

## Using statistical and machine learning approaches to describe estuarine tidal dynamics

Franziska Lauer \* and Frank Kösters 

Federal Waterways Engineering and Research Institute, Wedeler Landstraße 157, 22559 Hamburg, Germany

\*Corresponding author. E-mail: franziska.lauer@baw.de

 FL, 0009-0001-4780-735X; FK, 0000-0002-2302-5601

### ABSTRACT

Estuaries are ecologically valuable regions where tidal forces move large volumes of water. To understand the ongoing physical processes in such dynamic systems, a series of estuarine monitoring stations is required. Based on the measurements, estuarine dynamics can be described by key values, so-called tidal characteristics. The reconstruction and prediction of tidal characteristics by suitable approaches is essential to discover natural or anthropogenic changes. Therefore, it is of interest to inter- and extrapolate measured values in time and to investigate the spatial relationship between different stations. Normally, such system analyses are performed by deterministic numerical models. However, to facilitate long-term investigations also, statistical and machine learning approaches are good options. For a Weser estuary case study, we implemented three approaches (linear, non-linear, and artificial neural network regression) with the same database to enable the prediction of tidal extrema. Thereby we achieve an accuracy of 0.4–2.5% derivation (based on the RMSEs) while approximating measured values over 19 years. This proves that the approaches can be used for hindcast studies as well as for future analysis of system changes. Our work can be understood as a proof of concept for the practical potential of neural networks in estuarine system analysis.

**Key words:** feed-forward neural network, hindcast of tidal extrema (minimum and maximum), multiple regression modelling, Weser estuary, XAI

### HIGHLIGHTS

- Estuarine water-level characteristics (hydrograph) can be successfully approximated based on weighted machine-learning regression models.
- Comparison of different regression models with the same data set shows the potential of optimized neural networks.
- Initial feature testing and selection of ML variables provides information on physical processes driving estuarine water-level dynamics.

## 1. INTRODUCTION

Tidal hydrodynamics in estuaries and along coasts strongly influence both vulnerable ecosystems and economically important approach channels to sea ports. It is, therefore, not surprising that they have been thoroughly investigated by scientists (see Dyer 1997 for an overview). Although many physical processes are well understood, estuaries are highly dynamic systems and as such are still subject to ongoing research (MacCready & Geyer 2010). Changes in morphology and driving hydrodynamic forces frequently occur simultaneously. It is, therefore, not always straightforward to fully explain observations, e.g. *in situ* data of water level, on the basic principles of tidal hydrodynamics alone, but instead more complex explanations are needed (Hagen *et al.* 2022). However, these relatively small changes, in e.g. water level, are important to understand if the potential effects of climate change or anthropogenic measures are to be assessed.

Data from monitoring stations are needed to establish reliable information on the changes in tidal hydrodynamics. Since monitoring measurements are costly, the number of such stations is usually limited. The time series observations from these stations can be described by tidally driven key values, so-called tidal characteristic values (e.g. tidal high and tidal low water) (DIN 4049-3; Godin 1988; Lang 2003). A tidal characteristics analysis of water-level records is essential for a quantitative description of the estuarine hydrodynamic system. Tidal characteristic values are required for coastal engineering applications, e.g. to determine the design height of sea dikes.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence (CC BY 4.0), which permits copying, adaptation and redistribution, provided the original work is properly cited (<http://creativecommons.org/licenses/by/4.0/>).

Standard methods for predicting water levels are harmonic analysis (Kastens 2014; Hibbert *et al.* 2015) or hydrodynamic numerical modelling (NM). Recently, data-driven analyses have become increasingly important in hydrodynamics (Breiman 2001; Brunton *et al.* 2020) and, besides statistical methods, machine learning (ML) approaches are also employed more and more often. In many practical applications of water-level or wave-height forecasts, multiple ML variants are used, e.g. regression-based models (Günaydin 2008), artificial neural network (ANN) models (Bowles *et al.* 2005; Badejo & Udoudo 2014; Berbić *et al.* 2017), recurrent ANNs (NARX) (Arbian & Wibowo 2012; Rakshith *et al.* 2014; Tayel & Oumeraci 2018), or combinations of wavelet transformation plus ANNs (Erol 2011; Prahlada & Deka 2015; Dixit & Londhe 2016), to name a few. Comparisons of ML models with simpler statistical methods have been made, although not in research on hydrodynamics (Warner & Misra 1996; Li *et al.* 2001; Kumar 2005).

Despite the intensive efforts to reconstruct and predict the measurements of significant wave heights and water levels, important questions must be investigated and resolved more intensively. The length of data sets in many previous studies is limited to a relatively short time period and cannot reliably resolve multi-annual temporal effects. Hence, for more intra-seasonal variability, a comprehensive database is needed.

The applied ML models also need to be examined in detail, e.g. by comparing ANN models with further methods to determine whether MLR (Asma *et al.* 2012; Karimi *et al.* 2013), support vector machines (Berbić *et al.* 2017), or genetic programming (Ghorbani *et al.* 2010) could deliver similar results. Moreover, for ML methods and especially the approach with neural networks, the ANN settings (hyperparameters) need to be examined. Recent examples review the number of nodes in the hidden layers (Arbian & Wibowo 2012; Rakshith *et al.* 2014; Hibbert *et al.* 2015) or the activation function (Dorofki *et al.* 2012) or learning algorithm (Karimi *et al.* 2013). However, most of these investigations have been carried out as a trial-and-error procedure to search for the best combination. Multiple repetitions of ANN calculations, prior feature investigation and selection, or even hyperparameter optimization (HPO) have just become popular in recent years and slowly find their way into hydrodynamic applications. Therefore, some research groups developed software packages specifically for ML architecture optimization in environmental-related issues (Abbas *et al.* 2022; Amaranto & Mazzoleni 2023). Accurate design, appropriate management, and knowledge of relationships between the parameters affecting the performance of ML approaches are mandatory to make these methods more reliable (Huang 2003; Grabusts & Zorins 2015; Thomas *et al.* 2016).

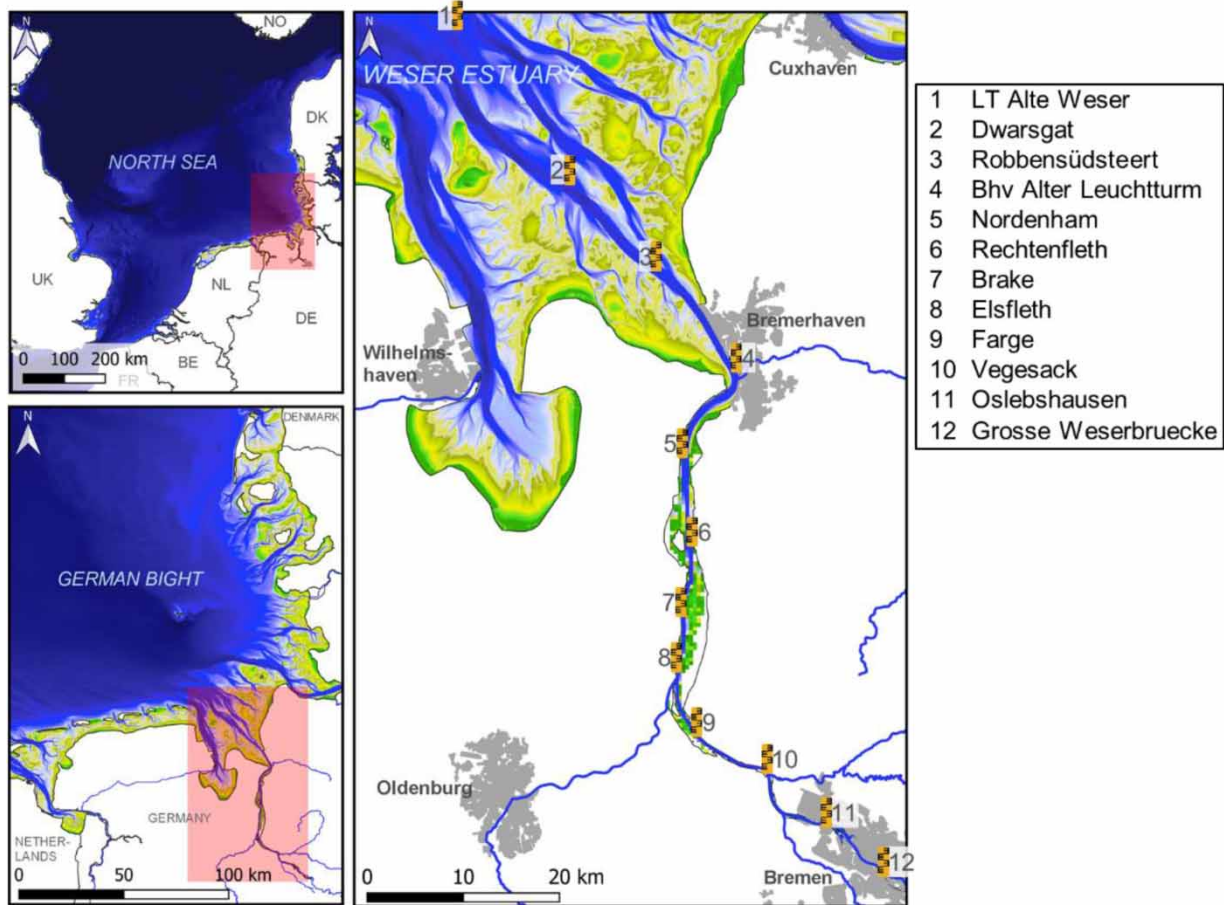
Hence, our goal in this study is to investigate how ML regression approaches can contribute towards developing a diagnostic and prognostic framework for hydrodynamic hindcast studies. This framework should then in the following allow us to carry out a reliable analysis of system changes. To do so, we developed a systematic approach to determine and compare how different statistical and ML approaches deal with exactly the same data set. An estuary located in the German Bight was selected as a study area for this purpose. With 12 monitoring stations operating in the Weser estuary area, parallel observations were made over a stretch of 120 km and covering 19 years. By assembling these hydrological data and comparing them with further features (predictors), we could investigate their functional relationships and select the appropriate predictors. To predict the water levels in the whole area, we specified the tidal extrema at 11 inner-estuary stations as targets and used measurements from stations further away as predictors. We implemented three regression methods (linear, non-linear regression (NLR) and ANN regression) with the same comprehensive database. For each regression method, we used the same two-step workflow with training and test cases to compare the output between the models, the stations and the targets. For the ANN approach, we performed an HPO to tune the ANN settings and then applied 100 ANNs at each station.

In our case, the main issue is how accurately the tidal characteristics can be reconstructed and predicted to discover system changes. This requirement for high accuracy can only be achieved if the different approaches are sufficiently investigated and optimized. Besides the inter- and extrapolation of measured values over a long time span, the spatial scale of the predictions is also of interest. For both tidal extrema – tidal high and low water – we achieved an accuracy of 0.4 – 2.5% derivation (based on root mean square error (RMSE) values) using the regression models while approximating the measured values over 11 stations and 19 years. Our work can be understood as a proof of concept for the practical potential of neural networks in estuarine system analysis.

## 2. MATERIALS AND METHODS

### 2.1. Study area

Our study area covers the tidally influenced part of the river Weser (Weser estuary) and the adjacent sea. The Weser estuary consists of the funnel-shaped outer estuary and a strongly channelized inner estuary (Figure 1). Tides enter the system from



**Figure 1** | Location of the Weser estuary and available water-level monitoring stations.

the North Sea and can propagate up to a tidal barrage (a weir) in Bremen. The tidal wave is distorted in the estuary by partial reflection, roughness, and shoaling effects (Lange *et al.* 2008).

Within the Weser Estuary, the local waterways and shipping administration (Waterways and Shipping Office (WSA) Weser-Jade-Nordsee) runs monitoring stations to continuously measure hydrological properties, e.g. water levels and conductivity. The processed data are available open source at <https://www.kuestendaten.de/Tideweser/DE/Startseite/Startseite-Portal-Tideweser-node.html> (last retrieved July 2023).

## 2.2. Database

### 2.2.1. Tidal characteristic values

The main properties of a tidal wave can be described by key values (e.g. tidal high water, tidal low water). These so-called tidal characteristic values are derived from water-level hydrographs. A single tide from such a hydrograph is defined from the first tidal low water to the second tidal low water. Some tidal characteristic values – tidal low water (tlw), tidal high water (thw), tide duration (td), ebb duration (ed), and flood duration (fd) – can be read directly from the hydrograph. Moreover, the tidal range (tr), tidal mean water (tmw), and tidal half water ( $t^{1/2w}$ ) can be calculated from the given data. Note that their occurrence time stamp is defined as the first low tide. This time stamp is also applied to the flood duration, while the ebb duration is dated to the occurrence time of the tidal high water. Details on the definition of tides can be found in the German DIN standard 4039 (1–3), and in Lang and Godin (DIN 4049-3; Godin 1988; Lang 2003). In one tide, eight tidal characteristic values (tlw, thw, tr,  $t^{1/2w}$ , tmw, fd, ed, td) can be extracted. For a semi-diurnal tidal cycle, this results in 704–708 individual values a year.

### 2.2.2. Target and predictor specification

Of these tidal characteristic variables, *tlw* and *thw* were selected as target variables for all further investigations as these are most commonly used in practical investigations. Both target variables were picked from 11 inner-estuary monitoring stations (nos. 2–12, Figure 1) to cover the tidal dynamics over the entire study area. Our aim is to reproduce these measured targets over the last 19 years precisely to within a few cm.

In order to achieve this objective with statistical approaches, suitable independent variables had to be found. These independent variables (also known as predictors) explain the investigated targets as well as possible without being influenced by them or other interdependencies.

Six tidal characteristic predictors from the Outer Weser estuary were taken as seaward boundary conditions (*tlw*, *thw*, *tr*, *tmw*, *ep*, *fp* from station no. 1, Figure 1). For the surface hydrology of the study area, the river discharge *Q* of the Weser and the main tributary (the river Hunte) were considered at the landward boundary. While a station 30 km upstream of the tidal barrier at Bremen is used for the river discharge of the Weser, the run-off of the tributary is located within the study area. Discharge of the Weser shows a wide range of variability (*Q* of 60 – 3,500 m<sup>3</sup>/s) and besides natural variability, it is also influenced by water management in the upper river sections. The daily run-off data were provided by the responsible authorities (WSA Verden) via the WISKI water management information system.

Furthermore, to consider meteorological aspects, the *u*- and *v*-component of the wind at 10 m height were used as predictors. These data were obtained from Germany's National Meteorological Service (Deutscher Wetterdienst, DWD) as hourly averaged data for the station Dwarsgat (station no. 2, Figure 1, Table 1).

### 2.2.3. Data assemblage

Assembling targets and predictors is critical in the preparation of all further analyses. The tidal low water of the tidal wave occurs first at the seaward station, so tidal parameters at stations further upstream can be related to this time stamp. In

**Table 1** | Overview of the target variables and predictors

Abbreviation	Meaning
<i>tlw</i> / <i>thw</i> gauge 12	Tidal low/high water target variable at station Grosse Weserbrücke
<i>tlw</i> / <i>thw</i> gauge 11	Tidal low/high water target variable at station Oslebshausen
<i>tlw</i> / <i>thw</i> gauge 10	Tidal low/high water target variable at station Vegesack
<i>tlw</i> / <i>thw</i> gauge 9	Tidal low/high water target variable at station Farge
<i>tlw</i> / <i>thw</i> gauge 8	Tidal low/high water target variable at station Elsflëth
<i>tlw</i> / <i>thw</i> gauge 7	Tidal low/high water target variable at station Brake
<i>tlw</i> / <i>thw</i> gauge 6	Tidal low/high water target variable at station Rechtenflëth
<i>tlw</i> / <i>thw</i> gauge 5	Tidal low/high water target variable at station Nordenham
<i>tlw</i> / <i>thw</i> gauge 4	Tidal low/high water target variable at station Bremerhaven Alter Leuchtturm
<i>tlw</i> / <i>thw</i> gauge 3	Tidal low/high water target variable at station Robbensüdsteert
<i>tlw</i> / <i>thw</i> gauge 2	Tidal low/high water target variable at station Dwarsgat
<i>tlw</i> gauge 1	Tidal low water predictor from station Leuchtturm Alte Weser
<i>thw</i> gauge 1	Tidal high water predictor from station Leuchtturm Alte Weser
<i>tr</i> gauge 1	Tidal range predictor from station Leuchtturm Alte Weser
<i>tmw</i> gauge 1	Tidal mean water predictor from station Leuchtturm Alte Weser
<i>fp</i> gauge 1	Flood duration predictor from station Leuchtturm Alte Weser
<i>ep</i> gauge 1	Ebb duration predictor from station Leuchtturm Alte Weser
<i>Q</i> dis.	Volumetric daily flow rate (discharge) predictor of the Weser River obtained in Intschede
<i>Q</i> trib.	Discharge predictor of the Weser tributary Hunte
<i>u</i> -wind gauge 2	East-west wind component predictor at station Dwarsgat
<i>v</i> -wind gauge 2	North-south wind component predictor at station Dwarsgat

addition, the daily discharge values are interpolated to this low-tide time series and the wind components are processed accordingly. This results in a data matrix with ten predictors with 13,410 data points (705 tides per year) over 19 years. For the two different targets, we slightly modified the  $10 \times 13,410$  data matrix by using the same predictor from station no. 1. These two data matrices were applied (target-sensitive) each time, while at each of the 11 stations of the inner estuary, all analyses were carried out individually. To train each model, a period of 3 years with 2,116 data points (July 2000 to June 2003) was extracted from the two data matrices and the 22 time series of the target variables. The selected training period covers a large number of maxima in both data matrices, e.g. different discharge situations, and therefore prevents extrapolation in the test step (Figure 2).

### 2.3. Feature investigation and selection of the predictors

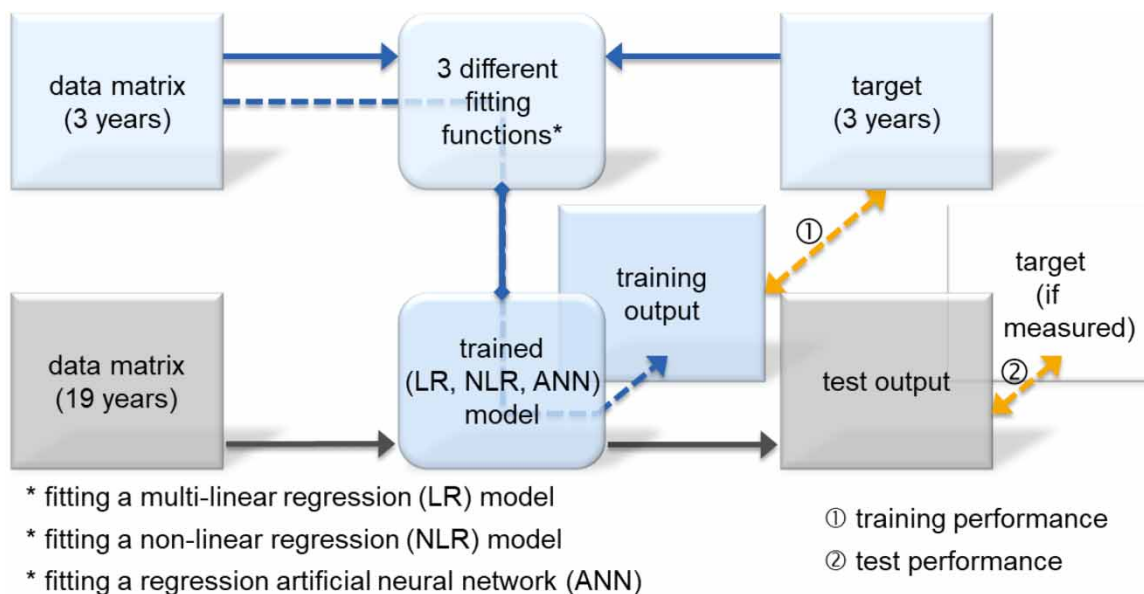
To test the independence of the 10 predictors, feature investigation was performed by calculating the variance inflation factor (VIF, Equation (1)). A common limit above which multicollinearity starts is a VIF of 5 (James *et al.* 2017). This indicates that the remaining predictors explain less than 80% of the variance for the tested predictor  $i$ . In detail, this can be seen if the VIF equation is converted to  $R^2$ :

$$\text{VIF}_i = \frac{1}{1 - R_i^2} \leq 5 \text{ is the same as } R_i^2 \leq 0.8 \quad (1)$$

where VIF is the variance inflation factor and  $R^2$  is the coefficient of the linear determination.

In conclusion, the best possible achievable value for the VIF is 1, which equals  $R^2 = 0$  and therefore no linear relationship exists between the tested predictors. To identify the extent of multicollinearity, the VIF was calculated for all 5,110 combinations. Due to this high figure in the results section, we only describe three specific cases: (I) the pairwise linear comparison (90 combinations), (II) the 'leave one out' variant with 10 combinations, and (III) one special case.

Further feature tests were carried out by associating the given predictors with the target variables. The relationships between the predictors and the targets were examined using statistical (filtering) methods to estimate the relevance of each predictor. Since some physical relationships may not be linear, other interconnections must also be considered. For



**Figure 2** | Two-step workflow (blue/grey) for three regression models (\* = LR, NLR, ANN). In the blue fields, the model fitting is done by training with values from 3 years (2,116 data from July 2000 to June 2003). In the grey fields, the final trained model is used to approximate data for a period of 19 years (13,410 data from January 2000 to December 2018). The performance (Ⓢ: training, Ⓢ: test) of both steps can be determined using the RMSE by comparing the actual measured values (target) to the individual regression model output. This workflow was used 22 times (2 targets at 11 stations) for each regression approach.

this purpose, in addition to station-wise linear Pearson correlation-based calculations (PCC), the relationships were also investigated via the ranking of the supervised relief algorithm.

The relief algorithm (Equation (2)) identifies and scores predictors based on the value difference between nearest neighbour instance pairs. If the predictor's response to the target variable is similar or the same each time, the predictor is upgraded ('hit'); conversely, the predictor is devalued ('miss'). This calculated weighting can then be used to create a ranking of the most important predictors. (Kira & Rendell 1992; Robnik-Sikonja & Kononenko 1997, 2003; Guyon & Elisseeff 2003)

$$[\text{idx}, \text{weights}] = \text{relief}(X, \text{tcv}_{\text{gauge } x}, k) \quad (2)$$

where  $X$  is the data matrix of the 10 predictors;  $\text{tcv}$  is the tidal characteristic value: either tidal high or low water (at station  $x$ );  $k$  is the integer for nearest neighbours;  $\text{idx}$  represents the indices of predictors ordered by predictor importance;  $\text{weights}$  define the weights of predictors.

Each relief analysis was compiled station- and target-wise and the resulting weights were ranked. Hence, each time the  $k$ -value of the relief algorithms went through a 1–200 for-loop, we observed that the weighting remained constant from a  $k$ -value of around 50–75 on.

#### 2.4. Regression model workflow for data approximation

The key aspect of this paper is the best possible approximation of the target variables with the chosen predictor data matrix. Three statistical methods were compared to achieve this objective: linear regression, NLR, and ANN regression. All three methods follow the same procedure, which is shown schematically in Figure 2. The schematic illustrates the distinction between training, where coefficients or node weights are adjusted (blue workflow, Figure 2), and the test case, where the weighted model is applied to replicate any data (grey workflow, Figure 2).

For the training process, a 3-year segment is extracted from the data matrix and set against each target variable separately. In these 22 regression model trainings, either the coefficients are adjusted (linear and non-linear regression) or, in the case of the neural networks, the weights of the nodes in the hidden layers are optimized (Figure 2, blue workflow). Training performance can be determined via the RMSE by comparing the final training output with the given target variable (Figure 2, yellow arrow (1)).

Here, the RMSE is an appropriate error index, as multiplication within the calculation emphasized large errors. In our case, it is important to describe all data with a similar (small) deviation, so penalizing outliers is beneficial to weed out bad regression models.

After training, the 3-year training data segment is integrated into the 19-year data frame, and the predictor data matrix is used in each weighted regression model to approximate the target variable over the entire time period (Figure 2, grey workflow). Here again, the performance comparison for testing is possible if, as in this case, measured target values are available (Figure 2, yellow arrow (2)). Thus, the test RMSE consists partly of the training RMSE plus the deviation of the unseen data.

For better comparability across the models and stations, the individual training or test RMSEs were divided by the range of measured variables and the results were stated in percent. These percentages were used to explain the precision of each model prediction. The choice of error metric potentially introduces to some degree a bias of our analyses. In order to rule that out, different error metrics have been tested (see Supplementary Material for a comparison with the Kling–Gupta efficiency metric (Gupta *et al.* 2009)).

This two-step workflow – training and test step – was employed with each of the three regression model approaches to carry out 22 analyses. The three fitting methods are discussed in the following sections.

##### 2.4.1. Multiple linear regression model

Multiple linear regression (MLR) is a statistical technique that uses several explanatory variables, which are also called predictors, to approximate the values of a target variable. Here, the same five preselected predictors are used in each analysis and provided with linear coefficients according to the target variable used. These five predictors are the tidal extrema ( $\text{tcv}$ ) and the tidal mean values ( $\text{tmw}$ ) at the seaward estuary limit, the wind at another seaward station ( $u, v$ ) and the river discharge at the landward limit ( $Q$ ). The general equation for each multiple linear regression model is:

$$\text{tcv}_{\text{gauge } x} = 1 + a * \text{tcv}_{\text{gauge } 1} + b * Q_{\text{discharge}} + c * \text{tmw}_{\text{gauge } 1} + d * u_{\text{gauge } 2} + e * v_{\text{gauge } 2} \quad (3)$$

where  $a, b, c, d, e$  represent the linear regression coefficients;  $tcv$  represents the tidal characteristic value: either tidal high or low water (at station  $x$  and no. 1);  $Q$  is the river discharge;  $tmw$  is the tidal mean water (at station no. 1);  $u, v$  is the wind vector components (at station no. 2).

This MLR equation (Equation (3)) was used in the two-step workflow of Figure 2 to identify the training and test performance for both target variables at any of the 11 inner-estuary stations.

#### 2.4.2. Multiple NLR model

A non-linear equation was tested as a second regression model approach. In contrast to the linear regression described above, a multiple non-linear regression model (MNLr) can specify arbitrary relationships between predictors and a target variable. In order to prevent overparameterization and to ensure simple cross-model comparison, a superscript regression coefficient was added only to the river discharge predictor. The idea of this equation is based on Niemeyer (1998) and was then extended with  $tmw$  by Kastens (2008), and now the wind vectors are included in our study. The equation for each NLR model is therefore as follows:

$$tcv_{gauge\ x} = a * tcv_{gauge\ 1} + b * Q_{discharge}^n + c * tmw_{gauge\ 1} + d * u_{gauge\ 2} + e * v_{gauge\ 2} + z \quad (4)$$

where  $a, b, c, d, e, z$  represent (linear) regression coefficients;  $n$  is the exponential regression coefficient;  $tcv$  is the tidal characteristic value: either tidal high or low water (at station  $x$  and no. 1);  $Q$  is the river discharge;  $tmw$  is the tidal mean water (at station no. 1);  $u, v$  represent wind vector components (at station no. 2).

This MNLr equation (Equation (4)) was applied in the same context of Figure 2 to identify the individual training and test performance for each target at each station.

#### 2.4.3. ML regression model

ML methods are a statistical alternative to the simple regression methods described above. Artificial neural networks (ANN) are particularly recommended as they inherently cover a wide range of regression procedures in the way the nodes are activated. The configuration of the activation functions, the learning algorithm as well as the chosen ANN topology (the number of nodes and hidden layers) determine how the data are fitted into an ML regression model.

The learning algorithm applied in an ANN can produce different node weights and correspondingly different results with each iteration, even though the same data set is used. In addition, assigning random numbers to the initial weights can force the learning algorithm to run into different extrema of the solution space. This makes it difficult to reproduce a single ANN model. We overcame this limitation by training 100 ANN models for each analysis and selecting the best two-thirds (66%). This considerably increases the robustness of the ANN approach.

Subsequently, the same five predictors were again used for each respective target variable. We decided to carry out the HPO of the activation functions and the number of nodes. To avoid information losses, it was necessary to prior limit the HPO by two specifications: (I) the neural networks used should consist of two hidden layers (and four layers in total) and (II) the number of nodes in the first hidden layer (HL) should not be smaller than in the second.

In practice, HPO using a Bayesian method then leads to the selected number of nodes as given in the two ANN equations below ((5):  $tlw$ -topology: 5–9–8–1; (6):  $thw$ -topology 5–13–11–1). The activation of the nodes was specified by sigmoid transfer functions and the weights were adjusted using the Levenberg – Marquardt learning algorithm.

$$ANN\ Model_{tlw} = data\ matrix_{5\ variables} - HL\ 1_{9\ nodes} - HL\ 2_{8\ nodes} - tlw_{gauge\ x} \quad (5)$$

$$ANN\ Model_{thw} = data\ matrix_{5\ variables} - HL\ 1_{13\ nodes} - HL\ 2_{11\ nodes} - thw_{gauge\ x} \quad (6)$$

where five variables represent  $tcv$  and  $tmw$  at station no. 1;  $Q$  is the discharge;  $u$ - and  $v$ -wind vectors at station no. 2. HL is the hidden layer of the ANN with the optimized number of nodes.

To train and test the ANN approach, we applied the same workflow as described above, changing only the number of computed models (100 instead of 1). For every ANN training, a common termination criterion is applied: With the sixth iteration, after the minimum validation error has been reached, the adjustment of the weights is stopped. Due to this abortion, the training output can never 100% match the observed values of the target variable. For ANN training, we chose a split of 80% (training), 10% (validation), and 10% (test). Further, each time, a cross-validation for partitioning of the 2,116 data was

performed, to ensure that each ANN used a different subset of the 2,116 training data to validate the different epochs and abortion criteria. Figure 2 shows that these 2,116 data are included later on in the test workflow (marked in grey), thus the 10% test data split in training is actually not necessary.

The test step (Figure 2) was conducted with the best two-thirds final weighted ANN models from training. For removing one-third of the ANN models, the station- and target-specific training RMSE served as a decision metric.

### 3. RESULTS

Within the results, we first look at the database by investigating the relevance of the predictors and selecting them (Section 3.1). Then we perform multiple regression analyses with these selected data to approximate the measured values from each target as well as possible (Section 3.2).

#### 3.1. Feature investigation and selection

##### 3.1.1. Independence of the assembled predictors

To estimate the independence of the given ten predictors, the linear correlation coefficient  $R^2$  and VIF (VIF; Equation (1)) were calculated for all 5,110 possible predictor combinations.

In order to determine the extent of multicollinearity, three subsets – (I) all 1 vs. 1 combinations (pairwise correlation); (II) all 9 vs. 1 combinations (leave-one-out variant) and (III) a special 2 vs. 1 combination – were examined in more detail below.

The pairwise  $R^2$ -calculation (subset I) and the individual pairwise VIF did not give indications of a high level of multicollinearity (no VIF values above 5). With 71% of the variance explained, the highest VIF (3.42) was found in the combination for the thw- to tmw-predictor.

On the other hand, looking at the VIF values for the leave-one-out variant (subset II) revealed that the data matrix does contain structural collinearities. In the six multi-linear regression (MLR) calculations, the left-out ebb- and flood duration (VIF of 1.6 and 1.8), the run-off of the main channel (VIF: 2.9), as well as the tributaries' run-off (VIF: 3.0) and the two wind components (VIF of 1.5 and 2.2) show that they are independent predictors. On the other hand, testing the predictors of the hydrograph (troughs and peaks, range, mean) reveals high VIF values in their individual 9 vs. 1 MLR calculations (tlw: 17.3, thw: 272.3, tr: 102.8, tmw: 178.9). In other words, the predictors of tlw, thw, tr, or tmw could be substituted by the remaining nine predictors of the predictor data matrix. Although the degree of multicollinearity is high here, in the associated single calculations, this could not be attributed to direct individual correlations.

One remarkable triangular constellation is the combination of 2 vs. 1 predictor for thw (VIF: 123.8), tr (VIF: 36.6) and tmw (VIF: 100.5). In all three MLR calculations, the remaining two predictors explain more than 97% of the variance of the tested predictor.

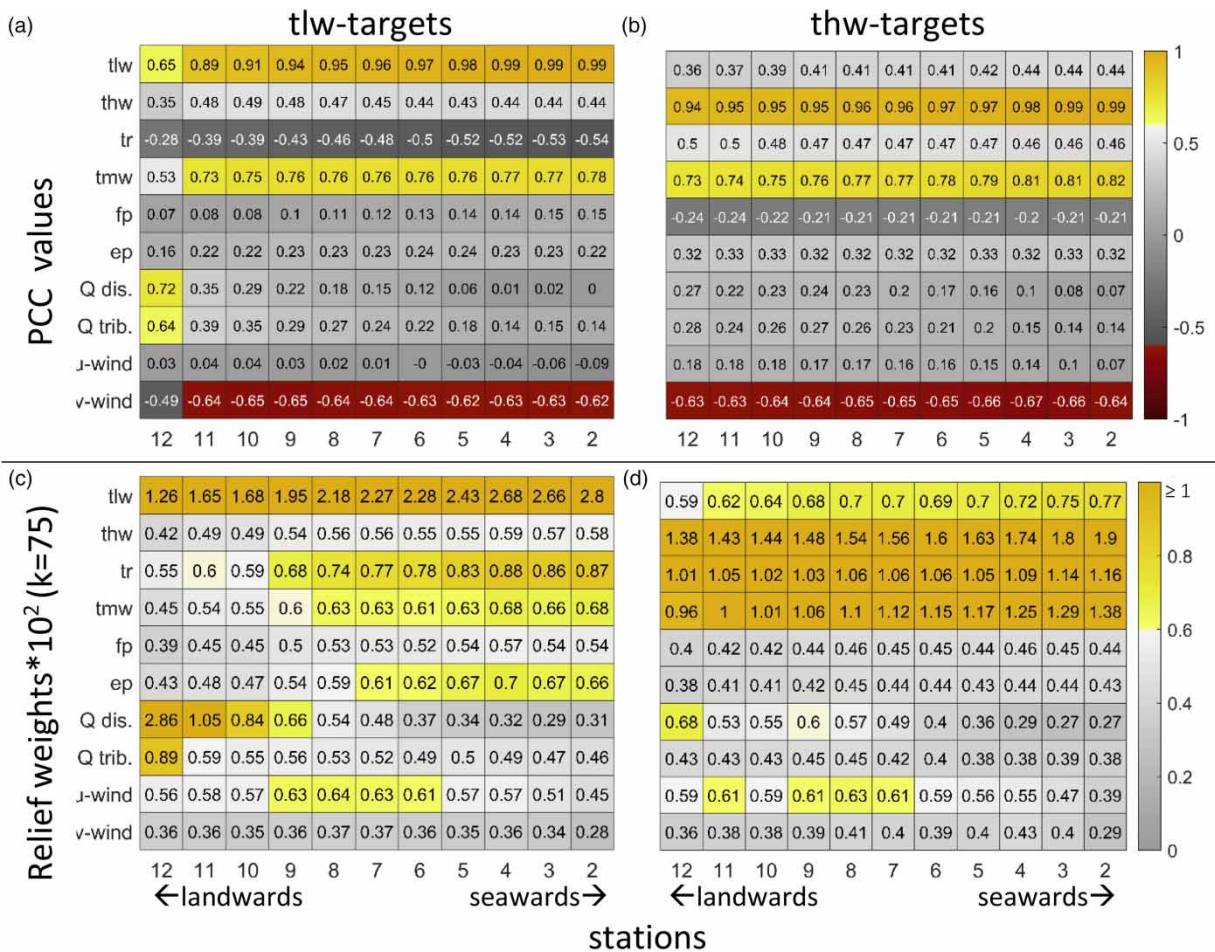
In summary, this means that the VIF calculations for the applied ten predictors data matrix revealed a certain degree of multicollinearity, which, however, only occurs in higher combinations ( $\geq 2$  vs. 1).

##### 3.1.2. Linear correlation between predictors and targets

Calculated Pearson correlation coefficients (PCC) for all inner-estuary stations (nos. 2–12) were displayed and compared (Figure 3(a) and 3(b)) to illustrate the direct dependence of the two target variables (tlw and thw) on the ten predictors.

This analysis shows that for most stations, both target variables are correlated with the same predictors. In detail of Figure 3(b), the target variables thw are strongly correlated with thw-predictor at station no. 1, which proves that these values move in unison. Hereby, the quality of the correlation decreases with increasing distance from the locations of this boundary condition. In addition to these apparent linear correlations, the PCC values are followed by correlations with the area-equal tidal mean water (tmw), the north–south wind vector ( $v$ -wind), and tidal range (tr) predictor. It is notable that the thw-rankings are almost the same for all stations (Figure 3(b)).

Figure 3(a) shows that these patterns are quite similar for the target variables of tlw. First of all, the tlw-targets are correlated with the tlw-predictor at station no. 1. The following correlations (tmw,  $v$ -wind, and tr) are largely the same as for the thw-targets. In contrast to thw and due to its geographic proximity, the up-estuary station no. 12 is strongly correlated with the river discharge predictor ( $Q$  dis.).



**Figure 3** | Pearson correlation coefficient (PCC, subplot (a,b)) and relief weights with 75 for the integer of the nearest neighbours ( $k = 75$ , subplot (c,d)) for each of the 11 inner-estuary stations (x-axes: no. 12 to no. 2) correlated with the given 10 predictors (y-axes). For better comparability, the relief weights have been enlarged by  $\cdot 10^2$ . Both correlation procedures were done column-wise separately for the respective tlw-target variables (left side: subplot (a,c)) and thw-target variables (right side: subplot (b,d)).

### 3.1.3. Rank analyses between predictors and targets (relief analysis)

Rank-based relief attribute calculations were carried out (Figure 3(c) and 3(d)) to compare the linear correlation PCC results of the previous section with further (unknown) functional relationships between predictors and targets. Further relevance of the predictors is revealed by examination of the relief calculations. Compared to the PCC analysis (Figure 3(a) and 3(b)), a higher cross-station heterogeneity is noticeable here, especially for the tlw-target variables (Figure 3(c)).

The following observations from the PCC can be confirmed using the relief calculations: (I) the high importance of the same predictor with the corresponding target variables (tlw with tlw, Figure 3(c); thw with thw, Figure 3(d)), (II) the inverse correlation of the opposite extrema (thw with tlw, Figure 3(c); tlw with thw, Figure 3(d)) and (III) the loss of importance of the river discharge predictor in the seaward direction. It is interesting to note that in the up-estuary stations, the river discharge predictor has become more relevant and is thus apparently subject to an NLR (Figure 3(d)).

In Figure 3(c), the arrangement of the other tidal characteristic predictors (tr, tmw, ep, fp) around the ranks 4–8 differs only in nuances in the relief weights. For example, the tidal range is preferred to the tidal mean water predictor because another functional relationship could possibly be better represented here. Looking at the flood and ebb duration, differences from previous linear PCC analyses become apparent. In the relief analyses, the ebb duration predictor is more decisive for the tlw-targets (Figure 3(c)). For the thw-targets, the flood duration gains more importance compared to the ebb duration predictor (Figure 3(d)).

A change in relevance is also observed for the wind component predictors. In contrast to the previous linear PCC analyses, the two wind predictors switch their position in the relief analyses (Figure 3(c) vs. 3(d)). While the predictor of the  $u$ -wind vector is now placed higher, the rankings for  $u$ -wind and  $v$ -wind are generally lower and more inhomogeneous than they were in the linear PCC analyses.

These correlation analyses provide important information for the regression modelling performed in the next step. In the following regression modelling, a global solution in predictor selection for each target variable could be wise, even though there might be better predictor combinations station-wise. Therefore, two aspects should be considered here: (I) the functional relationships that were ranked highest and (II) the subset that contained a suitable combination of an ideally low number of independent variables. Both analyses – Pearson correlation coefficient and relief attribute ranking – showed that the compilation of these 10 predictors is justified in terms of their relationships with the two target variables. The multicollinearity testing also demonstrated the independence of the predictors, although some combinations need to be selected with care.

Based on these analyses, we decided on the following predictors. The same-type predictor for the respective target variables has been shown to be essential. Computed from the data,  $tlw$  was the best predictor for the  $tlw$ -targets and  $thw$  for the  $thw$ -targets. Since  $tr$  and  $tmw$  can be considered almost equivalent, we have included the  $tmw$ -predictor instead of  $tr$ . In order to avoid a higher collinearity, no further predictors of the water-level hydrograph were selected. To cover another aspect, the river discharge with its non-linear connection should also be considered. Despite the strong fluctuations in the rankings, the two wind components were chosen because they bring an absolute added value with their low multicollinearity. The combinations of these five predictors show no striking VIF values in their respective calculations (VIF =  $tlw$ : 2.8/ $thw$ : 3.6,  $tmw$ : 4.3,  $Q$ : 1,  $u$ -wind: 1,  $v$ -wind: 2.2).

## 3.2. Application of multiple regression models

### 3.2.1. Regression modelling

Based on the previous predictor selection, three regression methods for the approximation of the two target variables were set up, calculated and evaluated – linear regression (LR), non-linear regression (NLR), and ANN regression.

The training of all regression model fittings made use of the same 3-year period for the selected five predictors and each target variable. The final training outputs given through the fitted regression coefficients or weighted nodes of the regression models were then compared with the measured values by calculating the root mean square errors (training RMSEs).

Figure 4 shows this training quality based on the RMSE values in cm for all regression models grouped by the two target variables  $tlw$  and  $thw$ . The effect of selecting only a subset (the best 66%) of the 100 ANNs, calculated separately for each of the 22 targets, is also shown.

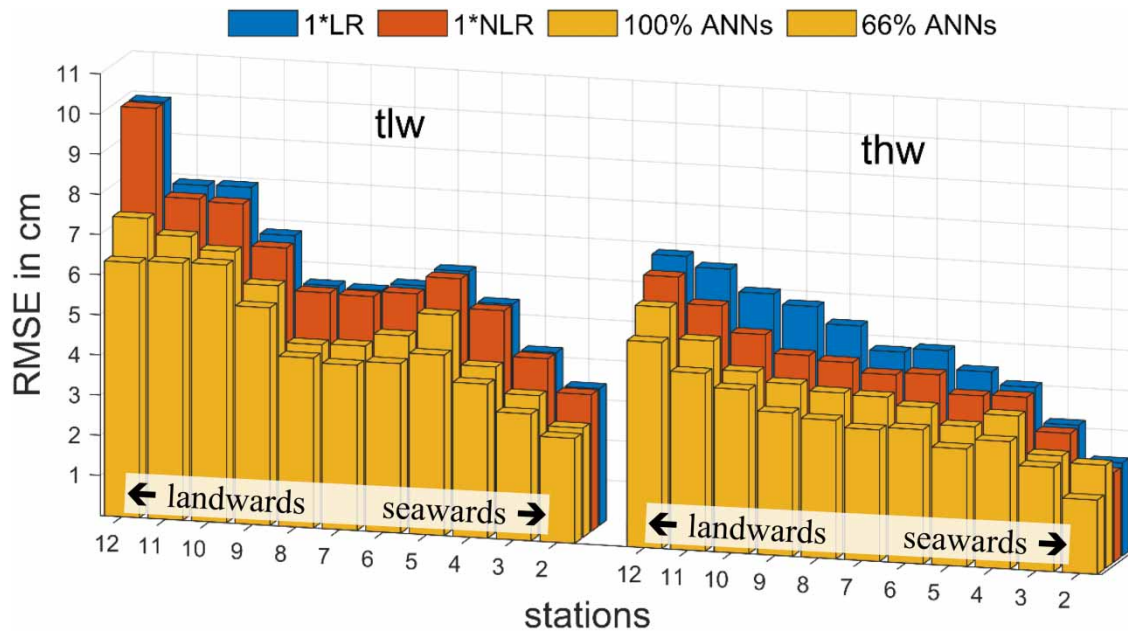
Figure 4 shows that for all stations the ANN approximations have lower error indices than the other two regression methods. Furthermore, the reduction from 100 ANNs to two-thirds seems reasonable, even if the differences between the yellow bars of the front row and the second row are not substantial (Figure 4, ANNs). For the  $thw$ -target variables on the right side of Figure 4, there is also a small bias between the linear and non-linear models. This bias occurred mainly at the landwards stations of the study area (Figure 4,  $thw$ ). On the other (left) side of Figure 4, however, the LR and NLR models do not differ (Figure 4,  $tlw$ ).

In addition, the station-wise comparisons between the  $tlw$ - and  $thw$ -targets reveal higher RMSE values for the approximation of  $tlw$ . Hence, all three regression variants explain the measured values of the tidal high water at a station slightly better than the measured values of the tidal low water (Figure 4, station-wise). Across the stations, a sea to land gradient is evident. For both targets, stations close to the seaward boundary can be better approximated than the landward stations.

In Figure 4, station nos. 6 – 4 for the  $tlw$ -targets and station nos. 5 – 3 for the  $thw$ -targets stand out. In this area, the Weser meanders and the cross-section changes significantly (see Figure 1), which might lead to other flow characteristics. With the given regression models these differences in the hydrodynamic behaviour of the river could not be resolved either.

### 3.2.2. Key value approximation

Having illustrated the approximation capabilities of the different regression models using the same predictor data matrix and the same targets, we now look at the applicability of the fitted models. For the reconstructions of the measured values, the predictor data matrix is inserted over 19 years into all final trained regression models to obtain an approximation over a time span of January 2000 to December 2018, referred to in the following as test outputs.



**Figure 4** | Station-wise representation of training quality for three regression models – linear (LR: blue), non-linear (NLR: red) and artificial neural networks (ANNs: yellow) regression. The RMSEs are shown in cm for the approximation of the two target variables: tidal low (tlw) and tidal high (thw) water. The regression coefficients respective ANN-nodes were fitted using a 3-year time period (July 2000 to June 2003).

The comparison of the test outputs with measured target variables is shown in Figure 5, based on the percentage RMSE. The front row of the bar charts in Figure 5 corresponds to the training data already known from Figure 4, and the back row additionally shows the RMSEs of the test outputs. The percentage RMSEs allow a clearer cross-target and cross-station comparison.

Figure 5 shows that the recurrent predictors from the boundaries of the study area can approximate the measured values of the whole area with a quality of 0.4 – 2.5%.

Three further observations from Figure 5 are already known from previous sections: (I) ANNs show the best approximation skill, i.e. lowest percent RMSE values, in each bar group; (II) station-wise, the target values of the thw can be better approximated than the values of the tlw-targets; and (III) in all models the performance is better at the seaward than at the landward stations. With a deviation of less than 1%, the approximation of both target variables works particularly well at the seaward stations nos. 2–4. Moreover, the bias between training and test output is negligibly small here. (Figure 5)

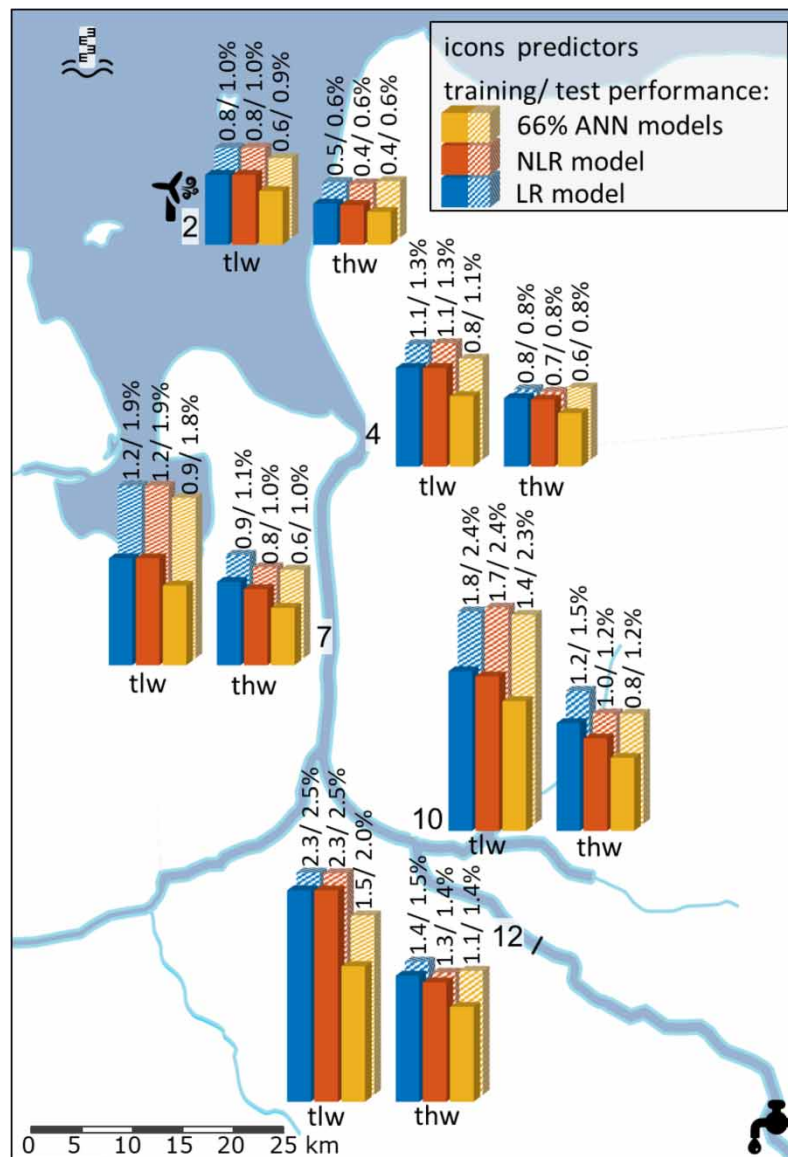
The gap described in Figure 4 between the ANN and LR/NLR outputs can also be seen in the second row of each bar chart of Figure 5 in the test step. At the most southern station no. 12, the test performance of 2% for the ANNs is actually better than the training performance of the two other regression methods (Figure 5, station no. 12/ tlw-target).

One final point to note from Figure 5 is that the inner-estuary stations, which are geographically farthest from the boundaries of the predictors, tend to be less well approximated (test RMSEs for stations nos. 7–10). Overfitting might be suspected if only the bias from training to test output is considered. However, this can be excluded due to the smaller difference (from training to test) for the other stations where the ANN settings were the same.

## 4. DISCUSSION

### 4.1. Hydrological interpretation of feature investigation and selection

In the results of Section 3.1, the different predictors and their functional relationship to the tlw- and thw-targets were analysed. The goal was to find a general approach that is valid for the whole estuary and for a time span of several decades. Predictors were chosen based on their relevance for estuarine processes, as described below. This approach allowed us to avoid feature redundancy.



**Figure 5** | Model performance for training (filled bars) and test (patterned bars) given by the RMSE values in % based on the station-specific measured range shown exemplarily at station nos. 2, 4, 7, 10, and 12; front row: selected percentage values for training (according to Figure 4); back row: proportional RMSEs, when testing the regression models with a data matrix of 19 years (13,410 data). The three icons show the location of the respective predictors.

For the hydrodynamic predictors, we learned that there is a high multicollinearity between tidal high water (thw), tidal range (tr), and tidal mean water (tmw). We selected the tmw-predictor instead of the tr-value for the subsequent regressing modelling to avoid redundancy and due to the importance of tmw in view of sea level rise. The tr-value is a way of describing the tidal energy in the system (Fickert & Strotmann 2007). The tmw-value (arithmetic mean of the hydrograph) describes physical processes related to the water level of the estuary that are based on non-tidal processes and can capture long-term trends. The functional relationship of tmw to other tidal parameters is not straightforward. When the mean water level is high, water depth increases and energy dissipation is potentially reduced. If the increase in tmw is forced from the sea, it leads to a larger tidal range while an increase forced from land (higher discharge) leads to a reduction in tr (Kastens 2014).

The linearity investigations for the river discharge ( $Q$ -predictor) are also interesting. The first impression from the PCC calculations and rankings shows only slight direct correlations with the tidal extrema targets. However, it is striking that the

$Q$ -predictor performs better in the relief rankings than in the PCC rankings, which indicates another functional relationship here. Since the discharge has a damping effect on the incoming tidal wave, changes in the tidal characteristics may be related to it. For example, if the discharge rises, the mean water can be expected to increase, while the tidal high water in the immediate vicinity of the tidal boundary may not rise as high (Kastens 2014). Further higher discharge downstream can force partial reflections of the tidal wave, which might increase the tidal range downstream (Zanke 2002). This suggests testing the  $Q$ -predictor in the subsequent regression modelling in different variants, i.e. with linear, exponential, and ANN calculations.

The meteorological features chosen in our study are the  $u$  and  $v$  components of wind direction and speed. This allows us to consider the wind components as separate predictors in order to understand the effect of a particular wind direction. The linear analyses showed that it takes a certain wind force to exert an effect on tidal extrema, and above a certain wind speed, this effect is not further amplified. However, the ranking analyses showed that these components are difficult to measure, which is why they were more often devalued by the relief algorithm. Even if wind gusts are averaged out in the hourly mean values, the wind variable is subject to a high natural variance in the measured values. In previous studies, it was pointed out that meteorological factors, such as wind speed, wind direction, and air pressure, are important for the prediction of water levels (Massel 1996; Asma *et al.* 2012). In an example by Bowles *et al.* (2005), in which wind is particularly important as a predictor, the ANN approach achieved an accuracy of 5.3 – 8.8 cm for the RSME values of water-level predictions. In our study region, the main effect of wind is the set-up of water levels in the German Bight, which is already considered by using  $tmw$  as a feature.

Overall, the feature investigations showed how the tidal extrema targets could be related to boundary conditions to reveal important relationships and physical processes (Godin 1972; Zanke 2002; Gönnert *et al.* 2004). Hence, the five selected predictors are well suited in terms of relevance and redundancy for the subsequent regression modelling.

#### 4.2. Comparing multiple (non-)linear regression models with ANNs

In the second part of the results, we evaluated how the three regression algorithms responded to the same predictor data matrix and targets. The difference between LR and NLR regression modelling depends on how the  $Q$ -predictor was included (linear versus exponential). It became already visible in the feature investigation results, and even more so in the training results, that the non-linear representation of discharge is particularly interesting for  $thw$  at the landward stations of the study area. The influence of a measured discharge value may become less important towards the seaward boundary. On the one hand, the ratio of discharge to tidal flow along the estuary changes due to different cross-sectional areas, and on the other hand, some time elapses before the discharge water volume arrives downstream. Here it might be worth considering a time lag of the  $Q$ -predictor. In all other respects, the LR and NLR results are almost identical. A pronounced improvement in data approximation was achieved when ANNs were used instead of LR and NLR models (Kastens 2008, 2014).

#### 4.3. Optimal use of ANNs

In our view, an important factor was the HPO in Section 2 (Materials and methods), which was used to select the best configurations for the neural networks. Such optimizations are often achieved by a random search (trial and error) (e.g. Arbian & Wibowo 2012; Rakshith *et al.* 2014; Hibbert *et al.* 2015). In order to obtain more robust results, we investigated the number of nodes and their activation functions intensively by using a sequential method (Bayesian optimization) and determined the best combination based on the RMSEs. Compared to the computationally intensive grid search and the uncertain ending of a random search, the use of Bayesian optimization for hyperparameter search is smart and fast. Similar to feature selection, individual optimal hyperparameter settings can be found station-wise; however, we chose a global solution for a better comparison across targets, stations, and methods.

In addition, we repeated the ANN calculations several times for each case. We found that some ANN models did not perform as well as others, but also that the best ANN model in training was not necessarily the best model when approximating the unseen data. It is also known from the literature that the best training model does not necessarily have the best generalization ability (Lawrence *et al.* 1998). For us, this means that multiple models per case are necessary and a selection of models after training is beneficial.

#### 4.4. Hindcast of tidal water levels

In the following, we try to compare the applicability of our fitted models with previously published results. However, there is one drawback: due to unknown absolute values, we can only compare RMSE values in cm, but cannot make relative

comparisons of prediction qualities. Nevertheless, [Badejo & Udoudo \(2014\)](#) applied an ANN approach in the Gulf of Guinea region (Nigeria) with comparable skills given by an RMSE of about 3.5 – 16.9 cm (converted from MSE into RMSE). For the reconstruction of water levels at five stations their study covered just a very short time period of 15 days. [Berbić \*et al.\*](#) depict similarly high deviations in the prediction of significant wave heights with an RMSE of 4.4 cm for a  $t + 1$  ANN approach and an RMSE of 16.6 cm for a  $t + 12$  ANN approach (Table 5 of [Berbić \*et al.\* 2017](#)). Further, Berbić's ANN approach performs better than their numerical analysis (NM) and their support vector machine approach (SVM). The approximations are more accurate when concentrating on key values such as high water, as we and [Hibbert \*et al.\* \(2015\)](#) did. Hibbert describes the deviation in the approximation of high water with 17.84 cm using a four-layer ANN approach (with two HLs). Hibbert's ANN approach was also better than their tested extended harmonic method. In contrast to these previous studies, our analysis ranges over a much longer time span of 19 years. We could show that an ANN approach can be valid for even such a long period and that an approximation of water levels could be obtained with comparable skill.

A possible shortcoming of our approach is that we do not consider physical processes with a potentially lagged effect, such as river discharge. That might be overcome in future studies by applying a different kind of ANN (NARX).

[Berkenbrink & Niemeyer \(2014\)](#) applied ANNs to approximate salinity in our study region of the Weser estuary. They found a bias between training and test case, which was argued to be due to the channel deepening of the Weser, which took place between their training and test case intervals ([Berkenbrink & Niemeyer 2014](#)). Our application, without known major interventions in the 19 years studied, shows that in all multiple regression modelling, there is a bias between training and unseen data. As [Badejo & Udoudo \(2014\)](#), [Berkenbrink & Niemeyer \(2014\)](#) also used only one HL with a very small number of nodes in their ANN approach, which may have impaired the ability of the neural networks to generalize to unseen data.

From our approach, we learned that we were able to obtain good approximations of unseen data near the boundaries of our study area. In the centre of the estuary, the approximation of the unseen data was not as straightforward, probably because some physical processes could not be covered.

## 5. CONCLUSIONS

The aim of the present research was to examine how well tidal characteristic values of the water level can be reconstructed and predicted in an estuarine setting. The results of this investigation show that over the entire estuary, both tidal extrema (tlw, thw) can be approximated within the given limits of  $\varnothing$  7 cm RMSE respective 1.5% of the absolute values. Another general finding emerging from this study is that it becomes more difficult to obtain reliable approximations the further one moves spatially away from the boundaries.

Our results show that Bayesian-optimized neural network regression is robust in obtaining the required high level of accuracy. Moreover, the neural network regression approach was more accurate than the LR and NLR approaches. The regression model comparison over the different stations shows that deviations between training and test sets cannot generally be explained by overfitting or anthropogenic interventions.

Based on our findings, it can be recommended for future studies of system changes to rely on optimized neural network regression to compare two test cases – an unaffected period before and a period after an intervention. Here, our work can be understood as a proof of concept for the practical potential of neural networks in estuarine system analysis.

In addition, the prior feature investigations and selection process contribute to a better understanding of the ongoing physical processes of the estuary. The highlighted correlations between features and targets are also useful in terms of estuarine system analysis.

## DATA AVAILABILITY STATEMENT

All relevant data are available from an online repository or repositories: [https://www.kuestendaten.de/DE/Services/Messreihen\\_Dateien\\_Download/Download\\_Zeitreihen\\_node.html](https://www.kuestendaten.de/DE/Services/Messreihen_Dateien_Download/Download_Zeitreihen_node.html) and [https://www.kuestendaten.de/DE/Services/Messreihen\\_Dateien\\_Download/Download\\_Zeitreihen\\_node.html](https://www.kuestendaten.de/DE/Services/Messreihen_Dateien_Download/Download_Zeitreihen_node.html).

## CONFLICT OF INTEREST

The authors declare there is no conflict.

## REFERENCES

- Abbas, A., Boithias, L., Pachepsky, Y., Kim, K., Chun, J. A. & Cho, K. H. 2022 AI4Water v1.0: An open-source python package for modeling hydrological time series using data-driven methods. *Geoscientific Model Development* **15** (7), 3021–3039.
- Amaranto, A. & Mazzoleni, M. 2023 B-AMA: A Python-coded protocol to enhance the application of data-driven models in hydrology. *Environmental Modelling & Software* **160**, 105609.
- Arbian, S. H. & Wibowo, A. 2012 Time series methods for water level forecasting of Dungun river in Terengganu, Malaysia. *International Journal of Engineering Science and Technology* **4** (4), 1803–1811.
- Asma, S., Sezer, A. & Ozdemir, O. 2012 MLR and ANN models of significant wave height on the west coast of India. *Computers & Geosciences* **49**, 231–237.
- Badejo, O. T. & Udouo, D. 2014 Modelling and prediction of water level for a coastal zone using artificial neural networks. *International Journal of Computational Engineering Research* **6** (4), 26–41.
- Berbić, J., Ocvirk, E., Carević, D. & Lončar, G. 2017 Application of neural networks and support vector machine for significant wave height prediction. *Oceanologia* **59** (3), 331–349.
- Berkenbrink, C. & Niemeyer, H. D. 2014 Analysis of salinity alterations due to estuarine waterway deepening by artificial neural networks. *Die Küste* **81**, 153–166.
- Bowles, Z., Tissot, P., Michaud, P. & Sadovski, A. 2005 Artificial neural network predictions of water levels in a Gulf of Mexico shallow embayment. *Revista de Matemática: Teoría Y Aplicaciones* **12** (1&2), 139–150.
- Breiman, L. 2001 Statistical modeling: The two cultures. *Statistical Science* **16** (3), 199–215.
- Brunton, S. L., Hemati, M. S. & Taira, K. 2020 Special issue on machine learning and data-driven methods in fluid dynamics. *Theoretical and Computational Fluid Dynamics* **34** (4), 333–337.
- Deutsche Norm 4049 Teil 3 Hydrologie – Teil 3: Begriffe zur Quantitativen Hydrologie (German DIN Standard: Terms for Quantitative Hydrology), pp. 80.
- Dixit, P. & Londhe, S. N. 2016 Prediction of extreme wave heights using neuro wavelet technique. *Applied Ocean Research* **58**, 241–252.
- Dorofki, M., Elshafie, A. H., Jaafar, O., Karim, O. A. & Mastura, S. 2012 Comparison of artificial neural network transfer functions abilities to simulate extreme runoff data. *International Conference on Environment, Energy and Biotechnology* **33**, 39–44.
- Dyer, K. R. 1997 *Estuaries – A Physical Introduction*, 2nd edn. John Wiley & Sons, Chichester, NY.
- Erol, S. 2011 Time-frequency analyses of tide-gauge sensor data. *Sensors* **11** (4), 3939–3961.
- Fickert, M. & Strotmann, T. 2007 Hydrodynamische Entwicklung der Tideelbe (Hydrodynamic development of the Tideelbe). *Von der Geoarchäologie über die Küstendynamik zum Küstenzonenmanagement. Coastline Reports* **9**, 59–68.
- Ghorbani, M. A., Khatibi, R., Aytak, A., Makarynsky, O. & Shiri, J. 2010 Sea water level forecasting using genetic programming and comparing the performance with artificial neural networks. *Computers & Geosciences* **36** (5), 620–627.
- Godin, G. 1972 *The Analysis of Tides*. University of Toronto Press, Toronto, Canada.
- Godin, G. 1988 *Tides. ANADYOMENE Edition*. ANADYOMENE, Ottawa, Canada.
- Gönnert, G., Isert, K., Giese, H. & Plüß, A. 2004 Charakterisierung der Tidekurve (Characterization of the tidal curve). *Die Küste* **68**, 99–141.
- Grabusts, P. & Zorins, A. 2015 The influence of hidden neurons factor on neural network training quality assurance. *Proceedings of the International Scientific and Practical Conference* **3**, 76–81.
- Günaydin, K. 2008 The estimation of monthly mean significant wave heights by using artificial neural network and regression methods. *Ocean Engineering* **35** (14–15), 1406–1415.
- Gupta, H. V., Kling, H., Yilmaz, K. K. & Martinez, G. F. 2009 Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology* **377** (1–2), 80–91.
- Guyon, I. & Elisseeff, A. 2003 An introduction to variable and feature selection. *Journal of Machine Learning Research* **3**, 1157–1182.
- Hagen, R., Winter, C. & Kösters, F. 2022 Changes in tidal asymmetry in the German Wadden Sea. *Ocean Dynamics* **72** (5), 325–340.
- Hibbert, A., Royston, S. J., Horsburgh, K. J., Leach, H. & Hisscott, A. 2015 An empirical approach to improving tidal predictions using recent real-time tide gauge data. *Journal of Operational Oceanography* **8** (1), 40–51.
- Huang, G.-B. 2003 Learning capability and storage capacity of two-hidden-layer feedforward networks. *IEEE Transactions on Neural Networks* **14** (2), 274–281.
- James, G., Witten, D., Hastie, T. & Tibshirani, R. 2017 *An Introduction to Statistical Learning: With Applications in R*, 2nd edn. Springer, New York, USA.
- Karimi, S., Kisi, O., Shiri, J. & Makarynsky, O. 2013 Neuro-fuzzy and neural network techniques for forecasting sea level in Darwin Harbor, Australia. *Computers & Geosciences* **52**, 50–59.
- Kastens, M. 2008 Tidewasserstandsanalyse in Ästuaren am beispiel der Elbe (Tidal water level analysis in estuaries using the example of the Elbe). *Die Küste* **72**, 145–169.
- Kastens, M. 2014 Statistical estuary data analysis in models and measurements – some methods and their limitations. *Die Küste* **81**, 185–201.
- Kira, K. & Rendell, L. A. 1992 A practical approach to feature selection. In: *Machine Learning Proceedings* (Sleeman, D. & Edwards, P., eds). Elsevier, Amsterdam, pp. 249–256.
- Kumar, U. 2005 Comparison of neural networks and regression analysis: A new insight. *Expert Systems with Applications* **29** (2), 424–430.
- Lang, G. 2003 Analyse von HN-Modell-Ergebnissen im Tidegebiet. *Mitteilungsblatt der Bundesanstalt für Wasserbau* **86**, 101–108.

- Lange, D., Müller, H., Piechotta, F. & Schubert, R. 2008 The weser estuary. *Die Küste* **74**, 275–287.
- Lawrence, S., Giles, C. L. & Tsoi, A. C. 1998 *What Size Neural Network Gives Optimal Generalization? Convergence Properties of Backpropagation. Technical Report*. University of Maryland.
- Li, S., Wunsch, D. C., O’Hair, E. & Giesselmann, M. G. 2001 [Comparative analysis of regression and artificial neural network models for wind turbine power curve estimation](#). *Journal of Solar Energy Engineering* **123** (4), 327–332.
- MacCready, P. & Geyer, W. R. 2010 [Advances in estuarine physics](#). *Annual Review of Marine Science* **2**, 35–58.
- Massel, S. R. 1996 Ocean surface waves: Their physics and prediction. *Advanced Series on Ocean Engineering* (Liu, P. L. F., ed.). World Scientific, Singapore.
- Niemeyer, H. D. 1998 Change of mean tidal peaks and range due to estuarine waterway deepening. In: *Proceedings of 26th Conference on Coastal Engineering*.
- Prahlada, R. & Deka, P. C. 2015 [Forecasting of time series significant wave height using wavelet decomposed neural network](#). *Aquatic Procedia* **4**, 540–547.
- Rakshith, S., Dwarakish, G. S. & Natesan, U. 2014 Tidal-level forecasting using artificial neural networks along the west coast of India. *Journal of the Japan Society of Civil Engineers* **2**, 176–187.
- Robnik-Sikonja, M. & Kononenko, I. 1997 An adaptation of relief for attribute estimation in regression. *Proceedings of the Fourteenth International Conference on Machine Learning* **5**, 296–304.
- Robnik-Sikonja, M. & Kononenko, I. 2003 [Theoretical and empirical analysis of relief and relief](#). *Machine Learning* **53**, 23–69.
- Tayel, M. & Oumeraci, H. 2018 Vorhersage extremer Sturmfluten mit Hilfe hydrodynamischer Modellierung und künstlicher neuronaler Netze (Predicting extreme storm surges using hydrodynamic modelling and artificial neural networks). *Die Küste* **86**, 341–367.
- Thomas, A. J., Walters, S. D., Gheytaoui, S. M., Morgan, R. E. & Petridis, M. 2016 [On the optimal node ratio between hidden layers: A probabilistic study](#). *International Journal of Machine Learning and Computing* **6** (5), 241–247.
- Warner, B. & Misra, M. 1996 [Understanding neural networks as statistical tools](#). *The American Statistician* **50** (4), 284–293.
- Zanke, U. C. E. 2002 *Hydromechanik der Gerinne und Küstengewässer: Für Bauingenieure, Umwelt- und Geowissenschaftler*. Vieweg + Teubner Verlag, Wiesbaden, Germany.

First received 19 December 2023; accepted in revised form 29 February 2024. Available online 13 March 2024