

КЛАСТЕРНЫЙ АНАЛИЗ СТАТИСТИЧЕСКИХ ПРОФИЛЕЙ ПРОСТРАНСТВЕННОГО РАСПРЕДЕЛЕНИЯ ГУМУСА

Н.И. Воробьев, к.т.н.¹, С.С. Ладан, к.б.н.²

¹*ФГБНУ ВНИИ сельскохозяйственной микробиологии, Санкт-Петербург
Россия, 196608, Санкт-Петербург, Пушкин, шоссе Подбельского, д. 3,
E-mail: nik.ivanvorobyov@yandex.ru*

²*ФГБНУ Всероссийский научно-исследовательский
институт агрохимии им. Д.Н. Прянишникова, Москва*

Показатели пространственного распределения содержания гумуса в почвах федеральных округов РФ образуют индивидуальные статистические профили регионов. Для изучения причин, деформирующих эти профили, был проведен кластерный анализ содержания гумуса в пашнях и залежах регионов РФ. Значения медианы и доверительного интервала профилей гумуса использовались для определения статистической дистанции полей регионов РФ и построения кластерных дендрограмм. Кластерный анализ показал, что регионы с близкими климатическими условиями, имеющие закономерно сходное распределение гумуса в почвах залежей, вследствие антропогенного давления характеризуются различными гумусовыми профилями пашен. Величина и направленность деформаций статистического профиля могут служить основанием для оценки рациональности влияния сельскохозяйственной деятельности на почвообразовательные процессы.

Ключевые слова: статистический профиль пространственного распределения гумуса в почве, почвы залежей и пашен, дендрограмма группировки полей регионов РФ.

Для цитирования: Воробьев Н.И., Ладан С.С. Кластерный анализ статистических профилей пространственного распределения гумуса// Плодородие. – 2021. – №5. – С. 33-36. DOI: 10.25680/S19948603.2021.122.09.

Колебания агрохимических характеристик в пределах одного и того же участка, хозяйства или региона при проведении любых почвенных исследований требуют комплекса статистических приемов по выявлению и установлению достоверности их изменений. Содержание гумуса – одна из базовых характеристик земель сельскохозяйственного назначения и определяется при любых почвенных исследованиях. Широкая представленность и информативность данного показателя сделали его релевантным фактором при проведении данного статистического исследования. Задачами исследования были проведение кластерного анализа статистических профилей пространственного распределения (СППР) гумуса в почвах ФО РФ и построение дендрограмм группировки по-

лей регионов РФ по СППР гумуса. Кластерный анализ является одним из алгоритмов аналитической системы искусственного интеллекта. Дендрограммы группировки регионов по агрохимическим характеристикам почв – это достоверные цифровые образы, которые могут использоваться для поиска в нейросетях системы искусственного интеллекта агротехнологий, обеспечивающих максимальную продуктивность сельскохозяйственных культур на полях ФО РФ.

Для получения СППР гумуса проводят агрохимический анализ почв (по несколько проб с каждого поля). В результате получают информацию в виде площадей, на которых содержание гумуса соответствует одному из 10 классов содержания гумуса (табл. 1).

1. Стохастические профили распределения (СПР) гумуса по площадям (% от общей площади залежей и пашен в регионах РФ) и по 10-ти классам содержания гумуса в почве (весовые %)

Федеральный округ РФ	Классы содержания гумуса в почве									
	<1	1,01-1,5	1,51-2,0	2,01-2,5	2,51-3,0	3,01-4,0	4,01-6,0	6,01-8,0	8,01-10,0	>10,0
	1	2	3	4	5	6	7	8	9	10
<i>Залежь</i>										
Центральный	0,00	9,13	24,47	21,09	14,53	14,02	16,27	0,33	0,16	0,00
Северо-Западный	2,70	8,26	10,93	16,21	13,41	36,96	11,52	0,00	0,00	0,00
Южный	0,00	19,85	14,42	7,07	14,57	44,09	0,00	0,00	0,00	0,00
Приволжский	0,00	0,00	0,00	0,00	0,43	0,52	19,13	48,42	31,13	0,37
Уральский	0,06	0,03	0,58	0,22	8,77	8,91	67,81	12,94	0,42	0,24
Сибирский	1,07	16,68	17,87	8,96	11,95	15,80	25,21	2,45	0,00	0,00
Дальневосточный	0,98	1,06	0,96	5,84	17,26	18,25	47,10	6,11	1,55	0,89
<i>Пашня</i>										
Центральный	0,25	4,19	9,39	10,38	9,79	11,75	38,22	15,64	0,28	0,11
Северо-Западный	0,83	3,21	10,17	21,96	18,05	24,12	15,41	5,03	0,71	0,53
Южный	2,98	5,44	10,91	8,72	11,48	35,97	21,13	3,37	0,00	0,00
Приволжский	1,35	5,14	9,85	9,13	8,94	16,17	26,72	16,52	5,32	0,86
Уральский	0,08	0,16	0,91	1,23	3,31	9,92	53,46	24,20	5,54	1,18
Сибирский	0,25	2,24	3,74	3,63	4,78	15,09	31,31	27,26	9,34	2,37
Дальневосточный	0,00	0,02	5,88	5,41	8,44	27,13	36,54	10,19	4,10	2,29

СППР гумуса характеризуется накопительной функцией распределения гумуса по классам, которая представляет собой зависимость площади полей от объеди-

ненных классов содержания гумуса (см. табл. 2). Например, объединенный 4-й класс СППР гумуса характеризует сумму площадей с содержанием гумуса в поч-

вах, соответствующих 1, 2, 3, 4 классам (см. табл. 1). Поэтому, для накопительной функции СППР гумуса в залежах Центрального ФО РФ относительное количество площадей залежей, соответствующих 4-му объединенному классу, равняется 0% (1-й класс, табл.1) +

9,13% (2-й класс, табл.1) + 24,47% (3-й класс, см. табл.1) + 21,09% (4-й класс, табл.1) = 54,69 % (табл. 2).

На рисунке 1 представлены две накопительные функции СППР гумуса для залежей Центрального и Приволжского ФО РФ, построенные в соответствии с данными таблицы 2.

2. Накопительные функции СППР гумуса по площадям (% от общей площади залежей и пашен в регионах РФ) и по 10-ти объединенным классам содержания гумуса в почве (весовые %)

Федеральный округ	Объединенные классы содержания гумуса в почве									
	<1	<1,5	<2,0	<2,5	<3,0	<4,0	<6,0	<8,0	<10,0	Любое
	1	1∪2	1∪2∪3	1∪...∪4	1∪...∪5	1∪...∪6	1∪...∪7	1∪...∪8	1∪...∪9	1∪...∪10
<i>Залежь</i>										
Центральный	0,00	9,13	33,61	54,69	69,23	83,24	99,51	99,84	100	100
Северо-Западный	2,70	10,97	21,90	38,11	51,52	88,48	100	100	100	100
Южный	0,00	19,85	34,27	41,34	55,91	100	100	100	100	100
Приволжский	0,00	0,00	0,00	0,00	0,43	0,95	20,09	68,51	99,63	100
Уральский	0,06	0,10	0,67	0,90	9,67	18,59	86,40	99,34	99,76	100
Сибирский	1,07	17,75	35,62	44,58	56,53	72,33	97,54	100	100	100
Дальневосточный	0,98	2,05	3,01	8,85	26,10	44,35	91,45	97,56	99,11	100
<i>Пашня</i>										
Центральный	0,25	4,44	13,82	24,20	34,00	45,75	83,97	99,61	99,89	100
Северо-Западный	0,83	4,03	14,20	36,16	54,20	78,32	93,73	98,76	99,47	100
Южный	2,98	8,41	19,32	28,05	39,53	75,50	96,63	100	100	100
Приволжский	1,35	6,49	16,33	25,46	34,41	50,58	77,30	93,82	99,14	100
Уральский	0,08	0,23	1,15	2,38	5,69	15,62	69,08	93,28	98,82	100
Сибирский	0,25	2,49	6,23	9,87	14,65	29,74	61,04	88,30	97,63	100
Дальневосточный	0,00	0,02	5,91	11,32	19,76	46,88	83,42	93,61	97,71	100

Накопительные функции СППР гумуса (см. рис. 1) можно характеризовать медианой (M_i) и доверительным интервалом (V_i), где i – порядковый номер ФО РФ (табл. 3). Причем, значения M_i и V_i будут представлены безразмерными десятичными числами в диапазоне 1,0 – 10,0. При определении M_i необходимо найти такое значение на оси абсцисс (см. рис. 1), при котором накопительная функция будет равна 50%. При определении V_i необходимо найти разницу двух значений на оси абсцисс (см. рис. 1), при которых накопительная функция будет равна 5 и 95% (см. пунктирные стрелки на рис. 1). Для определения статистической дистанции между СППР гумуса двух различных ОФ РФ необходимо на координатной плоскости найти Эвклидово расстояние между двумя точками, изображающими СППР гумуса в почвах сравниваемых ФО РФ (рис. 2).

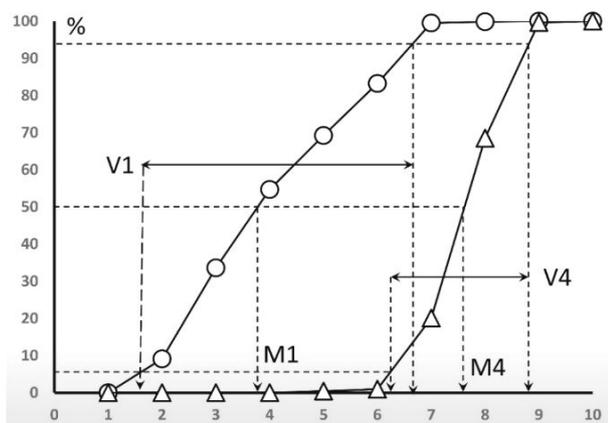


Рис. 1. Накопительные функции СППР гумуса в залежах Центрального (кружки) и Приволжского (треугольники) ФО РФ. M_1 и M_4 – безразмерные десятичные числа в диапазоне 1,0 – 10,0 – медианы СППР гумуса, связанные с 50%-ным значением накопительной функции СППР гумуса; V_1 и V_4 – безразмерные десятичные числа в диапазоне 1,0 – 10,0 – доверительные интервалы СППР гумуса, связанные с 5- и 95%-ными значениями накопительной функции СППР гумуса.

3. Медианы (M_i) и доверительные интервалы (V_i) накопительных функций СППР гумуса в залежах и пашнях ФО РФ, безразмерные десятичные числа в диапазоне 1,0 – 10,0

№ i	Федеральный округ	Залежь		Пашня	
		Медиана, M_i	Доверительный интервал, V_i	Медиана, M_i	Доверительный интервал, V_i
1	Центральный	3,78	5,2	6,11	5,6
2	Северо-Западный	4,89	5,3	4,77	5,2
3	Южный	4,59	4,6	5,29	5,6
4	Приволжский	7,62	2,6	5,96	6,5
5	Уральский	6,46	3,2	6,64	3,5
6	Сибирский	4,45	5,7	6,65	6,0
7	Дальневосточный	6,12	4,2	6,09	5,5

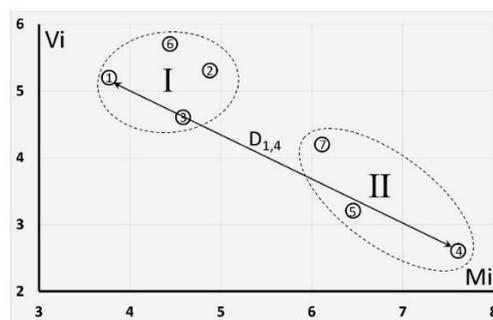


Рис. 2. Расположение на координатной плоскости точек, отображающих M_i и V_i СППР гумуса в залежах для каждого ФО РФ. Номера в точках соответствуют порядковым номерам ФО РФ в таблице 3. M_i – координата, равная медиане СППР гумуса, а V_i – координата, равная доверительному интервалу СППР гумуса. $D_{1,4}=4,64$ – максимальное Эвклидово расстояние между СППР гумуса для (1) Центрального ФО и (4) Приволжского ФО. Эллипс-образы представляют кластеры I и II, в которые объединились ФО РФ №1, 2, 3, 6 и №4, 5, 7 по СППР гумуса в почве.

Например, между точкой №1, изображающей СППР гумуса Центрального ФО РФ, и точкой №4, изображающей СППР гумуса Приволжского ФО РФ, наблюдается наибольшее Эвклидово расстояние ($D_{1,4}$), вычисляемое по формуле

$$D_{1,4} = \sqrt{(M1-M4)^2 + (V1-V4)^2} = \sqrt{(3,78-7,62)^2 + (5,2-2,6)^2} = 4,64$$

где M1 и M4 – медианы; V1 и V4 – доверительные интервалы СПР гумуса в залежах Центрального и Приволжского ФО РФ (см. табл. 3).

	1	2	3	4	5	6	7
1	0	1,11	1,01	4,64	3,34	0,84	2,54
2	1,11	0	0,76	3,84	2,62	0,60	1,65
3	1,01	0,76	0	3,63	2,34	1,11	1,58
4	4,64	3,84	3,63	0	1,31	4,43	2,19
5	3,34	2,62	2,34	1,31	0	3,21	1,06
6	0,84	0,60	1,11	4,43	3,21	0	2,24
7	2,54	1,65	1,58	2,19	1,06	2,24	0

	1	2	3	4	5	6	7
1	0	1,40	0,82	0,91	2,17	0,67	0,10
2	1,40	0	0,66	1,76	2,53	2,04	1,35
3	0,82	0,66	0	1,12	2,50	1,42	0,81
4	0,91	1,76	1,12	0	3,07	0,85	1,01
5	2,17	2,53	2,50	3,07	0	2,50	2,07
6	0,67	2,04	1,42	0,85	2,50	0	0,75
7	0,10	1,35	0,81	1,01	2,07	0,75	0

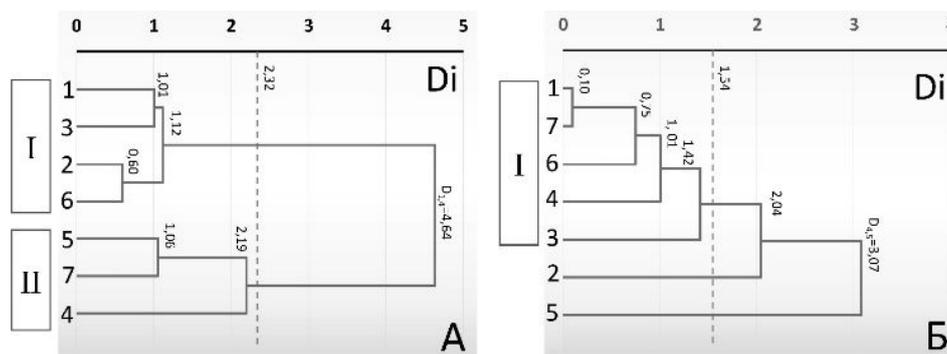


Рис. 3. Дендрогаммы кластеризации полей ФО РФ по СПР гумуса в залежах (А) и пашнях (Б). Номера в основаниях дендрогаммы соответствуют номерам ФО РФ из таблицы 3. А: в дендрогамму кластера I входят ФО РФ № 1, 2, 3, 6, а в дендро-образ кластера II – ФО РФ № 4, 5, 7. Б: в дендрогамму кластера I входят ФО РФ № 1, 3, 4, 6, 7, а ФО РФ № 2, 5 не сформировали дендрогамму кластера и не вошли в кластер I. Пунктирные линии ограничивают максимальные размеры кластеров.

Выводы. Дендрогамма кластеризации (рис. 3А) демонстрирует объединение в кластер I залежей по СПР гумуса ФО РФ №1 (Центральный), 2 (Северо-Западный), 3 (Южный), 6 (Сибирский) и объединение ФО РФ №4 (Приволжский), 5 (Уральский), 7 (Дальневосточный) в кластер II. Наиболее близкие СПР гумуса наблюдаются в залежах Северо-Западного и Сибирского ФО РФ ($D_{2,6}=0,60$). По всей видимости, близость СПР гумуса на этих территориях и объединение этих территорий в кластер I являются следствием сходных эколого-климатических условий в данных регионах (ежегодное количество осадков, средних температур и солнечной освещенности).

Дендрогамма кластеризации (рис. 3Б) демонстрирует объединение в кластер I пашен по СПВ гумуса ФО РФ № 1 (Центральный), 3 (Южный), 4 (Приволжский), 6 (Сибирский), 7 (Дальневосточный) в кластер I, а ФО РФ № 2 (Северо-Западный) и 5 (Уральский) не сформировали кластер и не вошли в кластер I. По всей видимости, именно хозяйственная деятельность на пашнях этих ФО РФ привела к тому, что стали существенно различаться по СПР гумуса дендрогаммы на рисун-

Используя приведенную формулу, можно заполнить цифрами все ячейки матрицы дистанций (D_i) СППР гумуса в залежах (1) и пашнях (2) ФО РФ.

Номера строк и столбцов матриц дистанций соответствуют номерам ФО РФ, представленным в таблице 3.

Кластерный анализ предполагает использование матриц дистанций (2), (3) и алгоритм кластеризации «дальний сосед» для построения дендрогаммы группировки полей регионов РФ (рис. 3). В результате на рисунке 3 получают дендрогаммы группировки полей регионов РФ, демонстрирующие группировку залежей и пашен ФО РФ по признаку различия СППР гумуса в их почвах. Максимальные размеры кластеров определяются половиной дистанции между наиболее удаленными ФО РФ по СППР признаку: $D_i < D_{1,4}/2=2,32$ (залежи) и $D_i < D_{4,5}/2=1,54$ (пашни). На рисунке 2 представлены эллипс-образы тех же кластеров I и II, что и на рисунке 3А.

При увеличении размерности пространства кластеризации (добавляя, например, еще сравнение по СППР рН) невозможно изобразить в четырехмерном пространстве кластеризации эллипс-образы кластеров. При многомерном кластерном анализе строятся только дендрогаммы группировки.

ках 3А и 3Б. Возможно, близкие климатические условия и зональные агротехнологии на территориях Северо-Западного и Уральского ФО РФ привели к особой динамике накопления гумуса в пашнях этих регионов.

Результаты применения кластерного анализа многомерных данных распределения агрохимических почвенных характеристик могут использоваться цифровыми системами искусственного интеллекта для планирования эффективного агрохимического управления продуктивностью различных сельскохозяйственных культур.

Литература

1. Воробьев Н.И., Свиридова О.В., Кутузова Р.С. Методические рекомендации по использованию граф-анализа в исследованиях систем, состоящих из биотических и абиотических компонентов (2-е издание). – СПб.: ГНУ ВНИИ сельскохозяйственной микробиологии, 2006. – 58 с.
2. Ганичева А.В., Ганичев А.В. Нечеткий кластерный анализ почв. В сб.: I Никитинские чтения "Актуальные проблемы почвоведения, агрохимии и экологии в природных и антропогенных ландшафтах". Материалы Международной научной конференции. Федеральное государственное бюджетное образовательное учреждение высшего образования «Пермский государственный аграрно-технологический университет имени академика Д.Н. Прянишникова», 2020. – С. 337-341.

3. Козлова А.А., Приставка А.А. Применение методов статистического анализа при изучении особенностей формирования и функционирования почв южного Предбайкалья. В сб.: Почва как связующее звено функционирования природных и антропогенно-преобразованных экосистем. Материалы V Международной научно-практической конференции, посвященной 90-летию кафедры почвоведения и оценки земельных ресурсов ИГУ и Дню Байкала. – Иркутск, 2021. – С. 119-123.
4. Мандель И.Д. Кластерный анализ. – М.: Финансы и статистика, 1988. – 176 с.
5. Матвеев Д.А., Воронаев В.В., Якушев В.В., Блохин Ю.И. и др. Состояние и перспективы создания новых методов количественной оценки внутриполевой изменчивости в точном земледелии// Агрофизика. – 2020. – № 1. – С. 59-70.
6. Мыслыва Т.Н., Кожеко А.В., Куцаева О.А. Особенности установления зон пространственной неоднородности в пределах землепользования для целей точного земледелия// Вестник Белорусской государственной сельскохозяйственной академии. – 2021. – № 1. – С. 192-199.
7. Мыслыва Т.Н., Куцаева О.А. Геостатистический анализ при оценке пространственного распределения агрохимических свойств почв земель сельскохозяйственного назначения. В сб.: Землеустройство, геодезия и кадастр: прошлое – настоящее – будущее. Сборник научных статей по материалам Международной научно-практической конференции, посвященной 95-летию землеустроительного факультета, 2020. – С. 204-213.
8. Скачкова С.А., Харитонова А.Е. Кластерный анализ качества почв Российской Федерации. В сб.: Статистические и системно-параметрические исследования. Сборник научных статей Международной научно-практической конференции. Министерство образования и науки РФ; Юго-Западный государственный университет, 2016. – С. 186-188.
9. Усманов Р.Р. Кластеризация систем обработки почвы и удобрений по агрохимическим и агрофизическим показателям. В сб.: Доклады ТСХА, 2020. – С. 60-64.
10. Фрид А.С. Параметры изменений показателей плодородия пахотных почв при различных уровнях антропогенных воздействий. В кн.: Научные основы предотвращения деградации почв (земель) сельскохозяйственных угодий России и формирования систем воспроизводства их плодородия в адаптивно-ландшафтном земледелии. – М., 2013. – С. 58-61.
11. Фрид А.С., Чуян О.Г., Соловченко В.Д., Тютюнов С.И. Оценка плодородия В кн.: Научные основы предотвращения деградации почв (земель) сельскохозяйственных угодий России и формирования систем воспроизводства их плодородия в адаптивно-ландшафтном земледелии. – М., 2013. – С. 17-34.
12. Холодов В.А., Ярославцева Н.В., Лазарев В.И., Фрид А.С. Интерпретация данных агрегатного состава типичных черноземов разного вида использования методами кластерного анализа и главных компонент// Почвоведение. – 2016. – № 9. – С. 1093-1100.
13. Экологическое нормирование и управление качеством почв и земель. А.Г. Барсегян и др.; С.А. Шоба, А.С. Яковлев, Н.Г. Рыбальский. – М., 2013.
14. Якушев В.П., Якушев В.В., Баденко В.Л., Матвеев Д.А., Чесноков Ю.В. Оперативное и долгосрочное прогнозирование продуктивности посевов на основе массовых расчетов имитационной модели агроэкосистемы в геоинформационной среде (обзор)// Сельскохозяйственная биология. – 2020. – Т. 55. – № 3. – С. 451-467.

CLUSTER ANALYSIS OF STATISTICAL PROFILES OF SPATIAL DISTRIBUTION OF HUMUS

Vorobyov N.I.¹, Ladan S.S.²

¹ All-Russia Research Institute for Agricultural Microbiology, Saint-Petersburg, RF

The indicators of the spatial distribution of the humus content in the soils of the federal districts of the Russian Federation form the individual statistical profiles of the regions. To study the reasons deforming these profiles, a cluster analysis of the humus content in arable and fallow soils in the RF regions was carried out. The values of the median and the confidence interval of the humus profiles were used to determine the statistical distance of the RF region soils and to construct cluster dendrograms. The cluster analysis showed that regions with similar climatic conditions have a similar distribution of humus in fallow soils. Cluster analysis showed that human agricultural activities change the humus profile of arable soil in the RF regions in different ways.

Key words: statistical profile of the humus spatial distribution in soil, an fallow and arable soils, dendrogram of the grouping of RF regions.