Modeling spatial structure of thermokarst lake fields in permafrost of Western Siberia based on satellite images

VYu Polishchuk1,3, IN Muratov2, YuM Polishchuk2,4

1 Institute of Monitoring of Climatic and Ecological Systems SB RAS (Tomsk, Russian Federation)
2 Yugra Research Institute of Information Technologies (Khanty-Mansiysk, Russian Federation)
3 Tomsk Polytechnic University (Tomsk, Russian Federation)
4 Institute of Petroleum Chemistry SB RAS (Tomsk, Russian Federation)

Corresponding author: Ildar Muratov (ildarmur@gmail.com)

Abstract

Deciphering the satellite images of medium and high spatial resolution of the northern territories of Western Siberia has been carried out using geoinformation system ArcGIS 10.3. Images of medium resolution Landsat-8 and high resolution Kanopus-V were used. Kanopus-V images alluded to determine the number and areas of small lakes, which are considered as intensive sources of methane emission into the atmosphere from thermokarst lakes. Data on the spatial characteristics of thermokarst lakes were obtained. Based on the integration of images of medium and high spatial resolution, a synthesized histogram of the distribution of lakes in a wide range of sizes was constructed, taking into account small lakes. The obtained histogram was approximated by a lognormal distribution law by the Pearson criterion with a probability of 0.99. Based on the geo-simulation approach, a new model of the spatial structure of the fields of thermokarst lakes is presented, taking into account the lognormal law of the lake size-distribution. Algorithms for modeling the spatial structure of the fields of thermokarst lakes are described. An example of modeling the field of thermokarst lakes with a lognormal law of their size-distribution is given. The practical applicability of the previously developed model with an exponential distribution of lakes in size, based on data from Landsat images, has been experimentally confirmed. The results can be used to obtain predictions of the dynamics of methane emissions from the thermokarst lakes of the Arctic zone of Northern Eurasia for the coming decades in the context of climate changes.

Keywords

geographic information systems, geo-simulation modeling, permafrost, remote sensing methods, satellite imagery, size-distribution of lakes, thermokarst lakes, Western Siberia
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 canned carbon in the vast frozen peat bogs of Northern Eurasia, can become a source of even more 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 a layer of permafrost in the northern territories of Eurasia and America. With the warming of the climate, a rise in temperature will lead to the melting of frozen rocks and the 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 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 (Pokrovsky et al. 2011) 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 (Holgersen and Raymond 2016), based on the use of the theoretical power law of the size distribution of lakes due to the lack of experimental data, require great doubts, since the power law is not supported by experimental data (Cael and Seekell 2016). 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 (1979), simulation modeling is a research method which can build an approximate model of a studied spatial objects. Simulation modeling is one of the most important mathematical modeling types which may be used for construction of sufficient model of the thermokarst lake fields with accuracy sufficient for current research. Low and Kelton (2004) claim that simulation modeling is used to construct models in cases where, firstly, there is no analytical solution or this solution is very complex and requires huge computer capacity and, secondly, the amount of experimental data about a modeled object is insufficient for statistical method. In such case a mathematical model is developed in simulation modeling.

Different groups of authors have introduced the special terms for modeling spatial objects. Berlyant et al. (1976), Tikunov (1997), Kuzmichenok (2003), Kovaleva (2008), Serdutskaya and Yatsishin (2009), Kulik and Yurofeev (2010), and Timonin (2010) have named this type by mathematico-cartographical modeling. Lawson and Denison (2002), Wang (2005), Poh-Chin et al. (2009) have introduced spatial modeling. Polishchuk and Tokareva (2010), Zhao and Murayama (2011) have introduced geo-simulation modeling.

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 such a model was considered in (Polishchuk and Polishchuk 2014), in which experimental data on the properties of lakes in the permafrost zone of Western Siberia were used, obtained from images of the average resolution of Landsat, in which small lakes are not visible. Therefore, to account for small lakes, it is necessary to attract high resolution images.

In connection with the above, the main goal of this work was to consider the issues of modeling the spatial structure of thermokarst lakes fields based on the integration of medium and high resolution images that take into account lakes of all sizes.

**Materials and methods**

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 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 fairly short period of the summer season (end of June – August) to minimize the effect of seasonal fluctuations in the water level in the lakes. During this period, the ice cover on the lakes completely disappears, preventing them from being excreted when interpreting images.

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-B (2.1 m) was carried out on a set of test sites, the map-layout of which in the different zones of permafrost in Western Siberia is shown in Fig. 1.

Test sites were chosen by us, as is customary in similar studies, in places where thermokarst lakes accumulate, i.e. in zones where conditions exist for the formation and development of foci of thermokarst processes (Viktorov et al. 2016). On each test site (TS), using satellite images, we determined from several hundred to several thousand thermokarst lakes.

The coastal boundaries of thermokarst lakes were determined from Landsat satellite images using the Fmask algorithm. This algorithm uses a number of parameters obtained from the spectral channels of Landsat satellite images to build assumptions about the nature of reflections of various geographical objects in different spectral ranges of Landsat-8. The algorithm is described in detail in (Zhu et al. 2015). The high-resolution satellite images of Kanopus-V were deciphered using the binary classification method. This method is based on the use of the pixel brightness threshold in images, which is used to select the boundaries of water objects. The processed panchromatic images have a strong contrast between
the water surface and the surrounding vegetation, which allows you to visually determine the threshold values for the classification of water.

The fragment of the deciphered space image, resulted on Fig. 2, illustrates the characteristic feature of the fields of thermokarst lakes, namely: the presence in the images of lakes varying in size – from very large to small and very small lakes. The processing of satellite images aimed at obtaining data on the number, location and areas of lakes was carried out using standard means of geographic information system ArcGIS 10.3.

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. A remote study of the shape of the boundaries of thermokarst lakes conducted in (Polishchuk and Polishchuk 2014) in different zones of the West Siberian permafrost (sporadic, discontinuous and continuous) showed that the error in estimating lake areas when replacing the boundaries of their real lakes by circles is relatively small (about 5% (Polishchuk and Polishchuk 2014)). This can serve as a basis for choosing a circle as a model of a lake in geo-simulation modeling of thermokarst lakes fields. And the formation of a geo-simulation model of fields of thermokarst lakes in the form of a set of random circles requires experimental data on the distribution of coordinates of the centers of lakes and the size-distribution of lakes. An analysis of histograms of the distribution of coordinates (latitude and longitude) of the location of the centers of lakes showed (Polishchuk and Polishchuk 2014) that empirical histograms of the distribution of coordinates of the centers of lakes correspond to the law of uniform density, according to the χ² criterion, with a probability of 95% (Polishchuk and Polishchuk 2014).

For the development of the model, the task of constructing histograms of the distribution of lakes by size (area), which would take into account all the lakes of the study area in a wide range of sizes – from tens of meters to tens of kilometers, becomes important. To construct such a histogram in (Polishchuk et al. 2018a) it was proposed to choose partial intervals of the histogram with an irregular pitch (according to a logarithmic law), namely: 20–50 m², 50–100 m², 100–200 m², etc. up to 200 km², which made it possible to present data on the distribution of lakes over intervals of their sizes rather compactly over 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 satellite images of both medium and high resolution. 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 (Polishchuk et al. 2018a). 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 at 78 test sites in all three permafrost zones represents small lakes (from 0.005 to 20 ha).

Fig. 2. Fragment of a decrypted Landsat-1 image (1973) with the image of field of thermokarst lakes
Results

Histogram of the distribution of the number and total area of lakes in size

In Fig. 3 shows the histograms of the distribution of the number and total area of lakes according to the results of the conducted research. The original histograms obtained separately from high and medium resolution images have an overlapping (overlapping) area of 5 intervals located in an area range of 0.5–20 ha. As shown in (Polishchuk et al. 2018b), a practically acceptable error of remote measurement of the area of lakes from medium-resolution (MR) images is achieved with lakes of 2 ha or more. Therefore, each synthesized histogram in Fig. 3 was obtained on the basis of the “stitching” of two initial histograms at the points corresponding to the value of the area of 2 hectares and marked on the graphs (Fig. 3) with vertical segments of a straight line.

The synthesized histograms of the distribution of the number of lakes and their total area in size (Fig. 3A, B, respectively) obtained as a result of such “stitching” are determined in a wide range of changes in the areas of lakes from 50 m² to 20,000 hectares, while at intervals of less than 2 hectares, data from high-resolution images (HR) are used, and in intervals of more than 2 hectares, data from MR images is used.

Lognormal distribution of lakes by size

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 synthesized histogram of the distribution of lakes (Fig. 3A), which showed (Polishchuk et al. 2018b) that the empirical distribution corresponds to a lognormal law. The compliance check was performed with help of the Excel software package using Pearson's criterion confirmed that the

![Fig. 3. The illustration of MR (Landsat-8, black columns) – HR (Kanopus-V, white columns) lake number (A) and area (B) histogram coupling within the total range of 20 to 2×10⁶ m² with and overlap in the range of 5×10⁴ to 10⁶ m². Vertical line marks the point of integration (stitching) of high and medium-resolution-based lake diagrams](image-url)
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.

According to (Kremer 2003), density of the probability of a lognormal size-distribution \( f(s) \) determined by the equation:

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

where \( S \) – area of a circle imitating a lake, \( \mu \) is the mathematical expectation, \( \sigma \) is the standard deviation.

**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 lake field:

1. The shape of the shoreline of the lake can be represented by the equation of a circle with the coordinates of the centers \( x_i, y_i \) and the area \( s_i \) (i is the serial number of the lake).
2. The coordinates of the centers of the circles are random variables, the distribution of which is determined by the law of uniform density.
3. The circles sizes are random variables whose distribution is determined by a lognormal law.
4. Spatial changes in the coordinates of the centers of the circles and their areas are statistically independent.

The geo-simulation model of a field of thermokarst lakes developed in accordance with these principles is a set of random circles (Fig. 4), whose statistical characteristics correspond to the above principles (1-4). In Fig. 4 shows the geometric interpretation of the model of thermokarst lake fields. For the geo-simulation, triples of numbers are used, representing the coordinates of the centers of the circles and the value of their areas. The coordinates of the points defining the boundaries of each circle are calculated as follows:

\[
x_k = R \cdot \cos \gamma + x, \tag{2}\]

\[
y_k = R \cdot \sin \gamma + y, \tag{3}\]

where \( x \) and \( y \) – coordinates of the center of the circle; \( x_k \) and \( y_k \) – coordinates of \( k \) – th point on the circle; \( \gamma \) – value of axial angle \( x \) and radius, directed from the center of the circle into \( k \) – th point on the circle; \( R \) – radius of the circle, computed using the following formula

\[
R = \sqrt{s / \pi} \tag{4}
\]

where \( s \) – area of the circle.

**Fig. 4.** Geometrical representation of fragment of the thermokarst lake field model as an aggregate of random circles

**Algorithms of modeling thermokarst lakes fields with lognormal size-distribution**

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

\[
f(x, y, s) \tag{4}
\]

where \( x \) and \( y \) – coordinates of circle center in a model. Consequently, the set of circles in the model of lake fields will be represented as a sequence of triples of random variables. In order to develop an algorithm for modeling thermokarst lake fields, it is necessary to take into account the type of \( x, y \) and \( s \) distribu-
tion laws and the statistical relationships between changes in the coordinates of lake centers and their areas, which, according to (Polishchuk and Polishchuk 2014), are statistically independent. With this in mind to simulate the fields of thermokarst lakes, the joint probability density (4) can be represented as:

\[ f(x, y, s) = f(x) \times f(y) \times f(s) \] (5)

where \( f(x) \) and \( f(y) \) of the probability density of a uniform distribution.

Taking into account equation (5), the generation of a sequence of random numbers that determine the characteristics of the location of the centers of circles is carried out using a pseudo-random number generator 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 (1), sequences of pseudo-random numbers are generated that satisfy the lognormal distribution law, in accordance with the equation obtained in (Dubner 2000):

\[ s_i = \exp(\mu + \sigma \times r) \] (6)

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

\[ r = \sum_{j=1}^{12} q_j - 6 \] (7)

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

The implementation of algorithms for modeling the spatial structure of the fields of thermokarst lakes described above was carried out using the C# programming language (Hejlsberg et al. 2012) in the Visual Studio development environment (Randolph et al. 2011). These software tools are universal means of implementing modern software products. Using these tools, in our work, a new software package of geo-simulation modeling of thermokarst lakes fields has been implemented, a generalized scheme of which is shown in Fig. 5. The following components are included in this complex:

1) data entry tools required for modeling thermokarst lakes;
2) the module for generating pseudorandom number variables distributed according to the normal law in accordance with equation (7);
3) a module for generating pseudorandom number variables distributed according to a lognormal law in accordance with equation (1);
4) means of displaying simulation results in a spreadsheet format in the form of files with the extension .xlsx and in the graphical format like Fig. 6.

**Fig. 5.** General structure of the software package for geo-simulation modeling of thermokarst lakes fields

**Fig. 6.** Result of modeling the field of thermokarst lakes with the lognormal size-distribution law
The result of the work of the geo-simulation software system is a field of model lakes, the areas of which are distributed according to a lognormal law. For illustration in Fig. 6 shows the fragment of a modeled image of a field of thermokarst lakes. In the simulation of this fragment, the number of model lakes 3000 and the parameters of the lognormal distribution \( M = 6.88 \) and \( D = 3.42 \), defined by experimental data based on satellite images for the permafrost zone of Western Siberia, were specified.

**Discussion**

The data presented in the “Results” section allows us to experimentally substantiate the applicability of the previously developed (Polishchuk and Polishchuk 2014) geo-simulation model of the spatial structure of thermokarst lakes fields with an exponential distribution of lake sizes, the parameters of which were determined from Landsat average resolution images (Polishchuk and Polishchuk 2014). In Fig. 3B shows the empirical distribution of the total area of lakes depending on their size, which shows that lakes with sizes from 2 to 500 hectares give the main part of the total area of lakes (about 80%).

The approximation of the truncated lakes distribution histogram, made using the Excel software package, showed that the empirical histogram of the distribution of lake areas in the size range 2-500 ha with a high coefficient of determination \( (r^2 = 0.72) \) corresponds to an exponential law. This can be considered as an experimental confirmation of the practical applicability of a previously developed model with an exponential distribution of lakes in size, obtained from Landsat image data. Such a simplified model can be used, for example, to estimate water reserves accumulated in thermokarst lakes of the Arctic zone of Russia, or to study and predict the dynamics of areas of thermokarst lakes under conditions of climatic changes and other tasks.

However, the model with the exponential distribution of lakes does not take into account small lakes, which are considered intensive sources of methane emissions. These small lakes are not found on medium-resolution Landsat images and therefore are not involved in developing a model with an exponential distribution of lake sizes. Therefore, in modeling the fields of thermokarst lakes in the permafrost zone, which require consideration of small lakes, for example, in estimating methane emissions, it is necessary to use a model with a log-normal distribution of lake size based on the sharing of high-resolution satellite images.

A visual comparison of the graphs in Figs 2, 6 shows that the model field of thermokarst lakes quite well reflects the texture and basic properties of the image of real lake fields in a satellite image. Therefore, computer geo-simulation modeling of thermokarst lakes fields with a lognormal lake size distribution can be used to study the spatial structure of lake fields in thermokarst-lake plains of other Arctic regions. For this, it is necessary to determine the parameters of the lognormal distribution of lakes in size for the territory under study based on the integration of medium and high resolution satellite images.

**Conclusion**

The article presents an approach to modeling the spatial structure of the fields of thermokarst lakes based on a geo-simulation model representing a set of random circles with a uniform distribution of the coordinates of their centers and a log-normal distribution of lake areas. An experimental substantiation of the lognormal distribution of lake sizes is given on the basis of the results of research on the empirical distribution of areas of thermokarst lakes in a very wide range of their sizes in the permafrost zone based on the joint use of satellite images of different spatial resolution obtained for the northern territories of Western Siberia. The results of checking the compliance of this law with the empirical histogram showed that the log-normal law corresponds to the experimental data, according to the Pearson criterion, at a significance level of 0.99.

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

The results can be used to obtain predictions of the dynamics of methane emissions from the thermokarst lakes of the Arctic zone of Northern Eurasia for the coming decades in the context of climate summit decisions (Paris, 2015).

**Acknowledgments**

The reported research was funded by Russian Foundation for Basic Research and the government of the Tomsk region of the Russian Federation, grants № 18-47-700001, № 18-45-860002 and № 18-45-703001.

**References**

- Polishchuk YuM, Bogdanov AN, Muratov IN, Polishchuk VYu, Lim A, Manasyppov RM, Shirokova LS, Pokrovsky OS (2018a) 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. Environment-


