UNIVERSITY OF RIJEKA
FACULTY OF ENGINEERING

Anna Maria Mihel

Estimation and prediction of discharges in tidal rivers and estuaries using machine learning

DOCTORAL THESIS

Rijeka 2025.

UNIVERSITY OF RIJEKA
FACULTY OF ENGINEERING

Anna Maria Mihel

Estimation and prediction of discharges in tidal rivers and estuaries using machine learning

DOCTORAL THESIS

Supervisor: Prof. Jonatan Lerga, PhD
Co-supervisor: Assoc. Prof. Nino Krvavica, PhD

Rijeka 2025.

SVEUČILIŠTE U RIJECI
TEHNIČKI FAKULTET

Anna Maria Mihel

Procjena i predviđanje protoka u rijekama pod usporom mora i prijelaznim vodama korištenjem strojnog učenja

DOKTORSKI RAD

Mentor: Prof. dr. sc. Jonatan Lerga
Komentor: Izv. prof. dr. sc. Nino Krvavica

Rijeka 2025.

Doctoral thesis supervisor: Prof. Jonatan Lerga, PhD (Faculty of Engineering, University of Rijeka)
Doctoral thesis co-supervisor: Assoc. Prof. Nino Krvavica, PhD (Faculty of Civil Engineering, University of Rijeka)

The doctoral thesis was defended on _____ at the University of Rijeka, Faculty of Engineering, Croatia, in front of the following Evaluation Committee:

1. Prof. Goran Mauša, PhD (Faculty of Engineering, University of Rijeka) - Committee Chair

2. Prof. Siniša Družeta, PhD (Faculty of Engineering, University of Rijeka)

3. Prof. Ljiljana Šerić, PhD (Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture, University of Split)

ACKNOWLEDGEMENTS

First of all, I would like to express my gratitude to my mentors, Prof. Jonatan Lerga, PhD, and Assoc. Prof. Nino Krvavica, PhD, for their professional and exceptional guidance, as well as for all the advice that steered me onto the right path during this doctoral study. Their patience, encouragement, and support have meant so much to me throughout this journey.

I would also like to express my gratitude to the entire Department of Computer Science for providing opportunities for academic growth and collaboration, as well as for enabling me to participate in various professional and educational activities, which I will always remember fondly.

I would also like to thank my boyfriend Goran and his family, with whom I shared many wonderful moments on numerous trips. These experiences restored my motivation. A big thanks also goes to everyone who has stood by my side and cheered me on over the years.

Finally, I owe my deepest gratitude to my parents, who have been my greatest source of strength and support throughout this entire academic journey and who have never once doubted me. Thank you for your unconditional love, support and patience - this is what motivated me not to give up, even in the most difficult moments. Mom, thank you for the countless meals that always kept me nourished, and Dad, thank you for being my personal taxi driver whenever I needed it.

ABSTRACT

Estimation and prediction of water discharges are particularly challenging in tidal rivers and estuaries as the water level is significantly affected by the nonlinear interaction between tides, storm surges, and river flow. Therefore, simple hydrological approaches, such as rating curves, are not effective for tidal rivers and estuaries; hence, as an alternative, this thesis proposes a machine learning solution.

The study was focused on developing a robust hybrid machine learning-based approach for the estimation and prediction of hydrological parameters, more precisely, river discharge and water level, in data-limited tidal rivers and estuaries. The proposed hybrid methodology, combining long short-term memory and attention mechanism, was compared to stand-alone long short-term memory and simple non-temporal machine learning models. The effectiveness of the proposed models was evaluated using simulated and measured water level and discharge data from the Neretva River estuary. Multiple station water level data were included as the only input features for the estimation problem, to test whether it was possible to achieve high accuracy of estimation for the most upstream station of the tidal reach, even without additional meteorological or salinity data. The results of the discharge estimation problem revealed the proposed hybrid solution to be the best-performing among the tested models, achieving the highest accuracy for both simulated and measured datasets, for all flow conditions in a microtidal environment. The second-best model was deemed a long short-term memory. This was confirmed by different evaluation metrics, a statistical significance test, and visual inspection. The conducted analysis demonstrated that the proposed hybrid solution provided better generalization and even extrapolation ability when trained on time-series data.

For the forecasting problem, a hybrid approach combining a convolutional neural network and long short-term memory was tested on another simulated dataset, for both

water level and discharge prediction. The forecasting scenarios that employed the specified hybrid approach demonstrated that, integrating the time-frequency representation improved prediction accuracy when non-stationary data was given as input. Likewise, by additionally employing feature engineering, the previously used time-series data results improved, especially for the extended forecasting horizon.

In addition, another scenario with a signal preprocessing technique, variational mode decomposition, was employed for the best-performing models and the estimation problem, aiming to enhance the estimation accuracy. The application of a signal preprocessing technique and the classification of its results as tidal components improved the estimation accuracy of long short-term memory and hybrid long short-term memory combined with attention mechanism for intra-daily discharge oscillations. However, regarding the intra-daily discharge variations, only the long short-term memory performance improved considerably. The hybrid long short-term memory combined with the attention mechanism model demonstrated competitive accuracy, with a Nash-Sutcliffe coefficient of efficiency above 0.970; however, its performance in improving estimation plateaued when trained on simpler and localized tidal components. Therefore, a simpler architecture, such as long short-term memory, yielded greater benefits when applying signal decomposition. The findings of this study indicate a strong potential for practical application, especially in the context of the numerous tidal rivers and estuaries, where reliance on expensive continuous discharge measurement systems could be substantially reduced by employing machine learning techniques.

Keywords: Discharge, Water level, Tidal rivers, Estuaries, Hybrid machine learning approaches, Signal preprocessing

PROŠIRENI SAŽETAK

Održivo upravljanje vodnim resursima zahtijeva temeljito razumijevanje dinamike toka u rijekama koje su pod usporom mora i u prijelaznim vodama. Sve češće i izraženije klimatske promjene, uz antropogene utjecaje, dodatno ističu potrebu za razvojem inovativnih pristupa koji omogućuju precizne i pouzdane procjene hidroloških parametara. U ovom doktorskom radu predstavljen je detaljan pregled literature koji ukazuje na sve širu primjenu modela strojnog učenja u hidrologiji, upravo zbog prednosti koje nude u modeliranju složenih i nelinearnih hidroloških procesa, ali također i inverznih problema, u odnosu na modele utemeljene na fizikalnim zakonima.

Problem procjene i predviđanja protoka predstavlja izazov za područja rijeka pod usporom mora i prijelazne vode budući da na razinu vode značajno utječe složena i nelinearna interakcija između plime i oseke, olujnih uspora i protoka rijeke. Zbog toga jednostavni hidrološki pristupi, poput protočnih krivulja, nisu dovoljno učinkoviti za takva područja. Naime, metode strojnog učenja predstavljaju alternative kojima je moguće modelirati složene riječne procese za područja rijeke koja su u fokusu ovog istraživanja.

U ovom radu dan je detaljan pregled dostupnih istraživanja koja se bave problemima procjene i predviđanja razine vode i protoka rijeke uz primjenu metoda strojnog učenja za područja rijeka pod usporom mora i prijelaznih voda. Kritična analiza, utemeljena na zaključcima iz prethodnih studija, ukazuje na to da hibridni pristupi koji proširuju strojno učenje dodatnim modelima i metodama, uključujući numeričke modele, druge modele strojnog učenja, optimizacijske metode i metode predobrade signala, potencijalno predstavljaju bolje rješenje, što je bila motivacija za provedeno istraživanje.

Područje istraživanja obuhvaća nizvodni tok rijeke Neretve, koji je pod utjecajem mikroplimnog Jadranskog mora, a obilježen je snažnom vertikalnom stratifikacijom slatke i slane vode te izraženim sezonskim karakterom protoka. Za vrijeme vodnog razdoblja

poveća se opasnost od poplava, dok za vrijeme sušnog, javlja se problem prodora slane vode. U svrhu procjene protoka na najuzvodnijoj stanici Metković korišteni su podatci o razini vode s s mareografske postaje i više hidroloških postaja duž rijeke Neretve od Opuzena do Metkovića, kako bi se ispitalo je li moguće postići zadovoljavajuću točnost za manja riječna područja pod usporom mora i prijelaznih voda i u slučajevima kada dodatni parametri poput meteoroloških podataka ili podataka o salinitetu nisu dostupni. Cilj ovog istraživanja bio je razviti matematičke modele temeljene na strojnom učenju za procjenu i predviđanje hidroloških parametara, kao što su protok i razine vode, za podatkovno ograničena područja.

Predložena hibridna metodologija, koja kombinira model duge kratkoročne memorije i mehanizam pažnje, testirana je i uspoređena sa samostalnim modelima strojnog učenja, kako vremenski ovisnima, tako i onima neovisnima o vremenskoj dimenziji. Učinkovitost predloženih modela analizirana je korištenjem simuliranih i izmjerenih podataka o razini vode i protoku za mikroplimnu rijeku Neretvu. Za problem predviđanja, hibridni pristup koji kombinira konvolucijsku neuronsku mrežu i model duge kratkoročne memorije testiran je na zasebnom simuliranom skupu podataka, s ciljem predviđanje razine vode i protoka. Osim toga, za modele s najboljim rezultatima u scenariju procjene protoka dodatno je primijenjena tehnika predobrade signala na razine vodostaja, koji su zatim klasificirani prema kategorijama plimnih komponenti, s ciljem poboljšanja točnosti procjene.

Predloženi hibridni pristup, pokazao se kao pristup s najboljim performansama među ispitivanim modelima za problem procjene protoka u cijelom rasponu vrijednosti, i za mjerene i za simulirane podatke. Kao drugi najbolje ocijenjeni model pokazao se model duge kratkoročne memorije, čije su performanse izražene kroz postotno pogoršanje, bile slabije za 14,28% prema korijenskoj srednjoj kvadratnoj pogrešci, 16,73% prema srednjoj apsolutnoj pogrešci, uz minimalna pogoršanja u pogledu Nash-Sutcliffe koeficijenta efikasnosti od oko 0,20% te Pearsonova koeficijenta korelacije od oko 0,10%, u odnosu na najbolje rangirani hibridni pristup. Usporedno s najboljim modelom, jednostavni modeli strojnog učenja koji ne uzimaju u obzir vremensku zavisnost značajno su odstupali u pogledu točnosti s prosječnim smanjenjem točnosti procjene za sve modele, pri čemu su vrijednosti korijenske srednje kvadratne pogreške, bile lošije za 45,73%, srednje apsolutne pogreške za 38,77%, Nash-Sutcliffe koeficijenta efikasnosti za 1,69% i Pearsonova koeficijenta korelacije za 0,81%. Osim mjera točnosti procjene, proveden je i test statističke

značajnosti te dani brojni grafički prikazi kojima su dodatno potvrđeni zaključci temeljeni na mjerama točnosti. Predloženi hibridni pristup nadmašio je ostale modele, pokazavši bolju sposobnost za generalizaciju i ekstrapolaciju, što je od osobite važnosti za kategorije protoka koje su bile najmanje zastupljene u korištenim skupovima podataka. Također, provedene su i analize doprinosa značajki prije i nakon postupka treniranja modela. Rezultati su pokazali kako razina vode na samom ušću rijeke ima značajan doprinos za procjenu protoka na najuzvodnijoj stanici Metković, uz razinu vode s iste lokacije.

Scenariji predviđanja u kojima je primijenjen hibridni pristup koji kombinira konvolucijsku neuronsku mrežu i model duge kratkoročne memorije dokazali su da uključivanje vremensko-frekvencijske domene može poboljšati točnost predviđanja nestacionarnih podataka te da primjena spektrograma predstavlja značajan potencijal za daljnja istraživanja. Dodatnom primjenom postupka inženjerstva značajki, unaprijeđeni su prethodni rezultati dobivenih na temelju modeliranja vremenskih serija, posebice za duže horizonte predviđanja.

Primjena tehnike predobrade signala i klasifikacija dobivenih plimnih komponenti dodatno su poboljšala točnost procjene pomoću modela duge kratkoročne memorije i hibridnog model duge kratkoročne memorije s mehanizmom pažnje u pogledu unutardnevne oscilacije protoka. Međutim, za među-dnevne varijacije protoka, značajno su se poboljšale samo performanse modela duge kratkoročne memorije, 21.99% za korijensku srednju kvadratnu pogrešku, 21.01% za srednju apsolutnu pogrešku, 1.24% za Nash-Sutcliffe koeficijent efikasnosti, te 0.30% za Pearsonov koeficijent korelacije, gledajući skup podataka za validaciju, u odnosu na scenarij gdje se koriste samo vremenski nizovi. Suprotno tome, hibridni pristup ostvario je bolje rezultate kada su korišteni izvorni podaci u formatu vremenskih nizova u odnosu na dekomponirane plimne komponente. Iako je hibridni pristup postigao nešto nižu, ali konkurentnu točnost, s Nash-Sutcliffe koeficijentom efikasnosti iznad 0.970, njegova učinkovitost u poboljšanju stabilnosti procjene nije se dodatno poboljšala kada je treniran na pojednostavljenim i lokaliziranim plimnim komponentama. Rezultati provedene studije imaju veliki potencijal za praktičnu primjenu, s obzirom na broj rijeka koje su pod usporom mora, kao i prijelaznih voda, gdje bi se potreba za skupim kontinuiranim mjerenjima protoka mogla smanjiti primjenom metoda strojnog učenja.

Ključne riječi: Protok, Razina vode, Rijeke pod usporom mora, Prijelazne

vode, Hibridni pristupi strojnog učenja, Predobrada signala

Contents

Chapter 1

INTRODUCTION

Contents

This chapter introduces the key concepts and structure of the thesis, establishing the foundation of the study. It begins by defining the context and motivation behind the research problem. Following this, the research aims, hypotheses, and scientific contributions are outlined. The chapter concludes with a brief overview of the research methodology and the overall thesis structure.

1.1.   Context and Motivation

Transitional areas of tidal rivers and estuaries are characterized by complex flows, primarily due to the nonlinear interaction between water levels and upstream discharge, as well as the downstream tidal influence. However, other forcing mechanisms also contribute to this complexity, as these areas are susceptible to extreme weather events due to more frequent climate changes. Over time, multiple factors of the forcing mechanism may exhibit variability and deviation from stationarity. These include atmospheric conditions

(such as air pressure and precipitation), internal water properties (such as temperature and salinity), and various other factors. Due to proximity to the sea, these areas are vulnerable to salinity intrusion, which has become a pronounced problem in recent decades, mainly due to the effects of climate change. The vulnerability of tidal rivers and estuaries, combined with their complex flow patterns, creates a constant and growing need for effective water resource management and protection strategies. The main goal is to mitigate the adverse effects of natural disasters, extreme climate events, and anthropogenic pressures. Continuous monitoring of hydrological parameters, particularly water level and discharge, is instrumental in achieving this goal.

Water level reflects the height of the water surface relative to a pre-established reference level [90], while discharge reflects the general response of the watershed (hydrologic processes) and the attenuation of flow (hydraulic processes) [40]. Unexpected changes in water level or discharge can negatively affect sectors sensitive to sudden fluctuations, such as agriculture, where it disrupts irrigation, and urban infrastructure, where it can cause flooding or an insufficient water supply. Therefore, continuous monitoring, combined with accurate estimation and prediction of these parameters, contributes to risk mitigation in these vulnerable sectors and enables early detection of potential hazards.

A problem that arises with continuous monitoring concerns the discharge parameter, also known as streamflow or flow rate. This is because water level measurements can be directly obtained from ground-based instruments, radars, or even satellite altimetry data. Although some measurement instruments are considered costly, low-cost alternatives are also widely available and utilized. However, obtaining continuous discharge records cannot be regarded as direct or straightforward, specifically for tidal rivers and estuaries, where river flow is not characterized as homogeneous or steady [91]. No ground-based instruments can be used to directly obtain discharge records, as they are estimated indirectly using various parameters, such as flow velocity and the cross-sectional area of the river. An example of such an instrument is the Horizontal Acoustic Doppler Current Profiler (H-ADCP). It measures the velocity field along horizontal acoustic beams, and discharge estimates are then obtained through integration of measured velocities over the channel's cross-sectional area. The downsides of these instruments are their cost, frequent maintenance, and inability to handle missing data or to reconstruct data outside of their estimation period when they were first installed.

An alternative solution, which is simple, cost-effective, and used in inland rivers for continuous discharge monitoring, represents the Stage-Discharge rating curve. It defines a functional dependence between river level and discharge at a given channel cross-section based on individual flow measurements. However, its utilization is inadequate in tidal rivers and estuaries due to the nonlinear relationship between tidal influence, discharge, and water level. [63] There is no one-to-one relationship or direct functional dependence between water level and discharge data in tidal rivers and estuaries, implying that a single water level may be associated with a range of discharge values. Therefore, a viable approach commonly employed to model complex river processes is the use of numerical models. The rationale for using these models resides in their ability to explain the underlying physical mechanisms. Although numerical approaches, such as physically based hydraulic models and conceptual hydrologic models, are commonly employed, they require discharge data as an input parameter. However, due to problems such as insufficient or lacking data, which are even more pronounced in data-scarce areas, data dependency becomes one of the crucial limitations of numerical models.

Due to the restrictions of previous approaches, machine learning was introduced as a powerful alternative for resolving the problem of discharge estimation and prediction. Machine learning has been extensively applied to numerous hydrological purposes, ranging from generating simulations [143] to modeling water quality [122], managing water resources [110], and many more. Furthermore, unlike simple statistical models, machine learning models effectively handle nonlinear and non-stationary data, which are prevalent when studying areas such as tidal rivers and estuaries. Additionally, machine learning models can be combined with other methodologies, such as signal processing techniques [76], optimization algorithms [23], or previously mentioned hydraulic and hydrologic models [28, 140].

Available review papers are mainly focused on inland rivers, however, only recently, a paper that provides a critical perspective on applying machine learning for hydrological analysis on water level and discharge in tidal rivers and estuaries was published by Mihel et al. [90]. Based on the findings of the review paper, it was concluded that enabling effective water level management in tidal rivers and estuaries has become a major challenge in recent years, mainly due to the influence of climate change and complex interactions between different factors, as well as anthropogenic activities. To ensure protection

from water-related risks and promote sustainability, various approaches are employed and considered essential, including monitoring, forecasting, and modeling. The review study highlighted two machine learning approaches as the basis for further research: advanced recurrent neural networks and hybrid models, which have demonstrated improved performance compared to stand-alone machine learning approaches. Advanced recurrent neural networks have been previously applied to problems related to water level and discharge in tidal rivers and estuaries, demonstrating considerable potential. This thesis investigates their use as a stand-alone approach and as part of a novel hybrid architecture, which has not yet been applied in the tidal reaches. Motivated by the recent popularity of attention mechanisms for sequence modeling and the limited number of studies that explore their applicability for tidal environments, but not for the problems at hand, this thesis introduces a novel approach that combines attention mechanisms with an advanced recurrent neural network. Signal preprocessing was included in another stage of the research to enhance estimation accuracy by decomposing the available water level observations into distinct frequency bands, by employing a method not previously applied in this context. The purpose of the decomposition was to capture meaningful tidal patterns and achieve higher estimation accuracy, compared to using time-series data.

1.2. Research Aims, Hypothesis, and Scientific Contributions

The thesis is focused on successfully achieving two research aims: (1) developing a robust hybrid machine learning approach capable of accurately estimating and predicting river discharges and water levels across diverse flow conditions, and (2) investigate physical processes in tidal rivers and estuaries, where traditional estimation methods often fall short due to complex interactions among tidal, river, and storm surge influences. In accordance with the declared research objectives, the hypotheses are as follows:

I. *The estimation and prediction of discharges and water levels can be achieved with high accuracy by using only water level data from multiple stations.*

II. *In complex environments, such as microtidal rivers and estuaries, discharge and water level reconstruction are more accurate when using machine learning models*

for time-series forecasting, compared to non-temporal machine learning models.

III. *Hybrid machine learning approach that integrates an attention mechanism outperforms other machine learning models for time-series forecasting in estimation and prediction of discharges and water levels.*

IV. *Decomposing water level signals into distinct frequency bands enhances the estimation accuracy for river discharges and water levels.*

The scientific contributions of this research are as follows:

– A machine learning-based approach for estimation and prediction of water levels and discharges in tidal rivers and estuaries;

– A hybrid machine learning model that integrates an attention mechanism to better identify critical features and improve estimation and prediction accuracy across flow conditions;

– Evaluation of the impact of tidal amplitude and storm surge on water level and river discharge estimation and prediction.

1.3. Research Methodology

The focus of the thesis is on two types of analysis: estimation and prediction of discharges and water levels. These analyses differ in that estimation focuses on reconstruction problems where missing data is present, thereby extending the current sequence. In contrast, prediction aims to forecast the target variable several hours, days, months, or more in advance. The research direction was based on a review paper [90], which provided a comprehensive overview and focused on the areas of tidal rivers and estuaries. As hybrid modeling was one of the most promising directions, this research investigated the potential of such an approach for the two types of analysis, for the Neretva River microtidal salt-wedge estuary.

The integration of the attention mechanism with an advanced, Recurrent Neural Network (RNN), as the proposed novel hybrid approach, was evaluated for the estimation analysis. The proposed model was compared with various machine learning approaches,

commonly employed for such or similar problems involving time-series analysis, across all flow conditions and evaluated using two datasets: one comprising measured records and the other simulated records. Measured data was provided by two governmental institutions, Croatian Waters and the Croatian Meteorological and Hydrological Service, while the simulated data was generated using a numerical model, STREAM 1D. The advantages and limitations of each approach were discussed, aiming to provide with such dual approach, insights into the models robustness and effectiveness.

Data preprocessing represents the first stage of the methodology, which involves identifying potential lags between the water levels at different stations and river discharge, correcting missing data, reducing the influence of high-frequency noise, and analyzing feature significance. The water level records provided by the Croatian Meteorological and Hydrological Service had previously undergone official verification to ensure data quality. However, discharge records contained some discrepancies, such as missing values, which were addressed through interpolation. In addition to missing data, the measured dataset contained high-frequency noise. Cross-correlation was employed to identify potential lags, after which the time-series data were corrected accordingly. After the data had been cleaned, feature significance analysis was conducted using correlation analysis and mutual information scores. The interactions between water level stations concerning the discharge parameter and their importance were analyzed. Correlation analysis is limited to finding only linear associations between the features using Pearson's correlation coefficient [27]; therefore, to surpass such limitation, mutual information was utilized alongside correlation. With mutual information, it is possible to identify the complete dependence between input and output features, both linear and non-linear [99]. Although such an analysis is useful, its results must be discussed from the hydraulic perspective, as contradictions may occur. Hence, the selection of a feature was dependent on correlation, mutual information analysis, and domain-specific knowledge. The final step in data preprocessing was determining the sliding window length for time-dependent models, ensuring that the window encompasses both daily and intra-day oscillations.

The second stage of the methodology entails the selection of machine learning models informed by the conducted literature review [90], whose performance is evaluated. More precisely, two categories were considered. The first category referred to simple non-temporal supervised ML models, such as Decision Trees (DT), Random Forest (RF),

Support Vector Regression (SVR) with radial basis function and sigmoid kernel, Light Gradient Boosting Machine (LGBM), and Extreme Gradient Boosting Machine (XGB). Although some of the mentioned models were applied for reconstructing discharge [117], models such as LGBM, which has been successfully used for estimating water levels in tidal rivers [90], and another popular gradient boosting model XGB [37], previously not employed for tidal reaches and water level and discharge estimation, were investigated. The rationale behind employing simple non-temporal machine learning models was their advantages when solving regression problems, such as short training time, low resource requirements, effectiveness in handling non-linear data, and transparency [47, 132]. The second category of machine learning models, classified as complex and time-dependent, represents time-series models, Long Short-Term Memory (LSTM), and LSTM combined with attention mechanism (LSTM-Attention). LSTM-Attention represents the proposed modeling approach that combines the benefits of the LSTM model with those of the attention mechanism by assigning different weights based on the relevance of input features [138]. The goal of the proposed hybrid methodology was to improve resilience to irrelevant and noisy data.

Additionally, these models capture temporal dependencies, a feature absent in simple non-temporal machine learning models. Developing simple non-temporal machine learning models was accomplished using the scikit-learn toolkit, while time-dependent models were built with PyTorch. Providing optimal performance for each model was crucial; therefore, a five-fold grid search cross-validation was performed. From the aspect of computational resources, a workstation available at the Faculty of Engineering equipped with an NVIDIA RTX A6000 with a GPU of 48GB was used to train the proposed machine learning models.

Evaluation of the models is the third phase of the methodology. Four metrics were employed for this purpose: root mean squared error, mean absolute error, Nash-Sutcliffe Efficiency coefficient, and Pearson's correlation coefficient. The selected metrics are commonly employed in water level and discharge estimation studies [90]. However, a detailed assessment of model performance was conducted from the perspective of different discharge categories: low, medium, high, and extremely high. The models adaptability was examined in relation to the distribution of data across different discharge categories. For such evaluation, a single metric was employed. Furthermore, a Taylor diagram was

used to examine the models' total error reduction and their ability to capture the inherent variation within the dataset. The step preceding the final evaluation phase was conducting a statistical significance analysis using a Wilcoxon signed-rank test, which employs the model's residuals to determine if the results of the models differ significantly from one another [26]. The final segment of the model evaluation focuses on enabling the model interpretability by employing Shapley Additive Explanations (SHAP) analysis [82] for simple non-temporal ML models and a feature occlusion test for the time-dependent models. These methods enabled the determination of feature contribution to the overall model's predictions. Feature occlusion was employed due to the limitations of SHAP analysis for the RNN-based architecture. Ascertaining feature redundancy, or necessity for providing accurate discharge prediction, was conducted by integrating additional features to the model, whose performance was then assessed based on the selected evaluation metrics.

Another hybrid approach was applied for discharge and water level forecasting, combining a convolutional neural network with an LSTM, using simulated data generated by a numerical model HEC-RAS. The model was evaluated on different input combinations, time-series data, and spectrograms. Feature engineering was also performed during the data preprocessing stage, as it offers significant advantages and potential prediction enhancements.

Including the signal processing method, variational mode decomposition, and testing the performance of time-series models, LSTM and LSTM-Attention, using the measured dataset, represents the final stage of the methodology. Decomposing water level signals into separate frequency bands was expected to improve the overall model performance. The decomposed mode functions were assigned to tidal constituent categories and, as such, were given to the models as input.

1.4. Thesis Overview

The thesis is structured in the following manner. In **Chapter 2**, an introduction to machine learning, and the theoretical basis of the applied models, both simple and time-dependent, is given.

A brief overview of the research area and data sources is given in **Chapter 3**. Specif-

ically, the chapter focuses on data acquisition, as measured data was been gathered from various instruments, and numerical models that generated simulated data.

Chapter 4 provides an overview of the literature focused on the hydrological analysis of water level and discharge, explaining the reason for employing machine learning for the discharge estimation problem in tidal rivers and estuaries, and classifies the machine learning models currently used by type of analysis and category of ML models. The chapter concludes by discussing the results of previous studies from a critical perspective.

In continuation of previous chapters, **Chapter 5** presents two hybrid approaches, and proposes an Advanced Recurrent Neural Network called Long Short-Term Memory with an Attention mechanism as the novel hybrid approach. This chapter also describes and outlines the techniques and methods used for data analysis and model development. Significance tests are discussed to ensure the validity of our results, along with techniques for providing model interpretability.

The results and discussion are presented in **Chapter 6**, where selected models are tested on acquired datasets and under different flow conditions, with an emphasis on the models' adaptability and generalization.

The last part of the dissertation, **Chapter 7**, contains research summary, conclusions, and recommendations for future research.

Chapter 2

AN INSIGHT INTO MACHINE LEARNING MODELS

Contents

This chapter provides an introduction to machine learning, and an overview of applied machine learning models divided into two categories: simple non-temporal and time-series models. The first category includes simple non-temporal supervised machine learning models: decision tree, random forest, support vector regression, light gradient boosting machine, and extreme gradient boosting machine. The second category presents

a recurrent neural network and its advanced variant, the long short-term memory.

2.1.   Machine Learning

Machine learning, a branch of artificial intelligence, encompasses computational methods that learn from experience [92]. This includes identifying patterns, trends, and relationships, so that decisions or predictions can be made. This branch focuses on developing learning algorithms and statistical methods that aim to be efficient and accurate, depending on the type of learning problem.

Frequent classes of learning problems refer to [92]: (1) Regression, whose aim is to predict a continuous output variable based on one or more input variables, by learning a function that maps input to output. Regression predictions are evaluated by employing different evaluation metrics, which calculate the magnitude of the difference between observed and predicted values [92]. Examples of such metrics are mean squared error and mean absolute error, and (2) Classification, whose aim is to predict a categorical output variable by assigning a category to each item [92]. These categories can be binary, when only two classes are present, or multiclass, when more than two classes are included in the dataset. Unlike regression, classification predictions are not evaluated based on the numerical closeness of values, but rather on their correct assignment to predefined classes using metrics such as accuracy and precision.

Different learning approaches can be distinguished depending on the type of data, namely supervised, unsupervised, and semi-supervised learning. In supervised learning, input-output pairs are known prior to training, indicating that the training dataset is labeled and the ground truth is available. The goal is to learn a function that maps inputs to outputs, which is then used to make predictions on unseen data. Typical learning problems in supervised learning include regression and classification. Unlike supervised learning, unsupervised learning handles unlabeled datasets, where the goal is to discover hidden patterns and structures in the data without access to the ground truth. Quantitatively evaluating the performance of such approaches can be challenging [92]. Clustering and dimensionality reduction are examples of learning problems addressed through unsupervised learning. Semi-supervised learning involves training on datasets that contain both labeled and unlabeled data, while also making predictions on unseen

data [92]. This approach is common in settings where data labeling is costly, despite the datasets being easily accessible. [92] An example of a strategy employed in semi-supervised learning is data augmentation, which aims to expand the labeled dataset.

The conducted research was based on solving a regression problem, regarding the estimation and prediction of discharge and water level, using a supervised learning approach, where input-output pairs were known in advance. Based on the literature review [90], the models described in the following sections were selected to serve as references to evaluate the proposed hybrid model for the estimation problem.

2.2. Simple Non-Temporal Machine Learning Models

2.2.1. Decision Tree

The Decision Tree (DT) represents the base structure of ensemble machine learning models. DT is defined as non-parametric algorithm, simple to understand, interpretable, and built in a supervised manner [52]. The regression approach for DT is based on predicting the continuous numerical value, unlike the classification approach, whose goal is to provide a class label [106]. Despite this difference in the output type, both regression and classification tasks rely on the same mechanism, where the output is determined at the leaf node. The main goal of DT is to reduce the output variable variance, and this is done by starting from the root node, which is then divided into the subsets of the data by employing a criterion for selecting the best split, such as mean squared error (MSE) or mean absolute error (MAE). Branches are generated with each split, and tree development ceases when a stopping criterion is met, which may pertain to maximum tree depth or other decision tree features. The present study utilized a classification and regression trees model for regression analysis. The selected criterion for splitting a node in the tree refers to the MSE, as it penalizes larger errors more heavily, and it is represented by the following mathematical expression:

$$\Delta MSE = MSE(S) - \left(\frac{|S_l|}{|S|} MSE(S_l) + \frac{|S_r|}{|S|} MSE(S_r) \right), \qquad (2.1)$$

where $|S_l|$ represents the number of samples in the left subset, $|S_r|$ number of samples in the right subset, $|S|$ total number of samples in the initial set, $MSE(S)$ initial error, $MSE(S_l)$ weighted MSE for the left subset, $MSE(S_r)$ weighted MSE for the right subset, and ΔMSE the splitting criterion in regression trees.

Despite the numerous advantages of the DT approach, it may encounter certain challenges. A primary concern pertains to overfitting. Overfitting can be partially mitigated using the pruning strategy, focusing on optimizing pruning parameters and performing validation on unseen data. This process removes certain leaf nodes, producing a shallower regression tree if the total prediction error remains within an acceptable range. In addition to reducing overfitting, it also reduces the time required to predict new data points [42].

2.2.2. Random Forest

Random Forest (RF) is an ensemble approach constructed of multiple DTs and was introduced in 2001 by Breiman [14]. Like DT, its application is possible for both regression and classification analysis. The more complex architecture of the RF model, in comparison to the previous DT, makes it more applicable for problems of a more complex nature while aiming to provide satisfactory accuracy. The development of the RF model is carried out through three general steps, shown in Figure 2.1. Data preparation represents the first step, with preprocessing and bootstrap sampling. With the second step, the construction of an RF model begins by generating multiple DTs with a random feature selection performed for every division. Lastly, the predictions of every DT are aggerated to produce the final prediction results, i.e., the averaged value of all constructed DTs [60], whose term can be expressed as [101]:

$$f^{RF}(x) = \frac{1}{N_{tree}} \sum_{n=1}^{N_{tree}} \hat{h}_n(x), \tag{2.2}$$

x is the input feature vector, N_{tree} represents the overall number of DT in a forest, $\hat{h}_n$ is the prediction of nth regression DT, and f is the overall prediction function of the RF obtained by averaging individual DT predictions.

RF differs from DT based on the techniques utilized to construct the mode, which are:

bagging (bootstrap aggregation), feature randomness, and binary recursive partitioning. Bagging is a technique that plays a key role when it comes to overfitting. The technique works on the principle of generating different subsets from the original dataset by random selection. This results in certain data points being included several times while, for example, others may not be included even once. The purpose of applying this approach is to reduce variance and achieve greater stability [85]. Every DT in RF is constructed using a different subset of data, i.e., a bootstrap sample. In this way, noise has a reduced impact on the prediction accuracy. Another advantage of RF is feature randomness, where a random set of features is used for every node split. Hence, the possibility of a high correlation between different trees is reduced. This is considered one of the more impactful limitations of DT compared to RF. Feature randomness technique thereby aims to improve the model's generalization ability on unseen data and its robustness.

The RF model is limited when it comes to extrapolating values that are outside the range of the training dataset [46]. In such situations, the model could either result in overestimations or underestimations. This stems from its dependence on historical data. Overfitting can occur when the dataset has a limited number of observations combined with overly deep individual trees. Another limitation is that the model's accuracy depends significantly on the distribution of the dataset. At the beginning of the RF model-building process, three parameters must be defined: the number of regression DTs, the independent variables that are randomly selected at the nodes, and the minimum number of observations that are required at the terminal node of each tree [74].
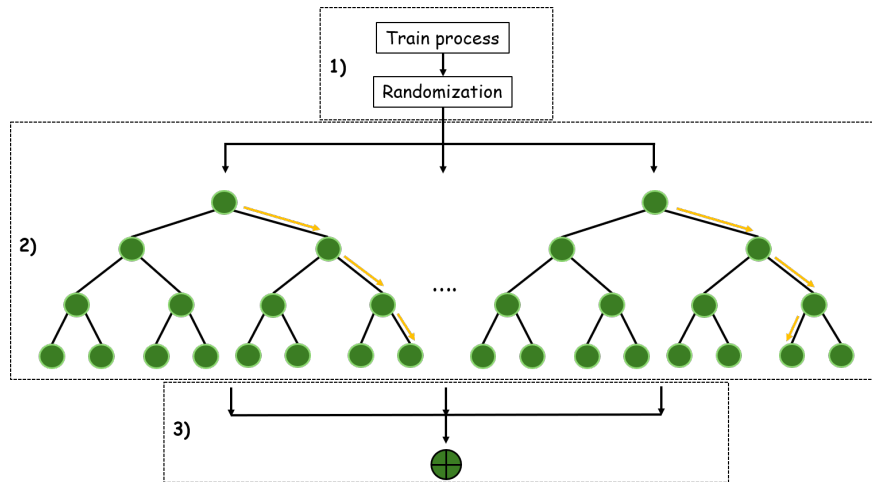
Figure 2.1: RF workflow: 1) Generate vectors with random elements, 2) Construct DTs, 3) Average all DTs' predictions

2.2.3. Support Vector Regression

Support Vector Regression (SVR) is a kernel-based machine learning approach specifically designed for regression problems. This represents an adaptation of the foundational SVM methodology, primarily employed for classification purposes, as presented by Vapnik in 1999 [123].

The operational scheme of the SVR approach is presented in Figure 2.2. From a more general perspective, SVR architecture has three main layers: input, hidden, and output. The input features construct the input layer, which is then given to the hidden layer where the kernel function is applied to process the given features. Finally, the kernel function outputs are passed to the output layer to provide the final predictions. The SVR approach is capable of addressing both linear relationships through the linear kernel and intricate nonlinear relationships using nonlinear kernel functions.
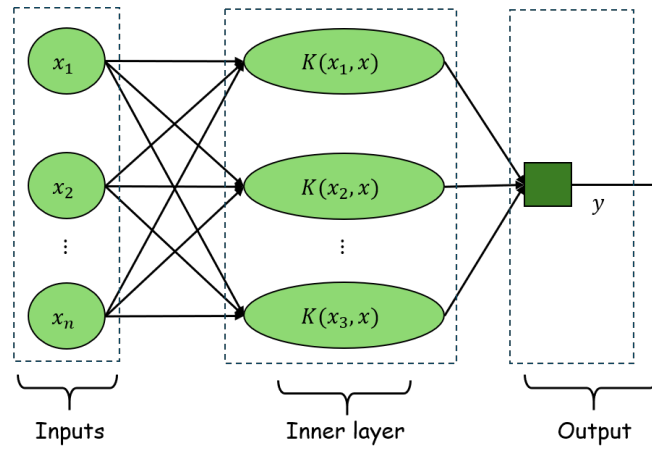
Figure 2.2: Operational scheme of the SVR model

Unlike the SVM, whose primary objective is to find the optimal hyperplane, SVR aims to determine the a regression function that provides the best data approximation within a predefined error tolerance (ε). The formulation of SVR, which considers the predefined error tolerance, can be defined as [48]:

$$\text{minimize} \quad \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{N}(\xi_i + \xi_i^*) \tag{2.3}$$

$$\text{subject to:} \quad \begin{cases} y_i - \left(\mathbf{w}^\top \phi(\mathbf{x}_i) + b \right) \leq \epsilon + \xi_i, \\ \mathbf{w}^\top \phi(\mathbf{x}_i) + b - y_i \leq \epsilon + \xi_i^*, \\ \xi_i, \xi_i^* \geq 0, \quad i = 1, \dots, N, \end{cases} \quad (2.4)$$

where $\mathbf{w}$ is the regression weight vector, C is the penalty parameter, ξ_i, ξ_i^* are nonnegative slack variables, $\phi(\cdot)$ is a nonlinear feature mapping function, b is the bias term, and ϵ is the insensitive loss parameter. The input $\mathbf{x}_i$ is a feature vector, the corresponding target value is y_i, and N represents the number of samples.

The regression function of an SVR model is defined as:

$$f(\mathbf{x}_i) = \mathbf{w}^\top \phi(\mathbf{x}_i) + b,$$

where $\mathbf{w}^\top \phi(\mathbf{x}_i)$ denotes the inner product between two vectors, b is the added bias term, and $f(\mathbf{x}_i)$ is the model prediction for the input vector $\mathbf{x}_i$.

Given that SVR utilizes a kernel-based methodology, the selection of kernels is influenced by the distribution of the data and the specific problem being addressed. Four primary types of kernels exist: linear, polynomial, RBF, and sigmoid. All of the kernels discussed, with the exception of the linear one, are capable of managing nonlinear data. However, given the complex nature involved, only two kernels were chosen: RBF and sigmoid. The use of an RBF kernel function in SVR facilitates the learning of intricate nonlinear relationships by transforming the input features into an infinite dimensional space. In contrast, the sigmoid kernel bears a closer resemblance to the activation function of neural networks, specifically the hyperbolic tangent function (*tanh*). The mathematical expression for the RBF and the sigmoid kernel can be written in the following manner [101, 117]:

$$K\left(x_i, x_j\right) = \exp\left(-\gamma \cdot \left\| x_i - x_j \right\|^2 \right), \gamma > 0, \quad (2.5)$$

$$K\left(x_i, x_j\right) = \tanh\left(\gamma \cdot x_i^T \cdot x_j + r \right), \quad (2.6)$$

where x_i and x_j are input vectors, γ is the scaling parameter, r is a bias term, tanh denotes the hyperbolic tangent function and exp the exponential function, $\left\| x_i - x_j \right\|^2$ is the squared Euclidean distance, and $x_i^T \cdot x_j$ is the inner product of two vectors.

2.2.4. Light Gradient Boosting Machine

Light Gradient Boosting Machine (LGBM) represents a notable advancement in ensemble machine learning techniques, having been introduced in 2017 by Ke et al. [65] to address the earlier challenges associated with high-dimensional features and large datasets. The LGBM approach distinguishes itself from other ensemble methods by including two additional techniques. The first technique, known as Gradient-based One-Side Sampling, filters out instances with small gradients, keeping only those that add to information gain. The second, Exclusive Feature Bundling, optimizes LGBM by clustering mutually exclusive or less frequent features, reducing their quantity. These methods boost LGBM performance while retaining satisfactory prediction accuracy and computational efficiency and also reduce the possibility of overfitting.

LGBM is based on leaf-wise, also known as best-first, tree growth strategy, shown in Figure 2.3. Opposite to level-wise tree growth, this strategy does not equally expand the tree structure at the same depth, i.e., uniformly, but is more focused on finding the leaf node that provides the highest error reduction and information gain, although this results in a deeper tree structure. It is often described as a greedy approach compared to a level-wise strategy. A leaf-wise strategy is effective in cases of large datasets and results in faster training [37] than a level-wise strategy. Besides the mentioned advantages, LGBM facilitates parallel training, resulting in the development of trees that can be considered efficient [120]. However, due to the existing tree structure imbalance, overfitting can pose a problem; therefore, appropriate regularization is essential.
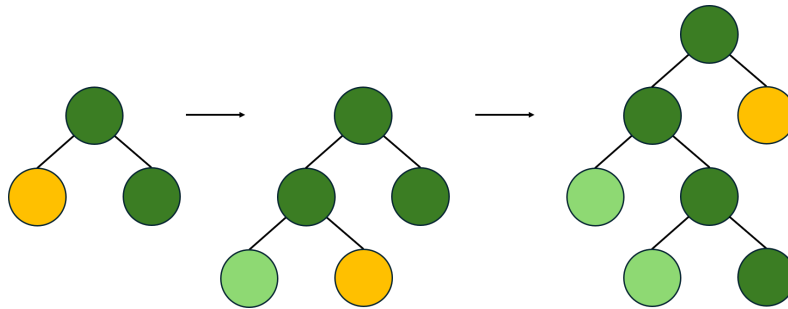
Figure 2.3: Step-by-step LGBM leaf-wise tree growth strategy

LGBM's objective function for improving the model's performance consists of two parts: the second-order Taylor approximation term and the regularization term. The first term aims to effectively update the model, while the second term controls the tree

complexity. The main aim of employing the mentioned loss functions is to enable fast learning. The expression for the LGBM objective function can be written as [43]:

$$J^{(t)} = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k), \tag{2.7}$$

where l is the loss function, y_i is the observed value, $\hat{y}_i$ is the estimated value, n is the number of samples, Ω represents regulation term, f_k the prediction function of k-th DT, K number of trees, t time step, and J overall objective function.

2.2.5. Extreme Gradient Boosting Machine

Extreme Gradient Boosting, often known as XGB, is yet another method within the ensemble gradient boosting framework. The present version of the XGB model represents an enhancement of the 2001 model introduced by Friedman [33]. In contrast to the traditional RF approach, which does not incorporate integrated regularization and tree pruning, both LGBM and XGB offer distinct benefits. The improvements lead to a notable decrease in model complexity and overfitting [98], as well as a reduction in training time. Furthermore, the XGB model encompasses CPU multithreading parallelization and block technology with the objective of enhancing performance [98].

Although both LGBM and XGB are machine learning approaches based on gradient boosting techniques, they differentiate in tree growth strategy. XGB employs a level-wise tree growth strategy as illustrated in Figure 2.4. This strategy creates a more balanced tree structure, where nodes are produced uniformly across the same level. The aforementioned strategy generates a model characterized by simpler and shallower tree architecture. As in LGBM, the goal is to minimize the loss function, which is conducted through node splitting at the entire tree level while finding the optimal split for every tree node. This means that all nodes are generated for every level, ensuring that each level of the tree is filled before proceeding to the next step. This is considered beneficial in situations where smaller datasets are provided, and a study by Gan et al. [37] has demonstrated how more consistent results can be achieved compared to the LGBM approach.

The objective function outlined in Formula 2.7 is applicable to the XGB approach as well, with the primary difference in the computation of gradients and their subsequent aggregation. LGBM employs a histogram-based approach to approximate the gradient
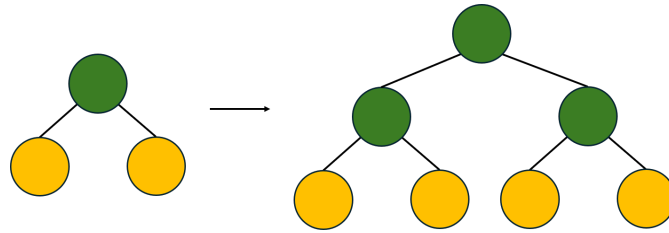
Figure 2.4: Step-by-step XGB level-wise tree growth strategy

(or first-order derivative) and Hessian (or second-order derivative), resulting in distinct bins for continuous data. In contrast, XGB calculates these values exactly for each data point, followed by aggregation at the leaf level. This allows for a meticulous optimization process to be carried out by the model.

2.3. Time-series Machine Learning Models

2.3.1. Recurrent Neural Networks

Recurrent Neural Networks (RNNs) are feedforward neural networks with one or more feedback loops, which enable them to process sequential data. RNNs were first developed in 1992 by Hopfield [59]. There are two key differences between RNNs and feedforward neural networks: (1) the sequence dynamic is captured through cycles in networks of nodes, and (2) retaining the state, which represents information from a long context window [77]. Today, they are employed for a wide variety of tasks, including regression tasks such as time-series forecasting, and classification tasks such as speech recognition and natural language processing.

RNNs input are sequences, denoted as $(x^{(1)}, x^{(2)}, , x^{(T)})$, where each $x^{(t)}$ is a real-valued vector. The output of an RNN is also a sequence, denoted as $(y^{(1)}, y^{(2)}, , y^{(T)})$. Additionally, apart from the sequence of input-output pairs, either the input or the output may represent only a single data point. If the utilized sequence is finite, the maximum time index is referred to as T. [77] RNNs introduce recurrent edges, which enable them to connect adjacent time steps. These recurrent edges form different cycles depending on the length, such as self-connections from a node to itself across time. There are two main equations which are calculated at every time step during the forward pass in a simple

RNN [77]:

$$h^{(t)} = \sigma\big(W^{hx}x^{(t)} + W^{hh}h^{(t-1)} + b_h\big), \tag{2.8}$$

$$\hat{y}^{(t)} = W^y h^{(t)} + b_y, \tag{2.9}$$

where t denotes time step, $x^{(t)}$ the data point at time t, $x^{(t-1)}$ the previous data point at time $t-1$, $h^{(t-1)}$ hidden node values in the networks previous state, $h^{(t)}$ hidden node values at time t, $\hat{y}^{(t)}$ the models prediction at time t, $W^{(hx)}$ matrix of conventional weights between the input and the hidden layer, $W^{(hh)}$ matrix of recurrent weights between the hidden layer and itself at adjacent time steps, $W^{(yh)}$ the weight matrix that maps the hidden state $h^{(t)}$ to the output space, $b^{(h)}$ and $b^{(y)}$ are the corresponding bias parameters, and σ is the sigmoid activation function.

A problem linked to the use of RNNs refers to the vanishing gradient problem. During backpropagation through time, the gradient becomes exponentially small, leading to difficulties in retaining long-term dependencies and the inability to learn correlations between distant time steps. The reasons behind this problem, as defined in the literature [103], are linked to standard nonlinear functions with a gradient being close to zero in most areas of the function, and because during backpropagation, gradients converge to zero rapidly, if the eigenvalues of the recurrent matrix become less than one. [103] Hence, the vanishing gradient problem motivated the development of the Long Short-Term Memory network, which is discussed in the following subsection.

2.3.2. Long Short-Term Memory

The RNN approach is commonly utilized for sequential data. However, it is important to address the fact that its ability to preserve the data over longer periods is limited, which is especially evident in the training phase as the vanishing gradient problem arises. To address this limitation, a novel approach known as Long Short-Term Memory (LSTM) was introduced by Hochreiter and Schmidhuber [56] in 1997. The presented model effectively mitigated the issue of vanishing gradient and demonstrated satisfactory accuracy for problems related to sequential data processing, confirmed by numerous studies across different areas of study [108].

Every LSTM network comprises multiple LSTM cells that serve as its basic building

block (see Figure 2.5). These cells are a part of the hidden layer within the LSTM network, and multiple cells can also be connected to form an LSTM layer. LSTM cell features two memory units: internal cell state (c_t) representing the long-term component, the candidate state ($\tilde{c}_t$), or new possible short-term memory component. In addition to memory units, three gating mechanisms are present: input (i_t), forget (f_t), and output (o_t). Gating mechanisms enable the LSTM network to manage the flow of information selectively, meaning this part of the network determines which information is stored, updated, or passed forward. A simpler explanation would define these gating mechanisms as switches, which are either turned on or turned off, depending on the input data [136]. This is possible by employing two types of activation functions: the sigmoid and hyperbolic tangent. Besides, the mentioned LSTM cell also contains a hidden state (h_t), which is defined as the output of each LSTM cell for a defined time step.
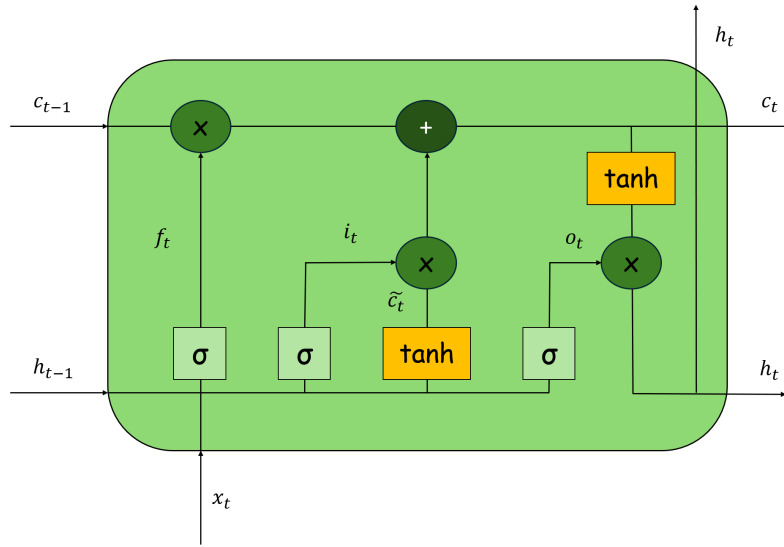
Figure 2.5: Internal structure of an LSTM cell

The flow of the information between gates occurs in the following manner. Using h_{t-1}, the previous time step hidden state, and x_t or the input of the current time step, the calculations for the three mentioned gating mechanisms and the candidate state $\tilde{c}_t$ are performed. Forget gate (expression shown in Eq. 2.10) applies a sigmoid function, which produces a value in the range between 0 and 1, that determines the quantity of the previous cell state (c_{t-1}) retention in the current, i.e., updated cell state (c_t). If the resulting value is closer to 1, the information is kept, while if the resulting value is closer

to 0, it is discarded.

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{2.10}$$

After this step, input gate calculations are performed, which (unlike the forget gate) determine whether new information will be added to the internal cell state (c_t) by employing a sigmoid function. If the result of the input gate is around 0, the information is ignored; however, if the resulting value is around 1, the new information is fully added to the current internal cell state (c_t). This enables the LSTM network to improve its ability to learn and retain important past information over time.

$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] + b_i\right) \tag{2.11}$$

Based on the output of the input gating mechanism, the candidate cell state $(\tilde{c}_t)$ is either added or ignored during the update of the internal cell state by using a hyperbolic tangent function, which converts the data into a range between -1 and 1. Therefore, the candidate cell state $(\tilde{c}_t)$ represents the potential new information to be incorporated into the cell state (c_t).

$$\tilde{c}_t = \tanh\left(W_n \cdot [h_{t-1}, x_t] + b_n\right) \tag{2.12}$$

The next phase represents the update of the cell memory, where the forget gate (f_t) determines how much of the previous cell state(c_{t-1}) is retained, while the input gate (i_t) task is to determine how much of the new information, i.e., candidate cell state $(Tildec_t)$ is included. The result represents an updated internal cell state (c_t) containing only vital information.

$$c_t = f_t \cdot c_{t-1} + i_t \cdot \tilde{c}_t \tag{2.13}$$

The last step involves the output gate (o_t) and generates a hidden state (h_t) that is either passed to the next time step or used for prediction. The primary function of the output gate (o_t) is to determine how much of the internal cell state (c_t) information will be used to generate a hidden state (h_t). If the value of the output gate (o_t) is close to 0,

only some small part of the information from the internal cell state (c_t) will be used to generate the hidden state (h_t). In contrast, if the output gate (o_t) is close to 1, most of the internal cell state (c_t) information will be used to generate a hidden state (h_t).

$$o_t = \sigma\left(W_o \cdot \left[h_{t-1}, x_t\right] + b_o\right) \tag{2.14}$$

$$h_t = o_t \cdot \tanh(c_t) \tag{2.15}$$

In Equations 2.10 - 2.15, the weights associated with the gating mechanisms and the candidate cell are denoted as W_f, W_i, W_o, and W_n. The corresponding biases are represented by b_f, b_i, b_o, and b_n, while the activation functions are indicated by σ for the sigmoid function and tanh for the hyperbolic tangent function [88].

After introducing basic machine learning models, more complex hybrid models, including the proposed novel one, are presented in Chapter 5. The following chapter provides an overview of the study area, datasets, and specifics regarding their collection and simulation.

Chapter 3

RESEARCH DOMAIN AND DATA SOURCES

Contents

This chapter provides an overview of the research area and the utilized datasets. The first part discusses the research area broadly, considering the full length of the Neretva River, followed by a more focused examination of the tidal section of the Neretva River, which serves as the primary area of interest. The second part comprises three datasets, whose primary purpose is to provide a more detailed performance assessment of the selected machine learning models. The field measurements are obtained from two governmental institutions, whereas the remaining two datasets are generated by employing different numerical modeling approaches. The aim of examining these datasets is to enhance understanding of the dynamics of river flow and its hydraulic properties.

3.1. Study Area

3.1.1. Neretva River: A General Perspective

The Neretva River, with a total length of 215 km and a catchment area of approximately 10,500 km^2, flows through two countries: Bosnia and Herzegovina (193 km of its course, with a catchment area of about 10,220 km^2) and Croatia (the last 22 km of its course, with a catchment area of about 280 km^2). The river originates in the Dinaric karst, more precisely near the Zelengora region, and drains into the Adriatic Sea in southern Dalmatia, near the town of Ploče.

Due to its unique geographical features and temperature properties, it is distinguishable from other rivers in Croatia. It is referred to as the eastern Adriatic coast's largest contemporary delta system [10]. Likewise, the river has the longest course within the eastern Adriatic basin. Climate changes related to air temperature and precipitation are usually negatively correlated. While, for example, air temperature follows an increasing trend of average seasonal and annual values, precipitation shows a slight decreasing trend by the same values. Air temperature varies depending on the season, with pronounced maritime characteristics. For example, summers are moderately warm with temperatures varying around 23.8 °C, and winters are moderately mild with temperatures around 7.4 °C. The smallest seasonal variations are observed between spring and autumn, with autumn temperatures being approximately 1.8 °C higher than those measured in spring. Regarding precipitation, the highest amount of annual precipitation, up to 70%, is characteristic of the colder part of the year, from October to March. Whereas for the period between April and September, the amount of precipitation is significantly reduced, precisely because of the dry periods and high air pressure. [35]

Hydropower plants are located in the upstream part of the river, which is often referred to as an inland region. According to the distance from the river mouth, there are eight hydropower plants: Čapljina (operational since 1979, reversible flow), Mostarsko blato (operational since 2010), Mostar (operational since 1985), Peć-Mlini (operational since 2004), Salakovac (operational since 1981), Grabovica (operational since 1981), Jablanica (operational since 1954), and Rama (operational since 1969) [79]. Hydropower plants assist in regulating the water regime during high waters, however their negative impact

cannot be ignored. Therefore, concerns such as increased erosion, reduced water quality, and many more are common.

However, the focus of this research is on the lower part of the Neretva River, specifically the downstream section, which can be described as an estuarine and tidal river section, i.e., the section of the river where the influence of tides is noticeable. The following subsection provides a more detailed discussion of the Neretva River's tidal reach.

3.1.2.   The Neretva River Estuary and Tidal River Section

The final 22 km of the Neretva River's tidal reach is located in Croatia (see Figure 3.1). This part of the Neretva River is classified as a microtidal salt-wedge estuary. The main characteristic of the microtidal salt-wedge estuary is a two-layer vertical structure containing both freshwater and saltwater, which are separated by a pycnocline, a region of strong vertical density gradient [29]. The thickness of the pycnocline in the Neretva River estuary is less than 50 cm, which is caused by the Adriatic Sea's exceptionally low tidal dynamics and strong freshwater inflow [66].

Precisely because of the proximity of the sea and the influence of hydropower plants in the upper reaches, as well as man-made structures, the water regime is altered, encountering problems such as a decrease in water quality and soil fertility/salinization for certain crops, and more frequent flooding during high waters. Therefore, conducting research in the specified area is important because it aims to provide better water resource management. This is expected to have a direct impact on agriculture and the protection of the local population from any potential hazards.

Hydrological processes have a pronounced seasonal character, with high flow common from October to April and low flow from May to September, typical for Mediterranean climate. During high flow periods, freshwater reserves are replenished in aquifers, while saltwater is pushed out from the river channel towards the river mouth. However, with high flow periods, the risk and damage of flooding increase due to the amount of rainfall, potential storm surges, and high sea levels. Low flow periods are associated with various problems related to salinization. As climate change is becoming more extreme, with higher temperatures during the summer and less frequent rainfall events, the saltwater can penetrate more than 20 km upstream of the river mouth, more precisely upstream
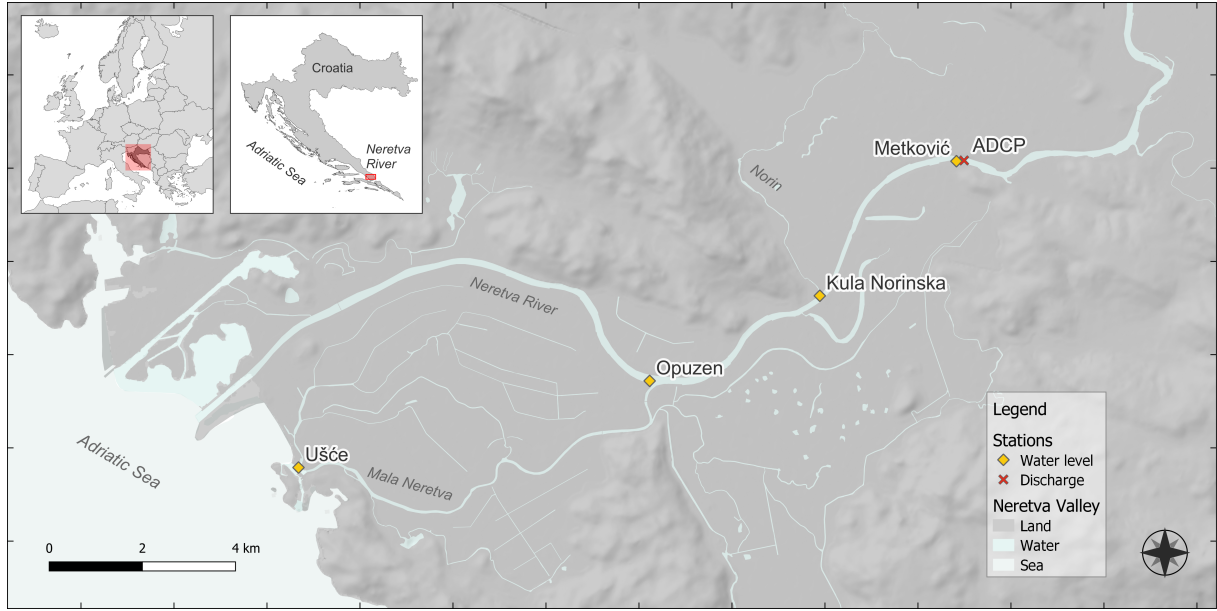
Figure 3.1: Geographical overview of the Neretva tidal river and locations of hydrological measurements of water level and discharge: a) Ušće (0 km from the mouth), b) Opuzen (11 km from the mouth), c) Kula Norinska (16 km from the mouth), d) Metković (22 km from the mouth) [88]

from Metković. Although coastal aquifers are available, irrigation in the river becomes limited due to the increase in salinity, negatively affecting agriculture [81].

3.2. Dataset Overview

Three sets of data from different sources were used for research purposes: (1) measured data obtained from hydrological stations, provided by Croatian Waters and the Croatian Meteorological and Hydrological Service, (2) simulated data generated by the STREAM 1D numerical model [66], and (3) simulated data generated by the HEC-RAS model. All datasets comprise water level records from four hydrological stations and discharge data from a single station. While measured and STREAM 1D simulated datasets span between 2016 and 2021, the HEC-RAS set encapsulates data between 2016 and 2019. The differences in data spans arise from the use of the HEC-RAS dataset in the initial phase of the research, representing a preliminary analysis. The records are presented in an hourly temporal resolution. Data resolution plays an important part in tidal reaches. To enable the effective capture of tidal dynamics, hourly data are mandatory, unlike, for example, inland reaches, where daily observations are often sufficient [55, 128]. A summarized

overview of all utilized datasets is presented in Appendix A in Tables A.1, A.2, and A.3.

3.2.1. Measured data: Acquisition and Correction

Three hydrological parameters are considered for the specified area: sea level, water level, and discharge. Sea levels are measured at the tidal station Ušće (Mala Neretva), situated near the mouth of the Neretva River, but outside the influence of river flow. Water level data is collected from three hydrological stations along the Neretva River, Opuzen (11 km from the river mouth), Norin (16 km from the river mouth), and Metković (20 km from the river mouth). Discharge data is gathered using three H-ADCPs positioned under the bridge in Metković. The stations of interest are shown in Figure 3.1.

Metković and Opuzen represent the two oldest hydrological stations on the Neretva River, which were established to measure water levels. From a historical point of view, the order of the established hydrological stations that became operational is the following: Metković in 1934., Opuzen in 1936., Ustava Ušće in 1976. The last station is Kula Norinska, established in 1986. The only station where sea level measurements have been available is the Ušće station since 1977. All stations, except Kula Norinska, have provided automated measurement reports since 2014, which are close to real-time [69].

Continuous monitoring of discharge, unlike water level, has only been available recently, more precisely, starting from 2015. The measuring equipment was positioned in the cross section under the road bridge, 90 m upstream from Metković hydrological station, where three H-ADCP were placed [66]. These measuring instruments provide a precise estimate of the discharge data. The estimations are obtained by integrating the velocity profile over cross sectional areas in bridge openings [88]. The overall discharge estimation in Metković is obtained by summing the discharge of three measuring devices. The data span six years (from 2016 to 2021), encompassing a wide range of hydrological conditions. In summer, negative discharge values were observed due to the low river flow and tidal currents. The highest discharge estimations are observed during the winter period, with values of around 1890 m^3/s, which is significantly larger, approximately five times the average annual discharge of 323 m^3/s. The reliability of the fixed acoustic instruments has already been pointed out in several studies [39, 57, 105]. The challenging aspects of these instruments are the acquisition cost and maintenance requirements.

It is not possible to estimate discharge at the Metković station directly from the water level records of its hydrological station. Such complexity is depicted in Figure 3.2, which illustrates the relationship, i.e., the correlation between water level and discharge, in both hourly and daily averaged formats, with a non-unique one-to-one relationship between the hydrological parameters, especially when focusing on lower and average flow scenarios. This can be attributed primarily to the impact of sea levels, but also the non-stationary flow dynamic, tidal propagation and influence of the backwater effect. Therefore, applying a simple approach such as Stage-Discharge rating curve is not appropriate, and other approaches should be considered to enable accurate flow estimation and prediction of river discharge at the Metković station.

(a) Hourly [88]   **(b)** Daily averaged
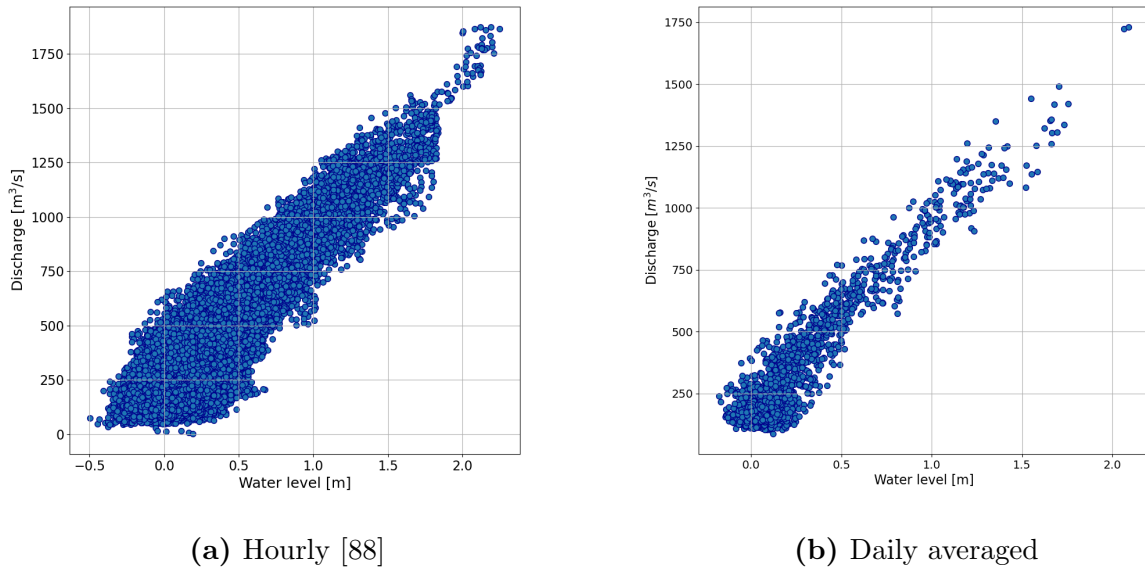
Figure 3.2: Stage-Discharge relationship at Metković station

The problem with measuring devices like the H-ADCP is the need for periodic calibration. Therefore, additional flow measurements are performed to ensure that the devices are correctly calibrated. However, in addition to this, different problems have a negative impact on the quality of the measured data, and some of these problems refer to non-physical oscillations, missing records, the complexity of low flows, but also deviations during data synchronization when several institutions collect them, and each of these institutions performs processing differently. Given the potential issues than can occur in the measured dataset, the methodology also includes simulated data, primarily to ensure reliability and robustness of proposed approaches.

3.2.2. Simulated Data

Two models were used for generating simulated data: HEC-RAS and STREAM, both of which are based on the one-dimensional (1D) shallow water equation (SWE). In a simpler model (HEC-RAS), the change in water level at individual stations along the Neretva River is the result of changes in sea level and river flow, assuming a homogeneous fluid of uniform salinity and density (freshwater only). In the second, more complex model (STREAM 1D), the change in water level at individual stations is the result of sea level changes, salt-wedge dynamics, and river flow, and it assumes a layered structure with two layers of different densities. The following subsections present details regarding the utilized models.

STRatified EstaArine Model

A numerical simulation was performed using a STRatified EstaArine Model (STREAM) model to generate the simulated dataset. STREAM is as a time-dependent and one-dimensional (1D) model based on the SWE, specifically developed for simulating water flow in microtidal estuarine regions [66]. The preference for 1D models, particularly two-layer shallow-water models, rather than the frequently used three-dimensional (3D) ones, primarily stems from their simplicity and lower computational demands, as well as their ability to adequately represent thin interfacial layers in strongly stratified salt-wedge estuaries [66]. The STREAM model has been effectively utilized in two Croatian estuaries and tidal rivers, the Rječina River and the Neretva River, demonstrating satisfactory accuracy in modeling the two-layer flow dynamics [66, 68, 70].

The governing equations for the STREAM 1D model are based on the following assumptions and simplifications [71]. The vertical salt-wedge structure is represented by two shallow layers of differing densities, separated by a pycnocline of negligible thickness, as illustrated in Figure 3.3. According to the standard SWE framework, it is assumed that velocity and density are uniform within each layer, vertical accelerations are insignificant, pressure distribution is hydrostatic, the channel bed has a gentle slope, and viscous effects such as friction and turbulence can be described using empirical relationships. The flow is primarily one-dimensional, and the channel geometry is defined by a sequence of irregular cross sections. The resulting governing equations form a coupled system of conservation

laws with source terms [66].

However, these equations are further extended to incorporate the effects of irregular channel geometry, along with detailed parameterizations of shear stress and entrainment. As a result, STREAM 1D can simulate the behavior of salt-wedges in channelized estuaries under varying sea levels and river discharge conditions [70].

(a) Cross section view

(b) Longnitudinal view
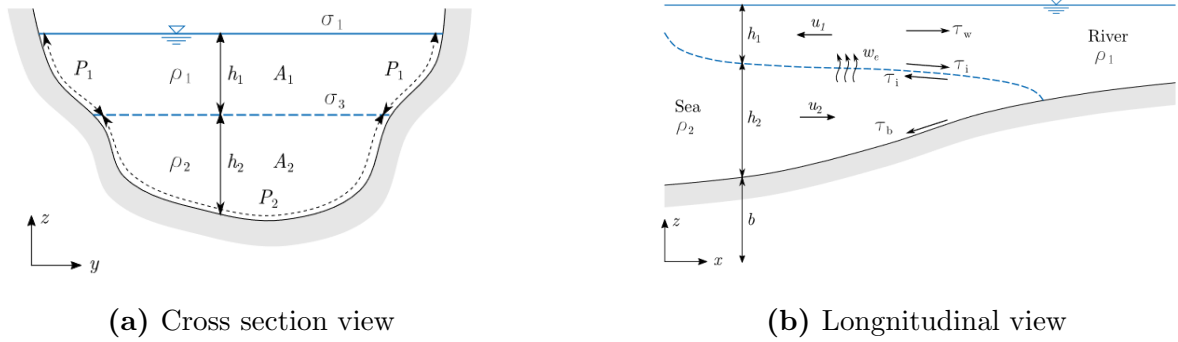
Figure 3.3: Characteristic two-layer system scheme of a STREAM 1D model for a salt-wedge estuary [66, 70]

The domain of the study area, which extends beyond the tidal limit (35 km from the river mouth), is based on the channel geometry derived from multiple cross sections (an example is illustrated in Figure 3.3a) within the Neretva tidal reach. The schematic representation of the cross section and longitudinal profile of a salt-wedge estuary is described by a set of physical and geometrical variables. The freshwater layer of constant density ρ_1 and thickness $h_1(x,t)$ flows over a salt-water layer of constant density ρ_2 and thickness $h_2(x,t)$. The ratio between the upper and lower layer density is denoted by $r = \rho_1/\rho_2$. The flow rate in each layer is defined as $Q_j(x,t) = u_j(x,t)\,A_j(x,t)$,, where $u_j(x,t)$ is the horizontal velocity, $A_j(x,t)$ is the cross sectional area, and index $j = 1, 2$ refers to the upper and lower layer, respectively. The bed elevation function is defined by $b(x)$, $\sigma_1(x,t)$ is the breadth at the free surface, and $\sigma_3(x,t)$ is the breadth at the interface between the layers. [68] The remaining variables, refer to the fluid velocity in each layer denoted by u_j, entrainment from the lower to the upper layer w_e, shear stress between the upper layer and channel bed T_w, shear stress at the interface between two layers T_i, and shear stress between the lower layer and channel bed T_b [66].

Performing the simulation requires defining boundary conditions and friction coefficients. The boundary conditions are defined for the upstream and downstream sections of the study area. The upstream boundary condition requires a time-series of discharge

data. This data is processed first by shifting the time-series by one hour and then appying a median filter with a three-hour window. The one hour shift is applied to account for the distance between the Metković station where discharges are measured and upstream boundary condition where the time-series is set. The median filter is applied to eliminate high-frequency noise and errors. The sea level measurements from the Ušče station are set at the downstream boundary condition (at the river mouth). Due to the two-layer structure, a second condition must be set at the downstream boundary condition. Its definition is based on the critical condition of two-layer flow that defines the depth of the interfaces between the upper and lower layer. A detailed description of the STREAM 1D implementation is given in the study by Krvavica et al. [66].

The numerical model adequately captures the complex dynamics of the estuary, as evident in the simulations (see Figure 3.4) that illustrate the typical behavior of the salt-wedge for different flow levels, ranging from low to high. Under low flow conditions, for example, approximately 125 m³/s (see figure 3.4a), the saltwedge can advance upstream of the Metković station. The outcome of such intrusion is nearly uniform, i.e., horizontal water level across the whole reach. The gradual advance of the saltwedge towards the river mouth is noticeable during average flow conditions, as shown in figures 3.4b and 3.4c with flow values around 335 and 570 m³/s. The gradient shows a smaller increase compared to the horizontal water surface over the tidal range, which is a consequence of the proximity of the saltwedge to the river mouth. A third possible outcome occurs under high flow conditions, characterized by a steep slope upstream from the river mouth, when the salt-wedge is strictly confined to the river mouth area. Figure 3.4d shows a high flow scenario characterized by the flushed saltwedge from the tidal area.

HEC-RAS

Another dataset was created using the Hydrologic Engineering Center-River Analysis System (HEC-RAS), a simulation software used for modeling water flow in various river environments [137]. The software supports the creation of 1D, 2D, or even a combined 1D/2D simulations, although this mainly depends on the modeling requirements. Also, HEC-RAS can effectively simulate steady or unsteady flow conditions. The 2D simulations are run by solving the SWE or the Diffusion Wave Equations.

Unlike the previous simulation using STREAM 1D, in which the flow is divided into

(a) 125 m^3/s

(b) 335 m^3/s

(c) 570 m^3/s

(d) 1003 m^3/s
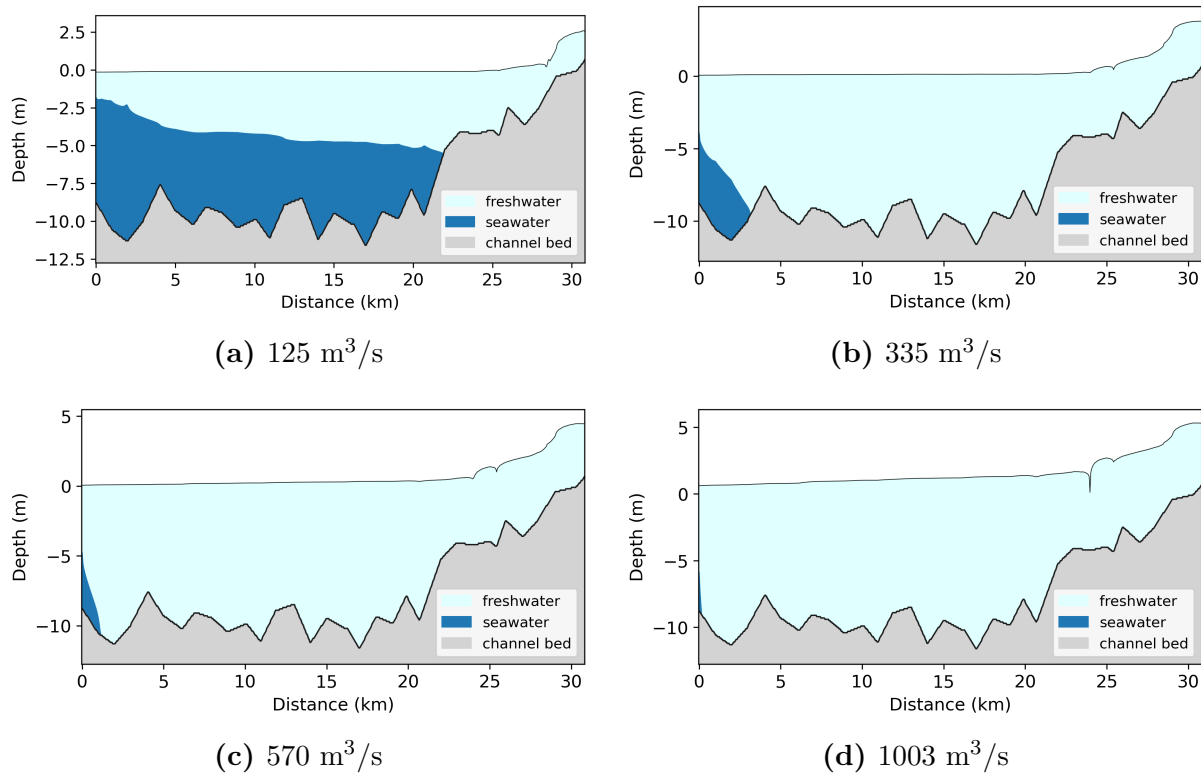
Figure 3.4: Changes in longitudinal profile of the River Neretva water levels according to different river discharges [88]

two layers by a pycocline, the employed HEC-RAS simulation is barotropic, meaning that salinity and temperature are constant throughout the entire considered channel (riverbed length of 46 km), and the fluid is characterized by equal density, i.e., it is homogeneous. To create such simulations, three prerequisites must be met. Firstly, the details about the channel geometry must be uploaded into the software, as well as two boundary conditions. Boundary conditions are defined for the lowest and highest points of the river channel. Hence, the measured sea levels must be provided for the downstream section of the river channel, while the flow rates must be provided for the upstream reach. Equaly, as in Subsections 3.2.1. and 3.2.2., water level and discharge data have been generated at four locations (Ušče, Opuzen, Norin, Metković).

After presenting the research area and the process of data collection and numerical simulations, the next chapter presents a literature review on machine learning approaches for hydrological analysis of water level and discharge.

Chapter 4

RELATED WORK ON MACHINE LEARNING METHODS

Contents

This chapter explores the application of machine learning approaches for discharge estimation and prediction in tidal rivers and estuaries. The reviewed studies primarily focus on the hydrological analysis of water levels and discharge. The literature is classified based on the type of analysis performed, and a critical discussion is provided, highlighting existing research gaps as well as the strengths and limitations of each category of models.

4.1. Machine Learning for Hydrological Analysis: A Modern Approach

Inverse problems, along with the nonlinear and complex hydrological processes characteristic of tidal rivers and estuaries, require robust solutions to ensure sustainability and mitigate the risk of potential disasters. These challenges can be addressed through the implementation of monitoring, modeling, and forecasting systems that remain effective even under extreme conditions. The development of such systems is essential for effective water resource management.

The application of machine learning for discharge estimation and prediction can be attributed to the limitations of existing approaches. In hydrology, modeling complex river processes is mainly accomplished using numerical models. Although machine learning is still to some degree considered a black-box approach, especially in the case of neural networks, the numerical models are substantially applied in the field of hydrology due to their advantage of providing adequate comprehension of rivers' physical mechanisms. There are two basic categories of numerical models that differ based on how they describe the physical processes: conceptual hydrologic and physically based hydraulic models [90]. The first category of conceptual hydrologic models represents a more simplified version of modeling complex physical laws, which does not require large datasets with numerous parameters and high spatial resolution. Hence, such models are deemed computationally efficient compared to the physically based hydraulic models. In contrast to such a simplified approach, physically based hydraulic models solve complex differential equations for the purpose of simulating the physical processes, thereby providing higher accuracy, unlike the previous category. Likewise, the second category requires more precise data in both temporal and spatial aspects, such as boundary conditions, river cross-section, and many more.

As physically based numerical models provide accurate simulations of complex river flow, they are of more interest for discussion. However, several difficulties may occur when using these models, with different sources. One of the probable issues is the data requirement aspect. Hydraulic numerical models require a wide range of input data, including not only hydrological time-series but also detailed information about channel

geometry, physical characteristics, and boundary conditions, which are often difficult to obtain. Likewise, another potential problem relates to the model calibration, which must be performed at appropriate intervals. The application of these models is additionally problematized due to the complex dynamics of tidal rivers and estuaries. The definition of boundary conditions depends on high-resolution data continuously collected, and a long dataset is necessary for performing tidal analysis.

Web of Science (WOS), a well-known bibliographic database of scientific papers, conferences, and other publications, has been used for advanced searches using specific relevant keywords and terms. Although our primary focus was on discharge parameters, we also provide an overview of research articles referring to another hydrological parameter, water level. These parameters also differ in the underlying purpose for their analysis. For example, water levels are mainly employed for developing early warning and flood forecasting systems. While, for example, discharge is utilized to enable the management of water resources and processes in river systems, as well as the development of models for reanalysis and prediction.

The search engine, based on the query defined in [90], presented 389 publications. However, it is important to emphasize that only research and review papers were considered. However, although the search had narrowed down the number of publications, additional manual checking of the filtered studies' abstracts was necessary to find those that specifically address the problems of modeling and predicting water levels and discharge in tidal rivers and estuaries. The search was further narrowed to articles published in the last 20 years. Out of a total of 389 studies, only 35 focused on the requested problems, while the remaining studies mostly dealt with other processes or parameters, including water quality, runoff, sediment transport, saltwater intrusion, and even biological indicators.

Until recently, there has been a lack of comprehensive studies that offer a literature review specifically addressing tidal rivers and estuaries, related to the estimation and prediction of hydrological parameters: water level and discharge. The available review papers focused mainly on the area of inland rivers and forecasting problems. Table 4.1 presents several examples of the review studies, highlighting their problem, output parameter, and period range.

Based on the 35 papers previously found for the problem at hand, a classification of machine learning categories for estimating water level and discharge was presented in a

Table 4.1: A brief insight into several review papers, their research focuses and the periods considered

Authors and publication year	Hydrological parameter	Type of analysis	Period
Yaseen et al. (2015) [135]	streamflow	modeling and forecasting	2000 - 2015
Zhang et al. (2018) [141]	streamflow	forecasting	2001 - 2017
Hamzah et al. (2020) [51]	streamflow	reconstruction	2000 - 2019
Zhu et al. (2020) [142]	lake water level	forecasting	2006 - 2020
Wee et al. (2021) [128]	water level	forecasting	2000 - 2020
Ibrahim et al. (2022) [61]	streamflow	forecasting	2009 - 2020
Ng et al. (2023) [94]	streamflow	forecasting	2017 - 2023

review paper by Mihel et al. [90], which focused specifically on the areas of tidal rivers and estuaries. The following five main categories were identified in the review: (a) Simple statistical approach, (b) Classifiers, kernel, and ensemble approach, (c) Shallow Neural Network (SNN) approach, (d) Recurrent Neural Network (RNN) approach, and (e) Hybrid approach. In addition to the classification of the machine learning approaches, the filtered studies that focus on water level and discharge as output features also differed in the type of analysis conducted. Therefore, three types of analysis were identified: forecasting, reconstruction, and establishing a stage-based discharge relationship. The summary of the literature is provided in Appendix B. In the following subsections, a brief overview of the papers pertaining to each analysis is presented. Every type of analysis was divided into three periods: the first published papers between 2000 and 2011 fall under the pioneering work, followed by the papers published between 2012 and 2017, which are referred to as early applications, and lastly, the recent advancement category, which encompasses publications between 2018 and 2023. However, another category is additionally added, which presents studies published since 2024, named the latest contributions. Therefore, instead of 35 studies covered in the review paper, this chapter presents an overview of 47 studies.

4.2. Forecasting Studies

This section presents publications that focus on the predictive modeling of hydrological parameters, employing historical data for model training and aiming to forecast unknown future values. A total of 27 publications have been found, where more than 80% of the publications (24 research papers) focused on forecasting water levels, while the re-

maining (3 research papers) were dedicated to discharge analysis. The classification of papers by utilized machine learning approaches is presented in Table 4.2, while additional information on the studies is included in Appendix B.1, Tables B.1 and B.2.

Two pioneering forecasting studies were published in 2003, aiming to predict water levels several hours in advance. The first paper by Supharatid [115] examined the use of a Multilayer Feed-Forward (MLFF) neural network and combined it with a Levenberg-Marquardt algorithm for the area of Chao Phraya River estuary located in Thailand. The study emphasized two points: the first related to the use of different training algorithms, where the proposed training algorithm resulted in a shorter training time, and the fact that having tidal constituents is not essential for obtaining accurate long-term forecasts. Chang and Chen [15] introduced a hybrid methodology to address the issue of water level prediction in the Tanshui River in Taiwan, where the effectiveness of a supervised Radial Basis Function Neural Network (RBFNN) combined with a fuzzy minmax clustering technique was evaluated. The model demonstrated satisfying accuracy for one-hour ahead forecasting during both average and extreme conditions, i.e., during events such as typhoons and floods. The proposed integration of unsupervised and supervised methods facilitated the automatic selection of model parameters.

After an almost nine-year gap, the next phase, referred to as early applications, began, with a study by Tsai et al. [121] where a hybrid approach was presented for the Tanshui River tidal region. The proposed approach combined Classification and Regression Trees (CART) with two variants of Multilayer Perceptron (MLP) Artificial Neural Networks (ANNs): one using standard activation functions, and one employing radial basis function (RBF) neurons in the hidden layer. However, although one of the proposed variations (CART-RBF) provided higher accuracy than other models used, the study emphasized two key limitations: one regarding the case-dependent nature of the model, and the other regarding the model's generalization being completely dependent on the given training set.

The first application of wavelet function integrated with a kernel approach, known as support vector machine, in the tidal reach of Tanshui River, located in Taiwan, was published by Wei [129] in 2012. The water level forecasting model demonstrated improvement in comparison to a Gaussian support vector machine approach during extreme weather related to typhoon events. The next year, another publication employed wavelet

functions in both continuous and discrete forms, but this time alongside a neural network. The mentioned study by Yang et al. [134] was conducted for the Yangtze River in the form of a multi-step ahead forecasting scenario, wherein the continuous and discrete wavelet transform combined with the neuro-fuzzy system (CWD-NF) exhibited minimal error accumulation, thereby demonstrating the model's efficacy in successfully filtering noise from the provided signals.

The pioneering study on discharge forecasting in the tidal region of the Mahakam River was published by Hidayat et al. [55], who demonstrated an additional use of the Levenberg-Marquardt algorithm for training a neural network, specifically an MLP model, expanding upon the findings of [115]. This study considered using historical discharges, wavelet analysis for tidal data, and a feature selection approach, which resulted in a model with satisfying performance up to 48 hours ahead. The continuation of ongoing analysis by Wei [130] on water level forecasting during typhoon events for the same tidal area as before, has been presented in 2015, but this time in a different manner, through an assessment of two classes of machine learning algorithms, namely lazy and eager techniques. Despite the comparative analysis of these classes' performance, no definitive conclusion could be drawn regarding a more suitable choice. However, when evaluating each class individually, it was observed that Locally Weighted Regression for lazy learners and SVR and ANN for eager learners outperformed the other models considered. Only a year later, a more detailed comparison was given by Pasupa and Jungjareantrat [96] for the Chao Phraya river, which featured a baseline model founded on Harmonic Analysis (HA) and various machine learning models. The analysis demonstrated that the baseline model, the harmonic method of tidal model, effectively represented the general trend, although it lacked precision when individual predictions were considered. Consequently, the evaluated machine learning models yielded better results. A particular focus was on the performance of SVR when paired with a radial basis function kernel, resulting in the lowest errors for the longest considered forecasting horizon.

Ahmed et al. [3] introduced a hybrid approach for the Karnaphuli River, which combined the SVR model with various kernel functions, including the commonly used radial basis function kernel, the sigmoid kernel and analysis of variance, known as ANOVA. The preprocessing stage integrated methods such as moving average (MA) and exponential moving average (EMA), followed by a resampling process where the daily data was

additionally classified into morning and night tide records. As the SVR model cannot directly handle missing records, they were imputed using the mean of all available tidal observations. Satisfying accuracy was achieved for all tested models, more precisely above 96%, calculated as the complement of the mean absolute error. The inclusion of water level data from multiple gauging stations to address the problem of backwater effect from the main channel was explored in a study by Sung et al. [113]. The study focused on a tributary known as Anyangcheon stream, which is connected to the main stream, the Han River. Various input feature combinations were evaluated using the MLP approach, from which it was concluded that including the main river water level records considerably improved the model's performance. Hence, MLP was proved to be a suitable and efficient alternative to complex physically based hydrological models.

Compared to the publications from the preceding two periods, the recent advancements period featured eight papers, which is noteworthy relative to the earlier total of 10 publications. The data indicate a continuous rise in publication trends, partially attributable to the intensifying effects of climate change.

The initial research employing LSTM for predicting hydrological parameters in tidal regions was carried out by Jung et al. [64]. The study examined the Han River region, drawing comparisons to prior research [113], which analyzed tributaries of the Han River and the water levels of the main river. The study's focus was on a bridge situated on the Han River, which becomes obstructed for both vehicles and pedestrians during flood events. The optimal hyperparameters for the LSTM model have been determined through sensitivity analysis, and various input sequence lengths were assessed during the model evaluation phase. A data input sequence of up to 1 hour yielded the best model performance, but only for shorter lead times. Yoo et al. [136] released another LSTM research on water level forecasts. The main difference between this study and the first [64] is the use of a weighted sum multiplied by a hyperbolic tangent ($tanh$) and a rectified linear activation unit ($ReLu$) instead of using the traditional activation function approach. The output of such an approach was the sum of these two components. Likewise, t-test and p-value were used to find relevant features, while sensitivity analysis was applied to obtain optimal hyperparameters, as in [64]. Results revealed that the hybrid activation function improved model performance over the traditional approach while maintaining acceptable accuracy for shorter forecasting horizons (up to 6 hours).

A tool known as non-stationary harmonic tidal analysis model (NS_TIDE), firstly proposed by Matte et al. [87], was further investigated in a study by Chen et al. [20] where the approach had been combined with auto-regression (AR) analysis with the main aim of improving the model accuracy for short-term forecasting for Yangtze River. The results show that the subtidal tides are adequately modeled when using the proposed hybrid approach, which was a major limitation of the stand-alone NS_TIDE model.

Liang et al. [76] introduced hybrid modeling for tidal level forecasting, integrating Empirical Wavelet Transform (EWT) with the machine learning approach Nonlinear Autoregressive with Exogenous inputs (NARX). EWT was used to preprocess data from four Pearl River estuary stations by generating intrinsic mode functions (IMFs) to forecast tidal levels using a time-series format. The suggested hybrid model was compared to baseline HA, Empirical Model Decomposition (EMD) combined with NARX, and Ensemble Empirical Mode Decomposition (EEMD) combined with NARX. EWT-NARX considerably outperformed other approaches in predicting tidal levels. EWT also avoided mode mixing, a limitation of EMD and EEMD, whose deconstructed signals had real physical meaning. The utilization of the LGBM in the area of tidal reaches, specifically for the purpose of water stage forecasting, has not yet been investigated. Consequently, Guo et al. [47] conducted a novel analysis in which the results of LGBM were compared to those of previously employed models, including RF, SVR, and MLP for the Lan-Yang River. In contrast to the grid search approach previously employed in various studies, a Bayesian optimization (BO) approach was selected to identify optimal model parameters. The study showed how the proposed model provided higher accuracy than others, deeming the proposed approach robust and efficient.

Storm surge forecasting requires a model that is both efficient and fast. Consequently, Chen et al. [17] conducted a comprehensive comparative analysis of the LSTM model's efficacy against simpler models, including Linear Regression (LR), Bayesian Ridge Regression (BRR), SVR, and Gradient Boosted Decision Tree, for the Yangtze River. Additional input features, including meteorological data and reference factors, were considered, in contrast to earlier studies. While some difficulties arose during extended forecasts, overall, the model demonstrated superior performance compared to simpler models. Zhang et al. [140] expands upon the work of Chen et al. [20], where the previous proposed hybrid approach was further evaluated and compared to a novel hybrid approach comprised of

NS_TIDE and a deep neural network, for the Pearl River. By employing the correlation analysis, the significance and strong relationship between NS_TIDE errors were confirmed. Various neural network architectures were considered and evaluated against the previously proposed AR, and it was found that a model combining the LSTM with the Feed-Forward Neural Network (FFNN) layer provided the best performance. Similarly, testing revealed that the effectiveness of including a dropout layer depended on the number of neurons.

A simple novel approach for water level forecasting that combined EEMD with stepwise regression was proposed by Chen et al. [22] for the tidal reach of Tanshui River, more precisely, Taipei Bridge. Only water level signals were considered as model input features, which were decomposed into IMFs, reconstructing the ocean and stream components. These two components were modeled and forecasted separately using stepwise regression, with the ocean using downstream data and the stream using data from upstream stations. The water level at Taipei Bridge was obtained by summing the results of these two component forecasts. Through evaluation, it was revealed that the model could achieve accurate forecasts even during extreme events, such as typhoons.

The first application of a graph convolutional recurrent network in tidal reaches was offered by Zhang et al. [139] to address the issue of storm surges, while the research area remained the same as in the previous study. The model consisted of Chebyshev Graph Convolution (Chebnet) [116], whose purpose was to extract spatial information, and Gated Recurrent Unit (GRU), which extracted temporal information. A comprehensive evaluation included various baseline models (LGBM, GRU, LSTM, and CNN-LSTM) to predict multiple station outputs. The proposed hybrid approach consistently outperformed baseline models even for extended forecasting horizons (up to 12 hours ahead). A study by Vu et al. [126] presents another forecasting application of an LSTM model. In contrast to earlier applications, this study incorporates previously omitted input features, including climate data and sea level at both local and global scales, facilitating long-term predictions extending several months into the future. Due to the problem of different scales, a frequency analysis was conducted and aimed at two objectives: optimizing lags and the correlation among the chosen features. The optimal input step (6 days) was sufficient for long-term forecasting, although, greater inaccuracies were observed during floods, than during drought events.

The latest contributions category consists of 7 studies. The first application of ML

analysis for the tidal reach, Río de la Plata Estuary, was presented in a study by Dato et al. [30]. The authors employed an MLFF model to forecast water levels up to four days ahead at two stations (intermediate and upstream sections of the estuary), using water levels, astronomical tide predictions, and meteorological data. Two types of models were applied, which are referred to as Forecast Unweighted and Forecast Weighted. The difference between the models was that for the unweighted scenario, the models were trained with real data sampling, i.e., unequal data distribution. Unlike this, the weighted case had data with equal distribution, which was useful for predicting extreme values. The findings indicate the effectiveness of the proposed model, revealing that the weighted scenario errors have nearly tripled for extended lead times. In contrast, the unweighted errors have only doubled, despite the weighted errors yielding lower errors.

An extension of previously applied AR [20], NARX was the focus of a study by Vidyalashmi et al. [124]. The NARX model was applied to forecast water level using three different combinations of inputs, one containing only water level data, the second containing additional discharge, and the last containing salinity. Through cross-correlation analysis, it was found that the inclusion of salinity in the third input combination largely influenced the result, from which it can be concluded that tides strongly influenced the water level. However, despite this, all models performed with satisfying accuracy. Gan et al. [36] tested two different implementations of the LGBM model for water level forecasting in the Yangtze Estuary up to two days in advance. The first model, denoted as LGBM1, represented a single regression model that employed a recursive approach, while the second model, LGBM2, generated multiple regression models depending on the maximum output step. Results indicate that LGBM2 outperformed LGBM1, especially for longer horizons, but at a cost of additional time. However, the accuracy of proposed models was still limited in the case of extreme events (typhoons), showcasing larger errors.

An approach consisting of an LSTM combined with a sequence-to-sequence (Seq2Seq) model was discussed in a study by Chen et al. [23]. The study aimed to predict discharge for three possible scenarios: short-term, mid-term, and long-term, in the Yangtze River using discharge, flow velocity gathered from three ADCP devices, and water level. Likewise, the model was trained and tested using different hydrological conditions. During the comparative analysis, the proposed model showed improvement over other approaches (LSTM, HA combined with Back-propagation and optimized with Particle Swarm Op-

timization (PSO-BP), and HA), with a consistent and rising percentage of prediction accuracy for longer lead times, with an emphasis on effectiveness in capturing peak values both during ebb tides and floods. A continuation of multi-station tide level prediction research [139] during storm surges was proposed by Shi et al. [109]. The authors used the same hybrid approach of graph convolution recurrent network (Cheb-GRU), but this time for a different tidal reach, the Yangtze Estuary. Cheb-GRU had greater accuracy than other models (LSTM, GRU, and CNN-LSTM) for the one-hour-ahead forecasting scenario. However, its performance decreased and resulted in instability for longer forecasting horizons. Likewise, the model was applied to the Alaska region to test its generalization ability for different areas under the tidal influence, with a larger dataset than initially used, showcasing its considerable performance improvements. Another study contribution referred to the tested topologies (simple and connected) to unweighted or weighted networks, where a more detailed network resulted in higher accuracy and a good fit.

The first mention of data assimilation was given in a study by Cremer et al. [28], where it represented a part of the employed hybrid approach for the tidal reach of the Elbe River. The Ensemble Kalman Filter (EnKF) represented the assimilation part, MIKE Flexible Mesh (FM), the numerical model, and LSTM, a machine learning approach. The LSTM was initially employed to produce synthetic data through forecasting. The synthetic data was integrated with the actual observations with EnKF, subsequently serving as input data for the numerical model. The proposed hybrid approach gave considerable improvements for shorter forecasting horizons (up to 4 hours ahead), while there was less difference in performance for longer horizons (up to 9 hours ahead) when compared to the stand-alone numerical model and the numerical model combined with data assimilation.

Almost four years after the publication of the first forecasting study on tidal reaches [76], which utilized signal preprocessing, EEMD, another study emerged. An EMD for tidal reaches, with various ML methods for water level prediction in the Yangtze Estuary, was proposed by Gao et al. [38]. The integrated modeling system EMD-ITG consisted of both LSTM and GRU models, which were applied separately, depending on the type of IMFs. Hence, for low to medium-frequency IMFs, the LSTM was used, while for high-frequency IMFs, the GRU was used. EMD-ITG outperformed different EMD ML combinations (EMD-LSTM, EMD-GRU, and EMD-CNN-LSTM), stand-alone models (LSTM, GRU, CNN-LSTM), and a numerical approach (NS_TIDE). Hence, ap-

plying EMD decomposition resulted in more accurate predictions, which showed greater resistance to the presence of noise and trend.

Author and publication year	Simple statistical approach	Classifiers, kernel and ensemble approach	SNN approach	RNN approach	Hybrid approach
Supharatid (2003) [115]			✓		
Chang and Chen (2003) [15]					✓
Tsai et al. (2012) [121]		✓	✓		✓
Wei (2012) [129]		✓			✓
Yang et al. (2013) [134]					✓
Hidayat et al. (2014) [55]			✓		✓
Wei (2015) [130]	✓	✓	✓		
Pasupa and Jungjareantrat (2016) [96]	✓	✓			
Ahmed et al. (2017) [3]		✓			✓
Sung et al. (2017) [113]			✓		
Jung et al. (2018) [64]				✓	
Yoo et al. (2020) [136]				✓	
Chen et al. (2020) [20]					✓
Liang et al. (2021) [76]					✓
Gou et al. (2021) [47]		✓	✓		
Chen et al. (2021) [17]	✓	✓		✓	
Zhang et al. (2023) [140]					✓
Chen et al. (2023) [22]					✓
Zhang et al. (2023) [139]				✓	✓
Vu et al. (2023) [126]				✓	
Dato et al. (2024) [30]			✓		
Vidyalashmi et al. (2024) [124]	✓				
Gan et al. (2024) [36]		✓			
Chen et al. (2024) [23]				✓	✓
Shi et al. (2024) [109]				✓	✓
Cremer et al. (2025) [28]					✓
Gao et al. (2025) [38]				✓	✓

Table 4.2: Forecasting study categorization

4.3.   Reconstruction Studies

The second category referred to the reconstruction analysis. This category addressed several potential problems regarding the estimation of hydrological parameters (water level and discharge): (a) at locations of interest (remote or ungauged), and (b) for missing time periods (ML imputation, hindcasting, upscaling current temporal resolution) and (c) optimization of complex water systems [90]. The reconstruction problem was the subject of 19 studies, of which 14 focused on water level and the remaining five on discharge as the model output. Table 4.3 presents published papers in chronological order, categorized according to defined machine learning approaches. A more detailed overview of studies is presented in the Appendix B.2, more precisely, Tables B.3 and B.4.

For the pioneering period, three publications were found that used simple ANNs, or more precisely, SNNs, for the purpose of water level reconstruction. The first study by Adib [1] used an MLP for the area of two rivers located in different countries, the Karun and the Severn River. The data used to train the MLP were generated using a numerical model, while the measured data were used to establish the regression relation. A comparison of the two approaches showed that the ANN resulted in adequate results, especially when it came to situations of extremely high water levels. It was also observed that the regression relation approach largely depended on the availability of data from the gauging stations, as well as on the length of the time-series when performing the calibration to determine the model parameters.

Aiming at the flood control problem, a paper by Wei and Hsu [131] explored the area of Tanshui River, where another SNN was applied, namely FFNN with a classical back-propagation (FFBP). As flood control was the primary objective of the research, data collected during typhoon events was utilized as the model's input, however, besides the main channel, tributary data was also considered. The rationale for utilizing an ANN, stems from the constraints of the physically based model, gradually varying unsteady flow model, which deals with the compound-complex channel (network or dendritic) system according to the multimode method of characteristics (CCCMMOC), which was inefficient for multireservoir operational modeling due to its complexity. The comparison between an FFBP and CCCMMOC revealed how the proposed ANN represented a suitable alternative. A somewhat different challenge was investigated in a study by Chinh et al. [24], for the Chiyoda Basin, regarding the problem of water distribution, mainly the irrigation and drainage system. A relationship between the amount of rainfall and water level was successfully modeled using an FFNN model for estimation of water level at two main channel river locations, where the water levels in the downstream section were largely influenced by tidal effects.

The following period of early applications was marked by two directions of conducted research, one which focused on the comparison and improvement of models, primarily simple ANN models and existing numerical models, and the other direction, where the problem of discharge reconstruction in tidal reaches was introduced for the first time. The first direction had three studies published sequentially, one after another, from 2012 to 2014. Chen et al. [19] was the first to compare FFNN with backpropagation to both

a 2D and 3D hydrodynamic model for the area of the Danshui River. The difference between the models was not as substantial, although ANN performed better for some locations. The following paper by Pierini et al. [97] conducted another comparison of FFNN with a backpropagation, but this time to a MOHID (Portuguese acronym for *MOdelao HIDrodinmica*) hydrodynamic model, for the Bahia Blanca River. This time the deviation of the hydrodynamic model was considerably larger than the proposed FFBP model, hence, the paper suggested further exploring the proposed architecture by feeding the network with outputs of the MOHID model. As the Danshui River region is vulnerable to extreme events such as typhoons, providing accurate simulations for the designated area is of critical importance. To address this challenge, Liu and Chung [78] presented a method that integrates a genetic algorithm with a traditional backpropagation neural network, and its efficacy was subsequently compared to that of a hydrodynamic model and a conventional neural network utilizing backpropagation. The results showed that the accuracy of the Genetic Algorithm Neural Network (GANN) model was consistently higher than the other two approaches for each location, and thus, it was concluded that the model meets the desired performance criteria.

The remaining research direction describes studies regarding the reconstruction of discharge, starting with an article by Gu et al. [44] for the Pudong New Area. The authors tested a hybrid approach, combining a numerical model (River Network Mathematical Model (RNMM)), a common neural network with a backpropagation algorithm, and a genetic algorithm aiming to optimize sluice management. The main operations involved in creating the proposed framework were as follows: the first part was training the ANN using the data generated by the RNMM model, and the second part involved the ANN, which acted as the fitness evaluator for the sluice operation solutions proposed by the Genetic algorithm (GA). The framework was found to attain sub-optimal rules for sluices, which represent practical operation strategies that efficiently balance multiple goals, while demonstrating notable robustness, speed, and flexibility. Another approach was presented by Hidayat et al. [55], which not only provided an innovative solution in the forecasting analysis, but also in the field of reconstruction. The study area remained the same, i.e., the Mahakam River, while the input data differed as wavelet analysis was applied to generate tidal components, which are used as inputs to the MLP. Hence, a hybrid approach was utilized. The analysis proved the feasibility of reconstructing discharge using solely water

level and tidal components, with only minor variations observed in extreme scenarios (both low and high). Subsequently, a study by Garel and Dalimonte [39] developed an MLP approach for freshwater discharge estimation in the Guardiana River, whose estuary is described as narrow. A single ADCP device was placed at the deepest area of the river where flow velocities are the highest. A quasi-stationary relationship between the maximum flow and discharge was identified using ADCP data. The proposed approach is used in cases where it is not possible to directly determine the entropy-based velocity ratio (Ω) due to the lack of data on the average flow velocity, with the ultimate goal of eliminating the need for demanding transverse measurements across the entire channel section. The performance of MLP was deemed satisfactory when the data represented all possible variations in river flow dynamics.

The recent advancement category of publications began with Bhar and Bakshi [12]. Their research focused on resolving the issue regarding data accessibility and availability at the downstream station located on the Hooghly River, using an FFBP model. To facilitate the station of interest with continuous data, as it provides records for only half a tidal cycle, data from upstream stations were selected as the model's inputs, and the model performance was evaluated across several combinations. It was observed that the performance does not improve with the inclusion of additional records from more distant upstream stations. Accordingly, only the stations closest to the one of interest were chosen as the optimal input combination. A satisfactory reconstruction accuracy of the model was achieved even during spring and neap cycles.

Following this, two studies were conducted by Guillou and Chapalain [45] and Gan et al. [37]. The first study compared multiple regression approaches (linear and polynomial) with a neural network, representing an alternative to previously used numerical models. The area of interest was the Elorn River. Unlike the previous study, input parameters such as the French tidal coefficient, atmospheric pressure, and wind speed were used, along with a discharge parameter to estimate the maximum water level. The study showed small differences in performance between the selected models, but the MLP still proved to be better, especially during conditions where extremely high water levels were encountered. However, all models experienced minor underestimations during those conditions. The second study involved a comparison between an ensemble approach, LGBM, and a numerical approach, NS_TIDE. The research focused on estimating water levels

at the downstream section of the Columbia River while using discharge data from both the main river and its tributary (Willamette River), as well as tidal data. Despite the pronounced distinctions in the nature of the tested approaches, their findings were found to be comparable. Likewise, a phase lag was detected during the flooding season for the LGBM model, resulting in increased prediction errors during these periods.

The compound flooding problem in Kapuas River was discussed in a study by Sampurno et al. [104]. The hybrid framework was employed, which consisted of two strategies: (1) employing a multiscale-hydrodynamic model, named SLIM 2D, for simulation of the compound flooding, and (2) where the data obtained as the output of the simulation is used for training simple machine learning models, as Multiple Linear Regression (MLR), RF, and SVR. This means that the accuracy of simple ML models is closely tied to the accuracy of a hydrodynamic model. In comparison, MLR performed the poorest, although, within the acceptable performance range, the RF predictions were the closest to the empirical observations. Around the same time, Thanh Hoan et al. [118] also suggested a hybrid modeling approach for addressing the problem of water level hindcasting for 18 water level stations on a daily scale in the Mekong River. Three bagging-based hybrid models (RF, Sequential Minimal Optimization (SOM), M5P) were chosen and assessed alongside the benchmark model Reduced Error Pruning Trees (REPT). The final results showed better accuracy, though not substantially better than the benchmark model. Nevertheless, despite the lack of notable enhancement in accuracy, they remained a more viable alternative compared to REPT.

Missing historical daily-averaged discharge records at the downstream station of the river have been successfully predicted using upstream water level records from multiple stations in the Mekong megadelta, as outlined in the study conducted by Thanh et al. [117]. A simple rating curve approach was compared with the results of various machine learning models, including DT, the ensemble method RF, kernel-based techniques such as SVR, Least Squares Support Vector Machine (LSSVM), and Gaussian Process Regression (GPR), as well as a statistical method, Multivariate Adaptive Regression Splines (MARS). Data preparation involved three stages: normalizing the data within a specified range, applying Fourier series fitting, and performing first-order differencing. The principle of input selection was based on the correlation coefficient and mutual information score. The MARS and RF models exhibited satisfactory and robust performance, even under

extreme conditions such as floods, droughts, and salinity intrusion.

The last reconstruction study discussed in the review paper by Mihel et al. [90] was published by Fei et al. [32]. In this study, the authors tested a different numerical model, the Hydrologic and Hydrodynamic Coupling Model (H2C), to further improve water level estimation, when coupled with machine learning methodologies (XGB and LSTM) in a hybrid architecture. The area of focus was the Tianhe-Zhuyin tidal reach, and the proposed model (H2C-XL) was compared with different approaches: a stand-alone numerical model, H2C, and hybrid variations combining numerical modeling with ML, such as the hydrologic-LSTM-hydrodynamic coupling model (HLHC) and the hydrologic-hydrodynamic-LSTM model (HHLC). The proposed model yielded better results than HLHC and demonstrated superior performance compared to the other two models. During flood periods, the most pronounced differences in model performance, as well as estimation errors, were observed.

The latest contribution category started with performance comparison of MLR, multiple non-linear regression, and ANN as the focus of a study by Lauer and Kösters [72], where they attempted to predict tidal extremes of Weser Estuary (minimum and maximum water levels for several stations) using tidal characteristics, discharge, and meteorological data. A variance inflation factor, station-wise linear Pearson correlation-based calculations, and a supervised Rank-based relief attribute calculations were employed to identify and select input features that are informative and non-redundant for the regression analysis. The combination of ANN and BO yielded greater accuracy relative to the other two models; nonetheless, the performance for seaward stations surpassed that of landward stations across all evaluated models. The previously discussed study by Dato et al. [30] also performed a hindcast analysis of water levels using an MLFF for the same tidal reach, as the initial step of its framework. The MLFF was trained on reanalysis data (using the same input features as in forecasting, but for a different time range). Four models were tested at two locations: Reanalysis Unweighted Buenos Aires, Reanalysis Weighted Buenos Aires, Reanalysis Unweighted Oyarvide, and Reanalysis Weighted Oyarvide. Unlike the forecasting scenario, where the results begin to diverge after a lead time of 24 hours, the results of hindcasting converge. A key limitation of hindcasting scenarios is that they cannot capture potential forecasts. However, the models achieved high accuracy in the reconstruction of both moderate and extreme events, with a delay

of less than one hour.

Li et al. [75] proposed a Deep Characteristic Learning (DCL) framework that enabled real-time flow monitoring using H-ADCP measurements and river cross-section characteristics for a location in the Shenzen River, under the influence of tides and backwater. The proposed framework enabled dimensionality reduction through Principal Component Analysis (PCA), adaptive ML algorithm selection (BP, Elman, RBF, Generalized Regression Neural Network (GRNN), and SVM), and hyperparameter optimization using a Genetic Algorithm Particle Swarm Optimization (GA-PSO). Frameworks' performance was compared to the stand-alone models included in it and traditional MLR models. DCL considerably outperformed other models in simulating discharge and was the only model to provide the closest discharge estimation of its minimal and maximal values.

Author and publication year	Simple statistical approach	Classifiers, kernel and ensemble approach	SNN approach	RNN approach	Hybrid approach
Adib (2008) [1]			✓		
Wei and Hsu (2008) [131]			✓		
Chinh et al. (2009) [24]			✓		
Chen et al. (2012) [19]			✓		
Pierini et al. (2013) [97]			✓		
Liu and Chung (2014) [78]			✓		✓
Gu et al. (2014) [44]					✓
Hidayat et al. (2014) [55]			✓		✓
Garel and Dalimonte (2017) [39]			✓		
Bhar and Bakshi (2020) [12]			✓		
Guillou and Chapalain (2021) [45]	✓		✓		
Gan et al. (2021) [37]			✓		
Sampurno et al. (2022) [104]		✓			✓
Thanh Hoan et al. (2022) [118]		✓			✓
Thanh et al. (2022) [117]	✓	✓			
Fei et al. (2023) [32]					✓
Lauer and Kösters (2024) [72]	✓		✓		
Dato et al. (2024) [30]			✓		
Li et al. (2025) [75]	✓	✓	✓		✓

Table 4.3: Reconstruction study categorization

4.4. Stage-Discharge Relationship Studies

Although the stage-discharge relationship is presented as a distinct category of analysis in tidal reaches, it closely aligns with the reconstruction problem, differing primarily in the input features considered. Stage-discharge analysis category is focused on studies

that mainly resolve the problem of discharge estimation through water level data, with exceptions of the opposite scenarios. The main difference between these two categories lies in the input data. The reconstruction problem leverages data from various stations, whereas the stage-discharge analysis is limited to data from the specific station of interest. Subsequently, examples of several prior studies can be compared to this category to some extent, those include studies by Hidayat et al. [55] and Thanh et al. [117]. While there is an abundance of research on the stage-discharge relationship within inland regions, as noted by various authors such as Bhattacharya and Solomatine [13] and Ajmera and Goya [4], there are only a limited number of studies focusing on tidal rivers and estuaries (see Appendix B.3, Table B.5). A categorization these stage-discharge studies is presented in Table 4.4.

Modeling a stage-discharge relationship is a relatively straightforward process, as previously mentioned, where predetermined rating curves are employed due to their simplicity and almost nonexistent cost. Under the almost steady flow conditions typical of inland rivers, the relationship between water level and discharge is relatively straightforward. In contrast, this relationship becomes considerably more complex in tidal rivers and estuaries, primarily due to tidal influences, but also due to other physical processes. Therefore, modeling discharge cannot depend solely on water level data from the exact location, but also requires tidal levels to be considered [50].

The initial proposition for establishing a stage-discharge relationship using an ANN in a tidally influenced Chao Phraya River, was given by Supharatid [114] in 2003. This study represented a unique example within tidal reaches, where the commonly applied stage-discharge relationship used to determine discharge was applied in the opposite manner, with the goal of predicting water level. The input features pertained to discharge and tidal data to predict the water stage using an MLFF. A simple stage-discharge relationship derived from using MLR and Multiple Power Regression (MPR) could not effectively capture the complex physical behavior related to the opposite influence of discharge and tides, which means that during periods of minimal discharge, the impact of tides becomes more pronounced. Three years later, another study emerged on establishing a stage-discharge relationship, but with a focus on a low-gradient tidal stream known as Isaac-Verot Coulee, where a more challenging scenario was observed with multiple rating curve loops and also complex channel geometry. Hence, Habib and Meselhe [50] had proposed

using two data-driven models, namely, an MLFF and a Nonparametric Local Regression (LOESS), for modeling such a complex relationship. It was found that including sufficient data from multiple stations had more impact on the models' performance than increasing their complexity. A greater emphasis was placed on ANN due to its superior ability to generalize, particularly in the context of extremely high discharge events. Nonetheless, both methodologies successfully replicated such a complex rating curve.

After another research gap, the stage-discharge topic resurfaced with a contribution from Wolfs and Willems [132], who explored the area of two rivers, Dender and Marke, and modeled discharge with both rating curve and ML approaches. Unlike previous studies, the area of interest was influenced by hysteresis. From the rating curve perspective, the Simple Rating Curve (SRC) and State-Dependent Parameter-Rating Curve (SDP-RC) were selected, while from the ML perspective, M5 model trees and MLP were used. Considering performance, complexity, and transparency, the recommended solution was SDP-RC, although M5 slightly outperformed it in terms of accuracy.

Author and publication year	Simple statistical approach	Classifiers, kernel and ensemble approach	SNN approach	RNN approach	Hybrid approach
Supharatid (2003) [114]			✓		
Habib and Meselhe (2006) [50]	✓		✓		
Hidayat et al. (2014) [55]			✓		✓
Wolfs and Willems (2014) [132]	✓		✓		
Thanh et al. (2022) [117]	✓	✓			

Table 4.4: Stage-discharge study categorization

4.5. Critical Analysis

Several challenges were identified in the literature that can generally be classified into two categories: one relating to the evolution of machine learning applications for the area of tidal rivers and estuaries, which refers to research gaps, and the other focusing on the advantages and limitations of the presented machine learning categories.

4.5.1. Research Gaps

The evolution of machine learning applications in tidal rivers and estuaries has encountered several impediments, such as (1) slow gain in popularity, (2) gaps in the frequency

of published studies, and (3) a limited number of available stage-discharge studies [90].

Insight into the slow growth in popularity until the 2000s, can be attributed to several factors. The most important factor relates to the limitations of the data available at the time and their quality. It was common until 2000 to manually collect hydrological data, which was time-consuming and problematic in areas with limited access. Likewise, discharge was previously estimated mainly using rating curves, which were the most cost-effective solution at that time. However, this led to inaccuracies as the river dynamics in tidal rivers and estuaries cannot be characterized as steady. Another aspect related to data refers to its temporal resolution, which even in current time represents a limitation; for example, available review papers [61, 135, 142] describe papers that use daily and longer temporal resolution, in dominant percentage (around 90%), while the remaining use hourly scale, which is required for capturing the unsteady dynamic of tidally influenced areas. Another reason was the limited computational power, which considerably impacted the use of machine learning approaches, especially until 2000; hence, simple statistical approaches were more frequently employed due to the mentioned. Based on the WoS search, 10 times the difference [90]. Another aspect highlighted the multidisciplinarity of the problem of interest and mistrust against the black-box methodologies, which raised questions about their reliability, as well as interpretability. This is one of the reasons why hydrological models and simple statistical methods were considered a more acceptable option.

The problem of gaps in the frequency of published studies over extended periods, for example, the absence of forecast analyses between 2003 and 2012, and of reconstruction studies before 2008, was also observed in previously published review papers dealing with general research problems in the field of inland rivers. Although such papers are more frequent, primarily due to the lower complexity of river dynamics, the problem of a lack of research remains pronounced. Consistent findings on publication frequency further corroborate this conclusion. Among the two review papers on water level analysis published in 2020, one references no more than four studies published prior to 2008 [142], whereas the other cites only five papers from the period between 2000 and 2012 [143]. A review paper [128] published the following year, 2021, also focused on water level parameter and exhibited a similar citation pattern, referencing a total of three studies published before 2008 and nine studies from 2003 to 2012 [128].

In addition to the quantitative representation of the studies, the problem also referred to the time scales used. Namely, most of the papers cited in the reviews use either daily or monthly resolution, which is not relevant for coastal areas, especially when dealing with semidiurnal oscillations. Hence, the identified gaps arise from challenges related to computing resources, data accessibility, and the challenges faced by simpler machine learning models in accurately representing complex tidal dynamics. The gap regarding the reconstruction studies prior to 2008 can be supported by a study [41], who highlighted that earlier research often overlooked missing values instead of addressing their reconstruction. This practice lasted until the emergence of more sophisticated machine learning techniques.

The remaining challenge referred to the limited number of studies (only five papers) that specifically deal with the problem of establishing the relationship between water level and discharge in tidal areas. This gap can be mainly attributed to complex flow regimes [90], which can include additional challenges such as reversible flow, tidal asymmetry, and related phenomena. Despite the growing trend of applying machine learning models to hydrological problems, challenges remain in gaining insight into the underlying physical processes, unlike numerical models, and also in unreliability during extrapolation beyond the training set [90]. Hence, the problem of establishing the relationship between water level and discharge still remains an under-researched problem.

The review paper also highlighted a problem of regional imbalance in the literature on tidal rivers and estuaries, as the majority of studies (37 out of 47) focus on Asian rivers, such as the Tanshui River in Taiwan, as well as the Yangtze and Chao Phraya Rivers. In contrast, relatively few studies have investigated tidal reaches in Europe and North and South America (11 out of 47), such as the Loire, Severn, and Columbia Rivers. Several factors contribute to this disparity, resulting in more intensive research in certain areas, such as those with higher population density and greater flood risks. The under-representation of European estuary systems further motivated the conducted research. Specifically, the Neretva River remains scarcely investigated in terms of water levels and discharges, despite its considerable socio-economic, as well as ecological importance to the surrounding area.

4.5.2. Strengths and Weaknesses of ML Categories

When discussing the advantages and limitations of categories of machine learning models, it is important to consider their ability to handle nonlinear and non-stationary data while providing computational efficiency and interoperability. This subsection provides a discussion on each type of ML approaches from several important perspectives, which is based on findings presented in a study by Mihel et al. [90]. Figure 4.1 illustrates the studies classified by the machine learning approach employed. The conclusion gained from the given visualization was that the most prevalent categories (around 50% of the reviewed studies) were SNNs and the hybrid approach. SNNs, due to their ability to model nonlinear relationships even without prior knowledge of physical processes in tidal environments, and hybrid models, as they can combine the advantages of different approaches and techniques (statistical methods, signal preprocessing, optimization algorithms, fuzzy clustering, ML architectures, numerical models) to achieve an increase in performance. The reason for the low representation of the other categories lies in the fact that the first and second categories of models are in most cases used as baselines, i.e. reference models for performance comparison, while the category of advanced RNNs, for example, has only since 2018 become popular for application in areas of tidal rivers and estuaries, which is followed by a small but steady increase.
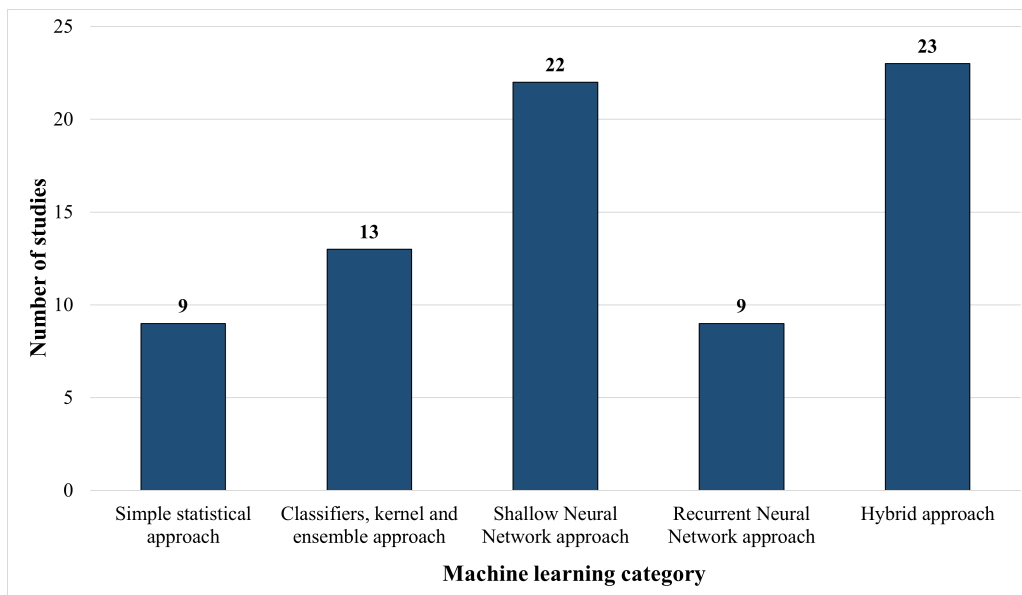
Figure 4.1: Classification of published studies by machine learning approach

The category of simple statistical approaches was used in only 20% of studies and

was rarely employed as a stand-alone approach (only two available studies), meaning this category of models was commonly used as a baseline for comparison with more advanced approaches and was even used as a component for hybrid modeling, as shown in [20]. By reviewing all published studies, several insights were found regarding the applied methods: (1) models based on linear assumptions (AR and MLR) are not efficient for capturing non-linearities, (2) those which are capable of capturing nonlinearities to some degree (MPR and BRR), (3) and other solutions which are more adaptable to nonlinearities but often struggle with either overfitting, generalization or are complex to interpret (LOESS and MARS). While the second category does not face the same challenge of managing non-linear data, it is more limited in terms of non-stationarity, as seen with DT and SVR in reconstruction or KNN in forecasting studies. Likewise, the ability to handle missing data can only be attributed to employed classifiers (DT) and ensembles (RF), not kernel methods. In terms of interpretability, classifiers typically feature the most simple architecture, placing them at the very top, followed by ensemble methods and then kernel approaches. However, when the performance of all these categories was compared to either a stage-discharge rating curve or a hydraulic analysis, which are recognized for their sensitivity to outliers and noise, considerable improvements were observed [96, 117]. Rating curve and HA necessitate recalibrations in response to geomorphic changes, whereas ML approaches do not incorporate such input parameters. Recalibration or retraining typically occurs only in extreme scenarios or when there are considerable deviations.

SNNs have often been shown to provide better results than the previous category, as they introduce nonlinearity through their activation functions and by using different algorithms. The most popular was the Levenberg-Marquardt, which enabled improved speed and efficiency. In the majority of reviewed articles, SNNs showed better generalization ability than hydrodynamic models, especially in cases of extreme events (peaks and typhoons). However, although this category of machine learning represents a step forward compared to the previous one, it still contains certain limitations that negatively affect the performance. This includes sensitivity when it comes to spatial variability, but also the size of the data sets. In addition, the training data set needs to be sufficiently representative of all potential scenarios, i.e., encompassing enough variation. Moreover, regarding forecasting, a pronounced decline in performance has been observed as the lead time increases.

The next category, representing mainly advanced RNNs, such as GRU and LSTM, has only recently become popular, more precisely in the last seven years. Unlike the previous two, this category's key advantage is the possibility of capturing long-term dependencies through gating mechanisms and memory cells. Hence, the problem with the vanishing gradient is adequately addressed and does not pose a major challenge like for ANNs, thereby enabling more accurate predictions. Because of their ability to learn from historical data, they can effectively identify complex relationships between discharge, tidal levels, and meteorological factors in tidal reaches, which is of crucial importance. Nevertheless, advanced RNNs are more computationally intensive than previous categories and, therefore, require more time and computer memory for model training, which can be even further extended if a larger grid of hyperparameters must be tuned. Additionally, the possibility of interpreting the gained result becomes challenging, which cannot be said for previous categories. Only one study pointed out the better results of advanced RNNs compared to a hybrid approach [23], which utilized an LSTM Seq2Seq model.

The last category of the machine learning approaches encompasses hybrid methodologies, with the initial article published during the period of pioneering work in 2003, followed by a continuous and steady publication rate during the other two periods, with approximately seven articles per period. Several studies have shown the potential of hybrid methods over stand-alone, and emphasized the improvement in performance [3, 20, 104, 134]. Several studies approached the hybrid modeling by integrating different numerical models, for example, NS_TIDE [20, 140] in forecasting analysis, or SLIM [104] and H2C [32] in reconstruction analysis. In all studies, the hybrid approach provided more accurate results than the stand-alone numerical model. This can be explained through a single key aspect. Although the numerical models can adequately represent flow conditions in tidal reaches, using machine learning models alongside them is additionally beneficial, as they are better suited for finding complex and nonlinear relationships within the data. This was highlighted in a study by Fei et al. [32], where XGB was employed to assess feature importance, and it was found that the inclusion of discharge data has a considerable impact on water level forecasting.

Signal preprocessing techniques have been applied in five hybrid-based studies, where three employ the Wavelet Transform and the remaining two use the EMD for capturing variations across daily, seasonal, and tidal scales. Although these techniques offer multi-

scale analysis, they differ in the same aspects. While WT uses already established wavelet functions that help regarding the presence of noise, more precisely, its isolation, the EMD is limited when presented with very noisy signals. As EMD directly decomposes the signal into various IMFs, it is often related to the problem of mode mixing [22]. However, this problem is partially resolved by incorporating white noise into an improvement variant of the same technique, EEMD, as in the study by Chen et al. [22]. All presented studies have shown considerable performance improvements, especially in facilitating longer horizons with acceptable accuracy than stand-alone ML approaches (SVM in [129]; LSTM, GRU, and LSTM-CNN in [38]). These approaches also capture rapid changes in processed signals, which is crucial for recognizing extreme occurrences quickly and reducing their effects. While EMD does not require parameter adjustments, WT and EEMD do. In terms of computational complexity, the methods may differ, but the complexity increases with the length of the time-series and the higher the temporal resolution. Therefore, WT and EMD are less demanding, while EEMD is the most demanding because it performs additional iterations with white noise.

A different type of hybrid models relates to various combinations of different architectures, statistical methods, fuzzy clustering, and optimization algorithms, ranging from early studies that integrated unsupervised and supervised learning approaches [15] to the latest optimized parallel computing frameworks [75]. A total of seven studies have been performed, with only one of them specifically addressing the discharge parameter. Chang and Chen [15] enabled faster model training by employing a single-layer neural network in its architecture, than a standard MLP with backpropagation. The proposed architecture gained accurate and reliable water level predictions even under extreme conditions. A combination of an ensemble CART and an RBF has demonstrated superior performance under tidal and typhoon conditions in Tsai et al. [121], surpassing the previously discussed RBFNN, which also yielded more accurate results than stand-alone ANN or CART in this study. RBF not only enhances model training time but also progressively improves accuracy, influenced by the specific combinations of architectures utilized. Similarly, ensembles, when utilized in conjunction with neural networks, enhance their generalization capability to a certain extent, as it is typically constrained by the varying representations of the data range. A combination of a statistical method (ANOVA) as kernel of an SVR gave similar results to its stand-alone version with an RBF kernel in

[3]. However, RBF was better for longer prediction horizons due to its noise resistance ability, as it focuses more on global patterns. At the same time, ANOVA was more adequate and precise for capturing short-term variations. Another study conducted by Thanh Hoan et al. [118] examined bagging-based hybrid machine learning methods, which have demonstrated superior accuracy for tidal reaches compared to stand-alone ML models. Parallel computing has only recently been introduced for tidal environments by Li et al. [75]. This approach not only improves adaptability through self-learning but also ensures that the size of the datasets does not limit model performance. Additionally, through dimensionality reduction, even the computational complexity becomes less of a disadvantage. Likewise, spatiotemporal feature extraction had been tested in two recent studies [109, 139]. The Cheb-GRU model was also found to be transferable based on the findings of Shi et al. [109] to other tidal areas. Its performance was strongly linked to the length of the dataset, i.e., a larger dataset has higher accuracy, which is consistent with the data and results from the other study.

4.5.3. Justification of Research Direction

Guided by the conclusions of previous studies, and the advantages and limitations of model categories, we focus primarily on hybrid modeling. In addition, with the emergence of advanced recurrent neural networks for hydrological modeling in the last five years in the tidal reaches, this study has tested the applicability of such approaches, either as stand-alone or as a part of a hybrid modeling scenario. Two research propositions were presented, one in the form of a review article by Mihel et al. [90], and another in an article by Shi et al. [109]. Both were related to the use of an attention mechanism, which until then had not yet been applied in tidal reaches, let alone under microtidal conditions.

At the time the review paper was published, the use of the attention mechanism in hydrology was notably limited, with only ten papers addressing the topic. Nevertheless, the trend of implementing the aforementioned mechanism has seen considerable growth over the past two years, with five studies currently available on the application of the attention mechanism in the tidal reaches [16, 54, 62, 83, 127]. Among these five studies, one published by Weber de Melo et al. [127] stands out as the most closely related to our problem, focusing on the application of the attention mechanism to develop an early

warning system for potential threats.

The research presented in the doctoral thesis has been published in several studies. The first paper represents the foundation for the given literature review [90], which had been broadened and included publications for the last two years. Based on the conclusions obtained from the literature review, three papers were been published. The first paper proposed a hybrid machine learning approach that integrates an attention mechanism combined with LSTM for the problem of discharge estimation [88]. The performance of this model had been compared to stand-alone models, including LSTM and simple non-temporal machine learning approaches, and evaluated on two dataset, one using measured data, and the other obtained through conducting a numerical simulation. The subsequent two conference publications [89, 91] focused on the CNN-LSTM hybrid model for the prediction of water level and discharge. A different dataset was used, also generated through a numerical simulation. Although CNN had previously been applied in three studies [38, 109, 139], its applicability showed potential for further exploration. Furthermore, an analysis employing signal preprocessing techniques was performed on the best models identified in the first study to enhance estimation accuracy. All of our aforementioned research was focused on the microtidal reach of the Neretva River. Details of the conducted studies are presented in subsequent chapters.

However, not only are the machine learning categories of importance, but the selection of evaluation metrics must also be discussed. Hence, based on the visualization presented
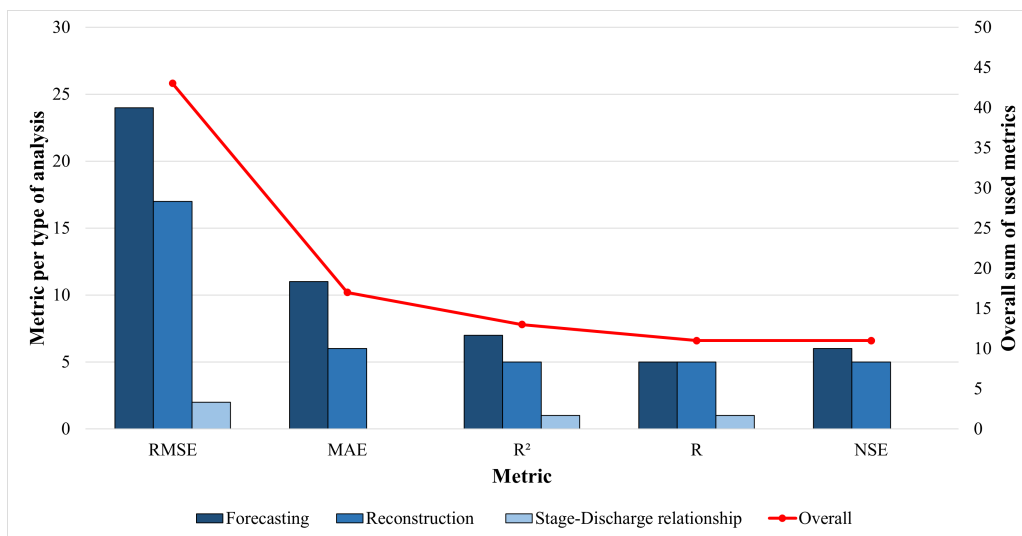
Figure 4.2: Most frequent evaluation metrics based on the reviewed studies

in Figure 4.2, the metrics with the highest applicability rate of all studies reviewed, considering all types of analysis, were considered.

After reviewing the relevant literature and gaining insights into the employed methodology, the next chapter provides an overview of the methodology, detailing the entire process from data preparation to model development.

Chapter 5

PROPOSED METHODOLOGY

Contents

This chapter outlines the techniques and methods employed for the purpose of discharge estimation in a Neretva tidal river, using the machine learning models discussed in Chapter 2 and the hybrid models presented in this chapter. Several phases are considered for enabling a systematic research process: data preparation, signal preprocessing and augmentation, model development consisting of training, optimization, and evaluation, and lastly, providing explainable and transparent models - with the aim of ensuring

reproducibility. During the data preparation phase, the collection and correction steps are addressed, and for signal preprocessing, decomposition and augmentation techniques are applied. After ensuring that the data has been adequately handled and prepared, the model development specifics regarding the model training and the optimization of hyperparameters are given. Following the model development, evaluation metrics and statistical significance tests are explained, emphasizing their previous application in the field of hydrology. The final phase of the methodology explores techniques that provide interpretability and transparency of the final models' results.

5.1. Data Preparation

Data preparation for training and testing the selected machine learning models was carried out in several steps. The first step was to examine whether a time lag exists between the input parameters of the hydrological stations' water levels and the discharge output variable at a single location. The inspection was carried out by employing a cross-correlation analysis, whose results are shown in Figures 5.1 and 5.2. Through visual inspection, it was concluded that a two-hour lag exists in both datasets (measured and simulated). The observed 2-hour lag between water levels and discharges at the same and downstream stations along the Neretva River estuary can be explained by tidal river dynamics, where frictional damping, channel convergence, and the interaction of tidal and riverine flows cause flow velocity (and thus discharge) to respond out of phase with tidal water level oscillations, typically resulting in phase lags of 1-3 hours [58]. As the downstream stations, Opuzen and Kula Norinska are located approximately 11 and 16 km from the river mouth, their proximity implies that the propagation time is less than 1 hour. Hence, the following step involved adjusting both time-series accordingly to resolve the time lag.

The second step of data preparation was to apply a splitting method, which would create a training dataset used by the model to learn patterns in the dataset, and the testing dataset which would be employed for evaluating the models' performance on new unseen data. As the range of our dataset spans from 2016 to 2021, and there were around 50,000 records, it was decided to employ a splitting approach where 80% of the data represented the training dataset, while the remaining 20% was the testing dataset. More
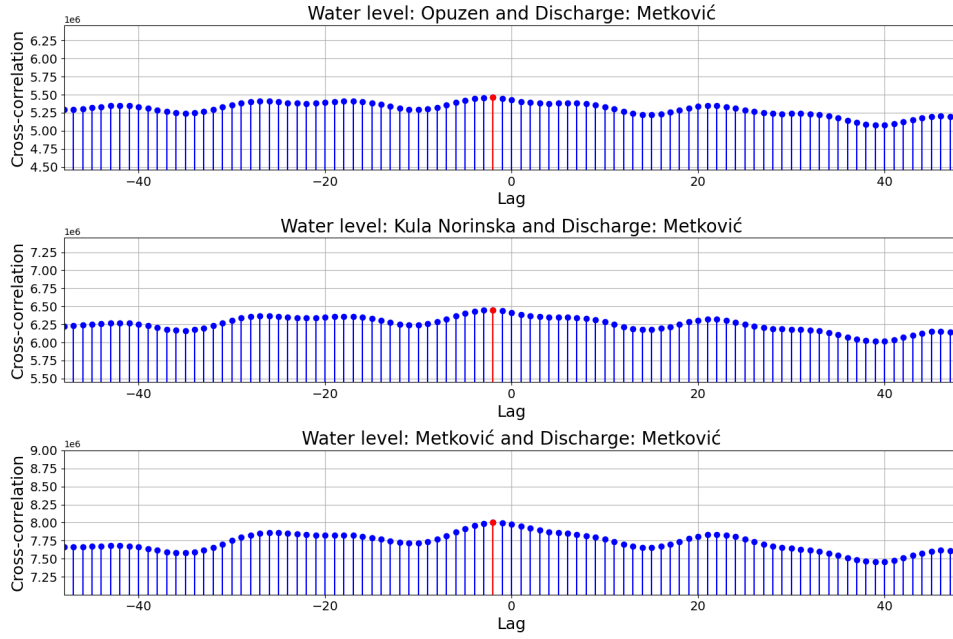
Figure 5.1: Two-hour time lag between multiple water level stations and a single discharge of the measured dataset [88]
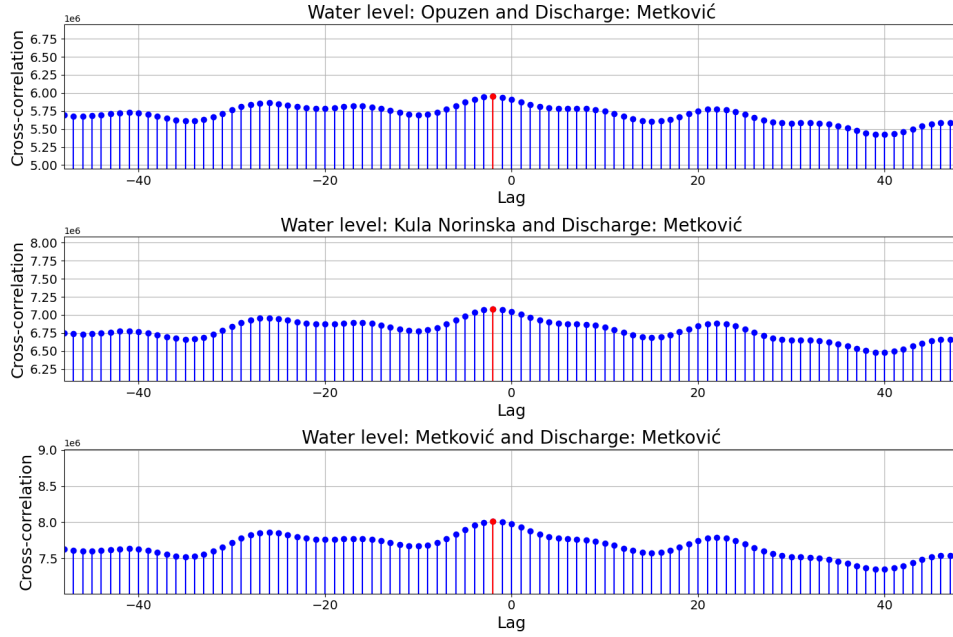
Figure 5.2: Two-hour time lag of the simulated dataset [88]

precisely, the training dataset contained records from January 2016 up to October 2020, while the testing from November 2020 up to December 2021. The remaining HEC-RAS dataset records ranged from 2016 to 2019, totaling approximately 30,000 records. The training data spanned from January 2016 to February 2019, while the testing data covered the period from March 2019 to December 2019.

The next step in data preparation was scaling. The rationale behind using the scaling method was the substantial difference in value ranges between water levels and discharge records. The scaling was performed by using *MinMaxScaler* from *scikit-learn* Python library developed for machine learning models. *MinMaxScaler* resizes the data to a pre-defined range of [0, 1], while maintaining the original data distribution. The mathematical expression for the *MinMaxScaler* approach is [129]:

$$\hat{x}_i = \frac{x_i - x_{min}}{x_{max} - x_{min}} \tag{5.1}$$

where x_i is the current observation, x_{min} the overall minimal value of a variable, x_{max} the overall maximal value of a variable, and $\hat{x}_i$ the normalized value of x_i after scaling.

This research investigates the predictive capability of both simple machine learning models (such as DT, RF, SVR, LGBM, and XGB) and time-series models (like LSTM and LSTM-Attention). However, these models differ in terms of the importance of data order. Time-series machine learning models require data to be arranged in chronological order, as this reflects the nature of real-world data. In contrast, simple machine learning models do not always rely on sequential order and can be trained even with randomly shuffled data. Hence, defining the adequate structure of input data for models is of critical importance. A sliding window approach pertaining to a 24-hour window was considered for predicting the discharge of the preceding hour. A 24-hour window of historical data was selected as it contains a full tidal cycle, including both high and low tide periods, hence properly representing the complex dynamic of the Neretva River flow. The window length selection can be further supported through the Adriatic sea dynamic characteristics of the tidal regime, which is referred to as a mixed tidal signal [67]. In such a tidal regime, daily or diurnal and half-daily or semidiurnal constituents are equally present and strong, without one being more dominant. Therefore, capturing both types of oscillations requires a window length of at least 24 hours.

5.2. Signal Preprocessing and Augmentation

5.2.1. Spectrogram-Based Time-Frequency Representation

Using only the time domain may be a too restrictive approach for modeling non-stationary signals in tidal rivers and estuaries, such as water level and discharge. As these hydrological parameters are influenced by extreme conditions, seasonal variations, tidal cycles, and more, gaining simultaneous insights into the time and frequency domains may enhance the understanding of such signals. While the Fourier transform (FT) enables transforming a time-series into a frequency domain signal, and back, it still lacks the ability to provide information on the energy density of a signal in both the time and frequency domains at the same time. Therefore, a well-known time-frequency distribution, Short-term Fourier transform (STFT), proposed by Gabor et al. [34] in 1946, was selected to produce its magnitude display spectrogram. A mathematical expression for STFT is [107]:

$$STFT_x(t, f) = \int_{-\infty}^{+\infty} x(\tau) \, g(\tau - t) \, e^{-j2\pi f\tau} \, d\tau \tag{5.2}$$

where t represents time, f frequency, $x(\tau)$ signal in time domain, $g(t-\tau)$ window function centered around t, $e^{-j2\pi f\tau}$ complex exponential function.

Spectrogram is derived from the previous Equation 5.2, which represents fundamental STFT, by calculating its squared magnitude. Therefore, the complete expression for STFT is presented in Equation 5.3 [89]:

$$SPEC_x(t, f) = \left| \int_{-\infty}^{+\infty} x(\tau) \, g(\tau - t) \, e^{-j2\pi f\tau} \, d\tau \right|^2 \tag{5.3}$$

where t represents time, f frequency, $x(\tau)$ signal in time domain, $g(t-\tau)$ window function centered around t, $e^{-j2\pi f\tau}$ complex exponential function.

The selected window function for performing analysis was the Hamming window. The main advantage of the utilized windowing function lies in minimizing possible spectral leakage. Likewise, as a 24-hour sliding window was utilized for input data, the dimension of each generated spectrogram was 24x12, as shown in a single example in Figure 5.3.
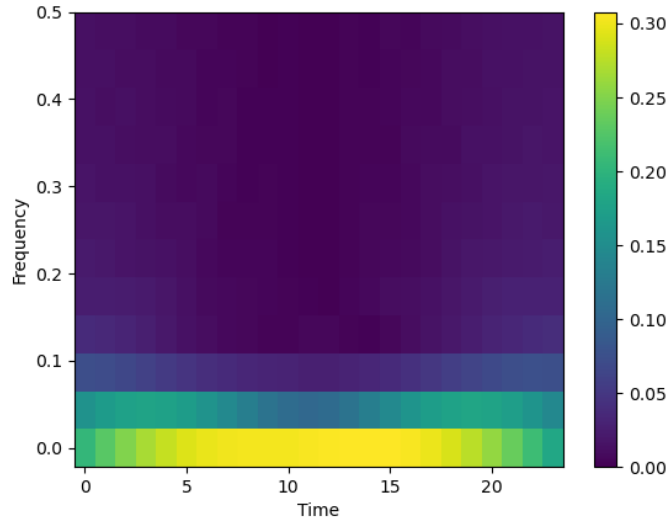
Figure 5.3: Example of a single water level spectrogram using STFT and Hamming window [89]

5.2.2. Variational Mode Decomposition

Variational Mode Decomposition (VMD), proposed by Dragomiretskiy and Zosso [31] in 2013, is a signal processing technique that decomposes the original input signal f into K sub-signals or modes, referred to as Intrinsic Mode Functions (IMFs) from EMD. VMD relies on three building blocks: (1) Wiener filtering, (2) Hilbert transform, and (3) frequency shifting. By combining these elements, VMD formulates a constrained variational problem, aiming to extract K band-limited modes in the following manner [31]:

$$\min_{\{u_k\},\{\omega_k\}} \left\{ \left\| \sum_k \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} \tag{5.4}$$

$$\text{s.t.} \quad \sum_k u_k(t) = f(t)$$

where k is the index of a mode, u_k is the kth mode function, $\{u_k\}$ represents set of all modes, ω_k is assigned central frequency of kth mode, $\{\omega_k\}$ is set of central frequencies of all modes, K is the total number of decomposed modes, $\delta(t)$ is Dirac distribution, $*$ is the convolution operator, t denotes time, j is the imaginary unit, $\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t)$ represent an expression of the Hilbert transform of $u_k(t)$ to an analytical signal, $e^{-j\omega_k t}$ enables frequency shifting to the baseband, and $f(t)$ is the original signal.

Hence, Dragomiretskiy and Zosso [31] decided to introduce the Lagrangian multiplier (α) and a quadratic penalty term, which converted the previously constrained variational problem to an unconstrained one. The quadratic penalty term primarily emphasizes the enhancement of reconstruction fidelity, whereas the Lagrangian multiplier is employed to enforce constraints rigorously [31]. Consequently, the formulation of the augmented Lagrangian multiplier can be expressed as [53]:

$$
\begin{aligned}
\mathcal{L}(\{u_k\}, \{\omega_k\}, \lambda) = \alpha \sum_k & \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \\
& + \left\| f(t) - \sum_k u_k(t) \right\|_2^2 + \left\langle \lambda(t), f(t) - \sum_k u_k(t) \right\rangle
\end{aligned}
\tag{5.5}
$$

where k is the index of a mode, u_k is the kth mode function, $\{u_k\}$ represents set of all modes, ω_k is assigned central frequency of kth mode, $\{\omega_k\}$ is set of central frequencies of all modes, K is the total number of decomposed modes, $\delta(t)$ is Dirac distribution, $*$ is the convolution operator, t denotes time, j is the imaginary unit, $f(t)$ represents original signal, $\lambda(t)$ is the Lagrange multiplier, α is the quadratic penalty factor, $\|f(t) - \sum_k u_k(t)\|_2^2$ represents a quadratic term, $\langle \lambda(t), f(t) - \sum_k u_k(t) \rangle$ is a term with Lagrangian multipliers, and $\mathcal{L}$ is the augmented Lagrangian.

The solution to the minimization problem outlined in Equation 5.4 involves performing a sequence of iterative sub-optimizations, using the alternate direction method of multipliers (ADMM). This process is repeated until convergence is achieved. Three steps are executed and repeated iteratively. These steps refer to mode update (see Equation 5.6), center frequency update (see Equation 5.7), and the Lagrangian multiplier λ (see Equation 5.8), presented in the given Algorithm 1. [31]

In the given Algorithm 1 k is the index of a mode, u_k is the k-th mode function, ω_k is k-th mode central frequency, λ Lagrangian multiplier, τ is step size parameter for the dual ascent update, n is ADMM iteration step, ϵ is convergence tolerance, t denotes time, f is the original signal in time domain, and $\mathcal{L}$ is the augmented Lagrangian.

Currently, no study has employed VMD in tidal rivers or estuaries for water level or discharge estimation or prediction. Only one paper by Chen et al. [21], was found through the WoS search. This study employed a combined approach of VMD with machine learning, focusing on ship traffic flow and congestion. An additional search identified four

---

Algorithm 1 ADMM optimization concept for VMD [31]

Initialize: $\{u_k^1\}$, $\{\omega_k^1\}$, λ^1, $n \leftarrow 0$
Repeat:
$n \leftarrow n + 1$
For $k = 1 : K$ **do**
Update u_k:

$$u_k^{n+1} \leftarrow \arg\min_{u_k} \mathcal{L}\Big(\{u_i^{n+1}\}_{i<k}, \{u_i^n\}_{i\geq k}, \{\omega_i^n\}, \lambda^n\Big) \tag{5.6}$$

End for
For $k = 1 : K$ **do**
Update ω_k:

$$\omega_k^{n+1} \leftarrow \arg\min_{\omega_k} \mathcal{L}\Big(\{u_i^{n+1}\}, \{\omega_i^{n+1}\}_{i<k}, \{\omega_i^n\}_{i\geq k}, \lambda^n\Big) \tag{5.7}$$

End for
Dual ascent:

$$\lambda^{n+1} \leftarrow \lambda^n + \tau\Big(f - \sum_k u_k^{n+1}\Big) \tag{5.8}$$

Until convergence:

$$\sum_k \frac{\|u_k^{n+1}\|_2^2 - \|u_k^n\|_2^2}{\|u_k^n\|_2^2} < \epsilon.$$

---

studies on the application of VMD in tidal rivers and estuaries; however, none of the studies integrated machine learning approaches.

5.3.   Hybrid Machine Learning Methods

5.3.1.   Convolutional Neural Network with Long Short-Term Memory

The utilized architecture for the forecasting task employs a hybrid approach combining a CNN and an LSTM. The justification for the application of this model refers to the use of spectrograms as input, independently and in combination with time-series. Such architecture had been previously applied for tidal reach, for example, [38, 109]. However, such an approach has not been applied to the area of microtidal river and discharge prediction, nor by integrating spectrograms. As the fundamental theory behind LSTM was previously discussed, the focus of this subsection is on CNN and its advantages when applied together with LSTM.

The first CNN architecture was proposed by LeCun et al. [73] in 2002, better known

as LeNet, applied for the handwritten digit recognition task. The multi-building blocks of CNN architecture refer to convolution, pooling, and fully connected layers, but also activation and loss functions.

A convolution layer can process data of different dimensionalities, ranging from 1D sequences and 2D matrices or images to 3D volumetric data and even 4D inputs such as video. For 1D input, such as time-series, it focuses on capturing local temporal patterns, while for 2D inputs, such as the spectrogram matrices used in this research, it operates similarly to image processing, extracting local features across both time and frequency dimensions. The convolutional layers are constructed of learnable filters, better known as kernels, and employ activation functions. By calculating the dot product between the inputs and kernels, a feature map is generated. Given the complex nature of tidal reach data, introducing nonlinearity is essential; therefore, the ReLu activation function is applied to each convolutional layer. The pooling layer performs sub-sampling of the feature maps generated by previous convolutional layers [6]. There are various types of pooling layers, such as min, max, average, and others. For the purpose of this research, a single Max Pooling layer was used. The final layer of a typical CNN architecture is a fully connected layer, which generates a vector representing the final regression predictions, i.e., continuous numerical values.

The selected architecture for water level and discharge forecasting consisted of two convolutional layers, a single Max Pooling layer, followed by LSTM and dropout, to further avoid overfitting, and lastly, a fully connected layer [89]. The first study on employing a CNN-LSTM for tidal reaches as a baseline model, integrated available spatial features, thus a more complex Chebnet model demonstrated significant advantages over CNN-LSTM [139]. In the absence of spatial features for the current research, CNN-LSTM serves as a suitable alternative, effectively extracting relevant local time-frequency and temporal patterns from the data. A holistic data analysis approach was achieved through this methodology.

5.3.2. Proposed Hybrid Model: Long Short-Term Memory with Attention Mechanism

The proposed architecture for this study is a hybrid modelling approach, which combines the previously discussed LSTM approach with an attention mechanism. The proposed model is referred to as the LSTM-Attention model. While the LSTM approach has been utilized in several studies concerning tidal rivers, particularly for water level estimation problems [17, 64, 126, 136, 139], the application of the attention mechanism has not yet been explored for this specific problem in the field of hydrology.

The attention mechanism is a technique specifically developed for processing sequential data. It is utilized as a part of the hybrid modelling approaches, for example, sequence-based architectures like Transformers or Encoder-Decoder models. Although the attention mechanism has been widely utilized in Natural Language Processing, its application in hydrology remains limited. This is further corroborated by a recent study by Mihel et al. [90], which found that only ten articles in the WoS database used the attention mechanism as part of a hybrid approach. LSTM offers advantages such as capturing long-term dependencies and temporal patterns within given sequential data while also, by including the attention mechanism, the model can further dive into the given data and identify the most relevant parts of the given input sequence. The attention mechanism serves as a means of enabling the model to selectively focus only on information that is considered important. Identifying relevant information can be rendered by dynamically assigning weights to different input parts [62].

Previously, we stated that the proposed hybrid architecture consists of two main types of layers: LSTM and the attention mechanism. Although LSTM layers generate hidden and cell states at each time step, only the hidden states are propagated to the attention mechanism (see Equation 2.15), whose role is to assign different weights to them. As previously discussed, these weights reflect the relevance of different sequence parts for the prediction assignment.

The attention mechanism consists of several steps. In the first step, the generated hidden states (h_t) of the LSTM layer are forwarded to the attention mechanism, after which a standard linear transformation is performed by multiplying them with a weight matrix (W_l). The multiplication result represents the hidden attention space to which the

LSTM hidden states are mapped to calculate the attention scores. Due to the complexity of the problem, the tangent hyperbolic activation function (tanh) is employed to introduce the nonlinearities into the model and constrain the multiplication result into a range between [-1, 1] (see Equation 5.9). To enable modelling complex, non-linear relationships and intricate patterns, which are commonly encountered in areas of interest, i.e., tidal reaches, where the interaction between tides, water level, and discharge can be described as highly complex, a *tanh* function was selected, as in [83]. The next step involves another linear transformation, which takes the previously generated vector u_t, which represents the transformed hidden state, and multiplies it with a transposed learnable weight vector v^T. This multiplication results in a scalar attention score for every timestep t (see Equation 5.10).

$$u_t = \tanh(W_l h_t + b_1) \tag{5.9}$$

$$e_t = v^T u_t \tag{5.10}$$

The second activation function, softmax, performs normalization of the given scalar attention scores (e_t) into a range of [0, 1], resulting in a probability distribution where the sum of all normalized attention scores equals 1 (see Equation 5.11). In this context, features associated with higher attention weights are considered more relevant to the final prediction than those with lower weights.

$$\alpha_t = \frac{\exp e_t}{\sum_{k=1}^{T} \exp(e_k)} \tag{5.11}$$

The normalized attention scores (α_t) are subsequently multiplied by the corresponding hidden states at each timestep (t) and then summed to produce a weighted sum of all timesteps, commonly referred to as a contextual vector (c) (see Equation 5.12). The weighted sum can be described as a comprehensive summary of information derived from the given sequence. The output of the attention mechanism, that is, the contextual vector, is forwarded to a fully connected layer, where it is multiplied with a learned weight matrix (W_{out}), which transforms the output in the adequate dimension and also produces the final prediction of the model (see Equation 5.13). The bias term (b_{out}) is also added as it helps the model better adjust to data while providing flexibility in modelling.

$$c = \sum_{t=1}^{T} \alpha_t h_t \tag{5.12}$$

$$\hat{y} = W_{out}c + b_{out} \tag{5.13}$$

5.4. Proposed Novel Model Development

5.4.1. Model Training and Optimization

Model training and validation phases were combined through a technique known as k-fold grid search cross-validation. The selected 80% of the training dataset were divided into folds, and all folds except one were utilized for model training. At the same time, the excluded fold served for validation purposes, i.e., finding the optimal hyperparameters of each model from the predefined range by testing all possible combinations. Hence, the final result of such a brute force paradigm gave us the best-fitted set of hyperparameters. The number of folds (k) was selected based on the provided amount of data and the adequate tradeoff between bias and variance.

Taking into consideration the dataset size and the tradeoff between bias and variance, and with an additional aim of balancing computational demands, a five-fold split was selected. Through each step of the performed cross-validation, four folds are used to train the model ($k-1$), while only one is used to evaluate the model performance. This process is performed for every possible combination of discrete hyperparameter values through grid search. When the best set of hyperparameters is found, the final result represents the average performance score obtained across all folds during the cross-validation. Hence, the main evaluation criteria used for refitting the models was the MSE.

However, performing k-fold cross-validation and grid search for *PyTorch* models is not directly possible, as is the case with *scikit-learn* library models, where these approaches are integrated into the mentioned framework. Therefore, using a library *skorch*, which acts as an intermediary between those libraries, was mandated. The compatibility with *scikit-learn* framework that enables the use of k-fold cross-validation and grid search in a standardized manner is enabled through a *NeuralNetRegressor* wrapper. The main advantage of such an approach is the ability to utilize already existing functions for either

optimization, evaluation, or others without needing to develop new methods.

PyTorch models require a few additional parameters to be defined before proceeding to the training process. These relate to selecting an optimization algorithm, a regularization technique, and a number of epochs. The literature overview [88] analysis revealed that Adaptive Moment Estimation (Adam) was the predominant optimizer for minimizing neural network loss function. Likewise, its advantages are further confirmed in a review study by Ahmed et al. [2], where it is used specifically in combination with the LSTM approach. However, for our problem, a slightly different Adam optimization variant was employed, namely, Nesterov-accelerated Adaptive Moment Estimation (Nadam). This optimizer integrates the favourable characteristics of the Adam optimizer with Nesterov momentum, aiming for accelerated convergence [125]. The selection of regularization methods and the number of epochs are intertwined. The early stopping strategy showcases the application of regularization, wherein model training ceases when the validation loss, i.e., the MSE, fails to decrease after a specified number of epochs. Such a strategy prevents the possibility of overfitting, minimizes training time, and is resource-efficient, even if the defined number of epochs is large. Employing an early stopping technique, especially during cross-validation, is highly useful to avoid and even prevent improper predictions [5]. The chosen step for early stopping was 15, and the total number of epochs was set to 500.

5.4.2. Evaluation Metrics and Statistical Significance Tests

In the evaluation phase, several performance metrics are considered to find the optimal hyperparameters of each model and validate the model performance on new unseen data. Those metrics are MSE, utilized as the main evaluation metric for finding the optimal model during the k-fold grid search cross-validation, followed by RMSE, MAE, NSE, and lastly, R. The aim is to obtain the lowest MSE, RMSE, and MAE, with the highest NSE and R value. While, for example, MSE and RMSE evaluate larger errors more strictly or give them greater weight, MAE provides an average error and, at the same time, better stability. Also, RMSE and MAE provide a calculation of the prediction error in measuring units of the output data. In parallel, two remaining metrics are applied, the aim of which is to assess the efficiency in terms of precision of the model through the NSE metric,

and to find the correlation between the actual and predicted data, which refers to the R metric. An additional metric had been employed for the forecasting, mean absolute percentage error (MAPE). The evaluation metrics mentioned above were selected to fulfil two key criteria, comprehensive model assessment and aligning with the standards for hydrological modelling, as outlined by Gupta et al. [49]. The expressions for evaluation metrics mentioned above are [88, 89]:

$$MSE = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n} \tag{5.14}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \tag{5.15}$$

$$MAE = \frac{\sum_{i=1}^{n}|y_i - \hat{y}_i|}{n} \tag{5.16}$$

$$NSE = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} \tag{5.17}$$

$$R = \frac{\sum_{i=1}^{n}(y_i - \overline{y})(\hat{y}_i - \overline{\hat{y}_i})}{\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2 \sum_{i=1}^{n}(\hat{y}_i - \overline{\hat{y}_i})^2}} \tag{5.18}$$

$$MAPE = \frac{\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \cdot 100}{n}, \tag{5.19}$$

where n is the number of observations, y_i is the observed value of i-th sample, and $\hat{y}_i$ is the predicted value of i-th sample, and $\bar{y}_i$ is the mean of the observed i-th sample.

To determine the optimal VMD parameters for decomposing signals, two metrics were selected: orthogonality index (OI) and reconstruction error ratio. Decomposition often suffers from problems such as mode mixing, when decomposed signals contain components of different frequencies. Therefore, to quantify the extent of such problems, OI was applied. The expressions for these metrics are given below [95, 112]:

$$IO = \frac{\sum_t \sum_{\substack{j=1 \\ i \neq j}}^{N+1} c_i(t) c_j(t)}{\sum_t x^2(t)} \tag{5.20}$$

where t is timestamp, $x(t)$ represents original signal, i and j are indexes of components, $c_i(t)$ and $c_j(t)$ are mode functions, and the expression $\sum_t x^2(t)$ represents the total energy of the original signal.

$$\text{Error} = \frac{\left\| \hat{\mathbf{X}}_{N \times L} - \mathbf{X}_{N \times L} \right\|_2}{\left\| \mathbf{X}_{N \times L} \right\|_2} \tag{5.21}$$

where $\mathbf{X}_{N \times L}$ is the original signal, and $\hat{\mathbf{X}}_{N \times L}$ is the reconstructed signal.

Statistical significance tests enable drawing conclusions concerning defined hypotheses related to one or more samples. The hypotheses can either be rejected or accepted based on the probability evaluation, whether about the relationship or difference in data and whether it pertains to chance. Based on the data distribution criteria, two categories of statistical significance tests can be found: parametric and nonparametric. Parametric tests are employed when the data is assumed to follow a distribution, most commonly the normal distribution. Examples of parametric tests are t-test, F test, and ANOVA. In contrast, nonparametric tests do not require prior knowledge of data distribution, and here, it is valid that if the criterion of normal distribution of data is not satisfied, nonparametric tests are applied [86]. Also, a significant advantage of applying these tests can be seen when it comes to extreme values since they are less sensitive to them. Examples of nonparametric tests are Mann-Whitney U, Kruskal-Wallis, and Wilcoxon signed-rank tests. As the models' residuals mostly deviate from normality (see Appendix C.2, Table C.3, and Figures C.1 and C.2 for more details), we adopt a nonparametric approach, specifically the Wilcoxon signed-rank test with Bonferroni correction.

Wilcoxon signed-rank test has already been applied in a previous study by Corazza et al. [26] in a similar manner as here, mainly due to its robustness regarding the t-test assumptions, i.e., deviations from them [25]. The nonparametric test was based on the absolute residuals of every model, which then provided a statistical assessment of whether the median of pairwise differences can be considered statistically significant or not, depending on the calculated p-values. The p-value threshold α was set to 0.05, consistent with the study by Corazza et al. [26]. If the statistical test result showed

that the p-value was less than the given threshold, it was possible to conclude that the difference between the two samples is statistically significant, and if otherwise, it was not. However, these results only provide additional acknowledgment of the previous results obtained through evaluation criteria and the fact that the difference between the absolute residuals is not due to chance.

Multiple testing introduces the problem of Type I errors; therefore, the Bonferroni correction was applied in order to reduce the risk of false positives of the conducted Wilcoxon Signed-Rank test performed for the models' residuals. The Bonferroni method adjusts the overall significance level by dividing it by the number of independent tests performed. The main advantage of this method is its straightforward application and simplicity [100].

5.5. Model Interpretability and Transparency

Nowadays, with the significant rise in countless variations of machine learning models, the need for providing complete model transparency has increased. Most of today's applied models can be considered black-box models, which is especially the case with Deep Neural Networks. Despite the potential for high accuracy, there is very little understanding of the model's internal operational mechanics. So, the need for human-understandable reasoning behind the models led to the development of explainable artificial intelligence (XAI) with the aim of satisfying ethics and regulatory demands. A significant increase in research interest related to XAI has only been observed since 2017 [8]. In the same paper, a taxonomy of different approaches is given based on their level of transparency and explainability, where three classes can be differentiated: transparent models, post-hoc explainability (model-agnostic and model-specific), and hybrid models. However, when providing a specific focus on the hydrological community, in the last five years, there has been a significant increase in the application of XAI, as reported by numerous studies [84]. An example of a review paper by Maier et al. [84] on XAI application in hydrology provided a different outlook on the application of XAI, more specifically a classification focused on different types of explanations, mainly: assessment of feature contributions, assessment of a model's adaptability to feature variations, and lastly, selection of relevant features. As XAI offers vital tools for explaining complicated mechanisms, our research

mainly focused on assessing feature contributions while including domain-specific knowledge to provide adequate result analysis and prevent misleading assumptions.

The first step of determining feature contribution was conducted in the data pre-processing stage by employing two Exploratory Data Analysis techniques: correlation analysis and mutual information score. The techniques differ based on the type of feature interdependency. While correlation analysis is mainly focused on finding a linear association between the features and the strength and direction of the relationship between them [119], the mutual information score is a broader concept of analysis, where the goal is to find any type of dependence between features, i.e., it is applied if complex nonlinear relationships between features are present [11]. The second stage of determining feature contribution was performed during the model evaluation phase by employing the SHAP technique for the category of simple machine learning models. In contrast, the feature occlusion test was used for the time-series machine learning models.

SHAP is a methodology founded on Sharpley values and concepts from cooperative game theory [133]. Unlike other XAI methods, in SHAP, the predictions are represented as a summation of every feature contribution [111]. More precisely, calculated SHAP values quantify the magnitude of every feature deviation, whether above or below, from the model's prediction relative to the baseline, i.e., average estimation. Furthermore, SHAP differentiates itself from other interpretability approaches by its mathematical framework since it effectively fulfils three critical properties: local accuracy, missingness, and consistency [82]. Another aspect that makes this method markedly different from others is the possibility of insight into local and global interpretations of model results. Global SHAP interpretations answer the question of the most important features for the overall model predictions. At the same time, local SHAP is more focused on the question of how every feature influences the model's individual predictions [7]. However, it is important to emphasize that global interpretations are derived from local ones. The local and global interpretation can be described by the following mathematical expressions [102]:

$$g(\mathbf{z}') = \phi_0 + \sum_{k=1}^{M} \phi_k(z'_k) \tag{5.22}$$

$$g(\mathbf{z}') = \phi_0 + \sum_{j=1}^{M} \phi_j z_j' \tag{5.23}$$

where ϕ_0 represents the baseline, z_k' and z_j' are the binary variables (0 = feature is included, 1 = feature is not included), $\phi_k(z_k')$ is the kth feature attribution, and ϕ_j is the feature attribution for j^{th} feature. The biggest difference between these formulations is related to the input z', as in local expression, it represents an alliance vector whose values can be either 0 or 1 ($\mathbf{z}' \in \{0,1\}^M$), while in global expression it represents all possible coalitions of features ($\mathbf{z}' \in [0,1]^M$).

The use of the SHAP technique is quite straightforward for the simple machine learning category, particularly for DT, RF, LGBM, and XGB, where the Tree Explainer method is employed. However, the issue arises with the SVR approach, as the necessary calculations cannot be performed across the entire range of predictions due to the computing power required and the substantial amount of time involved (regarding aspects such as the algorithm's complexity), but rather only on a subset of the data utilizing a model-agnostic Kernel Explainer. In contrast to Tree Explainer, Kernel Explainer can be resource-intensive for large datasets [9]. It operates under a black-box methodology, which can lead to biased outcomes if adequate representative data is not supplied. Additionally, it offers an estimation of feature contributions without incorporating kernel-specific transformations. A drawback of SHAP techniques pertains to any RNN-based layer, specifically in our case, LSTM. PyTorch LSTM implementation is not supported by the available SHAP framework; therefore, a different approach for determining feature contribution must be considered. Besides the aforementioned points, a potential issue that could adversely affect the SHAP interpretation is linked to the presence of highly correlated features. In such cases, the importance may not accurately represent the true relevance, thereby further limiting the effectiveness and precision of the SHAP method [93]. Therefore, it is possible to make false conclusions regarding certain features and whether they are truly indispensable or of negligible importance.

Feature occlusion analysis is a methodology also conducted in the post-evaluation phase, in which models were retrained from the beginning by systematically excluding variables that exhibit the highest correlation with other features. The selected approach is another form of providing model interpretability. The correlation was determined through

the correlation matrix, along with additional insights from SHAP regarding feature contribution across various models. The rationale for eliminating a feature that has a high correlation with another is that it contains nearly identical information; hence, evaluating the model to ascertain the redundancy of that feature might yield valuable insights into the problem at hand. This strategy is computationally intensive; nevertheless, it allows us to determine if the model can be simplified with minimal impact on performance and possibly a model that is easier to interpret and uses less time for the training process.

Based on the applied methodology, the subsequent chapter presents the results of the conducted estimation and prediction analysis, followed by a discussion regarding the findings.

Chapter 6

RESULTS AND DISCUSSION

Contents

This chapter presents an overview of the results and discussions related to three research directions: (1) discharge estimation, (2) the impact of signal decomposition on discharge estimation, and (3) prediction of discharge and water level. While measured data and STREAM 1D simulated data were used for directions (1) and (2), HEC-RAS simulated data was employed for direction (3).

6.1. Discharge Estimation

The discharge estimation problem has been investigated from two different perspectives: one by using a measured dataset and the second by using a simulated dataset generated by the STREAM 1D numerical model. Each of the datasets had been evaluated using different machine learning models, from simple non-temporal approaches, such as DT, RF, SVR, LGBM, and XGB, to more complex time-series models, LSTM, and the proposed hybrid modeling approach, LSTM-Attention. Each scenario used only water level information from four available stations as model inputs. The following subsection presents results and discusses models' performance by using different evaluation metrics and statistical significance tests, as well as identifying feature importance.

6.1.1. In-depth Analysis for Measured Data

The order of the models according to the accuracy shown on the test set is as follows: LSTM-Attention, LSTM, LGBM, XGB, RF, DT, SVR-rbf, and SVR-sigmoid. Appendix C.1 includes Table C.1 detailing selected hyperparameters, their tested ranges, and the optimal values identified via a five-fold grid search cross-validation method. According to the results of the evaluation metrics shown in Table 6.1, time-series models, LSTM-Attention and LSTM, in addition to providing the highest accuracy in discharge estimation, also show the highest consistency in performance, both in the training set and in the testing set. For LSTM-Attention, the differences in performance were 1.56% for RMSE, 5.39% for MAE, 3.59% for NSE, and 2.02% for R, while for LSTM 7.21% for RMSE, 13.24% for MAE, 3.61% for NSE, and 1.92% for R. Furthermore, it is important to emphasize that the training dataset does not contain the highest peak value. Nonetheless, even in such cases, common for real-world scenarios, the time-series models demonstrated an exceptional ability of data generalization and extrapolation. The largest oscillations in performance were observed for SVR-rbf, RF, and DT, whereas LGBM and SVR-sigmoid exhibited only moderate oscillations. The difference in performance between the best-performing model, LSTM-Attention, and the worst, SVR-sigmoid, was 30.12% for RMSE, 26.77% for MAE, 2.77% for NSE, and 4.25% for R. The difference in performance between the second-best-performing model, LSTM, and the worst, SVR-sigmoid, was 22.71% for

RMSE, 18.94% for MAE, 2.27% for NSE, and 4.15% for R. The difference between the two best performing models, LSTM-Attention and LSTM, was 9.59% in RMSE, 9.66% in MAE, 0.52% in NSE, and 0.10% in R.

Table 6.1: Evaluation of models on the training and testing dataset obtained from measured data (the two best-performing models for each dataset separately are emphasized in bold)

Method	Training				Testing			
	RMSE (m^3/s)	MAE (m^3/s)	NSE	R	RMSE (m^3/s)	MAE (m^3/s)	NSE	R
DT	59.203	42.828	0.933	0.966	76.147	54.368	0.955	0.980
RF	**55.812**	**40.674**	**0.941**	**0.970**	73.306	52.159	0.958	0.982
SVR-sigmoid	70.577	54.090	0.905	0.952	82.155	58.992	0.947	0.947
SVR-rbf	60.472	43.564	0.930	0.965	77.387	54.410	0.953	0.979
LGBM	59.572	43.360	0.932	0.966	73.130	52.260	0.958	0.983
XGB	**48.500**	**35.601**	**0.955**	**0.977**	73.212	51.836	0.958	0.982
LSTM	58.916	41.490	0.934	0.969	**63.495**	**47.821**	**0.969**	**0.988**
LSTM-Attention	56.509	40.874	0.939	0.969	**57.406**	**43.201**	**0.974**	**0.989**

Figure 6.1 displays the predicted versus observed plots for all machine learning approaches. The visualization only affirms the general results of evaluation metrics. Models with the widest dispersion of points, i.e., lowest accuracy in prediction, are DT, RF, SVRs (rbf and sigmoid), LGBM, and XGB. The same models encounter challenges when predicting values above 1500 m^3/s. Although most of these models tend to underestimate discharge above 1500 m^3/s, it is primarily the SVR approaches that show pronounced overestimates of these values. Different results are shown for LSTM-Attention and LSTM, whose points are closely scattered to the best-fit line, indicating a good fit, which is also supported by the R^2 metric. Nonetheless, LSTM-Attention exhibits a more balanced distribution of points, whereas LSTM shows minor deviations with points positioned below the best-fit line.

A more detailed assessment of predicted discharge values as categories is presented in Figure 6.2. Discharge values were categorized into four distinct groups: (1) low discharges are defined as values below 300 m^3/s; (2) medium discharges encompass values ranging from 300 to 1050 m^3/s; (3) high discharges include values between 1050 and 1500 m^3/s; and (4) extremely high discharges correspond to values of 1500 m^3/s and above. Low discharges encompass the largest proportion of values, approximately 60% of the entire dataset. The subsequent category accounts for around 30%, followed by a category
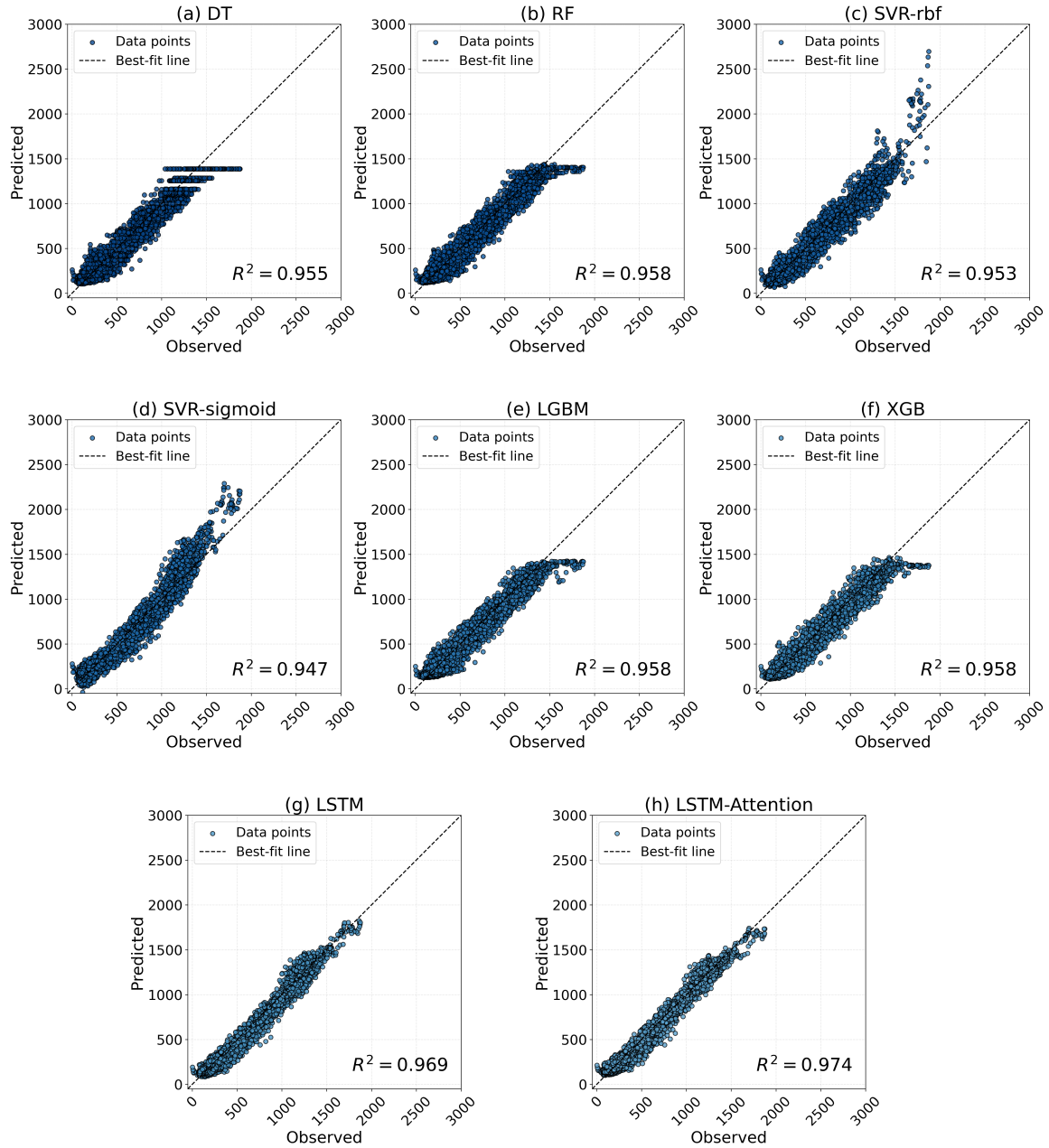
Figure 6.1: Measured data: Predicted versus observed

slightly under 10%, and the final category represents less than 1%. The frequency distribution of discharge categories is also illustrated in Figure 6.2. The smallest deviations in performance were observed for categories (1) and (2). The MAE of the first category was less than 20 m^3/s, the second between 60 and 80 m^3/s, while greater deviations apply to category (3), where MAE values were approximately in the range between 60 to 110 m^3/s, and the highest for category (4) whose MAE values deviate most considerably, starting from the best models whose MAE was below 100 m^3/s to the worst models whose

MAE was higher than 200 m³/s. In three out of four categories, LSTM-Attention had shown consistently better prediction accuracy, while in the (4) category, it was outperformed by less than 40 m³/s by LSTM. However, although some deviations are observed, LSTM-Attention and LSTM were still the only ones that provide less than 100 m³/s for the (4) category, which contains the most extreme discharge values with the lowest data frequency. Hence, the MAE below 100 m³/s is still considered satisfying.
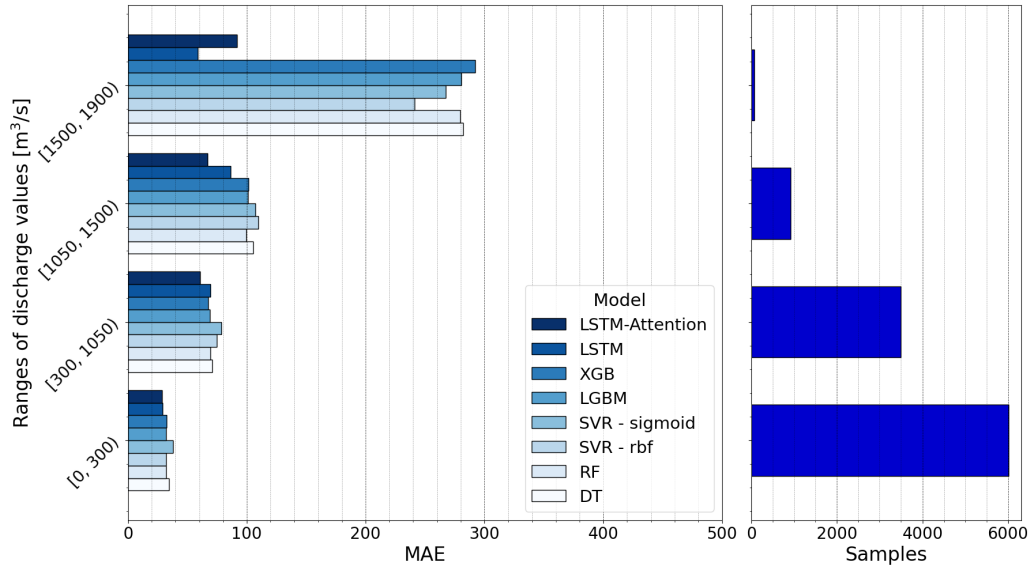
Figure 6.2: Measured data: MAE metric and frequency distribution of discharge categories

Another visual statistical description, the Taylor diagram, was included to generate a holistic assessment, as it employs three metrics: R, standard deviation, and centered RMSE. Figure 6.3 shows how each model performs in compliance with these three metrics. As the reference value of standard deviation for field measurements was 358.343, the models with the closest and lowest relative deviation were LSTM-Attention (-3.29%), LSTM (-3.28%), SVR-sigmoid (1.68%), and SVR-rbf (3.93%). Other approaches had relative deviations below -5%. However, although SVR-sigmoid had the closest standard deviation, it demonstrated the lowest level of performance in estimation ability and lack of fit. Hence, the models that provided an adequate balance between all metrics, especially superior performance in RMSE and R, were LSTM-Attention and LSTM.

An approach that enabled us to determine whether there are statistically significant differences between the evaluation metrics of simple non-temporal and time-series machine learning models, given in Table 6.1 and plots of predicted versus observed given
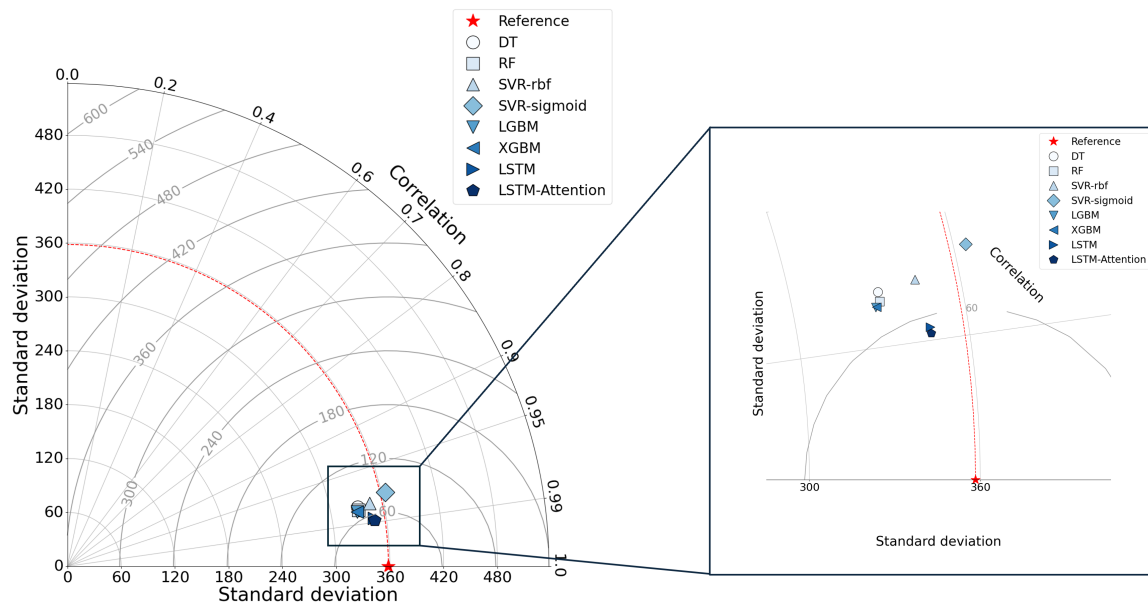
Figure 6.3: Measured data: Taylor diagram

in Figure 6.1, was selected. The results of the Wilcoxon signed-rank test are presented in Appendix C.2, Figure C.3 in matrix form with p-values. The analysis of the LSTM-Attention model's results revealed a statistically significant difference in performance when compared to simpler machine learning models and LSTM, which exhibited the closest performance in terms of both evaluation metrics and predicted versus observed outcomes. Similarly to the LSTM-Attention, LSTM also demonstrated a statistically significant difference in performance when compared to simple machine learning models. This analysis has confirmed the effectiveness of time-series models for discharge estimation problems using field measurements that still contain noise and irregularities. No significant performance differences were observed in the absolute error values between the LGBM and XGB models, as well as between the RF and LGBM models. Nonetheless, the relationship between RF and XGB models exhibited a minimal difference.

As the Wilcoxon signed-rank test does not account for potential Type I error arising from multiple comparisons, a Bonferroni correction was applied. The pairwise comparisons shown in Figure 6.4, led to the same conclusions as those of the Wilcoxon signed-rank test, indicating that the significance decisions remained unchanged after correction.

Determining the feature importance of input variables was carried out in two phases. During the first phase, which was carried out before training the model, the correlation

Figure 6.4: Measured data: Bonferroni correction performed on Wilcoxon Signed-Rank p-values

matrix and mutual information were used.

The correlation matrix provided in Appendix C.3 as Figure C.5 depicts a linear relationship between input features (water level stations) with the output feature (discharge feature measured at Metkovićć). The order of water level stations based on the correlation was: Metković (0.54), Kula Norinska (0.48), Opuzen (0.45) and lastly Ušće (0.21). The order of water level stations based on the calculated correlation coefficient was the same as their distance from the discharge Metković station. The highest correlation at Metković was expected as there is a direct relationship between discharge and local water level. Other stations exhibit moderate (Kula Norinska and Opuzen) to low correlation with the target station Metković. A strong relationship was found between adjacent water level stations, ranging from 0.96 for the relationship between Kula Norinska i Opuzen, to 0.62 for the relationship between Metković and Ušće. The result of the correlation analysis indicates that closer stations (Metković, located 90 m downstream from the H-ADCP devices, Opuzen, and Kula Norinska) were more informative for predicting discharge at Metković than the water level station located at the river mouth (Ušće).

As a correlation matrix is limited to finding linear interactions between the features, mutual information was employed (see Appendix C.3, Figure C.6). The same trend regarding the input feature predictors was also observed, where Metković was the most relevant feature, followed by a moderate importance of Kula Norinska, Opuzen, and the least important predictor, Ušće.

However, from the hydraulic viewpoint, the assumption that the downstream stations do not contribute to predicting the discharge of the upstream stations is not acceptable. Domain knowledge has already affirmed how tidal dynamics and sea levels influence both the water level and the flow pattern of the upstream stations in a tidal reach. Hence, excluding Ušće from our input features can cause potential issues as an oversimplified system.

The second phase of determining feature importance was conducted after model training. For simple machine learning models, SHAP analysis, whose results are presented in Appendix C.3, as Figures C.7a to C.7d, was applied. For all models, it was confirmed that the most considerable influence on final predictions had the Metković and Ušće water level stations. The gained insight into the importance of input features supported the previous claim on including downstream stations, which was also supported by domain knowledge.

As the SHAP approach had limitations regarding the time-series models developed in PyTorch, the feature occlusion method was utilized to determine feature contribution by testing it on different combinations of input features. Both LSTM-Attention and LSTM showed a notable decline in performance, around 50% and more when only water level from Metković station was used. This also supported the claim of using additional water level station data. The results indicated that additional input features were required to gain satisfying estimation accuracy. The largest improvement was observed when the water level from Ušće was used alongside Metković. Hence, this also confirms previous SHAP analysis results. By adding a single station, the results were improved, LSTM by 37.45% for RMSE (see Figure 6.5), and LSTM-Attention by 46.87% (see Figure 6.6). The remaining input features also showed improvements but in smaller percentages, less than 10%. Hence, the highest impact on the model estimations had the upstream and downstream stations, with a smaller impact by including midstream stations. A detailed assessment of the performed feature occlusion is provided in Appendix C.3, in Tables C.4 and C.5.
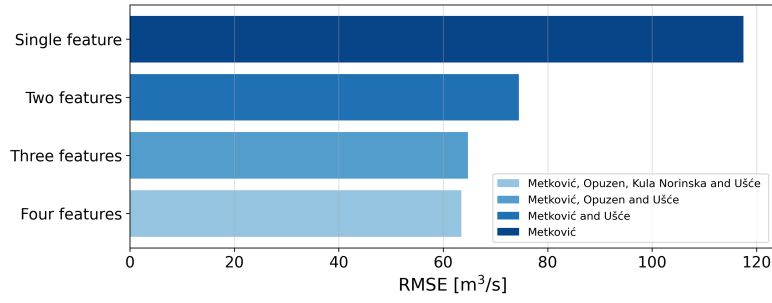
Figure 6.5: Measured data: Feature occlusion using LSTM [88]
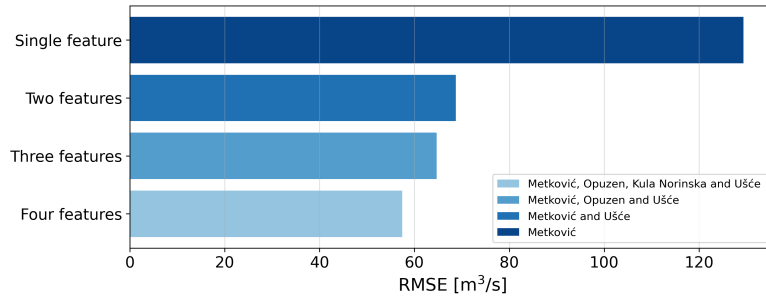
Figure 6.6: Measured data: Feature occlusion using LSTM-Attention [88]

6.1.2.   In-depth Analysis for Simulated STREAM 1D Data

As for the field measurements, the first step of the model evaluation phase involved the analysis of performance, whose results are presented in Table 6.2. The presented models, as previously, were optimized using a five-fold grid search cross-validation, and final hyperparameters, their tested ranges, and optimal values are located in the Appendix C.1, specifically Table C.2. Similarly to field measurements, the order based on the models performance on the testing set did not change drastically; only a slight deviation was observed. Hence, models arranged by decreasing estimation accuracy are LSTM-Attention, LSTM, XGB, RF, DT, SVR-sigmoid, LGBM, and SVR-rbf. Only LGBM showed reduced estimation performance on the simulated dataset than on field measurements, while SVRs consistently demonstrated their inferior performance, and time-series models performed the best. Differences in performance between the training and testing datasets have decreased; hence, starting from the highest estimation accuracy model, LSTM-Attention, here is the statistical overview. LSTM-Attention gave better RMSE in test set of about 2.11%, 0.99% for MAE, 1.01% for NSE, and 0.60% for R, also as LSTM 2.49% for RMSE, 1.86% for MAE, 1.51% for NSE, and 0.80% for R. A comparison between the LSTM-

Attention and SVR-rbf showed differences in performance by 57.97% in RMSE, 38.37% in MAE, 3.22% in NSE, and 1.42% in R. While with the second-best performing model, LSTM the differences were 50.97% in RMSE, 25.98% in MAE, 3.01% in NSE, and 1.32% in R. A comparison of the two best models showed that LSTM-Attention outperformed the LSTM approach by 14.28% in RMSE, 16.73% in MAE, 0.20% in NSE, and 0.10% in R.

Table 6.2: Evaluation of models on the training and testing dataset obtained from numerical simulation (the two best-performing models for each dataset separately are emphasized in bold)

Method	Training				Testing			
	RMSE (m^3/s)	MAE (m^3/s)	NSE	R	RMSE (m^3/s)	MAE (m^3/s)	NSE	R
DT	39.005	29.541	0.971	0.985	51.747	37.492	0.979	0.990
RF	33.542	24.919	0.979	0.989	48.729	34.669	0.981	0.991
SVR-sigmoid	47.756	37.784	0.957	0.978	53.010	40.978	0.978	0.990
SVR-rbf	39.712	29.797	0.970	0.985	70.128	36.554	0.962	0.983
LGBM	37.392	28.176	0.973	0.987	53.797	36.341	0.977	0.989
XGB	**31.147**	**23.761**	**0.982**	**0.991**	48.428	34.739	0.982	0.991
LSTM	35.240	27.561	0.976	0.988	**34.384**	**27.057**	**0.991**	**0.996**
LSTM-Attention	**30.094**	**22.753**	**0.983**	**0.991**	**29.473**	**22.530**	**0.993**	**0.997**

A better insight into the model's performance was gained through the predicted versus observed plot, presented in Figure 6.7. The same trend as in field measurements was observed, that the best models with the narrowest spread of points around the best-fit line were the time-series models (LSTM-Attention and LSTM), while remaining simple machine learning approaches produced a notably wider dispersion with either large underestimates for values above 1500 m^3/s (DT, RF, LGBM, and XGB), or overestimates (SVR variants). Regarding the same data range for time-series models, LSTM-Attention shows almost ideal alignment for the values above 1500 m^3/s, while LSTM was not as successful, with some deviations indicating overestimations. However, one specific characteristic was seen in all models, and it was the widest dispersion of points below 500 m^3/s. Therefore, the results of the evaluation metrics were further confirmed by the illustrated plot, which provided an insight into the models' performance on the overall discharge range.

Besides the predicted versus observed plot, a focus is also given to different discharge category ranges to determine how each model performs in different ranges, which is closely related to the data frequency of each category, as shown in Figure 6.8. Consistent with
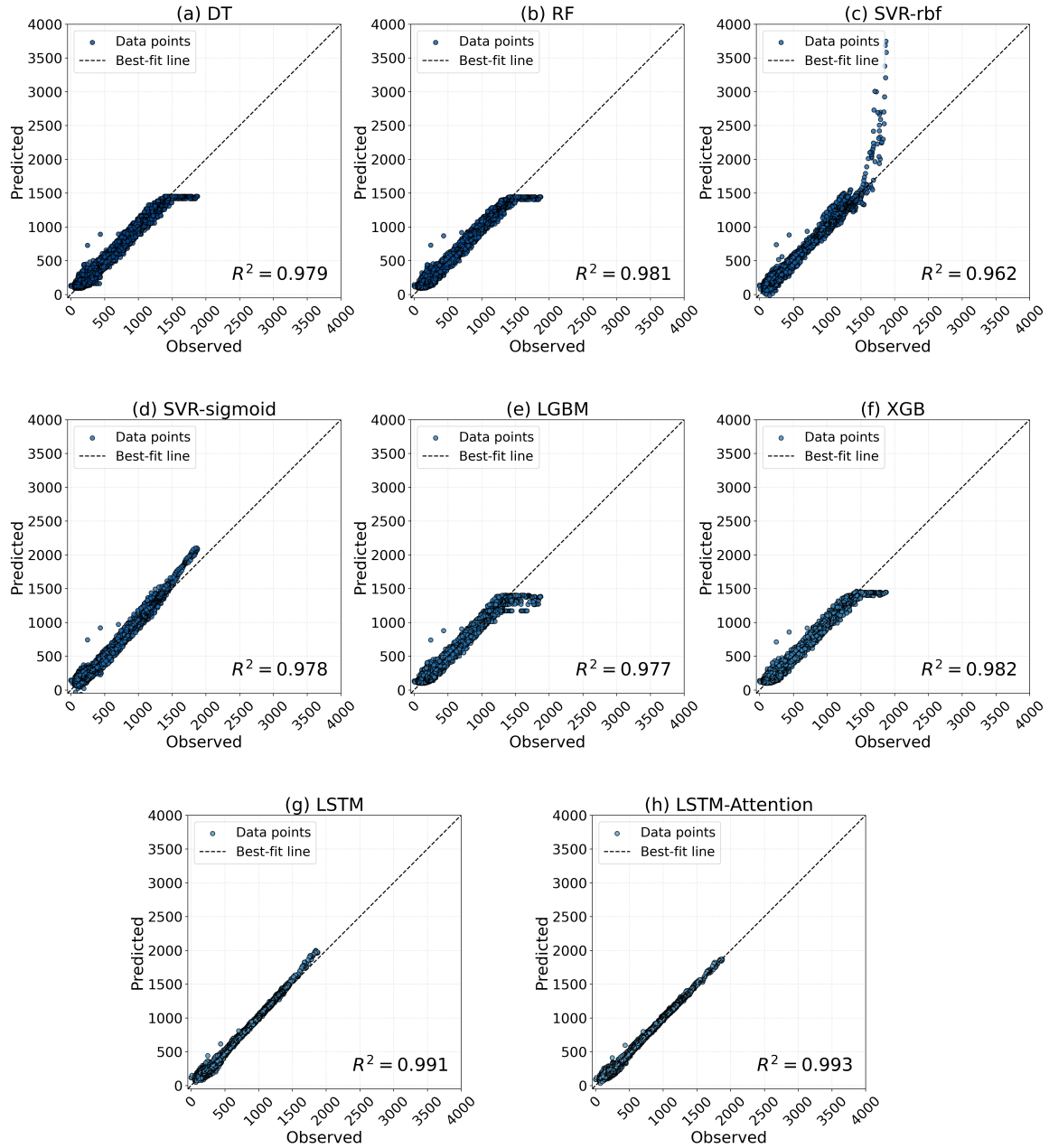
Figure 6.7: Simulated data: Predicted versus observed

field measurements, four categories of discharge are considered and evaluated, and in each category, LSTM-Attention outperformed all models, including LSTM, especially for the category of extremely high discharge values, while also achieving high estimation accuracy. The second-best model was LSTM, whose performance drastically declined for the last category; nonetheless, given the MAE remained below 100 m^3/s, its accuracy was still deemed satisfactory. Consequently, the low discharge categories exhibited MAE ranging from 20 to 40 m^3/s, the medium category from 20 to 50 m^3/s, and the high category

for time-series models below 40 m^3/s. In contrast, simple machine learning approaches yielded an MAE between 40 and 60 m^3/s. Finally, the last category demonstrated the lowest level of performance for simple machine learning models, with an MAE ranging from 160 to 460 m^3/s.
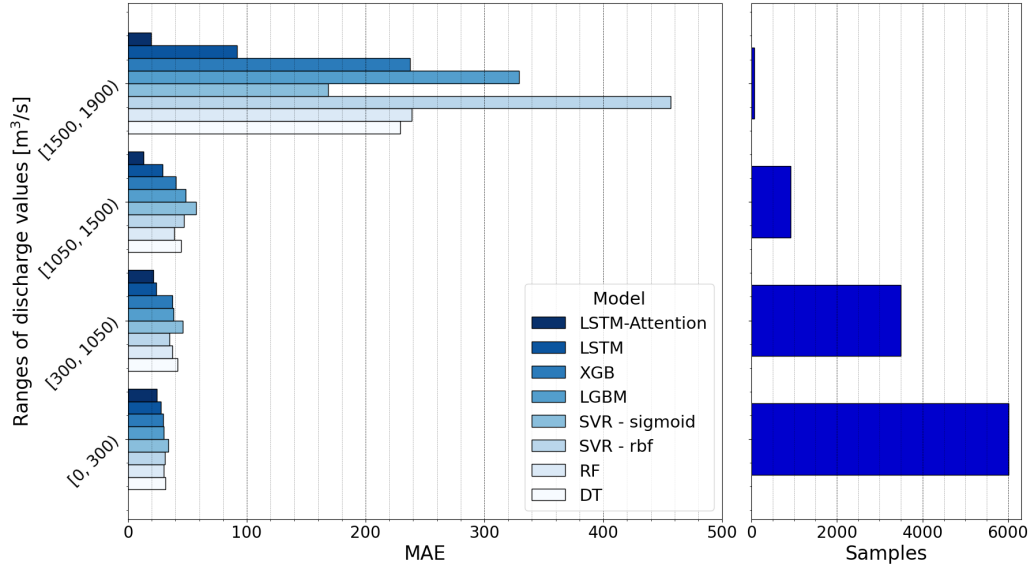
Figure 6.8: Simulated data: MAE metric and frequency distribution of discharge categories

Taylor's diagram in Figure 6.9 depicts how the selected models perform according to three metrics. The standard deviation reference for the STREAM 1D dataset was 358.275, and the order of the models from the closest to the furthest was: LSTM-Attention (-0.18%), XGB (-2.04%), RF (-2.04%), LSTM (2.10%), DT (-2.12%), LGBM (-2.77%), SVR-sigmoid (3.73%) and SVR-rbf (5.09%). These results are quite close to the results of RMSE and R; however, there are some minor inconsistencies. LSTM had a slightly higher relative deviation than XGB and RF. Nevertheless, the remaining metrics, which were noticeably better, suggest that these differences are more consequential. Additionally, the interchange in ranking between LGBM and SVR-sigmoid did not reveal any major performance disparities. Hence, the two best-performing models that exhibited balance in all considered metrics were LSTM-Attention, with almost ideal alignment with the standard deviation reference and LSTM.

Wilcoxon signed-rank (see Figure C.4 in Appendix C.2) test showed that the differences in results between LSTM-Attention and the other models were statistically significant. Similar results were observed for LSTM, but this did not extend to all simple
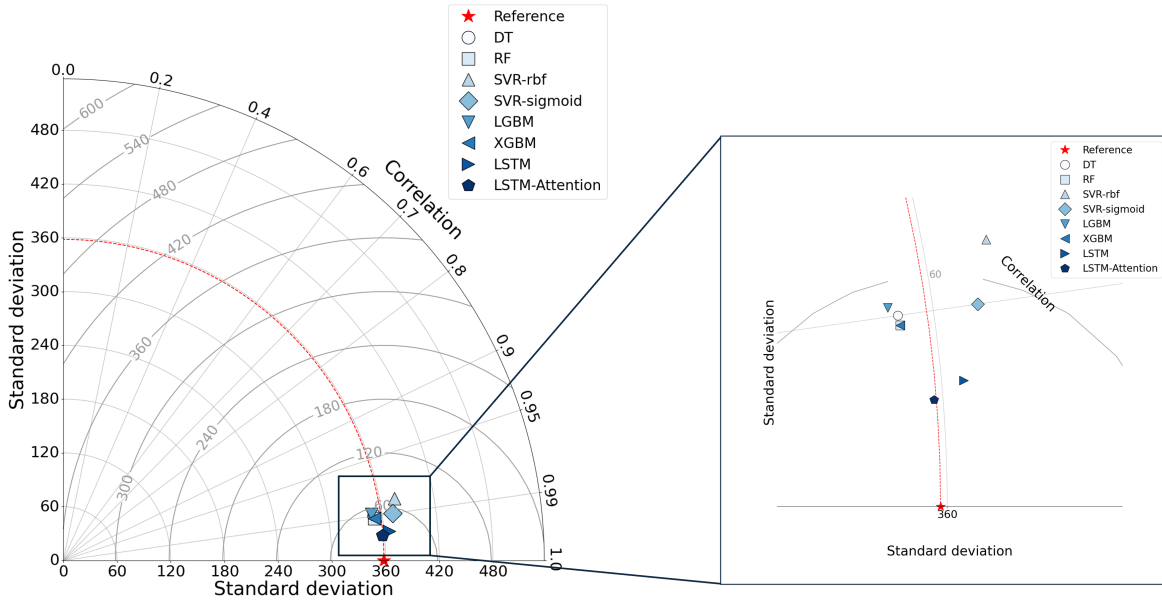
Figure 6.9: Simulated data: Taylor diagram

machine learning approaches. For example, the model pairs for which the p-value exceeded the predefined threshold were RF and XGB, SVR-rbf and XGB, and RF and SVR-rbf. Therefore, the null hypothesis could not be rejected for them, indicating that no statistically significant difference was observed between the models' residuals.

By applying a Bonferroni correction, a single difference in statistical significance was identified between SVR-rbf and LGBM models, indicating that the results of the models cannot be considered statistically different. Although the pairwise comparisons (Figure 6.10) led to conclusions largely consistent with those of the Wilcoxon signed-rank test, some deviations were observed, suggesting that the Wilcoxon signed-rank test may have resulted in a Type I error.

The feature importance for the simulated data was firstly assessed before training the models by examining the correlation matrix and mutual information, to gain insights into both linear and nonlinear relationships of input features with the output feature.

The correlation matrix, depicted in Appendix C.3 Figure C.8, gave results comparable to those of the measured data. The strengths of the linear relationship between the water level stations and discharge, derived from the most relevant predictors, were associated with the discharge from the target station. Hence, the water level station exhibiting the highest predictive power was Metković (0.56), followed by of Kula Norinska (0.53)
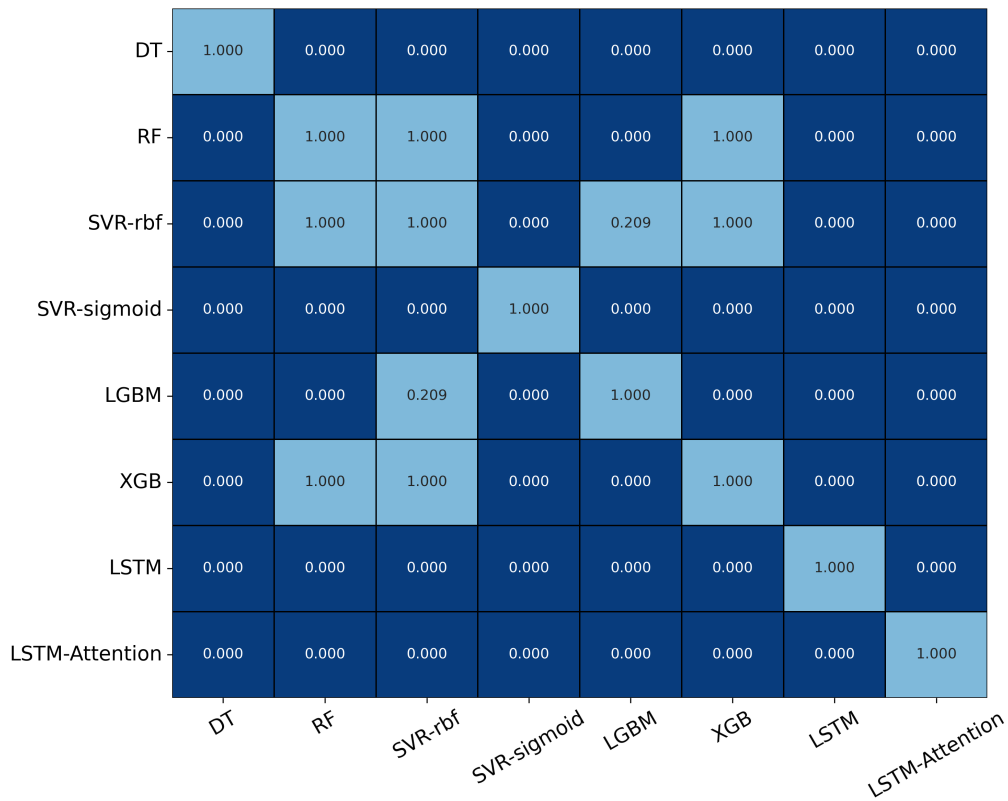
Figure 6.10: Simulated data: Bonferroni correction performed on Wilcoxon Signed-Rank p-values

and Opuzen (0.49) with moderate impact, while Ušće demonstrated the least predictive power (0.24). Focusing on input features, adjacent stations demonstrated the strongest correlation, specifically Metković and Kula Norinska (0.97), Kula Norinska and Opuzen (0.95). The lowest correlation was observed between Ušće and Kula Norinska (0.68), as well as Ušće and Metković (0.65). This visualization allowed us to determine that higher predictive power is associated with the nearest stations.

Mutual information showed a similar influence of features, with exactly the same order as in the correlation matrix, with slightly less impact of the downstream station Ušće, and a slightly increased impact of stations above the river mouth. The decision to remove the Ušće station data was rejected because although these methods show that its influence is almost negligible, considering domain knowledge, this claim cannot be valid. Therefore, for training the model, data from all water level stations, was used.

The SHAP analysis, presented in Appendix C.3 in Figures C.10a to C.10d, and feature occlusion, in Figures 6.11 and 6.12, were applied in the second stage of determining feature importance. The previous decision to use water levels from the Ušće station proved to

be of exceptional importance for model predictions, because it was precisely these water levels, together with water levels from Metković, that proved to be the two most important input features that had the greatest impact on the estimates of all models as depicted in SHAP analysis. The only deviation in these results was for the DT model, where Opuzen had a slightly higher impact, almost negligible, than Kula Norinska.

The feature occlusion, whose results for time-series models are presented as RMSE metric in Figure 6.11 for LSTM, and Figure 6.12 for LSTM-Attention, had only further validated SHAP analysis results concerning feature contribution. Hence, results indicated that including water level from Ušće improved the RMSE by around 73% for both models. By including the additional variable Opuzen, a performance decline of about 7.81% for LSTM, and an increase of 8.37% for LSTM-Attention were observed when compared to the two feature RMSE metric. The performance of LSTM declined by 9.87% with the addition of the Kula Norinska station compared to three feature scenario, whereas LSTM-Attention exhibited an increase in performance of approximately 11.37%. A comprehensive evaluation of the performed feature occlusion can be found in Appendix C.3, namely in Tables C.6 and C.7.
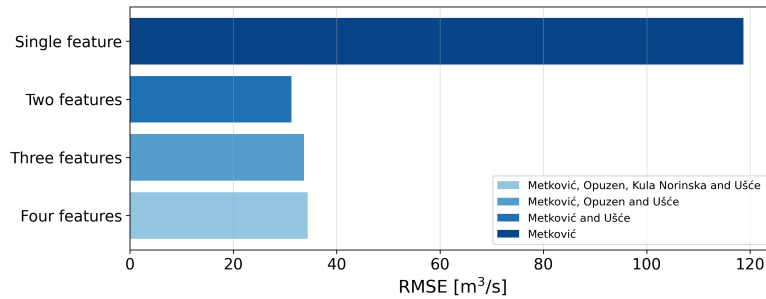
Figure 6.11: Simulated data: Feature occlusion using LSTM [88]
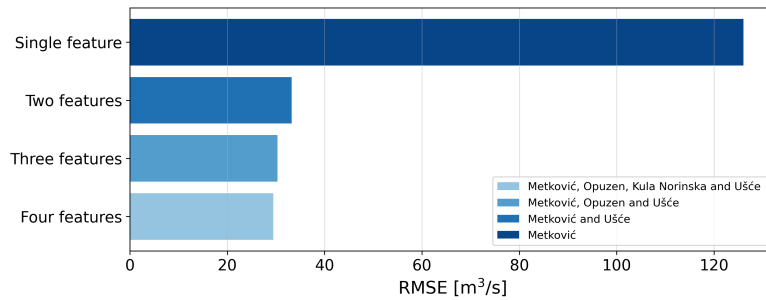
**Figure 6.12:** Simulated data: Feature occlusion using LSTM-Attention [88]

6.1.3.   A Comparative Analysis: Measured versus Simulated data

Previously, a separate analysis of machine learning model results was presented, one using solely field measurement (real-world conditions) and the other using solely simulated data generated by the STREAM 1D numerical model (ideal, noiseless conditions). To generate a clearer overview of this dual dataset approach and compare them, power spectral density plots, i.e., spectrograms, were created for every water level station (see Figure 6.13). Although these visualizations follow a similar trend, for field measurements, more variability was detected, as well as noise. Hence, with these two approaches, one primarily focused on assessing each machine learning model's robustness (field measurements) and the other used for performance validation (STREAM 1D dataset), a balanced evaluation was ensured.

Firstly, these approaches differed in data quality. STREAM 1D data represented a more controlled environment, where, unlike the scenario with measured data, where the presence of noise or possible irregularities poses a problem, it is possible to gain a deeper insight into the model's performance and estimation accuracy. As the measured data consisted of two datasets collected from two different agencies, some challenges were encountered. Features collected from one agency, water levels, passed the quality check, while discharge data from another agency encountered common problems. One issue related to missing data periods, mostly several hours, on one device, which was resolved with interpolation using the remaining two ADCPs. Another issue was the presence of high-frequency noise characteristics for ADCP devices, which can be attributed to the tidal currents and salt wedge dynamic influence in tidal environments, and outliers, which was resolved by applying a moving average of three hours.

Different evaluation metrics, statistical tests, and visual comparisons have led to several conclusions. LSTM-Attention and LSTM have shown great potential in accurately estimating discharge with high and satisfying accuracy, ranking as the two most effective models. LSTM-Attention performance was regarded as the superior choice.

The analysis revealed variations in the magnitude of metrics across results but also consistency in relative performance. Specifically, LSTM-Attention demonstrated improvement for the simulated dataset, achieving a 48.66% reduction in RMSE, a 47.85% reduc-
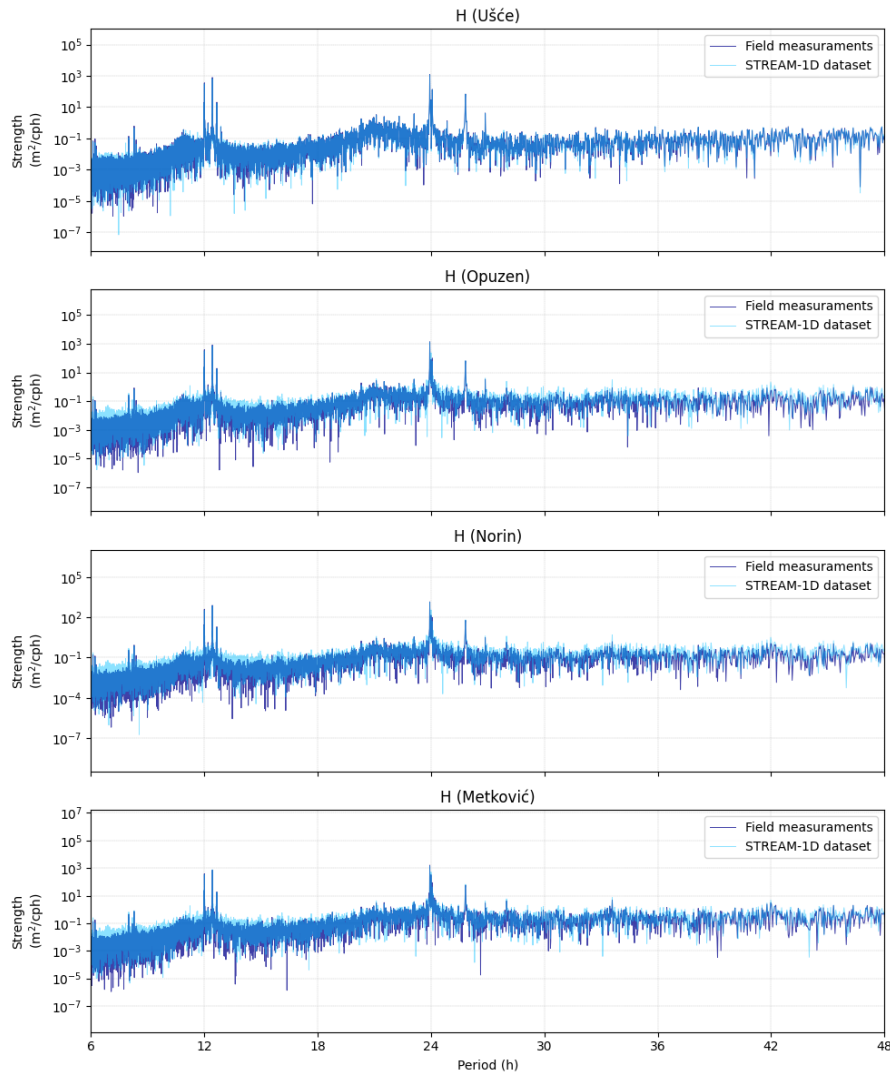
Figure 6.13: Comparison between simulated and measured data of water level power spectral density ranging from the tidal station to the most upstream part of the tidal reach

tion in MAE, a 1.95% increase in NSE, and a 0.81% increase in R compared to field measurements. Similarly, LSTM also demonstrated improvements with a lower RMSE of 45.85%, an MAE of 43.42% , a higher NSE 2.22%, and a R 0.81% for the simulated dataset.

Predicted versus observed plot validated the findings of general evaluation metrics, revealing small deviations in model performance between the two datasets. The LSTM-Attention demonstrated a more suitable fit for the discharge values exceeding 1500 m^3/s in the simulated dataset than for field measurements, although both exhibited sufficient accuracy.

The MAE assessment of discharge categories showed that LSTM-Attention outperformed all machine learning approaches in all four categories for the simulated data and three out of four categories for field measurements. The only instance in which the LSTM outperformed LSTM-Attention occurred within the extremely high discharge category, approximately 30 m^3/s. However, the LSTM-Attention results remained below 100 m^3/s, rendering the difference acceptable. Unfortunately, all simple machine learning model results considerably declined for the extremely high discharge category.

Models were evaluated holistically using a Taylor diagram for both datasets, which exhibited nearly identical reference standard deviations of 358.343 for measured data and 358.275 for simulated data. Time-series models demonstrated superior performance in minimizing total error, particularly highlighting the effectiveness of LSTM-Attention in capturing data variability. The LSTM approach demonstrated tendencies for overprediction when applied to simulated data, and simple machine learning models exhibited greater deviation from time-series models across all examined metrics.

The evaluation metrics were statistically analyzed by conducting the Wilcoxon signed-rank test with Bonferroni correction on the model residuals. In both scenarios, LSTM-Attention demonstrated a statistically significant improvement over other models, including LSTM, which gave the same results when compared to simple machine learning models. Differences were observed for the analysis of simple machine learning which indicated that certain models yield comparable results to each other, suggesting that the differences among them were not statistically significant.

Discharge time-series comparison of the results of the two most accurate models, LSTM-Attention and LSTM, on the test set for simulated (see Figure 6.14) and measured (see Figure 6.15) data showed a good overall agreement, with some minor discrepancies. February, the month with the highest discharge, had an excellent model agreement for the simulated data, with slightly lower accuracy for the measured data. However, daily oscillations of the same month are adequately captured for the measured data. In August, the month with the lowest discharge, in both scenarios, underestimations occured, with a more pronounced effect found for the measured data. This insight is probably a result of the influence of complex tidal-fluvial interactions. As stated in Matte et al. [87], the tidal oscillations can be amplified or dampened by the river discharge. The operation of the upstream hydroelectric power plants complicates the situation even more. Hence,

it is recommended in the future to apply an NS_TIDE model with a hybrid time-series model to enable robust capturing of high-frequency oscillations and also to investigate the impact of the time window length, which may limit the model in capturing diurnal oscillations.
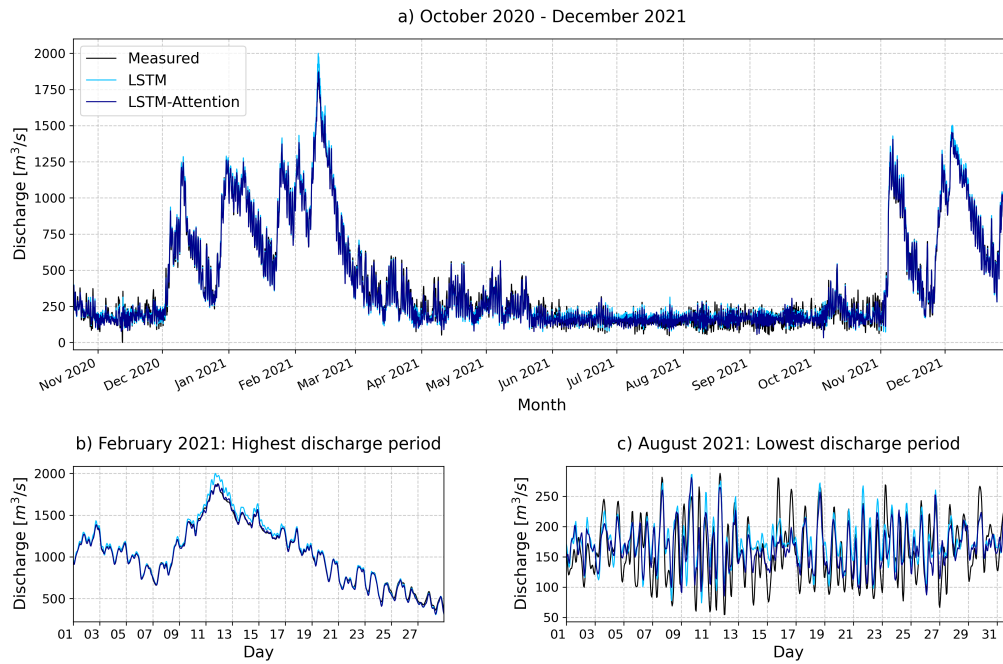
Figure 6.14: Time-series: STREAM 1D dataset

Previous studies by Habib and Meselhe [50] and Hidayat et al. [55] have successfully demonstrated that the selection of input variables more precisely water levels, from multiple locations in the tidal reaches, enhances model performance and accuracy. The selection of water levels from various locations was guided by the findings of prior research. This approach is particularly important in instances where additional data on the river's physical characteristics and meteorological information are unavailable. In addition, the length of the tidal reach, as observed in this study of a small salt-wedge estuary not exceeding 25 km from the river mouth, does not necessitate additional data. This finding aligns with the study conducted by Vu et al. [126], which examined a river hydrographic network of approximately 135,000 km.

The pre-training feature importance techniques, such as correlation matrix and mutual information, were limited and insufficient in their ability to identify predictors with the
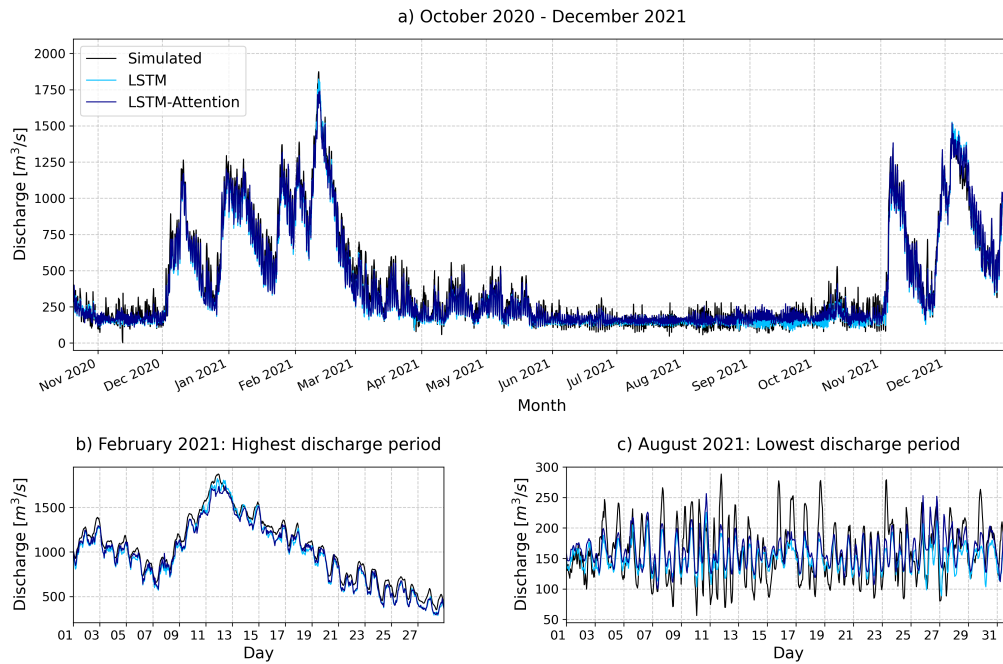
Figure 6.15: Time-series: Field measurements

highest influence on discharge estimation. Therefore, applying only these methods and relying on them for feature selection is not adequate for tidal reaches that embody complex hydrological systems. The post-training techniques, such as SHAP and feature occlusion results, have demonstrated alignment with domain knowledge regarding the influence of tides and sea levels on discharge. Hence, the prior issue to identify Ušće station as an important feature has been resolved. The combination of Metković and Ušće yielded noticeable improvements in estimation compared to relying solely on the water level data from the Metković station. Additionally, the inclusion of other features also showed improvements, but not as high. Hence, this led to the conclusion that all considered water level stations are relevant, in both measured and simulated data.

The study showed the suitability of the LSTM-Attention model for estimating discharge in tidal rivers and estuaries, using only water level data from multiple stations, from the river mouth up to the target discharge location. This approach offered several advantages that are not present in other models, including generalization and extrapolation ability beyond the training set. Additionally, it identifies critical features, effectively manages datasets with imbalanced data distribution, and consistently improves performance

in both simulated and measured data. The extrapolation in modeling environmental processes is of critical importance, mainly because of the increasing climate change that leads to more frequent extreme conditions absent in historical data. The decision to include simple machine learning models was based on previous studies, establishing them as baseline models. Some of the observed issues regarding these models involve extrapolation, as noted in [47], and data imbalance, as highlighted in [117]. Besides their limitations, they have been extensively applied because of their ability to handle nonlinear data, fast training, and also transparency. Although LSTM-Attention required more time for training, the computational cost was justified by the percentage improvement in estimation.

Although a direct comparison of the study results with previous studies is not possible due to differences in either data ranges (values and time), tidal river and estuary characteristics, selected input features, and the used machine learning models, a more general comparison was still feasible. According to Habib and Meselhe [50], the simple machine learning approaches showed lowest level of performance during the estimation of extreme discharge values, as was observed in our conducted study. As applying ANN had a positive impact on resolving such problems, the same can be said for advanced RNNs. A study by Wolfs and Willems [132] employed simulated data to avoid potential issues associated with measured data, as observed in our study. Hence, this was the rationale behind employing simulated data to validate the results of the measured data. Likewise, ANN generalization ability was closely tied to the size of the dataset, especially in the context of temporal dependencies, which motivated the use of advanced RNNs. Hidayat et al. [55] study differed in terms of used input features, which lacked water levels from the station of interest, in discharge ranges which were considerably higher than ours, and also the time span of the data, as our study encompassed six years of data, whereas theirs included less than one year. Hence, these differences can explain better results of our methodology when comparing studies based on the NSE metric. Thanh et al. [117] study differed in both temporal resolution, as they had available daily observations for an extended period above ten years, while our study had available hourly observations for a six-year period. Likewise, they employed several preprocessing techniques to remove trend and seasonality as well as the non-stationary nature of the data, and this is why their results of the simple machine learning models were higher than ours in the measured scenario. However, a major distinction between these two studies is also in the water level

stations, as they chose stations where tidal influence is not present, and in ours, it is. Another issue regarding extreme discharge values is in differences in frequencies of such events for their study, where we only have a single event above 1500 m^3/s. However, the results of the simulated dataset were quite close to the NSE of their yearly round value of DT, RF, and SVR. The last study by Vu et al. [126] showcased similar performance to our LSTM regarding high and extremely high discharge, i.e., flood periods, as its performance decreased in comparison to average conditions, even after including additional meteorological parameters. Nonetheless, we did not observe the same trend for LSTM-Attention, as its performance remains relatively stable and its estimation satisfying even during such infrequent events.

6.2. Signal Decomposition: Measured Data

This subsection represents a continuation of the research on discharge estimation, where earlier findings identified two time-series models, LSTM-Attention and LSTM, which were found to be the most effective and accurate. By employing the VMD method, the original time-series of all water levels were converted into mode functions, then classified as different categories of tidal constituents, and given to the previously best models in order to assess if the decomposition improves the discharge estimation process.

6.2.1. Analysis of VMD Mode Functions

The selection of VMD fundamental parameters for all water level stations, such as the number of mode functions (K), and the quadratic penalty term (α), was the first step of the experiment. The convergence tolerance was set to 1e-6, the bandwidth enforcement (τ) to 0, and other VMD parameters were set to default. Tested mode functions ranged between 2 and 15, whereas the quadratic penalty term ranged from 100 to 5000.

The first criterion that needed to be satisfied for the decomposition process, was the requirement that the minimum K captures all categories of tidal constituents. The classification was conducted based on the study by Lopes and Machado [80], where four categories were defined: long period (LP), diurnal (D), semidiurnal (S), and higher harmonics (HH).

Two examples are presented in Appendix D.1, as Figures D.1 and D.2, where classification of mode functions for the stations Ušće and Metković was performed with specified periods obtained from dominant frequencies. It was observed that if K is too small, more precisely, less than 5, it does not provide sufficient decomposition of the water level signal; therefore, the diurnal component is completely missing, which is referred to as under-decomposition. For K from 8 onward, over-decomposition occurs, where the signal is decomposed into a larger number of high-frequency modes whose content has no useful physical meaning. Therefore, K between 5 and 8 was further investigated for each water level station with different α in order to find the optimal combination.

Additionally, two metrics were used, the orthogonality index (OI) and the reconstruction error ratio. The results of these metrics are presented in Appendix D.1, more precisely from Figure D.3 to D.6. The reconstruction error for all stations was satisfactory for K=5 and K=6. The water level station Ušće had the highest reconstruction error, therefore α was set so that the reconstruction error would give a lower value. Therefore, α above 1000 was taken into further consideration. Regarding the OI metric, similarly to the previous case, only the Ušće station had the highest values among the other stations, and its best value could be obtained for K=5 and K=6, and for α=2000 and above. Since the goal is to find the minimum K, which satisfies all the necessary criteria, the selection process continued with K=5. The final step was a visual inspection of all signals decomposed in the frequency domain, where values below 5000 showed a wider frequency band, leading to spectral leakage, however, α=5000 did not show such problems for any station. Based on extensive research into possible optimal values, the final values of K=5 and α=5000 were determined. The conducted decomposition of signals is presented in Appendix D.2, from Figure D.7 to D.10.

After decomposition, the mode functions were classified into the previously mentioned categories. However, although decomposition had occurred, the residual remained. The problem of residual presence was addressed by further decomposing the residual using the VMD decomposition. This is because the residual maintained both diurnal and semidiurnal components. Despite this, as shown in Appendix D.2, Figure D.11, it did not lead to complete separation of the signal, and it remained strong. For this reason, the residual was summed with a high harmonic component, which is referred to as a residual mixed signal. The final summed and decomposed tidal constituents are presented in Appendix

D.3, as an example, for the tidal station Ušće (see Figure D.12) and the upstream target location Metković (see Figure D.13).

6.2.2. Comparison with Prior Research

Table 6.3 shows a comparison between the results obtained by training the model on time-series data and when using decomposed VMD mode functions. Although time-series data showed better results for the LSTM-Attention model, the opposite insights are shown when VMD is applied to the original input features.

Compared to the previous best LSTM-Attention model, LSTM with VMD, improved estimation results by 13.71% for RMSE, 12.56% for MAE, 0.72% for NSE, and 0.20% for R for the test set. In comparison to its variant trained on time-series data, the LSTM performance improved by 21.99% for RMSE, 21.00% for MAE, 1.24% for NSE, and 0.30% for R, also for the test set.

Evaluation metrics of the LSTM-Attention model, when time-series input variables were used in relation to the mode functions generated by VMD, revealed a minimal but present decline in performance. The model exhibited opposite behavior to that of LSTM when VMD was applied as a data preprocessing strategy. The RMSE decreased by 1.91%, the MAE by 0.83%, the NSE by 0.10%, and the R by 0.10%. Despite the existence of minor deviations, this performance decline was statistically negligible, as there has been no meaningful degradation regarding all considered evaluation metrics. Nevertheless, this overall summary of model performance necessitates a more detailed analysis; therefore, a predicted versus observed plot was used for a more thorough investigation of model performance.

Table 6.3: Time-series vs. Mode functions: LSTM performance comparison with LSTM-Attention on training and testing datasets

| Model	Input	Training set				Testing set			
		RMSE (m³/s)	MAE (m³/s)	NSE	R	RMSE (m³/s)	MAE (m³/s)	NSE	R
LSTM	H TS	58.916	41.490	0.934	0.969	63.495	47.821	0.969	0.988
	H MF + R	53.590	41.491	**0.945**	**0.974**	**49.533**	**37.776**	**0.981**	**0.991**
LSTM-Attention	H TS	**56.509**	**40.874**	0.939	0.969	**57.406**	**43.201**	**0.974**	0.989
	H MF + R	**52.433**	**36.586**	**0.948**	**0.976**	58.524	43.564	0.973	**0.990**

H = water level, TS = time-series, MF = mode functions, R = residual mixed signal

Figure 6.16 presents two models, LSTM and LSTM-Attention, trained on different

datasets: one using time-series data, and the other using generated mode functions together with the remaining residual. Only the LSTM model with classified tidal constituents showed the most tightly clustered dispersion of points around the best-fit line. Other models demonstrated that the best-fit line does not align with the center of the point distribution, except for LSTM-Attention, based on time-series data, indicating a trend of underprediction for discharge at the Metković location. According to the R^2 metric, the model that explained the least variance of the target output was LSTM when trained on time-series data, whereas the model with the highest variance was the same model, but trained on decomposed mode functions and mixed residual signal. The use of VMD here shows no negligible enhancement in the accuracy of LSTM-Attention, which maintained it comparable to that achieved with time-series data, as supported by the R^2 value. The model's performance declined considerably, even more than that of other models, for discharge values exceeding 1500 m^3/s.

Previous visualization showed how the models perform over the entire range of discharge values. Additionally, another assessment was provided that evaluated the models' performance within various discharge categories in relation to the MAE metric. The lowest discharge category, defined as below 300 m^3/s, demonstrated little variation in model performance, with the mean absolute error (MAE) that varied by less than 3 m^3/s. The best MAE was achieved by LSTM-Attention (27.86) with mode functions. The second discharge category, representing medium discharge values between 300 and 1050 m^3/s, displayed moderate variations in performance between the models, with the lowest MAE achieved for LSTM utilizing mode functions (43.26). The largest difference in model performance within this category was observed between the same model trained on time-series data, approximately 37.89%, while minor, for LSTM-Attention, ranging from 26.06% to 28.96%. The third category of high discharge values, between 1050 and 1500 m^3/s, resulted in almost equal MAE in LSTM with mode functions and LSTM-Attention with time-series data, with a difference of less than 2%. LSTM with time-series and LSTM-Attention with mode functions were similar in performance, with a difference of 6.48%. In the final category of extremely high discharge values exceeding 1500 m^3/s, the model exhibiting the lowest MAE was LSTM utilizing time-series data (58.809), followed by LSTM with mode functions (71.269), LSTM-Attention with time-series data (91.741), and finally, LSTM-Attention with mode functions (151.344). Although LSTM-Attention models pro-
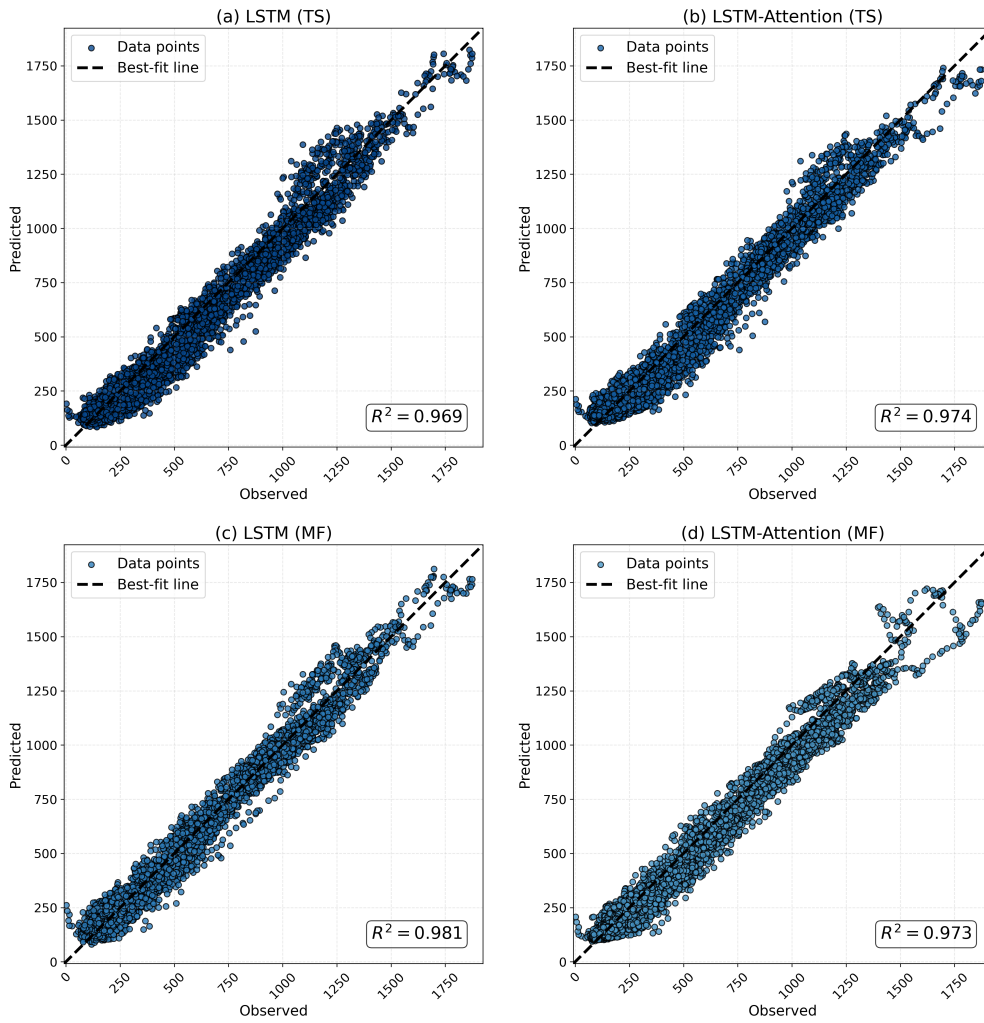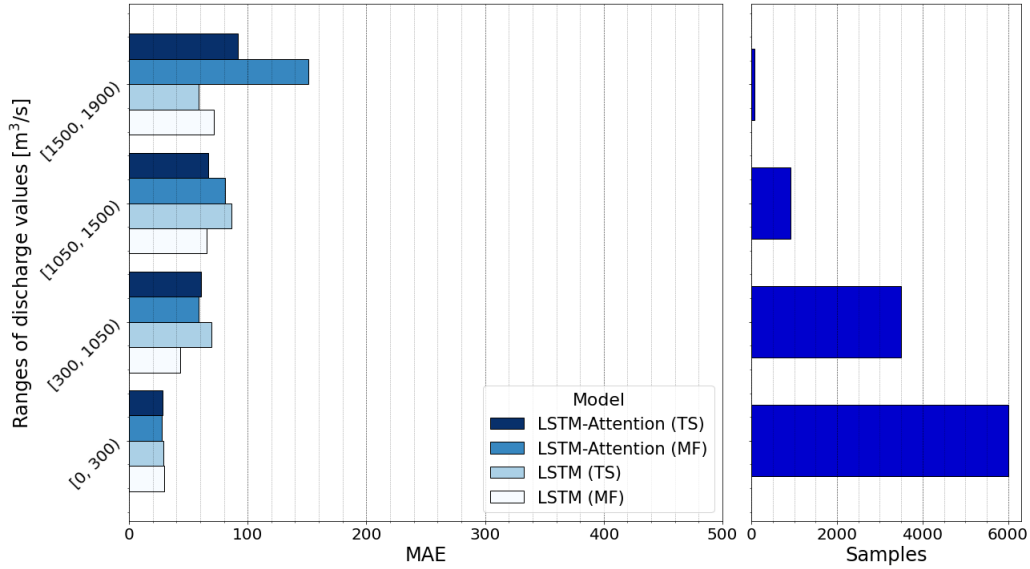
Figure 6.16: Time-series versus mode functions: Predicted versus observed

vided the best estimate of values for the predominant category of low discharges, LSTM with mode functions still proved to be the most consistent solution, providing stable performance through all discharge categories. LSTM-Attention with time-series was nearly comparable to that of LSTM with mode functions for extremely high discharges, where the difference in accuracy increased by 22.31%.

Figure 6.18 shows a comparison of the two best models, LSTM with VMD decomposition and LSTM-Attention with time-series, for the testing set, outlining the months of the highest (February) and the lowest (August) flow. During the month of highest discharge, LSTM aligns most accurately with the original trend; however, it fails to completely estimate the highest discharge peak, a limitation that was also discovered in the prior estimation study. However, in addition to the above, LSTM also contains fewer oscillations in the estimated discharge, compared to LSTM-Attention. Under low flow
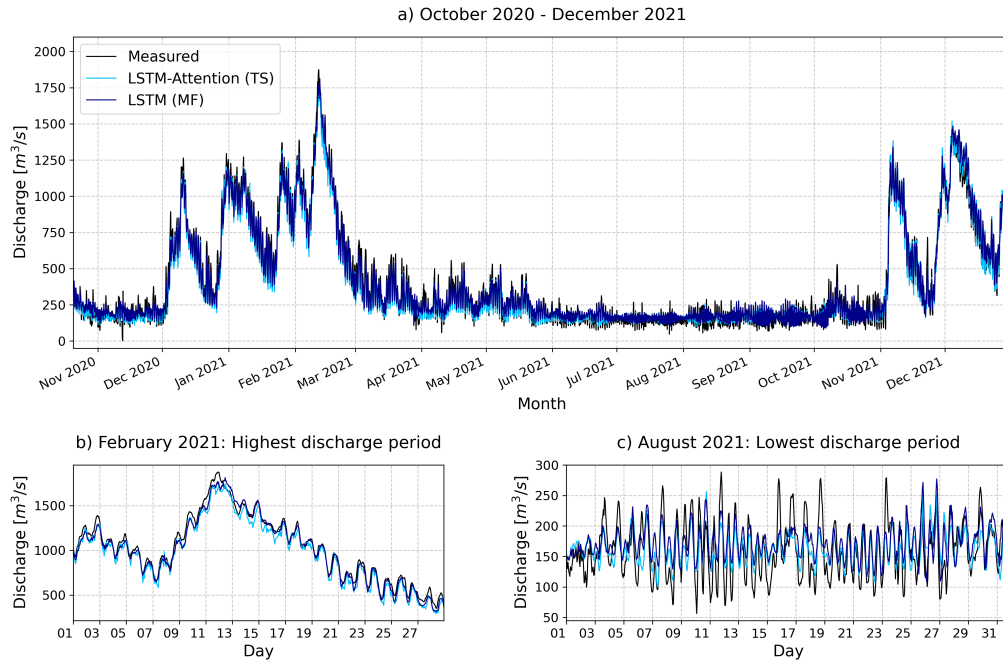
Figure 6.17: Time-series versus mode functions: MAE metric and frequency distribution of discharge categories

conditions, LSTM-Attention follows the original signal less accurately than its VMD-LSTM variant, while the opposite was observed when LSTM was trained on time-series data. This suggests that by decomposing the water level signal with VMD, even a simpler model like LSTM was able to accurately retain the predominant tidal periodicities in the model output, while minimizing the impact of short-term high-frequency variations. This modeling approach aligns with the hydrodynamic processes typical for tidal reaches.

In order to gain a deeper understanding of the model's performance, a Butterworth filter was applied to separate the discharge signal into interdaily variations, which are driven by river discharge, and intra-daily oscillations, which are driven by tides. Visualizations of the utilized filter are presented in Appendix D.4, in Figures D.14 and D.15.

Exploring the case of interdaily variations, LSTM-Attention effectively estimates these values, demonstrating satisfying performance with the time-series data. However, its performance declines when VMD decomposition is utilized. The opposite is valid for the LSTM model, where VMD decomposition has considerably improved its accuracy. Although it achieves the highest accuracy, it does not noticeably differ from the LSTM-Attention based on time-series.

In the context of intra-daily oscillations, the use of VMD decomposition has shown consistent improvements in both LSTM and LSTM-Attention, with a more pronounced effect on the LSTM model. LSTM-Attention struggles more with such rapid changes,

Figure 6.18: Time-series versus mode functions: Two best performing models

and, hence, making it more suitable for scenarios that involve interdaily variations, or only the use of raw time-series data.

The introduction of VMD decomposition showed both improvements, when it comes to the LSTM model, and stagnation, when it comes to the performance of the LSTM-Attention model. By comparing the models with different input combinations, we observed that the LSTM with mode functions provides the most consistent results for the entire range of discharge values at the Metković location. Such results can be attributed to the reduced spectral complexity of the data. Traditional time-series data contain complex and hidden patterns that are easy to detect and correctly modeled using the attention mechanism. This was also seen by comparing previously modeled LSTM and LSTM-Attention models that use time-series as input. However, as VMD separates the signals of each water level station, the signals become simpler and localized, and here, the previous considerable advantage of the LSTM-Attention model is reduced, and at the same time, the additional complexity of the model in this case is a topic for discussion.

6.3. Prediction: In-depth Analysis for HEC-RAS Data

Two studies were conducted using a hybrid approach, which combined a CNN with an LSTM model, for different input feature combinations. The first study was conducted for the purpose of discharge forecasting, where the use of spectrograms was compared to the time-series dataset approach. The study considered only water level data from multiple stations and discharge as the main input features. The second study addressed the problem of water level forecasting. It was even further extended by employing a feature engineering approach, which included additional input features. The following subsections present the study results and accompanying discussions.

6.3.1. Discharge Forecasting

Discharge forecasting was performed on three different datasets: (1) time-series, (2) spectrograms, and lastly, (3) a combination of time-series data and spectrograms. As defined in the Methodology section, the models were optimized using a five-fold grid-search cross-validation, whose ranges and optimal values are presented in Table E.1 of Appendix E.1. The proposed hybrid model consisted of two CNNs, a max-pooling layer, a single LSTM, and a fully connected layer. For this specific architecture, a modification of the Adam optimization algorithm, known as AdamW, was used. This optimization algorithm allows combining the benefits of the adaptive learning rate with decoupled weight decay regularization. The selected batch size was 256. In the conducted study, it was observed that the optimal values of output channels in CNN layers do not differ considerably for different input datasets. This suggests that the CNN part of the model has a similar capacity for local feature extraction, although it is implemented in a different way depending on the given type of input dataset. At the same time, the learning rate and LSTM hidden units were equal for all different input dataset combinations. The dropout rate was higher for time-series data than for spectrogram-based combinations, which can be partly attributed to the higher variance of the original raw time-series data compared to the other spectrogram data.

The evaluation metrics give us a general insight into the performance of the model

depending on the lead time and the used dataset on which the hybrid model was trained. The use of the time-frequency domain as a stand-alone approach or in combination with time-series has proven to be useful for a tidal river area characterized by a dynamic flow regime. For up to one hour ahead predictions, the smallest difference in evaluation metrics was observed for datasets containing spectrograms and a combination of spectrograms and time-series. The observed trend continues up to the longest forecasting horizon of 12 hours. The largest oscillations in performance were observed for the time-series data, which, for example, for one hour in advance, gave the largest differences compared to other input datasets. Its RMSE was lower by about 30-36%, MAPE by about 38-62%, and NSE by about 6-7%. Nonetheless, all models exhibited a notable decrease in performance as the lead time increased. The difference in performance for a lead time of 12 hours is between 4-6% for RMSE, 6-12% for MAPE, and 0.7-1% for NSE. This indicates that as the horizon increases, the difference in the performance of the better models has decreased by approximately 7 times for RMSE, 6 times for MAPE, and 2 times for NSE. Although it is worth noting that the largest drop in model performance occurs for forecasts between 3 and 6 hours in advance. Beyond this timeframe, performance metrics were still quite low, but the decrease in percentage was less pronounced.

Table 6.4: Hybrid model performance on the test dataset using three different input combinations

Input dataset	Lead time (h)	RMSE (m^3/s)	MAPE (%)	NSE
Time-series	1	27.670	8.652	0.988
	3	41.567	11.112	0.972
	6	59.229	14.013	0.943
	12	72.531	16.172	0.915
Spectrogram	1	**17.696**	**3.278**	**0.995**
	3	**32.413**	8.781	0.973
	6	**54.406**	**11.766**	**0.952**
	12	**68.532**	**14.187**	**0.924**
Spectrogram and time-series	1	19.309	5.293	0.994
	3	35.355	**8.743**	**0.980**
	6	56.603	13.223	0.948
	12	69.523	15.051	0.922

Figures 6.19 and 6.20 present a comparison of predicted and observed data across different lead times, ranging from one hour up to 12 hours. The best possible alignment can be observed in the one-hour scenario, where the data points are closely clustered. This indicates that the prediction achieved a satisfactory accuracy, with R^2 values ranging from 0.988 to 0.995. The last visualizations for the 12-hour ahead scenario revealed notable

deviations, particularly regarding high flow values, with the minimum R^2 of 0.915.
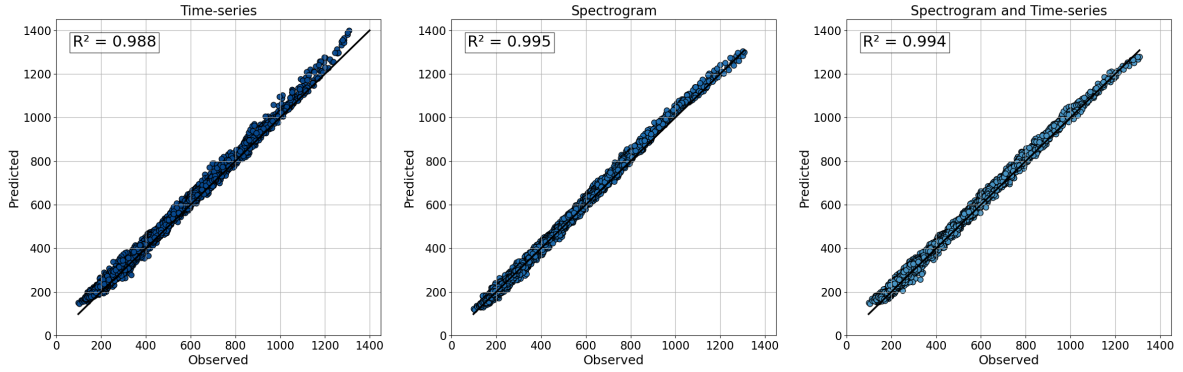

Figure 6.19: Predicted versus observed: 1-hour ahead
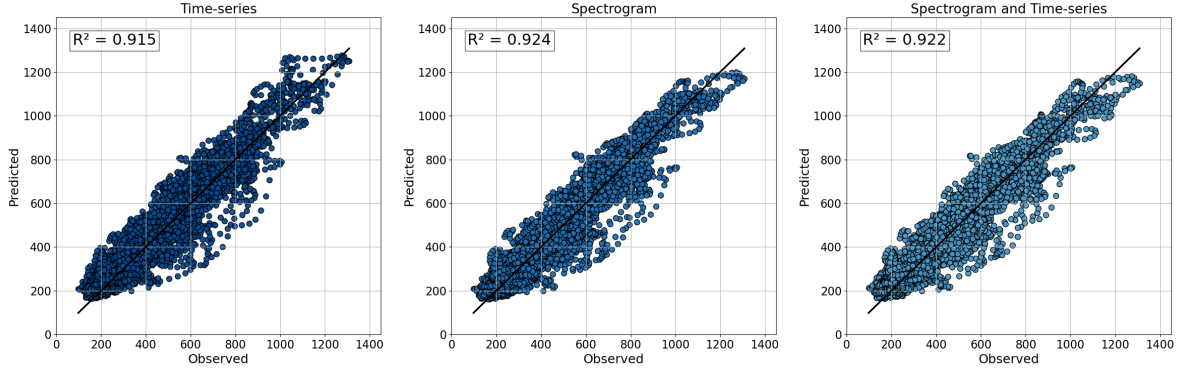
Figure 6.20: Predicted versus observed: 12-hours ahead

Figure 6.21 presents an additional visualization with an MAE metric, which was calculated according to the classification of discharge values. Three groups of discharge values were defined: low flows, which contain values below 300 m^3/s; medium or average discharge, whose values are in the range of 300 to 1050 m^3/s; and extreme discharges, whose values exceed 1050 m^3/s. The most represented categories are low and medium, with percentages of around 40-50%, while less than 5% of the values represented high discharge. Low discharge values were captured with the highest accuracy when spectrograms and the combination of spectrograms and time-series were used. However, with the extension of the prediction horizon, the performance of the time-series model approaches theirs. The same is also true for average discharge values, while, for example, the only notable difference was observed for high discharge, where the datasets containing the spectrogram exhibited a lower level of performance than the time-series.
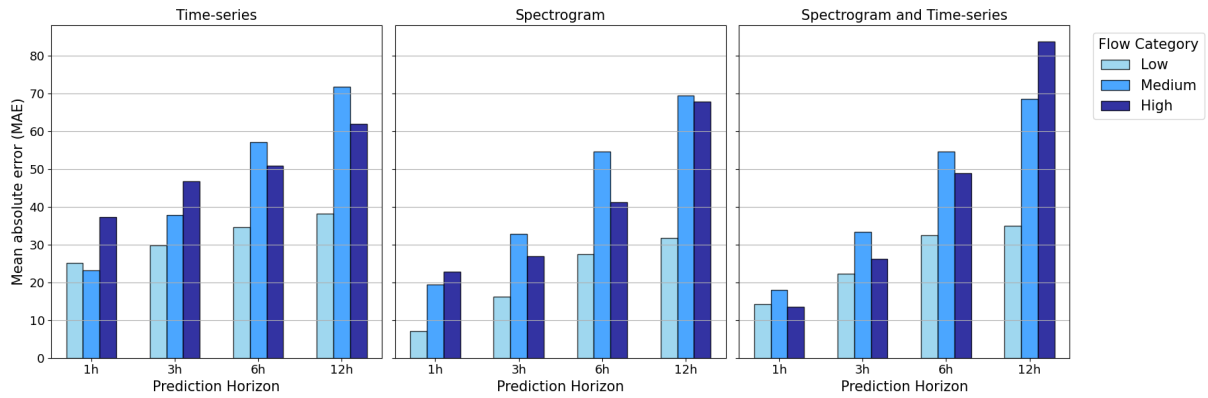
Figure 6.21: Evaluation of utilized datasets for the hybrid approach using MAE for various discharge categories

6.3.2. Water Level Forecasting

The second study focused on the application of the proposed hybrid model, but this time aimed at forecasting the water level of the station situated in the uppermost section of the tidal river, Metković. The architecture of the hybrid model remained consistent with the previous study, and had also been optimized under equal settings (ranges and optimal values are presented in Table E.2 of Appendix E.1. This study stands out from the previous one by broadening the dataset through a feature engineering approach. Alongside water level and discharge, a range of time-based features (hour, part of the day, day, week, month, season, year) and variables, including daily values, daily mean values, and a 24-hour rolling mean, were used as input data. Separate tests were conducted on four distinct approaches. Two of these approaches utilized a single type of input data, specifically time-series and spectrograms. The other two approaches integrated the previous two methods with separate feature engineering techniques. An examination of the relevance of individual variables was conducted utilizing mutual information, with the findings illustrated in Figure 6.22. The lowest impact was observed for almost all time-based features, except for month and summer, which is theoretically justified due to the monthly fluctuations in water levels influenced by weather conditions; for instance, during the summer months, the water levels are at their lowest.

Table 6.5 presents evaluation metrics of the test dataset for different combinations of composite features. As the lead time increases from 1- to 24 hours ahead, a decrease in performance is observed, similar to the previous study. The largest differences in
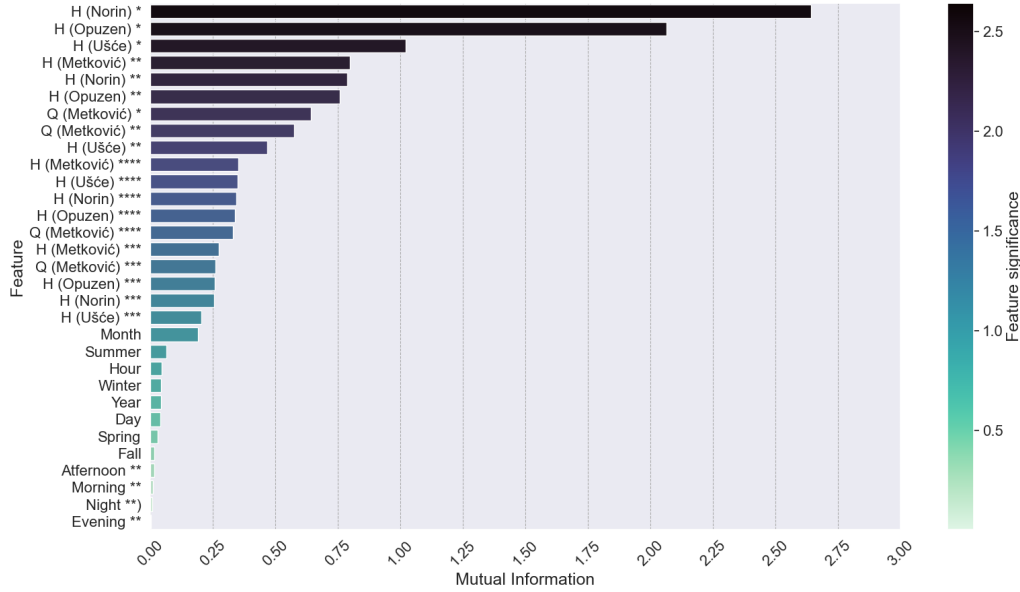
Figure 6.22: Extended dataset with feature engineering analyzed using mutual information (* denotes hourly scale, ** daily scale, *** daily mean, and **** rolling mean of 24-hours)

performance are observed for the 1-hour ahead scenario, where the order of the best-performing combination was time-series, spectrogram, time-series with feature engineering, and lastly, spectrogram with feature engineering, although the first two combinations have the slightest differences in performance. The percentage differences between the two best approaches are about 27% in terms of RMSE, around 15% in terms of MAE, and 0.3% in terms of NSE. Therefore, for the shortest lead time, feature engineering does not provide additional information to improve prediction performance; thus, stand-alone combinations are sufficient for accurately estimating rapid changes. However, this changes for the remaining lead times, as a slight difference in performance order was observed. Time-series and time-series with a spectrogram have switched places, meaning the order regarding the performance is as follows: time-series with feature engineering spectrogram, time-series, and lastly, spectrogram with feature engineering. Throughout all lead times, spectrogram-based features have shown consistency in performance. Similarly, the stand-alone combinations exhibit comparable performance, showing an average difference in longer lead times, with RMSE under 2%, MAE below 5%, and NSE less than 0.2%. Feature engineering, when combined with time-series, showed great potential for longer

lead times. However, when utilized alongside spectrogram data, it resulted in the lowest level of performance as the spectrogram by itself is sufficiently representative (in both the time and frequency domain).

Table 6.5: Hybrid model performance on the test dataset using four different input combinations

Composite features	Lead time (h)	RMSE (m^3/s)	MAE (m^3/s)	NSE
	1	**0.026**	**0.021**	**0.994**
Time-series	6	0.081	**0.052**	**0.946**
	12	0.096	0.069	0.925
	24	0.100	0.070	0.919
	1	0.055	0.040	0.975
Time-series with Feature Engineering	6	**0.073**	**0.052**	0.957
	12	**0.080**	**0.057**	**0.947**
	24	**0.093**	**0.064**	**0.928**
	1	0.033	0.024	0.991
Spectrograms	6	0.078	0.056	0.950
	12	0.095	0.067	0.926
	24	0.100	0.069	0.919
	1	0.081	0.060	0.946
Spectrograms with Feature Engineering	6	0.121	0.092	0.880
	12	0.112	0.081	0.896
	24	0.115	0.080	0.892

The previously discussed results are further supported by the presented visualizations in Figures 6.23 and 6.24. While in the first hour of forecasting, the clustering of points around the best-fit line is most satisfying for the time-series and spectrogram. Within the first hour, a combination of spectrogram analysis and feature engineering demonstrated inadequate performance, exhibiting a noticeable tendency to underpredict water level values. Consequently, this combination systematically underestimates water level values above 1.5 m, a pattern that persists and becomes even more pronounced across prolonged prediction horizons. Longer prediction horizons are considered to be more important, primarily for the implementation of early warning systems, facilitating prompt risk mitigation and strategic planning. Consequently, the optimal hybrid model, developed using a time-series dataset augmented with feature engineering, is considered the most effective solution for this particular issue.
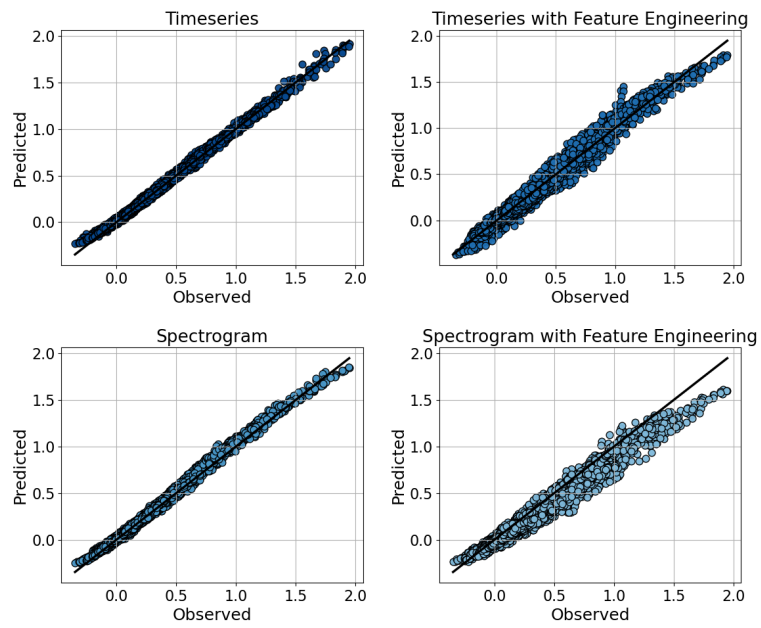
Figure 6.23: Simulated HEC-RAS data: Predicted versus observed for 1-hour ahead forecasting
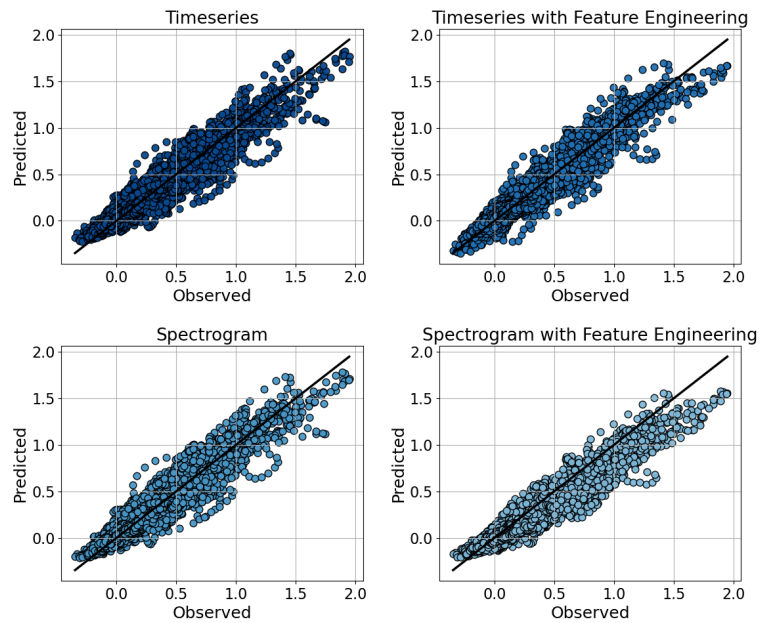
Figure 6.24: Simulated HEC-RAS data: Predicted versus observed for 24-hour ahead forecasting

Chapter 7

CONCLUSION

7.1. Summary of Contributions

In this thesis, two research aims were successfully achieved. The first objective was to develop a robust hybrid machine learning approach for estimating and forecasting river discharges and water levels across diverse flow conditions that range from low flows during the summer (between 50 and 350 m^3/s) to high flows during the winter (over 1050 m^3/s). Estimation of discharge had been successfully conducted using the proposed hybrid LSTM-Attention machine learning model on both measured and simulated data for the area of the microtidal river, Neretva. This was the first comprehensive analysis conducted for a tidal reach, aiming to provide a robust machine learning approach effective for both real-world and ideal scenarios. The results of the proposed approach were compared to those of the baseline time-series approach, LSTM, and various non-temporal machine learning approaches, DT, RF, SVR, LGBM, and XGB. Evaluation metrics and visual inspection have shown the highest performance improvement and estimation accuracy for the LSTM-Attention model for both considered datasets. The model achieved high estimation accuracy across all flow conditions, with an NSE of 0.993 for the simulated dataset and 0.974 for the measured dataset, demonstrating its ability to effectively identify critical features.

Likewise, the study demonstrated high accuracy by using only water level data from multiple locations in the tidal reach, both upstream and downstream, as model inputs. Hence, the LSTM-Attention estimation ability was unaffected by the lack of meteorologi-

cal data. However, such a scenario can only be considered and applied if the length of the tidal reach is limited, as in this case. Forecasting using a machine learning-based approach, CNN-LSTM, revealed that integrating the spectrogram into the discharge prediction resulted in improvements in accuracy. Additionally, by integrating feature engineering in water level prediction, the performance had improved for time-series data, particularly for the longest horizon. This work on estimation and forecasting contributes by developing two machine learning-based approaches. Improving discharge estimation accuracy and outperforming other machine learning approaches was achieved by integrating the attention mechanism within a hybrid model. Accurate forecasting of water level and discharge at the upstream discharge station, Metković, was achieved by implementing a hybrid CNN-LSTM model.

The second research aim, focusing on gaining insights into physical processes of tidal rivers and estuaries and thereby enhance estimation accuracy, was also achieved. The VMD decomposition technique was utilized for signal processing, and the resulting mode functions were classified into different categories of tidal components. These features were then used as inputs for the best-performing models of the prior aim, namely the LSTM-Attention and LSTM models. The results showed a significant increase in performance for the LSTM model that incorporated mode functions and the residual, in contrast to its variant based solely on time-series data. Nonetheless, a minor deviation from previous research on discharge estimation was observed regarding the LSTM-Attention model, whose performance remained almost unchanged, with an insignificant decline in performance. As a result, another contribution was achieved by decomposing water level signals into distinct frequency bands, leading to improved LSTM estimation accuracy for river discharge.

The research has several practical applications. Regarding water resource management, allocating sufficient water resources for irrigation purposes, enhancing flood forecasting abilities with greater precision of early warning systems by integrating such models, also for calibration of hydrological models, and improving flood risk assessment. Another possible application of this approach is in ecosystem management, by ensuring the suitability of water conditions, evaluating the impact of human activities, and even climate change in the designated area. Also, these approaches can be applied on a more global scale, specifically to other tidal rivers and estuaries, with better accuracy if less

pronounced stratification is present, for example, in well-mixed or partially-mixed estuaries. Overall, the conducted research can contribute to a more sustainable and efficient management of tidal rivers and estuary areas.

7.2. Future Work

With the increase in the number of studies in the field of hydrology and the increasing use of machine learning approaches, there is a significant variety of approaches which have not been tested so far for areas of tidal rivers and estuaries, which, unlike inland rivers, pose greater challenges for modeling due to the complex nature of tidal flow dynamics. Based on the reviewed literature and the approaches tested in this thesis, several suggestions for potential future research directions are outlined below.

Firstly, the application of the Physics Informed Neural Network machine learning technique could address potential issues of data scarcity. This methodology approximates PDE solutions by transforming the task of directly solving the governing equations into an optimization problem using a loss function. This ensures that estimations or predictions do not rely entirely on observational data, but are also physically consistent. Secondly, the use of alternative time-frequency distributions, including the quadratic class of time-frequency distributions, to better address the residual component that contains diurnal and semidiurnal patterns. Thirdly, employing an adaptive window length on time-series or for decomposed tidal components of water level signals may ensure that relevant temporal information would be adequately captured, covering both short- and long-term patterns. Additionally, considering different attention architectures could potentially result in a more effective solution for modeling decomposed tidal component signals, while also integrating additional hydrological parameters such as salinity and temperature.

BIBLIOGRAPHY

[1] Adib, A. (2008). Determining water surface elevation in tidal rivers by ann. In *Proceedings of the Institution of Civil Engineers-Water Management*, volume 161, pages 83–88. Thomas Telford Ltd.

[2] Ahmed, D., Hassan, M., and Mstafa, R. (2022). A Review on Deep Sequential Models for Forecasting Time Series Data. *Applied Computational Intelligence and Soft Computing*, 2022.

[3] Ahmed, F. I., Islam, M. R., Hossan, M. S., Rasel, R. I., and Sultana, N. (2017). River Tide Level Prediction: A Data Mining Approach for Hydrographie Time Series Data Analysis. In *2017 20th International Conference of Computer and Information Technology (ICCIT)*, pages 1–6. IEEE.

[4] Ajmera, T. K. and Goyal, M. K. (2012). Development of stage-discharge rating curve using model tree and neural networks: an application to Peachtree Creek in Atlanta. *Expert Systems with Applications*, 39(5):5702–5710.

[5] Almeida, J. (2002). Predictive non-linear modeling of complex data by artificial neural networks. *Current Opinion in Biotechnology*, 13:72–76.

[6] Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., and Farhan, L. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data*, 8:1–74.

[7] Antonini, A. S., Tanzola, J., Asiain, L., Ferracutti, G. R., Castro, S. M., Bjerg, E. A., and Ganuza, M. L. (2024). Machine Learning model interpretability using SHAP values:

Application to Igneous Rock Classification task. *Applied Computing and Geosciences*, 23:100178.

[8] Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R., et al. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information fusion*, 58:82–115.

[9] Aydoğan, B. and Aytekin, T. (2025). An in-depth analysis of kernelshap and samplingshap: assessing robustness, error, and efficiency: B. aydoğan, t. aytekin. *Knowledge and Information Systems*, pages 1–35.

[10] Balic, D. and Malvic, T. (2013). Pliocene-Quaternary stratigraphy and sedimentation at the Neretva River Mouth, on Croatian Adriatic Coast. *Geological Quarterly*, 57(2):233–241.

[11] Bendat, J. S. and Piersol, A. G. (2011). *Random data: analysis and measurement procedures*. John Wiley & Sons.

[12] Bhar, K. K. and Bakshi, S. (2020). Application of artificial neural network for predicting water levels in Hooghly estuary, India. *H2Open Journal*, 3(1):401–415.

[13] Bhattacharya, B. and Solomatine, D. P. (2005). Neural networks and M5 model trees in modelling water level-discharge relationship. *Neurocomputing*, 63:381–396.

[14] Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.

[15] Chang, F.-J. and Chen, Y.-C. (2003). Estuary water-stage forecasting by using radial basis function neural network. *Journal of Hydrology*, 270(1-2):158–166.

[16] Chen, C., Zhang, C., Tian, B., Wu, W., and Zhou, Y. (2023a). Mapping intertidal topographic changes in a highly turbid estuary using dense Sentinel-2 time series with deep learning. *ISPRS Journal of Photogrammetry and Remote Sensing*, 205:1–16.

[17] Chen, K., Kuang, C., Wang, L., Chen, K., Han, X., and Fan, J. (2021). Storm surge prediction based on long short-term memory neural network in the East China Sea. *Applied Sciences*, 12(1):181.

[18] Chen, W.-B., Liu, W.-C., and Hsu, M.-H. (2012a). Comparison of ANN approach with 2D and 3D hydrodynamic models for simulating estuary water stage. *Advances in Engineering Software*, 45(1):69–79.

[19] Chen, W.-B., Liu, W.-C., and Hsu, M.-H. (2012b). Predicting typhoon-induced storm surge tide with a two-dimensional hydrodynamic model and artificial neural network model. *Natural Hazards and Earth System Sciences*, 12(12):3799–3809.

[20] Chen, Y., Gan, M., Pan, S., Pan, H., Zhu, X., and Tao, Z. (2020). Application of auto-regressive (AR) analysis to improve short-term prediction of water levels in the Yangtze estuary. *Journal of Hydrology*, 590:125386.

[21] Chen, Y., Huang, M., Song, K., and Wang, T. (2023b). Prediction of Ship Traffic Flow and Congestion Based on Extreme Learning Machine with Whale Optimization Algorithm and Fuzzy c-Means Clustering. *Journal of Advanced Transportation*, 2023(1):7175863.

[22] Chen, Y.-C., Yeh, H.-C., Kao, S.-P., Wei, C., and Su, P.-Y. (2023c). Water level forecasting in tidal rivers during typhoon periods through ensemble empirical mode decomposition. *Hydrology*, 10(2):47.

[23] Chen, Z., Zong, Y., Wu, Z., Kuang, Z., and Wang, S. (2024). Prediction of discharge in a tidal river using the LSTM-based sequence-to-sequence models. *Acta Oceanologica Sinica*, 43(7):40–51.

[24] Chinh, L., Hiramatsu, K., Harada, M., and Mori, M. (2009). Estimation of water levels in a main drainage canal in a flat low-lying agricultural area using artificial neural network models. *Agricultural Water Management*, 96(9):1332–1338.

[25] Conover, W. J. (1999). *Practical nonparametric statistics*. john wiley & sons.

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

[27] Costa, V. (2017). Correlation and regression. *Fundamentals of statistical hydrology*, pages 391–440.

[28] Cremer, C. J. M., Mariegaard, J. S., and Andersson, H. J. (2025). A hybrid data assimilation and machine learning approach for enhancing operational forecasting in 2D hydrodynamic models. *Journal of Hydroinformatics*, 27(3):493–507.

[29] Cudaback, C. N. and Jay, D. A. (2000). Tidal asymmetry in an estuarine pycnocline: Depth and thickness. *Journal of Geophysical Research: Oceans*, 105(C11):26237–26251.

[30] Dato, J. F., Dinápoli, M. G., DOnofrio, E. E., and Simionato, C. G. (2024). On water level forecasting using artificial neural networks: the case of the Río de la Plata Estuary, Argentina. *Natural Hazards*, 120(11):9753–9776.

[31] Dragomiretskiy, K. and Zosso, D. (2013). Variational mode decomposition. *IEEE transactions on signal processing*, 62(3):531–544.

[32] Fei, K., Du, H., and Gao, L. (2023). Accurate water level predictions in a tidal reach: Integration of Physics-based and Machine learning approaches. *Journal of Hydrology*, 622.

[33] Friedman, J. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29:1189–1232.

[34] Gabor, D. (1946). Theory of communication. *The Journal of the Institution of Electrical Engineers - Part III: Radio and Communication Engineering*, 93(26):429–457.

[35] Gajić-Čapka, M., Gttler, I., Cindrić, K., and Branković, Č. (2018). Observed and simulated climate and climate change in the lower Neretva river basin. *Journal of Water and Climate Change*, 9(1):124–136.

[36] Gan, M., Chen, Y., Pan, S., Lai, X., Pan, H., Wen, Y., and Xia, M. (2024). An improved machine learning-based model to predict estuarine water levels. *Ocean Modelling*, page 102376.

[37] Gan, M., Pan, S., Chen, Y., Cheng, C., Pan, H., and 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(5):496.

[38] Gao, S., Feng, X., Xu, H., Wu, Y., and Feng, W. (2025). A hybrid deep learning model based on emd algorithm for non-stationary water level prediction of estuarine systems. *Estuarine, Coastal and Shelf Science*, 314:109128.

[39] Garel, E. and Dalimonte, D. (2017). Continuous river discharge monitoring with bottom-mounted current profilers at narrow tidal estuaries. *Continental Shelf Research*, 133:1–12.

[40] Ghorbani, M. A., Khatibi, R., Goel, A., FazeliFard, M. H., and Azani, A. (2016). Modeling river discharge time series using support vector machine and artificial neural networks. *Environmental Earth Sciences*, 75(8):1–13.

[41] Gill, M. K., Asefa, T., Kaheil, Y., and McKee, M. (2007). Effect of missing data on performance of learning algorithms for hydrologic predictions: Implications to an imputation technique. *Water Resources Research*, 43(7).

[42] Glassner, A. (2021). *Deep learning: a visual approach.* No Starch Press.

[43] Graf, R., Kolerski, T., and Zhu, S. (2022). Predicting Ice Phenomena in a River Using the Artificial Neural Network and Extreme Gradient Boosting. *Resources*, 11:12.

[44] Gu, Z., Cao, X., Liu, G., and Lu, W. (2014). Optimizing operation rules of sluices in river networks based on knowledge-driven and data-driven mechanism. *Water Resources Management*, 28(11):3455–3469.

[45] Guillou, N. and Chapalain, G. (2021). Machine learning methods applied to sea level predictions in the upper part of a tidal estuary. *Oceanologia*, 63(4):531–544.

[46] Guillou, N., Chapalain, G., and Petton, S. (2023). Predicting sea surface salinity in a tidal estuary with machine learning. *Oceanologia*, 65:318–332.

[47] Guo, W.-D., Chen, W.-B., Yeh, S.-H., Chang, C.-H., and Chen, H. (2021a). Prediction of River Stage Using Multistep-Ahead Machine Learning Techniques for a Tidal River of Taiwan. *Water*, 13(7):920.

[48] Guo, W.-D., Chen, W.-B., Yeh, S.-H., Chang, C.-H., and Chen, H. (2021b). Prediction of river stage using multistep-ahead machine learning techniques for a tidal river of Taiwan. *Water*, 13(7):920.

[49] Gupta, H., Kling, H., Yilmaz, K., and Martinez, G. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, 377:80–91.

[50] Habib, E. H. and Meselhe, E. A. (2006). Stage-Discharge Relations for Low-Gradient Tidal Streams Using Data-Driven Models. *Journal of Hydraulic Engineering*, 132(5):482–492.

[51] Hamzah, F. B., Mohd Hamzah, F., Mohd Razali, S. F., Jaafar, O., and Abdul Jamil, N. (2020). Imputation methods for recovering streamflow observation: A methodological review. *Cogent Environmental Science*, 6(1):1745133.

[52] Hannan, A. and 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.

[53] He, X., Luo, J., Zuo, G., and Xie, J. (2019). Daily runoff forecasting using a hybrid model based on variational mode decomposition and deep neural networks. *Water Resources Management*, 33:1571–1590.

[54] He, X., Shi, S., Geng, X., and Xu, L. (2022). Hierarchical attention-based context-aware network for red tide forecasting. *Applied Soft Computing*, 127:109337.

[55] Hidayat, H., Hoitink, A., Sassi, M., and Torfs, P. (2014). Prediction of Discharge in a Tidal River Using Artificial Neural Networks. *Journal of Hydrologic Engineering*, 19(8):04014006.

[56] Hochreiter, S. and Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9:1735–1780.

[57] Hoitink, A., Buschman, F., and Vermeulen, B. (2009). Continuous measurements of discharge from a horizontal acoustic Doppler current profiler in a tidal river. *Water Resources Research*, 45(11).

[58] Hoitink, A. F. and Jay, D. A. (2016). Tidal river dynamics: Implications for deltas. *Reviews of Geophysics*, 54(1):240–272.

[59] Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, 79(8):2554–2558.

[60] Huang, Y., Pan, J., and 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.

[61] Ibrahim, K. S. M. H., Huang, Y. F., Ahmed, A. N., Koo, C. H., and El-Shafie, A. (2022). A review of the hybrid artificial intelligence and optimization modelling of hydrological streamflow forecasting. *Alexandria Engineering Journal*.

[62] Janbain, I., Jardani, A., Deloffre, J., and Massei, N. (2023). Deep Learning Approaches for Numerical Modeling and Historical Reconstruction of Water Quality Parameters in Lower Seine. *Water*, 15(9):1773.

[63] Jones, A. E., Hardison, A. K., Hodges, B. R., McClelland, J. W., and Moffett, K. B. (2019). An expanded rating curve model to estimate river discharge during tidal influences across the progressive-mixed-standing wave systems. *Plos one*, 14(12):e0225758.

[64] Jung, S., Cho, H., Kim, J., and Lee, G. (2018). Prediction of water level in a tidal river using a deep-learning based LSTM model. *Journal of Korea Water Resources Association*, 51(12):1207–1216.

[65] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., and Liu, T.-Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. In *31st Conference on Neural Information Processing Systems (NIPS 2017)*, volume 2017-Decem, pages 3147–3155.

[66] Krvavica, N., Gotovac, H., and Lončar, G. (2021a). Salt-wedge dynamics in microtidal Neretva River estuary. *Regional Studies in Marine Science*, 43:101713.

[67] Krvavica, N., Gržić, M. M., Innocenti, S., and Matte, P. (2024). Impact of Storm Surge and Power Peaking on Tidal-Fluvial Dynamics in Microtidal Neretva River Estuary. *arXiv preprint arXiv:2411.13391*.

[68] Krvavica, N., Kožar, I., Travaš, V., and Ožanić, N. (2017a). 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(1):49–59.

[69] Krvavica, N., Lončar, G., Oskoruš, D., and Ružić, I. (2021b). A contribution to improving the system of transitional waters hydrological measurements: Hydraulic and spectral analyses of the Neretva River flow rate. *Hrvatske vode*, 29(118):255–274.

[70] Krvavica, N. and 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.

[71] Krvavica, N., Travaš, V., and Ožanić, N. (2017b). Salt-wedge response to variable river flow and sea-level rise in the Microtidal Rječina River Estuary, Croatia. *Journal of Coastal Research*, 33(4):802–814.

[72] Lauer, F. and Kösters, F. (2024). Using statistical and machine learning approaches to describe estuarine tidal dynamics. *Journal of Hydroinformatics*, 26(4):853–868.

[73] LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. (2002). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324.

[74] Li, B., Yang, G., Wan, R., Dai, X., and 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.

[75] Li, Y., Zhao, X., Wang, Y., and Zeng, L. (2025). Deep characteristic learning model for real-time flow monitoring based on H-ADCP. *Journal of Hydrology: Regional Studies*, 57:102115.

[76] Liang, B.-X., Hu, J.-P., Liu, C., and Hong, B. (2021). Data pre-processing and artificial neural networks for tidal level prediction at the Pearl River Estuary. *Journal of Hydroinformatics*, 23(2):368–382.

[77] Lipton, Z. C., Berkowitz, J., and Elkan, C. (2015). A critical review of recurrent neural networks for sequence learning. *arXiv preprint arXiv:1506.00019*.

[78] Liu, W.-C. and Chung, C.-E. (2014). Enhancing the predicting accuracy of the water stage using a physical-based model and an artificial neural network-genetic algorithm in a river system. *Water*, 6(6):1642–1661.

[79] Ljubenkov, I., Kvesić, D., and Erceg, J. (2024). Delta flood risk analysis: case study from the Neretva River (Croatia). *Estuarine Management and Technologies*, 1:69–93.

[80] Lopes, A. M. and Machado, J. A. T. (2017). Tidal Analysis Using Time–Frequency Signal Processing and Information Clustering. *Entropy*, 19(8):390.

[81] Lovrinović, I., Srzić, V., and 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.

[82] Lundberg, S. M. and Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30.

[83] Ma, J., Jia, C., Yang, X., Cheng, X., Li, W., and Zhang, C. (2020). A data-driven approach for collision risk early warning in vessel encounter situations using attention-BiLSTM. *IEEE access*, 8:188771–188783.

[84] Maier, H. R., Taghikhah, F. R., Nabavi, E., Razavi, S., Gupta, H., Wu, W., Radford, D. A., and Huang, J. (2024). How much X is in XAI: Responsible use of Explainable artificial intelligence in hydrology and water resources. *Journal of Hydrology X*, page 100185.

[85] Malek, N., Yaacob, W., Nasir, S., and Shaadan, N. (2022). Prediction of Water Quality Classification of the Kelantan River Basin, Malaysia, Using Machine Learning Techniques. *Water (Switzerland)*, 14:1067.

[86] Marusteri, M. and Bacarea, V. (2010). Comparing groups for statistical differences: how to choose the right statistical test? *Biochemia Medica*, 20(1):15–32.

[87] Matte, P., Jay, D. A., and Zaron, E. D. (2013). Adaptation of classical tidal harmonic analysis to nonstationary tides, with application to river tides. *Journal of Atmospheric and Oceanic Technology*, 30(3):569–589.

[88] Mihel, A. M., Krvavica, N., and Lerga, J. (2025). Regression-based machine learning approaches for estimating discharge from water levels in microtidal rivers. *Journal of Hydrology*, 646:132276.

[89] Mihel, A. M., Lenac, K., Krvavica, N., and Lerga, J. (2023). Discharge forecasting in coastal rivers using cnn-lstm hybrid approach. In *2023 International Symposium ELMAR*, pages 55–60.

[90] Mihel, A. M., Lerga, J., and Krvavica, N. (2024a). Estimating water levels and discharges in tidal rivers and estuaries: Review of machine learning approaches. *Environmental Modelling & Software*, 176:106033.

[91] Mihel, A. M., Pečnik, Š., Vrbančič, G., Lerga, J., and Krvavica, N. (2024b). Employing Feature Engineering for River Stage Forecasting to Improve Hybrid Model Performance. In *2024 International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*, pages 1–6. IEEE.

[92] Mohri, M., Rostamizadeh, A., and Talwalkar, A. (2018). *Foundations of machine learning*. MIT press.

[93] Nayebi, A., Tipirneni, S., Reddy, C. K., Foreman, B., and Subbian, V. (2023). WindowSHAP: An efficient framework for explaining time-series classifiers based on Shapley values. *Journal of Biomedical Informatics*, 144:104438.

[94] Ng, K., Huang, Y., Koo, C., Chong, K., El-Shafie, A., and Najah Ahmed, A. (2023). A review of hybrid deep learning applications for streamflow forecasting. *Journal of Hydrology*, 625:130141.

[95] Oprisan, S. A., Clementsmith, X., Tompa, T., and Lavin, A. (2023). Empirical mode decomposition of local field potential data from optogenetic experiments. *Frontiers in Computational Neuroscience*, 17:1223879.

[96] Pasupa, K. and Jungjareantrat, S. (2016). Water Levels Forecast In Thailand: A Case Study Of Chao Phraya River. In *2016 14th International Conference on Control, Automation, Robotics and Vision (ICARCV)*, pages 1–6. IEEE.

[97] Pierini, J. O., Lovallo, M., Telesca, L., and Gómez, E. A. (2013). Investigating prediction performance of an artificial neural network and a numerical model of the tidal signal at Puerto Belgrano, Bahia Blanca Estuary (Argentina). *Acta Geophysica*, 61(6):1522–1537.

[98] Piraei, R., Niazkar, M., Afzali, S., and Menapace, A. (2023). Application of Machine Learning Models to Bridge Afflux Estimation. *Water (Switzerland)*, 15:2187.

[99] Piras, D., Peiris, H. V., Pontzen, A., Lucie-Smith, L., Guo, N., and Nord, B. (2023). A robust estimator of mutual information for deep learning interpretability. *Machine Learning: Science and Technology*, 4(2):025006.

[100] Rice, T. K., Schork, N. J., and Rao, D. (2008). Methods for handling multiple testing. *Advances in genetics*, 60:293–308.

[101] Rodriguez-Galiano, V., Sanchez-Castillo, M., Chica-Olmo, M., and Chica-Rivas, M. (2015). Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. *Ore geology reviews*, 71:804–818.

[102] Saleem, R., Yuan, B., Kurugollu, F., Anjum, A., and Liu, L. (2022). Explaining deep neural networks: A survey on the global interpretation methods. *Neurocomputing*, 513:165–180.

[103] Salehinejad, H., Sankar, S., Barfett, J., Colak, E., and Valaee, S. (2017). Recent advances in recurrent neural networks. *arXiv preprint arXiv:1801.01078*.

[104] Sampurno, J., Vallaeys, V., Ardianto, R., and Hanert, E. (2022). Integrated hydrodynamic and machine learning models for compound flooding prediction in a data-scarce estuarine delta. *Nonlinear Processes in Geophysics*, 29(3):301–315.

[105] Sassi, M., Hoitink, A., and Vermeulen, B. (2011). Discharge estimation from H-ADCP measurements in a tidal river subject to sidewall effects and a mobile bed. *Water Resources Research*, 47(6).

[106] Sattari, M., Feizi, H., Colak, M., Ozturk, A., Apaydin, H., and Ozturk, F. (2020).

Estimation of sodium adsorption ratio in a river with kernel-based and decision-tree models. *Environmental Monitoring and Assessment*, 192:575.

[107] Sharma, G. K., Kumar, A., Rao, C. B., Jayakumar, T., and Raj, B. (2013). Short time Fourier transform analysis for understanding frequency dependent attenuation in austenitic stainless steel. *NDT & E International*, 53:1–7.

[108] Sherstinsky, A. (2020). Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network. *Physica D: Nonlinear Phenomena*, 404:132306.

[109] Shi, X., Chen, P., Ye, Z., Zhang, X., and Wang, W. (2024). Tide level prediction during typhoons based on variable topology in graph convolution recurrent neural networks. *Ocean Engineering*, 312:119228.

[110] Sit, M., Demiray, B. Z., Xiang, Z., Ewing, G. J., Sermet, Y., and Demir, I. (2020). A Comprehensive Review of Deep Learning Applications in Hydrology and Water Resources. *Water Science and Technology*, 82(12):2635–2670.

[111] Slater, L. J., Blougouras, G., Deng, L., Deng, Q., Ford, E., van Dijke, A. H., Huang, F., Jiang, S., Liu, Y., Moulds, S., et al. (2024). Challenges and opportunities of ml and explainable ai in hydrology.

[112] Sun, T., Zhou, M., and Chen, L. (2025). An Under-Sampled Line Array Element Signal Reconstruction Method Based on Compressed Sensing Theory. *Archives of Acoustics*.

[113] Sung, J. Y., Lee, J., Chung, I.-M., and Heo, J.-H. (2017). Hourly Water Level Forecasting at Tributary Affected by Main River Condition. *Water*, 9(9):644.

[114] Supharatid, S. (2003a). Application of a neural network model in establishing a stage-discharge relationship for a tidal river. *Hydrological Processes*, 17(15):3085–3099.

[115] Supharatid, S. (2003b). Tidal-level forecasting and filtering by neural network model. *Coastal Engineering Journal*, 45(01):119–137.

[116] Tang, S., Li, B., and Yu, H. (2024). Chebnet: efficient and stable constructions of deep neural networks with rectified power units via chebyshev approximation. *Communications in Mathematics and Statistics*, pages 1–27.

[117] Thanh, H. V., Binh, D. V., Kantoush, S. A., Nourani, V., Saber, M., Lee, K.-K., and Sumi, T. (2022). Reconstructing Daily Discharge in a Megadelta Using Machine Learning Techniques. *Water Resources Research*, 58(5):e2021WR031048.

[118] Thanh Hoan, N., Van Dung, N., Le Thu, H., Thuy Quynh, H., Al-Ansari, N., Van Phong, T., Trong Trinh, P., Duc Nguyen, D., Van Le, H., Bich Thi Nguyen, H., et al. (2022). Novel Time Series Bagging Based Hybrid Models for Predicting Historical Water Levels in the Mekong Delta Region, Vietnam. *CMES-Computer Modeling in Engineering & Sciences*, 131(3):1431–1449.

[119] Thomas, M. and Joy, A. T. (2006). *Elements of information theory*. Wiley-Interscience.

[120] Tian, Y., Zhang, Q., Huang, H., Huang, Y., Tao, J., Zhou, G., Zhang, Y., Yang, Y., and 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.

[121] Tsai, C.-C., Lu, M.-C., and Wei, C.-C. (2012). Decision Tree-Based Classifier Combined with Neural-Based Predictor for Water-Stage Forecasts in a River Basin During Typhoons: A Case Study in Taiwan. *Environmental Engineering Science*, 29(2):108–116.

[122] Tung, T. M., Yaseen, Z. M., et al. (2020). A survey on river water quality modelling using artificial intelligence models: 2000-2020. *Journal of Hydrology*, 585:124670.

[123] Vapnik, V. N. (2000). *The Nature of Statistical Learning Theory*. Springer New York, NY, 2 edition.

[124] Vidyalashmi, K., Nandana, J., Azhikodan, G., Priya, K., Yokoyama, K., Paramasivam, S. K., et al. (2024). Analysing the performance of the NARX model for forecasting the water level in the Chikugo River estuary, Japan. *Environmental Research*, 251:118531.

[125] Villeneuve, Y., Sguin, S., and Chehri, A. (2023). AI-Based Scheduling Models, Optimization, and Prediction for Hydropower Generation: Opportunities, Issues, and Future Directions. *Energies*, 16:3335.

[126] Vu, M., Jardani, A., Krimissa, M., Zaoui, F., and 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.

[127] Weber de Melo, W., Iglesias, I., and Pinho, J. (2025). Early warning system for floods at estuarine areas: Combining artificial intelligence with process-based models. *Natural Hazards*, 121(4):4615–4638.

[128] Wee, W. J., Zaini, N. B., Ahmed, A. N., and El-Shafie, A. (2021). A review of models for water level forecasting based on machine learning. *Earth Science Informatics*, 14(4):1707 1728.

[129] Wei, C.-C. (2012). Wavelet kernel support vector machines forecasting techniques: Case study on water-level predictions during typhoons. *Expert Systems with Applications*, 39(5):5189–5199.

[130] Wei, C.-C. (2015). Comparing lazy and eager learning models for water level forecasting in river-reservoir basins of inundation regions. *Environmental Modelling & Software*, 63:137–155.

[131] Wei, C.-C. and Hsu, N.-S. (2008). Multireservoir Flood-Control Optimization with Neural-Based Linear Channel Level Routing Under Tidal Effects. *Water Resources Management*, 22(11):1625–1647.

[132] Wolfs, V. and Willems, P. (2014). Development of discharge-stage curves affected by hysteresis using time varying models, model trees and neural networks. *Environmental Modelling & Software*, 55:107–119.

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

[134] Yang, J.-S., Yu, S.-P., and Liu, G.-M. (2013). Multi-step-ahead predictor design for effective long-term forecast of hydrological signals using a novel wavelet neural network hybrid model. *Hydrology and Earth System Sciences*, 17(12):4981–4993.

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

[136] Yoo, H., Kim, D., Kwon, H.-H., and Lee, S. (2020). Data Driven Water Surface Elevation Forecasting Model with Hybrid Activation FunctionA Case Study for Hangang River, South Korea. *Applied Sciences (Switzerland)*, 10:1424.

[137] Zainal, N. N. and Abu Talib, S. H. (2024). Review paper on applications of the hec-ras model for flooding, agriculture, and water quality simulation. *Water Practice & Technology*, 19(7):2883–2900.

[138] Zhang, L., Qin, H., Mao, J., Cao, X., and Fu, G. (2023a). High temporal resolution urban flood prediction using attention-based LSTM models. *Journal of Hydrology*, 620:129499.

[139] Zhang, X., Wang, T., Wang, W., Shen, P., Cai, Z., and Cai, H. (2023b). A multi-site tide level prediction model based on graph convolutional recurrent networks. *Ocean Engineering*, 269.

[140] Zhang, Z., Zhang, L., Yue, S., Wu, J., and Guo, F. (2023c). Correction of nonstationary tidal prediction using deep-learning neural network models in tidal estuaries and rivers. *Journal of Hydrology*, 622.

[141] Zhang, Z., Zhang, Q., and Singh, V. P. (2018). Univariate streamflow forecasting using commonly used data-driven models: literature review and case study. *Hydrological Sciences Journal*, 63(7):1091 1111.

[142] Zhu, S., Lu, H., Ptak, M., Dai, J., and Ji, Q. (2020). Lake water-level fluctuation forecasting using machine learning models: a systematic review. *Environmental Science and Pollution Research*, 27(36):44807 44819.

[143] Zounemat-Kermani, M., Matta, E., Cominola, A., Xia, X., Zhang, Q., Liang, Q., and 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.

List of Figures

List of Tables

LIST OF ABBREVIATIONS

ACC	accuracy level	AC	auto-correlation
ADCP	Acoustic Doppler Current Profiler	AIC	Akaikes Information Criterion
ANFIS	Adaptive Neuro Fuzzy Inference System	ANN	Artificial Neural Network
ANOVA	Analysis of Variance	AR	Auto Regressive
ARIMA	Auto Regressive Integrated Moving Average	ARMA	Auto Regressive Moving Average
BO	Bayesian optimization	BP	Back-propagation
BPNN	Back-propagation Neural Network	BPTT	Back-propagation Through Time
BRR	Bayesian Ridge Regression	CART	Classification And Regression Tree
CC	cross-correlation	CDW	Continuous and Discrete Wavelet Transformation
CFNN	Counter-propagation Fuzzy Neural Network	Chebnet	Chebyshev Graph Convolutional Network
CNN	Convolution Neural Network	CWT	Continuous Wavelet Transform
DCL	Deep Characteristic Learning	DT	Decision Tree
DWT	Discrete Wavelet Transform	EEMD	Ensemble Empirical Mode Decomposition
EI	efficiency index	EMA	Exponential Moving Average
EMD	Empirical Mode Decomposition	EnKF	Ensemble Kalman Filter
ENN	Emotional Neural Network	ETP	error of time-to-peak
FFBP	Feed-Forward Back Propagation	FFNN	Feed-Forward Neural Network
FT	Fourier Transform	GA	Genetic algorithm
GANN	Genetic Algorithm Neural Network	GBDT	Gradient Boosted Decision Tree
GGA	Generalized Genetic Algorithm	GPR	Gaussian Process Regression
GPU	graphical processing unit	GRNN	Generalized regression neural network
GRU	Gated Recurrent Unit	H-ADCP	Horizontal Acoustic Doppler Current Profiler

HA	Hydrological analysis	H2C	Hydrologic and Hydrodynamic Coupling model
H2C-XL	Hydrologic and Hydrodynamic Coupling model - XGBoost and LSTM	HHLC	Hydrologic-Hydrodynamic-LSTM Coupling model
HLHC	Hydrologic-LSTM-Hydrodynamic Coupling model	IMF	intrinsic mode function
KNN	K-Nearest Neighbour	LGBM	Light Gradient Boosting Machine
LOESS	Locally Weighted Least Squares	LR	Linear Regression
LS	Least Square	LSSVM	Least Squares Support Vector Machine
LSTM	Long short-term memory	LWR	Locally Weighted Regression
MA	Moving Average	MAD	mean absolute deviation
MAE	mean absolute error	MAPE	mean absolute percentage error
MARS	Multivariate Adaptive Regression Splines	ML	Machine Learning
MLFF	Multilayer Feed-Forward	MLP	Multilayer Perceptron
MLR	Multiple Linear Regression	MPR	Multiple Power Regression
MSE	mean squared error	Nadam	Nesterov-accelerated Adaptive Moment Estimation
NARX	Nonlinear Autoregressive with Exogenous input	NF	Neuro Fuzzy
NS_TIDE	nonstationary Tidal Harmonic Analysis	NSE, E	Nash-Sutcliffe efficiency coefficient
OI	orthogonality index	PAC	Partial auto-correlation
PCA	Principal Component Analysis	PE	peak error
PSO	particle swarm optimization	R, CC, PCC	correlation coefficient
R^2	coefficient of determination	$\bar{R}^2$	adjusted R^2
RBF	Radial Basis Function	RBFNN	Radial Basis Function Neural Network
ReLu	Rectified Linear Unit	REPT	Reduced Error Pruning Trees
RF	Random Forest	RMSE	root mean square error
RMSPE	root mean square percentage error	RNN	Recurrent Neural Network
RNMM	River Network Mathematical Model	SDP-RC	State Dependent Parameter-Rating Curve
Seq2Seq	sequence-to-sequence	SHAP	SHapley Additive exPlanations
SKI	skill index	SMO	Sequential Minimal Optimization
SNN	Shallow Neural Network	SRC	Single Rating Curve
SS	skill score	STFT	Short-time Fourier transform

STREAM	STRatified EstaArine Model	SVM	Support Vector Machine
SVR	Support Vector Regression	SWE	Shallow Water Equations
tanh	hyperbolic tangent	VMD	Variational Mode Decomposition
WMLP	Wavelet Multilayer Perceptron	WOS	Web of Science
WSVM	Wavelet Kernel Support Kernel Machine	WT	Wavelet Transform
XAI	explainable artificial intelligence	XGB	eXtreme Gradient Boosting

APPENDIX

A   Summary and Characteristics of the Datasets

Table A.1: Summary of measured dataset (2016-2021; CMHS: Croatian Meteorological and Hydrological Service; CW: Croatian Waters)

Statistic / Parameter	Ušće	Opuzen	Kula Norinska	Metković	Metković
Data source	CMHS	CMHS	CMHS	CMHS	CW
Hydrological parameter	water level	water level	water level	water level	discharge
Number of observations	52605	52605	52605	52605	52605
Min	-0.527	-0.289	-0.313	-0.496	1.633
Median	0.013	0.121	0.137	0.134	204.96
Mean	0.019	0.168	0.197	0.223	323.6
Standard deviation	0.152	0.234	0.264	0.337	264.482
Max	0.763	1.551	1.777	2.247	1874.137
Resolution	1h	1h	1h	1h	1h
Units	m	m	m	m	m^3/s

Table A.2: Summary of simulated data generated with the STREAM 1D numerical model (2016-2021)

Statistic / Parameter	Ušće	Opuzen	Kula Norinska	Metković	Metković
Hydrological parameter	water level	water level	water level	water level	discharge
Number of observations	52605	52605	52605	52605	52605
Min	-0.490	-0.499	-0.507	-0.511	0
Median	0.010	0.113	0.125	0.138	207.936
Mean	0.019	0.168	0.197	0.223	323.543
Standard deviation	0.152	0.234	0.264	0.337	264.482
Max	0.745	1.919	2.214	2.527	1876.449
Resolution	1h	1h	1h	1h	1h
Units	m	m	m	m	m^3/s

Table A.3: Summary of simulated data generated with the HEC-RAS numerical model (2016-2019)

Statistic / Parameter	Ušće	Opuzen	Kula Norinska	Metković	Metković
Hydrological parameter	water level	water level	water level	water level	discharge
Number of observations	35051	35051	35051	35051	35051
Min	-0.468	-0.454	-0.449	-0.437	64.904
Median	0.017	0.077	0.095	0.122	246.404
Mean	0.026	0.119	0.152	0.207	347.729
Standard deviation	0.142	0.217	0.251	0.309	222.862
Max	0.712	1.316	1.596	2.056	1434.299
Resolution	1h	1h	1h	1h	1h
Units	m	m	m	m	m^3/s

B Survey of Reviewed Publications Focusing on Tidal Rivers and Estuaries

B.1 Forecasting Hydrological Parameters

Table B.1: Listed reviewed publications from the period between 2000 and 2023 focused on water level forecasting. This summary is based on a review study by Mihel et al. [90], as well as additional studies published from 2024 to the present day, presenting details regarding the selected input and output features, the employed machine learning approach, applied evaluation metrics, and the temporal scale of input and output features.

Author/s	Method	Input	Input Time Scale	Output	Output Time Scale	Evaluation Metric
Supharatid (2003) [115]	MLFF	Scenario 1: water level	hourly,	tidal level	hourly, weekly, monthly	EI, RMSE, MAD
	MLFF	Scenario 2: tidal level	hourly	tidal level	hourly	EI, RMSE, MAD
Chang and Chen (2003) [15]	RBFNN	lunar month, lunar day, time, water level	hourly	water level	hourly	R, RMSE
Tsai et al. (2012) [121]	CART-ANN (MLP, RBF), benchmark models: CART, BPNN, RBFNN	precipitation, water level, historic releases	hourly	water level	hourly	MSE, MAE
Wei (2012) [129]	WSVM, SVM	water level, average precipitation, reservoir releases	hourly	water level	hourly SVM	RMSE
Yang et al. (2013) [134]	CDW-NF, CDW-ANN, CDW-LR	water level (average value of two time high-tide level)	daily	water level	daily	RMSE, MAE, R^2
Wei (2015) [130]	LWR, KNN, LR, SVR, ANN	water level, average precipitation, reservoir releases, tidal effects	hourly	water stage	hourly	CC, MAE, RMSE, AIC, computational efficiency
Pasupa and Jungjareantrat (2016) [96]	LR, KR, SVR, KNN, RF	water level	hourly	water level	hourly	RMSE
Ahmed et al. (2017) [3]	SVR (RBF, ANOVA) + (MA, EMA)	daily, morning, and night tide data	hourly	tide level	daily	MAE
Sung et al. (2017) [113]	MLP	water level, rainfall	hourly	water level	hourly	RMSE, R^2, NSE
Jung et al. (2018) [64]	LSTM	dam discharges, water level, predicted tide level	hourly	water level	hourly	RMSE, NSE
Yoo et al. (2020) [136]	LSTM	precipitation, discharge,	hourly	water surface elevation	hourly	RMSE, PE, NSE

Table B.1 – - Continued from previous page

Author/s Year	Method	Input	Input Time Scale	Output	Output Time Scale	Evaluation Metric
		tide level				
Chen et al. (2020) [20]	NS_TIDE + AR	discharge, tides	hourly	water level	hourly	RMSE
Liang et al. (2021) [76]	NARX+EWT, NARX+EEMD, NARX+EMD, HA	tidal level	hourly	tidal level	hourly	RMSE, R, MAPE, MAE
Guo et al. (2021) [47]	BO + SVR/RFR, /MLPR /LGBMR	rainfall, water level, tide	hourly	water level	hourly	NSE, R^2, MAE, RMSE, PWE, ETP
Chen et al. (2021) [17]	LSTM, BRR, GBDT, LR, SVR	meteorological data, water level, additional reference factors	hourly	water level	hourly	MAE, RMSE, ACC
Zhang et al. (2023) [140]	NS_TIDE + (LSTM+FNN +Q, LSTM+FNN, LSTM, AR)	discharge, water level,	hourly	water level	hourly	LOSS, RMSE, R^2
Chen et al. (2023) [22]	SMLR + EEMD	water level	hourly	water level	hourly	RMSE, R
Zhang et al. (2023) [139]	Cheb-GRU, Conv-LSTM, LSTM, GRU	tide level, meteorological data, time	hourly	tidal level	hourly	RMSE, MAE
Dato et al. (2024) [30]	MLFF	water level, meteorological data, predicted astronomical tide	hourly	water level	hourly	RMSE, TP, FP
Vidyalashmi et al. (2024) [124]	NARX	water level, discharge, salinity	hourly	water level	hourly	MSE, R
Gan et al. (2024) [36]	LGBM1, LGBM2	water level	hourly	water level	hourly	RMSE, MAE
Shi et al. (2024) [109]	Cheb-GRU, GRU, LSTM, ConvLSTM	tide level, meteorological data	hourly	tide level	hourly	RMSE, MAE, R^2
Cremer et al. (2025) [28]	LSTM + MIKE FM + DA, DA + MIKE FM, MIKE FM	water level, discharge, wind speed, wind direction	hourly	water level	hourly	URMSE
Gao et al. (2025) [38]	LSTM, GRU, LSTM-CNN, (LSTM, GRU, LSTM-CNN) +EMD, EMD-ITG, NS_TIDE	water level	hourly	water level	hourly	RMSE, NSE, MAE, MAPE

Table B.2: Listed reviewed publications from the period between 2000 and 2023 focused on discharge forecasting. This summary is based on a review study by Mihel et al. [90], as well as additional studies published from 2024 to the present day, presenting details regarding the selected input and output features, the employed machine learning approach, applied evaluation metrics, and the temporal scale of input and output features.

Author/s Year	Method	Input	Input Time Scale	Output	Output Time Scale	Evaluation Metric
Hidayat et al. (2014) [55]	MLP	water level, at-site historical discharge data, predicted tide level	hourly	discharge	hourly	RMSE, R^2, NSE
Vu et al. (2023) [126]	stacked LSTM	piezometer, sea level, air temperature, atmospheric pressure, precipitation, soil moisture, relative humidity, evaporation rate	daily	discharge	daily	R, R^2, RMSE
Chen et al. (2024) [23]	HA, PSO-BP, LSTM, LSTM+Seq2Seq	flow velocity, water level, discharge	half-hourly	discharge	half-hourly	RMSE, RSD, PCC

B.2   Reconstructing Hydrological Parameters

Table B.3: Listed reviewed publications from the period between 2000 and 2023 focused on water level reconstruction. This summary is based on a review study by Mihel et al. [90], as well as additional studies published from 2024 to the present day, presenting details regarding the selected input and output features, the employed machine learning approach, applied evaluation metrics, and the temporal scale of input and output features.

Author/s Year	Method	Input	Input Time Scale	Output	Output Time Scale	Evaluation Metric
Adib (2008) [1]	MLP	discharge, tide elevation, distance from the river mouth	daily	water level	daily	R^2
Wei and Hsu (2008) [131]	FFBP, CCCMMOC	water level, reservoir discharge, total lateral discharge, control-point levels, tributary discharge	hourly	water level	hourly	RMSE
Chinh et al. (2009) [24]	FFNN	water level, rainfall	minutes	water level	minutes	RMSE
Chen et al. (2012) [18]	BPNN, vertical 2D, 3D hydrodynamic models	water level, freshwater discharge	hourly	water level	hourly	RMSE, R, E
Pierini et al. (2013) [97]	BPNN,	ANN: tidal data, Numerical: water level, current, wind	hourly	tide level	hourly	RMSE, R, SKI
Liu and Chung (2014) [78]	BPNN, GANN	freshwater discharge downstream water level	hourly	water level	hourly	MAE, RMSE, PE
Bhar and Bakshi (2020) [12]	MLP	tide level	half-hourly	water level	single day	RMSE, E, R, MAPE
Guillou and Chapalain (2021) [45]	MLR, MPR, MLP	French tidal coefficient, atmospheric pressure, wind speed river discharge	hourly	water level maxima	hourly	MAE, R^2, RMSE
Gan et al. (2021) [37]	LGBM, NS_TIDE	river discharge, tide	hourly	water level	hourly	MAE, RMSE, CC, SS
Sampurno et al. (2022) [104]	SLIM 2D + MLR/SVM/RF	discharge, tide, weather parameters	hourly	water level	hourly	RMSE, NSE
Thanh Hoan et al. (2022) [118]	Bagging + RF/SMO/M5P, REPT	water level	daily	historical water level	daily	R^2, RMSE, MAE
Fei et al. (2023) [32]	H2C-XL, HLHC, HHLC, H2C	discharge, water level, tidal level, precipitation, evapotranspiration	hourly, daily	water level	hourly	NSE, KGE
Lauer and Kösters (2024) [72]	ANN + BO, MLR, MNLR	tidal low/high water levels, tidal range, tidal mean water	hourly	minimal and maximal tidal water levels	hourly	RMSE

Table B.3 – - Continued from previous page

Author/s	Method	Input	Input Time Scale	Output	Output Time Scale	Evaluation Metric
		level, flood and ebb duration, discharge, meteorological data				
Dato et al. (2024) [30]	MLFF	water level, meteorological data, predicted astronomical tide	hourly	water level	hourly	RMSE, TP, FP

Table B.4: Listed reviewed publications from the period between 2000 and 2023 focused on discharge reconstruction. This summary is based on a review study by Mihel et al. [90], as well as additional studies published from 2024 to the present day, presenting details regarding the selected input and output features, the employed machine learning approach, applied evaluation metrics, and the temporal scale of input and output features.

Author/s Year	Method	Input	Input Time Scale	Output	Output Time Scale	Evaluation Metric
Gu et al. (2014) [44]	RNMM + BPNN + GGA	boundary condition, opening degree and time, average stage of inner river, stage of outer river	hourly	discharge, stage	hourly	RMSE
Hidayat et al. (2014) [55]	WMLP	water levels, predicted tide levels, amplitude of tidal components	hourly	discharge	hourly	RMSE, R^2, NSE
Garel and Dalimonte (2017) [39]	MLP	velocity	daily	freshwater discharge	daily	-
Thanh et al. (2022) [117]	GPR, LSSVM, SVR, MARS, DT, RF	water stages	daily	discharge	daily	RMSE, R, NSE, MAE
Li et al. (2025) [75]	DCL, SVM, GRNN, RBF, Elman, BP, PCA-MLR, MLR-20, MLR, LR	river characteristics, cross-section shape, discharge, cell velocity data	not specified	discharge	not specified	RMSE, R^2

B.3 Establishing Stage-Discharge Relationship

Table B.5: Listed reviewed publications from the period between 2000 and 2023 focused on modeling the stage-discharge relationship. This summary is based on a review study by Mihel et al. [90], and a search for additional studies published from 2024 to the present day, presenting details regarding the selected input and output features, the employed machine learning approach, and applied evaluation metrics.

Author/s Year	Method	Input	Output	Evaluation Metric
Supharatid (2003) [114]	MLFF	tidal range, discharge	water level	EI, RMSE, MAD
Habib and Meselhe (2006) [50]	MLFF, LOESS	water level	discharge	RMSE, E
Wolfs and Willems (2014) [132]	SRC, SDP-RC, MLP, M5	local water level, local water level gradient	discharge	MAPE, RMSPE, R^2, $\bar{R}^2$
Hidayat et al. (2014) [55]	WMLP	water levels, predicted tide levels, amplitude of tidal components	discharge	RMSE, R^2, NSE
Thanh et al. (2022) [117]	GPR, LSSVM, SVR, MARS, DT, RF	water stages	discharge	RMSE, R, NSE, MAE

C   Estimation of Hydrological Parameters

C.1   Optimization of Hyperparameters for Discharge Estimation Task

Table C.1: Optimal hyperparameters for different machine learning models using measured data [88]

Model	Hyperparameter	Tested search range	Optimal value
DT	max_depth	$[10, 200]^*$	20
	$min\_samples\_leaf$	$[10, 100]^*$	40
	$min\_samples\_split$	$[10, 100]^*$	10
RF	max_depth	$[10, 50]^*$	10
	$min\_samples\_leaf$	$[10, 100]^*$	10
	$min\_samples\_split$	$[10, 100]^*$	10
	$n\_estimators$	$[10, 200]^*$	90
SVR - rbf	C	$0.001 \ x \ 10^n$ for n $\in \{0, 1, ..., 6\}$	1000
	γ^a	[0.0001, 0.0005, 0.001, 0.005, 0.01, 0.05, 0.1, 1]	1
	ε^b	[0.0001, 0.0005, 0.001, 0.005, 0.01, 0.05, 0.1]	0.01
SVR - sigmoid	C	$0.001 \ x \ 10^n$ for n $\in \{0, 1, ..., 6\}$	1000
	γ^a	[0.0001, 0.0005, 0.001, 0.005, 0.01, 0.05, 0.1, 1]	0.0005
	ε^b	[0.0001, 0.0005, 0.001, 0.005, 0.01, 0.05, 0.1]	0.05
LGBM	$learning\_rate$	[0.0001, 0.0005, 0.001, 0.005, 0.01, 0.05]	0.05
	max_depth	$[10, 50]^*$	10
	$n\_estimators$	$[10, 200]^*$	200
	num_leaves	$[10, 100]^*$	10
XGB	$learning \ rate$	[0.0001, 0.0005, 0.001, 0.005, 0.01, 0.05]	0.05
	max_depth	$[10, 50]^*$	10
	$n\_estimators$	$[10, 200]^*$	120
LSTM	$batch\_size$	64, 128, 256, 512	512
	$learning\_rate$	[0.0001, 0.0005, 0.001, 0.005, 0.01, 0.05]	0.01
	$hidden\_units$	$[8, 128]^{**}$	56
LSTM-Attention	$batch\_size$	64, 128, 256, 512	64
	$learning\_rate$	[0.0001, 0.0005, 0.001, 0.005, 0.01, 0.05]	0.001
	$hidden\_units$	$[8, 128]^{**}$	96

$^a gamma$; $^b epsilon$

Table C.1: Optimal hyperparameters for different machine learning models using measured data [88]

Model	Hyperparameter	Tested search range	Optimal value

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

Table C.2: Optimal hyperparameters for different machine learning models using STREAM 1D simulated dataset [88]

Model	Hyperparameter	Tested search range	Optimal value
DT	max_depth	$[10, 200]^*$	10
	$min\_samples\_leaf$	$[10, 100]^*$	10
	$min\_samples\_split$	$[10, 100]^*$	10
RF	max_depth	$[10, 50]^*$	20
	$min\_samples\_leaf$	$[10, 100]^*$	10
	$min\_samples\_split$	$[10, 100]^*$	20
	$n\_estimators$	$[10, 200]^*$	130
SVR - rbf	C	$0.001 \ x \ 10^n$ for n $\in \{0, 1, ..., 6\}$	1000
	γ^a	[0.0001, 0.0005, 0.001, 0.005, 0.01, 0.05, 0.1, 1]	1
	ε^b	[0.0001, 0.0005, 0.001, 0.005, 0.01, 0.05, 0.1]	0.01
SVR - sigmoid	C	$0.001 \ x \ 10^n$ for n $\in \{0, 1, ..., 6\}$	1000
	γ^a	[0.0001, 0.0005, 0.001, 0.005, 0.01, 0.05, 0.1, 1]	0.005
	ε^b	[0.0001, 0.0005, 0.001, 0.005, 0.01, 0.05, 0.1]	0.05
LGBM	$learning\_rate$	[0.0001, 0.0005, 0.001, 0.005, 0.01, 0.05]	0.05
	max_depth	$[10, 50]^*$	50
	$n\_estimators$	$[10, 200]^*$	200
	num_leaves	$[10, 100]^*$	100
XGB	$learning \ rate$	[0.0001, 0.0005, 0.001, 0.005, 0.01, 0.05]	0.05
	max_depth	$[10, 50]^*$	10
	$n\_estimators$	$[10, 200]^*$	140
LSTM	$batch\_size$	64, 128, 256, 512	64
	$learning\_rate$	[0.0001, 0.0005, 0.001, 0.005, 0.01, 0.05]	0.0005
	$hidden\_units$	$[8, 128]^{**}$	48
LSTM-Attention	$batch\_size$	64, 128, 256, 512	64
	$learning\_rate$	[0.0001, 0.0005, 0.001, 0.005, 0.01, 0.05]	0.0001
	$hidden\_units$	$[8, 128]^{**}$	112

$^a gamma; ^b epsilon$

Table C.2: Optimal hyperparameters for different machine learning models using STREAM 1D simulated dataset [88]

| Model	Hyperparameter	Tested search range	Optimal value

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

C.2   Statistical Tests

Two approaches were employed to either accept or reject the hypothesis that the data follows a normal distribution. Since the dataset contains more than 5,000 values, the D'Agostino-Pearson test was selected to test the normality assumption, with the results depicted in Table C.3. Based on the obtained p-value, all models demonstrated deviations from normality. Similarly, significant deviations in skewness and kurtosis were observed for all simple non-temporal ML models, with the exception of time-series models. To further confirm these findings, a histogram combined with kernel density estimation was employed. Through visual inspection of residuals of the LSTM-Attention model (Figures C.1 and C.2), an approximately normal distribution was observed. As the majority of the employed models do not exhibit a normal distribution, a nonparametric test was selected for residual comparison.

Table C.3: Normality test of residuals from machine learning models

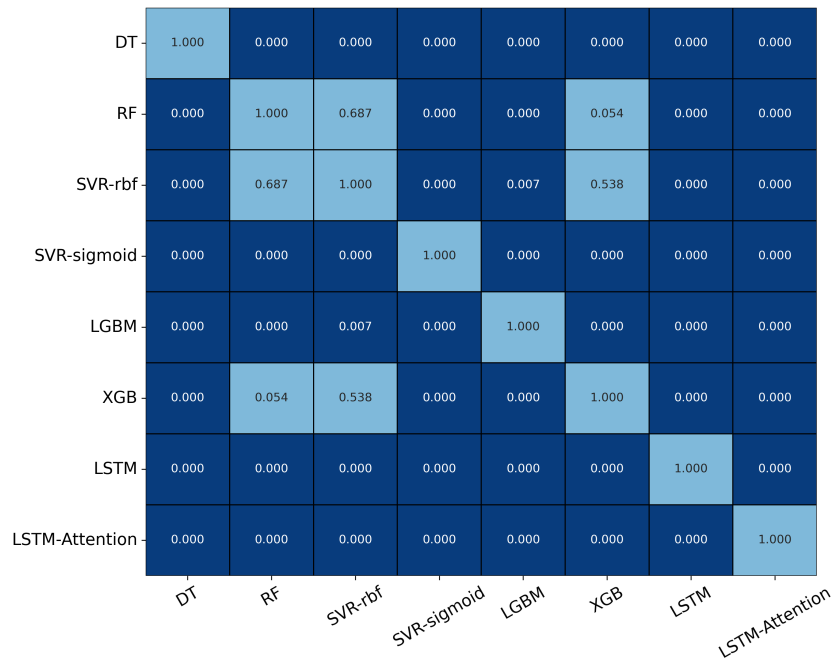| Model	Measured			Simulated		
	Skewness	Kurtosis	D'Agostino p-value	Skewness	Kurtosis	D'Agostino p-value
DT	0.676	6.150	0.000e+00	0.705	10.993	0.000e+00
RF	0.782	6.381	0.000e+00	0.940	13.962	0.000e+00
SVR-rbf	-0.590	11.710	0.000e+00	-10.309	210.397	0.000e+00
SVR-sigmoid	-0.543	6.946	1.492e-306	-0.503	5.231	3.232e-206
LGBM	0.841	6.416	0.000e+00	2.037	22.270	0.000e+00
XGB	0.866	7.137	0.000e+00	0.940	13.216	0.000e+00
LSTM	0.095	4.575	2.573e-84	0.167	3.885	2.099e-48
LSTM-Attention	0.286	4.307	4.948e-94	0.026	4.174	4.308e-56
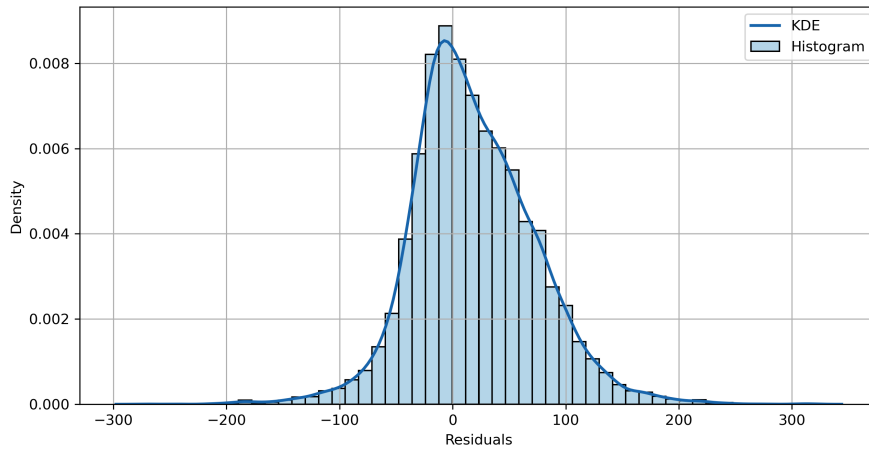
Figure C.1: Measured residual distribution of LSTM-Attention: Histogram and KDE visualization for normality assessment
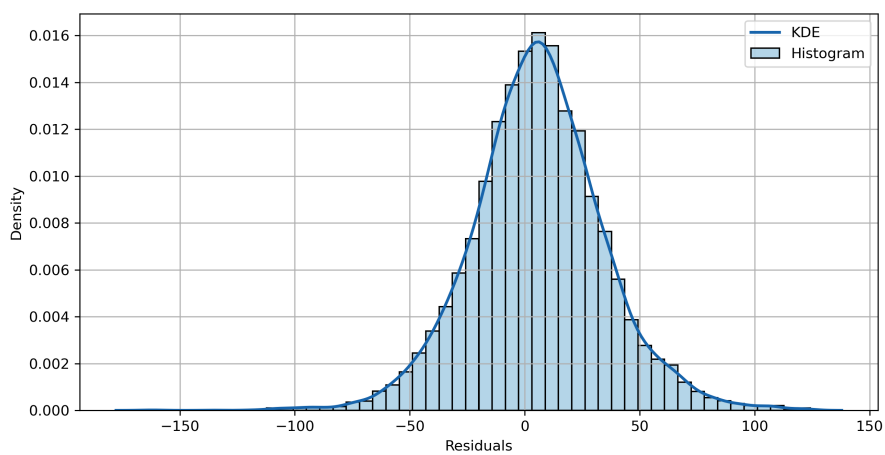
Figure C.2: Simulated residual distribution of LSTM-Attention: Histogram and KDE visualization for normality assessment
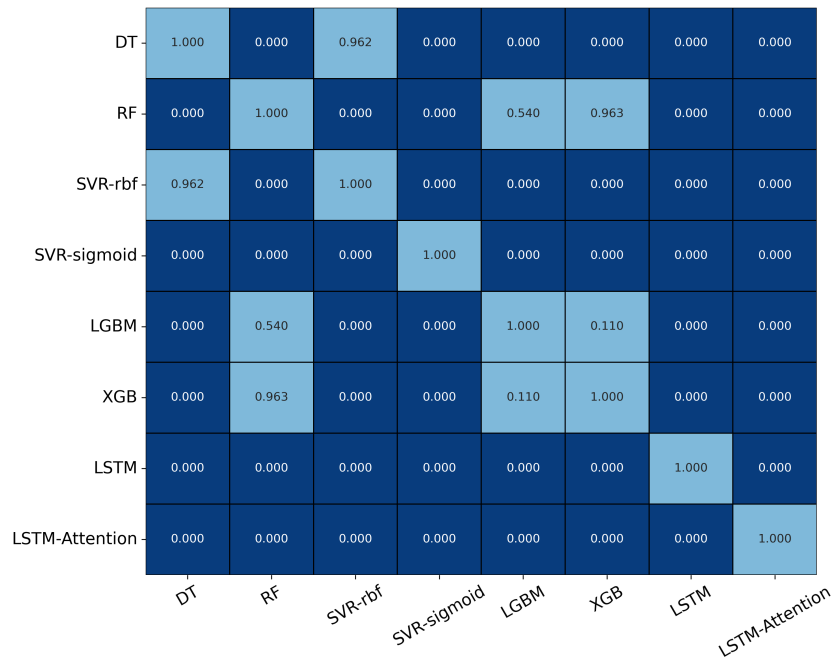
Figure C.3: Measured data: Matrix of Wilcoxon Signed-Rank p-values performed on model residuals

Figure C.4: Simulated data: Matrix of Wilcoxon Signed-Rank p-values performed on model residuals
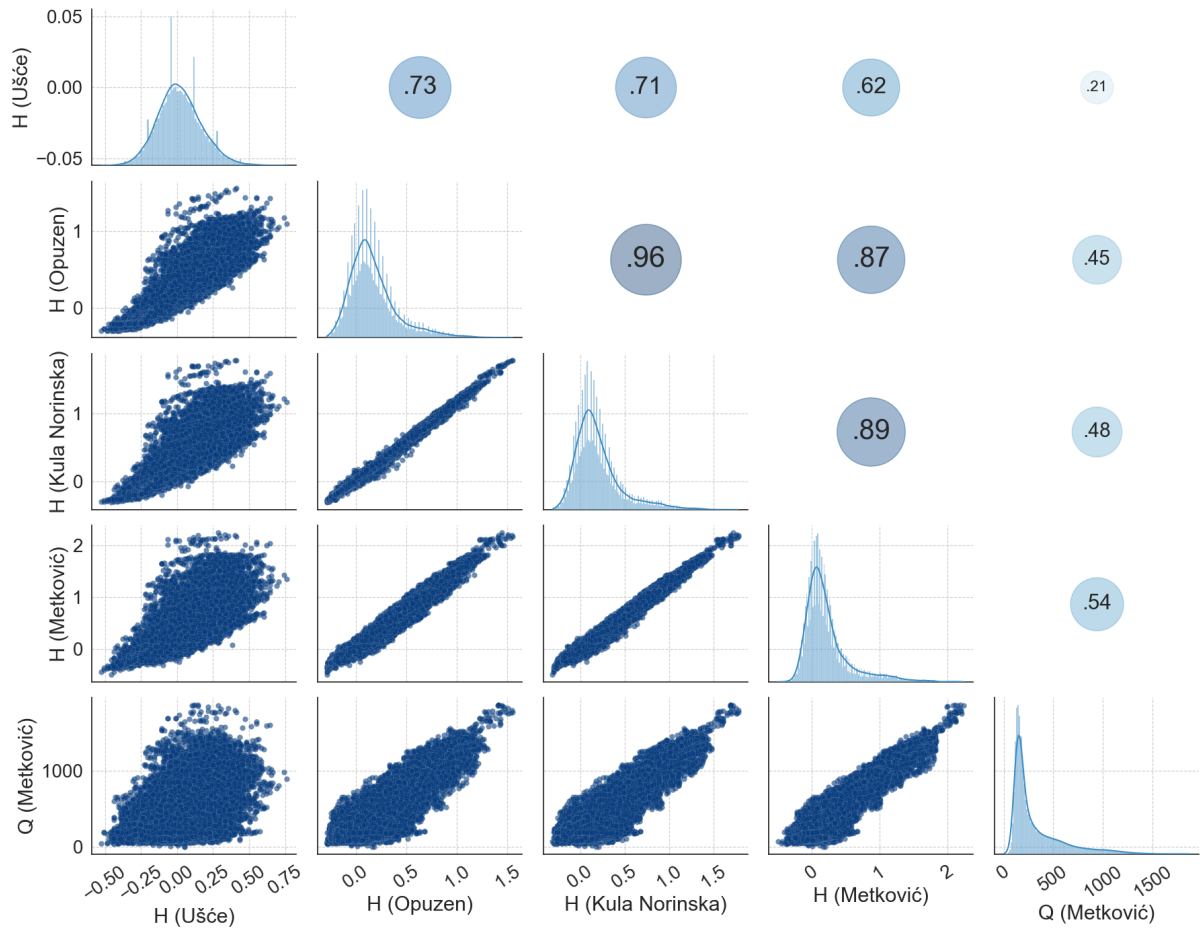
C.3   Feature Significance

Figure C.5: Measured data: Correlation matrix
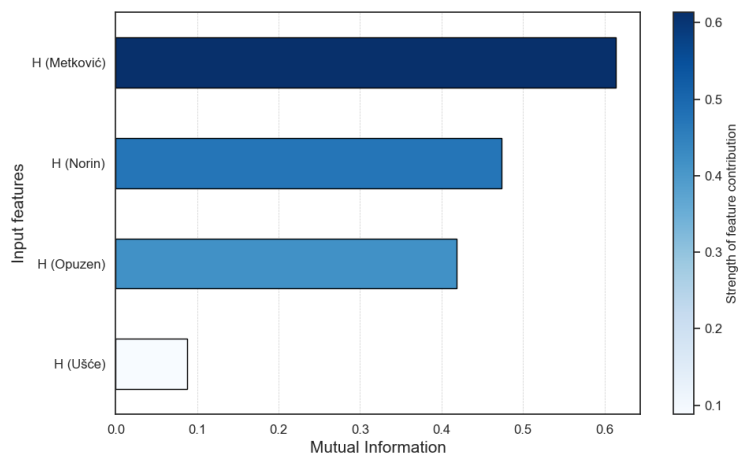
Figure C.6: Measured data: Mutual information

(a) DT

(b) RF

(c) LGBM

(d) XGB
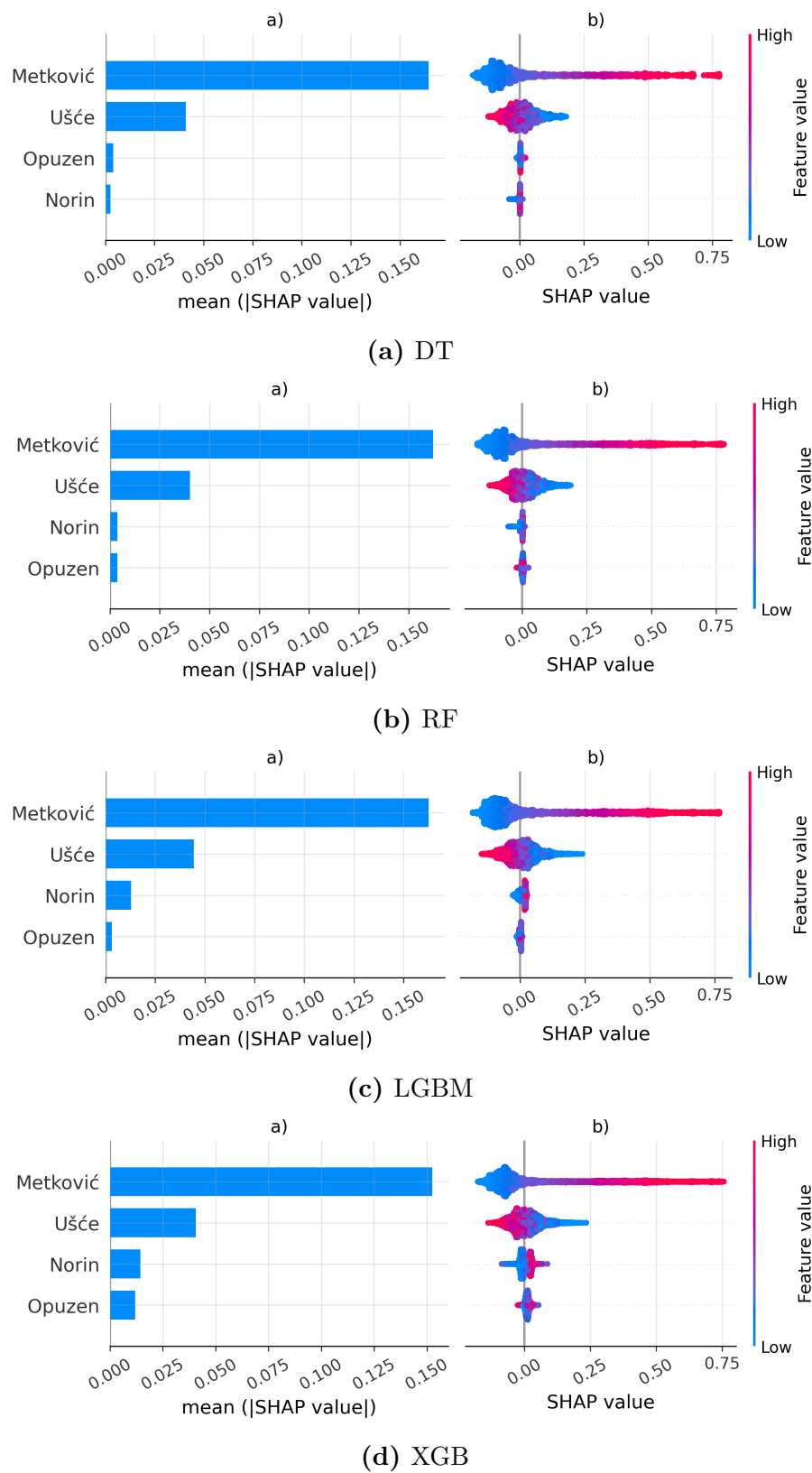
Figure C.7: Measured data: SHAP analysis [88]

Table C.4: Feature occlusion performed on measured data for the LSTM model for several scenarios. Single input scenario included only the water level from the target location, Metković. Second, the two-input scenario included the previous input combined with the tidal station data. The third scenario additionally included the Opuzen midstream station, and the fourth considered all available water level stations. [88]

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
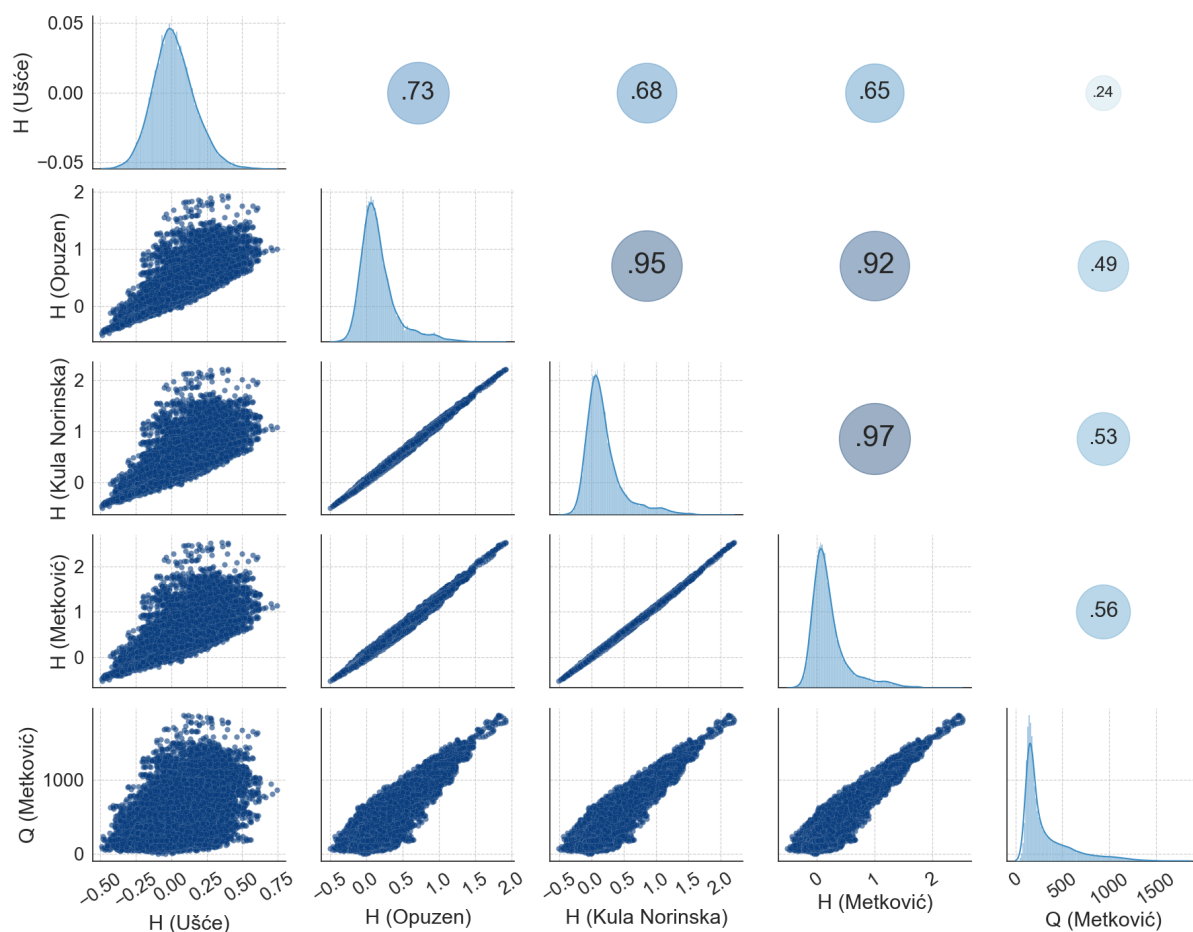[d] Input features: Ušće, Opuzen, Norin, and Metković

Table C.5: Feature occlusion performed on measured data for the LSTM-Attention model for several scenarios. Single input scenario included only the water level from the target location, Metković. Second, the two-input scenario included the previous input combined with the tidal station data. The third scenario additionally included the Opuzen midstream station, and the fourth considered all available water level stations. [88]

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
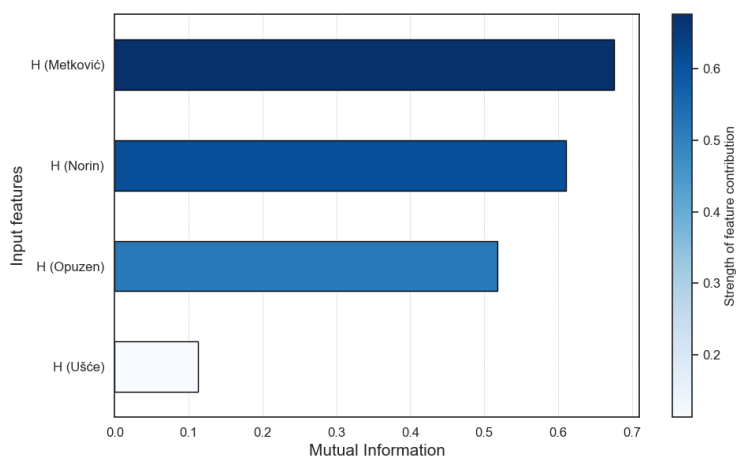
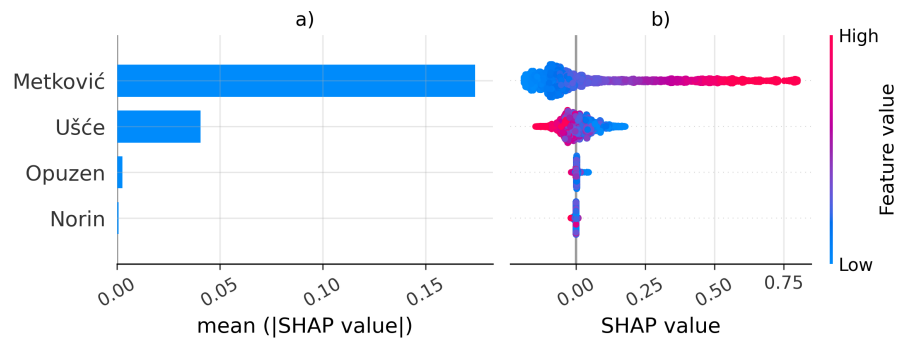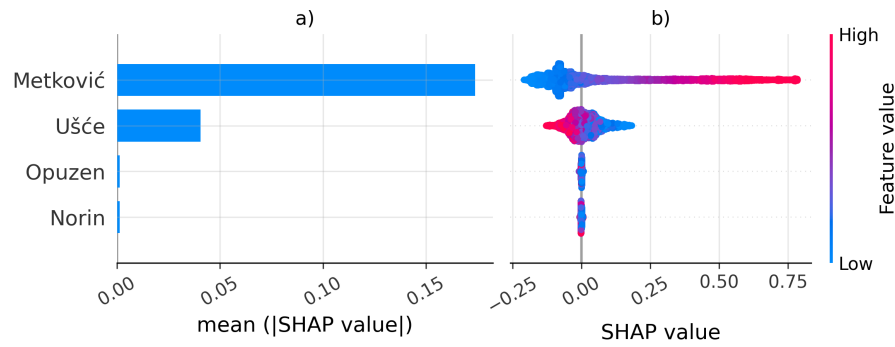**Figure C.8:** Simulated data: Correlation matrix
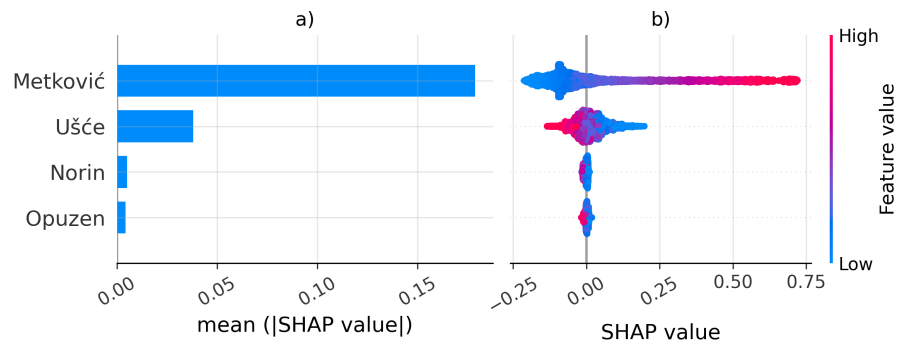

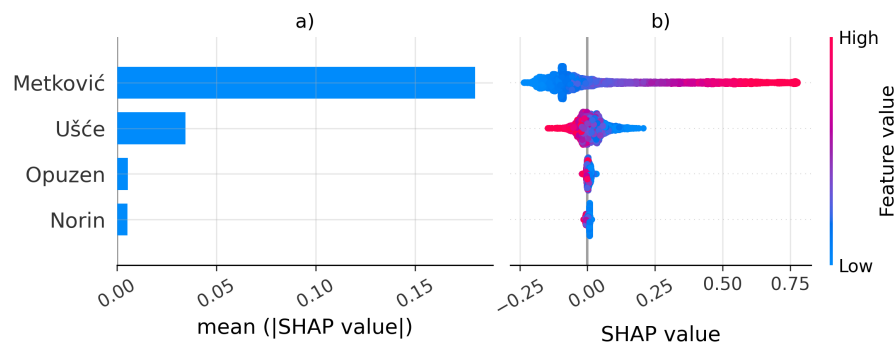**Figure C.9:** Simulated data: Mutual information

(a) DT

(b) RF

(c) LGBM

(d) XGB

Figure C.10: Simulated data: SHAP analysis [88]

Table C.6: Feature occlusion performed on simulated data for the LSTM model for several scenarios. Single input scenario included only the water level from the target location, Metković. Second, the two-input scenario included the previous input combined with the tidal station data. The third scenario additionally included the Opuzen midstream station, and the fourth considered all available water level stations. [88]

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 C.7: Feature occlusion performed on simulated data for the LSTM-Attention model for several scenarios. Single input scenario included only the water level from the target location, Metković. Second, the two-input scenario included the previous input combined with the tidal station data. The third scenario additionally included the Opuzen midstream station, and the fourth considered all available water level stations. [88]

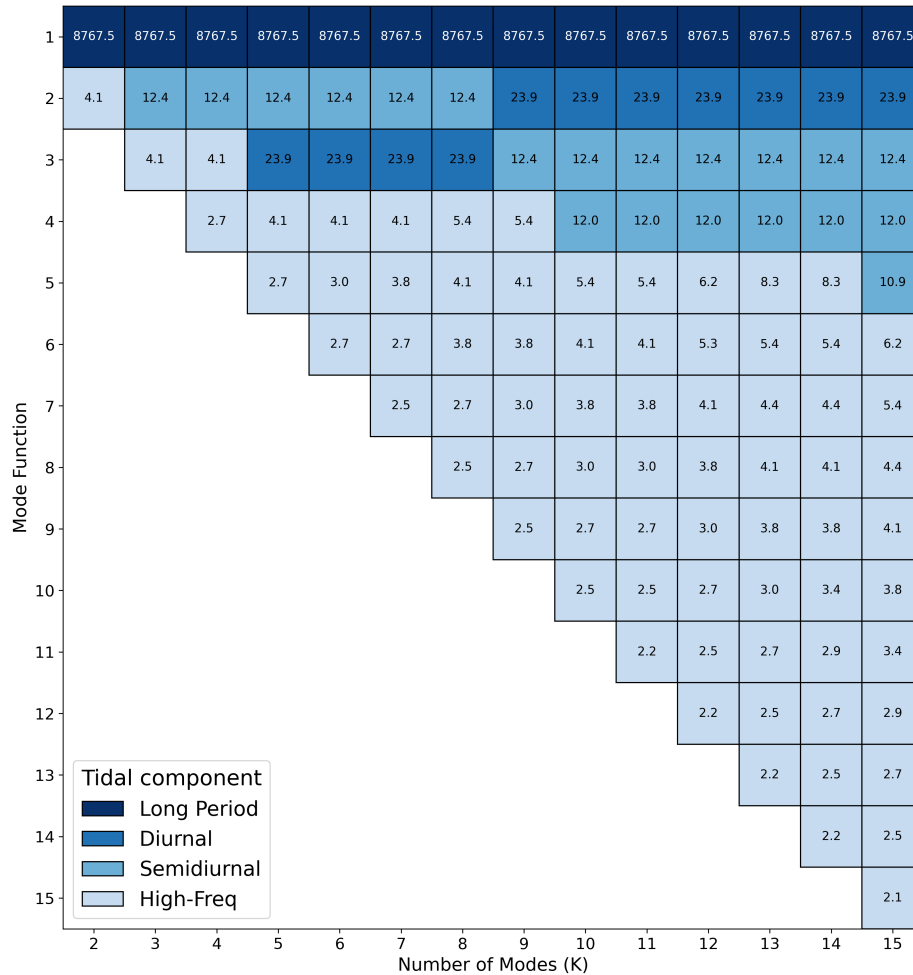| Metrics	Single Input[a]	Two Inputs[b]	Three Inputs[c]	Four Inputs[d]
RMSE	126.045	33.254	30.334	29.473
MAE	89.793	24.893	23.157	22.530
NSE	0.876	0.991	0.993	0.993
R	0.948	0.996	0.997	0.997

[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

D   Signal Processing Using VMD

D.1   Optimal VMD Parameters

Mode Function \ K	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	8767.5	8767.5	8767.5	8767.5	8767.5	8767.5	8767.5	8767.5	8767.5	8767.5	8767.5	8767.5	8767.5	8767.5
2	4.1	12.4	12.4	12.4	12.4	12.4	12.4	23.9	23.9	23.9	23.9	23.9	23.9	23.9
3		4.1	4.1	23.9	23.9	23.9	23.9	12.4	12.4	12.4	12.4	12.4	12.4	12.4
4			2.7	4.1	4.1	4.1	5.4	5.4	12.0	12.0	12.0	12.0	12.0	12.0
5				2.7	3.0	3.8	4.1	4.1	5.4	5.4	6.2	8.3	8.3	10.9
6					2.7	2.7	3.8	3.8	4.1	4.1	5.3	5.4	5.4	6.2
7						2.5	2.7	3.0	3.8	3.8	4.1	4.4	4.4	5.4
8							2.5	2.7	3.0	3.0	3.8	4.1	4.1	4.4
9								2.5	2.7	2.7	3.0	3.8	3.8	4.1
10									2.5	2.5	2.7	3.0	3.4	3.8
11										2.2	2.5	2.7	2.9	3.4
12											2.2	2.5	2.7	2.9
13												2.2	2.5	2.7
14													2.2	2.5
15														2.1

Tidal component
- Long Period
- Diurnal
- Semidiurnal
- High-Freq

Figure D.1: Ušće station: Example of tidal component classification of mode functions across tested range of K
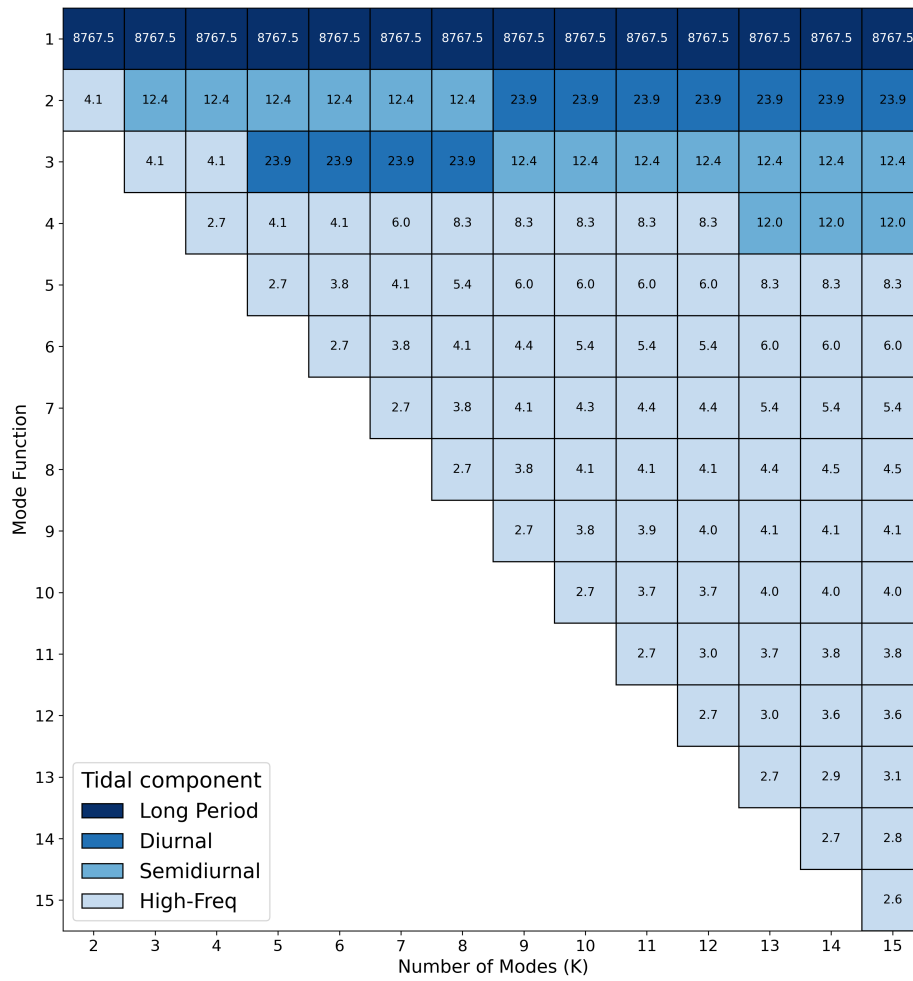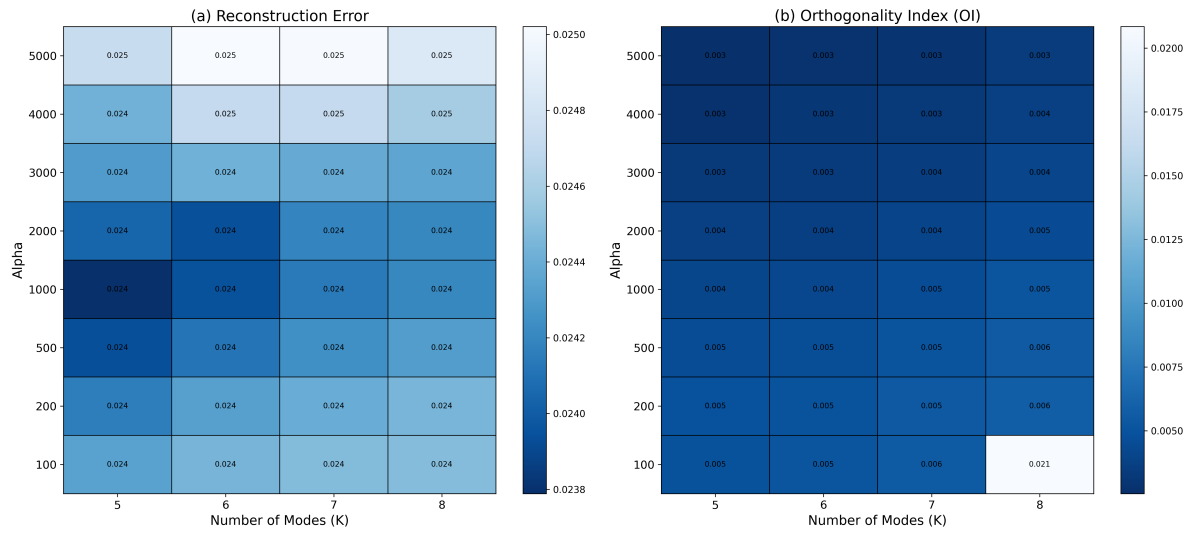
Figure D.2: Metković station: Example of tidal component classification of mode functions across tested range of K
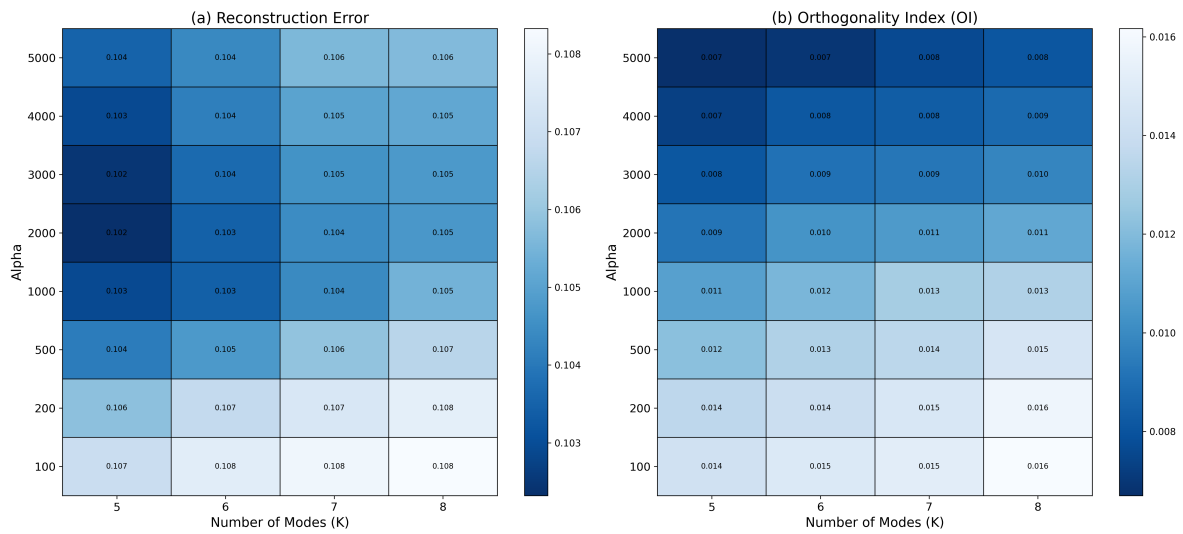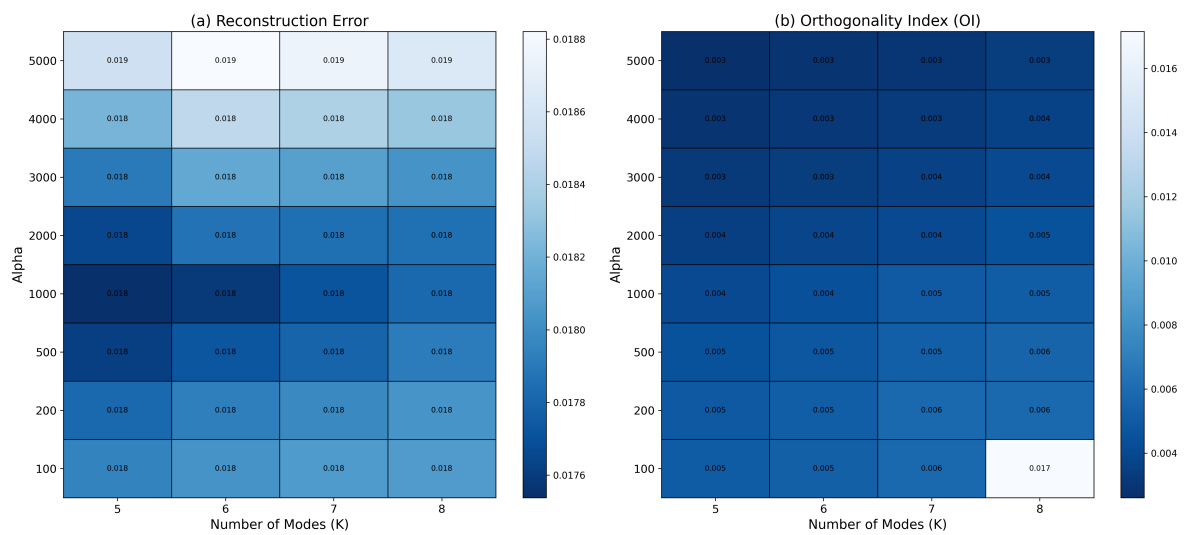

Figure D.3: Ušće station: Reconstruction error and OI for different K and α combinations
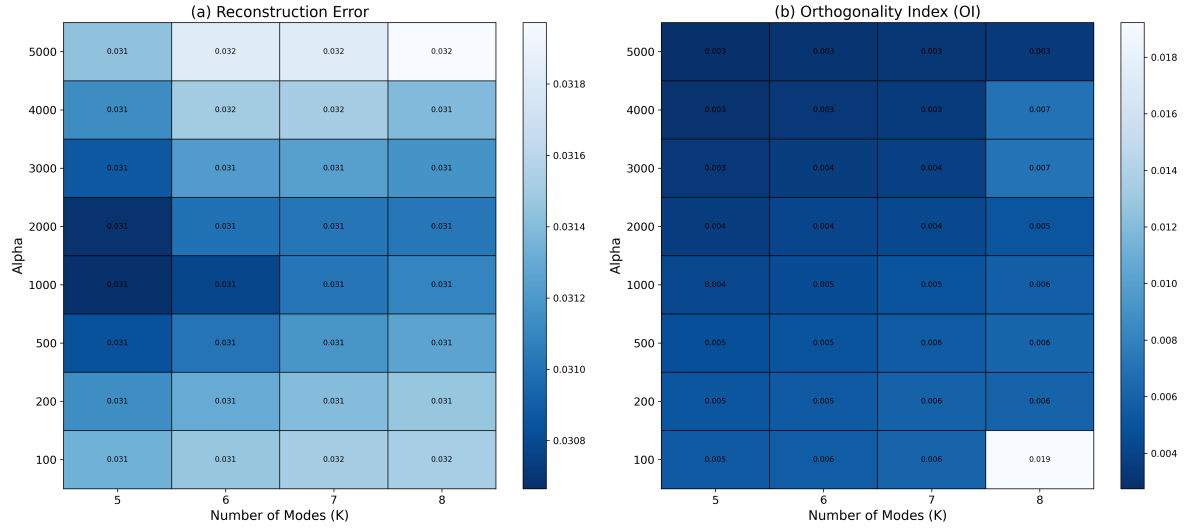
Figure D.4: Opuzen station: Reconstruction error and OI for different K and α combinations

Figure D.5: Norin station: Reconstruction error and OI for different K and α combinations

Figure D.6: Metković station: Reconstruction error and OI for different K and α combinations

D.2   Decomposed Original Water Level Signals
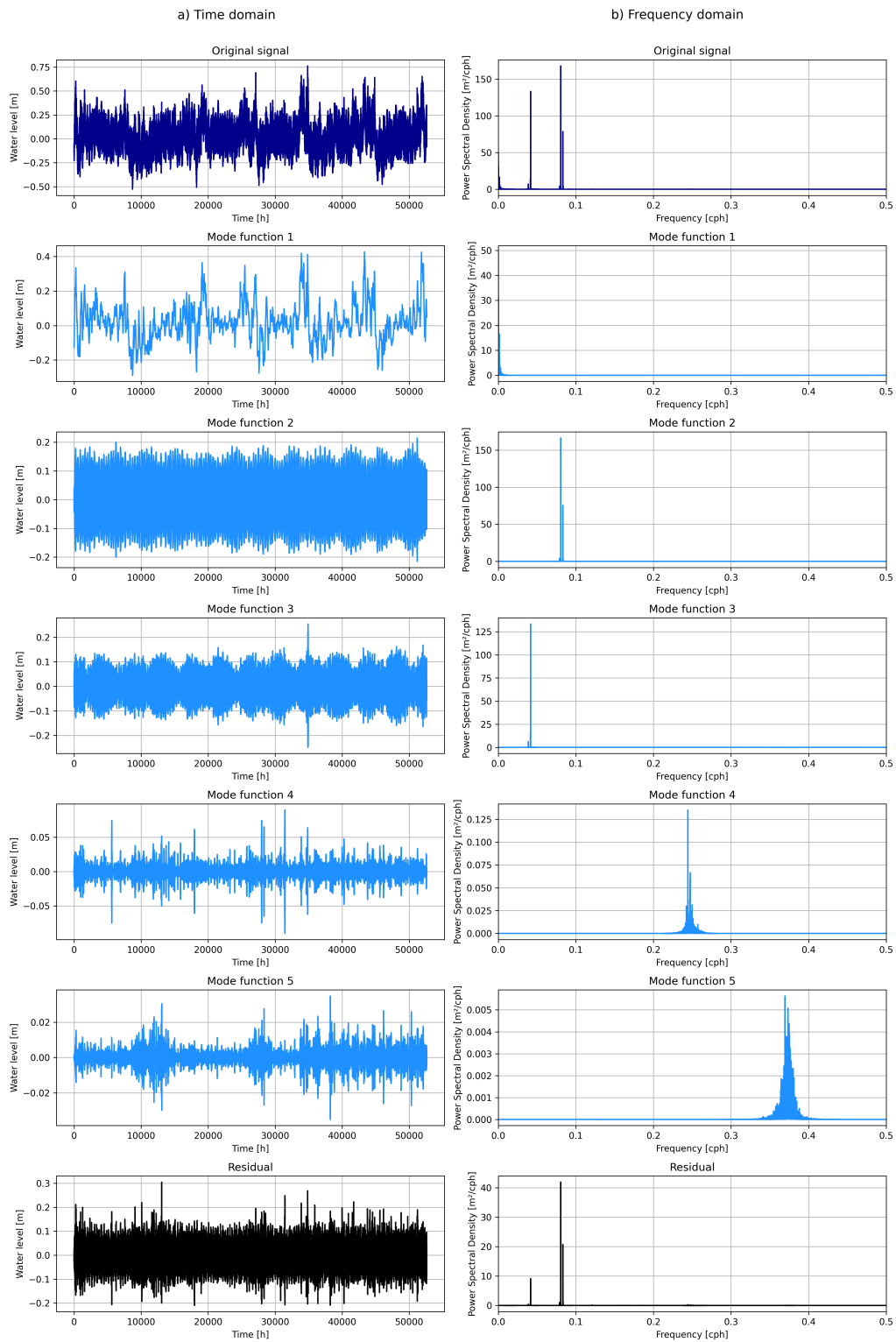
Figure D.7: Ušće station: VMD mode functions in time and frequency domain
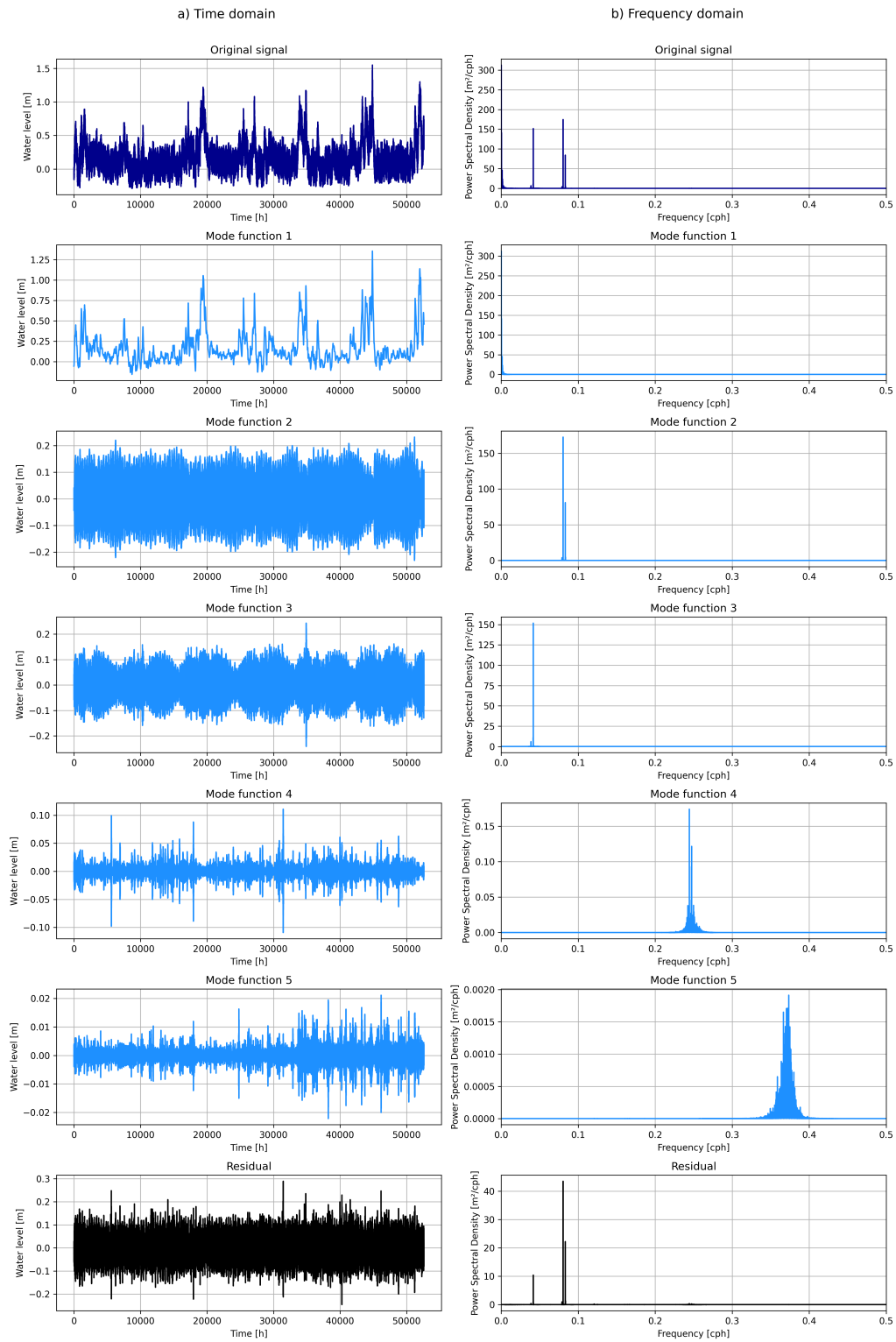
Figure D.8: Opuzen station: VMD mode functions in time and frequency domain
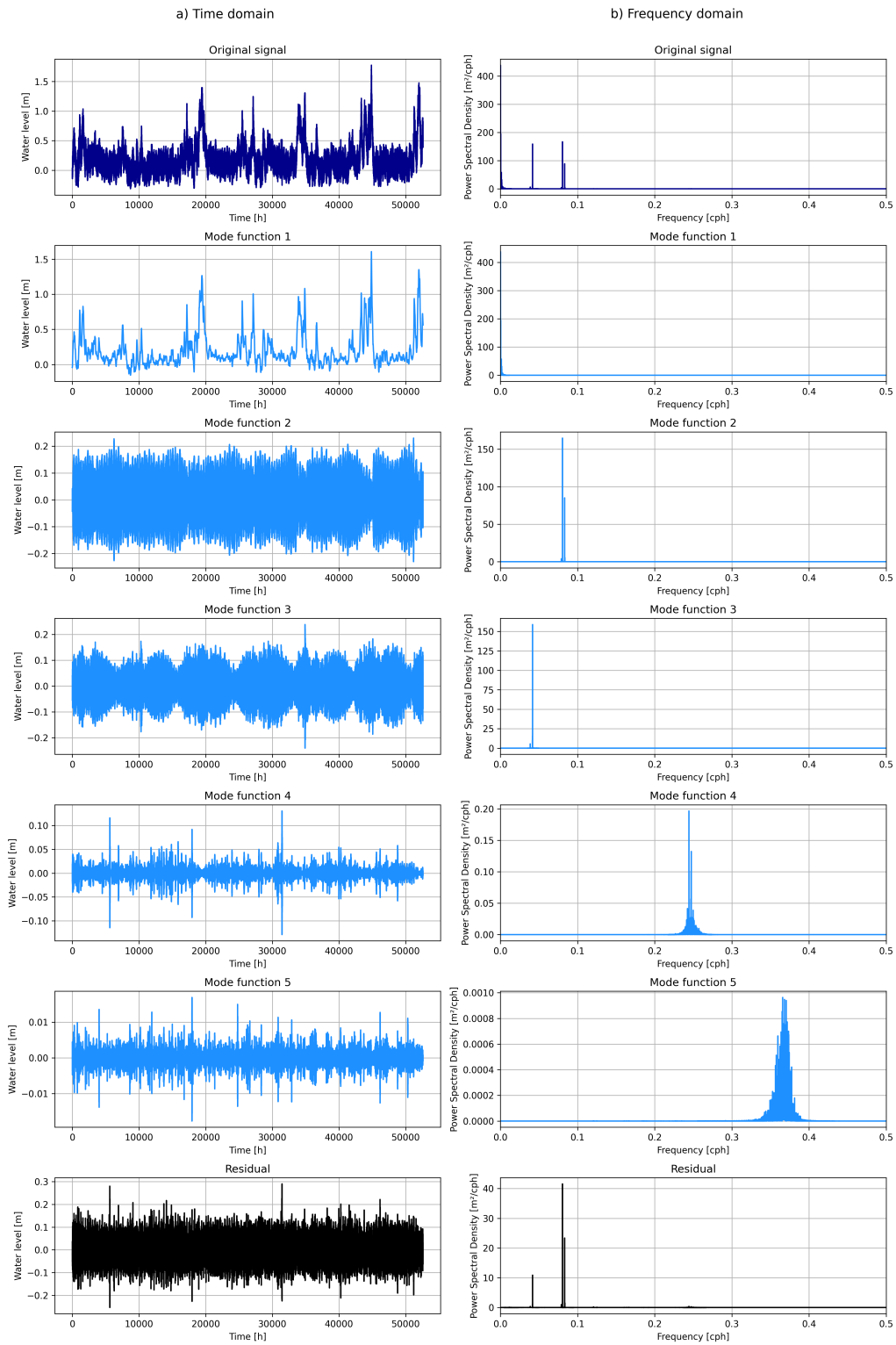
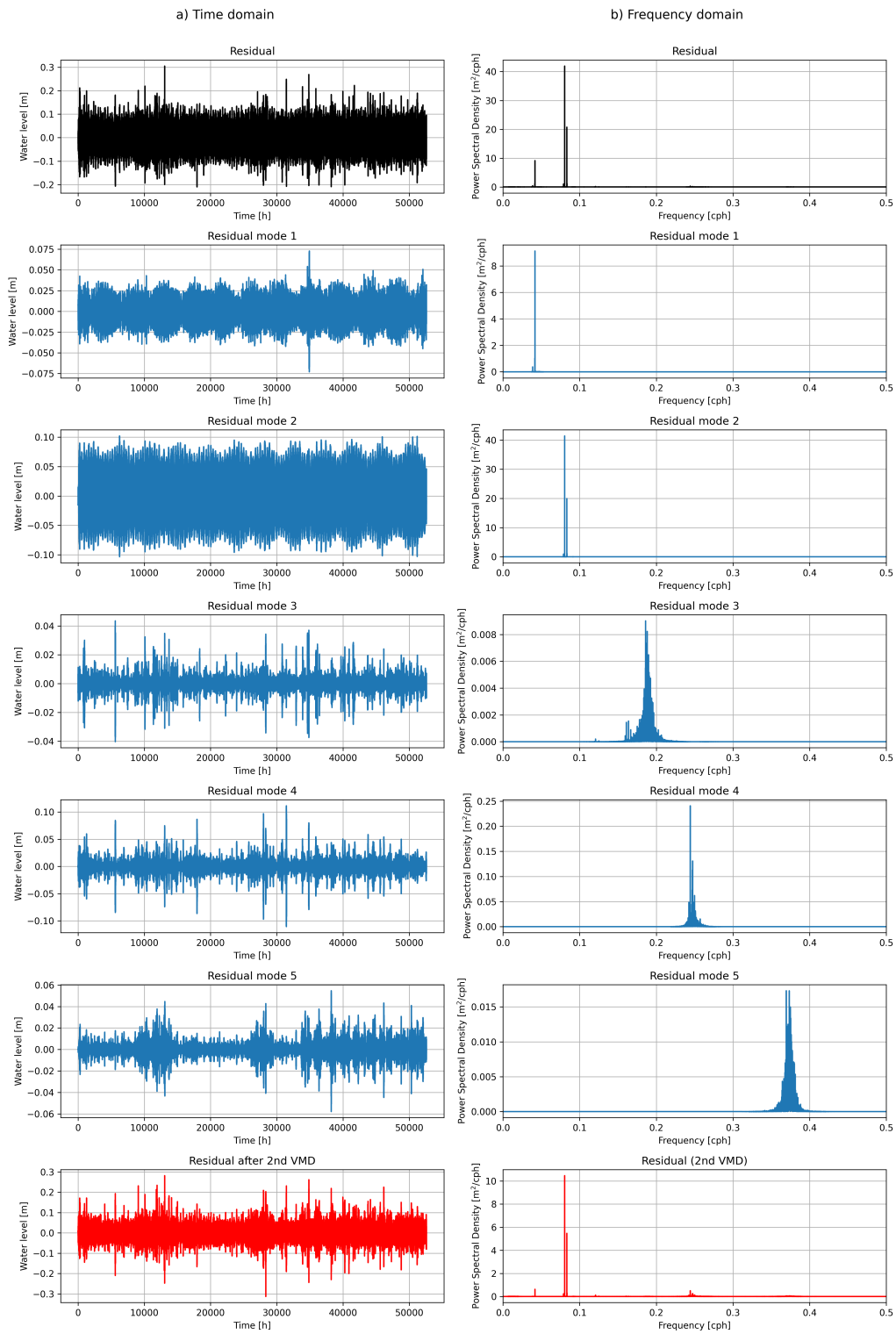**Figure D.9:** Norin station: VMD mode functions in time and frequency domain
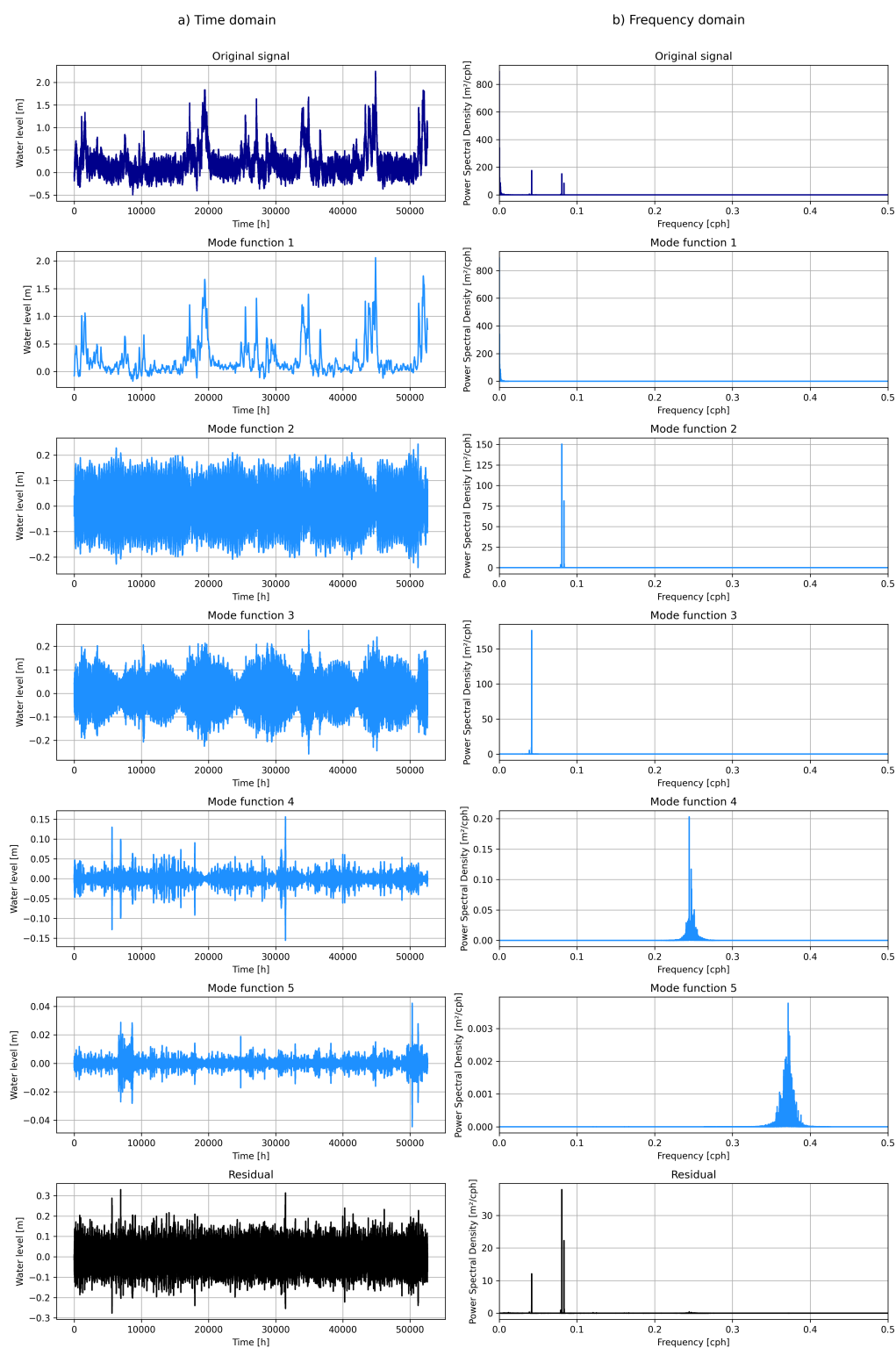
Figure D.10: Metković station: VMD mode functions in time and frequency domain

a) Time domain

b) Frequency domain

Figure D.11: Ušće station: Second-level VMD decomposition of the residual
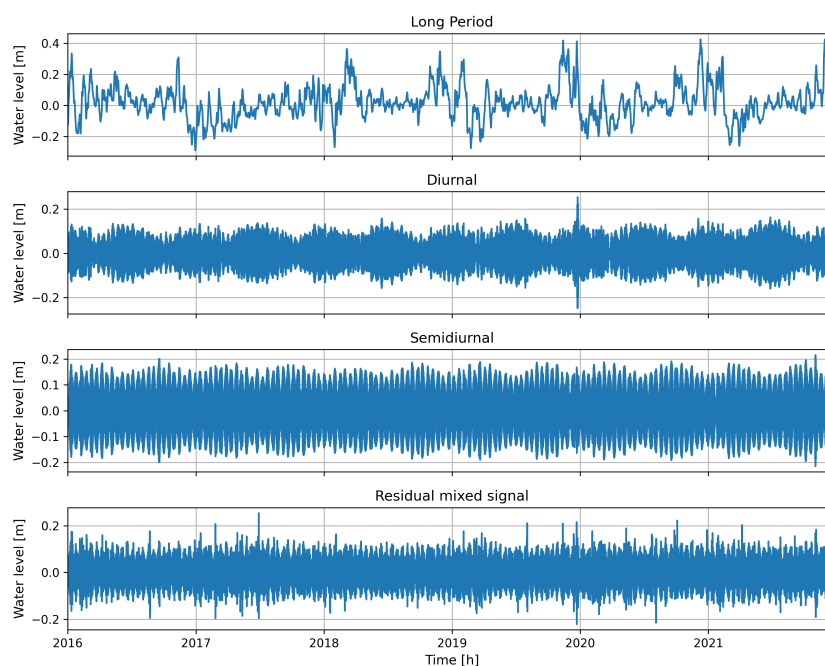
D.3 Tidal Constituents

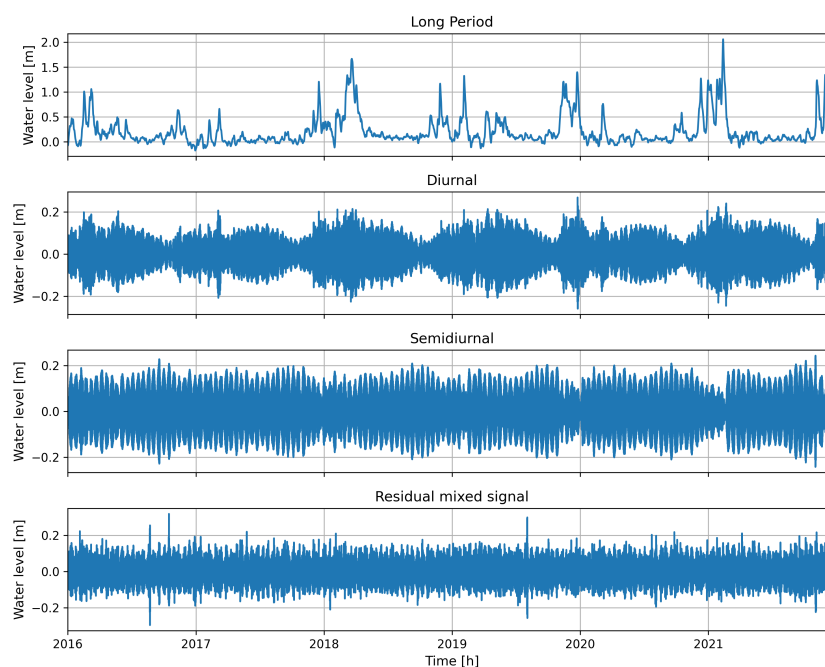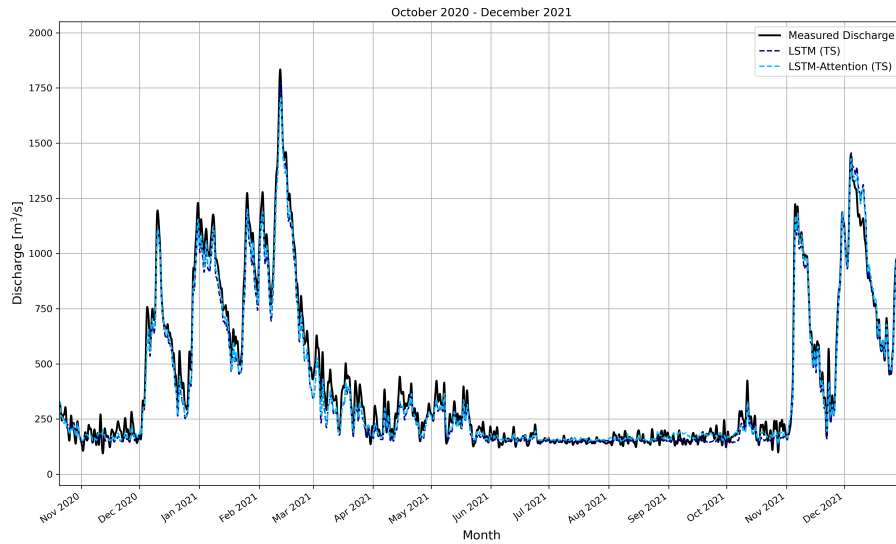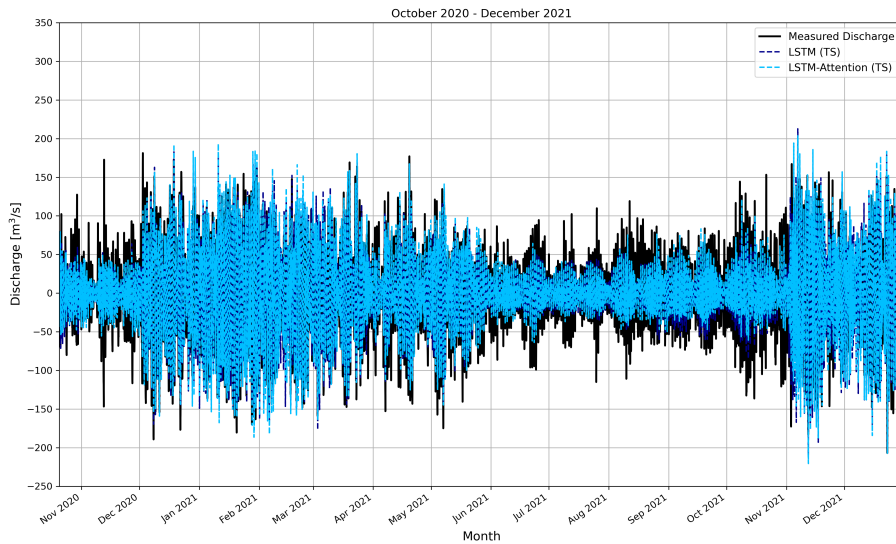**Figure D.12:** Ušće station: Tidal constituents

Figure D.13: Metković station: Tidal constituents

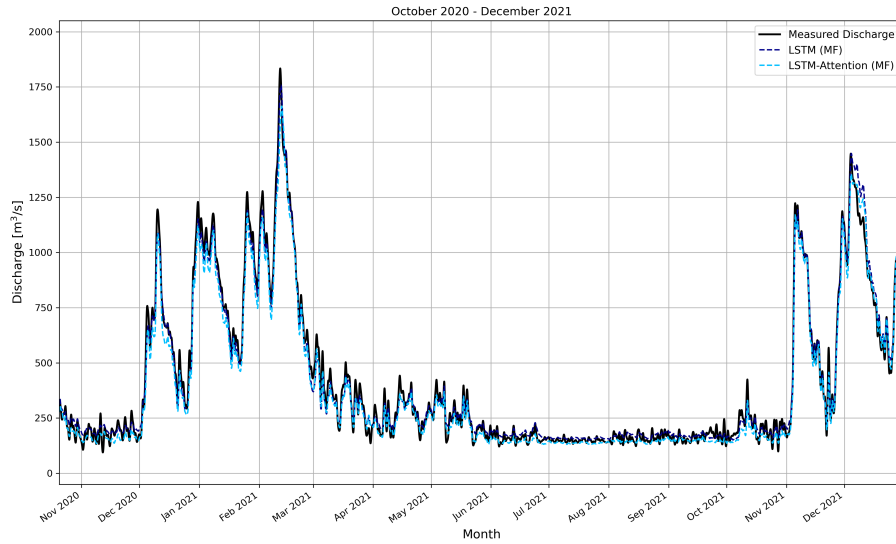## D.4   Intra-daily Oscillations Versus Interdaily Variations
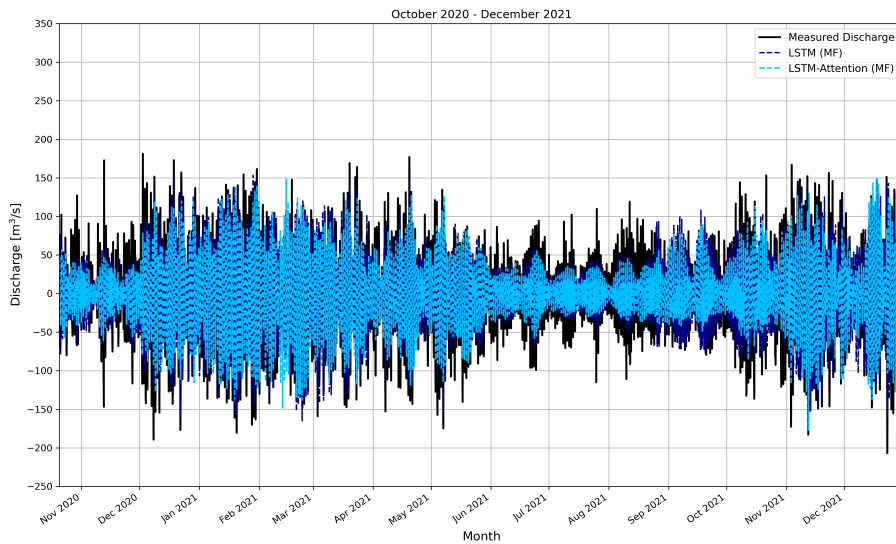

(a) Interdaily variations

(b) Intra-daily oscillations

Figure D.14: Butterworth filter (order=4) applied on time-series to examine models' performance on interdaily variations and intra-daily oscillations

(a) Interdaily variations

(b) Intra-daily oscillations

Figure D.15: Butterworth filter (order=4) applied on VMD classified mode functions and residual to examine models' performance on interdaily variations and intra-daily oscillations

E   Prediction of Hydrological Parameters

E.1   Optimization of Hyperparameters for Simulated Data Forecasting Task

Table E.1: Discharge Forecasting: Optimal hyperparameters obtained through k-fold grid search cross validation

Input dataset	Hyperparameter	Hyperparameter range	Optimal value
	Learning rate	[0.0001, 0.001, 0.01]	0.0001
	CNN - First layer	[64, 128, 256]	64
Time-series	*CNN - Second layer*	[64, 128, 256]	128
	LSTM hidden units	[64, 128, 256]	64
	Dropout	[0.1, 0.2, 0.3]	0.3
	Learning rate	[0.0001, 0.001, 0.01]	0.0001
	CNN - First layer	[64, 128, 256]	128
Spectrogram	*CNN - Second layer*	[64, 128, 256]	64
	LSTM hidden units	[64, 128, 256]	64
	Dropout	[0.1, 0.2, 0.3]	0.1
	Learning rate	[0.0001, 0.001, 0.01]	0.0001
	CNN - First layer	[64, 128, 256]	64
Spectrogram and Time-series	*CNN - Second layer*	[64, 128, 256]	64
	LSTM hidden units	[64, 128, 256]	64
	Dropout	[0.1, 0.2, 0.3]	0.1

Table E.2: Water Level Forecasting: Optimal hyperparameters obtained through k-fold grid search cross validation

Composite features	Hyperparameter	Hyperparameter range	Optimal value
	Learning rate	[0.0001, 0.001, 0.01]	0.0001
	CNN - First layer	[64, 128, 256]	64
Time-series	*CNN - Second layer*	[64, 128, 256]	64
	LSTM hidden units	[64, 128, 256]	128
	Dropout	[0.1, 0.2, 0.3]	0.3
	Learning rate	[0.0001, 0.001, 0.01]	0.001
	CNN - First layer	[64, 128, 256]	64
Time-series with Feature Engineering	*CNN - Second layer*	[64, 128, 256]	64
	LSTM hidden units	[64, 128, 256]	64
	Dropout	[0.1, 0.2, 0.3]	0.2
	Learning rate	[0.0001, 0.001, 0.01]	0.0001
	CNN - First layer	[64, 128, 256]	64
Spectrograms	*CNN - Second layer*	[64, 128, 256]	128
	LSTM hidden units	[64, 128, 256]	128
	Dropout	[0.1, 0.2, 0.3]	0.2
	Learning rate	[0.0001, 0.001, 0.01]	0.0001
	CNN - First layer	[64, 128, 256]	64
Spectrograms with Feature Engineering	*CNN - Second layer*	[64, 128, 256]	128
	LSTM hidden units	[64, 128, 256]	64
	Dropout	[0.1, 0.2, 0.3]	0.2

CURRICULUM VITAE

Anna Maria Mihel was born on March 21, 1996, in Levanger, Norway. Following the completion of her primary and secondary education, she enrolled in the undergraduate university study of Computing at the University of Rijeka, Faculty of Engineering, in 2014. In 2017, she enrolled in a graduate university study of Computing at the same faculty, with a specialization in Computer Systems. She completed her studies in 2020. In 2021, she began her postgraduate university doctoral study in the area of engineering sciences in the field of computer science, at the Faculty of Engineering and shortly after was employed as a research assistant by the faculty as a part of the program "Young Researchers Career Development Project - Training New Doctoral Students," funded by the Croatian Science Foundation (*DOK-2021-02-3306*).

During her postgraduate studies she participated in several projects, such as *"Compound flooding in coastal rivers in present and future climate"*, under the leadership of Assoc. Prof. Nino Krvavica, PhD, and *"Advanced Data Analysis Using Digital Signal Processing and Machine Learning Techniques"*, *"Computer-Aided Digital Analysis and Classification of Signals"*, and *"ABsistemDCiCloud"*, under the leadership of Prof. Jonatan Lerga, PhD.

She was also a part of the CEEPUS exchange program. This included her research stay at the Eötvös Loránd University, Faculty of Informatics, Department of Artificial Intelligence, and the University of Maribor, Faculty of Electrical Engineering and Computer Science, Intelligent Systems Laboratory.

LIST OF PUBLICATIONS

[1] Mihel, Anna Maria; Lerga, Jonatan; Krvavica, Nino. A brief introduction of machine learning techniques for discharge estimation in the areas of estuaries and tidal rivers. *My First Conference, 6th PhD Conference on Engineering and Technology Faculty of Civil Engineering*, University of Rijeka, 2022.

[2] Mihel, Anna Maria; Krvavica, Nino; Lerga, Jonatan. Long Short-Term Memory for Discharge Estimation in Coastal Neretva River. *2023 8th International Conference on Smart and Sustainable Technologies (SpliTech)*, 2023, pp. 16.

[3] Mihel, Anna Maria; Lenac, Kristijan; Krvavica, Nino; Lerga, Jonatan. Discharge Forecasting in Coastal Rivers Using CNN-LSTM Hybrid Approach. *2023 International Symposium ELMAR*, 2023, pp. 5560.

[4] Mihel, Anna Maria. Primjena hibridnog CNN-LSTM modela u svrhu predviđanja simuliranog protoka priobalne rijeke. *INFCON 2023*, 2023.

[5] Mihel, A. M.; Krvavica, N.; Lerga, J.; Oskoruš, D. Primjena umjetne inteligencije u procjeni hidroloških parametara priobalnih rijeka. *8. Hrvatska konferencija o vodama s međunarodnim sudjelovanjem - Hrvatske vode u proizvodnji hrane i energije - Zbornik Radova*, 2023, pp. 205212. Publisher: Hrvatske vode.

[6] Mihel, Anna Maria; Lerga, Jonatan; Krvavica, Nino. Estimating water levels and discharges in tidal rivers and estuaries: Review of machine learning approaches. *Environmental Modelling & Software*, Vol. 176, 2024, pp. 106033.

[7] Mihel, Anna Maria; Pečnik, Špela; Vrbančič, Grega; Lerga, Jonatan; Krvavica, Nino. Employing Feature Engineering for River Stage Forecasting to Improve Hybrid Model

Performance. *2024 International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*, 2024, pp. 16.

[8] Mihel, Anna Maria. Evaluacija točnosti predviđanja uzvodnog vodostaja priobalne rijeke do 24 sata unaprijed korištenjem kombiniranog pristupa inženjeringa značajki i hibridnog modela. *INFCON 2024*, 2024.

[9] Mihel, Anna Maria; Krvavica, Nino; Lerga, Jonatan. Regression-based machine learning approaches for estimating discharge from water levels in microtidal rivers. *Journal of Hydrology*, Vol. 646, 2025, pp. 132276.