Toward Privacy-Aware Environmental Monitoring of CO_2 and Air Pollutants in Southern Italy

Lelio Campanile(✉), Luigi Piero Di Bonito, Fiammetta Marulli, Antonio Balzanella, and Rosanna Verde

Department of Mathematics and Physics, Università degli Studi della Campania "L. Vanvitelli", Caserta, Italy
lelio.campanile@unicampania.it

Abstract. The increasing levels of CO_2 and air pollutants represent a major challenge to environmental sustainability and public health, particularly in regions characterized by complex geographic and socio-economic dynamics. This work proposes a study focused on the Southern Italy regions, where environmental vulnerabilities are displayed, along with a limited availability of high-granularity data. The main aim of this work is to build and provide a comprehensive and detailed dataset tailored to the region's unique needs, by leveraging datasets from EDGAR for greenhouse gases and air pollutants, integrated with demographic and territorial morphology data from ISTAT. The creation of composite indicators to monitor trends in emissions and pollution on a fine spatial scale is supported by the data set. These indicators enable initial insight into spatial disparities in pollutant concentrations, offering valuable data to inform targeted policy interventions. The work provided a foundation for next analytical studies, integrating different datasets and highlighting the potential for complex spatiotemporal analysis. The study provides a robust dataset and preliminary insights, enhancing the understanding of environmental dynamics in Southern Italy. Subsequent efforts will focus on extending this methodology to more extensive geographic contexts and incorporating real-time data for adaptive monitoring. The proposed framework also lays the groundwork for privacy-aware environmental monitoring solutions, enabling future integration with edge and IoT-based architectures while addressing privacy and data protection concerns.

Keywords: Greenhouses gas · Data aggregation · Air quality Monitoring · Data Analysis · IoT · Privacy-aware analytics

1 Introduction

The increasing concentration of CO_2 and air pollutants represents one of the most pressing challenges to environmental sustainability and public health worldwide. Anthropogenic activities, such as industrial production, urbanization, and vehicular emissions, have significantly accelerated the accumulation of greenhouse gases and other pollutants in the atmosphere, resulting in accelerated

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026
O. Gervasi et al. (Eds.): ICCSA 2025 Workshops, LNCS 15893, pp. 317–333, 2026.
https://doi.org/10.1007/978-3-031-97645-2_21

climate change and negative impacts on ecosystems. Beyond global concerns, the localized effects of pollution are profound, especially in regions with unique geographic and socio-economic characteristics. Southern Italy exemplifies such a region, where environmental vulnerabilities intersect with limited access to detailed and usable data. Although significant advancements in environmental monitoring have been made, the granularity of available datasets remains a critical bottleneck. Existing data on CO_2 emissions and air pollutants often lacks the spatial and temporal resolution needed to capture variations at municipal or regional levels. This limitation hinders the development of targeted interventions and complicates efforts to evaluate the effectiveness of mitigation strategies. Furthermore, the interplay between demographic factors, territorial morphology, and pollution intensities is inadequately represented in many studies, leaving critical gaps in understanding the socio-environmental dynamics of vulnerable areas like Southern Italy. To address these challenges, this study focuses on constructing a high-quality, comprehensive dataset tailored to the unique needs of the specific analysis target, as Southern Italy. The integration of greenhouse gas and air pollutant data from the EDGAR database with demographic and territorial morphology data from ISTAT (National Institute of Statitics for Italy) is intended to address the data deficit, thereby establishing a foundation for advanced environmental analyses. The dataset has been designed to facilitate the development of composite and evolutionary indicators that reflect pollution trends over time and space. These indicators are considered to be crucial for identifying spatial disparities in pollutant concentrations and their potential health impacts, thereby supporting informed decision-making for policymakers and stakeholders.

In addition to addressing environmental concerns, the increasing reliance on high-resolution and real-time environmental data collected through distributed IoT infrastructures raises important privacy considerations. Although the datasets employed in this study are collected from publicly available sources and do not include directly identifiable information, the integration of granular spatial, demographic, and temporal data may still pose risks related to re-identification and location-based profiling. As environmental monitoring systems evolve toward edge and cloud-based architectures with real-time data acquisition, the adoption of privacy-aware data processing methods becomes essential. This study anticipates these challenges and proposes a methodological foundation that can be extended with privacy-preserving techniques in future implementations.

A distinctive feature of this work is represented by the emphasis on the process adopted for integrating diverse and heterogeneous datasets to capture the multifaceted nature of environmental and demographic phenomena. Furthermore, this integration provides a unique opportunity to construct fine-grained indicators that account for spatial dependencies and temporal variations, offering deeper insights into the dynamics of pollution in Southern Italy. Preliminary analysis of the integrated dataset reveals significant heterogeneity in pollution levels across the region, underscoring the importance of localized interventions. For instance, coastal areas near industrial hubs and urban centers exhibit

higher concentrations of pollutants. These findings underscore the value of high-resolution data in identifying critical areas for intervention and in shaping region-specific policies to mitigate environmental and health impacts. In addition to addressing immediate challenges, this study establishes the foundation for future research that explores advanced methodologies for spatiotemporal analysis. By providing a robust data set and demonstrating its potential through preliminary findings, our objective is to inspire further investigations that incorporate real-time data streams and adaptive monitoring systems. Such advancements would enhance the granularity of environmental monitoring and enable dynamic responses to emerging threats, contributing to the long-term sustainability of vulnerable regions such as Southern Italy. The remainder of this paper is organized as follows. Section 2 provides an overview of related work, highlighting existing approaches to environmental monitoring and identifying gaps addressed by this study. Section 3 details the integrated methodology and data sources, describing the processes of data acquisition, pre-processing, and indicator development. The results of the preliminary dataset analysis are presented in Sect. 4. Section 5 discusses the results and evaluates the proposed indicators. Section 6 explores the implications of the findings and the limitations of the study. Finally, Sect. 7 concludes the paper by summarizing the main contributions and outlining potential applications of this research and describing the future works.

2 Related Works

Monitoring CO_2 emissions and air pollutants is a cornerstone of environmental research and policy development. Janssens-Maenhout et al. [12] propose an integrated system that combines satellite and in situ measurements to provide near-real-time updates on anthropogenic CO_2 emissions, facilitating global monitoring efforts. Similarly, Liu et al. [16] demonstrate the impact of near-real-time monitoring during the COVID-19 pandemic, revealing significant reductions in global CO_2 emissions, in [5] a security aspect is analyzed. The interplay between greenhouse gases and air pollutants is further explored by Lin et al. [15], who provide insights into their correlated emissions and implications for synergistic mitigation strategies. Cromar et al. [7] emphasize the importance of accurate air pollution monitoring networks, particularly for health research and patient care, where in [6] a specific southern Italy Campania region was investigated. Furthermore, Manisaldis et al. [17] offer a comprehensive review of the environmental and health impacts of air pollution, underscoring the critical need for robust monitoring systems.

In recent years, the integration of data and predictive models has emerged as a pivotal approach to addressing environmental challenges. High-granularity datasets, when coupled with advanced modeling techniques, offer unprecedented insights into pollutant dynamics and their impacts on ecosystems and human health. Govea et al. [9] discuss the integration of air quality and noise data in urban environments, demonstrating the potential of sensor networks for localized monitoring. Similarly, Superfund Research Program [10] highlights the impor-

tance of interoperability and data reuse to tackle complex environmental challenges effectively. These advancements are further complemented by efforts to estimate air pollution concentrations at a continental scale, as presented by Thomas et al. [19], who utilized data integration techniques to enhance resolution and accuracy.

The application of semantic middleware architectures for heterogeneous environmental data sources, explored by Akanbi and Masinde [1], underscores the necessity of integrating diverse datasets to achieve comprehensive environmental assessments. In the same vein, [8] propose leveraging grid computing frameworks to process and analyze large-scale environmental data, emphasizing the scalability of such approaches for global applications.

Indicators and predictive models play a crucial role in monitoring and forecasting environmental changes. In [2] the authors delve into artificial intelligence applications for environmental monitoring, showcasing the use of predictive models to evaluate air and water quality. Hybrid models combining empirical decomposition and neural networks, as discussed in [20], have been employed to improve the accuracy of water quality predictions. The integration of environmental data into models for vector-borne disease prediction, as demonstrated by [14], highlights the versatility of these methods across different domains.

Machine learning-based approaches have gained traction in advancing sustainability efforts. Mazumder et al. [18] propose a novel machine learning model for holistic environmental health monitoring, emphasizing the role of composite indicators in understanding complex interactions. These innovations pave the way for robust analytical frameworks capable of addressing multifaceted environmental issues.

The spatiotemporal analysis of environmental phenomena is integral to understanding pollutant behaviors over time and space. The utility of wireless sensor networks in conducting spatiotemporal analyses is evident in the work of [21], which highlights their role in monitoring air quality dynamics. A novel framework for spatiotemporal prediction using deep learning, introduced by [3], underscores the potential of advanced algorithms to forecast environmental data with high precision. Reviews on spatiotemporal event detection, such as [22], provide a comprehensive understanding of methodologies and applications in environmental monitoring.

Finally, the development of spatiotemporal LSTM models, as explored in [11], showcases the adaptability of these techniques to predict across multiple scales, addressing both temporal and spatial complexities.

These studies collectively underscore the advancements and challenges in integrating data, indicators, and spatiotemporal methodologies for environmental monitoring. By leveraging these approaches, researchers can develop comprehensive frameworks to inform policy interventions and foster sustainable development. Despite these advances, few existing works integrate high-resolution environmental data with socio-demographic variables in a privacy-aware manner. This study proposes to fill this gap.

3 Data Integration and Methodology

3.1 Data Sources

This analysis utilizes two principal data sources: the EDGAR database, which supplies comprehensive emissions data for greenhouse gases and air pollutants, and ISTAT datasets, which present demographic and territorial morphology information. The amalgamation of these information facilitates a more thorough environmental evaluation, permitting elevated geographical resolution with a specific emphasis on Southern Italy.

The Emissions Database for Global Atmospheric Research (EDGAR) is extensively utilized for tracking human emissions. This study employs the data retrieved from EDGAR v8.1 and 2024 GHG datasets, encompassing CO_2 emissions from 1970 to 2023, alongside the EDGAR v8.1_AP dataset, which offers comprehensive data on air pollutants including carbon monoxide (CO), ammonia (NH_3), nitrogen oxides (NO_x), particulate matter (PM_10, $PM_2.5$), and sulfur dioxide (SO_2) from 1970 to 2022. The emissions data is presented as annual gridmaps, providing significant insights into long-term trends and facilitating spatial disaggregation for localized research.

The ISTAT databases, in addition to emissions statistics, offer insights into population density, urbanization rates, land utilization, and socio-economic variables for all Italian municipalities.

This research concentrates on Southern Italy, in accordance with the project's aim to tackle environmental risks unique to this area. The region is marked by substantial pollution sources, elevated population density in metropolitan areas, and reliance on industrial and agricultural practices that affect CO_2 and air pollutant emissions. The scarcity of high-granularity data in this setting necessitates the utilization of sophisticated approaches for spatial disaggregation and data augmentation.

3.2 Data Processing and Integration

Integrating datasets from several sources necessitates a comprehensive preparation pipeline to guarantee consistency, correctness, and compatibility across varied spatial and temporal scales. The preliminary phase of data processing entails meticulous data cleansing and standardization, encompassing the resolution of inconsistencies in variable nomenclature, the imputation or elimination of absent values, and the synchronization of datasets to a unified temporal framework. Due to the extensive temporal range of the EDGAR datasets, meticulous alignment with ISTAT data is essential for preserving comparability and facilitating significant cross-referencing between emissions and demographic factors. This harmonization step improves data coherence and supports subsequent analytical procedures by maintaining temporal integrity across various sources.

A significant obstacle in integrating emissions data with demographic databases is the alignment of varying spatial resolutions. Emissions data are initially presented in a grid format, whereas demographic data adhere to administrative borders. Emissions values are correlated with ISTAT-defined territorial

units by GIS-based interpolation techniques, facilitating a cohesive integration of pollution and population data. This procedure guarantees that the dataset maintains a high degree of specificity while facilitating comparison analysis across various environmental indicators.

To improve spatial granularity, we utilized statistical downscaling methods that incorporate auxiliary factors such road network density, industrial activity levels, and weather data. These characteristics are associated with emissions and function as proxies to enhance spatial allocation. Furthermore, temporal interpolation techniques were utilized to address data deficiencies and provide uninterrupted time series, so enabling trend analysis across several decades. This method guarantees that data integration preserves historical consistency and geographic precision, facilitating comprehensive environmental evaluations.

In addition to integration, the processed datasets underwent validation methods to evaluate their compatibility with current environmental reports and official statistics. Statistical analyses of aggregated emissions data and national inventories were performed to confirm consistency with overarching trends. This validation process is essential for verifying the dependability of the produced indicators and ensuring their relevance in environmental policy discourse.

3.3 Methodological Framework for Environmental Indicators

This study employs a methodological framework to develop spatiotemporal environmental indicators, including the diversity of emission sources and the intricacies of demographic dynamics. These indicators function as analytical instruments to track pollution trends, evaluate environmental concerns, and furnish data-driven insights for policymakers. High-granularity indicators are essential for recognizing localized pollution patterns and their temporal changes, enabling a more focused approach to mitigation activities.

The development of environmental indicators adheres to a systematic technique. Composite indicators are established by amalgamating many factors, such as emissions statistics, population density, and land use patterns, to provide synthetic metrics that signify pollution exposure levels. Moreover, evolutionary indicators are intended to identify temporal trends, facilitating the detection of rising or falling pollution levels over time. Advanced interpolation techniques are utilized to estimate emissions at a more refined spatial resolution than the original records offer, hence improving the dataset's applicability for environmental monitoring.

This paper introduces multi-source data fusion techniques, which improve the granularity and interpretability of pollution indicators. By integrating emissions data with supplementary datasets—such as land cover information, traffic flow statistics, and meteorological observations—we develop enhanced indicators that yield more insights into pollution dynamics. These indicators are classified into absolute and relative measurements, facilitating comparison evaluations across geographies and various timeframes.

A crucial component of the methodological framework is the application of geostatistical modeling to enhance spatial estimations. The different spatial interpolation techniques were utilized to predict pollution levels in regions with limited observational data. These models enhance the precision of emissions estimates and improve the resolution of exposure assessment by utilizing spatial dependencies. Furthermore, Bayesian hierarchical modeling was utilized to integrate uncertainty estimation in the generated indicators, yielding confidence intervals that inform decision-making processes.

This work primarily emphasizes dataset construction and indicator formulation; future research will investigate the deployment of machine learning models to improve prediction capabilities. Possible avenues encompass the identification of emission hotspots through historical trends, the examination of significant environmental and socio-economic factors affecting pollution levels, and the creation of adaptive monitoring systems for real-time policy guidance. These developments would enhance the methodological framework and facilitate data-driven decision-making in environmental management.

This study's methodology establishes a basis for future progress in environmental monitoring. This research enhances the understanding of pollution dynamics in Southern Italy by integrating several data sources, utilizing new spatial approaches, and creating enhanced indicators. The findings establish a solid foundation for focused interventions and policy formulation, guaranteeing that environmental sustainability is prioritized in regional planning initiatives.

In addiction, this framework could evolve toward real-time monitoring using edge ad IoT-based systems, where data privacy and security becomes crucial. In future implementation, some privacy preserving and security improving techniques, such as differential privacy and spacial generalization, could be integrated

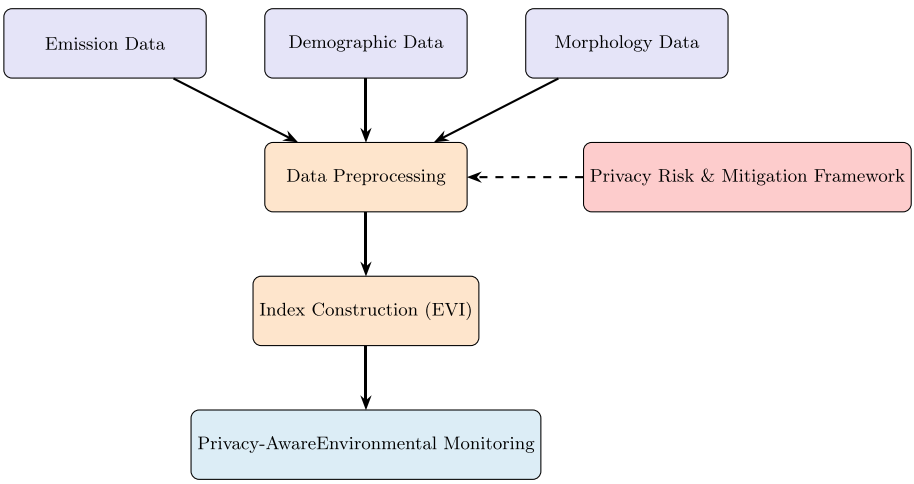

Fig. 1. Methodological workflow: from data acquisition to privacy-aware indicator computation.

to protect sensitive demographic and geographic patterns. These enhancements would mitigate risks of misuse of environmental data in sensitive areas. Figure 1 illustrates the methodological pipeline adopted for dataset integration, indicator construction, and analysis.

4 Preliminary Analysis

This initial analysis examines the integrated dataset, verifies its consistency, and derives essential information regarding the spatial distribution of air pollutants and greenhouse gas emissions in Southern Italy. The research amalgamates many data sources, encompassing emissions data from the EDGAR database, demographic and geographic information from ISTAT, and land cover attributes. This method facilitates a detailed environmental evaluation customized to the region's own geographical and socio-economic characteristics.

Southern Italy exhibits a diverse distribution of urbanized regions, rural areas, industrial hubs, and natural landscapes. The ISTAT dataset offers critical demographic and territorial morphology information, encompassing population density, urbanization rates, and land use attributes for each municipality. This dataset comprises essential properties such administrative codes, surface area, population statistics from 2021 and 2022, elevation data, and geographic coordinates. These criteria provide a thorough examination of socio-economic impacts on environmental conditions.

The land cover collection also includes essential information on surface classifications within Italian municipalities. It delineates manmade surfaces, wooded regions, shrublands, herbaceous flora, and aquatic bodies. These factors are essential for assessing environmental quality, as they affect the dispersion and absorption of air pollutants. Comprehending the spread of plant cover facilitates the evaluation of its mitigating impacts on CO_2 and pollution concentrations.

The EDGAR dataset offers comprehensive data on worldwide greenhouse gas and air pollutant emissions, provided at a high-resolution grid. It encompasses emissions of CO_2, NO_x, SO_2, CO, NH_3, and particulate matter, classified by sector and region. The dataset documents emissions in tons annually each coordinate grid point, providing insights into temporal and spatial trends. Through the filtration and extraction of pertinent data for Southern Italy, we rebuilt region-specific emission patterns.

We utilized GIS-based visualization techniques to study the spatial distribution of contaminants, creating maps to identify emission hotspots and spatial trends. The initial results reveal considerable discrepancies among various regions. Elevated levels of NO_x and CO are detected in urban and industrial areas, especially in Naples, Palermo, and Bari. In contrast, rural and poorly populated areas, such as inland Calabria and certain sections of Sardinia, demonstrate reduced pollution levels. This variation indicates a robust correlation among population density, economic activities, and emission levels.

To enhance comprehension of the relationship between socio-economic characteristics and air pollution, we conducted correlation studies. The findings demonstrate a robust positive association between population density and NO_x and CO emissions, validating the anticipated influence of urbanization on pollution levels. The proportion of artificial surfaces in a specific location correlates with elevated SO_2 and PM concentrations, underscoring the impact of industrial and transportation infrastructure on pollution levels. In contrast, regions with increased vegetation density generally have reduced concentrations of CO_2 and NO_x, indicating a possible ameliorative impact of green spaces on air quality.

The temporal examination of emissions, utilizing historical data from EDGAR, indicates significant trends. A significant fall in CO and SO_2 emissions was noted in 2020, corresponding with the decrease in industrial and transportation operations due to the COVID-19 pandemic. Nevertheless, emissions have subsequently increased, with CO_2 concentrations persisting in high-traffic areas, indicative of continued urban development and economic resurgence. Moreover, anomaly detection techniques revealed isolated emission surges in industrial areas, necessitating further examination to evaluate their environmental consequences.

The amalgamation of these datasets establishes a solid basis for environmental evaluation in Southern Italy. Linking emission levels with socio-economic and land cover data enables the construction of targeted environmental indicators and the identification of important places for intervention. The variations in pollution levels among several geographical regions highlight the necessity of tailored policy interventions. Subsequent study will concentrate on enhancing geographical models, integrating real-time monitoring systems, and utilizing machine learning methodologies for predictive environmental evaluations. These initiatives will augment the potential to formulate adaptive solutions for reducing pollution and enhancing sustainability in the region. The subsequent part will provide a comprehensive quantitative assessment of these initial findings utilizing sophisticated statistical methods and predictive modeling.

5 Results and Evaluation

This part provides a thorough study of the updated dataset, including methodological enhancements and significant findings regarding emissions distribution, socio-environmental correlations, carbon balance, and environmental fragility. By consolidating municipal-level data to align with emissions grid cells, we assured that demographic, land cover, and economic activity indicators could be effectively compared with emissions data. This enabled us to enhance our comprehension of the spatial and statistical correlations between CO_2 emissions and the wider environmental context of Southern Italy.

5.1 Dataset Refinement

The initial dataset posed a challenge: CO_2 emissions were reported at a grid resolution rounded to one decimal place, whereas municipal demographic and territorial data maintained exact geographic coordinates. Municipal data were consolidated inside their respective emissions grid cells to reconcile these datasets. This entailed aggregating the total population across all municipalities within a specified emissions coordinate.

- Calculating the total land area to precisely depict the magnitude of each emissions grid cell.
- Calculating population density by dividing the total population by the entire land area.
- Compiling land cover variables (tree cover, shrublands, herbaceous plants, artificial surfaces) to yield a significant depiction of the environmental composition of each grid cell.
- Computing carbon balance as the difference between total CO_2 emissions and the estimated carbon absorption capacity of vegetation within each grid cell.
- Classifying regions as either 'carbon deficit' or 'carbon surplus' based on whether their emissions exceeded or were offset by their natural absorption potential.

An extract of this dataset is shown in the following Table 1:

Table 1. Sample of Environmental Metrics per Grid Cell

Grid Cell	CO_2 Emissions (tons)	Carbon Balance	Carbon Status
36.7N, 14.8E	211047.42	−19685.82	Surplus
36.8N, 14.5E	13213.86	36928.49	Deficit
36.8N, 14.9E	13577.91	−5672.14	Surplus
37.0N, 15.2E	59084.27	43972.58	Deficit

This approach enables direct analysis of emissions data in conjunction with socio-economic and land use factors, hence assuring more trustworthy statistical interpretations.

5.2 Spatial Distribution of Emissions and Statistical Correlations

Our spatial study indicates that CO_2 emissions are primarily concentrated in urban and industrial regions, with Naples, Palermo, and Bari identified as major hotspots. This pattern conforms to expectations due to the elevated population density, industrial operations, and extensive transportation networks typical of these areas. In contrast, rural regions in Calabria and inland Sardinia demonstrate much reduced emissions, underscoring the idea that economic and

infrastructure activity, rather than mere population density, are the principal contributors to greenhouse gas output.

To investigate these associations further, we conducted correlation studies and regression modeling. The correlation heatmap in Fig. 2 demonstrates that CO_2 emissions exhibit weak correlations with population size, land area, population density, and land cover attributes.



Fig. 2. CO_2 Emissions vs Land Cover and Population

A regression analysis of CO_2 emissions and tree cover, depicted in Fig. 3, produced a R^2 value of 0.000003 (p-value = 0.96), signifying an absence of significant association between forest coverage and emissions. The correlation between CO_2 emissions and population density, shown in Fig. 4, yielded a R^2 of 0.00022 (p-value = 0.66), indicating that urbanization alone does not directly influence emission levels. Ultimately, land area demonstrated an unreliable correlation, yielding a R^2 of 0.000094 (p-value = 0.78). These findings underscore that industrial infrastructure, transportation systems, and energy generation are the principal sources of emissions, rather than mere settlement patterns or vegetation cover.

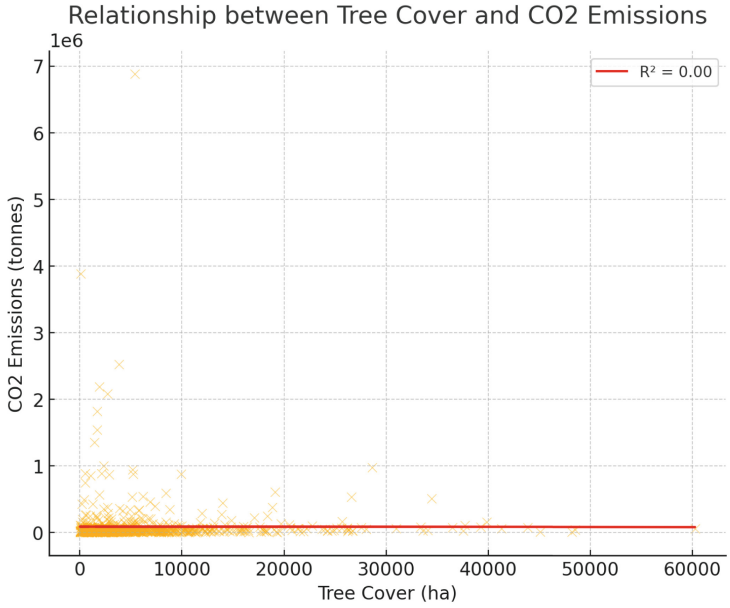

Fig. 3. Relationship between Tree Cover and CO_2 Emissions

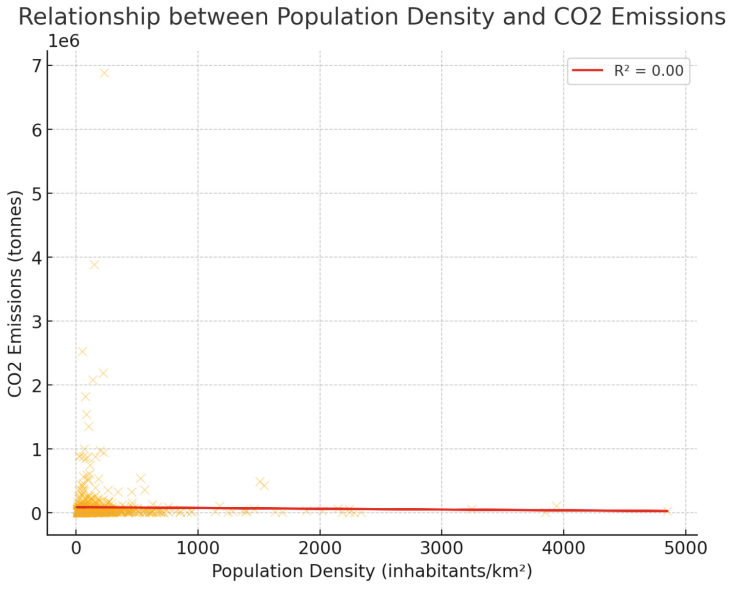

Fig. 4. Relationship between Population Density and CO_2 Emissions

5.3 Ecological Susceptibility Index

To assess sustainability, we calculated a carbon balance metric by evaluating the potential CO_2 sequestration of area flora. Using conventional estimates per hectare of carbon sequestration [4,13]:

- Forests sequester roughly 10 tons of CO_2 per hectare annually.
- Shrublands sequester approximately 3 tons per hectare annually.
- Grasslands sequester approximately 1 ton per hectare annually.

By aggregating the absorption capacity within each emission grid cell and juxtaposing them with documented emissions, we discerned areas that function under a carbon deficit (where emissions significantly surpass the natural sequestration potential) and carbon surplus zones where vegetation can mitigate emissions. The majority of regions have a shortfall highlighting the necessity for enhanced carbon management measures.

Furthermore, we developed an Environmental Vulnerability Index (EVI) to quantify the environmental stress levels experienced by each region. The EVI was calculated by integrating three primary factors: CO_2 emissions intensity, urbanization levels, and population density. All variables were normalized on a 0–1 scale to facilitate comparability, and the index was calculated as the weighted mean of these indicators. An elevated EVI score signifies increased exposure to environmental stressors, characterized by high emissions levels, dense urbanization, and significant population concentrations. This index facilitates an objective assessment of environmental vulnerability, identifying regions in need of immediate intervention. Significant EVI values were recorded in major metropolitan hubs and industrial centers, highlighting the unequal burden these regions experience regarding air pollution and climate effects.

The Carbon Balance metric represents the net difference between CO_2 emissions and the estimated carbon absorption capacity of vegetation in a given emissions grid cell. A positive carbon balance indicates that emissions exceed the sequestration potential, placing the region in a carbon deficit status. Conversely, a negative carbon balance suggests that local vegetation absorbs more CO_2 than is emitted, classifying the area as a carbon surplus region.

The Carbon Status classification provides a clear distinction between areas that contribute significantly to net CO_2 increases (deficit regions) and those that may act as potential carbon sinks (surplus regions). This classification enables targeted environmental policies by identifying areas where emissions reduction or reforestation efforts would be most effective.

These findings underscore the immediate need for specific initiatives to reduce emissions, especially in industrial areas and on transportation routes. Conventional strategies such as replanting may yield minimal advantages due to the tenuous correlation between vegetation cover and emission levels. More effective strategies may encompass limits on industrial emissions, sustainable urban mobility initiatives, and incentives for the adoption of renewable energy.

The subsequent study should focus on amalgamating real-time air quality surveillance with these results, utilizing machine learning algorithms to construct

prediction models for adaptive environmental policy modifications. This study establishes a foundation for evidence-based climate action and sustainable urban planning in Southern Italy by enhancing our comprehension of emissions patterns and their socio-environmental causes.

6 Discussion

The results of this study provide important insights into the spatial distribution and determinants of CO_2 emissions in southern Italy. The results show that emissions are mainly concentrated in industrialized and urban areas, with minimal correlation with population density and land cover characteristics. This suggests that industrial activities and transport infrastructure are the main sources of CO_2 emissions, rather than human settlement patterns or vegetation distribution.

The Environmental Vulnerability Index (EVI) highlights the disparities between emission sources and population exposure. The high-risk areas identified by the EVI align with metropolitan areas and industrial corridors, highlighting the need for targeted policies that focus on emission control, urban air quality improvement, and regulation of industrial activities. Policymakers should prioritize interventions targeting sectors with significant impact, including improvements in industrial energy efficiency, development of low-carbon transport networks, and rigorous monitoring of emissions in urban areas.

While the dataset integration and aggregation methods are robust, it is important to recognize their limitations. The application of grid-approximated emissions data introduces spatial uncertainty due to the aggregation of community-level data to fit the emissions grid framework. Future research should investigate higher resolution emissions data to improve spatial analyses and increase the accuracy of observed relationships. Another limitation is the omission of real-time emission variations due to seasonal and daily variations in industrial activity and traffic emissions. The integration of real-time monitoring data from IoT sensors with satellite-based remote sensing can improve the understanding of emission patterns in a dynamic and responsive manner.

Simply increasing forest cover is an insufficient countermeasure to offset high-emitting activities, especially in areas of high population density and industrialization. Interventions such as carbon pricing mechanisms, regulatory frameworks for industrial pollution, and investments in renewable energy should be implemented to achieve significant reductions in CO_2 emissions. Urban areas identified as high-risk zones in the EVI should be prioritized for initiatives to improve air quality, including the expansion of public transportation, the creation of pedestrian zones, and the development of green infrastructure. Policymakers should use these findings to set emission reduction targets that are spatially optimized to address the most pressing areas of concern.

This discussion highlights the importance of data-driven policy interventions in mitigating CO_2 emissions and enhancing environmental resilience. Thus, this study provides a basis for formulating targeted sustainability strategies by identifying high-risk areas and key emission drivers. Future research should prioritize

the integration of real-time emissions tracking, sophisticated machine learning models, and policy impact assessments to enhance mitigation strategies. Advancing these methods enhances our ability to develop resilient, low-carbon urban environments that are better prepared to meet the climate challenges of the coming decades.

7 Conclusions and Future Works

This study highlights the importance of implementing data-driven environmental policies that focus on reducing emissions at the source, offering a comprehensive evaluation of CO_2 emissions in Southern Italy, combining high-resolution emissions data alongside demographic, land use, and environmental indicators.

The spatial accuracy of emissions assessments and their correlation with socio-economic factors has been enhanced by structuring and refining the dataset to align municipal data with emissions grid cells. Preliminary findings indicate that emissions are predominantly affected by industrial activities and transportation networks, rather than by population density or vegetation cover.

The analysis revealed weak correlations between CO_2 emissions and tree cover, supporting the conclusion that source reduction of emissions is more effective than exclusive reliance on sequestration strategies. The Environmental Vulnerability Index (EVI) underscores the discrepancies between emission intensity and population exposure.

While the data integration methodology is robust, it is important to recognize its limitations. The spatial aggregation of emissions creates uncertainty, especially in areas characterized by significant spatial heterogeneity. Future research should investigate higher-resolution emissions datasets to improve the precision of spatial modeling. The study fails to include real-time emissions variations, which are crucial for comprehending seasonal and diurnal fluctuations. The incorporation of IoT-based monitoring systems alongside remote sensing data facilitates a more dynamic and responsive evaluation of emissions trends.

Finally, future developments of this framework will integrate privacy-enhancing technologies to ensure compliance with ethical and regulatory standards when deploying IoT-based environmental monitoring systems. It will also consider the use of formal privacy accounting techniques and the incorporation of DPIA evaluations for any new data source or processing module. The adoption of edge computing and real-time data collection pipelines must be balanced with robust privacy safeguards, enabling sustainable innovation that is both socially responsible and aligned with data protection principles.

Acknowledgements. This study was funded by the European Union - NextGenerationEU, in the framework of the GRINS - Growing Resilient, INclusive and Sustainable project (GRINS PE00000018 – CUP C93C22005270001). The views and opinions expressed are solely those of the authors and do not necessarily reflect those of the European Union, nor can the European Union be held responsible for them.

References

1. Akanbi, A.K., Masinde, M.: Semantic interoperability middleware architecture for heterogeneous environmental data sources. In: 2018 IST-Africa Week Conference (IST-Africa), p. 1. IEEE (2018)
2. Alotaibi, E., Nassif, N.: Artificial intelligence in environmental monitoring: in-depth analysis. Discov. Artif. Intell. **4**(1), 84 (2024)
3. Amato, F., Guignard, F., Robert, S., Kanevski, M.: A novel framework for spatio-temporal prediction of environmental data using deep learning. Sci. Rep. **10**(1), 22243 (2020). https://www.nature.com/articles/s41598-020-79148-7
4. Bai, Y., Cotrufo, M.F.: Grassland soil carbon sequestration: current understanding, challenges, and solutions. Science **377**(6606), 603–608 (2022)
5. Bobbio, A., Campanile, L., Gribaudo, M., Iacono, M., Marulli, F., Mastroianni, M.: A cyber warfare perspective on risks related to health IoT devices and contact tracing (2023). https://doi.org/10.1007/s00521-021-06720-1
6. Campanile, L., Cantiello, P., Iacono, M., Lotito, R., Marulli, F., Mastroianni, M.: Applying machine learning to weather and pollution data analysis for a better management of local areas: the case of Napoli, Italy (2021). https://www.scopus.com/inward/record.uri?eid=2-s2.0-85135227609&partnerID=40&md5=5a7c117fa01d0ba8d779b0e092bc0f63
7. Cromar, K.R., et al.: Air pollution monitoring for health research and patient care. An official American thoracic society workshop report. Ann. Am. Thorac. Soc. **16**(10), 1207–1214 (2019)
8. Fascista, A.: Toward integrated large-scale environmental monitoring using WSN/UAV/Crowdsensing: a review of applications, signal processing, and future perspectives. Sensors **22**(5) (2022). https://doi.org/10.3390/s22051824
9. Govea, J., Gaibor-Naranjo, W., Sanchez-Viteri, S., Villegas-Ch, W.: Integration of data and predictive models for the evaluation of air quality and noise in urban environments. Sensors (2024). https://www.mdpi.com/1424-8220/24/2/311
10. Heacock, M.L., et al.: Enhancing data integration, interoperability, and reuse to address complex and emerging environmental health problems. Environ. Sci. Technol. **56**(12), 7544–7552 (2022). https://doi.org/10.1021/acs.est.1c08383. pMID: 35549252
11. Hu, Y., O'Donncha, F., Palmes, P., Burke, M., Filgueira, R., Grant, J.: A spatio-temporal LSTM model to forecast across multiple temporal and spatial scales (2021). https://arxiv.org/abs/2108.11875
12. Janssens-Maenhout, G., et al.: Toward an operational anthropogenic CO_2 emissions monitoring and verification support capacity. Bull. Am. Meteor. Soc. **101**(8), E1439–E1451 (2020)
13. Jones, M., Donnelly, A.: Carbon sequestration in temperate grassland ecosystems and the influence of management, climate and elevated CO_2. New Phytol. **164**(3), 423–439 (2004)
14. Kiryluk, H.D., Beard, C.B., Holcomb, K.M.: The use of environmental data in descriptive and predictive models of vector-borne disease in North America. J. Med. Entomol. **61**(3), 595–602 (2024). https://doi.org/10.1093/jme/tjae031
15. Lin, X., et al.: An integrated view of correlated emissions of greenhouse gases and air pollutants in China. Carbon Balance Manage. **18**(1), 9 (2023)
16. Liu, Z., et al.: Near-real-time monitoring of global CO_2 emissions reveals the effects of the covid-19 pandemic. Nat. Commun. **11**(1), 5172 (2020)

17. Manisalidis, I., Stavropoulou, E., Stavropoulos, A., Bezirtzoglou, E.: Environmental and health impacts of air pollution: a review. Front. Public Health **8**, 14 (2020)
18. Mazumder, A., Engala, S.R., Nallapuraju, A.: Towards sustainable development: a novel integrated machine learning model for holistic environmental health monitoring. In: 2023 IEEE MIT Undergraduate Research Technology Conference (URTC), pp. 1–5. IEEE, October 2023. https://doi.org/10.1109/urtc60662.2023.10534927
19. Thomas, M.L., Shaddick, G., Simpson, D., de Hoogh, K., Zidek, J.V.: Data integration for high-resolution, continental-scale estimation of air pollution concentrations. arXiv preprint arXiv:1907.00093 (2019)
20. Wang, Z., Duan, L., Shuai, D., Qiu, T.: Research on water environmental indicators prediction method based on EEMD decomposition with CNN-BiLSTM. Sci. Rep. **14**(1), 1676 (2024)
21. Yasutani, R., Kitazumi, K., Narieda, S., Fujii, T., Umebayashi, K., Naruse, H.: Spatio-temporal analyses of environmental monitoring based on wireless sensor networks. In: 2021 IEEE Sensors. pp. 1–4 (2021). https://doi.org/10.1109/SENSORS47087.2021.9639557
22. Yu, M., et al.: Spatiotemporal event detection: a review. Int. J. Digit. Earth **13**(12), 1339–1365 (2020). https://www.tandfonline.com/doi/full/10.1080/17538947.2020.1738569