УДК 004.9+631

Б. М. Опанасюк
В. Б. Мокін

РОЗРОБКА ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ ГЕОСТАТИСТИИЧНОГО АНАЛІЗУ РАДІАЦІЙНОГО ЗАБРУДНЕННЯ В ЄВРОПІ

Вінницький національний технічний університет

Анотація

Проведено геостатистичний аналіз радіаційного забруднення в країнах Європи після аварії на Чорнобильській АЕС. Для вирішення поставленого завдання використано технологію SciKit-GStat та визначено оптимальні за різними критеріями моделі інтерполяції даних, використання яких дозволило створити інформаційну технологію на основі геостатистичних методів та підвищити ефективність аналізу.

Ключові слова: *моніторинг радіаційного забруднення, інтерполяція, геостатистичний аналіз, моніторинг, аналіз даних*

Abstract

A geostatistical analysis of radiation contamination in European countries after the Chernobyl accident was conducted. To solve the task, SciKit-GStat technology was used and the optimal data interpolation models were determined by different criteria, the use of which allowed creating information technology based on geostatistical methods and increasing the efficiency of the analysis.

Keywords: *radiation pollution monitoring, interpolation, geostatistical analysis, monitoring, data analysis.*

Вступ

Моніторинг радіаційного забруднення став критично важливим завданням після великих ядерних катастроф, таких як аварія на Чорнобильській атомній електростанції (1986 рік) та аварія на Фукусімі (2011 рік). У перші роки після Чорнобиля застосовувалися переважно ручні методи збирання даних про радіаційне забруднення, що вимагало значних людських і матеріальних ресурсів. Поступово, з розвитком інформаційних технологій (ІТ) цей процес автоматизувався і набув нових можливостей, що дозволило більш точно і швидко аналізувати дані та прогнозувати наслідки забруднень.

Використання сучасних методів аналізу даних, таких як штучні нейронні мережі та факторний аналіз, разом з інструментами геостатистичного аналізу та ГІС, дозволяє ефективно вирішувати завдання та досягати поставлених цілей у дослідженні радіаційного забруднення [1].

Метою дослідження є підвищення ефективності аналізу радіаційного забруднення в Європі шляхом створення інформаційної технології на основі геостатистичних методів.

Підготовка даних

Дослідження проводилось на основі даних з датасету Kaggle «Chemobyl Chemical Radiation / CSV / Country Data» (https://www.kaggle.com/datasets/brsdincer/chernobyl-chemical-radiation-csv-country-data). Було проведено геостатистичний аналіз, на основі якого розроблено інформаційну технологію, яка дозволяє покращити точність та ефективність моніторингу, прогнозування та управління радіаційними ризиками в Європі.

Зазначений вище датасет містить дані про рівень радіаційного фону радіонуклідів Йод-131, Цезій-134 та Цезій-137, що вимірювались на території сучасної Європи після катастрофи на Чорнобильській АЕС. Було обмежено множину ознак (рисунок 1) та вибрано прямокутну ділянку Європи, де – найбільша щільність станцій спостереження. Для цього була здійснена картографічна візуалізація даних за допомогою бібліотеки Folium (рисунок 2).

```
INFO:

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2051 entries, 0 to 2050
Data columns (total 7 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   PAYS         2051 non-null   object
 1   X            2051 non-null   float64
 2   Y            2051 non-null   float64
 3   Date         2051 non-null   object
 4   Iodine_131   2009 non-null   float64
 5   Caesium_134  1801 non-null   float64
 6   Caesium_137  1506 non-null   float64
dtypes: float64(5), object(2)
memory usage: 112.3+ KB
None
```

Рисунок 1 – Загальний вигляд оптимізованих даних

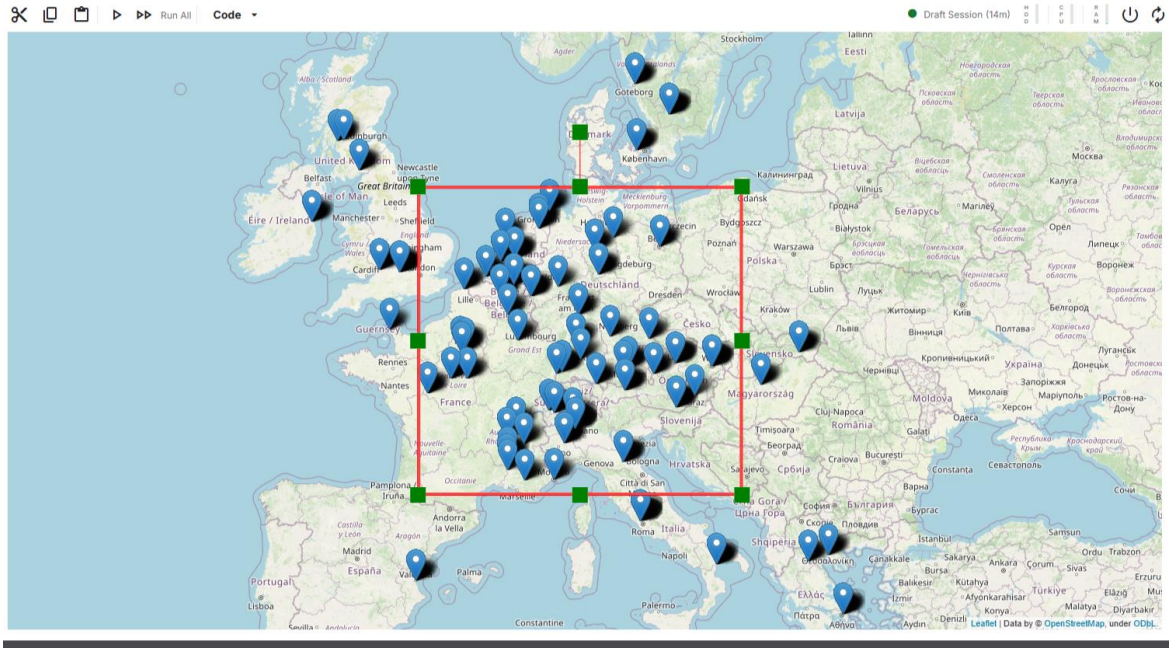
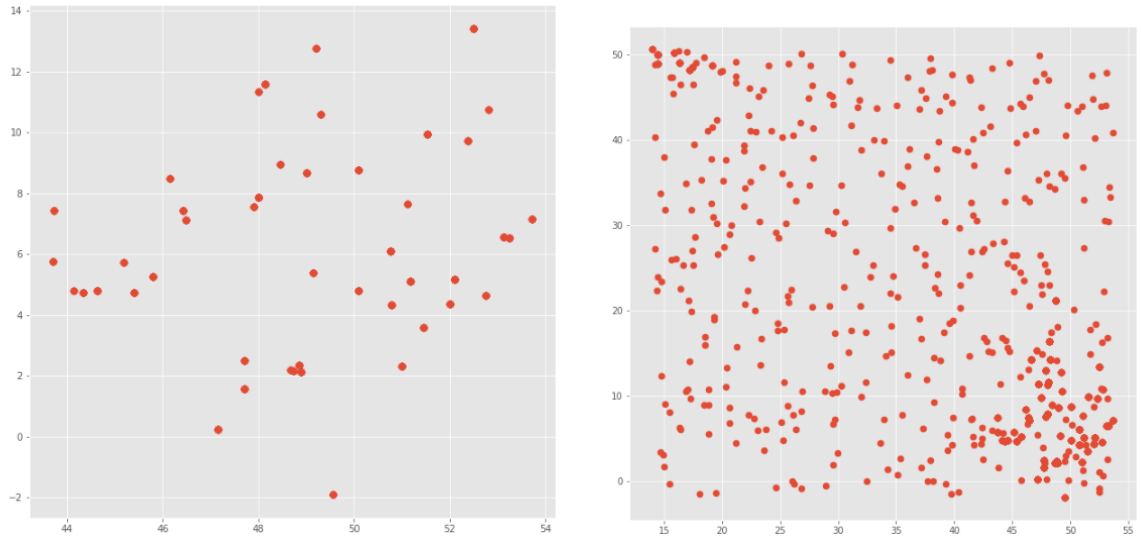


Рисунок 2 – Картографічне відображення вхідних даних

Виконавши базовий аналіз даних, приймаємо рішення що для здійснення геостатистичного аналізу нам необхідно відфільтрувати дані та використовувати лише дані країн, які мають найбільшу щільність станцій спостереження радіаційного фону. Відповідно, приймаємо рішення виконувати аналіз, враховуючи лише дані таких країн як Бельгія, Швейцарія, Німеччина, Франція, Нідерланди, Австрія та Чехія для забезпечення найкращої точності аналізу.

Часто у великих наборах даних зустрічаються записи з відсутніми значеннями. У контексті радіаційного моніторингу такі пропуски можуть бути викликані відсутністю показників із певних станцій або технічними проблемами під час передавання даних. Наше дослідження, на жаль не стало винятком, і саме тому було прийнято рішення використати інтерполяцію даних, що дозволить відновити відсутні значення на основі доступних даних [2].

Генеруємо діаграму розсіювання даних (рисунок 3, а) та помічаємо, що дані, розташовані достатньо щільно, проте, внаслідок географічних особливостей, збирання даних у певних ділянках та формах рельєфу не є можливим. Для корекції даних приймаємо рішення здійснити кластеризацію та передбачення даних за допомогою алгоритму K-Nearest Neighbor.



а) б)

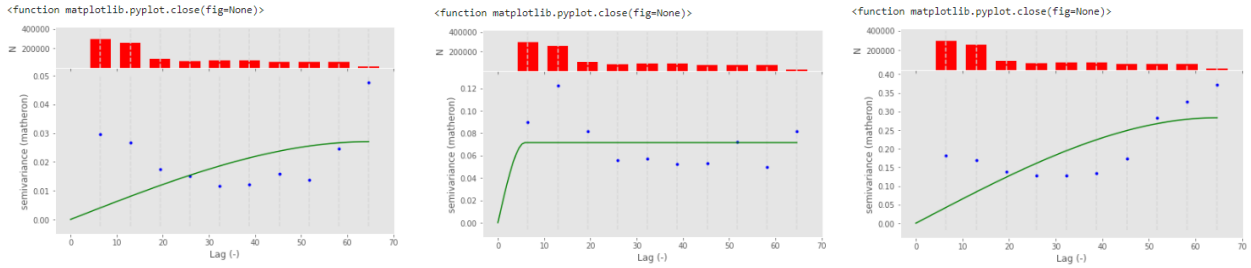
Рисунок 3 – Діаграма розсіювання даних: до (а) та після (б) передбачення допоміжних даних

Виконавши кластеризацію, розпочинаємо генерування даних, передбачених за допомогою K-nearest Neighbor алгоритму. Встановлюємо оптимальне значення 3-х сусідів для кожного із вже існуючих записів і виконуємо передбачення. Отримуємо показник точності передбачення у 96,5%, що свідчить про високий ступінь схожості даних до реальних замірів (рисунок 3,б).

Результати машинного навчання можна використовувати для побудови довгострокових прогнозів, що дозволяє визначити потенційні небезпеки, а також розробити плани щодо мінімізації впливу радіації на довкілля.

Геостатистичний аналіз

Геостатистичний аналіз є важливою частиною дослідження, оскільки він дозволяє виявити просторові закономірності та аномалії в даних про радіаційне забруднення. У нашому дослідженні використовуються інструменти бібліотеки SciKit-GStat, яка пропонує широкий набір функцій для побудови варіограм і моделювання просторових залежностей. На основі аналізу даних з отриманої діаграми розсіювання будуємо сферичні варіограми для основних досліджуваних ознак - Йод-131 (рисунок 4,а), Цезій-134 (рисунок 4,б) та Цезій-137 (рисунок 4,в).



а) б) в)

Рисунок 5 – Сферична варіограма по різних показнику: а) Йод-131; б) Цезій-134; в) Цезій-137

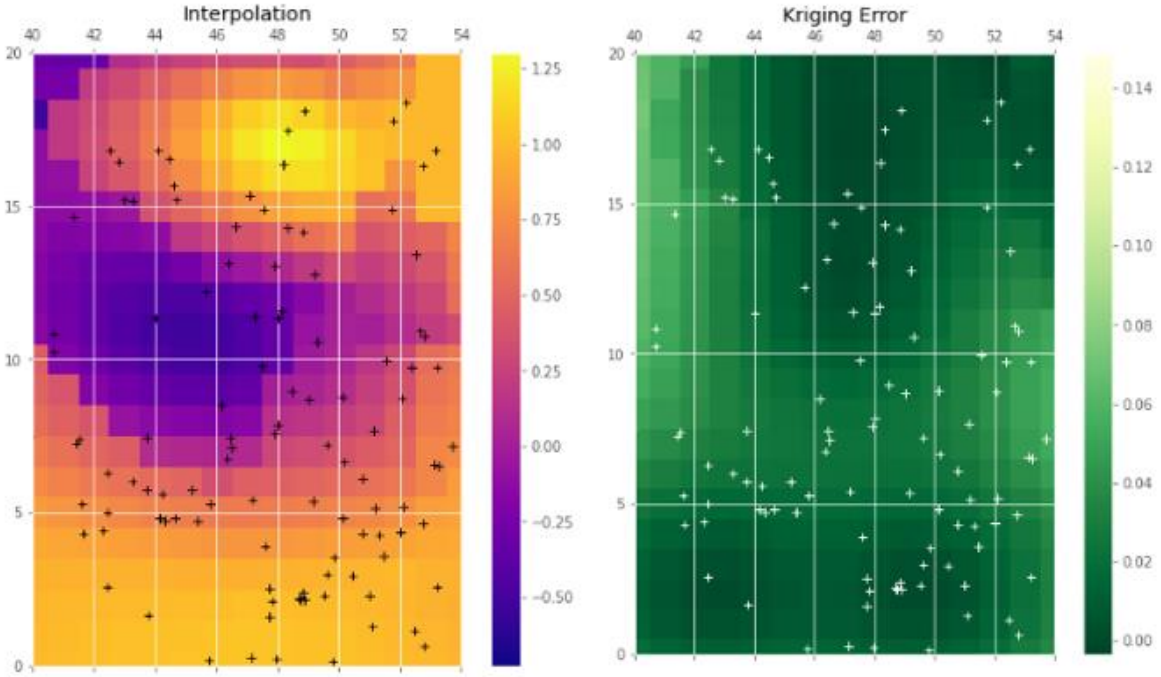
На даних рисунках можливо побачити наскільки далеко чи близько один від одного знаходяться станції відносно своєї широти.

На основі побудованих варіограм вимальовується модель просторових залежностей, яка дозволяє зробити більш точне визначення рівня радіації у невідомих точках.

Візуалізація даних та результати

Використовуючи дані, підготовані у попередніх кроках, будуємо графіки інтерполяції (рис. 5,а) та визначаємо помилки кригінгу (рис. 5,б), показники спостереження відображені у Bq/m³. З першого графіку (рис. 5,а) можна помітити області з найбільшим та найменшим рівнем радіаційного

забруднення. Ці дані можна використовувати, наприклад, для побудови маршрутів евакуації при потенційній техногенній катастрофі або ж для застосування термінових зусиль з деактивації ізотопів. Другий же графік (рис. 5,б) відображає невизначеність прогнозу зробленого за допомогою кригінг-методу геостатистичного інтерполювання та може бути застосований для розміщення додаткових засобів спостереження показників у місцях з найбільшим відхиленням.



а) б)

Рисунок 5 – Графіки помилок методів Interpolation (а) та Kriging (б)

Важливо, що для підвищення ефективності аналізу необхідно задавати значення не випадковим чином, а – на основі певної математичної функції, додаючи шум, щоб метод міг знаходити чинні закономірності [3]. Робимо геостатистичний аналіз на відповідних даних (рис. 6).

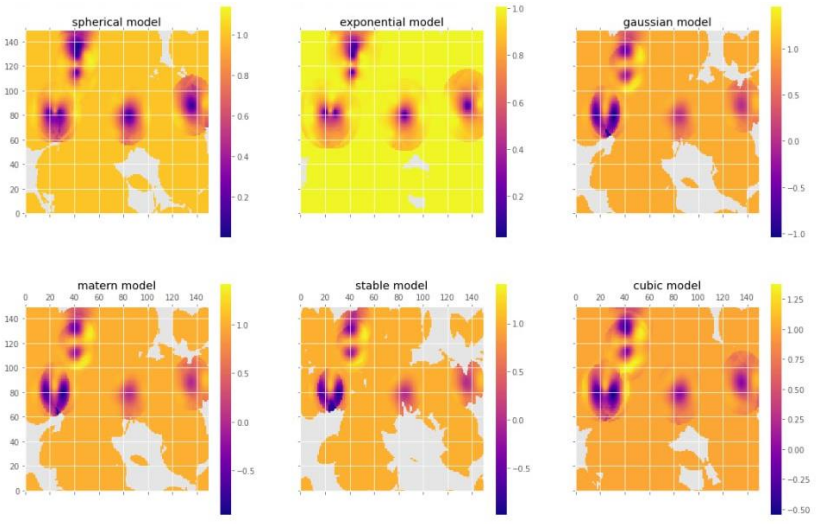


Рисунок 6 – Результат геостатистичного аналізу основними популярними моделями

Висновки

З графіків можна чітко прослідкувати рівень спостереження трьох основних нуклеотидів та рівні їх розповсюдження, що здійснювались протягом року після аварії на ЧАЕС. Також вдалось помітити

певну закономірність: практично всюди забруднення розсіялось рівномірно, за винятком міст та агломерацій, що є сильно озелененими та мають дещо нижчий рівень урбанізації (наприклад, Берлін, Гамбург, Париж, Амстердам тощо) – це й помітно на графіках специфічними потемніннями. «Чисті» зони можна з легкістю помітити й на відомих мапах, наданих джерелами Об'єднаного дослідного центру ЄС. Тому можна зробити висновок про те, що геостатистичний аналіз виконаний з винятковою точністю.

Геостатистичний аналіз показав, що найточніші показники можна помітити якраз у країнах, де станцій найбільше та розташовані вони найщільніше - це Німеччина, Австрія, Бельгія та Нідерланди. Найкраще ситуацію відображають «matern» та «stable» моделі.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Neteler, M., & Mitasova, H. (2002). Open source GIS: a GRASS GIS approach (Vol. 689). *Springer Science & Business Media*, 2013. – 435 p.

2. Goovaerts P. Kriging Interpolation. The Geographic Information Science & Technology Body of Knowledge. 2019. URL: https://gistbok.ucgis.org/bok-topics/kriging-interpolation (дата звернення: 05.12.2024).

3. Мокін В. Б. Удосконалення методу ординарного кригінгу геостатистичного аналізу для моделювання якості вод річок з урахуванням їх звивистості / В. Б. Мокін, Є. М. Крижановський, Ю. С. Семчук // Вісник Вінницького політехнічного інституту. — Вінниця : — 2011. — № 4. — С. 46–52.

Опанасюк Богдан Михайлович — студент групи 1ІСТ-23м, Вінницький національний технічний університет, Вінниця, 01-23-047.stud@vntu.edu.ua

Мокін Віталій Борисович – д-р техн. наук, проф., завідувач кафедри системного аналізу та інформаційних технологій, Вінницький національний технічний університет, Вінниця, e-mail: vbmokin@vntu.edu.ua

Opanasiuk Bohdan – student of group 1IST-23m, Vinnytsia National Technical University, Vinnytsia, 01-23-047.stud@vntu.edu.ua

Mokin Vitalii – Dr. Tech. Sciences, Prof., Head of the Department of System Analysis and Information Technologies, Vinnytsia National Technical University, Vinnytsia, e-mail: vbmokin@vntu.edu.ua