

Tidal Wave Transformation in the Mahakam Delta: Integrating Observations and Machine Learning

Muhammad Riza*, Nanda Khoirunisa, Zetsaona Sihotang, Idris Mandang

¹Department of Geophysics, Mulawarman University, Samarinda, 75123, Indonesia.

*Corresponding author. Email: Muhammad.riza@fmipa.unmul.ac.id

Manuscript received: 19 November 2025; Received in revised form: 30 June 2026; Accepted: 27 July 2026

Abstract

Tidal dynamics in tropical estuaries have made it difficult to predict water level variations and to understand the interaction between river discharge and ocean forcing. This study investigated tidal wave propagation, attenuation, and asymmetry in the Mahakam Delta, East Kalimantan, Indonesia. The objective was to quantify tidal wave transformation along the estuary-upstream transect and to evaluate the capability of machine learning to reproduce key hydrodynamic parameters. Water level time series with a 10-minute interval, collected from 12 July to 4 August 2018 at an estuary station and an upstream station, were analysed. A Random Forest regression model with temporal lag features predicted upstream water level with a coefficient of determination (R-squared) of 0.883 and a mean absolute error of 0.150 m; the most informative lag was 60 minutes. Analysis of 34 tidal cycles showed a mean attenuation coefficient of 0.76, indicating hyposynchronous conditions dominated by friction. A machine learning model for attenuation coefficient prediction reached an R-squared of 0.968 and a mean absolute error of 0.084, with estuary peak water level as the most influential predictor. The delta exhibited flood-dominant tidal asymmetry and negligible phase lag between estuary and upstream peaks. These findings demonstrated that the integration of data-driven modelling and hydrodynamic analysis can improve the prediction of water levels and support flood risk assessment in tropical deltaic systems.

Keywords: Deltaic Hydrodynamics; Hyposynchronous Estuary; Mahakam Delta; Random Forest Regression; Tidal Wave Propagation.

Citation: Riza, M., Khoirunisa, N., Sihotang, Z., & Mandang, I. (2026). Tidal Wave Transformation in the Mahakam Delta: Integrating Observations and Machine Learning. *Jurnal Geocelebes*, 10(2): 269–286, doi: 10.70561/geocelebes.v10i2.48409

Introduction

Tidal dynamics in tropical estuaries provide considerable obstacles for forecasting water level fluctuations and comprehending intricate hydrodynamic interactions. Efficient management of water resources, flood prediction, and coastal defense requires comprehension of the propagation, attenuation, and transformation of tidal waves as they move from estuaries to upstream river segments. The Mahakam Delta in East Kalimantan, Indonesia, illustrates these challenges as an extensive tropical river–estuary system whose hydrodynamics are governed by the combined effects of ocean tides, river flow,

and complex channel geometry (Van et al., 2020).

The metamorphosis of tidal waves in estuaries is regulated by two opposing physical forces. The geometric convergence of estuary morphology may enhance tidal energy, producing hypersynchronous systems characterized by higher waves, while bottom friction reduces tidal energy, resulting in smaller waves in hyposynchronous systems (Khojasteh et al., 2023). These changes in tidal-wave propagation can also generate tidal asymmetry, expressed as unequal flood and ebb durations and current intensities. Such asymmetry regulates

residual sediment transport and sediment redistribution among branching estuarine channels, with implications for channel siltation, tidal-flat accretion, and longer-term morphological change (Zhou et al., 2025). Seasonal fluctuations in river flow markedly affect tidal propagation by changing effective friction due to alterations in depth and the balance between pressure-gradient and frictional forces (Tenorio-Fernandez et al., 2019).

Classical harmonic analysis has limited ability to reproduce non-stationary estuarine water levels because tidal amplitudes and phases are modified by nonlinear interactions among river discharge, tidal forcing, water depth, and channel geometry. Although process-based hydrodynamic models can explicitly represent these interactions, they generally require detailed boundary conditions and extensive model configuration (Gan et al., 2024). Such modelling challenges are pronounced in the Mahakam Delta because of its complex channel network, variable topography, and interacting riverine and marine forcing across the river–delta–sea continuum (Van et al., 2020). Traditional models fail to accurately represent nonlinear tidal dynamics and demand significant computing resources. Recent advancements in machine learning, specifically Random Forest and deep learning frameworks, present promising alternatives by identifying nonlinear relationships within tidal systems while offering computational efficiency benefits for real-time forecasting (Zhou et al., 2025; Lim & Zohren, 2021). Temporal lag characteristics in machine learning models have shown significant effectiveness in predicting tides by integrating historical data at different time intervals, which represent wave velocity and propagation distance via estuarine systems (Han et al., 2024). Interpretability approaches like SHAP (SHapley Additive exPlanations) facilitate comprehension of black-box models by pinpointing the factors that most

substantially affect tidal dynamics (Kim et al., 2023; Lundberg & Lee, 2017).

Hydrodynamic processes and tidal dynamics within the Mahakam river-lake-delta system have been extensively modeled using multi-scale, finite-element simulations and hydraulic geometry frameworks (de Brye et al., 2011; Sassi et al., 2011; Van et al., 2016). Research has also successfully quantified how upstream river flow controls tidal propagation and water level profiles within the system's freshwater reaches (Sassi & Hoitink, 2013). More recent data-driven studies have applied neural-network architectures to tidal rivers and estuaries, including long short-term memory and feed-forward neural networks that incorporate river discharge to improve non-stationary tidal water-level predictions (Zhang et al., 2023). Within the Mahakam system, other recent studies have focused on the morphometric evolution and sediment infilling of the upstream Mahakam Lakes (Jati et al., 2026) or localized sediment dynamics (Rasimeng et al., 2019), a critical gap remains. Specifically, the integration of contemporary, modern machine learning methodologies has yet to be comprehensively explored for quantitative tidal analysis in the delta.

This study fills the gap by combining high-resolution observational data with machine learning methods to quantify tidal attenuation coefficients and their temporal variations, examine spatial and temporal patterns of tidal asymmetry, create predictive models for upstream water levels utilizing temporal lag features, and determine key factors influencing tidal wave propagation through feature importance evaluation. This study seeks to increase scientific comprehension and predictive accuracy of tidal dynamics in tropical deltaic systems by integrating extensive tidal cycle research with modern machine learning, hence offering improved

instruments for water resource management and flood prediction.

Methods

Research Location

This study was conducted in the Mahakam Delta, located on the eastern coast of Borneo Island, East Kalimantan, Indonesia (Figure 1). The delta occupies the low-lying coastal plain along the lower reach of the Mahakam River, bounded approximately by longitudes 117.15°E–117.68°E and latitudes 1.01°S–0.27°S. This area is characterized by a dense network of distributary channels, tidal creeks, and coastal wetlands that are strongly influenced by both river discharge and tidal forcing from the adjacent Makassar Strait.

Two tidal datasets were used to represent the upstream and downstream boundaries of the study domain. The upstream boundary is defined by an observed tide station (field observation site) located in the

upper delta at approximately 0.57746°S, 117.2747°E, capturing water-level variability in the fluvial-tidal transition zone. The downstream boundary is specified using modeled water-level data derived from tidal surge predictions provided by the Geospatial Information Agency (BIG), representing conditions near the estuary mouth at approximately 0.695061°S, 117.627144°E, where marine tidal influence is dominant. The modeled dataset was converted into a continuous time series using only records with valid timestamps and was used as a proxy for offshore tidal dynamics. Validation of the BRIN-associated tidal model against in situ observations in Indonesian coastal settings showed good agreement, with reported Willmott Skill Score 0.652 (Sufyan et al., 2023). These validation results indicate the model reproduces the main tidal variability with reasonable accuracy and can serve as a regional offshore tidal boundary signal rather than a direct substitute for estuarine point observations.

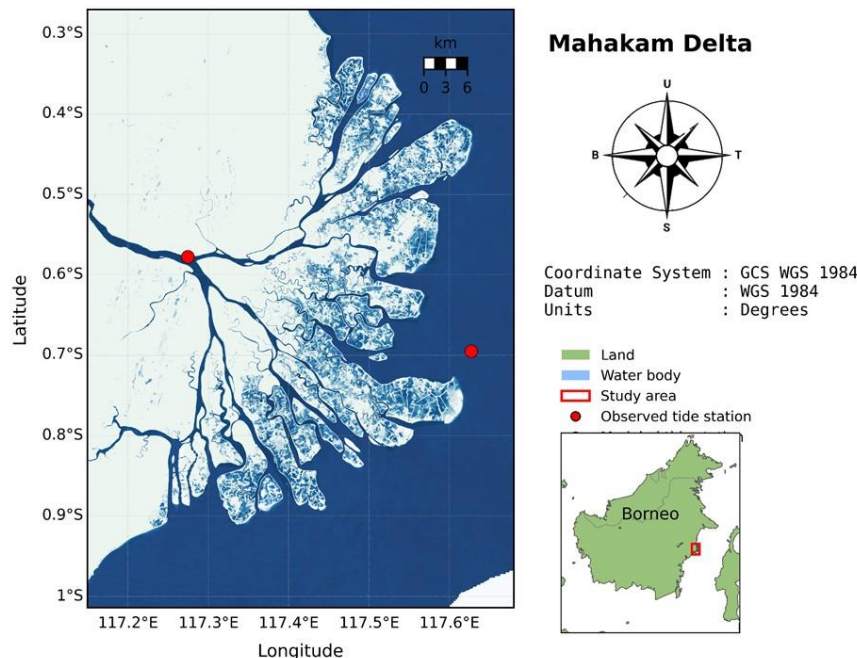


Figure 1. Research location in Mahakam Delta, East Kalimantan, Indonesia.

Data and Pre-processing

This study uses two water level (WL) time series:

1. Water Level at the Estuary Mouth

- a. Source: Model outputs from tidal surge predictions from BIG (Geospatial Information Agency).
- b. Format: Each row contains *longitude*, *latitude*, *date*, *time*, and *elevation* (m).

Only rows with complete date patterns (YYYY-MM-DD HH:MM:SS) are read and stored as a time series `wl_estuary` with a time index.

2. Water Level Upstream (Upper Delta)
 - a. Source: Data logger tidal station at Anggana.
 - b. Format: Data recorded in the data logger with 2-minute intervals.

Both time series are subsequently processed as follows:

- a. Resampled to 10-minute resolution (`RESAMPLE_RULE = "10min"`).
 - `wl_estuary` is linearly interpolated with respect to time.
 - `wl_upstream` is averaged within each 10-minute interval.
- b. Inner-joined so that only time points with complete data at both the mouth and upstream are retained (Hu et al., 2024).

Dataset Assembly and Lag Feature Engineering

To model upstream water level as a response to conditions at the estuary mouth, lag features are constructed from the estuarine water level with $\eta_e(t)$ is qater level at the estuary mouth at time t .

The use of estuarine-mouth water levels as predictors is physically reasonable because tidal signals propagate from the estuary mouth toward upstream locations with a measurable phase delay. The magnitude of this delay may vary spatially because of channel geometry, bathymetry, bottom friction, river discharge, and tidal forcing (Iglesias et al., 2019). Previous water-level observations also contain important information about the temporal evolution of hydrological systems and can therefore be used as lagged input variables in data-driven water-level models (Liu et al., 2022; Gan et al., 2024).

Accordingly, the program generates the following lagged features from the estuarine water-level series:

$$\begin{aligned} X(t) \\ = \eta_e(t), \eta_e(t - 10 \text{ min}), \eta_e(t - 20 \text{ min}), \dots, \eta_e(t - 60 \text{ min}) \end{aligned} \quad (1)$$

Lagged water-level features were constructed from antecedent estuarine water levels because historical water-level observations contain information on the temporal dependence of estuarine hydrodynamics and can improve short-term prediction performance (Gan et al., 2024).

Each lagged variable represents the water level observed at the estuary mouth at a particular time before the corresponding upstream observation. These variables allow the model to identify delayed relationships between changes occurring at the estuary mouth and the subsequent response of the upstream water level. Such lagged representations are particularly useful because water-level time series commonly exhibit temporal dependence and autocorrelation, meaning that current water-level conditions are partly related to previous observations (Gan et al., 2024).

Generating lagged variables through temporal shifting produces missing values in the first six observations because complete historical records are not available for these initial timestamps. Therefore, samples containing NaN values resulting from the shifting operation are removed. The remaining complete observations are then assembled into the final modelling dataset

Extraction of Tidal Cycles and Hydrodynamic Metrics

1. Tidal Peak Detection

After resampling and temporal synchronization, the two water-level time series were denoted as `wl_estuary` and `wl_upstream`. The variable `wl_estuary` represents the water level at the estuary mouth, whereas `wl_upstream` represents the water level at the upstream station. Both series were expressed in meters and

recorded at synchronized 10-minute intervals (Palmer et al., 2023).

- The estuary water level (wl_{estuary}) is smoothed using a rolling maximum with a window of 5 steps (± 50 minutes).
- A time point (t_{est}) is designated as a tidal peak at the estuary mouth if:

$$\eta_e(t_{\text{est}}) = \max_{t \in [t_{\text{est}} - 25 \text{ min}, t_{\text{est}} + 25 \text{ min}]} \eta_e(t) \quad (2)$$

High-water turning points were identified using a peak-based approach in which a tidal peak corresponds to a local maximum within a predefined temporal neighborhood (Palmer et al., 2023).

- For each estuary peak (t_{est}), an upstream peak (t_{up}) is identified as the time of maximum wl_{upstream} within a 3-hour window following the estuary peak:

$$t_{\text{up}} = \arg \max_{t \in [t_{\text{est}}, t_{\text{est}} + 3 \text{ hours}]} \eta_u(t) \quad (3)$$

where $\eta_u(t)$ is the upstream water level. This procedure represents the landward propagation of tidal waves, which may experience changes in amplitude and timing because of friction, channel geometry, water depth, and river discharge (Hoitink & Jay, 2016). For each estuarine high-water peak, the corresponding upstream peak was defined as the maximum upstream water level occurring within the subsequent three-hour search window. The window length was selected as a study-specific criterion to accommodate the expected landward propagation time.

2. Tidal Attenuation (Coefficient K)

Peak heights at the estuary mouth and upstream are defined as:

$$H_{\text{est}} = \eta_e(t_{\text{est}}), \quad H_{\text{up}} = \eta_u(t_{\text{up}}) \quad (4)$$

Tidal attenuation (attenuation coefficient) is defined as the ratio:

$$K = \frac{H_{\text{up}}}{H_{\text{est}}} \quad (5)$$

Cases where $H_{\text{est}} = 0$ are eliminated to avoid division by zero (Khojasteh et al., 2021; Xie & Wang, 2021).

3. Phase Lag

Phase lag between upstream and estuary mouth responses is measured as the time difference between the estuary peak and upstream peak, in minute:

$$\Delta\phi = \frac{t_{\text{up}} - t_{\text{est}}}{60 \text{ seconds/minute}} [\text{minutes}] \quad (6)$$

The high-water propagation lag was calculated as the difference between the occurrence time of the upstream and estuarine high-water peaks, expressed in minutes (Iglesias et al., 2019).

Flood-Ebb Asymmetry (FE)

To characterize the asymmetry between rising (flood) and falling (ebb) phases at upstream, a 12-hour segment is constructed around the upstream peak (Yu et al., 2020):

$$\text{segment} = [t_{\text{up}} - 6 \text{ hours}, t_{\text{up}} + 6 \text{ hours}] \quad (7)$$

A 12-hour analysis window centered on each upstream high-water peak was used as a study-specific approximation of one semidiurnal tidal cycle. Within this window, the preceding and subsequent low-water turning points were identified (Yu et al., 2020; Wünsche et al., 2024). Within this segment, the following are identified:

- $(t_{\text{min}}^{\text{before}})$: time of minimum water level before the peak (flood phase onset).
- $(t_{\text{min}}^{\text{after}})$: time of minimum water level after the peak (ebb phase end).

Flood duration is defined as the period from the preceding low water to the following high water, whereas ebb duration is the period from high water to the following low water (Wünsche et al., 2024):

$$T_f = t_{\text{up}} - t_{\text{min}}^{\text{before}}, \quad T_e = t_{\text{min}}^{\text{after}} - t_{\text{up}} \quad (8)$$

(both in minutes). Flood-ebb asymmetry is defined as (Wünsche et al., 2024):

$$FE = \frac{T_f}{T_e} \quad (9)$$

If $T_e \leq 0$ or is undefined, the cycle is excluded (marked as NaN) This ratio represents the inequality between the rising and falling tidal durations (Guo et al., 2018; Hu et al., 2024).

Local Tidal Range at the Estuary Mouth

For the same cycle, the local tidal range at the estuary mouth is computed within the 12-hour segment:

$$TR_{est} = \max(\eta_e(t)) - \min(\eta_e(t)) \in [t_{up} - 6 \text{ hours}, t_{up} + 6 \text{ hours}] \quad (10)$$

Local tidal range was defined as the difference between the maximum and minimum water levels within the identified tidal cycle (Khojasteh et al., 2021; Palmer et al., 2023). The 12-hour segment is used as a study-specific approximation for capturing one semidiurnal tidal cycle. Additionally, for each cycle:

a. Hour of day at estuary peak:

$$\text{hour_of_day} = \text{hour}(t_{est}) \quad (11)$$

The hour of day was extracted from the timestamp of each estuarine high-water peak and was included as an operational temporal predictor.

b. Estuary peak level: $\text{estuary_peak_level} = H_{est}$ (Sandbach et al., 2018; Xie & Wang, 2021).

Filtering of Tidal Cycles (Hydro Cycles Clean)

To focus on physically realistic and relevant cycles, the following filtering criteria are applied:

a. Minimum tidal range at estuary mouth:

$$TR_{est} \geq 1.0 \text{ m}$$

Cycles with an estuarine tidal range below 1.0 m were excluded using a

study-specific threshold intended to remove weak or poorly resolved tidal cycles.

b. Realistic tidal attenuation:

$$0 < K \leq 3.0$$

Positive tidal-range ratios were retained, while extremely large ratios were treated as potential peak-matching or data-quality outliers. The upper threshold was selected empirically and evaluated through sensitivity analysis.

c. Reasonable phase lag:

$$0 \leq \Delta\phi \leq 60 \text{ minutes}$$

Only non-negative peak lags of up to 60 minutes were retained. The upper limit was selected as a study-specific criterion based on station spacing, temporal resolution, and the expected order of high-water propagation times reported for shallow estuarine systems (Iglesias et al., 2019).

d. Removal of entire cycles with missing values (NaN) in `attenuation_K` or `phase_lag_min`.

This criterion ensures that the upstream tidal peak occurs after the estuary peak, consistent with the landward propagation of tidal waves (Iglesias et al., 2019). The maximum phase lag of 60 minutes is applied as a study-specific threshold based on the expected hydrodynamic response between the two stations.

Machine Learning Modeling of Upstream Water Level

To evaluate how far estuarine dynamics can explain upstream water level variations at the 10-minute scale, a Random Forest Regression model is employed with the following framework. Random Forest has been widely applied to hydrological water-level forecasting because it can represent complex and nonlinear relationships between predictor and response variables (Atashi et al., 2022; Nhu et al., 2020). Historical water-level observations can also provide relevant predictors for short-term

estuarine water-level modeling because water-level time series exhibit temporal dependence (Gan et al., 2024).

a. Model Structure

Target:

$$y(t) = \eta_u(t) \text{ (upstream water level)} \quad (12)$$

Features:

$$x(t) = [\eta_e(t), \eta_e(t - 10 \text{ min}), \dots, \eta_e(t - 60 \text{ min})] \quad (13)$$

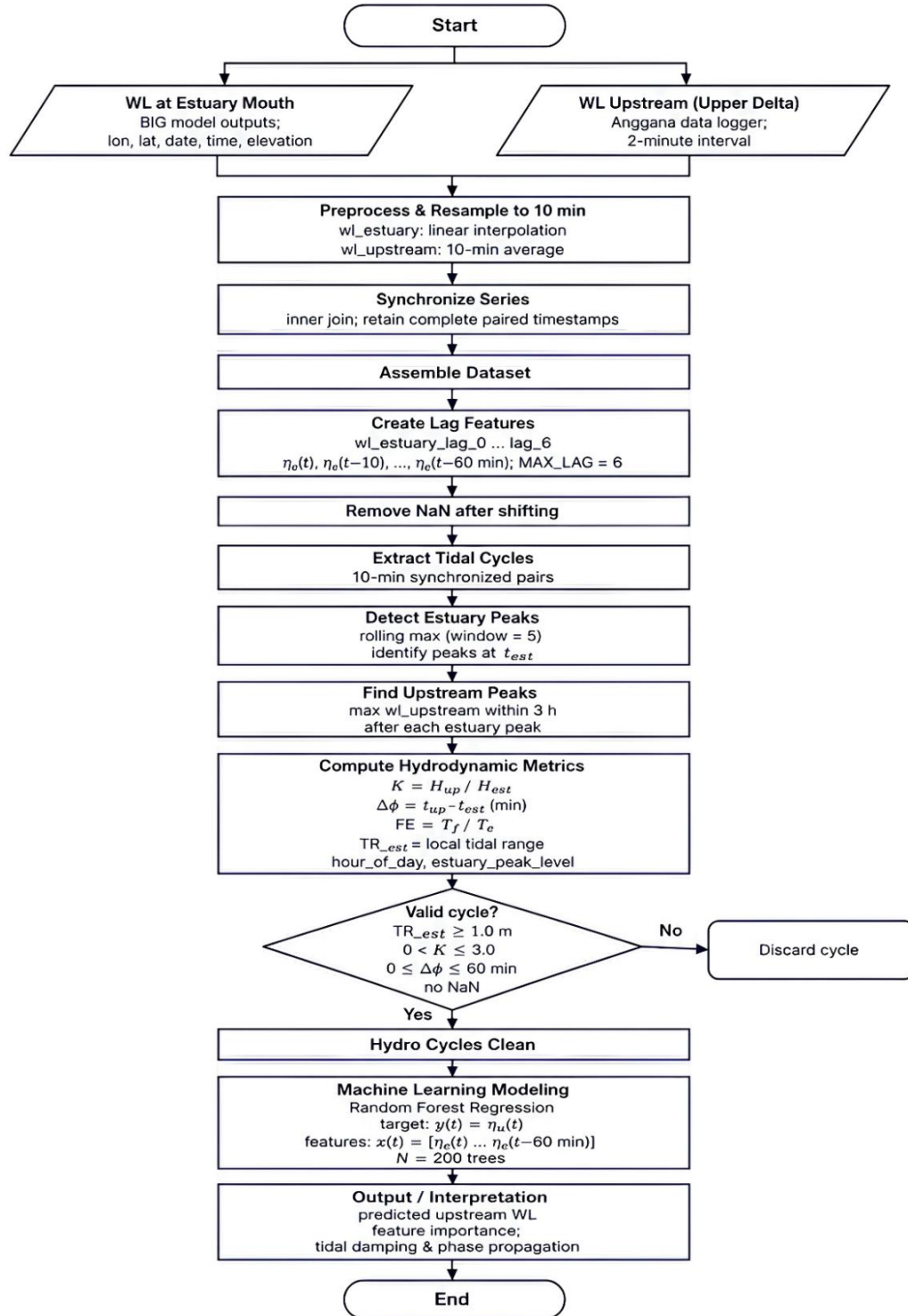


Figure 2. Data processing and machine learning workflow for estuarine and upstream tidal water level analysis.

The use of lagged water-level observations as model inputs is consistent with previous machine-learning studies showing that antecedent water levels can contribute substantially to subsequent water-level prediction (Nhu et al., 2020; Gan et al., 2024). The Random Forest model learns a non-linear function (James et al., 2021):

$$\hat{y}(t) = f(x(t)) \quad (14)$$

computed as the average prediction from N regression decision trees ($N = 200$):

$$\hat{y}(t) = \frac{1}{N} \sum_{i=1}^N f_i(x(t)) \quad (15)$$

where f_i is the i -th tree (James et al., 2021; Hosseinzadeh et al., 2023).

b. Hydrodynamic Interpretation

The ensemble approach provides robust predictions while permitting interpretation of feature importance. Lagged estuarine water levels serve as proxies for wave propagation and tidal signal transmission through the estuary-delta system. The model implicitly captures the phase relationships and tidal damping effects that characterize the hydrodynamics of the study region (Zhou et al., 2025).

Figure 2 illustrates the workflow used to develop a machine-learning model for predicting upstream water level from downstream observations (Sarafanov et al., 2020). Water-level data from the estuary

mouth and upper-delta stations are first processed resampled to 10-minute intervals, and synchronized using matched timestamps. The combined dataset is then used to generate lagged downstream water-level features, after which tidal cycles are identified and screened using physically plausible hydrodynamic criteria, including tidal range, time lag, and the absence of missing values. For each valid cycle, upstream peak water level and associated hydrodynamic metrics are calculated. Finally, these clean tidal-cycle data are used to train a Random Forest regression model, enabling prediction of upstream water-level peaks and supporting the interpretation of tidal damping and propagation along the estuary (Zounemat-Kermani et al., 2020).

Results and Discussion

Tidal Data Characteristics

This research analyzed tidal data with a temporal resolution of 10 minutes, gathered from July 12 to August 4, 2018, yielding a total of 3,303 observations in the Mahakam Delta estuary and upstream. The water levels in the Mahakam Delta estuary fluctuated between -0.970 and 1.355 m, whereas at the Mahakam Delta upstream, they varied between -0.488 and 1.690 m (Table 1). Figure 3 illustrates that the water levels at both sites exhibited a distinct semidiurnal tidal pattern throughout time.

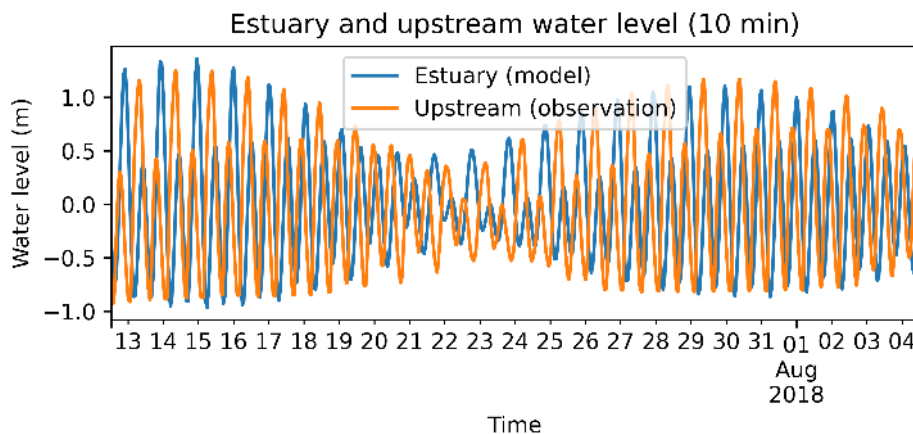


Figure 3. Timeseries estuary and upstream.

The Pearson correlation value between estuary and upstream water levels was -0.072, indicating a negligible relationship. This indicates a negligible direct linear correlation between the two locations. The scatter plot (Figure 4) illustrates that the data is distributed in a complex and non-linear manner. The data concentration exhibits an elliptical pattern that aligns with the tidal cycle.

Scatter of estuary and upstream water level

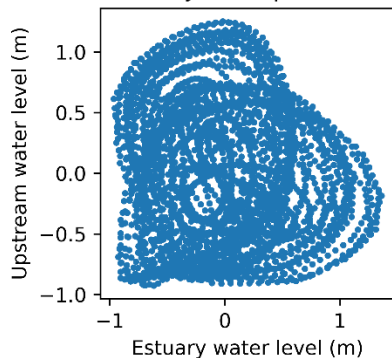


Figure 4. Scatter estuary and upstream.

Machine Learning Prediction with Temporal Lag

Table 1. Summary of Tidal Dynamics Research Results at Mahakam Delta Estuary and Upstream.

Parameter	Value
Temporal data resolution	10 minutes
Study period	July 12 - August 4, 2018
Number of observations	3,303 samples
Water level range at Mahakam Delta estuary	-0.970 to 1.355 m
Water level range at upstream Mahakam Delta	-0.488 to 1.690 m
R ² model	0.883
MAE model	0.150 m
Optimal temporal lag	60 minutes (6 intervals)
Number of tidal cycles analyzed	34 cycles
Mean attenuation coefficient (K)	0.762 ± 0.651
Mean tidal range at estuary	1.571 ± 0.342 m
Mean phase lag	0 minutes
Mean tidal asymmetry (Fe)	0.632 ± 0.599

The machine learning model using temporal lag characteristics achieved a coefficient of determination (R^2) of 0.883 and a Mean Absolute Error (MAE) of 0.150 m (Table 1). An R^2 value over 0.85

indicates that the model accounts for 88.3% of the variations in upstream Mahakam Delta water levels by using estuarine Mahakam Delta data with a temporal. This performance parallels experiments using a multilayer perceptron to forecast tides in estuaries, achieving R^2 values as high as 0.97 (Guillou & Chapalain, 2021).

The optimal time lag identified was 60 minutes (comprising 6 periods of 10 minutes), indicating that it takes about one hour for the tidal signal to propagate from the Mahakam Delta estuary to the upstream Mahakam Delta. Machine learning methodologies, particularly Recurrent Neural Networks (RNN) and hybrid systems, are the most promising approach to modeling complex physical processes in tidal rivers and estuaries. The efficacy of the lag-feature model confirms a substantial time-dependent link between tidal dynamics at the Mahakam Delta estuary and the resultant reactions at the upstream Mahakam Delta.

Tidal Cycle Analysis and Attenuation

An analysis of 34 tidal cycles revealed important characteristics of tidal wave propagation in the studied Mahakam Delta system. The tidal attenuation coefficient (K) exhibited a mean value of 0.762 ± 0.651 , however its distribution varied significantly (Figure 5).

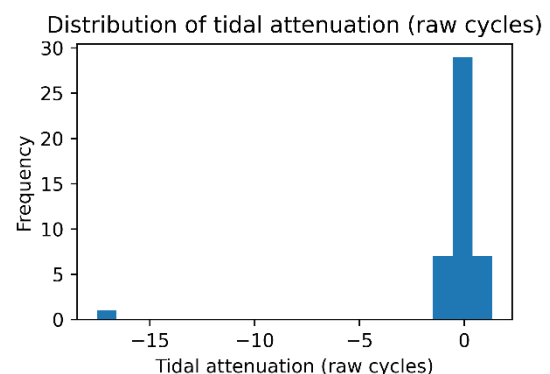


Figure 5. Distribution of tidal attenuation (raw cycles).

The histogram distribution indicates that the majority of cycles (about 43) had little attenuation (K around 0). The majority of

cycles have low K values, indicating a hyposynchronous condition. This indicates that frictional effects surpass the geometric convergence of the estuary, resulting in the attenuation of tidal waves as they propagate from the estuary to the upstream Mahakam Delta. In hyposynchronous deltas, bottom friction and riverbank resistance are the primary factors influencing tidal amplitude decrease in inland areas (Khojasteh et al., 2021). In hypersynchronous systems, geometric convergence dominates, leading to tidal amplification. The mean tidal range of the Mahakam Delta estuary was 1.571 ± 0.342 m, indicating that the region was in a mesotidal condition (Khojasteh et al., 2021). The alteration in tidal range results from the spring-neap cycles prevalent in semidiurnal tidal systems (Garel et al., 2021; Xie & Wang, 2021).

Phase Lag and Tidal Asymmetry

The research indicated that the mean phase lag between tidal maxima in the Mahakam Delta estuary and the upstream Mahakam Delta was 0 minutes for all cycles studied. This result suggests that the dominant belief that tidal wave propagation requires measurable travel time may need reconsideration. In the higher reaches of tidal rivers, river discharge may entirely govern incoming tidal attenuation. In some instances, river flow may postpone or alter tidal effects (Sassi & Hoitink, 2013). The mean tidal asymmetry value (Fe) was 0.632 ± 0.599 , indicating that the Mahakam Delta system analyzed had a greater propensity for flooding than for ebbing. Flood-dominant tidal asymmetry is characterized by a reduced flood duration compared to ebb length, resulting in heightened peak currents during the flood phase. Shallow-water impacts, friction, and/or inadequate exchange between deep channels and intertidal zones may asymmetrically affect tidal timing and amplitude (Spicer et al., 2023). Flood dominance often transports material into interior regions (upstream of the Mahakam Delta) during periods of slow to moderate

river flow (Friedrichs & Aubrey, 1994). Nonetheless, seasonal variations in river flow may significantly influence tidal asymmetry. For instance, elevated discharge during wet seasons may amplify tidal amplitudes in the lower reaches due to non-linear effects, while concurrently diminishing tidal amplitudes in the upstream Mahakam Delta due to increased friction damping (Yu et al., 2020).

Feature Importance in Water Level Prediction

The machine learning model for forecasting water levels in the upstream Mahakam Delta indicated that temporal lag elements derived from estuary data are crucial for identifying tidal propagation patterns. Figure 6 illustrates the outcomes of a feature importance study conducted using the Random Forest model. The characteristic with the greatest relevance value is lag 6 (60 minutes), at 0.508, followed by lag 0 (present observation), at 0.309. These two primary characteristics together account for 81.8% of the overall model relevance. Historical data with the greatest lag and real-time estuary data from the Mahakam Delta are important for forecasting upstream water level conditions (Anggraini et al., 2024). These findings align with previous studies indicating that including temporal lag factors can improve prediction accuracy in systems with complex temporal dynamics (Liu et al., 2025).

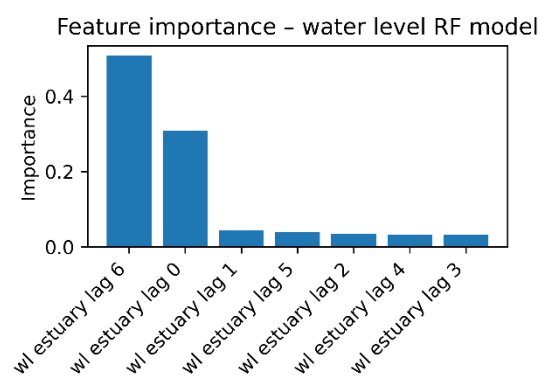


Figure 6. Feature importance-water level random forest model.

Tidal Attenuation Coefficient Modeling

The development of a machine learning model to predict the tidal attenuation coefficient yielded exceptional results, with a R^2 value of 0.968 and a mean absolute error (MAE) of 0.084 (Table 2). The precision of this forecast indicates that the model effectively simulated the movement of tidal waves from the Mahakam Delta estuary to the upstream Mahakam Delta.

The histogram of filtered attenuation distribution (Figure 7) exhibits a more concentrated pattern compared to the raw histogram. The majority of tidal cycles have attenuation coefficients ranging from 0 to 0.5, indicating that the Mahakam Delta experiences very hyposynchronous circumstances. Figure 8 presents a scatter plot depicting the relationship between attenuation and estuary tidal range. This indicates that the two variables are intricately connected. Generally, increased attenuation occurs more often when the tidal range is reduced, particularly within

the 1.0–1.5 m interval. Conversely, attenuation varies significantly at greater tidal ranges (> 2.0 m). This pattern indicates that the attenuation coefficient in the Mahakam Delta system is influenced by factors outside the tidal amplitude of the estuary. The interaction of river bottom topography, friction, and discharge also influences it (Khojasteh et al., 2023; Su & Jiang, 2023).

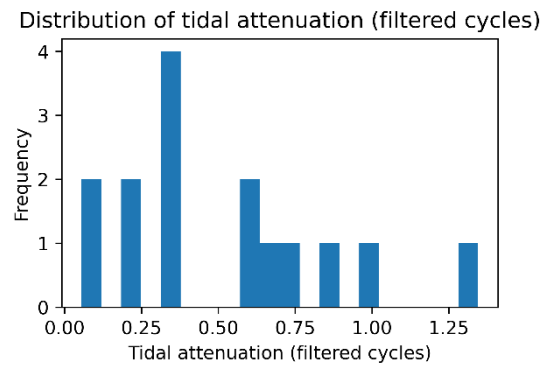


Figure 7. Distribution of tidal attenuation (filtered cycles).

Table 2. Performance of machine learning models for predicting tidal hydrodynamic parameters at Mahakam Delta.

Parameter	Model	R^2	MAE	RMSE
Water level (upstream)	Regression with lag features	0.883	0.150 m	—
Attenuation coefficient (K)	Random Forest	0.968	0.084	0.114
Tidal asymmetry (Fe)	Random Forest	0.786	0.318	0.114
Phase lag (minutes)	Classifier	0.0	—	—

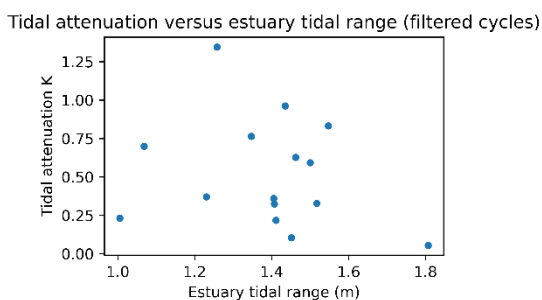


Figure 8. Tidal attenuation and estuary tidal range (filtered cycles)

Figure 9 illustrates an examination of feature importance for predicting attenuation. The paramount characteristic is the highest water level in the Mahakam Delta estuary, which has a significance value of 0.776. This element is far more critical than the others, indicating that the local tidal amplitude at the estuary is the

primary determinant influencing the magnitude of attenuation as waves propagate upwards. This conclusion aligns with studies demonstrating that peak water levels substantially affect the degree of shallow-water distortion and friction experienced by tidal waves (Su & Jiang, 2023).

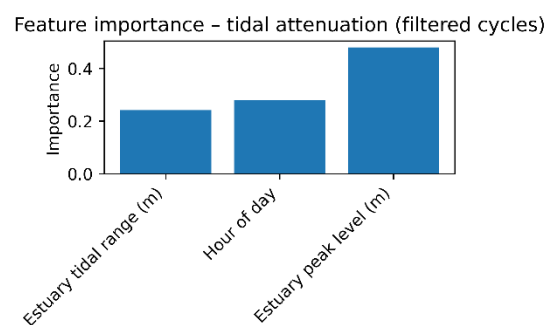


Figure 9. Feature importance-tidal attenuation.

Figure 9 presents the Random Forest feature-importance analysis for predicting the tidal attenuation coefficient. Estuary peak water level was identified as the most important predictor, with a feature-importance value of 0.776. This value was substantially higher than those of the other predictors, indicating that water-level conditions at the estuary boundary strongly influenced the landward transmission of the tidal signal. This interpretation is physically plausible because tidal propagation and attenuation in estuarine channels depend on the magnitude of tidal forcing and its interaction with bottom friction, river discharge, and channel geometry (Wang et al., 2021; Wu et al., 2022). This conclusion aligns with studies demonstrating that peak water levels substantially affect the degree of shallow-water distortion and friction experienced by tidal waves (Su et al., 2023).

Influence of Tidal Range on Attenuation

Table 3. Attenuation characteristics based on large tide vs. small tide groups at Mahakam Delta Estuary.

Characteristic	Large Tide	Small Tide
Number of cycles	24	10
Mean tidal range (m)	1.718 ± 0.236	1.218 ± 0.055
Median attenuation coefficient	0.379	0.956
Mean attenuation coefficient	0.625 ± 0.562	1.088 ± 0.771

The investigation, categorized by tidal range size (Table 3), revealed significant disparities in attenuation characteristics between large and small tides. The median attenuation coefficient was 0.379, whereas the mean was 0.625 during times of significant tidal fluctuations, with a mean range of 1.718 m (n=24 cycles). Conversely, during times of minor tides with a mean range of 1.218 m (n=10 cycles), the median attenuation coefficient increased to 0.956, while the mean reached 1.088. This indicates that damping was more pronounced during periods of reduced tidal amplitude. There are scientific

explanations for this peculiar phenomenon that first seem illogical. During high tide, the energy of the tidal wave increases, resulting in relatively significant amplitudes in the upstream reach despite the presence of friction. Conversely, during times of low tides, diminished wave amplitudes render friction effects comparatively more significant, resulting in elevated attenuation coefficients (Su & Jiang, 2023). This conclusion highlights the importance of data stratification based on spring-neap circumstances in the examination of complex tidal dynamics.

Phase Lag and Tidal Asymmetry Models

Presents a study of the significance of each feature in predicting phase lag. All characteristics have significance values around zero, aligning with empirical estimates indicating a mean phase lag of 0 minutes for all examined cycles. This indicates that certain circumstances in the Mahakam Delta during the research period resulted in simultaneous tidal maxima occurring at the estuary and upstream, despite the temporal delay of physical wave transmission (Zhou et al., 2025). The tidal asymmetry prediction model (Fe) performed adequately, although not as well as the attenuation model. The model exhibited a R^2 of 0.786 and a mean absolute error (MAE) of 0.318 (Table 2). The asymmetry model yields more varied predictions due to several variables influencing tidal asymmetry, including non-linear interactions among various tidal components, changes in bottom topography, and fluctuations in river flow. These results highlight the relative robustness of the attenuation model in capturing local tidal characteristics and their hydrodynamic implications. The attenuation model ($R^2 = 0.968$) outperforms the asymmetry prediction model ($R^2 = 0.786$), indicating that local tidal parametric properties provide more precise predictions of attenuation coefficients. This concerns the deterministic properties of tidal wave

propagation governed by continuity and momentum equations (Niesten et al., 2024).

Tidal asymmetry arises from complex overtide interactions and non-linear constituent interactions. This complicates predictions using linear or semi-linear data-driven models (Kim et al., 2024). The comprehensive synthesis of studies reveals that the Mahakam Delta system exhibits hydrodynamic characteristics typical of expansive tropical hyposynchronous estuaries. The significant influence of estuary tidal amplitude on attenuation coefficients indicates that bottom friction is the primary component governing this system (Khojasteh et al., 2023). As tidal waves go upstream, they dissipate energy due to friction, and their morphology alters as a result of shallow-water phenomena. This exacerbates the imbalance of flood dominance (Su & Jiang, 2023).

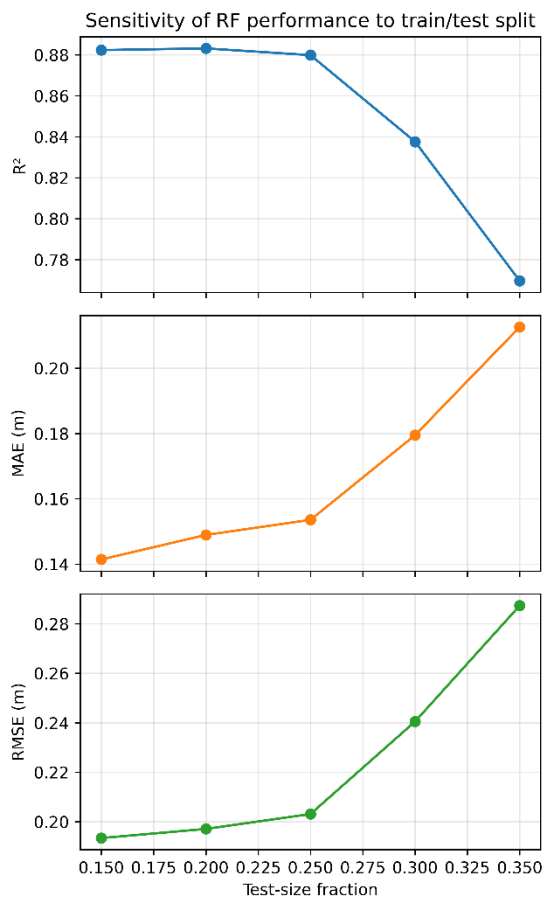


Figure 10. Sensitivity of Random Forest performance to different train/test split.

Figure 10 shows the sensitivity of the Random Forest model to different train/test partitions. The model achieved the highest performance at smaller test fractions, with R^2 values of 0.882, 0.883, and 0.880 for test fractions of 0.15, 0.20, and 0.25, respectively, and still remained relatively strong at a test fraction of 0.30 with $R^2 = 0.838$. However, performance declined further when the test fraction increased to 0.35, where R^2 dropped to 0.770, while MAE and RMSE increased from 0.142 m and 0.194 m at the 0.15 split to 0.207 m and 0.286 m at the 0.35 split, respectively.

These variations indicate that the estimated predictive performance was sensitive to the amount of data allocated to training and testing. In time-series forecasting, performance estimates can vary across testing periods because observations are temporally dependent and may exhibit non-stationary behavior; therefore, evaluation procedures should preserve temporal order and, where possible, examine performance across multiple testing periods (Bergmeir et al., 2018; Cerqueira et al., 2020). The relatively consistent performance across test fractions of 0.15–0.30 reduces, although does not eliminate, concern that the reported model skill resulted from a single favorable partition. Moreover, the simultaneous evaluation of R^2 , MAE, and RMSE provides a more comprehensive assessment than reliance on R^2 alone, because a high coefficient of determination does not necessarily imply low forecast error or reliable predictive performance (Hewamalage et al., 2023). Therefore, Figure 10 provides evidence of the model's temporal robustness within the observed study period, while also showing that its predictive accuracy was not uniform across all train/test partitions.

Conclusion

This study successfully integrated high-resolution observational data with machine learning techniques to investigate tidal wave propagation and attenuation in the

Mahakam Delta, East Kalimantan, Indonesia. The Mahakam Delta exhibits strongly hyposynchronous characteristics, with a mean attenuation coefficient of 0.762 ± 0.651 , indicating that frictional effects dominate over geometric convergence as tidal waves propagate upstream. The Random Forest model achieved exceptional predictive accuracy ($R^2 = 0.883$, MAE = 0.150 m) for upstream water level prediction using temporal lag features, with the optimal lag identified at 60 minutes. The attenuation coefficient prediction model ($R^2 = 0.968$, MAE = 0.084) identified estuary peak water level as the dominant factor controlling tidal wave energy dissipation, accounting for 77.6% of feature importance. The result that median attenuation is higher during small tides (0.956) than during large tides (0.379) indicates a non-linear interaction between wave energy and friction. Additionally, the Mahakam Delta exhibits flood-dominant tidal asymmetry ($Fe = 0.632 \pm 0.599$) with a mean phase lag of zero minutes between estuary and upstream tidal peaks.

These findings advance our understanding of tropical deltaic hydrodynamics and demonstrate the efficacy of machine learning for complex tidal systems. The integration of temporal lag features with feature importance analysis provides interpretable insights into physical wave propagation mechanisms while maintaining predictive power. The developed models offer practical applications for water resource management, flood forecasting, and navigation safety in the Mahakam Delta and comparable tropical estuarine systems. Future research should incorporate seasonal variations in river discharge, explore deep learning architectures for extended forecasting horizons, and validate model performance across multiple years to capture climate variability impacts. Enhanced understanding and predictive capabilities will support adaptive coastal management

in regions highly vulnerable to tidal flooding and climate change impacts.

Acknowledgements

Thanks to Geospatial Information Agency of Indonesia (BIG) for the open access data, so the authors can use it in this research. We also express our gratitude to who contributed to this research.

Author Contribution

Muhammad Riza conceptualized the research, performed data collection, designed the methodology, supervised the overall study, and led the manuscript writing and revision. Nanda Khoirunisa pre-processing, and analysis, prepared the figures and tables, and contributed to drafting the manuscript. Zetsaona Sihotang supported data acquisition and curation, conducted the literature review, assisted in interpreting the results, and contributed to manuscript editing. Idris Mandang provided scientific supervision, contributed to refining the research design, critically reviewed the results, and offered substantial suggestions during manuscript revision.

Conflict of Interest

The authors declare no conflict of interest.

References

- Anggraini, A. P. K., Trismadi, T., & Supriyadi, A. A. (2024). Predicting Water Levels from Environmental Parameters Using Random Forest Models. *International Journal of Engineering Continuity*, 4(1), 39–53. <https://doi.org/10.58291/ijec.v4i1.314>
- Atashi, V., Gorji, H. T., Shahabi, S. M., Kardan, R., & Lim, Y. H. (2022). Water level forecasting using deep learning time-series analysis: A case study of Red River of the North. *Water*, 14(12), 1971. <https://doi.org/10.3390/w14121971>
- Bergmeir, C., Hyndman, R. J., & Koo, B.

- (2018). A note on the validity of cross-validation for evaluating autoregressive time series prediction. *Computational Statistics & Data Analysis*, 120, 70–83. <https://doi.org/10.1016/j.csda.2017.11.003>
- Cerqueira, V., Torgo, L., & Mozetič, I. (2020). Evaluating time series forecasting models: An empirical study on performance estimation methods. *Machine Learning*, 109(11), 1997–2028. <https://doi.org/10.1007/s10994-020-05910-7>
- de Brye, B., Schellen, S., Sassi, M., Vermeulen, B., Kärnä, T., Deleersnijder, E., & Hoitink, T. (2011). Preliminary results of a finite-element, multi-scale model of the Mahakam Delta (Indonesia). *Ocean Dynamics*, 61(8), 1107–1120. <https://doi.org/10.1007/s10236-011-0410-y>
- Friedrichs, C. T., & Aubrey, D. G. (1994). Tidal propagation in strongly convergent channels. *Journal of Geophysical Research*, 99(C2), 3321–3336. <https://doi.org/10.1029/93JC03219>
- Gan, M., Chen, Y., Pan, S., Lai, X., Pan, H., Wen, Y., & Xia, M. (2024). An improved machine learning-based model to predict estuarine water levels. *Ocean Modelling*, 190, 102376. <https://doi.org/10.1016/j.ocemod.2024.102376>
- Garel, E., Zhang, P., & Cai, H. (2021). Dynamics of fortnightly water level variations along a tide-dominated estuary with negligible river discharge. *Ocean Science*, 17(6), 1605–1621. <https://doi.org/10.5194/os-17-1605-2021>
- Guillou, N., & Chapalain, G. (2021). Machine learning methods applied to sea level predictions in the upper part of a tidal estuary. *Oceanologia*, 63(4), 531–544. <https://doi.org/10.1016/j.oceano.2021.07.003>
- Guo, L., Brand, M., Sanders, B. F., Foufoula-Georgiou, E., & Stein, E. D. (2018). Tidal asymmetry and residual sediment transport in a short tidal basin under sea level rise. *Advances in Water Resources*, 121, 1–8. <https://doi.org/10.1016/j.advwatres.2018.07.012>
- Han, H., Kang, N., Yoon, J., & Hwang, S. (2024). Utilizing deep learning algorithm and high-resolution precipitation product to predict water level variability. *Journal of Korea Water Resources Association*, 57(7), 471–479. <https://doi.org/10.3741/JKWRA.2024.57.7.471>
- Hewamalage, H., Ackermann, K., & Bergmeir, C. (2023). Forecast evaluation for data scientists: Common pitfalls and best practices. *Data Mining and Knowledge Discovery*, 37, 788–832. <https://doi.org/10.1007/s10618-022-00894-5>
- Hoitink, A. J. F., & Jay, D. A. (2016). Tidal river dynamics: Implications for deltas. *Reviews of Geophysics*, 54(1), 240–272. <https://doi.org/10.1002/2015RG000507>
- Hosseinzadeh, P., Nassar, A., Boubrahimi, S. F., & Hamdi, S. M. (2023). ML-based streamflow prediction in the Upper Colorado River Basin using climate variables time series data. *Hydrology*, 10(2), 29. <https://doi.org/10.3390/hydrology10020029>
- Hu, S., Zhang, P., Cai, H., Ou, S., Liu, F., Lin, J., & Yang, Q. (2024). Long-term evolution and controlling factors of tidal duration asymmetry in the mainstream of the Zhujiang River Estuary, China. *Ocean & Coastal Management*, 256, 107289. <https://doi.org/10.1016/j.ocecoaman.2024.107289>

- Iglesias, I., Avilez-Valente, P., Bio, A., & Bastos, L. (2019). Modelling the main hydrodynamic patterns in shallow water estuaries: The Minho case study. *Water*, 11(5), 1040. <https://doi.org/10.3390/w11051040>
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An introduction to statistical learning: With applications in R*. Springer. <https://doi.org/10.1007/978-1-0716-1418-1>
- Jati, S. N., Irawan, D. E., Wijaya, D. D., & Puradimaja, D. J. (2026). Declining volume development index precedes basin infilling in Mahakam Lakes, Indonesia. *Frontiers in Earth Science*, 14, 1818351. <https://doi.org/10.3389/feart.2026.1818351>
- Khojasteh, D., Chen, S., Felder, S., Heimhuber, V., & Glamore, W. (2021). Estuarine tidal range dynamics under rising sea levels. *PLoS ONE*, 16(9), 1–25. <https://doi.org/10.1371/journal.pone.0257538>
- Khojasteh, D., Felder, S., Heimhuber, V., & Glamore, W. (2023). A global assessment of estuarine tidal response to sea level rise. *Science of the Total Environment*, 894, 165011. <https://doi.org/10.1016/j.scitotenv.2023.165011>
- Kim, D., Handayani, M. P., Lee, S., & Lee, J. (2023). Feature Attribution Analysis to Quantify the Impact of Oceanographic and Maneuverability Factors on Vessel Shaft Power Using Explainable Tree-Based Model. *Sensors*, 23(3), 1072. <https://doi.org/10.3390/s23031072>
- Kim, S. I., Seo, J. Y., Park, J.-H., Kim, P. J., Kim, I. G., Kim, K. M., & Ha, H. K. (2024). Imbalance of sediment transport in a meso-tidal bay: effect of tidal pumping and residual circulation. *Frontiers in Marine Science*, 11, 1362583. <https://doi.org/10.3389/fmars.2024.1362583>
- Lim, B., & Zohren, S. (2021). Time-series forecasting with deep learning: a survey. *Philosophical Transactions of the Royal Society A*, 379(2194), 20200209. <https://doi.org/10.1098/rsta.2020.0209>
- Liu, C.-H., Yang, T.-H., & Wijaya, O. T. (2022). Development of an Artificial Neural Network Algorithm Embedded in an On-Site Sensor for Water Level Forecasting. *Sensors*, 22(21), 8532. <https://doi.org/10.3390/s22218532>
- Liu, Z., Lu, Z., Zhu, W., Yuan, J., Cao, Z., Cao, T., Liu, S., Xu, Y., & Zhang, Z. (2025). Comparison of machine learning methods for predicting ground-level ozone pollution in Beijing. *Frontiers in Environmental Science*, 13, 1561794. <https://doi.org/10.3389/fenvs.2025.1561794>
- Lundberg, S. M., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. *31st Conference on Neural Information Processing Systems*, 1–10. <https://doi.org/10.48550/arXiv.1705.07874>
- Nhu, V.-H., Shahabi, H., Nohani, E., Shirzadi, A., Al-Ansari, N., Bahrami, S., Miraki, S., Geertsema, M., & Nguyen, H. (2020). Daily water level prediction of Zrebar Lake (Iran): A comparison between M5P, random forest, random tree and reduced error pruning trees algorithms. *ISPRS International Journal of Geo-Information*, 9(8), 479. <https://doi.org/10.3390/ijgi9080479>
- Nielsen, I., Huismans, Y., & Hoitink, A. J. F. (2024). Ebb-Dominant Mixing Increases the Seaward Sediment Flux in a Stratified Estuary. *Journal of Geophysical Research: Oceans*, 129(12), e2024JC021201. <https://doi.org/10.1029/2024JC021201>
- Palmer, K., Watson, C. S., Hunter, J. R.,

- Hague, B. S., & Power, H. E. (2023). An improved method for computing tidal datums. *Coastal Engineering*, 184, 104354. <https://doi.org/10.1016/j.coastaleng.2023.104354>
- Rasimeng, S., Mandang, I., & S. (2019). Identification of eroded sediment layer thickness and depth of {Mahakam} River at Tenggarong Bridge Area, Kutai Kartanegara East Kalimantan using Ground Penetrating Radar Method. *IOP Conf. Series: Earth and Environmental Science*, 279, 012040. <https://doi.org/10.1088/1755-1315/279/1/012040>
- Sandbach, S. D., Nicholas, A. P., Ashworth, P. J., Best, J. L., Keevil, C. E., Parsons, D. R., Prokocki, E. W., & Simpson, C. J. (2018). Hydrodynamic modelling of tidal-fluvial flows in a large river estuary. *Estuarine, Coastal and Shelf Science*, 212, 176–188. <https://doi.org/10.1016/j.ecss.2018.06.023>
- Sarafanov, M., Kazakov, E., & Borisova, Y. (2020). Modeling water level using downstream river water level observations and machine learning methods. *E3S Web of Conferences*, 163, 01009. <https://doi.org/10.1051/e3sconf/202016301009>
- Sassi, M. G., Hoitink, A. J. F., de Brye, B., Vermeulen, B., & Deleersnijder, E. (2011). Tidal impact on the division of river discharge over distributary channels in the Mahakam Delta. *Ocean Dynamics*, 61(12), 2211–2228. <https://doi.org/10.1007/s10236-011-0473-9>
- Sassi, M. G., & Hoitink, A. J. F. (2013). River flow controls on tides and tide-mean water level profiles in a tidal freshwater river. *Journal of Geophysical Research: Oceans*, 118(9), 4139–4151. <https://doi.org/10.1002/jgrc.20297>
- Spicer, P., Yang, Z., Wang, T., & Deb, M. (2023). Tidal energy extraction modifies tidal asymmetry and transport in a shallow, well-mixed estuary. *Frontiers in Marine Science*, 10, 1268348. <https://doi.org/10.3389/fmars.2023.1268348>
- Su, Y., & Jiang, X. (2023). Prediction of tide level based on variable weight combination of LightGBM and CNN-BiGRU model. *Nature*, 13, 9. <https://doi.org/10.1038/s41598-022-26213-y>
- Sufyan, A., Risandi, J., & Akhwady, R. (2023). Observasi Pasang Surut dan Arus di Muara Porong untuk Mendukung Pengembangan Pulau Lusi: Observation of Tide and Current in Muara Porong to Support Lusi Island Development. *Jurnal Teknologi Lingkungan*, 24(2), 200–206. <https://doi.org/10.55981/jtl.2023.986>
- Tenorio-Fernandez, L., Zavala-Hidalgo, J., & Olvera-Prado, E. R. (2019). Seasonal variations of river and tidal flow interactions in a tropical estuarine system. *Continental Shelf Research*, 188, 103965. <https://doi.org/10.1016/j.csr.2019.103965>
- Van, C. P., Gourgue, O., Sassi, M., Hoitink, A. J. F., Deleersnijder, E., & Soares-Frazão, S. (2016). Modelling fine-grained sediment transport in the Mahakam land–sea continuum, Indonesia. *Journal of Hydro-Environment Research*, 13, 103–120. <https://doi.org/10.1016/j.jher.2015.04.005>
- Wang, J., de Swart, H. E., & Dijkstra, Y. M. (2021). Dependence of tides and river water transport in an estuarine network on river discharge, tidal forcing, geometry and sea level rise. *Continental Shelf Research*, 225, 104476. <https://doi.org/10.1016/j.csr.2021.104476>
- Wu, Y., Hannah, C., Matte, P., O’Flaherty-Sproul, M., Mo, R., Wang, X., & MacAulay, P. (2022). Tidal

- propagation in the Lower Fraser River, British Columbia, Canada. *Estuarine, Coastal and Shelf Science*, 264, 107695.
<https://doi.org/10.1016/j.ecss.2021.107695>
- Wünsche, A., Becker, M., Fritzsche, R., Kelln, J., & Winter, C. (2024). The sensitivity of tidal asymmetry descriptors in the Ems estuary. *Ocean Dynamics*, 74, 613–627.
<https://doi.org/10.1007/s10236-024-01622-x>
- Xie, D., & Wang, Z. B. (2021). Seasonal Tidal Dynamics in the Qiantang Estuary: The Importance of Morphological Evolution. *Frontiers in Earth Science*, 9, 782640.
<https://doi.org/10.3389/feart.2021.782640>
- Yu, X., Zhang, W., & Hoitink, A. J. F. (2020). Impact of river discharge seasonality change on tidal duration asymmetry in the Yangtze River Estuary. *Scientific Reports*, 10, 6304.
<https://doi.org/10.1038/s41598-020-62432-x>
- Zhang, Z., Zhang, L., Yue, S., Wu, J., & Guo, F. (2023). Correction of nonstationary tidal prediction using deep-learning neural network models in tidal estuaries and rivers. *Journal of Hydrology*, 622, 129686.
<https://doi.org/10.1016/j.jhydrol.2023.129686>
- Zhou, J., Chen, L., Hu, T., Lu, H., Shi, Y., & Chen, L. (2025). The comparative study of machine learning agent models in flood forecasting for tidal river reaches. *Scientific Reports*, 15, 19130.
<https://doi.org/10.1038/s41598-025-04633-w>
- Zounemat-Kermani, M., Matta, E., Cominola, A., Xia, X., Zhang, Q., Liang, Q., & Hinkelmann, R. (2020). Neurocomputing in surface water hydrology and hydraulics: A review of two decades retrospective, current status and future prospects. *Journal of Hydrology*, 588, 125085.
<https://doi.org/10.1016/j.jhydrol.2020.125085>