

Київський національний університет будівництва і архітектури
Кафедра геоінформатики і фотограмметрії

Л.4 Геостатистика

доц., к.т.н. Горковчук Ю.В.

FACULTY
Geoinformation
systems and territory
management

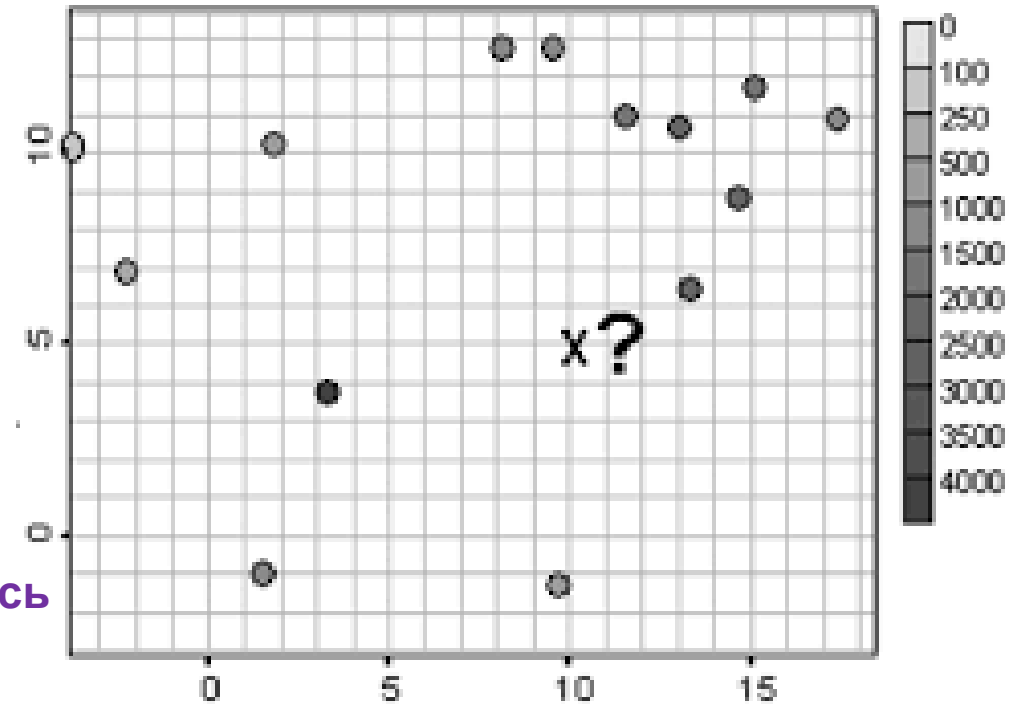


Концепція геостатистичного аналізу

- Є область, на якій проведено ряд вимірювань деякої величини Z – набір точок (x, y)
- Але є ділянки, не покриті вимірами, де значення величини Z необхідно визначити.
- Дані вимірювань дискретні та просторово неоднорідно розподілені

Необхідно

- оцінити значення в точці, де вимірювання не проводились
- створити карту, побудувати ізолінії
- знайти значення змінної, по якій мало вимірювань, використовуючи значення іншої корелятивною з нею змінної, по якій проведено багато вимірювань
- визначити ймовірність того, що значення змінної перевищать заданий поріг
- отримати набір рівно можливих стохастичних просторових реалізацій розподілу змінної



» Концепція геостатистичного аналізу

Геостатистика – це розділ статистики, що вивчає технології та методи для аналізу, оброблення і подання просторово-розподіленої (та/або просторово-часової) інформації на основі статистичних методів

Геостатистика моделює закономірності розподілу об'єктів, явищ і процесів в географічному просторі з врахуванням їх місцеположення

Предметом аналізу геостатистики є просторові змінні (або районовані змінні **regionalised variables**) – змінні з координатною прив'язкою: кількість опадів, щільність населення в деякій географічній області, ступінь забруднення ґрунту, тощо

Передбачається, що дані вимірювань $z(x_i)$ є реалізаціями випадкових змінних $Z(x_i)$, які описуються деякими функціями розподілу.

Випадкова функція визначається як набір зазвичай залежних між собою випадкових змінних $Z(x_i)$, по одній для кожного місця розташування x_i в даній області.

Будь-якому набору з N місць розташування $\{x_k, k=1, \dots, N\}$ можна поставити у відповідність N випадкових змінних $\{Z(x_1), \dots, Z(x_N)\}$, які характеризуються N -мірною умовною функцією розподілу:

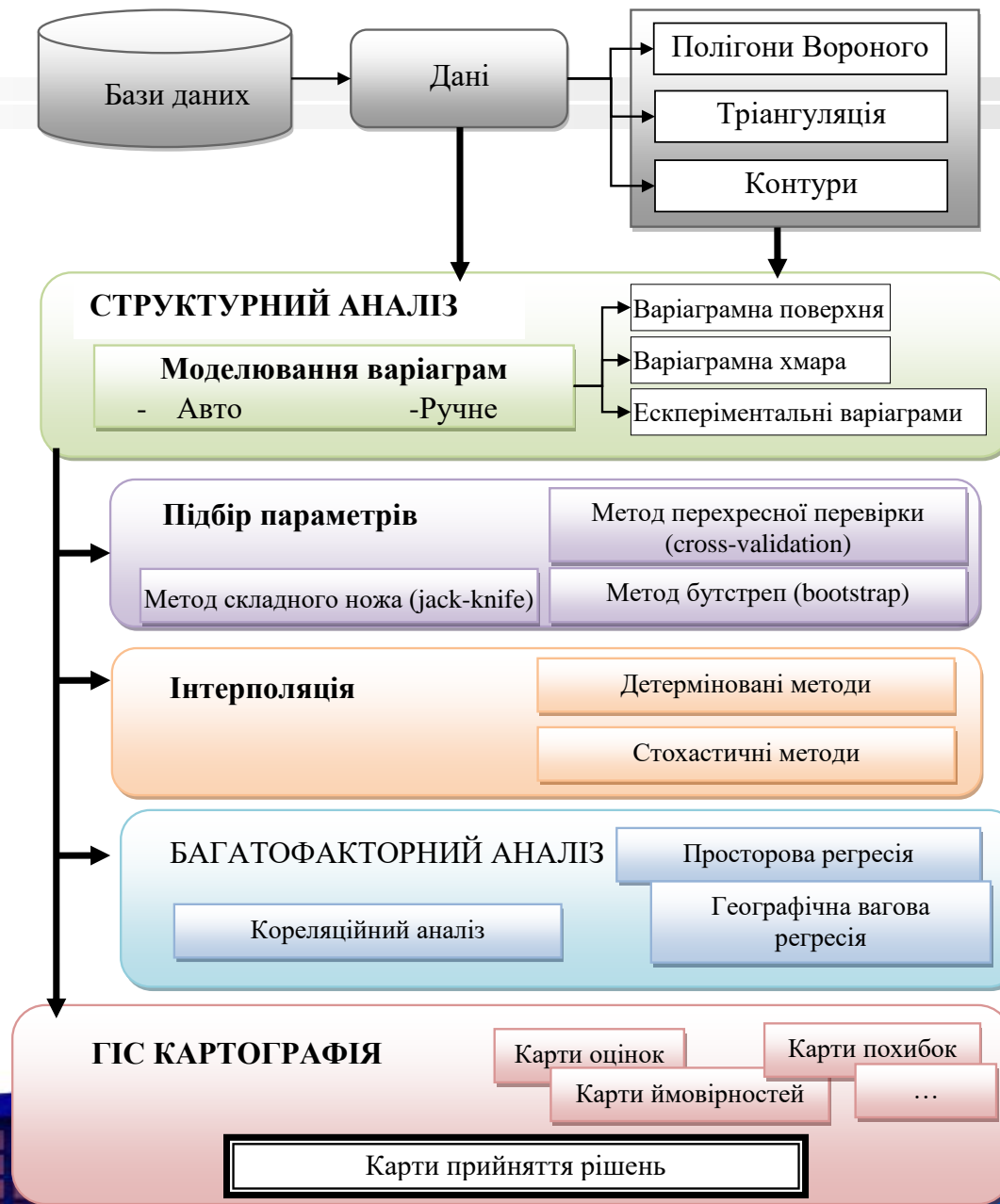
$$F(x_1, \dots, x_N; z_1, \dots, z_N) = \text{Prob} \{Z(x_1) \leq z_1, \dots, Z(x_N) \leq z_N\}$$



Концепція геостатистичного аналізу

Етапи геостатистичного аналізу

1. **Статистичний аналіз даних** - дослідження даних (визначення статистичних параметрів, перевірка на нормальність, виявлення тренда)
2. **Дослідження та моделювання просторової кореляції** - побудова варіограмної моделі
3. **Просторова оцінка** – інтерполяція, крігінг, тощо
4. **Підбір оптимальних параметрів моделювання** - оцінка точності моделювання
5. **Подання результатів геостатистичного аналізу** - карти ймовірностей, карти середніх оцінок



Статистичний аналіз даних

Основне завдання кореляційного аналізу - визначення зв'язку між випадковими змінними і оцінювання його інтенсивності та напрямку. Основне завдання регресійного аналізу є встановлення форми і вивчення залежності змінних.

- **Кореляція** (лат. correlatio - співвідношення) – це статистична залежність між випадковими величинами, що носить імовірнісний характер

Головні завдання кореляційного аналізу:

- 1) оцінка за вибірковими даними коефіцієнтів кореляції;
- 2) перевірка значущості вибіркових коефіцієнтів кореляції або кореляційного відношення;
- 3) оцінка близькості виявленого зв'язку до лінійного;
- 4) побудова довірчого інтервалу для коефіцієнтів кореляції.

- **Коефіцієнт кореляції** – міра залежності між випадковими величинами $[-1, 1]$
- **Індекс кореляції** – кореляційне відношення - $R \in [0, 1]$.

- **Регресія** - встановлення форми і вивчення залежності змінних

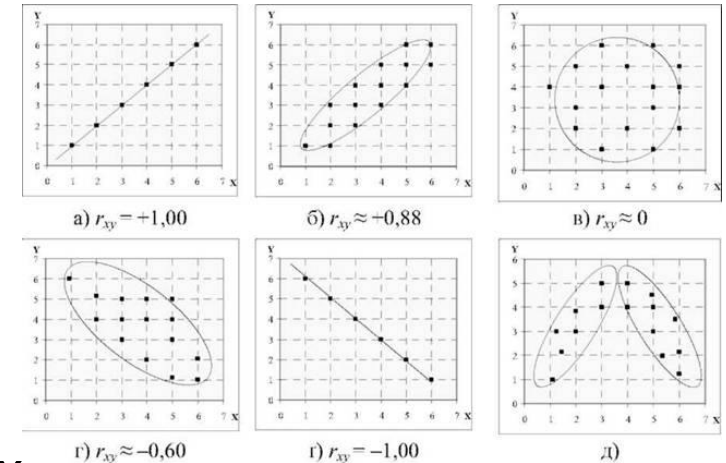
R дозволяє за величиною однієї ознаки (змінна x) знаходити середні (очікувані) значення іншої ознаки (змінна Y), зв'язаної з x кореляційно. Оскільки в дослідженнях конкретний вид взаємозв'язків невідомий, одне з головних завдань регресійного аналізу полягає у доборі відповідного виразу $Y = F(X)$, графік якого проходить через емпіричні точки (або досить близько до них) і таким чином зв'язує змінні x і Y .

- У прямокутній системі координат рівняння лінійної регресії:

$$y(x) = a_1x + b_1 \quad x(y) = a_2y + b_2$$

- **Головною властивістю рівняння регресії є те, що вона (регресія) мінімізує суму квадратів (дисперсію) відхилень точок на лінії від експериментальних даних**

$$\sum_{i=1}^n (y_i - y(x_i))^2 = \sum_{i=1}^n (\Delta y_i)^2 \rightarrow \min.$$



Статистичний аналіз даних

- **Просторова автокореляція (autocorrelation)** – це статистична міра, що описує як змінюється одна властивість з залежності від інших

Процес просторової автокореляції є вимірюванням ступеня, за яким набір просторових об'єктів та значення пов'язаних з ним даних можуть бути кластеризовані в просторі (додатня просторова автокореляція) або дисперговані (від'ємна просторова автокореляція)

Загальний вид просторової залежності :

$$\text{cov}(y_i, y_j) = E(y_i, y_j) - E(y_i)E(y_j) \neq 0, \quad \text{для } i \neq j$$

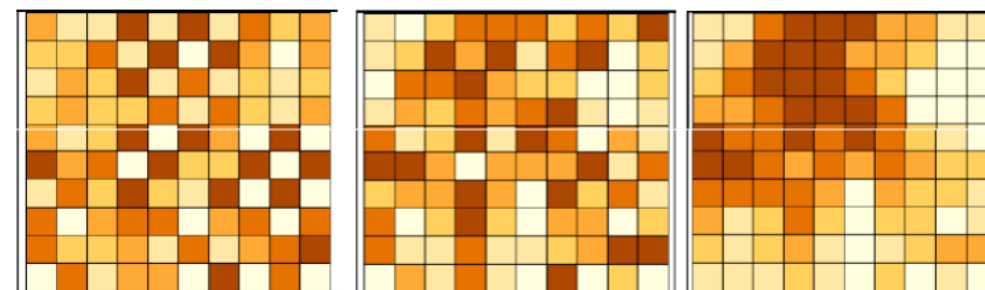
де E – математичне очікування події y_i, y_j .

В матричному виді просторова автокореляція має вид:

$$\Gamma_{ij} = \sum_{i=1}^n \sum_{j=1}^n W_{ij} Y_{ij}$$

де Γ є мірою просторової автокореляції для n просторово прив'язаних спостережень, що складається з просторової матриці ваг W_{ij} , що подає просторові відносини кожного місця i до всіх інших місцеположень j .

Матриця Y_{ij} подає непросторові відносини реалізацій змінної Y в місцях i з усіма іншими реалізаціями на всіх інших положеннях j .



Від'ємна просторова автокорреляція: поруч групуються одиниці, що володіють несхожими характеристиками

Випадкова

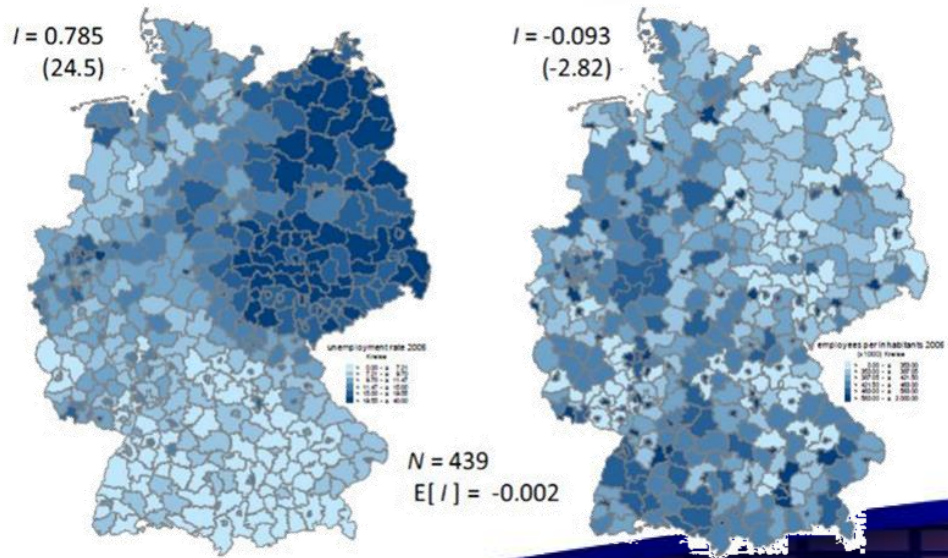
Додатня просторова автокорреляція: утворення кластерів у просторі спостережень з близькими показниками

Статистичний аналіз даних

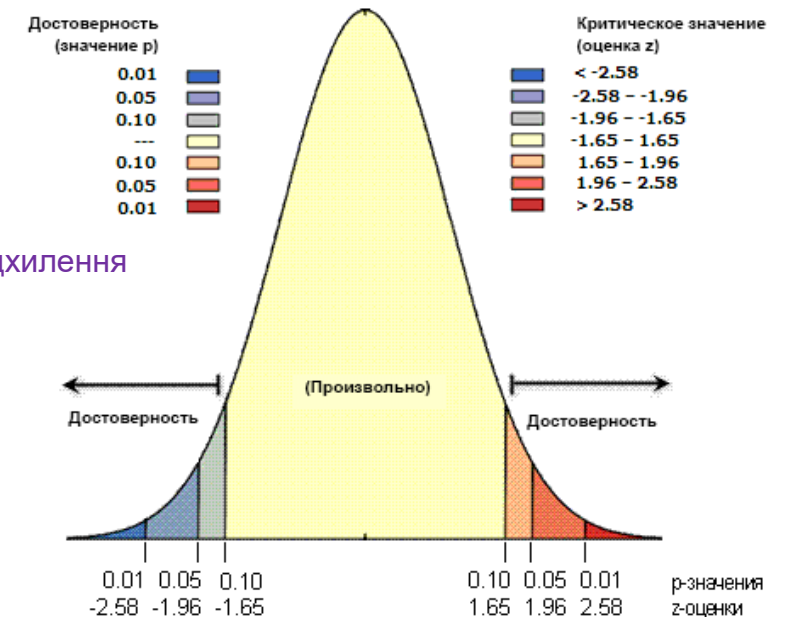
Показники автокореляції

- Глобальні показники оцінюють загальну структуру та тренд даних та є найбільш ефективними, коли просторові закономірності стійкі в межах області інтересу (Загальний індекс I Морана)
- Локальні показники оцінюють кожен об'єкт в контексті сусідніх об'єктів та порівнюють локальні ситуації з глобальною ситуацією

Діаграма Морана I



р-значення - вірогідність
Z-оцінка - стандартне відхилення



Статистичний аналіз даних

- **Просторова регресія** - метод просторового аналізу, що створює локальну модель змінної або процесу на основі рівняння регресії для кожного просторового об'єкту в наборі

Модель Географічно зваженої регресії має вид:

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^p \beta_k(u_i, v_i) \cdot x_{ik} + \varepsilon_i,$$

де (u_i, v_i) – місцеположення i -го об'єкту (координати i -ї точки);
 y_i – значення результативного признаку; x_{ik} – значення k -ї описової змінної для i -го об'єкту; $\beta_k(u_i, v_i)$ – невідомі коефіцієнти; ε_i – регресійні нев'язки; $i = 1, 2, \dots, n$; $k = 1, 2, \dots, p$.

Методи обчислювання вагових коефіцієнтів

- метод адміністративно-територіального поділу;
- метод рухомого фіксованого вікна;
- метод фіксованого ядра;
- метод адаптивних ядер.

Вектор оцінок коефіцієнтів для кожного місцеположення i обчислюється по формулі:

$$\hat{\beta}(u_i, v_i) = [X^T W(u_i, v_i) X]^{-1} X^T W(u_i, v_i) Y,$$
$$W(u_i, v_i) = \begin{pmatrix} w_{i1} & 0 & \dots & 0 \\ 0 & w_{i1} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & w_{i1} \end{pmatrix}$$

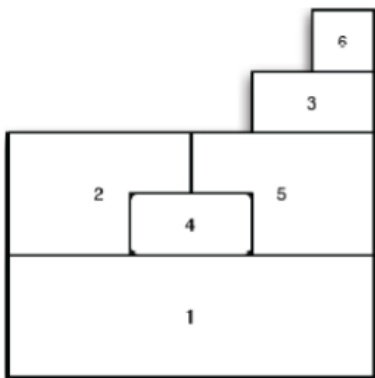
де $W(u_i, v_i)$ – матриця вагових коефіцієнтів розмірності $(n \times n)$, елементи якої визначають ступень впливу сусідів j на залежність місцеположення i .

Статистичний аналіз даних

Просторова матриця ваг – відображає припущення, що розподіл певних явищ та об'єктів (або їх інтенсивність) залежить від відстані між ними:

$$w_{ij} = 1 / d_{ij}^2$$

Тобто, вага, яку має клітина (i, j) є зворотною відстані d між двома точками i та j , (за методом обернених відстаней)



$$W = \begin{bmatrix} 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

$$W^* = \begin{bmatrix} 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

$$W = \begin{bmatrix} 0 & 1/3 & 0 & 1/3 & 1/3 & 0 \\ 1/3 & 0 & 0 & 1/3 & 1/3 & 0 \\ 0 & 0 & 0 & 0 & 1/2 & 1/2 \\ 1/3 & 1/3 & 0 & 0 & 1/3 & 0 \\ 1/4 & 1/4 & 1/4 & 1/4 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

Нормалізація матриці ваг по рядку

Варіаграма

- це статистичний момент другого порядку випадкової функції $Z(x)$, що застосовують для кількісного опису просторової безперервності:

$$\gamma(h) = \frac{1}{2n} \sum_{i=1}^n [z(x_i) - z(x_i + h)]^2$$

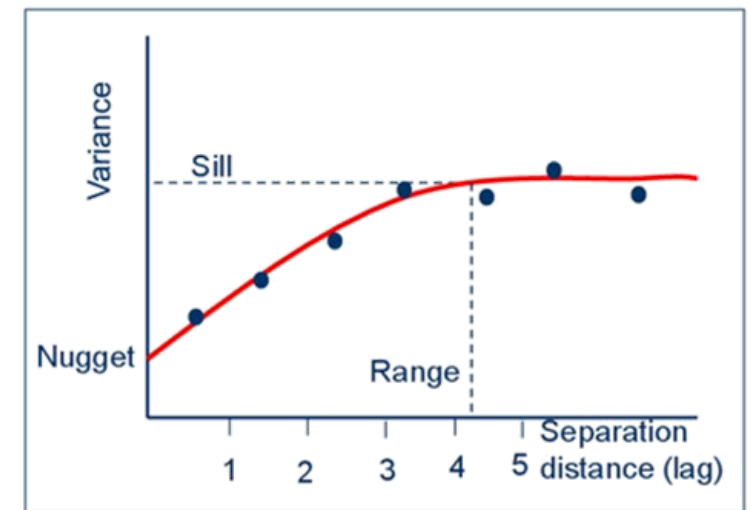
Варіаграма описує кореляційний зв'язок між двома випадковими величинами $Z(x)$ та $Z(x+h)$, що розташовані в просторі на відстані h . Сила зв'язку може змінюватися зі зміною відстані між точками та зміною напрямку вектора h . Для зафіксованого напрямку варіограма показує, як змінюються значення величини, що досліджується при збільшенні відстані між точками.

Елементи варіаграми

радіус кореляції (радіус залучення або просто радіус (**Range**)) –максимальні значення змінної, при яких варіаграма не збільшується, тобто втрачається залежність різниці значень у двох місцеположеннях від відстані між ними

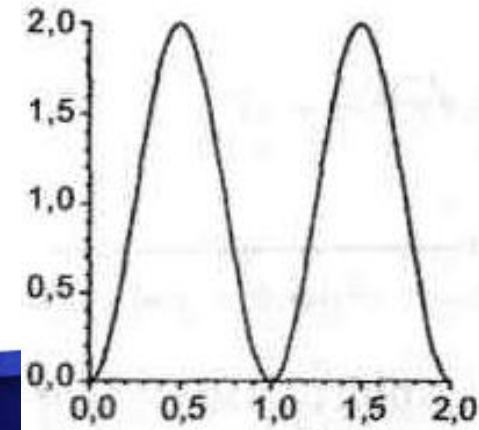
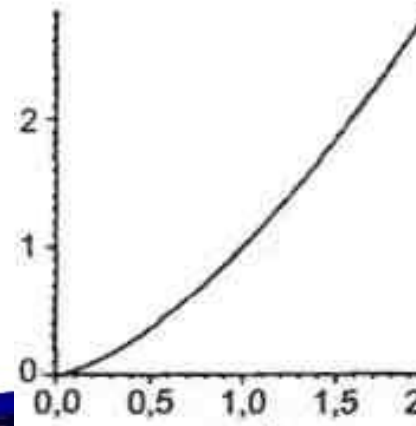
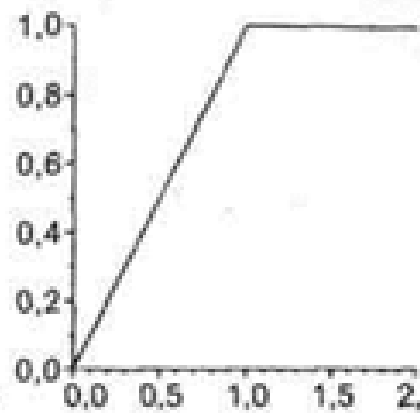
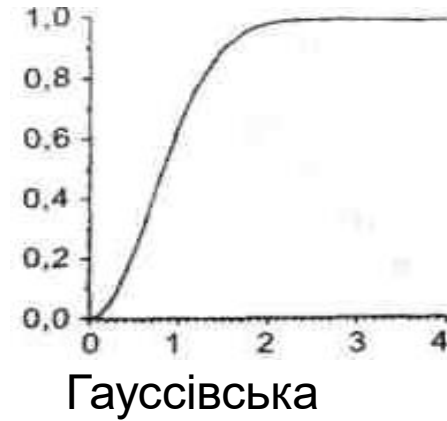
поріг (sill) - величина «насичення» варіограми

Оцінка просторово некорельованого шуму, (Nugget , англ. «самородок», (c_0)) - це залишкова варіація, тобто дисперсія похибок вимірювань, а також тих просторових змін, які мають характерний розмір, набагато менший, ніж крок випробування



Варіаграма

- Теоретичні варіограмні моделі



Вариограма

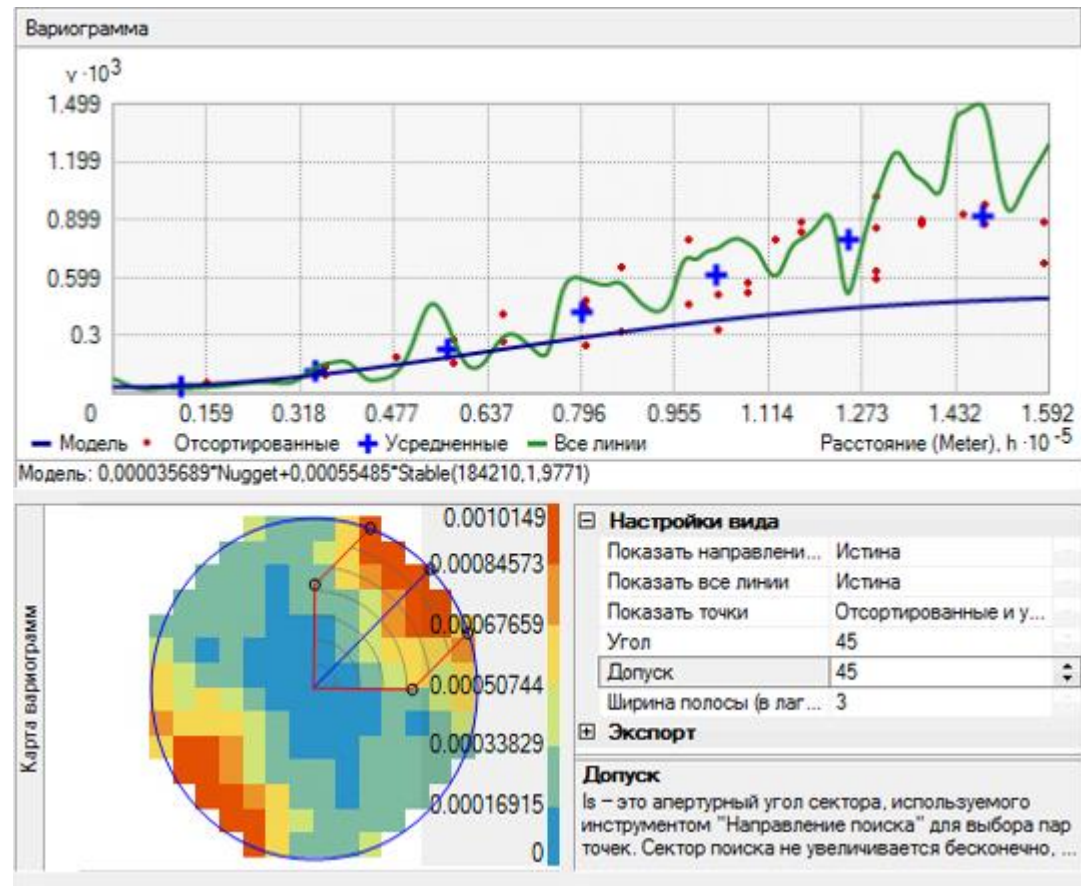
Моделювання варіограми - це ключовий крок між просторовим описом та інтерполяцією даних

Модель варіограми, за емпіричним даним, має:

- проходить через центр хмари об'єднаних в біни значень (червоні точки)
- проходить якомога ближче до середніх значень (сині перехрестя)
- проходить якомога ближче до ліній (зеленим)

Имейте в виду, что ваши знания явления могут диктовать форму модели, а также ее самородок, диапазон, частичный порог и значения анизотропии, даже если модель не слишком хорошо соответствует эмпирическим данным (помните, что эмпирические данные — просто образец реального явления, которое вы хотите моделировать, и не могут полностью представлять все его пространственные и статистические аспекты)

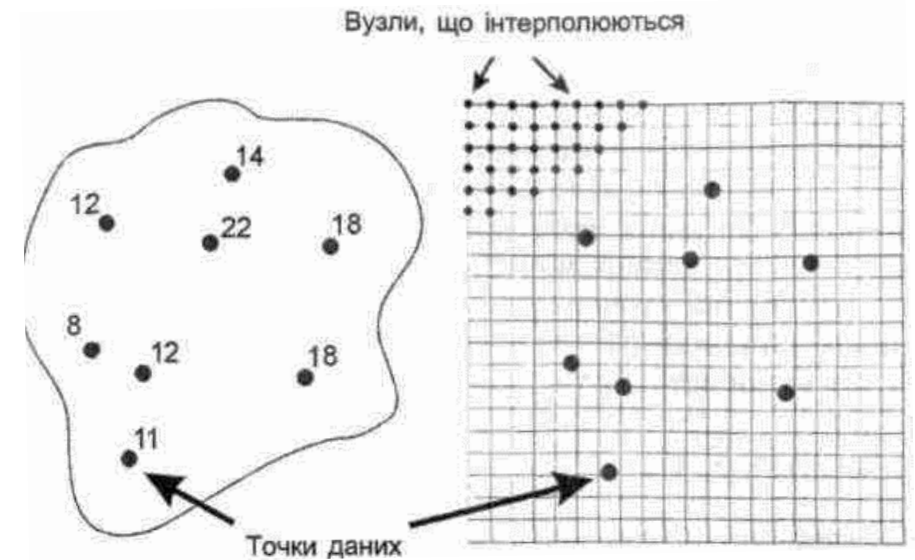
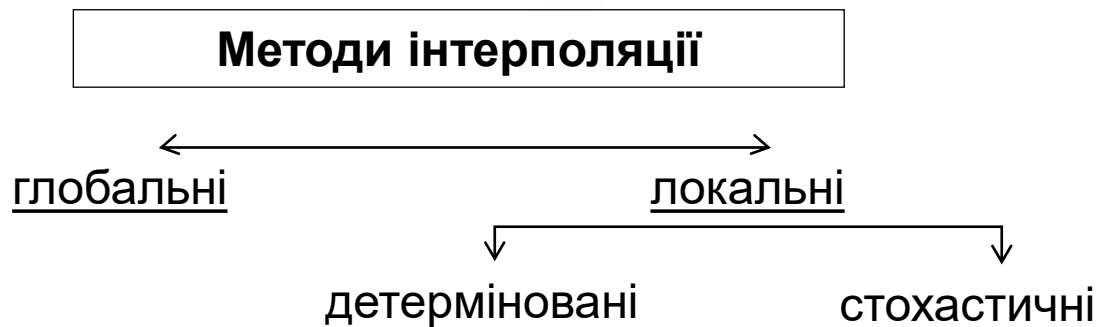
<https://desktop.arcgis.com/ru/arcmap/10.3/guide-books/extensions/geostatistical-analyst/fitting-a-model-to-the-empirical-semivariogram.htm>



Інтерполяція

- спосіб знаходження проміжних значень величини за наявним дискретним набором відомих значень

Завданням просторової інтерполяції є побудова на основі мережі вихідних точок суцільної поверхні з заданим розміром кроку сітки вузлів



процедура побудови регулярної прямокутної сітки числових значень на основі мережі нерегулярних точок одержала назву **gridding**, масив інтерпольованих по регулярній сітці даних — **grid**, вузли інтерпольованої сітки — **node**.

Інтерполяція

Методи

Глобальні

- класифікації з використанням зовнішньої інформації;
- поліноміальна регресія з геометричними координатами;
- регресійні моделі.

Локальні

детерміновані
розраховують значення точок на основі вимірних значень, що потрапляють в близькість до інтерпольованої точки, за допомогою математичних формул, які визначають згладженість результуючої поверхні

- метод найближчого сусідства (полігонів Тиссена-Вороного);
- метод середнього зважування обернено пропорційно відстані (дистанції);
- метод сплайнів.

стохастичні
ґрунтуються на статистичних моделях, що включають аналіз автокореляції (статистичні відносини між вимірними точками)

- геостатистичне моделювання/ крігінг

Вибір методу інтерполяції

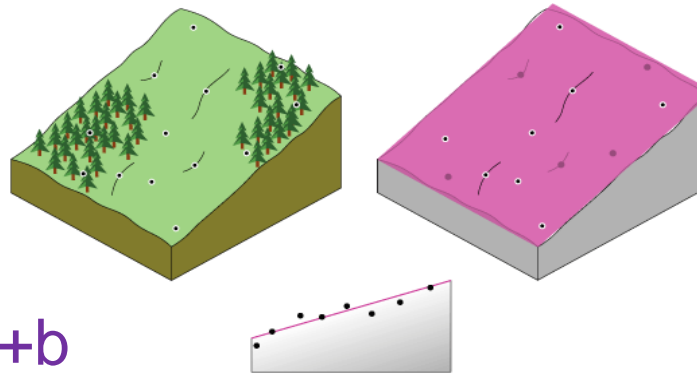
Кількість спостережень (точок)	Рекомендований метод інтерполяції
Близько 10	методи крігінга сплайн
Менш 250	крігінг із лінійною варіограмною моделлю
250-1000	лінійна інтерполяція крігінг
Понад 1000	крігінг обернена зважена відстань

Інтерполяція

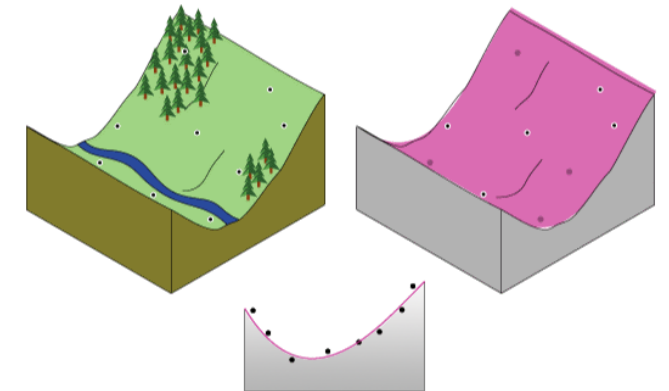
- Тренд - метод глобальної інтерполяція, який створює згладжену поверхню, що задана математичною (напр., поліноміальною) функцією для вхідних опорних точок. Вид тренду становлюють за графічним зображенням даних ряду, шляхом усереднення показників змінної

Два загальних типа тренд-інтерполяції

- Лінійний $y = a x + b$
- Логарифмічний $y = c \ln x + b$



Поліном 1-ого порядку



Поліном 2-ого порядку

» Интерполяція

- **Сплайн-інтерполяція** - ґрунтується на використанні для інтерполяції в околах даного вузла кускових поліноміальних функцій, які мають назву «функції сплайнів»

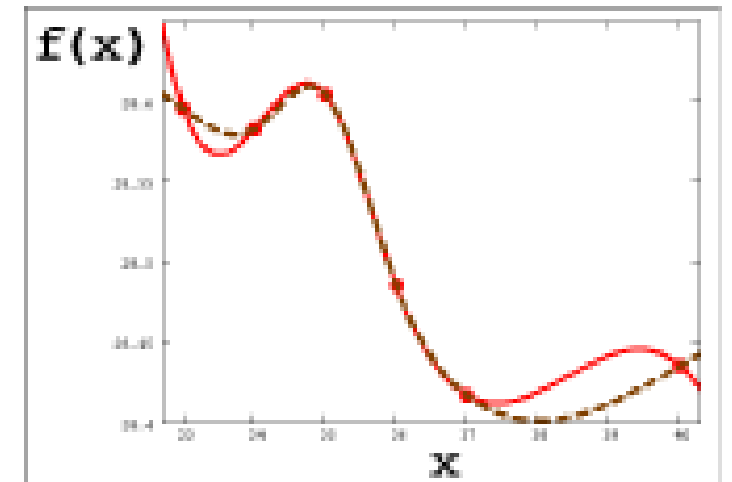
Типи сплайнів: лінійний, кубічний сплайн, сплайн Ерміта, Сплайн Катмулла-Рома, тощо

$$f(x) = ax^3 + bx^2 + cx + d$$

Коефіцієнти a , b , c , d розраховуються незалежно для кожного проміжку інтерполювання, виходячи зі значень y_i в сусідніх точках

+ висока швидкість
оброблення обчислювального
алгоритму
+ гладкість інтерпольованої
поверхні

- неможливість коректного
відображення різких змін у
поверхні
- наперед невідома точність
просторової інтерполяції



Інтерполяція

- Обернена зважена відстань Inverse Distance Method (IDM) - полягає в тому, що значення атрибута z в довільній точці простору, в якій не проводилися вимірювання, є **середнім зваженим по відстані із значень в точках вимірювань**, розміщених по сусідству в межах певного радіуса (або вікна) навкруги цієї точки. Ваги точок вимірювань є обернено пропорційними відстані до даної точки

$$z(x_j) = \frac{\sum_{i=1}^n z(x_i) d_{ij}^{-r}}{\sum_{i=1}^n d_{ij}^{-r}}$$

де x_j – точки (вузли) інтерполяції;

x_i – точки з відомими значеннями;

d_{ij} – відстані між точками з відомим значенням і точкою оцінювання;

r – показник ступеня інтерполяції;

n – кількість точок з відомим значеннями, що потрапляють в радіус оцінювання.

+ гнучкість

- відсутність методики обґрунтування виду вагової функції

+ легкість

- наперед невідома точність просторової інтерполяції

Інтерполяція

- Крігінг / Kriging — це метод знаходження кращої незміщеною лінійної оцінки (тобто найменшої дисперсії) значень точок

Всі моделі сімейства крігінга так чи інакше зводяться до лінійної регресійної оцінки:

$$V^*(x) - m(x) = \sum_{i=1}^n w_i(x) [V(x_i) - m(x_i)]$$

де $w_i(x)$ - ваги, які надають даним $V(x_i)$, які в свою чергу є реалізаціями просторової змінної V . Значення $m(x)$ та $m(x_i)$ є математичними очікуваннями (середніми) просторових змінних $V(x)$ та $V(x_i)$.

Кількість даних n , що використовуються для оцінювання, як і їх ваги можуть змінюватися в залежності від точки оцінювання x

Випадкова функція $V(x)$ зазвичай розкладається на дві компоненти: детерміністичний тренд $m(x)$ та випадкову нев'язку $R(x)$:

$$V(x) = R(x) + m(x)$$

Компонента нев'язки $R(x)$ моделюється як стаціонарна випадкова функція з нульовим математичним очікуванням $m_R(x)$ та коваріацією $C_R(h)$:

$$m_R(x) = E\{R(x)\} = 0 \quad \text{та} \quad \text{Cov}\{R(x), R(x+h)\} = E\{R(x)R(x+h)\} = C_R(h)$$

Таким чином, математичне очікування просторової змінної V в точці x буде дорівнювати значенню тренда:

$$E\{V(x)\} = m(x)$$

Інтерполяція

Крігінг

Умови:

- 1) Незміщені оцінки $V^*(x)$ в точці x (середнє значення похибки оцінювання дорівнює нулю): $R^*(x) = V^*(x) - V(x)$
- 2) Мінімізація варіації помилки, що дає «найкращу» в статистичному сенсі оцінку: $R^2(x) = E\{(R^*(x) - m_R(x))^2\}$

Розрізняють три типи Крігінгу, в залежності від моделі тренда $m(x)$:

1. Простий крігінг (Simple Kriging, SK) – середнє $m(x)$ є постійним та відомим в області дослідження S :

$$m(x) = m = \text{const, відомо } \forall x \in S$$

2. Звичайний крігінг (Ordinary Kriging, OK) – значення середнього в області невідомо, але постійно:

$$m(x) = m = \text{const, невідомо } \forall x \in A(x)$$

3. Універсальний крігінг (Universal kriging, UK) припускає, що невідоме середнє значення $m(x)$ гладко змінюється по всій області дослідження S . Компонента тренда моделюється як лінійна комбінація відомих функцій $f_k(x)$:

$$m(x) = \sum_{k=0}^k a_k(x) f_k(x)$$

де коефіцієнти $a_k(x)$ визначаються із системи рівнянь універсального крігінга

» Багатофакторний аналіз

- **предметом багатофакторного аналізу є дослідження впливу декількох різноманітних показників (факторів) на об'єкт дослідження з метою оцінити міру впливу кожного**
 - Функціональний - такий зв'язок, при якому будь-яке явище (об'єкт або його характеристика) повністю визначається однією або декількома факторами
 - Кореляційний - такий зв'язок явищ, при якому на кожне з них впливає велика кількість різноманітних чинників
 - За напрямком зв'язку прийнято розрізняти **пряму та зворотну форми зв'язку**
 - При аналітичному вираженні у статистиці розрізняють **прямолінійний і криволінійний зв'язки**

Методи багатокритеріальної оптимізації

призначення коефіцієнтів ваги кожного з критеріїв;

застосування узагальненого критерію цільової функції, який враховував би всі критерії з відповідними ваговими коефіцієнтами;

застосування методу експертних оцінок (метод Дельфи) та методів, які полягають у визначенні ступеня узгодженості між експертами впливу різних факторів на критерії, встановленні їх значущості, наприклад через коефіцієнт конкордації в порівнянні її з критерієм Пірсона та ін

GIS Картографія

- **Геовізуалізація** - проектування і генерація *геозображень* та іншої графіки на пристроях відображення (переважно на екрані дисплея) на основі вихідних цифрових даних та правил і алгоритмів їх перетворення та подання

- **Геозображення** - будь-яка просторово-часова масштабна генералізована модель просторових об'єктів або процесів, подана в графічній образно-знаковій формі (дво- або тривимірній, анімованій тощо).

Цифрова картографія

- цифрові карти
- Інтерактивні карти
- 3d сцени
- діаграми
- таблиці
- види з показниками часу
- види мережових відношень
- ...

Геоінформаційне картографування

- ведення картографічних баз даних
- тематичний моніторинг
- ведення кадастрів
- створення тематичних карт
- ...





Список літератури

1. Каневский М.Ф., Демьянов В.В., Савельева Е.А., Чернов С.Ю., Тимонин В.А. Элементарное введение в геостатистику серия Проблемы окружающей среды и природных ресурсов, № 11, ВИНТИ, Москва, 1999
2. В. Демьянов, Е. Савельева. Геостатистика. Теория и практика. Издательство «Наука», Москва, 2010, 327 стр., ISBN 978-5-02-037478-2
3. Getis, Arthur, and Jared Aldstadt. "Constructing the Spatial Weights Matrix Using a Local Statistic" (Создание матрицы пространственных весов с использованием локальных статистических показателей) *Geographical Analysis* 36(2): 90–104, 2004.
4. Справка ArcGIS 10.1 / <http://resources.arcgis.com/ru/help/main/10.1/index.html#/na/009z00000076000000/>