

ИСПОЛЬЗОВАНИЕ МЕТОДОВ КЛАСТЕР-АНАЛИЗА ДЛЯ ВЫДЕЛЕНИЯ СООБЩЕСТВ ОРГАНИЗМОВ (НА ПРИМЕРЕ ГИЛЬДИИ ГРЫЗУНОВ)

В работе рассматриваются особенности применения кластер-анализа, как одного из способов выделения сообществ организмов в континуальных местообитаниях. Обосновываются основные преимущества классификации, производимой на основании значений координатных осей многомерного шкалирования. Особое внимание уделяется проверке качества классификации и выделению переменных, имеющих ключевое значение при разделении множества на естественные группы.

Ключевые слова: сообщества, континуальное местообитание, гильдия грызунов, классификация, кластер-анализ.

Выделение сообществ живых организмов в континуальных местообитаниях в настоящее время составляет важный раздел теоретической и прикладной экологии. Многие современные методы многомерного анализа данных позволяют отобразить экологическую систему, определенную на основе множества ее элементов в системе координат конечной размерности в непрерывной форме [8]. Однако в природных условиях мы чаще всего имеем дело с дискретными множествами, отображать которые наиболее удобно, используя различные методы классификации.

Сложность классификации экологических сообществ заключается в том, что они, в отличие от таких категорий, как виды живых организмов — весьма неоднородные формирования, часто существующие в общем пространстве континуального местообитания. В то же время, при решении различных исследовательских задач приходится иметь дело с вполне конкретным местообитанием и населяющей его совокупностью организмов, которое исследуется лишь определенный период времени, т. е. с локальным или частным экологическим пространством [8; 13].

Сам процесс классификации сводится к образованию непересекающихся друг с другом подмножеств элементов (классов), выделенных из общего множества на основании определенного алгоритма. Классы выделяются на основе использования определенных логических отношений, которые бывают нескольких типов. Более подробно с ними можно ознакомиться в целом ряде работ [12, 1, 8 и др.].

Цель данной работы — демонстрация и обсуждение методов выделения типов сообществ организмов с использованием методов кластерного анализа.

Удобным объектом для целого ряда экологических исследований являются мелкие млекопитающие. Популярность мелких млекопитающих определяется их важной ценотической ролью в экосистемах, чувствительностью к изменениям среды, относительно коротким жизненным циклом и динамичностью популяционных процессов [2, 3].

В работе были использованы суммарные данные отловов мелких млекопитающих, которые выполнялись автором на территории Центрально-Лесного государственного биосферного природного заповедника (ЦЛГПБЗ) в июле — сентябре

2012 г. Полевые работы производились на трансекте общей протяженностью 2280 м, пересекающей различные типы коренных экосистем южной тайги. Трансекта отмечена 114 визированными точками, которые привязаны при помощи GPS в системе координат WGS 84 (UTM Zone 36 North). Расстояние между соседними точками составляет 20 метров. Точки имеют подробные геоботанические описания, включающие 275 переменных, прежде всего, фитоценотической среды, которые представлены комплексными характеристиками древостоя различных ярусов, характеристиками подлеска, травяно-кустарничкового и мохового ярусов. Также в описание включены некоторые геоморфологические и микроклиматические характеристики среды [5, 6]. Все этапы математической обработки данных были выполнены в программах MS Excel, Statistica 8.0 и Fracdim.

Одним из традиционных подходов в изучении структуры и функционирования сообществ является концепция гильдии, т. е. группы видов, которые используют набор ресурсов экосистемы сходным образом [2]. В связи этим, комплекс населяющих определенную территорию мелких млекопитающих принято разделять на гильдии грызунов и землероек [2]. В нашем случае наиболее репрезентативные данные получены для гильдии грызунов (711 экз.), поэтому демонстрацию подхода мы будем осуществлять именно на ее примере.

В фаунистический список мелких млекопитающих исследуемой территории включено 16 видов мышевидных грызунов [3], из которых 11 в последние годы довольно регулярно встречались в учетных отловах из лесных экосистем [4].

В наших отловах на трансекте было отмечено 7 видов грызунов. Среди них абсолютным доминантом является рыжая полевка (*Clethrionomys glareolus Schreber*) (87,4 %), довольно многочисленны красная полевка (*Clethrionomys rutilus Pall.*) (4,6 %), желтогорлая (*Apodemus flavicollis Meichior*) (4,6 %) и малая лесная (*Apodemus microps Kratochvil et Rosicky*) (2,6 %) мыши. Прочие виды, такие как темная полевка (0,4 %), лесной лемминг (0,26 %) и полевая мышь (0,14 %) очень редки и встречаются спорадически.

Классификацию гильдии грызунов можно проводить как на основе непосредственного распределения самих видов (прямая классификация), так и с использованием координат, полученных методом многомерного непараметрического шкалирования [6; 7; 8].

Одной из общих проблем метода классификации является выбор числа классов. Для корректного отображения числа классов необходимо осуществить оценку размерности анализируемого пространства. Если оценка размерности произведена тем или иным методом, то число классов должно быть равно $k=D+1$, где D — размерность.

Определить оптимальную размерность пространства возможно при помощи метода многомерного шкалирования. Не вдаваясь в процедуру самого метода, который достаточно подробно описан в целом ряде работ [7; 8; 9; 11], а также был подробно рассмотрен нами [6], напомним, что метод позволяет определить оптимальную размерность пространства по значению стресс-функции, которая возрастает по мере уменьшения размерности пространства в ходе процедуры многомерного шкалирования (Рис. 1.)

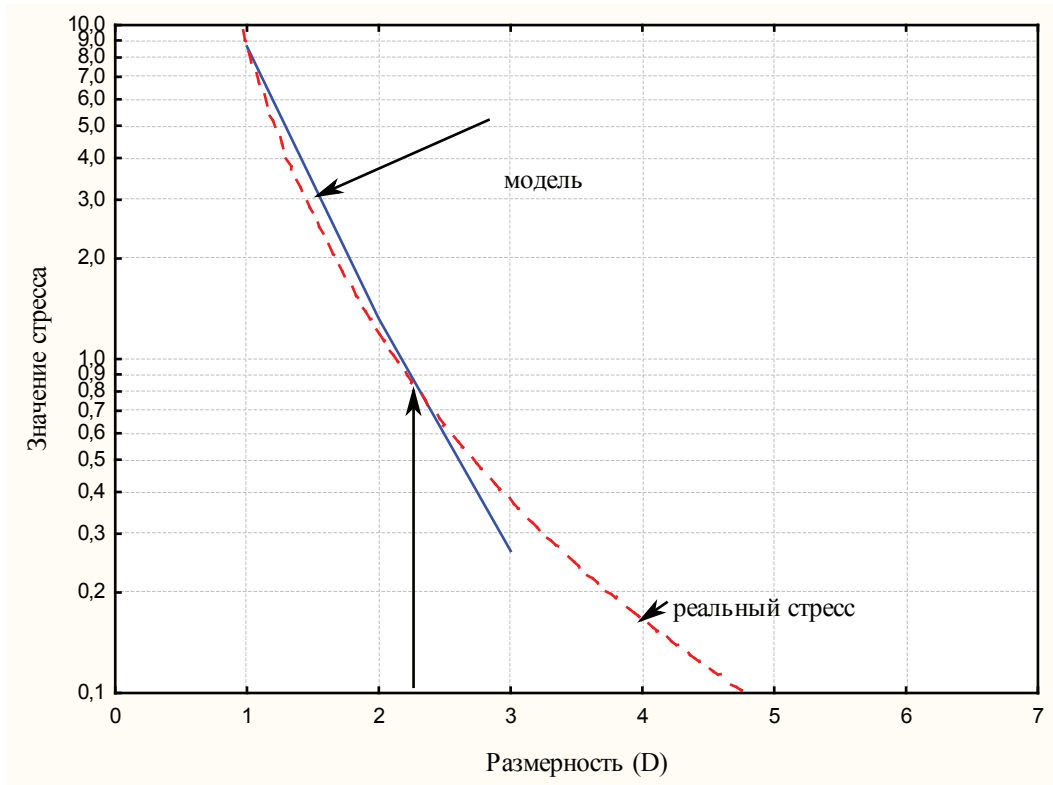


Рис. 1. Оценка размерности анализируемого экологического пространства по функции стресса, полученная в ходе многомерного шкалирования

Как видно из Рис. 1, оптимальная размерность пространства находится в промежутке, между 2 и 3. В случае такого промежуточного положения имеет смысл округление в сторону большей размерности. Таким образом, оптимальная размерность пространства равна трем. В результате, согласно приведенному выше алгоритму, получаем, что оптимальное число классов для анализируемой экологической системы равно четырем ($k = 4$).

Классификация сообществ грызунов может осуществляться как на основе собственно численности самих видов (прямая классификация), так и по значениям трех координатных осей, полученных методом многомерного шкалирования.

Прямая классификация, по большей части, строится на основе участия в классах самых широко распространенных видов и учитывает, в первую очередь, наиболее многочисленными из них. Недостатком такой классификации является то, что в ней по определению занижается роль редких видов, которые распределяются по остаточному принципу. Это приводит к тому, что особенности размещения редких видов в пространстве не всегда отображаются корректно.

Классификация, осуществляемая по осям многомерного шкалирования, изначально основана на матрице дистанций между видами без учета значений их абсолютной численности. В результате преобразования матрицы дистанций в ходе процедуры многомерного шкалирования мы получаем коэффициенты чувствительности

видов к координатным осям, которые отражают положение всех анализируемых видов в пространстве заданной размерности (Табл. 1).

Таблица 1

Коэффициенты чувствительности видов к координатным осям экологического пространства

Вид	Номер координатной оси			Индикатор гомогенных подобластей.		
	1	2	3	1	2	3
Рыжая полевка	0,20114	0,177960	-0,049692	+	+	-
Красная полевка	-0,53597	-0,550849	0,100983	-	-	+
Темная полевка	0,18031	-0,857306	-0,614200	+	-	-
Лесной лемминг	0,45144	0,859281	-0,647715	+	+	-
Малая лесная мышь	-0,43937	0,274308	0,770935	-	+	+
Желтогорлая мышь	-1,05260	0,248832	-0,124820	-	+	-
Полевая мышь	1,19505	-0,152226	0,564509	+	-	+

При этом каждый вид имеет свои собственные коэффициенты чувствительности к координатным осям и занимает свою подобласть экологического пространства [6]. Данные Табл 1. показывают, что каждый из представленных видов занимает свою подобласть экологического пространства и имеет свои коэффициенты чувствительности к координатным осям.

Таким образом, классификация на основе координатных осей многомерного шкалирования в гораздо большей степени отображает реальный характер размещения грызунов в пространстве.

На следующем этапе анализа была произведена классификация сообществ грызунов методом k-средних. Данный метод относится к т. н. девизионным методам классификации (классификация сверху) [8; 10]. В данном методе число классов определяет исследователь на основе знаний о размерности пространства. На первом этапе классификации алгоритм выбирает из всего множества k-точек, каждая из которых описана n-переменными. После того, как выбраны k-элементов системы, все остальные последовательно сравниваются по дистанции с каждым из k и присоединяются к тому из них, который оказывается ближе. В результате из множества образуется k-кластеров, с различным числом элементов в каждом. Для каждого кластера рассчитывается среднее значение переменных и абстрактную точку, которая соответствует центру тяжести кластера. Процедура носит итерационный характер и повторяется до тех пор, пока не будет найдена оптимальная конфигурация центров тяжести, вокруг которых максимально плотно группируются все элементы анализируемого множества [8].

На Рис. 2 показана структура населения грызунов для каждого из выделенных типов сообществ.

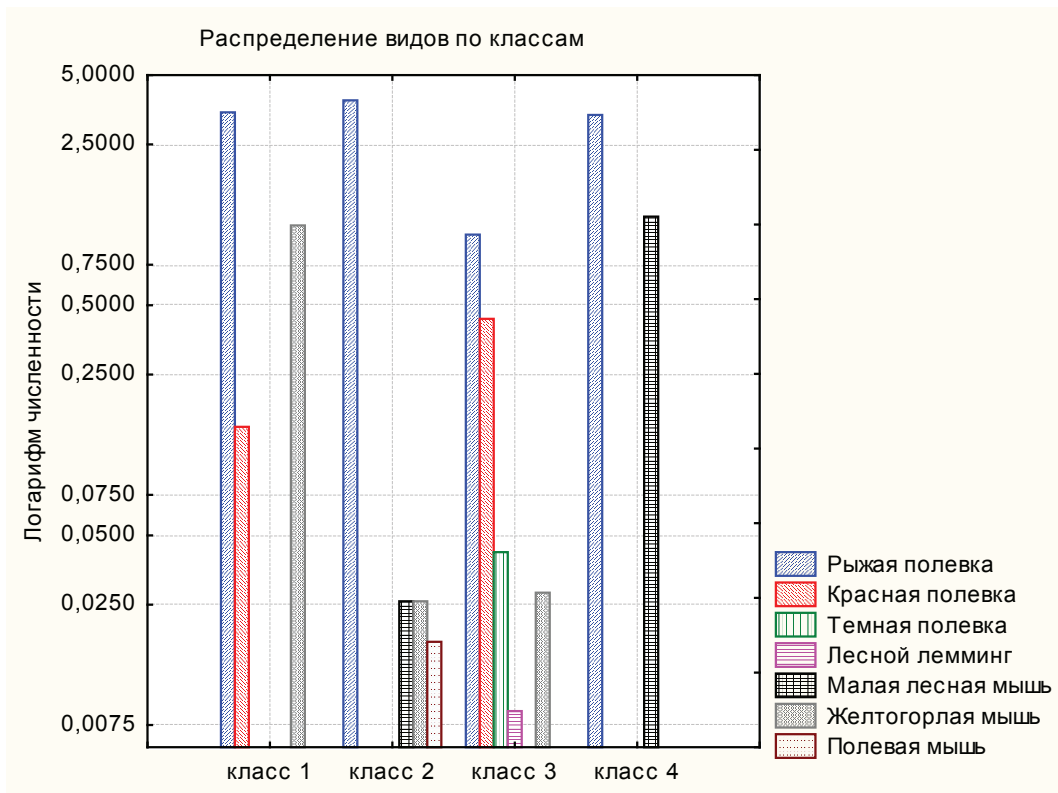


Рис. 2. Распределение грызунов по типам сообществ

В первом сообществе господствует рыжая полевка, а в качестве субдоминанта выступает желтогорлая мышь. С несколько более низкой численностью в нем представлена красная полевка.

Второй тип сообществ характеризуется сильным доминированием рыжей полевки и наличием сопутствующих видов: малой лесной, желтогорлой и полевой мышей, численность которых крайне низка.

Третий тип сообществ демонстрирует наибольшее видовое разнообразие. Здесь численность рыжей полевки несколько снижается, хотя она и остается доминирующим видом, а в качестве субдоминанта выступает красная полевка. Также в данном типе сообществ присутствуют темная полевка и желтогорлая мышь. Спорадически представлен лесной лемминг.

Четвертый тип сообществ характеризуется наибольшей бедностью видового состава и представлен всего двумя видами: рыжей полевкой и малой лесной мышью. Однако численность обоих видов весьма высока.

Надежность классификации была проверена при помощи дискриминантного анализа (метод Forward stepwise). Четыре класса выделяются на 98,2 % по структуре животного населения. Общая таблица ошибок (Табл. 2.) показывает, что дискриминантная функция выдает только 1,8 % ошибок, что позволяет считать полученную классификацию вполне надежной.

Таблица 2

Оценка качества классификации по частоте ошибочной дискриминации (строка таблицы — наблюдаемый класс; столбец — рассчитанный класс)

Класс	Процент корректной дискриминации	Класс 1	Класс 2	Класс 3	Класс 4
1	100,0000	27,00000	0,0000	0,00000	0,00000
2	98,2759	2,00000	114,0000	0,00000	0,00000
3	97,1831	2,00000	0,0000	69,00000	0,00000
4	100,0000	0,00000	0,0000	0,00000	14,00000
Всего	98,2456	31,00000	114,0000	69,00000	14,00000

Для выяснения роли каждого из видов грызунов в разделении исходного множества на четыре класса проводится общая оценка качества произведенной классификации по переменным (видам грызунов). Для надежности определения роли переменных были проведены два варианта пошагового дискриминантного анализа (Forward stepwise и Backward stepwise) которые, в данном случае, дали тождественные результаты (Табл. 3).

Таблица 3

Общая оценка качества дискриминантного анализа для четырех классов по 7 переменным. Модель включила 5 переменных. $P < 0,0000$

Переменная	Критерии					
	Вилкоксон лямбда	Частный Вилкоксон лямбда	F-критерий	p-level	Толеранс	Коэффициент детерминации
Желтогорлая мышь	0,125942	0,166374	367,4396	0,000000	0,776999	0,223001
Малая лесная мышь	0,089492	0,234140	239,8701	0,000000	0,835384	0,164616
Рыжая полевка	0,037220	0,562960	56,9306	0,000000	0,939823	0,060177
Красная полевка	0,027784	0,754155	23,9058	0,000000	0,886603	0,113397
Темная полевка	0,021555	0,972098	2,1049	0,100505	0,983978	0,016022

Исходя из данных классификации, ведущее значение при разделении классов по частному критерию Вилкоксон-лямбда, который указывает на вклад в дискриминацию конкретной переменной, и F-критерию имеет желтогорлая мышь, второе место занимает малая лесная мышь. Значения толеранса, который показывает степень избыточности переменной, показывает, что все представленные в таблице виды вносят существенный вклад в разделение классов и не один из них нельзя считать избыточным. Коэффициент детерминации показывает, что степень зависимости видов друг от друга очень низка.

Таким образом, можно полагать, что четыре типа сообществ, которые были выделены в ходе анализа, представляют собой естественные группы единого южно-таежного континуального местообитания. Классификация сообществ организмов, произведенная по значениям координат осей многомерного шкалирования, позволяет

адекватно отобразить особенности размещения сообществ в континуальных местообитаниях, поскольку при такой классификации учитываются, прежде всего, особенности размораживания организмов в экологическом пространстве, а не численный вклад в выборку того или иного вида.

Литература

1. Айвазян С. А., Бухштабер В. М., Енюков И. С., Мешалкин Л. Д. Прикладная статистика. Классификация и снижение размерности. М.: Финансы и статистика, 1989. 607 с.
2. Истомин А. В. Мелкие млекопитающие в региональном экологическом мониторинге (на примере Каспийско-Балтийского водораздела). Монография. Псков, 2008. 278 с.
3. Истомин А. В. Некоторые реакции биоты на изменение климата в лесных ландшафтах Каспийско-Балтийского водораздела // Вестник Российского государственного университета им. Иммануила Канта. Вып. 7.: Сер. Естественные науки. Калининград: Изд-во РГУ И. Канта, 2009. С. 15–22.
4. Истомин А. В. Влияние ветровалов на динамику сообществ мелких млекопитающих в естественных лесах южной тайги // Вестник Московского государственного университета леса — Лесной вестник. 2009. № 1 (64). С. 196–202.
5. Истомин А. В., Михалап С. Г. Градиентный подход и ГИС-анализ при изучении пространственной динамики популяций и сообществ организмов // Вестник Псковского государственного университета. Серия «Естественные и физико-математические науки» № 1. Псков: Изд-во ПсковГУ. 2012. С. 49–55.
6. Михалап С. Г. Некоторые подходы к описанию экологических ниш видов (на примере грызунов) // Вестник ПсковГУ: Серия «Естественные и физико-математические науки». Вып. 2. Псков. ПсковГУ, 2013. 220 с. С. 29–35.
7. Пузаченко Ю. Г., Кузнецов Г. В. Экологическая дифференциация грызунов сезонновлажных лесов Южного Вьетнама // Зоол. журн. Т. 77. Вып. 1. 1998. С. 117–132.
8. Пузаченко Ю. Г. Математические методы в экологических и географических исследованиях: Учебн. пособие для студ. вузов. М.: Изд. Центр «Академия», 2004. 416 с.
9. Пузаченко Ю. Г., Черенков С. Е. Принципы организации сообществ птиц смешанных лесов Подмоскovie // Сиб. экол. журн. Т. 14. № 4. 2007. С. 567–576.
10. Халафян А. А. STATISTICA 6. Статистический анализ данных. 3-е изд. Учебник. М.: ООО «Бином-Пресс», 2007. 512 с.
11. Holland S. M. Non – metric multidimensional scaling (MDS). Department of Geology. University of Georgia, Athens, GA 30602–2501. 2008. 8 p.
12. Ludwig J. A., Reynolds J. F. Statistical ecology: a primer on methods & computing. John Wiley & Sons, Inc. New York, NY, USA. 1988. 337 p.
13. Pulliam H. R. On the relationship between niche and distribution // Ecology Letters. 3. 2000. P. 349–361.

USE OF THE CLUSTER-ANALYSIS METHODS FOR DIFFERENTIATION OF BIOTIC COMMUNITIES BY THE EXAMPLE OF RODENT GUILD

S. Mihalap

The paper discusses the peculiarities of the cluster analysis as a mean for community segregation in continual habitats. The main advantages of the classification, which is produced on the values of the coordinate axes of the multidimensional scaling, are proved. The particular attention is paid to quality control of the classification and determination of the variables that are crucial for the separation of the set into the natural groups.

Key words: *community, continual habitat, guild of rodents, classification, cluster analysis*