ATLANTIS PRESS
Advances in Intelligent Systems Research, volume 174
Proceedings of the 8th Scientific Conference on Information Technologies for Intelligent Decision Making Support (ITIDS 2020)

The System of Geo-Simulation Modeling of Thermokarst Lakes Fields Based on the Log-Normal Distribution of Their Sizes

Vladimir Yu. Polishchuk*
Laboratory of geoinformation technologies,
Institute of Monitoring of Climatic and Ecological Systems
Tomsk, Russia
vy_polishchuk@hotmail.com

Yury M. Polishchuk
Center for space services,
Ugra Research Institute of Information Technologies
Khanty-Mansiysk, Russia
yupolishchuk@gmail.com

***Abstract*—The issues of modeling the fields of thermokarst lakes by the geo-simulation approach are considered. The permafrost zone is a zone of interest. The features of obtaining experimental data on the properties of real fields of thermokarst lakes from satellite images in the Arctic zone of Western Siberia are briefly described. Data on the distribution of random coordinates of the lakes centers and the distribution of lakes by their size are presented. Algorithmic issues of the geo-simulation modeling of lake fields are considered. The uniform distribution of the coordinates of the lakes centers is taking into account. It should be noted lakes sizes have lognormal distribution. The structure of the software package for modeling fields of thermokarst lakes is described. This is a set of software modules that provide input of model parameters, formation of pseudorandom numerical sequences and output of simulation results. The software package is implemented in Python 3.8. It allows displaying the results of modeling lake fields using modern geographic information systems. A fragment of the model field of thermokarst lakes is presented. It was obtained as a result of a computational experiment with the software package.**

Keywords*—*geo-simulation, software package, permafrost, thermokarst lakes, global warming, lognormal distribution law

I. Introduction

One of the main problems of our time has become the warming of the climate, creating global challenges to the world community in recent decades. The World Climate Summit held in connection with this (Paris, 2015) adopted recommendations to all countries of the world on the development of measures capable of preventing an increase in the average annual temperature of the Earth by more than 2°C until 2050. However, the development of measures to prevent a temperature increase in Russia, more than 60% of which is located in the permafrost zone, is impossible without the formation of well-grounded forecasts of changes in methane and carbon dioxide emissions, the development of which is necessary knowledge of the dynamics of fields of thermokarst lakes in permafrost. In this regard, the relevance of developing a geographic information technology for assessing and predicting the dynamics of fields of thermokarst lakes in permafrost zones is not in doubt.

The modeling issues of thermokarst lakes were considered in the works of West and Plow, Subin and Riley, Duguet, Hostetler, etc. [1]. Such models are convenient for studying processes in solitary lakes. However, they are not suitable for studying the spatiotemporal changes in the fields of thermokarst lakes. From the point of view of modeling the fields of thermokarst lakes on the territory of interest are methods of mathematical landscape morphology [2, 3]. Such methods are focused on the use of analytical models for modeling the dynamics of thermokarst territories. Methods and models of landscape mathematical morphology can be used to assess geoecological risks in the development of thermokarst processes in the permafrost areas. However, these are also not intended for the study of spatiotemporal changes in the fields of thermokarst lakes in a changing climate. In this regard, it becomes necessary to develop methods and tools for mathematical modeling of the dynamics of thermokarst processes for predicting changes of thermokarst lakes fields by means of computer experiments. Also we should to taking into account climatic changes.

Therefore, the use of simulation models [4] is CONSIDERED promising, allowing us to study the dynamics of the thermokarst lakes fields under modern global warming. Last years, within the framework of the methodology of simulation modeling, one of the new directions of computer modeling has been formed - geo-simulation modeling [5, 6]. This should be understanding as simulation of complex objects with a spatial structure by using methods and means of geoinformatics [7]. In [8], a geo-simulation model of the field of thermokarst lakes in the form of a set of random circles was proposed. The properties of this take into account the main statistical characteristics of real thermokarst lakes fields, determined from the experimental data of satellite measurements. In particular, in these studies it was found that the size distribution of lakes obeys an exponential law, which was taken into account when developing this model.

However, the model developed in [8, 9] was based on the results of remote studies of the properties of fields of thermokarst lakes from satellite images of medium resolution (30 m) Landsat, on which lakes of relatively large sizes (above 0.5 ha) are visible. In our more recent investigations of thermokarst lakes [10–13] using images of high (1–10 m) and ultrahigh resolution (less than 1 m), the presence in real fields of thermokarst lakes in permafrost (on the example of Western Siberia) of huge the number of small-sized lakes that are not found in medium resolution images. The methodology for integrating data from medium and high / ultra-high resolution images developed in [10, 11] made it possible to construct synthesized histograms of the distribution of lakes by area over a very wide range of sizes (from units m2 to tens of km2). An analysis of these histograms showed [12, 13] that the distribution of real lakes in the permafrost zone does not satisfy the exponential, but the lognormal law.

In connection with the above, it became necessary to modify the model and improve the system of geo-simulation modeling of fields of thermokarst lakes, aimed at taking into account the lognormal law of their size distribution. However, the issues of the program-algorithmic implementation of the system of geo-simulation modeling of thermokarst lakes fields taking into account the lognormal law are currently not considered enough, which determined the purpose of this work.

II. Experimental Data on Lakes Distribution Laws

The information basis for the experimental study of the properties of fields of thermokarst lakes is the data of remote measurement of the areas and perimeters of lakes from satellite images of the studied territory. The studies were conducted on the territory of all three permafrost zones (sporadic, discontinuous, and continuous) in Western Siberia based on satellite images of medium and high resolution. All images were taken in a rather short period of the summer season (end of June - August) in order to minimize the influence of seasonal fluctuations in the water level in the lakes. During this period, the ice cover on the lakes completely disappears, which does not allow them to be distinguished when decrypting the images.

Creating a geo-simulation model of the thermokarst lakes fields requires knowledge of the basic properties of these fields, which can be obtained experimentally from satellite images. An analysis of histograms of the distribution of the coordinates of the location of the centers of the lakes according to the data from the Landsat satellite showed [8,9] that the empirical histograms of the distribution of the coordinates of the lakes centers correspond, according to the χ^2 criterion, to the law of uniform density with a probability of 95% [14].

One of important problems for creating a geo-simulation model of the thermokarst lakes fields is the constructing histograms of the distribution of lakes by size. It would be taking into account all lakes in a wide range of sizes - from tens of meters to tens of kilometers. The issues of constructing such a histogram were considered in [11, 12], where it was proposed to select partial intervals of the histogram with irregular steps (according to the logarithmic law), namely: 20-50 m^2, 50-100 m^2, 100-200 m^2 and etc. up to 200 km^2. This made it possible to fairly compactly present data on the distribution of lakes in the intervals of their sizes in a very wide range of changes in the area of lakes.

We formed a synthesized histogram of the lakes distribution over areas in a very wide range of there sizes. It should be noted this was formed by using the methodology developed in [11] for integrating data on the number and area of water bodies obtained from satellite images of both medium and high resolution based on the combination (synthesis) of lakes data such as their number and sizes. In accordance with this technique, the synthesized histogram of the distribution of lakes by sizes was obtained by "stitching" two initial histograms, the first of which is based on data from medium resolution images (30 m) of Landsat-8 and represents large lakes (size from 0.5 to 20,000 ha). The second baseline histogram, obtained from data from high resolution images (2 m) of Canopus-B in 78 test plots in the permafrost zone, includes small lakes. A synthesized histogram of the size distribution of lakes, obtained above in the manner described above based on the results of studies in the permafrost zone of Western Siberia, is presented in [13].

According to [14], the probability density for the lognormal law of the distribution of the area of lakes (s) is determined by the equation:

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

where μ is the mean, σ is the standard deviation.

Estimates of the parameters μ and σ of the theoretical log-normal law of the size distribution of lakes obtained in [13] from sample data for the permafrost zone of Western Siberia take the values of mathematical expectation and variance 6.88 and 3.24, respectively. A verification of the correspondence between empirical and theoretical distributions using Pearson criterion [13] showed that the calculated χ^2 value is 0.23. Therefore, the synthesized histogram of the distribution of lake areas in a wide range of their sizes can be accepted corresponding to a lognormal law with a high probability of 0.99.

A correlation analysis of a multidimensional array of experimental data on changes in the coordinates of the lake centers and the areas of thermokarst lakes obtained from satellite images was carried out. The analysis showed that the cross-correlation coefficients between all the above indicators are quite small. This means [9] the statistical independence of the above indicators of the thermokarst lakes fields in the studied permafrost territory of Western Siberia.

III. Modeling Algorithms

The implementation of a mathematical model of the thermokarst lakes fields involves the creation of a generator of pseudo-random numerical sequences of triples, the first two of which are distributed according to a uniform law, and the third according to a log-normal law. In general, the mathematical model of fields of thermokarst lakes can be represented as follows:

$$f(x, y, s), \quad (2)$$

where x and y are the coordinates of the centers of the lakes, s is the area of the lakes.

Therefore, the set of circles is represented by the set of triples of random variables (x,y,s).To develop a modelling algorithm, it is necessary to take into account the absence of a statistical relationship between changes in the coordinates of location and sizes of lakes. The analysis carried out in [15] shows that random spatial changes in the areas and coordinates of the centers of lakes, as well as between the coordinates of the centers of lakes, are statistically independent. Therefore, to simulate the fields of thermokarst lakes, the joint probability density (3) can be represented as:

$$f(x,y,s) = f(x) \times f(y) \times f(s), \qquad (3)$$

where $f(x)$ and $f(y)$ are the probability densities of the uniform distribution, $f(s)$ is the probability density of the lognormal distribution law.

According to [14], the probability density for the lognormal law of the size lake distribution is determined by equation (1). In our case, estimates of mathematical expectation (M) and variance (D) for the lognormal distribution of lake areas in Western Siberia can be obtained from sample data in the form:

$$M = \sum_{i=1}^{m} (\ln s_i \times w_i), \qquad (4)$$

$$D = \sum_{i=1}^{m} ((\ln s_i)^2 \times w_i) - M^2, \qquad (5)$$

where m is the number of partial intervals of the empirical histogram, i is the number of the interval.

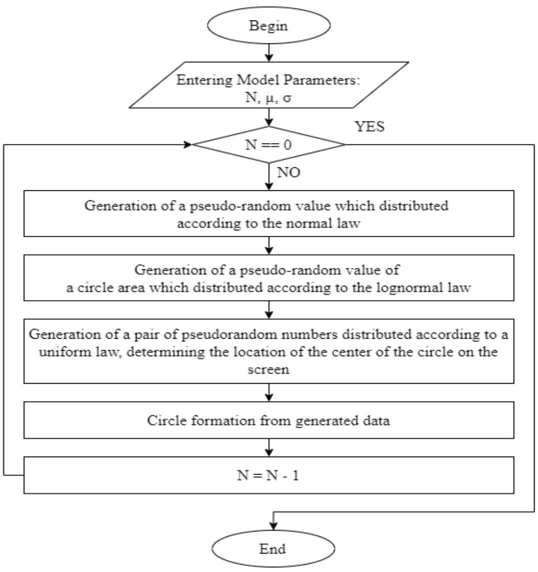

Fig. 1. Diagram of the algorithm for numerical modeling of thermokarst lakes fields.

Taking into account equation (3), a sequence of random numbers that determine the characteristics of the location of the circle centers is generated using a pseudo-random number sensor distributed according to the law of uniform density. To model the lakes with random sizes, the lakes areas are distributed by the law corresponding to equation (1). We will use the generation of a sequence of random numbers distributed according to the lognormal law according to the equation:

$$s_i = \exp(\mu + \sigma \times r_i), \qquad (6)$$

where r_i is a random number distributed according to the normal law and calculated by the formula:

$$r_i = \sum_{j=1}^{12} q_{ij} - 6, \qquad (7)$$

where q_{ij} is a random variable uniformly distributed on the segment [0,1].

In accordance with the above, the algorithm for the numerical simulation of the fields of the thermokarst lakes can be represented as the following sequence:

step 1. sets the number of simulated circles N;

step 2. the value of the lognormal distribution parameter μ is set;

step 3. the value of the parameter of the lognormal distribution σ is set;

step 4. a pseudo-random value is generated, distributed according to the normal law (7);

step 5. using the given parameters and the calculated pseudo-random value, the pseudo-random value of the circumference, distributed according to the log-normal law (6), is calculated by the formula;

step 6. a pair of pseudo-random numbers is generated, distributed according to a uniform law, determining the location of the center of the circle on the screen;

step 7. using the triple of numbers obtained in the previous steps 5 and 6, a circle is formed;

step 8. if the number of circles obtained is less than the number determined in step 1, then the algorithm repeats from step 4, otherwise the algorithm is completed.

Also, the algorithm for numerical modeling of fields of thermokarst lakes is shown in the form of a block diagram (Fig. 1).

IV. SOFTWARE IMPLEMENTATION OF MODELING SYSTEM

The structural diagram of the software package of geo-simulation modeling is presented in Fig. 2. The software package for geo-simulation modeling of thermokarst lakes fields was developed. It is a set of software modules providing input of model parameters, formation of pseudorandom numerical sequences and derivation of simulation results.

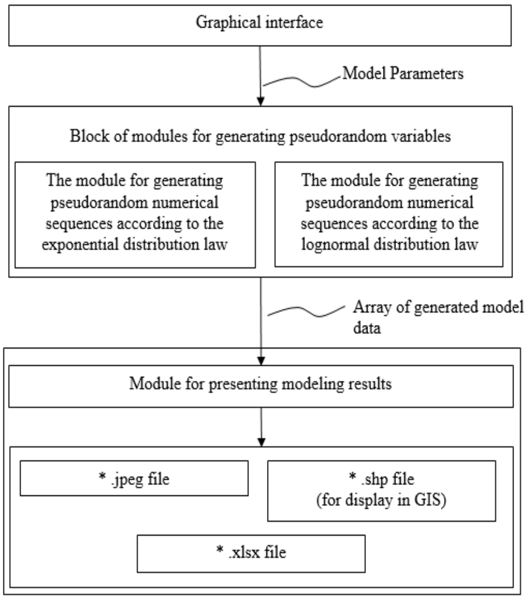

Fig. 2. A generalized scheme of a software package for geo-simulation modeling of thermokarst lakes fields.

In Fig. 2 diagram of the program complex is presented in the form of the following components:

- The graphical interface is intended for the user to select model parameters, the distribution law and the type of presentation of simulation results;
- The block of modules for generating pseudorandom variables is the main component of the simulation software package and is designed to generate random numerical sequences when implementing algorithms for the numerical simulation of fields of thermokarst lakes;
- The module for presenting simulation results is intended for converting simulation results into one of the following formats: Microsoft Excel (*.xlsx), vector format (*.shp), raster format (*.jpeg).

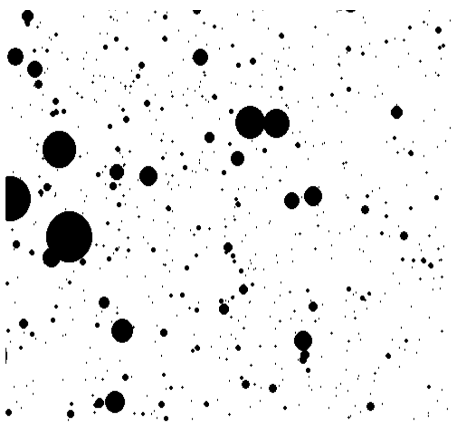

Fig. 3. A fragment of the model field of thermokarst lakes distributed according to the lognormal law.

Presented in the form of a diagram (Fig. 2), the software package is implemented in a high-level language - Python [16]. Python is an interpreted programming language. This means that the source code is converted in parts to the machine in the process of reading it by a special program - the interpreter. Python is a full-value and universal programming language which is widely used. The main, but not the only paradigm supported by it is object-oriented programming. In fact, the range of roles that Python can play as a multi-purpose programming language is virtually unlimited in scope. It can be used to implement, including software systems. Python interpreters are distributed freely under a license similar to the GNU General Public License [17].

To demonstrate the operation of the software package in Fig. 3 shows a fragment of the model field of thermokarst lakes. Numerical values of model parameters used for modeling: $M = 6{,}88$, $D = 3{,}42$ и $N = 3000$.

V. Conclusion

It is known [1] that thermokarst lakes in the Arctic regions are considered as one of the main sources of natural methane emissions. Of these, the most active methane generators are small lakes with sizes less than 500-1000 m^2. The methane concentration in the small lakes, according to [15, 18-19], is more than an order of magnitude higher than the methane concentration in larger lakes. Therefore, the use of the lognormal distribution law in the model makes it possible to take into account small lakes in modeling, what makes the developed geo-simulation modeling system a convenient tool for estimating methane reserves and emissions in thermokarst lakes. A geo-simulation modeling system can also be used to obtain predictive estimates of the dynamics of methane emissions from thermo-karst lakes for the coming decades.

Acknowledgment

The work was partially supported by grants under the RFBR projects No. 18-47-700001 and No. 18-45-860002.

References

[1] Modeling thermokarst lakes dynamics and carbon flux.–URL: http://www.docstoc.com/docs/35351317/Methods_of_thermokarst_lakes_modeling, free (accessed: 02.09.2020).

[2] A. S. Viktorov, The main problems of the mathematical morphology of the landscape. Moscow: Nauka, 2006, 252 p.

[3] A. S. Viktorov, T. V. Orlov, V. N. Kapralova, O. N. Trapeznikova, M. V. Arkhipov, S. A. Sadkov, A. V. Zverev, E. G. Panchenko, and P. V. Berezin, Mathematical morphology of cryolithozone landscapes, Moscow: RUDN, 2016, 232 p.

[4] N. N. Moiseev, “System analysis of the dynamic processes of the biosphere. The conceptual model of the biosphere,” Bulletin of the USSR Academy of Sciences, no. 2, pp. 47–54, 1979.

[5] Y. M. Polishchuk, and O. S. Tokareva, “Geoimitation modeling of atmospheric pollution zones as a result of gas burning in oil fields,” Information Systems and Technologies, no. 2, pp. 39–46, 2010.

[6] Y. Zhao, and Y. Murayama, “Urban dynamics analysis using spatial metrics geosimulation”, in Spatial analysis and

modeling in geographical transformation process, Y. Murayama, and R. Thapa, Eds. Dordrecht-Heidelberg-New York-London: Springer, The GeoJournal Library, 2011, vol. 100, pp. 153–168.

[7] E. G. Kapralov, A. V. Koshkarev, and V. S. Tikunov, Fundamentals of geoinformatics: textbook, manual for universities in 2 books, Moscow: Publishing Center “Academy”, 2004, 352 p.

[8] V. Y. Polishchuk, and Y. M. Polishchuk, Geo-simulation modeling of fields of thermokarst lakes in permafrost zones: monograph, Khanty-Mansiysk: UIP USU, 2013, 129 p.

[9] V. Y. Polishchuk, and Y. M. Polishchuk, “Modeling of thermokarst lake dynamics in West-Siberian permafrost,” in Permafrost: Distribution, Composition and Impacts on Infrastructure and Ecosystem, O. S. Pokrovsky, Ed. New York: Nova Science Publishers, 2014, vol. 6, pp. 205– 234.

[10] Y. M. Polishchuk, A. N. Bogdanov, and I. N. Muratov, “Methodological issues of constructing generalized histograms of the distribution of lake areas in the permafrost zone based on satellite images of medium and high resolution,” Current problems of remote sensing of the Earth from space, vol. 13, no. 6, pp. 224–232, 2016.

[11] Y. M. Polishchuk, A. N. Bogdanov, N. A. Bryksina, V. Y. Polishchuk, and I. N. Muratov, “Integration of satellite images of ultra-high and medium resolution for constructing histograms of the distribution of areas of thermokarst lakes in an extended range of their sizes,” Current problems of remote sensing of the Earth from space, vol. 15, no. 3, pp. 38–46, 2018.

[12] Y. M. Polishchuk, A. N. Bogdanov, N. A. Bryksina, I. N. Muratov, V. Y. Polishchuk, M. A. Kupriyanov, O. A. Baysalyamova, and V. P. Dneprovskaya, “Experience and results of remote research of cryolithozone lakes in Western Siberia from satellite images of various resolutions over a 50-year period,” Current problems of remote sensing of the Earth from the cosmos, vol. 14, no. 6, pp. 42–55, 2017.

[13] Y. M. Polishchuk, A. N. Bogdanov, and V. Y. Polishchuk, “The distribution of the areas of cryolithozone lakes in a wide range of their sizes according to satellite images of medium and high resolutions,” Bulletin of the Tomsk Polytechnic University. Geo-Resource Engineering, vol. 329, no. 3, pp. 16–25, 2018.

[14] N. S. Kremer, Probability Theory and Mathematical Statistics: Textbook for High Schools. Moscow: UNITY-DANA, 2004, 573 p.

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

[16] M. Lutz, Learning Python, 4th edition, trans. from English, Moscow–St-Petersburg: Symbol-Plus, 2011, 1280 p.

[17] GNU General Public License - URL: https://ru.wikipedia.org/wiki/GNU_General_Public_License (accessed: 02.15.2020).

[18] K. M. Walter, L. S. Smith, and F. S. Chapin, “Methane bubbling from northern lakes: present and future contributions to the global methane budget,” Philosophical Transactions of the Royal Society, vol. 365, pp. 1657–1676, 2007.

[19] M. A. Holgerson, and P. A. Raymond, “Large contribution to inland water CO_2 and CH_4 emissions from very small ponds,” Nature Geoscience Letters, vol. 9, pp. 222–226, 2016.