Scholar Publishing Group
Water Pollution Prevention and Control Project
https://doi.org/10.38007/WPPCP.2020.010104
ISSN 2790-2730 Vol. 1, Issue 1: 30-39

Evaluation on the Construction of Water Pollution Control Model Based on Random Forest

Roland Szabo[*]

Silesian University of Technology, 44-100 Gliwice, Poland

[*]*corresponding author*

Keywords: Random Forest, Water Pollution, Generalisation Error, Water Conservation

Abstract: With the continuous development of society, water pollution has become an important factor restricting people's survival and development, so it is necessary to establish a system that can effectively control water quality, ensure drinking water safety and reduce water waste. Water pollution control means to reduce water consumption, improve water quality and protect the ecological environment by designing an environmental protection measure that can effectively prevent water eutrophication. However, due to the insufficient attention paid to the assessment of water quality, many water bodies have been seriously affected. In order to solve the problems of traditional water pollution control models that rely too much on the subjective judgment and professional knowledge of experts and scholars, such as high difficulty in predicting water pollution problems and low prediction accuracy, based on reality, a stochastic forest algorithm was proposed to build control models to solve the problem. The divided interval samples were collected by computer, and then the samples were randomly classified by random forest algorithm. The calculated average generalization error was reduced by continuously introducing random variables. The final sample classification can more intuitively see the real-time state of water quality. Finally, according to the experimental data, the water pollution control model based on random forest proposed in this paper has an average increase of 13.1% in the four evaluation indicators compared with the traditional water pollution control model.

1. Introduction

Water pollution control means to protect the environment through various effective means and achieve the goal of sustainable development under the condition of abundant water resources and low per capita occupancy, so as to make pollutants migrate and transport in the water body. With the improvement of economic society and people’s living standards and the increasing awareness of environmental protection, more and more attention has been paid to the increasingly serious problem of water quality deterioration. At the same time, with the increasing requirements of human beings for water resources utilization and development technology, how to solve the water pollution problem through effective methods has become one of the serious challenges facing the world, so more relevant researchers are needed to explore this.

Some water pollution control experts have summarized the water pollution control schemes proposed in recent years, and put forward the development direction of water pollution control on this basis. By summarizing the problems of the deterioration of the earth’s ecology and combining the analysis of the impact of various social fields on the earth’s ecology, Delitic Ana found that water pollution was the main cause of the deterioration of the earth’s ecology, and determined that water pollution was the most serious problem in the field of ecological governance [1]. He Mingjing obtained a new plan for water pollution control by exploring the decontamination ability of biochar in water pollution control and combining the water pollution control effect, which opened up a new direction for water pollution control [2]. Rink Karsten collected sewage treatment and water quality data from various places through data integration environment, assessed the degree of water pollution in the collection area in combination with digital technology, emphasized the severe form of water pollution and provided a theoretical basis for the digitization of water pollution control scheme [3].

In order to explore the water pollution control scheme, Martini Sri proposed a new water pollution control scheme by introducing new technologies and methods in combination with membrane technology and adsorption method, laying a solid foundation for the future development of water pollution control [4]. Tony Maha A put forward a water pollution control plan based on the large amount of new adsorbed substances by combining the emerging low-cost adsorption technology and through the adsorption effect experiment on local water pollution, which promoted the development process of local water pollution control [5]. Ahmed Shahid, through research and investigation on local youth, combined with the local youth’s understanding and feedback on water pollution, analyzed the lack of local awareness of the serious hazards of water pollution, determined that waste dumping is the most important trigger factor in water pollution, and put forward prospects for the development of local water pollution control field [6].

Wang Linqing collected samples of local precipitation in recent years, and obtained the water-soluble composition data of local rainwater in recent years through the analysis of rainwater content. The data showed the severity of local water pollution problems and the source of water pollution problems, which opened a new direction for evaluating the effectiveness of water pollution control [7]. Combined with the construction of the difference model, Li He has adopted regional integration for local areas, and the experiment showed that the adoption of regional integration model can promote population aggregation and public management, thus effectively reducing transboundary water pollution, opening up new ideas and methods for water pollution control schemes [8]. Alizadeh Mohamad Javad obtained a more scientific and accurate water pollution assessment method by evaluating the local water quality and combining digital technology to compare and evaluate the water quality, laying a practical foundation in the assessment of water pollution control [9]. The above plans have summarized and analyzed the local water pollution problems in different degrees, but have not put forward a more complete solution.

Some other scholars and experts have explored how to improve the efficiency of water pollution control and the richness of monitoring and forecasting methods, hoping to get a more complete scheme. Bisht Anil Kumar can obtain the severity assessment data of local water pollution problems through the assessment of local water quality and the combination of artificial neural network technology. The analysis of the data can determine the feasibility of this technology in water pollution control [10]. Yang Huanhai reflected the severity of local water pollution problems by studying the normal growth of local aquatic products, and combined the prediction model based on long-term and short-term memory neural network. The analysis results showed that the technical model has high prediction accuracy and wide applicability as well as the reliability of the technology in water pollution control [11]. Barzegar Rahim studied the degree of water pollution in local lakes, and obtained data reflecting the degree of water pollution in local lakes by using the water quality detection system built by using neural network technology. The feasibility of this model was determined through analysis [12]. However, most of these studies cannot meet the current requirements for water pollution control research, so more in-depth exploration is needed.

The main purpose of this paper is to solve the problems of high difficulty and low accuracy of prediction of water pollution problems by relying too much on the subjective judgment and professional knowledge of experts and scholars under the traditional water pollution control model, summarize and analyze the existing traditional water pollution control model, and show its advantages and disadvantages. Finally, combined with the evaluation criteria of water pollution control model, a new water pollution control model is proposed. This model not only has more efficient collection efficiency and prediction effect, but also has better response ability to sudden water pollution problems.

2. Evaluation of Common Problems in Water Pollution Control Process

2.1. Industrial Water Pollution

With the progress of human science and technology and the need of social development, large-scale industrial production is essential, but in the process of industrial production, the neglected industrial wastewater is the main factor causing water pollution. Industrial wastewater is the main body of water pollution in the water area. It has the characteristics of large amount of pollution sources, complex composition and strong toxicity. The water pollution problem caused by industrial waste is the most difficult to treat and purify. Once the water area is polluted by industrial wastewater, it may suffer permanent pollution damage. Once such polluted water is drunk, it would cause serious consequences. In addition, industrial pollution also includes thermal pollution. Abnormal temperature rise would cause harm to organisms and components in natural water resources, resulting in biological loss, and the water area would lose its activity in the long run.

2.2. Agricultural Water Pollution

Agricultural water pollution is mainly agricultural wastewater generated in the process of livestock breeding. The composition of agricultural wastewater is very complex, such as fecal wastewater, pesticide wastewater and fertilizer wastewater. The purity of water area would be affected by the surrounding agriculture, resulting in a high content of pathogenic microorganisms and agricultural by-products. Water bodies that are injected with agricultural wastewater usually have blooms, red tides and other phenomena that make water bodies eutrophic, which would directly affect the ecological balance of water bodies and normal water use of people, thus affecting the local economic development.

2.3. Domestic Water Pollution

Domestic water pollution is usually produced by the use of household, formula, commercial and urban public equipment. Due to the complex composition of pollution sources, the wastewater generated by domestic water pollution would also contain various pollutants. In particular, the extensive use of commercial sewage and detergents has significantly increased the number and types of pollutants in domestic sewage. The more complex types of pollution sources cause the more serious anoxic phenomenon in the water. Drinking for a long time would cause great harm to human health.

The classification of common pollution sources of water pollution is shown in Figure 1.

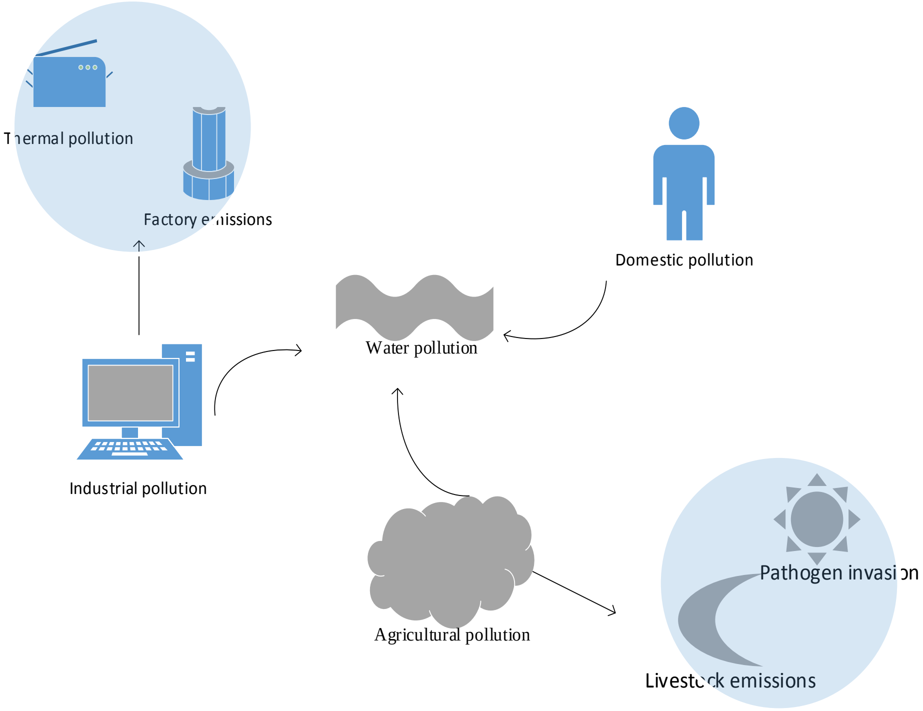


Figure 1. Classification of common sources of water pollution

3. Evaluation of Water Pollution Control Development

Water resources are the precious wealth given to mankind by nature and an important resource for human reproduction. However, due to people's wanton large-scale production activities and uncontrolled promotion of urbanization, a large amount of industrial wastewater and domestic garbage left in the process of social progress have caused serious pollution and impact on water resources. However, in the process of human production and life, harmful polluting substances invade into the water body. This phenomenon is called water pollution. Water pollution would not only reduce water resources, but also cause irreversible damage to the ecosystem. Therefore, it is necessary to take strict control measures on water pollution, monitor and predict the adverse effects of water pollution in a timely manner, and prevent the destruction of people's happy life.

In recent years, with the frequent occurrence of water pollution events, the water pollution problems that have been despised have been exposed to the public. The detailed classification, monitoring and prediction of water pollution problems are the main research directions in the field of water pollution control. In the traditional water resources control plan, most of them rely on the professional knowledge and subjective judgment of experts and scholars. It is difficult to find hidden dangers before the water pollution problem occurs, thus it is difficult to recover the loss of property and resources caused by the water pollution problem. In the process of water pollution control, it is also difficult to implement the formulated plan perfectly. On the basis of the tradition, based on the reality, this paper tentatively introduces the random forest algorithm to build an effective water pollution control model to monitor and predict the occurrence of water pollution problems and respond in time. The water pollution control structure is shown in Figure 2.

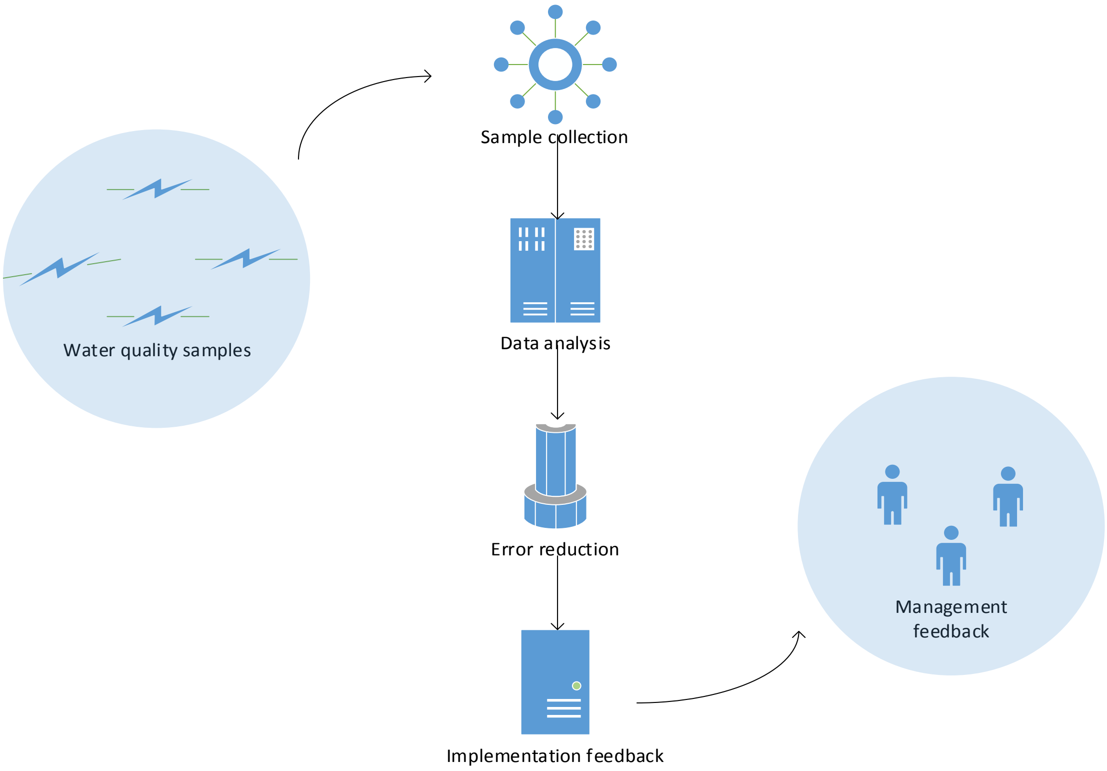


Figure 2. Water pollution control structures

4. Algorithm Evaluation of Random Forest

Stochastic forest is an efficient ensemble learning algorithm. Based on decision tree, it can effectively alleviate over-fitting, has high prediction accuracy and a broader tolerance for outliers and noise, and is widely used in various fields [13]. In the algorithm structure of random forest, after integrating the integrated learning environment based on the decision tree, the cluster selection of random attributes is introduced from the training process of the decision tree, and a relatively unified attribute classification is finally obtained [14]. Assuming that the sample set in the decision-making training process is obtained from the random attributes A and B, the prediction mean square generalization error $\sigma^{\sim}$ of the random forest can be calculated by Formula (1).

$$\sigma^{\sim} = \sigma_{A,B}(B-P(A))^2 \tag{1}$$

Among them, $P(A)$ is the correlation between random attributes and decision trees. When the number of decision trees tends to infinity, the mean square generalization error of random attributes can be obtained, which is recorded as I, as shown in Formula (2).

$$\sigma_{A,B}(B-x_{v_e}P(A,\tau_e))^2 \rightarrow I = \sigma_{A,B}(B-\sigma_\tau P(A,\tau))^2 \tag{2}$$

In Formula (2), x_{v_e} is the average function; τ_e is the random variable of the third decision tree, and σ_τ is the expected function. Therefore, the regression function of random forest can be calculated, as shown in Formula (3).

$$B = \sigma_{\tau} P(A, \tau) \tag{3}$$

After calculating the regression function of the random forest, the average generalization error of a decision tree in the random forest can be calculated by Formula (4), which is recorded as O.

$$O = \sigma_{\tau} \sigma_{A,B} (B - P(A, \tau))^2 \tag{4}$$

The random variable τ of different decision trees is independent of each other, so the generalization of the random forest regression prediction depends on the correlation of the residual and the accuracy of the decision tree. The error is reduced by introducing random variables to reduce the correlation [15]. To sum up, the algorithm cited by the water pollution control model based on random forest proposed in this paper makes the model run more efficiently. The running process of the algorithm is shown in Figure 3.

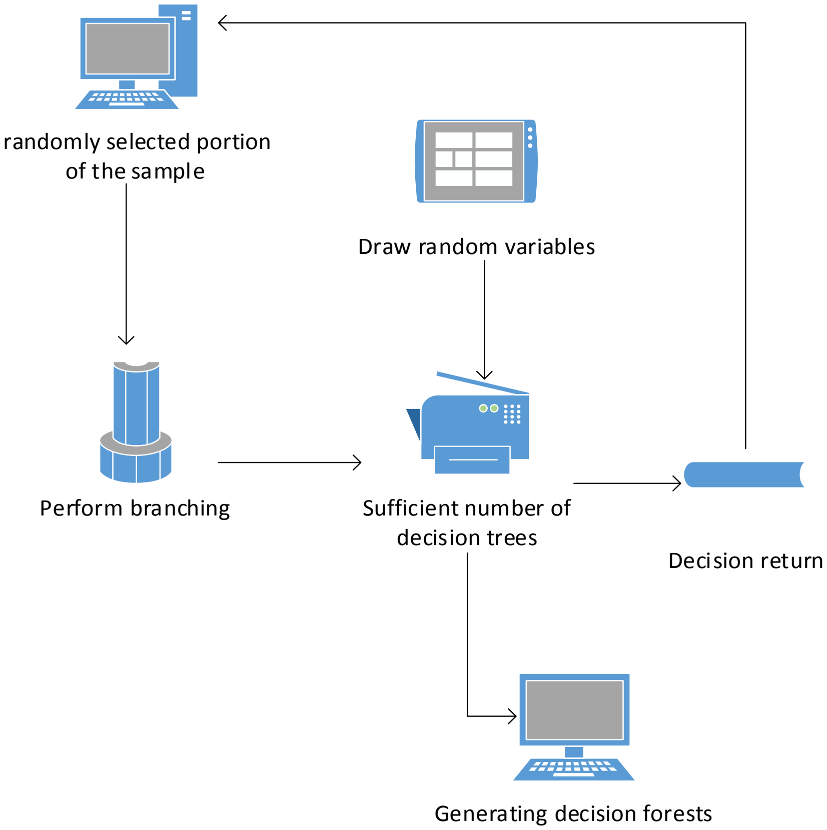

randomly selected portion of the sample
Draw random variables
Perform branching
Sufficient number of decision trees
Decision return
Generating decision forests

Figure 3. Random Forest algorithm flow

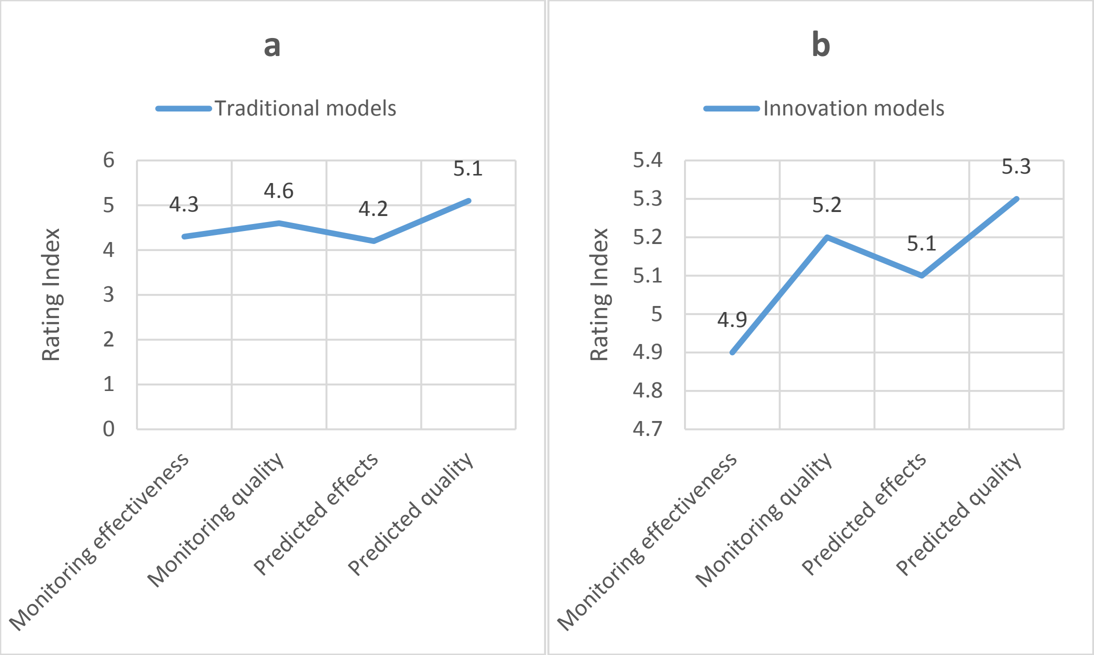
5. Experimental Evaluation of Water Pollution Control Model Based on Random Forest

With the progress of society and the development of science and technology, the quality of water resources is required to be higher and higher. The serious water pollution problem would seriously endanger the life safety of people drinking water. The monitoring and prediction of pollution sources in the water pollution problem is now the main research direction. In order to intuitively evaluate the performance and effect of the water pollution control scheme, the establishment of evaluation criteria is particularly important. As shown in Table 1, it is the classification and divergence of some evaluation standard indicators. In order to solve the existing water pollution problem and make the use of water resources sustainable, it is necessary to divide the water area in the region into zones, collect samples from the water area in each zone, and randomly classify the collected samples through the water pollution control model based on random forest proposed in this paper. The random attributes extracted from the sample are calculated by the random forest algorithm for the average generalization error, and then the random variables are continuously introduced to reduce the error. The final sample classification can effectively and accurately reflect the implementation status of water quality. If there is a strong deviation or trend of strong deviation in the status of water quality samples, the computer would send an alarm. Such a water pollution control model can effectively solve the problem that traditional water pollution control models rely too much on the professional knowledge and subjective judgment of experts and scholars, and can not monitor the water quality status in real time. Remedial measures are always taken after the water pollution problem has caused a certain impact, but it still needs some experiments to verify the effectiveness of the model.

Table 1. Evaluation criteria and their evaluation rules

Evaluation indicators	Rules of Conduct
Monitoring effectiveness	Speed of sample collection Source of sample collection
Monitoring quality	Water Quality Testing Difficulty of water collection

The first is to monitor and predict the water quality by dividing sample intervals in a certain area of water, and leave another area of water for blank comparison. After a period of testing, a questionnaire survey is conducted to the managers. The performance comparison analysis of the water pollution control model based on random forest and the traditional water pollution control model proposed in this paper is conducted according to the survey feedback and evaluation rules of the managers, as shown in Figure 4.



a. Performance analysis of traditional models

b. Performance analysis of innovative models

Figure 4. Comparative analysis of traditional and innovative models of water pollution control

As shown in Figure 4, Figure a shows the performance analysis of the traditional water pollution control model. The four performance indicators are monitoring effect, monitoring quality, prediction effect and prediction quality. The performance indexes of the four performance indicators were 4.3, 4.6, 4.2 and 5.1 respectively. Figure b shows the performance analysis of the innovative water pollution control model. The four performance indicators are monitoring effect, monitoring quality, prediction effect and prediction quality. The performance indexes of the four performance indicators were 4.9, 5.2, 5.1 and 5.3 respectively. According to the comparative analysis of the four performance indicators, the innovative water pollution control model was better than the traditional water pollution control model. Through the comparison of the two, the innovative water pollution control model proposed in this paper had an average improvement of 13.1% in four performance aspects compared with the traditional water pollution control model.

6. Conclusion

With the rapid development of computer technology and the progress of social industry, people have higher requirements for life. In the process of progress, the problem of water pollution that has been ignored by people has also become serious. In the process of people’s industrial production, agriculture, animal husbandry and life, the quality of water environment has been seriously affected. The frequent outbreak of water pollution events has caused immeasurable losses to people’s lives. In order to ensure the safety of people’s lives and property and minimize the occurrence of sudden environmental water pollution events, it is necessary to monitor and predict water resources in real time. In order to solve the problem that the traditional water pollution control mode relies too much on the subjective judgment and professional knowledge of experts and scholars, and cannot find and solve the hidden dangers of water pollution in time, and the solution to the water pollution problem is not perfect, an innovative water pollution control model was proposed by using the random forest algorithm and combining with the evaluation criteria of the water pollution control mode.

The model divided the water area to be operated into sample intervals, and then collects samples from different intervals. The collected samples were randomly classified by random forest algorithm and computer technology, and the calculated average generalization error was reduced by continuously introducing random variables. The final sample classification can reflect the state of water quality in real time. When the sample obtained had a huge and strong deviation, the computer would give a timely alarm to the manager and take countermeasures. In this paper, the model was used to set up an experiment in a certain water area. Through the evaluation and analysis of the experimental results, it was found that the water pollution control model with the introduction of random forest was based on the tradition, and its comprehensive performance was better than the traditional water pollution control model. It fundamentally solved some problems encountered in the practice of the traditional model, thus promoting the research process of the water pollution control model.

Funding

This article is not supported by any foundation.

Data Availability

Data sharing is not applicable to this article as no new data were created or analysed in this study.

Conflict of Interest

The author states that this article has no conflict of interest.

References

[1] Deletic Ana, Huanting Wang. Water pollution control for sustainable development. Engineering. (2019) 5(5): 839-840. https://doi.org/10.1016/j.eng.2019.07.013

[2] Mingjing He. Waste-derived biochar for water pollution control and sustainable development. Nature Reviews Earth & Environment. (2020) 3(7): 444-460.

[3] Rink Karsten. Virtual geographic environments for water pollution control. International Journal of Digital Earth. (2016) 11(4): 397-407. https://doi.org/10.1080/17538947.2016.1265016

[4] Martini Sri. Membrane technology for water pollution control: a review of recent hybrid mechanism. Jurnal Rekayasa Kimia & Lingkungan. (2020) 17(1): 83-96. https://doi.org/10.23955/rkl.v17i1.23610

[5] Tony Maha A. Low-cost adsorbents for environmental pollution control: a concise systematic review from the prospective of principles, mechanism and their applications. Journal of Dispersion Science and Technology. (2020) 43(11): 1612-1633.

[6] Ahmed Shahid, Saba Ismail. Water pollution and its sources, effects & management: a case study of Delhi. Shahid Ahmed and Saba Ismail. International Journal of Current Advanced Research. (2018) 7(2): 10436-10442.

[7] Linqing Wang. Water-soluble components in rainwater over Xi'an in northwest China: Source apportionment and pollution controls effectiveness evaluation. Atmospheric Pollution Research. (2019) 10(2): 395-403. https://doi.org/10.1016/j.apr.2018.08.011

[8] Li He, Juan Lu. Can regional integration control transboundary water pollution? A test from the Yangtze River economic belt. Environmental Science and Pollution Research. (2020) 27(22): 28288-28305. https://doi.org/10.1007/s11356-020-09205-1

[9] Alizadeh Mohamad Javad. Effect of river flow on the quality of estuarine and coastal waters using machine learning models. Engineering Applications of Computational Fluid Mechanics. (2018) 12(1): 810-823. https://doi.org/10.1080/19942060.2018.1528480

[10] Bisht Anil Kumar. Artificial neural network based water quality forecasting model for ganga river. International Journal of Engineering and Advanced Technology. (2019) 8(6): 2778-2785. https://doi.org/10.35940/ijeat.F8841.088619

[11] Huanhai Yang, Shue Liu. A prediction model of aquaculture water quality based on multiscale decomposition. Mathematical Biosciences and Engineering. (2020) 18(6): 7561-7579. https://doi.org/10.3934/mbe.2021374

[12] Barzegar Rahim, Mohammad Taghi Aalami, Jan Adamowski. Short-term water quality variable prediction using a hybrid CNN-LSTM deep learning model. Stochastic Environmental Research and Risk Assessment. (2020) 34(2): 415-433. https://doi.org/10.1007/s00477-020-01776-2

[13] Schonlau Matthias, Rosie Yuyan Zou. The random forest algorithm for statistical learning. The Stata Journal. (2020) 20(1): 3-29. https://doi.org/10.1177/1536867X20909688

[14] Georganos Stefanos, et al. Geographical random forests: a spatial extension of the random forest algorithm to address spatial heterogeneity in remote sensing and population modelling. Geocarto International. (2019) 36(2): 121-136. https://doi.org/10.1080/10106049.2019.1595177

[15] *Keerthan Kumar T G, Shubha C, Sushma S A. Sushma. Random forest algorithm for soil fertility prediction and grading using machine learning. Int J Innov Technol Explor Eng. (2019) 9(1): 1301-1304. https://doi.org/10.35940/ijitee.L3609.119119*