

# ГЕОПРОСТРАНСТВЕННАЯ СТАТИСТИКА

ЛЕКЦИЯ 3

# ПЛАН ЛЕКЦИИ:

2

- **1. Понятие геопространственной статистики и ее прикладные аспекты.**
- **2. Усредненный центр данных и индекс ближайшего соседства.**
- **3. Анализ и моделирование пространственной корреляции. Анализ кластеризации геопространственных данных.**
- **4. Вариография. Меры пространственной корреляции.**
- **5. Построение вариограмм. Моделирование вариограмм. Анизотропия вариограмм.**
- **6. Пространственный тренд и нестационарность. Моделирование пространственных взаимоотношений**

# Литературные и информационные ИСТОЧНИКИ

1. Геостатистика: теория и практика / В. В. Демьянов, Е. А. Савельева; под ред. Р. В. Арутюняна; Ин-т проблем безопасного развития атомной энергетики РАН. – М.: Наука, 2010. – 327 с.
2. Митчелл, Э. Руководство по ГИС-анализу / Э. Митчелл. – ESRI: 2000. – Ч. 1: Пространственные модели и взаимосвязи. – 170 с.
3. Домашняя страница проекта Quantum GIS [Электронный ресурс]. – Режим доступа: <https://www.qgis.org/en/site/>.
4. Официальный сайт ArcGIS Desktop [Электронный ресурс]. – Режим доступа: <https://desktop.arcgis.com/ru/desktop/>.

□ **1. Понятие геопространственной статистики  
и ее прикладные аспекты**

- **Геопространственная статистика** – это статистика с высокой степенью пространственного разрешения, являющаяся результатом тесной увязки статистической и геопространственной информации.
- *Анализ, проводимый в географическом контексте, приобретает все большее значение для понимания не только математического, но и пространственного распределения данных, а также для получения качественно новых данных из уже существующих.*

## □ Основные операции геопространственной

### статистики :

6

- статистическая обработка атрибутивных данных;
- описательная статистика выборки геопространственных данных;
- расширенные операции пространственной статистики.

- **Статистическая обработка атрибутов позволяет выполнять расчет значений новых атрибутивных полей на основе уже существующих.**
- *Пример - расчет соотношений кальция и магния в пределах рабочих участков землепользования сельскохозяйственного предприятия, осуществляемый на основе атрибутивных данных о содержании кальция и магния в почве каждого из рабочих участков, локализованных в пределах пахотных земель*



- 9
- **Функции описательной статистики** выборки геопространственных данных дают возможность выполнить расчет основных элементов описательной статистики массива атрибутивных данных векторных объектов (*максимальное, минимальное и среднее значения; медиану, среднеквадратическое отклонение, эксцесс и асимметрию, верхний и нижний квартили*), исследовать распределение и суммарную статистику набора геоданных, а также визуализировать данные с тем либо иным набором параметров описательной статистики.

**Расширенные операции геопространственной статистики** включают определение пространственного распределения геоданных и

10

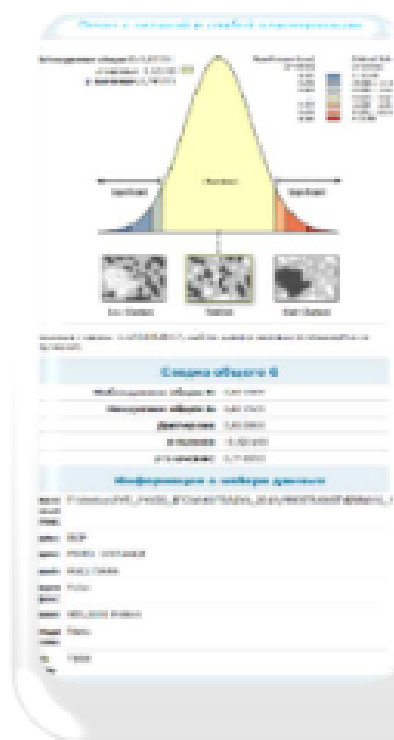
позволяют выполнить:

- анализ структурных закономерностей в распределении данных;
- измерение пространственного распределения данных;
- моделирование пространственных взаимосвязей между геоданными;
- расчет и анализ кластеризации данных;
- установление пространственной зависимости (кластеризацию или дисперсию объектов) по всему диапазону области наблюдения через заданные промежутки расстояний.

- Одним из вариантов применения методов геопространственной статистики для целей сельскохозяйственного производства и землеустройства может стать анализ пространственного распределения агрохимических свойств почв земель сельскохозяйственного назначения

# Функциональные возможности геопространственной статистики

12



Геопространственная статистика позволяет:

- суммировать ключевые характеристики распределения геоданных;
- идентифицировать статистически значимые пространственные кластеры (горячие точки, холодные точки) и пространственные выбросы;
- оценить общую картину кластеризации или дисперсии геоданных;
- разделить объекты на группы по схожести признаков;
- идентифицировать объекты со схожими характеристиками;
- моделировать пространственные отношения геоданных.

- **2. Усредненный центр данных и индекс ближайшего соседства**

- **Усредненный центр распределения данных** рассчитывается по атрибутам значений либо по значениям координат  $x$  и  $y$ .
- Вычисление координат усредненного центра распределения данных производится по формулам

$$\bar{X}_w = \frac{\sum x_i \cdot w_i}{\sum w_i};$$

$$\bar{Y}_w = \frac{\sum y_i \cdot w_i}{\sum w_i},$$

где  $X_w$  и  $Y_w$  – координаты усредненного центра распределения данных;

$x_i$  и  $y_i$  – координаты каждой точки, для которой известно значение атрибута;

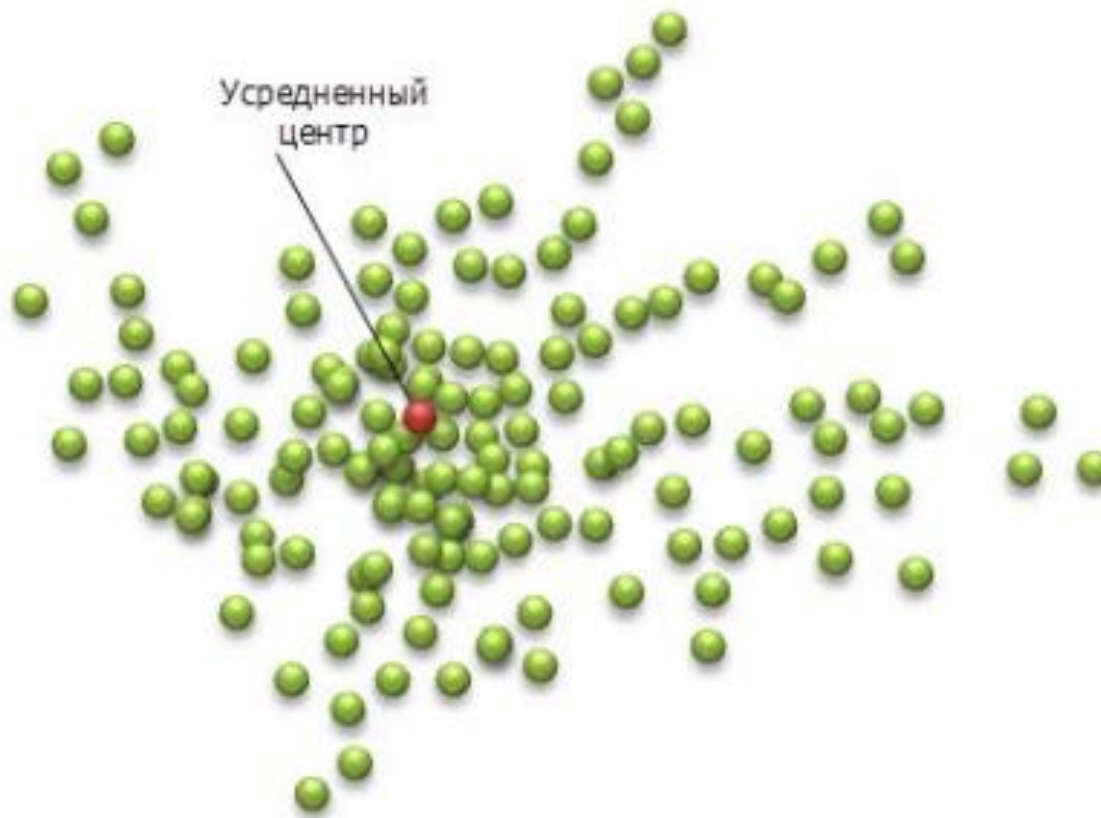
$w_i$  – значение весового атрибута в точке с координатами  $x_i$  и  $y_i$ .

## Результаты определения усредненного центра геоданных могут использоваться:

- для проведения криминалистического анализа при определении смещения среднего центра выявления краж в светлое время суток по сравнению с ночью для оптимального распределения ресурсов по предотвращению их осуществления;
- определения среднего центра наблюдаемости видов животных в пределах природоохранного объекта для обеспечения посетителей такого рода информацией;
- оценки среднего центра вызовов доставки еды с целью оптимизации размещения служб доставки;
- оценки среднего центра, взвешенного по количеству жителей младше 25 лет, чтобы определить идеальное местоположение центров по развитию досуга молодежи;
- установления усредненного центра возникновения случаев инфекционных заболеваний, чтобы определить вероятную начальную точку эпидемии.

# Усредненный центр распределения данных

16



- Пространственной вариацией усредненного центра данных является расчет эллипса стандартного отклонения или направленного распределения, выполняемый по формулам:

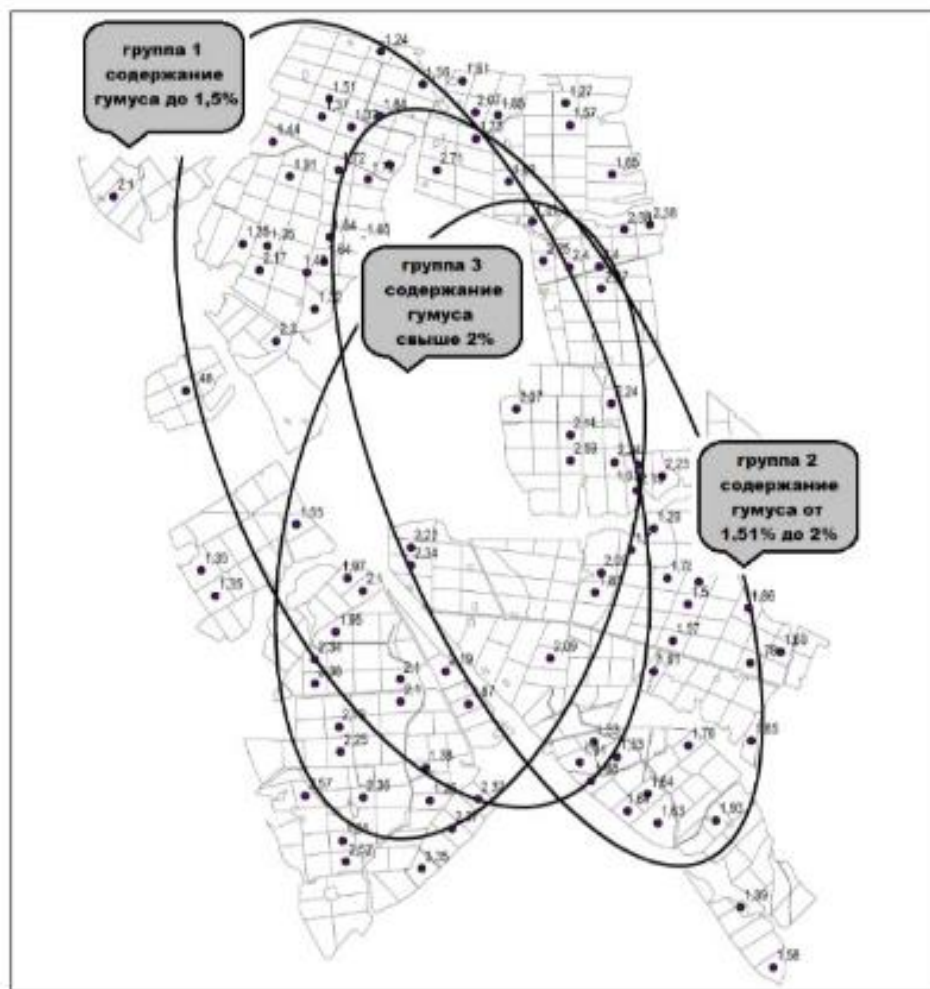
$$SDx_w = \sqrt{\frac{\sum w_i \cdot (x_i - \bar{x}_w)^2}{\sum w_i}};$$

$$SDy_w = \sqrt{\frac{\sum w_i \cdot (y_i - \bar{y}_w)^2}{\sum w_i}},$$

где  $SDx_w$  и  $SDy_w$  – среднеквадратичное отклонение по координатам  $x$  и  $y$  от среднего значения выборки геопространственных данных.

# Эллипсы распределения данных о содержании в почве гумуса для каждого из интервалов содержания

18



- Для оценки характера распределения данных и уровня их кластеризации вычисляют **величину индекса ближайшего соседства**.
- **Если данный показатель меньше единицы**, данные распределены не случайно и в них имеются кластеризированные области.
- **Если данный показатель больше единицы**, то данные распределены равномерно и явление кластеризации в них отсутствует.
- **Если индекс ближайшего соседства равен единице** – распределение данных является случайным и нулевая гипотеза о том, что данные распределены случайно и пространственно не связаны, не отвергается

### **3. Анализ и моделирование пространственной корреляции. Анализ кластеризации геопространственных данных**

□ **Пространственная корреляция** – зависимость

21

между значениями пространственно  
распределенной функции от взаимного  
расположения точек, которая оценивает  
пространственное распределение на основе  
местоположений пространственных объектов и  
их атрибутивных значений, используя величину  
общего или глобального индекса Морана ( $I$ ).

- Глобальный индекс Морана для нормально распределенных данных лежит в диапазоне от  $-1$  до  $+1$ .
- 1) величина индекса  $+1$  означает детерминированную прямую зависимость – группировку (кластеризацию) схожих (низких или высоких) значений и свидетельствует о наличии достоверной кластеризации данных;
- 2) величина индекса  $0$  означает абсолютно случайное распределение данных;
- 3) величина индекса  $-1$  означает детерминированную обратную зависимость – идеальное перемешивание низких и высоких значений, напоминающее шахматную доску, что свидетельствует о равномерном распределении данных и отсутствии их кластеризации

□ **Различают следующие концептуальные модели пространственных отношений:**

- – **модель 1:** близко расположенные соседние объекты оказывают большее влияние на вычисления для целевого объекта, чем удаленные объекты;
- – **модель 2:** близко расположенные соседние объекты оказывают большее влияние на вычисления для целевого объекта, чем удаленные объекты, однако угол наклона острее, влияние объектов уменьшается быстрее, и только ближайшие соседи оказывают существенное влияние;
- – **модель 3:** каждый объект анализируется в контексте соседних объектов в пределах указанного порогового расстояния, за пределами которого влияние соседних объектов не учитывается;
- – **модель 4:** объекты в пределах указанного порогового расстояния оказывают влияние на вычисления для целевого объекта, а влияние объектов за его пределами постепенно уменьшается;
- – **модель 5:** пространственные отношения определены путем вычисления матрицы пространственных весов.

□ Для определения величины фиксированного

расстояния или минимального расстояния окрестности

поиска соседства используют **пошаговую**

**пространственную автокорреляцию**, для которой

задают величину начального (*расстояние, на котором*

*необходимо начать анализ пространственной*

*автокорреляции*) и приращенного (*расстояние, на*

*которое необходимо увеличивать начальное*

*расстояние при каждой последующей итерации – лаг*)

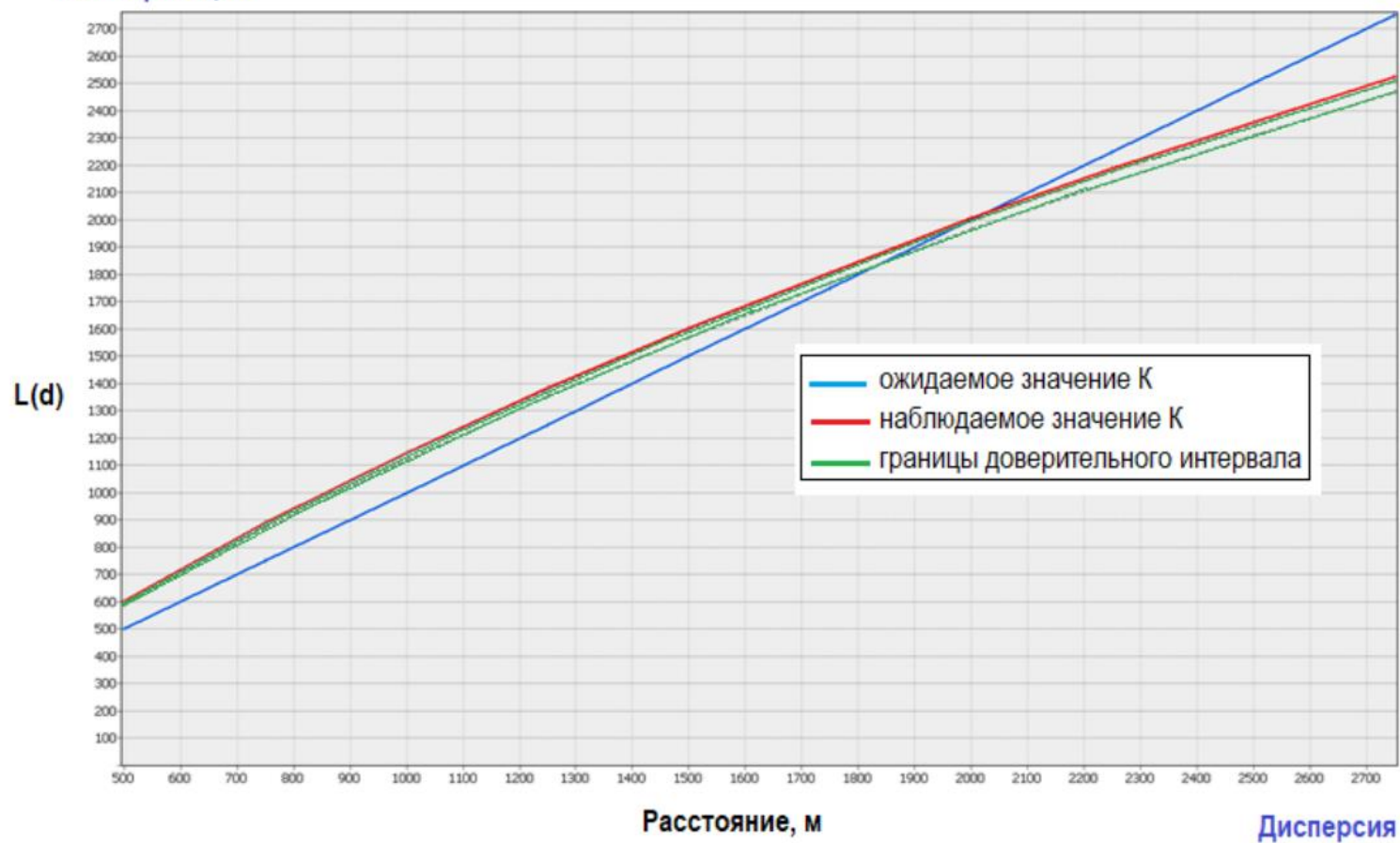
расстояний.

- Важным инструментом пространственной статистики и одним из способов анализа пространственных закономерностей в случайных точечных данных является **пространственный кластерный анализ на основе множественных расстояний, основанный на K-функции Рипли.**
- Данный вид анализа позволяет установить пространственную зависимость (кластеризацию или дисперсию объектов) по всему диапазону области наблюдения через заданные промежутки расстояний

# Результат выполнения анализа на основе множественных расстояний

26

кластеризация



## **4. Вариография. Меры пространственной корреляции**

- Под **вариографией** понимают анализ и моделирование пространственной корреляционной структуры данных.
- Анализ пространственной корреляционной структуры данных подразделяется на два этапа:
  - – этап 1: построение и интерпретация мер пространственной непрерывности на основе данных;
  - – этап 2: моделирование пространственной корреляционной структуры и построение теоретической функции, аппроксимирующей экспериментальные значения мер корреляции аналитической формулой.

- **Сущность вариографии** состоит в выявлении наличия корреляционной структуры в данных и ее описании.
- Определяется **эффективный радиус корреляции данных** (если он существует) – максимальное расстояние, на котором еще наблюдается зависимость между значениями в точках.
- **Конечной целью вариографии** является построение аналитической функции, описывающей пространственную корреляционную структуру данных для использования в геостатистических моделях интерполяции – модели вариограммы.

- **Ковариация** (covariance) – статистическая мера корреляции между двумя значениями  $Z(x)$  и  $Z(x + h)$  в точках, разделенных вектором  $h$ . Она характеризует степень похожести данных – *чем более похожи данные (ближе значения), тем больше значение ковариации.*
- **Полувариограмма** (semivariogram), **или вариограмма**, – вариация разницы значений переменной в двух точках как функция расстояния и направления между ними. Она характеризует степень различия данных в зависимости от расстояния между ними: *чем ближе значения данных (меньше разница между ними), тем больше значение вариограммы.*

- Существуют также статистические моменты, аналогичные вариограмме, но отличающиеся степенью, в которую возводится разница значений пар точек:
- – **мадограмма** (madogram) – модуль разницы значений переменной в двух точках как функции расстояния и направления между ними, позволяющий уменьшить влияние больших разбросов значений по сравнению с вариограммой.

- – **родограмма** (rodogram) – квадратный корень разницы значений переменной в двух точках как функции расстояния и направления между ними, позволяющий еще более снизить влияние значений с большим разбросом.
- – **дрейф** (drift) – зависимость средней разности значений функции точек от вектора расстояния между точками.

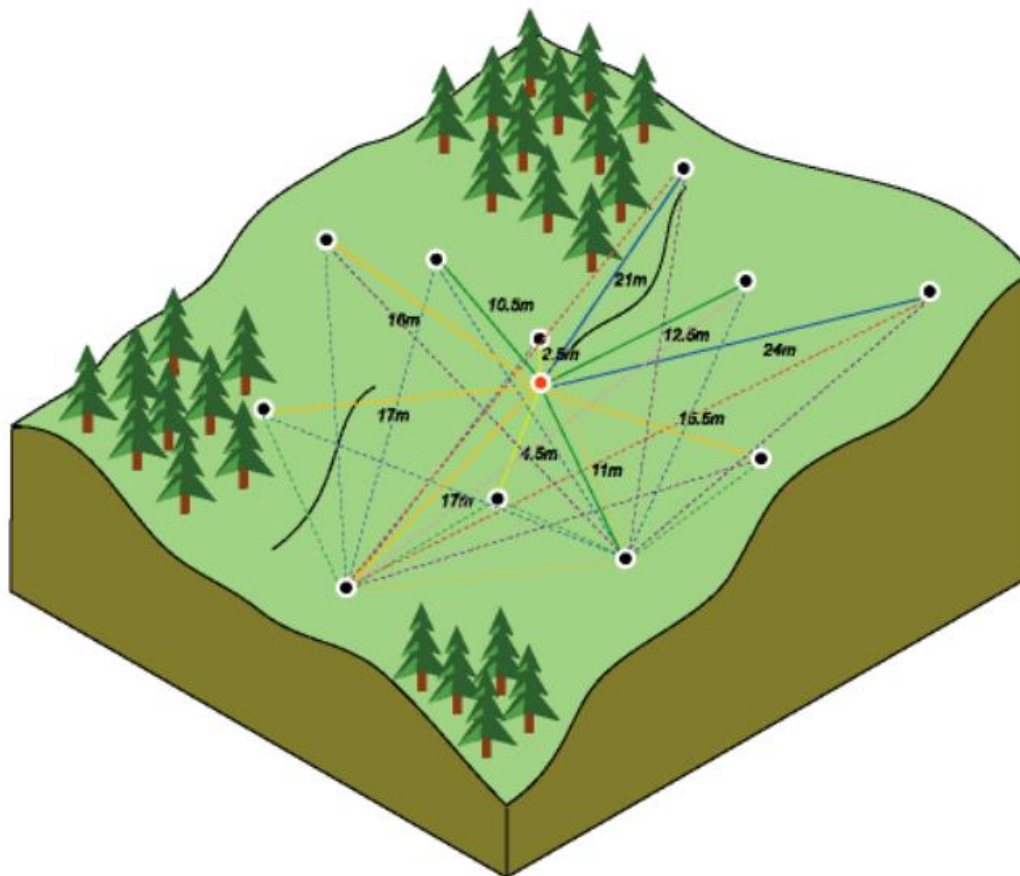
- **5. Построение вариограмм. Моделирование вариограмм. Анизотропия вариограмм**

**Для построения вариограммы** определяют квадрат разности между значениями для всех пар точек во всех местоположениях.

*На рисунке показан пример создания пары одного положения (красная точка) с другими одиннадцатью положениями. На основании полученных значений квадратов разностей между соседними парами точек строится график, на оси  $Y$  которого размещаются половины квадрата разности значений между соседними парами точек, а на оси  $X$  – интервал, разделяющий соседние пары точек.*

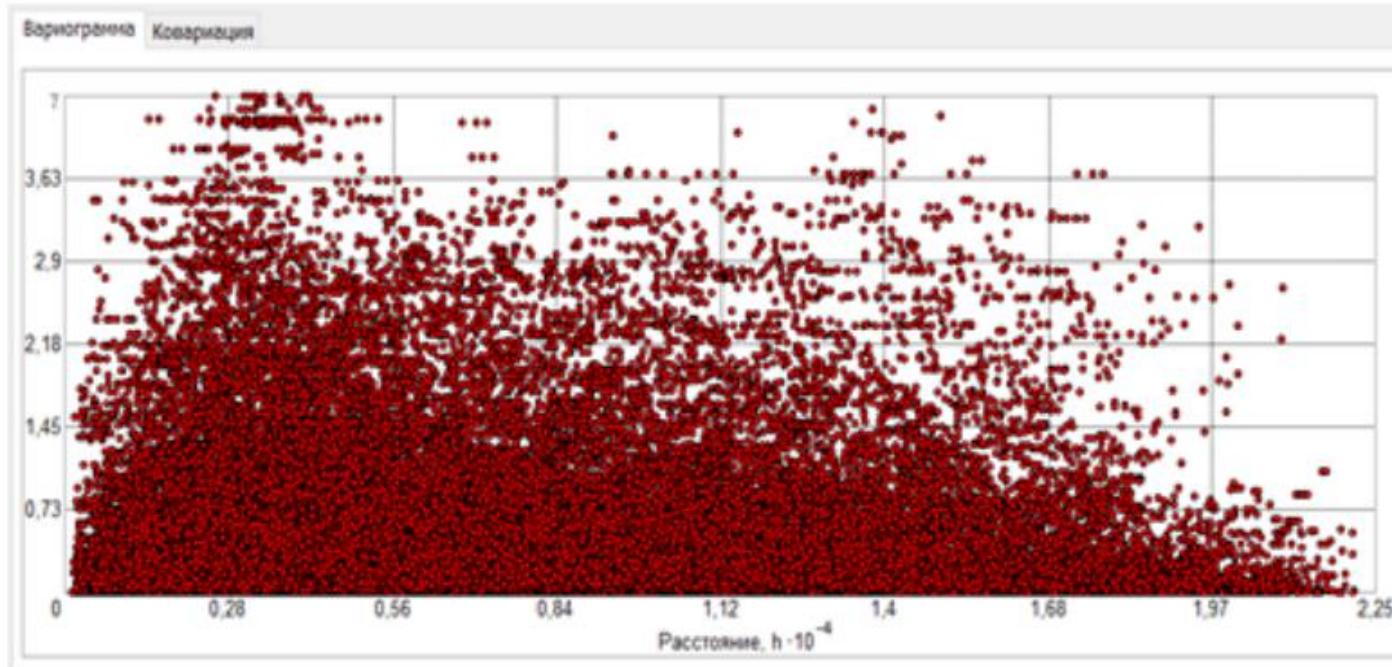
# Пример создания пар для точки с определенным местоположением с соседними точками

35



# Полученный график называют облаком вариограммы, или эмпирической вариограммой

36



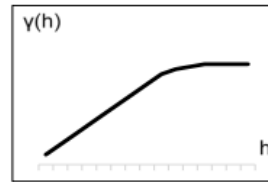
*Эмпирическая вариограмма – это диаграмма средних значений вариограммы на оси Y и расстояния (или лага) на оси X.*

Чтобы снизить число точек на эмпирической вариограмме, пары местоположений группируются на основе расстояния, которое отделяет их друг от друга. Процесс такого группирования называется **биннингом**.

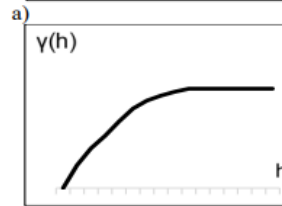
**Моделирование вариограммы** – процесс подбора математической модели, удовлетворяющей всем свойствам вариограммы и позволяющей описать ее для любого лага ( $a$ ) (расстояния) и направления.

**Наиболее часто используемыми типами моделей вариограммы** являются сферическая, круговая, экспоненциальная, гауссова и линейная, где  $c_0$  – самородок;  $c_1$  – порог вариограммы

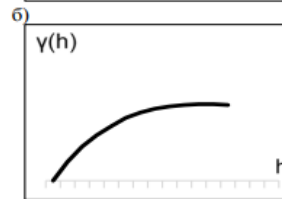
Общие формы и уравнения математических моделей, используемых для описания вариограммы (а – эффективный радиус корреляции):  
 а – сферическая; б – круговая; в – экспоненциальная; г – гауссова;  
 д – линейная



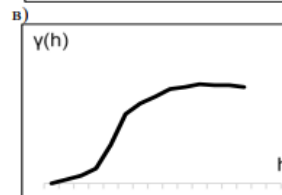
$$\gamma(h) = \begin{cases} c_0 + c \left( \frac{3h}{2a} - \frac{1}{2} \left( \frac{h^3}{a^3} \right) \right) & 0 < h \leq a \\ c_0 + c & h > a \\ 0 & h = 0 \end{cases}$$



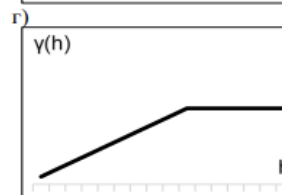
$$\gamma(h) = \begin{cases} c_0 + \frac{2c}{\pi} \left( \frac{h}{a} - \sqrt{1 - \left( \frac{h^2}{a^2} \right)} + \arcsin \left( \frac{h}{a} \right) \right) & 0 < h \leq a \\ c_0 + c & h > a \\ 0 & h = 0 \end{cases}$$



$$\gamma(h) = \begin{cases} c_0 + c \left( 1 - \exp \left( -\frac{h}{a} \right) \right) & h > 0 \\ 0 & h = 0 \end{cases}$$



$$\gamma(h) = \begin{cases} c_0 + c \left( 1 - \exp \left( -\frac{h^2}{a^2} \right) \right) & h > 0 \\ 0 & h = 0 \end{cases}$$



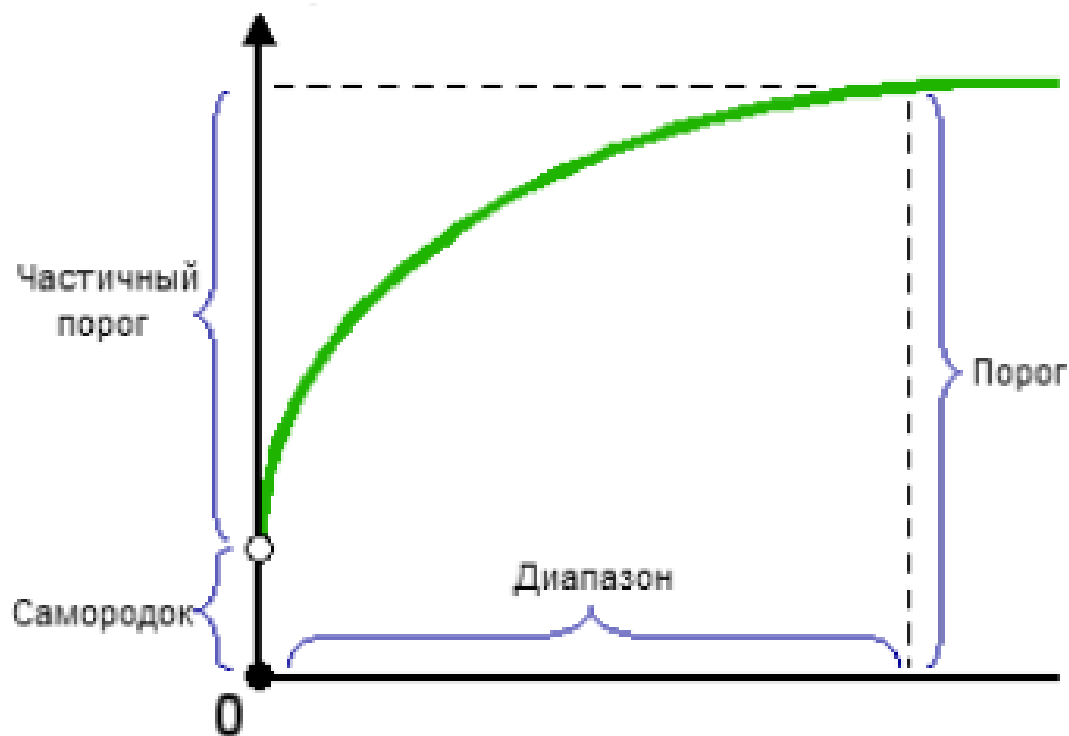
$$\gamma(h) = \begin{cases} c_0 + c \left( \frac{h}{a} \right) & 0 < h \leq a \\ c_0 + c & h > a \\ 0 & h = 0 \end{cases}$$

д)

**Эффективный радиус корреляции** – это порог вариограммы, или расстояние, на котором модель вариограммы выравнивается

39

## Характеристики модели вариограммы



Значение, в котором модель вариограммы достигает *диапазона* (значение на оси  $y$ ), называется **порогом**.

**Диапазон** – это расстояние, при котором кривая модели вариограммы достигает своего предельного значения (*порога*).

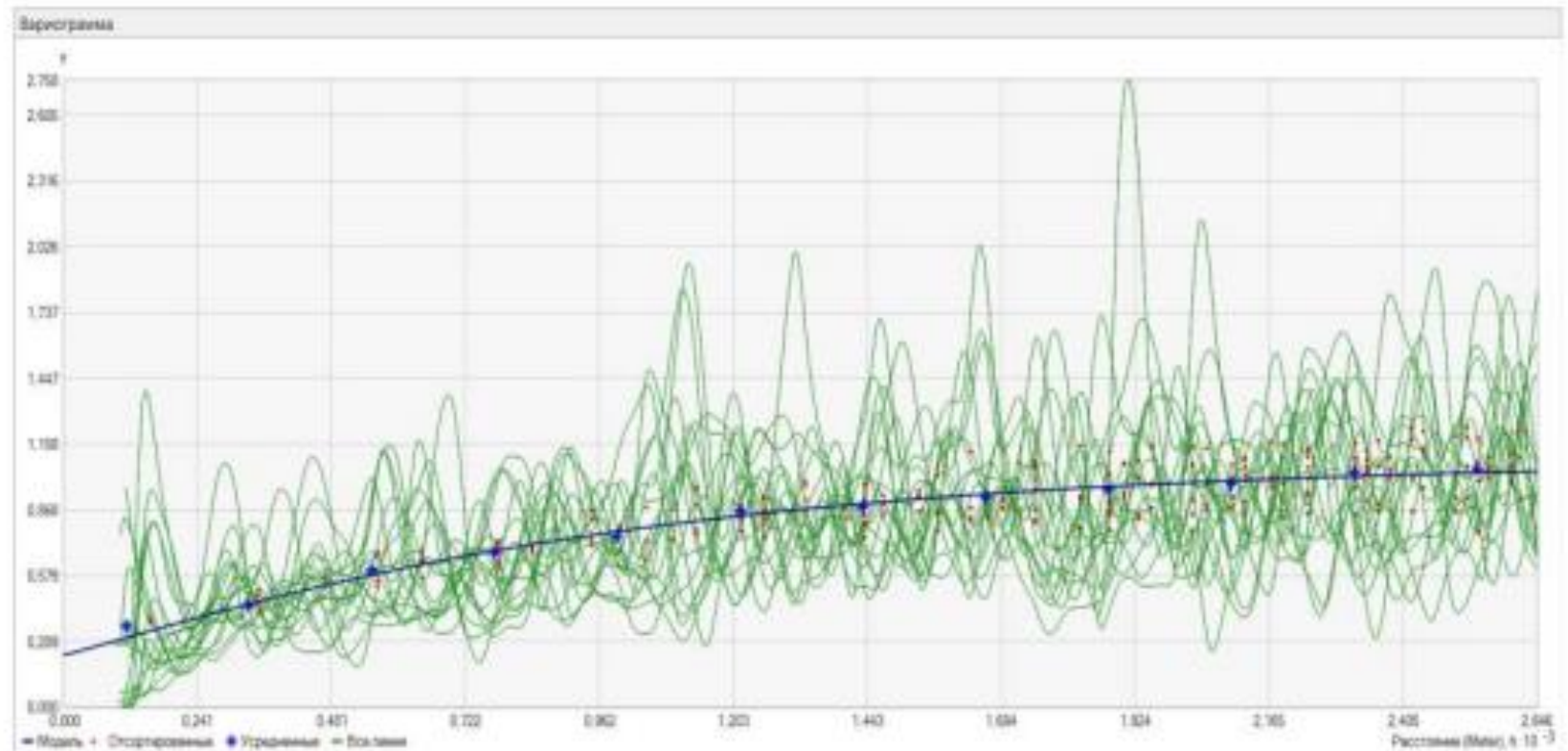
**Частичный порог** – это порог минус самородок.

**Самородок** – это погрешность измерения и (или) вариация на микроуровне (изменения в пространственном масштабе, слишком мелкие для обнаружения).

- Часто применяются **сферическая, экспоненциальная и гауссова модели** (или их комбинации).
- **Модель вариограммы (ковариации), подобранная к эмпирическим данным, должна:**
  - 1) проходить через центр облака объединенных в бины значений (красные точки);
  - 2) проходить как можно ближе к средним значениям (синие перекрестья);
  - 3) располагаться как можно ближе к середине локальных полиномов (зеленые линии)

# Модель вариограммы, подобранная к эмпирическим данным

42



Зависимость свойств изучаемых явлений от направлений в пространстве называют **анизотропией**.

43

**Анизотропия вариограммы** – изменение вариограммы и функций ковариации не только с изменением расстояния, но и с изменением направления.

Если **модель вариограммы** зависит только от расстояния между **точками**, она называется **изотропной**.

Если же **вариограмма** зависит и от **ориентации пары точек в пространстве**, то имеет место наличие **анизотропии**, что означает существование структур данных с различными пространственными характеристиками в различных направлениях.

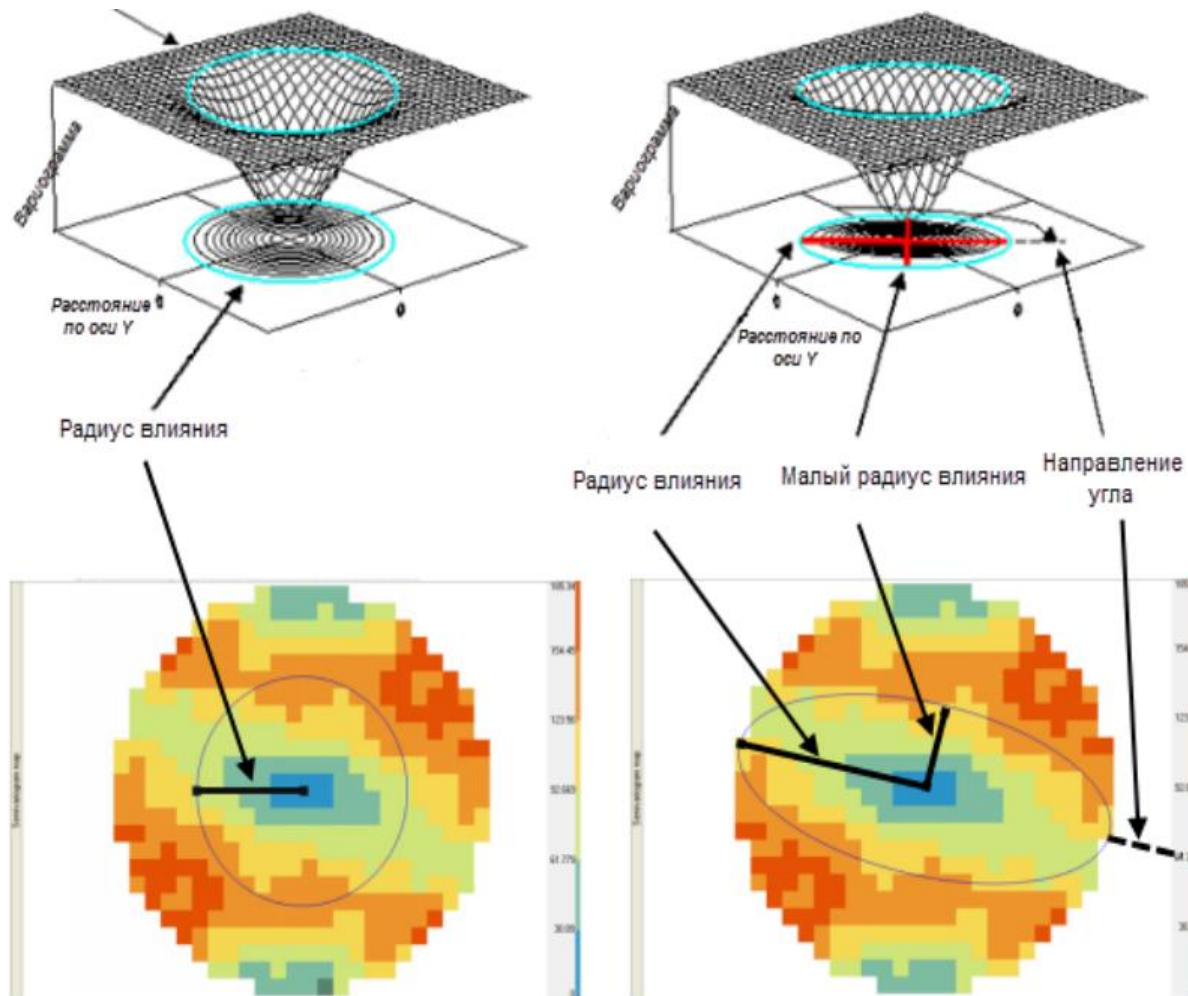
□ **Изотропная модель** вариограммы **одинакова во всех направлениях.**

44

□ **Анизотропная модель** достигает значения порога в одних направлениях **быстрее, чем в других.** Длина **длинной оси** при движении к порогу называется **большим радиусом**, длина **короткой оси** называется **малым радиусом**, также отображается **угол поворота** линии, которая формирует **большой радиус.**

# Анизотропия вариограммы

45



**В традиционной геостатистике анизотропию подразделяют на два класса: геометрическую и зонную (все остальные варианты анизотропии, кроме геометрической).**

**В случае геометрической анизотропии** изолинии вариограммы на вариограммной поверхности или вариограммной розе имеют **форму эллипса.**

**При негеометрической анизотропии** изолинии на вариограммной поверхности или вариограммной розе образуют **форму, отличную от эллипса.**

*Установить наличие анизотропии данных возможно посредством моделирования облака вариограммы в различных направлениях. О наличии данного явления будет свидетельствовать изменение формы облака с изменением направления поиска.*

□ **6. Пространственный тренд  
и нестационарность. Моделирование  
пространственных взаимоотношений**

- **Пространственный тренд** – это систематическое изменение наблюдаемой величины либо явления с изменением его местоположения (координаты).  
*(Изменение температуры воздуха в зависимости от высоты над уровнем моря)*
- Любая поверхность может состоять из двух основных компонентов:
  - 1) фиксированного (систематического) глобального тренда;
  - 2) случайной вариации на микроуровне.

*Глобальный тренд иногда называют структурой фиксированного среднего.*

Случайная вариация на микроуровне (которую иногда называют случайной ошибкой) может моделироваться в двух частях: **пространственная автокорреляция и эффект самородка.**

- Существуют геостатистические модели, позволяющие оценивать данные в присутствии пространственного тренда:
- – **кригинг с трендом** (универсальный кригинг) использует модель тренда как линейную комбинацию набора базисных функций. *Он прост в применении, не требует дополнительных настроек параметров, если правильно выбраны базисные функции;*
- – **кригинг с внешним дрейфом** использует дополнительные данные измерений коррелированной переменной в качестве модели тренда и позволяет достаточно точно оценить тренд при наличии данных дополнительной тренд-переменной во всех точках оценивания;

- – **локально меняющееся среднее** использует в качестве модели тренда локальное среднее значение, которое может быть получено с помощью метода движущегося окна;
- – **кригинг невязок с движущимся окном** подобен модели с локально меняющимся средним, однако вычислительно гораздо более сложен, поскольку предполагает подбор модели тренда и модели вариограммы в каждой локальной окрестности (окне);
- – **внутренняя случайная функция порядка  $k$**  использует моменты более высокого порядка, чем второй, совместно с вариограммой для моделирования трендов;
- – **моделирование нелинейного тренда** на разных масштабах с помощью искусственной нейронной сети.

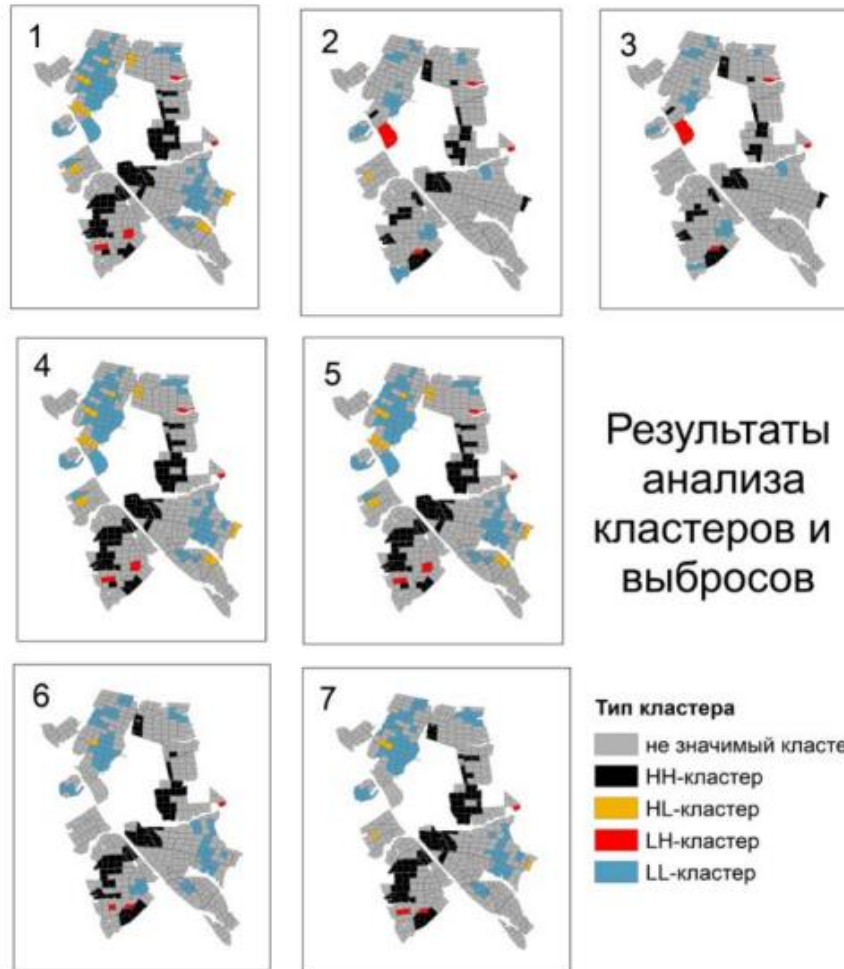
- **Моделирование пространственных взаимоотношений** должно отражать **внутренние отношения между пространственными объектами**, которые анализируются.

51

- **Внутренние отношения** между пространственными объектами, которые анализируются, могут быть описаны с помощью следующих **видов отношений**:
  - 1) обратно-взвешенное расстояние;
  - 2) обратно-взвешенные расстояние и площадь;
  - 3) фиксированный диапазон расстояний;
  - 4) зона индифферентности;
  - 5) смежность полигонов (по границе и (или) узлу);
  - 6) ближайшая окрестность  $K$ ;
  - 7) естественная окрестность (триангуляция Делоне);
  - 8) пространственно-временное окно;
  - 9) матрица пространственных весов

Результаты кластерного анализа при различных типах пространственных отношений: 1 – матрица пространственных весов; 2 – смежность полигонов по ребрам и узлам; 3 – смежность полигонов по ребрам; 4 – зона индифферентности; 5 – фиксированный диапазон расстояний; 6 – обратно-взвешенные расстояние и площадь; 7 – обратно-взвешенное расстояние

52



□ Спасибо за внимание!