

Turbidity assessment in coastal regions combining machine learning, numerical modeling, and remote sensing

Saeed Memari^{a,*}, Mantha S. Phanikumar^{a,b}, Vishnu Boddeti^c and Narendra N. Das^{a,d}

^a Department of Civil & Environmental Engineering, Michigan State University, East Lansing, MI 48824, USA

^b MSU AgBioResearch, East Lansing, MI 48824, USA

^c Department of Computer Science & Engineering, Michigan State University, East Lansing, MI 48824, USA

^d Department of Biosystems & Agricultural Engineering, Michigan State University, East Lansing, MI 48824, USA

*Corresponding author. E-mail: memarisa@msu.edu; saeed.memari@mines.edu

ABSTRACT

Machine learning models for water quality prediction often face challenges due to insufficient data and uneven spatial-temporal distributions. To address these issues, we introduce a framework combining machine learning, numerical modeling, and remote sensing imagery to predict coastal water turbidity, a key water quality proxy. This approach was tested in the Great Lakes region, specifically Cleveland Harbor, Lake Erie. We trained models using observed data and synthetic data from 3D numerical models and tested them against *in situ* and remote sensing data from PlanetLabs' Dove satellites. High-resolution (HR) data improved prediction accuracy, with RMSE values of 0.154 and 0.146 log₁₀(FNU) and R^2 values of 0.92 and 0.93 for validation and test datasets, respectively. Our study highlights the importance of unified turbidity measures for data comparability. The machine learning model demonstrated skill in predicting turbidity through transfer learning, indicating applicability in diverse, data-scarce regions. This approach can enhance decision support systems for coastal environments by providing accurate, timely predictions of water quality variables. Our methodology offers robust strategies for turbidity and water quality monitoring and has potential for improving input data quality for numerical models and developing predictive models from remote sensing data.

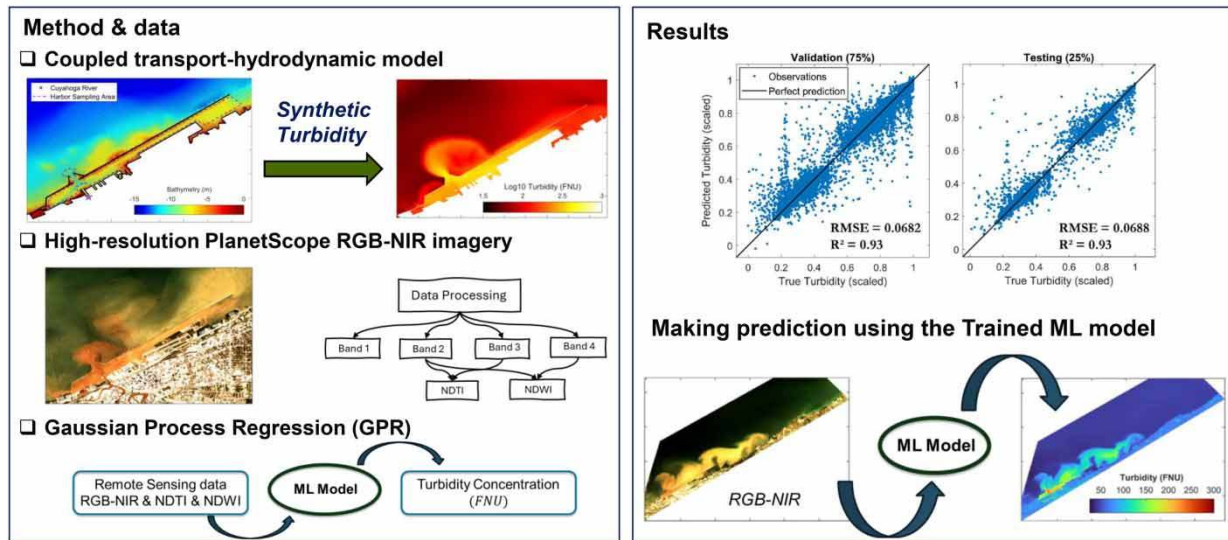
Key words: hydrodynamic modeling, machine learning, remote sensing, solute transport, transfer learning, water turbidity

HIGHLIGHTS

- Developed a framework combining machine learning, numerical modeling, and remote sensing for accurate turbidity prediction.
- Need for uniform turbidity measurement to improve monitoring and data comparability.
- Explored turbidity prediction in data-scarce areas via transfer learning.

GRAPHICAL ABSTRACT

Turbidity Assessment in Coastal Regions Combining Machine Learning, Numerical Modeling, and Remote Sensing



Conclusion: Our integrated approach significantly advances understanding and prediction of water turbidity. Leveraging numerical (synthetic) data for training machine learning models demonstrates significant potential, improving data diversity and availability, essential for boosting prediction accuracy and model applicability across different water bodies.

1. INTRODUCTION

Understanding the clarity and quality of water through turbidity, a key parameter influenced by particulate matter, is crucial for the health of aquatic ecosystems (Boyd & Tucker 2012; Water Quality & Health 2017). Turbidity reflects not only the physical conditions of water bodies but also hints at the presence of pollutants, making it an essential indicator for environmental monitoring. While measurements in nephelometric turbidity units (NTU), formazin nephelometric units (FNU), and milligrams per liter (mg/l) are common, the relationship between turbidity and suspended solids is complex, varying with particle characteristics (Bilotta & Brazier 2008; U.S. Geological Survey 2022). This underscores the need for reliable approaches to accurately gauge water quality and address the environmental challenges posed by elevated levels of turbidity.

Monitoring and predicting turbidity in coastal areas present a formidable challenge, necessitating the convergence of different methods, from field measurements to numerical modeling and remote sensing. Mechanistic models, particularly those based on three-dimensional conservation principles, have been used extensively to simulate the fate and transport of contaminants in water bodies (Pelletier *et al.* 2006; Nguyen *et al.* 2014, 2017; Safaie *et al.* 2020; Feizabadi *et al.* 2022; Memari & Phanikumar 2024a, 2024b). These models can simulate complex phenomena such as turbidity currents with reversing buoyancy (Sequeiros *et al.* 2009). However, their accuracy pivots on the realistic representation of forcing fields and boundary conditions, such as river discharge and turbidity (Jalón-Rojas *et al.* 2021; Zhu *et al.* 2022; Feizabadi *et al.* 2023). This suggests that while these models have furnished pivotal insights, their reliability and applicability necessitate further enhancements in the context of dynamic environments such as coastal waters.

In parallel with numerical modeling, remote sensing, particularly the use of satellite data, has emerged as a potent tool for assessing water quality parameters (Nelson *et al.* 2002; Olmanson *et al.* 2008; Topp *et al.* 2020; Li *et al.* 2022). This technology offers a synoptic view of the water bodies, allowing the consistent tracking of parameters, including turbidity. It exploits remote sensing indices such as the normalized difference turbidity index (NDTI) and normalized difference water index (NDWI) to infer turbidity levels from space (Gao 1996; Lacaux *et al.* 2007). However, despite its promise, remote sensing confronts several limitations, such as the inability to penetrate deep water bodies and the influence of atmospheric conditions on the received signal (Cunningham *et al.* 2013), necessitating supplementary approaches for water quality assessment.

In light of this, the fusion of machine learning with remote sensing and field measurement has gained momentum in recent years (Filisbino Freire da Silva *et al.* 2021). Machine learning models, equipped with the ability to discern complex patterns in data, have been utilized to predict outcomes related to water quality, including turbidity (Normandin *et al.* 2019; Li *et al.* 2021). For example, machine learning models such as random forests and decision trees have been effective in predicting turbidity and water quality from environmental factors (Anmala & Turuganti 2021; Venkateswarlu & Anmala 2023). These models, however, are data-sensitive and demand high-quality labeled data for training, which may not always be available. A significant limitation is the lack of spatial and temporal resolution in field measurements, which is critical in dynamic coastal environments. Similarly, the generalizability of these models hinges on the training data that encompasses a wide range of concentration values, which field sampling often lacks, especially for high turbidity samples.

The advent of transfer learning, a technique that allows the application of knowledge from related source domains to improve the performance of models in a target domain has shown promising results in overcoming these challenges (Farahani *et al.* 2021; Zhuang *et al.* 2021; Hee *et al.* 2022). Transfer learning techniques, coupled with machine learning algorithms, have shown potential in applications such as remote sensing for water quality assessment (Pu *et al.* 2019; Zhu *et al.* 2019; Gambin *et al.* 2021; Syariz *et al.* 2022; Arias-Rodriguez *et al.* 2023).

Despite these advancements, considerable gaps remain, particularly concerning data availability in a wide range of concentration values for training machine learning models and their generalizability. This calls for an integrated approach that leverages the strengths of numerical modeling, machine learning, and remote sensing. Our study addresses these gaps by utilizing synthetic data from well-tested numerical models to train machine learning models for turbidity prediction using remote sensing imagery in the visible bands. We also investigate the applicability of the trained models through transfer learning in different domains. This approach will allow us to predict turbidity in a unified concentration unit (FNU) for the study site and potentially for other coastal areas. The study contributes novel insights into the precision and limitations of numerical models, as turbidity can serve as a natural tracer. This holistic approach has the potential to transform our understanding of water quality in dynamic coastal environments.

2. MATERIALS AND METHODS

2.1. Study area

The focus of this study is on Lake Erie around the Port of Cleveland situated at the mouth of the Cuyahoga River in Cleveland, Ohio, United States (Figure 1). The Cuyahoga River has a significant historical background of contamination, characterized by numerous major fires resulting from the accumulation of industrial waste and oil slicks. One notable incident occurred in 1969, which acted as a catalyst for the environmental movement and subsequently led to the enactment of the Clean Water Act in 1972 (Adler 2002; Stradling & Stradling 2008).

The Port of Cleveland, located along the banks of the Cuyahoga River, has played a vital role in the economic development of the city. However, the port has also contributed to the pollution of the river through the discharge of wastewater and other industrial effluents. Recent endeavors addressed these concerns and diminished the adverse impacts on the river and its surrounding ecosystems. The port has implemented several initiatives aimed at enhancing water quality and safeguarding the health of the river and the Great Lakes ecosystem (Cleveland-Cuyahoga County Port Authority 2024).

Multiple factors contribute to the pollution and turbidity observed in the Cuyahoga River. Notably, the discharge of untreated or partially treated wastewater from municipal and industrial sources is a significant source of contamination. This encompasses sewage, stormwater runoff, and various types of water contaminated with chemicals, metals, and other pollutants. Another contributing factor is the release of untreated or partially treated industrial effluent from factories and other industrial facilities, which contains chemicals, metals, and other pollutants utilized or produced during the manufacturing process (American Rivers 2024). In addition to these pollution sources, stormwater runoff can transport sediment and debris into the river, exacerbating its turbidity. Moreover, the river is affected by nonpoint source pollution, including agricultural and urban runoff.

2.2. Mechanistic modeling details and data

2.2.1. Circulation and transport models

This study utilized the finite volume community ocean model (FVCOM) to conduct simulations of lake wide and nearshore circulation. FVCOM is a three-dimensional numerical model based on triangular unstructured grids, capable of prognostic

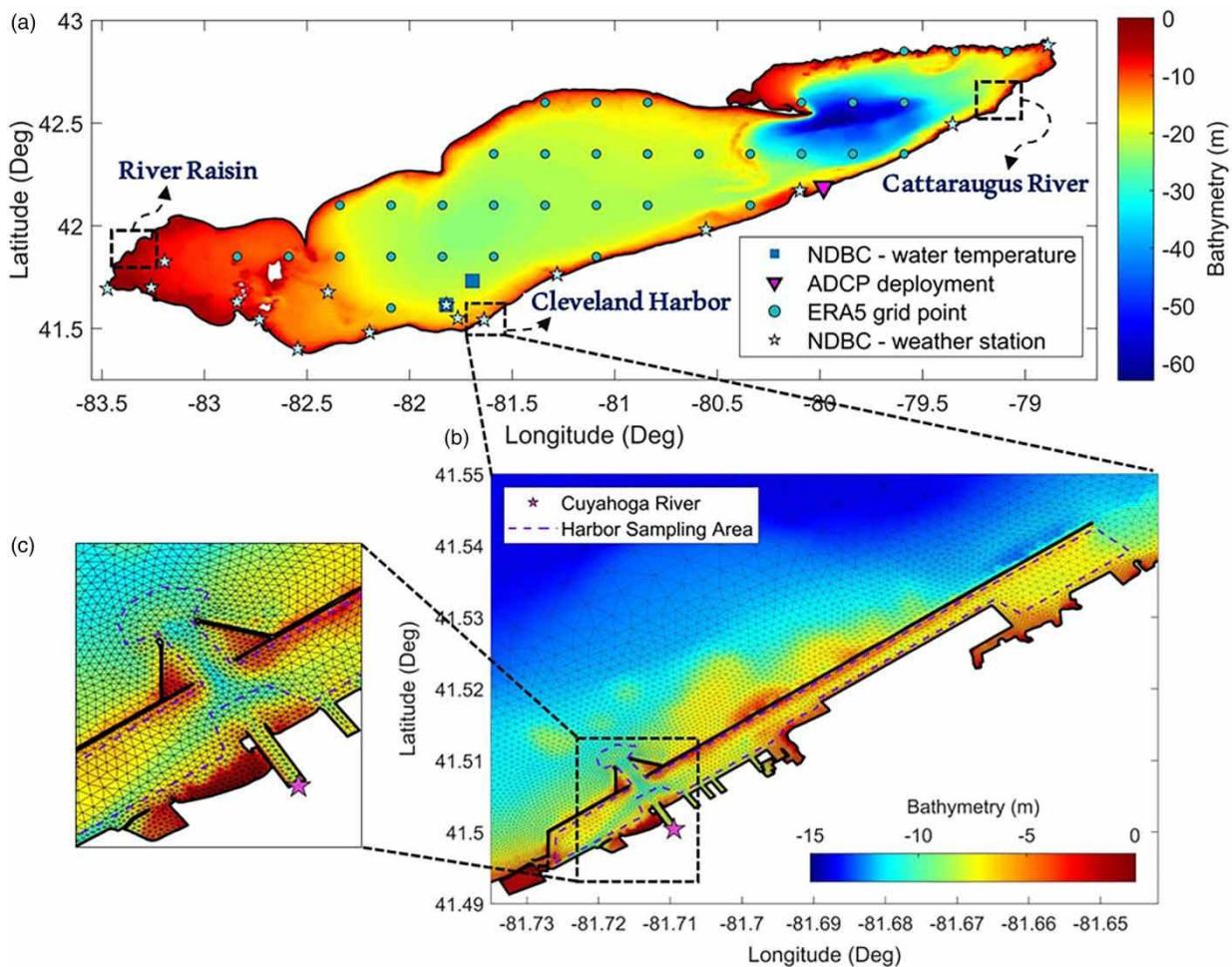


Figure 1 | (a) Bathymetric map of Lake Erie highlighting significant features, including NDBC meteorological and water temperature stations, ERA5 grid points, and the deployment site of the ADCP. (b) Bathymetric map of Cleveland Harbor displaying the numerical mesh, the Cuyahoga River mouth, and the sampling area within the harbor. (c) Close-up view of the river mouth and the primary opening of the harbor.

and free surface simulations of water circulation and transport in coastal and oceanic environments (Chen *et al.* 2003). It is widely employed for investigating hydrodynamic processes such as ocean currents and waves. FVCOM offers versatility in applications, including forecasting water levels and tides, modeling the movement of pollutants such as oil spills, and assessing the impacts of climate change on the ocean (Yang & Khangaonkar 2007; Chen *et al.* 2008; Lai *et al.* 2010; Ma *et al.* 2011; Memari & Siadatmousavi 2018). The model incorporates hydrostatic and Boussinesq approximations to solve the governing hydrodynamic equations. Vertical eddy viscosity and diffusivity are described using the Mellor-Yamada level 2.5 turbulence closure scheme (Mellor & Yamada 1982; Galperin *et al.* 1988), while the Smagorinsky turbulence closure model is employed for horizontal diffusion to determine coefficients for horizontal momentum and thermal diffusion (Smagorinsky 1963). These approaches allow for the dynamic calculation of horizontal and vertical mixing coefficients based on local conditions. This dynamic adjustment ensures that the mixing processes are accurately represented spatially and temporally within the model. Detailed equations for FVCOM can be found in Chen *et al.* (2003).

The upwind scheme was applied as it is particularly suitable for advection-dominant flows, such as those occurring within the harbor influenced by river flow. Moreover, careful attention was given to accurately representing the computational domain to minimize errors associated with steep gradients and numerical diffusion.

The hydrodynamic model simulation was conducted from 1 April 2019 to 30 July with a time step of 0.2 s. The simulation was initialized as a cold start with zero velocity. The initial temperature values were assigned as 0.1 °C at the surface and

linearly increased to 2.0 °C at the bottom of the lake. The first 2 months of the simulation (April and May) were designated as a spin-up period.

The coupled transport model describing the turbidity transport in space and time can be described by the following advection - dispersion equation:

In Equation (1), C represents turbidity in FNU, while u , v , and w denote the velocity components in the x , y , and z directions, respectively, measured in meters per second (m/s). K_H and K_V refer to the horizontal and vertical mixing coefficients (m^2/s), respectively. It should be emphasized that in our analysis, we made the assumption of no additional input or removal of turbidity once it enters the lake via the river. Consequently, any increase in turbidity (e.g., due to sediment resuspension) or decrease (e.g., due to settling of particles) was considered to have negligible impact on the overall turbidity levels. This assumption holds particularly true near the river mouth and within the harbor area, where the proximity to the river mouth and the presence of river-induced currents plays a significant role.

$$\frac{\partial C}{\partial t} + u \frac{\partial C}{\partial x} + v \frac{\partial C}{\partial y} + w \frac{\partial C}{\partial z} = \frac{\partial}{\partial x} \left(K_H \frac{\partial C}{\partial x} \right) + \frac{\partial}{\partial y} \left(K_H \frac{\partial C}{\partial y} \right) + \frac{\partial}{\partial z} \left(K_V \frac{\partial C}{\partial z} \right) \quad (1)$$

The coupled transport model run was initialized on 30 May 2019, and continued until 30 July 2019, with a time step of 0.2 s. The initial turbidity was set uniformly to 8 FNU across all computational nodes. Output from the model was saved at hourly intervals throughout the simulation period.

2.2.2. Domain discretization and bathymetry data

The computational domain encompassing Lake Erie and the Cleveland Harbor Area was discretized into 56,556 triangular elements and 30,071 nodes in the horizontal direction. The resolution of the triangular mesh in the horizontal plane varied based on the distance from the shoreline and bathymetry contours. In the harbor area and at the river mouth, the mesh resolution ranged from 10 to 40 m (Figure 1(b) and 1(c)), gradually increasing to 1,800 m in the central and deeper regions of the lake. The utilization of an unstructured triangular mesh facilitated precise representation of abrupt changes in the coastline shape, particularly within the harbor area. In the vertical direction, the computational domain was divided into 20 uniformly spaced sigma layers (21 levels). The resolution varied from a few centimeters near the shoreline to several meters in the off-shore and deep sections of the lake.

For most parts of the computational domain, 3-arc-second bathymetry data with a resolution of approximately 90 m, obtained from NOAA (National Geophysical Data Center 1999), were employed. While this resolution is considered accurate for the majority of lake and ocean models, it lacks the required precision within the harbor, near breakwaters, and at the river mouth where there are sharp bathymetry gradients and HR mesh. To address this issue, two additional sets of bathymetry data were utilized. The first was 10-m bathymetry data obtained from the United States Army Corps of Engineers, USACE, (*United States Army Corps of Engineers. Hydrographic Surveys 2023*), which covered most of the harbor and river mouth areas. The second consisted of electronic navigational charts (ENC) data from the NOAA's Office of Coast Survey (www.nautical-charts.noaa.gov) (*NOAA Electronic Navigational Charts (ENC) | InPort, n.d.*), which accounted for less than 1% of the total bathymetry data and specifically captured the bathymetry near the breakwater walls. This level of detail was necessary to ensure an accurate representation of the computational domain, given our focus on the circulation within and in proximity to the harbor area and the transport between the lake and the river. Additionally, the fine mesh resolution of 10–40 m applied inside and near the harbor further justified the inclusion of such detailed bathymetry. All bathymetry data were interpolated to the computational mesh using the natural neighbor method. This method, also known as Sibson interpolation, generates smooth surfaces by calculating a weighted average of neighboring data points based on the proportion of the overlapping Voronoi cells. It is particularly effective for scattered data and less sensitive to data point distribution, ensuring high accuracy in representing the bathymetric surface (Sibson 1981; Edelsbrunner & Shah 1992; Hiyoshi & Sugihara 2004).

2.2.3. Meteorological forcing data

Meteorological data from the National Data Buoy Center (NDBC, NOAA National Data Buoy Center 1971) and the ERA5 reanalysis dataset from the European Center for Medium-Range Weather Forecast (ECMWF, Hersbach *et al.* 2020) were combined to generate the forcing field for the mechanistic model (FVCOM model). To integrate the two meteorological datasets, hourly wind and air temperature data from the ERA5 grid points situated at least 15 km

away from any NDBC stations were extracted and merged with the NDBC data. The observed wind speeds from the NDBC stations were adjusted to a height of 10 m as described by Schwab & Morton (1984), who used a formulation that incorporates parameters such as the drag coefficient, stability length, and roughness length. The height-adjusted NDBC wind speeds were then combined with the ERA5 10-m wind speeds. On the other hand, no height correction was applied to the NDBC air temperature, as air temperature variations over short heights are negligible. The combined wind speeds and air temperatures were subsequently interpolated to the numerical mesh using the natural neighbor method.

For other forcing fields, including solar radiation (shortwave and longwave), cloud cover, and relative humidity, interpolation was performed solely from the ERA5 grid points over the lake, without integration with field observations. The decision to use an integrated approach combining both ERA5 and NDBC data was driven by the limited availability and spatial inconsistency of field observations over the lake, particularly on the northern side (Canadian side) of Lake Erie. By integrating the field observations with the ERA5 reanalysis dataset, we aimed to enhance the spatiotemporal resolution of the data while utilizing valuable field observations. This integrated method is expected to provide a more accurate representation of the forcing fields for the mechanistic model, compared to relying solely on either ERA5 or NDBC data.

2.2.4. River forcing data

Hourly observations of river discharge and turbidity were utilized to generate input (boundary condition) for the transport model. The data were obtained from the USGS stream gauge (#04208000, U. S. Geological Survey 2016) at the Cuyahoga River at Independence, OH, which is located 20.1 km away from the mouth of the river. However, the turbidity time series had continuous gaps due to missing data. To address this, interpolation methods were employed to predict the missing values. Classic interpolation methods proved ineffective due to the non-linear nature of turbidity values and continuous missing data. Therefore, Gaussian Process Regression (GPR), a machine learning technique (Wang & Jing 2022), was employed to predict the missing values and to generate hourly turbidity values as input for the mechanistic model.

GPR is a Bayesian nonparametric model suitable for interpolation and prediction (Rasmussen & Williams 2005; Wang & Jing 2022). It assumes that the underlying function describing the data follows a Gaussian process, which consists of random variables indexed by inputs with a joint Gaussian distribution. GPR provides a flexible and powerful approach to modeling complex datasets by defining a prior over functions and updating this prior with observed data to obtain a posterior distribution. This allows GPR to provide not only predictions but also uncertainty estimates, and this is particularly useful in environmental modeling.

One of the key advantages of GPR is its ability to handle noise in the data by explicitly incorporating it into the likelihood function, leading to more robust and reliable predictions (Rasmussen & Williams 2005). The kernel function, a crucial component of GPR, determines the covariance structure of the Gaussian process and can be tailored to capture the specific characteristics of the data, such as periodicity or smoothness. Commonly used kernels include the radial basis function (RBF) kernel and the Matérn kernel, each offering different properties suited to various types of data (Duvenaud 2014). By leveraging GPR, we effectively addressed the challenges posed by the non-linear and missing turbidity data, ensuring accurate and reliable input for our mechanistic model.

For the ML model setup, we used the ERA5 runoff and total precipitation data over the Cuyahoga watershed and combined with observed discharge and turbidity from the USGS gauge. This decision was made to increase the dimensionality of the data and enhance the predictive power of the ML model by using the variables that affect the turbidity values. As the river and turbidity data had already undergone quality checks by the USGS, no further cleaning or filtering was performed. Due to the significant difference in the order of magnitude within the turbidity range (1–1,000 FNU), we convert the linear scale of turbidity into a logarithmic scale (\log_{10}) for training our ML model. By making this conversion, we aimed to maintain consistency and improve the effectiveness of the ML model. The available data were divided into 80% for training the ML model and 20% for testing. To prevent overfitting and ensure robust validation of the ML model, we employed cross-validation techniques. The training data was divided into eight groups, allowing us to calculate validation scores for each group independently. This approach helped assess the performance and generalizability of the ML model across different subsets of the data. Considering the different units of the variables, data scaling was implemented prior to training the ML model. The trained ML model was then used to predict turbidity values in the missing data. The resulting turbidity time series, along with the observed river flux, served as the open boundary condition for the numerical transport model, capturing the behavior

of the Cuyahoga River. The flowchart for generating the hourly turbidity data and the coupled hydrodynamic-transport model is shown in Figure 2(a).

2.3. Remote sensing

In this study, remote sensing data from Planet Labs' PlanetScope Dove satellites (Planet Labs PBC 2019) were utilized to train a machine learning model based on the synthetic data from the mechanistic model. PlanetScope satellite constellation operates a fleet of small remote sensing satellites that capture HR images of the Earth's surface on a daily basis. These images, obtained through multispectral and hyperspectral sensors, offer a frequent revisit time (1–2 days) and high spatial resolution (~3 m), making them suitable for our research focused on a small area with dynamic circulation and transport patterns. It should be noted that the spectral resolution of Dove satellite data is comparatively lower than other remote sensing datasets like Landsat 8, Landsat 9, Sentinel 2, and Sentinel 3.

The specific data employed in our research were the surface reflectance (SR) product, which provides radiometrically calibrated and atmospherically corrected images. This product is available for orthorectified scenes captured by the sun-synchronous orbit Dove satellites. To achieve atmospheric correction, lookup tables generated using the 6SV2.1 radiative transfer code were employed, enabling the mapping of top-of-atmosphere (TOA) reflectance to bottom-of-atmosphere (BOA) reflectance. The SR product is provided as a 16-bit GeoTIFF image, with reflectance values scaled by a factor of 10,000. For atmospheric inputs, water vapor and ozone information were retrieved from MODIS (Moderate Resolution Imaging Spectroradiometer, <https://modis.gsfc.nasa.gov/>) near-real-time (NRT) data, while aerosol optical depth (AOD) input was determined from MODIS NRT aerosol data. By considering localized atmospheric conditions, the SR product ensures consistency and minimizes uncertainty in spectral response across various time points and locations. Further details can be found in Table 1.

Due to the disparity in spatial resolution between the mechanistic model and remote sensing data, preprocessing and data cleaning are necessary prior to their utilization. Firstly, we extract the SR data for each band within each triangular mesh. Subsequently, the z -score of each band's data is calculated individually, and data points exceeding an absolute z -score threshold of 3 are considered outliers and excluded (Hoaglin 2013). This approach ensures consistency among the data points. The accepted values within each triangular mesh are then averaged for each band, resulting in a single SR value

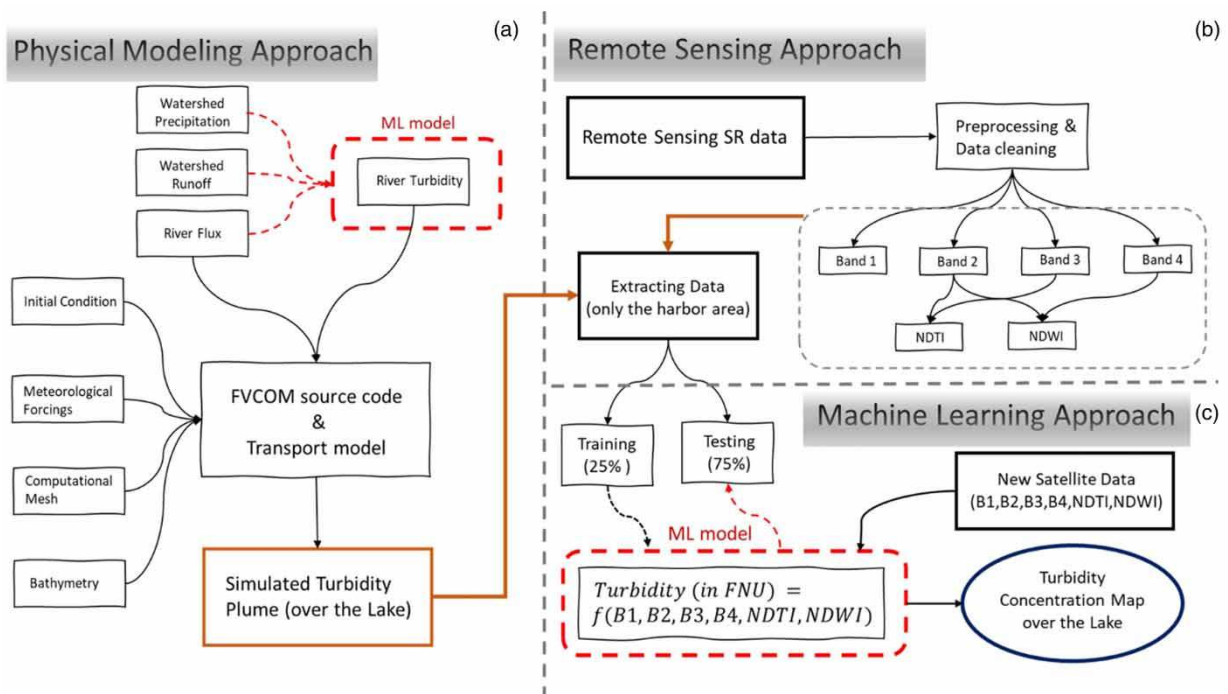


Figure 2 | The flowchart of (a) physical (mechanistic) modeling, (b) remote sensing, and (c) machine learning used in this research.

Table 1 | Details of the remote sensing images used in this research, including instrument type, spectral bands, pixel size, and atmospheric corrections

Instrument	Spectral Bands (nm)	Pixel Size (m)	Atmospheric Corrections
PS2	Blue: 455–515	3.0	Conversion to top-of-atmosphere (TOA) reflectance values using at-sensor radiance and supplied coefficients.
	Green: 500–590		
	Red: 590–670		Conversion to surface reflectance values using the 6SV2.1 radiative transfer code and MODIS NRT data.
	NIR: 780–860		

per triangular mesh. The flowchart for extracting and processing the remote sensing SR data is shown in Figure 2(b). Similarly, the NDTI and NDWI are computed using the equations: (Gao 1996; Lacaux *et al.* 2007).

$$NDTI = \frac{SR_{Red} - SR_{Green}}{SR_{Red} + SR_{Green}} \quad (2)$$

$$NDWI = \frac{SR_{Green} - SR_{NIR}}{SR_{Green} + SR_{NIR}} \quad (3)$$

2.3.1. Data preparation for machine learning

To develop the machine learning model to extract turbidity from satellite imagery, we first extracted the simulated turbidity for the uppermost 3 m of the water column from the transport model. This process was conducted concurrently with the time of image acquisition to ensure temporal alignment. These simulated turbidity values were then combined with the processed remote sensing data – specifically, blue, green, red, and near infrared bands, NDTI, and the NDWI (NDWI) – as depicted in Figure 3.

To ensure data quality for training the machine learning model, the scope was strictly limited to the data derived from the harbor sampling area (Figure 1(b)). This strategic selection was crucial in guaranteeing that the detected turbidity predominantly originated from the Cuyahoga River, thereby minimizing the influence of lake-wide turbidity and turbidity from other rivers (both upstream and downstream). This decision also capitalized on the resolution of the mechanistic and transport model, which, considering the data-intensive nature commonly associated with ML models, supplied ample training data samples.

The machine learning model was then developed using processed simulated turbidity and remote sensing data, specifically focusing on the harbor sampling area. Of the five images used for training the model, 75% of the data was used for training and validation, and the remaining 25% was dedicated to testing the ML model. This systematic allocation ensured a robust development and comprehensive evaluation of the ML model's performance. The trained ML model can be utilized to predict turbidity (in FNU) based on the remote sensing SR at four bands, as well as the NDTI and NDWI indexes. The flowchart for developing the ML model, based on the simulated turbidity and the remote sensing data, is shown in Figure 2(c).

2.3.2. Transfer learning: detecting turbidity in other regions of the lake

Transfer learning, briefly defined as the application of a model trained in one domain to a new, but related, domain, was employed in our study to enhance the adaptability of machine learning models across diverse regions (Zhuang *et al.* 2021; Syarif *et al.* 2022). We utilized the trained ML model, which was developed for predicting turbidity based on remote sensing images in the Cleveland Harbor area, to predict turbidity in other regions of Lake Erie, namely Cattaraugus River and River Raisin (Figure 1). This allowed us to assess the generalizability of the trained ML model across different regions of the lake.

2.4. Validation process

The performance of the mechanistic model in accurately describing water velocity was assessed by comparing model results for water current velocity with observed data obtained from a Nortek Aquadopp HR (2.0 MHz frequency) Profiler ADCP (Acoustic Doppler Current Profiler) deployed in Lake Erie at Erie, PA (coordinates: 42.1886 N and –79.9821 W), from 9 August to 10 September 2019 (Memari & Phanikumar 2024a). The instrument was deployed at a depth of 5.80 m

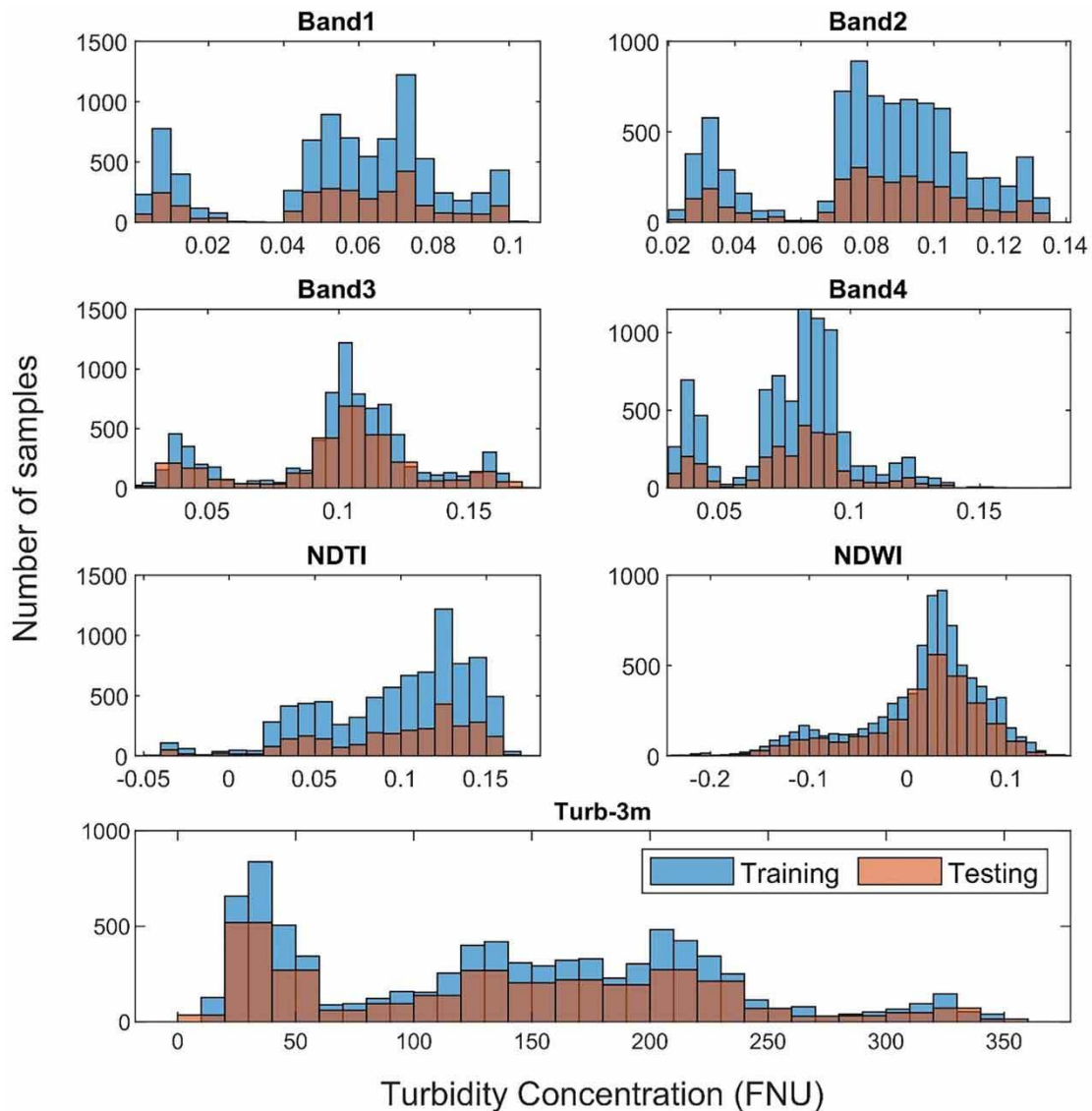


Figure 3 | Processed surface reflectance data at each band (band1, band2, band3, and band4), NDTI and NDWI indexes, and turbidity data used for training and testing the ML model.

from the surface and the measurement cell size was set to 0.25 m. Likewise, the accuracy of the mechanistic model in representing surface water temperature was evaluated using data from two NDBC buoys (#45164 and #45169), as illustrated in Figure 1.

2.4.1. Error metrics

Two metrics, RMSE (root mean square error) and R^2 (correlation coefficient), were used to evaluate the performance of the mechanistic model and machine learning models in comparison to the observed data. RMSE measures the average magnitude of differences between predicted (P_i) and observed (O_i) values, indicating the overall accuracy of the models. A lower RMSE signifies better agreement between predictions and observations. R^2 assesses the proportion of variance in the observed data explained by the models. Higher R^2 values indicate stronger correlations and capture a greater portion of the observed data variability. By considering both RMSE and R^2 , the models' accuracy, predictive capability, and ability to capture data patterns were comprehensively evaluated.

In Equations (4) and (5), P_i represents the predicted values, O_i represents the observed values, \bar{P} is the mean of the predicted values, and \bar{O} is the mean of the observed values, and n denotes the number of values (size of the vector).

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (P_i - O_i)^2}{n}} \quad (4)$$

$$R^2 = \frac{\sum_{i=1}^n (O_i - \bar{O})(P_i - \bar{P})}{\sqrt{\sum_{i=1}^n (O_i - \bar{O})^2} \sqrt{\sum_{i=1}^n (P_i - \bar{P})^2}} \quad (5)$$

3. RESULTS

3.1. Application of ML to river turbidity prediction

Figure 4(a) and 4(b) display scatter plots comparing the observed and ML-predicted turbidity for the training and test datasets. The validation scores yielded an RMSE of 0.154 log₁₀(FNU) and an R^2 value of 0.92, while the test data showed an RMSE of

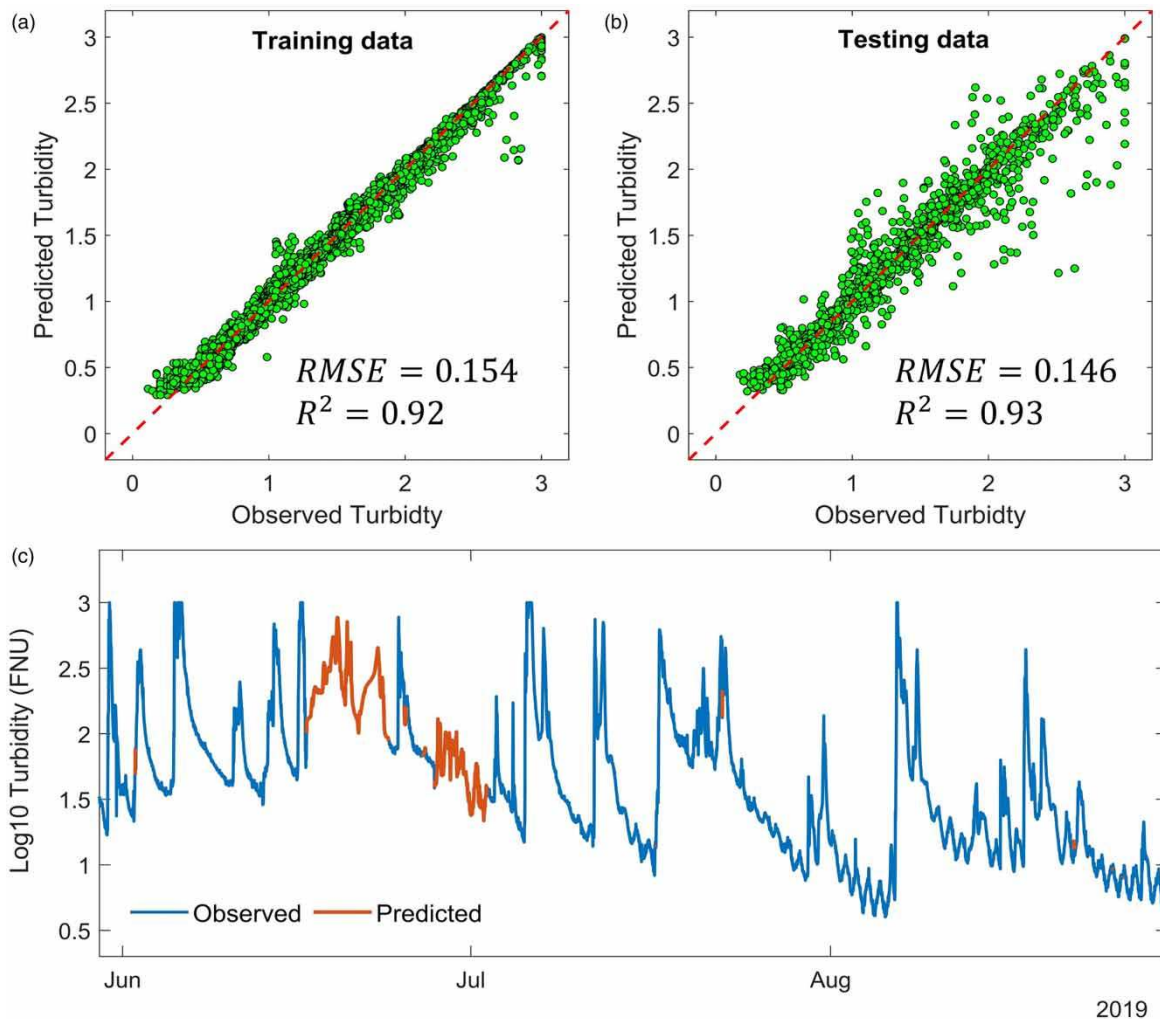


Figure 4 | Scatter plots of (a) training and (b) testing data with their performance metrics. (c) Log₁₀ turbidity values at the river mouth. The red line denotes the ML-predicted values and the blue line denotes the observed data.

0.146 $\log_{10}(\text{FNU})$ and an R^2 value of 0.93. In Figure 4(c), the ML-predicted turbidity time series (red line) exhibits an oscillating pattern similar to the observed data. Although there are no observations for the missing data period (red line in Figure 4(c)), we will compare the turbidity map near the harbor with a remote sensing image for this specific timeframe later. This comparison will help assess the extent to which the turbidity plume affected the harbor area.

3.2. Validation of the circulation model

The mechanistic model was tested using observed vertically averaged eastward (u) and northward (v) components of velocity at the ADCP deployment location. Figure 5(a) and 5(b) present the comparison between the simulated u -component and v -component of velocity against the observed data. Overall, the mechanistic model performed well in simulating the velocity components over time. Table 2 includes the RMSE and R^2 values as evaluation metrics. Although the RMSE for the v -component of velocity is smaller than that of the u -component, the R^2 value for the u -component is higher than that of the v -component. This difference can be attributed to the higher magnitude of the u -component compared to the v -component of velocity which resulted in a higher RMSE value for the u -component of velocity. Similarly, the smaller magnitudes of the v -component velocity values result in a reduced RMSE compared to the u -component.

Additionally, we assessed the mechanistic model's ability to accurately simulate the water temperature by comparing the simulated water surface temperature with observations from two NDBC stations (#45164 and #45169) close to the harbor

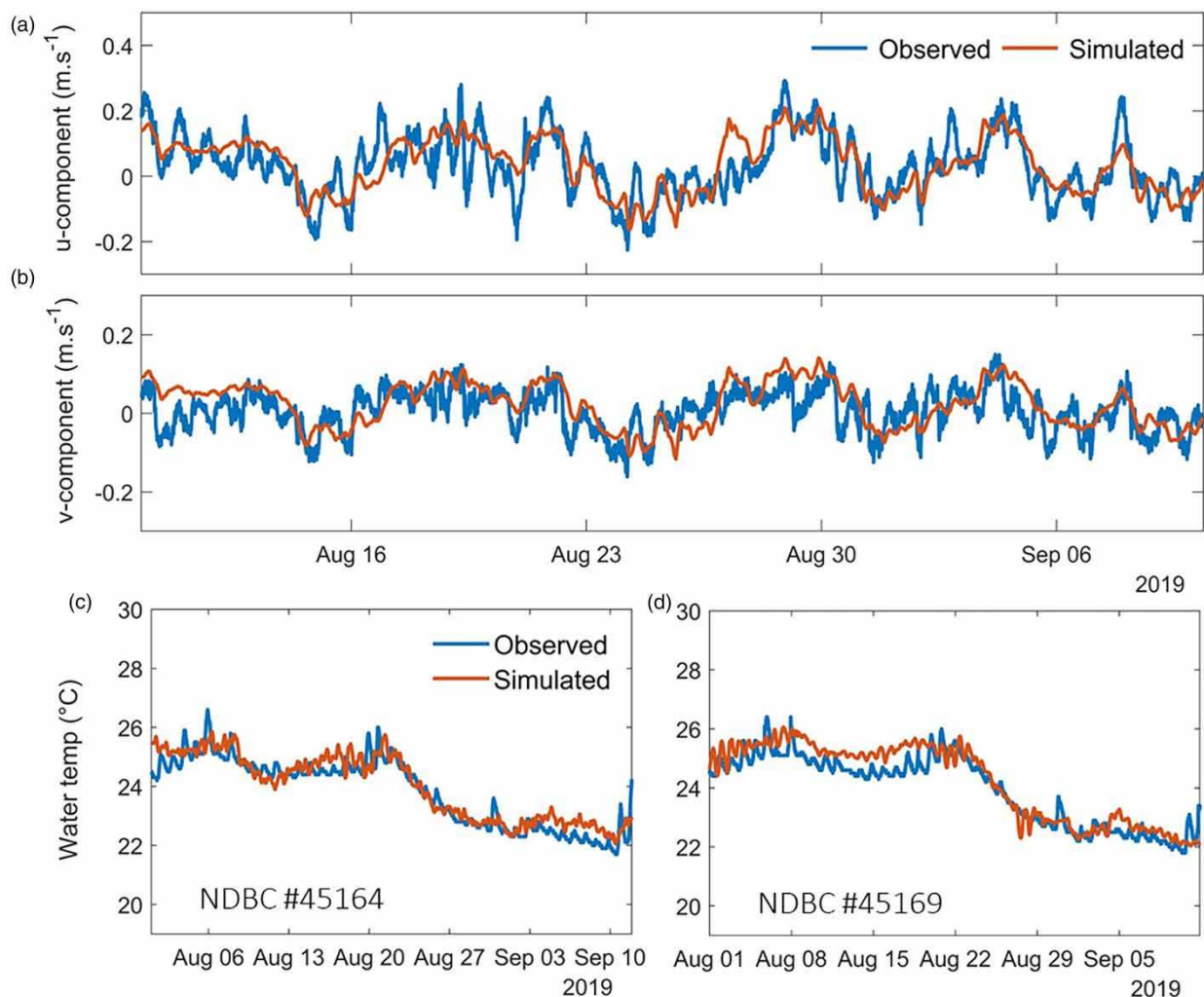


Figure 5 | Comparison between the vertically averaged simulated and observed (a) eastward and (b) northward components of velocity at the location of the ADCP deployment. Comparison between the simulated and observed (c, d) water surface temperature at the NDBC buoys.

Table 2 | Error metric scores between the simulated and observed water velocity and water surface temperature

Variable	RMSE	R^2
u (ms^{-1})	0.064	0.716
v (ms^{-1})	0.047	0.650
Water Temp ($^{\circ}\text{C}$), NDBC #45164	0.437	0.938
Water Temp ($^{\circ}\text{C}$), NDBC #45169	0.482	0.955

area. [Figure 5\(c\)](#) and [5\(d\)](#) demonstrate that the mechanistic model successfully simulated water surface temperature values and captured the overall trend. [Table 2](#) provides the RMSE and R^2 values for the comparison between the simulated and observed water surface temperatures. The RMSE value was found to be smaller than 0.5°C , and R^2 was greater than 0.93.

3.3. Validation of the transport model

During the simulation period, no *in situ* turbidity data were available for the harbor area. Therefore, a quantitative comparison between the simulated turbidity plume and the true turbidity values cannot be made. However, we can still assess the performance of the mechanistic model in simulating turbidity transport by comparing the spatial map of turbidity with remote sensing RGB images at different times. [Figure 6](#) shows the remote sensing RGB images at three different times along with their corresponding simulated turbidity plume.

In [Figure 6](#), we observe similarities between the extent and intensity of the turbidity plume in the simulated results and the remote sensing RGB images, particularly at a smaller scale within the harbor. While the similarities are notable within the harbor area, differences become more pronounced as we move away from the point source at the mouth of the river. Although the simulated turbidity closely resembles the observed data inside the harbor, variations become more significant outside the harbor area as the distance from the point source increases.

In [Figure 7](#), both the remote sensing RGB image and the simulated turbidity inside the harbor and its vicinity are depicted. The figure highlights the similarities observed between the simulated turbidity plume and the RGB image in areas outside the harbor. However, the accuracy or similarity gradually degrades as we move farther away from the harbor area. This discrepancy can be attributed to several factors, including the utilization of a larger mesh size outside the harbor, potential numerical diffusion errors associated with the mechanistic model, and the absence of other sources of turbidity, such as the Rocky River outlet depicted in the left corner of the images in [Figure 7](#).

3.4. Combined performance of circulation and transport models

The accuracy and reliability of turbidity predictions in this study are significantly influenced by the combined performance of the circulation and transport models. The circulation model provides essential velocity fields and boundary conditions that drive the transport model, directly influencing turbidity transport and dispersion. The validation of the circulation model using ADCP data (see [Figure 5\(a\)](#) and [5\(b\)](#)) confirmed its accuracy in simulating water currents, which is vital for reliable turbidity predictions. The transport model, utilizing these accurate velocity fields, effectively simulates the spatial and temporal distribution of turbidity. This combined approach was validated by comparing simulated turbidity plumes with remote sensing images, demonstrating strong agreement, especially within the harbor (see [Figures 6](#) and [7](#)).

3.5. Extracting turbidity maps from remote sensing images

In [Figure 8](#), a scatter plot is presented, illustrating the scaled simulated turbidity plotted against the ML-predicted scaled turbidity. For the validation dataset, the RMSE and R^2 values were 0.0682 and 0.93, respectively. Similarly, the testing scores yielded RMSE and R^2 values of 0.0688 and 0.93, respectively.

[Figure 9](#) shows the ML-predicted turbidity (in FNU) for two remote sensing images captured on 22 June 2019 and 23 June 2019. As depicted in [Figure 9](#), the turbidity appears to be higher within the harbor area on 22 June 2019. However, as we move farther away from the harbor and shift to the image captured on 23 June 2019, the turbidity decreases as we progress in both space and time.

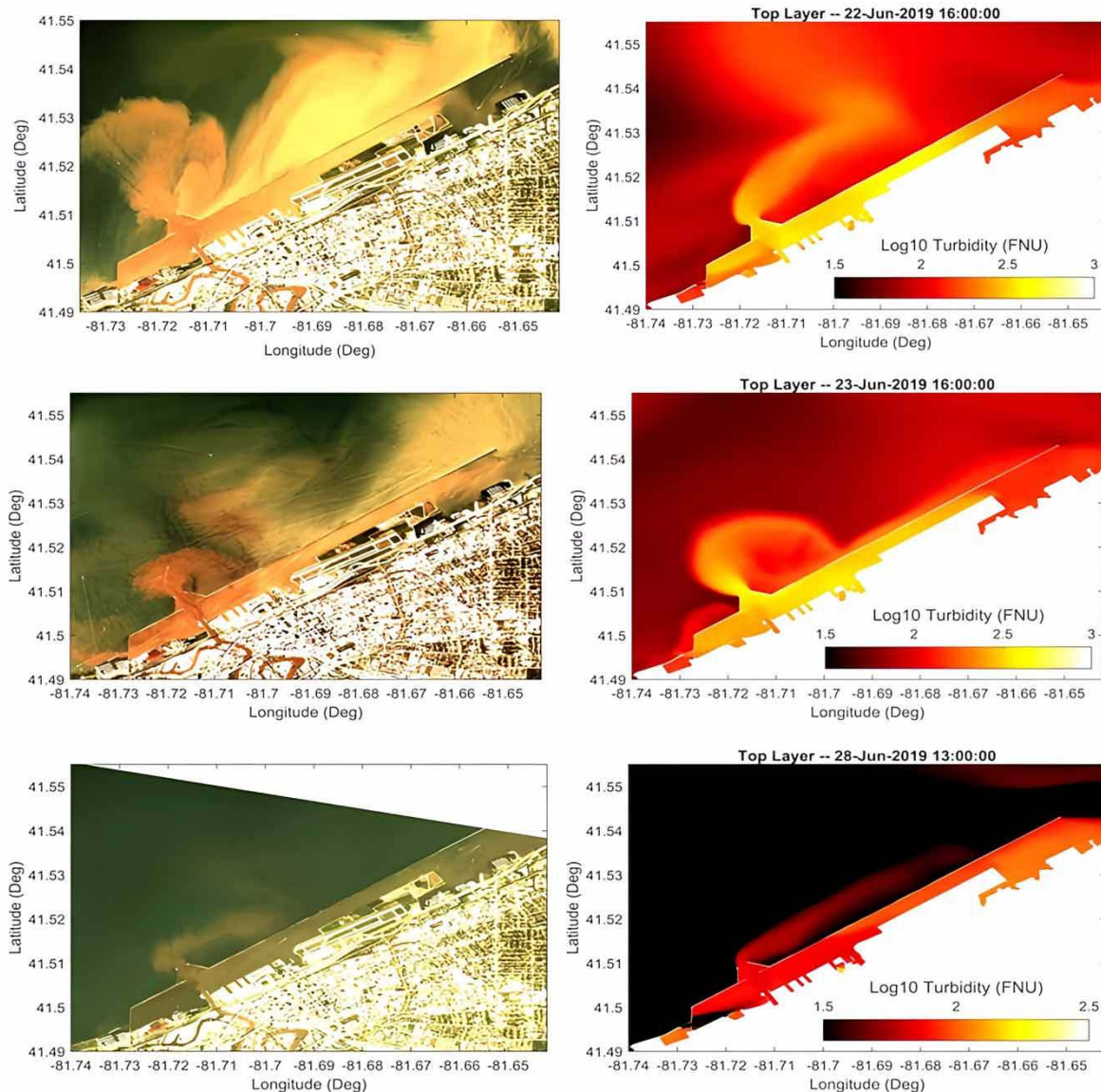


Figure 6 | Remote sensing RGB images (left side) and their corresponding simulated turbidity plume (right side) for images taken on 22, 23, and 25 June 2019.

3.6. Transfer learning: detecting turbidity in other regions of the lake

The results depicted in Figure 10, shown for the Cattaraugus and Raisin River plumes, present the RGB images in comparison with their corresponding turbidity plumes, as predicted by the ML model developed for the Cleveland Harbor area. There appears to be a robust correlation between high turbidity values and the corresponding RGB images for the Cattaraugus River and River Raisin.

Overall, the ML model demonstrated considerable skill in predicting turbidity through transfer learning, particularly when juxtaposed with the RGB images. However, it is worth noting that the model fell short in predicting accurate concentration levels as shown in Figure 10 (indicated by a red box). The absence of fine-tuning for these new sites likely contributed to this shortcoming.

The discrepancy might also be linked to the limitations of the training data. The model was originally trained on a limited number of images from a different region of the lake, which could have influenced the results. Additionally, variations in the concentration of dissolved and suspended compounds within the water might have affected the SR at the measured bands in these regions, thereby potentially contributing to the observed discrepancies.

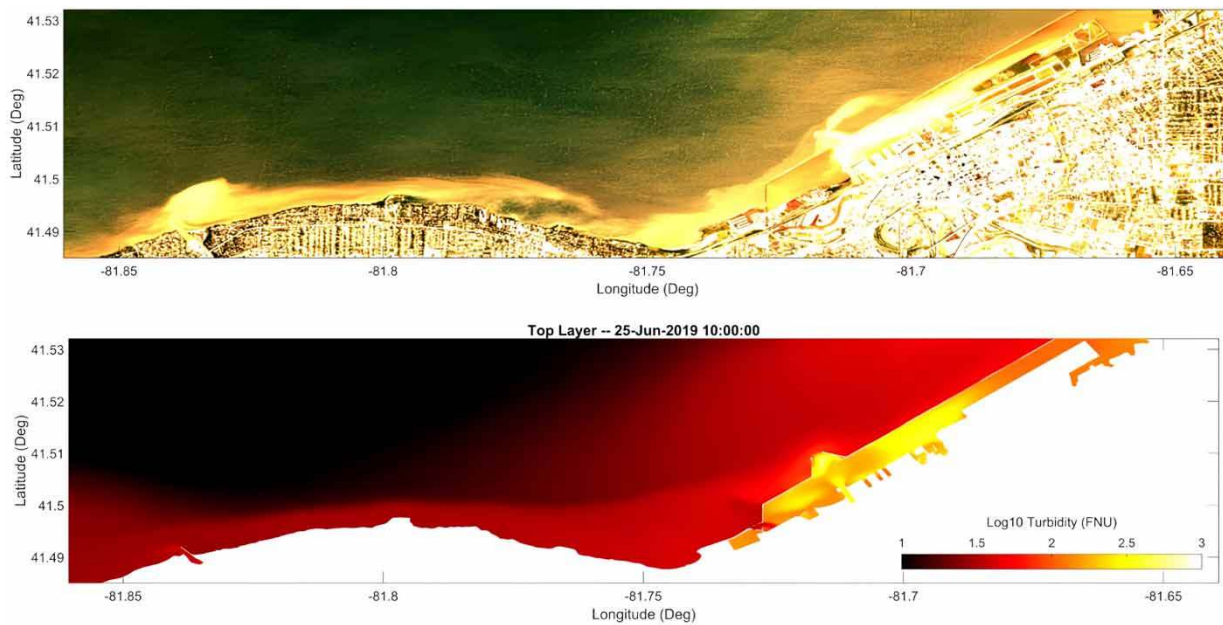


Figure 7 | (Top) RGB image of the remote sensing data and its corresponding (bottom) simulated turbidity map at 25 June 2019.

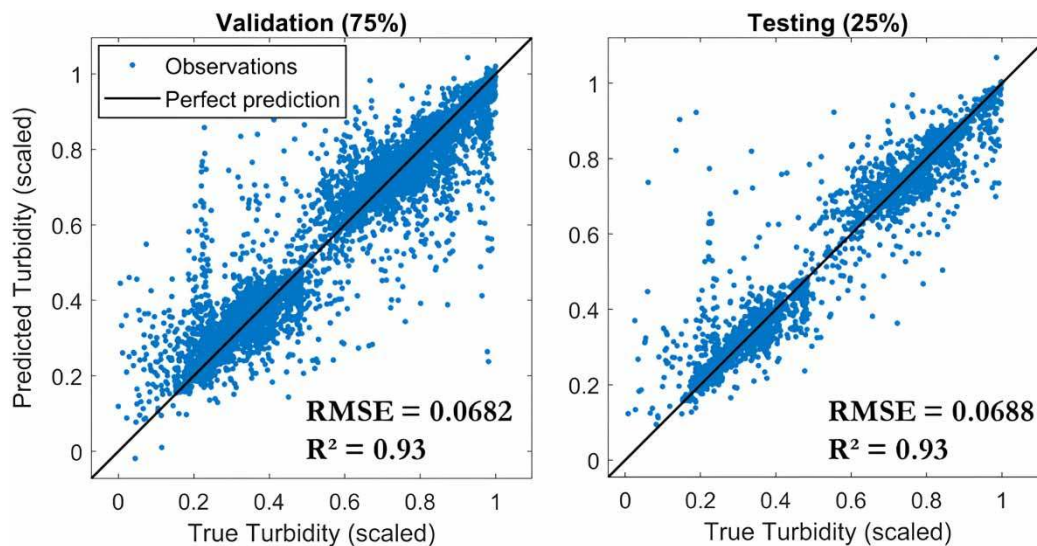


Figure 8 | Scatter plots of validation and testing data with their error metric scores.

4. DISCUSSION

The analysis of our results indicates that the most accurate results are obtained in the harbor area due to detailed domain discretization and precise boundary conditions (Sequeiros *et al.* 2009). This detailed discretization provides a clearer understanding of turbidity patterns, similar to the insights provided by HR Landsat 8 imagery in the Po River prodelta (Braga *et al.* 2017). Our results align with the findings of Braga *et al.* (2017), underscoring the importance of HR data for accurate prediction and analysis of turbidity. It should be noted that this approach can be implemented for the areas without coastal structures as long as the circulation and transport are resolved accurately.

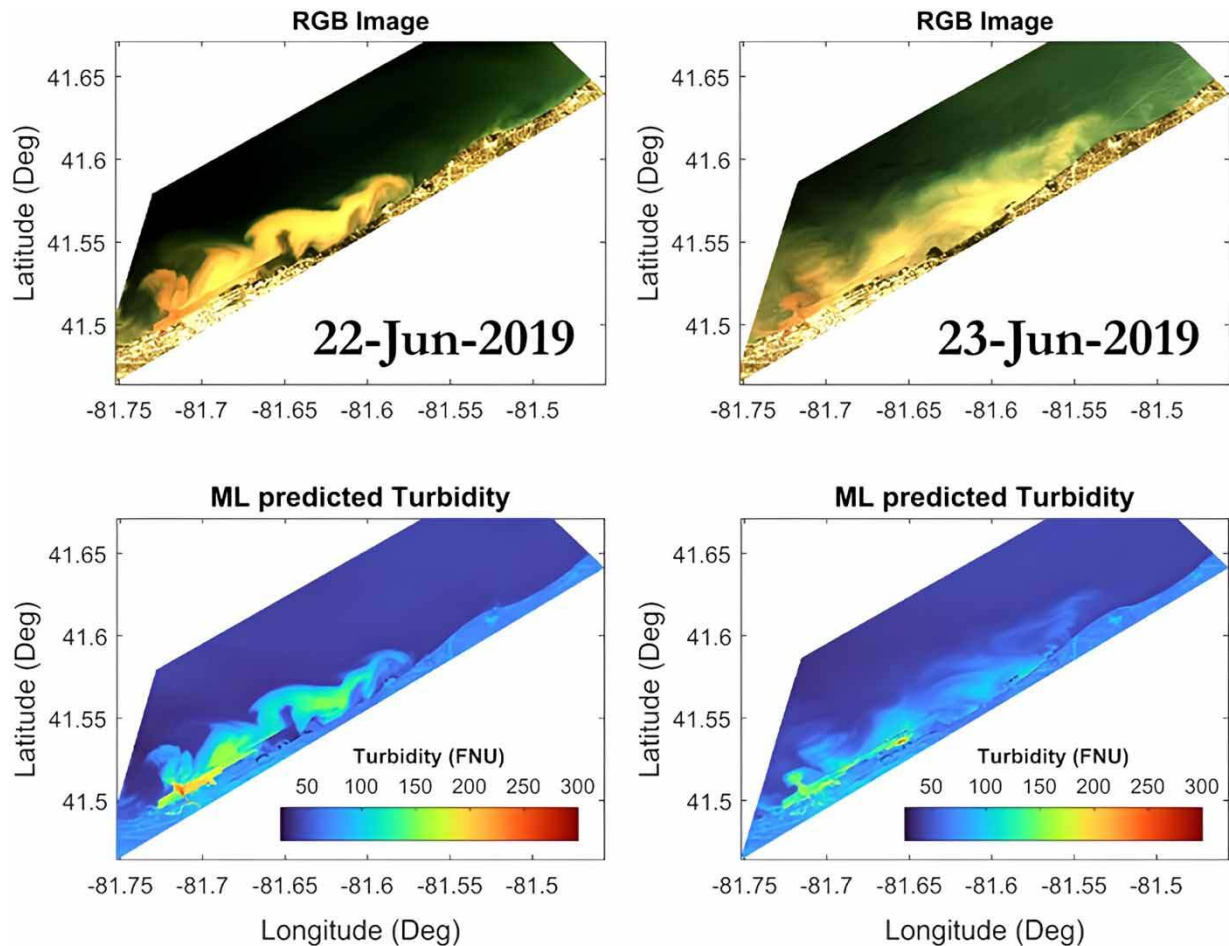


Figure 9 | (Top) Remote sensing RGB images and their (bottom) ML-predicted turbidity on 22 and 23 June 2019.

Significantly, our utilization of a mechanistic model, calibrated with ADCP and temperature data, allowed resolution of water dynamics, capturing the spatiotemporal variability of turbidity currents with high fidelity. This approach distinguishes our work from previous studies, as the ADCP data-driven calibration ensures a realistic representation of the underlying physical processes, enhancing the precision and accuracy of our predictions, while providing abundant simulated (synthetic) data for the machine learning model. By excluding this 3D mechanistic modeling component, we would sacrifice the spatiotemporal resolution, compromising the robustness and reliability of our assessments, and rendering our model susceptible to oversimplifications and poor performance.

Notably, our study also highlights the need for a unified measure of turbidity across the lake, which aligns with the method used by [Zhu *et al.* \(2022\)](#) in their study of the Great Lakes. By using the same unit as the USGS gauges, we can compare turbidity values at the river mouth with those in the lake, enhancing the comparability of data.

In contrast, the accuracy of results decreases outside the harbor area, mainly because of a larger mesh size which translates to larger numerical diffusion errors and the absence of other sources of turbidity in the model such as wastewater treatment plants, as well as mechanisms such as deposition and resuspension ([Felix 2002](#); [Schulz *et al.* 2018](#)). This mirrors the conclusions drawn by [Schulz *et al.* \(2018\)](#), where varying hydro- and meteorological conditions affected sediment fluxes in different locations.

As we moved farther from the river mouth and harbor area, we encountered issues related to mesh sensitivity and numerical diffusion. This suggests that an accurate representation of the domain and forcing fields, such as the wind field ([Beletsky *et al.* 2013](#)), and river boundary conditions is required to address the uncertainty of boundary conditions ([Hunt & Jones 2020](#)).

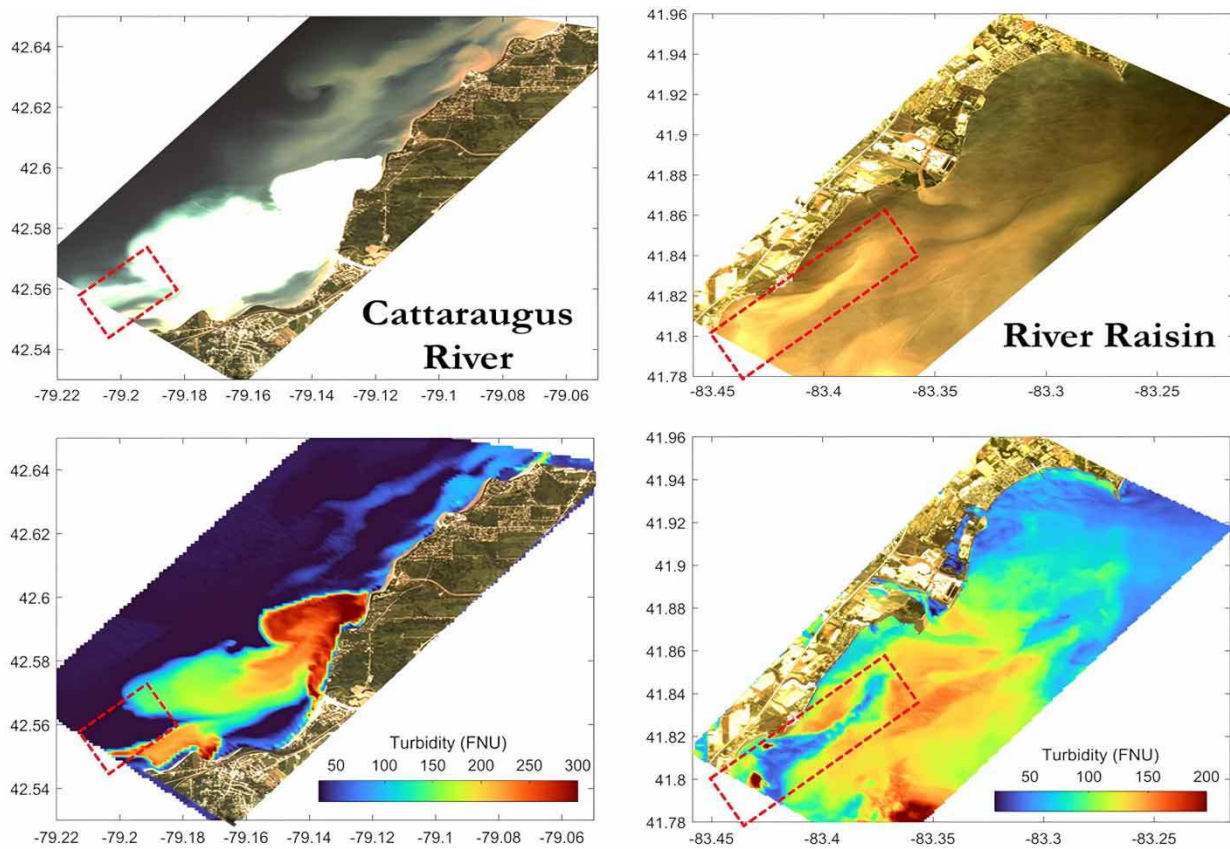


Figure 10 | (Top) Remote sensing RGB images for Cattaraugus River and River Raisin and their corresponding (bottom) ML-predicted turbidity plume.

Our study also recognized that turbidity is influenced by a variety of sources along the river, from the USGS gauge to the mouth of the river at the harbor. However, deriving turbidity from one remote sensing index poses challenges, such as a single NDTI value can correspond to multiple turbidity values (Garg *et al.* 2017, 2020). This supports the findings of (Zheng & DiGiacomo 2022), who utilized a simplified water clarity–turbidity index (CTI) to better capture major changes in water clarity/turbidity by including multiple variables from Visible Infrared Imaging Radiometer Suite (VIIRS) measurements, Secchi disk depth, and particulate backscattering coefficient.

The type of remote sensing instrument used has a significant impact on the results. If a model is trained on imagery from a specific instrument, it may produce inaccurate predictions when applied to images captured by other sensors such as Sentinel or Landsat (Le Fouest *et al.* 2015; Vanhellemont & Ruddick 2021). Yet, our approach proved effective when training machine learning models on different instruments, supporting the claims of (Saberioon *et al.* 2020; Filisbino Freire da Silva *et al.* 2021) regarding the potential of machine learning and satellite data in water quality prediction.

Notably, our study also highlights the need for a unified measure of turbidity across the lake and at the river mouth through *in situ* instruments, which aligns with the method used by Garg *et al.* (2017). By using the same unit as the USGS gauges, we can compare turbidity values at the river mouth with those in the lake, enhancing the comparability of data.

Finally, our study proposes using a GPR model to address uncertainty effectively, supporting the conclusions of Filisbino Freire da Silva *et al.* (2021) about the potential of machine learning in water quality assessment. This strategy, coupled with generating ample data for training machine learning models, can significantly improve our understanding of turbidity detection and monitoring, effectively addressing issues of water composition variability noted in our study and the research conducted by Normandin *et al.* (2019). This approach also aids in evaluating and refining the performance of mechanistic models in predicting turbidity accurately, building upon the work of Felix (2002) and Sequeiros *et al.* (2009).

Overall, our approach offers a promising direction for the development of robust, data-rich, and effective strategies for turbidity detection and monitoring. The synergy between *in situ* measurements, remote sensing imagery, and machine learning

algorithms can help in developing a more comprehensive understanding of turbidity dynamics in various marine and estuarine environments.

5. CONCLUSION

Our research aimed to address data limitations and enhance model generalizability for predicting water turbidity from remote sensing imagery by integrating mechanistic modeling, machine learning, and remote sensing. Our findings confirmed the efficacy of this integrated approach, showing significant improvements in both the precision and accuracy of turbidity predictions across a broad range of turbidity values.

We used a machine learning model to enhance hourly interpolation of river turbidity values during data gaps when there were no measurements of turbidity at the Cuyahoga River mouth. This was demonstrated to be effective in describing boundary conditions of the mechanistic model and improving its performance. Validation results showed strong correlations between observed and predicted turbidity, confirming the reliability of our models. Specifically, the machine learning model for river turbidity was validated not only by using performance metrics such as RMSE and R^2 but also by comparing the turbidity plume generated by the mechanistic model with remote sensing images at multiple times.

We bridged the gap in data availability for training machine learning models for predicting lake turbidity values from remote sensing images by using synthetic (simulated) turbidity data from the mechanistic model. This approach improved data availability across a wide range of concentration values. The accurate and abundant simulated data generated by the mechanistic model proved invaluable for training the machine learning model. This integration underscores the necessity for unified turbidity measures and highlights the impact of remote sensing data quality on prediction accuracy. This machine learning model not only performed well in predicting turbidity from the remote sensing images in the study site (Cleveland Harbor) but also demonstrated acceptable capability in predicting turbidity at other coastal areas, showing model transferability. However, fine-tuning of the machine learning model for new locations will further improve the predictions and generalizability of the model.

A number of decision support systems for coastal environments, including those for beach closures and harmful algal bloom (HAB) severity prediction, involve turbidity as a key variable. Our approach can be extended to enhance such systems by providing accurate and timely predictions of turbidity and other water quality variables of interest (e.g., bacteria, viruses, nutrients, HABs, etc.).

ACKNOWLEDGEMENTS

We thank Drs. Mary Anne Evans and Muruleedhara Byappanahalli, USGS and Mr Glen Black (USGS dive safety instructor) for their assistance with field data collection. The FVCOM model is available from the University of Massachusetts (<http://fvcom.smast.umassd.edu/>). ADCP field data used in this research are available on HydroShare (Memari & Phanikumar 2024a). The remote sensing data used in this research are available from Planet website (<https://www.planet.com/>). Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

DATA AVAILABILITY STATEMENT

All relevant data are available from the online repository HydroShare at the following link: <https://doi.org/10.4211/hs.5ee190b481c749fb8398f182742720f1>.

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

- Adler, J. (2002) *Fables of the Cuyahoga: Reconstructing A History of Environmental Protection*. Faculty Publications. Available from: https://scholarlycommons.law.case.edu/faculty_publications/191.
- American Rivers 2024 Available from: <https://www.americanrivers.org/media-item/american-rivers-names-clevelands-cuyahoga-river-of-the-year/> (accessed 20 February 2024).
- Anmala, J. & Turuganti, V. (2021) Comparison of the performance of decision tree (DT) algorithms and extreme learning machine (ELM) model in the prediction of water quality of the Upper Green River watershed. *Water Environment Research* **93** (11), 2360–2373. <https://doi.org/10.1002/wer.1642>.

- Arias-Rodriguez, L. F., Tüzün, U. F., Duan, Z., Huang, J., Tuo, Y. & Disse, M. (2023) Global water quality of inland waters with harmonized landsat-8 and sentinel-2 using cloud-computed machine learning. *Remote Sensing* **15** (5), Article 5. <https://doi.org/10.3390/rs15051390>.
- Beletsky, D., Hawley, N. & Rao, Y. R. (2013) Modeling summer circulation and thermal structure of Lake Erie. *Journal of Geophysical Research: Oceans* **118** (11), 6238–6252. <https://doi.org/10.1002/2013JC008854>.
- Bilotta, G. S. & Brazier, R. E. (2008) Understanding the influence of suspended solids on water quality and aquatic biota. *Water Research* **42** (12), 2849–2861. <https://doi.org/10.1016/j.watres.2008.03.018>.
- Boyd, C. E. & Tucker, C. S. (2012) *Pond Aquaculture Water Quality Management*. Springer Science & Business Media. Available from: https://books.google.com/books?hl=en&lr=&id=2r8GCAAQBAJ&oi=fnd&pg=PR11&dq=Pond+Aquaculture+Water+Quality+Management.+Springer+Science+%26+Business+Media.&ots=pKx64wXaGr&sig=S_OCUsGx1G8TQUeczgHoSkOUgI.
- Braga, F., Zaggia, L., Bellafore, D., Bresciani, M., Giardino, C., Lorenzetti, G., Maicu, F., Manzo, C., Riminucci, F., Ravaioli, M. & Brando, V. E. (2017) Mapping turbidity patterns in the Po river prodelta using multi-temporal Landsat 8 imagery. *Estuarine, Coastal and Shelf Science* **198**, 555–567. <https://doi.org/10.1016/j.ecss.2016.11.003>.
- Chen, C., Liu, H. & Beardsley, R. (2003) An unstructured grid, finite-volume, three-dimensional, primitive equations ocean model: Application to coastal ocean and estuaries. *Journal of Atmospheric and Oceanic Technology* **20**, 159–186. [https://doi.org/10.1175/1520-0426\(2003\)020<0159:AUGFVT>2.0.CO;2](https://doi.org/10.1175/1520-0426(2003)020<0159:AUGFVT>2.0.CO;2).
- Chen, C., Qi, J., Li, C., Beardsley, R., Lin, H., Walker, R. & Gates, K. (2008) Complexity of the flooding/drying process in an estuarine tidal-creek salt-marsh system: An application of FVCOM. *Journal of Geophysical Research* **113**. <https://doi.org/10.1029/2007JC004328>.
- Cleveland-Cuyahoga County Port Authority 2024 Port of Cleveland – The Premier Port on the Great Lakes. Available from: <https://www.portofcleveland.com/> (accessed 20 February 2024).
- Cunningham, A., Ramage, L. & McKee, D. (2013) Relationships between inherent optical properties and the depth of penetration of solar radiation in optically complex coastal waters. *Journal of Geophysical Research: Oceans* **118** (5), 2310–2317. <https://doi.org/10.1002/jgrc.20182>.
- Duvenaud, D. (2014) *Automatic Model Construction with Gaussian Processes*. <https://doi.org/10.17863/CAM.14087>.
- Edelsbrunner, H. & Shah, N. R. (1992) Incremental topological flipping works for regular triangulations. In *Proceedings of the Eighth Annual Symposium on Computational Geometry – SCG '92*, pp. 43–52. <https://doi.org/10.1145/142675.142688>.
- Farahani, A., Pourshojae, B., Rasheed, K. & Arabnia, H. R. (2021) *A Concise Review of Transfer Learning* (arXiv:2104.02144). arXiv. <https://doi.org/10.48550/arXiv.2104.02144>.
- Feizabadi, S., Rafati, Y., Ghodsian, M., Akbar Salehi Neyshabouri, A., Abdolahpour, M. & Mazyak, A. R. (2022) Potential sea-level rise effects on the hydrodynamics and transport processes in Hudson–Raritan Estuary, NY–NJ. *Ocean Dynamics* **72** (6), 421–442. <https://doi.org/10.1007/s10236-022-01512-0>.
- Feizabadi, S., Li, C. & Hiatt, M. (2023) A numerical experiment of cold front induced circulation in Wax Lake Delta: Evaluation of forcing factors. *Frontiers in Marine Science* **10**, 1228446.
- Felix, M. (2002) Flow structure of turbidity currents. *Sedimentology* **49** (3), 397–419. <https://doi.org/10.1046/j.1365-3091.2002.00449.x>.
- Filisbino Freire da Silva, E., Márcia Leão de Moraes Novo, E., de Lucia Lobo, F., Clemente Faria Barbosa, C., Tressmann Cairo, C., Almeida Noernberg, M. & Henrique da Silva Rotta, L. (2021) A machine learning approach for monitoring Brazilian optical water types using Sentinel-2 MSI. *Remote Sensing Applications: Society and Environment* **23**, 100577. <https://doi.org/10.1016/j.rsase.2021.100577>.
- Galperin, B., Kantha, L., Hassid, S. & Rosati, A. (1988) A quasi-equilibrium turbulent energy model for geophysical flows. *Journal of the Atmospheric Sciences* **45**, 55–62. [https://doi.org/10.1175/1520-0469\(1988\)045<0055:AQETEM>2.0.CO;2](https://doi.org/10.1175/1520-0469(1988)045<0055:AQETEM>2.0.CO;2).
- Gambin, A. F., Angelats, E., Gonzalez, J. S., Miozzo, M. & Dini, P. (2021) Sustainable marine ecosystems: Deep learning for water quality assessment and forecasting. *IEEE Access* **9**, 121344–121365. <https://doi.org/10.1109/ACCESS.2021.3109216>.
- Gao, B. (1996) NDWI – a normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment* **58** (3), 257–266. [https://doi.org/10.1016/S0034-4257\(96\)00067-3](https://doi.org/10.1016/S0034-4257(96)00067-3).
- Garg, V., Senthil Kumar, A., Aggarwal, S. P., Kumar, V., Dhote, P. R., Thakur, P. K., Nikam, B. R., Sambare, R. S., Siddiqui, A., Muduli, P. R. & Rastogi, G. (2017) Spectral similarity approach for mapping turbidity of an inland waterbody. *Journal of Hydrology* **550**, 527–537. <https://doi.org/10.1016/j.jhydrol.2017.05.039>.
- Garg, V., Aggarwal, S. P. & Chauhan, P. (2020) Changes in turbidity along Ganga River using Sentinel-2 satellite data during lockdown associated with COVID-19. *Geomatics, Natural Hazards and Risk* **11** (1), 1175–1195. <https://doi.org/10.1080/19475705.2020.1782482>.
- Hee, K., Cosa, A., Santhanam, N., Jannesari, M., Maros, M. & Ganslandt, T. (2022) Transfer learning for medical image classification: A literature review. *BMC Medical Imaging* **22**. <https://doi.org/10.1186/s12880-022-00793-7>.
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S. & Thépaut, J.-N. (2020) The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society* **146** (730), 1999–2049. <https://doi.org/10.1002/qj.3803>.
- Hiyoshi, H., Sugihara, K., (2004) Improving the global continuity of the natural neighbor interpolation. In: *Computational Science and Its Applications – ICCSA 2004* (Laganá, A., Gavrilova, M. L., Kumar, V., Mun, Y., Tan, C. J. K. & Gervasi, O., eds). Springer, pp. 71–80. https://doi.org/10.1007/978-3-540-24767-8_8.

- Hoaglin, D. (2013) *Volume 16: How to Detect and Handle Outliers*. Available from: <https://www.semanticscholar.org/paper/Volume-16%3A-How-to-Detect-and-Handle-Outliers-Hoaglin/d524a172b49e25f888376d662ee364aa77d99e8a>.
- Hunt, S. & Jones, H. F. E. (2020) The fate of river-borne contaminants in the marine environment: Characterising Regions of Freshwater Influence (ROFIs) and estuary plumes using idealised models and satellite images. *Marine Pollution Bulletin* **156**, 111169. <https://doi.org/10.1016/j.marpolbul.2020.111169>.
- Jalón-Rojas, I., Dijkstra, Y., Schuttelaars, H. M., Brouwer, R., Schmidt, S. & Sottolichio, A. (2021) Multi-decadal evolution of the turbidity maximum zone in a macrotidal river under climate and anthropogenic pressures. *Journal of Geophysical Research: Oceans* **126**. <https://doi.org/10.1029/2020JC016273>.
- Lacaux, J. P., Tourre, Y. M., Vignolles, C., Ndione, J. A. & Lafaye, M. (2007) Classification of ponds from high-spatial resolution remote sensing: Application to rift valley fever epidemics in Senegal. *Remote Sensing of Environment* **106** (1), 66–74. <https://doi.org/10.1016/j.rse.2006.07.012>.
- Lai, Z., Chen, C., Cowles, G. & Beardsley, R. (2010) A nonhydrostatic version of FVCOM: 1. Validation experiments. *Journal of Geophysical Research* **115**. <https://doi.org/10.1029/2009JC005525>.
- Le Fouest, V., Chami, M. & Verney, R. (2015) Analysis of riverine suspended particulate matter fluxes (Gulf of Lion, Mediterranean Sea) using a synergy of ocean color observations with a 3-D hydrodynamic sediment transport model. *Journal of Geophysical Research: Oceans* **120** (2), 942–957. <https://doi.org/10.1002/2014JC010098>.
- Li, J., Tian, L., Wang, Y., Jin, S., Li, T. & Hou, X. (2021) Optimal sampling strategy of water quality monitoring at high dynamic lakes: A remote sensing and spatial simulated annealing integrated approach. *Science of the Total Environment* **777**, 146113. <https://doi.org/10.1016/j.scitotenv.2021.146113>.
- Li, H., Yang, Q., Mo, S., Huang, J., Wang, S., Xie, R., Luo, X. & Liu, F. (2022) Formation of turbidity maximum in the Modaomen estuary of the Pearl River, China: The roles of mouth bar. *Journal of Geophysical Research: Oceans* **127** (12), e2022JC018766. <https://doi.org/10.1029/2022JC018766>.
- Ma, G., Shi, F., Liu, S. & Qi, D. (2011) Hydrodynamic modeling of Changjiang Estuary: Model skill assessment and large-scale structure impacts. *Applied Ocean Research* **33**, 69–78. <https://doi.org/10.1016/J.APOR.2010.10.004>.
- Mellor, G. L. & Yamada, T. (1982) Development of a turbulence closure model for geophysical fluid problems. *Reviews of Geophysics* **20** (4), 851–875. <https://doi.org/10.1029/RG020i004p00851>.
- Memari, S. & Phanikumar, M. S. (2024a) Acoustic Doppler Current Profiler (ADCP) Data Collected Near Erie, PA, Lake Erie, 2019, *HydroShare*. Available from: <https://hydroshare.org/resource/5ee190b481c749fb8398f182742720f1/>.
- Memari, S. & Phanikumar, M. S. (2024b) Assessing transport timescales in Lake Huron's Hammond Bay: The crucial role of the Straits of Mackinac's exchange flows. *Science of the Total Environment* **912**, 168777. <https://doi.org/10.1016/j.scitotenv.2023.168777>.
- Memari, S. & Siadatmousavi, S. M. (2018) Numerical modeling of heat and brine discharge near Qeshm desalination plant. *International Journal Of Coastal, Offshore And Environmental Engineering(Ijcoe)* **3** (4), 27–35. <https://doi.org/10.29252/ijcoe.1.4.27>.
- National Geophysical Data Center (1999) *Bathymetry of Lake Erie and Lake Saint Clair [Dataset]*. National Geophysical Data Center, NOAA. <https://doi.org/10.7289/V5KS6PHK>.
- Nelson, S., Soranno, P. A., Spence Cheruvilil, K., Batzli, S. & Skole, D. (2002) Regional assessment of lake water clarity using satellite remote sensing. *Education Journal of Limnology* **62**, 27–32. <https://doi.org/10.4081/jlimnol.2003.s1.27>.
- Nguyen, T. D., Thupaki, P., Anderson, E. J. & Phanikumar, M. S. (2014) Summer circulation and exchange in the Saginaw Bay-Lake Huron system. *Journal of Geophysical Research: Oceans* **119** (4), 2713–2734. <https://doi.org/10.1002/2014JC009828>.
- Nguyen, T. D., Hawley, N. & Phanikumar, M. S. (2017) Ice cover, winter circulation, and exchange in Saginaw Bay and Lake Huron. *Limnology and Oceanography* **62** (1), 376–393. <https://doi.org/10.1002/lno.10431>.
- NOAA Electronic Navigational Charts (ENC) | InPort. (n.d.). Available from: <https://www.fisheries.noaa.gov/inport/item/39976> (accessed 20 February 2023).
- NOAA National Data Buoy Center (1971) *Meteorological and Oceanographic Data Collected from the National Data Buoy Center Coastal-Marine Automated Network (C-MAN) and Moored (Weather) Buoys*. Available from: <https://www.ncei.noaa.gov/archive/accession/NDBC-CMANWx>.
- Normandin, C., Lubac, B., Sottolichio, A., Frappart, F., Ygorra, B. & Marieu, V. (2019) Analysis of suspended sediment variability in a large highly turbid estuary using a 5-year-long remotely sensed data archive at high resolution. *Journal of Geophysical Research: Oceans* **124** (11), 7661–7682. <https://doi.org/10.1029/2019JC015417>.
- Olmanson, L. G., Bauer, M. E. & Brezonik, P. L. (2008) A 20-year landsat water clarity census of Minnesota's 10,000 lakes. *Remote Sensing of Environment* **112** (11), 4086–4097. <https://doi.org/10.1016/j.rse.2007.12.013>.
- Pelletier, G. J., Chapra, S. C. & Tao, H. (2006) QUAL2Kw – a framework for modeling water quality in streams and rivers using a genetic algorithm for calibration. *Environmental Modelling & Software* **21** (3), 419–425.
- Planet Labs PBC 2019 Planet Application Program Interface: In Space for Life on Earth. Planet. <https://api.planet.com>.
- Pu, F., Ding, C., Chao, Z., Yu, Y. & Xu, X. (2019) Water-quality classification of inland lakes using landsat8 images by convolutional neural networks. *Remote Sensing* **11** (14), Article 14. <https://doi.org/10.3390/rs11141674>.
- Rasmussen, C. E. & Williams, C. K. I. (2005) *Gaussian Processes for Machine Learning*. The MIT Press. <https://doi.org/10.7551/mitpress/3206.001.0001>.
- Saberioon, M., Brom, J., Nedbal, V., Souček, P. & Císar, P. (2020) Chlorophyll-a and total suspended solids retrieval and mapping using Sentinel-2A and machine learning for inland waters. *Ecological Indicators* **113**, 106236. <https://doi.org/10.1016/j.ecolind.2020.106236>.

- Safaie, A., Weiskerger, C. J., Nguyen, T. D., Acrey, B., Zepp, R. G., Molina, M., Cyterski, M., Whelan, G., Pachepsky, Y. A. & Phanikumar, M. S. (2020) Modeling the photoinactivation and transport of somatic and F-specific coliphages at a Great Lakes beach. *Journal of Environmental Quality* **49** (6), 1612–1623. <https://doi.org/10.1002/jeq2.20153>.
- Schulz, E., Grasso, F., Le Hir, P., Verney, R. & Thouvenin, B. (2018) Suspended sediment dynamics in the macrotidal seine estuary (France): 2. Numerical modeling of sediment fluxes and budgets under typical hydrological and meteorological conditions. *Journal of Geophysical Research: Oceans* **123** (1), 578–600. <https://doi.org/10.1002/2016JC012638>.
- Schwab, D. J. & Morton, J. A. (1984) Estimation of overlake wind speed from overland wind speed: A comparison of three methods. *Journal of Great Lakes Research* **10** (1), 68–72.
- Sequeiros, O. E., Cantelli, A., Viparelli, E., White, J. D. L., García, M. H. & Parker, G. (2009) Modeling turbidity currents with nonuniform sediment and reverse buoyancy. *Water Resources Research* **45** (6). <https://doi.org/10.1029/2008WR007422>.
- Sibson, R. (1981) A brief description of natural neighbour interpolation. *Interpreting Multivariate Data* (Barnett, V., ed.). John Wiley & Sons, Chichester, pp. 21–36.
- Smagorinsky, J. (1963) General circulation experiments with the primitive equations: I. The basic experiment. *Monthly Weather Review* **91** (3), 99–164. [https://doi.org/10.1175/1520-0493\(1963\)091<0099:GCEWTP>2.3.CO;2](https://doi.org/10.1175/1520-0493(1963)091<0099:GCEWTP>2.3.CO;2).
- Stradling, D. & Stradling, R. (2008) Perceptions of the Burning River: Deindustrialization and Cleveland's Cuyahoga River. *Environmental History* **13** (3), 515–535.
- Syariz, M. A., Lin, C.-H., Heriza, D., Lasminto, U., Sukojo, B. M. & Jaelani, L. M. (2022) A transfer learning technique for inland chlorophyll-a concentration estimation using Sentinel-3 imagery. *Applied Sciences* **12** (1), Article 1. <https://doi.org/10.3390/app12010203>.
- Topp, S. N., Pavelsky, T. M., Jensen, D., Simard, M. & Ross, M. R. V. (2020) Research trends in the use of remote sensing for inland water quality science: Moving towards multidisciplinary applications. *Water* **12** (1), Article 1. <https://doi.org/10.3390/w12010169>.
- United States Army Corps of Engineers. Hydrographic Surveys. (2023). Available from: <https://navigation.usace.army.mil/Survey/Hydro>.
- U. S. Geological Survey. (2016) *USGS Water Data for the Nation*. <https://doi.org/10.5066/F7P55KJN>
- U.S. Geological Survey. (2022) *Turbidity – Units of Measurement*. Available from: <https://or.water.usgs.gov/grapher/fnu.html> (accessed 20 February 2024)
- Vanhellemont, Q. & Ruddick, K. (2021) Atmospheric correction of Sentinel-3/OLCI data for mapping of suspended particulate matter and chlorophyll-a concentration in Belgian turbid coastal waters. *Remote Sensing of Environment* **256**, 112284. <https://doi.org/10.1016/j.rse.2021.112284>.
- Venkateswarlu, T. & Anmala, J. (2023) Importance of land use factors in the prediction of water quality of the Upper Green River watershed, Kentucky, USA, using random forest. *Environment, Development and Sustainability*. <https://doi.org/10.1007/s10668-023-03630-1>.
- Wang, W. & Jing, B.-Y. (2022) Gaussian process regression: Optimality, robustness, and relationship with kernel ridge regression. *Journal of Machine Learning Research* **23** (193), 1–67.
- Water Quality and Health: Review of Turbidity. (2017). Available from: <https://www.who.int/publications-detail-redirect/WHO-FWC-WSH-17.01>.
- Yang, Z. & Khangaonkar, T. (2007) Development of A Hydrodynamic Model of Puget Sound and Northwest Straits. <https://doi.org/10.2172/1013954>.
- Zheng, G. & DiGiacomo, P. M. (2022) A simple water clarity-turbidity index for the Great Lakes. *Journal of Great Lakes Research* **48** (3), 686–694. <https://doi.org/10.1016/j.jglr.2022.03.005>.
- Zhu, Q., Shen, F., Shang, P., Pan, Y. & Li, M. (2019) Hyperspectral remote sensing of phytoplankton species composition based on transfer learning. *Remote Sensing* **11** (17), Article 17. <https://doi.org/10.3390/rs11172001>.
- Zhu, C., van Maren, D. S., Guo, L., Lin, J., He, Q. & Wang, Z. B. (2022) Feedback effects of sediment suspensions on transport mechanisms in an estuarine turbidity maximum. *Journal of Geophysical Research: Oceans* **127** (6), e2021JC018029. <https://doi.org/10.1029/2021JC018029>.
- Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H. & He, Q. (2021) A comprehensive survey on transfer learning. *Proceedings of the IEEE* **109** (1), 43–76. <https://doi.org/10.1109/JPROC.2020.3004555>.

First received 7 April 2024; accepted in revised form 21 August 2024. Available online 3 September 2024