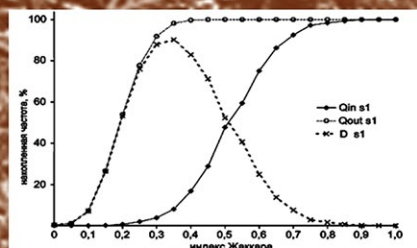
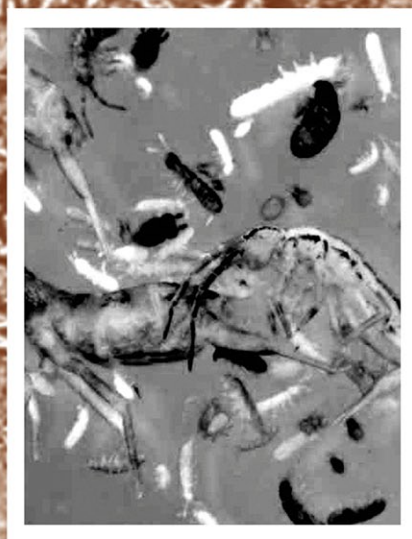


ЭТАЛОН СХОДСТВА: использование при сравнении состава и структуры сообществ



В.Н. Максимов
Н.А. Кузнецова



УДК 311.2:574

ББК 60.603+28.080.3

М17

Максимов В.Н., **Кузнецова Н.А.** Эталон сходства: использование при сравнении состава и структуры сообществ. М.: Товарищество научных изданий КМК. 2013. 89 с.

Предложен новый подход к решению проблемы адекватной оценки результатов при сравнении сообществ по составу и структуре с помощью традиционных индексов, для которых, как известно, отсутствуют статистические критерии. В качестве объекта рассмотрены сообщества почвенных коллембол, а также фитопланктона и макробентоса. Для примера выбраны широко используемые индексы: Жаккара (сравнение видового состава) и Шорыгина (сумма минимальных относительных обилий видов). Значения индексов сходства рассмотрены в связи с внутренней неоднородностью сообществ. В основе метода — переход от сравнения сообществ на уровне объединенных выборок (с одним значением индекса на выходе) к сравнению сообществ по данным отдельных проб (много значений индекса). Эмпирические распределения этих индексов для проб, взятых как из экологически различных, так и сходных сообществ, позволяют определить уровни значимости для принятия решения о сходстве их видового состава и структуры. Новый подход дает возможность обойтись без кластер-анализа и построения дендрограмм, порождающих разнообразие вариантов интерпретации данных. Для упрощения работы практикующего эколога разработан ускоренный метод создания эталона сходства, основанный на применении индекса фи-квадрат. Показано, как подобрать эталонную совокупность для использования этого индекса и сравнения сообществ как различных экосистем, так и анализа сезонных и межгодовых изменений сообществ в пределах одного биотопа.

Пособие предназначено для специалистов, работающих в области экологии сообществ.

*Работа была выполнена при финансовой поддержке РФФИ,
проект № 09-04-00328*

© Максимов В.Н., Кузнецова Н.А., 2013

© ИПЭЭ им. А.Н. Северцова РАН, 2013

© Товарищество научных изданий КМК,

издание, 2013

ISBN 978-5-87317-828-5

Содержание

Введение	3
Часть 1	8
1.1. Эталон сходства. Основные представления	8
1.2. Материалы и методы	11
1.3. Изменчивость индексов Жаккара и Шорыгина для эталонных совокупностей	14
1.4. Похвальное слово дендрограммам	22
Часть 2	24
2.1. Анализ матрицы сходства без построения дендрограмм	24
Часть 3	48
3.1. Ускоренный метод создания эталона сходства	48
3.2. Оценка эталона сходства по архивным данным на примере индекса фи-квадрат	52
Часть 4	62
4.1. Нужно ли суммировать исходные пробы при анализе сходства?	62
4.2. Что лучше — численность или биомасса? (небольшой, но поучительный пример)	73
Заключение	80
Литература	82
Приложение	83

ЧАСТЬ 1

1.1. Эталон сходства. Основные представления

В данной работе мы предлагаем способ решения данной проблемы, который можно назвать экспериментально-математическим, поскольку при его разработке мы не использовали строгие математические доказательства, а опирались на многолетний опыт сбора и анализа экологических данных. А именно: предлагается ввести понятие эталона сходства, при сравнении с которым для любой меры сходства или меры различия можно установить пределы, выход за которые следует считать указанием на отсутствие сходства или, соответственно, отсутствие различия. И если, скажем, таким пределом для индекса Жаккара окажется 0,45, то все пробы, у которых сходство по Жаккару меньше 45%, мы будем вправе считать принадлежащими сообществам, разным по видовому составу. Но при этом, выбрав вместо индекса Жаккара индекс Серенсена (по каким-то очень важным теоретическим соображениям!), мы заранее будем знать, что тот же вывод — об отличии видового состава — будет сделан, если индекс Серенсена окажется меньше 65%.

Конечно, из-за отсутствия четкого определения того, что такое сходство, вряд ли можно предложить эталон этого самого сходства, пригодный на все случаи жизни. Правда, в повседневном обиходе мы такой эталон, не задумываясь, применяем постоянно, говоря: «...похожи, как две капли воды», а для противоположного понятия такой же эталон предложил уже довольно давно А.С. Пушкин: «... стихи и проза, лёд и пламень не так различны меж собой...». Вряд ли, однако, эти эталоны могут послужить основой для метрологической проверки мер сходства и метрик, подобной проверке приборов для измерения физических величин. Зато для такой конкретной задачи, как сравнение видового состава сообществ, эталон сходства был предложен лет 25 тому назад, и мы позволим себе привести цитату из публикации, которая, к нашему сожалению, осталась не замеченной широкой научной общественностью: «...предположим, что у нас имеются две пробы, взятые в одно и то же время на одной и той же точке (станции). Теоретически следует ожидать, что видовой состав обеих проб должен быть полностью идентичен, а поэтому и значение индекса сходства, рассчитан-

ное для этих проб, должно быть в точности равно тому предельному значению данного индекса, которое он принимает при полном сходстве сравниваемых описаний. Однако очевидно, что вследствие случайных погрешностей при отборе проб, их обработке и при подсчете количества организмов и числа видов в пробе мы не получим абсолютного совпадения двух списков видов и их численностей для двух параллельных проб. Но тогда и рассчитанное эмпирическое значение индекса схождения не совпадёт с теоретически ожидаемым предельным значением... отклонение его значения от теоретически ожидаемого, связанное со случайными ошибками в определении видового состава, можно рассматривать, как своего рода статистическую погрешность.

Легко видеть, что любая совокупность параллельных проб может служить эталоном схождения, поскольку каждая из этих проб должна быть похожа на другую, «как две капли воды» (Максимов, 1984).

К сказанному уместно добавить некоторые рассуждения о содержании понятия «погрешность» в приложении к оценке числа видов и их обилия. Обычно в литературе, посвященной статистическим методам анализа экспериментальных данных, термины «ошибка», «погрешность» употребляются неизменно в сочетании со словом «измерение». Но измерение физической величины заключается в сравнении этой величины с некоторой величиной того же рода, принятой за единицу, а ошибкой измерения называют разницу между результатом измерения и истинным значением измеряемой величины. Это истинное значение неизвестно (иначе незачем было производить измерение), но любой экспериментатор твердо верит в то, что оно существует. Речь идет, конечно, о статике: динамические величины, т.е. скорости в полном соответствии с апорием Зенона не измеряют, а вычисляют, и у них нет «истинного значения», а есть либо «средняя скорость», либо «мгновенная скорость» (исключение составляет прямолинейное равномерное движение, которого в природе не бывает).

К сожалению, в отличие от физических величин, ни число особей в пробе, ни, тем более, число видов — никакого «истинного» значения не имеет. Да и сам процесс определения этих чисел — подсчет числа особей — нельзя назвать измерением в строгом смысле, поскольку в физической системе мер отсутствует такая единица измерения, как «штука» и более научнообразные термины «экземпляр» или «особь». Не случайно для такой величины, как плотность популяции, принимают размерность L^{-2} или L^{-3} . Положение усугубляется еще и тем, что процесс отбора пробы практически всегда производит нарушения в изучаемом объекте. Из-за этого, невозможно, скажем, взять повторную пробу почвы для определения количества населяющих ее беспоз-

воночных в точности в том же самом месте. В лучшем случае мы в состоянии отобрать пробу того же объема вплотную к той ямке, которая образовалась при отборе предыдущей. Нужно ли подробно объяснять, почему в двух соседних пробах число особей каждого вида будет различно, даже если мы ни разу не ошибемся при камеральной обработке проб, при подсчете особей, определении вида? Но если это понятно любому полевому экологу, правильно ли называть обнаруженные различия «ошибками измерения»?

Более серьезен 2-й вопрос: можно ли такие две пробы считать параллельными или повторными определениями (на привычном биологическом жаргоне — повторностями)? Не хотелось бы углубляться в дебри той дискуссии, которая недавно была реанимирована М.В. Козловым (2003), поэтому ограничимся чисто прагматическим суждением, справедливым, кстати, во всех случаях, когда подвергаются сомнению те или иные тонкости при использовании статистических методов в экспериментальных задачах: все зависит от цели конкретного экспериментального исследования. Если, например, мы отбираем несколько планктонных проб на одной и той же станции с борта неподвижного судна с целью оценить неравномерность (агрегированность, пятнистость) пространственного распределения планктона как раз в этот день и час именно на этой станции, то каждую из этих проб, несомненно, следует рассматривать, как независимую повторность. Если же цель исследования — получение сведений о сезонных изменениях в видовом составе и видовой структуре планктона в данном районе моря, то независимой повторностью будет, пожалуй, только серия проб, подобная первой, взятая на той же станции через определенный промежуток времени. Ясно, что все выводы, полученные после анализа собранного материала (с применением статистики или без оной), будут справедливы только для этой станции и для того периода времени, в пределах которого производили сбор материала.

Еще более серьезен 3-й вопрос: насколько список видов с оценками числа особей каждого из них (т.е. протокол обработки пробы) действительно характеризует видовой состав и структуру сообщества? Иначе говоря, насколько взятая нами проба представительна? Чтобы избежать бесплодных дискуссий и по этому поводу, договоримся, что выбор размера представительной пробы и необходимого количества таких проб осуществлен высококвалифицированными экологами с соблюдением всех рекомендаций, имеющихся в соответствующих методических руководствах. **Тогда, как мы и предлагаем, любая совокупность параллельных проб может служить эталоном сходства,**

поскольку каждая из этих проб должна быть похожа на другую, именно «как две капли воды», даже если это — пробы почвы, пробные площадки или содержимое дночерпателей.

Затруднение в использовании такого эталона в полевых экологических исследованиях связано с тем, что в них, в отличие от лабораторных экспериментов, параллельные пробы для определения видового состава сообщества отбирают крайне редко и лишь с какими-либо специальными целями. Вряд ли нужно подробно анализировать причины такого положения — трудоемкость камеральной обработки планктонных проб, определение видового состава фауны почвенных беспозвоночных, трудности отбора и анализа проб бентоса и т.п. слишком хорошо известны. В частности, в цитированной выше работе 1984 г. мы располагали только 8 параллельными пробами фитопланктона на каждом из двух створов исследованной реки. Поэтому полученные тогда результаты мы рассматривали, лишь как пример, иллюстрирующий **возможность** метрологической оценки не только индексов сходства, но и других характеристик экологических систем, таких, как индексы разнообразия, индексы устойчивости и т.п. В настоящее время мы располагаем гораздо более обширным материалом, который накоплен попутно с проведением обычных экологических исследований, а вовсе не в связи с какой-либо особой программой работ по экологической статистике или биометрической экологии.

1.2. Материалы и методы

Исходным материалом для исследования послужили результаты анализа видового состава коллембол в почвах и фитопланктона в морских водах. При всем очевидном различии объектов их объединяет примерно одинаковая статистическая надежность оценки обилия изучаемых организмов, так как и в том и в другом случае суммарное количество особей, подсчитанных в каждой единичной пробе, колебалось в пределах от нескольких десятков до 100–300 экземпляров, а средняя численность вида, определенная, как среднее геометрическое численности видов в пробе, усредненное по всем пробам на данной площадке, в большинстве случаев не превышала 5–6.

Видовой состав почвообитающих коллембол определяли в пробах, отобранных с помощью рамок 5×5 см на площадках размером 30×30 см. Количество особей каждого из обнаруженных видов подсчитывали после экстрагирования с помощью воронок Тулльгрена и фиксации по общепринятой методике (Гиляров, 1975). Для анализа были использованы совокупности по 36 проб, собранных на 5 пло-

щадках соснового леса на о. Силон в Дарвинском государственном заповеднике, характеристика которых дана в таблице 1.1.

Кроме того, в работе использованы материалы по коллемболам хвойных лесов, расположенных на территории Мордовского государственного заповедника (пойменная дубрава и ельник-черничник), юга Архангельской области (ельник-черничник, окрестности дер. Раменье Вельского р-на), Московской области (ельник-кисличник и зарастающий елями луг, Московская обл., стационар ИПЭЭ РАН имени А.Н. Северцова «Малинки»), а также Литвы (дубрава и

Таблица 1.1
Характеристика пробных площадок на о. Силон

Характеристика участков	Лишайник	Лишайник с пятном мха	Диффузная смесь лишайника и мха	Мох с пятном лишайника	Мох
Обозначение проб	L	LM	D	ML	M
Проективное покрытие лишайников, %	92	43	39	30	0
Проективное покрытие мхов, %	0	35	13	48	100
Мертвый покров, %	8	21	48	22	0
Видовой состав растений	<i>Cladina sylvatica</i>	<i>Cladina rangiferina</i> , <i>C. sylvatica</i> , <i>Dicranum undullatum</i>	<i>Cladina rangiferina</i> , <i>C. sylvatica</i> , <i>Pleurozium schreberi</i>	<i>Pleurozium schreberi</i> , <i>Dicranum undullatum</i> , <i>Cladina alpestris</i>	<i>Pleurozium schreberi</i>
Всего экз. коллембол	1612	2761	1641	2290	6646
Количество проб	33	36	36	36	35
Средняя численность вида в пробе	3,1	5,8	3,3	4,7	6,8
Среднее число видов, обнаруженных в пробе	26	26	20	21	21

Характеристика проб фитопланктона

Характеристика проб	Белое море (губа Чупа)			Карское море	
	июнь (подпробы)	август (подпробы)	август (пробы)	август (подпробы)	август (пробы)
Обозначение проб	ch51	ch35	ch50	k55	k20
Количество проб	51	35	50	55	20
Средняя численность вида в пробе	3,5	4,0	7,5	2,8	3,6
Среднее число видов, обнаруженных в пробе	10	12	19	12	30

ельник-черничник, окрестности г. Вильнюса). Особенности исследованных местообитаний и сроки учетов подробно описаны в книге (Кузнецова, 2005).

Пробы фитопланктона были отобраны в ходе экспедиционных работ на Белом и Карском морях сотрудниками кафедры общей экологии и гидробиологии МГУ более 30 лет тому назад. И в том и другом случае с борта судна, стоявшего на якоре, отбирали по 50 проб объемом 1 л из поверхностного горизонта, которые фиксировали раствором Люголя и концентрировали методом отстаивания. Видовой состав определяли в счетных камерах, просматривая по 5 камер из каждой параллельной пробы (Кольцова и др., 1971; Лихачёва и др., 1979). Некоторые характеристики совокупностей, использованных в анализе, приведены в таблице 1.2.

Из всего разнообразия индексов сходства мы выбрали для примера всего два. Один из них это уже упомянутый выше индекс Жаккара:

$$JCR = c/(a+b-c),$$

где a — число видов в списке А, b — число видов в списке В, а c — число видов, общих для обоих списков. Наверное, П. Жаккару и во сне не снилось, что лет 70 спустя хитроумные потомки обратят внимание на то, что этот индекс можно рассматривать, как меру пересечения двух множеств, одно из которых — список А, а другой — список В (Шитиков и др., 2005). Рискну предположить, что большинство геоботаников, до сих пор широко использующих индекс Жаккара (и не только геоботаников), делают это отнюдь не из уважения к теории множеств, а просто потому, что его смысл и без того совершенно ясен: это отношение

числа общих видов к суммарному числу видов в обоих списках. И, чтобы больше не возвращаться к этой теме, заметим, что благословение, исходящее от теории множеств, не спасло индекс Жаккара от разнообразной критики многочисленных авторов, предложивших позднее — на основе математически обоснованной, хотя и не всегда понятной, аргументации — другие индексы, смысл которых в некоторых случаях ничуть не менее прост и ясен без всяких математических премудростей.

Впрочем, наиболее часто критика индексов, подобных JCR, заключается в том, что они не учитывают различий в обилии видов, т.е. по современной компьютерной терминологии основаны на бинарных данных. Но, как было сказано выше, это лишь означает, что все такие меры близости оценивают **сходство видового состава** и не годятся для оценки **сходства видовой структуры** в сравниваемых пробах. Для этого существуют не менее многочисленные индексы, один из которых весьма популярен среди экологов. Он известен в отечественной литературе, как индекс Шорыгина (1939):

$$SHR = \Sigma \min(p_{i1}, p_{i2}),$$

где $\min(p_{i1}, p_{i2})$ — меньшее из двух относительных обилий i -го вида в сравниваемых пробах, $p_{ij} = n_{ij}/N_j$, если n_{ij} — численность i -го вида в j -й пробе, а $N_j = \Sigma n_{ij}$. Р. Уиттекер (1980), без ссылки на первоисточник, называет этот индекс «процентное сходство». Интересно, что, как заметил А.С. Константинов (1969), индекс Шорыгина, подобно индексу Жаккара, тоже можно рассматривать, как пересечение множеств. С нашей же точки зрения гораздо важнее то, что этот индекс можно легко рассчитать с помощью любого популярного статистического пакета, в котором есть программа кластер-анализа, предусматривающая среди прочих и расчет так называемого манхеттенского расстояния или метрики городского квартала (City Block Metric):

$$SHR = \Sigma |p_{i1} - p_{i2}|,$$

поскольку известно (Песенко, 1982), что $SHR = 1 - CBM/2$. Впрочем, не менее хорошо известно, что между индексами сходства и метриками, как мерами сходства или различия, нет принципиальной разницы с точки зрения того, какая из них более адекватна нашим представлениям о сходстве, как философской категории.

1.3. Изменчивость индексов Жаккара и Шорыгина для эталонных совокупностей

Для каждой из перечисленных выше совокупностей данных были рассчитаны матрицы индексов Жаккара и Шорыгина. Дополнительно, для проверки зависимости сходства проб от их размера, были со-