Statistical processing and analysis of the monitoring results for the Gulf of Finland 2001 - 2019

A I Kushnerov and A I Shishkin

Higher School of Technology and Energy of the Saint Petersburg State University of Industrial Technologies and Design, 4, Ivana Chernykh str., Saint-Petersburg, 198095, Russia

E-mail: kushnerov.a.i@yandex.ru

Abstract. The scientific work is dedicated to solving the problem of providing environmental monitoring data for the basin of the north-eastern part of the Gulf of Finland for the adoption of water protection decisions based on the geoinformation system. As part of the work, a relational database of hydroecological data was developed based on the results obtained in seasonal research expeditions for the period 2001 - 2019. Using the capabilities of statistical processing, we analyzed long-term series of observations on the main indicators of water quality.

1. Introduction

Scientific research expeditions - Bios-schools on the territory of St. Petersburg and the Leningrad Region are being held for the 70th time by the Ecological Club (IPO "ECPS and SBLR") using the technological park of the HSTE SPbSUITE. Systematic seasonal observations at the established control points made it possible to accumulate hydroecological monitoring data for 19 years. The wide coverage of studies of water bodies in the Gulf of Finland and Lake Ladoga allows a comprehensive assessment of the problems of the region [1, 2] and the most anthropogenic-loaded sections of the basins, the level of wastewater load [3].

When analyzing existing methods and computer programs for creating geoinformation databases of environmental monitoring, ArcGis v.10 and MO Access (2013) were selected. For statistical data processing, the computer program Statistica Ultimate Academic v.13 was used.

2. Target of research

The territory of the basin in question is part of the Northwestern Federal District and is located on the territory of two constituent entities of the Russian Federation - the Leningrad Region and the city of St. Petersburg. The water basin belongs to the Baltic Basin District, to the hydrographic unit 01.04.03 - the Neva and the rivers of the Ladoga Lake Basin and includes one water section (WS) 01.04.03.005 - rivers and lakes of the Gulf of Finland basin from the Russian border with Finland to the northern border of the delta Neva River [4].

Sampling and field studies are systematically carried out by the following water bodies like “Smolyachkov” stream, “Chernaya” river, “Gladyshevka” river, “Roshchinka” river, “Sestroretsk” reservoir, “Malaya Sestra» river, Gulf of Finland and others. At different periods of time, environmental
monitoring was carried out on the catchment area of the river basin “Neva” and the rivers of Lake Ladoga, as well as the river basin “Luga” and rivers of the southern Gulf of Finland (figure 1).

Figure 1. Map of the basin of rivers and lakes of the north-eastern part of the Gulf of Finland.

Annual seasonal studies are carried out at approximately 60 points of control. The total number of main control points is 181, and additional (from different depths, along sections) - 213 points.

3. Geoinformation database
At the initial stage of the project, the developed database was taken as the basis: “Hydroecological data of the basin of the north-eastern part of the Gulf of Finland by seasons for the period 2001 - 2018”, which is registered and has an author's certificate [5]. Using the MS Access program, the database has been improved and translated into a relational view. For this, all control points were systematized and a new structure was developed, which includes the following tables: “Water bodies”, “Control points”, “Indicator values”, “Expeditions”. In addition, the database was filled with the results of the spring and summer expeditions of 2019.

The developed database structure includes 50 of the following fields (columns): ID, year, month, day, point number, water body name, GPS, water temperature, air temperature, atmospheric pressure, smell, foami ness, turbidity, color, specific electrical conductivity, pH, acidity, alkalinity, hardness total, calcium, magnesium, chlorides, total iron, sulfates, nitrate nitrogen, ammonium nitrogen, nitrate nitrogen, copper, lead, phosphates, dissolved oxygen, BODs, oligochaete index, Woodywiss index, index Shannon, saprobity, in dex toxicity, degree of toxicity. All indicators were divided into the following groups: auxiliary parameters, characteristics of control points, hydrochemical, hydrobiological and hydrological. In total, the database contains more than 64 thousand numerical and alphabetic values.

Using the accumulated map material in ArcGis v. 10. The studied water bodies were vectorized. The main points of hydroecological control are plotted. The following layers were created: rivers,
lakes, control points, disturbance in the environment. At the control points, a cartographic and attribution database was combined. In addition to filling the GIS database of environmental monitoring, photographs were added from sampling sites and field studies. Thus, the developed software geoinformation complex will allow reasonably making environmental decisions [6, 7].

4. Research school (expedition)
The last 70th research school was held from August 2 to 11, 2019, the research program of which was focused on the professional conduct of public environmental control in order to clarify the points of control, assessment and prediction of the ecological state of water ecosystems in the Kurortny district of St. Petersburg and Vyborgsky district of Leningradskaya area.

As part of this expedition, 18 natural water bodies. School participants completed 9 exits to water bodies. The studies were carried out at established control points (61 control points: sampling and field studies), while only 259 samples were taken: of which 47 were water samples from the surface, 9 from the depth of water bodies, 56 water samples for toxicity analysis, 17 samples water for determination of BOD$_5$ (biological oxygen demand in 5 days), 5 samples of drinking water, 29 samples of zooplankton and 17 samples of bottom sediments (zoobenthos) for bioindication of aquatic ecosystems. A total of 50 indicators of water quality and parameters of water bodies were identified.

The main areas of research: hydrochemical according to research methods (colorimetry, turbidimetry, titrimetry, organoleptic and physico-chemical methods), hydrobiological (biotesting and bioindication by zooplankton and zoobenthos), hydrological and hydrometeorological characteristics.

All the results obtained were compiled into a common database with reference to the geographic information system of the Gulf of Finland basin.

5. Statistical processing and data analysis
For processing and analysis of large data arrays, statistical methods and models are used [8]. To determine descriptive statistics (n is the sample size, $X_{\text{min}}$ is the minimum value, $X_{\text{max}}$ is the maximum value, $X$ is the average value, $S_0$ is the standard deviation, F is the frequency of occurrence), the Statistica Ultimate Academic v.13 program was used. To assess the ecological state, water quality indicators such as pH, $O_2$, and BOD$_5$ were chosen. The samples were taken over the entire time period from 2001 to 2019, as well as selections by season: spring, summer, and autumn.

According to BOD$_5$, a total of 501 values were processed. The most common values were in the range of 2.73–4.08 mg/l, which corresponds to an excess of the MPC by 1.5–2 times. The smallest values starting with BOD$_5$ concentrations of 10.81 mg/l or more, i.e. high values were less common than concentrations close to zero. When analyzing the statistical density distribution of BOD$_5$ values of the general sample and by season, it was found that in summer, higher concentrations of BOD$_5$ (2.47 - 3.64 mg/l) are most common than in other seasons (autumn 1.47 - 2.72 mg/l; spring 1.5 - 2.96 mg/l). The results obtained confirm that in the summer period higher concentrations of dissolved organic substances are observed due to a decrease in water level, increase in temperature and a decrease in the content of dissolved oxygen for oxidation. The standard deviations ($S_0$) were also calculated. Descriptive statistics for all samples for BOD$_5$ are presented in table 1.

Table 1. Summary table BOD$_5$ by season for 2001 – 2019.

| №  | Data extraction       | n   | $X_{\text{min}}$ | $X_{\text{max}}$ | $X$   | $S_0$ |
|----|-----------------------|-----|-----------------|-----------------|------|------|
| 1  | 2001-2019             | 501 | 0.04           | 18.90           | 4.15 | 2.87 |
| 2  | 2001-2019 (spring)    | 161 | 0.04           | 18.90           | 4.02 | 2.96 |
| 3  | 2001-2019 (summer)    | 232 | 0.13           | 13.01           | 4.31 | 2.54 |
| 4  | 2001-2019 (autumn)    | 104 | 0.22           | 10.22           | 3.50 | 2.17 |
When comparing the average concentrations of BOD₃ by season, the highest value was found in the summer period (4.31 mg/l), and the lowest average in the fall (3.5 mg/l). The minimum concentration was in the spring season (0.04 mg/l), and the maximum in the spring season (18.9 mg/l).

Statistical processing of 1226 O₂ values for the period from 2001 to 2019 yielded the following results: \( X_{\text{max}} = 18.9 \text{ mg/l} \), \( X_{\text{min}} = 0 \text{ mg/l} \), \( \bar{X} = 9.15 \text{ mg/l} \). The most common values of 9.45 - 10.5 mg/l are acceptable, since they correspond to the MPC = 6 mg/l. The smallest values starting with O₂ of 14.7 mg/l or more. When analyzing the statistical density of O₂ values of the total sample and by season, it was revealed that in the autumn period, the highest concentrations of O₂ (11.97 - 13.16 mg/l) are most common, as well as in the spring season (11.82 - 13.16 mg/l). In summer, the most common values of O₂ are in the range of 8.65 - 9.45 mg/l. This can be explained by seasonal changes in temperature and, as a consequence, a change in the solubility of oxygen in water. Statistical processing for all samples is presented in table 2.

**Table 2. Summary table for the statistical processing of O₂ values.**

| № | Data extraction | n     | \( X_{\text{min}} \) | \( X_{\text{max}} \) | \( \bar{X} \) | \( S_0 \) |
|---|----------------|-------|-----------------------|-----------------------|-------------|---------|
| 1 | 2001-2019      | 1226  | 0.00                  | 18.90                 | 9.15        | 3.07    |
| 2 | 2001-2019 (spring) | 233  | 0.26                  | 14.40                 | 9.80        | 3.63    |
| 3 | 2001-2019 (summer) | 750  | 0.00                  | 18.90                 | 8.70        | 2.89    |
| 4 | 2001-2019 (autumn) | 238  | 1.30                  | 14.35                 | 9.80        | 2.65    |

When comparing the average concentrations of O₂ over the seasons, the highest value was established in the spring and autumn (9.8 mg/l), and the lowest average in the summer (8.7 mg/l). The minimum and maximum concentrations in the summer season were 0 mg/l and 18.9 mg/l, respectively.

Among 1648 pH values, the most frequently encountered values are 6.24 – 7.85, showing a neutral environment, with a MPC of 6.5 – 8.5. The smallest pH values are in the extremes of 4.64 - 5.44 and 9.19 - 10. When analyzing the statistical density of pH values of the general sample and by seasons, it was found that the most frequent pH values are in the range 6.5 - 7, which corresponds to a neutral medium shifted slightly acidic. Statistical processing for all samples is presented in table 3.

**Table 3. Summary table for statistical processing of pH values.**

| № | Data extraction | n     | \( X_{\text{min}} \) | \( X_{\text{max}} \) | \( \bar{X} \) | \( S_0 \) |
|---|----------------|-------|-----------------------|-----------------------|-------------|---------|
| 1 | 2001-2019      | 1648  | 4.46                  | 10.0                  | 7.10        | 0.77    |
| 2 | 2001-2019 (spring) | 402  | 5.00                  | 10.0                  | 6.81        | 0.58    |
| 3 | 2001-2019 (summer) | 924  | 4.46                  | 10.0                  | 7.29        | 0.83    |
| 4 | 2001-2019 (autumn) | 323  | 5.00                  | 8.3                   | 6.90        | 0.69    |

When comparing the average pH values for the seasons of the year, the highest value was established in the summer period (7.29), and the lowest average in the spring (6.81). At the same time, the minimum value was in the summer season (4.46), and the maximum in the spring and summer seasons (10).

When analyzing the maximum values of the selected indicators, the following contaminated water bodies were identified: “Smolyachkov” stream, “Chernaya” river, “Malaya Sestra” river, “Kamenka” river. Statistical processing and subsequent evaluation of the results made it possible to determine the most common values and compare them with the MPC: analysis showed that in summer, higher concentrations of BOD₃ are most often encountered than in other seasons.

The results obtained confirm that in summer there are higher concentrations of organic substances. O₂ values showed that in the summer they are minimal. When analyzing the pH values, it was found that the most common values are in a neutral medium, shifted to a slightly acidic side.
The developed base of hydroecological data on the studied water bodies based on the geographic information system will allow to accumulate, store and search for information, as well as more objectively conduct a comprehensive assessment of the state of water resources to justify water-related measures. The developed geoinformation database can be used as additions to the decision support system in the field of water resources management.

Acknowledgments
This work was financially supported by the Committee on Science and Higher Education of St. Petersburg under the project "Creating a Relational Hydroecological Database Based on the Results of Long-Term Monitoring of the Basin of the North-East Gulf of Finland on a Geoinformation Base " as part of a competition for the provision of subsidies to young scientists, young candidates of science universities, industry and academic institutions located in St. Petersburg in 2019.

The authors are grateful for the many years of research on environmental monitoring of water bodies to the entire team of the leading scientific and pedagogical school of St. Petersburg " Forecasting and environmental standardization of the load on natural ecosystems ": Epifanov A.V., Antonov I.V., Stroganova M.S., Kuvatov I.Zh., Barkhurev Kh.O., Permilovskaya K.K. and others. All studies were carried out and are ongoing under the guidance of professor, academician of the International Academy of Ecology, Man and Nature Protection Sciences Alexander I. Shishkin.

References
[1] Moiseenko T I, Dinu M I, Gashkina N A, Jones V, Khoroshavin V Y and Kremleva T A 2018 Present status of water chemistry and acidification under nonpoint sources of pollution across European Russia and West Siberia Environ. Res. Lett. 13 105007
[2] Berezina N A, Gubelit Y I, Polyak Y M, Sharov A N, Kudryavtseva V A, Lubimtsev V A, Petukhov V A, Shigaeva T D 2017 An integrated approach to the assessment of the eastern Gulf of Finland health: A case study of coastal habitats Journal of Marine Systems 171 159-71
[3] Epifanov A V, Shishkin A I, Antonov I V 2012 Rationing of Wastewater Discharge in Production of Pulp and its Products Using GIS Technology Pulp, Paper, Cardboard 1 66-73
[4] Vallius Henry 2015 Quality of the surface sediments of the northern coast of the Gulf of Finland, Baltic Sea Marine Pollution Bulletin 99 1–2 250-55
[5] Kushnerov A I, Antonov I V, Belomoev R P, Shishkin A I, Tihov S V 2019 Development of geoinformation database of hydroecological data on the basin of the north-eastern part of the finnish bay to support environmental protection Proc. XXIII Int. Bios-forum and Youth Bios-olympiad vol 2 ed A I Shishkin, A V Epifanov et al (St. Petersburg: Lubavich) p 86-93
[6] Zhang Jing, Zhang Jia, Du Xiangyang, Kang Hou and Qiao Minjuan 2017 An overview of ecological monitoring based on geographic information system (GIS) and remote sensing (RS) technology in China IOP Conf. Ser.: Earth Environ. Sci. 94 012056
[7] Zhilnikova N 2018 Geoinformation modelling system of natural technical complexes for simulation modelling and optimization of load distribution IOP Conf. Series: Materials Science and Engineering 450 062010
[8] Sannkov A V and Epifanov A V 2019 Development of a 3D geoinformation model of the Neva river for rationing anthropogenic load Proc. XXIII Int. Bios-forum and Youth Bios-olympiad vol 2 ed A I Shishkin, A V Epifanov et al (St. Petersburg: Lubavich) p 269-72