

**МНОГОМЕРНЫЙ СТАТИСТИЧЕСКИЙ АНАЛИЗ ЭКОСИСТЕМ НА ПРИМЕРЕ ДОННЫХ
СООБЩЕСТВ МАЛОЙ РАВНИННОЙ РЕКИ (ОБЗОР)**

**Шитиков Владимир Кириллович,
Зинченко Татьяна Дмитриевна
Институт экологии Волжского бассейна РАН,
445003, Россия, Самарская обл., Тольятти, ул. Комзина, д. 10,
zinchenko.tdz@yandex.ru**

многомерная ординация, анализ главных компонент, неметрическое шкалирование, канонический анализ, сообщества макрозообентоса; продольный градиент реки

Представлены методы многомерного анализа, позволяющие выполнять оптимальное проецирование данных с большим числом переменных в пространства с малой размерностью. Показана роль этих методов в экологии сообществ для обработки наблюдаемой информации с целью выявления значимых статистических связей таксономической структуры с характеристиками биотопов и факторами окружающей среды.

В качестве примера рассматривается сообщество донных организмов р. Уса, являющейся притоком Куйбышевского водохранилища. Подробно обсуждаются результаты анализа данных гидробиологической съемки в 10 местообитаниях шестью методами непрямо́й и прямо́й многомерной ординации, на основании которых формулируются выводы по экологии водотока. С использованием анализа РСА выделены главные компоненты мультиколлинеарного комплекса из 18 факторов окружающей среды, определяющих продольный градиент реки. На основе метода неметрического многомерного шкалирования (NMDS) выполнена кластеризация 87 видов макрозообентоса и установлена корреляционная связь таксономической структуры сообществ с абиотическими показателями. С использованием анализа избыточности (RDA) и построения многомерных деревьев классификации (MRT) установлены ведущие факторы, определяющие видовой состав донных организмов: скорость течения реки, содержание кислорода и фосфора, общая минерализация. Прокрустовый совместный анализ инерции (PCIA) показал, что статистическая связь видовой структуры с факторами среды носит «прерывистый» характер ($p = 0,165$) и на отдельных участках нуждается в привлечении дополнительных объясняющих переменных. На основе канонического анализа соответствий (CCA) показано, что изменчивость донных сообществ примерно в равной степени определяется продольным градиентом реки, многолетней и сезонной динамикой, а также прочими неучтенными факторами. Показаны основные перспективы и направления развития многомерных методов применительно к экологии сообществ.

**MULTIVARIATE STATISTICAL ANALYSIS OF ECOSYSTEMS ON THE EXAMPLE OF
BOTTOM COMMUNITIES OF A SMALL PLAIN RIVER (OVERVIEW)**

**Shitikov Vladimir Kirillovich, Zinchenko Tatiana Dmitrievna,
Institute of Ecology of the Volga River Basin of the Russian Academy of Sciences,
10, Komzina St., Togliatti, Samara Region, Russia, 445003,
zinchenko.tdz@yandex.ru**

multivariate ordination, analysis of principal components, nonmetric scaling, canonical analysis, macrozoobenthos communities; longitudinal gradient of the river, small plain river.

We present methods of multivariate analysis that allow us to perform optimal projection of data with a large number of variables into spaces with small dimensions. The role of these methods in the ecology of communities is shown to process the obtained long-term information in order to identify significant statistical links of the taxonomic structure with characteristics of biotopes and environmental factors. As an example, we consider the community of benthic organisms of the Usa river, which is a tributary of the Kuibyshev reservoir. The results of the analysis of hydrobiological survey data in 10 habitats are discussed in detail by six methods of indirect and direct multidimensional ordination, on the basis of which conclusions on the ecology of the watercourse are formulated. Using the Principal component analysis (PCA), the main components of the multicollinear complex of 18 environmental factors determining the longitudinal gradient of the river are identified. On the basis of the method of nonmetric multidimensional scaling (NMDS), 87 species of macrozoobenthos were clustered and correlation of the taxonomic structure of communities with abiotic indices was established. With the use of redundancy analysis (RDA) and the construction of multivariate regression trees (MRT), we have identified the leading factors determining the species composition of benthic organisms: river flow rate, oxygen and phosphorus content, total mineralization. Procrustic conjoint inertia analysis (PCIA) showed that the statistical relationship of the species structure with environmental factors is “intermittent” ($p = 0.165$) and in some areas requires additional explanatory variables. Based on canonical

correspondence analysis (CCA), it is shown that the variability of bottom communities is approximately equally determined by the longitudinal gradient of the river, the long-time and seasonal dynamics, and other unaccounted factors. We show the main perspectives and directions for the development of multivariate methods in relation to the analysis of the ecology of river bottom communities.

Введение

Экология сообществ рассматривает всё многообразие связей как внутри ассоциаций взаимодействующих организмов, образующих биоценозы, так и в процессе контакта экосистем с окружающей средой [7]. Количественная основа подобных биоиндикационных исследований была заложена Р. Уиттекером [49], разработавшим прямой градиентный анализ, который моделирует распределение отдельных ценопопуляций относительно осей 2-3 комплексных средообразующих факторов. Однако современный подход к изучению экосистем основан на статистическом анализе десятков и сотен разнообразных факторов (популяционных, фенотипических, генетических, экологических, химических, ландшафтно-географических), которые характеризуются значительной временной и пространственной изменчивостью, вызываемой сезонными, климатическими, антропогенными или иными причинами. Основная задача такого анализа – выявление значимости и механизмов воздействия всей совокупности факторов на структурно-функциональные особенности изучаемых сообществ, а также последующий прогноз характера их развития (с точки зрения повышения продуктивности и разнообразия) при различных сценариях природоохранной деятельности.

Анализ главных компонент [42] и дискриминантный анализ [22], разработанные выдающимися статистиками К. Пирсоном и Р. Фишером, положили начало широкому использованию многомерных статистических методов в различных сферах науки и техники. С 70-х годов XX века начался бурный рост числа этих методов и успешная их адаптация к обработке экологических данных [1, 18, 26, 38, 41]. При этом важное место в экологии сообществ заняла ординация – совокупность методов распознавания образов для анализа упорядоченности (нем. *Ordnung*) изучаемых объектов путем их оптимального проецирования в пространство малой размерности [5, 6, 8, 12, 23, 25, 39, 43]. Постановка задач ординации обусловлена, как правило, целью проводимого исследования и доступностью исходных данных.

Непрямая или «необъяснённая» ординация (*unconstrained ordination*) выполняется с использованием только таблицы «виды–местообитания» и ставит своей целью получить отображение в ортогональной системе координат структурных закономерностей изучаемого сообщества в форме графических проекций распределения популяций по биотопам. При этом обычно формируется отчётливая и хорошо интерпретируемая группировка видов и/или местообитаний, появляется возможность выявить закономерности взаимосвязи между выделенными группами или сравнить ординационные графики за разные периоды времени. Однако непрямую ординацию считают «пассивным путём» [32], который ограничивается лишь фиксацией некоторого распределения точек на плоскости. Возможно, что этой конфигурации точек можно приписать влияние неких предположительных экологических градиентов, но никаких формальных доказательств этому не даётся.

Прямая или объяснённая (*constrained*) ординация ставит своей целью связать внутреннюю изменчивость видовой структуры сообщества с теми или иными факторами внешних воздействий. Концептуально эта форма многомерного анализа, называемая канонической [48], позиционируется как расширение идей регрессионного анализа на моделирование многомерного отклика при одновременном использовании алгоритмов редукции данных. Формально речь идет о построении модели регрессии: $\mathbf{Y} = f(\mathbf{X})$, где \mathbf{Y} – матрица размером $n \times m$ содержит значения отклика y_{ij} (т.е. популяционные плотности j -го вида, $j = 1, \dots, m$, для каждого i -го местообитания, $i = 1, \dots, n$); \mathbf{X} – матрица размером $n \times q$, в которой определены значения объясняющих абиотических факторов x_{ik} , измеренные в каждой i -й точке пространства.

Поскольку переменные отклика \mathbf{Y} и объясняющих факторов \mathbf{X} , как правило, высоко коррелированы, возникают ещё более сложные вопросы количественной оценки вклада их взаимодействий в совокупную многомерную дисперсию (*Inertia*) данных. Это обусловило появление группы новых методов, основанных на анализе ковариаций (*covariance-based methods*). В таблице 1 приведена сводка наиболее известных многомерных методов статистического анализа, широко используемых в экологии, генетике и биохимии [22].

Таблица 1 [Table 1]

Обзор многомерных методов статистического анализа; данные:

Y – таблица популяционной плотности видов; D – матрица дистанций, X – таблица факторов среды

Overview of multivariate statistical methods of analyses;

Y – data set of abundance values of species; D – distance matrix, X – data set of environmental variables

Тип анализа Type of analyses	Данные Data sets	Аббревиатура Abbreviations	Наименование метода, авторы, год Name of methods, authors, year
Непрямая ординация / Unconstrained ordination	Y	PCA	Анализ главных компонент / Principal Component Analysis – Hotelling, 1933 [31]
		CA	Анализ соответствий / Correspondence Analysis – Hill, 1973 [29]
		DCA	Анализ соответствий с удаленной «аркой» / Detrended Correspondence Analysis – Hill, Gauch, 1979 [30]
	D	PCoA (MDS)	Анализ главных координат = Метрическое многомерное шкалирование / Principal Coordinate Analysis = Metric Multidimensional Scaling – Gower, 1966 [24]
NMDS		Неметрическое многомерное шкалирование / Non Metric Multidimensional Scaling – Kruskal, 1964 [33]	
Прямая ординация / Unconstrained ordination	X, Y	CCA	Канонический анализ соответствий / Canonical Correspondence Analysis – ter Braak, 1986 [47]
		RDA	Анализ избыточности / Redundancy Analysis – Rao, 1964 [44]
	D, Y	db-RDA	Анализ избыточности, основанный на дистанциях / distance-based Redundancy Analysis – Legendre, Anderson, 1999 [34]
Деревья Trees	X, Y	RGCCA	Обобщенный канонический корреляционный анализ с регуляризацией / Regularized Generalized Canonical Correlation Analysis – Tenenhaus, Tenenhaus, 2011 [46]
		MRT	Деревья многомерной регрессии / Multivariate regression trees – De'Ath, 2002 [17]
Дискриминантный анализ и PLS-регрессия / Discriminant Analysis and PLS-regression	X, Y	LDA	Линейный (канонический) дискриминантный анализ / Linear (Canonical) Discriminant Analysis – Fisher, 1936 [22]
		PLS	Семейство методов с использованием частных наименьших квадратов = проецирования латентных структур / Partial Least Squares = Projection to Latent Structures – Wold et al., 1983 [50];
		OPLS-DA	дискриминантный ортогональный анализ методом PLS / Orthogonal Partial Least Squares -Discriminant Analysis – Bylesjö et al., 2006 [14];
	PPLS-DA	дискриминантный приведенный анализ методом PLS / Powered Partial Least Squares -Discriminant Analysis – Liland, Indahl, 2009 [37];.	
D, Y	DPLS	дискриминантный анализ неоднородности методом PLS / Dissimilarity Partial Least Squares – Zeruza et al., 2012 [51]	
Дисперсионный анализ ANOVA	X, Y	ANOVA-PCA	Многомерный метод дисперсионного анализа с использованием главных компонент / ANOVA-Principal Component Analysis – Harrington et al., 2005 [27]
		rMANOVA	Многомерный ANOVA с регуляризацией / regularized Multivariate ANOVA – Engel et al., 2015 [21]
Ко-инерционный анализ / Co-Inertia Analysis	X, Y	CIA	Совместный инерционный анализ / Co-Inertia Analysis – Dolédec, Chessel 1994 [19]
		PCIA	Прокрустовый совместный инерционный анализ / Procrustean Co-Inertia Analysis – Dray et al., 2003 [20]
	X, Y > 2	MCIA	Множественный совместный инерционный анализ / Multiple Co-Inertia Analysis – Chessel, Hanafi, 1996 [15]
		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 [45]

Целью настоящей статьи явилась систематизация практических рекомендаций по проведению многомерного статистического анализа данных по экологии сообществ. Основой для нее явилась аналогичная работа М. Эвре с соавторами [28] в области химической экологии. Мы, разумеется, не могли продемонстрировать в статье использование всех многомерных процедур, представленных в табл. 1 и выбрали только шесть основных алгоритмов (рис. 1-6). Также за рамками нашего изложения

остались математические нюансы реализации каждого метода, и здесь фундаментальная монография братьев Лежандров [36] будет прекрасным пособием заинтересованному читателю.

Основной задачей было обозначение ключевых концепций основных методов, обсуждение их различий в плане достоинств и недостатков и показ особенностей интерпретации конкретных полученных результатов. Все методы, продемонстрированные нами, реализованы в программном обеспечении свободно распространяемой статистической среды R [40]. Примеры скриптов, позволяющих самостоятельно выполнить основные процедуры многомерного анализа, обсуждаемые ниже, подробно представлены в главе 9 электронной книги [10].

Материал и исходные предпосылки анализа

Выполнение процедур многомерного анализа иллюстрируются на примере, где в качестве объекта исследований выступает малая равнинная река Уса ($53^{\circ}26'32.2''$ с.ш., $48^{\circ}09'26.5''$ в.д.) длиной 76 км, протекающая у северо-западной оконечности Жигулевских гор и впадающая в Усинский залив Куйбышевского водохранилища. На протяжении ряда лет в вегетационный период на 10 постоянных станциях наблюдений реки отбирали гидробиологические пробы макрозообентоса, отдельные таксоны которого обладают признанными биоиндикационными свойствами и являются наиболее стабильными и разнообразными компонентами лотических экосистем [2]. В качестве факторов среды, формирующих продольный экологический градиент реки, одновременно производились измерения комплекса показателей, описывающих гидрологию реки, морфологию прилегающего рельефа, качество воды и её гидрохимический состав. Подробные сведения о технических аспектах проведения исследования и основные выводы о структурно-функциональных закономерностях изучаемого сообщества донных организмов представлены в статье [3].

Выбор конкретного статистического метода зависит, как правило, от характера решаемой биологической проблемы. Каждый из методов настроен на использование одного из двух типов исходных данных об анализируемом сообществе: (а) непосредственно таблицы обилия видов \mathbf{Y} и (б) предварительно рассчитанной матрицы дистанций \mathbf{D} . Первый вариант (а) используется, если ставится задача количественной оценки сравнительной значимости отдельных видов и получения информации об индивидуальных показателях численности таксонов. Методы (б), основанные на гипотезе «разделения на основе расстояния» (*isolation-by-distance*), становятся в центре внимания, если задача ограничивается оценкой общего таксономического сходства двух или более сообществ, пренебрегая при этом ролью отдельных видов. Как следствие, последующая декомпозиция матрицы \mathbf{D} приводит к группировке местообитаний только на основе используемой метрики расстояний, и уже невозможно непосредственно определить, какие конкретно виды обусловили именно такую кластеризацию.

Результаты многомерного анализа в значительной мере зависят от исходных предпосылок и способов предварительной обработки данных, включая учёт характера статистического распределения наблюдаемых показателей, механизмы их усреднения и трансформации, выбор метрики для матрицы расстояния и проч. Хотя большинство многомерных методов не предъявляют формальных требований к определённому закону распределения данных, форма их численного представления должна быть адекватна задачам исследования. Например, биомасса различных видов бентоса может различаться в десятки тысяч раз и непосредственное использование этого показателя при формировании матрицы \mathbf{Y} приведет к тому, что итоговая ординационная диаграмма будет целиком определяться численностью 1–2 видов крупных моллюсков. Если же выполнить обычную стандартизацию, приведя биомассу каждого вида к нулевому среднему и единичной дисперсии, то это также не будет способствовать выяснению сути дела, поскольку доминирующие и функционально важные виды будут иметь ту же относительную значимость, что и виды-маргиналы, играющие незначительную роль в экосистеме. Поэтому, чтобы «приглушить» чересчур сильную вариацию показателя, часто выполняют логарифмическую или иную трансформацию данных. Используя общие правила построения хорошо интерпретируемых ординаций [35], мы для нашего примера выполняли преобразование численности видов, приводящее к χ^2 -дистанции, которое является, по всей вероятности, наиболее разумным компромиссом при учёте как роли ведущих компонент, так и вклада редких (малочисленных) таксонов.

При обобщении многократно повторяющихся наблюдений часто возникает проблема: следует ли усреднять результаты проб, выполненные в разные годы, либо в течение сезонных периодов? Здесь трудно дать конкретные рекомендации: с одной стороны, учёт различных условий отбора проб позволяет составить «обобщённый портрет» экосистемы, однако, с другой стороны, анализ тренда многолетних или сезонных изменений видовой структуры сообществ может стать самостоятельной

задачей. В любом случае необходимо соблюдать правило «эквивалентности выборочного усилия», т.е. для каждого уровня группировки должно быть объединено примерно одинаковое число проб.

Данные, полученные при изучении экологии сообществ, могут интерпретироваться как бинарные (т.е. в формате «есть/нет»), количественные или композиционные. Ранее [9] нами был проанализирован экологический смысл различий между качественными и количественными мерами сходства. В дополнение к этому обратим внимание на проблему «двойного нуля», когда задаются резонным вопросом: можно ли считать две пробы подобными на том основании, что в них обеих отсутствует какой-то вид? В связи с этим при формировании матрицы дистанций мы рекомендуем использовать для количественных данных меру сходства Брея–Кёртиса или её «бинарный аналог» коэффициент Съеренсена, причём использование любых количественных или балльных оценок обилия видов всегда предпочтительно.

Композиционные данные представляют собой доли (%) отдельных компонентов, в сумме составляющие 100%. Это могут быть доли численности отдельных видов в пробе, проективные покрытия, химический состав почвы и проч. Анализ композиционных данных должен выполняться с особой осторожностью, поскольку эти переменные весьма зависят друг от друга и любые изменения приводят к нарушению пропорций, что может привести к «ложным корреляциям» – см. [13] для детальных пояснений.

Важнейшая задача в экологии сообществ состоит в идентификации экологических градиентов или биологически значимых кластеров, объединяющих местообитания, основные воздействующие факторы и таксоны организмов, составляющих изучаемое сообщество. Для этого используются различные по математической природе алгоритмы, шесть из которых мы представляем ниже в абзацах текста, пронумерованных цифрами **1-5**. Отметим, что для всех демонстрируемых методов многомерная природа экологических данных приводит к использованию в той или иной форме метода редукции, т.е. такого оптимального проецирования больших исходных матриц в пространство низкой размерности (в наших примерах – на плоскости), чтобы при этом была обобщена большая часть имеющейся информации.

Непрямая ординация.

Классическими методами непрямой ординации являются анализ главных компонент (PCA) и два алгоритма многомерного шкалирования (PCoA и NMDS), которые формируют в многомерном пространстве исходных данных новые оси оптимального проецирования и представляют обобщённую эмпирическую информацию в максимально компактной форме.

1. Матрица **X** факторов окружающей среды включает 18 разноплановых показателей, оценённых по данным многолетних наблюдений для 10 станций р. Уса. Рис. 1 содержит ординационную диаграмму, полученную методом PCA [4, 9, 22, 31], в форме корреляционного круга, стрелки которого соответствуют отдельным факторам. Концы стрелок соответствуют координатам новых латентных переменных, а косинус угла между стрелками и осями главных компонент равен коэффициенту корреляции между ними. В частности, максимальной положительной корреляцией с первой главной осью PCA1, объясняющей 73,5% общей вариации данных, обладают концентрации ионов NO_2^- , Fe^{2+} и ширина реки **Bs**, а максимальной отрицательной корреляцией – расстояние до устья **L**. Длина и интенсивность окраски стрелок соответствуют вкладу (*contribution*) каждой переменной в формирование редуцированного пространства.

Здесь легко сделать два важных вывода. Во-первых, все факторы среды составляют тесно связанный мультиколлинеарный комплекс, поскольку коэффициенты корреляции между ними (определяемые углами между соответствующими стрелками) близки по модулю к 1. Во-вторых, первая главная компонента легко интерпретируется как комплексный продольный градиент реки, обобщающий факторы различной природы, во многом определяемые пространственным расположением станций. Вторая главная компонента частично связана с локальными условиями загрязнения реки сельскохозяйственными стоками: крутизной склонов **GA**, площадью водосбора **MCA**, минерализацией **Min**, содержанием ионов NO_2^- .

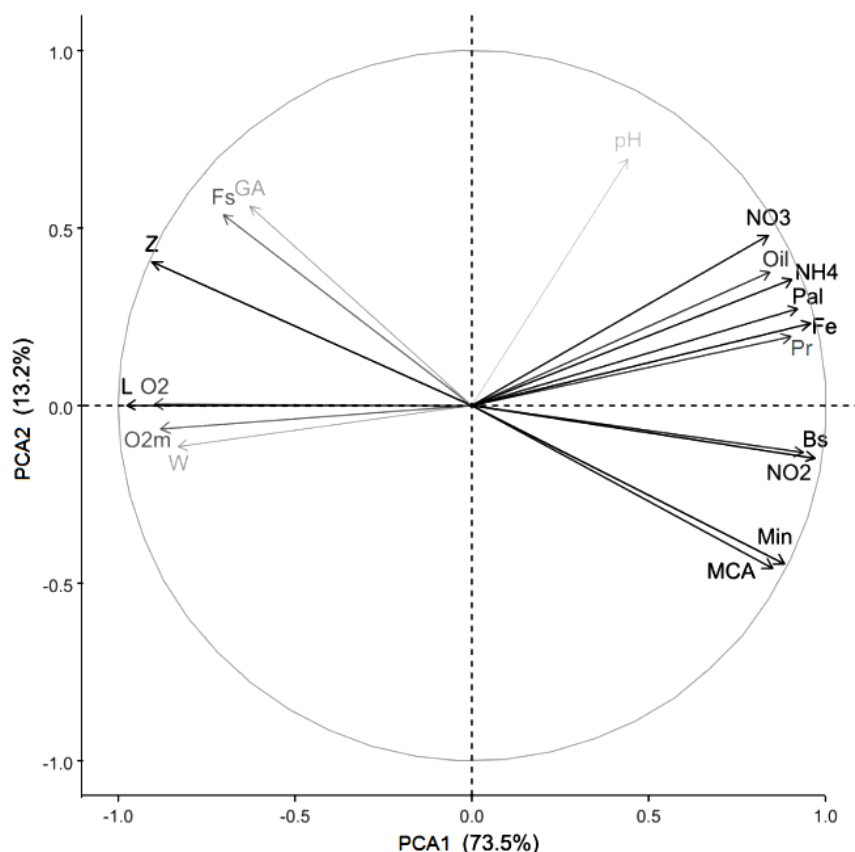


Рис. 1. Корреляционный граф показателей окружающей среды на станциях р. Уса, полученный методом главных компонент (PCA). Сокращения факторов среды приведены в табл. 2

2. Метод PCA основан на жёстко заданной процедуре вычисления собственных чисел и собственных векторов корреляционной матрицы, что при особой конфигурации данных может привести к нежелательным статистическим эффектам. Метод неметрического многомерного шкалирования NMDS [9, 16, 33, 43] позволяет использовать любую подходящую матрицу дистанций **D** и выполняет итерационную процедуру минимизации степени искажения взаимных расстояний между объектами при сокращении исходного многомерного пространства до 2-мерной плоскости (т. е. идет поиск наилучшей проекции для конкретных, даже иногда «зашумленных» эмпирических данных).

Представим результаты гидробиологического мониторинга р. Уса матрицей **Y**, содержащей трансформированные численности каждого из 87 обнаруженных видов и таксонов донных организмов на каждом из 10 участков водотока. После этого по формуле Брея-Кёртиса рассчитаем матрица **D** расстояний между каждой парой станций в многомерном пространстве видов. Ординационная диаграмма на рисунке 2 позволяет выделить в латентных шкалах NMDS1-NMDS1 две характерные группы станций в верхнем (ст. 01–04) и среднем (ст. 06–08) течении реки. С группировкой станций тесно связано положение видов макрозообентоса на диаграмме, которое определяется средневзвешенными координатами нескольких его возможных местообитаний.

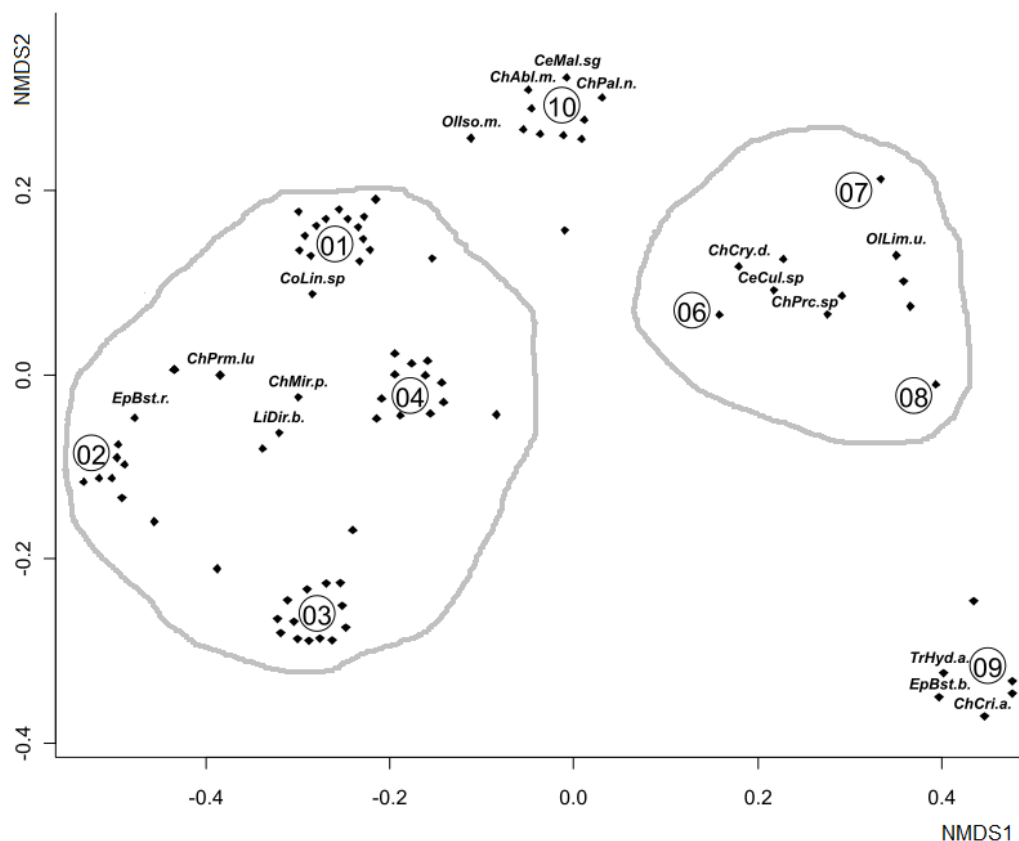


Рис. 2. Многомерная ординация донных сообществ, наблюдаемых на станциях {01 – 10} р. Уса методом неметрического многомерного шкалирования (NMDS). Коды видов расшифровываются по тексту.

Можно отметить следующие виды и таксоны, которые являются индикаторами обоих участков: в верховьях – *Baetis rhodani* (Pictet, 1843) (EpBst.r.), *Dicranota bimaculata* (Schummel, 1829) (LiDir.b.), *Micropsectra* gr. *praecox* (Meigen, 1818) (ChMir.p.), *Parametriocnemus lundbecki* (Johannsen, 1905) (ChPrm.lu), *Limnius* sp. (CoLin.sp), в среднем течении – *Culicoides* sp. (CeCul.sp), *Cryptochironomus defectus* (Kieffer, 1913) (ChCry.d.), *Procladius* sp. (ChPrc.sp), *Limnodrilus udekemianus* Claparède, 1862 (OILim.u.). Отметим, что в нижнем течении реки станция 09 характеризуется специфическими условиями аккумуляции биогенных веществ с поверхностным стоком водосборной площади, а ст. 10 в устьевой зоне реки испытывает влияние подпора эвтрофными водами Куйбышевского водохранилища. Каждая из них занимает на диаграмме особое положение и характеризуется относительно локальным составом донной фауны, где важное место занимают виды *Cricotopus* gr. *algarum* (Kieffer) (ChCri.a.), *Baetis bioculatus* (Linnaeus, 1758) (EpBst.b.), *Hydropsyche angustipennis* (Curtis, 1834) (TrHyd.a.) для ст. 09 и *Mallochohelea setigera* (Loew, 1864) (CeMal.sg), *Ablabesmyia monilis* (Linnaeus, 1758) (ChAbl.m.), *Paralauterborniella nigrohalteralis* (Malloch, 1915) (ChPal.n.), *Isohaetides michaelsoni* (Lastočkin, 1936) (OIIso.m.) для ст. 10.

Для оценки характера и силы влияния каждого физических показателя на видовую структуру донных сообществ, легко рассчитать коэффициенты корреляции между изученными факторами среды и NMDS-шкалами. С использованием рандомизационной процедуры [11] можно установить, что на уровне значимости $p = 0,05$ изменчивость видового состава макрозообентоса связана лишь с 5 показателями из 18 (табл. 2).

Таблица 2 [Table 2]

Оценка значимости связи отдельных факторов среды с ординационной структурой видов макрозообентоса (рис. 1,3,4)

Estimation of the significance of dependence of ordination structure of macrozoobenthos species (Fig. 1,3,4) from basic of environmental factors

Наименование показателей [Names of features]	Среднее значение [Mean values]	Коэффициенты корреляции [correlation coefficients]			<i>p</i> -значение [<i>p</i> -values]
		с осью [axis] S₁	с осью [axis] S₂	общий квадрат R²	
Расстояние до устья, км [Distance to a mouth, km] (L)	39,2±7,7	-0,926	-0,377	0,465	0,151
Течение, м/с [Current, m/s] (W)	0,279±0,079	-0,926	-0,376	0,679	0,03
Ширина, м [Width, m] (Bs)	12,89±1,83	0,995	0,097	0,483	0,127
Прозрачность, см [Transparency] (Pr)	43,89± 6,81	0,864	-0,503	0,598	0,064
Высота, м [Elevation, m] (Z)	78,8±10,9	-0,989	0,150	0,428	0,183
Площадь водосбора преобр. [Maximal dispersal area, m ²] (MCA)	0,994±0,005	1,000	-0,006	0,502	0,12
Освещенность склонов, % [Relative slope insolation, %] (Fs)	57,4±0,46	-0,986	0,168	0,114	0,733
Крутизна склонов, град [Slope steepness, degress] (GA)	0,829±0,294	-0,547	0,837	0,438	0,168
pH дна [pH a bottom] (pH)	8,181±0,038	0,317	-0,948	0,254	0,412
Кислород дна, % [Oxygen a bottom] (O2)	78,26±5,68	-0,876	-0,482	0,909	0,003
Кислород дна O ₂ , мг/л [Oxygen a bottom, mg/l] (O2m)	6,836±0,573	-0,862	-0,507	0,900	0,002
Фосфор общий PO ₄ ³⁻ -P, мг/л Total Phosphorus, mg/l (Pal)	0,412±0,118	0,969	0,245	0,652	0,045
Азот аммонийный NH ₄ ⁺ -N, мг/л [Ammonium Nitrogen, mg/l] (NH4)	0,176±0,049	0,988	0,152	0,583	0,077
Азот нитратный NO ₃ ⁻ -N, мг/л [Nitrat Nitrogen, mg/l] (NO3)	0,268±0,122	0,927	0,374	0,279	0,389
Азот нитритный NO ₂ ⁻ -N, мг/л [Nitrat Nitrogen, mg/l] (NO2)	0,023±0,004	0,999	0,036	0,360	0,257
Железо Fe ²⁺ , мг/л [Iron, mg/l] (Fe)	0,391±0,081	0,990	0,142	0,577	0,076
Общая минерализация, мг/л [Total mineralization, mg/l] (Min)	262±34	0,998	0,070	0,633	0,06
Нефтепродукты, мг/л [Oil product, мг/л] (Oil)	0,041±0,010	0,936	-0,353	0,287	0,38

Примечание: жирным шрифтом выделены факторы, влияющие со статистической значимостью $p < 0,05$, курсивом – при $p = 0,05 \div 0,1$. В круглых скобках указаны обозначения факторов для рис. 1, 3, 4.

Прямая ординация.

3. Большинство алгоритмов многомерного анализа связано с использованием аппарата линейной алгебры и, в частности, с нахождением собственных значений симметричной матрицы. Анализ избыточности (RDA) [9, 24, 48] является версией PCA с инструментальными переменными и одновременно расширением идей регрессионного анализа на случай многомерного отклика $Y = f(X)$. Таким образом, в результате RDA-анализа рассчитывается матрица канонических коэффициентов, взвешивающих влияние объясняющих переменных X , и они совместно с нагрузками главных компонент PCA определяют положение точек (т.е. видов и местообитаний) на ординационной диаграмме (рис. 3). На графике избыточности, кроме главных осей RDA1-RDA2, проводятся дополнительные оси физических градиентов, интерпретация которых подчиняется тем же правилам корреляционных отношений, что и рис. 1. Например, высокая корреляция между вектором ст. 02 и осями **W** и **O2** позволяет уточнить ведущие условия среды, определяющие состав донного сообщества на этой станции.

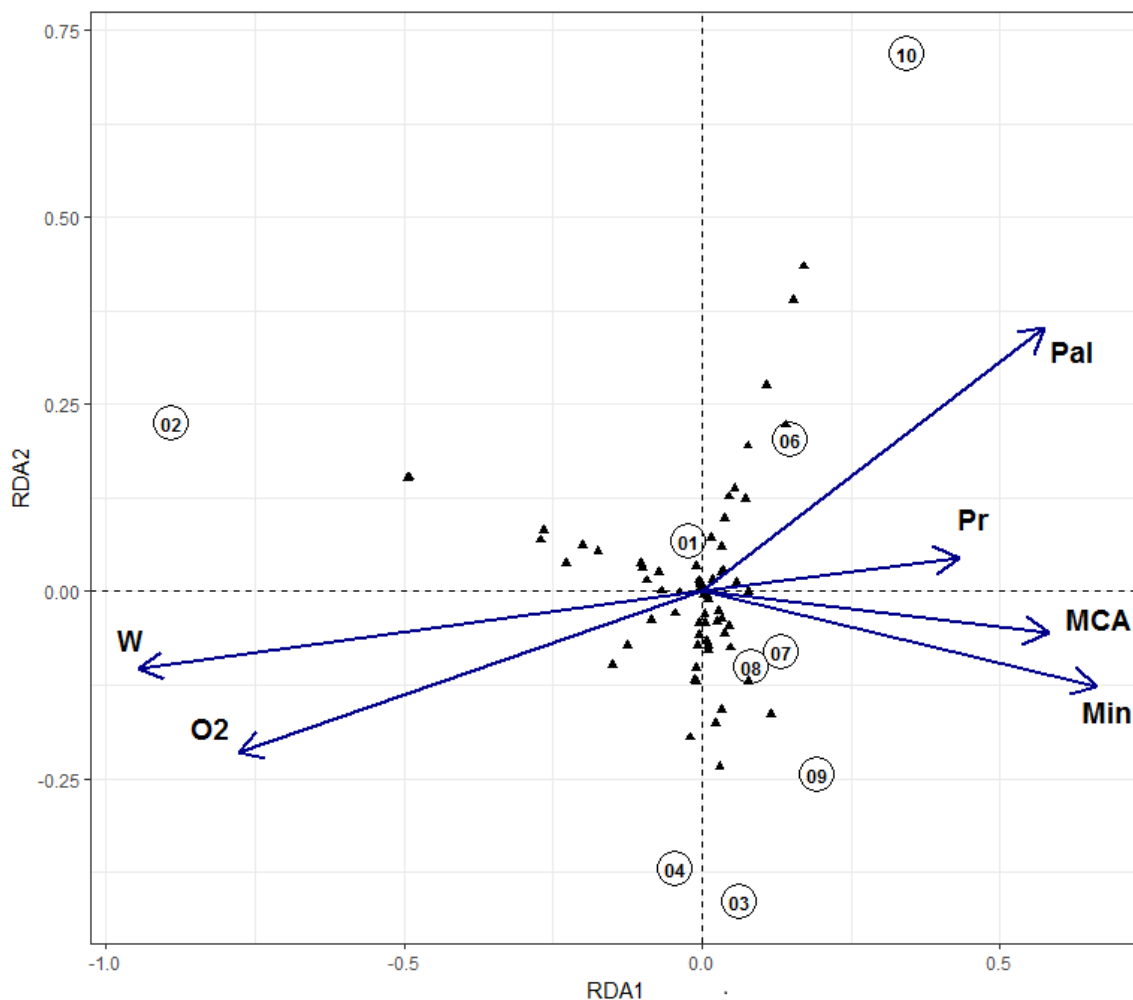


Рис. 3. Многомерная ординация станций наблюдения {01 – 10} на р. Уса методом анализа избыточности (RDA). Обозначения факторов среды приведены в табл. 2.

Отметим, что ординация на рис. 3 объясняет 100% статистической вариации наблюдаемых данных, поскольку 6 факторов среды и 2 главных компоненты позволяют точно определить положение всех 9 станций в многомерном пространстве. В данном случае это свидетельствует лишь о переусложнённости и потенциальной неустойчивости полученной модели. С экологической точки зрения ординация на рис. 3 видится менее удачной, чем на рис. 2, что определяется рядом причин: (а) недостаточно обоснованным предположением о линейном характере внутренних и внешних взаимодействий в экосистеме; (б) использованием корреляционной матрицы вместо более «экологичной» метрики Брея-Кёртиса; (в) мультиколлинеарностью пространства факторов среды. По последней причине статистически значимыми оказались лишь канонические коэффициенты для скорости течения **W**, содержания фосфора **Pal** и прозрачности **Pr**, а остальные факторы незначимы, поскольку попросту тесно связаны с ними.

4. Алгоритм построения деревьев многомерной классификации и регрессии (MRT) [10, 11, 17] не является проекционным методом и не формирует ординационной диаграммы, но позволяет эффективно осуществить отбор ключевых факторов среды. Деревья MRT формируются в результате рекурсивной процедуры разделения строк исходной таблицы многомерного отклика **Y** на все более и более гомогенные подмножества, в результате чего формируется древовидная иерархическая структура. С каждым узлом формируемого дерева связано значение одного из факторов **X**, найденное таким образом, чтобы минимизировать различия между точками в многомерном пространстве. Для первого узла дерева на рис. 4 таким граничным значением явилось насыщение кислородом **O2** = 94,4 и был выделена группа близких сообществ на ст. {02: 03}. На следующем шаге разбиения при **Min** = 309,5 образованы два кластера станций с различным уровнем минерализации. На графике показано, как последовательно снижается сумма дистанций от объединяемых объектов до центра их группировки.

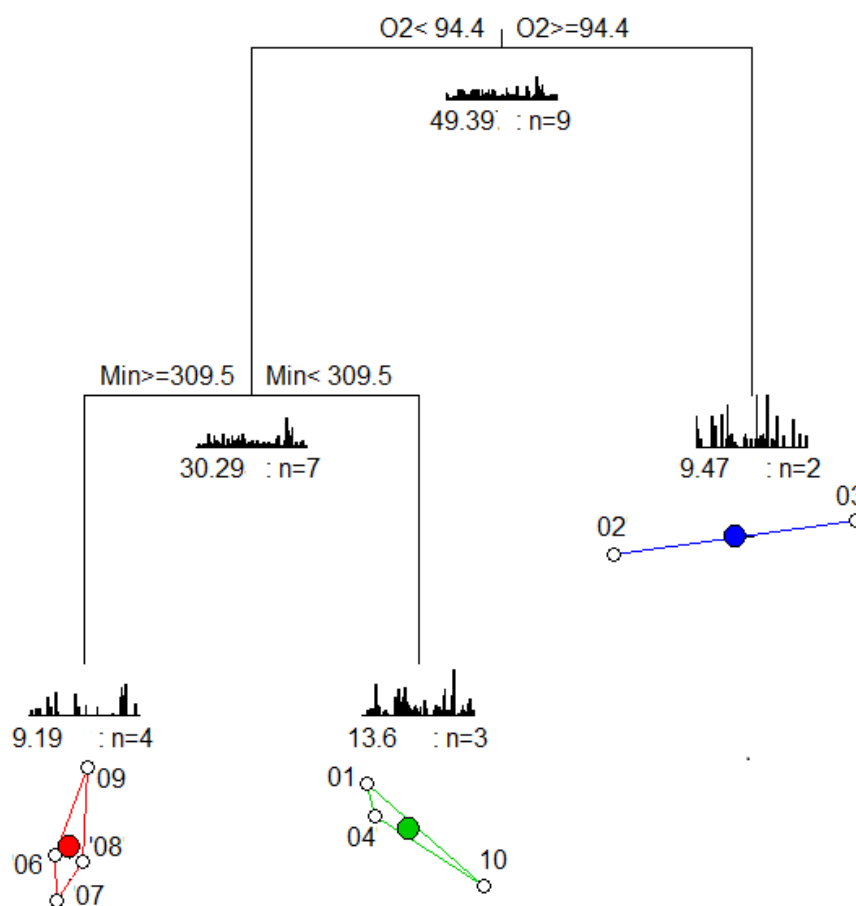


Рис. 4. Иерархическое многомерное дерево разбиений (MRT) донных сообществ по ключевым факторам среды (сокращения см. в табл. 2)

5. Если RDA можно назвать асимметричным анализом (т.е. выполняется поиск зависимости Y от X , но не наоборот), то совместный инерционный анализ (CIA) [19] совершенно равноправен по отношению к обоим наборам данных и с этой точки зрения является симметричным. На вход алгоритма CIA подаются две модели отображения таблиц Y и X в гиперпространстве, оптимальном относительно их максимальной вариации, после чего с использованием метода частичных наименьших квадратов (PLS) [50] выполняется поиск новых осей в направлении наибольшей совместной инерции (т.е. многомерной дисперсии) обобщаемых данных. Основное преимущество этого метода – гибкость реализации, поскольку общая модель может включать различные аналитические комбинации для сравнения Y и X . Например, если допускается, что видовая структура и факторы среды связаны между собой линейно, то используют анализ CIA PCA(X)-PCA(Y), т.е. сопоставляются две модели разложения на главные компоненты. Вариант CIA dbRDA(X)-dbRDA(Y) сравнивает две произвольных матрицы дистанций одинаковой размерности. Метод широко используется, например, для выявления изменений видового состава одного и того же участка реки до и после некоторого события (например, антропогенного воздействия), а множественный инерционный анализ MCIA позволяет исследовать динамику структуры сообществ за несколько последовательных периодов времени.

Прокрустовый совместный инерционный анализ PCIA [20] является одним из способов сравнения двух конфигураций точек методом «суперналажения». Для этого предварительно проводится подгонка геометрии пространства двух ординаций к сопоставимой форме путём масштабирования, центрирования и вращения. На рис. 5 показано, как передвинутся точки 9 станций р. Уса на RDA-диаграмме, если сходство между ними по обилию 87 видов донных организмов заменить вариацией 18 факторов среды. Эти изменения отражаются стрелками, начало которых указывает координаты местообитаний в биотическом пространстве, а конец – в экологическом. Можно сделать вывод, что видовой состав ст. 6, 7 и 9 в целом соответствует продольному градиенту условий среды, тогда как структуру донных сообществ ст. 1 и 10, по всей вероятности, определяют какие-то иные факторы, кроме отмеченных в табл. 2. С использованием методов Монте-Карло [11]

можно проверить нулевую гипотезу о равенстве нулю среднего прокрустового расстояния. В нашем случае $p = 0,165$, т. е. нельзя категорически утверждать, что структура сообщества определяется только наблюдаемыми показателями, поскольку влияние неучтённых факторов все же велико.

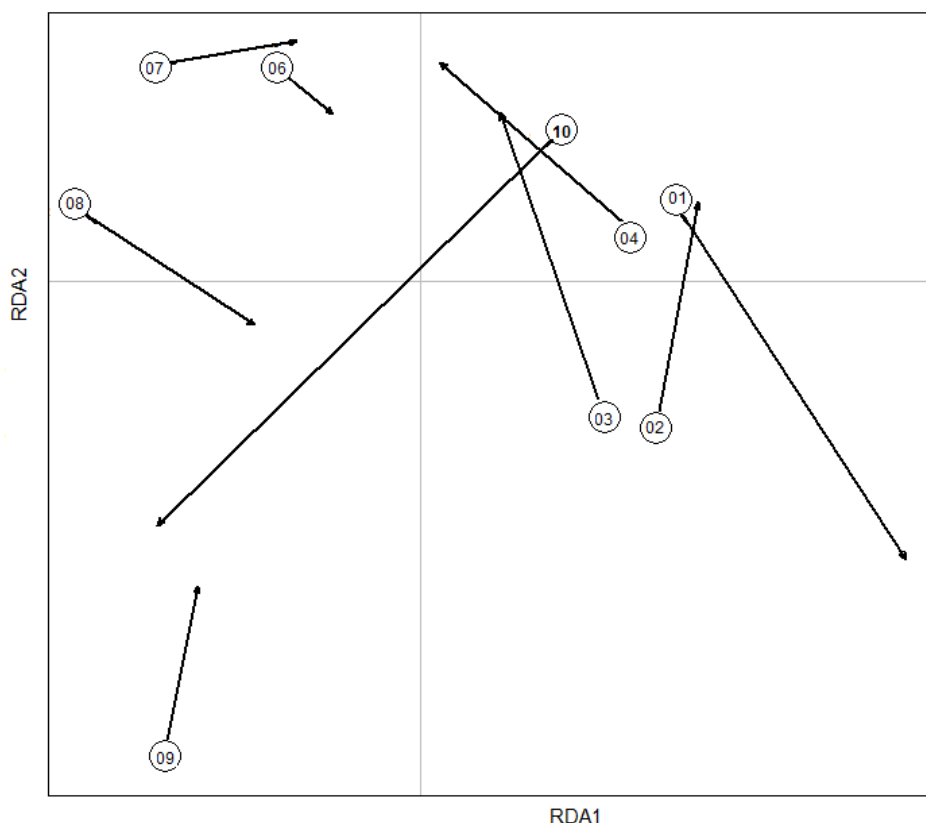


Рис. 5. Прокрустовые расстояния между ординациями станций наблюдения {01 – 10} на р. Уса, построенными по биотическим и абиотическим показателям

Разложение по главным компонентам в ходе СИА-анализа позволяет вычлнить значения факторных нагрузок отдельно для каждой переменной наборов Y и X . В нашем примере наибольшую ковариационную связь с факторами среды обнаруживают виды *Parametriocnemus lundbecki*, *Micropsectra gr. praecox*, *Cladotanytarsus mancus* (Walker, 1856), *Limnodrilus hoffmeisteri* Claparède, 1862, *Dicranota bimaculata*, *Oulimnius turbeculatus* (Müller, 1806). Ранжирование факторов среды по их ковариации с видовой структурой сообществ мало отличается от представленной в таблице 2.

6. Часто при анализе видовой структуры сообществ необходимо оценить вклад отдельных градаций нечисловых переменных. Если выше мы рассматривали «срез» гидробиологических наблюдений на р. Уса в ходе одной проведённой экспедиции, то для нового анализа матрицу Y сформируем из всех проб, в которых было обнаружено 158 видов донных организмов. Матрицу объясняющих переменных X составим из двух качественных переменных: ГОД, отражающий три периода отбора проб, и УЧАСТОК, определяющий пробы, относящиеся к верховьям, среднему течению и устьевому участку.

Будем использовать канонический анализ соответствий ССА, который преобразует исходные данные Y в матрицу χ^2 -расстояний, выполняет построение модели взвешенной линейной регрессии на переменные X , и, наконец, оценивает набор ординационных коэффициентов через разложение по сингулярным числам. Совокупная дисперсия многомерного отклика в нашем примере разлагается следующим образом: 24,4% связаны с первыми двумя «необъяснёнными» осями ССА1-ССА2 (рис. 6), а 22,6% – с компонентами, использующими независимые факторы X . Для качественных переменных последние не могут быть показаны в виде дополнительных осей (как на рис. 3), и поэтому представляются точками центроидов, координаты которых на рис. 6 обозначены прямоугольниками.

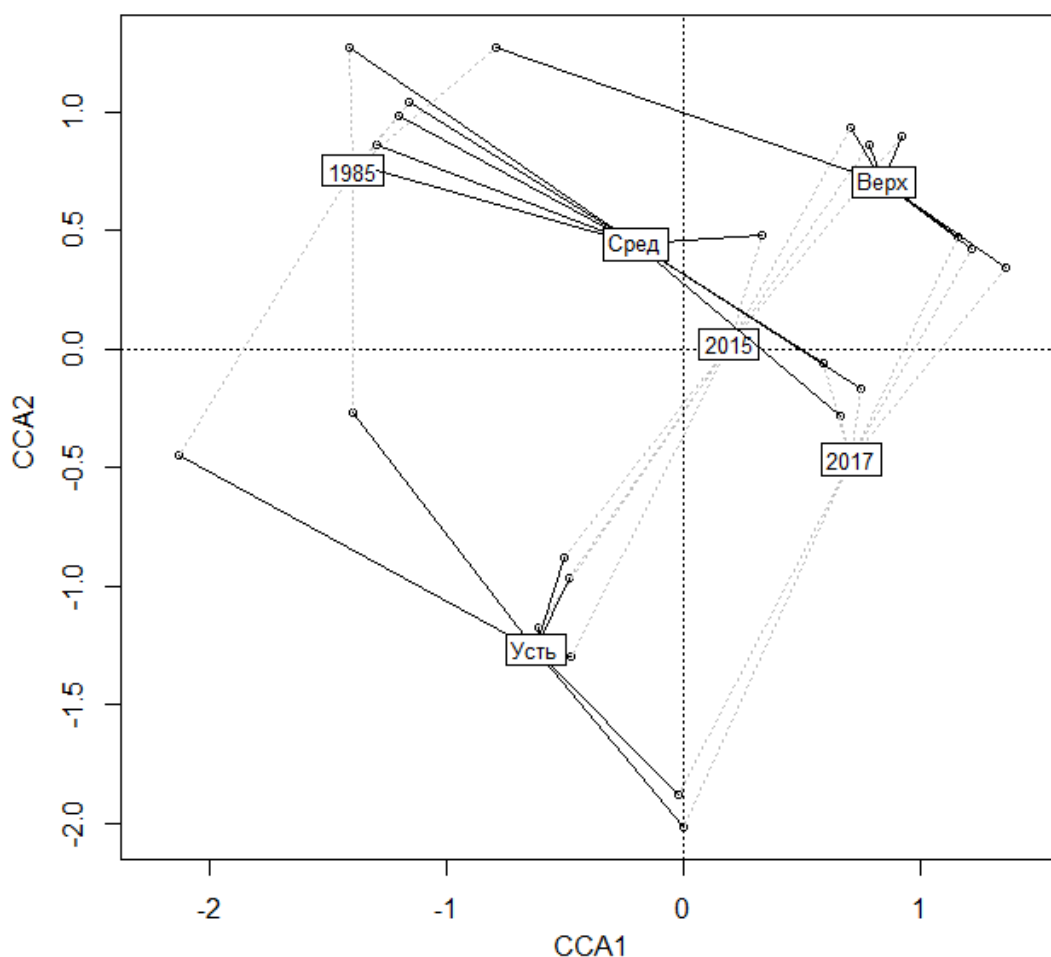


Рис. 6. Канонический анализ соответствий гидробиологических проб, взятых в разные годы и на разных участках р. Уса (верхнее, среднее течение и устьевой участок)

Разброс точек отдельных наблюдений относительно соответствующих центроидов позволяет судить о степени влияния каждого из факторов. Количественную оценку этому также даёт дисперсионный анализ с использованием рандомизационного теста [11], который показал статистически высоко значимое и приблизительно равное влияние обоих переменных ГОД и УЧАСТОК. Отметим при этом, что фактор ГОД отражает не только многолетнюю динамику донных сообществ, но и высокую стохастическую изменчивость видового состава, характерную для донных сообществ, пробы которых не могут быть признаны истинными статистическими повторностями.

Обсуждение проблемы

В целом, на основе всех блоков анализа 1–6 можно сделать следующие содержательные выводы относительно изучаемой экосистемы:

- от истоков к устью р. Уса существует продольный градиент условий среды, определяющий развитие донных сообществ;

- наиболее важными влияющими факторами являются насыщение кислородом, содержание фосфора и общая минерализация воды;

- продольный градиент носит, однако, «прерывистый» характер, нарушаемый локальными биотопическими условиями и неучтёнными факторами, что проявляется, например, в верхнем течении реки и на устьевых участках;

- пробы бентоса характеризуются высокой многолетней, сезонной и микробиотопической изменчивостью, т.е. мы никогда не имеем возможности «дважды войти в одну и ту же реку».

По мнению Н. Кенкеля [32], при использовании многомерных методов для статистической обработки наблюдений в биологических науках, «сохраняется значительная неопределённость в выборе оптимальной стратегии анализа, которая включает важные решения относительно преобразования данных, стандартизации переменных и общего методологического подхода. Эта стратегия должна основываться на тщательном и точном учёте цели исследования, структуры

обрабатываемых данных и основных статистических предпосылках. К сожалению, даже относительно свежие обзоры многомерных аналитических стратегий основываются в значительной степени на эмпирических представлениях сомнительной достоверности, ... которые появились в экологической литературе 1970-х годов и некритически увековечены по сей день. ...». Н.Кенкель ставил перед собой задачу «рассеять вводящие в заблуждение рекомендации и ошибочные обобщения относительной эффективности доступных многомерных методов...» [32].

Следует отметить, что Н. Кенкелю не удалось дать исчерпывающие рекомендации или в полной мере рассеять заблуждения. Само понятие «заблуждение» весьма вариативно: например, мы не согласны с негативным отношением автора к неметрическому многомерному шкалированию (NMDS), который, как показывает опыт [9, 16, 43], является наиболее устойчивым и интерпретируемым методом ординации в исследованиях экологии речных сообществ. Но, в любом случае, следует всячески содействовать пониманию того, что нет плохих методов, а есть неудачные примеры некорректного их использования в конкретных условиях. Выбор подходящих многомерных методов анализа требует гармоничного сочетания опыта и интуиции.

За рамками нашего изложения остались также такие новые перспективные методы как различные варианты дискриминантного анализа с использованием PLS, обобщенный ССА с регуляризацией (RGCCA), а также DIABLO, являющийся комбинацией PLS-DA и RGCCA который позволяет эффективно проводить анализ высоко коррелированных наборов данных (табл. 1).

Заключение

Сфера применения и арсенал алгоритмов многомерного анализа постоянно расширяется. Постепенно в исследованиях приобретает значение не сам факт использования того или иного многомерного метода, а тщательная и всесторонняя проверка адекватности построенной ординационной модели, для чего разработана целая система тест-статистик и кросс-проверочных алгоритмов. Из-за возрастающего масштаба исследований размер и разреженность исходных матриц постоянно увеличивается, что потребовало разработку специальных версий известных методов для работы с большими разреженными массивами данных.

Современные планы экспериментальных и природных исследований всё больше приобретают междисциплинарный характер и требуют тесного взаимодействия экологов, генетиков, химиков и биостатистиков. В этих условиях роль многомерных методов, позволяющих вести обработку массивов информации с большим числом разнородных переменных и, что особенно важно, выполнять качественную проверку достоверности построенных моделей с последующей их селекцией, становится очевидной и приоритетной в условиях климатических изменений и антропогенного воздействия.

Литература

1. Джонгман Р.Г.Г., тер Браак С.Дж.Ф., ван Тонгерен О.Ф.Р. Анализ данных в экологии сообществ и ландшафтов. М.: РАСХН. 1999. 306 с.
2. Зинченко Т.Д. Эколого-фаунистическая характеристика хирономид (Diptera, Chironomidae) малых рек бассейна Средней и Нижней Волги (атлас). Тольятти: Кассандра, 2011. 258 с.
3. Зинченко Т.Д., Головатюк Л.В., Шитиков В.К. Особенности пространственного распределения донных сообществ равнинной реки бассейна Средней Волги // Вестник Томского государственного университета. Биология. 2017. № 40. С. 163–180.
4. Миркин Б.М., Розенберг Г.С. Опыт применения метода главных компонент варьирования растительности // Экология. 1977. № 5. С. 29–37.
5. Миркин Б.М., Розенберг Г.С. Фитоценология. Принципы и методы. М.: Наука, 1978. 212 с.
6. Раменский Л.Г., Цаценкин И.А., Чижиков О.Н., Антипин Н.А. Экологическая оценка кормовых угодий по растительному покрову / М.: Сельхозгиз, 1956. 472 с.
7. Розенберг Г.С. Введение в теоретическую экологию (в 2-х т.). Тольятти: Кассандра, 2013. Т. 1. 556 с. Т. 2. 445 с.
8. Розенберг Г.С., Миркин Б.М. Эволюция и сравнение ординационных техник // Статистические методы анализа почв, растительности и их связи. Уфа: БФАН СССР, 1978. С. 95–117.
9. Шитиков В.К., Зинченко Т.Д., Розенберг Г.С. Макроэкология речных сообществ: концепции, методы, модели. Тольятти: Кассандра, 2012. 257 с.
10. Шитиков В.К., Мастицкий С.Э. Классификация, регрессия и другие алгоритмы Data Mining с использованием R. Электронная книга, 2017. 351 с. [Интернет ресурс] <http://www.ievbras.ru/ecostat/Kiril/Article/A44/DM.htm> (Дата обращения: 24.06.2018)
11. Шитиков В.К., Розенберг Г.С. Рандомизация и бутстреп: статистический анализ в биологии и экологии с использованием R. Тольятти: Кассандра, 2014. 314 с.

12. Bray J.R., Curtis J.T. An ordination of upland forest communities of southern Wisconsin // *Ecological monographs*, 1957. V. 27. P. 325–349.
13. Brückner A., Heethoff M. A chemo-ecologists' practical guide to compositional data analysis // *Chemoecology*. 2017. V. 27. P. 33–46.
14. 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.
15. Chessel D., Hanafi M. Analyses de la co-inertie de K nuages de points // *Revue de statistique appliquée* 1996. V. 44. P. 35–60.
16. Clarke K.R. Non-parametric multivariate analyses of changes in community structure // *Austral. J. Ecol.*, 1993. V. 18. P. 117–143.
17. De'Ath G. Multivariate regression trees: a new technique for modeling species environment relationships // *Ecology*. 2002. V. 83. P. 1105–1117.
18. Digby P.G.N., Kempton R.A. *Multivariate analysis of ecological communities*. London (UK): Chapman and Hall, 1987. 206 p.
19. Dolédec S., Chessel D. Co-inertia analysis: an alternative method for studying species–environment relationships. // *Freshwater Biology*. 1994. V. 31. P. 277–294.
20. Dray S., Chessel D., Thioulouse J. Procrustean co-inertia analysis for the linking of multivariate datasets // *Écoscience*. 2003. V. 10. P. 110–119.
21. 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.
22. Fisher R.A. The use of multiple measurements in taxonomic problems // *Annals of Eugenics*. 1936. V. 7. P. 179–188.
23. 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.
24. 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.
25. Greenacre M.J. *Theory and applications of correspondence analysis*. New York: Wiley, 1984. 364 p.
26. Greig-Smith P. *Quantitative Plant Ecology*. N.Y.; London: Acad. Press; Butter Worths Scientific Publ., 1957. 198 p.
27. 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.
28. 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.
29. Hill M.O. Reciprocal averaging: an eigenvector method of ordination // *Journal of Ecology*. 1973. V. 61. P. 237–249.
30. Hill M.O., Gauch H.G. Detrended correspondence analysis: an improved ordination technique // *Vegetatio*. 1980. V. 42. P. 47–58.
31. Hotelling H. Analysis of a complex of statistical variables into principal components // *Journal of Educational Psychology*. 1933. V. 24. P. 417–441.
32. Kenkel N.C. On selecting an appropriate multivariate analysis // *Canadian Journal of Plant Science*. 2006. V. 86. P. 663–676.
33. Kruskal J.B. Nonmetric multidimensional scaling: a numerical method // *Psychometrika*. 1964. V. 29. P. 115–129.
34. Legendre P., Anderson M.J. Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments // *Ecological Monographs*. 1999. V. 69. P. 1–24.
35. Legendre P., Gallagher E. Ecologically meaningful transformations for ordination of species data // *Oecologia*. 2001. V. 129. P. 271–280.
36. Legendre P., Legendre L. *Numerical Ecology*. Amsterdam: Elsevier Sci. BV, 2012. 1006 p.
37. Liland K.H., Indahl U.G. Powered partial least squares discriminant analysis // *Journal of Chemometrics*. 2009. V. 23. P. 7–18.
38. McCune B., Grace J.B. *Analysis of Ecological Communities*. Glenden Beach (Oregon): MjM Software, 2002. 285 p.
39. Minchin P.R. An evaluation of the relative robustness of techniques for ecological ordination // *Vegetatio*, 1987. V. 67. P. 1167–1179.
40. Oksanen J., Blanchet F.G., Kindt R. et al. *vegan: Community Ecology Package*. R package version 2.5–2. 2018. [Интернет ресурс] <https://cran.r-project.org/web/packages/vegan> (Дата обращения: 24.06.2018).
41. Orloci L. Geometric models in ecology. I. The theory and application of some ordination methods. // *Jour. Ecology*. 1966. V. 54. P. 193–215.

42. Pearson K. On lines and planes of closest fit to systems of points in space // *Philosophical Magazine*. 1901. V. 2. P. 559–572.
43. Prentice I. C. Non-metric ordination methods in ecology // *Journal of Ecology*, 1977. V. 65. P. 85–94.
44. Rao C.R. The use and interpretation of principal component analysis in applied research // *Sankhya*. 1964. Ser. A. V. 26. P. 329–358.
45. Singh A., Gautier B., Shannon C.P., Vacher M., Rohart F., Tebbutt S.J., Le Cao K-A. DIABLO-an integrative, multi-omics, multivariate method for multi-group classification // *BioRxiv*. 2016. N 067611. 50 p. [Интернет ресурс] <https://www.biorxiv.org/content/early/2018/03/20/067611> (Дата обращения: 24.06.2018) .
46. Tenenhaus A, Tenenhaus M. Regularized generalized canonical correlation analysis // *Psychometrika*. 2011. V. 76. P. 257–284.
47. ter Braak C.J. Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis // *Ecology*. 1986. V. 67. P. 1167–1179.
48. Ter Braak C.J.F. Canonical community ordination. Part I: Basic theory and linear methods // *Ecoscience*, 1994. V. 1. P. 127–140.
49. Whittaker R.H. Gradient analysis of vegetation // *Biological Reviews*. 1967. V. 42. P. 207–264.
50. Wold H. Partial least squares // *Encyclopedia of statistical sciences*. Kotz S., Johnson N. (eds). New York: Wiley, 1985. P. 581–591.
51. Zerzucha P., Daszykowski M., Walczak B. () Dissimilarity partial least squares applied to non-linear modeling problems // *Chemometrics and Intelligent Laboratory Systems*. 2012. V. 110. P. 156–162.

References

1. 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.
2. Zinchenko T.D. Ecological and faunal review of chironomids (Diptera, Chironomidae) small rivers in Middle and Low Volga basin (Atlas). Togliatty: Kassandra, 2011. 258 p. (in Russian).
3. Zinchenko T.D., Golovatyuk L.V., Shitikov V.K. Spatial distribution features of bottom communities of a plain river in the Middle Volga river basin // *Vestnik Tomskogo gosudarstvennogo universiteta. Biologia*. 2017. V. 40. P. 163–180. (in Russian).
4. Mirkin B.M., Rosenberg G.S. Experience of application of the method of principal components of vegetation variation // *Ecologia*. 1977. № 5. P. 29–37. (in Russian).
5. Mirkin B.M., Rosenberg G.S. *Phytocenology. Principles and methods*. M.: Nauka, 1978. 212 p. (in Russian).
6. Ramensky L.G., Tsatsenkin I.A., Chizhikov O.N., Antipin N.A. Ecological assessment of forage lands by vegetation cover. M.: Selchozgiz, 1956. 472 p. (in Russian).
7. Rozenberg G.S. *Introduction to the theoretical ecology* (in two volumes). Togliatty: Kassandra, 2013. V. 1. 556 p. V. 2. 445 p. (in Russian)
8. Rosenberg G.S., Mirkin B.M. Evolution and comparison of ordination techniques // *Statistical methods of soil and vegetation analysis and their connections*. Ufa: BFAN SSSR, 1978. P. 95 – 117. (in Russian).
9. Shitikov V.K., Zinchenko T.D., Rozenberg G.S. *Macroecology of river communities: concepts, methods, models*. Togliatty: Kassandra, 2012. 257 p. (in Russian).
10. Shitikov V.K., Mastitsky S.E. Classification, regression and other Data Mining algorithms using R. E-book, 2017. 351 c. [Internet resource] <http://www.ievbras.ru/ecostat/Kiril/Article/A44/DM.htm> (Accessed: 24.06.2018) (in Russian).
11. Shitikov V.K., Rozenberg G.S. Randomization and bootstrap: a statistical analysis in biology and ecology with R use. Togliatti: Kassandra, 2014. 314 p. (in Russian).
12. Bray J.R., Curtis J.T. An ordination of upland forest communities of southern Wisconsin // *Ecological monographs*, 1957. V. 27. P. 325–349.
13. Brückner A., Heethoff M. A chemo-ecologists' practical guide to compositional data analysis // *Chemoecology*. 2017. V. 27. P. 33–46.
14. 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.
15. Chessel D., Hanafi M. Analyses de la co-inertie de K nuages de points // *Revue de statistique appliquée* 1996. V. 44. P. 35–60.
16. Clarke K.R. Non-parametric multivariate analyses of changes in community structure // *Austral. J. Ecol.*, 1993. V. 18. P. 117–143.
17. De'Ath G. Multivariate regression trees: a new technique for modeling species environment relationships // *Ecology*. 2002. V. 83. P. 1105–1117.
18. Digby P.G.N., Kempton R.A. *Multivariate analysis of ecological communities*. London (UK): Chapman and Hall, 1987. 206 p.
19. Dolédec S., Chessel D. Co-inertia analysis: an alternative method for studying species–environment relationships. // *Freshwater Biology*. 1994. V. 31. P. 277–294.

20. Dray S., Chessel D., Thioulouse J. Procrustean co-inertia analysis for the linking of multivariate datasets // *Écoscience*. 2003. V. 10. P. 110–119.
21. 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.
22. Fisher R.A. The use of multiple measurements in taxonomic problems // *Annals of Eugenics*. 1936. V. 7. P. 179–188.
23. 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.
24. 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.
25. Greenacre M.J. Theory and applications of correspondence analysis. New York: Wiley, 1984. 364 p.
26. Greig-Smith P. Quantitative Plant Ecology. N.Y.; London: Acad. Press; Butter Worths Scientific Publ., 1957. 198 p.
27. 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.
28. 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.
29. Hill M.O. Reciprocal averaging: an eigenvector method of ordination // *Journal of Ecology*. 1973. V. 61. P. 237–249.
30. Hill M.O., Gauch H.G. Detrended correspondence analysis: an improved ordination technique // *Vegetatio*. 1980. V. 42. P. 47–58.
31. Hotelling H. Analysis of a complex of statistical variables into principal components // *Journal of Educational Psychology*. 1933. V. 24. P. 417–441.
32. Kenkel N.C. On selecting an appropriate multivariate analysis // *Canadian Journal of Plant Science*. 2006. V. 86. P. 663–676.
33. Kruskal J.B. Nonmetric multidimensional scaling: a numerical method // *Psychometrika*. 1964. V. 29. P. 115–129.
34. Legendre P., Anderson M.J. Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments // *Ecological Monographs*. 1999. V. 69. P. 1–24.
35. Legendre P., Gallagher E. Ecologically meaningful transformations for ordination of species data // *Oecologia*. 2001. V. 129. P. 271–280.
36. Legendre P., Legendre L. Numerical Ecology. Amsterdam: Elsevier Sci. BV, 2012. 1006 p.
37. Liland K.H., Indahl U.G. Powered partial least squares discriminant analysis // *Journal of Chemometrics*. 2009. V. 23. P. 7–18.
38. McCune B., Grace J.B. Analysis of Ecological Communities. Gleneden Beach (Oregon): MjM Software, 2002. 285 p.
39. Minchin P.R. An evaluation of the relative robustness of techniques for ecological ordination // *Vegetatio*, 1987. V. 67. P. 1167–1179.
40. Oksanen J., Blanchet F.G., Kindt R. et al. vegan: Community Ecology Package. R package version 2.5–2. 2018. [Интернет ресурс] <https://cran.r-project.org/web/packages/vegan> (Дата обращения: 24.06.2018).
41. Orloci L. Geometric models in ecology. I. The theory and application of some ordination methods. // *Jour. Ecology*. 1966. V. 54. P. 193–215.
42. Pearson K. On lines and planes of closest fit to systems of points in space // *Philosophical Magazine*. 1901. V. 2. P. 559–572.
43. Prentice I. C. Non-metric ordination methods in ecology // *Journal of Ecology*, 1977. V. 65. P. 85–94.
44. Rao C.R. The use and interpretation of principal component analysis in applied research // *Sankhya*. 1964. Ser. A. V. 26. P. 329–358.
45. 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. N 067611. 50 p. [Интернет ресурс] <https://www.biorxiv.org/content/early/2018/03/20/067611> (Дата обращения: 24.06.2018).
46. Tenenhaus A, Tenenhaus M. Regularized generalized canonical correlation analysis // *Psychometrika*. 2011. V. 76. P. 257–284.
47. Ter Braak C.J. Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis // *Ecology*. 1986. V. 67. P. 1167–1179.
48. Ter Braak C.J.F. Canonical community ordination. Part I: Basic theory and linear methods // *Ecoscience*, 1994. V. 1. P. 127–140.
49. Whittaker R.H. Gradient analysis of vegetation // *Biological Reviews*. 1967. V. 42. P. 207–264.
50. Wold H. Partial least squares // *Encyclopedia of statistical sciences*. Kotz S., Johnson N. (eds). New York: Wiley, 1985. P. 581–591.

51. Zerzucha P., Daszykowski M., Walczak B. () Dissimilarity partial least squares applied to non-linear modeling problems // Chemometrics and Intelligent Laboratory Systems. 2012. V. 110. P. 156–162.

УДК 574.2

ПРОБЛЕМА УСТОЙЧИВОСТИ ЦВЕТОЧНО - ДЕКОРАТИВНЫХ РАСТЕНИЙ И ГАЗОННЫХ ТРАВ К ПОВЫШЕННОМУ СОДЕРЖАНИЮ МЕДИ В ОКРУЖАЮЩЕЙ СРЕДЕ

Гладков Евгений Александрович¹, Ташлиева Илина Игоревна¹,

Долгих Юлия Ивановна², Гладкова Ольга Викторовна^{*}

**¹ФГБУН Институт Физиологии Растений им. К.А. Тимирязева РАН ²ФГБНУ ВНИИ
сельскохозяйственной биотехнологии
gladkovu@mail.ru**

декоративные растения, газон, медь

В статье рассмотрена проблема устойчивости декоративных растений к повышенному содержанию меди в окружающей среде. Рассмотрены технологии, позволяющие уменьшить воздействие меди на декоративные растения.

THE PROBLEM OF STABILITY OF FLOWER ORNAMENTAL PLANTS AND LAWN GRASSES TO HIGH LEVELS OF COPPER IN THE ENVIRONMENT

Gladkov E. A.¹, Tashlieva I.I.¹ Dolgikh Y. I.², Gladkova O. V.^{*}

¹Timiryazev Institute of Plant Physiology Russian Academy of Sciences,

²All-Russia Research Institute of Agricultural Biotechnology

lawn grass, ornamental plants, copper

The article deals with the problem of the resistance of ornamental plants to a high copper content in the environment. Considered technology to reduce the impact of copper on ornamental plants.

Красивоцветущие декоративные травянистые растения и газонные травы имеют важное значение в городском озеленении и декоративном садоводстве. Однако эти растения чувствительны к загрязнению почв тяжелыми металлами, среди которых медь обладает достаточно высоким уровнем фитотоксичности, существенно ограничивая использование декоративных растений. Концентрации меди в почве, которые ограничивают распространение растений и снижают декоративные качества согласно нашим исследованиям составляют 30-50 мг/ кг, в зависимости от вида. Эти концентрации меди ниже ПДК (55 мг/ кг) и ОДК(рН КС1 >5,5) (132 мг/ кг) [1]. Таким образом, медь оказывает неблагоприятное влияние на растения при относительно невысоком содержании в почве. Способы решения данной проблемы - уменьшение уровня загрязнения почвенного покрова медью и культивирование более устойчивых растений. Повысить жизнеспособность растительного покрова можно с использованием дикорастущих видов растений, однако эти растения не всегда обладают высокими декоративными свойствами. Другой подход - использование в городском озеленении новых технологий. Среди таких технологий особое место занимает клеточная селекция. Однако работы по получению растений, устойчивых к меди единичны [4]. Использование клеточной селекции подразумевает введение растений в культуру клеток. Нами были разработаны технологии получения каллусных культур различных декоративных растений [2,3], для ряда растений, например для декоративного льна (рис.1), это оказалось очень трудоемким процессом.