Environmental Investment Prediction Using Extended Belief Rule-based System and Evidential Reasoning Rule

Long-Hao Yang [a, c, e], Suhui Wang [d], Fei-Fei Ye [a, c, e], Jun Liu [c], Ying-Ming Wang [a, b, *], and Haibo Hu [e]

[a]Decision Sciences Institute, Fuzhou University, Fuzhou, Fujian, 350116, China

[b]Key Laboratory of Spatial Data Mining & Information Sharing of Ministry of Education, Fuzhou University, Fuzhou, 350116,China

[c]School of Computing, Ulster University at Jordanstown Campus, Newtownabbey BT37 0QB, Northern Ireland, UK

[d] School of Business Administration, Zhejiang University of Finance & Economics, Hangzhou, 310018, China

[e] Department of Electronic and Information Engineering, Hong Kong Ploytechnic University, Hong Kong, PR China

*The corresponding author: msymwang@hotmail.com

Abstract: A scientific environmental investment prediction plays a crucial role in controlling environmental pollution and avoiding the blind investment of environmental management. However, effective environmental investment prediction usually has to fact three challenges about diversiform indicators, insufficient data, and the reliability of prediction models. In the present study, a new prediction model is proposed using the extended belief rule-based system (EBRBS) and evidential reasoning (ER) rule, called ensemble EBRBS model, with the aim to overcome the above challenges for better environmental investment prediction. The proposed ensemble EBRBS model consists of two components: 1) multiple EBRBSs, which are constructed on the basis of not only using various feature selection methods to select representative indicators but also data increment transformation to enrich the training data; 2) an ER rule-based combination method, which utilizes the ER rule to accommodate the weights and reliabilities of different EBRBSs with the predicted outputs of these EBRBSs to have an integrated environmental investment prediction. A detailed case study is then provided for validating the proposed model via extensive experimental and comparison analysis based on the real-world environmental data about 25 environmental indicators for 31 provinces in China ranged from 2005 to 2018. The results demonstrate that the ensemble EBRBS model can be used as an effective model to accurately predict environmental investments. More importantly, the ensemble EBRBS model not only obtains a high accuracy better than some existing prediction models, but also has an excellent robustness compared with others under the situations of excessive indicators and insufficient data.

Keywords: Investment prediction; Ensemble model; Extended belief rule-based system; Evidential reasoning rule

Notations			
R_k	kth extended belief rules	$A_{i,j}$	jth reference value of the ith antecedent attribute
U_i	ith antecedent attributes	B_n	nth consequent of the consequent attribute
D	Consequent attribute	$\alpha_{i,j}^k$	Belief degree assigned to reference value $A_{i,j}$ in the kth rule
θ_k	Weight of the kth rule	β_n^k	Belief degree assigned to consequent B_n in the kth rule
δ_i	Weight of the ith antecedent attribute	β_n	Combined belief degree of consequent B_n
r_h	Reliability of the hth EBRBS	β_n^h	Combined belief degree of consequent B_n in the hth EBRBS
ω_h	Weight of the hth EBRBS	$m_{B_n,h}$	Basic probability mass of consequent B_n in the hth EBRBS
$u(A_{i,j})$	Utility values for reference value $A_{i,j}$	$\hat{m}_{B_n,e(i)}$	Combined basic probability mass of B_n in the ith aggregation
$u(B_n)$	Utility values for consequent B_n	p_{B_n}	Combined belief degree of B_n in the ensemble EBRBS model
$<x_t, y_t>$	tth input-output data pair	$f_h(\Delta x_k)$	Predicted output of the hth EBRBS to reply Δx_k
$<\Delta x_k, \Delta y_k>$	kth input-output data increment pair	$F(\Delta x_k)$	Predicted output of ensemble EBRBS model to reply Δx_k

1. Introduction

Environmental investment prediction is one of the most challenging tasks because of growing pollution emissions and ecological damages (Song *et al*., 2018; Ye *et al*., 2019a), and an accurate prediction of future environmental investment will

1

benefit resource allocation, pollution control, and cleaner production. Environmental investment prediction is also of vital significance to set a pollution control target for different decision-makers by the scientific investment planning. Although lots of researches have paid attention to investment prediction in the last few years under the aggravation of environmental pollution (Chen *et al.*, 2017; Coast *et al.*, 2017; Wang *et al.*, 2018), including expert survey-based, information fusion-based, or feature selection-based indicator selection, and time series-based or input-output relationship-based investment prediction, however, those studies are limited in different aspects and constructing an effective environmental investment prediction model is still facing a certain number of challenges.

Environmental indicator selection is the first challenge to construct an effective prediction model, and lots of previous studies based on singleton feature selection method have been done in the target of selecting a suitable set of representative indicators. However, due to the fact that there are various categories of indicators, *e.g.*, air pollution-related, water pollution-related, and economic development-related indicators, and these indicators play different roles in environmental investment prediction, the indicators selected by singleton feature selection method usually fail to reflect real representative indicators (Ye *et al.*, 2019a; Ye *et al.*, 2020a). Moreover, the indicator selection means the loss of indicator information because the unselected indicators have to be discarded in the model construction. Consequently, it is likely to neglect the information of important indicators on environmental investment prediction in the previous studies.

Limited number of data collected from historical environmental investments is the second challenge to construct an effective prediction model. This is because the input-output relationship-based prediction models have demonstrated their potential in environmental investment prediction over the time series-based models and they have a high requirement to initialize the parameters involved in the model construction. Consider that the experiences and knowledge of domain experts usually fail to ensure a desired accuracy due to the complicated process of environment investment prediction (Zheng *et al.*, 2020), sufficient environmental data should be collected and utilized to train the parameters of the input-output relationship-based prediction models. However, owing to the regional distinctiveness of environmental management, the public environmental data that can be used for constructing investment prediction models are relatively limited.

The use of reliability is the third challenge to construct an effective prediction model because there are many uncertain factors in environmental investment prediction, *e.g.*, imperfect modeling mechanism, unreasonable indicator selection, and incomplete indicator data. Hence, the reliability is necessary to be considered to quantify these factors into an overall measurement for a prediction model. More importantly, in those existing prediction models, much attention have been focused on the model weight, which indicates the importance of a model and is a subjective property, but none of them took into consideration the model reliability, which is used to measure the quality of model objectively. It is clear that reliability and weight characterize the properties of a model in different aspects and thus should be distinguished from each other in constructing an effective environmental investment prediction model.

According to the discussions above and the literature review shown in Section 2, three undisputed facts can be found as follows: (1) feature selection method is able to select representative environmental indicators from diversiform indicators, but it may bring some bias to the set of selected indicators; (2) environmental management data are relatively limited due to time and cost, especially for the indicator data that can be applied in environmental investment prediction, which will cause the over-fitting problem and decrease prediction models' accuracy; (3) existing studies for environmental investment prediction have not taken into consideration model's reliability in model construction, which mostly leads to inaccurate prediction results comparing with actual investments due to many uncertain factors related to indicators and data.

To overcome the above-mentioned challenges, in the present work, an ensemble environmental investment prediction model is proposed on the basis of the ensemble learning framework (Zhou, 2012; Yang *et al.*, 2020) to integrate multiple extended belief rule- based systems (EBRBSs) into an ensemble model using the evidential reasoning (ER) rule and it is so called *ensemble EBRBS model*, where the EBRBS is a rule-based system (Liu *et al.*, 2013) and has been successfully applied to various prediction problems, *e.g.*, bridge risk assessment (Yang *et al.*, 2017), pipeline leak detection (Yang *et al.*, 2018a), and environmental governance cost prediction (Wang *et al.*, 2020); the ER rule (Yang and Xu, 2013) is a generalization of Dempster-Shafer theory (DST) and it is able to consider both evidence weights and reliabilities in a coherent framework.

The construction of an ensemble EBRBS model consists of two processes: EBRBS-based construction process and ER rule-based combination process. The former process takes into consideration various kinds of feature selection methods and data increment transformation, indicating that diverse sets of representative environmental indicators should be selected and original environmental data should be transformed into data increments to enrich the training data; the latter process utilizes the ER rule to take into consideration both weights and reliabilities of different EBRBSs and also integrate the predicted outputs of these EBRBSs to predict environmental investments. Owing to these processes, the ensemble EBRBS model has the following advantages over other existing environmental investment prediction models:

(1) The ensemble EBRBS model is able to overcome the first challenge, because various kinds of feature selection methods are used together to select representative environmental indicators, indicating that diverse sets of representative indicators can be selected from different perspectives and therefore integrates the strengths of different feature selection methods into the proposed ensemble prediction model.

(2) The ensemble EBRBS model is able to overcome the second challenge owing to the data increment, which not only can enrich the scale of the training data used to approximate the complicated relationship between environmental indicators and investments, but also it is generated from a reversible data transformation process, so the ensemble EBRBS model can effectively predict environmental investments via the data increments.

(3) The ensemble EBRBS model is able to overcome the third challenge, because it is required to independently calculate the reliability and the weight of each EBRBS and reflect these reliabilities and weights on the predicted environmental investments using the ER rule. Moreover, the use of the ER rule can help the ensemble EBRBS model to obtain unique predicted investments from different EBRBSs constructed by diverse sets of representative indicators.

In order to demonstrate the effectiveness of the proposed ensemble EBRBS model, the environmental data about a total of 25 environmental indicators of 31 provinces in China ranged from 2005 to 2018 are used to provide an experimental study. The development procedure of an ensemble EBRBS model is introduced firstly based on EBRBS-based construction process and ER rule-based combination process, followed by the comparative analysis to demonstrate that the ensemble EBRBS model can accurately predict environmental investments better than some existing prediction models.

The novelties and contributions of this study are depicted below: (1) use of various feature selection methods to select representative indicators; (2) use of data increments to enrich training data; (3) taking into consideration prediction model's reliability; (4) combining the above three aspects to propose an ensemble model using the EBRBS and the ER rule.

The structure of this work is organized as follows: Section 2 is the literature review and challenges of environmental indicator selection and investment prediction. Section 3 proposes an ensemble environmental investment prediction model. Section 4 provides a case study to perform model validation. Section 5 finally concludes this study.

2. Literature Review and Challenges

In this section, the previous studies of environmental indicator selection and environmental investment prediction are reviewed firstly, and then the challenges of these studies are summarized to illustrate the necessity of this study.

2.1. Previous studies on environmental indicator selection

Previous studies on environmental indicator selection can be divided into three categories: expert survey-based, feature selection-based and information fusion-based indicator selection. The related literatures are reviewed as follows:

(1) Expert survey-based indicator selection. The main feature of the expert survey-based indicator selection is based on the analysis and experience of experts on environmental indicators to select representative indicators. So far, this way has been widely used in the field of environment-related indicator selection. For instance, according to the prior knowledge on environmental indicators, Chen *et al.* (2017) selected seven indicators, like the total volume of sulphur dioxide, chemical oxygen demand, and carbon dioxide, as environmental indicators to carry out the investment analysis of China's sustainable development; Song *et al.* (2019) investigated the impact of knowledge trade on sustainable development and environmental biased technical progress based on the environmental indicators selected by experts' experience; Zhang *et al.* (2019) chose environmental indicators also based on experts' experience to analyze the impact of environmental regulation and foreign investment behavior on the amount and intensity of carbon emissions; additionally, on the basis of the expertise of public health professionals, Mulenga and Siziya (2019) used indoor air quality measured by $PM_{2.5}$ and VOC to study the indoor air pollution-related respiratory health. From the above review, it was found that the experience of experts is a commonly used solution to select environmental indicators, but this way usually depends on experts' subjective judgment and therefore leads to some disputes on the objectivity and rationality of indicator selection.

(2) Information fusion-based indicator selection. The feature of the information fusion-based indicator selection is based on a composite operators/algorithm/function to integrate the information of different environmental indicators, thus it not only can reduce the number of indicators used in environmental investment prediction, but also ensure the integrity of environmental indicator information to the greatest extent. For instance, Jiang *et al.* (2018) presented a three-dimensional sustainability assessment methodology based on principal component analysis (PCA) method, in which the sustainability assessment is based on the integration of economic, environmental, and social indicators; Ye *et al.* (2019a) introduced the classical ER approach to integrate desirable and undesirable output indicators, respectively, for better environmental cost prediction. The results verified that the indicator integration could improve the accuracy of environmental cost prediction. Furthermore, for the air pollution management with interval uncertainty, Ye *et al.* (2020b) introduced the interval ER (IER) approach to integrate air pollution-related indicators. Recently, Deus *et al.* (2020) developed an aggregate indicator to assess the environmental impact of municipal solid waste management, in which the aggregate indicator is integrated from the per capita waste generation and the waste quality index using the geometric mean of the indicators. From the review of existing studies, it was found that although the information fusion can guarantee the integrity of indicator information, it is also easy to integrate useless indicators due to lack of pre-selection of effective indicators.

(3) Feature selection-based indicator selection. The main feature of the feature selection-based indicator selection is based on statistics or machine-learning techniques to select a subset of original environmental indicators as representative indicators. For instance, Salcedo-Sanz *et al.* (2018) analyzed and discussed the feature selection problems found in some renewable energy applications. They suggested that wrapper feature selection problem approaches are those mostly used in renewable energy prediction systems; Ye *et al.* (2019b) utilized PCA method to extract the top four principal components as

new indicators to construct a fuzzy rule-based system (FRBS)-based investment prediction model. In addition, the previous studies on the feature selection-based indicator selection also include using the interrelation of indicator (Niemeijer and de Groot, 2008; Roshani and Gerami, 2018) and using environmental impact to determine indicators (OECD, 2003; Issa *et al*., 2015). Recently, Wang *et al.* (2020) introduced correlation feature selection (CFS) method to select representative indicators for environmental governance cost prediction modeling. The results showed that the CFS method can effectively reduce model complexity and improve prediction accuracy. From the review of existing studies, it was clear that the feature selection-based indicator selection mainly focused on singleton feature selection method. To best of our knowledge, there are few studies on the application of multiple feature selection methods to select representative indicators.

2.2. Previous studies on environmental investment prediction

Investment prediction is the key factor that affects the efficiency of environmental management, but how many specific investments are needed in environmental management is still an urgent problem to be solved. For this reason, many studies have been done to propose an effective investment prediction model for improving environmental management. The related literatures are reviewed as follows:

(1) Time series-based investment prediction. Time series forecasting is a commonly used method in environmental investment prediction. Basically, by analyzing the time series data of previous environmental investments, the analogy of environmental investment can be carried out to predict the amount of environmental investment that may be achieved in the next period or years. For example, Valipour *et al*. (2013) introduced the auto regressive integrated moving average (ARIMA) model for the prediction of future investment scheme, but the prediction investment schemes proposed in their study only reflected the trend changes of investment in time series rather than the consideration of input-output relationship; Kaytez (2020) proposed a hybrid approach based on ARIMA model and least square support vector machine (LSSVM) for long-term forecasting of net electricity consumption. The results showed that the proposed approach is able to accurately predict net electricity consumption. As another common time series forecasting method, the grey model (GM) was also applied in the field of investment prediction, *i.e.*, Xu *et al*. (2017) developed a novel GM model with nonlinear optimized time response method to predict future energy consumption. They demonstrated that the proposed GM model is able to promote modeling accuracy; Chen *et al.* (2017) also provided an investment prediction scheme based on GM model and inverse data envelopment analysis (DEA) model for Chinese environmental governance in the next ten years. Recently, Tong *et al.* (2020) applied time series forecasting to predict the target-availability of China's investments for green growth. From the review of the above studies, the investment prediction processes were mainly based on time series forecasting, which ignored the influence of environmental pollution and economic development on investment prediction.

(2) Input-output relationship-based investment prediction. Input-output relationship-based investment prediction is a novel model which takes the indicators of economic development and environmental pollution as inputs and the indicators of environmental investment as outputs to construct prediction models. More importantly, this kind of model considers the relationship among regional economy, environmental pollution, and environmental investment in prediction process. In this respect, Ye *et al*. (2019a) proposed a genetic program model to construct a new environmental investment prediction scheme by considering the indicator integration and the risk coefficient of decision makers. The results showed that the proposed model has higher accuracy than GM model in investment prediction. Meanwhile, by using IF-THEN rules to describe the input-output relationship found in environmental investment prediction, fuzzy rule-based system (FRBS) and adaptive network-based fuzzy inference system (ANFIS) were applied to provide new solutions of input-output relationship

-based investment prediction modeling (Ye *et al.*, 2019b; Ye *et al.*, 2020a). The results further demonstrated that the input-output relation-based investment prediction outperforms the time series-based investment prediction. Recently, based on the same idea of using IF-THEN rules for investment prediction, EBRBS was introduced to construct input-output relationship-based investment prediction model, *i.e.*, Wang *et al.* (2020) proposed an EBRBS with joining learning for the investment prediction of environmental management. The results revealed that the EBRBS can be applied in investment prediction and has higher accuracy than other models; Wang and Ye (2020) used EBRBS as a core theory to propose a cost prediction model for the environmental governance of transportation industry under technological constraints. From the review of the above studies, the input-output relationship-based investment prediction is becoming the principal method and EBRBS is one of representative methodologies used for environmental investment prediction.

2.3. Challenges to face for developing an investment prediction model

According to the literature review shown in Sections 2.1 and 2.2, many studies have been done to construct models for environmental investment prediction. However, the following challenges still need attention.

Challenge 1: Various environmental indicators must be considered in prediction models.

A large number of environmental indicators, *e.g.*, waste gas, waste solid, and wastewater in a rough category, carbon dioxide, sulfur dioxide, chemical oxygen demand, and gross domestic product in a specific category, have been commonly used in previous studies. However, none of existing studies could contain all environmental indicators in a prediction model because too many indicators easily decrease the accuracy of the prediction model due to the curse of dimensionality. For this dilemma, a common strategy used in the previous studies was that a well-known feature selection method is introduced to select representative environmental indicators, but any feature selection method actually has its inherent weaknesses and it would also affect the accuracy of the prediction model.

Challenge 2: Constructing prediction models have to depend on a limited number of environmental data.

A limited number of environmental data is a serious but common problem in environmental investment prediction and this problem would result in the over-fitting and low accuracy of environmental investment prediction models because of lack of enough training data for model construction. Although many previous studies collected the environmental data of different regions into a whole to enlarge the scale of training data, *e.g.*, using the environmental data of Beijing, Fujian, and Xinjiang in China as training data to predict the environmental investment of Beijing, it would result in a new issue that the collected environmental data fail to reflect the differences in environmental investment between different regions and finally affect the accuracy of the prediction model.

Challenge 3: Reliability of prediction models should be considered in environmental investment prediction.

In the actual practice, various kinds of factors would affect the effectiveness and accuracy of environmental investment prediction, *e.g.*, the reliability of collected data, selected indicators and methodologies. All these reliabilities usually can be summarized into the reliability of the environmental investment prediction models. However, to best of our knowledge, previous studies have not yet taken in consideration the reliability of the models because the existing methodologies used for environmental investment prediction are lack of a necessary parameter to quantify reliabilities. In the case of providing an effective and accurate environmental investment prediction, it is necessary to design a new parameter to quantify these factors into an overall measurement of the reliability for a prediction model.

For the above three challenges, it is clear from the literature review shown in Sections 2.1 and 2.2 that existing studies still failed to construct an effective environmental investment prediction model, because all these attempts cannot meet the

requirement of the three challenges. Although the model based on EBRBS (Wang *et al*., 2020) has shown its advantages over other models in the terms of accuracy and efficiency, how to further enhance this model and overcome the three challenges needs to be studied. Therefore, the purpose of the present study is to propose a novel model based on EBRBS for better environmental investment prediction, as well as overcome the above three challenges.

3. Ensemble Environmental Investment Prediction Model

To overcome the three challenges detailed in Section 2.3, an ensemble environmental investment prediction model is proposed based on EBRBS and the ER rule. Firstly, a generic framework of the proposed model is introduced in Section 3.1, followed by the introduction of its two main modeling processes in Section 3.2 and Section 3.3. Finally, the environmental investment prediction process of the proposed model is introduced in Section 3.4.

3.1. Generic framework of ensemble EBRBS model

In this section, a generic framework based on the EBRBS (Liu *et al*., 2013) and the ER rule (Yang and Xu, 2012) is proposed to illustrate how to construct an ensemble environmental investment prediction model. As shown in Fig. 1, the ensemble model mainly consists of multiple EBRBSs and all these EBRBSs are combined by the ER rule, so the ensemble model is also called an *ensemble EBRBS model*. Additionally, it is clear from Fig. 1 that the construction of the ensemble EBRBS model includes two critical processes: EBRBS-based construction process and ER rule-based combination process, in which the target of the former one is to construct multiple EBRBSs and the target of the latter one is to combine the results from all these EBRBSs.
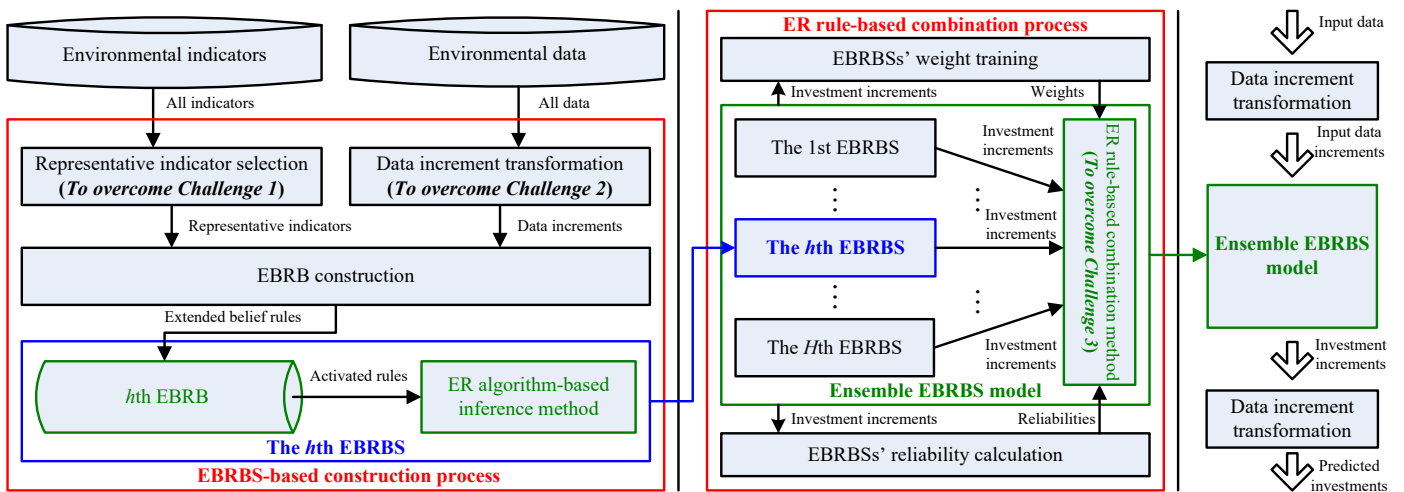

Fig. 1. Generic framework of ensemble EBRBS model for environmental investment prediction

From Fig. 1, the following descriptions are provided for the proposed ensemble EBRBS model:

(1) EBRBS-based construction process for constructing the ensemble environmental investment prediction model

Constructing an effective environmental investment prediction model usually depends on representative indicators and adequate training data. Hence, on the basis of EBRBS as the core foundation of environmental investment prediction model, feature selection methods are introduced to select representative indicators from original indicators, and a new data form, named data increment, is proposed to ensure adequate training data, in which the data increment can be generated by the increments of any two environmental data, so the total number of training data is $T\times(T-1)$ when there are T collected data, and the predicted environmental investments can be calculated according to the sum of previous investments and predicted data increments. According to the above-mentioned viewpoints, the EBRBS-based construction process is proposed for ensemble environmental investment prediction model and its details can be found in Section 3.2.

(2) ER rule-based combination process for constructing the ensemble environmental investment prediction model

Any kind of feature selection methods inevitably has its strength and weakness; thus it is usually a challenge to select an appropriate feature selection method. For this challenge, a smart strategy is to select indicators based on various kinds of methods so that all of them can take their respective advantages in selecting representative indicators. When there are H kinds of feature selection methods used to select representative indicators, H different EBRBSs should be constructed for environmental investment prediction. To make full use of these EBRBSs, an ER rule-based combination process is proposed as a cooperative scheme. More importantly, owing to the ER rule's advantages, the proposed ER rule-based combination process can consider both EBRBS's weight and reliability in a coherent framework. The details of the ER rule-based combination process can be found in Section 3.3.

(3) Environmental investment prediction based on the ensemble EBRBS model

Based on the EBRBS-based construction process and ER rule-based combination process, an ensemble EBRBS model can be constructed for environmental investment prediction and it includes H EBRBSs and the ER rule-based combination method. Considering that data increment is the basic data form used for constructing the ensemble EBRBS model, data increment transformation should be also included in the process of environmental investment prediction. In other words, when a new input data is provided for the ensemble EBRBS model, the first step is to transform the new input data into input data increments, the second step is to predict investment increments through H EBRBs and their ER algorithm-based inference method, the third step is to integrate the H predicted investment increments via the ER rule-based combination method, and the final step is to transform integrated investment increments into predicted investments.

3.2. EBRBS-based construction process for ensemble EBRBS model

In this section, the EBRBS-based construction process is introduced to illustrate how to construct an EBRBS based on excessive environmental indicators and insufficient environmental data, in which the related flowchart is showed in Fig. 2.
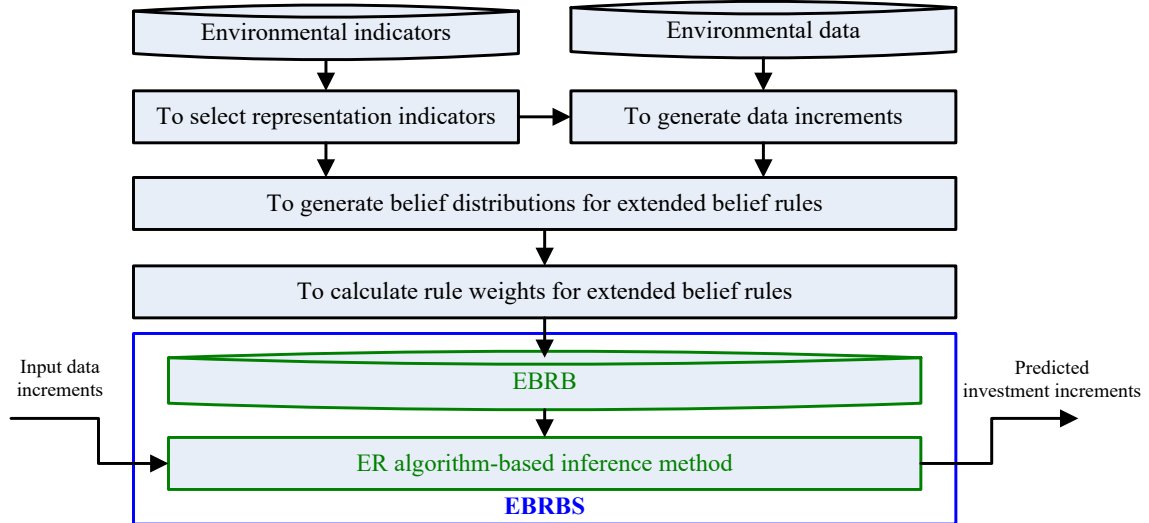

Fig. 2. Flowchart of the EBRBS-based construction process

From Fig. 2, there are four main steps to construct an EBRBS and their targets are related with representative indicators, data increments, and extended belief rules, respectively. Apart from the representative indicators, which are a subset of original environmental indicators, the data increments and extended belief rules are intermediate results during the EBRB construction. Thus, in order to provide a clear introduction of the EBRBS-based construction process, the definitions of data increment and extended belief rule are given as follows:

Definition 1 (Data increment). Consider a case of a M-dimensional function $y=f(x)$ with $x=\{x_1,…,x_M\}$ and its definition domain $[a, b]$, where a and b are all M-ary vectors, respectively. When there exists an input-output data pair $<x_0, y_0>$ in the

function $y=f(x)$, for any $x_1 \in [a, b]$, the data increment regarding the input and output data can be written as $\Delta x = x_1 - x_0$ and $\Delta y = f(x_1) - f(x_0) = f(x_0 + \Delta x) - f(x_0)$, respectively.

Definition 2 (Extended belief rule). Suppose that there are M antecedent attributes U_i ($i=1,\dots, M$) with each attribute having J_i reference values $A_{i,j}$ ($j=1,\dots, J_i$) and one consequent attribute D with N consequents B_n ($n=1,\dots, N$). Hence, the kth ($k=1,\dots, L$) extended belief rule R_k in the EBRB is written as:

$$R_k : IF\ U_1\ is\ \{(A_{1,j}, \alpha_{1,j}^k); j = 1,\dots,J_1\} \wedge \dots \wedge U_M\ is\ \{(A_{M,j}, \alpha_{M,j}^k); j = 1,\dots,J_M\}$$
$$THEN\ D\ is\ \{(B_n, \beta_n^k); n = 1,\dots,N\},\ with\ \theta_k\ and\ \{\delta_1,\dots,\delta_M\}. \tag{1}$$

where $\{(A_{i,j}, \alpha_{i,j}^k); j = 1,\dots,J_i\}$ and $\{(B_n, \beta_n^k); n = 1,\dots,N\}$ denote the belief distributions used to describe U_i and D in R_k; $\alpha_{i,j}^k (0 \le \alpha_{i,j}^k \le 1)$ and $\beta_n^k (0 \le \beta_n^k \le 1)$ denote the belief degrees of $A_{i,j}$ and B_n in R_k, and they must meet $\sum_{j=1}^{J_i} \alpha_{i,j}^k \le 1$ and $\sum_{n=1}^{N} \beta_n^k \le 1$; $\delta_i (0 < \delta_i \le 1)$ and $\theta_k (0 < \theta_k \le 1)$ denote the weight of U_i and R_k, respectively.

According to **Definition 1** and **Definition 2**, the corresponding steps of the EBRBS-based construction process are introduced as follows:

Step 1: To select representative indicators. Suppose that environmental investment D is related with a large number of original environmental indicators $\{U_i; i=1,\dots, +\infty\}$. In order to construct an effective investment prediction model, a feature selection method is introduced to select representative indicators. Without of loss generality, the selected representative indicators are signed as $\{U_i; i=1,\dots, M\}$. Hence, a corresponding EBRBS can be constructed, in which $\{U_i; i=1,\dots, M\}$ and D are regarded as antecedent and the consequent attributes of the EBRBS, respectively. Additionally, in order to describe these antecedent and consequent attributes, assume that there are J_i referential values $\{A_{i,j}; j=1,\dots, J_i\}$ for the ith antecedent attribute and N consequents $\{B_n; n=1,\dots, N\}$ for the consequent attribute D.

Step 2: To generate data increments. Suppose that there are T input-output environment data pairs $<x_t, y_t>$ ($t=1,\dots, T$) for the selected M antecedent attributes $\{U_i; i=1,\dots, M\}$ and the consequent attribute D, where $x_t=\{x_{t,1},\dots, x_{t,M}\}$. Considering that original input-output environment data pairs are usually insufficient for system modeling. Based on **Definition 1**, the data increments of any two environmental data pairs, e.g., $<x_t, y_t>$ and $<x_s, y_s>$ ($t, s=1,\dots, T; t \ne s$), are generated as follows:

$$\Delta x_{t,s} = x_t - x_s \tag{2}$$

$$\Delta y_{t,s} = y_t - y_s \tag{3}$$

Here, it is worth noting that the new set of training data has $T \times (T-1)$ input-output environmental data increment pairs. For the sake of descriptions, these $T \times (T-1)$ data increment pairs are denoted as $< \Delta x_k, \Delta y_k >$ ($k=1,\dots, L; L= T \times (T-1)$).

Step 3: To generate belief distributions for extended belief rules. Suppose that the basic parameters, including utility values $u(A_{i,j})$ ($j=1,\dots,J_i; i=1,\dots,M$) for antecedent attributes, utility values $u(B_n)$ ($n=1,\dots, N$) for consequent attribute, and attribute weights $\delta_i (i=1,\dots, M)$, are predetermined by using expert knowledge or prior information. Accordingly, L groups of belief distributions can be generated from L input-output environmental data increment pairs by using the utility-based information transformation technique (Yang 2001), in which the belief distribution of the ith antecedent attribute at the kth extended belief rule, namely $\{(A_{i,j}, \alpha_{i,j}^k); j = 1,\dots,J_i\}$, is generated from input data increment $\Delta x_{k,i}$ by:

$$a_{i,j}^k = \frac{u(A_{i,j+1}) - \Delta x_{k,i}}{u(A_{i,j+1}) - u(A_{i,j})}\ and\ a_{i,j+1}^k = 1 - a_{i,j}^k,\ if\ u(A_{i,j}) \le \Delta x_{k,i} \le u(A_{i,j+1}) \tag{4}$$

$$a_{i,t}^k = 0,\ for\ t = 1,\dots,J_i\ and\ t \ne j, j+1 \tag{5}$$

Similarly, the belief distribution of consequent attribute at the kth extended belief rule, namely $\{(B_n, \beta_n^k); n = 1,\dots,N\}$,

can be generated from output data increment Δy_k by using the utility-based information transformation technique.

Step 4: To generate rule weights for extended belief rules. Based on the belief distributions generated from L input-output data increment pairs, the rule weight of each extended belief rule should be calculated. Taking the kth ($k=1,...,L$) extended belief rule as an example, the similarity of rule antecedent (SRA) and the similarity of rule consequent (SRC) are calculated as follows:

$$SRA(R_l, R_k) = 1 - \max_{t=1,...,M} \left\{ \min \left\{ 1, \sqrt{\sum_{j=1}^{J_t} (\alpha_{t,j}^l - \alpha_{t,j}^k)^2} \right\} \right\}; l = 1,...,L; l \neq k \qquad (6)$$

$$SRC(R_l, R_k) = 1 - \min \left\{ 1, \sqrt{\sum_{n=1}^{N} (\beta_n^l - \beta_n^k)^2} \right\}; l = 1,...,L; l \neq k \qquad (7)$$

Then, the inconsistency degree of the kth extended belief rule is calculated as follows:

$$Incons(R_k) = \sum_{l=1, l\neq k}^{L} \left(1 - \exp \left\{ -\frac{\left(\dfrac{SRA(R_l, R_k)}{SRC(R_l, R_k)} - 1 \right)^2}{\left(\dfrac{1}{SRA(R_l, R_k)} \right)^2} \right\} \right) \qquad (8)$$

Finally, the rule weight of the kth extended belief rule is calculated as follows:

$$\theta_k = 1 - \frac{Incons(R_k)}{\sum_{j=1}^{L} Incons(R_j)} \qquad (9)$$

For the above EBRBS-based construction process, the following two key points should be provided:

(1) Basic parameters are vital importance to construct an effective EBRBS, but domain experts are not always available to provide an optimal value for each basic parameter. Hence, a better way is to train basic parameters using the parameter learning model (Yang *et al*., 2019) as follows:

$$\min MAE(\{\delta_i, u(A_{i,j}), u(B_n)\}) = \sum_{k=1}^{L} |\Delta y_k - f(\Delta \boldsymbol{x}_k)|$$

$$s.t. \begin{cases} 0 \leq \delta_i \leq 1; i = 1,...,M \\ u(A_{i,j}) \leq u(A_{i,j+1}); j = 1,...,J_i - 1; i = 1,...,M \\ u(A_{i,1}) = lb_i; u(A_{i,J_i}) = ub_i; i = 1,...,M \\ u(B_n) \leq u(B_{n+1}); n = 1,...,N - 1 \\ u(B_1) = lb; u(B_N) = ub \end{cases} \qquad (10)$$

where lb_i and ub_i are the lower and upper bounds of the ith input indicators, respectively; lb and ub are the lower and upper bounds of investment indicator, respectively; $\Delta \boldsymbol{x}_k$ and Δy_k are the kth ($k=1,..., L$) input data and investment increment; and $f(\Delta \boldsymbol{x}_k)$ is the predicted investment increment of the EBRBS for the given input data increment $\Delta \boldsymbol{x}_k$.

(2) In order to solve parameter learning models, various kinds of parameter learning algorithms have been proposed in previous studies, *e.g.*, DE algorithm, GA algorithm, and optimization toolbox at MATLAB. All of them are available to get a high-quality solution to optimize the value of basic parameters.

3.3. ER rule-based combination process for ensemble EBRBS model

In this section, the ER rule-based combination process is introduced to show how to construct an ensemble EBRBS model from the different EBRBSs constructed by different sets of the representative indicators selected by different feature selection methods, in which the related flowchart is showed in Fig. 3.
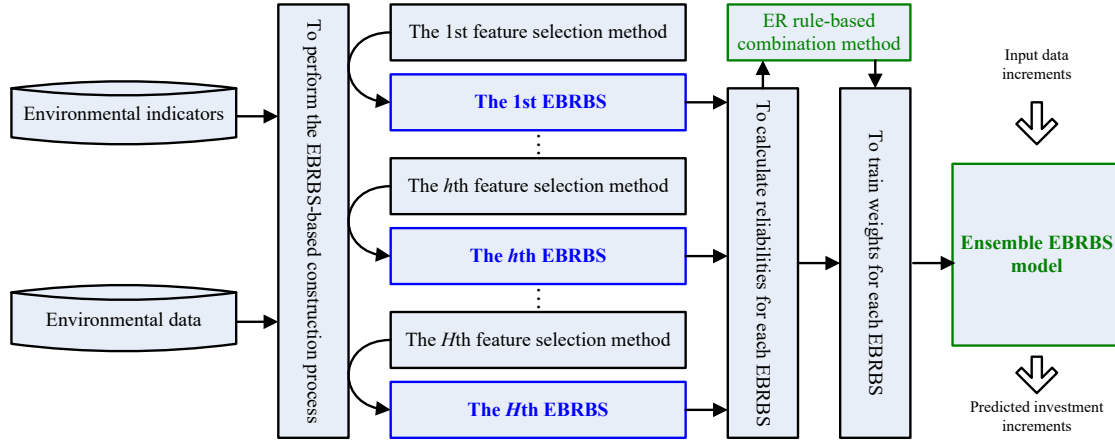
Fig. 3. Flowchart of the ER rule-based combination process

From Fig. 3, there are three main steps to construct an ensemble EBRBS model and their targets are related with EBRBSs and their weights and reliabilities, respectively. Apart from the EBRBSs, which can be constructed based on the steps shown in Section 3.2, the weights and reliabilities should be further calculated based on the EBRBS and/or the ER rule-based combination method. Thus, in order to provide a clear introduction of the ER rule-based combination process, the definitions of weights and reliabilities are given as follows respectively:

Definition 3 (Reliability of EBRBS). The reliability, as an objective property of an EBRBS, represents the capability of the EBRBS to provide a correct prediction of the considered environmental investment problem.

Definition 4 (Weight of EBRBS). The weight, as a subjective property of an EBRBS, represents the importance of the EBRBS on the process of using the ensemble prediction model to predict environmental investments.

Based on ***Definitions 3*** and ***4***, the detailed steps of the ER rule-based combination process are introduced as follows:

Step 1: To perform the EBRBS-based construction process. Suppose that there are H kinds of feature selection methods to select representative indicators from original environmental indicators $\{U_i; i=1,\ldots, +\infty\}$ for environmental investment D. Accordingly, H sets of representative environmental indicators can be selected and are further used to construct H EBRBSs according the EBRBS-based construction process shown in Section 3.2.

Step 2: To calculate reliabilities for each EBRBS. Suppose that there are L input-output data increment pairs $<\Delta x_k$, $\Delta y_k > (k=1,\ldots, L)$, and the predicted investment increments of the hth ($h=1,\ldots, H$) EBRBS is $f_h(\Delta x_k)$, in which the details of $f_h(\Delta x_k)$ can be found in Section 3.4. According to ***Definition 3***, the reliability of the hth EBRBS, denoted as r_h, can be calculated by using the error between actual data increments y_k and predicted data increments $f_h(\Delta x_k)$ as follows:

$$r_h = 1 - \frac{1}{L}\sum\nolimits_{k=1}^{L}\frac{\left|f_h(\Delta x_k) - \Delta y_k\right|}{\max_{i=1,\ldots,L}\{\Delta y_i\} - \min_{i=1,\ldots,L}\{\Delta y_i\}} \tag{11}$$

Step 3: To train weights for each EBRBS. Suppose that the predicted data increments of ensemble EBRBS model is $F(\Delta x_k)$ and Δx_k is related with the kth ($k=1,\ldots, L$) input-output data increment pair $<\Delta x_k, \Delta y_k >$, in which the details of $F(\Delta x_k)$ can be found in Section 3.4. According to ***Definition 4***, the weight of H EBRBSs, denoted as ω_h ($h=1,\ldots, H$), can be trained by the following optimization model:

$$\min MAE(\{\omega_h, u(B_n)\}) = \sum\nolimits_{k=1}^{L}\left|\Delta y_k - F(\Delta x_k)\right|$$
$$s.t. \begin{cases} 0 \le \omega_h \le 1; h=1,\ldots,H \\ u(B_n) \le u(B_{n+1}); n=1,\ldots,N-1 \\ u(B_1) = lb; u(B_N) = ub \end{cases} \tag{12}$$

11

where lb and ub are the lower and upper bounds of output data increments, respectively; $u(B_n)$ denotes the utility value of the nth ($n=1,\ldots, N$) consequent in the ensemble EBRBS model.

For the above-mentioned ER rule-based combination process, the following key points should be noted:

(1) The formula of calculating reliabilities shown in Eq. (17) is based on the prediction error of EBRBSs. When an EBRBS can correctly predict all input data increments, namely $f_h(\Delta x_k) = \Delta y_k$ ($t=1,\ldots, L$), the reliability of the EBRBS is $r_h=1$; Otherwise, the reliability of EBRBS is $0 \leq r_h < 1$.

(2) Based on the ER rule-based combination process, H EBRBSs and their reliabilities and weights can be obtained to construct an ensemble environmental investment prediction model by using the ER rule, which was proposed for evidence combination with the consideration of both evidences' weights and reliabilities.

3.4. Environmental investment prediction using the ensemble EBRBS model

Based on the EBRBS-based construction process shown in Section 3.2 and the ER rule-based combination process shown in Section 3.3, an ensemble EBRBS model can be constructed for environmental investment prediction. In this section, the process of using the ensemble EBRBS model to predict the investment of a given input data is introduced and the related flowchart is showed in Fig. 4.
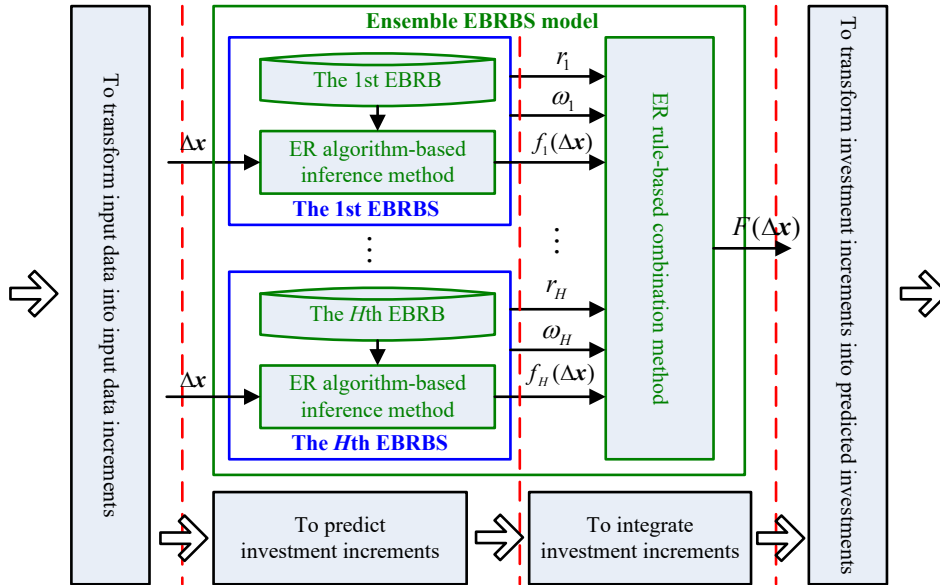

Fig. 4. Flowchart of environmental investment prediction process

From Fig. 4, there are four main steps to predict environmental investments based on the ensemble EBRBS model, the details of these steps are provided as follows:

Step 1: To transform input data into input data increments. Suppose that there are a new input data $x=\{x_i; i=1,\ldots, M\}$ and a historical input-output data pair $\langle x_k, y_k \rangle$. Hence, the input data increment of x and x_k can be calculated by using Eq. (2) and it is denoted as $\Delta x=\{\Delta x_i; i=1,\ldots, M\}$.

Step 2: To predict investment increments based on the hth ($h=1,\ldots, H$) EBRBS.

Step 2.1: To calculate activation weight for each extended belief rule. For the input data increment Δx, each input Δx_i should be transformed into a belief distribution, denoted as $S(\Delta x_i)=\{(A_{i,j}; a_{i,j}); j=1,\ldots, J_i\}$, using the utility-based information transformation technique (Yang 2001). Based on the belief distribution, the activation weight can be calculated for the kth ($k=1,\ldots, L$) extended belief rule as follows:

$$w_k = \frac{\theta_k \prod_{i=1}^{M}(S^k(\Delta x_i, U_i))^{\bar{\delta_i}}}{\sum_{l=1}^{L}(\theta_l \prod_{i=1}^{M}(S^l(\Delta x_i, U_i))^{\bar{\delta_i}})} \tag{13}$$

where $S^k(\Delta x_i, U_i)$ is the individual matching degree of Δx_i to the ith antecedent attribute in the kth extended belief rule; θ_k is the weight of the kth extended belief rule; $\bar{\delta_i}$ is the normalized weight of the ith antecedent attribute; and they can be calculated by:

$$S^k(\Delta x_i, U_i) = 1 - \sqrt{\frac{\sum_{j=1}^{J_i}(a_{i,j} - \alpha_{i,j}^k)^2}{2}} \tag{14}$$

$$\bar{\delta_i} = \frac{\delta_i}{\max_{j=1,\dots,M}\{\delta_j\}} \tag{15}$$

Step 2.2: To integrate activated rules for predicting investment increments. Based on activation weights, the activated rules which meet $w_k>0$ should be integrated using the analytical ER algorithm (Wang *et al.*, 2006):

$$\beta_n = \frac{\prod_{k=1}^{L}(w_k\beta_n^k + 1 - w_k\sum_{i=1}^{N}\beta_i^k) - \prod_{k=1}^{L}(1 - w_k\sum_{i=1}^{N}\beta_i^k)}{\sum_{i=1}^{N}\prod_{k=1}^{L}(w_k\beta_i^k + 1 - w_k\sum_{j=1}^{N}\beta_j^k) - (N-1)\prod_{k=1}^{L}(1 - w_k\sum_{j=1}^{N}\beta_j^k) - \prod_{k=1}^{L}(1 - w_k)} \tag{16}$$

Thus, the predicted investment increments of the hth EBRBS replied for input data increment $\Delta \boldsymbol{x}$ can be represented as follows:

$$f_h(\Delta \boldsymbol{x}) = \sum_{n=1}^{N}u(B_n)\beta_n + \frac{u(B_1) + u(B_N)}{2}(1 - \sum_{n=1}^{N}\beta_n) \tag{17}$$

where $u(B_n)$ is the utility value of the nth consequent in consequent attribute.

Step 3: To integrate investment increments based on H EBRBSs.

Step 3.1: To calculate the basic probability mass for each EBRBS. Based on Eq. (15), the combined belief distributions for H EBRBSs can be obtained and denoted as $\{(B_n, \beta_n^h); n=1,\dots, N\}$ ($h=1,\dots, H$). Afterwards, because the belief degree only exists in singleton consequents $B_1, \dots, B_N$, and global set $\Theta = \{B_1,\dots,B_N\}$, the basic probability masses for H EBRBSs can be calculated as follows:

$$m_{B_n,h} = \omega_h\beta_n^h, n = 1,\dots,N \tag{18}$$

$$m_{\Theta,h} = \omega_h(1 - \sum_{n=1}^{N}\beta_n^h) \tag{19}$$

$$m_{P(\Theta),h} = 1 - \omega_h \tag{20}$$

Step 3.2: To aggregate the basic probability mass of the first two EBRBSs. According to the reliability and basic probability masses of EBRBSs, the combined basic probability masses of the 1st and the 2nd EBRBSs can be calculated as follows:

$$\hat{m}_{\theta,e(2)} = [(1-r_2)m_{\theta,1} + (1-r_1)m_{\theta,2}] + \sum_{B\cap C=\theta}m_{B,1}m_{C,2}, \forall \theta \in \{B_1,\dots,B_N,\Theta\} \tag{21}$$

$$\hat{m}_{P(\Theta),e(2)} = (1-r_2)(1-r_1). \tag{22}$$

Step 3.3: To aggregate the basic probability mass of the remaining H-2 EBRBSs. The combined basic probability masses of remaining EBRBSs are calculated as follows:

$$\hat{m}_{\theta,e(i)} = [(1-r_i)m_{\theta,e(i-1)} + m_{P(\Theta),e(i-1)}m_{\theta,i}] + \sum_{B \cap C = \theta} m_{B,e(i-1)}m_{C,i}, \forall \theta \in \{B_1,...,B_N,\Theta\} \tag{23}$$

$$\hat{m}_{P(\Theta),e(i)} = (1-r_i)m_{P(\Theta),e(i-1)} \tag{24}$$

where

$$m_{\theta,e(i)} = \frac{\hat{m}_{\theta,e(i)}}{\sum_{\theta \in \{B_1,...,B_N,\Theta\}} \hat{m}_{\theta,e(i)} + \hat{m}_{P(\Theta)e(i)}} \tag{25}$$

$$m_{P(\Theta),e(i)} = \frac{\hat{m}_{P(\Theta),e(i)}}{\sum_{\theta \in \{B_1,...,B_N,\Theta\}} \hat{m}_{\theta,e(i)} + \hat{m}_{P(\Theta)e(i)}} \tag{26}$$

Hence, the combined belief degrees of H EBRBSs can be calculated by

$$p_\theta = \frac{\hat{m}_{\theta,e(L)}}{\sum_{n=1}^{N} \hat{m}_{D_n,e(i)} + \hat{m}_{\Theta,e(i)}}; \theta \in \{B_1,...,B_N,\Theta\} \tag{27}$$

Finally, based on utilities $u(B_n)$ ($n=1,...,N$) obtained from Eq. (18) in Section 3.3, the predicted investment increments replied for the input data increment Δx, denoted as $F(\Delta x)$, can be calculated as follows:

$$F(\Delta x) = \sum_{n=1}^{N} u(B_n)p_{B_n} + p_\Theta \frac{u(B_1) + u(B_N)}{2} \tag{28}$$

Step 4: To transform investment increments into predicted investments. Based on the historical input-output data pair $<x_k, y_k>$ and the predicted investment increments $F(\Delta x)$, the predicted investments can be obtained by $y_k + F(\Delta x)$.

For the above environmental investment prediction, Eqs. (17) - (26) are the specific formulas of the conventional ER rule introduced in (Yang and Xu, 2013), because the proposed ensemble EBRBS model focuses on the combination of the belief degrees regarding B_n ($n=1,..., N$), Θ, and $P(\Theta)$, instead of all elements of the power set of Θ.

4. Case Study

To verify the effectiveness of the proposed ensemble EBRBS model, the regional data of environmental management of 31 provinces in China are utilized to carry out an experimental study in this section.

4.1. Data source and indicator collection

A series of environmental management data related with 31 provinces in the mainland of China from 2005 to 2018 are used to verify the ensemble EBRBS model for environmental investment prediction. These data were derived from *China Statistical Yearbook* and *China Environmental Statistical Yearbook*, respectively, both of which are the most commonly used and reliable public database for the study of environmental management in China (Lee and Zhang, 2012; Wu *et al*., 2019; Cao *et al*., 2015).

According to the previous studies on the environmental managements of China (Chen *et al*., 2017; Frank *et al*., 2018; Li *et al*., 2013; Long *et al*., 2018; Ye *et al*., 2019a; Ye *et al*., 2019b; Ye *et al*., 2020a), 20 environmental input indicators and 5 environmental output indicators are collected to construct ensemble EBRBS models. The specific interpretations of these input and output indicators are shown in Table 1 and Table 2.

From Table 1, it can be found that the input indicators involve diverse types of environmental management indicators, *e.g.*, air pollution-related indicators, water pollution-related indicators, garbage and solid waste, and economic development -related indicators. These input indicators are the main indicators used in the previous studies of environmental management and investment prediction. In order to reduce the number of input indicators, all these input indicators are used as original environmental indicator to construct investment prediction models.

Table 1. Introduction of input indicators in environmental investment prediction

No.	Input indicator	Abbr.	Specific interpretation	Max	Min	Average
1	Sulfur dioxide	SO_2	Emission of sulfur dioxide	2,002,000	1000	662,937
2	Smoke and dust	SM	Emission of smoke and dust	1,797,683	1000	364,270
3	Carbon dioxide	CO_2	Emission of carbon dioxide	4,678	7.07	1096
4	Wastewater	WW	Total emission of wastewater	938,261	2685	201,795
5	Chemical oxygen demand	COD	Emission of chemical oxygen demand	198.25	1.38	53.6
6	Nitrous oxide	NO	Emission of nitrous oxide in wastewater	23.09	0.14	5.6
7	Petroleum emissions	PE	Petroleum emissions in wastewater	2,937	0.03	498
8	Lead emission	LE	Lead emission in wastewater	42,466	0.002	1,328
9	Volatile phenol	VP	Volatile phenol in wastewater	2,831	0.01	52.9
10	Mercury emission	ME	Mercury emission in wastewater	1,537	0.001	19.5
11	Cadmium emission	CE	Cadmium emission in wastewater	14,518	0.001	307.5
12	Arsenic emission	AE	Arsenic emission in wastewater	55,705	0.002	1622
13	Six valence inscriptions	SVI	Emission of six valence inscription	27,987	0.06	778
14	Garbage clean-up	GCU	Garbage removal and transportation volume	2,645	16.3	548
15	Industrial solid waste	ISW	Emission of industrial solid waste	35,372	0.2	5,464
16	Harmless treatment of garbage	HTG	Amount of garbage safely disposal	78,185	400	13,475
17	Garbage sanitary landfill	GSL	Amount of garbage sanitary landfill	43,337	349	9,182
18	gross domestic product	GDP	Value of gross domestic product	89,705	220	15,627
19	Total profit	TP	Total profit of Enterprises above Designated Size	10,574	-91.89	1,530.9
20	Operating profit	OP	Operating profit of Enterprises above Designated Size	10,499	-150	1,538

Table 2. Introduction of output indicators in environmental investment prediction

No.	Output indicator	Abbr.	Specific interpretation	Max	Min	Average
1	Energy consumption	EC	Total electricity consumption	5,959	9	1,397
2	Capital investment	CI	Fixed assets investment	55,203	162	10,306
3	Labor investment	LI	Total number of employees	1,973	15	469
4	Investment of pollution treatment	IPT	Investment in industrial pollution control	1,416,464	89	190,545
5	Investment of environmental protection	IEP	Government finance on environmental protection	458	0.13	80

From Table 2, it can be found that the output indicators involve diverse types of management investments because they are related to environmental pollution control, regional environmental management, and regional green production. To date, these output indicators have been commonly used in the previous studies for environmental investment prediction, *i.e.*, EC, CI and LI were the output indicators of investment prediction in (Chen *et al*., 2017; Ye *et al*., 2019a; Ye *et al*., 2019b; Ye *et al*., 2020a; Wang *et al*., 2020), IPT and IEP are also the two indicators that highly related to environmental pollution control and environmental protection. Hence, all these output indicators are used as environmental investments to construct models. Additionally, consider that each output indicator usually has not the same relationship to the input indicators shown in Table 1, the investment prediction of each output indicator is based on an independent prediction model. In other words, the prediction model is constructed by using one output indicator and its proprietary representative input indicators.

In order to compare the predicted investment of prediction models, the following three criteria including mean absolute

error (MAE), mean absolute percentage error (MAPE), and correlation coefficient (R) are utilized for comparative analysis.

$$MAE = \frac{\sum_{t=1}^{T} |z^t - y^t|}{T} \tag{29}$$

$$MAPE = \frac{1}{T} \sum_{t=1}^{T} \left| \frac{z^t - y^t}{y^t} \right| \times 100\% \tag{30}$$

$$R = \frac{\sum_{t=1}^{T} (z^t - \bar{z})(y^t - \bar{y})}{\sqrt{\sum_{t=1}^{T} (z^t - \bar{z})^2 \cdot \sum_{t=1}^{T} (y^t - \bar{y})^2}} \tag{31}$$

where

$$\bar{z} = \frac{1}{T} \sum_{t=1}^{T} z^t, \ \bar{y} = \frac{1}{T} \sum_{t=1}^{T} y^t \tag{32}$$

where z^t and y^t denote the predicted and real investments of the tth ($t=1,\ldots, T$) province at a certain year; T is the total number of provinces at the certain year. Here, the larger R and the smaller MAPE and MAE are considered to be a better performance for a prediction model.

4.2. Development procedure of ensemble EBRBS model

In this section, the procedures of developing an ensemble EBRBS model for environmental investment prediction are analyzed and they include EBRBS-based construction process and ER rule-based combination process.

4.2.1. EBRBS-based construction process for ensemble EBRBS model

To construct ensemble EBRBS models, the environmental data of input and output indicators from 2005 to 2017 of 31 provinces in China are used as the original training data and four kinds of feature selection methods are introduced to select representative input indicators, namely Pearson Correlation Coefficient-based, Simple Linear Regression-based, ReliefF Algorithm-based, and Correlation Coefficient Standard Deviation-based methods. The specific descriptions for each method are showed in Table 3.

Table 3. Introduction of four feature selection methods

Core of indicator selection	Abbr.	Descriptions
Pearson Correlation Coefficient	PCC	To evaluate the worth of each indicator by measuring the correlation (Pearson's) between it each class
Simple Linear Regression	SLR	The current set of attributes is applied to train a simple linear regression model iteratively by removing indicators, so that the performance can evaluate the worth of each indicator
ReliefF Algorithm	RA	To evaluate the worth of each indicator by repeatedly sampling a data and considering the value of a given attribute for the nearest data of the same and different classes
Correlation Coefficient Standard Deviation	CCSD	To evaluate the worth of each indicator by measuring its correlation coefficient and standard deviation

Hence, the EBRBS-based construction process for constructing ensemble EBRBS model is implemented as follows:

Step 1: To select representative indicators using four kinds of feature selection methods.

According to **Step 1** in Section 3.2, representative indicators should be selected for each investment. Taking Shanghai as an example, the representative indicators selected by the four kinds of feature selection methods for EC prediction are showed in Table 4. It shows clearly that the results of the four methods are quite different in the terms of selected representative indicators and their ranking. For example, the representative indicators selected by PCC method for the

investment prediction of Shanghai's EC are CO_2, AE, GDP, TP, and OP, and they are different from the representative indicators selected by SLR, RA, and CCSD methods. It is worth noting that CO_2 and OP are the mainly selected indicators for the investment prediction of Shanghai's EC, indicating that these two indicators are highly related to the EC of Shanghai. Additionally, the selected indicators shown in Table 4 also indicate that the ensemble EBRBS model can make full use of the advantages of different methods to select representative indicators.

Table 4. Four sets of representative indicators for EC prediction

Feature selection method	PCC	SLR	RA	CCSD
Representative indicators	CO_2, GDP, TP, OP, AE	SO_2, CO_2, GDP, TP, OP	CO_2, WW, PE, LE, VP	SO_2, CE, GCU, TP, OP

Step 2: To generate data increments for constructing EBRBSs.

Based on Eq. (2) and Eq. (3) at ***Step 2*** shown in Section 3.2, 13 the input-output environmental data pairs from 2005 to 2017 of Shanghai should be used to generate $13 \times 12 = 156$ input-output data increment pairs. In other words, 156 input-output data increments $<\Delta x_k, \ \Delta y_k>$ (k=1,…, 156) can be generated from 13 original input-output data pairs, so that the scale of training data used to construct the ensemble EBRBS model is enlarged from 13 to 156. For these input and output data increments, Table 5 provides a simple data statistic of the five representative indicators CO_2, GDP, TP, OP and AE selected by PCC method and the investment indicator EC.

Table 5. Data statistic of representative indicators for Shanghai's EC increment prediction

	Five representative indicators selected by PCC method					EC
	CO_2	GDP	TP	OP	AE	
Min	-521.70	-20,105.82	-1,974.35	-1,935.36	-147.58	-664.58
Max	521.70	20,105.82	1,974.35	1,935.36	147.58	664.58

Step 3: To generate belief distributions for 156 extended belief rules.

According to ***Step 3*** in Section 3.2, the basic parameters related to generate belief distributions should be optimized by the parameter learning model shown in Eq. (10) when the input indicators are used as antecedent attributes and the output indicator are used as consequent attributes in an EBRBS. Suppose that there are five reference values for all antecedent and consequent attributes, in which the reference values of antecedent attributes CO_2, GDP, TP, OP, and AE and the consequents of consequent attribute EC are set as {Very Low, Low, Medium, High, Very High}, the DE-based parameter learning algorithm (Yang *et al.*, 2019) is used to optimize basic parameters by setting the number of iterations and individuals as 200 and 20. Consequently, Table 6 shows the optimized basic parameters used for the investment prediction of Shanghai's EC.

Table 6. Optimized basic parameters of EBRBS for Shanghai's EC increment prediction

Indicator	Correspondent relationship	Weight	Very Low	Low	Medium	High	Very High
CO_2	Antecedent attribute	0.8500	-521.70	-200.63	23.50	156.89	521.70
GDP	Antecedent attribute	0.7967	-20,105.82	-9,088.30	-2,843.55	1,469.47	20,105.82
TP	Antecedent attribute	0.8363	-1,974.35	-1230.58	519.87	1314.26	1,974.35
OP	Antecedent attribute	0.8597	-1,935.36	-999.11	-801.84	151.08	1,935.36
AE	Antecedent attribute	0.6503	-147.58	-72.25	-1.18	79.61	147.58
EC	Consequent attribute	-	-664.58	-337.50	-124.83	367.68	664.58

Based on the basic parameters shown in Table 6, when the data increments of CO_2, GDP, TP, OP, AE, and EC are 63.17,

3055.20, 233.38, 227.42, 58.17, and 80.47, respectively, the corresponding belief distributions of these antecedent attributes and consequent attribute can be calculated using Eq. (4) and Eq. (5) and they are shown in Table 7. Similarly, for the 156 input and output data increments, 156 groups of belief distributions are calculated to generate 156 extended belief rules.

Table 7. Belief distributions of an extended belief rule for Shanghai's EC increment prediction

Indicator	Correspondent relationship	Very Low	Low	Medium	High	Very High
CO_2	Antecedent attribute	0.0000	0.0000	0.7026	0.2974	0.0000
GDP	Antecedent attribute	0.0000	0.0000	0.0000	0.9149	0.0851
TP	Antecedent attribute	0.0000	0.1637	0.8363	0.0000	0.0000
OP	Antecedent attribute	0.0000	0.0000	0.0000	0.9572	0.0428
AE	Antecedent attribute	0.0000	0.0000	0.2654	0.7346	0.0000
EC	Consequent attribute	0.0000	0.0000	0.4209	0.5791	0.0000

Step 4: To generate rule weights for 156 extended belief rules.

According to **Step 4** in Section 3.2, the rule weight of 156 extended belief rules should be calculated based on the belief distributions obtained from **Step 3**. After calculating SRA and SRC using Eqs. (6)-(7) and inconsistency degree using Eq. (8), the corresponding 156 rule weights can be obtained using Eq. (9), and their weight ranges are shown in Fig. 5. From Fig. 5, it is clear that the weight of almost extended belief rules, namely 103 of 156 rules, locates in the range [0.995, 1.000], which indicates that all these extended belief rules are important for the EBRBS constructed by using the representative indicators and the data increments.
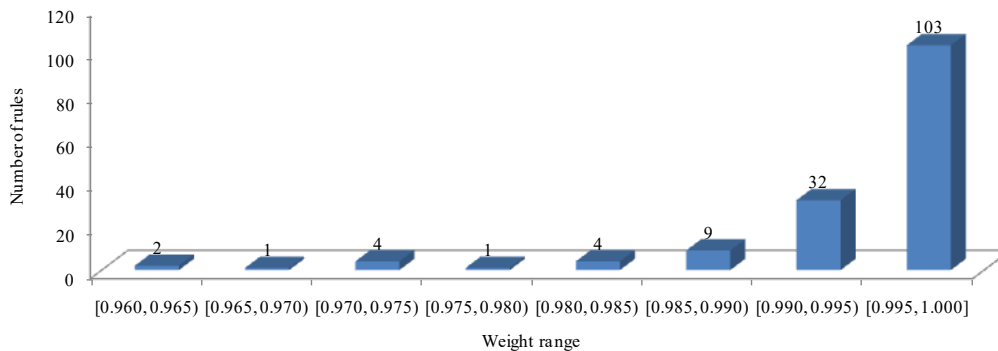

Fig. 5. Weight distribution of 156 extended belief rules

4.2.2. ER rule-based combination process for ensemble EBRBS model

In this section, the ER rule-based combination process is performed to construct ensemble EBRBS investments on the basis of the EBRBS-based construction process. The detailed steps are provided as follows:

Step 1: To construct four EBRBSs based on different feature selection methods.

Continuing with the increment prediction of Shanghai's EC to show **Step 1** in Section 3.3, apart from the EBRBS constructed in Section 4.2.1, another three EBRBSs can be constructed based on the representative indicators selected by SLR, RA, and CCSD methods shown in Table 4. In order to make differences among these four EBRBSs, which are signed as PCC-EBRBS, SLR-EBRBS, RA-EBRBS, and CCSD-EBRBS in the context below, Fig. 6 shows the predicted EC increments of four EBRBSs obtained from Eq. (13) to Eq. (17) and the actual EC increments regarding Shanghai. It can be found from Fig. 5 that: 1) the predicted EC increments of four EBRBSs and actual EC increments have a high fitting degree for 156 input-output data increment pairs; 2) four EBRBSs have different advantages in predicting EC increments, *i.e.*,

CCSD-EBRBS has a better fitting degree than other three EBRBSs between the 1st and the 32th training data; RA-EBRBS has a better fitting degree than other three EBRBSs between the 140th and the 156th training data.
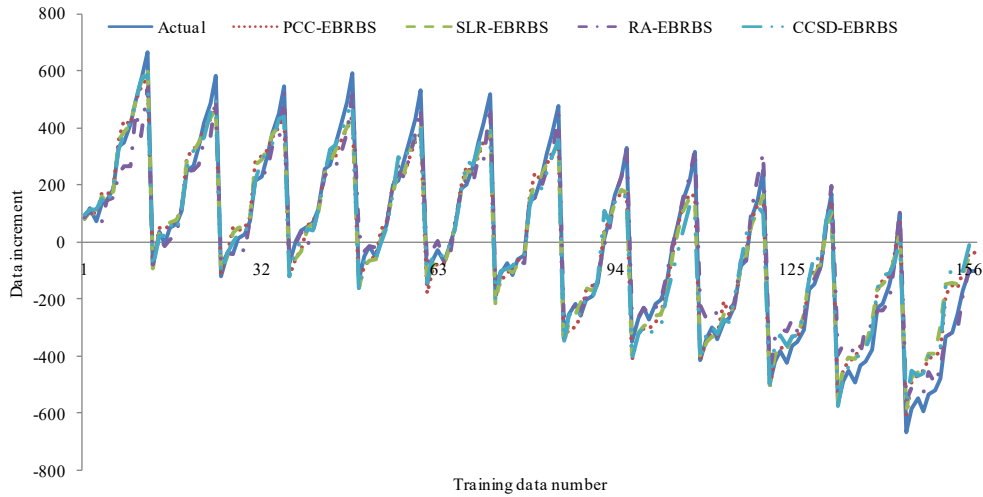

Fig. 6. Predicted Shanghai's EC increments of four EBRBSs

Step 2: To calculate reliabilities for four EBRBSs.

According to Eq. (11) at **Step 2** shown in Section 3.3, 156 input-output data increment pairs are used to calculate the reliability of PCC-EBRBS, SLR-EBRBS, RA-EBRBS, and CCSD-EBRBS. Fig. 7 shows the reliability of four EBRBSs and they are 0.9676, 0.9693, 0.9667 and 0.9684, respectively. These reliabilities indicate that the four EBRBSs have a good ability to accurately predict the EC increments of Shanghai.
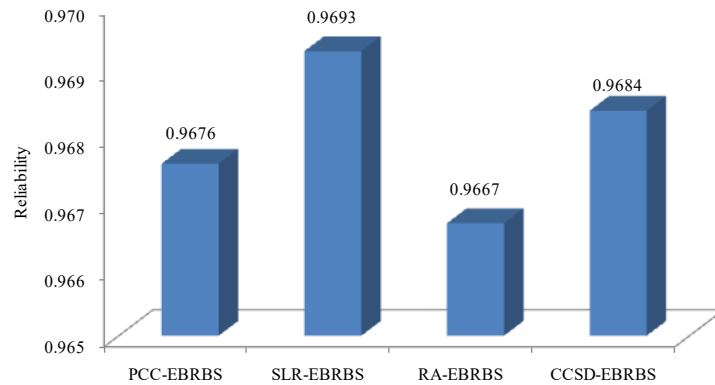
Fig. 7. Reliability of four EBRBSs for Shanghai's EC increment prediction

Step 3: To train the weights of four EBRBSs and the utilities of ensemble EBRBS model.

According to **Step 3** in Section 3.3, the weight of four EBRBSs and the utilities of ensemble EBRBS model should be optimized using the optimization model shown in Eq. (12). Here, the DE-based parameter learning algorithm (Yang *et al.*, 2019) is also used to obtain the optimized weights and utilities by setting the number of iterations and individuals as 200 and 20, respectively, in which Table 8 and Table 9 show the optimized weights and utilities.

Table 8. Optimized weights of four EBRBSs for Shanghai's EC increment prediction

Name	PCC-EBRBS	SLR-EBRBS	RA-EBRBS	CCSD-EBRBS
Weight	0.0051	0.1682	0.0697	0.0001

Table 9. Optimized utilities of ensemble EBRBS model for Shanghai's EC increment prediction

Name	Very Low	Low	Middle	High	Very High
Utility	-664.58	-325.02	-5.16	469.25	664.58

On the basis of the reliabilities shown in Fig. 7 and the optimized utilities and weights shown in Tables 8 and 9, the predicted EC increments of the ensemble EBRBS model can be obtained according to Eqs. (18)-(28) at **Step 3** shown in Section 3.4. Fig. 8 shows the predicted EC increments by the ensemble EBRBS model and the actual EC increments for Shanghai's 156 input- output data increment pairs. It can be found from Fig. 8 that the ensemble EBRBS model has a high fitting degree to predict the EC increments of Shanghai's 156 data.
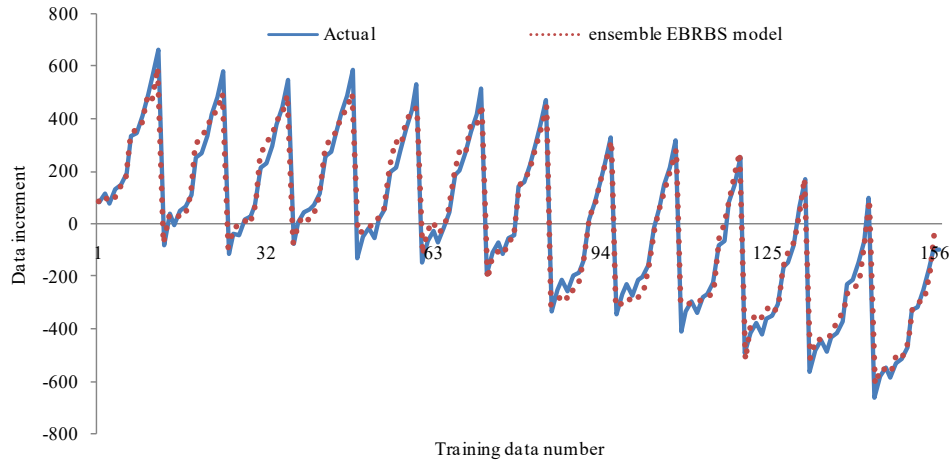
Fig. 8. Predicted Shanghai's EC increments of ensemble EBRBS model

Additionally, in order to further compare the ensemble EBRB models with four EBRBSs, the MAE of Shanghai's EC increments is calculated using 156 input-output data increment pairs. As shown in Fig. 9, the ensemble EBRBS model has the smallest MAE, namely 29.55, and PCC-EBRBS, SLR-EBRBS, RA-EBRBS, and CCSD-EBRBS have a larger MAE, namely, 43.12, 41.98, 40.77, and 44.3, respectively, indicating that the ER rule-based combination process is able to ensure the accuracy of the ensemble EBRBS model better than the EBRBS with singleton feature selection methods.
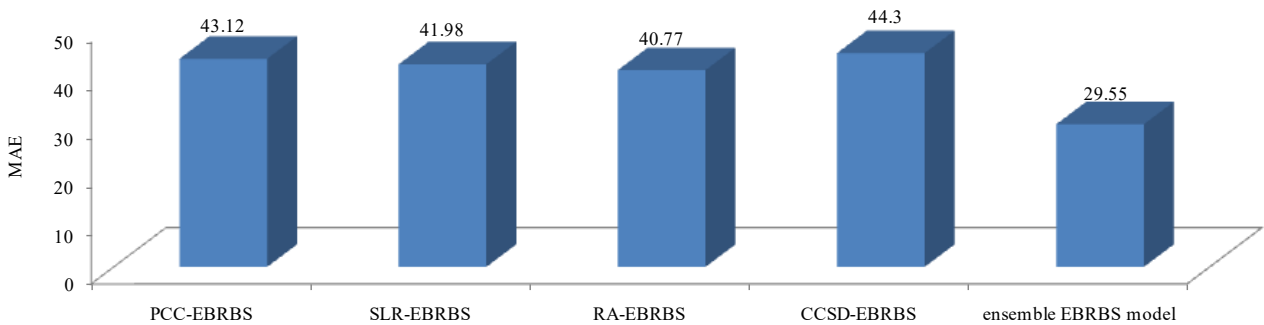

Fig. 9. MAE of Shanghai's EC increment prediction

4.3. Investment prediction results of ensemble EBRBS model

In this section, the investment prediction of five output indicators, namely, EC, CI, LI, IPT, and IEP, for 31 provinces in China is introduced to demonstrate the effectiveness of the ensemble EBRBS model. Through the same development procedure shown in Section 4.2, five ensemble EBRBS models can be constructed when the environmental data of input and output indicators from 2005 to 2017 are used as the training data. In order to identify the prediction performance of the ensemble EBRBS model on unseen data, the environmental data in 2017 and 2018 are used as historical input-output data pair and new input data, respectively. In other words, based on Section 3.4, when the input and output environmental data of 2017 are assumed to be $<x_{2017}, y_{2017}>$ and the input environmental data of 2018 to be x_{2018}, the input of the ensemble EBRBS model is $x_{2018}-x_{2017}$ and the predicted investment in 2018 is $F(x_{2018}-x_{2017})+y_{2017}$, in which $F(x_{2018}-x_{2017})$ is the

predicted investment increments of the ensemble EBRBS models. Additionally, in order to provide more comprehensive investment prediction results, Fig. 10 shows the number of input indicators to be selected as representative indicators in the process of constructing five ensemble EBRBS models for five investments, and Fig. 11 shows the comparison of actual investments and the predicted investments of the five ensemble EBRBS models for 31 provinces in 2018.
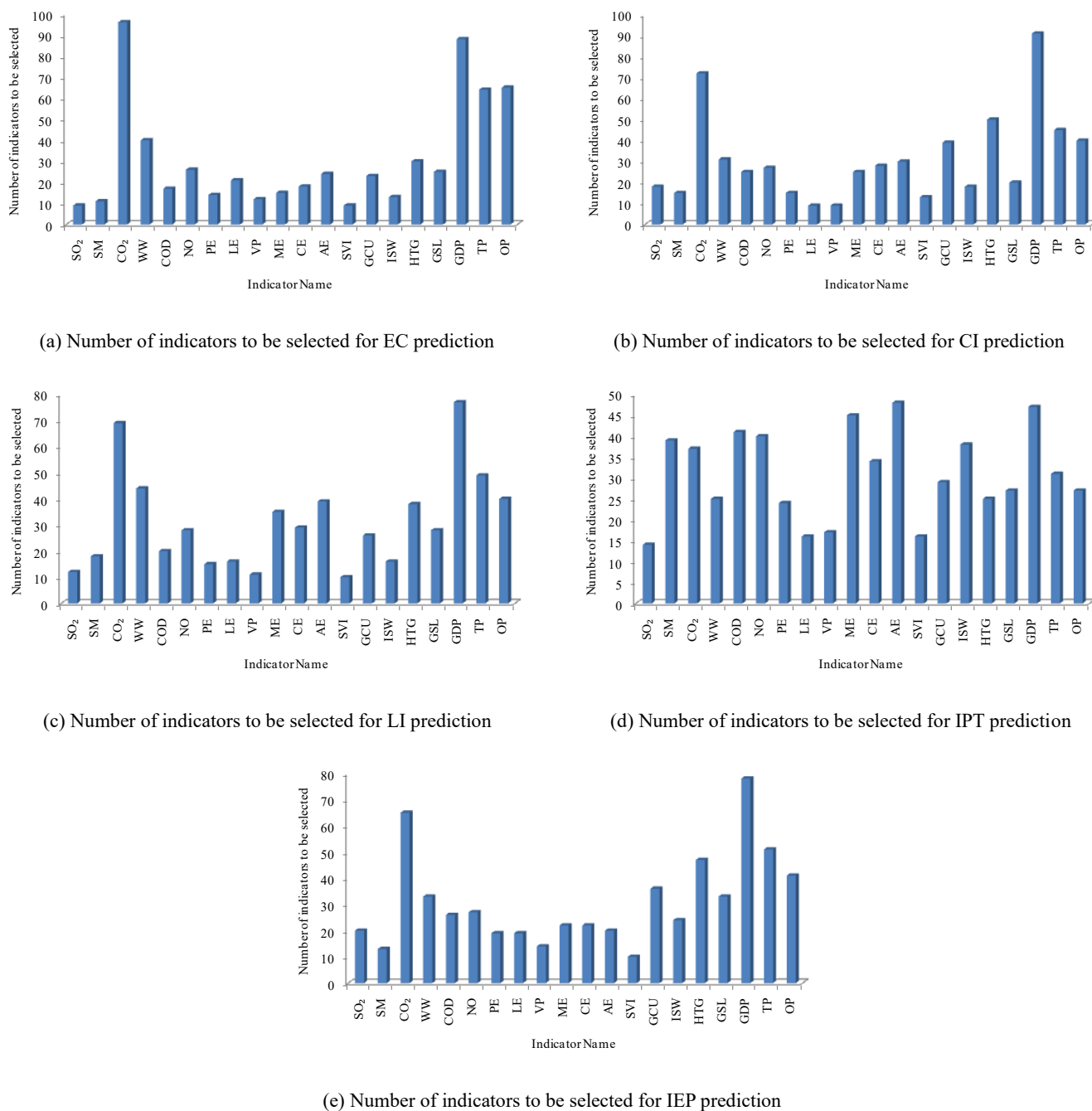

(a) Number of indicators to be selected for EC prediction

(b) Number of indicators to be selected for CI prediction

(c) Number of indicators to be selected for LI prediction

(d) Number of indicators to be selected for IPT prediction

(e) Number of indicators to be selected for IEP prediction

Fig. 10. Number of indicators to be selected in constructing ensemble EBRBS models

From Fig. 10, it is clear that all of twenty input indicators can be selected at least once for each investment prediction by the four feature selection methods, which means all of these input indicators have different influences on the prediction of environmental investments. However, in almost all previous studies, some of the twenty input indicators were selected to construct investment prediction models, which would decrease the accuracy of predicted investments. Hence, it is necessary to consider diverse indicators for environmental investment prediction. Additionally, the following three remarks are noted according to the order of input indicators shown in Fig. 10:

(1) The input indicators CO_2 and GDP are two most popular indicators selected for the prediction of five environmental

investments, which prove that the ensemble EBRBS model can provide a practical and scientific indicator selection because China government has been committed to reducing the emissions of CO_2 in a new round of environmental policies, and GDP is the main indicator to measure the economic development benefits. This is also in line with the current situation of China's environmental pollution emissions and economic development, that is, carbon emissions are one of the important sources of industrial pollution and the main cause of global warming, and improving economic development is the primary goal of developing countries. Therefore, the consideration of these two factors is also the key to effective investment in environmental management.

(2) The input indicators CO_2, GDP, TP, and OP are four popular indicators used for EC prediction. The main reason is that reducing carbon emission has always been a key consideration in the process of energy consumption, and energy consumption is also an important factor that significantly affects GDP and industrial outputs in the development of economic. Energy is one of the main driving forces of the current economic development. The rapid development of economy requires continuous consumption of energy, but at the same time, energy consumption also means the increase of carbon emissions. This explains why CO_2 and related economic indicators are representative input indicators for EC investment prediction.

(3) The indicators selected for IPT prediction are different from other investment predictions. The reason is that industrial production involves the discharge of many pollutants, such as waste gas, wastewater, and solid waste, thus industrial production has become one of important sources of environmental pollution in China. Therefore, the number of input indicators selected for IPT prediction is significantly higher than other environmental investments, and this phenomenon also further shows that in the current development process of China, the pollution control in industrial production has gradually become the key factor in environmental management. In the future, with the improvement of environmental treatment effect, the demand for the control of industrial pollution emission and the application of relevant green production technology will be higher.
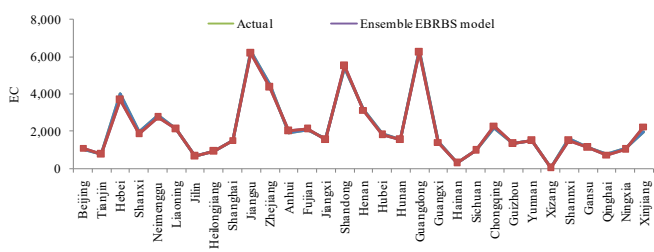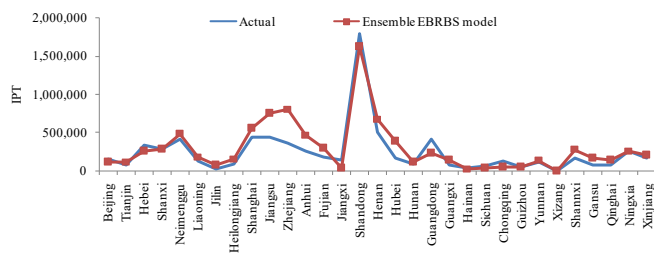

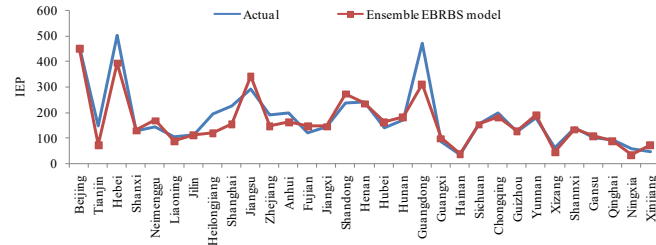(a) Comparison of predicted EC and actual investments
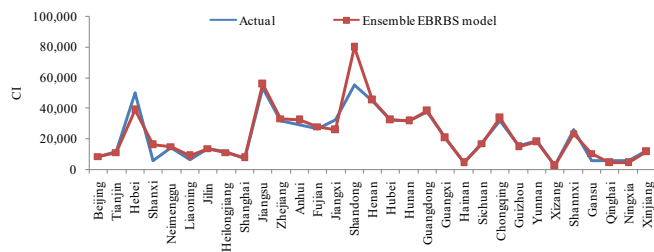
(b) Comparison of predicted CI and actual investments
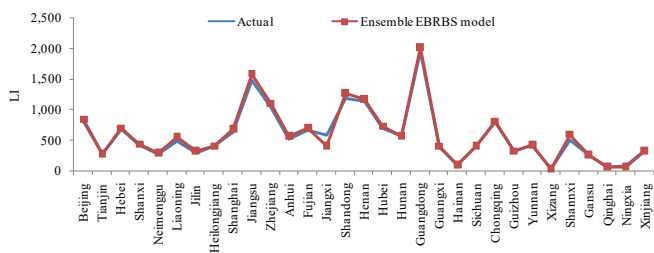
(c) Comparison of predicted LI and actual investments

(d) Comparison of predicted IPT and actual investments

(e) Comparison of predicted IEP and actual investments

Fig. 11. Comparison of predicted and actual investments of 31 provinces in 2018

From Fig. 11, it is obvious that the predicted investments of the five ensemble EBRBS models basically match the actual investments. Although the predicted investment of CI in Shandong provinces and the predicted investment of IPT in Jiangsu, Zhejiang, and Anhui provinces have a little different to the actual investment of those provinces, the predicted investment of the remaining provinces matches the actual investment accurately. From the view of regional difference, the investments of EC in Shandong, Guangdong, and Jiangsu provinces are obviously higher than other provinces, and the investments of EC in most western regions are lower than the eastern region in China, the reason is that the most eastern provinces have large population and industry, especially for the population and industry development in Shandong, Guangdong, and Jiangsu provinces, while the industry development of most western regions are far behind the other regions and the population density is also lower than the eastern regions.

Additionally, the investments of CI and LI shown in Fig. 11 indicate that most eastern provinces in China have higher investments than the most provinces located in Western China, which reveal that environmental investments have a great relationship with population, economy, and the government's emphasis on environmental management in China. It is worth noting that the investment of IPT in Shandong province is higher than the other provinces, this is because Shandong is a traditional industrial province, leading to the fact that its proportion of industry and output of industrial products are higher than most provinces. It is also one of China's three major economic centers, leading in industrial output, population and economic development. Moreover, Fig. 11(e) also indicates that the investment of IEP has a close relationship with regional economic and population, in which IEP in most economic development regions of Beijing, Hebei, Jiangsu and Guangdong are higher than the other provinces, The reason is that with the improvement of economic development level, the government will gradually increase the investment in environmental management, and policy makers will also pay more attention to environmental pollution control.

4.4. Comparative analysis of existing investment prediction models

In order to compare the accuracy of ensemble EBRBS model with some existing prediction models, including ARIMA -based model (Valipour *et al*., 2013), GM-based model (Chen *et al*., 2017), adaptive neural fuzzy inference system (ANFIS) -based model (Ye *et al*., 2020a), and EBRBS-based model (Wang *et al*., 2020), in which the former two models are the typical models based on time series and the latter two ones are the typical models based on input-output relationship in the previous studies. In addition, it is worth noting that, compared to environmental investment prediction using time series-based prediction model, the input-output relationship-based prediction models have been gradually applied in environmental investment prediction in recent years, and the ANFIS-based and EBRBS-based models are the recent models proposed for environmental investment prediction. Therefore, the above-mentioned four existing environmental investment prediction models are introduced to compare with the ensemble EBRBS model and their comparative results are shown in Table 10, in which the MAE, MAPE, and R are calculated by using Eq. (29) to Eq. (32) for 31 provinces in China for 2018.

Table 10. Comparison of different investment prediction models for five investments

	Prediction models	EC	CI	LI	IPT	IEP	Average rank
MAE	GM	666 (4)	19,613 (4)	651 (5)	2,161,347 (5)	445 (5)	4.6
	ARIMA	1,361 (5)	36,182 (5)	304 (4)	220,052 (3)	235 (4)	4.2
	ANFIS	202 (2)	2,684 (2)	73 (2.5)	257,401 (4)	39 (3)	2.7
	EBRBS	333 (3)	4,878 (3)	73 (2.5)	113,427 (2)	36 (2)	2.5
	Ensemble EBRBS	69 (1)	2,664 (1)	33 (1)	94,962 (1)	29 (1)	1
MAPE	GM	36% (4)	160% (4)	67% (5)	1,885% (5)	314% (5)	4.6
	ARIMA	71% (5)	203% (5)	42% (4)	136% (3)	138% (4)	4.2
	ANFIS	13% (2)	20% (2)	12% (2)	183% (4)	24% (3)	2.6
	EBRBS	18% (3)	28% (3)	15% (3)	58% (2)	20% (2)	2.6
	Ensemble EBRBS	4% (1)	17% (1)	5% (1)	54% (1)	16% (1)	1
R	GM	0.9213 (4)	0.5314 (5)	0.8404 (5)	0.8378 (3)	0.4258 (5)	4.4
	ARIMA	0.9644 (3)	0.9098 (3)	0.8945 (4)	0.9330 (1)	0.7786 (4)	3
	ANFIS	0.9835 (2)	0.9542 (1)	0.9607 (2)	0.5958 (5)	0.8990 (3)	2.6
	EBRBS	0.9063 (5)	0.8532 (4)	0.9351 (3)	0.8102 (4)	0.9185 (2)	3.6
	Ensemble EBRBS	0.9982 (1)	0.9471 (2)	0.9952 (1)	0.9242 (2)	0.9273 (1)	1.4

Table 10 shows the comparison of different models in investment prediction based on MAE, MAPE, R, and their ranks, where these existing models are constructed by using the same training and testing data and the top five input indicators. From Table 10, the predicted values of the ensemble EBRBS model are close to the actual values of five investments and better than other models, and the MAEs of the ensemble EBRBS model are 69, 2664, 33, 94962, and 29, respectively. The MAEs of EBRBS-based model are higher than those of the ensemble EBRBS model, which reveals the ability of the ensemble EBRBS model to overcome the challenges shown in Section 2.3. Moreover, the comparison of MAPEs and Rs also shows that the ensemble EBRB model produces satisfactory results compared to the existing four models, in which the MAPEs of GM-based and ARIMA-based models are much higher than the ensemble EBRBS models, especially for the prediction of CI and the prediction errors of IPT and IEP. Although the R of ANFIS-based model for CI prediction and ARIMA-based model for IPT prediction are slightly higher than that of the ensemble EBRBS model, the whole prediction accuracy of ensemble EBRBS model can outperform all listed studies. All in all, the above comparative analysis of different prediction models indicates that the ensemble EBRBS model is ability to accurately predict environmental investments and has a higher accuracy than other existing models.

5. Conclusions

In the present work, a novel ensemble EBRBS model based on EBRBS and the ER rule was proposed to enhance environmental investment prediction. The main components of the ensemble EBRBS model include multiple EBRBSs and an ER rule-based combination method. A case study based on the actual environmental data of 31 provinces in China illustrated the effectiveness of the proposed ensemble EBRBS model. The main contributions of this study can be further summarized as three aspects below:

(1) Considering that there are various categories of indicators related to environmental investment prediction and all these environmental indicators should be considered in constructing an effective prediction model, an ensemble framework

with different indicator selection methods was proposed on the basis of various sets of selected indicators to construct a number of EBRBSs. Hence, this ensemble framework can integrate the strengths of different feature selection methods and overcome the over-fitting problem caused by too many indicators used to construct prediction models.

(2) Because of the difficulty in collecting data for the construction of an effective prediction model, data increment was proposed to transform original environmental data into environmental data increments, so that there are sufficient training data for constructing EBRBSs and the ensemble EBRBS model. More importantly, the data increment transformation is a reversible process, so the ensemble EBRBS model can be used to predict environmental investments while the output of the EBRBS and ensemble EBRBS model is investment increments.

(3) The ER rule was used to integrate different EBRBSs into an ensemble model with consideration of their weights and reliabilities. Since there are many uncertain and objective factors in the actual environmental management and environmental data processing, it is necessary and reasonable to consider both reliability and weight of each EBRBS in the ensemble model. Consequently, the predicted investments of the ensemble model are consistent with actual environmental management and have a desired fitting degree compared to actual investments.

In the case study, the development procedure of the ensemble EBRBS model and the investment prediction of 31 provinces in China were provided, and the main conclusions can be summarized as follows:

Firstly, from the comparison between the predicted and actual environmental investments, it can be found that the ensemble EBRBS model has an excellent fitting degree between predicted and actual environmental investments in China, and the regional differences of the prediction results of 31 provinces in China are consistent with the regional differences of the actual environmental investments.

Secondly, in the target of further illustrating the accuracy of the ensemble EBRBS model, the prediction results of the ensemble EBRBS model were compared with the time series and input-output relationship-based prediction models that commonly used in the previous studies. The comparison results revealed that the ensemble EBRBS model has a powerful ability in predicting environmental investments compared to other existing prediction models.

Thirdly, the prediction results of the ensemble EBRBS model showed that the regional distribution of environmental investment in China is basically consistent with the regional distribution of economic development and population density, that is, the higher the level of economic development and population density are, the higher the government's investments in environmental governance have.

Finally, owing to the above contributions in environmental investment prediction using the ensemble EBRBS model, the decision makers, who are responsible for environmental management, obtain a scientific and reasonable tool for making regional environmental investment planning by considering representative indicators and model reliability under the limited number of environmental data.

Cross-regional characteristics are the important factors in environmental management, however, the regional polices, technological innovations, economic development, and other factors that may affect environmental investment prediction are not considered in the prediction model construction in this study. Thus, some future researches can be focused on investigating the influences of environmental regional factors on the investment predictions. Furthermore, regional cooperation is also a hot issue in pollution control and environmental protection in regional investment prediction, future researches can be also devoted to investment prediction under regional cooperation. Finally, the research objects of the present work have not yet involved the investment prediction within different industries, the ensemble model was just

limited to the environmental management of different regions. Therefore, the ensemble model proposed in this paper can be used in the investment prediction of different enterprises or industries in the future works.

Acknowledgements

Reference:

[1]   Cao H J, Fujii H, Managi S, 2015. A productivity analysis considering environmental pollution and diseases in China. *Journal of Economic Structures*, 4(6):11-25. https://doi.org/10.1186/s40008-015-0012-9

[2]   Costa-Campi M T, Garcia-Quevedo J, Martínez-Ros E, 2017. What are the determinants of investment in environmental R & D. *Energy Policy*, 104, 455-465. https://doi.org/10.1016/j.enpol.2017.01.024

[3]   Chen L., Wang Y. M., Lai F.J., Feng F., 2017. An investment analysis for China's sustainable development based on inverse data envelopment analysis. *Journal of Cleaner Production*, 142, 1638-1649. https://doi.org/10.1016/j.jclepro.2016.11.129

[4]   Deus R. M., Mele F. D., Bezerra B. S., Battistelle R. A. G., 2020. A municipal solid waste indicator for environmental impact Assessment and identification of best management practices, *Journal of Cleaner Production*, 242: 1-14. https://doi.org/10.1016/j.jclepro.2019.118433

[5]   Franke V. D. M., 2018. How knowledge enables governance: The coproduction of environmental governance capacity. *Environmental Science and Policy*, 87, 18-25. https://doi.org/10.1016/j.envsci.2018.05.016

[6]   Hamidi K., Sedghi H., Telvari A., Babazadeh H., 2018. Modeling climate variables of rivers basin using time series analysis (case study: Karkheh River basin at Iran). *Civil Engineering Journal*, 4, 78-86. DOI: 10.28991/cej-030970

[7]   Heslouina C., Perrot-Bernardet V., Cornier A., Perry N., 2017. A user oriented framework to support environmental performance indicators selection. Procedia CIRP, 61,709-714. https://doi.org/10.1016/j.procir.2016.11.211

[8]   Issa I. I., Pigosso D. C.A., McAloone T.C., Rozenfeld H., 2015. Leading product-related environmental performance indicators: a selection guide and database. *Journal of Cleaner Production*, 108, 321-330. https://doi.org/10.1016/j.jclepro.2015.06.088

[9]   Jiang Q., Liu Z., Liu W., Li T., Cong W., Zhang H., Shi J., 2018, A principal component analysis based three-dimensional sustainability assessment model to evaluate corporate sustainable performance, *Journal of Cleaner Production*, 187: 625-637. https://doi.org/10.1016/j.jclepro. 2018.03.255

[10]   Kaytez F., 2020, A hybrid approach based on autoregressive integrated moving average and least-square support vector machine for long-term forecasting of net electricity consumption, *Energy*, 197: 1-12. https://doi.org/10.1016/j.energy.2020.117200

[11]   Lee M, Zhang N, 2012. Technical efficiency, shadow price of carbon dioxide emissions, and substitutability for energy in the Chinese manufacturing industries. *Energy Economics*, 34 (5), 1492–1497. https://doi.org/10.1016/j.eneco.2012.06.023

[12]   Li X G, Yang J, Liu X J, 2013. Analysis of Beijing's environmental efficiency and related factors using a DEA model that considers undesirable outputs. *Mathematical and Computer Modelling*, 58, 956–960. https://doi.org/10.1016/j.mcm.2012.10.016

[13]   Liu J., Martínez L., Calzada A., Wang H., 2013. A novel belief rule base representation, generation and its inference methodology. *Knowledge-Based Systems*, 53, 129-141. https://doi.org/10.1016/j.knosys.2013.08.019

[14]   Long X, Chen B, Park B, 2018. Effect of 2008's Beijing Olympic Games on environmental efficiency of 268 China's cities. *Journal of Cleaner Production*, 172, 1423-1432. https://doi.org/10.1016/j.jclepro.2017.10.209

[15] Mulenga D., Siziya S., 2019. Indoor Air Pollution Related Respiratory Ill Health, a Sequel of Biomass Use. *SciMedicine Journal*, 1, 30-38. DOI: 10.28991/SciMedJ-2019-0101-5

[16] Niemeijer D. S., de Groot R., 2008. A conceptual framework for selecting environmental indicator sets. *Ecological indicators*, 8, 14-25. https://doi.org/10.1016/j.ecolind.2006.11.012

[17] OECD. OECD environmental indicators, development, measurement and use-Reference paper. Paris, 2003.

[18] Roshani A., Gerami M., 2018. New Rethinking on Managers' Competency Criteria and Success Factors in Airport Construction Projects. *Civil Engineering Journal*, 4, 26-32. DOI: 10.28991/cej-03091192

[19] Salcedo-Sanz S., Cornejo-Bueno L., Prieto L., Paredes D., Garcia-Herrera R., 2018, Feature selection in machine learning prediction systems for renewable energy applications, *Renewable and Sustainable Energy Reviews*, 90: 728-741. https://doi.org/10.1016/j.rser.2018.04.008

[20] Song M. L., Cao S. P., Wang S. H., 2019. The impact of knowledge trade on sustainable development and environment-biased technical progress. *Technological Forecasting & Social Change*, 144, 512-523. https://doi.org/10.1016/j.techfore.2018.02.017

[21] Song M. L., Peng J., Wang J. L., Dong L., 2018. Better resource management: An improved resource and environmental efficiency evaluation approach that considers undesirable outputs. *Resources, Conservation and Recycling*, 128, 197-205. https://doi.org/10.1016/j.resconrec.2016.08.015

[22] Tong C., Ding S., Wang B., Yang S. L., 2020, Assessing the target-availability of China's investments for green growth using time series prediction, *Physica A*, 537: 1-11. https://doi.org/10.1016/j.physa.2019.122724

[23] Valipour M., Banihabib M. E., Behbahani S. M. R., 2013. Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir. *Journal of Hydrology*, 476, 433-441. https://doi.org/10.1016/j.jhydrol.2012.11.017

[24] Wang H, Jiang G Z, Wang Y, Zhang H, Wang Y H., 2018. A two-stage optimization method for energy-saving flexible job-shop scheduling based on energy dynamic characterization. *Journal of Cleaner Production*, 188, 575-588. https://doi.org/10.1016/j.jclepro.2018.03.254

[25] Wang S., Ye F. F., 2020, Environmental Governance Cost Prediction of Transportation Industry by Considering the Technological Constraints, *Symmetry-BASEL*, 12: 1-15. https://doi.org/10.3390/sym12081352

[26] Wang Y. M., Ye F. F., Yang L. H., 2020. Extended belief rule based system with joint learning for environmental governance cost prediction. *Ecological Indicators*, 111, 1-14. https://doi.org/10.1016/j.ecolind.2020.106070

[27] Wu J, Li M J, Zhu Q Y, Zhou Z X, Liang L, 2019. Energy and environmental efficiency measurement of China's industrial sectors: A DEA model with non-homogeneous inputs and outputs. *Energy Economics*, 78, 468-480. https://doi.org/10.1016/j.eneco.2018.11.036

[28] Xu N., Dang Y. G., Gong Y. D., 2017. Novel grey prediction model with nonlinear optimized time response method for forecasting of electricity consumption in China. *Energy*, 118, 473-480. https://doi.org/10.1016/j.energy.2016.10.003

[29] Yang J. B., 2001, Rule and utility based evidential reasoning approach for multiattribute decision analysis under uncertainties, European Journal of Operational Research, 131(1): 31-61. DOI: 10.1016/S0377-2217(99)00441-5.

[30] Yang J. B., Xu D. L., 2013. Evidential reasoning rule for evidence combination. *Artificial Intelligence*, 205, 1-29. https://doi.org/10.1016/j.artint.2013.09.003

[31] Yang L. H., Wang Y. M., Chang L. L., Fu Y. G., 2017. A disjunctive belief rule-based expert system for bridge risk assessment with dynamic parameter optimization model. *Computers & Industrial Engineering*, 113, 459-474. https://doi.org/10.1016/j.cie.2017.09.027

[32] Yang L. H., Wang Y. M., Liu J., Martínez L., 2018a. A joint optimization method on parameter and structure for belief-rule-based systems. *Knowledge-Based Systems*, 142, 220-240. https://doi.org/10.1016/j.knosys.2017.11.039

[33] Yang L. H., Wang Y. M., Fu Y. G., 2018b. A consistency analysis-based rule activation method for extended belief-rule-based systems. *Information Sciences*, 445, 50-65. https://doi.org/10.1016/j.ins.2018.02.059

[34] Yang L. H., Liu J., Wang Y. M., Martínez L., 2019, New activation weight calculation and parameter optimization for extended belief rule-based system based on sensitivity analysis, *Knowledge and Information Systems*, 60(2): 837-878. https://doi.org/10.1007/s10115-018-1211-0

[35] Yang L. H., Ye F. F., Wang Y. M., 2020, Ensemble belief rule base modeling with diverse attribute selection and cautious conjunctive rule for classification problems, *Expert Systems with Applications*, 146: 1-14. DOI: 10.1016/j.eswa.2019.113161

[36] Ye F. F., Yang L. H., Wang Y. M., 2019a. A new environmental governance cost prediction method based on indicator synthesis and different risk coefficients. *Journal of Cleaner Production*, 212, 548-566. https://doi.org/10.1016/j.jclepro.2018.12.029

[37] Ye F. F., Yang L. H., Wang Y. M., 2019b. Fuzzy rule based system with feature extraction for environmental governance cost prediction. J*ournal of Intelligent & Fuzzy Systems*, 37, 2337-2349. DOI: 10.3233/JIFS-182628

[38] Ye F. F., Yang L. H., Wang Y. M., 2020a. A cost forecast method of environmental governance based on the input-output relationship and efficiency. *Control and Decision*, 35(4): 993-103. DOI: 10.13195/j.kzyjc.2018.0622. (In Chinese)

[39] Ye F. F., Yang L. H., Wang Y. M., 2020b. An interval efficiency evaluation model for air pollution management based on indicators integration and different perspectives. *Journal of Cleaner Production*, 245, 1-15. https://doi.org/10.1016/j.jclepro.2019.118945

[40] Zhang W., Li G. X., Uddin M. K., Guo S. C., 2020. Environmental Regulation, Foreign Investment Behavior, and Carbon Emissions for 30 provinces in China. *Journal of Cleaner Production*, 248: 1-11. https://doi.org/10.1016/j.jclepro.2019.119208

[41] Zheng S., He C., Hsu S. C. Sarkis J., Chen J. H., 2020, Corporate environmental performance prediction in China: An empirical study of energy service companies, Journal of Cleaner Production, 266: 1-16. https://doi.org/10.1016/j.jclepro.2020.121395

[42] Zhou Z. H. Ensemble Methods - Foundations and Algorithms, Taylor & Francis, 2012. DOI: 10.1109/MCI.2012.2228600