Scholar Publishing Group
Water Pollution Prevention and Control Project
https://doi.org/10.38007/WPPCP.2022.030305
ISSN 2790-2730 Vol. 3, Issue 3: 37-45

Drinking Water Pollution Prevention and Control Engineering Model Based on Random Forest Algorithm and Electronic Information Intelligence

Prusa Filip*

Univ Porto, P-4169007 Porto, Portugal

**corresponding author*

Keywords: Random Forest Algorithm, Electronic Information Intelligence Technology, Drinking Water Pollution, Prevention and Control Measures

Abstract: With the development of social and economic development, modern industry has made a huge contribution, but the pollution of drinking water (DW) sources has also developed to a more serious situation. The shortage of DW sources has intensified due to the large amount of industrial waste water discharged directly into the water, resulting in the pollution of DW sources. In order to prevent DW pollution, this paper uses a random forest (RF) algorithm and electronic information intelligence technology to build a water pollution early warning system, the system through the predictive power of the RF cannot meet the DW standards for intelligent warning water. This paper analyzes the current situation of DW pollution in Q city and finds that Q city has little supervision of polluted water sources in agriculture and water supply plants, so it proposes measures to prevent DW pollution in the hope that the citizens of Q city can use clean water resources.

1. Introduction

Water is the material basis for the survival and development of life and the hygiene of DW is linked to the health of all inhabitants. Having access to safe DW is the most basic need of life for people. Access to safe DW is the most basic need of people's lives, but, as society becomes more industrialised, the price of constantly improving material living standards is pollution and damage to the natural environment [1]. Water, soil, air, vegetation and other resources closely related to human survival by different degrees of infringement, especially water pollution problems are very serious, the quality of DW is declining, so it is necessary to strengthen the prevention and control of DW pollution [2].

Research on DW pollution prevention and control at home and abroad has made good achievements. Some developed countries have given sufficient attention to centralised wastewater treatment. At the same time, urban and rural sewage began to establish sewage collection networks for systematic collection and treatment. Since then, these countries have started primary and secondary treatment of wastewater, mainly through phosphorus removal, nitrogen removal, water quality improvement and safe DW technologies, which have been used to improve water quality and enhance the safety and security of DW [3-4]. In China, the initial use of laws and regulations to regulate and promote the protection of DW sources is very important for the state and localities to pay attention to DW safety and to make the protection of DW sources a top priority for environmental protection. Specifically, the "Three Rivers and Three Lakes" management project has effectively controlled water pollution in these major watersheds and prevented further deterioration of water quality through effective long-term management, legal disposal of a large number of highly polluting industrial enterprises, and the construction and operation of several sewage treatment plants [5-6]. In short, DW pollution is global in scope and its management should be of concern to people worldwide.

In this paper, we firstly elaborate the concept of RF algorithm, then design a RF based water pollution early warning electronic information intelligence system used to supervise DW quality, then take the DW pollution situation in Q city as an example, analyze its DW quality status, and finally propose measures to prevent DW pollution according to the pollution situation in Q city.

2. Basic Overview

2.1. RF Algorithm

The RF model is one of the more practical, and not too complex, machine learning models. In the vast majority of today's algorithms, the best classification results are achieved with high-precision models, even if some data does not have a significant impact on the model's judgement [7]. In addition, the RF model can be used with models that have high latitude characteristics without reducing data and filtering measures. Finally the model assesses the importance of each function in the rating process during the adjustment process and helps investors to perform an effective analysis.

The RF algorithm is a combination of Decision Trees and Bagging. Described as "representing the state of the art in integrated learning", RFs are easy to implement. RFs are characterised by two layers of randomness, sampling randomness and node randomness [8].

Sampling random: RFs are an extended variation of the Bagging integration algorithm that follows the principles of the self-help sampling method in sampling to reposition sampling, so that the training sample set is random [9].

Node randomisation: For regular decision trees, the root node is located at the top of the tree and contains all the original data. As the data at the root node is split, each node finds the best way to split the nodes, which requires passing all attribute variables to form different leaf nodes prior to splitting [10]. Adjustment by pruning operations is also required to reduce errors and over-fitting. For RFs, the selection of a subset of attributes in each decision tree is random by selecting a subset of thousands of attributes and then traversing all the attribute variables in the attribute subset to select the best attribute, i.e. the randomness of the nodes [11-12].

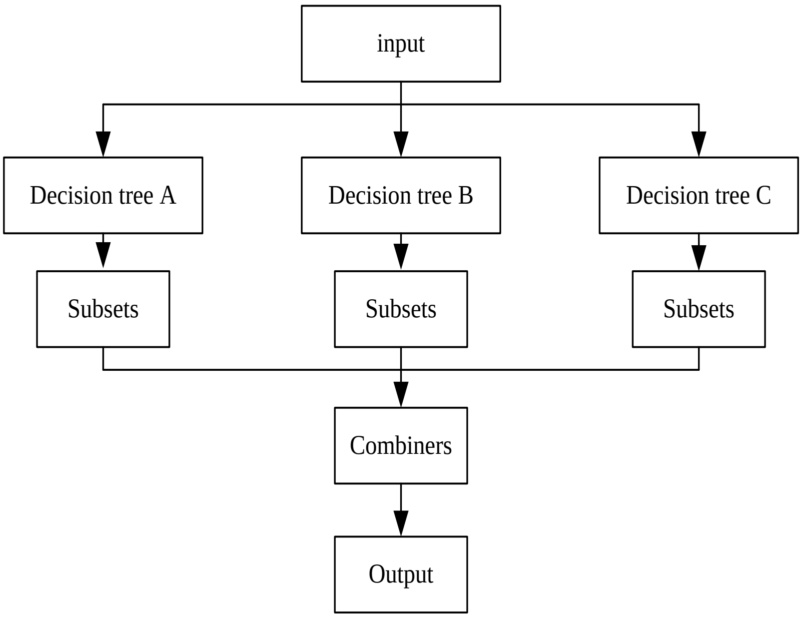

Figure 1. RF model

As Figure 1 shows the RF model, unlike individual decision tree algorithms, each base learner in a RF is complete. In constructing the ith base learner, assuming that the out-of-bag sample dataset is $O_i(x)$, the correct rate estimate $Q_i(x,y)$ expression is:

$$Q_i(x,y)=\frac{\sum_{i=1}^{k} I(h_i(x)=y_i(x,y)\in O_i(x))}{\sum_{i=1}^{k} I(h_i(x),(x,y)\in O_i(x))} \quad (1)$$

$$OOB\_error=1-Q_i(x,y) \quad (2)$$

$$p=(1-\frac{1}{m})^m \quad (3)$$

M is the sample training set, x and y are the samples to be classified, and $h_i(x)$ is the number of OOB samples that are correctly classified, p is the probability that the sample will not be drawn. OOB error is used for model tuning, and in general, the smaller the oob error, the higher the quality of the model.

2.2. Intelligent Water Pollution Early Warning System Based on RF and Electronic Information

With the rapid development of information technology such as the Internet of Things, 5G networks, artificial intelligence technology, big data and cloud computing, the intelligence of water pollution problem treatment has become the focus of people's thinking. For a certain water problem, building a complete and universal intelligent treatment system can effectively reduce, or even get rid of, human involvement and improve the efficiency of dealing with the problem [13]. For DW pollution, there is an even greater need for intelligent treatment. The RF algorithm is used to provide early warning of water pollution and, depending on the level of warning, considers the abnormal size and toxicity of the pollutant itself and the detection indicators of the pollutant in the water environment; for the water environment, the division of water bodies into surface water environmental quality standards is considered [14-15]. The system is mainly for emergency monitoring, which is part of DW pollution control, so the system can be used as part of emergency decision-making for DW pollution incidents.

3. Q City DW Pollution Status

3.1. Q City DW Source Pollution Characteristics

(1) Water source is influenced by geological structural factors

Table 1. Groundwater-type water sources exceed the water quality standards

Water quality exceedance factor	Number of water sources exceeding the standard	Proportion(%)
Total hardness	12	63.16
Sulfate	2	10.53
Nitrate	1	5.26
Ammonia nitrogen	4	21.05

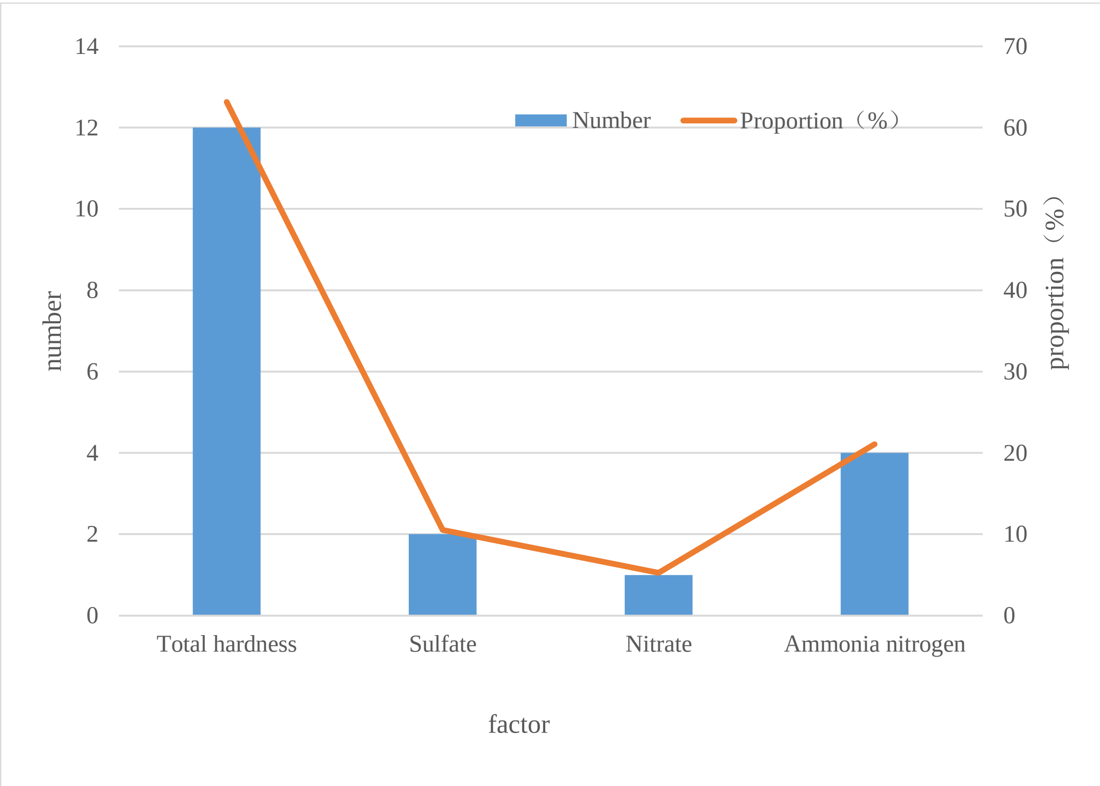

Figure 2. Water quality exceeds the standard

Most of the water source types in Q city are groundwater type, in addition to backup water sources and non-regional water supply for the lake and reservoir type water sources, other cities, towns (including townships) and rural DW sources are all groundwater type water sources. As

Table 1 and Figure 2, in the city's 75 groundwater-type water sources, a total of 19 water sources have exceeded water quality standards, including 12 water sources exceeded the total hardness index, accounting for 63.16% of the excess water sources; 2 water sources exceeded the sulfate index, accounting for 10.53%; 1 water source exceeded the nitrate index, accounting for 5.26%; 4 ammonia nitrogen exceeded the standard, accounting for 21.05%. It is mainly affected by geological structural factors, it can be seen that the Q City DW source pollution natural factors than man-made factors.

(2) The existence of agricultural surface pollution trend of water sources

Most of the agricultural land in the Q Water Protection Zone is arable, and of the 75 active water sources, 67 agricultural water sources are in the Zone, all of which are rich in nitrogen in agricultural soils. When rainwater dissolves nitrogen, nitrates leach into the soil and groundwater is contaminated. Only one water source currently exceeds the nitrate standard, but most of them already have high nitrate levels, and in the long term, Q's DW sources will face an increasingly serious trend of agricultural surface pollution.

(3) Water sources face the threat of long-term cumulative effects

Of the 85 water sources used, 54 are used for raising livestock in protected areas, mainly pig and chicken farming, mostly decentralised, as individual farmers have no fixed mode and method for farming methods and livestock manure disposal, resulting in long-term accumulation of livestock manure and other pollutants, causing groundwater pollution with rainwater washout, with ammonia nitrogen exceeding the standard as the main manifestation, with four water source wells exceeding the ammonia nitrogen index, livestock pollution has a significant impact on the long-term accumulation of these areas and should be reduced or eliminated as soon as possible to reduce or prevent the accumulation of livestock pollutants to pollute groundwater resources.

3.2. Agricultural Surface Source Water Pollution Analysis

Among the DW sources in Q city, only some reservoirs have some agricultural land in the upper watershed area, and other reservoirs are forest land. The area of agricultural land in the upstream of reservoirs is 197.6 hectares, and the main crops grown are 84.2 hectares of grain, 94.2 hectares of peanuts and 125.85 hectares of vegetable crops. According to the given emission coefficient statistics, the amount of agricultural surface source pollutants generated in the reservoir watershed is 39.7 tons of chemical oxygen demand, 14.6 tons of ammonia nitrogen, 5.7 tons of total phosphorus and 152.3 tons of total nitrogen. The pollutant generation of agricultural surface source in the reservoir watershed is shown in Table 2.

Table 2. Pollutants generated from agricultural surface sources in the DW source protection zone

Garden area	Farmland area	Planting area			Pollutant emissions			
		Grain	Vegetables	Peanut	COD	NH_3-N	TP	TN
484.3	197.6	84.2	125.8	94.2	34.5	14.6	5.7	152.3

3.3. Water Quality Pass Rate of Water Plants Needs to be Improved

The DW supply system in the city of Q is linked through a network of water sources, water plants and water supplies. In order to ensure the health and safety condition of DW, all aspects must meet the health requirements and the rings are closely linked. After having a qualified water source, the raw water treatment capacity of the plant should be equipped with certain standard water treatment technologies and the city's water supply network facilities should be constructed in a safe and reliable manner. In the raw water treatment of water treatment plants, effective measures should be taken to deal with the quality of different water sources and to understand the different causes of contamination to ensure the health of DW in the region. According to statistics, Q city centralized water supply plant workers 3587, of which 3371 rural accounted for, urban accounted for 216. Water quality testing of water supply plant, the water sample qualification rate of 76.34%, including urban water supply plant water sample qualification rate of 82.96%, the rural water supply plant water sample qualification rate of 75.42%, the water quality qualification rate is still a lot of room for improvement, see Figure 3.

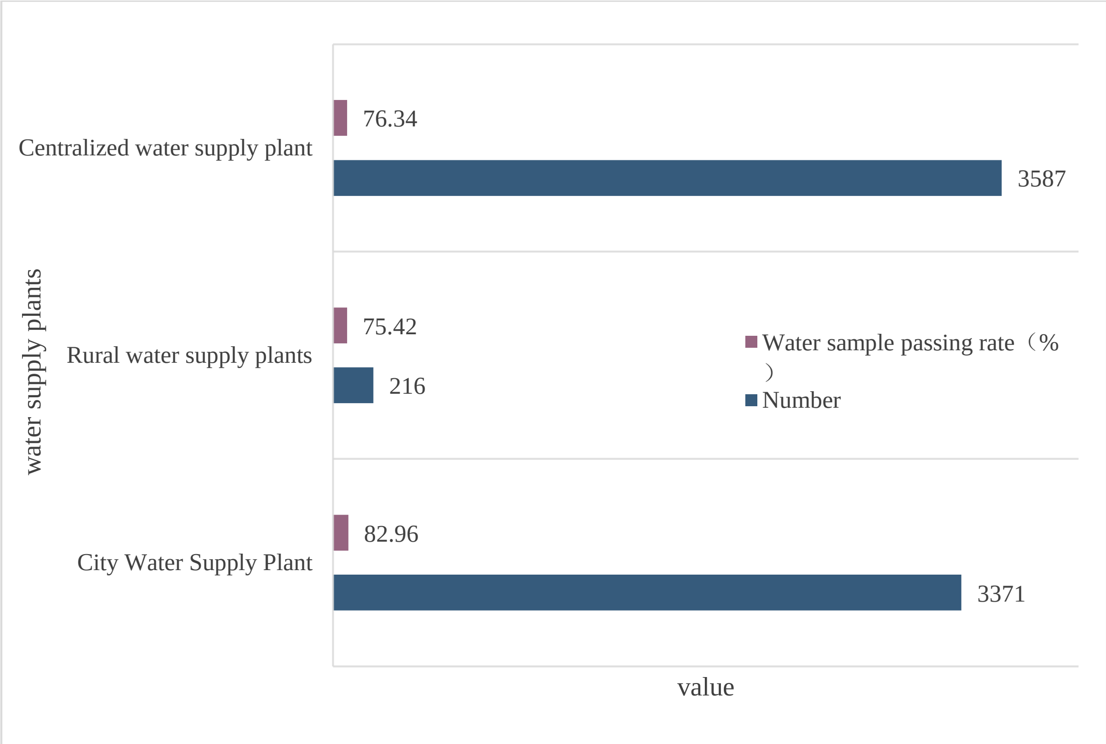

Figure 3. Water quality passing rate of water plant

4. Q City DW Pollution Prevention Management Measures and Suggestions

4.1. Unified Planning of DW Source Pollution Prevention and Control

Based on the many problems that exist in the current management of water pollution prevention and control in Q, it has become a priority to establish an independent management body for unified planning and deployment of water pollution prevention and control, integrating the resources of many relevant departments such as water conservancy, agriculture and fisheries, and giving administrative functions to this management body. The establishment of water pollution prevention and control management agencies at all levels, both urban and rural, creates an effective top-down, interlinked management system. In addition, the management of daily water use and the quality of water sources is organised within the management body [16]. For towns and cities (including towns and villages) the management of DW sources needs to be strengthened, and the 'subcontracting' system can be implemented by municipal officials or village officials, with each water source appointing a person to regularly check the condition of the water source and the water supply, forming a long-term record of water source inspections. See Figure 4 for an organisational chart.

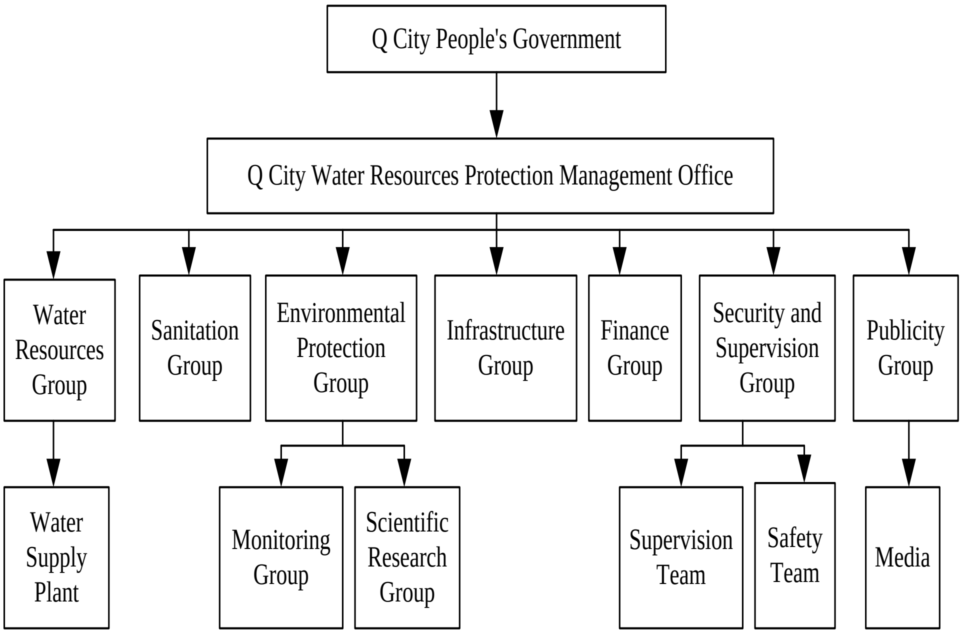

Figure 4. Organizational structure of DW source management in Q

Establishing an environmental pollution data reporting system to determine pollutant emissions is a fundamental task of integrated management, and each unit should regularly report the unit's raw material consumption to the Environmental Safety Technical Department, and the Ministry of Environmental Protection will confirm in detail the pollutant emissions generated by the unit based on a survey summarising the consumption of raw materials and relevant benchmark data.

4.2. Establish a Multi-Departmental Linkage Management Mechanism

Focus on the environmental protection of DW sources. The joint investigation team can conduct a comprehensive and detailed investigation of pollution sources around water sources from a number of departments, including environmental protection, urban construction and urban health, through special on-site meetings and office meetings, and rectify and implement the problems identified one by one. In order to establish and further improve the management and evaluation mechanism of the objective responsibility system and to strengthen coordination and cooperation according to their own needs, it is necessary to stop and take evidence of any discharge into the DW source and to promptly inform the environmental protection department. The relevant departments have their own responsibilities and are responsible for co-management and strengthening water source pollution management through joint enforcement and other means [17-18].

4.3. Promote the Construction of Water Source Infrastructure and Supporting Facilities

The construction of infrastructure at water sources is the basis of water source pollution prevention and management, and the completion of the infrastructure can more effectively and conveniently manage and protect water sources, which requires a degree of financial security. Water management agencies should unify planning and deployment, set appropriate standards to regulate water use and housing standards, and invest sufficient funds to complete the construction of standard housing for water sources in a phased and regular manner. Set up water source signs at the default water source location and note down basic information such as the name, number and date of creation of the water source. Set warning signs around the water source and standardise the size, specification, shape and other features of the signs, which must contain slogans reminding pedestrians and vehicles to drive carefully, along with a supervisory and management telephone number.

5. Conclusion

The Earth's water resources are mainly salt water, which is not directly drinkable, and three per cent of fresh water resources are mostly stored in the form of solid ice in the polar regions, so less than one per cent of water is actually available to humans. By these estimates, less than seven per cent of freshwater resources are currently available to mankind. In the current era of population explosion, the population and water consumption are geometrically increasing, the world is facing the problem of water shortage, especially China is one of the countries with extreme water shortage. In order to solve the problem of DW pollution, this paper on the current situation of DW pollution in Q city as an example, through the analysis of the city's groundwater quality exceeds the standard and the water quality of the water supply plant pass rate, put forward the corresponding water pollution prevention measures, in order to provide reference recommendations for other provinces and cities.

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] Fahimeh Motamedi, Horacio Perez Sanchez, Alireza Mehridehnavi, Afshin Fassihi, Fahimeh Ghasemi. Accelerating Big Data Analysis through LASSO-RF Algorithm in QSAR Studies. Bioinform. (2022) 38(2): 469-475. https://doi.org/10.1093/bioinformatics/btab659

[2] Valeria D'Amato, Rita Laura D'Ecclesia, Susanna Levantesi. ESG score prediction through RF algorithm. Comput. Manag. Sci. (2022) 19(2): 347-373. https://doi.org/10.1007/s10287-021-00419-3

[3] Mirna Nachouki, Mahmoud Abou Naaj. Predicting Student Performance to Improve Academic Advising Using the RF Algorithm. Int. J. Distance Educ. Technol. (2022) 20(1): 1-17. https://doi.org/10.4018/IJDET.296702

[4] Sridevi Subbiah, Kalaiarasi Sonai Muthu Anbananthen, Saranya Thangaraj, Subarmaniam Kannan, Deisy Chelliah. Intrusion detection technique in wireless sensor network using grid search RF with Boruta feature selection algorithm. J. Commun. Networks. (2022) 24(2): 264-273. https://doi.org/10.23919/JCN.2022.000002

[5] C. Venkata Narasimhulu. An automatic feature selection and classification framework for analyzing ultrasound kidney images using dragonfly algorithm and RF classifier. IET Image Process. (2021) 15(9): 2080-2096. https://doi.org/10.1049/ipr2.12179

[6] Musavir Hassan, Muheet Ahmed Butt, Majid Zaman. An Ensemble RF Algorithm for Privacy Preserving Distributed Medical Data Mining. Int. J. E Health Medical Commun. (2021) 12(6): 1-23. https://doi.org/10.4018/IJEHMC.20211101.oa8

[7] T. S. Deepu, V. Ravi. Modelling of interrelationships amongst enterprise and inter-enterprise information system barriers affecting digitalization in electronics supply chain. Bus. Process. Manag. J. (2022) 28(1): 178-207. https://doi.org/10.1108/BPMJ-12-2020-0554

[8] Kabelo Given Chuma, Mpho Ngoepe. Security of electronic personal health information in a public hospital in South Africa. Inf. Secur. J. A Glob. Perspect. (2022) 31(2): 179-195. https://doi.org/10.1080/19393555.2021.1893410

[9] Khoury Boutrous, Fatiha Nejjari, Viceng Puig. Reliability-Aware Zonotopic Tube-Based Model Predictive Control of a DW Network. Int. J. Appl. Math. Comput. Sci. (2022) 32(2): 197-211.

[10] Vinh Q. C. Tran, DucV. Le, Doekle R. Yntema, PaulJ. M. Havinga. A Review of Inspection Methods for Continuously Monitoring PVC DW Mains. IEEE Internet Things J. (2022) 9(16): 14336-14354. https://doi.org/10.1109/JIOT.2021.3077246

[11] Maria Benbouzid, Jamal Mabrouki, Mahmoud Hafsi, Driss Dhiba, Souad EI Hajjaji. Analysis and simulation of a reverse osmosis unit for producing DW in Morocco. Int. J. Cloud Comput. (2021) 10(5-6): 645-654. https://doi.org/10.1504/IJCC.2021.120400

[12] Diana Tapia-Pacheco, Laura Liliana Villa-Vazquez, Miguel Angel Perez-Angon. Research networks on the access of DW in Mexico City (2004-2018). Scientometrics. (2021) 126(3): 2557-2573. https://doi.org/10.1007/s11192-020-03569-4

[13] Ritsche Anne Kloosterman, Jan Peter Van Der Hoek. An integrated system approach to characterise a DW infrastructure system. Int. J. Crit. Infrastructures. (2020) 16(1): 1-22. https://doi.org/10.1504/IJCIS.2020.10027124

[14] S. Kavi Priya, G. Shenbagalakshmi, T. Revathi. Applied fuzzy heuristics for automation of hygienic DW supply system using wireless sensor networks. J. Supercomput. (2020) 76(6): 4349-4375. https://doi.org/10.1007/s11227-018-2341-6

[15] Soumia Bouzid, Messaoud Ramdani, Salah Chenikher. Quality Fuzzy Predictive Control of Water in DW Systems. Autom. Control. Comput. Sci. (2019) 53(6): 492-501. https://doi.org/10.3103/S0146411619060026

[16] Swati Chopade, Hari Prabhat Gupta, Rahul Mishra, Preti Kumari, Tanima Dutta. An Energy-Efficient River Water Pollution Monitoring System in Internet of Things. IEEE Trans. Green Commun. Netw. (2021) 5(2): 693- 702. https://doi.org/10.1109/TGCN.2021.3062470

[17] Amal Agarwal, Lingzhou Xue. Model-Based Clustering of Nonparametric Weighted Networks With Application to Water Pollution Analysis. Technometrics. (2020) 62(2): 161-172. https://doi.org/10.1080/00401706.2019.1623076

[18] Adam Niewiadomski, Marcin Kacprowicz. Type-2 Fuzzy Logic Systems in Applications: Managing Data in Selective Catalytic Reduction for Air Pollution Prevention. J. Artif. Intell. Soft Comput. Res. (2021) 11(2): 85-97. https://doi.org/10.2478/jaiscr-2021-0006