Simulation in the tasks of environmental monitoring of groundwater

T V Kozhevnikova¹, I S Manzhula¹* and L M Kondratieva²

¹Computing Center FEB RAS, Kim Yu Chena street, 65, 680000, Khabarovsk, Russia
²Institute of Water and Ecology Problems FEB RAS, Dikopoltceva street, 56, Khabarovsk, 680000, Russia

*E-mail: manzhula1994@gmail.com

Abstract. An algorithm has been developed that allows the processing of experimental data that are included in the external monitoring database of the Tunguska groundwater deposit. The problem of groundwater quality deterioration due to river filtration is exacerbated during severe floods. In the work, the selection of mathematical methods for solving the problems of simulation modeling is performed. The possibility of applying the methods of k-means cluster analysis, tree clustering and principal component analysis to extract similar objects from experimental data is shown. Mathematical methods are used to identify patterns and assess the impact of floods on the Amur river on the quality of groundwater in the Tunguska field, from the standpoint of multivariate analysis. The results of using the algorithm are presented on the example of a sample from the database on the content of aromatic compounds in groundwater samples from wells and the Penzenskaya Protoka. It was established that the Penzenskaya Protoka is isolated by indicators of the content of organic compounds, and the wells are grouped by the year of sampling, which indicates significant differences in the content of aromatic compounds in groundwater in the river filtration zone. The hypothesis of a significant impact of the 2013 flood on the Amur river was indirectly confirmed on the quality of groundwater.

1. Introduction

The globalization of the economy and the development of agricultural production are taking place on a larger scale compared with the ability to assess changes in the ecological state of natural resources under the influence of anthropogenic factors. The only event that is considered at the global level and recognized as a priority is the growing threat of climate change with an increase in the number of natural hazards [1, 2, 3], which can have a significant impact on the state of the land cover and the quality of surface and underground waters.

Over the past three decades, China has widely used groundwater for growing rice over large areas, which has led to significant changes in the quantity and quality of groundwater. This is especially true of coastal wetlands, where there is an active interaction of river and groundwater. Since the 1960s, large-scale degradation of floodplain soils has occurred on the Sanjiang plain, and about 80% of wetlands have disappeared due to a decrease in the depth of groundwater [4].

One of the acceptable ways to assess the consequences of human interaction with the natural environment is to monitor its change, analyze the state and predict possible consequences. All these components (observation, analysis and forecast) are the components of environmental monitoring,
which is carried out at various levels (global, regional and local). A thorough analysis of the current situation with obtaining reliable information for reliable prediction of changes in the state of the environment showed that the main problem is the uncertainty of our knowledge about natural processes [5].

In the last decade, it has been shown that the quantity and quality of surface waters is largely determined by global climate change and an increase in anthropogenic pressure. The catastrophic flood on the Amur river in 2013 was widely discussed in scientific publications from various positions: climatic, meteorological and hydrological. Intensive spring floods due to snowy winters and summer rain floods formed on almost all tributaries of the Amur river. The displacing flood from the western part of the basin took at its maximum the floods of the rivers of the eastern part of the basin, causing the cascade development of the flood [6].

In August 2013, when the water level in the Amur river reached 773 cm, the content of organic compounds increased significantly, among which aromatic compounds prevailed. Water during this period was characterized by maximum color for the entire observation period. Due to the extensive floodplain area, the water color values were much higher than during the floods in previous years.

In early September, the water level in Amur river near Khabarovsk stabilized and reached a historic maximum of 808 cm, while at Komsomolsk-on-Amur it continued to rise and reached its maximum of 916 cm on September 15. On the ridge of the flood, organic compound content and water color decreased, especially in the middle of the river, compared with coastal areas. However, the content of dissolved organic compounds on the left bank remained at a high level for a long time [7]. This was reflected in their increased content in the Penzenskaya Protoka, in which the natural waters of the Bureya river are characterized by a high content of humic substances, the precursors of aromatic compounds.

In some areas, the duration of flooding of the river floodplain to a depth of 2–4 m was more than 2 months. The width of the spills reached 20–30 km, capturing hayfields and pastures. The total duration of the flood near Khabarovsk was 115 days. As a result, the flooded vegetation was decomposed, and water-soluble products entered the underground aquifer with surface waters. Among the decomposition products of lignocellulose, toxic organic compounds are found, including methylated benzene derivatives (toluene, xylenes) and polycyclic aromatic hydrocarbons [8].

According to previous calculations by hydrogeologists, it was suggested that during the flood period of 2013, the active influence of river waters on the hydrodynamic conditions of the underground hydrosphere was not observed due to the presence of cover loams. At the same time, recommendations are made “on the reasonable removal of underground water intakes from the river contour” [9].

Despite the fact that numerous attempts are made to simulate the behavior of pollutants in groundwater in the river filtration zone, there are many unsolved problems. First of all, this is due to the multicomponent pollution of the aquifer and the complex dynamics of biogeochemical processes that occur during the interaction of water with rocks and organic compounds newly introduced with surface waters.

The paper proposes an algorithm for the study of groundwater (based on simulation) in order to classify objects (clusters, groups of wells) according to the presence of similar biochemical parameters, assuming that it is possible to distinguish a class of objects by distance from the Penza canal shore. This will prove that the surface water of the Amur river has an effect on the content of organic compounds in groundwater samples spreading in the Penza duct.

To understand natural processes, it is necessary to analyze the monitoring data of natural objects. In the process of analysis, it is required to identify similar objects, observing which you can better understand the laws by which changes in these objects occur. Note that classification is one of the fundamental processes in science. Quite often, there is a need to classify many objects according to several factors. For such a multidimensional classification, cluster analysis methods are used. Clustering can be considered a procedure that, starting to work with a particular data type, converts them into cluster data.
The most widespread are hierarchical agglomerative methods and iterative grouping methods. When using the methods of cluster analysis, it is quite difficult to give unambiguous recommendations on the preference for using certain methods.

In conditions of the need for a multivariate analysis of the studied groundwater indicators and low data structure, an effective method for identifying similar features (the amount of organic compounds in water samples) is cluster analysis — a multiple quantitative classification method. At the same time, the elements and their combinations are not important in themselves, but as indicators of the presence of biochemical indicators of water, which depend on the distance of the wells, the depth of the filters for groundwater sampling and the seasonality of observations.

Based on all of the above, the aim of the research is formulated: on the basis of simulation, to develop an algorithm for identifying spatio-temporal factors that can affect the quality of groundwater in the river filtration zone both in the process of long-term monitoring and during catastrophic floods.

2. Materials and methods
The territory of the Amur Region is part of the province of iron, manganese and silicon-containing fresh groundwater. In the interfluve of the Amur and Tunguska rivers, the Tunguska underground water deposit has been explored for water supply in Khabarovsk. According to their hydrochemical composition, these are sodium bicarbonate, low-mineralized (up to 200 mg/dm³) waters with a high content of iron and manganese [10].

On the territory of the Tunguska groundwater deposit, an observation groundwater monitoring network has been built [11], consisting of several well clusters located at different distances from the main Amur river bed and the left bank of the Penzenskaya Protoka (figure 1).

![Figure 1. Map of the location of well clusters located at different distances from the main Amur river bed and the left bank of the Penzenskaya Protoka.](image)

Longline groups, consisting of three compactly located wells, are equipped with filters 2 m long at different depths of the aquifer. Group 1 is located at a distance of 50 m from the water edge, group 2 – 300 m from the coast, group 3 – more than 1000 m from the coast (table 1).

To develop the algorithm, we used data on the content of aromatic compounds in groundwater samples from nine wells, as well as from the Penzenskaya Protoka [12], which are presented in table 2. For comparison, we used data obtained during the flood period of 2013 and post-flood in 2014. The content of aromatic compounds in groundwater was determined on a Shimadzu UV-3600 spectrophotometer at a wavelength of 275 nm and was expressed in arbitrary units of absorption [13].
Table 1. Characterization of groundwater sampling sites of the Tunguska field.

| Well groups | Distance from shore, m | Well number | Filter depth, m |
|-------------|------------------------|-------------|----------------|
| Group 1     | 50                     | K 1-1       | 14.7           |
|             |                        | K 1-2       | 24.7           |
|             |                        | K 1-3       | 34.7           |
| Group 2     | 300                    | K 2-1       | 13.7           |
|             |                        | K 2-2       | 26.7           |
|             |                        | K 2-3       | 37.7           |
| Group 3     | 1000                   | K 3-1       | 20.0           |
|             |                        | K 3-2       | 39.40          |
|             |                        | K 3-3       | 53.8           |

Table 2. Spatio-temporal dynamics of the content of aromatic compounds by spectral characteristics at 275 nm in groundwater of the Tunguska deposit in 2013–2014.

| Sampling location, well number | April 2013 | August 2013 | September 2013 | November 2013 | March 2014 | June 2014 | August 2014 | November 2014 |
|-------------------------------|------------|-------------|----------------|---------------|------------|------------|-------------|----------------|
| Penzenskaya Protoka K1-1      | 0.264      | 0.714       | 0.398          | 0.385         | 0.582      | 0.558      | 0.354       | 0.324          |
| K1-2                          | 0.137      | 0.540       | 0.386          | 0.357         | 0.265      | 0.269      | 0.201       | 0.156          |
| K1-3                          | 0.203      | 0.511       | 0.349          | 0.298         | 0.243      | 0.257      | 0.153       | 0.162          |
| K2-1                          | 0.188      | 0.422       | 0.333          | 0.187         | 0.187      | 0.197      | 0.106       | 0.124          |
| K2-2                          | 0.261      | 0.420       | 0.256          | 0.200         | 0.204      | 0.225      | 0.110       | 0.161          |
| K2-3                          | 0.169      | 0.386       | 0.197          | 0.194         | 0.194      | 0.165      | 0.154       | 0.087          |
| K3-1                          | 0.213      | 0.547       | 0.298          | 0.198         | 0.182      | 0.228      | 0.135       | 0.104          |
| K3-2                          | 0.034      | 0.158       | 0.231          | 0.147         | 0.025      | 0.205      | 0.303       | 0.017          |
| K3-3                          | 0.118      | 0.238       | 0.230          | 0.163         | 0.065      | 0.132      | 0.074       | 0.034          |
| K3-3                          | 0.142      | 0.411       | 0.281          | 0.205         | 0.084      | 0.102      | 0.126       | 0.085          |

Data are presented in conventional units of absorption.

It is assumed that the content of organic compounds in groundwater varies depending on the removal of wells from the coastline and the depth of sampling. During flooding of the river floodplain as a result of flooding, groundwater is contaminated with aromatic compounds [12].

In order to study the reliability of this hypothesis, the following research algorithm based on simulation modeling is proposed. In the process, the implementation of which necessitates the analysis of multidimensional data obtained during simulation experiments, in particular, the task of dividing data sets into disjoint subsets.

In the course of simulation experiments, a lot of observations are obtained, which must be divided into disjoint subsets (clusters) [14].

As initial data for simulation modeling, data on the content of aromatic compounds (table 2) are used, which must be divided into disjoint subsets (clusters). The objects for clustering are wells, the data on which contain chemical indicators and seasonality.

For the study, clustering methods were chosen, which are representatives of the main methodological approaches to dividing the initial set of objects into clusters: k-means, tree clustering,
and principal component analysis.

The simulation algorithm based on cluster analysis methods for visualizing the steps of processing experimental data is presented in figure 2.

Figure 2. Data processing algorithm.

To perform each of the presented types of cluster analysis, the data for the study are presented in table 3. It is assumed that data analysis is performed in the programming environment R [15].

Table 3. Labels for 20 data points by sampling location and year.

| Designations       | Observation period |
|--------------------|--------------------|
|                    | 2013   | 2014   |
| Penzenskaya Protoka| 1      | 11     |
| K1-1               | 2      | 12     |
| K1-2               | 3      | 13     |
| K1-3               | 4      | 14     |
| K2-1               | 5      | 15     |
| K2-2               | 6      | 16     |
| K2-3               | 7      | 17     |
| K3-1               | 8      | 18     |
| K3-2               | 9      | 19     |
| K3-3               | 10     | 20     |
3. Results
After a preliminary study of the data sample (checking homogeneity, calculating descriptive statistics), k-means clustering was carried out, the results of which are shown in figure 3.

Figure 3. Results of data analysis using the k-means method. The asterisk (*) indicates the center of the selected cluster: Cluster 1 (light blue dot) – wells 1 and 2 of the groups in 2013; Cluster 2 (green dots) – wells of 1 and 2 groups in 2014; Cluster 3 (four red dots) – wells of the middle layer of the aquifer 2 of the group and the wells of the upper and the middle layer of the aquifer of 3 groups for 2013 and 2014; Cluster 4 (black dots and red dot) – the wells of the lower layer of the aquifer of groups 2 and 3 for 2014; Cluster 5 (dark blue dot) – the Penzenskaya Protoka.

K-means method. Five clusters are set, the data for which are distributed as the well is removed from the Penzenskaya Protoka, taking into account the year of sampling. Objects (wells) were divided into five clusters. A clustering sign is the removal of wells from the Penzenskaya Protoka, which is consistent with the hypothesis put forward. Note that it is not possible to identify an unambiguous judgment on the effect of distance on the content of organic compounds in samples only by these results.

Tree clustering method. The results are shown in figure 4. According to the study, the following clusters can be distinguished:
1. Data from the K3-3 well for 2014.
2. Data mainly for 2014 for all wells of 1 and 2 groups, and the lower layer of the aquifer 3 of the group.
3. Data mainly 3 groups for 2013.
4. Data mainly for 2013 on the Penzenskaya Protoka and the first group.
5. Data on the 2nd group for 2013.

The results of the study of data by the tree clustering method are in full agreement with the notions of hydrological and biogeochemical processes that could occur in the river filtration zone during the flood and after this event.

Principal component analysis. According to the results of processing the data presented in figure 5, the Penzenskaya Protoka is distinguished as a separate object with specific indicators characteristic of surface water, in contrast to groundwater. It should also be noted that as a result of the data analysis, the wells of the first and second clusters were grouped into different clusters in accordance with the years of sampling. This confirms the hypothesis that the 2013 flood on the Amur river had a significant impact on the change in the content of aromatic compounds in groundwater in the river filtration zone, which, in turn, indicates significant changes in the hydrodynamic conditions of the underground hydrosphere. However, this conclusion requires a number of additional arguments and
scientific justifications, which is the goal of further research.

**Complete Linkage with Correlation-Based Distance**

![Figure 4. Tree clustering data analysis results.](image)

**Figure 4.** Tree clustering data analysis results.

**Figure 5.** Principal component analysis data analysis.

4. Conclusion
An algorithm has been developed that allows to process chemical indicators that are included in the database of external monitoring of the Tunguska groundwater deposit, carried out by the Institute of Water and Environmental Problems of the Far Eastern Branch of the Russian Academy of Sciences on the instructions of Municipal Unitary Enterprise Vodokanal in Khabarovsk [11].

The results of using the developed algorithm for the following mathematical methods of cluster analysis are presented: k-means, tree clustering, and principal component analysis. It was found that to assess the content of aromatic compounds in the used dataset of water samples, the best results were obtained using principal component analysis. With its help, the existing differences in the quality of groundwater in wells located at different distances from the coast of the Penzenskaya Protoka are proved; aromatic compounds were present in groundwater in the floodplain flood zone in post-flood 2014. In favor of the principal component analysis, it should be added that this is the only one of the considered methods that managed to identify differences in the content of aromatic compounds in groundwater in the river filtration zone in accordance with the years of sampling. Thus, the hypothesis
of the significant impact of the 2013 flood on the Amur river was indirectly confirmed on the quality of groundwater.

This algorithm is recommended for use by specialists — hydrologists, hydrochemists and hydrobiologists — for processing experimental data on monitoring water quality.

References
[1] Goswami B N, Venugopal V, Sengupta D, Madhusoodanan M S and Xavier P K 2006 Increasing trend of extreme rain events over India in a warming environment Science 314 1442–5
[2] Leng G Y, Tang Q H, Huang M Y and Leung L Y R 2015 A comparative analysis of the impacts of climate change and irrigation on land surface and subsurface hydrology in the North China Plain Reg. Environ. Chang. 15 251–63
[3] Qin N, Wang J, Yang G, Chen X, Liang H and Zhang J 2015 Spatial and temporal variations of extreme precipitation and temperature events for the Southwest China in 1960–2009 Geoenviron. Disasters 2 4
[4] Wang X, Zhang G and Xu Y J 2015 Impacts of the 2013 extreme flood in Northeast China on regional groundwater depth and quality Water 7 4575–92
[5] Kondratyev K I, Krapivin V F, Lakasa X and Savinykh V P 2006 Globalization and Sustainable Development: Environmental Aspects (St. Petersburg: Science)
[6] Danilov–Danilyan V I, Gelfan A N, Motovilov Yu G and Kalugin A S 2014 Catastrophic flood of 2013 in the Amur River basin: formation conditions, repeatability assessment, modeling results Water resources 41(2) 111–22
[7] Shesterkin V P 2016 Change in the chemical composition of the Amur waters during the period of the historical flood in 2013. Water resources 43(3) 287–96
[8] Suseela V and Tharayil N 2018 Decoupling the direct and indirect effects of climate on plant litter decomposition: Accounting for stress-induced modifications in plant chemistry Glob. Chang. Biol. 24 1428–51
[9] Kulakov V I 2017 Groundwater pollution in the Middle Amur artesian basin Izvestiya RGO 149(5) 36–47
[10] Kulakov V I 2011 Geochemistry of underground waters of the Amur Region (Khabarovsk: IEPP FEB RAS)
[11] Kulakov V V and Andreeva D V 2016 Dissolved groundwater gases of the Amur-Tunguska interfluve Pacific geology 35(2) 83–93
[12] Kondratyeva L M and Andreeva D I 2018 The effect of floods on the Amur River on the dynamics of organic substances in groundwater Water resources 45(6) 634–44
[13] Thomas O and Burgess C 2007 UV–visible spectrophotometry of water and wastewater (Elsevier)
[14] Magnus Ya R, Katyshev P K and Perezetskiy A A 2004 Econometrics (Moscow: Business)
[15] Kabakov R 2014 R in action. Analysis and visualization of data in the R program (DMK Press)