ATLANTIS PRESS
Advances in Intelligent Systems Research, volume 166
7th Scientific Conference on Information Technologies for Intelligent Decision Making Support (ITIDS 2019)

Modeling Spatial Structure of Thermokarst Lake Fields in Siberian Arctic by Satellite Imagery Based on Geo-Simulation and Heuristic Approaches

Yury Polishchuk
Yugra Research Institute of Information Technologies
Russia, Khanty-Mansiysk
yupolishchuk@gmail.com

Ildar Muratov
Yugra Research Institute of Information Technologies
Russia, Khanty-Mansiysk
ildarmur@gmail.com

Matvey Kupriyanov
Yugra Research Institute of Information Technologies
Russia, Khanty-Mansiysk
KupriyanovMA@spambox.ru

Vladimir Polishchuk
Tomsk Polytechnic University, Institute of Monitoring of Climatic and Ecological Systems SB RAS
Tomsk, Russia
liquid_metal@mail.ru

***Abstract*—On the basis of geo-simulation and heuristic approaches, a model of the spatial structure of thermokarst lakes fields has been developed, which allows to take into account the lognormal distribution of real lakes in size and tortuosity of their coastal boundaries. As a result, the model represents lakes of all sizes, including small lakes, which are considered as the most intense sources of methane emission into the atmosphere from thermokarst lakes. The log-normal size distribution is confirmed by experimental data based on the integration of data from images of medium and high spatial resolution in the northern territories of Western Siberia. Studies of the statistical properties of the tortuosity of the coastal boundaries of the lakes using high-resolution images of Kanopus-V showed that the distribution of the the degree of tortuosity corresponds to the normal law. The statistical characteristics of the degree of tortuosity of real lakes are determined. On the basis of the heuristic approach, a method of forming model objects of arbitrary shape with a given area and degree of tortuosity in the Cartesian coordinate system has been developed. An algorithm for modeling the spatial structure of the fields of thermokarst lakes is described.**

Keywords*—geo-simulation, heuristic modeling, remote sensing, thermokarst lakes, lake size-distribution, the tortuosity of coastal lakes, modeling algorithm, greenhouse gases, global warming

I. Introduction

The current global warming of the climate accelerates the degradation of permafrost. The permafrost is being a storehouse of canned carbon in the vast frozen peat bogs of Northern Eurasia, can become a cause of even more warming with the release of greenhouse gases. It will lead to the formation of new big challenges for the world community related to the violation of human-nature interaction. The development of measures to prevent an increase in the average annual temperature by more than 2 degrees by 2050 in accordance with the decisions of UN Climate Change Conference Paris 2015. These demand the formation of forecasts of the methane stock dynamics in the lakes of northern territories for the coming decades.

The dominant role in the accumulation of methane of small thermokarst lakes (with areas less than 0.01-0.05 ha) was established [1] in the permafrost zone of Western Siberia. However, the contribution of millions of such lakes to the global greenhouse effect due to small size has not been taken into account yet. Attempts to take them into account in estimating the total volume of world methane reserves in [2], based on the use of the theoretical power law of the lake size-distribution due to the lack of experimental data, raise great doubts, since the power law is not supported by experimental data [3]. This requires the development of methods and tools for modeling the dynamics of thermokarst lake fields. It will take into account a contribution of millions of small lakes to the total volume of methane reserves in a whole territory of Northern Eurasia.

According to Moiseev and Svirezhev [4], simulation modeling is one of the most important mathematical modeling types which may be used for constructing model of the thermokarst lakes fields with sufficient accuracy for current research. For modeling the spatial objects Polishchuk and Tokareva [5], Zhao and Murayama [6] have introduced geo-simulation modeling.

The spatial nature and high degree of complexity of the fields of thermokarst lakes as objects of modeling makes it necessary to use geo-simulation modeling [7]. The most important task is to develop a geo-simulation model of a field of thermokarst lakes, which is understood as a mathematical model that reproduces the spatial structure of fields of thermokarst lakes by simulating the shape, size and relative position of lakes in the study area taking into account the experimentally established statistical laws of their random location and size distribution. The development of a simplified model was considered in [8], which used experimental data on the properties of lakes in the permafrost zone of Western Siberia, obtained from images of an average resolution (30 m) of Landsat, on which small lakes are novisible. Therefore, for accounting to small lakes, you have to use high-resolution images.

In connection with the foregoing, the main goal of this work is a consideration of the issues of modeling the spatial structure of thermokarst lakes fields. It is based on the account of the tortuosity of the shape of the lake boundaries and the type of a law of the lakes distribution by their sizes, which is established on the basis of integrating medium and high resolution images that take into account lakes of all sizes, including small ones.

II. Statistical properties of the fields of thermokarst lakes in the cryolithozone of Western Siberia based on the space images

The informational basis for the experimental study of the properties of the fields of thermokarst lakes is the data of remote measurement of the areas of lakes from satellite images of the studied territory. The studies were carried out on the territory of all three permafrost zones (sporadic, discontinuous, continuous) in Western Siberia by remote method based on medium and high resolution satellite images taken in a relatively short period of time (2013 - 2015). All images were selected in a rather short period of the summer season (end of June - August) to minimize the effect of seasonal fluctuations of the lakes water level. During this period, there is no ice covering the lakes and preventing them from being detected under interpreting images process.

Since the medium resolution images of Landsat (30 m) provided a complete coverage of the study area, a mosaic of these images was used for research, which allowed studying the properties of hundreds of thousands of lakes. A study on high resolution images of Kanopus-V (2.1 m) was carried out on 78 test sites, which are situtated in the different zones of permafrost in Western Siberia are shown in [10].

Creating a geo-simulation model of thermokarst lakes fields requires knowledge of the basic properties of these fields, which can be obtained experimentally from satellite images. Analysis of the histograms of the coordinates (latitude and longitude) distribution of the lakes centers locations according to the data from the Landsat satellite showed [8] that the empirical histograms of such distributions correspond to the law of uniform density, according to the χ^2 criterion, with a probability of 95%, and the empirical size-distribution of lakes corresponds to exponential law by criterion χ^2, with a probability of 90%. A remote study of the shape of the boundaries of thermokarst lakes, conducted in [11,12], showed that in all zones of the West Siberian permafrost, the error in estimating lake areas when replacing the boundaries of real lakes with circles is relatively small (about 5% [11]). This means we can use a circle as a simplified model of a lake for geo-simulation modeling of thermokarst lakes fields. Consequently, the early geo-simulation model of the spatial structure of the fields of thermokarst lakes, previously developed in [8, 9], did not take into account the numerous small lakes and the tortuosity of the coastal boundaries of real lakes.

Therefore, the formation of a new geo-simulation model assumes the representation of the thermokarst lakes field in the form of random sets of flat figures with tortuous boundaries. To determine the properties of such figures, additional remote sensing studies are required to obtain experimental data on the tortuosity characteristics of the lake shorelines and the size-distribution of lakes in a wide range of sizes, including small lakes.

One of the important tasks is to build histograms of the size distribution of lakes, which would take into account all the lakes in a wide range of sizes - from tens of meters to tens of kilometers. To construct such a histogram in [13-15] it was proposed to choose partial intervals of the histogram with an irregular step (according to a logarithmic law), namely: 20-50 m 2, 50-100 m 2, 100-200 m 2, etc. up to 200 km 2. This allowed a rather compact presentation of data on the distribution of lakes at intervals of their sizes in a very wide range of changes in the areas of lakes.

Such histograms of the distribution of lake areas can be constructed only on the basis of the integration of data on the areas of water bodies obtained from both medium and high resolution of satellite images. The developed methodology for combining (synthesizing) data on the areas of lakes obtained from images of different spatial resolution in order to construct synthesized histograms of the distribution of areas of lakes in a very wide range of their sizes is described in [14,16]. In accordance with this methodology, the synthesized histogram of the distribution of lakes by area was obtained by “stitching” two initial histograms, the first of which is based on Landsat-8 data and represents large lakes (ranging in size from 0.5 to 20,000 hectares). The second initial histogram obtained from the data of the Kanopus-V images for 78 test sites in all three permafrost zones represents small lakes (from 0.005 to 20 ha). Another synthesized histogram of the total area distribution of lakes in the studied territory was constructed in the same way.

The synthesized histograms of the distribution of the number and total area of lakes, obtained in the manner described above according to the results of the studies conducted in the permafrost zone of Western Siberia, are presented in [10]. The determination of the type of the law of the lake size-distribution was carried out on the basis of an approximation of the obtained sinthesized histogram of the distribution of lakes, which showed [16,17] that the empirical distribution corresponds to a lognormal law. The conformity check was performed with help of the Excel software package using Pearson's criterion confirmed that the histogram of the lake size-distribution obtained in a wide range of their sizes corresponds to a lognormal law with a high probability of 0.99.

The analysis of another synthesized histogram of the distribution of the total lake area by lake sizes allows us to experimentally substantiate the applicability of the early geo-simulation model of the spatial structure of thermokarst lakes fields previously developed in [8] with the exponential size-distribution of lakes based on Landsat medium resolution images. Thus, the empirical distribution of the total areas of lakes by their sizes [10] shows that the bulk of the total area of lakes (about 80%) is made up of lakes with sizes ranging from 2 to 500 hectares. The approximation of the truncated (in the size range of 2–500 ha) synthesized histogram of the size-distribution of lakes showed with a high level of the coefficient of determination (R^2 = 0.72) its conformity to the exponential law.

An experimental study of the tortuosity of the coastlines of real lakes was carried out. According to satellite measurements, the values of the shape index (degree of the boundaries tortuosity) of the lakes were calculated by the formula [18]:

$$z = \frac{P^2}{4\pi s},$$

where *z* is the degree of tortuosity; *p* is the perimeter of the coastal boundary; *s* is the lake area.

The determination of the *z* value was carried out according to data from images of Kanopus-V on all 78 test sites. It is established that the distribution of the magnitude of the degree of tortuosity *z* corresponds to the normal distribution law. The mean values of z take the following values in different zones of permafrost: sporadic – 3.5; discontinuous – 3.38; continuous – 2.95.

III. Geo-simulation model of spatial structure of thermokarst lake fields

Based on the above, we can formulate the following fundamental principles that determine the essential properties of the model of the spatial structure of the thermokarst lakes fields:

1. The shape of the coastal boundaries of the lakes can be represented by flat figures with tortuous boundaries, which are described by the following characteristics: coordinates of the centers x_i, y_i; degree of tortuosity z_i; area s_i (i is the ordinal number of the lake).

2. The coordinates of the centers of flat figures are random variables, the distribution of which is determined by the law of uniform density.

3. The degree of tortuosity of the boundaries of flat figures is a random variable, the distribution of which is determined by a normal law.

4. The area of a flat figure with tortuous boundaries is a random variable which distributed by a lognormal law.

5. Spatial changes of the coordinates of the flat figures centers, the tortuosity degree of borders and their areas are statistically independent.

The geo-simulation model of the fields of thermokarst lakes is developed in accordance with these principles (1-5). The geometric interpretation of the proposed new geo-simulation model of thermokarst lakes fields is shown in Figure 1, which represents the fragment of the model of the spatial structure of the thermokarst lakes fields. Here is indicated: s_i - the area of the i-th model lake and i - the lake number.

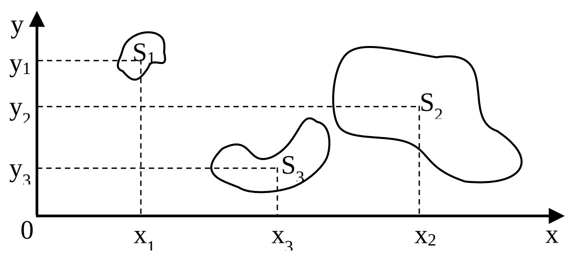

Fig. 1. Model representation of the fragment of the thermokarst lake field

To create flat figures using a heuristic approach, a technique of forming model objects has been developed. This technique allows generating objects of arbitrary shape with a given area and degree of tortuosity in the Cartesian coordinate system. The technique is performed as a sequence of the following steps:

1. Creating a polygon with a random number of vertices.

2. Connecting the vertices of the polygon with a continuous line using Bezier curves [19] allows us to create a model object with a random tortuosity of the contour boundary and a unit area.

3. Transformation of the model object to its original form with a given area while maintaining the initial value z.

IV. Algorithmic questions of modeling the thermokarst lake fields

Modeling the thermokarst lakes fields is based on the geo-simulation approach. Four random numbers are used, namely: representing the coordinates of the centers of flat figures, the degree of tortuosity of the boundaries and the size of their area and formed using the Monte Carlo method. In the general case, the mutual probability density of random coordinates of centers, degree of tortuosity, and areas of flat figures simulating lakes in a mathematical model of random fields of thermokarst lakes can be represented as:

$$f(x, y, z, s), \qquad (1)$$

where *x* and *y* – coordinates of flat figure center in a model; *z* – the degree of tortuosity of the flat figure boundaries; *s* – area of the flat figure imitating a lake.

Consequently, the set of flat figures in the model of lake fields will be represented as a sequence of four random variables. To develop an algorithm for modeling the fields of thermokarst lakes, it is necessary to take into account the form of the laws of distribution of x, y, z, s and statistical dependencies between changes in these quantities, which, according to experimental data [8, 12], can be taken statistically independent. With taking into account this to simulate the fields of thermokarst lakes the joint probability density (1) can be represented as:

$$f(x, y, z, s) = f(x) \times f(y) \times f(z) \times f(s), \qquad (2)$$

where *f(x)* and *f(y)* are the probability densities of the uniform distribution; *f(z)* is the probability density of the normal distribution; *f(s)* is the probability density of the lognormal distribution of the areas of flat figures, defined, according to [20], by the equation:

$$f(s) = \frac{1}{\sqrt{2\pi}s\sigma} exp\left(-\frac{(\ln s - \mu)^2}{2\sigma^2}\right), \qquad (3)$$

where μ is the mathematical expectation, σ is the standard deviation.

The generation of a sequence of random numbers determines the characteristics of the location of the circles centers with taking into account equation (2). It is carried out using a pseudo-random number generator which produces variables are distributed according to the law of uniform density. To simulate lakes with random sizes, the areas of which are distributed according to the lognormal law (3), sequences of pseudo-random numbers are generated and they correspond the lognormal distribution law, in accordance with the equation obtained in [21]:

$$s_i = \exp(\mu + \sigma \times r), \quad (4)$$

where r is a pseudo-random number distributed according to the normal law, calculated by the formula:

$$r = \sum_{j=1}^{12} q_j - 6, \quad (5)$$

where q_j is a random variable uniformly distributed on the interval [0,1].

Therefore, it is necessary to use under process developing of the geo-simulation system, both the software pseudo-random number generator producing the random variables distributed in accordance with the law of uniform density and the software pseudo-random number sequence generator producing the random variables distributed in accordance with the lognormal size distribution of lakes.

V. Conclusion

The article presents the approach to modeling the spatial structure of the thermokarst lakes fields based on a geo-simulation model as a set of random flat figures. The characteristics of them are the distribution of centers coordinates corresponds to the uniform distribution law, the distribution of the tortuosity degree corresponds to the normal distribution law and the distribution of areas corresponds to the lognormal distribution law. Experimental substantiation of the lognormal distribution of lakes by sizes is given on the basis of research results on the empirical distribution of thermokarst lakes in a very wide range of their sizes in the permafrost zone of Western Siberia based on the integration of satellite images with different spatial resolution. The results of checking the conformity of the area distribution law with the empirical histogram data showed that the lognormal distribution law corresponds to the experimental data, according to the Pearson criterion, at a significance level of 0.99. The procedure of modeling of the thermokarst lakes fields is briefly described. Each model lake is characterized by a four of numbers: the coordinates of the center, the tortuosity degree and the area of the lake. Statistical characteristics of the model lakes fields were obtained using images of medium and high spatial resolution.

The above analysis of the synthesized histogram of the distribution of the lakes total area by sizes showed that lakes are from 2 to 500 hectares make up to 80% from whole number of lakes in the studied area. This fact makes it possible to experimentally substantiate the applicability of the early geo-simulation model of the spatial structure of thermokarst lakes fields. This model uses exponential distribution law to describe a statistical characteristic of how thermokarst lakes distribute by their sizes and it based on medium-resolution images of Landsat [8]. Indeed, the approximation of a truncated (in the size range of 2–500 ha) synthesized histogram of the lakes distribution by their sizes corresponds to the exponential distribution law with a high level of the determination coefficient ($R^2 = 0.72$). This model of thermokarst lakes fields can be used for simplified estimation of water reserves accumulated in thermokarst lakes are located in the Arctic zone of Russia, or to study and predict the dynamics of thermokarst lakes under climatic changes and other tasks.

However, this early model does not take into account small lakes, which are considered intensive sources of methane emissions. These small lakes are not detected in the medium-resolution Landsat images and, therefore, are not take into account in the model uses exponential distribution law for lakes modeling. Consequently, under process modeling of thermokarst lakes fields in the permafrost zone it requires taking into account small lakes, using the lognormal distribution law. This based on the use of satellite images of high resolution in combination with the medium resolution images.

The results of work can be used to obtain predictions of the methane emissions dynamics from the thermokarst lakes are located in the Arctic zone of Northern Eurasia, for the coming decades, in the context of UN Climate Change Conference Paris.

Acknowledgment

The study was conducted with the financial support of the Department for Education and Youth Policy of the Khanty-Mansiysk Autonomous Okrug, the Government of Tomsk Oblast and, in partly, grants from the Russian Foundation for Basic Research, projects No. 18-47-700001, No. 18-45-860002, No. 18-45-703001 and No. 19-07-00282.

References

[1] O. S. Pokrovsky, L. S. Shirokova, S. N. Kirpotin, S. Audry, J. Viers, B. Dupre, "Effect of permafrost thawing on the organic carbon and metal speciation in thermokarst lakes of Western Siberia," Biogeosciences, vol. 8, 2011, pp. 565–583.

[2] M. A. Holgerson and P. A. Raymond, "Large contribution to inland water CO2 and CH4 emissions from very small ponds," Nature Geoscience Letters, vol. 9, 2016, pp. 222-226.

[3] B. B. Cael and D. A. Seekell, "The size-distribution of Earth's lakes," Sci. Rep. 29633, 2016.

[4] N. N. Moiseev and Y. M. Svirezhev, "System analysis of dynamic processes of the biosphere," Bulletin of the Academy of Sciences of USSR, vol. 2, 1979, pp. 47-54 (in Russian).

[5] Y. M. Polishchuk and O. S. Tokareva, "Geo-simulation modeling of air pollution zones as a result of the gas burning in oil fields," Information Systems and Technology, vol. 2, 2010, pp. 39-46 (in Russian).

[6] Y. Zhao and Y. Murayama, "Urban dynamics analysis using spatial metrics geo-simulation," Spatial analysis and modeling in geographical transformation process eds. Y. Murayama and R. Thapa (Dordrecht-Heidelberg-New-York-London: Springer, GeoJournal Library), chapter 10, 2011, pp. 153–167.

[7] Y. M. Polishchuk and V. Y. Polishchuk, "Using geo-simulation for predicting changes in the sizes of thermokarst lakes in northwestern Siberia," Kriosfera Zemli, vol. 20, 2016, pp. 29-36.

[8] V. Y. Polishchuk and Y. M. Polishchuk "Modeling of thermokarst lake dynamics in West-Siberian permafrost," Permafrost: Distribution, Composition and Impacts on Infrastructure and Ecosystems ed. O. Pokrovsky (New York: Nova Science Publishers), chapter 6, 2014, pp. 205-234.

[9] Y. M. Polishchuk and V. Y. Polishchuk, "Forecast of thermokarst lakes dynamics in permafrost based on geo-simulation modeling and remote sensing data", Proc. of Conf. "Mathematical and Information Technologies MIT-2016", ed, Y. Shokin and H Miloshevich et al (August 28 – September 5 2016, Vrnjacka Banja, Serbia – Budva, Montenegro): CEUR-Workshop Proceedings, vol. 1839, 2017, pp. 393-405.

[10] Y. M. Polishchuk, A. N. Bogdanov, I. N. Muratov, V. Y. Polishchuk, A. Lim, R. M. Manasypov, L. S. Shirokova and O. S. Pokrovsky, "Minor contribution of small thaw ponds to the pools of carbon and methane in the inland waters of the permafrost - affected part of the Western Siberian lowland," Environmental Research Letters, vol.13, 2018, pp. 1-16.

[11] V. Y. Polishchuk and Y. M. Polishchuk, "Remote studies of variability of the shape of coastal boundaries of thermokarst lakes in the permafrost of West Siberia," Study of Earth from space, vol. 1, 2012, pp. 61-64 (in Russian).

[12] V. Y. Polishchuk and Y. M., Polishchuk, "Geo-simulation modeling fields of thermokarst lakes in zone of permafrost," Khanty-Mansiysk: Yugra State University Press, 2013, p. 129, (in Russian).

[13] Y. M. Polishchuk, A. N. Bogdanov, V. Y. Polishchuk, R. M. Manasypov, L. S. Shirokova, S. N. Kirpotin and O. S. Pokrovsky, "Size distribution, surface coverage, water, carbon, and metal storage of thermokarst lakes in the permafrost zone of the Western Siberia lowland," Water, vol. 9, 2017, pp. 1-18.

[14] Y. M. Polishchuk, A. N. Bogdanov and I. N. Muratov, "Methodological issues of building generalized histograms of the distribution of lake areas in the permafrost zone based on medium and high resolution satellite images," Current problems in remote sensing of the Earth from space, vol. 13, 2016, pp. 224–232 (in Russian).

[15] Y. M. Polishchuk, A. N. Bogdanov, N. A. Bryksina, I. N. Muratov, V. Y. Polishchuk, "Integration of space images of ultra high and medium resolution for the construction of size-distribution histograms of thermokarst lakes in the extended range of their sizes," Current problems in remote sensing of the Earth from space, vol. 15, 2018, pp. 38-46 (in Russian).

[16] Y. M. Polishchuk, A. N. Bogdanov, N. A. Bryksina, V. Y. Polishchuk, I. N. Muratov, M. A. Kupriyanov, O. A. Baysalyamova and V. P. Dneprovsky, "Experience and results of remote sensing lakes in cryolithozone of Western Siberian by satellite images of different resolution for 50-year period," Current problems of remote sensing of Earth from space, vol. 14, 2017, pp. 42-55 (in Russian).

[17] Y. M. Polishchuk, A. N. Bogdanov, V. Y. Polishchuk, "Lake size-distribution in cryolithozone based on satellite images of medium and high resolution," Bulletin of the Tomsk Polytechnic University. Geo Assets Engineering, vol. 329, 2018, pp.16-25 (in Russian).

[18] N. Kremer, "Theory of probability and mathematical statistics: a textbook for high schools," Moscow: UNITY-DANA, 2003, p. 573 (in Russian).

[19] P. N. Dubner, "Handbook of statistical distributions", 2000, http://algolist.manual.ru/maths/matstat/logNormal/index.php (in Russian).

[20] X. A. Han, Y. C. Ma, X. L. Huang "The Cubic Trigonometric Bézier Curve with Two Shape Parameters," Applied Mathematics Letters, , Vol. 22, 2009, pp. 226-231.

[21] A.M. Berlyant, "A cartographic method of research," Moscow: Moscow University Press, 1978, p. 257.