

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ДЕРЖАВНИЙ ВИЩИЙ НАВЧАЛЬНИЙ ЗАКЛАД
«НАЦІОНАЛЬНИЙ ГІРНИЧИЙ УНІВЕРСИТЕТ»



ГЕОЛОГОРОЗВІДУВАЛЬНИЙ ФАКУЛЬТЕТ

Кафедра геоінформаційних систем

К.Л. Сергєєва

ГЕОСТАТИСТИКА

**МЕТОДИЧНІ РЕКОМЕНДАЦІЇ
ДО ЛАБОРАТОРНИХ ЗАНЯТЬ З ДИСЦИПЛІНИ**

для студентів спеціальностей

122 Комп'ютерні науки та інформаційні технології
(спеціалізація "Комп'ютерний еколого-економічний моніторинг"),

193 Геодезія та землеустрій
(спеціалізація "Геоінформаційні системи і технології")

Дніпро
НГУ
2016

Сергеєва К.Л.

Геостатистика. Методичні рекомендації до лабораторних занять з дисципліни для студентів спеціальностей 122 Комп'ютерні науки та інформаційні технології, 193 Геодезія та землеустрій / К.Л. Сергеєва ; М-во освіти і науки України, Нац. гірн. ун-т. – Дніпро: НГУ, 2016. – 63 с.

Автор:

К.Л. Сергеєва, канд. техн. наук.

Затверджено до видання редакційною радою ДВНЗ «НГУ» (протокол № 9 від 29.10.2016) за поданням кафедри геоінформаційних систем (протокол № 3 від 26.09.2016).

Методичні матеріали призначено для виконання лабораторних робіт з дисципліни «Геостатистика» студентами спеціальностей 122 Комп'ютерні науки та інформаційні технології та 193 Геодезія та землеустрій.

Розглянуто практичні рекомендації до розробки елементів програмних додатків статистичного аналізу просторових даних у середовищі Borland Delphi. Наведено етапи вирішення задачі інтерполяції за методом ординарного крігінгу. Подано рекомендації до виконання геостатистичної інтерполяції просторових змінних із застосуванням функціональних можливостей модулю Geostatistical Analyst пакету ArcGIS. Подано опис процесу інтерполяції та стохастичного моделювання просторових змінних у пакеті S-GEMS.

Наведено словник термінів з геостатистики.

Рекомендації орієнтовано на активізацію виконавчого етапу навчальної діяльності студентів.

Відповідальний за випуск: завідувач кафедри геоінформаційних систем, д-р техн. наук, проф. Б.С. Бусигін.

ЗМІСТ

| | |
|---|----|
| ПЕРЕДМОВА | 4 |
| Лабораторна робота 1. Статистичний аналіз просторових даних у середовищі Borland Delphi. Розрахунок найпростіших статистик | 6 |
| Лабораторна робота 2. Статистичний аналіз просторових даних у середовищі Borland Delphi. Кореляція. Регресія. Розрахунок статистик у ковзному вікні ... | 11 |
| Лабораторна робота 3. Вирішення задачі інтерполяції за методом ординарного крігінгу. | 14 |
| Лабораторна робота 4. Геостатистичний аналіз просторових даних в ArcGIS .. | 22 |
| Лабораторна робота 5. Вивчення розподілу даних..... | 38 |
| Лабораторна робота 6. Визначення ступеня забруднення ґрунтових вод м. Дніпро. | 48 |
| Лабораторна робота 7. Геостатистичний аналіз даних в S-GEMS..... | 50 |
| ПЕРЕЛІК ПОСИЛАНЬ | 57 |
| ДОДАТОК А. СЛОВНИК ТЕРМІНІВ З ГЕОСТАТИСТИКИ..... | 58 |
| ДОДАТОК Б. ВИМОГИ ДО ОФОРМЛЕННЯ ЗВІТУ З ЛАБОРАТОРНОЇ РОБОТИ .. | 61 |
| ДОДАТОК В. КРИТЕРІЇ ОЦІНЮВАННЯ ЛАБОРАТОРНИХ РОБІТ | 62 |

ПЕРЕДМОВА

Геостатистика як науковий напрямок виникла в 60-х роках ХХ сторіччя і представляє собою науку, що займається вивченням і оцінювання просторових (регіоналізованих) змінних за допомогою статистичних методів. Геостатистика утворює гілку просторової статистики і являє теорію оцінки просторових змінних.

Питання геостатистики охоплюють всі сфери діяльності людини, пов'язані із застосуванням просторових даних. До них відносяться: геологія, геофізика, навколишнє середовище (аналіз даних просторового розподілу різних видів живих організмів), екологія (аналіз даних по забрудненню ґрунтів), кліматологія (аналіз рівня опадів, температурних зрізів), гідрогеологія (моделювання гідрогеологічного осадового шару в рамках гідрогеологічної системи, зонування гідрогеологічного шару), електроспоживання (опис невизначеності прогнозу часового ряду по електроспоживанню), гірничорудна справа (аналіз і оцінка запасів при розробці родовищ корисних копалин – дорогоцінних металів, залізної руди, кольорових металів та ін.), епідеміологія та соціологія, рибне та лісове господарство тощо.

Лабораторний практикум спрямований на вивчення систем геостатистичного аналізу просторових даних – ArcGIS Geostatistical Analyst, S-GEMS, а також розробку елементів програмних засобів статистичного аналізу просторових даних в середовищі Borland Delphi. Методичні рекомендації містять перелік основних термінів і понять геостатистики.

Лабораторна робота № 1 присвячена створенню елементів програмного додатку, спрямованого на генерування вибірки значень просторової змінної, що містять детерміновану і стохастичну складову, побудову гістограми просторової змінної і розрахунок найпростіших характеристик її розподілу.

Лабораторна робота № 2 являється продовженням лабораторної роботи № 1 і передбачає генерування додаткової вибірки значень просторової змінної, виконання лінійною апроксимації, розрахунок коефіцієнту парної кореляції і статистик у «ковзному вікні».

У лабораторній роботі № 3 розглянуто приклад вирішення задачі інтерполяції за методом ординарного крігінгу.

У лабораторній роботі № 4 наведено рекомендації до виконання геостатистичної інтерполяції просторових змінних із застосуванням методу крігінг, реалізованого у модулі Geostatistical Analyst пакету ArcGIS.

Питання визначення та перетворення розподілу просторових даних розглянуто у лабораторній роботі № 5.

Лабораторна робота № 6 присвячена оцінці ступеня забруднення ґрунтових вод м. Дніпро із застосуванням геостатистичних методів.

Лабораторна робота № 7 спрямована на вивчення можливостей інтерполяції і стохастичного моделювання у пакеті S-GEMS.

У результаті виконання лабораторних робіт студент повинен:

- отримати знання закономірностей випадкових явищ і вміння застосовувати ймовірностно-статистичні методи для вирішення професійних завдань;

- оволодіти методами математичного та математико-статистичного моделювання у різних галузях.

Для успішного виконання лабораторних робіт студент повинен володіти теоретичними складовими дисципліни, а також знаннями, отриманими під час вивчення дисциплін «Інформатика та програмування» і «Теорія ймовірностей, ймовірнісні процеси та математична статистика».

Лабораторна робота 1

Статистичний аналіз просторових даних у середовищі Borland Delphi. Розрахунок найпростіших статистик

Об'єкт – просторові дані та випадкові функції.

Предмет – методи та програмні засоби статистичного аналізу просторових даних та випадкових функцій.

Мета – закріпити знання відносно статистичних методів аналізу просторових даних та випадкових функцій, побудови гістограми. Розробити у середовищі Borland Delphi програму побудови графіку випадкової функції та її гістограми, реалізувати можливість розрахунку найпростіших статистик.

Теоретичні положення

1. Побудова гістограми

Гістограма являє собою графічне відображення залежності частоти потрапляння елементів вибірки до відповідного інтервалу угруповання.

Компонент *Chart* середовища *Borland Delphi* дозволяє будувати діаграми та графіки. Є контейнером об'єктів *Series* типу *TChartSeries* – серій даних, що характеризуються різними стилями відображення. Кожна серія відповідає одній кривій на графіку. Властивості компоненту *Chart* наведені у табл. 1.1.

Таблиця 1.1

Властивості компоненту *Chart*

| | |
|--|--|
| <i>AllowPanning</i> | Визначає можливість користувача прокручувати видиму частину графіка під час виконання, натискаючи праву кнопку миші: <i>pmNone</i> – прокрутка заборонена; <i>pmHorizontal</i> – дозволена прокрутка тільки в горизонтальному напрямку; <i>pmVertical</i> – дозволена прокрутка тільки в вертикальному напрямку; <i>pmBoth</i> – дозволена прокрутка в обох напрямках. |
| <i>AllowZoom</i> | Дозволяє користувачеві змінювати масштаб зображення, вирізаючи фрагменти діаграми або графіка курсором миші. |
| <i>Title</i> | Визначає заголовок діаграми. |
| <i>Foot</i> | Визначає підпис під діаграмою. За замовчуванням відсутня. Текст підпису визначається підвластивістю <i>Text</i> . |
| <i>Frame</i> | Визначає рамку навколо діаграми. |
| <i>Legend</i> | Легенда діаграми – список позначень. |
| <i>MarginLeft</i> , <i>MarginRight</i> , <i>MarginTop</i> , <i>MarginBottom</i> | Значення лівого, правого, верхнього та нижнього полів. |

| | |
|--|---|
| <i>ButtomAxis</i> <i>LeftAxis</i> <i>RightAxis</i> | Визначають характеристики відповідно нижньої, лівої і правої осей. Доцільно зазначати ці властивості для графіків і деяких інших типів діаграм. |
| <i>LeftWall</i> <i>ButtomWall</i> <i>BackWall</i> | Визначають характеристики відповідно лівої, нижньої і задньої граней області тривимірного відображення графіка. |
| <i>SeriesList</i> | Список серій даних, що відображаються в компоненті. |
| <i>View3d</i> | Дозволяє або забороняє тривимірне відображення діаграми. |
| <i>View3dOptions</i> | Характеристики тривимірного відображення. |
| <i>Chart3DPersent</i> | Масштаб тривимірності (товщина діаграми, ширина ліній графіка). |

Редактор діаграм викликається:

- кнопкою з трьома крапками поруч із назвою властивості в інспекторі об'єктів;
- подвійним клацанням на компоненті *Chart* під час проектування форми;
- вибором команди *Edit Chart* в контекстному меню компонента *Chart* під час проектування форми.

Для завдання відображуваних значень треба використовувати методи серій *Series*:

Clear – очищає серію від нанесених раніше даних.

Add – дозволяє додати до діаграми нову точку.

Add(Const AValue:Double; Const ALabel:String; AColor:TColor),

де параметр *AValue* відповідає доданому значенню, параметр *ALabel* (необов'язковий) – назва, яка буде відображатися на діаграмі і в легенді, параметр *AColor* – колір.

AddXY – дозволяє додати нову точку в графік функції.

AddXY(Const AXValue, AYValue: Double; Const ALabel: String; AColor: TColor),

де параметри *AXValue* і *AYValue* відповідають аргументу і функції, параметри *ALabel* і *AColor* – ті ж, що і в методі *Add*.

Приклад побудови гистограми в середовищі *Borland Delphi*:

```

procedure TForm1.ShowHist;

var Hist:array of integer;
    Intervals,i, NOfIntervals:integer;
    T,dH:real;

begin
    //Зчитування кількості інтервалів гистограми
    try Intervals:=StrToInt(IntervalsEdit.Text);
    except
        ShowMessage('Числа введені невірно!')
    end;

```

```

HistSeries.Clear;
// Створення масиву висот стовпців гістограми
SetLength(Hist, Intervals+1);
// Ширина інтервалу
dH: = (OriginalSerie.MaxYValue-OriginalSerie.MinYValue)/Intervals;
// Розрахунок висот стовпців гістограми
For i:=0 to OriginalSerie.Count-1 do
begin
  NOfInterval:=round((OriginalSerie.YValue[i]-
  OriginalSerie.MinYValue/dH);
  Hist[NOfInterval]:=Hist[NOfInterval]+1;
end;

// Вивід гістограми на екран
For i:=0 to Intervals-1 do
HistSeries.AddXY(OriginalSerie.MinYValue+i*dH, Hist[i]);

```

2. Розрахунок статистик

1) Середнє значення вибірки випадкової величини обчислюється за формулою:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n},$$

де n – об'єм вибірки.

2) Дисперсією називається величина:

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}.$$

3) Стандартне відхилення розраховується за формулою:

$$\sigma = \sqrt{\sigma^2}.$$

4) Коефіцієнт асиметрії обчислюється за формулою:

$$\bar{M}_3 = \frac{\sum_{i=1}^n (x_i - \bar{x})^3}{n}, \quad A = \frac{\bar{M}_3}{\sigma^3},$$

де \bar{M}_3 – вибірковий центральний момент 3-го порядку.

Коефіцієнт асиметрії характеризує симетричність гістограми розподілу значень щодо модального значення. Для нормального розподілу $a = 1$.

5) Коефіцієнт ексцесу визначається наступним чином:

$$\bar{M}_4 = \frac{\sum_{i=1}^n (x_i - \bar{x})^4}{n}, \quad E = \frac{\bar{M}_4}{\sigma^4} - 3,$$

Коефіцієнт ексцесу характеризує відносну гостровершинність або пологість розподілу в порівнянні з нормальним розподілом. Якщо коефіцієнт ексцесу значимо відрізняється від 0, то функція щільності має або більш закруглений, або більш гострий пік, ніж пік щільності нормального розподілу.

6) Енергія обчислюється за формулою:

$$E_n = \frac{1}{n} \sum_{i=1}^n x_i^2.$$

7) *Ентропія* обчислюється за формулою:

$$E_t = - \frac{1}{n} \sum_{i=1}^n x_i \times \log_2 x_i.$$

Приклад розрахунку статистик у середовищі *Borland Delphi*:

```
// Розрахунок статистик
// Середнє значення
Xcp:=0;
For i:=0 to OriginalSerie.Count-1 do
  Xcp:=Xcp+OriginalSerie.YValue[i];
  Xcp:=Xcp/OriginalSerie.Count;
  MeanLabel.Caption:=FloatToStr(Xcp);
...
end;
```

Постановка завдання

У середовищі *Borland Delphi* створити програмний додаток, який дозволяє здійснювати статистичний аналіз даних. Для цього виконати наступні дії.

1. Сформуванати вибірку відповідно до варіанту формули (номер варіанта відповідає номеру за списком):

1. $Z(i) := A \cdot \sin(i \cdot B) + \sin(i \cdot C) + E$
2. $Z(i) := \sin(i^{1/2} \cdot A) + \cos(i \cdot B) + E$
3. $Z(i) := A \cdot \cos(i \cdot B) + B \cdot \sin(i) + E^2$
4. $Z(i) := A \cdot \cos(i \cdot E) + B \cdot \sin(i \cdot E)$
5. $Z(i) := \sin(i \cdot A) + \sin(i \cdot B) + \cos(i \cdot C) + E$
6. $Z(i) := \cos(i \cdot A) + \sin(i \cdot B) + 10 \cdot \cos(i \cdot C) + E$
7. $Z(i) := 10 \cdot \sin(i \cdot A) + 10 \cdot \sin(i \cdot B) + \cos(i \cdot C) + E$
8. $Z(i) := \sin(i^{1/3} \cdot A) + \cos(i \cdot E) + B$
9. $Z(i) := \sin(i \cdot A) + B \times \cos(i \cdot E)$
10. $Z(i) := 5 \cdot \sin(i \cdot E) + \cos(i \cdot B) + C \cdot \cos(i) + E$
11. $Z(i) := 2 \cdot \cos(i \cdot A) + \cos(i \cdot B) + \sin(E)$
12. $Z(i) := \cos(i^{1/2} \cdot A) + \cos(i \cdot E) + \sin(i + B)$
13. $Z(i) := A \cdot \cos(i \cdot E) + B \cdot \sin(i \cdot E)$
14. $Z(i) := A \cdot \sin(i \cdot B) + C \cdot \sin(i \cdot E) + \sin(i)$
15. $Z(i) := \cos(i \cdot A) + B \cdot \cos(i) + \sin(i \cdot C) + E$

де A, B, C – константи, вводяться користувачем;

E – випадкова складова, розраховується для кожної точки в інтервалі від $-Emax$ до $+Emax$ ($Emax$ – константа, вводиться користувачем).

2. Побудувати гістограму значень елементів вибірки. Кількість інтервалів значень задається користувачем.

3. Для створеного в пункті 1 набору даних провести розрахунок найпростіших статистик.

Варіанти розрахунку статистик:

| Варіант | Статистики | Варіант | Статистики | Варіант | Статистики |
|----------------|-------------------|----------------|-------------------|----------------|-------------------|
| 1 | 1,2,6 | 5 | 3,4,7 | 9 | 2,4,6 |
| 2 | 2,4,7 | 6 | 1,6,7 | 10 | 1,2,6 |
| 3 | 3,4,6 | 7 | 3,5,6 | 11 | 3,5,7 |
| 4 | 4,6,7 | 8 | 1,3,4 | 12 | 2,5,6 |

4. Відповідно до отриманих результатів виконання роботи зробити змістовні висновки. Оформити і захистити звіт.

Питання для підготовки до захисту лабораторної роботи

1. Що являє собою гістограма?
2. Наведіть основні властивості компоненту Chart.
3. Які методи компоненту Chart дозволяють додати до діаграми нову точку?
4. Наведіть визначення показників середнього значення, дисперсії, стандартного відхилення.
5. Що характеризують коефіцієнт асиметрії та коефіцієнт ексцесу?

Лабораторна робота 2

Статистичний аналіз просторових даних у середовищі Borland Delphi. Кореляція. Регресія. Розрахунок статистик у ковзному вікні

Об'єкт – просторові дані та випадкові функції.

Предмет – методи та програмні засоби статистичного аналізу просторових даних та випадкових функцій.

Мета – закріпити знання відносно методів розрахунку коефіцієнту лінійної кореляції та коефіцієнтів лінійної парної регресії. Отримати навички розрахунку статистик у ковзному вікні. Розробити у середовищі Borland Delphi програму обчислення коефіцієнта кореляції та побудови апроксимуючої кривої (лінії регресії) просторових даних та випадкових функцій.

Теоретичні положення

1. Розрахунок найпростіших статистик у ковзному вікні

Приклад реалізації обробника події *OnClick* для кнопки «Розрахунки у вікні» (рис. 2.1):

```
procedure TForm1.CulcWlnBucconClick(Sender: TObject);
var
  Xcp, Disp, Eng: real;
  WinSize, WinSizeDiv2: word;
  i, j: integer;
begin
  // Зчитування розміру вікна
  try
    WinSize:=ScrToInc(WinSizeEdic.Text);
  except
    ShowMessage(' Числа введені невірно!')
  end;
  // Зміна розміру вікна на непарним (якщо парний)
  if Odd(WinSize)=false then Inc(WinSize);
  // Розрахунок розміру половини вікна
  WinSizeDiv2:=Trunc(WinSize/2);
  DispWinSeries.Clear;
  MeanWinSeries.Clear;
  // Цикл по точках вихідної серії
  For i:=Trunc(WinSize/2) to OriginalSerie.Counc-1-Trunc(WinSize/2) do
  begin
    // Розрахунок середнього у вікні при поточному положенні
    Xcp:=0;
    For j:=i-WinSizeDiv2 to i+WinSizeDiv2 do
      Xcp:=Xcp+OriginalSerie.YValue[j];
    Xcp:=Xcp/WinSize;
    MeanWinSeries.AddXY(Xl+dX*i, Xcp);
    // Розрахунок дисперсії у вікні при поточному положенні
    Disp:=0;
    For j:=i-WinSizeDiv2 to i+WinSizeDiv2 do
      Disp:=Disp + sqr(OriginalSerie.YValue[j]-Xcp);
    Disp:=Disp/(WinSize-1);
    DispWinSeries.AddXY(Xl+dX*i, Disp)
  end;
end;
```

2. Розрахунок коефіцієнта лінійної кореляції

Коефіцієнт кореляції оцінює тісноту зв'язку між значеннями двох функцій (просторових змінних) і обчислюється за формулою:

$$r = \frac{\sum(A_i - \bar{A}) \times (B_i - \bar{B})}{\sqrt{\sum(A_i - \bar{A})^2 \times \sum(B_i - \bar{B})^2}}$$

де A_i – поточне значення першої просторової змінної;

B_i – поточне значення другої просторової змінної.

Значення коефіцієнта кореляції знаходяться в інтервалі від -1 до 1. Значення r , близькі по модулю до 1, означають тісний зв'язок; 0 – свідчать про повну відсутність зв'язку.

3. Апроксимація складних кривих

Апроксимація – заміна складних вимірних залежностей більш простими аналітичними функціями. В роботі розглядається найпростіший випадок – заміна значень псевдовипадкової величини (початкового графіка) прямою лінією.

Рівняння кривої має вигляд:

$$y = a * x + b,$$

де a і b – коефіцієнти, які потрібно обчислити за відомими значеннями вихідної залежності.

Під час виконання апроксимації ставиться умова мінімізації відхилень відомих значень від обчислених за апроксимуючою функцією. Ця задача вирішується за допомогою методу найменших квадратів.

Метод найменших квадратів

Нехай є набір значень y_i , кожне з яких знаходиться в точці x_i ($i = 1, 2, \dots, n$). Потрібно знайти залежність $y = f(x)$ за умови, що сума квадратів відхилень точки кривої від відповідної точки y_i є мінімальною:

$$\sum (f(x_i) - y_i)^2 \rightarrow \min.$$

Оскільки рівняння прямої має вигляд $y = a * x + b$, умову можна записати у вигляді

$$\sum (a * x_i + b - y_i)^2 \rightarrow \min.$$

Вирішення задачі методом найменших квадратів зводиться до розв'язання системи рівнянь:

$$a * s_{xx} + b * s_x - s_{yx} = 0; \quad a * s_x + b * n - s_y = 0,$$

де

$$s_{xx} = \sum (x_i * x_i); \quad s_{yx} = \sum (y_i * x_i); \quad s_x = \sum (x_i); \quad s_y = \sum (y_i).$$

Вирішуючи цю систему двох лінійних рівнянь щодо a і b , отримаємо остаточні формули для їх обчислення:

$$a = (s_{yx} * n - s_y * s_x) / (n * s_{xx} - s_x * s_x);$$
$$b = (s_{xx} * s_y - s_{yx} * s_x) / (n * s_{xx} - s_x * s_x).$$

Для побудови графіка цієї прямої необхідно розрахувати значення y в початковій і кінцевій точці області завдання вихідної функції. Для цього в

рівняння прямої $y = a * x + b$ потрібно підставити знайдені a і b , а також координати x першої та останньої точки початкового графіка.

Якість апроксимації оцінюється за ступенем відхилення відомих значень від графіка апроксимуючої прямої. Зазвичай обчислюється середньоквадратичне відхилення (СКО) σ :

$$\sigma^2 = \frac{1}{n-2} \sum_{i=1}^n (y_i - a * x_i - b)^2, \quad \sigma = \sqrt{\sigma^2}.$$

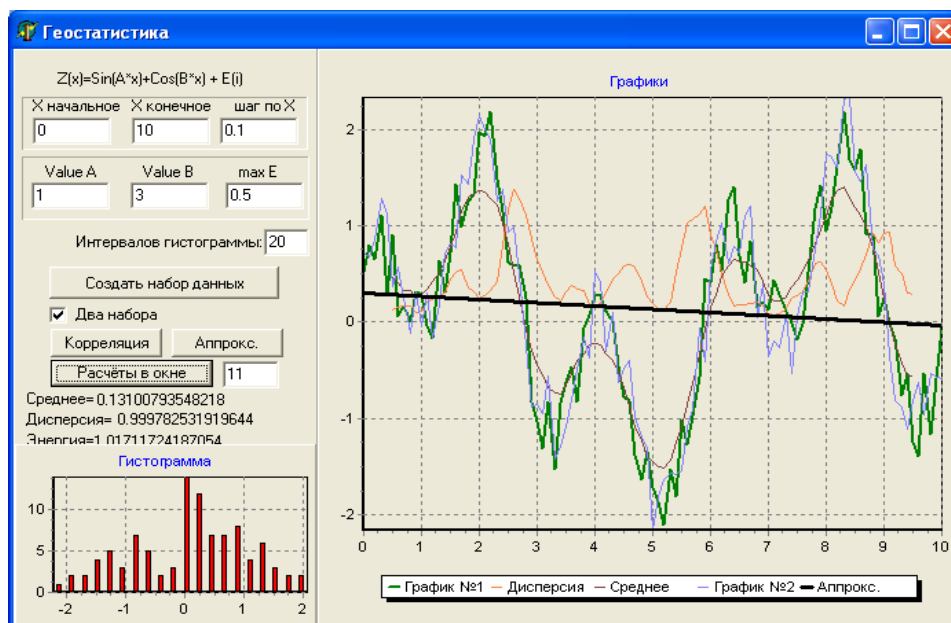


Рис. 2.1 Приклад оформлення інтерфейсу користувача

Постановка завдання

На підставі результатів, отриманих під час виконання лабораторної роботи 1, у середовищі *Borland Delphi* створити додаток, що дозволяє здійснювати статистичний аналіз даних. Для цього виконати наступні дії.

1. Розрахувати статистики в ковзному вікні і побудувати відповідні графіки.
2. Побудувати графіки двох функцій, які відрізняються значенням випадкової компоненти. Розрахувати загальний коефіцієнт кореляції для всіх значень, а також побудувати новий графік, що відображає значення коефіцієнта кореляції всередині ковзного вікна.
3. Побудувати графік апроксимуючої кривої (прямої лінії).
4. Відповідно до отриманих результатів виконання роботи зробити змістовні висновки. Оформити і захистити звіт.

Питання для підготовки до захисту лабораторної роботи

1. Що являє собою ковзне вікно для розрахунку статистик?
2. Надайте визначення коефіцієнта лінійної кореляції.
3. Які значення може приймати коефіцієнт кореляції?
4. Надайте визначення апроксимації.
5. Який показник характеризує якість апроксимації?

Лабораторна робота 3

Вирішення задачі інтерполяції за методом ординарного крігінгу

Об'єкт – просторові дані.

Предмет – метод ординарного крігінгу.

Мета – закріпити знання відносно понять варіограми та крігінгу. Отримати навички розв'язання системи рівнянь ординарного крігінгу

Теоретичні положення

Нехай на місцевості виміряли значення висоти в п'яти точках досліджуваного ландшафту – опорних точках (рис. 3.1).

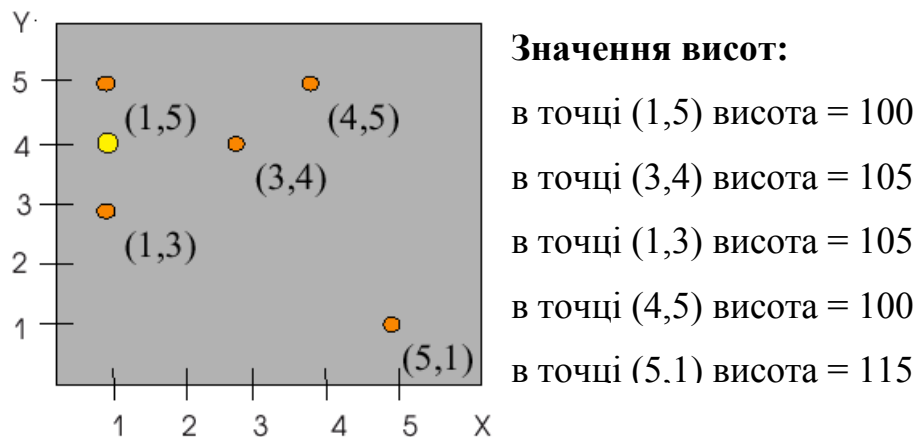


Рис. 3.1 Тестові дані по опорним точкам досліджуваного ландшафту

Для інтерполяції значення в точці з координатами $(1,4)$ (шукана точка), скористаємося методом ординарного крігінга. Модель ординарного крігінга визначається формулою:

$$Z(x) = m + e(x),$$

де x – положення точки і $Z(x)$ – значення виміряної величини для заданої точки; наприклад, для точки $Z(1,5) = 100$. Модель базується на постійному середньому m ($m = E\{Z(x)\} = \text{const}$, $E\{\cdot\}$ – математичне очікування) і випадкових помилках $e(x)$ з просторовою залежністю.

Вводиться припущення про характер просторової кореляції: варіація для кожної просторової орієнтації існує і залежить тільки від просторової орієнтації, а не від положення в просторі (функція варіограми):

$$\gamma(x, x + h) = \gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} (Z(x_i) - Z(x_i + h))^2,$$

де $N(h)$ – число пар, розділених вектором h .

Інтерполятор може бути отриманий як зважена сума даних,

$$\hat{Z}(x_0) = \sum_{i=1}^N \lambda_i Z(x_i),$$

де $Z(x_i)$ – вимірне значення в i -й точці, наприклад, $Z(1,5) = 100$;

λ_i – невідома вага для вимірюного значення в i -й точці;

x_0 – координати шуканої точки, наприклад, $(1,4)$;

$N = 5$ число опорних точок.

При виконанні інтерполяції для декількох точок, деякі з шуканих значень опиняться вище або нижче фактичних значень величин. В середньому, різниця між проінтерпольованими значеннями і фактичними значеннями повинна дорівнювати нулю. Така умова називається "умовою незміщеності інтерполятора". Для гарантії того, що інтерполятор є незміщений, сума ваг повинна дорівнювати одиниці.

$$\sum_{i=1}^N \lambda_i = 1.$$

Різниця між істинним значенням $Z(x_0)$ і інтерполятором $\hat{Z}(x_0)$ мінімізується:

$$(Z(x_0) - \hat{Z}(x_0))^2 \rightarrow \min,$$

а також мінімізується величина σ_e^2 – дисперсія похибок оцінювання.

З урахуванням цієї умови необхідно мінімізувати функцію

$$F = \sigma_e^2 + 2\mu \left(\sum_i a_i - 1 \right),$$

де μ – нове невідоме (множник Лагранжа).

Умовою мінімуму функції F є рівність нулю всіх часткових похідних по a_i і μ . рішення для мінімізації, обмежене умовою незміщеності, дає рівняння крігінга в матричному вигляді:

$$BA = C,$$

де

$$B = \begin{pmatrix} \gamma_1^1 & \gamma_1^2 & \dots & \gamma_1^N & 1 \\ \gamma_2^1 & \gamma_2^2 & \dots & \gamma_2^N & 1 \\ \dots & \dots & \dots & \dots & \dots \\ \gamma_N^1 & \gamma_N^2 & \dots & \gamma_N^N & 1 \\ 1 & 1 & \dots & 1 & 0 \end{pmatrix}, \quad A = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \dots \\ \lambda_N \\ \mu \end{pmatrix}, \quad C = \begin{pmatrix} \gamma_1^0 \\ \gamma_2^0 \\ \dots \\ \gamma_N^0 \\ 1 \end{pmatrix}.$$

Матриця **B** містить змодельовані значення варіограми для всіх пар опорних точок, де γ_i^j означає змодельовані значення варіограми, що базуються на відстані між двома опорними точками, визначеними як точки i та j .

Вектор **C** містить змодельовані значення варіограми для кожної пари опорної і шуканої точки, де γ_i^0 означає змодельовані значення варіограми, що базуються на відстані між i -ю точкою і шуканої точкою. Також оцінюється невідоме значення μ в векторі **A**.

1. Розрахунок емпіричної варіограми

Для розрахунку значень матриці **B** необхідно вивчити структуру даних шляхом створення емпіричної варіограми. На варіограмі показані значення

половини квадрата різниці для пар точок (відкладається по осі y) залежно від відстані між ними (відкладається по осі x).

Перший крок у створенні емпіричної варіограми – обчислення відстані і квадрата різниці між значеннями для кожної пари точок. Відстань між двома точками розраховується з використанням евклідовій відстані (табл. 3.1).

Таблиця 3.1

Розрахунок значень матриці **B** для тестових даних

| Точки | Розрахунок відстані | Відстань | 2•Варіограма | Варіограма |
|--------------|----------------------------|----------|--------------|------------|
| (1,5), (3,4) | $\sqrt{(1-3)^2 + (5-4)^2}$ | 2,236 | 25 | 12,5 |
| (1,5), (1,3) | $\sqrt{0^2 + 2^2}$ | 2 | 25 | 12,5 |
| (1,5), (4,5) | $\sqrt{3^2 + 0^2}$ | 3 | 0 | 0 |
| (1,5), (5,1) | $\sqrt{4^2 + 4^2}$ | 5,657 | 225 | 112,5 |
| (3,4), (1,3) | $\sqrt{2^2 + 1^2}$ | 2,236 | 0 | 0 |
| (3,4), (4,5) | $\sqrt{1^2 + 1^2}$ | 1,414 | 25 | 12,5 |
| (3,4), (5,1) | $\sqrt{2^2 + 3^2}$ | 3,606 | 100 | 50 |
| (1,3), (4,5) | $\sqrt{3^2 + 2^2}$ | 3,606 | 25 | 12,5 |
| (1,3), (5,1) | $\sqrt{4^2 + 2^2}$ | 4,472 | 100 | 50 |
| (4,5), (5,1) | $\sqrt{1^2 + 4^2}$ | 4,123 | 225 | 112,5 |

Для великих наборів даних (з великою кількістю вимірних опорних точок) число пар точок буде швидко зростати і стане некерованим. Тому проводиться угруповання пар точок; такий процес відомий як *бінінг*.

Бін – це заданий діапазон відстаней. Це означає, що всі точки, які знаходять на відстані від 0 до 1, згруповані в перший бін, а ті точки, які знаходяться на відстані від 1 до 2 метрів одна від одної, згруповані в другий бін і т.д. Береться середня емпірична дисперсія для всіх пар точок. У нашому прикладі, дані розподілені в п'ять бінів (табл. 3.2).

Таблиця 3.2

Бінінг емпіричної варіограми

| Розмір лага | Відстань | Середня відстань | Варіограма | Середня варіограма |
|-------------|-----------------|------------------|------------|--------------------|
| [1-2] | 1,414; 2 | 1,707 | 12,5; 12,5 | 12,5 |
| (2-3] | 2,236; 2,236; 3 | 2,491 | 12,5; 0; 0 | 4,167 |
| (3-4] | 3,606; 3,606 | 3,606 | 50; 12,5 | 31,25 |
| (4-5] | 4,472; 4,123 | 4,298 | 50; 112,5 | 81,25 |
| (5-...) | 5,657 | 5,657 | 112,5 | 112,5 |

Відобразимо на графіку емпіричної варіограми середню дисперсію залежно від середньої відстані для біна. Але значення емпіричної варіограми не можуть бути використані безпосередньо в матриці **B**; замість цього слід підібрати модель для емпіричної варіограми (рис. 3.2).

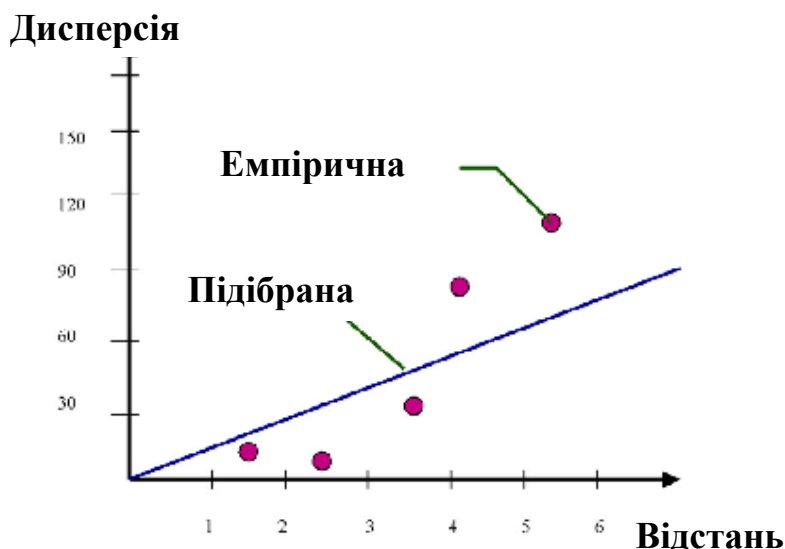


Рис. 3.2 Графічне відображення емпіричної варіограми і її моделі

Для простоти, підібрана модель – лінія регресії, побудована з використанням методу найменших квадратів, що має додатній кутовий коефіцієнт і проходить через початок координат ($y = kx$).

Формула для визначення дисперсії для будь-якої заданої відстані виглядає:

$$\text{Дисперсія} = \text{Кутовий коефіцієнт} * \text{Відстань}.$$

Відстань – це відстань між парами точок, позначається як h . У даному прикладі дисперсія для будь-якої відстані може бути визначена за формулою:

$$\text{Дисперсія} = 13.5 * h.$$

Тепер створимо матрицю значень **B** (табл. 3.3). Наприклад, значення γ_{12} для точок (1,5) і (3,4) у рівнянні дорівнює:

$$\text{Дисперсія} = 13,5 \cdot 2,236 = 30,19.$$

Таблиця 3.3

Розрахунок значень матриці **B** за даними емпіричної варіограми

| | | | | | | |
|--------------|--------------|--------------|--------------|--------------|--------------|---|
| | (1,5) | (3,4) | (1,3) | (4,5) | (5,1) | |
| (1,5) | 0 | 30,19 | 27,0 | 40,5 | 76,37 | 1 |
| (3,4) | 30,19 | 0 | 30,19 | 19,09 | 48,67 | 1 |
| (1,3) | 27,0 | 30,19 | 0 | 48,67 | 60,37 | 1 |
| (4,5) | 40,5 | 19,09 | 48,67 | 0 | 55,66 | 1 |
| (5,1) | 76,37 | 48,67 | 60,37 | 55,66 | 0 | 1 |
| | 1 | 1 | 1 | 1 | 1 | 0 |

Одиниці і нулі в нижньому рядку та крайньому правому стовпці визначені відповідно до умови незміщеності.

Формула обчислення значень матриці для ординарного крігінга наступна:

$$BA = C.$$

Тепер, коли побудовано матрицю **B**, необхідно знайти значення вектора **A**, що містить ваги, які будуть присвоєні вимірним значенням довкола шуканої точки. Тому виконаємо просту операцію з матричної алгебри і отримаємо таку формулу:

$$A = CB^{-1},$$

де B^{-1} – обернена матриця **B**. Обернену матрицю **B** одержуємо, виконавши операцію лінійної алгебри (табл. 3.4).

Таблиця 3.4

Розрахунок значень оберненої матриці **B**

| | | | | | |
|----------|----------|----------|----------|----------|----------|
| -0,02575 | 0,00704 | 0,0151 | 0,00664 | -0,00303 | 0,3424 |
| 0,00704 | -0,04584 | 0,01085 | 0,02275 | 0,0052 | -0,22768 |
| 0,0151 | 0,01085 | -0,02646 | -0,00471 | 0,00522 | 0,17869 |
| 0,00664 | 0,02275 | -0,00471 | -0,02902 | 0,00433 | 0,28471 |
| -0,00303 | 0,0052 | 0,00522 | 0,00433 | -0,01173 | 0,42189 |
| 0,3424 | -0,22768 | 0,17869 | 0,28471 | 0,42189 | -41,701 |

Далі, для шуканої точки будується вектор **C**. Наприклад, розглянемо точку (1,4). Обчислюємо відстань від точки (1,4) до кожної з точок з вимірними значеннями (1,5), (3,4), (1,3), (4,5) і (5,1). Виходячи з цих відстаней, визначаємо підібрану дисперсію за формулою

$$\text{Дисперсія} = 13.5 \cdot h,$$

виведеною раніше. Вектор **C** для точки (1,4) наведено в табл. 3.5.

Таблиця 3.5

Розрахунок значень вектору **C**

| Точка | Відстань | Вектор C для (1,4) |
|-------|----------|---------------------------|
| (1,5) | 1 | 13,5 |
| (3,4) | 2 | 27,0 |
| (1,3) | 1 | 13,5 |
| (4,5) | 3,162 | 42,69 |
| (5,1) | 5 | 67,5 |
| | | 1 |

Тепер, після того, як були розраховані матриця **B** і вектор **C**, обчислимо вектор ваг крігінга:

$$A = CB^{-1}.$$

Для цього скористаємося лінійною алгеброю. Ваги наведені в табл. 3.6.

Таблиця 3.6

Розрахунок значень ваг крігінга і результати інтерполяції

| Ваги | Значення | Результат |
|----------|----------|-----------------|
| 0,46757 | 100 | 46,757 |
| 0,09834 | 105 | 10,3257 |
| 0,46982 | 105 | 49,3311 |
| -0,02113 | 100 | -2,113 |
| -0,0146 | 115 | -1,679 |
| -0,18281 | | 102,6218 |

2. Виконання інтерполяції

Тепер, коли обчислено значення ваг, вага кожного виміряного значення множиться на це значення. Результати додаються і отримується шукане значення для точки (1,4) (рис. 3.3).

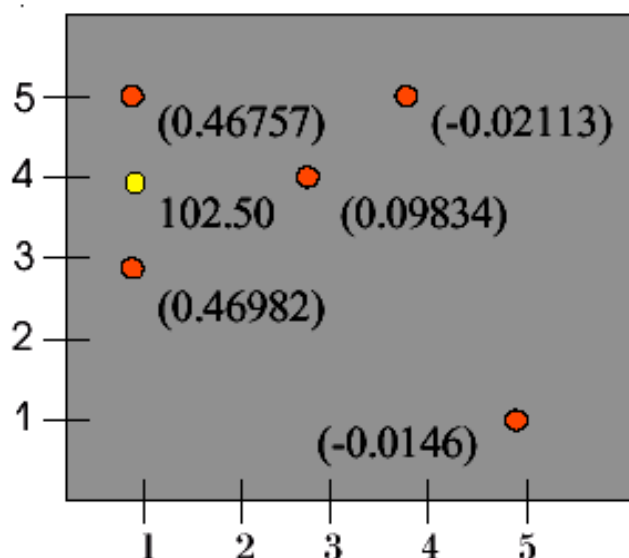


Рис. 3.3. Ваги крігінга і результати інтерполяції

Як і очікувалося, ваги зменшуються з відстанню, але таке зменшення визначено більш точно, оскільки при присвоєнні ваг враховувалася не тільки відстань між точками, але й просторовий розподіл даних. Виконана інтерполяція представляється досить достовірною.

3. Дисперсія крігінга

Одна із сильних сторін використання геостатистичного підходу – можливість обчислення статистичної міри помилки інтерполяції. Для цього необхідно помножити кожне значення вектора **A** на кожне значення вектора **C** і скласти їх разом, щоб отримати значення, відоме як *дисперсія* значень, отриманих з використанням крігінга. Корінь квадратний із дисперсії крігінга носить назву *стандартної помилки крігінга* (табл. 3.7).

Таблиця 3.7

Розрахунок дисперсії та стандартної помилки крігінга

| Вектор С | Ваги | Вектор С × Ваги |
|------------------------------------|----------|-----------------|
| 13,5 | 0,46757 | 6,312195 |
| 27,0 | 0,09834 | 2,65518 |
| 13,5 | 0,46982 | 6,34257 |
| 42,69 | -0,02113 | -0,90204 |
| 67,5 | -0,0146 | -0,9855 |
| 1 | -0,18281 | -0,18281 |
| Дисперсія крігінга | | 13,2396 |
| Стандартна помилка крігінга | | 3,6386 |

В даному випадку стандартна помилка крігінга дорівнює 3,6386.

Якщо припустити, що помилки розподілені за нормальним законом, 95 відсотків інтерпольованих значень будуть дорівнювати:

$$(\hat{Z}(S_0) - 1.96\sqrt{\sigma^2}, \hat{Z}(S_0) + 1.96\sqrt{\sigma^2})$$

Інтервал проінтерполірованного значення може бути інтерпретований таким чином. Якщо інтерполяція виконується знову і знову за однією і тією ж моделлю, для тривалого періоду в 95 відсотках випадків проінтерпольоване значення дорівнюватиме значенню в шуканій точці в межах заданого інтервалу. У нашому прикладі, інтервал інтерпольованого значення коливається від 95,49 до 109,75 ($102,62 + 1.96 * 3,64$).

Постановка завдання

На місцевості виміряли значення висоти в п'яти опорних точках досліджуваного ландшафту. Використовуючи ординарний крігінг, знайти проінтерполірованне значення в заданій точці. Оцінити результат.

| Номер варіанта | Номери опорних точок | Координати заданої точки | | Номер варіанта | Номери опорних точок | Координати заданої точки | |
|----------------|----------------------|--------------------------|---|----------------|----------------------|--------------------------|---|
| | | х | у | | | х | у |
| 1 | 13,7,22,5,31 | 1 | 0 | 7 | 16,8,23,25,35 | 1 | 2 |
| 2 | 14,1,10,27,6 | 7 | 0 | 8 | 31,2,29,24,13 | 6 | 2 |
| 3 | 18,21,5,7,34 | 9 | 0 | 9 | 15,19,12,4,32 | 10 | 2 |
| 4 | 1,15,10,32,30 | 3 | 1 | 10 | 32,1,28,23,18 | 1 | 3 |
| 5 | 5,24,27,7,14 | 5 | 1 | 11 | 34,26,6,8,17 | 7 | 3 |
| 6 | 17,10,26,36,19 | 9 | 1 | 12 | 35,27,12,19,14 | 0 | 4 |

Опорні точки:

| № | X | Y | Z | № | X | Y | Z | № | X | Y | Z |
|----|----|---|--------|----|----|---|--------|----|----|---|---------|
| 1 | 1 | 4 | -4,833 | 13 | 2 | 0 | 2 | 25 | 1 | 6 | -11,5 |
| 2 | 3 | 4 | -0,833 | 14 | 3 | 0 | 4,5 | 26 | 3 | 6 | -7,5 |
| 3 | 5 | 4 | 7,167 | 15 | 4 | 0 | 8 | 27 | 5 | 6 | 0,5 |
| 4 | 9 | 4 | 35,167 | 16 | 5 | 0 | 12,5 | 28 | 6 | 6 | 6 |
| 5 | 10 | 3 | 47 | 17 | 6 | 0 | 18 | 29 | 8 | 6 | 20 |
| 6 | 7 | 5 | 16,167 | 18 | 10 | 0 | 50 | 30 | 10 | 6 | 38 |
| 7 | 0 | 2 | -1,333 | 19 | 0 | 1 | -0,333 | 31 | 0 | 8 | -21,333 |
| 8 | 2 | 2 | 0,667 | 20 | 1 | 1 | 0,167 | 32 | 1 | 8 | -20,833 |
| 9 | 3 | 2 | 3,167 | 21 | 2 | 1 | 1,667 | 33 | 4 | 8 | -13,333 |
| 10 | 4 | 2 | 6,667 | 22 | 6 | 1 | 17,667 | 34 | 5 | 8 | -8,833 |
| 11 | 5 | 2 | 11,167 | 23 | 7 | 1 | 24,167 | 35 | 8 | 8 | 10,667 |
| 12 | 9 | 2 | 39,167 | 24 | 8 | 1 | 31,667 | 36 | 9 | 8 | 19,167 |

Для розрахунку проінтерпольованого значення створити програмний додаток у довільному програмному середовищі.

Відповідно до отриманих результатів виконання роботи зробити змістовні висновки. Оформити і захистити звіт.

Питання для підготовки до захисту лабораторної роботи

1. Який зміст поняття варіограма?
2. Що являє собою бінінг емпіричної варіограми?
3. Наведіть модель ординарного крігінга.
4. Як обчислюється стандартна помилка крігінга?
5. Наведіть основні етапи вирішення задачі інтерполяції за методом ординарного крігінга.

Лабораторна робота 4

Геостатистичний аналіз просторових даних в ArcGIS

Об'єкт – просторові дані.

Предмет – методи геостатистичного аналізу просторових даних.

Мета – ознайомитися з модулем Geostatistical Analyst пакету ArcGIS. Виконати дослідження набору вихідних даних та здійснити побудову поверхні засобами інструменту Geostatistical Wizard.

Теоретичні положення

В роботі розглянуто наступні набори даних:

| Набір даних | Опис |
|---------------------|---|
| <i>ca_outline</i> | Картографічна основа штату Каліфорнія |
| <i>ca_ozone_pts</i> | Контрольні точки виміру концентрацій озону, <i>ppm</i> (в проміле, $ppm = 0,0001\% = 10^{-6}$. Наприклад, якщо вказано, що масова частка речовини в суміші становить <i>15 ppm</i> , це означає, що на кожен кілограм суміші доводиться <i>15 мг</i> цієї речовини). |

Набір даних по озону (*ca_ozone_pts*) містить дані за 1996 р., що відображають максимальне значення концентрації озону в проміле (*ppm*), вибране з середніх значень спостережень за восьмигодинний період (вимірювання проводились в добовому режимі і групувалися в блоки по вісім годин).

Розташування даних: *C:\arcgis\ArcTutor\Geostatistics*.

1. Побудова поверхні з використанням параметрів, запропонованих за замовчуванням

1.1 Підключення модулю Geostatistical Analyst

В *ArcMap* виберіть меню *Інструменти*, потім рядок *Додаткові модулі* і у вікні відзначте галочкою опцію *Geostatistical Analyst*. Натисніть *Закрити*.

1.2 Підключення панелі інструментів модуля Geostatistical Analyst

Натисніть меню *Вид*, перейдіть на рядок *Панелі інструментів* і натисніть *Geostatistical Analyst* (рис. 4.1).

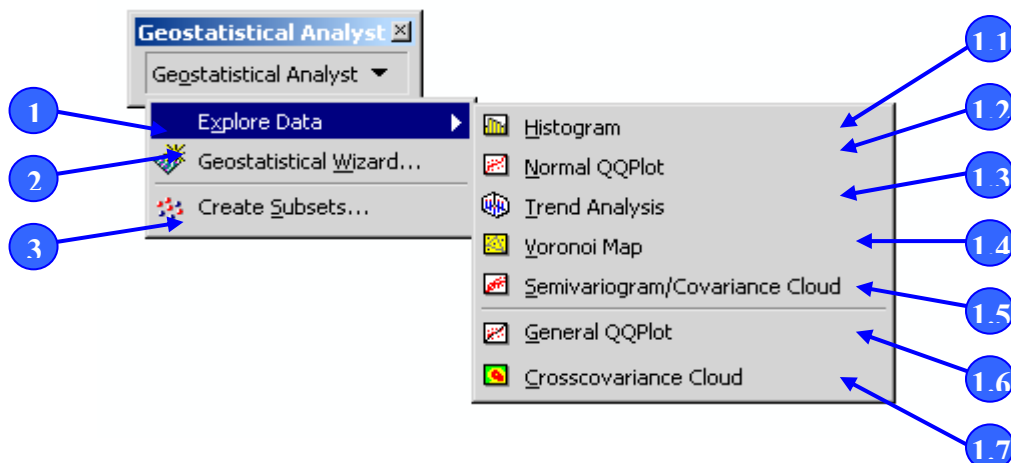


Рис. 4.1 Панель інструментів модуля *Geostatistical Analyst*

1. Explore Data (Дослідження даних) – вивчення статистики наборів даних. Ці інструменти можуть бути використані для дослідження даних незалежно від того, буде побудовано поверхню, чи ні. Інструменти дослідницького аналізу просторових даних, присутні у модулі *Geostatistical Analyst*, застосовуються для оцінки статистичних властивостей даних, таких як мінливість просторових даних, їх залежність і глобальні тренди.

1.1 Інструмент *Histogram* (Гістограма) створює гістограми частот для атрибутів в наборі даних, дозволяючи вивчати одномірний (для однієї змінної) розподіл даних за кожним атрибутом.

1.2 Інструмент *Normal QQPlot* (Нормальний графік Квантиль квантиль (КК)) дозволяє порівнювати розподіл даних із стандартним нормальним. Чим більш точно по точках можна побудувати пряму лінію, тим ближче розподіл до нормального.

1.3 Інструмент *Trend Analysis* (Аналіз тренду) дозволяє визначити відсутність / наявність трендів у вихідному наборі даних.

1.4 Інструмент *Voronoi Map* (Карта Вороного) кінцевої множини точок на площині являє таке розбиття площини, при якому кожна область цього розбиття утворює множину точок, більш близьких до одного з елементів множини, ніж до будь-якого іншого елементу множини.

1.5 Інструмент *Semivariogram / Covariance Cloud* (Хмара варіограми / коваріації) – це емпірична варіограма (на якій показані значення половини квадрата різниці між вимірними величинами) і коваріація для всіх пар опорних точок з набору даних; на хмарі варіограми ці значення показані як функція відстані між двома точками.

1.6 Інструмент *General QQPlot* (Загальний графік КК) використовується для оцінки схожості розподілів двох наборів даних. Загальний графік КК створюється шляхом нанесення значень даних, для яких сукупний розподіл має рівні значення.

1.7 Інструмент *Crossvariance Cloud* (Хмара взаємної коваріації) показує емпіричну взаємну коваріацію для всіх пар точок між двома наборами даних.

2. Geostatistical Wizard (Майстер операцій геостатистики) включає:

2.1 Моделювання варіограми

Геостатистичного аналізу даних відбувається в два етапи:

- моделювання варіограми або коваріації для аналізу властивостей поверхні;
- крігінг.

У модулі *Geostatistical Analyst* можливе використання цілого ряду методів, що базуються на крігінгу, включаючи методи ординарного, простого, універсального, індикаторного, імовірнісного та диз'юнктивного крігінга.

2.2 Інтерполяція поверхні та моделювання помилки

З використанням модуля *Geostatistical Analyst* можуть бути створені різні картографічні шари, включаючи карти проінтерпольованих значень, карти квантилів, карти ймовірностей і карти стандартної помилки інтерполяції.

2.3 Інтерполяція поверхні з використанням кокрігінга

Кокрігінг може бути використаний для поліпшення якості інтерполяції поверхні для однієї змінної шляхом врахування значень інших змінних за умови наявності просторової кореляції між ними.

3. Create Subsets (Створити піднабори) – перевірка моделі та діагностика. Вихідні дані можуть бути розбиті на два піднабори. Перший піднабір даних може бути використаний для створення моделі інтерполяції. Обчислені значення можуть бути потім зіставлені з відомими значеннями в pozostaх точках з використанням інструменту перевірки достовірності моделі.

2. Дослідження даних

2.1 Вивчення розподілу даних. Гістограма. Нормальний графік КК (Квантиль-квантиль)

Методи інтерполяції, що використовуються для побудови поверхні, дають кращі результати при нормальному розподілі даних (крива у формі дзвону). Якщо дані розподілені асиметрично (нерівномірно), можна скористатися можливістю перетворення даних для приведення їх до нормального розподілу. Інструмент *Гістограма* створює гістограми частот для атрибутів в наборі даних і вивчати одномірний розподіл даних по кожному атрибуту. Далі наведено етапи вивчення розподілу озону для шару *ca_ozone_pts* (рис. 4.2).

1. Виділіть шар *ca_ozone_pts*, перемістіть його в верхню частину таблиці змісту, потім за ним помістіть шар *ca_outline*.

2. На панелі інструментів *Geostatistical Analyst* перейдіть до опції *Дослідити дані* і потім виберіть рядок *Гістограма*.

3. В вікні *Шар* виберіть *ca_ozone_pts*.

4. У вікні *Атрибут* виберіть *OZONE*.

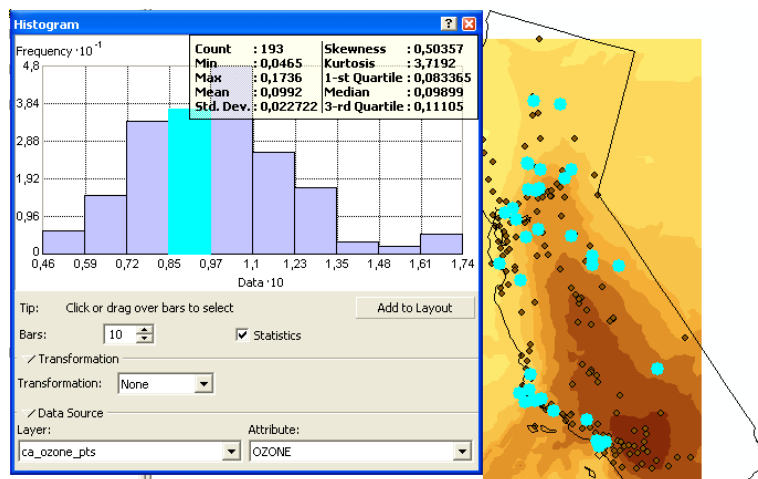


Рис. 4.2. Інструмент *Гістограма* модуля *Geostatistical Analyst*

Розподіл концентрації озону показано на гістограмі, на якій всі значення об'єднані в 10 класів. Висота стовпців пропорційна кількості даних з певними значеннями в кожному класі.

В цілому, важливими характеристиками розподілу є центральне значення, його розмах і асиметрія. Для швидкої перевірки характеру розподілу: якщо середнє і медіана приймають приблизно однакове значення, це є одним з підтверджень того, що дані розподілені за нормальним законом.

Гістограма на рис. 4.2 є унімодальною (одновершинною) і досить симетричною. Схоже, що дані мають розподіл, близький до нормального. Правий хвіст розподілу вказує на присутність відносно невеликої кількості елементів вибірки з великими значеннями концентрації озону.

5. Виберіть стовпчик гістограми зі значеннями концентрації озону від 0,162 до 0,175 ppm. Елементи вибірки, що мають такі значення, будуть виділені на карті. Зверніть увагу, що ці точки розташовані в районі Лос-Анджелеса.

6. Закрийте діалогове вікно.

Квантиль – число, таке, що задана випадкова величина не перевищує його лише з фіксованою ймовірністю (табл. 4.1).

- 0,25-квантиль називається першою (або нижньою) квантиллю;
- 0,5-квантиль називається медіаною або другою квантиллю;
- 0,75-квантиль називається третьою (або верхньою) квантиллю.

Таблиця 4.1

Квантилі стандартного нормального розподілу

| | | | | | | | | | |
|-----------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| <i>Ймовірність, %</i> | 99,99 | 99,90 | 99,00 | 97,72 | 97,50 | 95,00 | 90,00 | 84,13 | 50,00 |
| <i>Квантиль</i> | 3,715 | 3,090 | 2,326 | 2,000 | 1,960 | 1,645 | 1,282 | 1,000 | 0,000 |

Графік КК дозволяє порівняти розподіл даних із стандартним нормальним. Чим більш точно по точках можна побудувати пряму лінію, тим ближче розподіл до нормального (рис. 4.3).

1. На панелі інструментів *Geostatistical Analyst* виберіть опцію *Дослідити дані*, а потім рядок *Нормальний графік КК*.

2. У вікні *Шар* виберіть *sa_ozone_pts*.

3. У вікні *Атрибут* виберіть *OZONE*

На *графіку КК* квантилі двох розподілів розташовані по двох осях координат (наведені одна відносно одної). Для двох ідентичних розподілів *графік КК* представлятиме собою пряму лінію. Таким чином, нормальність розподілу даних по концентрації озону можна перевірити, використовуючи на *графіку* квантилі цих даних і квантилі стандартного нормального розподілу. *Графік* на рис. 4.3 дуже близький до прямої лінії. Якщо дані не відображають нормального розподілу ні на *гістограмі*, ні на *графіку КК*, перед тим як використовувати метод інтерполяції крігінг може виникнути необхідність привести їх до нормального розподілу.

2.2 Визначення глобальних трендів в даних

Якщо в даних існує тренд, він представляє собою не випадкову (детерміністську) складову поверхні, яка може бути описана будь-якою математичною формулою. Наприклад, пологий схил може бути представлений площиною. Долина може бути описана більш складною формулою (поліномом

другого порядку), яка має U-подібну форму. Однак у багатьох випадках ні схил не є правильною площиною, ні долина не має правильної U-подібної форми. Якщо поверхня тренда неадекватно відображає поверхню, можна відняти його і продовжити аналіз, моделюючи залишки або значення в опорних точках після вирахування тренда.

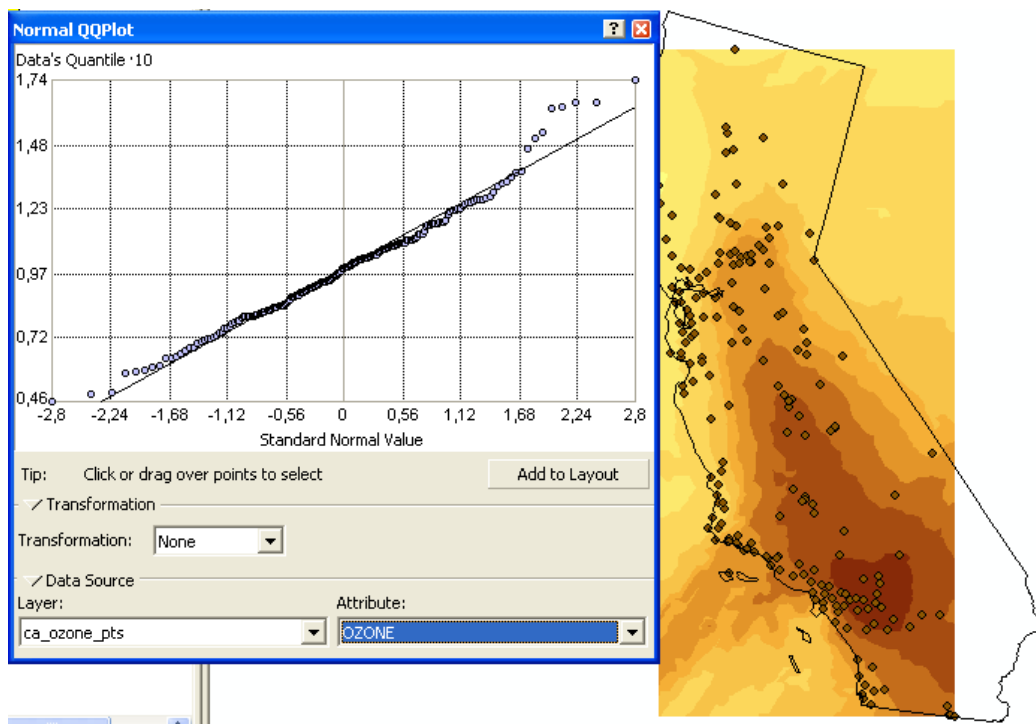


Рис. 4.3. Інструмент *Нормальний графік КК* модуля *Geostatistical Analyst*

Інструмент *Аналіз тренда* дозволяє визначити відсутність / наявність трендів у вихідному наборі даних (рис. 4.4).

1. На панелі інструментів *Geostatistical Analyst* виберіть *Дослідити дані*, а потім опцію *Аналіз тренда*.
2. Зі списку *Шар* виберіть *ca_ozone_pts*.
3. У вікні *Атрибут* виберіть *OZONE*.

Кожен вертикальний стовпець (відрізок) на діаграмі аналізу тренда представляє місце розташування, а його висота пропорційна значенню кожної точки з набору даних. Точки проєціюються на перпендикулярні площині, що відповідають напрямкам схід-захід і північ-південь. Через зпроєцьйовані точки проведена лінія (поліном), що щонайкраще описує їх розташування і моделює тренди в певних напрямках. Якщо лінія близька до прямої, це означає, що в даних немає тренду. Однак світло-зелена лінія на рис. 4.4 починається з низьких значень і збільшується в напрямку на схід, поки не почне знижуватися. Це вказує на те, що у даних, можливо, наявний сильний тренд в напрямку схід-захід і слабкіший в напрямку північ-південь.

Виявлений тренд, можливо, обумовлений тим фактом, що забруднення є низьким на узбережжі, а при русі всередину материка з'являються великі населені пункти, кількість яких знову зменшується в горах.

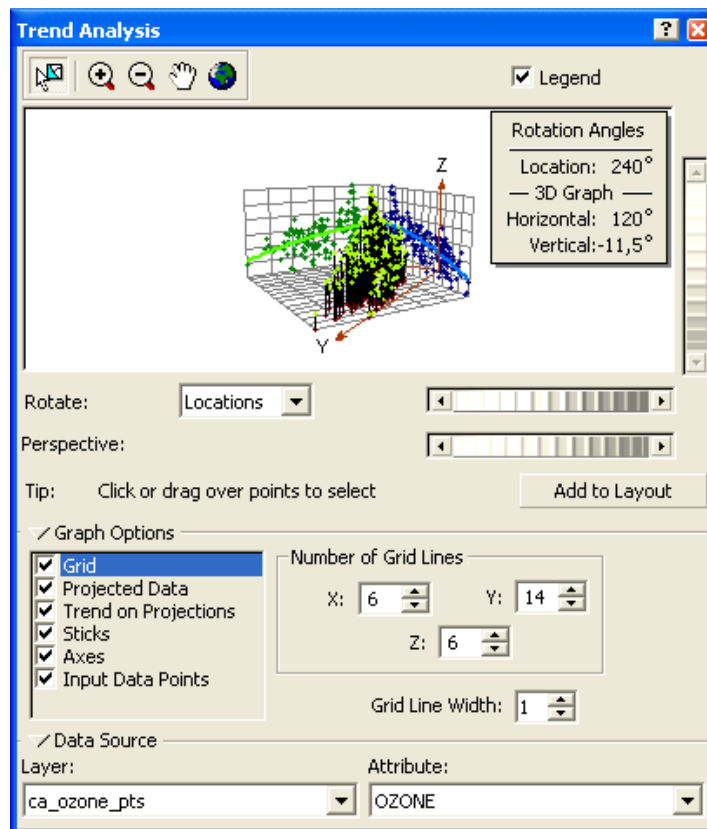


Рис. 4.4. Інструмент *Аналіз тренда* модуля *Geostatistical Analyst*

2.3 Просторова автокореляція і вплив за напрямками

1. На панелі інструментів *Geostatistical Analyst* виберіть *Дослідити дані*, а потім опцію *Semivariogram / Covariance Cloud*.

2. У вікні *Шар* виберіть *ca_ozone_pts*.

3. У вікні *Атрибут* виберіть *OZONE*.

Опція *Semivariogram / Covariance Cloud* дозволяє вивчити просторову автокореляцію між опорними точками. Просторова автокореляція базується на припущенні, що об'єкти, які розташовані найближче один до одного, найбільш схожі між собою. По осі *y* відкладаються значення варіограми, рівні квадрату різниці значень для кожної пари точок, а по осі *x* – відстань між парами точок (рис. 4.5).

Кожна червона точка на графіку *Semivariogram / Covariance Cloud* являє пару значень. Оскільки близькі за розташуванням точки повинні бути найбільше схожі, на варіограмі крайні ліві значення по осі *x* повинні мати низькі значення варіограми (знаходяться в нижній частині осі *y*). По мірі того, як відстань між парами точок збільшується (величини її зміщуються вправо по осі *x*), значення варіограми повинні також збільшуватися (зростати по осі *y*). Однак при певній відстані хмара вирівнюється, що вказує на те, що між парами точок за межами цієї відстані немає кореляції. Якщо для деяких точок, розташованих близько одна до одної (близько нуля по осі *x*), значення варіограми вищі за очікувані (велике значення по осі *y*), необхідно досліджувати ці пари точок на наявність неточностей в даних.

4. Інструментом вибору вкажіть область, в яку потрапляють точки інтересу, щоб виділити їх кольором.

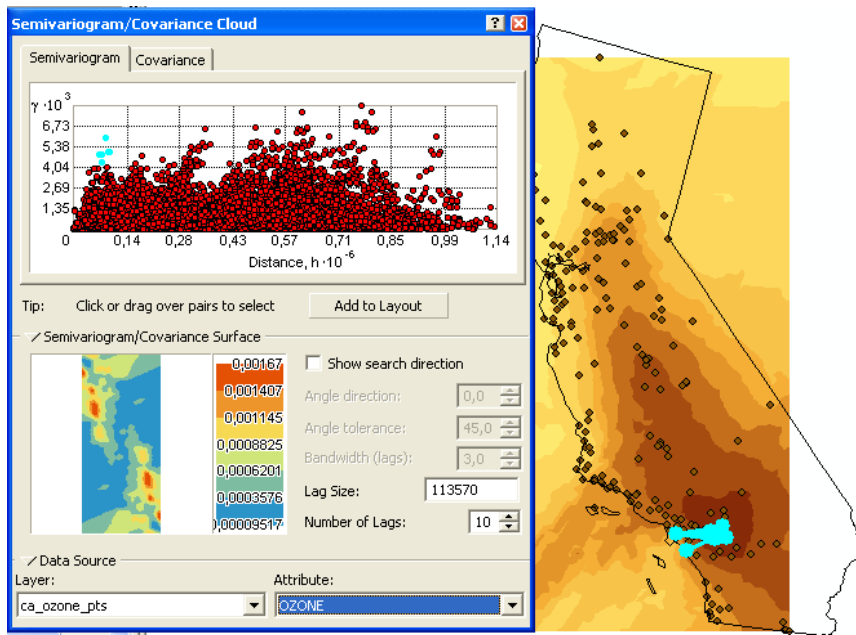


Рис. 4.5. Інструмент *Semivariogram / Covariance Cloud* модуля *Geostatistical Analyst*

Пари точок, обрані на варіограмі, виділені кольором на карті; відрізками показані зв'язки між точками, що утворюють пару.

Існує безліч причин, за якими значення відрізняються більшою мірою між точками в районі Лос-Анджелеса та іншими територіями. Одна з них – те, що в районі Лос-Анджелеса більше, ніж в інших районах, машин, що, безсумнівно, призводить до більшого забруднення і сприяє появі більш високих концентрацій озону в районі Лос-Анджелеса.

5. В меню *Вибірка (Selection)* виберіть опцію *Очистити вибрані об'єкти*.

Знаючи, що в наборі даних немає екстремальних (або помилкових) значень опорних точок і що розподіл є близьким до нормального, можна переходити до інтерполяції поверхні.

3. Картографування концентрації озону

1. На панелі інструментів *Geostatistical Analyst* виберіть опцію *Майстер операцій геостатистики*.

2. У вікні *Шар* виберіть *ca_ozone_pts*.

3. У вікні *Атрибут* виберіть *OZONE*.

4. У вікні *Методи* виберіть *крігінг*.

5. Натисніть *Далі*.

За замовчуванням будуть обрані *Ординарний крігінг* і *Карта проінтерпольованих значень*.

При дослідженні даних визначено, що в даних є глобальний тренд. Тренд найкращим чином може бути описаний поліномом другого порядку і має напрямок з південного сходу на північний захід. Цей тренд може бути представлений математичною формулою і видалений з даних. Після видалення тренда статистичний аналіз буде виконаний для залишків (складової поверхні, яка відповідає варіації на макрорівні). Тренд буде автоматично доданий назад

перед побудовою остаточної поверхні, і, таким чином, результат інтерполяції буде більш значущим.

6. У діалоговому вікні *Вибір геостатистичного методу* у розділі *Порядок тренда* виберіть *Другий*.

7. В діалозі *Вибір геостатистичного методу* натисніть *Далі*.

За замовчуванням, модуль *Geostatistical Analyst* картографує глобальний тренд для набору даних. Поверхня показує найбільш швидку зміну в напрямку з південного заходу на північний схід і більш плавну зміну в напрямку з північного заходу на південний схід (що призводить до утворення еліпса) (рис. 4.6).

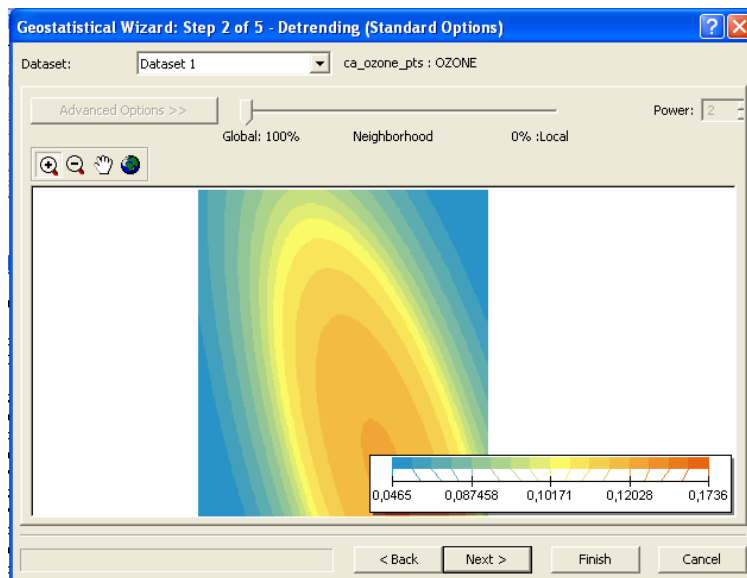


Рис. 4.6. Діалог дослідження глобального тренду в даних

Тренд в напрямку з південного заходу на північний схід, що характеризує якість повітря, може бути спричинений зосередженням озону між горами і узбережжям. Висоти і переважний напрямок вітру також є чинниками відносно низьких значень концентрації в горах і на узбережжі. Висока концентрація населення призводить до високих рівнів забруднення між горами і узбережжям. Тренд в напрямку з північного заходу на південний схід змінюється значно повільніше завдяки високій концентрації населення навколо Лос-Анджелеса і зменшенню його щільності в районі Сан-Франциско. Отже, можна обґрунтовано видалити ці тренди.

8. В діалозі *Видалення тренда* натисніть *Далі*.

3.1 Моделювання варіограми / коваріації

Метою моделювання за допомогою варіограми / коваріації є вибір кращої моделі, яка пройде через точки на варіограмі (жовта лінія на графіку).

Діалогове вікно *Моделювання варіограми / коваріації* дозволяє моделювати просторові взаємозв'язки в наборі даних. За замовчуванням розраховуються оптимальні параметри для сферичної моделі варіограми. Модуль *Geostatistical Analyst* спочатку визначає оптимальний розмір лага для згрупованих значень варіограми. *Розмір лага* – це розмір класу відстаней між точками, в який згруповані пари точок. Ця операція носить назву *бінінг (binning)*

– групування по класах. Скорочення розміру лага означає збільшення детальності моделі, що дозволяє вивчати локальні відмінності між сусідніми опорними точками (рис. 4.7).

9. Наберіть нове значення розміру лага, рівне 12000.

10. У вікні *Кількість лагів* введіть 10.

Зі зменшенням розміру лага підібрана варіограма (жовта лінія) швидко піднімається, а потім вирівнюється. *Радіус впливу (range)* – це відстань, на якій лінія вирівнюється. Таке вирівнювання варіограми вказує на те, що за межами значення радіусу впливу автокореляція невелика.

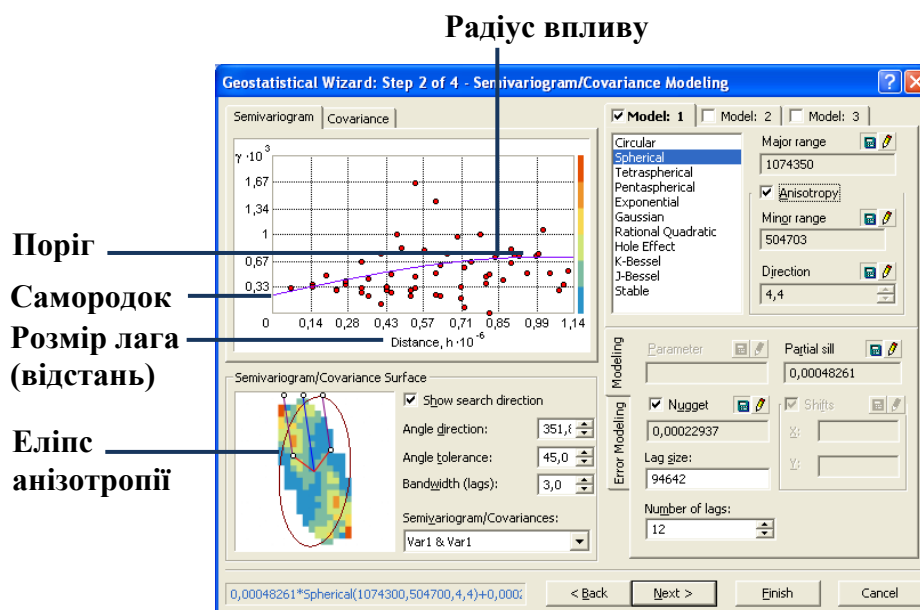


Рис. 4.7 Діалогове вікно *Моделювання варіограми / коваріації*

3.2 Варіограми за напрямками

У певних напрямках близько розташовані один до одного об'єкти можуть бути більш схожі, ніж в інших напрямках. Напрямки впливу носять назву *анізотропії* і можуть враховуватися модулем *Geostatistical Analyst*. Анізотропія може бути спричинена вітром, переносом, геологічною структурою та багатьма іншими процесами.

11. Відзначте галочкою опцію *Показати напрямок пошуку (Show Search Direction)*. В результаті відбудеться скорочення числа значень варіограми. Відобразатимуться тільки ті точки, які відповідають напрямку пошуку.

12. Утримуйте курсор на центральній лінії інструменту *Напрямок пошуку*. Міняйте напрямок інструменту пошуку. Тільки значення поверхні варіограми, розташовані в напрямку пошуку, відображаються на верхньому графіку варіограми.

13. Увімкніть опцію *Анізотропія*.

Еліпс на поверхні варіограми (рис. 4.7) вказує на радіус впливу варіограми в різних напрямках.

14. Введіть наступні параметри для інструменту *Напрямок пошуку* з тим, щоб покажчик напрямку збігся з малою віссю анізотропного еліпса:

- кут напрямку: 236,0;
- кут захвату: 45,0;
- ширина смуги (лаги): 3,0.

Форма кривої варіограми змінилась і тепер швидше досягає значення порога. Координати x і y подано у метрах, отже, радіус впливу в даному напрямку складає приблизно 74 км.

Введіть наступний параметр для інструменту *Напрямок пошуку* з тим, щоб покажчик напрямку збігся з великою віссю анізотропного еліпса: кут напрямки – 340,0.

Крива моделі варіограми зростає більш поступово, а потім вирівнюється. Радіус впливу в цьому напрямку становить 114 км.

Радіус впливу – це відстань, при якій модель варіограми досягає граничного значення (*порогу*). За межами цього радіуса впливу відмінності між точками при збільшенні розміру лага залишаються постійними. Лаг визначається відстанню між парами точок. Точки, віддалені на відстань лага, більшу, ніж радіус впливу, просторово не корелюють. *Самородок* представляє помилку вимірів та/або варіацію на мікрорівні.

15. Натисніть *Далі*.

3.3 Пошук сусідства

Зазвичай, щоб обмежити вихідні дані при інтерполяції задають область сусідства у формі кола (або еліпса). Точки, що потрапляють в задану околицю сусідства, використовуються при інтерполяції значень в точках, в яких вимірювання не проводилися. Крім того, щоб уникнути зсуву в певному напрямку, окружність (або еліпс) ділять на сектори, в кожному з яких обирається однакова кількість точок. Скориставшись діалогом *Пошук сусідства*, можна визначити кількість точок (максимум 200), радіус (або малу і велику осі еліпса), і кількість секторів круга (або еліпса), які будуть використані при виконанні інтерполяції.

Колір виділених точок вказує на те, які ваги будуть присвоєні кожній точці при обчисленні невідомих значень. У нашому прикладі, чотири точки (червоні) мають ваги більше 10 відсотків. Чим більше вага, тим більший вплив дана точка чинитиме в процесі виконання інтерполяції (рис. 4.8).

16. Клацніть мишею на графічному зображенні, щоб вибрати точку, для якої виконуватиметься інтерполяція (ця точка буде знаходитися на перехресті). Зверніть увагу на те, що змінилася вибірка опорних точок (поряд з присвоєними їм вагами). Саме ці точки будуть використані при обчисленні значення в шуканій точці.

17. Наберіть наступні координати в діалозі *Шукана точка*:

$X = -2044968$ і $Y = 208630,37$.

18. Відзначте галочкою опцію *Форма* і у вікні *Кут* наберіть 90. Зверніть увагу, як змінюється форма. Однак щоб врахувати напрямок впливу, змініть значення кута назад на 338,1.

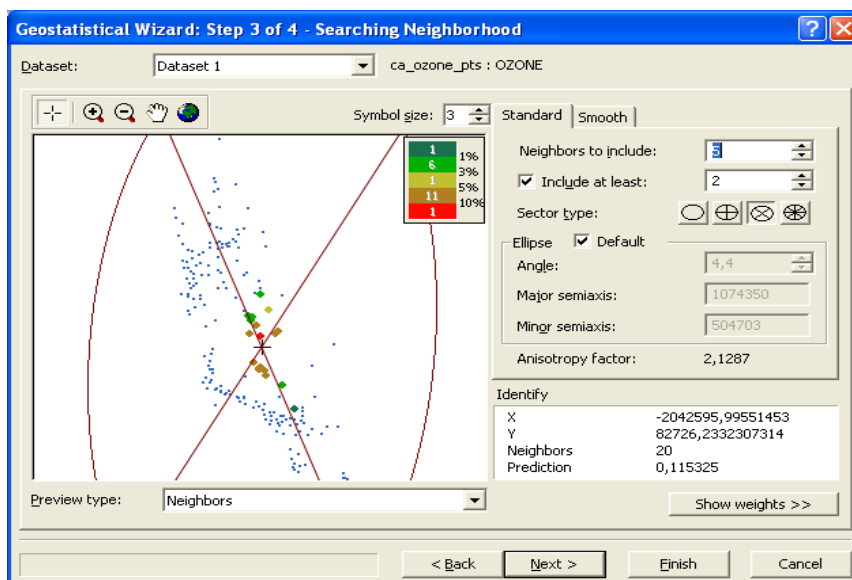


Рис. 4.8. Діалогове вікно *Пошук сусідства*

19. Відключіть опцію *Форма* – модуль *Geostatistical Analyst* буде використовувати параметри, задані за замовчуванням (раніше обчислені в діалозі *Варіограма / Коваріація*).

20. Натисніть *Далі* у діалозі *Пошук сусідства*.

Перед тим, як перейти до побудови поверхні, скористайтеся діалогом *Перехресна перевірка*, щоб виконати діагностику параметрів і визначити, наскільки доброю є модель.

3.4 *Перехресна перевірка*

Для всіх точок під час виконання перехресної перевірки виконується наступна операція: точки послідовно виключаються з вибірки, потім обчислюється значення в кожній точці з використанням позostalих даних, а потім виміряне і розраховане значення порівнюються. За сумарною статистикою перехресної перевірки можна визначити, чи підходить модель для створення карти.

Мета перехресної перевірки – допомогти прийняти обґрунтоване рішення про те, яка з моделей найбільш точно інтерполює значення. Для моделі, яка точно інтерполює значення, середня помилка повинна бути близька до 0, середньоквадратична помилка і середня стандартна помилки інтерполяції повинні мати найменше з можливих значення, а середньоквадратична нормована помилка повинна бути близька до 1.

Середня помилка – середнє відхилення розрахованих значень від фактичних. В даному випадку термін "помилка інтерполяції" використовується для позначення різниці між інтерпольованим і фактичним значенням.

21. Виберіть закладку *Графік КК* щоб переглянути графік *Квантиль-квантиль*. З даного графіка (КК) видно, що деякі значення розташовані трохи вище лінії, а деякі – трохи нижче лінії, але більшість точок знаходяться близько до прямої пунктирною лінії – це вказує на те, що помилки інтерполяції близькі до нормального розподілу.

22. Натисніть ОК.

Карта проінтерполірованих значень концентрацій озону з'явиться як верхній шар в таблиці змісту *ArcMap*. За замовчуванням, шару присвоюється назва методу крігінга, використаного для побудови поверхні (наприклад, *Ординарний крігінг*).

23. Клацніть на назві шару, щоб виділити його кольором, потім ще раз клацніть на ньому і змініть назву на *"видалений тренд"*.

24. Натисніть праву клавішу миші на створеному шарі *видалений тренд* і виберіть у відкривається меню опцію *Побудувати карту значень стандартної помилки інтерполяції*.

Стандартні помилки кількісно характеризують невизначеність інтерполяції для кожної точки побудованої поверхні. В 95 відсотках випадків справжнє значення поверхні дорівнюватиме *проінтерпольованому значенню ± 2 стандартні помилки інтерполяції*, якщо дані розподілені за нормальним законом. На поверхні стандартної помилки інтерполяції точки, розташовані ближче до точок вибірки, як правило, мають меншу помилку.

4. Порівняння моделей

Порівняння результатів двох побудованих і відображених на картах поверхонь дозволить на основі вивчення статистики перехресної перевірки прийняти зважене рішення про те, яка з них більш точно інтерполює значення концентрації озону (рис. 4.9).

1. Правою клавішею миші клацніть на шарі *"видалений тренд"*, виберіть опцію *Порівняти ...* Порівняйте цей шар з шаром, створеним на основі параметрів, запропонованих за замовчуванням.

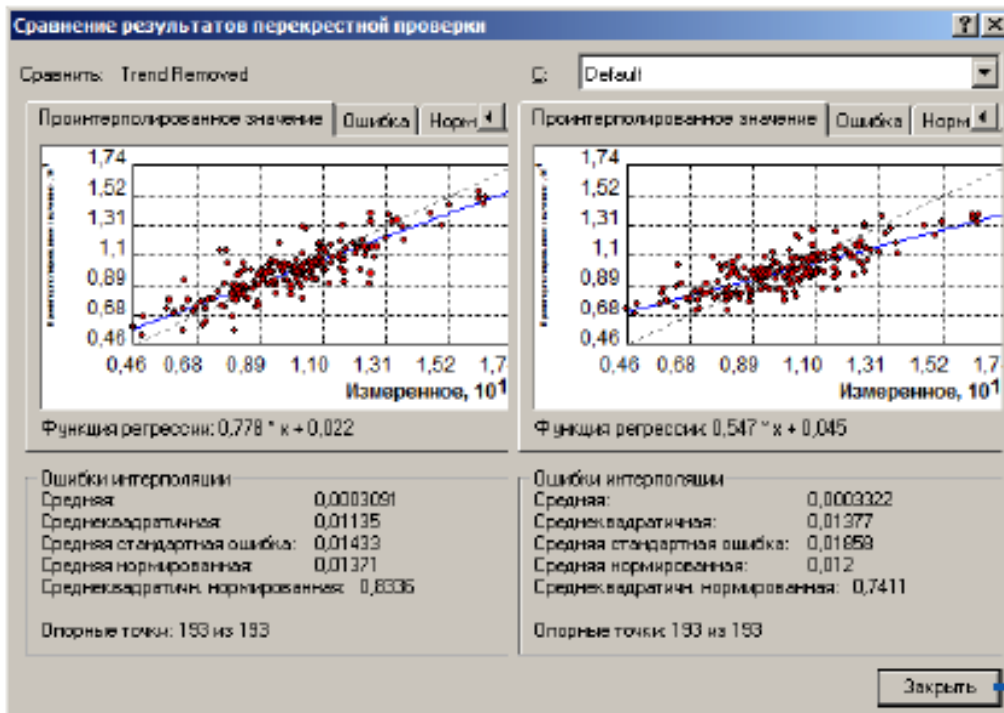


Рис. 4.9. Діалогове вікно *Перехресна перевірка*

Оскільки для шару з видаленим трендом середньоквадратична помилка інтерполяції менша, середньоквадратична нормована помилка інтерполяції ближча до одиниці, а середня помилка інтерполяції ближча до нуля, ніж для шару, побудованого з параметрами, запропонованими за замовчуванням, можна з упевненістю стверджувати, що модель з видаленим трендом є кращою і більш достовірною.

5. Оцінка протоколу рішень з використанням перевірки

Перевірка дозволяє оцінити отримані результати інтерполяції. Метою перевірки є отримання нормованих середніх помилок обчислень, приблизно рівних 0, невеликих середньоквадратичних помилок обчислень, середніх стандартних помилок, близьких до середньоквадратичних помилок обчислень, і нормованих середньоквадратичних помилок інтерполяції, приблизно рівних 1.

5.1 Створення піднаборів даних для використання їх в перевірці моделі

1. Додайте в *ArcMap* набір даних, для якого необхідно створити піднабори.
2. Виберіть опцію *Створити піднабори даних* (рис. 4.10).
3. Зі списку виберіть набір даних, для якого потрібно створити піднабір.
4. Натисніть *Далі*.
5. Необов'язковий крок: змініть місце розташування результуючого файлу бази геоданих на диску та / або його назву.
6. Необов'язковий крок: змініть назви піднаборів, запропоновані за замовчуванням.
7. Утримуючи движок, пересуньте його в потрібне положення.
8. Натисніть *Готово*.

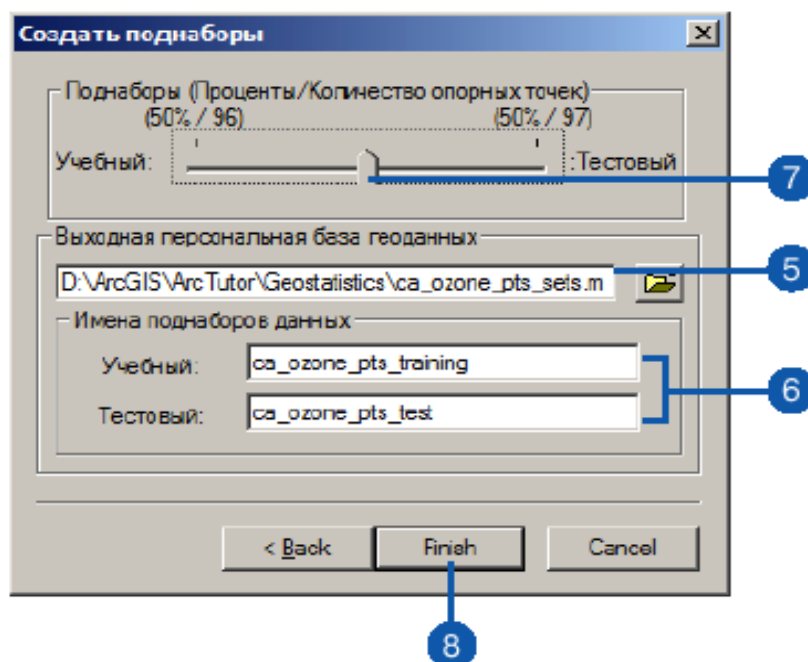


Рис. 4.10. Діалогове вікно *Створити піднабори*

У діалоговому вікні *Створити піднабори* (рис. 2.10) встановіть движок посередині, щоб розбити дані приблизно порівну між тестовим і навчальним набором.

5.2 Виконання перевірки

1. Натисніть на кнопку *Додати дані* і перейдіть до тестового і навчального наборів даних. Натисніть *Додати*.

2. Запустіть *Майстер операцій геостатистики*.

3. У вікні *Вхідні дані* виберіть навчальний набір даних.

4. У вікні *Атрибут* виберіть *OZONE*.

5. Відзначте опцію *Перевірка*.

6. У вікні *Вхідні дані* виберіть тестовий набір даних.

7. У вікні *Атрибут* виберіть *OZONE*.

8. Натискайте *Далі* в цьому і у всіх наступних діалогах доти, поки не дійдете до діалогу *Перевірка*.

9. Необов'язковий крок: збережіть таблицю з результатами перевірки в базі даних.

6. Картографування ймовірності перевищення критичного значення концентрації озону

Припустимо, що критичне значення концентрації озону становить $0,12$ ppm для восьмигодинного періоду, і необхідно дізнатися, чи існують точки, в яких це значення перевищено. Скористаємося методом індикаторного крігінга. Для цього методу не існує обмежень на нормальність розподілу даних. Значення даних перетворюються в серії нулів і одиниць відповідно до того, чи знаходяться значення даних нижче або вище критичного (порогового) рівня. Якщо порогове значення дорівнює $0,12$, будь-якому нижчому значенню буде присвоєно 0 , а значенням, що перевищують дану величину, буде присвоєно 1 .

1. На панелі *Geostatistical Analyst* виберіть *Майстер операцій геостатистики*...

2. У вікні *Вхідні дані (шар)* виберіть *ca_ozone_pts*.

3. У вікні *Атрибут* виберіть *OZONE*.

4. У вікні *Метод* виберіть *Крігінг*.

5. У діалозі *Виберіть вхідні дані і метод* натисніть *Далі*.

6. Виберіть *Індикаторний крігінг*; зверніть увагу, що обрано опцію *Карта ймовірності*.

7. Установіть первинне порогове значення рівним $0,12$.

8. Виберіть опцію *Перевищує*.

9. В діалозі *Вибір геостатистичного методу* натисніть *Далі*.

10. В діалозі *Вибір додаткових порогів* натисніть *Далі*.

11. Виберіть опцію *Анізотропія*, щоб врахувати спрямованість даних.

12. Задайте розмір лага рівним 25000 і кількість лагів, рівним 10 .

13. Натисніть *Далі* в діалозі *Моделювання варіограми / коваріації*.

14. Натисніть *Далі* в діалозі *Пошук сусідства*.

Блакитна лінія на діаграмі (рис. 4.11) показує порогове значення $0,12 \text{ ppm}$. Точкам зліва від цієї лінії присвоєно індикатор 0 , а точкам справа від лінії – індикатор 1 .

15. Прокрутіть таблицю вправо, поки не будуть відобразитися стовпці *Виміряні*, *Індикатор* і *Інтерполяція значень індикатора*.

16. Виділіть кольором рядок у таблиці із значенням індикатора, рівним 0 . Ця точка буде виділена зеленим кольором на графіку зліва від блакитної лінії, що відповідає пороговому значенню.

Стовпці з виміряними значеннями і значеннями індикатора показують дійсне і перетворене значення кожної точки вибірки. Проінтерпольовані значення індикатора можуть бути інтерпретовані як ймовірність перевищення порогових значень. Передбачувані значення індикатора обчислюються з використанням моделі варіограми на основі бінарних $(0,1)$ даних, отриманих індикаторним перетворенням вихідних даних. Перехресна перевірка послідовно виключає точку, а потім для кожної розраховує значення індикатора.

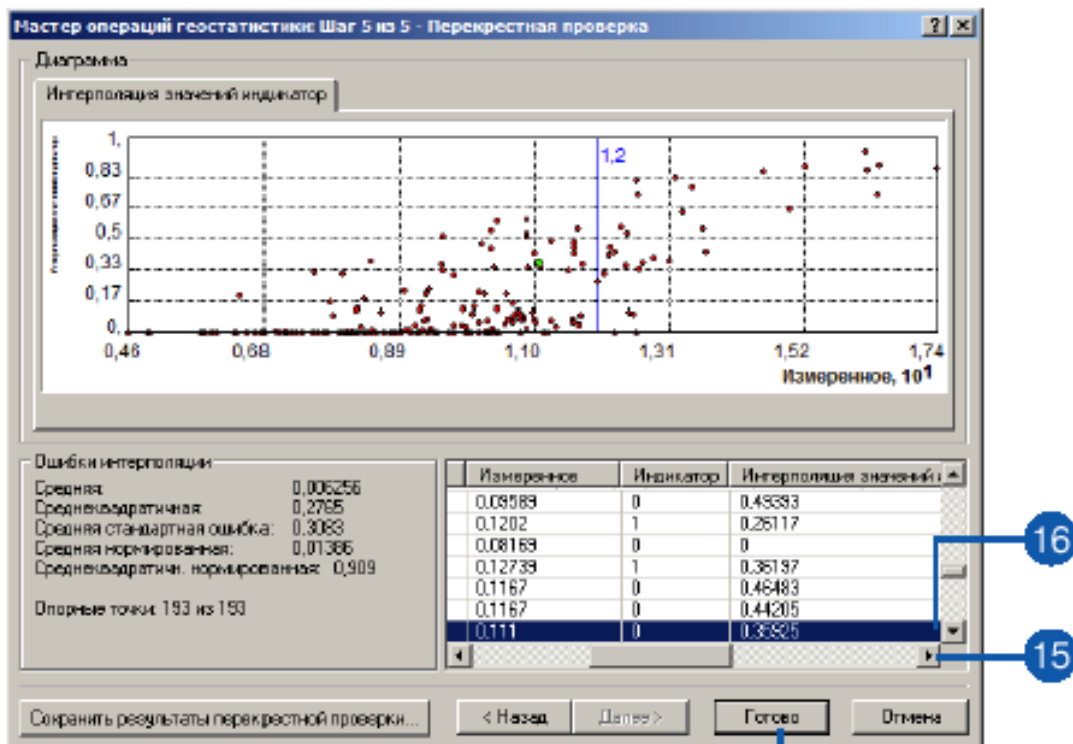


Рис. 4.11. Діалогове вікно *Перехресна перевірка (індикаторний крігінг)*

Наприклад, найбільше вимірне значення дорівнює $0,1736$. Якби вимірювання в цій точці в дійсності не проводилися, на основі методу індикаторного крігінга з імовірністю 85% було б визначено, що значення концентрації в даній точці перевищує порогове.

За отриманою картою видно, що ймовірність того, що значення концентрації в районі Лос Анджелеса (в середньому, складають менш $0,12 \text{ ppm}$ для кожного восьмигодинного інтервалу в році) перевищить встановлене порогове значення, досить висока.

Постановка завдання

Для набору даних по озону (*ca_ozone_pts*) виконати:

- побудову поверхні з використанням параметрів, запропонованих за замовчуванням;
- дослідження набору вихідних даних;
- картографування показників концентрації озону;
- порівняння результатів побудови поверхонь;
- оцінку протоколу рішень з використанням перевірки;
- картографування ймовірності перевищення порогового значення.

Відповідно до отриманих результатів виконання роботи зробити змістовні висновки. Оформити і захистити звіт.

Питання для підготовки до захисту лабораторної роботи

1. Наведіть перелік інструментів модулю Geostatistical Analyst.
2. Надайте визначення поняття «квантиль»?
3. Як визначити наявність та порядок тренду в даних?
4. Як здійснюється порівняння моделей варіограми?
5. Що представляє собою перехресна перевірка?

Лабораторна робота 5 Вивчення розподілу даних

Об'єкт – просторові дані.

Предмет – методи геостатистичного аналізу просторових даних.

Мета – закріпити знання відносно інструментів модуля Geostatistical Analyst пакету ArcGIS. Виконати дослідження набору вихідних даних з використанням методів моделювання і перетворення розподілів.

Теоретичні положення

1. Моделювання розподілів і визначення методів перетворень

1.1 Перетворення за методом Vox-Cox, арксинуса і логарифмічні перетворення

Модуль *Geostatistical Analyst* дає можливість використовувати декілька методів перетворень, включаючи метод *Vox-Cox* (також відомий як степеневі перетворення), логарифмічні перетворення і перетворення за методом арксинуса. Припустимо, що є дані спостережень в опорних точках $Z(s)$ і задане довільне перетворення $Y(s)=t(Z(s))$. Як правило, потрібно знайти перетворення, яке призведе до того, що значення $Y(s)$ будуть підчинятися закону нормального розподілу. Часто результатом перетворень є постійна дисперсія даних для всієї досліджуваної території.

Перетворення за методом *Vox-Cox* виглядає наступним чином:

$$Y(s) = (Z(s)^\lambda - 1) / \lambda, \lambda \neq 0$$

Припустимо, що дані складаються з підрахунків зустрічальності якогось явища. Для таких типів даних дисперсія часто співвідноситься з середнім значенням. Це означає, що якщо для частини досліджуваної території кількість підрахунків є малою, мінливість в цьому районі буде меншою, ніж мінливість в іншому районі, в якому кількість підрахунків є більшою. В цьому випадку, перетворення, що використовує квадратний корінь, допоможе зробити дисперсії більш постійними для всієї досліджуваної території і, крім цього, воно часто приводить дані до нормального розподілу. Перетворення за методом квадратного кореня є окремим випадком перетворення *Vox-Cox* ($\lambda = 0.5$).

Логарифмічне перетворення, по суті, є також окремим випадком перетворення за методом *Vox-Cox* при $\lambda = 0$; формула перетворення має такий вигляд:

$$Y(s) = \ln(Z(s)), Z(s) > 0$$

Наслідком логарифмічного перетворення є метод інтерполяції, відомий як логнормальний крігінг, тоді як для всіх інших значень λ відповідний метод інтерполяції відомий як трансгаусівський крігінг. Логарифмічне перетворення часто використовується в тих випадках, коли дані мають асиметричний розподіл і в них присутні великі значення. Ці значення можуть бути локалізовані на досліджуваній території, а логарифмічне перетворення допоможе зробити дисперсії більш постійними і нормалізувати дані.

Перетворення за методом арксинуса виглядає наступним чином:

$$Y(s) = \sin^{-1}(Z(s)), Z(s) \geq 0, Z(s) \leq 1$$

Перетворення за методом арксинуса може бути використане для відносних даних або даних, виражених у відсотках. Часто, коли дані відносні, дисперсія найменша для значень, близьких до 0 і 1, і найбільша, коли значення близькі до 0,5. Тоді перетворення за методом арксинуса часто призводить до того, що дисперсії постійні для всієї досліджуваної території і, крім того, часто його результатом є нормальний розподіл даних.

1.2 Перетворення за методом нормальних міток

Перетворення за методом нормальних міток впорядковує значення набору даних від найменших до найбільших і зіставляє ці класи з відповідними класами графіка нормального розподілу. Перетворення здійснюється наступним чином: значення беруться з відповідного класу нормального розподілу. Це можна бачити на діаграмах (рис. 5.1). На діаграмі рис. 5.1-а показана гістограма, яка зазвичай будується при проведенні дослідницького аналізу даних. По-іншому дані можна відобразити, скориставшись графіком кумулятивного розподілу, який показаний на рис. 5.1-б.

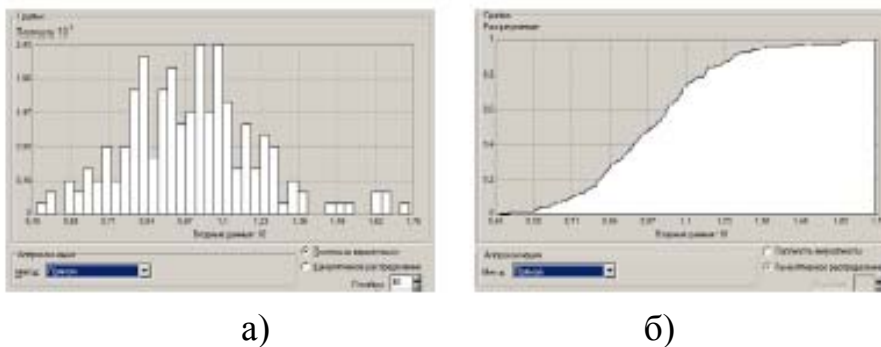


Рис. 5.1. Перетворення за методом нормальних міток: а – гістограма; б – графік кумулятивного розподілу

Щоб отримати перетворення за методом нормальних міток, треба скористатися кумулятивним розподілом. Візьміть кумулятивний розподіл спостережень і зіставте його з кумулятивним розподілом стандартного нормального розподілу. Це добре видно при розгляді процедури зіставлення графіків, показаний на рис. 5.2.

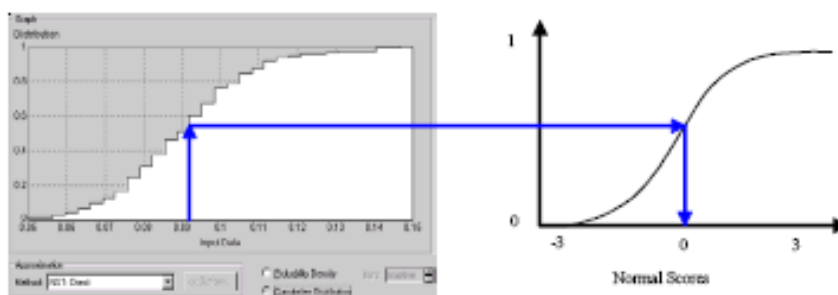


Рис. 5.2. Зіставлення кумулятивного розподілу спостережень з кумулятивним розподілом стандартного нормального розподілу

В даному прикладі значення вихідних даних, рівне 0,09 за методом нормальних міток, перетвориться в значення, приблизно рівне до нуля. У модулі *Geostatistical Analyst* можна використовувати три методи апроксимації: *прямий*, *лінійний* і *гаусова kern-функція*.

Прямий метод використовує спостережуваний кумулятивний розподіл; *лінійний метод* підбирає лінії між кожним "кроком" кумулятивного розподілу; *метод гауссової kern-функції* апроксимує імовірнісний розподіл шляхом підбору лінійної комбінації з кумулятивних розподілів щільності. Після виконання інтерполяції в перетвореному масштабі необхідно виконати зворотне перетворення, щоб привести обчислені значення до меншого масштабу.

Вибір методу апроксимації залежить від зроблених припущень і від ступеня згладжування апроксимації. *Прямий метод* – найменш згладжений і характеризується найменшою кількістю припущень; *лінійний метод* займає проміжне положення, а *метод гаусовських kern-функцій* характеризується найбільш згладженим зворотним перетворенням, але поряд з цим і найсерйознішими допущеннями (наприклад, розподіл даних може бути апроксимований за допомогою поєднання обмеженої кількості нормальних розподілів).

1.3 Використання перетворень (логарифмічного, за методом Vox-Cox і арксинуса)

Використання перетворень дозволяє зробити дисперсії постійними для всієї досліджуваної території та наблизити дані до нормального розподілу.

Скористайтеся *гістограмою* і *нормальними графіками КК* щоб, випробувавши різні методи перетворень, отримати нормальний розподіл даних. Деякі методи геостатистики вирішальним чином залежать від даних, що підкоряються нормальному розподілу – наприклад, *диз'юнктивний крігінг*, *карти квантилів* та ймовірності для *ординарного*, *простого* і *універсального крігінга*. Тому виконання перетворень може привести дані до розподілу, яке ближче до нормального, ніж у вихідних даних.

Опція *виконання перетворень* може бути використана, коли в діалозі *Виберіть вхідні дані і метод* ви обираєте методи *крігінга* і *кокрігінга* (рис. 5.3).

1. В діалозі *Вибір геостатистичного методу* виберіть потрібний метод перетворень з меню *Перетворення*.

2. Натисніть *Далі*.

3. Переходьте від діалогу до діалогу для того, щоб побудувати поверхню.

1.4 Використання перетворень за методом нормальних міток

За методом нормальних міток дані будуть перетворені і приведені до одновимірного нормального розподілу, що дозволить застосувати до них методи простого та диз'юнктивного крігінга. Бажано порівняти підбрану модель з емпіричної функцією кумулятивного розподілу для кожного з трьох методів апроксимації, використаних в перетвореннях за методом нормальних міток.

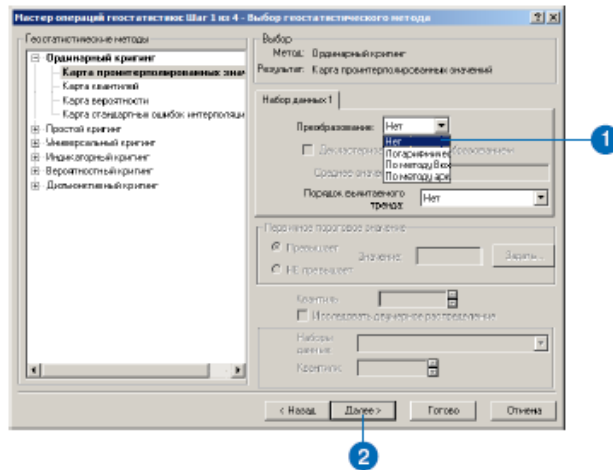


Рис. 5.3. Майстер операцій геостатистики: вибір методу перетворень

1.5 Моделювання розподілів

1. Для дослідження розподілу виберіть інтерполяцію за методом простого або диз'юнктивного крігінга.

2. В діалозі *Вибір геостатистичного методу* з меню *Перетворення* виберіть опцію *За методом нормальних міток*.

3. Натисніть *Далі*.

4. Щоб поміняти спосіб відображення графіка, виберіть опцію *Кумулятивний розподіл*.

5. Задайте кількість стовпців, які мають бути відображені на гістограмі.

6. У вікні *Вибір набору даних* виберіть інший набір даних (тільки для *кокрігінга* у випадку, коли у вас є два або більше наборів даних).

2. Перевірка на двовимірний нормальний розподіл

Диз'юнктивний крігінг вимагає дотримання умови підпорядкування даних двовимірному нормальному розподілу. Крім цього, для того, щоб створити карти ймовірності та квантилів, припустимо, що дані підпорядковуються повному багатовимірному нормальному розподілу. Щоб перевірити дані на відповідність одномірному нормальному розподілу можна скористатися *нормальними графіками КК*. Тепер перевіримо дані на відповідність двовимірному нормальному розподілу (жодна з цих перевірок не гарантує, що дані підпорядковуються повному багатовимірному нормальному розподілу, але іноді розумно висунути це припущення, ґрунтуючись на використанні інструментів діагностики). Розглянемо наступне ймовірнісне твердження:

$$f(p, h) = \Pr \text{ob}[Z(s) \leq z_p, Z(s + h) \leq z_p],$$

де z_p – стандартний нормальний квантиль для деякої ймовірності p .

Наприклад, відомі значення стандартних нормальних квантилів: при $p = 0,975$, $z_p = 1,96$; при $p = 0,5$, $z_p = 0$; а при $p = 0,025$, $z_p = -1,96$. Вираз для ймовірності, наведений вище, бере значення змінної Z в точці s та іншої змінної Z в якійсь точці $s + h$ і дає значення ймовірності того, що обидва цих значення менші за z_p . Ця формула ймовірності є функцією $f(p, h)$, яка залежить від p (і,

відповідно, z_p) і h . Функція буде також залежати від ступеня автокореляції між $Z(s)$ і $Z(s + h)$. Тепер припустимо, що $Z(s)$ і $Z(s + h)$ підпорядковуються двовимірному нормальному розподілу. Якщо значення автокореляції відомо, для функції $f(p, h)$ існують формули. Припустимо, що h – константа, і тільки значення p змінюється.

Припустимо, що значення p залишається постійним і функція $f(p, h)$ змінюється в залежності від h . Припустимо, що значення h дуже мале. В цьому випадку ймовірність $Prob[Z(s) \leq z_p, Z(s + h) \leq z_p]$ дуже близька до ймовірності $Prob[Z(s) \leq z_p] = p$. Далі, припустимо, що значення h дуже велике. В цьому випадку ймовірність $Prob[Z(s) \leq z_p, Z(s + h) \leq z_p]$ дуже близька до ймовірності $Prob[Z(s) \leq z_p] = Prob[Z(s + h) \leq z_p] = p^2$ (бо значення $Z(s)$ і $Z(s + h)$ практично незалежні). Таким чином, для фіксованого значення p , слід очікувати, що значення функції $f(p, h)$ змінюватимуться від p до p^2 . Тепер розглянемо вираз $f(p, h)$ як функцію обох параметрів p і відстані h (рис. 5.4).

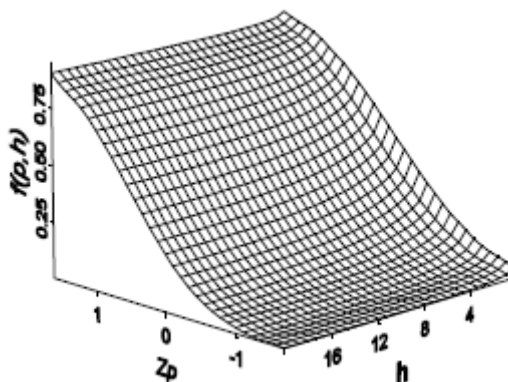


Рис. 5.4. Графічне відображення функції $f(p, h)$

Ця функція може бути конвертована в варіограми і коваріаційні функції для індикаторів. Якщо ймовірність

$$f(p, h) = Prob[Z(s) \leq z_p, Z(s + h) \leq z_p] = E[I(Z(s) \leq z_p), I(Z(s + h) \leq z_p)],$$

де I (твердження) – індикаторна функція, яка дорівнює 1, якщо твердження істинне, або 0, якщо твердження хибне, тоді коваріаційна функція для індикаторів з фіксованим значенням p виглядає наступним чином:

$$C_1(h; p) = f(p, h) - p^2$$

а варіограма для індикаторів з фіксованим p :

$$g_1(h; p) = p - f(p, h)$$

Отже, можна оцінити варіограму і коваріаційну функцію індикаторів вихідних даних і використовувати цю оцінку для отримання очікуваних варіограм і коваріаційних функцій індикаторів для різних значень p . Наприклад, вони можуть виглядати як на рис. 5.5.

Червоні точки в діалозі *Моделювання варіограми / коваріації* – це значення емпіричної коваріації і варіограми для індикаторних змінних. Зелена лінія – крива, побудована теоретичним способом, для індикаторної варіограми або індикаторної коваріації, яка припускає, що вихідні дані підкоряються двовимірному нормальному розподілу, а жовта лінія – модель, підібрана для

вихідних індикаторних даних. Таким чином, зелена лінія і жовта лінія повинні співпасти, якщо дані мають двовимірний нормальний розподіл.

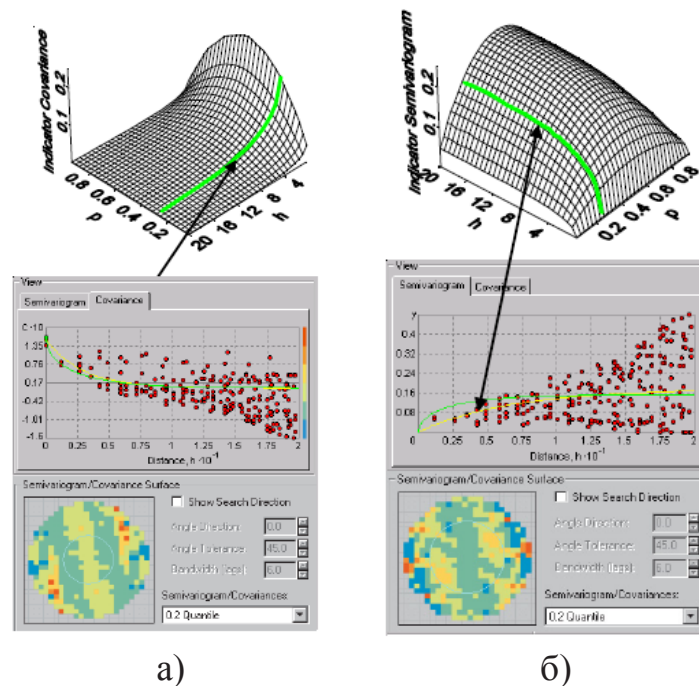


Рис. 5.5. Приклади графічного відображення функцій: а – коваріації; б – варіограми.

2.1 Перевірка двовимірного розподілу

В діалозі *Вибір геостатистичного методу* виберіть метод *крігінга* або *кокрігінга*, а потім метод *простого крігінга / кокрігінга* або *диз'юнктивного крігінга / кокрігінга*.

1. На панелі інструментів *ArcMap* натисніть кнопку *Додати дані* і додайте шар, дані якого ви хочете перевірити на двовимірний розподіл.

2. Запустіть *Майстер операції геостатистики*.

3. Виберіть метод *крігінга* або *кокрігінга*.

4. Натисніть *Далі* в діалозі *Виберіть вхідні дані і метод*.

5. Виберіть метод *простого крігінга / кокрігінга* або *диз'юнктивного крігінга / кокрігінга*.

6. Відзначте опцію *Дослідити двовимірний розподіл* і виберіть опцію *Перетворення за методом нормальних міток*.

7. Виберіть комбінацію наборів даних, яку ви хочете використовувати (тільки для *кокрігінга*).

8. Наберіть значення кількості перевірених квантилів.

9. Натисніть *Далі*.

10. Виберіть *Метод апроксимації*, задайте параметри і відзначте або опцію *Щільність ймовірності*, або *Кумулятивний розподіл* у діалозі *Перетворення за методом нормальних міток*. Натисніть *Далі*.

11. Задайте необхідні параметри в діалозі *Моделювання варіограми / коваріації* і натисніть *Далі*.

12. Вивчіть двовимірний розподіл в діалозі *Дослідження двовимірного розподілу*. Натисніть *Далі*.

3. Застосування декластеризації для даних, відібраних з різною щільністю

Дуже часто опорні точки не розподілені в просторі рівномірно або розподілені випадковим чином. З різних причин дані можуть відбиратися в одних районах з більшою щільністю, ніж в інших. Для правильного застосування перетворень за методом нормальних міток важливо, щоб гістограма (і кумулятивний розподіл) вибірки правильно відображала гістограму для всієї сукупності даних. Якщо дані нерівномірно розподілені в просторі і при цьому просторово автокорелюють, результуюча гістограма вибірки може не відповідати гістограмі для даних всієї сукупності (рис. 5.6).

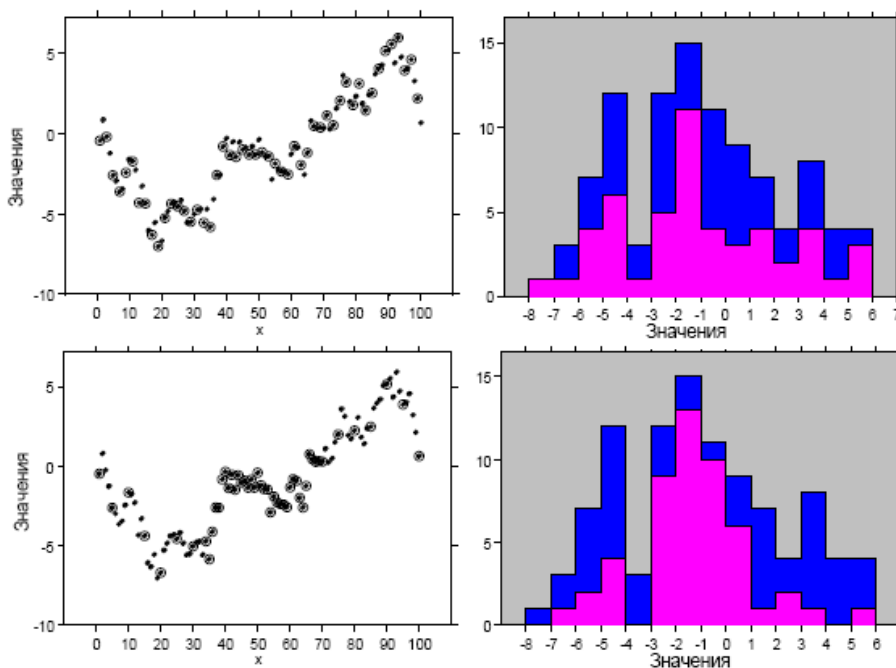


Рис. 5.6. Вибіркові гістограми

На лівому верхньому рисунку (рис. 5.6) вся сукупність значень в 100 точках вздовж лінії показана чорними крапками. Ці значення були отримані з використанням просторово корельованого процесу з постійним середнім значенням і сильною додатньою кореляцією. Вибірка – це всі інші точки, починаючи з першої, показані маленькими крапками. Праворуч показані гістограми: синім кольором – гістограма сукупності, малиновим кольором – гістограма вибірки. Оскільки вибірка дорівнює приблизно половині всієї сукупності, стовпчики гістограми вибірки повинні бути рівні половині висоти стовпчиків гістограми сукупності з деякими відхиленнями. На лівому нижньому рисунку (рис. 5.6) вибірка сформована з різною щільністю: до 34-ї точки відібрана кожна п'ята, потім до 70-ї відібрані всі точки, а потім до кінця – знову кожна п'ята. Зрештою вибірка знову включає половину точок генеральної сукупності. Різна щільність призводить до того, що вибірка містить високу частку значень, близьких до середнього, і, отже, стовпчики гістограми приблизно рівні стовпчикам гістограми для всієї сукупності значень від -3 до 1.

Внаслідок цього на гістограмі вибірки погано представлені дуже низькі і дуже високі значення. Одним з рішень, яке може бути використане для вибірки з різною густиною, є зважування даних; при цьому опорним точкам, розташованим на ділянках з високою щільністю, присвоюються менші ваги, а точкам, розташованим на території розріджено, присвоюються великі ваги (що призведе до збільшення висоти стовпчиків гістограми для низьких і високих значень даних).

Модуль *Geostatistical Analyst* дозволяє використовувати два методи. Метод, запропонований за замовчанням – це *декластеризація за осередками*. При використанні цього методу прямокутні осередки утворюють ґрид, поєднаний з опорними точками, і кожній точці осередку присвоюється вага, обернено пропорційна кількості точок, що потрапляють в цей осередок (рис. 5.7).

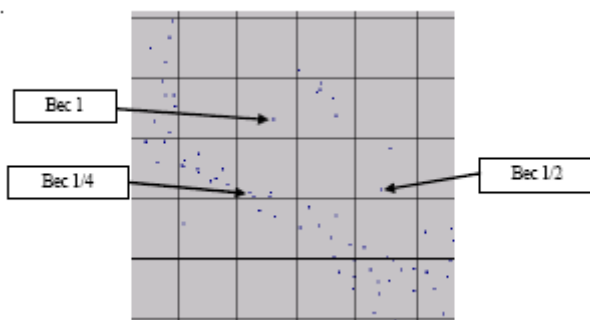


Рис. 5.7. Приклад декластеризації за осередками

Один з інструментів модуля *Geostatistical Analyst* дозволяє побудувати графік, на якому показано зважене середнє значення для всіх даних при різних розмірах осередків. Пропонується вибрати розмір осередку, що відповідає мінімальному зваженому середньому, якщо дані були відібрані з різною щільністю в районах з високими значеннями, і, навпаки, вибрати розмір осередку, який відповідає максимальному зваженому середньому, якщо дані були відібрані з різною щільністю в районах з низькими значеннями. Інша схема використовує полігональний метод, при якому для кожної опорної точки полігон визначається таким чином, щоб всі точки, що потрапляють всередину цього полігону, були б ближче до цієї точки, ніж до будь-якої іншої опорної точки (рис. 5.8).

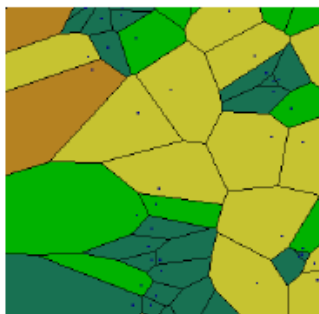


Рис. 5.8. Приклад декластеризації з використанням полігонального методу

Опорні точки показані маленькими точками, а навколо них намальовані полігони, колір яких відповідає розміру полігонів. Ідея полягає в тому, щоб

присвоїти кожній опорній точці вагу, пропорційну площі, яку вона "утворює". Проблема цього методу полягає в тому, що для крайових точок важко коректно визначити значення ваг. Граничним точкам часто можуть бути присвоєні завищені значення ваг, оскільки полігон простягається до границі області дослідження. У модулі *Geostatistical Analyst* границя – це прямокутник, наявність якого часто призводить до того, що граничним точкам присвоюються високі значення (рис. 5.9).

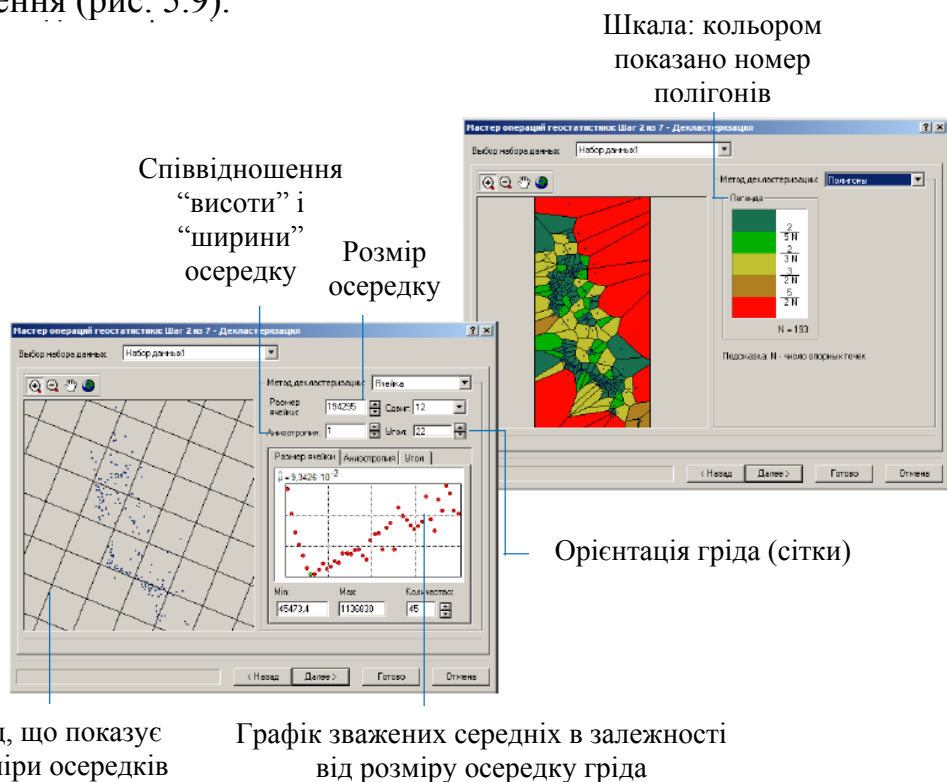


Рис. 5.9. Діалог декластеризації даних модуля *Geostatistical Analyst*

3.1 Виконання декластеризації за методом осередку (рис. 5.10)

1. В діалозі *Виберіть вхідні дані і метод* виберіть *крігінг* або *кокрігінг*.
2. Виберіть один з методів *крігінга* або *кокрігінга*: *ймовірнісний*, *диз'юнктивний* або *простий*.
3. В діалозі *Перетворення* виберіть *перетворення за методом нормальних міток*.
4. Виберіть опцію *Декластеризація* перед перетворенням.
5. Натисніть *Далі*.
6. В меню *Вибір набору даних* визначте набір даних, який ви хочете відобразити (тільки для *кокрігінга*).
7. Задайте необхідні параметри.
8. Вибирайте різні закладки для того, щоб переглянути діаграми, що відображають зміну вибірки залежно від розміру осередку, анізотропії та кута нахилу осередку.
9. Змініть розмір осередку, анізотропію, зміщення і кут нахилу осередку, щоб визначити екстремум графіка.

10. З меню *Метод декластеризації* виберіть *Полігони*, щоб відобразити полігони, за якими буде виконуватися декластеризація.

11. Натисніть *Далі*.

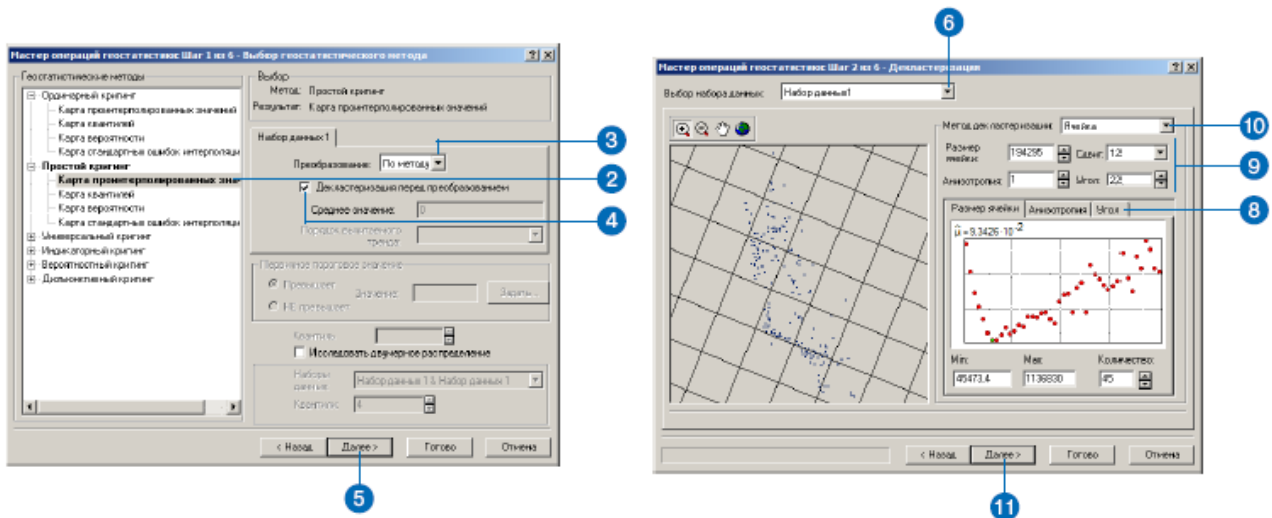


Рис. 5.10. Діалоги декластеризації даних за методом осередку

12. Із меню *Апроксимація* виберіть метод апроксимації, визначте параметри і відзначте опцію *Щільність ймовірності* або *Кумулятивний розподіл* у діалозі *Перетворення за методом нормальних міток*. Натисніть *Далі*.

13. В діалозі *Моделювання варіограми / коваріації* задайте необхідні параметри і натисніть *Далі*.

14. Визначте параметри в діалозі *Пошук сусідства* і натисніть *Далі*.

15. Вивчіть результати в діалозі *Перехресна перевірка* і натисніть *Готово*.

16. У діалозі *Інформація про результуючий шар* натисніть *ОК*.

Постановка завдання

Для набору даних по озону (*ca_ozone_pts*) виконати:

- перетворення даних за методом *Box-Cox*, арксинуса і логарифмічне перетворення;
- перетворення за методом нормальних міток;
- моделювання розподілів;
- перевірку на двовимірний нормальний розподіл;
- декластеризацію даних.

Відповідно до отриманих результатів виконання роботи зробити змістовні висновки. Оформити і захистити звіт.

Питання для підготовки до захисту лабораторної роботи

1. Наведіть методи моделювання і перетворення розподілів.
2. Що являє собою перевірка на двовимірний нормальний розподіл?
3. Розкрийте зміст поняття «декластеризація».
4. З якою метою застосовується декластеризація для даних?
5. Розкрийте сутність перетворення за методом нормальних міток.

Лабораторна робота 6

Визначення ступеня забруднення ґрунтових вод м. Дніпро

Об'єкт – дані про забруднення ґрунтових вод м. Дніпро.

Предмет – методи геостатистичного аналізу просторових даних.

Мета – закріпити знання відносно інструментів модуля Geostatistical Analyst пакету ArcGIS. Отримати навички практичного використання методів геостатистичного аналізу на прикладі дослідження реальних даних екологічного моніторингу ґрунтових вод м. Дніпро.

Теоретичні положення

В роботі розглядаються наступні набори даних:

| Набір даних | Опис |
|--------------------|---|
| <i>Border</i> | Картографічна основа м. Дніпро |
| <i>Well</i> | Контрольні точки виміру концентрацій забруднюючих речовин в ґрунтових водах |
| <i>Pollutants</i> | Таблиця, що містить дані про концентрацію забруднюючих речовин в контрольних точках – свердловинах (мг/дм ³). Поля таблиці: well – номер свердловини, cobalt – кобальт; selenium – селен; arsenic – миш'як; lead – свинець; barium – барій; aluminium – алюміній; strontium – стронцій; nitrate – нітрати; fluorine – фтор; copper – мідь; manganese – марганець; iron – залізо; chromium – хром; sulphate – сульфати; chlorine – хлор; phenol – фенол; complex – комплексний показник. |

Примітка: гранична допустима концентрація **кобальту** в питній воді складає ГДКв = 0,1 мг/дм³, а в водоймах рибогосподарського призначення – ПДКвр = 0,01 мг/ дм³; гранично допустима концентрація селену у питній воді становить 0,01 мг/ дм³; ГДК миш'яку – 0,05 мг/ дм³; ГДК свинцю – 0,03 мг/ дм³; ГДК барію – 0,1 мг/ дм³; ГДК алюмінію – 0,2 мг/ дм³; ГДК нітратів у воді становить 45 мг/ дм³.

Постановка завдання

1. Виконати перетворення розподілу значень вихідних даних.
2. Виконати перевірку даних на двовимірне нормальне розподіл.
3. Виконати декластерізацію мережі моніторингу.
4. Побудувати карту ймовірності перевищення ГДК першого із заданих речовин.
5. Побудувати карту забруднення ґрунтових вод за двома показниками з використанням методу інтерполяції кокрігінг.
6. Побудувати карти помилок інтерполяції.
7. Оформити звіт, зробити змістовні висновки.

Варіанти

| Номер комп'ютера | Речовина 1 | Речовина 2 | Номер комп'ютера | Речовина 1 | Речовина 2 |
|------------------|------------|------------|------------------|------------|------------|
| 1 | Кобальт | Хром | 7 | Стронцій | Барій |
| 2 | Селен | Фенол | 8 | Нітрати | Сульфати |
| 3 | Миш'як | Мідь | 9 | Фтор | Селен |
| 4 | Свинець | Залізо | 10 | Мідь | Кобальт |
| 5 | Барій | Фтор | 11 | Марганець | Хлор |
| 6 | Алюміній | Свинець | 12 | Залізо | Миш'як |

Питання для підготовки до захисту лабораторної роботи

Вказати, які інструменти модуля Geostatistical Analyst пакету ArcGIS дозволяють:

1. виділити на карті точки, значення атрибута яких знаходиться в заданому інтервалі;
2. виділити на карті пари точок, що розташовані на мінімальній відстані одна від одної та характеризуються низькими значеннями різниць атрибута;
3. визначити наявність або відсутність тренда у вихідному наборі даних;
4. побудувати просторовий розподіл ймовірностей перевищення заданого значення атрибута;
5. здійснювати порівняння моделей варіограми.

Лабораторна робота 7

Геостатистичний аналіз даних в S-GEMS

Об'єкт – дані гідрохімічних досліджень.

Предмет – методи геостатистичного аналізу та стохастичного моделювання просторових даних.

Мета – закріпити знання відносно методів геостатистичного аналізу просторових даних. Використовуючи інструментарій програмного забезпечення S-GeMS, оволодіти методами стохастичного моделювання.

Теоретичні положення

Програмне забезпечення *S-GeMS*¹ (*Stanford Geostatistical Modeling Software*) містить повний набір алгоритмів геостатистичного аналізу, моделювання просторових даних і 3D візуалізації. В *S-GeMS* не передбачені можливості статистичного аналізу даних, що вимагає застосування додаткового програмного забезпечення, наприклад *MS Excel*, *ArcGIS*.

Вихідні дані – дані гідрохімічних досліджень Тусонського басейну (1984 р.), надані у відкритий доступ дослідницькою групою університету Арізони. Вихідний файл (*Z:\Subjects\Геостатистика\SGeMS\tucwate.dat*) містить дані по 114 свердловинах. Вміст файлу доступний для перегляду в будь-якому текстовому редакторі.

Заголовок файлу виглядає наступним чином:

| | |
|------------------------|--|
| <i>ID no</i> | – номер свердловини; |
| <i>Easting</i> | – довгота; |
| <i>Northing</i> | – широта; |
| <i>Calcium</i> | – вміст кальцію (мг/л); |
| <i>Magnesium</i> | – вміст магнію (мг/м ³); |
| <i>Sodium</i> | – вміст натрію (мг/л); |
| <i>Potassium</i> | – вміст калію (мг/м ³); |
| <i>HCO₃</i> | – вміст кислих солей вугільної кислоти (мг/л); |
| <i>SO₄</i> | – оксид сірки (мг/л); |
| <i>Cl</i> | – вміст хлору (мг/л); |
| <i>NO₄</i> | – оксид азоту (мг/л). |

1. Завантаження даних

Для завантаження набору даних виберіть пункт об'єкта *Load Object ...* меню *Objects*. В якості типу об'єкта виберіть *point set* у випадіючому списку *Select Object*. Натисніть *Next*

Здайте ім'я для множини точок, наприклад, *Tucson_water*. Встановіть параметри – номери стовпців (*X* – стовпець 2, *Y* – стовпець 3 і відсутність даних по стовпцю *Z* (множину даних задано на площині). Для завершення завантаження даних натисніть кнопку *Finish* (рис. 7.1).

¹ <http://sgems.sourceforge.net>

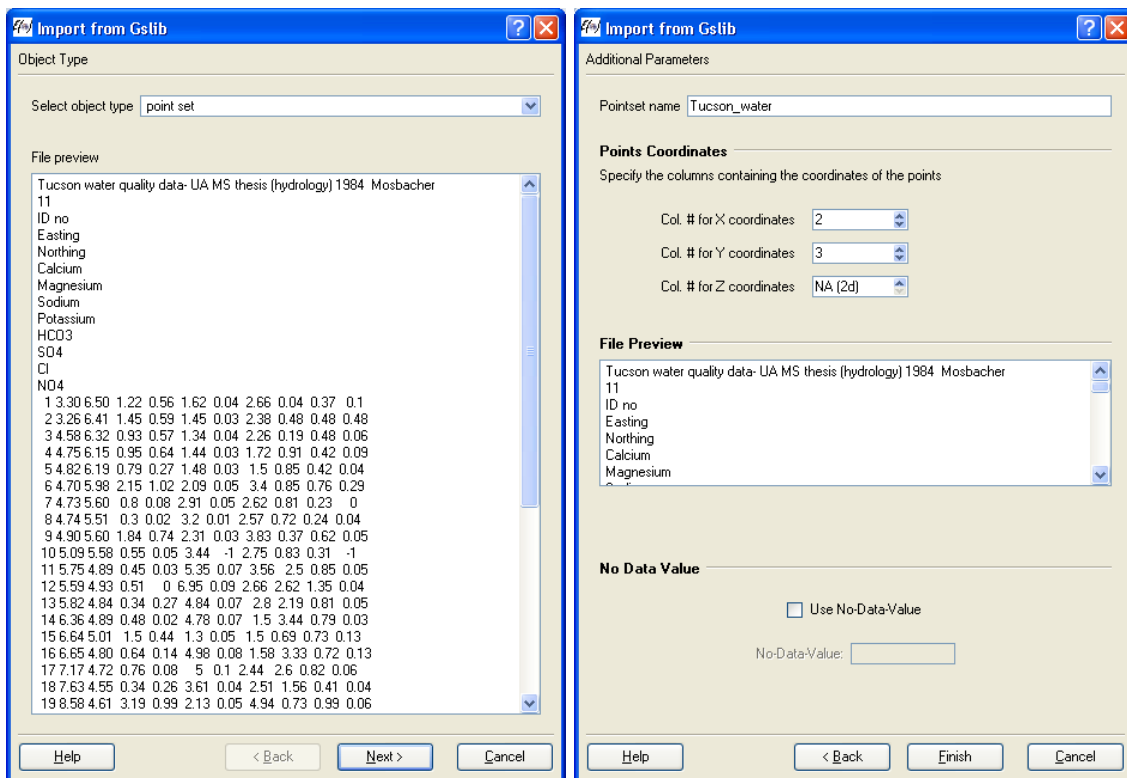


Рис. 7.1. Параметри завантаження даних

Зверніть увагу, що при відображенні *Commands Panel (View-Commands Panel)* (у нижній частині вікна *S-GeMS*) на закладці *S-GeMS Commands History* мається можливість доступу до історії команд *S-GeMS*. У цьому вікні будуть відображатися команди внутрішньою мовою *S-GeMS*. Після завантаження файлу на панелі команд *S-GeMS* з'явиться відповідна інформація (рис. 7.2).

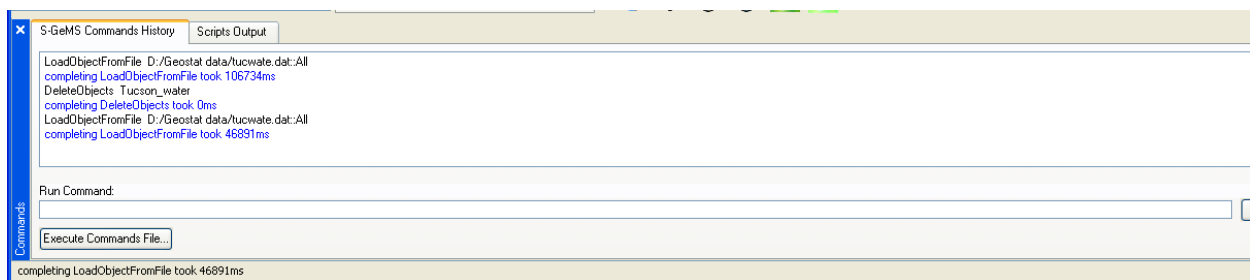


Рис. 7.2. Відображення історії команд *S-GeMS* у вікні *Commands Panel*

Даний засіб дозволяє відображати команди, пов'язані з кожною виконуваною дією, і згодом вводити окремі команди в текстове поле *Run Command* або перетворювати послідовності команд в сценарії і запускати сценарії за допомогою кнопки *Execute Commands File ...*

2. Попередній аналіз даних

Меню *Data Analysis* містить базові інструменти для дослідження розподілу даних. Для побудови гістограми розподілу значень вмісту HCO_3 в пробах води виберіть пункт *Histogram* даного меню, у списку *Object* вкажіть *Tucson_water*; в списку *Property* виберіть HCO_3 (рис. 7.3).

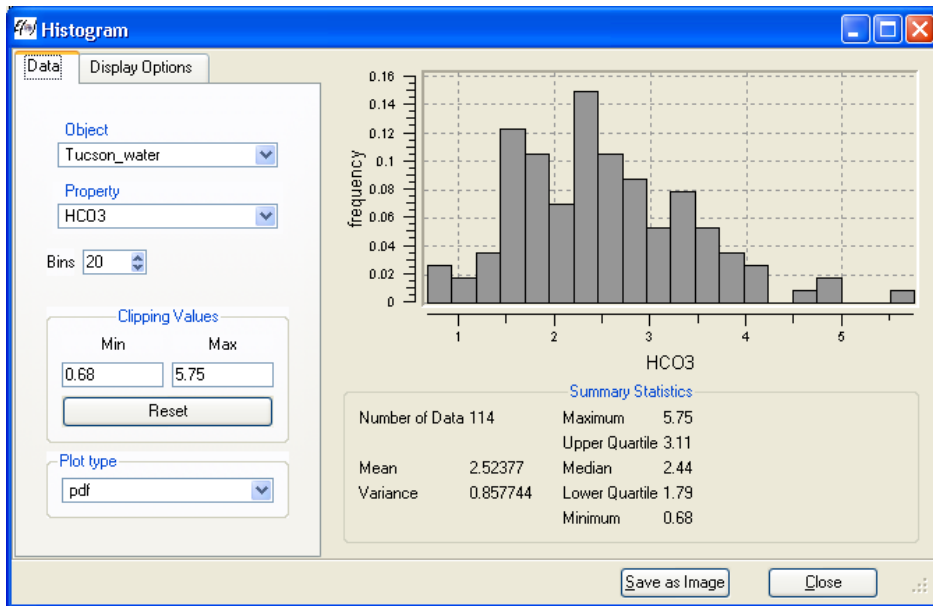


Рис. 7.3. Інструмент *Histogram*

Для побудови графіка емпіричної варіограми за даними замірів вмісту кислих солей вугільної кислоти виберіть пункт *Variogram* меню *Data Analysis*. У діалоговому вікні *Variogram Modeler* в розділі *Select Task* залиште опцію, зазначену за замовчуванням (*compute variogram from scratch*), у списку *Grid Name* виберіть *Tucson_water*, в *стисках Head and Tail Properties* – HCO_3 . Натисніть *Next* ...

У наступному діалоговому вікні задайте відповідні параметри для побудови однієї всенаправленої варіограми (з кутовим допуском 90 градусів і максимально широкою смугою пошуку опорних точок) і чотирьох варіограм за напрямками (0 , 45 , 90 , 135 градусів з кутовим допуском в $22,5$ градуса і смугою пошуку опорних точок шириною 100°). Побудова варіограми проводиться для 60 лагів. За замовчуванням величина лага (*lag spacing*) повинна становити $0,1^\circ$ з допуском $0,05^\circ$ (*lag tolerance*). Завдання таких значень параметрів дозволить врахувати розкид значень в свердловинах, розташованих на відстані $5,7^\circ$ (для порівняння: максимальна відстань між свердловинами по широті становить 6°) (рис. 7.4).

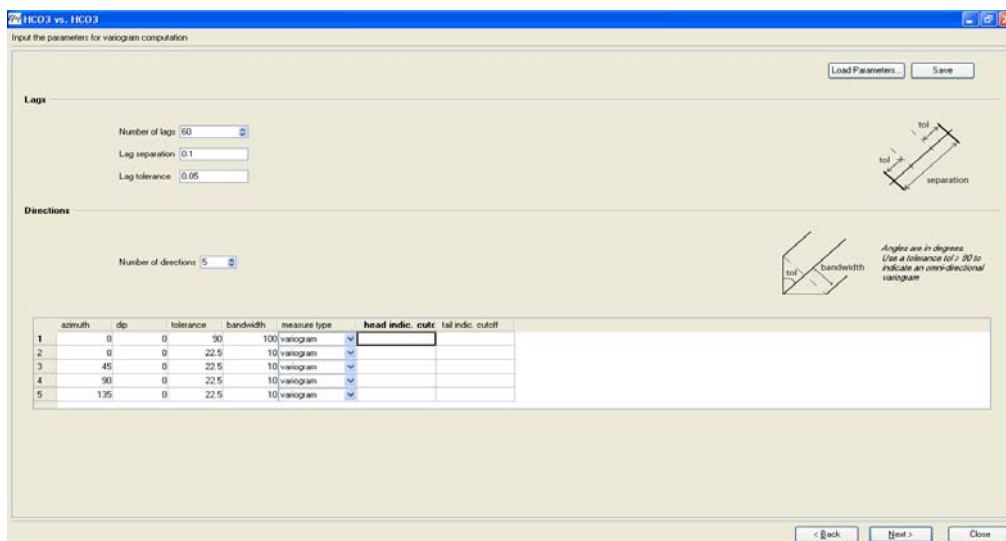


Рис. 7.4. Визначення параметрів побудови варіограми

У наступному діалоговому вікні виберіть меню *Window-Tile* для упорядкованого перегляду отриманих результатів (рис. 7.5).

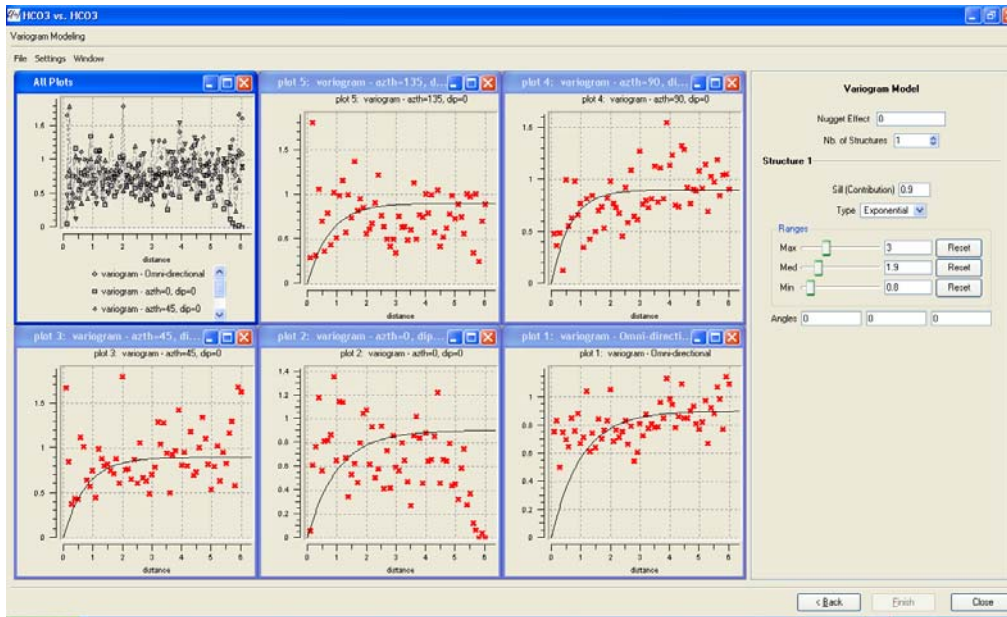


Рис. 7.5. Упорядкований перегляд варіограми за напрямками

Виберіть експоненційну модель варіограми. В якості плато варіограми вкажіть 0,8 (дисперсія проб). Здійсніть підбір моделі варіограми.

Збережіть на диску отриману модель.

3. Кригінг

Перед виконанням інтерполяції необхідно створити сітку, осередки якої в подальшому будуть містити отримані в результаті виконання процедури кригінга оцінки просторової змінної. Перейдіть на закладку *Info* і виберіть об'єкт *Tucson_water*. Множину даних об'єкта поміщено в прямокутник (*bounding box*) з параметрами мінімум: (0,31 і 0,8) і максимум: (11,1 і +6,5) в градусах широти і довготи. Для виконання інтерполяції необхідно створити сітку розміром 60 × 40 осередків. Задайте граничні значенні довготи (0,3 і 11,1) і широти (0,8 і 6,5). Розмір осередку сітки: 0,2°. Для цього виберіть меню *Objects – New Cartesian Grid* (рис. 7.6).

Створена сітка буде додана до списку об'єктів, доступних для візуалізації. Для виконання кригінга відобразіть панель *Algorithms (View-Algorithms)* і в розділі *Estimation* виберіть *kriging*. При цьому відобразиться прихована частина панелі *Algorithms*. Заповніть вкладки *General Data* і *Variogram*, як показано на рис. 7.7. Таким чином, оцінка вмісту кислих солей вугільної кислоти у вузлах сітки буде здійснюватися на основі ординарного кригінга з використанням еліпсоїда пошуку радіусом 10 градусів. Для оцінки невідомих значень будуть використовуватися значення в мінімум 2 і максимум 12 опорних точок в межах радіусу сусідства. Як параметри моделі варіограми (вкладка *Variogram*) використовуйте збережені дані моделі, створеної в пункті 2 (кнопка *Load existing model ...*). Натисніть кнопку *Run Algorithm*.

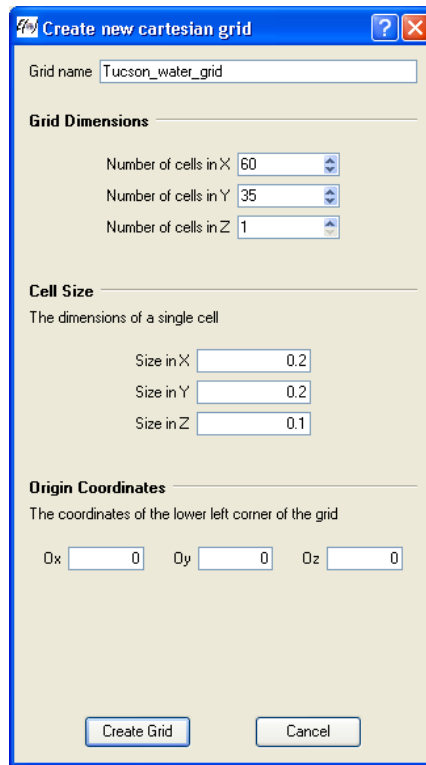


Рис. 7.6. Параметри створення сітки інтерполяції

Натисніть *Create Grid*.

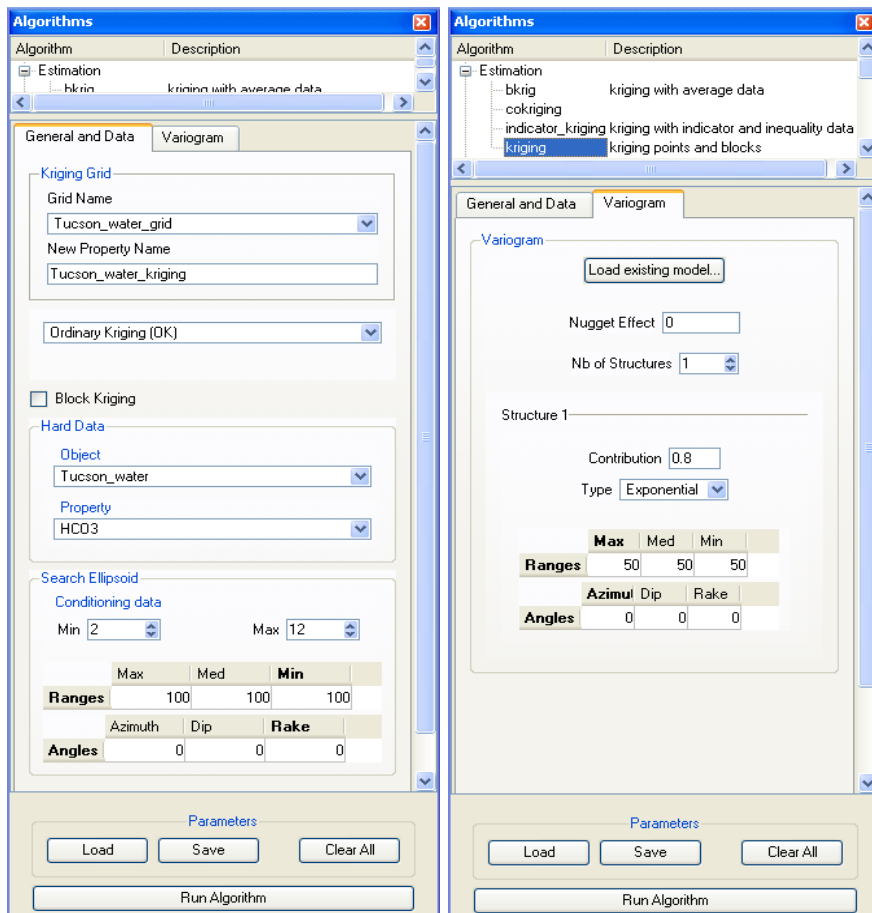


Рис. 7.7. Вкладки *General Data* і *Variogram* панелі *Algorithms*

4. Послідовне гаусівське стохастичне моделювання

В рамках даного розділу необхідно згенерувати п'ять моделей вмісту кислих солей вугільної кислоти з використанням послідовного гаусівського моделювання. Для цього перейдіть до розділу *Simulation – Sequential Gaussian Simulation* на панелі *Algorithms* і на вкладці *General* вкажіть необхідні параметри (рис. 7.8).

Число випадкових реалізацій дорівнює п'яти (*Nb of realizations*). Значення параметра *Seed* залишіть рівним заданому за замовчуванням. При виконанні послідовного гаусівського моделювання на попередньому етапі проводиться нормування значень даних в опорних точках для приведення їх до нормального розподілу (з нульовим середнім і одиничним стандартним відхиленням). Згенеровані моделі, таким чином, також будуть розподілені за нормальним законом. Під час виконання моделювання використовується простий крігінг у припущенні, що середнє значення досліджуваної величини (HCO_3) постійне і дорівнює нулю.

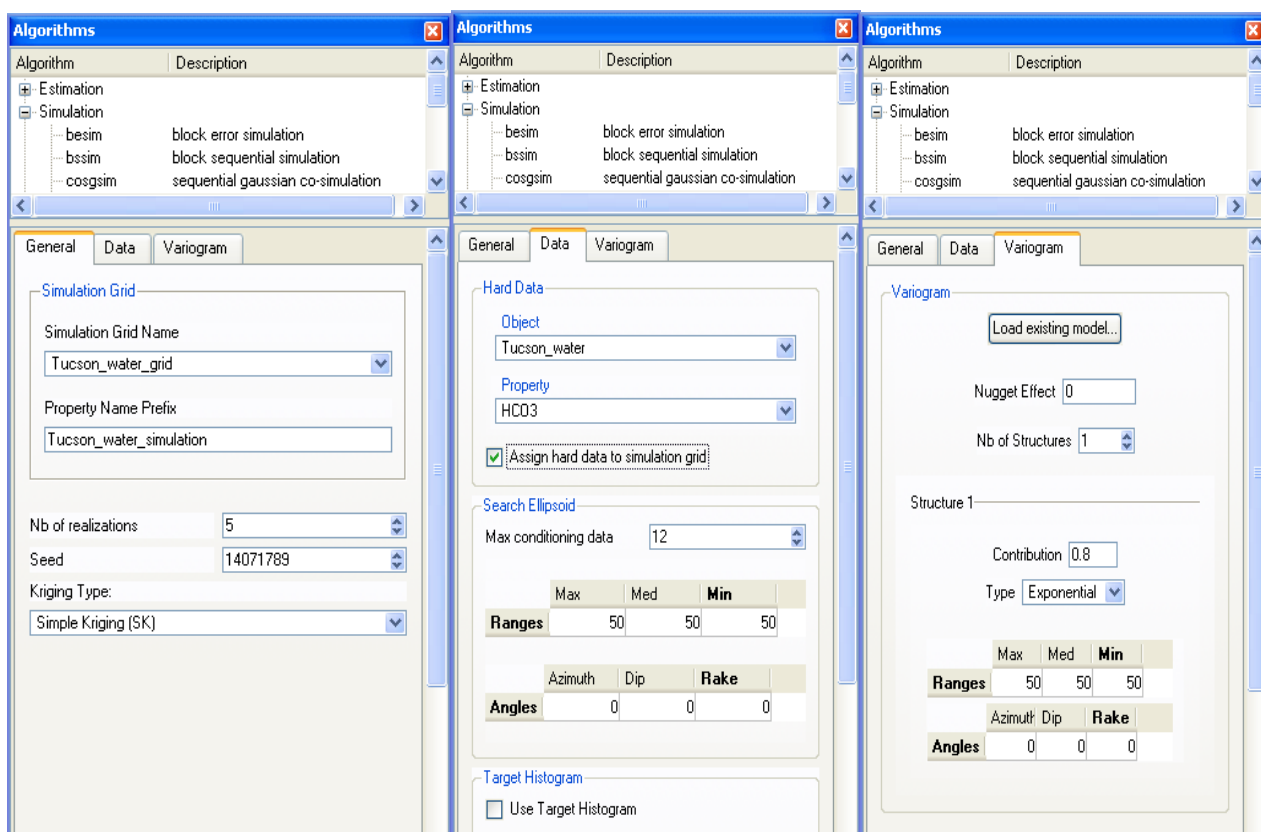


Рис. 7.8. Параметри виконання гаусівського стохастичного моделювання

На вкладці *Data* виберіть опцію *Use Target Histogram* і вкажіть розширені параметри (рис. 7.9).

Вибір опції *Use Target Histogram* дозволяє привести гістограму змодельованих даних HCO_3 у відповідність з гістограмою даних замірів в опорних точках. Таке перетворення передбачає виконання екстраполяції за межами розподілу, заданими користувачем в розділі *Lower Tail Extrapolation* і *Upper Tail Extrapolation*, на основі гістограми значень модельованої випадкової

величини. В нашому випадку граничні значення становлять: 0,68 і 5,75 – мінімальне і максимальне значення даних замірів у опорних точках.



Рис. 7.9. Параметри побудови стохастичних моделей

Після натискання кнопки *Run Algorithm* буде згенеровано п'ять реалізацій HCO_3 , доступних для перегляду на вкладці *Objects* у розділі *Tucson_water_grid*.

Постановка завдання

Для набору даних по озону (*tucwate.dat*) виконати:

- попередній аналіз даних (побудову гістограм і варіограм);
- інтерполяцію з використанням методу крігінг;
- послідовне гаусівське стохастичне моделювання.

Оформити звіт, зробити змістовні висновки.

Варіанти

| Номер комп'ютера | 1, 8 | 2, 9 | 3, 10 | 4, 11 | 5, 12 |
|------------------|---------|-----------|--------|-----------|-----------------|
| Показник | Calcium | Magnesium | Sodium | Potassium | SO ₄ |

Питання для підготовки до захисту лабораторної роботи

1. У чому полягає схожість та відмінності програмного забезпечення *S-GeMS* та модуля *Geostatistical Analyst* пакету *ArcGIS*?
2. На чому полягає застосування вікна *Commands Panel* програмного забезпечення *S-GeMS*?
3. Які існують методи попереднього аналізу даних?
4. Наведіть вихідні параметри інтерполяції за методом крігінг?
5. Якими є параметри побудови стохастичних моделей?

ПЕРЕЛІК ПОСИЛАНЬ

1. Johnston K. ArcGIS 9 Geostatistical Analyst / K. Johnston, J.M. Ver Hoef, K. Krivoruchko, N. Lucas // ESRI, 2003. – 306 с.
2. Армстронг М. Основи лінійної геостатистики / М. Армстронг // Фонтенбло: Центр геостатистики, 1998. – 149 с.
3. Давід М. Геостатистичні методи при оцінці запасів руд / М. Давід // Л.: "Недра", 1980. – 360 с.
4. Девіс Дж. Статистичний аналіз даних в геології / Дж. Девіс // СПб: Недра, 1990. – 319 с.
5. Дем'янов В.В. Геостатистики, теорія і практика / В.В. Дем'янов, О.О. Савельєва // М.: Наука, 2010. – 327 с.
6. Дюбрул О. Використання геостатистики для включення в геологічну модель сейсмічних даних / О. Дюбрул // European Association of Geoscientists and Engineers, 2002. – 295 с.
7. Дюбрул О. Геостатистика в нафтовій геології / О. Дюбрул // Москва-Іжевськ: Інститут комп'ютерних досліджень, НДЦ «Регулярна і хаотична динаміка», 2009 – 256 с.
8. Капутін Ю.Е. Гірничі комп'ютерні технології та геостатистика / Ю.Е. Капутін // СПб: Недра, 2002. – 424 с.
9. Ковалевський Є.В. Геологічне моделювання на основі геостатистики / Є.В. Ковалевський // European Association of Geoscientists and Engineers, 2012. – 117 с.
10. Матерон Ж. Основи прикладної геостатистики / Ж. Матерон // М.: Мир, 1968. – 407 с.

СЛОВНИК ТЕРМІНІВ З ГЕОСТАТИСТИКИ

Автокореляція – міра взаємної залежності між елементами єдиної послідовності спостережень, зсунутих на k одиниць (автокореляція k -го порядку).

Анізотропія – характеристика випадкової функції, яка визначає залежність функції варіограми від напрямку взаємного розташування опорних точок.

Варіограма – відображає характер варіації даних в парах опорних точок як функцію відстані між ними.

Варіографія – підбір просторово залежної моделі для даних.

Вибірка – кінцевий набір значень випадкової величини, отриманий в результаті спостережень за обмежений період.

Випадкова функція – функція, яка приймає невідомий вигляд в результаті випробувань.

Випадкове поле – випадкова функція, аргументом якої є координати простору.

Випадковий процес – випадкова функція, аргументом якої є час.

Випадковою називається подія, яка може або наступити при реалізації даного комплексу умов, або не наступити. Достовірні і неможливі події можуть розглядатися як крайні окремі приклади випадкових подій.

Внутрішня гіпотеза стаціонарності допускає, що прирощення просторової функції слабо стаціонарні, тобто середнє і дисперсія прирощень існують і не залежать від положення.

Генеральна сукупність – сукупність усіх можливих значень величини.

Геостатистика – наука, що займається вивченням і оцінювання просторових змінних за допомогою апарату теорії випадкових функцій.

Диференціальна функція розподілу (функція щільності розподілу ймовірності) характеризує ймовірність попадання вибіркового значення випадкової величини в заданий інтервал.

Достовірною називається подія, яка настає щоразу при реалізації даного комплексу умов.

Дрифт – характеристика випадкової функції – очікуване значення випадкової функції, що є константою (у випадку, якщо випадкова функція є стаціонарною, характеризується стаціонарністю другого порядку або задовольняє внутрішній гіпотезі стаціонарності) або залежне від просторових координат.

Екстраполяція – пошук значення досліджуваної характеристики за межами заданого інтервалу спостережень.

Ергодична випадкова функція – функція, одна реалізація якої на великому інтервалі еквівалентна великому числу реалізацій на обмеженому інтервалі і кожна реалізація якою характеризується приблизно однаковим середнім значенням.

Ефект основи в геостатистиці характеризує вплив розмірів блоків та обсягів проб на результати оцінювання невідомих значень просторової змінної (менші за розмірами блоки більш розсіяні, ніж великі).

Жорсткий інтерполятор – метод інтерполяції, який дає в опорній точці значення, рівне зміряному.

Закон розподілу – співвідношення, що встановлює зв'язок між можливими значеннями випадкової величини і відповідними їм ймовірностями.

Зона впливу – максимальна відстань, на якому між пробами ще існує кореляція.

Інтегральна функція розподілу виражає ймовірність того, що вибіркоче значення випадкової величини виявиться меншим за деяке граничне.

Інтерполяція – пошук значення досліджуваної характеристики між точками спостережень.

Кореляційний аналіз – припускає вивчення залежності між випадковими величинами з одночасною кількісною оцінкою ступеня не випадковості їх спільної зміни.

Крігінг – метод локальної оцінки значень просторової змінної як зваженої суми значень сусідніх опорних точок, враховує просторове розташування опорних точок навколо шуканої і присвоює значенням в опорних точках такі вагові коефіцієнти, які мінімізували б помилку оцінювання.

Незміщена оцінка – це точкова оцінка, математичне сподівання якої дорівнює оцінюваному параметру.

Неможливою називається подія, яка ніколи не настає при реалізації даного комплексу умов.

Перевірка – процедура порівняння моделей варіограми. Перевірка спочатку виключає з обчислень частину даних – "тестовий набір даних" – а потім використовує залишилися дані – "навчальний набір даних" – для створення моделей тренду й автокореляції, які будуть використані в інтерполяції.

Перехресна перевірка – процедура порівняння моделей варіограми. Перехресна перевірка використовує всі дані для оцінки моделі. Потім вона послідовно, по одній, видаляє опорну точку і інтерполірує для цього місцеположення відповідне значення. Обчислене та дійсне значення в опорній точці порівнюються. Така процедура повторюється для всіх точок.

Поріг варіограми – величина дисперсії проб. При досягненні порогу варіограма, як правило, вирівнюється, тобто більше не зростає.

Просторова змінна – змінна, що є функцією координат.

Радіус впливу варіограми – відстань, при якій модель функції варіограми починає вирівнюватися. Опорні точки, віддалені одна від одної на відстань, меншу, ніж радіус впливу, просторово корелюють, в той час як точки, віддалені одна від одної на відстань, більшу, ніж радіус впливу, ні.

Статистичною називається залежність між випадковими величинами, при якій зміна однієї з величин спричиняє зміна закону розподілу іншої величини.

Стаціонарна просторова змінна – просторова змінна, розподіл якої інваріантний до трансляції в просторі (тобто не залежить від місця розташування).

Стаціонарність другого порядку ("слабка" стаціонарність) – інваріантність до трансляції в просторі перших двох моментів розподілу просторової змінної (середнього і дисперсії (коваріації, варіограми)).

Умовна ймовірність – ймовірність однієї події за умови, що інша подія вже відбулося.

Умовний розподіл – розподіл випадкової величини при умові, що інша випадкова величина приймає певне значення.

Частковий поріг варіограми – значення порога, від якого віднято значення самородка.

ВИМОГИ ДО ОФОРМЛЕННЯ ЗВІТУ З ЛАБОРАТОРНОЇ РОБОТИ

Звіт оформлюється на аркушах формату А4 і повинен мати:

- титульну сторінку з назвою лабораторної роботи, прізвищем студента, номером групи та номером варіанта для робіт, в яких завдання по варіантах;
- мету роботи;
- постановку завдання лабораторної роботи;
- код програми з докладними коментарями (для робіт № 1-3);
- скріншоти основних результатів виконання кожного пункту завдання;
- висновки по роботі.

Звіт оформлюється відповідно до стандарту ДСТУ 3008 – 95. Документація. Звіти у сфері науки і техніки. Структура і правила оформлення.

КРИТЕРІЇ ОЦІНЮВАННЯ ЛАБОРАТОРНИХ РОБІТ

Максимальну оцінку виставляють за дотримання таких умов:

- повна відповідність звіту про виконання лабораторної роботи методичним рекомендаціям;
- володіння теоретичними відомостями, на яких базується предмет досліджень;
- загальна та професійна грамотність, лаконізм і логічна послідовність викладу матеріалу;
- відповідність оформлення звіту чинним стандартам (ДСТУ 3008 – 95. Документація. Звіти у сфері науки і техніки. Структура і правила оформлення) та наведеним вимогам.

Сергєєва Катерина Леонїдївна

ГЕОСТАТИСТИКА
МЕТОДИЧНІ РЕКОМЕНДАЦІЇ
ДО ЛАБОРАТОРНИХ ЗАНЯТЬ З ДИСЦИПЛІНИ

для студентів спеціальностей
122 Комп'ютерні науки та інформаційні технології
(спеціалізація “Комп'ютерний еколого-економічний моніторинг”),
193 Геодезія та землеустрій
(спеціалізація “Геоінформаційні системи і технології”)

Видано в редакції автора

Підписано до друку 20.10.2016. Формат 30×42/4.
Папір офсет. Ризографія. Ум. друк. арк. 3,5.
Обл.-вид. арк. 3,5. Тираж 20 пр. Зам. № ____

Державний вищий навчальний заклад
«Національний гірничий університет».
49005, м. Дніпро, просп. Д. Яворницького, 19