

Sustainable and Responsible Management

The Journal of Sustainable Business

ISSN 2724-4466



Volume No. 5, Issue No. 2 - 2024

Abruzzese Salute ETS

sustainable-and-responsible-management.org

Journal Information

Scientific Director

Michele Samuele Borgia

Managing and Responsible Editor

Maurizio Cirillo

Copy Editors

Stefano Cesinaro

Katia Di Tommaso

Chiara Mancini

ISSN

2724-4466

Online since 18th November, 2020

With authorization from the Court of Pescara no. 1 year 2020

Classification by Italian National Agency for the Evaluation of Universities and Research Institutes (ANVUR)

Scientific Journal

Governing Board Resolution nr. 184 of 27th July, 2023

DOI Prefix by Multilingual European DOI Registration Agency (mEDRA)

10.61013

Copyright Holder

Abruzzese Salute ETS

Via Umberto I, 78/80

65010 Cappelle sul Tavo (PE)

Italy



This work is licensed under the Creative Commons Attribution – Noncommercial – No Derivative Works 4.0 International License.

Exploring the impact of livestock farming on air quality in Lombardy

Simone De Luca*

Department of Socio-Economic, Management and Statistical Studies
University “G. d’Annunzio” of Chieti-Pescara
E-mail: simone.deluca002@phd.unich.it

Luigi Ippoliti

Department of Economics
University “G. d’Annunzio” of Chieti-Pescara
E-mail: luigi.ippoliti@unich.it

Pasquale Valentini

Department of Economics
University “G. d’Annunzio” of Chieti-Pescara
E-mail: pasquale.valentini@unich.it

Carlo Zaccardi

Department of Economics
University “G. d’Annunzio” of Chieti-Pescara
E-mail: carlo.zaccardi@unich.it

** Corresponding author*

Abstract

Pollution has become a major concern in recent years, since all its facets have a significant impact on ecosystems and human health. Air pollution is one of the most critical issues, with particulate matter (PM) being a key indicator of air quality. In particular, the formation of atmospheric PM₁₀ (i.e., particles with an aerodynamic diameter of 10 µm or less) depends on various factors, that is precursor gases, weather conditions, and human activities. Exploring the relationship between PM₁₀ and these factors is essential for understanding the dynamics of air pollution and developing effective mitigation strategies. Hence, this study investigates the spatial distribution of PM₁₀ and the other factors in Lombardy, Italy, and aims to identify spatial clusters of these variables. Specifically, the relation between PM₁₀ and ammonia (NH₃) is of primary importance, as a large portion of Lombardy is characterized by intensive livestock farming activities, which are a major source of NH₃ emissions in the region. The analysis is referred to the year 2020, and is based on data from the AgrImOnIA dataset, which includes information on air quality, meteorology, emissions, livestock, and land use. The investigation of the spatial distributions is conducted using spatial interpolation techniques, specifically Kriging, while the local Getis-Ord index is applied to recognize hot spots, that is clusters of municipalities with high pollution levels that require targeted interventions. The results show a positive correlation between PM₁₀ levels and NH₃ emissions, with hot spots identified in areas with high animal density. The findings

provide valuable insights into the dynamics of air pollution in Lombardy and highlight the need for sustainable agricultural practices to improve air quality in the region.

Keywords – Air quality; Livestock sector; Lombardy; Po Valley; PM₁₀; NH₃.

Paper type – Academic Research Paper

Sommario

L'impatto dell'allevamento sulla qualità dell'aria in Lombardia. – Negli anni recenti, l'inquinamento è diventato una delle principali preoccupazioni, poiché tutti i suoi aspetti hanno un impatto significativo sugli ecosistemi e sulla salute umana. L'inquinamento atmosferico è uno dei problemi più critici, poiché il particolato (PM) è un indicatore chiave della qualità dell'aria. In particolare, la formazione di PM₁₀ atmosferico (ovvero particelle con diametro aerodinamico pari o inferiore a 10 µm) dipende da vari fattori, ovvero gas precursori, condizioni meteorologiche e attività umane. Esplorare la relazione tra il PM₁₀ e questi fattori è essenziale per comprendere le dinamiche dell'inquinamento atmosferico e sviluppare strategie di mitigazione efficaci. Pertanto, questo studio indaga la distribuzione spaziale del PM₁₀ e degli altri fattori in Lombardia, Italia, e mira a identificare cluster spaziali di queste variabili. Nello specifico, il rapporto tra PM₁₀ e ammoniaca (NH₃) è di primaria importanza, poiché gran parte della Lombardia è caratterizzata da attività di allevamento intensivo, che rappresentano una delle principali fonti di emissioni di NH₃ nella regione. L'analisi si riferisce all'anno 2020 e si basa su dati tratti dal dataset AgrImOnIA, che include informazioni su qualità dell'aria, meteorologia, emissioni, bestiame e uso del territorio. L'indagine delle distribuzioni spaziali viene condotta utilizzando tecniche di interpolazione spaziale, nello specifico Kriging, mentre l'indice locale Getis-Ord viene applicato per riconoscere gli hot spot, ovvero cluster di comuni con elevati livelli di inquinamento che richiedono interventi mirati. I risultati mostrano una correlazione positiva tra i livelli di PM₁₀ e le emissioni di NH₃, con punti caldi identificati in aree ad alta densità di animali. I risultati forniscono preziose informazioni sulle dinamiche dell'inquinamento atmosferico in Lombardia ed evidenziano la necessità di pratiche agricole sostenibili per migliorare la qualità dell'aria nella regione.

Article history:

Received 29 August 2024

Received in revised form 23 October 2024

Accepted 1 November 2024

Available online

DOI 10.61013/2724-4466/5.2.81-100

1 Introduction

Air pollution poses a growing threat to the environment and human health, with profound implications for quality of life and ecosystem (European Environment Agency, 2021). A thorough understanding of the various pollutants involved is critical to develop effective mitigation strategies. Air pollutants are mainly divided into two categories: primary pollutants, which are emitted directly into the atmosphere, and secondary pollutants, which are formed by chemical reactions and micro-physical processes involving various precursor gases (Nenes, Pandis, Weber, & Russell, 2020).

Air quality analysis focuses primarily on the role of atmospheric particulate matter (PM), i.e. particles classified according to their aerodynamic diameter and considered key indicators in the presence of pollutants in the air. These fine dusts not only reflect the chemical composition of pollutants but also have direct implications for the health of the population. The study of their diffusion and concentration in the air is therefore crucial to understand the overall impact of this phenomenon.

According to Thunis et al. (2021), ammonia (NH_3) plays a crucial role among the most relevant precursor gases in the formation of atmospheric particulate matter.

The analysis proposed by Giannakis, Kushta, Bruggeman, and Lelieveld (2019) utilized annual simulations based on advanced meteorological models and global emission data to assess the impact of ammonia productions on fine particulate pollution in Europe. The research employed the Weather Research and Forecasting model coupled to Chemistry (WRF-Chem) to demonstrate whether ammonia reductions, in line with emission commitments starting from 2020, lead to an effective decrease in $\text{PM}_{2.5}$ (i.e., particulate matter with a diameter of 2.5 micrometers or less) concentrations in the air. The simulation reveals that such reduction in NH_3 productions showed relatively large contractions in $\text{PM}_{2.5}$ levels in Central-Western Europe and the United Kingdom. Furthermore, a peculiar aspect that emerged is that the economic benefit resulting from the avoided premature deaths in Europe amounts to 14.837 million Euros per year, an interesting element given the existing correlation between reductions in $\text{PM}_{2.5}$ and the decrease in premature mortality rates due to diseases linked to air pollution (Giannakis, Kushta, Bruggeman, & Lelieveld, 2019).

In Italy, the Lombardy region, along with other areas of the Po Valley, is of significant interest due to its intense agricultural activity, which substantially contributes to NH_3 emissions. These emissions play a critical role in the formation of PM_{10} (i.e., particulate matter with a diameter of 10 micrometers or less) and $\text{PM}_{2.5}$, as evidenced by ARPA Lombardia (ARPA Lombardia, n.d.) and the AgrImOnIA project (Agriculture Impact On Italian Air Quality; Fassò et al., 2023), both of which underscore the considerable impact of agricultural practices on air quality in northern Italy.

Further research has highlighted the significant role of intensive livestock farms in the production of ammonia and, consequently, in contributing to air pollution. In Italy, for example, intensive livestock

farms are identified as a major source of ammonia, which leads to the formation of particulate matter, surpassing even the transportation and industrial sectors in this regard. As shown in a recent 2022 Greenpeace map (Greenpeace, 2022), these farms are mainly concentrated in the Po Valley, with Lombardy leading the way. This study also revealed that nearly 90% of these farms received public funds, underscoring the complexity of the problem and the need for more sustainable policies.

An imperative approach involves exploring sustainable solutions to mitigate poor air quality and promote eco-friendly agricultural and industrial practices in areas that require greater attention. To this purpose, our research aims to propose an approach to analyse whether the diffusive behaviour of atmospheric particulate matter correlates with the dynamics of agricultural ammonia emissions, in order to explore the possible interaction between these two phenomena. Unlike previous investigations, which focused more on $PM_{2.5}$, our study focuses on PM_{10} . This choice is motivated by the significant presence of sensors for monitoring PM_{10} , as opposed to $PM_{2.5}$, which is monitored by a smaller number of stations. The larger PM_{10} database allows us to investigate air quality dynamics in greater detail, broadening our understanding of environmental implications. We will analyze whether significant variations in PM_{10} measurements can be attributed to intensive use of NH_3 in specific areas of Lombardy, especially in contexts of massive pig and bovine farming. The identification of similar spatial patterns or characteristic clusters among these variables could unveil new insights into the dynamics affecting air quality in the Po Valley.

Through the use of spatial data analysis and geostatistical techniques, we aim to gain a deeper understanding of the factors affecting air quality. Our approach not only offers new insights into the impact of agricultural livestock farming on air quality in this geographic area but also paves the way for more effective strategies to promote sustainable agricultural practices.

The paper is organized as follows: Section 2 presents an initial exploratory analysis of the data, while Section 3 offers a brief introduction to spatial interpolation using Kriging. Section 4 presents the results of spatial clustering through the *local G index*, and Section 5 concludes with a discussion.

2 The data

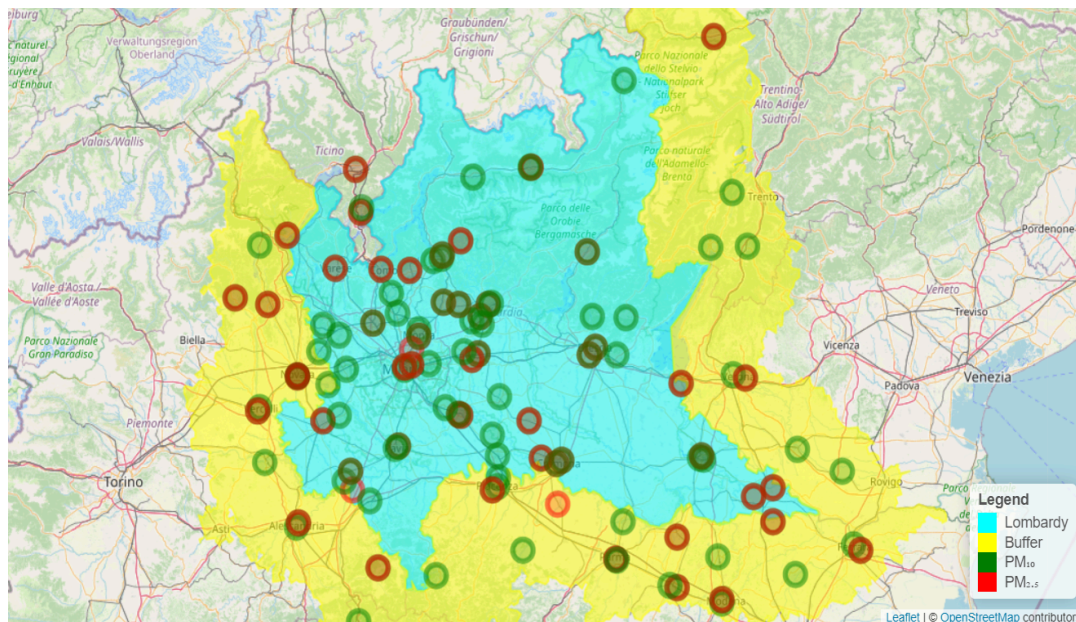
The data used in our study are from the open access dataset AgrImOnIA2 version 2.0.2, available on Zenodo (Fassò et al., 2022). This dataset was crucial for conducting our analysis and understanding the interaction between ammonia emissions and pollutant levels. In a preliminary step, researchers affiliated with this project processed and harmonized the data, aligning them to the same temporal and spatial

resolution to ensure consistency and reliability in the information collected. For specific details about the data harmonization, the reader is referred to the project documentation (Fassò et al., 2022).

The AgrImOnIA2 dataset includes daily surveys collected from 2016 to 2021 from five distinct fields of study that are essential for analyzing different aspects of the surrounding environment. These fields are: air quality (AQ, Air Quality), which monitors particles and noxious gases; meteorology (WE, Weather), which records data such as temperature, humidity and precipitation; air and sectoral emissions (EM, Emissions), which track pollution from various industries; livestock (LI, Livestock), which is important for studies on the impact of intensive livestock farming; and land use and land cover (LA, Land Use), which helps to understand changes in land use. The collection of these data provides a detailed picture of the natural and man-made environment.

The AQ data refer to measurements from 141 monitoring stations scattered throughout Lombardy, along with an area of 30.000 km^2 surrounding it. This additional area serves as a 0.3° buffer, included to improve the accuracy of estimates along the borders. Of these stations, 93 are located within Lombardy and 48 in the extended area. The location of the stations is shown in Figure 1.

Figure 1 – Map of Lombardy and buffer with stations equipped with PM_{10} and $\text{PM}_{2.5}$ sensors



Source: Authors' elaboration

Table 1 – Description of variables used in the analysis

Variable name	Variable description	Field of study	Source
Latitude	Latitude of the monitoring stations expressed in degrees		
Longitude	Longitude of the monitoring stations expressed in degrees		
Time	Daily observation timestamps, covering the period from 2016 to 2021		
AQ_PM10	Concentration of particulate matter with an aerodynamic diameter less than or equal to 10 micrometers (PM_{10}), measured in micrograms per cubic meter ($\mu g/m^3$)	Air Quality	(ARPA Lombardia 2023), (EEA, 2023)
WE_tot_precipitation	Total precipitation, including accumulated liquid and frozen water (such as rain, snow, and sleet) reaching the Earth's surface, measured in meters (m)	Weather	(Copernicus Climate Change Service, 2023)
EM_nh3_livestock_m	Emissions of ammonia (NH_3) from livestock activities, measured in milligrams per square meter (mg/m^2). Data collected to assess agricultural emission sources	Emission	(ECCAD, 2022)
EM_nh3_sum	Total emissions of ammonia (NH_3), covering sources such as livestock, soil, and burning, measured in milligrams per square meter (mg/m^2). Data used for comprehensive emission monitoring	Emission	(ECCAD, 2022)
EM_nox_traffic	Emissions of nitrogen oxides (NO_x) from road traffic, measured in milligrams per square meter (mg/m^2). Data obtained to evaluate transportation-related pollution on road	Emission	(ECCAD, 2022)
LI_pigs	Density of pigs within the municipality that has the closest centroid to the monitoring station, measured in square kilometers (km^2). The data help to understand the influence of pigs on emissions impacting air quality in the region	Livestock	(Italian Ministry of Health, 2021)
LI_bovine	Density of bovines within the municipality that has the closest centroid to the monitoring station, measured in square kilometers (km^2). The data help to understand the influence of bovines on emissions impacting air quality in the region	Livestock	(Italian Ministry of Health, 2021)
LA_hvi	High Vegetation Index in m^2/m^2 . This index is a key parameter for the study of land use and vegetation cover, reflecting the environmental conditions of the analyzed region	Land Use	(Muñoz, 2019)

Source: AgrImOnIA2 dataset

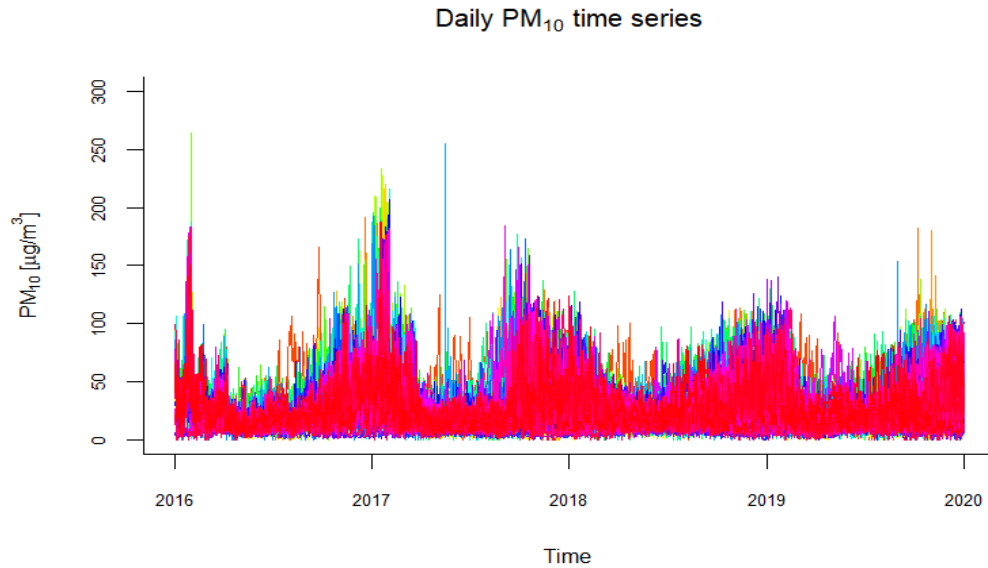
As part of our research, we decided to focus on a specific subset of variables among those available in the dataset, which we believe are particularly relevant for analyzing the interaction between NH_3 emissions and PM_{10} levels. In fact, this interaction is influenced by other variables related to human activities (such as agriculture and urban traffic) and natural factors (such as precipitation and vegetation). A detailed list with related descriptions of these selected variables is given in Table 1. By focusing on these features we are able to perform a more detailed and meaningful analysis, which we expect will contribute to accurate and informative results.

The dynamics the daily time series of PM_{10} and NH_3 are shown for all the monitoring sites in Figure 2 and Figure 3, respectively, from 2016 to 2020.

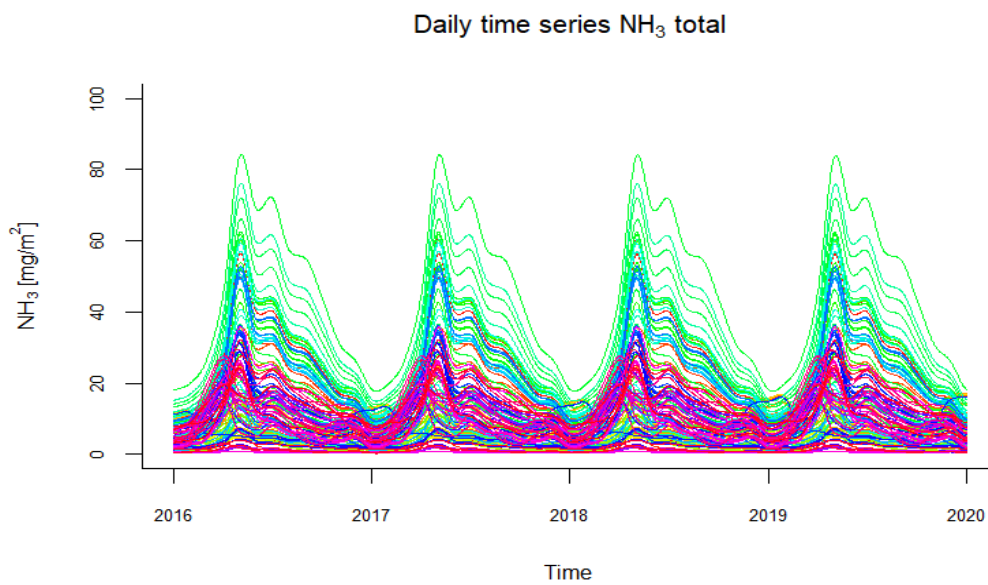
Both variables exhibit a clear periodicity, with PM_{10} levels increasing during the winter months and decreasing in the summer. In contrast, ammonia shows an opposite trend, with levels rising in the summer and declining during the winter period. However, since our focus is primarily on spatial analysis, we chose to aggregate the data annually (i.e., marginalizing time) to concentrate on the spatial distribution of PM_{10} and NH_3 .

As a result, Figure 4 illustrates the annual average of PM_{10} concentrations for three years (2016, 2018 and 2020) and for all stations in the study region. The points vary in size and color intensity to reflect the magnitude of the phenomenon. These plots offer an immediate and intuitive representation of the geographic distribution and severity of PM_{10} pollution, highlighting the most affected areas. However, because the data are only available at specific monitoring stations, spatial interpolation is necessary to gain a comprehensive understanding of the pollutant distribution across the entire region. Interpolation allows us to estimate pollutant levels at locations between these stations, providing a surface of PM_{10} concentrations. This process is also crucial for comparing the spatial distribution of PM_{10} with that of NH_3 . By interpolating PM_{10} data, we can better analyze and visualize how its distribution aligns or contrasts with the distribution of NH_3 , thereby enhancing our overall understanding of air quality patterns in the region.

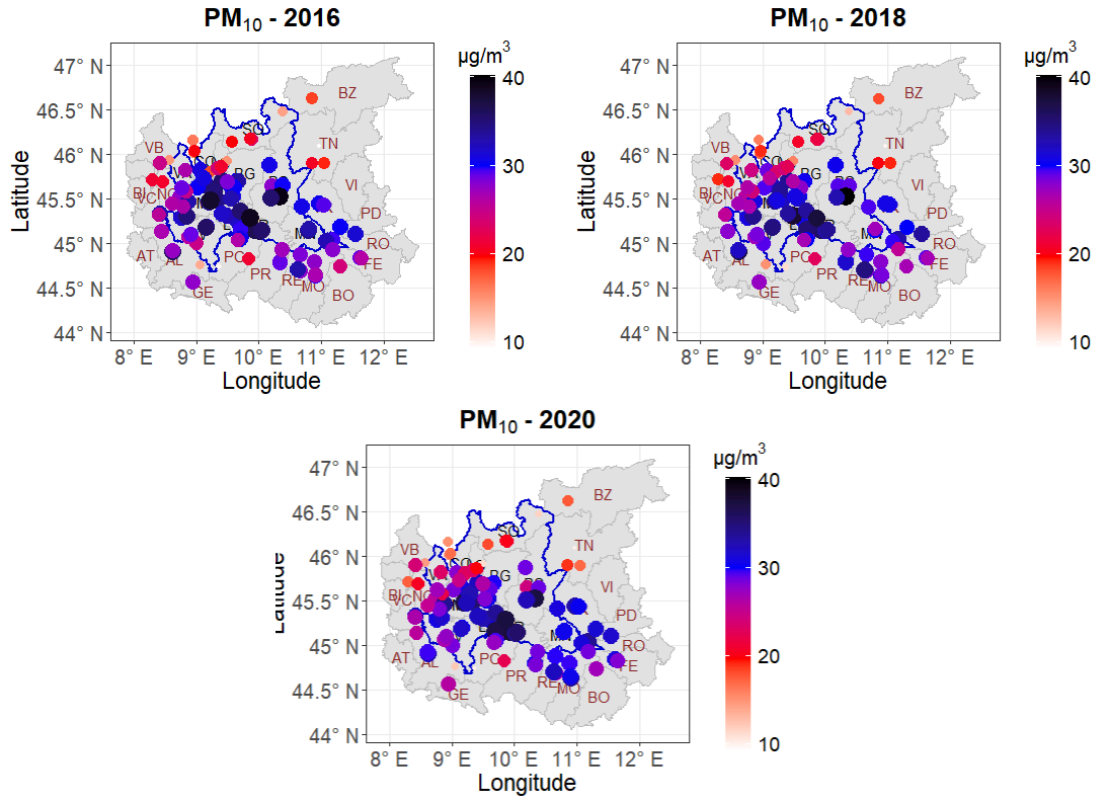
Our spatial analysis centers on data from the year 2020, chosen for its representativeness and the availability of comprehensive information on PM_{10} and NH_3 concentrations. This year reflects the environmental trends observed from 2016 to 2020, which were marked by relative stability. The analysis was also limited to 2020 due to the absence of ammonia data for 2021, which prevents a fair comparison with prior years.

Figure 2 – Daily time series of PM₁₀ levels for all monitoring stations in Lombardy

Source: Authors' elaboration

Figure 3 – Daily time series of total NH₃ levels for all monitoring stations in Lombardy

Source: Authors' elaboration

Figure 4 – Maps of average PM₁₀ concentration levels for each site in the years 2016, 2018, 2020

Note: Darker color and bigger point size indicate higher concentrations.

Source: Authors' elaboration

3 Spatial interpolation by Kriging

This section provides a brief introduction on Kriging for using this method to interpolate the values of continuous variables, such as PM₁₀ concentrations. This choice promotes an extension of the representation of the environmental variables under study, beyond the boundaries of the directly monitored sites.

If $z = (z(s_1), \dots, z(s_n))'$ are measurements observed at n spatial locations, $s_1, \dots, s_n$, from a real-valued spatial process $\{Y(s), s \in D\}$, with $D \subset \mathbb{R}^d$, a widely used representation of the random process is:

$$Z(s) = Y(s) + \epsilon(s)$$

where $\epsilon(s)$ is a sequence of uncorrelated Gaussian random variables, each with mean 0 and variance σ_ϵ^2 .

The process $Y(s)$ corresponds to the true (latent) spatial process vector of interest for which we further assume the following decomposition:

$$Y(s) = \mu(s) + \eta(s)$$

where $\mu(s)$, represents a spatial mean component and, $\eta(s)$, is a zero mean spatially-dependent random process with covariance matrix Σ_η . Usually, the mean, $\mu(s)$, is a parametrized function of spatial covariates, $X_j(s), j = 1, \dots, p$.

Predicting or interpolating the variable of interest, Y , at some unobserved location s_0 , it is of particular interest in spatial analysis. In practice, many different methods exist for spatial prediction, and the choice of a given method will depend on the specific characteristics of the data being analysed. Gaussian processes and Kriging (Cressie, 1993) are powerful tools for spatial interpolation and can be used to capture complex spatial structures and dependencies.

The application of Kriging, based on accurate spatial correlation between the sampling points, not only facilitates estimation, but also allows for detailed and uniform mapping of the distribution of environmental phenomena over the entire Lombardy region. This gives an extended and comprehensive view of the analyzed environmental impact, while also integrating the uncertainty estimates associated with each interpolation.

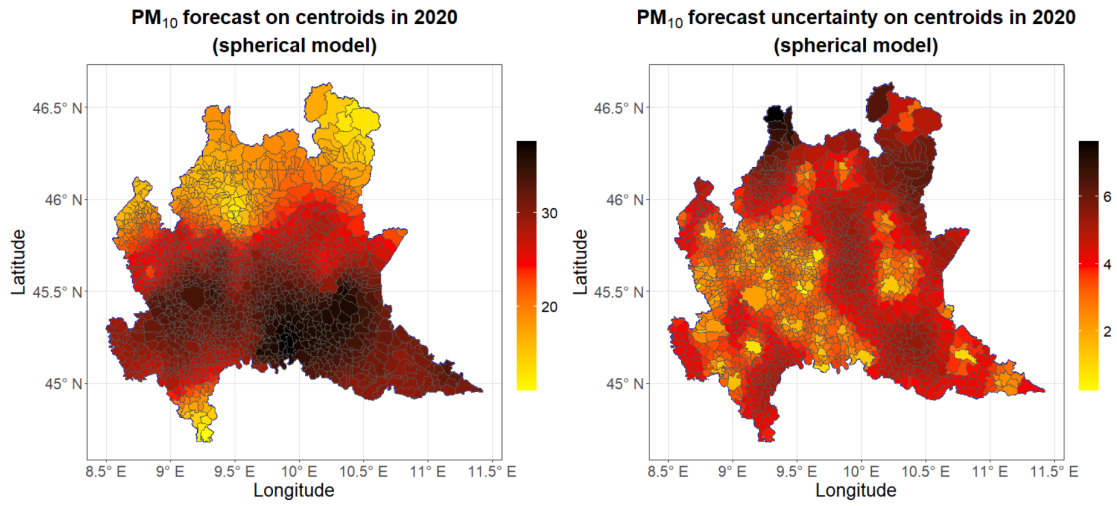
Mathematically, the Kriging prediction at the new site s_0 results in a linear combination of data,

$$\hat{Y}(s_0) = \sum_{i=1}^n \lambda_i z(s_i)$$

where $\lambda_i, i = 1, \dots, n$, are weighting coefficients which, under the choice of Universal Kriging (Cressie, 1993), depend on some trend coefficients, the covariance function used to represent the spatial dependence and the measurement noise.

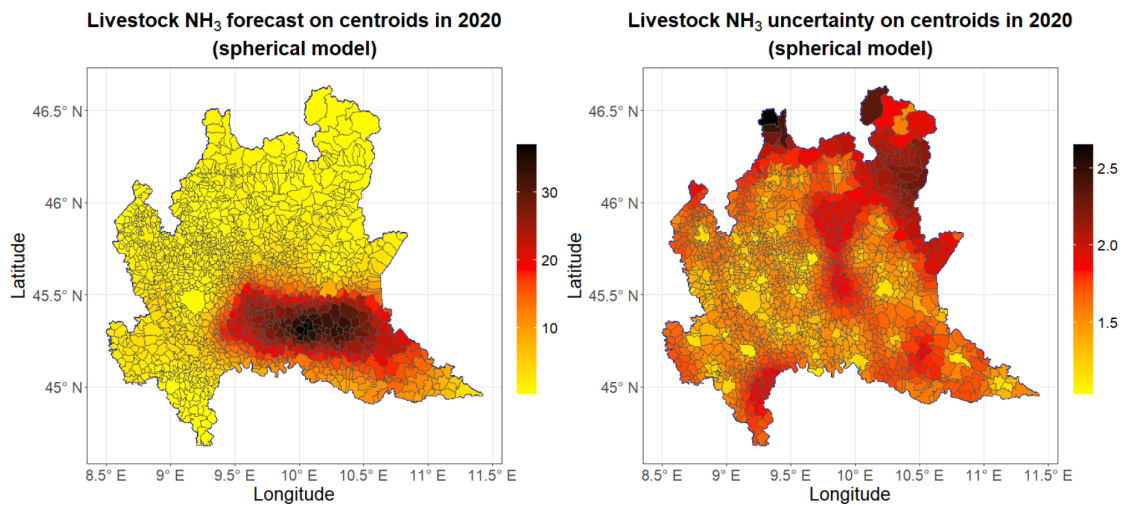
Figures 5 and 6 display the predictions at the 1503 centroids of the municipalities, along with the corresponding measures of uncertainty, for PM_{10} and livestock NH_3 for the year 2020. Predictions provided at the municipality level allow for the spatial alignment of PM_{10} data with NH_3 and facilitate a comparison between the two variables. These predictions were obtained by fitting a linear trend and a spherical covariance function with an included nugget effect. The critical parameters of the spherical model — namely *sill*, *range*, and *nugget* — were estimated using the maximum likelihood method (Cressie, 1993). All estimates were computed using the “geoR” package in R (Ribeiro & Diggle, 2001).

Figure 5 – Spatial distribution of PM_{10} in 2020 at the 1503 centroids of the municipalities of Lombardy



Source: Authors' elaboration

Figure 6 – Spatial distribution of NH_3 from livestock farming in 2020 at the 1503 centroids of the municipalities of Lombardy



Source: Authors' elaboration

As observed, the highest levels for both variables are found in the southeast part of the region. Additionally, the uncertainty in the predictions tends to be higher in areas with fewer monitoring sites.

4 Spatial clusters: the local G_i index

The examination of statistical dependence between observations assumes fundamental importance not only for interpolation purposes. A central principle is the notion of spatial autocorrelation, defined as the propensity of geographically contiguous observations to exhibit similar attributes. This phenomenon is crucial to the analysis of the spatial structure of datasets and finds its foundation in the First Law of Geography formulated by Tobler, which postulates that “everything is related to everything else, but the nearest elements exhibit a higher degree of correlation than those that are more distant” (Tobler, 1970). Spatial autocorrelation can manifest itself in the form of positive autocorrelation, characterized by the aggregation of similar observations, or negative autocorrelation, where observations close to each other have a high degree of dissimilarity.

To address these features in the data at the municipality level, distance metrics and weight matrices must be specifically calibrated to quantify the spatial relationships between observations. Within this analytical framework, the local spatial index of Getis and Ord assumes critical importance, as it facilitates the identification of spatial clusters of similar or dissimilar observations, contributing significantly to the exploration of the spatial structure of the data for the benefit of statistical modeling (Getis & Ord, 1996). The Getis-Ord G index is computed as a function of a distance or threshold predefined by the analyst and allows for a distinction between global-level analysis and local-level analysis. The global-level analysis is carried out through global G index, which provides a single value for the entire study domain. The local-level analysis is carried out through the local G_i index, which captures the presence of spatial autocorrelation at the local scale for each geographic unit or municipality.

There are two main variants of the local G_i index that tend to generate comparable results, despite their methodological differences. The first variant, denoted as G_i , excludes the autocorrelation of an observation with itself, thus precluding the value of the i -th variable from the autocorrelation computation and focusing only on the values of j -th adjacent features. In contrast, the second variant, identified as G_i^* , integrates the value of the i -th feature with those of neighboring observations, allowing the autocorrelation of a unit to be assessed by taking into account the interactions with its spatial neighbors.

For the development of the present study, a methodological choice aimed toward the use of the local index G_i was adopted. The formulation of the index G_i , employed in our investigation, is broken down as follows:

$$G_i = \frac{\sum_{j \neq i} w_{ij}(d)y_j}{\sum_{j \neq i} y_j}$$

where G_i is the Getis-Ord index for the spatial unit i -th, w_{ij} is the weight assigned to the pair of spatial units i and j based on their distance d or neighborhood relationship, and y_j is the value of the variable of interest for all neighbouring spatial units j other than i .

The formula under analysis highlights the mechanism by which the cross product between the y values at positions i and j is weighted by a distance coefficient w_{ij} . This coefficient is defined in such a way that it takes on the value of unity if the two positions are identical or are within a distance less than a given threshold limit d , and zero otherwise.

The investigation of the statistical significance of local autocorrelation is conducted through statistical inference, which makes it possible to perform homogeneous comparisons between different spatial units. In addition, the evaluative approach toward clustering patterns related to spatial concentration and their statistical significance provides a quantitative basis for interpreting the statistical significance of the identified spatial configurations.

Hence, the application of the local G_i index proves crucial in examining the spatial distribution of values within specific geographic areas, allowing one to distinguish between random distributional configurations and the marked presence of hot spots and cold spots. A hot spot is defined as an area where a significantly high concentration of high values is observed, indicative of an agglomeration of positive or high intensity events or attributes in the specific region. In contrast, a cold spot is identified as a prevalent area of low values, which denotes an accumulation of negative or low intensity events or attributes. The analysis made possible by the G_i index facilitates the identification of clusters of homogeneous (hot spots) or heterogeneous (cold spots) values, providing a detailed representation of spatial variations that contributes to the interpretation of complex phenomena.

The adoption of a Z test to assess the presence of hot spots and cold spots is justified through its ability to quantify the distance of local G_i index values from their expected value, assuming the null hypothesis of random distribution. In practical terms, the z -score transforms the local G_i index into a value that reflects the number of standard deviations of difference between the local observation and the expected mean under the null hypothesis. This transformation facilitates hypothesis testing, allowing us

to determine whether the observed spatial concentrations (hot spots or cold spots) are statistically significant or the result of chance. The null hypothesis (H_0) assumes that there is no spatial autocorrelation, implying that the spatial distribution of observed values is the result of chance. The alternative hypothesis (H_1), in contrast, suggests the existence of spatial autocorrelation, thus indicating the presence of spatial clusters of high values (hot spots) or low values (cold spots) (Abdulhafedh, 2017). The Z test takes the following form and distribution:

$$Z_i = \frac{G_i - E(G_i)}{\sigma_{G_i}} \sim N(0, 1)$$

where $E(G_i)$ is the expected value of the G_i index under the null hypothesis of no spatial autocorrelation, representing the mean expected in a random distribution, while σ_{G_i} is the standard deviation of the G_i index under the null hypothesis, indicating the expected variability of G_i in a context of random distribution.

The classification of clusters into hot spots, cold spots and insignificant areas is evaluated based on the following criteria:

- **Hot Spots:** An area is classified as a hot spot if it has a positive z -value and a p -value below the predefined significance threshold ($\alpha = 0.05$). This indicates a significant concentration of high values, higher than expected under the randomness assumption.
- **Cold Spots:** An area is identified as a cold spot if it has a negative z -value accompanied by a p -value below the significance threshold. This suggests a significant concentration of low values, implying a clustering of lower values than expected by chance.
- **Insignificant Areas:** Areas are considered insignificant if their p -values exceed the established significance threshold, indicating that the spatial concentration could result from random variations.

Applying this statistical framework for local G_i index analysis allows for accurate classification of areas in terms of spatial significance, providing a solid basis for interpreting the distribution of the phenomena studied. Identifying hot spots and cold spots provides valuable insights into areas requiring priority attention or specific interventions, while distinguishing insignificant areas helps avoid false interpretations of random patterns as significant phenomena. This methodology not only enriches spatial analysis in the context of air quality in Lombardy but also offers a replicable model for similar studies in other disciplines and geographic settings.

4.1 Spatial clusters

For a clearer and spatially contextualized understanding of the results, we provide a visual analysis using “cluster maps”, thus based on the application of the local spatial index G_i . This mapping will highlight classified areas, allowing us to quickly identify critical areas and thus assess the spatial organization of clusters.

The variables considered in the analysis are PM_{10} , total and livestock NH_3 emissions, pig and bovine density, NO_x traffic emissions, high vegetation index (HVI), and total precipitation. Each of these is relevant to study the spatial distribution of phenomena that contribute negatively or positively to air quality in Lombardy. Therefore, our analysis was conducted by calculating the G_i for the above variables, filtered according to the reference period (2020).

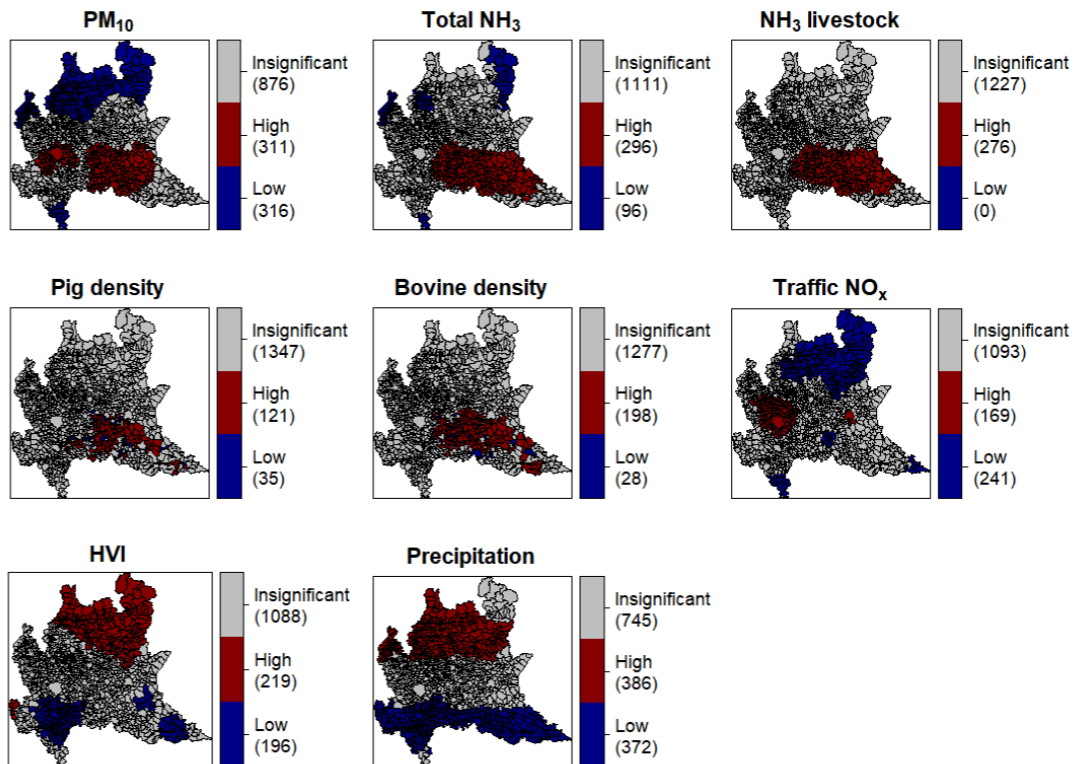
In order to apply this measure, we define a neighborhood structure by considering the matrix of spatial weights based on the “*k-nearest neighbors*” criterion: given the varied landscape of municipal areas in Lombardy, we adopt this matrix configuration in order to ensure that each unit has a uniform number of neighbors. This aims to prevent overestimation of units characterized by a larger number of neighboring localities. Considering some exploratory surveys conducted on the number of connections between municipalities, we decided to adopt a value of $k = 5$ and standardize the contiguity matrix by row.

Using the “localG()” function of R’s “spdep” library (Bivand, 2022), the software provides the local G_i measures for each area, including also the relative *expected value*, *z-scores* and associated *p-values*. In Figure 7, we report the results obtained from the “spplot()” function applied to the considered local index (Bivand, Pebesma, & Gomez-Rubio, 2013).

In general, the maps illustrate a notable spatial correspondence in the concentration levels of various environmental and agricultural variables. Specifically, high or low concentrations of PM_{10} , NH_3 emissions from the livestock sector, total NH_3 levels, animal species density (including pigs and bovine animals), and traffic emissions of nitrogen oxides (NO_x) tend to occur at the same geographic locations. This alignment suggests that these variables are closely linked spatially, possibly due to common influencing factors such as regional agricultural practices, traffic patterns, or environmental conditions. On the other hand, an opposite distribution is observed for the variables High Vegetation Index and total precipitation. Areas with high HVI typically indicate regions with dense vegetation, which could absorb pollutants and potentially lower the concentrations of PM_{10} and NH_3 emissions. Similarly, high total precipitation often results in the removal or dilution of pollutants, leading to lower observed concentrations. Thus, regions with high vegetation and high precipitation might show reduced levels of

pollutants compared to areas with less vegetation and lower precipitation. This contrast highlights the different influences these factors have on environmental and air quality conditions.

Figure 7 – Cluster map of air quality in Lombardy (2020) based on the local G_i index



Note: “High” denotes hot spots, “Low” denotes cold spots.

Source: Authors' elaboration

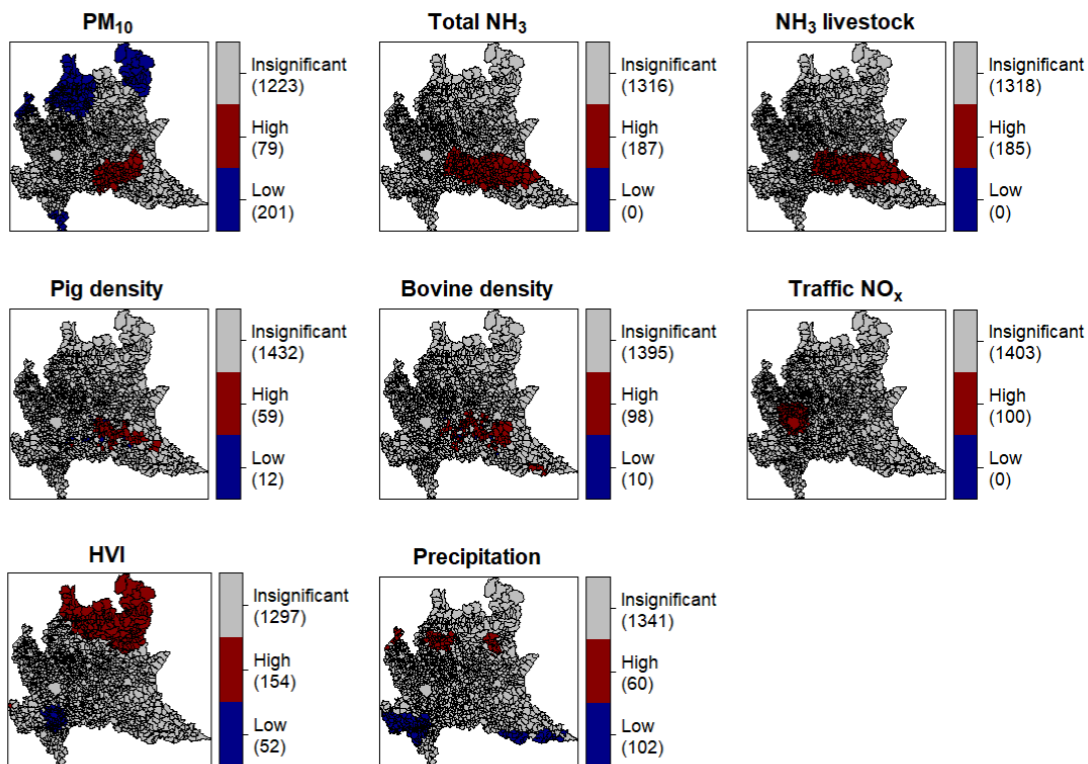
From the analysis conducted, a significant framework emerges that draws attention to the areas identified as hot spots, clearly highlighted in red in the maps. The relevant intensity in these specific areas underscores the importance of adopting targeted approaches to monitor high pollutant measurements in the area.

An important methodological challenge in local spatial autocorrelation statistics is selecting an appropriate p-value cut-off to effectively control the Type I error rate. To enhance the robustness of our findings and minimize the risk of Type I errors (false positives), we applied the False Discovery Rate

(FDR) adjustment (Benjamini & Hochberg, 1995) to the p-values. This adjustment was necessary to address the issue of multiple comparisons, which is common in spatial analysis using the local G index.

The FDR procedure offers a more flexible approach than the Bonferroni correction, as it controls the expected proportion of false positives among the rejected hypotheses. This less stringent method strikes a better balance between detecting true clusters and reducing false positives. Using this procedure, we recalculated the local G index for the same variables and extracted the associated p-values. The FDR correction was then applied following the Benjamini-Hochberg (1995) method, which limits the proportion of false positives among the significant results. Significant clusters were identified using a consistent α threshold of 0.01 across all variables, ensuring the detection of interesting spatial patterns while maintaining strict control over false positives.

**Figure 8 – Cluster map of air quality in Lombardy (2020)
based on the local G_i index corrected using False Discovery Rate (FDR)**



Note: “High” denotes hot spots, “Low” denotes cold spots.

Source: Authors' elaboration

The resulting map in Figure 8 reveals differences compared to the uncorrected map (Figure 7) for the local G_i index, particularly in the identification of new, refined spatial clusters. This highlights the importance of applying appropriate correction methods to enhance the reliability of hot spot and cold spot analysis in spatial studies.

5 Conclusions

This work aimed to investigate the spatial relationship between PM_{10} concentration levels and several variables representing meteorological conditions, vegetation, and emission levels of air pollutants related to human activities such as farming and traffic (i.e., NH_3 and NO_x , respectively). In particular, the study domain was the Lombardy region, and the territories located in its southeastern part (i.e., those belonging to the Po Valley and corresponding to the provinces of Mantova, Cremona, and Brescia) are characterized by intensive agricultural and livestock activities, a significant source of NH_3 emissions, which in turn contribute to the formation of atmospheric PM_{10} . Hence, ammonia emissions are a crucial factor to consider when analyzing air quality in this region. The first step in the analysis involved the alignment of the spatial scales of PM_{10} levels and NH_3 emissions. Universal Kriging was used to predict the values of these variables at the municipality level, and Figures 5 and 6 depicted that the highest levels of PM_{10} and livestock NH_3 emissions were concentrated in the southeastern part of Lombardy, corroborating the potential relationship between the two variables.

The second stage of the analysis focused on the spatial clustering, using the local G_i index, of eight variables, namely PM_{10} , total and livestock NH_3 emissions, pig and bovine density, NO_x traffic emissions, HVI, and total precipitation (see Figure 7). Furthermore, to minimize the risk of Type I errors (false positives), the FDR correction was applied (see Figure 8). Overall, these analyses revealed the presence of hot spots and cold spots in the region. Areas identified as hot spots, characterized by high levels of PM_{10} and NH_3 emissions, were found in the Central-South belt (within the Po Valley), where intensive agricultural and livestock activities are prevalent, and there is low vegetation and total precipitation. Another hot spot area was identified in the vicinity of Milan and adjacent municipalities, where NO_x emissions from traffic are more pronounced. This area is much smaller than the previous one, indicating that reducing ammonia emissions from farming activities could have a more significant impact on the air quality of the region than reducing NO_x emissions from traffic.

It is important to mention that there are alternative indices for the identification of spatial clusters, such as the local Moran I_i index (Anselin, 1995). In this study, the “localmoran()” function from the “spdep” package in R was also applied (Bivand, 2022), producing results similar to those obtained with the local G_i index, as shown in Figures 7 and 8. This consistency suggests that the spatial patterns identified are robust to the choice of index, reinforcing the validity of the conclusions drawn.

In summary, the analysis carried out in this work suggests that the spatial distribution of PM₁₀ concentrations in Lombardy is influenced by farming activities, as well as road traffic and environmental factors such as vegetation and precipitation. A limitation of this study is that the complexity of spatial interdependencies in spatial autocorrelation was underscored. Further research is needed to carry out an integrated approach to sustainability that addresses both air quality issues and the social impacts resulting from these environmental conditions. On the other hand, the results of this study can be used to inform policy decisions and interventions aimed at improving air quality and public health in Lombardy. In fact, this work highlights that strategies aimed at reducing ammonia emissions in the atmosphere are compelling in the areas identified as hot spots.

References

- Abdulhafedh, A. (2017). A novel hybrid method for measuring the spatial autocorrelation of vehicular crashes: Combining moran's index and getis-ord gi* statistic. *Open Journal of Civil Engineering*, 7, 208–221. <https://doi.org/10.4236/ojce.2017.72013>
- Anselin, L. (1995). Local indicators of spatial association—LISA. *Geographical Analysis*, 27, 93–115. <https://doi.org/10.1111/j.1538-4632.1995.tb00338.x>
- ARPA Lombardia (n.d.). *Ammoniaca*. Regione Lombardia. <https://www.arpalombardia.it/temi-ambientali/aria/inquinanti/ammoniaca/>
- ARPA Lombardia (2023). *Dati e indicatori: Aria*. <https://www.arpalombardia.it/dati-e-indicatori/aria/>
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1), 289–300.
- Bivand, R. (2022). R packages for analyzing spatial data: A comparative case study with areal data. *Geographical Analysis*, 54(3), 488–518. <https://doi.org/10.1111/gean.12319>
- Bivand, R. S., Pebesma, E., & Gomez-Rubio, V. (2013). *Applied spatial data analysis with R*. Springer. <https://doi.org/10.1007/978-1-4614-7618-4>
- Copernicus Climate Change Service (2023). *Climate change*. <https://climate.copernicus.eu/>
- Cressie, N. A. C. (1993). *Statistics for spatial data*. Wiley. <https://doi.org/10.1002/9781119115151>

- ECCAD (2022). *Cams-glob-ant*. <https://permalink.aeris-data.fr/CAMS-GLOB-ANT>
- European Environment Agency (2021). *Air quality in europe 2021* (technical report n. 15/2021). <https://www.eea.europa.eu/publications/air-quality-in-europe-2021>
- European Environment Agency (2023). *Air pollution*. <https://www.eea.europa.eu/en/topics/in-depth/air-pollution>
- Fassò, A., Rodeschini, J., Fusta Moro, A., Shaboviq, Q., Maranzano, P., Cameletti, M., et al. (2022). *Agrimonia: Open access dataset correlating livestock and air quality in the Lombardy region, Italy. agrimonia dataset (Version 2.0.2)*. <https://doi.org/10.5281/zenodo.6620530>
- Fassò, A., Rodeschini, J., Fusta Moro, A., Shaboviq, Q., Maranzano, P., Cameletti, M., et al. (2023). *Agrimonia: A dataset on livestock, meteorology and air quality in the Lombardy region, Italy. Scientific Data*, 10(1), 143. <https://doi.org/10.1038/s41597-023-02034-0>
- Getis, A., & Ord, J. K. (1996). *Local spatial statistics: An overview*. In P. Longley & M. Batty (Eds.), *Spatial analysis: Modelling in a GIS environment* (pp. 261-277). GeoInformation International.
- Giannakis, E., Kushta, J., Bruggeman, A., & Lelieveld, J. (2019). Costs and benefits of agricultural ammonia emission abatement options for compliance with european air quality regulations. *Environmental Science Europe*, 31, 93. <https://doi.org/10.1186/s12302-019-0275-0>
- Greenpeace (2022). *Allevamenti intensivi emettono ammoniaca, inquinano l'aria e ricevono soldi pubblici*. <https://www.greenpeace.org/italy/storia/16610/>
- Italian Ministry of Health (2021). *Banca dati nazionale (BDN) dell'anagrafe zootecnica [Italian veterinary information system]*. <https://www.vetinfo.it/>
- Muñoz, S. J. (2019). *ERA5-Land hourly data from 1950 to present. Copernicus climate change service (c3s) climate data store (cds)*. <https://doi.org/10.24381/cds.e2161bac>
- Nenes, A., Pandis, S., Weber, R. J., & Russell, A. (2020). Aerosol ph and liquid water content determine when particulate matter is sensitive to ammonia and nitrate availability. *Atmospheric Chemistry and Physics*, 20, 3249-3258. <https://doi.org/10.5194/acp-20-3249-2020>
- Ribeiro, P. J., & Diggle, P. (2001). Geor: A package for geostatistical analysis. *R-NEWS*, 1(2). <https://cran.r-project.org/web/packages/geoR/index.html>
- Thunis, P., Clappier, A., Beekmann, M., Putaud, J. P., Cuvelier, C., Madrazo, J., et al. (2021). Non-linear response of pm2.5 to changes in nox and nh3 emissions in the po basin (italy): Consequences for air quality plans. *Atmospheric Chemistry and Physics*, 21, 9309-9327. <https://doi.org/10.5194/acp-21-9309-2021>
- Tobler, W. R. (1970). A computer movie simulating urban growth in the detroit region. *Economic Geography*, 46, 234-240. <https://doi.org/10.2307/143141>