Università degli studi di Milano-Bicocca

Data Science Lab for Smart Cities

Final Essay

Urban Environment Analysis and PM10 Prediction Models in Milan

Authors:
Edoardo Fava - 851665 - e.fava4@campus.unimib.it
Samir Doghmi - 897358 - s.doghmi@campus.unimib.it

September 7, 2024

Contents

Abstract

1 Problem Description

1.1 Urbanisation and Smart cities

Urbanisation is one of the most prominent trends of the past century and key driver of development. Across high-income countries, more than 80% of the population lived in urban areas [1]. According to OECD [2], the spatial concentration of the population in metropolitan regions is expected to continue over the next two decades, reinforcing the past trend.

In this context, it has become crucial for policymakers to be able to control and better understand the cities they govern. The increasingly complex socio-economic processes and fast changes posed demands of flexible and suitable answers, ensuring effective governance. The concept of "Smart city" tries to respond to this need. A "Smart City" leverages various information technologies - including the Internet of Things (IoT), cloud computing, big data, and artificial intelligence (AI) - to facilitate the planning, construction, management and smart services of cities. The emergence of a network of sensors, cameras, cable and data centers allows city authorities to deliver essential services more quickly and efficiently [5]. The major goal of Smart Cities is improving the quality of life, reducing social inequality, protecting the environment, and promoting a more sustainable economic growth. These objectives are condensed in the eleventh United Nations Sustainable Development Goal: "Make cities and human settlements inclusive, safe, resilient and sustainable".

In this scenario, sustainability is a critical concept of societal health: Urban sustainability focuses on the persistence of a desirable outcome of urban environments over time [8]. Achieving this requires the efficient use of naturally available environmental resources and smart urban planning. As a result, sustainable cities are cities that manage to maximize energetic efficiency, reduce waste and pollution, support renewable energy, promote sustainable mobility, and preserve urban ecosystems.

Given the shift toward a more people-centric view, the need for smart governance has also become important. Greater citizen participation, access to public information as the growing acknowledgement of a city as a complex system of system [4] has led to the recognition of the central role of the social, economic and both institutional and not-institutional forces. With the advancement of ICTs, cities have more tools to improve certain aspects of the urban environments. On this logic, many institutions improved government services by reducing information asymmetries and promoting open data initiatives, guided by the principle that on the guideline that citizen empowerment leads to more virtuous behaviour. Information and awareness-raising measures become relevant to encourage more appropriate and positive lifestyles, facilitating urban interventions at the neighbourhood scale. These efforts also open up opportunities for collaboration with private agents, researchers and non-profit organizations, aiming to enhance and make more efficient traditional networks and services through digital solutions. Although these technological advantages bring new challenges, the potential for improving policy initiatives and increasing transparency offers significant benefits.

1.2 Urban air pollution and Particulate matter

City expansion and increasing complexity bring several risks to urban societies, especially in terms of environmental impacts. These constant pressures such as burning fossil fuels, residential heating and in general human footprint without proper accountability on possible effects, have triggered negative impacts on the environment. Air pollution stands out as one of the most pressing issues, with significant health and environmental consequences. The Institute For Health Metrics and Evaluation (IHME) ranked it as one of the leading threats to global health, second only to high blood pressure [25]. The growing threat of climate change and environmental degradation has pushed the adoption of measures and international compromises, such as the Paris Agreements (2015) to strengthen the global response to climate change and reduce anthropogenic emissions. Climate change influences air pollution by altering the frequency, severity, and duration of heat waves, air stagnation events, precipitation, and other meteorological conditions favourable to pollutant accumulation [9].

According to the World Health Organization (WHO), 4.2 million deaths are attributed to ambient air pollution each year [3]. it also estimated that in 2010, the annual economic cost of premature deaths from air pollution across the countries of the WHO European Region stood at US$ 1.431 trillion [10]. Particulate matter (PM) is one of the most harmful pollutants, causing many serious health effects. Short-term and long-term exposure contributes to disease through an increase in mortality rate, years of life lost and years lived with disability [11]. Both $PM_{2.5}$ - particles with an aerodynamic diameter equal to or less than $2.5\mu m$ - and PM_{10} - particles with a

diameter of equal to or less than $10\mu m$ are primarily originated by anthropogenic activities in urban areas [26]. In response to the growing number of local monitoring sites and mounting evidence of the harmful effects, the WHO updated its health-based guidelines for outdoor air quality in 2021. The new guidelines significantly lowered the recommended limits for average daily concentrations of particulate matter, setting the target for $PM_{2.5}$ at 15 $\mu g/m^3$ and PM_{10} at 45 $\mu g/m^3$. More stringent annual limits were set with PM_{10} at 15 $\mu g/m^3$ and $PM_{2.5}$ at 5 g/m^3. The spatial and temporal concentration of these pollutants in outdoor air varies according to the spatial distribution of the sources and their pattern of operation (e.g. daily or seasonal), the characteristics of the pollutants and their dynamics (dispersion, deposition, interaction with other pollutants), and meteorological conditions [7].
The deterioration of air quality questions the urban sustainability of smart cities, requiring a data-driven approach to effectively mitigate emissions at their source and capture pollutants in the air, while protecting public health.

1.3 Unique case of Milan

Delving into the details of our project, Milan is the largest and most populous city in northern Italy, with approximately a population of 1.4 million inhabitants in the municipality and over than 3 million inhabitants in the metropolitan area. Recently, IQAir, a Swiss real-time air quality website, ranked the city as the third most polluted city in the world [13]. Despite numerous criticisms of the assessment and its methodology, many studies have shown that northern Italy has a problem with air pollution. Currently, based on a ranking by the European Environment Agency, which orders cities from the cleanest to the most polluted depending on levels of $PM2.5$, the Metropolitan area of Milan is ranked 334th with a concentration of 19.7 $\mu g/m^3$ and it's in last place among cities with at least 1 million inhabitants [14].

The air quality in the city results from the interaction of heterogeneous factors. Geographically, it's located in the centre of the Padan plain in the Lombardy region, surrounded on three sides by large mountain ranges that severely restrict the circulation of large air masses. The climate of the Padan plain is therefore continental, characterised by rather cold winters, hot summers, infrequent rainfall, and generally high relative humidity.
The annual trend in PM_{10} concentrations, like the other pollutants, shows a pronounced seasonal dependence, with higher values in the winter period, due both to the poorer dispersive capacity of the atmosphere in the colder months and due to the anthropogenic emissions [12].
The latest update of the regional inventory of atmospheric emissions, managed by ARPA Lombardy, is dated 2021 and shows that road transport is the most relevant source of particulate matter emissions in the metropolitan area of Milan with a combined increase of 12% since 2019, as shown in Table 1. The primary contributors are emissions from diesel as well as vehicle brake, tyre, and road surface wear. Followed by "other sources" macro-sector, which has risen by 49%, including urban fires, wildfires, and the non-industrial combustion macro-sector (-7%), with residential heating, particularly from the consumption of woody biomass. Overall, emissions have increased by 0.52% from 2019.

Macro-sector description	PM10	PM2.5
Combustion in energy and trasformation industries	0.97%	1.21%
Non-industrial combustion plants	21.96%	26.80%
Combustion in manufacturing industry	6.76%	6.49%
Production processes	4.11%	3.63%
Solvent and other product use	6.45%	6.93%
Road Transport	44.50%	38.03%
Other mobile sources and machinery	1.59%	1.98%
Waste treatment and disposal	0.09%	0.11%
Agriculture	0.71%	0.33%
Other sources and sinks	12.86%	14.47%

Table 1: *Emission sources of the main pollutants.* ARPA, 2021

Eu Legislation, through Directive 2008/50/EC set a daily mean value limit for PM_{10} at 50 $\mu g/m^3$, permitting up to 35 days of exceedance per year, and an average annual value of $PM_{2.5}$ at 25 mug/m^3, without a daily binding constraint. Also, the EU has strengthened air quality standards closer to the guidelines of the World

Health Organization (WHO) by 2030 with the possibility of derogation until 2035. These updated standards will reduce the PM_{10} daily mean value at 20 $\mu g/m^3$ and $PM_{2.5}$ at 10 $\mu g/m^3$.

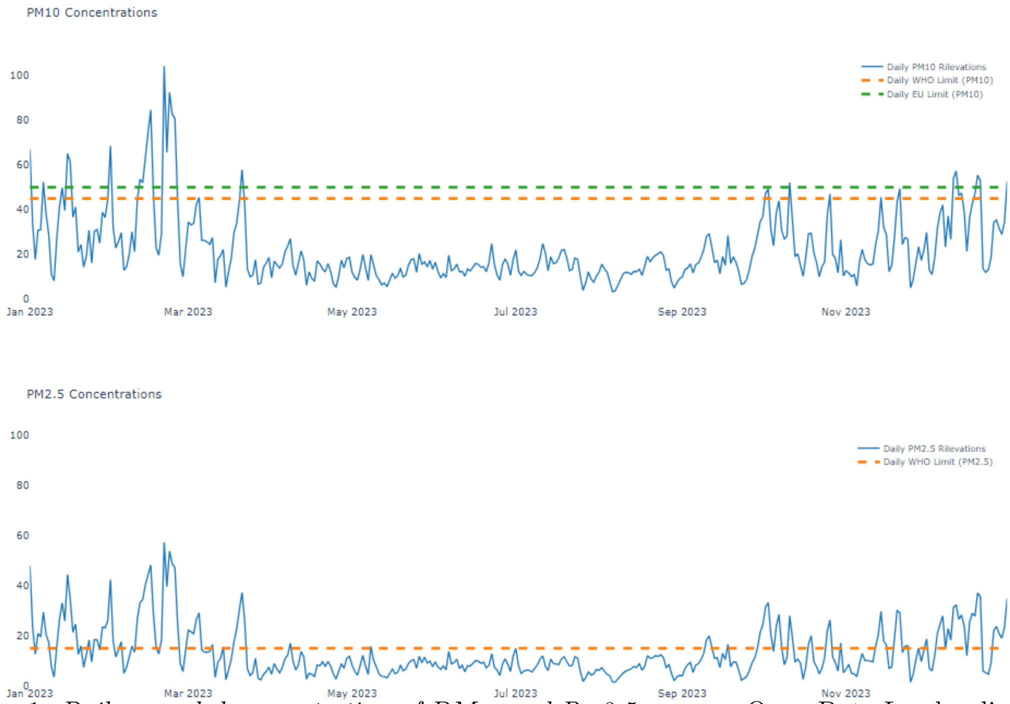


Figure 1: *Daily recorded concentration of PM_{10} and Pm2.5.* source: Open Data Lombardia, 2023

As shown in the figure above, during 2023, our period of analysis, air quality stations on average exceeded the daily EU limit of concentration of PM_{10} 22 times, with the station on Via Senato recording 49 days of violation. The situation is even worse for $PM_{2.5}$, where the daily concentration exceeded WHO guidelines for more than 107 days. However, the average annual concentration values for both PM_{10} and $PM_{2.5}$ remain within the EU limits and have decreased by 43% and 38%, respectively, since 2012.
It should also be noted that Milan's air pollution is partially due to emission sources located within the city. A study conducted in 2010 [16] by Ricerca sul Sistema Energetico highlights that approximately 65% of PM_{10} detected comes from extra-municipal contributions. Therefore, broader plans involving the metropolitan area and adjacent provinces are necessary.

1.3.1 PMs impact on communities

Short and long-term exposure to air pollution can lead to a wide adverse health effects depending on the duration, concentration of the exposure and the health status of the affected population. The people most affected by the health consequences of poor air quality are those who spend more time outdoors and are physiologically more vulnerable: children, the elderly, those with chronic illnesses, pregnant women, and unborn children. Focusing on individuals under 20 years of both sexes, the death rate due to air pollution exposure decreased by 86.55% between 1990-2021 but still remains the second cause of death as reported by IHME [25].

However, the quality of life for the entire population is compromised. World Health Organization (WHO) has found evidence of a relationship between particulate matter and lung cancer risk [15]. Based on this acknowledged evidence, they classified it as carcinogenic to humans (Group 1), as it is associated with an increase in genetic damage [17]. Short-term exposure to PM_{10} has been associated with the worsening of respiratory diseases, such as asthma, while the effects of long-term exposure are less clear. $PM_{2.5}$ with a less diameter is much more harmful because it is able to reach even the deep airways; both short and long-term exposure are associated with premature mortality, chronic bronchitis and heart causes [16].

City - Urban Audit Cities (LAU)	Tot Premature deaths	Premature deaths per 100,000 inhabitants	Years Of Life Lost
Brescia	356	154	3160
Milano (greater city)	4517	124	40100
Bolzano	90	84	798
Como	96	100	853
Gallarate	90	114	796
Bergamo	199	119	1764
Monza	214	125	1908
Milano	1882	127	16710
Varese	102	99	905
Busto Arsizio	115	115	1024
Saronno	56	118	494

Table 2: *Premature deaths and years of life lost in 2021 due to $PM_{2.5}$ exposure in Lombardy* European Environment Agency, 2021

Table 2 shows Brescia as the city with the highest premature death roll per 100,000 inhabitants in 2021, based on data retrieved from the European Environment Agency. The years of life lost (YLL) metric highlights the potential loss of years of life, considering factors such as age and sex. While Brescia has the highest rate, the Metropolitan area of Milan records the highest absolute numbers of premature deaths and YLL. In addition to its impacts on human health and social well-being, this severe disease burden results in reduced labour productivity, increased medical costs, and higher hospitalisation rates, which can put significant pressure on the national health care system.

1.3.2 Main mitigation policies

The City of Milan is part of the international C40 Cities Climate Leadership Group, an organization made up of mayors from greater cities around the world working together to promote urban decarbonization. According to a recent survey by ISTAT [18], only 11.5% of interviewed citizens are satisfied with the air quality, a very low figure compared to other European cities. As part of its local efforts, Milan has established the Piano Aria e Clima with the ambitious goal of becoming carbon neutral by 2050. The vision refers to a city that is clean and healthy, prosperous and modern, competitive and climate neutral [16]. In an effort to improve air quality and enhance the quality of life for its citizens, the city has adopted several policies aimed at reducing major sources of particulate matter. These include promoting the replacement of gas heating systems, renewing the vehicle fleet with fiscal incentives, and reducing traffic and speed limits within residential areas to 30 km/h, a measure introduced as part of the "Città 30" initiative to reduce the resuspension of particulate matter from road surfaces.

One of the main factors that has been developed in Smart Cities is Mobility. In 2012, Milan introduced a traffic restriction zone, Area C, covering 8.2 square kilometres, following a popular referendum with 79% approval, that involved the restriction of many Euro emission classes and a pollution charge that not only has decreased the PM emissions but also traffic accidents and congestion. While Area B covers almost the entire surface area of the municipality of Milan but with more lax restrictions, they contributed to reduction in traffic flow in the city centre, allowing pedestrians to reclaim public spaces and reallocating street space from fast to slow mobility. One instance, the Piazze Aperte program, aims to enhance urban public space as a social space: between 2018 and 2023, 42 tactical urbanism interventions have been completed, including 280 benches, 450 bike racks, 50 picnic tables, 38 ping pong tables, and 380 planters [20].

In tandem with these measures, policymakers have rethought urban mobility to reduce car ownership by strengthening public transportation and promoting alternative mobility options such as shared vehicles. The pandemic has enforced investment in cycle-pedestrian infrastructure (BiciBus, PediBus), with the city allocating 250 million euros to create 750 kilometres of new cycling corridors between the regional capital and the hinterland [21].

1.3.3 Nature based solutions

Many indicators can influence what people value in their living environment. In terms of life satisfaction, 84.2% of Milan residents reported being satisfied with living in their city in 2023 [18], with 15% of those surveyed indicating that the quality of life has improved compared to five years ago [18]. Nature-based solutions (NBS) have emerged in recent years, as a new approach to strengthen urban resilience and sustainability. European Union defines NBS as a strategically planned network of natural and semi-natural areas with other environmental features, designed and managed to deliver a wide range of ecosystem services [29]. These services are the benefits ecosystems provide, which humans rely on for their well-being [56]. Under this broad concept, urban forests — which comprise networks of forests, groups of trees and individual trees [30] — can provide important ecosystem services, including air purification (mitigation through carbon sequestration and storage) and adaptation to climate change, restoring of anthropized landscapes, enhancing protected areas and mitigating land consumption. In other words, they contribute to both pollutant removal and exposure reduction.
Beyond their environmental functions, they also have significant social and well-being benefits. Urban greenery reduces stress levels among the population, fosters interpersonal relationships, increases physical activity rates, and helps reduce mental health issues and aggression [23]. Clearly, green spaces serve multifunctional purposes combining social interaction, aesthetics and ecological functions. The city's policy-makers have created thirteen new parks, each exceeding 10,000 square metres, and have also revitalised existing green spaces to improve urban greenery. At present, the city's public tree heritage includes 240,000 trees (60% in parks and gardens) [31], with plans to plant 220,000 trees within 2030 in the municipality area. This effort is part of the broader urban forestation program, ForestaMI, which aims to plant three million trees across the entire Metropolitan City [22].

Another nature-based solution for increasing vegetation in urban areas is utilizing the outer surfaces of buildings and roofs. They offer the largest untapped area to provide additional green space in cities. For Vertical Greenery, building faces can be covered mostly by climbing vegetation which requires less maintenance (green-faced). The same cannot be said for living walls that are usually made up of potted plants and require more complex irrigation systems and maintenance. The air stagnation induced by the presence of tall buildings can potentially reduce the dispersion of pollutants at the pedestrian level without a persistent flow of wind, leading to increased health risks [46]. Vertical greenery could help mitigate this issue by lowering temperatures and reducing the heat island effect, which would also result in significant energy savings by decreasing the need for heating in winter and air conditioning in summer [22]. The "Milano Clever Cities" project, financed by EU, has pushed this direction with the installation of five green walls in a first phase concluded in 2023 [32]. Moreover, the city is promoting the creation of green roofs and facades by offering incentives and deductions to private entities. These initiatives are part of a broader project aimed at the renovation of the building sector, which is central to the ecological transition and improving energy efficiency in Milan's journey toward becoming fully carbon-neutral by 2040 [16].

1.3.4 Current perception of air quality information accessibility

A recent survey showed that Milan's citizens are aware of the quality of the air they breathe [40], yet 46% of them report that they rarely get information. The most common method for getting information is to query web search engines, followed by 31% who prefer online magazines, and 22% who directly consult institutional channels (Municipality website/di AMAT/Arpa). However, navigating the Amat and Arpa websites (visited 21st August 2024), we discovered that the daily air quality reports had not been updated respectively from 30th April 2024 [41] and 21st March 2024 [42]. Moreover, according to Article 18 of Legislative Decree 155/2010 [43], reporting $PM_{2.5}$ concentration levels is not mandatory, unlike PM_{10}.
Although the public is informed about PM_{10} level exceedances, both websites lack comprehensive information on air pollution exposure, symptoms to watch for, and recommended actions. This is significant, as 81% of those surveyed believe it is crucial for Milan's municipality to inform both residents and visitors about air quality [40]. In addition to details on the impact on human health, over than 80% expressed the need for information on how to minimize environmental impact and the sources of emissions [40].
Instead, at the European level, the Environment Agency's European Air Quality Index (EAQI) allows users to understand more about air quality where they live in real time. By accessing the interactive map available on the EAQI website or related app, users can find information about local air quality based on data from the nearest air monitoring station. As shown in the following table, the index is computed on five pollutants ranging from 1 (good) to 6 (extremely poor). For each pollutant, the index is computed separately according on concentration levels, as defined by World Health Organization [7], using 24-hour mean concentrations for PM_{10} and $PM_{2.5}$, and hourly mean concentrations for NO2, O3, and SO2.

Pollutant	Index level (based on pollutant concentrations in µg/m3)					
	Good	Fair	Moderate	Poor	Very poor	Extremely poor
Particles less than 2.5 µm (PM2.5)	0-10	10-20	20-25	25-50	50-75	75-800
Particles less than 10 µm (PM10)	0-20	20-40	40-50	50-100	100-150	150-1200
Nitrogen dioxide (NO2)	0-40	40-90	90-120	120-230	230-340	340-1000
Ozone (O3)	0-50	50-100	100-130	130-240	240-380	380-800
Sulphur dioxide (SO2)	0-100	100-200	200-350	350-500	500-750	750-1250

EAQI	Health messages	
	General population	Sensitive populations
Good	The air quality is good. Enjoy your usual outdoor activities	The air quality is good. Enjoy your usual outdoor activities
Fair	Enjoy your usual outdoor activities	Enjoy your usual outdoor activities
Moderate	Enjoy your usual outdoor activities	Consider reducing intense outdoor activities, if you experience symptoms
Poor	Consider reducing intense activities outdoors, if you experience symptoms such as sore eyes, a cough or sore throat	Consider reducing physical activities, particularly outdoors, especially if you experience symptoms
Very poor	Consider reducing intense activities outdoors, if you experience symptoms such as sore eyes, a cough or sore throat	Reduce physical activities, particularly outdoors, especially if you experience symptoms
Extremely poor	Reduce physical activities outdoors	Avoid physical activities outdoors

Figure 2: *The European Air Quality Index and related health messages for each band*

The polluted bands are displayed on a color scale ranging from green (indicating good air quality) to purple (indicating extremely poor air quality), reflecting the relative risk associated with short-term exposure to human health. The overall EAQI for each monitoring station is determined by the highest value among the five individual pollutant indices. For better interpretability, the index bands are complemented by health-related messages that provide recommendations for both the general population and sensitive populations.

1.4 Objective

ARPA Lombardia publishes real-time information from its sensor network daily, along with meteorological bulletins, to promote the culture of open data and support a people-centric approach to digital city trends. Meteorology significantly impacts the concentrations of air pollutants in the environment: temperature and solar radiation influence the heating and cooling requirements in homes. Likewise, wind speed, direction and humidity influence their dispersion.
In this work, we will exclusively consider the temporal year 2023, including meteorological and air quality monitoring station data, along with additional air quality data from non-official sensor stations obtained via the SensorCommunity platform. For this purpose, we have acquired daily data for the selected stations.

Our goal is to reliably predict the PM_{10} concentration at an arbitrary point placed within the city. We will employ machine learning models for this prediction, in particular boosted tree, dense neural network, and LSTM-based network. We evaluate their performances on R^2 and Mean Absolute Error (MAE).
Furthermore, we will assess that green areas lower air pollution. For this purpose, we define "green area" generally as both man-made and natural on-land vegetation in outdoor areas. To test these hypotheses, we will leverage spatial data such as buildings and greenery within several surrounding circles of different radii around the station. This data is sourced from OpenStreetMap (OSM), an open-source Volunteered Geographic Information (VGI) mapping project, which allows us to gain a clearer view of the city of Milan. Finally, we will build a map of the city with PM_{10} predictions every few hundred meters, and assess the impact of some modifications to the Milan urbanistic.

2 Data Analytics, Preparation for the Models and Prediction

2.1 Data Exploration and Preparation

As mentioned before, we retrieved the data regarding the air quality stations from two distinct sources. In the end, we will have the following monitoring stations:

- Milano Pascal Citta Studi
- Milano Verziere
- Milano v.Marche
- Milano v.Senato
- Milano v.Brera
- Milano v.Juvara
- MIlano lambrate
- Milano P.zza Zavattari

- 24644
- 32399
- 40256
- 44216
- 50128
- 70169
- 22851

where the stations in the first column are those obtained from Lombardia Open Data, and those in the second column from SensorCommunity. The stations in the first column, originating from the first source, already had official names. In contrast, we have retained the sensor IDs from the second source to identify the stations in the second column. In the Italian context, meteorological data from individual stations are considered representative for an area with a radius of 10-30 km [49]. To accurately associate meteorological data with air quality data, treating them as if they were sourced from the same location, we generated buffers with a 5-kilometre radius around each meteorological station and assigned the nearest air quality station's data accordingly. For plotting the buffers, we converted the EPSG:4326 coordinate system, which expresses distances in degrees of latitude and longitude, to the EPSG:32632 system, which uses meters. The results are shown in map (b).

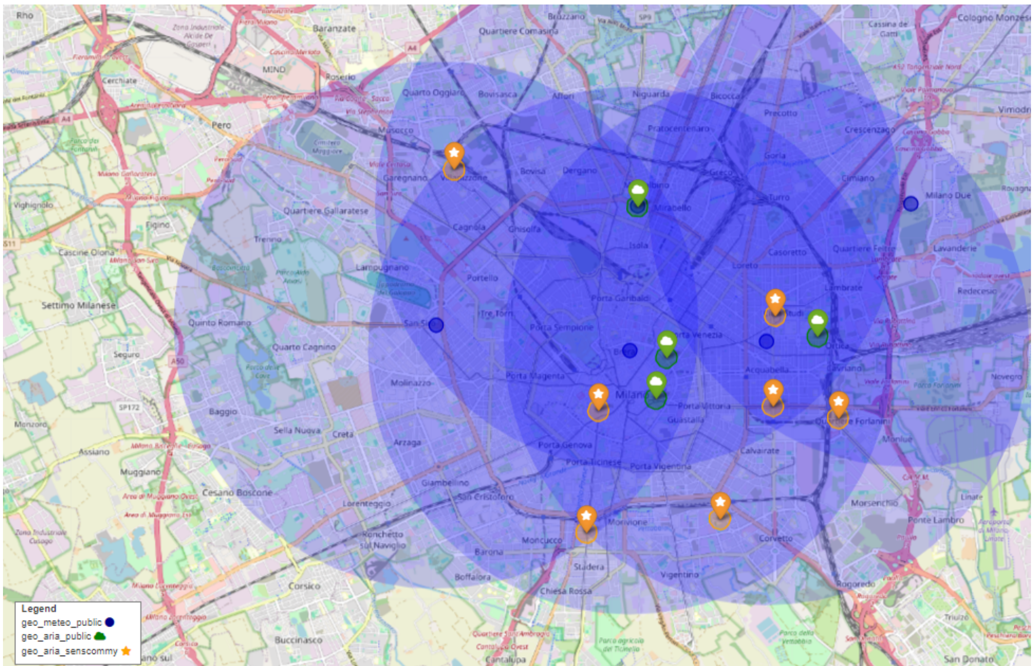


Figure 3: *Generation of buffers of 5km around meteorological stations*

On the map, the air quality stations from the institutional source are represented by green cloud icons, while the meteorological stations are shown with blue circle icons. The air quality stations from the unofficial source are depicted with yellow star icons. It is noteworthy that there is a single overlap: the Milano v.Marche station records both air and meteorological data.

From the unofficial sources, we have access to more comprehensive data than from the official sources. Both data sources are valid, the only difference is that SensorCommunity stations could be less precise due to the different hardware used (generally it is the fine dust sensor SDS011) [50]. In addition to PM_{10} measurements, which are common to both sources, we also have data on other chemical compounds, such as $PM_{2.5}$, benzene (C6H6), carbon monoxide (CO), nitrogen dioxide (NO2), tropospheric ozone (O3), and sulfur dioxide (SO2). Although these pollutants interact with PM_{10}, potentially contributing to the formation of secondary particulate matter [24], for the purpose of our work, we have decided to exclude them from our analysis.

Furthermore, on OpenStreetMap are provided a wide range of natural and artificial physical land features. To reduce the complexity and potential bias in our models, we decided to decrease the number of attributes by grouping them into broader categories based on the available descriptions. Therefore, we decreased the number of natural attributes from 21 to 6 by categorizing them according to their physical geographical features: trees, vegetation, forest, rock and sand. Similarly, we reduced the number of artificial attributes from 93 to 16 by grouping them based on their functional roles within the community, as shown in the figure below.

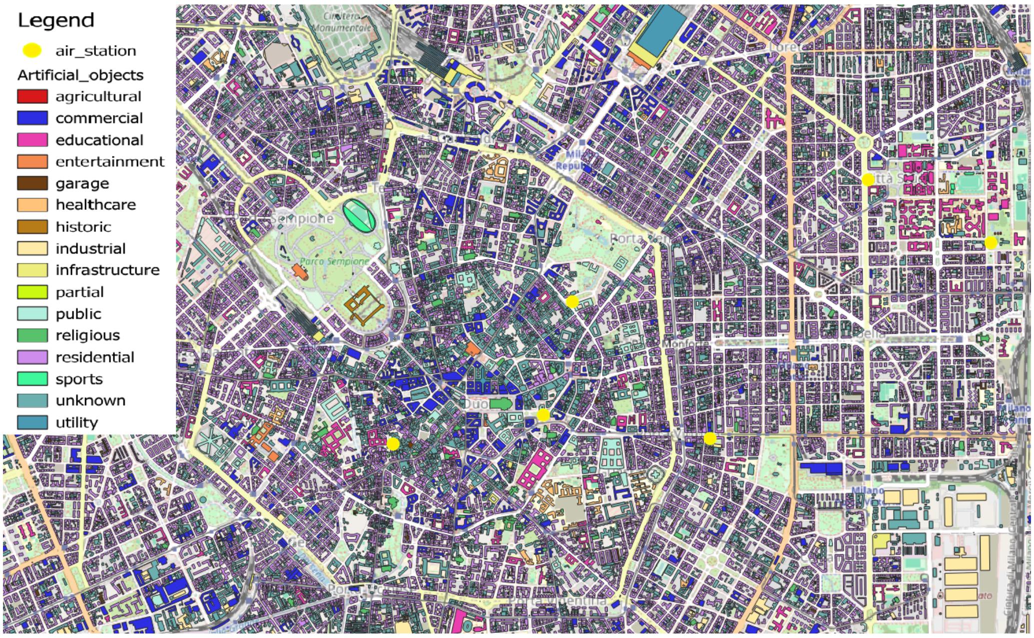


Figure 4: *Artificial objects retrieved from OpenStreetMap*

All the spatial elements are evaluated by the model based on the area. Regarding natural elements, we encountered non-bidimensional data types, so we had to estimate the area for such points or lines. We assigned a radius of $3.97m$ to each tree crown based on the average crown size of the trees managed by the Municipality of Milan, while a radius of $2.0m$ was assigned to non-trees vegetation based on our experience. Regarding buildings, we had not to estimate any area, but we made the decision to multiply the area by the number of floors. This choice obviously creates a bias in the data, that does not represent the area of a building anymore, but rather its livable area, as a proxy for the number of people using the building and also while also considering the space denied to the original green. This approach is based on findings from the literature, which indicate that air tends to stagnate at the base of buildings, potentially reducing the dispersion of pollutants [46].

2.2 Spatial features encoding: rings generation

Having obtained the list of stations located in the area and their coordinates, our goal is to determine the overall area of natural and artificial objects surrounding each station. Using the data retrieved from OpenStreetMaps, we generate several buffers of different sizes around each station, using the station's coordinates as the common center. These buffers have breakpoints at a radius of 150 meters for the near-station surroundings, 800 meters for the middle ring, and 2000 meters for the far elements. This approach allows us to quantify the natural and artificial objects within these rings of increasing distance from each station while ensuring that no ring extends beyond the city's perimeter.
Moreover, the data will not be recorded cumulatively. This means that as we expand the radius of the ring, we will only consider elements that are unique to each ring. So, we won't consider the circle with a 150-meter radius and the one with an 800-meter radius, but rather the ring with a 150-meter radius and the one with a radius from 150 to 800 meters. This method ensures that each object is counted only once within a specific

ring, reducing the risk of dependence between the features during the machine-learning phase. The generation of different rings around an air quality station is illustrated below.

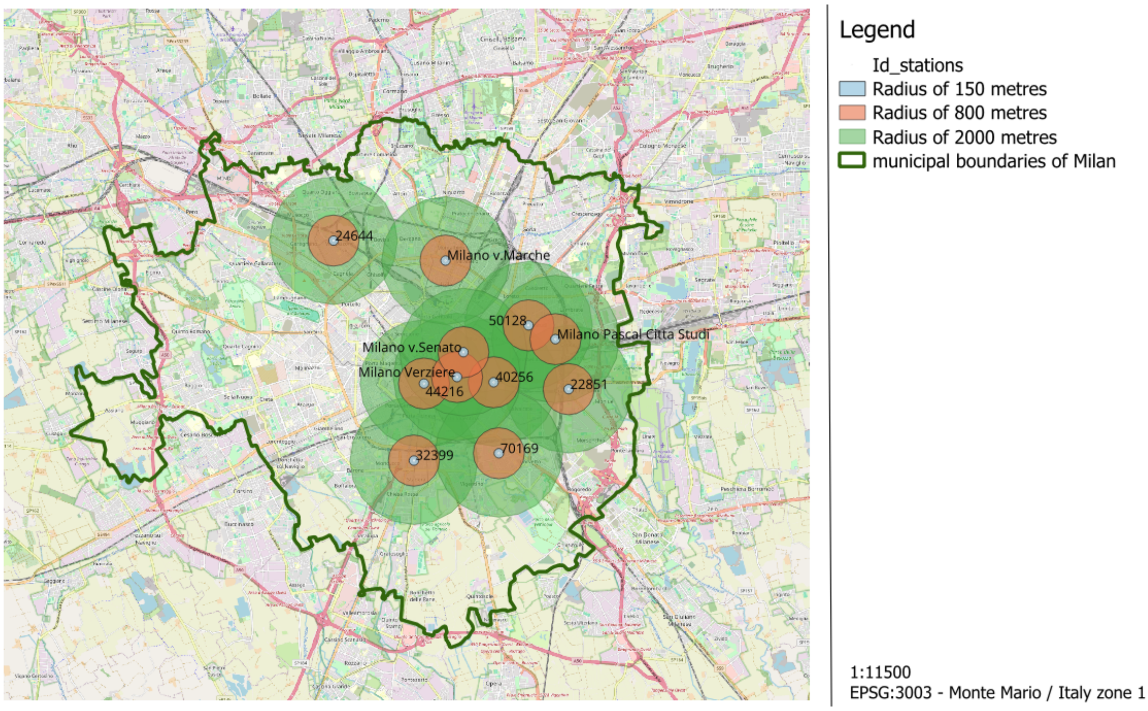


Figure 5: *different rings generation around an air quality station*

2.3 Machine Learning models

Quantifying exposure to air pollution is important in developing effective mitigation policies [51]. Approaches for air pollution modelling include deterministic modelling (based on atmospheric dynamics and chemical processes), statistical modelling, and machine learning modelling [52]. We do not have the expertise to create a deterministic model, so we evaluated the other two possibilities.
The statistical approach most specific for the task is Land use regression (LUR) models, using geospatial data to develop prediction models in environmental and health sciences [53]. Over the past decade, there has been a notable development around LUR models, especially regarding multi-source observations (fixed monitoring stations, mobile monitoring stations, satellites), a wider range of air pollutants, and the integration of advanced statistical methods [54].
Regarding the machine learning scenario, the evolution is as fast as that of the LURs. Nonlinear ML methods like the tree-based and SVMs successfully handle the data non-linearity with promising performances, but if they may be better than statistical models in handling space covariates, ML models are usually worse with the time dimension [55]. Deep Learning models are promising and increasingly popular, the main categories are RNN-based, CNN/RNN-based, and more recently transformer-based [55].
The reviewed literature contains many different models, especially depending on the available data and the covered area, but also varying much in the technique applied [33]. For example, we have the Aurora model from Bodnar et al., 20204 at Microsoft Research that is structured in a way suitable to model the Globe working with weather and climate data [58], at the same time we have MasterGNN from Hat at al., 2021 at Baidu Research that predicts together all the main kinds of pollution and at the same time all the main variables for weather conditions, based on Heterogeneous Recurrent Graph Neural Network and Multi-Adversarial Learning, tested successfully at a city level in Beijing and Shanghai [59].

Unfortunately, for our task we could not find any suitable model in the literature: we need one at a city level, daily temporal granularity, considering points of interest of different categories, predicting one pollutant with high spatial precision, not requiring traffic data (it's not freely available for Milan) nor advanced satellite-level data.

Then we proceeded to build our own model taking inspiration from the successful literature.
We developed three models of increasing complexity:

- Machine Learning: boosted tree (based on our previous experience on the same dataset and [60] and [61])
- Deep Learning dense-based: a simple dense neural network, to put the attention on the aggregated spatial surroundings features
- Deep Learning RNN-based: a network based on LSTMs and the previous point of this list we called Surrounding-Augmented LSTM SA-LSTM (referencing the work of [62] and [63])
- We did not explore the CNN/RNN-based solutions, combining the convolutional and the recurrent approaches in a single layer

The boosted tree model operates by building an ensemble of decision trees, where each tree corrects the errors of the previous ones. Each tree is trained on the residuals of the previous model, and the predictions are combined to produce the final output. This pure ML model is very explainable respect to the other deep-learning options we considered.

The dense network uses a decoder structure from the 102 covariates to one neuron at the end. The hidden layers have sizes of 64, 32, and 16 units, and all of them use ReLU activation. As a standard dense neural network, each neuron takes the previous layer's outputs and performs a weighted sum followed by the activation function. The multiple layers allow the model to learn relatively complex patterns.

Model	**R^2 on test set**	**MAE on test set**
Boosted Tree	49.00%	8.60
Dense NN	51.90%	8.16
SA-LSTM	84.59%	4.17

Table 3: Evaluation on our Milan dataset for each model

2.4 SA-LSTM

In the LSTM model, we based our predictions on sequences from the previous 7 days. To achieve this, we distinguish two types of inputs: time-dependent attributes, which include weather and date-related variables (36 covariates in total), and time-invariant attributes, which capture spatial information from the surroundings of the air quality station, aggregated by distance and category as explained in sections 2.1 Data Exploration and Preparation and section 2.2 Spatial features encoding (66 covariates in total). For each observation, the model ingested a sequence of 7 days of time-dependent data per station plus the time-invariant attributes, resulting in a data chunk with 318 values. Notably, We chose to not add the previous day's PM_{10} values in the inputs to avoid having the model predictions dependent on his own past predictions. This design ensures to forecast further into the future with high accuracy.

The two input types are processed differently in the initial stages: the time-dependent covariates are fed into an LSTM layer of 50 units, while the time-invariant covariates are processed through a Dense layer with 32 neurons and ReLU activation. The outputs from both layers are then concatenated and passed through a final Dense layer with 32 neurons and sigmoid activation.

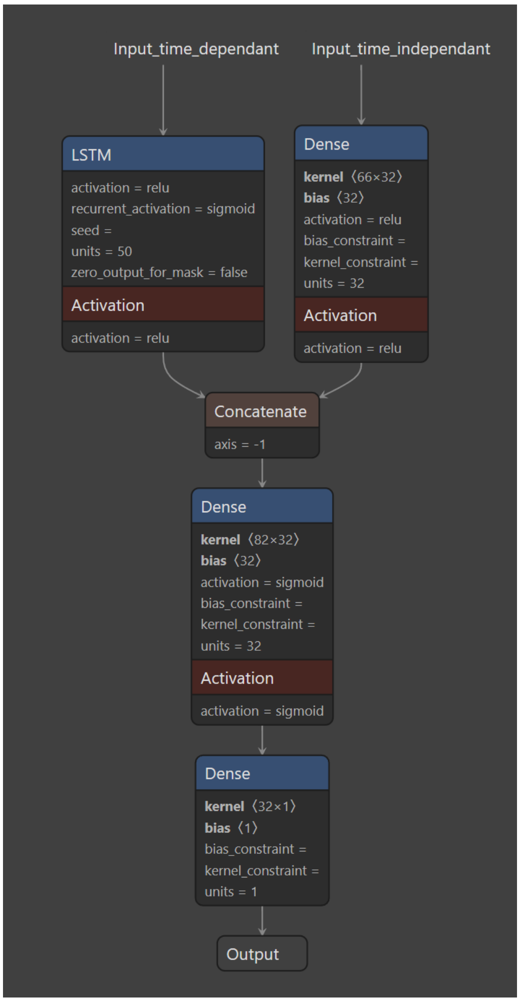


Figure 6: *Model architecture*

The dataset is split into training (80%), validation (10%), and testing (10%) sets. For the evaluation, we employed the R-squared (R2) metric and mean absolute error (MAE). To follow the general DNN guidelines, given the performance increases, we used a MinMaxScaler to adjust the values fed into the model. To ensure consistency in the evaluation phase, all MAE and R^2 values in this paper are computed after an inverse transform of the scaler on the network output values. After 600 epochs, the model achieved an MAE of 4.16, which is a state-of-the-art result [62]. There is slight overfitting, possibly due to the limited amount of data (just one year), but the gap is minimal, within a hundredths of a point.

2.5 Model use cases

The first use case is our primary objective: a high spatial precision prediction at a point in the city. It is obtained from the best model (LSTM) using 7 days of weather forecasts, the date, and the surrounding data of the chosen point. For this task, we made an example predicting PM_{10} pollution at City Life on the 15th of August with the following results:

- sunny: 81.76453 (previous days: sunny, sunny, sunny, sunny, sunny, sunny, sunny)
- cloudy: 72.670235 (previous days: sunny, sunny, cloudy, sunny, sunny, cloudy, cloudy)
- rainy: 42.609287 (previous days: sunny, sunny, cloudy, cloudy, rainy, cloudy, rainy)
- stormy: 35.58673 (previous days: sunny, sunny, sunny, sunny, rainy, rainy, stormy)

The high spatial precision is obtained thanks to the online process generating the surroundings covariates, lasting as little as 1.29 seconds on average for one point to process all green areas and buildings of the city. For this reason, it's possible to create a high-precision map of the pollution in Milan. Unfortunately, being the model based on the surrounding data with a radius of 2km, we are unable to provide reliable results for points of the city nearer than 2km to the city borders. With our resources (free Google Colaboratory workspace), we managed to get a map of pollution in the City with one point every 500 meters, but of course, it is possible to thicken the points grid more. Below is an example prediction for a sunny 30th of January after a sunny week.

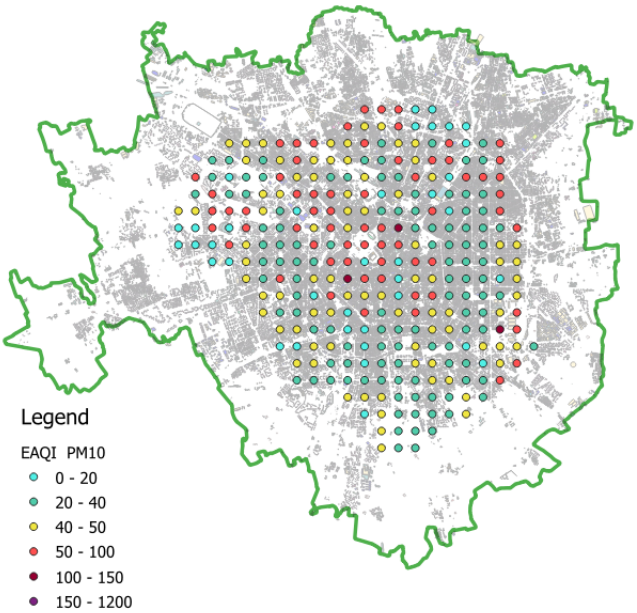


Figure 7: *PM prediction In Milan every 500 metres with color categorized according to EAQI*

As for the last objective, it is possible since we have control over the map data on which is based the surroundings covariates generation. We decided to assess with our model the impact on PM_{10} of the Scalo Farini project: a new 300 000 square meter area in the form of parks and trees [64]. Below there is a map depicting the difference of the pollution in Milan with and without the presence of the project, on the same day of the previous experiment.

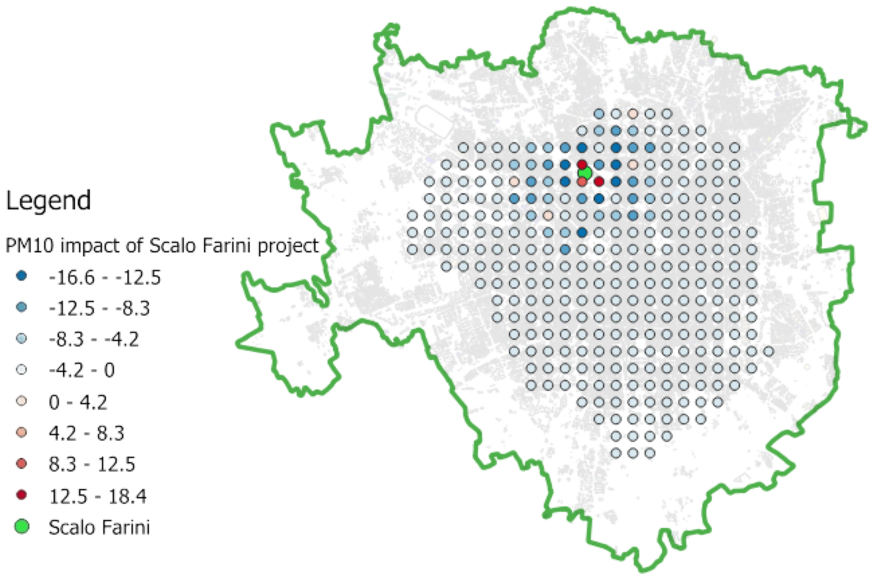


Figure 8: PM_{10} *impact of Scalo Farini*

The figure shows that the Scalo Farini project has a positive impact on the reduction of PM_{10} concentration values more in the distant areas than in the close proximity, as highlighted by the red-coloured dots. We hypothesise that this may be due to the lack of green areas calculated in the first 150-metre ring around the air quality stations, used as input for the model.

All the ways in which the tool can be used are interesting. Firstly, it is useful for pollution prediction tasks (from both private and public entities). We estimate the network to be effective also with other pollutants [62]. The high spatial granularity, mixed with an appealing user experience and interface, could enhance citizen interests in monitoring the air quality and consequently being more aware and possibly taking action.
Lastly, it is a very interesting tool for urbanistic and city planners, who may want to assess the impact on pollution for a project. In the example in the previous subsection we used a new green area, but the possibilities include new buildings of different kinds, other types of green areas and also the destruction of some points of interest.

3 Policy suggestions

3.1 Urban trees

We have already highlighted the multifaceted benefits of urban trees, as detailed in section 1.3.3, primarily through their ability to remove particulate matter from the air. This is mainly achieved through the process of dry deposition, where fine particles settle onto surfaces like leaves and are either absorbed by the plant or later removed by rain or dispersed by the wind [33]. The effectiveness of urban trees in providing ecosystem services is highly dependent on the specific traits of each tree species and the deposition velocity [33]. Determining the deposition velocity is complex as it is strongly affected by both external factors (meteorological conditions, pollutant concentration and composition) and tree traits [57]. The tree characteristics most relevant include canopy density (crown size, shape, and density), foliage longevity (e.g., leaf shape, surface properties, and physiology), and the emission of reactive compounds. For example, in a study [34], Weerakkody et al., 2018 identified several beneficial leaf traits for capturing particulate matter (PM), including small size due to their larger edge effect, complex shape, and hairy or waxy surfaces, which are more effective than leaves with a smooth surface. PM embedded in the wax layer cannot be resuspended leading to permanent removal of PM from the air [57].
Planting trees is one of the most straightforward and cost-effective nature-based solutions that can be implemented to improve urban air quality. In Milan's parks, Carpinus betulus is the most common tree species, accounting for 3.42% of all trees managed by the municipality [36]. In contrast, along the city's streets, the most common species is Platanus x acerifolia, representing 5.56% of the urban tree population [36]. However, these species does not have a significant impact on particulate matter sequestration [35].
A recent study by Zappitelli et al., 2023 [35] mapped and measured the capacity of urban greenery in Milan to sequester both PM_{10} and $PM_{2.5}$, using the AIRTREE multi-layer canopy model.This model integrates soil, plant, and atmospheric processes to simulate the exchange of atmospheric pollutants. As the same top-10 family tree ranking was obtained for both pollutants, we reported only the findings related to PM_{10}. The study found that the tree family that sequesters the highest amount of the pollutant is the evergreen Pinus family, with a value of 184.174 $PM_{10}\ g\,m^2\,y^{-1}$.

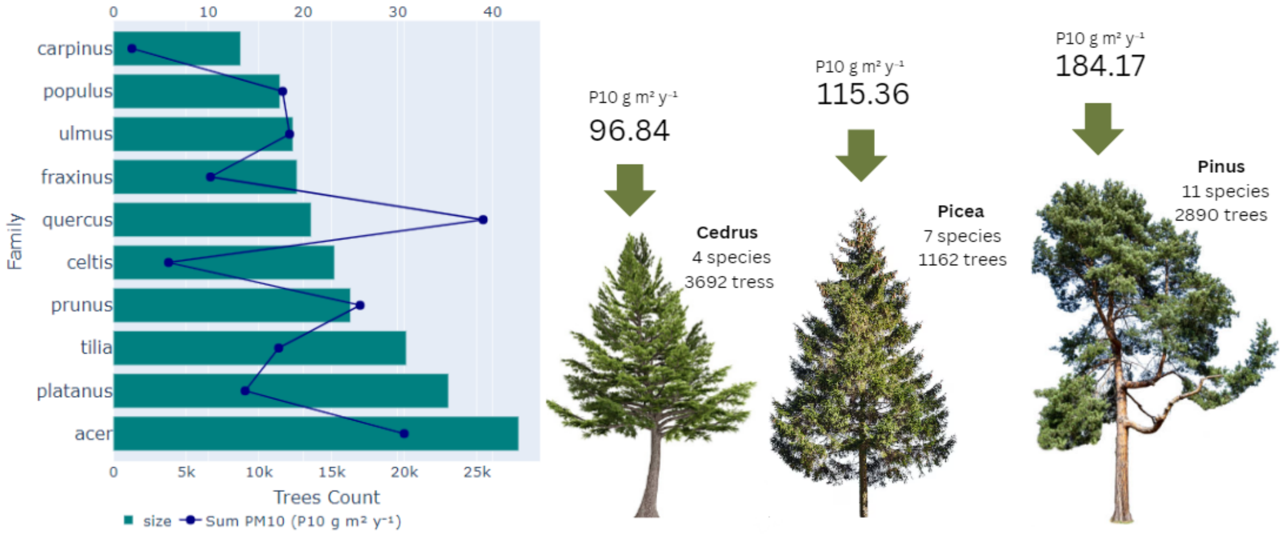


Figure 9: *PM_{10} removal indicator reported by tree family ([35]) compared to Municipality-managed tree population*

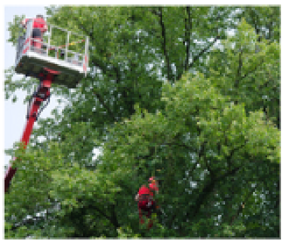
High sequestration values are also observed in the Picea and Cedrus families, with values of 115.363 PM_{10} $g\,m^2\,y^{-1}$ and 96.843 PM_{10} $g\,m^2\,y^{-1}$, respectively. As shown in figure 7, these three families account for a small portion of Milan's public tree inventory compared to more common species. Among individual species, the highest sequestration values are found in Cedrus libani (with only 119 trees in the municipality) and Taxodium distichum (322 trees). Conifers are particularly effective due to their dense, needle-like leaves, which create a larger surface area and a more complex leaf structure (larger area per cm^3, not per leaf), enhancing their ability to intercept and retain particulate matter [35]. Moreover, these are examples of evergreen trees, which lose and renew their leaves continuously throughout the year. Investing in the planting of evergreen species can help mitigate air pollution while promoting multiple ecosystem services.
People's health and well-being are closely linked to the state of the environment: especially on the social dimension and the quality of life perception, fostering a sense of place and enhancing connections with other people. The city should change its approach by resorting to trees – not just for aesthetic purposes but also for their functionality in addressing multiple factors – which can be an effective way of tackling air pollution and offer significant benefits to the community.

It's also important to acknowledge green area planning and management costs can be very challenging. Urban green spaces can be associated with ecosystem disservices, defined as perceived or actual negative impacts on human well-being [46]. These disservices include issues such as plant allergies and unpleasant smells, as well as tradeoffs like increased water consumption and wildfire risk, exacerbated by the increased intense heating due to climate change [37]. In addition, they need ongoing maintenance throughout their entire life cycle especially in times of increasing frequency of intense weather events, as seen in recent decades in Lombardy [48]. When thunderstorms and heavy windstorms occur, they are capable of knocking down a large number of trees in just a few minutes, damaging properties and utility services, leading to increased costs for municipalities to compensate for these incidents and restore the trees.

For policymakers, urban planners and stakeholders, it becomes important to handle these various aspects carefully to ensure the best tree species are selected for specific spatial planning contexts and local environments. Also, improper pruning practices could represent a threat to tree stability and a possible performance reduction, by decreasing the leaf area available. Muscan et al., 2024 [38], emphasizes that more frequent and less intense ordinary pruning scenarios are preferable for optimal tree management. In this regard, the municipality has recently adopted a policy of reduced mowing in areas where it does not interfere with the use of public spaces [44]. Designing green areas with high PM mitigation capacity requires integrating several considerations about ecosystem disservices and services into the decision-making process, as outlined in the table

tree roots lift the pavement

fallen tree after cloudburst



tree pruning

- Favourable physiological, morphological and micromorphological characteristics
- Low allergenicity
- High resistance to disease, and greater resilience to pollution and climate change
- Low maintenance requirements
- Respect for local biodiversity

Table 4: *considerations for the design of green areas with high PM_{10} sequestration capacity*

Developing practical indicators that measure these factors and incorporate them into public sector decision-making processes is crucial, in conjunction with many other considerations not covered in this work. Although many of these risks cannot be eliminated, a detailed risk assessment coupled with guidance on prevention and mitigation best practices can help reduce the loss of services provided by trees and the associated costs, ensuring continued benefits for citizens.

3.2 Increasing public participation

The City of Milan has been involving citizens in the maintenance of urban green spaces since 2012. This initiative primarily focuses on areas surrounding buildings, such as tree-lined parterres and flowerbeds [39]. To further promote and expand these efforts, more extensive awareness campaigns could be implemented to encourage greater citizen participation in the care and preservation of urban green spaces.

With the numerous agreements and ongoing initiatives – at least on paper – the municipality of Milan has demonstrated a proactive approach to enhancing environmental resilience while striving to improve the quality of life for its citizens. However, in line with expert recommendations, the promotion of environmental information and education as practical tools both enhance pro-environmental behaviour and contribute to the reduction of environmental impact [45].

Public acceptance of these initiatives can be challenging. For instance, a survey revealed that only 45% [40] of respondents are willing to change their habits to improve the air quality in their city, despite being aware of the harmful effects of air pollution on human health. This underlines the need for environmental initiatives that effectively translate pro-active environmental behaviour into the private sphere. Addressing these challenges involves not only improving accessibility to information and the benefits of trees, but also ensuring clarity of goals, maintaining continuous dialogue throughout the various project phases and understanding the needs of residents. Currently, according with "Sistema Informativo Valutazione Ambientale Strategica" (SIVAS), public information and participation methods are provided through the municipality's website and the SIVAS website. But, this approach may not be user-friendly for younger generations. For more direct, rapid, and impactful communication, new social marketing strategies should be adopted, such as promoting sustainable and healthy lifestyles, sharing messages about the various disservices and services provided by trees, and using social media to reach more people (although with more superficiality). All this could be an effective way to increase citizen awareness and involvement.

3.3 Future developments

Our analysis underlined how green areas are seen as crucial drivers to increase livability and make the territory more resilient. Enhancing public participation is essential to fostering a sense of belonging and responsibility among residents for their urban environment. At the same time, policymakers should address the public's need for accessible information. An institutional geographic web app, usable on both computers and smartphones, could greatly benefit urban residents. Similar to AirVisual by IQAir and EAQI, this app would offer real-time air quality data, detailed weather conditions for all locations within the municipality, and other relevant information, including guidance on minimizing environmental impact, identifying emission sources, and health-related advice.

Regarding our model, it could be used iteratively to create a pollution impact assessment for a new project in many different locations. This can be useful to optimize the best spot for that project (relative to decreasing as much as possible the impact on the pollution).

Our work demonstrates that predicting pollution in Milan with high precision is feasible, though it would be better to integrate additional data, such as traffic variables (that we couldn't find available for free) and information on major emission sources near specific locations within the city.

Lastly, it is possible to tune SA-LSTM to predict pollutants other than PM_{10}, and test it on cities different than Milan to assess if it really can compete with state-of-the-art models.

References

[1] *Ourworldindata article*

[2] OECD. *Regions at a Glance, 2013*

[3] *World Health Organization article*

[4] Tamara Kulesa, Susanne Dirks. *A Vision of Smarter Cities: How Cities Can Lead the Way into a Prosperous and Sustainable Future, 2009*

[5] Umang Singh, Ajith Abraham, Arturas Kaklauskas, Tzung-Pei Hong. *The Role of Smart Sensors in Smart City, 2021*

[6] G.V. Pereira, P. Parycek, E. Falco, R. Kleinhans. *Smart governance in the context of smart cities: A literature review, 2018*

[7] World Health Organization. *WHO global air quality guidelines, 2021*

[8] Emmanuel Adinyira, Samuel Oteng-Seifah, Theophilus Adjei-Kumi. *A Review of Urban Sustainability Assessment Methodologies , 2007*

[9] Daniel J. Jacob, Darrell A. Winner. *Effect of climate change on air quality, 2009*

[10] World Health Organization. *Economic cost of the health impact of air pollution in Europe, 2010*

[11] Bart Ostro, Joseph V. Spadaro, Sophie Gumy, Pierpaolo Mudu, Yewande Awe, Francesco Forastiere, Annette Peters. *Assessing the recent estimates of the global burden of disease for ambient air pollution: Methodological changes and implications for low- and middle-income countries, 2018*

[12] Concessioni autostradali Lombarde. *Progetto di monitoraggio ambientale*

[13] *IQAir article*

[14] *European Air Quality Index*

[15] World Health Organization. *Outdoor particulate matter exposure and lung cancer: a systematic review and meta-analysis*

[16] Municipality of Milan. *Piano aria e Clima, 2020*

[17] WHO. *Outdoor air pollution, 2016*

[18] Istat. *Qualità della vita nelle città italiane: un confronto europeo, 2023*

[19] Comitato per il Capitale Naturale. *Quinto rapporto sullo stato del capitale naturale in Italia, 2022*

[20] Municipality of Milan. *Piazze Aperte, 2023*

[21] Metropolitan City of Milan. *Progetto Biciplan, 2021*

[22] ForestaMI. *Report, 2020*

[23] FAO Alberitalia. *Report, 2020*

[24] Ministero dell'Ambiente e della sicurezza energetica. *Studio sul Particolato Secondario*

[25] Institute For Health Metrics and Evaluation. *report, 2021*

[26] Federico Karagulian, Claudio A. Belis, Carlos Francisco C. Dora, Annette M. Prüss-Ustün, Sophie Bonjour, Heather Adair-Rohani, Markus Amann. *Contributions to cities' ambient particulate matter (PM): A systematic review of local source contributions at global level, 2015*

[27] Upuli Perera, Ayanda Roji, Lauren Andres, Collins Adjei Mensah. *Enhancing quality of life through the lens of green spaces: A systematic review approach, 2016*

[28] Istat. *Ambiente urbano, 2022*

[29] *Article on European Union website*

[30] European Cooperation in science Technology. *Memorandum of Understanding for the implementation of the COST Action "European Network for the Integrative Approach of Urban Forestry" (INTUF) CA23148, 2024*

[31] Municipality of Milan. *Memorandum of Understanding for the implementation of the COST Action "European Network for the Integrative Approach of Urban Forestry" (INTUF) CA23148, 2024*

[32] *Clever Milano*

[33] Cristian Bodnar, Wessel P. Bruinsma, Ana Lucic, Megan Stanley, Johannes Brandstetter, Patrick Garvan, Maik Riechert, Jonathan Weyn, Haiyu Dong, Anna Vaughan, Jayesh K. Gupta, Kit Tambiratnam, Alex Archibald, Elizabeth Heider, Max Welling, Richard E. Turner, Paris Perdikaris. *Aurora: A Foundation Model of the Atmosphere, 2024*

[34] Weerakkody, U., Dover, J. W., Mitchell, Reiling. *Evaluating the impact of individual leaf traits on atmospheric particulate matter accumulation using natural and synthetic leaves. Urban For. Urban Green, 2018*

[35] Ilaria Zappitelli, Adriano Conte, Alessandro Alivernini, Sandro Finardi and Silvano Fares. *Species-Specific Contribution to Atmospheric*

Carbon and Pollutant Removal: Case Studies in Two Italian Municipalities, 2023

[36] Dataset available on Comune di Milano webpage. *Territorio: localizzazione degli alberi, 2023*

[37] Paulo Pereira, Francesc Baró. *Greening the city: Thriving for biodiversity and sustainability, 2022*

[38] Desirée Muscas, Fabio Orlandi, Roberto Petrucci, Chiara Proietti, Luigia Ruga, Marco Fornaciari. *Effects of urban tree pruning on ecosystem services performance , 2024*

[39] Article on MilanoToday. *Effects of urban tree pruning on ecosystem services performance , 2023*

[40] Mosaic report. *Risultati del questionario su informazione e percezione della qualità dell'aria a Milano , 2023*

[41] Agenzia Mobilità Ambiente e Territorio. *emissions map*

[42] Arpa Lombardia, *emissions map*

[43] *Legislative Decree 155/2010*

[44] Article on Comune di MIlano webpage. *Verde. In 54 parchi della città erba alta per favorire la biodiversità, contrastare le isole di calore e preservare il suolo dalla siccità, 2024*

[45] Genovaitė Liobikienė, and Mykolas Simas Poškus. *The Importance of Environmental Knowledge for Private and Public Sphere Pro-Environmental Behavior: Modifying the Value-Belief-Norm Theory, 2019*

[46] S. Palmieri, G. Durante, A. M. Siani, G. R. Casale. *The Atmospheric stagnation episodes and hospital admissions, 2008*

[47] C. M. Shackleton, S. Ruwanza, G. K. Sinasson Sanni, S. Bennett, P. De Lacy, R. Modipa, N. Mtati, M. Sachikonye and G. Thondhlana. *Unpacking Pandora's Box: Understanding and Categorising Ecosystem Disservices for Environmental Management and Human Wellbeing, 2016*

[48] Gaia Vaglio Laurin, Saverio Francini, Tania Luti, Gherardo Chirici, Francesco Pirotti, Dario Papale. *Satellite open data to monitor forest damage caused by extreme climate-induced events: a case study of the Vaia storm in Northern Italy, 2020*

[49] Massimo Crespi. *Caratteristiche e rappresentatività della metereologia di precisione nel contesto nazionale italiano, 2020*

[50] *Sensor.Community article*

[51] Xuying Ma, Ian Longley, Jay Gao and Jennifer Salmond. *Evaluating the Effect of Ambient Concentrations, Route Choices, and Environmental (in)Justice on Students' Dose of Ambient NO2 While Walking to School at Population Scales, 2020*

[52] Zhen Zhang, Shiqing Zhang, Caimei Chen, Jiwei Yuan. *A systematic survey of air quality prediction based on deep learning, 2024*

[53] Anna Mölter and Sarah Lindley. *Developing land use regression models for environmental science research using the XLUR tool – More than a one-trick pony, 2021*

[54] Xuying, Bin Zou, Jun Deng, Jay Gao, Ian Longley, Shun Xiao, Bin Guo, Yarui Wu, Tingting Xu, Xin Xu, Xiaosha Yang, Xiaoqi Wang, Zelei Tan, Yifan Wang, Lidia Morawska, Jennifer Salmond. *A comprehensive review of the development of land use regression approaches for modeling spatiotemporal variations of ambient air pollution: A perspective from 2011 to 2023, 2024*

[55] Zhen Zhang , Shiqing Zhang, Caimei Chen, Jiwei Yuan. *A systematic survey of air quality prediction based on deep learning, 2024*

[56] European Union. *Ecosystem services*

[57] Jenny Lindén, Malin Gustafsson, Johan Uddling, Ågot Watne, Håkan Pleijel. *Air pollution removal through deposition on urban vegetation: The importance of vegetation characteristics, 2023*

[58] Cristian Bodnar, Wessel P. Bruinsma, Ana Lucic, Megan Stanley, Johannes Brandstetter, Patrick Garvan, Maik Riechert, Jonathan Weyn, Haiyu Dong, Anna Vaughan, Jayesh K. Gupta, Kit Tambiratnam, Alex Archibald, Elizabeth Heider, Max Welling, Richard E. Turner, Paris Perdikaris. *Aurora: A Foundation Model of the Atmosphere, 2023*

[59] Jindong Han, Hao Liu, Hengshu Zhu, Hui Xiong and Dejing Dou . *Joint Air Quality and Weather Prediction Based on Multi-Adversarial Spatiotemporal Networks, 2021*

[60] Man Tat Lei, Joana Monjardino, Luisa Mendes, David Gonçalves and Francisco Ferreira. *Macao air quality forecast using statistical methods, 2019*

[61] Zhen Zhang, Shiqing Zhang, Caimei Chen, Jiwei Yuan. *A systematic survey of air quality prediction based on deep learning, 2024*

[62] Zhen Zhang, Shiqing Zhang, Caimei Chen, Jiwei Yuan. *A systematic survey of air quality prediction based on deep learning, 2024*

[63] Ghufran Isam Drewil, Riyadh Jabbar Al-Bahadili. *Air pollution prediction using LSTM deep learning and metaheuristics algorithms, 2022*

[64] Milano Città di stato article. *I 18 grandi progetti di rigenerazione che trasformeranno Milano e il suo hinterland*