Prediction and classification of suspended sediment and zooplankton signals from acoustic Doppler current profiler backscatter data using artificial neural networks

A Dwinovantyo\textsuperscript{1*}, S Solikin\textsuperscript{2}, H M Manik\textsuperscript{2}, T Prartono\textsuperscript{2} and Susilohadi\textsuperscript{3}

\textsuperscript{1} Research Center for Deep Sea, National Research and Innovation Agency (BRIN). Jl Y. Syaranamual Guru-Guru, Poka, Ambon 97233, Maluku, Indonesia
\textsuperscript{2} Department of Marine Science and Technology, Faculty of Fisheries and Marine Sciences, IPB University, Dramaga, Bogor 16680, West Java, Indonesia
\textsuperscript{3} Marine Geological Institute, Ministry of Energy and Mineral Resources Republic of Indonesia. Jl. Dr. Djunjunan No. 236, Bandung 40174, West Java, Indonesia

*E-mail: anggano@gmail.com

Abstract. Characterization of each underwater object has its challenges, especially for small objects. The process of quantifying acoustic signals for these small objects can be done using high-frequency hydroacoustic instruments such as an acoustic Doppler current profiler (ADCP) combined with the artificial intelligence (AI) technique. This paper presents an artificial neural network (ANN) methodology for classifying an object from acoustic and environmental data in the water column. In particular, the methodology was tuned for the recognition of suspended sediments and zooplankton. Suspended sediment concentration and zooplankton abundance, which extracted from ADCP acoustic data, were used as input in the backpropagation method along with other environmental data such as effects of tides, currents, and vertical velocity. The classifier used an optimal number of neurons in the hidden layer and a feature selection based on a genetic algorithm. The ANN method was also used to estimate the suspended sediment concentration in the future. This study provided new implications for predicting and classifying suspended sediment and zooplankton using the ADCP instrument. The proposed methodology allowed us to identify the objects with an accuracy of more than 95%.

Keywords: ADCP, artificial neural networks, classification, prediction, suspended sediment, zooplankton

1. Introduction

1.1. Background

One of the main problems in classifying suspended sediment and zooplankton from the acoustic signal is determining each object's signal characteristics. One of them is to distinguish the acoustic signal obtained while processing the raw data. The conventional method uses simple linear regression analysis to measure the concentration of suspended sediment. However, these techniques, in addition to being time-consuming, may result in low accuracy. The other limitation is that the results' linearity level with environmental conditions remains unknown [1]. The accuracy of the processing results also has limitations for prediction in the future. In addition, the characterization of the acoustic signal is
challenging to distinguish from zooplankton. Incorrect object classification might limit the reliability of zooplankton acoustic abundance estimation [2].

The previous research states that the problem can be solved statistically using an artificial intelligence (AI) approach [3]. Artificial neural network (ANN) is one of the AI methods that can be used. The principle of the ANN method is to use learning techniques in solving problems based on the input information, for example, acoustic backscattering data from acoustic Doppler current profiler (ADCP) instruments. Environmental parameters such as vertical velocity, current, and tide are inputted into the ANN model to get better accuracy results. The capabilities of the ANN model can be used to carry out learning and generate rules or operations from several examples or inputs entered and make predictions about possible outputs that will appear. One technique that is widely used in the ANN method is backpropagation. The advantage of using the ANN method compared to other AI methods is that it can study the input data structure that can produce weights [4] so that it can be used to separate acoustic signals from zooplankton [5] and suspended sediment [6].

In this paper, ANN methodology was applied to identify and classify suspended sediments and zooplankton. Compared to other literature, one of the novelties of this paper was to use a combination of environmental variables such as vertical velocity, current, and tidal. A prediction of suspended sediment concentration (SSC) using the ANN method was also performed. This research needs to be done to optimize the signal processing method of the ADCP instrument, especially in improving its ability to monitor suspended sediment conditions; and monitoring the presence, abundance, and behavior of zooplankton in the water column.

1.2. Aim

The purpose of this study was to predict the suspended sediment concentration and classify ADCP data to identify signal sources from suspended sediments and zooplankton using the ANN method. The acoustic signal classification was done by studying the acoustic backscatter pattern and vertical velocity from ADCP data.

2. Methodology

The schema of the framework in this research was data collection, pre-processing, feature selection, and classification (Figure 1). Raw data consists of in-situ observation data and acoustic backscatter data from the ADCP instrument. In-situ data consists of water samples, zooplankton samples, and time-series data on environmental conditions (temperature, salinity, pH, and tides).

![Figure 1. The workflow of the proposed methodology.](image-url)
ADCP data in echo intensity (EI) were processed and converted to mean volume backscattering strength (MVBS). Water current data were processed to obtain horizontal and vertical flow velocity profiles. The pre-processed data was integrated with in-situ observation data to speed up the ANN model's response and reduce the presence of noise features or low significance features. The classification results were obtained for the determination of suspended sediment and zooplankton.

2.1. Data Collection
The research location was in the Lembeh Strait, North Sulawesi Province, Republic of Indonesia. The mooring Nortek ADCP 750 kHz instrument was placed at the bottom of the water with coordinates 01°27'53.34" N - 125°14'05.56" E on April 3, 2016, and lifted on April 28, 2016. One of considerations for choosing this location was because of the input from the mainland such as activities at Port of Bitung. In addition, the oceanographic conditions of the Lembeh Strait are strongly influenced by tidal currents; the water mass transport from the Sulawesi Sea to the Maluku Sea and vice versa, which determines the dynamics of the aquatic environment. The average depth of the observation site was 20 m. Mobile ADCP, Teledyne RDI 307.2 kHz data acquisition, was carried out in a cross-section by the symbol (→) in Figure 2, from the northeast coast of Sulawesi Island to the west coast of Lembeh Island through the location where the ADCP mooring was placed. Data collection of water and zooplankton samples were indicated by the symbol (○) in Figure 1 when the mobile ADCP data acquisition was performed.

![Figure 2. Map of the location of mooring ADCP data collection points, water sampling location for suspended sediment and zooplankton samples (○), and mobile ADCP transect (→) in the Lembeh Strait](image)

2.2. Pre-processing
The conversion of raw data (counts) from ADCP was done to quantify suspended sediment concentrations (mg l⁻¹) and zooplankton (ind m⁻³). All analyses were carried out using MATLAB (Mathwork, Inc.). The ADCP instrument received the acoustic signal amplitude in units of count in echo intensity (E). Er is the noise level of all beams by calculating the minimum echo intensity value. The Er value used in this study was 40 counts for ADCP frequency of 307.2 kHz and 43 counts for ADCP frequency of 750 kHz. The value of acoustic backscatter (RB) can be calculated after knowing the value of two-way transmission loss (2 TLs) and echo level measured on the transducer (RL) and adding up the two values. According to previous research [7], the RB value can be calculated by the equation:

\[ RL = K_c (E - Er) \]
\[ Kc = \frac{127.3}{T} + 273 \]  
\[ TL = 20 \log_{10} R + 2 aR \]  
\[ RB = RL + 2 TL \]  
\[ SSC = 10^{(A + B \cdot RB)} \]  

Computing the conversion of EI to MVBS for zooplankton quantification based on sonar equations which can be written as follows [8, 9]:

\[ MVBS = C + 10 \log((T_x + 273.16)R^2) - L_{DBM} - P_{DBW} + 2aR + 10 \log(10^{k_c(\delta - \delta_r)/10} - 1) \]  

MVBS is the mean volume backscattering strength (dB m⁻¹), C is a constant involving several parameters depending on the ADCP instrument used (dB), \( T_x \) is the temperature measured at the transducer (°C). The \( L_{DBM} \) is logarithmic of the transmitted wavelength (log(m)), \( P_{DBW} \) is the logarithm of transmit power (log(W)), \( K_c \) is the conversion factor. The value of \( a \) is the absorption coefficient (dB m⁻¹) [10]. The estimation of zooplankton abundance is a rough estimate because the TS measurement is only done theoretically. The MVBS value for zooplankton analysis is used to estimate the abundance of zooplankton based on the equations that have been developed [11 - 13]:

\[ \{MVBS(f)\} = n T_s(f) \]  
\[ MVBS_f (dB) = 10 \log \rho \cdot TS_f (dB) \]  
\[ TS_f = 10 \log (\sigma (f_0, a_0) / 4\pi) \]  
\[ N(z) = \frac{<\sigma_0(f_0,a_0)/\sigma(f_0,a_0)>}{\sigma(f_0,a_0)} \]

The TS of zooplankton (dB) values used in this study were the theoretical TS obtained from the DWBA model [14], \( N(z) \) is the abundance of zooplankton at a certain depth, \( \sigma_v \) value is the ensemble average of single pings on the volume backscatter strength, \( \sigma \) is acoustic scattering cross-section with a specific frequency of equipment and size of zooplankton, \( f_0 \) is the frequency used in ADCP, and \( a_0 \) is individual zooplankton. The relationship between acoustically estimated zooplankton abundance and suspended sediment concentration was calculated using polynomial linear regression equations as follows:

\[ y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \cdots + \beta_k x_i^k + \epsilon_i \]  

Where \( y \) is acoustically estimated zooplankton abundance based on the theory and simulated MVBS and \( x \) is the suspended sediment concentration. The method was used to describe the relationship between zooplankton abundance and suspended sediments variables which are curvilinear.

2.3. Feature Selection

The feature selection process was a fundamental thing to do before being processed further by AI methods. This process aimed to speed up the learning process on the ANN model. In addition, feature selection could reduce noise features that can cause errors in the classification process. The principle of feature selection was to find elements that were a subset of the feature partition set that optimized classification performance. The method used was a filter approach by looking at the intrinsic properties of the input data used. The approach was based on feature statistics applicable in binary labeling of the data set [15].

The binary label identified two paired samples, zooplankton and suspended sediment, for each feature \( j \) and the absolute value \( |\delta| \) of the difference between the mean features of the two samples. The normalization process was carried out by calculating the number of variants used for selection. The high
value of $z_j$ supports the selection of feature $j$. Each value of $z_j$ was then weighed by the weight of $w_j$, defined as follows:

$$w_j = z_j * (1 - \alpha * \rho_j)$$

(12)

Where $\rho_j$ is the average absolute value of the Pearson correlation between candidate features $j$ and the others, $\alpha$ is a weighting factor between the values of 1 (weighted) and 0 (unweighted). The final output can be obtained from this equation in the form of a rating of the entire feature set based on weighted $w$.

2.4. Prediction and Classification

The ANN model was used to predict and classify suspended sediment and zooplankton by studying zooplankton’s movement patterns and behavior every day and calculating the weight of the relationship between zooplankton signals and suspended sediment. The method for determining the weight was the training process. The data used as input in the training process were time-series data of suspended sediment concentration, currents, tides, and acoustically zooplankton abundance. All input data were used to compare the best results in estimating suspended sediment concentration based on the highest correlation coefficient ($r$) [16]. The training process consisted of feedforward and backpropagation. In the feedforward process, it was necessary to have the number and size of the layer to be formed, the size of the subsampling, and the input data. The results of the feedforward process were in the form of weights that will be used to evaluate the backpropagation process.

ANN modeling was carried out using MATLAB, GMDH Shell DS, and Python software to select, sort data, and process all stages of the ANN simulation. Figure 3 describes the backpropagation model with input layers $x_1$, $x_i$, and $x_n$; hidden layers $z_j$, $z_i$, and $z_n$; and output layers $y_1$, $y_k$.

![Figure 3](image-url)

**Figure 3.** The model architecture we developed and proposed: multi-layer perceptron (MLP) with ANN method on (a) suspended sediment and (b) zooplankton, based on the results of feature selection analysis.

In Figure 3, several neurons were in the input layer and one node in the output layer. Between the input and output layers, there was a hidden layer with three neurons. The number of neurons in the input layer was equal to the number of attributes in the pattern to be recognized, while the number of neurons in the output layer was equal to the number of pattern classes. The input layer was in charge of forwarding the input and not doing the computations, but the hidden layer and the output layer computing for the learning process. The purpose of learning was to adapt the weight to produce the desired output. The vector notation of the input and output layers of the hidden and output layers is expressed in the following equation:

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_I \end{bmatrix} \quad y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_K \end{bmatrix} \quad o = \begin{bmatrix} o_1 \\ o_2 \\ \vdots \\ o_K \end{bmatrix}$$

(13)
Variables $I$, $J$, and $K$ represent the number of neurons in each layer. The number of hidden layers and neurons in each layer depends on the complexity of the problem. Usually, one hidden layer is sufficient for some applications. Meanwhile, trial and error is the best way to determine the number of neurons in the hidden layer.

The weight specification in the ANN architecture is described as each interconnected neuron, marked with an arrow in Figure 3. Each connection has a weight which later, the value of each weight will be different. The hidden and output layers have an additional input which is known as bias. Overall, 17 parameters in the training process undergo epoch changes to get the best results. The neurons in the input layer do not have an activation function.

In contrast, the neurons in the hidden and output layers have an activation function that sometimes differs, depending on the data or the problem of the data arrangement. The prediction model was developed based on standard backpropagation Levenberg-Marquardt (LM) training algorithm [6], defined as follows:

$$y_k(x_i) = x_i - [J^T J + \mu I]^{-1} J^T e$$

where $x_i$ is the vector of weights and biases of $i$, $J$ is Jacobian matrix which contains the first derivative of the network error regarding weights and biases, $e$ is the network error vector, and $I$ is the identity matrix. The greater the weight of a relationship, the more important the relationship between the two layers is.

In addition, the ANN model can classify suspended sediments and zooplankton using the result of the training data. The difference between ANN applications for static ADCP and mobile ADCP lies only in the information for each pixel of the ADCP data. In the case of suspended sediment and zooplankton classification, each pixel was an interrelated feature because it detected objects detected by ADCP, not an independent feature [3]. The accuracy of the network was evaluated by the mean squared error (MSE) and the coefficient of correlation, $R$.

### 3. Results and Discussion

#### 3.1. Prediction

The implementation of the backpropagation method began with the initiation of a network model consisting of three layers, including input with neurons (predictor attribute), a hidden layer with 50 neurons, output with ten neurons (target attribute). The variables used as network constructors include the learning rate coefficient of 0.1, the momentum coefficient of 0.5, the maximum number of epochs or iterations of 1000, and the target error of 0.1. The input data consisted of time series data of suspended sediment concentration, current velocity, and tides.

| Process / Step | Amount of data | MSE     | $r$   |
|----------------|----------------|---------|-------|
| Training       | 10030          | 4.3333  | 0.9840|
| Validation     | 2150           | 5.0108  | 0.9804|
| Testing        | 2150           | 3.5131  | 0.9873|

| Total Data     | 14330          |         |       |

Daily data on April 2-22, 2016, obtained from processing acoustic signals from ADCP to determine the concentration of suspended sediment totaled 14,330 data. Of the total for 21 days, 70% of the data, 10,030 data were selected for model training, 15% (2,150 data) were selected for model validation, and 15% (2,150 data) were selected for model testing. Subsampling of all data for training and validation
was done using the 'divideind' function, i.e., the data was divided by the bin depth of MATLAB. Table 1 was shown that the input data was correlated (as indicated by the value of $r$ close to 1). The input data was processed at each hidden layer. The error factor was minimized by repeating the training, validation, and testing processes until the ANN model was optimal.

The results were shown as a good correlation based on the minimum MSE and high $r$ values which shown $r>0.95$ between the observed and simulated suspended sediment concentration values. The evaluation for the model was based on the data simulation for 24 days. The accuracy of evaluating the model used, generated by the network for the training stage, was 98.41%. The accuracy of the validation data generated or the output of the trained network compared to the validation data target was 98.04%. The level of prediction accuracy using testing data with a model that has been trained and validated was 98.73%. Out of the 35 epochs carried out, the best network performance was at an error rate of 4.8581 in the 29th epoch. It might be concluded that the training results of the ANN model can be used to predict suspended sediment in the case study in the Lembeh Strait.

![Image](image.png)

**Figure 4.** Prediction of suspended sediment concentration from ANN simulation results for 7 days at 11.5-16 m of depth which is shown in red line color

The training data were then processed to predict the SSC based on time series data. The result was shown which was a strong spread in 11.5 – 16 m depth (Figure 4). The average trend of the distribution of suspended sediment concentrations was shown as a blue color representing model fit from ANN, grey as actual data, red line as a prediction of SSC, and red color as a trusted model range. The distribution of predicted suspended sediment concentration with the influence of tides and currents for seven days. We did not predict the abundance of zooplankton because zooplankton are living things that require more complex parameters so that they cannot be analyzed using ANN [17]. This was caused because ANN method was unable to fully capture the behavior of a nonlinear time series containing inherent complex components.
3.2. Classification

Analysis using ANN required data to be selected with linear forms such as suspended sediment concentration (mg l\(^{-1}\)) from previous research [1] and abundance of zooplankton (ind m\(^{-3}\)) which was calculated by the relationship between MVBS and TS values [14]. Using linear forms, the data from the training carried out on logarithmic data (dB) from MVBS and RB obtained a small correlation coefficient (\(r\)), around 0.1. The selection of input data was crucial before the training data simulation was carried out. Before the training process was carried out, data plots were carried out to observe the data distribution pattern. The estimation results of suspended sediment concentration and zooplankton abundance were compared across all-time series data. A comparison of the acoustically estimated zooplankton abundance and suspended sediment concentration was shown in Figure 5.

The scatter plot graph in Figure 5 shows the relationship between the estimated abundance of zooplankton and suspended sediment concentration from the time series data. On the curve, the variable value of suspended sediment concentration and zooplankton abundance decreases with a quadratic relationship between variables, as seen from the resulting curve forming a curved line. The concentration of suspended sediment and abundance of zooplankton had a correlation coefficient (\(r\)) of 0.6935, indicating that the relationship between the two was quite close.

The concentration of suspended sediment could affect the abundance of zooplankton in the water column. This relationship could be seen from the positive trend pattern on the quadratic line. If suspended sediment concentration increased, the abundance of zooplankton could also increase to a certain saturation point with a decreasing trend [18]. This line could decrease after the suspended sediment concentration was above 100 mg l\(^{-1}\). This decrease was related to the limitations of the acoustic method in detecting individual zooplankton when the suspended sediment concentration was very high. The amount of suspended sediment data was at a concentration of 60–70 mg l\(^{-1}\), as many as 952 data.

In theory, the calculation of zooplankton abundance might be identified by knowing the proportion of suspended sediment concentration [19]. However, this could not be done from the research results because many factors did not support it, such as the need for a higher frequency and an ADCP bin
resolution smaller than 1 m. The training data was carried out with 70% data, 15% data for validation, and 15% data for testing. The implementation of the neural network method began with the initiation of the suspended sediment data training model and the abundance of zooplankton from moored and mobile ADCP, as shown in Figure 6.

Figure 6. Distribution graph of ANN training results on zooplankton abundance and suspended sediment concentration

Figure 6 shows the distribution of the suspended sediment data on the x-axis and the abundance of zooplankton on the y-axis obtained, with colored dots indicating the accuracy and loss of training data that illustrated the accuracy of the training data. The result of the first model run shows a close relationship between suspended sediment and zooplankton, with the code ‘matched 1’. Re-training of the data was carried out by rearranging the epoch to get a more optimal model, shown with the code ‘matched 2’. This process was repeated until the optimal condition of the ANN model was obtained, shown by the orange dot. The trained data were then processed by validation and testing, as shown in Figure 7.

Figure 7. The results of processing training, validation, test, and overall data for (a) zooplankton abundance and (b) suspended sediments simulation using the ANN model
The training results on zooplankton abundance have a correlation coefficient of 0.9718, while the correlation coefficients at the validation and test stages are 0.9721 and 0.9671, respectively. The best data validation performance was obtained at the 145\textsuperscript{th} epoch. This training process took 10 hours to achieve the ideal model with a total of 14,300 data. To achieve ideal conditions, the ANN model required a training process with a learning rate that matched the ideal conditions achieved, obtained from a learning rate of 0.001. A learning rate value that was too high can cause an unstable algorithm that affects the classification results. On the other hand, it might cause the algorithm to take longer to reach the convergence and optimal conditions [4].

The distribution and unique characteristics of the abundance of zooplankton scattered in the temporal and spatial data were used to simulate the model in the classification network. This approach was used to select the input features of zooplankton, and the results of this ANN classification were then compared against the suspended sediment concentration data network [5]. The distribution graph of the training results of zooplankton abundance data was shown in the background signal and zooplankton data from the data population collection, identified from the different patterns between the abundance of the ANN model analysis and the original data of zooplankton abundance acoustically. The zooplankton data were extracted for the classification process using ANN. The data trained using the ANN was used to classify zooplankton, as shown in Figure 8.

![Graph](image)

**Figure 8.** The distribution of zooplankton resulting from the classification of sound signals from (a) moored ADCP time series data and (b) mobile ADCP spatial data using the ANN model

The results in Figure 8(a) show time series data of zooplankton abundance taken from static ADCP data. We can see the temporal distribution of zooplankton patterns that spread at a depth of 11.5 - 16 m. The results in Figure 8(b) show the abundance of zooplankton from mobile ADCP data in one cross to show the spatial data. The ANN model could determine and identify zooplankton populations with a correlation level of 97\% of all test data. The analysis of sound signals scattered by zooplankton has unique characteristics. It was retrained with input in vertical speed so that the ANN model could classify
these signals in the zooplankton category. The results obtained by the ANN model showed that there were scattered but consistent schools of zooplankton at a depth of 12 m.

The same treatment was carried out on the response of suspended sediments in scattering sound waves. The suspended sediment concentration was obtained from the conversion of the RB value to the empirical equation. The classification of suspended sediment was carried out with the previous analysis results, which states that the concentration of suspended sediment spread in the water column with different concentrations at each depth. In the previous stage, the suspended sediment concentration could be predicted using the ANN method. The suspended sediment was classified using the ANN method to isolate other scattering sources using different concentration characteristics. The purpose of the suspended sediment classification was to separate the suspended sediment and zooplankton to analyze the signal. A separation process was obtained to classify suspended sediment, zooplankton, or a mix of suspended sediment and zooplankton. The classification of suspended sediment from mooring ADCP and mobile ADCP was shown in blue in Figure 9.

![Figure 9](image_url)

**Figure 9.** Distribution of suspended sediment resulting from sound signal classification from (a) moored ADCP time series data and (b) mobile ADCP spatial data using the ANN model

The results in Figure 9(a) show time series data of SSC taken from static ADCP data, which has a concentration >85 mg l⁻¹. The suspended sediment spread along with the depth of water with variation of the concentration, so we only classify by defining the higher value of SSC. The results in Figure 9(b) show the SSC from mobile ADCP data to show the spatial data. The classification of suspended sediment at this stage involved input parameters of currents, tides, and suspended sediment concentration data. These parameters were trained to produce a beneficial learning outcome for the classification of suspended sediments. The training process was repeated several times until the optimal conditions were obtained, then the ANN classified the parameters and acoustic signals into a more efficient process.

The advantage of the ANN method in the classification of suspended sediment and zooplankton was to simulate many classification stages based on the parameters that were inputted simultaneously [15]. Time series data with criteria of abundance and high suspended sediment concentration indicated the
primary source of acoustic backscatter. The ANN method studies inputs such as suspended sediment concentrations, then related them to currents at the same time conditions, then tidal conditions, and finally connected with zooplankton data that have been identified and classified at the previous stage. The entire data could become data points that help AI determine what objects were at that depth. After being trained several times until the ANN model used was optimal, the classification of suspended sediment, zooplankton, and suspended sediment and zooplankton could be obtained in an area at one time. The training process that has been carried out produces weights that are evaluated statistically and used to calculate accuracy [5]. The classification results are presented in Figure 10.

![Distribution of suspended sediment, zooplankton, and the mix of suspended sediment and zooplankton resulting from sound signal classification from (a) mooring ADCP time series data and (b) moving ADCP spatial data using the ANN model](image)

**Figure 10.** Distribution of suspended sediment, zooplankton, and the mix of suspended sediment and zooplankton resulting from sound signal classification from (a) mooring ADCP time series data and (b) moving ADCP spatial data using the ANN model

The yellow color showed the object in the form of zooplankton. The blue color showed the suspended sediment object, and the red color showed the mixture of suspended sediment and zooplankton. The modeling performance using the ANN model was evaluated by statistical analysis of mean absolute error (MAE), root means square error (RMSE), coefficient of determination ($R^2$), and coefficient of efficiency (CE). Ideally, the MSE, RMSE, and MAE values should be zero, while $R^2$ and CE should be one. The error value can be zero, and the coefficient of determination value can be one of the observed and predicted values. The accuracy of the ANN model is shown in Table 2.

**Table 2.** Evaluation of the accuracy of the ANN model for the classification of suspended sediment and zooplankton

| No. | Classification Model | MAE    | RMSE   | $R^2$  | CE    | % Error |
|-----|----------------------|--------|--------|--------|-------|---------|
| 1   | Suspended Sediment   | $2 \times 10^{-7}$ | 1.4133 | 0.9840 | 0.9867 | 1.60    |
| 2   | Zooplankton          | $2 \times 10^{-7}$ | 2.6039 | 0.9711 | 0.9773 | 2.29    |
| 3   | Suspended Sediment and Zooplankton | $2 \times 10^{-7}$ | 0.8125 | 0.9883 | 0.9914 | 1.11    |
The overall ANN model used in the suspended sediment and zooplankton classification showed that the total predicted value was slightly higher than the observed value, indicated by a positive MAE value. The difference in RMSE, $R^2$, and CE values in the three models are almost the same, so it could be concluded that the ANN model contends with reasonable accuracy. The model also proved that the input to the ANN model from time-series data of suspended sediment concentration was correlated with tidal and current conditions. Time-series data of zooplankton abundance was related to the results of vertical velocity analysis obtained from ADCP measurements [20].

The accuracy of modeling using ANN was highly dependent on the hidden neurons shown from the parameters. The MAE, MSE, and RMSE values of the training might be lower along with the increase in the number of hidden neurons because the greater the number of hidden neurons, the difference in the output values in the model might be slightly different or the same as the actual data. The accuracy of the ANN model obtained has a high value with a percent error below 3%. These results provided a good relationship between suspended sediment objects and zooplankton in proving their relationship and presence in the water column with environmental influences.

The accuracy of the ANN model could be improved by adding other environmental parameters such as sunlight conditions, hydrodynamic conditions, and ground truth data of suspended sediment concentration and zooplankton abundance continuously. In other research, the ANN method was also used to classify small fish groups for the recognition of anchovy, sardine and horse mackerel by extracting the morphological of the fish [15]. This paper's feature selection and classification method improved the model using more complex data to increase measurement accuracy in future research.

4. Conclusion

An artificial neural model was successfully developed to predict and classify suspended sediment and zooplankton from the acoustic backscatter data of the ADCP instrument. The backpropagation ANN network with Levenberg-Marquardt training algorithm was used to train the ADCP acoustic data with environmental conditions such as vertical velocity, current velocity, and tides to obtain more accurate predictions and classifications. The main conclusions obtained in this study were found that the prediction result of ANN model from ADCP data combined with environmental data shown a good result with $r = 0.9873$. The ANN method also shown a successful classification for zooplankton and suspended sediment with have $r > 0.95$. The results of this work, which also consider two small objects detected from ADCP instruments, seemed to be promising and applicable in other areas of the Indonesian sea.

**Author's Contributions:** Angga Dwinovantyo contributed as the first author: wrote the main manuscript text, experimental design, data analysis, and interpretation. Others contributed as co-authors by providing a contribution in manuscript preparation.

**Acknowledgments:** The authors would like to thank Marine Geological Institute, Bandung, Indonesia, for supporting this project.

**Conflicts of Interest:** The authors declare that there are no conflicts of interest.

**References**

[1] Dwinovantyo A, Manik H M, Prartono T and Susilohadi S 2017 Quantification and analysis of suspended sediments concentration using mobile and static acoustic Doppler current profiler (ADCP) instruments. *Adv. Acoust. Vib.* 2017 4890421

[2] Lee K, Mukai T, Lee D J and Iida K 2014 Classification of sound-scattering layers using swimming speed estimated by acoustic Doppler current profiler. *Fish. Sci.* 80 1–11

[3] Demirci M, Unes F and Saydemir S 2015 Suspended sediment estimation using an artificial intelligence approach. *Sediment Matters* Springer, Cham. pp 83-95
[4] Abu-Mostafa Y S, Ismail M M and Lin H T 2012 Learning From data: a short course (California: AMLBook.com.) p 215
[5] Lavery A C and Stanton T K 2001 Acoustic classification of individual zooplankton using artificial neural network. J. Acoust. Soc. Am. 109(5) 2286
[6] Kumar D, Pandey A, Sharma N and Flugel WA 2016 Daily suspended sediment simulation using machine learning approach Catena 138 77-90
[7] Gartner J W 2004 Estimating suspended solids concentrations from backscatter intensity measured by acoustic Doppler current profiler in San Francisco Bay, California. Mar. Geol. 211 169-187
[8] Deines K L 1999 Backscatter estimation using broadband acoustic Doppler current profilers. In Proceedings of the IEEE Sixth Working Conference on Current Measurement (Cat. No.99CH36331) pp 249–253
[9] Mullison J 2017 Backscatter estimation using broadband acoustic Doppler current profilers - updated. In Proceedings of the ASCE Hydraulic Measurements & Experimental Methods Conference, Durham, NH. pp 1–5
[10] Francois R E and Garrison G R 1982 Sound absorption based on ocean measurements. Part II: Boric acid contribution and equation for total absorption. J. Acoust. Soc. Am. 72 1879–1890
[11] Holliday D V and Pieper R E 1995. Bioacoustical oceanography at high frequencies. ICES J. Mar. Sci. 52 (3-4) 279-296
[12] Roman M Holliday D and Sanford L 2001 Temporal and spatial patterns of zooplankton in the Chesapeake Bay turbidity maximum. Mar. Ecol. Prog. Ser. 213 215-227
[13] Brierley A S, Saunders R A, Bone D G, Murphy E J, Enderlein P, Conti SG and Demer DA 2006 Use of moored acoustic instruments to measure short term variability in abundance of Antarctic krill. Limnol. Oceanogr. Methods 4 18–29
[14] Dwinovantyo A, Manik H M, Prartono T, Susilohadi S and Mukai T 2019 Variation of zooplankton mean volume backscattering strength from moored and mobile ADCP instruments for diel vertical migration observation. Appl. Sci. 9 (9) 1851
[15] Aronica S, Fontana I, Giacalone G, Lo Bosco G, Rizzo R, Mazzola S, Basilone S, Ferreri R, and Bonanno A 2019 Identifying small pelagic Mediterranean fish schools from acoustic and environmental data using optimized artificial neural networks. Ecol. Inform. 50 149–161
[16] Targhi A T, Abbasszadeh S and Arabasadi Z 2017 A hybrid method for forecasting river suspended sediments in Iran. Int. J. River Basin Manag. 15 (4) 453-460
[17] Tealab A, Hefny H, and Badr A 2017 Forecasting of nonlinear time series using ANN. Future Comput. Inform. J. 2 (1) 38-47.
[18] Simmonds E J and MacLennan D N 2005 Fisheries Acoustics: Theory and Practice, Second Edition. (Oxford: Blackwell Science)
[19] Sawada K, Mukai T, Fukuda Y, and Matsuura T 2016 Comparison of zooplankton density estimated by acoustic inversion method and net sampling. J. Acoust. Soc. Am. 140 (4) 3243-3243.
[20] Mohn C, Denda A, Christiansen S, Kaufmann M, Peine F, Springer B, Turnewitsch R and Christiansen B 2018 Ocean currents and acoustic backscatter data from shipboard ADCP measurements at three North Atlantic seamounts between 2004 - 2015. Data Br. 17 237–245