

**МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ  
ПО ИСПОЛЬЗОВАНИЮ ГРАФ-АНАЛИЗА  
В ИССЛЕДОВАНИЯХ СИСТЕМ,  
СОСТОЯЩИХ ИЗ БИОТИЧЕСКИХ И  
АБИОТИЧЕСКИХ КОМПОНЕНТОВ**

(2-е издание)



Санкт-Петербург-Пушкин

2006

МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ПО ИСПОЛЬЗОВАНИЮ ГРАФ-АНАЛИЗА В ИССЛЕДОВАНИЯХ СИСТЕМ, СОСТОЯЩИХ ИЗ БИОТИЧЕСКИХ И АБИОТИЧЕСКИХ КОМПОНЕНТОВ. Воробьев Н.И., Свиридова О.В., Кутузова Р.С. (2-е издание, переработанное и дополненное) - СПб.: ГНУ ВНИИСХМ. 2006, 59 с.

Целью методических указаний является внедрение специального метода обработки многофакторных данных в исследованиях биологических систем. В опытах с десятками вариантов и сотнями характеристик граф-анализ предоставляет собой единственное средство получения информации о структуре и закономерностях функционирования биосистем. Этим методом можно обрабатывать одновременно как количественные, так и качественные экспериментальные данные. В помощь исследователю предлагаются вычислительные программы Clusters3.xls, Graph3.xls и Factor3.xls (вычислительная среда - Microsoft Excel, версии 2000-2003), которые выполняют все необходимые расчеты и предоставляют пользователю соответствующую информацию. Метод включает в себя стандартные процедуры кластерного, корреляционного и факторного анализов. Выходная продукция метода: дерево группировки вариантов; граф максимальных коэффициентов корреляции; направленный граф биотрансформации веществ и математическая модель дивергенции компонентов. Эти материалы дают объективную информацию о структуре биосистем и факторах, влияющих на биосистемные процессы.

Метод разработан сотрудниками ГНУ ВНИИСХМ: кандидатом технических наук Н.И.Воробьевым, кандидатами биологических наук О.В.Свиридовой и Р.С.Кутузовой. Основная идея метода была сформулирована в 1994 году при участии доктора биологических наук Ю.М.Возняковской. Методика разработана по материалам исследований, проводимых в 1993-2006 годах, и в рамках работ, выполняемых по гранту РФФИ 06-04-48800-а.

Рецензент доктор биологических наук, профессор Ю.В.Круглов.

Методические указания рассмотрены ученым советом ГНУ ВНИИ сельскохозяйственной микробиологии протокол № 1 от 08.02.2005 и рекомендованы к изданию.

© ГНУ ВНИИСХМ, 2006  
© Н.И.Воробьев, 2006, vorobyov@arriam.spb.ru  
© О.В.Свиридова, 2006  
© Р.С.Кутузова, 2006

## ВВЕДЕНИЕ

Современные исследования биосистем отличаются масштабом и длительностью проведения экспериментов. Информация, накопившаяся к настоящему времени, достигла таких объемов, что обработка данных без помощи компьютера становится практически невозможной. Проблема усугубляется еще и тем, что полезная информация, как правило, скрыта не только случайными факторами, но и сложными биосистемными взаимоисключающими процессами. При создании данной методики приходилось учитывать многомерность биосистем, использовать теоретические знания о биотических компонентах системы, а также учитывать то, что численности микроорганизмов и концентрации веществ могут за время эксперимента изменяться на несколько порядков. Все это предъявляет повышенные требования к разрабатываемой компьютерной системе обработки данных и требует ее теоретическое обоснование.

Предлагаемая методика (далее «Граф-анализ») является первым шагом на пути решения возникшей проблемы. Этим методом проводится систематизация данных и представление их в следующих формах:

- 1) дерево группировки вариантов (дендрограмма);
- 2) граф максимальных коэффициентов корреляций признаков;
- 3) направленный граф биотрансформации;
- 4) математическая модель дивергенции компонентов биосистемы.

Эти материалы позволяют исследователю сконцентрировать свое внимание на главных факторах, влияющих на биосистему, и получить представление о схеме биосистемных взаимодействий. На этапе биологической интерпретации дендрограммы и двух графов обнаруживаются противоречия, которые, с одной стороны, тестируют научные знания о биосистемах на их полноту, а с другой - формируют требования к последующим экспериментам.

В граф-анализе используются усредненные величины измеряемых характеристик, удовлетворяющих статистическому условию: отношение межвариантной дисперсии к случайной составляющей дисперсии превышает стандартную величину критерия Фишера. Результаты обработки данных, не удовлетворяющих этому условию, следует рассматривать как предварительные. Применение граф-анализа в этих случаях часто бывает все-таки оправдано важностью и полезностью получаемой информации.

В 1-м разделе рекомендаций вводятся базовые понятия граф-анализа и поясняются основные термины. Во 2-м разделе дается представление об основных выходных формах граф-анализа на примере опубликованных данных биологического эксперимента

(Р.С.Кутузова, Агрохимия, 2001). 3-й раздел является кратким руководством по использованию компьютерных программ: Clusters3.xls, Graph3.xls, Factor3.xls.

## 1. БАЗОВЫЕ ПОНЯТИЯ И ТЕРМИНЫ ГРАФ-АНАЛИЗА.

### 1.1. Прямоугольная матрица данных.

*Прямоугольной матрицей* данных называется группа ячеек, находящихся на листе <Данные> в программах Clusters3.xls, Graph3.xls и Factor3.xls (см. рис. 1) и содержащая исходную информацию о проведенных наблюдениях за биосистемой.

В ячейках электронной таблицы данные располагаются в определенном порядке.

В строках - данные, имеющие одну и ту же единицу измерения. Эти данные относятся к одной и той же измеряемой характеристике. Далее эту группу значений будем называть строка-признак, или просто - *признак*.

В столбцах - данные, содержащие величины признаков, которые относятся к одному варианту опыта. Эту группу значений будем называть столбец-вариант, или просто - *вариант*.

Каждый признак и вариант имеют порядковый номер: 2-й вариант, 3-й признак. На рис. 1 представлена таблица данных, содержащая 12 вариантов и 13 признаков.

Признаки могут быть как количественные (рис. 1, 1-11 признаки), так и качественные (рис. 1, 12-13 признаки). Численность микроорганизмов, количество гумуса, количество ферментов – все это *количественные признаки*. К *качественным признакам* относятся, например, вид микроорганизмов, растений, тип почв, факт внесения/невнесения удобрений.

Для включения качественного признака в матрицу данных рекомендуется воспользоваться системой, основанной на баллах. Возьмем, например, признак, характеризующий фактом внесения NPK (рис. 1, 12-й признак). В 3,4,7,8,11,12 вариантах опыта NPK вносили в почву. Поэтому в соответствующих ячейках строки-признака ставим 1 балл, а в остальных – 0 баллов. Аналогичным образом введены данные и для 13 признака. Таким образом, качественные признаки трансформируются в количественные признаки и обрабатываются в едином стиле совместно с количественными признаками.

Признаки, характеризующие биотические компоненты, называются *биотическими признаками*, а характеризующие абиотические компоненты – *абиотическими признаками*.

5

A	B													C	D	E	F	G	H	I	J	K	L	M	N	
1	Признаки													Варианты												
2	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12		
3	1	Аммонификаторы	14,5	27,8	13,2	14	10,9	15,4	15,2	16,6	17,9	13,6	22,7	23,9												
4	2	Аммилолитические	11,2	19,1	12	8,1	8,8	10	6,9	11,4	12,2	13,6	23,3	24,2												
5	3	Нитрификаторы 1 фазы	28,4	45,8	47,9	32,2	30,5	23,3	16,7	30	6,8	36	10,6	25												
6	4	Нитрификаторы 2 фазы	26	26	33,9	45,5	11	20,9	35,6	11,6	18,1	23,5	56,2	45,5												
7	5	Гуминдеструкторы	12,5	14,9	11,3	10,3	11	14,5	16,4	12,6	10,7	13,4	21,9	23,4												
8	6	Денитрификаторы	0,2	0,1	0,1	0,2	0,1	0,1	1,1	0,3	0,5	1,2	1,1	1,6												
9	7	Микромицеты	34	55	28	24	31	55	23	14	44	24	26	40												
10	8	Углерод гумуса, %	1,05	1,13	1,02	1,27	1,03	1,05	0,98	1,15	1,15	1,06	0,95	1,17												
11	9	Доступный углерод, %	0,131	0,141	0,211	0,314	0,14	0,12	0,294	0,151	0,126	0,142	0,12	0,151												
12	10	N-NH <sub>4</sub>	1,2	2,2	1,5	1,8	0,7	1,6	2,3	1,3	1	1,1	2,6	1,5												
13	11	N-NO <sub>3</sub>	0	2,4	2,1	2,8	0	0	5,5	2,5	0	0	3,9	1,4												
14	12	Внесение NPK	0	0	1	1	0	0	1	1	0	0	1	1												
15	13	Внесение навоза	0	1	0	1	0	1	0	1	0	1	0	1												
16																										

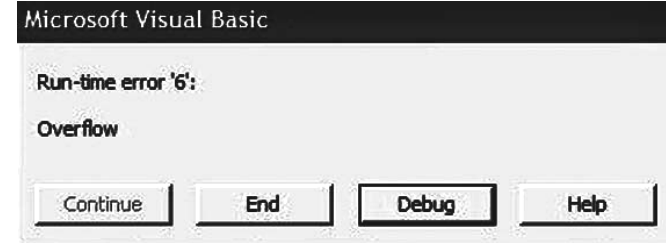
Рис. 1. Матрица данных, расположенная на листе <Данные> в программах Clusters3.xls, Graph3.xls и Factor3.xls.

### Основные правила по заполнению матрицы данных

Для повышения достоверности метода и устранения ошибок расчетов рекомендуется руководствоваться следующими правилами:

1. Нет необходимости вводить числа с большим числом ведущих или следующих нулей (например, 45600000,0 или 0,00000231). Рекомендуется выносить соответствующий множитель в размерность признака за пределы матрицы данных в описательную колонку признаков ( $4,56 \times 10^7$  – коэффициент  $10^7$  в размерность признака, а число 4,56 в ячейку матрицы данных;  $2,31 \times 10^{-6}$  – коэффициент  $10^{-6}$  в размерность признака, а число 2,31 в ячейку матрицы данных).

2. Строка-признак не должна состоять из одинаковых чисел. Для таких данных будут выдаваться сообщения об ошибке в программах Clusters3.xls, Graph3.xls и Factor3.xls.



Иными словами, из матрицы данных должны быть исключены строки-признаки с нулевой межвариантной дисперсией. Кроме того, данные признака должны удовлетворять условию: межвариантная дисперсия признака должна превышать случайную дисперсию в количестве раз, определяемое стандартным критерием Фишера.

3. В матрице данных не должно быть линейно зависимых признаков, то есть не следует в матрицу вводить данные, которые получены путем математического пересчета одних признаков в другие. Например, количество гумуса в почвенных образцах характеризуется тремя составляющими: гумин (негидролизующий гумусовый остаток); гуминовые кислоты и фульвокислоты. Если в матрицу данных введены признаки: общее количество гумуса, гуминовых кислот и фульвокислот, то не следует вводить признак, связанный с количеством гумина, так как этот признак вычисляется путем вычитания доли гуминовых кислот и фульвокислот из общего количества гумуса. Индикатором присутствия в данных таких зависимых признаков является равенство 1 или -1 коэффициента корреляции между ними. Если оставить связанные признаки в таблице данных, то снизится достоверность граф-анализа. В подобных случаях рекомендуется исключать один из зависимых признаков из матрицы данных.

4. В матрице данных также не должно быть одинаковых столбцов-вариантов (с одинаковыми значениями признаков). Один из повто-

ряющихся столбцов следует исключать из матрицы данных. Иначе информация о связях в объекте будет искажена и недостоверна.

5. Минимальное число вариантов рекомендуется выбирать, исходя из условия: число вариантов должно в 1,5-2 раза превышать число биотических признаков.

6. Число биотических признаков, связанных с биотическими компонентами биосистем, должно превышать число абиотических признаков, связанных с абиотическими компонентами.

7. Обратите внимание также на то, что при вводе десятичных чисел в Microsoft Excel вместо точки используется только ЗАПЯТАЯ.

### 1.2. Норма признака. Кластер-нормализация матрицы данных.

*Норма признака* используется в кластерном анализе для приведения таблицы данных к безразмерному виду (*кластер-нормализация данных*). Перед проведением этой операции вычисляется норма для каждого признака. Норма признака имеет ту же размерность, что и сам признак. Поэтому, после деления значений строки-признака на соответствующую норму эта строка будет содержать безразмерные величины.

*Нормой признака* называется разность максимального и минимального значений в строке-признаке. После деления значений признаков на соответствующую норму (*кластер-нормализация данных*) каждая строка признака будет содержать величины, обладающие следующими свойствами:

- 1) кластер-нормализованные значения признака являются безразмерными величинами;
- 2) разность любой пары значений из одной строки по модулю равна или меньше единицы.

Норма признака - всегда положительная величина. Математически вычисление нормы записывается следующим образом:

$$N_i = P_{i.MAX} - P_{i.MIN}, \quad (1)$$

где  $N_i$  – норма  $i$ -того признака;

$P_{i.MAX}$ ,  $P_{i.MIN}$  – максимальное и минимальное значение  $i$ -того признака (в  $i$ -той строке матрицы данных).

Например, для 1-го признака (рис. 1, численность аммонификаторов):

$$P_{1.MAX} = 27,8 \text{ (2-й вариант)}$$

$$P_{1.MIN} = 10,9 \text{ (5-й вариант)}$$

$$N_1 = 27,8 - 10,9 = 16,9 \text{ (норма 1-го признака).}$$

В табл.1 приведены нормы для всех признаков исходных данных (рис. 1). На рис. 2 представлена *кластер-нормализованная матрица*

исходных данных.

Таблица 1. Обобщенные характеристики признаков.

№	Признаки	Норма	Среднее	Девията
1	Аммонификаторы	16,90	17,14	16,78
2	Аммилолитические	17,30	13,40	19,04
3	Нитрификаторы 1 фазы	41,10	27,68	41,69
4	Нитрификаторы 2 фазы	45,20	29,48	46,98
5	Гуминдеструкторы	13,10	14,41	14,18
6	Денитрификаторы	1,50	0,55	1,80
7	Микромицеты	41,00	33,17	42,89
8	Углерод гумуса, %	0,32	1,08	0,302
9	Доступный углерод, %	0,194	0,17	0,22
10	N-NH <sub>4</sub>	1,90	1,57	1,89
11	N-NO <sub>3</sub>	5,50	1,72	6,03
12	Внесение NPK	1,00	0,50	1,73
13	Внесение навоза	1,00	0,50	1,73

### 1.3. Нормированное Евклидово расстояние – мера близости вариантов.

В кластерном анализе в качестве меры близости вариантов используется Евклидово расстояние, которое дает количественное представление о близости вариантов в пространстве признаков. Значение нормированного Евклидова расстояния для любой пары вариантов всегда больше 0 и меньше или равно 1.

Вычисляется *нормированное Евклидово расстояние* ( $E_{K1.K2}$ ) по следующей математической формуле:

$$E_{K1.K2} = \sqrt{\frac{1}{M_p} \cdot \sum_{i=1}^{M_p} (k_{i.K1} - k_{i.K2})^2}, \quad (2)$$

где  $k_{i.K1}$ ,  $k_{i.K2}$  - значения  $i$ -того признака для  $K1$ -го и  $K2$ -го вариантов соответственно из кластер-нормализованной матрицы данных (рис. 2);  $k_{i,j} = P_{i,j} : N_i$ ;

$M_p$  – число признаков в матрице данных;

$N_i$  – норма  $i$ -того признака (формула 1).

Для примера рассчитаем нормированное Евклидово расстояние между 1-м и 2-м вариантами нормализованной матрицы данных (рис. 2,  $M_p = 13$ ):

$$E_{1,2}^2 \cdot 13 = (0,858 - 1,645)^2 + (0,647 - 1,104)^2 + (0,691 - 1,114)^2 + (0,575 - 0,575)^2 + \\ + (0,954 - 1,137)^2 + (0,133 - 0,067)^2 + (0,829 - 1,341)^2 + (3,281 - 3,531)^2 + \\ + (0,675 - 0,727)^2 + (0,632 - 1,158)^2 + (0 - 0,436)^2 + (0 - 0)^2 + (0 - 1)^2 = 2,840$$

$E_{1,2} = \sqrt{2,840 : 13} = 0,467$  - нормированное Евклидово расстояние между 1-м и 2-м вариантами.

Программа Clusters3.xls вычисляет нормированные Евклидовы расстояния для всех комбинаций пар вариантов и предоставляет пользователю эту информацию на листе <Матрица> в виде *треугольной матрицы нормированных Евклидовых расстояний*.

На рис. 3 отмечены наиболее близко расположенные пары ( $E_{1,5} = 0,146$ , между 1-м и 5-м вариантами) и наиболее удаленные пары ( $E_{5,12} = 0,700$ , между 5-м и 12-м вариантами) вариантов опыта.

Согласно формуле (2), равенство значений всех признаков в сравниваемых вариантах приводит к тому, что Евклидово расстояние будет равно нулю\*.

Различают три диапазона Евклидовых расстояний:

- малое расстояние*  $0,0 < E \leq 0,3$ ;
- среднее расстояние*  $0,3 < E \leq 0,7$ ;
- большое расстояние*  $0,7 < E \leq 1,0$ .

Геометрически Евклидово расстояние связано с длиной прямой, соединяющей координатные точки вариантов в пространстве признаков. Значение Евклидова расстояния определяется путем деления этой длины на корень квадратный из общего числа признаков (размерности пространства признаков) (рис. 4).

Из рис. 4 видно, что варианты в пространстве признаков могут образовывать «сгустки» (*группы*) точек-вариантов, которые охватываются *кластер-сферой* (или *кластер-кругом* для плоскости). Кластер-сфера характеризуется наименьшим объемом, с одной стороны, а с другой – размер ее достаточен, чтобы охватить соответствующие точки-варианты. На рис. 4 представлен такой кластер-круг для 1, 3 и 4 вариантов в пространстве двух признаков. Размер кластер-круга (или кластер-сферы) характеризуется ее *диаметром* ( $\phi_{1,3,4}$ , рис. 4).

\* Согласно правилам заполнения матрицы данных (п.4), такие столбцы-варианты должны отсутствовать.

A	B		C																					
	D												E		F	G	H	I	J	K	L	M	N	
1	Варианты												1	2	3	4	5	6	7	8	9	10	11	12
2	0,858	1,645	0,781	0,828	0,645	0,911	0,899	0,982	1,059	0,805	1,343	1,414												
3	0,647	1,104	0,694	0,468	0,509	0,578	0,399	0,659	0,705	0,786	1,347	1,399												
4	0,691	1,114	1,166	0,783	0,742	0,567	0,382	0,73	0,165	0,876	0,258	0,608												
5	0,575	0,575	0,75	1,007	0,243	0,462	0,785	0,257	0,4	0,52	1,243	1,007												
6	0,954	1,137	0,863	0,786	0,84	1,107	1,252	0,962	0,817	1,023	1,672	1,786												
7	0,133	0,067	0,067	0,133	0,067	0,067	0,733	0,2	0,333	0,8	0,733	1,067												
8	0,829	1,341	0,683	0,585	0,756	1,341	0,561	0,341	1,073	0,585	0,634	0,976												
9	3,281	3,531	3,188	3,969	3,219	3,281	3,063	3,594	3,594	3,313	2,969	3,656												
10	0,675	0,727	1,088	1,619	0,722	0,619	1,515	0,778	0,649	0,732	0,619	0,778												
11	0,632	1,158	0,789	0,947	0,368	0,842	1,211	0,684	0,526	0,579	1,368	0,789												
12	0	0,436	0,382	0,509	0	0	1	0,455	0	0	0,709	0,255												
13	0	0	1	1	0	0	1	1	0	0	1	1												
14	0	1	0	1	0	1	0	1	0	1	0	1												
15	0	1	0	1	0	1	0	1	0	1	0	1												
16																								

Рис. 2. Кластер-нормализованная матрица исходных данных.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1													
2													
3													
4													
5													
6													
7													
8													
9													
10													
11													
12													
13													
14													
15													

Рис. 3. Треугольная матрица нормированных Евклидовых расстояний между вариантами опыта.

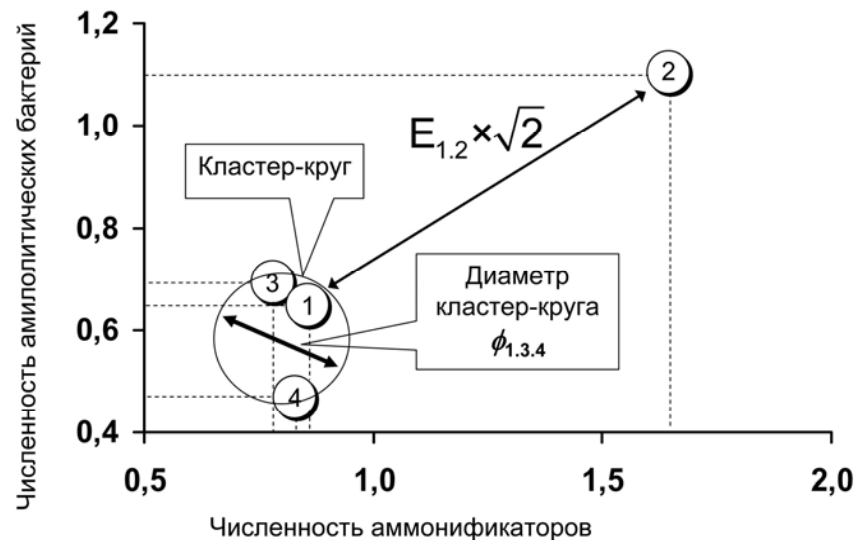


Рис. 4. Расположение вариантов кластер-нормализованной матрицы на координатной плоскости с осями, связанными с численностью аммонификаторов и амилотических бактерий. Цифры в кружках означают номер варианта из таблицы (рис. 2).

#### 1.4. Уровень близости (сходства) вариантов в группе.

Уровень близости (сходства) является величиной, равной разности: единица минус диаметр кластер-сферы, охватывающей данные варианты. Величина уровня близости ( $S_{K1.K2.K3...}$ ) рассчитывается по следующей формуле:

$$S_{K1.K2.K3...} = 1 - \phi_{K1.K2.K3...}, \quad (3)$$

где  $\phi_{K1.K2.K3...}$  – диаметр многомерной кластер-сферы, охватывающей  $K1, K2, K3, \dots$  варианты опыта в пространстве признаков.

Также как и для нормированного Евклидова расстояния, различают три диапазона уровней близости вариантов в группе:

высокий уровень близости  $0,7 < S \leq 1,0$ ;

средний уровень близости  $0,3 < S \leq 0,7$ ;

низкий уровень близости  $0,0 < S \leq 0,3$ .



Фактор-нормализованная матрица данных используется в корреляционном и факторном анализе. Если девиата равна 0 (то есть значения признака во всех вариантах равны между собой), то граф-анализ невозможен.\*

### 1.6. Коэффициент корреляции признаков – мера взаимозависимости признаков.

Парный коэффициент корреляции характеризует степень связанности двух признаков. Чем сильнее связь двух признаков, тем точнее изменения одного признака сопровождаются изменением другого признака. Чем слабее связь признаков, тем неопределеннее изменения одного признака сопровождаются изменением другого, то есть, тем больше доля случайной составляющей в варьировании признаков.

Если обратиться к геометрическому представлению данных фактор-нормализованной матрицы (рис. 5) в пространстве 1-го и 2-го признаков, то получим некоторое поле точек в этом пространстве. Расположение этих точек можно охарактеризовать прямой линией (*линией регрессии*), проходящей через начало координат с углом наклона  $\varphi$ . Наклон линии регрессии определяется условием «наименьших квадратов»: сумма квадратов отклонений вариантов от линии регрессии (см. рис. 6) наименьшая по величине.

Математически это условие записывается следующим образом:

$$Sum_{L1.L2} = \sum_{j=1}^{Mv} (f_{L2,j} - r_{L1.L2} \cdot f_{L1,j})^2 = \min, \quad (5)$$

где  $f_{L1,j}$ ,  $f_{L2,j}$  – значения L1-го и L2-го признаков соответственно для j-го варианта из фактор-нормализованной матрицы данных (рис.

$$5); f_{ij} = \frac{P_{ij} - q_i}{Dv_i};$$

$P_{ij}$  – значение i-го признака в j-том варианте опыта;

$q_i$  – среднее i-го признака;

$Dv_i$  – девиата i-го признака (4);

$Mv$  – число вариантов опыта;

$r_{L1.L2}$  – коэффициент корреляции L1-го и L2-го признаков.

\* Согласно правилам заполнения матрицы данных (п.2), признак с нулевой девиатой должен быть исключен из анализа.

Для определения коэффициента корреляции из условия (5) необходимо воспользоваться дифференциальным исчислением, то есть определить производную величины  $Sum_{L1.L2}$  от коэффициента корреляции  $r_{L1.L2}$  и приравнять ее нулю. В результате получаем простую формулу для вычисления коэффициента корреляции:

$$r_{L1.L2} = \sum_{j=1}^{Mv} f_{L1,j} \cdot f_{L2,j}. \quad (6)$$

Обозначения, использованные в формуле (6) такие же, как в формуле (5).

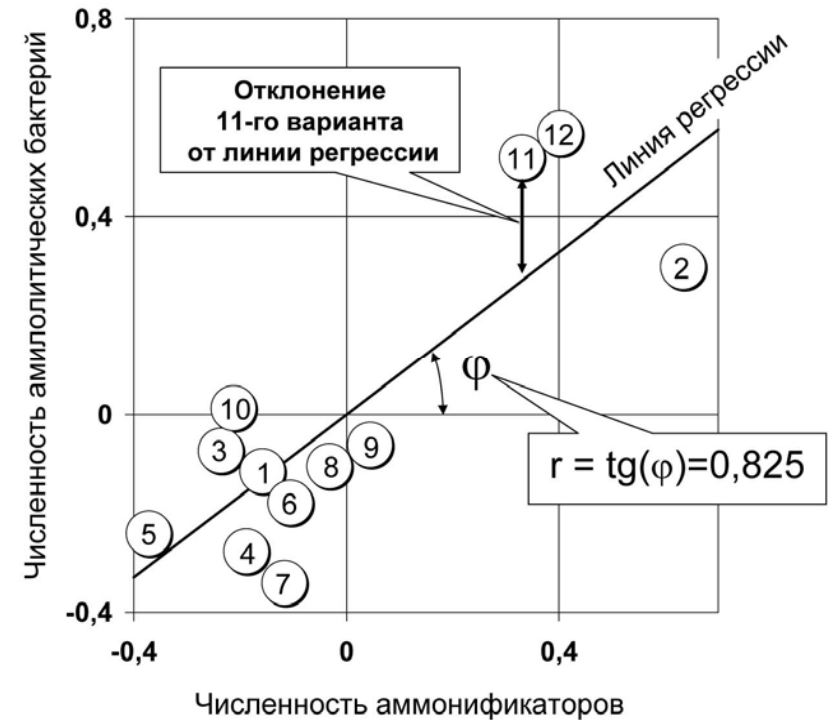


Рис. 6. Линейная регрессия фактор-нормализованных данных в пространстве двух признаков. Номера в кружках обозначают соответствующие варианты.

$\varphi$  – угол наклона линии регрессии.

$r$  – коэффициент корреляции признаков.

Для примера рассчитаем коэффициент корреляции между 1-м и 2-м признаками ( $r_{1,2}$ ):



$$r_{1,2} = (-0,157 \times -0,116) + (0,635 \times 0,299) + (-0,235 \times -0,074) + (-0,187 \times -0,278) + (-0,372 \times -0,242) + (-0,104 \times -0,179) + (-0,116 \times -0,341) + (-0,032 \times -0,105) + (-0,045 \times -0,063) + (-0,211 \times 0,011) + (0,331 \times 0,520) + (0,403 \times 0,567) = 0,825$$

В программе Graph3.xls на листе <Матрица> приводится треугольная матрица, содержащая коэффициенты корреляции всех комбинаций пар признаков (рис. 7).

На рис. 7 отмечены пары признаков с наибольшим, по модулю, ( $r_{1,2} = 0,825$ ; 1-й и 2-й признаки) и с наименьшим, по модулю, ( $r_{12,13}=0$ ; 12-й и 13-й признаки) коэффициентами корреляции.

Также как и для нормированного Евклидова расстояния, различают три диапазона коэффициентов корреляции признаков:

$$\text{сильная корреляция} \quad 0,7 < |r| \leq 1,0;$$

$$\text{средняя корреляция} \quad 0,3 < |r| \leq 0,7;$$

$$\text{слабая корреляция} \quad 0,0 < |r| \leq 0,3.$$

В граф-анализе значения коэффициентов корреляции используются для построения схемы (*графа максимальных корреляций*), состоящей из связей с максимальными коэффициентами корреляций признаков.

### 1.7. Линии Min-фактора и Мах-фактора в пространстве признаков.

В фактор-нормализованном пространстве признаков существуют две единственные линии, имеющие специальные *Min*-фактор и *Мах*-фактор, проходящие через начало координат и обладающие следующими свойствами (рис.8).

Линия *Min*-фактора располагается в пространстве признаков таким образом, что сумма квадратов проекций вариантов на нее является *наименьшей величиной* (рис. 8). Линия *Мах*-фактора располагается в пространстве признаков таким образом, что сумма квадратов проекций вариантов на нее является *наибольшей величиной*.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1														
2	Признаки	1	2	3	4	5	6	7	8	9	10	11	12	13
3	1		0,825	-0,070	0,383	0,659	0,313	0,462	0,154	-0,299	0,563	0,273	0,095	0,291
4	2			-0,059	0,512	0,795	0,530	0,253	-0,023	-0,475	0,362	0,077	0,167	0,182
5	3				-0,175	-0,330	-0,417	0,015	0,177	0,121	-0,135	-0,127	-0,065	0,363
6	4					0,637	0,512	-0,131	-0,047	0,366	0,712	0,568	0,631	-0,047
7	5						0,760	0,104	-0,251	-0,201	0,555	0,367	0,385	0,108
8	6							-0,226	-0,149	0,008	0,253	0,277	0,352	0,064
9	7								0,096	-0,431	0,028	-0,411	-0,592	0,175
10	8									0,240	-0,226	-0,173	0,067	0,620
11	9										0,321	0,620	0,572	-0,004
12	10											0,820	0,489	0,031
13	11												0,757	-0,115
14	12													0,000
15	13													

Рис. 7. Треугольная матрица парных коэффициентов корреляции признаков.

\*Прямые скобки означают, что коэффициент корреляции берется по модулю.

Математически свойства *Min*-фактора и *Max*-фактора записываются следующими формулами:

$$Fmin = \sum_{j=1}^{Mv} (w_1 \times f_{1,j} + w_2 \times f_{2,j} + \dots + w_{Mp} \times f_{Mp,j})^2 = min$$

для *Min*-фактора;

$$Fmax = \sum_{j=1}^{Mv} (w_1 \times f_{1,j} + w_2 \times f_{2,j} + \dots + w_{Mp} \times f_{Mp,j})^2 = max$$

для *Max*-фактора.

где  $f_{1,j}, f_{2,j} \dots f_{Mp,j}$  - значения 1, 2, ...  $Mp$  признаков соответственно для  $j$ -го варианта фактор-нормализованной матрицы данных (рис.5);

$w_1, w_2 \dots w_{Mp}$  - факторные нагрузки для 1, 2, ...  $Mp$  признаков;

$Mv$  - число вариантов опыта;

$Mp$  - число признаков.



Рис. 8. Расположение линий *Min*-фактора и *Max*-фактора в фактор-нормализованном пространстве признаков.

В аналитической геометрии доказано, что факторные нагрузки представляют собой значения косинусов некоторых углов

$$w_i = \text{Cos}(\varphi_i) \text{ и удовлетворяют условию: } \sum_{i=1}^{Mp} \text{Cos}^2(\varphi_i) = 1$$

Если все варианты в фактор-нормализованном пространстве располагаются на линии, перпендикулярной *Min*-фактору, то  $Fmin=0$  и факторные нагрузки будут определять жесткую детерминированную множественную связь признаков. В остальных случаях  $Fmin$  будет отражать долю случайной составляющей во множественной связи признаков. Для оценки допустимой величины случайной составляющей используют критерий Фишера, определяющий минимально допустимую величину отношения дисперсий *Max*-фактора к *Min*-фактору:

$$Disp_{FMIN} = Fmin / (Mv - 1) \quad \text{дисперсия } Min\text{-фактора};$$

$$Disp_{FMAX} = Fmax / (Mv - 1) \quad \text{дисперсия } Max\text{-фактора}.$$

Случайной составляющей можно пренебречь (то есть приравнять  $Fmin = 0$ ), если выполняется условие:

$$\frac{Disp_{FMAX}}{Disp_{FMIN}} > Fisher(q, Mv-1, Mv-1)$$

где  $q$  - уровень значимости (обычно выбирается равным 0,05).

$Fisher( , , )$  - функция для вычисления стандартного значения критерия Фишера.

Во вторую и третью позиции аргументов функции  $Fisher( , , )$  вводятся степени свободы варьирования для соответствующих дисперсий. В нашем случае обе дисперсии имеют одинаковое число степеней свободы варьирования и число это на единицу меньше, чем число вариантов.

## 2. ОСНОВНЫЕ ВЫХОДНЫЕ ИНФОРМАЦИОННЫЕ ФОРМЫ ГРАФ-АНАЛИЗА.

### 2.1. Дерево группировки вариантов (дендрограмма).

Данная выходная форма граф-анализа (дерево группировки вариантов, дендрограмма) используется для получения информации о главном экологическом факторе, влияющем на процессы в биосистеме.

Дендрограмма строится, используя нормированные Евклидовы межвариантные расстояния. Значения этих расстояний вычисляются в программе Clusters3.xls и помещаются на листе <Матрица> в треугольной матрице (рис. 3). Значения нормированных Евклидовых расстояний заносятся в те ячейки треугольной матрицы, которые находятся на пересечении соответствующих столбцов и строк. Например, значение Евклидова расстояния между 1-м и 2-м вариантами ( $E_{1,2} = 0,467$ ) находится на пересечении 1-й строки 2-го столбца треугольной матрицы (рис. 3).

Дендрограмма может быть построена тремя способами, используя такие алгоритмы, как:

- 1) «Дальний сосед»;
- 2) «Ближний сосед»;
- 3) «Среднее расстояние» (наименьший диаметр кластер-сферы).

Программа Clusters3.xls предоставляет возможность воспользоваться любым из этих трех алгоритмов.

На рис. 9 и 10 приведен пример, показывающий отличительные свойства алгоритмов группировки. На первом этапе построения дендрограммы три алгоритма сгруппировали одинаковые пары вариантов, образовавшие три кластера:

I кластер – варианты № 1 и 2,  $E_{1,2} = 0,173$ ;

II кластер – варианты № 3 и 4,  $E_{1,2} = 0,142$ ;

III кластер – варианты № 5 и 6,  $E_{1,2} = 0,262$ .

Различия алгоритмов обнаруживаются на следующем шаге построения дендрограммы, когда вариант № 7 подсоединяется к одному из трех кластеров.

По алгоритму «Ближний сосед» вариант №7 будет присоединен к I кластеру, так как  $E_{2,7} = 0,240$  является наименьшим расстоянием до ближайшего варианта №2 - «ближайшего соседа» - из этого кластера ( $E_{4,7} = 0,355$ ,  $E_{5,7} = 0,300$ ).

По алгоритму «Дальний сосед» вариант №7 будет присоединен к III кластеру, так как  $E_{6,7} = 0,375$  является наименьшим расстоянием до наиболее удаленного варианта №6 - «дальнего соседа» - из этого кластера ( $E_{1,7} = 0,414$ ,  $E_{3,7} = 0,404$ ).

По алгоритму «Среднее расстояние» вариант №7 будет присоеди-

нен ко II кластеру, так как диаметр кластер-круга (пунктирная линия, рис. 10)  $\phi_{3,4,7} = 0,375$  является наименьшим по сравнению диаметрами кластер-кругов: (сплошная линия)  $\phi_{1,2,7} = 0,414$ ; (двойная линия)  $\phi_{5,6,7} = 0,378$ .

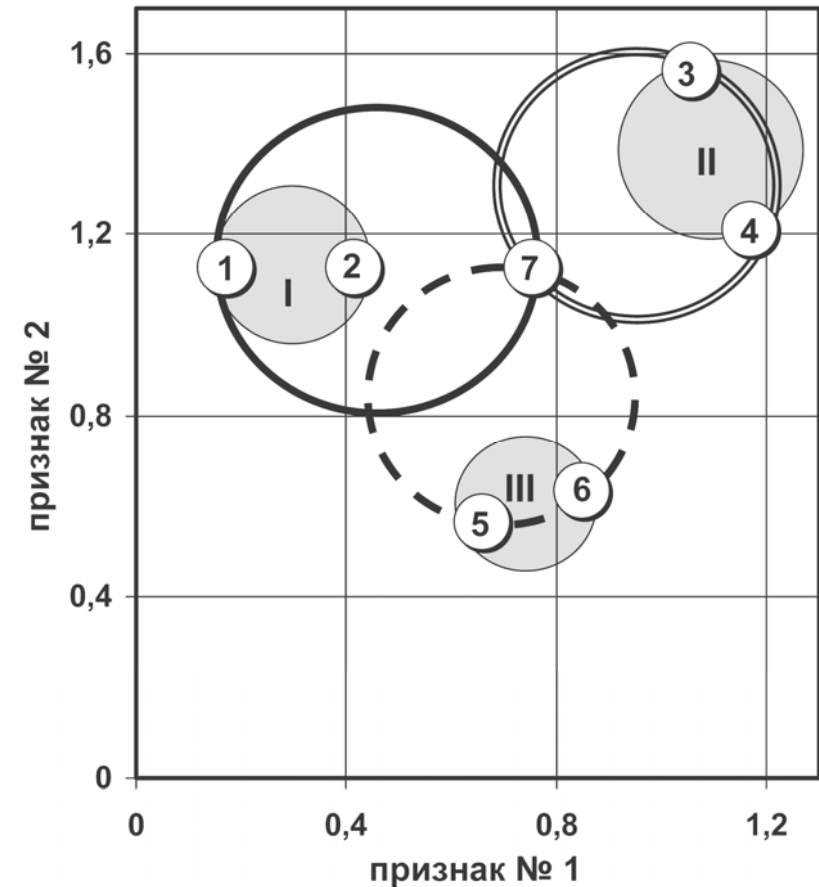


Рис. 9. Группировка варианта №7 в кластер-нормированном пространстве двух признаков:

- 1) с вариантами № 1 и 2 – по алгоритму «Ближний сосед» (кластер-круг – сплошная линия);
- 2) с вариантами № 3 и 4 – по алгоритму «Дальний сосед» (кластер-круг – двойная линия);
- 3) с вариантами № 5 и 6 – по алгоритму «Среднее расстояние» (кластер-круг – пунктирная линия).

На рис. 9 представлено расположение вариантов в двумерном кла-

стер-нормализованном Евклидовом пространстве признаков и необходимости других форм представлений особенностей группировки в этом случае нет. Проблемы с изображением характера группировки вариантов возникают при увеличении размерности пространства признаков. Для представления особенностей группировки вариантов опыта в пространстве трех и более признаков применяется специальная графическая форма – дерево группировки вариантов (дендрограмма). На рис 10А, 10Б, 10В представлены такие дендрограммы для 3-х алгоритмов группировки вариантов опыта, соответствующие исходным данным рис. 9.

Дендрограмма формируется из П-образных ломаных линий, образующих форму, похожую на дерево с ветвями. Дерево ветвится слева направо двоичным образом. Основания П-образных линий заканчиваются либо на отметке «уровень сходства = 1.0», либо на середине «полочки» следующей разветвляющейся П-образной линии. В основании П-образной линии, вблизи отметки «уровень сходства = 1.0» и правее ее, располагаются номера вариантов опыта. Полочка П-образной линии располагается напротив соответствующей отметки уровня сходства (см. рис.10А, формулу 3 из раздела 1.4). Каждая полочка П-образных линий охватывает соответствующую группу вариантов и указывает своим положением на уровень сходства (см. формулу 3).

**Ограничение на максимальный диаметр кластер-сферы для группы вариантов, соответствующих кластеру.**

Дендрограмма формируется в результате поэтапной группировки вариантов опыта. На первом шаге выявляются кластеры с наибольшим уровнем сходства, а затем каждый шаг группировки добавляет новые варианты в эти кластеры, что сопровождается увеличением диаметра соответствующей кластер-сферы. На последнем шаге объединяются все варианты в один кластер с диаметром кластер-сферы, близким к единице. Кластер, объединяющий все варианты, не содержит полезной информации. Более информативна группировка, возникающая на промежуточных этапах, то есть информативна группа, имеющая диаметр кластер-сферы, приблизительно равный 0,5. Учитывая это, кластером принято называть группу, диаметр кластер-сферы которой не превышает  $\phi \leq 0,5$ . Таким образом, дополнительный вариант включается в группу и результирующая группа останется кластером, если диаметр кластер-сферы не будет превышать величины 0,5.

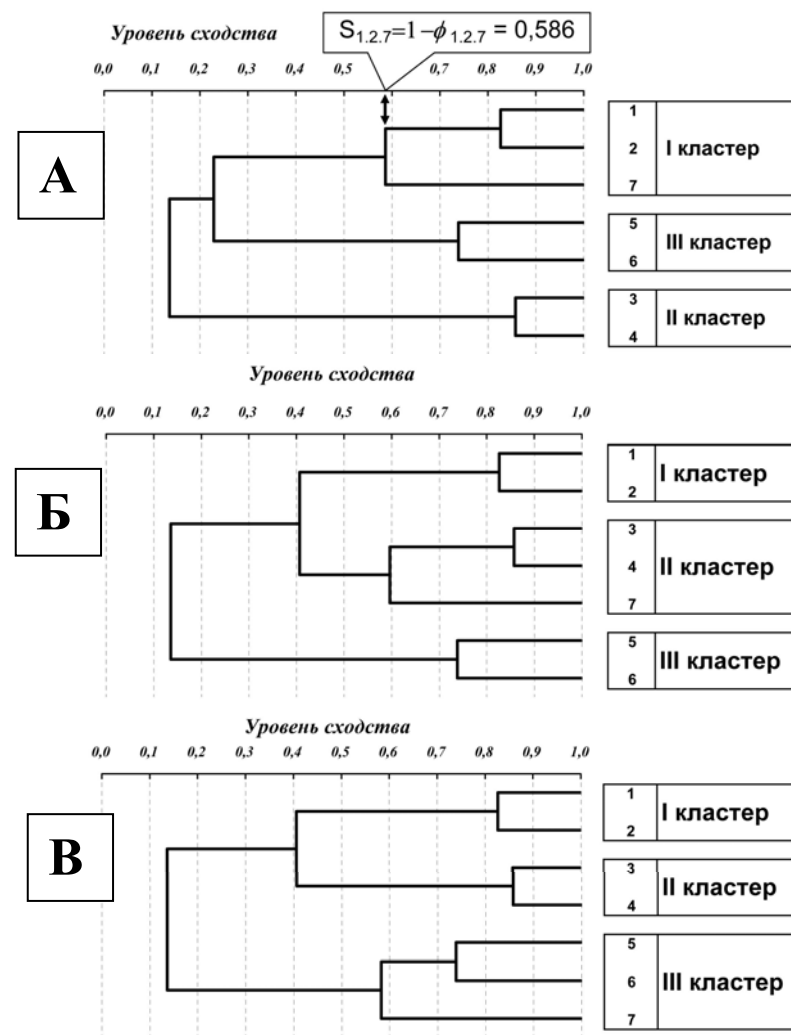


Рис. 10. Дендрограммы группировки вариантов опыта, исходные данные которого представлены на рис. 10. А – алгоритм построения «Ближний сосед». Б – алгоритм построения «Дальний сосед». В – алгоритм построения «Среднее расстояние».

На рис. 11 представлена дендрограмма для исходных данных рис. 1. Эта дендрограмма демонстрирует, что при включении варианта №2 в группу с вариантами № 1, 5, 6, 9, 10 диаметр результирующей кластер-сферы  $\phi_{1.2.5.6.9.10} > 0,6$ , то есть такую группировку вариантов нельзя называть кластером. Поэтому вариант №2 рассматривается как не-

зависимый, не входящий ни в одну группу.

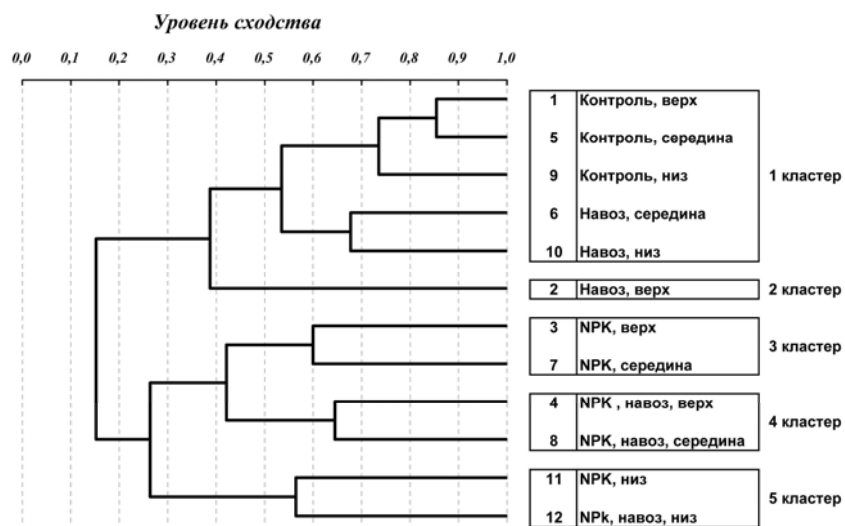


Рис. 11. Дендрограммы группировки вариантов опыта, исходные данные которого представлены на рис. 1. Алгоритм построения дендрограммы «Среднее расстояние».

### ИНФОРМАЦИОННЫЙ АНАЛИЗ.

#### Главный экологический фактор, определяющий группировку вариантов.

В рассматриваемом примере данных (рис. 1) биосистема почвенных микроорганизмов испытывает воздействие следующих экологических факторов:

- 1) применение НПК;
- 2) применение навоза;
- 3) воздействие элементов агроландшафта (вершина, середина и низ склона почвы).

Дендрограмма (рис. 11) показывает, что в 3-й, 4-й, 5-й кластеры вошли варианты, которые отличаются внесением в почву минеральных удобрений НПК, а в 1-й и 2-й кластеры вошли варианты без применения НПК. Это означает, применение НПК является главным экологическим фактором, влияющим на трансформационные процессы в почве. Остальные два экологических фактора имеют второстепенное значение и определяют меньшее различие вариантов внутри кластеров.

## 2.2. Граф максимальных коэффициентов корреляции признаков.

Данная выходная форма граф-анализа (*граф максимальных коэффициентов корреляции*, ГМК) используется для получения информации о структуре связей в биосистеме. В основу ГМК положена гипотеза о том, что непосредственные связи между компонентами биосистемы имеют наибольший по модулю коэффициент корреляции, а коэффициенты корреляции опосредованных связей меньше по модулю и вычисляются путем перемножения коэффициентов корреляции соответствующих непосредственных связей.

При ГМК используется треугольная матрица коэффициентов корреляции (рис. 7). Она вычисляется в программе Graph.xls по формуле (6) и помещается на листе <Матрица>. Значения коэффициентов корреляции заносятся в те ячейки матрицы, которые находятся на пересечении соответствующих столбцов и строк. Например, значение коэффициента корреляции ( $r_{1,2} = 0,825$ ) между 1-м и 2-м признаками находится на пересечении 2-го столбца и 1-й строки матрицы.

В программе Graph.xls ГМК формируется автоматически на листе <Граф> (рис. 12). Алгоритм построения ГМК представляет собой следующие последовательные операции.

*Шаг 1.* Ищется максимальный по модулю коэффициент корреляции (это  $r_{1,2} = 0,825$ , рис. 7) и в ГМК заносятся первые два объекта: прямоугольник с цифрой 1-го признака и прямоугольник с цифрой 2-го признака. Значение коэффициента корреляции помещается между ними.

*Шаг 2.* Ищется наибольший коэффициент корреляции среди оставшихся коэффициентов (это  $r_{10,11} = 0,820$ ). Данный коэффициент корреляции связывает 10-й и 11-й признаки (соответствующие им объекты появляются в ГМК).

*Последующие шаги.* Находим  $r_{2,5} = 0,795$ ,  $r_{5,6} = 0,760$  и т.д.

В результате, на листе <Граф> появляются все необходимые прямоугольники (или узлы ГМК) и связи между узлами ГМК (в виде значений соответствующих коэффициентов корреляции между признаками). Алгоритм построения графа на этом заканчивается.

Следует отметить, что ГМК (рис. 12) содержит только непосредственные связи между признаками. Он является ненаправленным графом, так как не отражает направление причинно-следственных связей между компонентами биосистемы.

### Коэффициенты корреляции признаков, связанных опосредованно.

Опосредованные связи имеют меньший по модулю коэффициент корреляции, чем непосредственные связи, и могут быть в первом приближении вычислены путем перемножения коэффициентов корреляции для промежуточных звеньев опосредованной связи.

Например, между 5-м и 10-м признаками коэффициент корреляции равен  $r_{5,10} = 0,555$  (см. рис. 7 и 12). Эти признаки в ГМК связаны через посредника (признак №4), то есть опосредованная связь состоит из двух звеньев с коэффициентами корреляции:  $r_{4,5} = 0,637$ ,  $r_{4,10} = 0,712$ . Произведение этих коэффициентов  $r_{4,5} \times r_{4,10} = 0,637 \times 0,712 = 0,455$ , что приблизительно равно  $r_{5,10}$ . Аналогичный результат получается для опосредованной связи между 7-м и 11-м признаками. Для этих признаков коэффициент корреляции равен  $r_{7,11} = -0,411$ , что приблизительно равно произведению  $r_{7,12} \times r_{11,12} = 0,757 \times -0,592 = -0,448$ .

Таким образом, чтобы получить величину коэффициента корреляции опосредованной связи двух признаков с помощью ГМК необходимо перемножить все коэффициенты корреляции, встречающиеся в цепи связей этих признаков. Это нетривиальное свойство ГМК демонстрирует его высокую информативность и достоверность с точки зрения представления схемы связей компонентов в биосистемах.

### ИНФОРМАЦИОННЫЙ АНАЛИЗ.

#### Минимаксное ограничение на уровень корреляционных связей в графе.

В рассматриваемом примере (рис. 12) в ГМК признак № 3 (численности нитрификаторов 1 фазы) не имеет связи ни с одним признаком, хотя максимальный по модулю коэффициент корреляции обнаруживается между признаками № 3 и 6:  $r_{3,6} = -0,417$ , то есть  $|r_{3,6}| < 0,5$ .

Такой уровень корреляции может возникнуть по разным причинам. Одной из причин может быть неполнота данных, то есть в данных отсутствует признак, посредством которого он связан с остальными. При этом вычисляются только коэффициенты корреляции опосредованных связей, имеющих заведомо низкий уровень. По этой причине можно предположить, что в ГМК (рис. 12) нитрификаторы 1 фазы (признак №3) не связаны ни с одним компонентом биосистемы.

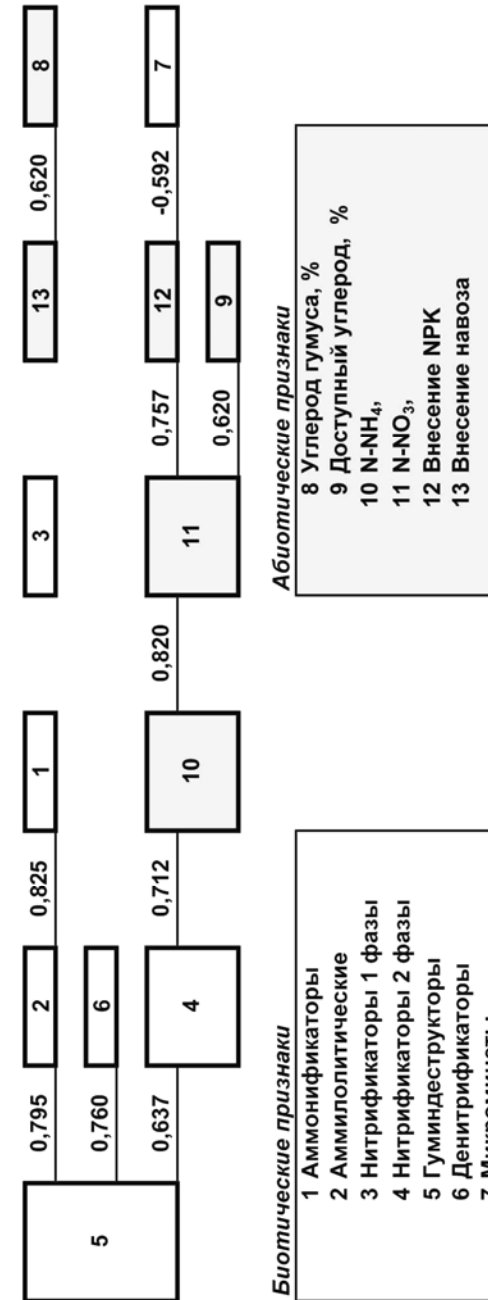


Рис. 12. Граф максимальных коэффициентов корреляции для биотических и абиотических признаков исходных данных (рис. 1).

При слабых корреляциях ( $|r| < 0,3$ ) вполне вероятно, что между данными компонентами связь вообще отсутствует и в ГМК она также не должна фиксироваться.

Учитывая это, при построении графа используется минимаксное ограничение на уровень значимости корреляционной связи в ГМК.

*Связь фиксируется в графе максимальных коэффициентов корреляции, если модуль коэффициента корреляции данного признака с одним из наиболее близких признаков (наиболее коррелированная пара признаков) превышает или равен 0,5.*

По этой же причине в ГМК (рис. 12) не включены связи признаков №8 и 13 с остальными признаками.

Наличие или отсутствие связей в ГМК, а также схема этих связей, важны для понимания процессов, происходящих в биосистеме. В рассматриваемом примере (рис. 12) выделяется блок взаимосвязанных микроорганизмов (признаки № 1, 2, 4, 5, 6). Отдельное независимое положение в графе занимают нитрификаторы 1 фазы (признак № 3) и микромицеты (признак № 7). Все это указывает на особую роль микроорганизмов в биотрансформации веществ и образование специальных трансформационных цепей в биосистемах.

Знак коэффициентов корреляции в ГМК также несет специфическую информацию. Например, численность микромицетов связана с признаком № 12 (внесение NPK, коэффициент корреляции отрицательный  $r = -0,592$ ). Это означает, что в вариантах с внесенным NPK численность микромицетов меньше, чем в вариантах без внесения NPK. Известно, что микромицеты играют важную роль в процессах синтеза почвенного гумуса. Следовательно, ГМК демонстрирует отрицательное влияние внесения NPK на гумусообразовательные процессы. Это же подтверждает обособленная связь в ГМК углерода гумуса (признак № 8) с признаком № 13 (внесение навоза).

### 2.3. Направленный граф биотрансформации веществ.

Данная выходная форма граф-анализа (*направленный граф биотрансформации веществ*, НГ) используется для получения информации о направленности процессов биотрансформации веществ в биосистеме и схеме трансформационных потоков в ней. При построении НГ компоненты биосистем делятся на два класса:

- 1) биотические компоненты – микроорганизмы, растения, животные и т.п.;
- 2) абиотические компоненты – различные вещества, которые синтезируются, разлагаются или поглощаются живыми организмами.

При функционировании биосистем каждый абиотический компонент подвергается множественному воздействию со стороны биотических компонентов. Поэтому значения признаков, соответствующие им являются функционально зависимыми от биотических признаков. При построении направленного графа внутрисистемных взаимодействий абиотическим признакам присваивается наименование: *признаки-функции*, а биотическим признакам – *признаки-аргументы*. Признаки-аргументы подразделяются, в свою очередь, на два подкласса: *синтез-аргументы* и *аргументы-деструкторы*, то есть микроорганизмы могут участвовать в цепи превращений веществ, как в качестве синтезаторов этих веществ, так и в качестве их деструкторов. Для того чтобы определить, к какому подклассу относится данный признак, воспользуемся факторным анализом (раздел. 1.7).

Математическая модель, описывающая динамику роста микробной популяции в присутствии питательного субстрата, представляет скорость изменения концентрации субстрата ( $V_S$ ) следующей формулой:

$$V_S = F(R_C, R_N, C_S, C_{C/N}) \cdot X, \quad (7)$$

где  $F(R_C, R_N, C_S, C_{C/N}) = \frac{R_C \cdot R_N}{\frac{R_C \cdot R_N}{C_S} + C_{C/N} \cdot R_C + R_N}$  – ресурсный

множитель, описывающий субстратное ингибирование процессов развития популяций;

$R_C, R_N$  – углеродный и азотный питательный ресурс;

$K_S$  – константа Михаэлиса, константа субстратного ингибирования;

$K_{C/N}$  – константа сопряженности углеродных и азотных потоков биотрансформации веществ;

$X$  – численность микроорганизмов.

Углеродный и азотные ресурсы будем оценивать по концентрациям аммонийного и нитратного азота в почве:

$$R_C = \alpha_C \cdot (N \cdot NH_4 + N \cdot NO_3); \quad (8)$$

$$R_N = \alpha_N \cdot N \cdot NO_3,$$

где  $\alpha_C, \alpha_N$  – постоянные коэффициенты.

После подстановки (8) в (7) получим формулу ресурсного множите-

ля в терминах значений признаков (рис. 1):

$$F(P_{10,i}, P_{11,i}, K_S, K_{C/N}) = \frac{(P_{10,i} + P_{11,i}) \cdot P_{11,i}}{K_S \cdot (P_{10,i} + P_{11,i}) \cdot P_{11,i} + K_{C/N} \cdot (P_{10,i} + P_{11,i}) + P_{11,i}}, \quad (10)$$

где  $K_S = \frac{\alpha_C}{C_S}$ ,  $K_{C/N} = \frac{\alpha_C}{\alpha_N} \cdot C_{C/N}$ . Коэффициент  $\alpha_C$ , который появляется в числителе (10) в виде сомножителя, игнорируется, так как в факторном анализе постоянные множители в данных не влияют на дивергенцию этих данных (см. формулу 12).

Учитывая это, в нашем примере, в факторном анализе в качестве признаков-аргументов используются признаки, образованные путем перемножения ресурсного множителя (10) на значения биотических признаков (признаков № 1-7). В результате к исходным данным (рис. 1) добавится еще 7 строчек, которые должны содержать следующие произведения:

$$P_{14,i} = P_{1,i} \times F(P_{10,i}, P_{11,i}, K_S, K_{C/N}), P_{15,i} = P_{2,i} \times F(P_{10,i}, P_{11,i}, K_S, K_{C/N}), \dots,$$

$$P_{19,i} = P_{6,i} \times F(P_{10,i}, P_{11,i}, K_S, K_{C/N}), P_{20,i} = P_{7,i} \times F(P_{10,i}, P_{11,i}, K_S, K_{C/N}).$$

Фактор-нормализация этих дополнительных признаков представлена на рис. 13.

В табл.2 приведены факторные нагрузки, рассчитанные с помощью программы Factor3.xls для признаков (рис. 1, 13). Можно показать, что синтез-аргументам соответствуют *отрицательные* факторные нагрузки ( $w_i < 0$ ), а аргумент-деструкторам – положительные ( $w_i > 0$ ). Для этого рассмотрим подробнее основное свойству *Min*-фактора (раздел 1.7). Согласно этому свойству, сумма произведений факторных нагрузок на значения фактор-нормализованных значений признаков (рис. 1, 13) для каждого варианта опыта приблизительно равна нулю. Например, для признак-функции №8 (углерод гумуса), признаков-аргументов № 14-20 (рис. 13) и варианта № 1 получим:

$$\begin{aligned} & (0,0306 \times -0,157) + (-0,444 \times -0,286) + (-0,0505 \times -0,263) + \\ & + (-0,1002 \times -0,301) + (-0,0925 \times -0,248) + (0,8012 \times -0,26) + \\ & + (-0,3381 \times -0,181) + (0,1566 \times -0,286) = -0,003 \approx 0 \end{aligned}$$

A	B	Варианты																					
		1	2	3	4	5	6	7	8	9	10	11	12										
1																							
2																							
3	14 Аммонификаторы x F(NH <sub>4</sub> , NO <sub>3</sub> , K <sub>S</sub> , K <sub>C/N</sub> )	-0,286	0,322	-0,035	0,068	-0,286	-0,286	0,087	0,087	0,087	0,087	0,459	0,459	0,459	-0,286	-0,286	-0,286	-0,286	-0,286	-0,286	0,512	0,512	0,020
4	15 Аммонификаторы x F(NH <sub>4</sub> , NO <sub>3</sub> , K <sub>S</sub> , K <sub>C/N</sub> )	-0,263	0,249	0,017	-0,012	-0,263	-0,263	0,051	0,051	0,051	0,051	0,152	0,152	0,152	-0,263	-0,263	-0,263	-0,263	-0,263	-0,263	0,743	0,743	0,117
5	16 Нитрификаторы 1 фазы x F(NH <sub>4</sub> , NO <sub>3</sub> , K <sub>S</sub> , K <sub>C/N</sub> )	-0,301	0,443	0,376	0,304	-0,301	-0,301	0,271	0,271	0,271	0,271	0,200	0,200	0,200	-0,301	-0,301	-0,301	-0,301	-0,301	-0,301	-0,024	-0,024	-0,063
6	17 Нитрификаторы 2 фазы x F(NH <sub>4</sub> , NO <sub>3</sub> , K <sub>S</sub> , K <sub>C/N</sub> )	-0,248	-0,004	0,029	0,246	-0,248	-0,248	0,500	0,500	0,500	0,500	0,500	0,500	0,500	-0,248	-0,248	-0,248	-0,248	-0,248	-0,248	0,602	0,602	0,003
7	18 Гуминдеструкторы x F(NH <sub>4</sub> , NO <sub>3</sub> , K <sub>S</sub> , K <sub>C/N</sub> )	-0,260	0,083	-0,033	0,015	-0,260	-0,260	0,587	0,587	0,587	0,587	0,587	0,587	0,587	-0,260	-0,260	-0,260	-0,260	-0,260	-0,260	0,552	0,552	0,066
8	19 Денитрификаторы x F(NH <sub>4</sub> , NO <sub>3</sub> , K <sub>S</sub> , K <sub>C/N</sub> )	-0,181	-0,144	-0,149	-0,096	-0,181	-0,181	0,728	0,728	0,728	0,728	0,728	0,728	0,728	-0,181	-0,181	-0,181	-0,181	-0,181	-0,181	0,471	0,471	0,165
9	20 Микромицеты x F(NH <sub>4</sub> , NO <sub>3</sub> , K <sub>S</sub> , K <sub>C/N</sub> )	-0,286	0,506	0,065	0,114	-0,286	-0,286	0,457	0,457	0,457	0,457	0,457	0,457	0,457	-0,286	-0,286	-0,286	-0,286	-0,286	-0,286	0,317	0,317	0,062
10																							

Рис. 13. Фактор-нормализованная матрица признаков, полученных путем произведения ресурсного множителя (10) на численности микроорганизмов.



Таблица 2. Факторные нагрузки, вычисленные с помощью программы Factor3.xls для 4-х наборов признаков.

№	Признаки-аргументы	Факторные нагрузки Min-фактора для признака-функции №			
		8	9	10	11
14	Аммонификаторы × $F(\text{NH}_4, \text{NO}_3, \text{K}_S, \text{K}_{\text{C/N}})^*$	-0,4440	-0,3970	0,4146	-0,3970
15	Аммилолитические × $F(\text{NH}_4, \text{NO}_3, \text{K}_S, \text{K}_{\text{C/N}})$	-0,0505	0,0846	0,0694	0,2931
16	Нитрификаторы 1 фазы × $F(\text{NH}_4, \text{NO}_3, \text{K}_S, \text{K}_{\text{C/N}})$	-0,1002	-0,1699	0,0981	-0,2101
17	Нитрификаторы 2 фазы × $F(\text{NH}_4, \text{NO}_3, \text{K}_S, \text{K}_{\text{C/N}})$	-0,0925	-0,2482	0,0790	-0,1448
18	Гуминдеструкторы × $F(\text{NH}_4, \text{NO}_3, \text{K}_S, \text{K}_{\text{C/N}})$	0,8012	0,7659	-0,8047	0,4762
19	Денитрификаторы × $F(\text{NH}_4, \text{NO}_3, \text{K}_S, \text{K}_{\text{C/N}})$	-0,3381	-0,3430	0,3683	0,0513
20	Микромицеты × $F(\text{NH}_4, \text{NO}_3, \text{K}_S, \text{K}_{\text{C/N}})$	0,1566	0,1639	-0,1553	0,2416
<b>Признаки-функции</b>					
8	Углерод гумуса, %	0,0306	-	-	-
9	Доступный углерод, %	-	0,1170	-	-
10	N-NH <sub>4</sub>	-	-	0,0110	-
11	N-NO <sub>3</sub>	-	-	-	0,6353
	$Disp_{\text{FMIN}}$	$1,1 \times 10^{-4}$	$1,2 \times 10^{-4}$	$7,3 \times 10^{-4}$	$5,3 \times 10^{-11}$
	$Disp_{\text{FMAX}}$	2,567	2,346	2,588	1,615
	Фактический критерий Фишера $Disp_{\text{FMAX}} / Disp_{\text{FMIN}}$	24254	20377	3563	$3,0 \times 10^{10}$
	Стандартный критерий Фишера $Fisher(0,05, 11, 11)$	2,82	2,82	2,82	2,82

\* Ресурсный множитель, заданный формулой (10).  $K_S=0,0135$ .  $K_{\text{C/N}}=9,55$ .

Для общего случая можно записать формулу, связывающую девиации признака-функции с девиациями признаков-аргументов:

$$W_{\text{FUN}} \times f_{\text{FUN},j} = (-W_{\text{ARG},1}) \times f_{\text{ARG},1,j} + \dots + (-W_{\text{ARG},L}) \times f_{\text{ARG},L,j} \quad (11)$$

где  $f_{\text{FUN},j}$  - значение признака-функции для j-того варианта фактор-нормализованной матрицы данных (рис.1);

$f_{\text{ARG},1,j}$ ,  $f_{\text{ARG},2,j}$  ...  $f_{\text{ARG},L,j}$  - значения 1, 2, ... L признаков-аргументов для j-того варианта фактор-нормализованной матрицы данных (рис.13);

$W_{\text{FUN}}$  - факторная нагрузка признака-функции (табл. 2);

$W_{\text{ARG},1}$ ,  $W_{\text{ARG},2}$  ...  $W_{\text{ARG},L}$  - факторные нагрузки для 1, 2, ... L признаков-аргументов (табл. 2).

Из формулы (11) следует, что при отрицательном значении факторной нагрузки признак-функция возрастает с ростом значения соответствующего признака-аргумента. При положительном значении факторной нагрузки признак-функция будет уменьшаться с возрастанием значения соответствующего признака-аргумента. Из этих рассуждений следует, что отрицательные факторные нагрузки указывают на процессы синтеза абиотического компонента, а положительные – на деструкцию.

### ИНФОРМАЦИОННЫЙ АНАЛИЗ.

#### Ограничение на уровень значимости связей в графе.

Из данных факторного анализа (табл. 2) следует, что деструкторами гумуса являются гуминдеструкторы (признак № 18) и микромицеты (признак № 20), а остальные биотические компоненты являются синтезаторами этого вещества. По отношению к аммонии (признак № 10, 3-й столбец таблицы № 2) роли микроорганизмов меняются. Для наглядного представления процессов синтеза/деструкции веществ в биосистеме по результатам факторного анализа можно построить направленный граф биотрансформации веществ (рис. 14).

Если факторная нагрузка признака-аргумента по модулю меньше, чем 0,05, то будем считать, что влиянием этого признака на абиотический компонент можно пренебречь. При построении направленного графа биотрансформации веществ такое *ограничение на уровень значимости связи* позволяет упростить вид графа и повысить его информативность.

Наличие или отсутствие связей в графе (рис. 14), а также схема и направление этих связей, важны для понимания процессов, происходящих в биосистеме. В рассматриваемом графе выделяется роль аммонификаторов и нитрификаторов 2 фазы (биотические признаки № 14, 17), которые выступают только в роли синтез-аргументов по отношению ко всем контролируемым в опыте абиотическим компонентам (за исключением аммония). У гуминдеструкторов выступают в основном в роли признака-деструктора, а аммоний они синтезируют.

Граф биотрансформации веществ демонстрирует присутствие в биосистеме признаков-аргументов обоих типов: синтезаторов и деструкторов веществ, что является по-видимому необходимыми атрибутами любой биосистемы. Наличие в биосистеме компонентов разных типов позволяет ей эффективно адаптироваться к изменяющимся экологическим условиям. Осуществляется это следующим образом. При изменении экологических условий в биосистеме происходит сдвиг в

численностях микроорганизмов. За этим следует изменение точки баланса процессов синтеза/деструкции и изменение концентрации веществ.

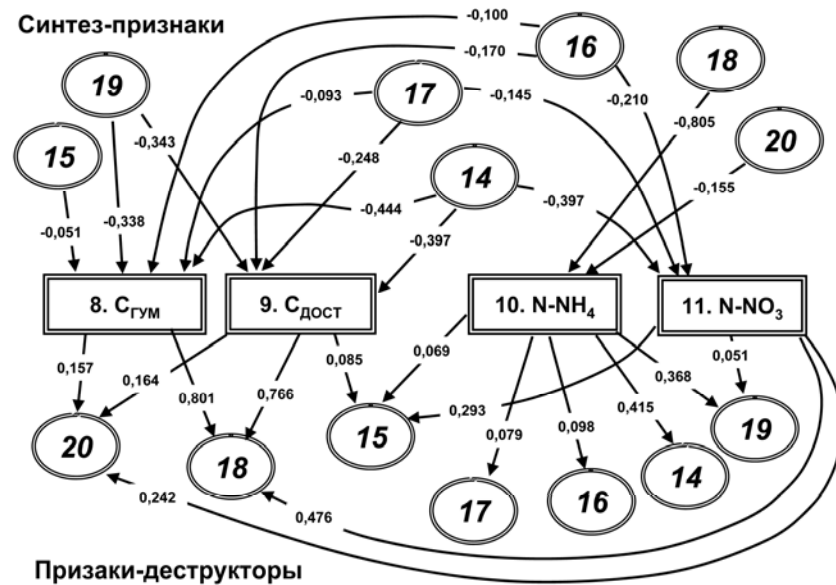


Рис. 14. Направленный граф биотрансформации веществ. Числа, расположенные около стрелок являются значениями факторных нагрузок соответствующих признаков-аргументов. Прямоугольники обозначают признаки-функции (абиотические компоненты биосистемы), а эллипсы - признаки-аргументы (биотические компоненты). В верхней части рисунка расположены синтез-признаки, а в нижней – признаки-деструкторы.

#### 2.4. Математическая модель дивергенции компонентов биосистемы.

Данная выходная форма граф-анализа (математическая модель дивергенции компонентов биосистемы) используется для получения информации о количественных закономерностях баланса процессов синтеза/деструкции в биосистеме при изменении экологических условий. Получаемая математическая модель представляет собой систему алгебраических уравнений, в которых представляется зависимость дивергенции абиотических признаков-функций от дивергенции биотических признаков-аргументов. Математическая модель позволяет по численностям микроорганизмов рассчитывать концентрации веществ в почве. Сравнивая расчетные параметры с данными опыта, удается

проанализировать механизмы адаптации биосистемы к изменению экологических условий.

Дивергенцией признака в  $j$ -ом варианте опыта ( $d_{i,j}$ ) называется относительная величина отклонения  $i$ -того признака от среднего значения:

$$d_{i,j} = P_{i,j} : q_i - 1, \quad (12)$$

где  $q_i = \frac{1}{Mv} \cdot \sum_{j=1}^{Mv} P_{i,j}$  - среднее арифметическое  $i$ -того признака;

$P_{i,j}$  - значение  $i$ -го признака для  $j$ -ого варианта исходной матрицы данной (рис. 1);

$Mv$  - число вариантов опыта.

Для построения математической модели дивергенции компонентов биосистемы необходимо провести: (i) функциональную идентификацию и (ii) параметрическую идентификацию модели. На первом этапе определяется вид, структура математической формулы модели, а на втором – значения постоянных коэффициентов формулы. Для определения вида этой формулы воспользуемся данными факторного анализа и преобразуем (11) с учетом (12):

$$d_{F,j} = b_{F,1} \times d_{1,j} + b_{F,2} \times d_{2,j} + \dots + b_{F,L} \times d_{L,j}, \quad (13)$$

где  $d_{F,j}$  - значение дивергенции признака-функции с  $F$ -номером для  $j$ -того варианта опыта;

$d_{1,j}, d_{2,j} \dots d_{L,j}$  - значения дивергенции 1, 2, ...  $L$  признаков-аргументов для  $j$ -того варианта опыта;

$$b_{F,i} = -\frac{q_i}{q_F} \times \frac{Dv_F}{Dv_i} \times \frac{w_i}{w_F} - \text{коэффициенты биотрансформации}$$

веществ для признака-функции с  $F$ -номером и  $i$ -того признака-аргумента;

$q_i, Dv_i, w_i$  - среднее арифметическое, девиата и факторная нагрузка для  $i$ -того признака-аргумента;

$q_F, Dv_F, w_F$  - среднее арифметическое, девиата и факторная нагрузка для признака-функции с  $F$ -номером.

На этапе функциональной идентификации модели осуществляется поиск такой математической композиции исходных признаков, при ко-

торой дисперсия *Min*-фактора будет наименьшей (раздел 1.7, табл. 2,  $Disp_{FMIN}$ ). Математическая теория популяционных процессов описывает динамику микробной трансформации веществ с помощью ресурсного множителя (см. формулу 10). Сравнение дисперсий *Min*-фактора для различных композиций признаков показало, что наименьшей дисперсией обладает комбинация признаков, представляющих собой произведение численностей микроорганизмов (признаки № 1-7) на ресурсный множитель (формула 10) при  $K_S=0,135$ .  $K_{C/N}=9,55$  (см. табл. 2,  $Disp_{FMIN} = 5,3 \times 10^{-11} - 7,3 \times 10^{-4}$ ).

Результаты параметрической идентификации математической модели (второй этап моделирования) представлены в табл. 3.

Таблица 3. Коэффициенты биотрансформации математической модели дивергенции компонентов.

№	Признаки-аргументы	Коэффициенты биотрансформации ( $b_{F,i}$ ) для признака-функции №			
		8. С <sub>ГУМ</sub>	9. С <sub>ДОСТ</sub>	10. NH <sub>4</sub>	11. NO <sub>3</sub>
14	Аммонификаторы $\times F(NH_4, NO_3, K_S, K_{C/N})^*$	1,1610	1,2708	-12,974	0,6282
15	Аммилолитические $\times F(NH_4, NO_3, K_S, K_{C/N})$	0,1215	-0,2489	-1,9970	-0,4263
16	Нитрификаторы 1 фазы $\times F(NH_4, NO_3, K_S, K_{C/N})$	0,2755	0,5719	-3,2284	0,3495
17	Нитрификаторы 2 фазы $\times F(NH_4, NO_3, K_S, K_{C/N})$	0,2095	0,6882	-2,1408	0,1984
18	Гуминдеструкторы $\times F(NH_4, NO_3, K_S, K_{C/N})$	-1,9005	-2,2242	22,842	0,6835
19	Денитрификаторы $\times F(NH_4, NO_3, K_S, K_{C/N})$	0,5597	0,6952	-7,2950	-0,0513
20	Микромицеты $\times F(NH_4, NO_3, K_S, K_{C/N})$	-0,4091	-0,5243	4,8534	-0,3818

\* Ресурсный множитель, заданный формулой (10).  $K_S=0,0135$ .  $K_{C/N}=9,55$ .

В результате получаем математическую модель дивергенции компонентов биосистемы, которая представляет собой систему нелинейных алгебраических уравнений и в матричном виде записывается следующим образом:

$$\begin{bmatrix} P_8 : q_8 - 1 \\ P_9 : q_9 - 1 \\ P_{10} : q_{10} - 1 \\ P_{11} : q_{11} - 1 \end{bmatrix} = \begin{bmatrix} b_{8,14} & b_{8,15} & b_{8,16} & b_{8,17} & b_{8,18} & b_{8,19} & b_{8,20} \\ b_{9,14} & b_{9,15} & b_{9,16} & b_{9,17} & b_{9,18} & b_{9,19} & b_{9,20} \\ b_{10,14} & b_{10,15} & b_{10,16} & b_{10,17} & b_{10,18} & b_{10,19} & b_{10,20} \\ b_{11,14} & b_{11,15} & b_{11,16} & b_{11,17} & b_{11,18} & b_{11,19} & b_{11,20} \end{bmatrix} \times \begin{bmatrix} P_1 \cdot F(P_{10}, P_{11}, K_S, K_{C/N}) : q_{14} - 1 \\ P_2 \cdot F(P_{10}, P_{11}, K_S, K_{C/N}) : q_{15} - 1 \\ P_3 \cdot F(P_{10}, P_{11}, K_S, K_{C/N}) : q_{16} - 1 \\ P_4 \cdot F(P_{10}, P_{11}, K_S, K_{C/N}) : q_{17} - 1 \\ P_5 \cdot F(P_{10}, P_{11}, K_S, K_{C/N}) : q_{18} - 1 \\ P_6 \cdot F(P_{10}, P_{11}, K_S, K_{C/N}) : q_{19} - 1 \\ P_7 \cdot F(P_{10}, P_{11}, K_S, K_{C/N}) : q_{20} - 1 \end{bmatrix} \quad (14)$$

где  $b_{F,j}$  – коэффициент биотрансформации  $i$ -того признака-аргумента по отношению к  $F$ -тому признаку-функции, представленный в табл. 3 ( $b_{9,16} = 0,5719$ );  
 $q_i$  – среднее арифметическое  $i$ -того признака, представленное в табл. 4.

С помощью модели (14) можно вычислить концентрации углерод- и азотсодержащих веществ, зная численности микроорганизмов. Следует отметить, что обратную задачу (поиск числа микроорганизмов по концентрациям веществ) с помощью этой модели решить не удастся, т.к. число абиотических признаков меньше биотических. Из этого следует, что для уверенного контроля за абиотическими компонентами с помощью математической модели необходимо, чтобы число биотических признаков было больше числа абиотических признаков\*. В нашем примере это условие выполняется: число биотических признаков равно 7, число абиотических признаков равно 4.

Таблица 4. Среднее арифметическое признаков, используемых в математической модели дивергенции компонентов биосистемы.

№	Признаки	Среднее арифметическое признака ( $q_i$ )
8	Углерод гумуса, %	1,084
9	Доступный углерод, %	0,170
10	N-NH <sub>4</sub>	1,567
11	N-NO <sub>3</sub>	1,717
14	Аммонификаторы $\times F(NH_4, NO_3, K_S, K_{C/N})^{**}$	3,111
15	Аммилолитические $\times F(NH_4, NO_3, K_S, K_{C/N})$	2,331
16	Нитрификаторы 1 фазы $\times F(NH_4, NO_3, K_S, K_{C/N})$	4,403
17	Нитрификаторы 2 фазы $\times F(NH_4, NO_3, K_S, K_{C/N})$	6,269
18	Гуминдеструкторы $\times F(NH_4, NO_3, K_S, K_{C/N})$	2,678
19	Денитрификаторы $\times F(NH_4, NO_3, K_S, K_{C/N})$	0,117
20	Микромицеты $\times F(NH_4, NO_3, K_S, K_{C/N})$	4,718

\*\* Ресурсный множитель, заданный формулой (10).  $K_S=0,0135$ .  $K_{C/N}=9,55$ .

### ИНФОРМАЦИОННЫЙ АНАЛИЗ.

#### Углеродный питательный ресурс и минерализация почвенного гумуса.

Математическая модель (14) позволяет не только рассчитывать концентрации веществ, но и анализировать трансформационные процессы в деталях. Одним из важных положений модели является пропорциональность концентрации минерального азота углеродному ресурсу почвы (формулы 8). Из этого следует, что в каждом варианте опыта углеродный ресурс, используемый микроорганизмами, пропорционален количеству минерального азота. На рис. 15 представлена

гистограмма распределения количества минерального азота по вариантам опыта, с помощью которого можно оценить величину используемого микроорганизмами углеродного ресурса.

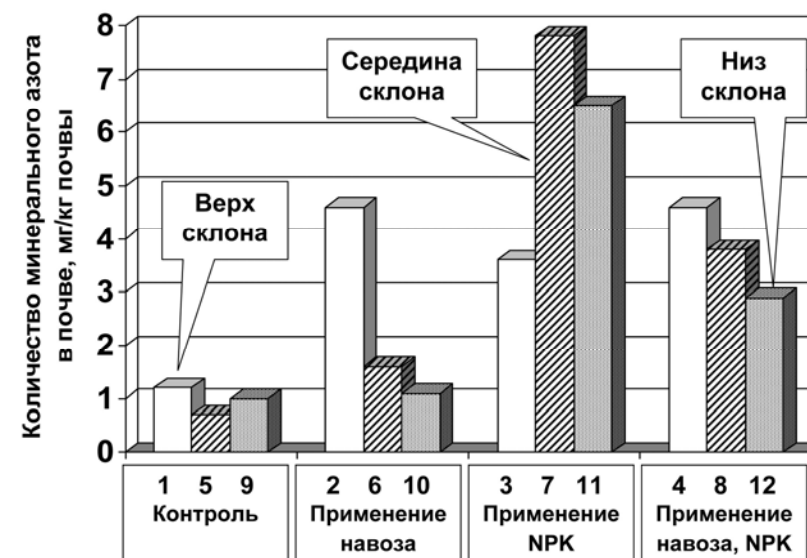
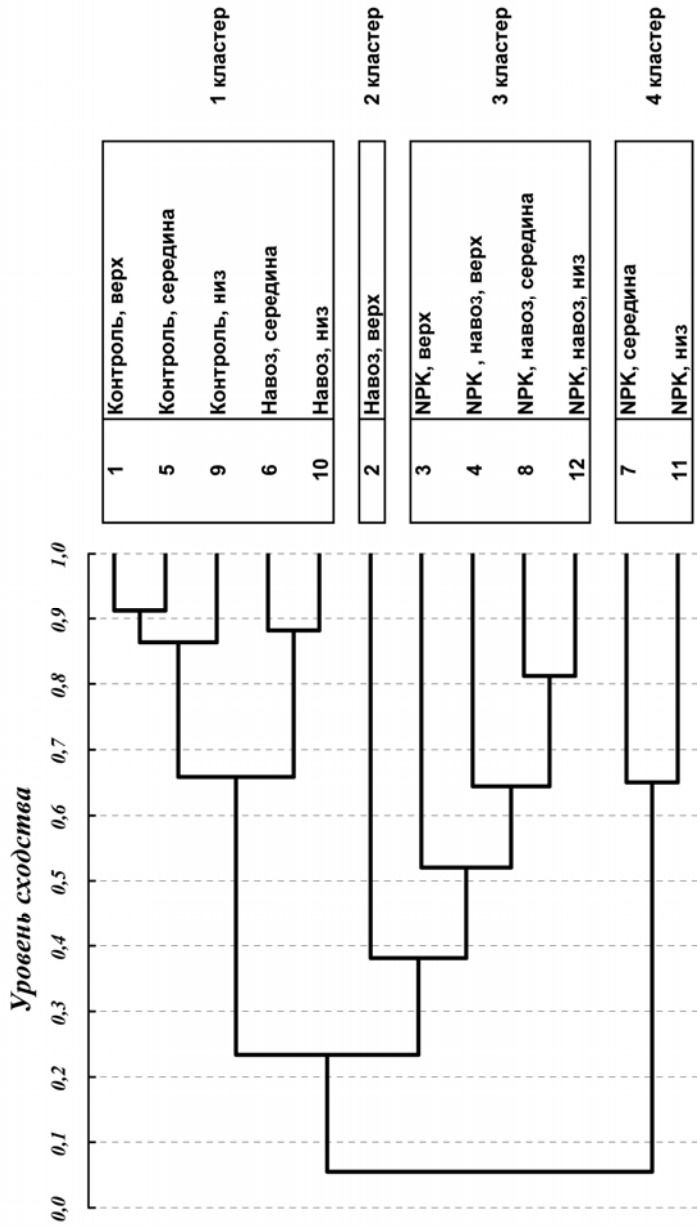


Рис. 15. Углеродный питательный ресурс микроорганизмов в вариантах опыта. Величина ресурса выражается косвенным образом через количество минерального азота в почве.

Из рис. 15 видно, что наибольший углеродный ресурс микроорганизмы используют в вариантах с применением одного NPK-удобрения. Это означает, что в этих вариантах микроорганизмами интенсивно разлагается гумусовые вещества почвы, то есть происходит его минерализация и снижение почвенного плодородия. В остальных вариантах микроорганизмами используется меньший по величине углеродный ресурс, то есть биосистема адаптируется к соответствующим экологическим условиям с меньшими энергетическими затратами.

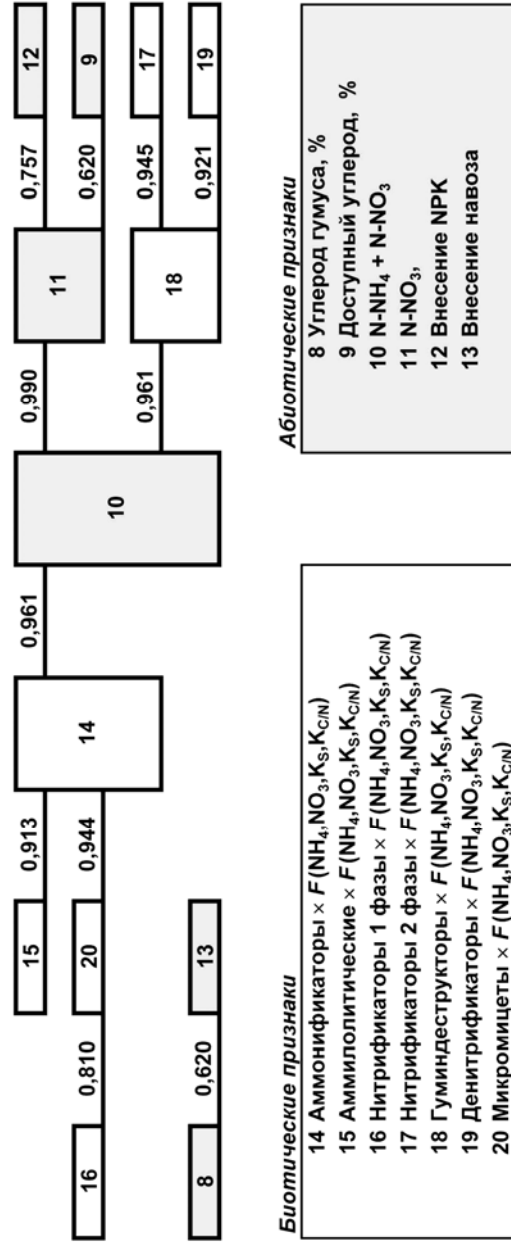
Использование признаков № 14-20 (рис. 13) вместо признаков № 1-7 (рис. 1) в кластерном и корреляционном анализе повышает информативность дендрограммы и графа максимальных коэффициентов корреляции, не внося существенных изменений в их структуры (сравните рисунки: 11 и 16, 12 и 17).

\* Правила заполнения матрицы данных (раздел 1.1, п.6).



41

Рис. 16. Группировка вариантов опыта при использовании вместо биотических признаков величин, равные произведению биотических признаков на ресурсный множитель (формула 10).



42

Рис. 17. Граф максимальных коэффициентов корреляции специальных биотических и абиотических признаков.

Число кластеров на дендрограмме (рис. 16) на единицу меньше, чем на дендрограмме (рис. 10) – 4 и 5 соответственно. Меньше и диаметр кластер-сферы соответствующих кластеров. Это является дополнительным подтверждением правильности математической модели (14). Это же подтверждает граф максимальных коэффициентов корреляции, демонстрируя более сильную корреляционную связь между специальными биотическими и абиотическими признаками (рис. 17).

### 3. КРАТКОЕ РУКОВОДСТВО ПО ИСПОЛЬЗОВАНИЮ ПРОГРАММ:

#### Clusters.xls, Graph.xls, Factor3.xls

##### 3.1. Общие правила работы с программы.

Руководство рассчитано на подготовленных пользователей вычислительной среды Microsoft Excel (версии 2000, XP). Пользователь должен уметь входить в эту среду и открывать файлы Clusters3.xls, Graph3.xls, Factor3.xls. Предполагается, что пользователь умеет вводить числовую информацию в ячейки электронной Excel-таблицы, сохранять данные и распечатывать необходимую информацию на принтере. Предлагаемые программы требуют умения работать как с клавиатурой, так и мышкой.

Прежде, чем открывать файлы, необходимо понизить уровень антивирусной защиты в Excel-среде. В настоящее время присутствие вируса в макросах Excel-программы является экзотической редкостью. Поэтому понижение уровня защиты не скажется на снижении безопасности компьютера. Если не понизить уровень защиты, то Excel отключит макросы, необходимые для проведения вычислений, и программы не смогут построить дендрограмму или граф. В крайнем случае, после выполнения необходимых расчетов в программах Clusters.xls и Graph.xls можно восстановить высокий уровень защиты Excel.

Для изменения уровня защиты необходимо открыть следующие контрольные панели среды Excel (рис. 18):

- 1) меню «Сервис»;
- 2) контрольную панель «Параметры», карточка «Безопасность»;
- 3) контрольная панель «Безопасность» (щелкнув мышкой по кнопке «Безопасность от макросов ...» на контрольной панели «Параметры»), карточка «Уровень безопасности»;
- б) щелкнуть мышкой в кружке рядом со словом «Низкая» (в кружке появиться точка).

После этого нужно вернуться назад в Excel-программу, нажимая на контрольных панелях кнопки «ОК».

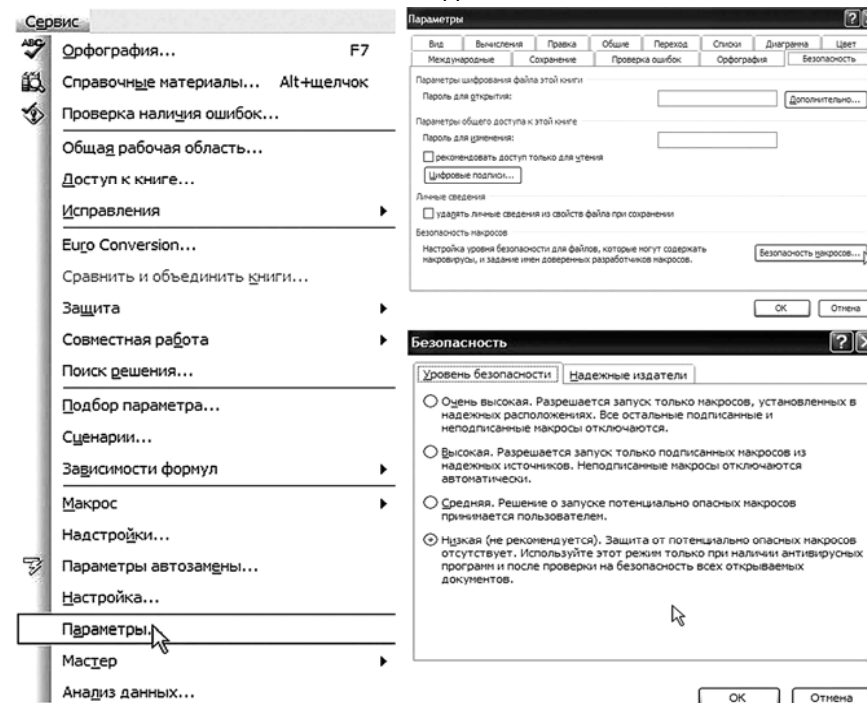


Рис. 18. Меню и контрольные панели, с помощью которых устанавливается низкий уровень защиты от макросов.

Вычислительная среда Excel допускает работу одновременно с двумя или тремя программами (Clusters3.xls, Graph3.xls, Factor3.xls). Это сокращает затраты времени на ввод матрицы данных в этих программах. Если матрица данных уже набрана в программе Clusters3.xls, то в программу Graph3.xls вводится копия матрицы. Для этого нужно выделить ячейки матрицы данных и скопировать их в буферную память Windows, перейти в другую программу, выделить в ней ячейки и вставить данные из буферной памяти Windows. На рис. 19 показано меню «Правка» с командами «Копировать» и «Вставить», с помощью которых выполняется эта операция.

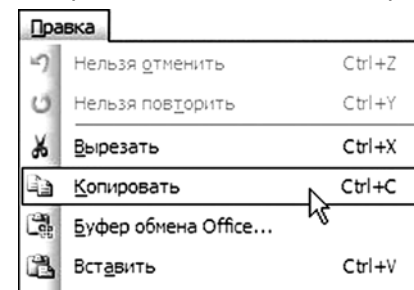


Рис. 19. Команды «Копировать» и «Вставить» в меню «Правка», которые можно использовать для переноса матрицы данных из одной программы в другую.

Чтобы избежать потери информации при работе с программами, рекомендуем периодически (через 15-20 минут) выполнять команду «Сохранить» из меню «Файл» (рис. 20).



Рис. 20. Команда «Сохранить» в меню «Файл», которую можно использовать для периодического сохранения вводимой и полученной информации.

### 3.2. Построение дендрограммы группировки вариантов с помощью программы Clusters3.xls.

После выполнения команды «Открыть» в меню «Файл» в Excel-среде загружается программа Clusters3.xls и высвечивается первый лист <Лист 1> программной книги. На переднем плане этого листа (только после первого запуска программы) выводится информационная панель со сведениями автора программы (рис. 21). Чтобы приступить к работе с программой, достаточно щелкнуть мышкой по кнопке «OK» этой панели и панель исчезнет с экрана.

Полное изображение первого листа <Лист 1> программы Clusters3.xls представлено на рис. 22. Пространство листа разделено на пять областей, соответствующих 5-ти последовательным этапам работы программы.

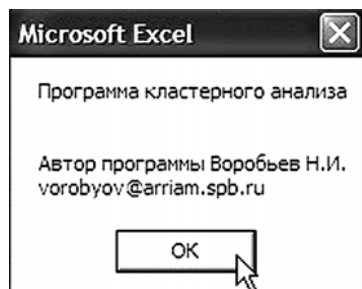


Рис. 21. Информационная панель, появляющаяся при старте программы Clusters3.xls. После щелчка по кнопке «OK» панель исчезает.

	A	B	C	D	E	F	G	H	I	J	K	L
1	<b>1. ЭТАП</b>											
2												
3	13	<<== Введите в эту ячейку число измеряемых характеристик объекта (признаков)										
4												
5	12	<<== Введите в эту ячейку число вариантов опыта										
6												
7	<b>2. ЭТАП</b>											
8												
9	Кнопка очистки листа "Данные"		<<== Нажмите эту кнопку один раз для очистки листа "Данные"! Внимание! Повторное нажатие на кнопку стирает все данные.									
10												
11	В выделенные ячейки на листе "Данные" введите числовые данные для расчета и вернитесь сюда, на "Лист 1".											
12	Внимание! Отсутствующие повторности соответствуют пустым клеткам таблицы данных.											
13												
14												
15	<b>3. ЭТАП</b>											
16												
17	Кнопка запуска вычислений "Ближний сосед"		<<== Нажмите эту кнопку для запуска вычислений по алгоритму "Ближний сосед" Внимание! Повторное нажатие на кнопку очищает лист <Матрица> и повторяет вычисления.									
18												
19												
20	Кнопка запуска вычислений "По среднему"		<<== Нажмите эту кнопку для запуска вычислений по алгоритму "По среднему" Внимание! Повторное нажатие на кнопку очищает лист <Матрица> и повторяет вычисления.									
21												
22												
23												
24	Кнопка запуска вычислений "Дальний сосед"		<<== Нажмите эту кнопку для запуска вычислений по алгоритму "Дальний сосед". Внимание! Повторное нажатие на кнопку очищает лист <Матрица> и повторяет вычисления.									
25												
26												
27	Результаты промежуточных вычислений будут находиться на листе "Результаты".											
28												
29												
30	<b>4. ЭТАП</b>											
31												
32	Кнопка запуска операции построения дендрограммы		<<== Нажмите эту кнопку для формирования дендрограммы. Внимание! Повторное нажатие на кнопку повторяет построение.									
33												
34												
35												
36	Дендрограмма находится на листе <Дендрограмма>											
37												
38	<b>5. ЭТАП</b>											
39												
40	Кнопка запуска копирования листов с результатами		<<== Нажмите эту кнопку для вызова диалоговой панели "Копирование листов"									
41												
42												
43												

Рис. 22. Первый лист <Лист 1> книги программы Clusters3.xls с изображением 5-ти этапов работы с данными и кнопками запуска вычислений.

### Порядок выполнения вычислений по программе.

#### 1-й этап

В ячейку «A3» листа <Лист 1> необходимо ввести\* число признаков матрицы данных, а в ячейку «A5» - число вариантов. На этом 1-й этап заканчивается.

#### 2-й этап

После того, как на 1-м этапе введены ненулевые числа вариантов и признаков следует щелкнуть мышкой по кнопке «Кнопка очистки листа "Данные"», которая располагается в области этого этапа (рис. 22). По команде, запускаемой этой кнопкой, произойдет очистка листа <Данные> и подготовка его к операции заполнения матрицы данных (рис. 1). Одновременно с переходом на лист <Данные> на передний план выводится информационная панель (рис. 23), которая напоминает, что

\* Необходимо вводить целые числа. Ввод числа в ячейку Excel-программы должен заканчиваться нажатием клавиши «Enter».

после ввода чисел в ячейки матрицы данных необходимо вернуться на первый лист <Лист 1> для выполнения действий следующего этапа. Щелчок мышью по кнопке «ОК» убирает эту панель с листа «Данные».

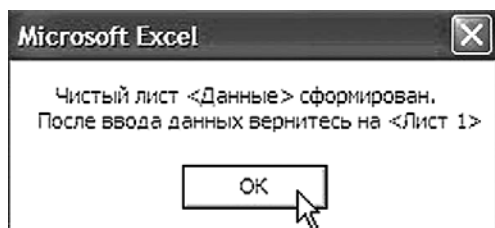


Рис. 23. Информационная панель, появляющаяся перед вводом данных в матрицу данных на листе «Данные»

Место, в котором располагаются ячейки матрицы данных, обведено прямоугольной рамкой. Сверху рамки располагается строка номеров столбцов, а слева – столбец номеров признаков. Руководствуясь этими номерами (*и правилами заполнения матрицы данных, раздел 1.1*) заполните числами эти ячейки.

После этого необходимо вернуться на первый лист программы <Лист 1>. Для этого необходимо щелкнуть мышью по ярлычку нужного листа <Лист 1>, расположенном в левом нижнем углу окна Excel-среды (рис. 24).

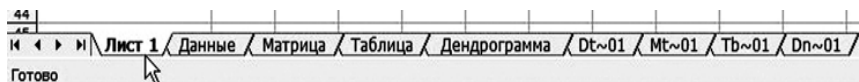


Рис. 24. Расположение ярлычков листов программы Clusters.xls в нижней части окна Excel-среды. Ярлычок листа, находящегося в данный момент на переднем плане, отличается по виду от остальных ярлычков. Щелчок мышью по ярлычку переводит соответствующий лист на передний план.

Второй этап работы с программой заканчивается щелчком мышью по ярлычку листа <Лист 1>, то есть возвратом на лист <Лист 1> (рис. 22).

### 3-й этап

На этом этапе выполняются вычисления по одному из алгоритмов построения дендрограмм: «Дальний сосед», «Среднее расстояние», «Ближний сосед». Для запуска вычислений по одному из алгоритмов необходимо щелкнуть мышью по соответствующей кнопке: «Кнопка запуска вычислений “Дальний сосед”», «Кнопка запуска вычислений “По среднему”», «Кнопка запуска вычислений “Ближний сосед”» (рис. 22). Пользователь может построить три дендрограммы и выбрать наиболее информативную среди них. Для этого нужно повторить 3,4,5 этапы, используя разные алгоритмы.

После запуска вычислений программа переходит на лист <Матрица>, формирует на этом листе треугольную матрицу нормированных Евклидовых расстояний (рис. 3) и осуществляет поиск закономерностей группировки вариантов опыта. В зависимости от быстродействия компьютера и от количества вариантов опыта на вычисления этого этапа тратится от десятков секунд до нескольких минут. В конце вычислений программа перейдет на лист <Лист 1> и на передний план выведет информационную панель (рис. 25), свидетельствующую об окончании вычислений.

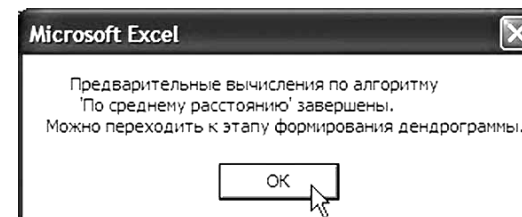


Рис. 25. Информационная панель, появляющаяся после завершения предварительных вычислений по алгоритму «По среднему расстоянию».

Необходимо щелкнуть мышью по кнопке «ОК», чтобы убрать информационную панель и перейти к вычислениям 4-го этапа.

### 4-й этап

На этом этапе запускается операция построения дендрограммы. Для запуска этой операции необходимо щелкнуть по кнопке «Кнопка запуска операции построения дендрограммы» (рис. 22). После выполнения этой операции на листе <Дендрограмма> появится изображение дендрограммы (рис. 11). Справа от дендрограммы расположен столбец номеров вариантов. Полученную дендрограмму можно распечатать на принтере или вставить в PowerPoint-презентацию, или в файл WinWord.

### 5-й этап (заключительный).

На этом этапе осуществляется сохранение листов: <Данные>, <Матрица>, <Таблица>, <Дендрограмма>. Сохранение листов сопровождается переименованием этих листов. После щелчка мышью по кнопке «Кнопка запуска копирования листов с результатами» (рис. 22) на передний план выводится диалоговая панель ввода имени копии листов (рис. 26). Если в поле «Номер копии листов» ввести число 101, то программа создаст копии листов с именами: <Dt~101> (копия листа <Данные>), <Mt~101> (копия листа <Матрица>), <Tb~101> (копия листа <Таблица>), <Dn~101> (копия листа <Дендрограмма>). Для осуществления операции копирования необходимо щелкнуть мышью по кнопке «Кнопка запуска копирования листов на диалоговой панели» (рис. 26).



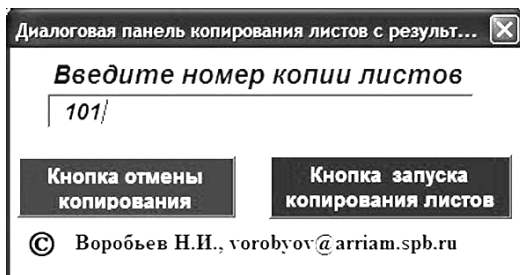


Рис. 26. Диалоговая панель для копирования листов с результатами вычислений. В поле «Номер копии листов» вводится целое число, которое входит в состав имени копируемых листов.

По завершении операции копирования в нижней части экрана Excel-программы появятся ярлычки соответствующих копий листов. Вычисления можно повторить, используя другие алгоритмы вычислений. Для этого достаточно повторить операции 3, 4 и 5 этапов. С новыми данными необходимо выполнить все этапы программы.

### 3.3. Построение графа максимальных корреляций признаков с помощью программы Graph3.xls

После выполнения команды «Открыть» в меню «Файл» в Excel-среду загружается программа Graph3.xls и высвечивается первый лист <Лист 1> программной книги. На переднем плане этого листа (только после первого запуска программы) выводится информационная панель со сведениями автора программы (рис. 27). Чтобы приступить к работе с программой, достаточно щелкнуть мышкой по кнопке «ОК» этой панели и панель исчезнет с экрана.

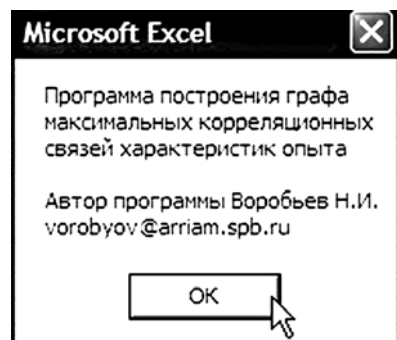


Рис. 27. Информационная панель, появляющаяся при запуске программы Graph3.xls. После щелчка по кнопке «ОК» панель исчезает.

Полное изображение первого листа <Лист 1> программы Graph3.xls представлено на рис. 28. Пространство листа разделено на пять областей, соответствующих 5-ти последовательным этапам работы программы.

	A	B	C	D	E	F	G	H	I	J	K
1	1. ЭТАП										
2											
3	13	<<== Введите в эту ячейку число измеряемых характеристик объекта (признаков)									
4											
5	12	<<== Введите в эту ячейку число вариантов опыта									
6											
7	2. ЭТАП										
8	Кнопка очистки листа "Данные"										
9	<<== Нажмите эту кнопку один раз для очистки листа "Данные"! Внимание! Повторное нажатие на кнопку стирает все данные.										
10											
11											
12											
13	3. ЭТАП										
14	Кнопка запуска предварительных вычислений										
15	<<== Нажмите эту кнопку для запуска вычислений по схеме "Дальний сосед". Внимание! Повторное нажатие на кнопку очищает лист <Матрица> и повторяет вычисления.										
16											
17											
18											
19	4. ЭТАП										
20	Кнопка запуска операции построения графа										
21	<<== Нажмите эту кнопку для формирования графа. Внимание! Повторное нажатие на кнопку повторяет построение. Граф находится на листе <Граф>										
22											
23											
24											
25	5. ЭТАП										
26	Кнопка запуска копирования листов с результатами										
27	<<== Нажмите эту кнопку для вызова диалоговой панели "Копирование листов"										
28											
29											
30											

Рис. 28. Первый лист <Лист 1> книги программы Graph3.xls с изображением 5-ти этапов работы с данными и кнопками запуска вычислений.

### Порядок выполнения вычислений по программе.

#### 1-й этап

В ячейку «A3» листа <Лист 1> необходимо ввести\* число признаков матрицы данных, а в ячейку «A5» - число вариантов. На этом 1-й этап заканчивается.

#### 2-й этап

После того, как на 1-м этапе введены ненулевые числа вариантов и признаков следует щелкнуть мышкой по кнопке «Кнопка очистки листа «Данные»», которая располагается в области этого этапа (рис. 28). По команде, запускаемой этой кнопкой, произойдет очистка листа <Данные> и подготовка его к операции заполнения матрицы данных (рис. 1). Одновременно с переходом на лист <Данные> на передний план выводится информационная панель (рис. 23), которая напоминает, что после ввода чисел в ячейки матрицы данных необходимо вернуться на первый лист <Лист 1> для выполнения действий следующего этапа. Щелчок мышью по кнопке «ОК» убирает эту панель с листа «Данные».

Место, в котором располагаются ячейки матрицы данных, обведено прямоугольной рамкой. Сверху рамки располагается строка номеров

\* Необходимо вводить целые числа. Ввод числа в ячейку Excel-программы должен заканчиваться нажатием клавиши «Enter».

столбцов, а слева – столбец номеров признаков. Руководствуясь этими номерами (*и правилами заполнения матрицы данных, раздел 1.1*) заполните числами эти ячейки.

После этого необходимо вернуться на первый лист программы <Лист 1>. Для этого необходимо щелкнуть мышью по ярлычку нужного листа <Лист 1>, расположенном в левом нижнем углу окна Excel-среды (рис. 24). Второй этап работы с программой заканчивается щелчком мышью по ярлычку листа <Лист 1>, то есть возвратом на лист <Лист 1> (рис. 28).

### **3-й этап**

На этом этапе выполняются предварительные вычисления. Для запуска вычислений необходимо щелкнуть мышью по кнопке «Кнопка запуска предварительных вычислений» (рис. 28). После запуска вычислений программа переходит на лист <Матрица>, формирует на этом листе треугольную матрицу парных коэффициентов корреляции (рис. 7) и осуществляет поиск максимальных по модулю коэффициентов корреляции. В зависимости от быстродействия компьютера и от количества вариантов опыта на вычисления этого этапа тратится от десятков секунд до нескольких минут. В конце вычислений программа перейдет на лист <Лист 1> и на передний план выведет информационную панель (рис. 29), свидетельствующую об окончании вычислений.

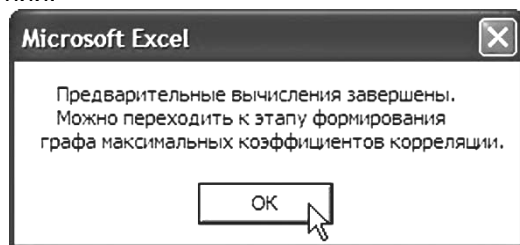


Рис. 29. Информационная панель, появляющаяся после завершения предварительных вычислений.

Необходимо щелкнуть мышкой по кнопке «OK», чтобы убрать информационную панель и перейти к вычислениям 4-го этапа.

### **4-й этап**

На этом этапе выполняются операции построения графа на листе <Граф>. Для запуска этой операции необходимо щелкнуть по кнопке «Кнопка запуска операции построения графа» (рис. 28). После выполнения этой операции на листе <Граф> появится изображение графа (рис. 11). Полученный граф можно распечатать на принтере или вставить в PowerPoint-презентацию, или в файл WinWord.

### **5-й этап (заключительный).**

На этом этапе осуществляется сохранение листов: <Данные>, <Матрица>, <Граф>. Сохранение листов сопровождается переимено-

ванием этих листов. После щелчка мышью по кнопке «Кнопка запуска копирования листов с результатами» (рис. 28) на передний план выводится диалоговая панель ввода имени копии листов (рис. 26). Если в поле «Номер копии листов» ввести число 101, то программа создаст копии листов с именами: <D~101> (копия листа <Данные>), <M~101> (копия листа <Матрица>), <G~101> (копия листа <Граф>). Для осуществления операции копирования необходимо щелкнуть мышью по кнопке «Кнопка запуска копирования листов на диалоговой панели» (рис. 26). По завершении операции копирования в нижней части экрана Excel-программы появятся ярлычки соответствующих копий листов.

### ***3.4. Построение направленного графа биотрансформации веществ и математической модели дивергенции компонентов биосистемы с помощью программы Factor3.xls.***

После выполнения команды «Открыть» в меню «Файл» в Excel-среду загружается программа Factor3.xls и высвечивается первый лист <Лист 1> программной книги. На переднем плане этого листа (только после первого запуска программы) выводится информационная панель со сведениями автора программы (рис. 30). Чтобы приступить к работе с программой, достаточно щелкнуть мышкой по кнопке «OK» этой панели и панель исчезнет с экрана.

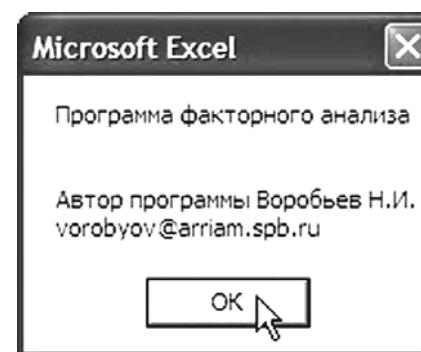


Рис. 30. Информационная панель, появляющаяся при запуске программы Factor3.xls. После щелчка по кнопке «OK» панель исчезает.

Полное изображение первого листа <Лист 1> программы Factor3.xls представлено на рис. 31. Пространство листа разделено на четыре области, соответствующие 4-м последовательным этапам работы программы.

### ***Порядок выполнения вычислений по программе.***

#### **1-й этап**

В ячейку «A3» листа <Лист 1> необходимо ввести\* число признаков матрицы данных, а в ячейку «A5» - число вариантов. На этом 1-й этап заканчивается.

	A	B	C	D	E	F	G	H	I	J	K
1	<b>1. ЭТАП</b>										
2											
3	13	<<== Введите в эту ячейку число измеряемых характеристик объекта (признаков)									
4											
5	12	<<== Введите в эту ячейку число вариантов опыта									
6											
7	<b>2. ЭТАП</b>										
8	Кнопка очистки листа "Данные"										
9	<<== Нажмите эту кнопку один раз для очистки листа "Данные"! Внимание! Повторное нажатие на кнопку стирает все данные.										
10											
11											
12	В выделенные ячейки на листе "Данные" введите числовые данные для расчета и вернитесь сюда, на "Лист 1".										
13	Внимание! Отсутствующие повторности соответствуют пустым клеткам таблицы данных.										
14											
15	<b>3. ЭТАП</b>										
16											
17	Кнопка запуска вычислений										
18	<<== Нажмите эту кнопку для запуска панели «Вычисления»										
19											
20											
21	<b>4. ЭТАП</b>										
22											
23	Кнопка запуска копирования листа "Данные"										
24	<<== Нажмите эту кнопку для вызова диалоговой панели «Копирование листа "Данные"»										
25											
26											

Рис. 31. Первый лист <Лист 1> книги программы Factor3.xls с изображением 4-х этапов работы с данными и кнопками запуска вычислений.

### 2-й этап

После того, как на 1-м этапе введены ненулевые числа вариантов и признаков следует щелкнуть мышкой по кнопке «Кнопка очистки листа «Данные»», которая располагается в области этого этапа (рис. 31). По команде, запускаемой этой кнопкой, произойдет очистка листа <Данные> и подготовка его к операции заполнения матрицы данных (рис. 1). Одновременно с переходом на лист <Данные> на передний план выводится информационная панель (рис. 23), которая напоминает, что после ввода чисел в ячейки матрицы данных необходимо вернуться на первый лист <Лист 1> для выполнения действий следующего этапа. Щелчок мышью по кнопке «ОК» убирает эту панель с листа «Данные».

Место, в котором располагаются ячейки матрицы данных, обведено прямоугольной рамкой. Сверху рамки располагается строка номеров столбцов, а слева – столбец номеров признаков. Руководствуясь этими номерами (*и правилами заполнения матрицы данных, раздел 1.1*) заполните числами эти ячейки.

После этого необходимо вернуться на первый лист программы <Лист 1>. Для этого необходимо щелкнуть мышью по ярлычку нужного

\* Необходимо вводить целые числа. Ввод числа в ячейку Excel-программы должен заканчиваться нажатием клавиши «Enter».

листа <Лист 1>, расположенном в левом нижнем углу окна Excel-среды (рис. 24). Второй этап работы с программой заканчивается щелчком мышью по ярлычку листа <Лист 1>, то есть возвратом на лист <Лист 1> (рис. 31).

### 3-й этап

На этом этапе выполняются вычисления факторных нагрузок *Min*-фактора в пространстве признаков, а также коэффициентов био-трансформации и средних признаков, которые используются в математической модели дивергенции компонентов биосистемы.

Для запуска вычислений необходимо щелкнуть мышью по кнопке «Кнопка запуска вычислений» (рис. 31). За щелчком по этой кнопке последует вывод на передний план диалоговой панели «Вычисления» (рис. 32). В верхнее поле необходимо ввести номер признака-функции, а в нижнее – номера признаков-аргументов, разделяя их запятыми.

Рис. 32. Диалоговая панель «Вычисления» с полями ввода номеров признака-функции и признаков-аргументов.

После ввода необходимой информации необходимо щелкнуть мышью по кнопке «Кнопка запуска вычислений» на диалоговой панели «Вычисления». За этими действиями последует переход на лист <Матрица>, на котором программа будет некоторое время проводить необходимые вычисления. В зависимости от быстродействия компьютера и от количества признаков-аргументов на вычисления этого этапа может быть потрачено от одной до нескольких десятков минут.

№	Признаки	Средние признаков	Девиацы признаков	Факторные нагрузки	Кэффициенты биотрансформации	Признак -> функция
8	Углерод думуса, %	1,08	0,302	0,2047	1	Признак -> функция
1	Аммонификаторы	17,14	16,785	-0,3385	0,4712	Признак -> сингез-аргумент
2	Аммилолитические	13,40	19,039	-0,0364	0,0349	
3	Нитрификаторы 1 фазы	27,68	41,693	-0,0006	0,0005	
4	Нитрификаторы 2 фазы	29,48	46,982	-0,1570	0,1342	
5	Гуминдеструкторы	14,41	14,178	0,7983	-1,1058	Признак -> аргумент-деструкция
6	Денитрификаторы	0,55	1,803	-0,4204	0,1748	
7	Микромицеты	33,17	42,891	-0,0585	0,0617	
	Дисперсия Min-фактора	0,0874				
	Дисперсия Max-фактора	2,41404				
	Фактический критерий Фишера	27,6198				
	Теоретический критерий Фишера (p=0.05)	2,81793				

Рис. 33. Таблица средних и девиат признаков, факторных нагрузок и коэффициентов биотрансформации для Min-фактора., созданная программой Factor3.xls.

В конце вычислений программа перейдет на лист <Данные>, на котором ниже исходных данных будет сформирована таблица со значениями средних признаков, факторных нагрузок *Min*-фактора и коэффициентами биотрансформации (рис. 33). Ниже таблицы выводятся значения дисперсий *Min*-фактора и *Max*-фактора, их отношения и стандартный критерий Фишера для этого отношения.

Третий этап работы с программой заканчивается щелчком мышью по ярлычку листа <Лист 1>, то есть возвратом на лист <Лист 1> (рис. 31).

#### 4-й этап (заключительный).

На этом этапе осуществляется сохранение листа <Данные>. Сохранение этого листа сопровождается его переименованием. После щелчка мышью по кнопке «Кнопка запуска копирования листа 'Данные'» (рис. 31) на передний план выводится диалоговая панель ввода имени копии этого листа (рис. 26). Если в поле «Номер копии листов» ввести число 101, то программа создаст копии листов с именами: <D~101> (копия листа <Данные>). Для осуществления операции копирования необходимо щелкнуть мышью по кнопке «Кнопка запуска копирования листов на диалоговой панели (рис. 26). По завершении операции копирования в нижней части экрана Excel-программы появятся ярлычки соответствующей копии листа.

## ЗАКЛЮЧЕНИЕ

Граф-анализ наиболее эффективен на начальных этапах исследования биосистем. Результаты анализа используются для минимизации числа факторов, воздействующих на биосистему, и для минимизации числа контролируемых характеристик. Привлечение математических моделей компонентов биосистемы значительно расширяет возможности метода, позволяя изучать поведение биосистем в целом и с точки зрения каждого частного процесса в ней.

## СПИСОК МЕТОДИЧЕСКОЙ ЛИТЕРАТУРЫ ПО ГРАФ-АНАЛИЗУ

1. Дюран Б., Оделл П. Кластерный анализ. - М.: Статистика, 1977.- 128 с.
2. Классификация и кластер /Под ред. Дж. Вэн Райзина. – М.: Мир, 1980.- 390 с.
3. Миркин Б.Г. Анализ качественных признаков и структур. – М.: Статистика., 1980. – 320 с.
4. Берж К. Теория графов и ее применение. – М.: Изд-во иностр. лит., 1962. – 319 с.
5. Оре О. Теория графов. – М.: Наука, 1968. – 324 с.
6. Смирнов Е.С. Таксономический анализ. – М.: Изд-во МГУ, 1969. –187 с.
7. Дубров А.М., Мхитарян В.С., Трошин Л.И. Многомерные статистические методы: Учебник. – М.: Финансы и статистика, 1998.- 352 с.

## ПУБЛИКАЦИИ, В КОТОРЫХ ПРИМЕНЯЛСЯ ГРАФ-АНАЛИЗ

8. Возняковская Ю.М., Попова Ж.П., Воробьев Н.И. Характеристика микробиологических показателей, используемых при определении уровня эффективного плодородия почвы. Сельскохозяйственная микробиология, № 5, 1994, с. 84-90.
9. N.Novikova, E.Pavlova, N.Vorobyov, E.Lemeshenko. Numerical Taxonomy of Rhizobium Strains from Legumes of the Temperate Zone. (1994), International Journal of Systematic Bacteriology, v. 44, N 4, pp. 734-742.
10. V.N.Pishchik, S.V.Tymofeyeva, A.A.Glavin, N.I.Vorobyov, and B.L.Rubenchik. Effect of N-Nitrosoamines and Nitrosationable Pesticide-Tetramethylthiuram Disulfide on Soil Microbiocenosis and Potato Yield: A Model System. Journal of Environmental Pathology, Toxicology and Oncology. N 13(4), 1994, p. 235-242.
11. A.A.Belimov, A.M.Kunakova, E.V.Gruzdeva, N.I.Vorobyev, A.P.Kozhemiakov, O.F.Khamova, S.M.Postavskaya, S.M.Sokova. Relationship between survival rates of associative nitrogen fixers on roots and yield response of plants to inoculation. FEMS Microbiology Ecology 17, (1995), p. 187-196.
12. В.Н.Пищик, С.В.Тимофеева, А.А.Главин, Н.И.Воробьев, Б.Л.Рубенчик, Л.А.Тиктин, И.И.Черняева. Влияние некоторых N-нитрозаминов на формирование почвенного микробоценоза в модельной системе. Экспериментальная онкология, № 17, 1995, с. 192-198.
13. В.Н.Пищик, И.И.Черняева, Н.И.Воробьев, А.М.Лазарев. О свойствах штаммов *Erwinia carotovora*, различных по вирулентности. Микробиология, 1996, том 65, № 2, с. 262-268.
14. И.Ю.Кирцидели, Н.И.Воробьев, О.М.Терешенков. Сообщества микромицетов из почв подзоны типичных тундр в районе северного побережья таймырского озера. Микология и фитопатология, т. 30, вып. 2, 1996, с.9-14.
15. Кутузова Р.С., Трибис Ж.М., Воробьев Н.И. Автотрофное окисление аммония как показатель нетто-минерализации азота в почве. - Почвоведение, 1997, № 10, стр. 1222-1229.

16. Р.С.Кутузова, М.В.Михалева, В.Ф.Смирнова, Н.И.Воробьев. Микробиологические аспекты применения медленнодействующего карбамидформальдегидного удобрения. Агрохимия, 1997, № 11, с. 43-51.
17. Белимов А.А., Иванчиков А.Ю. Воробьев Н.И. Роль доминирующей микрофлоры ризопланы ячменя во взаимодействии интродуцируемых диазотрофов с растением. Микробиология, 1998, т. 67, № 3, с. 409-415.
18. V.N.Pishchik, I.V.Mokrousov, A.M.Lazarev, N.I.Vorobyev, O.V.Narvskaya, I.I.Chernyaeva, A.P.Kozhemyakov and G.N.Koval. Biological properties of same nitrogen-fixing associative enterobacteria. Plant and Soil, 202, 1998, 49-59.
19. Воробьев Н.И. Алгоритм построения графа стохастических связей по матрице парных корреляций. /Ученые записки. Серия "Математика и информатика". Том 1. – СПб.: ЛГОУ. 1998, с. 19-26.
20. Г.В.Калько, Т.М.Лагутина, И.И.Новикова, Н.И.Воробьев. Кластерный анализ влияния микробов-антагонистов на динамику популяционной плотности фитопатогенного гриба *Fusarium oxysporum* на мембранных фильтрах в торфогрунте. Микология и фитопатология, т. 34, вып. 6, 2000, с. 71-77.
21. Калько Г.В., Лагутина Т.М., Воробьев Н.И.,Новикова И.И. Ингибирование микробами-антагонистами фитопатогенного гриба *Fusarium oxysporum* в торфогрунте. Микология и фитопатология, т. 35, 2001, вып. 3, с. 66-75.
22. Кутузова Р.С., Сирота Л.Б., Орлова О.В., Воробьев Н.И. Использование математического анализа для оценки микробиологического состояния почв агроландшафта и опыта. Агрохимия. 2001, № 1, с. 19-33.
23. Свиридова О.В., Михалева Л.В., Воробьев Н.И., Кочетков В.В. Разложение коры хвойных деревьев грибами и бактериями. Микология и фитопатология. 2001 т. 35, вып. 6, с.38-47.
24. Кутузова Р.С., Сирота Л.Б., Орлова О.В., Воробьев Н.И. Микробное сообщество и анализ почвенно-микробиологических процессов в дерново-подзолистой почве. Почвоведение, 2001, № 3, с 320-332.
25. Воробьев Н.И. Свиридова О.В. Граф биотрансформации С и N соединенной системой почвенных микроорганизмов. Экология и почвы. Избр. лекции 10-й Всероссийской школы. Том IV. Пущино. ОНТИ ПНЦ РАН, 1-5 октября 2001 г., с. 289-297.
26. Р.С.Кутузова, Л.Б.Сирота, О.В.Орлова, Н.И.Воробьев. Особенности почвенно-микробиологических процессов при внесении бамила в дерново-подзолистую почву. Агрохимия, 2002, № 5, с. 22-32.
27. Р.С.Кутузова, Н.И.Воробьев, Ю.В.Круглов. Структура микробного комплекса ризосферы пшеницы в условиях гербицидного стресса. Почвоведение, 2006, № 2, с. 220-227.

<b>ВВЕДЕНИЕ</b> .....	<b>3</b>
<b>1. БАЗОВЫЕ ПОНЯТИЯ И ТЕРМИНЫ ГРАФ-АНАЛИЗА</b> .....	<b>4</b>
1.1. ПРЯМОУГОЛЬНАЯ МАТРИЦА ДАННЫХ.....	4
1.2. НОРМА ПРИЗНАКА. КЛАСТЕР-НОРМАЛИЗАЦИЯ МАТРИЦЫ ДАННЫХ.....	7
1.3. НОРМИРОВАННОЕ ЕВКЛИДОВО РАССТОЯНИЕ – МЕРА БЛИЗОСТИ ВАРИАНТОВ. ...	8
1.4. УРОВЕНЬ БЛИЗОСТИ (СХОДСТВА) ВАРИАНТОВ В ГРУППЕ. ....	12
1.5. ДЕВИАТА ЗНАЧЕНИЙ ПРИЗНАКА – МЕРА ВАРИАБЕЛЬНОСТИ ПРИЗНАКА. ФАКТОР-НОРМАЛИЗАЦИЯ ИСХОДНЫХ ДАННЫХ. ....	13
1.6. КОЭФФИЦИЕНТ КОРРЕЛЯЦИИ ПРИЗНАКОВ – МЕРА ВЗАИМОЗАВИСИМОСТИ ПРИЗНАКОВ. ....	15
1.7. ЛИНИИ <i>Min</i> -ФАКТОРА И <i>Max</i> -ФАКТОРА В ПРОСТРАНСТВЕ ПРИЗНАКОВ. ....	17
<b>2. ОСНОВНЫЕ ВЫХОДНЫЕ ИНФОРМАЦИОННЫЕ ФОРМЫ ГРАФ-АНАЛИЗА</b> . ....	<b>21</b>
2.1. ДЕРЕВО ГРУППИРОВКИ ВАРИАНТОВ (ДЕНДРОГРАММА). ....	21
2.2. ГРАФ МАКСИМАЛЬНЫХ КОЭФФИЦИЕНТОВ КОРРЕЛЯЦИИ ПРИЗНАКОВ. ....	26
2.3. НАПРАВЛЕННЫЙ ГРАФ БИОТРАНСФОРМАЦИИ ВЕЩЕСТВ.....	29
2.4. МАТЕМАТИЧЕСКАЯ МОДЕЛЬ ДИВЕРГЕНЦИИ КОМПОНЕНТОВ БИОСИСТЕМЫ. ....	35
<b>3. КРАТКОЕ РУКОВОДСТВО ПО ИСПОЛЬЗОВАНИЮ ПРОГРАММ: CLUSTERS.XLS, GRAPH.XLS, FACTOR3.XLS</b> .....	<b>43</b>
3.1. ОБЩИЕ ПРАВИЛА РАБОТЫ С ПРОГРАММАМИ.....	43
3.2. ПОСТРОЕНИЕ ДЕНДРОГРАММЫ ГРУППИРОВКИ ВАРИАНТОВ С ПОМОЩЬЮ ПРОГРАММЫ CLUSTERS3.XLS. ....	45
3.3. ПОСТРОЕНИЕ ГРАФА МАКСИМАЛЬНЫХ КОРРЕЛЯЦИЙ ПРИЗНАКОВ С ПОМОЩЬЮ ПРОГРАММЫ GRAPH3.XLS.....	49
3.4. ПОСТРОЕНИЕ НАПРАВЛЕННОГО ГРАФА БИОТРАНСФОРМАЦИИ ВЕЩЕСТВ И МАТЕМАТИЧЕСКОЙ МОДЕЛИ ДИВЕРГЕНЦИИ КОМПОНЕНТОВ БИОСИСТЕМЫ С ПОМОЩЬЮ ПРОГРАММЫ FACTOR3.XLS. ....	52
<b>ЗАКЛЮЧЕНИЕ</b> .....	<b>56</b>
<b>СПИСОК МЕТОДИЧЕСКОЙ ЛИТЕРАТУРЫ ПО ГРАФ-АНАЛИЗУ</b> .....	<b>57</b>
<b>ПУБЛИКАЦИИ, В КОТОРЫХ ПРИМЕНЯЛСЯ ГРАФ-АНАЛИЗ</b> .....	<b>57</b>