JEE Journal of Ecological Engineering

Journal of Ecological Engineering 2022, 23(3), 74–83
https://doi.org/10.12911/22998993/145470
ISSN 2299–8993, License CC-BY 4.0

Received: 2021.12.11
Accepted: 2022.01.13
Published: 2022.02.01

Geospatial Assessment of Regression Analysis Between the Hydrocarbon Content in Surface Waters and Snow Cover on the Example of the Territories of the Far North of Russia

Martynova Natalia[1*], Budarova Valentina[1], Kravchenko Victoria[1]

[1] Industrial University of Tyumen, 38 Volodarskogo St. Tyumen, 652000 Russia

* Corresponding author's e-mail: natali.cherdanceva@mail.ru

ABSTRACT

The article presents the generalized results obtained from the analysis of oil pollution of surface waters in the fields of the Far North. The research considered the administrative territorial division of the Russian Federation, the territory of the Khanty-Mansi Autonomous Okrug – Yugra (KhMAO). The results of the study performed on the basis of field data on sampling for the year were presented. The influence of the hydrocarbon content in surface waters and snow cover was assessed. The aim of the work was to consider the snow cover as a natural source of pollutants, affecting the accumulation in surface waters and snow cover. The results obtained can be used for subsequent observations of snow cover and surface waters. The data obtained can serve as a basis for planning further research and developing the solutions for environmental protection in the Far North. The analysis of the dependencies between the indicators of hydrocarbon pollution in surface waters and snow cover was carried out using the methods of correlation and parametric multivariate regression analysis. The methods of geoinformation analysis and GIS technologies were also used in the work. It was revealed that the problem of the state of snow cover and its role as an indicator of atmospheric and soil pollution require further research. On the one hand, the snow cover detains metals, and polluted soil areas are formed locally, on the other hand, after the snow melts, the pollutants remaining on the surface with surface runoff enter rivers and are carried by the wind for quite long distances.

Keywords: regression analysis, hydrocarbon, surface waters, snow cover, GIS mapping services, ecology.

INTRODUCTION

In matters of climate protection, the Russian Federation positions itself as a developing country facing a dual task: economic development and environmental protection; therefore, in the process of promoting its overall modernization program, it declares environmental protection one of its main national goals, considering sustainable development as an important strategy, and implements the measures to prevent as well as control pollution and environmental protection.

The Russian Federation takes an active part in the processes of the Climate agenda, an impetus has been given at the federal level, and discussions have been launched at the level of the Government of the Russian Federation, the State Duma and at the level of economic sectors. Today, most of the instruments of global climate policy have been implemented or are under development in Russia, such as carbon regulation, renewable energy incentives, green financing mechanisms, green certificates market, and ESG taxonomy. Within the framework of the climate conference in Paris 2015, the goal was set for the Russian Federation: to reduce the polluting emissions by 2030 to 70–75% of the 1990 level, provided that the absorption capacity of forests is taken into account as much as possible. The main objectives of environmental protection include urban air quality control, improvement of surface water quality, and reduction of total carbon dioxide emissions.

The Arctic, the Arctic zone of the Russian Federation, the Russian Arctic, the Far North are all territories undergoing intensive industrial development, especially within the framework of the

development of the Earth's subsurface, that is, the development of oil and oil and gas fields. The oil industry occupies one of the first places in terms of technogenic impact, material intensity, and labor intensity in the fuel and energy complex.

As a result of half a century of development of oil and gas-bearing territories, the natural environment of the Far North has undergone significant transformations and disturbances, the reduction of the consequences of which is unlikely to be expected in the near future. In the current situation, the preservation and restoration of natural resources, the prevention of negative man-made impacts and the elimination of their consequences are urgent tasks of the environmental policy of the Far North.

The surface waters of the Far North are experiencing a powerful anthropogenic load associated with the active development in recent decades of the infrastructure of cities and the largest oil and gas production complex in Russia. As a result of man-made impact on the water bodies of the Far North, the state of surface waters is characterized as unfavorable. Thus, the Ob River in the areas within the Far North belongs to the category of "dirty". The Irtysh River belongs to one of the most polluted water bodies that require priority implementation of environmental measures. Many rivers of the Far North belong to the categories of "very polluted" and "dirty". Contamination of water bodies occurs with nitrite nitrogen, ammonium nitrogen, petroleum products, iron, copper, zinc, and manganese compounds.

Therefore, the assessment of surface water quality, identification of pollution sources, its scale and dynamics are the basis for making the most important management decisions in the field of environmental management.

The aim of the work was to consider the snow cover as a natural source of hydrocarbon pollutants, affecting the accumulation in surface waters and snow cover. In order to achieve this goal, the following tasks were solved:

- an overview of the soil pollution and snow cover problem under the conditions of the far North was made according to the available literature data;
- selection and chemical-analytical studies of samples of snow cover and surface waters were carried out;
- based on authors' own research results, a characteristic of the state of snow cover and soil in the studied territories was given.

Geoinformation systems allow creating information in digital form (Klemmer, 2021), which can then be used for continuous monitoring of environmental problems. Therefore, the geoinformation system ArcGIS Pro was adopted as a tool for solving the problems in the study.

MATERIAL AND METHODS

Research area

Environmental pollution by oil and petroleum products is one of the most important issues on the global climate agenda.

In order to analyze the oil pollution of surface waters in the fields of the Far North, the territory of the Khanty-Mansiysk Autonomous Okrug – Yugra (KhMAO), which is an administrative-territorial unit of the Russian Federation, was selected.

Khanty-Mansiysk Autonomous Okrug – Yugra is one of the world leaders in the production of hydrocarbons (Budarova et al., 2017; Zhelonkina et al., 2021; Khodzhaeva, 2019). Oil pollution is especially dangerous for aquatic ecosystems.

The areas of large and unique deposits of the Middle Ob region that have been developed for a long time are characterized by a very high degree of technogenic load by various environmentally unsafe industries and transport systems. Extreme technogenic load is recorded in most of the "old" large oil fields of Nizhnevartovsk, Surgut and Nefteyugansk districts. The absence or poor development of the extractive industry and transport communications in a significant part of the western territory of the district and the east led to a relatively favorable environmental situation within their borders – most of the Berezovsky and Beloyarsk districts, the east of the Nizhnevartovsk district were practically not affected by industrial influence (Kurakova and Chalov, 2020; Bogdanov et al., 2020).

This study (carried out in the laboratories of the Industrial University of Tyumen) was conducted on the territory of the Khanty-Mansi Autonomous Okrug – Yugra, located in the middle part of Russia and the Eurasian continent (Figure 1). From west to east, the territory of the region stretches for 1400 km from the eastern slopes of the Northern Urals almost to the banks of the Yenisei; from north to south – 900 km from the Siberian Uvalas to the Kondinsky taiga. The entire territory of Ugra belongs to the regions of the Far North.

Figure 1. Geographical location of the research area

Sampling

In this study, the samples were taken from the territory of the KhMAO (Figure 2).

The study area was divided into a grid for the distribution of values with a size of 20 000 m. Next, a random method was use for selecting the units of each grid element, which makes it possible to analyze and cover the sample sites throughout the study area (the study was carried out in the field, funded by the Industrial University of Tyumen).

As initial data for spatial reference, the methods of satellite-geodetic determination of the coordinates of sampling sites were used. The coordinate system adopted is the Geographic Coordinate System (GCS).

Timing of sampling data for the year

The surface data involved the data for April-May (high water), and in snow cover for March-April.

Using the functionality of the ArcGIS Pro geoinformation system, a number of models for establishing the dependence and independence of data was obtained. Going through all these stages of creating a correct regression model, as a result, different variants of data representation in models (with and without transformed variables) were obtained, which allow studying different aspects of the data obtained and the possibility to analyze the surfaces of coefficients. As a research, this allows obtaining and expanding the necessary knowledge to model the investigated process and (Chabuk, 2021; Fischer, 2009; Boori, 2021).

Spatial structural patterns were analyzed with the initial data (Schabenberger, 2017; Ebdon, 1985; Mitchell, 2005).

The following methods of analysis were used:

1. The average nearest neighborhood.
2. Spatial autocorrelation
3. High/Low clustering
4. Least Squares Method (OLS)

The method is the average neighborhood

The method of point distribution analysis - nearest neighbor analysis is a generally accepted procedure for determining the distance from each point to its nearest neighbor and comparing this value with the average distance between neighbors. The average distance of the nearest neighbor gives a measure of the sparsity of points in the distribution. This is valuable in itself, since in some cases point objects can conflict if they are located too close to each other (Nath, 2021).

The method is based on an algorithm for calculating the distance between the center of each object in space and the location of the center of its nearest neighbor. Next, it is necessary to bring to the average values of the distances between neighboring objects.

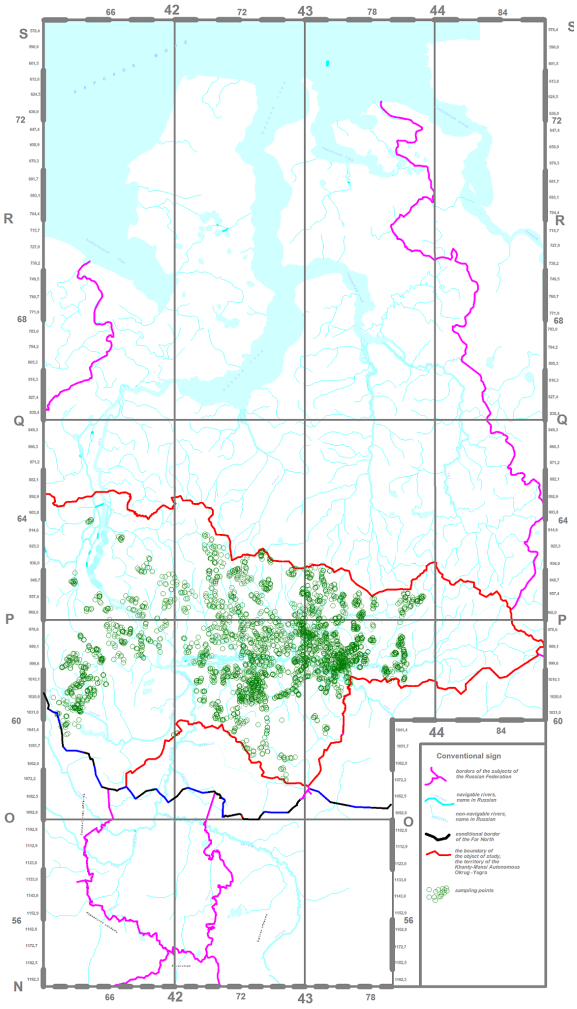


Figure 2. Geographical location of sampling points

On the basis of the results of the obtained values, it is possible to analyze the value of the obtained distance values with the theoretical data of a random distribution. In the software, the distances are either clustered based on the results or a dispersed distribution of objects is obtained.

The average nearest neighborhood is calculated using formula 1.

$$ANN = \frac{\sum_{i=1}^{n} d_i / n}{0.5 \Big/ \sqrt{\frac{n}{S}}} \quad (1)$$

where: d_i equals the distance between object i and its nearest neighbor, n corresponds to the total number of objects, S is the area of the minimum rectangle covering all objects.

The evaluation of the obtained data is calculated by the formula 2.

$$z = \frac{\sum_{i=1}^{n} d_i / n - 0.5 \Big/ \sqrt{\frac{n}{S}}}{M} \quad (2)$$

where: M is the RMS measurement error.

Spatial autocorrelation method

Spatial autocorrelation implements the basic principle of geography – close objects are more similar than distant ones.

The essence of the method is to measure spatial autocorrelation based simultaneously on the location of objects and their values. On the basis of the proposed set of objects and their associated attributes, an assessment is made whether there is clustering of objects or they are distributed scattered, or randomly.

On the basis of the analysis of the obtained half-dispersion graphs, it can be noticed that when the distance between the reference points is small, the half-dispersion is also small. This means that the measured values are close and, therefore, interrelated due to their spatial proximity. The half-dispersion increases along with the distance between the points, showing a rapid decline in the spatial correlation of values. Thus, the half-dispersion is a measure of the relationship of the measured values, depending on how close they are to each other.

High/Low clustering method

The High/Low clustering method is a multidimensional statistical procedure that collects the data containing the information about a sample of objects, and then arranges objects into relatively homogeneous groups.

The essence of the method is to divide the research area into cells and assign a degree of clustering of high or low values.

Least Squares Method (OLS)

Performs the Global Least Squares Method (OLS) for linear regression to predict or model a dependent variable based on its relationship with independent variables. It is one of the basic methods of regression analysis for estimating unknown parameters of regression models from sample data.

RESULT

The results of the study (the study was carried out in the laboratories of the Industrial University of Tyumen) provide the data obtained under the office conditions, by processing on laboratory equipment using the ArcGIS Pro software.

The Average Nearest Neighborhood method is used to calculate the range of distances to the number of neighboring objects. It returns the minimum, maximum, and average distances to the specified nth nearest neighbor (N is the input parameter) for a set of objects. As the tool works, messages are recorded.

Next, a step-by-step spatial autocorrelation was performed. It measures spatial autocorrelation for a series of distances and, if necessary, creates a line graph of these distances and the corresponding z-estimates. Z-scores reflect the intensity of spatial clustering; statistically significant and increasing peak z-scores indicate the distances at which spatial processes that provide spatial clustering are most pronounced. These peak distances often need to be used in tools with the Distance Range or Distance Radius parameter.

The first peak in surface waters is calculated at 110474.90 meters, whereas in snow cover – at 56388.62 meters. Results of step-by-step spatial autocorrelation shown in Figure 3 and 4.

Identification of the “hot spots”. Optimized analysis of “hot” points includes obtaining

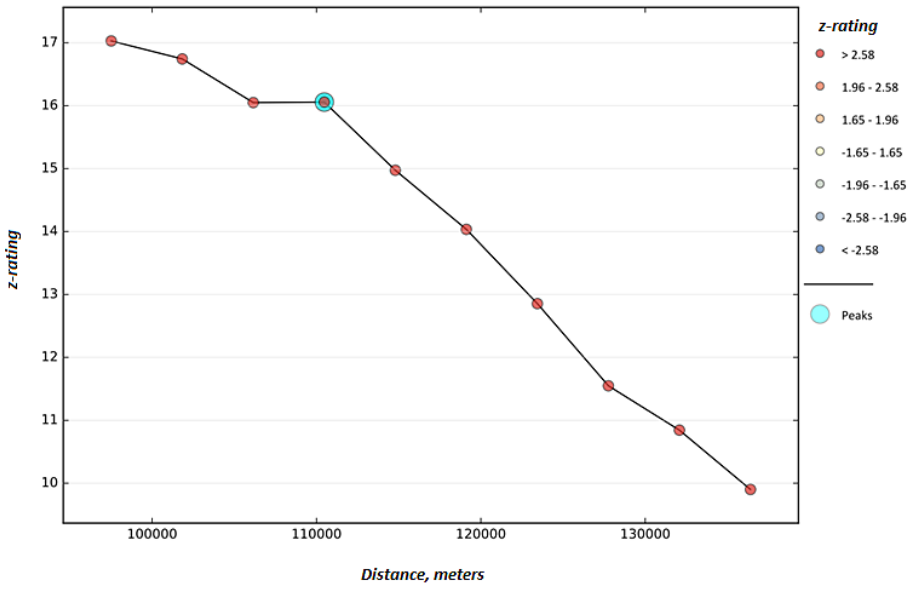


Figure 3. Spatial autocorrelation of surface waters

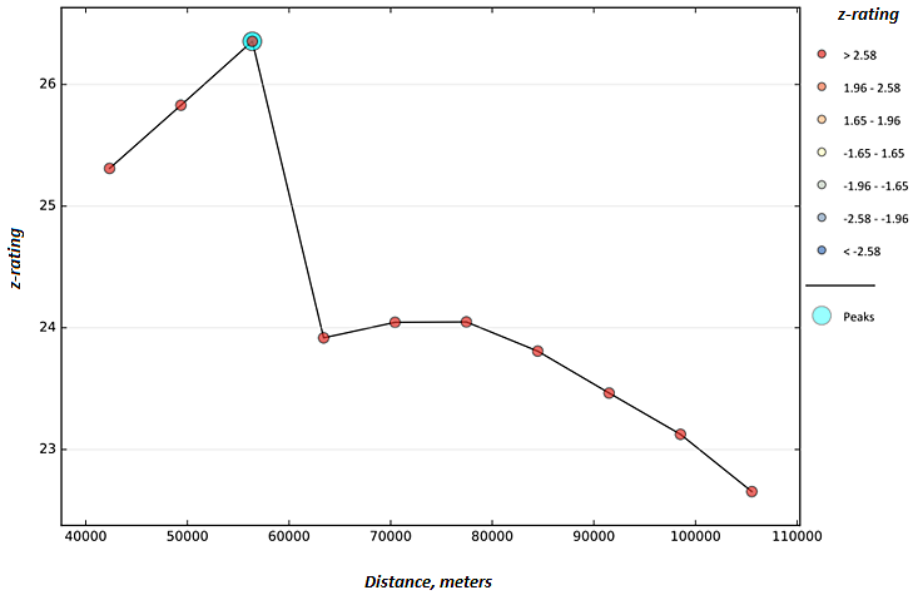


Figure 4. Spatial autocorrelation of bottom sediments

random points or objects with weights (points or polygons), creating a map of statistically significant "hot" points and "cold" points based on the Getis-Ord Gi statistical indicator in ArcGIS Pro. At the same time, the characteristics of the class of input objects are evaluated to obtain optimal results.

The optimal fixed distance band is based on the peak clustering found for surface waters is 34578.0 meters, for snow cover 29107.23 meters. The results of the neighborhood calculation are presented in Table 1.

In the space of the territory of the object of study, a grid was built, the edge of which is equal to 20,000 m. Further, the values for hydrocarbons in surface waters and snow cover were aggregated into the grid.

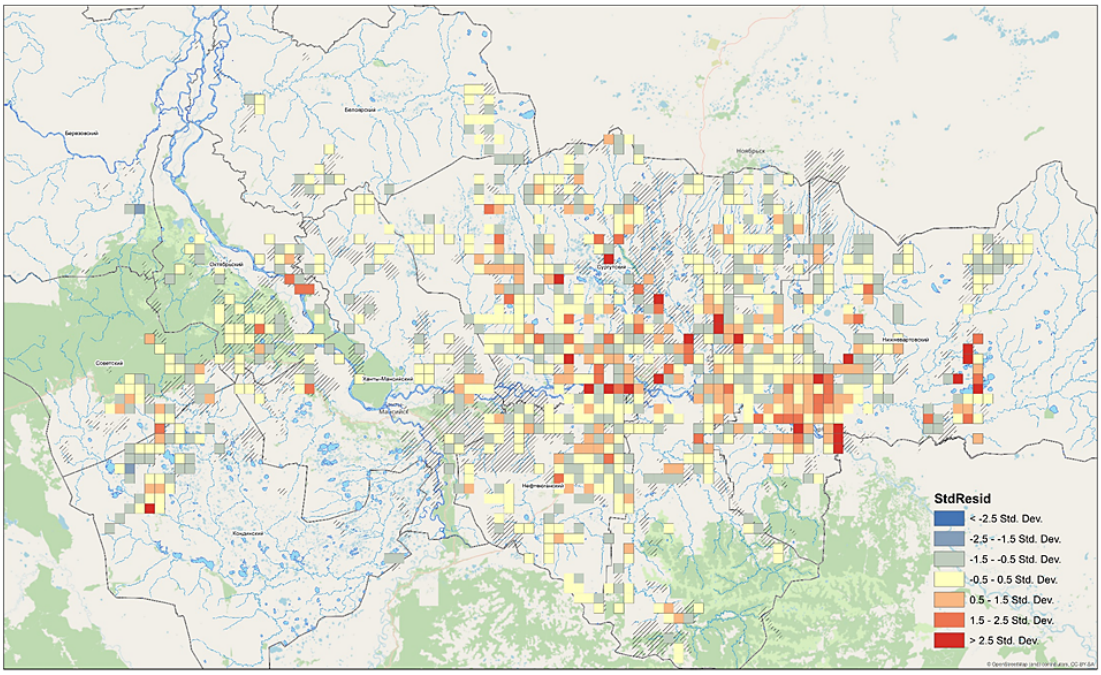
Using the Least squares method, a dependent variable was calculated for linear regression based on its relationship with independent variables. The independent variable is the concentration of hydrocarbons in surface waters, whereas the dependent variable is in the snow cover. Figure 5 shows the results of the least squares method.

In this case, the regression equation:

$$y = 0.124196x + 0.93644$$

Table 1. Threshold distances of surface waters and bottom sediments

Method for calculating steam distance	Threshold distance, m	
	Surface waters	Bottom sediments
Step-by-step spatial autocorrelation	110 474	56388.62
Optimized hotspot analysis	34 578	31 485
Calculate the range of distances to the number of neighboring objects	minimum 2616 average 15020 maximum 109885	minimum 8128 average 29107 maximum 146116



Summary of OLS results - Model Variables

Variable	Coefficient [a]	StdError	t-statistics	Probability [b]	Robust_SE	Robust_t	Robust_Pr [b]
Segment on the map	0,093644	0,004483	20,889250	0,000000*	0,004280	21,877020	0,000000*
Snow	0,124196	0,028591	4,343920	0,000019*	0,036728	3,381551	0,000766*

Diagnostics OLS

Input objects	Grid	Dependent variable:	Water
Number of observations	936	Akaike Information Criterion (Aic) [d]	-1633,820051
Multiple R-squared [d]	0,019803	Aligned R-squared [d]	0,018754
Connected F-statistics [e]	18,869645	Prob(>F), (1.934) degrees of freedom	0,000016*
Combined Wald statistics [e]	11,434890	Prob(>chi square), (1) degrees of freedom	0,000721*
Kenker Statistician (BP) [f]	13,540826	Prob(>chi square), (1) degrees of freedom	0,000233*
Jacques-Ber statistics [g]	3451,594087	Prob(>chi square), (1) degrees of freedom	0,000000*

Figure 5. Results of the least squares method for hydrocarbons in surface waters and snow cover

Six inspections were carried out:

1. The independent variable is statistically significant.
2. The coefficient of the independent variable is 0.124196, the sign of the relationship is positive.
3. The factor increasing the variance (VIF) in this case was not calculated.
4. Jacques-Ber statistics are statistically significant. The distribution of residuals has a positive asymmetry – the model is biased. The distribution of residuals is not normal, the model

is incorrect. The graph of residuals relative to the predicted dependent values of variables is structured (Figure 6, 7).

5. R2 is 0.018754, which indicates that the model is not correct.
6. No key independents found.

However, despite the failed checks, the Kenker test is statistically significant, so the investigated model can be improved by switching to geographically weighted regression. Within the framework of geographically weighted regression, a local form of linear regression used to model relations varying in space is obtained (Figure 8, 9).

The tool ArcGIS Pro revised the model characteristics obtained when running the least squares method tool and theoretically should have shown improved AICc and R-squared results. AICc decreased (it was -1,634, it became -1,745), which is a good indicator since a more

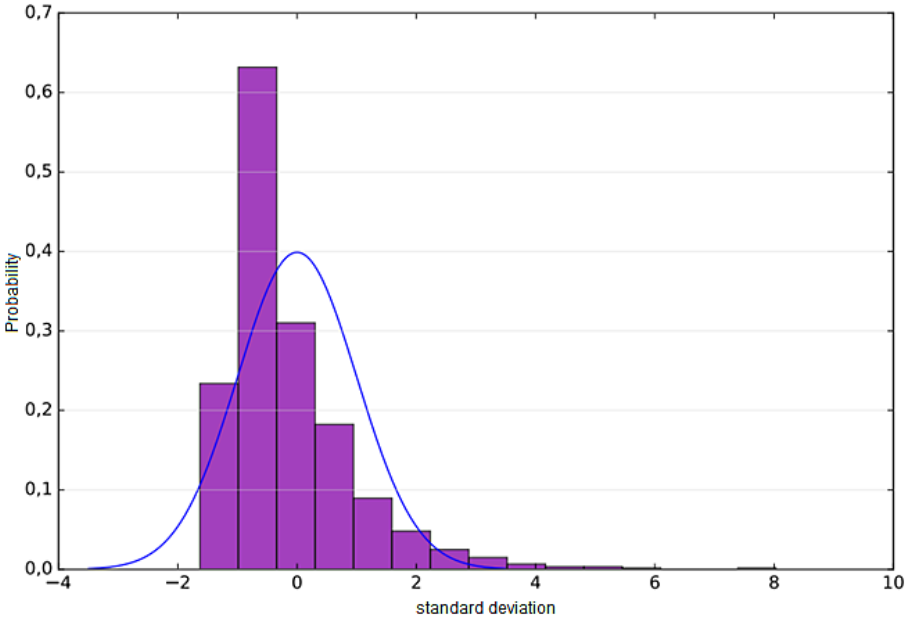


Figure 6. Histogram of standardized residuals

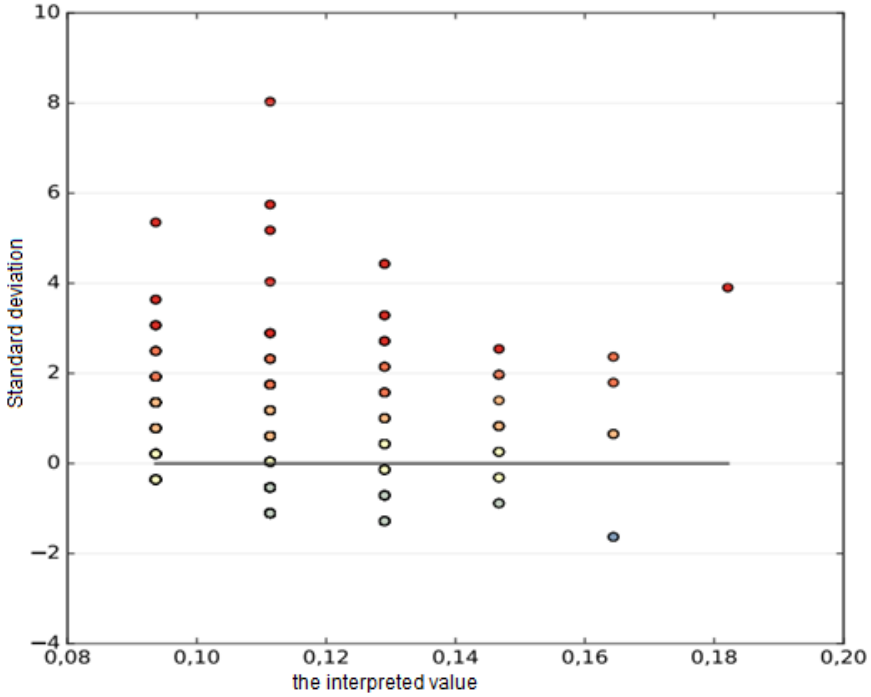


Figure 7. Graph of comparison of residuals with predicted values

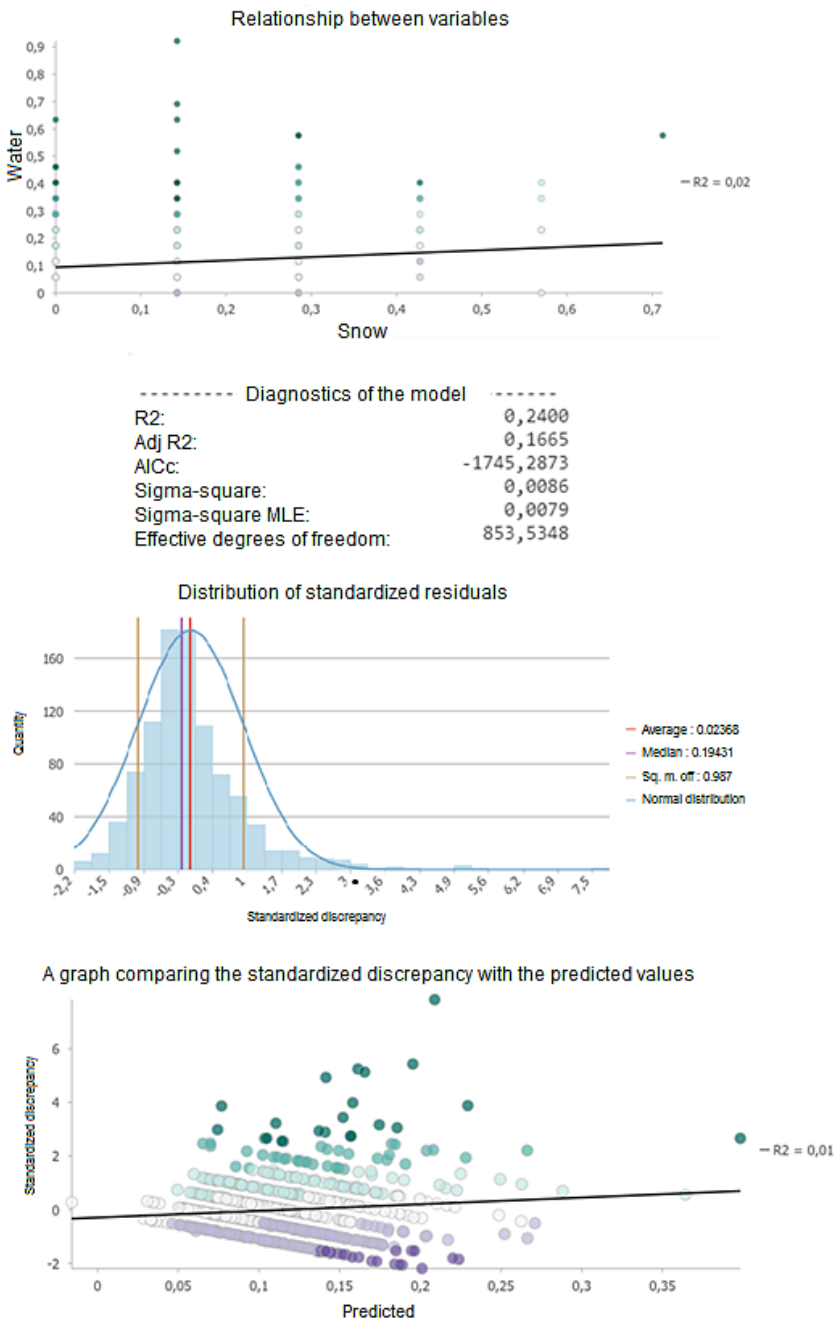


Figure 8. Geographically weighted regression

improved model should reduce the AICc value by more than 3 points. The R-squared value has increased (it was 0.02, it became 0.24), which means an improvement in the model, since now the 0.24 part of the dependent variable has been described by the model.

Attention should be paid to the R-squared indicator. On the Relationship graph, the distribution of all the residuals is seen, their form is structured, which means the model is incorrect. The relations themselves were tracked along the diagonal of the R-square line, where the maximum value corresponds to 0.40, and the minimum value is -0.03. However, for a reliable result, it is necessary that the R-squared indicator be equal to 0.5 or more, otherwise the model should not be trusted.

Next, using the Local bivariant relations method, two variables were analyzed for statistically significant relations using local entropy. Each object was classified into one of six categories based on the type of relationship. The output data can be used to visualize the areas where relationships between variables exist and to study changes in relationships within the study area (Figure 10).

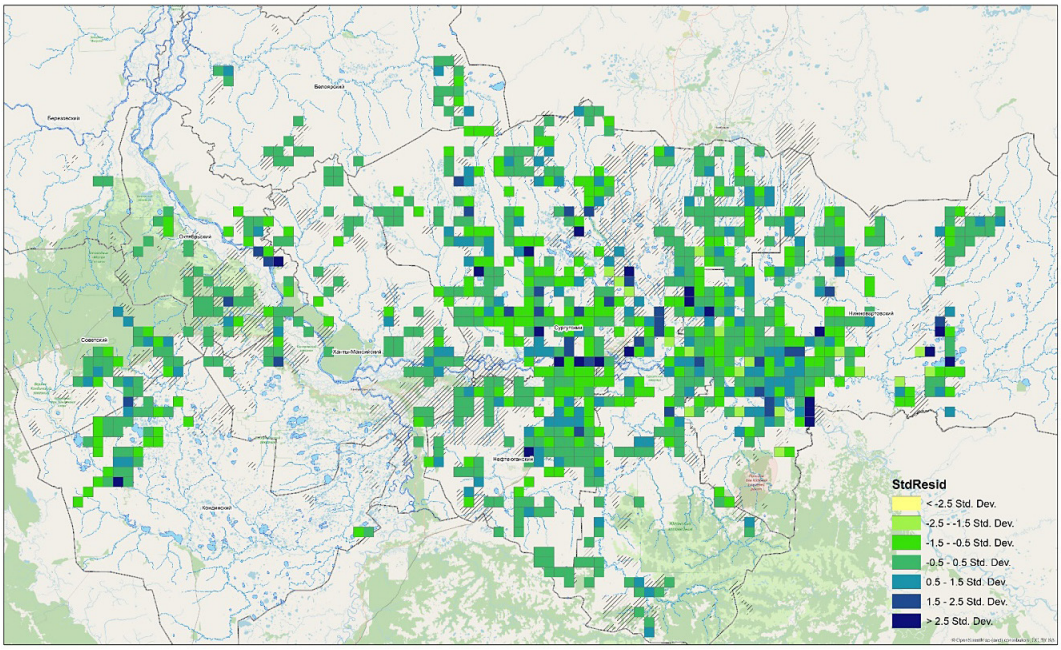
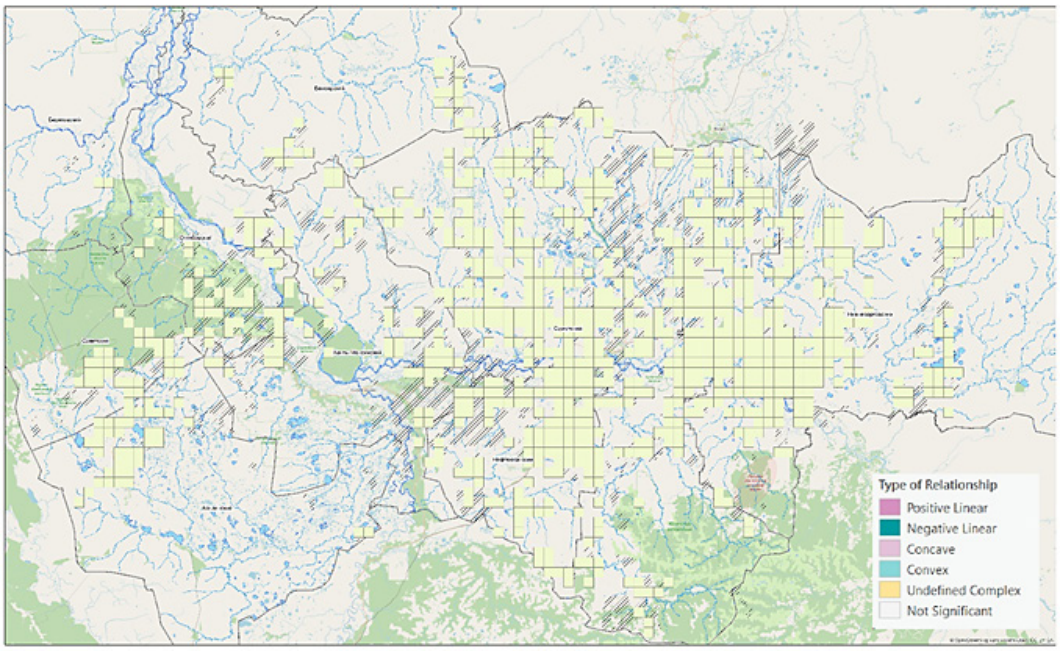


Figure 9. The final data model of geographically weighted regression



Category summary

Description	# of spatial objects	% of objects
Linear positive	0	0,00
Linear negative	0	0,00
Concave	0	0,00
Convex	0	0,00
Complex indefinite	0	0,00
Insignificant	936	100,00
Total	936	100,00

Figure 10. Visualization of the final data with a categorical summary

The tool showed that there are no statistically significant relationships when running with all possible variants of the number of neighbors, the number of permutations and confidence levels. Given the failed tests and the lack of reliability of the geographically weighted regression model, it is worth trusting the results of local bivariant relations and recognizing the model as statistically insignificant.

The dependent variable was the concentration of hydrocarbons in surface waters, and the independent concentration of hydrocarbon snow cover. All relationships between values are not significant.

CONCLUSIONS

The content of petroleum hydrocarbons in surface waters on the territory of KhMAO - Yugra, as a rule, does not exceed the maximum permissible values. This is significantly lower than 10–20 years ago, which indicates the effectiveness of the environmental policy carried out in the district. Regression analysis showed the presence of a statistically reliable dependence of the concentration of hydrocarbons in surface waters, as well as the concentration of hydrocarbon snow cover. As a confirmation, it can be highlighted that other studies have confirmed the data on the sources of pollution. Approximately half of the oil hydrocarbons in the fields of KhMAO - Yugra enter natural waters from man-made sources, whereas the remaining share is of natural origin.

The variety of sources of oil pollution makes it urgent to assess the impact of each of them on chemical runoff. The paper assesses the quality of water resources depends on hydrocarbon pollution.

Mathematical analysis of indicators of the composition of natural waters in deposits that differ in the intensity of man-made load provides ample opportunities for environmental assessment of large territories of the Far North.

REFERENCES

1. Bogdanov O., Shuvaev A., Abraeva T., Sabirianova R. 2020. Petroleum potential of the north-western part of the Khanty-mansiysk autonomous region KhmAO on the basis of petroleum system development history reconstruction. Paper presented at the Society of Petroleum Engineers - SPE Russian Petroleum Technology Conference 2019.
2. Boori M.S., Choudhary K., Paringer R., Kupriyanov A. 2022. Using RS/GIS for spatiotemporal ecological vulnerability analysis based on DPSIR framework in the republic of Tatarstan, Russia. Ecological Informatics, 67. DOI: 10.1016/j.ecoinf.2021.101490
3. Budarova V.A., Martynova N.G., Medvedeva Y.D., Budarov V.P. 2017. Geoinformation support at the facilities of the KHMAO - YUGRA oil and gas complex. In the collection: Oil and gas of Western Siberia. materials of the International Scientific and Technical Conference, 223–226.
4. Chabuk A., Al-Zubaidi H.A.M., Abdalkadhum A.J., Al-Ansari N., Ali Abed S., Al-Maliki A., Ewaid S. 2022. Application ArcGIS on modified-WQI method to evaluate water quality of the Euphrates river, Iraq, using physicochemical parameters. DOI: 10.1007/978-981-16-2380-6_58
5. Ebdon D. 1985. Statistics in Geography. Blackwell.
6. Fischer M., Getis A. 2009. Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications. DOI: 10.1007/978-3-642-03647-7.
7. Khodzhaeva G.K. 2019. Crude oil lines accident rate analysis in Nizhnevartovsk district KhMAO-ugra for years 2014–2018. Paper presented at the IOP Conference Series: Earth and Environmental Science, 381(1) DOI: 10.1088/1755-1315/381/1/012040
8. Klemmer K., Neill D.B. 2021. Auxiliary-task learning for geographic data with autoregressive embeddings. Paper presented at the GIS: Proceedings of the ACM International Symposium on Advances in Geographic Information Systems, 141–144. DOI: 10.1145/3474717.3483922
9. Kurakova A.A., Chalov R.S. 2020. Channel morphology and bank erosion in the lower reaches of the Ob River (within the Khmao-yugra autonomous district). Vestnik Moskovskogo Universiteta, Seriya 5: Geografiya, 2020(6), 41–50.
10. Mitchell A. 2005. The ESRI Guide to GIS Analysis. ESRI Press, 2.
11. Nath H., Rafizul I.M. 2022. Spatial variability of metal elements in soils of a waste disposal site in khulna: A geostatistical study. DOI: 10.1007/978-981-16-5547-0_3
12. Schabenberger O., Gotway C.A. 2017. Statistical methods for spatial data analysis. Statistical methods for spatial data analysis, 1–488. DOI: 10.1201/9781315275086
13. Zhelonkina E.E., Pafnutova E.G., Andreev A.A., Pafnutova I.D., Andreev K.A. 2021. Ecological assessment of wastewater on the environment by the example of khanty-mansiysk (KhMAO-Yugra). Paper presented at the IOP Conference Series: Earth and Environmental Science, 723(4). DOI: 10.1088/1755-1315/723/4/042031.