Scholar Publishing Group
Water Pollution Prevention and Control Project
https://doi.org/10.38007/WPPCP.2022.030402
ISSN 2790-2730 Vol. 3, Issue 4: 11-19

Water Environmental Pollution Risk Assessment and Water Pollution Dispersion Simulation by Integrating GIS and Random Forest Algorithm

Jinkuk Park*

Laboratoire de Maitrise des Energies Renouvelables (LMER), Bejaia 06000, Algerie

**corresponding author*

Keywords: Random Forest Algorithm, GIS Technology, Water Pollution Dispersion, Risk Contribution Rate

Abstract: Pollution poses a huge threat to the water ecosystem and human production and life, and emergency response to water pollution (WP) incidents has become an important issue. The simulation of the pollution dispersion process is particularly important, and the simulation results can provide an important reference for emergency rescue and emergency drills, helping to predict WP incidents and reduce the risk of accidents. With the development of computer technology and GIS technology, pollution dispersion simulation results can be expressed in numerical form or in the form of graphical data. In addition, this paper also combines the random forest (RF) algorithm to assess the risk of water environment pollution, analyze the characteristics of pollutant content in a river basin, and assess the health risk of groundwater quality in the basin. The results show that the contribution of carcinogenic risk of groundwater in the region is high for Cr in the dry period and as in the flat and abundant periods.

1. Introduction

In order to scale up urban construction, industries should be removed from key urban areas to improve urban and commercial development and promote economic growth. It is highly likely that the groundwater at these abandoned chemical sites is already contaminated and the contamination may spread widely, not only affecting the vital activities of flora, fauna and microorganisms around the contaminated sites, but also causing imbalances in existing ecosystems and threatening health [1-2].

The advent of GIS technology has led to a growing body of research on the use of GIS to construct water quality models, and the way in which water quality models are integrated with GIS

is progressing. For example, some scholars have used the WASPS reserve model as an example of combining model simulation with data communication to facilitate integration with GIS and visual representation into GIS [3]. Some scholars have begun to discuss how traditional water environment models can be combined with database technology and visual programming tools to create a mathematical modelling framework for water environment visualisation [4]. Some researchers have developed a series of water quality simulations and GIS-based technology assessment tools, pinning a complete water quality response system and reflecting the close relationship between water quality models and GIS [5]. The combination of WP dispersion process simulation and GIS technology can provide important reference and decision support in the disposal of WP events, therefore, in the coming period of time and WP dispersion related research needs to be urgently developed and in-depth.

This paper firstly introduces the concept and model of RF algorithm, and then proposes the dynamic visualization method of WP based on GIS; then takes a river basin as an example, analyzes its temperature, pH, DO and other pollutant content characteristics, and conducts risk assessment of groundwater quality in the basin; finally discusses the one- and two-dimensional models of river WP dispersion, constructs a GIS-based WP dispersion simulation system, and analyzes the The functions realized by the system are discussed.

2. Related Algorithms and Methods

2.1. RF Algorithm

RFs are a combination of various weak classifiers and a number of learning strategies developed from a bag that incorporates learning ideas. RFs inherit the properties of Bagging with some improvements. The RF algorithm is an integrated algorithm based on the decision tree algorithm, which can effectively improve the ability of model generation to test the forest [6]. Compared with the generalisation ability of single decision trees, the partitioning of several attributes of decision trees in the random forest set has been greatly improved. After obtaining a certain number of decision trees, the random forest algorithm votes on these decision tree generation results to decide the category that becomes the highest voted [7], and the results are expressed as:

$$H(x) = \arg\max_{Y} \sum_{i=1}^{k} I(h_i(x) = Y) \tag{1}$$

$$Gini(p) = 1 - \sum_{k=1}^{K} p_k^2 \tag{2}$$

Where H(x) is the final output result, hi(x) is a single decision tree, I is the sex function, and Y is the output variable. k is the decision tree category, p_k is the kth category probability, and Gini(p) is its Gini index. The training process of RF is the process of continuous classification and selection of the extracted feature vectors, based on superpixel blocks for feature extraction, each superpixel block contains the corresponding color, texture, shape and other features, these features form a feature set after sampling and other steps to form a single decision tree model, which grows into a RF model after continuous splitting [8-9].

2.2. GIS-Based Dynamic Visualization Method for WP

GIS visualization technology has been widely used with visualization features such as intuitive

information and simple operation, but GIS has limited ability to express regular spatial data [10]. At present, the dynamic visualization methods of GIS mainly include the following three:

Frame-by-frame animation method: generating time-series images with the same spatial interval and spatio-temporal reference data, continuously transforming the images at a given frequency, and using the principle of human visual retention to generate a visual image of continuous motion. This method is more common and less difficult, but requires a large amount of data [11]. In order to represent GIS in geo-temporal data, a high data resolution in time and space is required. Theoretically, the best visualisation is achieved when the spatial resolution reaches the maximum human eye resolution [12]. In fact, kind of ideal views are difficult to achieve, leading to large data surpluses, and views can be lost with the cost of information.

Dynamic map method (map animation method): This method must create dynamic visual changes and then express the order of objects of dynamic data. For representing spatio-temporal events with a degree of regularity, less data redundancy and a higher degree of continuity, this method has great advantages. However, it can only be used for vector data spaces. Furthermore, depending on the regularity of the spatial events, raster data or weakly structured data cannot be expressed [13-14].

Voxel-based frame animation method: In order to visualise 3D spatial data, these data are modelled on 3D GIS data based on time, images are generated, and then 3D data are expressed using a frame-by-frame animation method. This method reduces the computational stress of displaying the data when aggregated, but it is still essentially an animated representation of frames, with the data disappearing during visualisation [15].

3. Risk Assessment of Water Environment Pollution in a River Basin Based on RF Algorithm

3.1. Characteristics of Environmental Pollutant Content of River Water

There are 10 monitoring points in a river basin, of which R1, R2, and R3 are the upstream basin, R4, R5, and R6 are the midstream basin, and R7, R8, R9, and R10 are the downstream basin. Since temperature, pH, and dissolved oxygen (DO) are important factors affecting water quality and aquatic life, temperature, pH, and DO were measured at each monitoring site in this river basin, and their spatial distribution is shown in Figure 1.

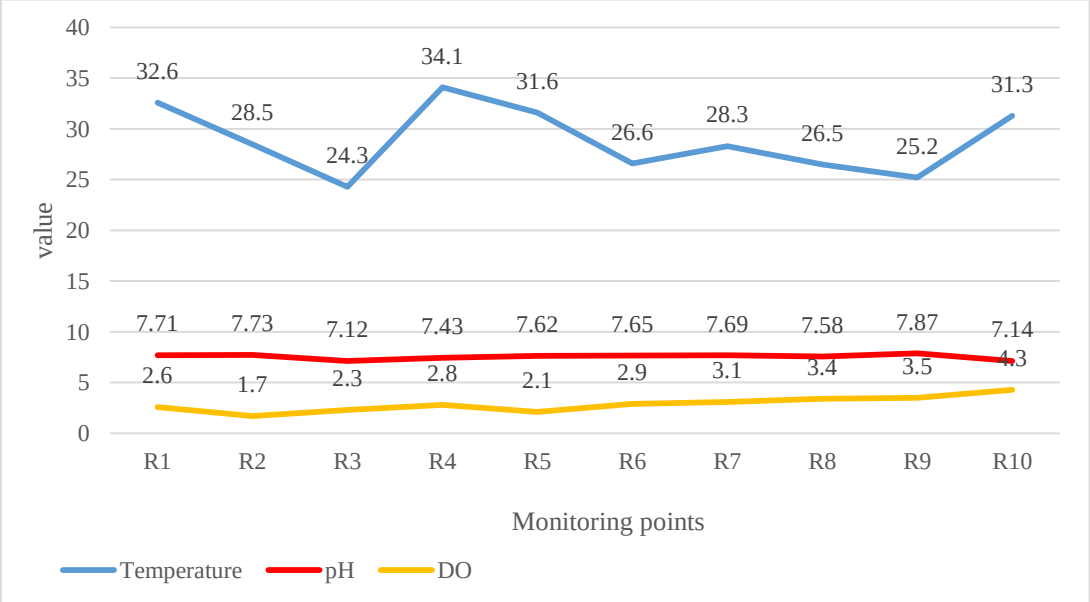

Figure 1. Spatial distribution map

Temperature directly affects the dissolution of pollutants in water and the rate of degradation of organic matter by microorganisms, etc., thus affecting the content of pollutants in the water body; in addition, temperature affects the groWPh and reproduction of aquatic organisms. The trend of temperature in this river basin is decreasing, then increasing, then decreasing and then increasing from upstream to downstream, and the temperature difference in midstream is significantly smaller than that in upstream and downstream.

The pH balance is the guarantee to maintain the virtuous cycle of aquatic ecosystem, too acidic or too alkaline will cause damage to the aquatic ecosystem. The trend of pH in this river is not obvious from upstream to downstream. The river is slightly alkaline overall, and the spatial variation of pH ranges from 7.10 to 7.90. The lowest pH values were found at points R3 upstream and R10 downstream of the river, and the pH values at these two points differed significantly from those at adjacent points, probably due to the influence of a small amount of acidic wastewater discharged into the vicinity at these two points.

Dissolved oxygen, which is related to water temperature, oxygen partial pressure and salinity in water, is an important condition for the survival of aquatic animals and plays an important role in improving water quality. The dissolved oxygen in this river went through two processes of decreasing and then increasing from upstream to downstream, and the change of dissolved oxygen in downstream was larger than that in upstream and midstream. The overall level of DO in the river was low, and the downstream was higher than the upstream and midstream, with a spatial variation range of 1.7-4.3 mg/L. The DO at point R2 in the upstream of the river was very low, below the surface water V standard, and was one of the most seriously polluted areas. In addition, the DO value at the midstream point R5 is also low, probably due to the discharge of organic pollutants into the vicinity, which also needs to be judged in conjunction with the index of organic pollutants. The downstream point R10 has the highest DO, which can reach the standard of surface water category IV, mainly due to the reduction of sewage discharged into the surrounding enterprises and the mixing and dilution effect of seawater.

3.2. Comprehensive Pollution Index Evaluation Results

The Nemero index method (P) is used to evaluate the water quality, although this evaluation method has advantages and disadvantages, it is widely used in China and the process is simple [16]. The evaluation model is :

$$P = \sqrt{\frac{p_{max}^2 + \overline{p}^2}{2}} \tag{3}$$

$\overline{p}$ is the average value of pollution index of all detection factors, and p_{max} is the maximum value of pollution index of each detection factor.

In this paper, the above method is used to evaluate the data of six monitoring points of the river in the dry, flat and abundant water periods respectively, and the results are shown in Figure 2.

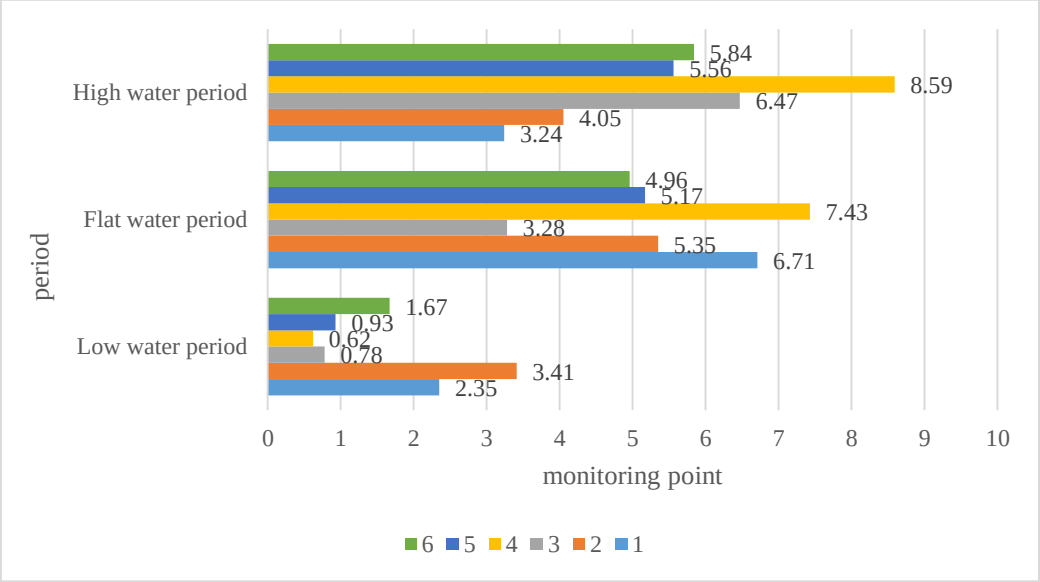

Figure 2. Pollution index evaluation

According to the data in Figure 2, it can be seen that each sampling site of the river was polluted to different degrees in three periods and the pollution level was high, then the groundwater quality of the river was also poor. There is a high risk of cancer if groundwater in the area is consumed. Heavy metals such as As, Cr and Ni were detected in the groundwater of the area, indicating that As, Cr and Ni are health risks. In contrast, Pb and Cd were not detected, indicating that Pb and Cd do not pose a health risk. As Table 1 and Figure 3 show the carcinogenic risk contribution(CRC) of As, Cr, Ni and other indicators for each period.

Table 1. Contribution rate of carcinogenic risk (%)

	Low water period	Flat water period	High water period
As	32.36	56.83	74.91
Cr	43.75	1.24	3.56
Ni	23.89	41.93	21.53

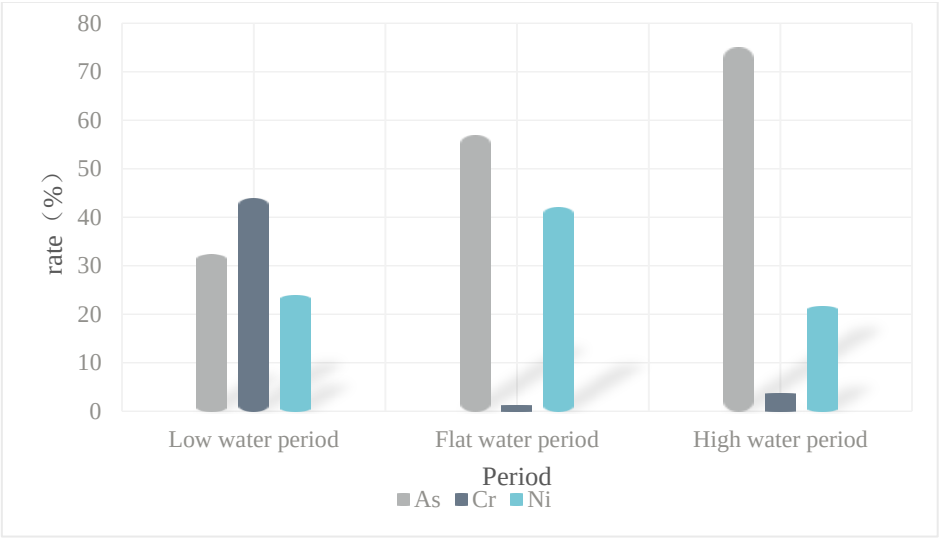

Figure 3. Risk assessment results in different periods

As shown in Figure 3, the CRC of As, Cr, and Ni were 32.36%, 43.75%, and 23.89% during the dry water period, and the highest CRC of Cr; during the flat water period, the CRC of As, Cr, and Ni were 56.83%, 1.24%, and 41.93%; during the abundant water period, the CRC of As, Cr, and Ni were 74.91%, 3.56% and 21.53%. The CRC of As was the highest and that of Cr was the lowest in the flat and dry water periods.

4. Modeling of WP Dispersion

4.1. Model Discussion

(1) River WP spreading model parameters discussion

The master data structure of the 1D river WP spreading model uses a 1D chain structure with a horizontal spacing unit size; the cell size in the 2D WP spreading model is the size of the pixels in the grid. The spatial resolution should be set to take into account the differences in search ranges, so blindly increasing the resolution will not improve the accuracy [17]. In the one- and two-dimensional river WP spreading model algorithm, the contaminant diffusion process is used as the change in contaminant concentration between cells, and it is necessary to ensure that the contaminant concentration is exchanged between all cells by less than 100%.

In the case of dispersed WP models in one- and two-dimensional rivers, the determination of diffusion coefficients and degradation coefficients (one-dimensional inclusion of diffusion, two-dimensional inclusion coefficients, pre-diffusion coefficients, joint diffusion coefficients) is related to the physical and chemical properties of the pollutant (solution, instability, etc.), and accurate diffusion coefficients and evapotranspiration coefficients need to be obtained through field tests [18].

(2) Discussion on the applicability of river WP dispersion models

From the perspective of model simulation, pollutants are mainly influenced by the flow of the river, the distribution along the river bank is negligible, so the one-dimensional WP dispersion model can be used for simulation; when the flow rate is small or even static, the spread of pollutants along the river bank is as important as the direction of the river, and the urgency cannot be ignored, this situation is chosen to use the two-dimensional WP dispersion model for simulation. And when the spatial scale of the study is large and the accuracy requirement is not high, the river width is negligible and a one-dimensional diffusion model can be used; when the spatial scale of the study is small, the river width cannot be neglected and if you want to achieve a certain accuracy you can use a two-dimensional diffusion model for simulation.

4.2. GIS-Based WP Diffusion Simulation System Construction

Based on the idea of hierarchical design, the watershed WP diffusion simulation system adopts a hierarchical C/S model architecture, including data layer, model layer, business layer and user interface layer. In order to improve the efficiency of data access, the system uses text files to store data on the client side, and the system realizes access to different types of spatial data through ArcEngine components. The overall structure of the system is shown in Figure 4.

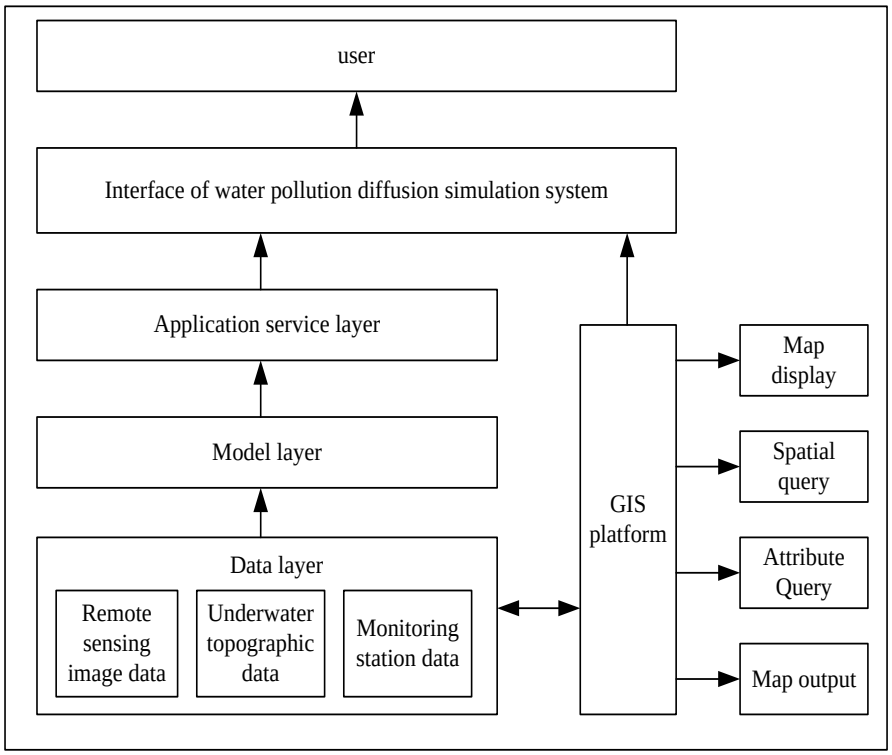

Figure 4. Overall structure of the system

WP dispersion simulation system mainly to achieve the following functions.

The hydrodynamic water quality mathematical modeling simulation and early warning forecast of each watershed, the simulation prediction of sudden WP accidents should have a basic reflection of the state of WP, such as the diffusion trajectory of pollutants, concentration, etc.

For sudden WP accident can make a rapid response. For sudden WP pollution accident, the efficiency and speed of model simulation has an important impact, requiring the simulation of hydrodynamic water quality model is best completed as soon as possible.

Through the organic combination of GIS and water quality mathematical model simulation technology, to achieve the visualization of changes in the hydrodynamic water quality of the watershed, the migration and transmission of pollutants changes; through the GIS-based time series dynamic simulation technology, to provide a visual demonstration of the evolution of the migration and transmission of pollutants in sudden WP accidents in the watershed.

5. Conclusion

WP incidents are extremely dangerous and can endanger people's health and safety in serious cases. In this work, we can help people prevent disease, ensure the safety of WP risks in river areas and analyse the role of heavy metal carcinogenic risks in water quality. The WP dispersion process simulation can predict the pollutant dispersion in a short period of time, understand the whole process and trend of continuous dispersion, assist relevant authorities in decision making and response to solve pollution events in a timely manner, and the GIS-based visualisation technology can more intuitively help decision makers to quickly understand the dispersion situation and cross the barriers of perception and understanding. This study is also highly scalable and extensible for the simulation and presentation of other geographic processes with spatial and temporal dynamics.

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] Pezhman Gholamnezhad, Ali Broumandnia, Vahid Seydi. An inverse model-based multiobjective estimation of distribution algorithm using Random-Forest variable importance methods. Comput. Intell. (2022) 38(3): 1018-1056. https://doi.org/10.1111/coin.12315

[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] Josalin Jemima J., D. Nelson Jayakumar, S. Charles Raja, Venkatesh P. Proposing a Hybrid Genetic Algorithm based Parsimonious RF Regression (H-GAPRFR) technique for solar irradiance forecasting with feature selection and parameter optimization. Earth Sci. Informatics. (2022) 15(3): 1925-1942. https://doi.org/10.1007/s12145-022-00839-y

[4] 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

[5] Syed Saqib Ali Kazmi, Mehreen Ahmed, Rafia Mumtaz, Zahid Anwar. Spatiotemporal Clustering and Analysis of Road Accident Hotspots by Exploiting GIS Technology and Kernel Density Estimation. Comput. J. (2022) 65(2): 155-176. https://doi.org/10.1093/comjnl/bxz158

[6] Victor Manuel Zezatti, Alberto Ochoa, Gustavo Urquiza, Miguel Basurto, Laura Castro, Juan Garcia. The Implementation of a Nickel-Electroless Coating in Heat Exchanger Pipes Considering the Problem of the Environmental Conditions ofthe Cooling Water Without Recirculation to Increase the Effectiveness Under Uncertainty. Int. J. Comb. Optim. Probl. Informatics. (2022) 13(4): 73-82.

[7] Mojtaba Barzegari, Liesbet Geris. Highly scalable numerical simulation of coupled reaction-Diffusion systems with moving interfaces. Int. J. High Perform. Comput. Appl. (2022) 36(2): 198-213. https://doi.org/10.1177/10943420211045939

[8] Hiromi Baba, Ryo Urano, Tetsuro Nagai, Susumu Okazaki. Prediction of self-diffusion coefficients of chemically diverse pure liquids by all-atom molecular dynamics simulations. J. Comput. Chem. (2022) 43(28): 1892-1 900. https://doi.org/10.1002/jcc.26975

[9] Konstantinos Petridis, Nikolaos Petridis. Diffusion of Innovations in Middle Eastern versus Western Markets: A Mathematical Computation Cellular Automata Simulation Model. Oper. Res. (2022) 22(2): 1597-1616. https://doi.org/10.1007/s12351-020-00598-y

[10] Angelika Zube, Dominik Kleiser, Alexander Albrecht, Philipp Woock, Thomas Emter, Boitumelo Ruf, lgor Tchouchenkov, Aleksej Buller, Boris Wagner, Ganzorig Baatar, Janko Petereit. Autonomously mapping shallow water environments under and above the water surface. Autom. (2022) 70(5): 482-495. https://doi.org/10.1515/auto-2021-0145

[11] Maria Gemel B. Palconit, Mary Grace Ann C. Bautista, Ronnie S. Concepcion II, Jonnel D. Alejandrino, Ivan Roy S. Evangelista, Oliver John Y. Alajas, Ryan Rhay P. Vicerra, Argel A. Bandala, EImer P. Dadios. Multi-Gene Genetic Programming of IoT Water Quality Index Monitoring from Fuzzified Model for Oreochromis niloticus Recirculating Aquaculture System. J Adv. Comput. Intell. Intell. Informatics. (2022) 26(5): 81 6-823. https://doi.org/10.20965/jaciii.2022.p0816

[12] Mohammad Najafzadeh, Farshad Homaei, Hadi Farhadi. Reliability assessment of water quality index based on guidelines of national sanitation foundation in natural streams: integration of remote sensing and data-driven models. Artif. Intell. Rev. (2021) 54(6): 4619-4651. https://doi.org/10.1007/s10462-021-10007-1

[13] Richard G. Bower, Benedict D. Rogers, Matthieu Schaller. Massively Parallel Particle Hydrodynamics at Exascale. Comput. Sci. Eng. (2022) 24(1): 14-25. https://doi.org/10.1109/MCSE.2021.3134604

[14] Arturo Vargas, Thomas M. Stitt, Kenneth Weiss, Vladimir Z. Tomov, Jean-Sylvain Camier, Tzanio V. Kolev, Robert N. Rieben. Matrix-free approaches for GPU acceleration of a high-order finite element hydrodynamics application using MFEM, Umpire, and RAJA. Int. JL High Perform. Comput. Appl. (2022) 36(4): 492-509. https://doi.org/10.1177/10943420221100262

[15] Maan Al-Zareer. Tunable hydrodynamic focusing with dual-neodymium magnet-based microfluidic separation device. Medical Biol. Eng. Comput. (2022) 60(1): 47-60. https://doi.org/10.1007/s11517-021-02438-3

[16] Mikolaj Marciniak. Hydrodynamic limit of the Robinson-Schensted-Knuth algorithm. Random Struct. Algorithms. (2022) 60(1): 106-116. https://doi.org/10.1002/rsa.21016

[17] David Abramov, Joseph N. Burchett, Oskar Elek, Cameron Hummels, J. Xavier Prochaska, Angus G. Forbes. CosmoVis: An Interactive Visual Analysis Tool for Exploring Hydrodynamic Cosmological Simulations. IEEE Trans. Vis. Comput. Graph. (2022) 28(8): 2909-2925. https://doi.org/10.1109/TVCG.2022.3159630

[18] Taher Abbasiasl, Hande Eda Sutova, Soroush Niazi, Gizem Celebi , Zeynep Karavelioglu , Ufuk Gorkem Kirabali , Abdurrahim Yilmaz , Huseyin Uvet , Ozlem Kutlu, Sinan Ekici, Morteza Ghorbani, Ali Kosar : A Flexible Cystoscope Based on Hydrodynamic Cavitation for Tumor Tissue Ablation. IEEE Trans. Biomed. Eng. (2022) 69(1): 513-524. https://doi.org/10.1109/TBME.2021.3100542