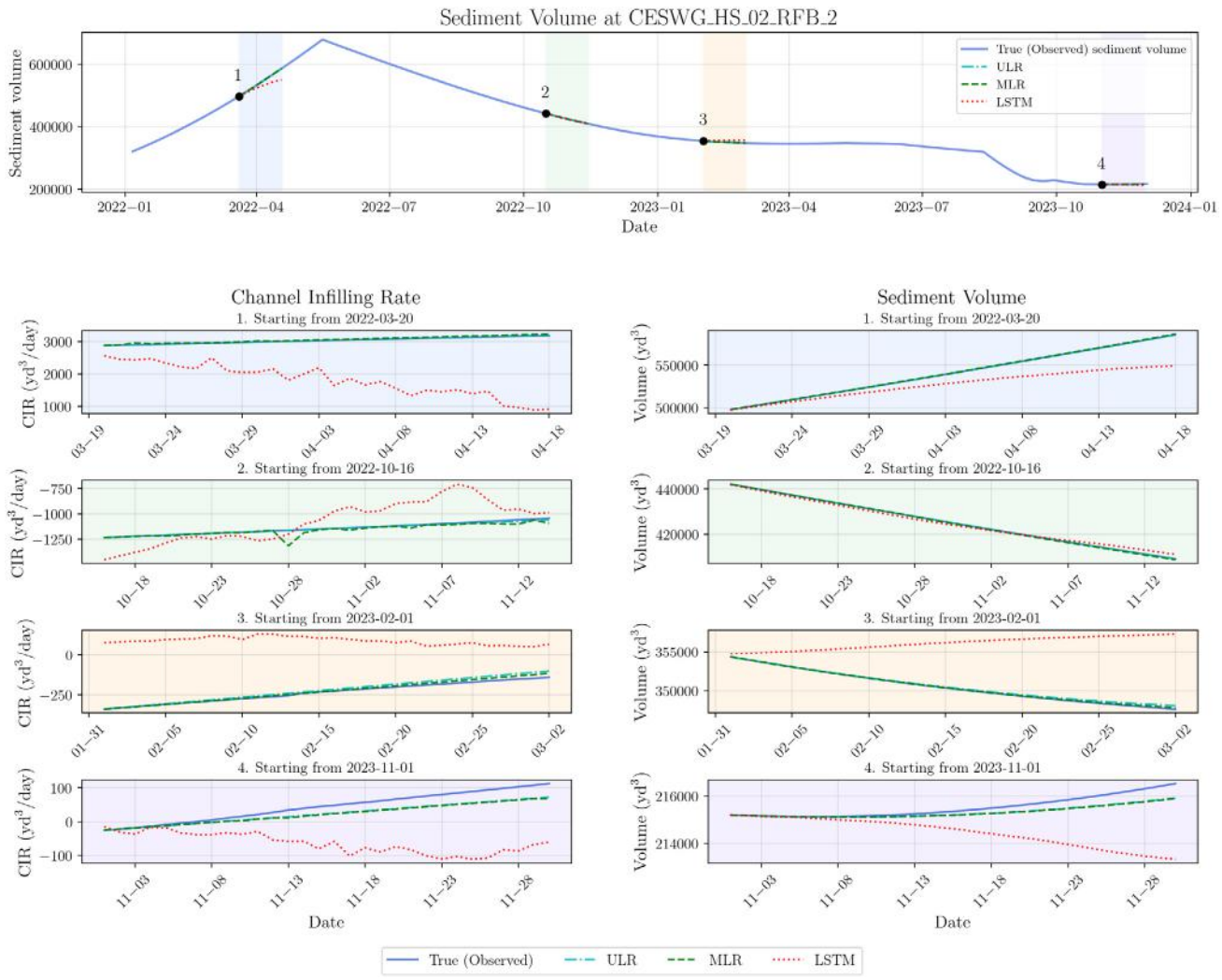


Fig. 15. HSC region: Forecasted CIR and sediment volume at a representative individual reach under the disaggregated strategy, compared across three modeling approaches. The top panel shows the observed sedimentation-volume time series over the test period, with four selected forecast windows highlighted by shaded regions and labeled by forecast start date. The lower panels show the detailed forecasts for these four windows: the left column presents the predicted CIR, and the right column presents the corresponding sedimentation-volume forecasts obtained by integrating the predicted CIR. Each row corresponds to one forecast window.

Table 2
HSC region: Evaluation metrics for sedimentation-volume predictions at reach-level for a representative reach (i.e., CESWG_HS_02_RFB_2), under the disaggregated modeling strategy, compared across different modeling approaches. Columns correspond to four selected forecast start dates and the full test set. For the full test set, the best value for each metric is highlighted in bold.

Method	Metric	2022-03-20	2022-10-16	2023-02-01	2023-11-01	Full test set
ULR	RMSE (yd^3)	68.56	26.18	197.12	282.07	**2226.55**
	nRMSE	0.0001	0.00006	0.0005	0.0013	0.0038
	MAPE (%)	0.007	0.003	0.041	0.09	0.27
MLR	RMSE (yd^3)	382.43	231.80	70.73	292.15	2301.74
	nRMSE	0.0007	0.0005	0.0002	0.0013	**0.0034**
	MAPE (%)	0.05	0.0425	0.01	0.09	**0.26**
LSTM	RMSE (yd^3)	17 671.27	1033.39	6315.08	1487.23	15 617.80
	nRMSE	0.0326	0.0024	0.0180	0.0069	0.0352
	MAPE (%)	2.49	0.21	1.61	0.49	2.79

stronger agreement between the observed and predicted CIR when using only the three most sensitive gauges. Consistent with these observations, the nRMSE and MAPE metrics computed for the full test set indicate that the three-gauge case consistently outperforms the eight-gauge case. This improvement can be attributed to the characteristics of the HSC region, where the inclusion of additional gauges may introduce noise rather than useful predictive information due to weak correlations between some gauges and the CIR data.

4.1.2. Comparison between disaggregated and aggregated strategies

Fig. 16 presents the predicted cumulative sedimentation volume using MLR for both Modeling *Scenario 1* (disaggregated strategy) and

Table 3
HSC region: Evaluation metrics for reach-level sedimentation-volume predictions at a representative reach (i.e., CESWG_HS_02_RFB_2) under the disaggregated modeling strategy, using different sets of sensitive gauges (MLR). Columns correspond to four selected forecast start dates and the full test set. Best values for the full test set are highlighted in bold.

Input	Metric	2022-02-06	2022-07-15	2023-02-03	2023-08-28	Full test set
Top 3 most sensitive gauges	RMSE (yd^3)	218.44	15.94	155.73	4539.63	**2301.74**
	nRMSE	0.0005	0.00002	0.0004	0.0190	**0.0034**
	MAPE (%)	0.04	0.001	0.03	1.54	**0.26**
Top 8 most sensitive gauges	RMSE (yd^3)	153.46	45.11	1288.92	11,558.70	3062.29
	nRMSE	0.0004	0.00008	0.0037	0.0484	0.0041
	MAPE (%)	0.03	0.006	0.28	3.55	0.29

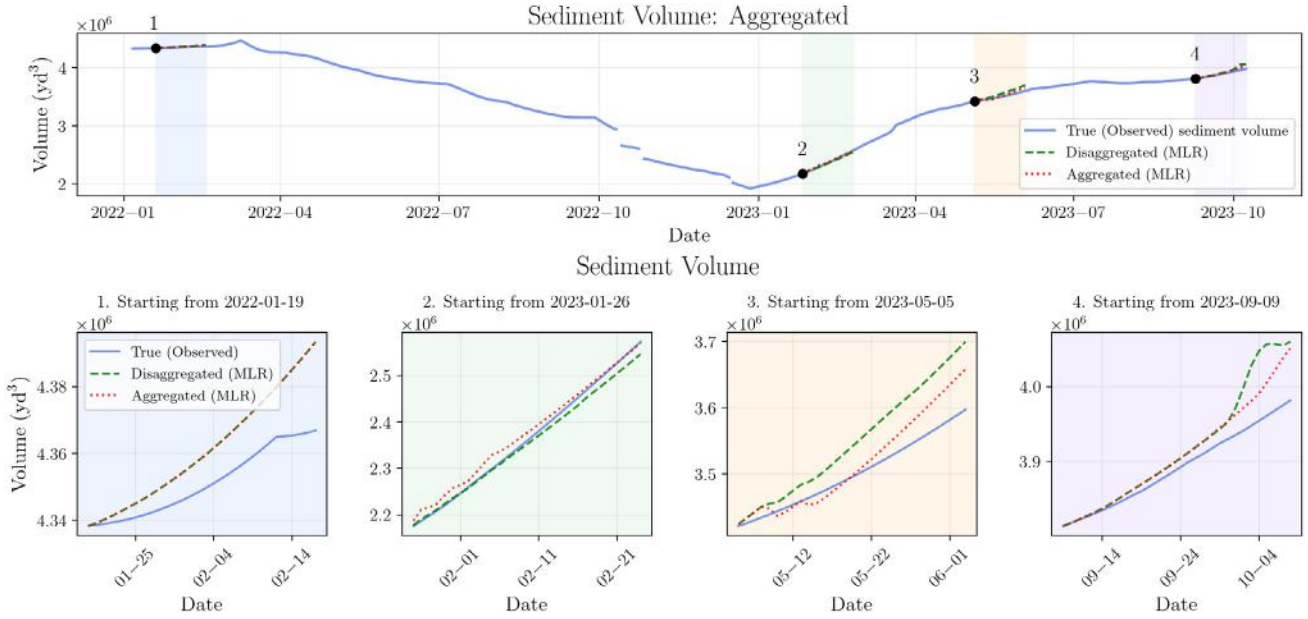


Fig. 16. HSC region: Forecasting of cumulative sedimentation volume aggregated over all reaches using two modeling strategies, illustrated for four selected forecast start dates using the MLR model with the three most sensitive gauges. The top row shows the overall sediment volume time series together with the forecast windows. The second row shows the corresponding zoomed-in views of sediment volume forecasts. Each column corresponds to one forecast window.

Modeling *Scenario 2* (aggregated strategy). The corresponding error metrics are reported in Table 4. Unlike the reach-level predictions, the full test set for cumulative volume predictions contains only 86 filtered time steps, spanning the period from January 1, 2022, to December 31, 2023. The limited size of the test dataset arises from the requirement to retain only the common test timesteps shared by both *Scenario 1* and *Scenario 2* in order to ensure a fair comparison. However, because the disaggregated strategy (*Scenario 1*) applies the IQR-filtering process separately to each reach, many participating reaches end up with a larger number of discarded time steps. Therefore, only the time steps retained across all participating reaches were used for testing, resulting in a reduced number of test time steps. As shown in Fig. 16 and Table 4, the cumulative volumes predicted for the two scenarios differ only marginally. Therefore, we conclude that cumulative volume prediction depends less on selecting the scenario and more on choosing the appropriate regression technique (ULR or MLR over LSTM-based models).

4.1.3. Model comparison under the aggregated strategy

From the results above, linear regression models demonstrate better overall predictive performance, which can be attributed to the limited number of bathymetric surveys and the resulting scarcity of informative data in the HSC region. Furthermore, the disaggregated and aggregated strategies have nearly identical prediction accuracy; however, the disaggregated strategy greatly reduces the number of available timesteps in the test set. Therefore, subsequent analyses in this section focus exclusively on the aggregated strategy, which provides comparable predictive performance while allowing a more flexible and expanded test set. Within this framework, more advanced LSTM-based models are introduced for further comparison. The following five approaches are compared. **(1) LSTM-Volume**: This model uses an LSTM architecture trained directly on cumulative sediment volume data. The input features are the same as those used in the other models. This model serves as a direct cumulative-volume forecasting baseline and is used to quantify the benefit of reformulating the prediction target as CIR. **(2) MLR**: The MLR model is trained to predict the CIR data. The cumulative sediment volume is then calculated from the predicted CIR values. **(3) LSTM**: The LSTM model is designed to forecast the CIR data, which is subsequently used to compute the cumulative sediment volume. **(4) CNN-LSTM**: This model combines CNN and LSTM architectures to predict the CIR data. The predicted CIR values are then employed to estimate the cumulative sediment volume. **(5) TCN-LSTM**: Integrating TCN and LSTM models, this approach predicts the CIR data from which the cumulative sediment volume is derived.

Fig. 17 presents four examples of the predicted CIR and the corresponding back-calculated cumulative sediment volume for the entire HSC region obtained using different models. Table 5 summarizes the associated performance metrics. The observed CIR and sediment volume trends are presented, along with a baseline model in which sediment volume is directly predicted using an LSTM model. The comparison with the LSTM-Volume baseline provides direct quantitative evidence for the advantage of the CIR-based formulation. For the HSC aggregated

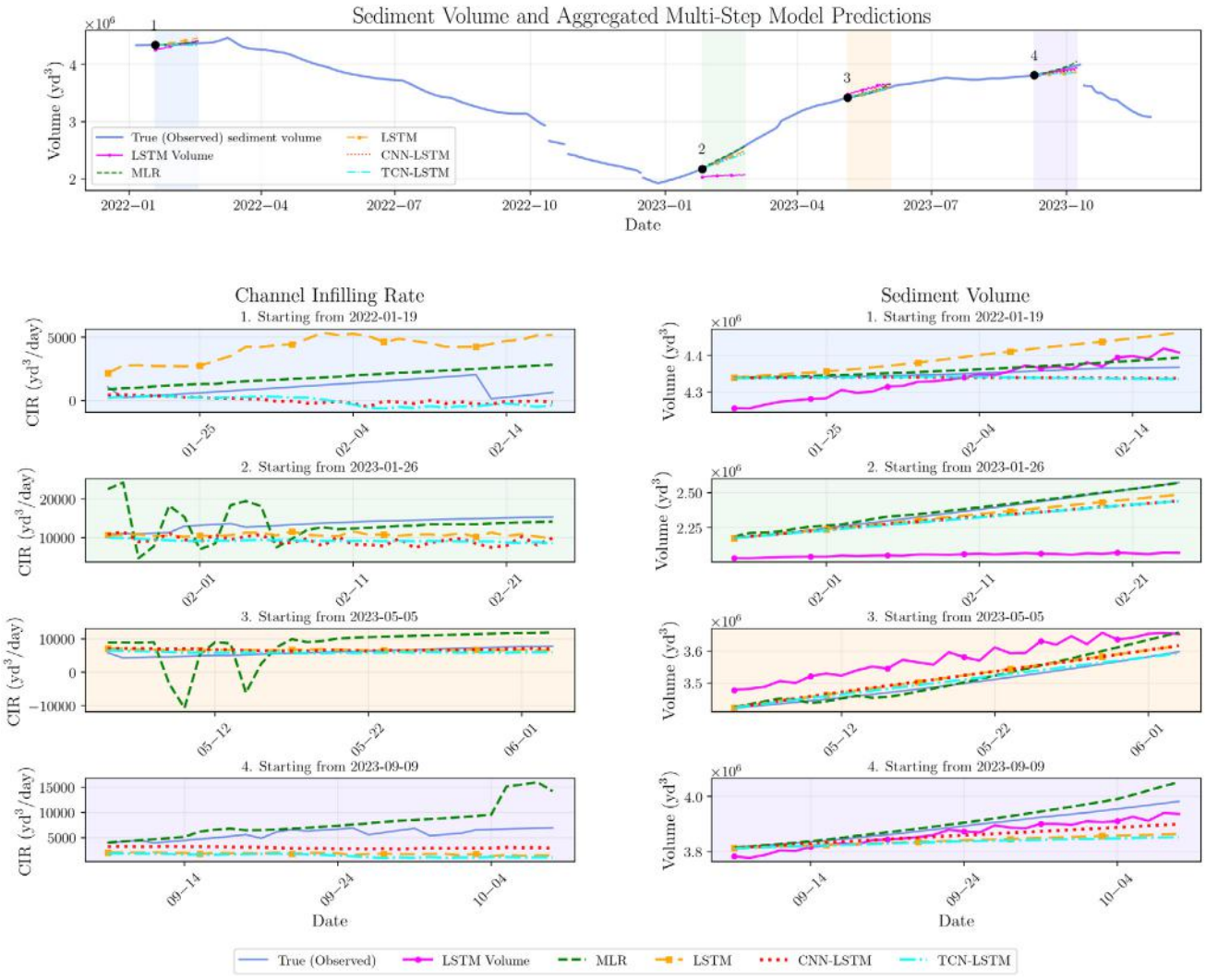


Fig. 17. HSC region: Forecasting of cumulative sedimentation volume under the aggregated modeling strategy, illustrated for four selected forecast start dates and compared across different modeling approaches. The top panel shows the observed aggregated sediment-volume time series over the test period, with four selected forecast windows highlighted by shaded regions and labeled by forecast start date. The lower panels provide detailed comparisons for these four forecast windows: the left column shows the predicted CIR, and the right column shows the corresponding sediment-volume forecasts obtained by integrating the predicted CIR. Each row corresponds to one forecast window.

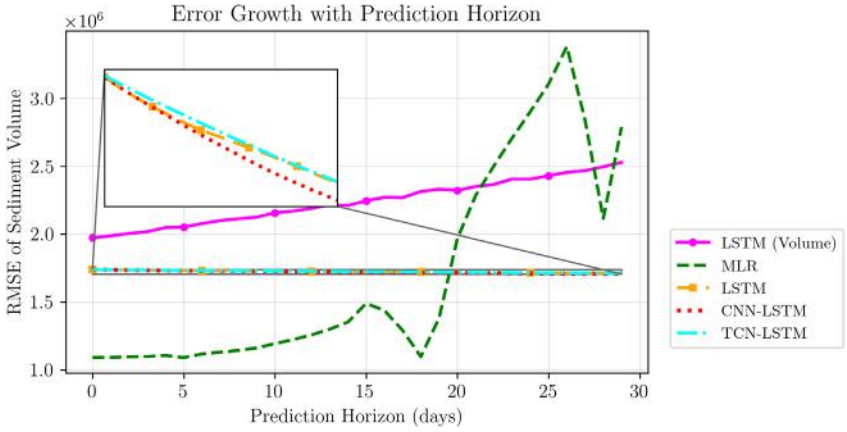


Fig. 18. HSC region: Cumulative sedimentation volume forecasting error growth with prediction horizon using different approaches.

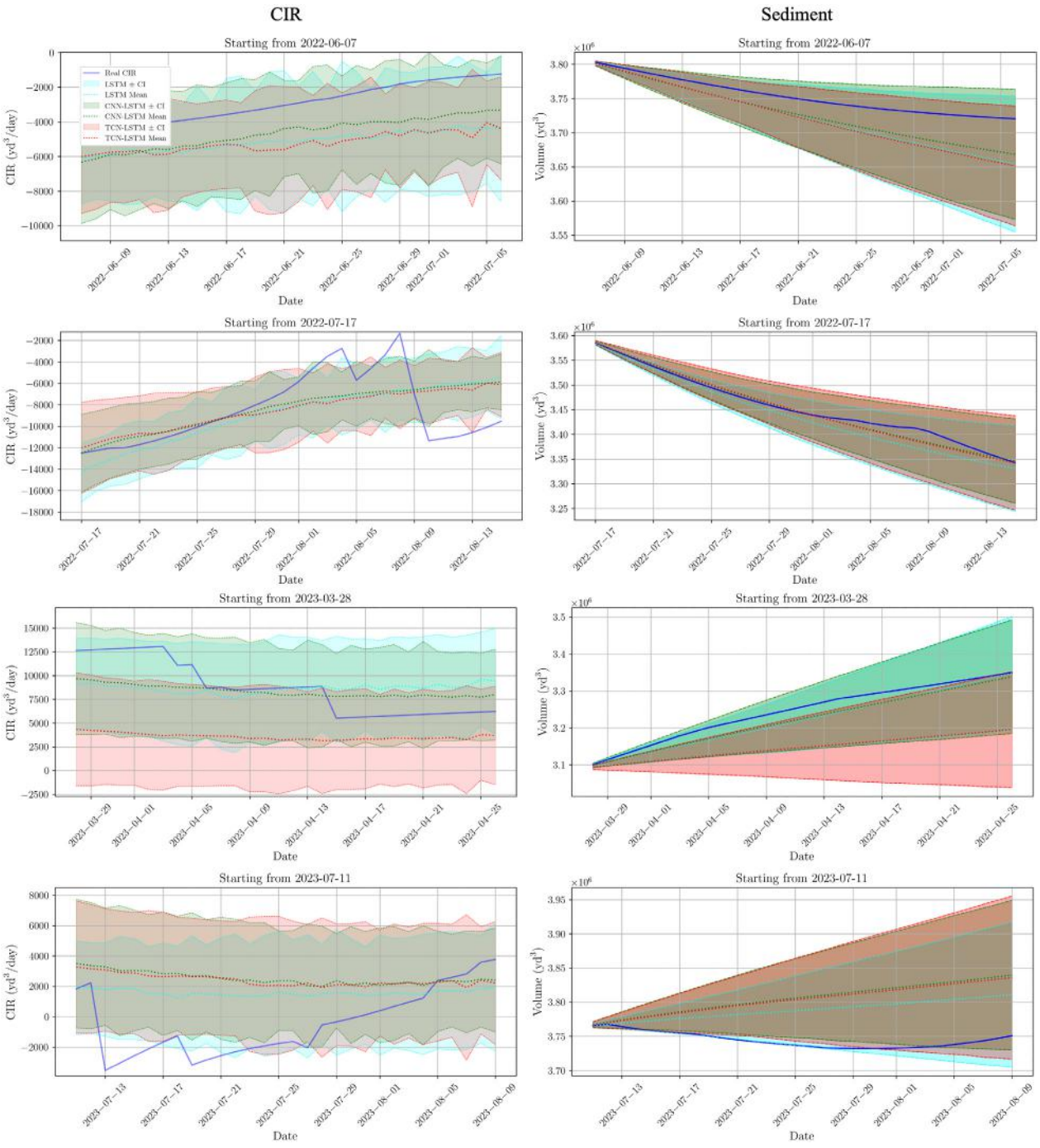


Fig. 19. HSC region: Comparison of CIR forecasts and the corresponding region-scale cumulative sediment volume predictions with uncertainty bounds across different modeling approaches. The left column shows CIR predictions, while the right column presents the associated cumulative sediment volume forecasts.

Table 4
HSC region: Evaluation metrics for cumulative sediment volume prediction under different modeling strategies using the three most sensitive gauges (MLR). Columns correspond to four selected forecast start dates and the full test set. Best values for the full test set are highlighted in bold.

Scenario	Metric	2022-01-19	2023-01-26	2023-05-05	2023-09-09	Full test set
Scenario 1 *(Disaggregated)*	RMSE (yd^3)	12,294	13,003	57,246	41,596	41,002
	nRMSE	0.0028	0.0054	0.0163	0.0106	**0.0074**
	MAPE (%)	0.23	0.41	1.37	0.67	**0.58**
Scenario 2 *(Aggregated)*	RMSE (yd^3)	12,294	15,907	26,843	26,116	**39,806**
	nRMSE	0.0028	0.0067	0.0076	0.0067	0.0087
	MAPE (%)	0.23	0.59	0.56	0.47	0.65

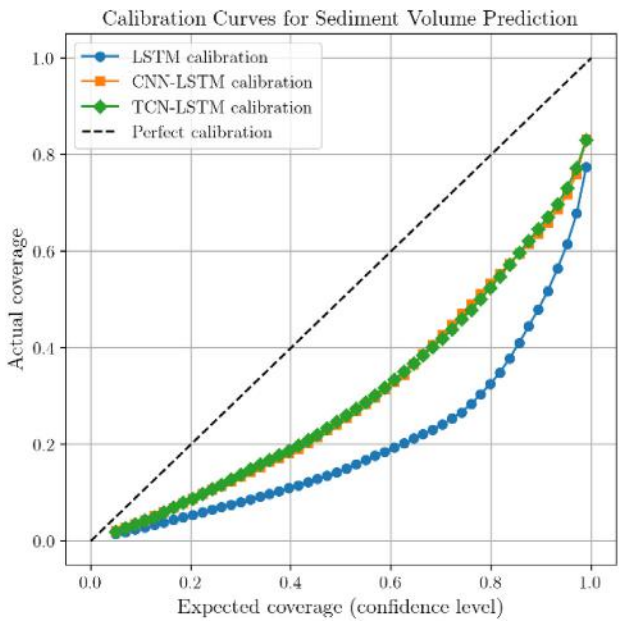


Fig. 20. HSC region: Cumulative sedimentation volume forecasting calibration curve.

case, direct cumulative-volume prediction using LSTM-Volume yields a full-test RMSE of 884,755 yd^3, nRMSE of 0.1167, and MAPE of 11.02%. In contrast, the best CIR-based model, MLR, reduces these errors to 39,806 yd^3, 0.0088, and 0.66%, respectively. This corresponds to reductions of approximately 95.5% in RMSE, 92.5% in nRMSE, and 94.0% in MAPE. These results indicate that modeling the rate process first and then integrating it into cumulative sediment volume substantially reduces error accumulation compared with direct volume prediction.

The left column of Fig. 17 highlights CIR forecasts, which serve as the primary drivers of cumulative volume predictions under the aggregated strategy. MLR produces comparatively smooth and linear CIR trends, whereas deep learning–based models (LSTM, CNN–LSTM, and TCN–LSTM) show more flexible temporal dynamics. These differences are important because even small systematic biases in CIR can accumulate over time and lead to large errors in the predicted sediment volume. The right column of Fig. 17 illustrates the resulting sediment volume forecasts. Consistent with Table 5, MLR achieves the best overall performance on the full test set, producing the lowest RMSE (39,806), nRMSE (0.0088), and MAPE (0.66). This result indicates that even with its simple structure, the linear model is able to produce more stable and robust volume estimates over long prediction horizons when using the aggregated strategy. Among deep learning models, CNN–LSTM outperforms both standard LSTM and TCN–LSTM, achieving lower full test set errors (RMSE = 93,143; nRMSE = 0.0174; MAPE = 1.43).

It should be noted that the LSTM Volume model shows the weakest generalization performance, with much higher errors on the full test set, suggesting that directly predicting cumulative volume without explicitly modeling CIR leads to greater error accumulation over time. In addition to the study above, the error growth is also of concern. Fig. 18 shows how sediment volume forecast error grows as the prediction horizon increases (day 0 to day 30) for the HSC region. Several findings are deduced from this figure:

- **Short-horizon vs. long-horizon trade-off (MLR):** MLR is much better than other models for near-term forecasts (0 to 20 days), but it becomes unreliable for longer horizons. This is consistent with a linear model producing small systematic biases in CIR that do not look great at day 0 to 20, but accumulate rapidly once you integrate over many days.
- **Why the three rate-based LSTM-based models look similar:** When the input information is limited and the evaluation is on aggregated volume, different architectures can end up learning very similar rate dynamics, especially if the true signal is smooth and the main challenge is avoiding bias rather than extracting complex patterns.
- **A practical strategy:** If your use case cares about relatively short-term forecasting (less than 20 days), MLR is the better choice. If you are more into longer planning horizons (e.g., larger than 30 days), use CNN-LSTM/TCN-LSTM/LSTM (rate-based) over MLR.

4.1.4. Uncertainty quantification of LSTM-based models

In this section, we focus on applying uncertainty quantification to the LSTM-based models (i.e., LSTM, CNN-LSTM, TCN-LSTM) for the HSC region using the aggregated strategy. Fig. 19 presents the UQ results for the HSC region using the LSTM-based models under the aggregated strategy. The left column shows CIR forecasts, while the right column shows the corresponding sedimentation-volume forecasts obtained by integrating the predicted CIR. Each row corresponds to one selected forecast start date. The shaded regions represent the prediction intervals estimated using MC-dropout, and the mean trajectories represent the corresponding model predictions. To quantitatively assess calibration, we computed the empirical coverage of the 95% confidence intervals for sedimentation-volume forecasts. For HSC, the empirical coverage values are 79.4%, 89.8%, and 89.2% for LSTM, CNN–LSTM, and TCN–LSTM, respectively. These results indicate that CNN–LSTM and TCN–LSTM provide better calibrated uncertainty estimates than the standard LSTM model in the sparse-survey HSC case. The lower coverage of the LSTM model suggests that its uncertainty bounds are relatively overconfident under data-scarce conditions.

Fig. 20 further evaluates the reliability of the probabilistic forecasts using calibration curves. The CNN-LSTM and TCN-LSTM models demonstrate calibration curves that lie closer to the perfect calibration line across a wide range of confidence levels, particularly in the mid-confidence range (approximately 0.3–0.8). In contrast, the LSTM model consistently performs worse than the other two, especially at higher confidence levels, indicating over-confident uncertainty estimates.

4.2. Southwest Pass: Forecasting under dense survey conditions

This section discusses the results and observations inferred from the prediction outcomes in the SWP region. Fig. 21 shows four reach samples of the sediment volume and CIR in the SWP region. The average number of surveys is noticeably higher than that observed in the HSC region. The following observations are drawn from the results obtained for the SWP region.

4.2.1. Disaggregated strategy

Fig. 22 illustrates the predicted sedimentation volume and CIR at the reach level for a representative reach (CEMVN_SW_05_SWP_01) in the SWP region at four different time instances. This reach is one of the most frequently dredged and most intensively surveyed locations within the SWP, making it a particularly informative test case for evaluating model performance under data-rich conditions. Table 6 reports the error metrics corresponding to these predictions. The full test set consists of 696 filtered time steps, covering the period from January 1, 2022, to December 31, 2023. Based on Fig. 22 and Table 6, we observe that LSTM proves better suited for the SWP region than ULR or MLR, which yield poor predictive performance with significantly larger errors. Since the SWP region is richly surveyed and captures nonlinearities in the sedimentation process, using an advanced regression technique with the capacity to learn such nonlinearities and patterns in the data (such as LSTM) is advantageous, particularly

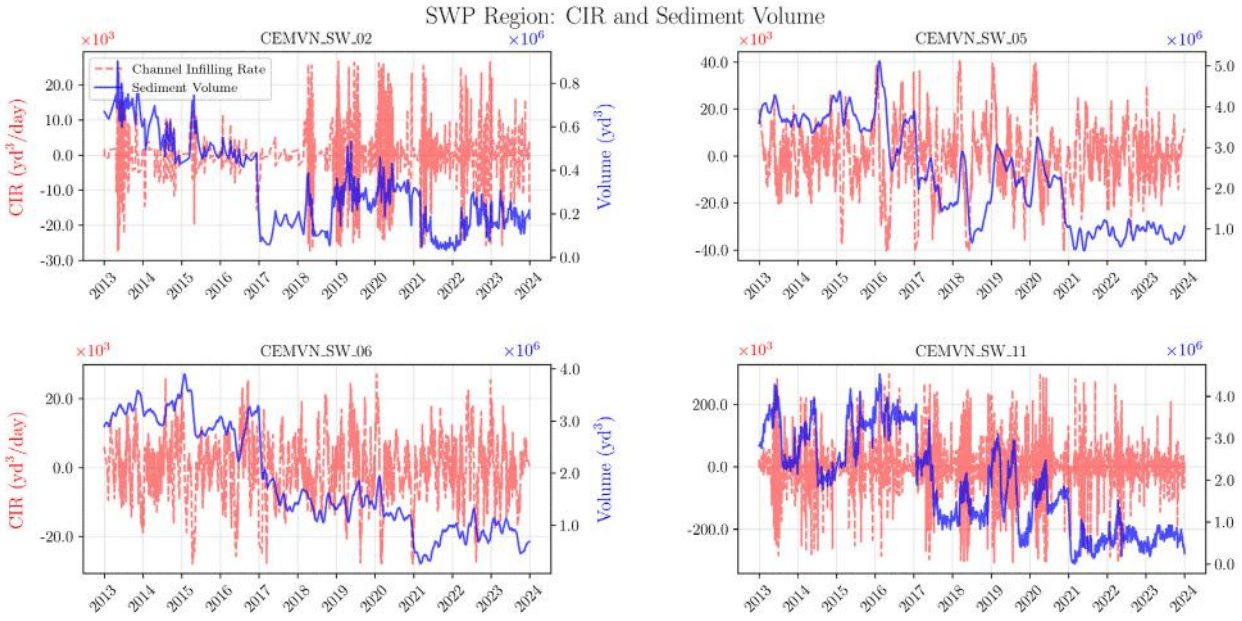


Fig. 21. Sedimentation volume and CIR for representative reaches in the SWP region (The legend for all subplots is identical to that of the top-left figure).

Table 5
HSC region: Evaluation metrics for cumulative sediment-volume prediction under the aggregated modeling strategy, compared across different modeling approaches. Columns correspond to four selected forecast start dates and the full test set. For the full test set, the best value for each metric is highlighted in bold.

Method	Metric	2022-01-19	2023-01-26	2023-05-05	2023-09-09	Full test set
LSTM volume	RMSE (yd^3)	40,876	328,469	78,217	29,865	884,755
	nRMSE	0.0094	0.1389	0.0224	0.0077	0.1167
	MAPE (%)	0.75	12.90	2.20	0.71	11.02
MLR	RMSE (yd^3)	12,294	15,907	26,843	26,116	**39,806**
	nRMSE	0.0028	0.0067	0.0077	0.0067	**0.0088**
	MAPE (%)	0.23	0.59	0.56	0.47	**0.66**
LSTM	RMSE (yd^3)	52,259	42,951	20,536	63,507	97,041
	nRMSE	0.0120	0.0182	0.0059	0.0163	0.0191
	MAPE (%)	1.02	1.37	0.55	1.34	1.57
CNN–LSTM	RMSE (yd^3)	16,128	66,579	20,589	43,083	93,143
	nRMSE	0.0037	0.0282	0.0059	0.0111	0.0174
	MAPE (%)	0.28	2.12	0.56	0.88	1.43
TCN–LSTM	RMSE (yd^3)	16,806	68,710	8832	69,497	97,350
	nRMSE	0.0039	0.0291	0.0025	0.0179	0.0191
	MAPE (%)	0.29	2.27	0.23	1.46	1.57

Table 6
SWP region: Evaluation metrics for sedimentation-volume predictions at the reach-level for a representative reach (i.e., CEMVN_SW_05_SWP_01), under the disaggregated modeling strategy, compared across different modeling approaches. Columns correspond to four selected forecast start dates and the full test set. Best values for the full test set are highlighted in bold.

Method	Metric	2022-03-20	2022-11-16	2023-04-25	2023-10-01	Full test set
ULR	RMSE (yd^3)	5,520,827	11,339,814	4,539,048	1,741,658	10,463,000
	nRMSE	7.5616	11.1807	4.4317	2.0111	7.8219
	MAPE (%)	590.62	887.58	340.00	151.30	612.07
MLR	RMSE (yd^3)	8,128,286	10,394,928	3,013,574	179,755	34,662,213
	nRMSE	11.1329	10.2491	2.9423	0.2076	11.6732
	MAPE (%)	898.81	796.47	192.88	18.16	864.63
LSTM	RMSE (yd^3)	76,599	92,002	186,806	15,158	**140,650**
	nRMSE	0.1049	0.0907	0.1824	0.0175	**0.1373**
	MAPE (%)	8.64	6.47	13.07	1.49	**11.49**

for capturing curvatures and inflection points in the sedimentation data. In contrast, a simple regression technique like linear regression is incapable of learning these patterns. Unlike the HSC region, where predictive superiority depended primarily on aligning with the nature of the interpolated training data, in the SWP region it arises from better learning of the underlying sedimentation dynamics. The following two conclusions are made considering LSTM as the chosen regression technique.

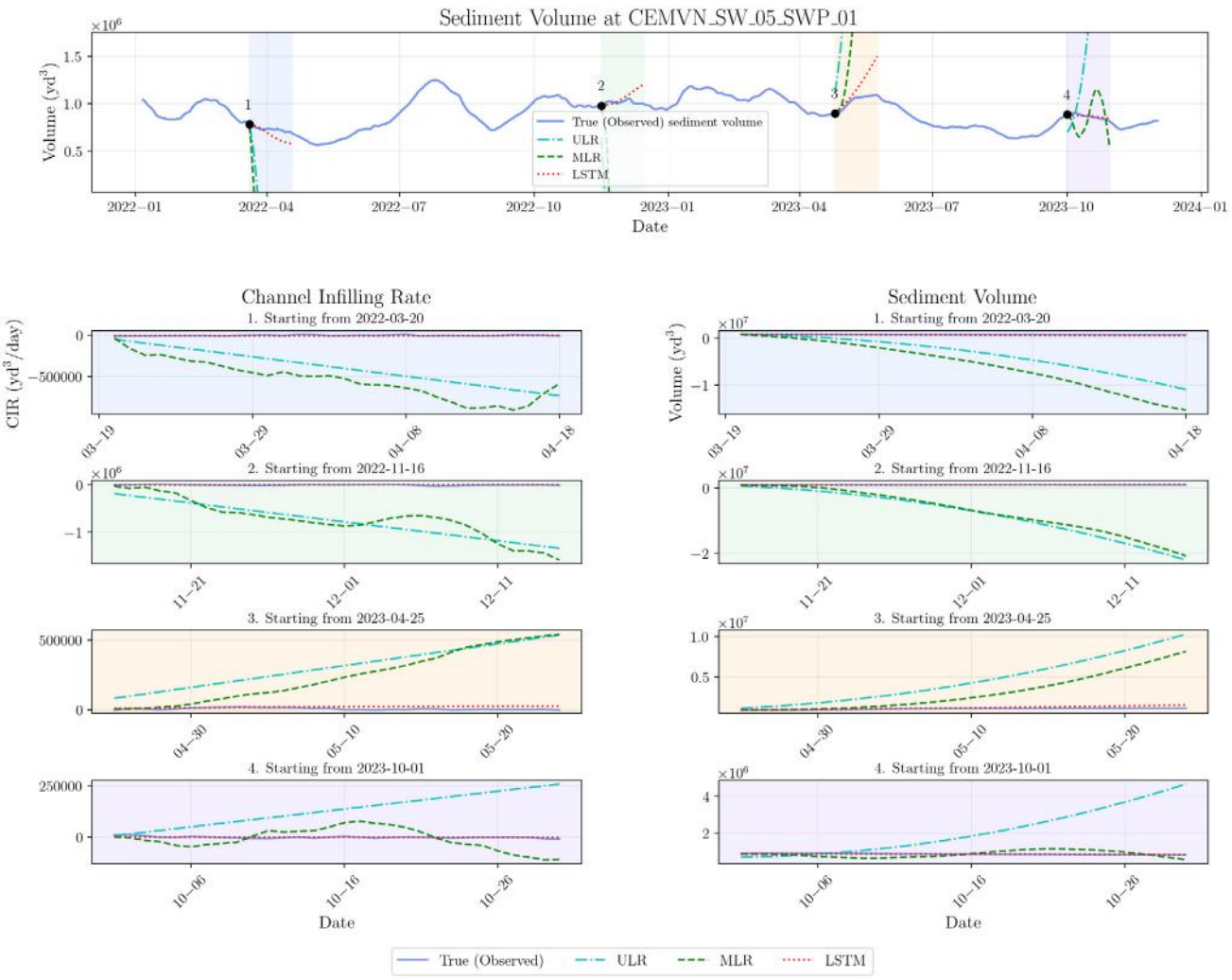


Fig. 22. SWP region: Forecasts of CIR and sediment volume at a representative individual reach (i.e., CEMVN_SW_05_SWP_01) under the disaggregated modeling strategy, compared across three modeling approaches. The top panel shows the observed sedimentation-volume time series over the test period, with four selected forecast windows highlighted by shaded regions and labeled by forecast start date. The lower panels show the detailed forecasts for these four windows: the left column presents the predicted CIR, and the right column presents the corresponding sedimentation-volume forecasts obtained by integrating the predicted CIR. Each row corresponds to one forecast window.

Table 7
SWP region: Evaluation metrics for reach-level sedimentation-volume predictions at a representative reach (i.e., CEMVN_SW_05_SWP_01) under the disaggregated modeling strategy using different sets of sensitive gauges (LSTM). Columns correspond to four selected forecast start dates and the full test set. Best values for the full test set are highlighted in bold.

Input	Metric	2022-02-06	2022-07-15	2023-02-03	2023-08-28	Full test set
Top 3 most sensitive gauges	RMSE (yd^3)	235,650	65,035	248,821	55,733	152,711
	nRMSE	0.2366	0.0555	0.2326	0.0800	0.1403
	MAPE (%)	16.89	4.78	19.97	6.47	12.15
Top 8 most sensitive gauges	RMSE (yd^3)	44,253	109,620	65,651	110,977	**140,650**
	nRMSE	0.0444	0.0936	0.0614	0.1593	**0.1374**
	MAPE (%)	3.84	7.16	5.50	11.93	**11.49**

The predicted sedimentation volume using the LSTM model at the reach level is evaluated under two configurations based on the most sensitive upstream gauges. In the first configuration, three gauges are used, while in the second configuration, eight gauges are incorporated. Table 7 summarizes the corresponding error metrics. The full test set consists of 696 filtered time steps, covering the period from January 1, 2022, to December 31, 2023. Based on the results summarized in Table 7, the eight-gauge configuration generally achieves better predictive performance over the full test period in terms of nRMSE and MAPE. However, when examining individual prediction instances, the performance differences between the two configurations are less consistent, with each configuration outperforming the other in some cases. Overall, incorporating a limited number of additional gauge observations appears beneficial. Since the LSTM model can effectively handle higher-dimensional inputs, and because the SWP dataset derived from frequent surveys captures the underlying sedimentation dynamics, the inclusion of sensitive upstream gauge information likely provides meaningful supplementary signals that enhance prediction accuracy.

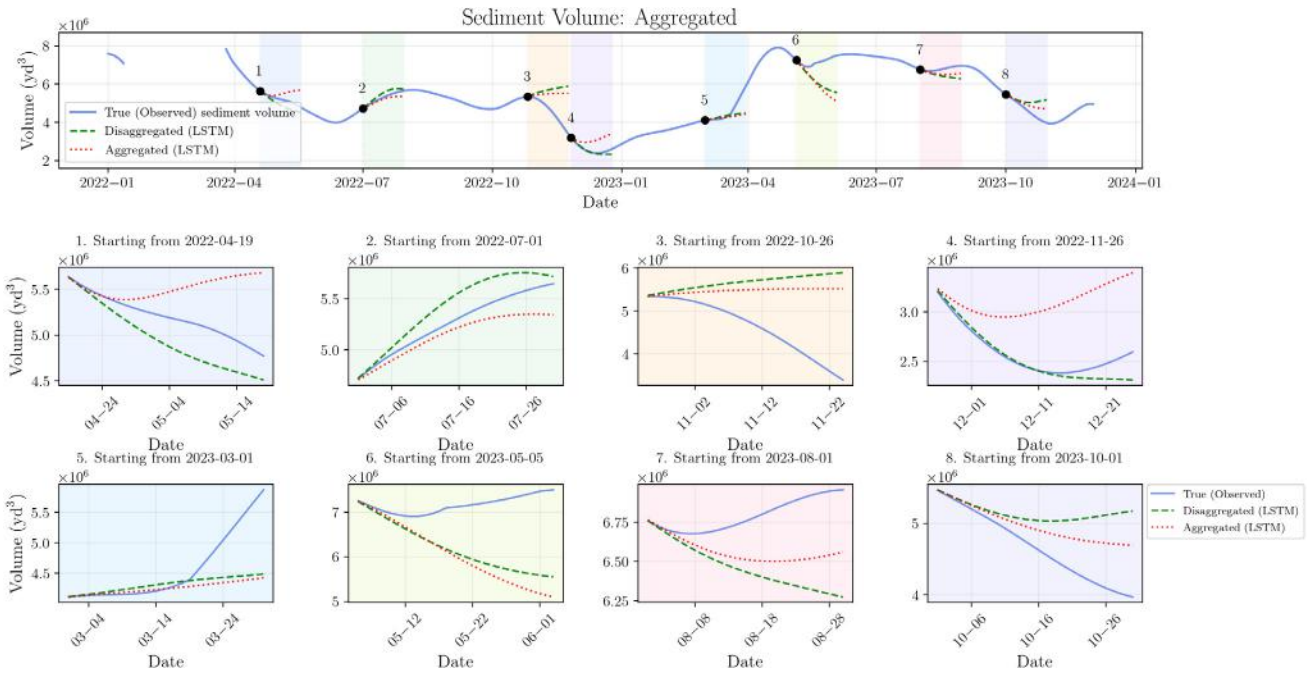


Fig. 23. SWP region: Forecasting of cumulative sedimentation volume aggregated over all reaches using two modeling strategies, illustrated for eight selected forecast start dates using the LSTM model with the eight most sensitive gauges. The top row shows the overall sediment volume time series together with the forecast windows. The second and third rows show the corresponding zoomed-in views of the sediment-volume forecasts. Each column corresponds to one forecast window.

Table 8
SWP region: Evaluation metrics for cumulative sediment volume predictions under two modeling strategies using the eight most sensitive gauges (LSTM). The eight forecast windows are shown in two blocks for readability, and the overall performance over the full test set is reported in the final column.

Scenario	Metric	2022-04-19	2022-07-01	2022-10-26	2022-11-26
Scenario 1 *(Disaggregated)*	RMSE (yd^3)	271,367	178,091	1,288,059	99,201
	nRMSE	0.0521	0.0338	0.2781	0.0382
	MAPE (%)	4.70	2.94	25.22	2.57
Scenario 2 *(Aggregated)*	RMSE (yd^3)	438,511	147,380	1,070,186	576,597
	nRMSE	0.0842	0.0280	0.2310	0.2225
	MAPE (%)	6.51	2.25	20.58	20.69

Scenario	Metric	2023-03-01	2023-05-05	2023-08-01	2023-10-01	Full test set
Scenario 1 *(Disaggregated)*	RMSE (yd^3)	516,189	1,165,344	399,398	624,958	**752,475**
	nRMSE	0.1132	0.1626	0.0588	0.1337	**0.1285**
	MAPE (%)	5.91	13.29	4.77	11.25	**11.27**
Scenario 2 *(Aggregated)*	RMSE (yd^3)	554,622	1,366,200	269,922	389,097	771,546
	nRMSE	0.1216	0.1907	0.0397	0.0832	0.1317
	MAPE (%)	6.07	15.04	3.27	7.13	11.38

4.2.2. Comparison between disaggregated and aggregated strategies

Fig. 23 illustrates the predicted cumulative sedimentation volume using LSTM for both Modeling *Scenario 1* (disaggregated strategy) and Modeling *Scenario 2* (aggregate strategy). Table 8 reports the corresponding error metrics. The full test set consists of 629 filtered time steps, covering the period from January 1, 2022, to December 31, 2023. Considering the full test set, *Scenario 1* performs marginally better than *Scenario 2*. Of the eight time instances reported, *Scenario 1* outperforms *Scenario 2* in four cases, while *Scenario 2* performs better in the remaining four. On balance, *Scenario 1* does slightly better overall. In this disaggregated scenario, each individual reach prediction generally maintains consistent accuracy, with errors showing little variation across neighboring reaches. For example, the prediction results for reach IDs SWP_04, SWP_05, and SWP_06 are relatively similar due to their spatial proximity and shared hydrological patterns. This level of accuracy is consistent across most reaches, though slightly worse for reaches such as SWP_01 and SWP_11, where data are more sparse. Therefore, similarly to the HSC region, for the SWP region, the clear conclusion is that cumulative volume prediction depends less on selecting the scenario and more on choosing the appropriate regression technique (variants of LSTM models over ULR or MLR).

4.2.3. Model comparison under the aggregated strategy

From the results above, LSTM-based models demonstrate better overall predictive performance, which can be attributed to the larger number of bathymetric surveys and the resulting rich data in the SWP region compared to the HSC region. Furthermore, the disaggregated and aggregated strategies have nearly identical prediction accuracy; Similar to the HSC region, the disaggregated strategy greatly reduces the number of available timesteps in the test set. Therefore, subsequent analyses in this section focus exclusively on the aggregated strategy. Within this framework, more advanced LSTM-based models are introduced for further comparison.

Fig. 24 presents four representative examples of the predicted CIR and the corresponding back-calculated cumulative sediment volume for the entire SWP region using different modeling approaches. Table 9

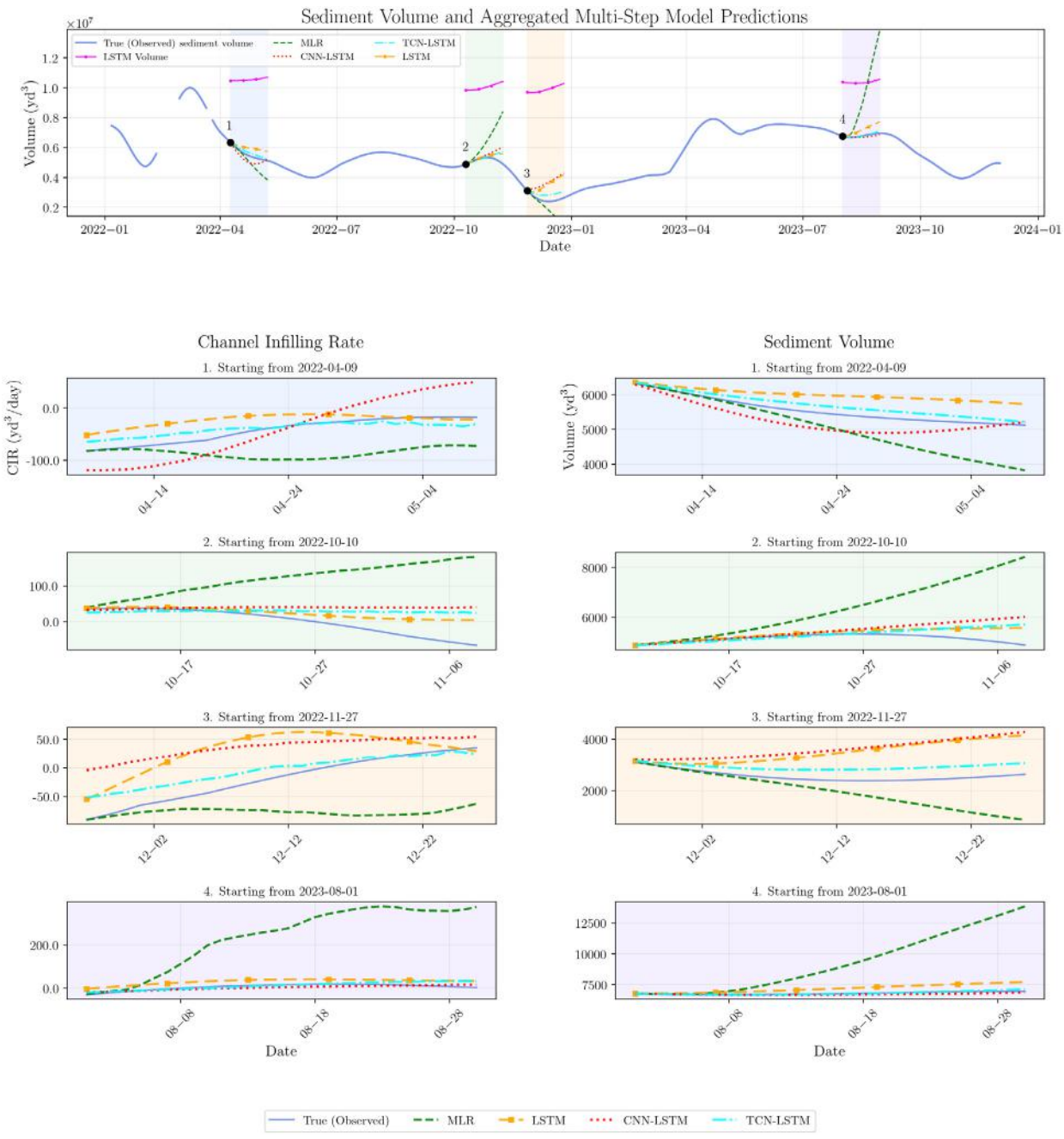


Fig. 24. SWP region: Forecasting of cumulative sedimentation volume under the aggregated modeling strategy, illustrated for four selected forecast start dates and compared across different modeling approaches. The top panel shows the observed aggregated sediment-volume time series over the test period, with four selected forecast windows highlighted by shaded regions and labeled by forecast start date. The lower panels provide detailed comparisons for these four forecast windows: the left column shows the predicted CIR, and the right column shows the corresponding sediment-volume forecasts obtained by integrating the predicted CIR. Each row corresponds to one forecast window.

summarizes the associated performance metrics. The figure shows the observed CIR and sediment volume evolution, together with model predictions obtained under the aggregated forecasting framework.

The left column of Fig. 24 highlights CIR predictions, which act as the primary drivers of cumulative sediment volume forecasts under the aggregated strategy. Similar to the HSC region, the MLR produces relatively smooth predictions, whereas the LSTM–based models (i.e., LSTM, CNN–LSTM, and TCN–LSTM) capture more complex temporal variability more accurately. The right column of Fig. 24 shows the corresponding sediment volume forecasts. The baseline LSTM Volume model shows huge deviations from observations across multiple starting times, which is consistent with its poor full test set performance reported in Table 9. It shows that the direct LSTM-Volume baseline gives a full-test RMSE of 4,997,062 yd^3, nRMSE of 1.0555, and MAPE of 106.66%. By comparison, the best CIR-based model, TCN-LSTM, reduces these errors to 764,065 yd^3, 0.1240, and 10.63%, respectively. This corresponds to reductions of approximately 84.7% in RMSE, 88.3% in nRMSE, and 90.0% in MAPE. These results further confirm that CIR-based forecasting provides a more stable and accurate prediction target than cumulative sedimentation volume, especially when datum

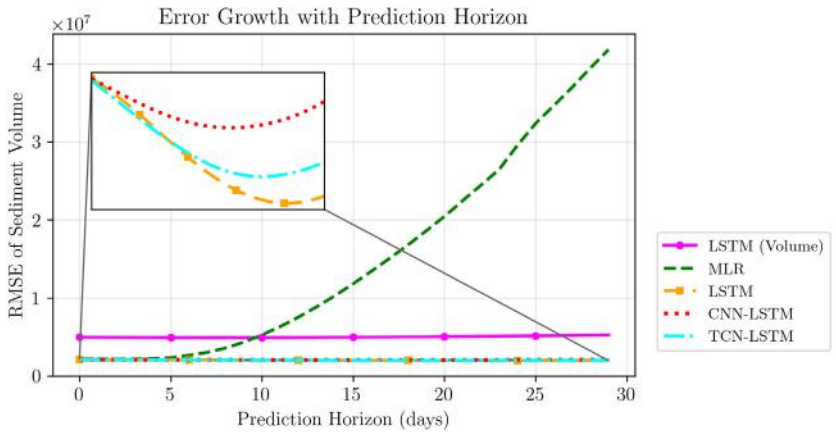


Fig. 25. SWP region: Cumulative sedimentation volume forecasting error growth with prediction horizon using different approaches.

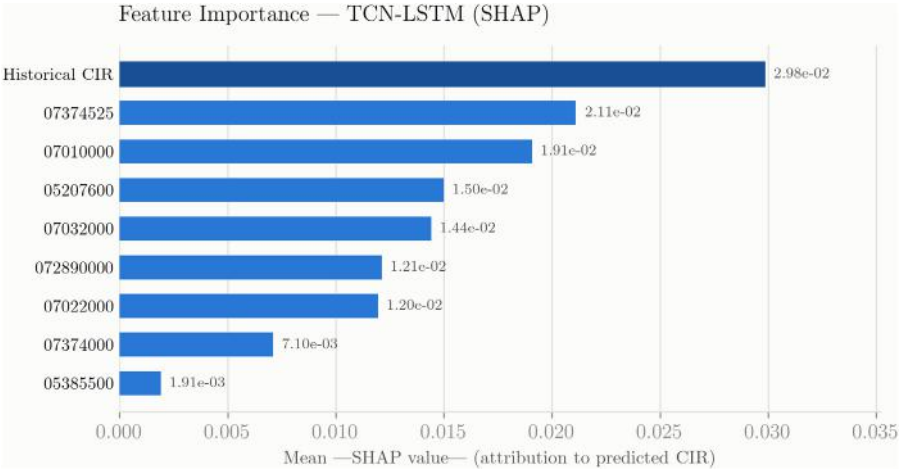


Fig. 26. SHAP-based feature importance for the TCN-LSTM model in the SWP aggregated case. The mean absolute SHAP value measures the average contribution of each input feature to the predicted CIR.

Table 9
SWP region: Evaluation metrics for cumulative sediment volume prediction under the aggregated modeling strategy, compared across different modeling approaches. Columns correspond to four selected forecast start dates and the full test set. Best values for the full test set are highlighted in bold.

Method	Metric	2022-04-09	2022-10-10	2022-11-27	2023-08-01	Full test set
LSTM volume	RMSE (yd^3)	5,010,233	4,888,630	7,326,433	3,575,668	4,997,062
	nRMSE	0.9021	0.9465	2.8484	0.5264	1.0555
	MAPE (%)	90.73	94.69	286.90	52.65	106.66
MLR	RMSE (yd^3)	662,289	1,599,133	807,892	3,344,661	1,922,644
	nRMSE	0.1193	0.3096	0.3141	0.4924	0.4692
	MAPE (%)	9.38	22.89	23.40	36.10	34.60
LSTM	RMSE (yd^3)	483,983	266,726	1,054,880	448,818	771,546
	nRMSE	0.0871	0.0516	0.4101	0.0660	0.1317
	MAPE (%)	8.22	3.56	37.04	5.78	11.38
CNN–LSTM	RMSE (yd^3)	319,847	444,350	1,149,633	85,591	843,126
	nRMSE	0.0576	0.0860	0.4470	0.0126	0.1409
	MAPE (%)	5.15	5.56	41.84	1.01	12.22
TCN–LSTM	RMSE (yd^3)	162,343	296,768	364,625	50,991	**764,065**
	nRMSE	0.0292	0.0575	0.1418	0.0075	**0.1240**
	MAPE (%)	2.78	3.69	13.64	0.46	**10.63**

shifts or nonstationary volume levels make direct volume prediction difficult. Among CIR-based aggregated models, linear regression models like MLR cannot predict the sediment volume accurately compared to LSTM-based models. Among deep learning models, TCN–LSTM achieves the best overall performance (RMSE = 764,065; nRMSE = 0.1240; MAPE = 10.63).

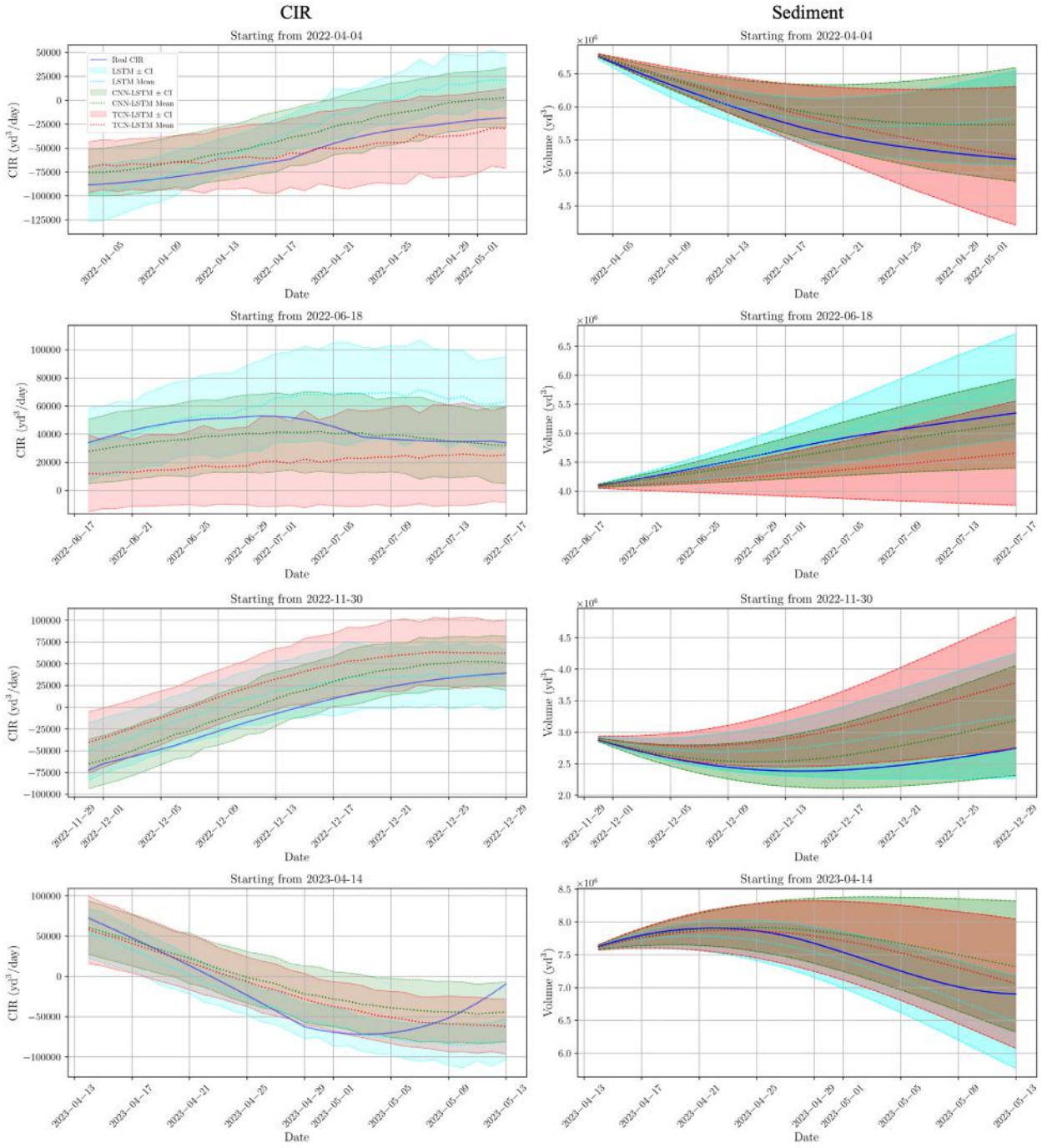


Fig. 27. SWP region: Comparison of CIR forecasts and the corresponding region-scale cumulative sediment volume predictions with uncertainty bounds across different modeling approaches. The left column shows CIR predictions, while the right column presents the associated cumulative sediment volume forecasts.

In addition to the study above, the error growth is also investigated. Fig. 25 shows how sediment volume forecast error grows as the prediction horizon increases (day 0 to day 30) for the SWP region. Several findings from this figure:

- **Linear regression is not suitable for this case:** From the results of MLR, the error increases drastically at longer horizons, reflecting strong error accumulation from small systematic biases in CIR. The low accuracy makes this method not suitable for regions like SWP.
- **Direct volume prediction produces large error:** It can be observed that the direct sediment volume prediction model fails to produce reliable results. This limitation arises because, during normalization, the training set does not include the test set values; owing to a datum shift, the test set values are almost entirely outside the range of the training data. As a result, the model is unable to generate accurate predictions. These findings indicate that direct sediment volume prediction is not suitable for this case study, unless data preparation and partitioning account for datum shifts and changes in maintained depth.
- **A practical strategy:** For regions like SWP that have rich survey data, it is recommended to use deep learning CIR prediction models, such as the TCN-LSTM model used in this case study.

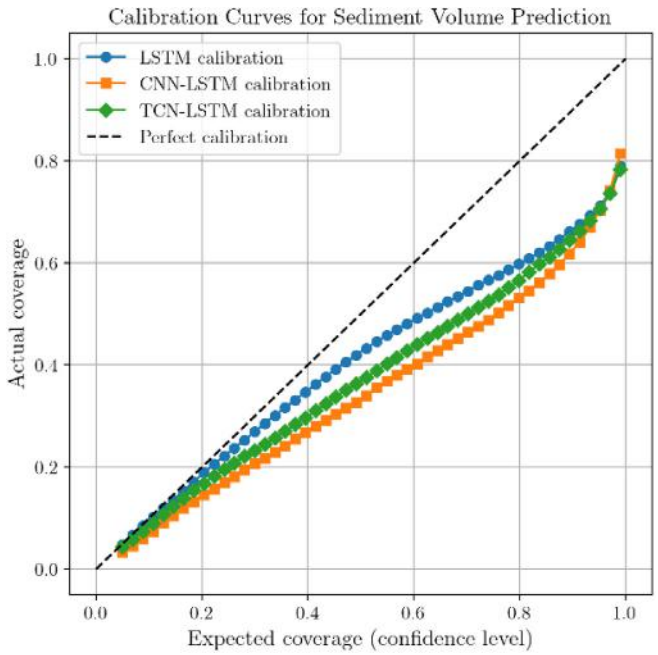


Fig. 28. SWP region: Cumulative sedimentation volume forecasting calibration curve.

Table 10
Summary of model performance under different survey-data conditions.

Region/data condition	Model	RMSE (yd^3)	nRMSE	MAPE (%)
HSC/sparse surveys	LSTM volume	884,755	0.1167	11.02
HSC/sparse surveys	MLR	**39,806**	**0.0088**	**0.66**
HSC/sparse surveys	LSTM	97,041	0.0191	1.57
HSC/sparse surveys	CNN–LSTM	93,143	0.0174	1.43
HSC/sparse surveys	TCN–LSTM	97,350	0.0191	1.57
SWP/dense surveys	LSTM volume	4,997,062	1.0555	106.66
SWP/dense surveys	MLR	1,922,644	0.4692	34.60
SWP/dense surveys	LSTM	771,546	0.1317	11.38
SWP/dense surveys	CNN–LSTM	843,126	0.1409	12.22
SWP/dense surveys	TCN–LSTM	**764,065**	**0.1240**	**10.63**

To further examine the interpretability of the best-performing deep learning model in the SWP region, we conducted a SHAP-based feature-importance analysis for the TCN-LSTM model under the aggregated strategy. Fig. 26 shows the mean absolute SHAP values of the input features with respect to the predicted CIR. The results indicate that historical CIR is the most important predictor, with a mean absolute SHAP value of 2.98×10^{-2}. This suggests that the TCN-LSTM model relies strongly on recent sedimentation-rate history when generating multi-step CIR forecasts. In addition to historical CIR, several upstream streamflow gauges also show contributions to the predicted CIR. The most influential streamflow gauges include 07374525 to 05385500, with mean absolute SHAP values ranging from 1.91×10^{-3} to 2.11×10^{-2}. These results indicate that the model does not rely solely on historical CIR information, but also extracts useful hydrologic signals from upstream discharge records. Therefore, by using the SHAP analysis after the model training and validation, we learn that the proposed model captures a physically meaningful relationship between upstream streamflow variability and downstream CIR response in the case study regions. It is worth noting that this analysis complements the GSA-based feature-selection procedure: GSA identifies influential upstream gauges before model training, whereas SHAP explains the relative contribution of historical CIR and selected streamflow inputs after the model has been trained.

4.2.4. Uncertainty quantification of LSTM-based models

In this section, we focus on applying uncertainty quantification to the LSTM-based models for the SWP region using the aggregated strategy. Fig. 27 presents the UQ results for the SWP region using the LSTM-based models under the aggregated strategy. Similar to the HSC case, the left column shows CIR forecasts and the right column shows the corresponding sedimentation-volume forecasts obtained by integrating the predicted CIR. Each row corresponds to one selected forecast start date. The shaded regions represent MC-dropout-based prediction intervals. Compared with HSC, the SWP region exhibits more complex nonlinear CIR dynamics, but the denser survey record provides more informative training data for both point prediction and uncertainty estimation. The uncertainty intervals generally capture the observed CIR and sedimentation-volume trends across the selected forecast windows. As expected, uncertainty in the sedimentation-volume forecasts increases with the prediction horizon because CIR uncertainty is accumulated during the integration process. For SWP, the empirical coverage values of the 95% confidence intervals for sedimentation-volume forecasts are 91.5%, 90.2%, and 92.9% for LSTM, CNN–LSTM, and TCN–LSTM, respectively. These values are closer to the nominal 95% confidence level than those observed for HSC, suggesting that the dense SWP survey record supports more reliable uncertainty estimation. Among the three models, TCN–LSTM provides the highest empirical coverage, which is consistent with its strong point-prediction performance in the SWP region.

Fig. 28 further evaluates the reliability of the probabilistic forecasts using calibration curves. Three models show similar results, while the LSTM is slightly closer to the perfect calibration curve. Compared to Fig. 20, the calibration curve indicates that the overall uncertainty quantification of the LSTM-based models in the SWP region is closer to the perfect calibration curve, suggesting reduced overconfidence relative to the results obtained for the HSC region.

4.3. Summary of model performance and data-availability effects

To provide a clearer comparison across the two study regions, Table 10 summarizes the full-test performance of the candidate models under two representative data-availability conditions. The HSC case represents a relatively sparse-survey setting, whereas the SWP case represents a dense-survey setting. For HSC, MLR achieves the best overall performance among the aggregated models, with a full-test RMSE of 39,806 yd^3, nRMSE of 0.0088, and MAPE of 0.66%. The larger errors produced by the LSTM-based models suggest that highly flexible models may not be advantageous when the reconstructed sedimentation-volume and CIR series are largely constrained by interpolation between infrequent surveys. Under such sparse-data conditions, the CIR series contains limited high-frequency observational information and often follows a more piecewise-linear structure, making simpler local regression models more stable. In contrast, for SWP, the TCN–LSTM model achieves the best overall performance, with a full-test RMSE of 764,065 yd^3, nRMSE of 0.1240, and MAPE of 10.63%. The denser SWP survey record better preserves nonlinear temporal variability associated with shoaling dynamics, providing a more suitable setting for LSTM-based models to learn temporal dependencies and nonlinear patterns. Therefore, the contrasting results between HSC and SWP indicate that model complexity should be matched to the information content of the available survey record, rather than interpreted as evidence that one model class is universally superior.

Fig. 29 further compares the reach-level forecasting performance of the optimal model in HSC and SWP. The results show that the proposed framework performs consistently across multiple individual reaches in both regions, indicating robustness to spatial variability in local sedimentation behavior. This suggests that the framework is transferable at the workflow level and can be adapted to multiple reaches within a navigation system. However, broader transferability

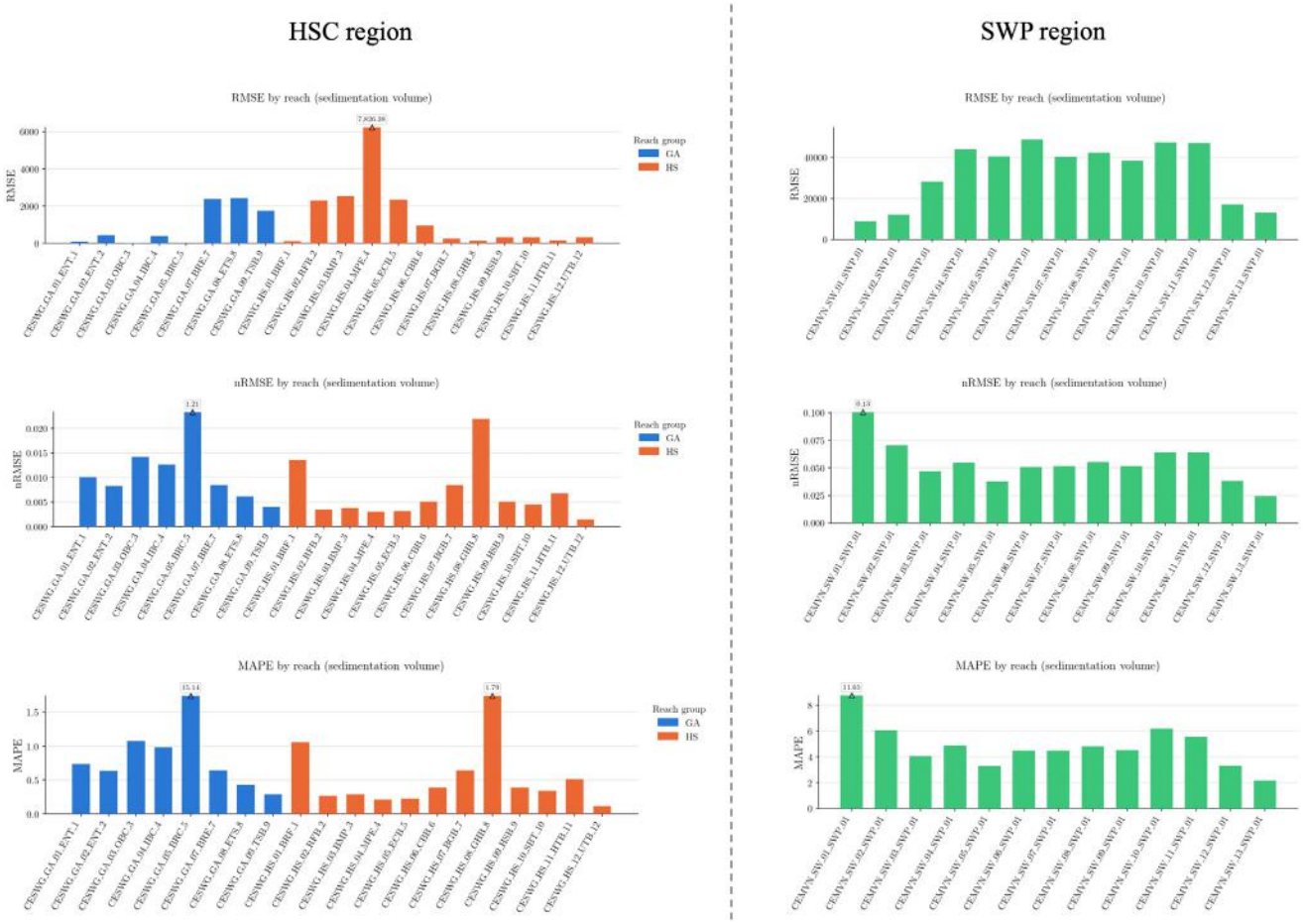


Fig. 29. Reach-level forecasting performance of the selected optimal model in the HSC and SWP regions. The left panels show the HSC results (GA: Galveston area; HS: Houston Ship Channel) and the right panels show the SWP results.

Table 11
Forecast-horizon sensitivity of the best-performing model in each region.

Region	Data condition	Selected model	Horizon	RMSE (yd^3)	nRMSE	MAPE (%)
SWP	Dense surveys	TCN–LSTM	30 days	764,065	0.1240	10.63
SWP	Dense surveys	TCN–LSTM	60 days	1,573,323	0.2576	22.76
SWP	Dense surveys	TCN–LSTM	90 days	2,141,280	0.3780	34.57

to other channels should still be interpreted with caution, because sediment dynamics, survey frequency, and hydrologic controls may differ substantially across systems. Therefore, application to a new channel requires local feature selection, model retraining, and independent validation.

A forecast-horizon sensitivity analysis was also conducted for the best-performing model in each region using 30-, 60-, and 90-day prediction horizons. The sample results for the SWP region are shown in Table 11. The results show that forecasting errors generally increase with longer horizons, which is expected because predicted CIR values are integrated over time to recover cumulative sedimentation volume. Therefore, small CIR biases can accumulate more strongly for 60- and 90-day forecasts. It should be noted that 60- and 90-day forecasts are not required for the current short-term dredging operation considered in this study, where the 30-day horizon is more directly aligned with operational sediment-management needs. Instead, the longer-horizon experiments are included as a sensitivity analysis to evaluate how the proposed framework behaves when the forecast window is extended. These results indicate that the proposed framework is most reliable for short-term operational forecasting, while longer-horizon forecasts remain feasible but should be interpreted with greater caution and accompanied by uncertainty estimates.

4.4. Computational requirements and model updating

The proposed framework is intended for operational short-term shoaling forecasting; therefore, its computational requirements are an important practical consideration. In the present implementation, the main computational cost is associated with two offline steps: GSA-based feature selection and model training. For the SWP region, using 79 discharge-related input variables, the GSA analysis required approximately 40 min on a MacBook Pro with an Apple M4 Pro chip and 32 GB of unified memory. After the influential streamflow gauges were selected, model training required approximately 20 min per reach. These costs are incurred during model development or recalibration rather than during routine forecasting. Once trained, the model can generate rolling short-term CIR and sediment-volume forecasts with negligible computational cost relative to the offline feature-selection and training stages.

The predictive models do not need to be retrained for every new forecast. Instead, retraining should be performed periodically when sufficient new bathymetric survey data have accumulated, or when major hydrological, morphological, dredging, or infrastructure changes occur. For data-rich regions such as SWP, model retraining can be conducted

more regularly, for example on a monthly or seasonal basis depending on operational needs. For data-sparse regions such as HSC, retraining may be less frequent and should be triggered after enough new survey observations are available to meaningfully update the CIR time series. Similarly, the GSA-based feature-selection step does not need to be repeated for every forecast, but should be recomputed during periodic recalibration or when substantial changes in hydrological conditions suggest that the dominant upstream controls may have changed.

5. Conclusion

5.1. Summary

This work proposed a CIR-based forecasting framework for sedimentation volume prediction in navigable river reaches, shifting the modeling focus from directly predicting cumulative volumes to forecasting the CIR. By reframing the problem in terms of CIR, the framework improves sensitivity to short-term sediment dynamics, reduces the influence of abrupt external interventions such as dredging, and improves interpretability relative to raw volume forecasts. The dual-modeling strategy, using linear regression in relatively data-scarce reaches such as the Houston Ship Channel (HSC) and LSTM-based models in data-rich reaches such as the Southwest Pass (SWP, which had on average 5 times more bathymetric surveys between 2012–2024 than HSC), demonstrated that model selection must align with survey density and data richness.

Results from HSC confirm that simpler linear regression approaches are more effective when bathymetric surveys are sparse, and interpolation dominates the data structure, whereas LSTM-based models yield poor performance under such conditions. Conversely, in SWP, where dense survey records capture non-linear sedimentation dynamics, LSTM-based models significantly outperformed linear models, successfully learning complex patterns and improving prediction accuracy. Sensitivity analysis further highlighted that while additional upstream gauge data added little value in HSC, it improved predictions in SWP by providing complementary hydrologic signals. Among them, the CNN-LSTM and TCN-LSTM architectures provide consistent, though small-scale improvements over the standard LSTM by incorporating convolutional feature extraction mechanisms that better capture local temporal patterns and multiscale dependencies in CIR. Moreover, differences between disaggregated (reach-level) and aggregated (system-level, consisting of numerous reaches) modeling strategies were marginal, indicating that forecasting accuracy depends more on the choice of regression technique than on scenario selection.

Beyond point predictions, this study integrates uncertainty quantification for all LSTM-based models using MC-dropout. The resulting predictive intervals provide valuable insights into forecast confidence and error growth with the prediction horizon. The Uncertainty Quantification (UQ) results indicate that CNN-LSTM and TCN-LSTM models exhibit slightly improved calibration and more stable uncertainty bounds compared to the standard LSTM, particularly for longer lead times. These findings highlight that even when point-prediction accuracy is similar, differences in uncertainty behavior can be critical for risk-aware decision-making in sediment management.

The practical significance of the proposed framework lies in its ability to support short-term sediment management under different data-availability conditions. The comparison between HSC and SWP indicates that the appropriate model choice depends on the information content of the available survey record. For data-sparse channels, simpler local regression models may provide more stable forecasts, whereas for data-rich channels, LSTM-based models can better exploit nonlinear temporal patterns in CIR. This data-adaptive model-selection strategy is directly relevant to real-world navigation-channel management, where survey frequency and data quality often vary across districts and reaches.

5.2. Limitations and future work

Although this study systematically evaluated multiple approaches to CIR forecasting across two regions, several limitations remain to be addressed in future research.

- First, the performance of the proposed framework depends on the availability and temporal density of bathymetric survey data. In data-sparse regions, such as HSC, the reconstructed daily sedimentation-volume and CIR series can be strongly influenced by interpolation between surveys. Under such conditions, deep learning models may not have sufficient reliable temporal information to learn robust nonlinear patterns and may be more prone to overfitting. This helps explain why simpler regression models performed better in HSC, whereas LSTM-based models were more effective in the data-rich SWP case.
- Second, the framework is demonstrated using two navigation channels. Although HSC and SWP represent contrasting survey-data conditions, the findings should not be interpreted as universal across all channels. Application to other systems requires local validation, recalibration, and reassessment of influential upstream predictors because sediment dynamics, hydrological forcing, dredging practices, channel morphology, and survey frequency may differ by region.
- Third, the current predictor pool focuses primarily on upstream streamflow gauges to maintain operational feasibility, since streamflow is routinely monitored and forecastable. As mentioned in Section 2.1, other variables, such as tides, winds, waves, salinity, turbidity, rainfall, and storm-event indicators, may further improve prediction accuracy when reliable observations and forecasts are available. Their systematic integration will be investigated in future work.
- Fourth, the IQR-based filtering step identifies abrupt CIR anomalies statistically rather than through explicit event attribution. As a result, some naturally occurring extreme sedimentation or scour events may be attenuated. Therefore, the present framework is mainly intended for regular short-term shoaling forecasting, while rare event-driven responses may require additional event-specific predictors or separate modeling strategies.
- Finally, the uncertainty quantification method employed in this paper (i.e., MC dropout) can only capture a subset of the uncertainty sources associated with CIR forecasting, primarily epistemic uncertainty arising from limited training data and uncertainty in the neural network parameters. Other sources of uncertainty remain unaccounted for, which may partly explain the overconfidence indicated by the calibration curves in Figs. 20 and 28. Therefore, developing more comprehensive uncertainty quantification methods that provide reliable and well-calibrated uncertainty bounds for shoaling-volume predictions is an important direction that is worth studying in the future.

It is also important to note that the superior performance of simpler linear models relative to advanced deep-learning models in the HSC region may not accurately reflect the real physical predictability of the model. Instead, this apparent advantage may arise from uncertainty in the reference values (i.e., the values used as ground truth), as the actual shoaling volume for each grid cell is unknown. For sparsely surveyed grid cells, as decribed in Section 3.2.1, CSAT uses linear temporal interpolation to estimate these values, which are then treated as ground truth. Consequently, the stronger performance of linear models may simply be an artifact of the method used to generate the reference data rather than evidence of their greater predictive capability. As additional survey data become available, CSAT-derived estimates should more closely approximate the actual shoaling volumes, potentially altering the conclusions of the model comparison. Therefore, improving the estimation of shoaling-volume reference values in data-sparse regions

within CSAT represents an important and valuable direction for future research. Overall, the study highlights the importance of adapting predictive models to local data conditions and sediment dynamics. By integrating CIR-based formulation, gauge-informed feature selection, and adaptive modeling, the proposed framework offers a practical step toward robust, data-driven shoaling forecasts. These results have direct implications for proactive dredging management, enabling longer lead times, more efficient allocation of resources, and improved resilience of navigation infrastructure. As forecasting performance is significantly dependent on the availability of past bathymetric data, the findings also shed light on the need for comprehensive surveying of waterway segments that are critical for supply chains, and the use of modern methods to complement traditional surveying methods, like crowdsourced (voluntary) bathymetry (Scully et al., 2025).

CRediT authorship contribution statement

Yichao Zeng: Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Mayank Chadha:** Writing – original draft, Visualization, Investigation, Data curation. **Dania Ammar:** Writing – original draft, Validation, Investigation, Formal analysis, Data curation, Conceptualization. **Dingbao Wang:** Writing – review & editing, Validation, Investigation, Formal analysis, Data curation. **Magdalena Asborno:** Writing – review & editing, Resources, Project administration, Formal analysis, Data curation. **Sarah Miele:** Writing – review & editing, Supervision, Resources, Project administration, Investigation, Funding acquisition, Formal analysis. **Charles J. McKnight:** Writing – review & editing, Resources, Project administration, Funding acquisition, Formal analysis. **Natalie P. Memarsadeghi:** Writing – review & editing, Resources, Project administration, Funding acquisition, Formal analysis, Conceptualization. **Michael A. Hartman:** Writing – review & editing, Software, Project administration, Funding acquisition, Formal analysis, Data curation. **Kenneth N. Mitchell:** Writing – review & editing, Supervision, Resources, Project administration, Investigation, Funding acquisition, Formal analysis. **Guga Gugaratshan:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition, Formal analysis. **Michael D. Todd:** Writing – review & editing, Supervision, Investigation. **Zhen Hu:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work was supported by the United States Army Corps of Engineers through the U.S. Army Engineer Research and Development Center Research Cooperative Agreement W912HZ24C0044. The support is gratefully acknowledged.

References

Afan, H.A., El-shafie, A., Mohtar, W.H.M.W., Yaseen, Z.M., 2016. Past, present and prospect of an artificial intelligence (AI) based model for sediment transport prediction. J. Hydrol. 541, 902–913.

Alp, M., Cigizoglu, H.K., 2007. Suspended sediment load simulation by two artificial neural network methods using hydrometeorological data. Environ. Model. Softw. 22 (1), 2–13.

Asborno, M., Broders, J., Mitchell, K.N., Hartman, M.A., Dunkin, L.M., 2023. Selection of hydrological variables to forecast shoaling with Xgb. Available at SSRN 4641922.

Asborno, M., Broders, J., Mitchell, K.N., Hartman, M.A., Dunkin, L.D., 2024. Forecasting sediment accumulation in the southwest pass with machine-learning models. J. Waterw. Port Coast. Ocean. Eng. 150 (2), 04023022.

Ayres, S.K., Pinkard, F., Kleiss, B., Division, M.V., 2014. Southwest Pass Outlets Bathymetry and Flow Distribution Assessment. Mississippi River Geomorphology and Potamology Report No. 5, U.S. Army Corps of Engineers, Mississippi Valley Division.

Bai, S., 2018. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. arXiv preprint arXiv:1803.01271.

Bain, R., Hartman, M., Godfrey, A.D., Sylvester, C., Smith, K., 2025. Corps shoaling analysis tool (CSAT) user guide. https://cirp.usace.army.mil/products/files/CSAT/CSAT_UserGuide_Draft.pdf. Draft version accessed on: 2025-06-26.

Bhattacharya, B., Solomatine, D.P., 2006. Machine learning in sedimentation modelling. Neural Netw. 19 (2), 208–214.

Byrnes, M.R., Baker, J.L., Li, F., et al., 2002. Quantifying potential measurement errors and uncertainties associated with bathymetric change analysis.

Chou, J.-S., Lin, J.-W., 2020. Risk-informed prediction of dredging project duration using stochastic machine learning. Water 12 (6), 1643.

Cigizoglu, H.K., Alp, M., 2006. Generalized regression neural network in modelling river sediment yield. Adv. Eng. Softw. 37 (2), 63–68.

Dodd, N., Stoker, A., Garnier, R., Vittori, G., De Los Santos, F., Brocchini, M., Soldini, L., Losada, M., 2008. Use of numerical models to study land-based sedimentation and subsequent nearshore morphological evolution. Coast. Eng. 55 (7–8), 601–621.

Dunkin, L., Coe, L., Ratcliff, J., 2018. Corps Shoaling Analysis Tool: Predicting Channel Shoaling. ERDC/CHL Technical Report No. ERDC/CHL TR-18-16, U.S. Army Engineer Research and Development Center, Coastal and Hydraulics Laboratory, https://apps.dtic.mil/sti/trecms/pdf/AD1081215.pdf. (Accessed 26 June 2025).

Dunkin, L.M., Saltus, C.L., Mitchell, K.N., Scully, B.M., 2015. Improved channel evaluation using automatic identification system and hydrographic survey data. In: Proc., Western Dredging Association and Texas A&M University Center for Dredging Studies "Dredging Summit and Expo".

Engelund, F., Anker, H., et al., 1967. Monograph on sediment transport in alluvial streams. (No Title).

Firat, M., Güngör, M., 2010. Monthly total sediment forecasting using adaptive neuro fuzzy inference system. Stoch. Environ. Res. Risk Assess. 24, 259–270.

Firat, M., Turan, M.E., Yurdusev, M.A., 2009. Comparative analysis of fuzzy inference systems for water consumption time series prediction. J. Hydrol. 374 (3–4), 235–241.

Gal, Y., Ghahramani, Z., 2016. Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. In: International Conference on Machine Learning. PMLR, pp. 1050–1059.

Ghimire, G.R., Hansen, C., Gangrade, S., Kao, S.-C., Thornton, P.E., Singh, D., 2023. Insights from dayflow: A historical streamflow reanalysis dataset for the conterminous United States. Water Resour. Res. 59 (2), e2022WR032312.

Graw, J., Wood, W., Phrampus, B., 2021. Predicting global marine sediment density using the random forest regressor machine learning algorithm. J. Geophys. Res.: Solid Earth 126 (1), e2020JB020135.

Guven, A., Kişi, Ö., 2011. Estimation of suspended sediment yield in natural rivers using machine-coded linear genetic programming. Water Resour. Manag. 25, 691–704.

Hartman, M.A., Mitchell, K.N., Dunkin, L.M., Lewis, J., Emery, B., Lenssen, N.F., Copeland, R., 2022. Southwest pass sedimentation and dredging data analysis. J. Waterw. Port Coast. Ocean. Eng. 148 (2), 05021017.

Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9 (8), 1735–1780. http://dx.doi.org/10.1162/neco.1997.9.8.1735.

Irish, J.L., White, T.E., 1998. Coastal engineering applications of high-resolution lidar bathymetry. Coast. Eng. 35 (1–2), 47–71.

Jain, S.K., 2001. Development of integrated sediment rating curves using ANNs. J. Hydraul. Eng. 127 (1), 30–37.

Jang, J.-S., 1993. ANFIS: Adaptive-network-based fuzzy inference system. IEEE Trans. Syst. Man Cybern. 23 (3), 665–685.

Jothiprakash, V., Garg, V., 2009. Reservoir sedimentation estimation using artificial neural network. J. Hydrol. Eng. 14 (9), 1035–1040.

Kabir, M., Kang, M.J., Wu, X., Hamidi, M., 2022. Study on U-turn behavior of vessels in narrow waterways based on AIS data. Ocean Eng. 246, 110608.

Kisi, O., 2005. Suspended sediment estimation using neuro-fuzzy and neural network approaches/estimation des matières en suspension par des approches neurofloues et à base de réseau de neurones. Hydrol. Sci. J. 50 (4).

Kisi, Ö., 2008. Constructing neural network sediment estimation models using a data-driven algorithm. Math. Comput. Simulation 79 (1), 94–103.

Kisi, O., 2012. Modeling discharge-suspended sediment relationship using least square support vector machine. J. Hydrol. 456, 110–120.

Kisi, O., Shiri, J., 2012. River suspended sediment estimation by climatic variables implication: Comparative study among soft computing techniques. Comput. Geosci. 43, 73–82.

Kress, M., 2021. Hydrosurvey data archive for federal navigation projects. https://data.gov/maritime/hydrosurvey-data-archive-for-federal-navigation-projects/index.html.

Li, C., Mahadevan, S., 2016. An efficient modularized sample-based method to estimate the first-order Sobol' index. Reliab. Eng. Syst. Saf. 153, 110–121. http://dx.doi.org/10.1016/j.ress.2016.04.012, URL: https://www.sciencedirect.com/science/article/pii/S0951832016300266.

Lohani, A., Goel, N.K., Bhatia, K.S., 2007. Deriving stage–discharge–sediment concentration relationships using fuzzy logic. Hydrol. Sci. J. 52 (4), 793–807.

Mateo-Pérez, V., Corral-Bobadilla, M., Ortega-Fernández, F., Rodríguez-Montequín, V., 2021. Determination of water depth in ports using satellite data based on machine learning algorithms. Energies 14 (9), 2486.

Melesse, A., Ahmad, S., McClain, M., Wang, X., Lim, Y., 2011. Suspended sediment load prediction of river systems: An artificial neural network approach. Agricult. Water. Manag. 98 (5), 855–866.

Nagy, H., Watanabe, K., Hirano, M., 2002. Prediction of sediment load concentration in rivers using artificial neural network model. J. Hydraul. Eng. 128 (6), 588–595.

Nourani, V., Baghanam, A.H., Adamowski, J., Kisi, O., 2014. Applications of hybrid wavelet–artificial intelligence models in hydrology: A review. J. Hydrol. 514, 358–377.

Partal, T., Cigizoglu, H.K., 2008. Estimation and forecasting of daily suspended sediment data using wavelet–neural networks. J. Hydrol. 358 (3–4), 317–331.

Port Houston, 2024. The Houston ship channel – maintaining a national economic engine. https://porthouston.com/the-houston-ship-channel-maintaining-a-national-economic-engine/. (Accessed 26 June 2025).

Port of Houston, 2025. Project overview - Port Houston: Houston ship channel expansion. https://www.expandthehoustonshipchannel.com/project-overview/. (Accessed 26 June 2025).

Pourzangbar, A., Brocchini, M., 2022. A new process-based, wave-resolving, 2DH circulation model for the evolution of natural sand bars: The role of nearbed dynamics and suspended sediment transport. Coast. Eng. 177, 104192.

Rajaee, T., 2010. Wavelet and neuro-fuzzy conjunction approach for suspended sediment prediction. CLEAN–Soil Air Water 38 (3), 275–286.

Rosati, J.D., Kraus, N.C., 2009. Rapid methods for estimating navigation channel shoaling. In: WEDA XXIX Technical Conference and 40th Texas A&M Dredging Seminar, Tempe, Arizona.

Sainath, T.N., Vinyals, O., Senior, A., Sak, H., 2015. Convolutional, long short-term memory, fully connected deep neural networks. In: 2015 IEEE International Conference on Acoustics, Speech and Signal Processing. ICASSP, IEEE, pp. 4580–4584.

Scully, B.M., Young, D.L., Bain, R.L., 2025. Crowdsourced bathymetry's role in a data fusion approach to measuring demand for waterway depth. In: Ports 2025: Port Engineering. Ports 2025, ASCE, pp. 260–270.

Uphues, C.F., da Silva, G.M., van Rooijen, A., Hesp, P.A., 2025. Modelling sediment dynamics in a high energy coastal embayment. Coast. Eng. 104924.

U.S. Army Corps of Engineers, 2024. Corps shoaling analysis tool (CSAT). https://cirp.usace.army.mil/products/csat.php.

U.S. Army Corps of Engineers, Galveston District, 2025. Houston ship channel improvement project. https://www.swg.usace.army.mil/Projects/Houston-Ship-Channel-Improvement-Project/. (Accessed 26 June 2025).

U.S. Army Corps of Engineers, Mobile District, 2025. eHydro. URL: https://www.sam.usace.army.mil/Missions/Spatial-Data-Branch/eHydro/. (Accessed 26 June 2025).

U.S. Department of Commerce, N.O., (NOAA), A.A., National Ocean Service, 2025. US coast pilot 5 (2025), Gulf of Mexico, Puerto Rico, Virgin Islands. https://nauticalcharts.noaa.gov/publications/coast-pilot/files/cp5/CPB5_WEB.pdf. (Accessed 26 June 2025). NOAA Coast Survey.

USACE, 2022. National channel framework. https://geospatial-usace.opendata.arcgis.com/maps/9227967a2748410983352b501c0c7b39/about. (Accessed 19 October 2022).

Yeates, E.M., Tavakoly, A.A., Mitchell, K.N., Dreaper, G.W., Afshari, S., 2020. Utilizing stream flows to forecast dredging requirements. USACE Rep. 1 (1), 1.

Zeng, Y., Ammar, D., Chadha, M., Wang, D., Asborno, M., Miele, S., McKnight, C.J., Memarsadeghi, N.P., Hartman, M.A., Mitchell, K.N., et al., 2026. A recursive shoaling forecasting framework with shifted delays and its application to the Southwest Pass in the United States. Mach. Learn. Appl. 100968.

Zhao, Y., Chadha, M., Barthlow, D., Yeates, E., Mcknight, C.J., Memarsadeghi, N.P., Gugaratshan, G., Todd, M.D., Hu, Z., 2024. Physics-enhanced machine learning models for streamflow discharge forecasting. J. Hydroinformatics 26 (10), 2506–2537.