*Desalination and Water Treatment*
www.deswater.com
doi: 10.5004/dwt.2021.27566

241 (2021) 223–229
November

# The use of GIS-based genetic algorithm in water pollution control planning

Lin Liu

*Faculty of Mathematics and Physics, Southwest Forestry University, Kunming 650224, China, email: linliu8213@163.com*

Received 25 June 2021; Accepted 23 September 2021

**ABSTRACT**

The study aims to ensure the coordinated development of different factors involved in water pollution control planning, such as environment, economy, and society. The model algorithm is combined with geographic information system (GIS). The effectiveness of the algorithm and the empirical knowledge of expert system are used to realize the optimal planning of water pollution. On the basis of previous studies, mining algorithm is used as a starting point, and the required parts are developed on the basis of genetic algorithm optimal planning scheme and BP (back propagation) water quality prediction and assessment through the constructive method. After that, the related research is used to verify the water pollution control planning method. The model algorithm is closely combined with GIS, and the model algorithm is combined with the empirical knowledge of expert system to realize the water pollution control planning. In this way, when the relevant data are sufficient, it will reduce a lot of simulation and experiment, greatly reduce the waste of human and material resources, and provide a shortcut to a certain extent.

*Keywords:* Genetic algorithm; Water pollution control; Geographic information system; Decision support

## 1. Introduction

With the rapid development of modern science and technology, the continuous development of industry leads to lots of waste discharged into rivers and lakes, which causes serious pollution of water bodies. In order to improve the water environment and reduce pollution, people begin to invest a lot of financial and material resources to build sewage treatment plants and to study the related technologies of water pollution control, which contributes to the purification and treatment of water pollution in various aspects [1]. However, with the development of water pollution prevention and control projects, the economic benefits of sewage treatment have not been improved with the in-depth research and improvement of water pollution treatment technology, and the cost of water pollution treatment has been constantly improved. The use of system analysis technology can greatly reduce the cost of water pollution control planning.

The planning of the water pollution control system is a subject that was put forward only after 1960. In contrast, the research of this subject has very obvious characteristics of the times. When linear and low-dimensional models are dealt with, it is very suitable to use conventional optimization methods, and this technology is also mature. However, for a nonlinear large-scale system of multi-level and complex regional water resources, the planning of a water pollution control system cannot meet the research in it [2].

The model algorithm is combined with a geographic information system (GIS); moreover, the model algorithm is combined with the empirical knowledge of an expert system to realize the water pollution control planning. In this way, when the relevant data are sufficient, the data model based on GIS and genetic algorithm (GA) can be directly used to achieve water pollution control planning, which will reduce a lot of unnecessary and complicated simulation and experiment, greatly reduce the waste of human and material resources, and provide a shortcut to a certain extent.

## 2. Method

### 2.1. Genetic algorithm

GA is a method of using the computer to simulate biological evolution. The algorithm is inspired by the study of biological evolution and implemented by a computer. It is to use the recombination and mutation of the current hypothesis to generate the later hypothesis, which is different from the general search hypothesis. The process of the genetic algorithm (Fig. 1) starts from the most primitive hypothesis bits, and the process is mainly divided into five stages: (1) the most primitive hypothesis population is generated by coding; (2) the individual fitness of the current population is calculated. If it is the optimal solution, it is necessary to stop; otherwise, it is necessary to continue to run the next step; (3) according to the prescribed selection skills, some groups with higher fitness are selected, while others are abandoned; (4) the crossover operator is applied to generate a new bit string; (5) it is necessary to randomly mutate these new populations and return to step (2).

### 2.2. Water quality assessment

Water quality assessment has its unique importance, it is the basis of water system planning and calculation of water basin capacity. The purpose of the water quality status assessment is to clarify the current water quality situation. Water quality assessment is based on the combination of water quality related monitoring data and field investigation results [3]. The factors causing environmental water quality problems are various, and the degree of water pollution is affected by the concentration of substances and other factors. These various factors also bring a certain amount of work for water quality assessment.

The relevant data obtained are compared. It is found that fecal coliform, dissolved oxygen (DO), permanganate index, chemical oxygen demand (COD), total phosphorus (TP), ammonia nitrogen and so on are the main influencing factors of water quality level. Therefore, the water quality level can be judged by selecting these relevant data indexes. A BP neural network with three layers of input layer, hidden layer and output layer is selected, and then the input layer of the BP neural network contains six nodes. The output layer is represented by only one node, so that the water quality category can be seen intuitively. Since the water quality level is from class I to inferior class V, considering that the range of activation function is [0,1], 0.1, 0.3, 0.5, 0.7, 0.9 and 1 are used to represent the water quality of class I and inferior class V respectively as the output of the neural network. The hidden layer has three nodes. Fig. 2 is the structure diagram of the whole model.

The establishment of the system usually needs to integrate GIS data, various program models and basic spatial processing functions (Fig. 3). However, the adaptability and efficiency of the system are largely determined by the system scheme [4]. The system integration schemes used by different application fields and different application developers are usually different, but the general integration schemes have some defects. For example, based on GIS, software and application analysis model are independent of each other. They open up data exchange mode through data access, which also leads to poor system integration to a certain extent. COMGIS can solve all kinds of problems well, and it can be separated from any computer language. It can run GIS function in any environment, use common development environment to realize professional model, and insert analysis control in other models, so as to realize efficient and seamless system integration [5].

In the water quality assessment and COMGIS prediction, the emphasis is to introduce BP and GA professional models, so that after the conversion of water quality assessment related data in the system, the data can be practiced and calculated by calling the neural network model space. This method improves the common data utilization efficiency and management problems. The establishment process is as follows:

The database will eventually be stored on physical devices such as computer hard disks. The access method and storage structure of the database are called the physical structure of the database, which depends on the database management system [6]. The main part of the database is the data table. After the water environment attribute information is analyzed and processed according to the actual needs, the main data table of the system is as follows (Table 1):

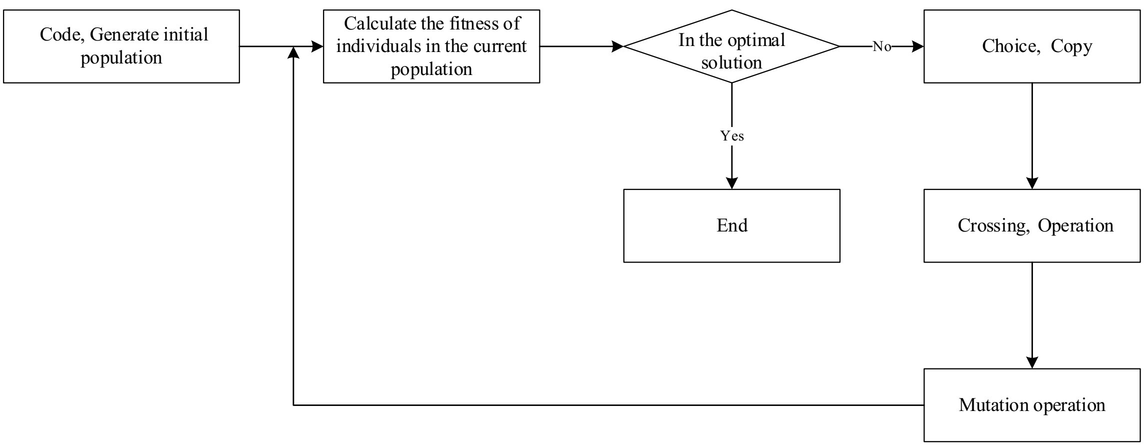


Fig. 1. Flow chart of basic genetic algorithm.

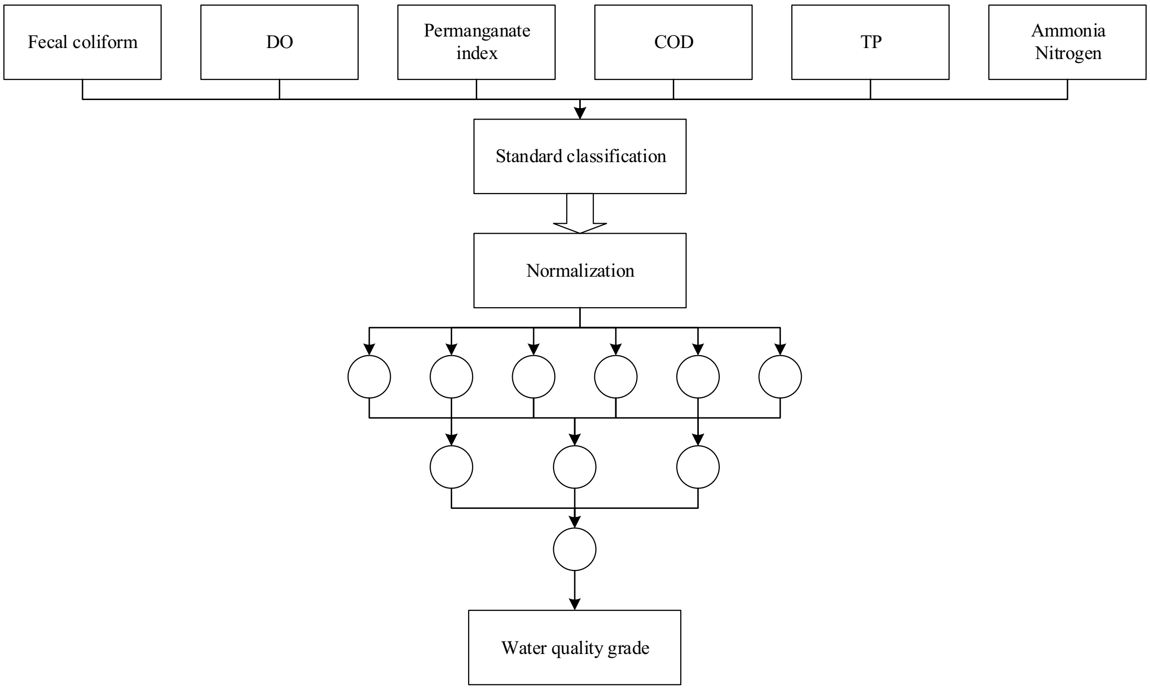


Fig. 2. Structure chart of water quality assessment model.

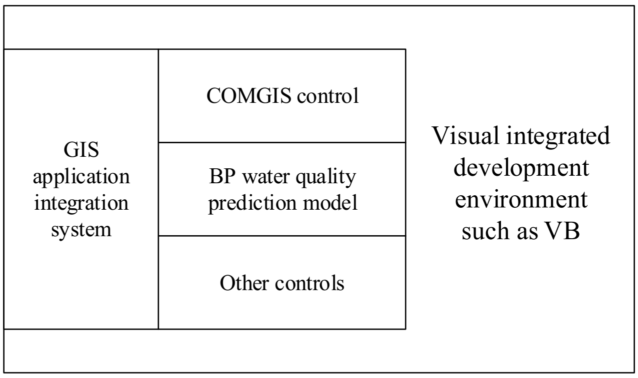


Fig. 3. Seamless integration between COMGIS and data mining.

Spatial database design consists of data structure design and data model. The data structure is usually to express the information of physical objects and organization objects, and the data model is to describe the information of physical objects [7].

With Di'er Songhua Jiang as an example, its water quality is evaluated. Parameter identification is performed by using relevant pollution source data (Table 2). The model is verified by the actual monitoring data in 2020, and the water quality mathematical model is determined. The results are as follows

### 2.3. Water pollution load forecasting

The first step in the whole process of water pollution control planning is water pollution load forecasting. The main function of GIS is to calculate the pollution load and forecast the planning level of the annual pollution load. Its components include population forecast, soil erosion, direct industrial discharge, municipal solid waste, municipal sewage and total pollution load calculation [8]. Among them, the calculation models of municipal sewage and municipal solid waste pollution load can also be expressed by the results of the population prediction model, which is convenient to predict the pollution load of municipal sewage and municipal solid waste in the planning level year.

The research of pollutant load in GIS mainly includes the following four aspects.

- Space management of pollution sources. The spatial data management function of GIS is used to manage pollution sources. On the existing digital map, the pollution source can be changed at any time according to the actual situation.
- Pollution source attribute management. Through the attribute data management function of GIS, the pollution source attributes of the identified location are described by programming.
- Calculation of pollution load. The models used to calculate the pollution load (including atmospheric dust, water waste, domestic sewage, industrial wastewater, and soil erosion) are used to calculate the time required for the pollution load.
- Pollution load query and interannual comparison. In daily work, it is often necessary to compare the changes of pollution load among years, to obtain the change degree of pollution load among years, and to find the appropriate method of pollution control.

### 2.4. Water environmental capacity

The fuzzy influence coefficient method mentioned by previous scholars is adopted. Comparatively speaking, water environment system is a huge and complex system that has the following characteristics: (1) because of the different distribution location of each pollution source and the amount of pollution discharged, the impact of pollution sources on water quality control is relatively small. In addition, the pollutants interact with each other after entering the water body. Therefore, it is difficult to accurately calculate and determine the impact of different sources; (2) in the plan, it is difficult to accurately predict the pollutant emissions of various pollution sources, which is ambiguous; (3) it is difficult to calculate all the pollution points that cause pollution to the river water body; (4) it is difficult to accurately simulate the calculation system in ordinary test, and the calculation process is more complex. Good results have been achieved in the application of the fuzzy influence coefficient method in the calculation of water environmental capacity [9].

According to the above water quality verification and identification, the water quality change of the control section of the river section is calculated after inputting the pollution load of a certain year.

$$A = \frac{\Delta C}{\Delta W} \tag{1}$$

where $A$ is the fuzzy influence coefficient [(mg/L)/(kg/d)]; $\Delta C$ is the change of water quality caused by the input of pollution load of a certain river section in a certain year (mg/L); $\Delta W$ is the pollution load of the river section in a certain year (kg/d).

The water environmental capacity of the water body in this river section is as follows:

$$R = \frac{C_S - C_0}{A} \tag{2}$$

where $R$ is the water environmental capacity of water body in a river section (kg/d); $C_S$ is the pollutant concentration specified by water quality target at control section (mg/d); $C_0$ is the input pollution load is zero in this reach (only background load from upstream).

This method regards the pollution source of river section as a "gray box" system, and does not study the spatial position of each water section and the relative size of water quality (because the pollution source of water is a very complex system, and it is difficult to use the relevant mathematical content to express). To a large extent, it reduces the rather complicated calculation process of calculating the coefficient between pollution sources after the coefficient of single pollution is calculated by the general influence coefficient method [10]. The calculation process of water environmental capacity can reflect the actual situation more accurately, which has certain guidance for the actual water pollution control plan and has strong operability.

## 3. Results and discussion

### 3.1. Analysis of scheme cost

The treatment of water pollution can bring many benefits besides the water body. The treatment of water pollution can improve the water quality of the basin, increase the diversity of its application, reduce the waste of water resources to a certain extent, improve the environment around the water area, increase the land value, and promote the production of industry and agriculture to a certain extent. These are the benefits of water pollution control. However, if there are benefits, there must be the corresponding "costs", including the construction cost and operation cost of sewage treatment facilities [11].

The year-consumption method is suitable for verifying whether the enterprise investment is effective or comparing schemes. The year-consumption method is one of the dynamic analysis methods of engineering economic analysis. Considering that the value of money will change with time, this method is in line with the objective economic law and is widely used at present. The calculation equation of this method is as follows:

$$C = KC_{\mathrm{rf}} + S \tag{3}$$

where $C$ is the annual cost (ten thousand yuan), that is, the sum of annual operation cost and annual depreciation cost of sewage treatment facilities; $K$ is the infrastructure investment (including loan interest); $C_{\mathrm{rf}}$ is the capital recovery factor; $S$ is the operation cost (ten thousand yuan/a).

The capital recovery factor can be calculated by Eq. (4).

$$C_{\mathrm{rf}} = \frac{i(1+i)^n}{(1+i)^n - 1} \tag{4}$$

where $i$ is the annual interest rate of investment expenses repayment; $n$ is the service life of treatment facilities.

#### 3.1.1. Cost of sewage treatment plant

In the scheme of water pollution control planning, different treatment methods are often used for different wastewater treatment plants, which can be divided into primary treatment or secondary treatment.

According to the construction experience of municipal sewage treatment plants in China and foreign countries, the relationship between the construction cost of secondary sewage treatment plants and the scale of sewage treatment plants is an exponential function, that is:

$$C = aQ^b \tag{5}$$

where $C$ is the construction cost of sewage treatment plant; $a,b$ is the constant, $b < 1$; $Q$ is the scale of sewage treatment plant.

Since $b < 1$, the larger the scale of sewage treatment plant is, the lower the unit water treatment investment is. In different periods, the unit investment and operation cost of sewage treatment plant are different. According to the investment and operation of municipal sewage treatment plants in recent years, the construction cost equation is obtained.

$$C = 1712.3Q^{0.7768} \tag{6}$$

Table 1
Attribute table

| Table name | Attribute field | Key fields |
|---|---|---|
| Water intake | Number, name, the river section that it belongs to, the quantity of water intake | Number |
| Monitoring section | Number, name, the river section that it belongs to | Number |
| Function section table | Number, name, water body name, starting point section, end point section, whole section length | Number |
| Pollution source table | Number, name, the river section that it belongs to, discharge volume and main components | Number |
| Monthly report of monitoring data | Section ID, flow rate, temperature, pH value, DO, $COD_5$, Hg, Pb, ammonia nitrogen, coliform group, water quality category, main pollution components | Section |

Table 2
Identification and verification of water quality parameters of the Di'er Songhua Jiang

| Area | Chemical oxygen demand | Volatile phenolic compounds | Ammonia nitrogen | Biochemical oxygen demand |
|---|---|---|---|---|
| Changchun | 0.85 | 1.407 | 4.13 | –3.01 |
| Beijing reservoir | –0.19 | 1.59 | –0.291 | –0.68 |
| Rishan reservoir | –0.318 | –0.33 | –0.243 | 0.73 |

where $C$ is the construction cost of sewage treatment plant, $Q$ is the scale of sewage treatment plant (ten thousand $m^3$/y).

The operation cost equation of sewage treatment plant is shown in Eq. (7).

$$C_R = 0.33Q \tag{7}$$

where $C_R$ is the operation cost of sewage treatment plant (10,000 yuan/y).

#### 3.1.2. Annual cost of industrial wastewater treatment

The types of industrial pollution sources are very complex, and the water pollution treatment process is not the same, so it is difficult to set the same cost equation as the municipal sewage in the city. According to the actual situation of each industrial wastewater treatment, the construction cost is estimated, and the operation cost is based on the domestic experience. The correlation equations shown in the table below (Table 3).

Therefore, for any industry, the annual cost equation can also be obtained. The cost of water pollution control plan is mainly composed of the above three aspects. It is worth noting that due to the different construction time and location of the project, the parameters in the relationship in the table will be different, and even form a cost calculation equation [12]. Therefore, when the actual cost of wastewater treatment is calculated, attention should be paid to the development and change of industrial wastewater treatment.

### 3.2. GIS implementation of optimal planning scheme based on GA

A code of the problem is selected; an initial population POP with N chromosomes is given. It suggests that the process of GA is very simple. Selection, crossover and mutation are the three main operators of GA, which make GA have different characteristics from other traditional methods. GA includes the following five basic elements: (1) parameter coding; (2) fitness function design; (3) initial population setting; (4) genetic operation design; (5) control parameter setting.

In order to clearly display the combination of different processing levels, and combine with GA coding rules, the real number coding method is selected. GA is used to optimize water pollution control plan [13]. The mathematical description of GA is as follows:

- Coding

If there are n pollution sources in a planning scheme, each pollution source will have three treatment levels, namely zero level treatment, first level treatment and second level treatment. In order to easily and quickly see the treatment of each pollution source, and to make the calculation easier, encoding with real numbers is considered, that is, using 1, 2, 3 to represent zero level treatment, first level treatment and second level treatment respectively, and using n real number strings {132123…23221} composed of 1, 2, and 3 to denote a combination of processing degree, that is, a chromosome.

- Fitness function

The algorithm takes the objective function value (the reciprocal of the total cost of each plan combination) as the fitness, and constructs the fitness function based on it. It means that the higher the total cost is, the smaller the applicability value is, the greater the possibility of eliminating it is [14].

- If the stop rule is satisfied, the algorithm stops; otherwise, the roulette probability is calculated according to the following equation:

$$P_i = \frac{f_i}{\sum_{j=1}^{N} f_i} \quad i = 1,2,3,\ldots,N \tag{8}$$

According to the probability distribution, some chromosomes are randomly selected from POP ($T$) to form a population;

$$\text{newpop}(t+1) = \left\{ \text{pop}_j(t) \middle| j = 1,2,\ldots,N \right\} \tag{9}$$

- By crossover, the crossover probability is PC, and a crosspop($t$ + 1) with N chromosomes is obtained.
- With a small probability P, a gene of a chromosome is mutated to form mutpop($t$ + 1): $t = t + 1$, a new population pop($t$) = mutpop($t$).

In the process of research, four optimal or near optimal schemes are proposed for different water quality conditions.

Then, the pollution load prediction model, water quality model and water quality assessment are called to predict and analyze the water quality after the implementation of each scheme [15]. Meanwhile, the above results are debugged to run GA, and a certain number of iterations are used to end the algorithm, so as to achieve the maximum environmental and economic benefits with less cost and get the optimal scheme. Finally, in the 24 running results, the results show that scheme 1 has 4 times, scheme 2 has 4 times, scheme 3 has 7 times, and scheme 4 has 9 times. Moreover, at the end of the run, the results tend to converge to scheme 3 and 4. Table 4 presents the specific results [16–19].

According to the calculation of cost analysis model, the annual cost of scheme 1 is 8.792 million yuan, that of scheme 2 is 18.293 million yuan, that of scheme 3 is 40.125 million yuan, and that of scheme 4 is 48.332 million yuan.

It suggests that the results of planning simulation scheme are very similar to the results of GA operation, which verifies the effectiveness, feasibility and efficiency of GA optimal planning [20–22].

Table 3
Relationship between annual cost of industrial water treatment and construction cost

| Types of wastewater | Relationship between annual operation cost $C_R$ and construction cost $C$ |
|---|---|
| Chemical industrial wastewater | $C_R = 0.21C$ |
| Containing wastewater | $C_R = 0.22C$ |
| Organic wastewater | $C_R = 0.21C$ |
| Electroplating wastewater | $C_R = 0.1165C$ |
| Phenolic wastewater | $C_R = 0.147C$ |
| Papermaking wastewater | $C_R = 0.1854C$ |
| Printing and dying wastewater | $C_R = 0.2505C$ |

Table 4
Results of GA running for 24 times

| Index | Cost (10,000 yuan) | Times |
|---|---|---|
| Scheme 1 | 879.2 | 4 |
| Scheme 2 | 1,829.3 | 4 |
| Scheme 3 | 4,012.5 | 7 |
| Scheme 4 | 4,833.2 | 9 |

## 4. Conclusion

The algorithm is combined with GIS to verify the original water quality prediction and assessment. On the one hand, the original water quality is evaluated. On the other hand, after each model is calculated, it will traverse the module again to generate a integrated solution. In the integrated solution, the water quality, water pollution load and water environmental capacity of an area are comprehensively considered and calculated in various ways, and a variety of combination calculations are obtained to obtain the initial optimal or suboptimal plan. Finally, the optimal value of revenue analysis is generated by the number of GA operations, and the global optimal value is obtained. By combining the advantages of the algorithm with the empirical knowledge of expert system, the optimal planning of water pollution can be realized. Therefore, in the case of more complete data collection, it can save a lot of simulation experiments, greatly reduce the waste of human and material resources, and provide a shortcut to a certain extent.

The water pollution load and water environment capacity involved are mostly based on the previous research results, without in-depth research on these two aspects. It is hoped that there will be in-depth innovative research on these two aspects in the future; in addition, due to data reasons, there is no forecast for the future. In the future work, it is hoped that with the support of data, a prediction about the future water pollution control planning can be made.

## Acknowledgement

- Funded by the Scientific Research Fund of Yunnan Education Department (No: 2021J0154).
- Funded by the Educational Science Research Project of Southwest Forestry University (No: YB202006).

## References

[1] Q. Wang, R. Liu, C. Men, L. Guo, Application of genetic algorithm to land use optimization for non-point source pollution control based on CLUE-S and SWAT, J. Hydrol., 560 (2018) 86–96.

[2] S.P. Duvvuri, J. Anmala, Fecal coliform predictive model using genetic algorithm-based radial basis function neural networks (GA-RBFNNs), Neural Comput. Appl., 31 (2019) 8393–8409.

[3] M. Tamer Ayvaz, A. Elçi, Identification of the optimum groundwater quality monitoring network using a genetic algorithm based optimization approach, J. Hydrol., 563 (2018) 1078–1091.

[4] M. Ehteram, H. Karami, S. Farzin, Reservoir optimization for energy production using a new evolutionary algorithm based on multi-criteria decision-making models, Water Resour. Manage., 32 (2018) 2539–2560.
[5] K. Li, Y. Chen, A Genetic Algorithm-based urban cluster automatic threshold method by combining VIIRS DNB, NDVI, and NDBI to monitor urbanization, Remote Sens., 10 (2018) 277, doi: 10.3390/rs10020277.
[6] A. Abdul Sattar, Preparation of novel hybrid (Almond Shell and *Pleurotus Sajor Caju*) biosorbent for the removal of heavy metals (nickel and lead) from wastewater, Water Conserv. Manage., 4 (2021) 1–7.
[7] A. Kadam, M. Rajasekhar, B. Umrikar, V. Bhagat, V. Wagh, R.N. Sankua, Land suitability analysis for afforestation in semi-arid watershed of Western Ghat, India: a groundwater recharge perspective, Geol. Ecol. Landscapes, 5 (2021) 136–148.
[8] H. Hc, S. Govindaiah, L. Srikanth, H.J. Surendra, Prioritization of sub-watersheds of the Kanakapura Watershed in the Arkavathi River Basin, Karnataka, India- using remote sensing and GIS, Geol. Ecol. Landscapes, 5 (2021) 149–160.
[9] J.D. Prasetya, D.H. Santoso, E. Muryani, T. Ramadhamayanti, B.A.S. Yudha, Carrying capacity of mercury pollution to rivers in the gold mining area of Pancurendang Village, Banyumas, J. Clean WAS, 5 (2021) 1–4.
[10] H.O. Nwankwoala, D.C. Okujagu, A review of wetlands and coastal resources of the Niger Delta: potentials, challenges and prospects, Environ. Ecosyst. Sci., 5 (2021) 37–46.
[11] B. Samuel, O. Temitope, D. Timothy, O. Olayinka, O. Olusegun, B. Emmanuel, Evaluation of the impacts of metals on soil samples, serum creatinine and blood urea nitrogen of residents in selected industrial communities in a developing country, Environ. Contam. Rev., 3 (2020) 40–47.
[12] N.S. Mohd, R.M. Mohamed, The initial ion effect of heavy metals adsorption by using hydrothermal carbonization banana peels, Environ. Contam. Rev., 3 (2021) 8–10.
[13] H. Tao, M. Li, M. Wang, G. Lü, Genetic algorithm-based method for forest type classification using multi-temporal NDVI from Landsat TM imagery, Ann. GIS, 25 (2019) 33–43.
[14] R. Geng, P. Yin, A.N. Sharpley, A coupled model system to optimize the best management practices for nonpoint source pollution control, J. Cleaner Prod., 220 (2019) 581–592.
[15] T. Hanák, M. Tuscher, O. Přibyl, Hybrid Genetic Algorithm-based approach for estimating flood losses on structures of buildings, Sustainability, 12 (2020) 3047, doi: 10.3390/su12073047.
[16] S. Jomthanachai, W. Rattanamanee, R. Sinthavalai, W.-P. Wong, The application of genetic algorithm and data analytics for total resource management at the firm level, Resour. Conserv. Recycl., 161 (2020) 104985, doi: 10.1016/j.resconrec.2020.104985.
[17] E. Dodangeh, M. Panahi, F. Rezaie, S. Lee, D. Tien Bui, C.-W. Lee, B. Pradhan, Novel hybrid intelligence models for flood-susceptibility prediction: meta optimization of the GMDH and SVR models with the genetic algorithm and harmony search, J. Hydrol., 590 (2020) 125423, doi: 10.1016/j.jhydrol.2020.125423.
[18] J. Yan, Z. Xu, Y. Yu, H. Xu, K. Gao, Application of a hybrid optimized BP network model to estimate water quality parameters of Beihai Lake in Beijing, Appl. Sci., 9 (2019) 1863, doi: 10.3390/app9091863.
[19] A.R.M. Towfiqul Islam, S. Talukdar, S. Mahato, S. Ziaul, K. Uddin Eibek, S. Akhter, Q.B. Pham, B. Mohammadi, F. Karimi, N.T. Thuy Linh, Machine learning algorithm-based risk assessment of riparian wetlands in Padma River Basin of Northwest Bangladesh, Environ. Sci. Pollut. Res., 28 (2021) 34450–34471.
[20] M. Bordbar, A. Neshat, S. Javadi, S. Mehdy Hashemy Shahdany, A hybrid approach based on statistical method and meta-heuristic optimization algorithm for coastal aquifer vulnerability assessment, Environ. Model. Assess., 26 (2021) 1–14.
[21] S. Khishtandar, Simulation based evolutionary algorithms for fuzzy chance-constrained biogas supply chain design, Appl. Energy, 236 (2019) 183–195.
[22] K. Elbaz, S.-L. Shen, A. Zhou, Z.-Y. Yin, H.-M. Lyu, Prediction of disc cutter life during shield tunneling with AI via the incorporation of a genetic algorithm into a GMDH-type neural network, Engineering, 7 (2021) 238–251.