

ОБ ИСПОЛЬЗОВАНИИ БИОИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ В СРАВНИТЕЛЬНОЙ ФЛОРИСТИКЕ

І. СХЕМНО-ЦЕЛЕВОЙ ПОДХОД. АБСОЛЮТНЫЕ МЕРЫ СХОДСТВА И РАЗЛИЧИЯ

© Б.И. Сёмкин¹, А.П. Орешко¹, М.В. Горшков²

¹*Тихоокеанский институт географии ДВО РАН, г. Владивосток*

²*Тихоокеанский государственный экономический университет, г. Владивосток*

Работа содержит основу схемно-целевого подхода в сравнительной флористике. Представлено аксиоматическое обоснование абсолютных мер сходства и различия. Рассмотрен принцип воспроизводимости результатов анализа структуры флор и его связь с развитием биоинформатики.

Ключевые слова: сравнительная флористика, математические методы, списки видов растений, таксономические спектры, дескриптивные множества, дескриптивные наборы, биоинформационные технологии, базы данных, схемно-целевой подход.

В сравнительной флористике для сопоставления различных фитохорий наиболее часто используются теоретико-графовые методы, включающие расчёт абсолютных мер сходства (мер пересечения) и различия, относительных мер сходства и различия, ориентированных и неориентированных графов, дендрограмм и оптимальных деревьев (Сёмкин, 1973а,б; 1987а,б; 2007; Шмидт, 1980; Юрцев, Сёмкин, 1980; Шлотгауэр, Столовникова, 1988; Шлотгауэр, Варченко, 1990; Малышев, 1999; 2002; Малышев, Байков, Доронькин, 1998; Крестов, Баркалов, Таран, 2004; Крестов, 2006; Сёмкин, Орешко, Горшков, 2005; 2007). Происходит интенсивное использование различных баз данных, компьютерных методов обработки информации и биоинформационных систем в связи с изучением биохорологического разнообразия растительного покрова (Abbott, Bisby, Rogers, 1985; Петросян, Букварёва, Марин, 1993; Сёмкин, 1994; Сёмкин, Варченко, Орешко, 1995; Сёмкин, Киселёв, Бочарников, Краснопеов, 1995; Сёмкин, Тимофеев, Варченко, Орешко, 1997; Компьютерные..., 1995; 1997а,б; Зверев, 1997; 1998; Пяк, Зверев, 1997; Марина, Марин, 1998; Юрцев, 1998; Информационные..., 2001; Крестов, 2001; 2005; 2006; Малышев, 2002).

Представление конкретных флор множествами и определение с помощью их флористических отношений (общности, сходства, различия, «целого-части», включения, невключения) и отдельных фракций видов растений (общие виды двух флор, дифференциальные виды одной флоры относительно другой, дифференциальные виды двух флор и т.д.) имеет большое значение для разработки количественных показателей необходимых для сравнительного анализа видовых списков, а также даёт возможность строгого обоснования флористики в целом (Тамарин, Шмидт, 1975; Сёмкин, 1977; 1983а,б; Юрцев, Сёмкин, 1980; Юрцев, 1982; 1987; Юрцев, Камелин, 1987а,б; 1991).

С использованием флористических баз данных и компьютерных программ появилась возможность сравнивать фитохории рангов провинций и областей, при этом обрабатывать видовые списки сосудистых растений размером в несколько тысяч таксонов (Малышев, 2002; Крестов, 2005; 2006). Естественно, такие массивы флористических данных невозможно опубликовать в открытой печати. В связи с этим встаёт вопрос о *воспроизводимости* (в узком смысле этого слова) результатов, полученных при обработке этих данных, т.е. возникает проблема проверки достоверности полученных в работе закономерностей. При невозможности проверить достоверность полученных в статье закономерностей, т.е. при отсутствии в статье данных, на основе которых были произведены последующие расчёты различных матриц, построены графы и др. и, в конечном счёте, установлены закономерности, отпадает и необходимость публикации такой статьи. Выход из создавшегося положения возможен в двух случаях. В первом случае, необходимо создать биоинформационную

систему, содержащую первичные данные и всю необходимую информацию для проверки предложенных в статье методов обработки данных.

При налаживании «быстрых» информационных технологий можно произвести расчёты по первичным данным и установить их достоверность (воспроизводимость в широком смысле слова). Во втором случае, необходимо так сжать данные, чтобы их можно было привести в публикуемой работе и по ним произвести все необходимые расчёты (воспроизводимость в узком смысле слова). Проблема сжатия флористических данных для опубликования их в статье будет специально рассмотрена нами в дальнейшем.

Следует отметить, что в большинстве работ по сравнительной флористике используются различные показатели сходства или различия для обработки флористических данных и определения флористических отношений в виде графов, дендрограмм и др. Редко приводятся обзоры по использованию математических методов во флористике (Шмидт, 1980; Малышев, 1987; Сёмкин, 1987б) и почти отсутствуют работы, посвящённые методологическим процедурам при установлении флористических закономерностей. В связи с этим в статье рассматривается проблема схемно-целевого подхода к установлению закономерностей и вопрос сжатия данных для решения проблемы воспроизводимости результатов.

В приложении приведены математические разработки по дескриптивным наборам (множествам) и аксиоматике абсолютных мер сходства (пересечения) и различия. Предлагаемый математический аппарат необходим для строгого обоснования количественных методов в сравнительной флористике.

1. Схемно-целевой подход анализа отношений включения и сходства

Сущность схемно-целевого подхода к установлению эмпирических закономерностей состоит в следующем (Сёмкин, Усольцева, 1982; Сёмкин, 1983а; 1987а): исходя из поставленной цели исследования и содержательного обсуждения проблемы, производится сбор экспериментальных данных по определённой схеме, согласующейся с дальнейшей математической обработкой данных. Предполагается, что прежде чем приступать к сбору и обработке данных, необходимо обсудить ряд предварительных этапов научных исследований (Orloci, 1978):

1. Формулировку проблемы.
1. Выбор математической модели.

2. Определение существенных переменных.

3. Выбор плана взятия проб и метода анализа данных.

Только после обсуждения этих этапов может быть начат сбор данных и их анализ, за которыми следуют содержательные интерпретации. Указанные этапы последовательно зависимы, в связи с чем решения, принимаемые на каждом этапе, ограничивают свободу применения алгоритмов на последующих этапах (Crovello, 1970; Dale, 1970; Orloci, 1978; Сёмкин, 1981).

Приведём схему установления отношений включения и сходства, использующуюся при геоботанических и флористических исследованиях (Сёмкин, 1983а; 1987а). Схема включает 4 основных этапа исследования (рис. 1): выбор цели исследования, сбор, анализ данных и интерпретацию результатов..

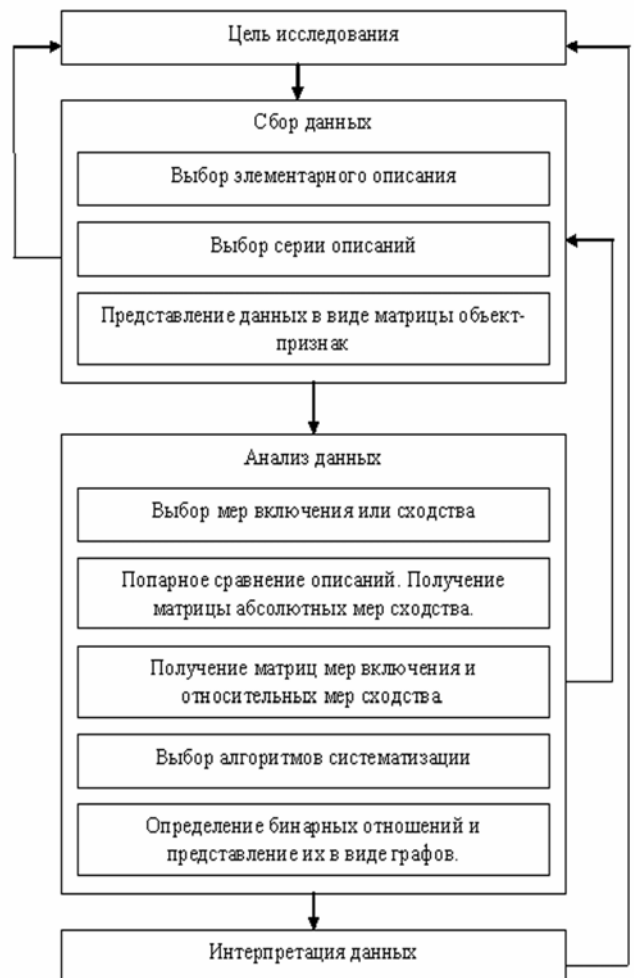


Рис. 1. Схема отношений включения и сходства (детерминистская схема)

Рассмотрим подробно указанные этапы исследования:

1. Выбор цели исследования – начальный и важный этап. Цели лучше всего формулируются теми исследователями, которые обладают большим опытом работы с конкретными объектами. Пока не существует принятых аргументов в защиту некоторого выбора цели. Однако важно чётко определить не всеобъемлющую цель. От выбора цели зависит выбор описаний объектов. Исходя из цели, определяют характерные масштабы признаков – пространственные и временные. Приведём некоторые цели исследования для сравнительной флористики:

- a) Анализ конкретных флор острова Сахалин;
- b) Анализ распределения натурализованных видов растений на острове Сахалин;
- c) Флористическое районирование острова Сахалин;
- d) Сравнительный анализ климатических зон по спектрам Раункиера;
- e) Сравнительный анализ высокогорных флор Чаткальского заповедника;
- f) Определение головной части семейственно-видового спектра для сравнения конкретных флор;
- g) Сравнение фитоохорий по семейственно-видовым спектрам.

2. Сбор данных. Для реализации выбранной цели необходимо собрать данные. Сначала производится выбор элементарного описания. Есть некоторые принципы, которым должны удовлетворять научные описания. Это принципы *воспроизводимости*¹ и *разделимости* описаний. Сущность этих принципов состоит в следующем. Первый принцип предполагает воспроизводимость описаний различными исследователями. Для этого следует чётко определить наименьшее количество признаков и их категорий, по которым производится описание объектов. Эти признаки и их категории должны быть доступны для восприятия многими исследователями. Затем выбирается элементарное описание, исходя из второго принципа – *разделимости*. Согласно этому принципу, элементарные описания одного и того же объекта должны быть более сходными, а описания существенно различных объектов различными. После выбора элементарного

описания, удовлетворяющего принципу воспроизводимости и делимости, производится серия описаний объектов по определённой схеме.

Для сравнимости результатов различных исследователей и их единообразия необходимо придерживаться определённых программ флористических исследований разной степени детальности (Юрцев, 1975; Юрцев, Камелин, 1987а,б; 1991).

3. Анализ данных. В результате выполнения процедуры сбора получаем матрицу данных «объект-признак», на основе которой производится анализ данных. При этом возникает проблема выбора мер включения и сходства, а также алгоритма систематизации данных. Меры включения и сходства есть функции от элементарных описаний, поэтому операции над описаниями должны быть согласованы с аналитическими выражениями этих функций. В свою очередь, в алгоритмах систематизации данных должны использоваться математические операции, которые не противоречат допустимым операциям над мерами включения и сходства. Проблема систематизации данных распадается на ряд последовательных задач. В некоторых случаях используются простые эвристические алгоритмы. Например, выбирается некоторый порог и все меры сходства (или включения) в матрице, имеющие численные значения большие или равные порогу, полагаются равными единице, а меньшие порога – равными нулю. В результате этой процедуры получается матрица смежности графа, которую можно интерпретировать как бинарное отношение на множестве описаний.

Наиболее часто для определения бинарных отношений используются алгоритмы оптимизации. При этом приходится решать три задачи: выбор и формулировка цели (выбор целевой функции), учёт ограничивающих факторов (учёт ограничителей) и реализация способа достижения цели при данных ограничениях.

4. Интерпретация результатов. В результате решения вышеуказанных задач получаем бинарные отношения, представляемые в виде графа и интерпретируемые исходя из содержания элементарных описаний и схемы сети связей. После содержательного анализа структур связей возможны некоторые дальнейшие уточнения целей исследования.

¹ Следует отметить, что требование получения одинаковых результатов различными исследователями и обеспечение их единообразия (Раменский, 1952 : 21) соответствует принципу воспроизводимости.

Традиционный подход к изучению отношений сходства в биологических науках состоит в том, чтобы сначала определить меры сходства, а затем исследовать взаимное расположение сходных объектов (Шрейдер, 1971). Для того чтобы ввести меры, измеряющие различные отношения, необходимо формально определить объекты, установить операции и отношения на них. Меры сходства, различия, включения определяются на формальных объектах (конечных множествах и дескриптивных наборах). Понятие «дескриптивный набор», отношения и операции пересечения и объединения двух и n наборов, а также различные меры на дескриптивных наборах широко используются в сравнительной флористике, геоботанике и экологии благодаря работам Б.И. Сёмкина (Сёмкин, 1973а; 2007; 2008). Проблема сравнительного анализа объектов, экспликация которых есть дескриптивные наборы, состоит в том, что отношения на этих объектах частичные и нетранзитивные, т.е. не все объекты сравнимы, и из сравнимости каждого из двух объектов с третьим ещё не следует их сравнимость между собой (нетранзитивность). В связи с этим приходится измерять степень сравнимости двух объектов и определять «наиболее сильные» связи для анализа взаимного расположения совокупности объектов. Для этих целей используется аксиоматический подход к введению первичных абстрактных понятий и разработка систем аксиом различных мер на совокупности первичных объектов. С помощью мер и задания пороговых величин строятся различные нетранзитивные (или даже транзитивные) отношения, содержательно интерпретируемые как эмпирические закономерности. Геометрически эти отношения изображаются в виде ориентированных и неориентированных графов, дендрограмм, оптимальных деревьев и др.

2. Дескриптивные множества и наборы

Математическое понятие «дескриптивное множество» (descriptive set) является экспликацией понятия «описания» (Сёмкин, 1973 а). Дескриптивное множество A определяется заданием весов $\mu_A(x_i) \geq 0$ для каждого элемента x_i ($i = 1, \dots, r$) множества X :

$$A = \left\{ \begin{array}{c} x_1, \dots, x_r \\ \mu_A(x_1), \dots, \mu_A(x_r) \end{array} \right\}.$$

Алгебраические операции и отношения для дескриптивных множеств определяются посредством введения операций и отношений над

весами – неотрицательными числами. В связи с этим, в случае фиксированного для определённого исследования элементов x_1, \dots, x_n множества X , можно использовать вместо дескриптивного множества эквивалентное понятие «дескриптивный набор», т.е. вектор, компонентами которого являются веса (Сёмкин, 1978 а, б; 1979; Сёмкин, Куликова, 1981; Сёмкин, Комарова, 1985). Основное требование, предъявляемое к дескриптивным наборам (см. приложение А) теорией измерений (Суппес, Зиннес, 1967) – однородность компонентов набора, т.е. каждая компонента набора должна быть измерена в одной и той же шкале отношений. Это свойство дескриптивных наборов позволяет находить сумму его компонентов.

Обозначая дескриптивный набор как $a = [\mu_A(x_1), \dots, \mu_A(x_r)]$ или $a = (a_1, \dots, a_r)$, свойство его однородности можно записать как $\lambda(a_1, \dots, a_r) = (\lambda a_1, \dots, \lambda a_r)$, $\lambda > 0$.

В геоботанике, экологии и флористике (Шенников, 1964; Василевич; 1969; Фёдоров, Гильманов, 1980) за веса обычно принимаются различные численные показатели: 0 и 1 – значения признака «присутствия вида на пробной площади или в фитоценозе»; число особей вида, приходящихся на единицу площади (численность); процент площади, занятой особями данного вида (покрытие); вес надземных частей вида на единицу площади или объёма (масса вида); число видов данной жизненной формы растений или их долей (процентов) видов в семействе или в более высоких таксонах; число видов географического элемента флоры и другие.²

Следует отметить формальные и содержательные различия обычных множеств и дескриптивных булевых наборов. Элементы обычных множеств *неупорядочены*, а компоненты булевых наборов упорядочены и их порядок обычно не изменяется в данном исследовании (это относится и к дескриптивным наборам с компонентами, принимающими неотрицательные числовые значения).

При объединении двух систем дескриптивных наборов (например, списков ви-

² Веса, измеренные в шкале порядка или в разных шкалах отношений, не могут составлять дескриптивный набор. Расчёт мер включения и сходства, как это делается в некоторых работах (Андреев, 1987), некорректен.

дов сосудистых растений двух флористических районов – Уссурийского и Даурского) необходимо заново упорядочить элементы объединённого дескриптивного набора и в дальнейших исследованиях этот порядок должен быть неизменным.

Множества типа $A = \{\text{солнце, разум, апельсин}\}$ которые любил приводить профессор Жегалкин (Шрейдер, 1971; Андреев, 1987), не удовлетворяют принципу однородности элементов и не могут быть представлены дескриптивным набором. Обычно, в сравнительной флористике при использовании таких понятий как «флористическое множество» (Камелин, 1987), «множество видов» (Юрцев, 1987; Юрцев, Камелин, 1987а,б) имеются в виду множества, элементы которых не произвольны, а организованы в целостные естественные совокупности.

3. Абсолютные меры сходства и различия

В сравнительной флористике (Юрцев, Сёмкин, 1980) подробно рассмотрены абсолютные и относительные меры сходства и различия двух множеств и двух дескриптивных (весовых) множеств, относительные меры включения и невключения для двух множеств. Для указанных мер приводится теоретико-множественная запись и даётся содержательная флористическая интерпретация. В дальнейшем мы будем сравнивать дескриптивные наборы, что позволяет одновременно объединить меры для случая множественной и дескриптивной интерпретации.

Абсолютная мера сходства двух множеств (Сёмкин, 1973 а; Юрцев, Сёмкин, 1980):

$$E(X, Y) = n(X \cap Y),$$

где $n(X \cap Y)$ – число общих элементов множеств X и Y легко обобщается на случай сравнения дескриптивных наборов и записывается в виде:

$$J(x, y) = v(x \wedge y),$$

где $v(x \wedge y)$ – мера конъюнкции двух дескриптивных наборов. Или в форме, удобной для её расчёта:

$$J(x, y) = \sum_{i=1}^r \min(x_i, y_i),$$

где x_i и y_i – веса первого и второго дескриптивного набора.

В приложении В приведена система аксиом, которой удовлетворяет абсолютная мера сходства двух дескриптивных наборов, а в приложении С показано, что в классе линейных мер сходства возможна только одна абсолютная мера

сходства (с точностью до постоянного множителя, определяющего размерность). Абсолютной мере сходства соответствует единственная мера различия (см. приложение С) (Юрцев, Сёмкин, 1980):

$$d(x, y) = v(x) + v(y) - 2v(x \wedge y),$$

где $v(x \wedge y)$, $v(x)$, $v(y)$ – вес конъюнкции двух дескриптивных наборов и веса наборов x и y соответственно. Для расчётов можно использовать меру различия в удобной форме:

$$d(x, y) = \sum_{i=1}^r |x_i - y_i|$$

Это хорошо известное в математике обобщённое расстояние Хэмминга. Использование абсолютных мер сходства, учитывающих не только присутствие, но и совместное отсутствие элементов в обоих сравниваемых множествах (Sokal, Sneath, 1963; Sneath, Sokal, 1973) ничего нового (кроме путаницы) не вносят в сравнительный анализ (Юрцев, Сёмкин, 1980), так как дуальная абсолютная мера различия не содержит совместно отсутствующие элементы и является коэквивалентной мере сходства, содержащей совместно отсутствующие элементы.

Легион абсолютных мер сходства содержащих совпадающие отсутствующие элементы (Sokal, Sneath, 1963; Sneath, Sokal, 1973) и построенные на их основе относительные меры сходства (или различия) заполнили все обзоры по использованию мер сходства в биологических исследованиях, оказываются ненужными и от них следует отказаться³.

Так как абсолютные меры различия, относительные меры сходства и различия, меры включения и невключения можно рассчитать на основе абсолютной меры сходства (включая веса сравниваемых дескриптивных наборов), то необходимо всегда (при отсутствии сводных списков или других первичных данных) приводить в публикациях мат-

³ При рассмотрении вероятностной и информационной интерпретации мер (Сёмкин, 1972) могут учитываться все клетки матрицы сопряжённости, однако некорректность, допущенная в работе (Sokal, Sneath, 1963), когда были объединены в одну формулу меры сходства и меры сопряжённости (совместимости) событий, позволила в последующих работах производить сравнение по отсутствующим элементам в описаниях.

рицу абсолютных мер сходства, а в общем случае матрицу мер конвергенций (Сёмкин, Горшков, 2008) (абсолютную меру совместимости событий в вероятностной интерпретации, абсолютную меру зависимости в информационной интерпретации, абсолютную меру пересечения площадей (Юрцев, 1982), абсолютную меру пересечения объёмов и другие). Следует отметить, что мы употребляем только понятие меры множества и считаем, что использование понятия мощности множества некоторыми исследователями (Константинов, 1969; Песенко, 1982) как аналога понятия меры множества некорректно, так как эти понятия эквивалентны только для конечных множеств.

На основе матрицы парных мер конвергенций, например с помощью редактора электронных таблиц Microsoft Excel, можно восстановить все результаты расчёта, приводимые в статье, если указаны соответствующие меры и алгоритмы для соответствующих объектов.

Матрица абсолютных мер конвергенций является основным документом, позволяющим воспроизвести результаты исследований при невозможности привести первичные данные из-за больших объёмов информации.

Заключение

Биоинформационные технологии в сравнительной флористике позволяют накапливать большие массивы данных о флорах различного ранга, производить обработку данных по определённой системе взаимосвязанных программ, изображать полученные закономерности графически и корректно составлять библиографию к работам по сравнительной флористике, используя информационно-справочную систему с комментариями.

Первостепенное значение в формировании биоинформационных систем имеет форма представления первичных данных и адекватное использование количественных показателей для сопоставления флор.

Нами предложено использовать элементарное представление флористических данных в виде дескриптивных множеств (наборов) и первичное сопоставление их посредством расчёта абсолютных мер сходства. Предложенная система аксиом мер конвергенции позволяет определить только меру сходства и на основе этой меры ввести относительные меры включения и сходства дескриптивных множеств.

ПРИЛОЖЕНИЯ

А. ДЕСКРИПТИВНЫЕ НАБОРЫ. ОТНОШЕНИЯ И ОПЕРАЦИИ НА НИХ

При определении отношений и операций на дескриптивных множествах обычно используются только численные значения весов.

В связи с этим вместо дескриптивных множеств можно использовать эквивалентное понятие – *дескриптивный набор*.

Обозначим дескриптивный набор как:

$$a = (\mu_A(x_1), \dots, \mu_A(x_r)) \text{ или} \\ a = (a_1, \dots, a_r).$$

На дескриптивных наборах задаётся отношение равенства и включения. Для каждой пары дескриптивных наборов $a = (a_1, \dots, a_r)$ и $b = (b_1, \dots, b_r)$ определим отношения равенства и включения следующим образом:

- $a = b$, тогда и только тогда, когда $a_i = b_i (i = 1, \dots, r)$;
- $a \leq b$, тогда и только тогда, когда $a_i \leq b_i (i = 1, \dots, r)$.

Алгебраические операции конъюнкции, дизъюнкции и разности двух дескриптивных наборов a и b определяются следующим образом:

$$a \wedge b = (\min(a_1, b_1), \dots, \min(a_r, b_r)); \\ a \vee b = (\max(a_1, b_1), \dots, \max(a_r, b_r)); \\ a \setminus b = (a_1 - \min(a_1, b_1), \dots, a_r - \min(a_r, b_r))$$

Вес дескриптивного набора $v(a)$ определяется как сумма весов:

$$v(a) = a_1 + a_2 + \dots + a_r.$$

Вес дескриптивного набора обладает следующими свойствами:

1. $v(\theta) = 0$, где $\theta = \underbrace{(0, \dots, 0)}_r$;
2. $v(a \wedge b) = \sum_{i=1}^r \min(a_i, b_i)$;
3. $v(a \vee b) = \sum_{i=1}^r \max(a_i, b_i)$;
4. $v(a \setminus b) = \sum_{i=1}^r a_i - \sum_{i=1}^r \min(a_i, b_i)$

В. ОПРЕДЕЛЕНИЕ КОНВЕРГЕНЦИИ ДВУХ ДЕСКРИПТИВНЫХ НАБОРОВ

Пусть E некоторое конечное множество дескриптивных наборов. Конвергенция в E есть отображение J произведения $E \times E$ в множество R действительных чисел, обладающее следующими свойствами:

1. $J(x, y) \geq 0, \forall x, y \in E$;
2. $J(x, y) = 0 \Leftrightarrow x \wedge y = \text{Ш}$;
3. $J(x, y) = J(y, x), \forall x, y \in E$;
4. $J(x, y) \leq J(x, x), \forall x, y \in E$;
5. $J(x, y) = J(x, x) \Leftrightarrow x \leq y$;
6. $J(\lambda x, \lambda y) = \lambda J(x, y), \lambda > 0$;
7. $J(x, y) + J(y, z) \leq J(x, z) + J(y, y), \forall x, y, z \in E$

С. АБСОЛЮТНЫЕ МЕРЫ СХОДСТВА И РАЗЛИЧИЯ

Легко проверить, что абсолютная мера сходства (мера пересечения) двух дескриптивных наборов $J(x, y) = \nu(x \wedge y) = c_{xy}$ удовлетворяет аксиомам мер конвергенций. c_{xy} может быть числом общих видов сравниваемых двух списков.

Приведём аксиомы для абсолютной меры сходства:

1. $E(x, y) \geq 0$;
2. $E(x, y) = E(y, x)$;
3. $E(x, y) = 0$ тогда и только тогда, когда $x \wedge y = \text{Ш}$;
4. $E(x, y) = E(a, b, c)$, где $a = \nu(x, x)$, $b = \nu(y, y)$, $c = \nu(x \wedge y)$ – меры конвергенции;
5. $E(\lambda a, \lambda b, \lambda c) = \lambda E(a, b, c)$, $\lambda > 0$.

Найдём все линейные функции мер сходства в общем виде:

$$E(x, y) = f(a, b, c) = \alpha_1 a + \alpha_2 b + \alpha_3 c + \alpha_4$$

где $\bar{\alpha}_1, \bar{\alpha}_2, \bar{\alpha}_3, \bar{\alpha}_4$ – постоянные коэффициенты.

Так как функция $f(a, b, c)$ должны удовлетворять аксиомам мер конвергенции, то для неё получены следующие соотношения:

1. $f(a, b, c) \geq 0$;
2. $f(\lambda a, \lambda b, \lambda c) = \lambda f(a, b, c)$;
3. $f(a, b, c) = f(b, a, c)$

$$4. \quad f(a, b, 0) = 0.$$

Эти свойства позволяют однозначно определять линейную функцию сходства. Действительно, из второго свойства следует, что $\bar{\alpha}_4 = 0$; из третьего свойства следует, что $\bar{\alpha}_1 = \bar{\alpha}_2$, из четвертого следует, что $\bar{\alpha}_1 = \bar{\alpha}_2 = 0$, из первого $\bar{\alpha}_3 > 0$.

Л и т е р а т у р а

Андреев В.Л. Анализ эколого-географических данных с использованием теории нечетких множеств. – Л., 1987. – 154 с.

Василевич В.И. Статистические методы в геоботанике. – Л., 1969. – 232 с.

Зверев А.А. Компьютерные информационные системы во флористических исследованиях // Состояние и перспективы развития Гербариев Сибири. Тез. докл. – Томск, 1997. – С. 23–25.

Зверев А.А. Сравнительный анализ флор с помощью компьютерной системы «IBIS» // Изучение биологического разнообразия методами сравнительной флористики: Матер. IV рабочего совещания по сравнительной флористике: Березинский биосферный заповедник, 1993. – СПб, 1998. – С. 284–288.

Информационные и телекоммуникационные ресурсы в зоологии и ботанике. – СПб, 2001. – 173 с.

Камелин Р.В. Процесс эволюции растений в природе и некоторые проблемы флористики // Теоретические и методические проблемы сравнительной флористики: Матер. II рабочего совещания по сравнительной флористике. Неринга, 1983. – Л, 1987. – С. 36–42.

Компьютерные базы данных в ботанических исследованиях: Тез. докл. II совещания (Санкт-Петербург, 17-19 апреля 1995 г.). – СПб, 1995. – 60 с.

Компьютерные базы данных в ботанических исследованиях: Сб. науч. тр. – СПб, 1997. – 113 с.

Компьютерные базы данных в ботанических исследованиях: Тез. докл. III совещания. – СПб, 1997. – 52 с.

Константинов А.С. Использование теории множеств в биогеографическом и экологическом анализе // Усп. соврем. биол. 1969. – Т. 67. – Вып. 1. – С. 99–108.

Крестов П.В. Использование флористической базы данных для фитогеографиче-

ских сравнений на Дальнем Востоке Азии // Информационные и телекоммуникационные ресурсы в зоологии и ботанике. — СПб, 2001. — С. 32.

Крестов П.В. Предложения к флористическому районированию северной Азии на основе сравнительного анализа флор на родовом уровне // Комаровские чтения. — Вып. 51. — Владивосток, 2005. — С. 15–56.

Крестов П.В. Растительный покров и фитогеографические линии Северной Пацифики: Автореф. дисс. ... д-ра биол. наук. — Владивосток, 2006. — 42 с.

Крестов П.В., Баркалов В.Ю., Таран А.А. Ботанико-географическое районирование острова Сахалин // Растительный и животный мир острова Сахалин (Материалы Международного сахалинского проекта). Ч.1. — Владивосток, 2004. — С. 67–90.

Мальшев Л.И. Основы флористического районирования // Бот. журн. 1999. — Т. 84. №. — С. 3–14.

Мальшев Л.И. Моделирование флористического деления Европы с помощью кластерного анализа // Бот. журн. 2002. — Т. 87. №. — С. 16–33.

Мальшев Л.И., Байков В.М., Доронькин В.М. Пространственное разнообразие родовой структуры во флоре Сибири // Изучение биологического разнообразия методами сравнительной флористики: Матер. IV рабочего совещания по сравнительной флористике: Березинский биосферный заповедник, 1993. — СПб, 1998. — С. 34–44.

Марина Л.В., Марин Ю.Ф. Пример использования информационной системы Висимского заповедника в ботанических исследованиях // Изучение биологического разнообразия методами сравнительной флористики: Матер. IV рабочего совещания по сравнительной флористике: Березинский биосферный заповедник, 1993. — СПб, 1998. — С. 273–284.

Песенко Ю.А. Принципы и методы количественного анализа в фаунистических исследованиях. — М., 1982. — 287 с.

Петросян В.Г., Букварёва Е.Н., Марин Ю.Ф. Интерактивная информационная система Висимского заповедника // Экология. 1993. №. — С. 79–81.

Пяк А.И., Зверев А.А. Опыт сравнительного анализа локальных флор с помощью прикладного статистического пакета Biostat // Бот. журн. 1997. — Т. 82. №. — С. 64–75.

Раменский Л.Г. О некоторых принципиальных положениях современной геоботаники // Бот. журн. 1952. — Т. 37. №. — С. 181–202.

Сёмкин Б.И. Общие принципы введения мер различия, сходства и разнообразия в биоценологии // Принципы и методы экспериментального изучения растительных сообществ: Тез. докл. и выступлений. — Л., 1972. — С. 12–16.

Сёмкин Б.И. Дескриптивные множества и их приложения // Исследование систем. Т. 1. Анализ сложных систем. — Владивосток, 1973а. — С. 83–94.

Сёмкин Б.И. О теоретико-множественных методах изучения растительных сообществ // Тез. докл. V делегатского съезда Всесоюз. бот. о-ва. — Киев, 1973б. — С. 210–211.

Сёмкин Б.И. Анализ структур фитocenотических данных: Автореф. дисс. ... д-ра биол. наук. — Тарту, 1977. — 49 с.

Сёмкин Б.И. Основы методов систематизации географических данных // Математические методы в экологии и географии. — Владивосток, 1978а. — С. 3–11.

Сёмкин Б.И. Принципы автоматической систематизации биоценотических описаний // Третья Всесоюзная конференция по биол. и мед. кибернетике. Т. 3. — М., 1978б. — С. 325–328.

Сёмкин Б.И. Эквивалентность мер близости и иерархическая классификация многомерных данных // Иерархические классификационные построения в географической экологии и систематике. — Владивосток, 1979. — С. 97–112.

Сёмкин Б.И. Методологические принципы исследования сложных систем в экологии и географии // Математическое моделирование природных систем. — Владивосток, 1981. — С. 5–9.

Сёмкин Б.И. Предисловие // Теоретико-графовые методы в биогеографических исследованиях. — Владивосток, 1983а. — С. 3–5.

Сёмкин Б.И. О постановке задач в структурной биоценологии // Теоретико-графовые методы в биогеографических исследованиях. — Владивосток, 1983б. — С. 5–12.

Сёмкин Б.И. Сравнительный многомерный анализ структурной организации

растительного покрова: Автореф. дисс. ... д-ра биол. наук. — Л., 1987а. — 34 с.

Сёмкин Б.И. Теоретико-графовые методы в сравнительной флористике // Теоретические и методические проблемы сравнительной флористики: Матер. II рабочего совещания по сравнительной флористике. Неринга, 1983. — Л., 1987б. — С. 149–163.

Сёмкин Б.И. Об эколого-географической структуре фиторазнообразия Дальнего Востока России // Природоохранные территории и акватории Дальнего Востока и проблемы сохранения биологического разнообразия: Матер. 2-ой науч. конф., посвящённой 60-летию со дня организации Уссурийского заповедника. — Владивосток, 1994. — С. 32–33.

Сёмкин Б.И. Количественные показатели для оценки односторонних флористических связей, предложенных Б.А. Юрцевым // Бот. журн. 2007. — Т. 92. № — С. 114–127.

Сёмкин Б.И. К методике анализа разновеликих множеств в сравнительной флористике // Комаровские чтения. — Вып. LVI. 2008. — С. 179–195.

Сёмкин Б.И., Горшков М.В. Система аксиом симметричных функций двух переменных и меры, измеряющие отношения сходства различия, совместимости и зависимости для компонентов биоразнообразия // Вестник ТГЭУ. 2008. № — С. 31–46.

Сёмкин Б.И., Комарова Т.А. Использование мер включения при изучении вторичных сукцессий (на примере послепожарных сообществ Южного Сихотэ-Алиня) // Бот. журн. 1985. — Т. 70. № — С. 89–97.

Сёмкин Б.И., Усольцева Л.А. Об использовании метода анализа разнообразий при исследовании взаимосвязи растительности и среды // Локальный мониторинг растительного покрова. — Владивосток, 1982. — С. 54–60.

Сёмкин Б.И., Куликова Л.С. Методика математического анализа списка видов насекомых в естественных и культурных биоценозах. — Владивосток, 1981. — 73 с.

Сёмкин Б.И., Варченко Л.И., Орешко А.П. Теоретические и методические подходы к учёту и охране биологического разнообразия Дальнего Востока с помощью баз данных // Компьютерные базы данных в ботанических исследованиях: Тез. докл. II совещания (Санкт-Петербург, 17–19 апреля 1995 г.). — СПб, 1995. — С. 52.

Сёмкин Б.И., Киселёв А.Н., Бочарников В.Н., Краснопеев С.М. Бассейновый подход в оценке биоразнообразия на основе ГИС-

технологий // Компьютерные базы данных в ботанических исследованиях: Тез. докл. II совещания (Санкт-Петербург, 17–19 апреля 1995 г.). — СПб, 1995. — С. 53.

Сёмкин Б.И., Орешко А.П., Горшков М.В. Биоинформационная система (БИС) «Биоразнообразии» и её использование // Матер. Всероссийской научно-практической конференции «Информационные технологии в науке, образовании и экономике». — Якутск, 2005. — С. 16–18.

Сёмкин Б.И., Орешко А.П., Горшков М.В. Биоинформационная система "Биоразнообразии" и её использование в научных и учебных целях // Матер. XIII научного совещания географов Сибири и Дальнего Востока (27–29 ноября 2007 г). Т.1. — Иркутск, 2007. — С. 195.

Сёмкин Б.И., Тимофеев И.В., Варченко Л.И., Орешко А.П. Сравнительный анализ флоры с использованием баз данных // Компьютерные базы данных в ботанических исследованиях. — СПб, 1997. — С. 90–93.

Супнес П., Зиннес Д. Основы теории измерений // Психологические измерения. — М., 1967. — С. 9–110.

Фёдоров В.Д., Гильманов Т.Г. Экология. — М., 1980. — 464 с.

Шенников А.П. Введение в геоботанику. — Л., 1964. — 447 с.

Шлотгауэр С.Д., Варченко Л.И. Сравнительный анализ конкретных флор высокогорий Западного Приохотья и Северного Приамурья: Препр. — Хабаровск, 1990. — 101 с.

Шлотгауэр С.Д., Столовникова М.А. Сравнительный анализ разновеликих флор Приохотья: Препр. — Хабаровск, 1988. — 101 с.

Шмидт В.М. Статистические методы в сравнительной флористике. — Л., 1980. — 176 с.

Шрейдер Ю.А. Равенство сходство порядок. — М., 1971. — 255 с.

Юрцев Б.А. Координация исследовательских программ разной интенсивности как подход к комплексному изучению биосферы // Изв. АН СССР. — Сер. биол. 1975. № — С. 618–623.

Юрцев Б.А. Флора как природная система // Бюлл. МОИП. Отд. биол. 1982. — Т. 87. — Вып. 4. — С. 3–22.

Юрцев Б.А. Флора как базовое понятие флористики: содержание понятия, подходы к изучению // Теоретические и методи-

ческие проблемы сравнительной флористики: Матер. II рабочего совещания по сравнительной флористике. Неринга, 1983. — Л., 1987. — С. 13–28.

Юрцев Б.А. Изучение и сохранение биологического разнообразия: вклад флористики // Изучение биологического разнообразия методами сравнительной флористики: Матер. IV рабочего совещания по сравнительной флористике: Березинский биосферный заповедник, 1993. — СПб, 1998. — С. 14–34.

Юрцев Б.А., Камелин Р.В. Программа флористических исследований разной степени детальности // Теоретические и методические проблемы сравнительной флористики: Матер. II рабочего совещания по сравнительной флористике. Неринга, 1983. — Л., 1987а. — С. 219–242.

Юрцев Б.А., Камелин Р.В. Очерк системы основных понятий флористики // Теоретические и методологические проблемы сравнительной флористики. — Л., 1987б. — С. 242–266.

Юрцев Б.А., Камелин Р.В. Основные термины и понятия флористики. — Пермь, 1991. — 80 с.

Юрцев Б.А., Семкин Б.И. Изучение конкретных и парциальных флор с помощью математических методов // Бот. журн. 1980. — Т. 65. № 12. — С. 1706–1718.

Abbott L.A., Bisby F.A., Rogers D.J. Taxonomic Analysis. Biology: computers, models and databases. — New York, 1985. — 336 p.

Crovello T.J. Analysis of character variation in ecology and systematics // Annu. Rev. Ecol. and Syst. 1970. — V. 58. — P. 55–98.

Dale M.B. On property structure, numerical taxonomy and data handing // Modern methods in plant taxonomy. — London, 1970. — P. 185–197.

Orloci L. Multivariate analysis in vegetation research. — Boston: University of Western Ontario. 2 ed., 1978. — 451 p.

Sneath P.H.A., Sokal R.R. Numerical taxonomy: The principles and practices of numerical classification. — San-Francisco, 1973. — 573 p.

Sokal R.R., Sneath P.H.A. Principles of numerical taxonomy. — San Francisco—London, 1963. — 359 p.

On the using bioinformatiоning technology in comparative floristic studies. I. The schematic purposeful approach. Absolute measures of similarity and dissimilarity

B.I. Semkin, A.P. Oreshko, M.V. Gorshkov

Pacific Institute of Geography FEB RAS, Vladivostok

Pacific State University of Economics, Vladivostok

Key words: comparative floristic analysis, mathematical methods, plant species lists, taxonomical spectra, descriptive sets, bioinformatiоning technologies, data bases, schematic purposeful approach.

The article contains a basis of the schematic purposeful approach in comparative floristics. The axiomatic basis of absolute measures of similarity and dissimilarity is presented. The principle of reproducibility of results of the analysis flora and its relation with development of bioinformatics is considered.

Ill. 1. Bibl. 67.