

# Экологические ниши, их современная интерпретация и моделирование с использованием пакета `ecospat`

Автор: Владимир Шитиков  
<https://stok1946.blogspot.com/>

Интенсивное развитие ГИС-технологий и дистанционных систем мониторинга, а также постоянно возрастающий объем общедоступных баз данных, обобщающих характеристики природных и антропогенных ландшафтов, динамику климатических изменений и широкий набор иных факторы окружающей среды, открывают новые возможности для изучения механизмов функционирования сообществ и их связи с условиями обитания популяций. За последние десятилетия мощным инструментом (макро)-экологических и биогеографических исследований стало моделирование распределения видов SDM (*Species Distribution Models*) и экологических ниш ENM (*Environmental Niche Models*). В наших предыдущих сообщениях [«Интерполяция и визуализация пространственных данных»](#) и [«Модели пространственного распределения видов»](#) мы сформировали компьютерную карту региона для визуализации результатов интерполяции и подробно рассмотрели ряд моделей SDM для таксонов макрозообентоса равнинных рек. Был также проиллюстрирован с помощью функций пакета `virtualspecies` один из подходов к построению ENM – метод виртуальных видов.

Однако на основе анализа некоторых публикаций экологического характера создается впечатление, что термин «экологическая ниша» трактуется излишне вольно, а основные теоретические концепции, разработанные в прошлом веке крупными учеными-экологами, не всегда принимаются во внимание. Целью настоящего сообщения является уточнение основных понятий экологических ниш и предпосылок их моделирования, а также анализ концептуальных различий SDM и ENM, которые аргументируются в публикациях (Soberón, 2007; Peterson, Soberón, 2012). В качестве примера нами рассматриваются модели экологических ниш и анализ их перекрытия для таксонов макрозообентоса, разработанные с использованием пакета `ecospat`.

## 1. Концепция экологических ниш, их классификация и моделирование

Область обитания вида обычно представляется на картах, в основе которых лежат наблюдения за временными и пространственными закономерностями, определяющими характер распределения отдельных организмов по изучаемой территории (Brown et al. 1996). Эти закономерности определяются тремя важнейшими группами факторов (Gaston, 2003; Guisan, Thuiller 2005): (а) пространственное распределение условий окружающей среды в диапазоне факторов, благоприятных для появления, выживания и размножения индивидов (в основном, в физиологическом смысле); (б) факторы биотической среды, определяющие идентичность и обилие "посредников" (таких, как опылители или переносчики семян), хищников, паразитов и конкурентов, а также доступность и динамику потребления ресурсов; и (в) способность видов к воспроизводству и расселению за счет собственных перемещений или с помощью внешних агентов,

Интерес к описанию, пониманию и прогнозированию географического и экологического распределения видов имеет давнюю историю (Wallace, 1860; Grinnell, 1917, 1924; Elton, 1927). Дж. Хатчинсон (Hutchinson, 1957, 1978), проанализировав эти работы, выделил два класса экологических ниш. **Ниши Гринелла** формируются в пространстве принципиально неинтерактивных переменных, необходимых для понимания крупномасштабных экологических и географических свойств видов и

отражающих условия окружающей среды, для которых конкуренция не имеет значения (Grinnell, 1917; Whittaker et al., 1973). Такие переменные были названы "сценопоэтическими" (*scenoproetic*), от греческих корней "постановка сцены". Другой класс – **ниши Эльтона** сосредоточены на учете всех типов биотических взаимодействий, которые оцениваются главным образом в локальных масштабах (Elton, 1927; MacArthur, 1969). Сюда включаются также интерактивные переменные среды, названные Хатчинсоном биономическими (*bionomic*), которые связаны с потребляемыми ресурсами, а население изучаемых сообществ может влиять на них и конкурировать за них. Различие между сценопоэтическими и биономическими переменными достаточно искусственно, поскольку некоторые из них могут в зависимости от масштаба играть важную роль при описании ниш обоих классов (например, солнечное излучение является и функцией географической широты, и фактором освещенности лесной подстилки).

Основополагающие идеи фундаментальных и реализованных ниш Хатчинсона отражают общие концепции  $n$ -мерного гиперпространства переменных и могут быть применены к нишам обоим классам, однако детали значительно меняются, поскольку проблемы измерения сценопоэтических и биономических осей совершенно разные. Отметим, что когда Хатчинсон (1957) определил нишу как совокупность всех "...состояний окружающей среды, которые позволили бы видам... существовать бесконечно", он не делал отличий между биотическими и сценопоэтическими переменными, хотя позже осознал их важность (Hutchinson, 1978).

Предположим, что область исследований  $\mathbf{G}$  разделена сеткой географических координат, ячейки которых имеют разрешение, соизмеримое с масштабом гринеллевских переменных (например,  $>10^3$  км<sup>2</sup>). Прежде, чем оценивать функции распределения анализируемого вида, необходимо принять соглашения об условиях абстрагирования каждой ячейки как "присутствие вида" с учетом масштаба и темпов демографических процессов. Этими условиями могут быть, например: "плотность населения превышает некоторый заданный порог" или "уровень рождаемости особей превышает уровень смертности". Часто оперируют неотрицательным параметром внутреннего темпа роста  $r$  (*intrinsic growth rate*) популяции, оцениваемом в предположении, что биотические факторы среды не налагают никаких ограничений.

Если для каждого вектора  $\mathbf{e}_j$ , являющегося  $j$ -й произвольной комбинацией переменных окружающей среды, известно значение внутреннего темпа роста  $r$ , то **фундаментальную нишу** Гринелла можно строго определить как  $\mathbf{N}_F = \{\mathbf{e}_j \mid r(\mathbf{e}_j) > 0\}$ , т.е. «набор всех возможных векторов  $\mathbf{e}$   $n$ -мерного пространства переменных среды, для которых внутренний темп роста численности вида положителен и/или популяционная плотность превышает заданный порог» (Soberón, 2007, p. 1118). Отметим, что нишу как среду обитания следует отличать от ниши как функции (Whittaker et al., 1973). Если рассматривать конкретное географическое пространство  $\mathbf{G}$ , то можно найти все ячейки в  $\mathbf{G}$  со значениями векторов переменных, принадлежащих  $\mathbf{N}_F$ :

$$\mathbf{J}_F = \{j \in \mathbf{G} \mid r(\mathbf{e}_j) > 0\}.$$

Этот набор ячеек, привязанный к сетке географических координат, целесообразно определить как **фундаментальную область распространения вида** (*fundamental area of distribution of a species*). На основе  $\mathbf{N}_F$  можно вычислить площадь  $\mathbf{J}_F$ , но теоретически допустима и обратная операция: если известна  $\mathbf{J}_F$ , то из ее ячеек можно извлечь комбинации векторов переменных среды для построения  $\mathbf{N}_F$ . Однако эта операция справедлива только в предположении, что  $\mathbf{J}_F$  является потенциальной областью распространения, и такие факторы, как перемещения особей и взаимодействия между видами, игнорируются.

Дж. Хатчинсон определил фундаментальную нишу, как объем  $n$ -мерного гиперпространства всех внутренних и внешних факторов, необходимых для выживания каждого вида в условиях отсутствия других видов, а **реализованную нишу** – как часть

фундаментальной ниши, в которой присутствие вида зафиксировано в условиях лимитированных ресурсов и межвидовых взаимодействий. Однако, даже в случае ниши Гринелла, сведение  $N_F$  к реализованной нише является сложной проблемой, начиная с того факта, что по определению не может существовать конкуренции за переменные среды. Следовательно, уменьшение площади  $N_F$  может происходить только за счет процессов в эльтоновской нише. Дж. Чейз и М. Лейболд (Chase, Leibold, 2003) подробно рассмотрели, как фундаментальная ниша Эльтона сокращается в ходе конкуренции за ресурсы при переходе к реализованной нише, которая также может быть выражена в пространстве ресурсных переменных. Они уточнили основные определения ниш с учетом того, что вид потенциально может существовать только там, где плотность ресурсов превышает минимальные потребности вида, чтобы поддерживать положительный (или нейтральный) чистый рост населения.

Таким образом, общий подход к моделированию экологических ниш основывается на системе уравнений популяционной динамики, и использует разложение суммарной скорости роста численности вида в заданной точке пространства (или каждой  $j$ -й ячейки на сетке) на три положительных или отрицательных компонента (Soberón, 2007):

- внутренний темп роста  $r(e_j)$ , не зависящий от интенсивности биотических взаимодействий и определяющийся (при наличии необходимых ресурсов) простым вектором крупномасштабных ( $>10^3$  км<sup>2</sup>) факторов среды гринеллевских ниш, таких как средняя температура, осадки, солнечная радиация, рельеф местности и т. д.;

- изменения темпов роста  $c(R_j)$ , обусловленные биотической средой и конкуренцией за ресурсы, в котором доминируют процессы ниши Эльтона с пространственным разрешением, соизмеримым с перемещениями индивидов (до 10 км<sup>2</sup>);

- компонент, связанный со структурой метапопуляции и динамикой дрейфа, заданный матрицей переходов  $T$ , которая определяет вероятности того, что в любой ячейке произойдет прием и отправка иммигрантов в течение определенного периода времени.

Эта упрощенная теоретическая схема позволяет определять ниши с точки зрения наборов параметров, а области распространения вида – с точки зрения наборов ячеек с фактическими или потенциальными демографическими свойствами популяций. Однако практическое моделирование ниш в зависимости от перечисленных трех групп факторов является чрезвычайно сложной задачей. Первая проблема связана с различным масштабом ниш. Ниша Эльтона имеет мелкочаеистую экологическую структуру и, в результате хищничества или конкуренции за ресурсы, многие местные популяции могут быть сокращены или даже уничтожены. Однако в географических масштабах, достаточно больших, чтобы нивелировать некоторую пространственную неоднородность, конкурентные виды могут сосуществовать в течение длительных периодов времени, и это никак не будет сказываться на размерах ниши Гринелла.

Вторая проблема заключается в том, что на практике результаты взаимодействий редко, если вообще когда-либо, хорошо предсказуемы по данным полевых наблюдений и экспериментов. Матрица  $R_j$  биологических параметров модели эльтоновской ниши определяет динамику использования ресурсов в ячейке  $j$ , их доступность и влияние вида на них, а также итог взаимодействия с каждым из видов-конкурентов. Она не может быть определена просто как набор чисел, поскольку для оценки ее компонентов требуются гипотезы о конкретных формах и моделях взаимодействий (Chase, Leibold, 2003).

Некоторые исследователи видят серьезные основания для беспокойства из-за отсутствия концептуальной строгости, так как во многих статьях описывается процесс моделирования распределения видов SDM, тогда как на самом деле рассматриваются модели ENM, связанные с экологическими нишами, и наоборот. Они считают, что

«термин ENM следует использовать только тогда, когда основное внимание уделяется оценке любого потенциального распределения в гипотетических условиях и обстоятельствах, но необходимо проявлять осторожность, чтобы отличать эти величины от их "существующих" или "реализованных" проявлений; с другой стороны, использование термина SDM должно включать шаги по преобразованию областей, оцененных как потенциальные, в фактические, с тем, чтобы точно реконструировать распределения» (Peterson, Soberón, 2012, p. 4).

Действительно, фундаментальная ниша Гринелла  $N_F$  далеко не полностью представлена географией изучаемой области, так как она представляет собой "конструкцию физиологических реакций" и распространяется на любые произвольные наборы условий  $e_j$ , даже находящиеся вне реального мира. Однако современный процесс разработки SDM включает в себя, по крайней мере, два перехода между пространствами: начальные стадии моделирования выполняются в географическом пространстве, далее, сама модель формируется в пространстве факторов окружающей среды, а результаты модели, как правило, визуализируются снова в географическом пространстве. Таким образом, подлинное "моделирование распределения видов" включает в себя как этап выделения экологической ниши, так и этапы оценки способности вида к распространению или его потенциал колонизации. Другое дело, что современная трактовка термина "экологическая ниша", к сожалению, все больше обесценивается: под ней понимается лишь простейшая фундаментальная ниша Гринелла, тогда как основополагающие теоретические представления Эльтона, Хатчинсона и МакАртура выносятся, как бы за скобки.

Несмотря на неоднозначность терминологии, моделирование ниш стало основной методологией количественной оценки связи структуры экологических сообществ с факторами окружающей среды. По состоянию на 18.02.2008 г. быстрый поиск в ISI Web of Science с использованием термов "*species distribution models*" или "*niche models*" или "*habitat models*" обнаружил 21 973 статей, 15% из которых посвящены биологии морских и внутренних вод (Thuiller et al., 2009). К настоящему времени этот объем, вероятнее всего, удвоился, однако в подавляющем большинстве построение моделей основывалось на легкодоступных географических и климатических переменных или факторах абиотической среды, которые имеют свою собственную внутреннюю динамику и не зависят от встречаемости или численности рассматриваемых видов. Наборы переменных, которые динамически связаны с численностью популяций, такие как лимитирующие пищевые ресурсы, влияние конкурентов или хищников и межвидовые взаимодействия в практике построения моделей почти не встречаются. Сместились и терминологические акценты: целью моделирования является, в основном, «географическое воплощение потенциальной ниши ("*potential niche*") как пересечение фундаментальной экологической ниши, физиологически определяющей диапазоны переменных среды, в которых вид может выжить, с набором факторов среды, которые фактически существуют на рассматриваемом реальном ландшафте» (Barve et al., 2011).

Постоянно разрабатываются новые методы и алгоритмы, позволяющие сделать моделирование как экологических ниш ENM, так и распределений видов SDM более удобным, полным, интерпретируемым и адекватным. Ранние модели SDM были ориентированы исключительно на географические измерения, а, как в случае пространственной эпидемиологии, вообще могли не иметь привязки к условиям окружающей среды. Однако многие современные методы (Guisan et al., 2017; Ovaskainen, Abrego, 2020; Шитиков и др., 2021a) в качестве основной цели прогнозирования ставят оценку степени различий между фактическим ареалом вида и его потенциальным распределением на основе статистического анализа комплекса внешних переменных.

Для оценки географических ареалов видов получил широкое распространение метод максимальной энтропии MaxEnt (Phillips et al., 2006; Лисовский и др., 2020; Шитиков и др., 2021б), который с байесовской точки зрения из всех возможных распределений вероятностей наилучшим образом представляет моделируемые данные. Итогом работы MaxEnt является обобщенная экспоненциальная функция, аргументами которой являются частные функции отдельных предикторов (линейные, квадратичные, множественные и др.) с настроенными коэффициентами  $\lambda$ , оценивающими вклад соответствующего экологического фактора, т.е. по сути, модель экологической ниши (Petrosyan et al., 2020). С другой стороны, связь между различными факторами в географическом и экологическом пространствах исследуют с помощью моделей ENM, построенных для совокупности видов, называемых искусственными или "виртуальными", с одинаковым и известным откликом по отношению к внешним воздействиям (Hirzel et al., 2001).

Одна из трудностей использования моделей SDM-ENM заключается в том, что количество доступных алгоритмов и программных средств велико и неуклонно растет, что затрудняет выбор наиболее подходящей методологии. Расхождения между различными технологиями могут быть очень большими, но, тем не менее, до сих пор в полной мере отсутствует не только единая теория, но и конкретные практические рекомендации. Это обусловлено как объективно существующим многообразием изучаемых экологических сообществ, природно-климатических зон, жизненных форм и техник проведения наблюдений, так и большим арсеналом разработанных методов компьютерной обработки и верификации моделей, выбор которых в значительной мере определяется субъективными вкусами исследователей. В частности, развернутый анализ результатов использования 33 моделей SDM на сообществах птиц, бабочек, деревьев и травянистой растительности (Norberg, 2019) обозначил лишь самые общие ориентиры.

Компьютерные решения по построению моделей SDM-ENM реализуются, в основном, в форме функций пакетов статистической среды R (*R Development Core Team*): *SDMvspecies* и *virtualspecies* для различных схем "виртуальных" видов, *dismo*, *ecospat*, *modler* для калибровки моделей MaxEnt и их сравнительных исследований, *ENMTML*, *ENMTools*, *NicheToolBox*, *BIOMOD2*, *ENMeval*, *spThin*, *kuenm*, *SDMTools*, *sdm* и многие другие пакеты для подготовки исходных данных, построения, прогнозирования, графического иллюстрирования, тестирования и оценки моделей экологических ниш с использованием различных алгоритмов. Альтернативой R может быть кросс-платформенное интерактивное приложение *NicheA* (*Niche Analyst*), выпущенное по публичной лицензии GNU и написанное на языке Java (Qiao et al., 2016).

## **2. Состав пакета *ecospat* и структура тестируемых исходных данных**

Декларируемой целью разработки пакета *ecospat* является предоставление новых инструментов и методов для поддержки пространственного анализа и моделирования экологических ниш и распределений видов в согласованной рабочей среде (Di Cola et al., 2017). Пакет содержит три группы функций: предварительный анализ данных (*Pre-modeling*), базовое моделирование и анализ построенных моделей (*Post-modeling*) – см. табл. 1. Предварительный анализ включает процедуры квантификации ниш видов и их сравнение с учетом различных областей или периодов времени, меры филогенетического разнообразия и другие функции исследования данных (например, обнаружение экстраполяции, *ExDet*). Базовое моделирование объединяет новый подход к построению ансамбля малых моделей (*ESM – ensemble of small models*) и различные реализации структуры пространственно-обусловленного моделирования сообществ видов (*SESAM – spatially-explicit modeling of species*



ecospat.ESM.EnsembleModeling	Оценивает и усредняет простые двумерные модели в прогнозирующий коллектив.
ecospat.ESM.Projection	Проецирует простые двумерные модели в новое пространство или время.
ecospat.ESM.EnsembleProjection	Проекты откалиброванный коллектив ESM в новое пространство или время.
ecospat.SESAM.prr	Использование структуры SESAM для прогнозирования состава сообщества с использованием правила "ранжирования вероятности".
ecospat.migclim	Реализует ограничения на распространение конкретных видов в моделях распределения видов
<i>Анализ моделей</i>	
ecospat.varpart	Выполняет разделение дисперсии для биномиальных GLM на основе девианс двух групп или прогнозирующих переменных.
ecospat.cons_Cscore	Проверяет неслучайные закономерности сосуществования видов и вычисляет С-индекс для всего сообщества и для каждой пары видов.

Модели экологических ниш будем строить на примере данных гидробиологической съемки донных сообществ бассейна Средней и Нижней Волги (Зинченко, 2011) в разные месяцы вегетационного периода 1990-2019 гг. Гидробиологическую съемку макрозообентоса проводили на 90 малых и 12 средних равнинных реках, притоках Куйбышевского, Саратовского и Волгоградского водохранилищ, в том числе, на 6 реках аридного региона бассейна оз. Эльтон. Всего было выделено  $S = 740$  видов и таксонов бентоса рангом выше вида. В наших предыдущих сообщениях блога (ссылки см. во введении) мы уже подробно останавливались на этом исследовании.

Для выполнения нижеприведенных скриптов в файле на общедоступном ресурсе [http://www.ievbras.ru/ecostat/Kiril/R/Blog/WB\\_data.RData](http://www.ievbras.ru/ecostat/Kiril/R/Blog/WB_data.RData) находятся точечные данные, привязанные к географическим координатам 132 участков рек, где выполнялись гидробиологические пробы. Объект ggplot2, воспроизводящий компьютерную карту региона для визуализации результатов представлен в файле по адресу: [http://www.ievbras.ru/ecostat/Kiril/R/Blog/WB\\_map.RData](http://www.ievbras.ru/ecostat/Kiril/R/Blog/WB_map.RData). Чтобы воспроизвести построение обсуждаемых моделей, необходимо загрузить эти файлы и разместить их в рабочем каталоге среды R.

```
load(file="niche_b.RData")
ls()
[1] "df.clim" "df_spe" "df_var" "dfR"
str(df_spe)
'data.frame': 132 obs. of 9 variables:
 $ Name      : chr  "Актушка" "Аманак" "Анлы" "Б. Вязовка" ...
 $ Участок   : chr  "малая" "малая" "малая" "малая" ...
 $ NamRiver  : Factor w/ 132 levels "Акту", "Аман", ...: 1 2 3 4 5 6 7...
 $ X         : num  49 52 52.3 50.2 50.8 ...
 $ Y         : num  53.4 53.7 53.9 52.5 52.2 ...
 $ ChPro.o.: int  2 0 2 0 0 0 0 0 0 ... # Prodiamesa olivacea
 $ OlTub.t.: int  2 2 0 3 2 3 1 2 0 ... # Tubifex tubifex
 $ CoBer.sp: int  0 0 0 0 0 0 0 0 3 ... # Berosus sp.
 $ CrAsl.a.: int  0 0 0 2 0 0 0 2 0 ... # Asellus aquaticus
```

```
str(df_var)
'data.frame': 132 obs. of 5 variables:
 $ NamRiver: Factor w/ 132 levels "Акты", "Аман",...: 1 2 3 4 5 6 7 8...
 $ Ground  : int  1 3 5 5 3 3 3 3 6 ...
 $ Miner   : num  860 440 860 2131 530 ...
 $ NH4     : num  0.386 0.18 0.386 0.72 0.69 ...
 $ O2      : num  92.4 87.5 95.5 91.7 140 ...
```

В таблице `df_spe` представлен список изученных водотоков с их географическими координатами и число проб, в которых встретились некоторые таксономические группы бентоса – личинки комара-звонца `ChPro.o.`, малощетинковые черви `OlTub.t.`, жуки-падальщики `CoBer.sp` и водяной ослик `CrAsl.a.` Таблица `df_var` содержит наблюдаемые гидрохимические показатели в тех же точках отбора проб – минерализация, насыщение кислородом `O2`, содержание ионов аммония `NH4` и категория донного грунта `Ground` (от 1 – песок или галька до 6 – черные илы).

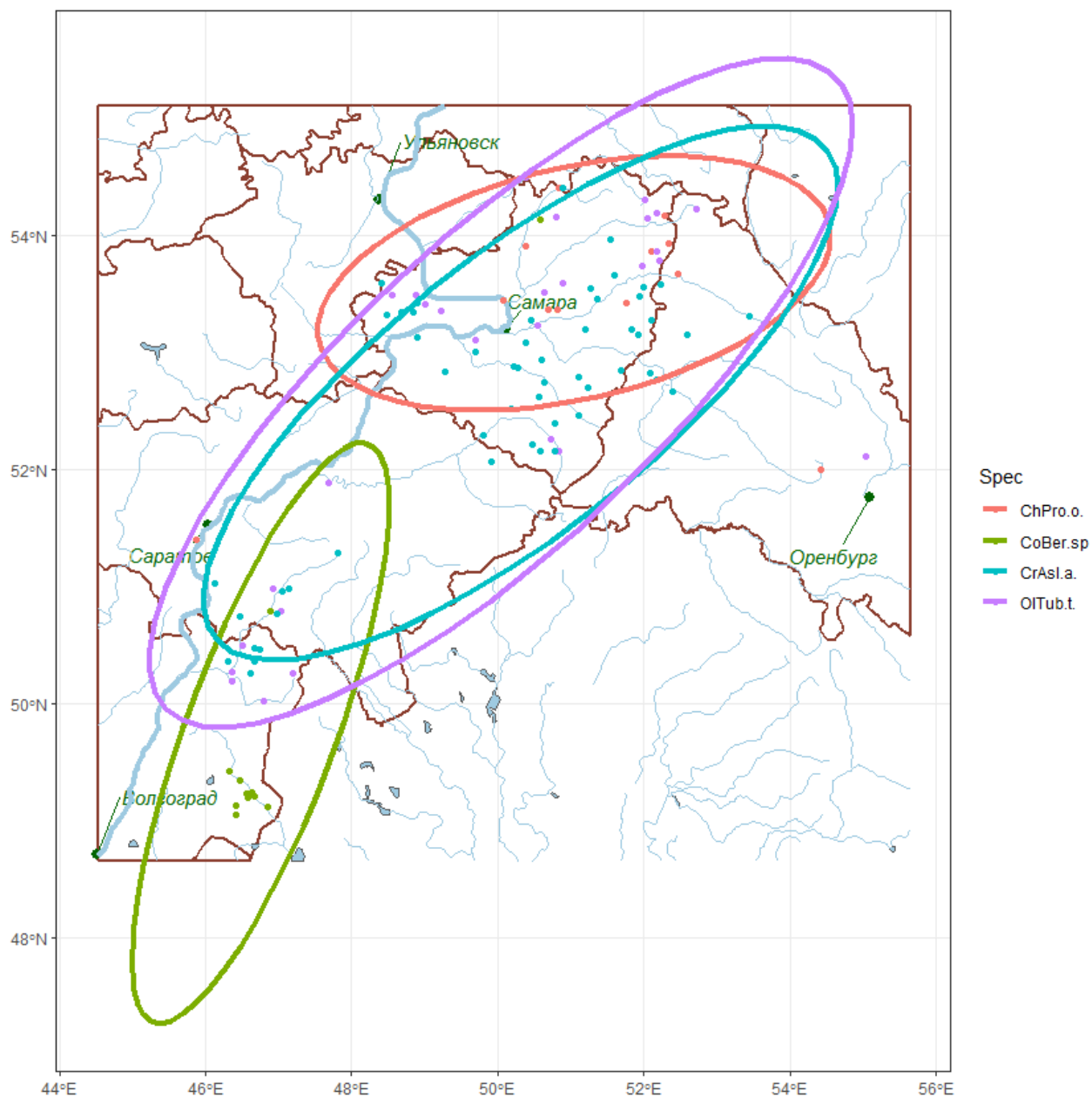
Кроме точечных данных из таблицы `df_var` в качестве предикторов для построения моделей ENM будем использовать биоклиматические и геофизические показатели, представленные в виде сеток (или матриц *grid*) на общедоступных серверах. Таблица `df.clim` содержит 22 таких показателя, вычисленных для географических координат каждого участка водотока. Естественно, что все биоклиматические данные представляют сильно коррелированный набор переменных и мы предварительно отобрали 7 базовых показателей с минимальным уровнем фактора инфляции дисперсии (VIF). Таблица `dfR` содержит данные семиканального растра  $155 \times 267$  по изучаемому географическому экстенду с этими показателями, смысл и обозначения которых будут представлены в таблице ниже.

Выполним визуализацию картосхемы исследуемого региона, нанесем на нее точки обнаружения отобранных видов макрозообентоса и построим для каждого из них 95%-ные эллиптические доверительные области двумерного нормального распределения:

```
load(file="WB_map.RData")
# Преобразование данных в «длинный» формат
library(tidyverse)
Df_1 <- gather(df_spe[, (4:9)], Spec, val, -X, -Y)
Df_1 <- Df_1[Df_1$val > 0, ]
Df_1$Spec <- as.factor(Df_1$Spec)

Basemap +
  geom_point(data = Df_1, aes(x=X, y=Y, color=Spec), size=1.5) +
  stat_ellipse(data = Df_1, aes(x=X, y=Y, color=Spec),
              size=1.5, type = "t") +
  theme_bw()
```





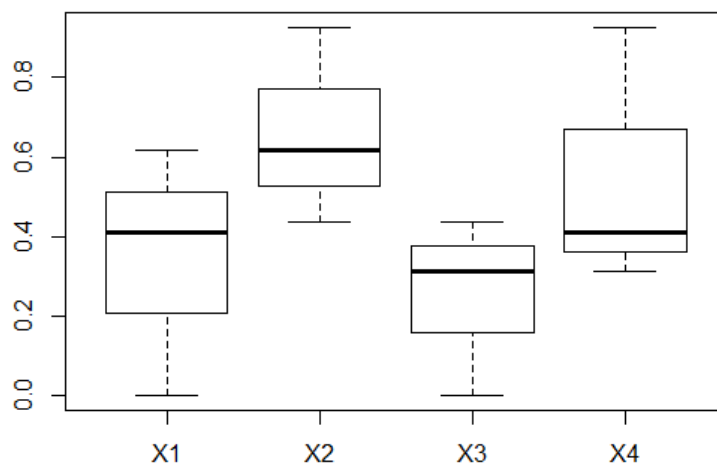
### 3. Функции предварительного анализа данных

Взаимная согласованность встречаемости видов оценивается большим количеством индексов (Gotelli, 2000), два из которых используется в пакете `ecospat`. Для подсчета индекса частоты совместных встреч количество участков, на которых обнаружена каждая из возможных пар видов ( $sp1$ ,  $sp2$ ), делится на число участков, на которых встретился самый редкий из них.

```
library(ecospat)
matr <- df_spe[, (6:9)]
matr[matr > 0] <- 1
matr <- matr[rowSums(matr) > 0, ]
# С удивлением обнаруживаем, что первые четыре вида попросту игнорируются
# в последующих расчетах, поэтому дублируем первых четыре столбца
matrix <- cbind(matr, matr)
```

```
ecospat.co_occurrences (data=matrix)
      ChPro.o. OlTub.t. CoBer.sp CrAsl.a.
ChPro.o. 1.0000000 0.6153846  0.0000 0.4102564
OlTub.t. 0.6153846 1.0000000  0.4375 0.9259259
CoBer.sp 0.0000000 0.4375000  1.0000 0.3125000
CrAsl.a. 0.4102564 0.9259259  0.3125 1.0000000
```

Из результатов видно, что виды ChPro.o. и CoBer.sp совместно не были обнаружены ни разу, а CrAsl.a. и OlTub.t. встречаются вместе в 93% случаев. Дополнительно выводится стандартная диаграмма «бокс-с-усами» с разбросом значений индекса для каждого из видов, из которой видно, что последняя пара таксонов имеет и наибольший средний индекс встречаемости.



Индекс C-заполнения (checkerboard score) оценивает среднее число фрагментов  $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$  и  $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$  для произвольной пары видов:  $CS = \sum_{ij} (n_i - n_{ij})(n_j - n_{ij})$ , где  $n_i$  и  $n_j$  – встречаемости видов  $i$  и  $j$ ,  $n_{ij}$  – одновременная встречаемость этой пары. Если значение стандартизированного размера эффекта SES больше 2 или меньше -2, то структурированность сообщества статистически значимо с вероятностью менее 0,05. Функция `ecospat.Cscore` возвращает средний индекс C-score для наблюдаемого сообщества (`ObsCscoreTot`) и рандомизованных матриц (`SimCscoreTot`), `SES.Tot` и соответствующие ему  $p$ -значения (`PVal.less` и `PVal.greater`). Информация по анализу C-score отдельных видов помещается в текстовый файл.

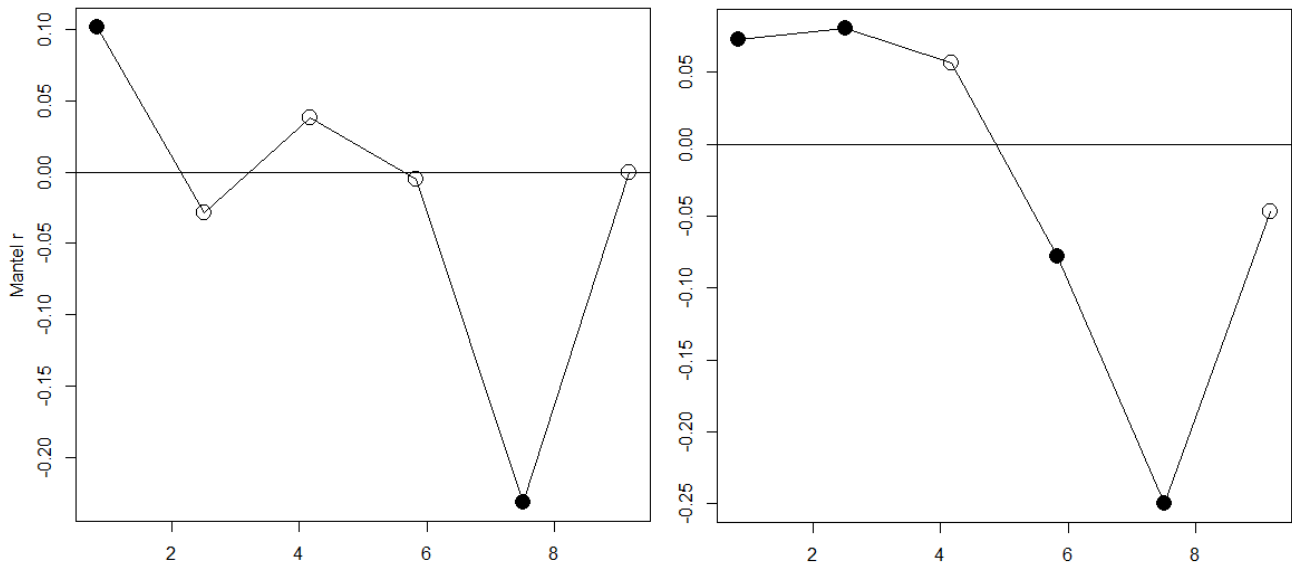
```
outpath <- getwd()
Res <- ecospat.Cscore(matr, 1000, outpath)
as.data.frame(Res)
  ObsCscoreTot SimCscoreTot PVal.less PVal.greater SES.Tot
1           609          576.478  0.999001  0.001998002  4.969678
```

```
# Помещено в файл Cscores.txt рабочего каталога
Sp1      Sp2      obs.C-sc exp.C-sc      SES_Cscore
ChPro.o. OlTub.t.    840    702      1.26317
ChPro.o. CoBer.sp   624    420      2.79490
ChPro.o. CrAsl.a.   874   1000      0.49148
OlTub.t. CoBer.sp   657    345      2.20430
OlTub.t. CrAsl.a.   120    560     -4.9968
CoBer.sp CrAsl.a.   539    423      0.78640
```

Поскольку в исходной матрице было найдено 609 сопряженных пар видов, когда как в 1000 рандомизованных матрицах их в среднем оказалось только 576, то это свидетельствует об отчетливых тенденциях взаимного исключения в сообществе. В то же время, для пары *OlTub.t.* – *CrAsl.a.* индекс *C-score* значительно ниже, чем для случайной организации связей, т.е. эти виды являются агрегированными.

Коррелограмма представляет собой график значений пространственной автокорреляции в зависимости от лага  $h$ . Типичной является коррелограмма, в которой значения  $I$  положительны для небольших расстояний, уменьшаются по мере увеличения лага до отрицательных значений и стабилизируются в точке  $h_{crit}$ , правее которой пространственную структуру можно считать статистически независимой. Стандартная ошибка выборочных значений и доверительные интервалы  $I$  могут быть найдены по формулам аппроксимации или методами Монте-Карло с использованием бутстрапа.

```
ecospat.mantel.correlogram(dfvar=df_spe,colxy=4:5, n=100, colvar=6:9,
                           max=10, nclass=7, nperm=100)
df_e <- cbind(df_spe[,4:5], df_var[,-1])
ecospat.mantel.correlogram(dfvar=df_e,colxy=1:2, n=100, colvar=3:6,
                           max=10, nclass=7, nperm=100)
```



*Сообщество из 4 видов макрозообентоса*

*Биотопические показатели*

*Черным цветом залиты статистически значимые величины автоковариации*

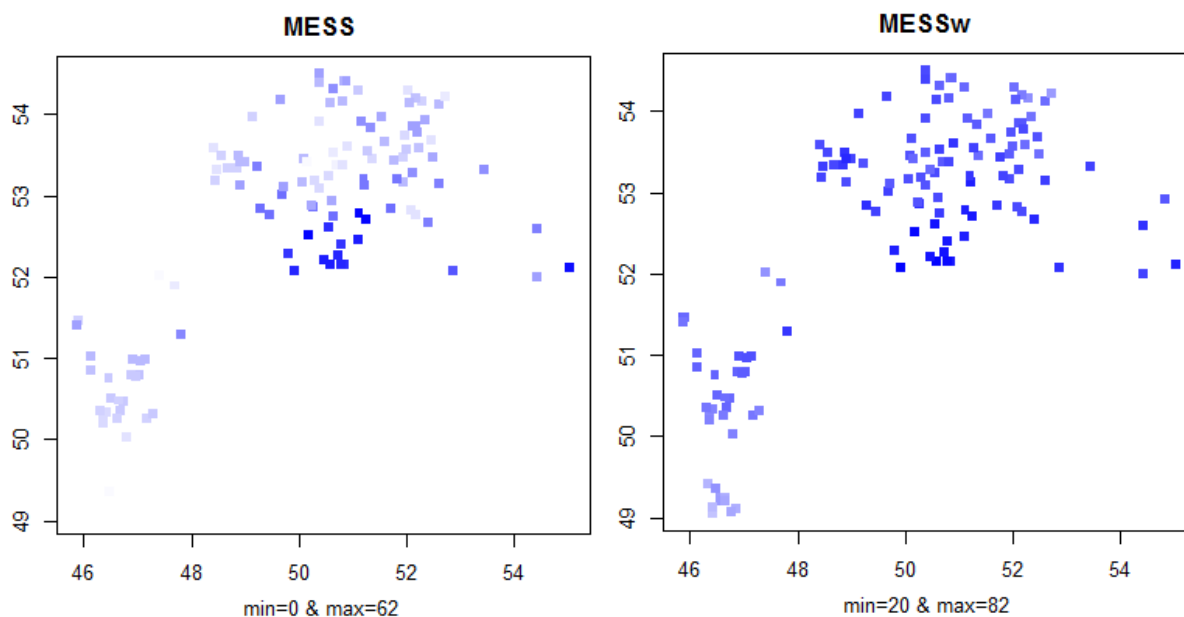
Определим теперь, сколько переменных в общей таблице геоклиматических показателей имеют коэффициент парной корреляции Пирсона менее 0.75.

```
x <- cor(df.clim, method="pearson")
ecospat.npred(x, th=0.75)
[1] 8
var_clim <- c("bio1", "bio3", "bio8", "bio15", "bio17", "alt", "tri")
df_m <- cbind(df_spe[,4:5], df.clim[,var_clim])
ecospat.mantel.correlogram(dfvar=df_m,colxy=1:2, n=100, colvar=3:9,
                           max=10, nclass=7, nperm=100)
```

Многомерная поверхность сходства окружающей среды (MESS – *Multivariate Environmental Similarity Surface*) измеряет сходство любой заданной точки с эталонным набором точек по отношению к выбранным переменным-предикторам. Она оценивает близости точки к распределению опорных точек, дает отрицательные значения для

разнородных точек и отображает эти значения по всей области прогнозирования (Elith et al., 2010). Функция `ecospat.mess` вычисляет два варианта карт – без использования весов (MESS) и с их использованием (MESSw). Данный пример показывает, что нет необходимости в расширении географического пространства для генерации фоновых точек, а пустая карта отрицательных значений свидетельствует о том, что ссылочная карта хорошо соответствует проецируемой (впрочем, в нашем случае иначе и быть не могло).

```
mess.object <- ecospat.mess (df_m, dfR, w="default")
ecospat.plot.mess (mess.object, cex=1, pch=15)
```



#### **4. Функции построения пространства ниш и оценка их перекрытия**

Как следует из изложенного в разделе 1, в подавляющем большинстве случаев под экологической нишей понимается фундаментальная ниша Гринелла. Один из распространенных подходов к моделированию пространства ниш заключается в оптимальном проецировании многомерной совокупности исходных коррелированных абиотических факторов на ортогональные оси пространства малой размерности. Для этого строят ординационные диаграммы в координатах первых двух главных компонент ( $PCA_1$  и  $PCA_2$ ) и калибруют их с использованием характера вариации переменных среды в пределах изучаемой области. На сетку ячеек, ограниченную минимальными и максимальными значениями главных компонент и определяющую экологическую изменчивость территории, наносят точки встречаемости анализируемых видов. Модель потенциальной экологической ниши рассматривается как некоторая область в пространстве латентных переменных, в которой плотность вероятности встречаемости вида соответствует заданным ограничениям.

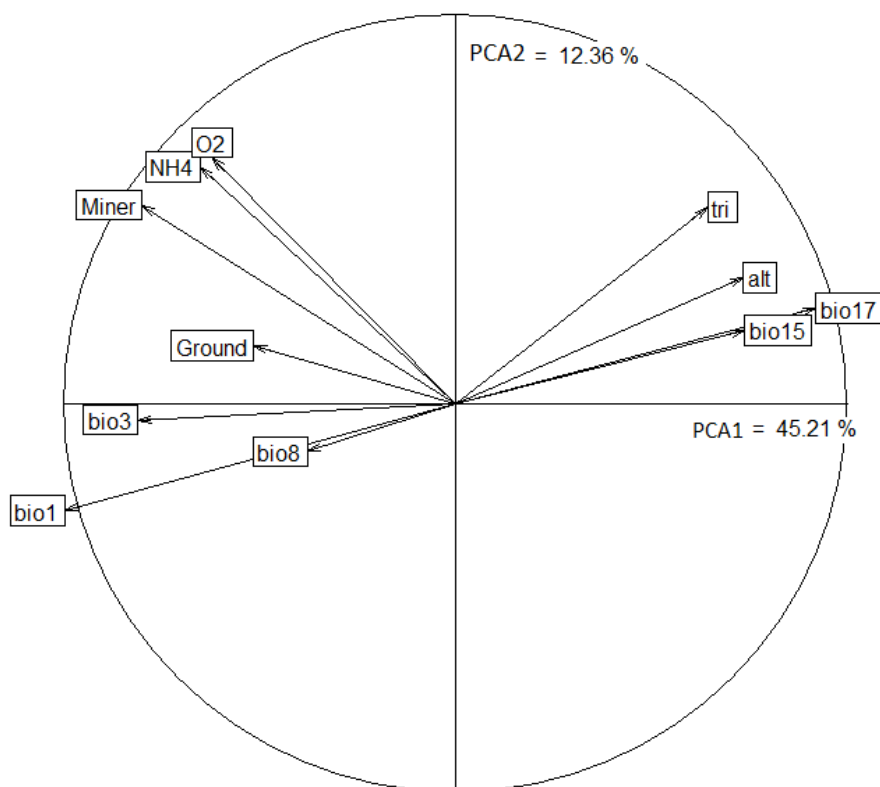
При использовании пакета `virtualspecies` для определения области экологической пригодности условий среды аппроксимируют гауссовыми кривыми распределение проекций точек встречаемости видов на каждую из осей  $PCA_1$  и  $PCA_2$ , после чего плотности вероятности обобщают для получения итоговых значений (Leroy et al., 2016). Экологическая ниша для такой простейшей функции ниши вида – эллипс на ординационной диаграмме с заданным уровнем доверительной вероятности и соответствующими диапазонами значений главных компонент.

Другой способ задать границы ниши, используемый в пакете `ecospat`, заключается в преобразовании бинарных переменных присутствия/отсутствия в

относительные плотности двумерного распределения путем сглаживания ядерными (*kernel*) функциями на сетке 100×100 с координатами PCA<sub>1</sub> и PCA<sub>2</sub> (Broennimann et al., 2012; Di Cola et al., 2017). После калибровки поверхности отклика *Z* в диапазоне [0, 1] экологическая ниша в такой интерпретации – некоторая нерегулярная область в пространстве двух главных компонент со значениями пригодности среды обитания (*environmental suitability*)  $z_{ij} > 0$ .

Первый этап построения поверхности ниши – анализ главных компонент с использованием полного набора из 11 переменных среды, измеренных в районе взятия гидробиологических проб:

```
Env <- cbind(df_var, df.clim[,var_clim])
Env$Miner <- log(Env$Miner)
# Анализ PCA с использованием функций пакета ade4
pca.env <- dudi.pca(Env[, -1], scannf=F, nf=2)
# главные компоненты для всей области исследования
scores.globclim <- pca.env$li
# PCA-нагрузки для всей области исследования
ecospat.plot.contrib(contrib=pca.env$co, eigen=pca.env$eig)
```



```
# Собственные числа, соответствующие главным компонентам
pca.env$eig
[1] 4.97342698 1.35988101 1.09108185 0.86996800 0.74167693 0.62506287
[7] 0.49281248 0.43122270 0.29641889 0.08665741 0.03179088
pca.env$co # Приведены в таблице
```

По результатам анализа главных компонент две главные оси максимальной дисперсии объясняют 57.5 % вариации данных, причем, как показано на корреляционном круге, основная ось PCA1 положительно коррелирует с высотой и количеством осадков, а отрицательно – с температурой и всем комплексом гидрохимических показателей. В следующей таблице приведены средние значения используемых показателей среды, их вклад (%) в модели прогнозирования MaxEnt и нагрузки на оси главных компонент PCA1- PCA2

Показатель	Среднее	Вклад	PCA1	PCA2
Среднегодовая температура (bio1)	51.9 ± 1.12	1.3	<b>-0.921</b>	-0.253
Изотермичность или отношение среднедневной разности температур к их годовой разности <i>min - max</i> (bio3)	21 ± 0.087	42.7	<b>-0.749</b>	-0.038
Средняя температура самой влажной четверти (bio8)	195 ± 3.9	6.36	-0.351	-0.113
Сезонная вариация осадков (bio15)	23.5 ± 0.266	3.91	<b>0.677</b>	0.174
Количество осадков в сухой четверти (bio17)	81.3 ± 1.16	1.72	<b>0.845</b>	0.226
Высота над уровнем моря (alt)	85.1 ± 5.1	36.2	<b>0.674</b>	0.299
Индекс шероховатости рельефа (tri)	68.3 ± 3.8	7.73	<b>0.591</b>	0.465
Категория грунта в баллах от 1 до 6 (Ground)			-0.476	0.137
Минерализация, г/л (Miner)	2.4 ± 0.62		<b>-0.739</b>	0.470
Содержание ионов аммония, г/л (NH4)	1.075 ± 0.31		<b>-0.603</b>	<b>0.561</b>
Насыщение кислородом, % (O2)	98.6 ± 2.5		<b>-0.574</b>	<b>0.583</b>

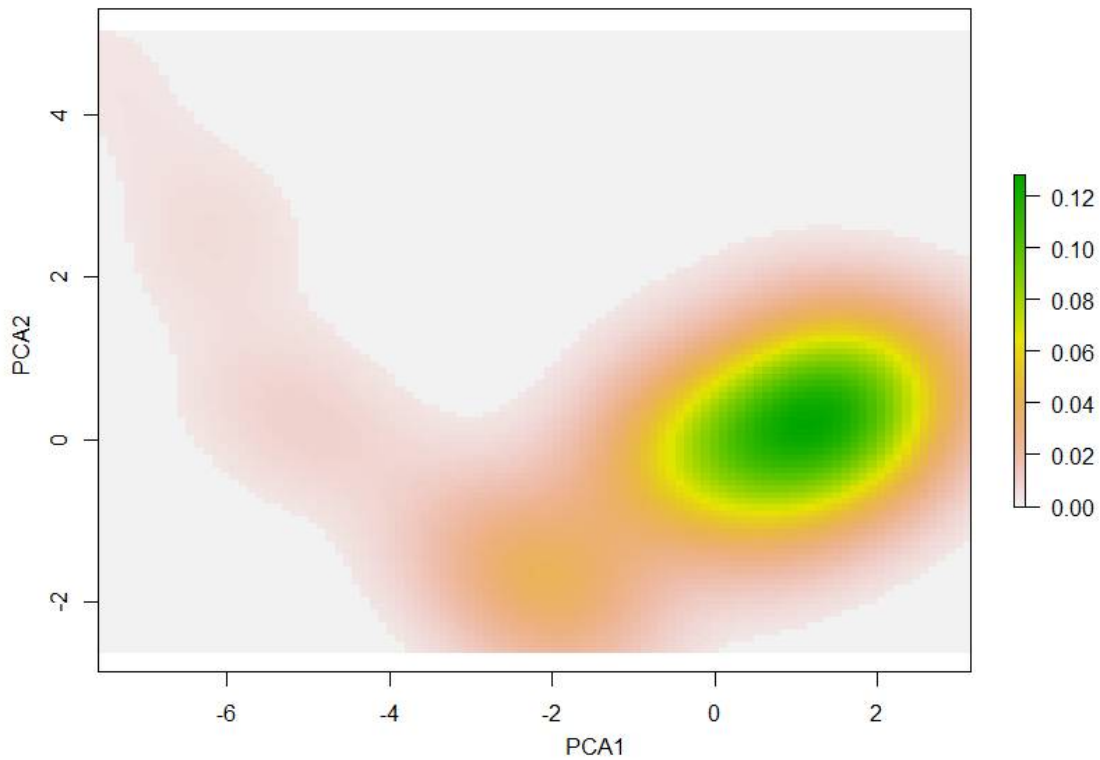
Выделим из общего массива значений главных компонент их подмножество, соответствующее точкам встречаемости анализируемого вида. Сделаем это для трех видов, представленных ранее, после чего для каждого из них сформируем объект, включающий совокупность компонентов класса `RasterLayer` (слой многоканального растра) на сетке из  $R=100 \times 100$  ячеек. Эти слои содержат сглаженные ядерными функциями нормированные и промежуточные значения пригодности среды обитания  $z_{ij} > 0$  (Broennimann et al., 2012). Поскольку обычно объем данных о видах невелик, для предсказаний плотности  $Z$  устанавливается слишком большая пропускная способность ядерной функции, и тогда многие ячейки растра будут принимать значения, близкие к 0. Установив  $th.sp > 0$ , можно удалить самые низкие плотности, предсказанные ядерной функцией по краям ниши: например, при  $th.sp = 0.05$  при сглаживании будет отброшено 5% маргинальных значений.

Сформируем растровые объекты для трех видов: *Prodiamesa olivacea* (ChPro.o.), *Tubifex tubifex* (OTub.t.) и *Berosus* sp. (CoBer.sp):

```
# главные компоненты для распределения каждого из трех выбранных видов
scores.sp.g1 <- suprow(pca.env, Env[which(df_spe$ChPro.o. >=1), -1])$li
scores.sp.g2 <- suprow(pca.env, Env[which(df_spe$OTub.t. >=1), -1])$li
scores.sp.g3 <- suprow(pca.env, Env[which(df_spe$CoBer.sp >=1), -1])$li
# Создание сетки для каждой из трех потенциальных ниш
grid.clim.g1 <- ecospat.grid.clim.dyn(glob=scores.globclim,
  glob1=scores.globclim, sp=scores.sp.g1, R=100, th.sp=0)
grid.clim.g2 <- ecospat.grid.clim.dyn(glob=scores.globclim,
  glob1=scores.globclim, sp=scores.sp.g2, R=100, th.sp=0)
grid.clim.g3 <- ecospat.grid.clim.dyn(glob=scores.globclim,
  glob1=scores.globclim, sp=scores.sp.g3, R=100, th.sp=0)
```

Построенная  $Z$ -поверхность ниши может быть визуализирована в двухмерном или трехмерном вариантах. Вся необходимая информация находится в слотах объекта `ecospat.grid.clim.dyn`:  $Z$  – грид для всей изучаемой территории;  $z$  – для ниши конкретного моделируемого вида:

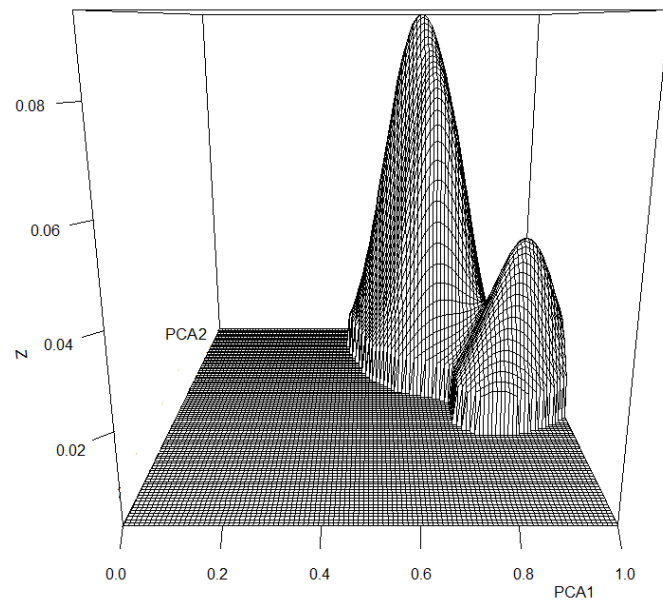
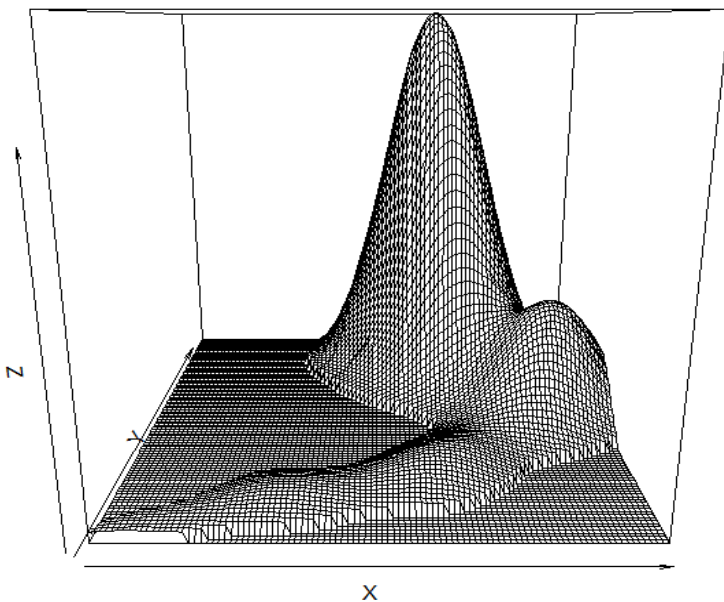
```
# диаграмма грида для всей изученной территории
X <- grid.clim.g1$Z
plot(X, xlab="PCA1", ylab="PCA2")
# пространственное распределение ниш каждого из трех выбранных видов
plot(grid.clim.g1$z)
plot(grid.clim.g2$z)
plot(grid.clim.g3$z)
```



```
# информацию из грида можно сохранить в виде матрицы
y <- raster::as.matrix(grid.clim.g1$Z)
y1 <- raster::as.matrix(grid.clim.g1$z)
y2 <- raster::as.matrix(grid.clim.g2$z)
y3 <- raster::as.matrix(grid.clim.g3$z)
```

Пространственные диаграммы можно построить с использованием функций `persp`:

```
persp(y)
persp(y2, xlab="PCA1", ylab="PCA2", ticktype = "detailed", nticks=6)
```

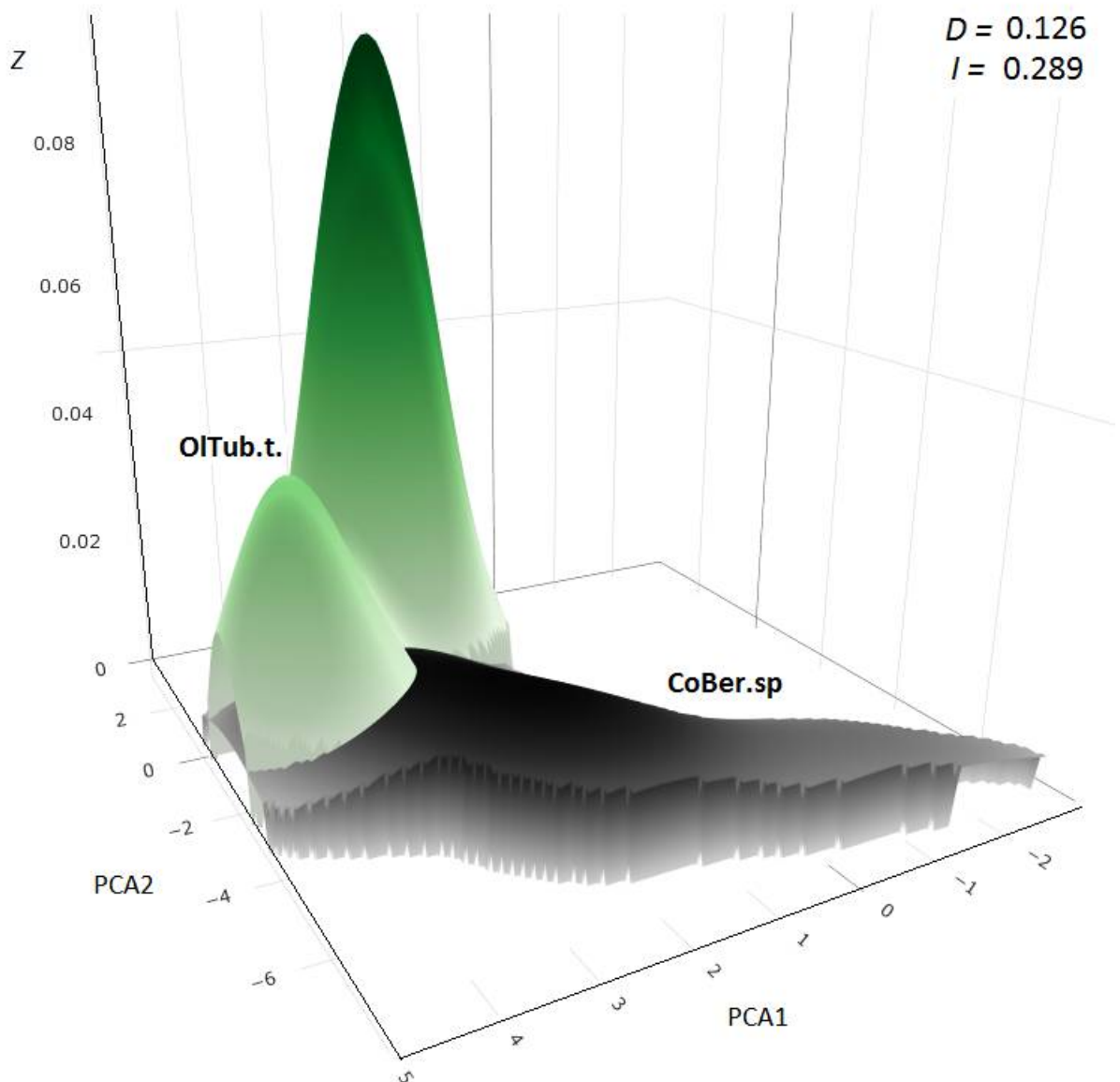


С использованием пакета `plot_ly` покажем перекрытие ниш *Tubifex tubifex* и *Berosus sp.*

```

library(plotly)
plot_ly(showscale = TRUE) %>%
  add_surface(z = ~y3, x=grid.clim.g3$x, y=grid.clim.g3$y,
             cmin = max(y3), cmax = min(y3), colorscale = "Greys",
             reversescale=TRUE) %>%
  add_surface(z = ~y2, x=grid.clim.g2$x, y=grid.clim.g2$y,
             cmin = max(y1), cmax = min(y1), colorscale = "Greens",
             reversescale=TRUE)

```



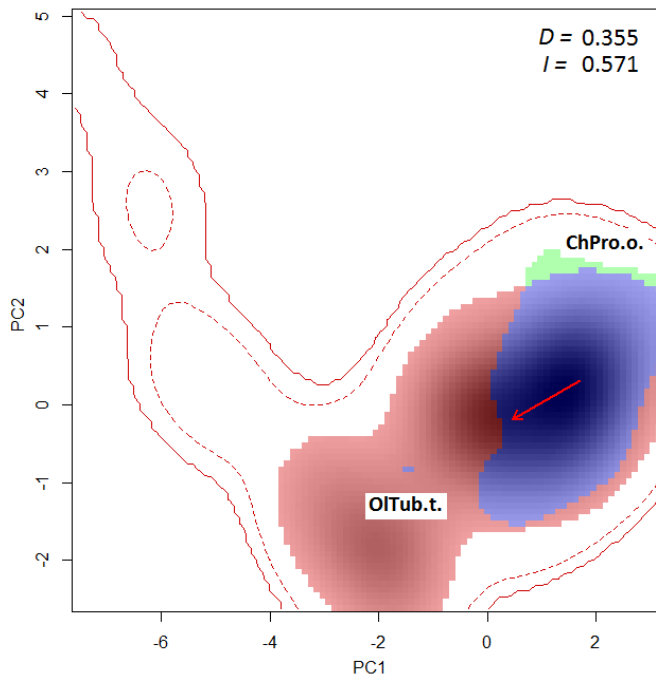
Построим ординационные диаграммы перекрытия ниш для каждой пары видов из трех, используя функцию `ecospat.plot.niche.dyn`. Красными контурными линиями обозначены границы изучаемой области (полной и с удалением 5% маргинальных участков). Красным и зеленым цветом закрашены ниши двух сравниваемых видов, а синим – совместно используемая область экологического пространства. Конец и начало стрелки показывают положение центроидов обеих ниш.

```

# Диаграммы перекрытия ниш для каждой пары сравниваемых видов
ecospat.plot.niche.dyn(grid.clim.g1, grid.clim.g2, quant=0.25, interest=2,
                      name.axis1="PC1", name.axis2="PC2")
ecospat.shift.centroids(scores.sp.g1, scores.sp.g2, scores.globclim,
                       scores.globclim )

```

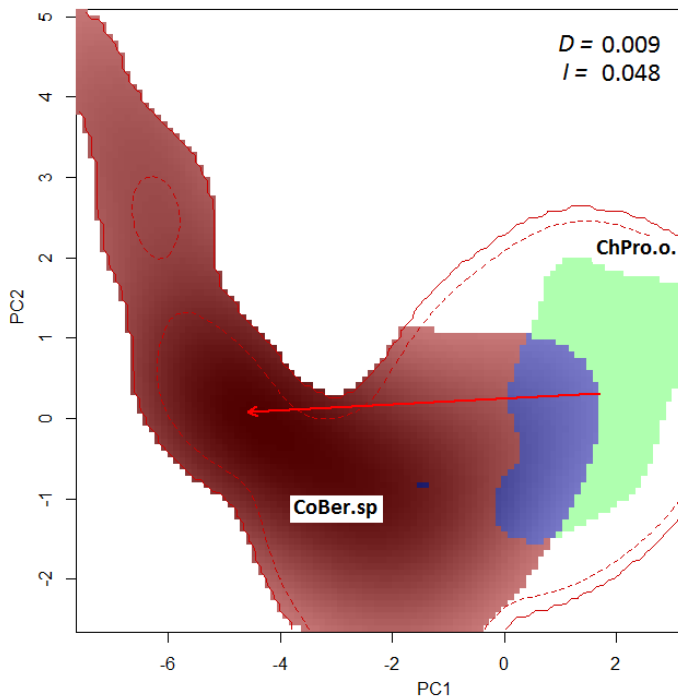




```

ecospat.plot.niche.dyn(grid.clim.g1, grid.clim.g3, quant=0.25, interest=2,
  name.axis1="PC1", name.axis2="PC2")
ecospat.shift.centroids(scores.sp.g1, scores.sp.g3, scores.globclim,
  scores.globclim )

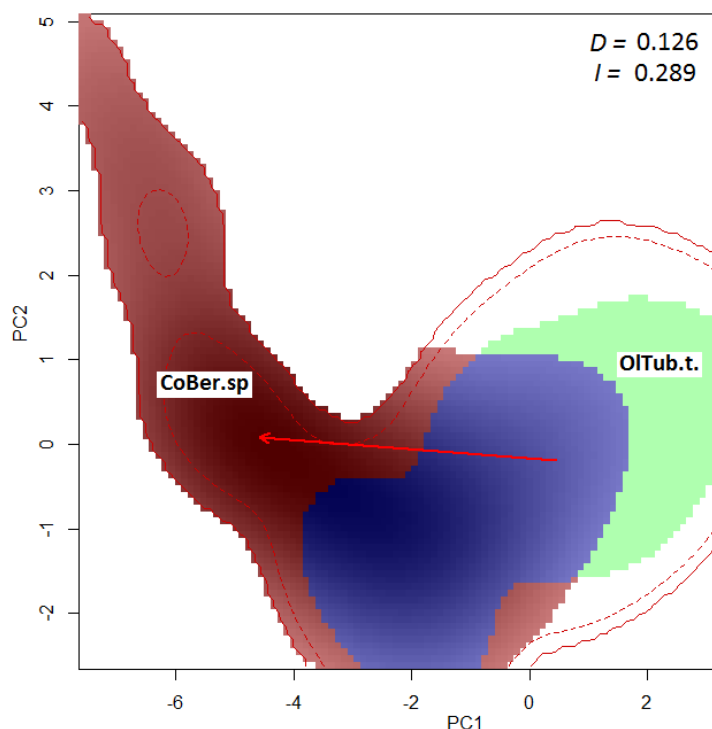
```



```

ecospat.plot.niche.dyn(grid.clim.g2, grid.clim.g3, quant=0.25, interest=2,
  name.axis1="PC1", name.axis2="PC2")
ecospat.shift.centroids(scores.sp.g2, scores.sp.g3, scores.globclim,
  scores.globclim )

```



Если ставится задача оценить долю перекрытия двух ниш, принадлежащих двум разным таксономическим группам, то используют такие индексы, как  $D$  Шенера (*Schoener*) и  $I$  Хеллингера (*Hellinger*). Эти индексы отражают меру совмещения ниш в заданном пространстве (Warren *et al.*, 2008) и колеблются между 0 (отсутствие перекрытия) и 1 (полное перекрытие):

$$D = 1 - 0.5 \left( \sum_{ij} |z_{1ij} - z_{2ij}| \right); \quad I = 1 - 0.5 \left( \sqrt{\sum_{ij} (z_{1ij}^{0.5} - z_{2ij}^{0.5})^2} \right).$$

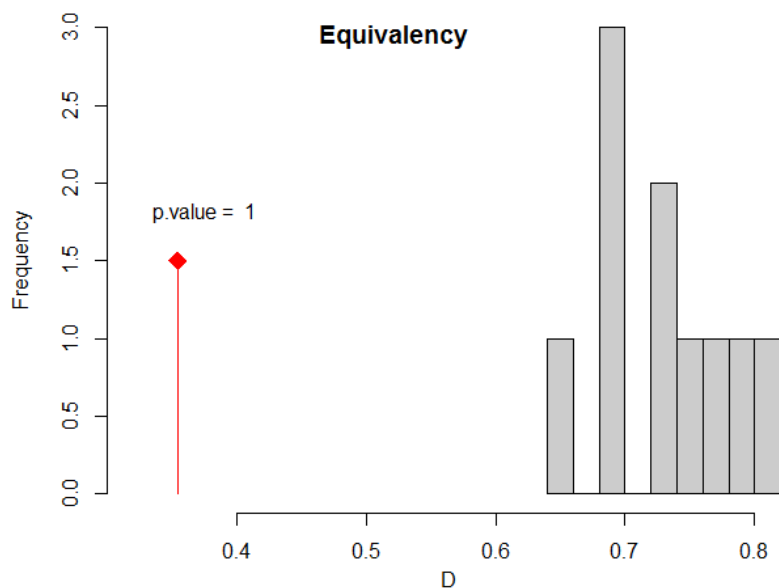
```
Res <- as.data.frame(ecospat.niche.overlap (grid.clim.g1,
      grid.clim.g2, cor = TRUE))
Res <- rbind(Res,as.data.frame(ecospat.niche.overlap (grid.clim.g1,
      grid.clim.g3, cor = TRUE)))
Res <- rbind(Res,as.data.frame(ecospat.niche.overlap (grid.clim.g2,
      grid.clim.g3, cor = TRUE)))
      D      I
1 0.355102712 0.57182408
2 0.009039101 0.04824245
3 0.126355538 0.28917921
```

Значения индексов перекрытия ниш  $D$  Шенера и  $I$  Хеллингера для трех сравниваемых видов приведены также в верхнем правом углу ординационных диаграмм перекрытия.

Статистическая значимость степени эквивалентности и сходства ниш оценивается методами рандомизации. Тест эквивалентности ниш определяет, остается ли перекрытие ниш постоянным при случайном перераспределении встречаемости обоих видов. Для этого все точки обнаружения видов объединяются и случайным образом многократно разбиваются на два набора данных того же объема, что и исходные выборки, и каждый раз рассчитывается статистика перекрытия ниши  $D$ . Этот процесс повторяется 100 раз (чтобы гарантировать, что нулевая гипотеза может быть отвергнута с высокой степенью достоверности), и строится гистограмма смоделированных значений. Если на построенной гистограмме наблюдаемое значение  $D$  попадает в 95%-й диапазон плотности нуль-моделей, то нулевая гипотеза об эквивалентности ниши не может быть отвергнута.

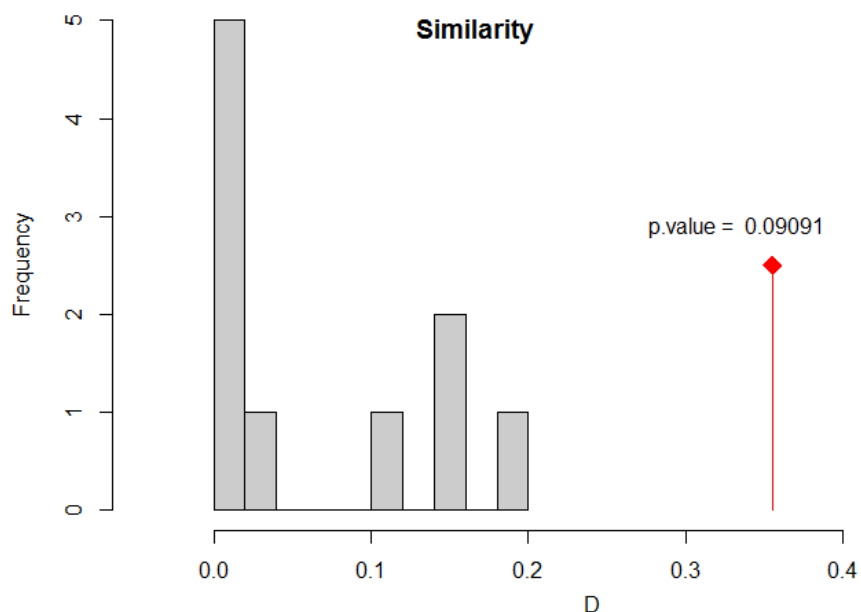
Выполним этот и последующие тесты для пары (ChPro.o.) – (OITub.t.)

```
eq.test <- ecospat.niche.equivalency.test(grid.clim.g1, grid.clim.g2,
  rep=10, alternative = "greater")
ecospat.plot.overlap.test(eq.test, "D", "Equivalency")
```



Тест на сходство ниш определяет, можно ли сходство двух экологических ниш объяснить случайными причинами. Для этого одна из ниш фиксируется, а вся наблюдаемая плотность встречаемости второго вида передвигается в случайно выбранную точку доступной области, которая является центром нового распределения. Этот процесс выполняется многократно и каждый раз вычисляются индексы перекрытия первой наблюдаемой ниши с нуль-моделью второй ниши. Если точка эмпирических значений  $D$  или  $I$  на гистограмме выходит за пределы 95%-го нуль-модельного диапазона, то ниши более похожи друг на друга, чем это ожидалось бы при случайных обстоятельствах.

```
sim.test <- ecospat.niche.similarity.test(grid.clim.g1, grid.clim.g2,
  rep=10, alternative = "greater", rand.type=2)
ecospat.plot.overlap.test(sim.test, "D", "Similarity")
```



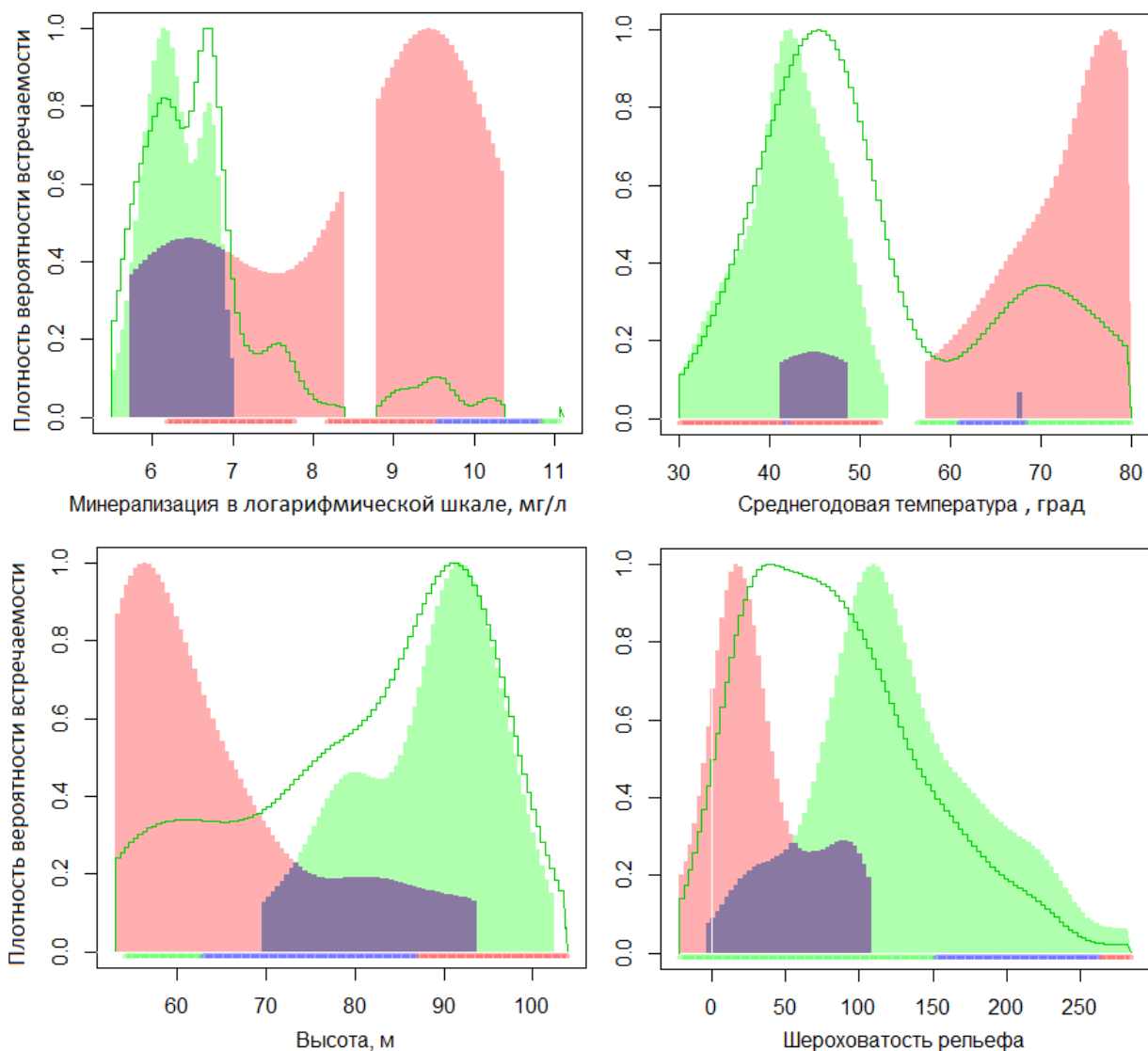
Дополнительную информацию о характере перераспределения видов между нишами получали путем разложения совокупной области на три категории (подобно известным диаграммам Венна): доля совместного использования (*expansion*) и доли пространства ниши, занимаемого каждым видом в отдельности (Di Cola et al., 2017).

```
ecospat.niche.dyn.index (grid.clim.g1, grid.clim.g2,  
                        intersection = 0.1)$dynamic.index.w  
  expansion  stability  unfilling  
0.43885990 0.56114010 0.01694598
```

Статистические тесты на эквивалентность и сходство каждой пары сравниваемых ниш показали с высоким уровнем значимости, что их различия не могут быть объяснены случайными причинами.

Дополнительно можно выяснить, какими конкретно переменными среды обусловлено различие экологических ниш. Построим графики плотности распределения по шкалам важнейших абиотических факторов встречаемости видов *Prodiamesa olivacea* (зеленым цветом) и *Berosus* sp. (оранжевым цветом), наиболее «контрастных» по условиям своего обитания. Синим цветом показана область совместной встречаемости. Зеленой контурной линией обозначена кривая плотности распределения для всех точек наблюдений

```
dfplot <- function(i, Nm)  
{  
  grid.clim.t.g1 <- ecospat.grid.clim.dyn(glob=as.data.frame(Env[,i]),  
    glob1=as.data.frame(Env[,i]),  
    sp=as.data.frame(Env[which(df_spe$ChPro.o==1),i]),  
    R=1000, th.sp=0)  
  grid.clim.t.g3 <- ecospat.grid.clim.dyn(glob=as.data.frame(Env[,i]),  
    glob1=as.data.frame(Env[,i]),  
    sp=as.data.frame(Env[which(df_spe$CoBer.sp==1),i]),  
    R=1000, th.sp=0)  
  ecospat.plot.niche.dyn(grid.clim.t.g1, grid.clim.t.g3, quant=0,  
    interest=2, name.axis1=Nm)  
}  
dfplot(i=3, Nm = "Минерализация в логарифмической шкале, мг/л")  
dfplot(i=6, Nm = "Среднегодовая температура, град ")  
dfplot(i=10, Nm = "Высота, м ")  
dfplot(i=11, Nm = "Шероховатость рельефа ")
```



## **5. Построение и анализ моделей распределения видов**

### *Ансамбль малых моделей (ESM - Ensemble of small models)*

Еще одной ключевой особенностью пакета `ecospat` является реализация стратегии ESM для моделирования распределения видов при наличии ограниченного числа точек встречаемости, как правило, менее 30 (например, в случае редких видов). ESM основан на создании ансамбля небольших моделей, каждая из которых использует две или более переменных и осуществляет свой частный прогноз. В конечном итоге совокупность прогнозов взвешивается по уровню компетентности каждой подмодели и вычисляется обобщенный результат. Считается, что формирование прогнозирующего коллектива позволяет снизить эффект переобучения и, соответственно, повысить точность всей модели.

Функции ESM пакета `ecospat` во многом интегрированы с пакетом `biomod2` (Thuiller et al., 2009) и позволяют выполнить подгонку коэффициентов серии небольших двумерных моделей различными методами (например, GLM, GAM, дискриминантный анализ, random Forrest, бустинг, Maxent и т.д.) с последующей их калибровкой. Для этого доступны различные индексы для количественной оценки производительности модели: AUC, TSS, Kappa, D Somers (= 2AUC – 1), Бойса и др. (Шитиков, Мастицкий, 2017). На втором этапе происходит усреднение прогноза ESM.

```

Env <- Env [,-1]
matr <- df_spe[, (6:9)]
matr[matr > 0] <- 1
xy <- df_spe[, (4:5)]
head(Env)
head(xy)
head(matr)
library(biomod2)
sp <- 1   ### Выполним моделирование распределения Prodiamesa olivacea
## Форматирование данных с использованием функции
# BIOMOD_FormatingData() пакета biomod2
myBiomodData <- BIOMOD_FormatingData( resp.var = as.numeric(matr[,sp]),
                                     expl.var = Env,
                                     resp.xy = xy,
                                     resp.name = colnames(matr)[sp])

### Калибровка простых двумерных моделей
my.ESM <- ecospat.ESM.Modeling( data=myBiomodData,
                               models=c('GLM','RF'),
                               NbRunEval=1,
                               DataSplit=70,
                               weighting.score=c("AUC"),
                               parallel=F)

```

Используемый перечень models ограничились двумя частными моделями из возможного списка 'GLM', 'GBM', 'GAM', 'СТА', 'ANN', 'SRE', 'FDA', 'MARS', 'RF', 'MAXENT.Phillips', 'MAXENT.Tsuruoka'. Для калибровки моделей выполняем только один проход (NbRunEval=1) разделения набора данных на обучающую (DataSplit=70) и экзаменующую последовательности.

```

### Оценка и объединение моделей в коллектив ESM
my.ESM_EF <- ecospat.ESM.EnsembleModeling(my.ESM,
                                          weighting.score = c("SomersD"), threshold=0)
## Веса простых двумерных моделей при их объединении их в коллектив ESM
my.ESM_EF$weights
## Проецирование в новое географическое пространство с использованием
## простых двумерных моделей
my.ESM_pred <- ecospat.ESM.Projection(ESM.modeling.output=my.ESM,
                                     new.env=Env)
### Проецирование в новое пространство с использованием коллектива ESM
my.ESM_EFpred <- ecospat.ESM.EnsembleProjection(
  ESM.prediction.output=my.ESM_pred,
  ESM.EnsembleModeling.output=my.ESM_EF)
## Оценки качества модели и коллектива ESMs
output.TH <- ecospat.ESM.threshold(my.ESM_EF, PEplot = TRUE)
output.TH

```

	model	sensitivity	specificity	Kappa	AUC	sensitivity.sd	specificity.sd	Kappa.sd
1	Full_GLM_ESM	0.8461538	0.8602151	0.6713836	0.8825476	0.05852969	0.03615262	0.06876368
2	Full_RF_ESM	1.0000000	0.9677419	0.9466019	0.9947615	0.00000000	0.01842065	0.03043355
3	Full_EF_ESM	0.9230769	0.9354839	0.8398058	0.9779432	0.04322704	0.02561288	0.05137792

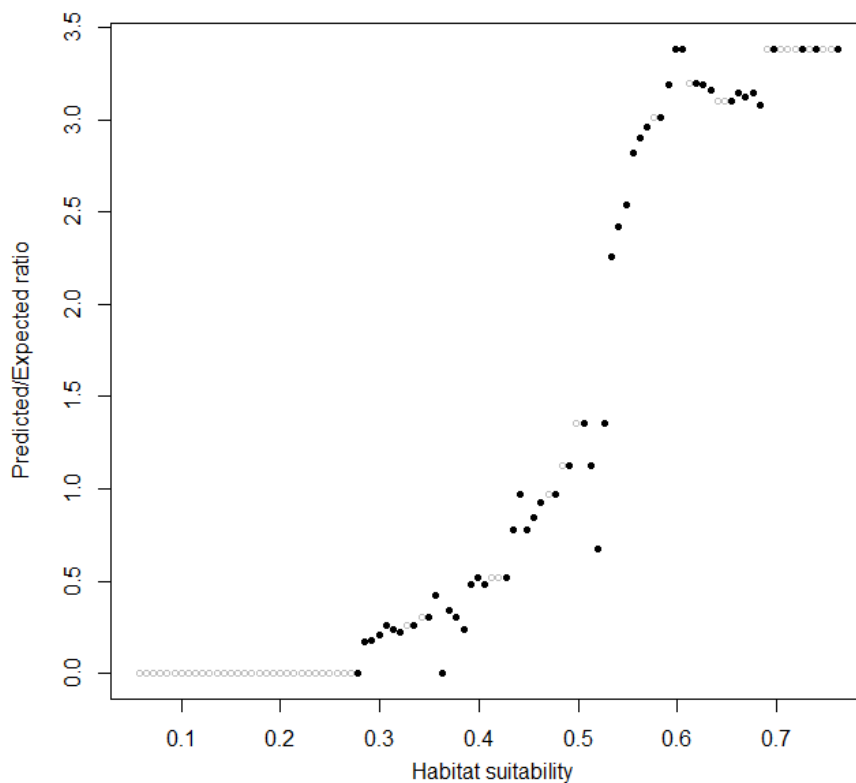
  

	AUC.sd	SomersD	Boyce	TSS	TSS.th	MPA1.0	MPA0.95	MPA0.90	Boyce.th.min	Boyce.th.max
1	0.03040430	0.7650951	NA	0.7063689	0.53	0.243	0.523	0.559	0.5109990	0.5109990
2	0.00419692	0.9895230	NA	0.9677419	0.38	0.400	0.520	0.591	0.3739505	0.3739505
3	0.01095091	0.9558864	NA	0.8585608	0.46	0.322	0.562	0.590	0.4805450	0.5232050

```

DATA <- as.data.frame (cbind(resp.var = my.ESM_EF$ESM.fit$resp.var,
                             pred = my.ESM_EF$ESM.fit[, "Full_EF_ESM"]/1000))
rownames(DATA) <- 1:nrow(my.ESM_EF$ESM.fit)
## Точки отсутствия вида, в которых прогноз обозначил наличие
Ind1 <- rownames(DATA[DATA$resp.var == 0 & DATA$pred >= 0.5232050,])
[1] "12" "67" "127"
## Точки наличия вида, в которых прогноз обозначил отсутствие
Ind2 <- rownames(DATA[DATA$resp.var == 1 & DATA$pred <= 0.4805450,])
[1] "41" "58" "95" "103" "105"

```



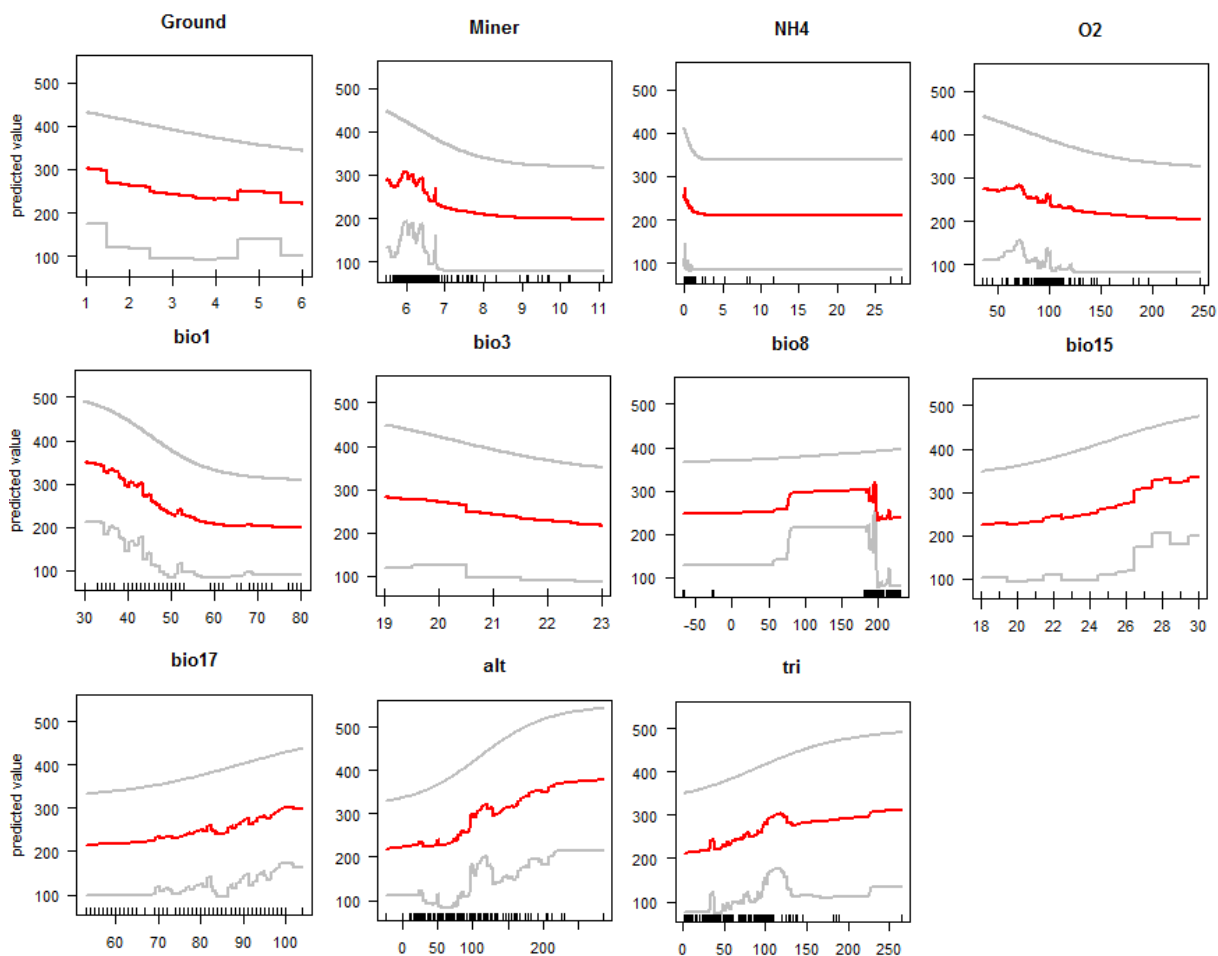
Объект, создаваемый функцией `ecospat.ESM.EnsembleModeling` включает вектор `fit` предсказанных значений отклика – пригодности среды обитания (*habitat suitability*) для точек исходной выборки. Две функции `...Projection` выполняют прогноз для произвольной области, заданной таблицей параметров среды. Одна из проблем анализа модели – как установить пороговое значение  $Z$ , обеспечивающее оптимальное классифицирование условий наличия/отсутствия вида. Разработчики пакета считают, что индекс Бойса (Boys) обеспечивает наиболее адекватную оценку порога в случаях, где точки отсутствия вида не несут вполне достоверной информации. Индекс Бойса отражает, насколько прогноз модели отличается от случайного распределения- наблюдаемых присутствий по интервалам отклика.

Обратим внимание, что на исходной совокупности прогноз частной моделью `random Forrest RF` по всей совокупности критериев оказался точнее, чем предсказание коллектива ESM. Остается только надеяться, что на "свежих" выборках эта ситуация изменится на противоположную.

Функции ESM пакета `ecospat` позволяют также оценить важность и характер влияния отдельных переменных среды на величину прогнозируемого отклика:

```
## относительная важность переменных среды в коллективе ESM
ecospat.ESM.VarContrib(my.ESM,my.ESM_EF)
      GLM      RF      ENS
Ground 0.09850391 0.08718873 0.09280562
Miner  0.09496770 0.10838209 0.10172315
NH4    0.06599796 0.06905701 0.06753849
O2     0.08527712 0.07157436 0.07837645
bio1   0.10183611 0.09759831 0.09970197
bio3   0.08762326 0.08616819 0.08689049
bio8   0.07626658 0.12273779 0.09966934
bio15  0.08911935 0.07950061 0.08427538
bio17  0.09292758 0.07882025 0.08582317
alt    0.11985719 0.10317730 0.11145725
tri    0.08762326 0.09579535 0.09173870
```

```
output.plot <- ecospat.ESM.responsePlot(ESM.EnsembleModeling.output =
my.ESM_EF, ESM.modeling.output = my.ESM)
```



### Ансамбль малых моделей (ESM - Ensemble of small models)

#### Пространственные прогнозы сообществ

В последние годы SDM все чаще используются для прогнозирования различных свойств сообщества, таких как видовое богатство, таксономический состав, функциональные и филогенетические закономерности, путем объединения прогнозов отдельных видов (S-SDMS, Дубюи и др. 2011, Калабрезе и др. 2014, Корд и др. 2014, Д'Амен и др. 2015с). Был предложен новый подход для улучшения прогнозов S-SDMS путем их увязки с прогнозами богатства (Guisan и Rahbek 2011, Д'Амен и др. 2015a, b).  $h$  является ограничением для прогноза S-SDM- исследования были предложены в качестве одного из шагов в рамках более масштабного "пространственно-явного моделирования совокупности видов" (SESAM). в дальнейшем он стремится реконструировать сообщества видов, реализуя серию последовательных расселений, местообитаний и биотических фильтров, а также - при необходимости - ранее описанные макроэкологические ограничения в качестве отдельных этапов моделирования (Guisan и Rahbek 2011).

SESAM, реализованный в ecospat, был протестирован на сегодняшний день путем применения трех последовательных шагов (D'Amen и др. 2015a, b): 1) расчет пригодности среды обитания для отдельных видов присутствия для каждой площадки - это, как правило, могут быть получены путем фитинга в СДМ для каждого вида (т. е. с-СДМ); этот этап представляет собой применение экологических интернет-фильтров на собрания общины; 2) вычислительное богатство прогноза-ний для каждого сайта -



богатство предсказания может быть получено в разные способы, например, путем суммирования вероятностей от отдельных видов предсказания для каждого сайта, либо фитинг прямой модели macroecological насыщенность (мэм); этот шаг предстает применением ограничения macroecological к среднее число видов, которые могут сосуществовать в рассматриваемой единице; 3) применение биотического правила для определения того, какие виды, потенциально присутствующие на участке, должны быть сохранены в окончательном прогнозе совокупности, чтобы соответствовать прогнозируемому значению богатства. последний шаг реализован в ecospat с помощью правила ранжирования вероятности (PRR), которое состоит в ранжировании видов в порядке убывания прогнозируемой вероятности присутствия (полученной из SDMs) и выборе видов из наиболее вероятных видов в списке до тех пор, пока сумма выбранных видов не достигнет ожидаемого значения богатства, полученного на шаге 2 (Д'Амен и др. 2015a, b).

### Список цитируемой литературы

*Franklin J.* Mapping Species Distributions: Spatial Inference and Prediction. Cambridge: Cambridge University Press, 2009. 320 p.

*Guisan A., Thuiller W., Zimmermann N.E.* Habitat Suitability and Distribution Models: With Applications in R. Cambridge: Cambridge University Press, 2017. 478 p.

*Peterson A.T., Soberón J., Pearson R.G., Anderson R.P., Martínez-Meyer E. et al.* Ecological Niches and Geographic Distributions (MPB-49). Princeton: Princeton Univ. Press, 2011. 328 p.

*Soberón J.* Grinnellian and Eltonian niches and geographic distributions of species. // *Ecol. Lett.* 2007. V. 10. P. 1115–1123.

*Peterson A.T., Soberón J.* Species distribution modeling and ecological niche modeling: Getting the concepts right. // *Nat. Conserv.* 2012. V. 10. P. 102–107.

*Brown J.H., Stevens G.C., Kaufman D.M.* The geographic range: Size, shape, boundaries, and internal structure // *Annual Review of Ecology and Systematics*, 1996. V. 27. P. 597–623.

*Gaston K.J.* The structure and dynamics of geographic ranges. Oxford University Press, Oxford, 2003. 266 p.

*Guisan A., Thuiller W.* Predicting species distribution: offering more than simple habitat models. // *Ecology Letters*. 2005. V. 8. P. 993–1009.

*Wallace A.R.* On the zoological geography of the Malay Archipelago. // *Proceedings of the Linnean Society of London*. 1860. V. 4. P. 172–184.

*Grinnell J.* Field tests of theories concerning distributional control. // *American Naturalist*. 1917. V. 51. P. 115–128.

*Grinnell J.* Geography and evolution. // *Ecology*. 1924. V. 5. P. 225–229.

*Elton C.* *Animal Ecology*. Sedgwick and Jackson, London, 1927.

*Hutchinson G.E.* Concluding remarks. // *Cold Spring Harb. Symp. Quant Biol*, 1957. V. 22. P. 415–427.

*Hutchinson G.E.* *An Introduction to Population Ecology*. Yale University Press, New Haven, 1978.

*Whittaker R.H., Levin S.A., Root R.B.* Niche, habitat, ecotope. // *Am. Nat.* 1973. V. 95. P. 321–338.

*MacArthur R.* The theory of the niche. / In: *Population Biology and Evolution* (eds K.C. Lewontin). Syracuse University Press, Syracuse, 1968. P. 159–176.

*Chase J.M., Leibold M.* *Ecological Niches: Linking Classical and Contemporary Approaches*. University of Chicago Press, Chicago and London, 2003. 212 p.

*Thuiller W., Lafourcade B., Engler R., Araújo M.B.* BIOMOD – a platform for ensemble forecasting of species distributions. // *Ecography*. 2009. V. 32. P. 369–373

Barve N., Barve V., Jiménez-Valverde A., Lira-Noriega A., Maher S., Peterson A.T., Soberón J., Villalobos F. The crucial role of the accessible area in ecological niche modeling and species distribution modeling // *Ecological Modelling*. 2011. V. 222(11). P. 1810-1819.

Ovaskainen O., Abrego N. *Species Distribution Modelling: With Applications in R*. Cambridge: Cambridge Univ. Press, 2020. 370 p.

Шутиков В.К., Зинченко Т.Д., Головатюк Л.В. Модели совместного распределения видов на примере донных сообществ малых рек волжского бассейна // *Журнал общей биологии*. 2021а. Т. 82, № 2, С. 143-154.

Phillips S.J., Anderson R.P., Schapire R.E. Maximum entropy modeling of species geographic distributions // *Ecol. Model.* 2006. V. 190. № 3–4. P. 231–259.

Лисовский А.А., Дудов С.В., Оболенская Е.В. Преимущества и ограничения использования методов экологического моделирования ареалов. 1. Общие подходы // *Журн. общ. биологии*. 2020. Т. 81. № 2. С. 123–134.

Шутиков В.К., Зинченко Т.Д., Головатюк Л.В. Модели максимальной энтропии и пространственное распределение донных сообществ Средней и Нижней Волги // *Российский журнал прикладной экологии*. 2021б. № 2, С. 10-16.

Hirzel A. H., Helfer V., Metral F. 2001. Assessing habitat-suitability models with a virtual species // *Ecol. Model.* V. 145. P. 111–121.

Norberg A., Abrego N., Blanchet F.G., Adler F.R., Anderson B.J. et al. A comprehensive evaluation of predictive performance of 33 species distribution models at species and community levels // *Ecol. Monogr.* 2019. V. 89. No 3. P. e01370.

Qiao H., Peterson A.T., Campbell L.P., Soberón J., Escobar L. NicheA: creating virtual species and ecological niches in multivariate environmental scenarios // *Ecography*. 2016. V. 39. P. 805–813.

Broennimann O., Fitzpatrick M.C., Pearman P.B., Petitpierre B., Pellissier L., Yoccoz N.G., Thuiller W., Fortin M.J., Randin C., Zimmermann N.E., Graham C.H., Guisan A. Measuring ecological niche overlap from occurrence and spatial environmental data // *Global Ecology and Biogeography*. 2012. V. 21. P. 481–497.

Di Cola V., Broennimann O., Petitpierre B., Breiner F.T., D'Amen M., Randin C., Engler R., Pottier J., Pio D., Dubuis A., Pellissier R.G., Mateo R.G., Hordijk W., Salamin N., Guisan A. Ecospat: an R package to support spatial analyses and modeling of species niches and distributions // *Ecography*. 2017. V. 40. P. 1–14.

Зинченко Т.Д. Эколого-фаунистическая характеристика хирономид (Diptera, Chironomidae) малых рек бассейна Средней и Нижней Волги (Атлас). Тольятти: Кассандра, 2011. 258 с.

Leroy B., Meynard C.N., Bellard C., Courchamp F. 2016. virtualspecies: an R package to generate virtual species distributions // *Ecography*. V. 39. P. 599-607

Warren D.L., Glor R.E., Turelli M. Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution // *Evolution*. 2008. V. 62. P. 2868–2883.

Gotelli N.J. Null model analysis of species co-occurrence patterns // *Ecology*. 2000. V. 81. P. 2606-2621.

Шутиков В. К., Мاستицкий С.Э. Классификация, регрессия и другие алгоритмы Data Mining с использованием R. Электронная книга. 2017. 351 с. URL: <https://stok1946.blogspot.com> (дата обращения 10.09.2021).

Шутиков В.К., Розенберг Г.С. Рандомизация и бутстреп: статистический анализ в биологии и экологии с использованием R.- Тольятти: Кассандра, 2014. 314 с.

D'Amen, M. et al. 2015a. Using species richness and functional traits predictions to constrain assemblage predictions from stacked species distribution models. – *J. Biogeogr.* 42: 1255 – 1266.

D'Amen, M. et al. 2015b. Predicting richness and composition in mountain insect communities at high resolution: a new test of the SESAM framework. – *Global Ecol. Biogeogr.* 24: 1443 – 1453.

D'Amen, M. et al. 2015c. Spatial predictions at the community level: from current approaches to future frameworks. – *Biol. Rev.* doi: 10.1111/brv.12222

*Golovatyuk L. V., Shitikov V. K., Zinchenko T. D.* Estimation of the Zonal Distribution of Species of Bottom Communities in Lowland Rivers of the Middle and Lower Volga Basin // *Biology Bulletin*. 2018. V. 45 (10). P. 1262–1268.

*Hijmans R.J., Cameron S.E., Parra J.L., Jones P.G., Jarvis A.* Very high resolution interpolated climate surfaces for global land areas // *International journal of Climatology*. 2005. V. 25. P. 1965-1978.

*Hirzel A.H., Hausser J., Chessel D., Perrin N.* (2002) Ecological-niche factor analysis: how to compute habitat suitability maps without absence data? *Ecology*, 83(7): 2027–2036

*Austin M., Belbin L., Meyers J., Doherty M., Luoto M.* Evaluation of statistical models used for predicting plant species distributions: role of artificial data and theory. *Ecological Modelling*. 2006. V. 199. P. 197–216.