

МЕТОДИКА ИССЛЕДОВАНИЙ

ОБЗОР СОВРЕМЕННЫХ ПРОГРАММНЫХ СРЕДСТВ, ИСПОЛЬЗУЕМЫХ ДЛЯ АНАЛИЗА ГЕОБОТАНИЧЕСКИХ ДАННЫХ

A REVIEW OF THE MODERN PROGRAMS FOR THE GEOBOTANICAL ANALYSIS

© А. Б. НОВАКОВСКИЙ
A. B. NOVAKOVSKY

Институт биологии Коми НЦ УрО РАН.
167928, г. Сыктывкар, ул. Коммунистическая, 28. E-mail: novakovsky@ib.komisc.ru

Представлен обзор наиболее распространенных программных средств, используемых для анализа геоботанических данных. Рассмотрены их возможности, используемые алгоритмы, достоинства и недостатки. Для всех программ указаны интернет ссылки на сайты разработчиков.

Ключевые слова: классификация растительности, ординация, кластеризация.

Key words: vegetation classification, ordination, clustering.

ВВЕДЕНИЕ

На сегодняшний день существует множество программных средств, разработанных для автоматизации обработки геоботанических данных. Однако специалисты-геоботаники зачастую слабо информированы о них. Трудности вызывает как понимание используемых алгоритмов, так и интерпретация полученных результатов. Такое положение объясняется рядом причин. Во-первых, языковой барьер: большинство программ и сопроводительной документации разработано в странах Европейского союза и США. Во-вторых, недостаточная математическая подготовка, что затрудняет понимание используемых алгоритмов. В-третьих, большинство программ являются коммерческими.

Цель настоящего обзора — познакомить читателя с основными математическими подходами, используемыми в геоботанических исследованиях; дать краткое описание возможностей наиболее распространенных программных средств, реализующих эти подходы; привести интернет ссылки, по которым можно было бы найти демонстрационные версии предлагаемых к обсуждению программ. В работе подробно рассматриваются следующие программные средства: TURBOVEG, TWINSPAN, CANOCO, SYNTAXON, CAP, JUICE, PC-ORD, BIOTAS и GRAPHS.

ОСНОВНЫЕ ПРИНЦИПЫ ОБРАБОТКИ ДАННЫХ

При обработке геоботанической информации используются 2 основных подхода — ординация данных и их классификация.

Принцип *ординации* заключается в представлении изучаемых объектов в виде точек многомерного пространства, где свойства выступают как их координаты. Использование данного принципа позволяет проецировать эти точки на пространство с уменьшенной размерностью (до 2- или 3-мерного) и имеет своей целью графически показать взаимное расположение исследуемых объектов (рис. 1). Одним из основных условий при таком проецировании является сохранение расстояний между объектами (или, если точнее, минимально возможное их изменение). Такое представление дает возможность визуального выделения существующих трендов в наборах данных и облегчает поиск факторов, наиболее сильно влияющих на исследуемые объекты. Кластеризация — это отдельная тема, на которой мы не будем подробно останавливаться в данной работе. Перечислим названия наиболее распространенных методов: многомерное шкалирование, анализ главных компонент, анализ соответствий, канонический анализ соответствий и т. д. (Уиллиамс, Ланс, 1986; Legendre, Legendre, 1998; Пузаченко, 2004). Описание большого количества

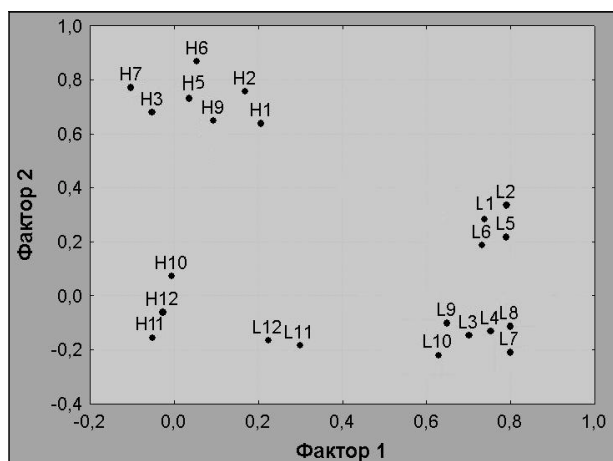


Рис. 1. Пример построения ординации.

An example of ordination.

методов ординации можно также найти в Интернете, например, на сайте Oklahoma State University: <http://www.okstate.edu/artsci/botany/ordinate>.

Под *классификацией* понимается задача разбиения всей совокупности рассматриваемых объектов на отдельные группы (классы) со сходными характеристиками (рис. 2) и определение взаимных отношений между ними. Существует несколько подходов в классификации. Первый, наиболее распространенный и математически формализованный, основывается на использовании мер сходства/расстояния. При таком подходе для всех пар рассматриваемых объектов рассчитывают степень их сходства между собой, а группы выделяют методами кластерного анализа. В геоботанике обычно используют коэффициенты сходства Жаккара, Сьеренсена, Стургена—Радулеску, Шимкевича—Симпсона и др., а также Евклидово расстояние, расстояние χ^2 и т. д. (Василевич, 1969; Шмидт, 1984; Лебедева, Кривоцуцкий, 2002). Кроме матриц сходств, в литературе описывается множество других методов кластеризации, основанных на оценке функций плотности статистического распределения, эвристических алгоритмах перебора, идея

математического программирования и др. (Ту, Гонсалес, 1978; Шитиков и др., 1987; Jongman et al., 1987; Ким, Мьюллер, 1989).

В фитоценологических исследованиях для классификации растительности часто используют довольно специфический метод, который основан на обработке валовых таблиц. Суть данного подхода заключается в последовательной перестановке строк и столбцов таблицы с целью нахождения блочно-диагонального вида. Такое представление позволяет выделить группы растений, четко характеризующие синтаксоны; другими словами, выделяется набор наиболее информативных признаков и одновременно производится разбиение множества объектов по этим признакам (Нешатаев Ю., 1987; Миркин и др., 2000). Этот подход к классификации, по сравнению с использованием мер сходства/различия, менее формализован и допускает большую свободу в его применении.

ОБЗОР ПРОГРАММ

TURBOVEG

<http://www.synbiosys.alterra.nl/turboveg>

Все описываемые ниже программы работают с массивами данных, которые требуется готовить специальным образом. Поэтому сначала остановимся на специализированной программе, предназначенной для ввода и хранения больших объемов фитоценологической информации — TURBOVEG (рис. 3), в которой количество хранимых элементов практически не ограничено. Основным предназначением TURBOVEG является функция преобразования накопленных данных в форматы, понятные для других программных продуктов, которые позволяют проводить более глубокую математическую обработку (TWINSPAN, SYNTAXON, JUICE, PC-ORD и др.).

Кроме хранения и преобразования данных, существует возможность проводить простые статистические расчеты, например: нахождение среднего

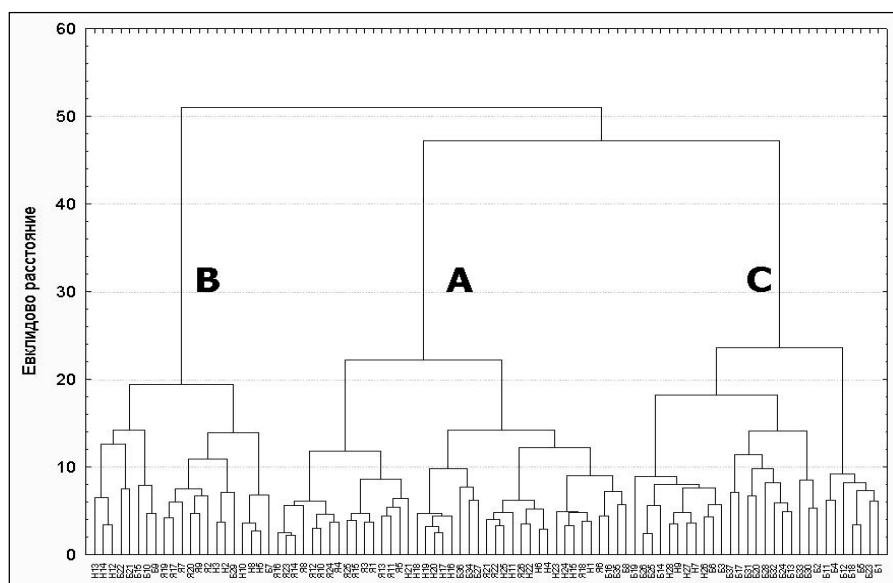


Рис. 2. Пример выделения кластеров.

An example of clustering.

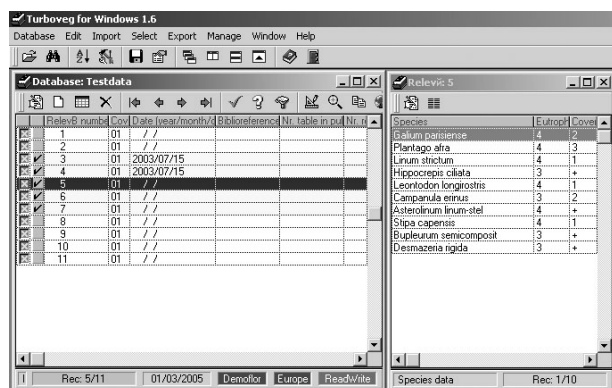


Рис. 3. Внешний вид главного окна программы TURBOVEG.

Слева — список геоботанических описаний, справа — список видов для выделенного описания

The TURBOVEG main window.

Relevé list and species list of a selected relevés are displayed on the left and on the right, respectively.

или максимального проективного покрытия всех видов, либо отдельных ярусов по выбранной группе описаний, расчет Евклидова расстояния и метрики χ^2 , построение графиков зависимостей между любыми характеристиками геоботанических описаний и т. д. (Hennekens, 1996; Hennekens, Schaminee, 2001).

В заключение можно сказать, что TURBOVEG обладает простым интерфейсом, позволяет хранить полную информацию о геоботанических описаниях. Основным предназначением программы является ввод, хранение и вывод накопленных данных. К сожалению, программа коммерческая, в свободном доступе находится только ее демонстрационная версия, которая не позволяет проводить операцию вывода данных. Реализованных же внутри программы алгоритмов обработки явно не достаточно.

TWINSPAN

Родоначальником среди программ, предназначенных для автоматической классификации растительности, можно считать TWINSAP, которая разрабатывалась в начале 1980-х гг. и предназначалась для автоматизации метода классификации Браун-Бланке. (Название TWINSAP получено как сокращение английских слов: **T**wo-**W**ay **I**ndicator **S**pecies **A**nalysis — двусторонний анализ на основе индикаторных видов.) Описание используемого алгоритма можно найти в работах: Hill, 1994; Legendre, Legendre, 1998.

Исходными данными для TWINSAP служит сводная таблица геоботанических описаний (столбцы таблицы — геоботанические описания, строки — виды растений). Результатом работы является валовая таблица, преобразованная к блочно-диагональному виду (рис. 4). Такое представление материала, как отмечалось выше, позволяет максимально выявить взаимосвязи между анализируемыми описаниями и встреченными видами и является основой для проведения классификации растительности. Результат отображается в виде текстового файла, в котором находящиеся справа и снизу от основной таблицы комбинации цифр «0» и «1» показывают принадлежность описаний и ви-

дов к полученным группам. При использовании терминологии Браун-Бланке можно сказать, что получившиеся группы описаний являются ассоциациями, а группы видов — дифференцирующими видами. Естественно, полученные результаты можно использовать только в качестве черновой заготовки, а для получения полноценной классификации требуется дальнейшая, ручная доводка валовой таблицы.

		11	1			
		219807	631245			
9	SOLI	SEM	211-3-	-----	0000	
5	PHRA	COM	53415-	-----	0001	
3	IVA	FRU	452113	-----	0010	
11	SPAR	PAT	3254-5	-----	0011	
8	SCIR	OLN	-1---5	3-----	01	
2	DIST	SPI	-54145	45-424	100	
1	ATRI	PAT	231--3	221411	101	
6	SALI	EUR	----22	-23411	1100	
10	SPAR	ALT	--142-	135553	1101	
7	SALI	VIR	-----	----34	1110	
12	SUAE	MAR	-----	----54	1110	
4	JUNC	GER	-----1	51----	1111	
			000000	111111		
			000001	001111		
			00011	010011		
			00101	0101		
			01			

Рис. 4. Пример валовой таблицы, обработанной с помощью программы TWINSAP (Hill, 1994).

Relevé table ordered and classified by TWINSAP (Hill, 1994).

Для лучшего понимания способов работы подобных алгоритмов остановимся на алгоритме TWINSAP более подробно.

Основной алгоритм классификации строится на основе так называемого принципа «дихотомии». Вся обработка ведется методом деления исходных данных на 2 части, поскольку любое разбиение можно представить как последовательность таких делений. В работе алгоритма можно выделить: 1) выделение групп сходных между собой геоботанических описаний; 2) выявление групп видов, наиболее тесно сопряженных (т. е. произрастающих совместно) друг с другом. Кратко алгоритм можно представить в виде следующих шагов:

1. Все множество описаний делится на 2 части*, каждая делится еще и еще раз, и так до тех пор, пока не получатся группы, состоящие из 3—4 элементов. Например, в нашем случае (рис. 4) первое деление дает следующие группы описаний: 12, 11, 9, 8, 10, 7 и 6, 3, 1, 2, 4, 5. В дальнейшем первая группа разбивается на 2 множества описаний: 12, 11, 9, 8, 10 и 7; а вторая группа: 6, 3 и 1, 2, 4, 5.

Примечание. * Наиболее сложный этап классификации состоит в разбиении множества описаний на 2 части (первый шаг). Для реализации данного действия используется следующий алгоритм:

1. При помощи одномерной ординации (используется метод взаимного усреднения — reciprocal averaging) проводится первичная расстановка описаний в направлении изменения флористического состава, т. е. описа-

ния упорядочиваются по градиенту изменения состава видов.

2. Все множество описаний делится на 2 части — это *первое приближение*.

3. На основании такого разбиения находят дифференцирующие виды, т. е. виды, тяготеющие только к одной из выделенных частей.

4. На основе выделенных дифференцирующих видов проводится новая, *улучшенная* расстановка описаний. При этом каждое описание принимает положение в соответствии с числом встретившихся в нем дифференцирующих видов; описания сортируются по убыванию этого числа.

5. Далее все множество описаний заново разделяют на 2 группы, с учетом найденных дифференцирующих видов.

6. Сортировка на основе небольшого количества наиболее сильно тяготеющих друг к другу дифференцирующих видов (*индикаторная сортировка*).

2. Полученные группы (классы) упорядочиваются таким образом, чтобы описания, относящиеся к близким (т. е. отличающимся наименьшим количеством элементов) классам оказались рядом.

3. Виды разбиваются на группы, сходные по встречаемости, аналогично первому пункту. В нашем примере в первую группу попали виды с номерами 9, 5, 13, 11, 8; а во вторую — 2, 1, 6, 10, 7, 12, 4.

4. Получившиеся группы видов упорядочиваются аналогично второму пункту.

Достоинства и недостатки программы. Фактически TWINSPAN является стандартом среди программ классификации растительности. Реализованный алгоритм позволяет достаточно быстро обрабатывать большие массивы данных. К недостаткам можно отнести как сложность работы с программой в части подготовки информации (данные хранятся в текстовой форме), так и отсутствие графического представления результатов.

CANOCO

<http://www.microcomputerpower.com/catalog/canoco.html>

Одной из первых среди программ, предназначенных для обработки геоботанических данных методами ординации, является CANOCO (CANONical Community Ordinations — каноническая ординация растительных сообществ). Программа используется как для выявления особенностей распределения растительных сообществ в соответствии с ведущими экологическими факторами, так и для подтверждения или опровержения результатов, полученных другими путями. В программе реализованы следующие методы:

- канонический анализ соответствий;
- принцип главных компонент;
- анализ избыточности.

В результате работы этих алгоритмов строятся ординационные диаграммы. Кроме графического результата, программа выдает корреляционные матрицы внешних факторов с осями ординационных диаграмм и внешних факторов друг с другом, что также немаловажно при интерпретации результатов. Подробное описание возможностей программы и принципов ее работы можно найти в работе: Jongman et al., 1987.

Для свободного скачивания в интернет доступна только версия CANOCO для DOS, с которой достаточно сложно работать, поскольку приходится

использовать разные запускаемые модули для обработки данных и графического отображения результатов.

SYNTAXON

<http://ramet.elte.hu/~podani/>

Одной из очень часто упоминаемых программ для проведения математической обработки геоботанических описаний является SYN-TAX, или SYNTAXON, которая предназначена для обработки геоботанических данных методами ординации и кластеризации и состоит из 3 модулей:

- модуль ординации, включающий в себя основные методы ординации (метрическая и не метрическая ординация, анализ соответствий, анализ избыточности, канонический корреляционный анализ и т. д.) и позволяющий графически представлять результаты анализа;

- модуль иерархической кластеризации (включает методы ближнего и дальнего соседа, кластеризации, основанной на энтропии и т. д.);

- модуль неиерархической кластеризации, в котором реализованы такие алгоритмы, как метод *k*-средних, метод нечеткой кластеризации и т. д.

Для определения мер сходства в программе предусмотрено использование 33 коэффициентов: качественные, количественные, ранговой и простой корреляций и т. д.

К достоинствам SYNTAXON можно отнести возможность графического представления получаемых результатов и использование большого числа коэффициентов сходства и различия, к недостаткам — отсутствие демонстрационной версии.

CAP

<http://www.pisces-conservation.com/indexsoftprog.html>

Одной из самых современных и обладающих простым интерфейсом пользователя является программа CAP (Community Analysis Package) (рис. 5). Программа предназначена для анализа растительных данных методами ординации, кластеризации и классификации. Ее отличительной особенностью является возможность работы с трансформирован-

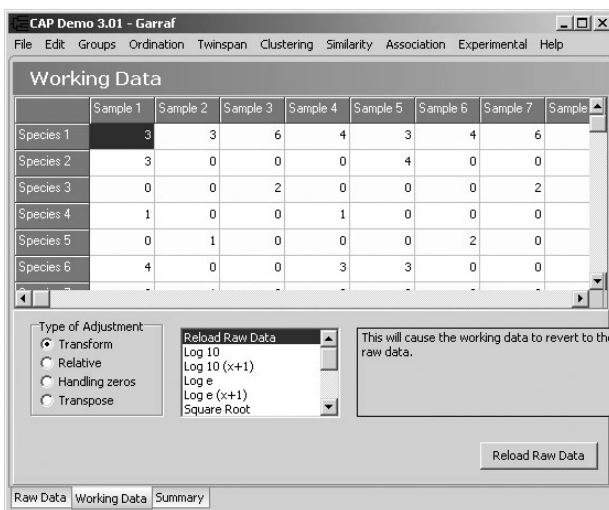


Рис. 5. Внешний вид главного окна программы CAP.

Закладка преобразования данных.

The CAP program main window with a data transforming page.

ными данными, что позволяет уменьшить влияние субъективного фактора. Причем легкость преобразованию дает возможность простого подбора наиболее подходящей формы трансформации. В программе реализованы следующие способы преобразования данных: логарифмический, экспоненциальный, синусоидальный, извлечения корня, изменение на определенную константу и т. д.

В САР реализованы следующие методы обработки данных:

- метод ординации: анализ главных компонент, анализ соответствий и т. д.;
- метод кластеризации: единичное присоединение, полное присоединение, присоединение по среднему и др.;
- отдельным пунктом меню вынесено использование алгоритма TWINSpan, предназначенного для проведения классификации растительных сообществ;
- существует возможность строить таблицы сходства с использованием качественных, количественных коэффициентов сходства и мер расстояния (всего 27 коэффициентов).

К положительным моментам можно отнести возможность графического отображения результатов работы алгоритмов и множества вспомогательных данных, таких как матрицы корреляций, статистические параметры для результатов и т. д.

JUICE

<http://www.sci.muni.cz/botany/juice.htm>

JUICE (рис. 6) предназначена для анализа и классификации больших массивов геоботанических данных как стандартными средствами, так и с помощью оригинального метода классификации «СОСТАИЛ» и оптимизирована для совместной работы с TWINSpan и TURBOVEG (Tichy, 2002, 2003). К достоинствам можно отнести возможность импорта данных из других форматов: текстовых файлов, формата RTF, Excel таблиц и баз данных Access. Обработка в JUICE может осуществляться как вручную (перестановкой строк и столбцов валовой таблицы), так и при помощи встроенных алгоритмов:

- сортировка строк и колонок валовой геоботанической таблицы по алфавиту, среднему проективному покрытию, выбранным жизненным формам или другим характеристикам;
- автоматическое построение синоптических таблиц;
- выявление индикаторных, доминантных и константных видов;
- возможность вычисления коэффициентов β -разнообразия;
- использование подходов других программных продуктов — проведение классификации методами TWINSpan и СОСТАИЛ.

Программа JUICE позволяет строить синоптические таблицы, проводить сортировку по описаниям и видам, а также классифицировать растительность различными методами (TWINSpan, СОСТАИЛ). Все алгоритмы могут работать с большими объемами информации (до 65 тыс. описаний). Программа является бесплатной. Кроме того, хорошо проработана система помощи. К недостаткам можно отнести отсутствие графического представления данных.

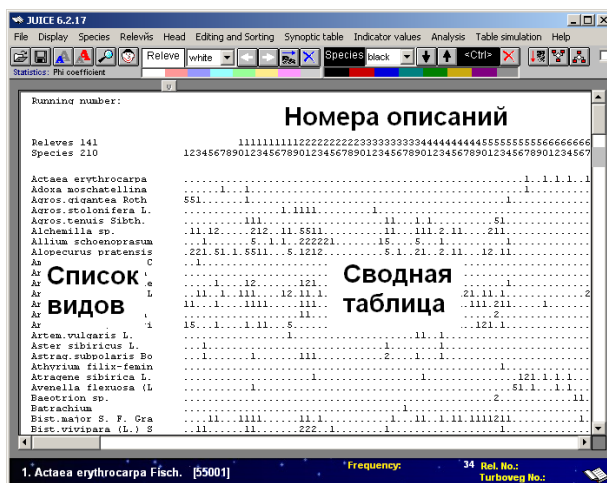


Рис. 6. Внешний вид валовой таблицы в программе JUICE.

The phytosociological table in JUICE

Остановимся подробнее на алгоритме классификации — СОСТАИЛ (Bruilheide, 1997, 2000). Этот метод похож на традиционный подход Браун-Бланке. Отличие заключается в использовании понятия «степень верности вида», которая определяется соотношением числа встреч данного вида в наборе описаний к общему количеству описаний.

Порядок выполнения обработки данных:

1. Реализация алгоритма возможна из двух положений: а) пользователь выбирает наборы описаний, являющиеся предварительно выделенными ассоциациями; б) выбираются наборы видов со сходными экологическими характеристиками, т. е. встречающиеся в одних экотопах.

2. Если алгоритм начинается из положения «а», то проводится расчет коэффициента верности для всех видов по отношению к выбранным группам описаний, виды с максимальными коэффициентами объявляются стартовыми наборами видов. Если алгоритм начинается из положения «б», то на основании изначально выбранных видов находят группы описаний, включающие эти виды*.

Примечание.* Существует несколько способов выделения групп описаний по набору дифференцирующих видов. Наиболее простой заключается в том, чтобы относить к такой группе все описания, в которых встретилось больше половины дифференцирующих видов.

3. Для всех видов, не вошедших в группы, рассчитывают коэффициент верности по отношению к выделенным наборам описаний.

4. Для каждой из групп видов добавляют вид с максимальным значением коэффициента верности, заново перераспределяют группы описаний на основании вновь полученных групп видов. Если проверка показывает, что новые группы описаний и видов однородны**, то добавление вида фиксируется, иначе вид изымают из набора.

Примечание.** Проверка групп видов и описаний на однородность состоит в следующем. Для всех видов, не вошедших в дифференцирующие группы, проводится расчет коэффициента верности по отношению к вновь полученным группам описаний. Если после пересчета возникает ситуация, при которой среди видов, не вошедших в группы, есть вид с коэффициентом верности большим, чем у какого-либо вида из дифференцирующей группы, то такая группа видов и описаний признается неоднородной.

5. Проводится проверка на однородность получившихся групп.
6. Шаги 3—5 повторяются до тех пор, пока либо все виды не будут в дифференцирующих группах, либо до момента, когда добавление любого вида приводит к неоднородности.
7. Последний шаг — комбинирование полученных дифференцирующих видов и соответствующих им наборов геоботанических описаний с целью нахождения растительных ассоциаций (групп ассоциаций, формаций и других единиц классификации растительности). Комбинирование дифференцирующих видов проводят с помощью логических операций (И, ИЛИ, НЕ). Например: к одной ассоциации относят все описания, в которых встретились дифференцирующие виды из группы *A*, совместно с видами из группы *B* или *C* и в которых нет видов из группы *D*.

PC-ORD

<http://home.centurytel.net/~mjm/>

Одним из наиболее распространенных и обладающих большими возможностями программных комплексов является PC-ORD, предназначенный для обработки валовых фитоценологических таблиц (рис. 7). В программе реализованы самые разнообразные алгоритмы, начиная от удаления пустых строк и столбцов и заканчивая классификацией растительности методом TWINSpan и нахождением индикаторных видов (McCune et al., 2002).

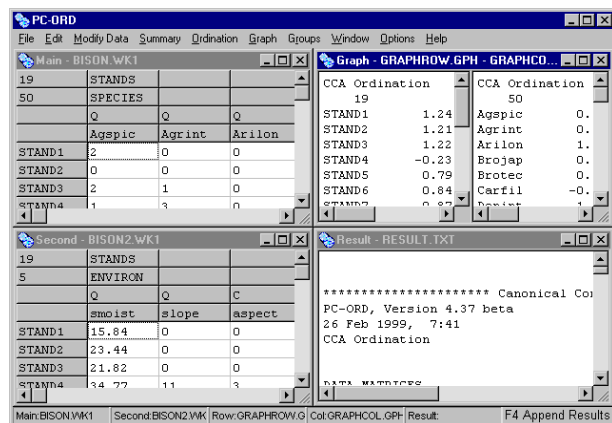


Рис. 7. Внешний вид PC-ORD.

Слева: сверху — валовая таблица геоботанических описаний, внизу — таблица дополнительных факторов; справа: сверху — текстовое представление построенных графиков, внизу — результаты обработки данных.

The PC-ORD main window.

Top left: phytosociological table; bottom left: secondary matrix of the additional factors; top right: text presentation of the created diagrams; bottom right: results of the data processing.

Важным достоинством PC-ORD является возможность использования различных экологических факторов при анализе данных. Их использование возможно как при проведении прямой ординации описаний, например, в каноническом анализе соответствий (Canonical Correspondence Analysis — CCA), так и при отображении результатов любого другого метода. Во втором случае значения выбранных экологических характеристик отображаются на построенных графиках и рисунках.

Всего реализованы такие методы ординации как принцип главных компонент, неметрическое шка-

лирование, ординация Брея—Куртиса (BRAY—CURTIS), ординация методом взаимного усреднения (reciprocal averaging), весовое усреднение (weighted averaging), анализ соответствий, несмещенный и канонический анализы соответствий. Кроме методов ординации, реализованы алгоритмы кластерного анализа: методы ближайшего и дальнего соседа, средневзвешенной кластеризации и метод Варда (Ward), метод кластеризации «гибкая бета» и др. В систему встроена программа автоматического приведения валовой таблицы к блочно-диагональному виду — TWINSpan.

Остановимся более подробно на методе выделения групп индикаторных видов — Indicator Species Analysis (Dufrene, Legendre, 1997; Mc Cune et al., 2002).

Алгоритм состоит из 3 шагов:

1. Пользователь задает разбиение всего множества геоботанических описаний, причем это разбиение может основываться как на результатах ручной обработки, так и результатах кластерного анализа или ординации.
2. Для каждого вида вычисляют индикаторное значение*, т. е. значение, которое определяет индикаторные свойства вида. Этот коэффициент будет максимальным для вида, встреченного только в описаниях одной группы, и минимальным для вида, встреченного во всех группах.

Примечание.* Индикаторное значение *i*-го вида по отношению к *j*-ой группе описаний рассчитывают следующим образом:

$$IndVal_{ij} = A_{ij} * B_{ij} * 100,$$

$$A_{ij} = \frac{NSp_{ij}}{NSp_i}; \quad B_{ij} = \frac{NSite_{ij}}{NSite_j}$$

NSp_{ij} — среднее проективное покрытие *i*-го вида в *j*-ой группе; NSp_i — сумма всех средних покрытий по группам; $NSite_{ij}$ — количество встреч *i*-го вида в *j*-ой группе; $NSite_j$ — количество описаний в *j*-ой группе. Индикаторное же значение *i*-го вида определяется как максимум среди всех индикаторных значений:

$$IndVal_i = \max(IndVal_{ij})$$

3. Проверяют статистическую значимость полученного коэффициента. Для этого проводят так называемую пермутационную (permutation) процедуру, т. е. все геоботанические описания случайным образом перемешиваются по группам и для всех видов заново считаются индикаторные значения. Так повторяется много раз (от 100 до 1000), в итоге для каждого вида получают набор индикаторных значений коэффициентов, для которых строят *z* статистику и проверяют ее достоверность.

Таким образом, для каждого вида определяют его приуроченность к априори заданным группам, и если индикаторное значение больше 25 % и оно является статистически значимым, то можно говорить об индикаторных свойствах данного вида.

BIOTAS

<http://www.ecostats.com/purchase/order.htm>

Остановимся на программе, которая не предназначена для проведения классификации растительности, однако может представлять определенный

интерес для геоботаников. Речь идет о программе BIOTAS, разработанной для визуализации и анализа пространственных данных. Другими словами, программа позволяет визуально отображать информацию о местах встреч тех или иных видов и накладывать карты соответствующей местности с целью нахождения взаимосвязей между рельефом, климатическими и другими условиями с встречаемостью тех или иных видов (рис. 8).

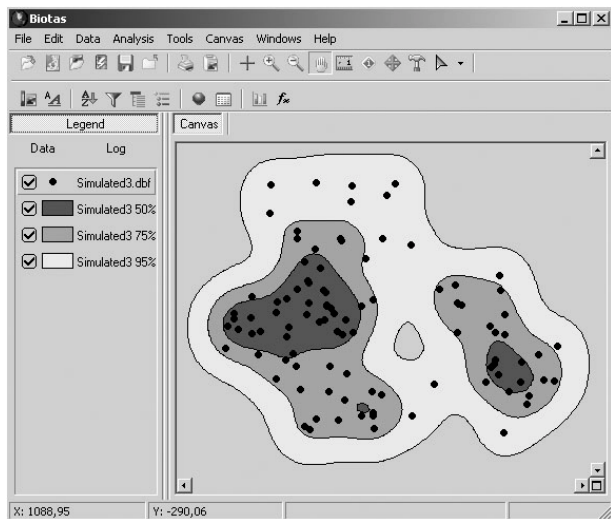


Рис. 8. Пример главного окна программы BIOTAS. Справа — построенные контуры встречаемости видов.

The BIOTAS main window; plotted areas of species occurrence on the right.

Кроме простого наложения, в BIOTAS реализованы и другие алгоритмы анализа пространственных данных:

- реализована возможность использования разнообразных фильтров (в том числе и случайного) для отбора информации для дальнейшего анализа;
- существует возможность анализа плотности пространственного распределения исследуемых объектов с помощью сеток плотности;
- введены алгоритмы проверки данных на возможность их кластеризации;
- в программе реализованы несколько алгоритмов, определяющих контуры распространения исследуемого вида с разными степенями достоверности (рис. 8);
- реализован алгоритм анализа пространственных данных в зависимости от временных параметров, например, его можно использовать при изучении миграции животных;
- введены функции определения линейной регрессии;
- в программе реализован целый ряд дополнительных функций, позволяющих рассчитывать статистические параметры (минимальное, максимальное и среднее значения, дисперсию, стандартное отклонение, медиану и др.), строить различные по форме гистограммы (обычные, круговые и т. п.).

ОТЕЧЕСТВЕННЫЕ РАЗРАБОТКИ

В России также предпринимались попытки создать программные комплексы и базы данных для хранения и автоматической обработки геоботанической информации.

К таким комплексам можно отнести: EСOPHYTO, IBIS, PRODRUMUS, KARTA и др. К большому сожалению, по этим разработкам очень мало информации, и сами программы недоступны через Интернет, так что всех заинтересовавшихся отсылаем к соответствующим статьям: Голуб и др., 1995; Зверев, 1995; Уланова, 1995; Заугольнова и др., 1995, 1999; Заугольнова, Ханина, 1996; Нешатаев В., 1997, 1999. Вопросам использования статистических подходов и баз данных в фитоценологии были посвящены специальные конференции в Ботаническом институте им. В. Л. Комарова РАН в 1995 и 1997 гг. и в Зоологическом институте РАН в 1999 и 2003 гг.

GRAPHS

Напоследок, остановимся на разработанном нами программном модуле GRAPHS. В качестве основы для модуля был взят программный продукт Microsoft Excel, входящий в состав пакета Microsoft Office, так как эта программа нашла широкое распространение у геоботаников для хранения данных и обладает гибким средством программирования VBA (Новаковский, 2004).

Модуль предназначен для проведения полуавтоматической классификации геоботанических описаний на основе использования коэффициентов сходства и графов, построенных с их использованием. Всего реализован расчет 12 коэффициентов, как качественных (Жаккара, Сьеренсена—Чекановского, Стургена—Радулеску и др.), так и количественных (Сьеренсена—Чекановского с учетом обилий, Шимкевича—Симпсона, ранговой корреляции Кэнделла и корреляции Пирсона и др.), а также коэффициентов сопряженности видов.

Для визуального отображения полученных результатов используются графы, вершинами которых являются рассматриваемые объекты (геоботанические описания или виды), а ребрами — коэффициенты сходства между ними (рис. 9). Та-

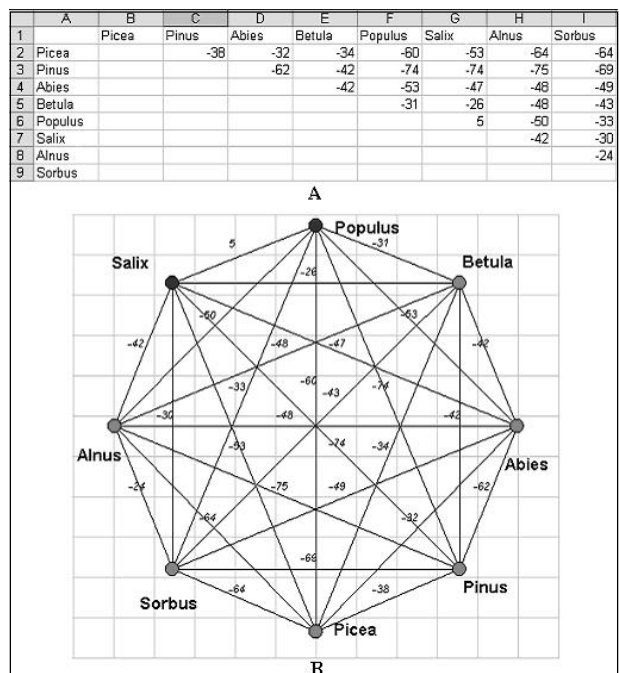


Рис. 9. Представление таблицы сходства (А) в виде графа (В) в программном модуле «GRAPHS».

A similarity table (A) and a graph (B) calculated and plotted in the «GRAPHS» module.

кой подход разрабатывался многими исследователями (Андреев, 1980; Кушнарев, 1983; Семкин и др., 1983; Шмидт 1984).

С целью облегчения восприятия полученных графов в модуле используются различные визуальные эффекты (изменение стиля, цвета, размеров вершин, длины ребер) и алгоритмы теории графов (разбиение на компоненты связности, построение дендритов и дендрограмм, представление в виде звезды). Описание используемых алгоритмов можно найти в работах, посвященных теории графов (Свами, Тхуластраман, 1984; Липский, 1988).

Для проверки работоспособности модуля проводилась его апробация на данных, собранных сотрудниками Института биологии Коми НЦ УрО РАН, Ботанического института им. В. Л. Комарова РАН и Санкт-Петербургского государственного университета.

Визуальное сравнение показало хорошее совпадение полученных графов с результатами ручной классификации (экспертной оценки), причем при использовании как метода Браун-Бланке, так и эколого-фитоценотического подхода.

В качестве иллюстрации остановимся на результатах обработки около 100 геоботанических описаний д. б. н. С. В. Дегтевой (Институт биологии Коми НЦ УрО РАН), выполненных в 10 административных районах Республики Коми. На основе этих данных с использованием программы Microsoft Excel была составлена сводная таблица для формации сероольховых лесов.

При классификации массива геоботанических описаний ассоциации выделяли на основании критерия участия в сложении нижних ярусов сообществ наиболее обильных видов (доминантов) и эколого-ценотических групп сопряженных видов. Всего

было выделено 10 ассоциаций, на каждую из которых пришлось от 3 до 18 описаний. Подробнее с их характеристиками и методикой классификации можно познакомиться в работе С. В. Дегтевой (2002).

Использование модуля «GRAPHS» дало возможность проверить, насколько экспертная оценка дифференциации растительных сообществ и их объединение в группы совпадает с результатами, полученными при независимой обработке с использованием количественных коэффициентов.

Для сравнения был построен граф в виде дендрограммы, которая строилась методом среднего. В качестве меры сходства использован коэффициент Сьеренсена—Чекановского. После построения дендрограммы вершины графа были окрашены в разные цвета в соответствии с экспертно выделенными ассоциациями (рис. 10). Граф наглядно показал, что выделенные автором группы описаний практически совпадают с результатами, получаемыми с помощью предложенного модуля.

На основании полученных результатов можно сделать вывод, что использование программного модуля «GRAPHS» значительно ускоряет процесс обработки описаний и дает вполне адекватную картину дифференциации описаний и возможности их интеграции в синтаксоны. Особенно полезной он может быть для первичного разбиения материала на гомогенные группы, объем которых соответствует ассоциациям при доминантном подходе. Для системы классификации Браун-Бланке получаемые с помощью модуля группы имеют объем примерно сравнимый с ассоциациями и субассоциациями (Новаковский, 2004).

К недостаткам модуля можно отнести невозможность отображения большого числа объектов,

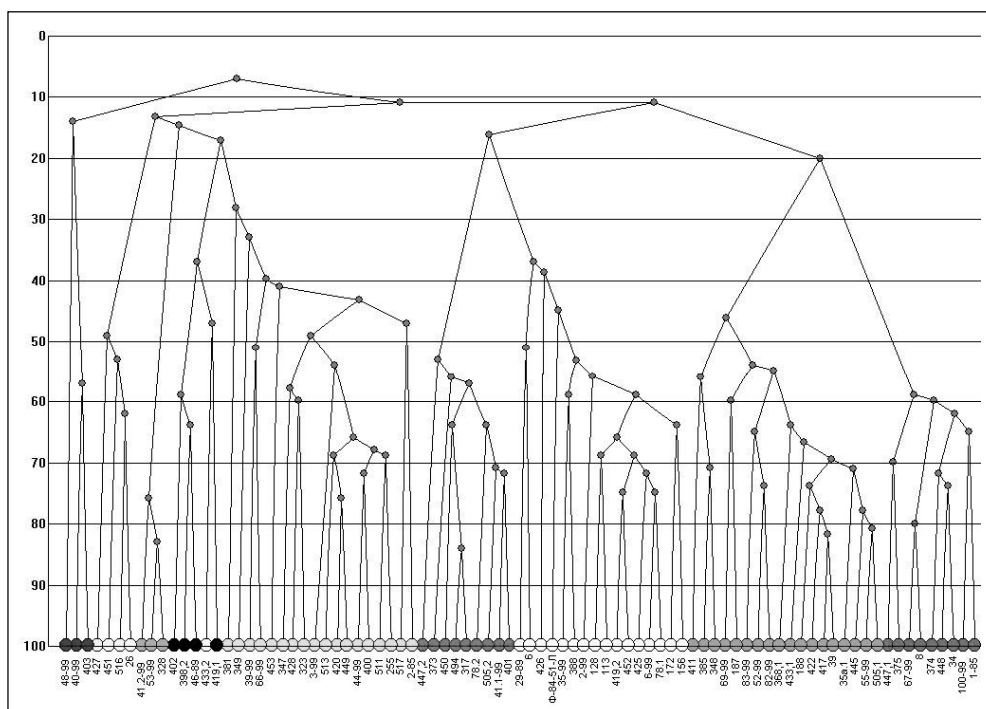


Рис. 10. Дендрограмма сходства по коэффициенту Сьеренсена—Чекановского.

По оси абсцисс — названия геоботанических описаний, по оси ординат — коэффициент сходства. Цветом выделены вершины графа, соответствующие описаниям из разных растительных ассоциаций.

The Sorensen similarity dendrogram.

Relevé names and similarity coefficient values are represented along the horizontal and the vertical axes, respectively. Graph vertexes correspondent to different associations of vegetation are marked by different colors.

например, для представления данных в виде окружности — не более 30 элементов, для дендрита и звезды — до 100, а для дендрограммы — до 200.

Выводы

На основе проведенного анализа существующих программных продуктов отметим следующие моменты. Наиболее распространенной программой, предназначенной для ввода, хранения и преобразования фитоценологической информации к другим форматам, является TURBOVEG (Hennekens, 1996; Hennekens, Schaminee, 2001). Она совместима с большинством наиболее распространенных продуктов обработки геоботанических данных. Кроме того, существуют отечественные разработки по базам данных, предназначенные для тех же целей: ЕСОРНУТО (Нешатаев В., 1997, 1999), PRODRUMUS (Заугольнова и др., 1995; Заугольнова, Ханина, 1996) и др. К сожалению, эти программы не получили достаточно широкого распространения и, по-видимому, в Интернете отсутствуют даже их демонстрационные версии.

Для целей ординации и кластеризации данных существует большое количество специализированных программ, таких как CANOCO, DECORANA и др. Однако их использование, по крайней мере, тех версий, которые нам известны, сопряжено с большими трудностями, поскольку они предназначены для работы в DOS режиме и ручной подготовки информации для анализа. Самой простой для пользователя программой, на наш взгляд, является CAP — «Community analysis package». Другой, очень распространенной, обладающей большими возможностями и достаточно легкой для использования, является программа PC-ORD. В ней, кроме различных методов ординации, реализованы дополнительные возможности обработки данных, например, анализ индикаторных видов (Dufrene, Legendre, 1997). В пользу применения PC-ORD говорит и очень хорошо разработанная система помощи, и изданная в 2002 г. книга с подробным описанием используемых методов (McCune et al., 2002).

Среди программ, предназначенных для классификации растительности методами обработки валовых геоботанических таблиц, можно еще раз выделить CAP и PC-ORD, реализующих обработку таблиц методом TWINSPAN. Кроме того, стоит отметить программу JUICE (Tichy, 2002, 2003), которая обладает простым графическим интерфейсом и развернутой системой online помощи, к тому же является бесплатной, и ее можно скачать из Интернета.

Напоследок отметим модуль GRAPHS, позволяющий отображать взаимосвязи между геоботаническими описаниями в графическом виде с использованием теории графов. Использование модуля целесообразно для визуализации данных и при не слишком большом количестве исследуемых объектов (Новаковский, 2004).

Работа выполнена при частичной финансовой поддержке проекта РФФИ-Урал (04-04-96014).

СПИСОК ЛИТЕРАТУРЫ

- Андреев В. Л. 1980. Классификационные построения в экологии и систематике. М. 142 с.
- Василевич В. И. 1969. Статистические методы в геоботанике. Л. 230 с.
- Голуб В. Б., Рухленко И. А., Халеев А. Е. 1995. Пакет программ для обработки геоботанических данных «ФИТОЦЕНОЛОГ» // II Совещ. «Компьютерные базы данных в ботанических исследованиях», СПб., БИН РАН, 1995: Тез. докл. СПб. С. 13—14.
- Дегтева С. В. 2002. Сероольшаники Республики Коми // Бот. журн. Т. 87. № 1. С. 107—121.
- Заугольнова Л. Б., Ханина Л. Г. 1996. Опыт разработки и использования баз данных в лесной фитоценологии // Лесоведение. № 1. С. 76—83.
- Заугольнова Л. Б., Ханина Л. Г., Комаров А. С. 1995. Информационно-аналитическая система для определения сукцессионного статуса лесных сообществ. Препринт. Пушкино. 51 с.
- Заугольнова Л. Б., Ханина Л. Г., Глухова Е. М. 1999. Создание базы данных и информационно-диагностической системы для определения синтаксономических адресов лесных сообществ Европейской России // Информационно-поисковые системы в зоологии и ботанике: Междунар. симпоз. СПб. С. 84. (Тр. ЗИН РАН; Т. 278).
- Зверев А. А. 1995. Опыт разработки интегрированной ботанической информационной системы (IBIS) // II Совещ. «Компьютерные базы данных в ботанических исследованиях». СПб., БИН РАН, 1995: Тез. докл. СПб. С. 19—20.
- Ким Дж.-О., Мьюллер Ч. У. 1989. Факторный анализ: статистические методы и практические вопросы // Факторный, дискриминантный и кластерный анализ / Под ред. И. С. Енюкова. М. С. 5—77.
- Кушнарев Е. Л. 1983. Применение графических методов анализа при изучении видовой структуры орнитосообществ по экологическому профилю // Теоретико-графовые методы в биогеографических исследованиях. Владивосток. С. 33—48.
- Лебедева Н. В., Криволуцкий Д. А. 2002. Биологическое разнообразие и методы его оценки. М., 432 с. <http://nature.ok.ru/biodiversity/book3.html>
- Липский В. 1988. Комбинаторика для программистов. М. 213 с.
- Миркин Б. М., Наумова Л. Г., Соломещ А. И. 2000. Современная наука о растительности. М. 263 с.
- Нешатаев В. Ю. 1997. Метод упорядочивания фитоценологических таблиц в Информационно-статистической системе ЕСОРНУТО // Междунар. конф. «Компьютерные базы данных в ботанических исследованиях»: Тез. докл. СПб. С. 73—75.
- Нешатаев В. Ю. 1999. Геоботаническая информационно-статистическая система ЕСОРНУТО // Организация научных исследований в заповедниках и национальных парках. М. С. 208—223.
- Нешатаев Ю. Н. 1987. Методы анализа геоботанических материалов. Л. 190 с.
- Новаковский А. Б. 2004. Возможности и принципы работы программного модуля «GRAPHS». Сыктывкар. 28 с. (Препринт; сер. «Автоматизация научных исследований»).
- Пузаченко Ю. Г. 2004. Математические методы в экологических и географических исследованиях. М. 416 с.
- Свами М., Тхуласираман К. 1984. Графы, сети и алгоритмы. М. 454 с.
- Семкин Б. И., Наумова В. В., Варченко Л. И., Борзова Л. М. 1983. Об использовании мер зависимости для определения степени ассоциации видов растений // Теоретико-графовые методы в биогеографических исследованиях. Владивосток. С. 20—32.
- Ту Дж., Гонсалес Р. 1978. Принципы распознавания образов. М. 416 с.

- Уиллиамс У. Т., Ланс Дж. Н. 1986. Методы иерархической классификации // Статистические методы для ЭВМ. М. С. 269—301.
- Уланова Н. Г. 1995. Математические методы в геоботанике. М. 109 с.
- Шутиков В. К., Розенберг Г. С., Зинченко Т. Д. 2003. Количественная гидроэкология: методы системной идентификации. Тольятти. 463 с. <http://www.tolcom.ru/kiril/Library/Book1/content0/content0.htm>
- Шмидт В. М. 1984. Математические методы в ботанике. Л. 287 с.
- McCune B., Grace J. B., Urban D. L.. 2002. Analysis of ecological communities. MjM Software Design, Oregon. 285 p.
- Bruilheide H. 1997. Using formal logic to classify vegetation // *Folia Geobot. Phytotax.* Vol. 32. N 1. P. 41—46.
- Bruilheide H. 2000. A new measure of fidelity and its application to defining species groups // *J. Veg. Sci.* Vol. 11. P. 167—178.
- Dufrene M., Legendre P. 1997. Species assemblages and indicator species: the need for a flexible asymmetrical approach // *Ecological monographs.* Vol. 67. N 3. P. 345—365.
- Hennekens S. M. 1996. TURBO(VEG). Software package for input, processing and presentation of phytosociological data. IBN-DLO. Wageningen, NL and University of Lancaster. Lancaster, UK. 59 p.
- Hennekens S. M., Schaminee J. H. J. 2001. TURBOVEG, a comprehensive database management system for vegetation data // *J. Veg. Sci.* Vol. 12. P. 589—591. <http://www.synbiosys.alterra.nl/turboveg/article.pdf>
- Hill M. O. 1994. DECORANA and TWINSpan for ordination and classification of multivariate species data: a new edition, together with supporting programs, in FORTRAN 77. Institute of Terrestrial Ecology. Huntingdon, UK. 58 p.
- Jongman R. H. G., Ter Braak C. J. F., Van Tongeren O. F. R. 1987. Data analysis in community and landscape ecology. Wageningen. 299 p.
- Legendre P., Legendre L. 1998. Numerical Ecology. 2nd ed. Amsterdam. 853 p.
- Tichy L. 2002. JUICE, software for vegetation classification // *J. Veg. Sci.* Vol. 13. P. 451—453. <http://www.sci.muni.cz/botany/juice/jvs.pdf>
- Tichy L. 2003. Formalized reproduction of an expert-based phytosociological classification: A case study of sub-alpine tall-forb vegetation // *J. Veg. Sci.* Vol. 14. P. 601—610. <http://www.sci.muni.cz/botany/juice/JVS2003b.pdf>

Получено 30 марта 2005 г.

SUMMARY

The paper reports about the most widespread software used for the analysis of vegetation data. Such software packages as TURBOVEG, TWINSpan, CANOCO, SYNTAXON, CAP, JUICE, PC-ORD, BIOTAS are briefly overviewed and the author's module «GRAPHS» is additionally described. Capabilities, using algorithms, merits and demerits are considered for all the programs discussed. The Web links to creator's pages are quoted where possible.