Wave: Jurnal Ilmiah Teknologi Maritim **Vol.18, No.02, Desember 2024: page: 76-90**
(Journal of Maritime Technology) p - ISSN: 1978-886X
e - ISSN: 2614-641X

A MODULAR TIDE LEVEL PREDICTION USING COMBINATION OF HARMONIC-ANALYSIS AND NONLINEAR AUTOREGRESSIVE EXOGENOUS (NARX) METHODOLOGY IN SEMARANG INDONESIA

Elvien Hastatomo Khasby[1,2], Anistia Malinda Hidayat[2], Usman Efendi[2], Pulung Nurtantio Andono[1], Catur Supriyanto[1], Arief Soeleman[1], Ferry Oktarisa[2], Aurel Dwiyana[2]

[1]University Dian Nuswantoro, Semarang, Indonesia

[2]Indonesian Agency for Meteorology Climatology and Geophysics (BMKG)

Article Info

Article history:

Received November 28, 2024

Revised May 16, 2025

Accepted May 24, 2025

Keywords:

Prediction

Tides

Harmonic analysis

Neural Network

NARX

ABSTRACT

Semarang is highly prone to tidal floods year-round, making tidal prediction using methods like Harmonic Analysis with Least Squares (HA-LS) crucial for disaster mitigation. However, this method is not effective enough because it has relatively high Root Mean Squared Error (RMSE) value of up to 11.43 cm with coefficient of determination (R^2) of around 0.727. The Nonlinear Autoregressive Exogenous (NARX) neural network model is then proposed to improve the accuracy of tide predictions. In general, the features used as input are divided into two types, namely atmospheric/weather data (temperature, pressure, direction, and wind speed) and estimated tide data from the HA-LS method. This research aims to analyze whether this type of neural network is suitable to be applied in Indonesia since the ocean and atmosphere condition might vary. Tidal observation data from the Tanjung Mas Maritime Meteorological Station is used as target data/actual data. Three scenarios are used by varying input types to find out which type of input produces the best performance of model prediction. Moreover, before feeding input data into the NARX neural network, all atmospheric data used as input are standardized using Z-score normalization, often called Min-Max scaling which can avoid the effect of outlier in the dataset. Based on these three scenarios, the use of combined atmospheric/weather data and tide estimates from HA-LS calculations as input to the NARX model produces the best predictive performance with the smallest RMSE value among all scenarios, approximately 5 cm, and the highest coefficient of determination (R^2) at about 0.974. These results indicate that NARX model can predict tides with high accuracy in Semarang.

Corresponding Author:

* Elvien Hastatomo Khasby

University Dian Nuswantoro, Semarang, Indonesia

Email: elvien.khasby@bmkg.go.id

INTRODUCTION

Semarang is one of the cities that is vulnerable to tidal flood (Ikhsyan, M. & Rintayani, 2017). The increase in land surface subsidence and the rise in sea level are the causes of the current tides flooding (Erlani & Nugrahandika, 2019). Therefore, information about tide level and its prediction in Semarang is needed to complete missing data in tide level or even provide warning of coastal flood. For further implementation, this can be used as part of disaster risk mitigation.

The gravitational force of the moon and the sun, along with the inertial centrifugal force needed for the relative motion of the Earth cause periodic changes in the surface of the sea, which are generally known as tides. Tides data is one of the important factors in marine meteorology. Therefore, Indonesian Agency for Meteorology, Climatology, and Geophysics (BMKG) installed more than 4500 Automatic Weather Station (AWS) in 24 different locations in Indonesia (Bagaskoro & Saputro 2023). However, one should remember that Indonesia is the largest archipelagic country which has the second longest coastline in the world (Alfahmi et al. 2019). This condition can illustrate the lack of AWS instrument needed to cover the whole country.

AWS is a collection of integrated meteorological sensors that automatically record weather data and produce electrical pulses that are gathered and transformed in a data recorder before being presented on a computer screen or translator. Ships that will sail and dock at ports can receive information on the weather and water conditions thanks to the use of maritime AWS data. Tide level data also has a vital role in giving information related to coastal flood and inundations, commonly known as rob flood. The northern area of Semarang is one of the most vulnerable areas to this type of flood therefore the continuous availability of tide level data is beyond important in Semarang.

AWS is installed at a height of 10 meters above the land surface and placed near the coast. This condition can cause constant exposure to the sea which makes water level sensor damage easily. Statistically, the technician of BMKG should repair and replace it in less than 2 years. The frequent damage of the sensor and the importance of tide level data for communities in general become the main reason behind why this research is conducted. We aim to predict and provide the tide level data by using the combination of Harmonic Analysis (HA) and Nonlinear Autoregressive Exogenous (NARX) neural network method.

Previously, it is commonly known that HA method can be used to handle missing data of tide level or simulate astronomical tide level (Rayson et al. 2021). In this approach, tidal forcing is modeled as a collection of spectral lines, or the sum of a finite collection of sinusoids at particular frequencies. These frequencies are determined by a number of different summations and differences of integer multiples of the six fundamental frequencies caused by planetary motions (Li et al. 2019). The effects of the earth's rotation (lunar day of 24:8 h), the moon's orbit around the earth (lunar month of 27 days), the earth's orbit around the sun (tropical year), periodicities in the location of lunar perigee (8.85 years), lunar orbital tilt (18.6 years), and the location of perihelion are all represented by these fundamental parameters (≈21,000 years). However, the error resulted in predicting tide level using this method only is still relatively high in this research.

The coastal city of Semarang, located on the north coast of Java, is significantly influenced by tidal dynamics that play a crucial role in its oceanographic and environmental characteristics. Tidal analysis reveals that the Semarang coast has a mixed, predominantly diurnal tidal regime, indicating that diurnal tides (one high and one low tide per day) dominate, though semi-diurnal components (two high and two low tides per day) are also present (Permana, Hariadi & Rochaddi 2012; Hakim, Suharyanto & Hidajat 2013; Rahili et al. 2023). This classification is primarily driven by the lunar declination cycle, which has a period of approximately 13.6 days, and the semi-diurnal influences from the new moon - full moon cycle, with a period of about 14.75 days. Additionally, the moon's perigee and the solar declination cycle, which peaks semi-annually around the June and December solstices, further modulate tidal heights, particularly influencing peak sea levels that can lead to coastal flooding (Djamaluddin et al. 2023).

The local bathymetry of the Java Sea, with its shallow depths averaging around 46 meters, plays a pivotal role in shaping these tidal dynamics. The shallow waters enhance nonlinear hydrodynamic effects, such as frictional and advective interactions, which amplify diurnal constituents like K1 and contribute to the complex tidal patterns observed in Semarang (Pugh 1987; Koropitan & Ikeda 2008). This bathymetric influence results in a mixed, predominantly diurnal regime, where K1 is amplified while semi-diurnal constituents, such as M2, are less prominent (Rahili et al. 2023). These tidal interactions are further complicated by geological factors, including alluvial deposits and rapid land subsidence, which intensify coastal flooding during high tide (Marfai & King 2008; Abidin et al. 2013).

Technology related to artificial intelligence has advanced quickly in recent years, including in marine science such as in the research of Yin et al. (2018). Therefore, as the second approach, authors try to implement neural network with NARX to get better results in the prediction of tides since tide level data can be categorized as time series in forecasting. Thus, the NARX model can offer improved learning effectiveness and increased prediction accuracy in time-series data (Wu et al. 2021). Moreover, NARX neural networks have also been used for modeling and prediction in a number of research fields, including those of Buevich et al. (2021). By this far, it is worth to notice that there is no research on NARX involving a dynamic regression network with static neurons and network output feedback that outperforms full regression neural networks, for predicting tide levels, except in Wu et al. (2021).

Besides adopting the research of Wu et al. (2021), this research also modifies the input data before feeding it into the NARX neural network. This research aims to analyze whether this type of neural network is suitable to be applied in Indonesia since the ocean and atmosphere condition might vary. Atmospheric parameters such as air temperature, air pressure, wind direction, and speed will be used as input data, together with tidal estimates which resulted from HA method in order to increase the performance of NARX neural network model for tide level prediction.

METHOD

Atmospheric data consists of air temperature, air pressure, wind speed and direction from AWS maritime, and direct observation done by the observer of Tanjung Emas Maritime Meteorological Station Semarang from February

2021 until June 2022 are used in this research. Two tidal data series were used in this study. The first dataset, covering the period from January 1, 2021 to January 31, 2021, was used to determine harmonic constants using UTide (Codiga 2011). A total of 29 tidal harmonic constants were derived using the UTide program. These harmonic constants were then applied to predict tidal levels from February 1, 2021, to June 30, 2022. In this case, the formula adopted from Sida li et al (2019) in tidal prediction using least square method is shown in Equation [1]. The second data series, spanning from February 1, 2021, to June 30, 2022, was used as the target data for tidal prediction. The data used in this study is recorded at an hourly frequency.

$$h(t) = Ho + \sum_i f_i H_i \cos[\, a_i t + (V_o + u)_i - k_i \qquad [1]$$

where:

h(t) : tide level estimated in time *t*

Ho : height of mean sea level above model datum

f_i : nodal factor for reducing mean amplitude to the required amplitude in the year of prediction which is used as input data

H_i : mean amplitude of *i*-th constituent during 18.6 year-period of lunar node cycle

a_i : speed of the *i*-th constituent in degrees per mean solar hour

t : time

$(V_o + u)$: local equilibrium phase of *i*-th constituent in the year of prediction

k_i : phase of *i*-th constituent relative to the local equilibrium phase

In the second attempt, authors used 3 different data type and named it as scenario 1 - 3. In the 1st scenario, only atmospheric data was used as input in NARX neural network model. In the 2nd scenario, only tide level data resulted from HA using Least Square method, later called as HA-LS, was used as input in NARX. In the 3rd scenario, the combination of atmospheric and estimated tide level from HA-LS are used as input data in NARX model. The third scenario was used to improve the prediction accuracy of tide level data (Wu et al. 2021) and later can be called as HA-NARX method. However, a modification approach was done to make it different from Wu et al. (2021) by normalizing input data. Moreover, before feeding data into the NARX neural network, all atmospheric data used as input were standardized using Z-score normalization, often called Min-Max scaling which can avoid the effect of outlier in the dataset. The following equation [2] is frequently used to perform Min-Max scaling:

$$X_{sc} = \frac{X - X_{min}}{X_{max} - X_{min}} \qquad [2]$$

where:

Xsc : normalized output value

X : current value which will be normalized

Xmin : minimum value in X dataset

Xmax : maximum value in X dataset

This type of standardization is also studied in the research Henderi (2021) about standardization and its effects on *K-means* clustering algorithm. Standardization of a dataset, part of data preprocessing approach, is a common requirement for many machine learning estimators because they might perform badly if the features do not look like standard normally distributed data. In neural network, the more representative the data, the better the neural network model performs.

A nonlinear autoregressive model for characterizing nonlinear discrete systems is a NARX neural network (Xie, T. & Liao 2018). In

Vol.18, No.02, Desember 2024: page: 76-90
p - ISSN: 1978-886X
e - ISSN: 2614-641X

nonlinear dynamic systems, it is the most popular kind of neural network and is effective at predicting time series (Zhao et al. 2020; Boussaada et al. 2018). As a result, NARX neural networks have been used in numerous disciplines to handle nonlinear sequence prediction issues. Additionally, NARX neural networks are better suited to the study and forecasting of time series data, such as tide level data, because they have a stronger mirroring power for nonlinear fitting than other neural networks (Wu et al. 2021).

The typical NARX neural network architecture is seen in Figure 1. A neural network's output typically serves as its input Figure 1(a), later known as parallel mode (closed loop). However, an open loop model of the series-parallel neural network represented in Figure 1(b) is established because the anticipated training output of a NARX neural network is known. The desired output is used as input in this mode. The NARX neural network is improved by this strategy in two ways: first, it becomes more accurate; and second, it is converted into a straightforward feedforward neural network that can make use of the modeling capabilities of a static neural network.

In Figure 1, *TDL* is an abbreviation of time delay, *y(t)* is the known expected output in time *t*, and *Y(t)* is the predicted tide data in time *t*. In this research, Levenberg–Marquardt (LM) algorithm is used to update bias and weight vector in order to minimize error until the desired error is reached.

Where *u* is a variable that is determined externally. The formula states that the input value *x(t)*, along with the prior output *y*, determines the value of *y(t)* at the subsequent instant *(t)*. In order to estimate tide level, one should specify the initial input parameter of NARX neural network. These parameters include the number of nodes in the input layer, hidden layer, and output layer, as well as the order in which the input and output are delayed.

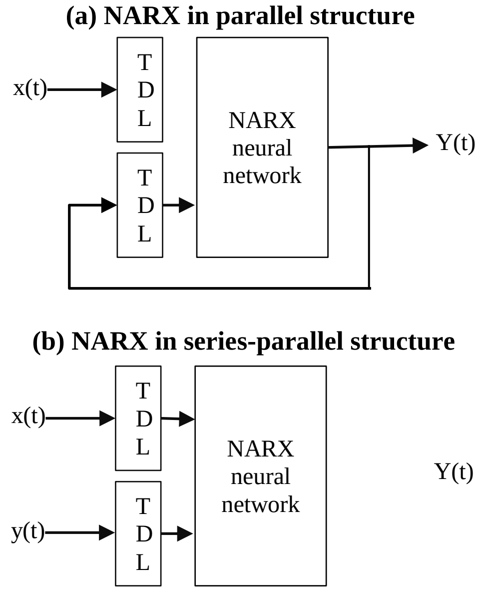
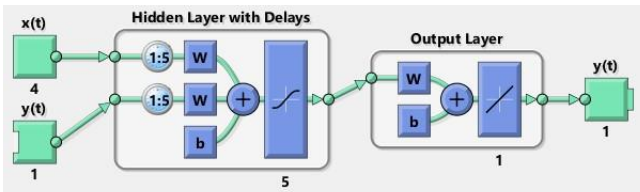


Figure 1. Two different structures of NARX neural network model

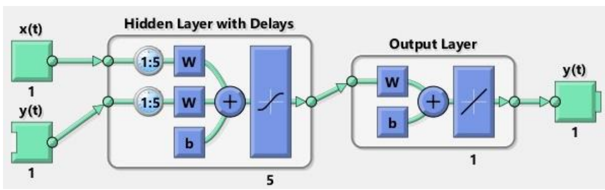
The following equation serves as an expression for the NARX neural network model.

$$y(t) = f(y(t-1), y(t-2), \ldots, y(t-n_y), u(t-1), u(t-2), \ldots u(t-n_u)) \quad [3]$$

(a) scenario 1



(b) scenario 2



(c) scenario 3

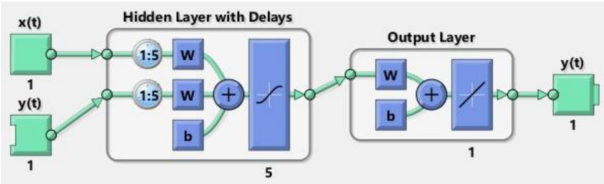


Figure 2. Architecture NARX

In this paper, we use four atmospheric parameters in scenario 1, therefore the input node is 4. While input node is 1 and 5 respectively for the use of scenario 2 and 3. The number of neurons in hidden layer is set to be 5 and the number of output node is set to be 1 for all scenarios. Moreover, the delay order is determined to be 1:5, applied in all three scenarios.

Figure 3 shows the third scenario which uses atmospheric data and tidal prediction data from the HA-LS method as inputs. This third experiment is expected to produce the best results. In order to evaluate the performance of the neural network model in tide level prediction, Mean Square Error (MSE) and Root Mean Square Error (RMSE) are used. MSE and RMSE are related to one and another and formulated in Equation [4] and [5] consecutively. Basically, RMSE is the rooted variant of MSE. MSE and RMSE are both essential elements of statistical models, especially in terms of regression. However, RMSE is used more often than MSE because this metric is expressed in the same units as the response variable when evaluating how well a model fits a dataset. Moreover, the mistakes provided by the RMSE are indicative of the amount of an "average" error rather than reporting in terms of the "average" of squared errors, as is the case for MSE (Tatachar 2021).

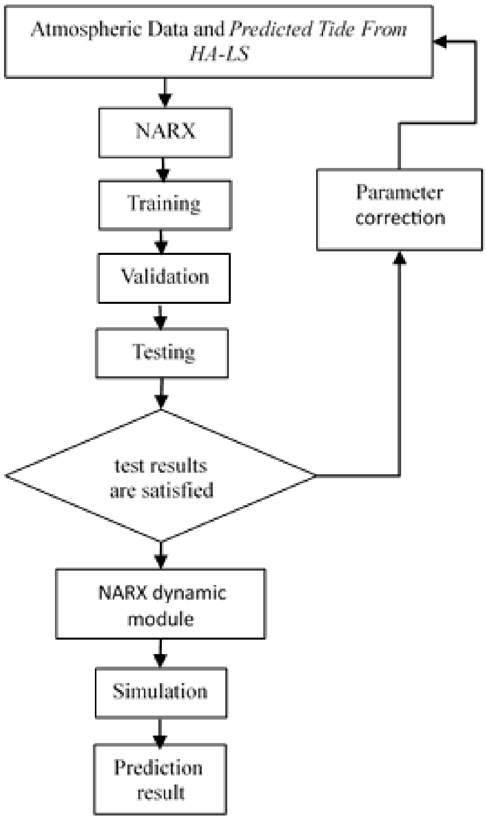


Figure 3. A flow diagram of general NARX neural network

The smaller the value, the better the performance of the prediction.

$$MSE = \sum (\hat{y}_i - y_i)^2 / n \quad [4]$$

$$RMSE = \sqrt{\sum (\hat{y}_i - y_i)^2 / n} \quad [5]$$

where:

$\hat{y}_i$: predicted value for the *i*-th observation

y_i : observed value for the i-th observation

n : sample size

In addition to MSE and RMSE, the coefficient of determination, also known as R-squared or R^2, is also used in this paper to evaluate the performance of prediction model. The coefficient of determination can be defined as the percentage of the variance of dependent variables that can be predicted from the independent variables. Research conducted by Chicco et al. (2021) showed that coefficient of determination R^2 is categorized more informative and accurate than Mean Absolute Percentage Error (MAPE). Moreover, it does not have interpretability restrictions on MSE and RMSE. Mathematical expression to calculate coefficient of determination R^2 is shown in Equation [6].

Commonly, the value of R^2 ranges between 0 and 1. When the value reaches 1, it means that the model has good performance in predicting the outcome. If the value of R^2 is between 0 and 1, it means that the model partially predicts the outcome. Meanwhile, when the value of R^2 is 0, it means that the model does not predict the outcome. However, if the value of R^2 is -∞, it illustrates the worst performance of the prediction model.

$$R^2 = 1 - \frac{\sum_{i=1}^{m}(Y_i - X_i)^2}{\sum_{i=1}^{m}(\bar{X} - X_i)^2} \quad [6]$$

$$\bar{X} = \frac{1}{m}\sum_{i=1}^{m} X_i \quad [7]$$

where:

Y_i : predicted *i*-th value

X_i : actual *i*-th value

$\bar{X}$: mean of actual value

m : sample size

RESULTS AND DISCUSSION

3.1. Preprocessing phase

The first step of the pre-processing process is Exploratory Data Analysis (EDA). This is a method to handle the presence of NaN (Not a Number) value and the outlier which exist within the dataset which can disturb the distribution data within the input data set of neural network model. Moreover, EDA needs to be done to make sure the data is complete and does not contain outliers which might disturb the prediction of neural network model. In data analysis, including neural network, EDA helps maximize the value of the data.

During the process of EDA, authors find that the atmospheric data in this research does not contain missing value, NaN value, or outlier. This happens because data collection was carried out directly and routinely by human resources of BMKG (primary data). Thus, further steps related to standardization can be taken. Min-Max scaling or z-score standardization method used for standardization in this paper, where the minimum value is set to be 0, while the maximum value is set to be 1. The module used to execute this standardization is MinMaxScaler from sklearn module in Python 3.9 application. Figure 3 shows how the atmospheric parameters vary before and after standardization.

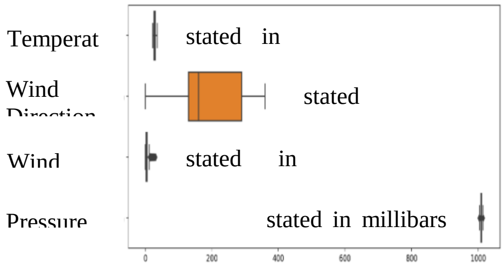
The features used in this study have different value ranges, for example temperature (tens value with a minimum value of 22.5°C and maximum value of 36°C), meanwhile air pressure range around thousands value with a minimum value of 1002.7mb and a maximum value of 1014.8mb. Therefore, feature scaling is required before including it into the ANN model. This standardization is applied to scale back or change, without distorting the differences in the range of values between variables and features so that they have a mean of 0 and a variance of 1. Standardization of data sets is a common requirement for many machine learning estimators because they may perform poorly if features are not visible like standard normally distributed data.

Before rescaling (Figure 3a), the average air temperature was 28.3°C with a minimum value of 22.5°C, while the maximum value is 36°C. Meanwhile, the average pressure value before rescaling showed 1009,2mb, with a minimum value of 1002,7mb and a maximum value of 1014,8mb. Another parameter, namely wind direction, shows an average value of 188° before rescaling, with a minimum value of 0° and a maximum value of 360°. Finally, the average wind speed before being scaled back shows 4 knots, with a minimum value of 0 knots and a maximum value of 31 knots.

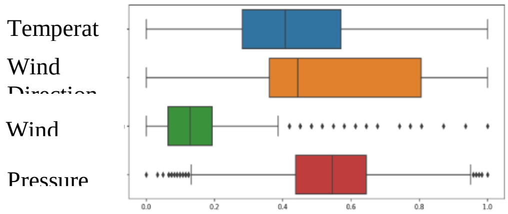
After rescaling (Figure 3b), the minimum value for all parameters is 0, while the maximum value is 1 without distorting the differences or distances between data in each parameter. This is evidenced by the different average values for each parameter. In the air temperature parameter, the temperature value is centered on an average value of 0.43, while the average pressure value is concentrated at 0.54. Quite different things are clearly observed in the distribution of pressure values, where there are several values that are less than the Quartile 1 or Q1 value (less than 0.18) and several values that are more than the Quartile 3 or Q3 value (more than 0.9), while the range of air temperature values is always in the range of Q1 and Q3 values. Wind direction values are concentrated at an average value of 0.52 and 0.13 for the average wind speed parameter.

3.2. Tides Prediction Using Atmospheric Data as Input NARX model

Atmospheric or weather input data used as input in the NARX model involves data on air temperature, wind direction, wind speed, and air pressure. These atmospheric factors are closely related to the process of wave formation, including tides. One potential global component that causes major displacements of the deep and near-surface masses of the Earth's material, including the observable changes in weather and climate, is constantly repeated lunar-solar tidal perturbation.



(a) Before standardization



(b) After standardization

Figure 4. Range of atmospheric dataset values (a) before and (b) after standardization.

3.2. Tides Prediction Using Atmospheric Data as Input NARX model

Atmospheric or weather input data used as input in the NARX model involves data on air temperature, wind direction, wind speed, and air pressure. These atmospheric factors are closely related to the process of wave formation, including tides. One potential global component that causes major displacements of the deep and near-surface masses of the Earth's material, including the observable changes in weather and climate, is constantly repeated lunar-solar tidal perturbation.

Other studies also stated that, in fact, the majority of fluid level sensors are pressure transducers. As a result, they document both the air pressure and the weight of the fluid column. In addition to this, Fourier analysis reveals distinct semi-diurnal and nocturnal variations in air pressure brought on by tides. Moreover, changes in air temperature always be followed by the change in air pressure. This represents

Wave: Jurnal Ilmiah Teknologi Maritim
(Journal of Maritime Technology)
Vol.18, No.02, Desember 2024: page: 76-90
p - ISSN: 1978-886X
e - ISSN: 2614-641X

how weather factors or atmospheric data can affect tides formation. Besides pressure, spectra of observed atmospheric temperature and winds were calculated in the transfer function which relates atmospheric elements with sea level tides.

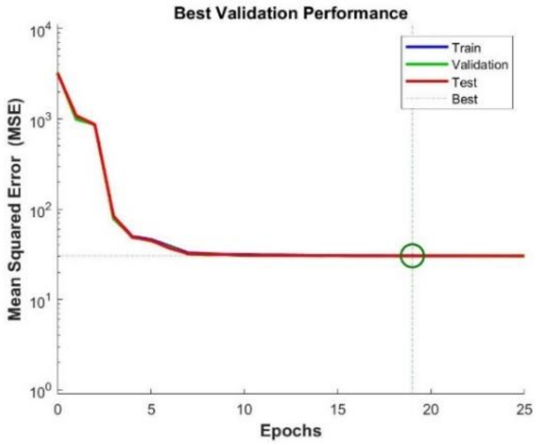


Figure 5. Mean squared error (MSE) trajectory during training until the model reaches its best performance after *i*-th iteration (epochs)

After training and validation, the best validation performance was reached after 19 iterations (epochs) with MSE value of 30.5 or RMSE value of 5.52 cm (Figure 5). This result is better compared to estimated tide level from HA using LS method which has MSE value of 130.57 and RMSE value of 11.43 cm.

Furthermore, when the validation process is carried out, the determinant value decreases slightly to 0.96702, meaning that the use of weather data influences the predicted value by around 96.7%. At this stage, it can be said that the performance of the model is slightly decreasing but the prediction equations produced by the model have more general predictive formula, considering that the NARX model not only remembers patterns in learning data, but also studies them to produce better tidal prediction values for the future. Therefore, when the prediction equations that have gone through the validation process are applied, the determinant value of the NARX model using testing data increases to 0.96843, a slight increase compared to the results of the training and validation processes.

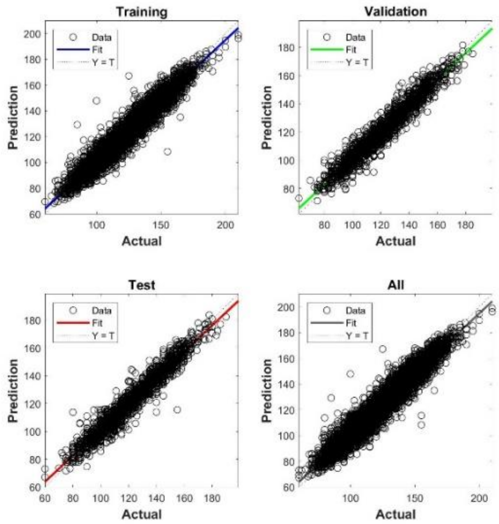


Figure 6. Prediction results from NARX using atmospheric data only as input. The diagram shows how the result of prediction varies to the actual data.

In the final stage, the NARX model has stored the learning memory. When the model is used to predict all data without giving initial data to be trained, it can produce a determinant value of 0.96754. This result is much better than the use of HA-LS method alone which showed a R^2 value of 0.727. This result highlights the significant role of atmospheric data in tidal variation and demonstrates that the model predicts tide levels in Semarang with high accuracy.

In the range of tidal values that are less than 70 cm, NARX neural network model tends to overestimate, meaning that the predicted tide values produced by the model tend to be higher than the actual values. Conversely, for tide values that are more than 160 cm, NARX neural network model tends to be underestimated, meaning that the model often predicts tide values that are lower than the actual values.

3.3. Tides Prediction Using Predicted Tide From HA-LS as Input NARX model

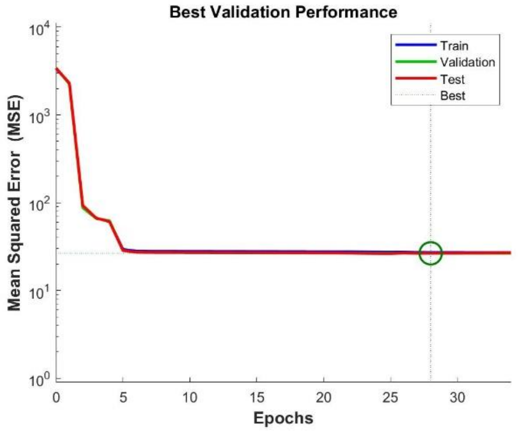


Figure 7. Mean squared error (MSE) trajectory during training until the model reaches its best performance after *i*-th iteration (epochs) in the 2nd scenario.

As discussed in previous section (3.2), the evaluation metrics show that estimated tide level data from harmonic analysis using least square method is worse compared to predicted tides data from NARX model in the 1st scenario, which used atmospheric data only as input. In this section, the estimated tides data will be used as input data in NARX neural network model or labeled as 2nd scenario.

In the 2nd scenario, the best validation performance after training and validation was reached after 28 iterations (epochs) with MSE value of 26.5 or RMSE value of 5.15 cm (Figure 7). Again, compared to the estimated tide level from the HA-LS approach, which has MSE values of 130.57 and RMSE values of 11.43 cm, this result is better. When this scenario is compared again to the 1st scenario which has RMSE value at about 5.52 cm, the use of estimated tides data shows a better performance for the prediction.

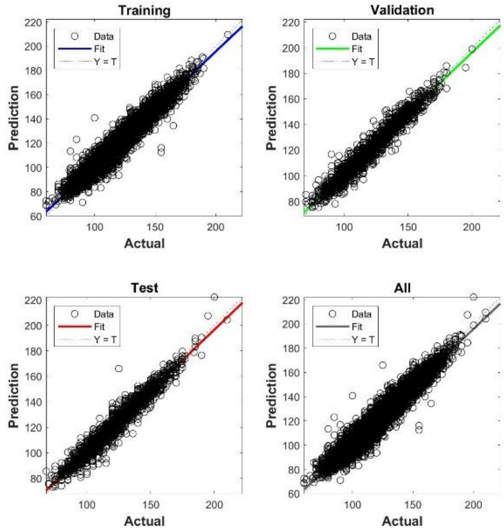


Figure 8. Prediction result from NARX using estimated tides data only as input (scenario 2). The diagram shows how the result of prediction varies to the actual data.

Figure 8 illustrates how prediction results compare to actual value in the 2nd scenario. The prediction outcomes of the NARX model have a coefficient determinant (R^2) of 0.97108 throughout the data learning phase (training data), indicating that the usage of approximated tides data affects the projected value by 97.1%. When compared to the NARX model's input of weather data alone (1st scenario), this result is comparatively better.

Additionally, after applying the validation process, the determinant value significantly rises to 0.97221, indicating that the data used has a 97.2% influence on the predicted value. Additionally, the model's post-validation prediction equations can lead to higher overall data generalization. The coefficient determinant value of the NARX model utilizing testing data thus climbs to 0.97257 when the prediction equations that have undergone the validation process are applied, which is a little increase compared to the results of the training and validation processes.

The estimated tides data from the HA-LS method have made a significant contribution to the prediction process, accounting for about 97.1% of it in the final stage when the NARX model, which has stored learning memory, is used to predict all data. This model can produce a determinant value of 0.97147 when used to predict tides in all data sets. Once more, these findings show that the NARX model outperforms the HA-LS method for tide prediction. It can be said that the NARX neural network model can improve the accuracy of estimated data from HA-LS technique.

The NARX model tends to overestimate, which means that the predicted tide values it generates tend to be higher than the actual values, in a range of tidal values that are less than 100 cm. On the other hand, the NARX model frequently underpredicts tide levels for values of more than 175 cm, resulting in predictions which are lower than the actual values. However, in general, if calculated as a whole, the NARX model in the 2nd scenario tends to slightly overestimate.

3.4. Tides Prediction Using a Combined of Atmospheric Data and Predicted Tide From HA-LS as Input NARX model

In the 3rd scenario, after training and validation, the best validation performance was attained after 24 iterations (epochs), with MSE value of 25.0 or RMSE value of 5.0 cm (Figure 9). Again, this result is substantially better than the estimated tide level from the HA-LS technique, which has MSE values of 130.57 and RMSE values of 11.43 cm. The usage of approximated tides data demonstrates a higher performance for the prediction in terms of MSE and RMSE when this scenario is compared to the 1st scenario and 2nd scenario. Compared to Wu et al. (2021), the RMSE value is smaller than tides prediction in the 3rd scenario using HA-NARX model in this research. However, in terms of determinant (R^2) that will be explained below (Figure 10), this approach performs as better as Wu et al. (2021), even better with overall R^2 value of 0.97, while the R^2 value in station ID 9414290 is approximately 0.94 in the research of Wu et al. (2021).

Figure 10 shows how the predicted tides data for the 3rd scenario compares to the actual tide values. The prediction results of the NARX model have a coefficient determinant (R^2) of 0.97469 during the data learning process (training data), which uses a combination of atmospheric data and estimated tidal data from HA-LS calculations as input (3rd scenario). This means that the combination of both data affects the prediction value up to 97.5%. This result is superior to that of the second scenario, which used only the estimated tidal data from the HA-LS computation and the meteorological data from the first scenario.

Additionally, after the validation procedure, the determinant value significantly rises to 0.97449, indicating that the total data has a 97.4% influence on the predicted value. Since the model not only recalls patterns in learning data but also examines them to create better tidal prediction values, it can be claimed that the performance of the model is currently somewhat declining but that the prediction equations it produces may have stronger generalization formulas.

The coefficient determinant value of the NARX neural network model is roughly 0.9722 throughout the processing of testing data. Furthermore, the coefficient determinant value is approximately 0.97427 when it is applied to predict all dataset series. This demonstrates that the integration of these data gives the NARX model stronger and more comprehensive learning resources, which ultimately improves prediction accuracy. The model virtually forecasts the tide level data at Semarang, with a coefficient determinant of more than 97.4%. This model performs significantly better than the 1st and 2nd scenarios, and even much better than estimated tides data from HA-LS calculation method.

In the range of tidal values that are less than 50 cm, the NARX model tends to overestimate, meaning that the predicted tide values produced by the model tend to be higher than the actual values. Conversely, for tide values that are more than 185 cm, the NARX model tends to be underestimated, meaning that the model often predicts tide values that are lower than the actual values.

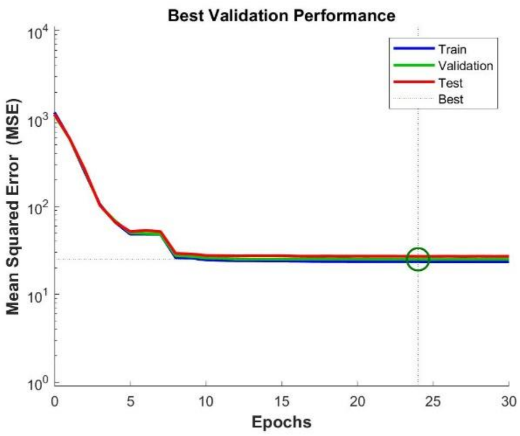


Figure 9. Mean squared error (MSE) trajectory during training until the model reaches its best performance after *i*-th iteration (epochs) in the 3rd scenario.

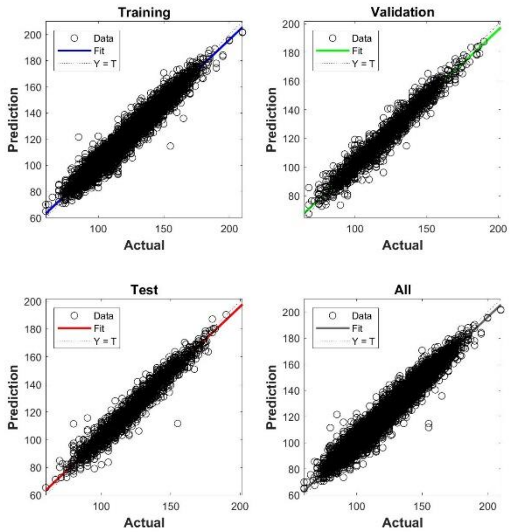


Figure 10. Prediction result from NARX using combined atmospheric data and estimated tides data as input (scenario 3). The diagram shows how the result of prediction varies to the actual data.

In the range of tidal values that are less than 50 cm, the NARX model tends to overestimate, meaning that the predicted tide values produced by the model tend to be higher than the actual values. Conversely, for tide values that are more than 185 cm, the NARX model tends to be underestimated, meaning that the model often predicts tide values that are lower than the actual values.

The high predictive accuracy of the HA-LS and NARX model combination shows significant potential for operationalization in coastal engineering and maritime planning, particularly in Semarang, where tidal flooding poses a persistent challenge. In the context of an early warning system, the model’s precise tidal forecasts can be integrated into a real-time monitoring platform at the Tanjung Mas Maritime Meteorology Station, enabling timely warnings for coastal flooding events, especially during peak flood periods. The model’s accurate

Vol.18, No.02, Desember 2024: page: 76-90
p - ISSN: 1978-886X
e - ISSN: 2614-641X

tidal predictions can be leveraged to optimize port scheduling at the busy Tanjung Mas Port in Semarang, reducing delays and improving navigation safety during tidal events. In addition, the model's output can inform design standards for coastal infrastructure, such as embankments or drainage systems by providing accurate tidal data. However, operationalizing the model requires addressing several challenges, such as ensuring consistent atmospheric data quality and developing an easy-to-use interface for stakeholders, to fully realize its potential in enhancing Semarang's coastal resilience.

CONCLUSION

A combined HA-NARX method, as a new approach for tide prediction, is applied in the tropical region of Semarang, Indonesia, to ensure continuous availability of tidal data. Analysis shows that the NARX model outperforms the HA method with the least squares approach in predicting tide levels in Semarang, based on MSE and RMSE values. Moreover, the use of the second scenario indicates that the NARX neural network model improves the accuracy of tide predictions compared to the HA-LS technique, as evidenced by three statistical indicators used in this paper.

The use of atmospheric data and estimated tides from HA-LS combined together as input in NARX model results in best performance of prediction with MSE and RMSE value of 25.0 and 5 cm, respectively. Moreover, the NARX model in the 3rd scenario performs very well in predicting the tide level, with a coefficient of determination of 0.974 for the entire dataset. This approach performs as better as Wu et al. (2021), even better with overall coefficient determinant value of 0.97, while the coefficient determinant value in station ID 9414290 is approximately 0.94 in the research of Wu et al. (2021). Further analysis of feature importance indicates that the most influential input is the estimated tidal data from the HA-LS method, while atmospheric data is the second most important input in the prediction process. Consistent with previous studies conducted in high-latitude regions, this study also finds that atmospheric data significantly contributes to tide prediction.

Despite its good performance, the model only uses one tidal station in Semarang, which may limit its application to other coastal areas with different tidal characteristics or environmental conditions. In addition, other factors such as land subsidence and sea level rise, which exacerbate flooding in Semarang, are not included in the model. Future research should explore the incorporation of additional inputs, such as land subsidence and sea level rise, and the addition of tidal stations to improve predictions and expand the model's scope.

ACKNOWLEDGEMENTS

This work used primary data from Indonesian Agency for Meteorology Climatology and Geophysics (BMKG).

REFERENCE

Abidin, H.Z., Andreas, H., Gumilar, I., Sidiq, T.P. & Fukuda, Y., 2013, 'Land subsidence in coastal city of Semarang (Indonesia): characteristics, impacts and causes', *Geomatics, Natural Hazards and Risk*, 4(3), 226–240.

Alfahmi, F., Boer, R., Hidayat, R., Perdinan & Sopaheluwakan, A., 2019, 'The Impact of Concave Coastline on Rainfall Offshore Distribution over Indonesian Maritime Continent', *Scientific World Journal*, 2019.

Bagaskoro, A. & Saputro, N.D., 2023,

'Pembuatan Aplikasi AWS Mobile di Stasiun Klimatologi Kelas 1 Jawa Tengah', *Prosiding Seminar Nasional …*, 2023, 606–613.

Boussaada, Z., Curea, O., Remaci, A., Camblong, H. & Bellaaj, N.M., 2018, 'A nonlinear autoregressive exogenous (NARX) neural network model for the prediction of the daily direct solar radiation', *Energies*, 11(3).

Buevich, A., Sergeev, A., Shichkin, A. & Baglaeva, E., 2021, 'A two-step combined algorithm based on NARX neural network and the subsequent prediction of the residues improves prediction accuracy of the greenhouse gases concentrations', *Neural Computing and Applications*, 33(5), 1547–1557.

Chicco, D., Warrens, M.J. & Jurman, G., 2021, 'The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation', *PeerJ Computer Science*, 7, 1–24.

Codiga, D.L., 2011, *Unified Tidal Analysis and Prediction Using the UTide Matlab Functions*, Narragansett.

Djamaluddin, T., Susanti, R.D., Aditiya, A., Pranowo, W.S., Mariyam, A.S., Hasanuddin, A.P., Mareta, L. & Rusnadi, I.E., 2023, *Astronomical tidal constituents potentially trigger coastal floods on the north coast of Java*, 040004.

Erlani, R. & Nugrahandika, W.H., 2019, 'Ketangguhan Kota Semarang dalam Menghadapi Bencana Banjir Pasang Air Laut (Rob)', *Journal of Regional and Rural Development Planning*, 3(1), 47.

Hakim, B.A., Suharyanto, S. & Hidajat, W.K., 2013, 'PENGARUH KENAIKAN AIR LAUT PADA EFEKTIFITAS BANGUNAN UNTUK PERLINDUNGAN PANTAI KOTA SEMARANG', *Buletin Oseanografi Marina*, 2(3), 81–93.

Henderi, H., 2021, 'Comparison of Min-Max normalization and Z-Score Normalization in the K-nearest neighbor (kNN) Algorithm to Test the Accuracy of Types of Breast Cancer', *IJIIS: International Journal of Informatics and Information Systems*, 4(1), 13–20.

Ikhsyan, N., Muryani, C. & Rintayani, P., 2017, 'Analisis Sebaran, Dampak dan Adaptasi Masyarakat terhadap Banjir Rob di Kecamatan Semarang Timur dan Kecamatan Gayamsari Kota Semarang', *Jurnal Geo Eco*, 3(2), 145–156.

Koropitan, A.F. & Ikeda, M., 2008, 'Three-dimensional modeling of tidal circulation and mixing over the Java Sea', *Journal of Oceanography*, 64(1), 61–80.

Li, S., Liu, L., Cai, S. & Wang, G., 2019, 'Tidal harmonic analysis and prediction with least-squares estimation and inaction method', *Estuarine, Coastal and Shelf Science*, 220(March), 196–208.

Marfai, M.A. & King, L., 2008, 'Tidal inundation mapping under enhanced land subsidence in Semarang, Central Java Indonesia', *Natural Hazards*, 44(1), 93–109.

Permana, H., Hariadi, H. & Rochaddi, B., 2012, 'KAJIAN KONDISI ARUS DAN SEBARAN SEDIMEN DASAR PADA SAAT MUSIM TIMUR DI PERAIRAN SEMARANG - DEMAK', *Journal of Oceanography*, 1(1), 121–128.

Pugh, D., 1987, *Tides, Surges and Mean Sea Level: A Handbook for Engineers and*

Vol.18, No.02, Desember 2024: page: 76-90
p - ISSN: 1978-886X
e - ISSN: 2614-641X

Scientists, John Wiley & Sons, Chichester.

Rahili, N., Hakim, B. Al, Ariyanto, D., Cholishoh, E., Aprijanto, Destianingrum, Prabawardani, R., Prijambodo, T. & Gumbira, G., 2023, 'Tidal Characteristics for Disaster Preparedness in the Port Area (Case Study: Port of Semarang, Central Java)', *ILMU KELAUTAN: Indonesian Journal of Marine Sciences*, 28(2), 181–188.

Rayson, M.D., Jones, N.L., Ivey, G.N. & Gong, Y., 2021, 'A Seasonal Harmonic Model for Internal Tide Amplitude Prediction', *Journal of Geophysical Research: Oceans*, 126(10), 1–29.

Tatachar, A. V, 2021, 'Comparative assessment of regression models based On model evaluation metrics', *International Research Journal of Engineering and Technology*, 8(9), 853–860.

Wu, W., Li, L., Yin, J., Lyu, W. & Zhang, W., 2021, 'A modular tide level prediction method based on a NARX neural network', *IEEE Access*, 9, 147416–147429.

Xie, H., Tang, H.A.O. & Liao, Y., 2018, 'Proceedings - International Conference on Machine Learning and Cybernetics', *Proceedings - International Conference on Machine Learning and Cybernetics*, 2(July), 690.

Yin, J.C., Perakis, A.N. & Wang, N., 2018, 'An Ensemble Real-Time Tidal Level Prediction Mechanism Using Multiresolution Wavelet Decomposition Method', *IEEE Transactions on Geoscience and Remote Sensing*, 56(8), 4856–4865.

Zhao, M., Zhang, R., Lin, C., Zhou, H. & Shi, J., 2020, 'Stochastic Model Predictive Control for Dual-Motor Battery Electric Bus Based on Signed Markov Chain Monte Carlo Method', *IEEE Access*, 8, 120785–120797.