

Міністерство освіти і науки України
Національний технічний університет
«Дніпровська політехніка»

Факультет інформаційних технологій
(факультет)

Кафедра системного аналізу та управління
(повна назва)

ПОЯСНЮВАЛЬНА ЗАПИСКА
кваліфікаційної роботи ступеня бакалавра

Здобувача вищої освіти Богун Валерія Вадимівна
академічної групи 124-21-1
спеціальності 124 Системний аналіз
за освітньо-професійною програмою Системний аналіз

на тему: «Системний аналіз і комп'ютерне моделювання просторового розподілу населення в регіоні»

Керівники	Прізвище, ініціали	Оцінка за шкалою		Підпис
		рейтинговою	інституційною	
кваліфікаційної роботи	<i>к.т.н., доц. Коряшкіна Л. С.</i>			
розділів:	3			
Інформаційно- аналітичний	<i>к.т.н., доц. Коряшкіна Л. С.</i>			
Спеціальний розділ	<i>к.т.н., доц. Коряшкіна Л. С.</i>			
Рецензент				
Нормоконтролер	<i>к.ф.-м.н., доц. Хом'як Т.В.</i>			

Дніпро
2025

ЗАТВЕРДЖЕНО:

завідувач кафедри
Системного аналізу та управління
(повна назва)

_____ к.т.н., доц. Желдак Т.А.
(підпис) (прізвище, ініціали)

« _____ » _____ 20__ року

ЗАВДАННЯ

на кваліфікаційну роботу
ступеня бакалавра

здобувачу вищої освіти Богун В.В. академічної групи 124- 21-1
спеціальності: 124 Системний аналіз
за освітньо-професійною програмою Системний аналіз
на тему «Системний аналіз і комп'ютерне моделювання просторового розподілу населення в регіоні»

затверджену наказом ректора НТУ «Дніпровська політехніка» від _____
05.05.2025 р. №336-с

Розділ	Зміст	Терміни виконання
1. Інформаційно-аналітичний розділ	<i>Проаналізувати методи просторового аналізу, джерела даних, методи кластеризації та кореляції. Визначити завдання дослідження.</i>	10.01.2025 – 01.03.2025
2. Спеціальний розділ	<i>Розробити математичне та програмне забезпечення. Провести підготовку та обробку геоданих і супутникових знімків, формування узгоджених масивів даних. Побудувати математичну модель кореляції та обґрунтувати вибір інструментів.</i>	01.03.2025 – 01.05.2025
3. Експериментально-аналітичний розділ	<i>Провести візуалізацію вхідних даних. Виконати розрахунки кореляційних коефіцієнтів (Пірсона, Спірмена, Кендалла). Побудувати графіки для порівняння показників та здійснити аналіз отриманих результатів.</i>	01.05.2025 – 31.05.2025

Завдання видано _____ доц. Коряшкіна Л. С.
(підпис) (прізвище, ініціали)

Дата видачі: 18.02.2025 р.

Дата подання до екзаменаційної комісії: _____ 13.06.2025 р.

Прийнято до виконання _____ Богун В. В.
(підпис студента) (прізвище, ініціали)

РЕФЕРАТ

Пояснювальна записка: 71 с, 7 рис., 3 додатки, 20 джерел.

Об'єктом дослідження є просторовий розподіл населення на основі супутникових знімків нічного освітлення.

Предметом дослідження є методи системного аналізу та комп'ютерного моделювання взаємозв'язку щільності населення зі світловим забрудненням в місті Берлін.

Мета дослідження є оцінювання щільності населення за рівнем світлового забруднення території шляхом аналізу супутникових даних з використанням статистичних методів і сучасних ГІС-технологій.

В інформаційно-аналітичному розділі проаналізовано підходи до аналізу просторового розподілу населення, методи описової статистики, машинного навчання, кластеризації, алгоритми виявлення кореляцій, а також програмні засоби обробки геопросторових даних. Досліджено наукові проєкти, які пов'язані з аналізом урбанізованих територій за супутниковими знімками.

У спеціальному розділі проведено підготовку вхідних даних для обробки. Зокрема, знайдено, очищено та зведено у формат координат (X, Y) та значень (Z) два незалежні набори: щільність населення міста Берлін та інтенсивність нічного освітлення за супутниковими знімками VIIRS. Масиви приведено до однакових географічних меж, що відповідають адміністративним межах міста.

В експериментально-аналітичному розділі реалізовано візуалізацію отриманих масивів у вигляді картографічного відображення, побудовано кореляційний аналіз між відповідними значеннями за допомогою коефіцієнта Пірсона. Здійснено перевірку гіпотези про наявність залежності між показниками, отримано графічні залежності та карти просторової кореляції.

Практична цінність отриманих результатів полягає у можливості швидкої оцінки щільності населення в урбанізованих регіонах за відкритими супутниковими даними.

Ключові слова: ЩІЛЬНІСТЬ НАСЕЛЕННЯ, VIIRS, DMSP-OLS, КОРЕЛЯЦІЙНИЙ АНАЛІЗ, ГІС-ТЕХНОЛОГІЇ.

ABSTRACT

Explanatory note: 71 p., 7 figures, 3 appendices, 20 sources.

The object of the study is the spatial distribution of the population based on satellite images of nighttime illumination.

The subject of the study is the methods of system analysis and computer modeling of the relationship between population density and light pollution in the city of Berlin.

The aim of the study is to assess population density based on the level of light pollution in the area through the analysis of satellite data using statistical methods and modern GIS technologies.

The information-analytical section analyzes approaches to studying the spatial distribution of the population, methods of descriptive statistics, machine learning, clustering, correlation detection algorithms, as well as software tools for geospatial data processing. Scientific projects related to the analysis of urbanized areas using satellite imagery are also studied.

In the special section, the input data for processing are prepared. Specifically, two independent datasets were collected, cleaned, and converted into coordinate format (X, Y) and values (Z): the population density of the city of Berlin and the intensity of nighttime illumination based on VIIRS satellite images. The datasets were aligned to the same geographical boundaries corresponding to the administrative borders of the city.

In the experimental-analytical section, the obtained datasets were visualized in the form of cartographic maps. Correlation analysis between the respective values was performed using the Pearson coefficient. The hypothesis regarding the presence of a relationship between the indicators was tested, graphical dependencies were obtained, and spatial correlation maps were developed.

The practical value of the obtained results lies in the possibility of rapid assessment of population density in urbanized regions using open satellite data.

Keywords: Population Density, VIIRS, DMSP-OLS, Correlation Analysis, GIS Technologies.

ЗМІСТ

ВСТУП.....	7
РОЗДІЛ 1. ІНФОРМАЦІЙНО-АНАЛІТИЧНИЙ РОЗДІЛ.....	10
1.1. Методи та алгоритми аналізу просторового розподілу населення	10
1.1.1 Просторовий аналіз та основні джерела даних.....	10
1.1.2 Методи описової статистики просторових даних	12
1.1.3 Кластеризація та математичні моделі просторового розподілу.....	14
1.1.4 Методи машинного навчання та програмні засоби для аналізу.....	19
1.1.5 Алгоритми виявлення кореляції між змінними	22
1.2. Аналіз існуючих наукових проектів	24
1.3 Змістовна постановка завдання	29
1.4 Висновки до першого розділу	31
РОЗДІЛ 2. СПЕЦІАЛЬНИЙ РОЗДІЛ	33
2.1 Вибір і підготовка даних.....	33
2.1.1 Вибір джерел даних про щільність населення.....	33
2.1.2 Первинне форматування та фільтрація геоданих	34
2.1.3 Опис джерела супутникових знімків VIIRS (особливості, переваги).....	35
2.1.4 Попередня обробка супутникових знімків (обрізка до меж міста)	37
2.2 Формування узгоджених масивів даних.....	39
2.3 Математичне моделювання	40
2.3.1 Опис методу кореляції Пірсона: математичний апарат	40
2.3.2 Особливості застосування до просторових даних	42
2.4.1 Мова програмування Python — обґрунтування вибору.....	44
2.4.2 Використані бібліотеки: pandas, numpy, matplotlib, geopandas, rasterio тощо	45
2.5 Висновки до розділу.....	47
РОЗДІЛ 3. ЕКСПЕРИМЕНТАЛЬНО-АНАЛІТИЧНИЙ РОЗДІЛ	50
3.1 Візуалізація вхідних даних	50
3.1.1 Відображення щільності населення (heatmap або contour plot).....	50
3.1.2 Відображення інтенсивності світлового забруднення.....	53
3.1.3 Порівняння за візуальними ознаками	55
3.2.1 Алгоритм узгодження координат і підрахунку.....	59
3.2.2 Розрахунок коефіцієнта Пірсона.....	60
3.2.3 Розрахунок коефіцієнтів кореляції Спірмена та Кендалла	61
3.2.4 Інтерпретація значення коефіцієнта	62
3.3.1 Побудова графіків для порівняння показників	63
3.4 Висновок до розділу 3.....	66
ВИСНОВКИ	68

	6
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	70
Додаток А. Відомість матеріалів кваліфікаційної роботи	72
Додаток Б. Відгук керівника кваліфікаційної роботи	73
Додаток В. Лістинг коду	74

ВСТУП

Упродовж останніх десятиліть урбанізація стала одним з основних процесів, що визначають розвиток суспільства на глобальному рівні. Зростання міст і урбанізованих територій супроводжується значними змінами не тільки в соціально-економічній структурі, але й у сфері навколишнього середовища. Міста, як центри концентрації населення, є основними джерелами антропогенного навантаження на екологічні системи, а швидкий розвиток інфраструктури сприяє погіршенню екологічної ситуації. Зміна екологічних параметрів, таких як якість повітря, рівень шуму, а також рівень світлового забруднення, є невід'ємною складовою процесів урбанізації, що потребує особливої уваги.

Одним з найбільш актуальних явищ, пов'язаних із зростанням урбанізаційних процесів, є світлове забруднення. Це феномен, що виникає в результаті надмірного чи нерегульованого використання штучного освітлення в нічний час, що, у свою чергу, призводить до порушення природних біоритмів живих організмів, змінює екосистеми та негативно впливає на здоров'я людини. У великих містах світлове забруднення часто стає настільки інтенсивним, що його не тільки важко ігнорувати, але й він стає частиною повсякденного життя городян. Відзначено, що найбільші концентрації світлового забруднення спостерігаються в місцях з високою щільністю населення, активним виробництвом та інтенсивним транспортним рухом. Тому вивчення взаємозв'язку між рівнем світлового забруднення та щільністю населення в урбанізованих територіях набуває особливої актуальності.

Актуальність цієї проблеми зростає в умовах глобальних змін, пов'язаних зі змінами клімату, зростанням чисельності населення та інтенсивною урбанізацією. У цих умовах важливо не тільки досліджувати рівень світлового забруднення, але й розробляти ефективні методи управління освітленням у міських умовах, щоб мінімізувати його негативний вплив на навколишнє середовище та здоров'я населення. Оскільки більшість світового

населення сьогодні проживає в міських агломераціях, вивчення таких проблем, як світлове забруднення, стає важливою частиною екологічних досліджень. Зокрема, для України, де проблема екології в міських районах є однією з ключових, ця тема набуває особливої ваги.

Технології, які дозволяють здійснювати моніторинг екологічних параметрів на великих територіях, є важливим інструментом для вирішення цієї проблеми. Сучасні методи дистанційного зондування Землі (ДЗЗ), зокрема супутникові знімки нічного освітлення, є потужним інструментом для вивчення світлового забруднення. За допомогою супутникових приладів VIIRS та DMSP-OLS можна не лише оцінити інтенсивність світлового забруднення на великих територіях, але й встановити кореляцію між рівнем освітлення та щільністю населення. Дані з цих приладів дозволяють детально досліджувати урбанізовані зони, визначати їх межі та ступінь впливу на екологічну ситуацію.

Особливу цінність у сучасних дослідженнях має інтеграція даних супутникового знімання з методами системного аналізу, математичного моделювання та геоінформаційних систем (ГІС). Застосування таких технологій дає змогу створювати точні моделі, що відображають взаємозв'язки між світловим забрудненням та демографічними показниками. Такі моделі дозволяють не тільки досліджувати вже наявні зв'язки, але й передбачати можливі сценарії розвитку ситуації в майбутньому, що є важливим для прийняття обґрунтованих рішень у сфері урбаністики та екології. Візуалізація таких моделей за допомогою сучасних програмних засобів, зокрема Python і відповідних бібліотек, дає можливість наочно продемонструвати результати аналізу та покращити процес прийняття рішень.

Метою цієї роботи є вивчення і моделювання кореляції між рівнем світлового забруднення та щільністю населення на прикладі міста Берлін, використовуючи супутникові дані та методи просторового аналізу. Це дослідження дасть змогу оцінити, як зміни в структурі розселення та інтенсивності освітлення впливають на екологічну ситуацію в урбанізованих зонах. Важливим етапом роботи є розробка та реалізація програмної моделі,

що дозволить на основі супутникових даних здійснювати просторовий аналіз і візуалізацію результатів. Це забезпечить можливість точного визначення зон з високим рівнем забруднення, що є важливим для управління міським середовищем і розробки заходів щодо зменшення впливу світлового забруднення на населення та навколишнє середовище.

Враховуючи актуальність та значущість теми, завданнями цієї роботи є: аналіз існуючих методик оцінки щільності населення за допомогою супутникових знімків, дослідження методів отримання даних про світлове забруднення, розробка математичної моделі для оцінки кореляції між інтенсивністю освітлення та щільністю населення, а також реалізація програмних засобів для візуалізації результатів. Використання таких підходів дасть змогу створити нові, більш ефективні методи для вирішення актуальних проблем урбанізації та екології.

Важливість цієї роботи також полягає в тому, що розроблені методи і моделі можуть бути використані не лише для аналізу ситуації в межах міста Берлін, а й адаптовані до інших урбанізованих регіонів. Це дозволить створити універсальний інструмент для дослідження взаємозв'язків між демографічними і екологічними факторами в контексті урбаністичних процесів. Результати дослідження матимуть значний практичний вплив на розробку політики екологічної безпеки та управління міським середовищем.

РОЗДІЛ 1. ІНФОРМАЦІЙНО-АНАЛІТИЧНИЙ РОЗДІЛ

1.1. Методи та алгоритми аналізу просторового розподілу населення

1.1.1 Просторовий аналіз та основні джерела даних

Просторовий розподіл населення — це ключовий аспект соціально-економічного аналізу, який визначає щільність, розміщення та взаємозв'язок між населеними пунктами в певній географічній області. Аналіз просторових структур дозволяє виявляти закономірності розселення, вивчати вплив інфраструктури, кліматичних умов, економічних факторів, а також екологічних чинників, зокрема світлового забруднення, яке в останні роки набуло особливого значення. Системний аналіз у цьому контексті дозволяє розглядати взаємозв'язки між компонентами складної соціально-екологічної системи — такими як щільність населення, нічна освітленість територій, транспортна інфраструктура, політика урбанізації тощо. У системному підході важливо не лише дослідити елементи системи, а й моделювати їхні взаємодії, впливи та зворотні зв'язки. Математичні та комп'ютерні методи уможливають формалізацію таких систем. Завдяки цифровим геоданим, сучасним обчислювальним алгоритмам і машинному навчанню, стало можливим проводити масштабні дослідження просторових даних і виявляти приховані патерни, які не були б очевидні при традиційному аналізі.

Однією з ключових задач у рамках цієї роботи є встановлення взаємозв'язку між щільністю населення і рівнем світлового забруднення. Ці дві величини можуть виявляти високий ступінь кореляції, проте лише математичне моделювання та аналітична оцінка дозволяють підтвердити або спростувати це припущення на основі конкретних даних. Варто також зазначити, що просторовий аналіз лежить в основі багатьох наукових дисциплін — географії, демографії, урбаністики, екології — і його методи інтегруються в сучасні міждисциплінарні дослідження. Таким чином, обрана тема є не лише актуальною, але й методично насиченою, що відкриває широкі можливості для застосування різних підходів аналізу та моделювання.

Для виконання якісного аналізу просторового розподілу населення необхідно використовувати достовірні джерела просторових даних, що забезпечують не лише координатну прив'язку об'єктів, а й атрибутивну інформацію, зокрема кількість населення, площу території, характеристики освітлення тощо. Основними типами джерел таких даних є: державні демографічні бази, супутникові спостереження, геоінформаційні системи (GIS), відкриті геодані та результати наукових досліджень.

1. Державні та міжнародні демографічні бази. Одним з найнадійніших джерел є дані національних статистичних служб, які проводять переписи населення. Наприклад, у Європі діє система Eurostat, що надає структуровані та агреговані дані по регіонах у форматі, придатному до просторового аналізу (наприклад, NUTS-рівні). Також використовуються дані Світового банку та Організації Об'єднаних Націй. У багатьох дослідженнях використовуються так звані ґридовані набори даних — це дані, розподілені по сітці (grid), що дозволяє оцінити щільність населення на клітинках, наприклад, 1×1 км. Прикладом є Global Human Settlement Layer (GHSL), Gridded Population of the World (GPW), WorldPop.

2. Супутникові знімки та освітленість. Світлове забруднення вивчається на основі даних з орбітальних супутників. Найбільш відомі джерела — це:

- DMSP-OLS (Defense Meteorological Satellite Program – Operational Linescan System) — історичні дані з нічного освітлення за останні 30 років.
- VIIRS (Visible Infrared Imaging Radiometer Suite) — більш сучасна система з вищою роздільною здатністю.

Ці зображення доступні через NASA Earthdata або NOAA, і широко використовуються в екологічному та урбаністичному аналізі.

3. Геоінформаційні системи (GIS). GIS-системи дають змогу інтегрувати просторові та атрибутивні дані, здійснювати аналіз, візуалізацію, трансформацію та екстракцію ознак. Найпопулярніші системи — QGIS, ArcGIS, а також бібліотеки для Python (наприклад, GeoPandas, Folium,

Rasterio). Вони дозволяють виконувати складні операції: overlay, spatial join, buffer, heatmap тощо.

4. Відкриті картографічні платформи. Одним з джерел допоміжної інформації є OpenStreetMap — глобальний відкритий картографічний проєкт. Його можна використовувати для аналізу забудованості, транспортної інфраструктури, типів землекористування.

5. Наукові агрегатори даних. У деяких випадках дослідники використовують агреговані бази, які поєднують різні шари даних: демографія + освітлення + висота + клімат. Наприклад, сайт SEDAC (Socioeconomic Data and Applications Center) або Google Earth Engine.

1.1.2 Методи описової статистики просторових даних

Описова статистика є фундаментальним інструментом початкового аналізу просторових даних, що дозволяє сформулювати базове уявлення про розподіл показників у межах досліджуваної території. Саме з цієї точки зору починається будь-який аналіз: перш ніж будувати складні моделі або застосовувати машинне навчання, необхідно виявити основні закономірності, оцінити наявність трендів, а також виявити потенційні аномалії. Просторові дані, зокрема щільність населення, мають певні особливості — вони характеризуються не лише чисельним значенням, а й географічним положенням, а тому класичні методи описової статистики потребують адаптації до двовимірного простору.

Першим кроком у цьому аналізі зазвичай є розрахунок базових показників: середнє значення щільності населення, медіана, мода, стандартне відхилення, коефіцієнт варіації, а також мінімальні та максимальні значення в межах географічного регіону. Наприклад, середнє значення щільності населення у великому місті може дати лише загальну оцінку, проте не дозволяє виявити сильний контраст між центральною частиною міста та околицями. Саме тому важливими є також методи просторового розподілу, які враховують

положення кожної ділянки — наприклад, коли кожна одиниця площі (піксель або полігон) має власне числове значення, яке може бути візуалізовано на карті.

Одним із найбільш поширених способів подання щільності населення є створення теплових карт (heatmaps), які візуалізують значення за допомогою колірної шкали. Такі карти дають можливість швидко виявити зони з найвищою концентрацією населення, візуально оцінити їхній розмір, протяжність та зв'язок з інфраструктурою. У поєднанні з іншими шарами — наприклад, освітленням територій — теплові карти дозволяють зробити перші припущення щодо потенційних взаємозв'язків.

Іншим важливим інструментом є ядерна оцінка щільності (Kernel Density Estimation, KDE), яка дозволяє побудувати гладке уявлення про розподіл населення, навіть якщо вихідні дані мають дискретний характер (наприклад, координати місць проживання або центри населених пунктів). Метод KDE працює за рахунок згладжування значень навколо кожної точки, у результаті чого утворюється континуальна поверхня, що відображає інтенсивність явища. Наприклад, KDE широко застосовується в урбаністиці для оцінки рівня забудованості, щільності населення, аварійності на дорогах та інших подій, прив'язаних до геолокації. У випадку з аналізом щільності населення KDE дозволяє краще побачити "гарячі зони" — регіони з концентрацією населення, навіть якщо вони не відображені чітко у табличних даних.

Картографічне подання результатів описової статистики є критично важливим у контексті просторового аналізу. Наприклад, картограми — це тематичні карти, на яких географічні регіони масштабуються відповідно до значення певного показника, зокрема щільності населення. Це дозволяє візуально підкреслити диспропорції між регіонами, показати, де відбувається концентрація населення, а де — його розрідження. Існує також підхід до побудови ізоліній (ізоплет), які з'єднують зони з однаковими значеннями щільності, що нагадує методика побудови висот на топографічних картах. Ці ізолінії дозволяють зробити оцінку зміни щільності в межах території без

різких переходів, що особливо корисно у регіонах із поступовою зміною густоти забудови.

Важливим аспектом є оцінка просторової автокореляції, яка вказує на те, наскільки значення показників в одному регіоні залежать від значень у сусідніх. Для цього використовуються такі статистичні показники як індекс Морана (Moran's I) та індекс ГEARІ (Geary's C). Позитивна автокореляція свідчить про те, що подібні значення щільності мають тенденцію групуватись у просторі (тобто високі значення розташовані поруч з іншими високими, і навпаки), а негативна — що сусідні значення, навпаки, сильно відрізняються. Це дає змогу глибше зрозуміти структуру просторового розподілу й дозволяє визначити зони високої внутрішньої неоднорідності.

З огляду на важливість візуалізації у просторовому аналізі, результати описової статистики часто поєднуються з інтерактивними картами, створеними за допомогою сучасних засобів візуалізації (наприклад, бібліотек Folium, Plotly або Dash у Python, Leaflet у JavaScript). Завдяки цьому результати стають не лише аналітично цінними, а й інтуїтивно зрозумілими навіть для користувачів, які не мають спеціальної статистичної підготовки.

1.1.3 Кластеризація та математичні моделі просторового розподілу

Кластеризація та сегментація просторових даних є потужними інструментами аналізу, які дозволяють виявити приховані закономірності у розподілі населення та зв'язках між щільністю населення і зовнішніми чинниками, такими як світлове забруднення. Ці методи належать до класу неконтрольованого навчання і базуються на поділі простору або об'єктів на групи, які мають спільні властивості. На відміну від описової статистики, що фіксує числові характеристики розподілу, кластеризація дозволяє перейти до виявлення структурованих закономірностей — наприклад, зон, у яких концентрація населення має схожі особливості, або в яких взаємозв'язок між густотою заселення і світловим навантаженням є подібним. Класичними

методами кластеризації є алгоритми k-середніх (k-means), ієрархічна кластеризація та DBSCAN. Алгоритм k-середніх широко застосовується для поділу просторових об'єктів (наприклад, пікселів супутникового знімка або одиниць адміністративного поділу) на k груп, кожна з яких має мінімальну внутрішньогрупову варіацію. У контексті щільності населення та світлового забруднення цей підхід дозволяє виділити регіони, де ці два показники мають подібний рівень або темп зростання. Наприклад, одна група може включати великі міста з високою щільністю населення та сильним освітленням, інша — передмістя з середніми значеннями, а третя — сільські райони з низькими показниками.

Проте просторові дані мають особливість — географічну залежність, тобто взаємозв'язок між сусідніми об'єктами. Стандартні алгоритми кластеризації цього не враховують, тому доцільно застосовувати модифіковані просторові алгоритми, які враховують просторову близькість об'єктів при формуванні кластерів. Один із таких підходів — спатіальна (просторова) кластеризація, яка враховує не лише значення атрибута (наприклад, щільність населення), але й координати кожної точки або об'єкта. У результаті створюються кластери, які водночас є і статистично однорідними, і просторово зв'язними. Це особливо корисно для аналізу урбанізованих територій, де контекст просторової близькості є критичним.

Алгоритм DBSCAN (Density-Based Spatial Clustering of Applications with Noise) дозволяє знаходити кластери довільної форми і є ефективним для виявлення висококонцентрованих регіонів населення серед розрідженого середовища. Його головна перевага — здатність працювати без чітко заданої кількості кластерів, а також ідентифікувати так званій "шум" — зони, що не входять у жодну групу. Такий шум часто відповідає малозаселеним територіям або природним зонам, де відсутнє світлове забруднення, що може стати важливим для побудови майбутніх екологічних або інфраструктурних моделей.

Сегментація, як метод поділу просторових даних, часто використовується у супутниковій зйомці та обробці зображень. У цьому контексті вона дозволяє виділити регіони зі схожою текстурою або спектральними характеристиками, які можуть відповідати різним типам заселення: щільно забудованим кварталам, промисловим зонам або зеленим насадженням. Завдяки сучасним алгоритмам машинного зору та глибокого навчання, такими як U-Net, Mask R-CNN чи SegNet, сегментація стала потужним інструментом у вивченні урбанізації, щільності заселення та екологічних змін у просторово-часовому вимірі. Наприклад, використання супутникових зображень нічного освітлення дозволяє провести сегментацію світлових плям та зіставити їх з картою населення, виявляючи кореляції та потенційні джерела перевищеного світлового навантаження.

У контексті міждисциплінарного аналізу, як у випадку зі щільністю населення та світловим забрудненням, ефективною є багатовимірна кластеризація, коли до аналізу залучаються одразу кілька параметрів: щільність, середній рівень освітлення, відстань до мегаполісів, рівень інфраструктурного розвитку тощо. Такий підхід дозволяє отримати комплексніші кластери, які відображають не лише поточний стан, а й потенційні вектори розвитку території або проблеми, пов'язані з перенавантаженням урбаністичного середовища. Проте для поглибленого аналізу і особливо — для побудови моделей прогнозування, оцінки впливів та оптимізації — необхідно перейти до застосування математичних моделей просторового розподілу, які дозволяють формалізувати взаємозв'язки між змінними, моделювати динаміку та структуру простору з використанням апарату аналітичної математики та чисельних методів.

Математичні моделі просторового розподілу зазвичай будуються з урахуванням таких основних компонентів: географічної локалізації, взаємодії між об'єктами (наприклад, населеними пунктами), часу та зовнішніх чинників (економічних, соціальних, екологічних). Одним із найбільш поширених підходів є використання диференціальних рівнянь, які описують зміну густини

населення або інших показників у просторі та часі. Наприклад, модель реакції-дифузії дозволяє описати, як густина населення змінюється під впливом локальної динаміки народжуваності/смертності (реакція) і міграції або розселення (дифузія). У таких моделях просторовий розподіл населення представлено як неперервне поле, а сам простір розглядається як континуум, що підлягає математичному опису.

Іншим підходом є стохастичне моделювання, де змінні описуються не лише як функції простору і часу, але й як випадкові величини з певними законами розподілу. Такий підхід особливо корисний при моделюванні урбаністичних процесів у великих просторових масштабах, де неможливо точно виміряти всі параметри системи. У контексті вивчення зв'язку між щільністю населення і світловим забрудненням стохастичні моделі можуть враховувати невизначеність у даних, коливання рівнів освітлення в залежності від пори року, погодних умов або типу забудови, а також відобразити імовірнісні сценарії зростання або зменшення щільності в окремих зонах.

Важливою групою є також регресійні просторові моделі, серед яких можна виділити просторову лінійну регресію, авторегресійні моделі SAR (spatial autoregressive) та моделі просторових похибок (SEM — spatial error models). У таких моделях щільність населення або рівень світлового забруднення можуть виступати як залежні змінні, а чинники, що на них впливають — як незалежні предиктори. Просторові регресійні моделі дозволяють врахувати той факт, що значення в сусідніх зонах не є незалежними, а мають просторову автокореляцію. Це означає, що висока щільність населення в одній зоні часто супроводжується подібними значеннями у сусідніх, що має бути враховано при статистичному аналізі, аби не отримати викривлених результатів. Такі моделі можуть використовуватись для кількісного аналізу впливу щільності населення на інтенсивність світлового забруднення, зокрема через побудову карт розподілу залишків, аналізу мультиколінеарності та оцінки значущості просторового компонента. Також варто згадати про моделі гравітаційного типу, які широко

застосовуються для опису переміщень населення, притягувальних ефектів мегаполісів або взаємодії між різними територіями. Ці моделі ґрунтуються на аналогії з фізичним законом тяжіння і припускають, що інтенсивність взаємодії між двома об'єктами пропорційна їх «масі» (наприклад, чисельності населення) і обернено пропорційна відстані між ними. У рамках дослідження, що стосується зв'язку між щільністю населення та світловим забрудненням, гравітаційна модель може бути використана для оцінки того, як світловий вплив мегаполісу поширюється на довколишні території з меншою щільністю населення, або як регіональна концентрація освітлення впливає на формування міської агломерації.

Ще один клас методів — агентно-орієнтовані моделі (ABM — agent-based modeling). Вони дозволяють моделювати взаємодію між індивідуальними агентами (мешканцями, забудовниками, адміністративними органами) у географічному середовищі. У цих моделях кожен агент має власні правила поведінки, що дозволяє враховувати складну динаміку на мікрорівні. Зокрема, це може бути корисно для моделювання розростання міст, процесів урбанізації або оцінки політик щодо обмеження світлового забруднення в умовах зростання щільності заселення. Агентно-орієнтовані моделі поєднують логіку імітаційного моделювання з географічними системами і вимагають значних обчислювальних ресурсів, однак дають змогу досліджувати сценарії, які важко або неможливо змоделювати за допомогою традиційних підходів.

Окремої уваги заслуговує інтеграція математичних моделей із геоінформаційними системами (ГІС), що дозволяє здійснювати просторову візуалізацію результатів моделювання та оперативну перевірку гіпотез. Моделі можуть бути реалізовані у вигляді сценаріїв у середовищах типу ArcGIS, QGIS або програмно через бібліотеки на Python (наприклад, GeoPandas, PySAL, rasterio), що забезпечує гнучкість у роботі з великими обсягами даних та проведення чисельних експериментів. Зокрема, розв'язки диференціальних рівнянь можна візуалізувати на карті, накладаючи їх на реальні географічні шари, що значно полегшує інтерпретацію та аналіз результатів.

1.1.4 Методи машинного навчання та програмні засоби для аналізу

У сучасному підході до аналізу просторових даних, особливо в контексті великих обсягів інформації та високої складності взаємозв'язків між змінними, усе більшої популярності набувають методи машинного навчання. Вони дозволяють моделювати не лише лінійні залежності, а й виявляти складні, приховані патерни у даних, що особливо цінно при вивченні взаємозв'язків, таких як вплив щільності населення на рівень світлового забруднення. Машинне навчання є гнучким інструментом, який добре масштабується, адаптується до структурованих і неструктурованих даних та демонструє високу продуктивність при розв'язанні завдань прогнозування, класифікації й виявлення аномалій.

Одним із базових інструментів у цьому контексті виступають регресійні моделі, адаптовані до просторових особливостей. Зокрема, просторово-зважена регресія (GWR, Geographically Weighted Regression) є ефективним методом, що дозволяє враховувати варіації впливу змінних у різних географічних точках. На відміну від класичних регресій, де параметри моделі є сталими по всьому простору, GWR дозволяє кожному об'єкту мати власний набір коефіцієнтів, що формуються з урахуванням просторового контексту. Це особливо актуально при вивченні регіональних відмінностей у розподілі щільності населення та світлового навантаження, оскільки ефект тієї самої змінної може бути різним у густозаселеній агломерації та на периферійній території. У задачах класифікації та структурного аналізу все частіше застосовуються алгоритми дерев рішень, а також їх ансамблеві варіанти — зокрема, Random Forest. Ці методи відомі своєю інтерпретованістю, здатністю працювати з категоріальними і числовими ознаками, а також вбудованою оцінкою важливості змінних. У просторовому аналізі Random Forest може бути використано для побудови моделей, які виявляють фактори впливу на рівень світлового забруднення, визначають території підвищеного ризику або

класифікують типи зон за інтенсивністю нічного освітлення. Залучення таких методів також дозволяє враховувати численні взаємозв'язки між просторовими, соціальними та економічними змінними без необхідності жорсткої апріорної формалізації.

Не менш перспективним напрямом є застосування нейронних мереж, зокрема згорткових нейронних мереж (Convolutional Neural Networks, CNN), які довели свою ефективність у роботі з візуальними та просторовими даними. У випадку досліджень світлового забруднення CNN дозволяють обробляти супутникові знімки нічного освітлення, виявляти аномальні зони, ідентифікувати закономірності в розподілі джерел світла та навіть будувати моделі передбачення розширення урбаністичних зон на основі історичних зображень. Навчання таких мереж потребує великих обсягів даних і обчислювальних ресурсів, але вони здатні виявляти складні нефіксовані структури, які важко описати традиційними математичними моделями.

Ключовою перевагою машинного навчання є його здатність об'єднувати різноманітні джерела інформації — географічні координати, статистичні показники, растрові зображення, часові ряди тощо — в єдину модель, що враховує багатовимірність простору. Такі підходи відкривають нові горизонти для міждисциплінарних досліджень, зокрема дозволяють поєднувати знання з географії, демографії, урбаністики, екології та комп'ютерних наук. У випадку теми, присвяченої аналізу щільності населення і світлового забруднення, машинне навчання не лише допомагає будувати точніші моделі, а й виявляти нетривіальні взаємозв'язки, які залишаються поза увагою класичного статистичного підходу.

Варто зазначити, що ефективне застосування методів машинного навчання в просторовому аналізі неможливе без використання спеціалізованих програмних засобів, які підтримують обробку геоданих, інтеграцію з ГІС, машинне навчання та візуалізацію результатів. Найбільш поширеними платформами для таких цілей є Python з його геоаналітичними та ML-бібліотеками, R з глибокими інструментами просторової статистики, а також

QGIS — візуальне середовище з розширеними можливостями картографії і попередньої обробки даних. Їх детальний огляд буде здійснено у наступному підпункті.

У контексті виконання системного аналізу просторового розподілу населення та рівня світлового забруднення, вибір програмного інструментарію є критично важливим. Сучасні дослідження у цій сфері передбачають роботу з великими масивами геопросторових даних, супутниковими знімками, статистичними показниками, а також виконання складних обчислень — від базової фільтрації до глибокого навчання. Тому необхідно використовувати гнучкі, модульні та добре підтримувані програмні платформи. Одними з найбільш популярних рішень є Python, R та QGIS — кожен із них має свої сильні сторони та підходить для різних етапів дослідження.

Мова програмування Python стала стандартом де-факто у сфері аналізу даних, і зокрема — у просторовому аналізі. Її відкритий екосистемний характер та активна спільнота розробників сприяють стрімкому розвитку інструментів. Бібліотека GeoPandas надає потужні можливості для роботи з геометрією об'єктів, маніпуляції з шейп-файлами, фільтрації та побудови просторових зв'язків. Для роботи з растровими даними, включаючи супутникові знімки нічного освітлення, активно використовується Rasterio, що дозволяє читати, обробляти та трансформувати геоприв'язані растри. Водночас бібліотека PySAL — це цілий набір інструментів для просторової статистики, аналізу автокореляції, побудови моделей GWR тощо. А для побудови моделей машинного навчання Python має scikit-learn, що забезпечує базовий набір алгоритмів, а також інтегрується з TensorFlow чи PyTorch для більш складних завдань.

У той же час, мова R залишається провідним інструментом у сфері академічного просторового аналізу та статистичного моделювання. Вона має глибоко продуману систему пакетів, серед яких виділяються sf (Simple Features), sp та raster, що дозволяють здійснювати складні просторові обчислення, працювати з великими наборами геоданих і створювати якісну

картографічну візуалізацію. Пакет `tmap` забезпечує створення інтерактивних або статичних карт з багатим функціоналом кастомізації, що особливо корисно для наукових публікацій або аналітичних звітів. Значною перевагою R є його орієнтація на глибокий статистичний аналіз, що дозволяє поєднувати просторові залежності з часовими рядами, факторними моделями чи байєсівськими підходами.

Для візуалізації та попередньої обробки просторових даних, а також для роботи з користувацькими шарами, широко використовується QGIS — потужна десктопна геоінформаційна система з відкритим кодом. QGIS підтримує широкий спектр форматів даних, дозволяє інтерактивно накладати й редагувати шари, використовувати готові плагіни, а також запускати скрипти на Python через інтегроване середовище. Її перевагою є можливість швидкої побудови тематичних карт, геопросторової агрегації та створення моделей без необхідності глибокого програмування. Для дослідника, який хоче швидко перевірити гіпотези або візуалізувати результати моделі, QGIS є незамінним інструментом.

Інтеграція цих інструментів — Python, R та QGIS — дозволяє побудувати повноцінний цикл аналітики: від збору та очищення даних до побудови складних моделей і формування кінцевих візуалізацій. Наприклад, можна провести первинну обробку супутникових знімків у QGIS, завантажити дані у Python для машинного аналізу, а потім експортувати результати до R для глибшого статистичного тестування. Такий підхід забезпечує не лише гнучкість, а й відтворюваність аналізу, що особливо важливо у науковій роботі.

1.1.5 Алгоритми виявлення кореляції між змінними

При побудові просторових моделей, особливо в контексті взаємозв'язку таких складних факторів, як щільність населення та світлове забруднення, важливо не лише візуалізувати дані, а й виявити реальні залежності між ними. Програмні засоби, розглянуті вище, надають можливість не тільки обробляти

й моделювати дані, але й проводити формальний кореляційний аналіз, який дозволяє встановити силу та характер взаємозв'язку між змінними.

Найпоширенішими статистичними підходами для цього залишаються класичні коефіцієнти кореляції: Пірсона, Спірмена та Кендалла. Коефіцієнт Пірсона застосовується у випадках, коли маємо справу з нормальним розподілом та лінійною залежністю між змінними. Він є зручним при первинному аналізі числових даних, наприклад, коли потрібно оцінити, чи збільшується рівень світлового забруднення зі зростанням щільності населення в певних регіонах. У випадках, коли розподіл не є нормальним або присутні значні викиди, застосовуються непараметричні аналоги — коефіцієнти Спірмена і Кендалла, які засновані на рангах і краще пристосовані до нестандартних даних або нелінійних залежностей. Це особливо актуально при роботі з неоднорідними територіями — наприклад, сільською місцевістю поблизу великих міст, де щільність може коливатися, а джерела освітлення — мати зовсім іншу природу.

Однак просте обчислення кореляцій для чисельних векторів не враховує просторовий характер даних. Саме тому для геоінформаційного аналізу часто використовують крос-кореляційні методи, які дозволяють дослідити, як зміна одного показника в певній локації впливає або корелює з іншим показником у сусідніх локаціях. Це надзвичайно важливо, адже, наприклад, яскраво освітлене місто може створювати світлове забруднення, яке поширюється за межі самого населеного пункту, формуючи зональний ефект у навколишніх районах. Просторово-зважена крос-кореляція дозволяє виявити такі тенденції й уточнити моделі, беручи до уваги не лише значення змінних, а й їхнє просторове розташування. Коли ж ідеться про багатовимірні залежності, де аналізується кілька десятків параметрів одночасно (наприклад, щільність, висотність забудови, тип дорожньої мережі, економічна активність, рівень світлового забруднення тощо), варто звертатися до методів зниження розмірності. Головні компонентні аналізи (PCA) та *t-distributed stochastic neighbor embedding* (t-SNE) є потужними інструментами для візуалізації

багатовимірних просторових даних. PCA дозволяє звести складні взаємозалежності до кількох головних компонент, що максимально пояснюють варіацію в даних, і таким чином дає змогу виявити, які саме фактори (з-поміж усіх) найбільше впливають на досліджуване явище. t-SNE, у свою чергу, є більш гнучким підходом, що дозволяє краще передавати локальні структури даних, зберігаючи відстані між точками у високовимірному просторі. Це особливо цінно при візуалізації кластерів або локальних аномалій у міських чи приміських районах.

Використання вищезазначених алгоритмів кореляційного аналізу, з урахуванням їх просторових характеристик, дозволяє отримати більш обґрунтовану та інтерпретовану модель взаємодії факторів. І що важливо — ці підходи органічно поєднуються з програмними засобами, розглянутими раніше: зокрема, бібліотеки Python (наприклад, `scipy.stats`, `seaborn`, `sklearn`) та R (`cor`, `corrplot`, `Hmisc`, `factoextra`) мають вбудовані засоби для швидкого обчислення коефіцієнтів, побудови графіків кореляційних матриць і виконання зниження розмірності.

1.2. Аналіз існуючих наукових проєктів

У цьому підпункті буде розглянуто приклади актуальних наукових досліджень, які стосуються просторового аналізу щільності населення, впливу урбанізації та проблеми світлового забруднення. Особливу увагу буде приділено роботам, у яких застосовано математичне моделювання, алгоритми машинного навчання, ГІС-аналіз або інші методи обробки просторових даних. Кожне з досліджень буде коротко охарактеризовано з точки зору методології, використаних інструментів, обсягів та типу даних, а також висновків, до яких дійшли автори. Це дозволить краще зрозуміти, які підходи вже показали себе ефективними, в яких напрямках ще є простір для розвитку, а також як саме можна адаптувати існуючий досвід до конкретної задачі аналізу взаємозв'язку щільності населення зі світловим забрудненням. Зібраний огляд також

допоможе обґрунтувати вибір методів, інструментів і структури моделі в межах даного дослідження.

Одним із цікавих прикладів застосування математичного моделювання у сфері просторового аналізу є дослідження під назвою "Modeling population density using land cover data", в якому пропонується інноваційний підхід до прогнозування розподілу населення на основі даних про земельний покрив. Основна ідея авторів полягає в тому, що інформація про типи землекористування (наприклад, орні землі, лісові масиви, урбанізовані території тощо) вже містить достатню кількість просторових характеристик, які корелюють з густиною населення. Це дозволяє зменшити залежність моделі від великої кількості вхідних змінних, використовуючи лише дані перепису населення та карту землекористування. У дослідженні було запропоновано модель CPDM (China Population Density Model), яка застосовувалася для моделювання щільності населення Китаю з просторовим розширенням у 1 км. Цікаво, що автори підходили до урбанізованих і сільських територій по-різному: для сільських зон було використано зважені лінійні моделі з багатофакторними коефіцієнтами, тоді як для міст — експоненційна модель спаду, що враховує розмір міста та відстань від центру. Такий зональний підхід у межах агроекологічних регіонів дозволив зменшити рівень похибок за рахунок локальної адаптації моделі.

Результати моделювання виявилися високоточними: середня відносна похибка становила 3,13% у сільських районах і 5,26% — у міських. Це значно краще порівняно з моделлю Landscan, яка показувала більші відхилення на рівні адміністративних одиниць. Модель CPDM виявилася не тільки ефективною, але й обчислювально простою, оскільки вимагає мінімум вхідних даних. Такий підхід демонструє потенціал використання обмеженого, але релевантного набору геопросторових параметрів для точного прогнозування розселення, що особливо актуально для регіонів із обмеженим доступом до деталізованих демографічних баз.

Іншим показовим прикладом більш складного та технологічно просунутого підходу до просторового моделювання є дослідження, присвячене провінції Сичуань у Китаї, де автори використали алгоритм Random Forest із застосуванням різних методів оптимізації гіперпараметрів. Метою роботи було створення детального розподілу населення з просторовою роздільністю 1 км, використовуючи багатоджерельні просторові дані, зокрема супутникові знімки, дані нічного освітлення, висоти будівель, POI (точки інтересу) та інші ознаки. Це дозволило згенерувати якісний і більш реалістичний шар населення, адаптований до локальних характеристик території. Особливу увагу в дослідженні приділено аналізу того, як методи оптимізації параметрів моделі (включаючи байєсівську оптимізацію з використанням гаусових процесів і градієнтного бустингу) впливають на якість моделі. У порівнянні з більш простими підходами, такими як випадковий пошук або пошук по сітці, саме байєсівська оптимізація дозволила досягти найвищих показників точності — найвищого коефіцієнта детермінації (R^2) і найменших похибок (MAE, RMSE) під час валідації на рівні волостей.

Крім високих чисельних результатів, робота демонструє важливість глибокого аналізу впливу вхідних ознак на результат — аналіз важливості ознак, проведений за допомогою SHAP-значень, показав, що найбільший вплив мають POI, пов'язані з повсякденним життям (магазини, лікарні, школи), а також нічне освітлення й штучні покриття. Ці дані, як виявилось, значно краще передбачають населеність території, ніж традиційні змінні.

Також підкреслено, що моделювання сільських територій залишається викликом — через меншу кількість структурованих даних, слабо виражені антропогенні ознаки та нижчу якість супутникових зображень. Проте навіть за таких умов модель перевершила відкриті глобальні набори даних — такі як GPW, LandScan та WorldPop — за точністю на локальному рівні, що свідчить про потенціал індивідуального налаштування моделей під конкретний регіон.

Ще один важливий напрям у дослідженнях стосується використання мультиджерельних супутникових і просторових даних для моделювання

щільності населення без прямої залежності від даних перепису. У дослідженні, присвяченому місту Хэфэй, запропоновано модель, яка інтегрує водонепроникні поверхні (Impervious Surfaces), дані нічного освітлення (NTL) і точки інтересу (POI) для побудови оцінки щільності населення з високою точністю. Модель базується на поетапній багатовимірній лінійній регресії з подальшим перенесенням параметрів на піксельний рівень (500 м), що дозволяє отримати деталізоване просторове представлення розподілу населення. На відміну від моделей, які використовують одну змінну, автори довели, що комбінування кількох слабо корельованих джерел (низький VIF і кореляція < 0.9) значно покращує прогностичну здатність моделі. Використана методика виявила сильний зв'язок між щільністю населення та показниками урбанізованості, що підтверджено високими значеннями скоригованого R^2 (до 0.852) та низькими помилками MAE і RMSE після 10-кратної крос-валідації.

Застосування моделі на піксельному рівні дозволило отримати просторову карту, яка значно точніше, ніж дані перепису, відображає внутрішні відмінності у населеності в межах адміністративних одиниць. Візуальне порівняння з супутниковими знімками високої роздільності підтвердило реалістичність результату — населення чітко концентрується в забудованих зонах, уникаючи неточностей, властивих грубому адміністративному узагальненню.

Загалом, дослідження демонструє ефективність простої, але грамотно інтегрованої багатовимірної моделі для отримання щільності населення в міжпереписні періоди, що особливо актуально для стрімко урбанізованих або дано-дефіцитних територій. Цей підхід може слугувати цінним інструментом для оперативного моніторингу змін, міського планування і надзвичайного реагування.

Ще одним підходом до просторової деталізації населення стало моделювання місько-сільських меж на основі уточненого розподілу щільності населення. У дослідженні на прикладі провінції Гуйчжоу (Китай) запропоновано двоетапну методику, що поєднує моделі машинного навчання

та геоінформаційний аналіз. Першим етапом було отримання високодеталізованих даних (30 м) про просторовий розподіл населення шляхом поєднання моделі випадкового лісу та дазиметричного картування з використанням різноманітних джерел: нічного освітлення (VIIRS/DMSP), землекористування, POI, топографічних та кліматичних характеристик, дорожньої інфраструктури та водних об'єктів. Другим етапом стало виділення меж міських і сільських територій за допомогою методу точки розриву, застосованого до отриманої карти щільності населення. Отримані результати показали високу точність — понад 93% симуляційної відповідності у моделі просторового розподілу населення та 88% точності у виділенні меж, підтверджену як стратифікованою випадковою перевіркою, так і зіставленням з супутниковими зображеннями високої роздільності. Найкращі результати досягнуті для густонаселених міських центрів, а межі малих міст виявилися складнішими для точного визначення.

Окремо підкреслено значущість POI, які мали найвищу вагу серед усіх предикторів (25.3%), що вказує на тісний зв'язок між функціональною урбанізацією і розподілом населення. Це дає змогу не лише уточнити межі, а й зрозуміти структуру урбанізованих ядер.

Метод визнано універсальним і практично застосовним завдяки використанню відкритих даних і відносній простоті реалізації. Він відкриває можливості для регулярного оновлення просторових меж міських агломерацій, що критично важливо в умовах динамічної урбанізації та відсутності актуальних переписів.

Іншим напрямом досліджень стало вдосконалення методів оцінки щільності населення з використанням супутникових знімків нічного освітлення DMSP-OLS. У межах дослідження, проведеного для провінції Гуандун (Китай), було запропоновано методику, що дозволяє зменшити вплив ефекту «засвітки» (overglow) міських територій і забезпечити роздільне моделювання міського та сільського населення. Для цього придатні до життя території були виділені на основі класифікації землекористування MODIS, що

дозволило виключити непридатні освітлені ділянки (водойми, болота, снігові покриви тощо). Далі території класифікувались як міські або сільські, що дало змогу застосовувати різні моделі для кожного типу середовища. Для міських територій використовувалася степенева модель, побудована на очищених зображеннях нічного освітлення, тоді як у сільських районах була запропонована нова модель SAHSI (slope-adjusted human settlement index), яка поєднувала показники рельєфу (DEM), вегетаційної активності (EVI) та освітленості (NTL).

Метод показав підвищену точність у порівнянні з іншими поширеними індексами, такими як VANUI чи HSI-EVI, як у міських, так і сільських умовах. Зокрема, у міських районах середня відносна помилка (MRE) зменшилась до 55.14%, а в сільських — до 71.23%, що свідчить про здатність моделі SAHSI враховувати складну морфологію рельєфу.

Отримана карта щільності населення (з роздільністю 1 км) відобразила характерні просторові закономірності, включно з високою концентрацією населення в дельті Жемчужної ріки та зниженням щільності у гірських районах. Дослідження підтвердило, що комбіноване використання нічного освітлення та землекористування є ефективним інструментом для оцінки щільності населення у відсутності або недостатності переписних даних.

1.3 Змістовна постановка завдання

Змістовна постановка завдання та мета кваліфікаційної роботи магістра полягає у математичному аналізі та комп'ютерному моделюванні взаємозв'язку між щільністю населення та світловим забрудненням на прикладі міста Для досягнення поставленої мети, постановку завдання було розбито на два глобальні етапи:

1. Математичне моделювання та аналітичне дослідження кореляції між щільністю населення та світловим забрудненням.

2. Програмна реалізація та візуалізація результатів дослідження з використанням інструментів комп'ютерного аналізу (Python).

На першому етапі розглянуто та проаналізовано математичні підходи до вивчення зв'язку між просторовими змінними. Основну увагу приділено статистичним методам оцінки кореляції між даними: коефіцієнтам кореляції Пірсона та Спірмена, регресійному аналізу, нормалізації значень, метрикам відстані та методам кластеризації, які можуть слугувати додатковим інструментом просторової інтерпретації.

Для самого процесу математичного аналізу сформульовано наступну постановку завдання:

Необхідно дослідити залежність між значеннями нічного освітлення (яке відображає світлове забруднення) та щільністю населення.

Потрібно визначити:

1. Чи існує статистично значущий кореляційний зв'язок.
2. Який тип зв'язку є найхарактернішим (лінійний, логарифмічний, експоненційний тощо).
3. Якими є аномалії — ділянки, де спостерігається невідповідність між високою/низькою щільністю населення та інтенсивністю освітлення.
4. Яку точність має модель кореляції за основними метриками (R^2 , середньоквадратична похибка, р-значення).

На другому етапі передбачається практична реалізація математичної моделі з використанням сучасних інструментів комп'ютерного аналізу, зокрема Python (бібліотеки Pandas, NumPy, GeoPandas, Matplotlib, Seaborn, Scikit-learn тощо).

Основні задачі цього етапу:

1. Збір даних:

- супутникові знімки нічного освітлення (наприклад, VIIRS/DMSP-OLS);
- статистичні дані про щільність населення;

2. Обробка даних:

- прив'язка знімків до координат;
- агрегація та нормалізація даних;
- злиття демографічної та супутникової інформації в єдину таблицю.

3. Проведення кореляційного аналізу:

- обчислення коефіцієнтів кореляції;
- побудова моделей залежності;
- статистична оцінка моделей.

4. Візуалізація результатів:

- тематичні карти з накладенням інтенсивності освітлення та щільності населення;
- графіки розсіювання, кореляційні діаграми, теплові карти;
- анімації або інтерактивні графіки (за можливості).

Результатом розробки має стати серія графічних і картографічних візуалізацій, що ілюструють виявлені закономірності, а також інтерпретація отриманих даних у вигляді статистичного висновку.

1.4 Висновки до першого розділу

Світлове забруднення — це складне багатоаспектне явище, яке, попри свою наочність, часто недооцінюється з точки зору впливу на соціальне, екологічне та економічне життя міст. Одним з ефективних інструментів для його аналізу є використання супутникових знімків нічного освітлення в поєднанні з просторовими даними щодо щільності населення.

У даному розділі було проаналізовано основні методи й алгоритми, що дозволяють здійснити системний аналіз просторового розподілу населення та дослідити його взаємозв'язок зі світловим забрудненням. Розглянуто джерела геопросторових даних, методи описової статистики, підходи до кластеризації, а також математичні моделі, що застосовуються у просторовому аналізі. Також

було розглянуто сучасні програмні засоби та методи машинного навчання, які можуть бути використані для аналізу даних дистанційного зондування Землі.

Особливу увагу приділено алгоритмам виявлення кореляційних зв'язків між змінними, що є ключовими для досягнення мети дослідження — виявлення статистичних залежностей між рівнем світлового забруднення та щільністю населення у межах міста Також проведено аналіз актуальних наукових проєктів, що використовують подібні підходи до вивчення просторових процесів. Це дозволило сформулювати бачення актуального рівня розвитку галузі, визначити найефективніші інструменти та підходи, які доцільно застосувати у подальшій роботі.

Постановка завдання, виконана наприкінці розділу, враховує як математичну складову дослідження (розробка моделей і алгоритмів аналізу кореляцій), так і потенційно програмну реалізацію (застосування Python та спеціалізованих бібліотек для візуалізації і моделювання процесів).

Таким чином, перший розділ створює аналітичне підґрунтя для подальшого математичного моделювання та експериментальної перевірки гіпотез щодо взаємозв'язку щільності населення зі світловим забрудненням у міському середовищі.

РОЗДІЛ 2. СПЕЦІАЛЬНИЙ РОЗДІЛ

2.1 Вибір і підготовка даних

2.1.1 Вибір джерел даних про щільність населення

Для проведення аналізу просторового розподілу населення було обрано високоточне джерело даних — ресурс WorldPop, який на сьогодні визнаний одним із провідних провайдерів глобальних демографічних даних. Цей проєкт реалізується спільними зусиллями кількох міжнародних наукових установ, включно з Університетом Саутгемптона, Університетом Луїсвілля, Університетом Намюра, а також Центром міжнародних досліджень інформаційних наук Землі при Колумбійському університеті. Джерело підтримується фондом Білла і Мелінди Гейтсів, що свідчить про його авторитетність та значущість у сфері аналізу населення.

У межах цього дослідження було обрано саме дані по території Німеччини, а конкретніше — масив даних, що охоплює місто Берлін. Дані надані у форматах GeoTIFF та ASCII XYZ, що дозволяє здійснювати гнучку обробку як у геоінформаційних системах (ГІС), так і в середовищі програмування Python. Для подальшого аналізу був використаний саме формат XYZ, де кожен запис складається з трьох змінних: x (довгота), y (широта) та z (оцінена кількість людей на квадратний кілометр).

Дані мають просторову роздільну здатність 30 секунд дуги, що еквівалентно приблизно одному кілометру на екваторі. Це забезпечує доволі високий рівень деталізації, придатний для просторового аналізу міського середовища. Проекція, в якій подано дані — Geographic Coordinate System, WGS84 — є стандартом для глобальних геоданих і сумісна з більшістю ГІС-платформ.

Особливу увагу варто звернути на метод, за допомогою якого було здійснено моделювання щільності населення. У цьому випадку мова йде про *dasymmetric redistribution* на основі алгоритму Random Forest. Це означає, що замість простого усереднення чи прив'язки до адміністративних меж, модель

використовує набір вхідних ознак (наприклад, використання землі, топографічні особливості, дані супутникових знімків), щоб з більшою точністю передати реальний розподіл населення.

Таким чином, вибір саме цього джерела дозволив забезпечити достатню точність і деталізацію для цілей дослідження, а також надав змогу гнучко інтегрувати дані в подальші етапи моделювання, що включають візуалізацію, статистичну обробку та порівняльний аналіз зі знімками нічного освітлення.

2.1.2 Первинне форматування та фільтрація геоданих

Після отримання вихідного масиву у форматі ASCII XYZ було здійснено попередню обробку даних з використанням мови програмування Python. Основною метою цього етапу було приведення масиву у вигляд, придатний до географічного аналізу, фільтрації та візуалізації. Для цього було використано низку бібліотек, серед яких — `pandas`, `geopandas`, `shapely` та `matplotlib`, які дозволяють ефективно працювати з просторовими даними у середовищі Python.

На першому етапі дані були імпортовані з CSV-файлу, в якому зберігалася інформація про щільність населення у вигляді координат (довгота та широта) та числового значення — кількості осіб на квадратний кілометр. Використовуючи функцію `Point` з бібліотеки `shapely`, кожен запис було перетворено у геометричний об'єкт, після чого створено гео-датасет з допомогою `GeoDataFrame`. Для координатної системи було обрано стандарт EPSG:4326 (WGS84), що є міжнародним стандартом для глобального геопросторового аналізу.

На наступному етапі обробки були завантажені межі адміністративних одиниць Німеччини у форматі `shapfile`. З цього шару було виокремлено полігон, що відповідає межах міста Берлін. Цей полігон використовувався як маска для обрізання — відбір лише тих точок, які просторово знаходяться всередині визначених меж. Операцію фільтрації було реалізовано за

допомогою методу `within()`, який дозволяє перевірити, чи належить певна точка заданій геометричній області.

Результатом цієї операції стало створення нової підмножини даних, яка містить лише ті записи, що відповідають території Берліна. Ці дані були збережені у форматі Excel-файлу для подальшої зручності обробки, перевірки та аналізу. Отримана таблиця вже є зручною основою для наступних етапів дослідження, таких як накладання супутникових знімків нічного освітлення та подальший статистичний аналіз. Важливо зазначити, що на цьому етапі було забезпечено як просторову, так і тематичну відповідність даних. Була проведена ретельна перевірка на наявність пропущених або аномальних значень, що могло б вплинути на результати подальшого аналізу. Крім того, застосування програмних засобів дозволило забезпечити масштабованість підходу, що може бути використано і для інших регіонів із подібною структурою геоданих.

2.1.3 Опис джерела супутникових знімків VIIRS (особливості, переваги)

Одним із найважливіших джерел для аналізу світлового забруднення та, відповідно, вивчення непрямих ознак щільності населення є супутникові знімки VIIRS. Повна назва цього інструменту — Visible Infrared Imaging Radiometer Suite. Це сучасна супутникова система, що є частиною програми спостереження за Землею «Suomi NPP» (National Polar-orbiting Partnership), яку було запущено у 2011 році як спільний проєкт NASA та NOAA. Однією з головних особливостей VIIRS є його здатність з високою точністю фіксувати не лише денні, а й нічні зображення поверхні планети, включаючи слабе штучне освітлення, розсіяне світло в містах, а також природні явища — наприклад, полярне сяйво чи пожежі. На відміну від попередніх супутникових систем, таких як DMSP-OLS (Defense Meteorological Satellite Program - Operational Linescan System), які мали обмежену роздільну здатність і більшу

похибку, VIIRS надає знімки високої якості з роздільною здатністю приблизно 750 метрів. Цей рівень деталізації дозволяє точно фіксувати просторове розташування джерел нічного світла в межах населених пунктів. Крім того, VIIRS має кращу калібровку сенсорів, що забезпечує стабільність та надійність вимірів у динаміці. Завдяки цьому знімки VIIRS часто використовують у наукових дослідженнях, пов'язаних з урбанізацією, енергоспоживанням, моніторингом стихійних лих, а також у проєктах оцінки динаміки населення.

Однією з переваг VIIRS є можливість отримання глобальних даних у відкритому доступі. Це дозволяє дослідникам з усього світу вільно завантажувати знімки і використовувати їх у своїх проєктах без додаткових ліцензій чи обмежень. Зображення надаються у форматі GeoTIFF, сумісному з більшістю геоінформаційних систем, а також підтримуються платформами обробки супутникових даних, зокрема в Python через бібліотеки rasterio, GDAL, georandas тощо. Важливо також зазначити, що дані VIIRS мають щоденну періодичність оновлення, що дозволяє здійснювати майже реальний моніторинг змін в урбанізованих територіях, наприклад у зв'язку з розвитком міста, змінами електроосвітлення або навіть надзвичайними ситуаціями.

У рамках цього дослідження було використано VIIRS-знімки, завантажені з інтерактивної онлайн-платформи Light Pollution Map (<https://www.lightpollutionmap.info>). Цей сайт є популярним джерелом візуалізації та завантаження даних про світлове забруднення по всій планеті. Інтерфейс сайту дозволяє користувачу легко вибрати необхідну територію, масштабувавши карту до меж конкретного міста, наприклад, Берліна. Далі можна завантажити супутникові знімки в різних форматах, зокрема GeoTIFF або PNG з геоприв'язкою, що дозволяє надалі перетворити їх у числові масиви для подальшого аналізу.

Особливістю платформи Light Pollution Map є інтеграція даних з кількох супутникових джерел та їх перетворення у зручну для користувача форму. Наприклад, платформа дозволяє бачити як абсолютні значення світлового

забруднення (у ніт або $\text{mag}/\text{arcsec}^2$), так і порівняльні шкали інтенсивності освітлення, що суттєво полегшує первинний візуальний аналіз. Завдяки зручному API та регулярному оновленню даних, ресурс широко використовується не тільки в наукових дослідженнях, а й у прикладних задачах — зокрема у міському плануванні, екологічному моніторингу та навіть у сфері астрономії.

У ході підготовки до практичного дослідження було завантажено знімки повної поверхні Землі. Однак для зосередження на об'єкті дослідження — місті Берлін — ці дані було обрізано до меж адміністративної території міста, аналогічно до попереднього етапу з обробкою даних щільності населення. Це забезпечило можливість прямого порівняння двох джерел інформації в одній системі координат, а також дозволило уникнути зайвого навантаження на систему під час візуалізації та обчислень. Загалом, вибір саме VIIRS як джерела супутникових знімків був зумовлений кількома чинниками: по-перше, це висока точність та деталізація; по-друге, регулярність оновлення даних; по-третє — вільний доступ через зручну онлайн-платформу. Крім того, важливим є те, що методології обробки знімків VIIRS уже підтвердили свою ефективність у попередніх дослідженнях, зокрема у проектах, що стосуються визначення урбанізованих територій або непрямого вимірювання щільності населення.

2.1.4 Попередня обробка супутникових знімків (обрізка до меж міста)

Після отримання супутникових знімків VIIRS із платформи Light Pollution Map наступним етапом дослідження стала їх попередня обробка. Метою цього кроку було виділення лише тієї частини супутникового зображення, яка відповідає географічним межам міста Берлін. Це дозволяє сфокусувати аналітичну увагу виключно на об'єкті дослідження, зменшити об'єм вхідних даних та оптимізувати подальші обчислення, включно з візуалізацією, розрахунком кореляції та збереженням у форматах, придатних

до обробки. Оскільки VIIRS-знімки охоплюють великі території — у даному випадку майже всю територію Німеччини або навіть ширше — необхідно було обрізати зображення до точних меж Берліна. Для цього було використано шейп-файл адміністративного поділу Німеччини (`gadm41_DEU_2.shp`), який містить геометричну інформацію щодо регіонів другого рівня — зокрема, земель та міст. Із цього файлу за допомогою бібліотеки `geopandas` було виділено полігон, який відповідає території Берліна, шляхом фільтрації рядків, у яких значення поля `"NAME_2"` дорівнювало `"Berlin"`.

Далі було здійснено ключову операцію — обрізку растрового зображення (GeoTIFF-файлу) на основі геометрії полігону Берліна. Для цього використовувалася бібліотека `rasterio`, яка забезпечує роботу з геопросторовими растровими даними. Насамперед необхідно було впевнитися, що система координат полігону Берліна співпадає з системою координат растрового знімка, оскільки в іншому випадку обрізка виконуватиметься некоректно. З цією метою було здійснено перетворення (проекцію) полігону до тієї ж системи координат, що й растр — зазвичай це WGS84 або інша проекція, вказана в метаданих растрового зображення.

Наступним кроком стало перетворення геометрії полігону у формат, сумісний із `rasterio.mask.mask()` — функцією, яка виконує обрізку растрового зображення по заданій геометрії. Ця функція приймає GeoJSON-подібну структуру координат, яку було отримано методом `__geo_interface__` на об'єкті `berlin.unary_union`. Об'єднання (`unary union`) гарантує, що навіть якщо місто складається з кількох полігонів (що трапляється, наприклад, у випадку з окремими адміністративними ділянками), обрізка відбудеться коректно по всій території. Функція `mask()` повертає обрізаний масив зображення та відповідну трансформацію (геоприв'язку), яка описує, як зображення розташоване у просторі. Після цього було оновлено метадані растрового файлу (висота, ширина, трансформація) та здійснено збереження обрізаного знімка до нового GeoTIFF-файлу — `ber_pd_2020_1km.tif`. Така структура дозволяє

надалі працювати лише з тими даними, які безпосередньо стосуються Берліна, не втрачаючи при цьому просторової точності.

На цьому етапі варто підкреслити, що проведення обрізки супутникового знімка не лише сприяє ефективнішій обробці, а й забезпечує коректність подальшого аналізу. Без цього кроку неможливо було б прямо порівнювати супутникові дані з демографічною інформацією, оскільки вони охоплювали б різні території. Крім того, збереження обрізаного знімка у вигляді окремого файлу дозволяє використовувати його у різних модулях проєкту незалежно, забезпечуючи гнучкість у подальших сценаріях роботи — наприклад, у побудові теплових карт, аналізі розподілу інтенсивності освітлення, чи розрахунках просторових статистичних метрик.

2.2 Формування узгоджених масивів даних

Після отримання та попередньої обробки двох основних джерел інформації — щільності населення та рівня світлового забруднення — наступним кроком стало приведення цих даних до єдиного формату, придатного для подальшого аналізу, зокрема для оцінки їх взаємозв'язку за допомогою статистичних методів.

Масив із щільністю населення, отриманий із WorldPop, уже мав зручну структуру: кожен рядок представляв координати точки (довгота та широта), а також числове значення щільності — кількість людей на квадратний кілометр. Після фільтрації за межами міста Берлін цей масив можна було безпосередньо використовувати для аналітичних розрахунків, оскільки кожна точка однозначно прив'язана до місця в просторі й має достовірне, кількісне значення.

Натомість дані світлового забруднення були отримані у вигляді супутникового знімка (растрового зображення), що охоплював територію всієї планети. Зображення було обрізане до точних меж Берліна, після чого виконано просте, але ефективне перетворення: для кожного пікселя

визначалося його просторове розташування (координати) та значення яскравості. Це значення відповідало рівню штучного освітлення й було представлене у вигляді цілих чисел від 0 до 11, де 0 означає повну темряву, а 11 — максимальну зафіксовану інтенсивність освітлення. Незважаючи на обмежену кількість градацій у зображенні VIIRS, таке представлення дозволяє охопити загальні тенденції та концентрації джерел світлового забруднення у місті. Оскільки географічні межі обох масивів (щільності населення та світлового забруднення) були однаковими, а крок сітки — співставним, стало можливим формування двох паралельних наборів точок із координатами та значеннями, які можна безпосередньо порівнювати між собою.

Таким чином, у результаті обробки було сформовано два узгоджених масиви — по суті, таблиці, де кожен елемент містить координати (x , y) і відповідне значення: або щільність населення, або рівень світлового забруднення. Це забезпечило можливість прямого порівняння значень у відповідних точках простору й створило надійну основу для статистичного аналізу, зокрема для обчислення коефіцієнта кореляції Пірсона, що стане предметом наступного розділу.

2.3 Математичне моделювання

2.3.1 Опис методу кореляції Пірсона: математичний апарат

Одним із ключових завдань даного дослідження є визначення ступеня взаємозв'язку між двома кількісними показниками — щільністю населення та рівнем світлового забруднення. Для оцінки такого зв'язку найдоцільніше використовувати **коефіцієнт кореляції Пірсона** — один з основних інструментів математичної статистики, який дозволяє кількісно описати лінійну залежність між двома змінними.

Суть методу

Коефіцієнт кореляції Пірсона (позначається як r) є безрозмірною величиною, що змінюється в інтервалі від -1 до $+1$:

значення $r = +1$ свідчить про **ідеальну пряму (позитивну) кореляцію**: при зростанні одного показника інший також зростає лінійно;

$r = -1$ — **ідеальна зворотна (негативна) кореляція**: при зростанні одного показника інший зменшується;

$r = 0$ — **відсутність лінійної залежності** між змінними.

В інших випадках значення r свідчить про наявність слабкої, середньої або сильної кореляції, яку можна оцінити за емпіричною шкалою:

$|r| < 0.3$ — слабка кореляція

$0.3 \leq |r| < 0.7$ — середня кореляція

$|r| \geq 0.7$ — сильна кореляція

Математична формула

Коефіцієнт Пірсона обчислюється за наступною формулою:

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

де:

x_i, y_i — значення першої змінної (в нашому випадку — щільність населення) у точці i ;

y_i, u_i — значення другої змінної (рівень світлового забруднення) у тій самій точці;

\bar{x}, \bar{y} — середні значення відповідних змінних;

n — кількість спостережень (у нашому випадку — кількість точок-пікселів у межах міста).

Формула по суті представляє собою відношення коваріації між двома змінними до добутку їх стандартних відхилень. Такий підхід дозволяє оцінити не лише напрям, але й силу зв'язку між змінними.

Умови застосування. Для коректного застосування методу Пірсона необхідно, щоб дані відповідали кільком базовим умовам:

- обидві змінні повинні бути кількісними (що у нашому випадку виконується — як щільність населення, так і рівень освітленості мають числову природу);

- змінні мають бути приблизно нормально розподілені (хоча в реальних умовах цей критерій часто порушується, Пірсон усе одно залишається надійним індикатором взаємозв'язку, особливо за великої вибірки);
- дані мають бути незалежними (кожна пара точок — незалежне спостереження).

Переваги та недоліки.

Коефіцієнт Пірсона є простим у реалізації та інтерпретації, що робить його особливо зручним для початкового аналізу. Його основна перевага — здатність швидко виявити наявність (чи відсутність) лінійної залежності між змінними.

Водночас метод має і певні обмеження. Зокрема, він не враховує нелінійні зв'язки, які можуть існувати між змінними. Крім того, він чутливий до викидів — тобто до окремих значень, що значно відрізняються від інших і можуть викривити результати.

2.3.2 Особливості застосування до просторових даних

Застосування методу кореляції Пірсона до просторових даних, зокрема таких, що мають географічну прив'язку, вимагає додаткового осмислення та врахування низки особливостей. Просторові дані відрізняються від класичних табличних змінних тим, що кожна точка спостереження має не лише числові значення, але й координатне положення у певному фізичному просторі. Це означає, що такі дані схильні до просторової автокореляції, тобто значення в сусідніх точках можуть бути подібними не випадково, а саме через географічну близькість.

Визначення просторової автокореляції. Просторова автокореляція — це явище, коли значення змінної в певному місці залежить від значень у навколишніх точках. Наприклад, у містах спостерігається скупчення населення в певних районах, а в інших — навпаки, дуже низька щільність. Аналогічно, світлове забруднення зазвичай концентрується поблизу

густонаселених чи індустріально активних зон, створюючи великі світлі «плями» на супутникових знімках. Це означає, що дані не є незалежними одне від одного, а така залежність є прямим порушенням одного з ключових припущень класичної кореляції Пірсона.

Просторове згладжування. Ще однією особливістю є різна роздільна здатність масивів, що порівнюються. У нашому випадку щільність населення подана з високою деталізацією, тобто дані чітко розділені по ділянках розміром приблизно 1 км², тоді як знімок світлового забруднення — менш деталізований, і має лише 11 рівнів яскравості. Це обумовлює різну чутливість до локальних коливань, а тому при обробці таких даних важливо або узгодити їх просторову роздільність, або обережно інтерпретувати результати.

У нашій роботі було прийнято рішення обрізати обидва масиви до однакової географічної області — меж міста Берлін. Таким чином забезпечується порівнюваність простору, в якому значення зчитуються. Для вирівнювання структури даних було здійснено перетворення зображення супутникового знімка в масив пікселів, де кожен піксель має координати та значення освітленості. Оскільки площа покриття обох масивів є однаковою, можливо зіставити точки в одному масиві з відповідними точками в іншому.

Ще один важливий аспект — агрегування чи збереження початкової структури. Якщо в одному масиві пікселів значно більше, ніж в іншому, то застосування Пірсона без нормалізації або згладжування призведе до нерівномірного внеску деяких ділянок у загальний результат. У нашому випадку, на щастя, це питання було знято ще на етапі попередньої обробки: обидва масиви були приведені до спільного вигляду — координат + значення, з однаковою кількістю точок, що дозволяє уникнути втрати даних або спотворення аналізу.

Інтерпретація результатів у геопросторовому контексті. У класичній статистиці високий коефіцієнт кореляції Пірсона автоматично свідчить про сильний зв'язок між змінними. У випадку ж просторових даних необхідно пам'ятати, що навіть невисоке значення кореляції може бути значущим,

оскільки геопросторові змінні часто мають складну структуру і не завжди утворюють просту лінійну залежність. Таким чином, при аналізі результатів слід зважати не лише на числове значення r , але і на просторову логіку: чи розташовані зони високої освітленості в районах з високою щільністю населення? Чи є відмінності між центральною частиною міста та передмістям? Відповіді на ці питання допомагають глибше зрозуміти характер взаємозв'язку та уникнути помилкових висновків.

2.4.1 Мова програмування Python — обґрунтування вибору

У рамках даного дослідження як основний інструмент для обробки, аналізу та візуалізації даних було обрано мову програмування Python. Такий вибір зумовлений рядом об'єктивних причин, які стосуються як функціональних можливостей мови, так і її переваг у сфері геоінформаційного аналізу, статистичного моделювання та роботи з великими масивами даних.

По-перше, Python — це мова загального призначення, що вирізняється лаконічним синтаксисом та високою читабельністю коду. Завдяки цьому вона є зручною для швидкої розробки прототипів, проведення експериментів та реалізації складних аналітичних підходів. Це особливо актуально для наукових досліджень, де часто необхідно адаптувати підходи або повторно запускати розрахунки з новими параметрами.

По-друге, екосистема Python включає в себе потужні спеціалізовані бібліотеки, які суттєво полегшують роботу з просторовими, растровими та табличними даними. Зокрема, у цьому проєкті активно використовувались такі пакети як:

`pandas` — для обробки табличних даних;

`geopandas` — для роботи з географічними об'єктами (точками, полігонами) та просторовими операціями;

`rasterio` — для обробки та аналізу супутникових знімків у форматах GeoTIFF;

`matplotlib` — для побудови графіків та візуалізацій;

`scipy.stats` — для реалізації статистичних методів, зокрема кореляційного аналізу.

Ще однією важливою перевагою є те, що Python широко підтримується науковою спільнотою. Багато сучасних досліджень у галузях геоінформатики, урбаністики, екології та демографії базуються саме на Python, що дозволяє використовувати перевірені методики та приклади з відкритих джерел. Додатково, Python є платформонезалежним, тобто може працювати як на Windows, так і на Linux чи macOS, що забезпечує гнучкість у виборі середовища для обчислень. Також варто зазначити, що Python добре підходить для автоматизації обробки даних. У випадку з супутниковими знімками та просторовими шарами, де обсяг інформації є значним, можливість реалізувати пакетну обробку або циклічний аналіз значно пришвидшує роботу дослідника.

Загалом, Python виступає як універсальний інструмент, що об'єднує в собі аналітичну гнучкість, потужність бібліотек та простоту реалізації. У цьому дослідженні саме він дозволив поєднати супутникові знімки, просторові шари та демографічну інформацію в єдину систему для подальшого статистичного аналізу та візуалізації результатів.

2.4.2 Використані бібліотеки: `pandas`, `numpy`, `matplotlib`, `geopandas`, `rasterio` тощо

У процесі реалізації даного дослідження активно використовувався широкий спектр бібліотек мови Python, які забезпечили повний цикл роботи з даними — від попередньої обробки й візуалізації до статистичного аналізу та просторової фільтрації. Кожна з цих бібліотек виконувала свою окрему функцію, дозволяючи ефективно взаємодіяти як з растровими, так і з векторними геоданими, а також здійснювати глибоку аналітичну обробку.

Однією з ключових бібліотек стала `pandas` — універсальний інструмент для роботи з табличними структурами даних. Вона дозволила завантажити великі масиви інформації, зокрема у форматі CSV та Excel, а також здійснювати сортування, фільтрацію, агрегацію та обчислення над стовпцями. `Pandas` забезпечила зручний спосіб інтегрувати координатні значення та значення щільності населення, а також додати результати просторових вибірок.

Бібліотека `numpy` була використана переважно для низькорівневих обчислень та роботи з багатовимірними масивами. Вона дозволила швидко і зручно працювати з растровими даними, що зберігалися у вигляді числових матриць, та проводити базові математичні операції. Зокрема, `numpy` був корисний при обробці зображень, нормалізації даних і підготовці масивів до аналізу.

Бібліотека `geopandas` стала ключовою для обробки просторових даних у векторному форматі (точки, полігони, лінії). Саме за допомогою цієї бібліотеки здійснювалося завантаження `shapefile` з межами німецьких регіонів, вибірка полігону міста Берлін, а також просторове фільтрування населення та точок пікселів, що потрапляли всередину цього полігону. `GeoDataFrame`, який створюється за допомогою `geopandas`, дозволяє комбінувати геометричну інформацію з табличною структурою, що особливо зручно при візуалізації та експорту результатів.

Ще однією фундаментальною бібліотекою була `rasterio`, що забезпечує гнучку роботу з геопросторовими растровими файлами (`GeoTIFF` тощо). Саме через `rasterio` здійснювалося відкриття супутникових знімків формату `VIIRS`, обрізка за полігоном Берліну за допомогою методу `mask`, а також витяг числових значень світлового забруднення для подальшого порівняння з щільністю населення. Завдяки підтримці різних систем координат і автоматичному управлінню метаданими, `rasterio` є незамінним інструментом для роботи з супутниковими даними.

Для побудови графіків, візуалізації розподілу даних та створення діаграм використовувалась `matplotlib` — одна з найстаріших і найнадійніших бібліотек для наукової графіки. З її допомогою було побудовано гістограми, діаграми розсіювання та графіки, що демонструють зв'язок між світловим забрудненням та щільністю населення. В окремих випадках також використовувалась `seaborn` — візуальна надбудова над `matplotlib`, яка дозволяє створювати більш стилізовані й інформативні графіки з мінімальними зусиллями. Вона використовувалась для побудови теплових карт і графіків кореляції з автоматичним відображенням коефіцієнтів.

Крім того, для зчитування та розпакування файлів, зокрема у форматі `.gz`, було використано модулі `gzip` та `shutil`. Вони дозволили здійснити первинну обробку великих файлів, отриманих із відкритих джерел, і розпакувати їх у придатному для аналізу вигляді.

І, нарешті, бібліотека `scipy`, а точніше її модуль `scipy.stats`, була використана для реалізації методу кореляції Пірсона. Саме завдяки їй здійснювався розрахунок коефіцієнтів залежності між двома масивами чисел — щільністю населення та рівнем світлового забруднення в межах міста Берлін.

Таким чином, сукупність вищезгаданих бібліотек створила повноцінне середовище для виконання складного міждисциплінарного аналізу, що поєднує просторову аналітику, роботу з супутниковими знімками та статистичне моделювання.

2.5 Висновки до розділу

Проведена у другому розділі робота стала основою для всієї подальшої частини дослідження. Було виконано значну частину підготовки, що охоплює збір, відбір, трансформацію та просторову синхронізацію даних. Ретельне відношення до цього етапу дало змогу уникнути багатьох типових помилок у роботі з геоданими — зокрема, неузгодженостей у системах координат,

розбіжностей у межах регіонів або невідповідностей у форматах вхідних файлів. Для аналізу щільності населення використано високоточний масив, сформований за допомогою Random Forest-методики з розподіленням на основі реальних статистичних та географічних джерел. Цей масив має формат XYZ, де X та Y — це координати, а Z — чисельне значення щільності в певній точці, виражене в людях на квадратний кілометр. Масив був обрізаний геометрією міста Берлін, що дозволило сфокусуватись лише на необхідному регіоні без надмірних і зайвих даних з інших територій Німеччини.

Своєю чергою, дані про світлове забруднення були отримані із супутникових знімків VIIRS. Це растрові дані, де яскравість світіння вночі для кожного пікселя визначає рівень штучного освітлення на поверхні землі. На відміну від щільності населення, де значення можуть сягати кількох тисяч, знімки світлового забруднення мають всього 11 градацій значення (від 0 до 11), що вказує на нижчу деталізацію. Тим не менш, зображення було обрізане до тих самих меж, що й масив населення, що забезпечило цілісність у порівнянні. Попри нижчу роздільну здатність світлового знімка, його географічне вирівнювання з масивом населення дозволило перетворити кожен піксель у координату з відповідним числовим значенням — показником яскравості. Таким чином, обидва масиви, хоч і отримані з різних джерел, зрештою були представлені в уніфікованому вигляді: координати + значення. Це дозволяє надалі працювати з ними математично, накладати один на одного, порівнювати, будувати графіки та обчислювати ступінь зв'язку між ними.

У межах цього розділу також було задіяно широкий спектр програмних інструментів, серед яких GeoPandas, Rasterio, Pandas, NumPy та інші. Вони дозволили зручно обробляти як векторні, так і растрові дані, виконувати просторові фільтрації, маскування, проектування та підготовку до аналізу.

Цей баланс між кількісною точністю та просторовою цілісністю є ключовим для досягнення валідних результатів. Дані було приведено до такого формату, який дозволяє уникнути помилок при аналізі та забезпечити релевантність результатів математичного моделювання. Таким чином, логічно

завершивши розділ 2, ми повністю підготували ґрунт для розділу 3, де буде представлений власне аналіз взаємозв'язків та інтерпретація отриманих числових результатів.

РОЗДІЛ 3. ЕКСПЕРИМЕНТАЛЬНО-АНАЛІТИЧНИЙ РОЗДІЛ

3.1 Візуалізація вхідних даних

3.1.1 Відображення щільності населення (heatmap або contour plot)

Візуалізація просторових даних відіграє ключову роль у виявленні закономірностей, які можуть бути неочевидними при роботі з табличними форматами. Саме з цією метою було побудовано теплову карту (heatmap) щільності населення міста Берлін, що дозволяє наочно представити, як розподілено населення по території.

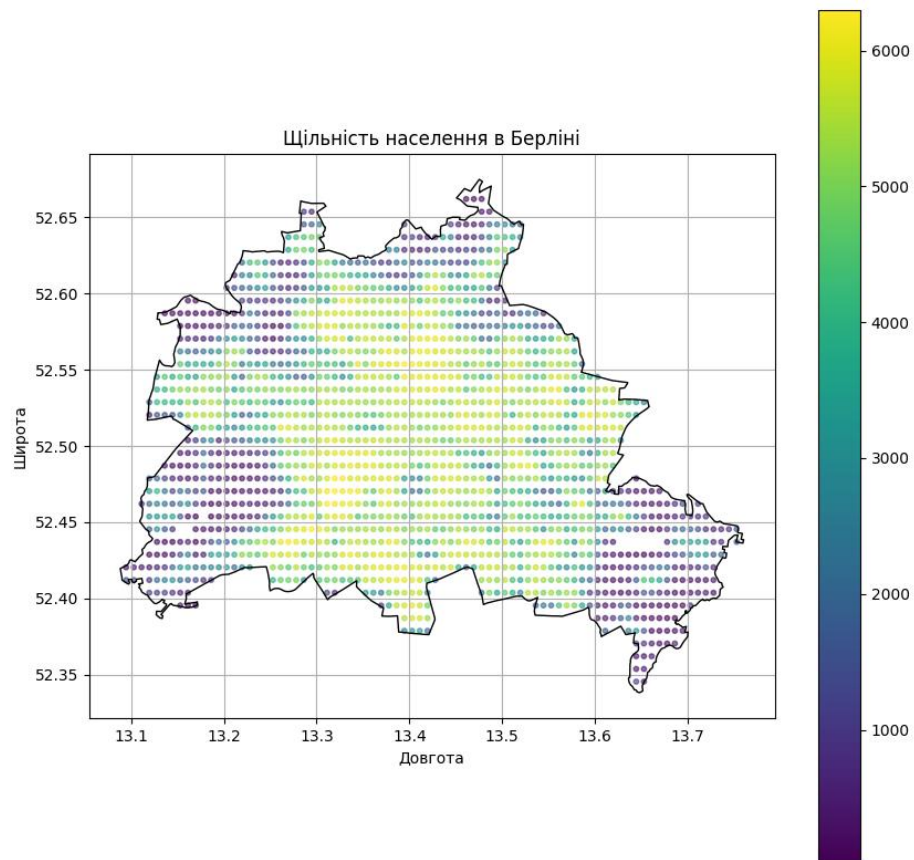


Рисунок 3.1 – Щільність населення Берліну точками

В основі карти лежить масив, отриманий шляхом обрізки всіх точок із загальнонімецького датасету до меж міста. Кожна точка має координати та значення щільності населення в людей на квадратний кілометр. Для побудови графіку використовувалася бібліотека `matplotlib`, яка дозволяє швидко та ефективно виводити геопросторову інформацію. Було прийнято рішення

відобразити карту у вигляді теплового зображення, де більш насичені тони відповідають вищій щільності населення, а світлі або холодні — нижчій.

Перевага такого підходу полягає в його інтуїтивній зрозумілості. Навіть не маючи спеціальної підготовки, глядач може швидко побачити, де саме зосереджене населення. Так, наприклад, центральні частини Берліна — зокрема райони Мітте, Пренцлауер-Берг, Фрідріхсайд — відображаються як зони з високим рівнем заселеності. Натомість периферійні райони, зокрема лісові або індустріальні зони, демонструють значно нижчу щільність.

Було також протестовано альтернативний підхід — побудова контурного графіку (contour plot). Цей метод дозволяє відобразити щільність у вигляді ізоліній, тобто ліній однакового значення. Такий тип графіку є надзвичайно корисним при аналітичному вивченні змінних, оскільки надає змогу простежити градієнти зміни щільності та визначити, де саме відбуваються переходи між більш і менш населеними ділянками. Проте, порівняно з тепловою картою, контурний графік менш виразний для нефахового користувача.

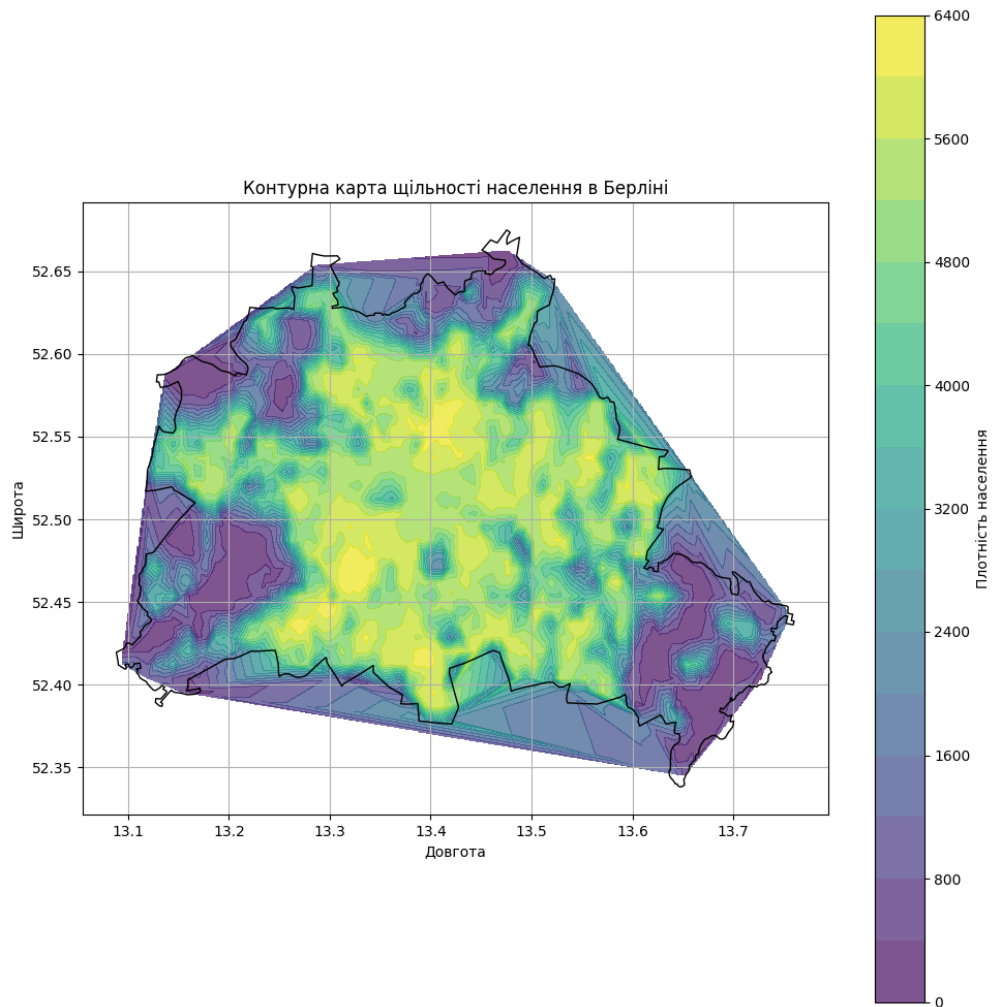


Рисунок 3.2 – Контурна карта щільності населення Берліну

З технічного боку, для побудови графіків дані було попередньо згруповано у формат координатної сітки, яка відповідала просторовому розташуванню пікселів. У разі необхідності заповнення пропущених значень або згладжування меж між зонами було застосовано методи інтерполяції на базі `scipy` або `seaborn`.

Не менш важливим є те, що побудовані візуалізації не лише інформативні, але й створюють підґрунтя для подальшого порівняння з іншими просторовими шарами — зокрема, з картою світлового забруднення. Адже наявність візуального шару дає змогу здійснювати як якісні, так і кількісні порівняння у наступних підрозділах дослідження.

Таким чином, візуалізація щільності населення відіграє подвійну роль: з одного боку — це спосіб представлення даних, а з іншого — важливий

інструмент аналізу, який дозволяє робити первинні висновки ще до застосування математичних методів. Візуальна картина одразу підказує, що найбільш насичені зони проживання у місті цілком збігаються з історично сформованими урбанізованими ядрами.

3.1.2 Відображення інтенсивності світлового забруднення

Після обрізки глобального супутникового знімка VIIRS до меж міста Берлін, стало можливим перейти до візуалізації просторового розподілу світлового забруднення в межах міської території. Це стало одним із ключових етапів дослідження, адже саме на основі цієї візуалізації відбувалося подальше порівняння з даними про щільність населення.

Для побудови карти світлового забруднення використовувався растр з попередньо обрізаними даними, де кожному пікселю відповідає значення інтенсивності освітлення у нічний час. Палітра кольорів у цій візуалізації була підібрана таким чином, щоби чітко відобразити градацію інтенсивності: від темно-фіолетових тонів для слабкоосвітлених районів — до яскраво-жовтих та зелених відтінків, які відображають найвищі рівні світлового навантаження.

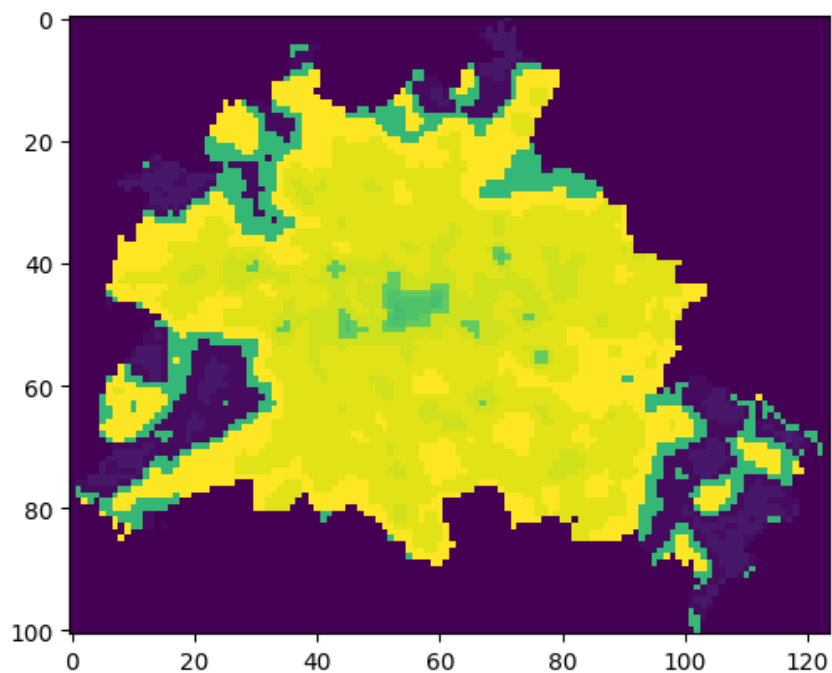


Рисунок 3.3 – Карта світлового забруднення

При побудові карти використовувалась функція `show()` з бібліотеки `rasterio.plot`, яка дозволяє швидко та якісно відобразити двовимірну структуру растрового знімка. Також додатково використовувалась бібліотека `matplotlib`, яка дала змогу оформити карту згідно з науковими вимогами — з відповідною легендою, підписами осей, масштабом та кольоровою шкалою. Це полегшило не лише інтерпретацію результатів, але й порівняння з іншими картографічними шарами.

При аналізі отриманої карти світлового забруднення стає помітно, що максимальна інтенсивність зосереджена у центральній частині Берліна — там, де сконцентрована адміністративна, торгова, транспортна та культурна інфраструктура. До таких зон відносяться райони Мітте, Фрідріхсхайн, частково Шарлоттенбург. Ці зони на карті мають найвищі значення — 10 або 11 за шкалою індексації зображення, що є характерним для густонаселених урбанізованих територій з активним нічним освітленням.

Водночас можна побачити великі ділянки з мінімальним освітленням — лісові зони, міські парки, периферійні житлові масиви з низькоповерховою забудовою. Тут інтенсивність освітлення коливається в межах 0–3, і на карті вони відображаються холодними кольорами. Залежно від особливостей інфраструктури, інколи навіть окремі дороги чи промислові об'єкти можуть мати вищу інтенсивність, ніж довколишні житлові райони.

Загалом, карта світлового забруднення створює деталізовану картину нічного освітлення міста, яка є як візуально зрозумілою, так і зручною для подальшого аналітичного аналізу. Варто зазначити, що хоча дані мають нижчу роздільну здатність порівняно з картою населення, вони все ж дозволяють визначити ключові зони концентрації світлового навантаження. Це дає підстави для побудови кореляційних моделей, які можуть описати зв'язки між щільністю населення та інтенсивністю нічного освітлення в межах урбаністичного простору.

3.1.3 Порівняння за візуальними ознаками

Порівняння щільності населення та рівня світлового забруднення за візуальними ознаками є важливим етапом попереднього аналізу, що дозволяє сформулювати первинні гіпотези ще до математичної обробки. Візуалізація дозволяє побачити розподіли, виявити просторові патерни, а також зіставити структурні особливості двох різнорідних типів даних у контексті міського простору.

Під час побудови теплової карти щільності населення було помітно, що основні осередки концентрації населення зосереджені у центральних та східних частинах Берліна. Тепловий відтінок на таких ділянках поступово переходить від чорного до білого кольору, що відповідає високим значенням густоти. Натомість у периферійних зонах — лісах, паркових зонах, промислових територіях — кольорова палітра майже відсутня, що візуально підкреслює відсутність постійного населення.

Супутникові знімки світлового забруднення (на основі даних VIIRS) мають іншу стилістику, однак дають змогу простежити подібну просторову структуру. Центральні райони — зокрема Мітте, Кройцберг, Шарлоттенбург — виділяються яскравими червоними пікселями, що візуально відповідає високим рівням освітленості. Проте, на відміну від карти населення, тут зустрічаються зони підвищеної яскравості навіть у місцях із незначною щільністю — як-от транспортні вузли, залізничні станції, території аеропортів або великі торговельні комплекси.

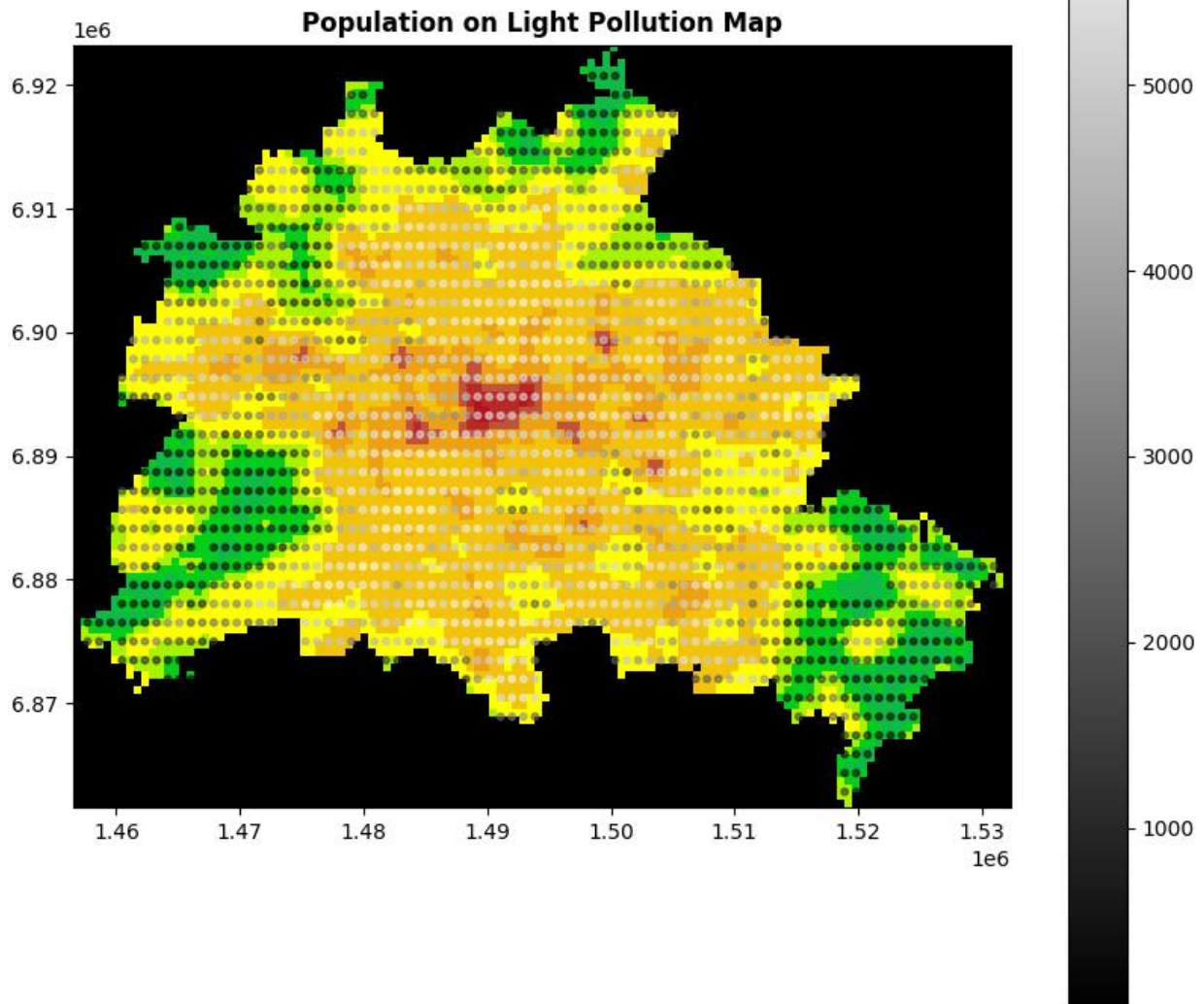


Рисунок 3.4 – Карта світлового забруднення та щільності населення

При візуальному накладанні обох карт можна спостерігати як зони прямої відповідності (висока щільність — високе освітлення), так і певні розбіжності. Наприклад, деякі густозаселені житлові квартали, особливо в спальних районах, мають помірний рівень освітленості — ймовірно, через меншу інтенсивність нічної активності або специфіку забудови. Водночас окремі зони високої яскравості на знімках світлового забруднення взагалі не мають відповідника у карті щільності — як уже зазначалося, це можуть бути нежитлові об'єкти. Загалом, візуальний аналіз підтверджує наявність зв'язку між щільністю населення та світловим забрудненням, але також підкреслює,

що цей зв'язок не є лінійним та однозначним. Важливим є також фактор інфраструктури, типу забудови, функціонального призначення території. Візуальна подібність контурів між двома картами служить своєрідною передумовою для більш глибокого статистичного аналізу, зокрема розрахунку коефіцієнтів кореляції.

Також на карті можна спостерігати місця із середнім рівнем світлового забруднення, але з мінімальним або нульовим населенням. Це можуть бути великі відкриті простори, парки, площі або інфраструктурні об'єкти. Характерним прикладом є Темпельхофер Фельд — громадський парк на місці колишнього аеропорту. Ця територія має добре освітлену інфраструктуру та великі відкриті площі для відпочинку, катання на велосипедах і пікніків. На карті світлового забруднення ця зона чітко виділяється світло-жовтим кольором, що підтверджує її середню яскравість, однак теплокарта населення демонструє тут повну відсутність густоти — територія позначена чорним кольором. Саме через наявність таких "аномальних зон" ми не отримуємо максимально високого значення коефіцієнта кореляції Пірсона, адже цей показник є лінійним і не враховує локальні просторові виключення, які суперечать загальній тенденції.

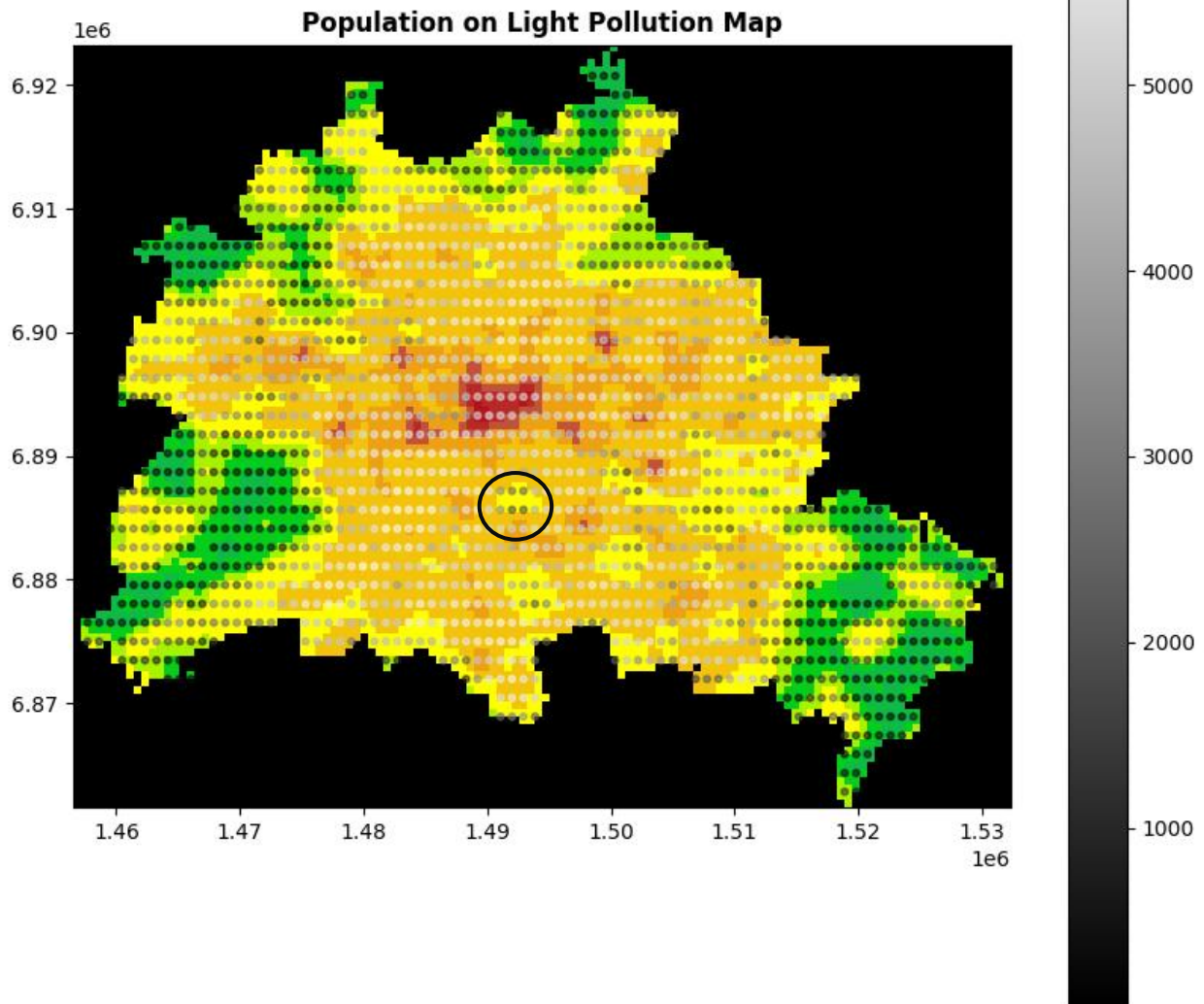


Рисунок 3.5 – Зони середнього забруднення з низьким населенням

У підсумку, саме візуальні інструменти стали першою і дуже інформативною сходинкою до формулювання основних закономірностей між населенням і рівнем штучного освітлення. Це дозволило швидко і наочно виявити зони найбільшого збігу або, навпаки, розбіжностей, які згодом можуть бути проаналізовані як винятки або підтвердження загальної тенденції.

3.2.1 Алгоритм узгодження координат і підрахунку

Перш ніж здійснювати безпосередній розрахунок кореляції між щільністю населення та інтенсивністю світлового забруднення, необхідно було забезпечити коректне узгодження просторових координат двох джерел даних. Обидва масиви — карта населення та обрізаний фрагмент супутникового знімка VIIRS — мають геопросторову природу, однак їх формат представлення та тип покриття суттєво відрізняються. Саме тому важливим етапом став процес синхронізації їхньої просторової прив'язки. Щільність населення була представлена як масив точок із заданими координатами (широта, довгота) та відповідним значенням щільності у кожній точці (кількість осіб на квадратний кілометр). Ці дані вже були приведені до меж міста Берлін та представлені у вигляді таблиці з координатами кожного осередку. Таким чином, кожен рядок у таблиці — це фактично центр одного квадратного кілометра, з числовим значенням щільності в ньому.

З іншого боку, супутникові дані VIIRS були представлені як растрове зображення, де кожен піксель мав певну координатну прив'язку та значення світлового навантаження — від 0 до 11. У цьому випадку не було масиву точкових координат, а лише поле значень, яке покривало простір.

Узгодження здійснювалося шляхом вибірки значень з растра за точками щільності населення. Для цього використовувалася функція `sample_gen()` з модуля `rasterio`, яка дозволяє отримати значення пікселя растра в конкретних географічних координатах. Вхідними даними були координати точок з таблиці населення, а виходом — значення світлового забруднення, які припадали на ці точки. Координатні системи обох шарів були приведені до єдиного стандарту — CRS EPSG:4326 або до CRS самого растра (залежно від початкової прив'язки супутникового знімка), щоб уникнути зміщень або помилок у геопросторовому проектуванні. Без цієї операції було б неможливо точно прив'язати координати населення до пікселів растру. Після зчитування світлового значення для кожної точки, ці значення були додані до основного

датафрейму як окрема колонка `light_pollution`. Потім із таблиці були вилучені всі рядки, де значення світлового забруднення було відсутнім (NaN), що могло статися, якщо точка випадала за межі растра або в зонах без зафіксованого світла.

Таким чином, на завершення цього етапу ми отримали два синхронізовані масиви однакового розміру: кожен елемент мав координати, значення щільності населення та значення світлового забруднення. Це дало змогу перейти до етапу математичного аналізу та розрахунку коефіцієнта кореляції між цими показниками.

3.2.2 Розрахунок коефіцієнта Пірсона

Після узгодження географічних координат щільності населення та інтенсивності світлового забруднення було побудовано два масиви з просторовою прив'язкою. У кожній точці — незалежно від її географічного положення — є пара значень: чисельність населення та відповідний рівень нічної освітленості, отриманий із супутникових знімків VIIRS. Завдяки цьому з'явилася можливість перевірити наявність статистичної залежності між цими показниками. Для аналізу ступеня зв'язку між змінними було використано коефіцієнт кореляції Пірсона, класичний інструмент статистичного аналізу, який дозволяє оцінити лінійну залежність між двома числовими величинами. Обчислення проводилося за допомогою функції `pearsonr()` з модуля `scipy.stats`.

Перед обчисленням з таблиці було виключено всі рядки, що містили пропущені або некоректні значення. Це важливо, адже наявність NaN значень могла б спотворити підсумковий результат. Далі два масиви — `light_pollution` (світлове забруднення) та `Z` (щільність населення) — були подані на вхід функції.

У результаті розрахунку було отримано:

Коефіцієнт кореляції Пірсона: 0.753

p-значення: 1.9e-308

Такий рівень коефіцієнта (близький до 0.8) свідчить про високу позитивну кореляцію між інтенсивністю світлового забруднення та густотою населення. Це означає, що в цілому, із зростанням чисельності населення на певній території, зростає й рівень нічного освітлення, що є логічним висновком у межах урбаністичного простору: щільно заселені райони зазвичай мають активну інфраструктуру, вуличне освітлення, транспортні вузли, комерційні площі тощо. Надзвичайно низьке р-значення (практично нульове) вказує на високу статистичну значущість результату. Інакше кажучи, вірогідність того, що така кореляція виникла випадково — майже відсутня. Це дозволяє з впевненістю стверджувати про наявність реального, не випадкового зв'язку між двома змінними.

Разом із тим важливо пам'ятати, що коефіцієнт Пірсона показує лише лінійну залежність. У деяких випадках існуючі зв'язки можуть бути нелінійними, або мати складну просторову природу, пов'язану, наприклад, із функціональними відмінностями районів (промислові, житлові, парки тощо). Візуалізація даних та аналіз залишків моделі можуть допомогти в подальшій інтерпретації результатів.

Таким чином, обчислення коефіцієнта Пірсона не лише підтвердило гіпотезу про кореляцію між щільністю населення та світловим забрудненням у межах міста Берлін, а й надало об'єктивну основу для глибшого аналізу урбаністичних процесів.

3.2.3 Розрахунок коефіцієнтів кореляції Спірмена та Кендалла

Для комплексного аналізу взаємозв'язку між щільністю населення та світловим забрудненням, окрім традиційного коефіцієнта кореляції Пірсона, було проведено розрахунок двох інших популярних статистичних показників — рангової кореляції Спірмена та коефіцієнта Кендалла.

На відміну від коефіцієнта Пірсона, який оцінює лінійну залежність між двома змінними, коефіцієнти Спірмена та Кендалла враховують монотонні (не

обов'язково лінійні) зв'язки, будуючись на ранговому порядку значень. Це дозволяє краще розуміти залежності у випадках складних або нелінійних даних, якими є просторові показники щільності населення та інтенсивності світлового забруднення.

Розрахунки проводились з використанням бібліотеки SciPy. Результати показали, що коефіцієнт Спірмена становить 0.354 при дуже низькому р-значенні ($< 10^{-51}$), що свідчить про статистично значущий монотонний зв'язок між показниками. Коефіцієнт Кендалла дорівнює 0.226 ($p < 10^{-36}$), також підтверджуючи наявність помірної кореляції.

Варто зазначити, що значення Спірмена та Кендалла менші за коефіцієнт Пірсона (0.753), що може свідчити про те, що зв'язок між щільністю населення і світловим забрудненням у межах Берліну є переважно лінійним, хоча і присутні деякі нелінійні елементи або винятки у даних.

Таким чином, аналіз кореляцій за різними методиками дозволяє отримати повнішу картину взаємозв'язку між просторовими показниками, підкреслюючи важливість комплексного підходу при роботі з геопросторовими даними.

3.2.4 Інтерпретація значення коефіцієнта

Отримані значення коефіцієнтів кореляції Пірсона, Спірмена та Кендалла дозволяють зробити низку важливих висновків про взаємозв'язок між щільністю населення та інтенсивністю світлового забруднення в межах міста Берлін.

Коефіцієнт Пірсона, який дорівнює 0.753, свідчить про досить сильний лінійний зв'язок між двома змінними. Це означає, що в районах з вищою щільністю населення зазвичай спостерігається й більш інтенсивне світлове забруднення. Такий результат відповідає очікуванням, адже густонаселені міські райони часто характеризуються підвищеним використанням штучного освітлення, рекламою, транспортом та іншими джерелами світла.

З іншого боку, коефіцієнти Спірмена (0.354) та Кендалла (0.226) є помірними за величиною, що вказує на існування певних відхилень від суто лінійної залежності. Ці рангові кореляції оцінюють монотонність зв'язку, тому їх відносно нижчі значення можуть сигналізувати про те, що взаємозв'язок не є абсолютно однорідним по всій території. Можливі локальні аномалії, такі як парки, водні об'єкти або інші зони з низьким освітленням, де щільність населення не завжди корелює із світловим забрудненням.

Також варто враховувати, що дані світлового забруднення мають меншу деталізацію (всього 11 рівнів інтенсивності), що могло вплинути на заниження рангових кореляцій. З іншого боку, дані про щільність населення мають більш широкий діапазон значень і більшу просторову дискретність, що сприяє вищому коефіцієнту Пірсона. Статистично дуже низькі р-значення для всіх трьох кореляцій свідчать про те, що виявлені зв'язки є статистично значущими і не випадковими. Це підсилює довіру до отриманих результатів та підтверджує існування реальної взаємозалежності між щільністю населення і світловим забрудненням у місті Берлін.

Узагальнюючи, можна сказати, що отримані коефіцієнти кореляції підтверджують гіпотезу про те, що зростання щільності населення супроводжується збільшенням світлового забруднення, однак ця залежність має певні просторові варіації, які слід враховувати у подальшому аналізі і моделюванні.

3.3.1 Побудова графіків для порівняння показників

Для кількісного аналізу взаємозв'язку між щільністю населення та рівнем світлового забруднення було побудовано розсіювальний графік (scatter plot). На цьому графіку по осі X відкладається значення світлового забруднення, а по осі Y — щільність населення у відповідних осередках (пікселях або зонах).

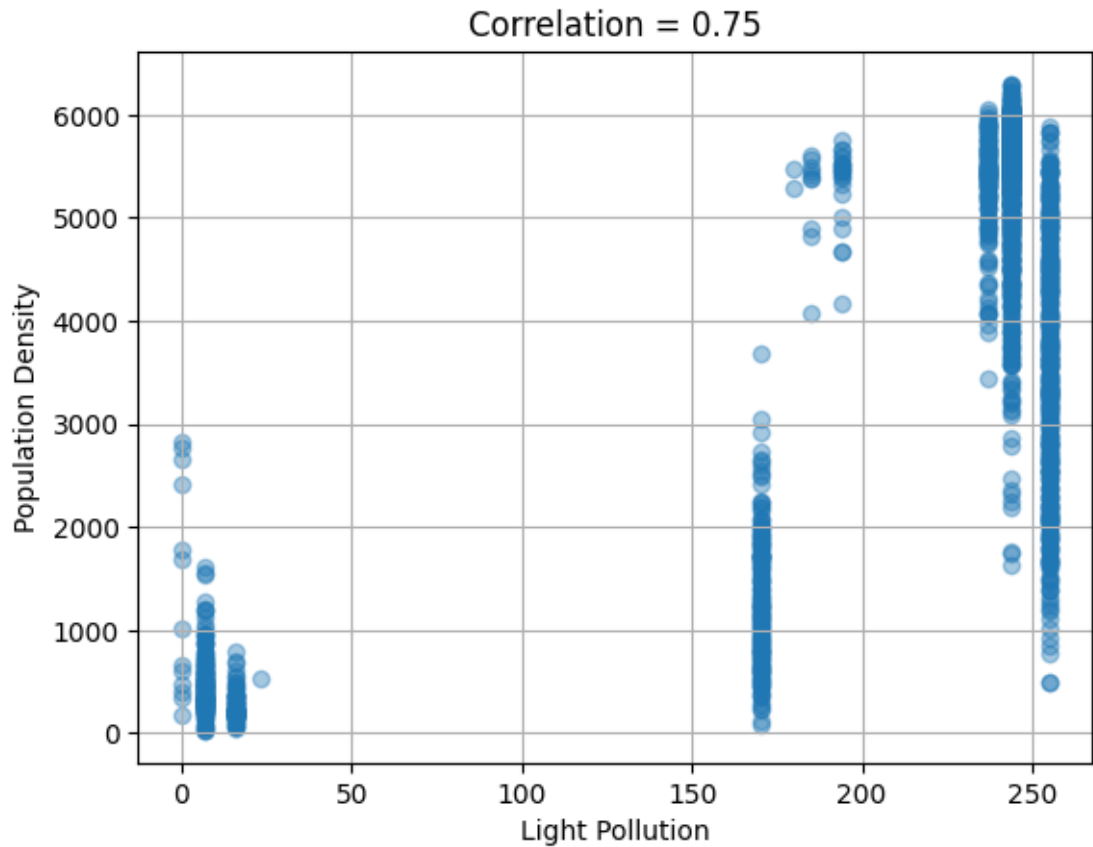


Рисунок 3.5 – Графік коефіцієнту Пірсона

У заголовку графіка наведено значення коефіцієнта Пірсона (0.75), що вказує на досить сильний позитивний лінійний зв'язок між цими двома змінними. Проте, сам графік демонструє, що ці дані мають складну структуру — з кластерами та скупченнями точок, зокрема у зонах високої освітленості.

Слід зазначити, що шкали для щільності населення (0–6000 осіб на квадратний кілометр) і світлового забруднення (0–255 одиниць цифрової яскравості, але ефективно ~0–11) суттєво відрізняються за діапазоном, що ускладнює інтерпретацію графіка. Через це більшість точок на графіку згущуються у вузькому вертикальному діапазоні. Подальша нормалізація значень або логарифмічне перетворення може допомогти краще відобразити структуру розподілу.

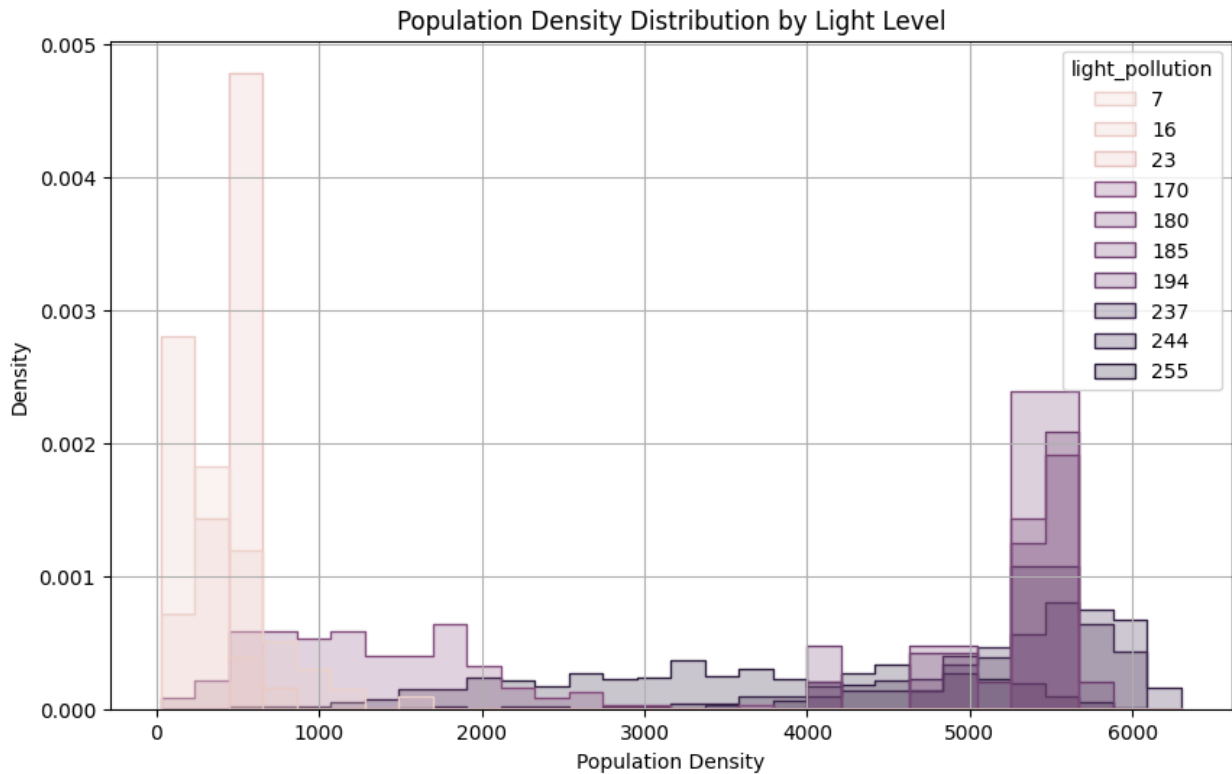


Рисунок 3.6 – Графік щільності населення від освітленості

На другому графіку представлено розподіл щільності населення залежно від рівня освітленості. Тут видно, що найвищі щільності населення спостерігаються переважно при максимальних рівнях світлового забруднення (значення 237–255). Водночас при низькому рівні освітлення (7, 16, 23) щільність має чіткий пік у районі до 1000 осіб, що підтверджує, що в темних зонах проживає менше населення. Така сегментація підтримує гіпотезу про існування функціональних зон: густозаселених і добре освітлених центрів та менш населених, слабоосвітлених периферій.

Попри деякі обмеження у візуалізації, ці графіки є наочним підтвердженням зв'язку між освітленістю території та її заселеністю. Вони також показують наявність аномальних точок — таких як освітлені, але нежитлові території (аеропорти, парки, ТЦ), що можуть знижувати значення кореляції.

3.4 Висновок до розділу 3

У цьому розділі було проведено порівняльний аналіз щільності населення та рівня світлового забруднення в межах міського простору Берліна. Використання картографічної візуалізації, розсіювального графіка та розподілів дозволило не лише оцінити кореляцію між цими показниками, але й виявити просторові патерни та потенційні аномалії. Основною виявленою закономірністю є позитивний зв'язок між щільністю населення та рівнем світлового забруднення, що підтверджується значенням коефіцієнта кореляції Пірсона (0.75). Це свідчить про те, що у густозаселених районах, як правило, спостерігається і вищий рівень нічного освітлення, що є логічним з огляду на щільність забудови, транспортну інфраструктуру та комерційну активність.

Однак цей зв'язок не є лінійним і має винятки, зумовлені особливостями міського середовища. Наприклад, було виявлено зони з високим рівнем освітлення, але низькою або відсутньою щільністю населення — як-от парки, великі транспортні вузли, аеропорти або колишні промислові зони. Типовим прикладом є Темпельхофер Фельд — великий громадський простір на місці колишнього аеропорту, який має високий рівень освітлення (світло-жовтий відтінок на карті), але повну відсутність постійного населення (чорний колір на карті щільності). Такі території суттєво впливають на розрахунок кореляції, знижуючи її значення, оскільки коефіцієнт Пірсона чутливий до викидів і передбачає лінійність.

Крім того, при інтерпретації результатів слід враховувати потенційні джерела похибки наприклад - сезонність: супутникові знімки світлового забруднення можуть бути зібрані у різні пори року, що впливає на інтенсивність нічного освітлення — наприклад, у зимовий період темна частина доби триває довше, тому рівень світіння виявляється вищим.

У підсумку, аналіз підтвердив наявність зв'язку між населенням і світловим забрудненням, але також показав, що цей зв'язок опосередкований рядом інших чинників: інфраструктурою, типом території, функціональним

зонуванням, а також фізичними та технічними умовами збору даних. Подальші дослідження можуть включати врахування типів землекористування, часу доби, буденної та святкової активності, щоб побудувати більш точну модель.

ВИСНОВКИ

У ході виконання дипломної роботи було реалізовано комплексний підхід до оцінювання просторового розподілу населення за супутниковими знімками нічного освітлення. Основну увагу зосереджено на поєднанні методів системного аналізу, математичного моделювання, статистичних методів та сучасних ГІС-технологій для вивчення взаємозв'язку між щільністю населення та рівнем світлового забруднення в межах міста Берлін.

У першому розділі здійснено критичний огляд наукових джерел, присвячених аналізу просторового розподілу населення, методам обробки геоданих, кластеризації, алгоритмам виявлення кореляцій, а також аналізу урбанізованих територій на основі супутникових знімків. Окрему увагу приділено огляду сучасних підходів, які використовуються в міжнародних дослідженнях з подібної тематики.

У другому розділі наведено процес підготовки вхідних даних — зокрема, дані про щільність населення та супутникові знімки VIIRS, що характеризують рівень нічного освітлення. Виконано приведення масивів до єдиного формату, обрізку до адміністративних меж міста Берлін, формування узгоджених наборів просторових даних. Здійснено математичне обґрунтування вибору методу кореляційного аналізу, описано програмне забезпечення та інструменти, які використовувались для реалізації дослідження.

У третьому розділі представлено результати практичного застосування описаних методик — виконано візуалізацію просторових масивів у вигляді картографічних зображень, проведено розрахунок коефіцієнтів кореляції Пірсона, Спірмена та Кендалла для перевірки гіпотези про наявність зв'язку між досліджуваними показниками. Побудовано графіки залежностей та карти просторової кореляції, які підтверджують статистично значущий зв'язок між щільністю населення та інтенсивністю нічного освітлення.

Узагальнюючи результати дослідження, можна зробити висновок, що використання супутникових знімків нічного освітлення дозволяє ефективно оцінювати щільність населення в урбанізованих регіонах. Запропонований підхід має високу практичну цінність для оперативного аналізу, може бути масштабований на інші міста та регіони і застосований у сфері містобудування, демографічних досліджень та моніторингу урбанізаційних процесів.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Kyba C.C.M. et al. Artificially lit surface of Earth at night increasing in radiance and extent // *Science Advances*. – 2017. – Vol. 3(11). – e1701528. – DOI: 10.1126/sciadv.1701528.
2. Román M.O. et al. VIIRS night-time lights: New developments and applications for environmental and socio-economic studies // *Remote Sensing of Environment*. – 2018. – Vol. 210. – P. 113–132. – <https://doi.org/10.1016/j.rse.2018.03.015>.
3. Baugh K. et al. Monitoring of urban dynamics via VIIRS Day/Night Band radiometry // *Remote Sensing*. – 2013. – Vol. 5(9). – P. 4961–4978.
4. Pearson K. Note on regression and inheritance in the case of two parents // *Proceedings of the Royal Society of London*. – 1895. – Vol. 58. – P. 240–242.
5. Python Software Foundation. Python 3 Documentation [Електронний ресурс]. – 2024. – Режим доступу: <https://docs.python.org/3/>.
6. McKinney W. Data Structures for Statistical Computing in Python // *Proceedings of the 9th Python in Science Conference*. – 2010. – P. 51–56.
7. Virtanen P. et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python // *Nature Methods*. – 2020. – Vol. 17. – P. 261–272.
8. Hunter J.D. Matplotlib: A 2D Graphics Environment // *Computing in Science & Engineering*. – 2007. – Vol. 9(3). – P. 90–95.
9. OpenStreetMap contributors. Population density data (Berlin area) [Електронний ресурс]. – 2024. – Режим доступу: <https://www.openstreetmap.org/>.
10. Statistisches Bundesamt (Destatis). Statistical data on Berlin population density [Електронний ресурс]. – 2023. – Режим доступу: <https://www.destatis.de/>.
11. NOAA National Centers for Environmental Information. VIIRS Nighttime Lights Annual Composites [Електронний ресурс]. – 2024. – Режим доступу: <https://ngdc.noaa.gov/eog/viirs/>.

12. Соколова І.А. Методи аналізу просторових даних у геоінформаційних системах. – Київ: Наукова думка, 2021. – 256 с.
13. Коваленко В.П. Основи обробки статистичних даних у Python для природничих наук. – Харків: ХНУ, 2020. – 198 с.
14. Light Pollution Map. VIIRS 2015 Overlay [Електронний ресурс]. – 2024. – Режим доступу: <https://www.lightpollutionmap.info/>.
15. Elvidge C.D. et al. VIIRS night-time lights // International Journal of Remote Sensing. – 2017. – Vol. 38(21). – P. 5860–5879.
16. Ma T., Zhou Y., Pei T., Haynie S., Fan J. Night-time light derived estimation of spatio-temporal characteristics of urbanization dynamics using DMSP/OLS satellite data // Remote Sensing of Environment. – 2014. – Vol. 158. – P. 453–464.
17. Li X., Zhou Y. Urban mapping using DMSP/OLS stable night-time light: A review // International Journal of Remote Sensing. – 2017. – Vol. 38(21). – P. 6030–6046.
18. Zhang Q., Schaaf C.B., Seto K.C. The vegetation adjusted NTL urban index: A new approach to reduce saturation and increase variation in nighttime radiance data // Remote Sensing of Environment. – 2013. – Vol. 129. – P. 32–41.
19. UN-Habitat. World Cities Report 2020: The Value of Sustainable Urbanization. – 2020.
20. Esri & ArcGIS Blog. Working with Population Density Data and Remote Sensing in GIS. – 2022.
21. Кваліфікаційна робота бакалавра [Електронний ресурс] : методичні рекомендації для здобувачів ступеня бакалавра освітньо-професійної програми «Системний аналіз» зі спеціальності 124 Системний аналіз / уклад.: Т.А. Желдак, Т.В. Хом'як, А.В. Малієнко ; М-во освіти і науки України, Нац. техн. ун-т «Дніпровська політехніка». – Дніпро : НТУ «ДП», 2025. – 32 с.

22. Коряшкіна, Л., Малієнко, А., Станіна, О., Шевченко, Ю., Кодола, Я. (2025). Системний аналіз та оптимальний вибір комплексу заходів для підвищення безпеки на підприємстві. *Information Technology: Computer Science, Software Engineering and Cyber Security*, 72–80, doi: <https://doi.org/10.32782/IT/2025-2-7>
23. Кваліфікаційна робота бакалавра [Електронний ресурс] : методичні рекомендації для здобувачів ступеня бакалавра освітньо-професійної програми «Системний аналіз» зі спеціальності 124 Системний аналіз / уклад.: Т. А. Желдак, Т. В. Хом'як, А. В. Малієнко ; М-во освіти і науки України, Нац. техн. ун-т «Дніпровська політехніка». – Дніпро : НТУ «ДП», 2025. – 32 с. url: <https://ir.nmu.org.ua/handle/123456789/170863>
24. Коряшкіна, Л. С., Станіна, О. Д., & Шевченко, Ю. О. (2024). Практикум з диференційних рівнянь.