C O M M O N S   D E E D

저작자표시-비영리-변경금지 **2.0** 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:

저작자표시. 귀하는 원저작자를 표시하여야 합니다.

비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.

변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 이용허락규약(Legal Code)을 이해하기 쉽게 요약한 것입니다.

Disclaimer

Master's Thesis

User-Friendly Reliable ANN Model for Stability Number of Rock Armors and Tetrapods

실무자들을 위한 신뢰성 있는 사석 및
테트라포드 피복재의 안정수 인공신경망 모델

February 2018

Graduate School of Civil and Environmental Engineering
Seoul National University

In-Chul Kim

ABSTRACT

The stability number of rubble mound breakwaters determines the appropriate weight of armor units of concrete or rock required to resist the wave condition. Therefore, the prediction of suitable stability number is necessary for the stability of the breakwaters. Many empirical formulas have been developed for the stability number since Hudson (1959). To improve the empirical formulas which had significant differences between observed data and prediction data, the machine learning, ANN in particular, has been used during the last two decades. However, most of ANN models did not deal with reliability assessment such as confidence interval. In addition, they are seldom used by practicing engineers probably because most of them did not provide them with an explicit calculation method. In this study, to solve these problems, bootstrap resampling technique was used to make the information or assessment of the reliability in prediction. Also, Excel files made with the by-products of the ANN model such as weights and biases are provided, so that practicing engineers can easily use ANN model.

Keywords: Tetrapod, armor stone, rock armor, stability number, machine learning, artificial neural networks, assessment of the reliability, confidence interval

Student Number: 2016-21249

Table of Contents

List of Tables

List of Figures

List of Symbols

Latin Uppercase

A Eroded cross-sectional area of the profile

D_n Nominal size for Tetrapod

D_{n50} Nominal size for rock armor

H_s Significant wave height

K_D Stability coefficient

N Number of incident waves

N_a Number of Tetrapods per m^2

N_{od} Number of units displaced out of the armor layer

N_s Stability number

$\left(N_{s,pred}^{i}\right)_j$ Output of j-th ANN model for i-th data

$N_{s,pred}^{i}$ Predicted stability number for i-th data

$\overline{N}_{s,pred}$ Average of predicted stability number

$N_{s,obs}^{i}$ i-th observed stability number

$\overline{N}_{s,obs}$ Average of observed stability number

$\left(N_{s,pred}^{i}\right)_\alpha$ $100 \cdot a$ % quantiles of 500 ANN model outputs for i-th data

$\left(\bar{N}_{s,pred}\right)_{\alpha}$ Average of $100 \cdot a$ % quantiles of 500 ANN model outputs

M Mass of the armor unit

P Permeability coefficient

R Correlation coefficient

R_c Crest elevation

S Dimensionless damage level

SS Spectrum shape

T_m Mean wave period

T_p Peak wave period

I_a Index of agreement

W_{50} Average mass of graded rubble

WF_i Weight factor in RMSE

<u>Latin Lowercase</u>

a_j^h Output in hidden layer

a_j^o Output in output layer

b_j^h Bias in hidden layer

b_j^o Bias in output layer

$c_{confidence}$	Confidence interval coefficient
f^h	Activation function in hidden layer
f^o	Activation function in output layer
g	Gravity
h	Water depth
n	Number of elements in data set
n_j^h	State in hidden layer
n_j^o	State in output layer
s_{0m}	Mean wave steepness
$s_{N_s}^i$	Standard deviation of 500 ANN model outputs for *i*-th data
$\overline{s}_{N_s}$	Average of standard deviation of 500 ANN model outputs
$w_{j,i}^h$	Weights in hidden layer
$w_{j,i}^o$	Weights in output layer
$x_{\min}$	Minimum of input variables
$x_{\max}$	Maximum of input variables
x_i	Input variables
x_i^n	Normalized input variables

Greek Uppercase

Δ	Relative buoyant density

Greek Lowercase

α	Slope angle of structures
ξ_m	Surf-similarity parameter
ρ	Mass density of water
ρ_a	Mass density of rock or concrete
ϕ	Packing density
ϕ_{SPM}	Packing density given in the Shore Protection Manual

A prime (′) indicates the scaled parameter to $H_s = 1$ m

CHAPTER 1. INTRODUCTION

1.1 Background

The stability number is the most important parameter describing the relation between the wave condition and the stability of the rubble mound breakwaters. Since the stability number determines the suitable weight of armor units of concrete or rock required to resists the wave condition, the prediction of stability number is very important. For rock armor, well-known empirical formulas were developed by Hudson (1959) and Van der Meer (1987a). Even if these formulas were developed from a number of experimental data, significant differences between the observed stability numbers and the predicted ones were a fatal drawback. To overcome the drawback, the machine learning methods were introduced: Artificial neural networks (ANN) models of Mase et al. (1995), Kim and Park (2005), Balas et al. (2010), and Lee et al. (2016); fuzzy model of Erdik (2009); and M5' model tree of Etemad-Shahidi and Bonakdar (2009). For Tetrapod, there are several empirical formulas: Hudson (1959), De Jong (1996), Van der Meer (1988b), and Suh and Kang (2012).

Most of the empirical formulas showed significant differences in prediction. Most of the formulas using machine learning did not consider reliability, especially confidence interval. Also, they are hard to be used by user since researchers might not provide them with an explicit calculation method.

1.2 Previous studies

1.2.1 The formulas of stability

Hudson (1959) developed the empirical formula by the experiments involved building an armor unit slope and exposing it to increasing wave heights. Van der Meer (1987a) designed formula for rock armor including the influence of wave period, number of waves, and the permeability of the core. For Tetrapod, Van der Meer (1987b) proposed formula for surging breaker type, including the influence of wave period, storm duration, and damage level. De Jong’s (1996) formula for plunging breaker type additionally considered the influence of crest elevation and packing density. The Suh and Kang’s (2012) formula is applicable to breakwaters with various slope angles and packing density, and it is also applicable to low-crested breakwaters.

1.2.2 The reliability of ANN prediction

Artificial neural networks (ANN) are data driven modelling techniques that have been widely used and are useful techniques that can be utilized when the physical mechanism between variables is complex or unknown. However, there are some reasons why the reliability of ANN prediction should be considered. First of all, ANN is a black box modelling, that is, it cannot be explained how learning from input data was done, therefore, ANN prediction should be accompanied by information regarding its reliability. In addition, ANN are a kind of interpolation or extrapolation since the data-patterns are used to calibrate ANN. Thus, it would be another reason why taking into account the reliability of prediction is necessary.

Several studies have taken into account the reliability of ANN prediction in coastal structures. Van Gent and Van den Boogaard (1998) showed ANN prediction of the horizontal forces for the design of vertical breakwaters and the prediction method that has information on the reliability of ANN prediction. The information is reliability-intervals around a prediction value depending on the significant level. Also, Van Gent et al. (2007) developed ANN prediction method for the wave overtopping and confidence intervals of it. Van Gent et al. (2007) used resampling techniques to assess the uncertainty of the ANN prediction, using the set of 500 ANN to provide the estimation value and the uncertainty of the model with regard to the accuracy of its prediction.

1.3 Objectives and research overview

The ultimate objective of this research is to suggest two ANN prediction methods for the stability number which has reliability assessment such as confidence interval. One is for rock armor and the other is for Tetrapod. The ANN prediction methods would be given as Excel files to be easily used by practicing engineers.

This study proceeds in the following order. First, input parameters were selected for the rock armor and Tetrapod. Next, the data for the rock armor and Tetrapod were scaled to $H_s = 1$ m satisfying Froude similitude law, so that the model created by the experimental data can be used directly for the prototype conditions. Third, the test data set should be extracted from the entire data set, and test data set should not be used in training and only used in test. Thus, the entire data set is classified into the candidates for training and validation data (resampling data set) and the test data set. Fourth, training and validation data set should be selected to make ANN model. In this process, the resampling technique was used to guarantee the representativeness of training and validation data. Resampling was performed 500 times to satisfy a sufficient number of times and bootstrap resampling was utilized. Fifth, the numbers of hidden neurons should be determined for each ANN models of rock armor and Tetrapod. The numbers of hidden neurons were determined before the performance of the model in the validation data set decreases, considering RMSE (Root Mean-Square Error), IOA (Index of Agreement), and R (correlation coefficient) values. The method is to prevent the ANN model from overfitting with training data set. Next, ensemble average of 500 values using 500 ANN models was used as a predicted value. For confidence interval, two methods were used. One is to use the standard deviation of 500 values and the other is to use the quantiles of 500 values. After constructing ANN models, to verify that the selection of input parameters and the number of model are reasonable, two tests about them were conducted. After finishing verification of ANN prediction model, performance evaluation of method was conducted with the entire and test data set. Finally, the Excel files were made with weight and bias of 500 ANN models, making engineers use easily.

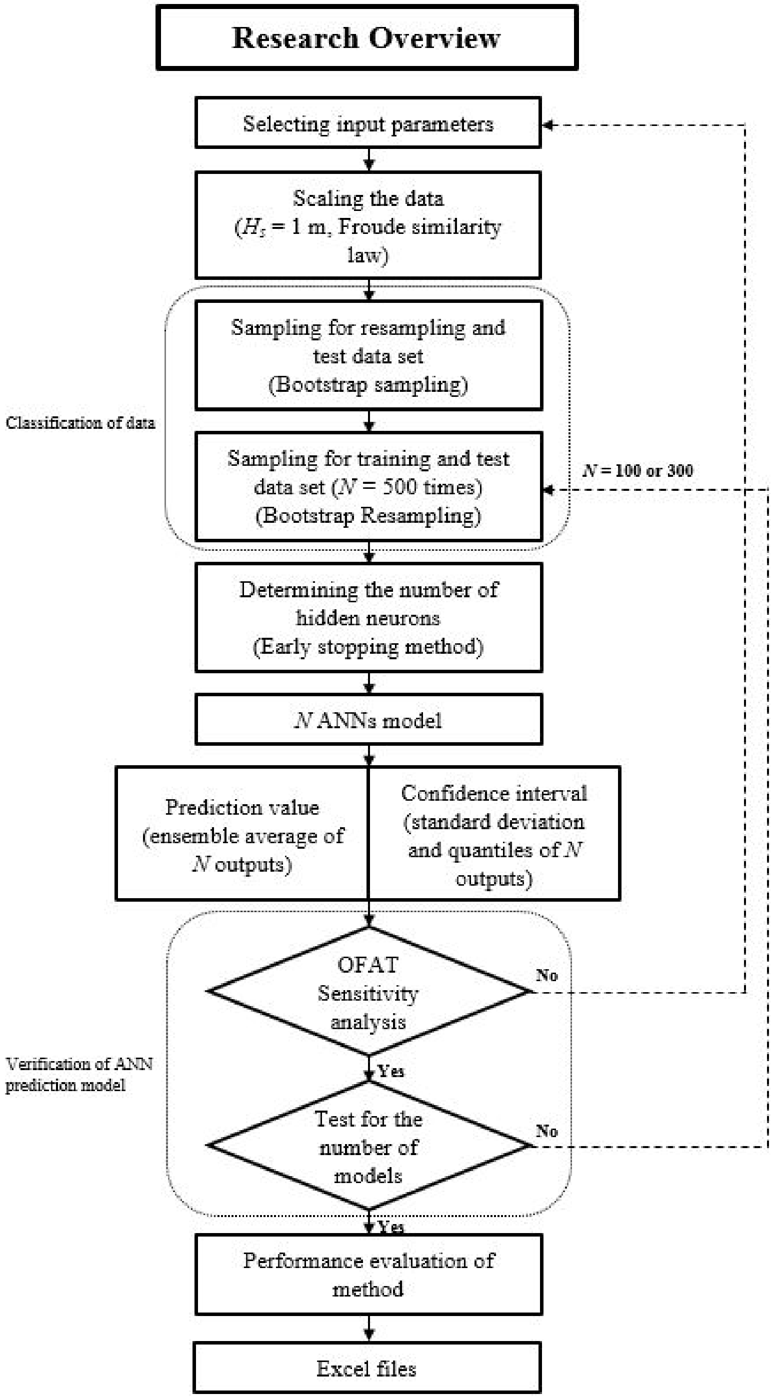


Figure 1.1 Research overview

CHAPTER 2. THEORETICAL BACKGROUNDS

2.1 Stability number

The minimum mass of individual armor units for a rubble mound foundation can be calculated by a formula of the Hudson type (Goda, 2009):

$$M = \frac{\rho_a}{N_s^{\,3}\Delta^3} H_s^{\,3} \qquad (2.1)$$

Therefore, the value of the stability number depends on not only the shape of the armor units, but also the wave conditions and mound dimensions. The nominal diameter of armor units is defined as follows:

$$D_n = \left(M / \rho_a\right)^{1/3} \qquad (2.2)$$

Thus, the nominal diameter of armor units is calculated by

$$D_n = \frac{H_s}{\Delta N_s} \qquad (2.3)$$

where Δ is the relative mass density of the stone or concrete in water and is calculated by

$$\Delta = \rho_a / \rho - 1 \qquad (2.4$$

In rock armor, the nominal diameter of armor units is the value of the average mass of graded rubble:

$$D_{n50} = \left(W_{50} / \rho_a\right)^{1/3} \qquad (2.5$$

2.2 Parameters

Rock Armor

(1) The significant wave height, H_s is incident wave height at the toe of structure, and the effect of wave height, H_s, and relative buoyant density, Δ, can be shown by $H_s / \Delta D_n$.

(2) The dimensionless damage level, *S*, is defined as follows:

$$S = A / D^2{}_{n50} \qquad (2.6$$

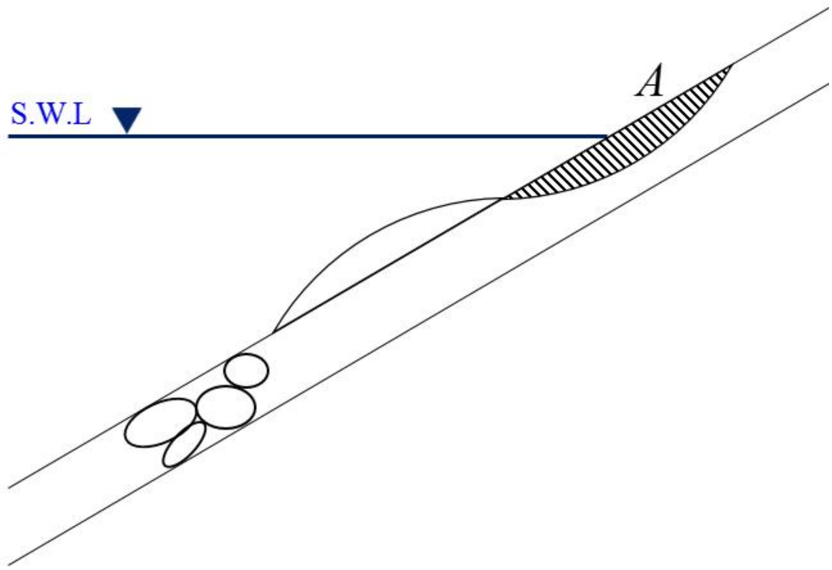


Figure 2.1 Eroded cross-sectional area of the profile

where A is the eroded cross-sectional area of the profile (see Figure 2.1). The dimensionless damage level physically means the number of damaged cubic stones. The criteria of "no damage" and "failure" were defined differently according to the slope angle, α . Before reaching the "no damage" criterion or the "failure" criterion, more damage occurs due to the presence of a larger amount of stones around the water level for lower slope angle, α .

(3) The effect of storm duration on damage was very critical, and the damage level, S, can be expressed as a square root of the number of waves, $\sqrt{N}$. Thus, the influence of the number of waves on stability can be expressed as $S/\sqrt{N}$.

(4) The surf similarity parameter, ξ_m, is related to slope angle, wave height, and wave period:

$$\xi_m = \frac{\tan\alpha}{\sqrt{s_{0m}}} \quad (2.7$$

and the wave steepness is given by the mean wave period:

$$s_{0m} = \frac{2\pi H_s}{g T_m^{\ 2}} \quad (2.8$$

The surf similarity parameter determined plunging waves, $\xi_m < 3$, and surging waves, $\xi_m > 3$, and the stability number according to surf similarity parameters in plunging waves and surging waves showed different tendencies. Minimum stability occurred at the transition from plunging to surging waves, called collapsing waves.

(5) Since the armor grading has little or no effect on the stability, the armor layer can be represented by the nominal size of the rock armor unit, D_{n50}.

(6) Van der Meer (1988a) used T_m because the influence of the spectral shape, *SS*, on stability was very small and might be negligible in that case. Therefore, T_m was used as the value for the period in this study, and *SS* and T_p were excluded from input parameters for rock armor.

(7) A permeability coefficient, P, was introduced to take into account the permeability of the structure, and P was defined by $P = 0.1$ for the impermeable core, 0.5 for the permeable core, and 0.6 for the homogeneous structure. In the both cases of plunging and surging waves, more permeable structure was more stable.

(8) The stability of the structure at a water depth of 0.2 m in front of the structure was higher than that of 0.8 m, because breaking of high waves occurred better on the foreshore with a water depth of 0.2 m.

Tetrapod

(1) The significant wave height, H_s is incident wave height at toe of the structure, and the effect of wave height, H_s, the nominal diameter, D_n, and relative buoyant density, Δ, can be represented by $H_s / \Delta D_n$, and the wave height can describe damage to the breakwater independent of the water depth.

(2) The dimensionless damage level, S, including displacement and settlement, does not take into account the porosity of the armor layer. Thus, another description of damage was suggested by Van der Meer (1988b) for breakwaters with an armor layer of concrete elements: the number of units displaced out of the armor layer, N_{od}. Van der Meer (1988b) described the influence of the storm duration for Tetrapod, longer storm duration causes more damage:

$$\sqrt{\frac{N_{od}}{\sqrt{N}}} = \text{constant} \tag{2.9}$$

(3) De Jong (1996) found that the influence of crest elevation, R_c, is also important. More wave overtopping occurred at low-crested structure compared with high-crested structures, causing more wave energy dissipation in all segments of the breakwater, therefore, low-crested structures are more stable.

(4) The packing density, ϕ, can be represented as follows (Shore Protection Manual, 1984):

$$\phi = N_a D_n^{\ 2} \tag{2.10}$$

where N_a is the number of armor units per m^2. The normal packing density for Tetrapod in the Shore Protection Manual (1984) is 1.04. Higher packing density results in higher stability.

(5) Van der Meer (1987b) showed that an increase in wave period will increase the stability of the Tetrapod for high-crested structures, however, Burger (1995) reported that the influence of the wave period would be different depending on the crest freeboard.

(6) Suh and Kang (2012) showed that the slope angel, α, increases the stability of the structure.

2.3 Artificial neural networks (ANN)

Artificial neural networks (ANN) model is a data-based regression model that analyzes nonlinear relations between input and output variables based on a large amount of data. ANN is a modeling technique that can be used alternatively when there is sufficient experimental data, but the physical relationship between variables is too complex or unknown, such as the stability number.

The structure of ANN is shown in Figure 2.2. Among the various models of neural networks, MLP (Multi-Layer Perceptron) is the most widely used model for data analysis. MLP is a feed-forward neural network consisting of an input layer, a hidden layer, and an output layer. Within the structure of ANN, there is an information processing units called a neuron, which is connected by different weights indicating the strength of the relationship between input and output variables. In a feed-forward neural network, the neurons are connected only in a forward direction, from one layer to the next layer, and the neurons of the same layer do not connect each other. The number of neurons in the input layer is equal to the number of input parameters and the number of neurons in the output layer is equal to the number of output parameters, which is one in this study. The number of hidden neurons in hidden layer should be determined to construct suitable ANN model. This will be explained in Section 3.3.2. In Figure 2.2, R and S represent the number of input variables (input neurons) and hidden neurons, respectively.

The propagation process is as follows. First, each input and output variables are normalized to transform the range of values to [min, max] = [-1, 1]. This normalization is done by:

$$x_i^n = \frac{x_i - (x_{\min} + x_{\max})/2}{(x_{\max} - x_{\min})/2} \quad (2.11$$

where x_i^n is the normalized x_i , $x_{\min}$ and $x_{\max}$ are the minimum and maximum of input variables, respectively. Second, this normalized input variable, x_i^n, is calculated as state, n_j^h, with the weights, $w_{j,i}^h$, and bias, b_j^h, as in Equation 2.12:

$$n_j^h = \sum_{i=1}^{R} w_{j,i}^h x_i^n + b_j^h \text{ for } j = 1,2,3,\dots,S \quad (2.12$$

and the state, n_j^h, is sent to the activation function, f^h, and the hidden neuron's output is calculated as a_j^h. The activation function in the neural network has several functions according to the input layer data. Sigmoid function was used as activation function in hidden layer in this study:

$$a_j^h = f^h\left(n_j^h\right) = \tanh\left(n_j^h\right) \quad (2.13$$

The same process was done in the output layer, the state of output layer was calculated as follows:

$$n_1^o = \sum_{j=1}^{S} w_{1,j}^o a_j^h + b_1^o \quad (2.14$$

and the linear function was used as activation function in output layer:

$$a_1^o = f^o\left(n_1^o\right) = n_1^o \quad (2.15$$

Training in artificial neural networks refers to the process of minimizing the error function defined by the RMSE (Root Mean-squared Error) while continuously modifying the weights. Training of MLP was performed using conjugate gradient algorithms. This algorithm leads to the fastest convergence and the process of search is performed along conjugate directions. The Powell-Beale restarts algorithm was used as an optimization algorithm, which is one of the conjugate gradient algorithms.

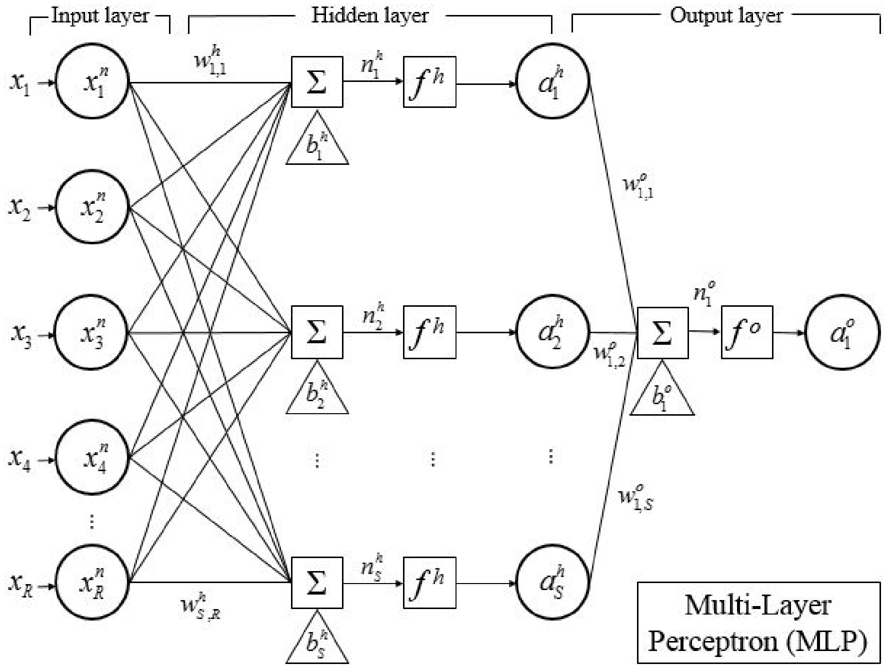


Figure 2.2 Structure of Artificial Neural Networks

CHAPTER 3. METHODOLOGY

3.1 Data

Rock Armor

The 579 experimental data set of Van der Meer (1988a) was used for the ANN model of rock armor. Table 3.1 shows the parameters for rock armor. H_s, T_m, ξ_m, S, N, h, P, and $\cot\alpha$ were selected as the input parameters.

Table 3.1 Parameters for rock armor

Parameter	Mean	Standard deviation	Minimum	Maximum
H_s (m)	0.147	0.145	0.046	1.18
T_m (s)	2.195	0.710	1.240	4.40
ξ_m	2.979	1.416	0.670	7.58
S	7.401	6.310	0.320	23.97
N	1981	1001	1000	3000
h (m)	0.925	0.8	0.2	5
P	0.266	0.204	0.1	0.6
$\cot\alpha$	3.054	1.166	2	6
N_s	2.095	0.633	0.79	4.38

Tetrapod

The experimental data sets of Van der Meer (1987b), De Jong (1996), and Suh and Kang (2012) were used for the ANN model of Tetrapod, and the total number of data is 286. Table 3.2 shows the parameters for Tetrapod. H_s, T_m, ξ_m, N_{od}, N, R_c, ϕ, and $\cot\alpha$ were selected as the input parameters.

Table 3.2 Parameters for Tetrapod

Parameter	Mean	Standard deviation	Minimum	Maximum
H_s (m)	0.165	0.041	0.087	0.266
T_m (s)	1.782	0.425	1.036	2.990
ξ_m	3.246	0.705	2.331	6.860
N_{od}	0.640	0.927	0	5.752
N	1642	953	427	3078
R_c (m)	0.209	0.088	-0.071	0.300
φ	1.012	0.030	0.880	1.020
$\cot\alpha$	1.547	0.192	1.333	2.000
N_s	2.760	0.569	1.660	5.025

3.2 Preparation of data

The input and output parameters in the database should be scaled using Froude similitude law because the final objectives was to suggest method applicable to all kinds of scale. Therefore, all the parameters in the data sets for rock armor and Tetrapod were scaled to $H_s = 1\,\mathrm{m}$. A prime indicates the scaled parameter to $H_s = 1\,\mathrm{m}$, and the process of scaling parameters are as follows.

First, the scaled length parameter such as H_s', R_c', and h' were obtained by multiplying $1/H_s$ to H_s, R_c, and h. Second, the scaled time parameter such as T_m' were derived by multiplying $1/H_s^{0.5}$ to T_m. Third, the dimensionless parameter were used without scaling. Since every value of H_s' in data sets became 1, this scaling method reduced the number of input parameter by one. Tables 3.3 and 3.4 show the scaled data distribution for rock armor and Tetrapod, respectively.

Table 3.3 Scaled parameters for rock armor

Parameter	Mean	Standard deviation	Minimum	Maximum
H_s'	1.0	0.0	1.0	1.0
T_m'	6.305	2.175	3.173	12.730
ξ_m'	2.979	1.416	0.670	7.580
S'	7.401	6.310	0.320	32.970
N'	1981	1001	1000	3000
h'	7.062	2.647	1.384	17.354
P'	0.266	0.204	0.1	0.6
$\cot\alpha'$	3.054	1.166	2	6
N_s'	2.095	0.633	0.79	4.38

Table 3.4 Scaled parameters for Tetrapod

Parameter	Mean	Standard deviation	Minimum	Maximum
H_s'	1.0	0.0	1.0	1.0
T_m'	4.503	1.292	3.161	9.227
ξ_m'	3.246	0.705	2.331	6.860
N_{od}'	0.640	0.927	0	5.752
N'	1642	953	427	3078
R_c'	1.404	0.714	-0.574	3.072
Φ'	1.012	0.030	0.880	1.020
$\cot\alpha'$	1.547	0.192	1.333	2.000
N_s'	2.760	0.569	1.660	5.025

3.3 ANN configuration

3.3.1 Classification of data

The training process is adjusting the ANN model's configuration, which is adjusting the weight and bias of ANN model using training set. The process of sampling for training, validation, and test data sets from the original data set must precede in the training process, and the sampling process should be performed in a reasonable way. Since sampling directly affects the performance of the ANN model, the biased training data set creates a wrong model and the biased test data set may leads to weird evaluation of model performance. Therefore, reasonable sampling technique should be taken. In this study, bootstrap sampling (Efron, 1979) was used for the sampling method. A Bootstrap sampling is a random selection of N data out of the N original data with replacement. The probability that a data in the original data set is not included in the sampling is $(1-1/N)^N$, which is $1/e$ for large N.

The entire data set were divided into training, validation, and test data set. To accurately evaluate the performance of the model, test data set were not used in training and only used in test. In this study, two bootstrap sampling procedures were performed. The first is to classify the entire data set into the candidates for training and validation data (resampling data set) and the test data set. The data selected once or more from the entire data set is considered as the element of resampling data set, and the data which has not been extracted is included in test data set. Since first sampling is only for classifying data, not for constructing model, the data selected several times in the resampling data set are regarded as a single data.

Kolmogorov-Smirnov (KS) test is carried out to verify that the test data set represents the entire data set. The second sampling is to classify the resampling data set into training and validation data set. The data selected once or more in the sampling process is considered as the training data, and the data set which has not been extracted is used as the validation set. In the second bootstrap sampling procedure, the number of times data is chosen is reflected in the error function RMSE as a weight factor, WF_i (see Equation 3.2).

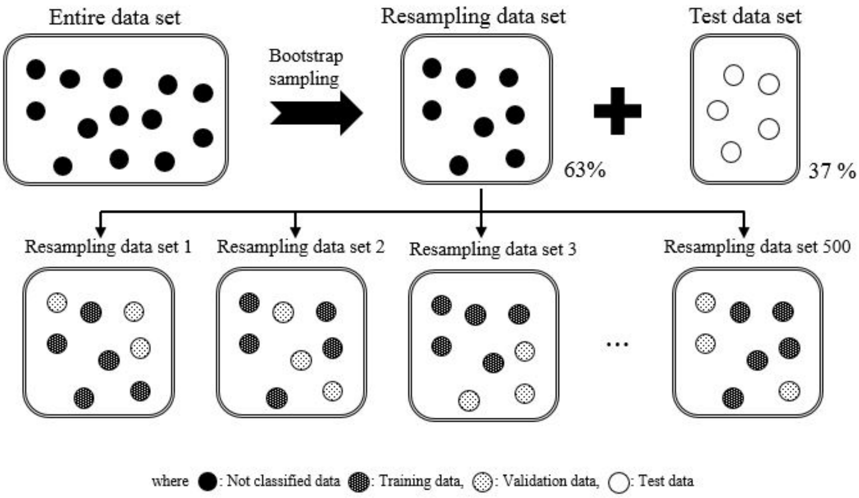


Figure 3.1 Classification of data

3.3.2 The number of Hidden neurons

It is important to determine the number of hidden neurons in hidden layer, because the number of hidden neurons might have a great impact on the performance and complexity of ANN model. When training the ANN model, weights and bias are determined so that the error function set based on the training data is minimized. At this point, it is important to prevent the overfitting that the ANN model is biased to the training data set. In this study, early stopping (Heskes, 1997) was used to prevent overfitting by using the validation data set. This method is a way to stop training before the error function of the ANN model for the validation data set increases, i.e. to determine the number of hidden neurons. The values of R (correlation coefficient) and IOA (index of agreement) were considered to determine the number of hidden neurons as well as the existing error function, RMSE (Root Mean-Square Error).

The value of R is calculated by

$$R = \frac{\sum_{i=1}^{n}\left(N_{s,pred}^{i} - \bar{N}_{s,pred}\right)\left(N_{s,obs}^{i} - \bar{N}_{s,obs}\right)}{\sqrt{\sum_{i=1}^{n}\left(N_{s,pred}^{i} - \bar{N}_{s,pred}\right)^{2}}\sqrt{\sum_{i=1}^{n}\left(N_{s,obs}^{i} - \bar{N}_{s,obs}\right)^{2}}} \quad (3.$$

where $N_{s,pred}^{i}$ is the predicted stability number for *i*-th input in training, validation or test data set, $\bar{N}_{s.pred}$ is average of predicted stability number, $N_{s,obs}^{i}$ is *i*-th observed stability number in training, validation or test data set, and $\bar{N}_{s.obs}$ is average of observed stability number.

The value of RMSE is calculated by

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} WF^{i}\left(N^{i}_{s,pred} - N^{i}_{s,obs}\right)^{2}} \tag{3.2}$$

where n is the number of elements in training, validation or test data and WF^{i} is the number of times i-th data was chosen

The value of IOA is calculated as follows (Willmott, 1981).

$$I_a = 1 - \frac{\sum_{i=1}^{n}\left(N^{i}_{s,pred} - N^{i}_{s,obs}\right)^{2}}{\sum_{i=1}^{n}\left[\left|N^{i}_{s,pred} - \bar{N}_{s.obs}\right| + \left|N^{i}_{s,obs} - \bar{N}_{s.obs}\right|\right]^{2}} \tag{3.3}$$

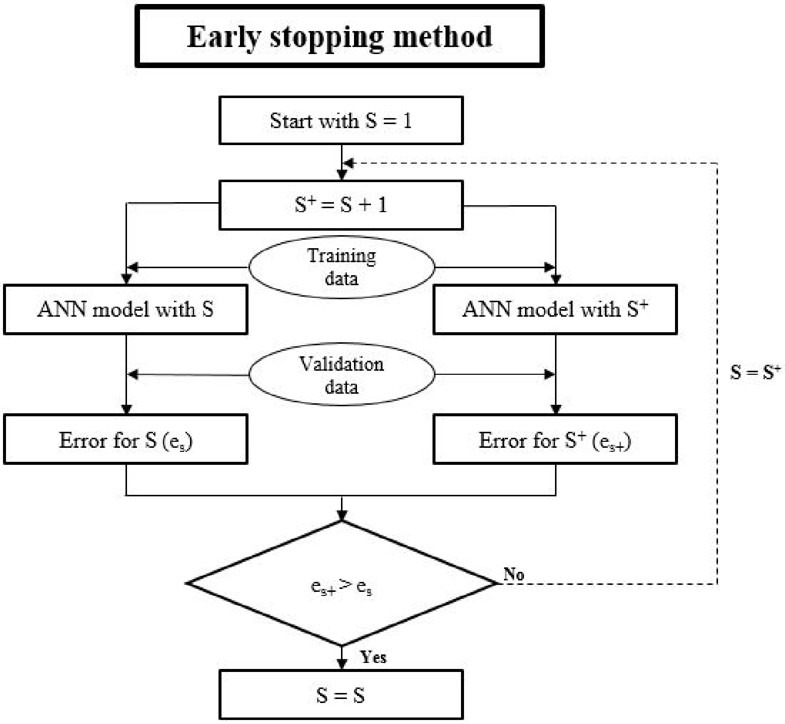


Figure 3.2 Flow chart of early stopping method

3.4 Uncertainty assessment

3.4.1 Bootstrap resampling

There are several reasons why reliability should be taken into account in developing ANN model. First, since the neural network model is a kind of 'black box', information regarding its reliability or uncertainty should be provided. Second, when calibrating the ANN model, namely, adjusting the weight and bias of the ANN model, a training data set is used. However, since this training data set is a sample, not population, it is impossible to cover all cases. Therefore, ANN can be seen as a kind of interpolation or extrapolation, and thus reliability must be taken into account (Van Gent and Van den Boogaard, 1998). In addition, since the stability number is a function of various random variables, the stability number is also a random variable. The fact that the stability number is a random variable may result in inconsistency of the stability number when similar input data are used. Hence, the information of reliability such as confidence interval is necessary to cover the inconsistency of the output.

Van den Boogarrad et al. (2000) suggested resampling techniques to construct the training data set and validation data set. The process of dividing the original data set into training and validation data set is repeated many times, and ANN model is also generated for each training and validation data set. The techniques give several advantages in developing ANN model. First, when constructing ANN model with a single training data set, the model can be biased with the specific training data set. The likelihood for such a model to be biased disappears because different training and validation data sets are used to adjust ANN model's configuration. In other words, the representativeness of training and validation data set can be guaranteed. Also, a large number of ANN models make the same number of prediction values, outputs, and the set of outputs can give the information or assessment of the reliability in prediction.

3.4.2 Prediction value and confidence interval

Van Gent (2007) suggested the estimation method and how to provide the reliability (or uncertainty) of the ANN model with regard to the accuracy of prediction by using bootstrap resampling. 500 ANN models can be constructed by the 500 resample of the data set, and the ensemble average of 500 ANN model outputs is prediction value in this study:

$$N^i_{s,pred} = \frac{1}{500}\sum_{j=1}^{500}\left(N^i_{s,pred}\right)_j \quad (3.4$$

where $N^i_{s,pred}$ is the prediction value for i-th data and $\left(N^i_{s,pred}\right)_j$ is the output of j-th ANN model for i-th data.

There are two methods to provide confidence interval: one used the standard deviation of 500 ANN model outputs, assuming the 500 ANN model outputs satisfy a Gaussian distribution, and the other used the quantiles. The variance of 500 ANN model outputs is calculated by

$$s_{N_s}^{i\ 2} = \frac{1}{500-1}\sum_{j=1}^{500}\left[\left(N_{s,pred}^{i}\right)_j - N_{s,pred}^{i}\right] \tag{3.5}$$

First, $100\cdot(1-\alpha)\%$ confidence interval using the standard deviation is calculated by

$$\left[N_{s,pred}^{i} - c_{confidence}\cdot s_{N_s}^{i}, N_{s,pred}^{i} + c_{confidence}\cdot s_{N_s}^{i}\right] \tag{3.6}$$

where $N_{s,pred}^{i}$ is the ensemble average of 500 ANN model output for *i*-th data, $s_{N_s}^{i}$ is the standard deviation of 500 ANN model outputs for *i*-th data, and the value of $c_{confidence}$ is selected such that for no more than 100α % of all $500\cdot$(The number of data) ANN model outputs satisfying following condition:

$$\left|\left(N_{s,pred}^{i}\right)_j - N_{s,pred}^{i}\right| \geq c_{confidence} s_{N_s}^{i} \tag{3.7}$$

Second, $100\cdot(1-\alpha)\%$ confidence interval using the quantiles is calculated by

$$\left[\left(N_{s,pred}^{i}\right)_{\alpha/2}, \left(N_{s,pred}^{i}\right)_{1-\alpha/2}\right] \tag{3.8}$$

where $\left(N^i_{s,pred}\right)_{a/2}$ and $\left(N^i_{s,pred}\right)_{(1-a)/2}$ are $100\cdot(\alpha/2)$ % and $100\cdot(1-\alpha/2)$ % quantiles for i-th data, respectively.

The fact that the standard deviation and quantiles of each data were used results in the uneven graph. For smoother graph, $100\cdot(1-\alpha)\%$ average confidence intervals were also calculated as follows. $100\cdot(1-\alpha)\%$ average confidence interval using standard deviation is given by

$$\left[N^i_{s,pred} - c_{confidence}\cdot\overline{s}_{N_s}, N^i_{s,pred} + c_{confidence}\cdot\overline{s}_{N_s}\right] \tag{3.9}$$

where $\overline{s}_{N_s}$ is the average of standard deviation of 500 ANN model outputs, and calculated by

$$\overline{s}_{N_s} = \frac{1}{n}\sum_{i=1}^{n} s^i_{N_s} \tag{3.10}$$

where n is the number of elements in the entire data set, and $s^i_{N_s}$ is standard deviation of 500 ANN model outputs for i-th data.

$100\cdot(1-\alpha)\%$ average confidence interval using quantiles is given by

$$\left[N_{s,pred}^{i} - \bar{N}_{s,pred} + \left(\bar{N}_{s,pred}\right)_{\alpha/2}, N_{s,pred}^{i} + \bar{N}_{s,pred} - \left(\bar{N}_{s,pred}\right)_{(1-\alpha)/2} \right] \quad (3.1$$

where $\left(\bar{N}_{s,pred}\right)_{\alpha/2}$ and $\left(\bar{N}_{s,pred}\right)_{(1-\alpha)/2}$ are the average of $100\cdot(\alpha/2)\%$ and $100\cdot(1-\alpha/2)\%$ quantiles, respectively.

3.5 Verification of ANN prediction method

To verify that ANN prediction method were reasonably constructed, the following two steps were performed. First, sensitivity analysis was performed with the generated models to confirm that the input parameters for the ANN model are appropriately selected. In this study, the One-Factor-At-a-Time (OFAT) method was used to figure out the influence of the input parameters on the stability number. The OFAT methods are a way to confirm the results in output when one parameter changes and the remaining variables are fixed at the baseline values. For rock armor, the mean value was used as the baseline value for H_s, T_m, S, and ξ_m. On the other hand, the most frequent value was used for N, P, h, and cot α. For Tetrapod, the mean value was used as the baseline value for H_s, T_m, ξ_m, N_{od}, N, and R_c. On the other hand, the most frequent value was used for ϕ and cotα. Second, to make sure that the number of models was properly selected, the number of models increased from 100 to 500 and the results were compared.

3.6 Performance evaluation of method

Two data sets were used in performance evaluation of method. First, in order to compare the previous methods and present method, the entire data was used. In general, when developing an empirical formula, all the experimental data set used in development of the empirical formula was used in performance evaluation. Therefore, it is reasonable that present method was evaluated with the entire data to compare the previous methods and present method. Second, the test data set was used since the test data set was never used in constructing model. Test data set was separated from the training and validation test data sets, therefore the performance evaluation with the test data set is more objectively.

Not only three statistical parameters explained in Section 3.3.2, R, RMSE, and IOA, but also NSE (Nash-Sutcliffe efficiency) and MBE (Mean Bias Error) were calculated and compared to evaluate the performance of methods.

(1) R ranges from -1 to 1 and a value of 1 and -1 indicate perfect linear relationships between two variables (see Equation 3.1). On the other hand, a value of 0 implies no linear relationships between two variables. However, the values are often not related with the differences between two variables.

(2) RMSE ranges from 0 to ∞, with 0 corresponding to the ideal model (see Equation 3.2). Even if RMSE shows the differences between two variables, it cannot show the linearity between variables.

(3) NSE is the normalized error by the variance of the observed values, which ranges from $-\infty$ to 1, with 1 corresponding to the perfect model (see Equation

3.12). NSE is highly sensitive to extreme values due to the squared differences. The value of NSE is calculated as follows (Nash and Sutcliffe, 1970).

$$NSE = 1 - \frac{\sum_{i=1}^{n}\left(N_{s,pred}^{i} - N_{s,obs}^{i}\right)^2}{\sum_{i=1}^{n}\left(N_{s,obs}^{i} - \bar{N}_{s.obs}\right)^2} \tag{3.12}$$

where $N_{s,pred}^{i}$ is the prediction of ANN model for i-th input in entire or test data set, $N_{s,obs}^{i}$ is i-th observed stability number in entire or test data set, and $\bar{N}_{s.obs}$ is the average of observed stability number.

(4) IOA represents the ratio of mean squared error and the potential error, which range from 0 to 1, with 1 corresponding to the ideal model (see Equation 3.3). IOA is sensitive to differences between two variables as well as to proportionality. IOA is also highly sensitive to extreme values.

(5) MBE represents the average of all the errors, which refer how biased model is and whether the model under-predict or over-predict: Positive value and negative value indicate over-predicted and under-predicted, respectively:

$$MBE = \frac{1}{n}\sum_{i=1}^{n}\left(N_{s,pred}^{i} - N_{s,obs}^{i}\right) \tag{3.13}$$

where n is the number of entire or test data set, $N^i_{s,pred}$ is the prediction of ANN model for i-th input in entire or test data set, and $N^i_{s,obs}$ is i-th observed stability number in entire or test data set.

CHAPTER 4. RESULTS

4.1 Classification of data

For rock armor, total 579 data were separated into 364 data for the candidates for training and validation data set (resampling data set) and 215 data for the test data set. For Tetrapod, total 286 data were separated into 180 data for the candidates for training and validation data set (resampling data set) and 106 data for the test data set. According to Kolmogorov-Smirnov (KS) test of which significance level is assumed as 0.05, the null hypothesis is that the entire, resampling, test data sets come from the same distribution and is not rejected for all parameters. Figures 4.1-4.9 and 4.10-4.18 compare probability mass function of all parameters in the entire, resampling, and test data sets for rock armor and Tetrapod, respectively.

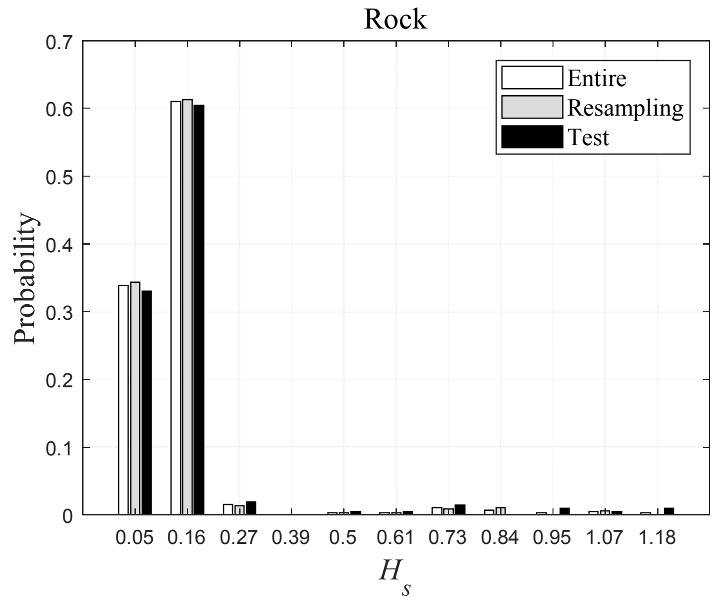


Figure 4.1 PMF of H_s for rock armor

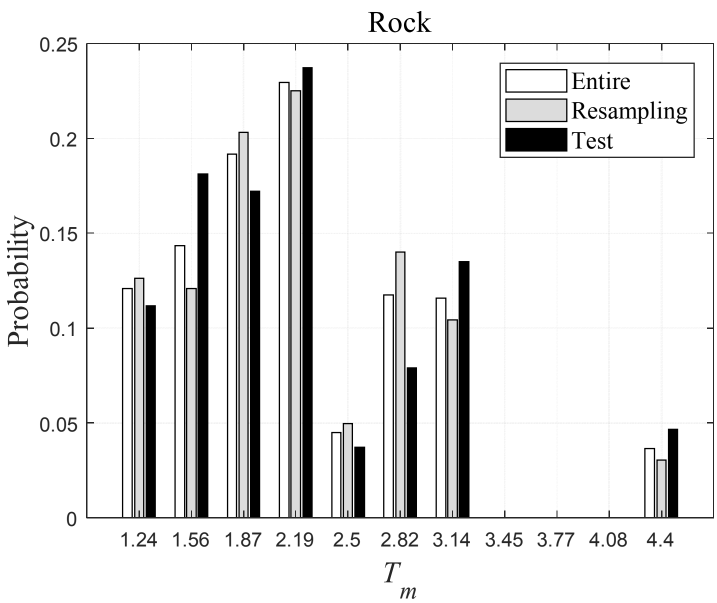


Figure 4.2 PMF of T_m for rock armor

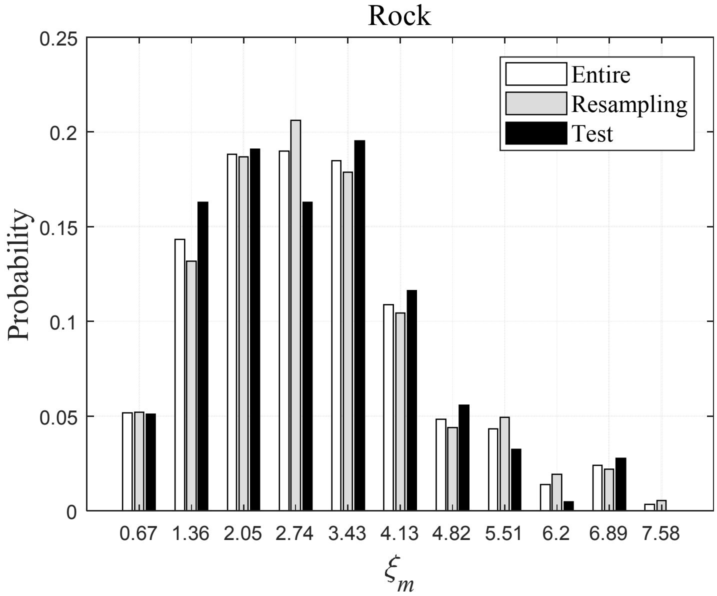


Figure 4.3 PMF of ξ_m for rock armor

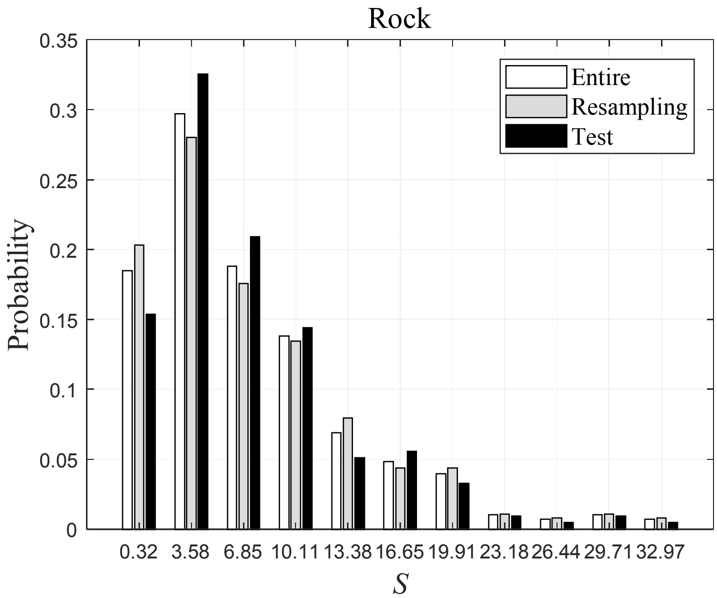


Figure 4.4 PMF of S for rock armor

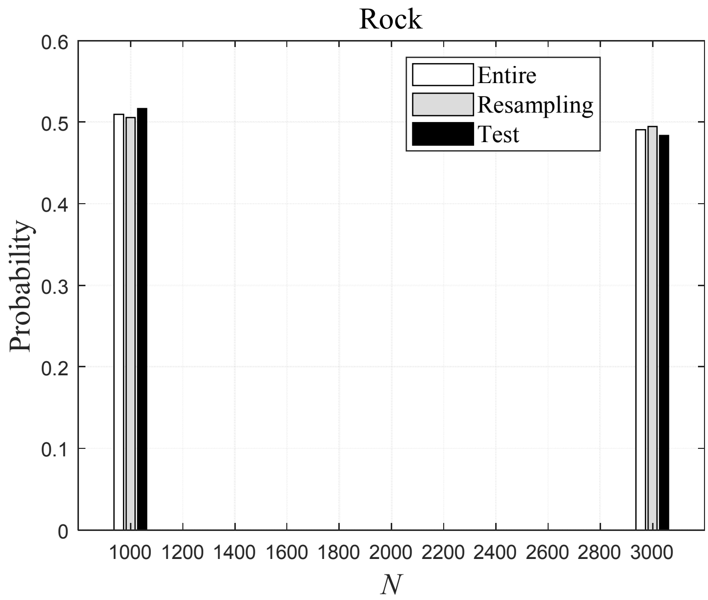


Figure 4.5 PMF of *N* for rock armor

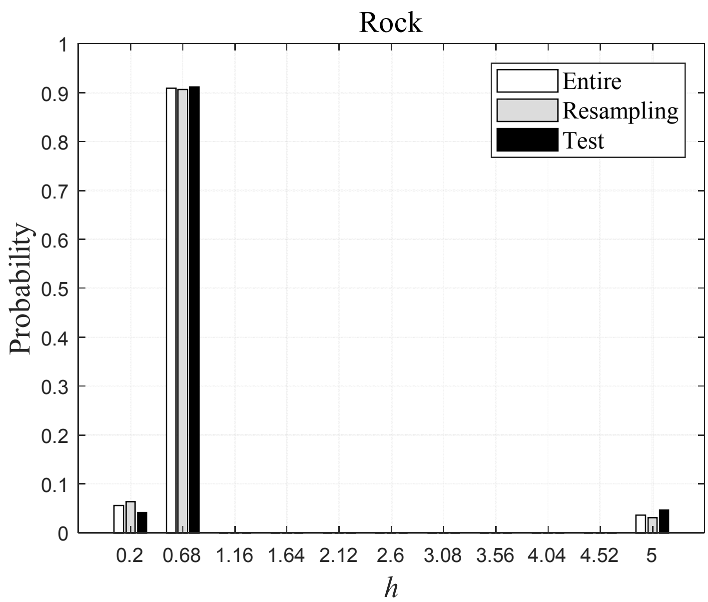


Figure 4.6 PMF of *h* for rock armor

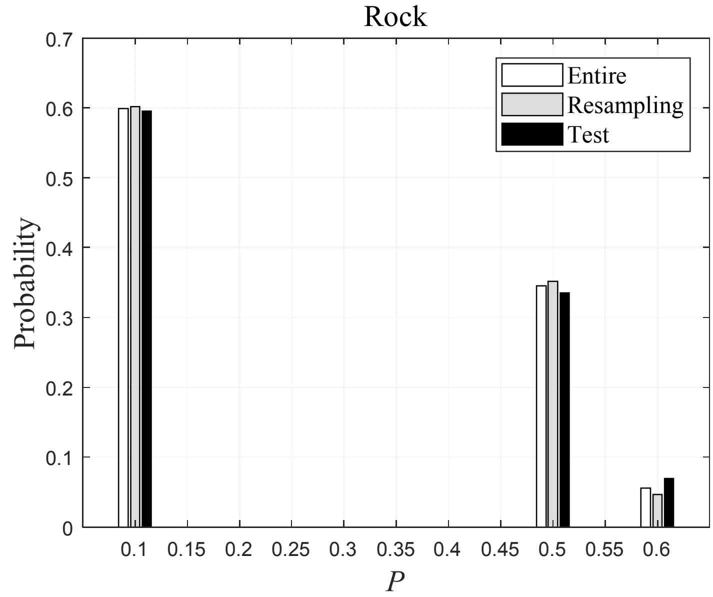


Figure 5.7 PMF of *P* for rock armor

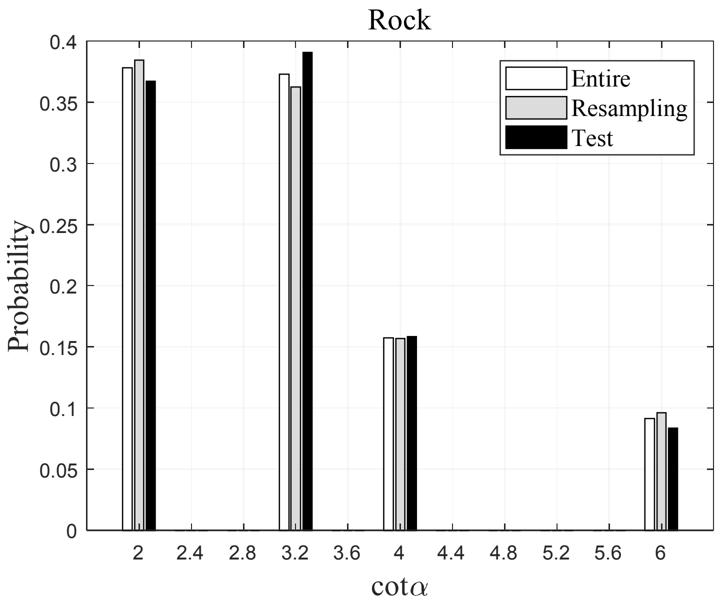


Figure 6.8 PMF of cot*α* for rock armor

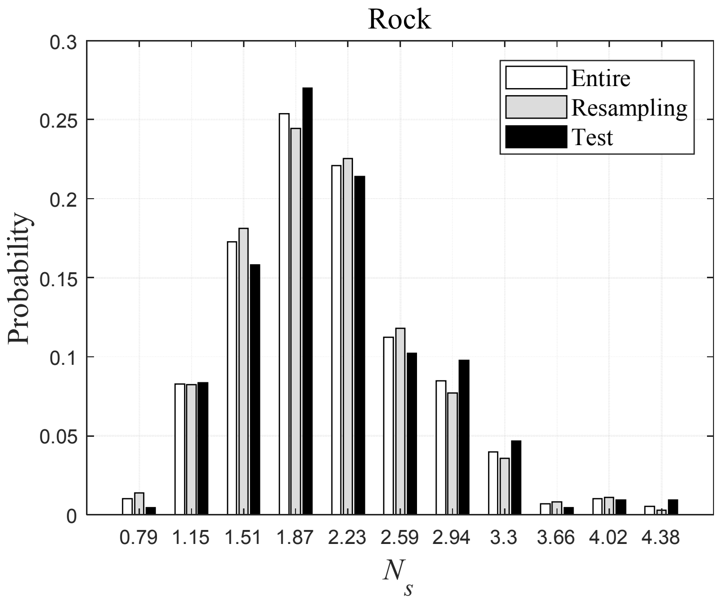


Figure 7.9 PMF of N_s for rock armor

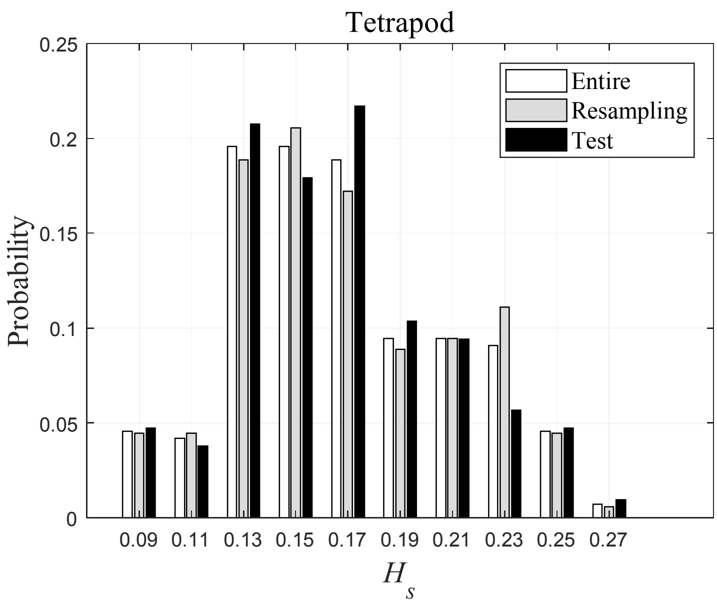


Figure 8.10 PMF of H_s for Tetrapod

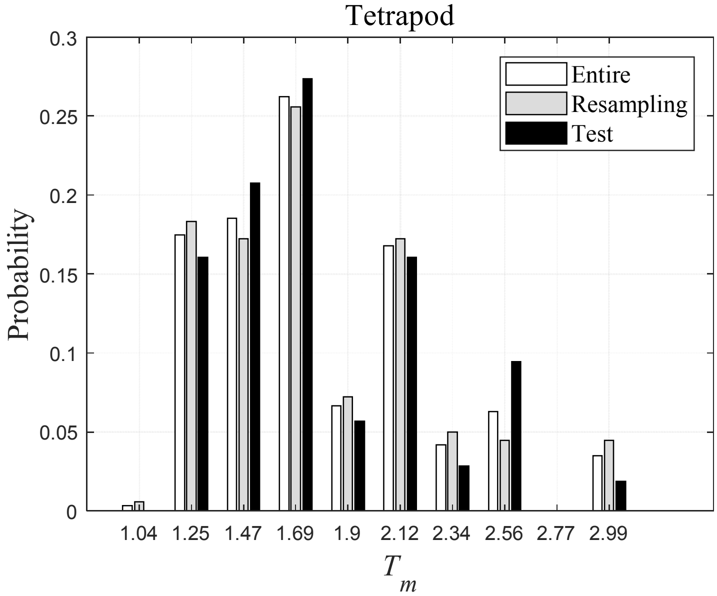


Figure 4.11 PMF of T_m for Tetrapod

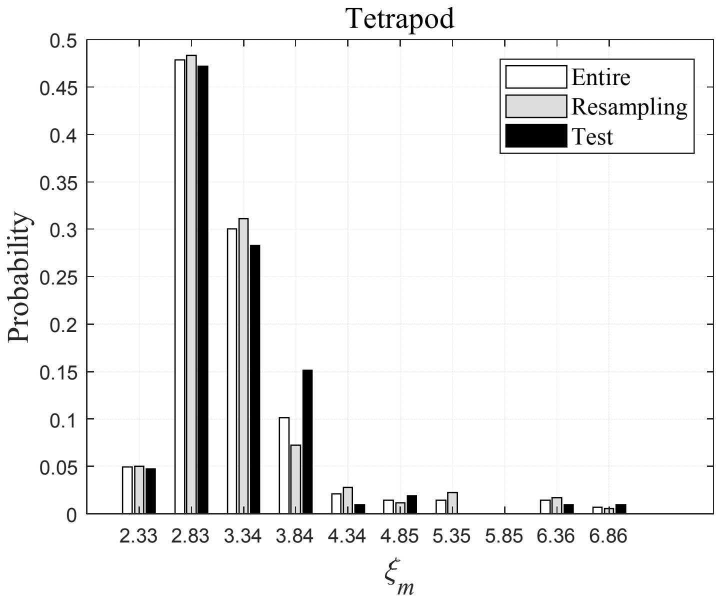


Figure 4.12 PMF of ξ_m for Tetrapod

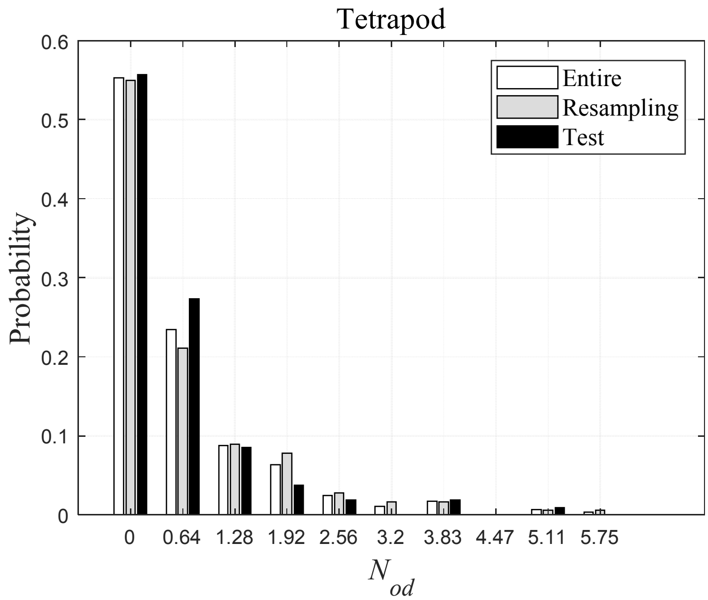


Figure 4.13 PMF of N_{od} for Tetrapod

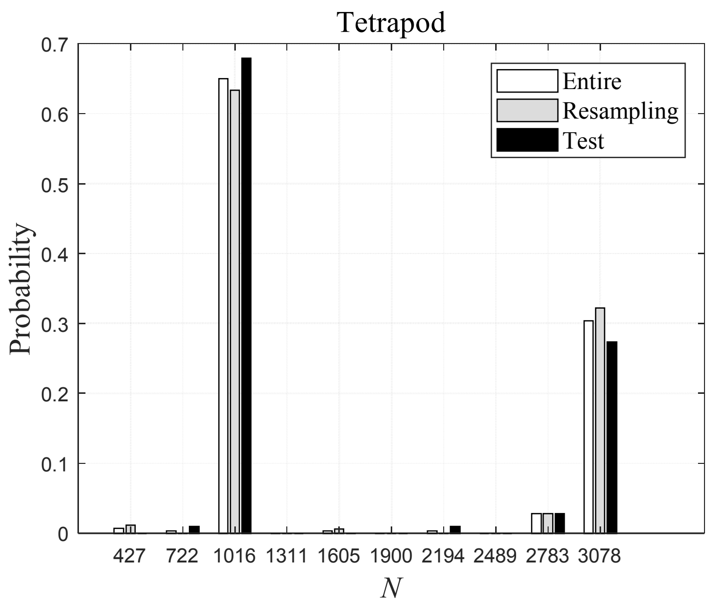


Figure 4.14 PMF of N for Tetrapod

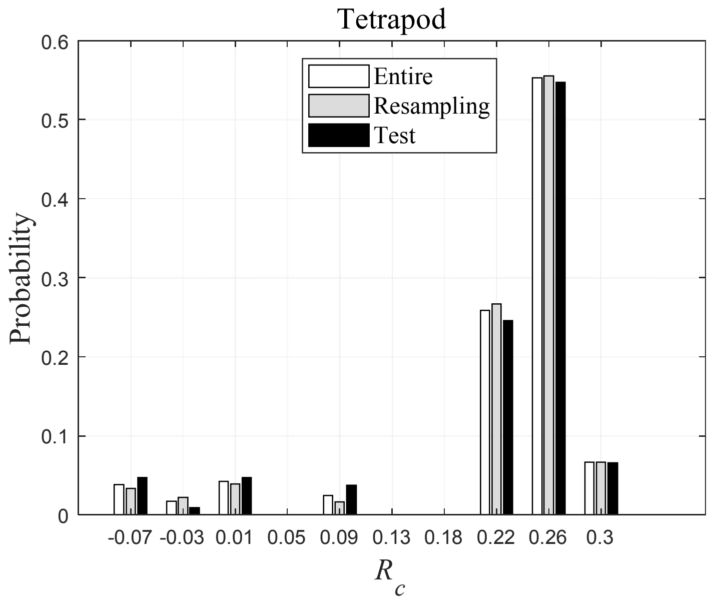


Figure 4.15 PMF of R_c for Tetrapod

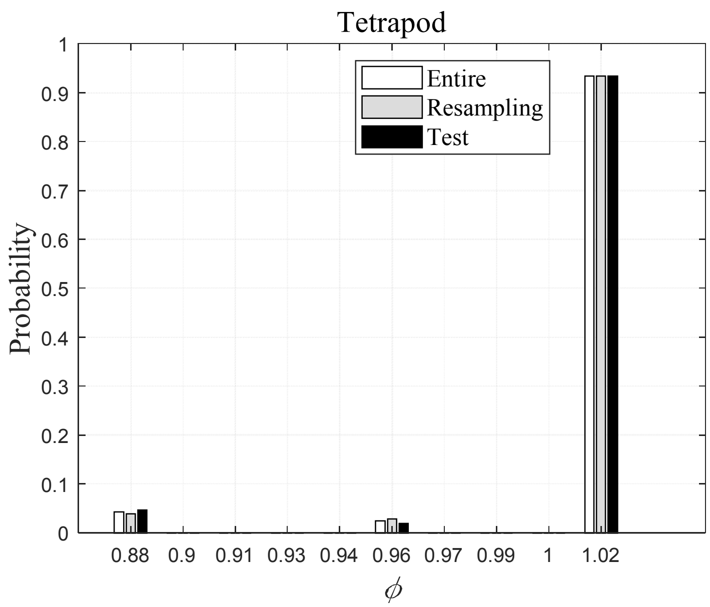


Figure 4.16 PMF of φ for Tetrapod

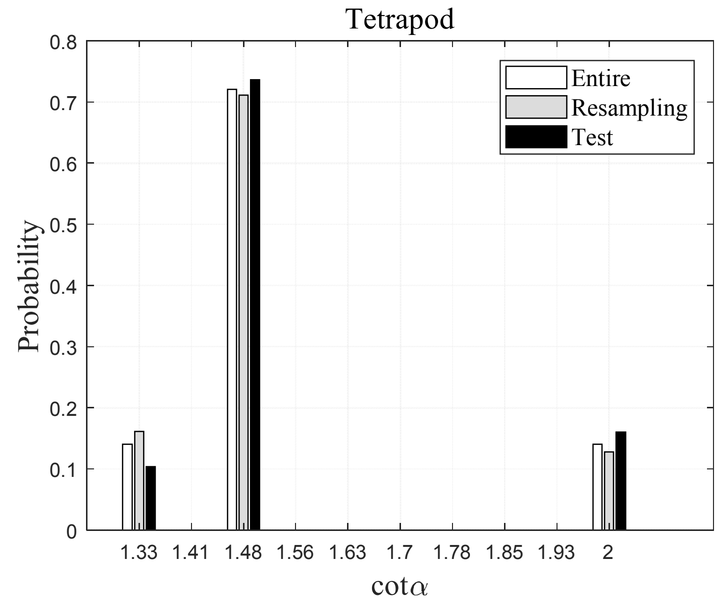


Figure 4.17 PMF of cot*α* for Tetrapod

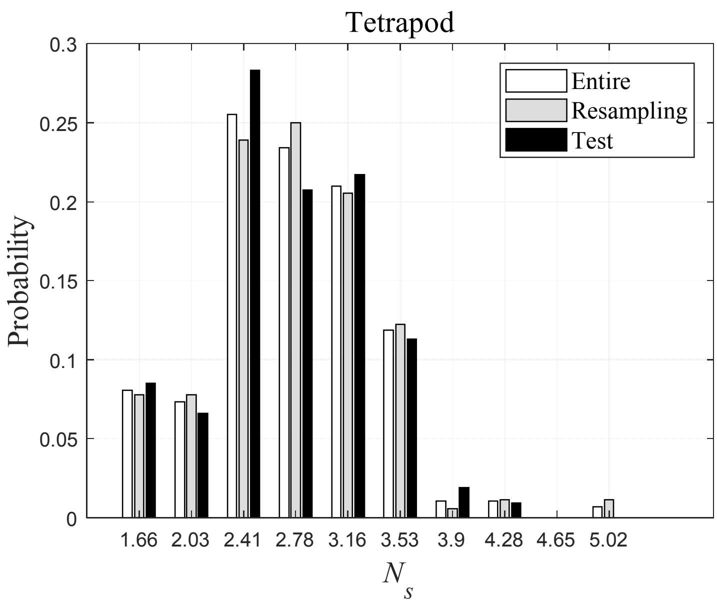


Figure 9.18 PMF of N_s for Tetrapod

4.2 The number of hidden neurons

The values of R, IOA, and RMSE were calculated for 500 ANN models as the number of hidden neurons increased. Then, the average values of R, IOA, and RMSE for the 500 ANN models were plotted against the number of hidden neurons.

Rock Armor

Figures 4.19, 4.20, and 4.21 show the performance of the ANN model for different number of hidden neurons. It can be found that the performance of the ANN model on the validation data set starts to worsen when the number of hidden neurons is 7. Thus, the optimal number of hidden neurons for rock armor was determined as 6.

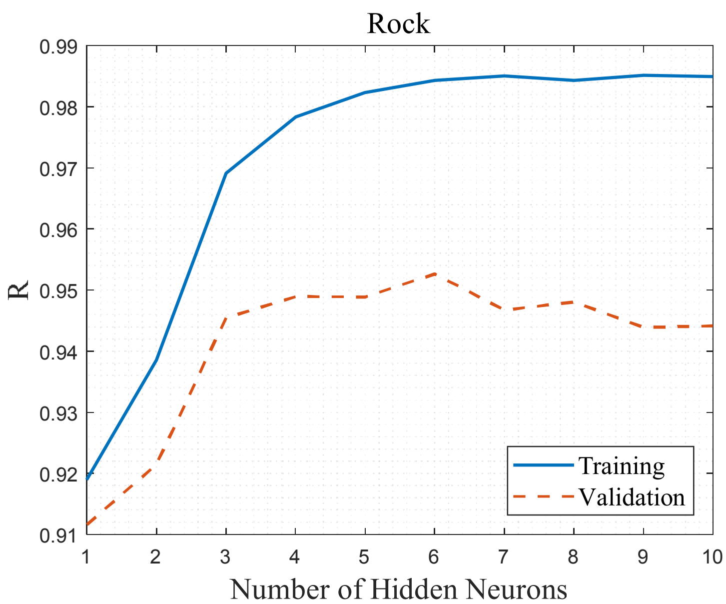


Figure 4.19 Average values of R for the ANN models for rock armor

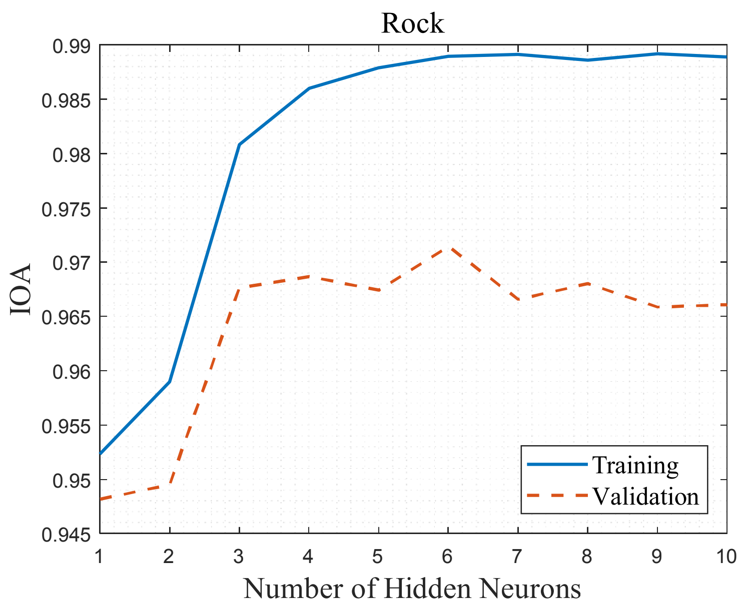


Figure 4.20 Average values of IOA for the ANN models for rock armor

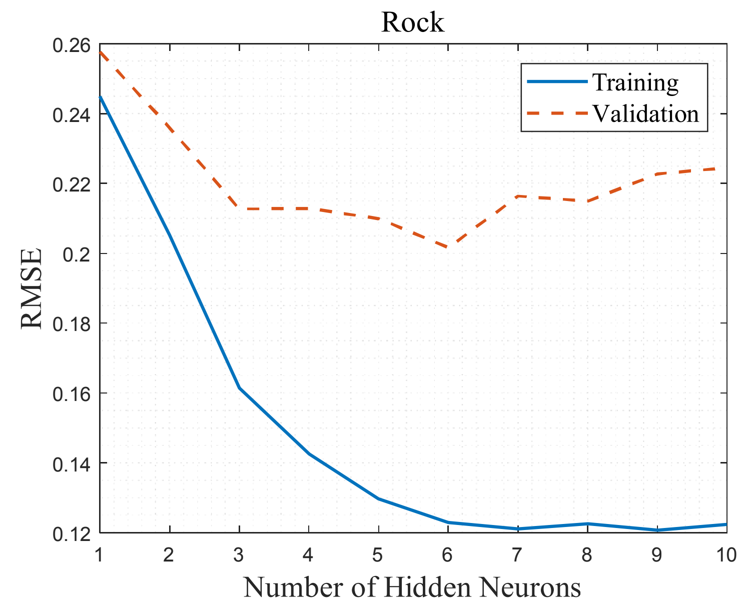


Figure 4.21 Average values of RMSE for the ANN models for rock armor

Tetrapod

Figures 4.22, 4.23, and 4.24 show the performance of the ANN model for different number of hidden neurons. It can be found that the performance of the ANN model on the validation data set starts to worsen when the number of hidden neurons is 4. Thus, the optimal number of hidden neurons for Tetrapod was determined as 3.

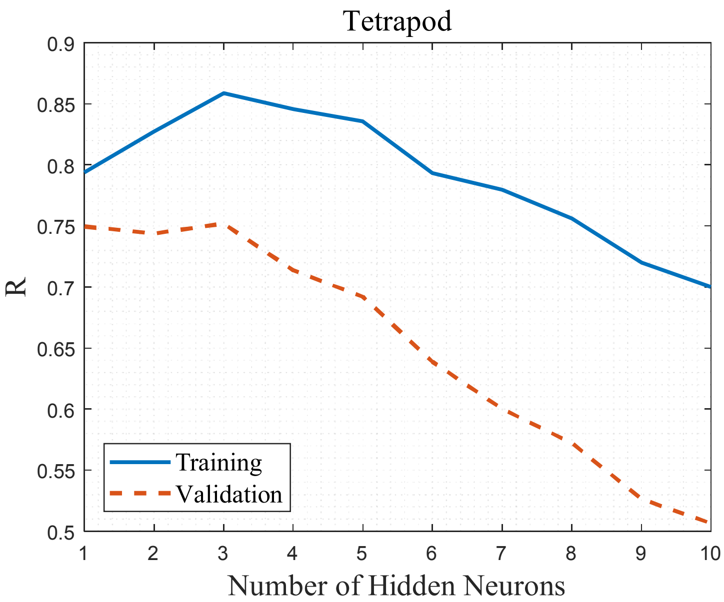


Figure 4.22 Average values of R for the ANN models for Tetrapod

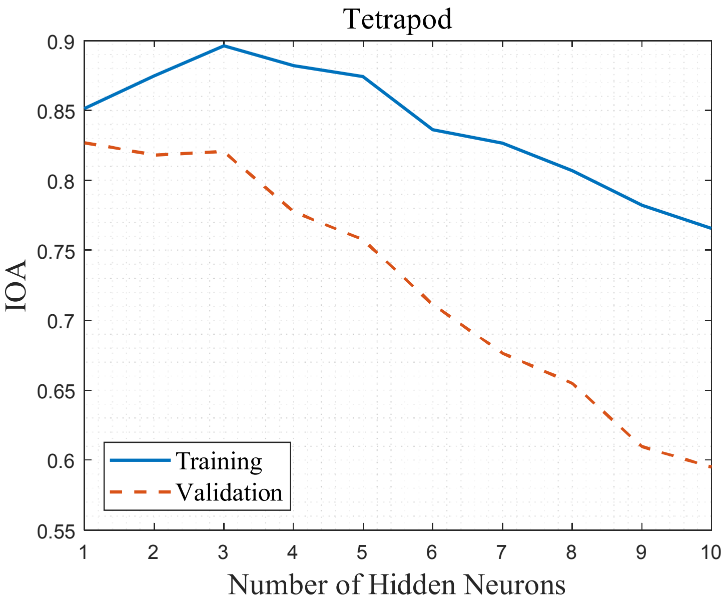


Figure 4.23 Average values of IOA for the ANN models for Tetrapod

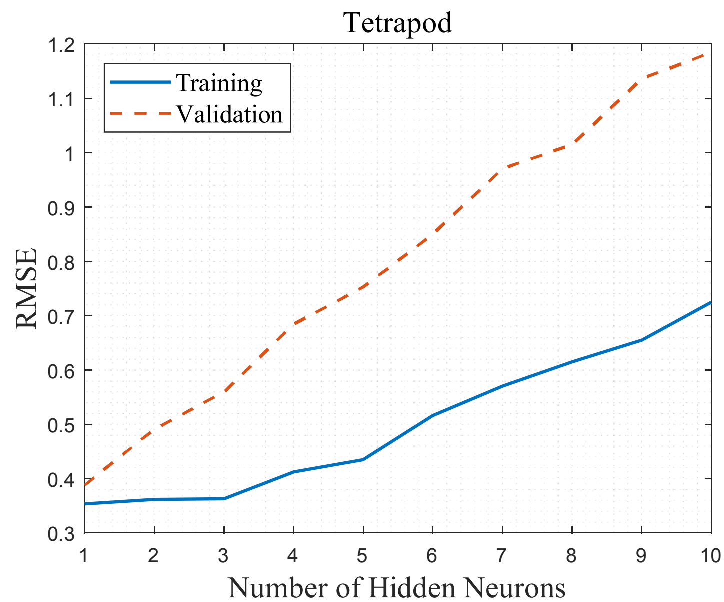


Figure 4.24 Average values of RMSE for the ANN models for Tetrapod

4.3 Prediction value and confidence interval

As stated in Section 3.4.2, the ensemble average of 500 ANN model outputs was used as prediction value for the stability number, and the confidence interval was calculated by using standard deviation and quantiles of 500 ANN model outputs.

Rock Armor

Figures 4.25, 4.26, and 4.27 show the prediction value and 90% confidence using standard deviation and quantiles, respectively. And Figures 4.28 and 4.29 show average confidence interval using standard deviation and quantiles, respectively. Since confidence level is 90%, 1.56 was selected as the value of $c_{confidence}$ and 5% and 95% quantiles were used. Also, 90% average confidence intervals using standard deviation and quantiles were calculated as follows, respectively.

$$\left[N_{s,pred}^{i}-0.219, N_{s,pred}^{i}+0.219\right] \quad (4.1$$

and

$$\left[N_{s,pred}^{i}-0.207, N_{s,pred}^{i}+0.225\right] \quad (4.2$$

The blue circle, the green triangle, and the red inverted triangle indicate the prediction value, the upper lower confidence interval, respectively.

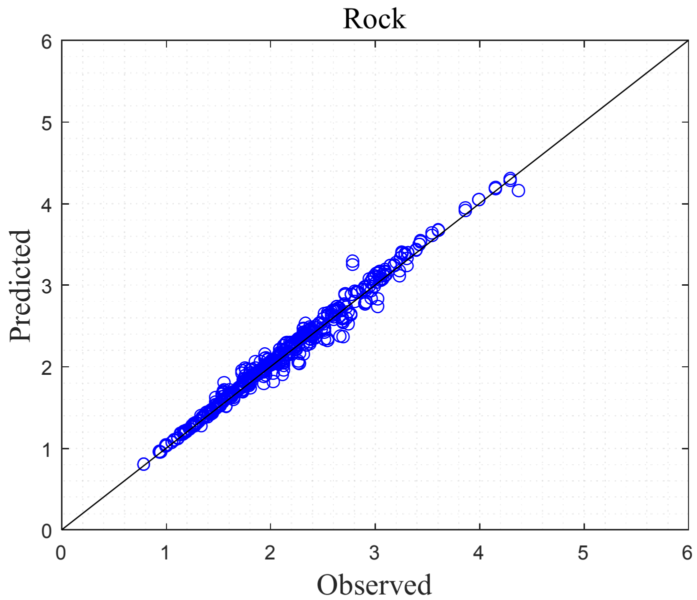


Figure 4.25 Predicted values versus observed values for rock armor

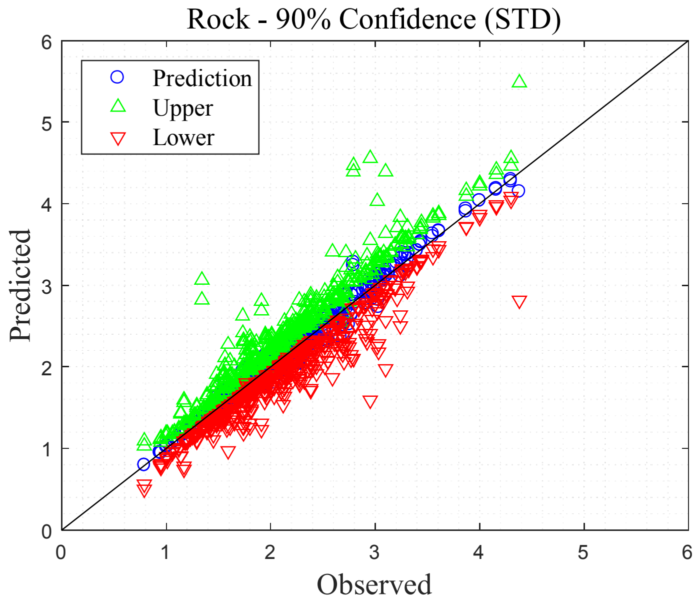


Figure 4.26 90% confidence interval using standard deviation for rock armor

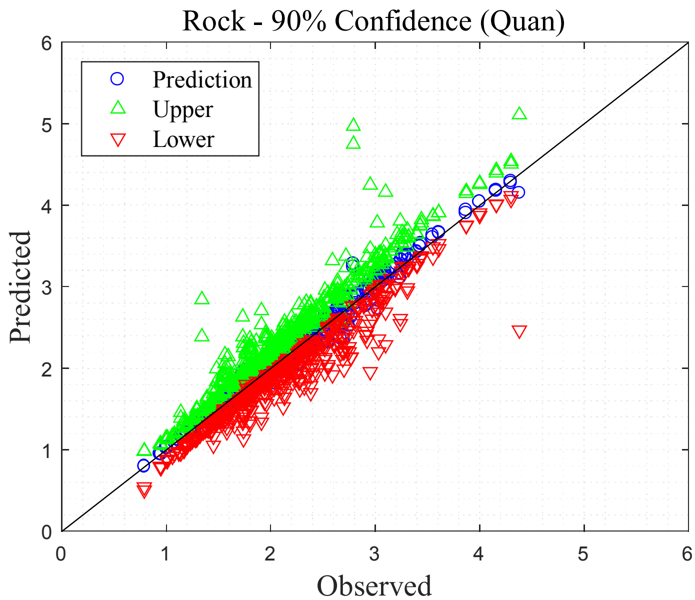


Figure 4.27 90% confidence interval using quantiles for rock armor

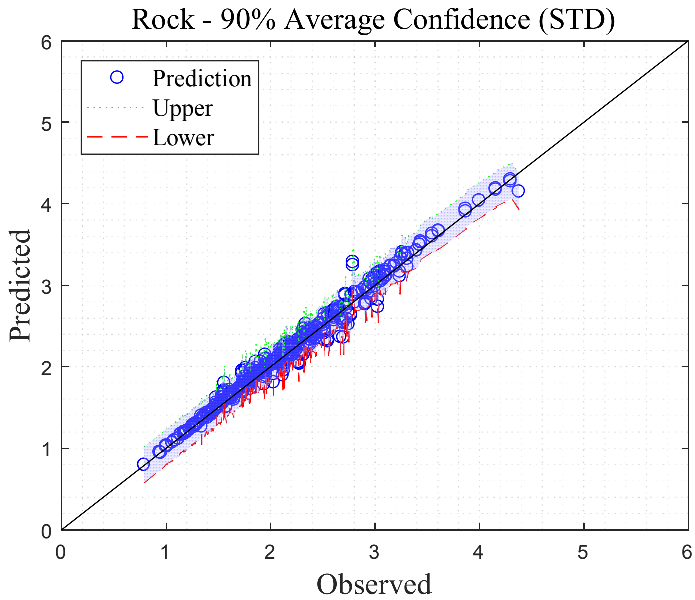


Figure 4.28 90% average confidence interval using standard deviation for rock armor

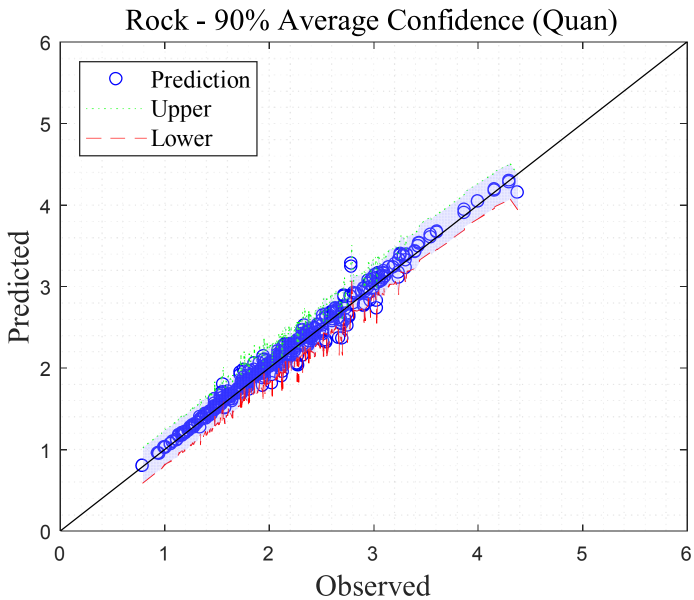


Figure 4.29 90% average confidence interval using quantiles for rock armor

Tetrapod

Figures 4.30, 4.31, and 4.32 show the prediction value and 90% confidence using standard deviation and quantiles, respectively. And Figures 4.33 and 4.34 show average confidence interval using standard deviation and quantiles, respectively. Since confidence level is 90%, 1.3 was selected as the value of $c_{confidence}$ and 5% and 95% quantiles were used. Also, 90% average confidence intervals using standard deviation and quantiles were calculated as follows, respectively.

$$\left[N^i_{s,pred} - 0.61, N^i_{s,pred} + 0.61\right] \quad (4.3$$

and

$$\left[N^i_{s,pred} - 0.407, N^i_{s,pred} + 0.485\right] \quad (4.4$$

The blue circle, the green triangle, and the red inverted triangle indicate the prediction value, the upper confidence interval, and the lower confidence interval, respectively.

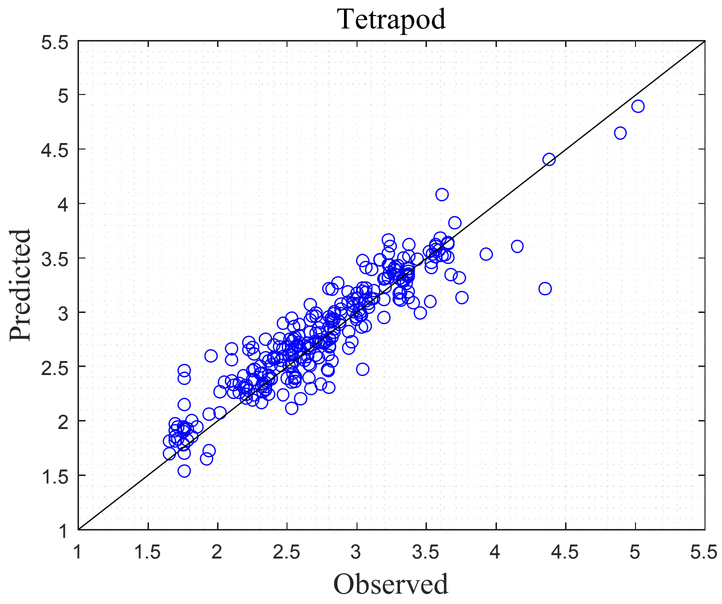


Figure 4.30 Predicted values versus observed values for Tetrapod

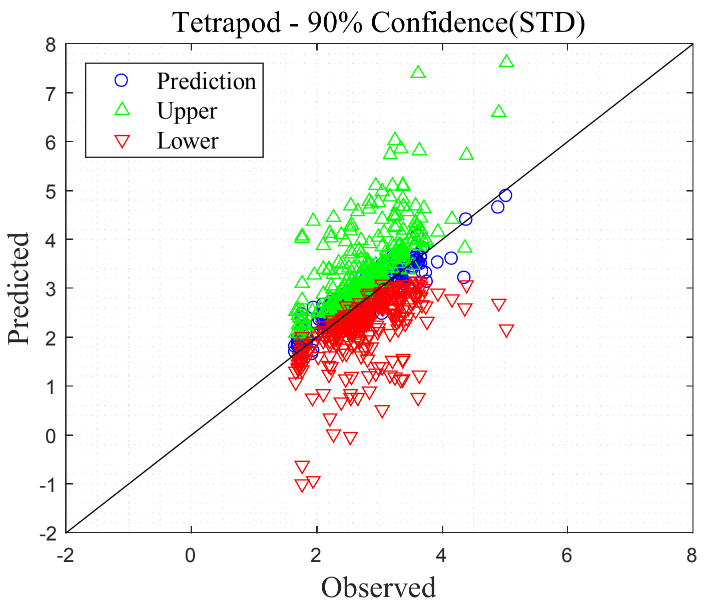


Figure 4.31 90% confidence interval using standard deviation for Tetrapod

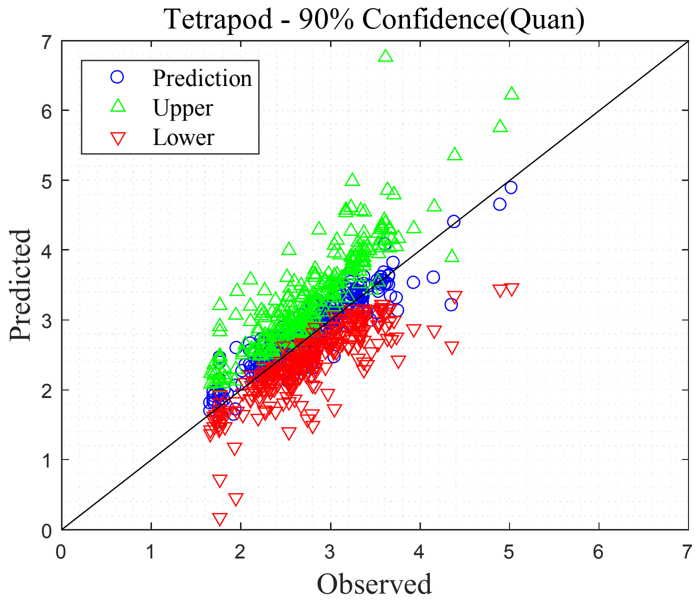


Figure 4.32 90% confidence interval using quantiles for Tetrapod

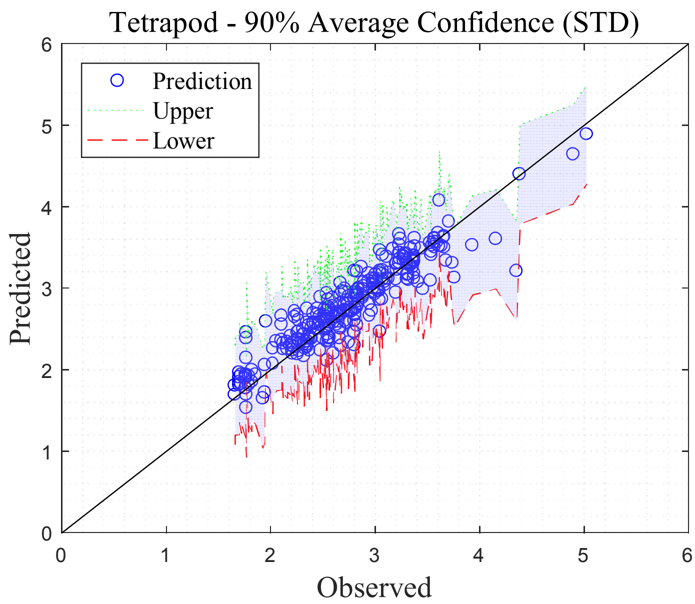


Figure 4.33 90% average confidence interval using standard deviation for Tetrapod

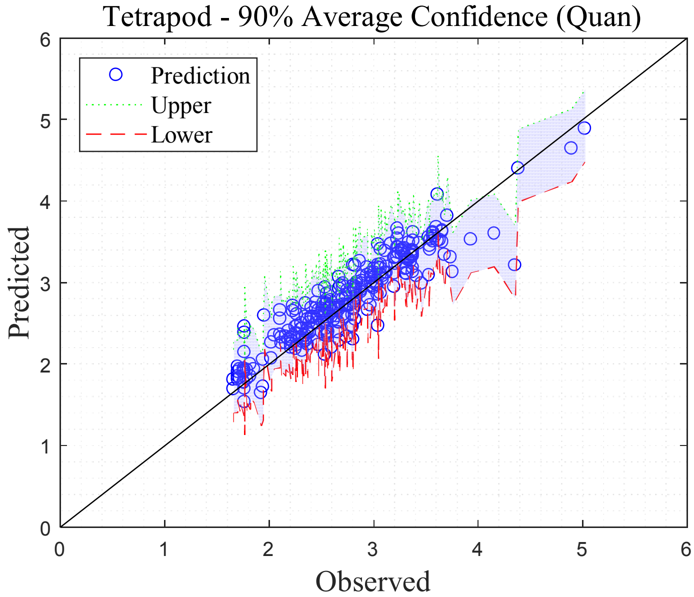


Figure 4.34 90% average confidence interval using quantiles for Tetrapod

4.4 Verification of ANN prediction model

4.4.1 Sensitivity analysis

Tables 4.1 and 4.2 show the results of OFAT method for rock armor and Tetrapod, respectively. The range and baseline of input, and range of stability numbers were given for the data which were illustrated in Section 3.1, and the mean value and standard deviation of stability number were calculated for the same data. The number of wave and slope angle of structure were shown as the least sensitive parameters for rock armor and Tetrapod, respectively. However, both variables are significant parameters related to the stability number, and thus, the number of wave and slope angle were included in the empirical formulas of Van der Meer (1987a) and Suh and Kang (2012), respectively. In addition, the more input parameters were, the more accurate the prediction was (See Tables 4.3 and 4.4). Therefore, it is better not to exclude the two parameters from the input parameters.

Table 4.1 Results of sensitivity analysis for rock armor

Parameter	Range	Baseline	Range of N_s	Mean of N_s	STD of N_s
H_s (m)	0.046-1.18	0.147	1.125-5.998	2.200	0.872
T_m (s)	1.24-4.4	2.195	2.482-2.515	2.497	0.010
ξ_m	0.67-7.58	2.98	2.489-2.883	2.567	0.095
S	0.32-32.97	7.40	2.452-2.519	2.486	0.018
N	1000 or 3000	1000	2.479-2.493	2.486	0.007
h (m)	0.2, 0.4, 0.8, or 5	0.8	0.715-5.177	2.545	0.614
P	0.1,0.5, or 0.6	0.1	2.298-2.637	2.434	0.104
cotα	2,3,4, or 6	2	2.493-2.539	2.518	0.021

Table 4.2 Results of sensitivity analysis for Tetrapod

Parameter	Range	Baseline	Range of N_s	Mean of N_s	STD of N_s
H_s (m)	0.087-0.266	0.165	2.656-2.986	2.939	0.068
T_m (s)	1.036-2.990	1.782	2.385-4.346	2.990	0.335
ξ_m	2.33-6.86	3.25	0.741-4.354	3.017	0.412
N_{od}	0-5.752	0.64	2.434-3.877	2.808	0.318
N	427-3078	1641.9	2.898-3.084	3.009	0.078
R_c (m)	-0.072-0.3	0.209	2.677-4.071	2.983	0.319
Φ	0.88, 0.95, or 1.02	1.02	2.386-2.971	2.943	0.112
cotα	1.33, 1.5, or 2	1.5	2.945-3.053	2.979	0.031

Table 4.3 Comparison with model except for *N* for rock armor

Model	R	RMSE	NSE	IOA	MBE
Model without exception	0.993	0.079	0.984	0.996	0.011
Model except for *N*	0.990	0.096	0.977	0.994	0.015

Table 4.4 Comparison with model except for cot α for Tetrapod

Model	R	RMSE	NSE	IOA	MBE
Model without exception	0.924	0.221	0.849	0.958	0.020
Model except for *N*	0.923	0.224	0.844	0.957	0.025

4.4.2 The number of models

Tables 4.5 and 4.6 show comparison of the performance of the ANN method in the entire data increasing the number of model from 100 to 500 for rock armor and Tetrapod, respectively. As can be seen in the Tables 4.5 and 4.6, the number of models cannot make a significant difference in performance. When calculating the confidence interval using quantile, the interval for various confidence levels can be presented as the number of models increases, therefore, 500 models used as used in this study.

Table 4.5 Performance increasing the number of model for rock armor

The number of models	R	RMSE	NSE	IOA	MBE
100	0.993	0.078	0.985	0.996	0.009
300	0.993	0.078	0.985	0.996	0.010
500	0.993	0.079	0.984	0.996	0.011

Table 4.6 Performance increasing the number of model for Tetrapod

The number of models	R	RMSE	NSE	IOA	MBE
100	0.920	0.229	0.838	0.955	0.025
300	0.922	0.223	0.846	0.957	0.020
500	0.924	0.221	0.849	0.958	0.020

4.5 Performance evaluation of method

As stated in Section 3.6, the entire and test data sets were used in performance evaluation of method, and five statistical parameters were calculated and compared with those of the previous methods.

Rock Armor

Table 4.7 shows the performance of the ANN method in the entire data and test data set, and the performance in the entire data is better than that of test data set. Although the test data set were not used in constructing model, it has fine performance in the test data set. Table 4.8 shows the comparison of the previous methods and present method with the entire data set, and the performance of present method improved than that of the other methods.

Table 4.7 Performance in the entire and test data set for rock armor

Data	R	RMSE	NSE	IOA	MBE
Entire data	0.993	0.079	0.984	0.996	0.011
Test data	0.990	0.102	0.985	0.994	0.021

Table 4.8 Comparison of the previous methods and present method for rock armor

Author	R	RMSE	NSE	IOA	MBE	Remarks
Van der Meer (1987a)	0.907	0.272	0.815	0.948	0.045	Empirical formula
Mase et al. (1995)	0.91	-	-	-	-	Including data of Smith et al. (1929)
Kim and Park (2005)	0.902-0.952	-	-	-	-	Including data of low-crested structures
Balas et al. (2010)	0.906-0.968	-	-	-	-	ANN-PCA hybrid models
Lee et al. (2016)	0.973	-	-	0.948	-	HS-ANN hybrid model
Present method	0.993	0.079	0.984	0.996	0.011	-

Tetrapod

Table 4.9 shows the performance of the ANN method in the entire data and test data set, as like for rock armor, it shows good performance in the test data set. Table 4.10 shows the comparison of the previous methods and present method with the entire data set, and it can be seen that the present method is much more accurate than the empirical formulas. There are two reasons why the prediction in Tetrapod is less accurate than that in rock armor. First, the characteristics by the shape of Tetrapod results in less accurate prediction compared to the rock armor. Although the shape of Tetrapod increases the permeability which leads to increasing absorption of incident wave, reducing reflection and overtopping, there are limitations because of legs broken off and broken pieces causing further damage within the armor layer (USACE, 2006). Also, uncertainty from those limitations makes prediction of stability number less accurate. Secondly, as explained in Section 3.1, while single data set of Van der Meer (1988a) were used in evaluation and constructing of method for rock armor, the data sets that were obtained from different sources were used for Tetrapod.

Table 4.9 Performance in the entire and test data set for Tetrapod

Data	R	RMSE	NSE	IOA	MBE
Entire data	0.924	0.221	0.849	0.958	0.020
Test data	0.882	0.261	0.770	0.933	0.025

Table 4.10 Comparison of the previous methods and present method for Tetrapod

Author	R	RMSE	NSE	IOA	MBE	Remarks
Van der Meer (1988b)	0.791	0.371	0.501	0.875	-0.038	Surging breaker, Cotα = 1.5
De Jong (1996)	0.824	0.348	0.679	0.898	0.014	Plunging breaker, Cotα = 1.5
Suh and Kang (2012)	0.805	0.420	0.455	0.878	0.048	Regardless of breaker type and slope angle
Present method	0.924	0.221	0.849	0.958	0.020	Regardless of breaker type and slope angle

4.6 How to use Excel files

For the convenience of practicing engineers, Excel files are attached to the following web site (stabilitynumber.blogspot.com). Excel files were made with weights and biases of the ANN model for rock armor and Tetrapod, respectively.

1. Click on the words "Download".

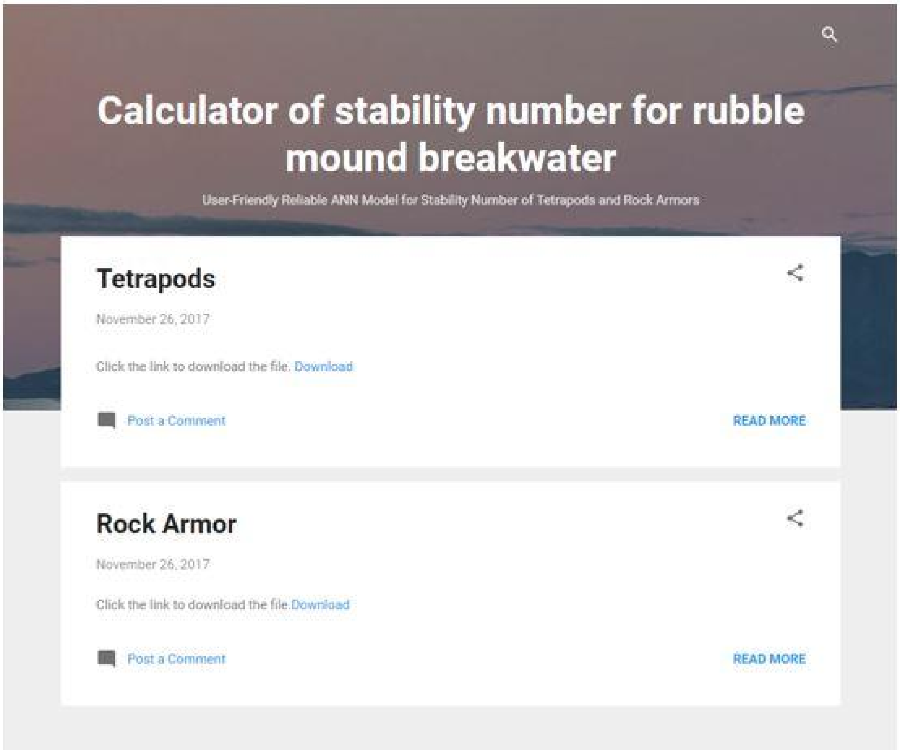
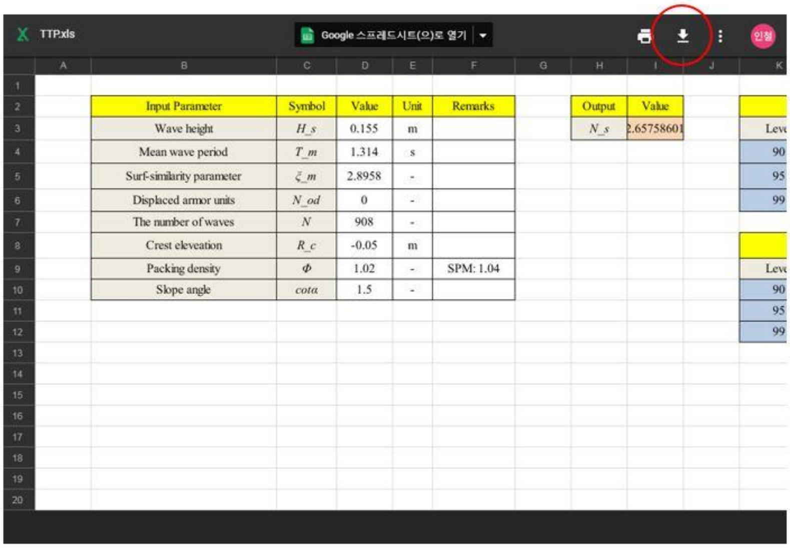


Figure 4.35 How to download Excel files 1

2. Click on the download icon and save the files.

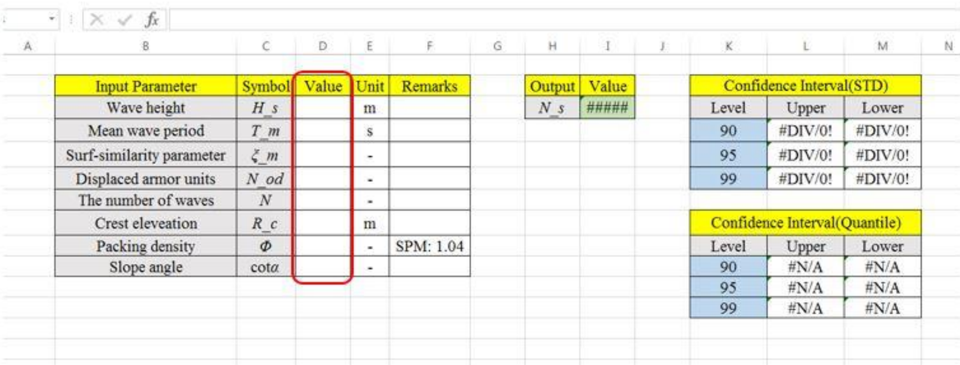


Input Parameter	Symbol	Value	Unit	Remarks
Wave height	H_s	0.155	m	
Mean wave period	T_m	1.314	s	
Surf-similarity parameter	ξ_m	2.8958	-	
Displaced armor units	N_od	0	-	
The number of waves	N	908	-	
Crest eleveation	R_c	-0.05	m	
Packing density	Φ	1.02	-	SPM: 1.04
Slope angle	$cot\alpha$	1.5	-	

Output	Value
N_s	2.65758601

Figure 4.36 How to download Excel files 2

3. At the “N_s” tap, enter the input values.



Input Parameter	Symbol	Value	Unit	Remarks
Wave height	H_s		m	
Mean wave period	T_m		s	
Surf-similarity parameter	ξ_m		-	
Displaced armor units	N_od		-	
The number of waves	N		-	
Crest eleveation	R_c		m	
Packing density	Φ		-	SPM: 1.04
Slope angle	cota		-	

Output	Value
N_s	#####

Confidence Interval(STD)		
Level	Upper	Lower
90	#DIV/0!	#DIV/0!
95	#DIV/0!	#DIV/0!
99	#DIV/0!	#DIV/0!

Confidence Interval(Quantile)		
Level	Upper	Lower
90	#N/A	#N/A
95	#N/A	#N/A
99	#N/A	#N/A

Figure 4.37 Entering the input values

4. Output value, prediction value of stability number and confidence intervals corresponding to the input values would be shown.

Input Parameter	Symbol	Value	Unit	Remarks
Wave height	H_s	0.155	m	
Mean wave period	T_m	1.314	s	
Surf-similarity parameter	ξ_m	2.896	-	
Displaced armor units	N_od	0	-	
The number of waves	N	908	-	
Crest eleveation	R_c	-0.05	m	
Packing density	Φ	1.02	-	SPM: 1.04
Slope angle	cotα	1.5	-	

Output	Value
N_s	2.658

Confidence Interval(STD)		
Level	Upper	Lower
90	3.589065	1.726107
95	3.764071	1.551101
99	4.114081	1.201091
Confidence Interval(Quantile)		
Level	Upper	Lower
90	3.577911	1.893773
95	3.75394	1.756541
99	4.591433	1.118321

Figure 4.38 Output value and confidence intervals

CHAPTER 5. CONCLUSIONS

5.1 Research summary

To construct the rubble mound breakwaters, the stability number should be the most significant parameter to be considered, because the stability number determines the appropriate weight of armor units of concrete or rock. There have been a lot of studies to predict the stability number from empirical formulas to the ANN models. Several machine learning formulas such as ANN models have been provided to predict the stability number more accurately, but most ANN models were difficult for practicing engineers to use. In this study, for the convenience of practicing engineers, the way to obtain the stability number was provided by simply entering the input variable through the Excel file. In addition, the reliability assessment, which was not covered by most existing ANN models, was considered: the confidence interval corresponding to the confidence level. Although more studies are necessary for how to use the confidence interval, we can suggest following application method of confidence interval. (1) When calculated confidence interval is wide, there should be more examination in application of the prediction value. (2) When more conservative design is required, upper limit might be more applicable, and vice versa. Lastly, even if it is a formula created through experiment data, it is possible to directly apply the prototype values because it is a formula created after scaling the parameters by using the Froude similitude law.

5.2 Research limitations and future study

There are research limitation in this study that should be improved in the future. First of all, even if the fact that the data set which is used in test and not used in training process gives a guarantee of validity in test process, if there is data set that is not related with this study, the evaluation of the method could be carried out more precisely and objectively. Also, if there are observed prototype data set in the field, we can prove whether the scaling of data through Froude similitude law is effective. Lastly, in the figures showing the confidence interval (Figures 4.26, 4.27, 4.31, and 4.32), it can be seen that some upper limits or lower limits are greatly deviated from $y = x$ graph. Although the reason for this phenomenon is supposed that the physical properties of the entire data is different from those of the specific data, further study is needed to figure out the exact reason.

REFERENCES

Balas, C. E., Koc, M. L., and Tür, R. (2010). "Artificial neural networks based on principal component analysis, fuzzy systems and fuzzy neural networks for preliminary design of rubble mound breakwaters." *Applied Ocean Research* 32, 425-433.

Burger, G. (1995). "Stability of low-crested breakwaters". Final Proceedings, EU research project Rubble mound breakwater failure modes, MAST, 2.

De Jong, R. J. (1996). "Wave transmission at low-crested structure, Stability of Tetrapod at front, crest and rear of a low-crested breakwater." Master thesis, Delft Univ. of Technology, Netherlands.

Erdik, T. (2009). "Fuzzy logic approach to conventional rubble mound structures design." *Expert Systems with Application* 36, 4162-4170.

Etemad-Shahidi, A., and Bonakdar, L. (2009). "Design of rubble-mound breakwaters using M5' machine learning method." *Applied Ocean Research* 31, 197-201.

Efron, B., and Tibshirani, R. J. (1994). "An introduction to the bootstrap." CRC press.

Goda, Y. (2009). "Random seas and design of maritime structures." World scientific.

Heskes, T. (1997). "Practical confidence and prediction intervals." *Advances in neural information processing systems*, 176-182.

Hudson, R. Y. (1959). "Laboratory investigation of rubble-mound breakwaters." *J.*

Waterw. Harbors Div. 85(WW3), 93-121.

Kim, D. H., and Park, W. S. (2005). “Neural network for design and reliability analysis of rubble mound breakwaters.” *Ocean Engineering* 32, 1332-1349.

Lee, A., Geem, Z., and Suh, K.-D. (2016). “Determination of optimal initial weights of an artificial neural network by using the harmony search algorithm: Application to breakwater armor stones.” *Applied Sciences* 6(6), 164; doi:10.3390/app6060164.

Mase, H., Sakamoto, M., and Sakai, T. (1995). “Neural network for stability analysis of rubble-mound breakwaters.” *J. Waterway, Port, Coastal, Ocean Eng.* 121(6), 294-299.

Smith, W. G., Kobayashi, N., and Kaku, S. (1992). “Profile changes of rock slopes by irregular waves.” *Proc., 23rd Int. Conf. Coastal Engineering*, ASCE, Reston, VA, 1559-1572.

Suh, K.-D., and Kang, J., (2012). “Stability formula for Tetrapod.” *J. Waterway, Port, Coastal, Ocean Eng.,* 138(3), 261-266.

U.S. Army Coastal Eng. Res. Center., (1984). “Shore Protection Manual, 3rd ed.” U.S. Government Printing Office, Washington, D.C., USA.

U.S. Army Coastal Eng. Res. Center., (2006). “Coastal Engineering Manual” U.S. Government Printing Office, Washington, D.C., USA.

Van der Meer, J. W. (1987a). “Stability of breakwater armor layers − Design formulae.” *Coastal Engineering* 11, 93-121.

Van der Meer, J. W. (1987b). “Stability of rubble mound breakwaters, stability formula for breakwaters armoured with Tetrapod." *Report on Basic Research, H462 Volume II*, Delft Hydraulics Laboratory, The Netherlands.

Van der Meer, J. W. (1988a). *Rock slopes and gravel beaches under wave attack.* Delft Hydraulics, Communication No. 396.

Van der Meer, J. W. (1988b). "Stability of cubes, Tetrapod and accropode." *In: Design of breakwaters*, Thomas Telford, London, 71-80.

Van den Boogaard, H. F. P., Mynett, A. E., and Heskes, T. M. (2000). "Resampling techniques for the assessment of uncertainties in parameters and predictions of calibrated models." *Proc. Hydro informatics Conference*, IOWA.

Van Gent, M. R., and van den Boogaard, H. F. (1998). "Neural network modelling of forces on vertical structures." *ASCE. Proc. ICCE 1998*(2), 2096-2109.

Van Gent, M. R., van den Boogaard, H. F., Pozueta, B., and Medina, J. R. (2007). "Neural network modelling of wave overtopping at coastal structures." *Coastal Engineering*, 54(8), 586-593.

Willmott, C. J. (1981). "On the validation of models." *Physical Geography* 2, 184-194.

국문초록

실무자들을 위한 신뢰성 있는 사석 및 테트라포드 피복재의 안정수 인공신경망 모델

서울대학교 대학원

건설환경공학부

김 인 철

경사식 방파제의 안정수는 주어진 해안 환경에서의 콘크리트나 사석에 대한 적정 중량을 결정하기 때문에, 방파제의 안정성에 있어서 중요한 변수 중 하나이다. 경사식 방파제의 안정수를 계산하는 방법과 공식에 대한 연구는 1950년대부터 지금까지 계속되어 오고 있다. 1959년도의 Hudson 식을 시작으로 여러 실험 자료를 통한 실험식들이 제안되었다. 그러나 대부분의 실험식들은 실제 값과 예측 값에 큰 차이를 보이기 때문에, 최근 20년동안 기계 학습을 통한, 특히 인공신경망을 통한 연구가 활발하게 진행되고 있는 상황이다. 그러나 대부분의 인공신경망을 이용한 방법들은 신뢰성에 대한 정보를 제공하고 있지 않다. 또한, 입력자료를

이용하여 출력 값을 계산하는 과정에 대한 명확한 계산 방법을 제시하고 있지 않기 때문에, 현장에서 바로 적용하기 힘든 경우가 대부분이다. 본 연구에서는 이러한 문제점들을 해결하기 위해, 우선 bootstrap resampling을 사용한 인공신경망 모델을 제안하였다. bootstrap resampling을 사용하면 예측 값의 신뢰성에 관련된 정보를 제공할 수 있다. 또한, 실무자들이 현장에서 적용할 수 있게 인공신경망 모델의 가중치(weight)와 편향(bias)을 이용하여 만든 엑셀 파일을 제공하였다.

Keywords: 테트라포드, 사석, 안정수, 기계 학습, 인공 신경망, 신뢰성 평가, 신뢰 구간

학번: 2016-21249