

# Multi-objective optimization of wave break forest design through machine learning

Jie Ren, Zengchuan Dong, Wei Xu, Qilai Zhang, Rensheng Shi, Hansong Zhu and Biao Sun

## ABSTRACT

Planting trees on a floodplain along a river is a practical and ecological method for embankment protection. Optimization of wave break forest is also a new concept on wave attenuation studies. In this study, we carried out physical experiments to obtain fundamental data and proposed the Cluster Structure Preserving Based on Dictionary Pair for Unsupervised Feature Weighting model (CDUFW) for multi-objective wave break forest design. Physical experiments were designed with considering the effects of different planting configurations on wave attenuation in three scenarios: (1) the equilateral triangle arrangement with different row spacings; (2) different arrangements with the same density; (3) different tree shapes with the same row spacing. The physical experiment condition was typically defined according to the field research of the study area. Then, a multi-objective weighting model for wave break forest design optimization was based on the scheme set of physical experiment outputs using the proposed CDUFW model. Physical experiments showed that different arrangement modes take advantage of the wave attenuation effect of different forest widths. The CDUFW model performed well in finding the effective, economic and reasonable scheme. The proposed model is excellent in data mining and classification, and can be applied to many decision-making and evaluation fields.

**Key words** | dictionary learning, physical experiment, spectral clustering, vegetation, wave attenuation

Jie Ren (corresponding author)

Zengchuan Dong

Wei Xu

Rensheng Shi

Hansong Zhu

Biao Sun

College of Hydrology and Water Resources, Hohai University,

No. 1, Xikang Road, Nanjing 210098,

China

E-mail: [jessieren2016@163.com](mailto:jessieren2016@163.com)

Qilai Zhang

School of Computer Science and Technology,

Tianjin University,

No. 135, Yaguan Road, Tianjin 300350,

China

Biao Sun

Water Resources Department of Heilongjiang

Province, No. 4, Wenzhong Road, Harbin 150001,

China

## INTRODUCTION

Many ancient civilizations were born on the banks of rivers. Humans are certain to rely on water resources, but the excessive use of water resources also raises concerns. People have been paying more attention to the protection of the ecological environment of rivers rather than unlimited demands. Since the 1950s, runoff restoration has received increasing attention and construction of ecological runoff has become an international trend. Concrete dams and artificial revetments, which may lead to the loss of biodiversity and productivity, have been demolished in many countries (Auerbach *et al.* 2014; Wohl *et al.* 2015; Huq *et al.* 2017) and replaced by environmentally friendly ecologically

specimen banks (ESBs) (Everaert *et al.* 2013; Küster *et al.* 2015; Movalli *et al.* 2017). Planting wave break forests on the floodplain is a type of ecological embankment which effectively combines wave prevention and ecosystem creation. It is generally believed that wave break forests should be planted in coastal zones. However, more and more studies focus on planting wave break forests on river floodplains, such as Huaihe River in China (Zhang & Song 2013), Noordwaard in the Netherlands (Reinout *et al.* 2010) and so on. Planting wave break forests on floodplains can not only reduce the destruction of levees by wind and waves and prevent levees from landslide and bank collapse,

but also protect soil from erosion, maintain the balance of land and water ecosystems, and improve the environment of levees (Duarte *et al.* 2013; Temmerman *et al.* 2013).

The study of wave break forests has been a great interest in many aspects and covers various topics, including hydrodynamics of vegetated channels (Nepf 2012; Kumar 2014; Kalra *et al.* 2017), the relationship among flow alteration, riparian vegetation and ecosystems (Shafroth *et al.* 2017; Datta 2018), and wave attenuation by vegetation (Knutson *et al.* 1982; Kobayashi *et al.* 1993; Nepf 2012; Silinski *et al.* 2018), etc. Wave attenuation by vegetation has been investigated through both physical experiments and numerical models (Reinout *et al.* 2010; Anderson & Smith 2014; Peruzzo *et al.* 2018). Most of them focus on individual characteristics such as vegetation width, vegetation density (i.e. the amount of vegetation per square meter), vegetation flexibility, or water depth. It has been widely accepted that wave attenuation by vegetation is related to plant characteristics (width, density, and stiffness) and hydrodynamic conditions (water depth, wave period, and wave height). Wave heights are exponentially attenuated in plant regions, and the fastest subtractions occur in the first few to 10 meters of the plant belt (Knutson *et al.* 1982). The effect of unsubmerged plants on wave propagation is greater than that of submerged plants because the flow velocity of the upper part of the water is very high, and the stem of the unsubmerged plant occupies the whole water body, which will have a great influence on the upper water body compared with the submerged plant (Augustin *et al.* 2009). The higher density of vegetation and higher plant stiffness leads to higher wave energy dissipation (Möller *et al.* 2014; Silinski *et al.* 2018).

The configurations of vegetation can also influence the attenuation effect (Maza *et al.* 2015). The equilateral triangle arrangement, which has been diffusely used according to our field investigation of Nenjiang River, has seldom been studied and compared with other configurations. The difference of these arrangements on the effect of wave attenuation needs further study. It is necessary to elucidate specific studies on different engineering construction areas rather than studying the mechanism through experiments (Alexander & Allan 2006; Everaert *et al.* 2013). Furthermore, the available land area and funds are different for each project and there is a contradiction among the available

investments, the available land area and the effect of the wave attenuation. Therefore, a specific optimization scheme of vegetation should be proposed for each project.

Compared with the subjective empowerment and the objective weighting method based on entropy theory, some machine learning theories, such as the spectral clustering and dictionary learning, may be more suitable for the wave break forest scheme optimization. There are few examples of the research on the layout of the wave break forest, hence, the subjective weighting method, such as expert weighting, may lead to great subjective arbitrariness. Moreover, the method of mutual information (MI) and its variants, as the development of information entropy, has been widely used in many fields (Alfonso *et al.* 2013; Vergara & Estévez 2014; Taormina & Chau 2015). The unsuitability of this method is that methods based on MI theory would have a preference on the attribute which has more possible values (Quinlan 1986) or which has fewer possible values (Quinlan 1993). However, the scheme set to be optimized is an incomplete scheme set of artificial design, and the numbers of values of different indicators are different. These methods may cause a large deviation from actuality. To circumvent the above problems, we introduced a novel method based on dictionary learning and spectral clustering. Dictionary learning can capture the intrinsic feature space of data by learning an over-complete dictionary (Zhu *et al.* 2016) and spectral clustering has been demonstrated (Zhao *et al.* 2010; Liu *et al.* 2016) as an effective approach to classify the data by exploring the clustering structure. It is widely accepted that the information from different views is relevant and complementary. Clustering structure information and intrinsic feature space characterize the data from different perspectives, respectively. To more accurately describe the data, we explore the complementary information between different views by combining spectral clustering and dictionary learning. Based on this motivation, we proposed the Cluster Structure Preserving Based on Dictionary Pair for Unsupervised Feature Weighting model (CDUFW).

The purpose of this paper is to put forward a method that can be used to optimize the design scheme of prodigal forest under multi-objective conditions. The physical experiments were projected on the Baidajie Dike in Nenjiang

River basin of China and the CDUFW model was applied to the optimization of wave break forest design of the basin.

## EXPERIMENT SETUP

Physical experiments were established to determine the influence of vegetation configuration and tree shape on wave attenuation. Experiments simulated the situation of waves striking the wave break forest along the transverse direction of the river. Experiment elements such as water depth, wave height, and tree dimensions were all scaled to the experiment channel based on the actual situation of the research area.

### Study area

Baidajie Dike, which is located at the mainstream of Nenjiang River, is the largest tributary of Songhua River. Nenjiang River ( $45^{\circ}27' - 51^{\circ}38' \text{ N}$ ,  $119^{\circ}52' - 126^{\circ}30' \text{ E}$ ) is the boundary river among Heilongjiang Province and Inner Mongolia Autonomous Region, Jilin Province in North-eastern China. It flows 1,370 km with a drainage area of  $2.97 \times 10^5 \text{ km}^2$ . The precipitation of Nenjiang River is mainly concentrated during June–September, accounting for 82% of the annual precipitation, which leads to a wide water surface, with the water level rising over the floodplain (Li et al. 2012). The mainstream of the Nenjiang River forks 10 km upstream of Baidajie Dike and joins together 10 km downstream of it. Baidajie Dike is on the left bank of the left tributary of Nenjiang River, named Tuoli River, in Tailai County, Qiqihar City. This segment of river meanders and zigzags with wetlands and yoke lakes on the floodplain and the average inclination of the slopes is 0.6‰. The flood prevention standard of Baidajie Dike was 30-year reoccurrence before 2016 and is expected to be raised to 50-year reoccurrence in the next few years. The sketch of Baidajie Dike is shown in Figure 1 and the hydrologic characteristics and wave features of Baidajie Dike are shown in Table 1. The water depth of floodplain in Table 1 is the average water depth of floodplain of the designed flood with 2% frequency. The average wave height and the average wave period were calculated by the Putian experimental station formula, which is recommended by Specification JTS145-2-2013 of China. By this specification, the one-tenth wave

height and the significant wave height were then calculated according to the average wave height and the average wave period. Poplars and willows are discretely planted on the floodplain along Baidajie Dike. The total length of vegetation along the direction of the river on the floodplain of Baidajie Dike is 2 km and the current vegetation area on the floodplain is 8–14 m wide. The spacing between trees is about 2.5 m, in equilateral triangle configuration according to our field investigation.

### Initial condition validation

All experiments were run under both regular wave condition and irregular wave condition. The water depth and average period in Table 1 were scaled and then taken as the initial input of the wave generator system. The wave heights at the location of sensor 1 (Figure 8(a) in Appendix A, available with the online version of this paper) were measured with the sampling frequency (100 Hz) to validate experiment initial conditions for both regular and irregular wave condition. The description and schematic of the experimental apparatus and setup are presented in Appendix A. The system automatically iterated and modified if the error rate of wave heights between the measured value and the set value in Table 1 exceeded 5%, otherwise the parameters were stored to generate waves. The elements of regular and irregular waves were counted in accordance with Specification JTS145-2-2013 of China. Runs of the regular wave were validated by one-tenth wave height  $H_{1/10}$ , which is the arithmetic mean of multiple wave heights, i.e.:

$$H_{1/10} = \frac{1}{N/10} \sum_{i=1}^{N/10} H_i \quad (1)$$

where  $H_i$  represents the  $i$ th measured wave height value. For irregular wave runs, the significant wave height is calculated according to the statistical data of the wave height, which is described by the following expression:

$$H_{1/3} = \frac{1}{N/3} \sum_{j=1}^{N/3} H_j \quad (2)$$

where  $H_j$  is the value of wave height counted by zero-up crossing method.

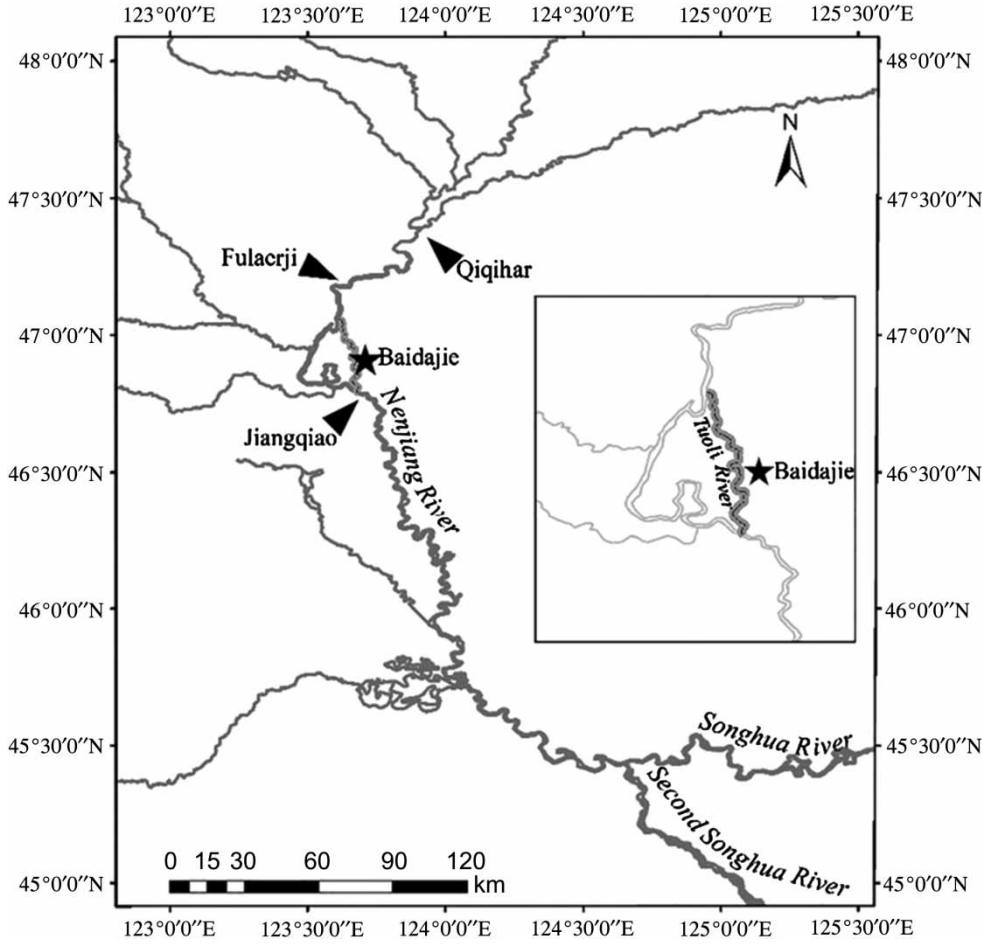


Figure 1 | Sketch of Baidajie Dike.

Table 1 | Hydrologic characteristics and wave features of 50 years recurrence in Baidajie

Water depth of floodplain $H(m)$	Average wave height $\bar{H}(m)$	One-tenth wave height $H_{\frac{1}{10}}(m)$	Significant wave height $H_s(m)$	Average period $T(s)$
2.83	0.57	1.157	0.91	3.01

Experiment scheme

After the incident wave condition was validated, the wave attenuation effects of different schemes were studied. Experiments were conducted to investigate the effects of row spacing, arrangement and tree shape on the wave attenuation of the wave break forest separately. Four kinds of simulation tree with different tree heights, trunk radiuses and crown radiuses were made based on the actual

condition of vegetation on the flood plain of Baidajie Dike, along with the consideration of the changes in plant growth (Table 2). Planting costs for each shape of tree are also listed in Table 2.

Experiment scenarios of different row spacings (Figure 2) were carried out with the width of vegetation

Table 2 | Parameters of different shapes of tree

Tree shape	Trunk height $H_T(m)$	Crown height $H_C(m)$	Trunk radius $R_T(m)$	Crown radius $R_C(m)$	Planting cost $C$ (CNY/individual plant)
T1	0.26	0.35	0.025	0.25	100
T2	0.26	0.35	0.015	0.25	50
T3	0.26	0.35	0.015	0.17	50
T4	0.16	0.08	0.007	0.13	3

ranging from 1 to 5 m to determine the wave attenuation effect of equilateral triangle arrangement. Trees of type T3 were taken as the experimental tree. If the difference of the two runs of a certain experiment scheme exceeded 5%, the third run would be taken and the average of the closest two values was taken as the final value.

Then, experiments of three different configurations (i.e. equilateral triangle, square and quincunx) of vegetation were carried out, taking T3 as the experimental tree too. The sketch of different configurations is shown in Figure 3. The space between two trees was 25 cm in equilateral triangle, square configurations and 30 cm in quincunx mode. Under these three configurations, 18, 16 and 18 trees can be planted in a square of  $10 \times 10$  cm. This makes the density of vegetation approximately equal under the three arrangements. Because rigid plants are much larger than flexible plants, it is almost impossible to make the density of rigid vegetation in different arrangements completely consistent as in the studies of flexible plants. Each kind of configuration with different widths of vegetation ranging from 1 to 5 m was measured to determine the difference among them. Each scheme was run twice and the average value was taken if the deviation of the two detected values was

within 5%. Another one run was taken if the deviation of the two detected value exceeded 5%.

The influence of different kinds of simulation tree was then experimented under the equilateral triangle configuration with a plant space of 25 cm. Different forest widths ranging from 1 to 5 m were also taken into consideration and at least two runs for each circumstance ensured the quality of the experiment.

The attenuation effects of all experiments were evaluated by wave attenuation coefficient  $K$ , defined as follows:

$$K = \frac{H_0 - H_v}{H_0} \times 100\% \quad (3)$$

where  $H_0$  is the wave height without vegetation and  $H_v$  is the wave height behind vegetation.

#### OPTIMIZATION MODEL BASED ON CDUFW

The physical experiments above provide a certain wave attenuation coefficient for a certain configuration with certain forest width and tree shape in each scheme. On the

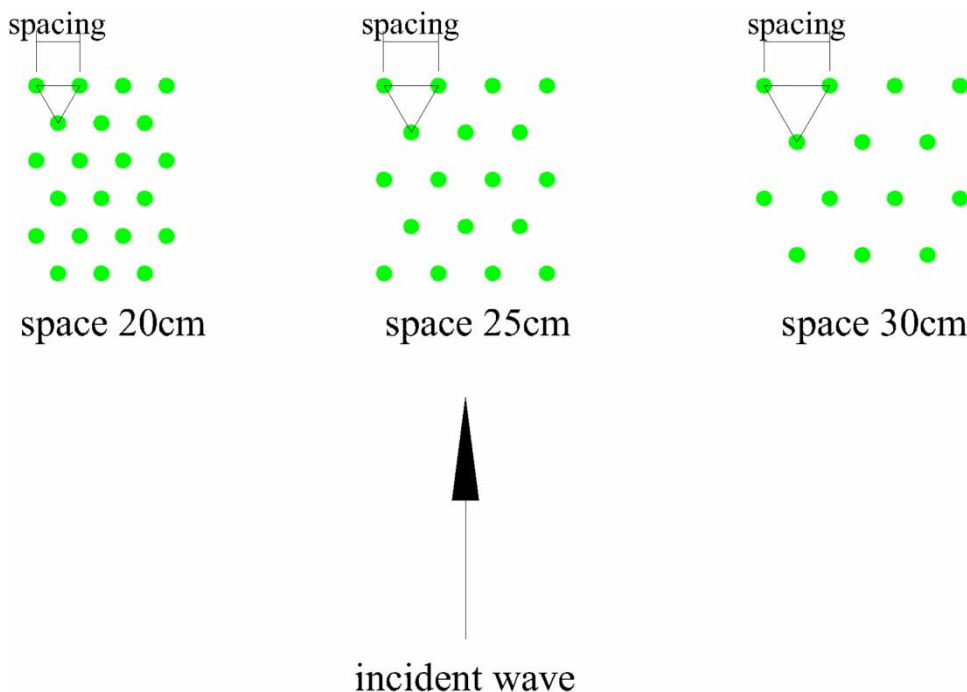


Figure 2 | Sketch of equilateral triangle arrangement with different spacings.

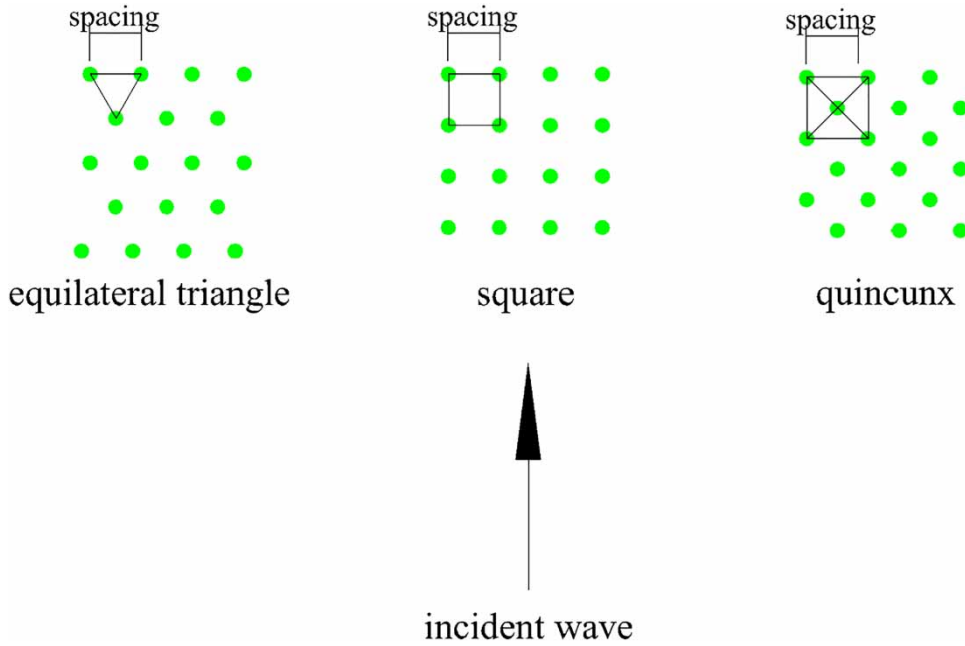


Figure 3 | Sketch of different configurations.

one hand, these schemes vary in investment and land use. On the other hand, the provided investment and the expected wave attenuation effect are different in each project. So, it is necessary to carry out a multi-objective optimization according to the actual situation. Multi-objective evaluation usually gives weight to different indexes first, and then gives each scheme a score by some weighted calculation method. An adaptive weighting model that learns the importance of data features based on dictionary learning and clustering is introduced in this section. This method can be used to obtain the importance of different factors, so as to select the optimal solution.

### Preliminaries of CDUFW

Matrices are written as bold capital letters and vectors are denoted by bold lowercase. Let  $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \in \mathbf{R}^{m \times n}$  be the data matrix with  $m$  features  $r$  samples.  $\|\mathbf{X}\|_F$  is the Frobenius norm of  $\mathbf{X}$ ,  $\text{Tr}(\mathbf{X})$  is the trace of  $\mathbf{X}$ .  $\mathbf{I}$  is the identity matrix. The F-norm is defined as:

$$\|\mathbf{X}\|_F = \sqrt{\text{Tr}(\mathbf{x}^T \mathbf{x})} = \sqrt{\sum_{i=1}^m \sum_{j=1}^n x_{ij}^2} \quad (4)$$

The goal of clustering is to partition the instances into  $k$  clusters  $\{\mathbf{P}_1, \mathbf{P}_2, \dots, \mathbf{P}_K\}$  and the clustering model under the framework of matrix factorization using the prevalent relaxation calculation method (Tang & Liu 2012) is as follows:

$$\begin{aligned} \min \|\mathbf{X} - \mathbf{U}\mathbf{F}^T\|^2 \\ \text{s.t. } \mathbf{F}^T \mathbf{F} = \mathbf{I}, \mathbf{F} \geq 0 \end{aligned} \quad (5)$$

where  $\mathbf{U} \in \mathbf{R}^{m \times k}$  is representation matrix.  $\mathbf{F} = [\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_n]^T \in \mathbf{R}^{m \times k}$  is the cluster indicator matrix and each entry  $f_{ij}$  in  $\mathbf{F}$  indicates whether  $\mathbf{x}_i \in \mathbf{P}_j$ . If  $\mathbf{x}_i \in \mathbf{P}_j$ ,  $f_{ij} = 1$ ; otherwise,  $f_{ij} = 0$ . The cluster indicator matrix  $\mathbf{F}$  reveals the discriminative information of data.

Compared with traditional matrix factorization, dictionary learning has less reconstruction error and the data can be better reconstructed due to the overcomplete dictionary. From the viewpoint of matrix reconstruction, dictionary learning was used to replace matrix factorization:

$$\min \|\mathbf{X} - \mathbf{D}\mathbf{A}\|_F^2 + \alpha \|\mathbf{A}\| \quad (6)$$

where  $\mathbf{D}$  is the dictionary and  $\mathbf{A}$  is the intrinsic feature matrix (also called representation coefficients matrix),  $\alpha \|\mathbf{A}\|$  is the regularization item imposed on  $\mathbf{A}$ .

According to the principle of clustering, similar samples have similar data structures, while the dissimilar samples should be divided into different clusters. Thus, the cluster structure information can be converted into the following optimization problem:

$$\begin{aligned} & \min_{\mathbf{F}} \text{Tr}(\mathbf{F}^T \mathbf{L} \mathbf{F}) \\ & \text{s.t. } \mathbf{F}^T \mathbf{F} = \mathbf{I} \end{aligned} \quad (7)$$

where  $\mathbf{L} \in \mathbf{R}^{n \times n}$  is the Laplacian matrix and  $\mathbf{L} = \mathbf{D} - \mathbf{S}$ .  $\mathbf{S} \in \mathbf{R}^{n \times n}$  is the instance similarity matrix learned by the kernel function.

### Algorithms of CDUFW

Clustering structure can indicate affiliation relations of samples and dictionary learning can reflect the data distribution with fewer constraints. The label space and sample space are different descriptions of the same data set. The coupled dictionary learning separately based on samples and pseudo-labels generated by spectral clustering was used to explore the intrinsic data information. In the coupled dictionary learning, it was expected to minimize the error between the two dictionary learnings continuously. Then, the projection matrix can be obtained, which reveals the importance of each feature. Thereunder, a novel unsupervised feature empowerment algorithm CDUFW was proposed.

As mentioned above, a data matrix is written as  $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \in \mathbf{R}^{m \times n}$ , the proposed method is formulated as:

$$\begin{aligned} & \min_{\mathbf{D}_1, \mathbf{D}_2, \mathbf{A}} \|\mathbf{X} - \mathbf{D}_1 \mathbf{A}\|_F^2 + \|\mathbf{F}^T - \mathbf{D}_2 \mathbf{A}\|_F^2 + \alpha \|\mathbf{A}\| + \beta \text{Tr}(\mathbf{F}^T \mathbf{L} \mathbf{F}) \\ & \text{s.t. } \mathbf{F}^T \mathbf{L} \mathbf{F} = \mathbf{I}, \mathbf{F} \geq 0 \end{aligned} \quad (8)$$

where  $\mathbf{F} \in \mathbf{R}^{n \times k}$  is the cluster indicator matrix,  $\mathbf{L} \in \mathbf{R}^{n \times n}$  is the Laplacian matrix,  $\mathbf{A} \in \mathbf{R}^{p \times n}$  is the intrinsic feature matrix,  $p$  is the size of intrinsic feature space.  $\mathbf{D}_1 \in \mathbf{R}^{m \times p}$  and  $\mathbf{D}_2 \in \mathbf{R}^{k \times p}$  are representation dictionaries to describe the data distribution where  $k$  is the number of clusters the instances are partitioned into. It can be seen that the model developed by Equation (8) consists of two parts, including dictionary

learning  $\|\mathbf{x} - \mathbf{D}_1 \mathbf{A}\|_F^2 + \|\mathbf{F}^T - \mathbf{D}_2 \mathbf{A}\|_F^2 + \alpha \|\mathbf{A}\|$  and spectral clustering  $\beta \text{Tr}(\mathbf{F}^T \mathbf{L} \mathbf{F})$ .

There are four variables,  $\mathbf{D}_1$ ,  $\mathbf{D}_2$ ,  $\mathbf{A}$ ,  $\mathbf{F}$ , involved in Equation (8) to be optimized. A selective minimization strategy was used to solve this problem. To ensure that the objective function in Equation (6) was convex, only one variable was updated at one time and others were fixed. Then, the updating process is iterative until the model converges. The iteration steps are shown in Appendix B (available with the online version of this paper).

After the four variables were updated and the optimized solution of Equation (8) was obtained, the projection matrix  $\mathbf{E}$  which reveals the importance of each feature was calculated as follows:

$$\mathbf{E} = (\mathbf{A} \mathbf{X}^T (\mathbf{X} \mathbf{X}^T + \varepsilon \mathbf{I}))^T \quad (9)$$

where  $\varepsilon$  is a small-valued variable to avoid anomalies in the inversion of the matrix and is defined as  $10^{-6}$ . The weight of each feature can then be defined as:

$$\begin{aligned} w &= (w_1, w_2, \dots, w_n) \\ w_i &= \frac{\|e_i\|_2}{\sum_{i=1}^m \|e_i\|_2} \end{aligned} \quad (10)$$

where  $e_i$  is the  $i$ -th row of  $\mathbf{E}$ .

The details about CDUFW above are summarized as follows:

**Algorithm 1** Cluster Structure Preserving Dictionary Pair for Unsupervised Weighting (CDUFW).

**Input:** A data matrix  $\mathbf{X}$ , the size of intrinsic feature space  $p$ , loss variation ratio  $\sigma$ , the maximum iteration  $N$  and the parameters  $\alpha, \beta$

**Output:** The weight array  $\mathbf{w}$ ;

- 1: Initialize all variables.
- 2: **do**
- 3: Update  $\mathbf{D}_1, \mathbf{D}_2$ .
- 4: Update  $\mathbf{A}$ .
- 5: Update  $\mathbf{F}$ .
- 6: **until** up to the maximum iterations or loss variation ratio
- 7: calculate  $\mathbf{E}$ .
- 8: calculate  $\mathbf{w}$ .
- 9: **return**  $\mathbf{w}$ .

The convergence of each variable guarantees the convergence of the objective function, which has been intensively studied (Gorski et al. 2007). The time complexity of the proposed function (Equation (6)) was biquadratic.

**Pre-process and post-process of the CDUFW**

The magnitude of values under different factors varies greatly. Therefore, the raw data obtained by experiment and calculation need to be processed to form the data matrix before being brought into the CDUFW model. The membership functions were used to normalize the raw data under each indicator in the pre-process. The pre-process steps are as follows:

- (1) Determine the features of interest and the original data matrix **O**.

As was mentioned above, the decision makers should consider the effect of wave elimination effect, economic input and land investment to choose a certain scheme of wave break forest. These can be transformed into three features: wave attenuation coefficient *K*, the width of forest and the cost (being the number of plants multiplied by the cost of planting per tree). Thus, the original data matrix can be formed as follows:

$$O = \begin{bmatrix} o_{11} & o_{12} & \dots & o_{1n} \\ o_{21} & o_{22} & \dots & o_{2n} \\ o_{31} & o_{32} & \dots & o_{3n} \end{bmatrix} \tag{11}$$

where  $o_{ij}$  is the value of the scheme  $j$  to feature  $i$ .

- (2) Calculate relation degree and form the data matrix **X**.

The wave attenuation coefficient belongs to the benefit type while the others belong to the cost type. The benefit type means that it is expected to be larger, while the cost type indicates that the feature value is expected to be lower. The calculation formula of relation degree is shown in Equations (12) and (13) for benefit type and cost type, respectively:

The benefit type:  $x_{ij} = (o_{ij} - o_{i\min}) / (o_{i\max} - o_{i\min})$  (12)

The cost type:  $x_{ij} = (o_{i\max} - o_{ij}) / (o_{i\max} - o_{i\min})$  (13)

where  $o_{ij}$  is the relative membership of scheme  $j$  to the optimal value of the feature  $i$ ,  $O_{i\min}$  is the minimum of eigenvalues of all schemes to feature  $i$ ,  $O_{i\max}$  is the maximum of values of all schemes to feature  $i$ . Then the data matrix **X** can be obtained:

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ x_{31} & x_{32} & \dots & x_{3n} \end{bmatrix} \tag{14}$$

After the data matrix **X** was obtained, the weight of each feature can be worked out by CDUFW with Algorithm 1 mentioned above, then post-processes were needed to choose a better scheme. The distance between scheme and best goal along with the distance between scheme and worst goal were taken as a criterion to evaluate each scheme. The post-process steps are as follows:

- (1) Define the relative greatest scheme and the relative worst scheme.

In the multi-objective optimization, the comparison is limited to the chosen  $n$  schemes, so the best and the worst are relative concepts. In the alternative  $n$  schemes, the relative best scheme **G** and relative worst scheme **B** were defined according to their relation degree (Zhou et al. 2007):

$$G = (g_1, g_2, g_3)^T, g_i = \bigcup_{j=1}^n x_{ij} \tag{15}$$

$$B = (b_1, b_2, b_3)^T, b_i = \bigcap_{j=1}^n x_{ij}$$

- (2) Calculate relative optimum membership degree of each scheme.

Relative optimum membership of scheme  $j$  is defined by weighted Euclidean distance and least square method:

$$c_j = \left\{ 1 + \frac{\sum_{i=1}^3 [w_i (g_i - x_{ij})]^2}{\sum_{i=1}^3 [w_i (x_{ij} - b_i)]^2} \right\}^{-1} \tag{16}$$

The higher the value of  $c_j$ , the better the scheme. The scheme with the highest value of  $c_j$  was selected as the best one.



## RESULTS AND DISCUSSION

### Results of physical experiments

In the section regarding physical experiment, three factors, including space of equilateral triangle arrangement, different arrangements and different tree shapes, were taken into consideration. The results of these three factors are shown in Figures 4–6 respectively.

As shown in Figures 4–6, the trend of experimental results under regular wave conditions and irregular wave conditions are consistent with the increase of forest width, and in all experiments the wave attenuation coefficient increases with the increase of forest width, which is in line with previous research results (Reinout et al. 2010; Hoque et al. 2018).

Considering its extensive use in practical engineering, the equilateral triangle configuration was especially studied (Figure 4). The wave reduction effect decreases with the increase of plant spacing, which is consistent with the

conclusion that the wave elimination effect increases with the increase of vegetation density in previous studies (Möller et al. 2014; Peruzzo et al. 2018). However, in practical engineering, the use of distance is more easily accepted by workers and is convenient for construction.

The comparison of three different arrangements (Figure 5) shows that the square configuration which is commonly used in the experiment appeared to have a sharp turn after 3 m, and the increase of the wave attenuation coefficient decreased. It could be due to the less projected area of the trees in the direction of the incoming wave of the square configuration than others. It is apparent from Figure 3 that the projection area of the trunks and crowns of the quincunx and equilateral triangle arrangement is twice that of the square arrangement. It appears that the wave attenuation effect of the quincunx arrangement is better than that of the equilateral triangle arrangement, but the slope of the equilateral triangle curve is greater than that of the quincunx by 3 m. Thus, it could be speculated that

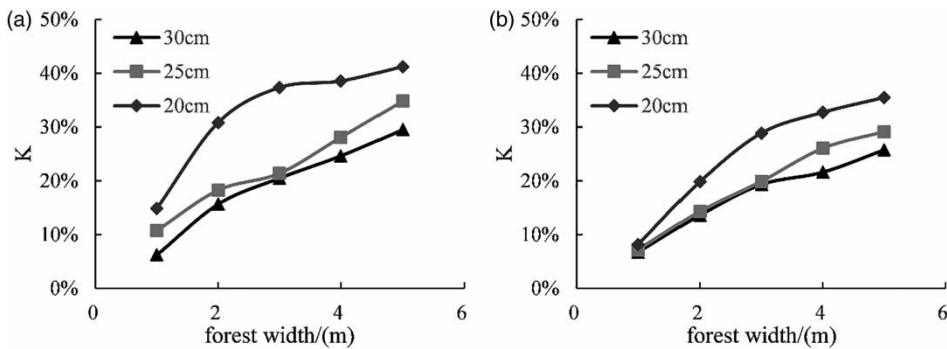


Figure 4 | Wave attenuation coefficient for three spacing of equilateral triangle arrangement and the five forest widths under regular wave (a) and irregular wave (b).

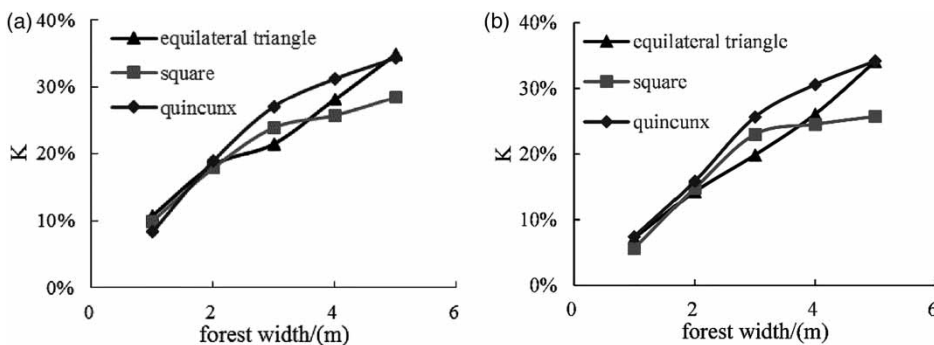
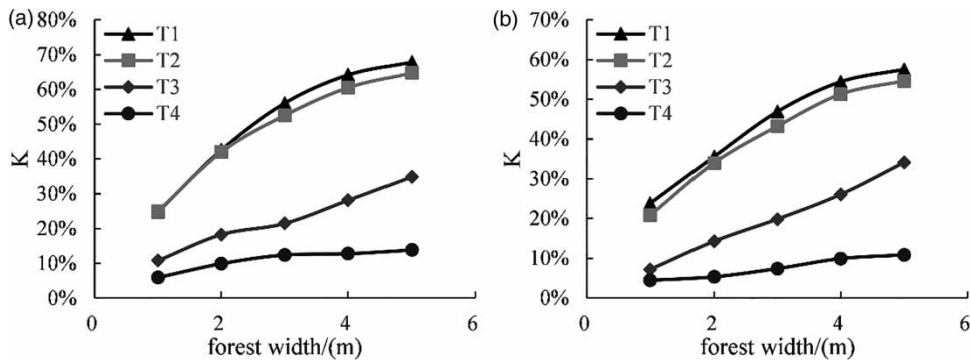


Figure 5 | Wave attenuation coefficient for three arrangement and the five forest widths under regular wave (a) and irregular wave (b).



**Figure 6** | Wave attenuation coefficient  $K$  for four kinds of tree shapes and the five forest widths under regular wave (a) and irregular wave (b).

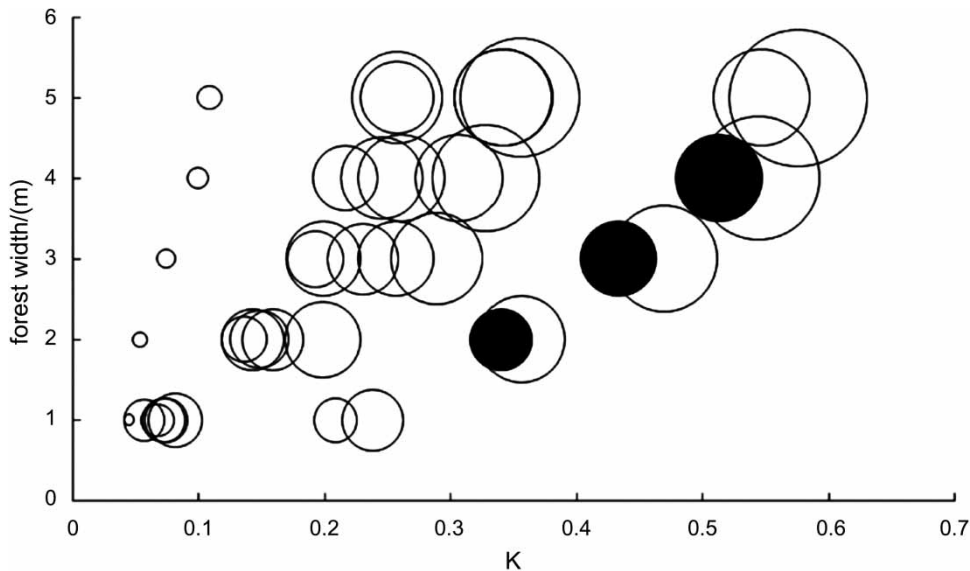
the wave attenuation effect of the equilateral triangle arrangement is the best when the forest width is more than 5 m on the laboratory scale (i.e. 50 m in reality). The stationary point of the curve of the square and quincunc arrangement appears at the forest width of 3 m while the curve of the equilateral triangle arrangement has not yet appeared a stationary point, so more experiments are needed for future study.

According to Figure 6, it is evident that the wave dissipating effect differs in tree shapes. Both wave attenuation coefficients of T1 and T2 are much higher than those of T3 and T4. If we divide these four kinds of trees into two categories, one is T1 and T2, the other is T3 and T4, the main difference between the two is the crown radius on the basis of Table 2. It is conceivable that under the same water level, the principal influence factor of wave attenuation effect is the crown radius of the tree. The difference between T1 and T2 is rather small, and it could be speculated that the contribution of trunk radius to wave attenuation is very small. The diameters of the tree trunk and tree crown are consistent with the area per unit height of each vegetation stand normal to wave direction as described in a previous study (Kobayashi et al. 1993). It can be found from Figure 6 that the effect of tree trunks and canopy on the wave attenuation varies greatly, so it is not recommended to simplify rigid plants to uniformly rigid cylinders. When the rigid plant density is defined as the number of trees per unit area, special consideration should be given to changes in crown radius. Moreover, it is interesting that the initial difference between T3 and T4 is small, but as the width of forest increases, the gap became larger, which is due to

the fact that the wave attenuation effect of T4 has hardly increased along the course. The total height of the tree of T4 is 0.24 m while the water depth of this experiment is 0.28 m, which means trees of type T4 are completely submerged under water. It can be inferred that when the vegetation is completely submerged, the wave effect rarely increases with the increase of forest width. This phenomenon is in accordance with previous studies by Augustin et al. (2009). Generally speaking, the branches and stems produce turbulence and the vortex in the stems draws energy from the average flow (Nepf 2011). The orbital velocity of water particles in the water column is gradually decreasing from the surface of water to the ground. When the vegetation is completely submerged, it cannot impede the top portion of water, thus the wave energy decrease is lower.

## Results of CDUFW

To make the decision more realistic, all valid physical experiment results under irregular wave condition were chosen to form a design set. The design set consisting of 40 schemes is presented in the form of a bubble diagram (Figure 7). This diagram comprehensively reflects the three aspects of each scheme. It expresses the cost of a scheme by its bubble size, the wave attenuation coefficient by the horizontal axis, and the forest width by the vertical axis. It is obvious that several schemes with different costs and forest widths can achieve similar wave attenuation effects, and it can be inferred that schemes represented by solid black bubbles are better than the hollow circles.



**Figure 7** | Bubble diagram of design set for the wave break forest where the schemes represented by solid bubbles are better than hollow bubbles.

A bubble diagram (Figure 7) consisting of the 40 schemes was built to verify the rationality of the results. The matrix generated by the design set was taken as the original data matrix  $\mathbf{O}$  and the calculation was carried on as described above under 'Algorithms of CDUFW'. The specific steps are shown in Appendix B. It is calculated that the weight of forest width, cost and wave attenuation coefficient was 0.3892, 0.3298 and 0.2810, respectively. The results showed that when making decisions on the planting layout of the wave break forest, it seems that we should not persist in the pursuit of excessive wave attenuation coefficient, which will lead to a great increase in planting area and planting cost. Schemes with accredited attenuation coefficient and lower cost were preferred. The best scheme chosen by the proposed model is T2 of an equilateral triangle with a space of 0.25 m arrangement in 2 m forest width. The chosen scheme is exactly one of these three black solid bubbles whose wave attenuation coefficient is 0.3417. The best scheme chosen by the model is exactly one of the four possible optimal schemes we inferred from Figure 7, which provides compelling evidence that the proposed model is valid and scientific. As mentioned earlier, the scheme set is just based on the experiment we have already carried out, so the best is a relative concept here. With the deepening of the research on the mechanism of wave breaking, more diverse and more efficient schemes

can be proposed. Even if the number of schemes is increased and the number of indicators considered increases, this model can still give reasonable choices.

## CONCLUSIONS

To study the influence of different factors on the wave attenuation effect in a practical engineering background, we carried out physical experiments. In total, 40 kinds of schemes were carried out and wave attenuation effect was evaluated by wave attenuation coefficient  $K$ . The laboratory study focused on vegetation effect under equilateral triangle configuration, a vegetation effect comparison among three different configurations and a comparison among four different tree shapes. The trend of wave reduction of equilateral triangle arrangement and the wave effect changes under different densities are similar to that of other permutations, but different arrangement modes take advantage on the wave attenuation effect of different forest widths. The effects of water depth, canopy and tree trunk on the wave attenuation effect of vegetation have been verified in the experiment of different tree shapes.

At the aim of choosing the best one in various wave break forest layout design schemes, an objective weighting model (CDUFW) was proposed, which is the first study in

the field of wave break forest to our knowledge. The dictionary learning and spectral clustering have been combined to reflect the data structure information and  $l_{2,1}$ -norm was used to quantify the significance of features. The proposed can overcome the mentioned defects of the entropy weighting method. It is effective even when there is no correlation between indicators. This model was applied to the data set of physical experimental data and obtains a reasonable optimization result. In future research, a more comprehensive data set can be acquired as input to the model so that better optimization schemes can be obtained.

The experimental design in this study was more closely related to engineering practice, which is of great significance to the design of river vegetation in other regions. The optimization model of vegetation design scheme is also proposed according to the actual demand, so as to achieve the objective effect of wave dissipation effectively, which can be used for reference by other river management decision makers. The CDUFW model is superior in its ability of intrinsic information mining and its convergence and computational complexity, and it can be applied in other multi-objective decision-making fields of water resources management.

## ACKNOWLEDGEMENTS

This work was supported by the Applied Technology Research and Development Program of Heilongjiang Province under Grant No. GZ16B031 and No. GZ16B035.

## REFERENCES

- Alexander, G. G. & Allan, J. D. 2006 Stream restoration in the Upper Midwest, U.S.A. *Restor. Ecol.* **14**, 595–604.
- Alfonso, L., He, L., Lobbrecht, A. & Price, R. 2013 Information theory applied to evaluate the discharge monitoring network of the Magdalena river. *J. Hydroinform.* **15** (1), 211–228.
- Anderson, M. E. & Smith, J. M. 2014 Wave attenuation by flexible, idealized salt marsh vegetation. *Coast. Eng.* **83**, 82–92.
- Auerbach, D. A., Deisenroth, D. B., McShane, R. R., McCluney, K. E. & Poff, L. R. 2014 Beyond the concrete: accounting for ecosystem services from free-flowing rivers. *Ecosyst. Svcs.* **10**, 1–5.
- Augustin, L. N., Irish, J. L. & Lynett, P. 2009 Laboratory and numerical studies of wave damping by emergent and near-emergent wetland vegetation. *Coast. Eng.* **56** (3), 332–340.
- Datta, D. 2018 Assessment of mangrove management alternatives in village-fringe forests of Indian Sunderbans: resilient initiatives or short-term nature exploitations? *Wetlands Ecol. Manage.* **26** (3), 399–413.
- Duarte, C. M., Losada, I. J., Hendriks, I. E., Mazarrasa, I. & Marbà, N. 2013 The role of coastal plant communities for climate change mitigation and adaptation. *Nat. Clim. Change* **3** (11), 961–968.
- Everaert, G., Pauwels, I. S., Boets, P., Verduin, E., Haye, M. A. A. D., Blom, C. & Goethals, P. L. M. 2013 Model-based evaluation of ecological bank design and management in the scope of the European water framework directive. *Ecol. Eng.* **53** (3), 144–152.
- Gorski, J., Pfeuffer, F. & Klamroth, K. 2007 Biconvex sets and optimization with biconvex functions: a survey and extensions. *Math. Methods Oper. Res.* **66** (3), 373–407.
- Hoque, A., Husrin, S. & Oumeraci, H. 2018 Laboratory studies of wave attenuation by coastal forest under storm surge. *Coast. Eng. J.* **60** (2), 225–238.
- Huq, N., Bruns, A., Ribbe, L. & Huq, S. 2017 Mainstreaming ecosystem services based climate change adaptation (EBA) in Bangladesh: status, challenges and opportunities. *Sustainability* **9** (6), 926.
- Kalra, T. S., Aretxabaleta, A., Seshadri, P., Ganju, N. K. & Beudin, A. 2017 Sensitivity analysis of a coupled hydrodynamic-vegetation model using the effectively subsampled quadratures method. *Geosci. Model Dev. Discuss.* **10** (12), 1–28.
- Knutson, P. L., Brochu, R. A., Seelig, W. N. & Inskeep, M. 1982 Wave damping in *Spartina alterniflora* marshes. *Wetlands* **2** (1), 87–104.
- Kobayashi, N., Raichle, A. W. & Asano, T. 1993 Wave attenuation by vegetation. *J. Waterway Port Coast. Ocean Eng.* **119** (1), 30–48.
- Kumar, B. 2014 Flow prediction in vegetative channel using hybrid artificial neural network approach. *J. Hydroinform.* **16** (4), 839–849.
- Küster, A., Becker, P. R., Kucklick, J. R., Pugh, R. S. & Koschorreck, J. 2015 The international environmental specimen banks – let's get visible. *Environ. Sci. Pollut. Res.* **22** (3), 1559–1561.
- Li, H., Xie, M. & Jiang, S. 2012 Recognition method for mid- to long-term runoff forecasting factors based on global sensitivity analysis in the Nenjiang river basin. *Hydrol. Process.* **26** (18), 2827–2837.
- Liu, H., Shao, M. & Fu, Y. 2016 Consensus guided unsupervised feature selection. In: *Thirtieth AAAI Conference on Artificial Intelligence*, February 12–17, Phoenix, Arizona, USA. AAAI Press, Phoenix, Arizona, pp. 1874–1880.
- Maza, M., Lara, J. L. & Losada, I. J. 2015 Tsunami wave interaction with mangrove forests: a 3-d numerical approach. *Coast. Eng.* **98**, 33–54.

- Möller, I., Kudella, M., Rupprecht, F., Spencer, T., Paul, M., Wesenbeeck, B. K. V., Wolters, G. & Kai, J. 2014 [Wave attenuation over coastal salt marshes under storm surge conditions](#). *Nat. Geosci.* **7** (10), 727–731.
- Movalli, P., Dekker, R., Koschorreck, J. & Treu, G. 2017 [Bringing together raptor collections in Europe for contaminant research and monitoring in relation to chemicals regulations](#). *Environ. Sci. Pollut. Res.* **24** (31), 24057–24060.
- Nepf, H. M. 2011 [Flow over and through biota](#). In: *Treatise on Estuarine and Coastal Science* (E. Wolanski & D. McLusky, eds). Academic Press, Waltham, MA, pp. 267–288.
- Nepf, H. M. 2012 [Hydrodynamics of vegetated channels](#). *J. Hydraul. Res.* **50** (3), 262–279.
- Peruzzo, P., Serio, F. D., Defina, A. & Mossa, M. 2018 [Wave height attenuation and flow resistance due to emergent or near-emergent vegetation](#). *Water* **10** (4), 402.
- Quinlan, J. R. 1986 [Induction of decision trees](#). *Mach. Learn.* **1** (1), 81–106.
- Quinlan, J. R. 1993 *C4.5: Programs for Machine Learning*. Morgan Kaufmann, San Mateo, CA.
- Reinout, O. D., Augustijn, D., Wijnberg, K., Dekker, F. & Mindert, V. D. 2010 [Modelling Wave Attenuation by Vegetation with Swan-veg; Model Evaluation and Application to the Noordwaard Polder](#). Master thesis, Water Engineering and Management, University of Twente, Delft.
- Shafroth, P. B., Schlatter, K. J., Gomez-Sapiens, M., Lundgren, E., Grabau, M. R., Ramírez-Hernández, J., Rodríguez-Burgueño, E. & Flessa, K. W. 2017 [A large-scale environmental flow experiment for riparian restoration in the Colorado river delta](#). *Ecol. Eng.* **106**, 645–660.
- Silinski, A., Schoutens, K., Puijalon, S., Schoelynck, J., Luyckx, D., Troch, P., Meire, P. & Temmerman, S. 2018 [Coping with waves: plasticity in tidal marsh plants as self-adapting coastal ecosystem engineers](#). *Limnol. Oceanogr.* **63** (2), 799–815.
- Tang, J. & Liu, H. 2012 [Unsupervised feature selection for linked social media data](#). In: *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Association for Computing Machinery, Beijing, China, pp. 904–912.
- Taormina, R. & Chau, K. W. 2015 [Data-driven input variable selection for rainfall-runoff modeling using binary-coded particle swarm optimization and extreme learning machines](#). *J. Hydrol.* **529**, 1617–1632.
- Temmerman, S., Meire, P., Bouma, T. J., Herman, P. M., Ysebaert, T. & De Vriend, H. J. 2013 [Ecosystem-based coastal defence in the face of global change](#). *Nature* **504** (7478), 79–83.
- Vergara, J. R. & Estévez, P. A. 2014 [A review of feature selection methods based on mutual information](#). *Neural Comput. Appl.* **24** (1), 175–186.
- Wohl, E., Bledsoe, B. P., Jacobson, R. B., Poff, L. R., Rathburn, S. L., Walters, D. M. & Wilcox, A. C. 2015 [The natural sediment regime in rivers: broadening the foundation for ecosystem management](#). *Bioscience* **65** (4), 358–371.
- Zhang, M. & Song, Z. 2013 [Calculation on wave dissipation performance for different wave break forest structures](#). *Adv. Sci. Technol. Water Resour.* **33** (6), 40–43.
- Zhao, Z., Wang, L. & Liu, H. 2010 [Efficient spectral feature selection with minimum redundancy](#). In: *Twenty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2010*. AAAI Press, Atlanta, Georgia, USA, July 11–15.
- Zhou, H. C., Zhang, G. H. & Wang, G. L. 2007 [Multi-objective decision making approach based on entropy weights for reservoir flood control operation](#). *J. Hydraul. Eng.* **38** (1), 100–106.
- Zhu, P., Hu, Q., Zhang, C. & Zuo, W. 2016 [Coupled dictionary learning for unsupervised feature selection](#). In: *Thirtieth AAAI Conference on Artificial Intelligence*, February 12–17, Phoenix, Arizona, USA. AAAI Press, Phoenix, Arizona, pp. 2422–2428.

First received 27 June 2018; accepted in revised form 18 November 2018. Available online 30 January 2019