

Quantitative methods for reconstruction of the Holocene paleoclimatic characteristics and vegetation changes based on palynological data from lake and peat sediments



Novenko E.Yu.^{1*}, Andreev R.A.^{1,2}

¹ Institute of Geography of the Russian Academy of Sciences, Staromonetny lane, 29, Moscow, 119017, Russia

² Lomonosov Moscow State University, Faculty of Geography, Leninskie Gory, 1, Moscow, 119991, Russia

ABSTRACT. The paper presents the results of a methodological study of the quantitative reconstruction of paleoclimatic characteristics (mean January and July temperatures, mean annual temperature and precipitation) and forest cover within a 20 km radius of the study site using palynological data from lake and peat sediments. The following quantitative methods have been tested: the best modern analog technique, transfer functions as weighted averaging, weighted averaging partial least square, and the Random Forest method, a meta estimator that fits a number of decision tree classifiers on various sub-samples of the dataset. A database of subfossil pollen assemblages for northern central Siberia (north of 60°C) was used as a training data set for model construction, containing 174 pollen assemblages, climatic characteristics for the sampling area, and calculations of forest cover based on remote sensing data. Leave-one-out cross validation was applied for testing of the methods. The results showed that all models had the highest coefficients of determination (R^2), the smallest errors and the lowest uncertainty for the reconstruction of the mean July temperature. The developed methods were applied to two key sites in Central Siberia, located in the vicinity of Igarka and the settlement of Tura (Krasnoyarsk Region). The results obtained showed similar trends in the changes in palaeoclimatic characteristics reconstructed by different methods, but a rather large variation in temperature and precipitation values at some time intervals.

Keywords: paleoclimate, the best modern analog technique, transfer functions, Random Forest method, Central Siberia

For citation: Novenko E.Yu., Andreev R.A. Quantitative methods for reconstruction of the Holocene paleoclimatic characteristics and vegetation changes based on palynological data from lake and peat sediments // Limnology and Freshwater Biology. 2024. - № 4. - P. 544-549. DOI: [10.31951/2658-3518-2024-A-4-544](https://doi.org/10.31951/2658-3518-2024-A-4-544)

1. Introduction

Quantitative vegetation and climate reconstructions are an important part of the retrospective analysis of the natural environment. Data derived from diatom, chironomid and isotope geochemical analyses of palaeoarchives from marshes and lakes were used to calculate quantitative parameters of the paleoenvironment. Reconstructions of climate and vegetation cover based on the results of pollen analysis are commonly used. However, as numerous studies have shown, the use of a data base consisting of palynological assemblages from contrasting geographical regions negatively affects the accuracy of quantitative reconstructions. A number of researchers concluded that the

dataset should be regionally constrained and that the methods should be tested on a regional database.

The objectives of the present study are to develop and test quantitative methods for paleoclimate reconstructions from pollen data of mean July and January temperatures, mean annual temperature and precipitation, and total forest coverage within a 20 km radius of the study site for the area of northern Central Siberia.

2. Material and methods

The following quantitative reconstruction methods have been tested for reconstructing forest cover and palaeoclimate: the Best Modern Analogue Technique

*Corresponding author.

E-mail address: lenanov@mail.ru (E.Yu. Novenko)

Received: June 11, 2024; **Accepted:** July 01, 2024;

Available online: August 26, 2024

© Author(s) 2024. This work is distributed under the Creative Commons Attribution-NonCommercial 4.0 International License.



(MAT), based on a statistical search for fossil pollen assemblages of their closest modern analogues; Weighted Averaging (WA), based on the construction of a transfer function based on the relationship between the ecological requirements of taxa and a climatic conditions; Weighted Averaging Partial Least Square (WAPLS), which also constructs a transfer function but uses a different statistical method; and Random Forest (RF), which builds models based on decision trees. Table 1 shows the parameters of the models used. Calculations of climatic characteristics and testing of methods by leave-one-out validation are implemented in the form of scripts in the R programming language. The analogue package was used for the MAT (Simpson and Oksanen, 2021), for WA and WAPLS – rioja (Juggins, 2023), the RF package randomForest (Liaw and Wiener, 2003). A database of subfossil pollen assemblages for the north of Central Siberia (north of 60°N), which contains 174 sites, was used as a training dataset for constructing models. The database contains information on climatic characteristics in the form of a table for each surface sampling point (the mean temperature in July, January, the mean annual temperature, °C; average annual precipitation, mm; forest coverage within a radius of 20 km, %). The pollen assemblages of 142 surface samples were taken from the Eurasian Modern Pollen Database-2 (Davis et al., 2020). In addition, the results of earlier palynological analyses of surface samples from the Igarka, Tura, Turukhansk and Vanavara regions (32 spectra) were used. The sampling area of the surface samples contained in the database covers the vicinity of Norilsk and the Putorana plateau, the North Siberian Lowland in the Khatanga region and the Central Yakutia. Climate indicators (for the last 30 years) for these areas were taken from the database of the All-Russia Research Institute of Hydrometeorological Information - World Data Center (RIHMI-WDC) for weather stations.

The current forest coverage within a radius of 20 km around the surface sampling point is determined based on data from the WorldCover 2020 project (ESA WorldCover project, 2021). The data represent layers of landscape cover for almost the entire land area with a spatial resolution of 10 m, classified based on satellite imagery Sentinel-1 and Sentinel-2. An independent statistical and visual assessment of World Cover project

shows the high accuracy of the classification (82.1% for Asia, ESA WorldCover project, 2021). The determination of forest cover within a radius of 20 km around the surface sampling site was performed in the QGIS software version 3.16, as well as using an additional program in the R language.

The tests of the accuracy of the reconstruction methods (standard error of prediction, RMSEP) and the coefficient of determination (R^2), as well as the selection of the most appropriate parameters of the models, were carried out by leave-one-out cross validation. The results of the verification are shown in Table 1.

3. Results and discussion

According to obtained results, MAT was shown to have the highest cross-validated reconstruction accuracy, outperforming the other methods for both indicators and all parameters. The smallest errors and the smallest variance, as well as the best coefficient of determination, were achieved for the July temperature for all models. Reconstructions of the mean annual temperature show a relatively good accuracy. Annual precipitation is worse determined than the other indicators. For the calculation of T_{Jan} and T_{ann} , as well as P_{ann} , MAT and RF give better results than WA and WAPLS. The coefficient of determination (R^2) for all parameters in MAT and RF show a high proportion of the explained variance (from 74 to 87% respectively).

However, the strong difference with other models for T_{Jan} , T_{ann} and P_{ann} raises doubts about the reliability of these estimates, which requires further tests. In general, the leave-one-out cross-validation results show the ability of the models used to reliably reconstruct trends in the dynamics of the mean July temperature and forest coverage, and somewhat less reliably for T_{ann} and other characteristics.

The tested four methods of reconstruction of climatic characteristics and forest coverage were used to analyze the Holocene changes in vegetation and climate in Central Siberia. The calculations are based on fossil pollen assemblages from lacustrine and peat sediments from two peatlands located near the town of Igarka (Novenko et al., 2023) and the settlement of Tura. The results obtained showed similar trends in the changes

Table 1. Results of leave-one-out cross validation

Methods	Climatic characteristics									
	RMSEP					R^2				
	T_{Jul}	T_{Jan}	T_{ann}	P_{ann}	Forest cover	T_{Jul}	T_{Jan}	T_{ann}	P_{ann}	Forest cover
Best Modern Analogue Technique (MAT)	1,09	2,18	0,92	49,8	0,18	0,84	0,78	0,87	0,81	0,68
Weighted Averaging (WA)	1,24	3,33	1,43	79,2	0,18	0,78	0,48	0,68	0,47	0,66
Weighted Averaging Partial Least Square (WAPLS)	1,15	3,17	1,5	86,3	0,18	0,81	0,53	0,65	0,39	0,66
Random Forest (RF)	1,24	2,34	1,06	51,4	0,2	0,78	0,74	0,82	0,78	0,61

in paleoclimatic characteristics reconstructed by different methods, but a rather large variation in temperature and precipitation values at some time intervals.

With the exception of individual peaks, the best correspondence between the various methods was found for the reconstruction of the mean July temperature in, the largest range of values for precipitation. To refine the reconstruction, it is necessary to expand the dataset of surface pollen assemblages.

Acknowledgements

The studies were supported by grant of the Ministry of Science and Higher Education of Russian Federation (agreement № 075-15-2024-554 of 24.04.2024).

Conflict of interests

The authors declare that there is no conflict of interests.

References

Davis B.A.S, Chevalier M., Sommer P. et al. 2020. The Eurasian Modern Pollen Database (EMPD), version 2. *Earth System Science Data* 12 (4.): 2423–2445.

ESA WorldCover project. 2021. [Internet resource]. Access: <https://viewer.esa-worldcover.org/worldcover/> (date of access 2.02.2022)

Juggins S. 2023. rioja: Analysis of Quaternary Science Data. R package version 1.0-6. [Internet resource]. Access: <https://cran.r-project.org/package=rioja> (data of access 2.10.2023)

Liaw A., Wiener M. 2003. Classification and Regression by randomForest. *R News* 2/3: 18-22. [Internet resource]. Access: <https://CRAN.R-project.org/doc/Rnews/>. (data of access 5.06.2023)

Novenko E., Rudenko O., Mazei N. et al. 2023. Effects of Climate Change and Fire on the Middle and Late Holocene Forest History in Yenisei Siberia. *Forests* 14: 2321. [10.3390/f14122321](https://doi.org/10.3390/f14122321)

Simpson G.L., Oksanen J. 2021. analogue: Analogue and weighted averaging methods for palaeoecology. R package version 0.17-6, [Internet resource]. Access: <https://cran.r-project.org/package=analogue> (data of access 12.09.2023)

WorldCover Product Validation Report | V2.0. // ESA WorldCover. 2022. [Internet resource]. Access: https://worldcover2021.esa.int/data/docs/WorldCover_PVR_V2.0.pdf (data of access 7.10.2023)

Количественные методы реконструкций палеоклиматических характеристик и изменений растительности в голоцене по палинологическим данным из озерных и болотных отложений

Краткое сообщение

LIMNOLOGY
FRESHWATER
BIOLOGYНовенко Е.Ю.^{1*}, Андреев Р.А.^{1,2}¹ Институт географии РАН, Старомонетный переулок, д. 29, Москва, 119017, Россия² Московский государственный университет имени М.В. Ломоносова, географический факультет, Ленинские горы, д.1, Москва, 119991, Россия

АННОТАЦИЯ. В статье представлены результаты разработки количественных методов реконструкции палеоклиматических характеристик (средние температуры января и июля, среднегодовая температура, среднегодовые осадки) и доли лесопокрытой площади в радиусе 20 км (лесистости) по палинологическим данным из озерных и болотных отложений. Протестированы следующие методы количественных реконструкций: метод лучших аналогов, взвешенная нелинейная регрессия, регрессия методом наименьших частичных квадратов, и метод случайного леса (Random Forest), строящий модели на основе деревьев принятия решений. В качестве тренировочной выборки для построения моделей использована база данных субфоссильных спорово-пыльцевых спектров для севера Средней Сибири (севернее 60° с.ш.), которая содержит 174 точки, климатические характеристики для территории отбора проб и расчеты лесистости по данным дистанционного зондирования Земли. Для всех моделей наименьшие ошибки и небольшой их разброс, а также наивысшие коэффициенты детерминации получены для реконструкции средней температуры июля. Разработанные методики были применены для двух ключевых участков в Средней Сибири, расположенных в окрестностях г. Игарка и пос. Тура (Красноярский край). Полученные результаты показали сходные тренды изменений палеоклиматических характеристик, реконструированных различными методами, но довольно большой разброс значений температур и осадков на отдельных хроносрезках.

Ключевые слова: палеоклимат, лесистость, метод лучших аналогов, трансферная функция, метод случайного леса, Средняя Сибирь

Для цитирования: Новенко Е.Ю., Андреев Р.А. Количественные методы реконструкций палеоклиматических характеристик и изменений растительности в голоцене по палинологическим данным из озерных и болотных отложений // Limnology and Freshwater Biology. 2024. - № 4. - С. 544-549. DOI: 10.31951/2658-3518-2024-A-4-544

1. Введение

Количественные реконструкции характеристик растительности и климата – важное направление в ретроспективном анализе развития компонентов природной среды. Для расчёта количественных параметров среды в прошлом используют данные диатомового, хирономидного, изотопно-геохимического анализов палеоархивов болот и озёр. Широко распространены реконструкции климата и растительного покрова по данным спорово-пыльцевого анализа. Но как показали многочисленные исследования, использование палинологических спектров

из контрастных физико-географических регионов отрицательно влияет на точность количественных реконструкций. Это обусловлено изменчивостью состава спектров, приуроченных к различающимся зонально и секторно территориям. Исследователи приходят к выводу, что выборка должна быть регионально ограничена, а затем протестирована на региональной базе данных. Цель представленного исследования состояла в разработке количественных методов реконструкции палеоклиматических характеристик (средние температуры января и июля, среднегодовая температура, среднегодовые осадки) и доли лесопокрытой площади в радиусе

*Автор для переписки.

Адрес e-mail: lananov@mail.ru (Е.Ю. Новенко)

Поступила: 11 июня 2024; Принята: 01 июля 2024;

Опубликована online: 26 августа 2024

© Автор(ы) 2024. Эта работа распространяется под международной лицензией Creative Commons Attribution-NonCommercial 4.0.



20 км (лесистости) по палинологическим данным для территории севера Средней Сибири.

2. Материалы и методы

Для реконструкции лесистости и палеоклимата протестированы следующие методы количественных реконструкций: метод лучших аналогов (Modern Analogue Technique, MAT), основанный на математическом поиске для ископаемых спектров наиболее близких современных аналогов; взвешенная нелинейная регрессия (weighted averaging, WA), основанная на построении трансферной функции из расчёта экологических приуроченностей таксонов к климатическому показателю; регрессия методом наименьших частичных квадратов (weighted averaging partial least square, WAPLS), также строящая трансферную функцию, но с помощью иной статистического метода; а также метод случайного леса (Random Forest, RF), строящий модели на основе деревьев принятия решений. В таблице 1 приведены параметры использованных моделей.

Расчеты и тестирование методов путем перекрёстной проверки реализованы в виде скриптов на языке программирования R. Для MAT использован пакет analogue (Simpson and Oksanen, 2021), для WA и WAPLS – rioja (Juggins, 2023), RF – пакет randomForest (Liaw and Wiener, 2003).

В качестве тренировочной выборки для построения моделей использована база данных субфоссильных спорово-пыльцевых спектров для севера Средней Сибири (севернее 60° с.ш.), которая содержит 174 точки. В базе данных в виде таблицы для каждой точки отбора поверхностных проб приведена информация о климатических характеристиках (средняя температура июля, января, среднегодовая температура, °C; среднегодовые осадки, мм; лесистость в радиусе 20 км, %). Спорово-пыльцевые спектры 142 поверхностных проб взяты из базы данных Eurasian Modern Pollen Database-2 [Davis et al., 2020]. Помимо них использованы результаты палинологического анализа поверхностных проб из районов Игарки, Туры, Туруханска и Ванавары (32 спектра), полученные нами ранее. Территория

отбора поверхностных проб, содержащихся в базе данных, охватывает окрестности Норильска и плато Путорана, Северосибирскую низменность в районе Хатанги и Центральнуюкутскую низменность. Климатические показатели (за последние 30 лет) для данных точек взяты из базы данных МЦД-ВНИИГМИ для метеостанций Игарка, Тура, Туруханск и Ванавара.

Современная лесопокрытая площадь в радиусе 20 км вокруг точки отбора поверхностных проб определена на основе данных проекта WorldCover 2020 (ESA WorldCover project, 2021). Данные представляют собой слои ландшафтного покрова для территории почти всей суши с пространственным разрешением 10 м, классифицированные на основе съемки со спутников Sentinel-1 и Sentinel-2. Независимая статистическая и визуальная оценка WorldCover показывает высокую точность проведённой классификации (82,1% для Азии, ESA WorldCover project, 2021). Определение лесистости в радиусе 20 км вокруг места отбора поверхностных проб выполнено в ПО «QGIS» версии 3.16, а также с помощью вспомогательной программы на языке R.

Оценка точности методов реконструкции (расчёт среднеквадратической ошибки предсказания, RMSEP) и их объясняющей способности (коэффициент детерминации R^2), а также подбор наиболее подходящих параметров моделей произведены методом перекрёстной проверки с исключением. Результаты проверки приведены в таблице 1.

3. Результаты и их обсуждение

Наивысшие результаты при перекрёстной проверке показал MAT, имеющий преимущество перед другими методами по обоим показателям и для всех параметров. Для всех моделей наименьшие ошибки и небольшой их разброс, а также наивысшие коэффициенты детерминации получены для $T_{июл}$. Реконструкции среднегодовой температуры показали относительно средний результат. Годовые осадки $P_{год}$ реконструированы хуже остальных показателей. Лесистость рассчитана всеми моделями с примерно одинаковой точностью. Для

Таблица 1. Результаты перекрёстной проверки с исключением (leave-one-out cross validation)

Метод	Параметры									
	RMSEP					R2				
	$T_{июл}$	$T_{январь}$	$T_{ср.год}$	$P_{год}$	Лесистость	$T_{июл}$	$T_{январь}$	$T_{ср.год}$	$P_{год}$	Лесистость
Метод лучших аналогов (MAT)	1,09	2,18	0,92	49,8	0,18	0,84	0,78	0,87	0,81	0,68
Взвешенная нелинейная регрессия (WA)	1,24	3,33	1,43	79,2	0,18	0,78	0,48	0,68	0,47	0,66
Регрессия методом наименьших частичных квадратов (WAPLS)	1,15	3,17	1,5	86,3	0,18	0,81	0,53	0,65	0,39	0,66
Случайный лес (RF)	1,24	2,34	1,06	51,4	0,2	0,78	0,74	0,82	0,78	0,61

показателей $T_{\text{янв.}}$ и $T_{\text{ср. год}}$, а также $R_{\text{год}}$ MAT и RF показывают результаты лучше, чем WA и WAPLS. Коэффициенты детерминации для всех параметров у MAT и RF показывают высокую долю объяснённой дисперсии (от 74 до 87%). Однако сильная разница с другими моделями для $T_{\text{янв.}}$, $T_{\text{ср. год}}$ и $R_{\text{год}}$ вызывает сомнения в достоверности данных оценок, что требует дальнейших тестов. В целом, результаты проверки отражают способность использованных моделей надёжно отражать тренды динамики средней температуры июля и лесистости и несколько менее надёжно для $T_{\text{ср. год}}$.

Протестированные 4 метода реконструкции климатических характеристик и лесистости применены для анализа изменений растительности и климата в Средней Сибири в голоцене. Расчеты приведены по ископаемым спорово-пыльцевым спектрам из разрезов торфяных болот: крупнобугристого болота в районе Игарки (Novenko et al., 2023) и болота Горное (Тура). Полученные результаты показали сходные тренды изменений палеоклиматических характеристик, но довольно большой разброс значений температур и осадков на отдельных хроносрезях. За исключением, отдельных пиков, наилучшее соответствие между различными методами выявлено для средней температуры июля, самый большой разброс значений – для осадков. Для уточнения реконструкции необходимо расширять базу поверхностных спорово-пыльцевых спектров.

Благодарности

Исследование выполнено при поддержке гранта Минобрнауки РФ (Соглашение № 075-15-2024-554 от 24.04.2024).

Конфликт интересов

Авторы заявляют об отсутствии конфликта интересов.

Список литературы

Davis B.A.S, Chevalier M., Sommer P. et al. 2020. The Eurasian Modern Pollen Database (EMPD), version 2. *Earth System Science Data* 12 (4.): 2423–2445.

ESA WorldCover project. 2021. [Internet resource]. Access: <https://viewer.esa-worldcover.org/worldcover/> (дата обращения 2.02.2022)

Juggins S. 2023. rioja: Analysis of Quaternary Science Data. R package version 1.0-6. [Internet resource]. Access: <https://cran.r-project.org/package=rioja> (дата обращения 2.10.2023)

Liaw A., Wiener M. 2003. Classification and Regression by randomForest. *R News* 2/3: 18-22. [Internet resource]. Access: <https://CRAN.R-project.org/doc/Rnews/>. (дата обращения 5.06.2023)

Novenko E., Rudenko O., Mazei N. et al. 2023. Effects of Climate Change and Fire on the Middle and Late Holocene Forest History in Yenisei Siberia. *Forests* 14: 2321. [10.3390/f14122321](https://doi.org/10.3390/f14122321)

Simpson G.L., Oksanen J. 2021. analogue: Analogue and weighted averaging methods for palaeoecology. R package version 0.17-6, [Internet resource]. Access: <https://cran.r-project.org/package=analogue> (дата обращения 12.09.2023)

WorldCover Product Validation Report | V2.0. // ESA WorldCover. 2022. [Internet resource]. Access: https://worldcover2021.esa.int/data/docs/WorldCover_PVR_V2.0.pdf (дата обращения 7.10.2023)