Modeling the spatial structure of thermokarst lake fields in the cryolithozone of Northern Eurasia accounting lognormal law of their size distribution

V Y Polishchuk\textsuperscript{1,3}, I N Muratov\textsuperscript{2}, and Y M Polishchuk\textsuperscript{2,4}

\textsuperscript{1}Institute of Monitoring of Climatic and Ecological Systems SB RAS, Akademichesky Ave., 10/3, Tomsk, 634055, Russia
\textsuperscript{2}Yugra Research Institute of Information Technologies, Mira, 151, Khanty-Mansiysk, 628011, Russia
\textsuperscript{3}Tomsk Polytechnic University, Lenina Ave., 30, 634050, Tomsk, Russia
\textsuperscript{4}Institute of Petroleum Chemistry SB RAS, Akademichesky Ave., 4, Tomsk, 634055, Russia

Abstract. Power and exponential lake size distributions are now routinely used for modeling the spatial structure of thermokarst lake fields. Based on a combination of satellite images of moderate and high spatial resolution, a synthesized lake size distribution histogram is constructed, which is approximated by a lognormal distribution law using the Pearson criterion with a probability of 0.99. This distribution law takes into account small lakes which are considered as intensive sources of methane emission into the atmosphere from thermokarst lakes. Based on a geo-simulation approach, a model of the spatial structure of thermokarst lake fields is proposed taking into account the lognormal lake size distribution. Algorithms for modeling the spatial structure of thermokarst lake fields are presented. An example of modeling a field of thermokarst lakes with a lognormal lake size distribution is given.

1. Introduction
The current global warming of the climate, most clearly manifested in the northern latitudes of the planet, accelerates the degradation of permafrost. Permafrost, being a storehouse of conserved carbon in the vast frozen peat bogs of Northern Eurasia, can become a source of even greater warming with the release of greenhouse gases, which will lead to the formation of new big challenges for the world community related to the violation of human-nature interaction. Indeed, carbon is currently in a bound state as an organic matter in the permafrost layer in the northern territories of Eurasia and America. With the climate warming, a temperature rise will lead to the melting of frozen rocks and additional release of methane as a product of the vital activity of microorganisms recycling thawed organic matter, which can make an additional tangible contribution to the climate warming.

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 yet been taken into account. Attempts to take them into account in estimating the total volume of world methane reserves in a recently published paper [2], based on the use of the theoretical power law of the size distribution of lakes due to lack of experimental data, raise great doubts, since the power law is not supported by experimental data [3]. 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 the World Summit on Climate (Paris, 2015) calls for the formation of forecasts of the dynamics of lake methane stock in the lakes of northern territories for the coming decades. This required the development of methods and tools for modeling the dynamics of thermokarst lake fields that would allow for the contribution of millions of small lakes to the total amount of methane reserves in the vast territories of Northern Eurasia.

According to Moiseev and Svirezhev [4], simulation modeling is a research method which can build an approximate model of spatial objects being studied. Simulation modeling is one of the most
important mathematical modeling types which may be used to construct an efficient model of thermokarst lake fields with an accuracy that is sufficient for the current research. Low and Kelton [5] claim that simulation modeling is used to construct models in cases where, first, there is no analytical solution or this solution is very complex and requires huge computer capacity and, second, the amount of experimental data about an object being modeled is insufficient for the statistical method. In this case a mathematical model is developed in simulation modeling.

Various groups of authors introduced special terms for modeling of spatial objects. Berlyant et al [6], Tikunov [7], Kuzmichenok [8], Kovalev [9], Serdutskaya and Yatsishin [10], Kulik and Yurofeev [11], and Timonin [12] have named this type by mathematical-cartographical modeling. Lawson and Denison [13], Wang [14], and Poh-Chin et al [15] introduced the term of spatial modeling. Polishchuk and Tokareva [16], Zhao and Murayama [17] introduced the term of geo-simulation modeling.

All above types of modeling are aimed at studying spatial objects using spatial data whose analysis is based on spatial analysis methods implemented with the help of modern geo-information systems (GIS-analysis). In our opinion, the most suitable general term for the above-mentioned types of modeling (mathematical-cartographical, spatial, geo-information, geo-simulation, etc.) is the term ‘geo-simulation modeling’, which is defined as the creation of a model and model application to objects with spatial structure. Problems of creating a geo-simulation model of thermokarst lake fields are studied by Polishchuk V. and Polishchuk Yu. in [18, 19]. Nowadays the power and exponential laws of the lake size distribution based on data from Landsat space images are used for modeling the spatial structure of thermokarst lakes fields. These images have a spatial resolution of 30 m that does not allow studying small lakes. The main goal of the paper is using the lognormal law to take into account small lakes in modeling the spatial structure of thermokarst lake fields.

2 Geo-simulation model of spatial structure of thermokarst lake fields

The creation of a geo-simulation model of thermokarst lake fields requires knowledge of the basic properties of these fields, which can be obtained experimentally. Because of the inaccessibility of the northern territories of Eurasia, thermokarst experimental studies were carried out by remote sensing. For the remote study, twenty-nine test sites (TSs) were chosen in different zones of the West-Siberian permafrost (sporadic, discontinuous, and continuous ones). The remote study of the shape of the boundaries of thermokarst lakes was carried out using satellite images. Investigations performed at test sites in sporadic, discontinuous, and continuous permafrost showed that the error in estimating the lake areas while replacing their real lake boundaries by a circle is comparatively small (about 5% [19,20]). It may be the reason for choosing a circle as a model of lake in geo-simulation modeling of thermokarst lake fields. In addition, the formation of a geo-simulation model of thermokarst lake fields in the form of a population of random circles requires experimental knowledge of the distribution of the coordinates of lake centers and the distribution of lake sizes (lake size distribution). To establish the regularities of the distribution of random coordinates of lakes, satellite images were used. Analysis of histograms of the distribution of latitude and longitude values of the locations of lake centers given by Polishchuk V. and Polishchuk Yu. In [18] showed that experimental regularities of the distribution of the coordinates of lake centers correspond to the law of uniform density according to the criterion $\chi^2$ with a probability of 95% [19].

Hence, the following fundamental principles determining substantial properties of the model of spatial structure of thermokarst lake fields can be formulated:

1. Lake coastline shapes can be represented by a circle equation with the coordinates of centers $x_i$, $y_i$ and the area $s_i$ ($i$ is the lake serial number).
2. Spatial changes in the position of the centers of circles and their areas are statistically independent.
3. Random distribution of the circle centers’ coordinates $x_i, y_i$ ($i = 1, n$) is governed by a uniform law.
A model of a thermokarst lake field is a population of random circles (Fig. 1) whose statistical characteristics correspond to the above principles (1-3). Figure 1 presents a geometrical interpretation of the model of thermokarst lake fields. Using a triplet of numbers \((x, y, s)\) representing the value of the center of the circle coordinates and the value of its area, the coordinates of the points determining the borders of each circle are calculated as follows:

\[
\begin{align*}
    x_k &= R \cos \gamma + x_k, \\
    y_k &= R \sin \gamma + y_k,
\end{align*}
\]

where \(x\) and \(y\) are the coordinates of the center of the circle;

\(x_k\) and \(y_k\) are the coordinates of the \(k\)–th point on the circle;

\(\gamma\) is the value of the axial angle \(x\) and the radius directed from the center of the circle to the \(k\)–th point on the circle

\(R\) is the radius of the circle computed using the formula \(R = \sqrt{s/\pi}\),

where \(s\) is the area of the circle.

Major elements in the model description are the characteristics of lake shapes, the parameters of their random location on the surface, and the specific form of the size distribution law of lakes. The properties of the model fields of thermokarst lakes will depend substantially on the form of the distribution law of lakes in size (areas). The form of the size distribution law can be determined experimentally from satellite images.

At present, the lakes inventory data are used in studying the regularities of lake size distribution both at planetary [21] and regional [3, 19] levels based on the results of remote measurements of areas and the number of lakes from satellite imagery of the average spatial resolution of Landsat (resolution: 30 m).

3 Modeling of thermokarst lake fields
In the general case, the mutual density of the probabilities of random coordinates of the centers and areas of circles imitating lakes in a mathematical model of random thermokarst lake fields can be presented in the form:

\[f(x, y, s),\]

where \(x\) and \(y\) are the coordinates of the circle center in a model;

\(s\) is the area of a circle imitating a lake.

Consequently, the aggregate of circles in the model of lake fields will be presented as a set of groups of three random values \((x, y, s)\). To develop an algorithm for modeling thermokarst lake fields, it is necessary to take into consideration the forms of the distribution law values of \(x, y\) and \(s\).
and the statistical connections between changes in the lake coordinates and their areas which are statistically independent according to [19]. Hence, together with using software generators for the distribution of pseudo-random numbers, the software realization of an imitation model of thermokarst lake fields includes creating a generator for pseudo-random number sequences distributed in accordance with the lake size distribution law determined based on satellite images.

We should consider a geo-simulation model of the spatial structure of a thermokarst lake field that is an aggregate of circles and reflects the state of the thermokarst lake field whose geometrical representation is shown in Fig. 1. According to the above, the coordinates \( x \) and \( y \) are distributed according to a uniform law. Based on Landsat satellite imagery in [19] it is shown that lake areas are distributed according to an exponential law. Therefore, a one-parameter exponential law was chosen for describing thermokarst lake size distribution in the following form:

\[
f(s) = \lambda \times e^{-\lambda s},
\]

where \( \lambda \) is a parameter of the distribution law.

The value of the parameter \( \lambda \) can be determined with the help of experimental data. According to Wentzel [22], the mathematical expectation of a random quantity according to the distribution law is determined in the form

\[
M(s) = \frac{1}{\lambda}.
\]

Using, as an estimate of the mathematical expectation \( M(s) \), the value of the average lake area estimated in accordance with experimental data in the following form:

\[
\bar{S} = \frac{1}{n} \sum_{i=1}^{n} S_i, \quad i = 1, n.
\]

we find that the parameter \( \lambda \) can be determined with the help of experimental data by the formula

\[
\lambda = \frac{1}{\bar{S}}.
\]

Testing of the correspondence of the exponential law of lake area distribution given by Eq. (1) with experimental histograms was carried out in [19]. It is shown that at all test sites in studied territory this law corresponds to experimental data in accordance with the criterion \( \chi^2 \) with an average probability of 90\%. Consequently, the exponential law of lake size distribution in the form (1) according to areas does not contradict the experimental data. An analysis of the experimental size distribution of lakes according to their areas has shown that \( \lambda \) at all test sites in a continuous permafrost varies in the range of 0.037 – 0.071 and in a discontinuous permafrost, in the range of 0.034 – 0.086 with average values of 0.054 and 0.060, respectively. The estimation has shown that the error of determination of the average values of lake areas by modeling with the use of experimental data is 17\%. This may be regarded as a suitable result of modeling thermokarst lake fields for predicting thermokarst lake field dynamics.

However, this model does not take into account small lakes (with sizes less than 0.2-0.5 ha) which have an increased concentration of methane in the lake water. These small lakes are not detected in Landsat images and, therefore, are not taken into account in the model within the exponential distribution of lakes in size, which makes it necessary to use satellite imagery of high and ultra-high resolution.

In this connection, investigations of the distribution of lakes in the cryolithic zone of Western Siberia were carried out using high-resolution space images of Canopus-B (resolution: 2 m) in combination with Landsat-8 (30 m) images. The processing of space images carried out using the standard tools of the geographic information system ArcGIS 10.3 was aimed at obtaining data on the number and areas of lakes with significantly different sizes for the purpose of constructing a histogram of the distribution of lakes in a very wide range of their sizes. The construction of such a histogram of the distribution of the lake areas in an extremely wide range of changes in their sizes (from units of m\(^2\) to hundreds of km\(^2\)) from satellite imagery was carried out using a technique described in [23], which proposed a three-stage procedure for constructing a histogram based on the integration of satellite data.
on the areas and the number of lakes. As a result, a single synthesized histogram of lake distribution over areas (Fig. 2A) was obtained in a wide range of sizes, resulting from the synthesis of two initial histograms of the lake distribution obtained separately from high (HR) and medium (MR) resolution images.

**Figure 2.** Illustration of MR (Landsat-8, black columns) – HR (Kanopus-V, white columns) lake number (A) and area (B) histogram coupling within a total range of 20 to $2 \times 10^8$ m$^2$ with an overlap in the range of $5 \times 10^3$ to $10^6$ m$^2$. Vertical line marks the point of integration (stitching) of high- and medium-resolution-based lake diagrams.

For the construction of such a histogram, partial intervals with an irregular pitch (logarithmic law) were chosen, namely: 2-5 m$^2$, 5-10 m$^2$, 10-20 m$^2$, 20-50 m$^2$, etc. up to 200 km$^2$, which made it possible to provide data on the distribution of lakes over intervals of their sizes quite compactly in a very wide range of changes in the lake areas. The synthesized histogram taking into account both large and small lakes obtained on the basis of integration of data from satellite images of different resolution allows approximation by a lognormal distribution law.

According to [24], the probability density for the lognormal distribution law of the area of lakes ($s$) is determined by the equation
where $\mu$ is the expectation, and $\sigma$ is the standard deviation.

In our case, estimates of the mathematical expectation (M) and variance (D) for the lognormal distribution of the lake areas in the permafrost zone were obtained from experimental sample data in [23] in the form $M = 6.88$ and $D = 3.42$. Testing of the compliance of the empirical and theoretical distributions, performed with the Excel software package, using the Pearson test showed that the synthesized histogram of the distribution of the lake areas obtained over a wide range of their sizes corresponds to a lognormal law with a probability of 0.99. Fig. 2B presents the empirical distribution of the total area of lakes according to their sizes, which shows that lakes with sizes from 2 to 500 ha give the bulk of the total area of lakes (about 80%).

Based on the above geometric interpretation of the model of the spatial structure of the thermokarst lake fields, where the circles are characterized by three random variables, namely, the two coordinates of the centers and area, we briefly describe the procedure of modeling the fields of thermokarst lakes in accordance with the lognormal distribution. The values of the coordinates of the centers are calculated using a pseudo-random number generator distributed according to the law of uniform density, as described in [19]. Calculation of the pseudorandom value characterizing the area of the circle is performed using a generator of pseudorandom numbers distributed according to the lognormal law in accordance with formula (5). The result of the procedure is the modeled field of model lakes whose area values obey the lognormal distribution law.

For illustration, Fig. 3 shows the result of modeling of a simulated image of the field of thermokarst lakes. Here, the lognormal distribution parameters $M = 6.88$ and $D = 3.42$, which characterize the permafrost zone in the territory of Western Siberia, were used. When modeling the fragment, the number of model lakes 3000 is specified.

![Figure 3. Result of modeling a field of thermokarst lakes with a lognormal size distribution.](image)

4. Conclusions
This paper described an approach to modeling the spatial structure of thermokarst lakes on the basis of a geo-simulation model that represents a set of random circles with a uniform distribution of the coordinates of their centers and a lognormal size distribution. An experimental substantiation of the lognormal distribution of the lake size was given on the basis of the results of a study on an empirical distribution of thermokarst lake areas in a very wide range of their sizes in the permafrost zone based on a joint use of space images of various types of spatial resolution. The results of testing the compliance of this law with an empirical histogram showed that the lognormal law fits the experimental data, according to the Pearson criterion, at a significance level of 0.99.

A procedure was briefly described for modeling a field of thermokarst lakes, where each model lake is characterized by a triplet of numbers: the coordinates of the center and the area. A fragment of
a simulated field of thermokarst lakes was presented. The statistical characteristics of the model field of lakes were obtained from images of moderate and high spatial resolution.

The above-obtained results can be used to predict the dynamics of thermokarst lake fields in the permafrost zone of Northern Eurasia for the coming decades in the context of decisions of the climatic summit of Paris, 2015.

Acknowledgements
This work was supported by the RFBR under grants no. 18-47-700001, no. 18-45-860002 and no. 18-45-703001

References
[1] Pokrovsky O S, Shirokova L S, Kirpotin S N, Audry S, Viers J and Dupre B 2011 Effect of permafrost thawing on the organic carbon and metal speciation in thermokarst lakes of Western Siberia Biogeosciences 8 565-83
[2] Holgerson M A and Raymond P A 2016 Large contribution to inland water CO2 and CH4 emissions from very small ponds Nature Geoscience Letters 9 222-26
[3] Cael B B and Seekell D A 2016 The size-distribution of Earth's lakes Sci. Rep. 29633
[4] Moiseev N N and Svirzhev Y M 1979 System analysis of dynamic processes of the biosphere Bulletin of the Academy of Sciences of USSR 2 47-54(in Russian)
[5] Low A M and Kelton W D 2004 Simulation: Classic Computer Science (St. Petersburg-Kiev: Publishing Group BHV) p 847 (in Russian)
[6] Berlyant A M, Zhukov V T and Tikunov V S 1976 Mathemattico-cartographic modeling in the "creation - the use of maps" Bulletin of Moscow State University 235-43 (in Russian)
[7] Tikunov V 1997 Simulation in cartography: studies for high schools (Moscow: Moscow State University Press) p 405 (in Russian)
[8] Kuzmichenok V A 2003 Mathematics and cartographic modeling of possible changes in water resources and glaciers of Kyrgyzstan to climate change Bulletin of the Kyrgyz-Russian Slavic University 6 33-41(in Russian)
[9] Kovalev T M 2008 News of Altai University 1 58-62 (in Russian)
[10] Serdutskaya L F and Yatishin A V 2009 Technoeociology: Mathematical and cartographic modeling (Moscow: Publishing house LIBROKOM) p 232 (in Russian)
[11] Kulik K N and Yurofeev V G 2010 Computer mathematics and cartographic modelling agrolesolandshafts based on aerospace information Reports of Russian Academy of Agricultural Sciences 1 52-4 (in Russian)
[12] Timonin S A 2010 Mathematics and mapping and GIS modeling of demographic processes in the regions of the Russian Federation Bulletin of Moscow State University Ser. Geography 5 11-18 (inRussian)
[13] Lawson A and Denison D 2002 Spatial cluster modeling (Boca Raton-London-New-York: CRC Press / Chapman and Hall) 287
[14] Wang X 2005 Integrating GIS, simulation models and visualization in traffic impact analysis Computers, Environment and Urban Systems 29 471–96
[15] Poh-Chin Lai, Fun-Mun So and Ka-Wing Chan 2009 Spatial epidemiological approaches in disease mapping and analysis (Boca Raton-London-New-York: CRC Press / Tailor and Francis Group) 184
[16] Polishchuk Y and Tokareva O 2010 Geosimulation modeling of air pollution zones as a result of the gas burning in oil fields Information Systems and Technology 2 39-46 (in Russian)
[17] Zhao Y and Murayama Y 2011 Urban dynamics analysis using spatial metrics geosimulation Spatial analysis and modeling in geographical transformation process eds Y Murayama and R Thapa (Dordrecht-Heidelberg-New-York-London: Springer, GeoJournal Library) 100 153–68
[18] Polishchuk V and Polishchuk Y 2014 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) 6 205-34

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

[20] Polishchuk V and Polishchuk Y 2012 Remote studies of variability of the shape of coastal boundaries of thermokarst lakes in the permafrost of West Siberia Study of Earth from space 1 61–4 (in Russian)

[21] Verpoorter C, Kutser T, Seekel D and Tranvik L 2014 A global inventory of lakes based on high resolution satellite imagery Geophys. Res. Lett. 41 1–7

[22] Wentzel E S 2002 Probability theory (Moscow: Vyschaya School) p 575 (in Russian)

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

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