

Главный ботанический сад имени Н. В. Цицина РАН

Московская школа на Юго-Западе №1543

Исследование применимости метода
биоклиматического моделирования к мохообразным
на примере рода *Tetraphis*

Анастасия Захарова

Научный руководитель:

д.б.н. Михаил Станиславович Игнатов

Москва, 2019

Содержание

Введение	3
Материалы и методы	7
Метод	7
Материалы	8
<u><i>Tetraphis pellucida</i></u>	9
<u><i>Tetraphis geniculata</i></u>	10
Результаты и обсуждение	11
Моделирование общего распространения <i>T. pellucida</i>	11
<i>Оценка пригодности моделей</i>	11
<i>Построенные модели</i>	12
<i>Анализ вклада переменных</i>	14
Моделирование распространения <i>T. pellucida</i> на территории России	18
<i>Построенные модели</i>	19
<i>Оценка пригодности моделей</i>	21
<i>Сравнение спрогнозированного распространения <i>T. pellucida</i> на территории России с реальным</i>	22
<i>Анализ вклада переменных</i>	24
Моделирование распространения <i>T. pellucida</i> по всему миру, только файл gar100	27
Моделирование общего распространения <i>T. geniculata</i>	28
<i>Построенные модели</i>	29
<i>Анализ вклада переменных</i>	31
Заключение	35
<i>Отбор моделей</i>	35
<i>Обсуждение распространения <i>T. pellucida</i></i>	35
<i>Обсуждение распространения <i>T. geniculata</i></i>	36
<i>Наиболее важные выявленные предикторы</i>	
Выводы	38
Благодарности	38
Литература	39
Приложения	43

Введение

Определение закономерностей в распространении видов – одна из главных задач современных ботанических исследований. Проанализировав связь между местообитаниями вида и факторами среды на изучаемой территории, можно смоделировать схему распространения данного вида (Sandarov, Pisarenko, 2018).

Для предсказания ареалов организмов в зависимости от климата используются географические модели (SDM – Species Distribution Models) (Jeschke, Strayer, 2008). SDM – это вероятностные модели, статистически связывающие распространение видов и сообществ с средой, в которой эти организмы находятся (Guisan, Zimmerman, 2000).

Еще в 1807 Humboldt и Bonpland показали, что климат важен в объяснении моделей распространения животных и растений. С тех пор проводилось много работ, в которых климат вместе с другими факторами среды использовался для выявления основных закономерностей в распределении растительности по всему миру (Guisan, Zimmerman, 2000).

Помимо того, что метод географического моделирования - важный инструмент экологии, поскольку он помогает исследовать взаимодействия между видами и окружающей средой, этот метод также используется для проверки биогеографических гипотез (Guisan, Zimmerman, 2000). Например, такая работа выполнена Mourell и Ezcurra (1996), целью которых было описать закономерности географического распространения основных растительных форм аргентинских видов кактусов. Для тестирования были выдвинуты гипотезы о:

1. Благоприятных условиях окружающей среды. Предполагалось, что экстремальные средовые условия способствуют благополучию вида меньше, чем промежуточные, поскольку при предельных условиях виду требуются «дорогостоящие» адаптивные механизмы.
2. Временных изменениях в климате. Чем более местообитания сезонны, тем большее количество видов может в них сосуществовать; при непредсказуемых варьированиях климата наибольшее видовое разнообразие наблюдается при срединных значениях его составляющих.
3. Гетерогенности местообитаний. Чем более они гетерогенны, тем больше экологических ниш может быть доступно за счет различий между микроместообитаниями. Таким образом в них может встречаться большее количество видов.
4. Ограниченной энергии и продукции. Согласно этой гипотезе, доступная энергия направлена на поддержание существования определенного количества видов, а видовое богатство зависит от первичной продукции – меры потока энергии.

Результаты данного исследования показали значительные различия в особенностях распространения разных растительных форм. Для столчатых видов кактусов верна гипотеза о благоприятных условиях среды, поскольку их распространение, по-видимому, ограничено низкими температурами. Северные, субтропические биогеографические регионы мало схожи с южными, умеренными. Видовое богатство резко падает выше 34° южной широты и выше высоты в 3000 м, за

исключением рода *Oreocereus*, встречающегося и на высоте 4000 м; стебли представителей данного рода полностью покрыты длинными шипами и волосками, по-видимому, служащими для теплоизоляции, препятствуя замерзанию растений, а также в качестве защиты от механических воздействий.

Распространение шаровидных видов сильно связано с количеством осадков, выпадающих летом, а кроме того, с характеристиками почвы, к примеру, со степенью ее каменистости. Шаровидные кактусы встречаются в относительно теплых регионах с обильными летними дождями. Видовое богатство шаровидных кактусов наблюдается в регионах с высотной зональностью, являющейся мерой гетерогенности местообитаний. Видовое богатство этого типа кактусов обуславливается переменами климата.

Широта, будучи косвенным средовым фактором, послужила лучшим предсказателем для богатства столбчатых кактусов, однако трудно использовать этот факт, поскольку широта строго коррелирует с переменными, обозначающими количество безморозных дней и дождливых летних дней, и является составным индикатором, свидетельствуя как о благоприятных условиях окружающей среды, так и о вариациях климата.

Для опунциевых кактусов наиболее точным предсказателем видового богатства оказалось количество летних осадков, индекс аридности и среднегодовая температура, то есть подтвердилась гипотеза о вариациях климата.

Закономерности в видовом богатстве столбчатых, опунциевых и шаровидных кактусов оказалась наиболее связана с климатическими вариациями (сезонностью климата); количество летних осадков – наиболее подходящая предсказательная переменная.

И сама продукция, и такие переменные, как эвапотранспирация (испарение влаги с самого растения или с почвы), среднегодовая температура и среднегодовое количество осадков – важные факторы, влияющие на продуктивность, - не существенно связаны с видовым богатством. Это может быть результатом того, что продуктивность, связанная с эвапотранспирацией, оказалась не сильно колеблющейся внутри ареала кактусов Аргентины. Кроме того, большинство кактусов часто встречается в низкопродуктивных местообитаниях.

Моделирование используется и для оценивания степени влияния преобразованной территории и других измененных факторов среды на распространение организмов (Guisan, Zimmerman, 2000). (Kienast et al., 1998; Lischke et al., 1998). Исследование Kienast et al. (1998) посвящено оцениванию изменений в растительности, обусловленных климатом. В нем использовался метод оценки экологического риска (ERA), который может быть направлен на определение вероятности того, что произойдет какое-то событие или же на выявление возможных изменений в эволюции экосистемы. Последнее сильно зависит от количественных и качественных моделей, а для этого подходит биогеографическое моделирование.

Также моделирование используется для дополнения флористических и фаунистических атласов, предсказания возможных ареалов инвазивных видов, составление умозаключений о палеоклимате на

основе распространения видов (климат раннего и среднего Миоцена - Kershaw, 1997; климат Голоцена – Dimitriadis, Cranston, 2001) (Guisan, Zimmerman, 2000; Jeschke, Strayer, 2008).

Моделирование инвазивных видов было проведено как для относительно небольших участков территории (Йосемитский национальный парк - Underwood et al., 2004; североамериканские прерии – Anderson et al., 2006), так и для материков (напр., Северной Америки - Strayer, 1991; Австралии – Sindel, Michael, 1992), и для всей территории Земли (напр., Thuiller et al., 2005).

В ботанических работах было использовано множество различных SDM: для прогнозирования пространственного распространения наземных видов растений (напр., Shao & Halpin, 1995 – деревья и кустарники), водных растений (напр., Lehmann et al., 1997 – Рогоз и Кувшинка), растительных сообществ (Brzeziecki et al., 1993 – 71 тип лесных сообществ), типов растительности (напр., Van de Rijt et al., 1996) и видового богатства (Mourell & Ezcurra, 1996; Heikkinen, 1996). Также биогеографическое моделирование использовалось в зоологических исследованиях (Aspinall, 1992; Mastrotillo et al., 1997).

Работ по моделированию распространения мохообразных проведено гораздо меньше, (напр., Désamoré et al. 2012; Song et al. 2015; Číhal et al., 2017). Вместе с тем, мхи – удобный объект для климатических исследований в силу таких свойств, как пойкилогидридность, чувствительность к высоким температурам, высокая способность к распространению и значительно меньшее видовое разнообразие по сравнению с сосудистыми растениями. Из подобных работ значительная доля была проведена не для России. Исследование Désamoré et al. (2012) заключалось в определении влияния климата в настоящем и прошлом на распространение и генетическое разнообразие мхов умеренного климата и роли Средиземноморья в эволюции этих мхов. С этой целью был использован SDM-метод и метод молекулярной филогеографии. В качестве модельного объекта был выбран *Homalothecium sericeum* Schimp. как типичный представитель европейской флоры умеренного климата.

Работа Song et al. (2015) проводилась, в том числе, и для выяснения видового богатства и разнообразия мха рода *Didymodon* Hedw., выбранного как включающий наибольшее количество видов в семействе *Pottiaceae*, в Тибете. В России моделированию мхов посвящена одна работа (Sandanov, Pisarenko, 2018), в которой в качестве модельного объекта использовался мох *Crossidium squamiferum* (Viv.) Jur., Laumb.-Fl. Oesterr.-Ung. Все виды рода *Crossidium* распространены в аридных областях мира («Флора...», 2003), как и кактусы, моделирование распространения видов которых описано выше. Для рода *Tetraphis*, напротив, характерны влажные и даже сырые местообитания; мы хотели посмотреть, насколько хорошо прогнозируются ареалы видов с такими климатическими предпочтениями.

Во всех указанных в предыдущем абзаце исследованиях использовалась программа Maxent – современный метод предсказания, использующий неполную информацию, известную о распространении изучаемого вида (Phillips et al., 2006). Мы решили проверить, насколько эта программа применима к моделированию двух видов мха рода *Tetraphis*.

Этот род относится к семейству Tetraphidaceae порядка Tetraphidales, одного из наиболее древних представителей листостебельных мхов. Для представителей этого семейства моделирование еще не

проводилось. В семействе Tetraphidaceae два рода – *Tetraphis* Hedw. и *Tetrodontium* (Schwägr.). *Tetrodontium* – мелкий мох со стеблем в 0.5-1 мм длиной, встречающийся на скалах, обычно в глубоких нишах («Флора...», 2017), что делает его неудобным модельным объектом, поскольку для построения моделей используются общеклиматические данные.

Tetraphis – широко распространенный род, включающий два вида – *T. pellucida* Hedw. и *T. geniculata* Girg. ex Milde. Распространение *Tetraphis pellucida* преимущественно совпадает с зоной хвойных лесов Голарктики. К югу от таежной зоны Голарктики этот вид встречается редко, преимущественно в горах до Южной Европы, Турции, Тибета, Сычуаня, Тайваня, высокогорий севера Южной Америки. В России вид встречается часто в пределах ареала ели и особо обычен в темнохвойных лесах, хотя местами нередок и в сырых борах и хвойно-широколиственных лесах, а также каменистых берегах. В то же время, *T. pellucida* практически отсутствует в зоне вечной мерзлоты: его нет на большей части территории Якутии, где единичные находки его известны только в березняках на более теплых склонах южной экспозиции. В степной зоне, а также в зоне широколиственных лесов он редок и встречается главным образом в пойменных болотах и в березовых колках. («Флора...», 2017).

В Китае *T. pellucida* встречается на самом севере Внутренней Монголии, в граничащей с ней провинцией Хэйлунцзян, на границе с Северной Кореей, в Сычуане, Шэньси, Гуйчжоу, Юньнани, в Тибетском автономном районе; на севере Синьцзян-Уйгурского района – на границе с Россией, Монголией и Казахстаном (<http://www.tropicos.org/Image/100009090?projectid=22>).

В США и в Канаде данный вид отмечен в большинстве штатов и провинций, хотя его нет в некоторых центральных южных штатах: в Небраске и к югу от нее; его также нет в Северной Дакоте, в Неваде и в Юте. Особенно много находок *T. pellucida* на восточном побережье Северной Америки (http://www.efloras.org/object_page.aspx?object_id=85115&flora_id=1).

Растет *T. pellucida* на гнилой древесине (на пнях чаще, чем на поваленных стволах), на гумусе и на торфе под корнями упавших деревьев (обычно в заболоченных лесах) и изредка на затененных выходах песчаника («Флора...», 2017).

Tetraphis geniculata описан из Японии. В Евразии вид встречается только на Востоке. В России – на Сахалине; на Курильских островах – на Шикотане, Кунашире, Итуруп; в Приморье и на юге Хабаровского края. На материке очень редок («Флора...», 2017).

За пределами России *T. geniculata* встречается в Японии, Корее. В Китае вид известен на северо-востоке Внутренней Монголии и на границе провинции Цзилинь с Северной Кореей (<http://www.tropicos.org/Image/100009088?projectid=22>).

В Северной Америке *T. geniculata* встречается как на восточном, так и на западном побережьях материка, полностью отсутствует в ее центральных районах. На западном побережье наиболее северные его находки – на Аляске; также *T. geniculata* встречается в западных провинциях Канады – в Юконе, Британской Колумбии и в Альберте. Наиболее южные находки – в Вашингтоне, Орегоне и в Айдахо. На восточном побережье *T. geniculata* встречается в Канаде – в Ньюфаундленде и Лабрадоре, в Квебеке и

других восточных провинциях. Наиболее южные находки *T. geniculata* на восточном побережье США – в штате Нью-Йорк (http://www.efloras.org/object_page.aspx?object_id=84746&flora_id=1).

Растет *T. geniculata* также на сильно разложившейся древесине и по характеру местообитаний и биотопов, по-видимому, не отличается от *T. pellucida* («Флора...», 2017).

Т. Герцог в 1926 году показал, что ареалы видов мохообразных сопоставимы с размерами ареалов таксонов покрытосеменных более высокого ранга – секций и родов. Кроме того, для мхов характерна высокая степень дизъюнктивности, то есть среди них чаще по сравнению с остальными высшими растениями встречаются виды, ареалы которых представляют собой участки, ограниченные друг от друга территориями суши или океана в сотни и тысячи километров. Наиболее распространенная причина возникновения «разорванных» ареалов – разделение исходного сплошного ареала, связанного с переменами климата («Древнейшие на суше», 1984).

Как следует из вышеизложенного, *T. geniculata* – дизъюнктивный вид, поскольку его ареал разделен на несколько частей, между которыми – участок суши – центральная часть Северной Америки, разделяющая западное и восточное побережья материка, на которых встречается *T. geniculata*, и Тихий океан, разграничивающий популяции этого мха в западной части Северной Америки и на востоке Евразии.

Цель нашей работы - исследовать применимость метода биоклиматического моделирования в программе Maxent для представителей рода мохообразных *Tetraphis*.

Задачи:

1. Собрать данные о распространении двух видов данного рода
2. Провести статистический анализ в программе Maxent
3. Сравнить получившиеся модели и выбрать из них оптимальные для предсказания местонахождений представителей каждого из двух видов
4. Сравнить реальное текущее распространение *T. pellucida* и *T. geniculata* с предсказанным

Материалы и методы

Метод

Моделирование мы проводили в программе Maxent (Phillips et al., 2006). Суть работы Maxent – оценка вероятности найденного распространения по максимальной энтропии этого распространения, то есть поиск наиболее выровненного географического распределения присутствия вида (поиск всех возможных территорий с подходящими для вида условиями).

В качестве входных данных используется файл, содержащий координаты находок вида (точки регистрации, TP), распространение которого прогнозируется, и файл с предикторами формата asc,

основанный на базе данных WorldClim2 (Fick, Hijmans, 2017; <http://worldclim.org/version2>). В начале моделирования происходит отбор определенного числа ячеек растра (фоновых точек, ФТ), при этом некоторые ТР могут попасть в ФТ. Значения факторов среды в ФТ и в ТР используются в обучении модели, которое происходит путем их трансформации в «функции предикторов» (ФП) разными математическими методами. Итогом работы MaxEnt является расчет нормализованной экспоненциальной функции, в которую в качестве аргумента входит сумма всех ФП, каждый из которых имеет свой коэффициент λ , назначаемый самой программой. Предсказанные ареалы строятся программой для заданной территории и цветами выделаются области, где вид встречается с большей вероятностью. Вклад отдельных предикторов в построение модели характеризуется различными параметрами, отчасти обсуждаемыми ниже. Подробное описание работы Maxent можно найти в статье Phillips et al., 2006.

Выбор нами программы Maxent основывался на следующих ее преимуществах:

1. Для моделирования в Maxent нужны лишь данные о присутствии видов.
2. Взаимосвязи между парами предикторов (растровых географических данных, описывающих изменчивость средовых факторов в пространстве) могут исказить выходные модели; Maxent может в некоторых случаях это учитывать, минимизируя влияние взаимно коррелирующих предикторов при анализе.
3. У вероятностного распространения, предсказанного Maxent есть четкое математическое объяснение, благодаря чему модели можно легко трактовать.
4. В настоящее время метод максимальной энтропии широко используется в области моделирования, являясь наиболее популярным.

Материалы

В качестве материала для моделирования распространения *T. pellucida* и *T. geniculata* на всей территории Земли мы использовали данные о находках этих видов, полученные из базы данных GBIF (<https://www.gbif.org/>) и Arctoa (<http://arctoa.ru/Flora/basa.php>). Для каждого из видов, таким образом, было сделано по одному файлу, одновременно содержащему координаты находок и с GBIF, и с Arctoa. Файл по *Tetraphis pellucida* включал 17298 точек, по *Tetraphis geniculata* – 282.

Распространение *T. pellucida* мы моделировали как для всей планеты, так и отдельно для России. Моделирование распространения по всей территории Земли было сделано, чтобы понять, в какой степени метод максимальной энтропии применим к данному виду. Моделирование распределения этого вида по России проводилось более детально, и, кроме того, учитывались данные не только базы WorldClim2 (Fick, Hijmans, 2017; <http://worldclim.org/version2>), но и также данных повторяемости облачности (Wilson, Jetz, 2016), рассчитанные по дистанционным данным; глобальным данным по гетерогенности среды (Tuanmu, Jetz, 2015); глобальным данным по почвенно-литологическим условиям (Shangguan et al., 2016; Hengl et al., 2017). Использовалось 479 точек находок *T. pellucida*, взятых с сайта

ArcToa. Эти точки были получены из файла с частично (с шагом в 100 км) удаленными координатами находок *T. pellucida* по России.

Tetraphis pellucida

Для моделирования распространения данного вида в мире дополнительно на основе исходного файла было сделано два разреженных, из которых часть точек была убрана. В одном файле шаг разреживания был 100 км (rar100), в другом – 250 (rar250). Сделано это было, чтобы избежать выбора ФТ преимущественно из агрегаций ТР, которые неизбежно присутствуют в хорошо исследованных районах (например, в Великобритании) и которых почти нет в малонаселенных и на порядки слабее исследованных районах Азиатской России. Степень разреживания данных была определена исходя из площади изучаемой территории. В итоге в файле rar100 получилось 470 точек, в rar250 – 174.

Разреживание производилось в программе ArcGis 10.3 с использованием инструмента "delete identical". Этот инструмент находит одинаковые записи на основе значений входных полей, затем удаляет все, кроме одной идентичной записи из каждого набора. Сравнение проводилось по местоположению точек находок, с параметром Допуск XY (XY Tolerance) равным 100 км. В результате из исходного облака точек сохранились только те, которые удалены друг от друга не менее чем на 100 км. Выборка rar250 делалась так же, но с шагом 250 км.

Также, чтобы исключить неслучайный выбор ФТ, мы использовали файл коррекции, bias file, для нивелирования различий в степени изученности мхов разных районов России, любезно предоставленный С.В. Дудовым. Создание файла коррекции происходит путем присвоения разным ячейкам раstra разных значений. Чем присвоенное значение больше, тем больше вероятность отбора соответствующей ячейки в ФТ, значения факторов среды в которых используются в обучении модели. (Phillips et al., 2006). В нашей работе он получен следующим образом. Из базы данных мхов России (Ivanov et al., 2017) и GBIF были взяты все точки находок мхов с географическими координатами. Вокруг них штатными средствами были построены буферные зоны 20 км. Пикселям в пределах этих буферных зон было присвоено значение 1, за их пределами - "no data". Таким образом, обучение модели велось только в пределах этих буферных зон вокруг точек сбора мхов. Подробнее про файл коррекции можно прочитать в статье El-Gabbas, F. Dormann, 2018.

Мы запускали в анализ файл с координатами находок вида, подлежащий обработке, в формате csv, а также файл в формате asc, содержащий растровые географические данные – предикторы (переменные, «слои», environmental layers), каждый из которых описывает определенный параметр окружающей среды на исследуемой территории. Все изначально установленные параметры Maxent при обработке данных по *T. pellucida* мы оставляли по умолчанию, кроме maximum iterations, который мы приняли во всех запусках равным 1500 исходных вместо 500, т.к. это значение по умолчанию годится лишь для простых моделей; replicates (репликации) - количество построений моделей с использованием одного файла с координатами, мы изменили с 1 на 5, как и в остальных случаях моделирования;

replicated run type поменяли с crossvalidate на subsample. Кросс-валидация подразумевает разделение группы входных данных на две равные по количеству точек и построение модели по каждой из этих групп отдельно. Оставшаяся группа используется для тестирования качества модели. При subsample есть тестовая и обучающая выборка, полученные из точек с исследуемой территории; процент тестовых точек выбирается исследователем. Мы выбрали 30% тестовых точек, соответственно 70% участвовало в обучающей выборке; так же мы разделили и выборки, использовавшиеся в других случаях моделирования, следуя большинству публикаций, в которых решаются сходного рода задачи, например Mateo et al. (2016). В конце обработки данных из трех серий репликаций (т.е. репликаций (а) по файлу со всеми точками, (б) rar100 и (в) rar250) были выбраны три наиболее удачные (см. моделирование 1).

По значениям AUC (area under the ROC-curve), наиболее часто используемым для оценки качества полученной модели Maxent, мы выбрали один файл из трех использованных, с помощью которого была построена наилучшая прогнозная модель, и по нему было проведено отдельное моделирование, чтобы иметь возможность сравнить получившиеся результаты с результатами моделирования *T. geniculata*, для которого известно гораздо меньшее число точек, в мире. В обоих случаях изменялся параметр сложности модели - regularization multiplier (RM). Поскольку, смоделировав распространение *T. pellucida* по территории России, мы выяснили и подтвердили, что AUC test и AUC training принимают наибольшие значения, если не увеличивать параметр сложности модели, при данном моделировании RM мы приняли равным 0.5 и 1.

При анализе в Maxent файла с координатами находок *T. pellucida* на территории России мы меняли параметр сложности модели – regularization multiplier (RM). Мы использовали RM = 0.5, 1, 1.5, 2.5, 5. Для каждого из значений RM был выполнено по пять репликаций, из которых было выбрано по одной наилучшей репликации для составления выводов о качестве модели, которую каждая из этих репликаций представляет.

Tetraphis geniculata

Для *T. geniculata* мы моделировали распространение только для всего мира, поскольку для России оно компактное, ограниченное только небольшой частью Дальнего Востока. Как и при моделировании распространения *T. pellucida*, мы использовали bias-файл. Параметры обработки данных использовались те же, что и в случае моделирования распространения *T. pellucida* по территории России, и вся обработка в целом проводилась так же; мы смотрели, при каком значении RM получается наиболее достоверная модель, однако координаты находок *T. geniculata* не разреживались.

Результаты и обсуждение

1. Моделирование общего распространения *T. pellucida*

При проведении этого моделирования оценивалась изменчивость качества прогнозирования моделей при разных типах разреживания файлов (неразрезанный файл, rar100, rar250).

Оценка пригодности моделей

Главный показатель качества модели, выбранный нами - AUC, площадь под ROC (receiver-operating characteristic) – кривой (рис. 1). AUC интерпретируется как показатель вероятности того, что случайно выбранные точки из TP предсказаны моделью лучше, чем любые случайно выбранные точки (любые ФТ).

Параметры трех наиболее удачных моделей (по наибольшему значению AUC и наименьшему числу оmissions (неверно предсказанных присутствий вида)) показаны в результирующей таблице (см. табл. 1).

Как видно из таблицы 1, разреживание значительно увеличивает AUC, из-за чего улучшается способность модели предсказывать распространение вида. Значения AUC test моделей 2 и 3 выше, чем по модели 1, однако разреживание с шагом в 250 км дает худшие показатели и по AUC test, и по test omission по сравнению с разреживанием с шагом в 100 км. AUC test модели 2 больше AUC test модели 1 на ~5%, а модели 3 на ~1,4% и равна 0.918; это говорит о том, что модель 2 обладает очень хорошей предсказательной способностью, большей по сравнению с моделями 1,3.

номер модели	1	2	3
файл для обработки	Т.р. все точки	Т.р. rar100	Т.р. rar250
AUC training (площадь под ROC - кривой, построенной по точкам обучающей выборки)	0,8699	0,927	0,9196
AUC test (площадь под ROC - кривой, построенной по точкам тестовой выборки)	0,8678	0,9177	0,904
Fixed cumulative value 1 training omission (омиссия по точкам обучающей выборки при фиксированном совокупном значении 1)	0,0089	0,0062	0,0085
Fixed cumulative value 1 test omission (омиссия по тестовым точкам при фиксированном совокупном значении 1)	0,0058	0	0,02
Fixed cumulative value 5 training omission (омиссия по точкам обучающей выборки при фиксированном совокупном значении = 5)	0,0341	0,0433	0,0169
Fixed cumulative value 5 test omission (омиссия по тестовым точкам при фиксированном совокупном значении = 5)	0,0423	0,0362	0,06
Fixed cumulative value 10 training omission (омиссия по точкам обучающей выборки при фиксированном совокупном значении = 10)	0,0949	0,0743	0,0508
Fixed cumulative value 10 test omission (омиссия по тестовым точкам при фиксированном совокупном значении = 10)	0,1095	0,058	0,08

Табл. 1. Результирующая таблица по моделям общего распространения *T. pellucida*.

На рис. 1 представлен график AUC, полученный по модели 2. Обычно значения AUC по тренировочным данным выше, чем по тестовым, а AUC test является реальной мерой качества

прогнозирования модели. Площадь AUC всегда меньше 1, а случайное распределение равно 0.5. По оси абсцисс на графике AUC (рис. 4) отложена «специфичность» - доля верно предсказанных отсутствий представителей *T. pellucida*, а по оси ординат – «чувствительность» - доля верно предсказанных присутствий. Поскольку Maxent не использует данные об отсутствии, специфичность оценивается с помощью случайно отобранных точек из ФТ. Красная и синяя кривые ROC показывают, насколько хорошо модель описывает тренировочные и тестовые данные соответственно.

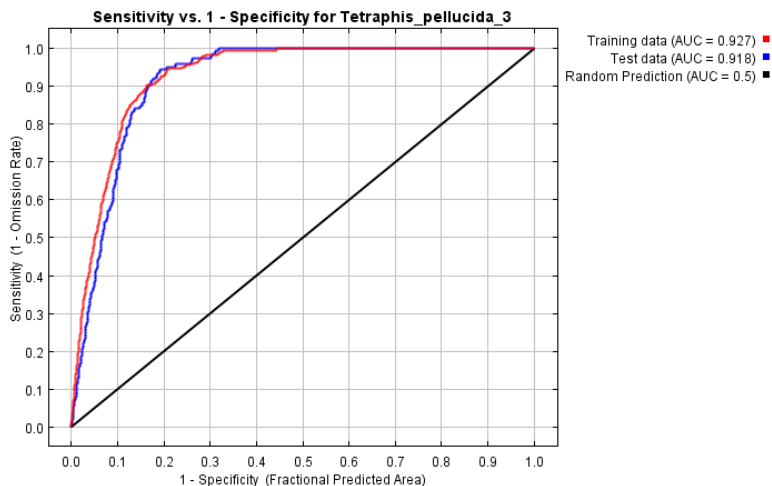


Рис. 1. График, показывающий AUC test и AUC training по модели 2 (файл rar100).

Построенные модели

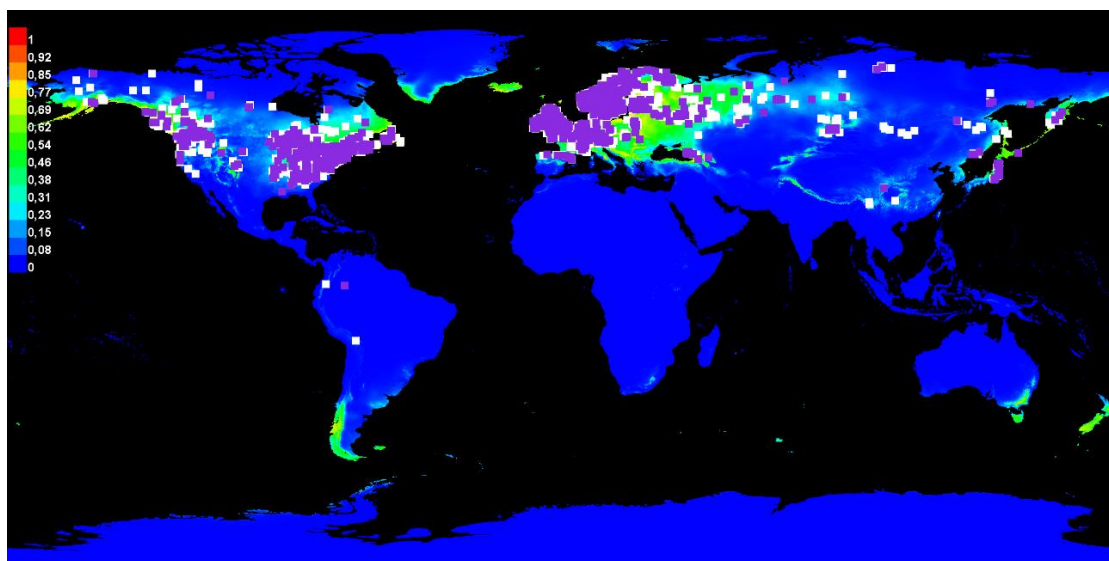


Рис. 2. Карта распространения *T. pellucida* по всему миру, модель 1 (AUC training = 0,8699; AUC test = 0,8678).

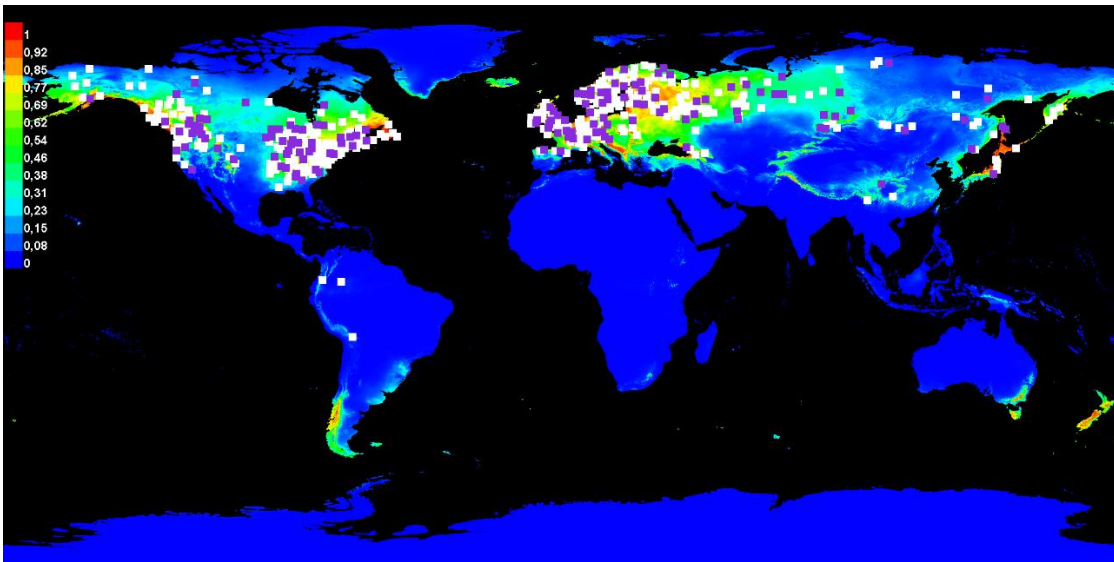


Рис. 3. Карта распространения *T. pellucida* по всему миру, модель 2 (AUC training = 0,927; AUC test = 0,9177).

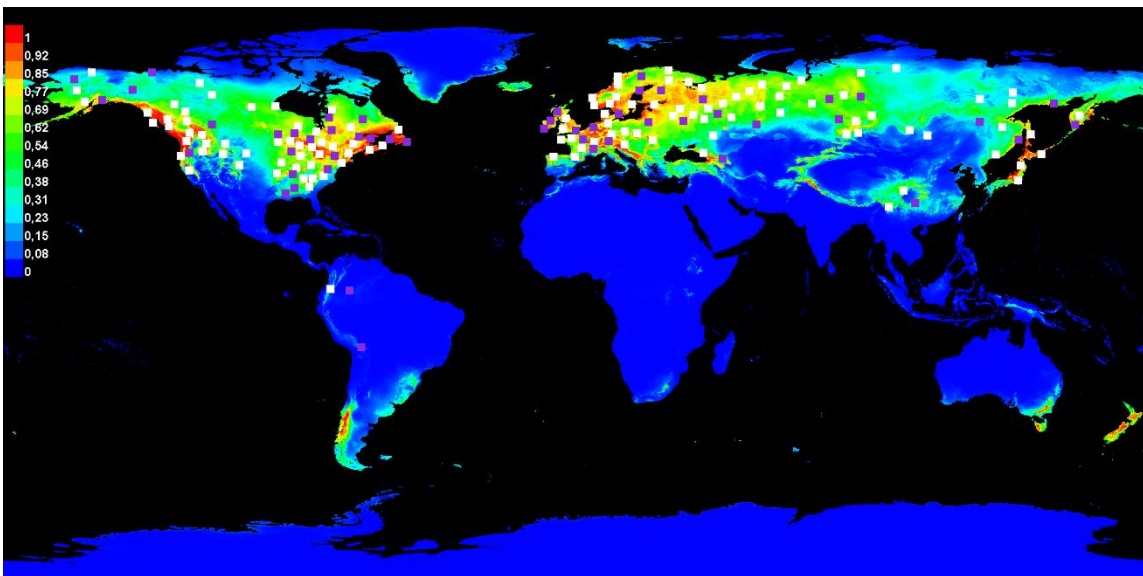


Рис. 4. Карта распространения *T. pellucida* по всему миру, модель 3 (AUC training = 0,9196; AUC test = 0,904).

На полученных картах (рис. 2,3,4) показано предсказанное распространение *T. pellucida*. Цветами обозначена вероятность встречаемости данного вида в разных частях мира. Синий цвет свидетельствует о том, что обнаружить *T. pellucida* на соответствующей территории маловероятно, красный – о том, что вероятность найти вид высока. Белые квадраты показывают местонахождения из обучающей выборки, на основе них была построена модель; фиолетовые квадраты относятся к тестовой части выборки.

Из карт, построенных по трем моделям видно, что первая (по неразрезанному файлу) не предсказывает значительную часть точек в Сибири. Третья модель (rar250), наоборот, выделяет зеленым области возможного распространения вида, в которых он отсутствует или крайне редок. Вторая модель (rar100) наилучшим образом совпадает с имеющимися данными о распространении вида, так что в данном случае критерий AUC, который здесь наивысший и для тренировочной выборки, и для тестовой, можно признать хорошо работающим.

Если сравнить полученную карту по модели 2 (рис. 3) с реальными находками *T. pellucida* в Северной Америке (см. ниже), можно заметить, что по карте прослеживается отсутствие этого вида в штатах Среднего Запада (относительно засушливом районе в ветровой тени Скалистых гор), что гораздо хуже видно из обычных карт, таких как, например, во Флоре Северной Америки (рис. 5).

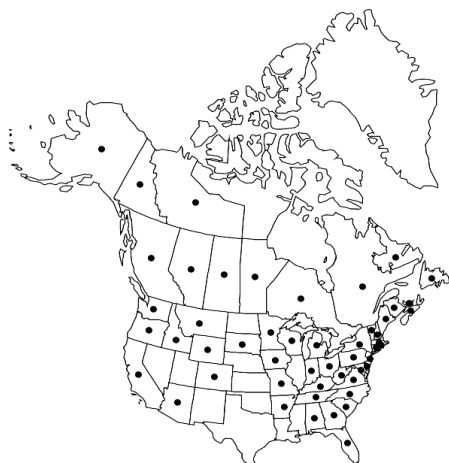


Рис. 5. Находки *T. pellucida* в Северной Америке (http://www.efloras.org/object_page.aspx?object_id=85115&flora_id=1)

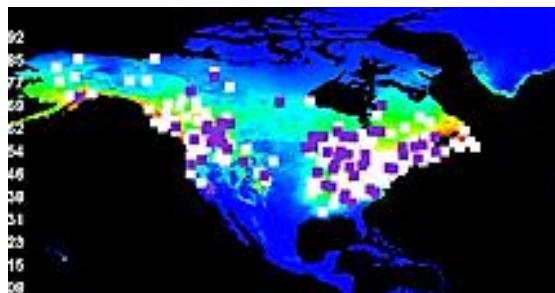


Рис. 3. Карта распространения *T. pellucida* по всему миру, модель 2. Показана только Северная Америка.

Анализ вклада переменных

Variable	Percent contribution	Permutation importance
wc2.0_bio_10m_14.tif - Количество осадков в наиболее сухой месяц	46	6.6
wc2.0_bio_10m_11.tif - Средняя температура наиболее холодного квартала	31.2	12.3
wc2.0_bio_10m_01.tif - Среднегодовая температура	6.7	12.1
wc2.0_bio_10m_05.tif - Максимальная температура наиболее теплого месяца	4.9	40.4
wc2.0_bio_10m_12.tif - Годовые осадки	2.6	6.3
wc2.0_bio_10m_08.tif - Средняя температура наиболее влажного квартала	1.8	0.8
wc2.0_bio_10m_10.tif - Средняя температура наиболее теплого квартала	1.4	2.5
wc2.0_bio_10m_07.tif - Среднегодовая амплитуда температур	0.8	1.8
wc2.0_bio_10m_09.tif - Средняя температура наиболее сухого квартала	0.8	1
wc2.0_bio_10m_19.tif - Количество осадков в наиболее холодный квартал	0.7	2.4
wc2.0_bio_10m_03.tif - Изотермальность (отношение среднегодовой температуры к среднегодовой амплитуде температур умножено на 100)	0.6	1.7
wc2.0_bio_10m_15.tif - Сезонность осадков (коэффициент вариации)	0.5	1.8
wc2.0_bio_10m_04.tif - Температурная сезонность	0.4	0.6
wc2.0_bio_10m_17.tif - Количество осадков в наиболее сухой квартал	0.4	0.7

ws2.0_bio_10m_06.tif - Минимальная температура наиболее холодного месяца	0.3	3.9
ws2.0_bio_10m_02.tif - Среднесуточная амплитуда температур	0.3	0
ws2.0_bio_10m_13.tif - Количество осадков наиболее влажного месяца	0.2	3.5
ws2.0_bio_10m_16.tif - Количество осадков наиболее влажного квартала	0.2	0.4
ws2.0_bio_10m_18.tif - Количество осадков наиболее теплого квартала	0.2	1.2

Табл. 2. Процентный вклад переменных в построение модели общего распространения *T. pellucida*, показан вклад переменных и их важность при пермутации (%).

Оценка вклада каждого из предикторов в построенную модель представлена в результирующей таблице Maxent (табл. 2), в первом столбце которой указаны названия используемых при анализе переменных (с добавленной нами расшифровкой переменных), а в двух оставшихся - по-разному выраженная важность каждой из переменных; показаны усредненные значения для пяти проведенных репликаций по выбранной модели 2. Для всех переменных использовались геоданные с разрешением 10 мин.

При обучении модели, в процессе «тренировки», отслеживаются переменные, вносящие вклад в ее построение. С каждой итерацией Maxent увеличивает вклад этих переменных, в результате чего изменяется коэффициент λ при ФП. В конце обучения вклады переменных в построение модели пересчитываются в проценты и отображаются в столбце “percent contribution”. Интерпретировать этот столбец необходимо с осторожностью, т. к. значения в нем могут быть разными в зависимости от того, какой путь «был избран» программой для построения модели, а также в случае корреляции переменных между собой.

Третий столбец (permutation importance), в отличие от второго, не зависит от путей, которыми была построена модель, а от окончательной модели и показывает уменьшение AUC training при случайном изменении значений среды в точках обучающей выборки – и в ФТ, и в ТР. Поэтому, хотя столбец permutation importance тоже чувствителен к корреляции данных, этому столбцу стоит придавать большее значение.

По второму столбцу табл. 2 можно заметить, что наибольший процентный вклад в построение модели вносят две переменные - ws2.0_bio_10m_14 - количество осадков наиболее сухого месяца (46%) и ws2.0_bio_10m_11 - средняя температура наиболее холодного квартала (31,2%). Значения прироста у остальных переменных гораздо меньше. Третий столбец свидетельствует о том, что AUC training уменьшится на 40.4% при изменении значений переменной ws2.0_bio_10m_05 - максимальной температуры наиболее теплого месяца в точках обучающей выборки. Важность этого параметра понятна с учетом экологии вида, который не выдерживает длительного пересыхания.

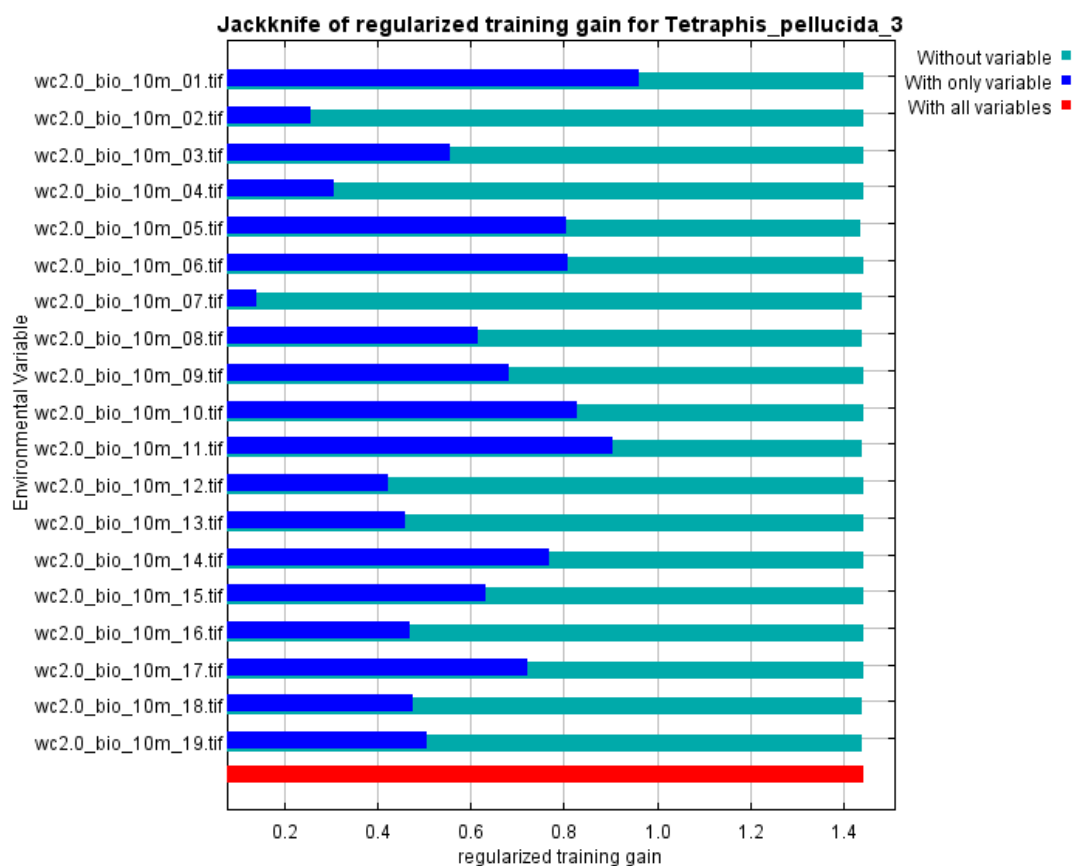


Рис. 6. Диаграмма, показывающая результат jackknife-теста по обучающим данным; модель 2 (J1).

Также для оценки вклада переменных были также проделаны jackknife-тесты (рис. 6), показывающие вклад в модель каждой из переменных, возникший в результате математических преобразований значений в исследуемых точках. Проводятся они с помощью построения моделей без каждой из переменных, моделей только с конкретной переменной и модели, использующей все переменные; на выходе получается столбчатая диаграмма. Длина синей линии отражает прирост модели, когда использовался только соответствующий предиктор. Чем длиннее синяя линия, тем важнее соответствующая переменная при построении модели сама по себе. Длина бирюзовой линии показывает прирост модели, исключая этот предиктор при построении. Чем короче бирюзовая линия, тем больше уникальной информации содержит исключенная переменная, и тем хуже получается модель в ее отсутствие.

По J1 можно сделать вывод о том, что наибольший вклад по тренировочным точкам при использовании в построении модели одной переменной обеспечивают переменные wc2.0_bio_10m_01 - среднегодовая температура и wc2.0_bio_10m_11 - средняя температура наиболее холодного квартала. Эти переменные находятся на третьем и втором местах соответственно в столбце percent contribution табл. 2, а пермутация обеих этих переменных уменьшает AUC training более, чем на 12%. Наименьший прирост происходит за счет wc2.0_bio_10m_07 - годовая амплитуда температур, вклад которой в построение модели составляет 0.8%. Можно сделать вывод еще и о том, что никакие из предикторов не

содержат информации, которой бы не было в других; исключение любого из них не привело бы к искажению модели.

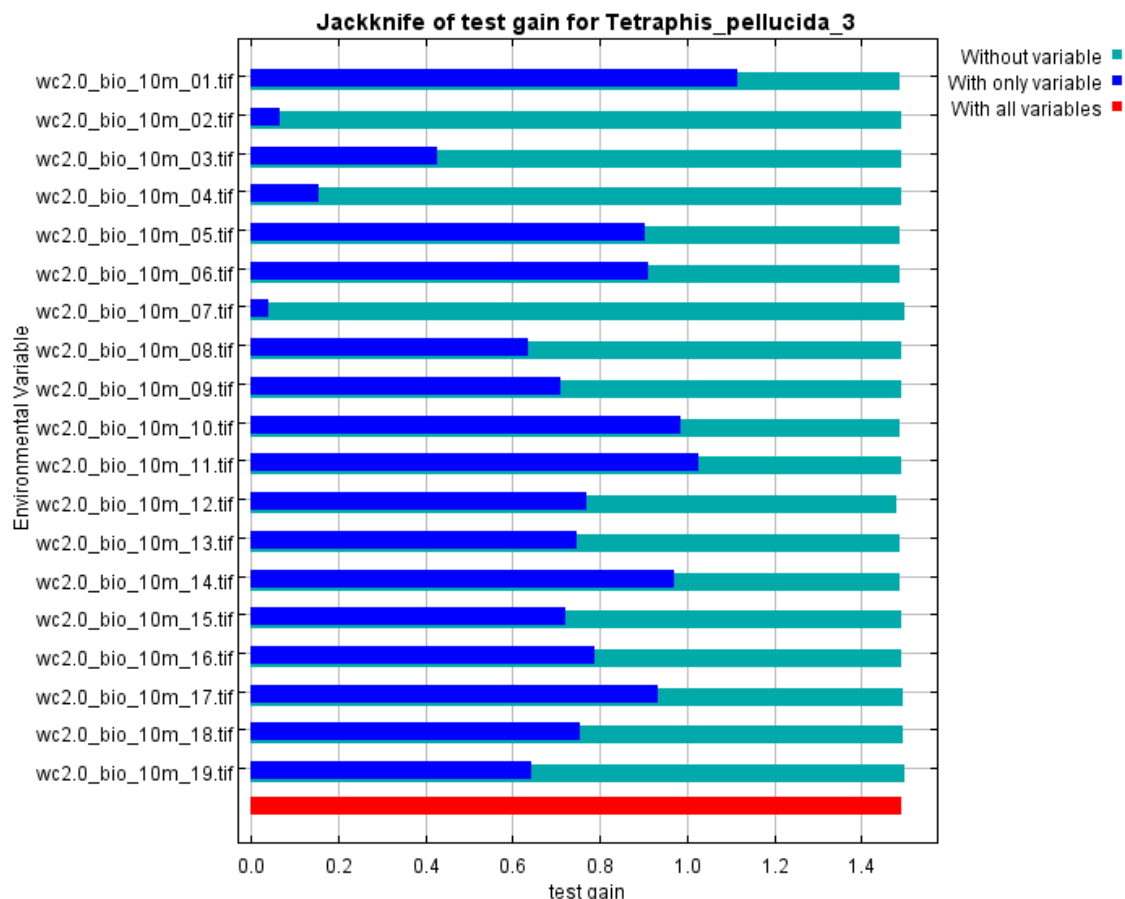


Рис. 7. Диаграмма, показывающая результат jackknife-теста по тестовым данным; модель 2 (J2).

J2 подтверждает, что самая важная переменная при построении модели сама по себе - wc2.0_bio_10m_01 - среднегодовая температура, по тестовым данным она увеличивает прирост даже больше, чем по тренировочным, а переменные, слабо увеличивающие прирост по тренировочным точкам (wc2.0_bio_10m_07 - годовая амплитуда температур, wc2.0_bio_10m_02 - среднесуточная амплитуда температур, wc2.0_bio_10m_04 - температурная сезонность), по тестовым увеличивают прирост еще слабее.

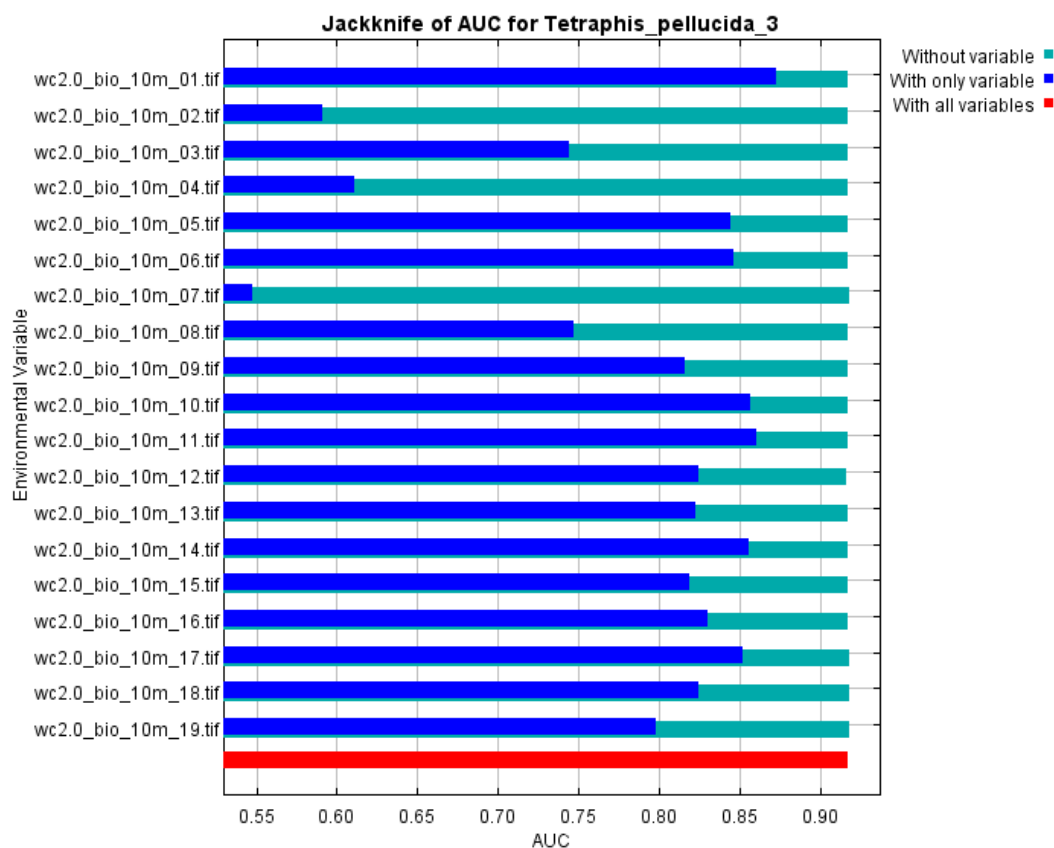


Рис. 8. Диаграмма, показывающая результат jackknife-теста по AUC test; модель 2 (J3).

Интересно, что незначительное влияние и на AUC test вышеупомянутых *wc2.0_bio_10m_07*, *wc2.0_bio_10m_02*, *wc2.0_bio_10m_04* соотносится с их незначительным уменьшением AUC training при пермутации (см. табл. 2); *wc2.0_bio_10m_02* и вовсе не влияет на площадь под ROC-кривой.

2. Моделирование распространения *T. pellucida* на территории России

При моделировании распространения *T. pellucida* на территории России мы изменяли параметр regularization multiplier (RM) (в таблицах - Reg_mult) для построения оптимальной модели. RM отвечает за степень сложности модели, то есть за то, насколько построенная модель будет близка к тренировочным данным, на основе которых проводилось обучение. По умолчанию он принят за единицу. Чем RM выше, тем более «простой» получается модель и тем более «плавным» оказывается распространение, представленное на картах выходных моделей, полученная модель менее «привязана» к тренировочным данным. И наоборот, если этот параметр ниже 1, предсказанное распространение более локализованное, «островное», оно более точно соответствует тренировочным образцам; строится «сложная» модель. Однако это может привести и к излишнему обучению модели, в результате чего она будет плохо описывать тестовые данные, к тому же, в целом, при низких значениях RM AUC больше, чем при высоких.

Построенные модели

По распространению *T. pellucida* в России были построены следующие модели:

Модель 1

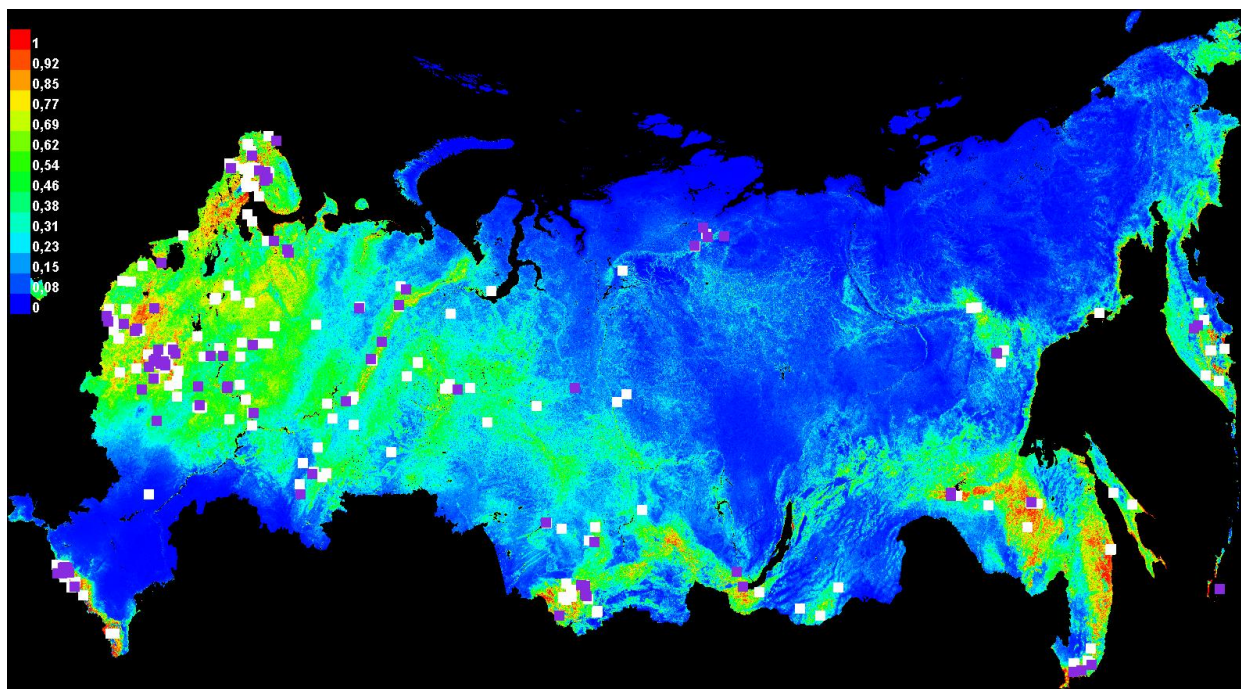


Рис. 9. Карта, полученная при моделировании распространения *T. pellucida* на территории России; regularization multiplier = 0.5; модель 1 (AUC training = 0,8491; AUC test = 0,819).

Модель 2

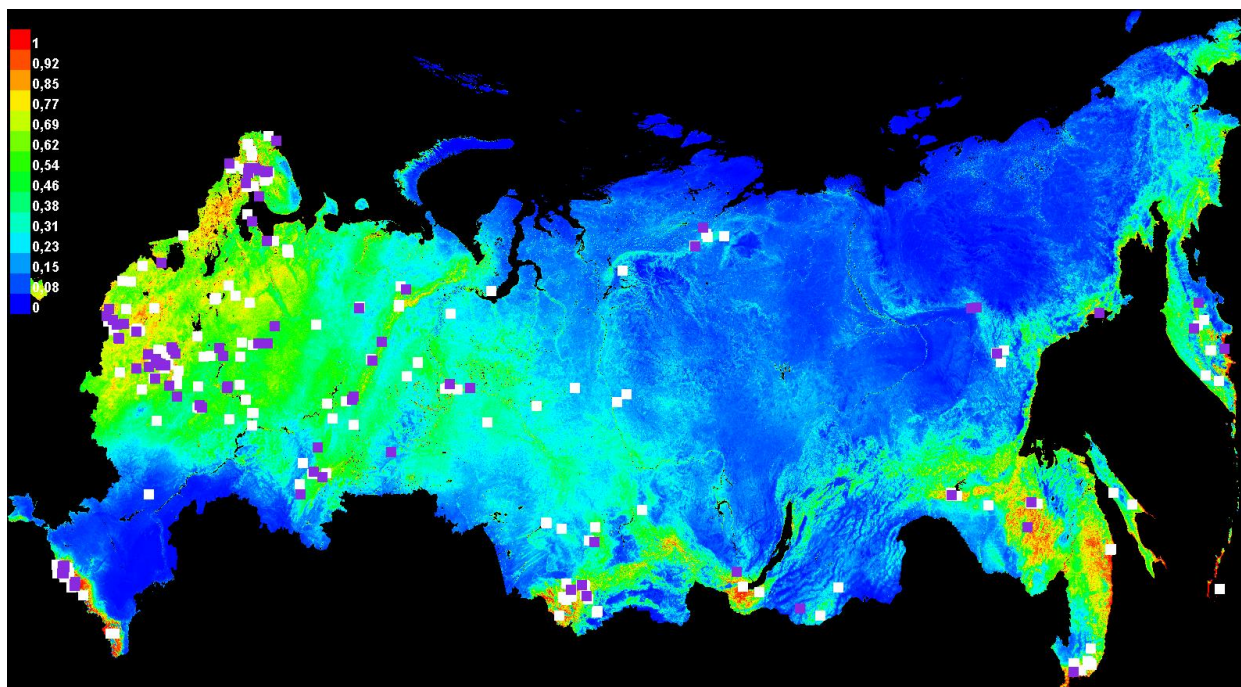


Рис. 10. Карта, полученная при моделировании распространения *T. pellucida* на территории России; regularization multiplier = 1; модель 2 (AUC training = 0,8173; AUC test = 0,8154).

Модель 3

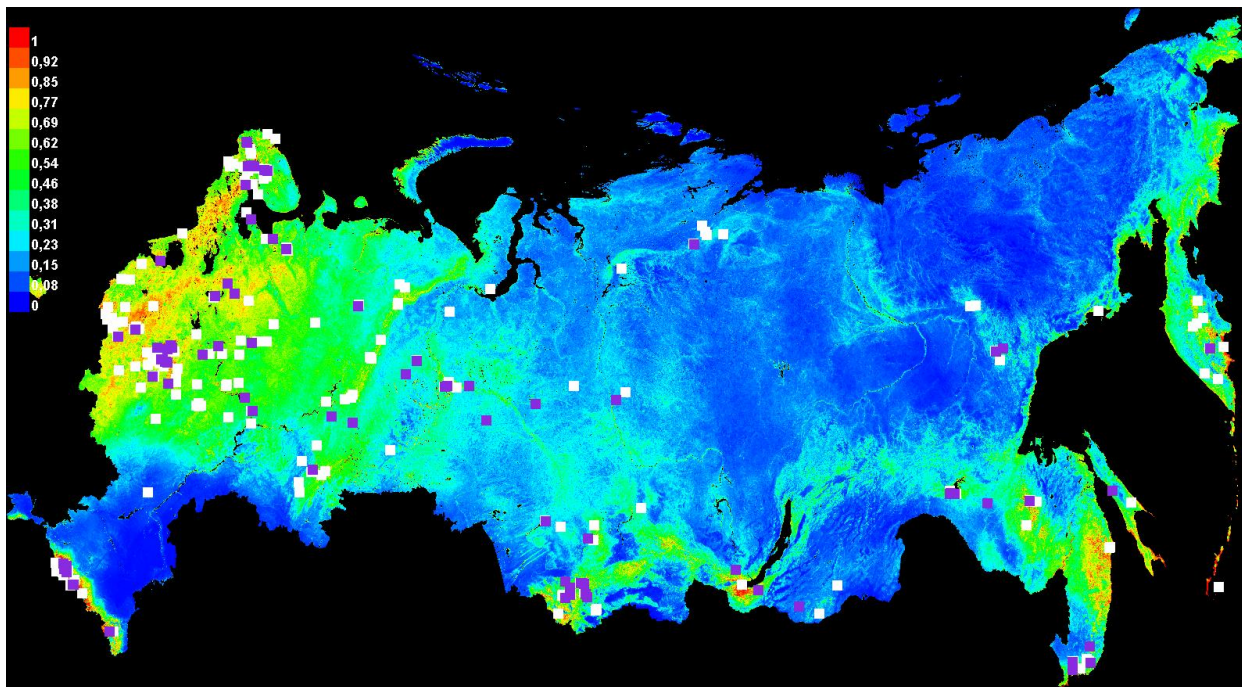


Рис. 11. Карта, полученная при моделировании распространения *T. pellucida* на территории России; regularization multiplier = 1,5; модель 3 (AUC training = 0,8083; AUC test = 0,7743).

Модель 4

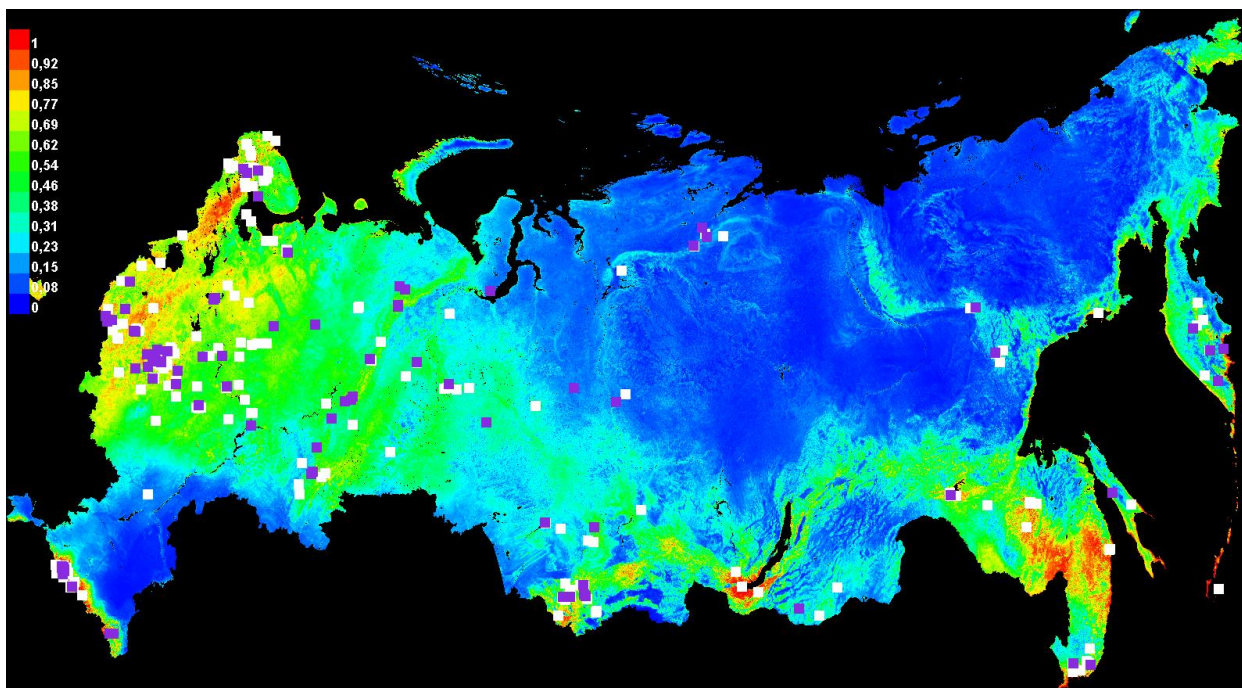


Рис. 12. Карта, полученная при моделировании распространения *T. pellucida* на территории России; regularization multiplier = 2.5; модель 4 (AUC training = 0,7976; AUC test = 0,7554).

Модель 5

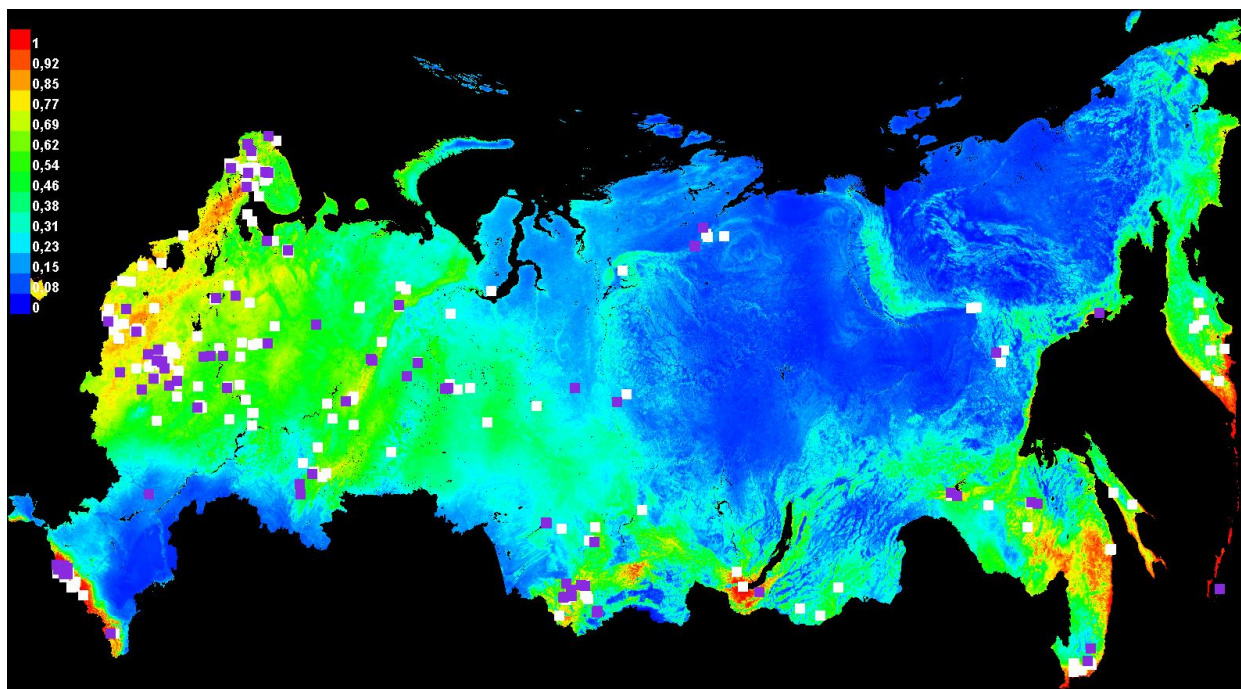


Рис. 13. Карта, полученная при моделировании распространения *T. pellucida* на территории России; regularization multiplier = 5; модель 5 (AUC training = 0,7817; AUC test = 0,7572).

Оценка пригодности моделей

№ модели	1	2	3	4	5
Reg_mult (параметр сложности модели)	0,5	1	1,5	2,5	5
AUC training (площадь под ROC - кривой, построенной по точкам обучающей выборки)	0,8491	0,8173	0,8083	0,7976	0,7817
AUC test (площадь под ROC - кривой, построенной по точкам тестовой выборки)	0,819	0,8154	0,7743	0,7554	0,7572
Fixed cumulative value 1 training omission (омиссия по точкам обучающей выборки при фиксированном совокупном значении 1)	0,0051	0,0051	0	0,0051	0,0103
Fixed cumulative value 1 test omission (омиссия по тестовым точкам при фиксированном совокупном значении 1)	0,012	0	0	0	0
Fixed cumulative value 5 training omission (омиссия по точкам обучающей выборки при фиксированном совокупном значении = 5)	0,0359	0,0359	0,0205	0,0359	0,0462
Fixed cumulative value 5 test omission (омиссия по тестовым точкам при фиксированном совокупном значении = 5)	0,0482	0,0241	0,0361	0,0843	0,0602
Fixed cumulative value 10 training omission (омиссия по точкам обучающей выборки при фиксированном совокупном значении = 10)	0,0462	0,0821	0,0769	0,0615	0,0769
Fixed cumulative value 10 test omission (омиссия по тестовым точкам при фиксированном совокупном значении = 10)	0,0964	0,0602	0,0723	0,1325	0,0843

Табл. 3. Результирующая таблица по распространению *T. pellucida* на территории России, демонстрирующая значения рассматриваемых параметров для моделей 1,2,3,4,5 (см. карты на рис. 9-13).

На основе результирующей таблицы (табл. 3) за лучшую была принята первая модель (RM = 0.5), поскольку AUC test у этой модели наибольший.

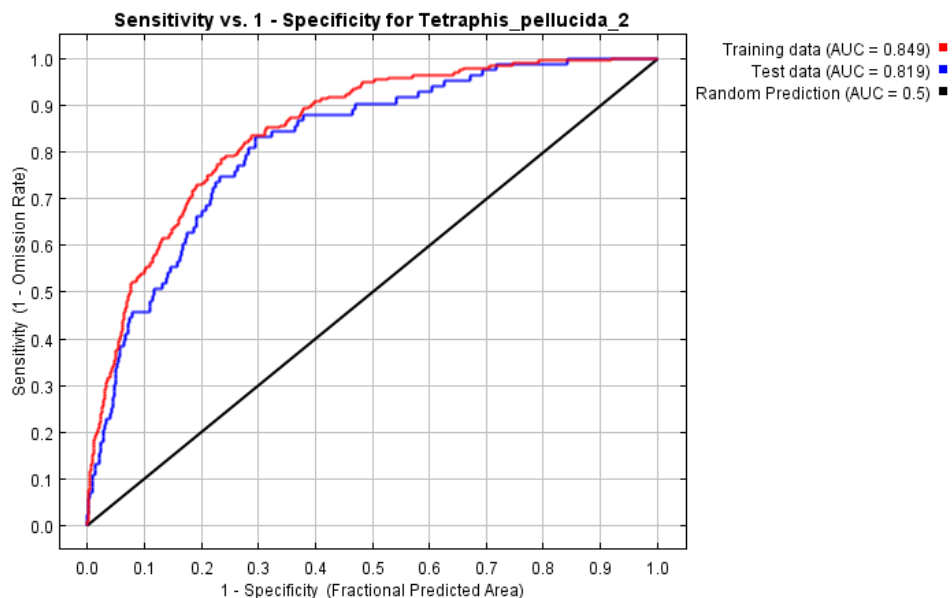


Рис. 14. График, отображающий ROC и AUC, построенные для модели 1.

Из графика на рис. 14 видно, что значение $AUC_{test} = 0.819$, а $AUC_{training} = 0.849$; каждое из этих значений расценивается как хорошее.

Сравнение спрогнозированного распространения *T. pellucida* на территории России с реальным

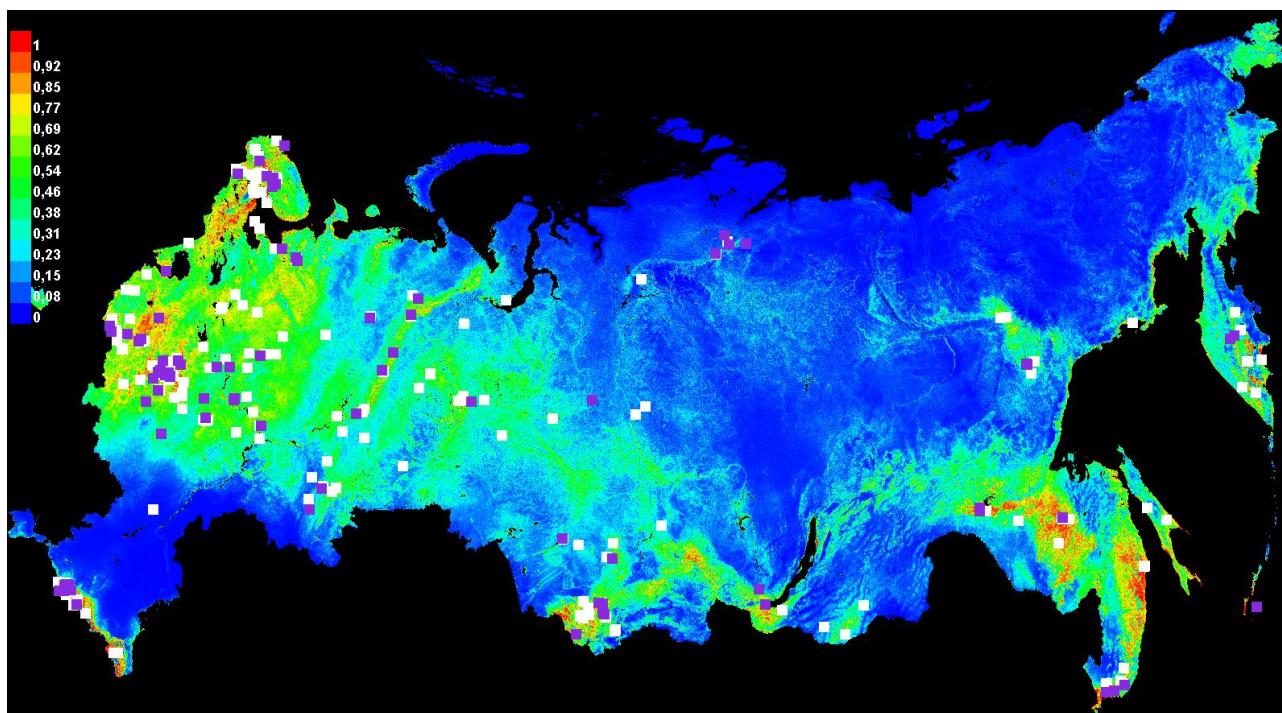


Рис. 9. Карта, полученная при моделировании распространения *T. pellucida* на территории России; regularization multiplier = 0.5; модель 1 ($AUC_{training} = 0,8491$; $AUC_{test} = 0,819$).

резко континентальный, и большая часть которой охвачена вечной мерзлотой (см. рис. 16), *T. pellucida* отсутствует, что показано на рис. 9.

Спрогнозированные территории присутствия *Tetraphis pellucida* полностью подтверждаются рис. 15. Этот вид предсказывается на юго-западе Кольского полуострова, где преобладает тайга и лесотундра; в южной части Кавказа – на юго-западном склоне Большого Кавказа растут в том числе хвойные леса; в европейской части России. Кроме того – в Ханты-Мансийском, Ямало-Ненецком районах, в средней части Красноярского района; на юге Бурятии, в Алтайском крае и в близлежащих районах; в Хабаровском крае, в Приморье и на юго-востоке Камчатки, на Сахалине и на южных Курилах – на Шикотане, что согласуется с рис. 15 и с находками.

(http://81.17.153.132/rbf/maps_sp1.pl?where_genera=268&sname_id=1223&view=view)



Рис. 16. Карта распространения вечной мерзлоты в России

(https://www.google.com/search?q=%D0%BA%D0%B0%D1%80%D1%82%D0%B0+%D1%80%D0%B0%D1%81%D0%BF%D1%80%D0%BE%D1%81%D1%82%D1%80%D0%B0%D0%BD%D0%B5%D0%BD%D0%B8%D1%8F+%D0%B2%D0%B5%D1%87%D0%BD%D0%BE%D0%B9+%D0%BC%D0%B5%D1%80%D0%B7%D0%BB%D0%BE%D1%82%D1%8B+%D0%B2+%D1%80%D0%BE%D1%81%D1%81%D0%B8%D0%B8&sxsr=ACYBGNSuHZt1mHSOd84vXPfYurS0q6gNXw:1568577309355&source=lnms&tbm=isch&sa=X&ved=0ahUKewiQq87qzdPKAhXifZoKHdGJD3EQ_AUIEigB&biw=1366&bih=625#imgrc=DqtRhBkMh5fAtM:)

Анализ вклада переменных

Variable	Percent contribution	Permutation importance
wc2.0_bio_30s_07.tif - Среднегодовая амплитуда температур	31.7	8.9
wc2.0_bio_30s_13.tif - Количество осадков наиболее влажного месяца	19.6	15.9
wc2.0_bio_30s_01.tif - Среднегодовая температура	10	7.6
wc2.0_bio_30s_05.tif - Максимальная температура наиболее теплого месяца	7.9	5.2
modcf_meanannual.tiff.tif - Среднегодовая облачность	7.2	2.3
wc2.0_bio_30s_15.tif - Сезонность осадков (коэффициент вариации)	5.6	26.4

wc2.0_bio_30s_19.tif - Количество осадков наиболее холодного квартала	5.2	19
wc2.0_bio_30s_03.tif - Изотермальность (отношение среднегодовой температуры к среднегодовой амплитуде температур умножено на 100)	3.3	3.8
cv_01_05_1km_uint16.tif - Гетерогенность ландшафта – коэффициент вариации	2.5	3.3
modcf_intraannualsd.tiff.tif - Межгодовая вариация облачности	2.4	3.5
bdricm_m_250m_ll.tif - Мощность рыхлых отложений	2.2	1.3
wc2.0_bio_30s_02.tif - Среднесуточная амплитуда	1.8	1.9
shannon_01_05_1km_uint16.tif - Гетерогенность ландшафта – индекс Шеннона	0.6	0.9

Табл. 4. Процентный вклад переменных в построение модели распространения *T. pellucida* в России и их важность при пермутации (%).

Для определения важности переменных при построении модели распространения *T. pellucida* на территории России, как и по всей планете, мы использовали результирующую таблицу (табл. 4) и jackknife-тесты (рис. 17-19). Для всех переменных данные были взяты с разрешением в 30 сек.

Судя по вкладу отдельных предикторов (percent contribution), наибольший вклад в модель 1 вносят переменные wc2.0_bio_30s_07 – среднегодовая амплитуда температур, wc2.0_bio_30s_13 - количество осадков наиболее влажного месяца и wc2.0_bio_30s_01 – среднегодовая температура.

Наиболее значимыми предикторами в столбце «permutation importance» оказались wc2.0_bio_30s_15 – сезонность осадков, wc2.0_bio_30s_19 - количество осадков в наиболее холодный квартал и wc2.0_bio_30s_13 - количество осадков наиболее влажного месяца.

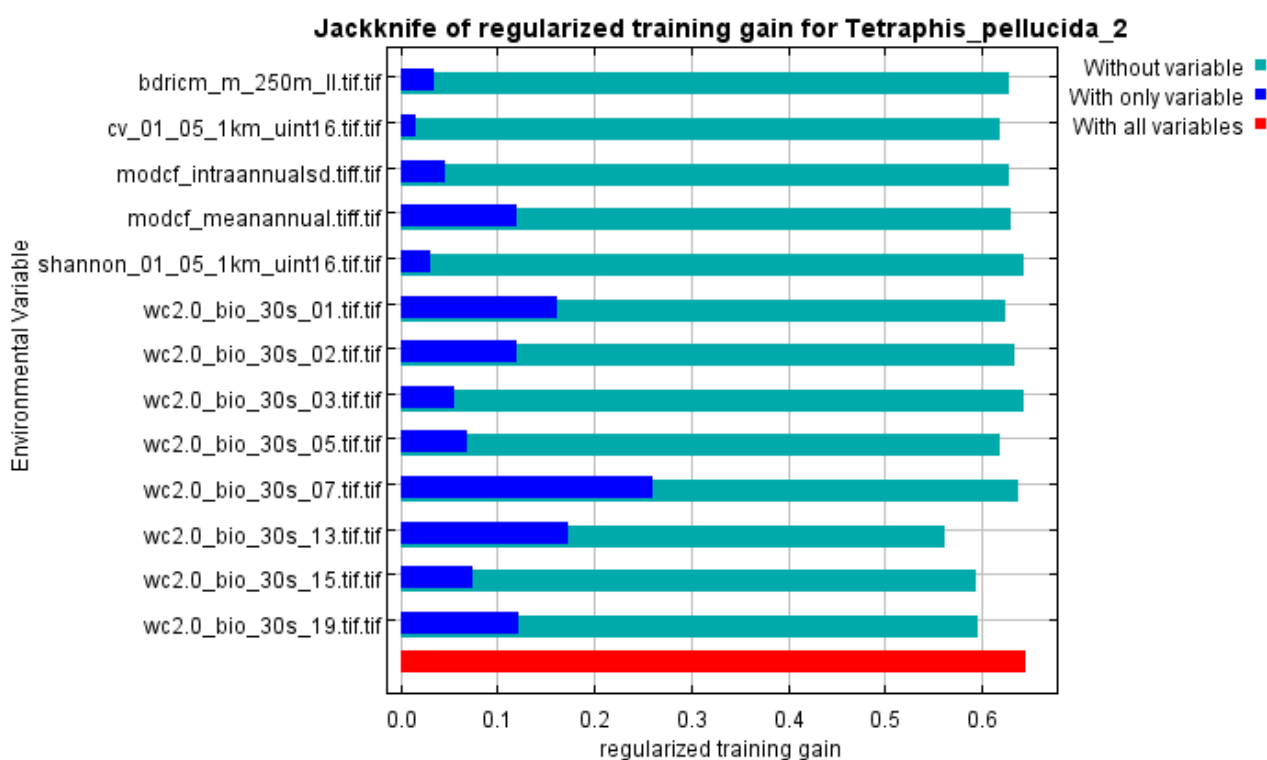


Рис. 17. Диаграмма, показывающая результат jackknife-теста по *T. pellucida* для обучающих данных (J1).

J1 свидетельствует о том, что наиболее бесполезная переменная для обучения модели распределения *T. pellucida* на территории России – cv_01_05_1km_uint16 - степень гетерогенности местообитаний, а наиболее удачно характеризующая тренировочные данные переменная - wc2.0_bio_30s_07 - среднегодовая амплитуда температур. Она несет наибольшее количество информации по сравнению с другими переменными; она же вносит и наибольший процентный вклад в построение модели (см. табл. 4).

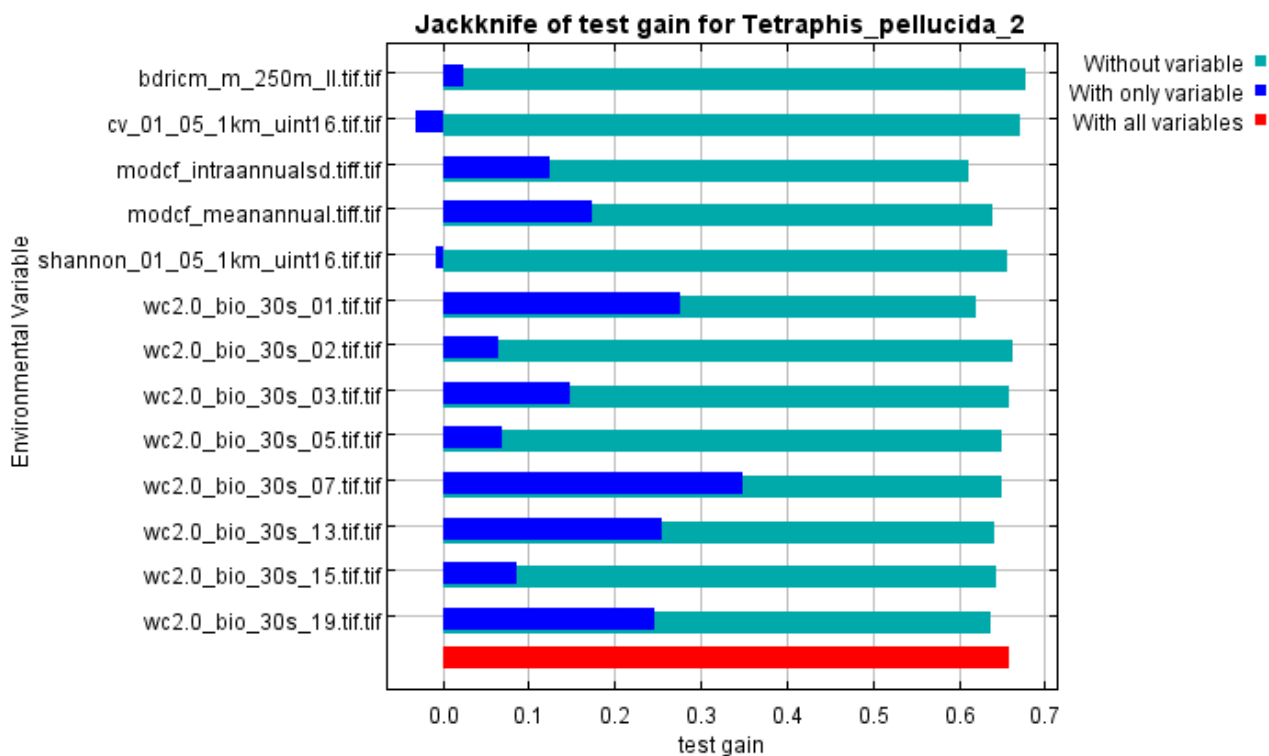


Рис. 18. Диаграмма, показывающая результат jackknife-теста по *T. pellucida* для тестовых данных (J2).

По J2 видно, что тестовый прирост при использовании каждой из двух переменных - cv_01_05_1km_uint16 - степень гетерогенности местообитаний и shannon_01_05_1km_uint16 – индекс Шеннона - отрицательный. Это значит, что модель получилась хуже, чем нулевая и что эти предикторы не подходят для прогнозирования распространения *T. pellucida* на территории России. Важность wc2.0_bio_30s_07 сохраняется.

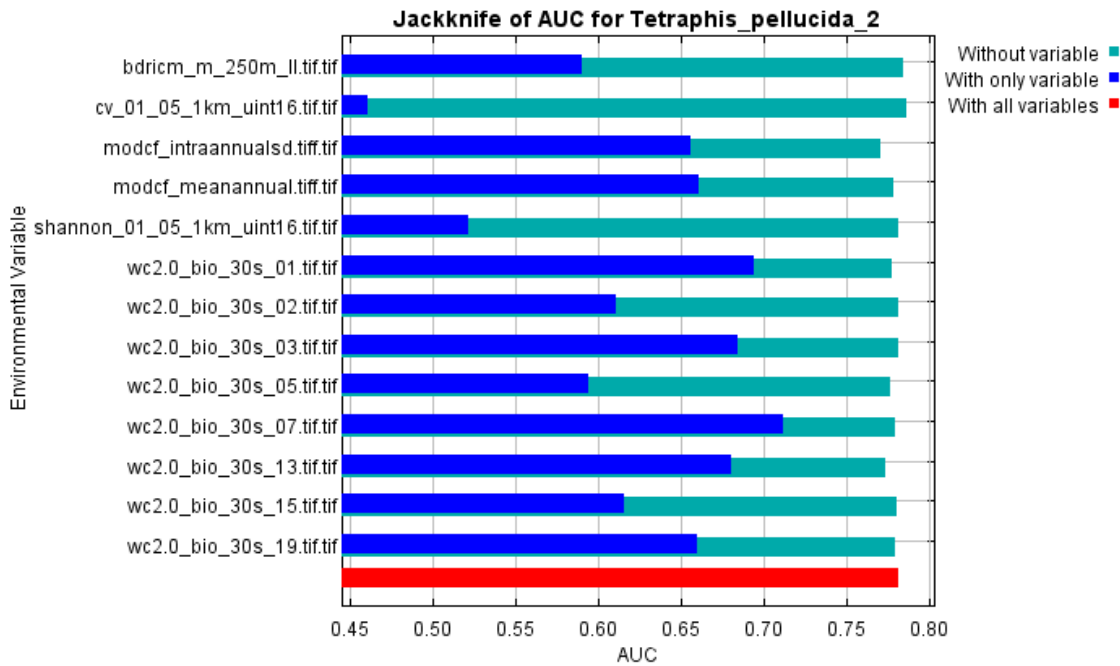


Рис. 19. Диаграмма, показывающая результат jackknife-теста для AUC по тестовым данным по *T. pellucida* (J3).

Наиболее важная переменная для увеличения test AUC - wc2.0_bio_30s_07, значимость которой показана также в табл. 4 и на двух предыдущих диаграммах. Бесполезность и неприменимость предикторов cv_01_05_1km_uint16 и shannon_01_05_1km_uint16 можно заметить и на J3, помимо J2.

Различные оценки значимости отдельных параметров, по-видимому, отражают сложные зависимости и скоррелированность между собой отдельных предикторов. Наиболее надежным, как обычно считается, является параметр "permutation importance", который подчеркивает важность предикторов, связанных с поддержанием постоянно влажности в местообитаниях *T. pellucida*.

1.2 Моделирование распространения *T. pellucida* по всему миру, только файл rar100

номер модели	1	2
Reg_mult	0,5	1
AUC training (площадь под ROC - кривой, построенной по точкам обучающей выборки)	0,934	0,9278
AUC test (площадь под ROC - кривой, построенной по точкам тестовой выборки)	0,9319	0,9255
Fixed cumulative value 1 training omission (омиссия по точкам обучающей выборки при фиксированном совокупном значении 1)	0,0062	0,0062
Fixed cumulative value 1 test omission (омиссия по тестовым точкам при фиксированном совокупном значении 1)	0	0,0072
Fixed cumulative value 5 training omission (омиссия по точкам обучающей выборки при фиксированном совокупном значении = 5)	0,0279	0,0372
Fixed cumulative value 5 test omission (омиссия по тестовым точкам при фиксированном совокупном значении = 5)	0,029	0,0217
Fixed cumulative value 10 training omission (омиссия по точкам обучающей выборки при фиксированном совокупном значении = 10)	0,0743	0,0712
Fixed cumulative value 10 test omission (омиссия по тестовым точкам при фиксированном совокупном значении = 10)	0,0797	0,087

Табл. 5. Результирующая таблица, полученная при моделировании с использованием файла rar100.

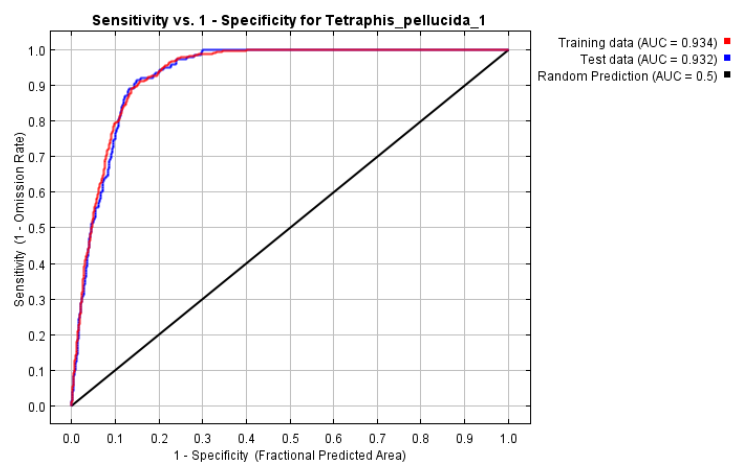


Рис. 20. График, отображающий ROC и AUC, построенные для модели 1 (RM = 0.5).

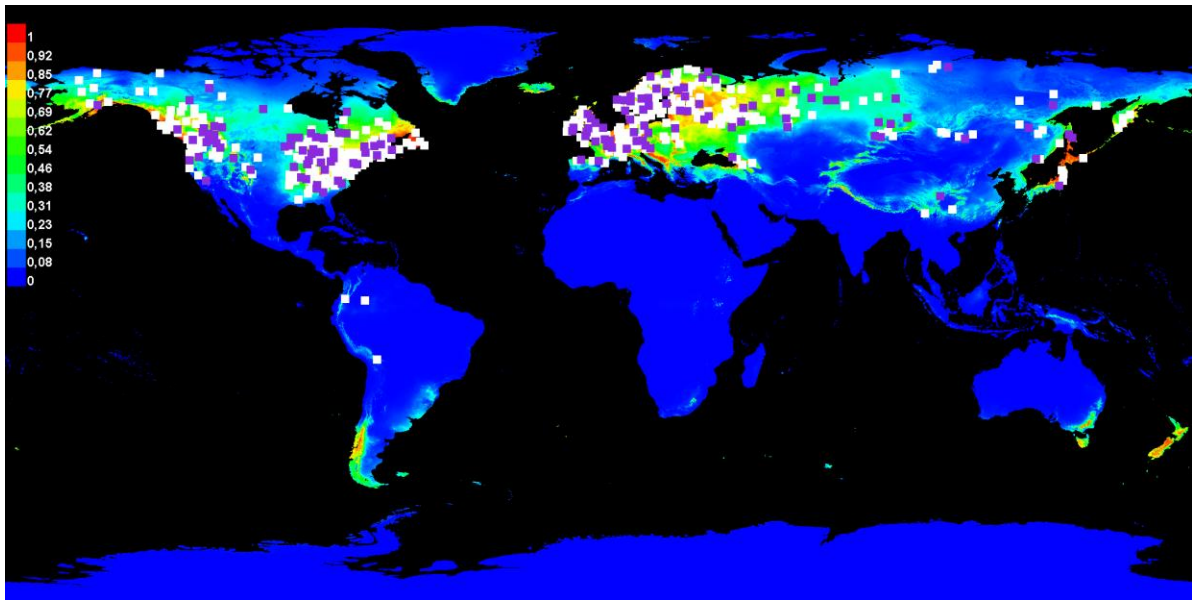


Рис. 21. Карта распространения *T. pellucida* по всему миру; файл rar100; модель 1 (AUC training = 0,934; AUC test = 0,9319).

Результаты дополнительного моделирования, проведенного с использованием файла rar100, как и результаты по двум выше описанным моделированиям, подтверждают тенденцию AUC быть больше при меньших значениях RM.

Хотя значения AUC, полученные при использовании файла rar100 и больше, чем при использовании других файлов с координатами распространения *T. pellucida* по всему миру, значения AUC по *T. geniculata* больше (см. ниже).

3. Моделирование общего распространения *T. geniculata*

В результате обработки данных о находках *T. geniculata* мы получили пять моделей, параметры пригодности которых представлены в табл. 6

Исходя из того, что AUC test наибольшая у модели 1 и равна 0,9858 (RM = 0.5), эта модель была выбрана в качестве наилучшей для прогноза распространения *T. geniculata* по миру.

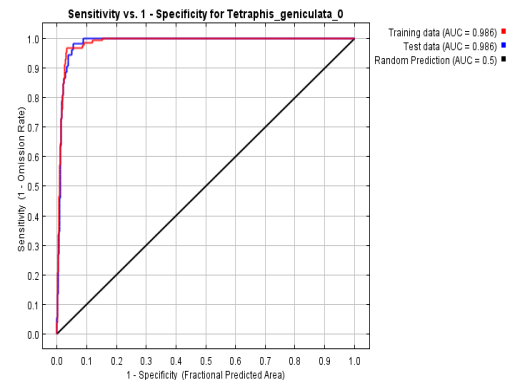


Рис. 22. График, отображающий ROC и AUC, построенные для модели 1. AUC training = 0,9863; AUC test = 0,9858

Значения обеих AUC модели 1, как и других моделей, полученных при моделировании 3, очень близки к единице; крайне высокая прогнозная способность модели.

номер модели	1	2	3	4	5
Reg_mult	0,5	1	1,5	2,5	5
AUC training (площадь под ROC - кривой, построенной по точкам обучающей выборки)	0,9863	0,9855	0,9833	0,9824	0,9821
AUC test (площадь под ROC - кривой, построенной по точкам тестовой выборки)	0,9858	0,9853	0,9811	0,9814	0,973
Fixed cumulative value 1 training omission (омиссия по точкам обучающей выборки при фиксированном совокупном значении 1)	0,0081	0,0081	0,0081	0,0081	0
Fixed cumulative value 1 test omission (омиссия по тестовым точкам при фиксированном совокупном значении 1)	0	0,0192	0	0	0
Fixed cumulative value 5 training omission (омиссия по точкам обучающей выборки при фиксированном совокупном значении = 5)	0,0323	0,0242	0,0242	0,0161	0
Fixed cumulative value 5 test omission (омиссия по тестовым точкам при фиксированном совокупном значении = 5)	0,0192	0,0385	0	0,0192	0,0577
Fixed cumulative value 10 training omission (омиссия по точкам обучающей выборки при фиксированном совокупном значении = 10)	0,0323	0,0484	0,0484	0,0161	0,0081
Fixed cumulative value 10 test omission (омиссия по тестовым точкам при фиксированном совокупном значении = 10)	0,0577	0,0577	0,0769	0,0385	0,0577

Табл. 6. Результирующая таблица, демонстрирующая результаты по наиболее удачным моделям распространения *T. geniculata* по всему миру

Построенные модели

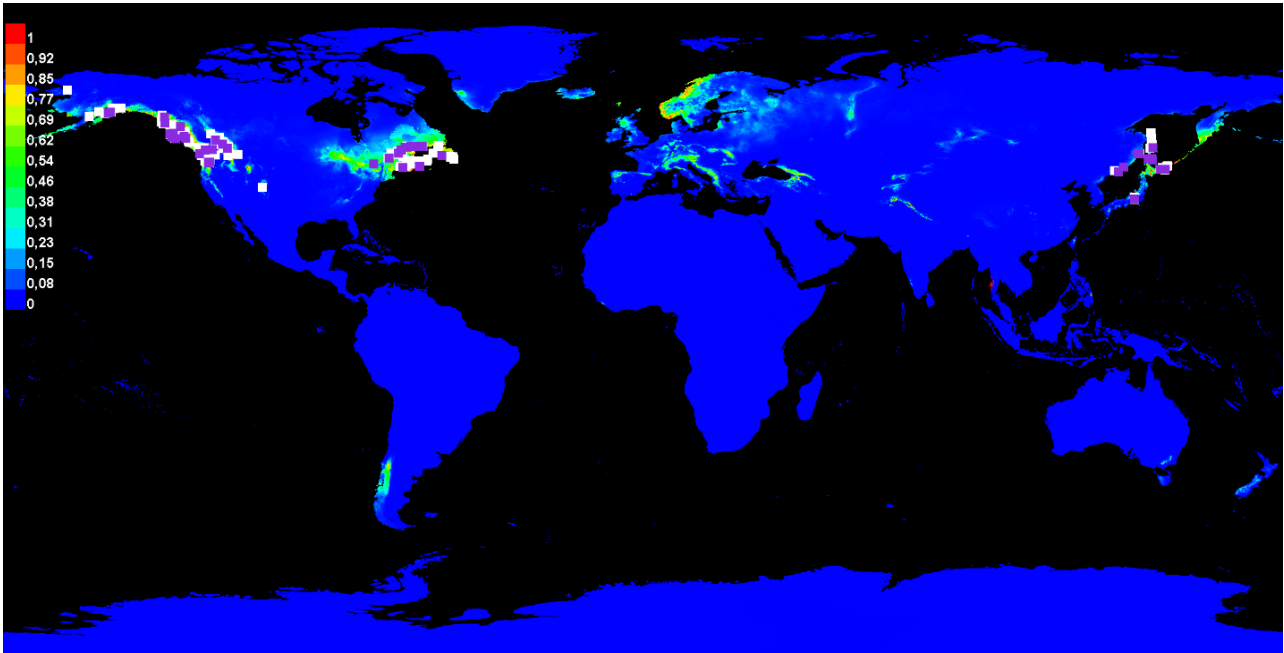


Рис. 23. Карта, полученная при моделировании распространения *T. geniculata* на всей территории планеты; regularization multiplier = 0.5, модель 1 (AUC training = 0,9863; AUC test = 0,9858).

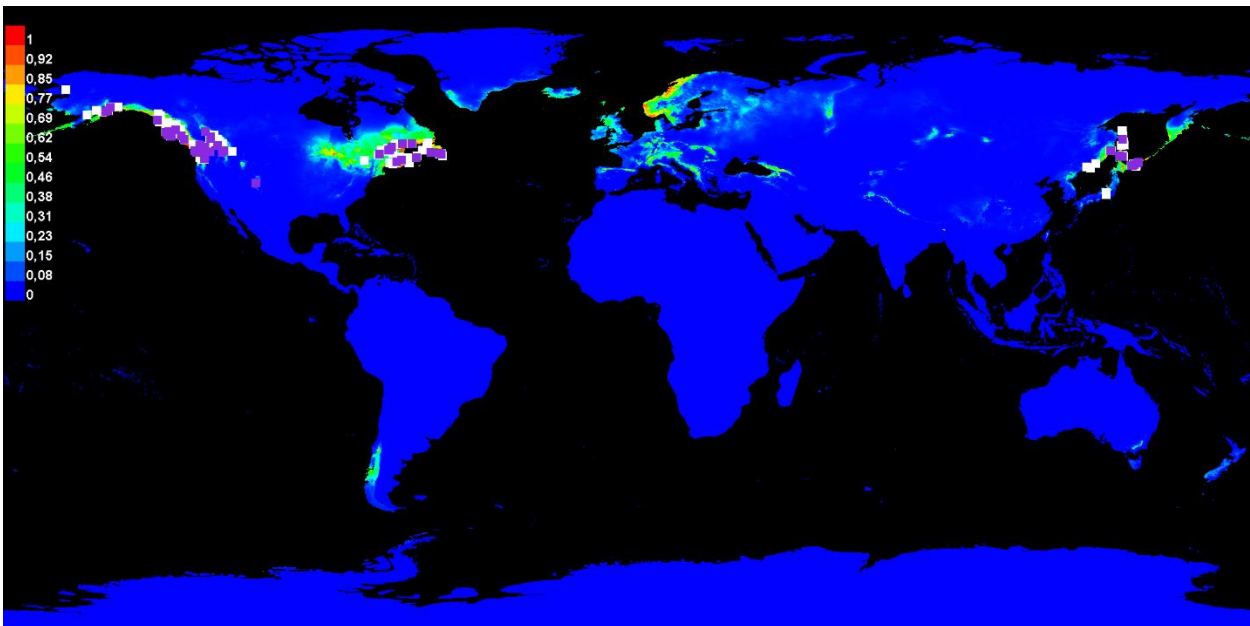


Рис. 24. Карта, полученная при моделировании распространения *T. geniculata* на всей территории планеты; regularization multiplier = 1, модель 2; (AUC training = 0,9855; AUC test = 0,9853)

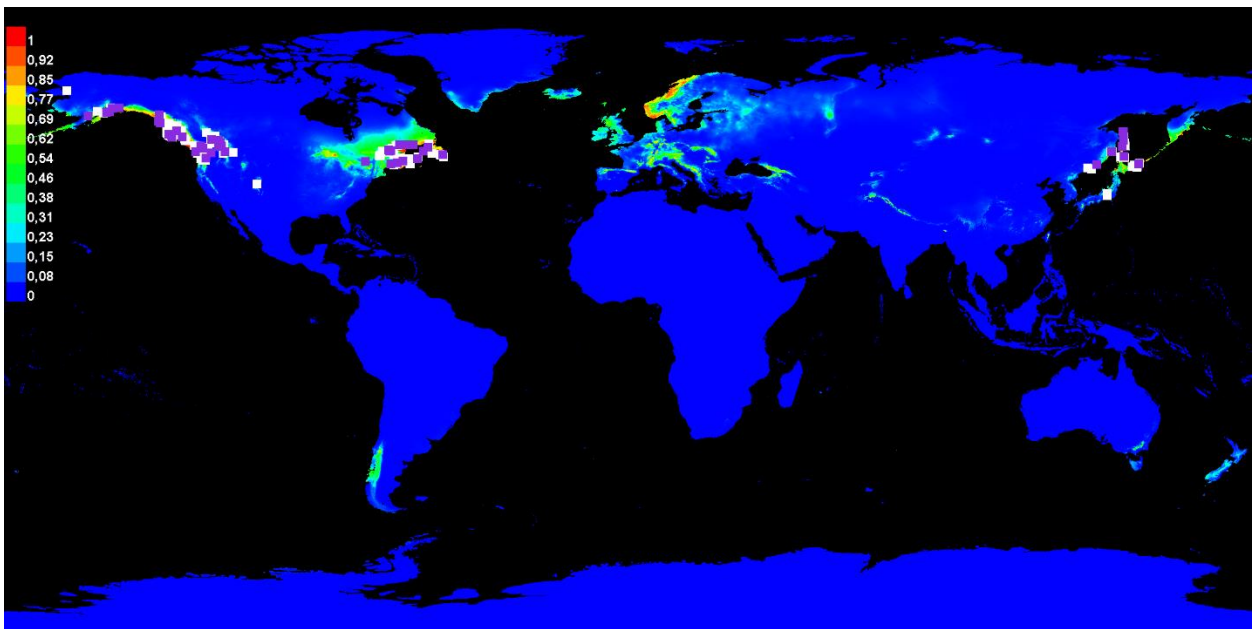


Рис. 25. Карта, полученная при моделировании распространения *T. geniculata* на всей территории планеты; regularization multiplier = 1.5, модель 3 (AUC training = 0,9833; AUC test = 0,9811).

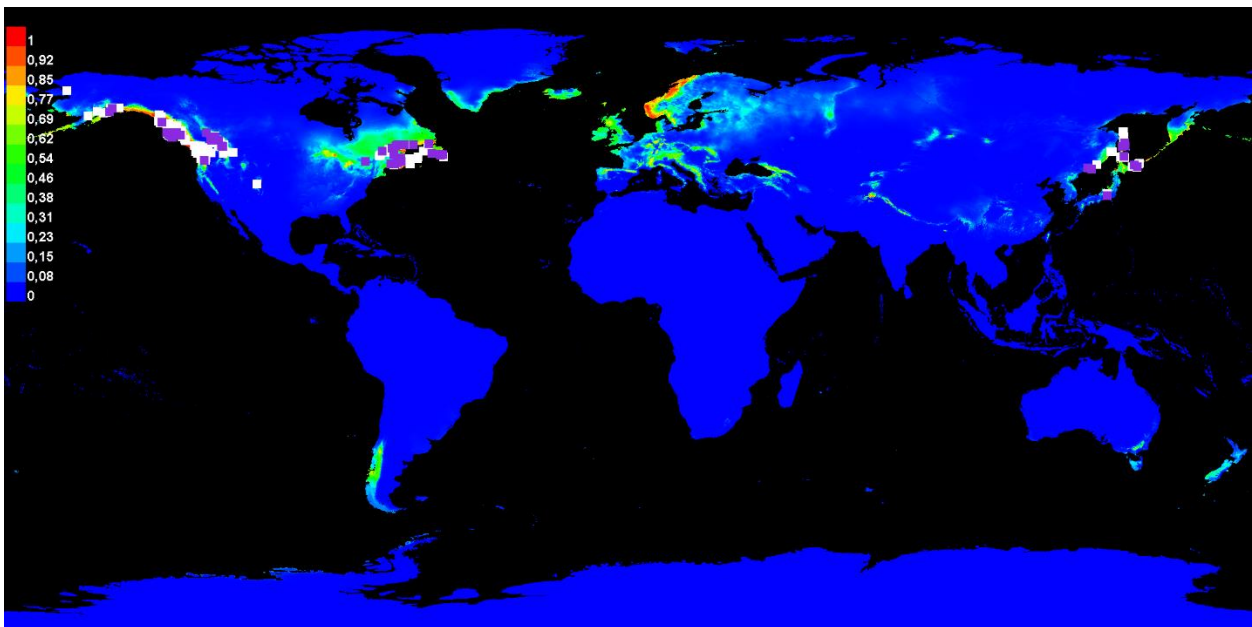


Рис. 26. Карта, полученная при моделировании распространения *T. geniculata* на всей территории планеты; regularization multiplier = 2.5, модель 4 (AUC training = 0,9824; AUC test = 0,9814).

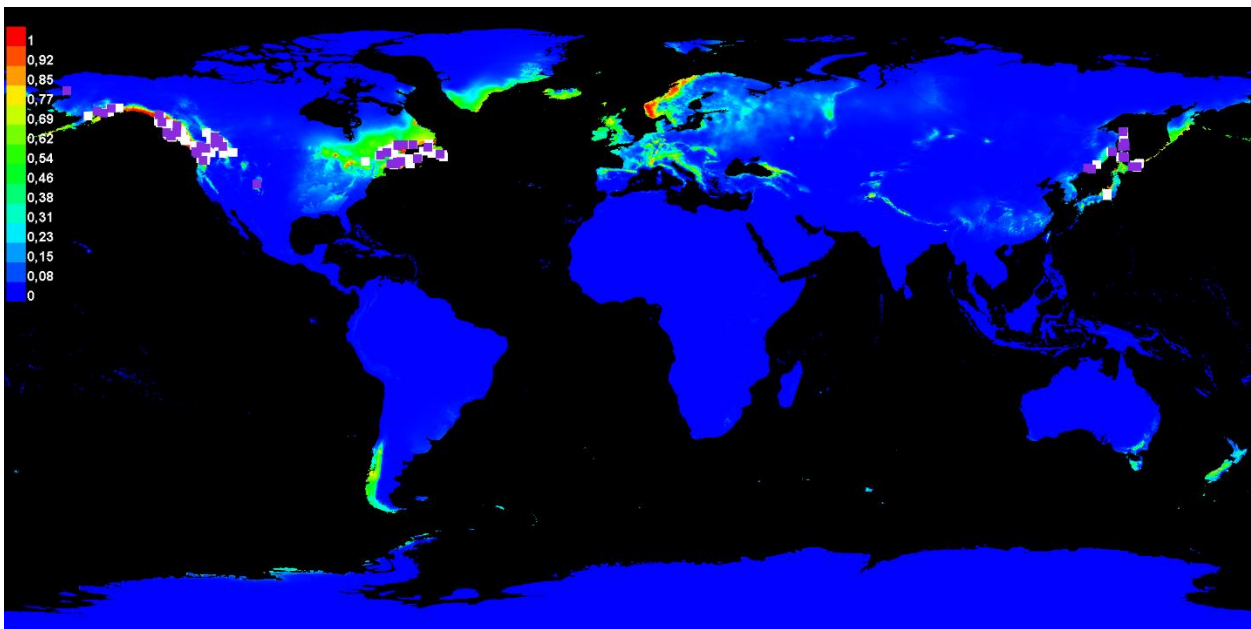


Рис. 27. Карта, полученная при моделировании распространения *T. geniculata* на всей территории планеты; regularization multiplier = 5, модель 5 (AUC training = 0,9821; AUC test = 0,973).

Анализ вклада переменных

Variable	Percent contribution	Permutation importance
wc2.0_bio_10m_14.tif - Количество осадков наиболее сухого месяца	32.2	3.7
wc2.0_bio_10m_01.tif - Среднегодовая температура	26.3	3.8
wc2.0_bio_10m_19.tif - Количество осадков наиболее холодного квартала	7.5	4.3
wc2.0_bio_10m_12.tif - Годовые осадки	7.2	14.9
wc2.0_bio_10m_17.tif - Количество осадков наиболее сухого квартала	5.3	6.7
wc2.0_bio_10m_04.tif - Температурная сезонность	4.9	2.1
wc2.0_bio_10m_10.tif - Средняя температура наиболее теплого квартала	3.4	8.6
wc2.0_bio_10m_05.tif - Максимальная температура наиболее теплого месяца	3.1	2.6
wc2.0_bio_10m_08.tif - Средняя температура наиболее влажного квартала	2.5	6.1
wc2.0_bio_10m_03.tif - Изотермальность (отношение среднегодовой температуры к среднегодовой амплитуде температур умножено на 100)	1.3	9.5
wc2.0_bio_10m_11.tif - Средняя температура наиболее холодного квартала	1.2	19.6
wc2.0_bio_10m_09.tif - Средняя температура наиболее сухого квартала	1.2	7.3
wc2.0_bio_10m_15.tif - Сезонность осадков (коэффициент вариации)	1	0.3
wc2.0_bio_10m_07.tif - Среднегодовая амплитуда температур	0.8	0.9
wc2.0_bio_10m_13.tif - Количество осадков наиболее влажного месяца	0.7	2.7
wc2.0_bio_10m_16.tif - Количество осадков наиболее влажного квартала	0.5	0.9
wc2.0_bio_10m_06.tif - Минимальная температура наиболее холодного месяца	0.3	4.2

wc2.0_bio_10m_02.tif - Среднесуточная амплитуда температур	0.3	0.2
wc2.0_bio_10m_18.tif - Количество осадков наиболее теплого квартала	0.2	1.6

Табл. 7. Процентный вклад переменных в построение модели общего распространения *T. geniculata* и важность при пермутации (%).

Столбец таблицы 7, показывающий процентный вклад, свидетельствует о том, что самые значимые переменные - wc2.0_bio_10m_14 - количество осадков в наиболее сухой месяци wc2.0_bio_10m_01 - среднегодовая температура; permutation importance - wc2.0_bio_10m_11 – средняя температура наиболее холодного квартала и wc2.0_bio_10m_12 - годовые осадки.

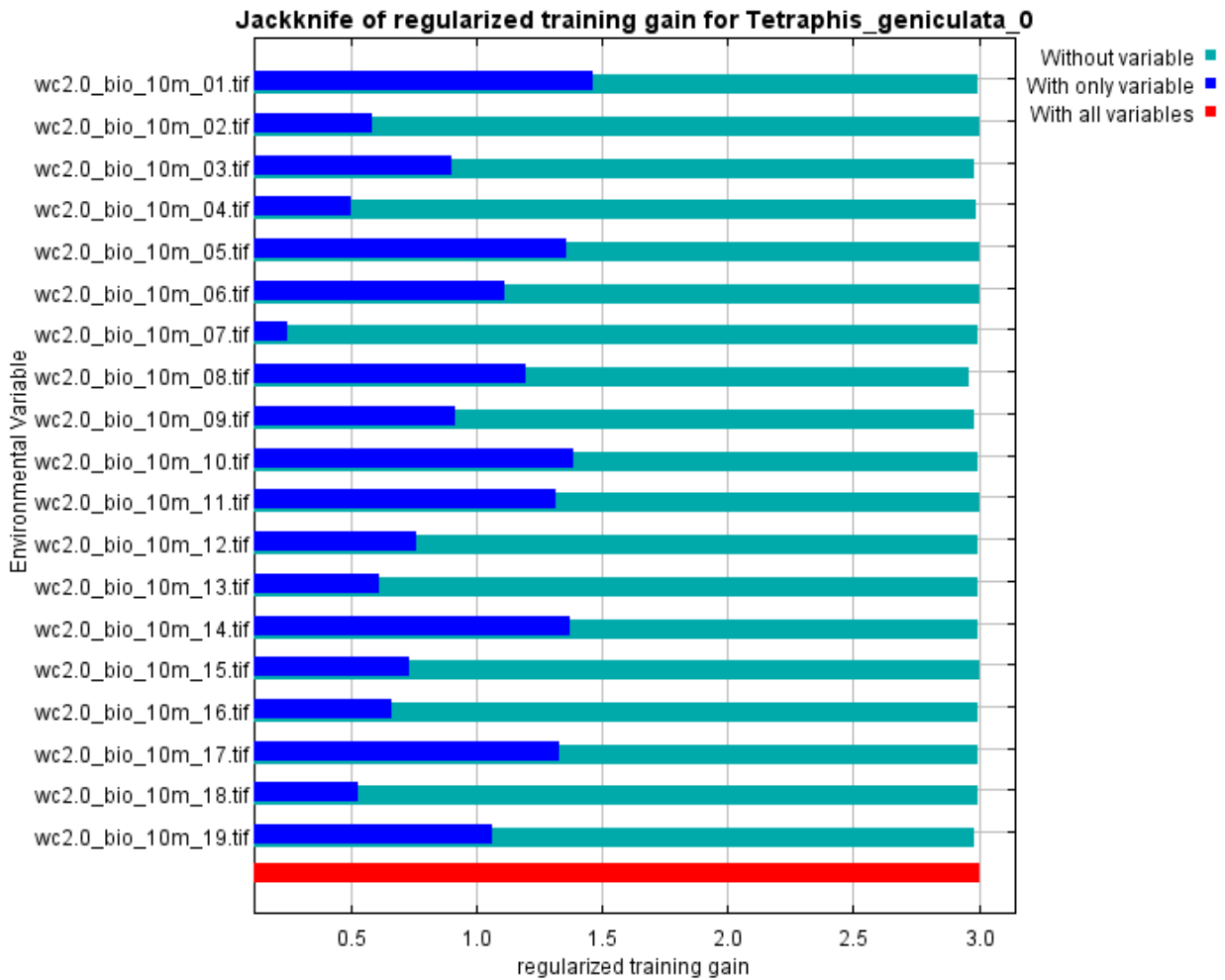


Рис. 28. Диаграмма, показывающая результат jackknife-теста по *T. geniculata* для обучающих данных (J1).

По J1 можно сделать вывод о том, что среди всех использованных предикторов нет ни одного, в котором бы содержалась информации, отсутствующая в других. Что касается важности предикторов при построении модели только с использованием одного из них, шесть предикторов вызывают сопоставимый прирост, это - wc2.0_bio_10m_01 - среднегодовая температура, wc2.0_bio_10m_05 - максимальная температура наиболее теплого месяца, wc2.0_bio_10m_10 - средняя температура наиболее теплого квартала, wc2.0_bio_10m_11 - средняя температура наиболее холодного квартала, wc2.0_bio_10m_14 - количество осадков в наиболее сухой месяц, wc2.0_bio_10m_17 - количество осадков в наиболее сухой

квартал. Среди этих переменных нельзя выделить какую-то одну, с помощью которой в отсутствие других можно было бы построить хорошую обучающую модель, хотя наибольший прирост достигается за счет wc2.0_bio_10m_01. Наименьший прирост при обучении модели – за счет предиктора wc2.0_bio_10m_07 - годовая амплитуда температур.

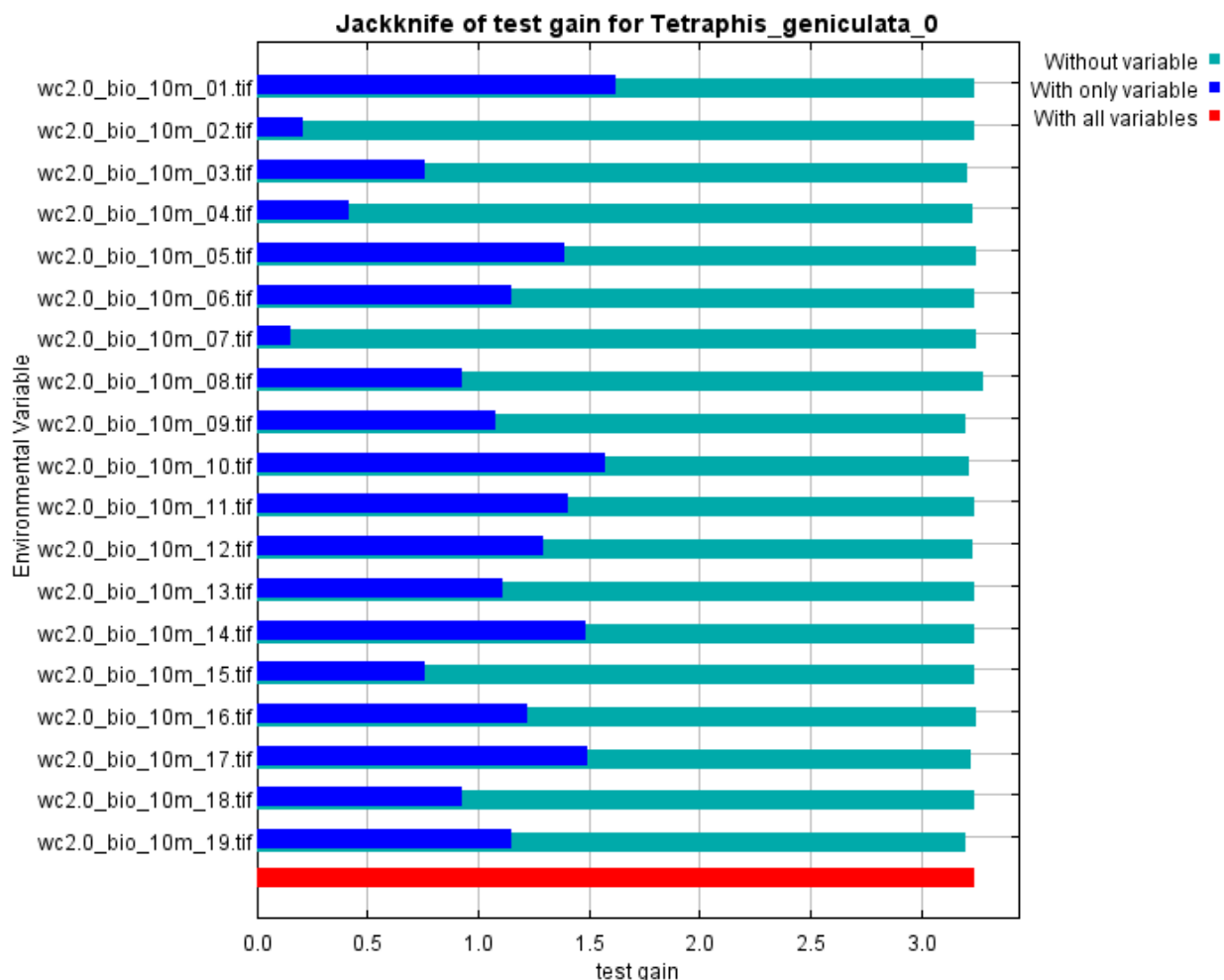


Рис. 29. Диаграмма, показывающая результат jackknife-теста по *T. geniculata* для тестовых данных (J2).

Предиктор, вклад которого как по тестовым, так и по тренировочным данным наибольший - wc2.0_bio_10m_01 - среднегодовая температура. Из прочих переменных имеющих большой вклад в J2 - такие же, как на J1 – это wc2.0_bio_10m_05, wc2.0_bio_10m_10, wc2.0_bio_10m_14, wc2.0_bio_10m_17, то есть максимальная температура наиболее теплого месяца, средняя температура наиболее теплого квартала, количество осадков в наиболее сухой месяц, количество осадков в наиболее сухой квартал. Наименьший прирост, помимо переменной wc2.0_bio_10m_07, вызывается переменной wc2.0_bio_10m_02 - среднесуточная амплитуда температур.

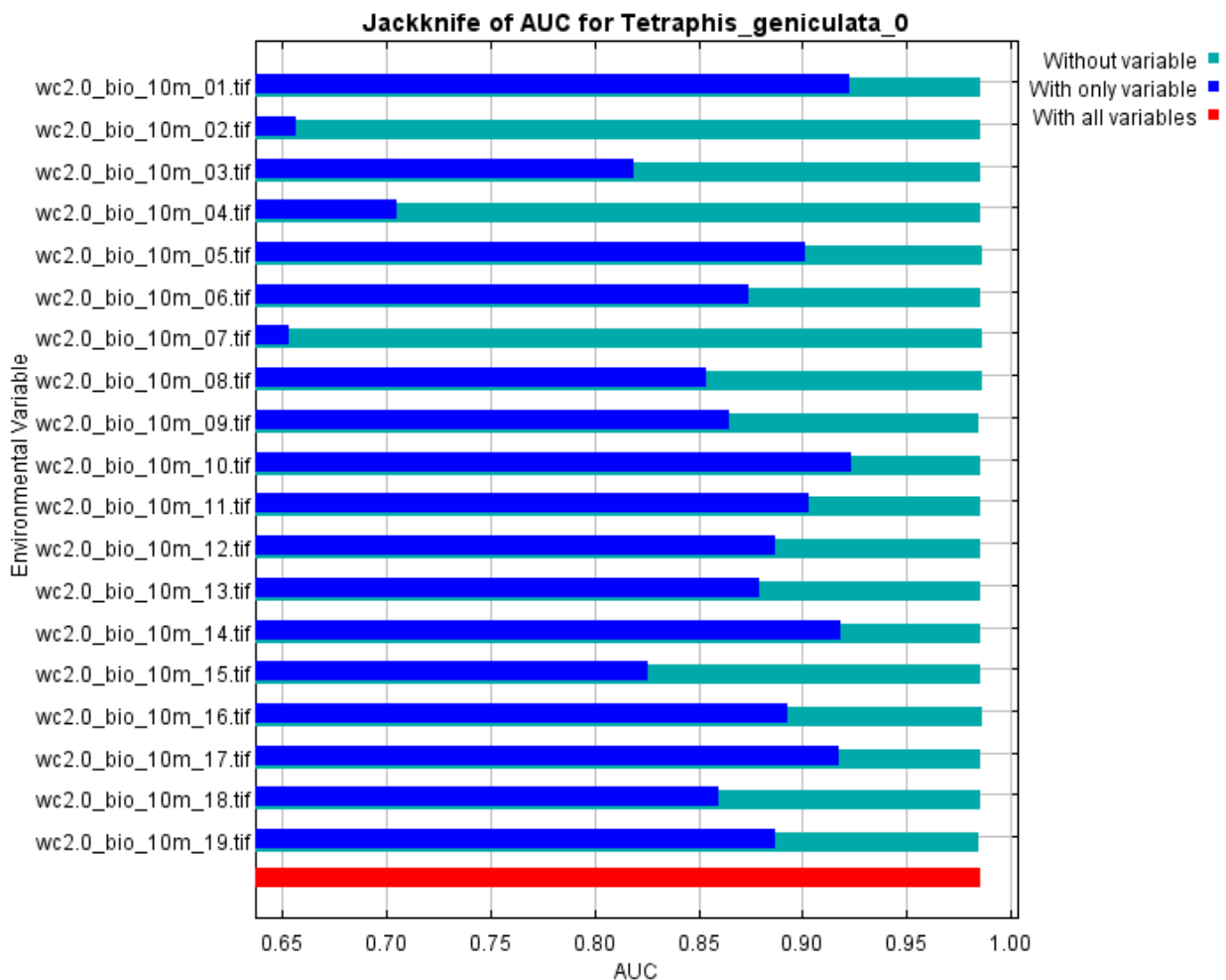


Рис. 30. Диаграмма, показывающая результат jackknife-теста для AUC test по *T. geniculata* (J3).

Диаграмма, показывающая AUC test по каждой из переменных, подтверждает результаты, показанные на J1 и J2 – наибольший вклад в построение модели, судя по этим трем диаграммам имеют четыре вышеуказанные переменные. Наименее увеличивающие прирост предикторы – отмеченные выше wc2.0_bio_10m_07, wc2.0_bio_10m_02 и wc2.0_bio_10m_04 - температурная сезонность. Если рассмотреть третий столбец табл. 2, можно заметить, что при пермутации этих трех переменных довольно незначительно уменьшается AUC training (соответственно на 0.9, 0.2, 2.1%). На основе этого можно сделать вывод о том, что использовать при моделировании общего распространения *T. geniculata* годовой амплитуды температур, среднесуточной амплитуды температур и температурной сезонности в качестве предикторов – не лучший выбор.

Так же, как и в случае с *Tetraphis pellucida*, наиболее существенными для объяснения распространения вида представляются два предиктора - средняя температура наиболее холодного квартала и годовые осадки. Действительно, *T. geniculata* имеет гораздо более южное (по сравнению с *T. pellucida*) распространение, которое более определено тяготеет к районам с океаническим климатом.

Заключение

Отбор моделей

Предсказанные с помощью программы Maxent ареалы, которые мы получили для двух видов, *T. pellucida* и *T. geniculata*, по большей части соответствуют реальным. Значения AUC всех выбранных моделей больше 0.8, причем AUC по *T. geniculata* > 0.98, что говорит о высокой прогностичности этих моделей.

Мы подтвердили, что значения обеих AUC выше у более сложных пространственных моделей: во всех случаях моделирования с изменением параметра сложности (RM) лучшие модели получались при RM = 0.5. Разреживание, сделанное для файла, содержащего точки находок *T. pellucida* по всему миру с шагом в 100 км (моделирование 1), оказывает положительное влияние на значения AUC для тренировочных и тестовых выборок (см. табл. 1).

Обсуждение распространения *T. pellucida*

Результаты нашей работы показали, что Maxent применим к моделированию распространения мхов рода *Tetraphis*. Предсказанные территории отсутствия *T. pellucida* соответствуют реальным: этот вид не предсказывается в зоне вечной мерзлоты, для него не характерен континентальный климат; на карте распространения *T. pellucida* в Северной Америке заметно, что этот вид отсутствует в южных центральных штатах США (см рис. 3).

T. pellucida предсказывается в Европе, на Востоке России, в Китае, в Северной Америке, что также соответствует его реальному распространению, однако, он предсказывается и в местах, где его нет: в Чили, на юго-востоке Австралии, в Тасмании и Новой Зеландии. В Чили, вероятно, так происходит из-за наличия там подходящих для *T. pellucida* влажных хвойных лесов. Отсутствие вида на этих неверно предсказанных территориях объясняется невозможностью переноса диаспор с Востока Евразии на столь дальние расстояния и, к тому же, наличием естественной преграды – Тихого океана («Древнейшие на суше», 1984).

Обсуждение распространения *T. geniculata*

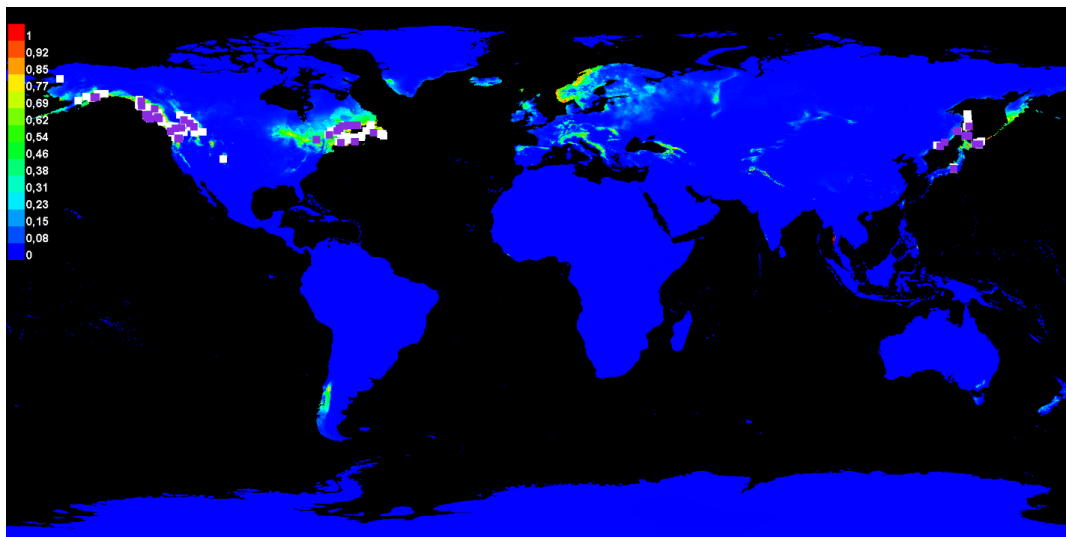


Рис. 23. Карта, полученная при моделировании распространения *T. geniculata* на всей территории планеты; regularization multiplier = 0.5, модель 1.

Таким образом, в случае *T. geniculata* мы имеем значительное расхождение между его предсказанным и реальным ареалами. Предсказанное распространение в Европе не подтверждается. Объяснение этому можно связать с тем, что распространение в данном случае определяется не только климатическими факторами, но и другими причинами. Из известных примеров подобного распространения можно привести распространение родов семейства Taxodiaceae, к примеру – *Sequoia* Endl. и *Taxodium* Rich.

Европейская древесная флора эпохи плиоцена помимо многих родов, типичных для современных европейских лесов таких, как *Quercus* L., *Carpinus* L., *Fagus* L., *Pinus* L., *Picea* A. Dietr., *Abies* Mill., содержала и роды, встречающиеся сейчас в Восточной Азии и/или в восточной части Северной Америки, в том числе и *Sequoia* и *Taxodium*. Эти деревья принадлежат к так называемой аркто-третичной флоре, впервые выделенной Gardner и Ettinghausen в 1869, которая во время неогена существовала на территории Северного полушария в Северной Америке, Европе и Азии. Последующее исчезновение термофильных высокоствольных хвойных, таких как *Sequoia*, *Taxodium*, *Cryptomeria* и др. во время плейстоцена объясняется гипотезой, выдвинутой еще в 1850х годах американским ботаником А. Греем. Холодные фазы в течение плиоцена и ледниковый период в последующем плейстоцене в совокупности с западно-восточными горными цепями, такими, как Пиринеи, Альпы, Карпаты, Кавказ и Средиземное море сформировали барьеры к южному отступлению многих видов аркто-третичной геофлоры, в результате чего произошло массовое вымирание этих видов в Европе. В горных цепях и долинах юго-восточной Азии, к примеру, Юньнань и Северной Америки, например, в Скалистых горах, Аппалачах, напротив, был возможен уход с северо-запада на юго-восток или с севера на юг, достигающий низких широт за счет отсутствия барьеров в виде морей и тем самым позволяющий деревьям умеренного климата распространиться к югу на непокрытых льдом территориях или переждать неблагоприятный период и

распространиться к северу в период межледниковья. Из-за западно-восточных барьеров на относительно холодных территориях в Европе вымерло большое количество родов деревьев, которые сейчас встречаются в составе вечнозеленых лесов на юго-востоке Китая. Они были сменены деревьями «умеренных смешанных мезофитных лесов». Многие таксоны исчезли уже с начала четвертичного периода, однако другие, в том числе *Sequoia* и *Taxodium* сохранились дольше, затем окончательно исчезнув в течение раннего и среднего четвертичного периода (Birks, Tinner, 2016).

На полученной карте распространения *T. geniculata* предсказывается на западной части Скандинавского полуострова, омываемой Норвежским морем. На самом деле этого вида, в отличие от *T. pellucida*, там нет, а предсказание его в этом регионе обуславливается, вероятно, наличием подходящих местообитаний – таежных лесов из сосны и ели.

Рис. 23 показывает выраженную дизъюнктивность предсказанного распространения *T. geniculata*, которое соответствует реальности (см. рис. 23 и 31).



Рис. 31. Находки *T. geniculata* в Северной Америке.
(http://www.efloras.org/object_page.aspx?object_id=84746&flora_id=1)

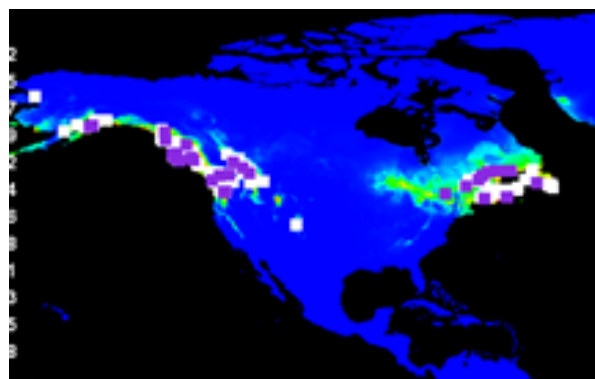


Рис. 23. Показано распространение *T. geniculata* только на территории Северной Америки

Наиболее важные выявленные предикторы

Как и во многих работах с использованием Maxent, определение вклада отдельных факторов оказывается довольно противоречивым, различающимся при применении разных критериев. Так, при моделировании общего распространения *T. pellucida* и *T. geniculata* jackknife-тест выявляет особо значительный вклад среднегодовой температуры. Действительно, общая конфигурация ареала *T. pellucida* циркумбореальная, в целом определяющаяся температурным режимом (средняя температура воздуха в июле от +14° до +18°C). Однако тот факт, что для модели того же вида для России этот предиктор имеет не первостепенное значение говорит о невысокой ценности этого теста. Показательно и то, что для России распространение *T. pellucida* больше определяется годовой амплитудой температур (что соответствует редкой встречаемости в зоне вечной мерзлоты), тогда как в мировом распространении по результатам теста jackknife этот предиктор имеет наименьший вклад.

Результирующие таблицы (см. разделы «Анализ вклада переменных») свидетельствуют о том, что у моделей, построенных по обоим изучаемым видам на весь мир, наибольший вклад имеет количество осадков в наиболее сухой месяц. Это легко объяснить тем, что представители рода *Tetraphis* растут преимущественно в постоянно влажных местообитаниях и не выносят длительного пересыхания. Модель *T. pellucida* в России также имеет наибольший вклад этой переменной.

Наиболее важные переменные, выявленные с помощью “permutation importance” для *T. pellucida* получились разные для мира и для России. Для мира это `wc2.0_bio_10m_05` - максимальная температура наиболее теплого месяца, а для России - `wc2.0_bio_30s_15` - сезонность осадков. Можно интерпретировать максимальную температуру наиболее теплого месяца также как параметр, косвенно определяющий отсутствие длительного пересыхания во время летней засухи, однако, правильной будет признать, что определение вклада отдельных предикторов не является наиболее сильной стороной данного метода, по крайней мере при использовании такого относительно небольшого числа параметров. В целом же программа Maxent обладает значительным потенциалом для изучения распространения видов и требует дальнейшего изучения приемов ее использования.

Выводы

Используя в качестве модельных объектов оба вида рода *Tetraphis*, мы выяснили и подтвердили, что программу Maxent можно использовать для моделирования распространения мхов, однако в случаях моделирования таких видов, как *T. geniculata*, когда распространение вида определяется не только климатическими факторами, но и историческими, ледниковым периодом, Maxent не учитывает никакие параметры, кроме климатических, в результате чего часть ареала прогнозируется неверно. По результатам моделирования были выбраны оптимальные модели для предсказания возможных местонахождений *Tetraphis pellucida* и *Tetraphis geniculata*, и с помощью разных параметров была оценена их прогнозная способность. Наиболее удачными моделями оказались пространственно сложные, «островные» модели. В случае общего моделирования распространения *T. pellucida*, лучшая модель получилась при обработке файла, разреженного с шагом в 100 км.

Благодарности

Я благодарю Михаила Станиславовича Игнатова за научное руководство, Сергея Валерьевича Дудова за помощь в получении данных для обработки и консультирование по методу работы, Марию Алексеевну Колесникову за помощь в работе с данными и Сергея Менделевича Глаголева за организацию практики.

Литература

1. База данных «Флоры мхов России». Иванов О. В., Колесникова М. А., Афонина О.М., Акатова Т.В., Баишева Э. З., Белкина О. А., Безгодов А. Г., Чернядьева И. В., Дудов С.В., Федосов В. Э., Игнатова Е. А., Иванова Е. И., Кожин М. Н., Лапшина Е. Д., Нотов А. А., Писаренко О. Ю., Попова Н. Н., Савченко А. Н., Телеганова В. В., Украинская Г. Ю., Игнатов М. С.; v. 26; http://arctoa.ru/ru/Archive-ru/26_1/Arctoa26_001_010.pdf (версия 03.09.2019).
2. Бардунов Л. В. «Древнейшие на суше», 1984; Академия наук СССР, Сибирское отделение. Издательство «Наука». Стр. 1-39.
3. Игнатов М.С., Игнатова Е.А., 2017; Флора мхов России; *Tetraphis.*; v. 2; стр. 115 -123; <http://arctoa.ru/Flora/taxonomy-ru/flora-2017-2023/Tetraphidaceae.pdf#Tetraphis> (версия 29.08.2018).
4. Игнатов М.С., Игнатова Е.А., 2003; Флора мхов средней части европейской России; v. 1; стр. 300 – 301.
5. Anderson, R.P., Peterson, A.T., Egbert., S.L., 2006. Vegetation-index models predict areas vulnerable to purple loosestrife (*Lythrum salicaria*) invasion in Kansas.; *Southwest. Nat.*; v. 51; p. p. 471–480.
6. Aspinall, R., 1992. An inductive modeling procedure based on Bayes' theorem for analysis of pattern in spatial data.; *Int. J. Geogr. Inf. Syst.*; v. 6; p. p. 105–121.
7. Birks, H. J. B., Tinner, W., 2016. Past forests of Europe.; In: SanMiguel-Ayanz, J., de Rigo, D., Caudullo, G., Houston Durrant, T., Mauri, A. (Eds.); *European Atlas of Forest Tree Species.*; p. p. 36-39.
8. Brzeziecki, B., Kienast, F., Wildi, O., 1993. A simulated map of the potential natural forest vegetation of Switzerland.; *J. Veg. Sci.*; v. 4; p. p. 499–508.
9. Číhal L, Kaláb O, Plášek V., 2017. Modeling the distribution of rare and interesting moss species of the family Orthotrichaceae (Bryophyta) in Tajikistan and Kyrgyzstan.; *Acta Soc Bot Pol.*; v. 86; p. p. 1-15.
10. Désamoré, A. E., Laenen, B. Ctech, M., Papp, B., Hedenäs, L., Mateo, R.G, Vanderpoorten, A., 2012. How do temperate bryophytes face the challenge of a changing environment? Lessons from the past and predictions for the future.; *Global Change Biology*; v. 18; p. p. 2915–2924.
11. Dimitriadis, S., Cranston, P.S., 2001. An Australian Holocene climate reconstruction using Chironomidae from a tropical volcanic maar lake.; *Palaeogeogr. Palaeoecol.*; v. 176; p. p. 109–131.

12. El-Gabbas, A., Dormann C. F., 2018. Improved species-occurrence predictions in data-poor regions: using large-scale data and bias correction with down-weighted Poisson regression and Maxent.; *Ecography*; v. 41; p. p. 1161–1172.
13. Fick, S. E., Hijmans, R. J., 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas.; *International Journal of Climatology*; v. 37; p. p. 4302–4315.
14. Guisan, A., Zimmermann, N., 2000. Predictive habitat distribution models in ecology. – *Ecological Modelling*; v. 135; p. p. 147–186.
15. Heikkinen, R.K., 1996. Predicting patterns of vascular plant species richness with composite variables: a meso-scale study in Finnish Lapland.; *Vegetatio*; v. 126; p. p. 151–165.
16. Hengl, T., de Jesus, J. M., Heuvelink, G. B. M., Gonzalez, M. R., Kilibarda, M., Blagotić, A., Shangguan, W. , Wright, M. N. , Geng, X., Bauer-Marschallinger, B., Guevara, M. A., Vargas, R., MacMillan, R. A., Batjes, N. H., Leenaars, J. G. B., Ribeiro, E., Wheeler, I., Mantel, S., Kempen, B., 2017. SoilGrids250m: Global gridded soil information based on machine learning.; *PLOS ONE*; v. 12; p. p. 1- 40.
17. Jeschke, J. M., Strayer, D. L., 2008. Usefulness of Bioclimatic Models for Studying Climate Change and Invasive Species; *Ann. N.Y. Acad. Sci.*; v. 1134; p. p. 1–24.
18. Kershaw, A.P., 1997. A bioclimatic analysis of early to middle Miocene brown coal floras, Latrobe Valley, south-eastern Australia; *Aust. J. Bot.*; v. 45; p. p. 373–387.
19. Kienast, F., Wildi, O., Brzeziecki, B., 1998. Potential impacts of climate change on species richness in mountain forests – an ecological risk assessment. *Biol. Conserv.*; v. 83; p. p. 291–305.
20. Kunzmann, L., Kvaček, Z., Mai, D. H., Walther, H., 2009. The genus *Taxodium* (Cupressaceae) in the Palaeogene and Neogene of Central Europe.; *Review of Palaeobotany and Palynology*; v. 153; p. p. 153–183.
21. Lehmann, A., Jaquet, J.-M., Lachavanne, J.-B., 1997. A GIS approach of aquatic plant spatial heterogeneity in relation to sediment and depth gradient, Lake Geneva, Switzerland.; *Aquat. Bot.*; p. p. 347–361.

22. Lischke, H., Guisan, A., Fischlin, A., Williams, J., Bugmann, H., 1998. Vegetation responses to climate change in the Alps-Modeling studies.; In: Cebon, P., Dahinden, U., Davies, H., Imboden, D., Jaeger, C. (Eds.); A View from the Alps: Regional Perspectives on Climate Change. MIT Press, Boston; p. p. 309–350.
23. Mastrorillo, S., Lek, S., Dauba, F., Belaud, A., 1997. The use of artificial neural network to predict the presence of small-bodied fish in a river.; *Freshw. Biol.*; v. 38; p. p. 237–246.
24. Mateo, R. G., Broennimann, O., Normand, S., Petitpierre, B., Araújo, M. B., Svenning, J.-C., Baselga, A., Fernández-González, F., Gómez-Rubio, V., Muñoz, J., Suarez, G. M. Luoto, M., Guisan, A., Vanderpoorten, A., 2016. The mossy north: an inverse latitudinal diversity gradient in European bryophytes.; *Scientific Reports*; v. 6, p. p. 1-9.
25. Mourell, C., Ezcurra, E., 1996. Species richness of Argentine cacti: A test of biogeographic hypotheses.; *Journal of Vegetation Science*; v. 7; p. p. 667-680.
26. Phillips, S., Andreson, R., Schapire., R., 2006. Maximum entropy modeling of species geographic distributions.; *Ecological Modelling*.; v. 190; p. p. 231–259.
27. Sandanov, D. V., Pisarenko. O. Y., 2018. Bioclimatic modeling of *Crossidium Squamiferum* (Viv.) Jur. (Pottiaceae, Bryophyta) distribution; *Arctoa*; v. 27; p. p. 29–34.
28. Shao, G., Halpin, P.N., 1995. Climatic controls of eastern North American coastal tree and shrub distributions.; *J. Biogeogr.*; v. 22; p. p. 1083–1089.
29. Sindel, B.M., Michael, P.W., 1992. Spread and potential distribution of *Senecio madagascarensis* Poir (fireweed) in Australia.; *Aust. J. Ecol.*; v. 17; p. p. 21–26.
30. Song, S., Liu, X., Bai, X., Jiang, Y., Zhang, X., Yu, C., Shao., X., 2015. Impacts of environmental heterogeneity on moss diversity and distribution of *Didymodon* (Pottiaceae) in Tibet, China.; *PLOS ONE*; v. 10; p. p. 1-17.
31. Strayer, D.L., 1991. Projected distribution of the zebra mussel, *Dreissena polymorpha*, in North America.; *Can. J. Fish. Aquat. Sci.*; v. 48; p. p. 1389–1395.
32. Thuiller, W. et al.; 2005. Niche-based modelling as a tool for predicting the risk of alien plant invasions at a global scale.; *Global Change Biol.*; v. 11; p. p. 2234– 2250.

33. Tuanmu, M.-N., Jetz, W., 2015. A global, remote sensing-based characterization of terrestrial habitat heterogeneity for biodiversity and ecosystem modelling.; *Global Ecology and Biogeography*; v. 24, №11; p. p. 1329–1339.
34. Underwood, E.C., Klinger, R., Moore, P.E., 2004. Predicting patterns of non-native plant invasions in Yosemite National Park, California, USA.; *Diversity Distrib.*; v. 10; p. p. 447–459.
35. Van de Rijt, C.W.C.J., Hazelhoff, L., Blom, C.W.P.M., 1996. Vegetation zonation in a former tidal area: a vegetationtype response model based on DCA and logistic regression using GIS.; *J. Veg. Sci.*; v. 7; p. p. 505–518.
36. Wang, L., Chen, J., & Shangguan, Z., 2016. Photosynthetic characteristics and nitrogen distribution of large-spike wheat in Northwest China.; *Journal of Integrative Agriculture*; v. 15; p. p. 545–552.
37. Wilson, A. M., & Jetz, W., 2016. Remotely Sensed High-Resolution Global Cloud Dynamics for Predicting Ecosystem and Biodiversity Distributions.; *PLOS Biology*; v. 14, №3; p. p. 1 – 20.

Приложение

номер модели	1	2	3	4	5
reg_mult	0,5	1	1,5	2,5	5
Training AUC	0,8491	0,8173	0,8083	0,7976	0,7817
Test AUC	0,819	0,8154	0,7743	0,7554	0,7572
Minimum training presence cumulative threshold	0,1861	0,6571	1,5375	0,7594	0,9606
Minimum training presence Cloglog threshold	0,0224	0,0594	0,096	0,0737	0,0966
Minimum training presence area	0,9197	0,8947	0,8595	0,9069	0,9095
Minimum training presence training omission	0	0	0	0	0
Minimum training presence test omission	0	0	0,012	0	0

Табл. 8. Результирующая таблица по распространению *T. pellucida* в России.

номер модели	1	2	3	4	5
reg_mult	0,5	1	1,5	2,5	5
Training AUC	0,8491	0,8173	0,8083	0,7976	0,7817
Test AUC	0,819	0,8154	0,7743	0,7554	0,7572
Maximum test sensitivity plus specificity cumulative threshold	27,0525	31,4156	16,8627	23,2638	20,1592
Maximum test sensitivity plus specificity Cloglog threshold	0,4187	0,4592	0,3137	0,4119	0,3986
Maximum test sensitivity plus specificity area	0,2954	0,2876	0,4597	0,392	0,4615
Maximum test sensitivity plus specificity training omission	0,1641	0,2513	0,1231	0,1692	0,1333
Maximum test sensitivity plus specificity test omission	0,1687	0,1928	0,1084	0,1928	0,1325

Табл. 9. Результирующая таблица по распространению *T. pellucida* в России.

Таблицы с параметрами оценки моделей, полученных по распространению *T. pellucida* в России (табл. 10-14.2)

Reg_mult	AUC training	AUC test	Fixed cumulative value 1 area	Fixed cumulative value 5 area	Fixed cumulative value 10 area
0,5	0,8491	0,819	0,8168	0,6311	0,5145

Табл. 10. Таблица с параметрами оценки модели 1.

Fixed cumulative value 1 Cloglog threshold	Fixed cumulative value 5 Cloglog threshold	Fixed cumulative value 10 Cloglog threshold	Fixed cumulative value 1 training omission	Fixed cumulative value 1 test omission	Fixed cumulative value 5 training omission	Fixed cumulative value 5 test omission	Fixed cumulative value 10 training omission	Fixed cumulative value 10 test omission
0,0571	0,1544	0,2331	0,0051	0,012	0,0359	0,0482	0,0462	0,0964

Табл. 10.2. Таблица с параметрами оценки модели 1.

Reg_mult	AUC training	AUC test	Fixed cumulative value 1 area	Fixed cumulative value 5 area	Fixed cumulative value 10 area
1	0,8173	0,8154	0,8671	0,6906	0,5669

Табл. 11. Таблица с параметрами оценки модели 2.

Fixed cumulative value 1 Cloglog threshold	Fixed cumulative value 5 Cloglog threshold	Fixed cumulative value 10 Cloglog threshold	Fixed cumulative value 1 training omission	Fixed cumulative value 1 test omission	Fixed cumulative value 5 training omission	Fixed cumulative value 5 test omission	Fixed cumulative value 10 training omission	Fixed cumulative value 10 test omission
0,0726	0,1621	0,2371	0,0051	0	0,0359	0,0241	0,0821	0,0602

Табл. 11.2. Таблица с параметрами оценки модели 2.

Reg_mult	AUC training	AUC test	Fixed cumulative value 1 area	Fixed cumulative value 5 area	Fixed cumulative value 10 area
1,5	0,8083	0,7743	0,8917	0,7148	0,5823

Табл. 12. Таблица с параметрами оценки модели 3.

Fixed cumulative value 1 Cloglog threshold	Fixed cumulative value 5 Cloglog threshold	Fixed cumulative value 10 Cloglog threshold	Fixed cumulative value 1 training omission	Fixed cumulative value 1 test omission	Fixed cumulative value 5 training omission	Fixed cumulative value 5 test omission	Fixed cumulative value 10 training omission	Fixed cumulative value 10 test omission
0,0808	0,1551	0,2252	0	0	0,0205	0,0361	0,0769	0,0723

Табл. 12.2. Таблица с параметрами оценки модели 3.

Reg_mult	AUC training	AUC test	Fixed cumulative value 1 area	Fixed cumulative value 5 area	Fixed cumulative value 10 area
2,5	0,7976	0,7554	0,8893	0,7135	0,5912

Табл. 13. Таблица с параметрами оценки модели 4.

Fixed cumulative value 1 Cloglog threshold	Fixed cumulative value 5 Cloglog threshold	Fixed cumulative value 10 Cloglog threshold	Fixed cumulative value 1 training omission	Fixed cumulative value 1 test omission	Fixed cumulative value 5 training omission	Fixed cumulative value 5 test omission	Fixed cumulative value 10 training omission	Fixed cumulative value 10 test omission
0,0832	0,1756	0,2563	0,0051	0	0,0359	0,0843	0,0615	0,1325

Табл. 13.2. Таблица с параметрами оценки модели 4.

Reg_mult	AUC training	AUC Test	Fixed cumulative value 1 area	Fixed cumulative value 5 area	Fixed cumulative value 10 area
5	0,7817	0,7572	0,9069	0,743	0,6245

Табл. 14. Таблица с параметрами оценки модели 5.

Fixed cumulative value 1 Cloglog threshold	Fixed cumulative value 5 Cloglog threshold	Fixed cumulative value 10 Cloglog threshold	Fixed cumulative value 1 training omission	Fixed cumulative value 1 test omission	Fixed cumulative value 5 training omission	Fixed cumulative value 5 test omission	Fixed cumulative value 10 training omission	Fixed cumulative value 10 test omission
0,0979	0,1983	0,2768	0,0103	0	0,0462	0,0602	0,0769	0,0843

Табл. 14.2. Таблица с параметрами оценки модели 5.