Information system for diagnosing pollutants based on data analysis algorithms[1]

Yurii I. Molorodov

Institute of commputational technologies SB RAS
ICT SB RAS, Novosibirsk, Russia, yumo@ict.sbras.ru

Abstract. The paper shows the influence of factors, processes and consequences of pollution of components of urban ecosystems on water pollution and hydro chemical composition of water of small rivers and the Ob river basin. The structure of the prototype of an interactive information system (IAS), which provides for the possibility of obtaining a forecast for the next five years, is described.

Keywords: pollution, pollutants, information resources, control of pollution consequences.

1 Introduction

For at least two millennia, water quality has been steadily deteriorating and reaches levels of pollution where the use of water for various purposes is severely restricted or the water may be harmful to humans. This deterioration is related to socio-economic development within the river basin, but atmospheric transport of pollutants over long distances has now changed this picture: even remote areas can be exposed to indirect pollution.

The problem of pollution of urbanized territories is relevant all over the world. It is most acute in large industrial cities and their agglomerations, where, as a result of strong technogenic pollution of the natural environment, urban-geochemical anomalies are formed with a high accumulation of organic and inorganic pollutants in ecosystems. Cities serve as powerful sources of migration from air and water flows of pollutants to geochemically associated territories, where pollutants are included in regional and global biogeochemical cycles.

Sources and nature of pollution, water resources and the environment in General, cities are diverse and depend on the specifics of industrial production, natural conditions, etc. therefore, scientific and practical interest in studying the factors, processes and consequences of pollution of components of urban ecosystems does not weaken. The results of these studies are necessary for a comprehensive assessment of environmental risks of urban ecosystem functioning, development of measures to improve the quality of life in cities and reduce their negative impact on the environment.

2 Data sources for analysis and operation

Specialists of the Ministry of natural resources and ecology of the Novosibirsk region (MNR&E NR) regularly monitor the quality of the atmospheric environment and measure the degree of pollution of the atmosphere and water of water bodies. They state the increase in pollution every year. These circumstances were the reason for the start of work on creating an interactive information analytical system (IAS). Data from long-term observations for the period 2006 to 2018 are stored on the MNR&E NR`s Web resource, which is open to the public (https://dlh.nso.ru/). They are the result of monitoring changes in pollutants in water resources, in order to further analyze quantitative and qualitative indicators of the state of surface water bodies in the Novosibirsk region.

The goal of our work is to develop an architecture and create a pilot version of the IAS designed for integration and advanced processing of data processing of field observations, measurements and laboratory studies.is carried out.

3 Leading technologies and standards

Currently, there are few rivers in the world, that are not polluted by human waste. Fertilizers and pesticides from agricultural land enter rivers with sewage. And, they also get water from sewers and drainage ditches. Some factories drain streams of dirty water into rivers and lakes. Pollution of rivers and lakes with nitrate fertilizers is growing on the planet almost every week. Unfortunately, even if you ban the use of nitrate fertilizers tomorrow, the situation will worsen. Nitrates slowly, for many years, seep through the earth into the beds of rivers or lakes. Dirty sewage and fertilizers enter lakes and reservoirs and cause rapid growth of ooze-algae that suffocate river fauna and flora.

The source of problems from the past include pathogens, oxygen balance, eutrophication, and heavy metals. Currently, they are recognized, investigated, and the necessary measures to prevent pollution are identified.

Current time problems are of a different nature. These are, on the one hand, traditional point and more extensive sources of pollution (nitrates) and widespread pollution problems (synthetic organic matter), and, on the other hand, problems of the "third generation" associated with global cycles (acid rain, climate change) [2].

In the past, water pollution in developing countries was mainly caused by the discharge of untreated wastewater. These problems are now more complex as a result of the production of hazardous waste products and the rapidly increasing use of pesticides in agriculture. In fact, water pollution is now a more serious problem in some developing countries, at least in developing new industries, than in developed countries. Unfortunately, developing countries generally lag far behind in controlling their main sources of pollution. As a consequence, the state of the environment in developing countries is constantly deteriorating. There are a huge number of microbial agents, elements and chemical compounds that cause water pollution. They can be divided into the following categories: microorganisms, organic compounds that can be destroyed biologically, suspended particles, nitrates, salts, heavy metals, fertilizers, and organic micro-pollutants [3].

3.1 Proposed approach for solving the problem

To ensure effective measures to address the problem of water pollution, a spatial and temporal analysis of the level of water pollution is necessary, including the study of toxicity indicators based on geoinformation technologies. The analysis procedure includes taking samples, identifying the level of contamination, and examining the data obtained. A modern interactive information system has been created, to work with the collected material. Using the Leaflet library for visualization and analysis [7]. In addition to modern tools for diagnosing pollutants, algorithms for working with an interactive map, have been developed based on data analysis algorithms, with the ability to display sample collection sites and information about research results.

Leaflet is an open source library written in JavaScript designed for displaying maps on websites. Supports most mobile and desktop platforms that support HTML5 and CSS3.

Along with OpenLayers and Google Maps, the API is one of the Most popular JavaScript map libraries used on major sites such as Flickr, Foursquare, Craigslist, and Google maps. Data.gov, IGN, Wikimedia projects, OpenStreetMap, Meetup, WSJ, MapBox, CloudMade, CartoDB, and others.

Leaflet allows a developer who is not familiar with GIS to easily display raster maps consisting of small tile fragments, with possibly additional layers overlaid on top of the main one. Layers can be interactive, such as displaying a hint when you click on a marker.

In addition to displaying information on a geographical map, sometimes there is a need for statistical analysis accumulated over several years. The Chart JS library is used for this purpose (https://www.chartjs.org/) written in JavaScript. It provides a variety of pre-created charts, such as combined histograms, bar charts, sector charts, geosystems, and others.

4 Information system architecture and its functioning

The architecture of an information system is defined by its components, their functions, and their interaction. The system is based on client-server technology and consists of a client part, a server, and a JSON data structure. A client-server is a computing or network architecture in which tasks, or network load, are distributed between service providers, called servers, and service customers, called clients. In the described architecture, the client is the browser (user), and the server is the web server. The system logic is distributed between the server and the client, data is stored in JSON format, and information is exchanged over the network. The advantage of this approach is that clients do not depend on the user's specific operating system, so the system is a cross-platform service.

An important component of the system is the client part of the application (Fig. 1). The map service is based on an interactive map from the Leaflet library (http://leafletjs.com).

It was decided to use JSON, as the data format for exchange between the client and the server, since:

- has built-in support in popular browsers released after 2009;
- native JavaScript format.

Figure 1 shows the architectural components of the data access system and the scheme of their interaction.

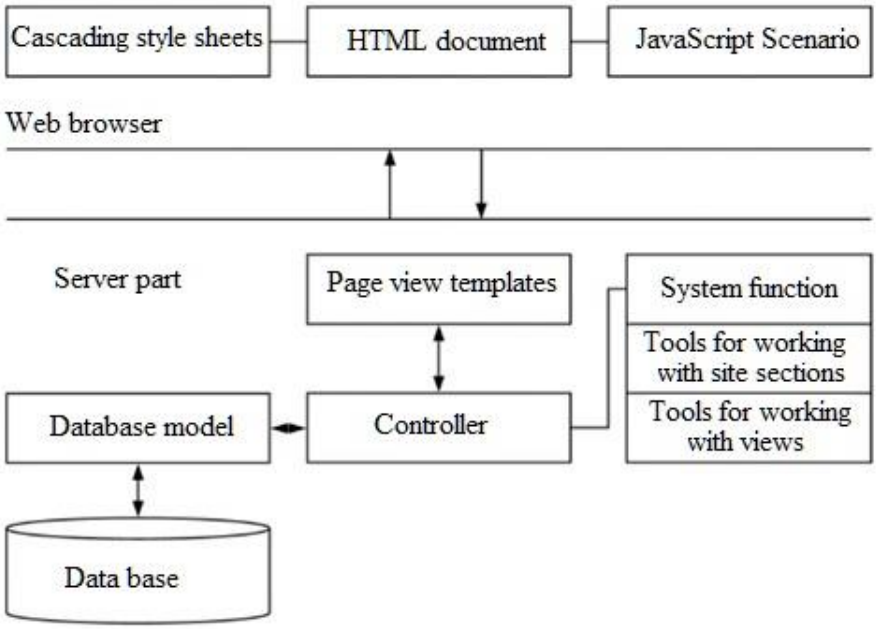

Figure 1. Architecture of data management system from external sources.

Data in JSON format, according to RFC 4627, can be: JavaScript objects, arrays, quoted strings, numbers, Boolean values, null.

The format of the data structure JSON:

Id: Number, Name: String, Coordinate: Array,

UKISV: Object {Years: Array, Value years: Array},

Description: String …

To get an object view, it is enough to use native JavaScript functions for converting from a string view to an object view, using the deserialization operation.

Data in JSON format, according to RFC 4627, can be: JavaScript objects, arrays, quoted strings, numbers, Boolean values, null.

4.1 Main modules of the information system

The IAS functionality provides a set of software implementations made in the form of separate modules. Let's highlight 8 of them (Fig. 2).

1. The research data processing module unloads the structure from a JSON file.
2. The geographical data processing module uploads the coordinates of the Novosibirsk region borders from a GeoJSON file.
3. The data processing module on the server is responsible for uploading data from the data structure.
4. User tracking module-reacts to the user's action and selects the desired action depending on the logical value.
5. The visualization module is responsible for displaying information on the screen.
6. The report module is responsible for creating and processing dynamic reports.
7. The query module is a component responsible for displaying information about the dynamics of changes in water pollution from time to time.
8. The forecast module is responsible for displaying information based on the research forecast for five years ahead.

Each module performs an independent function. If you need to make changes to a module, you just need to change a certain function. Any module can be expanded, and improved, and new functionality can be added.

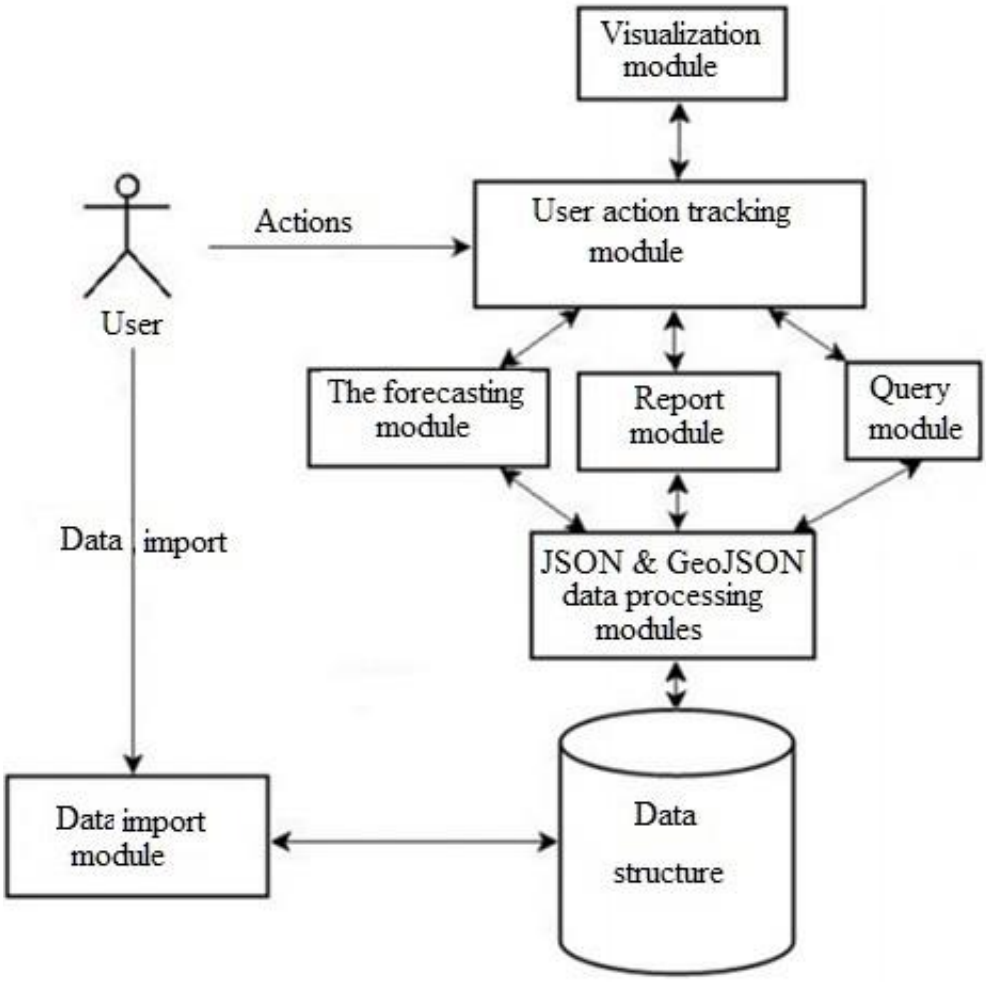

Figure 2. Information system`s modules

Research data processing module. The current state of most surface water bodies, and coastal areas does not meet current environmental and urban planning requirements. The change in the natural regime and the unfavorable state of most water bodies are affected by the visibility area [4]:

- Anthropogenic loads – releases of wastewater, discharges of pollutants, placement of objects in water protection zones and coastal protection strips, etc.
- Natural factors-climate changes lead to a decrease in the water content of many rivers in the territory, a prolonged low-water cycle for the Ob river, an increase in salinity of surface waters, drying up of drainless lakes, as well as rotting water plants, lack of oxygen [5];
- Man – made reasons causing the deterioration of runoff formation in watersheds and water treatment facilities.

In this regard, a module for processing water quality data was created based on the research of the MNR&E NR. After processing is completed, the visualization module creates a line graph where you can see the indicators of the specific combinatorial index of water pollution (ICIWP) [1] for one year.

The following gradation of water quality is accepted depending on the values of ICIWP (Table 1.).

Table 1. Gradations of quality.interaction.

Quality class	***Characteristics***	***ICIWP without regard CWP****
1 class	*conditionally pure*	*1*
2 class	*slightly polluted*	*(1-2)*
3 class	*contaminated*	*(2-4)*
category «А»	*contaminated*	*(2-3)*
category «Б»	*very polluted*	*(3-4)*
4 class	*dirty*	*(4-11)*
category «А»	*dirty*	*(4-6)*
category «Б»	*dirty*	*(6-8)*
category «В»	*very dirty*	*(8-10)*
category «Г»	*very dirty*	*(8-11)*
5 class	*extremely dirty*	*(11- ∞)*

*Critical water pollution indicators (CWP).

With the help of the geographical data processing module, the border of the Novosibirsk region is formed. The boundary coordinates are stored in a special data structure of the GeoJSON format. After processing, the visualization module adds borders to the map through layers**.**

The forecasting module. This module processes data series and analyzes them using linear approximation. The forecast is made using the least squares method.

Approximation of a function *f(x)* is called finding a function (approximating function) *g(x)* that is close to the given function. Criteria for proximity of functions can be different [8].

If the approximation is based on a discrete set of points, the approximation is called point or discrete. If the approximation is performed on a continuous set of points (segment), the approximation is called continuous or integral. An example of such an approximation is the decomposition of a function into a Taylor series, that is, replacing some function with a power polynomial.

The most common type of point approximation is interpolation-finding intermediate values of a quantity from an existing discrete set of known values. Let be a discrete set of points called interpolation nodes, as well as the values of the function at these points. You need to build a function *g(x)* that passes closest to all the specified nodes. Thus, the proximity criterion of the function is $g(x_i)=y_i$.

The function *g(x)* is usually chosen as a polynomial, which is called an interpolation polynomial. If the polynomial is one for the entire interpolation area, it is said that the interpolation is global. If the polynomials are different between different nodes, we talk about piecewise or local interpolation. Having found an interpolation polynomial, we can calculate the values of the function between nodes, as well as determine the value of the function even beyond the specified interval (extrapolation) [9].

User help. To launch the web application, the user enters the URL of the information system in the string. On the client side, the browser renders the resulting code and presents the user with a set of visual controls: an interactive map, navigation tools, zooming, and switching the map type (Fig. 3).

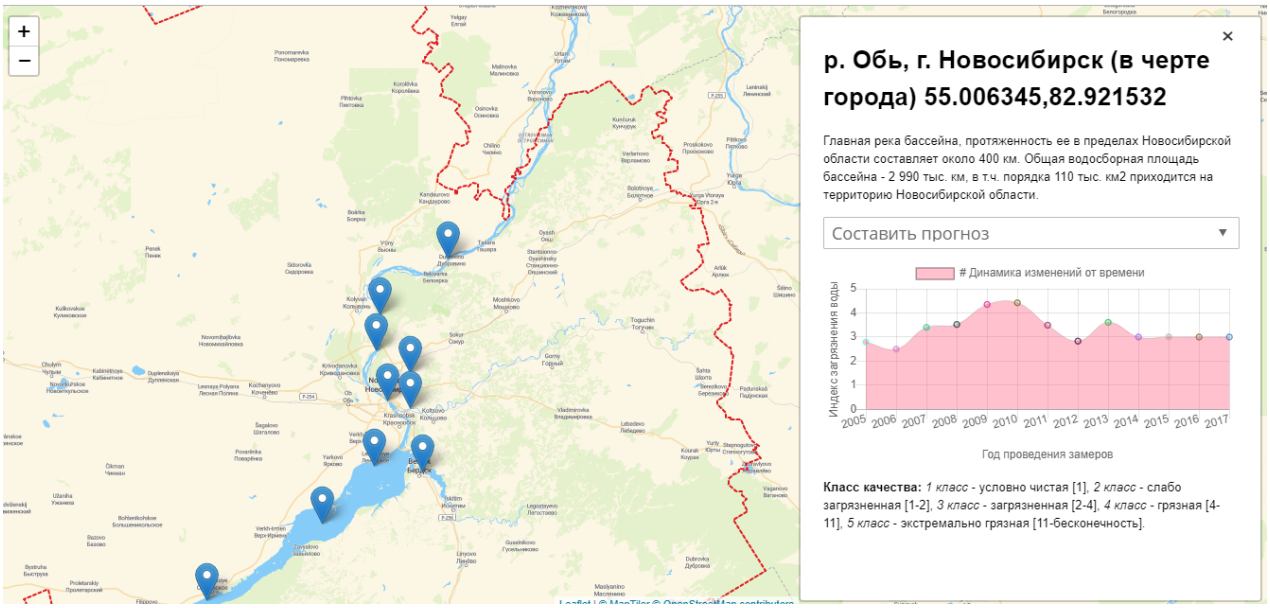

Figure 3. Interactive map with a list of water sampling locations

The interactive map is the main component of the system interface. Interactive mode allows you to easily move around the map and change its scale. Detailed information is displayed through the user interface, which is graphically and logically divided into three Windows. The window on the left shows the map and the scale that can be changed, for this purpose, the corresponding buttons are created in the upper left window. There are markers on the map that mark the places where research works are carried out. Clicking on the marker results in information about the data received on the site (Fig. 3).

The upper-right window shows a hierarchical list of all regions of Russia and Kazakhstan. Here you can select a country, field work area and get information about it. Objects located on the map have different attribute information that POPs up as a window when you click on the marker (Fig. 3). This is the physical name of the object, its administrative name, research period, coordinates, quality class according to ICIWP, a brief description of the reservoir. By clicking on the "Make a forecast" selection window (Fig. 4), it is possible to predict the level of contamination based on previous samples.

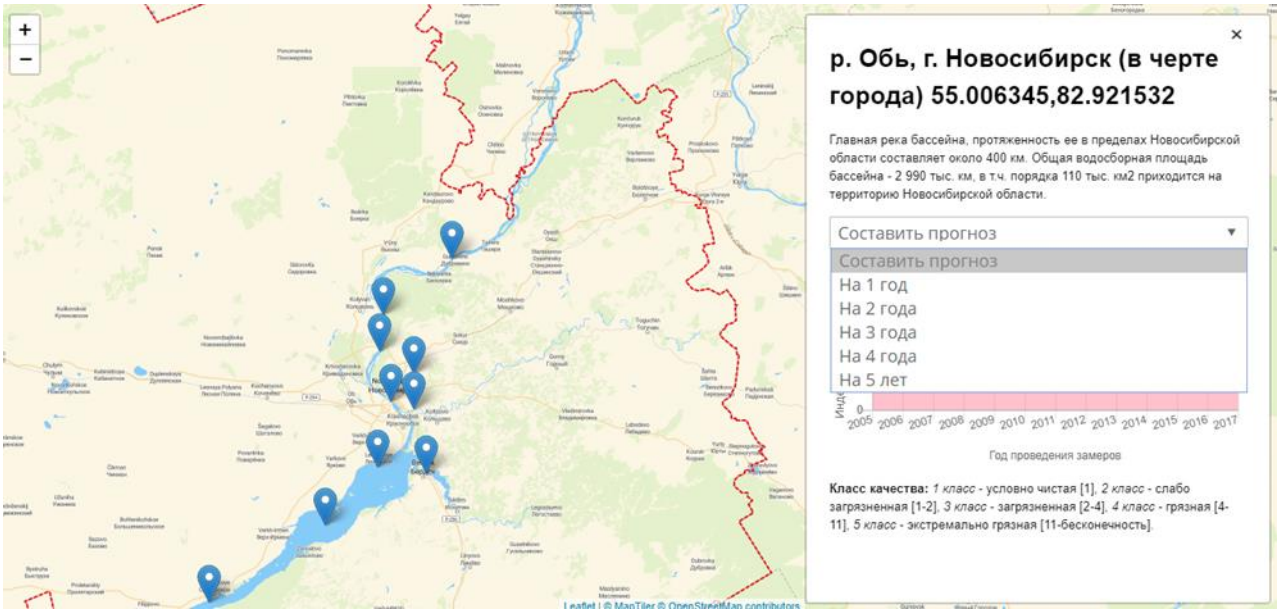

Figure 4. Make a forecast

After the analysis of the information the system will display the result as a "trend Line" (trendline is a geometric display of the average values of the analyzed indicators), also when you hover over the graph, you detailed rates ICIWP and the level of pollution (figure 5).

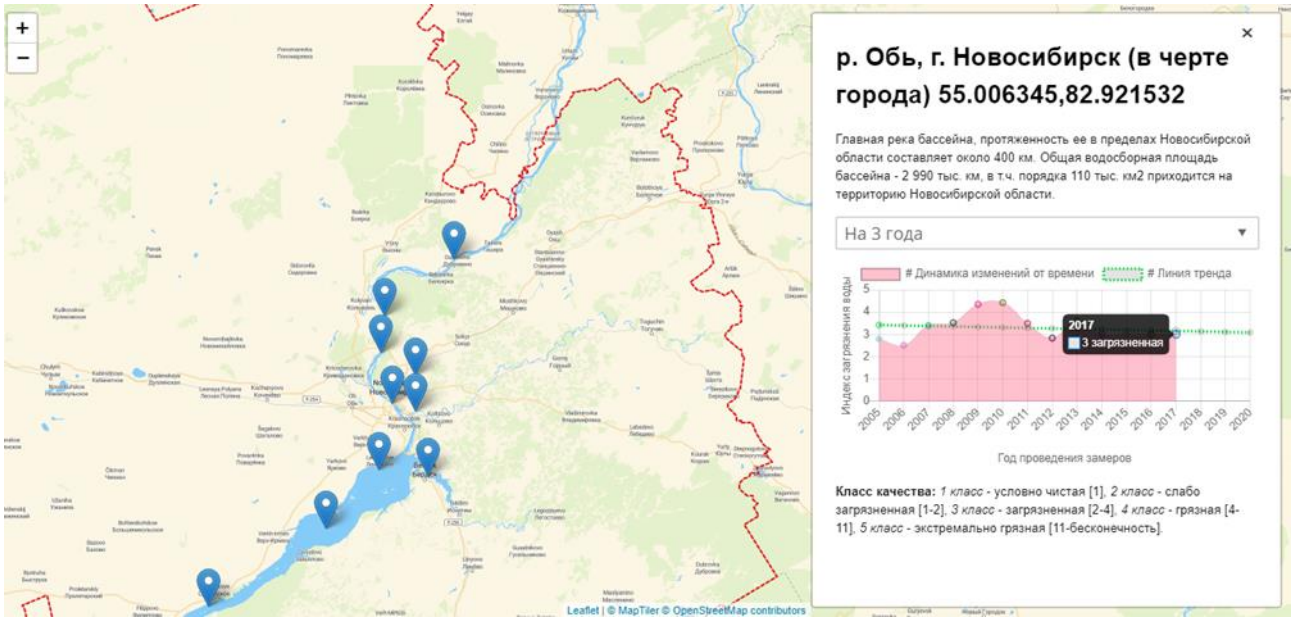

Figure 5. Trend line

5 Conclusion

A software package was developed based on the modern Leaflet library and methods of mathematical analysis of observation data to work with the results of long-term monitoring of the ecological state of water bodies in the Novosibirsk region. It allows you to quickly and efficiently analyze time series of data using statistical analysis algorithms, which indicates its great potential for scientific research. A software package was created based on the modern Leaflet library and methods of mathematical analysis of observation data to work with the results of long-term monitoring of the ecological state of water bodies in the Novosibirsk region. It allows you to quickly and effectively analyze time series of data using statistical analysis algorithms, which indicates its great potential for scientific research.

Acknowledgements. A project 2.3 “Creating models, algorithms, and integrated systems” supported this work.

References

[1] Ministry of natural resources and ecology of the Novosibirsk region [Electronic resource]. URL: https://dlh.nso.ru/
[2] Vorontsov A. I., Nikolaevskaya N. G. Issues of ecology and protection of the water environment. - Moscow: In-fra-M, 2011. - 98 p. (in russian)
[3] Danilov-Danilyan V. I., Losev K. S. Ecological challenge and sustainable development. - Moscow: Progress-Tradition, 2010. – 233 p. (in russian)
[4] Orlov D. S. Ecology and protection of the hydrosphere in chemical pollution: Studies. Handbook / Orlov D. S., Sadovnikova L. K., Lozanovskaya I. N.-M.: Higher school, 2012. – 167 p. (in russian)
[5] Protasov V. F. Ecology, health and protection of the water environment in Russia: Studies and help. Handbook / V. F. Protasov. - Moscow: Finance and statistics, 2012. – 289 p. (in russian)
[6] Sitarov V. A., Pustovoitov V. V. Social ecology. - Moscow: publishing center "Akade-Miya", 2010. – 236 p. (in russian)
[7] Ruzhnikov G. M. et al. Modern technologies of information and analytical assessment. SO RAMS. 2012. Vol. 32, no. 6. 57. Pp. 55-59. (in russian)
[8] Approximation [Electronic resource]. URL: https://ru.wikipedia.org/wiki/Аппроксимация (Yes-TA addresses: 05.02.2019) (in russian)
[9] Interpolation [Electronic resource]. URL: https://ru.wikipedia.org/wiki/Интерполяция (date accessed: 05.02.2019) (in russian)
[10] Method of least squares [Electronic resource]. URL: https://ru.wikipedia.org/wiki/Метод_наименьших_квадратов. (in russian)