



ЛЕКЦИЯ 2. ОСНОВНЫЕ ПОНЯТИЯ И ТЕРМИНЫ ГЕОМАТИКИ

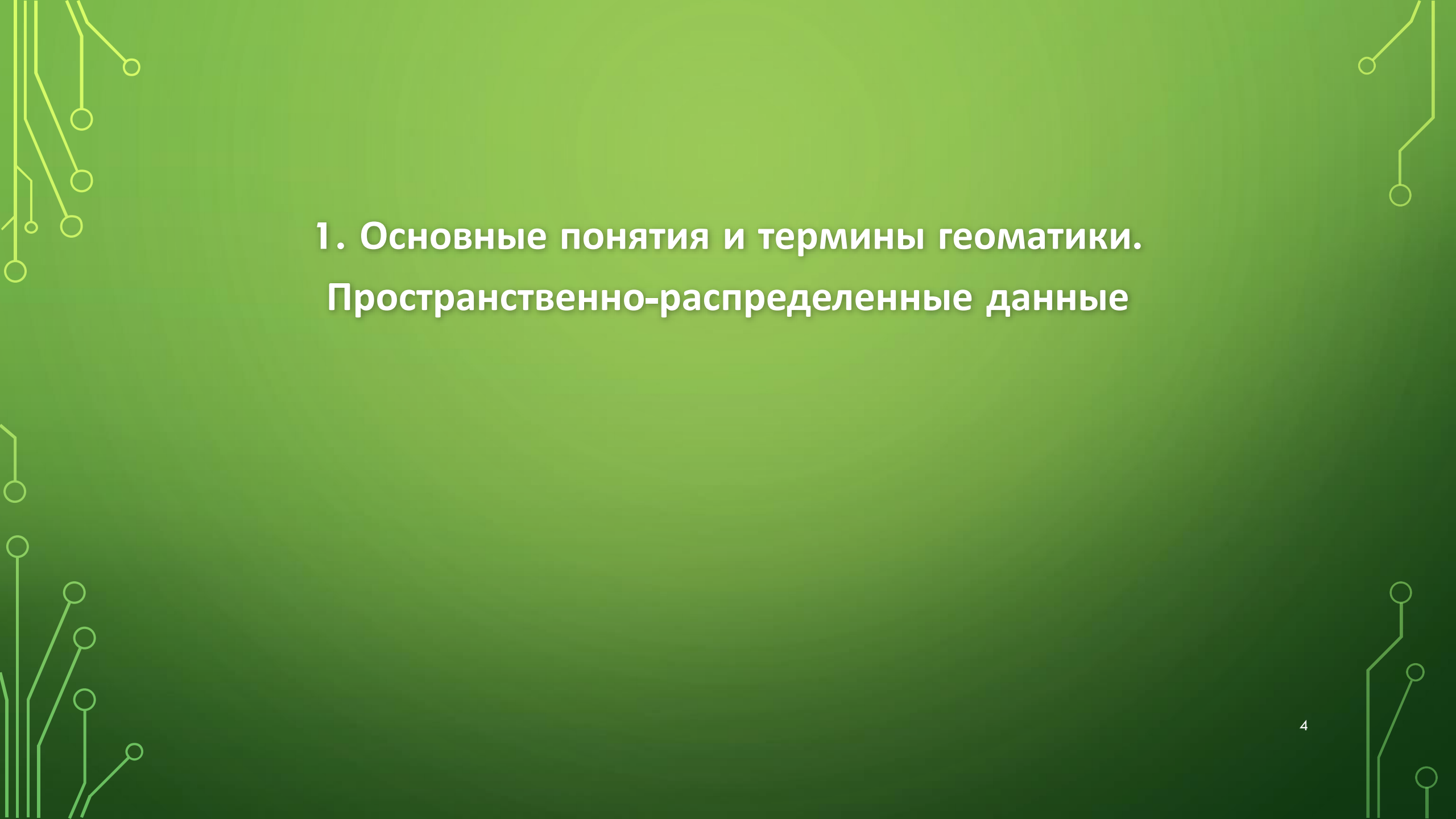
ТЕМА 1

ПЛАН ЛЕКЦИИ

- 1. Основные понятия и термины геоматики. Пространственно-распределенные данные.
- 2. Метрика в пространстве. Пространственное разрешение.
- 3. Сеть мониторинга и кластерность. Декластеризация данных.
- 4. Пространственная непрерывность данных. Стационарность данных.
- 5. Геостатистическое оценивание данных. Проверка качества модели методом кросс-валидации

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- 1. Введение в методы анализа данных по окружающей среде / М. Ф. Каневский [и др.] // Элементарное введение в геостатистику. Сер. Проблемы окружающей среды и природных ресурсов. – 1999. – № 11. – С. 1–12.
- 2. 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.
- 3. Геостатистический анализ данных в экологии и природопользовании (с применением пакета R): учеб. пособие / А. А. Савельев [и др.]. – Казань: Казанский университет, 2012. – 120 с. 21
- 4. Митчелл, Э. Руководство по ГИС-анализу / Э. Митчелл. – ESRI: 2000. – Ч. 1: Пространственные модели и взаимосвязи. – 170 с.
- 5. Морисита , М. Иб-индекс, мера рассеивания людей / М. Морисита // Исследования по экологии населения. – 1962. – Вып. 4(1). – С. 1–7.
- 6. Самсонов , Т. Е. Визуализация и анализ географических данных на языке R [Электронный ресурс] / Т. Е. Самсонов. – М.: Географический факультет МГУ, 2017. – Режим доступа: <https://tsamsonov.github.io/r-geo-course>

The background is a dark green gradient. In the corners, there are decorative elements consisting of light green lines that resemble circuit traces or data paths, ending in small circles.

1. Основные понятия и термины геоматики. Пространственно-распределенные данные

- Любые **данные** представляют собой информацию, представленную в виде, пригодном для обработки автоматическими средствами при возможном участии человека. Данные о пространственных объектах и (или) их наборах называют **пространственно-распределенными данными** (геоинформационными, геопространственными данными, геоданными.)
- Если кроме информации о местоположении присутствует информация о времени получения данных, то данные называют **пространственно-временными.**

Примеры пространственно-распределенных данных:

- распаханность земель сельскохозяйственного значения в пределах административного района;
- плотность загрязнения почвы радионуклидами;
- средняя стоимость земельного участка в пределах административно-территориальной единицы.

Примеры пространственно-временных данных:

- плотность загрязнения почвы по турам агрохимических обследований;
- кадастровая стоимость земель сельскохозяйственного назначения по турам кадастровой оценки;
- динамика количества земель крестьянских (фермерских) хозяйств за определенный период.

- **Растровая модель пространственных данных** – цифровое представление пространственных объектов в виде совокупности ячеек раstra (пикселей) с присвоенными им значениями класса объекта.
- Она предполагает позиционирование объектов с указанием их положения в соответствующей раstrу прямоугольной матрице единообразно для всех типов пространственных объектов (точек, мультиточек, линий, полилиний, полигонов и поверхностей, аннотаций и объектов размеров)

- **Каждой ячейке** такой растровой модели **соответствует** одинаковый по размерам, но **разный** по характеристикам **участок поверхности геопространственного объекта**.
- Данная модель разбивает всю изучаемую территорию (или поверхность) на отдельные элементы регулярной сетки или ячейки растра, каждая из которых содержит только одно атрибутивное значение.
- **Основное назначение растровых моделей геопространственных данных – непрерывное отображение поверхности.**

ПРИМЕРЫ РАСТРОВЫХ МОДЕЛЕЙ ДАННЫХ: ТОПОГРАФИЧЕСКАЯ КАРТА (А);
СПУТНИКОВЫЙ СНИМОК (Б); ОРТОФОТОМОЗАИКА (В)



а



б



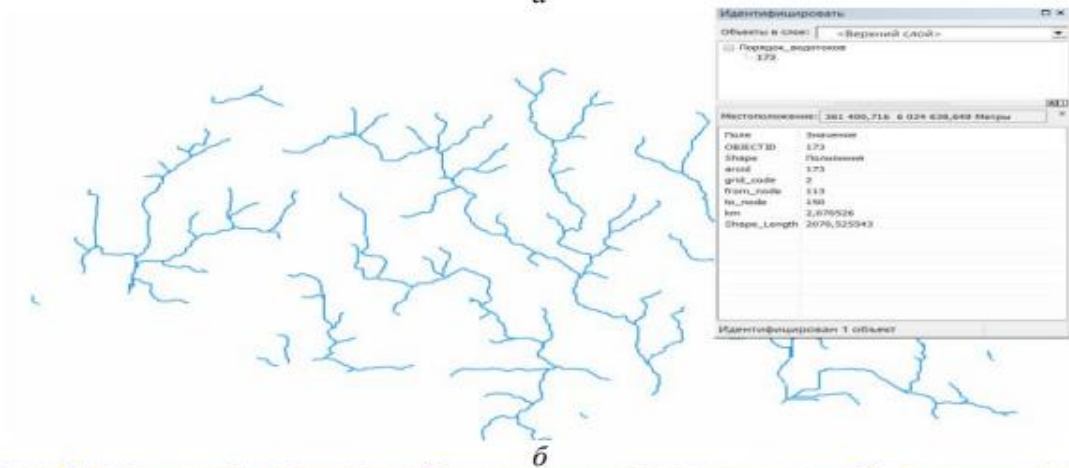
- Векторная модель пространственных данных – цифровое представление пространственных объектов в виде множества дискретных объектов с заданными характеристиками.

Векторные модели бывают:

- топологическими,
- нетопологическими



а



б



Примеры векторных моделей данных:
 полигональный слой (а);
 линейный слой (б);
 точечный слой (в)

- **Векторная топологическая модель** состоит из описания отдельных объектов, а также из описаний топологии – отношений отдельных объектов между собой.
- В векторных топологических моделях все объекты имеют определенное заданное размещение относительно друг друга и не могут произвольно размещаться в пространстве.

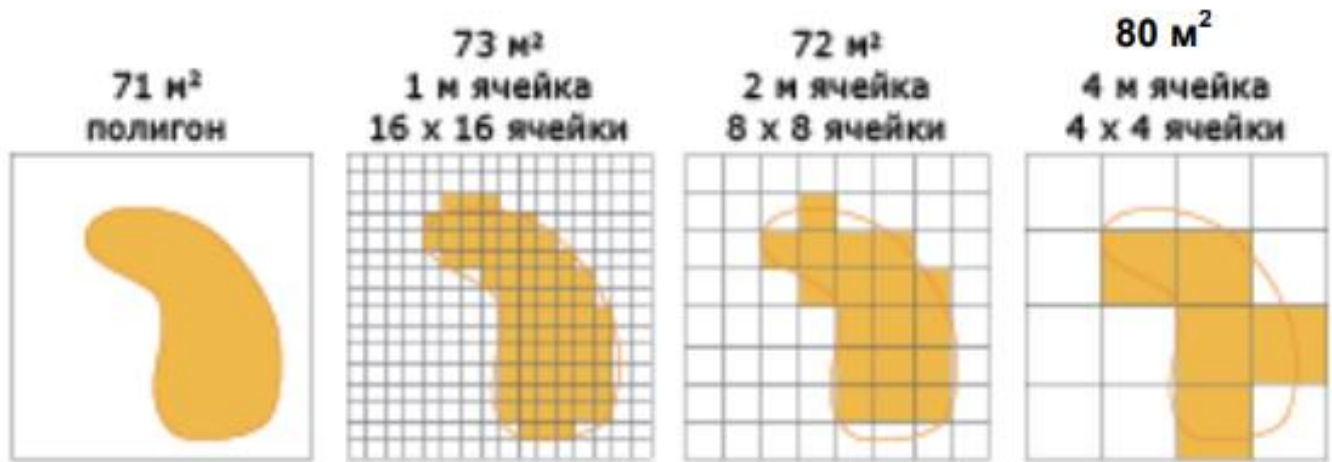
Векторная топологическая модель данных содержит три основных типа объектов: **узлы, дуги, полигоны.**

- **Векторная нетопологическая модель** содержит только информацию о пространственном местоположении объекта (переменной).
- В **векторных нетопологических** моделях все объекты являются полностью независимыми друг от друга и могут произвольно размещаться в пространстве.
- Пространственная информация представляется наборами элементарных графических примитивов – **точек, линий, полигонов**

- Наиболее широко распространенной векторной нетопологической моделью пространственно-распределенных данных является **шейп-модель**.
- В шейп-модели допустимы четыре вида данных: **точки, линии, полигоны и мультиточки**.
- В пределах одного слоя, представленного в виде шейп-модели, допускаются **объекты только одного вида**.
- Соответствующие слои называются **точечными, линейными, полигональными (площадными) и мультиточечными**.

- Геоматика оперирует регионализированными данными – измерениями, обладающими координатной привязкой.
- Координатная привязка может быть:
 - – **пространственной**, определяющей географическое положение измерения (географические координаты) в пространстве или его относительное положение по отношению к другим объектам (специальная координатная система для определенной местности);
 - – **временной**, определяющей время проведения измерения (абсолютное или относительное);
 - – **пространственно-временной**, т. е. и пространственной, и временной одновременно

- Одним из ключевых свойств пространственно-распределенных данных является их **пространственное разрешение**.
- Под **пространственным разрешением** понимается наименьший размер особенности, которую могут отражать данные и пространственные оценки



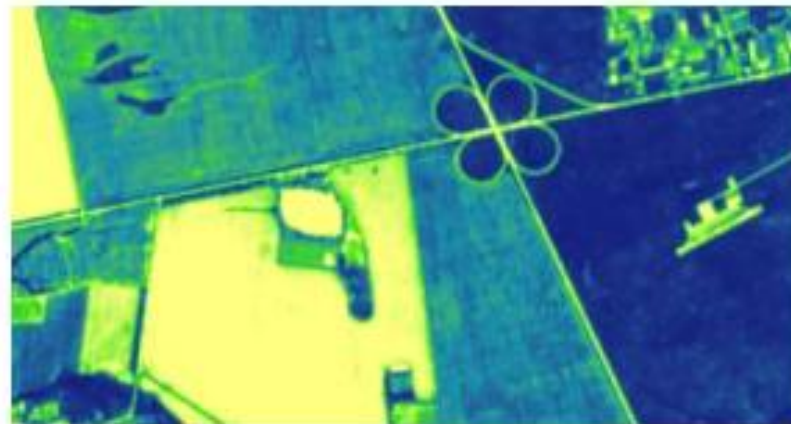
- **Меньший размер ячейки**
- **Выше разрешение**
- **Выше пространственная точность объектов**
- **Медленнее отображение**
- **Медленнее обработка**
- **Больше размер файла**

- **Больше размер ячейки**
- **Ниже разрешение**
- **Ниже пространственная точность объектов**
- **Быстрое отображение**
- **Быстрая обработка**
- **Меньше размер файла**

Изображение с высоким (слева) и низким (справа) пространственным разрешением



Размер ячейки 10 м, масштаб 1:50 000



Размер ячейки 10 м, масштаб 1:10 000

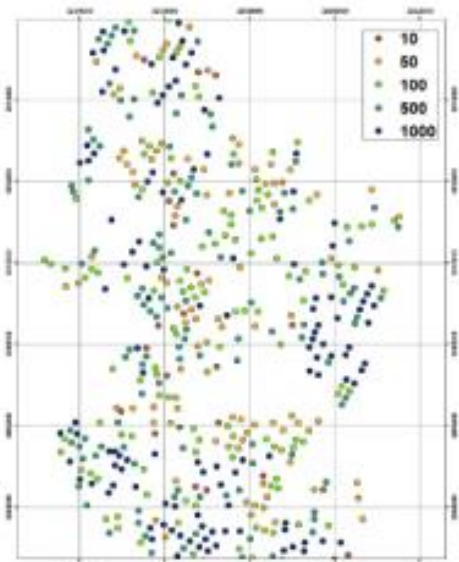
Влияние масштаба на отображение растровых данных с одинаковым пространственным разрешением

- **3. Сеть мониторинга и кластерность.
Декластеризация данных**

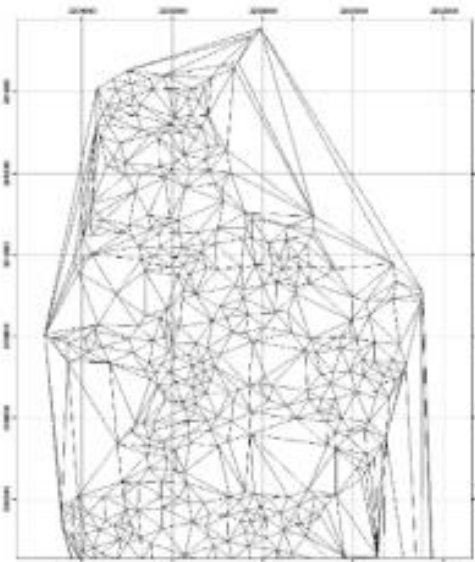
- **Мониторинг** - постоянное отслеживание какого-либо процесса или явления для установления его соответствия первоначальным предположениям или желаемому результату.
- **Цель мониторинга** – постоянный или периодический контроль за объектом наблюдения, предусматривающий возможность предупреждения негативных или нежелательных изменений состояния объекта.
- В процессе мониторинга создается **мониторинговая сеть**, состоящая из пунктов, в которых с установленной периодичностью выполняются те или иные наблюдения

- Мониторинговые сети бывают **регулярными** и **нерегулярными**.
- Если сеть мониторинга имеет зоны с заметно более высокой плотностью измерений, чем остальная область, то сеть мониторинга **кластерная**.
- **Кластер (cluster)** – область повышенной плотности точек измерений пространственной функции.

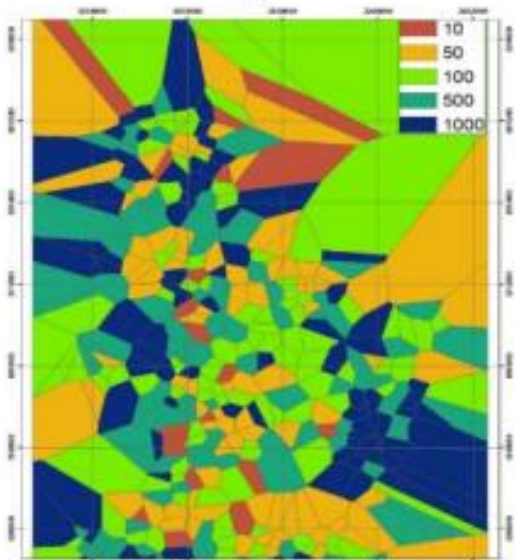
- Для визуализации сети наблюдений используют ряд способов:
- нанесение точек (пунктов наблюдений) на плоскость по пространственным координатам, при этом цвет нанесенной точки может соответствовать измеренной в ней величине (рис. а);
- для визуализации кластерной структуры сети мониторинга используется триангуляция Делоне – система треугольников с вершинами в точках измерений, непересекающимися ребрами и минимальным количеством тупоугольных треугольников - векторный способ отображения поверхности (рис. б).
- полигоны Вороного(разбиение Тиссена, ячейки Дирихле) или области влияния (рис. в)
- линейная интерполяция, в пределах которой значения функции вычисляются согласно геометрическим принципам (рис. г)



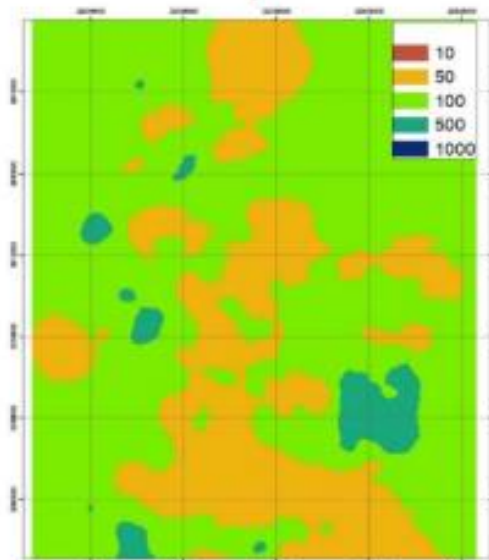
а



б



в



г

Диаграмма
 расположения:
 точек измерений (а),
 триангуляция сети
 мониторинга (б),
 полигоны Вороного (в)
 и контуры данных
 измерений по
 триангуляции (г)

- **4. Пространственная непрерывность данных.
Стационарность данных**

- **Пространственная непрерывность**, которая означает, что близко расположенные в пространстве измерения, скорее всего, будут иметь близкие значения, является важным свойством пространственного распределенных данных.
- Пространственная непрерывность описывается с помощью **корреляционных и ковариационных функций** (статистических моментов), выражающих меру этой непрерывности

- Для количественного описания пространственной непрерывности и моделирования пространственной корреляции, применяется **вариограмма** – статистический двухточечный момент второго порядка, характеризующий степень различия данных в зависимости от расстояния между ними (зависимость квадрата разности значений функции в точках от вектора расстояний между точками)

- **Пространственная непрерывность** непосредственно тесно связана со стационарностью данных.
- **Стационарность** (постоянство) – свойство процесса либо явления не изменять свои характеристики со временем.
- **Пространственная стационарность** в строгом смысле означает, что распределения случайной величины в двух различных зонах области распределения являются идентичными.
- **Пространственная нестационарность** заключается в меняющемся характере функции распределения данных в зависимости от местоположения точек измерения.

- **5. Геостатистическое оценивание данных. Проверка качества модели методом кросс-валидации**

- **Геопространственная статистика** – раздел математической статистики, который связан с численным описанием переменных, распределенных в географическом пространстве и времени.
- **Геостатистическое оценивание данных** - оценка статистических параметров, которые определяют пространственное и (или) временное распределение и зависимость соответствующих переменных.

- **Геостатистическое оценивание данных предусматривает:**
 - определение наличия ошибок и выбросов в данных;
 - оценку базовых статистических закономерностей;
 - проведение корреляционного анализа при наличии нескольких переменных.

- Для этого определяются:
- 1) минимальное и максимальное расстояния окрестности поиска ближайшего соседства, дающие возможность подобрать оптимальную величину окрестности поиска при пространственном моделировании (инструмент геообработки «Пошаговая пространственная автокорреляция»);
- 2) глобальный индекс Морана (I), позволяющий определить, имеет ли место явление кластеризации по отношению к атрибутивным данным, каковыми являются сведения о физико-химических и агрохимических свойствах почвы (инструмент геообработки «Пространственная автокорреляция»);

- 3) **общий индекс Getis-OrdG** для оценки общей структуры и тренда геоданных, а также степени кластеризации высоких и (или) низких значений выборки (инструмент геообработки «Высокая (Низкая) кластеризация»);
- 4) **индекс Getis-OrdG**, позволяющий установить наличие кластеризации данных с высокими и низкими значениями и выполнить анализ кластеров и выбросов для установления отличий атрибутивных значений каждого рабочего участка от его окрестности, а окрестности рабочего участка – от остальной территории (инструмент геообработки «Анализ горячих точек»)

- Дальнейший пространственный анализ предполагает исследование и моделирование пространственной корреляции между данными по одной или нескольким переменным.
- Пространственная корреляция — зависимость между значениями пространственно распределенной функции от взаимного расположения точек.

- Основной геостатистической моделью, которая в том или ином виде используется во всех методах геостатистики, является **кригинг**.
- **Кригинг (kriging)** – геостатистический метод пространственной интерполяции, основанный на линейной регрессии, обладает несмещенностью и минимальной дисперсией оценки.

- **Валидация** – проверка качества работы модели с помощью данных, неиспользованных для ее настройки.
- **Кросс-валидация (cross-validation)** – наиболее простой и часто используемый метод подбора оптимальных параметров модели интерполяции с помощью оценки значения в точке измерения без учета самого измерения в этой точке.

- **Выполняется кросс-валидация следующим образом:**
 - – из базы данных временно изымается одна точка, и для нее проводится оценка значения;
 - – полученное значение сравнивается с известным, и вычисляется невязка – разница между измеренными и оцененными значениями: $\Delta Z(x) = Z(x) - Z^*(x)$;
 - – первые два шага проводятся для всех точек базы данных. Полученные невязки $\Delta Z(x)$ могут быть графически представлены в виде карты (карты невязок), по которой можно посмотреть, в каких зонах метод интерполяции срабатывает лучше, а в каких хуже.

- Точность методов интерполяции определяют по величине средней ошибки (ME), среднеквадратичной ошибки (RMSE), средней нормированной ошибки (MSE), средней стандартизированной ошибки (ASE) и среднеквадратичной нормированной ошибки (RMSS):

$$ME = \frac{\sum_{i=1}^N (Q_i - S_i)}{N};$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (Q_i - S_i)^2}{N}};$$

$$MSE = \sqrt{\frac{\sum_{i=1}^N \frac{(Q_i - S_i)}{\sigma^2(S_i)}}{N}};$$

$$ASE = \sqrt{\frac{\sum_{i=1}^N \sigma^2(S_i)}{N}};$$

$$RMSSE = \frac{RMSE}{\Delta},$$

где Q_i – наблюдаемое значение;

- S_i – предсказанное значение;
- N – объем выборки;
- σ – дисперсия выборки;
- Δ – диапазон, равный разнице между максимальным и минимальным наблюдаемыми значениями.

