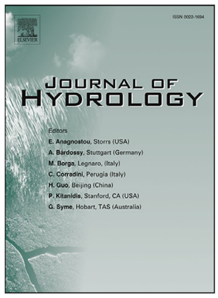
Journal Pre-proof



Regression-based machine learning approaches for estimating discharge from water levels in microtidal rivers

Anna Maria Mihel, Nino Krvavica, Jonatan Lerga

PII: S0022-1694(24)01672-X
DOI: https://doi.org/10.1016/j.jhydrol.2024.132276
Reference: HYDROL 132276

To appear in: *Journal of Hydrology*

Received date : 19 July 2024
Revised date : 13 October 2024
Accepted date : 16 October 2024

Please cite this article as: A.M. Mihel, N. Krvavica and J. Lerga, Regression-based machine learning approaches for estimating discharge from water levels in microtidal rivers. *Journal of Hydrology* (2024), doi: https://doi.org/10.1016/j.jhydrol.2024.132276.

Highlights

Regression-based Machine Learning Approaches for Estimating Discharge from Water Levels in Microtidal Rivers

Anna Maria Mihel,Nino Krvavica,Jonatan Lerga

- First comparison of ML models for reconstruction of discharge in microtidal rivers.
- LSTM-based models accurately estimate river discharge using only water levels from multiple stations.
- LSTM-Attention model can predict river discharge under all flow conditions even for unbalanced inputs.
- Using ML we can reconstructs discharge for data-limited tidal rivers.

Regression-based Machine Learning Approaches for Estimating Discharge from Water Levels in Microtidal Rivers

Anna Maria Mihel[a,b], Nino Krvavica[b,c] and Jonatan Lerga[a,b,*]

[a]Department of Computer Engineering, Faculty of Engineering, University of Rijeka, Vukovarska 58, 51000 Rijeka, Croatia
[b]Center for Artificial Intelligence and Cybersecurity, University of Rijeka, Radmile Matejcic 2, 51000 Rijeka, Croatia
[c]Department of Hydrologic Engineering, Faculty of Civil Engineering, University of Rijeka, Radmile Matejcic 3, 51000 Rijeka, Croatia

ARTICLE INFO

Keywords:
machine learning
discharge
regression
microtidal river
STREAM 1D
LSTM

ABSTRACT

The challenges of managing water resources in tidal rivers, exacerbated by climate change and anthropogenic impacts, require innovative approaches for accurate estimation of hydrological parameters. In tidal rivers and estuaries, water levels depend primarily on river discharge and tidal dynamics. Microtidal estuaries are particularly complex due to the strong stratification and two-layer structure, which also affect the water level. This study investigates the potential of machine learning (ML) models for estimating discharge in the Neretva River, Croatia, using only water level data from multiple stations. Comparative analyzes were performed between simple supervised models - Decision Tree (DT), Random Forest (RF), Support Vector Regression (SVR), Light Gradient Boosting Machine (LGBM), Extreme Gradient Boosting (XGB) - and time series models - Long Short-Term Memory (LSTM) and LSTM-Attention. Both simulated and measured data sets were used for this purpose. The results show that time series models perform satisfactorily in the assessment of discharge and overcome the challenges faced by simple supervised models, especially under high flow scenarios. Overall, LSTM-Attention proves to be the best model when analyzing all error metrics with superior performance over the entire range of discharge values. It surpasses the overall LSTM model performance, with a percentage increase of above 9% in RMSE and MAE, above 0.2% in NSE, and above 0.1% in R for both simulated and measured datasets.

1. Introduction

Tidal rivers and estuaries are transitional zones where freshwater comes into contact with the marine environment (Cai et al., 2015). These environments are characterized by complex flows involving the exchange of energy and materials (Chen et al., 2023; Du et al., 2023). Tides, influenced by the gravitational forces of the sun and moon and Earth's rotation, significantly affect water levels and flow dynamics in coastal regions (Chen et al., 2023).

Additionally, climate factors such as wind, temperature, air pressure, and precipitation play a crucial role, contributing to extreme weather events like storm surges, typhoons, and hurricanes (Chen et al., 2022). Anthropogenic activities, including modifications to floodplains, channels, urbanization, and agriculture, further impact tidal river dynamics (Vercruysse and Grabowski, 2021; van Maren et al., 2023). Recent studies have also emphasized the significant influence of wastewater on water quality and discharge volume (Dutta et al., 2021; Kundu et al., 2022; Panagopoulos and Giannika, 2024). Given these complexities, effective water management is essential for providing

annamaria.mihel@riteh.uniri.hr (A.M. Mihel); nino.krvavica@uniri.hr (N. Krvavica); jonatan.lerga@riteh.uniri.hr (J. Lerga)
ORCID(s): 0000-0002-3697-9471 (A.M. Mihel); 0000-0001-5014-5476 (N. Krvavica); 0000-0002-4058-8449 (J. Lerga)

timely information on environmental and flood risks, enabling informed decision-making and adaptation measures in tidal rivers and estuaries.

Effective water management depends on hydrologic monitoring of both water level and river discharge. While measuring water levels is relatively simple and cost-effective, discharge measurements are often limited. Continuous monitoring of river discharge is usually based on establishing relationships between water level and discharge or on direct measurements of water velocity using radar sensors or acoustic Doppler current profilers (ADCP). Unlike inland rivers, where discharge can be determined directly from known water levels, tidal rivers pose a particular challenge. In tidal environments, water levels are influenced by the river flow, storm surges and tidal dynamics, as well as their non-linear interactions (Hidayat et al., 2014; Wolfs and Willems, 2014). It is therefore difficult to establish reliable relationship between water level and discharge in tidal rivers (Habib and Meselhe, 2006; Lee et al., 2021).

Radar sensors only measure surface water velocity and are therefore not suitable for tidal rivers due to the nonuniform velocity profiles resulting from the complex flow patterns, including bidirectional and return flows. ADCPs can capture velocity profiles, but their implementation can be technically complex and expensive, especially in regions with limited financial resources or difficult field conditions, including harsh weather conditions (Habib and Meselhe, 2006; Lee et al., 2021; Thanh et al., 2022). In addition, while the installation of ADCP stations provides velocity profiles and reliable discharges, it does not adress the need for long data sets. In cases where such devices have only recently been installed, or where data are missing, the only viable solution is discharge reconstruction. Reconstruction in tidal rivers can be done using numerical modeling or data-driven approaches, usually based on artificial intelligence (AI) (Lee et al., 2021; Thanh et al., 2022).

Discharge estimation has gained considerable attention over the past decade, driven by the increasing frequency of extreme events due to climate change. However, most studies in this field focus primarily on forecasting, as highlighted in several recent review papers. These studies cover a broad range of topics, from general perspectives on artificial intelligence (AI) (Yaseen et al., 2015) to optimization and hybrid modeling approaches (Ibrahim et al., 2022; Ng et al., 2023). In particular, they emphasize the effectiveness of Long Short-Term Memory (LSTM) networks, whether as part of hybrid models (Ng et al., 2023; Khatun et al., 2023; Sabzipour et al., 2023; Mohanty et al., 2024), stand-alone models with improved accuracy (Li et al., 2023), or even for transfer learning applications (Khoshkalam et al., 2023).

However, current methods for estimating discharge from water levels in tidal rivers and estuaries using artificial intelligence (AI) are limited by a lack of comprehensive studies. A recent review by Mihel et al. (2024) addressed this gap, examining the use of machine learning (ML) in tidal environments for forecasting and reconstructing water levels and river discharges. The review identified research gaps, along with the strengths and limitations of ML models, and noted the surprisingly low number of studies investigating these challenges. This review highlighted advanced Recursive Neural Networks (RNNs) and hybrid models as promising approaches for modeling complex, nonstationary, and nonlinear time series data, especially for capturing long-term dependencies. The study also proposed several future research directions, including hybrid models combining LSTM and attention mechanisms. Building on the findings of this comprehensive review of machine learning in tidal rivers and estuaries (Mihel et al., 2024), we present an overview of the most significant contributions in this field.

One of the first attempts to establish a relationship between water level and discharge in a tidal river was made by Habib and Meselhe (2006). The authors focused on the low-gradient Isaac-Verot Coulee tidal river in southwest Louisiana. Using artificial neural networks (ANN) and Loess regression methods, the study included multiple water level monitoring stations, accounting for the backwater effect and lagged stages. The ANN model showed a better generalization ability compared to the Loess model, especially at higher discharge values.

Hidayat et al. (2014) conducted a study in which a multilayer perceptron (MLP) in combination with the Levenberg-Marquardt optimization algorithm was used to hindcast and forecast the discharge of the Mahakam River in Indonesia. The study faced challenges in extreme discharge conditions that affected the reliability of the model, which relied on multiple water level stations in the hindcasting scenario and a single upstream location in the forecasting scenario.

Similarly, Wolfs and Willems (2014) investigated the relationship between water level and discharge in two Belgian rivers, the Marke and the Dender. The study compared four approaches: single and state-dependent parameter rating curves (SRC and SDP-RC), ANN and M5' model trees. The SDP-RC was preferred despite the complexity of the model, as it best represents the complex behavior of river hysteresis.

A recent paper by Thanh et al. (2022) addressed the problem of discharge reconstruction in the Mekong megadelta in Vietnam. Using observations from two upstream stations and discharge data from one downstream station, six models were evaluated: M5' Decision Tree, Random Forest (RF), Support Vector Regression (SVR), Least Squares Support Vector Regression (LSSVR), Gaussian Process Regression and Multivariate Adaptive Regression Spline (GPR) and Multivariate Adaptive Regression Spline (MARS). RF and MARS showed superior performance, especially for extreme events, while the results of GPR, LSSVR and SVR were considered only adequate. The DT model was rejected due to atypical oscillations.

The newest study by Vu et al. (2023) explored the area of the Loire-Bretagne river system, incorporating 18 stations as forecasting targets. However, besides including the main river stations (five of them), they also considered tributary stations and stations that are not connected to the Loire river system, such as Bretagne sub-basin rivers stations, and single station that belongs to the Charente River. A stacked LSTM was tested, which included both hydrological and meteorological data variables sampled daily.

Despite the rapid progress of artificial intelligence and its growing popularity in hydrologic engineering (Zounemat-Kermani et al., 2020), the application of AI for discharge estimation in tidal rivers has not yet been sufficiently investigated (Mihel et al., 2024). This study aims to fill this gap by evaluating the efficiency of different machine learning (ML) methods for discharge estimation in a microtidal Neretva River estuary (Croatia) using water level data from multiple stations.

This study is the first to apply various ML models to estimate discharge from water levels in a microtidal river with strong stratification and two-layered saltwedge profile, which significantly affect discharge dynamics (Krvavica et al., 2021). Eight ML models, from simple supervised algorithms to complex time-dependent models, were tested and compared using both measured and simulated data, which is a novel approach in this field. Motivated by the limited use of ML in tidal river discharge estimation, as highlighted by a recent review (Mihel et al., 2024), this study introduces time-dependant ML models for the first time under microtidal conditions. Building on previous findings, we assess two approaches: an advanced recursive neural network (RNN) and a hybrid model. LSTM was chosen for its proven accuracy, efficiency, and robustness, while a novel hybrid model combining LSTM and the attention mechanism is introduced to enhance predictive accuracy by identifying critical features. The models' performance was tested under various flow conditions and seasons, characterized by both quasi-steady and oscillatory patterns.

The structure of this paper is as follows: The first section contains a brief literature review and introduces the research topic. The second section describes the research area, data collection, and simulation methods. The third section gives an overview of the ML models and their underlying principles. The fourth section describes the methodology, including the steps for data processing and model construction. Section five presents the results based on statistical and graphical analyses, followed by a discussion in section six. The final section contains conclusions and suggestions for future improvements.

2. Data and Study Site

This section gives a brief overview of the Neretva River, focusing on the lower part of the river and highlighting the main hydrological issues. It also describes the data collection process for the measured and simulated data sets.

2.1. Neretva River Estuary

The Neretva River, the largest river on the eastern Adriatic coast, flows through Bosnia and Herzegovina and Croatia. Our study focuses on its alluvial delta, which is classified as a saltwedge estuary (Krvavica and Ružić, 2020). Climate change has a significant impact on this region, as the river plays a crucial role for agriculture and irrigation, but is also an important source of flooding (Gajić-Čapka et al., 2018). The Neretva River has a typical seasonal regime characterized by a low flow from May to September and a high flow from October to April (Krvavica et al., 2021). The Mediterranean climate in the region, which is favorable for fruit and vegetable cultivation, poses a threat to the agriculture as a results of more frequent saltwater intrusion, which affects plant growth, especially in the summer. Factors such as reduced river inflows due to insufficient rainfall and the dynamics of the coastal aquifer exacerbate the problem of salinization (Zovko et al., 2018; Lovrinović et al., 2023). On the other hand, winter and early spring seasons are associated with flood events, that can be exacerbated by storm surges and high sea levels, resulting in increased damages.

The study area covers the last 23 km of the Neretva River in Croatia, which is exposed to the influences of hydropower plants upstream and tidal dynamics downstream due to its proximity to the sea (Figure 1). The influence of upstream inflow refers to several hydropower plants and man-made structures such as dams and reservoirs, whose construction has significantly altered the natural hydrological regime of the river (Ljubenkov and Vranješ, 2012). Furthermore, salinity intrusion is particularly evident due to sea level rise and reduced freshwater inflow (Krvavica et al., 2021; Lovrinović et al., 2023).

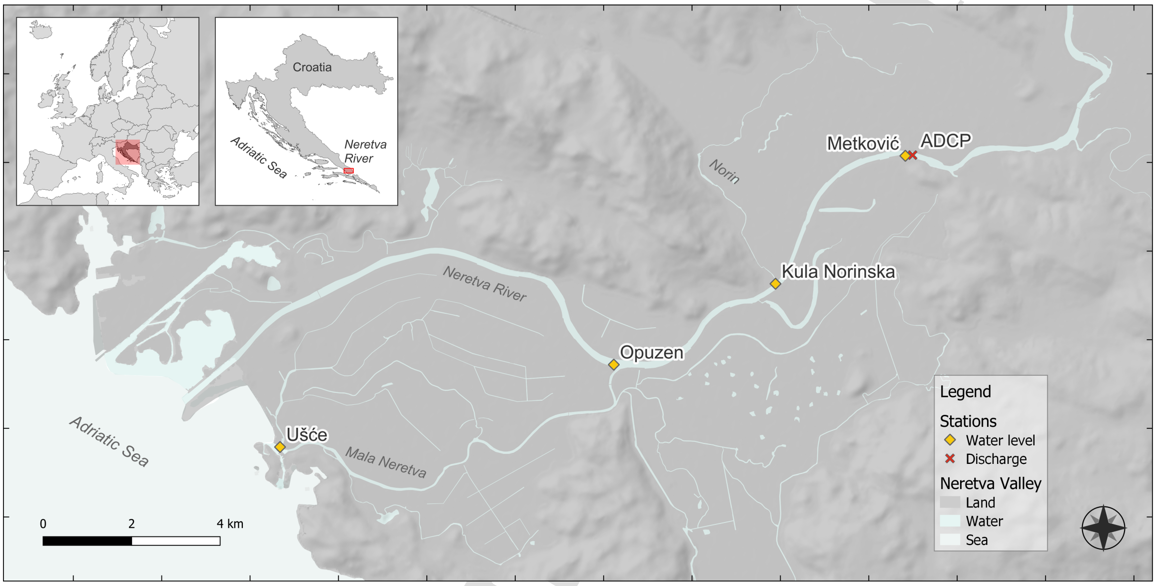


Figure 1: The Neretva River Estuary location, with water level and discharge stations.

Due to the microtidal nature of the Adriatic Sea, tidal amplitudes are low, with storm surges regularly exceeding the daily amplitude ranges. The highest recorded sea level at the Neretva River mouth (in the period 1977-2018) was 120 cm a.s.l., and the lowest -25 cm a.s.l. (Krvavica et al., 2021). The mean astronomical tidal amplitudes in the Adriatic Sea, however,do not exceed 30 cm (Medvedev et al., 2020). The Neretva River estuary, which is characterized by strong stratification throughout the year, has a pycnocline thickness of less than 50 cm (Krvavica et al., 2021). Such strong stratification is a result of the extremely low tidal dynamics of the Adriatic Sea.

2.2. Measured Data

The measurements of hydrological parameters were carried out by the Croatian Meteorological and Hydrological Service and Croatian Waters. These include sea levels at the tidal station Ušće, as well as water levels at hydrological stations Opuzen (11 rkm), Norin (16 rkm) and Metković (20 rkm), River discharge at Metković is measured with Horizontal Acoustic Doppler Current Profiler (H-ADCP) devices. Figure 1 shows the water level stations and a single discharge station (ADCP). The data set spans six years (2016-2021) with hourly intervals.

Discharge monitoring in the Neretva River estuary has only recently been implemented (since the end of 2015). The discharges are continuously measured with three H-ADCP devices installed under the bridge in Metković (Krvavica et al., 2021). The H-ADCP devices estimate the discharge by integrating the velocity profile over three cross-sectional areas in each bridge opening. The total discharge is calculated by adding the values measured by the individual devices. The data set (2016-2021) covers a wide range of hydrological conditions, from negative discharges (tidal currents and low river flow) in summer to peak winter discharges of 1890 m^3/s, with an average annual discharge of 323 m^3/s. The maximum water levels in Metković reached 2.25 m a.s.l during the observed period. High water levels are the result of several factors: high sea level and high river flow, the operation of hydropower plants upstream and the interaction between surface and subsurface flow and runoff during extreme rainfall.

The relationship between discharge and water level at the Metković station is shown in Figure 2. A relatively strong correlation between the two hydrological parameters is observed, but with a noticeable dispersion of points around the mean. Therefore, the discharge rating curve is not suitable here. Consequently, our hypothesis is that the use of simple and advanced ML models can provide valuable insights into the dynamics of the river and reveal the relationship between discharge and water level at different stations.

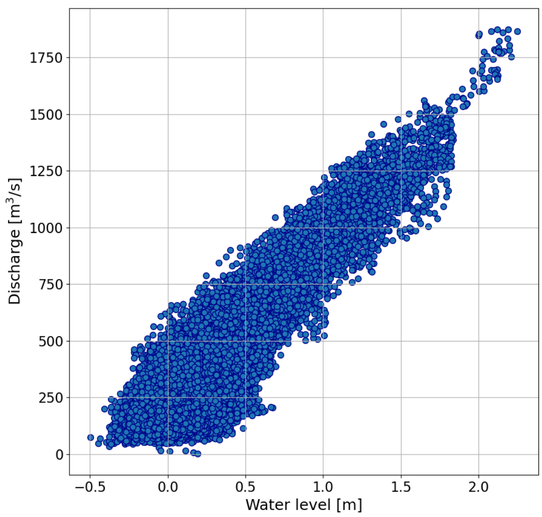


Figure 2: Stage-discharge relationship for the Neretva River in Metković based on hourly data.

Although additional measurements are regularly taken at the Metković station to ensure proper calibration of the H-ADCP devices, several issues affect the quality of the discharge time series data. Problems such as missing data, nonphysical oscillations, and the complexity of low water flow affect the quality of the discharge time series data. Discrepancies in synchronization can also arise due to data collection and processing by different institutions.

To solve the first problem, we filled the gaps in the time series by establishing a non-linear correlation between the three H-ADCP devices. Other two problems can be minimized by careful noise filtering. In this study, however, we chose to perform numerical simulations and evaluate the selected ML methods on both measured and simulated data. A rationale behind this choice is that we need a controlled data set that accurately represents the main hydraulic processes without being masked by gaps, errors, noise and potential time shifts that occur in heterogeneous data sets. This approach ensures a more reliable and robust evaluation of the ML techniques under investigation.

2.3. Simulated Data

The simulated water levels at the stations are generated using the STRatified EstuArine Model (STREAM), a one-dimensional, time-dependent numerical model developed specifically for microtidal estuaries (Krvavica et al., 2017). STREAM has already shown good performance in modeling the two-layer flow dynamics in the Rječina and Neretva rivers (Krvavica et al., 2017, 2021; Krvavica and Ružić, 2020). This approach, which is characterized by the computational efficiency of 1D shallow water models, proves to be more appropriate than its 3D counterparts. It is simple, but at the same time effectively captures the dominant hydraulic processes that occur in the two-layer flow in saltwedge estuaries.

The model domain is defined by the channel geometry generated from the cross-sections of the Neretva estuary, extending from the river mouth to 35 km upstream, exceeding the tidal limit. At the downstream boundary, a time series of the total water level is set based on the measured sea levels at the Ušće station, while the interface between the upper and lower layers is defined based on a critical two-layer flow condition. A time series of river discharge is set at the upstream boundary. It is important to note that the discharge time series is subjected to a two-stage processing. First, the values are shifted by one hour to account for the distance of 15 km between the Metković station and the upstream boundary. Then, a median filter with a 3-hour window is applied to eliminate high-frequency noise and single value errors.

The friction coefficients are calibrated by minimizing the error between the simulated and measured data sets. In contrast to single-layer shallow water equations, the calibration process involves not only the determination of the optimal friction coefficient between the fluid and the riverbed, but also the determination of the interfacial friction coefficient between two fluids of different densities. First, the river bed friction is calibrated using only high flow conditions (predominantly single-layer case). Next, the interfacial friction is calibrated for the entire data set, which also includes the conductivity measurements at the Metković station. More details about the calibration, setup and efficiency of the model can be found in the earlier study by Krvavica et al. (2021). After calibration, the simulated data set shows good agreement with the measured water levels at the Opuzen, Kula Norinska, and Metković stations.

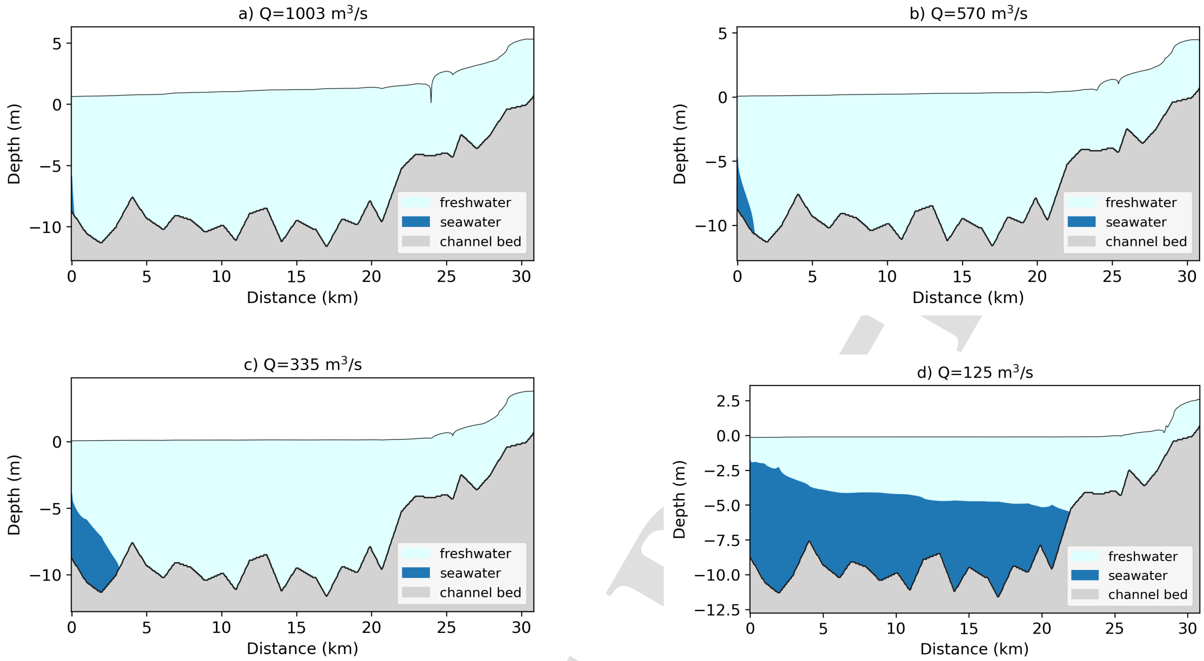
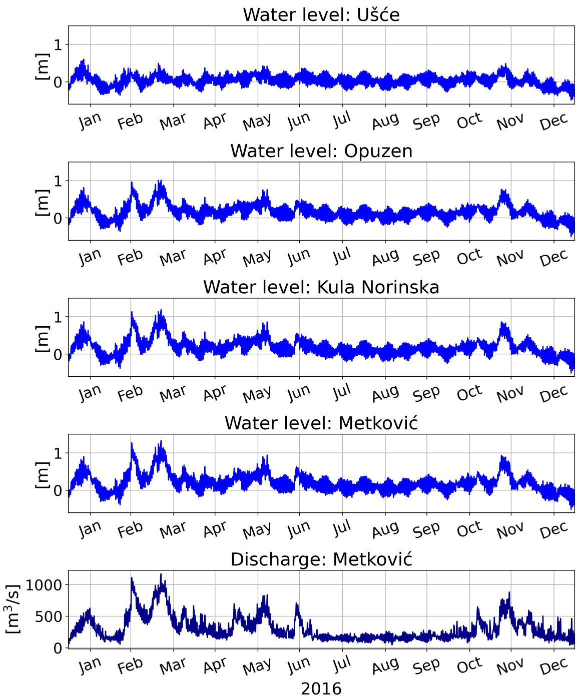


Figure 3: Longitudinal profile of the River Neretva water levels for river discharges: a) 1003 m^3/s, b) 570 m^3/s, c) 335 m^3/s, d) 125 m^3/s

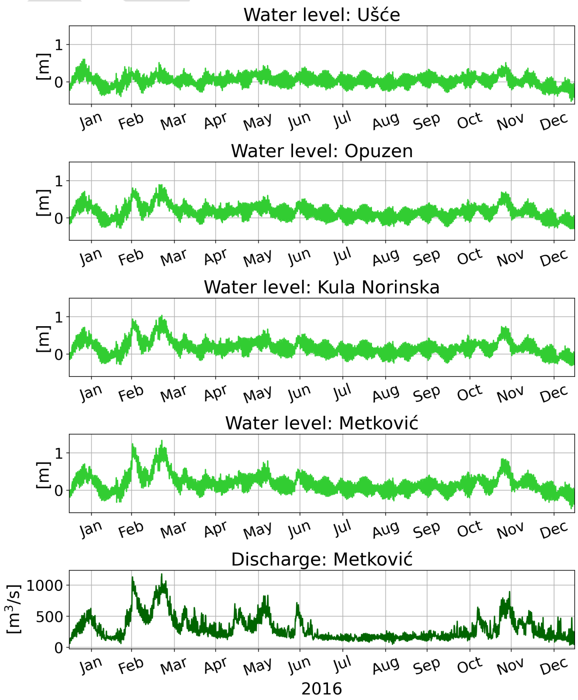
Figure 3 illustrates the salinity profiles for different river discharges. The simulations show the typical behavior of the saltwedge at high to low discharges and justify the use of the numerical model and its ability to represent the complex dynamics of the estuary. At high flows, the saltwedge is completely flushed out of the estuary and the water surface has a relatively steep gradient along the entire river channel. At high to medium flows, the saltwedge is confined to the river mouth and the water surface has a slightly lower gradient along the entire river channel. At medium to low flows, the saltwedge begins to intrude upstream, and the water surface has a very low gradient for the first 20 kilometers. Finally, at low flows, the saltwedge intrudes upstream beyond the Metković station, resulting in a horizontal water surface along the entire length of the saltwedge.

Estimating river discharge from water levels in tidal rivers and estuaries it particularly challenging due to the complex intercation between sea level and river flow, along with their non-linear interactions. Sea levels are influenced by a combination of tides and storm surges, and these factors interact with river flow to produce variable water levels at a given station (Matte et al., 2014). In microtidal estuaries, where tidal ranges are small, the complexity increases further because of the strong stratification and the presence of a two-layer flow structure. In these conditions, the water column consists of an upper freshwater layer and a lower salt-wedge layer (Krvavica et al., 2021). Along the reach where this two-layer structure is present, the total water level tends to remain nearly constant, regardless of changes in river discharge (see Figure 3). This is because variations in river flow primarily affect the depth of the lower saltwater wedge rather than the overall water level. Consequently, during periods of low river flow, a wide range of discharges can occur without significant changes in surface water level. This phenomenon makes it particularly challenging to estimate discharge from water levels alone in microtidal estuaries, as traditional methods based on a single water level may not capture the underlying changes in flow conditions.

The simulated and measured data sets cover a continuous period of six years (2016-2021) at hourly intervals. The simulated data set distinguishes between discharge values for the lower saltwater and upper freshwater layers. Therefore, the total discharge at the Metković station is a result of the summation of these two values. We should note that the approach presented in this study could be used to estimate only freshwater discharge, which is more relevant information for water management; however, to be consistent with the measured data, we chose to use the total discharge. Figure 4 shows the time series of water level and discharge at four locations in 2016.



(a) Simulated



(b) Measured

Figure 4: Water levels and discharge data for a period of one year (2016): a) simulated and b) measured

3. Brief Overview of ML Models

In the current study, eight different ML models were implemented, namely:

- Decision tree (DT),
- Random forest (RF),
- Support vector regression (SVR) with radial basis and sigmoid function,
- Light gradient boosting machine (LGBM),
- Extreme gradient boosting (XGB),
- Long short-term memory (LSTM),
- LSTM-Attention.

The section provides a brief theoretical background of each ML model.

3.1. Decision Tree (DT)

The decision tree (DT), a non-parametric method, belongs to the supervised learning approach (Hannan and Anmala, 2021), which can be applied to regression and classification problems (Sattari et al., 2020) depending on the type of dependent variable. The DT method can be described as an easy-to-interpret method that provides satisfactory accuracy (Thanh et al., 2022), relatively fast execution time and good short-term prediction performance (Malek et al., 2022) and is therefore frequently used in studies.

In this paper, an optimized classification and regression tree (CART) model is used for the problem of regression, in which the dependent value is predicted based on multiple independent variables. The regression approach differs from classification in that it does not generate the classes of dependent variables but the response value for each new observation with respect to the dependent variable, where the splitting of the trees is based on the squared residual minimization principle (Choubin et al., 2018). However, this approach may encounter some problems, such as overfitting (which can be solved with the pruning technique whose purpose is to reduce the tree size) and linear regression loss.

3.2. Random Forest (RF)

Random Forest (RF) is an extension of the previously mentioned DT approach and, like DT, can be used for both regression and classification problems. However, unlike DT, it falls into the category of ensemble methods, as in this particular case it consists of multiple regression decision trees whose prediction results are combined and averaged to provide the final estimates (Huang et al., 2023).

The selected RF implementation includes a bagging method (bootstrap aggregation) that solves the problem of overfitting while providing higher stability and variance reduction (Malek et al., 2022). In addition to the bagging method, RF also utilizes feature randomness and binary recursive partitioning to create each decision tree in a forest. This gives the created trees their independence, ability to deal with missing values, and other advantages. One of the problems with RF is that the training data set does not contain the values that the model predicts on the unseen data set (Guillou et al., 2023). Three parameters must be defined for the construction of RF models (Li et al., 2016). These parameters are the number of regression DTs, the randomly selected independent variables at the nodes and the minimum observations required at the end node of each tree.

3.3. Support Vector Regression (SVR)

Support Vector Regression (SVR) is an adapted variant of the original Support Vector Machines (SVM), which are kernel-based and were proposed by Vapnik (Vapnik, 2000), with the primary goal of solving classification problems (Guillou et al., 2023). The SVR approach aims to find the optimal hyperplane for the data based on the predefined error or threshold that can be considered acceptable by such a model. In this work, a ϵ-SVR was chosen as it considers the range of error insensitivity.As already mentioned, this method is kernel-based, which means that the data is transformed into a higher dimension to perform the separation (Guillou et al., 2023). Here, two different SVR kernel functions are used and tested for nonlinear category problems: the radial basis function and the sigmoid function.

3.4. Light Gradient Boosting Machine (LGBM)

Light Gradient Boosting Machine (LGBM), published in 2017 (Ke et al., 2017), is based on a DT algorithm that uses leaf-wise tree growth to increase the training speed (Gan et al., 2021) (Gan et al., 2021) and enables parallel training that leads to efficient tree growth (Tian et al., 2022). The above model can solve various problems such as regression and classification (e.g. binary, multiclass, and lambda) (Tran et al., 2021). The tree is built depending on the node that can provide the largest error reduction, which characterizes it as more greedy than the standard approach in gradient boosting methods of growing trees in stages. In addition to leaf-wise tree growth, LGBM is also known for its use of gradient-based one-sided sampling (GOSS), histogram-based algorithm, and exclusive feature bundling (EFB). By using such methods, the LGBM can reduce the probability of overfitting and boost computational efficiency.

3.5. Extreme Gradient Boosting (XGB)

Extreme Gradient Boosting (XGB) is another ensemble ML model that is based on DTs and simultaneously incorporates techniques such as tree pruning and regularisation with the aim of overfitting prevention (Piraei et al., 2023). More precisely, it is a further development of the model that was already developed in 2001 by Friedman. In contrast to the LGBM, XGB is based on a level-wise tree growth process in which the DTs are consistently constructed equally for each level in order to reduce the estimation error of the previous stage. This is quite efficient when a small data set is available and leads to more stable results than it would be possible with the LGBM model (Gan et al., 2021). XGB enables CPU multithreading parallelization, provides efficient and fast processing of large data sets, and integrates block technology (Piraei et al., 2023).

3.6. Long Short-Term Memory (LSTM)

Due to the previously encountered recurrent neural network (RNN) restriction regarding the vanishing gradient, a more efficient approach, named the Long Short-Term Memory (LSTM), was introduced by Hochreiter and Schmidhuber (1997). LSTM represents an improved version of the RNN that was able to overcome the previous limitations and proved to be even more efficient for sequential data processing, leading to impressive results in various fields (Sherstinsky, 2020).

Each LSTM cell consists of three gates, namely input (i_t), forget (f_t) and output (o_t), and it also contains the internal cell state (c_t) representing its memory, the candidate state ($\tilde{c}_t$), a hidden state (h_t) and two activation functions: the sigmoid and the hyperbolic tangent. The above gating mechanisms in the LSTM allow the network to selectively control the information flow as they act as switches that can be turned on or off based on the input data (Yoo et al., 2020), replacing the typically used activation functions (Lindemann et al., 2021). All expressions for the gates and the states are shown in Eq. 1 - 6.

$$f_t = \sigma\left(W_f \cdot \left[h_{t-1}, x_t\right] + b_f\right) \tag{1}$$

$$i_t = \sigma\left(W_i \cdot \left[h_{t-1}, x_t\right] + b_i\right) \tag{2}$$

$$\tilde{c}_t = \tanh\left(W_n \cdot \left[h_{t-1}, x_t\right] + b_n\right) \tag{3}$$

$$c_t = f_t \cdot c_{t-1} + i_t \cdot \tilde{c}_t \tag{4}$$

$$o_t = \sigma\left(W_o \cdot \left[h_{t-1}, x_t\right] + b_o\right) \tag{5}$$

$$h_t = o_t \cdot \tanh(c_t) \tag{6}$$

where W_f, W_i, W_o and W_n are the weights, b_f, b_i, b_o and b_n are the biases of the forget, input, output gate and candidate cell, $sigma$ and $\tanh$ are the activation functions.

3.7. LSTM-Attention

A hybrid LSTM-Attention model was also used in the study as an improved time series technique. This model combines the advantages of the LSTM architecture, which include capturing long-term dependencies and patterns, with the advantages of the attention mechanism, which allows the model to focus on the most important parts of the input sequence. The purpose of the LSTM in the model is to recognize the correlation between the time steps of different features and select those that can be considered relevant and from which an accurate estimate can be obtained.

The attention mechanism works on the principle of determining the importance of each used feature in the data set, i.e. its weight, for each time step in order to provide an appropriate estimate of the output feature at the specified time step. This type of attention mechanism has some similar properties to the global attention mechanism, where the entire input sequence is processed, just like the LSTM, to generate weights for each element in the sequence, since the entire sequence is of importance, even though it requires more computational resources, unlike local attention (Luong et al., 2015). However, our mechanism is based on the hidden states that are generated by the LSTM and for which the attention weights are generated; therefore, this mechanism can be categorized as content-based attention. Two activation functions are used in the attention mechanism: the hyperbolic tangent (tanh), which adds non-linear properties, i.e. it is useful in modeling complex relationships as in the problem at hand, and the *softmax* activation function, which generates a probability distribution of values to be able to provide emphasis on the relevant information of a given sequence. Finally, the obtained weighted sum of hidden states is passed to a single feed-forward layer to generate the final discharge estimates.

4. Methodology

4.1. Data Processing

Data processing consisted of lag correction, splitting and normalisation. A cross-correlation has been performed between the input and output variables to account for the lag between the water levels at different stations and river discharge. A two-hour time delay was found between all water level stations and upstream discharge, thus the time series were shifted accordingly. Cross-correlation results are shown in Figures A.1 and A.2 of Appendix A.

The data set is divided into a training data set (80% of observations, from January 2016 to October 2020) and a test data set (20% of observations, from November 2020 until December 2021). The training data set is used for the learning process, while the test data set is used to evaluate the performance of the models on unseen data. Before training the model, the input (water levels) and output (discharge) data were scaled to a range between 0 and 1 using *MinMaxScaler* from *scikit-learn* (an ML library for Python). This normalization process counteracts the differences in the variable ranges and preserves the primary data distribution. This approach is based on the following mathematical expression:

$$\hat{x} = \frac{x_i - x_{min}}{x_{max} - x_{min}} \tag{7}$$

which requires three pieces of information, the current observation x_i, the overall minimal x_{min} and maximal x_{max} value of a variable.

Unlike models that do not depend on the order of the data and the interaction between the variables, especially non-time series models (DT, RF, SVR, LGBM, and XGB), where shuffling the data does not improve performance, time series models require an appropriate ordering of the data in time. We selected two time series models, LSTM and LSTM-Attention, which capture the temporal dependencies through the sequential data. However, for the models to capture these dependencies effectively, they require a certain structure of the input data. Therefore, a sliding 24-hour window for the water level data is used as input for predicting the discharge of the last hour as it covers one full cycle of high and low tide, thereby, the daily river flow dynamic is successfully captured in 24 hours of historical records. This is due to the fact that the Adriatic Sea has a mixed tidal signal, with equally strong diurnal and semi-diurnal constituents. Hence, the 24-hour window is sufficient for capturing both types of oscillation.

4.2. Model Training and Optimization of Hyperparameters

Models were trained on 80% of each data set using five-fold cross-validation with grid search to optimize hyperparameters. The chosen number of equally separated k-folds was five due to the tradeoff between bias and variance and the moderate size of the available data set. This approach minimizes overfitting as each hyperparameter combination is evaluated in a separate fold during training.

The simple principle of k-fold cross-validation is shown in Figure 5. The first step of this method is to determine the number of folds with which the models are first trained and then tested. For this particular problem, we used a five-fold split, where the first four folds are used to train and the last fold is used to test the performance of the model. The overall result is obtained by taking the average of all the divisions created.

The only obstacle to using k-fold cross-validation and the search network approach is that they are not available for Pytorch models. Nevertheless, the library *skorch* facilitates the use of wrappers for Pytorch models, such as *NeuralNetRegressor*. This library acts as an intermediary between Pytorch models and the *scikit* library. It enables the use of existing training, optimization and evaluation functions instead of developing new functions for these purposes. The main metric for refitting the models was MSE, continued by other scoring metrics such as RMSE, MAE, NSE, and R.

The next step was to select a suitable optimization algorithm. The Adam optimization algorithm is often used in combination with LSTM due to its many advantages (Ahmed et al., 2022). For the current problem, however, a different Adam variant is used, namely Nesterov-accelerated adaptive moment estimation (Nadam). Nadam is a hybrid optimization algorithm that uses both the Adam optimizer and the Nesterov momentum to achieve faster convergence (Villeneuve et al., 2023). In addition to the chosen optimization algorithm, a regularization technique for early stopping

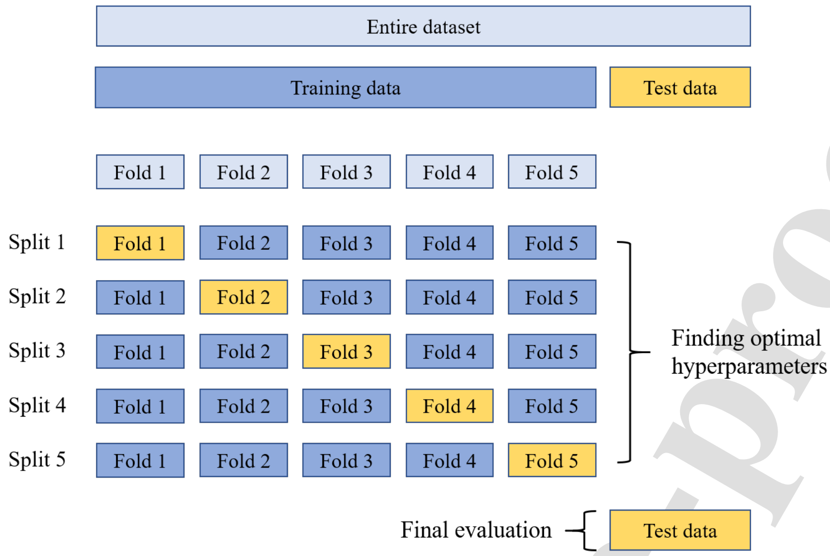


Figure 5: Principle of k-fold cross-validation

was used, which primarily aims to prevent overfitting of the model. The selected number of epochs before the model stops learning, i.e., the validation loss or MSE stagnation, was equal to 15. This is considered essential in cross-validation (Almeida, 2002) as it prevents critical errors, i.e. inaccurate estimates. We have chosen 500 as the number of epochs.

4.2.1. Performance Metrics

The mean squared error (MSE) serves as a guide for the selection of the optimal hyperparameters and the final model evaluation, accompanied by the Nash-Sutcliffe efficiency (NSE), the root mean squared error (RMSE), the mean absolute error (MAE) and the correlation coefficient (R), which are calculated as follows:

$$MSE = \frac{\sum_{i=1}^{n}(Q_i^{obs} - Q_i^{pred})^2}{n} \tag{8}$$

$$NSE = 1 - \frac{\sum_{i=1}^{n}(Q_i^{obs} - Q_i^{pred})^2}{\sum_{i=1}^{n}(Q_i^{obs} - \overline{Q^{obs}})^2} \tag{9}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Q_i^{obs} - Q_i^{pred})^2}{n}} \tag{10}$$

$$MAE = \frac{\sum_{i=1}^{n}\left|Q_i^{obs} - Q_i^{pred}\right|}{n} \tag{11}$$

$$R = \frac{\sum_{i=1}^{n}(Q_i^{obs} - \overline{Q^{obs}})(Q_i^{pred} - \overline{Q^{pred}})}{\sqrt{\sum_{i=1}^{n}(Q_i^{obs} - \overline{Q^{obs}})^2 \sum_{i=1}^{n}(Q_i^{pred} - \overline{Q^{pred}})^2}} \tag{12}$$

The focus is on low MSE, RMSE and MAE values. The goal is to achieve high NSE and R values that approach unity to enable robust evaluation of model performance. The selected evaluation metrics MSE, NSE, RMSE, MAE and R are used to quantify model performance. While RMSE and MAE provide insights into the prediction errors in discharge units, NSE facilitates the comparison with the mean observed values. The correlation coefficient R measures the linear relationship between predicted and observed values. The selection of these metrics complies with the standards for hydrological modeling and ensures a comprehensive assessment of the effectiveness of the models (Gupta et al., 2009).

4.2.2. Analysis of statistical significance

After evaluating the performance metrics of each ML model, we compared their residuals to determine which models performed statistically better than others. To achieve this, we applied the Wilcoxon signed-rank test, following a similar approach to that used in the study by Corazza et al. (2013). The null hypothesis of the Wilcoxon test asserts that there is no difference between the models' predictions, i.e., their absolute residuals. The analysis focuses on the p-value as the key statistical measure. Based on the commonly accepted p-value threshold of 0.05, there are two possible scenarios, one that rejects the null hypothesis if the p-value is less or equal to 0.05, indicating statistically significant results, and the other possibility where there is no sufficient evidence to reject the null hypothesis where p-value is greater than 0.05. In simpler terms, when p is less than 0.05 we conclude that the compared models differ in their predictive capability.

4.2.3. Model interpretability and feature importance

The importance of providing transparency in AI models has grown, along with their increasing application in hydrology. Consequently, in recent years, the use of Explainable AI (XAI) methods has significantly risen. A study by Maier et al. (2024) emphasized the general application of such methods and categorized them into three groups: identification of critical features, estimation of feature contributions, and evaluation of the adaptability of a model to variations in features. Although XAI provides the necessary tools for clarifying complex processes, a careful analysis of its results based on domain-specific knowledge must be explored to avoid incorrect conclusions.

In this study, we focus on estimating feature contributions. First, we conducted two initial analyses to establish the relationships between the input and output variables. For this purpose, we applied correlation and mutual information (Bendat and Piersol, 2011; Thomas and Joy, 2006). Additionally, after obtaining the predictions from different ML models, the feature contributions were estimated using the SHapley Additive exPlanations (SHAP) method for simpler ML models (Lundberg and Lee, 2017) and Feature Occlusion Test for time-series models.

SHAP assigns each feature an importance value based on its contribution to the model's predictions, by breaking down the predicted output into the sum of feature contributions. This method is grounded in concepts from cooperative game theory (CGT) (Xu et al., 2024). SHAP supports both global and local interpretations, facilitating a deeper insight into individual predictions and the model's overall behavior. SHAP values follow three key properties: local accuracy, missingness, and consistency, enabling them to provide not only the required transparency of the model but also explanations that are consistent and accurate. (Lundberg and Lee, 2017)

SHAP is easily applicable to simple ML models, but it application is limited for SVR (RBF and sigmoid) and LSTM-based models. Although SHAP provides KernelExplainer as a model-agnostic solution, it is restricted for SVR models and inadequate for LSTM-based models. KernelExplainer has high computational requirements, limited interpretability for large datasets, does not encompass kernel-specific transformations, and may be biased if given a limited range of background data. Likewise, KernelExplainer allows only 2D input format (number of observations, features), while LSTM requires 3D (number of observations, timesteps, features). Ignoring the time-steps by averaging them hinders the SHAP capacity to handle time-dependent patterns. The current SHAP implementation is not compatible with RNN-based model layers for PyTorch. Another restriction of SHAP is related to highly correlated features. The presence of such features limits the SHAP's ability to find features with the greatest impact on model predictions (Nayebi et al., 2023). Hence, determining whether the importance is overstated or understated is impossible.

An alternative to SHAP for time-series models is a feature occlusion test. The feature occlusion test is a relatively simple technique used to assess the importance of individual features in a machine learning model by systematically occluding, or removing individual features, and then observing how the model's performance changes.

5. Results

The performance of the eight ML models on the training and test data sets are summarized in Table 1. As expected, all models generally performed better in estimating discharge from the simulated data set than from the measured data set, as the latter is subject to noise and inconsistencies even after preprocessing. The hyperparameters selected during the optimization process are shown in Table B.1 of Appendix B.

Two time-dependent models, LSTM and the hybrid LSTM-Attention, consistently achieved better results for all evaluation metrics, which are highlighted in bold in Table 1. The improvements of the LSTM-Attention model over the LSTM are moderate but consistent across different metrics. On the other hand, the SVR models, especially those with radial basis function and sigmoid kernels, showed suboptimal performance. The ranking of the models from best to worst based on the performance of the simulated data set is as follows: LSTM-Attention, LSTM, XGB, RF, DT, SVR-sigmoid, LGBM and SVR-rbf. A similar ranking applies to the measured data set, with the LSTM models at the top and the SVR models at the bottom.

LSTM and LSTM-Attention maintained their relative performance rank across both data sets. However, it is important to consider the data-specific differences in the results. For example, when comparing the performance of LSTM-Attention on simulated and measured data sets, there is a significant increase in accuracy for the simulated data set — 48.66% for RMSE, 47.85% for MAE, 1.95% for NSE and 0.81% for R. LSTM shows a similar trend with improvements of 45.85% for RMSE, 43.42% for MAE, 2.22% for NSE and 0.81% for R for the simulated data set.

In addition to the significant difference in estimation accuracy between simulated and measured data, the persistence of discrepancies between the evaluation metrics of the training and testing sets was also observed. While significant differences are seen for simple ML models, the same does not apply to the time series ML models. The largest differences were observed for SVR with the rbf kernel for the simulated data, and SVR with the sigmoid kernel for the measured data. This leads to the conclusion that certain models lack generalization ability and require more diversity in the data, which is not the case when it comes to the discharge of tidal rivers and estuaries, especially when the available data set is limited. Therefore, the preference is towards applying models that are not limited by the previous obstacles and enable a more robust and reliable discharge estimation, such as LSTM-Attention and LSTM.

Table 1: Performance indicators for all considered models.

	Simulated Data				**Measured Data**			
Model	**RMSE (m^3/s)**	**MAE (m^3/s)**	**NSE**	**R**	**RMSE (m^3/s)**	**MAE (m^3/s)**	**NSE**	**R**
				Training				
DT	39.005	29.541	0.971	0.985	59.203	42.828	0.933	0.966
RF	33.542	24.919	0.979	0.989	55.812	40.674	0.941	0.970
SVR - rbf	39.712	29.797	0.970	0.985	60.472	43.564	0.930	0.965
SVR - sigmoid	47.756	37.784	0.957	0.978	70.557	54.090	0.905	0.952
LGBM	37.392	28.176	0.973	0.987	59.572	43.360	0.932	0.966
XgBoost	31.147	23.761	0.982	0.991	48.500	35.601	0.955	0.977
LSTM	35.240	27.561	0.976	0.988	58.916	41.490	0.934	0.969
LSTM-Attention	30.094	22.753	0.983	0.991	56.509	40.874	0.939	0.969
				Testing				
DT	51.747	37.492	0.979	0.990	76.147	54.368	0.955	0.980
RF	48.729	34.669	0.981	0.991	73.306	52.159	0.958	0.982
SVR - rbf	70.128	36.554	0.962	0.983	77.387	54.410	0.953	0.979
SVR - sigmoid	53.010	40.978	0.978	0.990	82.155	58.992	0.947	0.947
LGBM	53.797	36.341	0.977	0.989	73.130	52.260	0.958	0.983
XgBoost	48.428	34.739	0.982	0.991	73.212	51.836	0.958	0.982

Table 1: Performance indicators for all considered models.

	Simulated Data				**Measured Data**			
LSTM	**34.384**	**27.057**	**0.991**	**0.996**	**63.495**	**47.821**	**0.969**	**0.988**
LSTM-Attention	**29.473**	**22.530**	**0.993**	**0.997**	**57.406**	**43.201**	**0.974**	**0.989**

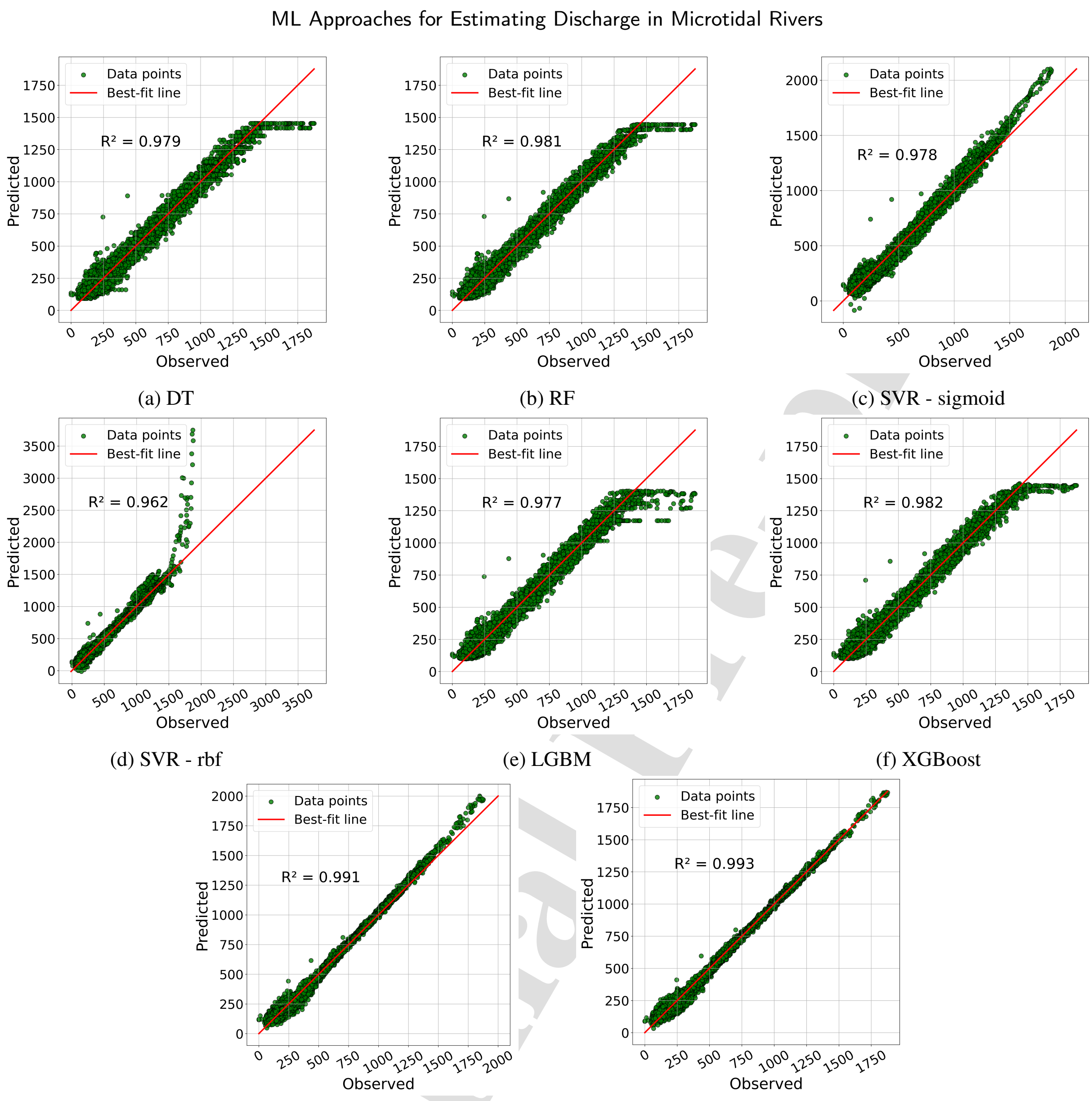
Figures 6 and 7 visually compare the predicted discharge values with the observed discharge values for all models and highlight their scatter from the best-fit line. It is noticeable that most models had problems estimating high discharge values (above 1500 m^3/s), leading to both over- and under-predictions. This is most likely a result of data imbalance, as there is a larger sample size for lower discharge values. In particular, DT, RF, LGBM and XGB tended to underestimate high discharge values, while the SVR models tended to overestimate. The simple supervised model also had problems with low discharge values (below 500 m^3/s), while LSTM and LSTM-Attention showed a very good agreement. Overall, time-dependent ML models showed a more balanced prediction profile, with LSTM-Attention achieving the best agreement.

When comparing time-dependant models, LSTM and LSTM-Attention, we can clearly see a much larger spread of values for the measured data set compared to the simulated data set. In Figures 6 and 7, we see a smaller scatter of points for the LSTM-Attention model for the entire range of discharge values, even for the discharge extremes above 1500 m^3/s. However, larger errors are observed for the LSTM model due to its limitation of overpredicting extremes which result in the largest errors in estimations. To summarize, LSTM-Attention and LSTM are the best models, which is confirmed by both comprehensive error metrics (Table 1) and visual inspection of the results (Figures 6 and 7). These models show high accuracy and reliability in predicting discharge in tidal rivers for different scenarios.

A Wilcoxon signed-rank test was conducted to further analyze the evaluation metrics presented in Figure C.1 of Appendix C. The test aimed to determine whether there are statistically significant differences between the time-series models in comparison to other models, for both the simulated and measured dataset.

The LSTM-Attention model demonstrated the best performance, and this is confirmed by the Wilcoxon test results. The p-values for LSTM-Attention were statistically significant when compared to LSTM and simpler ML models, with enough evidence to reject the null hypothesis. Similarly, the LSTM, also demonstrated significant performance improvements over simpler models. This shows that the time-series models used are effective and reliable for both datasets, even with noise present. However, when simple ML models are compared, there are instances when the differences between the models' predictions do not always differ substantially.

For the simulated dataset, differences between RF, SVR-rbf, and XGBoost were minor, with p-values close to the threshold, indicating that their prediction capabilities are not significantly different. In the measured dataset,



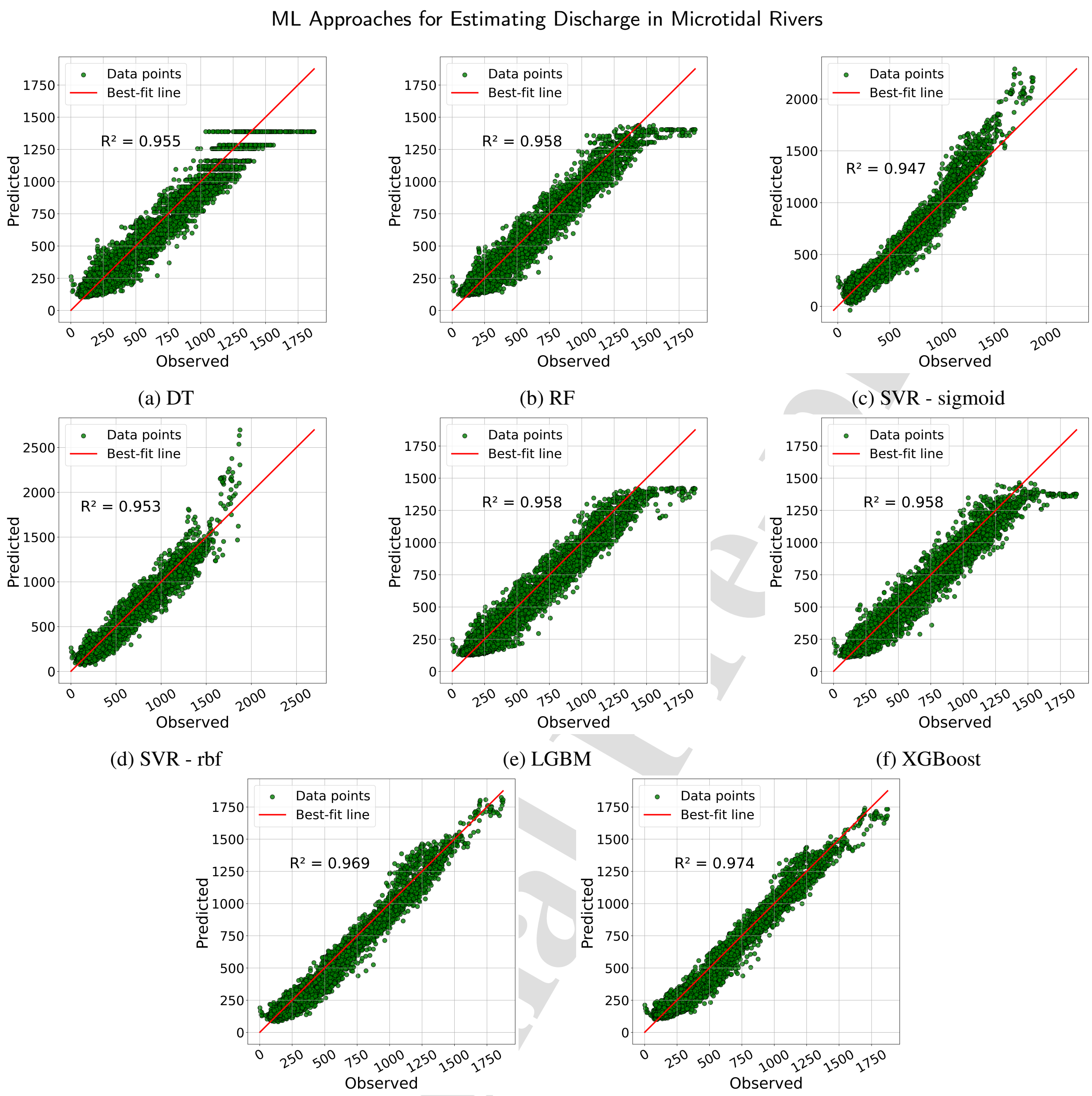
(a) DT (b) RF (c) SVR - sigmoid

(d) SVR - rbf (e) LGBM (f) XGBoost

(g) LSTM (h) LSTM-Attention

Figure 6: Predicted versus observed discharges for simulated data set

LightGBM and XGBoost showed similar performance, with larger p-values in comparisons like RF vs. LGBM and RF vs. XGBoost, indicating no significant difference between models.



(a) DT (b) RF (c) SVR - sigmoid

(d) SVR - rbf (e) LGBM (f) XGBoost

(g) LSTM (h) LSTM-Attention

Figure 7: Predicted versus observed discharges for measured data set

6. Discussion

6.1. Data Challenges and Limitations

Application of ML in hydrology typically includes problems and challenges with data quality. In our case, the water level data obtained by the Croatian Meteorological and Hydrological Service agency at hourly intervals did not pose any challenges or issues. This is because those measurements had already undergone the official quality check. Hence, no additional corrections were required. However, the same does not apply to the discharge data.

Discharge values were obtained via a different agency, Croatian Waters. Discharges are estimated by summing the flow rates calculated from water level and velocity profiles measured by three horizontal ADCP devices installed under a bridge. Due to various reasons, such as malfunctions or maintenance services, the devices had short periods of missing data, with lengths of several hours. Those short periods were filled by applying an interpolation on measurements obtained by the other two devices. Additionally, the discharge data also contained high-frequency noise characteristic for ADCP measurements, especially in tidal rivers and estuaries, where tidal currents and saltwedge dynamics have a noticeable influence. Therefore, a moving average of three hours was applied to the measured data to remove high-frequency noise and outliers. An additional level of control was imposed by investigating simulated data, generated by a numerical model.

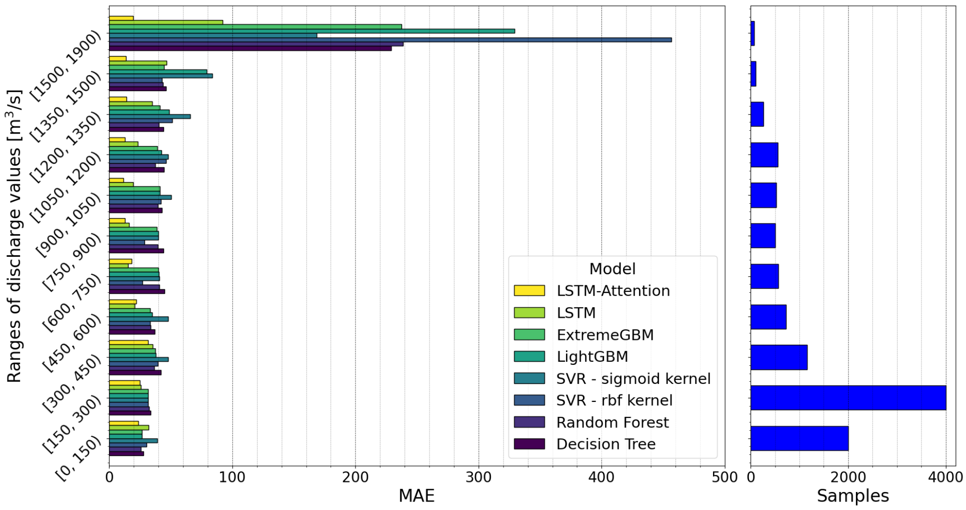
In this study, we used both simulated and observed datasets to ensure a comprehensive analysis of the system's dynamics. The observed dataset contains some noise and irregularities inherent to field measurements, which can affect the accuracy of model evaluation. On the other hand, the simulated dataset is clearer, free from such inconsistencies, providing a controlled environment for comparison. By comparing the spectrograms in Figure D.1 of Appendix D., we observe that while both datasets show similar trends in the strength spectrum, the observed data has more variability and noise across different periods. This variability helps assess the model's robustness under real-world conditions, while the simulated data allows us to validate the model's performance under ideal, noiseless conditions. This dual approach ensures that our model can handle both ideal and real-world scenarios effectively.

6.2. Detailed Assessment of Models Accuracy

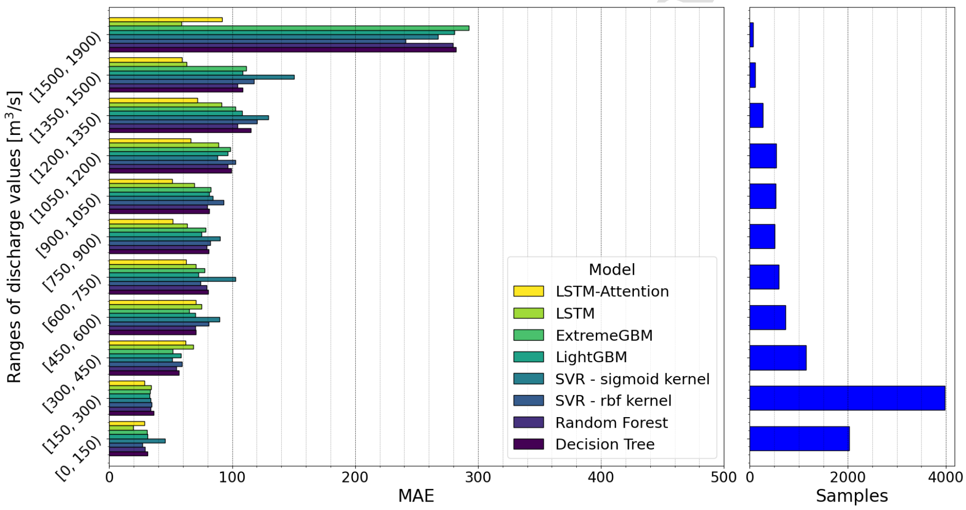
The accuracy of all considered ML models is discussed in more detail by evaluating their accuracy for different discharge ranges - specifically at low, medium, high and extremely high discharge range. Figure 8 illustrates the metric of mean absolute error (MAE) for all ML models, where the discharges are categorized into bins with an interval of 150 m^3/s, except the values above 1500 m^3/s. The reason is that a very small number of samples refer to extreme flows, which represent a single event in the test data set. The figure also includes a histogram for each discharge range describing the frequency of each interval in the test data set.

The results confirm that most ML models achieved better results for the simulated discharges compared to the measured ones. An exception is the SVR-rbf model, which shows poor performance for high discharges in the simulated data set. Furthermore, the following conclusions can be drawn:

- Low discharges (below 300 m^3/s): All models show similar behavior with minimal MAE, with LSTM and LSTM-Attention proving to be the best fit for both measured and simulated data sets. These results are expected given that the majority of the data (about 60%) falls within the mentioned range.



(a) Simulated



(b) Measured

Figure 8: Calculated MAE metric and histogram for different discharge ranges, for: a) simulated data set and b) measured data set.

- Medium discharges (300 to 1050 m^3/s): The models show comparable accuracy for the measured data set. For the simulated data set, however, LSTM and LSTM-Attention perform better than the other models. The second largest category, pertaining up to 30% of data, shows the beginning of the larger differences in performance between the utilized models. This is especially evident towards the higher end of the range.

- High discharges (1050 to 1500 m^3/s): LSTM and LSTM-Attention perform significantly better than the other models for both data sets. The trend of increasing MAE error also continues in this value category, which contains about 10% of the total data set.

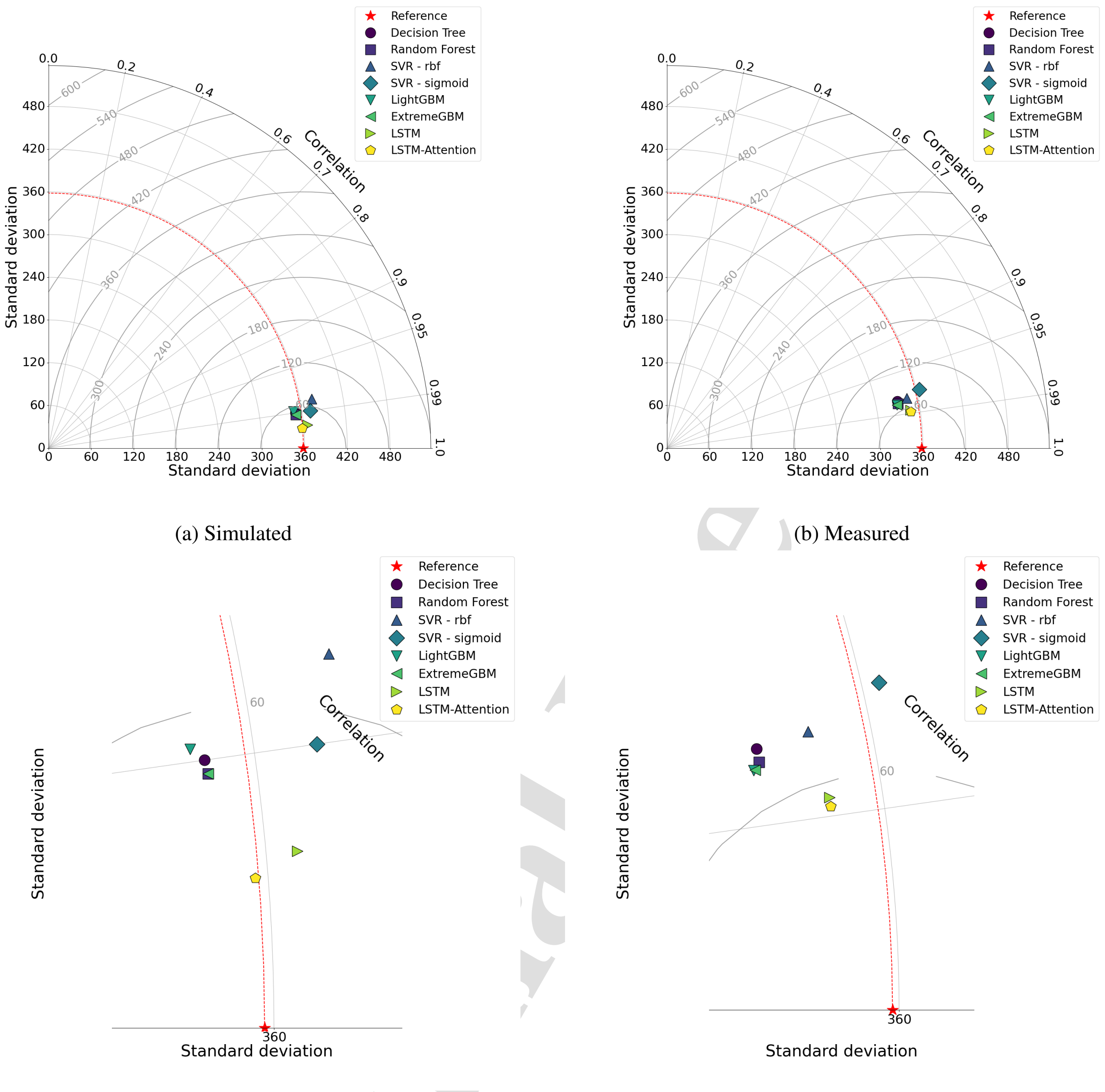
- Extremely high discharges (over 1500 m^3/s): LSTM and LSTM-Attention perform several times better than simple ML for both data sets, with MAE error lower than 100 m^3/s. Less than 10% of the data is contained in this range. Simple ML models' error is above 160 m^3/s for the simulated and 200 m^3/s for the measured data set, thereby indicating the biggest disparities in accuracy.

These results can be explained by examining the histogram in Figure 8, which shows that high discharges are less frequent than medium discharges and extremely high discharges are very rare. It is well known that traditional ML models have difficulty predicting rare values due to training biases where models prioritize common values, leading to suboptimal performance on rare events (Chawla et al., 2002). Unbalanced distributions and a lack of representative samples for rare values can impair the ML ability to generalize effectively. However, time-dependent ML models seem to deal with this problem quite successfully. Therefore, both LSTM and LSTM-Attention exhibit a remarkable ability to generalize and extrapolate on new unseen data.

When comparing the two time-dependent ML models, LSTM-Attention performs slightly better than LSTM. For the measured data set, LSTM-Attention improves the estimate in 7 out of 11 discharge ranges, with lower accuracy observed only for the extremely high discharge range. Similarly, the results of the simulated data set show an improvement in 9 out of 11 discharge ranges, with insignificantly lower accuracy in the two medium discharge ranges than LSTM, more precisely, 9.83% as mean improvement. Taking into account the performance of the model in all categories, we can conclude that unlike other models, where the data distribution significantly affects the accuracy of the model, this is not the case with the novel LSTM-Attention model, and that is exactly why it is considered an adequate model for predicting discharge independent of river dynamics and extreme conditions.

In addition, a comprehensive comparison illustrated in the Taylor diagram (Figure 9) provides a holistic assessment of several statistical metrics, including standard deviation, R, and centered RMSE. The standard deviation of the simulated and measured data set is almost identical, 358.28 and 358.34, respectively. In the simulated scenario (Figure 9a), LSTM and LSTM-Attention show comparable performance based on the R metric, with a slight difference of only 0.1%. However, LSTM has a slightly higher standard deviation (365.8) than LSTM-Attention (357.6) and a higher RMSE value, indicating a possible tendency to over-predict extreme values.

Based on the visualization of the simulated data set, the following can be concluded: The ability to minimize the total error is better for the hybrid LSTM-Attention model, and at the same time it is better at capturing data variability. In the scenario with the measured data set (Figure 9b), LSTM-Attention and LSTM show superior performance based on the R and RMSE metrics. The inclusion of the standard deviation metric demonstrates the adequacy of both models (LSTM-Attention: 346.5, LSTM: 346.6) as their standard deviation closely matches the reference standard deviation of the measured data set (358.28), although not as close as in the simulated scenario (358.34).



(a) Simulated

(b) Measured

(c) Simulated (detail)

(d) Measured (detail)

Figure 9: Taylor diagram for all considered ML models for: a) simulated data set, b) measured data set, c) simulated data set (detail), and d) measured data set (detail).

Figures 10 and 11 show the time series of the observed and estimated discharge by the two best models, LSTM and LSTM-Attention, for the simulated and measured data set, respectively. Overall, both models show good agreement with the simulated data over the entire discharge range (Figure 10). A closer look at the maximum values in February 2021 shows excellent agreement between the estimated and simulated discharges. However, at low discharges in August 2021, both models seem to underestimate the amplitude of the daily oscillations reflecting the tidal dynamics, although they capture the daily mean discharge.

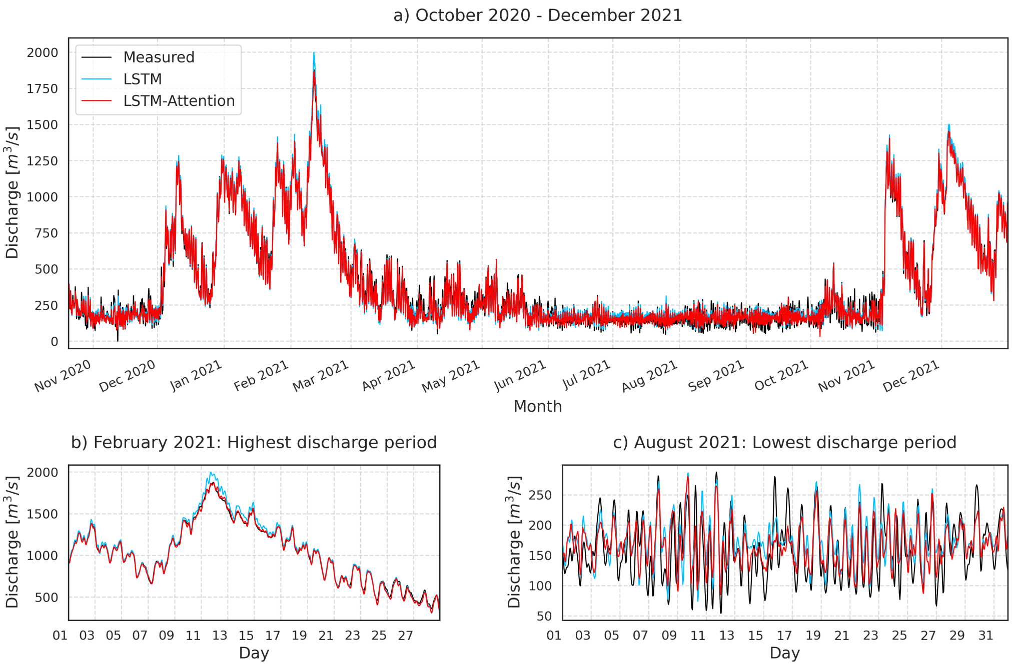


Figure 10: Simulated and estimated time series of discharges for the test data set: a) Oct 2020 - Dec 2021, b) Feb 2021, c) Aug 2021.

Similar results are found for the measured data set, with both models showing good overall agreement (Figure 11). In this case, the models have a slightly lower accuracy for the period of maximum flow in February 2021, although they can accurately capture the daily oscillations. Also for the period of low flow in August 2021, both models underestimate the oscillations even more than in the case of the simulated data set. This is most likely related to complex tidal-fluvial interactions, where tidal oscillations can be dampened or amplified by the river discharge (Matte et al., 2014). In this case, the operation of upstream hydropower plants further complicates the interaction between tidal and fluvial waves. Therefore, to capture the high-frequency oscillations, a hybrid time-dependant model that integrates a non-stationary harmonic analysis, such as NS_TIDE (Matte et al., 2014), could give more robust results for low flow conditions. Another potential cause of the lower performance in capturing the daily oscillations during the low flow period could be attributed to the 24-hour window of both models required to capture the maximum flows. Therefore, further investigation of the influence of the window length on the performance of the LSTM and LSTM-Attention models is needed for this type of problem.

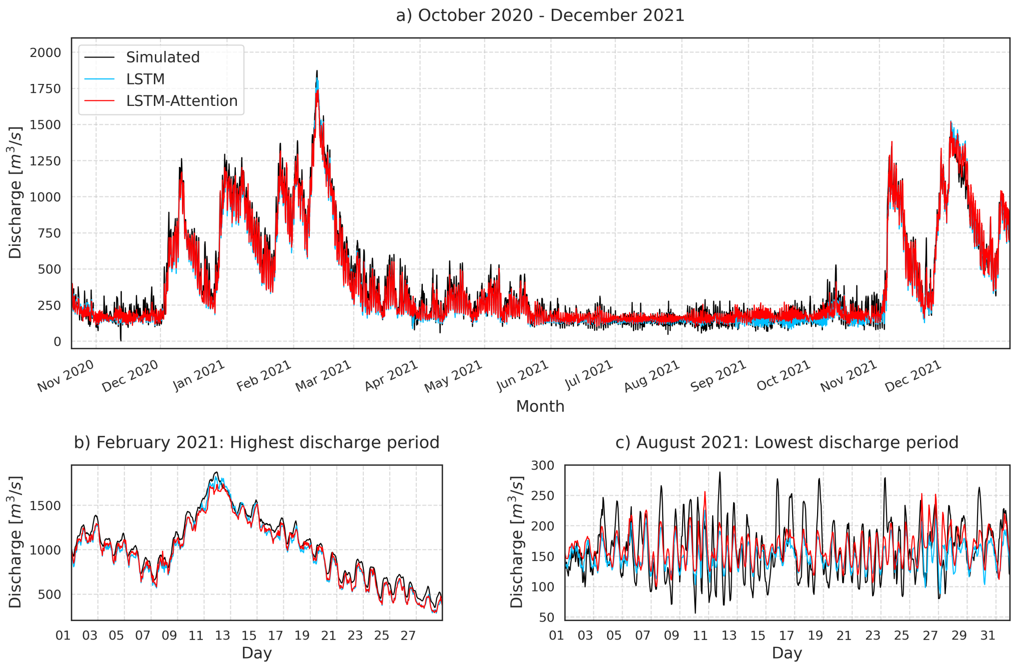


Figure 11: Measured and estimated time series of discharges for the test data set: a) Oct 2020 - Dec 2021, b) Feb 2021, c) Aug 2021.

6.3. Exploring Feature Importance

The decision to use only water level data from multiple stations as input variables for machine learning models is based on the previous research conducted by Habib and Meselhe (2006) and Hidayat et al. (2014). These studies demonstrated the usefulness of incorporating water level data from multiple stations. This approach is particularly useful in situations where other parameters, such as meteorological or physical data, may not be readily accessible. Additionally, because the research is focused on a small salt-wedge estuary within a 25 km distance from the river mouth, meteorological factors such as temperature, air pressure, precipitation, wind, and others parameters were deemed negligible. This is in contrast to a recent study by Vu et al. (2023), which investigated the entire basin of Loire-Bretagne, encompassed by a hydrographic network exceeding 135,000 km.

The feature importance of different water level stations where examined using correlaion, mutual information, SHAP and feature occlusion. The correlation matrix in Figures E.1 and E.2 of Appendix E. presents the relationship between water levels at four stations—Ušće, Opuzen, Kula Norinska, and Metković—and discharge measured at the Metković station. Water levels at Ušće show a low correlation with discharge at Metković with a correlation coefficient of 0.24. Water levels at Opuzen have a moderate relationship with discharge at Metković, with a correlation coefficient of 0.49. Kula Norinska station shows a correlation of 0.53 with discharge at Metković, indicating a slightly stronger relationship compared to Opuzen, but still not highly predictive on its own. Water levels at Metković exhibit the strongest correlation with discharge at the same location, with a coefficient of 0.56. Higher correlation is expected due to a direct relationship between local water levels and discharge. Much higher correlations are found between water levels at adjacent stations. For example, Opuzen and Kula Norinska show a very high correlation (0.95), as do Kula Norinska and Metković (0.97). These values suggest strong interdependencies between water levels at these locations. Overall, the results highlight that while all water level measurements contribute to the prediction of discharge at Metković, the local water levels at Metković and nearby stations (Kula Norinska and Opuzen) provide more predictive power than those at more distant stations like Ušće.

However, this approach is limited in accounting for the non-linear interactions between the features. Therefore, mutual information was also applied, shown in Figure F.1 of Appendix F., which enables detecting both linear and non-linear dependencies between variables. The mutual information score measures the dependence between each water level (input feature) and discharge, with higher values indicating a stronger relationship. In both simulated and measured datasets, the results are in line with the correlation analysis, water level at Metković is the most important predictor, reflecting its proximity to the discharge measurement point. Water levels at Kula Norinska rank second, and water levels at Opuzen have moderate importance, more so in measured data than simulated. Water levels at Ušće show the least importance in both datasets, having minimal influence on discharge. Overall, local stations like Metković and Kula Norinska dominate discharge predictions, with downstream stations contributing less.

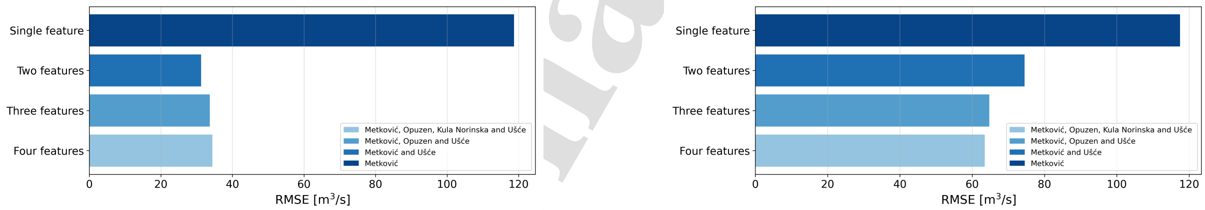
However, this assumption contradicts hydraulic principles, as tidal dynamics and sea levels have a critical impact on the water levels at the upstream stations, as well as on the flow patterns. Ignoring this feature may result in an oversimplified system, reducing the accuracy and reliability of the model's predictions. Based on the domain knowledge, we justify the inclusion of the sea level data from the tidal station.

Applying the SHAP method to simple ML models (DT, RF, LGBM, and XGB), Figures G.1 to G.4 in section G.1 of Appendix G., and Figures G.1 to G.4 in section G.2 of Appendix G., enabled us to thoroughly understand the influence of individual variables on discharge prediction, both at local and global significance levels. This study showed that the most significant variables in estimating tidal river discharge are the water level at the tidal station Ušće and the hydrological station for which the prediction is made (Metković). This supports our argument that the tidal station has an important effect on upstream water levels.

Due to the limitations of the SHAP approach for time-series models, the feature importance of the LSTM-based model was evaluated using feature occlusion and testing its performance on different combinations of input features. Performance declined in nearly all scenarios when features were removed, as shown in Tables H.2 to H.4 of Appendix H. The only exception was the LSTM model evaluated on the simulated dataset (Table H.1). Using only the water level from the Metković station resulted in the worst performance for all time-series models, highlighting the necessity of including additional features. Based on SHAP analysis, the water level at the tidal station Ušće proved to be the most important next to the Metković station. For scenario including these two stations, performance improved significantly—by 37-47% for the simulated dataset and around 73% for the measured dataset in terms of RMSE. This underscores the need to include water level data from the tidal station.

Additional variables, such as water levels at Opuzen and Kula Norinska, also improved the model. For Opuzen, the LSTM-Attention model showed an 8% improvement for the simulated data and a 5-11% improvement for the measured data. However, the LSTM model's performance decreased by 7% with the inclusion of Opuzen for the simulated data. Including Kula Norinska led to a 2% improvement in the LSTM-Attention model for the simulated data and a 1-11% improvement for the time-series models in the measured data. Again, the LSTM model's performance declined by 2% when Kula Norinska was added. These RMSE differences based on input feature combinations are visualized in Figures 12 and 13. While midstream tidal stations also improved the model, their impact was smaller compared to the two-variable scenario.

In conclusion, while basic statistical methods like the correlation and mutual information can identify significant variables, their interpretation is limited in complex hydrological systems like tidal rivers and estuaries. By applying SHAP, we were able to assess the importance of individual features for simpler ML models. For time-series models, however, a different approach was needed due to the specific limitations of RNN-based models. Testing various combinations of input variables with LSTM-based models highlighted the importance of including all stations to improve prediction accuracy, especially when using measured data.



(a) Simulated

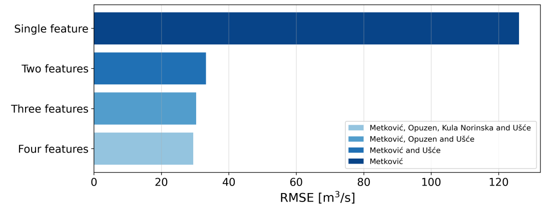
(b) Measured

Figure 12: LSTM prediction error with different input features

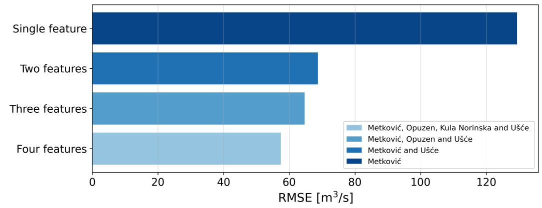
6.4. Suitability and Benefits of LSTM-Attention Model

The inclusion of the attention mechanism significantly improved model performance across all predicted discharge ranges. This enhancement is due to several advantages of the LSTM-Attention model, which will be explained in this subsection.

The LSTM-Attention model demonstrated better extrapolation abilities, leading to enhanced generalization performance. This is particularly important in modeling environmental processes, where extreme event frequency



(a) Simulated



(b) Measured

Figure 13: LSTM-Attention prediction error with different input features

and severity increase over time. Historical data may not adequately capture these changes, making models capable of effective generalization essential. The attention mechanism's ability to improve extrapolation has been observed in other studies (Yang et al., 2024).

Data-driven models often struggle to predict values outside their training set, leading to overfitting or underfitting, as noted in previous research Forghanparast and Mohammadi (2022). For example, in a tidal reach prediction study (Guo et al., 2021), RF, SVR, and LGBM models—also used in our study—struggled with extrapolating values beyond the training data. Our study confirms these findings, as evident in the predicted vs. observed plot and MAE metrics. DT and RF are limited by their reliance on local patterns, while SVR's extrapolation depends heavily on data quality. LGBM focuses on complex interactions, and XGB excels at minimizing loss, making it more reliable for extrapolation. Tree-based models generally face challenges when predicting outside the training range.

The introduction of the attention mechanism mitigates this limitation in LSTM models by allowing them to focus on relevant non-local information for predictions. While LSTM retains information from earlier time steps, the attention mechanism assigns varying weights to hidden states based on their relevance, improving adaptability to non-stationary patterns. This feature makes the model more resilient to irrelevant and noisy data, a key limitation of simple ML models, which cannot identify temporal or sequential dependencies.

Previous studies also highlight the attention mechanism's robustness in handling noise and outliers, with minimal performance drops Li et al. (2024). Similarly, our study shows that LSTM is more resilient to noise compared to simpler ML models, with SVR being the least resilient.

Data imbalance is a critical issue in ML model development, as observed in Thanh et al. (2022), where predicting discharge for less frequent ranges was challenging. This trend is also evident in our study, especially for high (1000-1500 m^3/s) and extremely high discharges (above 1500 m^3/s).

Despite the limitations of simple ML models, they were included in this study due to their advantages in solving regression problems, including handling non-linear data, transparency, resource efficiency, and fast training. Kernel-based SVR is optimal for smaller datasets, while DT, RF, XGB, and LGBM perform better with large-scale, high-dimensional data (Nagaradjane et al., 2024). When selecting a model, factors such as data size, complexity, and non-stationarity must be considered. While DT and SVR struggle with non-stationary data, ensemble models like RF, XGB, and LGBM adapt well to changing patterns.

Model training time is another important factor. In our study, the order of training time for hyperparameter optimization was RF, LSTM-Attention, SVR (RBF and sigmoid kernel), LSTM, LGBM, XGB, and DT. Although LSTM-Attention had longer training times, its performance improvement justifies the computational cost.

While simple ML models have their benefits, they fall short in predicting discharge in tidal reaches. LSTM-Attention offers several advantages, including identifying critical features, generalization, extrapolation, and handling imbalanced datasets. Compared to the stand-alone LSTM, this method enhances overall model performance, providing a broader range of benefits.

6.5. Comparison with Prior Research

The results obtained are in agreement with findings from earlier studies. While it is not possible to directly compare models and measurements, a general comparison can be performed, indicating that our results are expected. According to Habib and Meselhe (2006), simple statistical ML approaches encounter difficulties in estimating extreme discharge values, unlike neural networks. Although our study is related to utilizing advanced RNNs and hybrid models, it confirms the previous statement.

Wolfs and Willems (2014) utilized only simulated data to avoid common uncertainties, which are unavoidable when having measured data. We have also emphasized this fact. However, the ranges of utilized water levels and discharge are quite different from ours, as water levels are four times higher and discharge is around ten times lower. One of the indications that led us to employ advanced RNNs is the limited ability of ANN when a small dataset is available. Its generalization power drops, and its interpretation is not as straightforward as that of decision trees.

Study by Hidayat et al. (2014) significantly differed in the range of discharge values, but not in a visualized water level station of Tenggarong, which contains the same range as our multiple utilized stations. The distinction between our methodology and theirs pertains to the water level station at the discharge station of interest. However, their dataset was smaller, consisting of less than two years, whereas ours spanned over a six-year period. The better performance of our models can be attributed to these mentioned impediments. The simulated dataset had an RMSE improvement of half than Hidayat et al. (2014), around 1.5%, when compared to the overall discharge range, and the same percentage for the simulated dataset, around 3%.

The results of Thanh et al. (2022) can be partially compared with ours. Comparison is only viable for two metrics, NSE and R because the water level and discharge range are ten times higher. The obtained results of the models that were also used in our work (DT, RF, and SVR) deviate less than 1% from the mentioned metrics for the simulated data set, which indicates the reliability of the methods used and the possibility of generalization, even though these are different analyzed tidal rivers. Also, the data distribution significantly differs because, unlike our data, where there is only one event with extreme discharge values above 1500 m^3/s, the representation of such events is much higher in the previous study. Therefore, the RF model does not encounter significant difficulties when estimating those values compared to ours.

When further investigating the performance of the LSTM model, the estimation precision is greater for the lowest discharge values. In contrast, it is larger for flood periods in both simulated and measured cases, as in Vu et al. (2023), although they utilized additional meteorological parameters. Therefore, the omission of additional parameters did not decrease performance in our case. However, the same conclusions are not drawn for the hybrid LSTM-Attention model, particularly in the simulated scenario, as there are no significant disparities in the model's performance during the described extreme events.

7. Conclusion

The complex nature of tidal flow dynamics in rivers poses a major challenge to the development of effective water management systems and timely risk warning protocols, especially in the face of ongoing climate change and anthropogenic impacts. This study investigates the possibility of estimating discharge in microtidal rivers such as the Neretva in Croatia using only water level data and ML models. To fully evaluate the performance of the model, we conducted tests with simulated and measured data sets. The study compares the performance of six simple supervised models (DT, RF, SVR-rbf, SVR-simgoid, LGBM, XGB) with two time series models (LSTM and LSTM-Attention) using different statistical and graphical evaluation methods. The main findings and accomplishments obtained from this study are as follows:

- The potential of advanced RNN and hybrid modeling (Mihel et al., 2024) led to the application of LSTM and a novel LSTM-Attention model with the main goal of improving discharge estimation for a microtidal river.
- Discharge in a tidal reach of microtidal rivers can be estimated using only water level data from various locations, either upstream or/and downstream, which is inline with several previous studies (see a review paper by Mihel et al. (2024)). Investigated ML models exhibit sufficient accuracy without incorporating additional meteorological data. For that reason, including additional data is unnecessary and does not impair the model's performance, if the length of the tidal reach is limited, in particular less than 25 km as in the present case.

- The study presents the first comprehensive analysis between simple and complex time series ML models, with a particular focus on the novel LSTM-Attention model for a microtidal river. The results show that the ML time series models are reliable and accurate in assessing river discharge based on water levels in both simulated and measured scenarios. In contrast, simple supervised models struggled and faced significant challenges in discharge estimation, except at low flow conditions, where they showed satisfactory accuracy. Time series models had minimal problems at high discharges and showed the lowest errors in both statistical and graphical analyses.
- LSTM-Attention and LSTM showed the least scatter of points around the best-fit line, with LSTM-Attention showing a better fit in most discharge ranges, which was confirmed by the MAE metric across different discharge ranges. The Taylor diagram confirmed these conclusions and showed that LSTM-Attention and LSTM achieved the most favorable combination of several statistical metrics. Consequently, LSTM-Attention proved to be the preferred model due to its good correlation, reasonable deviation from the standard deviation of the observed data and minimal overall residuals in both scenarios.
- Based on extensive model analyses, we can, therefore, conclude that the LSTM-Attention model provides the most reliable results for scenarios that include both simulated and measured flow values and are characterized by numerous oscillations during both high and low flow periods.
- Accurate estimates during flood periods are essential for timely flood warnings, risk mitigation and public safety. This novel approach is suitable for reconstructing discharges in microtidal rivers at locations where discharge monitoring stations have only recently been established or to fill missing data.

The machine learning models provide valuable insights that can be translated into practical recommendations for water resource management and environmental assessment. Unlike traditional methods that rely heavily on direct discharge measurements, which can be logistically challenging and costly in tidal rivers and estuaries, our methodology reconstructs river discharges from available water level data. This approach uses advanced machine learning techniques to account for the complex interactions between tidal influences and river flows, offering accurate and reliable discharge estimates.

For water resource management, the discharge estimates can inform decision-making processes related to water allocation, drought mitigation, and flood risk management. Reliable discharge data can support the sustainable management of water resources for irrigation, ensuring that planned allocations align with the natural variability of the system. Furthermore, this approach can significantly enhance flood forecasting capabilities by providing real-time, high-resolution discharge data where only water level measurements are available. For instance, integrating these models into early warning systems can enable authorities to predict and respond to flood events with greater precision, ensuring timely and effective evacuation plans and flood control measures. Moreover, the reconstructed discharge data can be used to calibrate hydrological models, improving flood risk assessments and management of flood control infrastructure.

In the context of ecosystem management, accurate discharge data can help monitor and conserve aquatic habitats more effectively. These data can be used to assess the suitability of water conditions for various species, identify critical thresholds, and evaluate the impacts of human activities or climate change on ecosystems. For example, discharge data can guide the implementation of environmental flows and the restoration of degraded habitats, ensuring that water resources are managed in a way that supports biodiversity and ecosystem health. This work contributes to a more integrated and sustainable management of these dynamic and ecologically important water bodies.

Generalization ability of our tested approach presents a pilot study which can be applied to other tidal rivers and estuaries globally, where different ML models, simple and time series, have been evaluated on two different datasets, one obtained through a conducted simulation, while the other through measurements and estimations of government agencies. For any tidal river whose river flow dynamic can be precisely explained by solving a system of partial differential equations (PDE), our analyzed machine learning architecture results and their performance can be partially compared, even though different hydro-meteorological conditions can be present.

It is reasonable to assume that the proposed ML method will perform even better in coastal rivers and estuaries with less pronounced stratification (well-mixed and partially-mixed estuaries), as micro-tidal estuaries have a week relationship between water levels and discharge at low values (due to the two-layer salt-wedge structure). Conversely, in complex and divergent estuaries and deltas with numerous tributaries, there may be certain constraints if a network of hydrological stations is not sufficiently densely distributed. In addition, the method may have a lower level of prediction in large rivers that are characterized by high flow and sections with long distances between adjacent stations, as the river flow may undergo significant changes between adjacent stations. In such circumstances, it will likely be necessary to incorporate other meteorological parameters, such as wind and precipitation, or direct surface inflows during rainfall episodes. Furthermore, it is probable that the introduction of an additional groundwater level parameter is necessary for rivers that exhibit a significant interaction with groundwater in order to accurately determine the hydrological parameter using this method.

Based on our results, we can make several suggestions for future research in this area. Tracking the length of saltwedge intrusion length and its influence on water level dynamics could certainly improve the predictive capabilities of ML models. This could be achieved by combining numerical modeling with ML algorithms. The results could be further improved by decomposing the tidal signals into harmonic constituents and the residual, which could be used as individual inputs for ML models. Another way to improve the results is to include additional hydrological parameters such as temperature and salinity to account for their seasonal variability. Finally, time-frequency distributions, wavelet analyses or non-stationary harmonic analyses could be used for this purpose by developing a hybrid ML method.

Acknowledgment

This work was fully supported by the EU Horizon 2020 project INNO2MARE under the number 101087348, the Croatian Science Foundation project 4SEAFLOOD ("Compound Flooding in Coastal Croatia under Present and Future Climate", IP-2022-10-7598), and University of Rijeka projects uniri-iskusni-tehnic-23-83, uniri-iskusni-tehnic-23-74, uniri-iskusni-tehnic-23-11, and uniri-zip-2103-4-22. We would like to thank Croatian Waters for their support in providing measurement data.

CRediT authorship contribution statement

Anna Maria Mihel: Methodology, Validation, Investigation, Writing – Original Draft Preparation, Visualization. **Nino Krvavica:** Data curation, Conceptualization, Methodology, Validation, Formal Analysis, Investigation, Resources, Writing – Original Draft Preparation, Writing – Review & Editing, Supervision, Project Administration. **Jonatan Lerga:** Conceptualization, Methodology, Validation, Formal Analysis, Investigation, Resources, Writing – Original Draft Preparation, Writing – Review & Editing, Supervision, Project Administration, Funding Acquisition.

Appendix A. Lag Shift

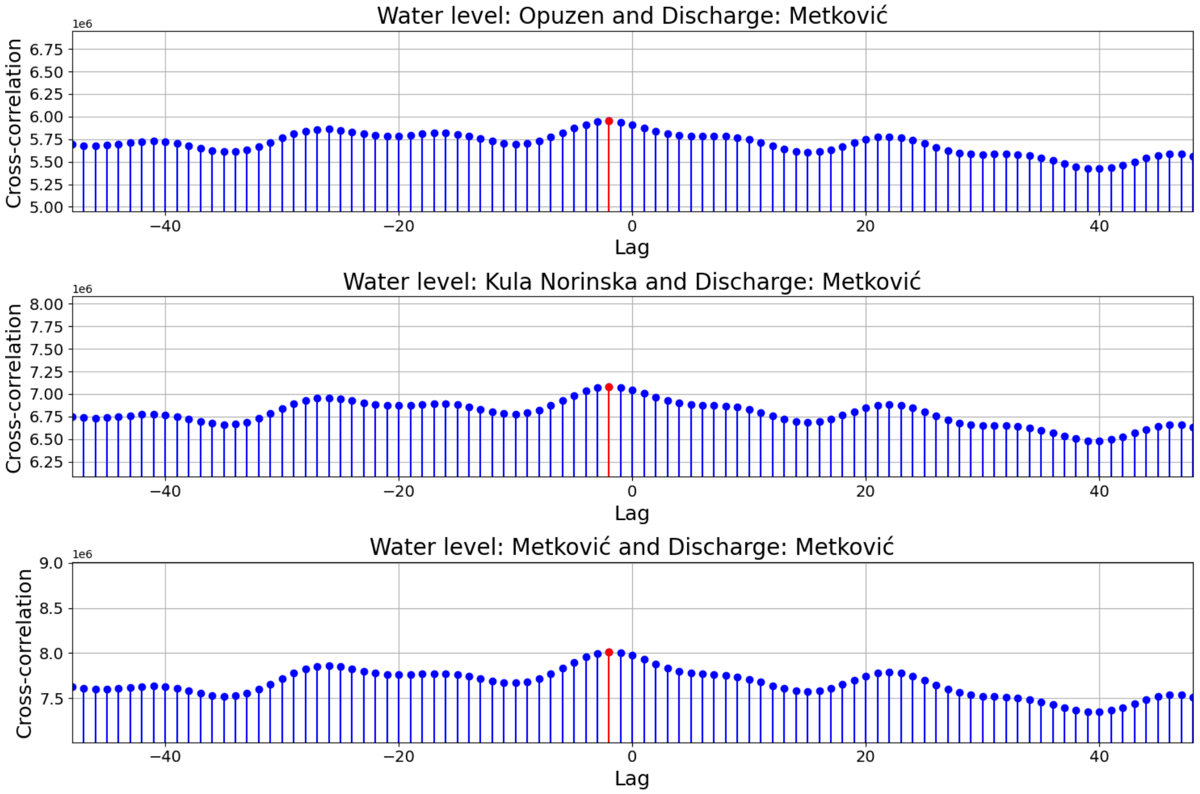


Figure A.1: Cross-correlation between simulated water levels at different stations and discharges at the Metković station.

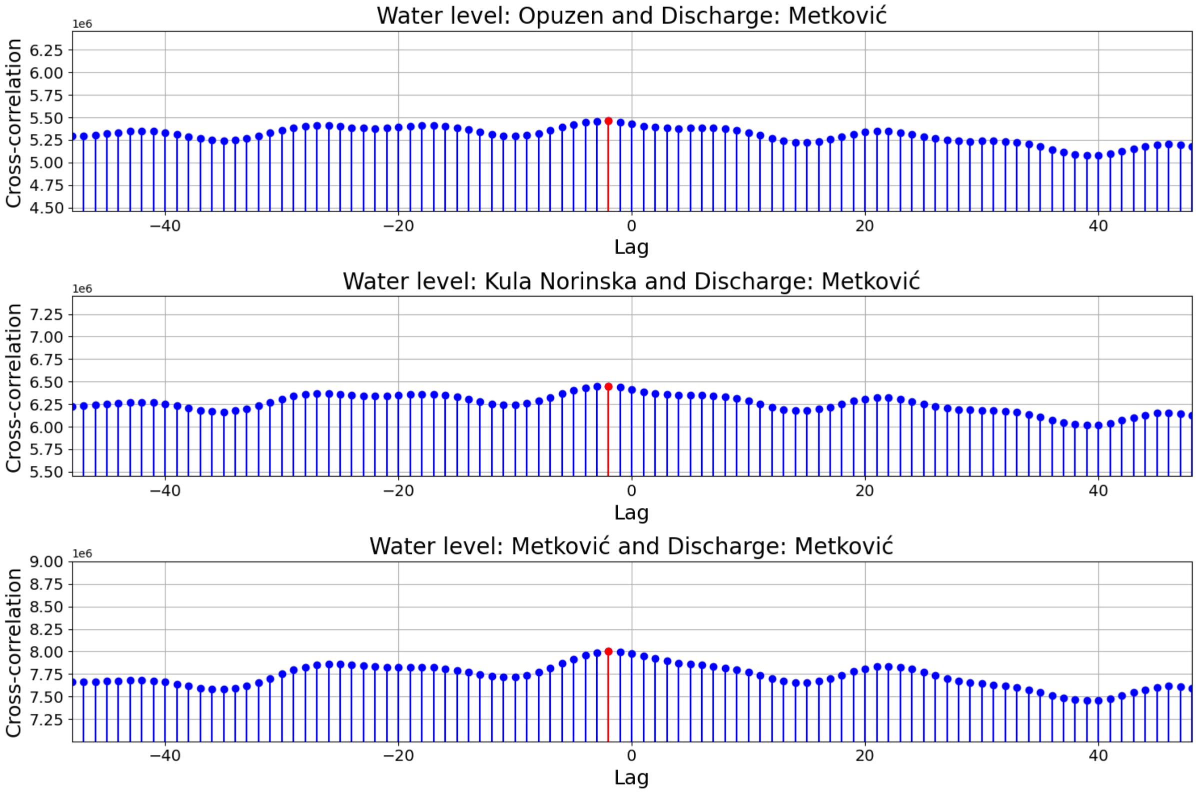


Figure A.2: Cross-correlation between measured water levels at different stations and discharges at the Metković station.

769 ## Appendix B. Optimization of Models Hyperparameters

Table B.1: Optimal hyperparameters for different machine learning models using simulated and measured data sets.

Model	Hyperparameter	Search range	Simulated best fit	Measured best fit
DT	*max_depth*	[10, 200]*	10	20
	min_samples_leaf	[10, 100]*	10	40
	min_samples_split	[10, 100]*	10	10
RF	*max_depth*	[10, 50]*	20	10
	min_samples_leaf	[10, 100]*	10	10
	min_samples_split	[10, 100]*	20	10
	n_estimators	[10, 200]*	130	90
SVR - rbf	C	0.001 x 10^n for n ∈ {0, 1, ..., 6}	1000	1000
	γ^a	[0.0001, 0.0005, 0.001, 0.005, 0.01, 0.05, 0.1, 1]	1	1
	ε^b	[0.0001, 0.0005, 0.001, 0.005, 0.01, 0.05, 0.1]	0.01	0.01
SVR - sigmoid	C	0.001 x 10^n for n ∈ {0, 1, ..., 6}	1000	1000
	γ^a	[0.0001, 0.0005, 0.001, 0.005, 0.01, 0.05, 0.1, 1]	0.005	0.0005
	ε^b	[0.0001, 0.0005, 0.001, 0.005, 0.01, 0.05, 0.1]	0.05	0.05
LGBM	*learning_rate*	[0.0001, 0.0005, 0.001, 0.005, 0.01, 0.05]	0.05	0.05
	max_depth	[10, 50]*	50	10
	n_estimators	[10, 200]*	200	200
	num_leaves	[10, 100]*	100	10
XGB	*learning rate*	[0.0001, 0.0005, 0.001, 0.005, 0.01, 0.05]	0.05	0.05
	max_depth	[10, 50]*	10	10

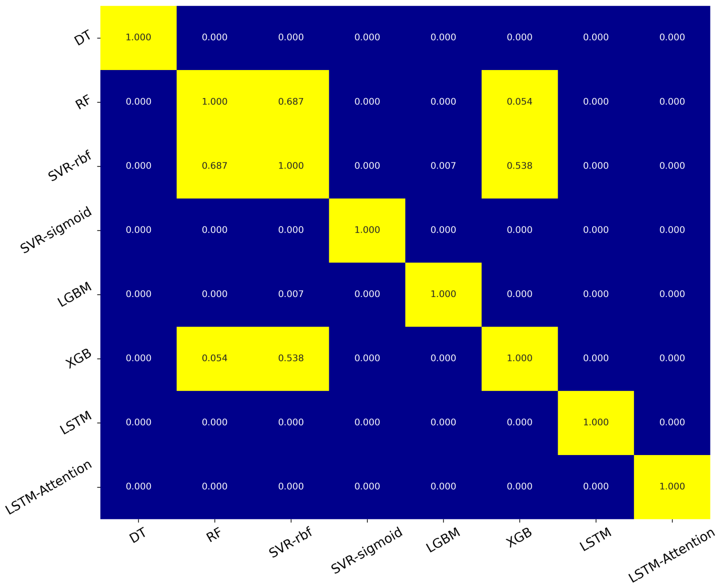
Table B.1: Optimal hyperparameters for different machine learning models using simulated and measured data sets.

Model	Hyperparameter	Search range	Simulated best fit	Measured best fit
	n_estimators	[10, 200]*	140	120
LSTM	*batch_size*	64, 128, 256, 512	64	512
	learning_rate	[0.0001, 0.0005, 0.001, 0.005, 0.01, 0.05]	0.0005	0.01
	hidden_units	[8, 128]**	48	56
LSTM-Attention	*batch_size*	64, 128, 256, 512	64	64
	learning_rate	[0.0001, 0.0005, 0.001, 0.005, 0.01, 0.05]	0.0001	0.001
	hidden_units	[8, 128]**	112	96

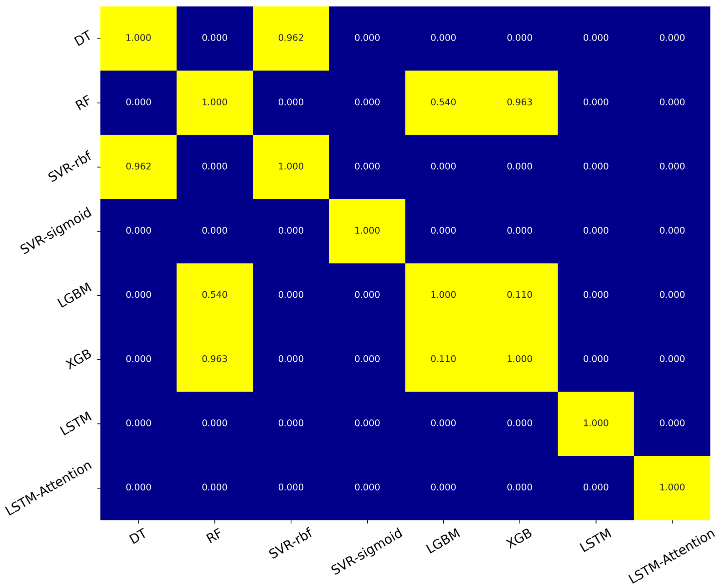
$^{a}gamma$; $^{b}epsilon$

$^{*}step = 10$; $^{**}step = 8$

Appendix C. Wilcoxon Signed-Rank Test



(a) Simulated



(b) Measured

Figure C.1: Wilcoxon Signed-Rank test when comparing errors from different models. The matrix shows p-values (rounded to the third decimal) for: a) simulated data set, and b) measured data set.

771 # Appendix D. Power Spectral Density

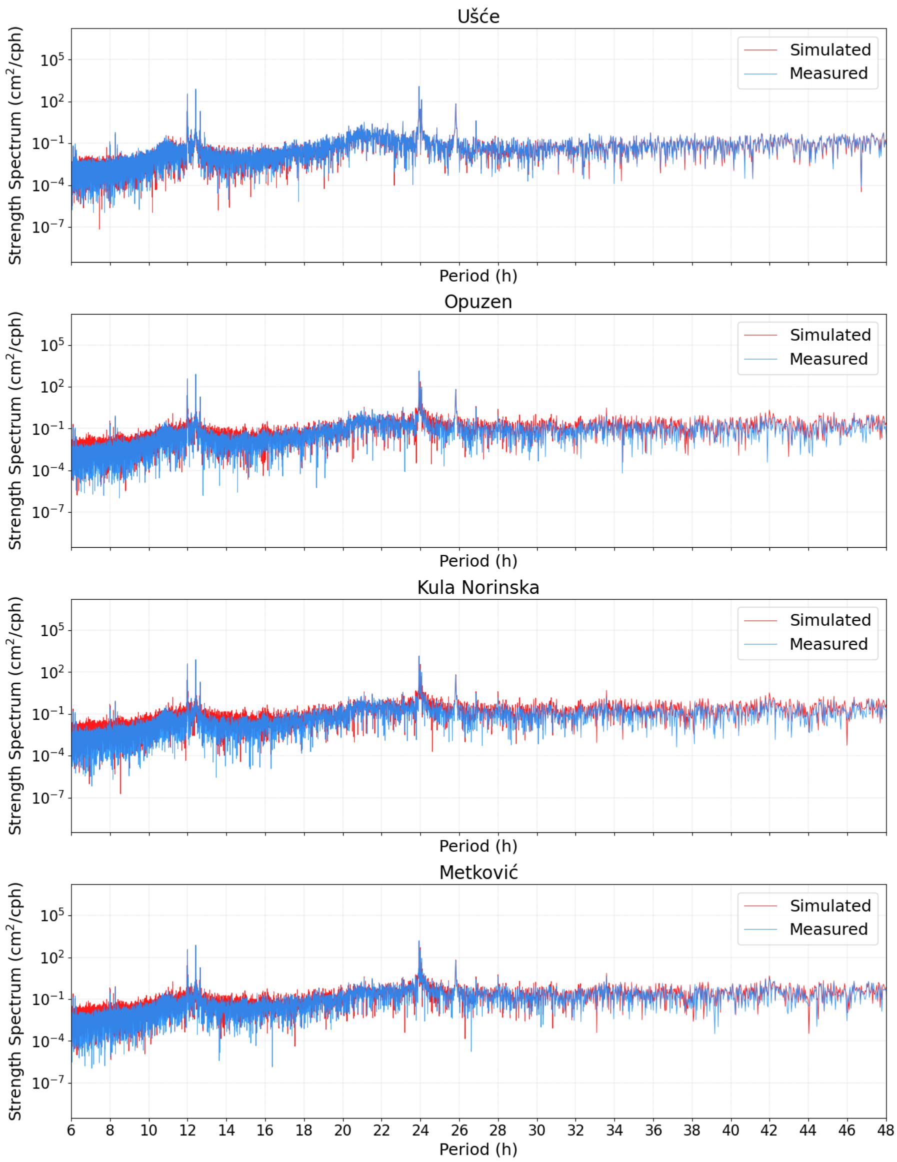


Figure D.1: Power spectral density for the water levels from the tidal station to the most upstream part of the tidal reach, their distribution over different periods (from 6 to 48 h), and comparison between simulated and measured datasets.

772 **Appendix E. Correlation Matrix**

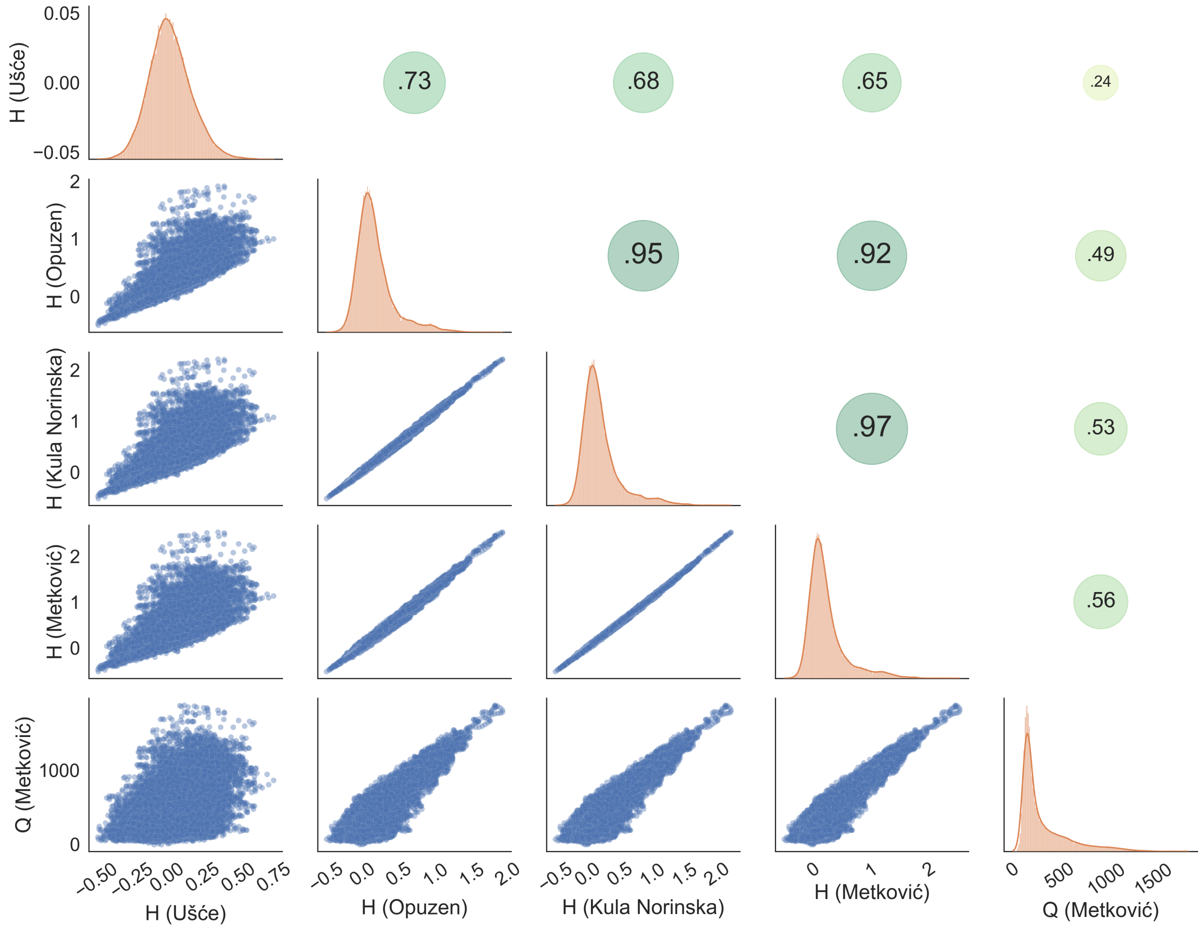


Figure E.1: Correlation matrix for simulated water level and discharge data at different stations (Ušće, Opuzen, Kula Norinska, and Metković). The parameter H denotes water level, while Q denotes discharge. Lower left panels show correlation plots, diagonal panels show distribution plots for each parameter, and upper right panels show correlation coefficients.

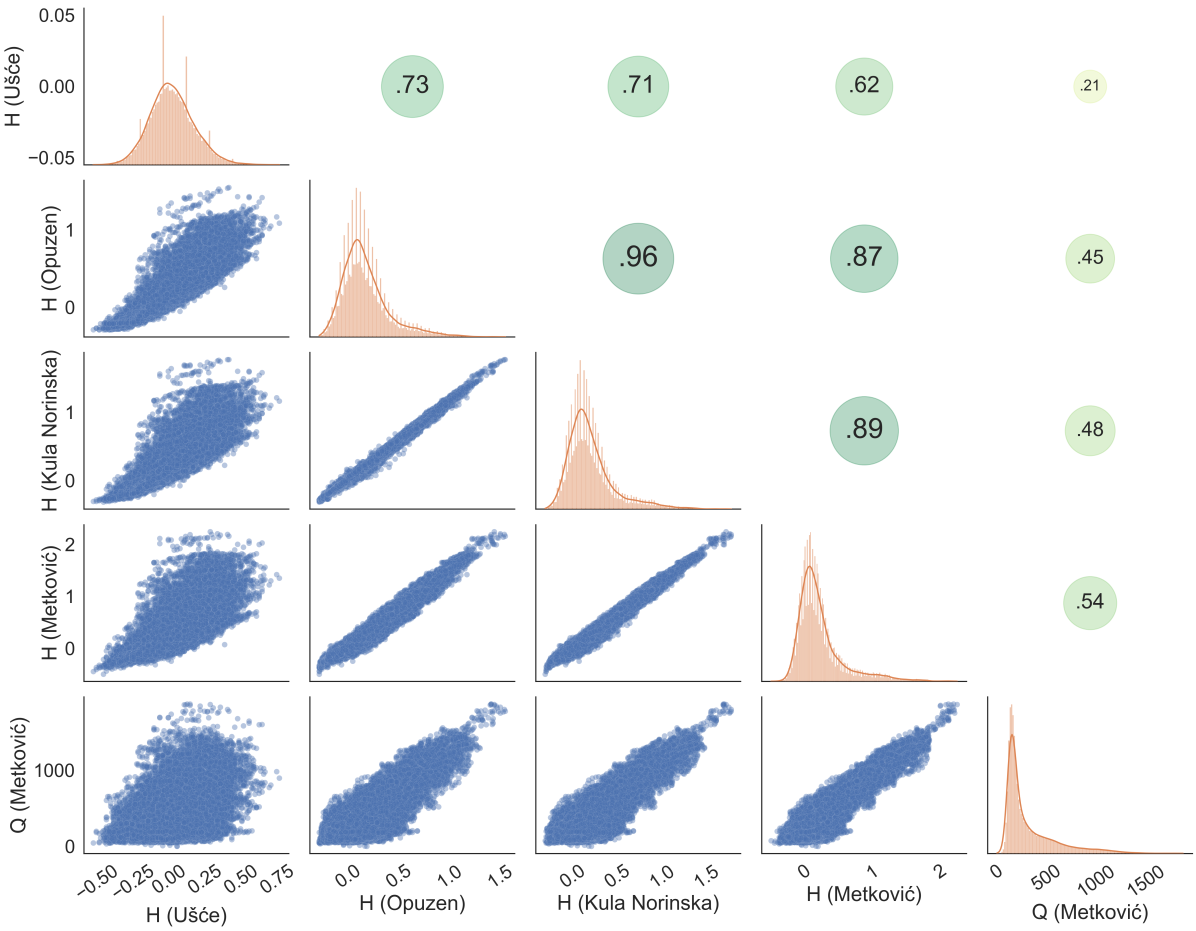


Figure E.2: Correlation matrix for measured water level and discharge data at different stations (Ušće, Opuzen, Kula Norinska, and Metković). The parameter H denotes water level, while Q denotes discharge. Lower left panels show correlation plots, diagonal panels show distribution plots for each parameter, and upper right panels show correlation coefficients.

773 # Appendix F. Feature Importance Based on Mutual Information

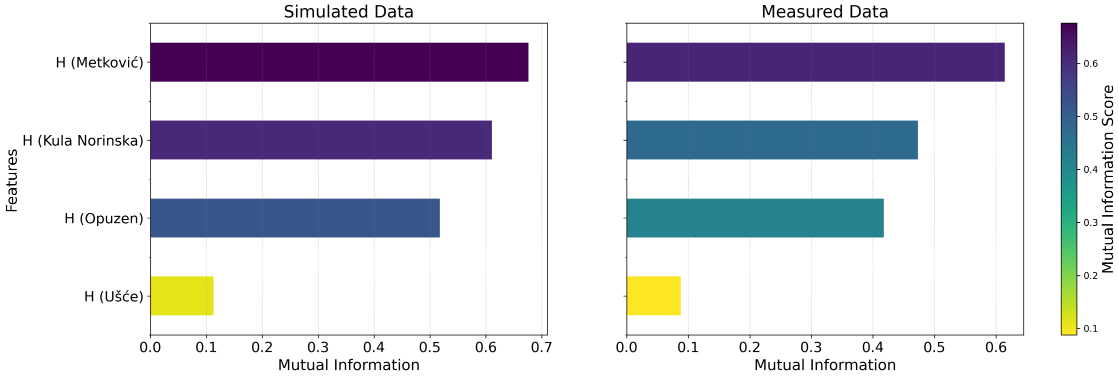


Figure F.1: Feature importance based on mutual information for discharge estimation with ranking of features from the most to the least important for two datasets: (a) Simulated and (b) Measured
The parameter H denotes water level at a station in brackets.

Appendix G. Explainable Artificial Intelligence: SHAP method

This section provides SHAP analysis for simple ML models, with global and local explainability. The global analysis provides the overall feature importance, while local is focused on each prediction and their contribution to estimation.

G.1 Simulated data scenario

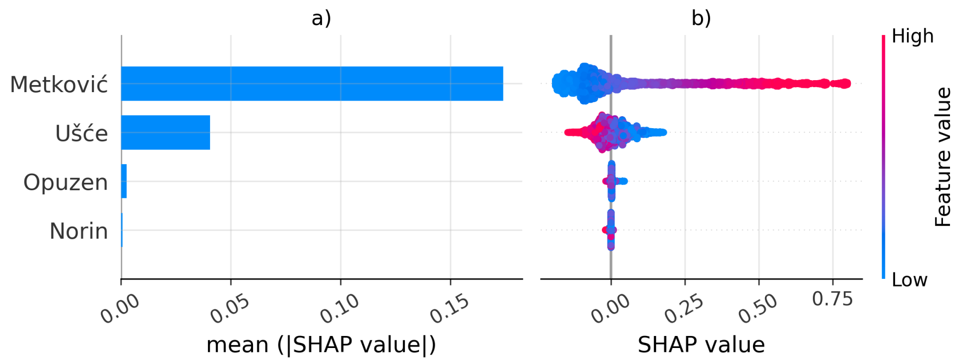


Figure G.1: DT SHAP analysis: a) Global and, b) Local Interpretation

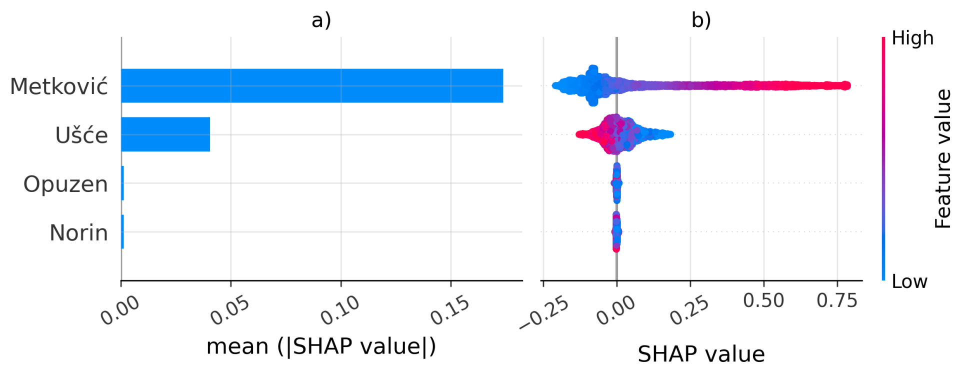


Figure G.2: RF SHAP analysis: a) Global and, b) Local Interpretation

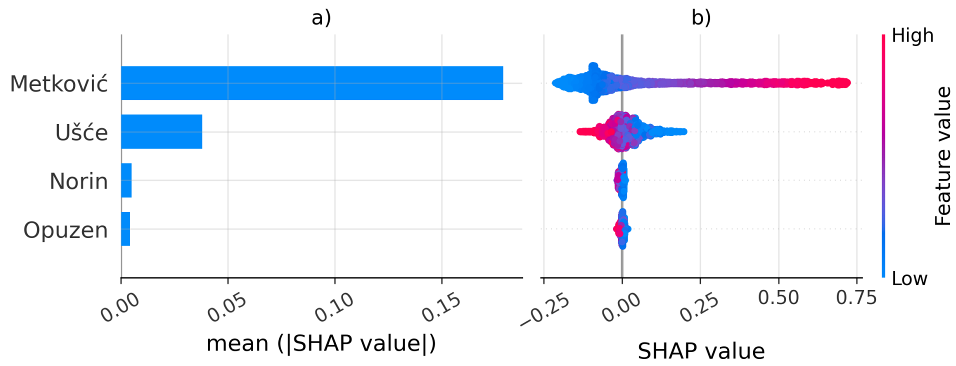


Figure G.3: LGBM SHAP analysis: a) Global and, b) Local Interpretation

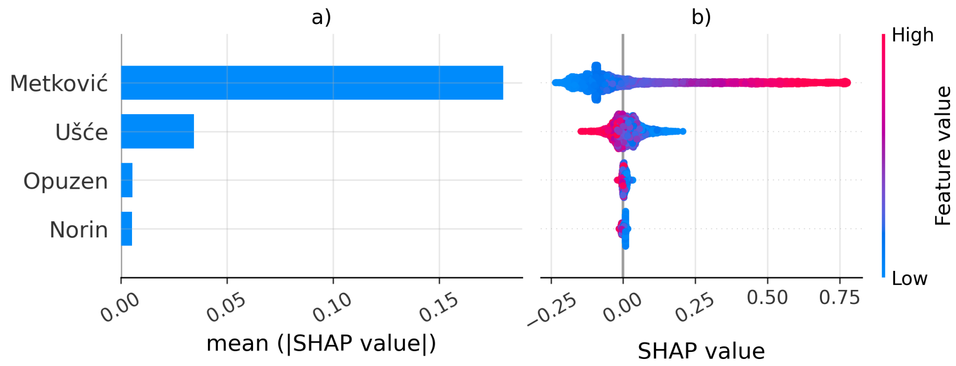


Figure G.4: XGB SHAP analysis: a) Global and, b) Local Interpretation

779 **G.2 Measured data scenario**

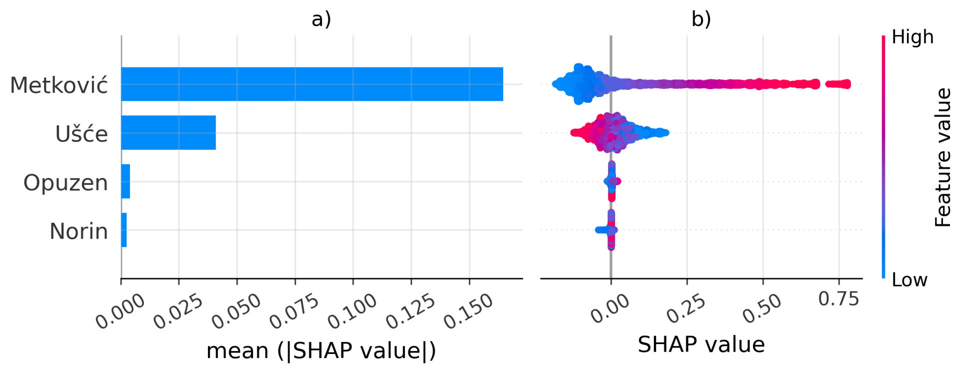


Figure G.5: DT SHAP analysis: a) Global and, b) Local Interpretation

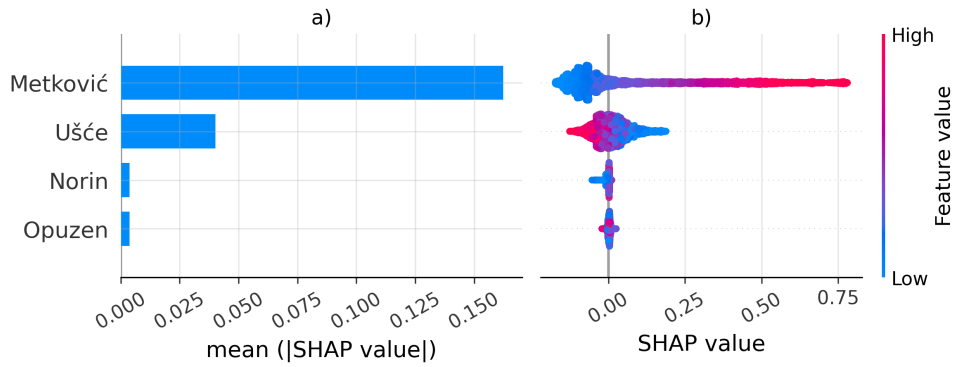


Figure G.6: RF SHAP analysis: a) Global and, b) Local Interpretation

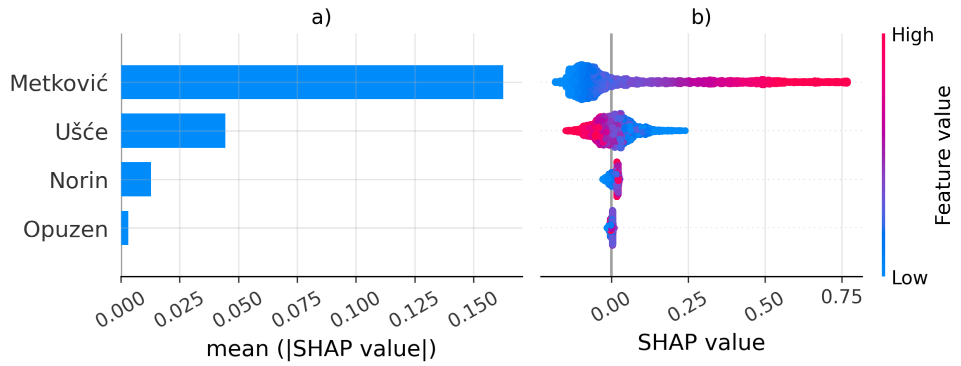


Figure G.7: LGBM SHAP analysis: a) Global and, b) Local Interpretation

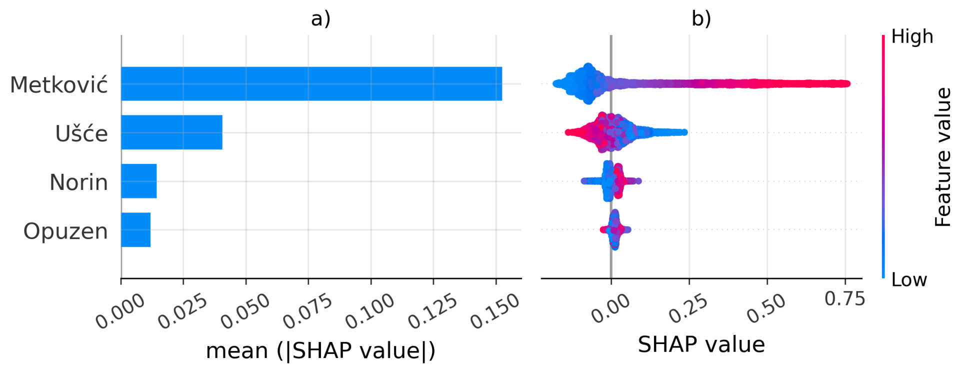


Figure G.8: XGB SHAP analysis: a) Global and, b) Local Interpretation

Appendix H. Feature Occlusion

Metrics	Single Input[a]	Two Inputs[b]	Three Inputs[c]	Four Inputs[d]
RMSE	118.727	31.267	33.708	34.384
MAE	83.864	24.125	25.894	27.054
NSE	0.890	0.992	0.991	0.991
R	0.962	0.996	0.996	0.996

[a] Input feature: Metković
[b] Input features: Ušće and Metković
[c] Input features: Ušće, Opuzen, and Metković
[d] Input features: Ušće, Opuzen, Norin, and Metković

Table H.1
Effect of feature occlusion on LSTM model performance on simulated dataset. Four different scenarios are tested, the first contains only the target station data, the second contains the target and a tidal station, the third contains the target, one midstream and a tidal station, and the fourth contains the target, two midstreams and a tidal station.

Metrics	Single Input[a]	Two Inputs[b]	Three Inputs[c]	Four Inputs[d]
RMSE	117.506	73.497	64.744	63.495
MAE	84.356	54.454	48.684	47.495
NSE	0.892	0.958	0.967	0.969
R	0.963	0.982	0.987	0.988

[a] Input feature: Metković
[b] Input features: Ušće and Metković
[c] Input features: Ušće, Opuzen, and Metković
[d] Input features: Ušće, Opuzen, Norin, and Metković

Table H.2
Effect of feature occlusion on LSTM model performance on measured dataset. Four different scenarios are tested, the first contains only the target station data, the second contains the target and a tidal station, the third contains the target, one midstream and a tidal station, and the fourth contains the target, two midstreams and a tidal station.

Metrics	Single Input[a]	Two Inputs[b]	Three Inputs[c]	Four Inputs[d]
RMSE	126.045	33.254	30.334	29.473
MAE	89.793	24.893	23.157	22.530
NSE	0.876	0.991	0.993	0.993
R	0.948	0.996	0.997	0.997

[a] Input feature: Metković
[b] Input features: Ušće and Metković
[c] Input features: Ušće, Opuzen, and Metković
[d] Input features: Ušće, Opuzen, Norin, and Metković

Table H.3
Effect of feature occlusion on LSTM-Attention model performance on simulated dataset. Four different scenarios are tested, the first contains only the target station data, the second contains the target and a tidal station, the third contains the target, one midstream and a tidal station, and the fourth contains the target, two midstreams and a tidal station.

Metrics	Single Input[a]	Two Inputs[b]	Three Inputs[c]	Four Inputs[d]
RMSE	129.384	68.744	64.709	57.406
MAE	91.337	52.539	48.626	43.201
NSE	0.870	0.963	0.967	0.974
R	0.959	0.985	0.986	0.989

[a] Input feature: Metković
[b] Input features: Ušće and Metković
[c] Input features: Ušće, Opuzen, and Metković
[d] Input features: Ušće, Opuzen, Norin, and Metković

Table H.4
Effect of feature occlusion on LSTM-Attention model performance on measured dataset. Four different scenarios are tested, the first contains only the target station data, the second contains the target and a tidal station, the third contains the target, one midstream and a tidal station, and the fourth contains the target, two midstreams and a tidal station.

References

Ahmed, D., Hassan, M., Mstafa, R., 2022. A review on deep sequential models for forecasting time series data. Applied Computational Intelligence and Soft Computing 2022. doi:10.1155/2022/6596397.

Almeida, J., 2002. Predictive non-linear modeling of complex data by artificial neural networks. Current Opinion in Biotechnology 13, 72–76. doi:10.1016/S0958-1669(02)00288-4.

Bendat, J.S., Piersol, A.G., 2011. Random data: analysis and measurement procedures. John Wiley & Sons.

Cai, H., Savenije, H.H., Zuo, S., Jiang, C., Chua, V.P., 2015. A predictive model for salt intrusion in estuaries applied to the yangtze estuary. Journal of Hydrology 529, 1336–1349.

Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P., 2002. Smote: synthetic minority over-sampling technique. Journal of artificial intelligence research 16, 321–357.

Chen, K., Kuang, C., Wang, L., Chen, K., Han, X., Fan, J., 2022. Storm surge prediction based on long short-term memory neural network in the east china sea. Applied Sciences (Switzerland) 12, 181. doi:10.3390/app12010181.

Chen, Y.C., Yeh, H.C., Kao, S.P., Wei, C., Su, P.Y., 2023. Water level forecasting in tidal rivers during typhoon periods through ensemble empirical mode decomposition. Hydrology 10, 47. doi:10.3390/hydrology10020047.

Choubin, B., Darabi, H., Rahmati, O., Sajedi-Hosseini, F., Kløve, B., 2018. River suspended sediment modelling using the cart model: A comparative study of machine learning techniques. Science of the Total Environment 615, 272–281. doi:10.1016/j.scitotenv.2017.09.293.

Corazza, A., Di Martino, S., Ferrucci, F., Gravino, C., Sarro, F., Mendes, E., 2013. Using tabu search to configure support vector regression for effort estimation. Empirical Software Engineering 18, 506–546.

Du, Y., Cheng, Z., You, Z., 2023. Morphological changes in a macro-tidal estuary during extreme flooding events. Frontiers in Marine Science 9. doi:10.3389/fmars.2022.1112494.

Dutta, D., Arya, S., Kumar, S., 2021. Industrial wastewater treatment: Current trends, bottlenecks, and best practices. Chemosphere 285, 131245.

Forghanparast, F., Mohammadi, G., 2022. Using deep learning algorithms for intermittent streamflow prediction in the headwaters of the colorado river, texas. Water 14, 2972.

Friedman, J., 2001. Greedy function approximation: A gradient boosting machine. Annals of Statistics 29, 1189–1232. doi:10.1214/aos/1013203451.

Gajić-Čapka, M., Güttler, I., Cindrić, K., Branković, C., 2018. Observed and simulated climate and climate change in the lower neretva river basin. Journal of Water and Climate Change 9, 124–136. doi:10.2166/wcc.2017.034.

Gan, M., Pan, S., Chen, Y., Cheng, C., Pan, H., Zhu, X., 2021. Application of the machine learning lightgbm model to the prediction of the water levels of the lower columbia river. Journal of Marine Science and Engineering 9, 496. doi:10.3390/jmse9050496.

Guillou, N., Chapalain, G., Petton, S., 2023. Predicting sea surface salinity in a tidal estuary with machine learning. Oceanologia 65, 318–332. doi:10.1016/j.oceano.2022.07.007.

Guo, W.D., Chen, W.B., Yeh, S.H., Chang, C.H., Chen, H., 2021. Prediction of river stage using multistep-ahead machine learning techniques for a tidal river of taiwan. Water 13, 920.

Gupta, H., Kling, H., Yilmaz, K., Martinez, G., 2009. Decomposition of the mean squared error and nse performance criteria: Implications for improving hydrological modelling. Journal of Hydrology 377, 80–91. doi:10.1016/j.jhydrol.2009.08.003.

Habib, E., Meselhe, E., 2006. Stage - discharge relations for low-gradient tidal streams using data-driven models. Journal of Hydraulic Engineering 132, 482–492. doi:10.1061/(ASCE)0733-9429(2006)132:5(482).

Hannan, A., Anmala, J., 2021. Classification and prediction of fecal coliform in stream waters using decision trees (dts) for upper green river watershed, kentucky, usa. Water (Switzerland) 13, 2790. doi:10.3390/w13192790.

Hidayat, H., Hoitink, A., Sassi, M., Torfs, P., 2014. Prediction of discharge in a tidal river using artificial neural networks. Journal of Hydrologic Engineering 19, 04014006. doi:10.1061/(ASCE)HE.1943-5584.0000970.

Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Computation 9, 1735–1780. doi:10.1162/neco.1997.9.8.1735.

Huang, Y., Pan, J., Devlin, A., 2023. Enhanced estimate of chromophoric dissolved organic matter using machine learning algorithms from landsat-8 oli data in the pearl river estuary. Remote Sensing 15, 1963. doi:10.3390/rs15081963.

Ibrahim, K.S.M.H., Huang, Y.F., Ahmed, A.N., Koo, C.H., El-Shafie, A., 2022. A review of the hybrid artificial intelligence and optimization modelling of hydrological streamflow forecasting. Alexandria Engineering Journal 61, 279–303. URL: https://www.sciencedirect.com/science/article/pii/S111001682100346X, doi:https://doi.org/10.1016/j.aej.2021.04.100.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.Y., 2017. Lightgbm: A highly efficient gradient boosting decision tree, in: 31st Conference on Neural Information Processing Systems (NIPS 2017), pp. 3147–3155.

Khatun, A., Sahoo, B., Chatterjee, C., 2023. Two novel error-updating model frameworks for short-to-medium range streamflow forecasting using bias-corrected rainfall inputs: Development and comparative assessment. Journal of Hydrology 618, 129199.

Khoshkalam, Y., Rousseau, A.N., Rahmani, F., Shen, C., Abbasnezhadi, K., 2023. Applying transfer learning techniques to enhance the accuracy of streamflow prediction produced by long short-term memory networks with data integration. Journal of Hydrology 622, 129682.

Krvavica, N., Gotovac, H., Lončar, G., 2021. Salt-wedge dynamics in microtidal neretva river estuary. Regional Studies in Marine Science 43, 101713. doi:10.1016/j.rsma.2021.101713.

Krvavica, N., Kožar, I., Travaš, V., Ožanić, N., 2017. Numerical modelling of two-layer shallow water flow in microtidal salt-wedge estuaries: Finite volume solver and field validation. Journal of Hydrology and Hydromechanics 65, 49–59.

Krvavica, N., Ružić, I., 2020. Assessment of sea-level rise impacts on salt-wedge intrusion in idealized and neretva river estuary. Estuarine, Coastal and Shelf Science 234, 106638. doi:10.1016/j.ecss.2020.106638.

Kundu, D., Dutta, D., Samanta, P., Dey, S., Sherpa, K.C., Kumar, S., Dubey, B.K., 2022. Valorization of wastewater: A paradigm shift towards circular bioeconomy and sustainability. Science of the Total Environment 848, 157709.

Lee, M., Yoo, Y., Joo, H., Kim, K., Kim, H., Kim, S., 2021. Construction of rating curve at high water level considering rainfall effect in a tidal river. Journal of Hydrology: Regional Studies 37, 100907. doi:10.1016/j.ejrh.2021.100907.

Li, B., Yang, G., Wan, R., Dai, X., Zhang, Y., 2016. Comparison of random forests and other statistical methods for the prediction of lake water level: A case study of the poyang lake in china. Hydrology Research 47, 69–83. doi:10.2166/nh.2016.264.

Li, H., Zhang, L., Zhang, Y., Yao, Y., Wang, R., Dai, Y., 2024. Water-level prediction analysis for the three gorges reservoir area based on a hybrid model of lstm and its variants. Water 16, 1227.

Li, J., Yuan, X., Ji, P., 2023. Long-lead daily streamflow forecasting using long short-term memory model with different predictors. Journal of Hydrology: Regional Studies 48, 101471.

Lindemann, B., Müller, T., Vietz, H., Jazdi, N., Weyrich, M., 2021. A survey on long short-term memory networks for time series prediction, in: 14th CIRP Conference on Intelligent Computation in Manufacturing Engineering, 15-17 July 2020, pp. 650–655. doi:10.1016/j.procir.2021.03.088.

Ljubenkov, I., Vranješ, M., 2012. Numerical model of stratified flow - case study of the neretva riverbed salination (2004) | numerički model uslojenog tečenja - primjer zaslanjivanja korita rijeke neretve (2004). Gradjevinar 64, 101–112. doi:10.14256/jce.639.2011.

Lovrinović, I., Srzić, V., Aljinović, I., 2023. Characterization of seawater intrusion dynamics under the influence of hydro-meteorological conditions, tidal oscillations and melioration system operative regimes to groundwater in neretva valley coastal aquifer system. Journal of Hydrology: Regional Studies 46, 101363. URL: https://www.sciencedirect.com/science/article/pii/S2214581823000502, doi:https://doi.org/10.1016/j.ejrh.2023.101363.

Lundberg, S.M., Lee, S.I., 2017. A unified approach to interpreting model predictions, in: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (Eds.), Advances in Neural Information Processing Systems 30. Curran Associates, Inc., pp. 4765–4774. URL: http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf.

Luong, M.T., Pham, H., Manning, C.D., 2015. Effective approaches to attention-based neural machine translation, in: Conference Proceedings - EMNLP 2015: Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics (ACL). pp. 1412–1421. doi:10.18653/v1/d15-1166.

Maier, H.R., Taghikhah, F.R., Nabavi, E., Razavi, S., Gupta, H., Wu, W., Radford, D.A., Huang, J., 2024. How much x is in xai: Responsible use of "explainable" artificial intelligence in hydrology and water resources. Journal of Hydrology X , 100185.

Malek, N., Yaacob, W., Nasir, S., Shaadan, N., 2022. Prediction of water quality classification of the kelantan river basin, malaysia, using machine learning techniques. Water (Switzerland) 14, 1067. doi:10.3390/w14071067.

van Maren, D., Beemster, J., Wang, Z., Khan, Z., Schrijvershof, R., Hoitink, A., 2023. Tidal amplification and river capture in response to land reclamation in the ganges-brahmaputra delta. Catena 220, 106651. doi:10.1016/j.catena.2022.106651.

Matte, P., Secretan, Y., Morin, J., 2014. Temporal and spatial variability of tidal-fluvial dynamics in the st. lawrence fluvial estuary: An application of nonstationary tidal harmonic analysis. Journal of Geophysical Research: Oceans 119, 5724–5744.

Medvedev, I.P., Vilibić, I., Rabinovich, A.B., 2020. Tidal resonance in the adriatic sea: Observational evidence. Journal of Geophysical Research: Oceans 125, e2020JC016168.

Mihel, A.M., Lerga, J., Krvavica, N., 2024. Estimating water levels and discharges in tidal rivers and estuaries: Review of machine learning approaches. Environmental Modelling Software 176, 106033. URL: https://www.sciencedirect.com/science/article/pii/S136481522400094X, doi:https://doi.org/10.1016/j.envsoft.2024.106033.

Mohanty, A., Sahoo, B., Kale, R.V., 2024. A hybrid model enhancing streamflow forecasts in paddy land use-dominated catchments with numerical weather prediction model-based meteorological forcings. Journal of Hydrology 635, 131225.

Nagaradjane, A., Candassamy, A., Dhamodharan, K., Shanmugaraj, A.K., 2024. Smart heart disease prediction and amalgamation tracking system. Engineering review 44, 14–24. doi:10.30765/er.2110.

Nayebi, A., Tipirneni, S., Reddy, C.K., Foreman, B., Subbian, V., 2023. Windowshap: An efficient framework for explaining time-series classifiers based on shapley values. Journal of Biomedical Informatics 144, 104438.

Ng, K., Huang, Y., Koo, C., Chong, K., El-Shafie, A., Najah Ahmed, A., 2023. A review of hybrid deep learning applications for streamflow forecasting. Journal of Hydrology 625, 130141. URL: https://www.sciencedirect.com/science/article/pii/S0022169423010831, doi:https://doi.org/10.1016/j.jhydrol.2023.130141.

Panagopoulos, A., Giannika, V., 2024. A comprehensive assessment of the economic and technical viability of a zero liquid discharge (zld) hybrid desalination system for water and salt recovery. Journal of Environmental Management 359, 121057.

Piraei, R., Niazkar, M., Afzali, S., Menapace, A., 2023. Application of machine learning models to bridge afflux estimation. Water (Switzerland) 15, 2187. doi:10.3390/w15122187.

Sabzipour, B., Arsenault, R., Troin, M., Martel, J.L., Brissette, F., Brunet, F., Mai, J., 2023. Comparing a long short-term memory (lstm) neural network with a physically-based hydrological model for streamflow forecasting over a canadian catchment. Journal of Hydrology 627, 130380. URL: https://www.sciencedirect.com/science/article/pii/S0022169423013227, doi:https://doi.org/10.1016/j.jhydrol.2023.130380.

Sattari, M., Feizi, H., Colak, M., Ozturk, A., Apaydin, H., Ozturk, F., 2020. Estimation of sodium adsorption ratio in a river with kernel-based and decision-tree models. Environmental Monitoring and Assessment 192, 575. doi:10.1007/s10661-020-08506-9.

Sherstinsky, A., 2020. Fundamentals of recurrent neural network (rnn) and long short-term memory (lstm) network. Physica D: Nonlinear Phenomena 404, 132306. doi:10.1016/j.physd.2019.132306.

Thanh, H., Binh, D., Kantoush, S., Nourani, V., Saber, M., Lee, K.K., Sumi, T., 2022. Reconstructing daily discharge in a megadelta using machine learning techniques. Water Resources Research 58. doi:10.1029/2021WR031048.

Thomas, M., Joy, A.T., 2006. Elements of information theory. Wiley-Interscience.

Tian, Y., Zhang, Q., Huang, H., Huang, Y., Tao, J., Zhou, G., Zhang, Y., Yang, Y., Lin, J., 2022. Aboveground biomass of typical invasive mangroves and its distribution patterns using uav-lidar data in a subtropical estuary: Maoling river estuary, guangxi, china. Ecological Indicators 136, 108694. doi:10.1016/j.ecolind.2022.108694.

Tran, D., Tsujimura, M., Ha, N., Nguyen, V., Binh, D., Dang, T., Doan, Q.V., Bui, D., Ngoc, T.A., Phu, L., Thuc, P., Pham, T., 2021. Evaluating the predictive power of different machine learning algorithms for groundwater salinity prediction of multi-layer coastal aquifers in the mekong delta, vietnam. Ecological Indicators 127, 107790. doi:10.1016/j.ecolind.2021.107790.

Vapnik, V.N., 2000. The Nature of Statistical Learning Theory. 2 ed., Springer New York, NY.

Vercruysse, K., Grabowski, R., 2021. Human impact on river planform within the context of multi-timescale river channel dynamics in a himalayan river system. Geomorphology 381, 107659. doi:10.1016/j.geomorph.2021.107659.

Villeneuve, Y., Séguin, S., Chehri, A., 2023. Ai-based scheduling models, optimization, and prediction for hydropower generation: Opportunities, issues, and future directions. Energies 16, 3335. doi:10.3390/en16083335.

Vu, M., Jardani, A., Krimissa, M., Zaoui, F., Massei, N., 2023. Large-scale seasonal forecasts of river discharge by coupling local and global datasets with a stacked neural network: Case for the loire river system. Science of the Total Environment 897.

Wolfs, V., Willems, P., 2014. Development of discharge-stage curves affected by hysteresis using time varying models, model trees and neural networks. Environmental Modelling and Software 55, 107–119. doi:10.1016/j.envsoft.2014.01.021.

Xu, Y., Lin, K., Hu, C., Wang, S., Wu, Q., Zhang, J., Xiao, M., Luo, Y., 2024. Interpretable machine learning on large samples for supporting runoff estimation in ungauged basins. Journal of Hydrology 639, 131598.

Yang, Y., Gao, Y., Wang, Z., Li, X., Zhou, H., Wu, J., 2024. Multiscale-integrated deep learning approaches for short-term load forecasting. International Journal of Machine Learning and Cybernetics , 1–16.

Yaseen, Z.M., El-Shafie, A., Jaafar, O., Afan, H.A., Sayl, K.N., 2015. Artificial intelligence based models for stream-flow forecasting: 2000–2015. Journal of Hydrology 530, 829–844.

Yoo, H., Kim, D., Kwon, H.H., Lee, S., 2020. Data driven water surface elevation forecasting model with hybrid activation function-a case study for hangang river, south korea. Applied Sciences (Switzerland) 10, 1424. doi:10.3390/app10041424.

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.

Zovko, M., Romić, D., Colombo, C., Iorio, E.D., Romić, M., Buttafuoco, G., Castrignanò, A., 2018. A geostatistical vis-nir spectroscopy index to assess the incipient soil salinization in the neretva river valley, croatia. Geoderma 332, 60–72. doi:10.1016/j.geoderma.2018.07.005.

Declaration of interests

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

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: