

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение
высшего образования

«ПЕНЗЕНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»

Пензенский педагогический институт им. В.Г. Белинского

Факультет
физико-математических и естественных наук

Кафедра
«Зоология и экология»

Направление подготовки

06.03.01 «Биология»

Профиль

«Биоэкология»

БАКАЛАВРСКАЯ РАБОТА

на тему:

**«ЭКОЛОГИЧЕСКИЕ АСПЕКТЫ В РАСПРОСТРАНЕНИИ
MYOTIS NATTERERI В УСЛОВИЯХ ЕВРОПЕЙСКОЙ ЧАСТИ
РОССИИ»**

Студент

Цыбулькин Дмитрий Викторович

(подпись, дата)

Руководитель

Смирнов Д.Г.

(подпись, дата)

Нормоконтролёр

Быстракова Н.В.

(подпись, дата)

Работа допущена к защите (протокол заседания кафедры от _____ № ____)

Заведующий кафедрой

Титов С.В.

(подпись)

Работа защищена с отметкой _____ (протокол заседания ГЭК от _____ № ____)

Секретарь ГЭК

Быстракова Н.В.

(подпись)

Пенза 2019

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ	3
ГЛАВА 1. АНАЛИЗ БИОРАЗНООБРАЗИЯ И ПОСТРОЕНИЕ МОДЕЛЕЙ РАСПРОСТРАНЕНИЯ ВИДОВ (ОБЗОР ЛИТЕРАТУРЫ)	5
ГЛАВА 2. МАТЕРИАЛ И МЕТОДЫ.....	8
2.1. Объект исследования.....	8
2.2. Исходные данные.	8
2.3. Используемые инструменты.....	12
2.4. Принцип работы алгоритма в программе MaxEnt.	12
2.5. Методика построения модели.....	13
2.6. Методика определения лимитирующих факторов.	14
ГЛАВА 3. ФИЗИКО-ГЕОГРАФИЧЕСКОЕ ОПИСАНИЕ ТЕРРИТОРИИ ИССЛЕДОВАНИЯ.....	18
ГЛАВА 4. РЕЗУЛЬТАТЫ	20
4.1. Распространение ночницы Наттерера.	20
4.2. Определение карты экологических ниш ночницы Наттерера.	20
4.3. Определение главных лимитирующих факторов.	24
ГЛАВА 5. ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ	34
ВЫВОДЫ.....	36
БЛАГОДАРНОСТИ	37
СПИСОК ЛИТЕРАТУРЫ	38

ВВЕДЕНИЕ

Сегодня, в биологии существует множество задач, связанных с изучением мест обитания живых организмов, а также лимитирующих факторов, которые влияют на их жизнь. В процессе таких исследований, учёные, как правило, ведут журналы записей находок, которые содержат географические координаты точек встречи видов. При визуализации таких данных на карте, у исследователя могут возникнуть вопросы о причинах такого распространения, о том, где могут быть наиболее похожие и следовательно благоприятные условия для существования организма в настоящем или будущем времени.

Для ответа на такие вопросы, в наши дни актуально использовать один из методов, находящийся на стыке таких наук, как: биология, география и анализ данных — метод максимальной энтропии, реализованный в прикладной программе MaxEnt (Maximum Entropy Species Distribution Modelling) (URL: <https://github.com/mrmaxent/Maxent>). Явными достоинствами этого метода можно считать: возможность использования метода для любого вида живого существа, независимо от его таксономической группы; наличие понятного пользовательского интерфейса, что снижает порог входа для исследователей в области биологии, которые не используют языки программирования в работе для моделирования с нуля; возможность исследовать несколько видов, с целью найти закономерности в их жизни на одной территории или определить для каждого из них свои ключевые лимитирующие факторы. Помимо этого, при наличии готовой, хорошо оптимизированной модели и получении новых данных мест находок, возможно делать новые прогнозы. Данные, получаемые в выводе программы, можно в дальнейшем использовать для оценки модели, проанализировать модель на статистическую значимость, используя различные инструменты анализа или визуализировать полученную информацию разными способами.

Целями этой работы являются: моделирование карты пригодных мест обитания ночницы Наттерера (*Myotis nattereri*) на территории Европейской части России и определение экологических факторов, влияющих на распространение.

Задачи:

1. Создать базу данных мест находок ночницы Наттерера на территории России.
2. Определить экологические факторы для использования в процессе моделирования пригодных мест обитания вида.
3. Построить модели пригодных мест обитания вида в пределах Европейской России.
4. Определение ключевых экологических факторов, определяющих распространение и оптимальные условия обитания вида в пределах исследуемой территории.

Работа выполнена на кафедре «Зоология и экология» Пензенского государственного университета.

ГЛАВА 1. АНАЛИЗ БИОРАЗНООБРАЗИЯ И ПОСТРОЕНИЕ МОДЕЛЕЙ РАСПРОСТРАНЕНИЯ ВИДОВ (ОБЗОР ЛИТЕРАТУРЫ)

В настоящее время, актуально использовать различные методики с заложенным в них математическим аппаратом во множестве научных областей, с целью подтвердить существующие гипотезы, а также, обнаружить новые факты, которые могут быть значимы при текущем положении дел. Для изучения ареалов видов, на текущий период времени существует не один подобный метод для моделирования карты оптимальных условий вида. Так, в работе Пузаченко (Пузаченко и др., 2010), автор рассматривает ряд программ, таких как BioGeomancer или DesktopGARP и других, используемых для анализа биоразнообразия и построения моделей распространения видов. В настоящее время, в практике исследования размещения видов используется огромный арсенал статистических методов: методы группировки на основе дистанции, метод максимальной энтропии, многомерная и логистическая регрессии, факторный и дискриминантный анализ, искусственные нейронные сети и генетические алгоритмы. У каждого из них есть достоинства и недостатки.

В этой работе используется метод максимальной энтропии, потому что среди всех методов, он наиболее популярный в текущий промежуток времени и его цитируемость растёт с каждым годом. Причём сейчас, количество отечественных работ по данному методу составляет единицы. При использовании такого метода важно понимать, что при работе алгоритма, исследователь получает не реальную карту ареала вида, а карту вероятностей встречи вида на определённой территории, построенную на основе факторов среды. При этом условии, метод можно рассматривать как способ получить карту экологической ниши объекта исследования, по Хатчинсону (Hutchinson, 1957).

В результате появления метода максимальной энтропии и подобных ему, перед исследователями открывается ряд областей, где алгоритм может оказаться очень полезным, а именно: планирование и создание заповедных зон и особо охраняемых природных территорий; изучение экологии, эволюции видов; исследование инвазивных видов; эпидемиология и распространение заболеваний.

Метод характеризуется тем, что для прогноза необходимы только географические данные мест находок и растровые географические слои, содержащие в себе данные окружающей среды. Эти источники данных должны быть в одинаковых системах координат и иметь одинаковые размеры ячеек и охваты. Таким образом, модель работает с данными, описывающими только наличие вида на определённой территории, без информации о том, где этот вид может быть ещё или его не может быть вовсе (Phillips и др., 2006).

Данные, представленные растрами, могут представлять разную информацию: антропогенные, биотические и абиотические факторы. При выборе переменных рекомендуется выбирать те, которые в большей степени, так или иначе, способны прямо или косвенно влиять на жизнь объекта исследования. Так, например, в работе (Рахматуллина и др., 2017) для исследования оптимальных мест роста и развития Сосны обыкновенной (*Pinus sylvestris*) использовались данные близости к автомобильным дорогам. Фактор сыграл высокую значимость и благоприятные условия были предсказаны в местах, достаточно удалённых от дорог, что характеризуется различными видами загрязнений, наравне с высокой газопоглолительной способностью сосны и негативно сказывается на её росте. То есть, авторам удалось, используя метод машинного обучения, точно выявить антропогенный фактор окружающей среды, который, по своей сути, является переходом от более благоприятных условий к менее благоприятным для развития организма.

В работе Сабденовой А.А. с соавторами (Сабденова и др., 2016), можно отметить другую важную особенность метода. Авторы использовали метод максимальной энтропии для моделирования экологических ниш Марала (*Cervus elaphus sibiricus*) – представителя благородных оленей Евразии на территории национального заповедника. Мотивация использования алгоритма заключалась в том, что полевые исследования распространения вида затруднены в связи с горной местностью заповедника и суровым климатом высокогорья. Тем не менее, используя алгоритм, исследователи смогли построить достаточно хорошую модель, которая показывала хороший результат прогноза.

На основе имеющейся информации, можно сделать небольшой вывод о том, что метод является актуальным в использовании, а правильно построенная модель, которая способна делать хороший прогноз, может быть использована с различными практическими целями.

ГЛАВА 2. МАТЕРИАЛ И МЕТОДЫ

2.1. Объект исследования.

Ночница Наттерера – летучая мышь отряда Рукокрылые (Chiroptera), семейства Гладконосые – (Vespertilionidae). Ареал обитания вида охватывает Европу от Великобритании до Урала. На севере граница ареала проходит от южной части Скандинавского полуострова до Италии, Балканского полуострова, Кавказа и Ближнего Востока. Основными местами обитания являются островная древесно-кустарниковая растительность, на опушках лесов рядом с тихими стоячими водоёмами или водотоками, а также каменных расщелинах и скалах в склонах по берегам рек с выраженными карстовыми формами рельефа, в деревянных элементах построек человека. Характерными кормовыми биотопами вида являются небольшие заливы средних и крупных рек, а также озёр, хорошо защищённые от ветра. Летают невысоко от поверхности воды. Пищу могут собирать с субстрата. В рацион питания входят различные насекомые: двукрылые, мелкие чешуекрылые, ручейники, жесткокрылые, мелкие бабочки, пауки. (Beck, 1995). Среди ночниц, обитающих на территории России, ночница Наттерера выделяется сильной чувствительностью к плохой погоде. При слабом морозящем дожде или при порывистом прохладном ветре, эта ночница прекращает охоту. Ведёт оседлый образ жизни, зимует в различных подземных убежищах естественного и искусственного происхождения. Полёт манёвренный, но медленный. (Стрелков, 1963; Ильин, Смирнов, 2000; Демянчик, 2001).

2.2. Исходные данные.

Для проведения исследования были взяты данные мест находок ночниц Наттерера на территории России. Эти сведения (координаты) рукокрылых, полученные из кадастровой базы данных кафедры зоологии и экологии ПГУ и литературных источников (Ильин и др., 2001 г.). Количество записей

составило 154 единицы. Эти данные были собраны приблизительно за период времени с 1890-х по 2010-е годы.

В дополнение к данным о координатах ночницы Наттерера, с портала «Млекопитающие России» (URL: <http://rusmam.ru/>) были взяты места всех находок ночниц по территории России, их количество равно 1531 единице. Все эти данные представляют собой файлы, формата CSV и содержат таблицу, куда входят: названия видов, широта и долгота локаций (рис. 1).

	A	B	C
1	SPECIES	LATITUDE	LONGITUDE
2	M.nattereri	56.1666	51.1333
3	M.nattereri	57.2441	50.9596
4	M.nattereri	58.5666	49.6500
5	M.nattereri	55.604	48.950
6	M.nattereri	55.290	49.287
7	M.nattereri	55.272	49.269
8	M.nattereri	55.364	48.250
9	M.nattereri	55.241	49.217
10	M.nattereri	54.052	41.933

Рис. 1. Пример CSV файла с таблицей мест находок ночницы Наттерера.

Также, для исследования были взяты географические растровые слои. В эти данные входят биоклиматические переменные, которые можно свободно получить с открытого ресурса «WorldClim» (URL: <http://www.worldclim.org/bioclim>). Состав этих переменных представлен в таблице 1.

Таблица 1. Биоклиматические переменные.

Название переменной	Описание
bio_1	Средняя температура за год
bio_2	Средняя амплитуда температуры за сутки (разность максимальной и минимальной температуры)
bio_3	Изотермичность (частное между bio_2 и bio_7, умноженное на 100)
bio_4	Температура за сезон (произведение стандартного отклонения на 100)

bio_5	Максимальная температура наиболее жаркого месяца в году
bio_6	Минимальная температура наиболее холодного месяца в году
bio_7	Средняя амплитуда температуры за год (разность bio_5 и bio_6)
bio_8	Средняя температура наиболее влажного квартала в году
bio_9	Средняя температура наиболее сухого квартала в году
bio_10	Средняя температура наиболее жаркого квартала в году
bio_11	Средняя температура наиболее холодного квартала в году
bio_12	Общее количество осадков за год
bio_13	Число осадков за наиболее влажный месяц в году
bio_14	Число осадков за наиболее сухой месяц в году
bio_15	Число осадков за сезон в году в году
bio_16	Число осадков наиболее влажного квартала в году
bio_17	Число осадков наиболее сухого квартала в году
bio_18	Число осадков наиболее жаркого квартала в году
bio_19	Число осадков наиболее холодного квартала в году

Помимо биоклиматических переменных, для модели были взяты космические снимки Земли, сделанные при помощи спутника Terra, использующего сканирующую систему MODIS (Moderate Resolution Imaging Spectroradiometer) (URL: <https://modis.gsfc.nasa.gov/data/>). Эти снимки содержат в себе данные отражательной способности поверхности в различных спектрах. Снимки сделаны в 2004 году за период с января до октября, не считая, ноябрь и декабрь. Они описывают обобщённые среднемесячные данные. Для каждого месяца использованы 7 наборов переменных спектральных яркостей, которые соответствуют каналам сканирующей системы спутника. В таблице 2 отображены взаимосвязи

каждого канала и фактора окружающей среды (Книжников и др., 2004; Лисовский, Оболенская, 2014).

Таблица 2. Взаимосвязь спектральных каналов сканирующей системы и природных факторов.

Номер канала	Фактор, с которым есть связь
1	Фотосинтез, активность поглощения хлорофилла
2	Отражательная способность органелл растительной клетки
3	Типы лесов и снеговые условия
4	Состояние растительности и снеговые условия
5	Наличие влаги в растениях
6	Наличие влаги в растениях
7	Наличие влаги в почве и экосистеме

Дополнительно, для прогноза были взяты такие переменные, как высота над уровнем моря (переменная – alt), и крутизна склонов (переменная – krutizna), которые характеризуют рельеф исследуемой местности. Такие переменные могут быть синтезированы с использованием QGIS (URL: <http://www.qgis.org/ru/site/>). Общее количество признаков составило 84 единицы.

Исходные растровые данные представляют собой матрицы, где каждая ячейка содержит конкретную информацию о факторе среды. Например, для температуры – это цифра, обозначающая температуру в градусах; для высоты – это значение, определяющее высоту в метрах или километрах в конкретной местности. Каждый растр обладает следующими свойствами: разрешение, охват, размер ячеек, ширина и высота матрицы, значение NODATA – число, в ячейках, обозначающее отсутствие данных. В работе каждый растр обладал следующими свойствами: шириной равной 5298 колонок, высотой равной 1846 строк, границей по оси абсцисс равной 2594130,3985, границей по оси ординат равной 4584333,6735, размером ячейки 2000 в метрах, значение отсутствия данных -9999. Все переменные принадлежат к количественным непрерывным переменным.

2.3. Используемые инструменты.

Для общего обзора и визуализации географических данных, в работе использовалась программа QGIS (URL: <http://www.qgis.org/ru/site/>), а также, встроенный в неё модуль GRASS (Geographic Resources Analysis Support System) (URL: <https://grass.osgeo.org/>). Для преобразований содержимого растров, использовался язык программирования Python 3.6.5 (URL: <https://www.python.org/>). Это один из наиболее популярных инструментов для написания алгоритмов и анализа данных, его используют в биоинформатике, системной биологии, аналитической биологии и множестве других дисциплин. Интегрированная среда разработки, в которой он применялся – Jupyter Notebook (URL: <https://jupyter.org/>). Для открытия и сохранения растров использовалась библиотека rasterio (URL: <https://pypi.org/project/rasterio/>), для работы с массивами и для некоторых расчетов – библиотека numpy (URL: <https://www.numpy.org/>); библиотека pandas (URL: <https://pandas.pydata.org/index.html>) – для работы с табличными данными и библиотека машинного обучения – SciKit-learn (URL: <https://scikit-learn.org/stable/>) для использования метода главных компонент, инструмент gdal (URL: <https://pypi.org/project/GDAL/>), содержащий методы из одноимённой GIS системы, модуль affine (URL: <https://pypi.org/project/affine/>) – для преобразования номера ячейки матрицы в географическую координату и наоборот, а также, библиотека pyproj (URL: <http://pyproj4.github.io/pyproj/v2.2.0rel/>) для трансформирования географических точек из одной системы отсчёта в другую. Непосредственно, для создания и обучения модели, использовалась прикладная программа MaxEnt (URL: <https://github.com/mrmaxent/Maxent>).

2.4. Принцип работы алгоритма в программе MaxEnt.

Главная задача алгоритма заключается в том, чтобы сделать карту, с распределёнными на ней значениями вероятности, таким образом, что в местах, где вид присутствует, вероятность будет равна 1, а в других ячейках

необходимо определить эту вероятность. Изначально, алгоритм «не знает» эту вероятность, но «знает» окружающие условия для расчёта этой вероятности и условия, при которых вероятность равна 1. По умолчанию, алгоритм не может поставить в исследуемых ячейках вероятность равную 0, потому что это обусловлено отсутствием информации о том, были ли записи находок в этой местности или нет. Тогда, согласно принципу максимальной энтропии, о том, что для выбранной территории и факторов среды, присущих ей, наиболее характерное распределение вероятности будет при максимальной энтропии. Вследствие этого, программа присваивает всем исследуемым ячейкам, значение вероятности 0,5. Далее, на основе того, где было найдено существо, для факторов окружающей среды вычисляются коэффициенты значимости. Полученные переменные используются в процессе обучения модели для корректировки исходного значения вероятности (Phillips и др., 2006).

2.5. Методика построения модели.

Для построения модели был создан файл bias или файл смещения. Он содержит в себе данные мест находок всех ночниц на территории России. Для его создания, использовался один географический слой, по свойствам соответствующий исследуемой области. Алгоритм заключался в том, чтобы каждой исследуемой ячейке была присвоена единица, а в точках, где были места находок и вокруг них, на территории, представляющей собой буферную зону радиусом в 5 километров, было присвоено значение 10. Этот файл необходим для того, чтобы задать алгоритму исследованную территорию с характерным биотопом. Таким ячейкам программа будет отдавать предпочтение с условием соответствия факторов в процессе обучения алгоритма. В итоге, влияние неравномерного отбора данных на определённых территориях будет сведено к минимуму (Phillips, Dudík, 2008; Дудов, 2016).

Итого, все входные параметры, которые были указаны в программе MaxEnt:

- Непосредственно, точки находок ночницы Наттерера и исходные растровые слои, описывающие факторы окружающей среды;
- Указан «bias» файл;
- Тип выходящего формата для карты оптимальных условий – raw (исходный);
- Множитель L1-регуляризации равный 1, по умолчанию;
- Количество фоновых точек по умолчанию равное 10000, но так как количество ячеек в исходной матрице значительно больше, параметр выбирался случайно;
- Количество репликаций равное 5, с используемым типом – кроссвалидация или скользящий контроль, при котором используется метод k-folding для разбиения выборки на k-кластеров (Воронцов, 2004);
- Количество итераций обучения алгоритма и достижения оптимальной модели равно 500, по умолчанию.

2.6. Методика определения лимитирующих факторов.

Для решения задачи определения ключевых факторов, определяющих оптимальные условия для обитания вида необходимы географические слои данных, между которыми корреляция минимальна. Исходные слои, использованные в первой задаче, достаточно сильно коррелируют между собой. К тому же, они обладают конкретным недостатком – в различных местах на карте присутствуют пропуски.

В этой ситуации нельзя заполнить пропуски средними значениями, рассчитанными на основе ближайших ячеек, потому что сами растры имеют не одинаковые границы, где заканчиваются данные среды и начинается константа NODATA. Для решения этой проблемы, было решено, чтобы

количество пропусков для всех слоёв было одинаково. Для этого был реализован наиболее оптимальный алгоритм, который использовал один исходный слой, куда заносились пропуски из всех слоёв, затем расположения этих пропусков применялись для всех оставшихся слоев. Предварительно, каждая матрица значений растров была преобразована в одномерный массив. В результате при удалении всех значений пропусков из массивов, их длина стала одинаковой.

Так как задача заключается в прогнозировании таких переменных, которые вносят наибольший вклад, то эти переменные должны быть независимы друг от друга, то есть, не должны быть сильно коррелированными. Подойти к решению этой задачи можно различными способами, один из них – удаление некоторых сильно коррелированных переменных из набора. Однако, достаточно сложно точно понять, какие именно переменные следует удалять, кроме того, удалённые факторы могут являться одними из ключевых лимитирующих факторов по результатам моделирования. Поэтому, для решения этой проблемы использовался метод главных компонент (Smith, 2002).

Метод главных компонент – довольно популярный метод в статистике, анализе данных, биоинформатике и других науках. Предназначен для редукции размерности и снижения корреляции исходной выборки переменных. Суть метода заключается в аппроксимации исходного набора переменных до такой формы в пространстве, что собственные вектора или айген-вектора этой формы будут являться, главными компонентами, описывающими исходные данные. При проекции на такие вектора сохраняется наибольшее количество информации от исходных данных. (Tipping, Bishop, 1999).

В результате получились новые массивы, описывающие растровые данные, практически, не коррелирующие между собой. Причём количество компонент, которые было необходимо получить, равно количеству исходных

переменных, потому что в этой задаче целью является не редукция признаков, а избавление от зависимостей в данных.

У полученных главных компонент можно отметить следующие атрибуты:

- Общая процентная сумма того, на сколько полученные компоненты объясняют исходные переменные равна 100 процентам;
- Расчетная ковариация шума по вероятностному распределению модели главных компонент равна 0 (Bishop, 2006).

Получившиеся массивы были преобразованы к исходному виду. После этого, оптимальным вариантом для возвращения к исходному размеру и заполнения массивов пропусками, является обращение к исходным массивами и изменение тех ячеек, которые не содержат пропуски, а именно замена их значения на значение полученных компонент. В ином случае, при попытке заполнения готовых массивов пропусками, процесс является очень сложным с точки зрения структур данных и занимает большее количество времени. В конечном итоге получаются матрицы, исходного размера и с одинаковыми атрибутами, которые можно использовать в программе MaxEnt.

Для построения модели, предсказывающей лимитирующие факторы среды, использовались точки находок ночницы Наттерера, вместо исходных слоев, использовались, полученные ранее, главные компоненты. Все остальные параметры были такими же, как и при определении карты экологической ниши.

После завершения обучения алгоритма оценка его работы проводилась стандартными средствами программы MaxEnt, а именно:

- График оmissии — это зависимость дробного значения от кумулятивного порога для двух случаев: для кривой, показывающей долю предсказанной территории, а также для кривых пропусков, то

есть, неверно предсказанных значений на тестовых и тренировочных данных в процессе прогноза.

- ROC-кривая и AUC – типичный пример статистического метода, используемого для оценки работы бинарного классификатора. ROC-кривая – это график отражающий соотношение между объектами тестовой выборки, которые в процессе обучения были верно классифицированы, в случае MaxEnt – такие ячейки раstra или площади поверхности Земли, в которых был была находка и алгоритм предсказал его наличие, а также ошибочно классифицированные места, где находка была, но программа предсказала его отсутствие. AUC – это площадь под ROC-кривой. Она отражает количественное значение ROC-кривой. Как правило, чем больше значение этой площади, тем лучше работает алгоритм и тем выше его прогнозная способность (Davis, Goadrich, 2006);
- Jackknife тест – это статистический метод передискретизации, полезный в задачах оценки смещения и дисперсии выборки. Является линейной аппроксимацией метода bootstrap (Cameron, Trivedi, 2005). В выводе MaxEnt, jackknife-тест представляет из себя линейные диаграммы для каждого признака и оценка его вклада в модель в двух случаях: при отсутствии этого признака в процессе обучения и при обучении, используя только этот признак.
- Response Curves – графики кривых ответов каждой переменной. Отражают зависимость степени оптимальных условий в прогнозе модели от диапазона значений каждого фактора среды;

ГЛАВА 3. ФИЗИКО-ГЕОГРАФИЧЕСКОЕ ОПИСАНИЕ ТЕРРИТОРИИ ИССЛЕДОВАНИЯ

В качестве модельной территории, где рассматривается распространение ночницы Наттереры, нами взята Европейская часть России. Она располагается на Восточно-Европейской равнине. Её географическими границами являются: с севера – Ледовитый океан; с востока – Уральские горы; с юга – Каспийское море, Чёрное море, Кавказские горы; с запада – Балтийское море.

Большая часть территории Восточно-Европейской равнины имеет равнинный рельеф. На северо-западе платформы располагается Балтийский щит, здесь она приподнята. Балтийский щит связан с возвышенными равнинами Кольского полуострова, Карелии и Хибинами. Приподнятый фундамент является основой Высокого Заволжья и Среднерусской возвышенности. Формирование Приволжской возвышенности зависит от тектонических поднятий отдельных частей платформы. Северные низменности: Прикаспийская и Печорская появились в результате того, что равнины множество раз подвергались приходу моря.

В северной части равнины главными возвышенностями являются – Валдайская и Смоленско-Московская. Они образовались из-за накопления моренных отложений.

Средняя температура на территории Русской равнины: в январе от 4°, 5°C (например, на Черноморском побережье) до -40°, -48°C (на востоке Республики Саха, где абсолютный минимум температуры составляет -66,8°); в июле температура находится в диапазоне от 0°, 1°C (на островах Карского моря) до 24-26°C (Республика Калмыкия, где абсолютный максимум достигает 45,4°C). (Россия в цифрах, 2017)

На большей части территории умеренный климат. С запада на восток растёт континентальность. Отлично выражена географическая зональность. У побережья Баренцева моря преобладает мохово-лишайная тундра, которая к югу сменяется на обширную лесную зону, простирающуюся от Прибалтики до Урала. От Карпат до Урала тянется зона лесостепи, которая затем переходит в степную зону. Прикаспийская низменность занята полупустынями и участками песчаных пустынь. (Горкин, 2006).

ГЛАВА 4. РЕЗУЛЬТАТЫ

4.1. Распространение ночницы Наттерера.

В процессе сбора данных о координатах мест находок летучей мыши на территории России, была создана база данных, содержащая эту информацию. На рис. 2 визуализировано содержимое этой базы находок (координаты отмечены красными точками). Как видно из иллюстрации, вид расселён по территории Европейской части очагово.

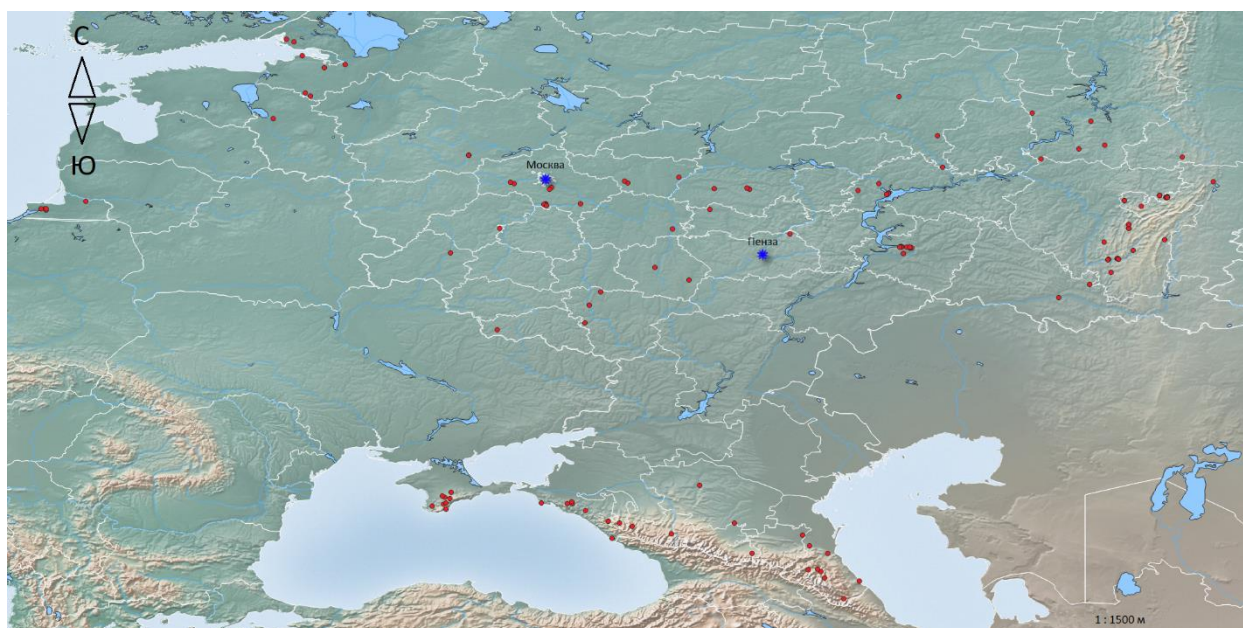


Рис. 2. Точки находок ночницы Наттерера на карте.

4.2. Определение карты экологических ниш ночницы Наттерера.

В результате работы алгоритма было получено 5 репликаций или вариантов обучения модели. Репликации были необходимы для того, чтобы понять, насколько стабильна модель. По полученным данным, можно заключить, что она действительно стабильная. Так, на графике оmissии (рис. 3) можно увидеть, что с повышением кумулятивного порога, количество предсказанной территории снижается. А также, при повышении кумулятивного порога, растёт количество пропусков. В целом, среднее число пропусков по тестовым данным почти совпадает с предсказанным числом пропусков, по определению кумулятивного порога.

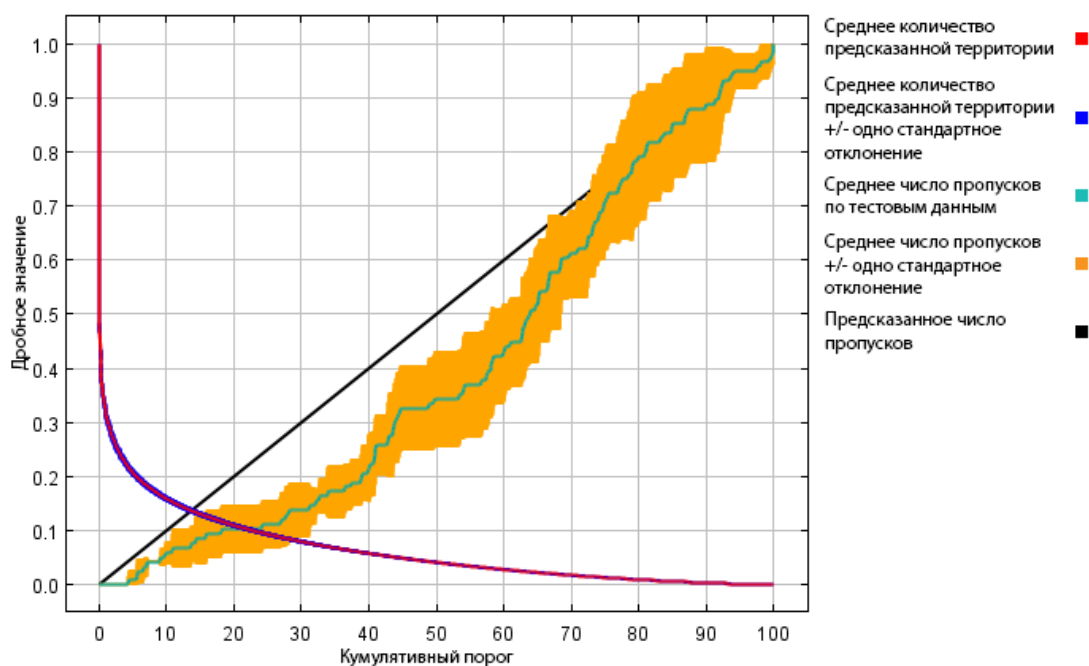


Рис. 3. График средней оmissии и предсказанной территории для ночницы Наттерера.

При изучении графика, на котором изображена общая ROC-кривая для всех репликаций (рис. 4), можно обнаружить что модель обладает отличной прогностической способностью. Так, средняя площадь под ROC-кривой равна 0.958. Поэтому, мы можем говорить о том, что в 95.8 процентах случаев, модель верно предскажет наличие вида в конкретной точке. При этом, стандартное отклонение AUC равно 0.006.

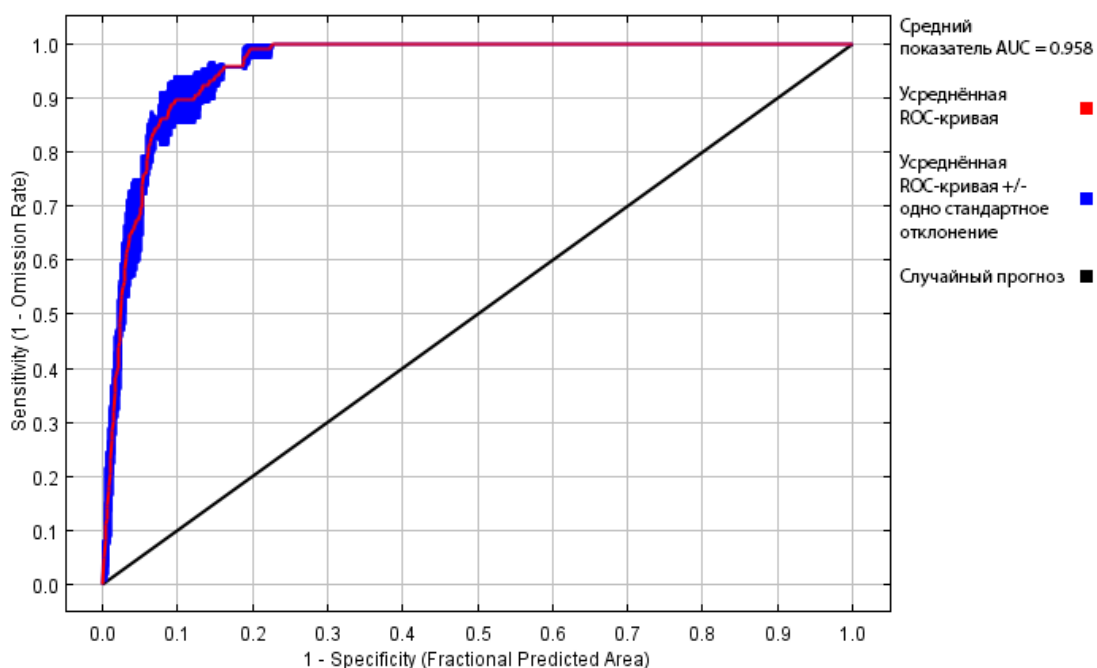


Рис. 4. График ROC-кривой.

Также, можно увидеть само изображение с прогнозом оптимальных территорий для летучей мыши (рис. 5).

Таким образом, по информации, заложенной в карте, можно отметить то, что на основе заложенных в модель факторов, ночница Наттерера способна жить на достаточно большой территории Европейской части России. Тем не менее, она является стенобионтом, по отношению к некоторым факторам. Земными границами у данного распределения вероятностей являются: на юге – Закавказье и Прикаспийская низменность, на востоке – Зауралье, на севере – Печёрская низменность и Кольский полуостров.

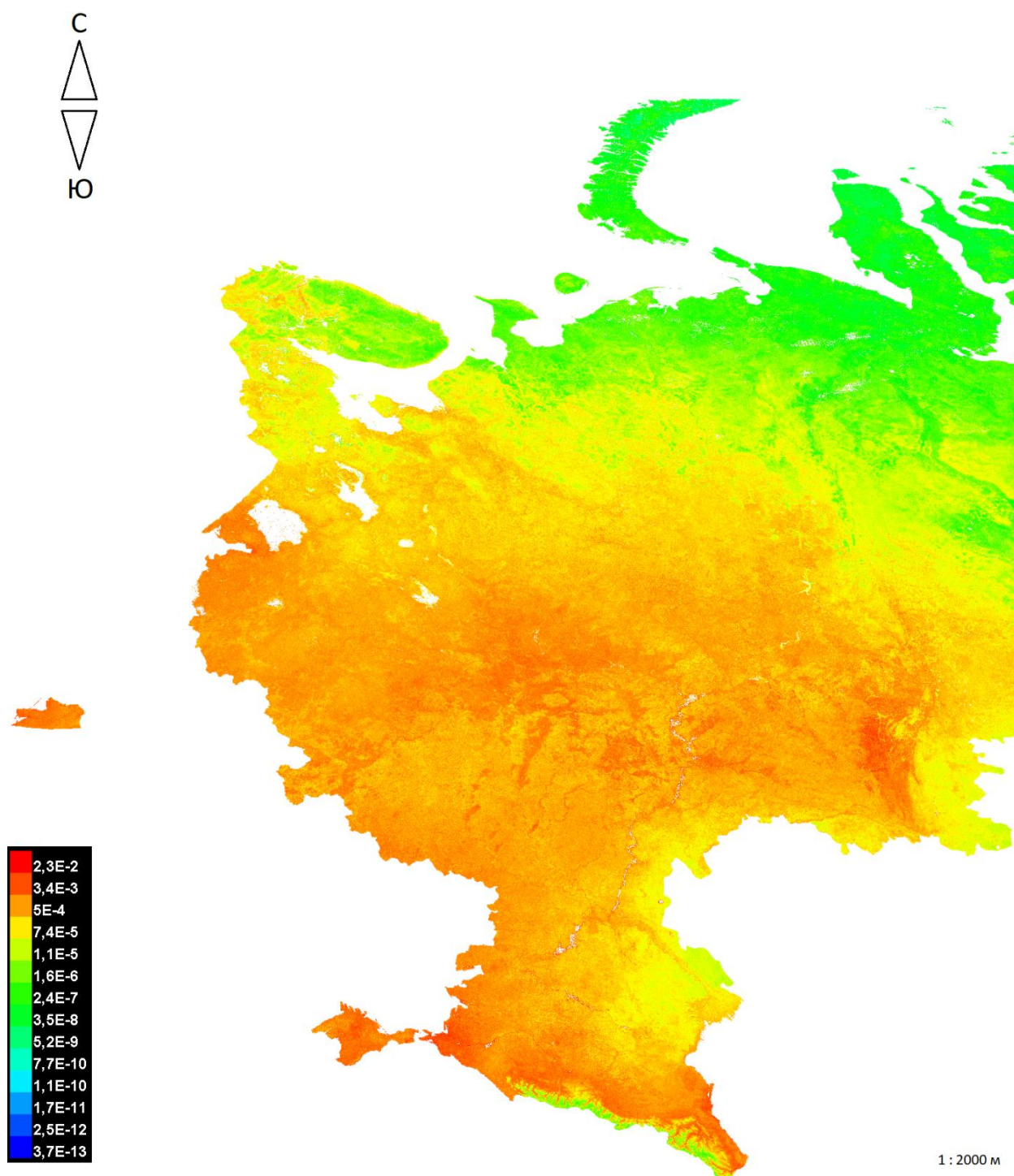


Рис. 5. Карта благоприятных мест обитания для ночницы Наттерера на территории Европейской части России. Градация перехода цвета от синего к красному означает переход от менее благоприятных условий к наиболее благоприятным соответственно.

4.3. Определение главных лимитирующих факторов.

Для оценки модели, прогнозирующей ключевые лимитирующие факторы, также, как и в методике построения карты оптимальных условий для вида, были использованы: график оmissии (рис. 6) и ROC-кривая (рис. 7).

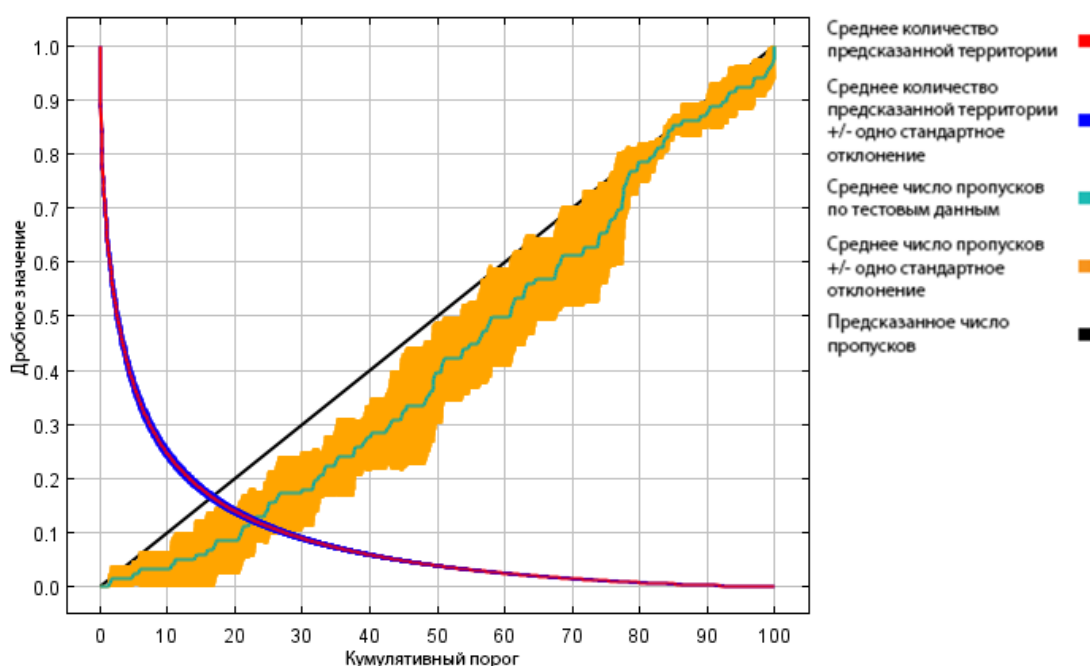


Рис. 6. График средней оmissии и предсказанной территории для ночницы Наттерера.

На графике оmissии видно, что кривая пропусков, полученных в процессе обучения модели, находится достаточно близко к предсказанной кривой пропусков.

Оценка точности модели, рассчитанная с использованием ROC-кривой, имеет достаточно высокое значение. Средняя площадь под ROC-кривой равна 0.942 со средним стандартным отклонением равным 0.016.

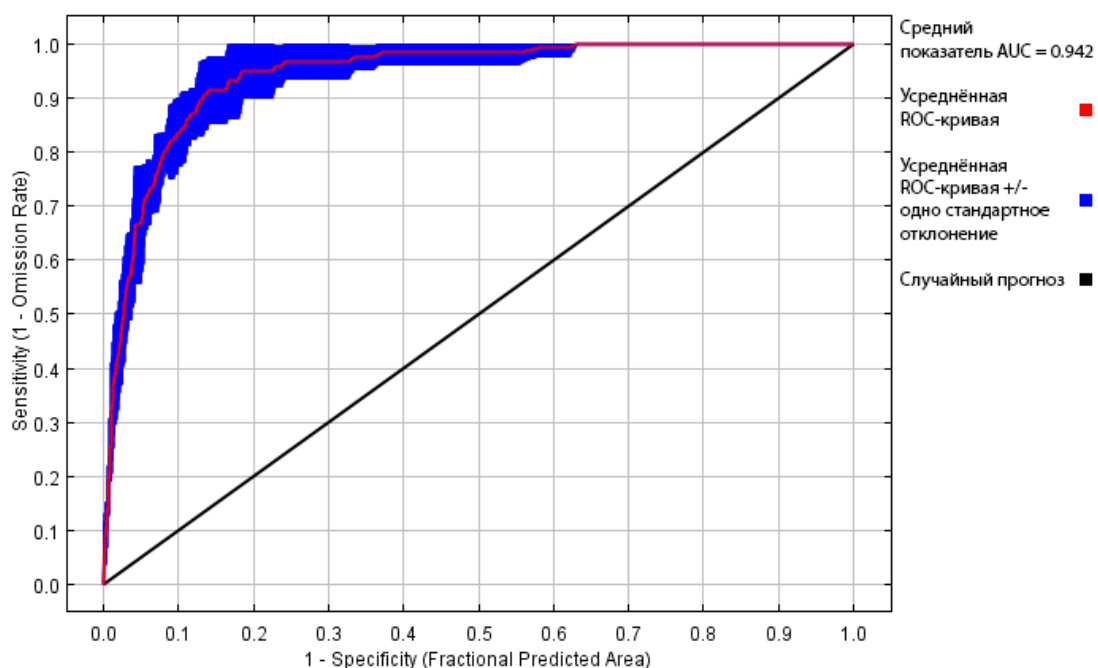


Рис. 7. График ROC-кривой.

Для определения ключевых лимитирующих факторов для ночницы Наттерера используется таблица вклада переменных, которая была взята из вывода программы MaxEnt (табл. 3). В первой колонке находятся наиболее важные переменные, во второй – процент их вклада в прогноз. Среди 84 компонент были отобраны только те, процент вклада которых больше единицы.

Таблица 3. Проценты вклада каждой переменной.

Переменная	Процент вклада, %
componenta_2	30.1
componenta_63	13.3
componenta_0	11.2
componenta_4	7.2
componenta_1	6.9
componenta_70	3.3

componenta_75	3.2
componenta_15	2
componenta_3	2
componenta_82	1.6
componenta_10	1.3
componenta_9	1.2
componenta_5	1.1
componenta_44	1.1

Несмотря на то, что полученные компоненты представляют собой абстрактные величины, каждая компонента в таблице вклада переменных взаимосвязана с исходными данными. В таблице 4 представлено соответствие каждой компоненты, которая сделала наибольший вклад в модель и исходной переменной (табл. 1, 2). Эти взаимосвязи были определены по наибольшему коэффициенту влияния и взяты из атрибута components_ из результатов работы метода главных компонент. Порядок взаимосвязей в таблице соответствует их процентному вкладу в модель в таблице 3.

Таблица 4. Взаимосвязь главных компонент и исходных переменных.

Исходная переменная	Главная компонента
bio_4	componenta_2
bio_9	componenta_63
krutizna	componenta_0
bio_3	componenta_4
bio_3, bio_16, 05.band3, 05.band4	componenta_1
bio_7, bio_2	componenta_70
bio_15, bio_1	componenta_75
02.band2, 02.band6, 02.band7	componenta_15
bio_4	componenta_3
bio_3	componenta_82
10.band5, 10.band3, 10.band2	componenta_10
04.band6, 03.band6	componenta_9
03.band2, 03.band1	componenta_5
04.band1, 04.band3	componenta_44

Как видно из таблицы, можно заключить, что на первом месте стоит фактор, процент вклада которого равен 30.1 %. Он обозначает среднюю температуру за сезон. Вторая переменная отражает среднюю температуру наиболее сухого квартала в году с процентным вкладом равным 13.3 %. На третьем месте переменная, отражающая крутизну склонов со значимостью в 11.2 %. После неё следует изотермичность с вкладом в 7.2 %. Затем идёт компонента, которая взаимосвязана с несколькими факторами, а именно: изотермичность, число осадков наиболее влажного квартала в году, 3 и 4 спектральные каналы за май 2004 года, вклад которых 6.9 %. Далее, на 6 и 7 месте следуют средняя амплитуда температуры за год, средняя амплитуда

температуры за сутки, а также, число осадков за сезон и средняя температура за год, по 3.3 % и 3.2 % соответственно. На 8 месте стоит компонента, связанная со 2, 6 и 7 спектрами за февраль с процентом вклада 2. Далее, на 9 и 10 местах следуют компоненты, уже интерпретированные предыдущими компонентами, а именно: средняя температура за сезон и изотермичность, с вкладами в 2 % и 1.6 % соответственно. В конце остаются переменные, вклад которых колеблется около 1 %. Это различные спектры света за февраль, апрель и октябрь.

Для проверки значимости каждой из переменных, использовался jackknife-тест. Он проводился для тренировочных данных (рис. 8), тестовых (рис. 9) и для тестовых данных, после оценки прогноза с помощью ROC-кривой (рис. 10).

На первом графике (рис. 8) можно заметить, что наиболее влиятельным фактором, вложившимся в модель, является `componenta_2`, то есть, средняя температура за сезон. Важно то, что при упущении этой переменной из модели, снижается и регуляризованный прирост. А также, при обучении модели, используя только эту переменную, прирост максимален среди всех переменных. Это можно интерпретировать как то, что эта переменная, по сравнению с другими содержит наибольшее количество полезной информации для обучения алгоритма и, при этом, содержит больше информации, которой нет в других аргументах. На втором месте по содержанию важной информации, является `componenta_63` – средняя температура самого сухого квартала. После этого, можно выделить такие, наиболее важные переменные, как: `componenta_4`, `componenta_70`, `componenta_5`, `componenta_0`, `componenta_1`, `componenta_72`. Последняя компонента, не присутствующая в таблице значимости, взаимосвязана с `bio_19`, число осадков самого холодного квартала, имеет вклад в менее чем 1 %. При отсутствии этих переменных в процессе обучения, их снижение

прироста менее значительно, чем с предыдущими факторами, тем не менее, они обладают достаточно полезной информацией, влияющей на обучение.

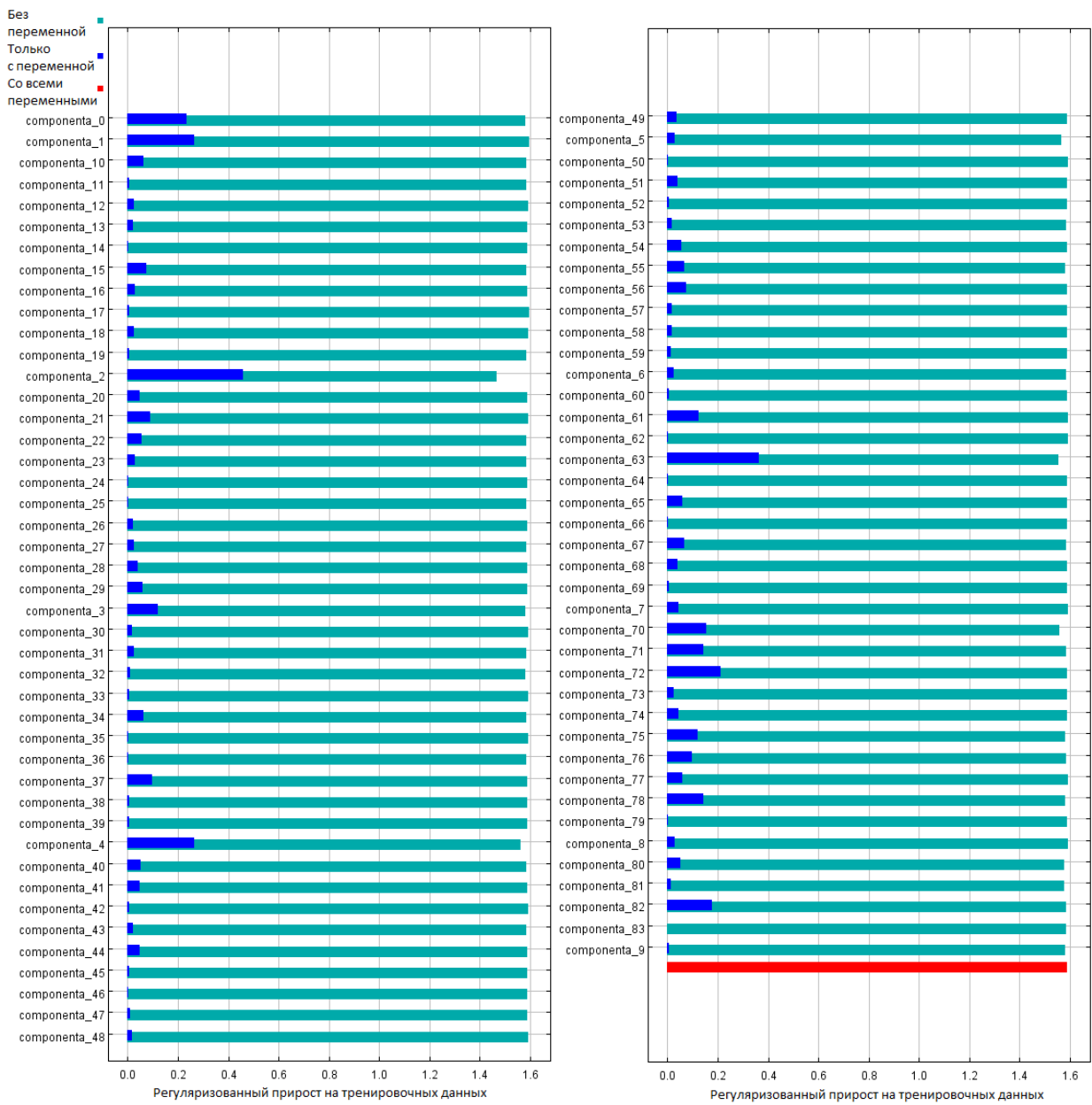


Рис. 8. Результат jackknife-теста для главных компонент на тренировочном наборе данных.

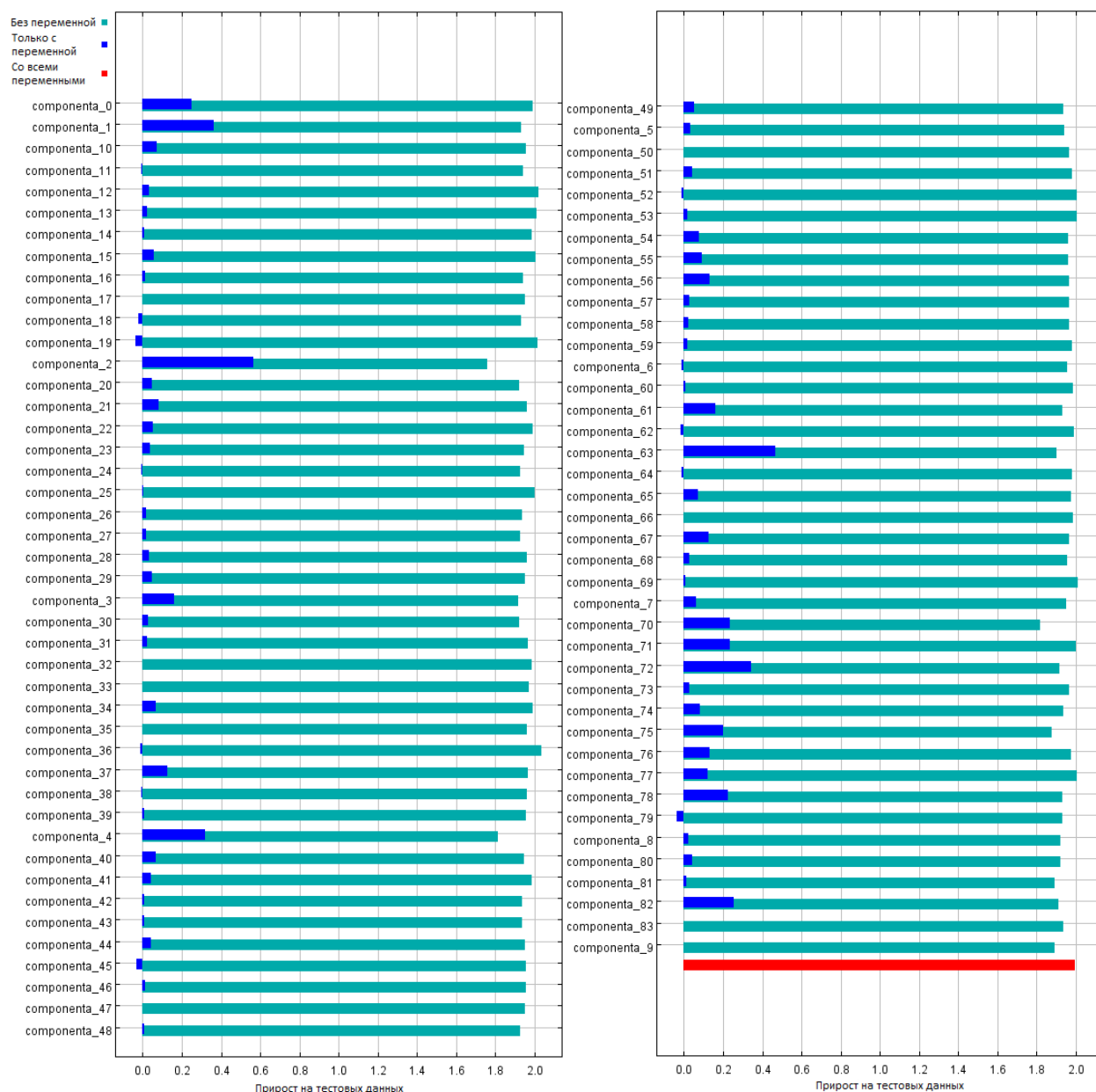


Рис. 9. Результат jackknife-теста для главных компонент на тестовом наборе данных.

На втором графике (рис. 9) график отражает значимость переменных на тестовых данных. Ситуация здесь, похожа на результаты предыдущего jackknife-теста, однако, есть и различия. Основными переменными, содержащими наиболее важную информацию, также являются componenta_2 и componenta_63. За ними следуют: componenta_0, componenta_1, componenta_4, componenta_72, componenta_82, componenta_70 и componenta_71. По количеству информации же, значимыми являются (в порядке уменьшения влияния): componenta_2, componenta_63, componenta_1,

componenta_72, componenta_4, componenta_82, componenta_78, componenta_70 и componenta_71. Компонента 71, присутствующая в результате теста, имеет вклад менее 1 % и представляет собой среднюю температуру самого влажного квартала в году.

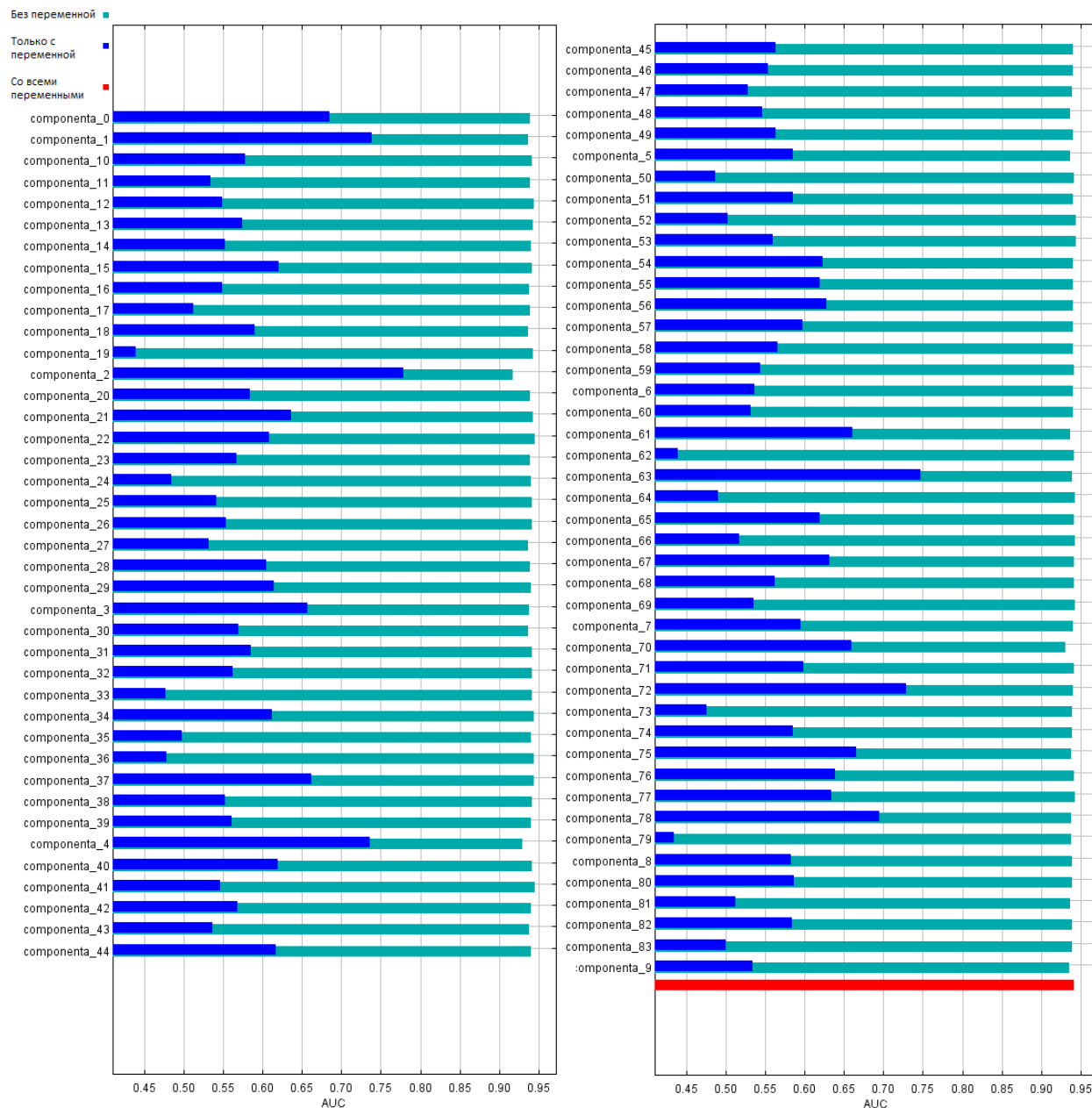


Рис. 10. Результат jackknife-теста для главных компонент на тестовом наборе данных после оценки модели ROC-кривой.

Последний график с jackknife-тестом, проведённым для переменных, после оценки модели площадью под ROC-кривой отражает итоговый результат важности каждой переменной. Здесь можно отметить, что факторы: componenta_2, componenta_63, componenta_1, componenta_4

представляют наиболее значимую информацию. В результате использования только этих переменных поодиночке, AUC находится в пределах значения 0,75. То есть, только при использовании этих переменных, алгоритм с точностью в 75 % будет делать верные прогнозы. По количеству информации, входящей в одни факторы, которой нет в других, можно выделить следующие переменные: componenta_2, componenta_4, componenta_70.

В целом, результат последнего jackknife-теста практически полностью подтверждает результаты двух предыдущих тестов.

Последним методом, определяющим влияние факторов среды на оптимальные условия, являются графики кривых ответов (Response curves) (рис. 11). Были взяты графики из модели, прогнозирующей карту экологических ниш вида. Отобраны только те переменные, вклад которых был наиболее значимым в задаче прогноза ключевых лимитирующих факторов, а кривые построены с учётом корреляции исходных данных.

На этих кривых по оси абсцисс изменяются значения лимитирующего фактора, а по оси ординат – пригодность условий. Кривые отмечены красным цветом, они усреднены по пяти репликациям. Синим оттенком обозначены кривые +/- одно стандартное отклонение.

В общем, кривые достаточно хорошо показывают границы для значений факторов, которые определяют пригодность мест обитания.

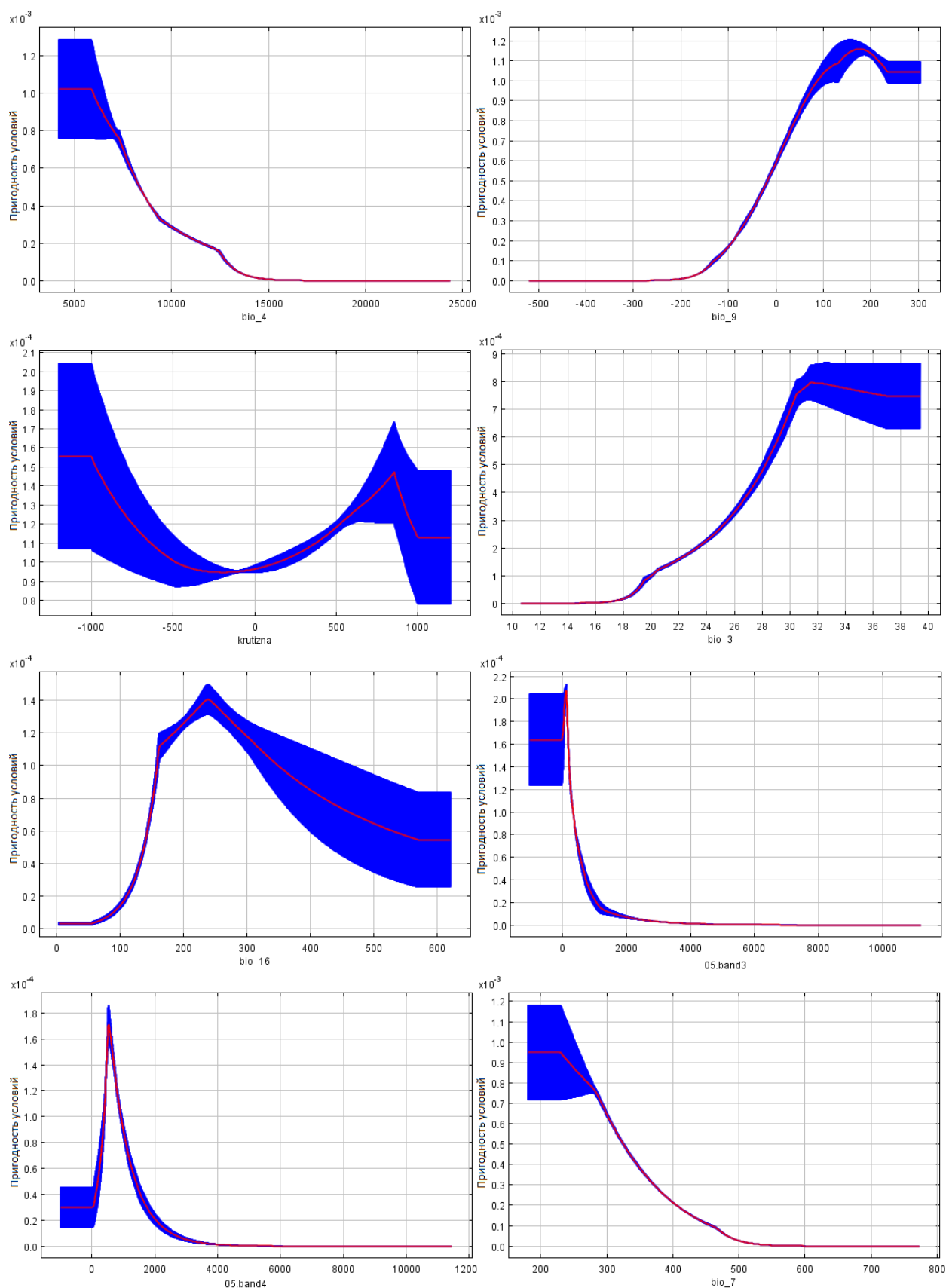


Рис. 11. Графики кривых ответов лимитирующих факторов.

ГЛАВА 5. ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

Подводя итог результатов работы, можно упомянуть, что алгоритм, заложенный в программе MaxEnt, практически оправдал свою значимость. Хорошо построенная модель делает достаточно правильные прогнозы, которые можно оценить, используя различные статистические метрики качества.

На основе полученной карты оптимальных условий ночницы Наттерера (рис. 5), можно сделать вывод, что летучая мышь способна жить на значительной площади Европейской части России. Снижение пригодности мест обитания на протяжённости с севера страны до Западно-Сибирской равнины можно объяснить переходом с юга от умеренной климатической зоны до севера к субарктическому поясу. Помимо этого, низкие показатели можно заметить на Прикаспийской низменности, что можно связать с переходом к резко-континентальному климату и отсутствием подходящих биотопов для летучей мыши, из преобладания пустынных местностей.

Данные, полученные в результате моделирования экологических ниш с целью найти самые важные лимитирующие факторы, в полной мере подтверждают статистическую значимость карты мест пригодности обитания. Так, можно отметить, что наиболее важным фактором для вида является температура, а именно, её колебания за длительные периоды времени в году: кварталы и сезоны. Что также подтверждается на графиках кривых ответов исходных температурных переменных (рис. 11). На них можно заметить, что величина пригодности территории резко снижается в определённых диапазонах значения фактора среды. Такая же ситуация и с факторами, определёнными количеством осадков и космическими снимками, взаимосвязанными с снеговыми условиями и растительностью. Важно отметить, что степень пригодности условий зависит и от крутизны склонов. Так, наиболее благоприятные условия определены в диапазонах, где

крутизна максимальна и наоборот – там, где поверхность земли наиболее ровная, пригодность принимает наименьшее значение.

Таким образом, модель является действительно правдивой и результат её работы полностью соответствует реальным наблюдениям, полученным в природе.

Помимо этого, работа делает дополнительный вклад в исследования этого вида летучих мышей на территории России.

ВЫВОДЫ

1. В начале работы были собраны все существующие, научно-подтверждённые данные мест находок ночницы Наттерера по Европейской части России;
2. В процессе работы были получены географические слои, отражающие фундаментальные явления природы, куда входит информация о таких факторах, как: температура, количество осадков, фотосинтетическую способность растений, снежный покров, количество влаги. Эти географические слои были использованы для моделирования карты благоприятных условий жизни ночницы Наттерера, а также для синтеза не коррелирующих главных компонент для анализа лимитирующих факторов.
3. В результате изучения метода максимальной энтропии, были построены наиболее оптимальные модели для достижения целей работы, оценка качества которых имеет очень высокое значение.
4. В конечном итоге, в результате работы были получены целевые сведения как о местах с самыми благоприятными факторами для жизни летучей мыши, так и главные факторы природы, влияющие на её существование.

БЛАГОДАРНОСТИ

Автор работы выражает благодарность коллективу кафедры «Зоология и экология» ПГУ, научному руководителю Д.Г. Смирнову за предложенную тему работы, а также, ведущему научному сотруднику Зоологического музея МГУ А.А. Лисовскому, за рекомендации по использованию метода.

СПИСОК ЛИТЕРАТУРЫ

1. Воронцов К. В., 2004. Комбинаторный подход к оценке качества обучаемых алгоритмов. — Математические вопросы кибернетики / Под ред. О. Б. Лупанов. — М.: Физматлит. — Т. 13. — С. 5–36.;
2. Горкин А. П., 2006 г. «География» современная иллюстрированная энциклопедия / Росмэн-Пресс, 685 с.;
3. Демянчик В. Т., Демянчик М. Г., 2001. Рукокрылые Беларуси: справочник-определитель // Брест: Изд-во С. Лаврова;
4. Дудов С. В., 2016. Моделирование распространения видов по данным рельефа и дистанционного зондирования на примере сосудистых растений нижнего горного пояса хр. Тукурингра (Зейский заповедник, Амурская область) // Журнал общей биологии. — Т. 77. — №. 2. — С. 122-134;
5. Ильин В. Ю., Смирнов Д. Г., 2000. Особенности распространения оседлых видов рукокрылых (Chiroptera: Vespertilionidae) на востоке Русской равнины и в смежных регионах // Экология. — Т. 2. — С. 118-124;
6. Ильин В. Ю. и др., 2002. Материалы к кадастру рукокрылых (Chiroptera) Европейской России и смежных регионов // Пенза: Пензенский гос. пед. ун-т;
7. Книжников Ю. Ф., Кравцова В. И., Тутубалина О. В., 2004. Аэрокосмические методы географических исследований: Учеб. для студ. высш. уч. заведений. М.: Изд. центр “Академия”. — 336 с.;
8. Лисовский А. А., Оболенская Е. В., 2014. Исследование ареалов мелких млекопитающих Юго-Восточного Забайкалья методом моделирования экологической ниши // Журнал общей биологии. — Т. 75. — №. 5. — С. 353-371;
9. Пузаченко Ю. Г., Желтухин А. С., Сандлерский Р. Б., 2010. Анализ пространственно-временной динамики экологической ниши на

- примере популяции лесной куницы (*Martes martes*) // Журнал общей биологии. – Т. 71. – №. 6. – С. 467-487;
- 10.Портал «Млекопитающие России» [Электронный ресурс] // URL: <http://rusmam.ru/> (дата обращения 01.06.2019 г.);
- 11.Рахматуллина И. Р., Рахматуллин З. З., Латыпов Э. Р., 2017. Моделирование условий произрастания и анализ вклада факторов в формирование высокобонитетных насаждений сосны обыкновенной (*Pinus sylvestris* L.) в программе MaxEnt (на примере Бугульминско-Белебеевской возвышенности в пределах Республики Башкортостан) // Природообустройство. – №. 3;
- 12.Россия в цифрах., 2017: Краткий статистический сборник / Росстат- М., 2017 - 511 с.;
- 13.Сабденова А. А. и др., 2016. Применение метода компьютерного моделирования Maxent для анализа распространения марала *Cervus elaphus sibiricus* на территории национального парка «Колсай колдери» // Инновационные технологии в науке и образовании. – №. 2. – С. 13-19;
- 14.Стрелков П.П., 1963. Рукокрылые. В кн.: Млекопитающие фауны СССР. М.-Л.: АН СССР. С. 159-161;
- 15.Affine. V. 2.2.2 [Электронный ресурс] // URL: <https://pypi.org/project/affine/> (дата обращения 01.06.2019 г.);
- 16.Beck A., 1995. Fecal analyses of European bat species // *Myotis*. – Т. 32. – №. 33. – С. 109-119;
- 17.Bishop C. M., 2006. Pattern recognition and machine learning. – springer;
- 18.Cameron A. C., Trivedi P. K., 2005. Microeconometrics: methods and applications. – Cambridge university press;
- 19.Davis J., Goadrich M., 2006. The relationship between Precision-Recall and ROC curves // Proceedings of the 23rd international conference on Machine learning. – ACM. – С. 233-240;

- 20.GDAL. V. 2.4.0 [Электронный ресурс] // URL: <https://pypi.org/project/GDAL/> (дата обращения 01.06.2019 г.);
- 21.GRASS GIS – Geographic Resources Analysis Support System [Электронный ресурс] // URL: <https://grass.osgeo.org/> (дата обращения 01.06.2019 г.);
- 22.Hutchinson G.E., 1957. Concluding Remarks // Cold Spring Harbor Symposium on Quantitative Biology «Population Studies: Animal Ecology and Demography», V. 22, P. 415-427;
- 23.Jupyter Notebook [Электронный ресурс] // URL: <https://jupyter.org/> (дата обращения 01.06.2019 г.);
- 24.Maxent software for species habitat modeling. V. 3.4.1 [Электронный ресурс] // URL: <https://github.com/mrmaxent/Maxent> (дата обращения 01.06.2019 г.);
- 25.MODIS – Moderate Resolution Imaging Spectroradiometer [Электронный ресурс] // URL: <https://modis.gsfc.nasa.gov/data/> (дата обращения 01.06.2019 г.);
- 26.Numpy. V. 1.16.2 [Электронный ресурс] // URL: <https://www.numpy.org/> (дата обращения 01.06.2019 г.);
- 27.Pandas. V. 0.24.1 [Электронный ресурс] // URL: <https://pandas.pydata.org/index.html> (дата обращения 01.06.2019 г.);
- 28.Phillips S. J., Anderson R. P., Schapire R. E., 2006. Maximum entropy modeling of species geographic distributions // Ecological modelling. – Т. 190. – №. 3-4. – С. 231-259;
- 29.Phillips S. J., Dudík M., 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation // Ecography. – Т. 31. – №. 2. – С. 161-175;
- 30.PyProj. V. 1.9.6 [Электронный ресурс] // URL: <http://pyproj4.github.io/pyproj/v2.2.0rel/> (дата обращения 01.06.2019 г.);
- 31.Python – programming language. V. 3.6.5 [Электронный ресурс] // URL: <https://www.python.org/> (дата обращения 01.06.2019 г.);

- 32.QGIS. V. 3.4.4 [Электронный ресурс] // URL: <http://www.qgis.org/ru/site/>
(дата обращения 01.06.2019 г.);
- 33.Rasterio. V. 1.0.24 [Электронный ресурс] // URL:
<https://pypi.org/project/rasterio/> (дата обращения 01.06.2019 г.);
- 34.Scikit-learn. V. 0.18.1 [Электронный ресурс] // URL: <https://scikit-learn.org/stable/> (дата обращения 01.06.2019 г.);
- 35.Smith L. I., 2002. A tutorial on principal components analysis;
- 36.Tipping M. E., Bishop C. M., 1999. Mixtures of probabilistic principal component analyzers // Neural computation. – Т. 11. – №. 2. – С. 443-482;
- 37.WorldClim – Global Climate Data. Free climate data for ecological modeling and GIS [Электронный ресурс] // URL:
<http://www.worldclim.org/bioclim> (дата обращения 01.06.2019 г.).