

# АРХЕОЛОГИЯ ВОЛГО-ДОНСКОГО МЕЖДУРЕЧЬЯ

УДК 902.2(470.61):912.648

DOI: 10.58529/2782-6511-2023-2-1-6-23

## **Сарычев Дмитрий Владимирович**

старший преподаватель кафедры геоэкологии и мониторинга окружающей среды, факультет географии, геоэкологии и туризма, Воронежский государственный университет (Россия, Воронеж)

ORCID: 0000-0002-3755-0108

E-mail: sarychev.geo@gmail.com

## **Земцов Григорий Леонидович**

кандидат исторических наук, доцент кафедры отечественной и всеобщей истории института истории, права и общественных наук Липецкого государственного педагогического университета имени П. П. Семёнова-Тян-Шанского (Россия, Липецк)

ORCID: 0000-0002-7177-7242

E-mail: grizem@rambler.ru

## **Предиктивное археологическое моделирование методом максимальной энтропии (на примере поселений поздне римского времени в бассейне Верхнего Дона)\***

**АННОТАЦИЯ.** В статье рассмотрена методика предиктивного моделирования потенциальных местоположений археологических памятников с помощью геоинформационных технологий и машинного обучения на основе метода максимальной энтропии (MaxEnt). Объектом моделирования являлись места поселений поздне римского времени типа Каширки-Седелок, существовавшие в бассейне Верхнего Дона с середины III по начало IV в. н. э. В целях исследования была подготовлена географическая база данных, включающая сведения о локализации 169 известных археологических памятников данного типа в исследуемом регионе — координаты их местоположений послужили для обучения и валидации предиктивной модели. В качестве предикторов использовались ГИС-слои, созданные на основе тематических карт и цифровой модели рельефа, прямо или косвенно отражающие палеогеографические условия ландшафта, в частности рельеф, подстилающие породы, климат, гидрографию и почвенный покров бассейна Верхнего Дона в исследуемый период. В результате созданная модель проанализировала более 7 миллионов участков размером 90 x 90 м общей площадью 59,4 тыс. км<sup>2</sup> на предмет их потенциальной пригодности для размещения поселений типа Каширки-Седелок и выявила среди них 29 860 га наиболее пригодных участков (менее 1 % от исследуемой территории). По тестовой выборке археологических памятников была показана высокая статистическая надежность созданной модели (AUC = 0,915) и ее прогностическая эффективность для археологических исследований (Прирост Квамме KG = 0,97). Результаты проведенного моделирования используются для планирования и проведения полевых археологических исследований в регионе.

**КЛЮЧЕВЫЕ СЛОВА:** предиктивное археологическое моделирование, АРМ, максимальная энтропия, MaxEnt, поздне римское время, археологические памятники типа Каширки-Седелок, ГИС

\* Исследование выполнено при финансовой поддержке Всероссийской общественной организации «Русское географическое общество», проект № 14/2022-Р «Поиск раннеславянских археологических памятников в бассейне Дона с помощью геоинформационных технологий и искусственного интеллекта».

# ARCHAEOLOGY OF THE VOLGA-DON INTERFLUVE REGION

UDC 902.2(470.61):912.648

DOI: 10.58529/2782-6511-2023-2-1-6-23

## **Dmitry V. Sarychev**

Voronezh State University (Russia, Voronezh)

ORCID: 0000-0002-3755-0108

E-mail: sarychev.geo@gmail.com

## **Grigory L. Zemtsov**

Candidate of Historical Sciences, Lipetsk State Pedagogical University (Russia, Lipetsk)

ORCID: 0000-0002-7177-7242

E-mail: grizem@rambler.ru

## **Predictive Archaeological Modeling with MaxEnt (Late Roman Settlements in The Upper Don Basin Case)**

**ABSTRACT.** The paper is devoted to an archaeological predictive modeling technique with GIS and machine learning based on maximal entropy algorithm (MaxEnt). Archaeological sites of the late roman time presented by settlements of Kashirka-Sedelok type (mid-3rd — early 4th century AD) in the Upper Don basin were the modeling objects. We found out that there are 169 known sites of such type in the study region, and collected all available information on their locations into a geodatabase including our own field survey data. The coordinates of these sites split into training and testing subsets served for developing a predictive model with MaxEnt. Environmental GIS-layers based on a digital elevation model and thematic maps were the model predictors. We made such predictors to reflect the paleogeography of the Upper Don basin in the study time period by choosing relatively stable landscape features of relief, bedrock, climate, hydrography and soils of the area. The trained model analyzed through more than 7 million cells of 90x90 m with overall area of 59 400 sq. km to identify potentially suitable areas for the Kashirka-Sedelok settlements and found 29 860 ha the most suitable areas (less than 1 % of the study area) as the result. The efficiency of the conducted modeling, evaluated on the testing sites, is high based on metrics of receiver operating characteristic (AUC = 0,915) and Kvamme's Gain (KG = 0,97). The results of modeling serve as an agenda for the upcoming archaeological surveys in the region.

**KEYWORDS:** archaeological predictive modeling, APM, maximal entropy, MaxEnt, late roman time, archaeological sites of the Kashirka-Sedelok type, GIS

## ВВЕДЕНИЕ

Геоинформационные системы, дистанционное зондирование и машинное обучение становятся все более важными инструментами в археологии. Одним из приоритетных направлений их использования является предиктивное археологическое моделирование, то есть выявление потенциальных мест расположения древних памятников<sup>1</sup>. Предполагается, что размещение археологических пунктов было детерминировано географическими условиями, некоторые из них остались неизменными, изменились слабо или известным образом. Если такие условия формализуемы в виде растровых или векторных геоинформационных слоев (карт), то с их помощью можно выявлять закономерности размещения исследуемых памятников и вероятные местоположения еще не открытых объектов. Такое предиктивное моделирование можно выполнять различными методами, от простого визуального анализа картографического материала до технологий искусственного интеллекта<sup>2</sup>.

Большинство подобных исследований посвящены изучению определенных территорий, на которых анализируются памятники сразу нескольких эпох или культур<sup>3</sup>. В результате такого моделирования выявляемая зона потенциального расположения разновременных археологических объектов, как правило, оказывается слишком обширной и неоднородной. В свете возникновения более прогрессивных методов<sup>4</sup> и накопления более детальных археологических данных возникает возможность эффективного применения предиктивного моделирования для изучения размещения археологических памятников в узких пространственно-временных срезах. В данной статье мы подробнее раскрываем первый для отечественной археологии опыт предиктивного моделирования методом максимальной энтропии<sup>5</sup>: в новой работе расширена территория исследования, использованы новые археологические и геоинформационные данные, улучшен методический аппарат моделирования.

Исследование выполнено на примере выявления поселений середины III — начала IV в. н. э. — памятников типа Каширки-Седелок — в бассейне Верхнего Дона. Это показательный объект для моделирования: данные памятники представлены исключительно селищами с очень тонким культурным слоем, кратковременность поселений позволяет рассматривать их как узкохронологический феномен, а их картографию — как практически одномоментный срез политической географии древности. В настоящее время на исследуемой территории обнаружено более 150 памятников типа Каширки-Седелок. Их систематические раскопки не проводились, а были связаны преимущественно с охранно-спасательными работами в зонах хозяйственного освоения<sup>6</sup>. Однако характер размещения поселений позднееримского времени на территориях, изученных в ходе сплошных археологических разведок, позволяет предположить, что значительное количество поселений типа Каширки-Седелок еще не обнаружено в бассейне Верхнего Дона.

<sup>1</sup> Judge W. J., Martin D. W., Sebastian L. Quantifying the present and predicting the past: theory, method, and application of archaeological predictive modeling. Washington, D. C., 1989; GIS and archaeological site location modeling. Boca Raton, 2006.

<sup>2</sup> Глебова А. Б., Сергеев И. С., Быков Н. И. Геоинформационный анализ с построением модели нейронной сети для прогноза размещения археологических объектов в ландшафтах Юго-Восточного Алтая // Вестник Санкт-Петербургского университета. Науки о Земле. 2022. Т. 67, № 3. С. 491–514.

<sup>3</sup> Зайцева Е. А. Прогнозирование расположения объектов археологии с применением ГИС и ДЗЗ // Труды IV (XX) всероссийского археологического съезда в Казани: сборник научных статей. В 5 т. Казань, 2014. Т. IV. С. 399–401; Создание предиктивной модели для поиска археологических объектов на Соловецком архипелаге / Мартынов А. И. [и др.] // Геодезия, картография, геоинформатика и кадастры. От идеи до внедрения: сборник научных статей. СПб, 2017. С. 448–455.

<sup>4</sup> Predictive modeling for archaeological site locations: Comparing logistic regression and maximal entropy in north Israel and north-east China / Wachtel I. [et al.] // Journal of Archaeological Science. 2018. No. 92. P. 28–36.

<sup>5</sup> Сарычев Д. В., Земцов Г. Л., Фабрициус Е. В. Моделирование размещения археологических памятников методом максимальной энтропии (на примере поселений позднееримского времени Липецкой области) // Археология и геоинформатика. Четвертая Международная конференция. Тезисы докладов. М., 2019. С. 85–86; Сарычев Д. В., Земцов Г. Л., Фабрициус Е. В. Выявление археологических памятников позднееримского времени в Липецкой области методом максимальной энтропии // Геоинформационное картографирование в регионах России: сборник научных статей. Воронеж, 2020. С. 302–308.

<sup>6</sup> Земцов Г. Л., Сарычев Д. В., Фабрициус Е. В. Картография поселений позднееримского времени в бассейне реки Быстрая Сосна // Гуманитарные исследования Центральной России. 2019. № 4 (13). С. 7–10; Население Верхнего Подонья в позднееримское время: жертва или агрессор? / Земцов Г. Л. [и др.] // Stratum plus. Археология и культурная антропология. 2019. № 4. С. 257–264; Микрорегион поселений «Плющань» позднееримского времени на территории Верхнего Подонья / Земцов Г. Л. [и др.] // Российская археология. 2020. № 2. С. 60–71.

Таким образом, цель данного исследования состояла в выявлении наиболее вероятных местоположений поселений типа Каширки-Седелок в бассейне Верхнего Дона с помощью геоинформационных технологий и машинного обучения методом максимальной энтропии на основе известных локаций памятников данного типа. Для этого решались следующие задачи:

1. Создание географической базы данных известных археологических памятников типа Каширки-Седелок в бассейне Верхнего Дона.
2. Подготовка цифровых карт (растровых ГИС-слоев), прямо или косвенно характеризующих палеогеографические условия исследуемой территории.
3. Моделирование с помощью алгоритма MaxEnt потенциального размещения поселений типа Каширки-Седелок в бассейне Верхнего Дона.

Исследование имеет прикладную направленность и призвано сфокусировать поисковые археологические работы по выявлению поселений поздне римского времени на наиболее вероятные локации. Информационно-методическая основа работы способствует более полному изучению и сохранению историко-археологического наследия исследуемого региона и может быть масштабирована на иные территории и археологические культуры.

## МАТЕРИАЛЫ И МЕТОДЫ

### *Теоретические основы метода максимальной энтропии*

Метод максимальной энтропии (алгоритм MaxEnt) в анализе географического распространения объектов — это технология машинного обучения, основанная на принципе минимизации относительной энтропии в пространстве предикторов между плотностью распределения по точкам объекта и плотностью, рассчитанной по точкам фона<sup>7</sup>. Такой подход может служить для прогнозирования размещения археологических памятников<sup>8</sup>, более того, показано превосходство MaxEnt над логистической регрессией, традиционно используемой для подобного рода задач,<sup>9</sup> а также над алгоритмами GLM, GAM и Random Forest<sup>10</sup>. Для отечественной практики предиктивного археологического моделирования метод максимальной энтропии является новым<sup>11</sup>, поэтому считаем целесообразным описать его подробнее.

Суть алгоритма MaxEnt сводится к сравнению математических ожиданий для переменных-предикторов с учетом допущения, что любая система естественным образом стремится к состоянию наивысшей энтропии. Математическое ожидание от переменной (например, такой, как «высота местоположения» или «расстояние до водотока» и др.) соотносится с понятием ее среднего арифметического. Первый шаг к определению вероятностного распространения исследуемых объектов заключается в сравнении математических ожиданий всех переменных. Если обозначить распределение точек известных памятников как  $p$ , а любой предиктор (например, расстояние до ближайшего водотока от каждой точки) — как  $x$ , тогда, соответственно,  $M p(x)$  — есть математическое ожидание от  $x$  для  $p$ . Пусть  $P$  — это искомое реальное распределение, включающее также еще не выявленные памятники, тогда для каждой переменной  $x$  вероятно следующее уравнение:

$$M p(x) = M P(x).$$

Данное условие может выполняться для бесконечного множества вариантов теоретически возможного распределения памятников. Для того, чтобы выбрать одно, наиболее вероятное

<sup>7</sup> A statistical explanation of MaxEnt for ecologists / Elith J. [et al.] // Diversity and Distributions. 2011. No. 17. P. 43–57; Phillips S. J., Anderson R. P., Schapire R. E. Maximum entropy modeling of species geographic distributions // Ecological Modeling. 2006. Vol. 190/3-4. P. 231–259; Phillips S. J., Dudík M. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation // Ecography. 2008. Vol. 31, iss. 2. P. 161–175.

<sup>8</sup> Maxent Modeling of Ancient and Modern Agricultural Terraces in the Troodos Foothills, Cyprus / Galletti C. S. [et al.] // Applied Geography. 2013. No. 39. P. 46–56; Howey M. C., Palace M. W., McMichael C. H. Geospatial modeling approach to monument construction using Michigan from AD 1000–1600 as a case study // Proceedings of the National Academy of Sciences. 2016. Vol. 113, iss. 27. P. 7443–7448; Banks W. E. The application of ecological niche modeling methods to archaeological data in order to examine culture-environment relationships and cultural trajectories // Quaternaire. 2017. Vol. 28, no. 2. P. 271–276; Rafuse D. J. A Maxent Predictive Model for Hunter-Gatherer Sites in the Southern Pampas, Argentina // Open Quaternary. 2021. Vol. 7., no. 6. P. 1–21.

<sup>9</sup> Predictive modeling for archaeological site locations...

<sup>10</sup> Advancing predictive modeling in archaeology: An evaluation of regression and machine learning methods on the Grand Staircase-Escalante National Monument / Yaworsky P. M. [et al.] // PLoS ONE. 2020. Vol. 15, iss. 10. P. 1–22.

<sup>11</sup> Сарычев Д. В., Земцов Г. Л., Фабрициус Е. В. Моделирование размещения археологических памятников...

из них, применяется принцип максимальной энтропии. Так, если  $y$  — свободная переменная, выражающая размещение точек (памятников), тогда информационную энтропию ( $H$ ) этой системы можно описать следующей формулой:

$$H(P) = - \sum P(y) \ln P(y).$$

Следовательно, в соответствии с принципом максимальной энтропии, из всех теоретически подходящих распределений следует выбрать одно, с наиболее высоким значением информационной энтропии.

Представленная математическая концепция энтропии отражает степень «неупорядоченности», хаотичности в системе. Основное допущение состоит в том, что все явления естественным образом стремятся к беспорядку, если не прилагать соответствующих усилий по их упорядочению. В контексте археологии таким усилием можно считать то, что места для поселений люди выбирали не случайным образом. При моделировании мы учитываем это обстоятельство путем выбора такого потенциального распределения, математическое ожидание которого совпадает с ожиданием от распределения известных памятников. А посредством выбора распределения с максимальной энтропией мы сводим к случайности все ненаправленные влияния, то есть выбираем то потенциальное распределение, которое при прочих равных условиях было бы наиболее вероятным в естественной системе.

Таким образом, чтобы применить метод максимальной энтропии для выявления размещения археологических памятников при анализе по пикселям растровых карт, необходимо выполнить следующие этапы: 1) рассчитать распределение  $p$  по пикселям, в которых отмечены искомые памятники; 2) вычислить математические ожидания для  $p$  в отношении всех рассматриваемых параметров  $x$  (предикторов); 3) определить набор всех распределений с математическими ожиданиями, как и у  $p$ , и из этого набора выявить распределение с максимальной информационной энтропией.

Пусть  $n$  — это число всех пикселей на карте. Тогда для каждого пикселя имеем:  $p(y) = 1/n$ , если в данном пикселе отмечен памятник и  $p(y) = 0$  — если памятник отсутствует. Важно, что сумма всех значений  $p$  будет составлять ровно единицу. На втором этапе мы рассчитываем математическое ожидание для каждого параметра, причем числитель дроби для категориальных параметров получается суммированием значений в пикселях, а для непрерывных — путем интегральных вычислений. Третий этап можно выполнить с помощью метода максимального подобия (maximal likelihood), при котором искомое распределение с максимальной энтропией выявляется в ходе итеративного процесса аппроксимаций<sup>12</sup>.

В результате моделирования по данному алгоритму мы получаем длинную последовательность очень маленьких чисел, которую принято переводить в логарифмическую шкалу и визуализировать в виде растровых карт с количественным фоном или градиентной заливкой. По таким картам удобно определять места с высокой и низкой вероятностью нахождения памятников и исследовать общие закономерности их размещения на рассматриваемой территории.

Более детальное описание параметров и работы алгоритма MaxEnt можно найти в специальных статьях<sup>13</sup>. Однако из вышеприведенного описания метода можно сделать простой вывод, что для моделирования размещения археологических объектов алгоритму MaxEnt необходима обучающая выборка известных местонахождений исследуемых объектов и растровые слои-предикторы, характеризующие географические свойства территории, на которой необходимо выявить потенциальные места поселений.

#### *Данные об объекте моделирования*

Моделирование в данном исследовании проводили для территории бассейна Верхнего Дона — северная лесостепная часть бассейна реки Дон, чаще всего выделяемая от его истока до устья реки Воронеж<sup>14</sup>. Площадь исследуемой территории составила 59,4 тыс. км<sup>2</sup> (см. *цв. вклейку, рис. 1*).

<sup>12</sup> Predictive modeling for archaeological site locations...

<sup>13</sup> Phillips S. J., Anderson R. P., Schapire R. E. Op. cit.; Phillips S. J., Dudík M. Op. cit.; Лисовский А. А., Дудов С. В. Преимущества и ограничения использования методов экологического моделирования ареалов. 2. MaxEnt // Журнал общей биологии. 2020. Т. 81, № 2. С. 135–146.

<sup>14</sup> Долина Дона: природа и ландшафты. Воронеж, 1982.

В середине III — начале IV в. н. э. в бассейне Верхнего Дона существовало население, в материалах которого переплетались черты киевской и черняховской археологических культур. Поселения этой общности были объединены А. М. Обломским в группу археологических памятников типа Каширки-Седелок<sup>15</sup>. Они были распространены от верхнего течения р. Красивая Меча (Тульская область) на севере до р. Девицы (Воронежская область) на юге. Большая часть выявленных памятников этого типа сосредоточена в западной части бассейна Верхнего Дона, на территории современной Липецкой области.

Известные памятники типа Каширки-Седелок были неукрепленными поселками, состоявшими из 3–7 «усадеб», включавших жилые и хозяйственные постройки. Поселки занимали в среднем площадь до 2 га, часто были вытянуты на несколько сотен метров вдоль ближайшего водотока. Тонкий культурный слой, оставленный этими поселениями, свидетельствует об их кратковременном существовании<sup>16</sup>.

Памятники типа Каширки-Седелок характеризуются широким спектром ландшафтных условий их местонахождений<sup>17</sup> и отсутствием четких индикаторов на местности, что обуславливает дополнительную сложность при попытке их предиктивного моделирования. Наиболее типичным местоположением памятников этого типа можно считать сниженные мысовидные участки надпойменных террас малых рек и суходолов, очевидно, обводненных во времена существования поселений, с доступом к более удобным для земледелия угодьям. Отмечается приуроченность большинства рассматриваемых памятников к ручьям и малым рекам второго — четвертого порядков с дистанцией между поселениями вдоль водотоков в 5–7 км; в меньшем количестве и с более разреженным расположением известны памятники этого времени на берегах крупных рек — Дона, Быстрой Сосны и Красивой Мечи<sup>18</sup>.

В результате анализа археологических отчетов и публикаций нами были собраны сведения о местоположении 169 известных памятников типа Каширки-Седелок в бассейне Верхнего Дона (см. *цв. вклейку, рис. 1*). Эти сведения были внесены в специально подготовленную географическую базу данных. Каждый памятник обозначен в виде полигона и точки (условного центра поселения). Все пункты разделены на 4 класса по точности определения их географических координат: 1 — «точно», установлено по ГНСС-навигатору, проверено авторами на местности (51 памятник); 2 — «точно», координаты взяты из отчетов, без проверки на местности (27 памятников); 3 — «приблизительно», по описаниям местоположения из отчетов, без проверки на местности (76 памятников); 4 — «не точно», местоположение в отчете описано неясно (15 памятников). Для обучения предиктивной модели использовали выборку 78 памятников, координаты которых определены с наибольшей точностью (классы точности 1 и 2), остальные точки с менее надежной географической привязкой — 91 памятник (классы точности 3 и 4) — использовали в целях тестирования модели.

Неравномерность в распределении известных памятников, связанная с отличиями в методах и объемах полевых археологических обследований, может приводить к пространственной автокорреляции данных, искажающей результаты моделирования<sup>19</sup>. Для коррекции этого явления в модели применяли специальный растр предвзятости выборки (bias grid). Он был создан согласно методическим рекомендациям<sup>20</sup>, путем оценки плотности ядер (kernel density) в радиусе 5 километров от каждого памятника. Выбор такой дистанции опирается на данные о плотности размещения поселений типа Каширки-Седелок в исследуемом регионе. Фоновые точки для моделирования, или так называемые точки «псевдоотсутствия» поселений, в количестве 25 тысяч штук, также были взяты в обозначенной пятикилометровой зоне рандомно, но при этом вне точек и полигонов известных памятников.

Необходимо отметить, что количество учтенных археологических памятников типа Каширки-Седелок в регионе превышает количество выявленных здесь поселений данного типа, так как части некоторых поселений (как правило, «усадьбы»), были открыты и зарегистрированы как

<sup>15</sup> Обломский А. М. О ритмах развития лесостепного Поднепровья и Подонья в позднеримское и гуннское время // Археология Центрального Черноземья и сопредельных территорий: сборник научных статей. Липецк, 1999. С. 127–134.

<sup>16</sup> Население Верхнего Подонья в позднеримское время...; Восточная Европа в середине I тысячелетия н. э. М., 2007. С. 73–74. (Раннеславянский мир. Археология славян и их соседей: Сер. науч. публикаций; Вып. 9.)

<sup>17</sup> Восточная Европа в середине I тысячелетия н. э. С. 80.

<sup>18</sup> Земцов Г. Л., Сарычев Д. В., Фабрициус Е. В. Указ. соч.

<sup>19</sup> The importance of correcting for sampling bias in MaxEnt species distribution models / Kramer-Schadt S. [et al.] // Diversity Distribution. 2013. Vol. 19, iss. 11. P. 1366–1379.

<sup>20</sup> A statistical explanation of MaxEnt...; Phillips S. J. A Brief Tutorial on Maxent. 2017. Available at: [http://biodiversityinformatics.amnh.org/open\\_source/maxent/](http://biodiversityinformatics.amnh.org/open_source/maxent/) (accessed: 10.01.2023).

отдельные археологические памятники в разное время. Поэтому для моделирования использовали не все точки: в случае попадания двух или более точек в одну ячейку раstra (пиксель) — засчитывали их как одну точку.

#### *Подготовка слоев-предикторов*

Как следует из вышеприведенного описания территории и объекта моделирования, от поселений типа Каширки-Седелок не осталось каких-либо заметных следов в современном рельефе. Их расположение, как и у большинства археологических памятников, связано с гидрографической сетью, но других явных закономерностей в размещении не прослеживается. В связи с этим при попытке предиктивного моделирования локаций данного типа приходится опираться на их возможные связи с параметрами рельефа, подстилающих геологических пород, гидрографии, свойствами почв и климата, которые не изменились за прошедшее время, изменились слабо или предсказуемым образом. Если такие параметры удастся формализовать в виде карт — растровых геоинформационных слоев с пространственным разрешением, достаточным для дифференциации мест поселений и фоновых территорий, то эти слои могут быть использованы в качестве предикторов для модели.

Конкретный набор слоев-предикторов в каждом случае моделирования формируется с учетом специфики географической зоны, исследуемой археологической культуры и ее локаций, доступности и качества пространственных данных, а также особенностей алгоритма моделирования. Так, в горах морфометрические параметры рельефа — абсолютная высота, уклон, экспозиция и др. — будут более значимы для размещения поселений, чем на равнинах. Для моделирования стоянок палеолитических культур, в отличие от более поздних, существенным будет доступность сырья для каменных орудий, то есть геологический фактор. Карты палеоландшафтных реконструкций с высоким пространственно-временным разрешением — редкость, поэтому предикторами чаще служат цифровые модели рельефа (ЦМР) и их производные. Наконец, метод моделирования также определяет состав предикторов, в частности, многие алгоритмы чувствительны к эффекту мультиколлинеарности, что требует отсеивания коррелирующих слоев.

Для отсеивания несущественных предикторов и преодоления мультиколлинеарности в MaxEnt встроен хорошо зарекомендовавший себя метод L1-регуляризации<sup>21</sup>. Его улучшение за счет предварительного отбора предикторов дополнительными математическими способами маловероятно<sup>22</sup>, поэтому для MaxEnt нет необходимости заранее исключать по одному предиктору из каждой коррелирующей пары или выделять главные компоненты факторным анализом, что является популярными методами для устранения мультиколлинеарности. Такие математические процедуры необходимы в том случае, если по результатам моделирования планируется анализ вклада факторов в модель, — в MaxEnt он выполняется эвристическими методами, которые чувствительны к взаимно коррелирующим предикторам. Независимо от применяемых математических подходов, подчеркивается важность изначального экспертного отбора релевантных предикторов<sup>23</sup>.

Таким образом, учитывая особенности территории, объекта и метода моделирования, а также обеспеченность пространственными данными, мы подготовили 23 геоинформационных слоя-предиктора, прямо или косвенно отражающих палеогеографические условия бассейна Верхнего Дона. При этом геоморфологические и гидрографические предикторы созданы на основе цифровой модели рельефа MERIT DEM<sup>24</sup>; геологические предикторы — на основе карт четвертичных отложений масштаба 1:500 000; климатические предикторы — по палеоклиматическим данным PaleoClim<sup>25</sup>; почвенные предикторы — по данным SoilGrids<sup>26</sup> (таблица).

<sup>21</sup> Hastie T., Tibshirani R., Friedman J. H. The elements of statistical learning: data mining, inference, and prediction. New York, 2009.

<sup>22</sup> Modelling and predicting fungal distribution patterns using herbarium data / Wollan A. K. [et al.] // Journal of Biogeography. 2008. Vol. 35, iss. 12. P. 2298–2310.

<sup>23</sup> Elith J., Leathwick J. R. The contribution of species distribution modelling to conservation prioritization // Spatial Conservation Prioritization: Quantitative Methods & Computational Tools. Oxford, 2009. P. 70–93.

<sup>24</sup> A high accuracy map of global terrain elevations / Yamazaki D. [et al.] // Geophysical Research Letters. 2017. Vol. 44, iss. 11. P. 5844–5853.

<sup>25</sup> PaleoClim, high spatial resolution paleoclimate surfaces for global land areas / Brown J. [et al.] // Scientific Data. 2018. Vol. 5, iss. 1. Article number: 180254.

<sup>26</sup> SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty / Poggio L. [et al.] // Soil. 2021. Vol. 7, iss. 1. P. 217–240.

Таблица

**Сведения об исходных пространственных данных и производных геоинформационных слоях-предикторах (\* – жирным шрифтом выделены категориальные предикторы)**

Исходные данные, их пространственное разрешение / масштаб	Названия слоев-предикторов, единицы измерения	Краткое обозначение в модели*
Цифровая модель рельефа MERIT DEM. 3 угловых сек. (~ 90 м/пиксель)	Абсолютная высота, м над у. м.	elev
	Уклон местности, °	slope
	Экспозиция склона	<b>aspect</b>
	Профильная кривизна	profsc
	Индекс топографической позиции	TPI
	Топографический индекс влажности	TWI
	Индекс тепловой нагрузки	HLI
	Индекс ветровой экспозиции	WEI
	Форма рельефа	<b>geomorphon</b>
	Близость водотока, м	strprox
	Близость места слияния водотоков, м	confprox
	Высота над ближайшим водотоком, м	REM
Геологические карты Липецкой, Курской, Орловской, Тульской, Рязанской, Тамбовской и Воронежской областей, М 1:500 000	Генетический тип четвертичных отложений	<b>Q</b>
	Близость выходов дочетвертичных отложений, м	D
Палеоклиматические данные PaleoClim, Anthropocene v1.2b. 30 угловых сек. (~1 км / пиксель)	Среднегодовая температура, °С	bio_1
	Температурная сезонность	bio_4
	Годовая амплитуда температур, °С	bio_7
	Годовые осадки, мм/год	bio_12
	Сезонность осадков	bio_15
Данные о почвенном покрове SoilGrids. 250 м/пиксель	Доля глинистых частиц (< 0,002 мм) в мелкоземной фракции, %	clay
	Доля частиц песка (> 0,05 мм) во фракции мелкозема, %	sand
	Доля илистых частиц (≥ 0,002 мм и ≤ 0,05 мм) в мелкоземной фракции, %	silt
	Содержание органического углерода почвы в мелкоземной фракции, г/кг	soc

В качестве основы для создания большинства предикторов была выбрана цифровая модель рельефа MERIT DEM ([http://hydro.iis.u-tokyo.ac.jp/~yamada/MERIT\\_DEM](http://hydro.iis.u-tokyo.ac.jp/~yamada/MERIT_DEM)) с пространственным разрешением около 90 м/пиксель. Более детальные открытые ЦМР с разрешением около 30 м/пиксель, доступные на исследуемый регион, в частности SRTM DEM V3, ASTER GDEM V3, AW3D30, COPERNICUS GLO-30 и FABDEM, также были проанализированы на предмет их использования в исследовании, однако они оказались менее пригодными вследствие больших искажений высот и форм рельефа под участками древесно-кустарниковой растительности или артефактов

от коррекции таких участков. Дополнительным преимуществом ЦМР MERIT является ее корректность для гидрологического моделирования<sup>27</sup>.

Тайлы «n50e035» и «n50e040» ЦМР MERIT были собраны в единое покрытие и обрезаны по охвату исследуемой территории, полученный растр спроецирован в прямоугольную систему координат WGS 84 / UTM zone 37N с применением билинейной интерполяции для пересчета ячеек — в таком виде ЦМР использовали как слой-предиктор «абсолютная высота», а также на ее основе рассчитали производные растры: «уклон местности» (slope), «экспозиция склона» (aspect), «профильная кривизна» (profc), «индекс топографической позиции» (TPI), «формы рельефа» (geomorphon), «топографический индекс влажности» (TWI), «индекс тепловой обеспеченности» (HLI) и «индекс ветровой экспозиции» (WEI). Посредством таких предикторов сделана попытка учесть в модели потенциально значимые для размещения поселений топографические и микроклиматические условия местности, определяемые главным образом рельефом и потому относительно стабильные во времени. При этом, если абсолютная высота, уклон и экспозиция уже традиционно используются в археологическом моделировании<sup>28</sup> и достаточно понятны, то остальные предикторы важно пояснить.

Растры уклона и экспозиции рассчитывали в градусах (значения экспозиции затем переводили в классы по восьми румбам) на основе алгоритма Зевенбергера-Торна<sup>29</sup>. Тот же алгоритм использовался при расчете *профильной кривизны* (Profile curvature — «profc») — показателя геометрической кривизны поверхности относительно нормальной линии уклона. Положительные значения кривизны профиля присваиваются пикселям, приходящимся на выпуклые части склонов; отрицательные значения — на вогнутые участки склонов; близкие к нулю — на субгоризонтальные поверхности. По картам профильной кривизны хорошо дифференцируются речные террасы, что важно в данном исследовании.

Визуально похожим показателем, часто служащим для простой классификации рельефа по формам, является *индекс топографической позиции* (topographic position index — TPI). В отличие от профильной кривизны, TPI основан на сопоставлении высоты в каждой ячейке ЦМР со средней высотой соседних ячеек в заданном радиусе, но интерпретация значений индекса схожа с кривизной: положительный TPI показывает места с преобладающими высотами над окружением, например, холмы; отрицательный — низменности по отношению к окрестностям, например, долины, а околонулевые значения — равнины или области с изменчивым уклоном<sup>30</sup>. В исследовании использовали мультимасштабную версию TPI.

Более сложную классификацию *форм рельефа* получили с помощью выделения так называемых «геоморфонов» (geomorphon). Этот метод основан на алгоритме компьютерного зрения, который сегментирует ЦМР на 10 классов: низменности, долины, подошвы склонов, балки, склоны, бровки склонов, отроги, хребты, вершины и плоскоместья<sup>31</sup>. Подобные автоматически генерируемые карты форм рельефа успешно применялись в археологическом предиктивном моделировании<sup>32</sup>.

*Топографический индекс влажности* (topographic wetness index — TWI), также известный как составной топографический индекс (CTI), является функцией от уклона местности и вышележащей водосборной площади<sup>33</sup>. TWI показывает потенциал скопления влаги на различных участках в зависимости от окружающих форм рельефа, тем самым «подсвечивая» территории, предрасположенные к переувлажнению грунтов или подтоплениям и потому менее пригодные для размещения поселений.

<sup>27</sup> A high accuracy map of global terrain elevations.

<sup>28</sup> Judge W. J., Martin D. W., Sebastian L. Op. cit.

<sup>29</sup> Zevenbergen L. W., Thorne C. R. Quantitative analysis of land surface topography // *Earth Surface Processes and Landforms*. 1987. Vol. 12, iss. 1. P. 47–56.

<sup>30</sup> Guisan A., Weiss S. B., Weiss A. D. GLM versus CCA spatial modeling of plant species distribution // *Plant Ecology*. 1999. No. 143. P. 107–122.

<sup>31</sup> Stepinski T., Jasiewicz J. Geomorphons — a new approach to classification of landform // *Proceedings of Geomorphometry, Redlands*, 2011. P. 109–112.

<sup>32</sup> Verhagen P., Drăguț L. Object-based landform delineation and classification from DEMs for archaeological predictive mapping // *Journal of Archaeological Science*. 2012. Vol. 39, no. 3. P. 698–703.

<sup>33</sup> Terrain attributes: estimation methods and scale effects / Moore I. D. [et al.] // *Modeling Change in Environmental Systems*. London, 1993. P. 189–214.

*Индекс тепловой обеспеченности* (heat load index — HLI) — показатель, взвешивающий экспозицию и уклон так, что результирующие максимальные значения соответствуют юго-западным, а минимальные — северо-восточным склонам, для соответствия замеченному факту, что юго-западные склоны теплее даже юго-восточных при одинаковой инсоляции<sup>34</sup>. Отмечена более высокая информативность HLI в сравнении с показателями уклона и экспозиции для моделирования археологических локаций Кукутени в Румынии<sup>35</sup>.

*Индекс ветровой экспозиции* (wind exposition index — WEI) — показывает усредненный эффект от ветров всех направлений через заданный интервал азимутов (15°, по умолчанию). Это безразмерный индекс: значения выше 1 показывают продуваемые участки, ниже 1 — подветренные укрытия<sup>36</sup>. На равнинах в зимний период влияние ветра, как было показано<sup>37</sup>, может являться важным параметром для размещения поселений.

Описанные предикторы, за исключением HLI, рассчитывали с помощью соответствующих инструментов в SAGA GIS 7.8.2, используя значения «по умолчанию» для всех не упомянутых выше опций настроек. Для расчета HLI, не представленного на сегодня в SAGA, использовали специальную функцию из набора инструментов Gradient&Geomorphometric Modeling 2.0 для ArcGIS Pro.

Также по ЦМР MERIT с помощью инструментов гидрологического моделирования в ArcGIS Pro 3.0.3 был построен векторный слой палеогидрографической сети: многие памятники типа Каширки-Седелок выявлены по суходолам, но во времена существования этих поселений, очевидно, в близлежащих руслах текли ручьи или реки. Эмпирическим путем, по морфометрии бассейнов таких суходолов, было установлено, что для соответствия картографии известных памятников достаточные водотоки должны были формироваться при вышележащей водосборной площади приблизительно от 300 га и более — такой порог был установлен для генерации сети «палеоводотоков». В наши дни эти бассейны дренируются, как правило, временными потоками, то есть использование современной гидрографической сети в модели подразумевало бы, что многие из поселений типа Каширки-Седелок не имели непосредственного доступа к проточной воде. Таким образом, на основе полученного слоя палеогидрографической сети, который в целом является более разветвленной версией современной гидрографической сети, были рассчитаны производные растровые слои-предикторы: «близость водотока» (strprox), «близость места слияния водотоков» (confprox) и «высота над ближайшим водотоком» (REM).

В дополнение к описанным предикторам, все же являющимся связанными одной исходной ЦМР, предпринята попытка учесть также геологические, климатические и почвенные факторы по альтернативным источникам данных. Так, ареалы подстилающих пород на бассейн Верхнего Дона были оцифрованы по геологическим картам четвертичных отложений Липецкой, Курской, Орловской, Тульской, Рязанской, Тамбовской и Воронежской областей масштаба 1:500 000, изданным в 1998 году Межрегиональным центром по геологической картографии Министерства природных ресурсов Российской Федерации. На основе полученного в результате оцифровки векторного слоя сгенерировано два предиктора: категориальный растр «генетический тип четвертичных отложений» (Q) и континуальный растр «близость выходов дочетвертичных отложений» (D) по ячейкам, соответствующим ЦМР.

Возможное влияние климатических факторов на размещение исследуемых поселений учитывали в модели на основе палеоклиматических данных PaleoClim (<http://www.paleoclim.org/>). Они предоставляют по 19 биоклиматических переменных на 12 временных периодов от приблизительно 3,3 млн лет назад до современного периода в пространственном разрешении от 0,5

<sup>34</sup> McCune B., Keon D. Equations for potential annual direct incident radiation and heat load // Journal of Vegetation Science. 2002. Vol. 13, iss. 4. P. 603–606.

<sup>35</sup> Nicu I. C., Mihu-Pintilie A., Williamson J. GIS-Based and Statistical Approaches in Archaeological Predictive Modelling (NE Romania) // Sustainability. 2019. Vol. 11, iss. 21. P. 1–13.

<sup>36</sup> Gerlitz L., Conrad O., Böhner J. Large scale atmospheric forcing and topographic modification of precipitation rates over High Asia — a neural network based approach // Earth System Dynamics. 2015. No. 6. P. 1–21.

<sup>37</sup> Чечушков И. В., Валиахметов И. А. «Ветер крепчает»: моделирование скорости ветра при формировании объяснительных моделей системы расселения // Археология и геоинформатика: электронное издание. 2021. Вып. 10. Режим доступа: <https://www.archaeolog.ru/media/periodicals/agis/AGIS-10/Chechushkov/page1.html> (дата обращения: 10.01.2023).

до 10 угловых минут<sup>38</sup>. Мы использовали данные за временной период «Anthropocene, v1.2b», по сути, характеризующие климат на основе скорректированных инструментальных наблюдений за последние полвека. Было выбрано 5 основных биоклиматических переменных в максимально доступном разрешении 30 угловых секунд (~ 1 км/пиксель): среднегодовая температура (bio\_1), температурная сезонность (bio\_4), годовая амплитуда температур (bio\_7), годовые осадки (bio\_12) и сезонность осадков (bio\_15).

Неоднородности почвенного покрова и возможную приуроченность к ним древних поселений попытались учесть с помощью данных SoilGrids (<https://soilgrids.org/>). Сервис предоставляет картографические модели для 11 свойств почв, включая параметры механического состава, содержания углерода, азота, водородный показатель, емкость катионного обмена. При этом прогнозные карты даны с пространственным разрешением 250 м/пиксель на шесть интервалов глубины: 0–5, 5–15, 15–30, 30–60, 60–100 и 100–200 см<sup>39</sup>. Из обозначенных параметров выбрали, вероятно, наиболее стабильные во времени показатели: доля глинистых частиц (clay), частиц песка (sand) и илистых частиц (silt) в мелкоземной фракции, а также содержание органического углерода почвы в мелкоземной фракции (soc). Использовали показатели для слоя 30–60 см, т. к. он наиболее соответствует средней глубине концентрации находок эпохи бронзы и раннего железного века в исследуемом регионе.

Исходные значения в растрах выбранных слоев климатических и почвенных предикторов были переведены в стандартные единицы измерений путем умножения на коэффициенты пересчета, указанные в спецификациях данных. Все слои-предикторы были приведены к единому пространственному охвату и разрешению посредством обрезки по границе исследуемой территории и пересчету по сети ячеек, соответствующей основе ЦМР, что позволило создать однородный набор растров-предикторов. Подготовку и анализ пространственных данных, использованных в исследовании, а также все картографические работы осуществляли с примирением геоинформационных сред QGIS 3.26, SAGA GIS 7.8.2 и ArcGIS Pro 3.0.3.

#### *Настройки модели и оценка эффективности моделирования*

Моделирование по методу максимальной энтропии выполняли в программе MaxEnt версии 3.4.4.<sup>40</sup> Входными данными служили подготовленные слои-предикторы (табл.) и обучающие выборки известных локализаций археологических памятников типа Каширки-Седелок. В качестве мест «псевдоотсутствия» использовались 25 000 случайных фоновых точек в 5-километровой буферной зоне от точек и полигонов известных памятников.

В целях отладки моделей выполняли их многократное воспроизведение с бутстрэп-генерацией новых обучающих и тестовых выборок (bootstrap replication). Такой подход оптимален при моделировании на основе малых выборок местонахождений<sup>41</sup>. Итоговые модели рассчитывались итеративно по достижению уровня сходимости  $10^{-5}$  (convergence threshold) или ограничения в 500 итераций. Наилучшую с технической точки зрения комбинацию настроек модель подбирали по минимальному скорректированному информационному критерию Акаике (AICc) с помощью ENMeval 2.0.4 в среде R<sup>42</sup>. В частности, определяли наилучший коэффициент регуляризации (regularization multiplier) и комбинацию числовых признаков (feature types) — линейных (linear), квадратичных (quadratic), их производных (product) и нелинейных (hinge) — для аппроксимации связей с предикторами<sup>43</sup>.

Оценку прогностической способности моделей проводили по стандартному критерию AUC (площадь под ROC-кривой). Данный непараметрический показатель отражает надежность

<sup>38</sup> PaleoClim, high spatial resolution paleoclimate surfaces...

<sup>39</sup> SoilGrids 2.0: producing soil information...

<sup>40</sup> Phillips S. J., Dudik M., Schapire R. E. Maxent software for modeling species niches and distributions (Version 3.4.4). Available at: [http://biodiversityinformatics.amnh.org/open\\_source/maxent/](http://biodiversityinformatics.amnh.org/open_source/maxent/) (accessed: 10.01.2023).

<sup>41</sup> Phillips S. J. Op. cit.

<sup>42</sup> ENMeval 2.0: redesigned for customizable and reproducible modeling of species' niches and distributions / Kass J. M. [et al.] // *Methods in Ecology and Evolution*. 2021. Vol. 12, iss. 9. P. 1602–1608.

<sup>43</sup> Merow C., Smith M. J., Silander J. A. A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter // *Ecography*. 2013. Vol. 36, iss. 10. P. 1058–1069.

модели в распознавании точек анализируемой выборки как «присутствие» или «фон» для моделируемого объекта. AUC измеряется в диапазоне от 0 до 1, при этом значения выше 0.7 говорят о приемлемом, 0.8 — хорошем и 0.9 — отличном качестве моделирования<sup>44</sup>. На основе обучающего и тестового наборов точек рассчитывали, соответственно,  $AUC_{train}$  и  $AUC_{test}$ . Последний показатель использовали как основной, однако они оба не являются достаточными, ввиду взаимосвязи из-за применения бутстрэппинга и недостатков самого подхода AUC<sup>45</sup>.

Дополнительную оценку качества моделей проводили по непрерывному индексу Бойса (CBI) в ENMeval 2.0.4, который специально разработан для случаев моделирования без реальных точек «отсутствия»<sup>46</sup>. По сути, CBI является коэффициентом ранговой корреляции Спирмена между долей местонахождений объекта на участках с заданным диапазоном спрогнозированной пригодности и ожидаемой долей местонахождений, в соответствии с долей площади этих участков от исследуемой территории<sup>47</sup>. Чем ближе значения CBI к +1, тем лучше смоделированное распределение согласуется с распределением тестовых местонахождений; чем ближе к -1 — тем вероятнее модель прогнозирует отсутствие объекта там, где он на самом деле отмечен; близкие к нулю значения индекса говорят о работе прогноза на уровне случайности.

Значимость предикторов в построении модели определяли по показателю «важность при пермутации». Он состоит в том, что поочередно для каждого предиктора случайным образом искажают значения в ячейках раstra для местонахождений памятников и точек фона, затем моделирование повторяют с таким «мутировавшим» предиктором и вычисляют, насколько из-за него падает AUC модели. Величина этого падения в процентах и есть рассматриваемый показатель, чем он выше — тем больше вклад соответствующего предиктора в модель. Значение параметров среды для размещения поселений также исследовали методом «складного ножа» (jackknife test), при котором моделирование дополнительно повторяют в двух вариантах для каждого предиктора: только по нему и, наоборот — по всем остальным, без его участия. Сравнение получаемых пар моделей, например, по приросту их объяснительной способности, позволяет оценить важность предикторов для итоговой модели<sup>48</sup>.

Результаты моделирования выводили в формате «Cloglog» (комплементарная лог-лог модель), при котором пиксели выходного раstra выражают потенциальную пригодность местности для размещения поселений типа Каширки-Седелок в континуальной вероятностной шкале от 0 до 1. Пересчет этих значений в бинарную форму, с целью выделения условно пригодных территорий и расчета метрик эффективности моделирования, выполняли по порогу вероятности, выше которого 90 % тестовых точек верно распознавалось моделью — то есть критерий 10-го перцентиля, часто используемый в подобном моделировании<sup>49</sup>.

На основе полученной бинарной формы прогноза потенциальных мест расположения исследуемых памятников рассчитали основную метрику эффективности предиктивного археологического моделирования — «Прирост Квамме» (Kvamme's Gain — KG), которая сопоставляет долю выделяемой моделью пригодной площади к доле попадающих в эту площадь археологических памятников из общего числа памятников на исследуемой территории. Соответственно, чем ближе значения KG к +1, тем меньше выделяемая моделью площадь пригодных территорий и большее количество точек в нее попадает; 0 и близкие к нему значения показателя характеризуют работу модели на уровне случайности; отрицательные значения KG говорят о том, что модель прогнозирует места без искомым объектов<sup>50</sup>.

<sup>44</sup> Baldwin R. A. Use of Maximum Entropy Modeling in Wildlife Research // Entropy. 2009. Vol. 11, no. 4. P. 854–866.

<sup>45</sup> Lobo J. M., Jimenez-Valverde A., Real R. AUC: a misleading measure of the performance of predictive distribution models // Global Ecology & Biogeography. 2008. Vol. 17, iss. 2. P. 145–151.

<sup>46</sup> Evaluating the ability of habitat suitability models to predict species presences / Hirzel A. H. [et al.] // Ecological Modeling. 2006. Vol. 199, iss. 2. P. 142–152.

<sup>47</sup> Evaluating resource selection functions / Boyce M. S. [et al.] // Ecological Modeling. 2002. No. 157. P. 281–300.

<sup>48</sup> Phillips S. J., Anderson R. P., Schapire R. E. Op. cit., Phillips S. J., Dudík M. Op. cit.

<sup>49</sup> Liu C., White M., Newell G. Selecting thresholds for the prediction of species occurrence with presence-only data // Journal of Biogeography. 2013. Vol. 40, no. 4, pp. 778–789.

<sup>50</sup> Judge W. J., Martin D. W., Sebastian L. Op. cit. P. 329.

### РЕЗУЛЬТАТЫ МОДЕЛИРОВАНИЯ

С помощью метода максимальной энтропии мы создали модель потенциального размещения археологических памятников позднеримского времени типа Каширки-Седелок в бассейне Верхнего Дона. Оценка полученной модели по критерию AUC составила 0,969 по обучающей выборке точек и 0,915 — по тестовой выборке точек, что свидетельствует о высокой статистической надежности модели. Аналогично высокие статистические показатели эффективности проведенного моделирования получены по непрерывному индексу Бойса (CBI = 0,89) и приросту Квамме (KG = 0,97).

В качестве основного результата моделирования был получен растровый слой на всю территорию бассейна Верхнего Дона с разрешением около 90 м/пиксель. Для каждой ячейки этого слоя на основе совокупного вклада использованных предикторов был рассчитан процент вероятной пригодности для размещения поселений типа Каширки-Седелок. Методом 10-го перцентиля мы получили пороговое значение на уровне 69 % для бинаризации непрерывного растра пригодности, то есть в бинарной модели пиксели с вероятностью пригодности от 70 % и выше считали пригодными участками для размещения поселений рассматриваемого времени, а до 69 % включительно — непригодными территориями. При этом 10 % из известных поселений типа Каширки-Седелок, фактически существовавших, но при таком подходе попадающих в условно непригодные зоны, считали нетипичными поселениями. Соответственно, мы выделили около 4 тысяч отдельных потенциально пригодных участков размерами приблизительно от 1 га до 2 тыс. га и общей площадью 29 862 га, что составляет менее 1 % от исследуемой территории (см. *цв. вклейку, рис. 2*).

Выявленные места потенциального расположения поселений сосредоточены преимущественно по долинам рек Дона, Красивой Мечи, Сосны, Сновы, а также небольших правобережных притоков реки Воронеж. Часть из этих территорий, например, расположенных в долине р. Дон, севернее впадения в него Красивой Мечи, а также по рекам Снова, Репец, Олым и Пальна, представляются перспективными для проведения специальных разведочных работ.

Оценка использованных в моделировании предикторов методами «складного ножа» и «важности при пермутации» показывает, что наибольший вклад в модель обеспечили суммарно слои «близость водотока», «высота над ближайшим водотоком» и «генетический тип четвертичных отложений», при этом не наблюдается выраженного преобладания вклада хотя бы одного из предикторов. Необходимо отметить, что мы намеренно не приводим количественные оценки вклада предикторов в модель, так как эти показатели носят условный характер, с учетом использования в анализе большого количества заведомо коррелирующих между собой слоев<sup>51</sup>.

### ОБСУЖДЕНИЕ

Проведенное моделирование в целом показывает существование в регионе достаточного количества мест, похожих на локации известных поселений типа Каширки-Седелок, что согласуется с результатами предыдущих работ<sup>52</sup>. Однако существенная часть выявленных потенциально пригодных территорий не была заселена вследствие ряда возможных причин, среди которых недостаток популяционного ресурса, пропуск пригодных мест вследствие необходимого рассредоточения поселений, табуированность мест из-за религиозных или суеверных воззрений, контроль территорий со стороны конкурирующих групп населения и др. В связи с подобными причинами следует быть сдержанными в ожиданиях к результатам предиктивного археологического моделирования.

Кратковременность существования поселений типа Каширки-Седелок и их неукрепленность позволяют предположить, что люди не рассматривали эти селища как окончательные, а могли перемещаться в поиске более удачных мест. Вероятно, освоение исследуемой территории этими культурными группами происходило с запада на восток по рекам бассейна Быстрой Сосны, а за-

<sup>51</sup> A statistical explanation of MaxEnt...; Phillips S. J., Dudík M. Op. cit.

<sup>52</sup> Сарычев Д. В., Земцов Г. Л., Фабрициус Е. В. Выявление археологических памятников...

тем на север и юг вдоль Дона и его притоков, при этом заселение потенциально пригодных мест могло идти в порядке очереди, что также не учитывается в модели.

Часть территорий отнесена моделью к пригодным ошибочно из-за недоучета каких-либо лимитирующих факторов, погрешностей входных данных и самого метода моделирования. Так, недостатком моделирования в MaxEnt и подобных пиксельно-ориентированных подходах можно считать существенную изменчивость результатов моделирования в зависимости от выбираемого уровня дискретности анализа (размера пикселя); а также недостаточный учет контекста, то есть окружения и взаиморасположения объектов местности.

Проведенное моделирование, несмотря на высокие статистические показатели качества, заведомо не учитывает многие важные параметры, например, биотические факторы — состояние и типы растительности, охотничье-промысловых ресурсов и т. п. — во времена существования искомым объектов. Поскольку с момента исчезновения в регионе поселений типа Каширки-Седелок прошло более полутора тысяч лет, мы имеем возможность включить в модель только такие параметры, которые возможно реконструировать посредством палеогеографических моделей или использовать косвенные предикторы, что и было продемонстрировано в данном исследовании. В любом случае модель будет несовершенна из-за невозможности учесть в ней важные, но уже недоступные параметры среды. Несмотря на это, выполненное моделирование несет прогностическую ценность и помогает в планировании и проведении полевых археологических исследований, она может итеративно улучшаться за счет обнаружения новых локаций поселений типа Каширки-Седелок и появления новых слоев-предикторов с лучшим пространственным и временным разрешением.

В настоящее время по результатам моделирования планируются и проводятся полевые разведочные археологические работы для сбора фактического материала и полевой верификации модели. Проверены единичные участки, есть находки исследуемого времени, однако количества обследованных локаций пока недостаточно для получения статистически значимых оценок. Эта работа продолжается, и по мере накопления сведений планируются дополнительные научные публикации по результатам полевой проверки модели.

В ходе разведочных работ ученый-археолог полагается не только на знания о ландшафтном расположении памятников — немаловажную роль в этом процессе играет интуиция исследователя, которую пока нельзя полностью заменить формальными методами моделирования. Позитивистский подход по созданию универсальных законов и механизации человеческого фактора, на наш взгляд, не является состоятельным, что особенно касается наук, изучающих человека. В то же время объединение геоинформационных технологий, дистанционного зондирования, методов машинного обучения и широкого набора пространственных данных обеспечивает эффективный познавательный и прогностический аппарат, способный существенно дополнять данные классических методов археологии.

## **ЗАКЛЮЧЕНИЕ**

В данном исследовании мы детально рассмотрели методику предиктивного археологического моделирования на основе алгоритма MaxEnt, впервые для российской археологии, и проиллюстрировали ее применение на примере поиска возможных местоположений поселений типа Каширки-Седелок по явным и скрытым факторам размещения памятников этого типа. При этом была создана географическая база данных, включившая информацию о 169 археологических памятниках типа Каширки-Седелок, известных в бассейне Верхнего Дона. В целях работы нами также подготовлен полезный для дальнейших археолого-географических исследований региона набор из 23 геоинформационных слоев-предикторов, характеризующих с высоким пространственным разрешением палеогеографические условия, в том числе геолого-геоморфологическое строение, климат, гидрографию и почвенный покров бассейна Верхнего Дона. На основе этих данных с помощью MaxEnt в бассейне Верхнего Дона выявлено около 300 км<sup>2</sup> (менее 1 % от исследуемой территории) наиболее пригодных территорий для размещения поселений типа Каширки-Седелок, перспективных для проведения поисковых археологических работ. Статистические оценки качества модели, выполненные по метрикам AUC, CBI и Kvatme's Gain, показали высокую надежность проведенного моделирования.

Дальнейшая полевая проверка результатов моделирования позволит оперативно корректировать разработанную модель и повышать ее прогностический потенциал. Исследование выполнено с применением оригинальных методических подходов, обеспечивает информационную основу для планирования поисковых археологических работ и в конечном счете служит более полному выявлению, изучению и сохранению историко-археологического наследия региона.

#### БЛАГОДАРНОСТИ

Авторы благодарят Фабрициус Е. В. и Гончарова В. О. за помощь при подготовке археологических данных; доцента кафедры исторической геологии и палеонтологии Воронежского государственного университета, к. г.-м. н. Д. А. Иванова за помощь в подготовке и интерпретации геологических данных; Н. И. Лозбенева, научного сотрудника Почвенного института им. В. В. Докучаева, научного сотрудника ИПЭЭ им. А. Н. Северцова РАН, к. б. н. Цурикова С. М. и директора Ботанического сада ВГУ им. проф. Б. М. Козо-Полянского, к. с.-х. н. Воронина А. А. за помощь в интерпретации почвенных данных; к. г. н. Нестерова Ю. А., к. г. н. Горбунова А. С. и Корецкого Н. А. за консультации по геоморфометрическим параметрам; а также редакцию Историко-географического журнала в лице главного редактора, к. и. н. Михеева М. В., за помощь в подготовке и улучшении рукописи статьи.

#### References

Baldwin R. A. Use of Maximum Entropy Modeling in Wildlife Research. *Entropy*, 2009, vol. 11, no. 4, pp. 854–866. DOI: 10.3390/e11040854 (In English).

Banks W. E. The Application of Ecological Niche Modeling Methods to Archaeological Data in Order to Examine Culture-Environment Relationships and Cultural Trajectories. *Quaternaire*, 2017, vol. 28, no. 2, pp. 271–276. DOI: 10.4000/quaternaire.7966 (In English).

Boyce M. S., Vernier P. R., Nielsen S. E., Schmiegelow F. A. Evaluating Resource Selection Functions. *Ecological Modeling*, 2002, no. 157, pp. 281–300. DOI: 10.1016/S0304-3800(02)00200-4 (In English).

Brown J., Hill D., Dolan A., Carnaval A. C., Haywood A. M. PaleoClim, High Spatial Resolution Paleoclimate Surfaces for Global Land Areas. *Scientific Data*, 2018, vol. 5, iss. 1, article number: 180254. DOI: 10.1038/sdata.2018.254 (In English).

Chechushkov I. V., Valiakhmetov I. A. [“The Wind Is Getting Stronger”: Modeling Wind Speed in The Formation of Explanatory Models of Settlement Systems]. *Arkheologiya i geoinformatika* [Archaeology and geoinformatics], 2021, vol. 10. Available at: <https://www.archaeolog.ru/media/periodicals/agis/AGIS-10/Chechushkov/page1.html> (accessed: 10.01.2023). DOI: 10.25681/IARAS.2021.978-5-94375-342-8.21 (In Russian).

*Dolina Dona: priroda i landshafty* [Don Valley: Nature and Landscapes]. Voronezh: Tsentr-Chernozemnoye kn. izd-vo Publ., 1982. (In Russian).

Elith J., Leathwick J. R. The Contribution of Species Distribution Modelling to Conservation Prioritization. *Spatial Conservation Prioritization: Quantitative Methods & Computational Tools*. Oxford: Oxford University Press, 2009, pp. 70–93. (In English).

Elith J., Phillips S. J., Hastie T., Dudik M., Chee Y. E., Yates C. J. A Statistical Explanation of MaxEnt for Ecologists. *Diversity and Distributions*, 2011, no. 17, pp. 43–57. DOI: 10.1111/j.1472-4642.2010.00725.x (In English).

Galletti C. S., Ridder E., Falconer S. E., Fall P. L. Maxent Modeling of Ancient and Modern Agricultural Terraces in the Troodos Foothills, Cyprus. *Applied Geography*, 2013, no. 39, pp. 46–56. DOI: 10.1016/j.apgeog.2012.11.020 (In English).

Gerlitz L., Conrad O., Böhner J. Large Scale Atmospheric Forcing and Topographic Modification of Precipitation Rates Over High Asia — A Neural Network Based Approach. *Earth System Dynamics*, 2015, no. 6, pp. 1–21. (In English).

*GIS and Archaeological Site Location Modeling*. Boca Raton: CRC Press Taylor & Francis Group, 2006. (In English).

Glebova A. B., Sergeev I. S., Bykov N. I. [Geoinformation Analysis with the Construction of a Neural Network Model for Predicting Location of Archaeological Monuments in the Landscapes of South-

Eastern Altai]. *Vestnik Sankt-Peterburgskogo universiteta. Nauki o Zemle* [Vestnik of Saint Petersburg University. Earth Sciences], 2022, vol. 67, no. 3, pp. 491–514. DOI: 10.21638/spbu07.2022.306 (In Russian).

Guisan A., Weiss S. B., Weiss A. D. GLM Versus CCA Spatial Modeling of Plant Species Distribution. *Plant Ecology*, 1999, no. 143, pp. 107–122. DOI: <https://doi.org/10.1023/A:1009841519580> (In English).

Hastie T., Tibshirani R., Friedman J. H. *The Elements of Statistical Learning: Data Mining, Inference, And Prediction*. 2<sup>nd</sup> ed. New York: Springer-Verlag, 2009. (In English).

Hirzel A. H., Le Lay G., Helfer V., Randon C., Guisan A. Evaluating the Ability of Habitat Suitability Models to Predict Species Presences. *Ecological Modeling*, 2006, vol. 199, iss. 2, pp. 142–152. DOI: 10.1016/j.ecolmodel.2006.05.017 (In English).

Howey M. C., Palace M. W., McMichael C. H. Geospatial Modeling Approach to Monument Construction Using Michigan from AD 1000–1600 as a Case Study. *Proceedings of the National Academy of Sciences*, 2016, vol. 113, iss. 27, pp. 7443–7448. DOI: 10.1073/pnas.1603450113 (In English).

Judge W. J., Martin D. W., Sebastian L. *Quantifying the Present and Predicting the Past: Theory, Method, and Application of Archaeological Predictive Modeling*. Washington, D. C.: Bureau of Land Management, 1989. (In English).

Kass J. M., Muscarella R., Galante P. J. et al. ENMeval 2.0: Redesigning for Customizable and Reproducible Modeling of Species' Niches and Distributions. *Methods in Ecology and Evolution*, 2021, vol. 12, iss. 9, pp. 1602–1608. DOI: 10.1111/2041-210x.13628 (In English).

Kramer-Schadt S., Niedballa J., Pilgrim J. D. et al. The Importance of Correcting for Sampling Bias in MaxEnt Species Distribution Models. *Diversity Distribution*, 2013, vol. 19, iss. 11, pp. 1366–1379. DOI: 10.1111/ddi.12096 (In English).

Lissovsky A. A., Dudov S. V. [Advantages and Limitations of Application of the Species Distribution Modeling Methods. 2. MaxEnt]. *Zhurnal obshchey biologii* [Journal of General Biology], 2020, vol. 81, no. 2, pp. 135–146. DOI: 10.31857/S0044459620020049 (In Russian).

Liu C., White M., Newell G. Selecting Thresholds for the Prediction of Species Occurrence with Presence-Only Data. *Journal of Biogeography*, 2013, vol. 40, no. 4, pp. 778–789. (In English).

Lobo J. M., Jimenez-Valverde A., Real R. AUC: A Misleading Measure of the Performance of Predictive Distribution Models. *Global Ecology & Biogeography*, 2008, vol. 17, iss. 2, pp. 145–151. DOI: 10.1111/j.1466-8238.2007.00358.x (In English).

Martynov A. I., Medvedev A. A., Alekseenko N. A., Koshutin R. A. [Creation of a Predictive Model for the Search for Archaeological Objects in the Solovetsky Archipelago]. *Geodeziya, kartografiya, geoinformatika i kadastry. Ot idei do vnedreniya: sb. nauch. statey* [Geodesy, Cartography, Geoinformatics and Cadastres. From Idea to Implementation: Collection of Sci. Articles]. Saint Petersburg: Politekhnik Publ., 2017, pp. 448–455. (In Russian).

McCune B., Keon D. Equations for Potential Annual Direct Incident Radiation and Heat Load. *Journal of Vegetation Science*, 2002, vol. 13, iss. 4, pp. 603–606. DOI: 10.1111/j.1654-1103.2002.tb02087.x (In English).

Merow C., Smith M. J., Silander J. A. A Practical Guide to MaxEnt for Modeling Species' Distributions: What It Does, and Why Inputs and Settings Matter. *Ecography*, 2013, vol. 36, iss. 10, pp. 1058–1069. DOI: 10.1111/j.1600-0587.2013.07872.x (In English).

Moore I. D., Gessler P. E., Nielsen G. A., Petersen G. A. Terrain Attributes: Estimation Methods and Scale Effects. *Modeling Change in Environmental Systems*. London: Wiley, 1993, pp. 189–214. (In English).

Nicu I. C., Mihu-Pintilie A., Williamson J. GIS-Based and Statistical Approaches in Archaeological Predictive Modelling (NE Romania). *Sustainability*, 2019, vol. 11, iss. 21, pp. 1–13. DOI: 10.3390/su11215969 (In English).

Oblomsky A. M. [On the Rhythms of Development of the Forest-Steppe Dnieper and Don Region in the Late Roman and Hun Times]. *Arkheologiya Tsentral'nogo Chernozem'ya i sopredel'nykh territoriy: sb. nauch. statey* [Archaeology of the Central Chernozem Region and Adjacent Territories: A Collection of Sci. Articles]. Lipetsk: Izd-vo Voronezh. ped. in-ta Publ., 1999, pp. 127–134. (In Russian).

Phillips S. J., Anderson R. P., Schapire R. E. Maximum Entropy Modeling of Species Geographic Distributions. *Ecological Modelling*, 2006, vol. 190, iss. 3–4, pp. 231–259. DOI: 10.1016/j.ecolmodel.2005.03.026 (In English).

Phillips S. J., Dudík M. Modeling of Species Distributions with Maxent: New Extensions and a Comprehensive Evaluation. *Ecography*, 2008, vol. 31, iss. 2, pp. 161–175. DOI: 10.1111/j.0906-7590.2008.5203.x (In English).

Poggio L., Sousa L. M., Batjes N. H., Heuvelink G. M., Kempen B., Ribeiro E., Rossiter D. SoilGrids 2.0: Producing Soil Information for the Globe with Quantified Spatial Uncertainty. *Soil*, 2021, vol. 7, iss. 1, pp. 217–240. DOI: 10.5194/soil-7-217-2021 (In English).

Rafuse D. J. A Maxent Predictive Model for Hunter-Gatherer Sites in the Southern Pampas, Argentina. *Open Quaternary*, 2021, vol. 7, no. 6, pp. 1–21. DOI: 10.5334/oq.97 (In English).

Sarychev D. V., Zemtsov G. L., Fabritsius E. V. [Identification of Archaeological Sites of the Late Roman Period in the Lipetsk Region Using the Maximum Entropy Method]. *Geoinformatsionnoye kartografirovaniye v regionakh Rossii: sb. nauch. statey* [Geoinformation Mapping in the Regions of Russia: A Collection of Sci. Articles]. Voronezh: Tsifrovaya poligrafiya Publ., 2020, pp. 302–308. (In Russian).

Sarychev D. V., Zemtsov G. L., Fabritsius E. V. [Modeling the Location of Archaeological Sites by the Maximum Entropy Method (On the Example of the Late Roman Settlements in the Lipetsk Region)]. *Arkheologiya i geoinformatika. Chetvertaya Mezhdunar. konf. Tezisy dokladov* [Archaeology and Geoinformatics. Fourth International Conf. Abstracts of Reports]. Moscow: IA RAN Publ., 2019, pp. 85–86. (In Russian).

Stepinski T., Jasiewicz J. Geomorphons — A New Approach to Classification of Landform. *Proceedings of Geomorphometry*. Redlands: Esri Press, 2011, pp. 109–112. (In English).

Verhagen P., Drăguț L. Object-Based Landform Delineation and Classification from Dems for Archaeological Predictive Mapping. *Journal of Archaeological Science*, 2012, vol. 39, no. 3, pp. 698–703. DOI: 10.1016/j.jas.2011.11.001 (In English).

*Vostochnaya Evropa v seredine I tysyacheletiya n. e.* [Eastern Europe in the Middle of the 1<sup>st</sup> Millennium AD]. Moscow: Institut arkhologii RAN Publ., 2007. (Early Slavonic world. Archaeology of the Slavs and their neighbors: Series of scientific publications; iss. 9.) (In Russian).

Wachtel I., Zidon R., Garti S., Shelach-Lavi G. Predictive Modeling for Archaeological Site Locations: Comparing Logistic Regression and Maximal Entropy in North Israel and North-East China. *Journal of Archaeological Science*, 2018, no. 92, pp. 28–36. DOI: 10.1016/j.jas.2018.02.001 (In English).

Wollan A. K., Bakkestuen V., Kauserud H., Gulden G., Halvorsen R. Modelling and Predicting Fungal Distribution Patterns Using Herbarium Data. *Journal of Biogeography*, 2008, vol. 35, iss. 12, pp. 2298–2310. DOI: 10.1111/j.1365-2699.2008.01965.x (In English).

Yamazaki D., Ikeshima D., Tawatari R. et al. A High Accuracy Map of Global Terrain Elevations. *Geophysical Research Letters*, 2017, vol. 44, iss. 11, pp. 5844–5853. DOI: 10.1002/2017GL072874 (In English).

Yaworsky P. M., Vernon K. B., Spangler J. D., Brewer S. C., Coddling B. F. Advancing Predictive Modeling in Archaeology: An Evaluation of Regression and Machine Learning Methods on the Grand Staircase-Escalante National Monument. *PLoS ONE*, 2020, vol. 15, iss. 10, p. e0239424. DOI: 10.1371/journal.pone.0239424 (In English).

Zaitseva E. A. [Forecasting the Location of Archeological Objects Using GIS and Remote Sensing of the Earth] *Trudy IV (XX) Vserossiyskogo arkhelogicheskogo s'yezda v Kazani: sb. nauch. statey. V 5 t.* [Proceedings of the 4<sup>th</sup> (20<sup>th</sup>) All-Russian Archaeological Congress in Kazan: A Collection of Sci. Articles. In 5 vols.]. Kazan: Otechestvo Publ., 2014, vol. 4, pp. 399–401. (In Russian).

Zemtsov G. L., Sarychev D. V., Fabricius E. V., Goncharov V. O. [Population of the Upper Don Region in the Late Roman Time: Victim or Aggressor?]. *Stratum plus. Arkheologiya i kul'turnaya antropologiya* [Stratum plus. Archaeology and Cultural Anthropology], 2019, no. 4, pp. 257–264. (In Russian).

Zemtsov G. L., Sarychev D. V., Fabritsius E. V. [Cartography of Late Roman Settlements in the Basin of the River Bystraya Sosna]. *Gumanitarnyye issledovaniya Tsentral'noy Rossii* [Humanities Researches of the Central Russia], 2019, no. 4 (13), pp. 7–10. DOI: 10.24411/2541-9056-2019-00001 (In Russian).

Zemtsov G. L., Sarychev D. V., Goncharov V. O., Fabritsius E. V. [Plyshchan Microregion of Settlements of the Late Roman Period in the Upper Don Region]. *Rossiiskaia arkheologiya* [Russian Archaeology], 2020, no. 2, pp. 60–71. DOI: 10.31857/S086960630005677-3 (In Russian).

Zevenbergen L. W., Thorne C. R. Quantitative Analysis of Land Surface Topography. *Earth Surface Processes and Landforms*, 1987, vol. 12, iss. 1, pp. 47–56. DOI: 10.1002/ESP.3290120107 (In English).

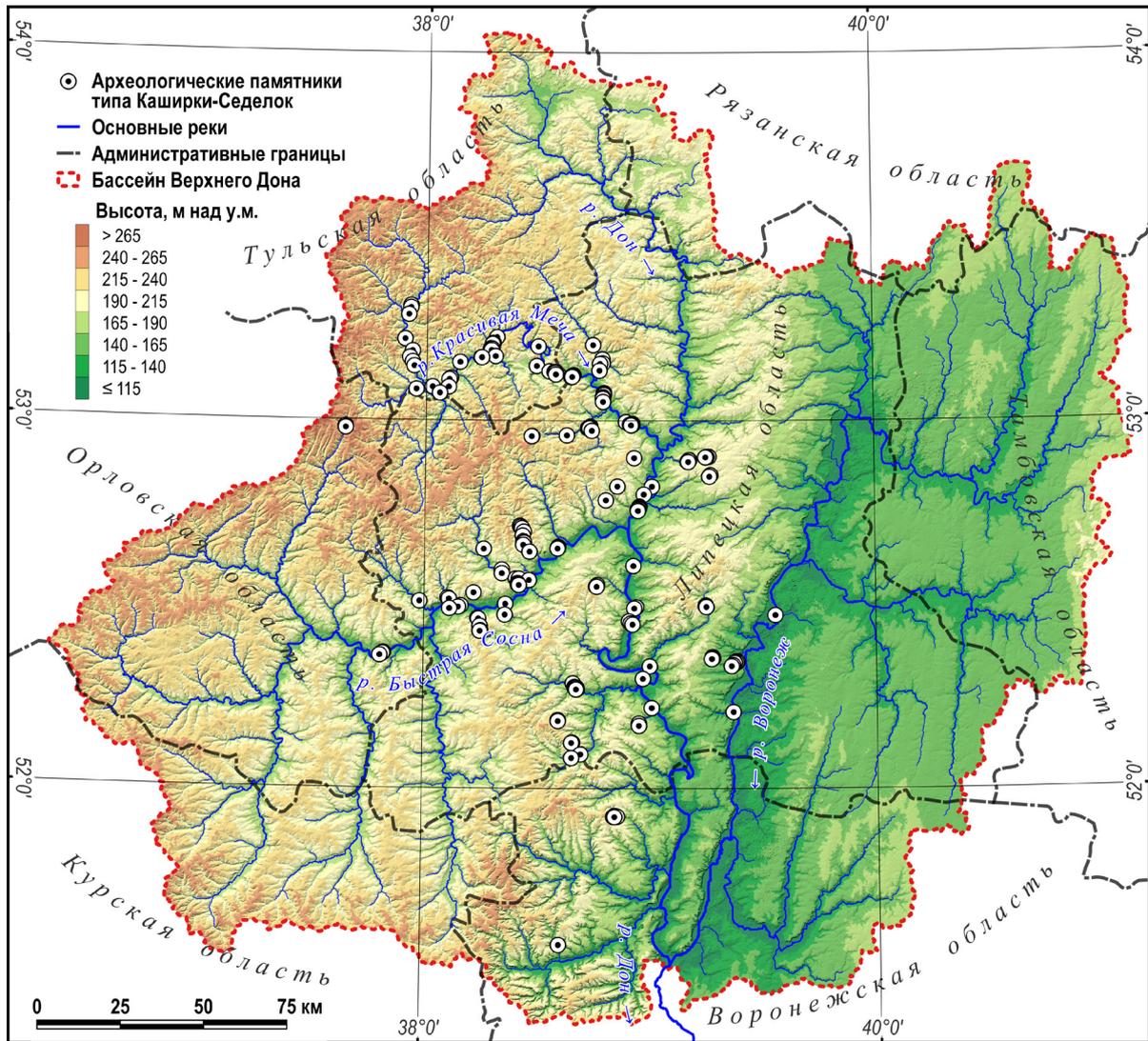


Рис. 1. Карта исследуемой территории — бассейн Верхнего Дона. Точками отмечены археологические памятники сер. III — нач. IV в. н. э. типа Каширки-Седелок

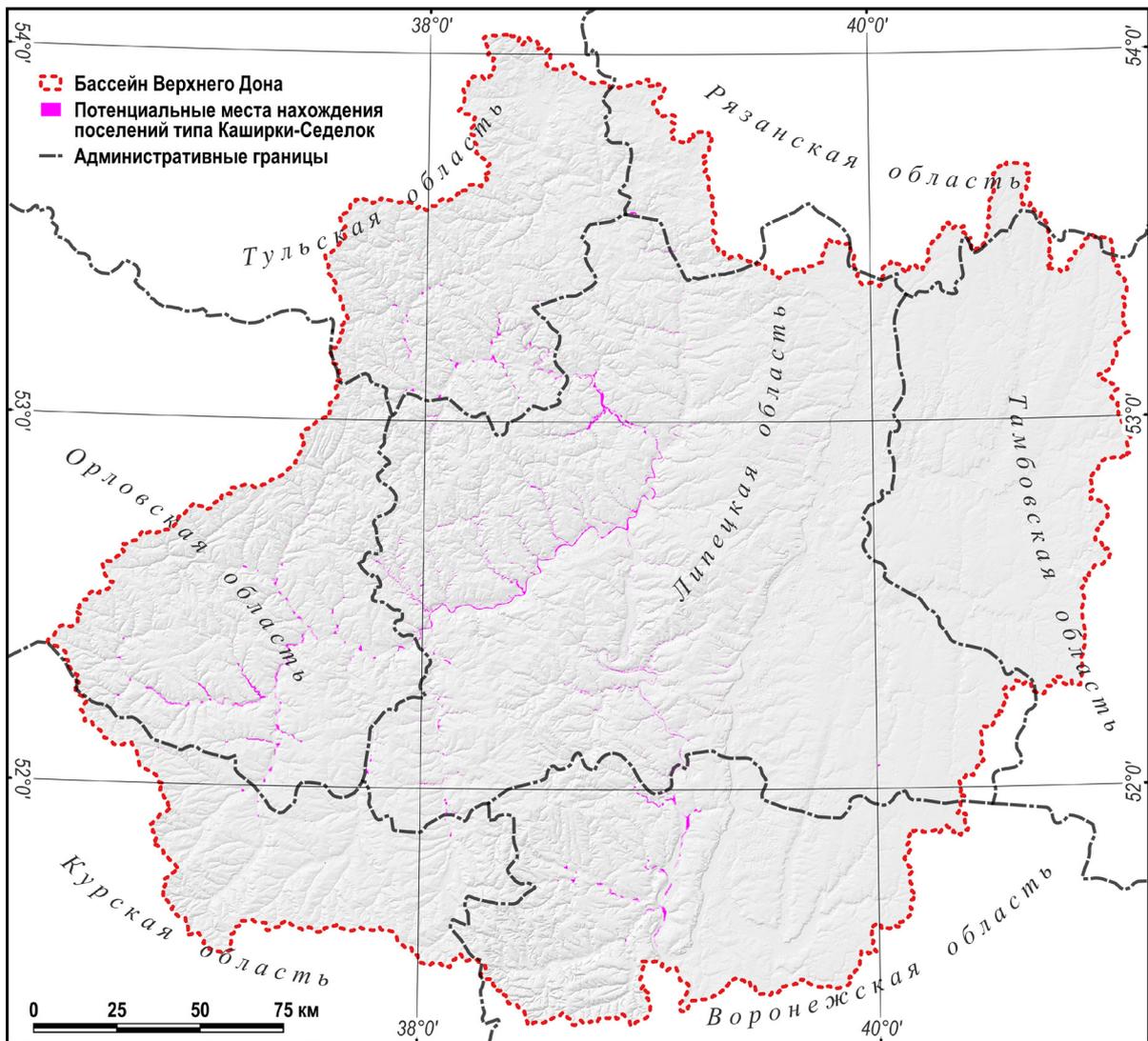


Рис. 2. Карта потенциально пригодных территорий для размещения поселений типа Каширки-Седелок в бассейне Верхнего Дона по результатам предиктивного археологического моделирования методом максимальной энтропии (MaxEnt)