

# Инструкция по использованию надстройки ExStatR

Разработчики:  
Новаковский А.Б., Сабитов Д.А

*novakovsky@ib.komisc.ru,*  
*sabitovdemon@mail.ru*

©Институт биологии Коми НЦ  
УрО РАН

Сыктывкар, 2017

## Оглавление

Название раздела	Стр.
1. Системные требования, совместимость и установка.....	3
2. Начало работы.....	3
3. Пункт меню "Настройки..." .....	4
4. Подготовка данных для анализа.....	5
5. Выбор данных для анализа.....	6
6. Пункт меню "Метод главных компонент..." .....	7
7. Пункт меню "Неметрическое многомерное шкалирование..." .....	10
8. Табличное представление результатов ординации.....	12
9. Пункт меню "Кластерный анализ..." .....	13
10. Наложение дополнительных факторов на дендрограмму или ординационную диаграмму.....	15
11. Пункт меню "Теплокарты..." .....	18
12. Список дополнительных пакетов и их установка.....	21
13. Список R файлов (скриптов) надстройки ExStatR.....	22
14. Литература.....	23

## 1. Системные требования, совместимость и установка

Для работы надстройки ExStatR необходимы следующие компоненты установленные на компьютере пользователя: 32 или 64 разрядная версия Microsoft Excel из поставки Microsoft Office (2003, 2007, 2010); свободно распространяемая программа статистической обработки "R" версии 3.2.2. или выше (скачать можно по адресу <https://cran.r-project.org>) и дополнительные пакеты расширяющие функционал R (см. раздел. 12).

Общий объем программы R, дополнительных статистических пакетов и самой надстройки ExStatR – около 150 мб. на жестком диске.

*Примечание.* При установке надстройки на систему Windows XP возможно возникновение ошибок с русскими названиями. Рекомендуется использовать латиницу во всех обозначениях.

## 2. Начало работы

Надстройка ExStatR не является самостоятельной программой и не обладает собственным интерфейсом. ExStatR подключается к Microsoft Excel. Доступ к функциям осуществляется через дополнительный пункт меню Excel рис. 1 и 2.

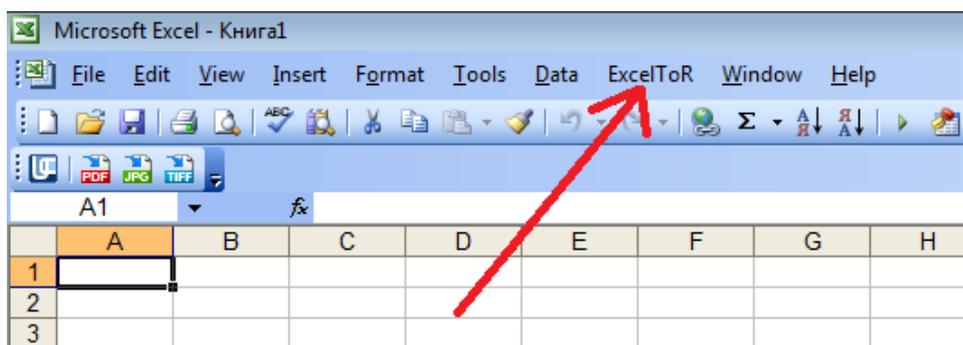


Рис. 1. Расположение пункта меню ExStatR для версии Microsoft Excel 2003 (и более ранних)

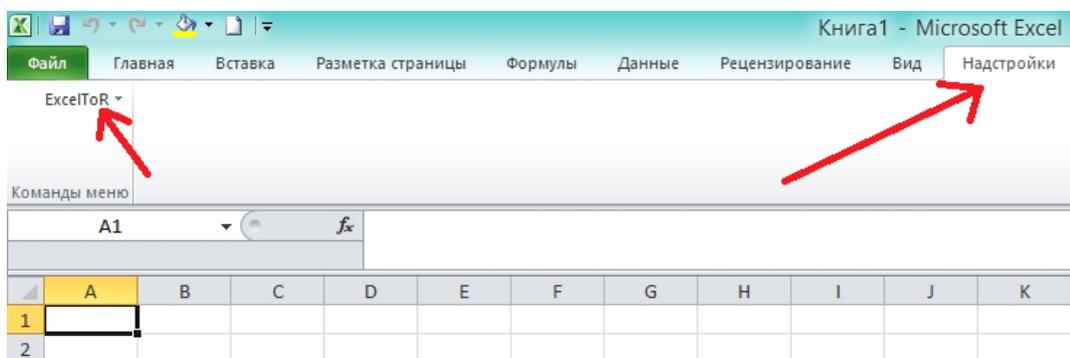


Рис. 2. Расположение пункта меню ExStatR для версии Microsoft Excel 2007 и 2010

После выбора в меню пункта "ExStatR" появляется основное меню надстройки (рис. 3), которое содержит полный список реализованных функций. На сегодняшний день в надстройке реализованы следующие алгоритмы обработки данных:

- Метод главных компонент (Principal component analysis – PCA)
- Неметрическое многомерное шкалирование (Nonmetric multidimensional scaling – NMS)
- Кластерный анализ (cluster analysis)
- Построение теплокарт (Heatmaps).

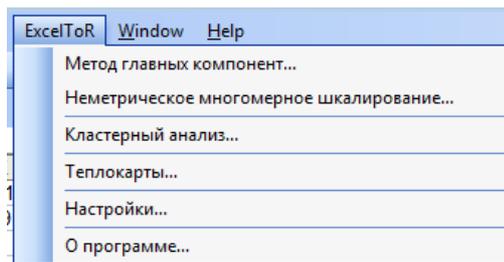


Рис. 3. Основное меню надстройки

Кроме перечисленных методов анализа в основном меню находятся пункты: "Настройки..." и "О программе...".

### 3. Пункт меню "Настройки..."

В меню "настройки" (рис. 4) доступны следующие пункты:

- "*Путь к файлу RScripts.exe*" – путь к файлу RScripts.exe из статистического пакета R, который отвечает за выполнение скриптов для всех типов анализа. По умолчанию файл RScripts.exe находится в папке "C:\Program files\R\номер версии R\bin\i386 (или x64)\". Папка "i386" создается при установке 32 битной версии, "x64" – при установке 64 битной версии программы R.

**Примечание.** В самой папке bin также расположен файл RScripts.exe. Его использовать нельзя. При выборе этого файла, в дальнейшем, при выполнении любого анализа, будет возникать ошибка!

- "*Путь к скриптам R*" – выбор папки где расположены скрипты выполняемые в программе R (см. список скриптов, раздел 13). По умолчанию папка со скриптами расположена внутри папки ExStatR. Обычно это путь "C:\Program Files (x86)\ExStatR\RScripts"
- "Размер изображения (в пикселях)" – задает размер результирующих диаграмм. Минимальные размеры 100x100 пикселей. Можно выбрать любые другие, комфортные для пользователя размеры. (Рекомендуется установить размеры изображения в два раза меньше, чем установленное разрешение экрана).

- "Язык" – выбор языка интерфейса. На текущий момент поддерживается два языка интерфейса: русский и английский.

*Примечание.* По умолчанию, используется английский язык. Поэтому при возникновении ошибок в работе модуля, интерфейс переключается на английский язык.

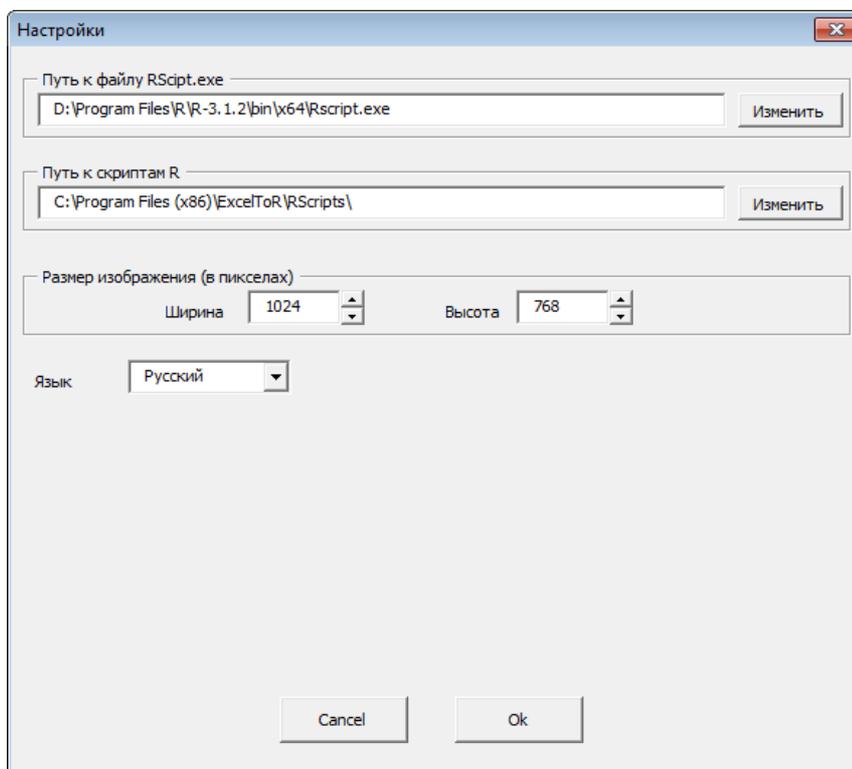


Рис. 4. Внешний вид диалогового окна "Настройки"

#### 4. Подготовка данных для анализа

В качестве исходных данных используется стандартная Excel таблица. Столбцы таблицы представляют собой перечень анализируемых объектов, строки – их характеристики (или наоборот строки – объекты, столбцы – характеристики). В ячейках таблицы ставятся числовые значения параметров.

Для удобства пользователей области содержащие названия объектов и данные для анализа определяются автоматически. Для корректного определения этих областей желательно придерживаться следующей схемы: названия объектов – одна верхняя строка, названия свойств – одна левая колонка (рис. 5).

*Примечание.* Пустая клетка интерпретируется надстройкой как нулевое значение. Наличие текстовых значений вне верхней строки и левой колонки интерпретируется как ошибка в данных.

По умолчанию (если пользователь не выделил какую-либо область), в анализ включаются все данные, расположенные на текущем листе Excel. При выделении пользователем области данных, все вычисления ведутся только с этой областью.

Microsoft Excel - Ординация и кластерный анализ.xls

File Edit View Insert Format Tools Data ExcelToR Window Help

M25

Species	42-2004	181-2008	67-2004	18-2004	19-2004.2	63-2008	17-2008	42-2009-2	24-2004	40-2004
Abies sibirica	1	1	3	3	3	2	3	4	5	1
Abietinella abietina	1	1	0	0	0	1	0	0	3	0
Aconitum septentrionale	2	2	0	5	0	0	2	0	0	0
Actaea erythrocarpa	0	0	0	0	0	0	1	0	2	5
Adoxa moschatellina	1	1	2	0	5	0	0	0	5	0
Agrostis tenuis	0	0	0	0	0	5	0	2	0	0
Alchemilla sp.	0	0	1	1	0	0	0	0	0	0
Alnus incana	0	0	0	0	0	0	0	0	0	1
Anemonastrum biarmense	0	0	0	0	2	2	0	0	2	0
Angelica archangelica	0	0	0	2	0	0	2	0	0	1
Angelica sylvestris	0	0	1	1	0	2	0	0	2	0
Antennaria dioica	0	0	0	0	0	0	0	1	2	1
Anthoxantum alpinum	1	1	0	0	3	0	0	0	0	0
Asplenium viride	0	0	0	0	0	2	0	2	0	0
Aster alpinus	0	0	0	0	0	0	0	0	2	0
Betula nana	0	0	0	0	0	0	1	0	0	0

Анализируемые объекты

Свойства объектов

Числовые данные

Рис. 5. Таблица с данными, подготовленными для анализа

## 5. Выбор данных для анализа

Большинство методов анализа (кластерный анализ, метод главных компонент и неметрическое шкалирование) состоит из двух диалоговых окон: выбор данных для анализа и настройки самого метода.

На рис. 6 приведено диалоговое окно выбора данных для анализа.

- "Элементы для анализа" – список названий объектов, которые будут использоваться при построении диаграмм.
- "Диапазон подписей" – диапазон ячеек в которых содержатся названия элементов. По умолчанию, при выбранном переключателе "Использовать колонки", это верхняя строка, при "Использовать строки" – левая колонка.
- "Диапазон данных" – область, содержащая числовые данные для анализа.
- Переключатель использовать колонки/строки – пользователь определяет строки или столбцы таблицы считать за анализируемые объекты.

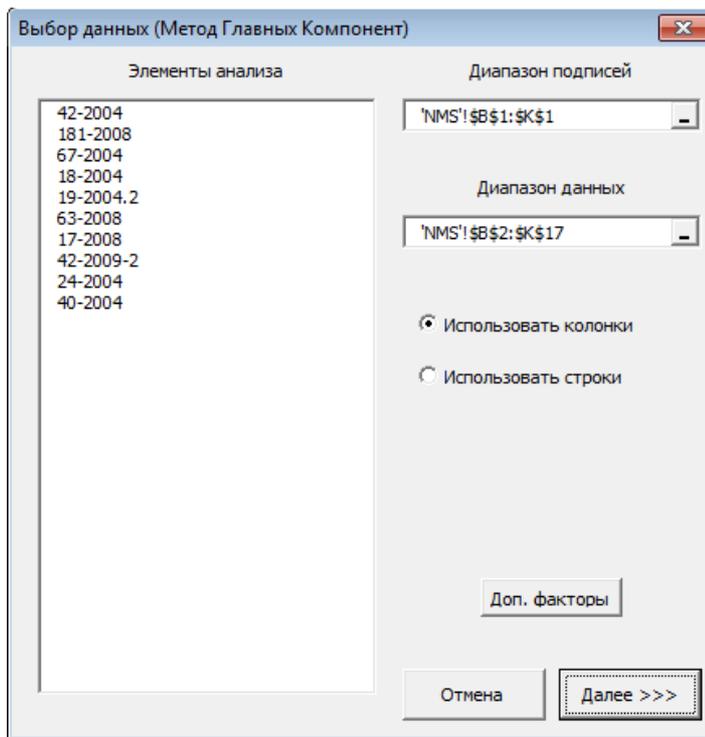


Рис. 6. Диалоговое окно выбора элементов для анализа. (Используется в методах ординации PCA и NMS и кластерном анализе)

Кнопка "Доп. факторы" вызывает диалоговое окно наложения дополнительных факторов на результирующую диаграмму (см. раздел 10)

## 6. Пункт меню "Метод главных компонент..."

Этот пункт реализует непрямую ординацию методом главных компонент (PCA – Principal component analysis). Анализ состоит из двух диалоговых окон: окно выбора данных (см. раздел 5) и окно настроек PCA анализа (рис. 7).

Детальное описание метода можно найти в литературе (см. раздел 14).

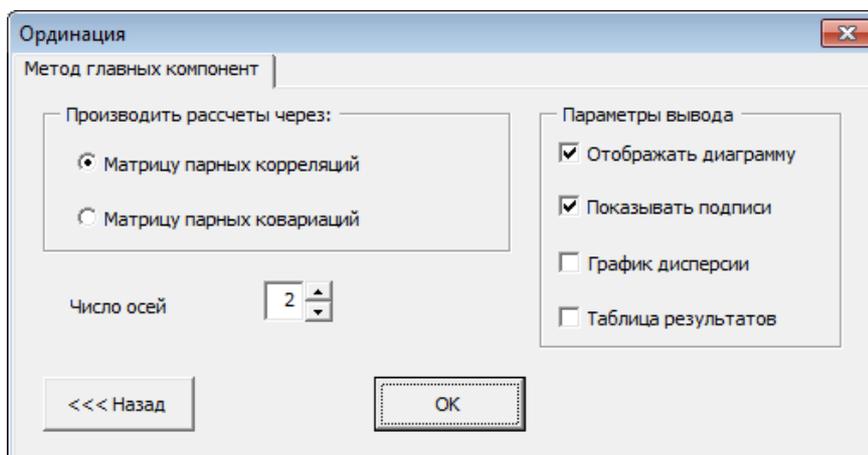


Рис. 7. Диалоговое окно настроек для PCA анализа (метод главных компонент)

Окно настроек PCA ординации содержит следующие пункты:

- Производить расчеты через:
  - Матрицу парных корреляций – для поиска компонент используется корреляционная матрица. Т.е. связанность признаков друг с другом определяется при помощи коэффициента корреляции Пирсона.
  - Матрицу парных ковариаций – для поиска компонент используется ковариационная матрица. Связанность признаков определяется при помощи коэффициента ковариации.

*Примечание.* Принципиального различия между этими способами расчетов нет. Тогда возникает вопрос, какую матрицу лучше использовать для построения главных компонент: корреляционную или ковариационную. Однозначного ответа здесь нет, и вряд ли можно привести какие-либо рекомендации со статистической точки зрения в пользу одного из вариантов. Возможно, из двух способов оценки связности признаков следует выбирать тот, для которого на первые несколько компонент приходится максимальная доля дисперсии всего комплекса анализируемых признаков (рис. 8), однако это соображение является условным и не может служить универсальной рекомендацией для ее применения на практике.

- Число осей – задает количество осей PCA ординации (минимальное значение 2, максимальное – 6).

*Примечание.* Вне зависимости от количества выбранных осей, результирующая диаграмма строится для первых двух осей ординации. Заданное количество осей влияет на "таблицу результатов" (см. раздел 8).

- Параметры вывода
  - Отображать диаграмму – отобразить результирующую диаграмму (рис. 8).
  - Показывать подписи – отображать или нет на результирующей диаграмме названия объектов (при большом числе анализируемых объектов > 30 рекомендуется отключать эту опцию).
  - График дисперсии – доля объясненной вариабельности (дисперсии) признака объясненная соответствующей осью ординации (собственные значения матрицы). Чем число выше, тем большую долю дисперсии объясняет соответствующая ось (рис. 9). Этот график используется для определения количества осей PCA ординации (подробнее можно найти в Интернете: правило Кайзера и правило сломанной стрости).
  - Таблица результатов – см. раздел 8.

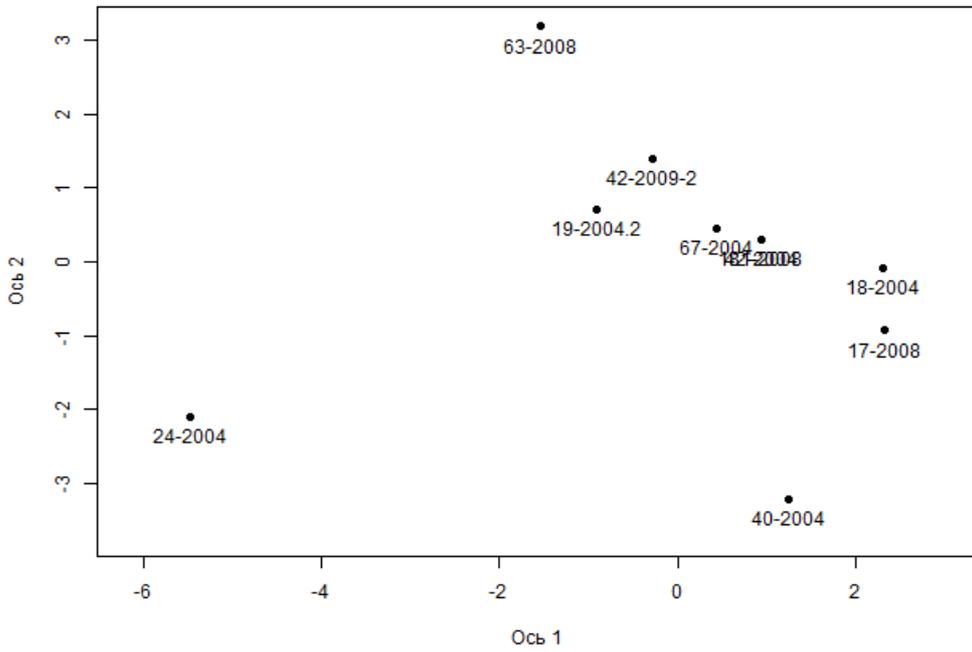


Рис. 8. Результирующая диаграмма PCA ординации

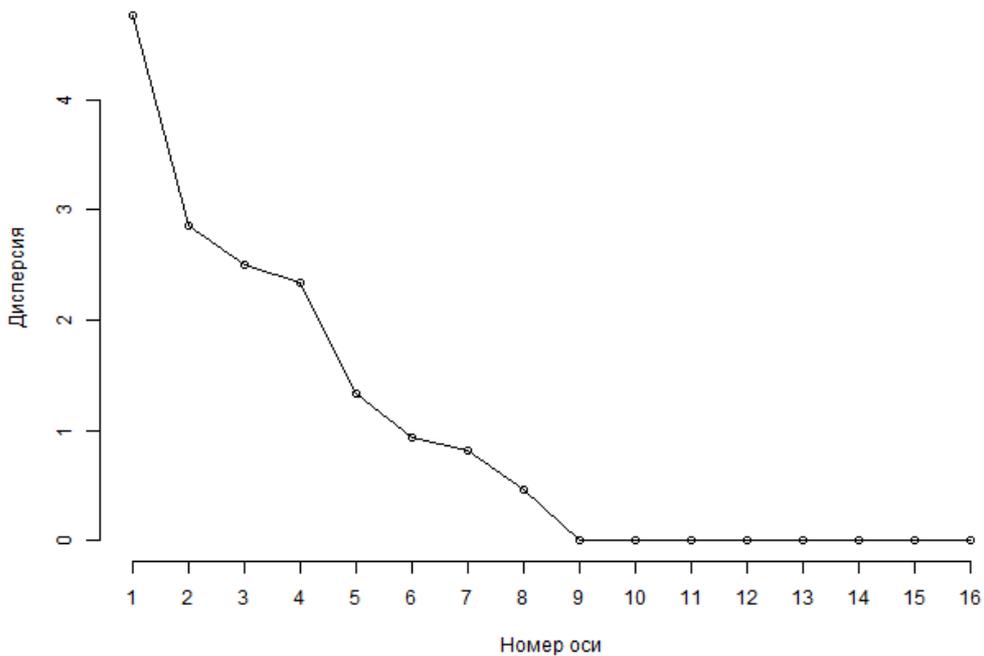


Рис. 9. График дисперсии для каждой из осей PCA ординации (доли объясненной вариации)

*Примечание.* Доля объясненной вариации определяется как соотношение дисперсии для конкретной оси к общей сумме дисперсий по всем осям.

## 7. Пункт меню "Неметрическое многомерное шкалирование..."

В этом пункте реализована ординация методом неметрического многомерного шкалирования (NMS – non metric multidimensional scaling). Аналогично PCA ординации, анализ состоит из двух диалоговых окон: окно выбора данных (см. раздел 5) и окно настроек NMS анализа (рис. 10).

Детальное описание метода можно найти в литературе (см. раздел 14).

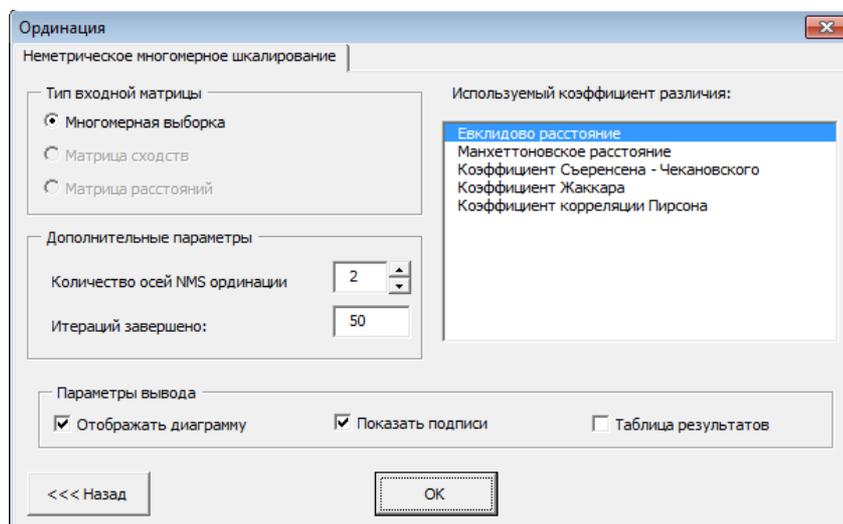


Рис. 10. Диалоговое окно настроек NMS анализа (многомерное неметрическое шкалирование)

Окно настроек NMS ординации:

- Тип входной матрицы:
  - Многомерная выборка – на вход подается стандартная таблица с данными, на основе которой строится матрица различий. При этом для построения используется один из коэффициентов различия.
  - Матрица сходств – на вход подается квадратная симметричная матрица, диагональные элементы которой равны 1, а все недиагональные элементы варьируются от 0 до 1. Для проведения NMS анализа матрица сходств автоматически преобразуется в матрицу различий.
  - Матрица расстояний – на вход подается квадратная симметричная матрица, диагональные элементы которой равны 0, а все недиагональные элементы варьируются от 0 до 1.
- Дополнительные параметры:
  - Количество осей NMS ординации – задает количество осей NMS ординации (минимальное значение 2, максимальное – 6).
  - Итераций завершено – задание фиксированного максимального количества итераций (шагов) алгоритма. Небольшое значение числа итераций ускоряет NMS анализ, однако снижает его точность.

- Используемый коэффициент различия – список коэффициентов сходства/различия при расчете расстояний между объектами. На сегодняшний день в настройке реализованы: Евклидово и манхеттовское расстояния, количественные коэффициенты сходства Жаккара и Сьеренсена-Чекановского, коэффициент корреляции Пирсона.
- Параметры вывода:
  - Отображать диаграмму – Отображение результирующей NMS диаграммы (рис. 11).
  - Показывать подписи – отображать или нет на результирующей диаграмме названия объектов (при большом числе анализируемых объектов > 30 рекомендуется отключать эту опцию).
  - Таблица результатов – вывод результирующей матрицы с координатами объектов (см. раздел 8).

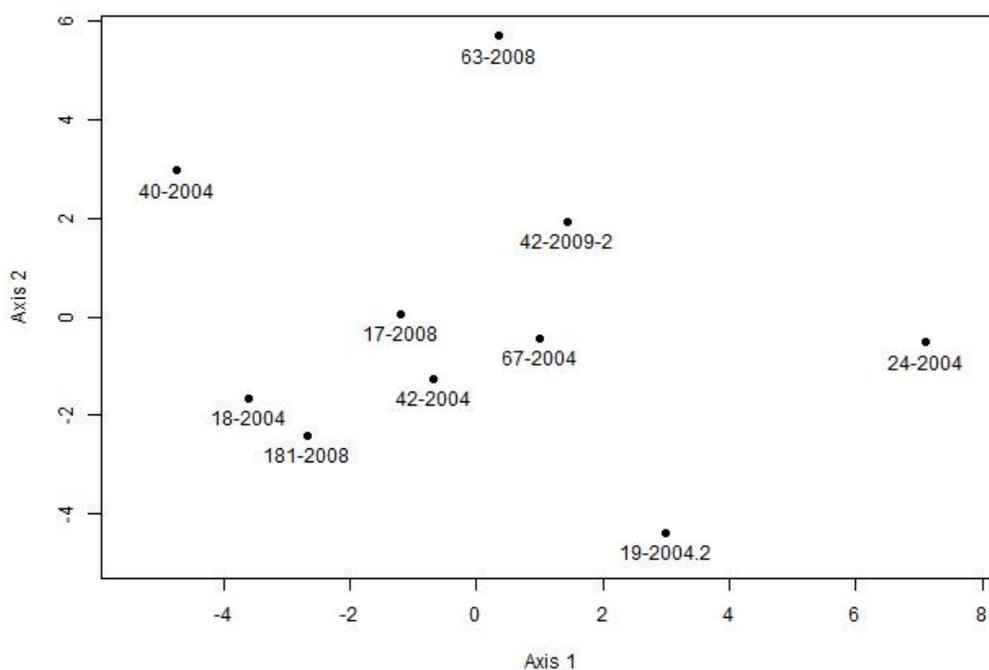


Рис. 11. Результирующая диаграмма NMS ординации

**Примечание.** Ординация методом NMS не работает в случае если в исходной матрице существуют два или более идентичных объекта, т.е. расстояние между которыми по выбранному коэффициенту различия равно 0. В этом случае возникнет ошибка (см. рис. 12).

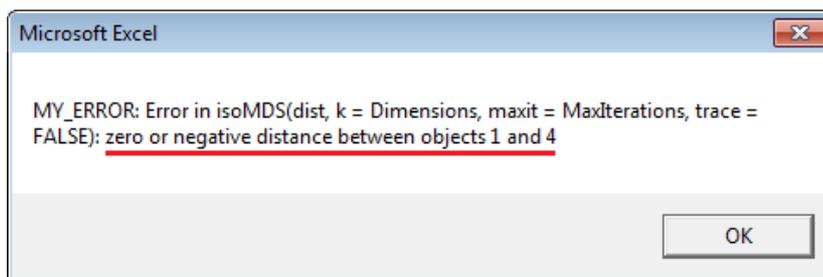


Рис. 12. Пример сообщения об ошибке вызванной нулевым расстоянием между объектами (в рассматриваемом примере объекты № 1 и 4 оказались полностью идентичными)

## 8. Табличное представление результатов ординации

Наряду с графическим представлением результатов PCA или NMS ординации существует возможность их табличного отображения (рис. 13).

	A	B	C
1	Результаты анализа главных компонент		
2	Используется матрица парных корреляций		
3			
4		PCA 1	PCA 2
5	Значения собственных векторов (дисперсия по осям)	4.624689	2.699556
6	Доля объясненной дисперсии от общей суммы (в процентах)	27.20406	15.87974
7			
8	42-2004	0.80932	-0.22458
9	181-2008	1.562143	1.153166
10	67-2004	0.564014	-0.48835
11	18-2004	2.264484	-1.90679
12	19-2004.2	-1.1467	-0.33939
13	63-2008	-1.1776	3.729382
14	17-2008	1.885857	-1.01031
15	42-2009-2	-0.2976	1.662671
16	24-2004	-5.50271	-1.8424
17	40-2004	1.038798	-0.73341
18			
19	Корреляция дополнительных факторов с осями ординации		
20	Фактор 1	-0.48685	-0.19734
21	Фактор 2	-0.30056	-0.62811

Рис. 13. Табличное представление результатов PCA ординации

Результирующая таблица содержит следующие сведения:

Первая строка. Общее название проведенного анализа (на текущий момент анализ главных компонент или NMS ординация).

Вторая строка. Для PCA ординации – какая матрица использовалась при расчетах корреляций или ковариаций. Для NMS ординации – что служило источником данных. Матрица сходства или различия или же использовался коэффициент различия.

Далее, для PCA анализа отображаются собственные значения матрицы и их доля от общей суммы собственных значений (доля объясненной дисперсии).

*Примечание.* NMS ординация строится другим методом, который не использует собственные вектора и собственные значения, поэтому подобной информации для NMS анализа не выводится.

Затем отображается список названий анализируемых объектов и их координаты на ординационной диаграмме.

Последний пункт – "Корреляция дополнительных факторов с осями ординации". Подробнее см. раздел 10.

## 9. Пункт меню "Кластерный анализ..."

Здесь реализованы методы иерархического кластерного анализа. Аналогично PCA и NMS кластерный анализ состоит из двух диалоговых окон: выбор элементов для анализа (раздел 5), и окно настроек кластерного анализа (рис. 13).

Детальное описание метода можно найти в литературе (см. раздел 14).

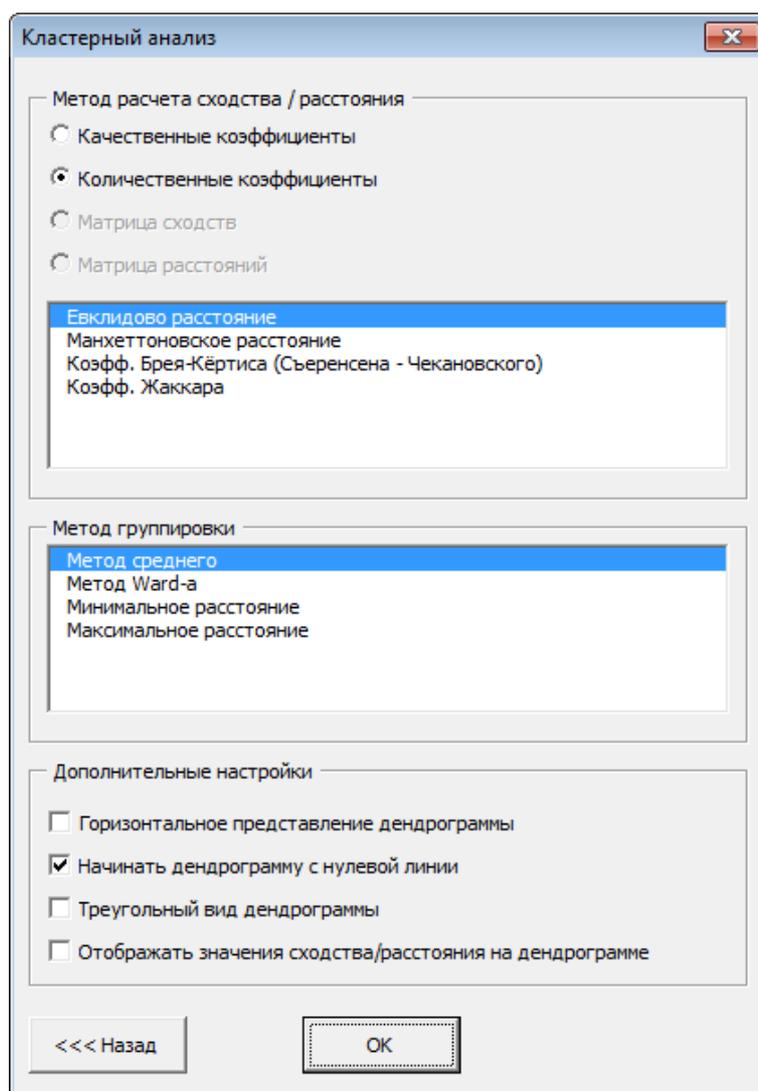


Рис. 13. Диалоговое окно настроек кластерного анализа

Настройки для кластерного анализа:

- Метод расчета сходства/расстояния:
  - Качественные коэффициенты – используются коэффициенты которые учитывают только присутствие или отсутствие какого либо признака. Реализованы коэффициенты Жаккара и Сьеренсена-Чекановского.
  - Количественные коэффициенты – коэффициенты которые учитывают количественные характеристики признаков. Реализованы: Евклидово и манхеттоновское расстояния, количественные коэффициенты сходства Жаккара и Сьеренсена-Чекановского.
  - Матрица сходств – на вход подается квадратная симметричная матрица, диагональные элементы которой равны 1, а все недиагональные элементы варьируются от 0 до 1. Для проведения кластерного анализа матрица сходств автоматически преобразуется в матрицу расстояний.
  - Матрица расстояний – на вход подается квадратная симметричная матрица, диагональные элементы которой равны 0, а все недиагональные элементы варьируются от 0 до 1.
  
- Метод группировки:
  - Метод среднего (UPGMA – unweighted pair-group method using arithmetic averages, average-linkage clustering)
  - Метод Ward-a (Ward's method)
  - Минимальное расстояние (Single-linkage clustering)
  - Максимальное расстояние (Complete-linkage clustering)

*Примечание.* Рекомендуется использовать метод среднего или, для более четкого выделения кластеров, метод Ward-a.

- Дополнительные настройки. Настройки влияют исключительно на внешний вид результирующей дендрограммы (см. рис. 14):
  - Горизонтальное представление дендрограммы
  - Начинать дендрограмму с нулевой линии
  - Треугольный вид дендрограммы
  - Отображать значения сходства/расстояния на дендрограмме – пункт указывает, отображать или нет полученные для каждого элемента дендрограммы (каждого объединения) коэффициенты сходства/расстояния. Значения отображаются в узлах дендрограммы.

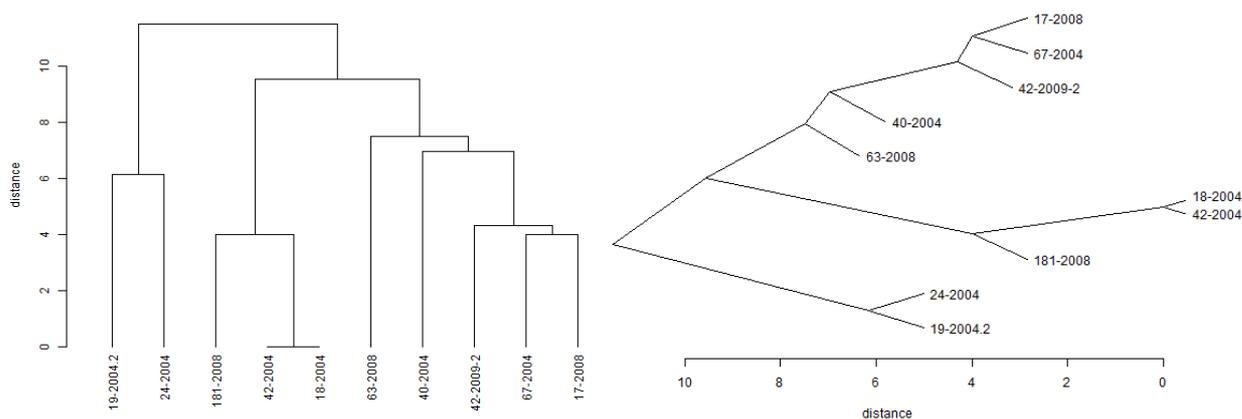


Рис. 14. Различие внешнего вида результирующих кластерных дендрограмм для одного и того же набора данных с полностью идентичными способами анализа. Слева – вертикальное отображение, представление в виде прямоугольников, элементы начинаются с 0. Справа – горизонтальное отображение, представление в виде треугольников, подвешенные элементы

## 10. Наложение дополнительных факторов на дендрограмму или ординационную диаграмму

На результаты кластерного или ординационного анализа можно наложить дополнительные факторы (рис. 15). Реализовано два способа наложения дополнительных факторов: в виде раскраски объектов (используется при кластерном анализе и ординации) (рис. 18, 19), либо в виде векторов отражающих корреляцию между дополнительными факторами и осями ординации (используется только для PCA и NMS ординации) (рис. 19).

*Примечание.* Раскраска позволяет отобразить один дополнительный фактор, наложение векторов – неограниченное количество дополнительных факторов.

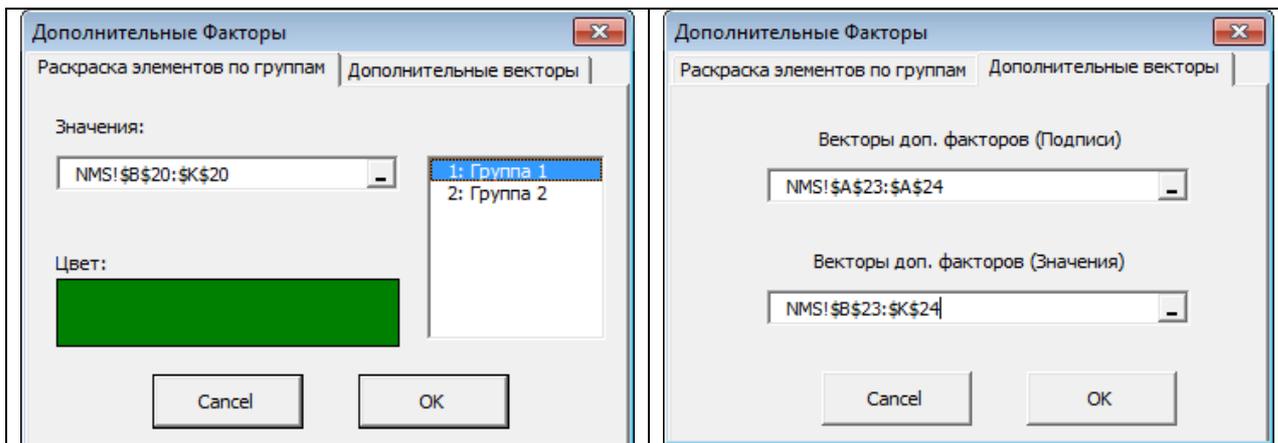


Рис. 15. Диалоговое окно "Дополнительные факторы". Слева – диалоговое окно задания раскраски объектов, справа – окно наложения дополнительных данных в виде векторов

В качестве входных данных для раскраски вводится одна строка/столбец с числовыми или символьными обозначениями цветов (рис. 15). Цвета для разных групп задаются автоматически, однако пользователь может их изменить, для этого надо нажать мышкой на прямоугольнике "Цвет".

Следует обратить внимание, что количество значений в поле "раскраска" должно совпадать с количеством анализируемых объектов. Кроме того, следует учитывать, что выбрано в качестве анализируемых объектов строки или столбцы. Если названия объектов находятся в верхней строке, то номера для раскраски также должны находиться в строке (рис. 16). Если названия объектов – левая колонка, то данные для раскраски также должны быть расположены в виде колонки (рис. 17).

*Примечание.* Количество цветов (групп) ограничено 15.

	A	B	C	D	E	F	G	H	I	J	K
1	Species	42-2004	181-2008	67-2004	18-2004	19-2004.2	63-2008	17-2008	42-2009-2	24-2004	40-2004
2	Abies sibirica	1	0	3	3	3	2	3	4	3	1
3	Abietinella abietina	1	0	0	0	0	1	0	0	3	0
4	Aconitum septentri	2	0	0	5	0	0	2	0	0	0
5	Actaea erythrocar	0	0	0	0	0	0	1	0	2	5
6	Adoxa moschatelli	1	1	2	0	5	0	0	0	5	0
7	Agrostis tenuis	0	0	0	0	0	5	0	2	0	0
8	Alchemilla sp.	0	0	1	1	0	0	0	0	0	0
9	Alnus incana	0	2	0	0	0	0	0	0	0	1
10	Anemonastrum biarmie	0	0	0	0	2	2	0	0	2	0
11	Angelica archangelica	0	0	0	2	0	0	2	0	0	1
12	Angelica sylvestris	0	2	1	1	0	2	0	0	2	0
13	Antennaria dioica	0	0	0	0	0	0	0	1	2	1
14	Anthoxantum alpinum	1	0	0	0	3	0	0	0	0	0
15	Asplenium viride	0	0	0	0	0	2	0	2	0	0
16	Aster alpinus	0	0	0	0	0	0	0	0	2	0
17	Betula nana	0	2	0	0	0	0	1	0	0	0
18											
19		Раскраска объектов									
20	Группы	1	2	2	1	2	2	1	2	2	1
21											
22		Дополнительные факторы									
23	Фактор 1	0.4	0.3	0.3	0.4	1.2	0.6	0.9	0.3	0.9	0.2
24	Фактор 2	5	3	9	7	8	3	6	5	12	15

Рис. 16. Расположение названий объектов и дополнительных факторов при выборе пользователем "Использовать колонки"

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Species	42-2004	181-2008	67-2004	18-2004	19-2004.2	63-2008	17-2008	42-2009.2	24-2004	40-2004	Группы	Фактор 1	Фактор 2		
2	Abies sibirica	1	1	3	3	3	2	3	4	5	1	1	0.4	5		
3	Abietinella abietina	1	0	0	0	0	1	0	0	3	0	2	0.3	3		
4	Aconitum septentrionale	2	0	0	5	0	0	2	0	0	0	3	0.3	9		
5	Actaea erythrocarpa	0	0	0	0	0	0	1	0	2	5	3	0.4	7		
6	Adoxa moschatellina	1	1	2	0	5	0	0	0	5	0	2	1.2	8		
7	Agrostis tenuis	0	0	0	0	0	5	0	2	0	0	1	0.6	3		
8	Alchemilla sp.	0	0	1	1	0	0	0	0	0	0	1	0.9	6		
9	Alnus incana	0	2	0	0	0	0	0	0	0	1	2	0.3	5		
10	Anemonastrum biarmense	0	0	0	0	2	2	0	0	2	0	3	0.9	12		
11	Angelica archangelica	0	0	0	2	0	0	2	0	0	1	2	0.2	15		
12	Angelica sylvestris	0	2	1	1	0	2	0	0	2	0	2	0.4	18		
13	Antennaria dioica	0	0	0	0	0	0	0	1	2	1	3	0.8	20		
14	Anthoxanthum alpinum	1	0	0	0	3	0	0	0	0	0	3	1.2	8		
15	Asplenium viride	0	0	0	0	0	2	0	2	0	0	1	2.1	5		
16	Aster alpinus	0	0	0	0	0	0	0	0	2	0	1	0.9	14		
17	Betula nana	0	2	0	0	0	0	1	0	0	0	1	0.7	12		

Рис. 17. Расположение объектов, их названий и дополнительных факторов при выборе пользователем "Использовать строки"

Для наложения на ординационную диаграмму дополнительных факторов в виде векторов требуется задать две области: область подписей и область значений (рис. 15). Область подписей содержит названия векторов, которые будут отображаться на диаграмме. Значения – сами числовые данные дополнительных векторов.

*Примечание.* Здесь следует обращать внимание на то, что является объектами для анализа строки или колонки (рис. 16, 17). В зависимости от этого меняются области для ввода подписей и значений дополнительных факторов.

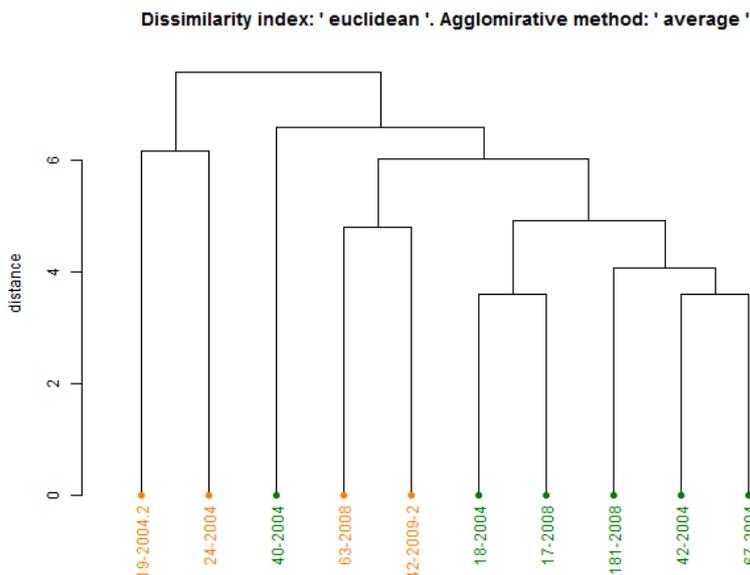


Рис. 18. Пример использования окраски элементов для кластерного анализа

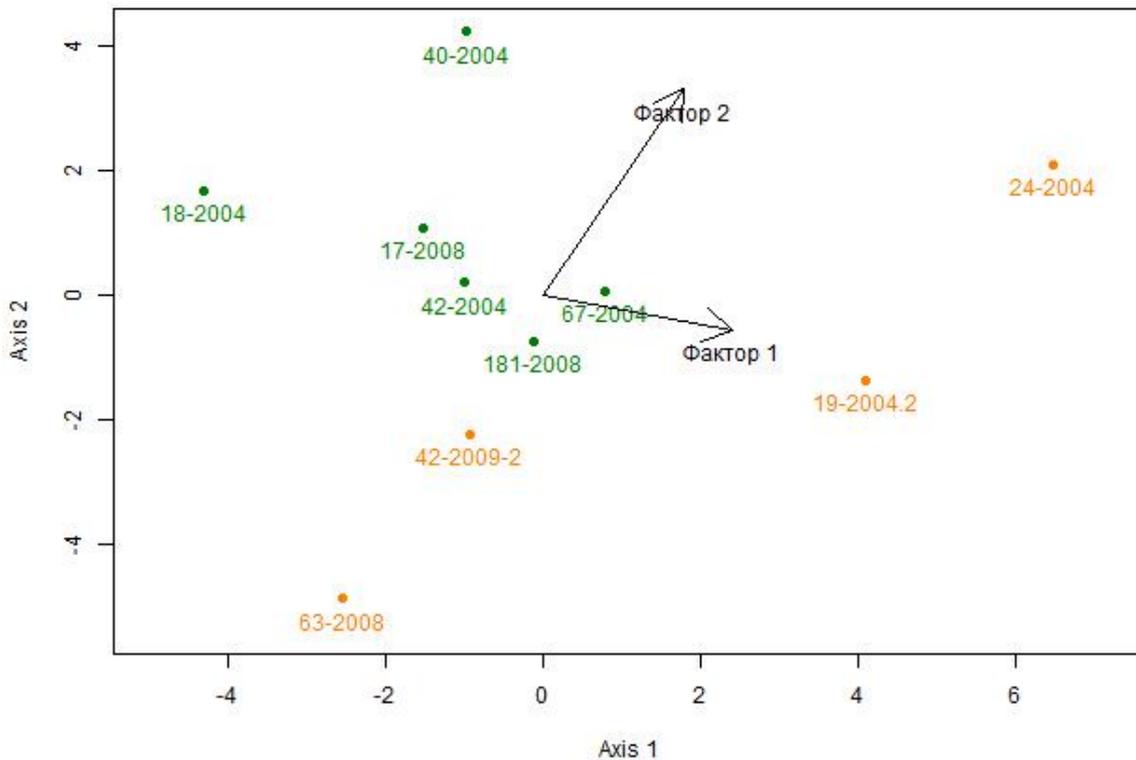


Рис. 19. Пример раскраски объектов и наложения дополнительных факторов на результирующую ординационную диаграмму

Коэффициенты корреляции между дополнительными факторами и осями ординации отображаются в результирующей таблице.

## 11. Пункт меню "Теплокарты..."

Построение "теплокарт" (рис. 20) скорее является способом визуализации данных, а не специализированным методом их анализа. Его суть состоит в том, что строится прямоугольная сетка, в клетках которой цветом отображаются числовые значения. Минимальным значениям соответствует один цвет, максимальным – другой. Промежуточные значения окрашиваются по цветам градиента.

Мы использовали этот подход для визуализации температурных данных (рис. 21).

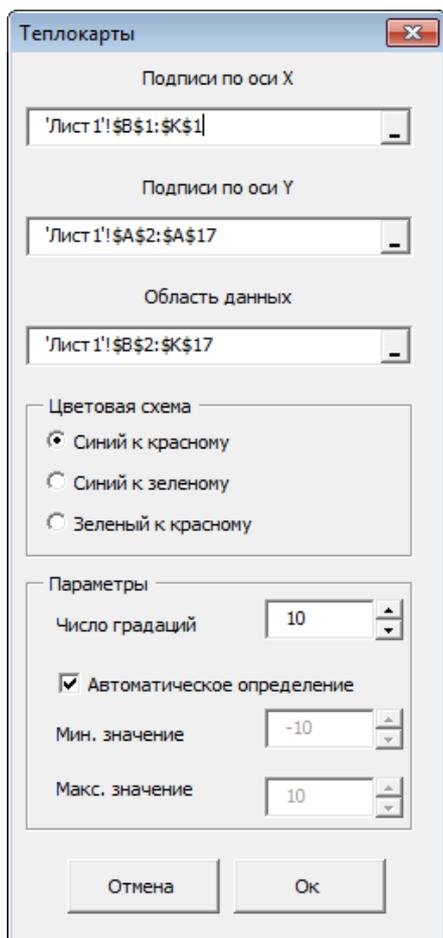


Рис. 20. Диалоговое окно настроек для построения "Теплокарт"

Окно настроек для "Теплокарт":

- Подписи по оси X, подписи по оси Y – области содержащие названия осей X и Y на результирующей диаграмме.

*Примечание.* В качестве подписей по оси X необходимо выбрать одну строку, по оси Y – одну колонку

- Область данных – область на листе Excel которая будет использоваться для построения диаграммы.

*Примечание.* В области данных могут находиться только числовые значения.

- Цветовая схема (Синий к красному, синий к зеленому, зеленый к красному) – задание цветов соответствующих минимальному и максимальному значениям.
- Параметры:
  - Число градаций – число промежуточных градаций цвета от минимального до максимального (увеличение числа градаций делает переходы между цветами более плавными).
  - Автоматическое определение – при выбранном пункте автоматически в выделенной области данных определяется минимальное и максимальное значения. При отключении этого пункта мини-

мальное и максимальное значения требуется задать самостоятельно.

- Мин. значение, макс. значение – ручное задание минимального и максимального значения для построения теплокарты.

*Примечание.* Если какое-либо значение в области данных выпадет за диапазон от минимального значения к максимальному, то соответствующая точка будет окрашена в белый цвет.

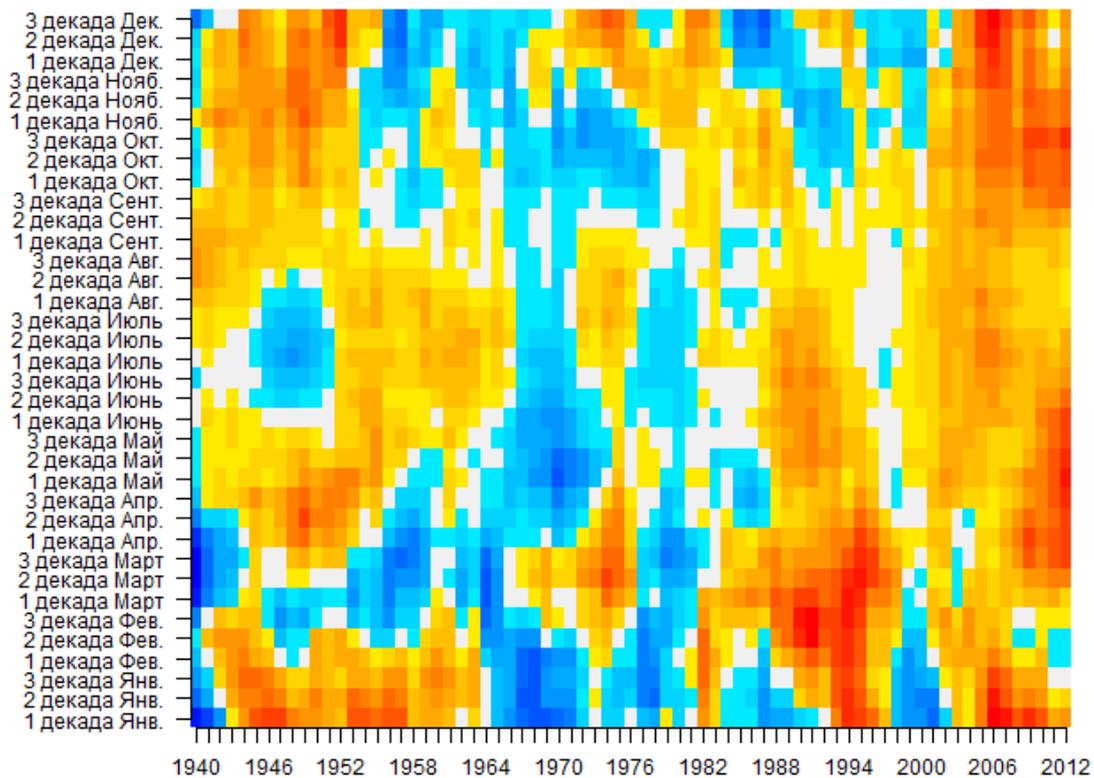


Рис. 21. Результат использования метода "Теплокарты" для визуализации температурных отклонений от средних значений для европейской части России

## 12. Список дополнительных пакетов и их установка

Для полноценной работы надстройки ExStatR необходимо в программу R установить некоторые дополнительные пакеты.

*Примечание*, для установки необходим доступ к сети Интернет.

Установка дополнений:

- Запустите программу R.
- В появившемся окне вставьте команды:  
`install.packages("XML")`  
`install.packages("vegan")`  
`install.packages("Hmisc")`  
`install.packages("colorRamps")`  
`install.packages("dendextend")`
- нажмите enter.
- Через некоторое время возникнет окно: "выберете зеркало CRAN (CRAN mirror)" – тот адрес, откуда будет проводиться скачивание пакетов. Выберете Russia (выбор не принципиален, можно использовать и любой другой адрес).
- После этого необходимо подождать пока компьютер скачает и установит указанные пакеты (может занять до 5-10 минут, в зависимости от скорости Интернет соединения и скорости компьютера).

Общий объем пакетов – около 75 мб.

### 13. Список R файлов (скриптов) надстройки ExStatR

Название файла	Описание
CalculationRoutines.R	Функция запуска одного R скрипта из другого
ClusterAnalysis.R	Скрипт реализации алгоритмов кластерного анализа
ExportRoutines.R	Набор функций для записи результатов анализа в XML файл
HeatMap.R	Скрипт построения теплокарт
InitializationRoutines.R	Начальный скрипт инициализации (определяет пути к входному и выходному XML файлам и их наличие в указанных папках)
MDS2.R	Скрипт реализации алгоритма неметрического многомерного шкалирования
PCA.R	Скрипт реализации метода главных компонент (общий файл)
pcaRoutines.R	Вычислительная часть метода главных компонент
xmlOutputError.R	Скрипт вывода информации об ошибках выполнения в среде R
xmlParsingRoutines.R	Набор функций извлечения данных из XML файлов

## 14. Литература

Кластерный анализ:

- Jongman R.H.G., Ter Braak C.J.F., Van Tongeren O.F.R. Data analysis in community and landscape ecology. Wageningen, 1987. 299 p.
- Ward J.H. Hierarchical grouping to optimize an objective function // Journal of the American statistical association, 1963. Vol. 58, № 301. P. 236-244.
- Песенко Ю.А. Принципы и методы количественного анализа в фаунистических исследованиях. М.: Наука, 1982. 287 с.
- Ту Дж., Гонсалес Р. Принципы распознавания образов. М., 1978. 416 с.
- Уиллиамс У.Т., Ланс Дж.Н. Методы иерархической классификации // Статистические методы для ЭВМ / Под ред. К. Энслейна, Э. Рэлстона, Г.С. Уилфа. 1986. С. 269-301.
- Факторный, дискриминантный и кластерный анализ. / Под ред. И.С. Енюкова. М., 1989. 216 с.
- Шитиков В.К., Розенберг Г.С., Зинченко Т.Д. Количественная гидроэкология: методы системной идентификации. Тольятти, 2003. 463 с. <http://www.ievbras.ru/ecostat/Kiril/Library/Book1/Content0/Content0.htm>

Ординация методом главных компонент:

- Андрукович П. Ф. Применение метода главных компонент в практических исследованиях. М: Издательство московского университета, 1973. 124 с.
- Введение в многомерный статистический анализ / пер. с англ. / Т. Андерсон. / М: Физматгиз, 1963. 500 с.
- Терехина А.Ю. Анализ данных методами многомерного шкалирования. М: Наука, 1986. 168 с.

Ординация методом многомерного неметрического шкалирования:

- Cox T. F., Cox M. A.A. Multidimensional scaling (2<sup>nd</sup> edition). Chapman and Hall, 2001. 294 p.
- Kruskal, J. B. Nonmetric multidimensional scaling: A numerical method // Psychometrika, 1964. Vol. 29(2). P. 115-130
- Shepard, R. N. Analysis of proximities: Multidimensional scaling with an unknown distance function I & II // Psychometrika, 1962. Vol. 27. P. 125-140, 219-246.
- Многомерное шкалирование / пер. с англ. / М. Дэйвисон. / М: Финансы и статистика, 1988. 245 с.
- Статистические методы для ЭВМ / пер. с англ. / К. Энслейн, Э. Рэлстон, Г.С. Уилф / М: Наука, 1986. 464 с.
- Толстова Ю.Н. Основы многомерного шкалирования: учебное пособие. М: КДУ, 2006. 160 с.