

# МЕТОДИЧЕСКИЕ ВОПРОСЫ ПРОГНОЗИРОВАНИЯ ДИНАМИКИ ТЕРМОКАРСТОВЫХ ОЗЕР АРКТИКИ НА ОСНОВЕ РАНДОМИЗИРОВАННОГО МАШИННОГО ОБУЧЕНИЯ

Академик РАН, проф., д.т.н. Попков Ю.С.<sup>1</sup>, проф., д.т.н. Мельников А.В.<sup>2</sup>,  
проф., д.ф.-м.н. Полищук Ю.М.<sup>2</sup>

<sup>1</sup> ФГБУН "Институт системного анализа ФИЦ ИУ РАН", Москва, 117312, Россия,

<sup>2</sup> АУ «Югорский НИИ информационных технологий», г. Ханты-Мансийск, 628011, Россия.

Проведен анализ состояния проблем моделирования пространственно-временной динамики полей в условиях современных климатических изменений. Показано, что аналитические методы исследования динамики термокарстовых процессов в отдельных озерах не пригодны для исследования пространственно-временных изменений полей термокарстовых озер. Геоимитационный метод моделирования, пригодный для исследования динамики полей термокарстовых озер, не обеспечивает достаточно высокую точность прогнозирования. Рассмотрены вопросы применения нового подхода к прогнозированию пространственно-временной динамики полей в условиях современных климатических изменений, основанного на методах и алгоритмах энтропийно-рандомизированного машинного обучения. Проведен анализ экспериментальных результатов дистанционного исследования динамики полей термокарстовых озер в арктической зоне мерзлоты. На основе данных со спутниковых снимков в зоне мерзлоты Западной Сибири выявлена тенденция сокращения площадей озер в последние десятилетия. Регрессионный анализ геокриологических и климатических данных за указанный период времени показал, что сокращение площадей озер может быть объяснено в основном повышением приземной температуры и изменением уровня осадков. Проведенные дистанционные исследования в зоне мерзлоты позволили также установить логнормальный закон распределения озер по размерам, учет которого в рамках энтропийно-рандомизированного подхода позволяет определить пространственную структуру модельных полей озер.

## 1. Введение

Особую важность рандомизированный подход представляет для решения задач прогнозирования динамики накопления парниковых газов в термокарстовых озерах арктической зоны в связи с их влиянием на глобальные климатические изменения, что может явиться основой разработки и функционирования систем адаптации к меняющимся условиям среды обитания на различных управленческих уровнях. Потепление климата в ближайшие десятилетия будет приводить к таянию мерзлых пород и дополнительному высвобождению углекислого газа, а также метана как продукта жизнедеятельности микроорганизмов, перерабатывающих оттаявшее органическое вещество. Это способно внести дополнительный ощутимый вклад в потепление климата, что вызывает озабоченность мировой общественности. Осознание этого послужило поводом для проведения Мирового Климатического саммита (Париж, 2015), принявшего рекомендации всем странам мира по разработке мер, способных не допустить повышение среднегодовой температуры Земли более чем на 2 °С до 2050 года. Разработка таких мер на региональном уровне для арктических регионов невозможна без формирования обоснованных прогнозов объемов эмиссии метана и углекислого газа, которые могут быть основаны на знаниях о пространственно-временной динамике полей озер на территории регионов.

## 2. Вопросы математического моделирования и прогнозирования полей термокарстовых озер

Рассмотрим вопросы моделирования пространственно-временной динамики полей термокарстовых озер. Обзор результатов исследования термокарстовых процессов с использованием ряда аналитических моделей, приведенный в [1], показал, что эти модели позволяют достаточно эффективно исследовать процессы в одиночных термокарстовых озерах. Использование аналитических моделей для моделирования пространственно-временной структуры полей термокарстовых озер, характерных для арктических территорий, в литературе не рассматриваются. В связи с этим определенным интерес представляют разработанные в [2] методы математической морфологии ландшафта, позволяющие проводить точный теоретико-математический анализ динамики термокарстово-озерных равнин. Методы и модели математической морфологии ландшафта могут использоваться для оценки геоэкологических рисков при развитии термокарстовых процессов в условиях продолжающегося потепления климата. Однако эти методы также не позволяют исследовать пространственно-временные изменения полей термокарстовых озер.

В [3] рассмотрены методы и средства моделирования динамики полей термокарстовых озер на основе геоимитационного подхода, позволяющие прогнозировать динамику полей термокарстовых озер и объемов накопления метана и углекислого газа в озерах Арктики. Однако разработанные средства геоимитационного моделирования использовали традиционные математические методы (в частности, регрессионный анализ данных о взаимосвязи климатических и геокриологических изменений), что не позволило обеспечить достаточно высокую точность результатов моделирования и прогнозирования динамики термокарстовых озер. Так, исследования адекватности геоимитационной модели на основе сравнения с экспериментальными данными для арктической зоны Западной Сибири показали [1], что погрешность моделирования в среднем составила около 18%. Поэтому возникает необходимость в поиске нового подхода к моделированию пространственно-временной динамики полей термокарстовых озер, который позволил бы получать более точные прогнозные оценки.

По нашему мнению, наиболее перспективным для разработки эффективной модели представляется подход, основанный на методах рандомизированного машинного обучения, которые показали высокую эффективность в решении проблем прогнозирования мировой экономики, демографии и др. [4]. Предлагаемые методы позволяют получить энтропийно-оптимальную апостериорную плотность распределения вероятностей оцениваемых параметров модели при «наихудших», в терминах энтропии, шумах измерений. Такой подход позволяет получать состоятельные и эффективные оценки параметров модели в условиях ограниченных объемов эмпирических данных.

Проблема извлечения полезной информации, под которой понимается оценивание параметров модели и прогнозирование, с использованием реальных данных, является одной из основных в современной науке. В разных научных дисциплинах разработаны многочисленные методы решения этой проблемы, основанные на определенных для каждого метода гипотезах о свойствах данных, собранных в процессе нормального функционирования исследуемого объекта. Наиболее продвинутыми в этом направлении являются математическая статистика, эконометрика, финансовая математика, теория управления, распознавание образов и другие научные дисциплины. Методы, разработанные в указанных научных дисциплинах, базируются на двух группах фундаментальных гипотез, одна из которых относится к моделям, а другая - к данным. Относительно моделей предполагается, что они имеют вполне определенные параметры (будем называть их детерминированными), значения которых неизвестны и не доступны прямому измерению. Вторая группа гипотез, касающихся данных, более существенна, так как эти гипотезы формулируются в терминах статистических свойств массивов данных, например, достаточное их количество, свойство выборки из генеральной совокупности, нормальность плотностей распределения вероятностей. Наличие этих свойств и предположений, касающихся моделей, в конкретных задачах проверить практически невозможно. Такая ситуация характерна, прежде всего, для задач, в которых объемы массивов реальных данных скудны, а сами данные содержат ошибки. Это полностью относится к задачам моделирования и прогнозирования динамики полей термокарстовых озер. В этих условиях оценивание значений характеристик (параметров) модели осуществляется по малому количеству не вполне достоверных данных. Если относиться к ним как к случайным объектам, то и оценки характеристик модели приобретают свойства случайных переменных.

Поэтому естественным образом возникает предложение рассматривать параметры модели как случайные величины. Это предложение приводит к трансформации модели с детерминированными параметрами в модель со случайными параметрами, которую будем называть рандомизированной моделью (РМ). Характеристиками РМ являются функции плотности распределения вероятностей (ПРВ) параметров. Поэтому для РМ нужно определять на основе экспериментальных данных оценки ПРВ параметров, а не оценки их значений. Имея эти оценки, можно использовать РМ для прогнозирования, если уметь генерировать последовательности случайных векторов с заданными ПРВ.

В данной работе предлагается использовать принцип максимизации информационной энтропии как меры неопределенности. Ее максимизация гарантирует получение наилучших решений при максимальной неопределенности. Поскольку информационная энтропия характеризует неопределенность, связанную не только со случайными параметрами РМ, но с шумами наблюдений, то последнее ее качество гарантирует получение наилучших оценок для максимально неопределенных (в единицах энтропии) шумов. Таким образом, получаемые в результате максимизации информационной энтропии оценки ПРВ можно трактовать как робастные.

Прогнозирование в виде рандомизированных ансамблей – новая парадигма в прогнозировании, альтернативная классическому понятию детерминированного прогноза. Рандомизированный ансамбль содержит прогнозные траектории с оптимальными оценками параметров. Здесь оптимальность понимается в энтропийных терминах. Важная проблема в рандомизированном прогнозировании – генерация последовательностей случайных векторов с заданной ПРВ (в частности, энтропийно-оптимальной ПРВ). Насколько нам известно, методология рандомизированного прогнозирования как в мире, так и в нашей стране разработана недостаточно.

Применительно к прогнозированию пространственно-временной динамики термокарстовых озер в условиях климатических изменений разработана методика рандомизированного моделирования динамики площадей озер, основанная на генерации оптимизированного ансамбля предсказательных траекторий. Предлагается динамическая модель марковского типа со случайными параметрами, характеризующая изменение размеров озер под влиянием изменений среднегодовых температур воздуха и годовой суммы осадков. Для идентификации модели разработана процедура энтропийно-рандомизированного машинного обучения, позволяющая построить оценки плотностей распределения вероятностей параметров и шумов измерений с использованием данных со спутниковых снимков.

### 3. Экспериментальные данные о пространственно-временной динамике полей термокарстовых озер по спутниковым снимкам

Информационной основой для построения рандомизированной модели динамики полей термокарстовых озер являются экспериментальные данные об изменениях численности и площадей термокарстовых озер в арктической зоне. Получение этих данных ввиду труднодоступности и высокой степени заболоченности территорий в зоне мерзлоты невозможно без применения дистанционных методов. Исследуемая территория в арктической зоне Западной Сибири имеет площадь около 1,05 млн. км<sup>2</sup>. Эта территория включает три характерные подзоны: островная (южная), прерывистая (средняя) и сплошная (северная). Озера в зоне мерзлоты характеризуются чрезвычайно широким интервалом изменений их размеров - от единиц м<sup>2</sup> до сотен тысяч гектаров. В наших дистанционных исследованиях [5-8] использовались оптические и радарные снимки среднего разрешения Landsat, ERS-2, ENVISAT, снимки высокого разрешения Канопус-В, БКА, Alos и снимки сверхвысокого разрешения Ресурс-П, QuickBird, GeoEye-1. Для минимизации влияния сезонных колебаний уровня воды в озерах снимки выбирались в достаточно короткий период летнего сезона (июль - август). В этот период полностью исчезает ледовый покров на озерах, мешающий их выделению при дешифрировании снимков. Обработка космических снимков, проведенная с использованием стандартных средств геоинформационных систем ArcGIS и ENVI, была направлена на получение данных о динамике численности и площадей озер и их пространственном распределении.

Анализ временных рядов суммарных площадей термокарстовых озёр на каждом тестовом участке позволил определить линейные тренды изменений средней площади озёр. Коэффициент линейного тренда, согласно [9], на разных ТУ за период более 40 лет, начиная с 1973 г., изменяется в пределах от – 0,28 га/год до +0,003 га/год со средним значением – 0,09 га/год, т.е в среднем в зоне мерзлоты Западной Сибири наблюдается сокращение площади озер. Так, за период исследований, начиная с 1973 г., общая площадь озер в зоне мерзлоты Западной Сибири в среднем сократилась [3,8] приблизительно на 3 и 8 % в сплошной и прерывистой подзонах мерзлоты соответственно.

На основе анализа взаимосвязи динамики площадей термокарстовых озер с изменениями среднегодовой температуры воздуха и годовой суммы осадков, проведенного в [9] методом регрессионного анализа, найдено уравнение множественной линейной регрессии в следующем виде:

$$\bar{S} = c_0 + c_1 * x_1 + c_2 * x_2 + c_3 * x_3, \quad (1)$$

где  $\bar{S}$  - усредненная по всем исследованным ТУ за определенный год площадь озер (зависимая переменная уравнения регрессии);  $c_i$  - регрессионные коэффициенты ( $i = 0, \dots, 3$ );  $x_1$  – среднегодовая температура воздуха,  $x_2$  – годовая сумма осадков,  $x_3$  – время (годы).

Сравнивая значения коэффициентов регрессии  $c_1$  и  $c_2$  (для температуры и осадков соответственно), приведенные в [9], видим, что влияние изменений среднегодовой температуры на

величину средней площади термокарстовых озер на порядок превышает влияние изменений уровня осадков. Поэтому можно сделать вывод о том, что установленное дистанционными исследованиями сокращение размеров озер в зоне мерзлоты определяется в основном повышением среднегодовой температуры, а влияние изменения уровня осадков менее значимо.

Рассмотрим результаты исследования пространственной структуры полей озер. Исследования статистики параметров местоположения озер проведены на исследуемой территории с использованием снимков Landsat. Результаты исследования законов распределения координат центров термокарстовых озёр приведены в [3]. Анализ показал, что гистограммы распределения координат центров озер на плоскости, согласно критерию  $\chi^2$ , соответствуют закону равномерной плотности с вероятностью 95 %.

Важным показателем пространственной структуры полей озер является распределение озер по размерам. Дистанционным исследованиям пространственной структуры полей озер в зонах мерзлоты различных регионов мира посвящены большое число работ, в которых отмечается, что распределение озер по размерам соответствует степенному закону. В большинстве проведенных исследований использовались космические снимки среднего разрешения Landsat (разрешение 30 м), на которых однако не обнаруживаются озера малых размеров. Такие озера обладают повышенной концентрацией метана, значительно превышающей его концентрацию в озерах больших размеров [10]. Поэтому для изучения закономерностей распределения термокарстовых озер в широком диапазоне размеров с учетом вклада малых озер были применены спутниковые снимки сверхвысокого разрешения в сочетании со снимками среднего разрешения. С использованием полученных данных о числе и площадях озер была построена гистограмма распределения площадей озер в чрезвычайно широком диапазоне изменения их размеров (от 5 до  $2 \times 10^8$  м<sup>2</sup>). Статистический анализ гистограммы распределения озер по размерам с использованием критерия Пирсона показал, что гистограмма распределения соответствует логнормальному закону с высокой вероятностью 0,99. Заметим, что оценки запасов метана и углекислого газа в малых озерах (и особенно в сверхмалых) при обычно используемом для этих целей степенном законе распределения оказываются значительно завышенными по сравнению с логнормальным законом [7].

Исследование проводилось при финансовой поддержке частично грантов Российского фонда фундаментальных исследований по проектам № 19-07-00282, № 18-45-860002, № 18-45-703001, 18-47-860013 и № 18-47-700001.

### Литература

1. Полищук В.Ю., Полищук Ю.М. Геоимитационное моделирование полей термокарстовых озер в зонах мерзлоты. Ханты-Мансийск: УИП ЮГУ, 2013. – 129 с.
2. Викторов А.С. Основные проблемы математической морфологии ландшафтов. М.: Наука, 2006. 252 с.
3. Polishchuk V.Y., Polishchuk Y.M. Modeling of thermokarst lake dynamics in West-Siberian permafrost // New York: Nova Science Publishers. 2014. № 6. P. 205-234.
4. Попков Ю.С., Попков А.Ю., Дубнов Ю.А. Рандомизированное машинное обучение при ограниченных объемах данных: От эмпирической вероятности к энтропийной рандомизации. – М.: ЛЕНАНД. – 320 с.
5. Полищук Ю.М., Богданов А.Н., Брыксина Н.А. Опыт и результаты дистанционного исследования озёр криолитозоны Западной Сибири по космическим снимкам различного разрешения за 50-летний период // Современные проблемы дистанционного зондирования Земли из космоса 2017. Т. 14. № 6. С. 42-55.
6. Полищук Ю.М., Богданов А.Н., Брыксина Н.А. Интеграция космических снимков сверхвысокого и среднего разрешения для построения гистограмм распределения площадей термокарстовых озёр в расширенном диапазоне их размеров // Современные проблемы дистанционного зондирования Земли из космоса. 2018. Т. 15. № 3. С. 38-46.
7. Polishchuk Y.M., Bogdanov A.N., Muratov I.N., Polishchuk V.Y., Lim A., Manasypov R.M., Shirokova L.S., Pokrovsky O.S. 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. 2018. Vol. 13. P. 1-16.
8. Shiklomanov A.I., Lammers R.B., Lettermaier D.P., Polishchuk Y.M., Savichev O.G., and Smith L.C. Hydrological Changes: Historical Analysis, Contemporary Status, and Future Projections. In book: Regional Environmental Changes in Siberia and Their Global Consequences. Eds.

- P.Ya.Groisman and G.Gutman, Dordrecht – Heidelberg – New-York - London: Springer, 2013, pp. 111-154.
9. Полищук В.Ю. Анализ взаимосвязи климатических изменений и термокарстовых процессов в зоне многолетней мерзлоты Западной Сибири // Оптика атмосферы и океана. 2017. Т. 30. № 03. С. 237–242.
  10. Pokrovsky O.S., Shirokova L.S., Kirpotin S.N., Audry S., Viers J., Dupre B. Effect of permafrost thawing on the organic carbon and metal speciation in thermokarst lakes of Western Siberia // Biogeosciences. 2011. vol. 8. P. 565–583.

## **METHODICAL ISSUES OF FORECASTING THE DYNAMICS OF THERMOKARST ARCTIC LAKES BASED ON RANDOMIZED MACHINE LEARNING**

Yuri S. Popkov<sup>1</sup>, Andrey V. Melnikov<sup>2</sup> and Yuri M. Polishchuk<sup>2</sup>

<sup>1</sup>Institute of System Analysis, Federal Research Centre «Informatics and Control” of RAS, Moscow, 117312, Russia

<sup>2</sup>Ugra Research Institute of Information Technology, Khanty-Mansiysk, 628011, Russia

The analysis of the state of the problems of modeling the spatio-temporal dynamics of fields under the conditions of modern climate changes is carried out. It is shown that analytical methods for studying the dynamics of thermokarst processes in individual lakes are not suitable for studying spatio-temporal changes in the fields of thermokarst lakes. The geo-simulation modeling method suitable for studying the dynamics of fields of thermokarst lakes does not provide high accuracy of forecasting. The problems of applying a new approach to the prediction of the spatio-temporal dynamics of fields under the conditions of modern climatic changes based on entropy-randomized machine learning are considered. An analysis of the experimental results of a remote study of the dynamics of thermokarst lakes fields in the Arctic permafrost is carried out. Based on data from satellite images in the permafrost of Western Siberia, a tendency toward a reduction in the lake sizes in recent decades has been revealed. Regression analysis of geocryological and climatic data for the indicated period of time showed that the reduction in the lake sizes can be explained mainly by an increase in surface temperature and changes in the level of precipitation. The remote studies in the permafrost zone also made it possible to establish the lognormal law of the lake size-distribution, taking into account the entropy-randomized approach allows us to determine the spatial structure of the model fields of lakes.