

ГОСУДАРСТВЕННОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ
«ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»

**ФУНДАМЕНТАЛЬНЫЕ
И ПРИКЛАДНЫЕ АСПЕКТЫ
СОВРЕМЕННОЙ БИОЛОГИИ**

**Первая Всероссийская
молодежная научная конференция, посвященная
125-летию биологических исследований
в Томском государственном
университете**

(Томск, 6–9 октября 2010 г.)



Издательство Томского университета

2010

ПРИМЕНЕНИЕ МЕТОДОВ МНОГОМЕРНОГО ШКАЛИРОВАНИЯ В СРАВНИТЕЛЬНОЙ ФЛОРИСТИКЕ¹

А.С. Мочалов

Раскрывается общее понятие метода многомерного шкалирования и приводится пример его применения при сравнении флор.

APPLICATION OF MULTIDIMENSIONAL SCALING IN COM- PARATIVE FLORISTICS

A.S. Mochalov

Reveals the general concept of multidimensional scaling method and an example of its use in comparing floras.

В связи с возросшими возможностями использования компьютеров в биологических исследованиях большое значение получают неметрические методы машинного анализа плохо определенных так называемых «рыхлых» структур и представление информации о них в форме, приспособленной для восприятия исследователем [1]. К этой области относятся, в частности, методы многомерного шкалирования, позволяющие «изображать» систему объектов с матрицей расстояний или сходства произвольной природы и размерности совокупностью точек в M -мерном пространстве с минимальным искажением, в смысле того или иного критерия, первоначальной матрицы расстояний. При M , равной, 2 объекты представляются в виде точек на плоскости, что очень удобно для визуального анализа. Первоначально эти методы были разработаны для финансовых расчетов, наиболее подробно суть многомерного шкалирования изложена М. Дейвисоном [2]. В биологии этот способ применим, в первую очередь, для анализа корреляционных матриц, так как позволяет изобразить отношения между признаками в наглядной форме и служит хорошим дополнением к другим, более формальным средствам их исследования [1].

Допустим, имеется конечное множество объектов X и матрица R расстояний или мер сходства между ними, а также произвольное представление S объектов множества X в виде точек на плоскости с координатами $x_1, y_1, \dots, x_N, y_N$. Введем расстояние между точками i и j :

$$c_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} .$$

¹ Исследование выполнено при поддержке РФФИ (проект № 10-04-90835 – моб_ст).

Определим критерий различия между множеством X и его представлением на плоскости в виде

$$H(X, C) = \sum_{i,j} (f(r_{ij}) - c_{ij})^2,$$

где f – некоторое монотонное преобразование.

В методах многомерного шкалирования ищется такое представление C^* , для которого функция H принимает наименьшее возможное значение. Это приводит к задаче минимизации H как функции многих переменных $x_1, \dots, x_N, y_1, \dots, y_N$. Формальные способы расчета достоверности в данном случае отсутствуют, и размерность пространства представлений можно выбирать произвольно. Формальных критериев выбора размерности не существует [1].

Помимо этого, для получения более наглядных и убедительных результатов при обработке матрицы дивергенции можно использовать одновременно методы K -средних и многомерного шкалирования [1]. Весьма наглядные результаты показывает метод многомерного шкалирования в комбинации с корреляционным анализом. В частности, используя этот алгоритм, можно показывать наиболее характерные для видов признаки (Ефимов, устное сообщение). При этом принадлежность исследуемого образца к тому или иному виду рассматривается как дополнительный признак (0 – образец не принадлежит виду, 1 – образец принадлежит к виду).

Перспективы применения методов многомерного шкалирования очень высоки, они служат достойной альтернативой не только более популярным пока дискриминантному и факторному анализам, результаты которых во многих случаях не пересекаются с многомерным шкалированием, но и методам парногруппового среднеарифметического связывания в сравнительной флористике, результаты которого в значительной степени дополняют. Нами использовалось двумерное шкалирование при сравнительном анализе птеридофлор Урала. Все расчеты проводились в программе PAST, для построения более наглядных графиков в некоторых случаях использовался пакет программ STATISTICA 6.0.

При сравнительном анализе флор (птеридофлор) методами многомерного шкалирования таблица присутствия-отсутствия видов представляется в виде нулей и единиц (0 – отсутствие вида во флоре, 1 – присутствие вида во флоре). Полученная матрица транспонируется таким образом, что для каждой конкретной флоры (в нашем случае птеридофлоры) присутствие или отсутствие каждого вида становится признаком, имеющим соответствующее значение 0 или 1. Используя метод двумерного шкалирования, выбираем меру сходства (коэффициент Жаккара или меру включения Симпсона) и получаем координаты каждой конкретной флоры на осях X и Y и минимальное расстояние между ними, по которым строим график и оцениваем связь флор (птеридофлор). Методы многомерного шкалирования более чем другие удачно подходят для реализации идей Б.А. Юрцева [3, 4] о неравнозначности видов во флорах. В перспективе при таком подходе к анализу не обязательно ограничиваться лишь абсолютными значениями (0 или 1), вполне возможен сравнительный анализ с оцененным весом того или иного вида в каждой конкретной флоре. Например, уникальные виды для конкретной флоры будут иметь больший вес по сравнению с банальными или наоборот.

Анализируя сравнение птеридофлор Урала и прилегающих равнинных территорий, можно сделать вывод о более показательном отображении результатов многомерного шкалирования по сравнению с парногрупповым среднеарифметическим связыванием и более точные результаты (рис. 1, 2).

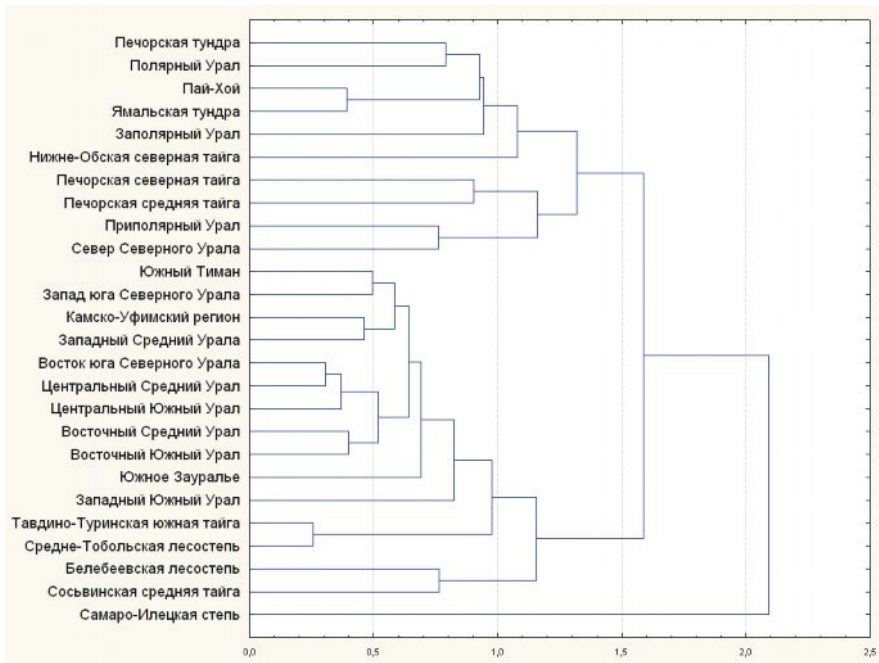


Рис. 1. Граф сходства районов птеридофлоры Урала, построенный методом парногруппового среднеарифметического связывания с использованием коэффициента Жаккара

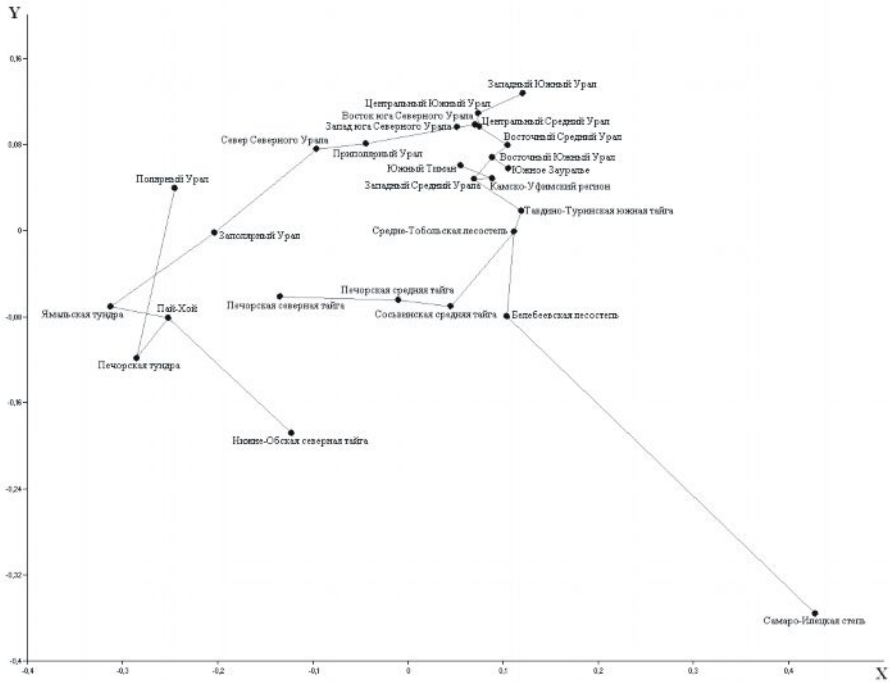


Рис. 2. Граф сходства районов птеридофлоры Урала, построенный методом двумерного шкалирования с использованием коэффициента Жаккара

Так при парногрупповым среднеарифметическом связывании птеридофлоры Белебеевской лесостепи и Сосьвинской средней тайги попадают в один кластер, что достаточно трудно объяснить. Но при многомерном шкалировании видно, что связь этих птеридофлор осуществляется только через птеридофлору Средне-Тобольской лесостепи, что вполне логично.

Литература

1. *Ефимов В.М., Ковалева В.Ю.* Многомерный анализ биологических данных: Учеб. пособие. Томск: ТГУ, 2005. 95 с.
2. *Дэйвисон М.* Многомерное шкалирование. М.: Финансы и статистика, 1988. 254 с.
3. *Юрцева Б.А.* Эколого-географическая структура биологического разнообразия и стратегия его учета и охраны // Биологическое разнообразие: подходы к изучению и сохранению. СПб., 1992. С. 7–21.
4. *Юрцева Б.А.* Предисловие // Изучение биологического разнообразия методами сравнительной флористики. СПб.: НИИХ СПбГУ, 1998. С. 3–9.