

Многомерный статистический анализ экологических сообществ (обзор)

© 2019. В. К. Шитиков, д. б. н., с. н. с.,
Т. Д. Зинченко, д. б. н., профессор, зав. лабораторией,
Институт экологии Волжского бассейна РАН,
445003, Россия, Самарская обл., Тольятти, ул. Комзина, 10,
e-mail: stok1946@gmail.com

Современный подход к изучению экологии сообществ предполагает интеграцию и совместную обработку больших массивов наблюдений. При этом обычно привлекаются разнообразные показатели (популяционные, фенотипические, генетические, химические, ландшафтно-географические), которые характеризуются значительной временной и пространственной изменчивостью. Целью проводимого статистического анализа является выявление значимых связей таксономической структуры с характеристиками биотопов и факторами окружающей среды. В его основе обычно лежат многомерные методы, позволяющие выполнять оптимальное проецирование данных с большим числом переменных в пространствах с малой размерностью.

В статье подробно рассматривается эволюция алгоритмов многомерного анализа, начиная с классической непрямой ординации на основе главных компонент (PCA) до современных интегрированных симметричных методов, нашедших применение в *omics*-технологиях. Выделяется класс алгоритмов, основанных на вычислении матриц дистанций, таких как метрическое (PCoA) и неметрическое (NMDS) многомерное шкалирование, а также рассматриваются достоинства и недостатки их использования. Обсуждается зависимость результатов анализа избыточности (RDA) и канонического анализа соответствий (CCA) от характера распределения исходных данных и даются рекомендации по их предварительной трансформации.

Показана роль таких симметричных методов, как двухблочный алгоритм частных наименьших квадратов (2B-PLS) и анализ совместной инерции (CIA), которые путём разложения по осям многомерной ковариации позволяют установить, какие виды из разных комплексов наблюдений в наибольшей мере ассоциированы друг с другом. Прокрустов анализ (PCIA) широко используется для выявления изменений видового состава изучаемого региона до и после некоторого события (например, антропогенного воздействия). Обобщённые алгоритмы прокрустова и канонического анализа (GPA, RGCCA, DIABLO) работают с большим числом таблиц и позволяют исследовать динамику структуры сообществ за несколько последовательных периодов времени, а также наилучшим образом сформировать консенсусную конфигурацию.

Даны ссылки на многочисленные примеры использования методов ординации в отечественной и зарубежной литературе. Показаны основные перспективы развития многомерных методов в экологии сообществ.

Ключевые слова: многомерная ординация, анализ главных компонент, анализ соответствий, неметрическое шкалирование, канонический анализ, двухблочные алгоритмы, прокрустов анализ, анализ совместной инерции.

Multivariate statistical analysis of ecological communities (review)

© 2019. V. K. Shitikov ORCID: 0000-0002-8385-1913*
T. D. Zinchenko ORCID: 0000-0002-3808-4700*
Institute of Ecology of the Volga River Basin of the Russian Academy of Sciences,
10, Komzina St., Togliatti, Samara Region, Russia, 445003,
e-mail: stok1946@gmail.com

The modern approach to the study of communities' ecology involves the integration and joint processing of large arrays of observations. This usually involves a variety of indicators (population, phenotypic, genetic, environmental, chemical, landscape and geographical) which are characterized by significant temporal and spatial variability. The purpose of the analysis is to identify significant statistical relationships of the taxonomic structure with the characteristics of biotopes and environmental factors. It is based on multivariate methods that allow optimal projection of data with a large number of variables into low-dimensional spaces.

The article focuses on the evolution of algorithms for multivariate analysis, starting with the classic unconstrained ordination based on principal components (PCA) up to modern integrated symmetric methods used in *omics* technologies. A class of algorithms, such as metric (PCoA) and non-metric (NMDS) multidimensional scaling, based on the calculation

of distance matrices, is distinguished and the advantages and disadvantages of their use are considered. The dependence of the results of redundancy analysis (RDA) and canonical correspondence analysis (CCA) on the distribution law of the empirical data is discussed and recommendations for their preliminary transformation are given.

It is shown the role of such symmetric methods as the two-block algorithm of partial least squares (2B-PLS) and the co-inertia analysis (CIA), which allow to establish by decomposition on axes of multidimensional covariations what species from different complexes of observations are most associated among themselves. Procrustean analysis (PCIA) can be widely used to identify changes in the species composition of the study region before and after some event (e.g., anthropogenic impact). Generalized Procrustean algorithms and canonical analysis (GPA, RGCCA, DIABLO) allow you to work with a large number of tables and explore the dynamics of community structure for several sequential periods of time or to form a consensus configuration by the best way.

References to numerous examples of the use of ordination methods in domestic and foreign literature are given. The main prospects and directions of development of multidimensional methods in relation to the ecology of communities are shown.

Keywords: multivariate ordination, principal component analysis, correspondence analysis, non-metric scaling, canonical analysis, two-block algorithms, Procrustean analysis, co-inertia analysis.

Современный подход к изучению экосистем основан на статистическом анализе множества разнообразных факторов (популяционных, фенотипических, генетических, химических, ландшафтно-географических), которые характеризуются значительной временной и пространственной изменчивостью, обусловленной сезонными, климатическими, антропогенными или иными причинами. Основными задачами анализа экологии сообществ являются оценка значимости межвидовых связей внутри ассоциаций взаимодействующих организмов и выявление механизмов воздействия всей совокупности внешних факторов на структурно-функциональные особенности изучаемых биоценозов [1]. Конечная цель исследования – прогноз динамики развития экосистем (с точки зрения повышения продуктивности и разнообразия) при различных сценариях природоохранной деятельности.

Анализ главных компонент [2] и дискриминантный анализ [3] положили начало широкому применению многомерных статистических исследований в различных сферах науки и техники. С 70-х годов XX века начался активный рост арсенала этих методов и успешная их адаптация к обработке экологических данных [4, 5]. При этом важное место в экологии сообществ заняла **ординация** – совокупность методов распознавания образов для анализа упорядоченности (нем. *Ordnung*) изучаемых объектов путём их оптимального проецирования в пространство малой размерности [6, 7].

В XXI веке сфера применения алгоритмов многомерного анализа постоянно расширяется. Решаются всё более сложные проблемы количественной оценки соответствия одновременно нескольких многомерных массивов данных, а также вклада взаимодействий между факторами в совокупную дисперсию в услови-

ях мультиколлинеарности переменных. Это обусловило появление группы новых методов, основанных на симметричном анализе ковариаций (*covariance-based methods*), прокрустовом анализе, канонических корреляциях и т. д. [8]. Бурное развитие *omic*-исследований потребовало разработки мощных алгоритмов выделения биомаркеров с использованием латентных структур [9].

Целью настоящей статьи явилась систематизация практических рекомендаций по проведению многомерного статистического анализа данных в экологии сообществ, что аналогично работе [8] в области химической экологии. Главной задачей было обсуждение ключевых концепций отдельных методов, сфер их применения, выявленных достоинств и недостатков, а также особенностей интерпретации получаемых результатов.

Все представленные нами методы реализованы в программном обеспечении свободно распространяемой статистической среды R [10]. Примеры скриптов, позволяющих самостоятельно выполнить основные процедуры многомерного анализа, можно найти в работах [8, 11], а детальное описание используемого математического аппарата – в монографии [12].

Классификация методов ординации и интерпретация результатов

Все методы многомерной ординации сводятся к синтезу некоей оптимальной информационной структуры, состоящей из осей новых латентных переменных, которые наилучшим образом «объясняют» общую вариацию набора данных. Поскольку 2–3 главных оси концентрируют основную долю дисперсии и являются ортогональными, то появляется возможность визуализации многомерного

облака точек, отображая их, например, на плоскости с минимальными искажениями при проецировании.

В общем случае ординация выполняется с использованием одной или нескольких таблиц с одинаковым числом строк, представляющих объекты, на которых проводились наблюдения (местообитания, биотопы, створы рек, временные периоды и проч.). Основной является таблица, которая описывает изучаемое экологическое сообщество и в своих столбцах содержит значения популяционной плотности того или иного вида, обнаруженного в каждом местообитании (матрица **Y**). Для каждой пары строк матрицы наблюдений **Y** могут быть рассчитаны расстояния в многомерном пространстве видов, которые составляют матрицу дистанций **D**. Наконец, для каждого местообитания могут быть определены значения параллельно измеряемых внешних абиотических факторов (матрица **X**). Основное различие между отдельными многомерными методами заключается в том, сколько и какие матрицы исходных данных используются в анализе.

Непрямая или «необъяснённая» ординация (*unconstrained ordination*) выполняется с использованием только одной матрицы **Y** «виды – местообитания» и ставит своей целью получить отображение в ортогональной системе координат структурных закономерностей изучаемого сообщества в форме графических проекций распределения популяций по биотопам. Результатом ординации обычно является совмещённая диаграмма видов и/или местообитаний в координатах двух главных осей, на которой обычно легко выделить кластеры этих объектов, оценить хорошо интерпретируемые закономерности взаимосвязей между выделенными группами или сравнить ординационные графики за разные периоды времени.

Прямая или объяснённая (*constrained*) ординация ставит своей целью связать внутреннюю изменчивость видовой структуры сообщества **Y** с теми или иными факторами внешних воздействий **X**. Концептуально эта форма многомерного анализа, называемая канонической [13], рассматривается как расширение идей регрессионного анализа для моделирования многомерного отклика $Y = f(X)$ при одновременном использовании алгоритмов редукции данных. Выполнение прямой ординации для интерпретации всегда более предпочтительно, но требует сбора необходимой исходной информации. Следует уточнить, что непрямая ординация всегда является одним из этапов прямого анализа, поэтому их нельзя

противопоставлять ни в методическом, ни в вычислительном аспекте.

Результатом прямой ординации является проведение на ординационной диаграмме дополнительных осей («стрелок»): каждая из них соответствует одному из внешних факторов среды из матрицы **X**, а её длина отражает уровень воздействия на структуру экологического сообщества. Интерпретация ординационных диаграмм основана на понятии *корреляционного круга*: угол между двумя стрелками факторов (или между стрелкой и осью главных координат) указывает на величину корреляции между этими двумя переменными. Аналогичную оценку корреляций можно выполнить для каждого вида или местообитания, если мысленно соединить стрелкой соответствующую точку с началом координат. Чем ближе друг к другу концы сравниваемых стрелок, тем теснее связь между соответствующими объектами и/или переменными [14].

Если прямая ординация является асимметричным анализом (т. е. выполняется поиск зависимости **Y** от **X**, но не наоборот), то разработаны методы анализа, совершенно равноправного по отношению к обоим наборам данных, и с этой точки зрения названного *симметричным*. Если на вход алгоритма подаются две матрицы типа **Y** «виды – местообитания», то их можно сопоставить между собой, оценив степень расхождений одних и тех же точек, и отобразить эти изменения на специальной ординационной диаграмме. Таким образом, в частности, можно оценить модификацию видового состава одного и того же региона до и после некоторого события (например, антропогенного воздействия). Обобщённые алгоритмы прокрустова и канонического анализа позволяют работать с большим числом таблиц и исследовать динамику структуры сообществ за несколько последовательных периодов времени.

В таблице приведена сводка наиболее известных многомерных методов статистического анализа, широко используемых в экологии, генетике и биохимии. Полностью библиографические описания приведены в источнике [8].

Непрямая и прямая ординации

Классическими методами не прямой ординации являются анализ главных компонент и анализ соответствий, использующие непосредственно таблицы обилия видов **Y**, а также два алгоритма многомерного шкалирования, основанные на предварительно рассчитывае-

Обзор многомерных методов статистического анализа
Overview of multivariate statistical methods of analyses

Тип анализа Type of analyses	Данные Data sets	Аббревиатура Abbreviations	Наименование метода, авторы, год Name of method, authors, year
Непрямая ординация / Unconstrained ordination	Y	PCA	Анализ главных компонент / Principal Component Analysis – Goodall, 1954 [15]
		CA	Анализ соответствий / Correspondence Analysis – Hill, 1973 [16]
		DCA	Анализ соответствий с удалённой «аркой» / Detrended Correspondence Analysis – Hill, Gauch, 1979 [17]
	D	PCoA (MDS)	Анализ главных координат = Метрическое многомерное шкалирование / Principal Coordinate Analysis = Metric Multidimensional Scaling – Gower, 1971 [18]
		NMDS	Неметрическое многомерное шкалирование / Non-metric Multidimensional Scaling – Anderson, 1971 [19]
Прямая ординация / Constrained ordination	X, Y	CCA	Канонический анализ соответствий / Canonical Correspondence Analysis – ter Braak, 1986 [20]
		RDA	Анализ избыточности / Redundancy Analysis – Rao, 1964 [21]
	X, D	db-RDA	Анализ избыточности, основанный на дистанциях / Distance-based Redundancy Analysis – Legendre, Anderson, 1999 [22]
	X, Y > 2	RGCCA	Обобщённый канонический корреляционный анализ с регуляризацией / Regularized Generalized Canonical Correlation Analysis – Tenenhaus, Tenenhaus, 2011 [23]
Дискриминантный анализ и PLS-регрессия / Discriminant Analysis and PLS-regression	X, Y	LDA	Линейный (канонический) дискриминантный анализ / Linear (Canonical) Discriminant Analysis – Fisher, 1936 [3]
		PLS	Регрессия с использованием частных наименьших квадратов = проецирования латентных структур / Partial Least Squares = Projection to Latent Structures – Wold et al., 1983 [24]
		OPLS-DA	Дискриминантный ортогональный анализ методом PLS / Orthogonal Partial Least Squares Discriminant Analysis – Bylesjö et al., 2006 [25]
		2B-PLS	Двублочный метод частных наименьших квадратов / Two-Block Partial Least Square – Sampson et al., 1989 [26]
	X, Y > 2	DIABLO	Интегрированный анализ данных с использованием латентных структур / Data Integration Analysis for Biomarker discovery using a Latent component for Omics studies – Singh et al., 2016 [27]
Дисперсионный анализ ANOVA	X, Y	ANOVA-PCA	Многомерный метод дисперсионного анализа с использованием главных компонент / ANOVA-Principal Component Analysis – Harrington et al., 2005 [28]
		rMANOVA	Многомерный ANOVA с регуляризацией / Regularized Multivariate ANOVA – Engel et al., 2015 [29]
Анализ совместной инерции / Co-Inertia Analysis	X, Y	CIA	Анализ совместной инерции / Co-inertia Analysis – Dolédec, Chessel, 1994 [30]
		PCIA	Прокрустовый анализ совместной инерции / Procrustean Co-inertia Analysis – Dray et al., 2003 [21]
	X, Y > 2	MCIA	Множественный анализ совместной инерции / Multiple Co-inertia Analysis – Chessel, Hanafi, 1996 [32]
		GPA	Обобщенный прокрустовый анализ / Generalized Procrustes analysis – Gower, 1975 [33]

Примечание: Y – таблица популяционной плотности видов; D – матрица дистанций, X – таблица факторов среды.
Note: Y – data set of abundance values of species; D – distance matrix, X – data set of environmental variables.

мой матрице дистанций D . Все они выполняют примерно одну и ту же задачу: сформировать в многомерном пространстве исходных данных новые оси оптимального проецирования и представить обобщённую информацию о видовой структуре сообщества в максимально компактной и удобной для визуализации форме. Выбор конкретного метода зависит, как правило, от характера решаемой биологической проблемы и статистических особенностей распределения данных в матрице Y .

Анализ главных компонент PCA [2, 15] работает корректно, если показатели обилия видов имеют приблизительно нормальное (по крайней мере, симметричное) распределение и связаны между собой линейными отношениями. Большое количество нулевых значений, характерное для многовидовых ассоциаций, сильно вредит достоверности результатов анализа.

Факториальный *анализ соответствий* CA [8, 16] работает с таблицами сопряжённости, каждая клетка которых содержит нестандартизованные частоты (число экземпляров каждого вида) или метки присутствия-отсутствия вида 0/1. При большом числе видов, характерном для многих экологических сообществ, расстояния между точками в пространстве, образуемом осями соответствия CA, могут быть серьёзно искажены по сравнению с фактическими расстояниями между объектами в исходной таблице данных. Разработаны методы компенсации таких искажений, например, «эффекта арки» [17], но они попутно могут «сглаживать» и реально существующие зависимости.

Методы PCA и CA непосредственно основаны на эмпирических таблицах популяционной плотности и используются для сравнительной оценки значимости отдельных видов. Однако, если задача ограничена оценкой общего видового сходства, пренебрегая при этом ролью отдельных таксонов, то целесообразно использовать методы, основанные на матрице расстояний D . Они не требуют явных предположений о характере распределения данных и формируют стабильные, хорошо интерпретируемые диаграммы.

Определённой проблемой использования таких методов как *анализ главных координат* (PCoA) [12, 18] и *неметрического многомерного шкалирования* (NMDS) [19] является отсутствие общепринятой методики выбора формулы для оценки расстояний. По результатам проведённого анализа [14, 34] при формировании матрицы дистанций можно

порекомендовать для количественных данных меру сходства Брея–Кёртиса или её «бинарный аналог» – коэффициент Сьёренсена для качественных данных.

Такие методы прямой ординации как *анализ избыточности* (RDA) [13] и *канонический анализ соответствий* (CCA) [20] формируют модель зависимости многомерного отклика Y от набора количественных независимых переменных X . Вначале многомерная изменчивость данных видовой структуры Y разлагается на две составляющие: (а) объясняемую вариацию, вызываемую влиянием внешних переменных, и (б) остаточную или необъясненную дисперсию. Далее реализуются два отдельных анализа, формирующих оси латентных координат в направлении максимума обеих оценок дисперсий: PCA в случае RDA или CA в случае CCA. Если массив обилия видов не отвечает предположениям, постулируемым для этих методов, то его рекомендуется преобразовать в матрицу D и использовать *анализ избыточности на основе матрицы расстояний* (db-RDA) [22].

Дискриминантный анализ с использованием регрессии на основе *частных наименьших квадратов* PLS-DA [24] выполняется, если массив X состоит из одной качественной переменной (группирующего фактора). В отличие от классического дискриминантного анализа, алгоритм позволяет обрабатывать большое количество коррелированных переменных (больше, чем число объектов). Полученная модель может выполнять прогнозирование, т. е. рассчитывать вероятности отнесения произвольного объекта к каждой из групп.

Симметричный анализ двух или нескольких матриц

Прокрустовый анализ совместной инерции PCIA [31] сравнивает две матрицы сообществ с произвольным составом видов, но с одинаковым набором местообитаний, либо две ординации, выполненные отдельно по каждой таблице. На сформированной факториальной плоскости, состоящей из двух осей, каждый объект отображается двумя точками, соединёнными стрелкой, обозначающей переход биотопа из начального состояния в конечное. Чем короче стрелки на диаграмме, тем больше это указывает на совпадение структуры сообществ, описанных двумя таблицами.

Обобщённый прокрустовый анализ GPA [33] обрабатывает несколько матриц со-

обществ с произвольным составом видов, но с одинаковым набором местообитаний. Цель анализа состоит в том, чтобы наилучшим образом сформировать *консенсусную конфигурацию*, т. е. «среднюю» из частных ординаций, соответствующих каждой таблице.

Двухблочный метод частных наименьших квадратов 2B-PLS [26] и *анализ совместной инерции* CIA [30] используются для анализа соответствия между двумя матрицами сообществ с произвольным составом видов, но с одинаковым набором местообитаний. Методы формируют информационную структуру (т. е. набор латентных осей), которая оптимально ориентирована относительно многомерной ковариации между двумя таблицами.

Обобщённый канонический корреляционный анализ с регуляризацией RGCCA [23] и *интегрированный анализ данных* с использованием латентных структур DIABLO [27, 32] применяется для обобщения информации из несколько матриц сообществ с произвольным составом видов, но одинаковым набором местообитаний. Цель анализа чаще всего состоит в том, чтобы определить, какие виды из разных таблиц в наибольшей мере ассоциированы между собой.

Заключение

Многомерные методы ординации, описанные выше, обеспечивают эффективное решение трёх важнейших задач в экологии сообществ: (а) идентификацию экологических градиентов или биологически значимых кластеров, объединяющих местообитания и таксоны организмов, составляющих изучаемое сообщество; (б) выделение основных внешних факторов, определяющих изменчивость структуры сообществ; (в) сопоставление двух или нескольких массивов данных, содержащих описания изучаемой экосистемы, сделанных в разное время или с использованием разных таксономических групп, с оценкой статистической значимости имеющихся структурных изменений.

Однако большое количество представленных выше статистических методов создают значительную неопределённость в выборе оптимальной стратегии анализа, чему способствуют «вводящие в заблуждение рекомендации и ошибочные обобщения относительной эффективности доступных многомерных методов...» [7]. Например, не всегда оправдано негативное отношение некоторых авторов к неметрическому многомерному шкалированию (NMDS), который, как показывает опыт [14],

является наиболее устойчивым и интерпретируемым методом ординации в исследованиях по экологии речных сообществ. За последнее время развивается система современных статистических тестов, позволяющих выполнять количественную проверку достоверности построенных моделей с последующей селекцией лучших из них. Использование методов Монте-Карло, рандомизации и кросс-проверки, которые остались за рамками нашего обзора, позволяют придать самому понятию «заблуждение» более корректный доказательный смысл [11, 35].

Современные планы экспериментальных и природных исследований всё больше приобретают междисциплинарный характер и предполагают тесное взаимодействие экологов, генетиков, химиков и биостатистиков. В этих условиях роль многомерных методов, позволяющих вести обработку массивов информации с большим числом разнородных переменных становится очевидной и приоритетной.

References

1. Rozenberg G.S. Introduction to the theoretical ecology (in two volumes). Togliatty: Kassandra, 2013. V. 1. 556 p. V. 2. 445 p. (in Russian).
2. Hotelling H. Analysis of a complex of statistical variables into principal components // Journal of Educational Psychology. 1933. V. 24. P. 417–441. doi: 10.1037/h0071325
3. Fisher R.A. The use of multiple measurements in taxonomic problems // Annals of Eugenics. 1936. V. 7. P. 179–188. doi: 10.1111/j.1469-1809.1936.tb02137.x
4. Digby P.G.N., Kempton R.A. Multivariate analysis of ecological communities. London (UK): Chapman and Hall, 1987. 206 p.
5. Jongman R.H.G., ter Braak C.J.F., van Tongeren O.F.R. Data analysis in community and landscape ecology. Wageningen (The Netherlands): Pudoc, 1987. 299 p.
6. Orloci L. Geometric models in ecology. I. The theory and application of some ordination methods // Jour. Ecol. 1966. V. 54. P. 193–215. doi: 10.2307/2257667.
7. Kenkel N.C. On selecting an appropriate multivariate analysis // Canadian Journal of Plant Science. 2006. V. 86. P. 663–676. doi: 10.4141/P05-164.
8. Hervé M.R., Nicolè F., Lê Cao K.A. Multivariate analysis of multiple datasets: a practical guide for chemical ecology // Journal of Chemical Ecology. 2018. V. 44. P. 215–234. doi: 10.1007/s10886-018-0932-6.
9. Meng C., Zeleznik O.A., Thallinger G.G., Kuster B., Gholami A.M., Culhane A.C. Dimension reduction techniques for the integrative analysis of multi-omics data // Briefings in Bioinformatics. 2016. V. 17. P. 628–641. doi: 10.1093/bib/bbv108.

10. Rohart F., Gautier B., Singh A., Lê Cao K.-A. mixOmics: An R package for 'omics feature selection and multiple data integration // *PLoS Comput. Biol.* 2017. V. 13. No. 11. doi: 10.1371/journal.pcbi.1005752.
11. Shitikov V.K., Mastitsky S.E. Classification, regression and other Data Mining algorithms using R. E-book, 2017. 351 p. [Internet resource] <https://stok1946.blogspot.com/> (Accessed: 24.07.2018) (in Russian).
12. Legendre P., Legendre L. Numerical ecology. Amsterdam: Elsevier Sci. BV, 2012. 1006 p.
13. Ter Braak C.J.F. Canonical community ordination. Part I: Basic theory and linear methods // *Ecoscience*. 1994. V. 1. P. 127–140. doi: 10.1080/11956860.
14. Shitikov V.K., Zinchenko T.D., Rozenberg G.S. Macroecology of river communities: concepts, methods, models. Togliatti: Cassandra, 2012. 257 p. (in Russian).
15. Goodall D.W. Objective methods for the classification of vegetation. III. An essay in the use of factor analysis // *Austral. J. Bot.* 1954. V. 2. P. 304–324. doi: 10.1071/BT9540304.
16. Hill M.O. Reciprocal averaging: an eigenvector method of ordination // *Journal of Ecology*. 1973. V. 61. P. 237–249. doi: 10.2307/2258931.
17. Hill M.O., Gauch H.G. Detrended correspondence analysis: an improved ordination technique // *Vegetatio*. 1980. V. 42. P. 47–58. doi: 10.1007/BF00048870.
18. Gower J.C. Statistical methods of comparing different multivariate analyses of the same data // *Mathematics in the archaeological and historical sciences* / Ed. P. Tautu. Edinburgh: Edinburgh University Press, 1971. P. 138–149.
19. Anderson A.J.B. Ordination methods in ecology // *Journal of Ecology*. 1971. V. 59. P. 713–726. doi: 10.2307/2258135.
20. ter Braak C.J. Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis // *Ecology*. 1986. V. 67. P. 1167–1179. doi: 10.2307/1938672.
21. Rao C.R. The use and interpretation of principal component analysis in applied research // *Sankhya*. 1964. Ser. A. V. 26. P. 329–358.
22. Legendre P., Anderson M.J. Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments // *Ecological Monographs*. 1999. V. 69. P. 1–24. doi: 10.1890/0012-9615(1999)069[0001:DBRATM]2.0.CO;2.
23. Tenenhaus A., Tenenhaus M. Regularized generalized canonical correlation analysis // *Psychometrika*. 2011. V. 76. P. 257–284. doi: 10.1007/s11336-011-9206-8.
24. Wold H. Partial least squares // *Encyclopedia of statistical sciences* / Eds. S. Kotz, N. Johnson. New York: Wiley, 1985. P. 581–591.
25. Bylesjö M., Rantalainen M., Cloarec O., Nicholson J.K., Holmes E., Trygg J. OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification // *Journal of Chemometrics*. 2006. V. 20. P. 341–351. doi: 10.1002/cem.1006.
26. Sampson P.D., Streissguth A.P., Barr H.M., Bookstein F.L. Neurobehavioral effects of prenatal alcohol: part II. Partial least squares analysis // *Neurotoxicology and Teratology*. 1989. V. 11. P. 477–491. doi: 10.1016/0892-0362(89)90025-1.
27. Singh A., Gautier B., Shannon C.P., Vacher M., Rohart F., Tebutt S.J., Le Cao K.-A. DIABLO-an integrative, multi-omics, multivariate method for multi-group classification // *BioRxiv*. 2016. No. 067611. 50 p. doi: 10.1101/067611.
28. Harrington P.B., Vieira N.E., Espinoza J., Nien J.K., Romero R., Yergey A.L. Analysis of variance – principal component analysis: a soft tool for proteomic discovery // *Analytica Chimica Acta*. 2005. V. 544. P. 118–127. doi: 10.1016/j.aca.2005.02.042.
29. Engel J., Blanchet L., Bloemen B., van den Heuvel L.P., Engelke U.H.F., Wevers R.A., Buydens L.M.C. Regularized MANOVA (rMANOVA) in untargeted metabolomics // *Analytica Chimica Acta*. 2015. V. 899. P. 1–12. doi: 10.1016/j.aca.2015.06.042.
30. Dolédec S., Chessel D. Co-inertia analysis: an alternative method for studying species–environment relationships // *Freshwater Biology*. 1994. V. 31. P. 277–294. doi: 10.1111/j.1365-2427.1994.tb01741.x.
31. Dray S., Chessel D., Thioulouse J. Procrustean co-inertia analysis for the linking of multivariate datasets // *Écoscience*. 2003. V. 10. P. 110–119. doi: 10.1080/11956860.2003.11682757.
32. Chessel D., Hanafi M. Analyses de la co-inertie de K nuages de points // *Revue de statistique appliquée*. 1996. V. 44. P. 35–60.
33. Gower J.C. Generalized procrustes analysis // *Psychometrika*. 1975. V. 40. P. 33–51. doi: 10.1007/BF02291478.
34. Shitikov V.K., Zinchenko T.D., Abrosimova E.V. Statistical analysis of the results of multivariate ordination as exemplified by data on river benthic communities // *Russian Journal of Ecology*. 2012. V. 43. P. 137–141. doi: 10.1134/S1067413612010146.
35. Shitikov V.K., Rozenberg G.S. Randomization and bootstrap: a statistical analysis in biology and ecology with R use. Togliatti: Cassandra, 2014. 314 p. (in Russian).