

ПРОБЛЕМА ОТСУТСТВИЯ “ОТСУТСТВИЙ”: ПОДХОД ЭНГЛЕРА–ХЕНГЛА В МОДЕЛИРОВАНИИ ПРОСТРАНСТВЕННОГО РАСПРЕДЕЛЕНИЯ ВИДОВ

© 2023 г. С. С. Огурцов*, **, @

* Центральнo-Лесной государственный природный биосферный заповедник,
пос. Заповедный, 32, Нелидовский городской округ, Тверская область, 172521 Россия

** Институт проблем экологии и эволюции им. А.Н. Северцова РАН, Ленинский проспект, 33, Москва, 119071 Россия

@E-mail: etundra@mail.ru

Поступила в редакцию 02.10.2023 г.

После доработки 08.10.2023 г.

Принята к публикации 08.10.2023 г.

Рассмотрены возможности создания искусственных точек отсутствия (псевдоотсутствий) для построения моделей пространственного распределения видов. Описан подход, предложенный Робинoм Энглeром и адаптированный Томиславoм Хенглом, который учитывает индексы пригодности местообитаний и расстояния до точек присутствия для создания точек псевдоотсутствия. На примере черники в Центральнo-Лесном заповеднике и его охрaнной зоне сравниваются обобщенные линейные модели, построенные на основе подхода Энглeра-Хенгла, пройденных маршрутов и расстояний до точек присутствия, а также модель, построенная по методу максимальной энтропии. Полученные результаты свидетельствуют о превосходстве модели на основе подхода Энглeра-Хенгла как с точки зрения оценок качества, так и с точки зрения реалистичности построенных карт пространственного распределения.

Ключевые слова: пригодность местообитаний, пространственное распределение, распространение вида, псевдоотсутствия, GLM, MaxEnt

DOI: 10.31857/S1026347023600814, **EDN:** HODFBI

В моделировании пространственного распределения/распространения видов (Species Distribution Modelling, SDM) большую популярность приобрели методы на основе данных только о присутствии вида (presence-only) на территории исследований (Guisan *et al.*, 2017). Полевой сбор первичных данных для построения таких моделей значительно проще, а сами они требуют намного меньше допущений. Тем не менее, качество этих моделей далеко не всегда превосходит подходы на основе данных о присутствии/отсутствии (presence/absence) (Engler *et al.*, 2004; Duque-Lazo *et al.*, 2016). Большинство из них имеют тенденцию переоценивать область потенциального распространения вида (Engler *et al.*, 2004). Данные только о присутствии отличаются от данных о присутствии/отсутствии тем, что они указывают места, где наблюдался изучаемый вид, но не могут использоваться для определения мест, где он не встречался (Dettmers, Bart, 1999).

Целый ряд таких техник, как обобщенные линейные модели (Generalized Linear Models, GLM; Mladenoff *et al.*, 1999; Glenz *et al.*, 2001), обобщенные аддитивные модели (Generalized Additive

Models; Guisan, Zimmermann, 2000) и деревья классификации (Classification Trees; Jerina *et al.*, 2003) построены на алгоритмах работы с данными типа “присутствие/отсутствие”, что является их ограничением. Очень часто исследователь обладает лишь данным о присутствии, что является препятствием для использования техник “присутствия/отсутствия”. В то же время принято считать, что последние техники моделирования более надежны при наличии достоверных данных об отсутствиях (Chefaoui, Lobo, 2008; Duque-Lazo *et al.*, 2016).

Одним из возможных способов компенсации отсутствий является использование информации из всей (или только части) области исследования, называемой “фоном” (background) и объединение их с существующими данными о присутствии (Elith *et al.*, 2006; Peterson *et al.*, 2011). Такой подход использования данных называется “присутствие/фон” (presence/background). Фоновые точки призваны отразить условия окружающей среды, в которых существует вид (доступные условия), чтобы противопоставить их условиям, в которых вид был обнаружен. Данный подход используется

в таких техниках как максимальная энтропия (Maximum Entropy, MaxEnt; Phillips *et al.*, 2006), ENFA (Ecological Niche Factor Analysis; Hirzel *et al.*, 2002) и GARP (Genetic Algorithm for Rule-set Prediction) (Stockwell, Peters, 1999; Elith *et al.*, 2006; Peterson *et al.*, 2011). При этом фоновые точки генерируются случайным образом и почти никак не контролируются исследователем (в случае ENFA) или контролируются лишь частично посредством поправочного файла-маски (bias-file) (в случае MaxEnt) (Phillips *et al.*, 2009).

Еще одним подходом является способ объединения данных о присутствии с так называемыми псевдоотсутствиями (pseudo-absences). Такая схема исследования (далее дизайн) называется соответственно “присутствие/псевдоотсутствие” (presence/pseudo-absence). Точки псевдоотсутствия используются для того, чтобы отразить условия, в которых с большой вероятностью вид отсутствует (Stockwell, Peters, 1999; Zaniewski *et al.*, 2002). Надежность псевдоотсутствий зависит от характеристики самого вида (Hirzel *et al.*, 2001), его численности (Kéry *et al.*, 2010) и способов сбора данных (MacKenzie, Royle, 2005). Если вероятность надежных псевдоотсутствий для растений достаточно высока, то для мобильных животных (в частности, млекопитающих) она может быть не столь высокой (Boyce *et al.*, 2002; Johnson *et al.*, 2006; Smulders *et al.*, 2010). При этом всегда остается вероятность того, что среди них могут встретиться ложные отсутствия (Franklin, 2009). Таким образом, псевдоотсутствия это не достоверные, но предполагаемые отсутствия, а фоновые точки — это случайная комбинация всех условий среды. Подход с псевдоотсутствиями используют в тех случаях, когда исследователь обладает данным лишь о присутствии, но хочет использовать “presence/absence” моделирование (Chefaoui, Lobo, 2008). Это особенно актуально при построении функций выбора ресурсов (Resource Selection Function, RSF, Manly *et al.*, 2002), которые, как правило, опираются на обобщенные линейные модели.

Способ создания точек псевдоотсутствия до сих пор является дискуссионным вопросом (Kobler, Adamic, 2000; Keating, Cherry, 2004; Franklin, 2009; VanderWal *et al.*, 2009; Ghoddousi, 2010; Barbet-Massin *et al.*, 2012; Renner *et al.*, 2015). В настоящей работе рассматривается перспективный для ряда исследовательских задач вариант генерации псевдоотсутствий — подход Энглера-Хенгла. Изначально Робин Энглер с коллегами предложили генерировать псевдоотсутствия на основе уже построенных моделей “присутствия/фона” (Engler *et al.*, 2004). Это было развитием идей Элизабет Заниевски и ее коллег, которые впервые предложили сначала строить карту пригодности местообитаний (на основе GAM), а уже потом на ее основе создавать псевдоотсутствия и строить новую регрессионную модель (Zaniewski *et al.*, 2002). В от-

личие от Заниевски с соавторами, которые для первой GAM использовали полностью случайные точки, Энглер для этого строил модель ENFA, а уже потом на основе ее карты генерировал псевдоотсутствия, которые затем добавлялись к исходным данным о присутствиях и использовались для построения GLM (Engler *et al.*, 2004). Результатом ENFA (или любой другой модели с “присутствием/фоном”) являлась карта индексов пригодности местообитаний (Habitat Suitability Index, HSI в шкале от 0 до 100%). Отображая области с низким HSI, возможно оценить, где вероятность появления вида очень мала, а затем случайным образом сгенерировать новый набор взвешенных точек псевдоотсутствий в этих областях. На втором этапе с помощью этих созданных точек и известных присутствий строилась уже итоговая модель “присутствия/отсутствия” (например, GLM или GAM) (Zaniewski *et al.*, 2002; Engler *et al.*, 2004; Jiménez-Valverde *et al.*, 2008; Hengl *et al.*, 2009). Таким образом, выявление непригодных местообитаний с помощью “профильных” (т.е. “присутствие/фон”) техник позволяет генерировать надежные псевдоотсутствия для построения моделей “присутствия/отсутствия” (Chefaoui, Lobo, 2008). Доказано, что GLM, которые построены с использованием точек псевдоотсутствий, взвешенных на основе карт ENFA, превосходят таковые модели, построенные лишь на случайных точках псевдоотсутствий (Engler *et al.*, 2004).

При таком подходе возникает проблема с определением правила, по которому происходит создание псевдоотсутствий. Так, Энглер с коллегами выбирали точки псевдоотсутствия только в непригодных областях согласно построенной карте ENFA. Непригодные области определялись по заданному пороговому значению — 0.3, согласно Энглеру с соавт. (Engler *et al.*, 2004). Чуть позже Томислав Хенгл с коллегами признали такое сочетание факторного анализа и GLM наиболее перспективным, поскольку в нем используется лучшее из двух техник (Hengl *et al.*, 2009). Эти авторы развили идею Энглера, предложив вычислительную процедуру, которая объединяет оценку плотности (ядерное сглаживание), факторный анализ экологических ниш (ENFA) и геостатистику (регрессионный кригинг). При таком подходе псевдоотсутствия отбирались не только на основе информации о пригодности местообитаний, но и их географическом положении (Hengl *et al.*, 2009). Известно, что точки псевдоотсутствия, созданные вдали (в первую очередь, в экологическом пространстве) от точек присутствия увеличивают объяснительную способность моделей (Chefaoui, Lobo, 2008). Помимо этого Хенгл с коллегами предложили выбирать псевдоотсутствия не только лишь в непригодных местообитаниях, а пропорционально индексу пригодности местообитаний. Применение точек псевдоотсут-

ствия, полученных с помощью этого метода, можно рассматривать как промежуточный подход между “presence-only” и “presence/absence” дизайнами, который особенно полезен, когда точные данные об отсутствии недоступны (Chefaoui, Lobo, 2008; Sillero, Barbosa, 2021).

В настоящей работе сравниваются возможности подхода Энглера-Хенгла с другими популярными способами создания точек псевдоотсутствий для GLM/GAM, а также с дизайном “присутствие/фон” модели MaxEnt. В отличие от Энглера и Хенгла за основу создания псевдоотсутствий была взята не ENFA, а сама MaxEnt. Для создания моделей были использованы 456 точек встреч черники (*Vaccinium myrtillus* L., 1753), собранных автором за период с 2008 по 2022 гг. Данный вид был выбран по нескольким причинам. Во-первых, он является типичным кустарничковым растением для территории исследований. Во-вторых, его распространение в значительной степени связано с характерными типами еловых (*Picea abies* (L.) Н. Karst., 1881) и сосновых (*Pinus sylvestris* L., 1753) лесов. Таким образом, при хорошем знании территории исследований распространение черники будет достаточно легко проверить по построенным картам. В-третьих, подход Энглера-Хенгла был успешно опробован на мобильном и эвритопном виде (*Ursus arctos* L., 1758) ранее (Огурцов, 2023). По этой причине было принято решение сосредоточиться на неподвижном и достаточно стенотопном виде. Данная работа продолжает серию исследований по моделированию пригодности местообитаний, начатых Ю.Г. Пузаченко и его коллегами (Желтухин и др., 2009; Пузаченко и др., 2010, 2011 и др.).

Цель исследования — проверить эффективность подхода Энглера-Хенгла для генерации точек псевдоотсутствия в моделировании пространственного распределения видов на примере черники. Для достижения цели решались следующие задачи: 1) создать ряд наборов точек псевдоотсутствия черники по различным правилам; 2) на основе этих наборов построить GLM; 3) сравнить качество построенных GLM между собой, а также с моделью на основе MaxEnt.

МАТЕРИАЛ И МЕТОДЫ

Территория исследований

Исследование проводили в Центрально-Лесном государственном природном биосферном заповеднике (ЦЛГЗ) и его охранной зоне (рис. 1). Территория представляет собой южнотаежный ландшафт, в прошлом частично трансформированный человеческой деятельностью. Заповедник расположен на юго-западе Валдайской возвышенности (Россия, Тверская обл., 56°26′–56°31′ с.ш., 32°29′–33°01′ в.д.); это грядово-холмистая возвы-

шенная равнина с высотами от 238 до 267 м над ур. м. Его площадь составляет 24 421 га, а площадь охранной зоны вокруг него — 46 694 га. Для дренированных частей моренных гряд характерны южнотаежные широколиственно-еловые леса и их производные, для озерных и флювиогляциальных равнин — бореальные еловые леса. На пологих склонах моренных гряд и в понижениях озерных равнин развиты верховые болота (рис. 1). Для территории характерен умеренно-континентальный климат. Более половины осадков выпадает в виде дождя в летне-осенний период; остальное количество — в виде снега зимой—весной. Средняя температура воздуха составляет +4.21°C (в июле +16°C, в январе –10°C), среднегодовое количество осадков — 730.9 мм (Пузаченко и др., 2016).

Общий принцип моделирования

В настоящем исследовании было проверено 4 различных дизайна моделирования на основе двух разных техник. В первом случае использовалась MaxEnt (далее модель “MaxEnt”). Во втором случае строилась GLM с применением подхода Энглера-Хенгла (Engler-Hengl, EH; далее модель “EH”). В третьем случае строилась GLM, но точки псевдоотсутствия отбирались согласно весам расстояний (distance) до точек присутствия (модель “dist”). Такой подход в большей степени соответствует дизайну на основе радиальных буферов вокруг точек встреч, но без необходимости выбора радиуса буферной зоны. В четвертом случае строилась GLM, но точки псевдоотсутствия отбирались в пределах пройденных маршрутов (треков), где не было зафиксировано присутствия черники (модель “tracks”). Таким образом, сравнивался подход Энглера-Хенгла с двумя другими популярными приемами создания псевдоотсутствий (расстояния/буферы и маршруты), а также с самой MaxEnt. В общем виде этапы подготовки данных, построения и сравнения моделей представлены на рис. 2. Вся подготовительная работа, анализы и моделирование проводились в программе RStudio 1.1.447 с использованием языка R 4.0.3 (R Core Team, 2020).

Предикторные переменные окружающей среды

На первом этапе был подготовлен набор предикторов окружающей среды (рис. 2, этап 1). В качестве них использовались вегетационные индексы (ВИ), полученные на основе данных мультиспектральных спутниковых снимков, морфометрические характеристики, полученные на основе цифровой модели рельефа (ЦМР), и типы растительного покрова (Coudun, Gégout, 2007; Roberts *et al.*, 2014; Nielsen *et al.*, 2016; Pentnerian *et al.*, 2019; McClelland *et al.*, 2020). В качестве данных мультиспектральной спутниковой съемки для

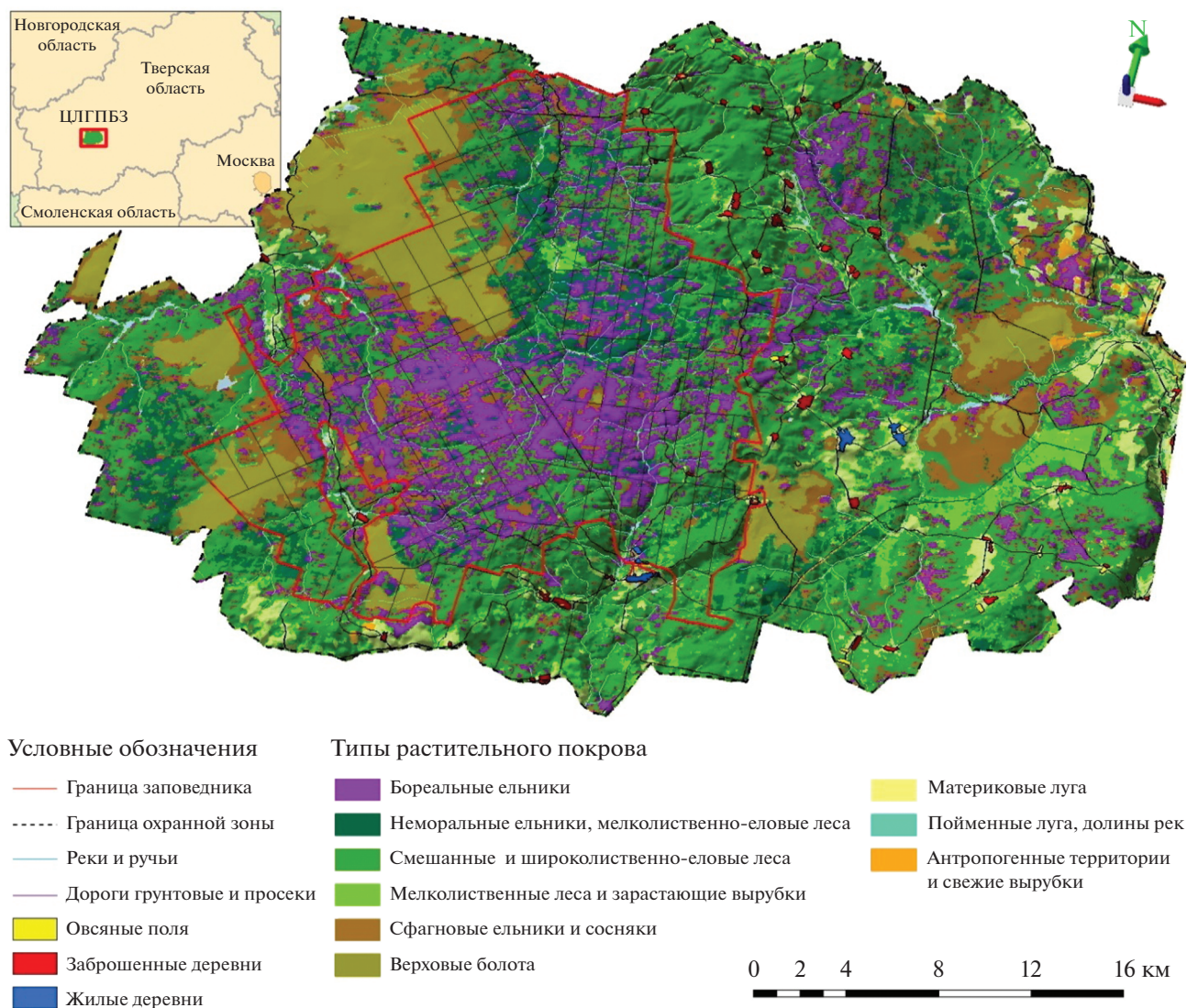


Рис. 1. Территория Центрально-Лесного государственного природного биосферного заповедника (ЦЛГПБЗ) и его охранной зоны с отображением типов растительного покрова согласно классификации сцен спутника Landsat 8.

расчета ВИ использовались все безоблачные сцены спутника Landsat 8 OLI-TIRS (Landsat, 2019; <http://earthexplorer.usgs.gov>) за период с апреля по сентябрь с 2013 по 2020 гг. Всего было отобрано 9 сцен за разные даты: 26.04.2014, 06.06.2014, 10.09.2014, 29.04.2015, 25.09.2017, 07.05.2018, 11.08.2018, 19.05.2019, 11.06.2019. Из вегетационных индексов применялись усовершенствованный ВИ (Enhanced Vegetation Index, EVI; Huete *et al.*, 2002) и стандартизованный индекс разностей увлажненности (Normalized Difference Moisture Index, NDMI; Vermote *et al.*, 2016). Дополнительно был рассчитан ортогональный ВИ влажности (wetness), являющийся мерой оценки комбинированной влажности почвы и растительного покрова (Baig *et al.*, 2014). Его рассчитали путем специального преобразования Tasseled Cap Transformation

(Kauth, Thomas, 1976) с поправками для сенсора OLI спутника Landsat 8 (Baig *et al.*, 2014).

Морфометрические характеристики рельефа рассчитывались на основе глобальной ЦМР (SRTM 1 Arc-Second Global, <http://www.earthexplorer.usgs.gov>) с пространственным разрешением 30 м. Это абсолютная высота (“altitude”), экспозиция северных (“northness”) и восточных склонов (“eastness”), а также составной топографический индекс (Compound Topographic Index, “СТІ”; Moore *et al.*, 1991). Перед расчетом морфометрических характеристик ЦМР была проецирована в проекцию UTM 36N (Sillero, Barbosa, 2021).

К типам растительного покрова были отнесены следующие 7 категорий: 1) бореальные ельники, 2) неморальные ельники и мелколиственно-еловые леса, 3) смешанные и широколиственно-еловые ле-

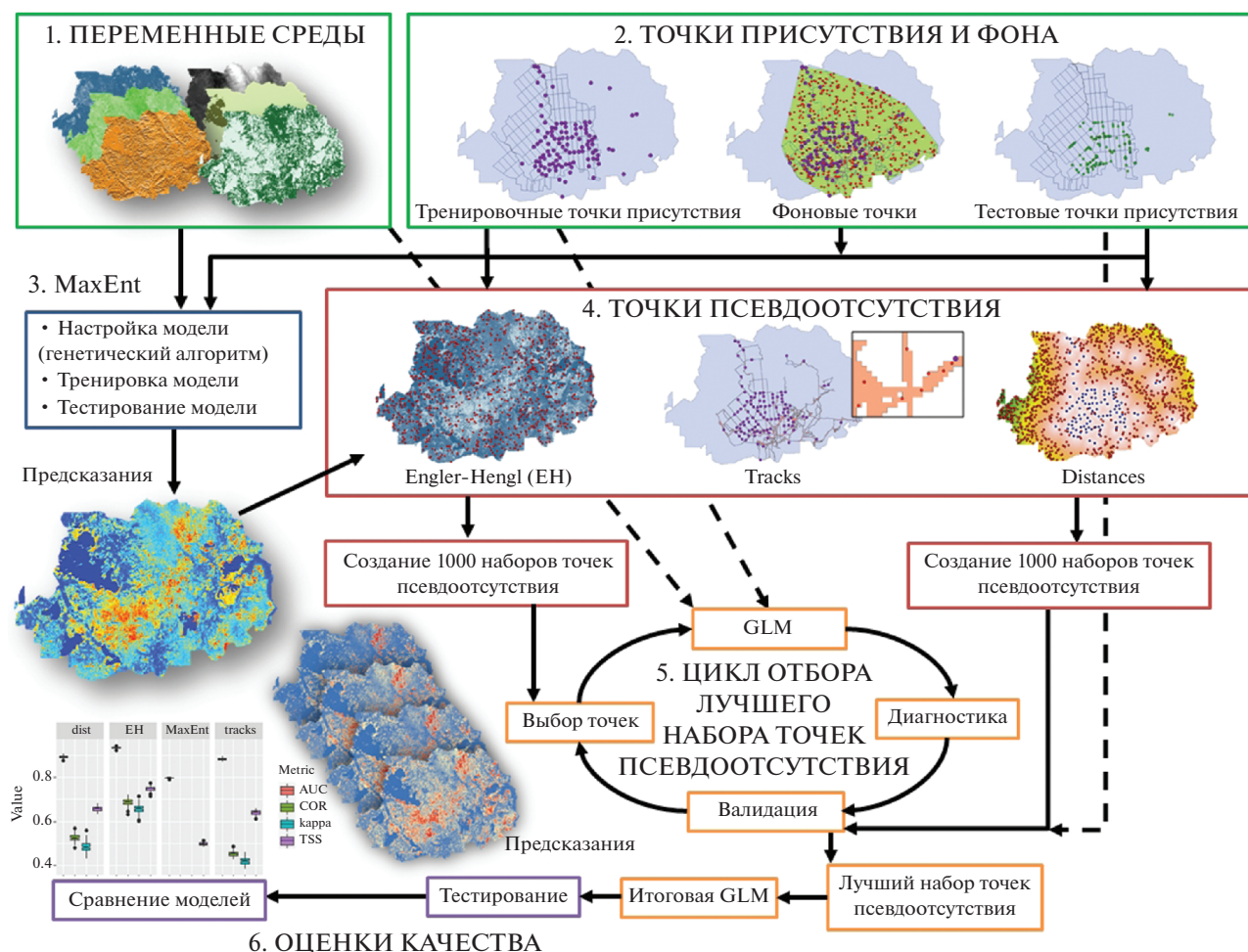


Рис. 2. Этапы подготовки данных, построения и сравнения моделей пространственного распределения черники на территории Центрально-Лесного заповедника и его охранный зоны на основе MaxEnt и GLM для трех дизайнов создания точек псевдоотсутствия.

са, 4) молодые вторичные мелколиственные леса, зарастающие вырубки, гари и ветровалы, 5) сфагновые ельники и сосняки, 6) верховые болота, 7) материковые луга. Все данные были получены на основе карты типологии растительного покрова, построенной по результатам визуального полуавтоматического дешифрирования сцен спутника Landsat 8 OLI (рис. 1). Дешифрирование выполнялось с помощью набора инструментов “Image Classification” в геоинформационной системе (ГИС) ArcMap 10.6.1 (Esri Inc.) методом классификации максимального правдоподобия (Pal, Mather, 2003).

Все расчеты переменных окружающей среды были проведены в ГИС ArcMap и SAGA 7.7.1 (Conrad *et al.*, 2015). Переменные были приведены к единому географическому экстенду и системе координат WGS84 в проекции UTM 36N с разрешениями 30 м. Все переменные прошли проверку на мультиколлинеарность (Araújo *et al.*, 2019; Silero, Barbosa, 2021). Сначала была выполнена ран-

говая корреляция Спирмена и составлена матрица, на основании которой среди них были выбраны кандидаты на исключение (при $r > |0.7|$). Дополнительно был проведен VIF-тест (Variance Inflation Factor). В качестве порога было выбрано значение $VIF = 10$ (Zuur *et al.*, 2009; Duque-Lazo *et al.*, 2016). Для тестов и проверок на мультиколлинеарность применяли R-пакеты “car” (Fox, Weisberg, 2018), “ecospat” (Broennimann *et al.*, 2018) и “usdm” (Naimi *et al.*, 2014).

Точки присутствия вида

Сбор данных о распространении черники проводился во время учетных маршрутов по территории исследований согласно общим принципам цифрового картографирования растений (Nielsen *et al.*, 2003, 2016; Shores *et al.*, 2019). Помимо присутствия отмечали также обилие (проективное покрытие) кустарничка в радиусе 5 м по шкале Браун-Бланке (Shores *et al.*, 2019). В настоящей

работе в качестве присутствий включали лишь те точки, где проективное покрытие составляло от 3 до 5 баллов (т.е. от 25 до 100%). Сбор данных проводился с помощью специально разработанной формы в мобильном приложении ArcGIS Quick-Capture (Esri Inc.) с заданной минимальной точностью геопривязки 4.6 м.

Чтобы снизить эффекты пространственной погрешности при сборе данных, все точки регистрации черники были разрежены (Coudun, Gégout, 2007; Aiello-Lammens *et al.*, 2015; Petrosyan *et al.*, 2019, 2023) по расстоянию 605 м. Оценка кластеризованности точек проводилась с помощью расчета среднего индекса ближайшего соседа (Average Nearest Neighbor Index, ANNI). Значения индекса меньше 1 указывают на кластерность, а больше 1 — на дисперсность (Evans *et al.*, 2021). Разреженные точки, удовлетворяющие дисперсному распределению ($ANNI > 1$), использовались как тренировочные для построения моделей (рис. 2, этап 2). После разреживания осталось 107 таких точек ($ANNI = 1.12$, $z = 2.37$, $p < 0.05$). Пространственное разреживание выполнялось с помощью R-пакета “spThin” (Aiello-Lammens *et al.*, 2015). Расчет ANNI проводился, используя R-пакет “spatialEco” (Evans *et al.*, 2021).

Для независимого тестирования использовались 206 точек встреч черники, взятые из архива заповедника. Все эти точки были собраны независимо от автора сторонними исследователями за тот же период времени.

Моделирование с помощью MaxEnt

На третьем этапе для построения моделей использовалась MaxEnt (рис. 2, этап 3), как наиболее оптимальная техника для данных только о присутствии (Phillips *et al.*, 2006; Elith *et al.*, 2011; Merow *et al.*, 2013; Phillips *et al.*, 2017). Для этого применялась одноименная программа Maxent 3.4.1 (Phillips *et al.*, 2018). Для исправления погрешности неравномерного сбора данных (Merow *et al.*, 2013; Syfert *et al.*, 2013) был создан корректирующий файл (bias file) на основе минимального выпуклого полигона (Minimum Convex Polygon, MCP), охватывающего все пройденные маршруты (Maiorano *et al.*, 2015). В пределах этого полигона с помощью процедуры рандомизации было создано 1000 различных наборов 10000 точек, которые в дальнейшем указывались MaxEnt как фоновые точки (background samples) для сравнений (Phillips, Dudík, 2008). Во всех случаях использовалось максимум 1000 итераций и порог сходимости (convergence threshold) 10^{-5} .

Выбор оптимальных настроек гиперпараметров модели MaxEnt производился с помощью генетического алгоритма (Vignali *et al.*, 2020). В этом случае размер популяции составлял 20, доля луч-

ших моделей — 40%, доля случайных моделей — 20%. Шанс “мутации” во время “кроссинговера” был установлен как 40%. Число “поколений” равнялось 5 (Vignali *et al.*, 2020; 2020a). Для проверки (validation) таких моделей применялась кросс-валидация по методу пространственных блоков (spatial block k-fold cross-validation) с размером блока 820 м и $k = 10$ при помощи R-пакета “blockCV” (Valavi *et al.*, 2019), т.е. в каждом случае использовалось 10 реализаций (всего 10000 моделей). Размер блока был определен по медиане размаха пространственной автокорреляции переменных среды (Valavi *et al.*, 2019). Лучшая комбинация гиперпараметров MaxEnt определялась по значениям показателя площади под кривой (Area Under the Curve, AUC) ROC-графика (Receiver Operating Characteristic), информационного критерия Акаике (Akaike’s Information Criterion, AIC) и показателя True Skill Statistic (TSS). После выбора оптимальных настроек и построения модели проводилась процедура редукции переменных пошаговым удалением наименее важных предикторов с оценкой полученных моделей по значению тестовой TSS (Vignali *et al.*, 2020).

Для каждого набора фоновых точек была построена модель с оптимальными настройками. Оценка их качества проводилась по независимым данным, согласно “золотому стандарту” (Araújo *et al.*, 2019), на основе полностью сторонних тестовых (архивных) данных. Оценочными метриками выступили тестовая TSS, тренировочная AUC (AUC_{train}), тестовая AUC (AUC_{test}) и их разница (AUC_{diff}). Показатель TSS рассчитывался по порогу, максимизирующему его значение (Guisan *et al.*, 2017; Vignali *et al.*, 2020a). Итоговая модель была построена на основе того набора фоновых точек, при котором значения AUC_{test} и TSS были максимальные, а AUC_{diff} — минимальная. Определение оптимальных настроек модели с помощью генетического алгоритма, построение моделей, их оптимизация и оценка качества были выполнены с помощью R-пакетов “SDMtune” (Vignali *et al.*, 2020, 2020a) и “dismo” (Hijmans *et al.*, 2017).

Итоговая карта строилась на основе формата представления данных “cloglog” (Merow *et al.*, 2013). В этом случае предсказания были интерпретированы как относительные вероятности присутствия (relative probabilities of presence), где 0 — минимальная вероятность присутствия (прогнозируемое отсутствие), 1 — максимальная вероятность присутствия (прогнозируемое присутствие) (Merow *et al.*, 2013).

Модели были построены как для авторских, так и для архивных точек. Во втором случае в качестве тестовых присутствий выступали авторские данные.

Создание точек псевдоотсутствия

Для создания точек псевдоотсутствия использовался алгоритм рандомизации, как наиболее простой и не требующий особых допущений (Keating, Cherry, 2004; Barbet-Massin *et al.*, 2012). Для создания карты весов псевдоотсутствий полученная карта относительных вероятностей присутствия на основе MaxEnt (рис. 2, этап 3) переводилась в шкалу от 0 до 100 (в соответствии со шкалой HSI), а затем инвертировалась. В дальнейшем был использован подход Хенгла с соавт. (Hengl *et al.*, 2009), которые сформулировали гипотезу и предложили правило создания точек псевдоотсутствия. Правило основано на предположении, что распространение вида по территории есть функция пригодности его местообитаний (или вероятности присутствия). Таким образом, распределение вероятностей (τ) может быть применено для генерации точек псевдоотсутствия как:

$$\tau(x) = [100\% - \text{HSI}(x)]^2. \quad (1)$$

Квадратная степень гарантирует, что большинство точек псевдоотсутствия будут находиться ближе к границе низких значений HSI. При этом псевдоотсутствия будут приблизительно следовать распределению Пуассона. Чтобы учесть влияние расстояний от точек присутствия, добавляется вторая часть уравнения $d_R(x)$, которая представляет собой нормализованные расстояния (в диапазоне от 0 до 100%), т.е. отношения расстояний до ближайшей точки присутствия (x) к максимальному расстоянию. В конечном итоге мы получаем следующую формулу:

$$\tau(x) = \left[\frac{d_R(x) + (100\% - \text{HSI}(x))}{2} \right]^2. \quad (2)$$

Полученная формула определяет условия отбора точек псевдоотсутствия на основе двух условий: 1) чем больше расстояние от ближайшей точки присутствия, тем больше вероятность отсутствия; 2) вероятность отсутствия обратно пропорциональна значению индекса пригодности местообитаний (вероятности присутствия). Таким образом, в областях с низкими значениями HSI точек псевдоотсутствия будет больше, а в областях с высокими значениями HSI — меньше. По данной формуле для каждого пикселя оцениваются “веса отсутствия” $\tau(x)$, по которым в дальнейшем строится карта. Такие условия соответствуют принципу создания псевдоотсутствий, используемому Лобо с соавт. (Lobo *et al.*, 2010), при котором учитывается экологическая непригодность и расстояния до точек присутствия. Вместе с этим такой способ исправляет погрешности подхода Энглера, где использовались не HSI как таковые, а лишь категории пригодности. Тем не менее, поскольку данная формула основана на допущениях, а полевая верификация точек не

проводилась, нельзя гарантировать реальные отсутствия в искусственно-сгенерированных точках. В связи с этим в дальнейшем такие точки рассматриваются как псевдоотсутствия (рис. 2, этап 4).

В то же время, мозаичность территории исследований достаточно высокая, поэтому параметр расстояния не может гарантировать создание надежных псевдоотсутствий. В связи с этим было решено изменить формулу Хенгла (2), убрав из нее параметр расстояния (1), и создать на ее основе аналогичную карту “весов отсутствия” (дизайн “ЕН”). Еще одним изменением стало использование карты на основе модели MaxEnt, а не ENFA в качестве карты весов.

После этого выполнялась процедура случайного создания точек псевдоотсутствия в количестве, десятикратно превышающем число точек присутствия по правилам построенной карты весов отсутствия. Поскольку точки генерируются случайным образом, и одна генерация может сильно отличаться от другой в виду небольшого числа точек относительно общего числа ячеек грида (773 712), была выполнена 1000 таких симуляций согласно Энглеру (Engler *et al.*, 2004).

Для создания точек псевдоотсутствия на основе пройденных маршрутов (дизайн “tracks”) и расстояний (дизайн “dist”) было также сгенерировано 1000 их наборов в числе, десятикратно превышающем число точек присутствия, но на основе масок пройденных маршрутов и расстояний соответственно. Предварительно из маски маршрутов были вырезаны все ячейки грида, куда попадали точки присутствия. Веса расстояний рассчитывались согласно предварительно построенной карте эвклидовых расстояний до точек присутствия (рис. 2, этап 4).

Для тестирования итоговых моделей необходимо получить набор независимых точек псевдоотсутствия (т.е. отличных от тех, на которых обучалась модель). Для этого повторялись описанные выше действия, но уже для тестовых (архивных) точек присутствия. В итоге были созданы такие же 1000 наборов точек псевдоотсутствия для всех трех дизайнов, но построенные уже по архивным точкам (рис. 2).

Построение GLM

На основе сгенерированных наборов точек псевдоотсутствия и точек присутствия строились 1000 моделей GLM (одна модель на каждую генерацию точек) со связующей функцией на основе логитов. Для упрощения процедуры построения моделей в рамках поставленных задач не рассматриваются взаимодействия предикторов и группирующая переменная.

Для каждой построенной модели проводился расчет основных оценок ее качества: каппы Ко-

Таблица 1. Оценки качества и выбранный порог для лучших моделей пространственного распределения черники на территории Центрально-Лесного заповедника и его охранной зоны для дизайнов GLM на основе подхода Энглера-Хенгла (EH), пройденных маршрутов (tracks), расстояний до точек присутствия (dist), а также для максимальной энтропии (MaxEnt)

Модель	Порог	Кappa	TSS	COR	AUC
EH	0.26	0.71	0.76	0.72	0.94
tracks	0.35	0.45	0.65	0.49	0.90
dist	0.17	0.56	0.67	0.57	0.91
MaxEnt	0.42	—	0.49	—	0.79

Примечание: Кappa — каппа Коэна, TSS — True Skill Statistic, COR — коэффициент точечной бисериальной корреляции, AUC — показатель площади под ROC-кривой.

эна (Cohen's Kappa), TSS, коэффициента точечной бисериальной корреляции (COR) и AUC. Затем выбиралась та модель, для которой значения этих показателей были максимальными (лучшая модель). Все оценки рассчитывались на основе независимых тестовых данных. Таким образом, был получен набор точек псевдоотсутствия в местах, максимально отличных от присутствий (рис. 2, этап 5). Оценка качества проводилась при помощи R-пакетов "PresenceAbsence" (Freeman, Moisen, 2008), "dismo", "ecospat" и "biomod2" (Thuiller *et al.*, 2020). Расчет AUC проводился при помощи R-пакета "pROC" (Robin *et al.*, 2020). После выбора лучшего набора точек псевдоотсутствия строилась лучшая GLM для каждого дизайна.

Сравнение моделей

Для оценок дискриминации использовалось 5 порог-зависимых показателей: доля правильных предсказаний (correct classification rate, CCR), отношение ошибочной классификации (misclassification rate, MR), чувствительность (sensitivity, SE), специфичность (specificity, SP) и отношение шансов (odds ratio, OR). Для общего сравнения моделей был использован набор стандартных показателей: AUC, TSS, Kappa и COR (рис. 2, этап 6) (Duque-Lazo, Navarro-Cerrillo, 2017; Petrosyan *et al.*, 2023). Известно, что TSS и Kappa являются предпочтительными оценкам качества в моделях такого рода (Allouche *et al.*, 2006; Phillips, Elith, 2010; Penteciani *et al.*, 2019). Поскольку Kappa и COR подходят только для техник "присутствия/отсутствия" (Guisan *et al.*, 2017) для MaxEnt они не рассчитывались.

Для определения порога, который лучше всего различает присутствия и псевдоотсутствия для GLM (Shores *et al.*, 2019), выбирался тот, который максимизирует значение показателя Kappa, путем автоматического перебора всех возможных порогов от 0 до 1 с шагом 0.01 (Araújo *et al.*, 2005; Guisan *et al.*, 2017). При построении бинарных карт для MaxEnt в качестве порога для разделения

предсказаний на бинарные классы (присутствие/отсутствие вида) был использован порог maxSSS (maximum Sum of Sensitivity plus Specificity; Liu *et al.*, 2013; Recio *et al.*, 2020).

РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

Согласно общим оценкам качества модель "EH" оказалась лучшей (табл. 1). Для нее отмечены максимальные значения всех показателей (рис. 3). При выбранном пороге 0.12 значение максимальной Kappa составило 0.71. Это значительно выше, чем у других GLM при соответствующих максимизирующих порогах (рис. 4). Показатель TSS оказался еще выше (0.76) (рис. 3, 4), как и коэффициент бисериальной корреляции (COR = 0.72) (рис. 3). Значения AUC для всех GLM были очень высокие (рис. 3). В отличие от них модель на основе MaxEnt продемонстрировала наименьшие значения TSS и AUC (табл. 1).

Если рассматривать более подробно оценки качества дискриминации для моделей на основе GLM по выбранным порогам (см. табл. 1), то для модели "EH" во всех случаях они оказались лучше, чем для других сценариев (табл. 2). Так, для нее оказалась максимальная доля правильных предсказаний (CCR = 0.95) и минимальное отношение ошибочной классификации (MR = 0.05) (табл. 2). Эта же модель продемонстрировала лучшие значения чувствительности и специфичности. Она правильно распознавала 70% независимых точек присутствия и практически все независимые точки псевдоотсутствия (SP = 0.98). Особенно сильно модели различались по значению отношения шансов. Значение OR для модели "EH" намного превосходило таковое для других моделей (табл. 2).

На картах, построенных по непрерывным значениям предсказаний, наблюдается в целом одинаковый паттерн пространственного распределения черники по территории исследований. В то же время для модели "dist" отмечена наименьшая доля области распространения кустарничка, в то

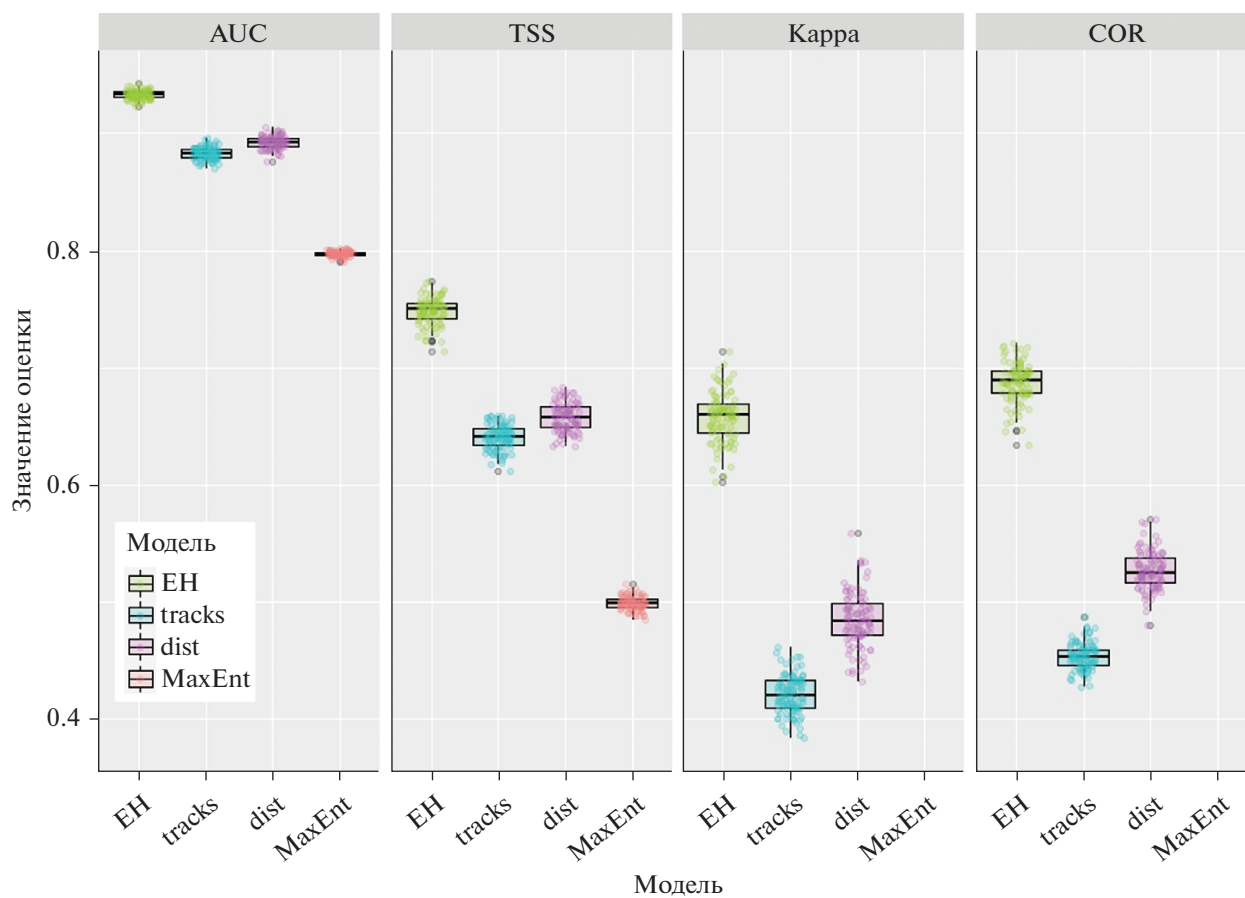


Рис. 3. Оценки качества моделей пространственного распределения черники на территории Центрально-Лесного заповедника и его охранный зоны для дизайнов GLM на основе подхода Энглера-Хенгла (EH), пройденных маршрутов (tracks), расстояний до точек присутствия (dist), а также для максимальной энтропии (MaxEnt).

время как для модели “MaxEnt” предсказания оказались наименее определенными (рис. 5). Наиболее реалистичные карты дали модели “EH” и “tracks”.

Если рассмотреть классифицированные карты по выбранным пороговым значениям предсказаний, то разница для GLM будет не столь заметна

(рис. 5). Для модели “EH” доля потенциальной области распространения черники составила 11.16% от всей территории исследований, для модели “tracks” – 11.82%, а для модели “dist” – 10.50%. Таким образом, дизайн моделирования GLM слабо влиял на предсказание общей площади распространения вида. В то же время по моде-

Таблица 2. Показатели качества дискриминации лучших моделей пространственного распределения черники на территории Центрально-Лесного заповедника и его охранный зоны по порогу, максимизирующему показатель Карра, для дизайнов GLM на основе подхода Энглера-Хенгла (EH), пройденных маршрутов (tracks) и расстояний до точек присутствия (dist)

Показатель качества	Модель		
Оценочный показатель	EH	tracks	dist
Доля правильных предсказаний (Correct classification rate, CCR)	0.95	0.88	0.92
Отношение ошибочной классификации (Misclassification rate, MR)	0.05	0.12	0.08
Чувствительность (Sensitivity, SE)	0.7	0.7	0.67
Специфичность (Specificity, SP)	0.98	0.90	0.95
Отношение шансов (Odds ratio, OR)	114.90	20.34	34.14

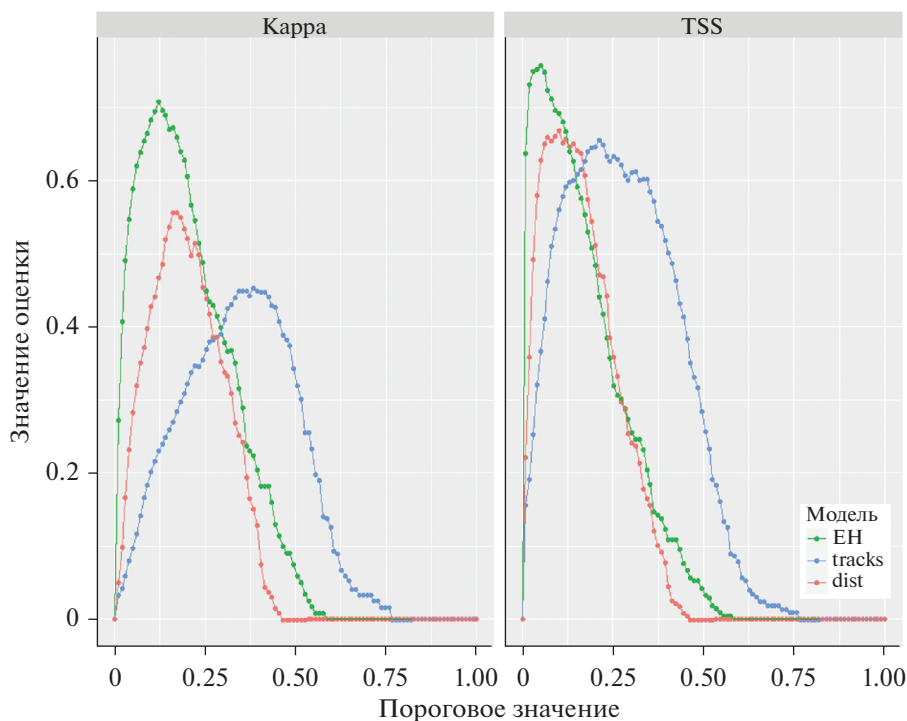


Рис. 4. Вариация значений Кappa и TSS в зависимости от пороговых значений для моделей пространственного распределения черники на территории Центрально-Лесного заповедника и его охранный зоны для дизайнов GLM на основе подхода Энглера-Хенгла (ЕН), пройденных маршрутов (tracks), расстояний до точек присутствия (dist).

ли MaxEnt доля территории потенциального присутствия черники оказалась больше всего и составила 31.69% (рис. 5).

ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

Большинство SDM предполагает, что вид, так или иначе, обитает в пригодных местообитаниях (где встречаются его точки присутствия) и не обитает в непригодных местах (где встречаются его точки отсутствия). Тем не менее, на практике такие отношения между экологической нишей и распределением вида не столь однозначны (Lobo *et al.*, 2010). Вид может отсутствовать в принципиально пригодных условиях и, наоборот, присутствовать в непригодных (Hirzel, Le Lay, 2008).

Ложная регистрация отсутствий вида (falls absences, FA) может быть вызвана разными причинами. Хирзель и Ле Лэй определили среди них следующие: географические барьеры или препятствия, мешающие нормальному распределению особей по потенциально пригодной территории; локальное вымирание, вызванное изменениями окружающей среды или стохастическими факторами; малая площадь территории, неспособная обеспечить жизнеспособную популяцию; альтернативные местообитания, которые могут использовать виды-генералисты, меняя их в течение одного сезона; биотические взаимодействия (сук-

цессионные стадии, конкуренция, хищничество и т.п.) (Hirzel, Le Lay, 2008). В случае животных их присутствие в подходящем местообитании также может быть не зарегистрировано из-за попытки животного избежать наблюдателей, либо даже из-за своеобразного поведения отдельных особей (Mertzanis *et al.*, 2008). Лобо с коллегами определяли такие отсутствия как случайные (contingent absences), т.е. вид отсутствует в потенциально пригодных экологических и климатических условиях в виду действия некоторых ограничивающих сил (Lobo *et al.*, 2010). Подобные просчеты неизбежны и являются существенной проблемой SDM-исследований (Le Maitre *et al.*, 2008; Lobo *et al.*, 2010; Piédallu *et al.*, 2017). Если такие ошибки наиболее вероятно ожидаемы, то следует применять дизайн только с данными о присутствии вида (Chefaoui, Lobo, 2008; Hirzel, Le Lay, 2008).

В отличие от ложных (случайных) отсутствий, истинные отсутствия (true absences, TA) определяются исключительно непригодными экологическими или климатическими условиями для вида. Лобо с соавторами называли их экологическими отсутствиями (environmental absences; Lobo *et al.*, 2010). Даже в этом случае причины их возникновения могут быть различными. Виды могут действительно отсутствовать, потому что среда обитания не подходит или потому, что они еще не колонизировали всю подходящую область или их отсутствие яв-

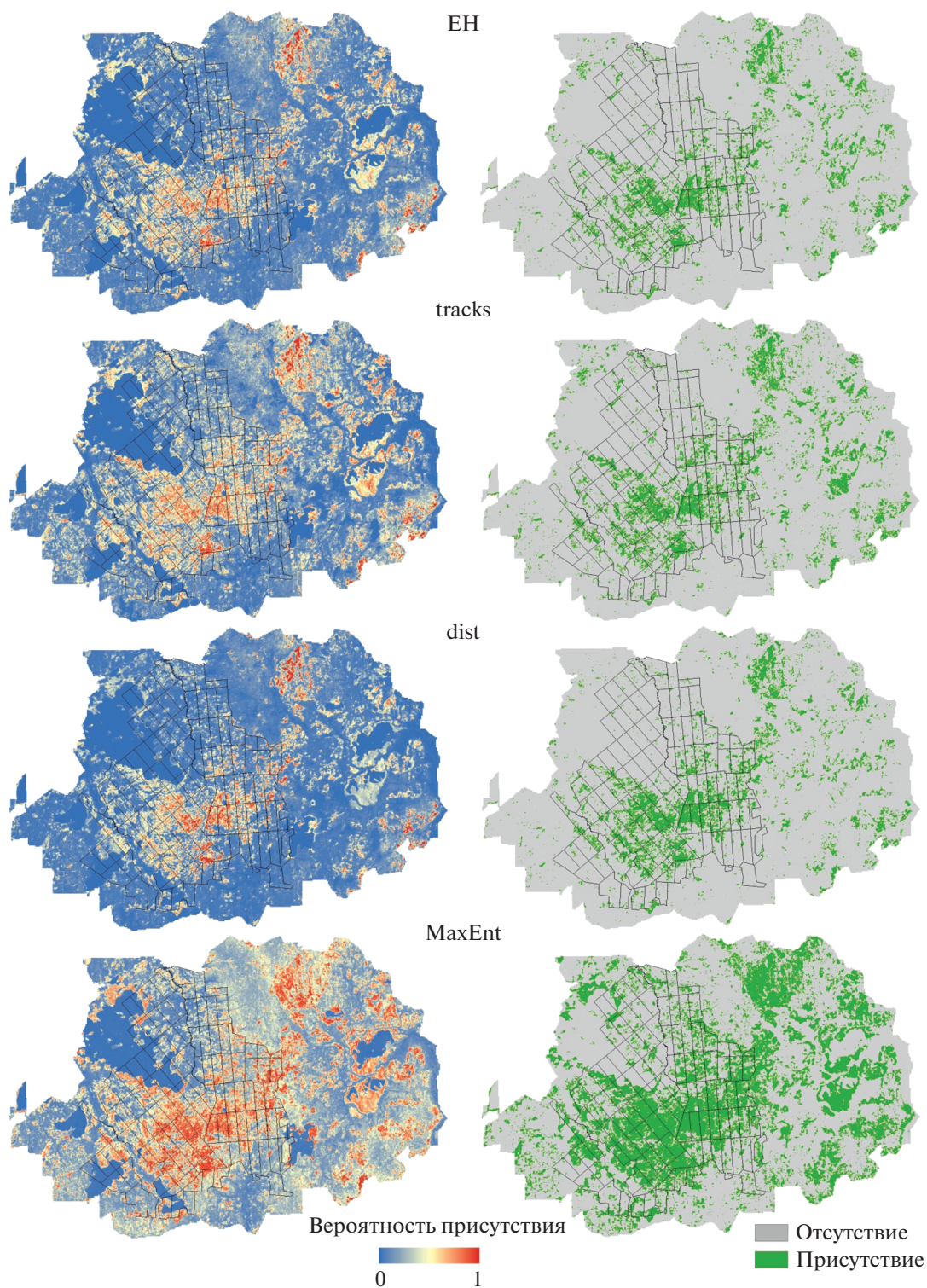


Рис. 5. Карты пространственного распределения черники на территории Центрально-Лесного заповедника и его охранный зоны для дизайнов GLM на основе подхода Энглера-Хенгла (ЕН), пройденных маршрутов (tracks), расстояний до точек присутствия (dist), а также для максимальной энтропии (MaxEnt). Слева карты непрерывных вероятностей присутствия, справа — классифицированные по выбранному порогу.

ляется временным (Martin *et al.*, 2005). По всем этим причинам отсутствие наблюдения в данном месте не может быть надежно интерпретировано как истинное отсутствие, поэтому в большинстве случаев можно полагаться только на данные о присутствии. Недостаток надежных данных об отсутствии является одной из основных проблем, которые должны решить экологи, занимающиеся построением SDM (Hirzel *et al.*, 2002; Soberón, Peterson, 2005). В большей степени это касается животных (особенно мобильных), а не растений.

Существует еще и третий тип отсутствий, в который объединены все ошибки при сборе пространственных данных, связанные с его неполнотой и предвзятостью. Они называются методологическими отсутствиями (methodological absences). Этот тип отсутствий может представлять собой наиболее важный источник неопределенности для изучения закономерностей и процессов, лежащих в основе географического распределения видов (Lobo *et al.*, 2010). Таким образом, в соответствии с концепцией ниш, предложенной Собероном и Накамурой (Soberón, 2007; Soberón, Nakamura, 2009), случайные отсутствия будут за пределами реализованной, но внутри фундаментальной ниши, экологические отсутствия будут за пределами как реализованной, так и фундаментальной ниши, а методологические отсутствия будут как в реализованной, так и в фундаментальной нише.

Вероятность возникновения каждого типа отсутствий на территории варьирует в зависимости от пространственной и экологической удаленности местоположения от условий, преобладающих в известных точках присутствия. Из этого следует, что экологические отсутствия будут более вероятны в тех местах, где экологические условия находятся далеко от комплекса условий, в которых был обнаружен вид. И наоборот, случайные отсутствия будут более вероятными в территориально удаленных местах с благоприятными условиями окружающей среды, тогда как вероятность обнаружения методологических отсутствий будет выше в экологически благоприятных местах, расположенных ближе всего к известным точкам присутствия (Lobo *et al.*, 2010).

Подход создания псевдоотсутствий Энглера-Хенгла позволяет максимально сократить вероятность возникновения случайных и методологических (т.е. ложных) отсутствий, повышая при этом вероятность возникновения экологических (т.е. истинных) отсутствий. При смещении отношения отсутствий от ложных к истинным, повышаются оценки качества моделей, рассчитанные на основе матрицы сопряженности (Guisan *et al.*, 2017). Этим же занижается вероятность возникновения присутствий в малопригодных местах (т.е. ложных присутствий; false presences, FP) и повышается вероятность возникновения присутствий в высоко-

пригодных местах (т.е. истинных присутствий; true presence, TP).

Существуют разные способы создания точек псевдоотсутствия. Самый распространенный из них — это случайная выборка из большого набора ячеек (сайтов) всей исследуемой области (Stockwell, Peters, 1999; Kobler, Adamic, 2000; Franklin, 2009). Такой метод рандомизации является наиболее простым и не требует особых допущений (Keating, Cherry, 2004). Кроме того, этот метод подходит для обычных видов, признаки присутствия которых легко идентифицировать. По сути, в таком виде данный отбор ничем не отличается от подхода “presence/background”. При этом в некоторых случаях ячейки, соответствующие присутствиям, могут быть удалены из случайного набора фона, или может быть исправлена некоторая систематическая ошибка выборки путем создания корректирующей области выбора (Phillips *et al.*, 2009; Barbet-Massin *et al.*, 2012). Известно, что географическая или экологическая область, в которой производится выборка этих псевдоотсутствий, существенно влияет на производительность модели (Zaniewski *et al.*, 2002; VanderWal *et al.*, 2009; Barbet-Massin *et al.*, 2012).

Есть множество других способов создания псевдоотсутствий, например, за пределами определенного буфера вокруг присутствий (Hirzel *et al.*, 2001; VanderWal *et al.*, 2009; Ghoddousi, 2010) или по определенной маске (Elith *et al.*, 2010), с помощью “point-process” моделей (Renner *et al.*, 2015), отбором на основе пороговых значений наиболее значимых предикторов (Le Maitre *et al.*, 2008) и другие (Zaniewski *et al.*, 2002; Engler *et al.*, 2004; Chefaoui, Lobo, 2008; Phillips *et al.*, 2009; Barbet-Massin *et al.*, 2012). Многие из этих способов, в частности случайный выбор точек по территории исследования, были подвергнуты обоснованной критике (Lobo *et al.*, 2010).

Согласно полученным результатам модель с псевдоотсутствиями, построенными на основе подхода Энглера-Хенгла, оказалась значительно лучше других моделей, в том числе и MaxEnt. При этом эта модель была без учета расстояний до точек присутствия. Для нее наблюдались максимальные значения оценочных показателей, в том числе и отношения шансов (OR). Отношение шансов показывает отношение произведений всех истинных предсказаний (TP · TA) к произведению всех ложных (FP · FA). Чем оно больше, тем в модели в целом больше истинных (т.е. правильных) предсказаний. Дизайны на основе треков и расстояний оказались значительно хуже. Наиболее реалистичные карты получились также для модели на основе подхода Энглера-Хенгла.

Как правило, оценки качества моделей MaxEnt превосходят таковые GLM (например, Guisan *et al.*, 2007; Nath *et al.*, 2019). При этом известно,

что значения AUC для моделей MaxEnt обычно завышены относительно таковых для GLM (Guisan *et al.*, 2007), поэтому сравнивать эти две техники моделирования только по показателю AUC не совсем корректно. В то же время, если для построения GLM используются реальные отсутствия, то качество таких моделей превосходит MaxEnt (Duque-Lazo *et al.*, 2016). В нашем случае, используя сгенерированные псевдоотсутствия по методу Энглера-Хенгла, удалось построить GLM, которая значительно превосходила MaxEnt не только по значению AUC, но и по TSS.

Построенные карты пространственного распределения черники наглядно демонстрируют, что способ построения GLM на основе подхода Энглера-Хенгла существенно сужает область потенциального присутствия вида. Модель MaxEnt, наоборот, предсказывает наибольшую область пространственного распределения черники, вовлекая не только ее типичные местообитания, но и много смежных биотопов, где присутствие черники маловероятно. Итоговая площадь территории, потенциально пригодной для кустарничка, оказалась вдвое большей, чем по оценкам GLM. Таким образом, MaxEnt строит менее определенные и более “размытые” карты, завышая общие вероятности относительного присутствия. Это может вызывать сложности для выделения областей потенциального присутствия достаточно стенопотных или редких видов. Подобное поведение модели вполне логично и предсказуемо, ведь значения экогеографических переменных в точках присутствия соотносятся с таковыми в случайно выбранных точках фона, а не в точках отсутствия. В конечном итоге это приводит к сильной переоценке реальной области распространения вида. В GLM, напротив, значения переменных окружающей среды в точках присутствия противопоставляются таким значениям в максимально удаленных в экологическом пространстве точках.

Выбор числа точек псевдоотсутствия также является дискуссионным вопросом. Как отмечают Манли с соавторами очень важно минимизировать ошибки выборки, отбирая данные таким образом, чтобы они были полностью репрезентативными для исследуемой области (Manly *et al.*, 2002). Это означает, что достаточно большое количество точек выбирается случайным образом из ландшафта, чтобы контрастировать с точками присутствия. МакДональд предлагает использовать на несколько порядков больше псевдоотсутствий, чем присутствий при применении экспоненциальной модели (McDonald, 2003). Барбет-Массин с коллегами установили, что для GLM и GAM наилучшие результаты предсказаний достигаются при использовании большого количества псевдоотсутствий (например, 10000) с одинаковыми весами присутствия и отсутствия (Barbet-

Massin *et al.*, 2012). Это согласуется с выводами относительно числа точек в MaxEnt (Phillips, Dudík, 2008). Шефауи и Лобо также советуют брать большее число точек псевдоотсутствий, но предупреждают от экстремальных значений (Chefaoui, Lobo, 2008). В этом случае оптимальным считается число псевдоотсутствий в 10 раз большее, чем число точек присутствий (Chefaoui, Lobo, 2008; Jiménez-Valverde *et al.*, 2008). Влияние числа псевдоотсутствий на качество моделей на основе подхода Энглера-Хенгла является актуальной проблемой для следующих исследований.

ЗАКЛЮЧЕНИЕ

Гизан с коллегами подчеркивали два ключевых правила для уменьшения последующих проблем со статистическим методом, который использует точки присутствия в сочетании с точками псевдоотсутствий (Guisan *et al.*, 2017). Во-первых, важно ограничить область исследования реалистичной областью. Было показано, что использование большого, но нереалистичного размера для выборки псевдоотсутствий отрицательно влияет на модель и ее прогнозы (VanderWal *et al.*, 2009; Elith *et al.*, 2010). Во-вторых, случайная выборка псевдоотсутствий – это стратегия с наименьшими предположениями, и ее следует использовать по умолчанию, если нет веских аргументов в пользу другого, более специфичного подхода (Barbet-Massin *et al.*, 2012). Используемый подход Энглера-Хенгла с весами MaxEnt позволяет учесть оба этих правила. Область отбора точек псевдоотсутствий имеет четкие экологические ограничения, а стратифицированный дизайн создания точек позволяет выявлять наиболее реалистичные места потенциального отсутствия вида. На основе такого подхода удастся строить модели очень хорошего качества согласно независимым оценкам, которые превосходят таковые на основе только лишь MaxEnt. В то же время использование параметра расстояния в условиях мозаичного ландшафта весьма спорно и требует дополнительных проверок.

СПИСОК ЛИТЕРАТУРЫ

- Желтухин А.С., Пузаченко Ю.Г., Сандлерский Р.Б. Оценка качества местообитаний животных на основе учетов следовой активности и дистанционной информации // Сибирский экологический журн. 2009. № 3. С. 341–351.
- Огурцов С.С. Моделирование пригодности местообитаний бурого медведя *Ursus arctos* (Linnaeus, 1758) на основе функции выбора ресурсов в мозаичных ландшафтах южной тайги. Автореф. дис. канд. биол. наук. М.: Институт проблем экологии и эволюции им. А.Н. Северцова РАН, 2023. 26 с.
- Пузаченко Ю.Г., Желтухин А.С., Сандлерский Р.Б. Анализ пространственно-временной динамики эко-

- логической ниши на примере популяции лесной куницы (*Martes martes*) // Журн. общей биологии. 2010. Т. 71(6). С. 467–487.
- Пузаченко Ю.Г., Кузьмин С.Л., Сандлерский Р.Б. Количественная оценка параметров ареалов (на примере представителей рода *Rana*) // Журн. общей биологии. 2011. Т. 72(5). С. 339–354.
- Пузаченко Ю.Г., Желтухин А.С., Козлов Д.Н., Кораблев Н.П., Федяева М.В., Пузаченко М.Ю., Сиунова Е.В. Центрально-Лесной государственный природный биосферный заповедник. Научно-популярный очерк. Издание 2-е. Тверь: ООО “Печатня”, 2016. 80 с.
- Aiello-Lammens M.E., Boria R.A., Radosavljevic A., Vilela B., Anderson R.P. spThin: An R package for spatial thinning of species occurrence records for use in ecological niche models // *Ecography*. 2015. V. 38. P. 541–545. <https://doi.org/10.1111/ecog.01132>
- Allouche O., Tsoar A., Kadmon R. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS) // *J. Applied Ecology*. 2006. V. 43. P. 1223–1232. <https://doi.org/10.1111/j.1365-2664.2006.01214.x>
- Araújo M.B., Anderson R., Barbosa A.M., Beale C., Dormann C., Early R., Garcia R., Guisan A., Maiorano L., Naimi B., O'Hara R., Zimmermann N., Rahbek C. Standards for distribution models in biodiversity assessments // *Science Advances*. 2019. V. 5. eaat4858. <https://doi.org/10.1126/sciadv.aat4858>
- Araújo M.B., Pearson R.G., Thuiller W., Erhard M. Validation of species-climate impact models under climate change // *Global Change Biology*. 2005. V. 11. P. 1504–1513. <https://doi.org/10.1111/j.1365-2486.2005.01000.x>
- Baig M.H.A., Zhang L., Shuai T., Tong Q. Derivation of a tasselled cap transformation based on Landsat 8 at-satellite reflectance // *Remote Sensing Letters*. 2014. V. 5(5). P. 423–431. <https://doi.org/10.1080/2150704X.2014.915434>
- Barbet-Massin M., Jiguet F., Alberti C.H., Thuiller W. Selecting pseudo-absences for species distribution models: how, where and how many? // *Methods in Ecology and Evolution*. 2012. V. 3. P. 327–338. <https://doi.org/10.1111/j.2041-210X.2011.00172.x>
- Boyce M.S., Vernier P.R., Nielsen S.E., Schmiegelow F.K.A. Evaluating resource selection functions // *Ecological Modelling*. 2002. V. 157(2–3). P. 281–300. [https://doi.org/10.1016/S0304-3800\(02\)00200-4](https://doi.org/10.1016/S0304-3800(02)00200-4)
- Broennimann O. Package ‘ecospat’, version 3.0. *Spatial Ecology Miscellaneous Methods*. 2018. 107 p.
- Chefaoui R.M., Lobo J.M. Assessing the effects of pseudo-absences on predictive distribution model performance // *Ecological Modelling*. 2008. V. 210. P. 478–486. <https://doi.org/10.1016/j.ecolmodel.2007.08.010>
- Conrad O., Bechtel B., Bock M., Dietrich H., Fischer E., Gerlitz L., Wehberg J., Wichmann V., Boehner J. System for Automated Geoscientific Analyses (SAGA) v. 2.1.4 // *Geoscientific Model Development*. 2015. V. 8. P. 1991–2007. <https://doi.org/10.5194/gmd-8-1991-2015>
- Coudun C., Gégout J.-C. Quantitative prediction of the distribution and abundance of *Vaccinium myrtillus* with climatic and edaphic factors // *J. Vegetation Science*. 2007. V. 18. P. 517–524. [https://doi.org/10.1658/1100-9233\(2007\)18\[517:QPOT-DA\]-2.0.CO;2](https://doi.org/10.1658/1100-9233(2007)18[517:QPOT-DA]-2.0.CO;2)
- Dettmers R., Bart J. A GIS modeling method applied to predicting forest songbird habitat // *Ecological Applications*. 1999. V. 9. P. 152–163. [https://doi.org/10.1890/1051-0761\(1999\)009\[0152:AGMMAT\]2.0.CO;2](https://doi.org/10.1890/1051-0761(1999)009[0152:AGMMAT]2.0.CO;2)
- Duque-Lazo J., Navarro-Cerrillo R.M. What to save, the host or the pest? The spatial distribution of xylophage insects within the Mediterranean oak woodlands of Southwestern Spain // *Forest Ecology and Management*. 2017. V. 392. P. 90–104. <https://doi.org/10.1016/j.foreco.2017.02.047>
- Duque-Lazo J., van Gils H., Groen T.A., Navarro-Cerrillo R.M. Transferability of species distribution models: The case of *Phytophthora cinnamomi* in Southwest Spain and Southwest Australia // *Ecological Modelling*. 2016. V. 320. P. 62–70. <https://doi.org/10.1016/j.ecolmodel.2015.09.019>
- Elith J., Graham C.H., Anderson R.P., Dudík M., Ferrier S., Guisan A., Hijmans R.J., Huettmann F., Leathwick J.R., Lehman A., Li J., Lohmann L.G., Loiselle B.A., Manion G., Moritz C., Nakamura M., Nakazawa Y., Overton J.M.M., Peterson A.T., Phillips S.J., Richardson K., Scachetti-Pereira R., Schapire R.E., Soberón J., Williams S., Wisz M.S., Zimmermann N.E. Novel methods improve prediction of species' distributions from occurrence data // *Ecography*. 2006. V. 29 (2). P. 129–151. <https://doi.org/10.1111/j.2006.0906-7590.04596.x>
- Elith J., Phillips S.J., Hastie T., Dudík M., Chee Y.E., Yates C.J. A statistical explanation of MaxEnt for Ecologists // *Diversity and Distributions*. 2011. V. 17. P. 43–57. <https://doi.org/10.1111/j.1472-4642.2010.00725.x>
- Elith J., Kearney M., Phillips S. The art of modelling range-shifting species // *Methods in Ecology and Evolution*. 2010. V. 1. P. 330–342. <https://doi.org/10.1111/j.2041-210X.2010.00036.x>
- Engler R., Guisan A., Rechsteiner L. An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudoabsence data // *J. Applied Ecology*. 2004. V. 41(2). P. 263–274. <https://doi.org/10.1111/j.0021-8901.2004.00881.x>
- Evans J.S., Murphy M.A., Ram K. Package ‘spatialEco’. Spatial analysis and modelling utilities. Version 1.3-7. 2021. 181 p.
- Fox J., Weisberg S. An R Companion to Applied Regression, Third Edition. Thousand Oaks CA: Sage Publications Inc., 2018. 608 p.
- Franklin J. Mapping species distributions. Cambridge: Cambridge University Press, 2009. 320 p. <https://doi.org/10.1017/CBO9780511810602>
- Freeman E.A., Moisen G. PresenceAbsence: An R package for presence-absence model analysis // *J. Statistical Software*. 2008. V. 23(11). P. 1–31. <https://doi.org/10.18637/JSS.V023.I11>
- Ghoddousi A. Habitat suitability modelling of the Brown bear *Ursus arctos* in Croatia and Slovenia using telemetry data. Master of Science thesis. Imperial College London, 2010. 71 p.
- Glenz C., Massolo A., Kuonen D., Schlaepfer R. A wolf habitat suitability prediction study in Valais (Switzerland) // *Landscape and Urban Planning*. 2001. V. 55. P. 55–65. [https://doi.org/10.1016/S0169-2046\(01\)00119-0](https://doi.org/10.1016/S0169-2046(01)00119-0)
- Guisan A., Thuiller W., Zimmermann N.E. Habitat suitability and distribution models. Cambridge: Cambridge University Press, 2017. 462 p. <https://doi.org/10.1017/9781139028271>

- Guisan A., Zimmermann N. Predictive habitat distribution models in ecology // *Ecological Modelling*. 2000. V. 135. P. 147–186.
[https://doi.org/10.1016/S0304-3800\(00\)00354-9](https://doi.org/10.1016/S0304-3800(00)00354-9)
- Guisan A., Zimmermann N.E., Elith J., Graham C.H., Phillips S., Peterson A.T. What matters for predicting the occurrences of trees: techniques, data, or species' characteristics? // *Ecological Monographs*. 2007. V. 77(4). P. 615–630.
<https://doi.org/10.1890/06-1060.1>
- Hengl T., Sierdsema H., Radovic A., Dilo A. Spatial prediction of species' distributions from occurrence-only records: combining point pattern analysis, ENFA and regression-kriging // *Ecological Modelling*. 2009. V. 220. P. 3499–3511.
<https://doi.org/10.1016/j.ecolmodel.2009.06.038>
- Hijmans R.J., Phillips S., Leathwick J., Elith J. Package 'dismo'. Species Distribution Modeling. Version 1.1-4. 2017. 68 p.
- Hirzel A.H., Hausser J., Chessel D., Perrin N. Ecological-niche factor analysis: how to compute habitat-suitability maps without absence data? // *Ecology*. 2002. V. 83(7). P. 2027–2036.
[https://doi.org/10.1890/0012-9658\(2002\)083\[2027:ENFAHT\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2002)083[2027:ENFAHT]2.0.CO;2)
- Hirzel A.H., Le Lay G. Habitat suitability modelling and niche theory // *Journal of Applied Ecology*. 2008. V. 45. P. 1372–1381.
<https://doi.org/10.1111/j.1365-2664.2008.01524.x>
- Hirzel A.H., Helfer V., Metral F. Assessing habitat-suitability models with a virtual species // *Ecological modelling*. 2001. V. 145. P. 111–121.
[https://doi.org/10.1016/S0304-3800\(01\)00396-9](https://doi.org/10.1016/S0304-3800(01)00396-9)
- Huete A., Didan K., Miura T., Rodriguez E., Gao X., Ferreira L.G. Overview of the radiometric and biophysical performance of the MODIS Vegetation Indices // *Remote Sensing of Environment*. 2002. V. 83. P. 195–213.
[https://doi.org/10.1016/S0034-4257\(02\)00096-2](https://doi.org/10.1016/S0034-4257(02)00096-2)
- Jerina K., Debeljak M., Dzeroske S., Kobler A., Adamic M. Modeling the Brown bear population in Slovenia. A tool in the conservation management of a threatened species // *Ecological modelling*. 2003. V. 170. P. 453–469.
[https://doi.org/10.1016/S0304-3800\(03\)00245-X](https://doi.org/10.1016/S0304-3800(03)00245-X)
- Jiménez-Valverde A., Gómez J.F., Lobo J.M., Baselga A., Hortal J. Challenging species distribution models: the case of *Maculinea nausithous* in the Iberian Peninsula // *Annales Zoologici Fennici*. 2008. V. 45. P. 200–210.
<https://doi.org/10.5735/086.045.0305>
- Johnson C.J., Nielsen S.E., Merrill E.H., McDonald T.L., Boyce M.S. Resource selection functions based on use-availability data: theoretical motivation and evaluation methods // *J. Wildlife Management*. 2006. V. 70(2). P. 347–357.
[https://doi.org/10.2193/0022-541X\(2006\)70\[347:RS-FBOU\]2.0.CO;2](https://doi.org/10.2193/0022-541X(2006)70[347:RS-FBOU]2.0.CO;2)
- Kauth R.J., Thomas G.S. The Tasseled Cap – a graphic description of the spectral- temporal development of agricultural crops as seen by Landsat // *Proceedings of the Symposium on Machine Processing of Remotely Sensed Data*. US, Indiana, West Lafayette: Purdue University, 1976. P. 4B41–4B51.
- Keating K., Cherry S. Use and interpretation of logistic regression in habitat selection studies // *Journal of Wildlife Management*. 2004. V. 68. P. 774–789.
[https://doi.org/10.2193/0022-541X\(2004\)068\[0774:UAIOLR\]2.0.CO;2](https://doi.org/10.2193/0022-541X(2004)068[0774:UAIOLR]2.0.CO;2)
- Kéry M., Gardner B., Stoeckle T., Weber D., Royle J.A. Use of spatial capture-recapture modeling and DNA data to estimate densities of elusive animals // *Conservation Biology*. 2010. V. 25. P. 356–364.
<https://doi.org/10.1111/j.1523-1739.2010.01616.x>
- Kobler A., Adamic M. Identifying brown bear habitat by a combined GIS and machine learning method // *Ecological Modelling*. 2000. V. 135. P. 291–300.
[https://doi.org/10.1016/S0304-3800\(00\)00384-7](https://doi.org/10.1016/S0304-3800(00)00384-7)
- Landsat 8 (L8) Data Users Handbook. Version 5.0. Department of the Interior U.S. Geological Survey. 2019. 106 p.
- Le Maitre D.C., Thuiller W., Schonegevel L. Developing an approach to defining the potential distributions of invasive plant species: a case study of *Hakea* species in South Africa // *Global Ecology and Biogeography*. 2008. V. 17. P. 569–584.
<https://doi.org/10.1111/j.1466-8238.2008.00407.x>
- Liu C., White M., Newell G. Selecting thresholds for the prediction of species occurrence with presence-only data // *J. Biogeography*. 2013. V. 40. P. 778–789.
<https://doi.org/10.1111/jbi.12058>
- Lobo J., Jimenez-Valverde A., Hortal J. The uncertain nature of absences and their importance in species distribution modelling // *Ecography*. 2010. V. 33 (1). P. 103–114.
<https://doi.org/10.1111/j.1600-0587.2009.06039.x>
- MacKenzie D.I., Royle J.A. Designing occupancy studies: general advice and allocating survey effort // *J. Applied Ecology*. 2005. V. 42. P. 1105–1114.
<https://doi.org/10.1111/j.1365-2664.2005.01098.x>
- Maiorano L., Boitani L., Monaco A., Tosoni E., Ciucci P. Modeling the distribution of Apennine brown bears during hyperphagia to reduce the impact of wild boar hunting // *European J. Wildlife Research*. 2015. V. 61(2). P. 241–253.
<https://doi.org/10.1007/s10344-014-0894-0>
- Manly B.F.J., McDonald L.L., Thomas D.L., McDonald T.L., Erickson W.P. Resource Selection by Animals. Statistical Design and Analysis for Field Studies. Second Edition. Kluwer Academic Publishers, 2002. 221 p.
<https://doi.org/10.2307/5247>
- Martin T.G., Wintle B.A., Rhodes J.R., Kuhnert P.M., Field S.A., Low-Choy S.J., Tyre A.J., Possingham H.P. Zero tolerance ecology: improving ecological inference by modelling the source of zero observations // *Ecology Letters*. 2005. V. 8. P. 1235–1246.
<https://doi.org/10.1111/j.1461-0248.2005.00826.x>
- McClelland C.J.R., Coops N.C., Kearney S.P., Burton A.C., Nielsen S.E., Stenhouse G.B. Variations in grizzly bear habitat selection in relation to the daily and seasonal availability of annual plant-food resources // *Ecological Informatics*. 2020. V. 58. P. 101116.
<https://doi.org/10.1016/j.ecoinf.2020.101116>
- McDonald T.L. Estimation of resource selection functions when used and available samples overlap // *Resource selection methods and applications* / Ed. Huzurbazar S. Madison, Wisconsin: Omnipress, 2003. P. 35–39.
- Merow C., Smith M.J., Silander J.A.Jr. A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter // *Ecography*. 2013. V. 36. P. 1058–1069.
<https://doi.org/10.1111/j.1600-0587.2013.07872.x>
- Mertzanis G., Kallimanis A.S., Kanellopoulos N., Sgardelis S.P., Tragos A., Aravidis I. Brown bear (*Ursus arctos*) habitat

- use patterns in two regions of northern Pindos, Greece – management implications // *J. Natural History*. 2008. V. 42. P. 301–315.
<https://doi.org/10.1080/00222930701835175>
- Mladenoff D.J., Sickley T.A., Wydeven A.P. Predicting grey wolf landscape colonization: logistic regression models vs. new field data // *Ecological Applications*. 1999. V. 9. P. 37–44.
[https://doi.org/10.1890/1051-0761\(1999\)009\[0037:PGWRLR\]2.0.CO;2](https://doi.org/10.1890/1051-0761(1999)009[0037:PGWRLR]2.0.CO;2)
- Moore I.D., Grayson R.B., Ladson A.R. Digital terrain modelling: A review of hydrological, geomorphological, and biological applications // *Hydrological Processes*. 1991. V. 5(1). P. 3–30.
<https://doi.org/10.1002/hyp.3360050103>
- Naimi B., Hamm N., Groen T.A., Skidmore A.K., Toxopeus A.G. Where is positional uncertainty a problem for species distribution modelling // *Ecography*. 2014. V. 37. P. 191–203.
<https://doi.org/10.1111/j.1600-0587.2013.00205.x>
- Nath A., Sinha A., Lahkar B.P., Brahma N. In search of Aliens: Factors influencing the distribution of *Chromolaena odorata* L. and *Mikania micrantha* Kunth in the Terai grasslands of Manas National Park, India // *Ecological Engineering journal*. 2019. V. 131. P. 16–26.
<https://doi.org/10.1016/j.ecoleng.2019.02.012>
- Nielsen S.E., Boyce M.S., Stenhouse G.B., Munro R.H.M. Development and testing of phenologically driven grizzly bear habitat models // *Écoscience*. 2003. V. 10. P. 1–10.
<https://doi.org/10.1080/11956860.2003.11682743>
- Nielsen S.E., Larsen T.A., Stenhouse G.B., Coogan S.C.P. Complementary food resources of carnivory and frugivory affect local abundance of an omnivorous carnivore // *Oikos*. 2016. V. 126(3). P. 1–12.
<https://doi.org/10.1111/oik.03144>
- Pal M., Mather P.M. An assessment of the effectiveness of decision tree methods for land cover classification // *Remote Sensing of Environment*. 2003. V. 86(4). P. 554–565.
[https://doi.org/10.1016/S0034-4257\(03\)00132-9](https://doi.org/10.1016/S0034-4257(03)00132-9)
- Penteriani V., Zarzo-Arias A., Novo-Fernández A., Bombieri G., López-Sánchez C.A. Responses of an endangered brown bear population to climate change based on predictable food resource and shelter alterations // *Global Change Biology*. 2019. V. 25(3). P. 1133–1151.
<https://doi.org/10.1111/gcb.14564>
- Peterson A.T., Soberón J., Pearson R.G., Anderson R.P., Martínez-Meyer E., Nakamura M., Araújo M.B. *Ecological Niches and Geographic Distributions*. Princeton: Princeton University Press, 2011. 316 p.
- Petrosyan V., Osipov F., Bobrov V., Dergunova N., Nazarenko E., Omelchenko A., Danielyan F., Arakelyan M. Analysis of geographical distribution of the parthenogenetic rock lizard *Darevskia armeniaca* and its parental species (*D. mixta*, *D. valentini*) based on ecological modelling // *Salamandra*. 2019. V. 55(3). P. 173–190.
- Petrosyan V., Dinets V., Osipov F., Dergunova N., Khlyap L. Range Dynamics of Striped Field Mouse (*Apodemus agrarius*) in Northern Eurasia under Global Climate Change Based on Ensemble Species Distribution Models // *Biology*. 2023. V. 12(7). P. 1–30.
<https://doi.org/10.3390/biology12071034>
- Phillips S.J., Anderson R.P., Dudík M., Schapire R.E., Blair M.E. Opening the black-box: an open-source release of Maxent // *Ecography*. 2017. V. 40. P. 887–893.
<https://doi.org/10.1111/ecog.03049>
- Phillips S.J., Anderson R.P., Schapire R.E. Maximum entropy modeling of species geographic distributions // *Ecological Modelling*. 2006. V. 190. P. 231–259.
<https://doi.org/10.1016/j.ecolmodel.2005.03.026>
- Phillips S.J., Dudík M. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation // *Ecography*. 2008. V. 31. P. 161–175.
<https://doi.org/10.1111/j.2007.0906-7590.05203.x>
- Phillips S.J., Dudík M., Elith J., Graham C.H., Lehmann A., Leathwick J., Ferrier S. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data // *Ecological Applications*. 2009. V. 19(1). P. 181–197.
<https://doi.org/10.1890/07-2153.1>
- Phillips S.J., Dudík M., Schapire R.E. Maxent software for modeling species niches and distributions (Version 3.4.1). 2018. Accessed at http://biodiversityinformatics.amnh.org/open_source/maxent/.
- Phillips S.J., Elith J. POC plots: calibrating species distribution models with presence-only data // *Ecology*. 2010. V. 91. P. 2476–2484.
<https://doi.org/10.1890/09-0760.1>
- Pié d'allu B., Quenette P.Y., Bombillon N., Gastineau A., Miquel C., Gimenez O. Determinants and patterns of habitat use by the brown bear *Ursus arctos* in the French Pyrenees revealed by occupancy modelling // *Oryx. Fauna & Flora International*. 2017. V. 53(2). P. 1–10.
<https://doi.org/10.1017/s0030605317000321>
- R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria, 2020. Accessed at <https://www.R-project.org/>.
- Recio M.R., Knauer F., Molinari-Jobin A., Huber D., Filacorda S., Jerina K. Context-dependent behaviour and connectivity of recolonizing brown bear populations identify transboundary conservation challenges in Central Europe // *Animal Conservation*. 2020. V. 24(1). P. 73–83.
<https://doi.org/10.1111/acv.12624>
- Renner I.W., Elith J., Baddeley A., Fithian W., Hastie T., Phillips S.J., Popovic G., Warton D.I. Point process models for presence-only analysis // *Methods in Ecology and Evolution*. 2015. V. 6. P. 366–379.
<https://doi.org/10.1111/2041-210X.12352>
- Roberts D.R., Nielsen S.E., Stenhouse G.B. Idiosyncratic responses of grizzly bear habitat to climate change based on projected changes in their food resources // *Ecological Applications*. 2014. V. 24(5). P. 1144–1154.
<https://doi.org/10.1890/13-0829.1>
- Robin X., Turck N., Hainard A., Tiberti N., Lisacek F., Sanchez J.-C., Müller M., Siegert S., Doering M. Package ‘pROC’, version 1.16.2. Display and Analyze ROC Curves. 2020. 95 p.
- Shores C.R., Mickle N., Graves T.A. Mapping a keystone shrub species, huckleberry (*Vaccinium membranaceum*), using seasonal colour change in the Rocky Mountains // *International J. Remote Sensing*. 2019. V. 40(15). P. 5695–5715.
<https://doi.org/10.1080/01431161.2019.1580819>
- Sillero N., Barbosa A.M. Common mistakes in ecological niche models // *International J. Geographical Information Science*. 2021. V. 35(2). P. 213–226.
<https://doi.org/10.1080/13658816.2020.1798968>

- Smulders M., Nelson T.A., Jelinski D.E., Nielsen S.E., Stenhouse G.B.* A spatially explicit method for evaluating accuracy of species distribution models // *Diversity and Distributions*. 2010. V. 16. P. 996–1008.
<https://doi.org/10.1111/j.1472-4642.2010.00707.x>
- Soberón J.* Grinnellian and Eltonian niches and geographic distributions of species // *Ecology Letters*. 2007. V. 10(12). P. 1115–1123.
<https://doi.org/10.1111/j.1461-0248.2007.01107.x>
- Soberón J., Peterson A.T.* Interpretation of models of fundamental ecological niches and species' distributional areas // *Biodiversity Informatics*. 2005. V. 2. P. 1–10.
<https://doi.org/10.17161/bi.v2i0.4>
- Soberón J., Nakamura M.* Niches and distributional areas: concepts, methods, and assumptions // *Proceedings of National Academy of Science. USA*. 2009. V. 106. P. 19644e19650.
<https://doi.org/10.1073/pnas.0901637106>
- Stockwell D., Peters D.* The GARP modelling system: problems and solutions to automated spatial prediction // *International J. Geographical Information Sciences*. 1999. V. 2. P. 143–158.
<https://doi.org/10.1080/136588199241391>
- Syfert M.M., Smith M.J., Coomes D.A.* The effects of sampling bias and model complexity on the predictive performance of MaxEnt species distribution models // *PLoS ONE*. 2013. V. 8. P. e55158.
<https://doi.org/10.1371/journal.pone.0055158>
- Thuiller W., Georges D., Engler R., Breiner F.* biomod2: ensemble platform for species distribution modeling. R package version 3.4.6. 2020. 105 p.
- Valavi R., Elith J., Lahoz-Monfort J.J., Guillera-Aroita G.* blockCV: An R package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models // *Methods in Ecology and Evolution*. 2019. V. 10. P. 225–232.
<https://doi.org/10.1111/2041-210X.13107>
- VanDerWal J., Shoo L.P., Johnson C.N., Williams S.E.* Abundance and the environmental niche: Environmental suitability estimated from niche models predicts the upper limit of local abundance // *American Naturalist*. 2009. V. 174. P. 282–291.
<https://doi.org/10.1086/600087>
- Vermote E., Justice C., Claverie M., Franch B.* Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product // *Remote Sensing of Environment*. 2016. V. 185. P. 46–56.
<https://doi.org/10.1016/j.rse.2016.04.008>
- Vignali S., Barras A.G., Arlettaz R., Braunisch V.* SDM tune: An R package to tune and evaluate species distribution models // *Ecology and Evolution*. 2020. V. 10(20). P. 11488–11506.
<https://doi.org/10.1002/ece3.6786>
- Vignali S., Barras A., Braunisch V.* Package ‘SDM tune’. Species distribution model selection. Version 1.1.3. 2020a. 70 p.
- Zaniewski A.E., Lehmann A., Overton J.McC.* Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns // *Ecological Modelling*. 2002. V. 157. P. 261–280.
[https://doi.org/10.1016/S0304-3800\(02\)00199-0](https://doi.org/10.1016/S0304-3800(02)00199-0)
- Zuur A.F., Ieno E.N., Walker N.J., Saveliev A.A., Smith G.M.* Mixed Effects Models and Extensions in Ecology with R. New York: Springer, 2009. 574 p.
<https://doi.org/10.1007/978-0-387-87458-6>

Absence of the “Absences”: the Engler-Hengl Approach in Species Distribution Modelling

S. S. Ogurtsov^{1, 2, #}

¹Central Forest State Nature Biosphere Reserve, Zapovednyi, 32, Nelidovo District, Tver Region, 172521 Russia

²Severtsov Institute of Ecology and Evolution, Russian Academy of Sciences, Leninsky prospect, 33, Moscow, 119071 Russia

#e-mail: etundra@mail.ru

The possibilities of creating artificial absence points (pseudo-absences) to build species distribution models are considered. An approach proposed by Robin Engler and adapted by Tomislav Hengl is described, which takes into account habitat suitability indices and distances to presence points to create pseudo-absences. Using the example of bilberries (*Vaccinium myrtillus*) in the Central Forest Nature Reserve and its buffer zone, generalized linear models based on the Engler-Hengl design, traveled tracks and distances to presence points are compared, as well as a model built using the maximum entropy method. The results obtained indicate the superiority of the model based on the Engler-Hengl approach both in terms of quality assessments and in terms of the realism of the constructed spatial distribution maps.

Keywords: habitat suitability, spatial distribution, species distribution model, pseudo-absences, GLM, MaxEnt