The Effect of Territorial Servitization on Air Quality: Empirical Analysis of PM2.5 Concentration in Italian Municipalities

Giulia Palma[1] and Francesco Scotti[2*]

[1] *Dipartimento di Scienze Sociali, Politiche e Cognitive, Università di Siena, 53100 Siena, Italy*
[2]*Department of Management, Economics and Industrial Engineering, Politecnico di Milano, Via Lambruschini, 4/B, 20156, Milan, Italy.*
[*]*corresponding author: francesco.scotti@polimi.it*

Abstract

The challenge of air pollution continues to affect global public health and environmental quality. This paper explores the extent to which territorial servitization may contribute to reducing PM2.5 levels in Italian municipalities. Based on the application of machine learning techniques alongside traditional regression analyses, we show that a 1% growth in the concentration of knowledge-intensive business services reduces PM2.5 levels by percentages between 0.11% and 0.47%. We also clarify how a wide range of socioeconomic and meteorological variables govern air quality. Moreover, we estimate the municipal environmental efficiency through a data envelopment analysis approach, showing that better scores are achieved by territories with higher servitization and quality of institutions. Our findings, which are stable across different geographical macro-areas, suggest that both the economic structure and the capabilities of local administration may drive air pollution dynamics. Such results are essential for policymakers aiming to design effective interventions targeting local environmental sustainability.

Keywords:— PM2.5, servitization, air quality, machine learning, regression

1 Introduction

Carbon emissions have substantial implications for health (Ferro et al., 2024), tourism (Zhang et al., 2020), labour participation (Li and Li, 2022), productivity (Zivin and Neidell, 2012), and safety levels (Bondy et al., 2020). As air pollution does not entail only environmental losses but significantly affects economic systems' sustainable development, reducing carbon emissions has become a central focus of global policy agendas, particularly in light of international commitments such as the Paris Agreement and the United Nations' Agenda 2030 (Bel and Teixidó, 2020, Tolonen, 2024). Consequently, understanding the drivers of air quality has emerged as a critical area of investigation for policymakers and researchers.

Air pollution levels are influenced by a complex interplay of factors, including geographical characteristics, meteorological conditions, and the economic structure of a region. Extant studies have widely discussed how the wind (Li et al., 2017), atmospheric pressure (Danek et al., 2022), temperature (Dawson et al., 2007), relative humidity (Lou et al., 2017), spatial topography (Guo et al., 2024), and urban morphology (Augusto et al., 2024) affect the air quality with well-consolidated evidence. Nonetheless, there remains a notable gap in understanding how the economic framework of a locality impacts the local level of carbon emissions.

Our paper aims to fill this gap by investigating how recently emerged economic models may contribute to decoupling economic growth from environmental degradation, addressing the pressing need for sustainable development. In particular, we focus on the promising approach of territorial *servitization*, defined as the capability of territories to produce output from knowledge-intensive business services (KIBS) (Lafuente et al., 2017). As territorial servitization involves the introduction of value-adding services into traditional operations (Cusumano et al., 2015), this transition has the potential to substantially lower emissions by reducing dependence on material goods, energy, and raw materials (Rothenberg, 2007, Roos, 2015). Furthermore, servitization tends to substitute physical input factors (e.g., machinery, and equipment) with service factors such as intellectual, human, and technological capitals that are usually clean and imply lower levels of carbon emissions (Yusliza et al., 2020, Hao et al., 2021). Finally, servitization is usually technically supported by a process of digital transformation that assists enterprises in facilitating a circular economy and reducing resource waste (Wilts and Berg, 2018, Paschou et al., 2020, Bressanelli et al., 2024).

Despite its potential benefits, empirical evidence examining the impact of servitization on air quality remains limited to applications at the firm level in terms of energy conservation, consumption, and waste management (Wang et al., 2023, Vaillant and Lafuente, 2024).[1] Conversely, the bulk of theoretical and empirical studies discusses the economic impact of shifting towards service-based models (Lafuente et al., 2017, Bellandi

[1]See also Menon et al. (2024) for a systematic review on the relationship between servitization and both the environmental and economic performance of firms.

and Santini, 2019, Lombardi et al., 2022), while neglecting the effects of servitization on air pollution measured at the territorial level.

Our paper addresses this research objective by analysing how territorial servitization influences air pollution at the municipal level in Italy while controlling for a rich set of meteorological, economic, and demographic factors. Specifically, we robustly investigate the main drivers of PM2.5 emissions using machine learning and regression methods. We estimate Random Forest and XGBoost models and provide alternative metrics on the significance of factors influencing emission levels. These models utilize feature importance metrics and Shapley Values to clarify the relevance of individual features. Additionally, we strengthen the empirical evidence obtained with a basic Ordinary Least Squares (OLS) approach through Ridge and Lasso regression techniques.

We demonstrate that a 1% increase in the concentration of knowledge-intensive service sectors leads to a reduction in PM2.5 levels ranging from 0.11% to 0.47%. Additionally, we provide a comprehensive analysis of the influence of various socioeconomic and meteorological factors on air quality. For instance, lower PM2.5 levels are experienced by municipalities with higher rainfall levels and wind speed. Moreover, wealthier regions characterized by higher per capita income and a larger share of the active population are associated with poorer air quality.

We then assess municipal environmental efficiency using a data envelopment analysis framework, revealing that better scores are achieved by areas located in the Centre and South of Italy. In particular, regions characterized by higher levels of servitization and stronger institutional quality achieve superior environmental efficiency scores, meaning that both economic structure and the administrative capabilities of local governments play a crucial role in shaping air pollution dynamics. Due to the heterogeneity of air quality across Italian macro-areas, we repeat our analysis for municipalities located in the North, Centre, and South of Italy, obtaining consistent results.

The relevance and timeliness of our research cannot be overstated. As cities grapple with increasing pollution levels and their associated health impacts, understanding the economic drivers behind these trends is essential for developing effective policy interventions. Furthermore, our work contributes to an emerging body of literature that seeks to integrate economic theory with environmental science, providing insights that can inform sustainable urban planning and policy decisions. This is particularly significant in light of the European Union's increasingly stringent PM2.5 regulations and the United Nations' 11^{th} Sustainable Development Goal, which emphasizes the creation of sustainable and resilient cities.

This article is structured as follows: in Section 2, we review relevant literature and previous studies that inform our research. Section 3 provides a comprehensive overview of the datasets utilized in our study, detailing their sources and relevant characteristics. Moreover, it describes the analysed context. Section 4 outlines the various regression and machine learning techniques employed to investigate the relationship between PM2.5 levels and the identified influencing factors. In Section 5, we present our findings regarding the main drivers of PM2.5 levels, highlighting the significant impact of meteorological factors as well as socioeconomic and production structure variables. Section 6 interprets these results in the context of existing literature, providing insights into the implications of our findings for policy and practice. Finally, we conclude with a summary of our key findings and their relevance to future research and policy initiatives aimed at improving air quality in Section 7.

2 Related Work

Air quality has gained momentum among policymakers' agendas. Indeed, achieving carbon abatement has become a pivotal mission to comply with international targets and effectively deal with environmental issues such as climate change. Consistently, the academic literature has widely scrutinized the main factors influencing the emissions level measured across territories (IPCC, 2014, European Commission, 2018).

Air quality is a complex phenomenon influenced by an interplay of natural and anthropogenic factors. Meteorological conditions create a dynamic and context-specific framework for air quality by determining the dispersion, dilution, and deposition of pollutants (Danek et al., 2022, Nguyen et al., 2024, Tian et al., 2024). For instance, strong winds can transport contaminants away from their source, reducing local concentrations, while stagnant conditions can exacerbate pollution levels (Xie et al., 2022, Jin et al., 2024). Rainfall acts as a natural cleansing mechanism by washing particulate matter and gaseous pollutants out of the atmosphere through wet deposition, with long-duration heavy precipitation having the most pronounced impact (Fan et al., 2004, Guo et al., 2016, Lin et al., 2025). Humidity and temperature levels also affect air quality, as they impact on chemical reactions related to the formation of secondary pollutants (Zhang et al., 2017, Łowicki, 2019, Zender-Świercz et al., 2024).

Recent studies provide additional empirical support for these dynamics. For example, Bamola et al. (2024) find that the PM2.5/PM10 ratio is negatively correlated with wind speed and temperature, but positively associated with relative humidity, with the highest pollution levels occurring under low wind speeds and high relative humidity in winter and post-monsoon seasons. Similarly, Li et al. (2025) demonstrate that both the intensity and duration of precipitation events significantly enhance PM2.5 removal through wet deposition, though certain conditions (e.g., drizzle following clean episodes) may conversely exacerbate pollution due to secondary aerosol formation. Using a high-resolution modeling framework, Fang et al. (2024) confirm that lower wind speeds, lower temperatures, and reduced boundary layer heights contribute to PM2.5 accumulation, particularly in winter. The work of Nguyen et al. (2024) further supports these findings, showing that PM2.5 concentrations are most sensitive to humidity and temperature, especially during pollution episodes, where high relative humidity and low temperatures synergistically elevate PM2.5 levels. Additionally, Fattah et al. (2023) report strong negative correlations between PM2.5 and meteorological factors across South Asian cities, with humidity and temperature being the most influential parameters. Similar spatially heterogeneous results are observed by Jin et al. (2022), who find wind speed and temperature to be more strongly associated with PM2.5 in northern and northwestern China, whereas humidity dominates in southern regions. Notably, studies using process-based models also reveal that temperature modulates vertical mixing and heterogeneous chemistry affecting PM2.5, while wind speed and planetary boundary layer height influence the dilution capacity of the atmosphere (Ma et al., 2021). Overall, these findings reinforce the notion that meteorological parameters exert substantial and multifaceted control over PM2.5 pollution dynamics, with varying mechanisms across spatial and seasonal contexts.

Also, socioeconomic factors such as the level of wealth within a region strongly correlate with air quality, albeit in complex ways. High-income regions experiencing rapid industrialization, high production levels as well as high life standards, can spike their emissions level due to increased reliance on energy-intensive sectors. However, they also benefit from advanced pollution control technologies and stringent environmental regulations, potentially constraining pollutant emissions per capita (Grossman and Krueger, 1995).

Finally, the economic structure of a region also plays a critical role in shaping its air quality. Regions dominated by manufacturing and heavy industries tend to experience higher pollution levels due to the emission-intensive nature of these activities (Cole et al., 2005). In contrast, a higher presence of knowledge-intensive sectors—such as information technology, research and development, and finance—often corresponds to improved air quality. These sectors are less reliant on fossil fuels and generate fewer direct emissions, contributing to a cleaner environment (Vona et al., 2018). Moreover, knowledge-intensive industries often drive innovation in clean energy technologies and sustainable practices, creating spillover effects that benefit other sectors (Jaffe et al., 2005). Encouraging a shift toward such industries can thus serve as an effective strategy for long-term pollution reduction.

In this direction, the economic framework of servitization refers to the growing trend of firms providing more comprehensive market offerings, or "bundles," that integrate goods, services, support, self-service, and knowledge to enhance customer value. The core principle of these market strategies is that manufacturing firms enhance the value of their physical products by reducing the focus on production activities but integrating services (Vandermerwe and Rada, 1988). Several authors have assessed how servitization reshapes the productive configuration of manufacturing firms (Bellandi and Santini, 2019), improves productivity in local value chains (Lombardi et al., 2022), and increases the economic profit, especially in small firms (Neely, 2008). Furthermore, it has been highlighted how servitization can activate the interaction between manufacturing and knowledge-intensive business sector, thus fueling territorial development (Lafuente et al., 2017) through the growth of local markets and positive repercussions on economic activities (Gomes et al., 2019).

Despite this wide literature on the economic consequences of servitization (Sforzi and Boix, 2019, Friesenbichler and Kügler, 2022), extant studies do not have widely clarified how the growth of service sectors with respect to manufacturing industries may contribute to reducing air pollution. Limited exceptions are represented by Wang et al. (2023), Palafox-Alcantar et al. (2024), and Vaillant and Lafuente (2024), who discuss how servitization may improve energy transition, conservation, as well as waste and emission reduction by fostering the digital transformation, the optimization of the production structure and the adoption of green technological innovation. Moreover, other authors discuss how servitization may contribute to sustainable circular economy business models (Bressanelli et al., 2024, Sgambaro et al., 2024, Stabler et al., 2024). Although such benefits are well-documented when focusing on firms, we still neglect scientific evidence on the impact of servitization on air quality measured at the municipality level.

Territorial servitization introduces value-adding services that can lower emissions by reducing reliance on material goods (Roos, 2015). Moreover, it often substitutes physical inputs with clean service factors like intellectual, human, and technological capital, while digital transformation supports a circular economy and reduces resource waste (Yusliza et al., 2020, Hao et al., 2021, Paschou et al., 2020, Bressanelli et al., 2024). As a consequence, our research hypothesis is that larger territorial servitization might be associated with better air quality.

Overall, by disentangling how servitization may affect the concentration of $PM2.5$ emissions while controlling for a robust set of socioeconomic and meteorological factors, we aim to offer valuable insights for policymakers on how specific adjustments to the local economic structure may contribute to improving the local air quality.

3 Data

3.1 Dependent Variable

Our dependent variable is the average municipal-level PM2.5 emissions measured over the period 2014–2020.[2] PM2.5 is selected since it is a pollutant directly monitored according to the United Nations' 11^{th} Sustainable Development Goal, which emphasizes the creation of sustainable and resilient cities.

Furthermore, focusing on PM2.5 is particularly appropriate for a municipal-level analysis because it is a local pollutant with limited atmospheric dispersion compared to more volatile gases such as CO_2, NO_x, or SO_x. While these gases can travel long distances and are influenced by regional or even global atmospheric processes, PM2.5 concentration is highly sensitive to local emission sources and to local meteorological conditions (Choi et al., 2021, Zhao et al., 2021, Roostaei et al., 2024). This makes PM2.5 a more spatially specific indicator of air quality, well-suited to capturing the impact of socioeconomic, demographic, and environmental variables at the municipal scale. Our focus aligns with a growing body of research that has analyzed the relationship between economic, social, meteorological, and demographic factors and air quality using PM2.5 levels as the dependent variable. Indeed, previous research has investigated the role of urbanization and demographic patterns (Chen et al., 2018, Shi et al., 2021), industrial controls (Zhang et al., 2017), and the interaction of anthropogenic and natural drivers (Wang et al., 2018) in shaping PM2.5 concentrations. These works confirm that PM2.5 serves as a key environmental outcome through which broader socioeconomic and environmental dynamics can be examined.

Additionally, PM2.5 is a primary determinant of mortality (González et al., 2015, Tarín-Carrasco et al., 2021, Beloconi and Vounatsou, 2023), thus being monitored by the European Environment Agency (EEA) as a benchmark for the European Commission's environmental targets. Its relevance for public health, combined with its responsiveness to local factors, has led to its widespread use in urban and environmental studies aimed at informing place-based policy interventions (Li et al., 2021, Zhang et al., 2019). Thus, the choice of PM2.5 allows for both a scientifically grounded and policy-relevant investigation of air quality determinants at the local level.

We construct our dependent variable using raster data disclosed by EEA.[3] Specifically, we utilize annual air pollutant concentration grids with a spatial resolution of 1km×1km, generated through an advanced "regression-interpolation-merging mapping" technique. This approach integrates air quality monitoring data with external

[2] We stop our main analysis in 2020 since the following years for which PM2.5 information is available (e.g., 2021 and 2022) are significantly affected by the COVID-19 pandemic and we prefer to exclude this period from our study. However, Figures B4, B10, and Table C4 provide robustness checks where data are averaged over the period 2014-2022 thus including the COVID-19 pandemic.

[3] Input data and related information are available at the following link: `https://www.eea.europa.eu/en/datahub/datahubitem-view/82700fbd-2953-467b-be0a-78a520c3a7ef`.

covariates, such as topographical and demographic factors, to enhance spatial accuracy and reliability (Horálek et al., 2024).

The spatial resolution of EEA data does not align precisely with the boundaries of Italian municipalities. To address this mismatch, the annual average PM2.5 concentration is estimated through an *Area-Weighted Interpolation* method applied to the pollutant concentration grids (Pebesma and Bivand, 2023, Chapter 5). In particular, we calculate the PM2.5 concentration for the j-th municipality ($Y(M_j)$) as:

$$Y(M_j) = \sum_{i=1}^{n_C} \gamma_{ij} Y(C_i) \tag{1}$$

More in detail, $\{C_i\}_{i=1}^{n_C}$ constitutes a collection of spatially indexed cells in which a variable $\{Y(C_i)\}_{i=1}^{n_C}$ is measured. In our paper, C_i indicates the $i-$th pixel in the EEA grid, and $Y(C_i)$ represents the recorded value of PM2.5 concentration. Moreover, $\{M_j\}_{j=1}^{n_M}$ constitutes the spatial grid of Italian municipalities where we project the Y variable. Finally, γ_{ij} is the fraction of the surface shared between C_i and M_j. Moreover, given that the raster cells are significantly smaller than the overall study area and that the values $Y(C_i)$ can be assumed to remain constant within each pixel at this spatial resolution, the weights γ_{ij} can be computed as:

$$\gamma_{ij} = \frac{|C_i \cap M_j|}{|C_i|} \tag{2}$$

where $C_i \cap M_j$ is the spatial overlap between the i-th pixel and j-th municipality and $|\cdot|$ indicates the surface of the underlying region.

Even if in Italy, the national statistical office (ISTAT) does not disclose data related to PM2.5 emissions at the municipal level and other international repositories (e.g., the Copernicus Atmosphere Monitoring Service) offer real-time air quality data but at a coarse resolution, we verify the quality of our estimates. In particular, ISTAT provides information about the maximum PM2.5 concentration observed in a year and about the number of days in which the PM2.5 concentrations exceed health limits. We thus assess the correlation between our data aggregated at the province level (average PM2.5 concentration) with these two variables. Figure A2 highlights the strong positive relationship between our data and the other two variables. In particular, we observe a Spearman correlation ρ equal to 0.81 (p-value = 0.000) when comparing average PM2.5 concentration with the maximum PM2.5 concentration in a year. Moreover, we find a Spearman correlation ρ equal to 0.71 (p-value = 0.000) when comparing average PM2.5 concentration with the number of days in which the PM2.5 concentration exceeds health limits.

3.2 Explanatory Variables

We aim to explain the air quality at the municipal level through a wide set of meteorological and socioeconomic variables.

Our key variable of interest is represented by the territorial *Servitization* since we expect that a higher concentration of KIBS activities may shrink the pollutants concentration. In particular, we measure territorial servitization as the difference between the number of employees in KIBS sectors and manufacturing activities, normalized by the total labour workforce. Our definition of KIBS sectors is based on the work by Friesenbichler and Kügler (2022) who identify high-tech (NACE rev.2 sectors 59, 60, 61, 62, 63, 72), market (NACE rev.2 sectors 50, 51, 69, 70, 71, 73, 74, 78, 80), financial (NACE rev.2 sectors 64, 65, 68), and other (NACE rev.2 sectors 58, 75, 84, 85, 86, 87, 88, 90, 91, 92, 93) services. Concerning manufacturing activities, we refer to Section C of the NACE rev. 2, including codes from 10 to 33.

In terms of socioeconomic factors, we also consider the income per capita (*Income pc*) since a higher level of wealth might be associated to more intensive production activities and stronger pollutant concentration. As the local production structure might influence the air quality, we also include an indicator of related variety (*RV*) measuring the economic variety within each business sector with a two-digit detail level. Such a variable is expected to capture the strength of economic connections between sub-sectors, as knowledge spills over primarily between related products (Frenken et al., 2007). In formula:

$$RV_j = \sum_{g=1}^{G} P_{g,j} * H_{g,j} \tag{3}$$

where:

$$P_{g,j} = \sum_{i \in S_{g,j}} p_i \tag{4}$$

$$H_{g,j} = \sum_{i \in S_{g,j}} \frac{p_{i,j}}{P_{g,j}} * \log_2\left(\frac{1}{p_{i,j}/P_{g,j}}\right) \tag{5}$$

and $p_{i,j}$ is the share of the two-digit sector i in municipality j, $S_{g,j}$ is a one-digit sector aggregation, with $G = 1, ..., G$.

We then include the *Maqi* index, which evaluates the quality of municipal administrations by examining several dimensions concerning bureaucratic quality and capacity, local politicians' valence attributes, and local governments' economic and fiscal performance (Cerqua et al., 2025). We expect the quality of municipal institutions may contribute to improving the efficiency of the local business environment, thus affecting the level of emissions.

Concerning meteorological factors, we include the *Wind speed* and the level of *Rainfall* as they may favor a dispersion of alternative pollutant agents (Chai et al., 2014, Zhang and Tan, 2016). Furthermore, we consider *Humidity* since it may increase atmospheric turbulence, resulting in a faster spreading of contaminants (Zhang et al., 2017).

Table 1: We show descriptive statistics for our dependent and explanatory variables.

	Q1	Mean	Q3	Sd
PM2.5	8.373	9.977	11.432	2.437
Income pc	12,173.990	13,647.850	14,807.880	2,080.689
RV	-8.343	-6.572	-4.705	2.409
Population	1,225	8,083.246	6,787	28,894.380
Inactive population	51.031	56.838	60.667	9.347
Wind speed	2.172	2.548	2.834	0.562
Humidity	73.773	76.609	79.881	4.404
Rainfall	8.342	9.679	10.412	1.513
Maqi	100.938	103.118	105.377	3.086
Servitization	0.137	0.198	0.242	0.091

Finally, we employ demographic factors related to the total number of inhabitants *Population* of the analysed municipalities to account for size effects. Moreover, we encompass the *Inactive population*, denoting the percentage of inactive population, which refers to individuals with an age below 14 and above 65 years old.

Table 1 provides an overview of the descriptive statistics for the dependent variable and our economic, demographic, and meteorological regressors. All these factors are measured as the average of the values observed over the time frame 2014-2020 to avoid that specific fluctuations in some years may significantly affect our results. Our sample is consistent with the dataset recently published by Amaddeo et al. (2024), disclosing several information about the economic, human, social, and physical capital of Italian municipalities.

We also apply winsorization by replacing observations below the 2.5$^{\text{th}}$ percentile with the 2.5$^{\text{th}}$ percentile value, and those above the 97.5$^{\text{th}}$ percentile with the 97.5$^{\text{th}}$ percentile value for all our variables. This approach has been recently applied in the literature (see e.g., Flori and Scotti (2025)). We also run our model keeping the original dataset without winsorization. Results were very stable and are available upon request.

3.3 The Italian Context

We focus on Italy as our study area, since it offers a particularly compelling case for examining the drivers of PM2.5 due to its persistently poor air quality relative to other European countries. According to the European Environment Agency (EEA), in 2021, Italy reported one of the highest average concentrations of PM2.5 among

EU member states, reaching 13 μg/m^3, well above the World Health Organization's recommended annual threshold of 5 μg/m^3 and among the worst values in Western Europe (Chiarini et al., 2020, 2021).[4] Italy's geographic and climatic diversity provides an ideal context for analyzing spatial variation in air quality. The country spans from the Alpine regions in the north to the Mediterranean coasts in the south, encompassing a wide range of topographic and climatic conditions.

Northern territories, particularly the Po Valley, which includes Lombardy, Veneto, and Emilia-Romagna, are characterized by dense industrial activity, high population density, and a basin-like morphology that traps pollutants, especially during winter months marked by thermal inversions and stagnant atmospheric conditions. In contrast, central and southern areas benefit from milder Mediterranean climates, greater wind intensity, and more effective atmospheric dispersion, which tend to mitigate pollution accumulation. Appendix A provides detailed information related to climate, economic, social, and demographic characteristics of Italian municipalities.

Table A1 confirms that provinces such as Cremona and Milan consistently report PM2.5 concentrations close to the European Commission's limit of 25 μg/m^3. By contrast, municipalities in the Centre and South exhibit significantly lower pollution levels, averaging around 10 μg/m^3, with provinces from Sardinia and Abruzzi recording the best air quality conditions (see Tables A2 and A3). These patterns are consistent with aggregated data reported by ISTAT.[5]

Climatic differences also play a role in shaping pollution exposure. The South exhibits the highest average wind speed (2.55 m/s), likely contributing to more efficient pollutant dispersion, along with higher humidity levels. In contrast, the North receives more rainfall but suffers from lower air circulation (see Tables A4–A6). The structure of local economies also varies: the North shows the lowest servitization index (0.167), indicating a heavier reliance on manufacturing, whereas the South exhibits a relatively higher shift toward service-oriented activities (0.198). The North records the highest per capita income (€19,424), reflecting stronger economic development compared to the Centre (€17,054) and the South (€13,648). The North's higher related variance (RV) signals greater sectoral interdependence, reinforcing the complexity of its economic dynamics. While the Centre shows a slightly higher average population, driven by large municipalities such as Rome, the North and South display more balanced distributions, with lower standard deviations.

Overall, these structural, climatic, and geographic differences jointly contribute to the spatial disparities in air quality outcomes observed across Italian municipalities.

[4] Air pollution concentration data are available at: `https://www.eea.europa.eu/publications/europes-air-quality-status-2023`.

[5] See the ISTAT BES 2020 Report: `https://www.istat.it/it/files/2021/10/BES-Report-2020.pdf`.

4 Methods

To investigate the determinants of PM2.5 concentration across municipalities, this study adopts a multi-step empirical strategy combining machine learning, regularized regression, and efficiency analysis (see Figure 1). The first step employs machine learning techniques to identify the main socioeconomic, meteorological, and environmental factors associated with PM2.5 levels. These methods are particularly suited to capture complex, nonlinear relationships and interactions within high-dimensional data, offering a data-driven approach to variable selection. To corroborate and interpret the findings from machine learning, the analysis then turns to regularized regression methods—namely ridge and lasso—which allow for greater transparency in estimating the direction and magnitude of associations while mitigating issues of multicollinearity and overfitting. Finally, building on the insights into PM2.5 drivers, a non-parametric Data Envelopment Analysis (DEA) is used to assess the relative efficiency of municipalities in achieving lower levels of air pollution given their specific structural characteristics. This combined approach enables both the identification of key pollution drivers and the evaluation of how effectively local contexts translate structural conditions into environmental outcomes.

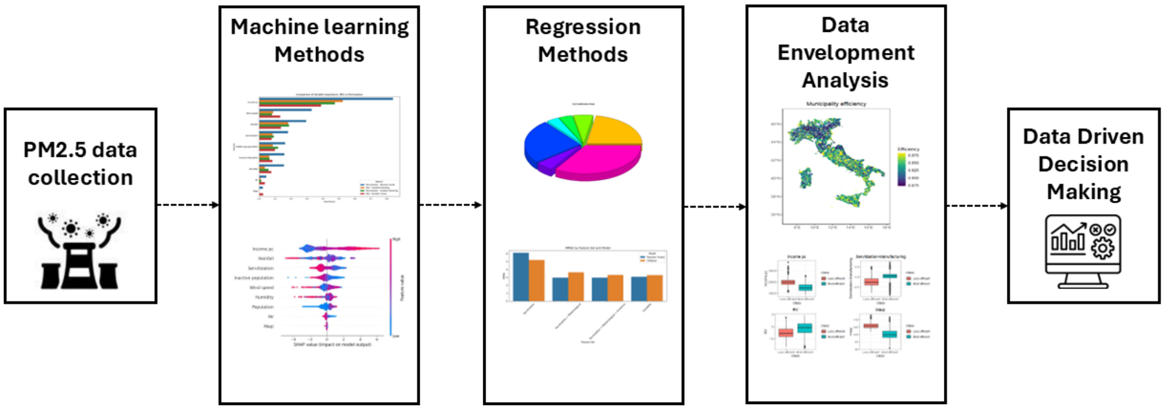

Figure 1: The methodological framework.

4.1 Machine Learning Methods

In the context of our analysis, machine learning methods play a crucial role in achieving predictive accuracy and capturing complex relationships between PM2.5 levels and various influencing factors. Unlike traditional regression techniques, machine learning algorithms can model non-linear relationships and interactions among predictors without requiring explicit specification of these relationships.

Indeed, traditional parametric models impose restrictive assumptions such as linearity and additivity, limiting their ability to capture these intricate dependencies unless explicitly modeled through interaction or polynomial terms, which rapidly increase model complexity and risks overfitting. In contrast, machine learning algorithms inherently approximate the unknown underlying function $f(\mathbf{x})$ mapping predictors $\mathbf{x}$ to PM2.5

concentrations by adaptively partitioning the feature space and aggregating multiple weak learners. This non-parametric, data-driven approach allows for automatic detection and modeling of high-order interactions and non-linearities without prior specification, thus providing a more flexible and accurate representation of the pollutant generation and dispersion processes.

This flexibility allows for improved performance, particularly in high-dimensional datasets where traditional models may struggle.

Thus, in this context, ensemble machine learning methods such as Random Forest and XGBoost are particularly well-suited because they inherently model non-linear relationships and complex feature interactions without requiring prior specification. Random Forest, through its aggregation of multiple decision trees built on bootstrapped samples, effectively reduces variance and enhances robustness against overfitting, which is crucial when dealing with noisy environmental data. XGBoost, a gradient boosting framework, further improves predictive accuracy by sequentially correcting errors of previous trees and incorporating regularization techniques to prevent overfitting. Both methods also provide interpretable measures of feature importance, enabling a deeper understanding of the drivers of air pollution. Their proven success in various environmental modeling tasks and ability to handle high-dimensional data make them ideal choices for accurately predicting PM2.5 levels in Italian municipalities.

Formally, let us denote the PM2.5 concentration at location i as y_i, and the vector of predictors (e.g., meteorological variables, servitization index, demographic controls) as $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{ip})$.

Ensemble machine learning methods like Random Forest and XGBoost implicitly model a more general function:

$$y_i = f(\mathbf{x}_i) + \varepsilon_i,$$

where $f(\cdot)$ belongs to a rich class of non-linear functions constructed by aggregating decision trees and ε_i is an error term. Each decision tree partitions the predictor space into regions with approximately constant response values, effectively capturing complex interactions and non-linearities without requiring explicit model specification.

For example, Random Forest builds T trees, each trained on a bootstrap sample and a random subset of features, and predicts via averaging:

$$\hat{y}_i^{RF} = \frac{1}{T} \sum_{t=1}^{T} h_t(\mathbf{x}_i),$$

where $h_t(\cdot)$ is the prediction from the t-th tree. This averaging reduces variance and improves generalization. XGBoost, on the other hand, sequentially builds trees to correct residual errors, optimizing an objective function with regularization:

$$\hat{y}_i^{XGB} = \sum_{m=1}^{M} h_m(\mathbf{x}_i),$$

where each h_m is a tree added to minimize the loss plus a regularization term controlling complexity.

Among the various machine learning techniques available, we implement Random Forest and XGBoost due to their robustness, interpretability, and effectiveness in regression tasks. In particular, in our analysis, we employ the Random Forest model with specific parameters: 100 trees (or estimators) and a random state of 42.[6]

The choice of 100 trees is made based on empirical evidence suggesting that having a larger number of trees generally leads to improved performance by reducing variance without significantly increasing bias. The averaging effect of multiple trees helps mitigate overfitting, especially in high-dimensional spaces where noise can adversely affect predictions. While increasing the number of trees can improve accuracy, it also increases computational cost.

XGBoost is employed since it provides regularization parameters that help prevent overfitting, making it particularly suitable for our analysis where interpretability and accuracy are paramount. Similarly to the Random Forest model, we implement the XGBoost model with 100 estimators and a random state of 42.

Both Random Forest and XGBoost rely on the whole set of explanatory variables introduced in Section 3.2 including servitization as the main factor of interest and socioeconomic, meteorological, and demographic controls.

We investigate the key factors influencing PM2.5 levels based on our analysis using both Mean Decrease Impurity (MDI) and permutation importance methods for the Random Forest and Gradient Boosting models. These techniques provide complementary insights into feature significance, allowing us to assess the impact of various predictors on air quality. To evaluate feature importance, we first calculate MDI, which measures the average decrease in impurity brought by each feature across all trees in the model. This method is particularly useful for understanding which features are most influential in splitting the data during tree construction. We also employ permutation importance, which assesses how the model's performance changes when the values of a specific feature are randomly shuffled. This approach provides a more direct measure of how much each feature contributes to the model's predictive accuracy. We perform 30 repetitions for each permutation to ensure statistical reliability in our results.

In addition to using MDI and permutation importance methods, we further enhance our understanding of feature significance by employing Shapley Values. They provide a unified measure of feature importance that

[6] The random state parameter, set to 42, ensures reproducibility of results by controlling the randomness involved in bootstrapping samples and feature selection during tree construction.

quantifies the contribution of each feature to the model's predictions. By calculating Shapley values, we can not only determine which features are important but also understand how they influence the predictions in a more interpretable manner. Indeed, positive Shapley values indicate that a feature contributes positively to the prediction, while negative values suggest a detrimental effect. This allows us to visualize not only the importance of features but also their directionality in influencing PM2.5 levels.

The suitability of these machine learning models for PM2.5 prediction is well established. For example, Makhdoomi et al. (2025) evaluate four ensemble methods, including Random Forest and XGBoost, and find them highly effective in forecasting PM2.5 with over 95% of predictions within the range of observed concentrations. Yang et al. (2024) use Random Forest and XGBoost to assess the relative importance of meteorological and satellite-based predictors, finding XGBoost achieves the highest accuracy. Similarly, Peng et al. (2022) apply XGBoost to PM2.5 data, demonstrating its superior performance in capturing daily and nighttime pollution patterns. Kim et al. (2022) find gradient boosting models outperform chemical transport models, confirming the reliability of tree-based ML in short-term air quality prediction. Even in studies exploring advanced models like deep learning or Bayesian frameworks, such as Lin et al. (2025), tree-based methods like Random Forest and XGBoost remain highly competitive, offering an effective balance of interpretability, flexibility, and predictive strength.

4.2 Regression methods

In a second step, we utilize Ordinary Least Squares (OLS) regression, along with its regularized variants, Ridge and Lasso approaches. These methods are chosen for their ability to handle multicollinearity and improve prediction accuracy in the presence of numerous predictors. The regression framework is defined by the following equation:

$$Y_i = \alpha + \beta * Servitization_i + \gamma * E_i + \delta * D_i + \theta * M_i + \epsilon_i,$$

where Y_i represents the observed level of PM2.5 for the i-th observation, $Servitization_i$ is our variable of interest and represents the relative percentage of employees in KIBS sectors with respect to those employed in manufacturing activities in municipality i, as introduced in Subsection 3.2. Moreover, E_i is a vector of economic factors including income per capita (*Income pc*), the related variety indicator (*RV*), and the *Maqi* index explained in Subsection 3.2. In addition, D_i is a set of demographic factors encompassing the total population (*Population*) and the inactive population, while M_i takes into account meteorological factors such as *Wind Speed*, relative humidity (*Humidity*), and the *Rainfall* level. Finally, ϵ_i captures the error term.

We rely on the Ridge regression, since it helps reduce variance at the cost of introducing some bias. This trade-off is crucial in high-dimensional datasets where traditional OLS estimates may become unstable. We also utilize Lasso regression since it also performs variable selection by shrinking some coefficients exactly to zero. This feature makes Lasso particularly useful for identifying significant predictors in high-dimensional datasets. We employ cross-validation to determine optimal values for the tuning parameter α, using a range of values generated logarithmically from 10^{-4} to 10^{4}. This approach ensures that we explore a broad spectrum of regularization strengths while avoiding overfitting.

The choice of OLS, Ridge, and Lasso regression stems from their complementary strengths. While OLS provides unbiased estimates, it can be unstable in high-dimensional settings. Ridge regression mitigates this issue through regularization, reducing variance without eliminating predictors. Lasso regression not only reduces variance but also performs variable selection, leading to more interpretable models. Overall, we rely on these methods since they are widely used in environmental and policy-oriented econometric research due to their interpretability, predictive strength, and robustness to multicollinearity and model uncertainty.

The use of OLS is well established in the literature for modeling linear relationships between variables, particularly when interpretability and hypothesis testing are prioritized. For instance, Vaillant and Lafuente (2024) adopts OLS to assess the environmental performance implications of digital servitization, emphasizing its clarity and flexibility in analyzing firm-level survey data. Similarly, Zhao et al. (2022) employ an extended OLS framework to disentangle the effects of urbanization, income inequality, and energy efficiency on carbon emissions. Lasso and Ridge regression are increasingly applied to improve model performance in high-dimensional settings, as seen in Aller et al. (2021), who uses Cluster-LASSO for variable selection and robustness. These techniques have been effectively applied in similar studies exploring environmental determinants and policy impacts (Wang et al., 2023, Yao et al., 2023), supporting their relevance for our empirical strategy.

4.3 Data envelopment analysis

Finally, we investigate the municipal efficiency in achieving a certain level of air quality given a specific socioeconomic structure. We do this by estimating the efficiency score based on the Data Envelopment Analysis (DEA) framework (Charnes et al., 1978, Banker et al., 1984). DEA is a mathematical programming technique rooted in microeconomic theory that operates without parametric assumptions, accommodates multiple inputs and outputs, and is used to evaluate the relative technical efficiency of decision-making units (DMUs). The core idea of DEA is to construct an empirical, piecewise-linear production boundary by determining the best linear combinations of various DMUs (such as municipalities in this context) and evaluating these synthesized or “virtual” DMUs against the actual ones in the dataset.

Technical efficiency θ_j is estimated according to Equations (1)–(4) (variable returns to scale (VRS)):

$$\min_{\theta,\lambda} \theta_j,\ subject\ to: \tag{6}$$

$$-\frac{y_{m,j}}{\theta_j} + \sum_{i=1}^{J} \lambda_i y_{m,i} \geq 0, m = 1, ..., M, \tag{7}$$

$$x_{n,j} - \sum_{i=1}^{J} \lambda_i x_{n,i} \geq 0, n = 1, ..., N, \tag{8}$$

$$\lambda_i \geq 0, i = 1, ..., J\ and\ \sum_{i=1}^{J} \lambda_i = 1 \tag{9}$$

where j denotes the decision making unit (DMU), m is the index for output, λ_j is the weight for the j^{th} municipality and n is the index for input (e.g. $y_{m,j}$ is the m^{th} output and $x_{n,j}$ is the n^{th} input of the selected j^{th} DMU). In particular, our output is the inverse of PM2.5 levels of Italian municipalities.[7] In terms of input, we take into account the same set of explanatory variables introduced in Section 3.2, including economic, demographic, as well as meteorological factors.

We also do some robustness checks by incorporating emission-specific inputs to better isolate the environmental efficiency dimension in the DEA analysis. In particular, we incorporate meaningful proxies that capture key aspects of emission intensity. Specifically, we include the number of vehicles per municipality as a proxy for transportation density and the total expenditure for energy consumption at the municipal level.[8]

Furthermore, given our choice of inputs, the measured environmental efficiency of municipalities may partly capture structural differences—such as lower industrial intensity—rather than reflecting genuine efficiency in pollution reduction.

To address this potential bias, we conduct a robustness check by re-estimating the DEA model after excluding the economic variables most directly associated with industrial structure and economic capacity—specifically, servitization, RV, and income per capita. This approach allows for a clearer separation of environmental efficiency from underlying structural economic characteristics.

We apply an "output-oriented" DEA, meaning that a real DMU realizing a higher output (level of PM2.5) with the same amount of inputs is defined as "dominated" and is positioned at the interior of the empirical efficiency frontier. By evaluating all DMUs in the dataset, a production frontier is established that encloses the observed data. Radial technical efficiency scores, ranging from 0 to 1, are then determined. These scores indicate how far each DMU is from the efficiency boundary, with those on the frontier achieving the maximum score of 1. DEA offers several benefits that make it a valuable tool for evaluating municipal spending efficiency. One key advantage is its ability to handle multiple inputs simultaneously, effectively simplifying complex data into a single efficiency measure. Additionally, unlike regression-

[7] We have inverted the orientation of the PM2.5 concentration to reflect the correct interpretation of the problem. Indeed, municipalities with lower levels of PM2.5 experience better air quality and should be considered as more efficient, ceteris paribus, economic, demographic, and meteorological factors.

[8] Data on total energy expenditure are not available at the municipal level and are reported by the National Statistical Office only at the regional scale. We therefore allocate regional energy expenditure to municipalities proportionally to their share of the regional population.

based methods, DEA provides greater flexibility by not requiring a predefined functional form for the process contributing to PM2.5 generation or making assumptions about error distributions.

Due to the sensitivity of traditional DEA to outliers and measurement errors, resulting in biased efficiency estimates, we apply the Robust DEA (R-DEA) method by Simar and Wilson (1998, 2002, 2011) to overcome these limitations. We estimate the robust (bias-corrected) environmental efficiency scores through 100 bootstrap resampling iterations.

5 Results

In this section, we present the findings from our analysis of the factors influencing PM2.5 levels, as well as an evaluation of the performance of the machine learning models employed in our study. The results are organized into three main subsections. Subsections 5.1 and 5.2 discuss the significant variables affecting PM2.5 concentrations based on machine learning and regression analyses, respectively. Subsection 5.3 assesses the determinants of environmental efficiency of Italian municipalities.

5.1 The main drivers of PM2.5 based on machine learning methods

Figure 2 indicates a hierarchy of variables that significantly affect air quality, ranked from most to least important based on MDI and permutation methods for the Random Forest and Gradient Boosting models.

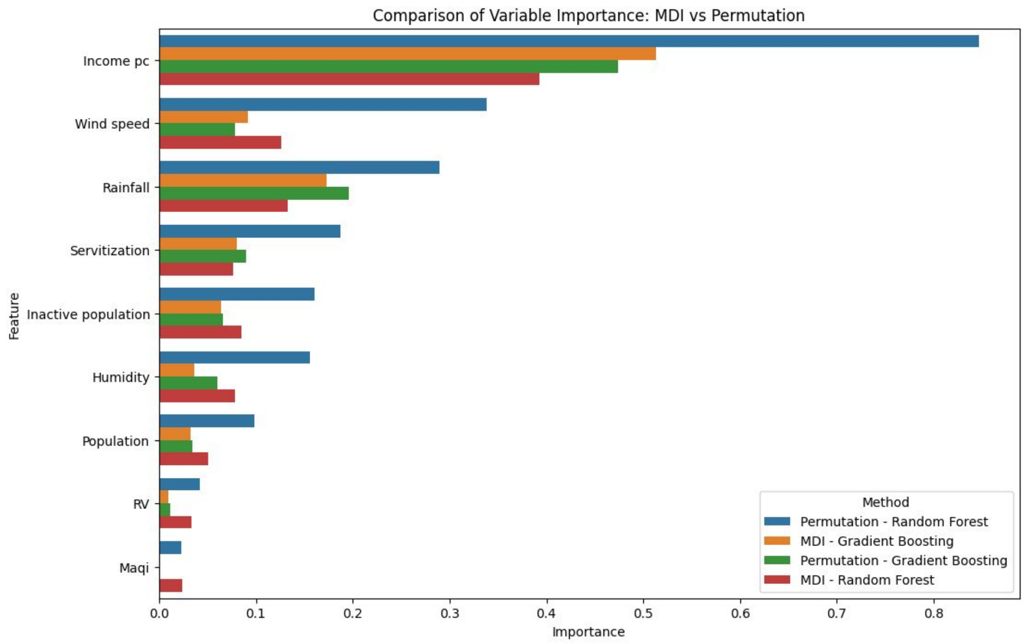

Figure 2: Comparison of Variable Importance: MDI vs Permutation.

The analysis reveals that *income per capita* is the most significant driver of PM2.5 levels with feature importance above 0.4 for all models. Interestingly, it also has the strongest impact on air quality according to Shapley values (see Figure 3). In particular, higher income levels correlate with elevated emissions. This finding aligns with existing literature, which suggests that regions with higher economic activity tend to experience worse air quality due to intensified production processes, as well as increased industrial activity and vehicle usage (Hao et al., 2016, Sapkota and Bastola, 2017).

We also find consistent results regarding the importance of meteorological factors on air quality. Indeed, *wind speed* and *rainfall* emerged as crucial factors. Both increased wind speeds and precipitation levels are associated with improved dispersion of pollutants, since they can effectively wash out airborne particles, leading to lower concentrations of pollutants in the atmosphere (Chai et al., 2014, Zhang and Tan, 2016). We rather obtain a lower impact of *humidity* on air quality, with higher levels of this feature slightly reducing the PM2.5 levels (Zhang et al., 2017, Łowicki, 2019).

However, also the local socioeconomic structure influences air pollution dynamics. Indeed, the variable *servitization* shows an interesting trend where areas with higher service-oriented economic structures experience better air quality. This supports our hypothesis that a greater concentration of services relative to manufacturing activities tends to reduce emissions. This result confirms previous evidence obtained by Wang et al. (2023) showing how servitization contributes to improving energy conservation and emission reduction. Similarly, it is in line with the results of Vaillant and Lafuente (2024), who highlight how digital servitization may help shrink energy consumption and waste generation.

Conversely, factors *RV* and *Maqi* have minimal influence on PM2.5 concentrations, often showing low statistical significance in our analysis. This suggests that local economic variety and bureaucratic quality do not play a key role in determining air quality dynamics at the municipal level in Italy. Finally, higher *inactive population* is associated with lower PM2.5 concentration, whereas the opposite holds for the variable *population*. These results suggest that areas with older citizens and a lower number of inhabitants experience lower air pollution, probably due to more limited households consumption and activities that generate emissions.

We perform three robustness checks with respect to our main analysis on the whole sample of Italian municipalities. First, we repeat our analysis by including in our sample data over the period 2014-2022, thus considering also the COVID-19 pandemic. Interestingly, we observe that the main patterns are confirmed even if meteorological variables assume higher importance with respect to economic and demographic factors (see Figure B4 and B10). This result corroborates the high relevance of human activities in shaping PM2.5 concentration and provides an insight about the fact that some effects in air quality drivers remained visible even after the end of the most critical phases of the pandemic.

Second, we include middle age population as an additional covariate. Figures B5 and B11 confirm our main results in terms of Feature Importance and Shapley values according to both Random Forest and XGBoost models. In particular, according to both machine learning and regression models, middle age population has a slightly positive relationship with PM2.5 concentration, meaning that a higher portion of people in the range 35-60 years old is associated with lower air quality. This result may be explained by the fact that municipalities with more middle-aged workers often see higher levels of commuting and private vehicle use, leading to increased traffic-related PM2.5 emissions. Additionally, this age group tends to drive economic activity and consumption, further contributing to pollution through energy use, heating, and industrial demand.

Third, we substitute the variable servitization with the factor *Servitization variation*, computed as the average variation in employment in KIBS over the period 2014-2020. Figures B6 and B12 suggest a negative and significant relationship between PM2.5 concentration and servitization variation. Indeed, a higher growth in employment in KIBS sectors is associated with larger air quality improvement. However, notice how the relevance of servitization variation is lower with respect to servitization. This result suggests that PM2.5 concentration is more linked with employment levels in KIBS rather than with their variation.

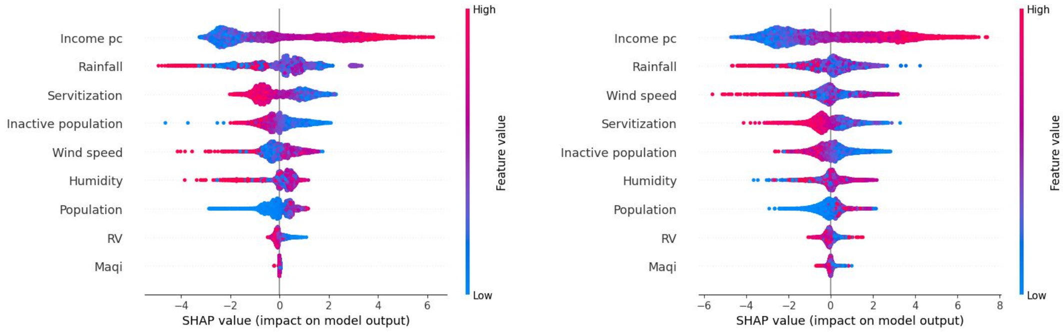

Figure 3: SHAP summary plot for Random Forest Model and Gradient Boosting Model illustrating the contribution of each feature to the model's predictions. Each point represents an individual instance from the dataset. The horizontal position of each point indicates the SHAP value, which quantifies the impact of that feature on the model's output for that instance. Positive SHAP values indicate that the feature pushes the prediction higher, while negative SHAP values indicate the feature pushes the prediction lower. The vertical axis represents the features, ordered by decreasing feature importance as determined by the mean absolute SHAP value across all instances. The color of each point represents the value of the feature for that instance; red indicates a high feature value, and blue indicates a low feature value. Therefore, the plot reveals both the overall importance of each feature and its relationship with the model's output. For example, a cluster of red points on the positive side of the x-axis indicates that high values of that feature tend to increase the model's prediction. The spread of SHAP values on the y-axis reveals the variance/interaction effects with other features in the model.

We also evaluate the predictive performance of our machine learning models by employing several metrics, including Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Adjusted R^2. In particular, we estimate several specifications of our Random Forest and XGBoost models to predict PM2.5 concentrations using a series of nested feature sets. The base specification includes servitization, the variable of main interest, progressively enriched with meteorological, economic, and demographic controls.

Figure B13 clearly shows that models relying solely on servitization perform poorly across all metrics, particularly in terms of predictive accuracy (high MSE/RMSE and low Adjusted R^2). The inclusion of meteorological and economic variables significantly improves performance, especially for Random Forest. However, it is the complete model—incorporating all variable blocks—that consistently achieves the best overall performance, balancing low error rates and high explanatory power across both algorithms.

These findings underscore the importance of accounting for a comprehensive set of covariates when modeling air pollution: while servitization plays a role, omitting meteorological and economic conditions would severely limit the model's explanatory capacity and precision.

To assess the robustness of our findings, we repeat our analysis across different geographical macro-areas, splitting our sample among municipalities located in the North, Centre, and South of Italy. This is particularly relevant, given the heterogeneous socioeconomic structure in Italy and the North-South divide (González, 2011, Rungi and Biancalani, 2019) (see Appendix B).

We observe that in the North of Italy, socioeconomic features still have a pivotal role in governing air quality. For instance, higher income per capita correlates with larger PM2.5 concentration, whereas a higher servitization is clearly achieved in less polluted territories (see Figure B7). Conversely, both in the Centre and South of Italy, we observe a stronger influence of demographic and meteorological factors on air quality (see Figures B8, and B9). This may be due to the relatively balanced local economic structure, which exhibits limited variability and, consequently, may not provide sufficient explanatory power for PM2.5 levels.

5.2 The main drivers of PM2.5 based on regression methods

We aim to corroborate our previous results obtained through machine learning methods through OLS, Lasso, and Ridge regressions.

Table 2: Results for the OLS, Ridge, and Lasso regressions. Models are applied to all Italian municipalities.

Variable	OLS	Ridge	Lasso
Income pc	0.0007***	2.6156***	1.9956***
	(0.0000)	(0.0729)	(0.0568)
RV	-0.3058***	-0.7892***	-0.0319
	(0.0251)	(0.0789)	(0.0434)
Population	-0.0000	-0.1078	
	(0.0000)	(0.1186)	
Inactive population	-0.0792***	-0.7242***	-0.0833
	(0.0060)	(0.0614)	(0.0542)
Wind speed	0.0548	0.0199	
	(0.1164)	(0.0707)	
Humidity	0.0518***	0.1861***	
	(0.0120)	(0.0564)	
Rainfall	-0.3441***	-0.6854***	
	(0.0259)	(0.0551)	
Maqi	-0.0714***	-0.2103***	
	(0.0185)	(0.0583)	
Servitization	-0.4698***	-0.1126***	-0.3811***
	(0.1238)	(0.0250)	(0.0642)
Observations	7,035	7,035	7,035

Notes: *p $<$ 0.1, **p $<$ 0.05, ***p $<$ 0.01. Variables for which no coefficient is reported were not selected by the Lasso regression, indicating that their estimated coefficients are equal to zero.

These analyses confirm the key role of *income per capita* in driving the air quality at the local level, with larger level of wealth usually associated with higher PM2.5 concentration. Similarly, municipalities with stronger rainfall experience better air quality according to both OLS and Ridge regressions, while this variable is not significant according to the Lasso model.

We find some differences with respect to machine learning methods for the other meteorological and socioeconomic factors. Concerning the former, regression models tend to associate stronger statistical significance to humidity with respect to wind speed. Moreover, in terms of socioeconomic variables, *RV* and *Maqi* become relevant drivers of PM2.5 concentration with areas experiencing higher levels of economic variety and institutional quality subject to better air quality. We rather confirm the pivotal role of *servitization*, as all our regression models provide a strong indication in favour of lower levels of PM2.5 in case of a stronger presence of service activities with respect to manufacturing sectors, in line with our main research hypothesis. In particular, across our specifications we find that a 1% increase in the concentration of KIBS sectors is associated with a reduction in terms of PM2.5 levels between -0.47 and -0.11.

When we consider demographic variables, we find that a higher presence of old citizens is associated with better air quality, whereas limited statistical evidence is obtained for the total population, suggesting that PM2.5 dynamics are not necessarily affected by the size of the municipality.

Consistently with section 5.1, we verify the stability of our results through three robustness checks. First, extending the sample to 2014–2022 to account for the COVID-19 period, confirms the main patterns, with meteorological variables gaining importance (see Table C4). Second, adding the middle-aged population share as an explanatory variable supports its positive association with PM2.5, likely due to greater commuting and consumption (see Table C5). Third, replacing servitization with its variation shows a negative relationship with PM2.5, though its explanatory power is lower than the level-based indicator (see Table C6).

Our results tend to be stable also across different geographical macro-areas (see Appendix C). For instance, servitization is significant in all our model specifications with coefficients ranging between -0.38 and -0.08 (see Tables C1, C2, and C3). We also obtain that higher related variety and institutional quality are associated with better air quality according to the OLS and Ridge models, whereas the *RV* and *Maqi* variables are not statistically relevant in the Lasso regression. These results may suggest that socioeconomic factors still drive air quality dynamics in specific spatial macro-areas. In terms of meteorological factors, we find robust results for rainfall and humidity, with the former having a negative effect and the latter a positive one. These effects are statistically significant in the OLS and Ridge regression models, while the Lasso model yields a null coefficient, meaning that they do not have a relevant impact on PM2.5 levels. Mixed results emerge in terms of the wind speed variable. The coefficient is not statistically significant according to the Ridge and Lasso regressions in all macro-areas. Conversely, according to OLS estimates, higher wind speed determines lower PM2.5 concentration in the Centre and South of Italy, whereas it reduces air quality in the North.

5.3 Efficiency analysis on PM2.5

Figure 4 highlights the geographical distribution of the efficiency computed as described in Section 4.3. Higher scores are obtained by municipalities located in the Centre and South of Italy, meaning that they tend to achieve lower emissions levels with respect to other territories displaying a comparable economic structure. In particular, the top five provinces with the highest average efficiency levels are Rieti, Caltanissetta, Catanzaro, Cosenza, and Reggio di Calabria, with values ranging between 0.950 and 0.954. Conversely, territories in the North of Italy exhibit the lowest scores with Reggio Emilia, Treviso, Monza e della Brianza, Milano, and Vicenza constituting the bottom five provinces in terms of efficiency with scores in the range 0.893-0.899.

Environmental efficiency provides complementary information to the PM2.5 level. In particular, environmental efficiency allows us to identify those municipalities obtaining a lower level of PM2.5 while maximizing the exploitation of local socioeconomic factors. This implies that not necessarily municipalities with higher levels of air quality are associated with larger environmental efficiency. Indeed, the two variables achieve a correlation equal only to 0.46.

To investigate which inputs characterize the most efficient units, we divide Italian municipalities in two groups. We mainly focus on economic factors, since we are interested in understanding how alternative socioeconomic structures

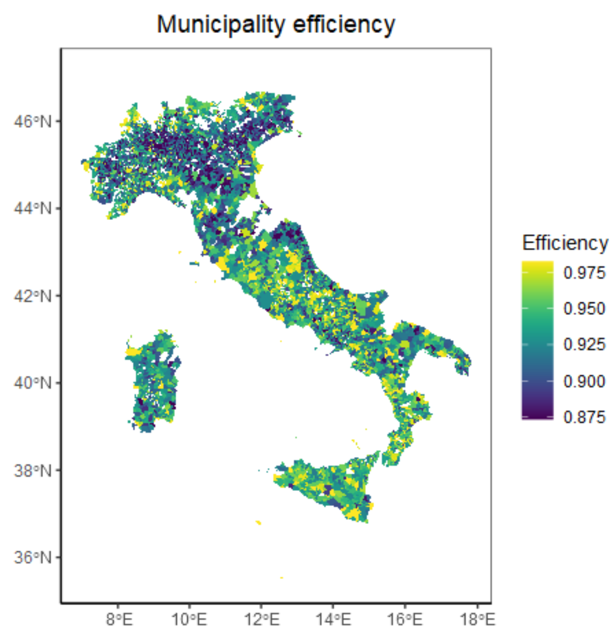

Figure 4: Municipal environmental efficiency.

affect environmental efficiency. Less (most) efficient territories are defined as administrative in the lower (upper) tail of the efficiency distribution, below the 0.2 (above the 0.8) quantile. Figure 5 plots the distribution of the economic variables employed in the efficiency analysis separating the most efficient (in blue) and the less efficient (in red) units, using as a threshold the 20^{th} percentile of the distribution of efficiency scores. We also perform t-test to formally compare whether most and less efficient municipalities display statistically significant different features (see Table D1).

We can see that the most efficient units have greater values of *servitization*. These results highlight how a higher presence of knowledge-intensive sectors with respect to manufacturing activities is associated not only with better air quality but also with better environmental efficiency. This means that given the same socioeconomic factors and meteorological variables, a municipality can obtain a lower concentration of PM2.5 in case it experiences higher servitization.

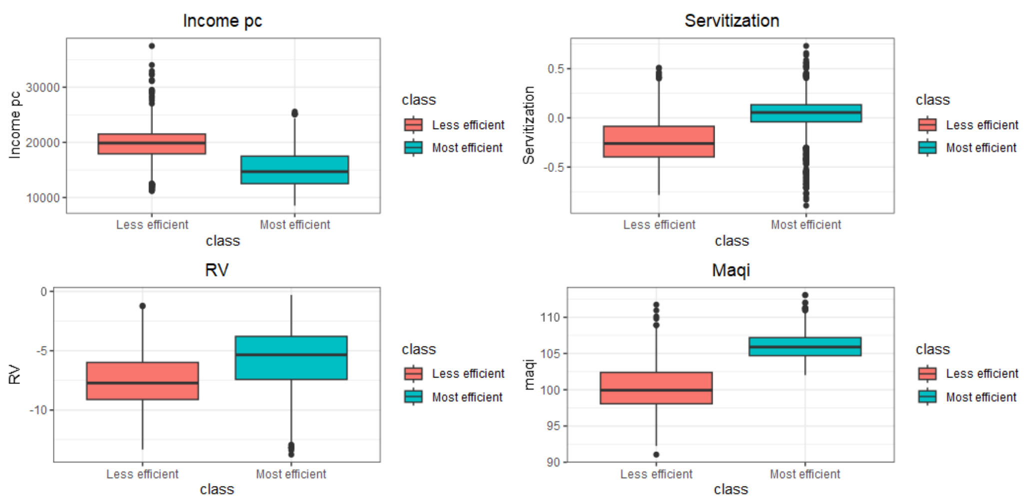

Figure 5: Distribution of socioeconomic variables of most and less efficient municipalities.

We obtain an interesting result also concerning the variable *Maqi*. Indeed, while this variable is not a significant driver of the PM2.5 concentration, especially according to machine learning methods, it has a key role in determining the environmental efficiency of Italian municipalities. Indeed, we observe a clear separation in the distribution of the *Maqi* index between more and less efficient municipalities, with better environmental performance associated with higher institutional quality. Such a result suggests that municipalities with better administrative capabilities can create a better local business environment that may smooth processes and operations, thus contributing to the improvement of local air quality. This finding fuels an emerging debate in the literature, which highlights how high-quality institutions should be regarded as essential inputs for implementing sound legislation, which, when effectively enforced, can help mitigate local environmental degradation (Bernauer and Koubi, 2009, Lau et al., 2014, Ali et al., 2019).

We obtain that richer areas with lower economic variety also exhibit smaller efficiency scores, confirming that the socioeconomic structure of a municipality has a key role in explaining the local environmental efficiency.

As explained in Section 4.3, we also further test robustness by incorporating emission-specific inputs into the DEA model and excluding economic variables most directly associated with local industrial structure to better isolate environmental efficiency.

Figures D4 and D5, as well as Table D2 support our main findings, suggesting that our results are not driven by the specific input mix. In particular, we do not observe statistically significant differences between most and least efficient municipalities in terms of number of vehicles and energy expenditures at the municipality level. Moreover, we confirm that most efficient municipalities are poorer areas, with better institutional quality, higher presence of KIBS sectors, and a more interconnected local economy. Such results hold even when excluding income per capita, servitization, and RV from the input mix.

We repeat the same analysis across geographical macro-areas in Appendix D. Interestingly, we confirm that the *Maqi* index mainly drives environmental efficiency, with better-performing territories exhibiting higher institutional quality (see Figures D1, D2, and D3). Larger servitization still characterizes more efficient areas in the North and Centre of Italy, whereas differences are flattened in the South. Less efficient municipalities tend to be wealthier places in terms of income per capita, except in the Centre of Italy, whereas the related variety does not tend to be a key driver of environmental efficiency across the analysed macro-areas.

6 Discussion

The bulk of the literature on servitization highlights its positive effects on firms' competitiveness based on the introduction of advanced digital technologies, increasing product-service value delivery through smart solutions (Kohtamäki et al., 2020, Favoretto et al., 2022, Kohtamäki et al., 2022, Yang et al., 2024). Conversely, our findings contribute to the emerging evidence on the role of servitization in fostering improvements in non-financial performance. A recent stream of research discusses how the Internet of Things (IoT), Big Data, and Artificial Intelligence (AI) may enable the implementation of the circular economy (CE) paradigm into firms business models underpinning novel opportunities in terms of value discovery, value realization, and value optimization capabilities (Bressanelli et al., 2018, Zheng et al., 2018, Sjödin et al., 2023). Our analysis complements such literature focusing on the impact of servitization on the environmental performance at the territorial level. Our results are consistent with previous findings showing that servitization may promote energy conservation and emissions reduction at the firm level (Wang et al., 2023). Moreover, our findings are in line with former evidence revealing that digital servitization may help shrink energy consumption and waste generation (Vaillant and Lafuente, 2024). While we find that the sectoral composition of local economy drives air quality based on the significant negative impact of servitization on PM2.5 levels, we find that RV is not a key determinant of emissions at the municipality level. Such a result is consistent with several studies indicating that a shift towards manufacturing and industrialized sectors tends to raise pollution, while a less evident impact concerns the level of diversification of the labour market structure (You et al., 2015, Kolcava et al., 2019, Zhao et al., 2022).

Our analysis also suggests a positive relationship between income per capita and PM2.5 levels, meaning that territories with higher wealth experience stronger air degradation. Theoretically, the state of the environment non-trivially depends on the level of economic development since an increase in economic activity is likely to raise pollution, but wealthier territories may adopt more stringent environmental standards and adopt more innovative green technologies. However, several studies obtain that income represents a pivotal driver of emissions (Bae et al., 2017, Özokcu and Özdemir, 2017, Aller et al., 2021).

Territories with high-quality institutions are more likely to have stricter environmental policies and respect international environmental agreements aimed at reducing emissions (Dutt, 2009, Tamazian et al., 2009, Tamazian and Rao, 2010). We confirm such evidence since in our analysis, the maqi index has a negative relationship with PM2.5 concentration and positively contributes to greater environmental efficiency.

Concerning demographic factors, smaller municipalities with a higher presence of inactive people exhibit higher air quality, supporting that urbanization and metropolitan expansion processes may be associated with environmental degradation (Sadorsky, 2014, Salahuddin et al., 2016, Aller et al., 2021).

Our findings corroborate a growing body of literature highlighting the nuanced but significant roles of meteorological conditions in shaping PM2.5 dynamics (Jin et al., 2022, Páez-Osuna et al., 2022, Sirithian and Thanatrakolsri, 2022).

Several studies emphasize the role of wind in facilitating pollutant dispersion and thus lowering PM2.5 levels. For instance, Bamola et al. (2024) observe a negative association between wind speed and PM2.5/PM10 ratios, with high wind speeds (>10 m/s) contributing to cleaner air, especially in summer. Similar effects are noted by Hu et al. (2022), who find that wind speed explains up to 63% of the variation in cyclists' PM2.5 exposure, confirming its major role in reducing localized pollution levels. The dispersion potential of wind is further supported by Fang et al. (2024), who link low wind speed and low planetary boundary layer height with elevated PM2.5 levels during winter pollution episodes.

Rainfall, particularly in the form of light and prolonged precipitation, has been shown to facilitate wet deposition of particulate matter. Li et al. (2025) provide compelling evidence that longer and more intense rain events significantly enhance the removal efficiency of PM2.5 and its constituents, with the impact being most pronounced under low-wind and high-humidity conditions.

Our finding of a positive association between relative humidity and PM2.5 levels is likewise supported by several studies. Nguyen et al. (2024) find that humidity has the most dominant influence on PM2.5 concentration during pollution episodes (ρ = 0.45), while Fattah et al. (2023) also rank humidity as the most influential meteorological factor across major South Asian cities. The mechanism likely involves hygroscopic growth of aerosols and enhanced secondary particle formation under moist conditions, as discussed by Ma et al. (2021).

Taken together, our insights reinforce the importance of incorporating economic, social, demographic, and weather patterns into air quality management and modeling strategies.

7 Conclusions

In this study, we investigate the factors influencing PM2.5 levels using machine learning techniques as well as regression and data envelopment analyses. Our findings reveal that areas characterized by a higher concentration of KIBS sectors tend to experience lower PM2.5 levels, supporting our hypothesis regarding the benefits of servitization in reducing emissions.

In particular, our results indicate that a 1% rise in the share of KIBS sectors within the local economy is linked to a decrease in PM2.5 concentrations, with estimated effects ranging from –0.47 to –0.11. These findings fuel the debate on the role of servitization in helping territories meet international climate targets. According to OECD estimates, emissions in Italian regions should yearly decrease by about 1.9% to achieve the EU's goal of cutting emissions by 55% by 2030 relative to 1990 levels. Similarly, alternative policy scenarios developed by E3Mlab and IIASA (2016) to meet the 2030 climate targets project annual greenhouse gas (GHG) emission reductions for Italy between 2.4% and 4.0%, depending on energy efficiency strategies. These estimates align with broader European and global studies, which suggest that an annual carbon emission reduction of approximately 2.0% is necessary to stay on track with climate commitments (Fragkos et al., 2017). Overall, our results suggest that servitization may support the transition towards a low-carbon economy.

Moreover, several other critical drivers of PM2.5 concentrations emerge, with income per capita and meteorological factors such as wind speed and precipitation, representing significant predictors. Specifically, higher income levels correlate with increased PM2.5 concentrations, while greater wind speeds and precipitation are associated with improved air quality.

We then examine environmental efficiency in Italian municipalities, showing that the Centre and South generally achieve better efficiency scores, while the North lags behind. More environmentally efficient municipalities have greater servitization but lower income per capita and Maqi index. Knowledge-intensive sectors enhance both air quality and efficiency. Institutional quality significantly impacts efficiency by fostering sustainable development patterns. Richer areas with lower economic diversity tend to have lower efficiency, underscoring the role of socioeconomic structures.

The implications of these findings are substantial for policymakers aiming to improve air quality.

The heterogeneities observed across Italian macro-areas suggest that one-size-fits-all approaches are unlikely to be effective, and that regional differences should instead be acknowledged and valorized when designing environmental policy. In the industrialized North, where income levels are higher and the economy is more manufacturing-oriented, expanding knowledge-intensive business services could offer a promising path for mitigating air pollution. Our findings indicate that stronger servitization is associated with lower PM2.5 levels, suggesting that shifting toward cleaner, service-based activities may help reduce environmental pressures without hindering economic performance. In contrast, in the South—where pollution levels are generally lower and the economy is already more service-driven—greater policy emphasis on improving institutional quality may prove more effective. The positive association between the Maqi index and environmental efficiency, particularly in southern municipalities, indicates that strengthening administrative capacity could support more effective local environmental governance.

Importantly, both the expansion of KIBS and the enhancement of local institutional quality are not only relevant for environmental performance, but may also serve as catalysts for broader socioeconomic development. In this sense, promoting servitization and institutional capacity at the local level may advance not only environmental goals but also the broader objective of sustainable and inclusive development. These findings reinforce the need for differentiated, place-based policy strategies that integrate environmental, economic, and institutional dimensions.

Despite our efforts to adopt a methodologically rigorous approach, this study has some limitations. First, although PM2.5 is a key air quality indicator widely monitored by international organizations due to its significant health implications, our analysis focuses solely on this single pollutant. Future research could be enriched by including additional hazardous pollutants, thereby offering a more comprehensive picture of local air quality dynamics. Second, while the selection of variables in our empirical models aligns with the existing literature, the analysis may be refined by incorporating a broader set of explanatory variables or a more detailed set of inputs for the DEA, better capturing the factors contributing to PM2.5 emissions and ensuring stronger internal consistency. Third, the analysis is based on cross-sectional data. Incorporating longitudinal information would allow for the investigation of temporal trends and the assessment of variation during critical periods such as the COVID-19 pandemic. This could shed further light on the role of human activities by comparing the influence of explanatory variables under both normal and crisis conditions. Finally, the study is limited to a single country. Broadening the geographical scope could reveal spatial heterogeneity in the drivers of PM2.5 and improve the generalizability of the findings.

References

Ali, H. S., V. Zeqiraj, W. L. Lin, S. H. Law, Z. Yusop, U. A. A. Bare, and L. Chin (2019). Does quality institutions promote environmental quality? Environmental Science and Pollution Research 26, 10446–10456.

Aller, C., L. Ductor, and D. Grechyna (2021). Robust determinants of co2 emissions. Energy Economics 96, 105154.

Amaddeo, E., M. Arena, A. S. Bergantino, G. Bonaccorsi, A. Buongiorno, A. Clemente, A. Flori, M. Intini, F. Scotti, V. M. Urbano, et al. (2024). A set of multidimensional indicators to assess the resilience and attractiveness of italian provinces and municipalities (2010–2022 panel data). Data in Brief 57, 111042.

Augusto, B., D. Lopes, S. Rafael, M. C. Coelho, and J. Ferreira (2024). Assessing the impact of different urban morphology scenarios on air pollutant emissions distribution. Science of The Total Environment 950, 175341.

Bae, J. H., D. D. Li, and M. Rishi (2017). Determinants of co2 emission for post-soviet union independent countries. Climate Policy 17(5), 591–615.

Bamola, S., G. Goswami, S. Dewan, I. Goyal, M. Agarwal, A. Dhir, and A. Lakhani (2024). Characterising temporal variability of pm2. 5/pm10 ratio and its correlation with meteorological variables at a sub-urban site in the taj city. Urban Climate 53, 101763.

Banker, R. D., A. Charnes, and W. W. Cooper (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. Management science 30(9), 1078–1092.

Bel, G. and J. J. Teixidó (2020). The political economy of the paris agreement: Income inequality and climate policy. Journal of Cleaner Production 258, 121002.

Bellandi, M. and E. Santini (2019). Territorial servitization and new local productive configurations: the case of the textile industrial district of prato. Regional Studies 53(3), 356–365.

Beloconi, A. and P. Vounatsou (2023). Revised eu and who air quality thresholds: Where does europe stand? Atmospheric Environment 314, 120110.

Bernauer, T. and V. Koubi (2009). Effects of political institutions on air quality. Ecological economics 68(5), 1355–1365.

Bondy, M., S. Roth, and L. Sager (2020). Crime is in the air: The contemporaneous relationship between air pollution and crime. Journal of the Association of Environmental and Resource Economists 7(3), 555–585.

Bressanelli, G., F. Adrodegari, M. Perona, and N. Saccani (2018). Exploring how usage-focused business models enable circular economy through digital technologies. Sustainability 10(3), 639.

Bressanelli, G., N. Saccani, and M. Perona (2024). Are digital servitization-based circular economy business models sustainable? a systemic what-if simulation model. Journal of Cleaner Production 458, 142512.

Cerqua, A., C. Giannantoni, F. Zampollo, and M. Mazziotta (2025). The municipal administration quality index: the italian case. Social Indicators Research, 1–34.

Chai, F., J. Gao, Z. Chen, S. Wang, Y. Zhang, J. Zhang, H. Zhang, Y. Yun, and C. Ren (2014). Spatial and temporal variation of particulate matter and gaseous pollutants in 26 cities in china. Journal of Environmental Sciences 26(1), 75–82.

Charnes, A., W. W. Cooper, and E. Rhodes (1978). Measuring the efficiency of decision making units. European journal of operational research 2(6), 429–444.

Chen, J., C. Zhou, S. Wang, and S. Li (2018). Impacts of energy consumption structure, energy intensity, economic growth, urbanization on pm2. 5 concentrations in countries globally. Applied energy 230, 94–105.

Chiarini, B., A. D'Agostino, E. Marzano, and A. Regoli (2020). The perception of air pollution and noise in urban environments: A subjective indicator across european countries. Journal of environmental management 263, 110272.

Chiarini, B., A. D'Agostino, E. Marzano, and A. Regoli (2021). Air quality in urban areas: Comparing objective and subjective indicators in european countries. Ecological Indicators 121, 107144.

Choi, S.-Y., S.-W. Park, J.-Y. Byun, and Y.-J. Han (2021). Characteristics of locally occurring high pm2. 5 concentration episodes in a small city in south korea. Atmosphere 12(1), 86.

Cole, M. A., R. J. Elliott, and K. Shimamoto (2005). Industrial characteristics, environmental regulations and air pollution: an analysis of the uk manufacturing sector. Journal of environmental economics and management 50(1), 121–143.

Cusumano, M. A., S. J. Kahl, and F. F. Suarez (2015). Services, industry evolution, and the competitive strategies of product firms. Strategic management journal 36(4), 559–575.

Danek, T., E. Weglinska, and M. Zareba (2022). The influence of meteorological factors and terrain on air pollution concentration and migration: a geostatistical case study from krakow, poland. Scientific Reports 12(1), 11050.

Dawson, J., P. Adams, and S. Pandis (2007). Sensitivity of pm 2.5 to climate in the eastern us: a modeling case study. Atmospheric chemistry and physics 7(16), 4295–4309.

Dutt, K. (2009). Governance, institutions and the environment-income relationship: a cross-country study. Environment, Development and Sustainability 11, 705–723.

E3Mlab and IIASA (2016). Technical report on member state results of the euco policy scenarios.

European Commission, E. (2018). A clean planet for all. a european long-term strategic vision for a prosperous, modern, competitive and climate neutral economy.

Fan, S.-M., L. W. Horowitz, H. Levy, and W. J. Moxim (2004). Impact of air pollution on wet deposition of mineral dust aerosols. Geophysical research letters 31(2).

Fang, C., X. Li, J. Li, J. Tian, and J. Wang (2024). Impact of meteorological conditions on pm2. 5 pollution in changchun and associated health risks analysis. Atmosphere 15(5), 616.

Fattah, M. A., S. R. Morshed, A.-A. Kafy, Z. A. Rahaman, and M. T. Rahman (2023). Wavelet coherence analysis of pm2. 5 variability in response to meteorological changes in south asian cities. Atmospheric Pollution Research 14(5), 101737.

Favoretto, C., G. H. Mendes, M. G. Oliveira, P. A. Cauchick-Miguel, and W. Coreynen (2022). From servitization to digital servitization: How digitalization transforms companies' transition towards services. Industrial marketing management 102, 104–121.

Ferro, S., A. Palma, C. Serra, and M. Stafoggia (2024). Beyond birth: The medium-term health impact of prenatal exposure to air pollution. Journal of Environmental Economics and Management, 103009.

Flori, A. and F. Scotti (2025). When the intensity of trading meets compliance requirements: An assessment for firms operating within the eu ets. Energy Economics, 108542.

Fragkos, P., N. Tasios, L. Paroussos, P. Capros, and S. Tsani (2017). Energy system impacts and policy implications of the european intended nationally determined contribution and low-carbon pathway to 2050. Energy Policy 100, 216–226.

Frenken, K., F. Van Oort, and T. Verburg (2007). Related variety, unrelated variety and regional economic growth. Regional studies 41(5), 685–697.

Friesenbichler, K. S. and A. Kügler (2022). Servitization across countries and sectors: Evidence from world input-output data. Economic Systems 46(3), 101014.

Gomes, E., O. F. Bustinza, S. Tarba, Z. Khan, and M. Ahammad (2019). Antecedents and implications of territorial servitization. Regional Studies 53(3), 410–423.

González, A., G. Daly, P. Pinch, N. Adams, V. Valtenbergs, M. C. Burns, and H. Johannesson (2015). Indicators for spatial planning and territorial cohesion: Stakeholder-driven selection approach for improving usability at regional and local levels. Regional Studies 49(9), 1588–1602.

González, S. (2011). The north/south divide in italy and england: Discursive construction of regional inequality. European Urban and Regional Studies 18(1), 62–76.

Grossman, G. M. and A. B. Krueger (1995). Economic growth and the environment. The quarterly journal of economics 110(2), 353–377.

Guo, L.-C., Y. Zhang, H. Lin, W. Zeng, T. Liu, J. Xiao, S. Rutherford, J. You, and W. Ma (2016). The washout effects of rainfall on atmospheric particulate pollution in two chinese cities. Environmental pollution 215, 195–202.

Guo, W., Y. Yang, J. Zhang, K. Han, Y. Yang, Q. Chen, S. Li, and Y. Zhu (2024). Effects of valley topography on ozone pollution in the lanzhou valley: A numerical case study. Environmental Pollution 363, 125225.

Hao, Y., H. Chen, and Q. Zhang (2016). Will income inequality affect environmental quality? analysis based on china's provincial panel data. Ecological indicators 67, 533–542.

Hao, Z., C. Liu, and M. Goh (2021). Determining the effects of lean production and servitization of manufacturing on sustainable performance. Sustainable Production and Consumption 25, 374–389.

Horálek, J., L. Vlasáková, M. Schreiberová, J. Marková, P. Schneider, P. Kurfürst, F. Tognet, J. Schovánková, O. Vlček, and D. Damašková (2024, February). European air quality maps for 2020. PM10, PM2.5, Ozone, NO2, NOx and Benzo(a)pyrene spatial estimates and their uncertainties. ETC HE Report 2022/12.

Hu, H., Q. Chen, Q. Qian, X. Zhou, Y. Chen, and Y. Cai (2022). Field investigation for ambient wind speed and direction effects exposure of cyclists to pm2. 5 and pm10 in urban street environments. Building and Environment 223, 109483.

IPCC (2014). Summary for policymakers. in: Climate change 2014: Mitigation of climate change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Jaffe, A. B., R. G. Newell, and R. N. Stavins (2005). A tale of two market failures: Technology and environmental policy. Ecological economics 54(2-3), 164–174.

Jin, H., X. Chen, R. Zhong, and M. Liu (2022). Influence and prediction of pm2. 5 through multiple environmental variables in china. Science of The Total Environment 849, 157910.

Jin, M.-Y., L.-Y. Zhang, Z.-R. Peng, H.-D. He, P. Kumar, and J. Gallagher (2024). The impact of dynamic traffic and wind conditions on green infrastructure performance to improve local air quality. Science of the Total Environment 917, 170211.

Kim, B.-Y., Y.-K. Lim, and J. W. Cha (2022). Short-term prediction of particulate matter (pm10 and pm2. 5) in seoul, south korea using tree-based machine learning algorithms. Atmospheric Pollution Research 13(10), 101547.

Kohtamäki, M., V. Parida, P. C. Patel, and H. Gebauer (2020). The relationship between digitalization and servitization: The role of servitization in capturing the financial potential of digitalization. Technological Forecasting and Social Change 151, 119804.

Kohtamäki, M., R. Rabetino, V. Parida, D. Sjödin, and S. Henneberg (2022). Managing digital servitization toward smart solutions: Framing the connections between technologies, business models, and ecosystems.

Kolcava, D., Q. Nguyen, and T. Bernauer (2019). Does trade liberalization lead to environmental burden shifting in the global economy? Ecological Economics 163, 98–112.

Lafuente, E., Y. Vaillant, and F. Vendrell-Herrero (2017). Territorial servitization: Exploring the virtuous circle connecting knowledge-intensive services and new manufacturing businesses. International Journal of Production Economics 192, 19–28.

Lau, L.-S., C.-K. Choong, and Y.-K. Eng (2014). Carbon dioxide emission, institutional quality, and economic growth: empirical evidence in malaysia. Renewable energy 68, 276–281.

Li, F., J. Yan, Y. Wei, J. Zeng, X. Wang, X. Chen, C. Zhang, W. Li, M. Chen, and G. Lü (2021). Pm2. 5-bound heavy metals from the major cities in china: Spatiotemporal distribution, fuzzy exposure assessment and health risk management. Journal of Cleaner Production 286, 124967.

Li, X. and Y. Li (2022). The impact of perceived air pollution on labour supply: Evidence from china. Journal of Environmental Management 306, 114455.

Li, X., Y. Ma, Y. Wang, N. Liu, and Y. Hong (2017). Temporal and spatial analyses of particulate matter (pm10 and pm2. 5) and its relationship with meteorological parameters over an urban city in northeast china. Atmospheric research 198, 185–193.

Li, Y., L. Zhou, H. Liu, S. Liu, M. Feng, D. Song, Q. Tan, H. Jiang, S. Zuoqiu, and F. Yang (2025). Disparities in precipitation effects on pm2. 5 mass concentrations and chemical compositions: Insights from online monitoring data in chengdu. Journal of Environmental Sciences 156, 421–434.

Lin, Y.-C., Y.-T. Lin, C.-R. Chen, and C.-Y. Lai (2025). Meteorological and traffic effects on air pollutants using bayesian networks and deep learning. Journal of Environmental Sciences 152, 54–70.

Lombardi, S., E. Santini, and C. Vecciolini (2022). Drivers of territorial servitization: An empirical analysis of manufacturing productivity in local value chains. International Journal of Production Economics 253, 108607.

Lou, C., H. Liu, Y. Li, Y. Peng, J. Wang, and L. Dai (2017). Relationships of relative humidity with pm 2.5 and pm 10 in the yangtze river delta, china. Environmental monitoring and assessment 189, 1–16.

Łowicki, D. (2019). Landscape pattern as an indicator of urban air pollution of particulate matter in poland. Ecological indicators 97, 17–24.

Ma, S., M. Shao, Y. Zhang, Q. Dai, and M. Xie (2021). Sensitivity of pm2. 5 and o3 pollution episodes to meteorological factors over the north china plain. Science of the Total Environment 792, 148474.

Makhdoomi, A., M. Sarkhosh, and S. Ziaei (2025). Pm2. 5 concentration prediction using machine learning algorithms: an approach to virtual monitoring stations. Scientific Reports 15(1), 8076.

Menon, R. R., A. Bigdeli, A. Adem, A. Schroeder, M. Awais, T. Baines, G. Battisti, N. Driffield, S. Fouad, and M. Roeder (2024). Unpacking the triple nexus: Environmental performance, economic performance and servitization–a systematic review and theoretical reflections. Journal of Cleaner Production, 142459.

Neely, A. (2008). Exploring the financial consequences of the servitization of manufacturing. Operations management research 1, 103–118.

Nguyen, G. T. H., L. T. La, H. Hoang-Cong, and A. H. Le (2024). An exploration of meteorological effects on pm2. 5 air quality in several provinces and cities in vietnam. Journal of Environmental Sciences 145, 139–151.

Özokcu, S. and Ö. Özdemir (2017). Economic growth, energy, and environmental kuznets curve. Renewable and Sustainable Energy Reviews 72, 639–647.

Páez-Osuna, F., G. Valencia-Castañeda, and U. A. Rebolledo (2022). The link between covid-19 mortality and pm2. 5 emissions in rural and medium-size municipalities considering population density, dust events, and wind speed. Chemosphere 286, 131634.

Palafox-Alcantar, P., C. McElroy, P. Trotter, R. Khosla, A. Thomas, and R. Karutz (2024). Servitization for the energy transition: The case of enabling cooling-as-a-service (caas). Journal of Cleaner Production 482, 144190.

Paschou, T., M. Rapaccini, F. Adrodegari, and N. Saccani (2020). Digital servitization in manufacturing: A systematic literature review and research agenda. Industrial Marketing Management 89, 278–292.

Pebesma, E. and R. Bivand (2023). Spatial Data Science: With applications in R. Chapman and Hall/CRC.

Peng, J., H. Han, Y. Yi, H. Huang, and L. Xie (2022). Machine learning and deep learning modeling and simulation for predicting pm2. 5 concentrations. Chemosphere 308, 136353.

Roos, G. (2015). Servitization as innovation in manufacturing—a review of the literature. The handbook of service innovation, 403–435.

Roostaei, V., F. Gharibzadeh, M. Shamsipour, S. Faridi, and M. S. Hassanvand (2024). Vertical distribution of ambient air pollutants (pm2. 5, pm10, nox, and no2); a systematic review. Heliyon 10(21).

Rothenberg, S. (2007). Sustainability through servicizing. MIT Sloan management review.

Rungi, A. and F. Biancalani (2019). Heterogeneous firms and the north–south divide in italy. Italian Economic Journal 5(3), 325–347.

Sadorsky, P. (2014). The effect of urbanization on co2 emissions in emerging economies. Energy economics 41, 147–153.

Salahuddin, M., K. Alam, and I. Ozturk (2016). The effects of internet usage and economic growth on co2 emissions in oecd countries: A panel investigation. Renewable and Sustainable Energy Reviews 62, 1226–1235.

Sapkota, P. and U. Bastola (2017). Foreign direct investment, income, and environmental pollution in developing countries: Panel data analysis of latin america. Energy Economics 64, 206–212.

Sforzi, F. and R. Boix (2019). Territorial servitization in marshallian industrial districts: the industrial district as a place-based form of servitization. Regional Studies 53(3), 398–409.

Sgambaro, L., D. Chiaroni, and A. Urbinati (2024). The design and servitization of products according to the circular economy principles: An ecosystem perspective in the building industry. Journal of Cleaner Production 454, 142322.

Shi, K., Y. Wu, and L. Li (2021). Quantifying and evaluating the effect of urban expansion on the fine particulate matter (pm2. 5) emissions from fossil fuel combustion in china. Ecological Indicators 125, 107541.

Simar, L. and P. W. Wilson (1998). Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric frontier models. Management science 44(1), 49–61.

Simar, L. and P. W. Wilson (2002). Non-parametric tests of returns to scale. European Journal of Operational Research 139(1), 115–132.

Simar, L. and P. W. Wilson (2011). Inference by the m out of n bootstrap in nonparametric frontier models. Journal of Productivity Analysis 36, 33–53.

Sirithian, D. and P. Thanatrakolsri (2022). Relationships between meteorological and particulate matter concentrations (pm2. 5 and pm10) during the haze period in urban and rural areas, northern thailand. Air, Soil and Water Research 15, 11786221221117264.

Sjödin, D., V. Parida, and M. Kohtamäki (2023). Artificial intelligence enabling circular business model innovation in digital servitization: Conceptualizing dynamic capabilities, ai capacities, business models and effects. Technological Forecasting and Social Change 197, 122903.

Stabler, D., H. Hakala, T. Huikkola, and A.-L. Mention (2024). Aligning servitization and circularity: The role of institutional confluence in sustainable business models. Journal of Cleaner Production 462, 142666.

Tamazian, A., J. P. Chousa, and K. C. Vadlamannati (2009). Does higher economic and financial development lead to environmental degradation: evidence from bric countries. Energy policy 37(1), 246–253.

Tamazian, A. and B. B. Rao (2010). Do economic, financial and institutional developments matter for environmental degradation? evidence from transitional economies. Energy economics 32(1), 137–145.

Tarín-Carrasco, P., U. Im, C. Geels, L. Palacios-Peña, and P. Jiménez-Guerrero (2021). Contribution of fine particulate matter to present and future premature mortality over europe: A non-linear response. Environment international 153, 106517.

Tian, X., C. Zhang, H. Liu, B. Zhang, C. Lu, P. Jiao, and S. Ren (2024). Research on air quality in response to meteorological factors based on the informer model. Sustainability 16(16), 6794.

Tolonen, E. (2024). Keeping up with the decarbonization: Conceptualizing and investigating incumbents' responses to transition pressures in the post-paris world. Journal of Cleaner Production 476, 143520.

Vaillant, Y. and E. Lafuente (2024). Digital versus non-digital servitization for environmental and non-financial performance benefits. Journal of Cleaner Production 450, 142078.

Vandermerwe, S. and J. Rada (1988). Servitization of business: Adding value by adding services. European management journal 6(4), 314–324.

Vona, F., G. Marin, D. Consoli, and D. Popp (2018). Environmental regulation and green skills: an empirical exploration. Journal of the Association of Environmental and Resource Economists 5(4), 713–753.

Wang, J., W. Wang, H. Wu, and Y. Liu (2023). Exploring the effects of manufacturing servitization on enterprise energy conservation and emissions reduction moderated by digital transformation. Energy Economics 122, 106706.

Wang, S., X. Liu, X. Yang, B. Zou, and J. Wang (2018). Spatial variations of pm2. 5 in chinese cities for the joint impacts of human activities and natural conditions: A global and local regression perspective. Journal of cleaner production 203, 143–152.

Wilts, H. and H. Berg (2018). The digital circular economy: Can the digital transformation pave the way for resource-efficient materials cycles? Technical report, Wuppertal Institut für Klima, Umwelt, Energie.

Xie, J., T. Sun, C. Liu, L. Li, X. Xu, S. Miao, L. Lin, Y. Chen, and S. Fan (2022). Quantitative evaluation of impacts of the steadiness and duration of urban surface wind patterns on air quality. Science of The Total Environment 850, 157957.

Yang, Y., Z. Wang, C. Cao, M. Xu, X. Yang, K. Wang, H. Guo, X. Gao, J. Li, and Z. Shi (2024). Estimation of pm2. 5 concentration across china based on multi-source remote sensing data and machine learning methods. Remote Sensing 16(3), 467.

Yang, Z., Y. Zhang, and T. Zhang (2024). Leveraging digitalization and servitization to improve financial performance. Production Planning & Control 35(15), 2007–2020.

Yao, S., R. Xie, F. Han, and Q. Zhang (2023). Labor market distortion and air pollution: An empirical analysis based on spatial effect modeling. Journal of Environmental Management 337, 117743.

You, W.-H., H.-M. Zhu, K. Yu, and C. Peng (2015). Democracy, financial openness, and global carbon dioxide emissions: heterogeneity across existing emission levels. World Development 66, 189–207.

Yusliza, M. Y., J. Y. Yong, M. I. Tanveer, T. Ramayah, J. N. Faezah, and Z. Muhammad (2020). A structural model of the impact of green intellectual capital on sustainable performance. Journal of Cleaner Production 249, 119334.

Zender-Świercz, E., B. Galiszewska, M. Telejko, and M. Starzomska (2024). The effect of temperature and humidity of air on the concentration of particulate matter-pm2. 5 and pm10. Atmospheric Research 312, 107733.

Zhang, C. and Z. Tan (2016). The relationships between population factors and china's carbon emissions: does population aging matter? Renewable and Sustainable Energy Reviews 65, 1018–1025.

Zhang, L., Y. Cheng, Y. Zhang, Y. He, Z. Gu, C. Yu, et al. (2017). Impact of air humidity fluctuation on the rise of pm mass concentration based on the high-resolution monitoring data. Aerosol and Air Quality Research 17(2), 543–552.

Zhang, N., R. Ren, Q. Zhang, and T. Zhang (2020). Air pollution and tourism development: An interplay. Annals of Tourism Research 85, 103032.

Zhang, Q., Y. Zheng, D. Tong, M. Shao, S. Wang, Y. Zhang, X. Xu, J. Wang, H. He, W. Liu, Y. Ding, Y. Lei, J. Li, Z. Wang, X. Zhang, Y. Wang, J. Cheng, Y. Liu, Q. Shi, L. Yan, G. Geng, C. Hong, M. Li, F. Liu, B. Zheng, J. Cao, A. Ding, J. Gao, Q. Fu, J. Huo, B. Liu, Z. Liu, F. Yang, K. He, and J. Hao (2019, December). Drivers of improved $PM_{2.5}$ air quality in China from 2013 to 2017. Proceedings of the National Academy of Sciences 116(49), 24463–24469.

Zhao, J., Q. Jiang, X. Dong, K. Dong, and H. Jiang (2022). How does industrial structure adjustment reduce co2 emissions? spatial and mediation effects analysis for china. Energy Economics 105, 105704.

Zhao, Y.-F., J. Gao, Y.-J. Cai, J.-J. Wang, and J. Pan (2021). Real-time tracing vocs, o3 and pm2. 5 emission sources with vehicle-mounted proton transfer reaction mass spectrometry combined differential absorption lidar. Atmospheric Pollution Research 12(3), 146–153.

Zheng, P., T.-J. Lin, C.-H. Chen, and X. Xu (2018). A systematic design approach for service innovation of smart product-service systems. Journal of cleaner production 201, 657–667.

Zivin, J. G. and M. Neidell (2012). The impact of pollution on worker productivity. American Economic Review 102(7), 3652–3673.

A The Italian Context

Tables A1 and A2 show the worst and best five provinces in terms of PM2.5 concentration, respectively. Table A3 exhibits the average PM2.5 levels across macro areas.

Figure A1 exhibits the geographical distribution of the average PM2.5 concentration at the municipal level.

Tables A4, A5, and A6 display the descriptive statistics in municipalities located in the North, Centre, and South of Italy.

Figure A2 highlights the relationship between our data on the average PM2.5 concentration aggregated at the province level and two variables: (i) the maximum PM2.5 concentration in a year and (ii) the number of days in which the PM2.5 concentration exceeds health limits.

Table A1: We show the 5 provinces with the highest PM2.5 concentration in Italy.

Province	PM2.5
Cremona	23.40966
Milano	23.02079
Lodi	22.65597
Padova	22.39508
Monza e della Brianza	22.38707

Table A2: We show the 5 provinces with the lowest PM2.5 concentration in Italy.

Province	PM2.5
Sondrio	6.523928
Sassari	6.629870
Nuoro	6.911106
Belluno	6.934527
L'Aquila	7.021705

Table A3: We show the average PM2.5 concentration in the North, Centre, and South of Italy.

Macro-area	PM2.5
Centre	10.113187
North	15.945477
South	9.976538

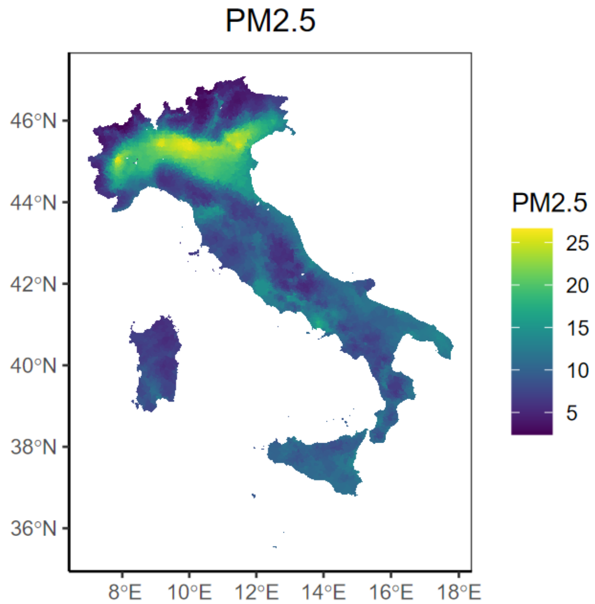

Figure A1: We show the geographical distribution of the average PM2.5 [$\mu g/m^3$] concentration at the municipal level.

Table A4: Descriptive statistics for municipalities located in the North of Italy.

	Q1	Mean	Q3	Sd
PM2.5	10.946	15.945	20.794	5.904
Income pc	17,767.920	19,424.050	20,952.530	2,737.798
RV	-8.405	-6.386	-4.279	2.643
Population	993	6,762.864	6,055	31,871.590
Inactive population	53.993	59.580	63.203	8.727
Wind speed	1.866	2.136	2.384	0.369
Humidity	71.267	74.199	77.596	4.291
Rainfall	9.852	11.351	12.235	2.259
Maqi	101.357	103.270	105.349	2.995
Servitization	0.103	0.167	0.208	0.097

Table A5: Descriptive statistics for municipalities located in the Centre of Italy.

	Q1	Mean	Q3	Sd
PM2.5	8.387	10.113	11.766	2.330
Income pc	15,644.550	17,053.710	18,142.150	2,263.546
RV	-9.019	-7.092	-5.136	2.572
Population	1,449	12,968.350	9,421	97,599.160
Inactive population	54.861	59.654	63.644	7.474
Wind speed	1.865	2.189	2.401	0.470
Humidity	74.265	76.452	78.790	3.167
Rainfall	9.367	10.399	10.863	1.314
Maqi	101.309	103.177	105.379	3.084
Servitization	0.125	0.184	0.228	0.084

Table A6: Descriptive statistics for municipalities located in the South of Italy.

	Q1	Mean	Q3	Sd
PM2.5	8.373	9.977	11.432	2.437
Income pc	12, 173.990	13, 647.850	14, 807.880	2, 080.689
RV	-8.343	-6.572	-4.705	2.409
Population	1, 225	8, 083.246	6, 787	28, 894.380
Inactive population	51.031	56.838	60.667	9.347
Wind speed	2.172	2.548	2.834	0.562
Humidity	73.773	76.609	79.881	4.404
Rainfall	8.342	9.679	10.412	1.513
Maqi	100.938	103.118	105.377	3.086
Servitization	0.137	0.198	0.242	0.091

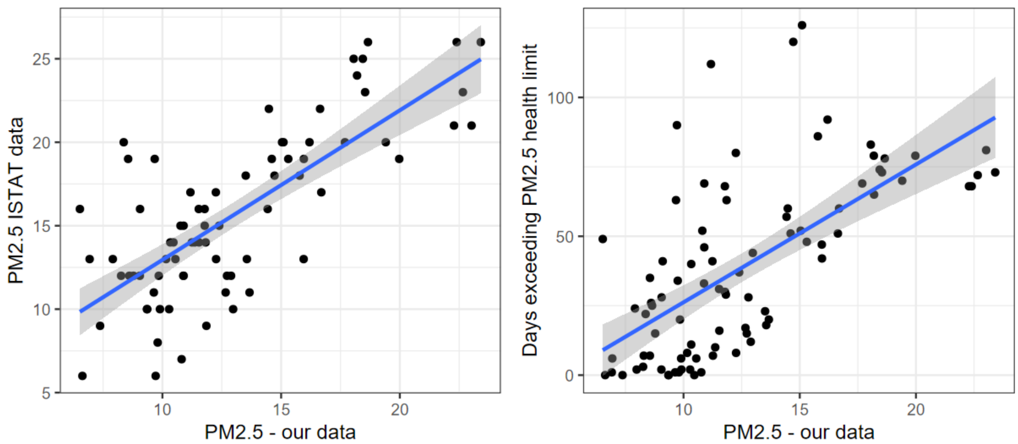

Figure A2: The left panel shows the interplay between our data on the average PM2.5 concentration aggregated at the province level and the maximum PM2.5 concentration in a year. The right panel shows the interplay between our data on the average PM2.5 concentration aggregated at the province level and the number of days in which the PM2.5 concentration exceeds health limits.

B Additional results on machine learning methods

Figures B1, B2, and B3 show the feature importance for Random Forest and XGBoost models applied on municipalities located in the North, Centre, and South and Islands of Italy.

Figures B4, B5, and B6 display the feature importance for Random Forest and XGBoost models in three robustness checks. The first includes in the analysis also the period 2020-2022 thus allowing us to assess the impact of the COVID-19 pandemic on the drivers of PM2.5. The second repeats our main analysis including Middle age population as an additional control variable. The third encompasses the variation of employment in KIBS sectors (*Servitization variation*) as our main interest variable.

Figures B7, B8, and B9 highlight the Shapley values for Random Forest and XGBoost models applied on municipalities located in the North, Centre, and South and Islands of Italy.

Figures B10, B11, and B12 exhibit the Shapley values for Random Forest and XGBoost models on our robustness checks. Specifically, they illustrate the feature contributions during the COVID-19 pandemic period, and after the inclusion of the middle-aged population and servitization variation as additional explanatory variables.

Finally, Figure B13 shows alternative performance metrics for Random Forest and Gradient Boosting Models for alternative model specifications.

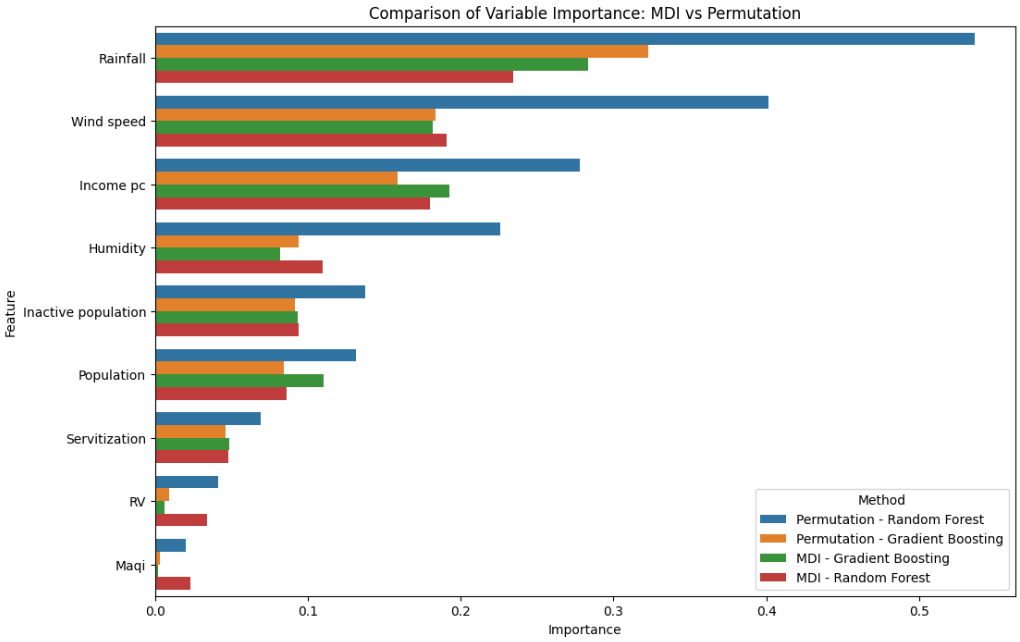

Figure B1: Comparison of Variable Importance: MDI vs Permutation. Random Forest and XGBoost models are applied on all municipalities in the North of Italy.

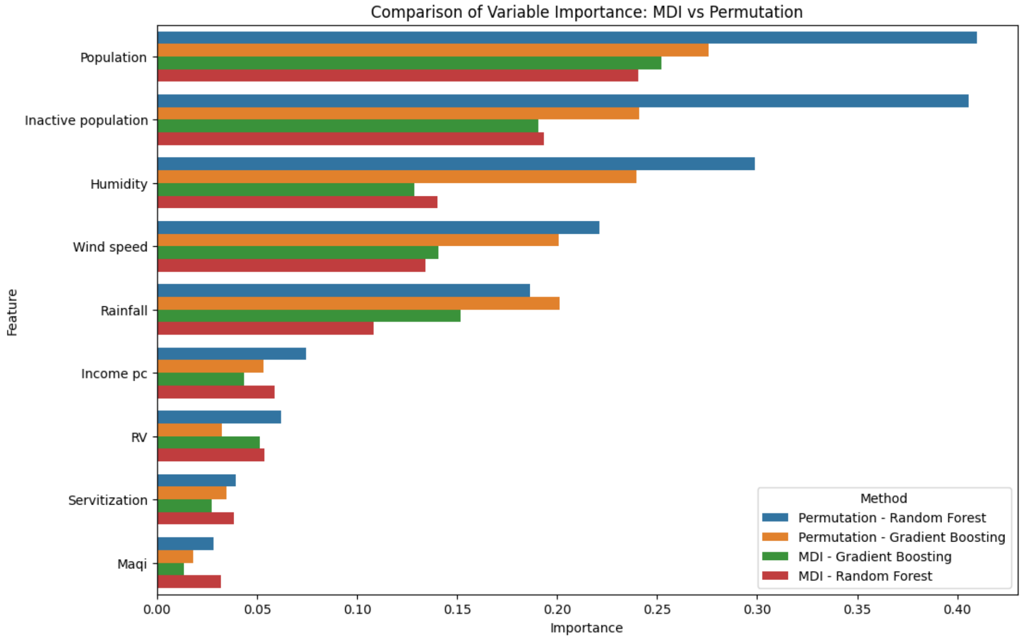

Figure B2: Comparison of Variable Importance: MDI vs Permutation. Random Forest and XGBoost models are applied on all municipalities in the Centre of Italy.

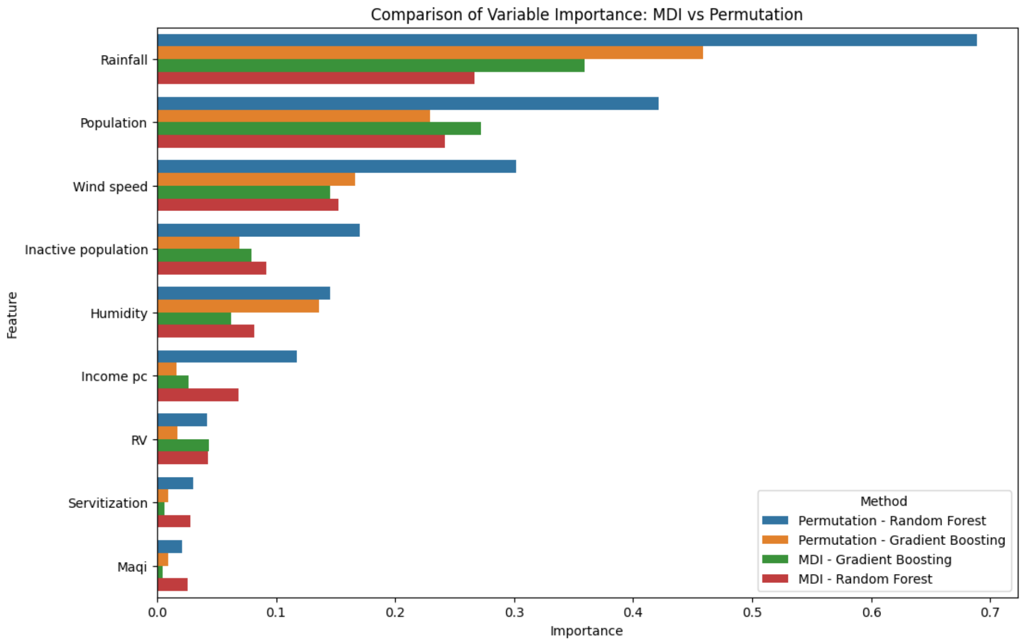

Figure B3: Comparison of Variable Importance: MDI vs Permutation. Random Forest and XGBoost models are applied on all municipalities in the South and Islands of Italy.

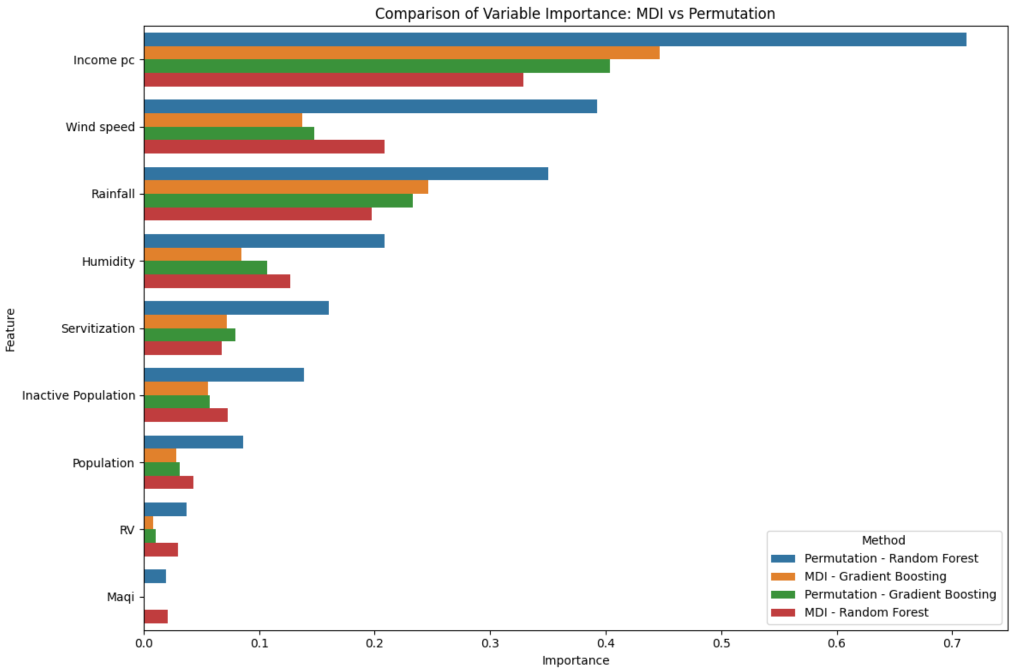

Figure B4: Comparison of Variable Importance: MDI vs Permutation. Random Forest and XGBoost models are applied on all Italian municipalities. Data are averaged over the time frame 2014-2022, thus also including the COVID-19 pandemic.

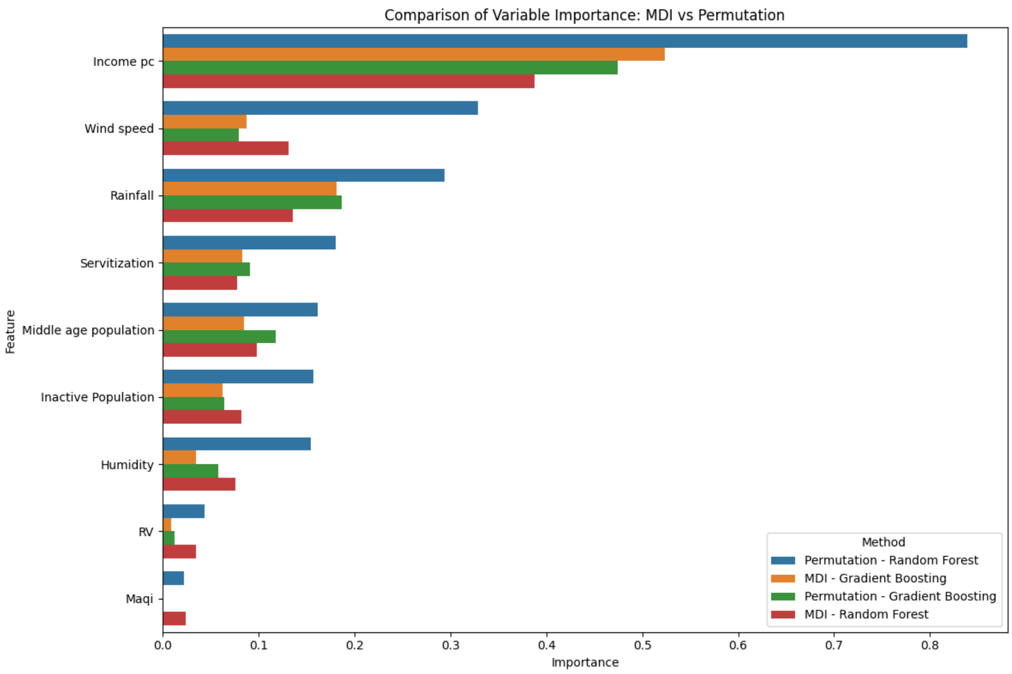

Figure B5: Comparison of Variable Importance: MDI vs Permutation. Random Forest and XGBoost models are applied on all municipalities. We also include the middle-aged population as an explanatory variable.

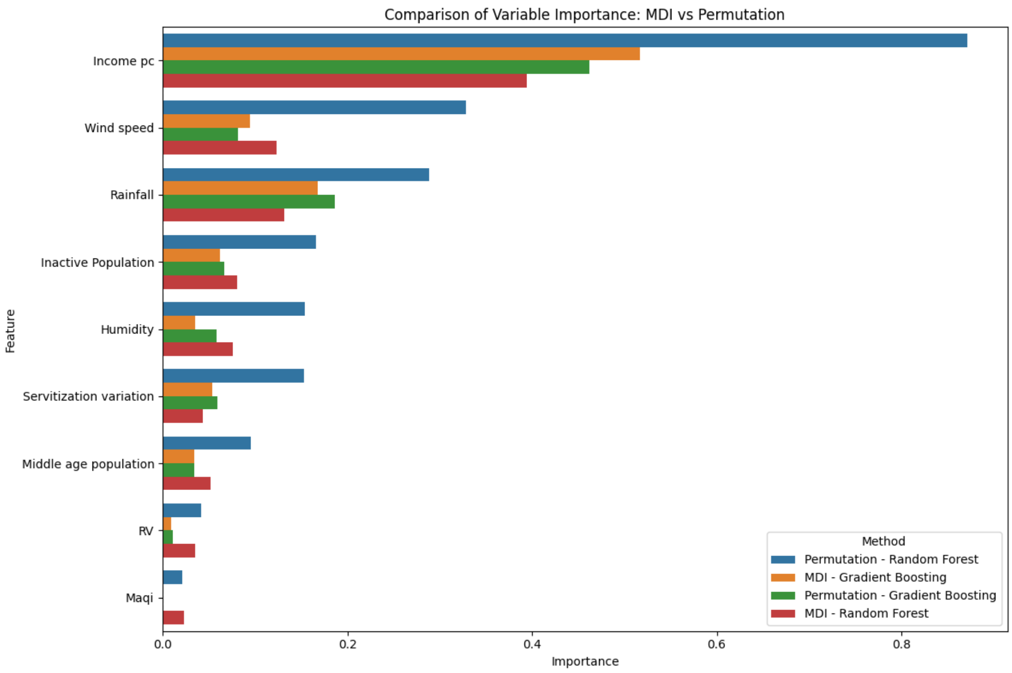

Figure B6: Comparison of Variable Importance: MDI vs Permutation. Random Forest and XGBoost models are applied on all municipalities. We include the percentage variation in KIBS sectors over the period 2014-2020 rather than the portion of employees in KIBS sectors as an explanatory variable.

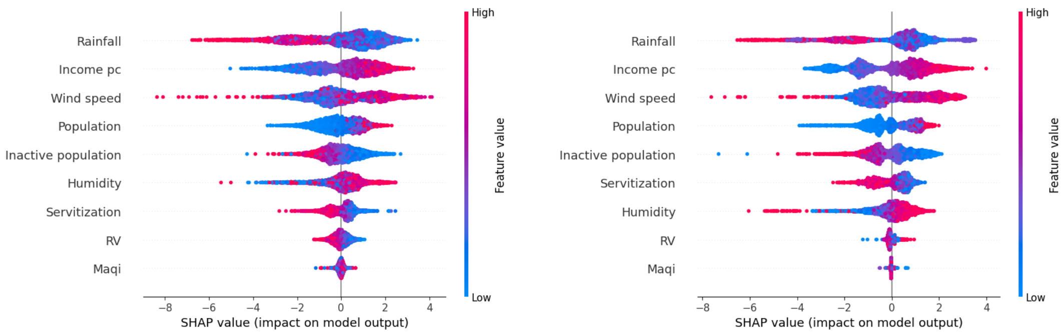

Figure B7: SHAP Summary Plot for Random Forest Model and Gradient Boosting Model applied on municipalities in the North of Italy.

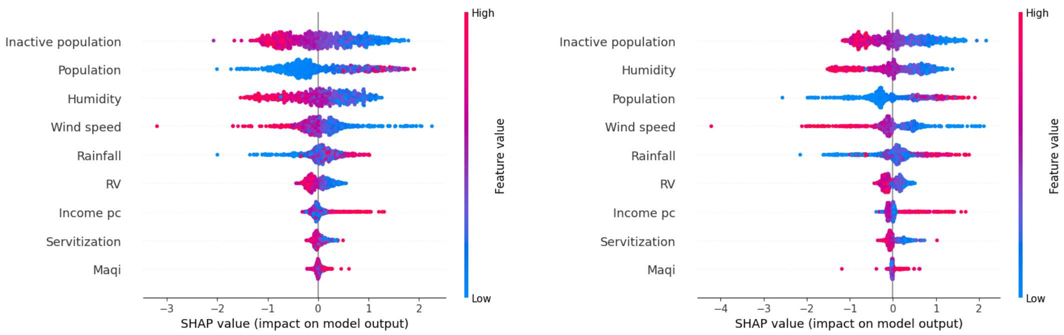

Figure B8: SHAP Summary Plot for Random Forest Model and Gradient Boosting Model applied on municipalities in the Centre of Italy.

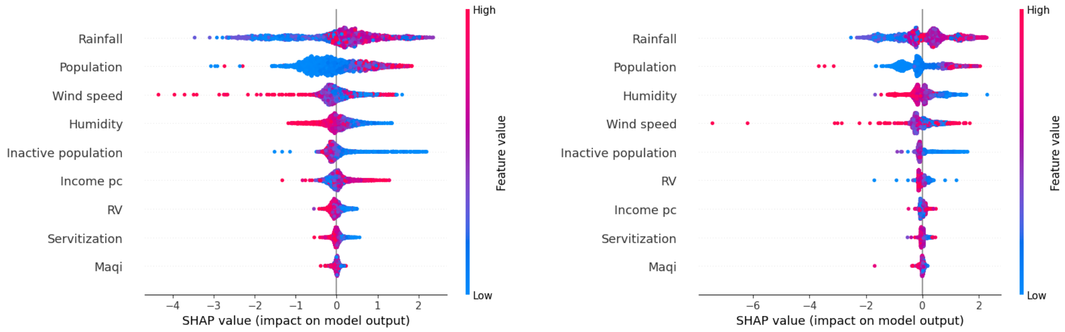

Figure B9: SHAP Summary Plot for Random Forest Model and Gradient Boosting Model applied on municipalities in the South and Islands of Italy.

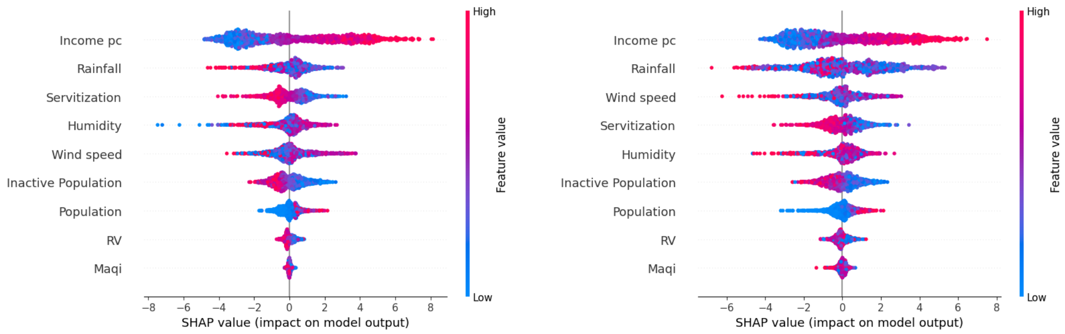

Figure B10: SHAP Summary Plot for Random Forest Model and Gradient Boosting Model applied on municipalities. Data are averaged over the time frame 2014-2022, thus also including the COVID-19 pandemic.

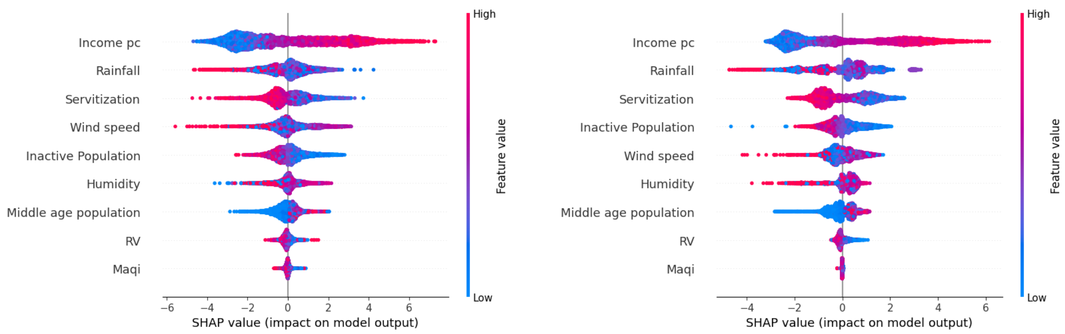

Figure B11: SHAP Summary Plot for Random Forest Model and Gradient Boosting Model applied on all municipalities. We also include the middle-aged population as an explanatory variable.

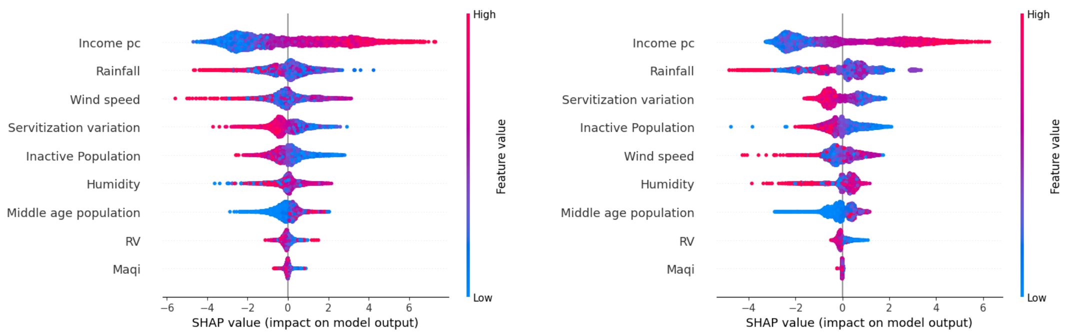

Figure B12: SHAP Summary Plot for Random Forest Model and Gradient Boosting Model applied on all municipalities. We include the percentage variation in KIBS sectors over the period 2014-2020 rather than the portion of employees in KIBS sectors as an explanatory variable.

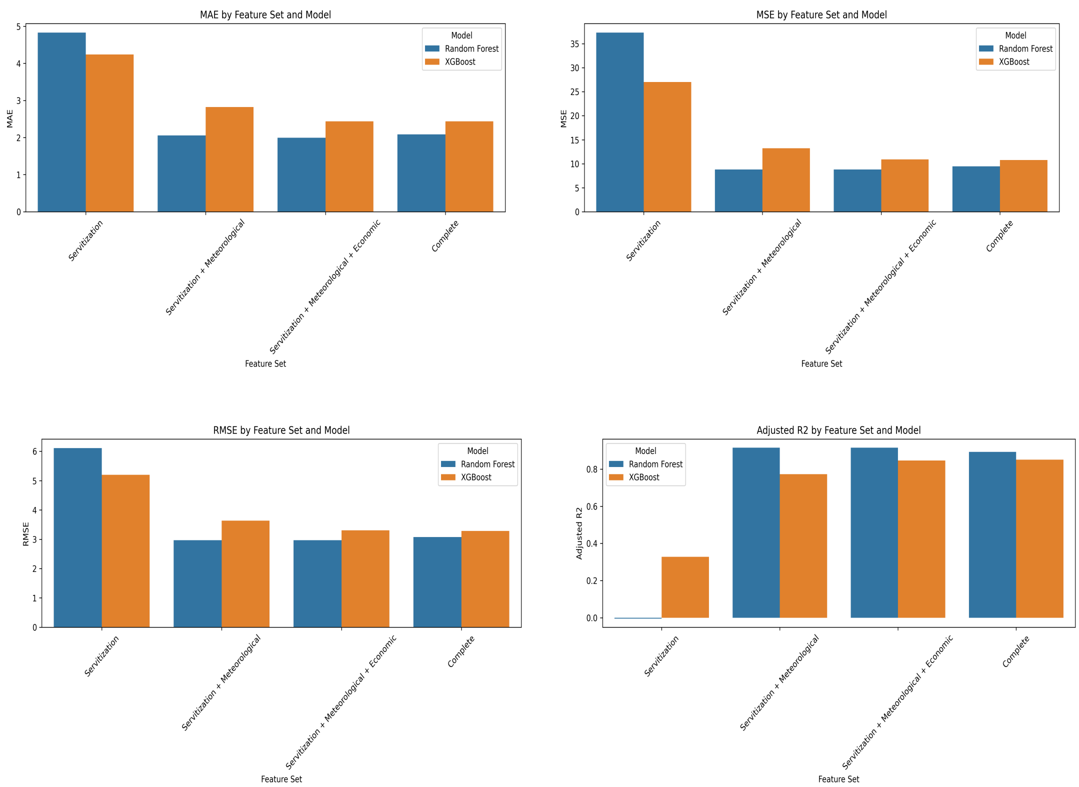

Figure B13: Performance Metrics for Random Forest and Gradient Boosting Models for alternative model specifications.

C Additional results on OLS, Ridge, Lasso

Tables C1, C2, and C3 show the results of our OLS, Ridge, and Lasso regressions for municipalities located in the North, Centre, and South and Islands of Italy.

Furthermore, Tables C4, C5, and C6 show the results of our OLS, Ridge, and Lasso regressions on our robustness checks. Specifically, these models analyse the impact of meteorological, economic, social, and demographic factors on PM2.5 concentration during the COVID-19 pandemic period, and after the inclusion of the middle-aged population and servitization variation as additional explanatory variables.

Finally, Figure C1 illustrates the percentage influence of the coefficients estimated in Table 2 through OLS, Ridge, and Lasso regressions. We compute these weights as the ratio between the specific coefficients and the sum of all variables included in the models.

Table C1: Results for the OLS, Ridge, and Lasso regressions. Models are applied to municipalities in the North of Italy.

Variable	OLS	Ridge	Lasso
Income pc	0.0006***	2.6177***	2.0000***
	(0.0000)	(0.0730)	(0.0583)
RV	-0.3953***	-0.7912***	-0.0305
	(0.0362)	(0.0783)	(0.0403)
Population	0.0000	-0.1151	
	(0.0000)	(0.1385)	
Inactive population	-0.1349***	-0.7241***	-0.0796
	(0.0094)	(0.0604)	(0.0540)
Wind speed	1.4256***	0.0205	
	(0.2194)	(0.0706)	
Humidity	0.1848***	0.1847***	
	(0.0180)	(0.0567)	
Rainfall	-0.6875***	-0.6871***	
	(0.0346)	(0.0561)	
Maqi	-0.0523*	-0.2130***	
	(0.0283)	(0.0614)	
Servitization	-0.3719***	-0.1123***	-0.3754***
	(0.1048)	(0.0240)	(0.0660)
Observations	3,753	3,753	3,753

Notes: *p $<$ 0.1, **p $<$ 0.05, ***p $<$ 0.01. Variables for which no coefficient is reported were not selected by the Lasso regression, indicating that their estimated coefficients are equal to zero.

Table C2: Results for the OLS, Ridge, and Lasso regressions. Models are applied to municipalities in the Centre of Italy.

Variable	OLS	Ridge	Lasso
Income pc	0.0001*	2.6128***	1.9944***
	(0.0000)	(0.0715)	(0.0565)
RV	-0.2857***	-0.7874***	-0.0312
	(0.0305)	(0.0774)	(0.0423)
Population	0.0000	-0.1062	
	(0.0000)	(0.1099)	
Inactive population	-0.0790***	-0.7238***	-0.0843
	(0.0086)	(0.0633)	(0.0541)
Wind speed	-1.7905***	0.0169	
	(0.1367)	(0.0701)	
Humidity	0.2161***	0.1840***	
	(0.0208)	(0.0563)	
Rainfall	-0.2520***	-0.6865***	
	(0.0458)	(0.0557)	
Maqi	0.0162	-0.2073***	
	(0.0224)	(0.0568)	
Servitization	-0.0870***	-0.2127***	-0.3771***
	(0.0210)	(0.0669)	(0.0610)
Observations	857	857	857

Notes: *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$. Variables for which no coefficient is reported were not selected by the Lasso regression, indicating that their estimated coefficients are equal to zero.

Table C3: Results for the OLS, Ridge, and Lasso regressions. Models are applied to municipalities in the South and Islands of Italy.

Variable	OLS	Ridge	Lasso
Income pc	0.0000	2.6189***	1.9957***
	(0.0000)	(0.0722)	(0.0575)
RV	-0.4285***	-0.7896***	-0.0326
	(0.0240)	(0.0762)	(0.0438)
Population	0.0000	-0.1066	
	(0.0000)	(0.1177)	
Inactive population	-0.0226***	-0.7214***	-0.0809
	(0.0052)	(0.0605)	(0.0551)
Wind speed	-0.3788***	0.0263	
	(0.0828)	(0.0722)	
Humidity	0.0731***	0.1888***	
	(0.0099)	(0.0558)	
Rainfall	-0.2972***	-0.6846***	
	(0.0296)	(0.0578)	
Maqi	-0.0364**	-0.2109***	
	(0.0145)	(0.0578)	
Servitization	-0.1407***	-0.1129***	-0.3770***
	(0.0489)	(0.0363)	(0.0641)
Observations	2,425	2,425	2,425

Notes: *p $<$ 0.1, **p $<$ 0.05, ***p $<$ 0.01. Variables for which no coefficient is reported were not selected by the Lasso regression, indicating that their estimated coefficients are equal to zero.

Table C4: Results for the OLS, Ridge, and Lasso regressions. Models are applied to all municipalities in Italy. Data are averaged over the time frame 2014-2022, thus also including the COVID-19 pandemic.

Variable	OLS	Ridge	Lasso
Income pc	0.0007***	2.4848***	1.8958***
	(0.0001)	(0.0729)	(0.0568)
RV	-0.2905***	-0.7497***	-0.0303
	(0.0251)	(0.0789)	(0.0434)
Population	-0.0000	-0.1024	
	(0.0002)	(0.1186)	
Inactive population	-0.0752***	-0.6880***	-0.0791
	(0.0061)	(0.0614)	(0.0542)
Wind speed	0.0658	0.0239	
	(0.1164)	(0.0707)	
Humidity	0.0596***	0.2140***	
	(0.0121)	(0.0564)	
Rainfall	-0.3785***	-0.7539***	
	(0.0259)	(0.0551)	
Maqi	-0.0678***	-0.1998***	
	(0.0185)	(0.0583)	
Servitization	-0.4463**	-0.1070	-0.3620***
	(0.2238)	(0.0650)	(0.0642)
Observations	7,035	7,035	7,035

Notes: *p $<$ 0.1, **p $<$ 0.05, ***p $<$ 0.01. Variables for which no coefficient is reported were not selected by the Lasso regression, indicating that their estimated coefficients are equal to zero.

Table C5: Results for the OLS, Ridge, and Lasso regressions. Models are applied to all municipalities in Italy. We also include the middle- aged population as an explanatory variable.

Variable	**OLS**	**Ridge**	**Lasso**
Income pc	0.0007***	2.6156***	1.9956***
	(0.0000)	(0.0729)	(0.0568)
RV	-0.3058***	-0.7892***	-0.0319
	(0.0251)	(0.0789)	(0.0434)
Middle age population	-0.0021**	-0.1450	0.0001
	(0.0009)	(0.1190)	(0.0002)
Inactive population	-0.0792***	-0.7242***	-0.0833
	(0.0060)	(0.0614)	(0.0542)
Wind speed	0.0548	0.0199	
	(0.1164)	(0.0707)	
Humidity	0.0518***	0.1861***	
	(0.0120)	(0.0564)	
Rainfall	-0.3441***	-0.6854***	
	(0.0259)	(0.0551)	
Maqi	-0.0714***	-0.2103***	
	(0.0185)	(0.0583)	
Servitization	-0.4698***	-0.1126***	-0.3811***
	(0.1238)	(0.0250)	(0.0642)
Observations	7,035	7,035	7,035

Notes: *p $<$ 0.1, **p $<$ 0.05, ***p $<$ 0.01. Variables for which no coefficient is reported were not selected by the Lasso regression, indicating that their estimated coefficients are equal to zero.

Table C6: Results for the OLS, Ridge, and Lasso regressions. Models are applied to all municipalities in Italy. We also include the servitization variation as an explanatory variable.

Variable	OLS	Ridge	Lasso
Income pc	0.0006***	2.6171***	1.9836***
	(0.0000)	(0.0729)	(0.0568)
RV	-0.2984***	-0.8011***	-0.0297
	(0.0254)	(0.0794)	(0.0432)
Middle age population	-0.0001*	-0.1139	
	(0.0000)	(0.1198)	
Inactive population	-0.0754***	-0.7116***	-0.0871
	(0.0063)	(0.0622)	(0.0539)
Wind speed	0.0603	0.0241	
	(0.1151)	(0.0712)	
Humidity	0.0540***	0.1795***	
	(0.0118)	(0.0557)	
Rainfall	-0.3369***	-0.6932***	
	(0.0255)	(0.0560)	
Maqi	-0.0679***	-0.2187***	
	(0.0182)	(0.0581)	
Servitization variation	-0.3191**	-0.0946**	-0.2873**
	(0.1420)	(0.0458)	(0.1235)
Observations	7,035	7,035	7,035

Notes: *p $<$ 0.1, **p $<$ 0.05, ***p $<$ 0.01. Variables for which no coefficient is reported were not selected by the Lasso regression, indicating that their estimated coefficients are equal to zero.

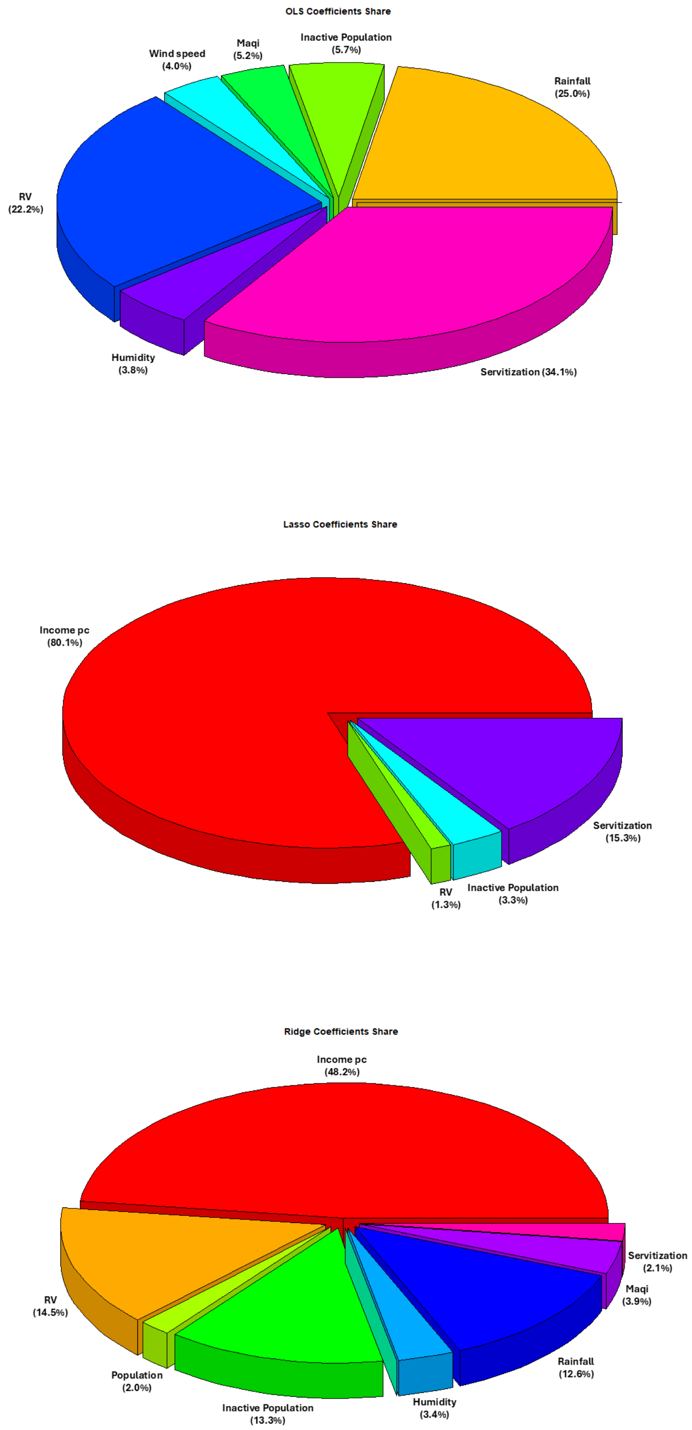

Figure C1: We show the percentage influence of the coefficients estimated in Table 2 through OLS, Ridge, and Lasso regressions. We compute these weights as the ratio between the specific coefficients and the sum of all variables included in the models.

D Additional results on Data Envelopment Analysis

Figures D1, D2, and D3 plot the distribution of the economic variables of most efficient (in blue) and the less efficient (in red) units.

Table D1 displays the p-values of t-tests comparing whether most and less efficient municipalities display statistically significant differences in socioeconomic features.

Finally, Figures D4, and D5 plot the distribution of the economic variables of most efficient (in blue) and the less efficient (in red) units when the set of variables used as input in DEA includes energy and transport variables or exclude industry density variables.

Table D1 displays the p-value of t-tests comparing whether most and less efficient municipalities display statistically significant different socioeconomic features with Reference to Figures D4 and D5.

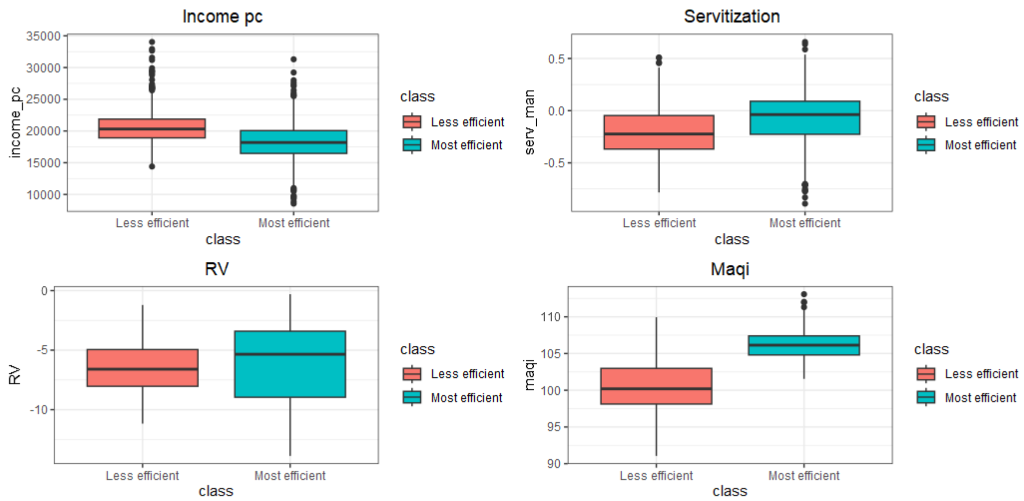

Figure D1: Distribution of socioeconomic variables of most and least efficient municipalities in the North of Italy.

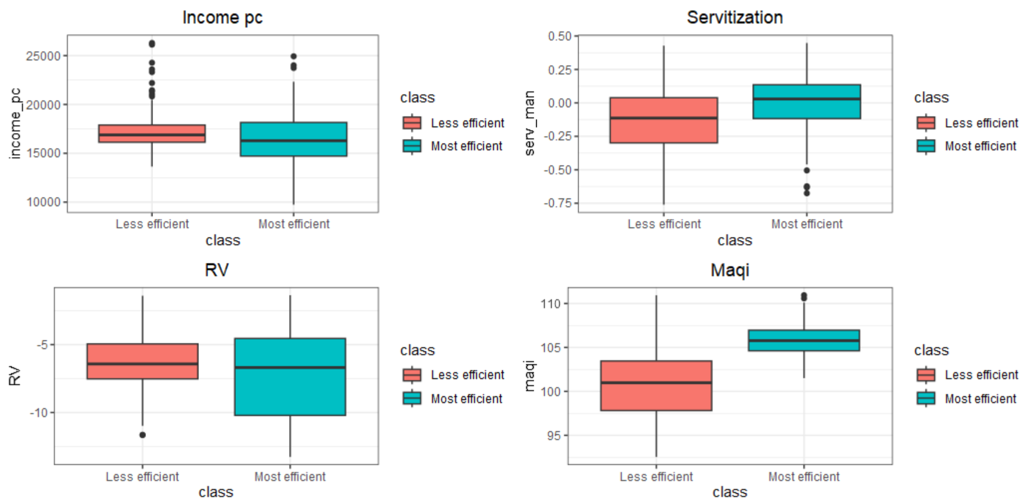

Figure D2: Distribution of socioeconomic variables of most and least efficient municipalities in the Centre of Italy.

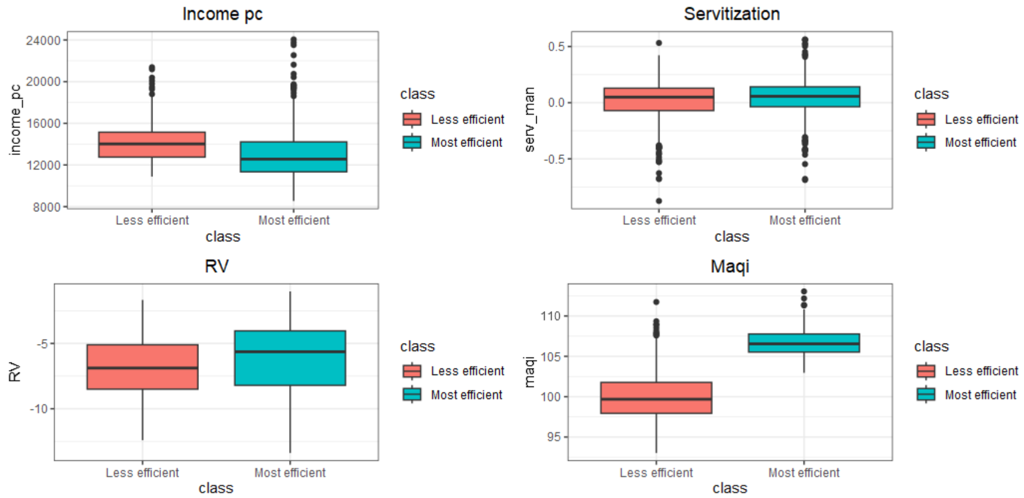

Figure D3: Distribution of socioeconomic variables of most and least efficient municipalities in the South of Italy.

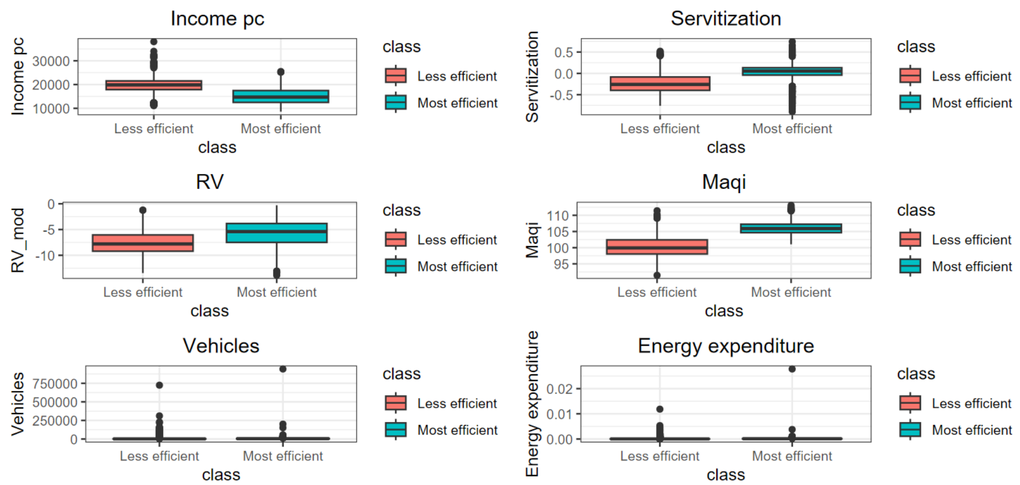

Figure D4: Distribution of socioeconomic variables of most and least efficient municipalities in Italy. The set of variables used as input in DEA includes energy and transport variables.

Table D1: We show the p-value of t-tests on socioeconomic features of most and less efficient municipalities as described in section 5.3. *Notes:* *p < 0.1, **p < 0.05, ***p < 0.01.

	Italy	North	Centre	South
Income pc	0.000***	0.012**	0.134	0.039**
Servitization	0.000***	0.009***	0.048**	0.126
RV	0.008***	0.154	0.101	0.212
Maqi	0.000***	0.000***	0.000***	0.000***

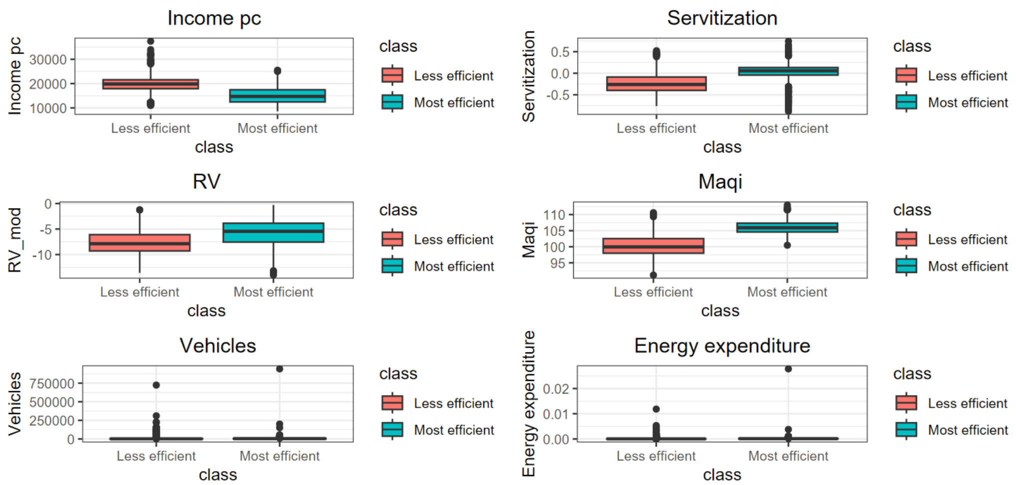

Figure D5: Distribution of socioeconomic variables of most and least efficient municipalities in Italy. The set of variables used as input in DEA excludes industry density variables.

Table D2: P-values of t-tests comparing most and less efficient municipalities. *Notes:* $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

	Energy-Transport input	No industry density
Income pc	0.000***	0.000***
Servitization	0.000***	0.000***
RV	0.011**	0.014**
Maqi	0.000***	0.000***
Vehicles	0.990	0.973
Energy Expenditures	0.258	0.251