

УДК 631.474

## ГЕОСТАТИСТИЧЕСКИЙ АНАЛИЗ КАК ИНСТРУМЕНТ ОЦЕНКИ ПРОСТРАНСТВЕННОГО РАСПРЕДЕЛЕНИЯ СВОЙСТВ ПОЧВ ЗЕМЕЛЬ СЕЛЬСКОХОЗЯЙСТВЕННОГО НАЗНАЧЕНИЯ

*д-р с.-х. наук Т.Н. МЫСЛЫВА; О.А. КУЦАЕВА*

*(Белорусская государственная сельскохозяйственная академия, Горки)*

*Представлены результаты использования методов геостатистического анализа для оценки характера пространственного распределения данных о содержании в почве гумуса, подвижных фосфора и калия, а также рН почвенного раствора посредством применения инструментов модуля «Пространственная статистика» ArcGIS версии 10.5. Установлено наличие достоверной кластеризации исследуемых показателей, поскольку величина индекса Морана колеблется от 0,197827 до 0,360388; величина фактического общего индекса Getis-OrdG превышает его теоретическое значение, а величина z-оценки > 2,58. Максимальным количеством пространственных выбросов с высокими и низкими значениями характеризуются данные о содержании в почве подвижных фосфора и калия – 344 и 369 соответственно. В целом количество кластеров с низкими значениями в 1,3 раза превышает количество кластеров с высокими значениями.*

**Ключевые слова:** *геостатистический анализ, почва, кластеризация, индекс Морана, Getis-OrdG, Getis-OrdGi\*.*

**Введение.** Мониторинг состояния земельных ресурсов, оценка и прогноз их изменений под воздействием антропогенных и природных факторов – неотъемлемая часть комплекса мероприятий по рациональному использованию земель, сохранению и воспроизводству почвенного плодородия [1]. Выявление характеристик пространственной структуры агрохимических и физико-химических показателей может стать основой для оценки плодородия почв и базой для разработки рациональной политики управления окружающей средой в условиях интенсификации аграрного производства и усиления антропогенного прессинга на биосферу. Исходя из этого существует потребность в получении адекватной информации о пространственно-временном распределении агрохимических свойств почвы как в пределах отдельных регионов и природно-территориальных комплексов, так и в пределах локальных и микролокальных территорий, что особенно важно при внедрении элементов системы точного земледелия [2]. Геостатистика является эффективным методом изучения пространственного распределения характеристик почвы, их несогласованности и уменьшения дисперсии ошибок оценки и затрат на выполнение агрохимических исследований [3–5]. Одним из возможных вариантов применения методов геопространственной статистики для целей сельскохозяйственного производства и землеустройства может стать анализ пространственного распределения агрохимических и физико-химических свойств почв земель сельскохозяйственного назначения. Однако, несмотря на достаточно широкое применение в различных отраслях, методы геопространственного анализа, впрочем, как и методы геостатистического анализа в целом, по-прежнему не нашли широкого использования в аграрной сфере и землеустройстве Беларуси.

**Цель и методика исследования.** Цель данного исследования – установление возможности применения методов геопространственного анализа для оценки неоднородности пространственного распределения агрохимических и физико-химических свойств почвы в пределах землепользования РУП «Учхоз БГСХА». Геостатистический анализ данных выполнялся с помощью функциональных возможностей наборов инструментов «Анализ структурных закономерностей» и «Расчет кластеризации» модуля «Пространственная статистика» программного продукта ArcGIS версии 10.5. Для анализа использовались данные, полученные из материалов агрохимического обследования территории РУП «Учхоз БГСХА», выполненного в 2018 году УКПП «Могилевская областная проектно-изыскательская станция агрохимизации». Общая площадь обследованной территории составила 8342,1 га. Почвенный покров представлен преимущественно дерново-подзолистыми супесчаными на водно-ледниковых супесях и дерново-подзолистыми суглинистыми почвами на лессовидных суглинках [6].

Шейп-файл с размещением земельных участков в пределах территории выполнения исследования создан в среде ГИС ArcGIS версии 10.5 по результатам оцифровки планово-картографических материалов, полученных при проведении агрохимического обследования территории хозяйства в 2018 году.

Глобальный индекс Морана (I) рассчитывали по следующей формуле (1):

$$I = \frac{n \sum_{i=1}^n \sum_{j=i}^n w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\left[ \sum_{i=1}^n \sum_{j=i}^n w_{ij} \right] \cdot \left[ \sum_{i=1}^n (y_i - \bar{y})^2 \right]} \quad (1)$$

В формуле (1)  $n$  – количество единиц в выборке;  $w_{ji}$  – вес пространственной связи между  $i$ -той и  $j$ -той единицей выборки;  $y_i$  – атрибутивное значение для  $i$ -той единицы выборки;  $\bar{y}$  – выборочное среднее значение атрибута.

Величину индекса Getis-OrdGi\* рассчитывали по формуле (2):

$$\text{Getis-OrdGi}^* = \frac{\sum_{j=i}^n w_{i,j} x_j - \bar{X} \sum_{j=1}^n w_{i,j}}{\sqrt{\frac{n \sum_{j=1}^n w_{i,j}^2 - \left(\sum_{j=1}^n w_{i,j}\right)^2}{n-1}}}, \quad (2)$$

где  $x_j$  – атрибутивное значение объекта наблюдений;  $j$ ;  $w_{ij}$  – пространственный вес между объектами  $i$  и  $j$ ;  $n$  – общее число объектов.

**Основная часть.** Для исследования пространственного распределения данных о содержании в почве гумуса, подвижных фосфора и калия и pH почвенного раствора посредством инструментов модуля «Пространственная статистика» определяли: *минимальное и максимальное расстояния окрестности поиска ближайшего соседства*, дающие возможность подобрать оптимальную величину окрестности поиска при пространственном моделировании; *глобальный индекс Морана*, позволяющий определить, имеет ли место явление кластеризации по отношению к атрибутивным данным, каковыми в нашем случае являются сведения о физико-химических и агрохимических свойствах почвы; *общий индекс Getis-OrdG* для оценки общей структуры и тренда геоданных, а также степени кластеризации высоких и/или низких значений выборки; *индекс Getis-OrdGi\**, позволяющий установить наличие кластеризации данных с высокими и низкими значениями.

Для определения величины фиксированного расстояния или минимального расстояния окрестности поиска соседства между значениями того или иного почвенного показателя использовали инструмент «Пошаговая пространственная автокорреляция», в диалоговом окне которого задавали величину начального (расстояние, на котором необходимо начать анализ пространственной автокорреляции) и приращенного (расстояние, на которое необходимо увеличивать начальное расстояние при каждой последующей итерации) расстояний. В нашем случае величина расстояния, на котором необходимо начать анализ пространственной автокорреляции, составила 550 м, тогда как величина приращения, установленная эмпирическим путем, колебалась от 50 м для содержания в почве pH почвенного раствора до 100 м для содержания в почве гумуса и до 250 м для содержания в почве подвижных фосфора и калия. При выполнении пошаговой пространственной автокорреляции выделялись десять интервалов расстояний, равномерно распределенных по всему экстенду. Для каждого интервала рассчитывался глобальный индекс Морана и интервал, для которого данный индекс будет наибольшим, рекомендовался как оптимальное расстояние для окрестности поиска. В результате получали граф, на котором отмечены минимальное и максимальное расстояния окрестности поиска ближайшего соседства (рисунок 1).

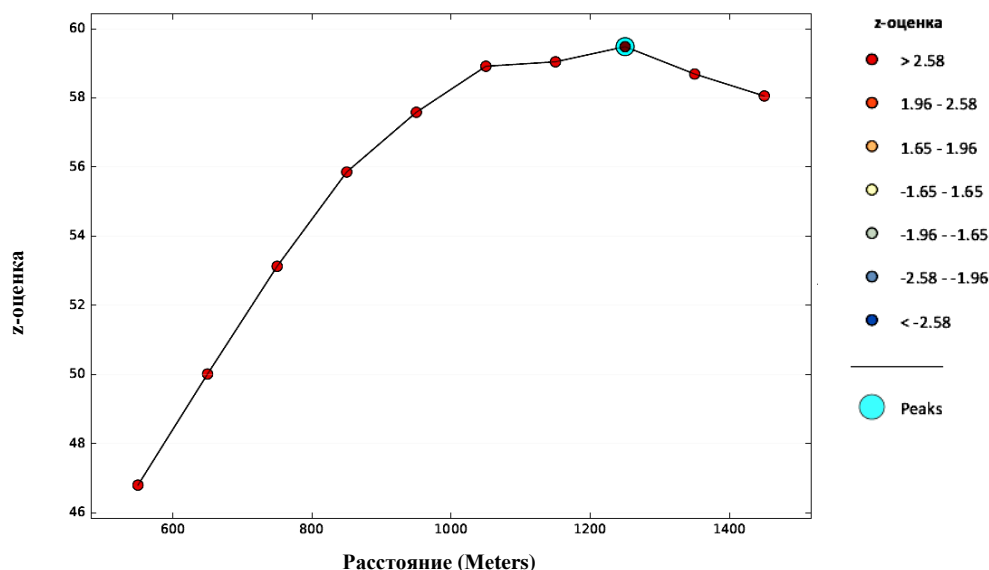


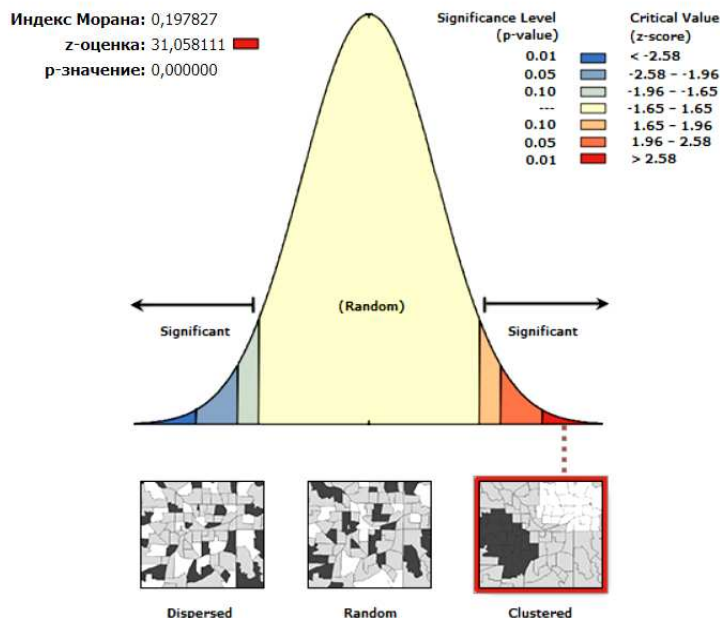
Рисунок 1. – Графическая интерпретация величины минимального расстояния окрестности поиска соседства между значениями содержания гумуса в почве

В таблице 1 приведены результаты определения минимального расстояния окрестности поиска ближайшего соседства для всех рассматриваемых показателей. Полученные результаты целесообразно использовать при подборе параметров модели вариограммы при пространственном моделировании свойств почвы интерполяцией методом кригинга, а также при установлении величины лага.

Таблица 1. – Минимальное и максимальное расстояния окрестности поиска ближайшего соседства

Название показателя	Расстояние окрестности поиска, м		Минимальное расстояние окрестности поиска		Максимальное расстояние окрестности поиска	
	минимальное	максимальное	дисперсия	z-оценка	дисперсия	z-оценка
pH <sub>KCl</sub>	700	850	0,000067	28,841475	0,000045	29,273034
Гумус, %	1050	1250	0,000030	59,854804	60,061236	0,000022
P <sub>2</sub> O <sub>5</sub> , мг/кг	2550	2550	0,000006	89,678997	0,000006	89,678997
K <sub>2</sub> O, мг/кг	2050	2050	0,000009	71,995204	0,000009	71,995204

Для того чтобы определить, имеет ли место явление кластеризации по отношению к атрибутивным данным, каковыми в нашем случае являются сведения о различных показателях почвы, рассчитывалась величина глобального индекса Морана. Глобальный индекс Морана для нормально распределенных данных лежит в диапазоне от -1 до +1. Величина индекса +1 означает детерминированную прямую зависимость – группировку схожих (низких или высоких) значений – и свидетельствует о наличии кластеризации данных; величина индекса 0 означает абсолютно случайное распределение данных; величина индекса -1 означает детерминированную обратную зависимость – идеальное перемешивание низких и высоких значений, напоминающее шахматную доску, что свидетельствует о равномерном распределении данных. Данный индекс является мерой пространственной автокорреляции и характеризует наличие или отсутствие пространственной автокорреляции геоданных. По результатам его определения генерируется графический отчет, в котором приводится величина глобального индекса Морана, рассчитанная для выборки по значениям атрибута, а также величина z-оценки, позволяющие судить о характере распределения данных (рисунок 2).



Заданная z-оценка 31.0581109664, вероятность меньше 1%, что полученный тип распределения - кластеризован - может быть результатом случайного выбора.

Рисунок 2. – Графическая интерпретация величины глобального индекса Морана, рассчитанной для pH<sub>KCl</sub> почвы

Для исследуемых показателей величина глобального индекса Морана колеблется в пределах от 0,197827 до 0,360388, следовательно, данные об агрохимических и физико-химических свойствах почвы в пределах исследуемой территории распределены не случайно и кластеризованы. Поскольку величина

z-оценки во всех случаях превышает 2,58, то с вероятностью 99% можно утверждать, что кластеризованный тип распределения данных не является случайным (таблица 2).

Таблица 2. – Результаты определения величины глобального индекса Морана

Название показателя и объем выборки	Фактический глобальный индекс Морана	Ожидаемый глобальный индекс Морана	Дисперсия	z-оценка	Тип распределения данных
pH <sub>KCl</sub> , n = 1622	0,197827	-0,000617	0,000041	31,058111	Кластеризованы с вероятностью 99%
Гумус, %, n = 1636	0,277465	-0,000612	0,000019	63,163917	
P <sub>2</sub> O <sub>5</sub> , мг/кг, n = 1630	0,204838	-0,000614	0,000005	91,269389	
K <sub>2</sub> O, мг/кг, n = 1634	0,221134	-0,000612	0,000008	79,800517	

Следует также отметить, что при расчете глобального индекса Морана учитывались не координаты пространственных данных, а их атрибутивные значения, каковыми являются значения содержания в почве гумуса, фосфора, калия и величины pH.

Степень кластеризации значений для каждой выборки данных определяли посредством вычисления общего индекса Getis-OrdG, с помощью которого оценивали общую структуру и тренд геоданных (таблица 3).

Таблица 3. – Результаты определения величины общего индекса Getis-OrdG

Название показателя и объем выборки	Фактический общий индекс Getis-OrdG	Ожидаемый общий индекс Getis-OrdG	z-оценка	p-значение	Тип кластеризации данных
pH <sub>KCl</sub> , n = 1622	0,018350	0,018242	4,463522	0,000008	Кластеризация высоких значений
Гумус, %, n = 1636	0,038475	0,036907	10,536313	0,000000	
P <sub>2</sub> O <sub>5</sub> , мг/кг, n = 1630	0,132254	0,126622	6,587257	0,000000	
K <sub>2</sub> O, мг/кг, n = 1634	0,090315	0,087229	5,583397	0,000000	

По результатам определения величины фактического общего индекса Getis-OrdG установлено, что во всех случаях имеет место его превышение над ожидаемым значением. Судя по величине z-параметра, которая превышает 2,58, с вероятностью 99% можно утверждать, что установленный сильно кластеризованный тип распределения данных с высокими значениями является не случайным, о чем свидетельствует и величина p-значения, не превышающая 0,000008 при допустимом значении 0,01. В качестве параметра концептуализации пространственных отношений был выбран *fixed\_distance\_band*, при котором каждый объект анализировался в контексте соседних объектов в пределах указанного порогового расстояния (см. таблицу 1), за пределами которого влияние соседних объектов не учитывалось.

Анализ «горячих точек» является качественным методом исследования и инструментом предварительной оценки геоданных, который ориентирован на определение приоритетных областей («горячих точек») и позволяет установить наличие кластеризации данных с высокими и низкими значениями [7]. Этот метод позволяет исследовать: 1) все объекты изучаемой области данных; 2) окрестность каждого из объектов; 3) изучаемую область данных. Цель анализа – определение наличия у окрестности объекта статистически значимых отличий изучаемого атрибута от всей области значений. Если в окрестности объекта значение изучаемого атрибута выше, чем в изучаемой области, объект является «горячей точкой», если ниже – «холодной». Его выполняют посредством определения величины индекса *Getis-OrdGi\** – статистического показателя, рассчитываемого для каждого пространственного объекта в наборе данных. Однако следует отметить, что при расчете индекса *Getis-Ord Gi\** учитываются не атрибутивные значения отдельных объектов, а атрибутивные значения их окрестностей, которые рассчитываются для каждого объекта и сравниваются со значениями в остальной области исследований.

Полученные итоговые z-оценки и p-значения свидетельствуют о том, в какой области пространства кластеризуются объекты с высокими или низкими значениями. При этом выполняется анализ каждого объекта в контексте соседних объектов. Чтобы быть статистически значимой «горячей» либо «холодной» точкой, объект должен иметь высокое или низкое значение и быть окружен другими объектами с такими же высокими или низкими значениями. Локальная сумма значений для объекта и его соседей сравнивается пропорционально с суммой всех объектов; когда локальная сумма сильно отличается от ожидаемой локальной суммы и когда это отличие является слишком большим, чтобы быть результатом случайного процесса, получается статистически значимая z-оценка. Высокое значение z-оценки и небольшая величина p-значения для функции указывают на наличие статистически значимой «горячей точки». Низкое отрицательное значение z-оценки и небольшая величина p-значения указывают на наличие статистически значи-

мой «холодной точки». Чем выше (или ниже) показатель  $z$ , тем более интенсивна кластеризация данных. Значения  $z$ -оценки, близкие к нулю, означают, что пространственная кластеризация данных отсутствует.

В результате выполнения анализа горячих точек определены статистически значимые пространственные кластеры высоких значений (горячих точек) и низких значений (холодных точек) для всех изучаемых показателей (рисунок 3).

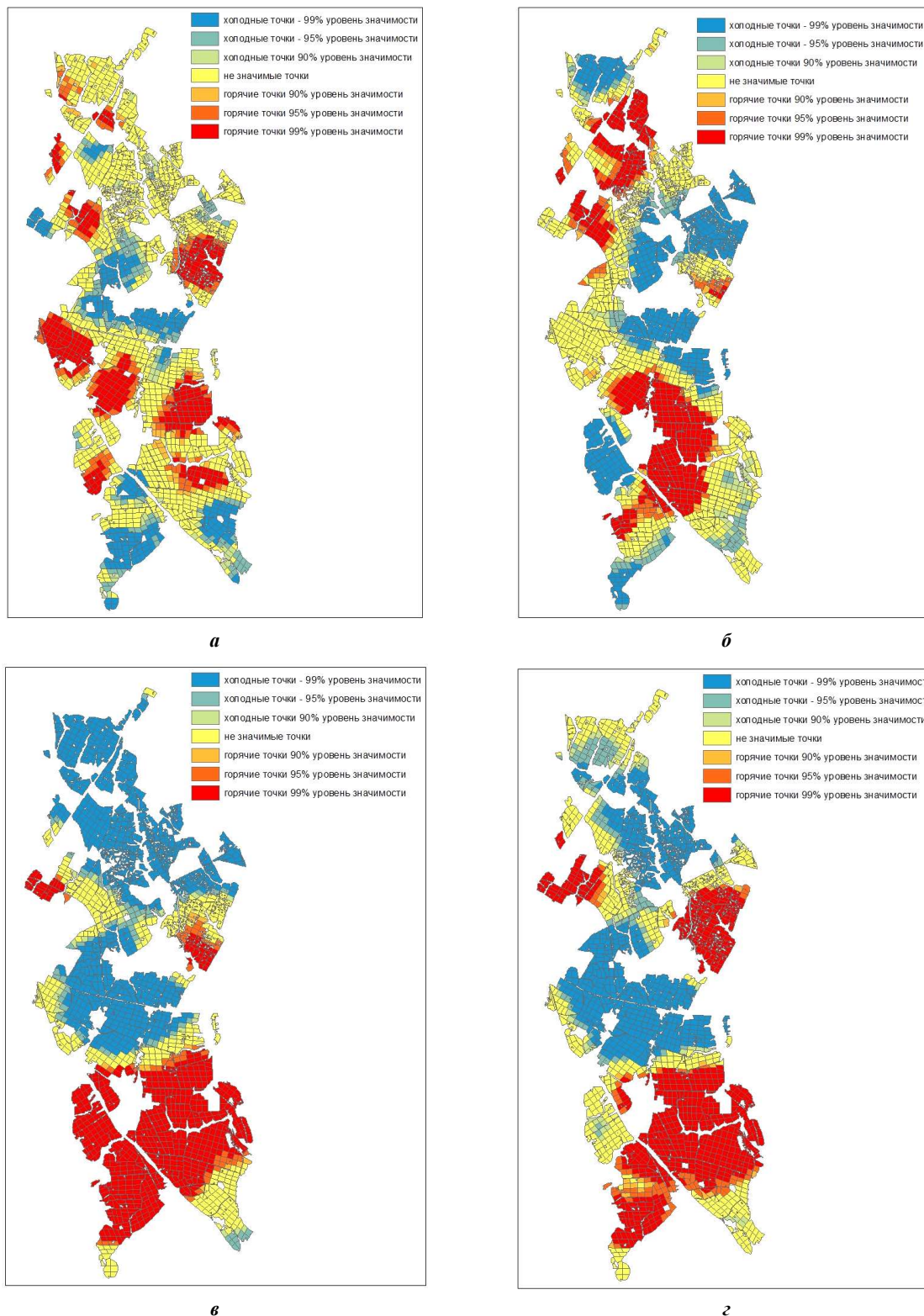


Рисунок 3. – Результаты анализа «горячих точек»:  
 а – pH<sub>KCl</sub>; б – содержание гумуса; в – содержание фосфора; г – содержание калия

Данная информация дает возможность визуально идентифицировать где, именно в пределах территории исследований будут наблюдаться максимальные и минимальные значения изучаемых свойств почв.

Анализ кластеров и выбросов идентифицирует концентрации высоких и низких значений, концентрации и пространственные выбросы данных об агрохимических и физико-химических свойствах почвы (рисунок 4). Данный анализ позволяет установить, где проходят наиболее четкие границы между контурами с высоким и низким содержанием того либо иного элемента в почве и есть ли в пределах исследуемой территории аномально высокие или аномально низкие значения показателей, которые можно отнести к пространственным выбросам.

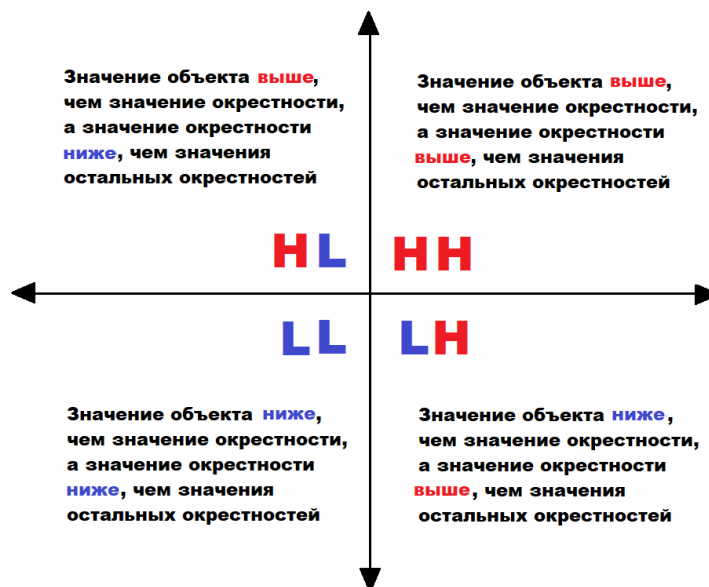


Рисунок 4. – Схема интерпретации результатов анализа кластеров и выбросов данных

Пространственные выбросы в нашем случае могут быть обусловлены как несовершенством методики отбора проб почвы, так и несовершенством применяемых методов картографирования результатов агрохимических исследований. По нашему мнению, при обнаружении участков-выбросов необходимо провести дополнительное их обследование с целью уточнения ситуации. Также необходимо исключать данные участки из набора данных при выполнении моделирования пространственного распределения того либо иного показателя посредством метода интерполяции. В частности, применение методов кригинга требует наличия данных, распределение которых нормально либо близко к нормальному. С другой стороны, наличие выбросов может быть обусловлено и объективными факторами, например, применением различных доз минеральных удобрений в пределах отдельных участков. Этим, в частности, может быть объяснена пестрота пространственного распределения подвижных фосфора и калия.

В результате выполнения анализа кластеров и выбросов по результатам оценки величины локального индекса Морана, z-оценки и p-значения установлено наличие пространственных выбросов высоких и низких значений для всех исследуемых показателей (таблица 4).

Таблица 4. – Результаты выполнения анализа кластеров и выбросов

Название показателя и объем выборки	Тип кластера							
	NN-кластер		NL-кластер		LN-кластер		LL-кластер	
	1	2	1	2	1	2	1	2
pH <sub>KCl</sub> , n = 1622	399	2022,49	141	718,65	106	536,79	262	1313,16
Гумус, %, n = 1636	335	1727,89	68	343,47	152	765,47	499	2565,58
P <sub>2</sub> O <sub>5</sub> , мг/кг, n = 1630	429	2169,89	169	843,63	175	890,89	564	2889,87
K <sub>2</sub> O, мг/кг, n = 1634	409	2087,80	158	810,0	211	1068,40	472	2402,13
1 – количество рабочих участков, входящих в кластер, шт.; 2 – площадь рабочих участков, входящих в кластер, га								

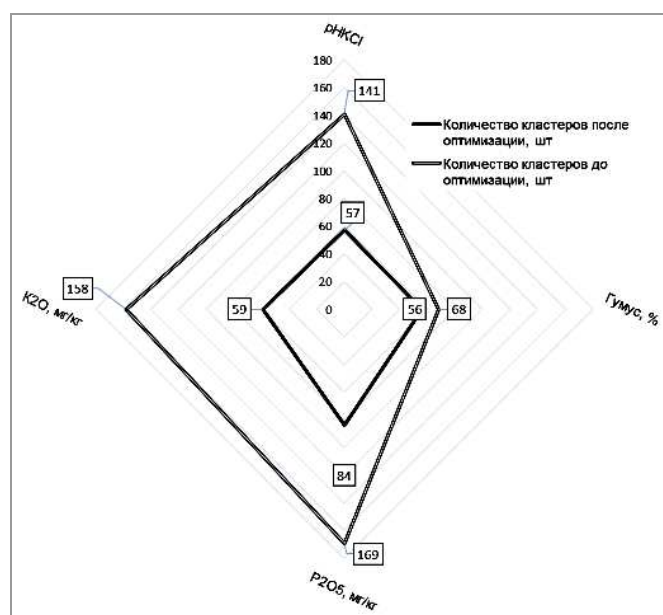
Максимальным количеством пространственных выбросов с высокими и низкими значениями характеризуются данные о содержании в почве подвижных фосфора и калия – 344 и 369 соответственно.



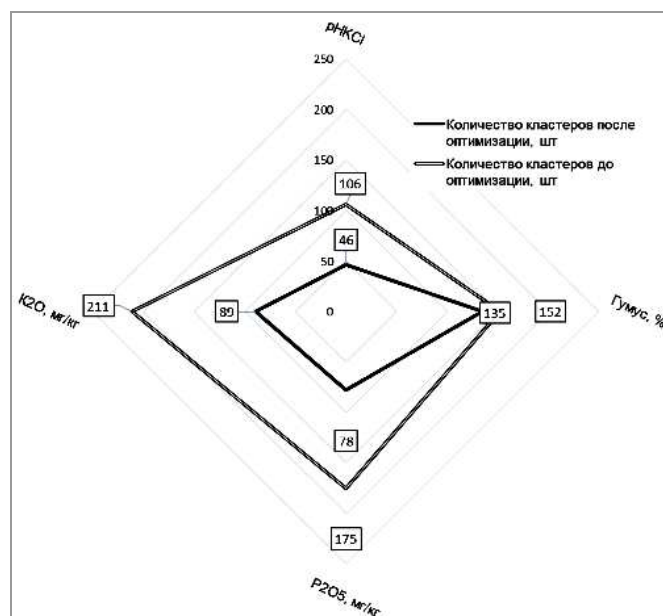
В целом количество кластеров с низкими значениями в 1,3 раза превышает количество кластеров с высокими значениями.

Для получения оптимальных результатов при определении пространственных выбросов выполнен оптимизированный анализ, при котором вычисления локального индекса Морана производилось автоматически с использованием параметров, извлеченных из характеристик входных данных. В результате оптимизации кластерного анализа количество выбросов с высокими и низкими значениями значительно сократилось (рисунок 5), а максимальное количество выбросов с высокими значениями составило 84 кластера для данных о содержании фосфора. Максимальным количеством выбросов с низкими значениями характеризуются данные о содержании в почве гумуса (135 кластеров).

При выполнении данного анализа также уточнено оптимальное расстояние окрестности поиска ближайшего соседства, которое составило, м:  $pH_{KCl}$  – 683; гумус – 1280; фосфор подвижный – 890; калий подвижный – 889. Уточненное расстояние окрестности поиска целесообразно использовать при пространственном моделировании в качестве одного из параметров модели вариограммы.



а



б

Рисунок 5. – Количество кластеров с высокими (а) и низкими (б) выбросами до и после оптимизации

**Заключение.** Использование геостатистического анализа данных о  $pH_{KCl}$  почвы и содержания в ней гумуса, подвижных фосфора и калия позволяет:

- выявить и математически оценить пространственное распределение агрохимических и физико-химических показателей почвы;
- изучить пространственную автокорреляцию данных и определить местоположения в области исследования с аномальными значениями;
- оценить кластеризацию данных о свойствах почвы и определить местоположения кластеров в пространстве;
- выполнить визуализацию кластеров путем построения карты локального индикатора пространственной ассоциативности;
- установить наиболее четкие границы между плодородными и мало плодородными землями и определить менеджмент-зоны с различной степенью плодородия по комплексу показателей.

Полученная информация может использоваться для осуществления внутрихозяйственных землеустроительных мероприятий при внедрении как системы точного земледелия в целом, так и отдельных ее элементов.

#### ЛИТЕРАТУРА

1. Мыслыва, Т.Н. Практические аспекты использования методов геопространственного анализа в землеустройстве и земельном кадастре / Т.Н. Мыслыва, О.А. Куцаева // Земля Беларуси. – 2018. – № 3. – С. 14–18.
2. Якушев, В.В. Точное земледелие: теория и практика / В.В. Якушев. – СПб. : ФГБНУ АФИ, 2016. – 364 с.
3. Behera, S.K. Spatial distribution of surface soil acidity, electrical Conductivity, soil organic carbon content and exchangeable potassium, calcium and magnesium in some cropped acid Soils of India / S.K. Behera, A.K. Shukla // Land Degrad. Dev. – 2015. – Vol. 16. – P. 71–79.
4. Geostatistical interpolation of object counts collected from multiple strip transects ordinary kriging versus finite domain kriging / H. Saito [et al.] // Stoch. Environ. Res. Risk Assess. – 2005. – Vol. 19. – P. 71–85.
5. Spatial variability of soil organic carbon in the forestlands of northeast China / L. Liu, H. Wang, W. Dai [et al.] // J. Forest. Res. – 2014. – Vol. 25 (4). – P. 867–876.
6. Мыслыва, Т.Н. Геостатистический анализ пространственного распределения агрохимических свойств почв земель сельскохозяйственного назначения / Т.Н. Мыслыва, Ю.А. Белявский // Приемы повышения плодородия почв и эффективности удобрения : материалы междунар. науч.-практ. конф., посвящ. памяти ученых: А.И. Горбылевой, Ю.П. Сиротина и В.И. Тюльпанова. – Горки : БГСХА, 2019. – С. 101–103.
7. Дидиков, А.Е. Применение метода «Анализ горячих точек» для повышения эффективности управления предприятием и выявления значимых экологических проблем / А.Е. Дидиков, М.Е. Алексева // Экономика и экологический менеджмент. – 2015. – № 3. – С. 273–282.

Поступила 10.12.2019

#### GEOSTATISTICAL ANALYSIS AS A TOOL FOR EVALUATING THE SPATIAL DISTRIBUTION OF AGROCHEMICAL, PHYSICAL AND CHEMICAL PROPERTIES OF SOILS OF AGRICULTURAL PURPOSES

T. MYSLYVA, O. KUTSAEVA

*The results of using the methods of geostatistical analysis to assess the nature of the spatial distribution of data on the content of humus, mobile phosphorus and potassium in the soil and the pH of the soil solution applying the tools of the “Spatial statistics” module of ArcGIS version 10.5 are presented. The presence of reliable clustering of the studied indicators was established, since the value of the Moran index ranges from 0.197827 to 0.360388; the value of the actual overall Getis-OrdG index exceeds its theoretical value, and the value of the z-score is > 2.58. Data on the content of mobile phosphorus and potassium in the soil are characterized by the maximum amount of spatial emissions with high and low values – 344 and 369, respectively. In general, the number of clusters with low values is 1.3 times higher than the number of clusters with high values.*

**Keywords:** geostatistical analysis, soil, clustering, Moran index, Getis-OrdG, Getis-OrdGi \*.