Prediction of Transmission Coefficients of Regular Wave Attenuation by Emergent Vegetation

Shangpeng Gong
Changsha University of Science & Technology
JIE CHEN (chenjie166@163.com)
Changsha University of Science & Technology
Changbo Jiang
Changsha University of Science & Technology
Fei He
University of Western Australia
Zhiyuan Wu
Changsha University of Science & Technology

Research Article

Keywords: Prediction, ANNs, MNLR, emergent vegetation

Posted Date: May 19th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-429711/v1

Prediction of transmission coefficients of regular wave attenuation by emergent vegetation

Shangpeng Gong[1], Jie Chen[1,2,✉], Changbo Jiang[1,2], Fei He[3]& Zhiyuan Wu[1,2]

[1]School of Hydraulic Engineering, Changsha University of Science & Technology, Changsha, China. [2]Key Laboratory of Dongting Lake Aquatic Eco-Environmental Control and Restoration of Hunan Province, Changsha, China. [3]School of Civil, Environmental and Mining Engineering, The University of Western Australia, Perth, Australia.✉email: chenjie166@163.com

Transmission coefficient (K_t) for wave attenuation by vegetation is essential parameter for predicting the wave height. In this paper, based on the experimental data of three kind of artificial vegetation model, genetic programming (GP), artificial neural networks (ANNs) and multivariate non-linear regression (MNLR) were used to analyze the dimensionless factors including Ursell number (Ur), relative width (R_B) relative height (α) and volume fraction (φ). The proposed GP formulae were compared with MNLR and ANNs. The predictions of GP models were in good agreement with measured data, and outperformed MNLR equations. Otherwise, GP and ANNs were used to obtain the weight of each factor. The results can provide a reference for the artificial planting of the three plants.

The emergent vegetation represented by mangroves is widely distributed in middle and low latitudes, which can effectively attenuate wave height and protect the coastline from the action of storm surges[1,2]. The attenuation law of waves in the nearshore vegetation zone is a major concern in the study of wave dissipation[3,4]. The wave absorbing ability of vegetation can be expressed by the transmission coefficient (K_t). On the one hand, as the sea level rises, the occurrence of extreme waves such as storm surges is increasing. The K_t can be predicted according to the parameters such as incident wave height, wave period, vegetation density, and the disaster warning can be carried out in the area with large transmission coefficient. On the other hand, with

the strengthening of local environmental protection awareness, measures for artificial mangrove planting have increased, and the area of mangroves has generally shown a trend of first decreasing and then increasing in some countries. The sensitivity analysis of the factors that affect K_t is carried out to obtain the parameters that have a greater impact on the waves. This can provide guidance for the planting of mangroves and achieve the greatest economic effect.

Multivariate non-linear regression (MNLR) analysis is the most traditional approaches to establish prediction formulas[5,6]. However, that requires active assumption of the objective function, which requires background knowledge about the relationship between input and output. Furthermore, some formulas are complicated in form and do not indicate the actual physical process, and the prediction accuracy needs to be improved. Predictive methods such as artificial neural networks (ANNs) and genetic programming (GP) can effectively deal with complex multivariate nonlinear problems, and have been used to estimate hydraulic characteristics[7–9]. The main advantage of using GP for symbolic regression is that there is no need to specify the size and shape of the approximation function in advance, and the specific knowledge of the problem can be included in the search process through an appropriate mathematical function[10]. ANNs cannot obtain a definite formula like GP, but it has excellent performance in sensitivity analysis. Therefore, in this paper, we demonstrated the applicability of GP in the prediction of K_t. In order to compare the accuracy of the prediction results, the prediction results obtained by GP, ANNs and MNLR methods were compared with experimental data. Otherwise, Sensitivity analysis was carried out by using GP and ANNs in order to get the weightings of dimensionless parameters.

Methods

Transmission coefficient

The normally accepted dimensionless parameter for evaluating the performance of the vegetation is the transmission coefficient (K_t)[4,6], which is the ratio of the

transmitted wave height to the incident wave height[5,11]. K_t can range from 0 to 1, where 0 illustrates no transmission and 1 illustrates complete transmission. The larger the K_t, the weaker the wave-dissipating ability of vegetation. K_t can be defined as a function of wave, fluid, and structure properties.

Experimental study is the main approach to understand the mechanism of wave attenuation by vegetation[12,13]. Chen et al[6] obtained an empirical formula (Eq. 1) of K_t which consists of platform, stems and roots through multivariate regression analysis based on the 244 test cases.

$$K_t = C_{platform}\left(\frac{H}{L},\frac{W}{L},\frac{H}{D}\right)\bullet C_{root}\left(\frac{W}{L},\frac{H}{D},\frac{h_r}{D},\frac{h_r b_v a}{D}\right)\bullet C_{steam} \tag{1}$$

He et al[5] studied the performance of the stem, root, and canopy on regular wave attenuation by vegetation, laboratory experiments were conducted testing seven vegetation models, five wave heights, four vegetation densities and four water depths. The coefficient of determination, R^2, between the K_t and dimensionless parameters is 0.822.

$$K_t = -3.92\times10^{-9}\underbrace{\left(\frac{L^2H_0}{h^3}\right)^{0.18}}_{Hydrodynamics}\underbrace{\left(\frac{h_s}{h}\right)^{-44.524}}_{Stem}\underbrace{\left[\left(\frac{h_r}{h}\right)^{-0.02}\right]^j}_{Root}\underbrace{\left[\left(\frac{h_c}{h}\right)^{62.40}\right]^j}_{Canopy}\underbrace{\left(\frac{B}{L}\right)^{-96.37}}_{Vegetation\ length}\underbrace{(\varphi)^{4.528}}_{Stem\ density}+0.777 \tag{2}$$

Wave attenuation is related to hydrodynamic factors and the characteristics of vegetation. Phan et al[14] studied the effect of plant density on wave attenuation by simulating plants with cylinders. Augustin et al[15] studied wave height attenuation in plant zones by changing plant density. Akgul et al[16] studied the wave reduction and hydrodynamic factors of emergent plants. Blackmar et al[17] studied the attenuation law of wave height in heterogeneous plant zones through wave flume experiment and numerical simulation. John et al[18] pointed out that wave attenuation was related to vegetation type, height, density and stiffness as well as wave characteristics. Paul et al[19] studied the wave attenuation effect of submerged plants and claimed that wave attenuation was closely related to the length, density and stiffness of seaweed leaves. The research of Gedan et al[20] and Möller et al[21] show that the effect of wave attenuation depends on the submerged water depth and incident wave energy. Huang et al[22] found that K_t decreased with the increase of distribution density. Referring to

the previous research, and combined with three experimental set-ups. A general function expression representing K_t is given by:

$$K_t = f(h, H_i, L, h_v, B, \varphi) \tag{3}$$

The wave length of a laboratory scale regular wave can be calculated by:

$$L = \frac{gT^2}{2\pi} \tanh kh \tag{4}$$

The volume fraction can be calculated by the following formula:

$$\varphi = \frac{V_s}{V} \tag{5}$$

where V_s is the volume of water, V is the volume of a vegetated area at still water.

The parameters of Eq. (3) were expressed in terms of dimensionless parameters using Buckingham's π theorem. The dimensionless change of the influencing factors in the relationship can be obtained as follows:

$$K_t = f(\frac{L^2 H}{h^3}, \frac{B}{L}, \frac{h_v}{h}, \varphi) \tag{6}$$

For brevity, Ursell number (U_r) is equal to L^2H/h^3, relative width (R_B) is equal to B/L, relative height (α) is equal to h_v/h.

$$K_t = f(U_r, R_B, \alpha, \varphi) \tag{7}$$

Test and training data

The experiments of regular wave attenuation by three kind of emergent vegetation were conducted in a rectangular flume with dimension of 40.0 m long× 0.5 m wide × 0.8 m deep (Fig. 1), in the Changsha University of Science & Technology, China. The planned regular waves were generated by a piston-type wave generator on the left side of the flume, and measured by the wave gauges (G1~G3). G1 and G2 were placed in front of the vegetation to measure incident waves and reflected waves by using the approach of Goda and Suzuki[23], and G3 was placed behind the vegetation to measure transmission wave. A porous, clival wave absorber was constructed at the opposite end to reduce reflection. The vegetation model was placed in the middle of the flume.

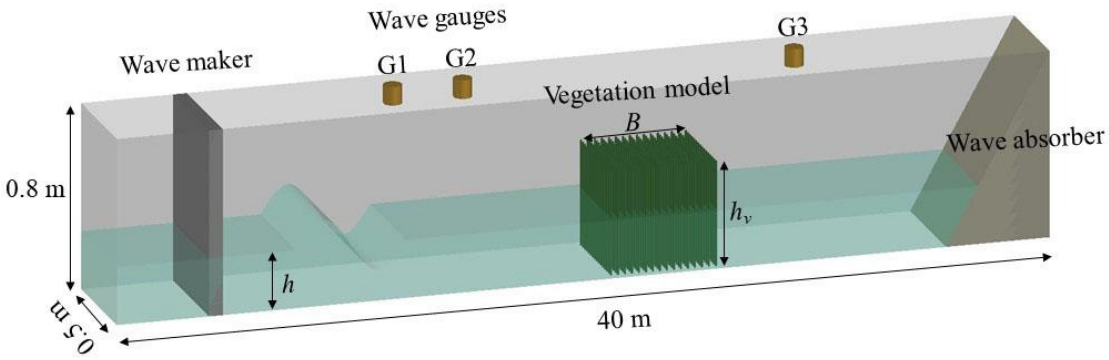


Figure 1. Experimental setup.

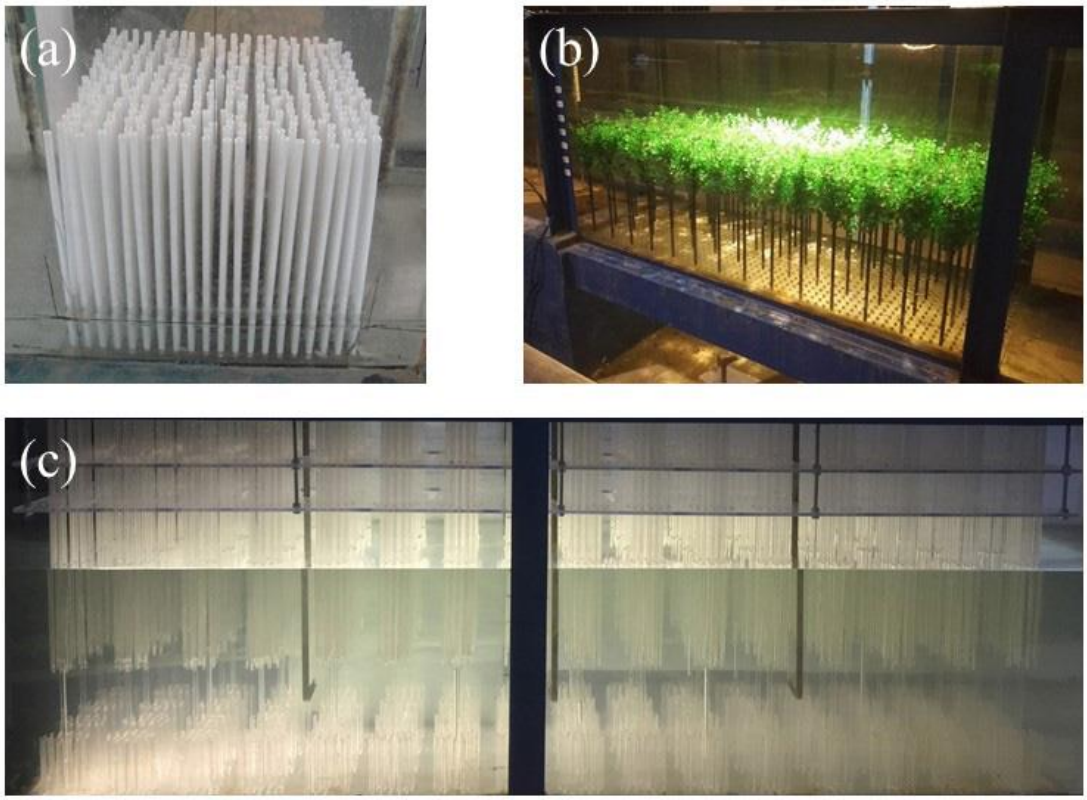


Figure 2. (a) Rigid cylinder model (Model 1), (b) Plastic synthetic vegetation model (Model 2), (c) Cylinder array model (Model 3).

The rigid cylinder model (Fig.2a), which consisted of rigid cylindrical PVC sticks, was placed on a plexiglass panel with a length of 50 cm, a width of 48 cm and a thickness of 1.6 cm. The diameter of the cylinder was 1.0 cm and the height was 50 cm. The plastic synthetic vegetation model (Fig.2b), which consisted of hemispherical roots, round rod-shaped trunks, and leaf-like canopy, was placed on a plexiglass panel with a length of 150 cm, a width of 48 cm and a thickness of 0.8 cm. The Cylinder array model (Fig.2c), which consisted of plexiglass tube, was placed on a plexiglass panel with a length of 200 cm, a width of 48 cm and a thickness of 0.8 cm. Layout

and detailed geometric parameters of the three models (Fig.2) were published in Gong et al[5].

For Model 1, A total of 3 regular waves cases were carried out, with water depth of 25.0 cm, wave period of 1.0 s and three wave heights were tested under water depth condition (Table 1). A total 12 runs were conducted covering four water heights from 25.0 to 45.0 cm and three wave height for each water depth for Model 2. Similarly, a total of 48 experiments were conducted in Model 3.

Table 1. Hydrodynamic conditions of vegetation model

Model	Water depth h (cm)	Wave height H (cm)	Wave period T (s)	Number of types of arrangement
Model 1	25.0	2.7,4.5,6.5	1.0	10
Model 2	25.0, 30.0	3.0,5.0, 9.0	1.0	4
	35.0, 45.0	5.0, 9.0, 12.0		
Model 3	25.0, 45.0	3.0, 5.0, 7.0, 9.0	1.0,1.5,2.0	4
	30.0, 35.0	5.0, 7.0, 9.0, 12.0		

Poor quality data leads to poor quality mining results, and the difference of order of magnitude will lead to the dominant position of the larger magnitude attribute and slow down the convergence speed of iteration. However, Eureqa and SPSS both depend on the sample distance is sensitive to the order of magnitude of the data, while each factor usually has different order of magnitude and dimensional order of magnitude. Therefore, decide whether to normalize scale and offset of the dimensionless data according to the prompts of the Eureqa.

Dimensionless data processing mainly solves the comparability of the data. The index values are all at the same quantity level, which can be used for comprehensive evaluation and analysis. Table 2 summarizes the main parameters of the three experimental databases.

Table 2. Summary of physical vegetation properties.

Models	Parameters	Scope
Model 1	K_t	[0.61,0.86]
	U_r	[2.72,7.49]
	R_B	[3.84]
	α	[2]
	φ	[0.04,0.15]
Model 2	K_t	[0.21,0.83]
	U_r	[1.32,9.02]
	R_B	[1.01,1.15]
	α	[1.00,1.80]
	φ	[0.003,0.008]

	K_t	[0.44,0.97]
	U_r	[1.31,60.54]
Model 3	R_B	[0.52,1.54]
	α	[1.00,1.80]
	φ	[0.017,0.039]

Genetic programming

Genetic programming (GP) is suitable for inductive mathematical models. It is robust, simple and universal, and has a strong ability to solve complex nonlinear problems. Such as the prediction of non-breaking wave induced scour depth at the trunk section of breakwaters[8]. Furthermore, based on laboratory experimental data, Koç et al[24] suggested that GP has good potential in solving complex problems in the field of coastal engineering. GP is an adaptive global optimization search algorithm, and its search process is a random search algorithm in essence, but the genetic operation in the evolution process can make the path jump to different subspaces randomly, which makes its spatial ergodicity better than the traditional heuristic search. Giustolisi[9] highlighted that the determination of channel Chèzy resistance coefficient using GP is more accurate than traditional formula. Keijzer et al[20] used the GP to determine the plant resistance equation and compared the theoretical research method with the data driven results based on GP, and the results showed that GP could obtain more concise and accurate relationships. Therefore, the GP method was used to obtain the formula for calculating the wave attenuation effect of plants, and the sensitivity analysis was carried out.

GP analysis for K_t forecasting process by Eureqa[26,27]. Table 3 shows the target expression, mathematical building-blocks, error metric, row weight and data splitting used in this study. The main procedures are shown as follows:

Step 1: Data entry. Enter dimensionless numbers into the Eureqa by column, and give names to each variable on the row named name. Each row representing a set of measurements or values that are in some sense simultaneous.

Step 2: Data preparation. The difference in figure of variables is minor, and standardization can be decided according to the prompts of the software.

Step 3: Search definition. Editing the formula to target expression and selecting the appropriate mathematical building-blocks (i.e. arithmetic, trigonometric,

exponential). In general, we need to select as many operators as possible to get a more comprehensive expression, but this may lead to subjective overlearning. The mathematical building-blocks are selected based on the formula of He et al. (2019) in this paper. The default setting is able to minimize the mean absolute error (MAE) and it performs well in most cases, which is proved by the results obtained. By default, Eureqa will randomly shuffle the data and then split it into training and validation data sets based on the total size of the data. The training set is used to generate and optimize solutions, and the validation set is used to test how well those models generalize to new data. Eureqa also uses the validation data to filter out the best models to display in the Eureqa interface.

Step 4: start and stop search. When the program starts to run, generation will be generated continuously and will not stop automatically. The search does not stop until the MAE is less than 0.02 and remains constant or varies little over a long period of time. Table 1 shows the target expression, mathematical building-blocks, error metric, row weight and data splitting used in this study.

Table 3. The characteristics of employed GP models using Eureqa.

Parameter description	Setting of parameters
Target expression	$K_t = f(U_r, R_B, \alpha, \varphi)$
Mathematical building-blocks	c, x, +, - ,× ,÷ , sin, cos
Error metric	$MAE = \frac{1}{N}\sum_{i=1}^{N}\lvert y - f(x)\rvert$
Row weight	none
Data splitting	Treat all data points equally

Artificial neural networks

ANNs is a mathematical model that simulates the brain nervous system for complex information processing based on the main functions of the human brain. Such systems learn to perform tasks by considering examples, and usually do not need to write any task-specific rules. Given sufficient complexity of training networks, neural networks can represent arbitrary nonlinear functions. Network is composed of neurons. They are arranged into three basic input, hidden and output layer (Fig.3). ANNs gives relative weights between neurons relay, continuously adjust the weights by using the algorithm, so as to get the minimum prediction error and prediction precision is given, and is widely applied in the parameter sensitivity analysis.

ANNs analysis for K_t forecasting process by IBM SPSS Statistics 24, and the main procedures are shown as follows:

Step 1: The Multilayer Perceptron (MLP) procedure produces a predictive model for one or more dependent (target) variables based on values of the predictor variables. The neural network model used in this work is the MLP network which is the most common neural network model and have been used in many engineering relevant predictions. The MLP networks are generally good at fitting non-linear time series data and can capture the non-linear patterns of time series.

Step 2: Selecting Kt as a dependent variable, RH, RB, α and φ as covariates.

Step 3: Partition the data. Randomly assign cases based on relative numbers of cases, training 70%, test 30%. Automatic architecture selection.

Step 4: Selecting Architecture. Using one hidden layer, and the numbers of units is automatically.

Step 5: Selecting the output content (independent variable importance analysis and predicted by observed chart in the network performance group).

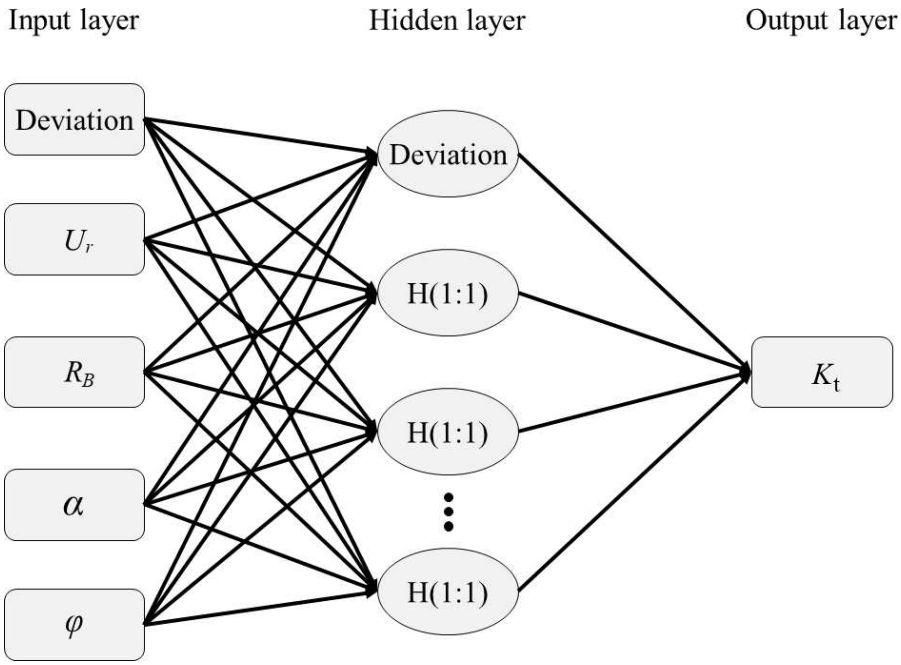


Figure 3. The ANNs structure used in this study.

Results and Discussion

Prediction of rigid cylinder model

A total of 30 dataset of Model 1 has been used for the prediction of the K_t using the GP, ANNs and MNLR models. K_t as the dependent variable, and the Ursell number (U_r), relative width (R_B), relative height (α) and volume fraction (φ) as the independent variables.The same process as He et al[5], the power exponential function as shown in Eq. (8) is obtained by using the MNLR method. The solution equation of GP method as shown in Eq. (9). However, there is no way for ANNs to produce specific and mathematical equations for the prediction of the phenomena.

$$K_t = 1.121\left(\frac{HL^2}{h^3}\right)^{-0.095}\left(\frac{B}{L}\right)^{0.678}\left(\frac{h_v}{h}\right)^{-0.051}\varphi^{-0.172} \tag{8}$$

$$K_t = 0.0799U_r^{\,2}\varphi - 0.94U_r\varphi - 0.00581U_r + 0.777\varphi + 0.926 \tag{9}$$

As indicated in Fig.4, the coefficients of determination (R^2) of MNLR, GP and ANNs are 0.75, 0.95 and 0.75, respectively. The empirical formula gives a better prediction for the approach of GP than others.

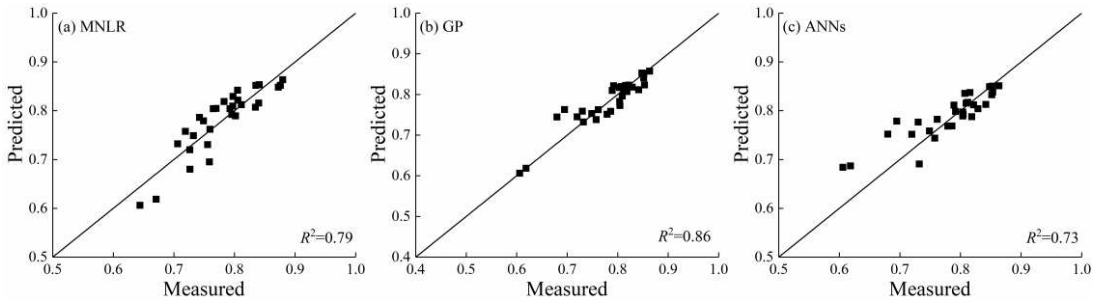


Figure 4. Comparison between the measured dataset from rigid cylinder model and predicted value using (a) non-linear regression, (b) GP and (c) ANNs.

Prediction of plastic synthetic vegetation model

A total of 48 dataset of Model 2 has been used for the prediction of the K_t using the GP, ANNs and MNLR models. To explore the differences between MNLR, GP and ANNs, treating K_t as the dependent variable, Ursell number (U_r), relative width (R_B), relative height (α) and volume fraction (φ) as the independent variables, the result of MNLR method as shown in Eq. (10). Similarly, the solution equation of GP method as shown in Eq. (11).

$$K_t = 0.098\left(\frac{HL^2}{h^3}\right)^{-0.094}\left(\frac{B}{L}\right)^{21.565}\left(\frac{h_v}{h}\right)^{-3.903}\varphi^{-0.29} \tag{10}$$

$$K_t = 16.9\varphi \times \sin\left(42.6\alpha\right) + 32R_B - 5.65\alpha - 34\varphi - 25.7 \tag{11}$$

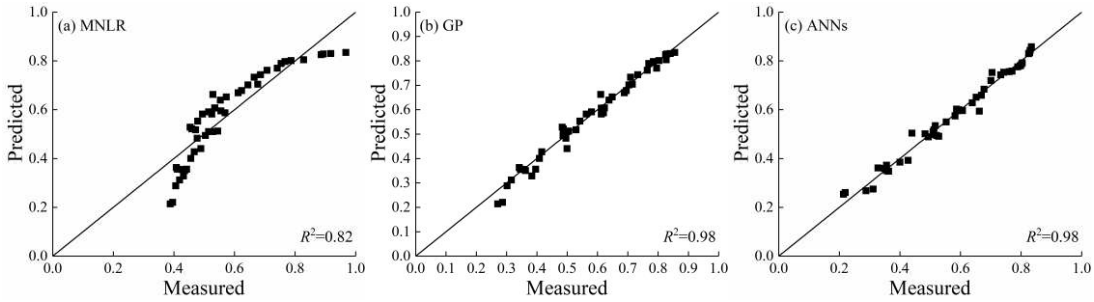


Figure 5. Comparison between the measured dataset from plastic synthetic vegetation model and predicted value using (a) MNLR, (b) GP and (c) ANNs.

Fig. 5 illustrates the relationship between the measured predicted K_t of the MNLR result, and its determination coefficient (R^2) is 0.75. Some data points deviate far from the fitting line, indicating that the formula obtained by using the MNLR has a poor fitting degree to the measured value. The relationship between the measured value and the predicted value of the GP results shown in Fig. 5b, and its R^2 is as high as 0.98. Compared with Fig. 5a, the data points and fitting line are closer to the comparison analysis results show that compared with the traditional MNLR method and ANNs method (Fig. 5c). Therefore, the genetic programming method can obtain more accurate prediction formula.

Prediction of cylinder array model

A total of 384 dataset of Model 3 has been used for the prediction of the K_t using the GP, ANNs and MNLR models. the solution equation of MNLR as shown in Eq. (12). Similarly, the solution equation of GP method as shown in Eq. (13).

$$K_t = 0.4\left(\frac{HL^2}{h^3}\right)^{-0.218}\left(\frac{B}{L}\right)^{-0.58}\left(\frac{h_v}{h}\right)^{0.609}\varphi^{-0.227} \tag{12}$$

$$K_t = 1.08 + 0.306\alpha + 1.83\times10^{-4}U_r^{\,2} - 0.0195U_r - 0.416R_B - 6.52\varphi \tag{13}$$

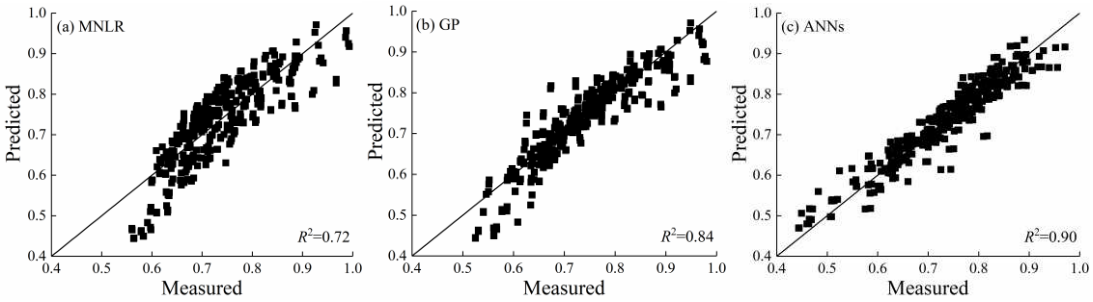
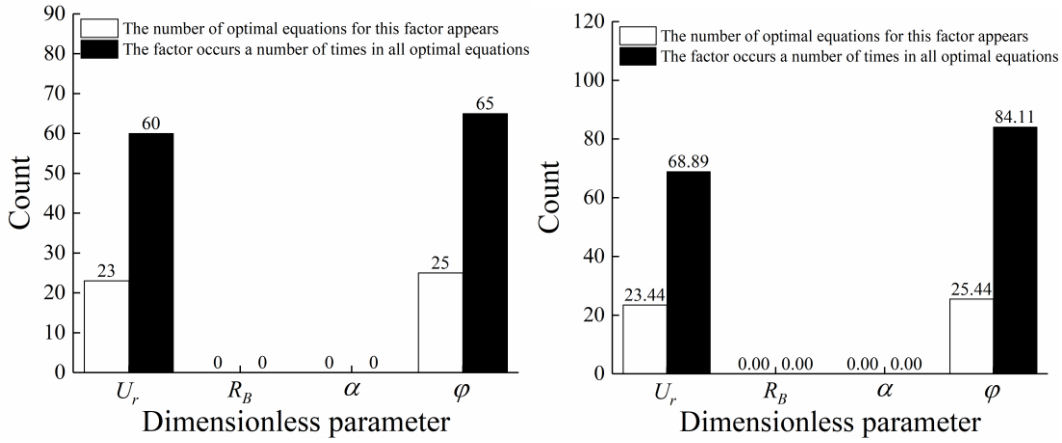


Figure 6. Comparison between the measured dataset from cylinder array model and predicted value using (a) MNLR, (b) GP and (c) ANNs.

As indicated in Fig. 6, the coefficients of determination of the three methods are similar. The R^2 of GP methods is higher than that of MNLR prediction.

Sensitivity analysis by genetic programming

The dimensionless factors may or may not appear in an optimal formula of a certain size, and they may occur once or more. The frequency of the dimensionless parameters appearing in 10 operations was counted. Fig. 7 shows the statistical result about each factors in all occurrences in the optimal equation and the factors of the optimal number of equations. The more the dimensionless parameters appear, the more closely they are related to the transmission coefficient.



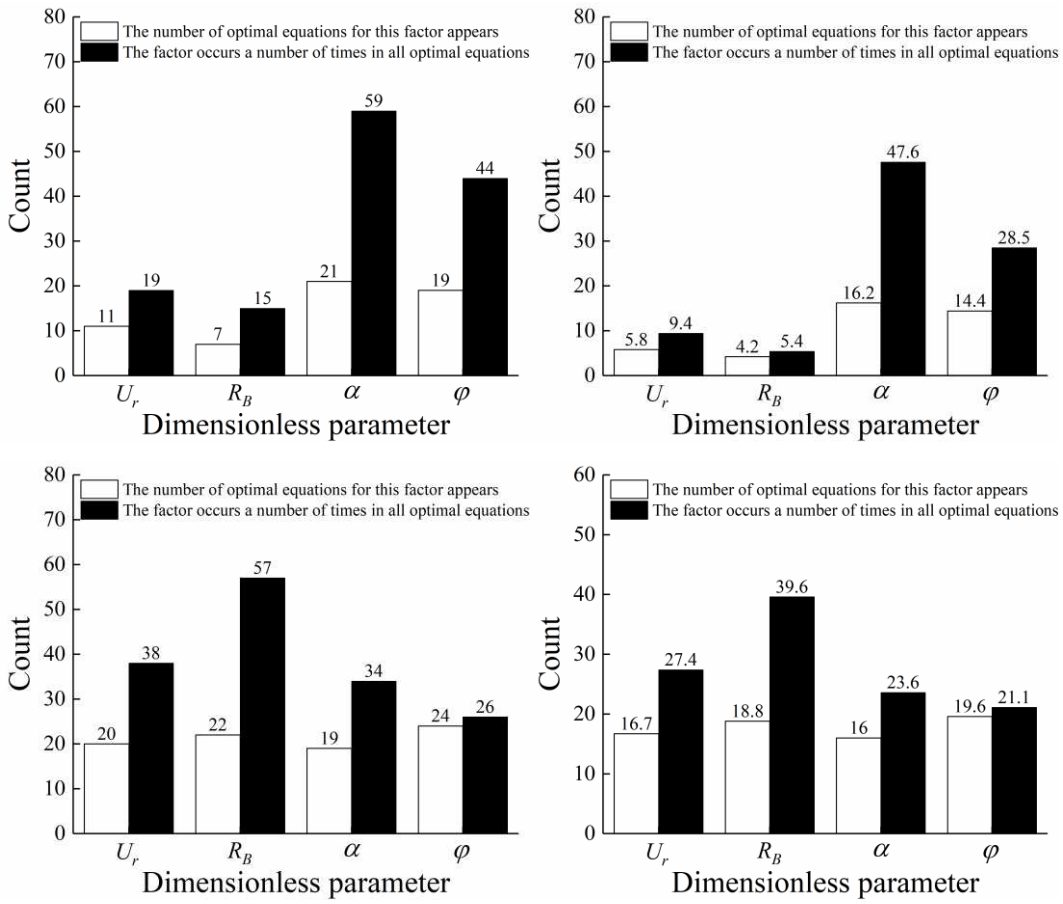


Figure 7. The mean values of each factor index were fitted for 10 times of the (a-b) Model 1, (c-d) Model 2 and (e-f) Model 3.

Sensitivity analysis by artificial neural networks

SPSS statistical analysis tool was used to analyze the experimental data. The analysis results are shown in Table 4.

Table 4. The importance of regular wave independent variables(ANNs)

Model	Independent variables	Importance	Normalized importance
Model 1	Ur	0.221	28.4%
	R_B	-	-
	α	-	-
	φ	0.779	100.0%
Model 2	Ur	0.080	18.1%
	R_B	0.292	65.8%
	α	0.443	100.0%
	φ	0.185	41.7%
Model 3	Ur	0.446	100.0%
	R_B	0.285	63.9%
	α	0.172	38.5%
	φ	0.097	21.7%

Comparison between the results obtained by using GP and ANNs methods

Table 5 ranks the weights of dimensionless factors. Generally speaking, the

conclusions (Table 5) of GP and ANNs are similar, and GP performs well in sensitivity analysis. As for the two methods of GP and ANNs, different results appear, it is probably due to the fact that the results of GP are randomly generated, leading to slight inconsistency in the optimal equation table of each fitting.

Table 5. Significance order of influencing factors of each model

Model	GP	ANNs
Model 1	$\varphi > U_r$	$\varphi > U_r$
Model 2	$\alpha > \varphi > U_r > R_B$	$\alpha > R_B > \varphi > U_r$
Model 3	$R_B > U_r > \alpha > \varphi$	$U_r > R_B > \alpha > \varphi$

It can be seen from Table 5 that the influence of dimensionless factors of different plant models on the K_t is not consistent. The sensitivity analysis results of the GP and the ANNs show that the volume fraction (φ) has a dominant effect on the wave attenuation effect of Model 1, and the influence of the Ursell number (U_r) is relatively small. Model 1 ignored the influence of crown and root, and had no density change in vertical direction. This model can be used to characterize rigid plants with unprotruding surface roots, large and unbending trunk, such as the black pine forest[28]. When planting this kind of plant, the planting density can be improved to achieve better wave reduction effect.

For the generalized plant model considering the effect of root, stem and leaves, different wave conditions and dimensionless factors have different effects. The effect of relative height (α) on the wave attenuation of plants is obviously dominant. Compared with the Model 1, it can be found that the influence factors of relative model height are improved, because the model takes into account the influence of plant crown and root, and the change of water depth will lead to different plant parts tobe submerged under wave motion. This model is similar to the normal mangrove with dense branches and well-developed roots. When planting this kind of vegetation, if it is used to protect the safety of coastal people, etc. during extreme events such as storm surge, the vegetation height can be improved to achieve better wave reduction effect. If it is mainly used to resist wind and waves and promote sediment deposition, vegetation of different tree ages and heights can be planted appropriately to achieve better wave reduction effect.

For the quantitative model considering the effect of root, stem and leaf. In regular

waves, the effect of U_r on wave attenuation was dominant. Compared with the Model 2 considering the effect of rhizome and leaves, it can be found that the influence factors of relative model width are improved. Tang et al[29] pointed out that coastal vegetation could significantly reduce wave height and wave run-up, and wave run-up would decrease monotonously with the increase of vegetation width. The model was considered to be similar to the coastal mangroves with dense foliage but relatively hard wood, respiration roots or strut roots protruding upward, such as sea mulberry and loam forest. Similarly, when planting this kind of plant, the planting width can be improved to achieve better wave reduction effect.

Conclusions

This paper studies the non-flooded rigid model, considering the influence of the generalized vegetation model and considering the effects of the Ursell number, relative width, the relative height and volume fraction, adopts the transmission coefficient size measure effect of wave absorption by vegetation. Using non-linear regression, genetic programming and artificial neural networks, formula fitting and weight analysis were carried out for each factor, and the following conclusions were obtained:

(1) The formulas of transmission coefficient and Ursell number, relative width, the relative height and volume fraction of different plant forms were established. The inner relationship between the transmission coefficient of flora and the factors in regular wave is revealed.

(2) Genetic programming method is useful in understanding and establishing new formulas, being capable of analyzing and predicting data effectively, and provide new ideas and methods for solving complex problems related to plant wave elimination.

References

1. Dalrymple, R. A., Kirby, J. T. & Hwang, P. A. Wave Diffraction Due to Areas of Energy Dissipation. *Journal of Waterway, Port, Coastal, and Ocean Engineering* **110**, 67–79 (1984).
2. Mazda, Y. *et al.* Drag force due to vegetation in mangrove swamps. *Mangroves &*

Salt Marshes **1**, 193–199 (1997).
3. Liu, P. L.-F. *et al.* Periodic water waves through an aquatic forest. *Coastal Engineering* **96**, 100–117 (2015).
4. Hoque, A., Husrin, S. & Oumeraci, H. Laboratory studies of wave attenuation by coastal forest under storm surge. *Coastal Engineering Journal* **60**, 225–238 (2018).
5. He, F., Chen, J. & Jiang, C. Surface wave attenuation by vegetation with the stem, root and canopy. *Coastal Engineering* **152**, 103509 (2019).
6. Chen, X., Chen, Q., Zhan, J. & Liu, D. Numerical simulations of wave propagation over a vegetated platform. *Coastal Engineering* **110**, 64–75 (2016).
7. Abbaspour, A., Farsadizadeh, D. & Ghorbani, M. A. Estimation of hydraulic jump on corrugated bed using artificial neural networks and genetic programming. *Water Science & Engineering* **000**, 189–198 (2013).
8. Pourzangbar, A. *et al.* Prediction of non-breaking wave induced scour depth at the trunk section of breakwaters using Genetic Programming and Artificial Neural Networks. *Coastal Engineering* **121**, 107–118 (2017).
9. Giustolisi, O. Using genetic programming to determine Chèzy resistance coefficient in corrugated channels. *Journal of Hydroinformatics* **6(3)**, (2004).
10. Lee, J., Kim, S., Lee, S., Kang, D. & Lee, J. Prediction of added resistance using genetic programming. *Ocean Engineering* **153**, 104–111 (2018).
11. Alamailes, A. & Turker, U. Using Analytical Approach to Estimate Wave Transmission Coefficient in Floating Structures. *J. Waterw. Port Coast. Ocean Eng.* **145**, 04019010 (2019).
12. Ismail, H., Abd Wahab, A. K. & Alias, N. E. Determination of mangrove forest performance in reducing tsunami run-up using physical models. *Nat Hazards* **63**, 939–963 (2012).
13. Chen, J., Duan, Z., Jiang, C. & Guan, Z. Laboratory study on tsunami erosion and deposition under protection of rigid emergent vegetation. *Nat Hazards* **92**, 995–1010 (2018).
14. Phan, K. L., Stive, M. J. F., Zijlema, M., Truong, H. S. & Aarninkhof, S. G. J. The effects of wave non-linearity on wave attenuation by vegetation. *Coastal Engineering* (2019) doi:10.1016/j.coastaleng.2019.01.004.
15. Augustin, L. N., Irish, J. L. & Lynett, P. Laboratory and numerical studies of wave damping by emergent and near-emergent wetland vegetation. *Coastal Engineering* **56**, 332–340 (2009).
16. Akgul, M. A., Yilmazer, D., Oguz, E., Kabdasli, M. S. & Yagci, O. The effect of an emergent vegetation (i.e. Phragmistes Australis) on wave attenuation and wave kinematics. *Journal of Coastal Research* **65**, 147–152 (2013).
17. Blackmar, P. J., Cox, D. T. & Wu, W.-C. Laboratory Observations and Numerical Simulations of Wave Height Attenuation in Heterogeneous Vegetation. *J. Waterway, Port, Coastal, Ocean Eng.* **140**, 56–65 (2014).
18. John, B. M., Shirlal, K. G. & Rao, S. Effect of Artificial Vegetation on Wave Attenuation – An Experimental Investigation. *Procedia Engineering* **116**, 600–606 (2015).

19. Paul, M. *et al.* Plant stiffness and biomass as drivers for drag forces under extreme wave loading: A flume study on mimics. *Coastal Engineering* **117**, 70–78 (2016).
20. Gedan, K. B., Kirwan, M. L., Wolanski, E., Barbier, E. B. & Silliman, B. R. The present and future role of coastal wetland vegetation in protecting shorelines: answering recent challenges to the paradigm. *Climatic Change* **106**, 7–29 (2011).
21. Möller, I. *et al.* Wave attenuation over coastal salt marshes under storm surge conditions. *Nature Geosci* **7**, 727–731 (2014).
22. Huang, Z., Wu, T.-R., Chen, T.-Y. & Sim, S. Y. A possible mechanism of destruction of coastal trees by tsunamis: A hydrodynamic study on effects of coastal steep hills. *Journal of Hydro-environment Research* **7**, 113–123 (2013).
23. Goda, Y. & Suzuki, Y. Estimation of Incident and Reflected Waves in Random Wave Experiments. in *International Conference on Coastal Engineering* (1976). doi:10.9753/icce.v15.47.
24. Koç, M., Balas, C. & Koç, D. Stability assessment of rubble-mound breakwaters using genetic programming. *Ocean Engineering* **111**, 8–12 (2016).
25. Keijzer, M., Baptist, M., Babovic, V. & Uthurburu, J. R. Determining equations for vegetation induced resistance using genetic programming. in *Genetic and Evolutionary Computation Conference, GECCO 2005, Proceedings, Washington Dc, Usa, June* 1999–2006 (2005). doi:10.1145/1068009.1068343.
26. Schmidt, M. & Lipson, H. Distilling Free-Form Natural Laws from Experimental Data. *Science* **324**, 81–85 (2009).
27. Yuan, F. & Mueller, T. Identifying models of dielectric breakdown strength from high-throughput data via genetic programming. *Sci Rep* **7**, 17594 (2017).
28. Suwa, R. Evaluation of the wave attenuation function of a coastal black pine Pinus thunbergii forest using the individual-based dynamic vegetation model SEIB-DGVM. *Journal of Forest Research* **18**, 238–245 (2013).
29. Tang, J., Zhao, C. & Shen, Y. Numerical investigation of the effects of coastal vegetation zone width on wave run-up attenuation. *Ocean Engineering* **189**, 106395 (2019).

Acknowledgements

The study is financially supported by the National Natural Science Foundation of China (Grant No. 51839002 & 51979014). Partial support also comes from the Hunan Education Department Scientific Research Projects of China (Grant No. 18A123).

Author contributions

Conceptualization: Shangpeng Gong, Jie Chen. Methodology: Shangpeng Gong, Fei He. Writing—original draft preparation: Shangpeng Gong. Writing—review and editing: Zhiyuan Wu. Supervision: Changbo Jiang, Zhiyuan Wu

Competing interests

The authors declare no competing interests.

A statement about the experimental study involving plants

Our experimental study did not involve real plants (either in the wild or in cultivation). We used artificial plant model for the experiments, and the parameters of the plant model were referred to journal papers (cited in the manuscript). The material of our plant model complies with relevant institutional, national, and international guidelines and legislation.

Additional information

Correspondence and requests for materials should be addressed to Jie Chen.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Figures

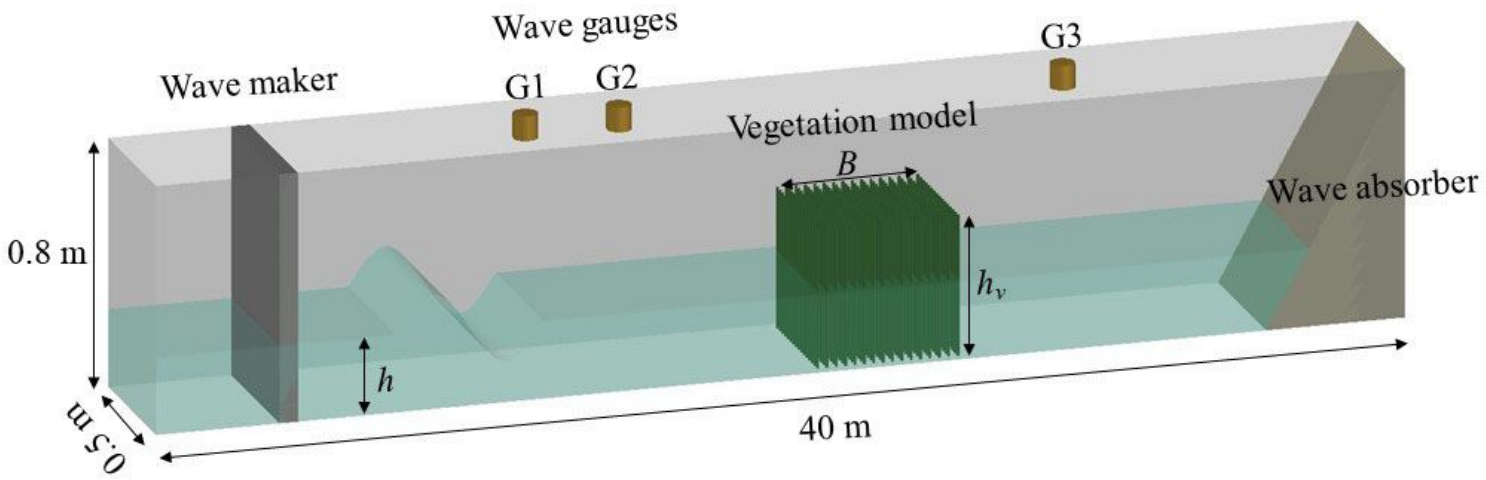


Figure 1

Experimental setup.

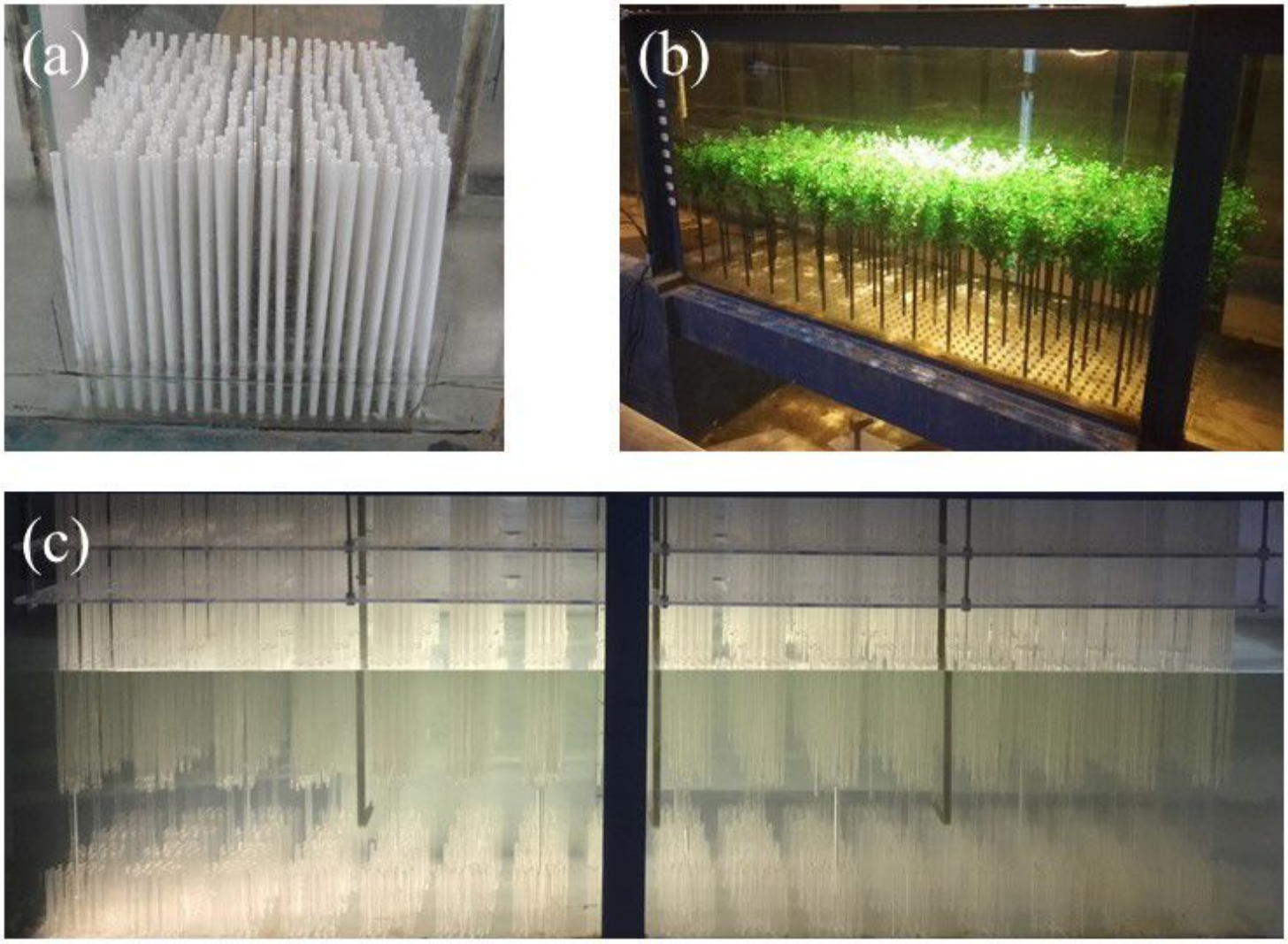


Figure 2

(a) Rigid cylinder model (Model 1), (b) Plastic synthetic vegetation model (Model 2), (c) Cylinder array model (Model 3).

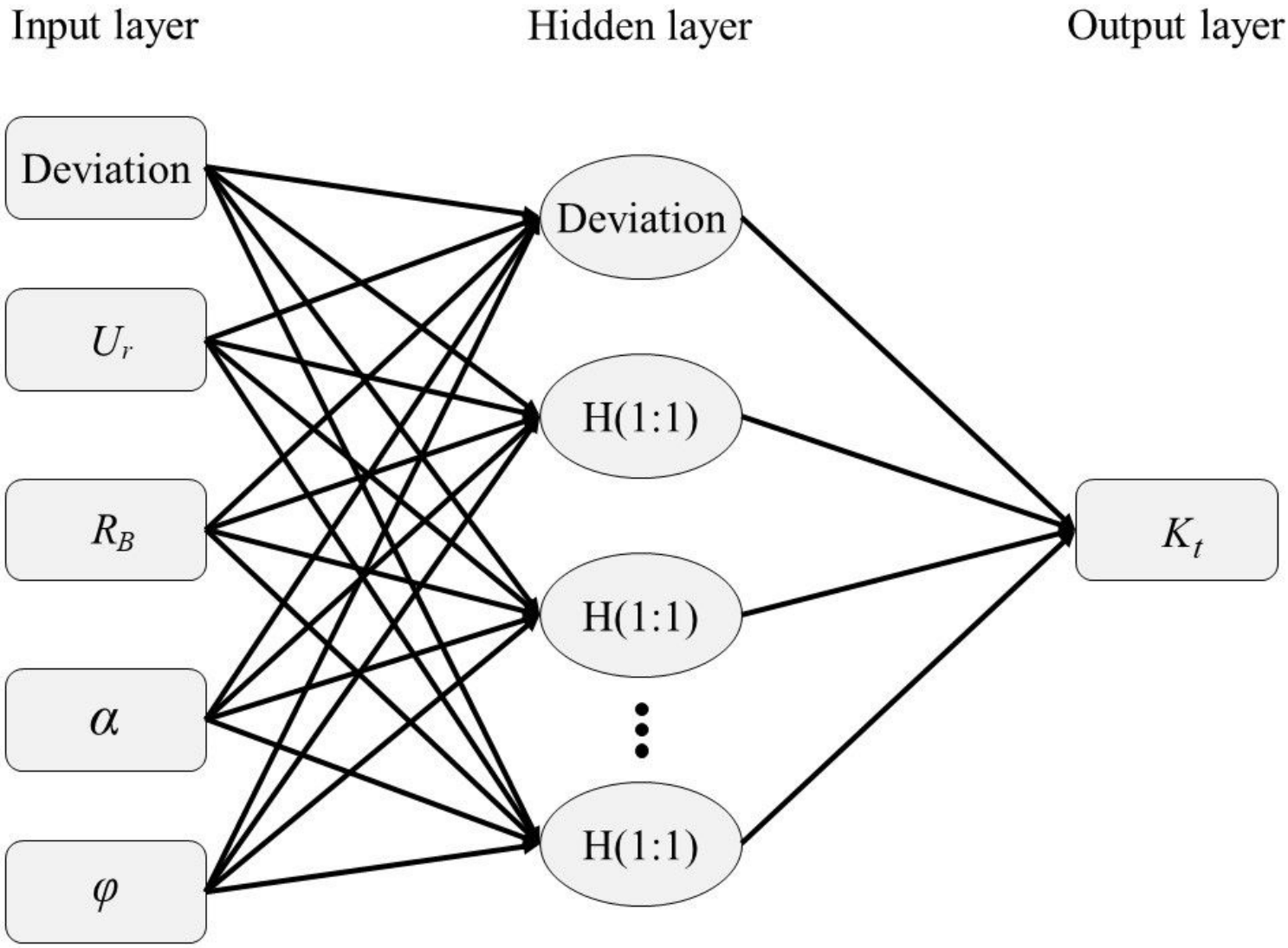


Figure 3

The ANNs structure used in this study.

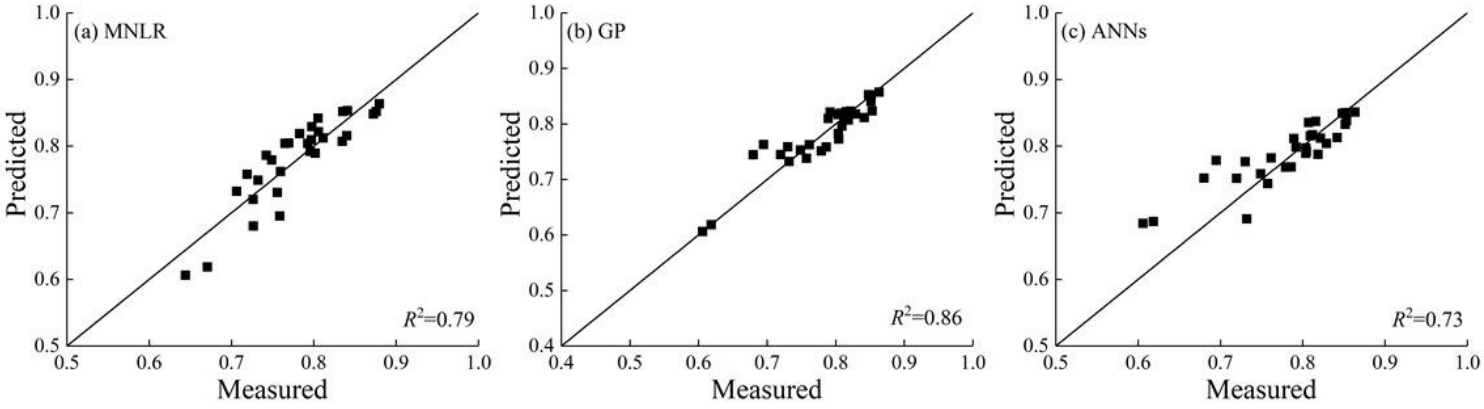


Figure 4

Comparison between the measured dataset from rigid cylinder model and predicted value using (a) non-linear regression, (b) GP and (c) ANNs.

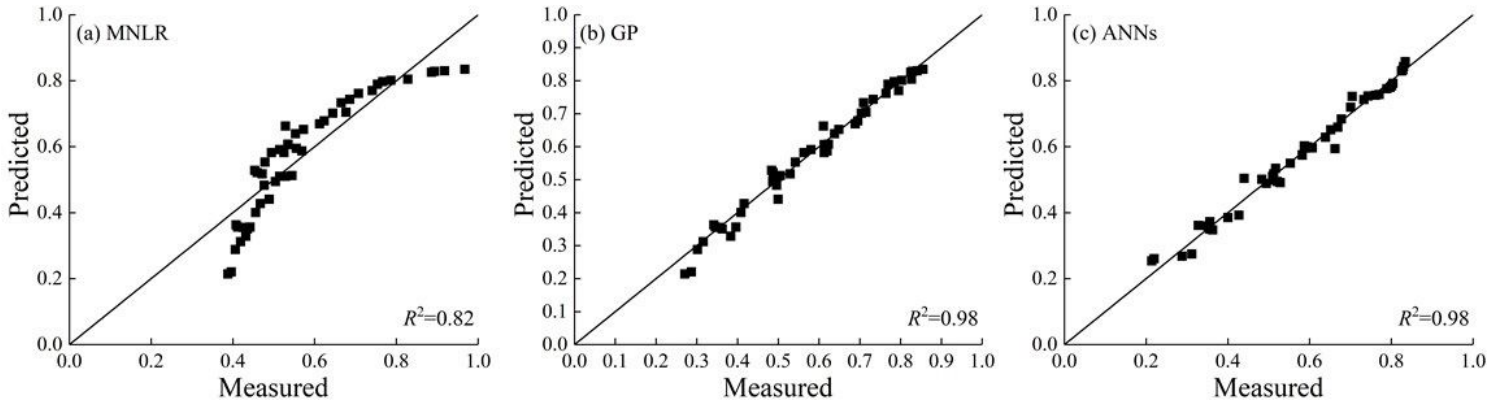


Figure 5

Comparison between the measured dataset from plastic synthetic vegetation model and predicted value using (a) MNLR, (b) GP and (c) ANNs.

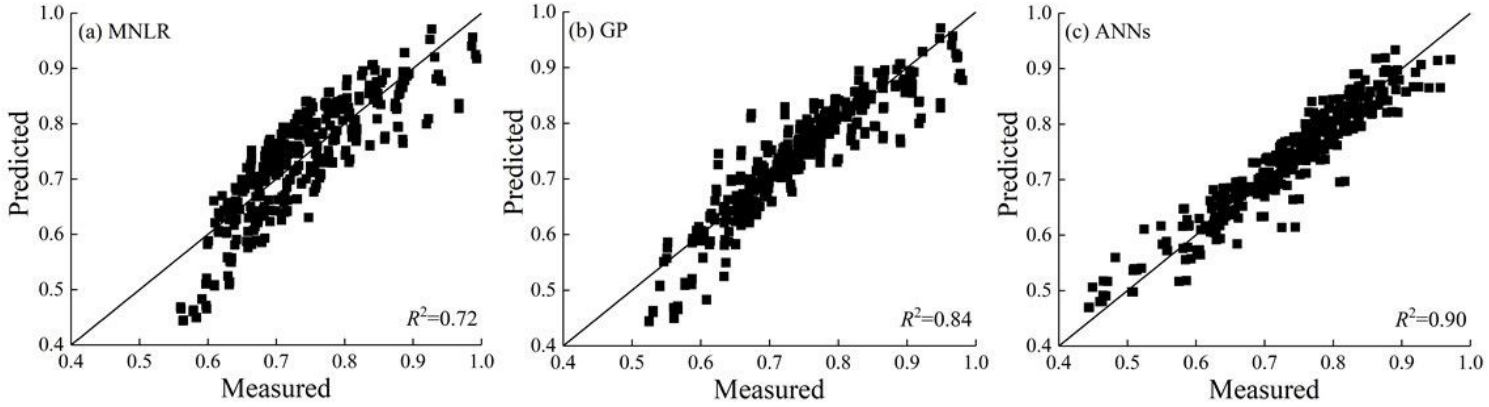


Figure 6

Comparison between the measured dataset from cylinder array model and predicted value using (a) MNLR, (b) GP and (c) ANNs.

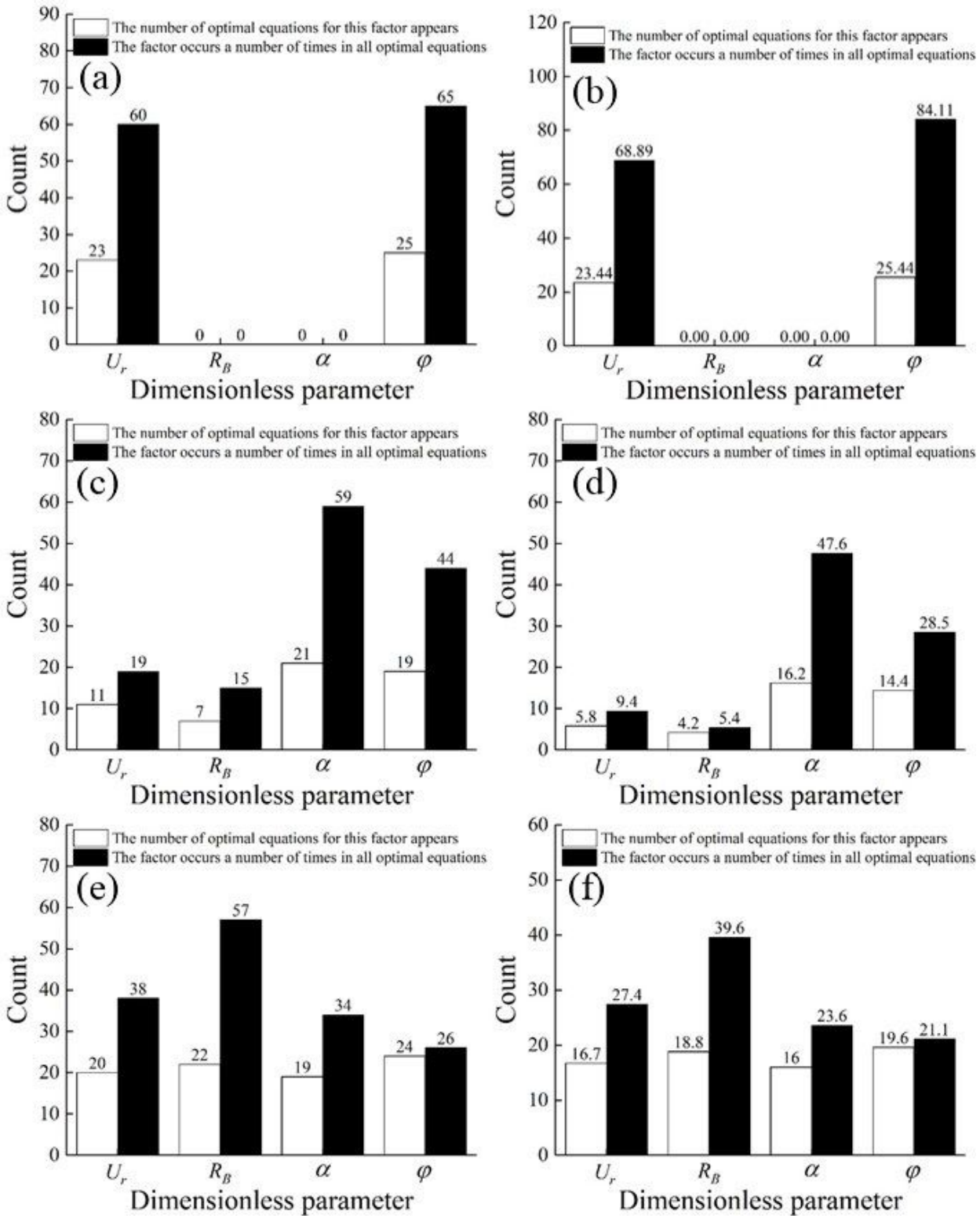


Figure 7

The mean values of each factor index were fitted for 10 times of the (a-b) Model 1, (c-d) Model 2 and (e-f) Model 3.