

# ЧИСЕЛЬНА АПРОКСИМАЦІЯ ТА ПРОГРАМНА РЕАЛІЗАЦІЯ ПОШУКУ АНОМАЛІЙ В ЕКОЛОГІЧНИХ ДАНИХ МОВОЮ PYTHON

Державний університет «Житомирська політехніка»

## Анотація

У роботі досліджено застосування методу найменших квадратів для лінійної апроксимації екологічних часових рядів та його програмну реалізацію мовою Python. Особливу увагу приділено алгоритмізації процесу виявлення аномальних значень експозиції забруднюючих речовин на основі математичного критерію відхилення. Розроблена програма забезпечує обчислення параметрів тренду та автоматизовану ідентифікацію нетипових викидів.

**Ключові слова:** метод найменших квадратів, лінійна апроксимація, виявлення аномалій, екологічні дані, Python.

## Abstract

The paper investigates the application of the least squares method for the linear approximation of ecological time series and its software implementation in Python. Special attention is given to the algorithmization of the process of detecting anomalous values of pollutant exposure based on a mathematical deviation criterion. The developed program provides the calculation of trend parameters and the automated identification of atypical emissions.

**Keywords:** least squares method, linear approximation, anomaly detection, ecological data, Python.

## Вступ

Сьогодні моніторинг обсягів викидів шкідливих речовин в атмосферу є одним із ключових завдань для оцінки екологічної безпеки [1]. Для аналізу таких часових рядів і виявлення нетипових відхилень ефективно застосовуються чисельні методи лінійної апроксимації, зокрема метод найменших квадратів [2].

Метою роботи є розроблення та програмна реалізація мовою Python алгоритму апроксимації екологічних даних для виявлення аномальних значень викидів.

## Результати дослідження

Для побудови математичної моделі використано відкриті дані Світового банку щодо середньорічної експозиції дрібних твердих частинок PM<sub>2.5</sub> (мікрограми на кубічний метр) в Україні за період з 1990 по 2020 рік (Таблиця 1).

Таблиця 1. Динаміка середньорічної експозиції PM<sub>2.5</sub> в Україні (1990–2020 рр.) за даними Світового банку [3]

Рік (x)	PM <sub>2.5</sub> , мкг/м <sup>3</sup> (y)	Рік (x)	PM <sub>2.5</sub> , мкг/м <sup>3</sup> (y)
1990	33.06	2006	21.38
1991	32.76	2007	20.86
1992	32.40	2008	20.42
1993	31.99	2009	20.04
1994	31.53	2010	19.71
1995	31.03	2011	19.44

1996	30.47	2012	18.68
1997	29.82	2013	18.88
1998	29.10	2014	18.43
1999	28.34	2015	17.79
2000	27.55	2016	16.24
2001	26.57	2017	15.66
2002	25.32	2018	17.10
2003	24.02	2019	15.58
2004	22.84	2020	14.90
2005	21.97	-	-

Для виявлення довгострокового тренду та ідентифікації аномальних відхилень застосовано лінійну апроксимацію методом найменших квадратів (МНК). Рівняння тренду шукається у вигляді

$$y = ax + b \quad (1)$$

де  $y$  – значення експозиції PM2.5,  $x$  – рік спостереження.

Згідно з МНК, невідомі параметри  $a$  та  $b$  визначаються з умови мінімізації суми квадратів відхилень реальних даних від теоретичної прямої:

$$S = \sum_{i=1}^n (y_i - (ax_i + b))^2 \rightarrow \min \quad (2)$$

Розв'язок відповідної системи рівнянь дозволяє обчислити шукані коефіцієнти за формулами:

$$a = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \quad (3)$$

$$b = \frac{\sum_{i=1}^n y_i - a \sum_{i=1}^n x_i}{n} \quad (4)$$

Для виявлення аномалій розроблено алгоритм мовою програмування Python з використанням бібліотеки NumPy [4]. Аномальним вважається рік, у якому модуль різниці між реальним та теоретичним значенням перевищує заданий поріг чутливості  $\Delta$ . Фрагмент програмної реалізації наведено на рисунку 1.

```
# Дані спостережень (роки та значення PM2.5)
x = np.arange(1990, 2021)
y = np.array([33.06, 32.76, 32.40, 31.99, 31.53, 31.03, 30.47, 29.82,
              29.10, 28.34, 27.55, 26.57, 25.32, 24.02, 22.84, 21.97,
              21.38, 20.86, 20.42, 20.04, 19.71, 19.44, 18.68, 18.88,
              18.43, 17.79, 16.24, 15.66, 17.10, 15.58, 14.90])

n = len(x)

# Обчислення коефіцієнтів тренду a та b за МНК
sumX = np.sum(x)
sumY = np.sum(y)
sumX2 = np.sum(x**2)
sumXY = np.sum(x * y)

a = (n * sumXY - sumX * sumY) / (n * sumX2 - sumX**2)
b = (sumY - a * sumX) / n

# Виявлення аномальних значень викидів
delta = 1.5 # поріг чутливості відхилення
for i in range(n):
    y_teor = a * x[i] + b
    if abs(y[i] - y_teor) > delta:
        print(f"Аномалія: {x[i]} рік. Відхилення: {abs(y[i] - y_teor):.2f} мкг/м³")
```

Рис. 1. Фрагмент програми

В результаті виконання програми було отримано рівняння лінійного тренду  $y = -0.65x + 1332.01$ . Візуалізацію фактичних даних, побудованої лінії тренду та ідентифікованих аномалій наведено на рисунку 2.

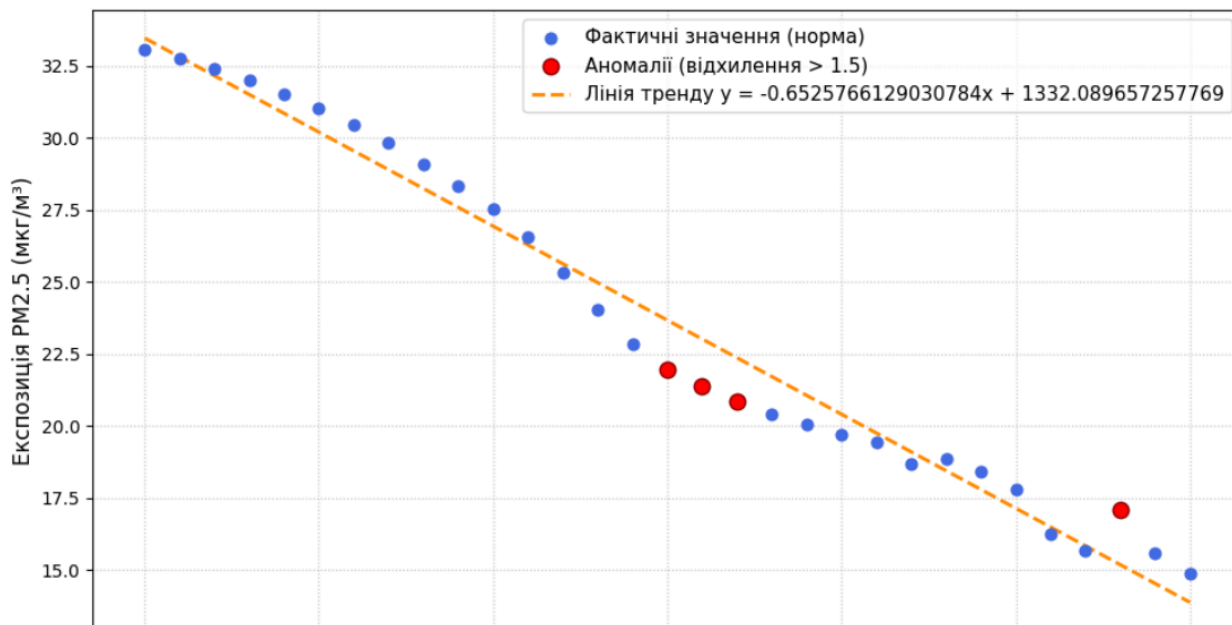


Рис. 2. Динаміка середньорічної експозиції PM2.5 в Україні та лінія тренду

Встановлений поріг чутливості  $\Delta = 1.5$  мкг/м<sup>3</sup> дозволив виявити найбільш суттєві відхилення показників, які зведені у таблицю 2.

Таблиця 2. Виявлені аномальні відхилення

Рік	Фактичне значення PM2.5, мкг/м <sup>3</sup>	Теоретичне значення тренду, мкг/м <sup>3</sup>	Модуль відхилення, мкг/м <sup>3</sup>
2005	21.97	23.67	1.70
2006	21.38	23.02	1.64
2007	20.86	22.37	1.51
2018	17.10	15.19	1.91

### Висновки

У роботі розроблено алгоритм та його програмну реалізацію мовою Python для аналізу екологічних часових рядів методом найменших квадратів. Застосування лінійної апроксимації до даних щодо експозиції PM2.5 в Україні дозволило встановити стійкий макротренд до зниження рівня забруднення протягом 1990–2020 років. Використання математичного критерію відхилення дало змогу автоматизовано ідентифікувати аномальні роки (зокрема, різке підвищення показників у 2018 році та період послідовних відхилень 2005–2007 років)

### СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Кириєнко П. Г., Варламов Є. М., Квасов В. Є., Лобов С. О. Організація моніторингу за станом атмосферного повітря у м. Харків. Екологічна безпека та природокористування. 2023. № 4 (48). С. 81–90.
2. Задачин В. М. Чисельні методи: навчальний посібник. – Харків : ХНУ, 2012. – 312 с.

3. World Bank: World Development Indicators. URL: <https://databank.worldbank.org/source/world-development-indicators> (дата звернення: 20.06.2026).
4. NumPy Documentation. URL: <https://numpy.org/doc/> (дата звернення: 20.06.2026).

**Шмідт Анатолій Євгенович** — аспірант факультету гірничої справи, природокористування та будівництва, Державний університет «Житомирська політехніка», м. Житомир, e-mail: [phde2251\\_shaye@student.ztu.edu.ua](mailto:phde2251_shaye@student.ztu.edu.ua)

Науковий керівник: **Герасимчук Людмила Олександрівна** — кандидат сільськогосподарських наук, доцент кафедри загальної екології, Державний університет «Житомирська політехніка», м. Житомир

**Shmidt Anatolii Ye.** — PhD student, Faculty of Mining, Environmental Management and Construction, Zhytomyr Polytechnic State University, Zhytomyr, email: [phde2251\\_shaye@student.ztu.edu.ua](mailto:phde2251_shaye@student.ztu.edu.ua)

Supervisor: **Herasymchuk Liudmyla O.** — Cand. Sc. (Agr.), Associate Professor of the Department of General Ecology, Zhytomyr Polytechnic State University, Zhytomyr