

БОТАНИЧЕСКАЯ ИНФОРМАТИКА

УДК 581.9:58.08:51

ОБ ИСПОЛЬЗОВАНИИ БИОИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ В СРАВНИТЕЛЬНОЙ ФЛОРИСТИКЕ

III. ОТНОСИТЕЛЬНЫЕ МЕРЫ СХОДСТВА И РАЗЛИЧИЯ ДЕСКРИПТИВНЫХ МНОЖЕСТВ

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

¹Тихоокеанский институт географии ДВО РАН, г. Владивосток

² Тихоокеанский государственный экономический университет, г. Владивосток

Работа содержит аксиоматическое обоснование мер сходства и мер различия для дескриптивных множеств (дескриптивных наборов). Рассматривается проблема согласования мер включения и мер сходства. Предлагается отдавать предпочтение одной мере процентного сходства (различия), которая в частном случае совпадает с мерой Браун-Бланке и двойственной ей мерой различия (расстоянием Юрцева). Рассмотрены методы систематизации флористических данных: метод построения оптимального дерева, метод диагонализации (метод диаграмм Чекановского), метод построения графов сходства, включения-сходства (с использованием пороговых величин) и метод построения дендрограмм.

Ключевые слова: сравнительная флористика, математические методы, меры сходства и различия, списки видов растений, таксономические спектры, дескриптивные множества, дескриптивные наборы, биоинформационные технологии, согласование мер, свойство двойственности мер, оптимальное дерево, диагонализация матрицы, диаграмма Чекановского, граф сходства, граф включения-сходства, дендрограмма.

Дуализм понятий «сходство» и «различие» интерпретируется следующим образом: «чем больше сходство объектов, тем меньше их различие» и даёт возможность рассматривать меры сходства и меры различия как двойственные друг другу и использовать их в равной мере для анализа отношений сходства и различия на флористических данных. Следовательно, сходство и различие объектов может быть выражено только *одной* мерой сходства или мерой различия. Определение одного коэффициента сходства-различия путём взятия частного от меры сходства и меры различия (Kulczynski, 1927; Песенко, 1982) или обратного отношения (Ekman, 1940), а также разности меры сходства и меры различия (Stugren, Rădulescu, 1961) или разности меры различия и меры сходства (Мальшев, 1972) некорректно и далее нами рассматриваться не будет.

Принцип дуализма или двойственности мер сходства и мер различия позволяет формулировать двой-

ственные свойства и доказывать справедливость двойственных предложений (теорем). Например, все аксиомы мер сходства являются двойственными и в отношении к мерам различия. Следовательно, можно формулировать аксиомы только для мер сходства, а затем через них определять двойственные свойства мер различия или наоборот.

1. ОТНОСИТЕЛЬНЫЕ МЕРЫ СХОДСТВА ДЕСКРИПТИВНЫХ МНОЖЕСТВ (ДЕСКРИПТИВНЫХ НАБОРОВ)

Рассмотрим общепринятые свойства относительных мер сходства двух дескриптивных наборов, которые послужили экспликацией для системы аксиом (Приложение A_1):

1. Значение мер сходства должны быть положительными и ограниченными, т.е. изменяться в интервале $[0, 1]$;
2. Мера сходства принимает минимальное значение равное 0 при отсутствии общих признаков у сравниваемых объектов;
3. Мера сходства принимает максимальное значение равное 1 при полном совпадении признаков у сравниваемых объектов;
4. Мера сходства симметрична;
5. Мера сходства измеряется в абсолютной шкале, т.е. является однородной функцией нулевой степени относительно двух переменных дескриптивных наборов;
6. Относительная мера различия определяется как разность между 1 и относительной мерой сходства (см. Приложение A_2).

Наиболее часто в сравнительной флористике используется коэффициент Сёрнсена

(Braun-Blanquet, 1951; Dagnelie, 1960; Greig-Smith, 1964; Макфедьен, 1965; Williams, Lance, 1965; Češka, 1966; Roux, Roux, 1967; Василевич, 1969; Шеляг-Сосонко, Куквица, 1969; Gounot, 1969; Макаревич, 1971; Заки, Шмидт, 1972; Красилов, 1972; Romaniszyn, 1972; Воронов, 1973; Goodall, 1973; Кожевников, 1974; Mueller-Dombois, Ellenberg, 1974; Дажо, 1975; Одум, 1975; Тамарин, Шмидт, 1975; Чернов, 1975; Трасс, 1976; Goldsmith, Harrison, 1976; Семкин, Комарова, 1977; Абакумов, Филиппова, 1978; Быков, 1978; Захаров, 1978; Миркин, Розенберг, 1978; Dzwonko, 1978; Семкин, 1979; Норин, 1979; Сергиенко, 1979; Уиттекер, 1980; Шмидт, 1980; Юрцев, Сёмкин, 1980; Юрцев, Секретарёва, 1983; Баркалов, 1981; Гусарова, Семкин, 1982; Марина, 1982; 1983; Несис, 1982; Песенко, 1982; Пийн, 1982; Суханов, 1982; Зайцев, 1984; Кафанов, Жуков, Федотов, 1984; Пушкарь, Короткий, Семкин, 1984; Уланова, 1995) и коэффициент Жаккара (Jaccard, 1901; 1902; 1908; 1912; 1928; Harper, 1915; Алёхин, 1925; Беклемишев, 1931; 1961; 1970; Сукачев, 1931; Быков, 1953; 1957; 1973; 1978; 1983; Жадовский, 1925; Sokal, Sneath, 1963; Шенников, 1964; Greig-Smith, 1964; Hagmeier, Stults, 1964; Кац, Кац, Кипиани, 1965; Češka, 1966; Hagmeier, 1966; Maddocks, 1966; Valentine, 1966; Грейг-Смит, 1967; Макаревич, 1967; 1971; Roux, Roux, 1967; Mello, Buzas, 1968; Александрова, 1969; Василевич, 1969; Gounot, 1969; Ниценко, 1971; Levandovsky, Winter, 1971; Красилов, 1972; Сёмкин, 1972 а, б; 1978; Воронов, 1973; Goodall, 1973; Sneath, Sokal, 1973; Толмачев, 1974; Mueller-Dombois, Ellenberg, 1974; Дажо, 1975; Лашинский, Рогинская, 1975; Тамарин, Шмидт, 1975; Чернов, 1975; Шаткаускас, 1975; Леме, 1976; Трасс, 1976; Галанин, 1977; 1979; 1980; Ребристая, 1977; Андреев, 1979, 1980; Норин, 1979; Сергиенко, 1979; Неронов, Лушекина, 1980; Шмидт, 1980; Юрцев, Сёмкин, 1980; Сапегин, 1981; Кусакин, Растанов, 1982; Песенко, 1982; Алексашина, Штина, 1984; Уланова, 1995).

Упомянутые коэффициенты входят в семейство мер сходства, упорядоченных параметром τ (Приложение А₂):¹

$$K_{\tau}(a, b) = \frac{K_0(a, b)}{1 + \tau - \tau K_0(a, b)},$$

где $a = (a_1, \dots, a_r)$, $b = (b_1, \dots, b_r)$ – дескриптивные

наборы;

$$K_0(a, b) = \frac{2 \sum_{i=1}^r \min(a_i, b_i)}{\sum_{i=1}^r a_i + \sum_{i=1}^r b_i}.$$

Меры сходства K_{τ} эквивалентны (Сёмкин, Двойченков, 1973). При $\tau=1$ получаем меру сходства:

$$K_1(a, b) = \frac{\sum_{i=1}^r \min(a_i, b_i)}{\sum_{i=1}^r \max(a_i, b_i)} = \frac{\sum_{i=1}^r \min(a_i, b_i)}{\sum_{i=1}^r a_i + \sum_{i=1}^r b_i - \sum_{i=1}^r \min(a_i, b_i)}.$$

¹ График зависимости $K_{\tau}(a, b)$ от $K_0(a, b)$ будет аналогичен графику $K_{\tau}(a; b)$ от $K_0(a; b)$ (Сёмкин, Орешко, Горшков, 2008 б).

Совокупность мер сходства, встречающихся в сравнительной флористике, геоботанике и экологии упорядочиваются с помощью двух параметров τ и η (Сёмкин, 1979; 2007):

$$K_{\tau; \eta}(A, B) = \left[\frac{K_{\tau}^{\eta}(A; B) + K_{\tau}^{\eta}(B; A)}{2} \right]^{1/\eta},$$

где
$$K_{\tau}(B; A) = \frac{n(A \cap B)}{(1 + \tau)n(A) - \tau n(A \cap B)},$$

$$K_{\tau}(A; B) = \frac{n(A \cap B)}{(1 + \tau)n(B) - \tau n(A \cap B)},$$

$$-\infty < \eta < +\infty, \quad -1 < \tau < \infty.$$

Для булевых наборов, имеющих компоненты равные единице и нулю, меры сходства $K_{1; -1}$ и $K_{0; -1}$ совпадают с коэффициентом Жаккара и коэффициентом Сёренсена соответственно.

2. МЕРЫ РАЗЛИЧИЯ ДЕСКРИПТИВНЫХ НАБОРОВ

Мера различия $F(a, b)$ определяется как дополнение меры сходства до 1, т.е.:

$$F(a, b) = 1 - K(a, b).$$

Мера различия $F_1(a, b)$, двойственная мере сходства Жаккара $K_{1; -1}(a, b)$ и мера различия $F_{0; -1}(a, b)$, двойственная мере сходства Сёренсена $K_{0; -1}(a, b)$ соответственно равны:

$$F_{1; -1}(a, b) = \frac{\sum_{i=1}^r a_i + \sum_{i=1}^r b_i - 2 \sum_{i=1}^r \min(a_i, b_i)}{\sum_{i=1}^r a_i + \sum_{i=1}^r b_i - \sum_{i=1}^r \min(a_i, b_i)},$$

$$F_{0; -1}(a, b) = \frac{\sum_{i=1}^r a_i + \sum_{i=1}^r b_i - 2 \sum_{i=1}^r \min(a_i, b_i)}{\sum_{i=1}^r a_i + \sum_{i=1}^r b_i}.$$

Мера различия $F_{1; -1}(a, b)$ является расстоянием (Romaniszyn, 1970; 1972; Levandovsky, Winter, 1972; Сёмкин, 1973). Мера различия $F_0(a, b)$ не удовлетворяет неравенству треугольника и не является расстоянием.

Мера различия

$$F_{0; \infty}(a, b) = \frac{\sum_{i=1}^r a_i + \sum_{i=1}^r b_i - 2 \sum_{i=1}^r \min(a_i, b_i) + \left| \sum_{i=1}^r a_i - \sum_{i=1}^r b_i \right|}{\sum_{i=1}^r a_i + \sum_{i=1}^r b_i + \left| \sum_{i=1}^r a_i - \sum_{i=1}^r b_i \right|}$$

является расстоянием, которое для дескриптивных наборов совпадает с расстоянием Юрцева (Сёмкин, 2007).

Из семейства мер различия $F_{\tau}(a, b)$:

$$F_{\tau}(a, b) = 1 - K_{\tau}(a, b), \quad -1 < \tau < \infty,$$

расстоянием являются меры различия при $\tau \geq 1$ (Приложение А₃).

Следует отметить, что как меры сходства, так и двойственные им меры различия могут использоваться для сравнительного анализа флор.

3. СОГЛАСОВАНИЕ МЕР СХОДСТВА И МЕР ВКЛЮЧЕНИЯ

При сравнительном анализе смешанных по видовому обилию флор (как разновеликих, так и одновеликих) определяется отношение сходства-включения, которое графически изображается в виде смешанных графов включения-сходства (ориентированных и неориентированных). При этом возникает проблема согласования мер сходства и мер включения. Мера сходства и меры включения взаимно согласованы друг с другом при пороге $\delta \geq 0$ тогда и только тогда, когда выполняется соотношение $K(a, b) \geq \delta \Leftrightarrow K(a; b) \geq \delta, K(b; a) \geq \delta$.

Если выполняется только одностороннее соотношение, т.е. если

$K(a; b) \geq \delta, K(b; a) \geq \delta \rightarrow K(a, b) \geq \delta$, то мера сходства одностороннее согласована с мерами включения. Если не выполняется и условие односторонней согласованности, то меры сходства и меры включения называются несогласованными.

Пример 1. Мера сходства Жаккара $K_{1;-1}(a, b)$ не согласована с мерами включения $K_{0;-1}(a; b)$ и $K_{0;-1}(b; a)$.

Пример 2. Мера сходства $K_{1;-1}(a, b)$ односторонне согласована с мерами включения $K_1(a; b)$ и $K_1(b; a)$.

Пример 3. мера сходства Браун-Бланке $K_{0;-\infty}(a, b)$ взаимно согласована с мерами включения $K_0(a; b)$ и $K_0(b; a)$.

Следует особо отметить, что класс мер $K_{\tau;-\infty}(a, b)$ взаимно согласован с мерами включения $K_{\tau}(a; b)$ и $K_{\tau}(b; a)$, т.е. взаимно согласованные меры составляют только класс эквивалентных мер Браун-Бланке.

Вводится также двойственное определение согласованности мер различия с мерами не включения. Мера различия $F(a, b)$ взаимно согласована с мерами не включения $F(a; b)$ и $F(b; a)$ при пороге $\delta \leq 1$ тогда и только тогда, когда выполняется соотношение $F(a, b) \leq 1 - \delta \Leftrightarrow F(a; b) \leq 1 - \delta, F(b; a) \leq 1 - \delta$.

Отметим, что мера различия Юрцева $F_{0;-\infty}(a, b)$ взаимно согласована с мерами не включения $F_{0;-\infty}(a; b)$ и $F_{0;-\infty}(b; a)$.

4. ПРИМЕРЫ АНАЛИЗА ОТНОШЕНИЙ СХОДСТВА КОНКРЕТНЫХ ФЛОР

4.1 Сравнительный анализ конкретных флор островов Верхотурова и Карагинский (Берингово море)

Сравнительный анализ шести конкретных флор островов Верхотурова и Карагинский произведён на основе матрицы пересечений (Баркалов, Кожевников, Харкевич, 1986). По матрице пересечений (табл. 1, над диагональю) нами рассчитана матрица мер сходства Сёренсена (табл. 1, под диагональю), матрица мер включения (табл. 2),

Таблица 1

Матрица числа общих видов (матрица пересечений) для каждой пары конкретных флор (над диагональю) и матрица мер сходства (%) Сёренсена (под диагональю) островов Верхотурова и Карагинский

	1	2	3	4	5	6
1		192	173	184	199	160
2	289		204	208	221	172
3	69	271		205	209	172
4	64	78	253		220	152
5	65	76	77	277		173
6	68	78	76	76	298	
6	58	65	53	57	62	260

Примечание: 1 – о. Верхотурова; о. Карагинский; 2 – бассейн р. Маркелова; 3 – западный макросклон г. Высокая; 4 – западный и восточный макросклон г. Туманная; 5 – западный и восточный макросклон г. Перевал; 6 – окрестности бывших посёлков Ягодное и Островной.

Таблица 2

Матрица мер включения (%) конкретных флор островов Верхотурова и Карагинский (см. обозначения к табл. 1)

	1	2	3	4	5	6
1		66	60	64	69	55
2	100		75	77	82	63
3	71	100		81	83	54
4	68	81	100		79	55
5	66	75	74	100		58
6	67	74	70	74	100	
6	62	66	52	58	67	100

Таблица 3

Матрица мер сходства (%) Браун-Бланке (над диагональю) и матрица расстояний (%) Юрцева (под диагональю) островов Верхотурова и Карагинский (см. обозначения к табл. 1)

	1	2	3	4	5	6	
1		×	66	60	64	67	55
2	34		×	75	75	74	63
3	40	25		×	74	70	52
4	36	25	26		×	74	55
5	33	26	30	26		×	58
6	45	37	48	45	42		×

матрица мер сходства Браун-Бланке (табл. 3, над диагональю) и матрица расстояний Юрцева (табл. 3, под диагональю). При пороге $\delta \geq 70\%$ на основе матриц сходства Сёрнсена и Браун-Бланке поострен граф включения-сходства (рис. 1), а по матрице мер различия Сёрнсена (дополнение меры сходства до 1) и по матрице расстояний Юрцева построены оптимальные деревья (рис. 2 а, б соответственно).

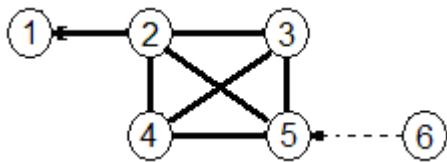


Рис. 1. Граф включения-сходства, построенный при пороге включения $\delta \geq 70\%$ по матрице мер сходства Сёрнсена (табл. 1, под диагональю) и матрице мер сходства Браун-Бланке (табл. 3, над диагональю). Пунктирной линией обозначено максимальное включение

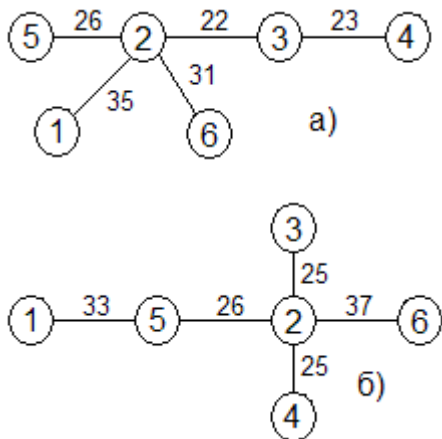


Рис. 2. Оптимальное дерево (дендрит), построенное по матрице мер различия Сёрнсена (а) и по матрице мер различия Юрцева (б)

Как видно, графы включения-сходства для матриц мер Сёрнсена и Браун-Бланке при пороге $\delta \geq 70\%$ совпали. Содержательная интерпретация графов включения-сходства достаточно проста: конкретные флоры о. Карагинского (2, 3, 4, 5) сходны между собой (образуют полный подграф или «ядро»), а конкретная флора (6) максимально включается в конкретную флору (5). В конкретную флору о. Верхотурова (1) включается конкретная флора (2) о. Карагинского. Конкретные флоры (1) и (6) составляют «периферию».

Оптимальные деревья, построенные двумя способами, несколько различны. Однако, как «ядро» сходных конкретных флор, так и конкретных флор «периферии» в обоих оптимальных деревьях составляют одинаковые конкретные флоры: (2, 3, 4, 5) и (1, 6) соответственно.

4.2 Сравнительный анализ конкретных флор европейского северо-востока на основе широтных спектров

Относительные широтные спектры представляют собой одновеликие дескриптивные множества, сумма весов элементов которых равна 1 (или 100%). В этом случае может быть использована только одна мера сходства – мера процентного сходства (Юрцев, Семкин, 1980), т.е. $E(A, B) = m(A \cap B)$ или двойственная ей мера различия $D(A, B) = 1 - m(A \cap B)$.

Таблица 4
Широтные спектры 19 конкретных флор европейского северо-востока (по: Ребристая, 1977)

Подзоны ²	№	Конкретная флора	α_1	α_2	α_3	α_4
А	1	о. Георга	63	37	-	-
	2	П-ов Адмиралтейства	54	40	3	3
В	3	Маточкин Шар	52	31	15	2
С	4	Хабарово	46	27	17	10
	5	Амдерма	42	29	17	12
	6	Хуптпэ	31	28	24	17
	7	Усть-Кара	36	26	23	15
D	8	Хальмер-Ю	20	27	24	29
	9	Верховья Кары	18	31	24	27
	10	Гнетью	16	28	26	30
	11	Ониндомусюр	15	15	28	42
	12	М. Сюрней	16	16	28	40
	13	Нямдою	17	15	27	41
	14	Саваю	14	16	30	40
	15	Морею	15	16	29	40
Е	16	Воркута	13	16	25	46
	17	Пымвашор	6	15	22	57
F	18	Лёк-Воркута	7	13	25	55
	19	Сивая Маска	3	10	23	64

Примечание: А – высокоарктические тундры; В – арктические тундры; С – северные гипоарктические тундры; D – южные гипоарктические тундры, северная мелкоерниковая полоса; Е – южные гипоарктические тундры, южная крупноерниковая полоса; F – лесотундра; α_1 – арктические виды; α_2 – арктоальпийские виды; α_3 – гипоарктические виды; α_4 – бореальные виды.

² Подзона «южные гипоарктические тундры» не всегда делится на две полосы (Юрцев, Толмачев, Ребристая, 1978).

Матрица сходства, рассчитанная по широтным спектрам 19 конкретных флор европейского северо-востока
(по: Ребристая, 1977), по диагонали 100% сходства (×)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	×	91	83	73	71	59	62	47	49	44	30	32	32	30	31	29	21	20	13
2	91	×	88	79	77	65	68	53	55	50	36	38	38	36	37	35	27	26	19
3	83	88	×	90	88	76	79	64	66	61	47	49	49	47	48	46	38	37	30
4	73	79	90	×	96	85	89	74	72	70	57	59	59	57	58	56	48	47	40
5	71	77	88	96	×	88	91	76	76	73	59	61	61	59	60	58	50	49	42
6	59	65	76	85	88	×	95	88	87	85	71	73	73	71	72	70	60	61	53
7	62	68	79	89	91	95	×	84	82	80	68	70	70	68	69	67	58	58	51
8	47	53	64	74	76	88	84	×	96	96	83	85	85	83	84	82	72	73	65
9	49	55	66	72	76	87	82	96	×	95	81	83	83	81	82	80	70	71	63
10	44	50	61	70	73	85	80	96	95	×	86	88	87	86	87	84	73	75	66
11	30	36	47	57	59	71	68	83	81	86	×	98	98	97	98	95	85	87	78
12	32	38	49	59	61	73	70	85	83	88	98	×	98	98	99	94	83	85	76
13	32	38	49	59	61	73	70	85	83	87	98	98	×	96	97	94	84	86	77
14	30	36	47	57	59	71	68	83	81	86	97	98	96	×	99	94	83	85	76
15	31	37	48	58	60	72	69	84	82	87	98	99	97	99	×	94	83	85	76
16	29	35	46	56	58	70	67	82	80	84	95	94	94	94	94	×	89	91	82
17	21	27	38	48	50	60	58	72	70	73	85	83	84	83	83	89	×	96	92
18	20	26	37	47	49	61	58	73	71	75	87	85	86	85	85	91	96	×	91
19	13	19	30	40	42	53	51	65	63	66	78	76	77	76	82	92	91	×	×

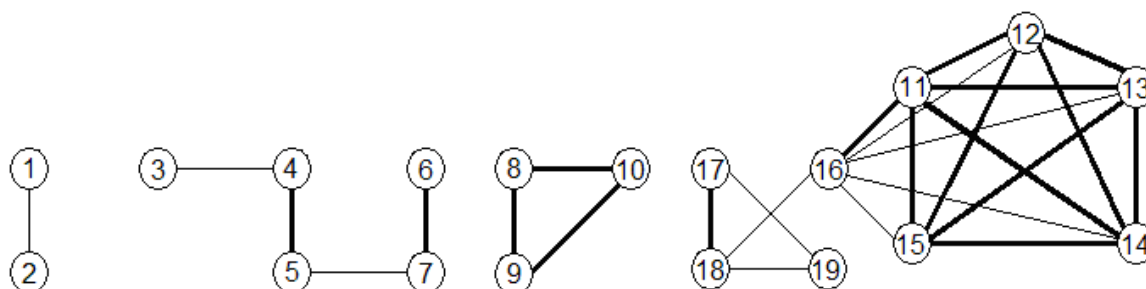


Рис. 3. Граф сходства широтных спектров 19 конкретных флор европейского северо-востока
(при порогах: $\delta \geq 95\%$ (жирные линии), $\delta \geq 90\%$ (тонкие линии))

Для иллюстрации сравнительного анализа нами использовались 19 конкретных флор, взятых из 6 ботанико-географических подзон тундровой зоны по данным О.В. Ребристой (1977). По широтным спектрам (табл. 4) с помощью меры процентного сходства рассчитана матрица сходства (табл. 5), на основе которой проиллюстрированы методы сравнительного анализа.

4.2.1 Метод построения графа сходства

Построение неориентированного графа происходит на основе симметричной матрицы сходства или различия. Выбирается порог сходства δ (или различия $1-\delta$) и те объекты (описания), мера сходства между которыми больше или равна порогу соединяются произвольной ли-

нией (или ребром). В результате получается граф сходства, вершины которого изображают объекты (описания), а рёбра указывают на отношение сходства между объектами (описаниями).

Обычно строятся несколько графов сходства при различных пороговых величинах. Из этих графов выбирают только те, которые характеризуют наиболее «сильные» связи. Возможно совмещение нескольких графов на одном рисунке посредством выделения рёбер графов линиями разной толщины, штриховки, цвета и т.д.

Этот метод под названием «сетевого анализа» и «плеяд Терентьева» широко использовался в геоботанике, экологии и систематике (Терентьев, 1959; 1960; Александрова, 1969; Василевич, 1969; McIntosh, 1973; Песенко, 1982; Галанин, 1991) и в настоящее время также не потерял своего значения (Беликович, 2001 а,б; Кожевников, Кожевникова, 2004).

В нашем случае по матрице мер сходства (табл. 5) наиболее сильные связи выделяются при порогах $\delta \geq 95\%$ и $\delta \geq 90\%$ (рис. 3). При этом получается граф, имеющий 5 компонент связности, т.е. рассматриваемые конкретные флоры разбиваются на 5 групп.

4.2.2 Метод диаграмм Чекановского (диагонализация)

И. Чекановский (Czekanowski, 1909) применил «метод дифференциального анализа» К. Пирсона для выделения антропологических групп (Макаревич, 1971). Польский геоботаник С. Кульчинский (Kulczynsky, 1927) использовал этот метод для разграничения растительных сообществ. Впоследствии этот метод стали называть «методом Чекановского» и под этим названием он часто упоминался в геоботанических и экологических работах (Ružička, 1958; Норин, 1965; 1971; Грейг-Смит, 1967; Макаревич, 1967; 1971; Александрова, 1969; Василевич, 1969; McIntosh, 1973; Леме, 1976; Песенко, 1982).

Иногда используется термин «диагонализация» (Браверман и др., 1971; Песенко, 1982) для метода приведения квадратной матрицы с помощью перестановки рядов (соединение перестановки строк с такой же перестановкой столбцов матрицы) к матрице, в которой наибольшие значения элементов расположены вдоль главной диагонали. Нами предлагается использовать метод «транзитивного замыкания» совместно с методом «диагонализации» для матриц сходства (различия) (Сёмкин, 1987).

Алгоритм расчёта прост: матрицу сходства «умножают» саму на себя $K \times K = K^2$ способом «строка-столбец» по формуле:

$$K_{ij}^2 = \max[\min(K_{i1}, K_{1j}), \min(K_{i2}, K_{2j}), \dots, \min(K_{in}, K_{nj})],$$

где $i = 1, \dots, n; j = 1, \dots, n$. Повторяя указанную процедуру, получаем матрицу $K^2 \times K^2 = K^4$, затем $K^4 \times K^4 = K^8$ и т.д. Процедура умножения матриц прекращается в случае получения идемпотентной матрицы, т.е. матрицы K обладающей свойством $K^2 = K$.

Для данного случая при возведении матрицы сходства (табл. 5) в 16 степень получаем матрицу транзитивного замыкания, в которой остаётся только

переставить 12 и 13 столбец (а также соответствующие строки), чтобы получить окончательный результат (табл. 6).

В матрице транзитивного замыкания осталось только 9 различных по значению элементов: 88, 90, 91, 92, 95, 96, 98, 99, 100. Эти элементы могут быть закодированы различными способами: различной штриховкой, приданием ячейкам матрицы различной окраски и др. (Сёмкин, 1987).

4.2.3 Метод построения дендрограмм

Методы построения дендрограмм подробно рассмотрены в работе Б.И. Сёмкина (1987), там же приведены дендрограммы для конкретных флор европейского северо-востока, построенные методом взвешенного среднего арифметического связывания (см. рис. 7б в работе (Сёмкин, 1987)). Эта же дендрограмма (рис. 4) воспроизведена нами для сравнения различных методов.

Следует отметить, что при построении дендрограммы методами агломеративных иерархических процедур используется алгоритм, при котором приходится выбирать максимальное число в матрице сходства (исключая диагональные элементы), а если таких чисел несколько, то предлагается брать одно из них по определённой правилу (Сёмкин, 1987). В результате такой процедуры определяется только одна дендрограмма. Однако если при выборе максимального элемента в матрице оказывается несколько таких элементов, то можно построить и несколько дендрограмм, беря поочерёдно каждый из них. Для исключения этого случая необходимо при расчёте мер сходства оставлять три, четыре или более знаков после целого числа, чтобы в матрице сходства все элементы по численной величине были различны. Таким образом определяется только одна дендрограмма. Данный способ был апробирован с помощью программы Statistica 6.0: нами были получены однозначные результаты при условии использования аналогичных мер сходства (различия) и способа построения дендрограммы.

4.2.4 Обсуждение результатов

Совокупность конкретных флор европейского северо-востока разбивается на две совокупности (рис. 4): A – арктические флоры с преобладанием арктических и аркто-альпийских видов и B – арктические флоры с преобладанием аркто-альпийских и бореальных видов. Эти типы флор были выделены В.Н. Андреевым (1954). В свою очередь группа A распадается на две подгруппы: A_1 – высокоарктические флоры и A_2 – арктические флоры, которые были выделены А.И. Толмачёвым (1932). Группа B также распадается на две подгруппы: B_1 – гипоарктические флоры и B_2 – бореальные флоры.

Матрица транзитивного замыкания, рассчитанная по широтным спектрам 19 конкретных флор европейского северо-востока (по: Ребристая, 1977), × – значения равные 100%

	1	2	3	4	5	6	7	8	9	10	11	13	12	14	15	16	17	18	19
1	×	91	88	88	88	88	88	88	88	88	88	88	88	88	88	88	88	88	88
2	91	×	88	88	88	88	88	88	88	88	88	88	88	88	88	88	88	88	88
3	88	88	×	90	90	90	90	88	88	88	88	88	88	88	88	88	88	88	88
4	88	88	90	×	96	91	91	88	88	88	88	88	88	88	88	88	88	88	88
5	88	88	90	96	×	91	91	88	88	88	88	88	88	88	88	88	88	88	88
6	88	88	90	91	91	×	95	88	88	88	88	88	88	88	88	88	88	88	88
7	88	88	90	91	91	95	×	88	88	88	88	88	88	88	88	88	88	88	88
8	88	88	88	88	88	88	88	×	96	96	88	88	88	88	88	88	88	88	88
9	88	88	88	88	88	88	88	96	×	96	88	88	88	88	88	88	88	88	88
10	88	88	88	88	88	88	88	96	96	×	88	88	88	88	88	88	88	88	88
11	88	88	88	88	88	88	88	88	88	88	×	98	98	98	98	95	91	91	91
13	88	88	88	88	88	88	88	88	88	88	98	×	98	98	98	95	91	91	91
12	88	88	88	88	88	88	88	88	88	88	98	98	×	99	99	95	91	91	91
14	88	88	88	88	88	88	88	88	88	88	98	98	99	×	99	95	91	91	91
15	88	88	88	88	88	88	88	88	88	88	98	98	99	99	×	95	91	91	91
16	88	88	88	88	88	88	88	88	88	88	95	95	95	95	95	×	91	91	91
17	88	88	88	88	88	88	88	88	88	88	91	91	91	91	91	91	×	96	92
18	88	88	88	88	88	88	88	88	88	88	91	91	91	91	91	91	96	×	92
19	88	88	88	88	88	88	88	88	88	88	91	91	91	91	91	91	92	92	×

Гипоарктические флоры выделялись А.И. Толмачёвым (Толмачёв, 1932; Ребристая, 1977).

Следовательно, при одноуровневой классификации выделяются хорошо известные типы арктических флор (Толмачёв, 1932; Ребристая, 1977): высокоарктические, арктические и гипоарктические флоры. Подгруппа A_1 делится на две части: A_{21} – арктические флоры с преобладанием арктических видов (с присутствием бореальных видов) и A_{22} – арктические флоры с равным участием арктических и аркто-альпийских видов.

Подгруппа B_1 делится на две части: B_{11} – смешанная бореально-арктическая флора и B_{12} – смешанная аркто-бореальная флора. Эти типы флор были также выделены В.Н. Андреевым (Андреев, 1954; Ребристая, 1977).

Результаты, полученные тремя вышеприведёнными методами, в целом не противоречат друг другу и позволяют дать согласованные интерпретации. Для дескриптивных наборов, компоненты которых выражены в долях или в процентах, все меры сходства и включения сводятся только к одной мере процентного сходства или двойственной ей мере процентного различия и мерам им эквивалентным. В сравнительной флористике следует чаще использовать относительные процентные спектры (таксономические, географические, экологические, и другие).

ЗАКЛЮЧЕНИЕ

В сравнительной флористике наиболее часто используют меры сходства Сёрнсена и Жаккара, а также их обобщения на случай количественных показателей. Из мер сходства (различия), удовлетворяющих условию взаимного согласования, можно указать только меру сходства Браун-Бланке и двойственную ей меру различия (расстояния) Юрцева, а также эквивалентные им меры.

В общем случае (при наличии разновеликих и одновеликих по числу видов флор) преимущество отдаётся взаимно согласованным мерам сходства Браун-Бланке и различия (расстоянию) Юрцева, благодаря которым возможно построение смешанных графов включения-сходства.

В рассмотренных нами примерах по сравнительному анализу конкретных флор мы обошлись только двумя мерами сходства и двойственными им мерами различия – мерой процентного сходства (различия) и мерой сходства Браун-Бланке (расстоянием Юрцева). Причем, как доказано в Приложении А₃, последние приведённые меры являются частными случаями мер процентного сходства (различия).

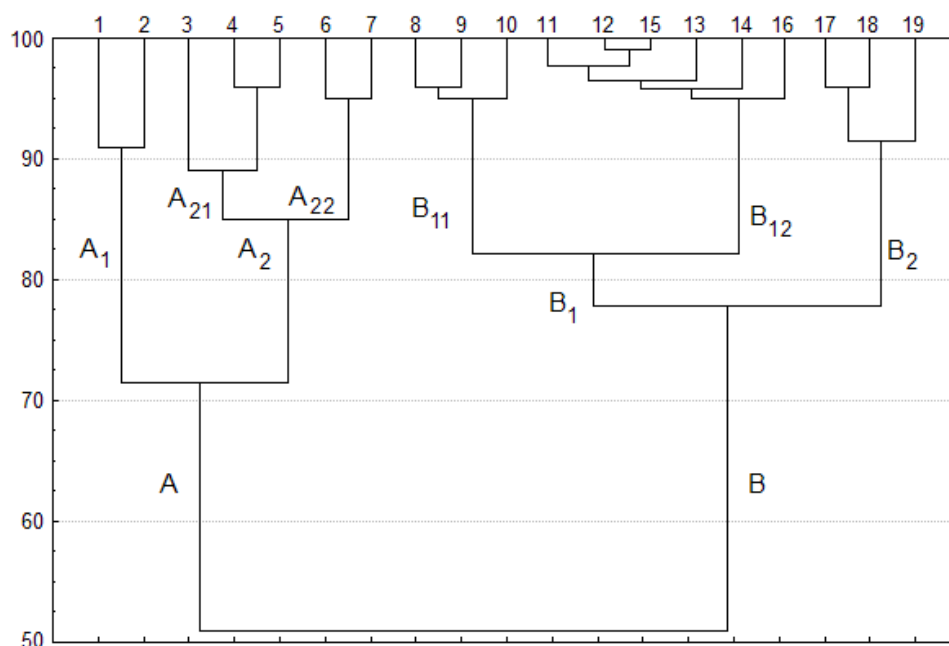


Рис. 4. Дендрограмма для конкретных флор европейского северо-востока, построенная методом взвешенного среднего арифметического связывания:

- A – арктические флоры с преобладанием арктических и аркто-альпийских видов;
- B – арктические флоры с преобладанием аркто-альпийских и бореальных видов;
- A_1 – высокоарктические флоры;
- A_2 – арктические флоры;
- B_1 – гипоарктические флоры;
- B_2 – бореальные флоры;
- A_{21} – арктические флоры с преобладанием арктических видов (с присутствием бореальных видов);
- A_{22} – арктические флоры с равным участием арктических и аркто-альпийских видов;
- B_{11} – смешанная бореально-арктическая флора;
- B_{12} – смешанная аркто-бореальная флора.

В принципе, для парного сравнения флористических признаков вполне достаточно использовать только одну меру процентного сходства или двойственную ей меру различия, удовлетворяющую аксиомам расстояния.

Аксиоматический подход к выделению мер сходства и различия позволяет отбросить большое количество «коэффициентов сходства», которые не удовлетворяют системе аксиом и не могут считаться мерами сходства (различия). Впервые показана возможность однозначного определения классов мер сходства (различия), удовлетворяющих установленным свойствам, исходя только из системы аксиом.

Широко используемые в сравнительной флористике, геоботанике и экологии методы систематизации данных – метод оптимального дерева, метод диагонализации (диаграмм Чекановского), метод построения графов сходства (с помощью пороговых величин) и метод дендрограмм вполне пригодны для установления эмпирических закономерностей. Эти методы, с использованием биоинформационных технологий, позволяют не только быстро обрабатывать огромные массивы информации, но и представлять результаты расчётов в приемлемой графической форме, удобной для интерпретации полученных закономерностей.

ПРИЛОЖЕНИЯ

A_1 КЛАСС ОТНОСИТЕЛЬНЫХ МЕР СХОДСТВА, ПРЕДСТАВЛЯЕМЫХ ДРОБНО-ЛИНЕЙНЫМИ ФУНКЦИЯМИ

Систему аксиом для меры сходства дескриптивных множеств A и B (Сёмкин, 2007) дополним ещё одной аксиомой:

1. $0 \leq K(A, B) \leq 1$ (аксиома ограничения);
2. $K(A, B) = K(B, A)$ (аксиома симметричности);
3. $K(A, B) = 0$ тогда и только тогда, когда $A \cap B = \emptyset$ (аксиома минимального сходства);
4. $K(A, B) = 1$ тогда и только тогда, когда $A=B$ (аксиома максимального сходства);
5. $K(A, B) = f(m(A), m(B), m(A \cap B)) = f(a, b, c)$ (мера сходства есть функция только трёх переменных a, b и c).

Будем искать меру сходства $K(A, B)$ в виде дробно-линейной функции от a, b и c :

$$K(A, B) = f(a, b, c) = \frac{\alpha_1 a + \alpha_2 b + \alpha_3 c + \alpha_4}{\beta_1 a + \beta_2 b + \beta_3 c + \beta_4},$$

где $a=m(A)$, $b=m(B)$, $c=m(A \cap B)$, a_i, b_i ($i=1, 2, 3, 4$) – искомые постоянные.

Из аксиом меры сходства (Сёмкин, 2007) и аксиом мер конвергенции (Сёмкин, Орешко, Горшков, 2008 а) получаем:

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

Можно показать, что из свойств (1–5) однозначно определяется дробно-линейная функция от a , b и c с точностью до параметра τ . Укажем ряд следствий, полученных из свойств (1–5): (2) $\Rightarrow \alpha_4 = \beta_4 = 0$; (3) $\Rightarrow \alpha_1 = \alpha_2 = \alpha$ и $\beta_1 = \beta_2 = \beta$; (4) $\Rightarrow \alpha = 0$; (5) $\Rightarrow \alpha_3 = 2\beta + \beta_3$.

В результате получим:

$$f(a, b, c) = \frac{c}{\left(\frac{\beta}{2\beta + \beta_3}\right)(a+b) + \left(\frac{\beta_3 c}{2\beta + \beta_3}\right)} = \frac{c}{\frac{2\beta}{2\beta + \beta_3} \frac{(a+b)}{2} + \frac{\beta_3 c}{2\beta + \beta_3}}$$

Обозначим $\frac{2\beta}{2\beta + \beta_2} = 1 + \tau$, тогда

$$\frac{\beta_3}{2\beta + \beta_2} = -\tau. \text{ Окончательно получим:}$$

$$f(a, b, c) = \frac{c}{\frac{(a+b)}{2}(1+\tau) - \tau c} \text{ или}$$

$$f_\tau(a, b, c) = \frac{\frac{2c}{a+b}}{(1+\tau) - \tau \frac{2c}{a+b}} = \frac{K_0}{(1+\tau) - \tau K_0},$$

где $K_0 = \frac{2c}{a+b}$. Из (1) $\rightarrow -1 < \tau < \infty$.

A₂ КЛАСС ОТНОСИТЕЛЬНЫХ МЕР РАЗЛИЧИЯ, ПРЕДСТАВЛЯЕМЫХ ДРОБНО-ЛИНЕЙНЫМИ ФУНКЦИЯМИ

Будем искать меру сходства $F(A, B)$ в виде дробно-линейной функции от a, b и c :

$$F(A, B) = g(a, b, c) = \frac{\alpha_1 a + \alpha_2 b + \alpha_3 c + \alpha_4}{\beta_1 a + \beta_2 b + \beta_3 c + \beta_4},$$

где $a=m(A)$, $b=m(B)$, $c=m(A \cap B)$, a_i, b_i ($i=1, 2, 3, 4$) – искомые постоянные.

Из аксиом меры различия (Сёмкин, 2007) и аксиом мер конвергенции (Сёмкин, Орешко, Горшков, 2008а) получаем:

6. $0 \leq g(a, b, c) \leq 1$;
7. $g(\lambda a, \lambda b, \lambda c) = g(a, b, c)$;

$$8. \quad g(a, b, c) = g(b, a, c)$$

$$9. \quad g(a, b, 0) = 1;$$

$$10. \quad g(c, c, c) = 0.$$

Из соотношений (1-5) следует семейство мер различия, упорядоченных параметром τ .

Укажем ряд следствий, полученных из свойств

$$(1-5): (2) \Rightarrow \alpha_4 = \beta_4 = 0;$$

$$(3) \Rightarrow \alpha_1 = \alpha_2 = \alpha \text{ и } \beta_1 = \beta_2 = \beta;$$

$$(4) \Rightarrow \alpha = \beta; (5) \Rightarrow \alpha_3 = 2\alpha.$$

В результате получим дробно-линейную функцию:

$$g(a, b, c) = \frac{\alpha(a+b) - 2\alpha c}{\alpha(a+b) + \alpha_4 c} = \frac{a+b-2c}{a+b + \left(\frac{\alpha_4}{\alpha}\right)c}, \alpha > 0.$$

Вводя обозначение $\left(\frac{\alpha_4}{\alpha}\right) = -2\varphi_\tau$, где

$$\varphi_\tau = \frac{\tau}{1+\tau}, \tau > -1 \text{ получим меру различия двух}$$

множеств A и B :

$$F(A, B) = \frac{a+b-2c}{a+b-2\varphi_\tau c}, \varphi_\tau = \frac{\tau}{1+\tau}, -1 < \tau < \infty.$$

Или в теоретико-множественных обозначениях:

$$F_\tau(A, B) = \frac{m(A) + m(B) - 2m(A \cap B)}{m(A) + m(B) - 2\varphi_\tau m(A \cap B)}.$$

Мера различия $F_\tau(A, B)$, $-1 < \tau < \infty$ есть расстояние при $\tau \geq 1$, ($\varphi_\tau \geq 1/2$). Например, мера различия

$$F_1(A, B) = \frac{m(A) + m(B) - 2m(A \cap B)}{m(A) + m(B) - m(A \cap B)}$$

расстояние, а мера различия

$$F_0(A, B) = \frac{m(A) + m(B) - 2m(A \cap B)}{m(A) + m(B)}$$

расстоянием не является. Доказательство приведено в работе Б.И. Сёмкина (1973).

A₃ МЕРА СХОДСТВА БРАУН-БЛАНКЕ И ЕЁ СВЯЗЬ С МЕРОЙ ПРОЦЕНТНОГО СХОДСТВА

Рассмотрим универсальное множество U , состоящее из r элементов – $\{x_1, \dots, x_r\}$. Тогда с помощью характеристической функции можно выделить следующие подмножества:

$$A \leftrightarrow \mu_A(x_i) = \begin{cases} 1, & x_i \in A \\ 0, & x_i \in U \setminus A \end{cases}, \quad B \leftrightarrow \mu_B(x_i) = \begin{cases} 1, & x_i \in B \\ 0, & x_i \in U \setminus B \end{cases}$$

$$A \cap B \leftrightarrow \mu_{A \cap B}(x_i) = \begin{cases} 1, & x_i \in A \cap B \\ 0, & x_i \in U \setminus B \end{cases}$$

$$\text{Тогда} \quad n(A) = \sum_{i=1}^r \mu_A(x_i), \quad n(B) = \sum_{i=1}^r \mu_B(x_i),$$

$$n(A \cap B) = \sum_{i=1}^r \mu_{A \cap B}(x_i) = \sum_{i=1}^r \min(\mu_A(x_i), \mu_B(x_i)).$$

$$K(p, q) = \sum_{i=1}^r \min(p_i, q_i) = \sum_{i=1}^r \min\left[\frac{\mu_A(x_i)}{n(A)}, \frac{\mu_B(x_i)}{n(B)}\right] = n(A \cap B) \min\left[\frac{1}{n(A)}, \frac{1}{n(B)}\right] = \frac{n(A \cap B)}{\max[n(A), n(B)]}$$

Аналогично можно показать, что процентное различие дескриптивных наборов p и q есть удвоенное расстояние Юрцева. Действительно,

$$d(p, q) = \sum_{i=1}^r p_i + \sum_{i=1}^r q_i - 2 \sum_{i=1}^r \min(p_i, q_i) = 2 - 2 \frac{n(A \cap B)}{\max[n(A), n(B)]} = 2 \left(1 - \frac{n(A \cap B)}{\max[n(A), n(B)]}\right) = 2 d_Y$$

Л и т е р а т у р а

Абакумов В.А., Филиппова Л.М. Гидробиологические методы контроля загрязнения водной среды // Опыт и методы экологического мониторинга. Пушино, 1978. С. 96–101.

Александрова В.Д. Классификация растительности. Обзор принципов классификации и классификационных систем в различных геоботанических школах. – Л., 1969. – 275 с.

Алексахина Т.И., Штина Э.А. Почвенные водоросли лесных биогеоценозов. – М., 1984. – 148 с.

Алехин В.В. Фитосоциология (учение растительных сообществ) и её последние успехи у нас и на Западе // Методика геоботанических исследований. Л.–М., 1925. С. 7–95.

Андреев В.Л. Системы-классификации в биогеографии и систематике (детерминистские методы) // Иерархические классификационные построения в географической экологии и систематике. Владивосток, 1979. С. 3–59.

Андреев В.Л. Классификационные построения в экологии и систематике. – М., 1980. – 142 с.

Андреев В.Н. Растительный покров восточно-европейской тундры и мероприятия по его использованию и преобразованию: Автореф. дисс. д-ра биол. наук. – Л., 1954. – 701 с.

Баркалов В.Ю. Сосудистые растения острова Шумшу и Парамушир (северные Курилы): Автореф. дисс. канд. биол. наук. – Томск, 1981. – 26 с.

Баркалов В.Ю., Кожевников А.Е., Харкевич С.С. Сосудистые растения островов Верхотурова и Карагинский (Берингово море) и охрана их

Проведём нормировку характеристической функции:

$$p_i = \frac{\mu_A(x_i)}{n(A)}, \quad i = 1, \dots, r; \quad q_i = \frac{\mu_B(x_i)}{n(B)}, \quad i = 1, \dots, r.$$

$$\text{Очевидно, что} \quad \sum_{i=1}^r p_i = 1, \quad \sum_{i=1}^r q_i = 1.$$

Рассмотрим дескриптивные наборы

$p = (p_1, \dots, p_r)$, $q = (q_1, \dots, q_r)$. Определим процентное сходство между дескриптивными наборами p и q :

генофонда // Комаровские чтения. Владивосток, 1986. Вып. 33. С. 110–168.

Беклемиев В.Н. Основные понятия биоценологии в приложении к животным компонентам наземных сообществ // Труды по защите растений. Т. 1. Вып. 2. 1931. С. 278–358.

Беклемиев В.Н. Термины и понятия, необходимые при количественном изучении популяций эктопаразитов и нидиколов // Зоол. журн. 1961. Т. 40. Вып. 2. С. 149–158.

Беклемиев В.Н. Биоценологические основы сравнительной паразитологии. – М., 1970. – 502 с.

Беликович А.В. Ландшафтная флористическая неоднородность растительного покрова (на примере модельных районов Северо-Востока России). – Владивосток, 2001а. – 248 с.

Беликович А.В. Растительный покров северной части Корякского нагорья. – Владивосток, 2001б. – 420 с.

Браверман Э.М., Дорофеев А.А., Лушельский В.Я., Мучник И.Б. Диагонализация матрицы связи и выявление скрытых факторов // Проблемы расширения возможностей автоматов. М., 1971. С. 42–79.

Быков Б.А. Геоботаника. – Алма-Ата, 1953. – 457 с.

Быков Б.А. Геоботаника. – 2-е изд., испр. Алма-Ата, 1957. – 367 с.

Быков Б.А. Геоботанический словарь. – 2-е изд., переработ. и доп. Алма-Ата, 1973. – 214 с.

Быков Б.А. Геоботаника. – 3-е изд., перераб. Алма-Ата, 1978. – 288 с.

Быков Б.А. Экологический словарь. – Алма-Ата, 1983. – 216 с.

- Василевич В.И.* Статистические методы в геоботанике. – Л., 1969. – 232 с.
- Воронов А.Г.* Геоботаника. – М., 1973. – 384 с.
- Галанин А.В.* Анализ распределения растений по типам местообитаний (на примере района равнинно-горного ландшафта среднего течения р. Рау-Чуа Западной Чукотки) // Ботан. журн. 1977. Т. 62. № 8. С. 1089–1101.
- Галанин А.В.* Экологополюсическая структура флоры хребта Куркуре (Восточный Алтай) // Ботан. журн. 1979. Т. 64. № 10. С. 1041–1413.
- Галанин А.В.* Флора и растительность Усть-Чаунского биологического стационара (Западная Чукотка) // Ботан. журн. 1980. Т. 65. № 9. С. 1174–1187.
- Галанин А.В.* Флора и ландшафтно-экологическая структура растительного покрова. – Владивосток, 1991. – 272 с.
- Грейг-Смит П.* Количественная экология растений. – М., 1967. – 360 с.
- Гусарова И.С., Сёмкин Б.И.* Сравнительный анализ конкретных флор сублиторали Север-западной части Тихого океана // Биология шельфовых зон мирового океана. Часть I (Тез. докл. 3 Всесоюз. конф. по морской биол.). Владивосток, 1982. С. 18–19.
- Дажо Р.* Основы экологии. – М., 1975. – 415 с.
- Жадовский А.Е.* Метод Жаккара (Jaccard) в изучении растительных сообществ // Методика геоботанических исследований. Л.-М., 1925. С. 97–117.
- Заки М.А., Шмидт В.М.* О систематической структуре флор южного Средиземноморья. I. Методика и анализ структуры 5 региональных и 11 локальных флор // Вестник ЛГУ. Биология. 1972. № 9. Вып. 2. С. 57–69.
- Зайцев Г.Н.* Математическая статистика в экспериментальной ботанике. – М., 1984. – 424 с.
- Захаров Ю.Д.* Раннетриасовые аммоноидеи Востока СССР. – М., 1978. – 224 с.
- Кафанов А.И., Жуков В.Е., Федотов П.А.* Состав и распределение летней фауны амфипод (*Amphipoda, Gammaridea*) на литорали бухты Витязь // Гидробиологические исследования заливов и бухт Приморья. Владивосток, 1984. С. 52–68.
- Кац Н.Я., Кац С.В., Китиани М.Г.* Атлас и определитель плодов и семян, встречающихся в четвертичных отложениях СССР. – М.: Наука, 1965. – 367 с.
- Кожевников Ю.П.* Анализ флоры Телекойской рощи и её окрестностей (Центральная Чукотка) // Ботан. журн. 1974. Т. 59. № 7. С. 967–979.
- Кожевников А.Е., Кожевникова З.В.* Эффективность охраны сосудистых растений Приморья и Приамурья на заповедных территориях // Вестник ДВО РАН. № 4. 2004. С. 8–22.
- Красилов В.А.* Палеоэкология наземных растений (основные принципы и методы). – Владивосток, 1972. – 210 с.
- Кусакин О.Г., Ростомов С.А.* Биогеографическая структура фауны равнооких ракообразных шельфа западного Сахалина и острова Монерон // Морская биогеография. Предмет, методы, принципы районирования. М., 1982. С. 176–184.
- Лащинский Н.Н., Ронгинская А.В.* Особенности флористического состава травяных сосновых лесов Нижнего Приангарья // Изв. Сиб. отд. АН СССР. Сер. биол. 1975. Вып. 2. № 10. С. 50–60.
- Леме Ж.* Основы биогеографии. – М., 1976. – 309 с.
- Макаревич В.Н.* Применение метода Чекановского при изучении динамики флористического состава лугового сообщества // Ботан. журн. 1967. Т. 52. № 11. С. 1630–1639.
- Макаревич В.Н.* Применение метода Чекановского при первичной обработке геоботанических описаний // Методы выделения растительных ассоциаций. Л., 1971. С. 125–140.
- Макфедьен Э.* Экология животных. Цели и методы. – М.: Мир, 1965. – 375 с.
- Мальшев Л.И.* Флористические спектры Светского союза // История флоры и растительности Евразии. Л., 1972. С. 17–40.
- Марина Л.В.* Опыт сравнительного анализа высокогорных флор речных бассейнов хребта Куркуре (Восточный Алтай) // Ботан. журн. 1982. Т. 67. № 3. С. 285–292.
- Марина Л.В.* Сравнительный анализ двух высокогорных флор Восточного Алтая (по их внутриландшафтным подразделениям): Автореф. дисс. канд. биол. наук. – Л., 1983. – 23 с.
- Миркин Б.М., Розенберг Г.С.* Фитоценология. Принципы и методы. – М., 1978. – 211 с.
- Неронов В.М., Лушеккина А.А.* Региональная биогеография Монгольской Народной Республики (анализ опубликованных схем районирования) // Итоги науки и техники. Биогеография. М., 1980. Т. 2. С. 41–64.
- Несис К.Н.* Зоогеографическое положение Средиземного моря // Морская биогеография. Предмет, методы, принципы районирования. М., 1982. С. 270–306.
- Ниценко А.А.* Растительная ассоциация и растительное сообщество как первичные объекты геоботанического обследования. – Л., 1971. – 183 с.
- Норин Б.Н.* О синузильном сложении растительного покрова лесотундры // Ботан. журн. 1965. Т. 50. № 6. С. 745–764.
- Норин Б.Н.* Использование коэффициентов сходства для классификации микрогруппировок лесотундры // Методы выделения растительных ассоциаций. Л., 1971. С. 206–225.
- Норин Б.Н.* Структура растительных сообществ Восточно-европейской лесотундры. – Л., 1979. – 200 с.

- Одум Ю. Основы экологии. М., 1975. – 740 с.
- Песенко Ю.А. Принципы и методы количественного анализа в фаунистических исследованиях. – М., 1982. – 287 с.
- Пийн Т.Х. Напочвенные лишайники и их местообитания на мысе Челюскин (Таймыр) // Структура растительности полярных пустынь и болот. Труды по ботанике. Тарту, 1982. Т. 59. С. 22–38.
- Пушкарёв В.С., Короткий А.М., Сёмкин Б.И. Кластерный анализ флористических и фитоценологических описаний ископаемых диатомовых водорослей сообществ голоцена Сихотэ-Алиня // Ботан. журн. 1984. Т. 69. № 11. С. 1532–1542.
- Ребристая О.В. Флора востока Большеземельской тундры. – Л., 1977. – 334 с.
- Сапегин Л.М. Структура и изменчивость луговых фитоценозов. – Минск, 1981. – 100 с.
- Сергиенко В.Г. Материалы к анализу конкретных флор полуострова Канин // Вестник ЛГУ. 1979. № 21. Вып. 4. С. 37–45.
- Сёмкин Б.И. Об аксиоматическом подходе определению мер различия и квазиразличия на семействах множеств // Информационные методы в системах управления измерением и контролем. Т. 1. Владивосток, 1972а. – С. 23–26.
- 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., 1972б. С. 12–16.
- Сёмкин Б.И. Deskриптивные множества и их приложения // Исследование систем. Т. 1. Анализ сложных систем. Владивосток, 1973. С. 83–94.
- Сёмкин Б.И. Основы методов систематизации географических данных // Математические методы в экологии и географии. Владивосток, 1978. С. 3–11.
- 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. Теоретико-графовые методы в сравнительной флористике // Теоретические и методические проблемы сравнительной флористики: Материалы II рабочего совещания по сравнительной флористике. Неринга. 1983. Л., 1987. С. 149–163.
- Сёмкин Б.И. Количественные показатели для оценки односто-ронних флористических связей, предложенных Б.А. Юрцевым // Ботан. журн. 2007. Т. 92. № 4. С. 114–127.
- Сёмкин Б.И., Двойченко В.И. Об эквивалентности мер сходства и различия // Исследование систем. Т. 1. Анализ сложных систем. Владивосток, 1973. С. 95–104.
- Сёмкин Б.И., Комарова Т.А. Анализ фитоценологических описаний с использованием мер включения (на примере растительных сообществ долины реки Амгуэмы на Чукотке) // Ботан. журн. 1977. Т. 62. № 1. С. 54–63.
- Сёмкин Б.И., Орешико А.П., Горшков М.В. Об использовании биоинформационных технологий в сравнительной флористике. I. Схемно-целевой подход. Абсолютные меры сходства и различия // Бюлл. БСИ ДВО РАН [Электронный ресурс]: науч. журн. / Ботан. сад-институт ДВО РАН. – Владивосток, 2009. Вып. 3. С. 102–111 – <http://botsad.ru/jurnal/number3.htm>
- Сёмкин Б.И., Орешико А.П., Горшков М.В. Об использовании биоинформационных технологий в сравнительной флористике. II. Меры включения дескриптивных множеств и их использование // Бюлл. БСИ ДВО РАН [Электронный ресурс]: науч. журн. / Ботан. сад-институт ДВО РАН. – Владивосток, 2009. Вып. 4. С. 58–70. – <http://botsad.ru/jurnal/number3.htm>
- Сукачёв В.И. Руководство к исследованию типов лесов. – М.–Л., 1931. – 328 с.
- Суханов В.В. Модель распределения видового обилия на литорали островной гряды // Морская биогеография. Предмет, методы, принципы районирования. М., 1982. С. 52–75.
- Тамарин П.В., Шмидт В.М. Сравнительный анализ некоторых коэффициентов сходства // Успехи биометрии. Л., 1975. С. 45–54.
- Терентьев П.В. Метод корреляционных плеяд // Вестн. ЛГУ. 1959. № 9. С. 35–43.
- Терентьев П.В. Дальнейшее развитие метода корреляционных плеяд // Применение математических методов в биологии. Т. 1. Л., 1960. С. 42–58.
- Толмачёв А.И. Флора центральной части Восточного Таймыра, 1–3 // Труды Полярной комиссии. Л., 1932. Вып. 8. С. 1–126.
- Толмачёв А.И. Введение в географию растений. – Л., 1974. – 244 с.
- Трасс Х.Х. Геоботаника. История и современные тенденции развития. – Л., 1976. – 252 с.
- Уиттекер Р. Сообщества и экосистемы. – М., 1980. – 327 с.
- Уланова Н.Г. Математические методы в геоботанике. – М., 1995. – 109 с.
- Чернов Ю.И. Природная зональность и животный мир суши. – М., 1975. – 224 с.
- Шаткаускас А.В. Некоторые географические особенности восточноевропейских широколиственных лесов // Ботан. журн. 1975. Т. 60. № 5. С. 646–662.
- Шеляг-Сосонко Ю.Р., Куковица Г.С. О степени сходства травостая степных и лесных формаций Подолии // Количественные методы анализа растительности. Материалы 2-го Всесоюз. совещ.: Применение количественных методов при изучении структуры растительности. БИОМАТ. Тарту, 1969. С. 157–160.
- Шенников А.П. Введение в геоботанику. – Л., 1964. – 447 с.

- Шмидт В.М.* Статистические методы в сравнительной флористике. – Л., 1980. – 176 с.
- Юрцев Б.А., Секретарёва Н.А.* Тнэквеемская чозениевая роща на севере Нижнеанадырской низменности // Эколого-ценотические и географические особенности растительности. (К 100-летию В.В. Алёхина). М., 1983. С. 232–248.
- Юрцев Б.А., Сёмкин Б.И.* Изучение конкретных и парциальных флор с помощью математических методов // Ботан. журн. 1980. Т. 65. № 12. С. 1706–1718.
- Юрцев Б.А., Толмачёв А.И., Ребростая О.В.* Флористическое ограничение и разделение Арктики // Арктическая флористическая область. Л., 1978. С. 9–104.
- Braun-Blanquet J.* Pflanzensoziologie Grundzüge der Vegetationskunde. – Wien, 1951. – 631 p.
- Czekanowski J.* Zur differential Diagnose der Neandertalgruppe // Korrespbl. Dtsch. Ges. Anthropol. 1909. Bd 40. S. 44–47.
- Češka A.* Estimation of the mean floristic similarity between and within sets of vegetational releves // Folia geobot. phytotax. Vol. 1. № 2 (1–92). 1966. P. 93–100.
- Dagnelie P.* Contribution à l'étude des communautés végétales par l'analyse factorielle // Bull. Serv. Carte phytogéograph. ser. B. № 5. 1960. P. 7–71, 93–195.
- Dzwonko Z.* Application of Jaccard's and Sørensen's formulas in numerical comparison and classification of phytosociological records // Zexz. nauk. UJ. 1978. № 493. P. 23–28.
- Ekman S.* Begründung einer statistischen Methode in der regionalen Tiergeographie nebst einer Analyse der paläarktischen Steppen- und Mästenfauna // Nova acta reg. soc. sci. upsal. ser. 4, 1940. Vol. 12, fasc. 2. P. 1–115.
- Goldsmith F.B., Harrison C.M.* Description and analysis of vegetation // Methods in Plant Ecology. Oxford-London-Edinburgh-Melbourne, 1976. P. 85–155.
- Goodall D.W.* Sample Similarity and Species Correlation // Handbook of Vegetation science. Part 5. Ordination and Classification of vegetation. London, 1973. P. 107–156.
- Gounot M.* Méthodes d'étude a quantitative de la vegetation. – Paris, 1969. – 314 p.
- Greig-Smith P.* Quantitative plant ecology. 2nd Ed. – London, 1964. – 242 p.
- Hagmeier E.M.* A numerical analysis of the distribution patterns of North American mammals. II. Re-evaluation of the provinces // Syst. Zool. 1966. Vol. 15. № 4. P. 279–299.
- Hagmeier E.M., Stults C.D.* A numerical analysis of the distributional patterns of North American mammals // Syst. Zoology. 1964. Vol. 13. № 3. P. 125–155.
- Harper R.M.* Vegetation types in Florida // Geol. Surv. 7th Ann. Rept. 1915. P. 135–188.
- Jaccard P.* Distribution de la flore alpine dans le Bassin des Dranses et dans quelques regions voisines // Bull. Soc. Vaudoise sci. Natur. 1901. V. 37. Bd. 140. S. 241–272.
- Jaccard P.* Gezetre der Pflanzenvertheilung in der alpinen Region. Auf Grund statistisch-floristischer Untersuchungen // Flora. 1902. Bd. 90. S. 351–377.
- Jaccard P.* Nouvelles recherches sur la distribution florale // Bull. Soc. Vaudoise sci. Natur. 1908. V. 44. Bd. 140. – S. 223–270.
- Jaccard P.* The distribution of the flora in the alpine zone // New Phytol. 1912. V. 11. P. 37–50.
- Jaccard P.* Die statistische-floristische Methode als Grundlage der Pflanzensociologie // Handb. Biol. Arbetsmenth. (Aberhalden), 1928. V. 11. № 5. P. 165–202.
- Kulczynsky S.* Zespoly roślin w Pienach // Bull. intern. acad. polon. sci. lett. Cl. sci. math. natur. ser. B. 1927. S. 2. P. 57–203.
- Levandowsky M., Winter D.* Distance between sets // Nature. V. 234. № 5323. 1971. P. 34–35.
- Maddocks R.F.* Distribution patterns of living and subfossil podocopid ostracodes in the Nosy Bé area, northern Madagascar: Univ. Kansas Paleont. Contr. 1966. – 72 p.
- McIntosh R.P.* Matrix and plexus techniques // Handbook of Vegetation science. Part 5. Ordination and Classification of vegetation, London, 1973. P. 158–191.
- Mello J.F., Buzas M.A.* An application of cluster analysis as a method of determining biofacies // J. Paleont. 1968. V. 42. № 3. P. 747–758.
- Mueller-Dombois D., Ellenberg H.* Aims and methods of vegetation ecology. – New York, 1974. – 547 p.
- Romaniszyn W.* An attempt at interpreting agglomerative tendencies of animals based on definition of similarity and distance // Wiad. Ekol. 1970. Bd. XVI. Zes. 4. – S. 306–327.
- Romaniszyn W.* Uwagi Krytyczne o difiniciti Soresena; metodzie Renkonene obliczania Współczynnikow podobienstwa zbiorow // Wiad. Ekol. 1972. Bd. XVIII. Zes. 4. S. 375–380.
- Roux G., Roux M.* A propos du quelques méthodes de classification en phytosociologie // Rev. statist. appl. 1967. V. 15. № 2. P. 59–72.
- Ružička M.K.* Anwendung mathematischer-statistischer Methoden in der Geobotanik (sintetische Bearbeitung von Aufnahmen) // Biologia. Bratislava. 1958. Roč. 13. č. 9. S. 647–661.
- 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.
- Stugren B., Rădulescu M.* Metode matematice in zoogeografia regionala // Stud. biol. 1961. T. 12. № 1. P. 8–24.
- Valentine J.W.* Numerical analysis of marine molluscan ranges on the extratropical Northeastern

Pacific Shelf // Limnology and Oceanography. 1966. V. 11. № 2. P. 198–211.

Williams W.T., Lance G.N. Logic of computer-based intrinsic classifications // Nature. 1965. V. 207. № 4993. P. 159–161.

ABOUT THE USE OF BIOINFORMATION TECHNOLOGIES IN COMPARATIVE FLORISTIC STUDIES. III. RELATIVE MEASURES OF SIMILARITY AND DISSIMILARITY OF DESCRIPTIVE SETS

B.I. Semkin¹, A.P. Oreshko¹, M.V. Gorshkov²

¹*Pacific Institute of Geography FEB RAS, Vladivostok*

²*Pacific State University for Economics, Vladivostok*

Key words: comparative floristics, statistic methods, similarity-dissimilarity measures, plant species lists, taxonomic spectrae, descriptive sets, bioinformation technology, optimal tree, matrix diagonalization, Czekanovskiy diagram, similarity graph, inclusive-similarity graph, dendrogram.

The axiomatic basis of measures of similarity and dissimilarity for descriptive sets is submitted. The problem of the coordination of measures of inclusion and measures of similarity is considered. It is offered to prefer one measure of percentage similarity (dissimilarity), which in that specific case coincides with a Braun-Blanquet measure and dual to it by a measure of dissimilarity (Yurtsev distance). The methods of systematization of floristic data are considered: a method of construction of optimum tree, method of diagonalization (Czekanowski method of the diagram), method of construction graphs of similarity and inclusion-similarity (with use of limit values), method of construction tree diagram.

Ill. 4. Tabl. 6. Bibl. 123