Politecnico di Torino

Politecnico di Torino

Environmental and land engineering

A.y. 2025/2026

Graduation session: March 2026

Chemical Composition and variability of PM_{10} in the Po Valley:

Spatial Patterns, Seasonal Behavior and Meteorological Drivers

Supervisors:

Marco Ravina
Deborah Panepinto

Candidate:

George Costin Bortun

Abstract

This thesis analyses the chemical composition and variability of PM10 in the Po Valley, within the framework of the LIFE PrepAIR project; The study aims to investigate the PM10 trends, its seasonality and its chemical composition, exploring the influence of emission sources and meteorological drivers, verifying the compliances with the current and future targets and limits defined by the European directives and World Health Organization guidelines. The analysis examined daily samples collected between 2018 and 2019 at five urban and rural background stations, regarding inorganic ions, trace elements, organic and elemental carbon (OC/EC), and levoglucosan. The Analysis of PM composition showed that the main components in all the sites were secondary inorganic species and organic carbon, with pronounced seasonal variability, associated with winter atmospheric stagnation and seasonal differences in emission patterns. Urban – rural comparisons showed distinct compositional variability. Urban sites revealed a more pronounced signature of primary combustion sources, while the rural site exhibited a higher contribution of secondary components, consistent with a stronger influence of secondary aerosols formation and reduced proximity to direct sources. Trace metals were also considered, confirming the compliance with legislative limits while revealing spatial differences with higher nickel levels at the rural site and relatively higher copper concentrations in the urban environment, suggesting distinct sources-related influences. Meteorological analysis on the Torino site confirmed the role of temperature, radiation, precipitation, wind speed and stagnation in controlling PM10 and its component variability. The correlation analysis revealed the differences in linearity, monotony and strength between the studied relationships, depending on the component and the meteorological parameter chosen. A data – driven regression learning approach was subsequently applied. Despite the relatively limited size and dimensionality of the available dataset, ensemble models demonstrated satisfactory capabilities in reproducing PM10 variability and in confirming the patterns identified in the descriptive analysis. Predictive performances improved when the model was trained with the meteorological data and short-term lag variables were included, suggesting the relevance of persistence effects in PM10 variability. Overall, the study offers an integrated assessment of PM10 chemical composition in the Po Valley, linking spatial variability and seasonal behaviour to source related processes and atmospheric controls.

Contents

Chapter 1

Definition and classification of PM

Figure 1: Turin air pollution - (ENVI.info, 2024)

For many people living in the Po Valley, low air quality is not an abstract concept but a real daily experience, particularly during winter months when the air often feels dense and difficult to breathe.

This perception reflects a well documented environmental reality: the Po Valley is one of the most affected regions in Europe in terms of air pollution, due to its complex orography, high population density and significant anthropogenic emission sources. This poses substantial risk to the health of its inhabitants.(European Environment Agency, 2025)

Among regulated air pollutants, the most widespread are nitrogen oxides NO_x, Ozone O_3 and Particulate Matter PM. The aim of this chapter is to introduce PM, describing its definition, classification, formation mechanism and health effects.

1.1. Definition

According to the European Environmental Agency (EEA), PM is a generic term used to indicate solid or liquid particles suspended in the atmosphere, often indicated also as aerosols; this includes a wide range of particles both from natural and anthropogenic sources, as mineral dust, combustion derived particles, organic compounds, biological material.(European Environment Agency, 2025)

Usually PM is classified by its aerodynamic properties, that allow to study its transport, fate in the environment and possible risk to human health. The key parameter used to classification is the aerodynamic diameter. "The aerodynamic diameter of a particle is defined as the diameter of a sphere, whose density is 1g/cm³ (density of water), which settle in still air at the same velocity as the particle in question" (*WHO Global Air Quality Guidelines*, 2021)

For regulatory and monitoring purposes, PM is divided into size fractions such as PM_{10}, $PM_{2.5}$ and PM_1, in order to refer to particles with an aerodynamic diameter less than or equal to 10 μm, 2,5 μm, 1μm, respectively.

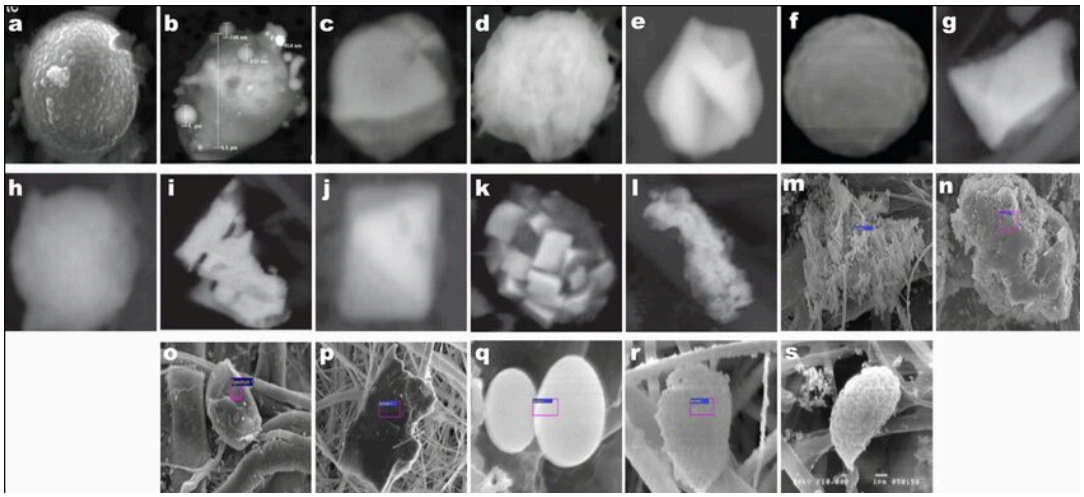
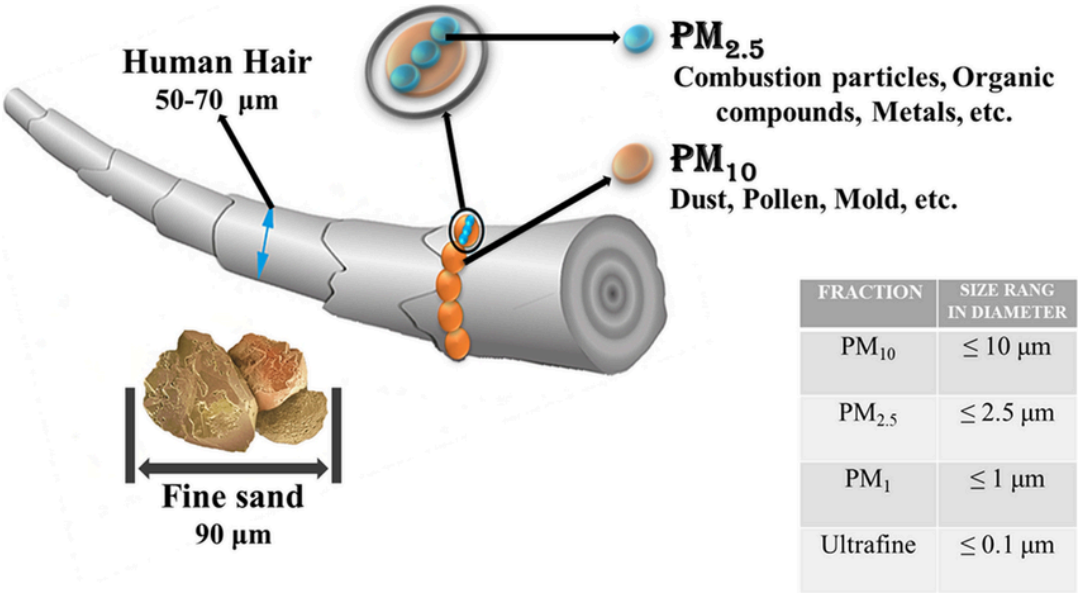


Figure 2: Morphological features of atmospheric particulate matters - (Sharma & Taneja, 2022)

To provide a reference scale, the diameter of a human hair lies between 50 and 100 μm and the comparison with PM is visible in the following figure:



FRACTION	SIZE RANG IN DIAMETER
PM_{10}	≤ 10 µm
$PM_{2.5}$	≤ 2.5 µm
PM_1	≤ 1 µm
Ultrafine	≤ 0.1 µm

Figure 3: Different types of particulate matters and their size comparison -(Kulkarni et al., 2011)

The Whitby model is one of the most fundamental frameworks used to classify atmospheric aerosols as a trimodal distribution: it is shown that PM consist in three distinct size modes or peaks, each resulting from different physical and chemical processes and with a different mass and number distribution (Figure 4).

- Nuclei mode (0.005 - 0.1 μm): contains the highest number of particles but has the lowest contribute to the total mass. They are formed through photochemical reactions and combustion processes. These particles are transient, they tend to coagulate rapidly because of their small size and high mobility, growing into the accumulation mode.
- Accumulation mode (0.1 - 2 μm): it's the bulk of the fine particle mass and it's the most stable category of particles since they are too large to coagulate but too small to settle out easily by gravity. They result from the coagulation of nuclei mode particles and from condensation of vapors. The main components of this class of particles are sulfates, nitrates, ammonium and organic/elemental carbon.
- Coarse mode (> 2 μm): it represents a significant portion of the total aerosol mass but has the lowest number of particles. these are originated mainly from mechan-

ical processes like wind erosion, volcanic eruptions and sea-wave breaking. This class of particles are less stable than the previous one, since they tend to sediment easily because of their relatively high mass. (Kulkarni et al., 2011)

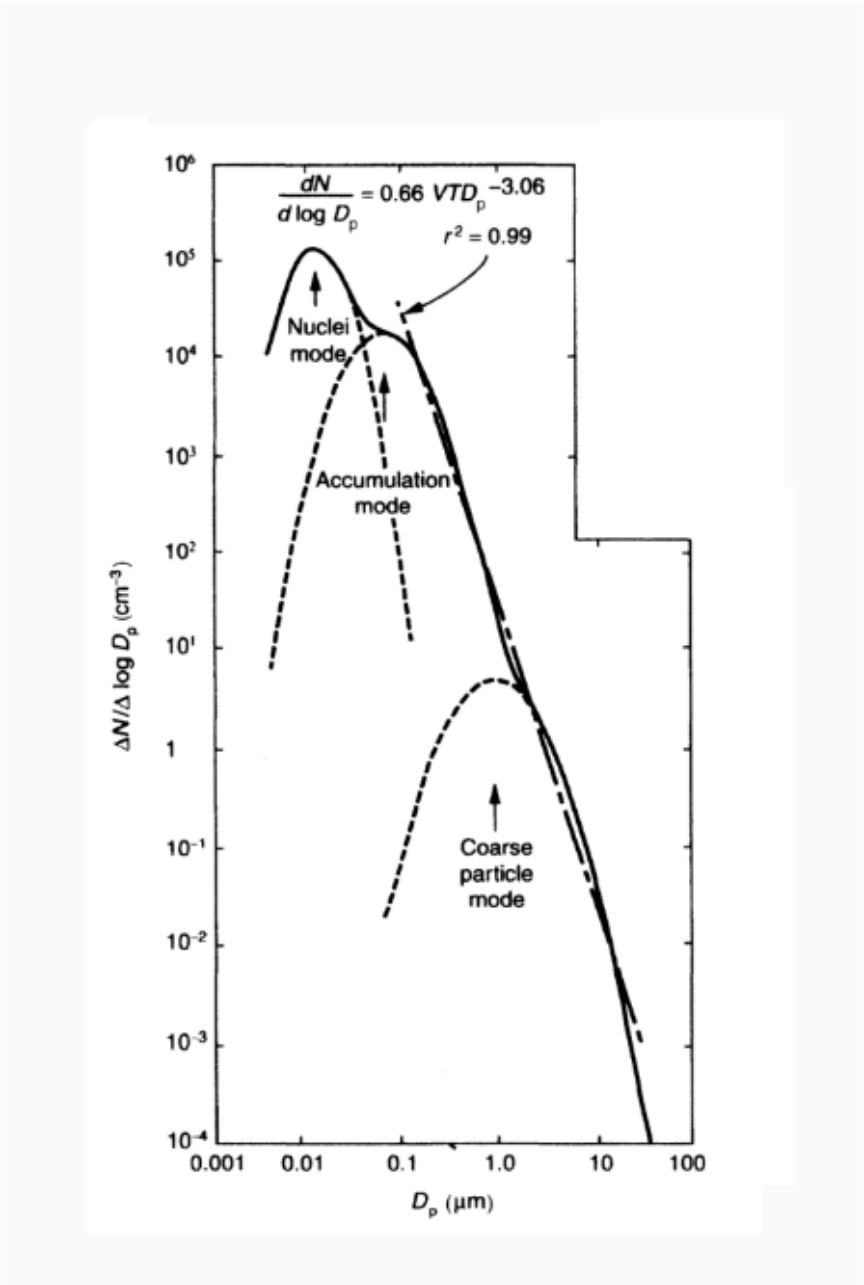


Figure 4: Number distribution of an average urban model aerosol - (Kulkarni et al., 2011)

1.2. Chemical composition and sources

Particulate matter is a chemically heterogenous mixture and its composition reflects emission sources and atmospheric transformation processes. PM components are in fact classified conventionally as primary or secondary species. The first ones are directly emitted into the atmosphere from natural or anthropogenic sources, while the second ones are formed directly in atmosphere through chemical reactions: from gaseous precursors to gas-to-particle conversion and growth processes (Putaud et al., 2010; Zhang et al., 2015)

1.2.1. Primary PM

Primary PM often are the core for further chemical growth and , among them, the most relevant components are:

- Elemental Carbon (EC): a tracer for incomplete combustion processes, in particular for vehicular traffic, usually the concentration of EC in PM is lower in rural areas and increases significantly in kerbside. EC is largely confined to the fine fraction and it's associated with the accumulation mode.
- Mineral dust (Crustal material): originated from the suspension of soil or dust from the ground, the concentration of this component can be particularly high in southern Europe due to wind transport from African deserts. Crustal material is mainly associated with the coarse mode.
- Sea Salt: naturally derived from sea spray, the concentration therefore is higher near the coasts. As the crustal material is mainly associated with the coarse mode.
- Primary organic matter(OM): it is often a major component of PM10 at almost every sites and it's originated by combustion and biomass burning. It's largely present in the fine particulate fraction, related to the accumulation mode.

1.2.2. Secondary PM

secondary inorganic aerosols constitute a major fraction of $PM_{2.5}$ and are formed through chemical reactions between gases, mainly Sulfur Dioxide SO_2, Nitrogen Oxides NO_x and Ammonia NH_3 ; the main secondary PM components are:

- Nitrate (NO_3^-): Formed from NO_x (whose sources are mainly traffic and industry), in form of nitric acid (HNO_3) reacting with ammonia (NH_3) to form ammonium nitrate (NH_4NO_3). These reactions involve other oxidants such as VOC_s, Ozone (O_3), hydroperoxyl radicals, organic peroxy radicals and halogens oxides. These

reactions are temperature dependant making the concentration of nitrates higher during winter.
- Sulfate (SO_4^{2-}) : resulting from the oxidation of SO_2 , usually emitted from industrial activity.
- Ammonium (NH_4^+) as a result of the neutralization reactions between ammonia (NH_3), often emitted by agricultural sources, and sulfuric (H_2SO_4) or nitric (HNO_3) acids already present in the atmosphere.
- Secondary organic aerosols (SOA): formed from the oxidation of Volatile Organic Compounds (VOC_s). These reactions form a complex mixture, often difficult to fully characterize.

(LIFE PREPAIR Project, 2023; Putaud et al., 2010; Raffaelli et al., 2020)

1.2.3. Dynamic evolution and interactions

It is important to see PM formation as the result of a dynamic continuum of processes, rather than a static combination of components (Figure 5) . It begins with primary emissions and new particle formation: at molecular scale, gas-phase species undergo nucleation, resulting in molecular clusters and subsequently new particles. They may interact between them and bigger particles primary emitted as well. As particles age their chemical composition evolves through various reactions and also their optical properties may change, enhancing their ability to act as cloud nuclei (CCN) or ice nuclei (IN). During their lifetime, PM can be transported from local to regional scales, depending on the interaction with meteorological conditions; they can also be removed by dry deposition (gravitational settling) and wet deposition (scavenging by precipitation).

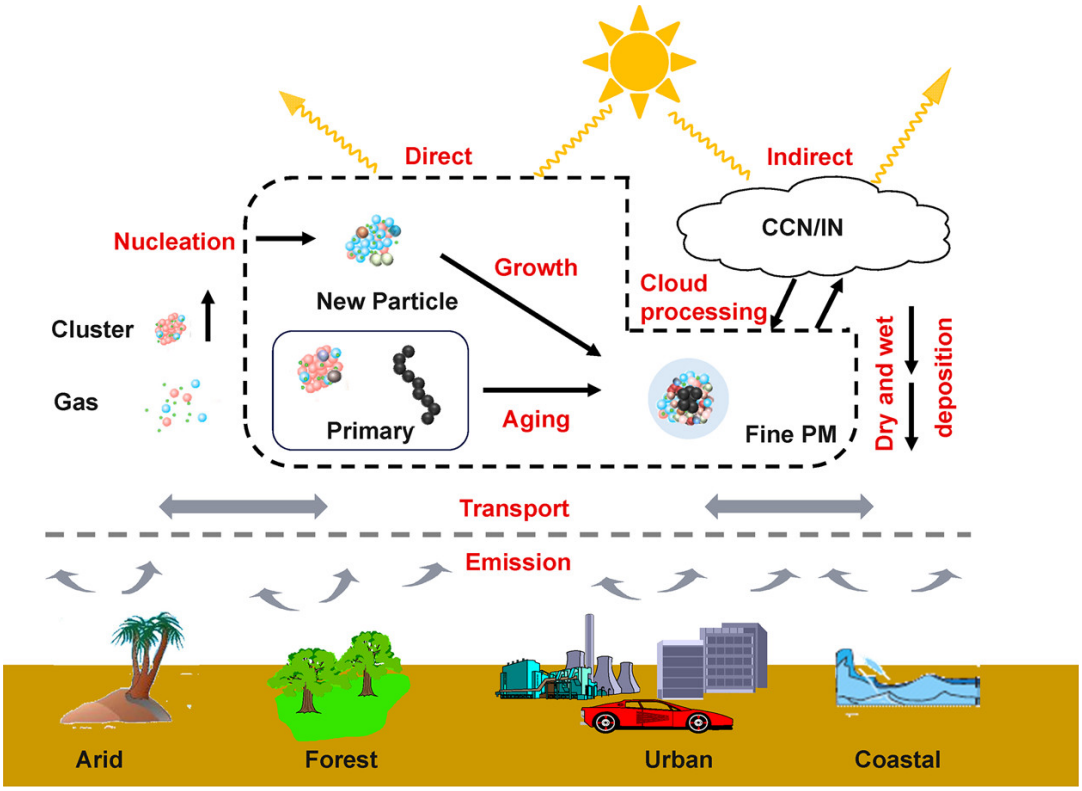


Figure 5: PM formation and evolution scheme - (Zhang et al., 2015)

1.2.4. Global observations of PM chemical composition

In the last two decades, many monitoring networks were created in order to improve the measurements of air pollutants as PM and its chemical characterization, both at regional and global scales. Among the global networks, SPARTAN (Surface PARTiculate mAtter Network) represent one of the most remarkable coordinated projects, collecting data since 2013 about $PM_{2.5}$ mass and chemical composition at ground level in multiple continents (North and South America, Europe, Africa, East and South Asia. From the result published by Snider et al. (2015), summarized in Figure 6, the components identified are:

- ANO_3 Ammonium nitrate
- ASO_4 Ammoniated sulfate
- CM crustal material
- EBC equivalent black carbon
- TEO trace elemental oxides

- RM residual matter, mainly associated with organics
- PWB particle-bound water

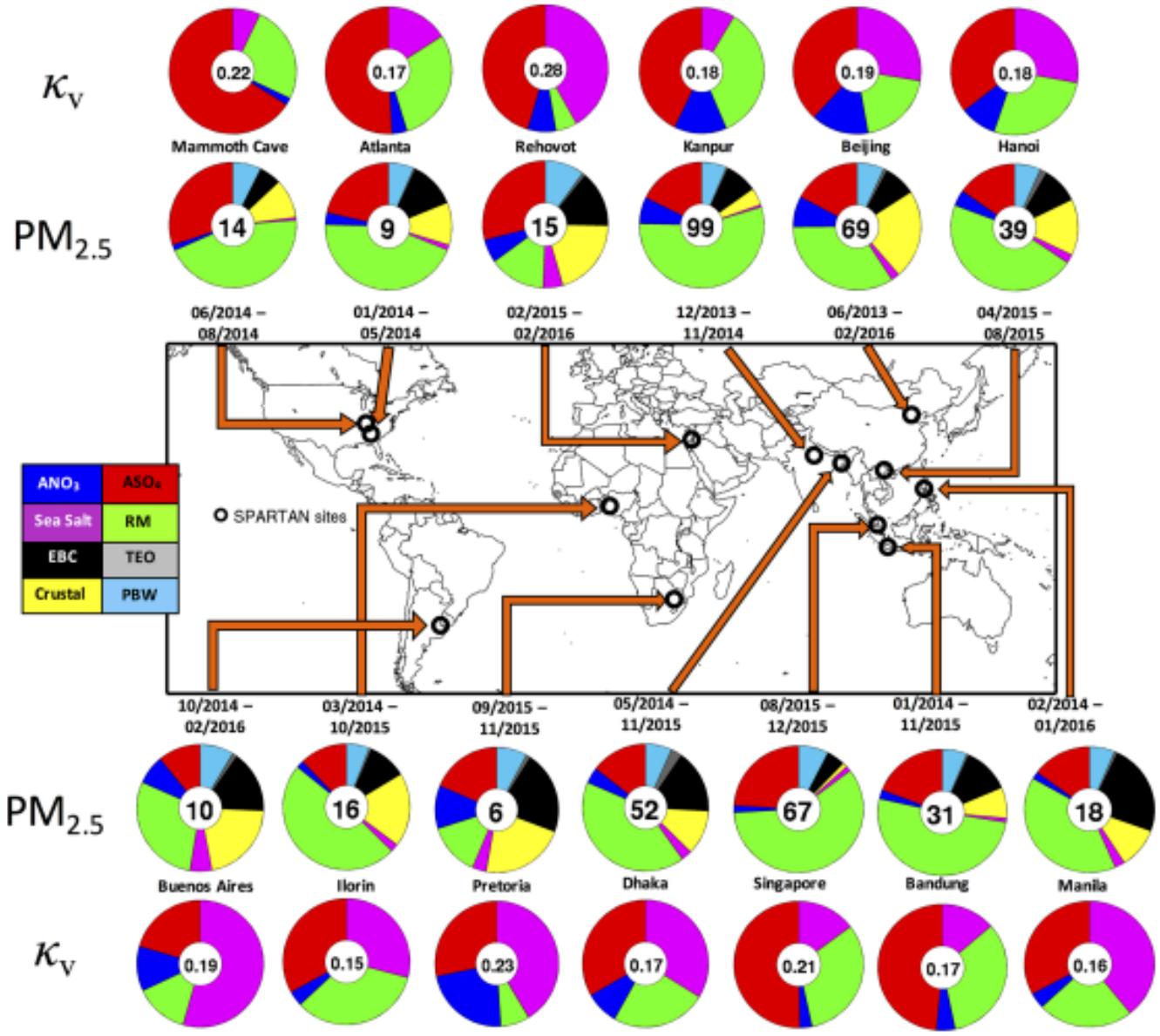


Figure 6: Map of the chemical composition of SPARTAN sites - (Snider et al., 2016)

Although the relative contributions vary geographically, ammoniated sulfate, crustal material and equivalent black carbon frequently represent major fractions of the total $PM_{2.5}$ mass. Residual matter, related to organics and unknown material, also represents a significant share, but it is also the fraction characterized by a higher uncertainty due to its heterogeneous nature.

At a regional scale, the European network EMEP (European Monitoring and Evaluation Programme (EMEP), 2026), during intensive measurements periods, indicates that the chemical composition of PM varies significantly depending on geographical location and especially on season, as shown in Figure 7. For the $PM_{2.5}$, which represent on average the 70% approximately of total PM_{10} mass, the dominant components are:

- SIA (Secondary Inorganic Aerosols) which include nitrates, sulfates and ammonium; the first ones are more relevant during winter while the second ones during summer.
- Carbonaceous matter - Organic carbon, which contribution increases significantly due to residential wood burning and reduce atmospheric dispersion.

In contrast, PM_{10} shows a stronger influence of coarse natural components, mineral dust in fact can represent between 32% and 42% of PM_{10} mass during summer, especially in southern Europe. (Aas et al., 2012)

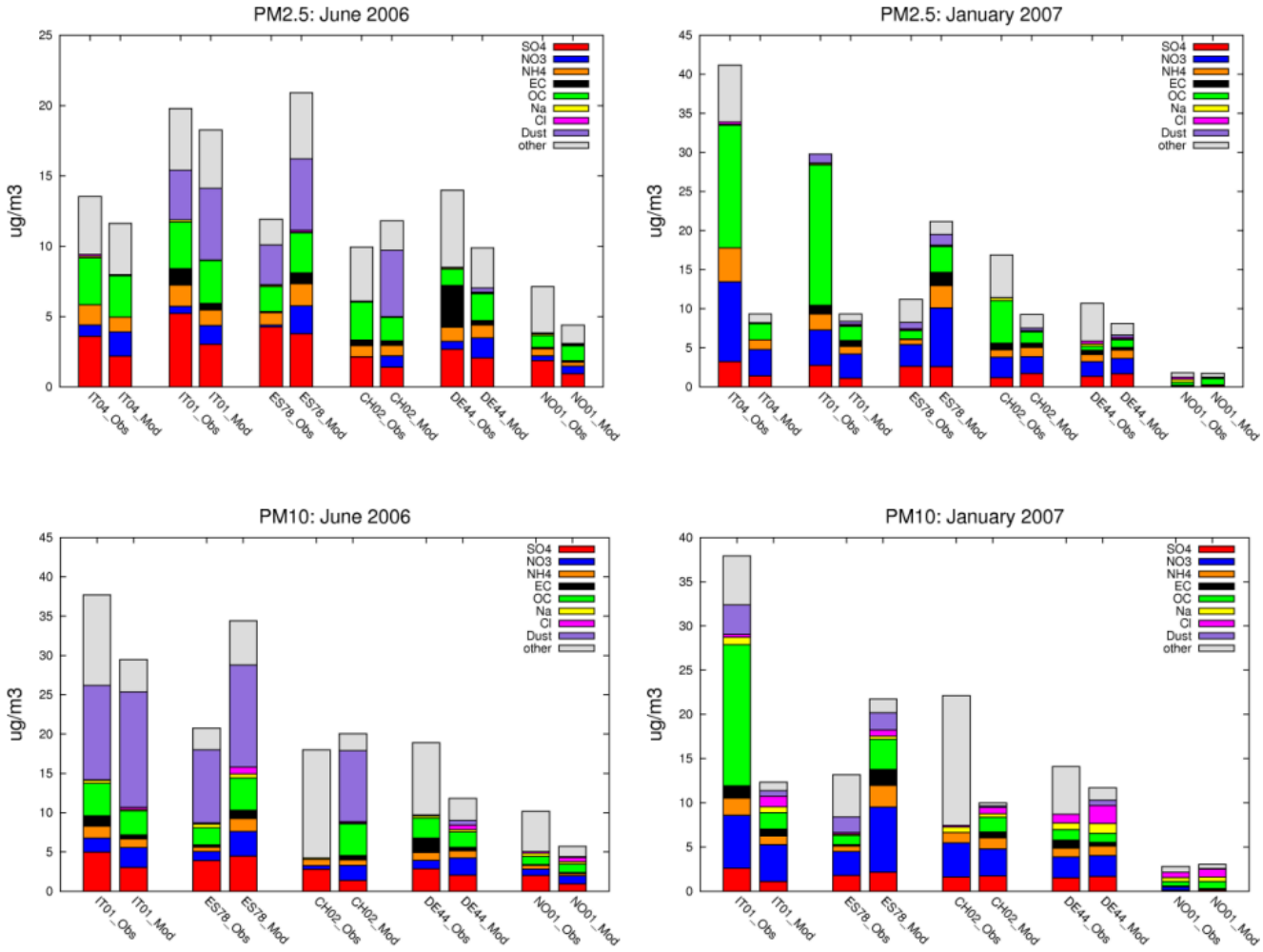


Figure 7: Spatial and seasonal variability across the EMEP sites - (Aas et al., 2012)

1.3. Current regulations and health effects

One of the most studied aspects of air pollution, it's its concentration. In fact, it's essential to understand the link between exposure (define by concentration and duration) and health effects. This link it's the baseline from which the policymakers are able to define guidelines and limit values in order to reduce the adverse effect of pollution on health.

According to the latest "Air Quality Guidelines" published by the World Health Organization (WHO) in 2021, the recommended levels that should not be exceeded are reported in Table 1:

	PM10	**PM2.5**
Annual average	15	5
Daily average	45	15

Table 1: WHO suggested limits for PM, all the values are expressed in $\mu g/m^3$.

These are not to be intended as zero-risk values, but they are set on the base of the synthesis of many epidemiological studies aimed at statistically quantifying the increase in mortality related with a unit increase in pollution concentration. The WHO recommendations strongly influence national legislation, even thought legally binding limits present usually higher values. (*WHO Global Air Quality Guidelines*, 2021)

The European Union has recently set lower limits (Table 2) with the new Directive (2024/2881) adopted on 23 October 2024, moving toward a closer alignment with WHO guidelines. This directive also introduced the planning of a strengthened monitoring system and enhanced citizen protection tools. These targets have to be respected by Member States by 2030 (*Direttiva (UE) 2024/2881 del Parlamento europeo e del Consiglio*, 2024).

	PM10	**PM2.5**
Annual average	from 40 to 20	from 25 to 10
Daily average	from 50 (35 exc.) to 45 (18 exc.)	introduced at 25 (18 exc.)

Table 2: PM limits (all values are expressed in $\mu g/m^3$) introduced by the new European Directive (2024/2881). "exc." indicates the maximum number of exceedances per year of the 24-hour limit value.

Although EU limits remain higher then the WHO suggestions,their implementation is crucial to reduce the adverse effects that PM has on our health.

According to WHO, air pollution is considered a primary risk for non-contagious-diseases (NCD) and many studies confirm it as a high mortality burden (mainly due to PM2.5), that can be approximated to 8.8 million premature deaths per year globally; that could also be seen as a Loss of Life Expectancy (LLE) of 2,9 years due to this exposure. In order to have a measure Tabacco smoke has a LLE of 2.2 years, while all the form of violence combined reach a LLE of 0.3. (Lelieveld et al., 2020)

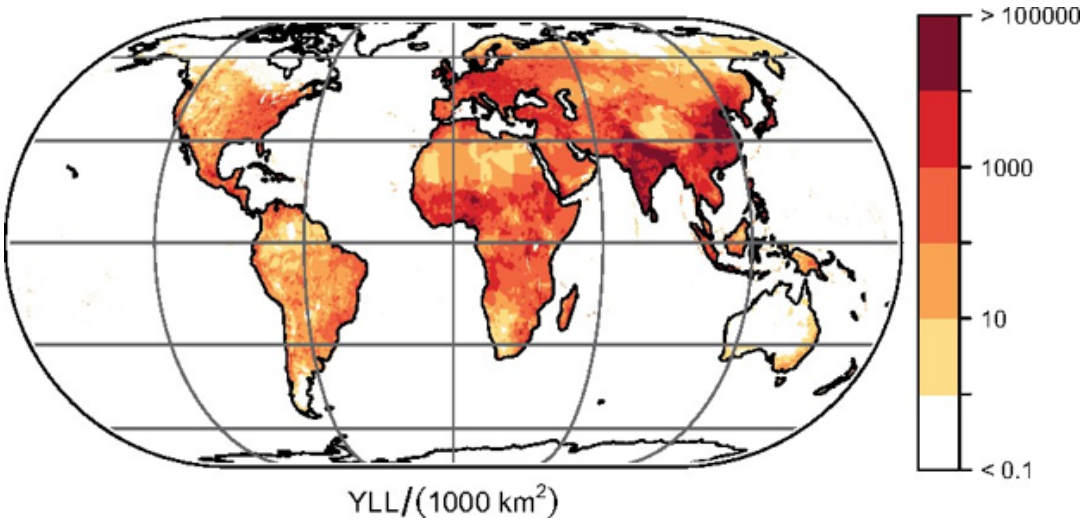


Figure 8: Annual years of life lost from air pollution (units per 1000 km². The global total is 233 million per year (Lelieveld et al., 2020)

Particle size is fundamental when investigating the health impacts of PM.

PM_{10} is the thoracic fraction that remains in the upper airways, causing asthma and bronchitis, while PM2.5 is the respirable fraction, that reaches the alveoli and consequently the bloodstream causing systemic inflammation. Ultra fine particles UFPs ($< 0.1\ \mu m$) may even translocate to the brain leading to neurodegenerative issues. (Kelly & Fussell, 2012)

Long term exposure seems to be a "high certainty" driver of mortality from ischemic heart disease, stroke, chronic obstructive pulmonary disease(COPD) and lung cancer, but it has been demonstrated that also 24-hours peaks in PM2.5 and PM10 are associated with a higher cardiovascular and respiratory mortality. (*WHO Global Air Quality Guidelines*, 2021)

Beyond particle size, toxicity and health risks are strongly influenced by the chemical composition, as shown in Figure 9.

Every component can lead to different specific adverse effect/ diseases:

- Transition metals such as zinc(Zn), copper(Cu), vanadium(V), iron (Fe) and nickel (Ni), can lead to oxidative stress, cellular damage and inflammation in both pulmonary and cardiovascular tissues. Other species particularly dangerous are lead (Pb), it can impact the nervous system, Cadmium (Cd), it may induce renal damage and allergic responses, Arsenic (As), that has been associated with cardiovascular diseases and acute respiratory infections, especially in children. Arsenic (As), Chromium(Cr), Nickel (Ni) and Lead(Pb) are also associated with carcinogenic risks. (Guo et al., 2022; Schwarze et al., 2006)
- Soluble Organic Compounds, Polycyclic aromatic hydrocarbons, in particular those related to diesel exhaust particles, are well known carcinogens
- Crustal and Mineral particles, derived from soil erosion or road abrasion (e.g. quartz, asbestos), can increase respiratory morbidity and stimulate the release of pro-inflammatory cytokines. (Schwarze et al., 2006)

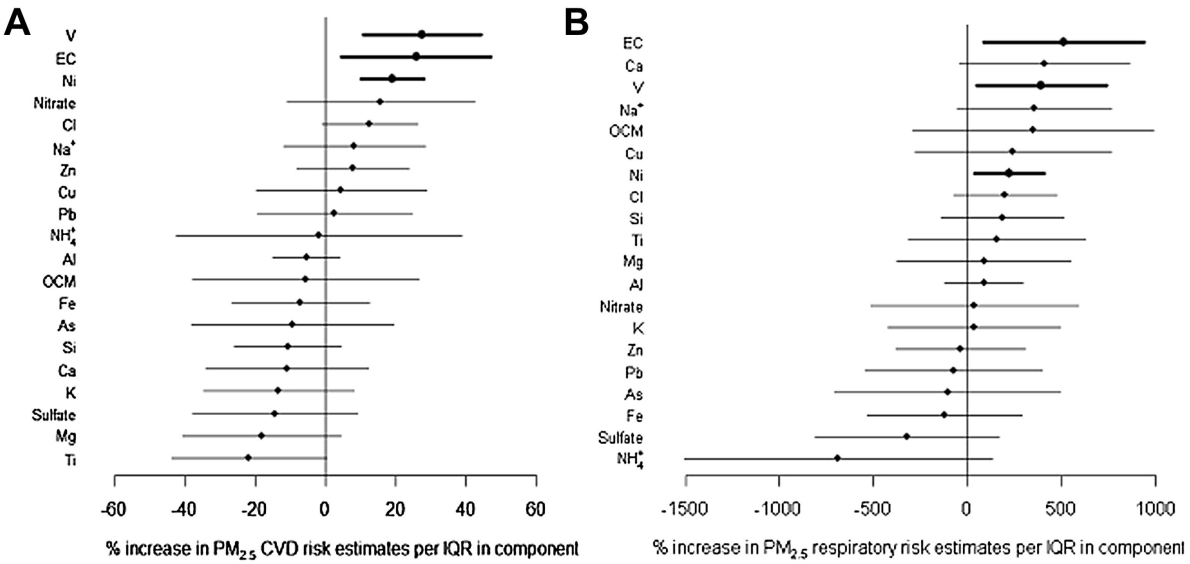


Figure 9: Association between PM2.5 chemical components and cardiovascular (A) and respiratory (B) hospitalization risk, espressed as percent increase per interquartile range (IQR) increase in component fraction, the vertical line indicates no effect (Kelly & Fussell, 2012)

Chapter 2

Methodology

2.1. Objective of the thesis

The aim of this thesis is to investigate the chemical composition and variability of PM_{10} in the Po Valley within the framework of the LIFE PrepAIR project. The result of interest is an integrated assessment of particulate matter composition, spatial differences between urban and rural environments and the relationship with the meteorological conditions. To achieve this goal, the overall objective was structured into the following sub-targets:

- Characterization of the main components of PM, including inorganic ions, trace elements, organic and elemental carbon and levoglucosan, evaluating their trends and average concentrations.
- Comparison of urban and rural sites results, in order to assess the differences in primary and secondary contributions and the spatial variability of PM.
- Investigation of the role of the meteorological parameters in controlling PM_{10} and its chemical components
- Verification of the compliances with current and future European limits and WHO guidelines values.
- Application of a data-driven regression learning approach to support the interpretation of the observed patterns in chemical composition and to explore the predictive range for PM_{10} overall concentration.

2.2. Geographical context: the Po Valley

Figure 10: 3D satellite view of the Po Valley source: google earth

The Po Valley is a unique and complex geographic and environmental context in Northern Italy, characterized by a particular topography, that, as said already in the introduction, creates a set of challenges impacting directly the air quality.

The Po valley can be described as a predominantly flat basin enclosed by major topographic reliefs: the Alps to the North and West, and the Apennines and various hilly ranges to the South and East. These structures act like natural barriers for air masses exchanges and its role is decisive for the microclimate of the region, characterized by frequent periods of low wind speed or total calm. in fact it is the area with the lowest mean wind speed in Europe with values between 2 and 2.5 m/s that during winter reach 1.5 m/s (at 10 m height from ground level) (Figure 11). (Caserini et al., 2017)

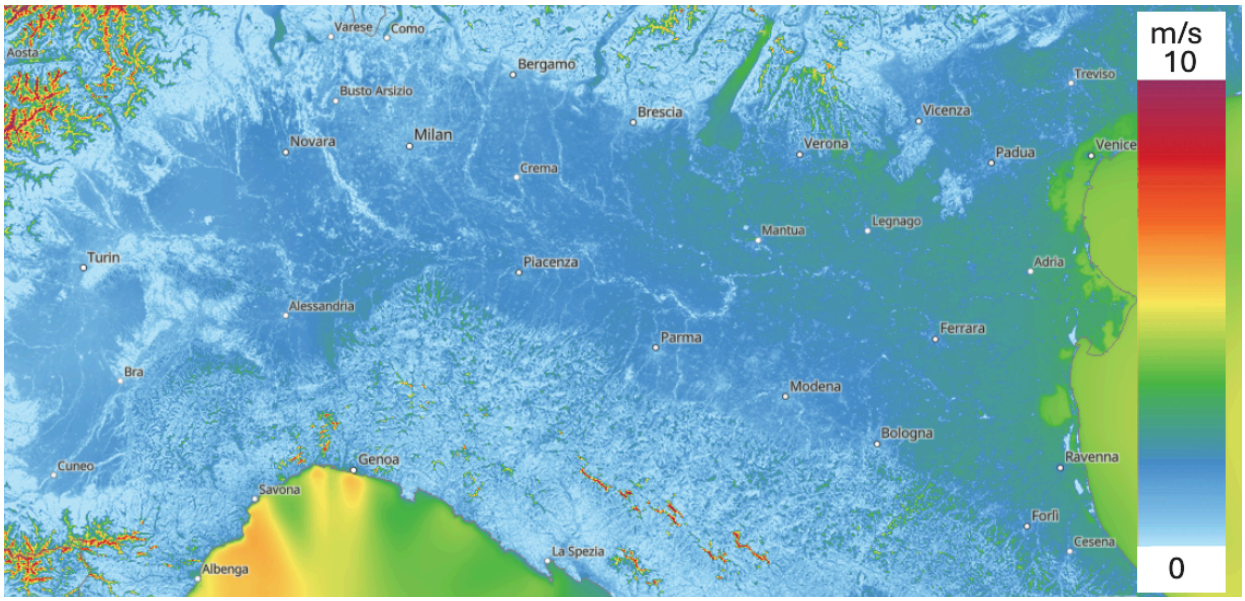


Figure 11: Mean wind speed at 10 m of height from https://globalwindatlas.info/en/

Furthermore the area is characterized by moderate precipitations (approximately 700 - 900 mm mean annual precipitation), therefore the contribution of wet deposition appears to be limited, particularly during winter, when the precipitation level is lower, as shown in Figure 12 and Figure 13.

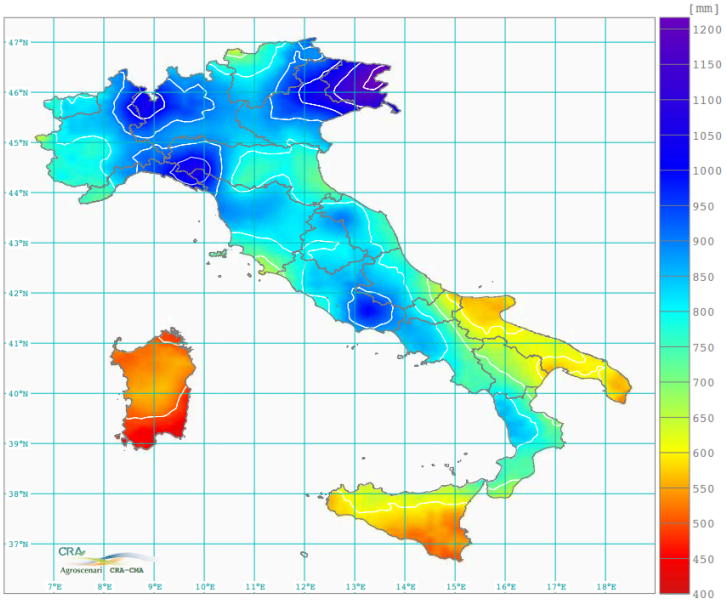


Figure 12: Mean annual precipitation (1981 - 2010) - Atlante italiano del clima e dei cambiamenti climatici

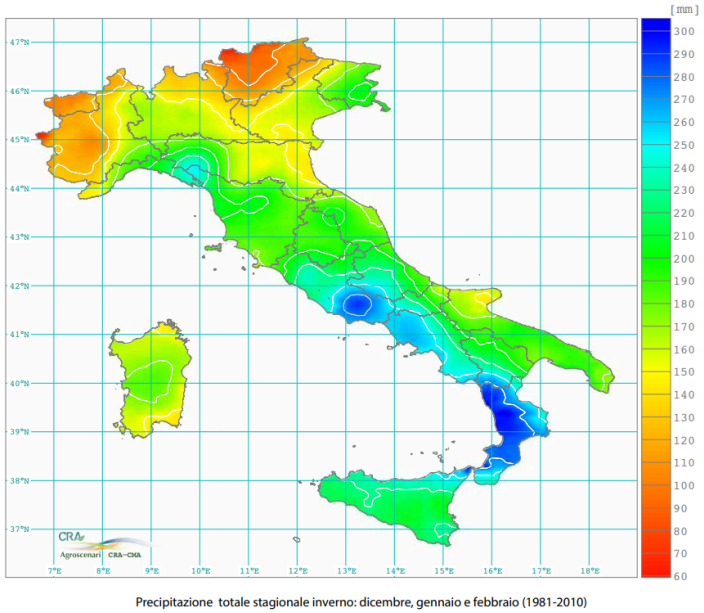


Figure 13: Mean winter precipitation (1981 - 2010) - Atlante italiano del clima e dei cambiamenti climatici

In order to fully explain the persistence of pollution episodes, the vertical structure of the lower atmosphere should be considered; in particular the behavior of the Planetary Boundary Layer (PBL), which plays a decisive role in stagnation episodes.

The PBL is the lowest part of the troposphere directly influenced by surface processes such as heating, cooling and friction. Within it, the Mixing layer (ML) defines the height available for pollution dispersion, consequently the volume available for pollution dilution.

During daytime, the ground surface heated by sunlight enhances turbulence and the PBL expands, leading to a higher pollutant dispersion, while at night, surface cooling stabilizes the atmosphere, the mixing layer disappear, followed by the formation of stable layers near the ground.(Figure 14) (Julaha et al., 2024)

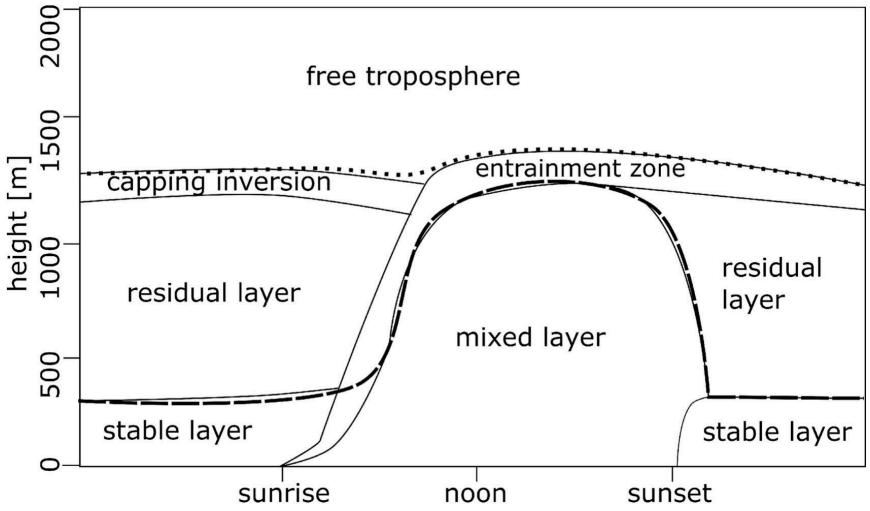


Figure 14: typical PBL structure across the day (Julaha et al., 2024)

In this context, it is necessary to take into account thermal inversions, layers in which temperature increases with height, acting as a cap suppressing vertical mixing. The mixing layer is usually capped by a thermal inversion which prevents further upward mixing.

In the Po Valley, winter anticyclonic conditions frequently lead to persistent thermal inversions and shallow boundary and mixing layers. During these periods, the PBL rarely exceeds approximately 400-500 m. Consequently primary pollutants are trapped near the surface, and the confinement of precursor gases enhances secondary aerosols formation processes.

The combined effect of low wintertime PBL, frequent and persistent thermal inversions, weak horizontal ventilation and limited wet deposition create a low dispersion system, and that's explain why this area becomes easily one of the most polluted in Europe, especially during wintertime.

This critical scenario should also be viewed in the context of climate change. Rising temperatures may increase the frequency of thermal inversions and reduce precipitations, leading to even more limited atmospheric dispersion. Thus if emissions do not decrease in the coming years, air quality is likely to further deteriorate (Caserini et al., 2017).

2.3. PrepAir: project targets and monitoring network and systems

The data analyzed in the following chapter were collected by the monitoring network established under the European project PrepAir within the framework of the program LIFE. The sampling campaigns were carried out as part of the project activities, therefore no sampling activities were conducted within this thesis work, which is based on the analysis of the dataset collected during the PrepAIR monitoring campaigns. The main objective of the project was to implement tools and coordinated actions to improve air quality in the Po Valley, being, as said in the introduction, one of the most critical in Europe from this point of view.

The project ran from February 1st 2017 to January 31 2024, and during this period the main lines of actions were the assessment of air quality variations, the environmental effect monitoring, sector specific mitigation actions and the most relevant for this thesis, the establishment of a dedicated monitoring network focused on the chemical characterization of PM10. (LIFE PREPAIR Project Consortium - Action D, 2023)

The monitoring network was designed to represent different environments, with the so-called "special stations" distributed in different types of sites:

- Urban background sites:
- **Torino**: the station is located in a public park in the southern part of the city. It was chosen since it possesses a wide range of measuring instruments suitable for source apportionment studies.
- **Milano**: the station is located in the eastern part of the city within the "Citta Studi" university area. It is located in a playground approximately 130 meters from a major road.
- **Vicenza**: the station is located in a residential area inside a small park, in the north east part of the city. It is situated downwind of the city center and quite close to an industrial/commercial area.
- **Bologna**: the station is located in the north-western part of the city, between the city center, the highway and the airport.
- Rural background site:

- **Schivenoglia**: located in the province of Mantova, it is far from major direct emission sources.
- Suburban and Alpine sites: Cavallermaggiore and Aosta (later added, that have not been considered for this analysis)

(LIFE Integrated Project PREPAIR, 2023)

Figure 15: Location of the special stations: urban and rural sites

The chemical species analyzed within the PrepAIR monitoring network include a wide range of elements, inorganic ions and carbonaceous components:

- Elements (Atomic Number Z > 11): Aluminum (Al), Silicon (Si), Phosphorus (P), Sulfur (S), Chlorine (Cl), Potassium (K), Calcium (Ca), Titanium (Ti), Vanadium (V), Chromium (Cr), Manganese (Mn), Iron (Fe), Nickel (Ni), Copper (Cu), Zinc (Zn), Bromine (Br), Rubidium (Rb), and Lead (Pb)
- Cations: Sodium (Na^+), Ammonium (NH_4^+), Potassium (K^+), Magnesium (Mg^{2+}), and Calcium (Ca^{2+}).
- Anions: Chloride (Cl^-), Nitrite (NO_2^-), Bromide (Br^-), Nitrate (NO_3^-), Phosphate (PO_4^{3-}), and Sulfate (SO_4^{2-}).
- Carbonaceous Compounds: Organic Carbon (OC) and Elemental Carbon (EC) (soot). Their sum is referred to as Total Carbon (TC).
- Sugars: levoglucosan, which serves as a unique tracer for biomass (wood) combustion.

2.3.1. PM10 sampling and sample analysis

To collect PM_{10} sample, every site was equipped with two low volume samplers, operating at 2.3 m^3/h collecting the particles on 47 mm diameter filters using two different types of membranes : quartz fiber filters, used for carbonaceous components, ions and sugars (as levoglucosan) and mixed cellulose esters or teflon filters, dedicated to the determination of heavier elements (Z > 11). (LIFE Integrated Project PREPAIR, 2023)

To chemically characterize the PM samples, different analytic techniques were used, depending on the target species.

- TOT/TOR "Thermo-Optical Analysis,was applied for the distinction between Organic Carbon (OC) and Elemental Carbon (EC); it consists in heating the sample in two distinct phases and the carbon released in each phase it's converted in methane and measured by a flame ionization detector. This analysis produces a thermo-gram in which the split point between OC and EC is determined based on optical correction. (Khan et al., 2011)
- X-Ray Fluorescence (XRF), was used to identify and quantify the heavier elements on the Teflon or mixed cellulose filters.
- Ion chromatography, was employed to measure anions, as nitrates and sulfates, cations, as ammonium, sodium and potassium, and levoglucosan. The PM sample is dissolved in water and then injected into a Methronm 881 Compact IC pro, where the components flow with different velocities, enabling the detection for each of them, with conductivity detectors for ions and a pulsed amperometric detector for levoglucosan. (LIFE Integrated Project PREPAIR, 2023)

Chapter 3

Data Analysis

The analysis was performed using Matlab, considering the data that go from the 28th of April 2018 until the 31st of December 2019; The only exception is Torino, since the measurements started later that year (September). The sections of total PM concentration, Chemical composition and Urban and Rural background comparison are related to all the five stations, while the meteorological correlations were considered only for the Torino site.

3.1. Total PM concentration

The first studied aspect was the trend of total PM_{10} for every station, underlying the limits (current and future) and the number of exceedances.

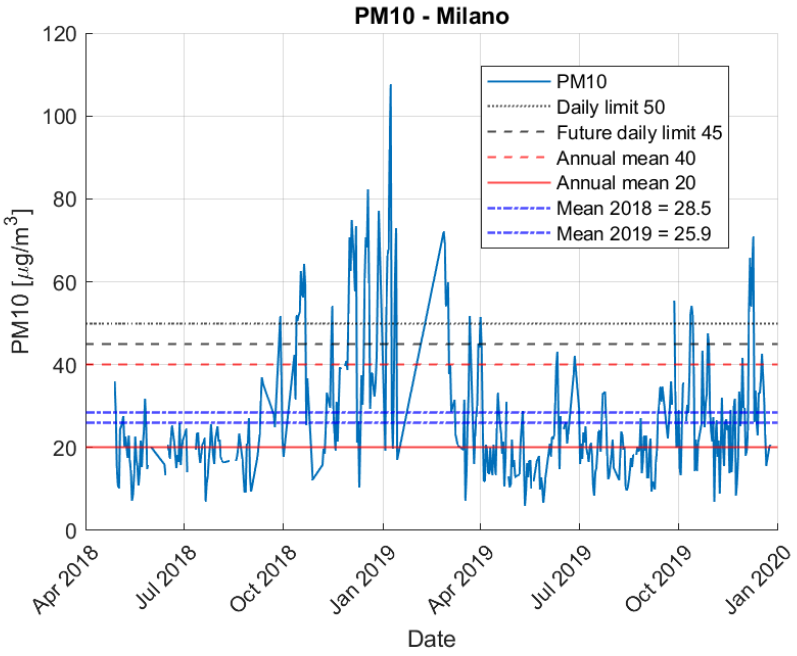


Figure 16: Total PM_{10} trend and limits - Milano station

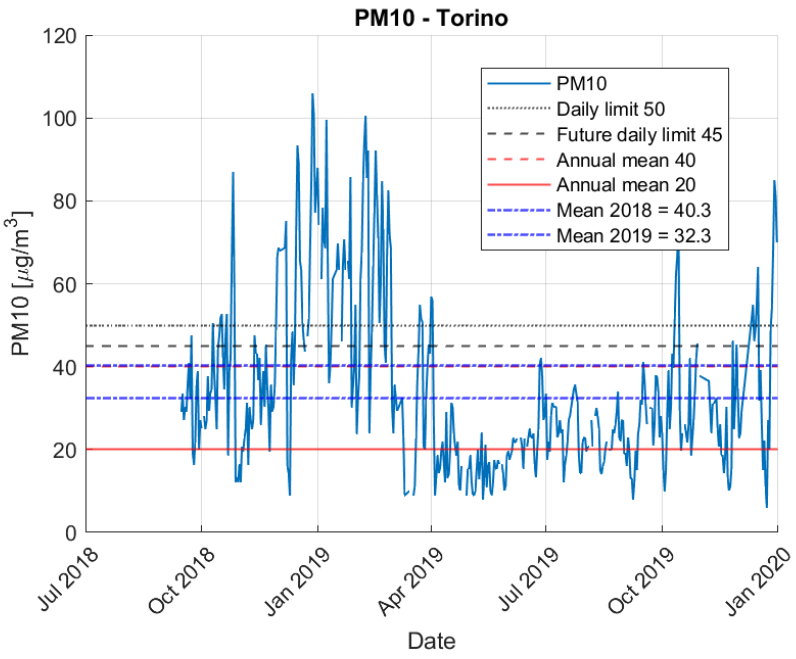


Figure 17: Total PM_{10} trend and limits - Torino station

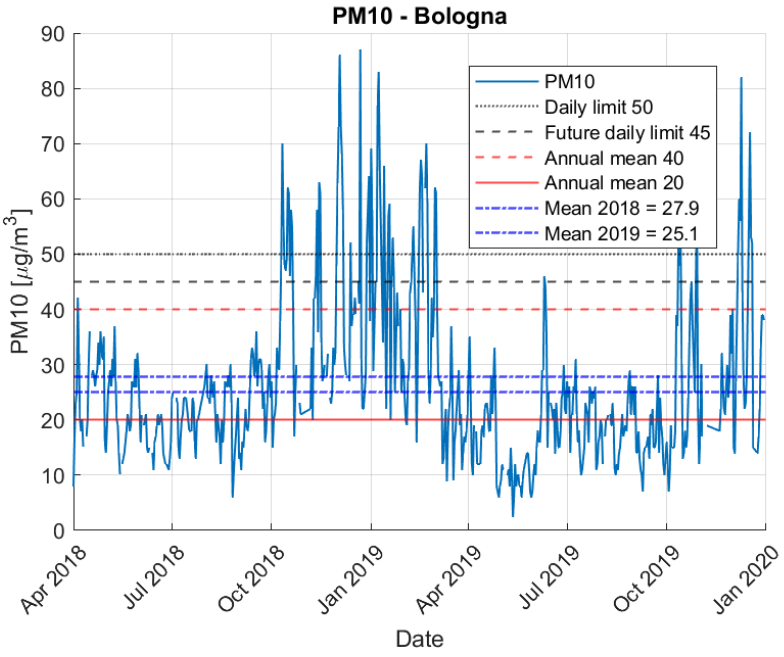


Figure 18: Total PM_{10} trend and limits - Bologna station

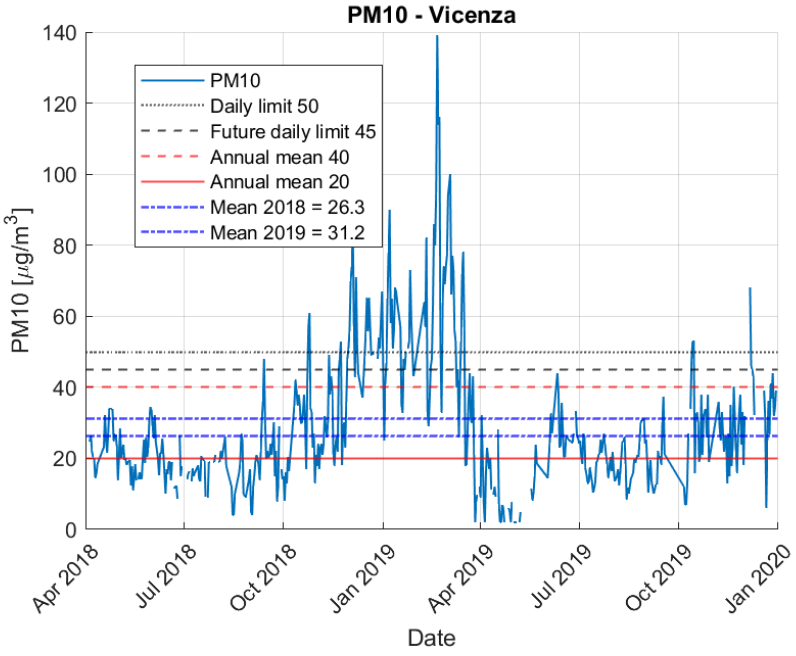


Figure 19: Total PM_{10} trend and limits - Vicenza station

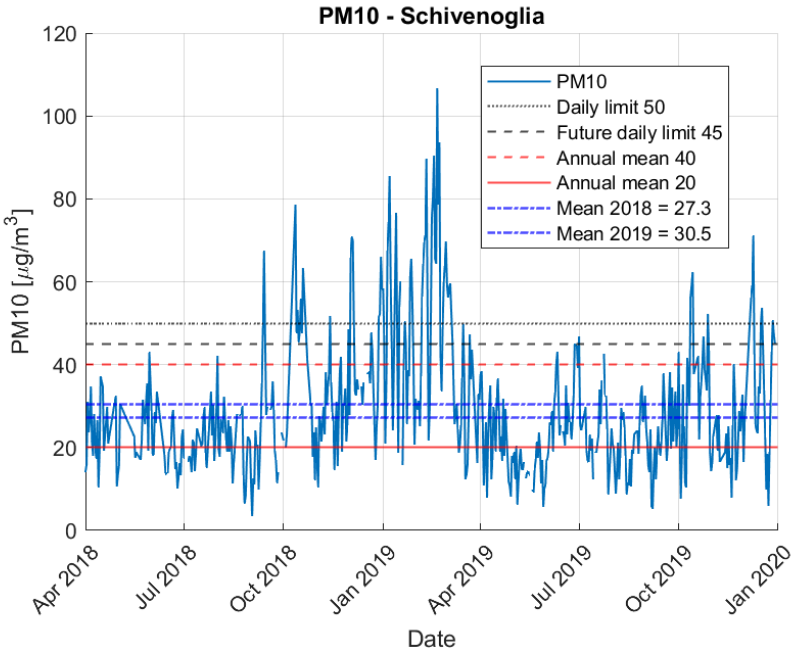


Figure 20: Total PM_{10} trend and limits - Schivenoglia station

	2019 Old Directive	2019 New Directive
Milano	23	26
Torino	58	70
Bologna	29	36
Vicenza	48	55
Schivenoglia	43	54

Table 3: Number of daily exceedances in 2019 under the old (50 μg/m³ - 35 exc.) and new (45 μg/m³ - 18 exc.) EU Directive limits.

All the sites respect the existing limits in terms of annual mean, except for the site of Torino during 2018 (but it is important to underline that the result could be biased by the fact that only data from September are available); while none complies with the one introduced by the new directive (*Direttiva (UE) 2024/2881 del Parlamento europeo e del Consiglio*, 2024). Regarding the permitted annual (2019) exceedances of the daily limits, just Bologna and Milano respect the existing legislation while none of the sites complies with the new ones.

These results were compared to studies that have analyzed the trend of PM10 for more years. ARPA Lombardia has published a report showing the mean annual concentrations and the number of exceedances (Figure 21 and Figure 22).(ARPA Lombardia, 2025) These result present some differences between the ones calculated previously (higher annual average and number of exceedances) but the main reason for such discrepancy may be the number of station considered, in fact ARPA Lombardia report mean and exceedances based on data coming from different stations across the city.(ARPA Lombardia, 2025). The most relevant aspect of this report is that both the annual mean and number of exceedances are decreasing over the years, despite being still far from the limits that will be imposed from 2030 by the EU.

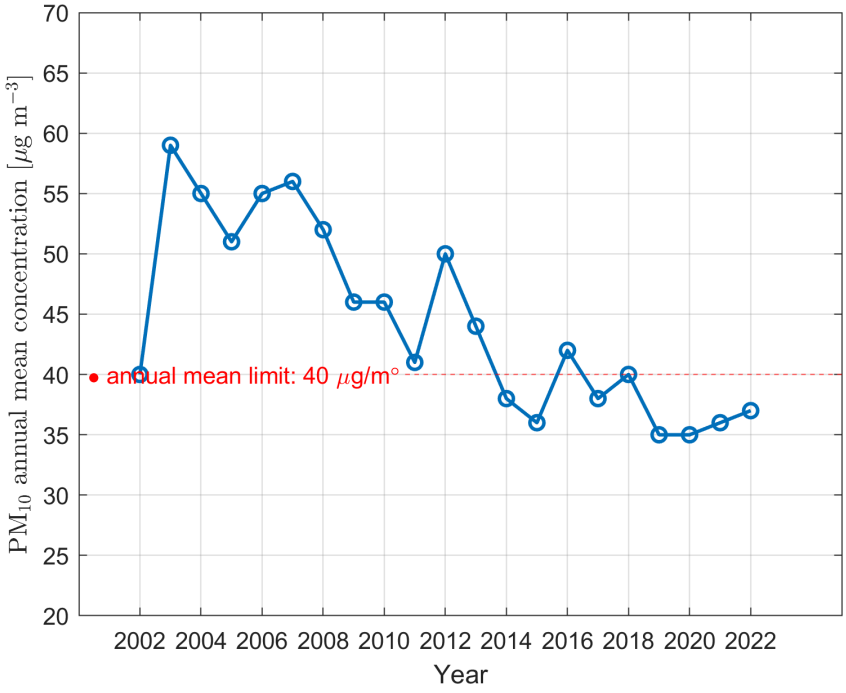


Figure 21: PM_{10} trend in Milano - plot from ARPA Lombardia data

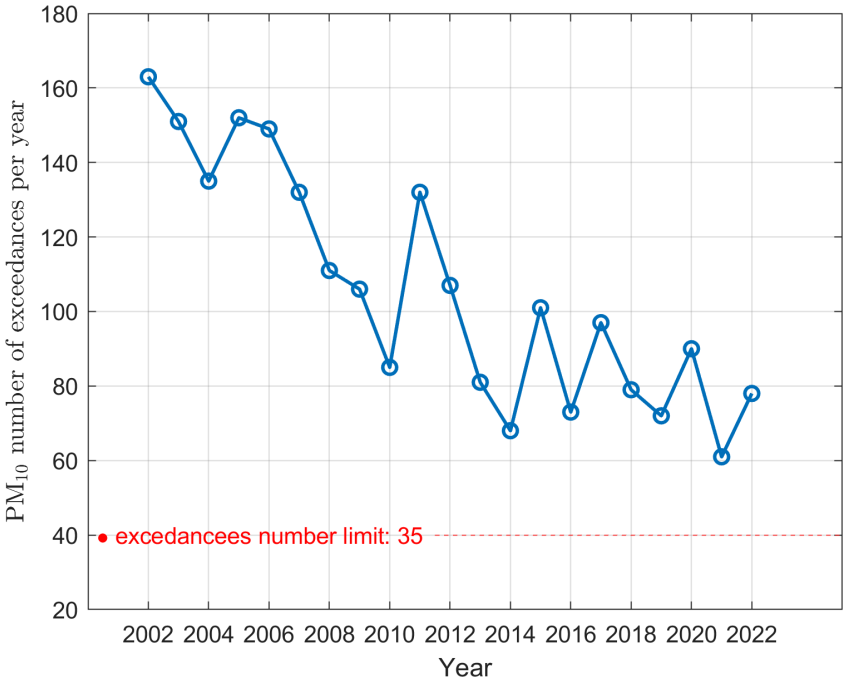


Figure 22: PM_{10} exceedances trend in Milano - plot from ARPA Lombardia data

This comparison and analysis has been done also with data reported by ARPA Piemonte in "Uno sguardo all'aria 2023" (the latest one available). Also in this case the report included various stations across the metropolitan area and the province of Torino. (ARPA Piemonte et al., 2024)

STAZIONE	Rendimento strumentale. 2023 (% giorni validi)	PM10 – VALORE MEDIO ANNUO Valore limite annuale: 40 µg/m³										PM10- NUMERO DI SUPERAMENTI del valore limite di 24 ore (50 µg/m³)									
		2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
Baldissero (B)	94%	14*	17	14	11	8	18*	22	20*	18	**16**	6*	8	1	1	0	5*	22	15*	9	**0**
Beinasco (B)	96%	30	33	29	36	28	27	28	27	29	**25**	47	68	52	88	41	49	59	39	36	**21**
Borgaro (B)	97%	-	-	-	-	-	-	-	27	31	**26**	-	-	-	-	-	-	-	33	39	**25**
Borgaro	-	31*	35	31	38	30	26	30	28	-	**-**	44*	71	54	90	42	28	51	37	-	**-**
Carmagnola	85%	36	41	37	45	36	35	36	33*	37	**31**	82	107	73	122	69	69	81	49*	82	**39**
Ceresole Reale (B)	84%	5	7	9	11	11*	9*	10*	10	11*	**10***	0	0	0	6	0*	1*	1*	3	0*	**0***
Chieri (B)	93%	-	-	-	-	-	-	28	28	31	**27**	-	-	-	-	-	-	55	40	48	**33**
Collegno	94%	32	36	32	40	33	30	35	32	36	**28**	61	81	61	102	56	50	85	57	74	**35**
Druento	92%	19	23	21	27	22	19	21*	21	21	**19**	11	23	22	41	15	10	11*	7	10	**6**
Ivrea - Liberazione (B)	99%	-	-	-	-	-	-	25	23	26	**22**	-	-	-	-	-	-	48	33	41	**22**
Ivrea - Liberazione	-	23	28	26	31*	25	24*	27	-	-	**-**	30	55	41	60*	28	29*	43	-	-	**-**
Leini (B)	94%	25	36	30	34	33*	26	29	26	27	**23**	35	84	57	79	28*	44	59	38	33	**17**
Oulx	89%	17	18*	16	18	18	15	16	15	19	**17**	5	7*	0	8	1	0	1	0	1	**0**
Pinerolo	92%	-	21	23	26	21	19	18	17	19	**19**	-	11	15	40	11	5	12	10	9	**4**
Settimo (B)	82%	-	-	-	-	-	-	-	31	35	**33***	-	-	-	-	-	-	-	55	77	**55***
Settimo	-	34	39	35	44*	36	34	35	34	-	**-**	81	98	70	99*	65	63	83	64	-	**-**
Susa	87%	16	18	17	22	18	15	16	17	18	**16**	1	11	10	27	6	1	3	5	4	**0**
To - Consolata	88%	35	40	35	43	33*	28	36*	28*	33	**32**	75	93	75	108	55	45	84*	30*	57	**53**
To - Grassi	89%	43*	52*	42	47	40*	38	41	36	40	**35**	77*	75*	89	112	76*	83	98	75	98	**66**
To - Lingotto (B)	88%	31	36	32	38	31	28	31	29	34	**27**	56	85	60	101	45	50	72	46	66*	**35**
To - Lingotto	84%	32	38	34	39	28	27	30	26	28	**24***	59	86	62	92	39	48	67	40	37	**19***
To - Rebaudengo (B)	95%	40	43	37	46	39	34	36	33	37	**33**	94	101	74	118	87	71	88	65	86	**63**
To - Rubino (B)	95%	32	36*	30*	37	30*	28	28	26*	31	**24**	55	53*	46*	91	36*	46	53	26*	47	**26**
To - Rubino	93%	31	36	32	38	29	28*	32	30	32	**27**	58	84	65	97	33	42*	66	57	58	**37**

(*) Rendimento strumentale inferiore all'85%

Figure 23: Total PM_{10} trend and exceedances - stations across the city and province of Torino - (ARPA Piemonte et al., 2024)

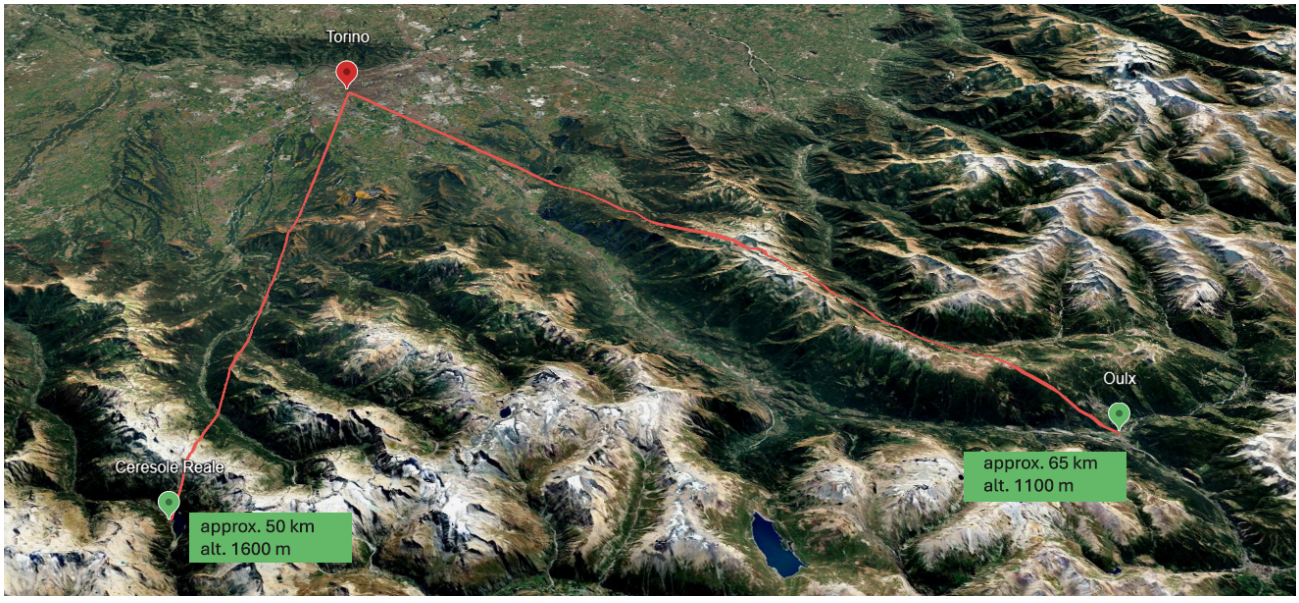


Figure 24: Position of Ceresole Reale and Oulx, compared to Torino - Google Earth

In this case, the most relevant aspect is the difference in values between the data reported by stations located within the city and the ones resulting by stations located in more remote zones, as the stations of Ceresole Reale and Oulx. They have the lowest values both for mean concentration and exceedances. This represents strong evidence that geographical location play a decisive role in the observed concentrations, even at relatively small spatial scales. As for the example of Milano, the trends of mean annual concentration and annual number of exceedances are decreasing but the pace of this trend is again too slow to meet the new directive targets.

3.2. Chemical composition

3.2.1. Average composition by site

The first aspect studied regarding chemical composition was the average percentage of all the components, for all the five stations.

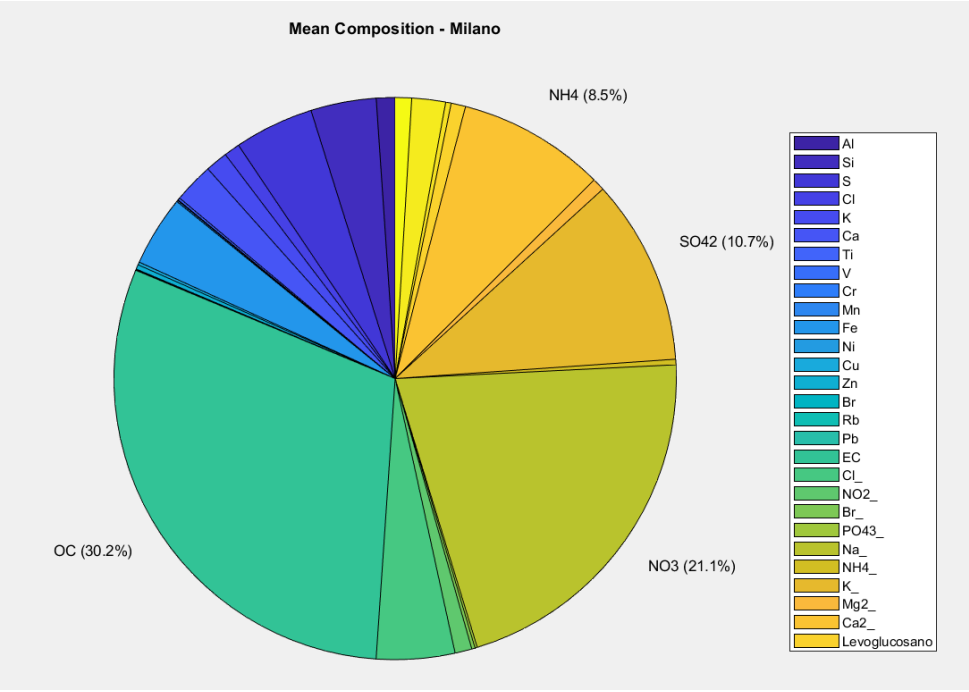


Figure 25: Average chemical composition - Milano

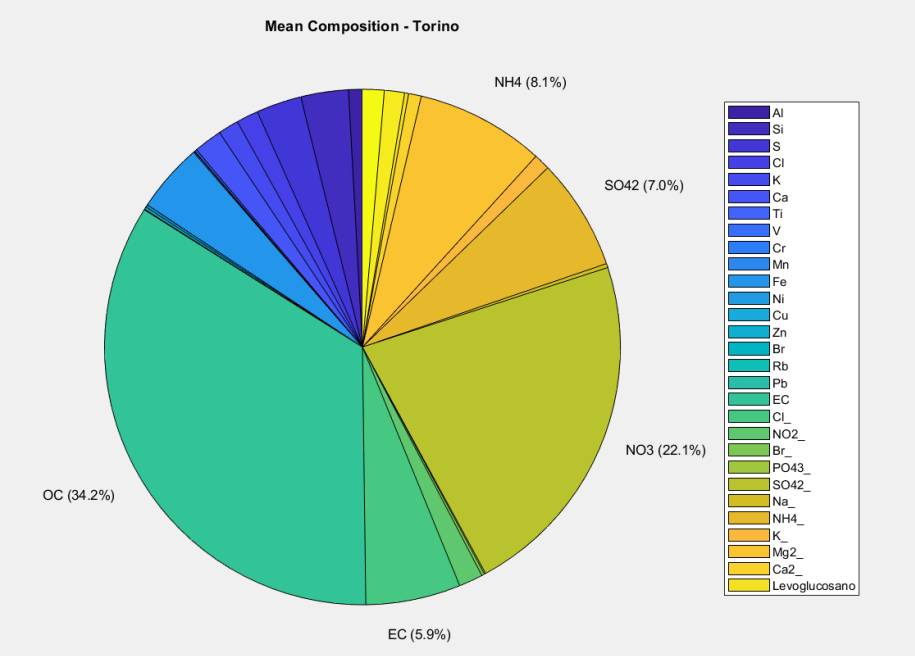


Figure 26: Average chemical composition - Torino

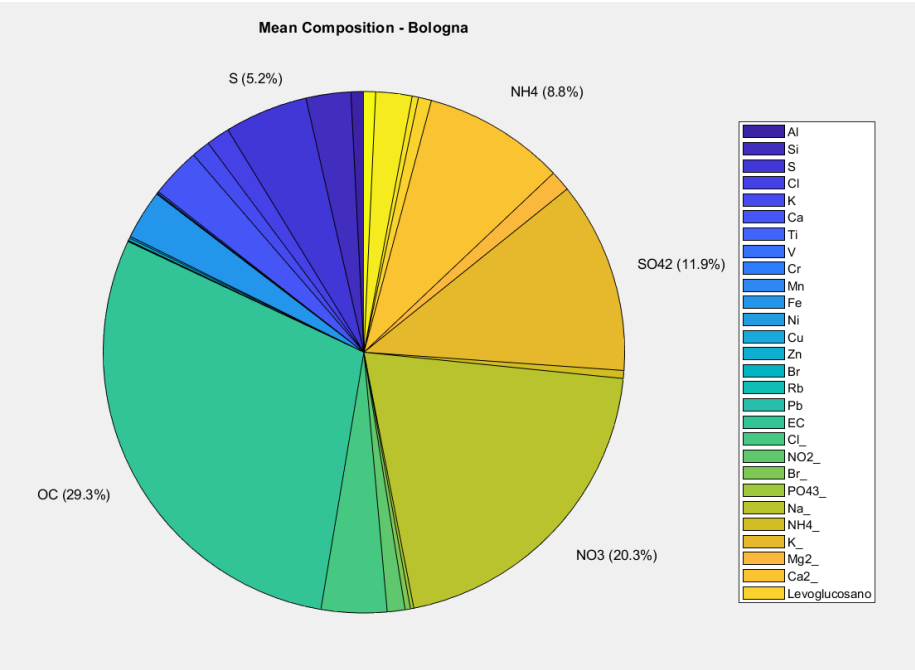


Figure 27: Average chemical composition - Bologna

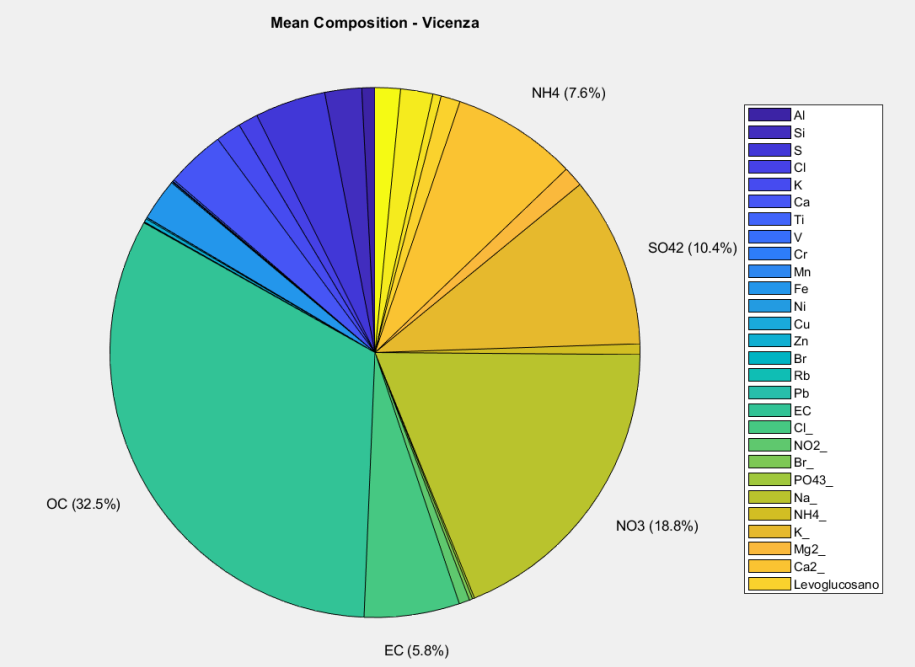


Figure 28: Average chemical composition - Vicenza

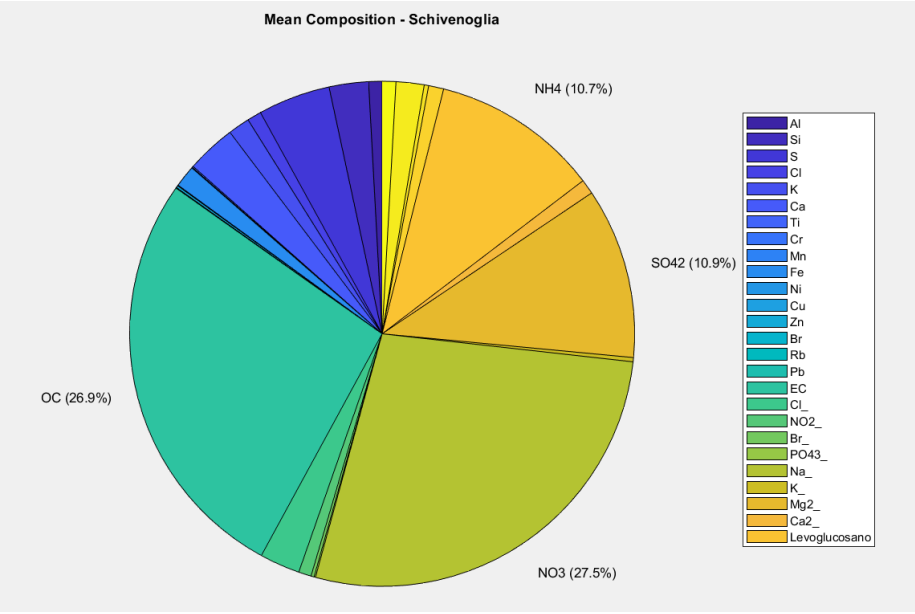


Figure 29: Average chemical composition - Schivenoglia

In all the sites the main components are: Organic Carbon OC, Nitrates NO_3^-, Sulfates SO_4^{2-}, and Ammonium NH_4^+, as expected from literature review.

The percentages are similar except for Schivenoglia, that shows a lower value for OC and a higher one of Nitrates. This is explainable by the fact that nitrates are mainly associated with secondary particulate matter formation, while OC both of primary and secondary and secondary PM is more likely to be found in rural sites where direct emission are lower compared to a city background.

3.2.2. Mass closure analysis

In order to have a more complete view of the chemical composition of PM_{10}, a mass chemical closure analysis was performed. The reconstructed mass (RM) was calculated by grouping the measured species into major chemical classes mainly used in aerosols studies: organic matter (OM), elemental carbon (EC), secondary inorganic aerosols (SIA), mineral matter (MIN), sea salt (SS) and trace elements (TE); crustal elements were converted to their oxide forms to estimate mineral matter, while organic matter was calculated by using a conversion factor of 1.6 to organic carbon. Secondary inorganic aerosol was obtained by the sum of nitrates, ammonium and non-sea salt sulfates (Terzi et al., 2010). The reconstructed mass was then compared to the total value of PM_{10}.

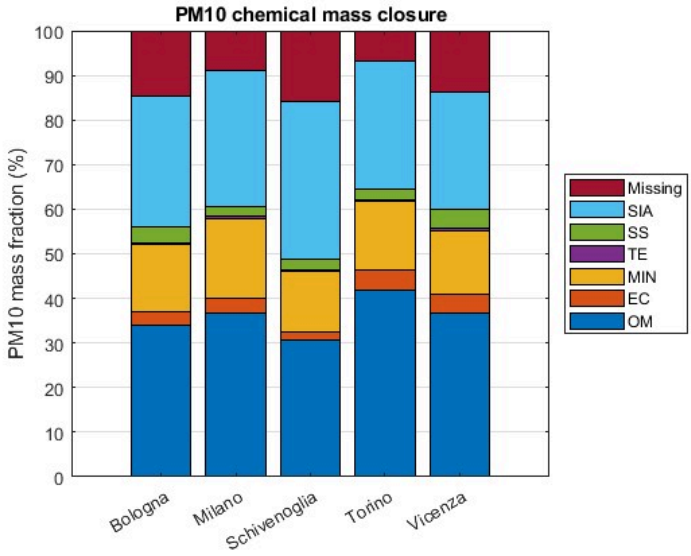


Figure 30: Mass closure analysis plot for every site

The result shows that the reconstructed mass explain the majority of the total PM_{10}, with closure values ranging from 84% to 93%. The remaining missing mass can be attributed to various factors, mainly the presence of particle-bound water retained on the filter, the uncertainties of the OC - OM conversion factor and the presence of unmeasured organic compounds. (Terzi et al., 2010)

3.2.3. Seasonal changes

The result of Bologna for the year 2019 has been divided in spring-summer and fall-winter in order to be compared with the study made by Matta et al (2002) regarding the site of Bologna but referred to the year 2000.

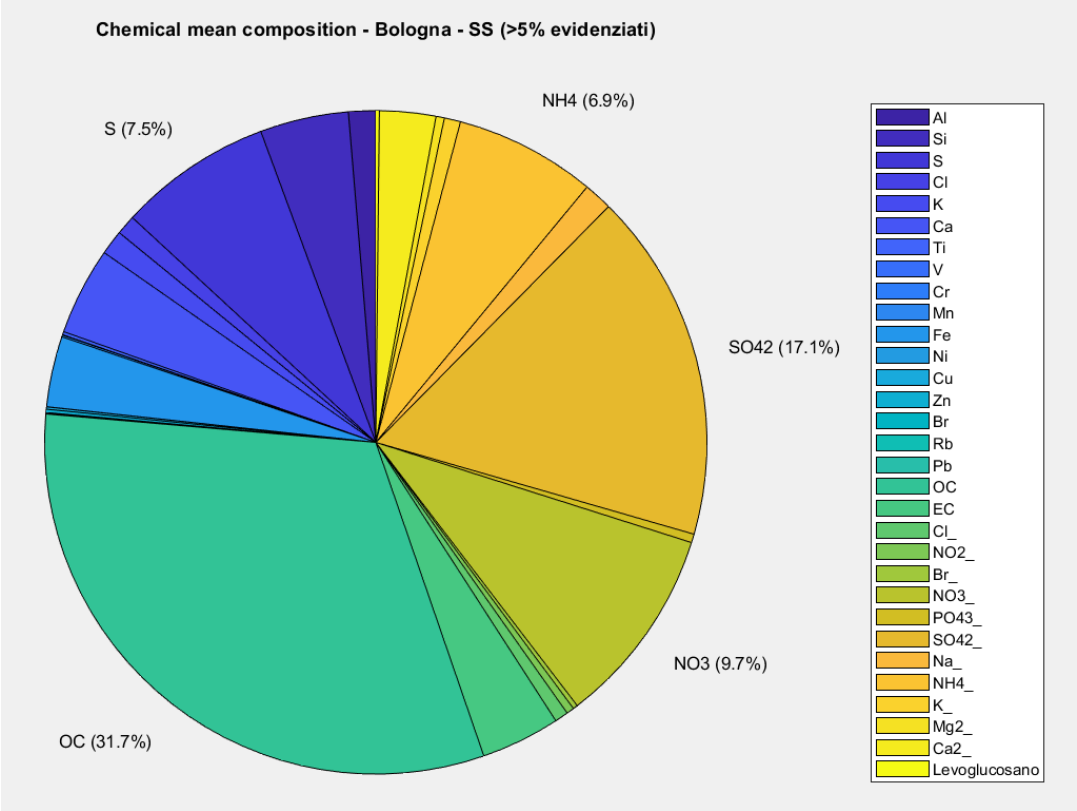


Figure 31: Average chemical composition spring-summer 2019 Bologna

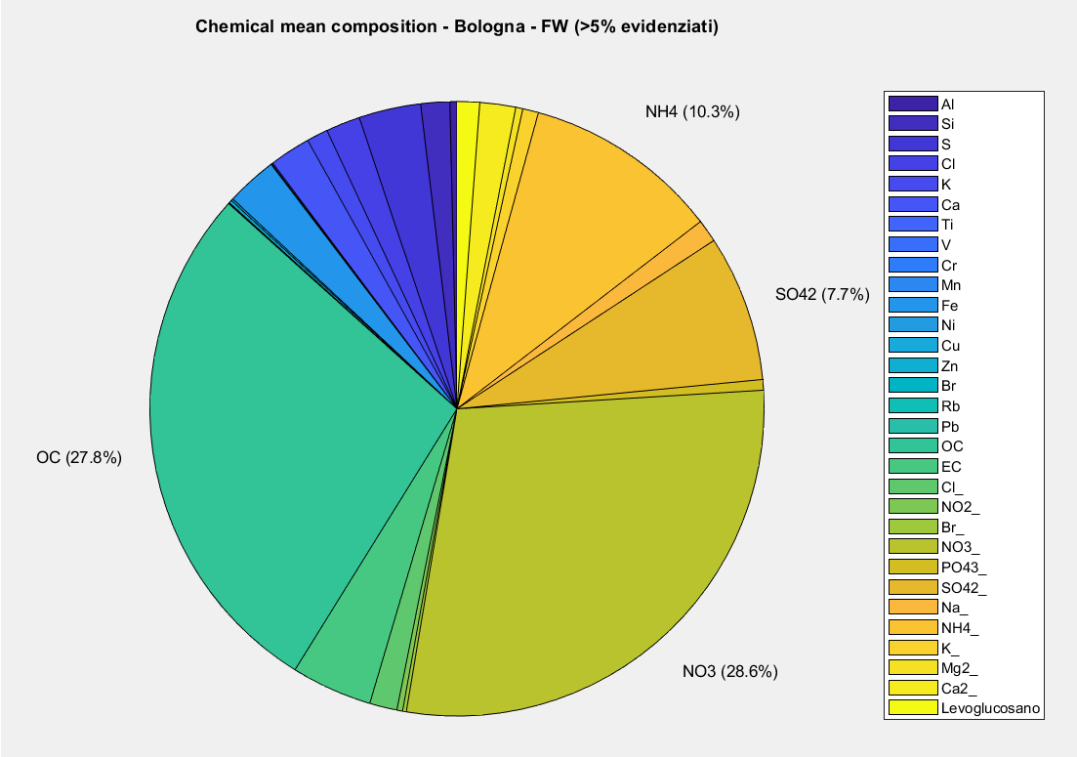


Figure 32: Average chemical composition fall winter 2019 Bologna

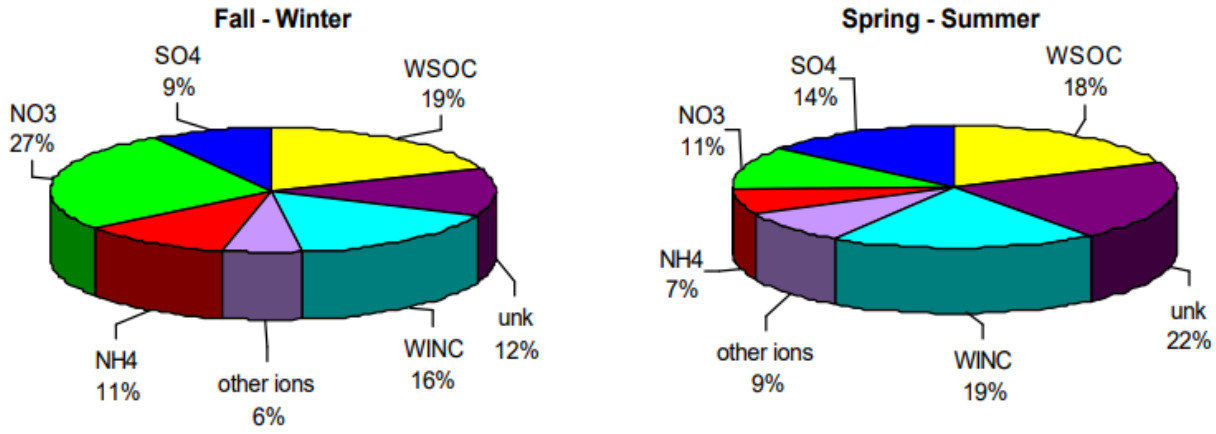


Figure 33: Average chemical composition spring-summer fall winter 2000 Bologna with the contribution of nitrates, solfates, ammonium, other ions, water soluble and insoluble organic species (WSOC and WINC). (Matta et al., 2002)

The percentages are very similar for both the periods they have been analyzed, (considering the sum of WSOC and WINC as the OC in the PrepAIR dataset). This show that the emissive scenario in Bologna did not undergo significant changes, in terms of sources.

The seasonal changes in the percentages of nitrates, sulfates and ammonium are reasonably caused by the seasonality of weather, in fact the concentration of both nitrates and sulfates are the result of a the equilibrium between gaseous and solid/liquid phases. Nitrates are promoted by higher humidity and lower temperatures, while sulfates are enhanced by photochemical reactions, consequently they present higher percentages in periods with a higher solar exposure (Seinfeld & Pandis, 2006).

3.3. Urban and Rural background comparison

One of the most interesting difference presented by the chosen sites is their background; In order to see how it can influence the concentration and composition of PM, the data from Schivenoglia have been compared to the data of the others sites. Every comparison has a different specific focus: carbonaceous components, levoglucosan, secondary inorganic compounds and metal traces.

3.3.1. Carbonaceous Components

As said in first chapters, carbonaceous material constitutes a major fraction of PM_{10}. It can be divided into Organic Carbon (OC) and Elemental Carbon (EC).

EC is a primary pollutant, since it's directly emitted from incomplete combustion processes, and the main sources are traffic and biomass combustion. While OC is a mixture of compounds, some primary, from traffic and biomass burning and some secondary, resulting from chemical reactions (gas to particle conversion) in the atmosphere. A large part of the OC detected, usually, comes from long range transport (aged aerosol or particles originated in distant regions, an example of source could be far wildfire). (Saarikoski et al., 2008)

Trend in Vicenza and Schivenoglia were studied to see the seasonality of both EC and OC (Figure 34, Figure 35). In both the cases, there is a clear seasonality, expected because of the use of biomass based heating systems used during the winter. The seasonality seems more marked for the EC trend, an hypothesis could be that since OC is made of both primary and secondary particles, even when direct emissions are lower, the concentration is still sustained by the secondary component.

An interesting way of studying the differences in sources is to study the OC/EC ratio. Every source has a typical value range:

- traffic ~ 0.71
- secondary organic aerosol ~ 3.3
- Biomass combustion ~ 6.6
- Long term transport ~ 12

The differences in ratios between Vicenza and Schivenoglia are showed in Figure 36 and Figure 37

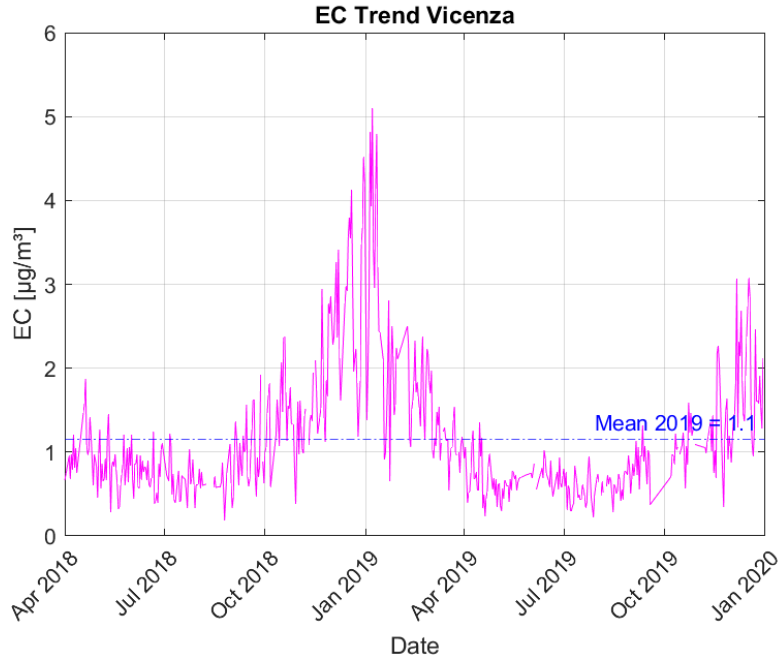


Figure 34: EC trend in Vicenza and average annual values

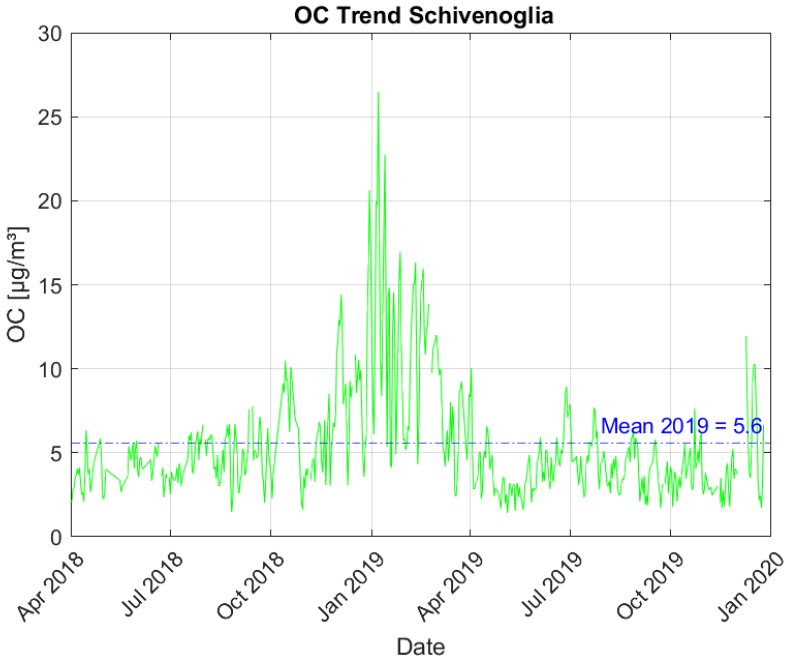


Figure 35: OC trend in Schivenoglia and average annual values

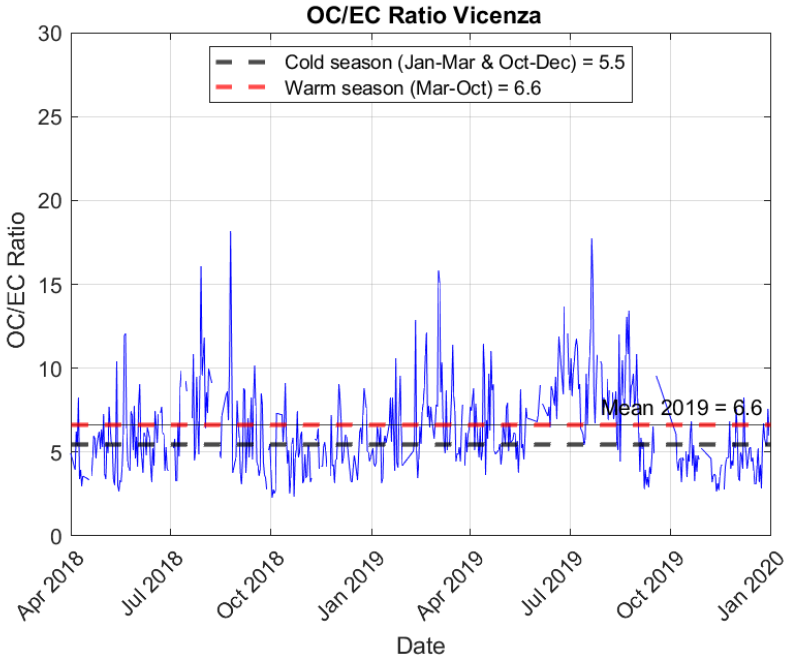


Figure 36: OC/EC trend in Vicenza and average annual values

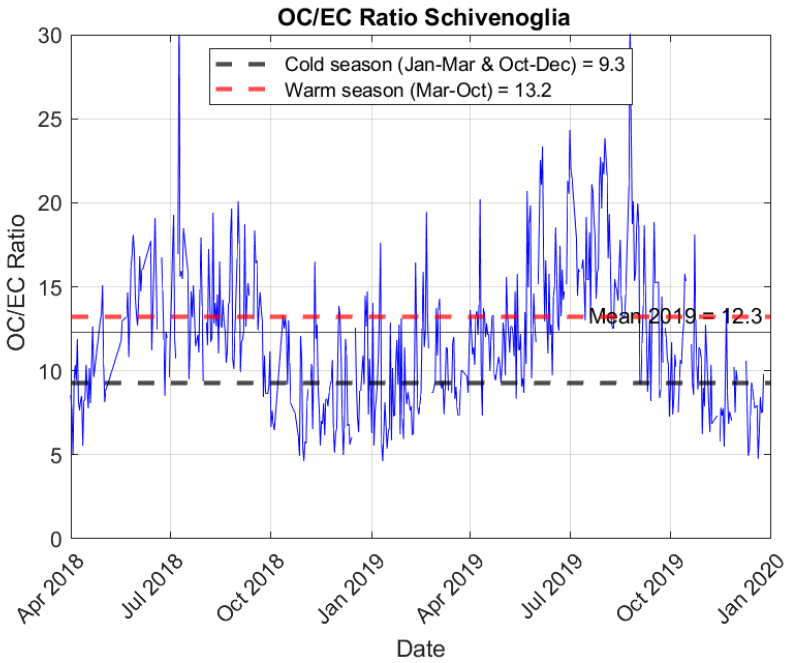


Figure 37: OC/EC trend in Schivenoglia and average annual values

The OC/EC ratio appears to be higher in Schivenoglia, explainable with a major contribution of secondary PM and long term transport. During the cold season in both the graphs, the ratio tends closer to the signature value of the ratio related to biomass burning (compared to the mean annual value) (Saarikoski et al., 2008).

3.3.2. Levoglucosan

Another key tracer of incomplete combustion it's levoglucosan, a sugar anhydride directly produced by cellulose combustion. Because of this peculiarity, it's considered a key signature for burning plant material, making it more specifically related to a source than EC or OC. In addition, compared to other wood smoke compounds, it is highly resistant to chemical degradation and consequently more stable during atmospheric transport. levoglucosan is associated with fine particles, in fact its concentration is usually the same in PM_{10} and $PM_{2.5}$. (Jordan et al., 2006; Simpson et al., 2004). Levoglucosan has been compared to Potassium (K) trend, since it's aswell related to combustion, but in contrast to levoglucosan, it is also related to soil dust, sea salt, traffic non-exhaust emissions (brake dust, tire wear and road dust), industrial emissions. (Pachon et al., 2013)

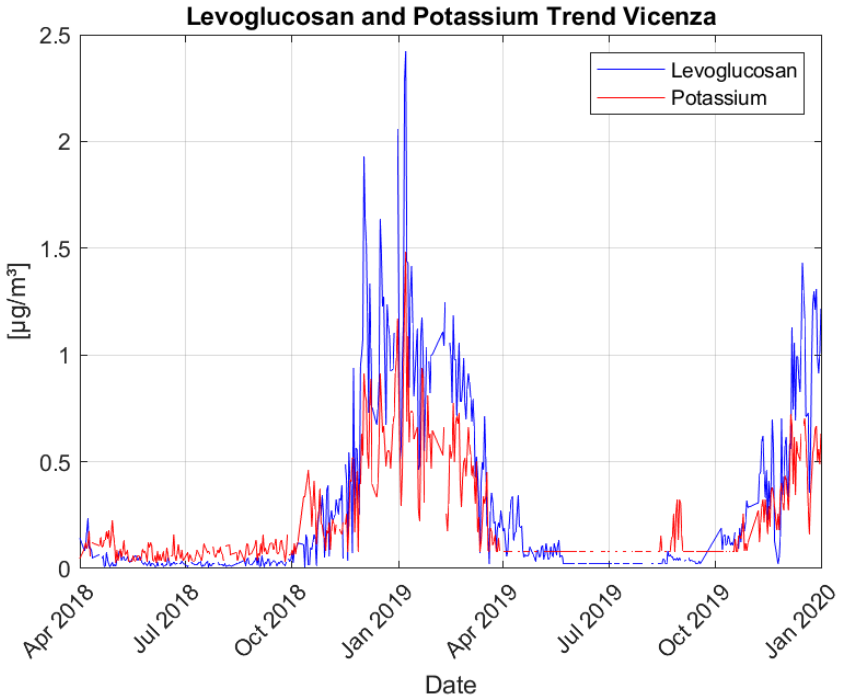


Figure 38: levoglucosan and potassium trend in Vicenza

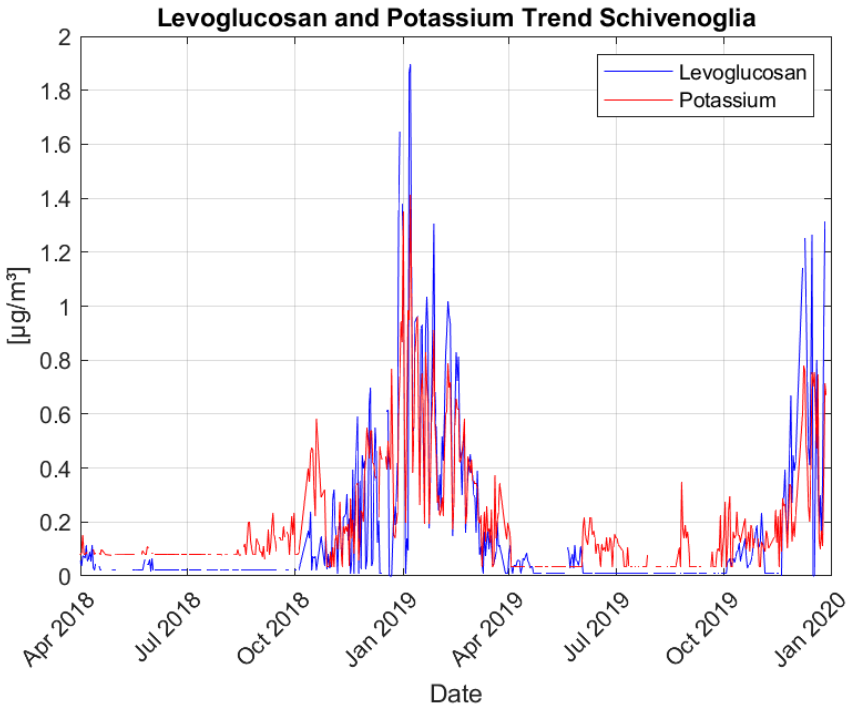


Figure 39: levoglucosan and potassium trend in Schivenoglia

As expected the peaks regard the cold season, but during these periods levoglucosan shows higher values than potassium; although potassium comes from many sources, it has a low factor of emission, estimated to be around 5.7 mg/g of PM (Pachon et al., 2013); levoglucosan, instead, has a factor of emission of 140mg/g of PM for biomass smoke (Simpson et al., 2004). The fact that both values of potassium and levoglucosan are generally higher in Vicenza, is likely related to a higher population.

3.3.3. Secondary Inorganic Compounds

As said in the introduction, Secondary inorganic compounds are usually a major fraction of PM (both 2.5 and 10). This fraction it's made of salts resulting from reactions of neutralization between the precursors (ammonia, nitrogen oxides and sulfur dioxide) (Figure 40).

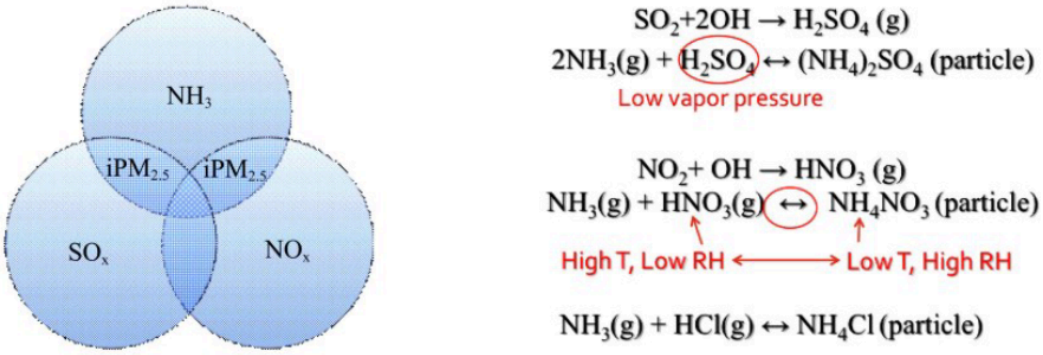


Figure 40: Neutralization reactions that forms inorganic $PM_{2.5}$ (Wang-Li, 2015)

Each of the precursors is emitted by different sources:

- Ammonia NH_3 is mainly related to agriculture and livestock
- Nitrogen oxides NO_x is usually emitted by traffic and heating
- Sulfur dioxide SO_2 derives principally from industrial activities as energy production (where fossil fuels are used) (LIFE Integrated Project PREPAIR, 2023)

It is already documented that the highest concentrations of the secondary PM usually are detected in sites that are distant from the sources since these reactions occur in timescales of hours and days and regional space scales. (Seinfeld & Pandis, 2006)

Usually in literature the inorganic secondary component is calculated by the sum of the following ions: nitrates, sulfates and ammonium (it is necessary to underline that sometimes a minor amount of ammonium chloride can be present). In this case, Milano was taken as example of urban background site. As expected the secondary component is higher in the rural site; in both the graphs (Figure 41 and Figure 42) the peaks are more likely related to nitrates, favored by low temperature and high relative humidity, while during the warm season, the main component of the secondary inorganic PM is plausibly made of sulfates.

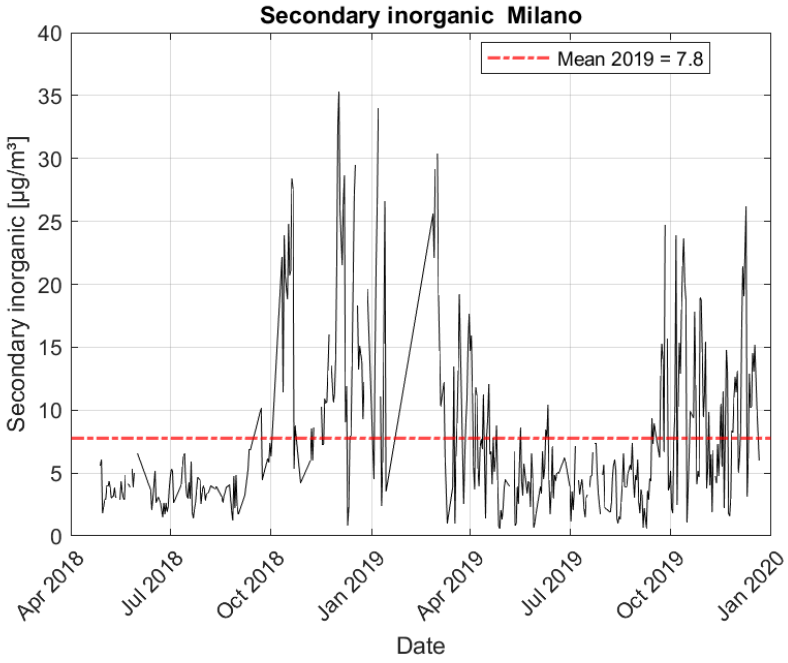


Figure 41: Secondary inorganic PM in Milano with the 2019 average annual value

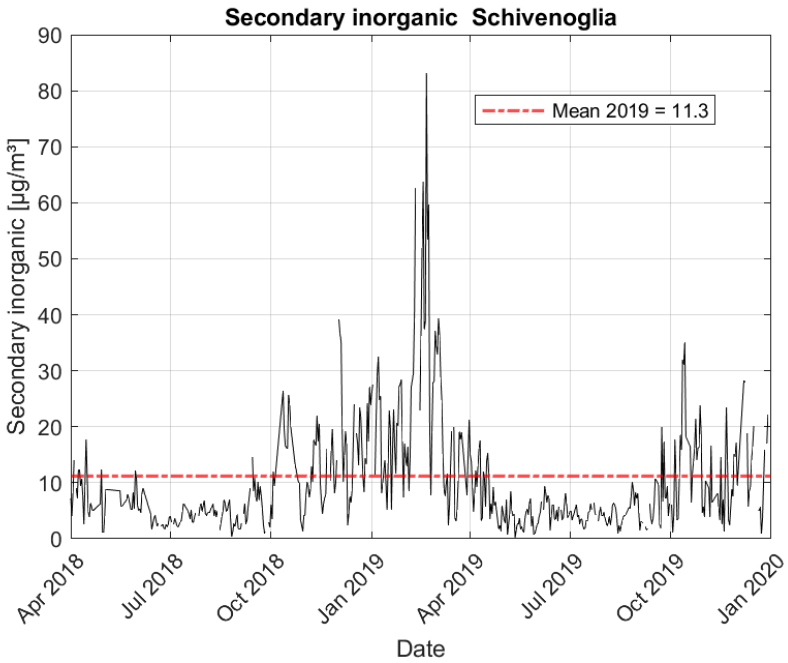


Figure 42: Secondary inorganic PM in Schivenoglia with the 2019 average annual value

3.3.4. Metals traces

Metal traces have been chosen as the last comparison between urban and rural background, since their high toxicological impact, in particular lead, arsenic, nickel and cadmium. (Guo et al., 2022). Since the PrepAir dataset included only Nickel and lead, cadmium and arsenic were taken from ARPA Piemonte and ARPA Lombardia, respectively for Torino and Schivenoglia (from the same stations). The annual mean for the two sites was compared with the legislative limits (Decreto Legislativo 13 agosto 2010, n. 155.), in order to verify their compliance Table 4.

Specie	Torino (2019)	Schivenoglia (2019)	Limite/Target	Unità
As	0.7077	0.5670	6.0000	ng/m³
Cd	0.1329	0.0850	5.0000	ng/m³
Ni	1.6340	2.5820	20.0000	ng/m³
Pb	0.0047	0.00553	0.5000	µg/m³

Table 4: Mean annual values of trace metals in Torino and Schivenoglia for the year 2019, compared with the limits

A part from the verified compliance, another interesting aspect considered was the Nickel value showed by the site of Schivenoglia, compared to the one of Torino's site, almost the double. Nickel main sources are, traffic (in general fossil fuels combustion), crustal material, naval emissions and industrial activities (in particular the one related to the production of steel). Given this context, an urban background would be the one expected with higher values of Nickel, but after a brief research of the locations of the steel production plants in Italy, it's more plausible for Schivenoglia to be more impacted by the relatively close industrial sources and consequently to present higher values of Nickel. (Figure 43)

Figure 43: Location of steel production plants in northern Italy, compared to the position of Schivenoglia - https://federacciai.it/

To confirm that industrial emissions are likely the ones to cause that value at the Schivenoglia site, the trend of copper (Cu) has been considered both for Torino and Schivenoglia, since it's one of the key tracers for vehicular traffic (It's directly emitted by brake wear and generally regards non-exhaust emissions)(LIFE PREPAIR Project Consortium - Action D, 2023). The comparison confirms that plausibly the higher value of Nickel in Schivenoglia is attributable to industrial activity.

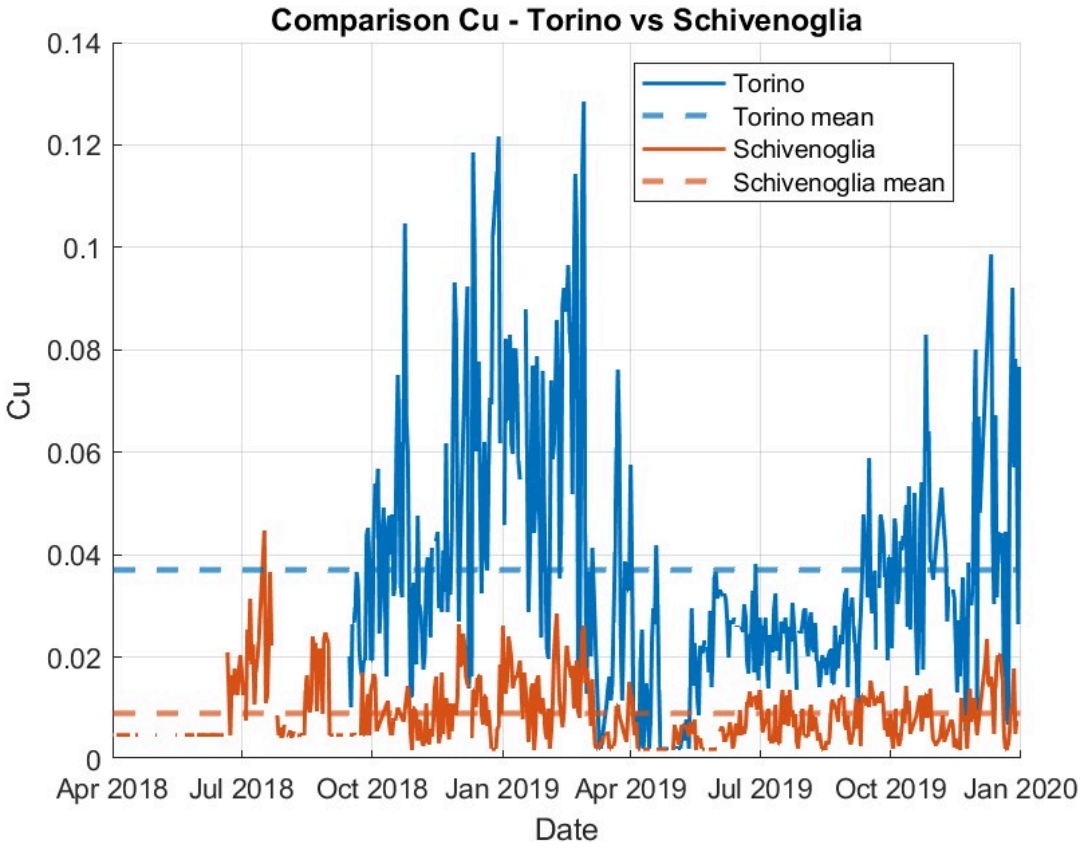


Figure 44: Comparison of Cu trends and mean values for both the sites of Torino and Schivenoglia

3.4. Meteorological correlations

This part of the analysis investigates the correlations between selected meteorological variables and total PM and components concentration. In every case Pearson's and Spearman's correlation coefficient were calculated. During literature review, they turned out to be ideal for paired atmospheric data; Pearson's coefficient measures the strength and direction of a linear correlation between two continuous variables and it is calculated on the original data values but is sensitive to outliers. On the other hand, Spearman's coefficient evaluates the monotonic relationship between two variables, by applying the Pearson's coefficient to the ranked values of the data. Spearman's coefficient is usually preferred in these types of studies, since it's less sensitive to outliers and does not require assumptions about the data distribution and linearity. It should be underlined that very high values of correlation (tending towards 1 or −1) are not expected, since PM variability results from

the interaction of multiple processes, and paired correlations tends to describe just partial relationships within this complex system. (Wilks, 2011)

In this analyses, for the graphical part, time series were smoothed using a moving average in order to reduce short-term variability and enhance the underlying trends, especially for meteorological data.

3.4.1. Temperature vs levoglucosan, EC and total PM_{10}

The first relationship examined was between average daily temperature and levoglucosan, EC and total PM_{10}.

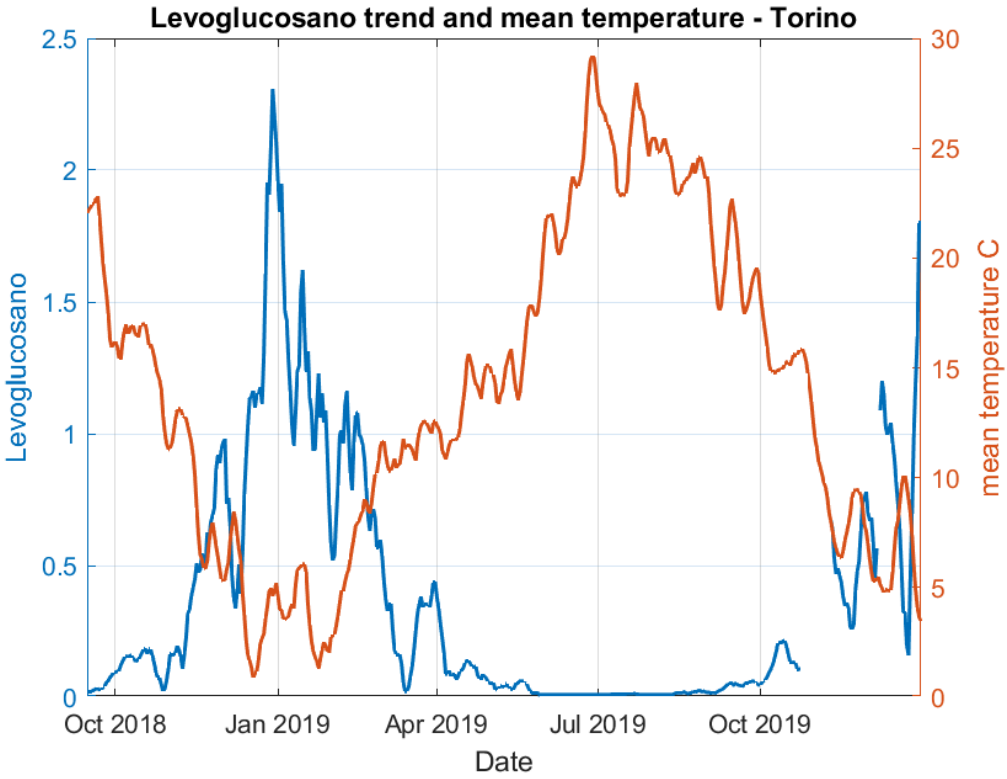


Figure 45: Compared trend between temperature and levoglucosan

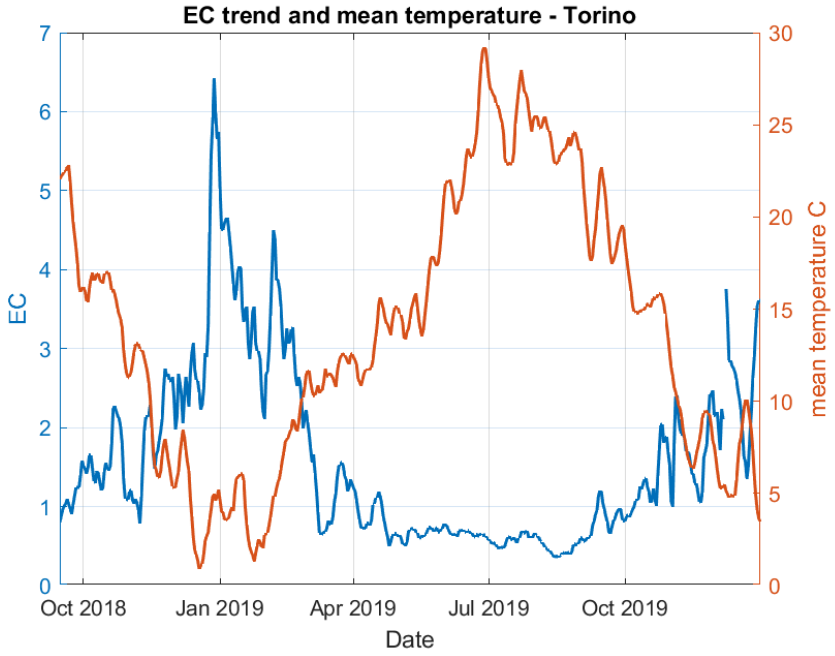


Figure 46: Compared trend between temperature and EC

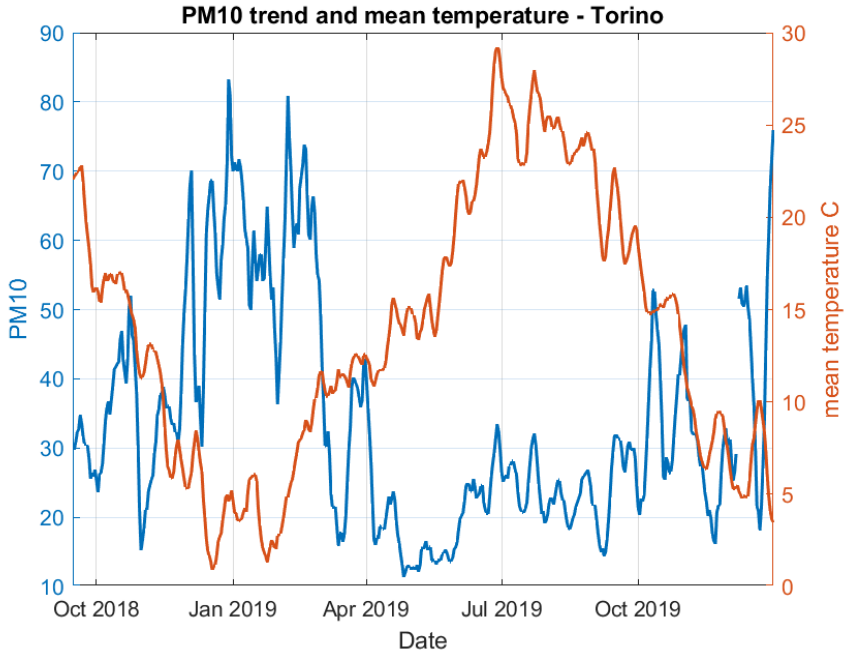


Figure 47: Compared trend between temperature and PM_{10}

	Pearson	Spearman
PM10	−0.48	−0.41
EC	−0.63	−0.68
levoglucosan	0.10	−0.71

Table 5: Pearson's and Spearman's correlation coefficients for the temperature and respectively PM10, EC, levoglucosan.

As discussed previously, all the species analyzed present seasonality but it is more marked for levoglucosan, which is linked mainly to one source (biomass burning) than for PM_{10} whose value is related to various sources. The correlations as expected are negative: increasing temperatures corresponds to reduced heating demand and consequently lower combustion-related emissions. Looking at Table 5, EC and PM_{10} show more consistent Pearson-Spearman agreement, indicating a linear response to temperature, this is probably due to the fact they integrate multiple sources. On the other hand levoglucosan exhibit the strongest monotonic relationship with temperature (Spearman = 0.71). The near-zero Pearson coefficient compared to the Spearman's value indicates a non linear relationship, in fact levoglucosan concentrations increase sharply during cold season, producing a threshold-like seasonal pattern, in total contrast with the linearity assumption required by the Pearson's coefficient.

3.4.2. Average wind speed

Mean wind speed was correlated only with the PM_{10}, since wind acts primarily as a mechanical dispersion mechanism, affecting mainly dilution and not chemical transformation.

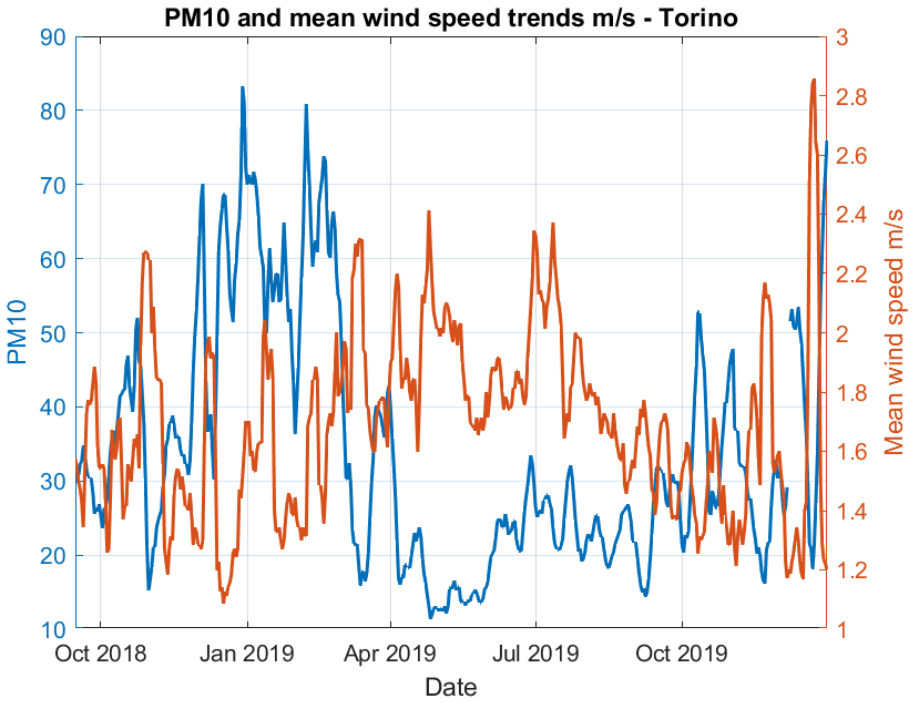


Figure 48: Compared trend between wind mean speed and PM10

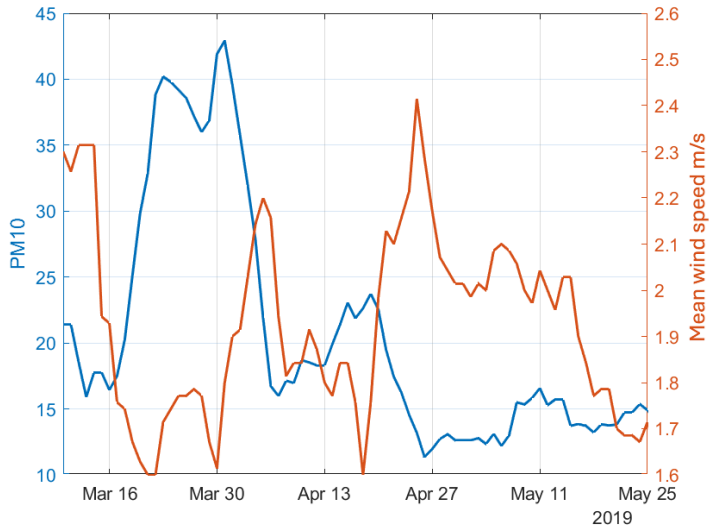


Figure 49: Compared trend between wind mean speed and PM10 - zoom April/May

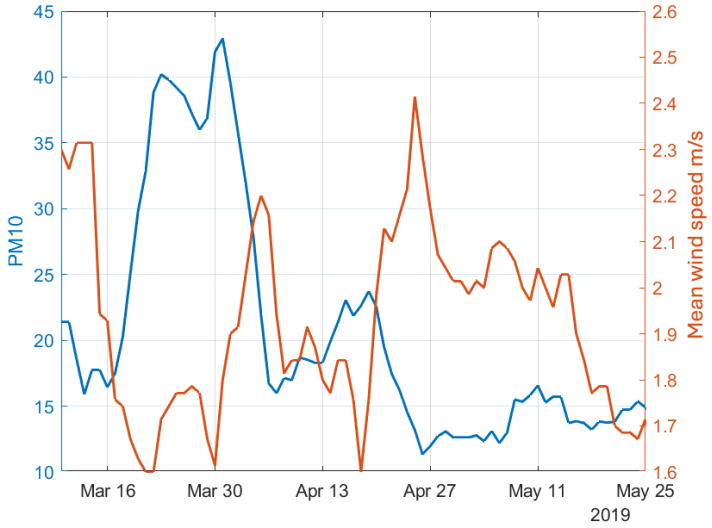


Figure 50: Compared trend between wind mean speed and PM10 - zoom Jun/Aug

	Pearson	**Spearman**
PM10	−0.48	−0.41

Table 6: Pearson and Spearman correlation coefficients for the wind and PM10.

The similarity of the two coefficients indicate a linear and monotonic relationship, we can interpret wind as gradual and continuous process of PM removal. The limited values of correlations can be explained by the fact that wind speed shows aswell limited values (as described in the introduction) and it's not the only meteorological component influencing the PM_{10} trend.

3.4.3. Total precipitation

Precipitation, unlike the wind, acts as a more selective removal mechanism; this mechanism is called wet deposition and according to the "Falling drop" model, it depends on both the characteristics of rain (drop's size, precipitation's intensity and droplets velocity) and PM, in particular, its size. (Xing Gao, 2025)

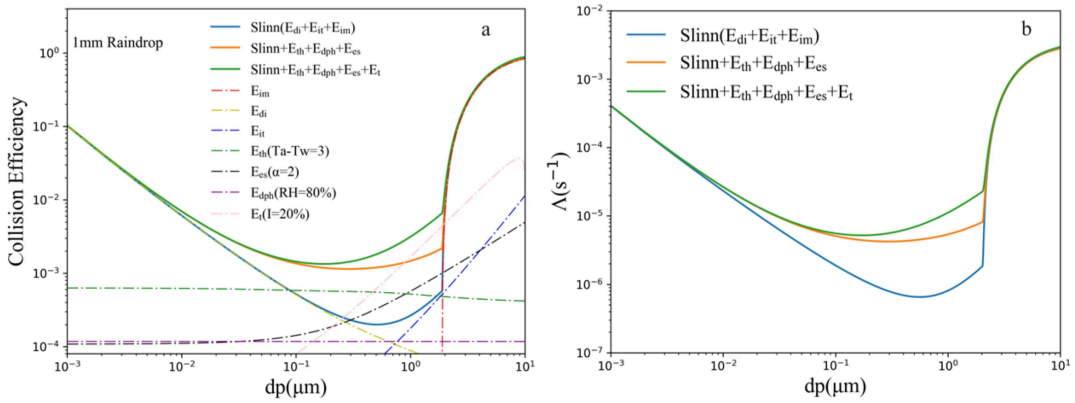


Figure 51: Falling drop model; on the left, the relationship between the collision efficiency and the particle size of the aerosol, on the right,the relationship between the scavenging coefficient and the particle size of the aerosol (Xing Gao, 2025)

According to Figure 51, there is a minimum in scavenging efficiency, between 0.1 and 1 μm, no matter which formula it is used (the lines correspond to different formulas used to calculate the efficiency, taking into account different cases). The range indicated is the one regarding fine particles (accumulation mode): SO_4^{2-}, NO_3^- and NH_4^+, along with EC, OC and trace elements as lead (Pb).(Kulkarni et al., 2011)

The main crustal fraction's components (Aluminum, Titanium, Silicium and Calcium), Lead and Secondary Inorganic PM were correlated with the total daily precipitation.

	Pearson	**Spearman**
Crustal fraction	−0.23	−0.49
Lead	−0.18	−0.33
Secondary inorganic	−0.02	−0.17

Table 7: Pearson and Spearman correlation coefficients for the total precipitation and Crustal fraction, Inorganic secondary fraction and Lead.

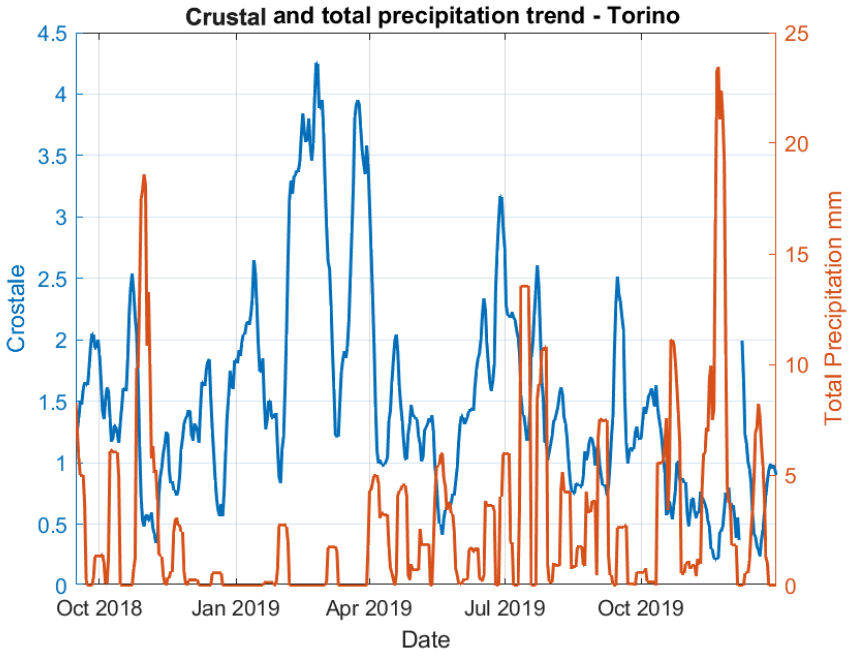


Figure 52: Trend comparison between crustal components and total precipitation

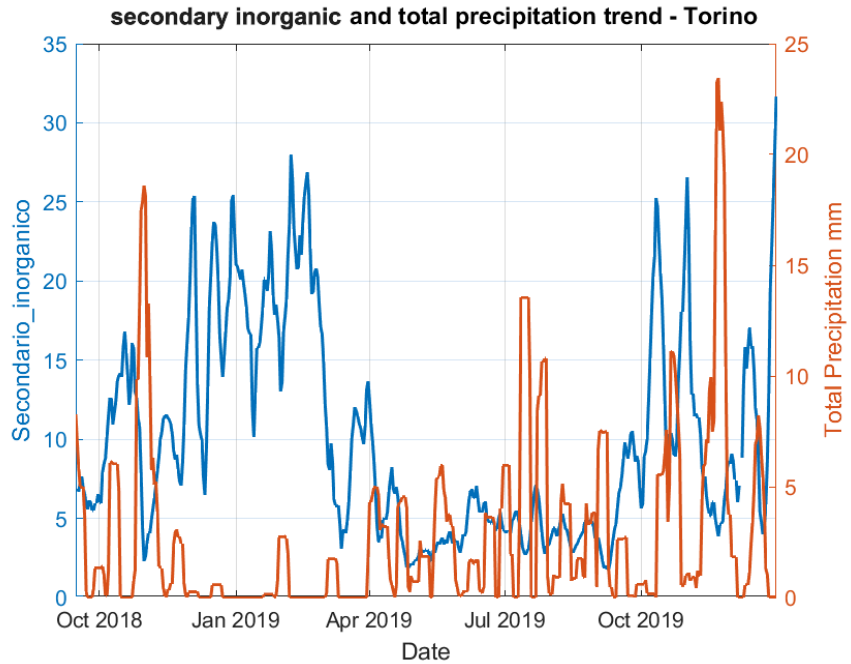


Figure 53: Trend comparison between Secondary inorganic PM and total precipitation

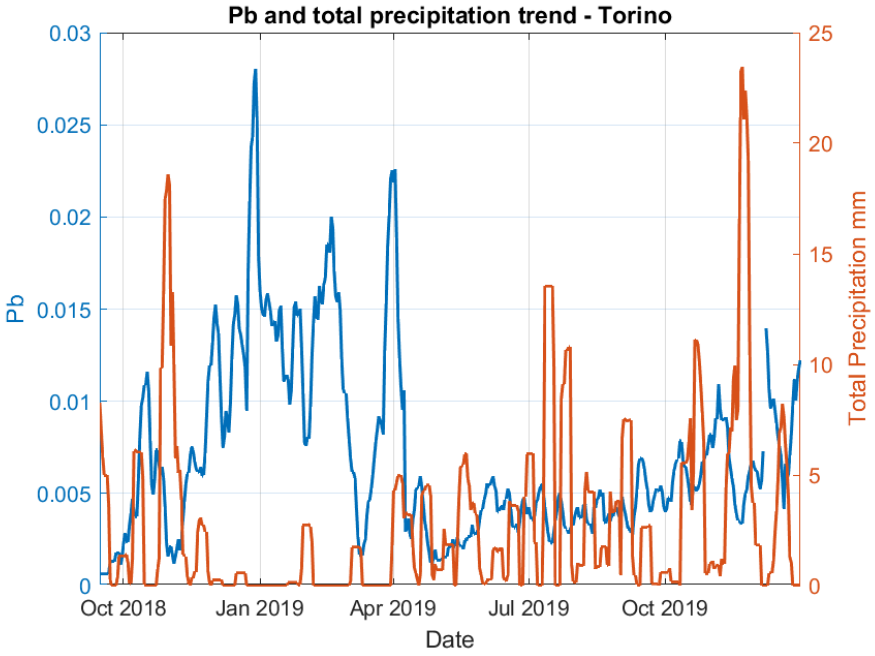


Figure 54: Trend comparison between Lead and total precipitation

From Table 7, Figure 52, Figure 53 and Figure 54 it is clear that precipitations have low influence for some components of PM as the secondary inorganic fraction and the Lead, as expected, while it has a visible impact on the Crustal component. The values of correlations shows that the relationships analyzed are quite far from being linear and even the Spearman's coefficient for the crustal fraction reaches just 0.49, plausibly linked with the mid-low values of the total daily precipitation, as for the wind.

3.4.4. Total radiation

Considering the total radiation (W/m^2), it influences both the mixing height and the chemistry of the PM (photochemical reactions linked to components as nitrates). To observe this different impact correlations with different components were chosen: total PM, Ammonium, Nitrates.

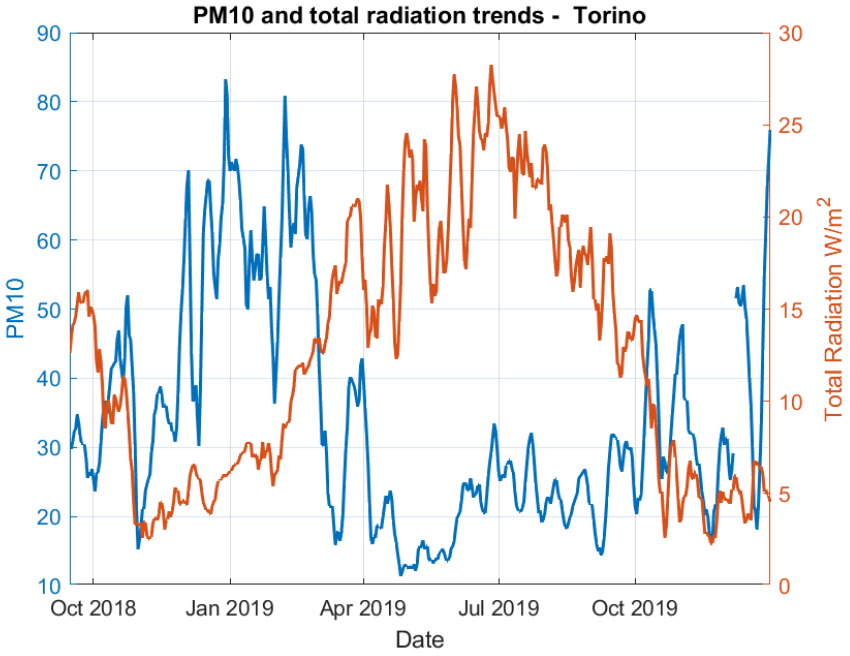


Figure 55: Trend comparison between PM_{10} and total radiation

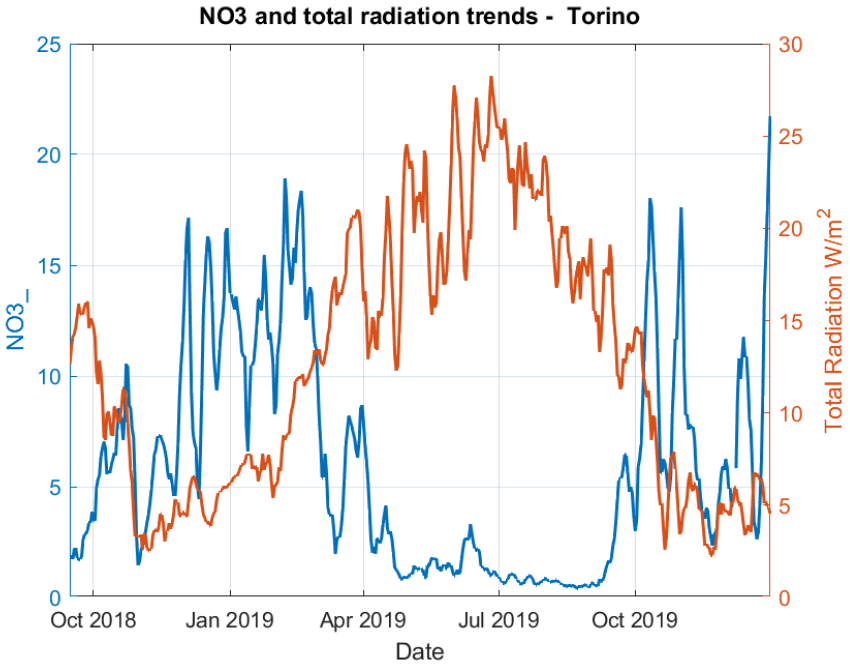


Figure 56: Trend comparison between Nitrates and total radiation

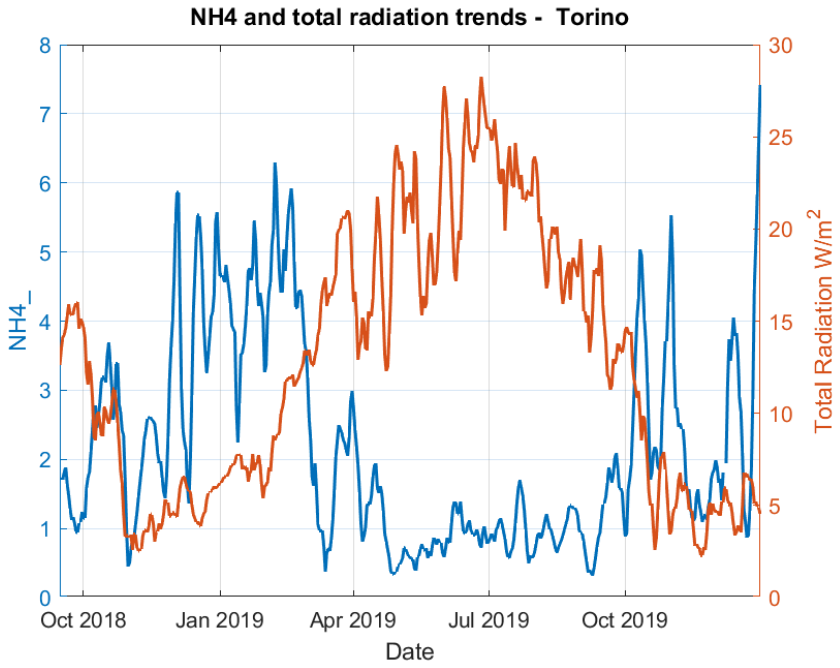


Figure 57: Trend comparison between Ammonium and total radiation

	Pearson	**Spearman**
PM10	−0.39	−0.39
Ammonium	−0.11	−0.51
Nitrates	−0.48	−0.59

Table 8: Pearson and Spearman correlation coefficients for the Total radiation and PM10, Ammonium, Nitrates

The dilution effect is reflected by the negative (linear) correlation with PM_{10}, while the interaction with the stability of secondary aerosols as nitrates and ammonium is plausibly reflected by the strong negative correlation values of Spearman's coefficient; in fact as remarked in the introduction, ammonium nitrate, under warm, high radiation conditions, tends to partially dissociate into gaseous species. PM_{10} correlation is weaker likely because it depends on multiple processes and sources.

This interpretation is also supported by the negative relationship observed between PM_{10} and mixing height, confirming the dominant link between mixing height and PM concentration (Figure 58).

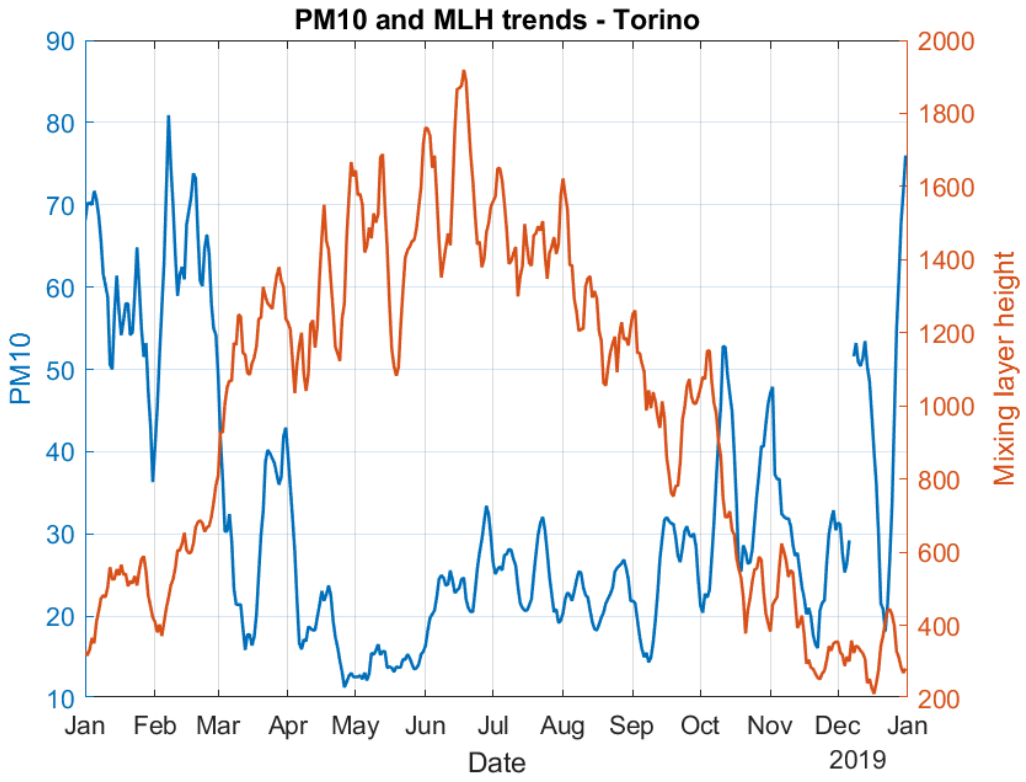


Figure 58: Trend comparison between PM_{10} and Mixing Layer Height, Spearman : −0.46, Pearson: −0.48

In conclusion, it should be noted that solar radiation, temperature and mixing height co-vary seasonally, therefore it should be considered that part of the studied correlations may reflect the underlying seasonal cycle rather than a total causal relationship.

3.5. Machine learning models

3.5.1. Use of ML models for air quality

Machine learning (ML) is an interesting tool used nowadays in the field of air quality modeling; it is becoming indispensable, since traditional models struggle to capture the non linear relationships between atmospheric pollutants and their sources.

These models are used both for estimation (determination of the concentration of pollutant at the same time step) and forecasting (prediction based on historical time-series).

The algorithms applied are various (e.g. Ensembling learning, Neural Networks, Support Vector Machines and Multiple Linear Regression) and depending on the task some are more effective than others (Masih, 2019).

Overall the main used in the field of air quality are the Ensembling Learning ones; they in fact are less sensitive to overfitting, manage non-linearity, missing data and have overall a superior performance (Masih, 2019; Rybarczyk & Zalakeviciute, 2018). This techniques "train" multiple predictors to address the same problem combining their result into a single prediction, the core principle is that a group of "weak" predictors provides better accuracy and generalization than any individual model (or component). This is achieved by the introduction of randomness or stochasticity which ensures that the individual predictors are diverse and do not make all the same errors.

The randomness can be introduced in different ways, in order to build the ensemble:

- Bagging, in this case the stochasticity is added primarily in the training data, creating several different subsets of the original database by taking random samples with replacements (bootstrapping); then a separate predictor is generated for each of these random datasets, and finally the output is calculated with the average of all the predictions or through a majority vote, based on the task required (regression or classification respectively).
- Boosting, randomness is introduced through sequential training, rather than parallel data sampling. Consequently each new model in the sequence is designed to specifically address the incorrectly predicted observations from the previous one; usually it is more accurate compared to bagging, even though is more sensitive to noise.

Other models used in fewer studies include linear regression models and artificial neural networks. Linear regression models describe the relationship between predictors and the response variable through a linear combination of input variables, while neural networks are nonlinear model that can represent more complex relationships through layers of interconnected processing units (neurons) and weighted connections, adjusted during the training phase.(Masih, 2019; Rybarczyk & Zalakeviciute, 2018)

The goodness of fit of these models are evaluated with different parameters:

- R^2 Coefficient of determination, a dimensionless index, ranging from 0 (no correlation) to 1 (perfect correlation) and represent the fitting degree of a model. The average values for ensemble models range between 0.6 and 0.8 depending on dataset, predictors and spatial context.
- MAE Mean Absolute Error, that measure the average magnitude of the difference between the predicted values and the observed ones, its range depend on the scale of the data.
- RMSE Root Mean Square Error, it's similar to MAE but gives a higher weight to larger errors
- RAE Relative Absolute Error, it is the variance of the model, useful to compare models using different scale data

(Masih, 2019; Rybarczyk & Zalakeviciute, 2018)

For this thesis, the Regression Learner App of Matlab was used, with a focus on "Ensemble of Trees" models; the data of interest used for every case, were put together in a table after being selected (only the year 2019 was considered) and normalized. The results showed in the next sub-sections, regards the ensemble model that for every case had reached the best result (Bagged or Boosted). Model performance during training was evaluated using 5-fold cross-validation: this means that 2019 dataset was portioned into five subsets, each of these was used for validation and the other four for training, at every iteration; validation metrics were then averaged across folds. After validation the model was retrained using all the 2019 data. For every case also meteorological data were used; in order to reduce the possibility of over fitting, all the available meteorological data were mutually correlated (Spearman's coefficient) and consequently only few of them were chosen, the ones that showed the lowest correlations:

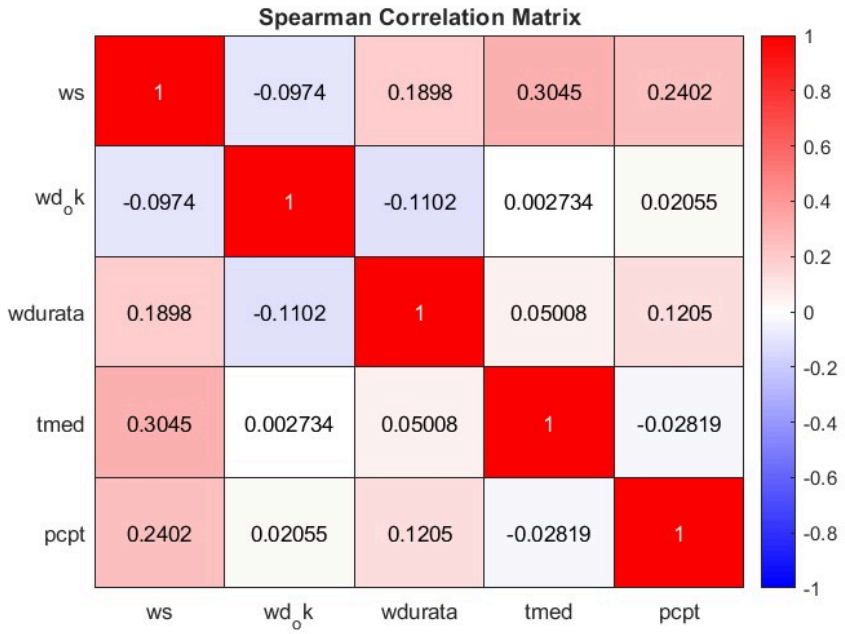


Figure 59: Matrix of correlation

Abbreviation	Variable	Unit
ws	Wind Speed	m s^{-1}
wd_ok	Wind Direction	degrees
wdurata	Wind Duration	h
tmed	Mean Air Temperature	°C
pcpt	Precipitation	mm

Table 9: Used meteorological variables, their abbreviations and unit of measure

3.5.2. Use of Ensembles of Trees models for estimation.

An interesting use of these models is the estimation of components or pollutants that are not regularly or easily measured at the sampling stations, starting from data that are more quickly available. The first component chosen was levoglucosan: in fact usually it is not measured. For this example,the selected predictors were the meteorological data (Table 9), PM_{10}, Copper (key parameter for traffic) and EC (it includes traffic and combustion as sources). The best results were achieved by the Bagged tree model:

Metric, **Training (Validation)**
RMSE: 0.11549
R^2: 0.69
MSE: 0.013337
MAE: 0.05709

Table 10: Training results of the model - levoglucosan estimation

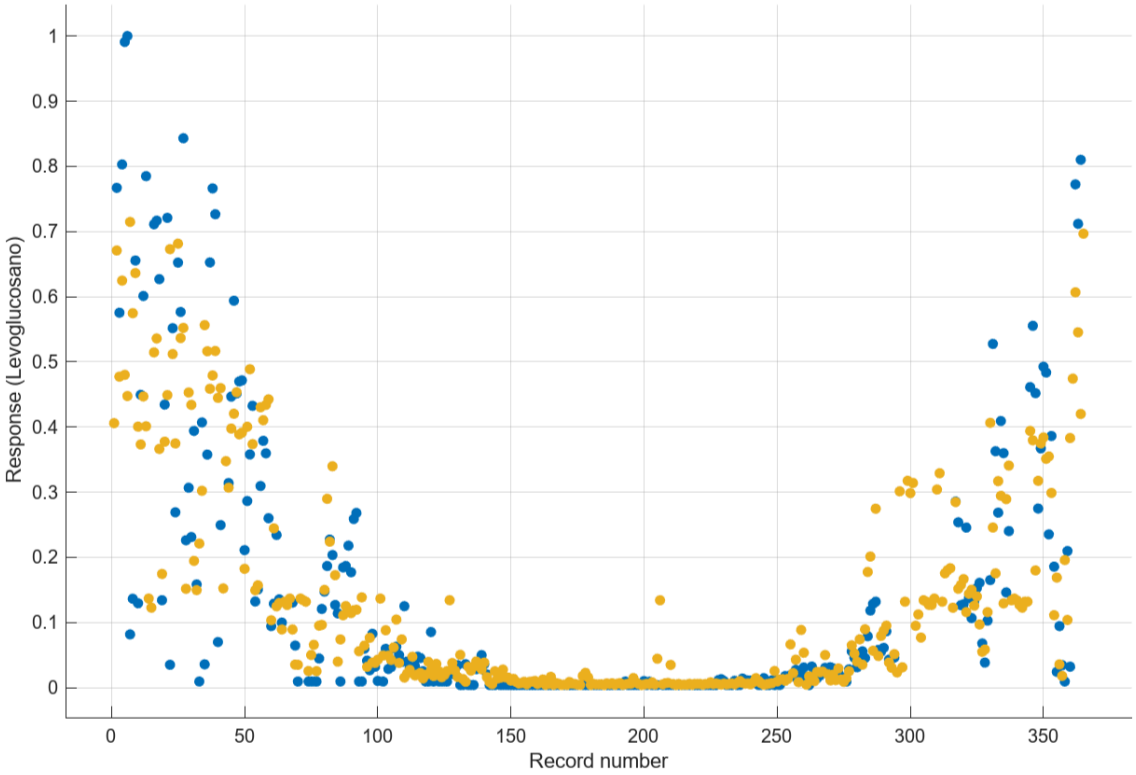


Figure 60: Response plot - true (blue dots), predicted (yellow dots) - for levoglucosan concentration

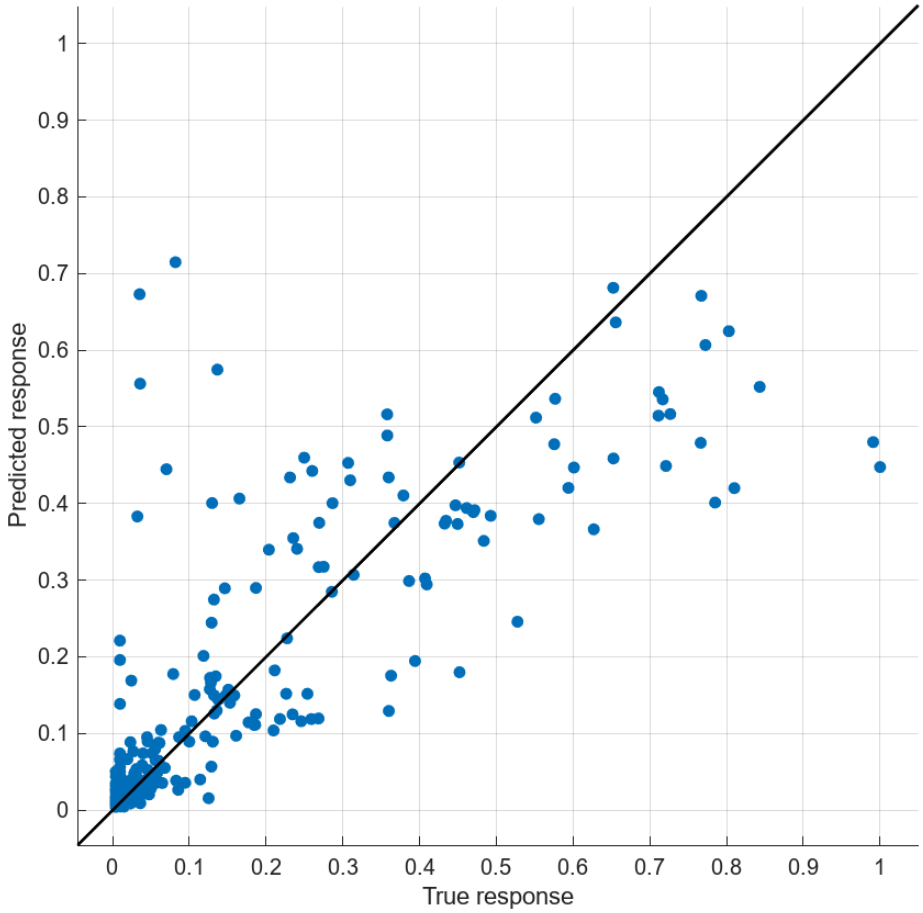


Figure 61: Predicted (line) vs observations (dots) plot for levoglucosan concentration

The model shows a satisfactory performance in reproducing levoglucosan concentrations, being the R^2 approximately 0.7, showing that a major part of the observed variability is captured. The seasonal pattern is effectively reproduced (Figure 60). The match between observed and predicted vales is strong at low concentrations, while dispersion increases at intermediate values, showing an underestimation at high concentrations, the model smooths extreme events (Figure 61). This could be justified by the use of a relatively limited dataset with few peaks. Improvements could include extending the dataset to multiple years in order to capture a higher number of events with elevated concentrations, refining the predictors and improving emission proxies. Without these adjustments the model is still quite far from being a good substitute of the analytical measure of levoglucosan, for now it's more a supportive estimation tool.

The second attempt was made by trying to estimate OC from meteorological variables (Table 9), BTEX (Benzene, toluene, xylene), NO_x, Ozone and PM_{10}. This trial was made since BTEX are part of VOC_s (volatile organic compounds) and they can be precursors of the secondary fraction of OC, or at least can be an indicator of the presence of other VOC_s that are aswell precursors. The ozone is a key parameter indicating photochemical activity resulting in a higher transformation of VOC_s into PM. Meanwhile NO_x are a proxy of traffic and combustion, linked to the primary component of OC. (Seinfeld & Pandis, 2006)

Metric, **Training (Validation)**
RMSE: 0.066639
R^2: 0.87
MSE: 0.0044407
MAE: 0.047227

Table 11: Training results of the model - OC estimation

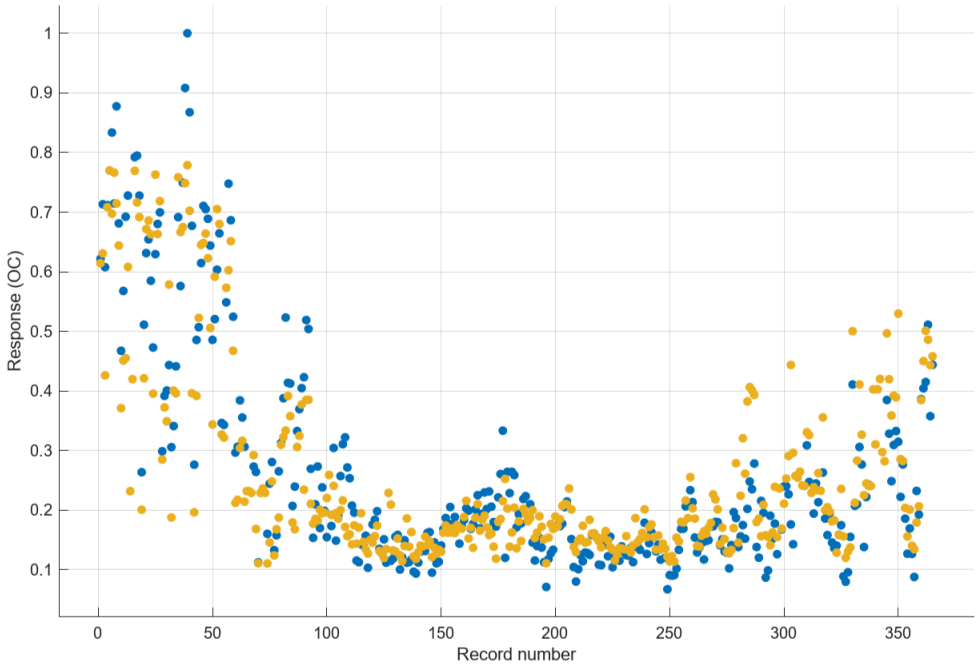


Figure 62: Response plot - true (blue dots), predicted (yellow dots) for OC concentration

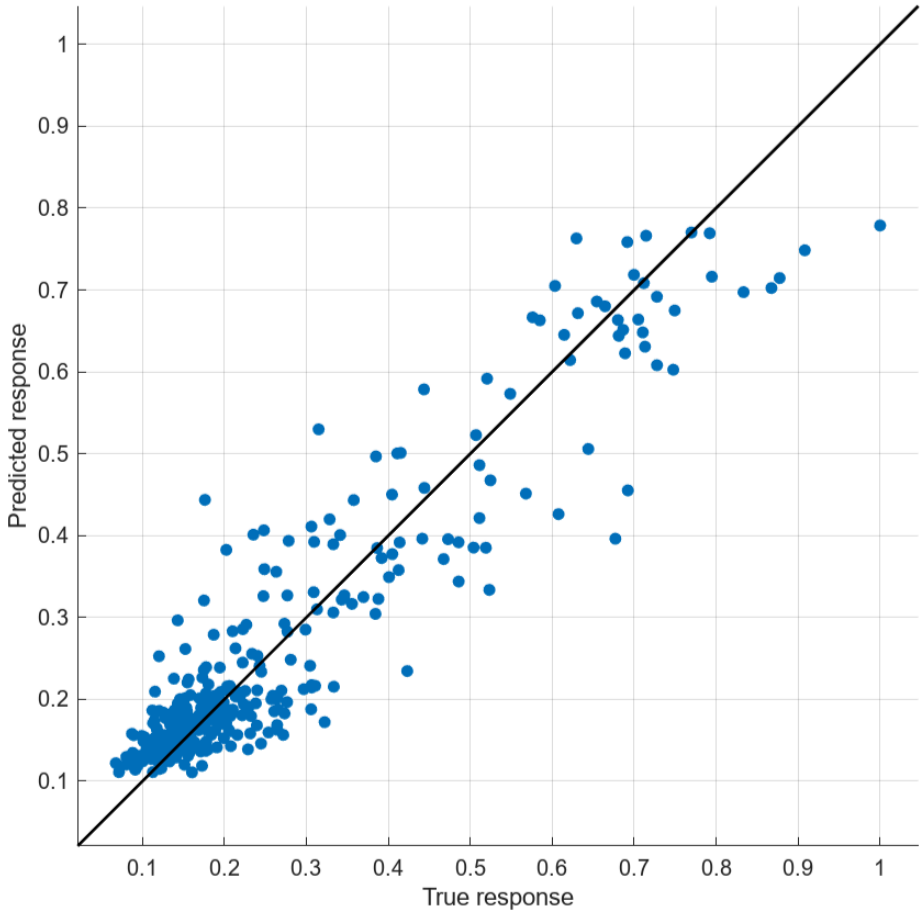


Figure 63: Predicted (line) vs observations (dots) plot for OC concentration

In this case the performance of the bagged trees model is definitely improved, compared to the levoglucosan case, with a R^2 of 0.87 and a relatively low RMSE. The model explains a large fraction of the observed variability. This indicates that BTEX, NO_x,O_3, and meteorological variables are a physically meaningful and informative set of predictors for organic carbon.

In the temporal response plot (Figure 62) the model reproduce well the seasonal variability without exhibiting strong smoothing of high values, suggesting that both primary and secondary OC are effectively represented. The limitation in this case could be, again, the limited dataset, especially for generalizability.

The last estimation trial was made for Inorganic Secondary Particulate, using the meteorological variables in Table 9, Ozone, NO_x, Ammonia and PM_{10}. in this case

ammonia and NO_x are linked to the presence of nitrates in the particulate, while the ozone is used again as an indicator of photochemical activity, linked to the equilibrium between sulfates and nitrates.

Metric, **Training (Validation)**
RMSE: 0.068939
R^2: 0.86
MSE: 0.0047514
MAE: 0.043001

Table 12: Training results of the model - Secondary inorganic particulate estimation

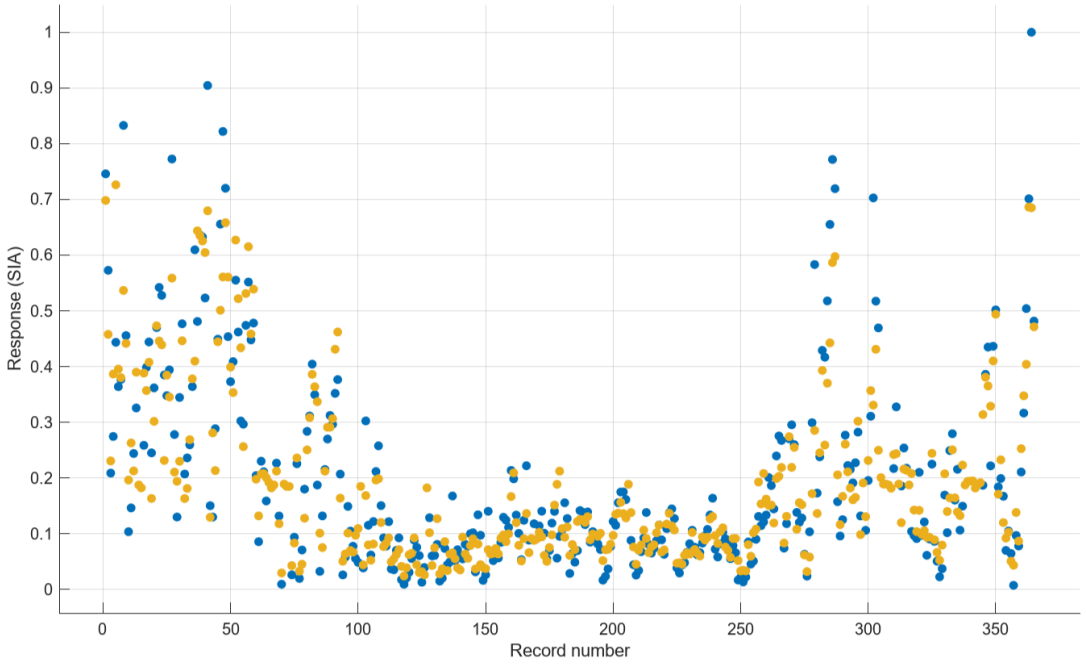


Figure 64: Response plot - true (blue dots), predicted (yellow dots) for secondary inorganic particulate concentration

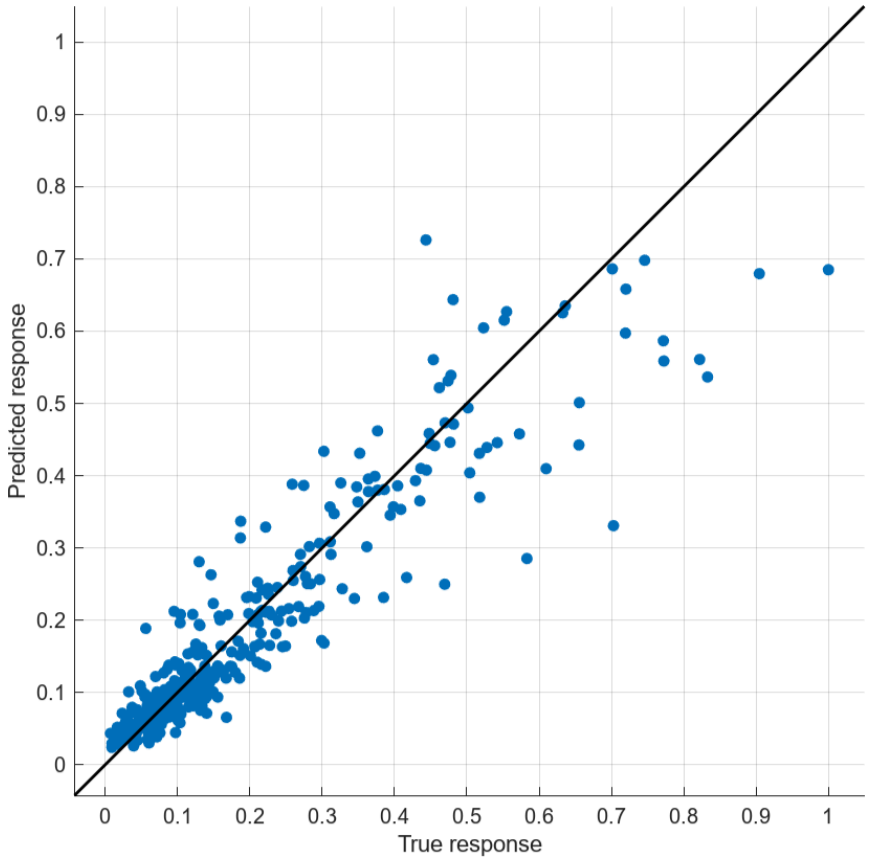


Figure 65: Predicted (line) vs observations (dots) plot for secondary inorganic particulate concentration

In this case the best performance was obtained by the boosted trees model (the difference is minimal considering the same R^2 and a RMSE 0.001 lower). Also in this case the model shows a strong performance (Table 12). The variability of the secondary inorganic aerosols (SIA) is well captured by it (Figure 64), although some dispersion remains during higher concentration episodes. Slightly scattered high values are expected considering their secondary nature and the limited number of predictors used.

A peculiarity of ensemble model is that they allow to perform the feature importance analysis, in order to better understand the contribution of every predictor to the model output. (Masih, 2019; Rybarczyk & Zalakeviciute, 2018) The results are shown in Figure 66, Figure 67, Figure 68.

For the levoglucosan model, the dominant predictor is Elemental Carbon (EC), while for both the OC and SIA models, the most important predictor is PM_{10}, suggesting that all of the models mainly rely on emission-related proxies rather than on meteorological variables and consequently dispersion processes to estimate the target's concentrations. In the SIA model the PM_{10} predictor seems to have almost the totality of the importance and this is consistent with the fact that SIA represent one of the largest fractions of PM, while in the OC model also benzene and xylene show a considerable contribute, indicating that OC variability is closely associated with total particulate levels (being part of it) but also with the presence of gaseous precursors.

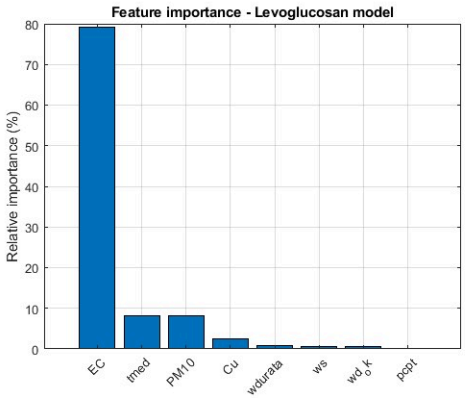


Figure 66: Importance feature plot - Levoglucosan estimation model)

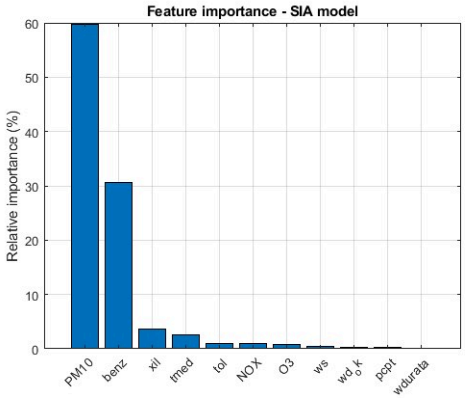


Figure 67: Importance feature plot - OC estimation model)

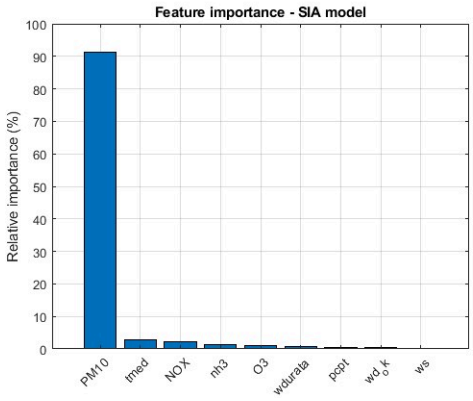


Figure 68: Importance feature plot - SIA estimation model)

These three examples suggest that ensemble trees methods are particularly suited for integrated, heterogeneous and multi-process variables as OC and SIA, and less effective for highly specific and episodic tracers as levoglucosan.

3.5.3. Predictive analysis

For the predictive part different models were trained with 2019 data and tested with 2021 data.It's fundamental to highlight that the testing data were normalized with 2019 data in order to achieve consistency and avoid data leakage. For every trial both boosted trees and bagged trees were trained but only the best one was chosen for the testing part (Although for every case they reached very similar values). The first trial was made with only meteorological predictors; (since the available variables for 2021 were less than the ones available for 2019, just those in common were considered, among the one used in Table 9: ws, tmed, pcpt).

Metric, **Training (Validation)**
RMSE: 0.14430
R^2: 0.47
MSE: 0.020822
MAE: 0.10408

Table 13: Training results of the model (Bagged trees) - PM_{10} prediction with only meteorological variables

Metric, **Testing**
RMSE: 0.16250
R^2: 0.27
MSE: 0.026405
MAE: 0,11879

Table 14: Testing results of the model (Bagged trees) - PM_{10} prediction with only meteorological variables

the R^2 is already relatively low in the training and it gets lower in the testing (Table 13,Table 14); this is also confirmed by the low alignment of the predicted and observed values in Figure 69 , meaning that the model has a weak generalization capacity, likely due to the limited amount of informations and for the absence of any emission proxy. The trial with only meteorological variables was done also with all the available meteorological data (relative humidity, total precipitation, total solar radiation, mean temperature, maximum temperature, minimum temperature, wind speed, calm wind fraction, wind gust days), but the testing performance was worse (R^2 approximately 0.24), plausibly because of the overfitting introduced by correlated variables.

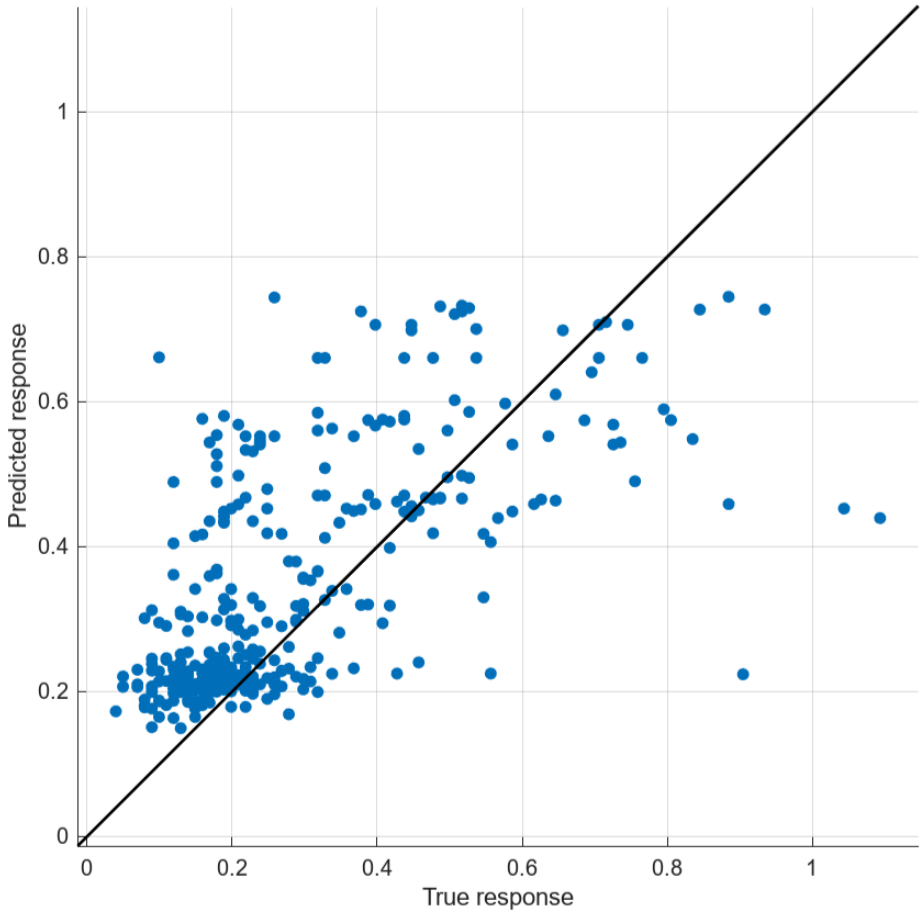


Figure 69: Predicted (line) vs observations (dots) plot for PM_{10} concentration with only meteorological predictors

The second trial was made with the introduction of the values of PM_{10} from the previous day, in order to simulate the memory of the atmospheric system, especially in the Po Valley, characterized by a low magnitude of dispersion.

Metric, **Training (Validation)**
RMSE: 0.10767
R^2: 0.70
MSE: 0.011594
MAE: 0.07339

Table 15: Training results of the model (Boosted trees) - PM_{10} prediction with meteorological variables and one day of lag in PM_{10} data

Metric, **Testing**
RMSE: 0.0967
R^2: 0.74
MSE: 0.0094
MAE: 0,0658

Table 16: Testing results of the model (Boosted trees) - PM_{10} prediction with meteorological variables and one day of lag in PM_{10} data

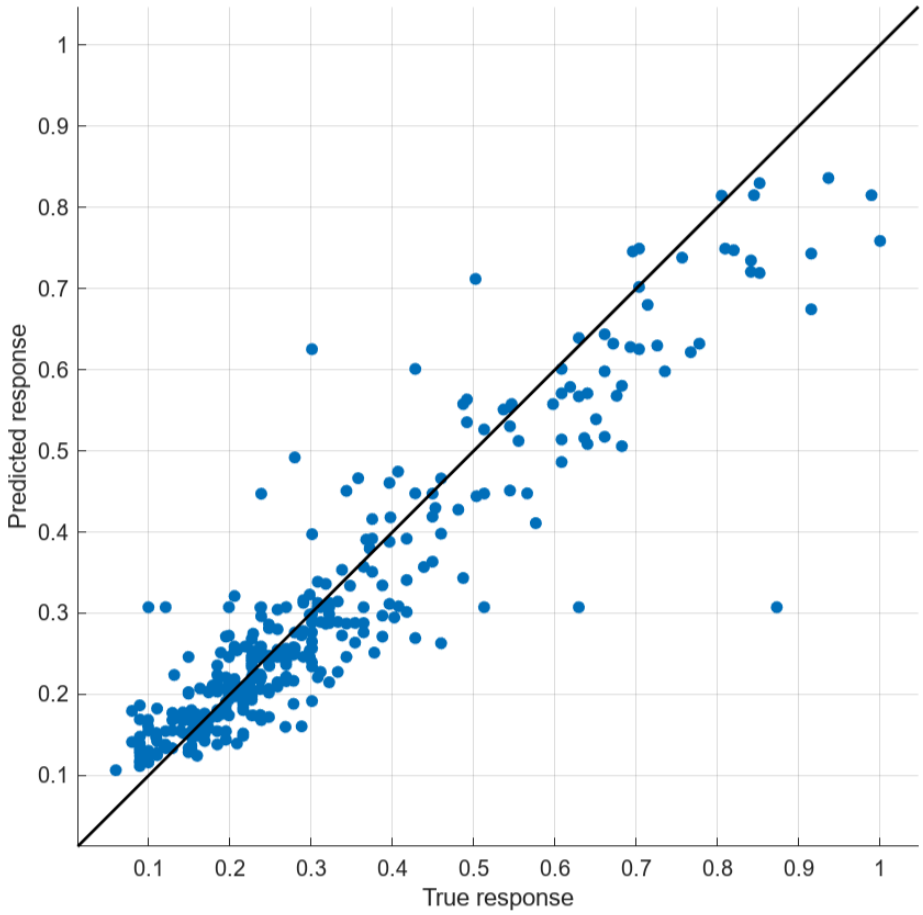


Figure 70: Predicted (line) vs observations (dots) plot for PM_{10} concentration with meteorological predictors and one day of lagged PM_{10} values

In this case, from both the metrics of learning and testing (Table 15 and Table 16) and the plot of observed and predicted responses in Figure 70, it is evident that the model is more accurate, this was expected because of the inclusion of the lag of

PM_{10} among the predictors, but it's still surprising the slight improvement of the metrics (Table 16) in the testing phase. This suggests good annual generalization capacity and could reflect a less variable concentration distribution compared to training year and a stable persistence of PM_{10}.

The third and final trial was made with the introduction of two days of lag values of PM_{10} in order to see if this addition improves the model's performances.

Metric, **Training (Validation)**
RMSE: 0.10730
R^2: 0.71
MSE: 0.011512
MAE: 0.0743

Table 17: Training results of the model (Boosted trees) - PM_{10} prediction with meteorological variables and two days of lag in PM_{10} data

Metric, **Testing**
RMSE: 0.09922
R^2: 0.73
MSE: 0.00984
MAE: 0,0683

Table 18: Testing results of the model (Boosted trees) - PM_{10} prediction with meteorological variables and two days of lag in PM_{10} data

While the learning phase parameters (Table 17) are similar to those of the model with one day of PM_{10} lag (Table 15), the testing phase has slightly lower performances (Table 18), likely explainable with the effect of overfitting, in fact the two values of lags are very correlated between them and they can be perceived by the model as noise.

Before selecting the final model for the predictive analysis, additional regression models were tested (available in the MATLAB regression learner), using the predictor configuration that showed the best performances (meteorological variables and lagged PM_{10}). These included linear regression and neural network models.

Model	RMSE	MSE	MAE	R^2
Linear Regression (Interactions)	0.0880	0.0077	0.0645	0.79
Narrow Neural Network	0.1084	0.0117	0.0766	0.68
Boosted Trees	0.0967	0.0094	0.0658	0.74
Bagged Trees	0.0992	0.0098	0.0683	0.73

Table 19: Comparison of regression algorithms testing phase (2021), predictors: meteorological variables and lagged PM_{10}.

Linear regression showed slightly better performances in RMSE, MAE and R^2, suggesting that relationships between the predictors and PM_{10} is mainly linear in the present dataset. The inclusion of lagged PM_{10} introduces a strong factor of auto-regression, allowing relatively simple models to capture a substantial portion of the temporal variability.

Nevertheless, ensemble tree models were used for the subsequent predictive analysis because of their flexibility in capturing potential non-linear relationships between predictors, as well as their widespread use in air quality modelling studies. (Masih, 2019; Rybarczyk & Zalakeviciute, 2018)

The last step of the analysis with ML models was to apply the best one so far to all the years available (2021 - 2024).

Metric	2021	2022	2023	2024
RMSE	0.0967	0.0958	0.0840	0.0930
R^2	0.7400	0.7500	0.7200	0.7500
MSE	0.0094	0.0091	0.0071	0.0087
MAE	0.0658	0.0646	0.0613	0.0660

Table 20: Table of the testing metrics for the different years tested

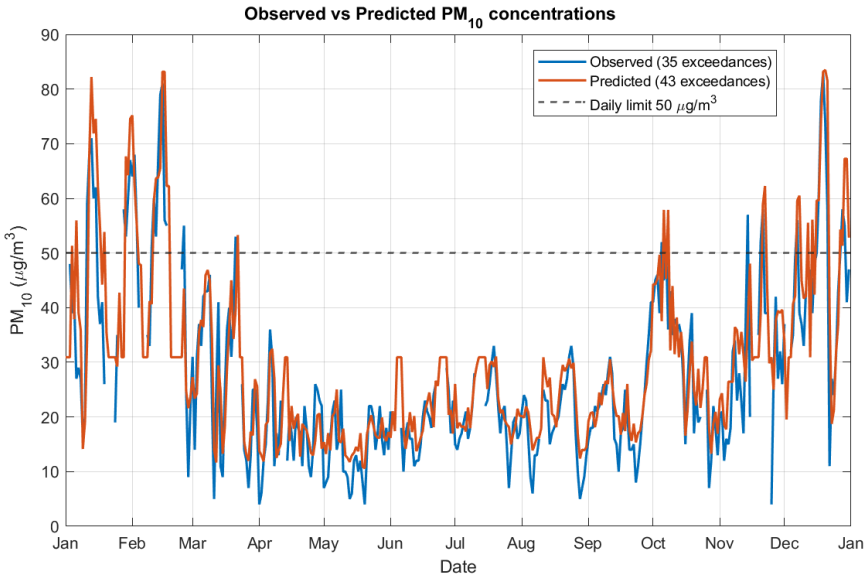


Figure 71: temporal plot of observed and predicted PM_{10} concentrations with the current daily limit - year 2023

Overall the model continues demonstrating a stable inter-annual performance. R^2 values are consistently between 0.72 and 0.75, with relatively low and stable error metrics (RMSE and MAE). Considering the error metrics, the best performance is achieved with 2023 data (Figure 71), and the model maintains acceptable predictive skills across all years, with results in line with values reported in literature for comparable data approaches (Masih, 2019; Rybarczyk & Zalakeviciute, 2018).

After the metrics analysis, the number of predicted and observed exceedances (of current daily limits) was compared for every year.

Metric	2021	2022	2023	2024
Observed exceedances (>50 µg/m³)	46	66	35	54
Predicted exceedances (>50 µg/m³)	53	66	43	62
Hits	35	48	26	45
Misses	11	18	9	9
False alarms	18	18	17	17

Table 21: Comparison between observed and predicted PM_{10} exceedances (daily limit 50 µg/m³). Hits represent correctly predicted exceedances, misses represent observed exceedances not captured by the model, and false alarms represent predicted exceedances not observed in the measurements.

The model manages to capture the majority of exceedances Table 21 (between 73% and 83%); although some exceedances days are not captured, while some additional are predicted, the results indicate that the model is able to reproduce the general occurrence of high-concentration episodes.

Given the limited dataset and the lack of advanced inputs as transport models outputs, detailed emission inventories or satellite-data, the achieved predictive performance can be considered satisfactory .

Chapter 4

Discussion

4.1. PM_{10} levels and regulatory implications

The Data analysis confirms the persistence of critical air quality conditions in the Po Valley. Although the annual mean values generally comply with the limits set by the current European legislation, the number of daily exceedances and the comparison with the new limits introduced by the Directive 2024/2881 highlight a scenario which is far from meeting future targets.

The seasonal variability observed in the dataset clearly reflects the well known pattern typical of the Po Valley, with higher PM_{10} concentrations during the winter, linked to the combination of increase in emission and low dispersion conditions. The comparison with historical datasets from regional agencies indicates that the concentration of particulate matter have gradually decreased over the last years, reflecting the effectiveness of emission reduction policies and technological improvements. However the observed reduction rates seems to be insufficient to achieve the targets introduced by the new European Air Quality Directive, which moves closer to the WHO guidelines.

4.2. Chemical composition of PM_{10}

The chemical composition analysis reveals that the dominant components of PM_{10} at all sites are organic carbon (OC) and secondary inorganic aerosols (SIA), mainly composed of nitrate (NO_3^-), sulfate(SO_4^{2-}) and ammonium (NH^+_4). These results are consistent with previous studies conducted across Europe and Worldwide (Aas et al., 2012; Putaud et al., 2010; Snider et al., 2016). The mass closure analysis further supports this interpretation and the reconstructed PM_{10} mass explains between approximately 84% and 93% of the measured particulate mass across the investigated sites, showing that the main chemical fractions were successfully captured by the analytical measurements, confirming the consistency of these results with the previous consulted studies (Terzi et al., 2010). A clear seasonal pattern is also visible from the chemical composition analysis, in fact, nitrates tend to increase during

winter while sulfates are relatively high during summer. This trend was expected after the review of previous analyses conducted across Europe (Aas et al., 2012) and of literature (Seinfeld & Pandis, 2006).

4.3. Urban - rural differences and source implications

The comparison between urban and rural sites provides important insights about the influence of emission sources and atmospheric processes on PM_{10}.

Urban sites generally show stronger contributions from primary PM linked to combustion sources, in particular traffic and residential heating, while the rural site exhibits a higher relative contribution of secondary components, reflecting the typical timescales and space scales of secondary aerosols formation processes. These findings are consistent with previous studies conducted across the world (Snider et al., 2016).

The analysis of carbonaceous components and levoglucosan supports this interpretation. Seasonal patterns in EC,OC and consequently their ratio show a clear increase during cold season largely attributable to residential biomass combustion. Higher ratios in the rural site show a major contribute to long term transport particles.

Metal traces were also taken into account, highlighting differences between urban and rural environments. Higher copper concentrations observed in urban sites are linked to non-exhaust traffic emission while higher nickel concentrations observed at the rural site suggest the influence of nearby industrial activities, consistent with the known industrial emission sources in the region.

4.4. Role of meteorology

The analysis has confirmed that meteorology plays a key role in controlling the variability of PM_{10} concentrations in the region. The correlation analysis performed showed how temperature, wind speed, precipitation, solar radiation and the height of the mixing layer influence the pollutant trends consistently with previous studies analyzed (Caserini et al., 2017; Kirešová & Guzan, 2022). Temperature shows a negative correlation with PM_{10} values, reflecting seasonal emission patterns and atmospheric stability conditions. Wind speed also shows a negative correlation with

PM_{10} indicating its role in decreasing pollution concentrations as well as precipitations, but for both of them the relatively weak correlations observed suggest that they cannot fully explain the PM variability especially for a region characterized by weak winds and relatively low precipitations (in particular during winter). Solar radiation appears to influence PM variability through multiple mechanisms. Increased radiation, in fact, increase the height of the PBL, and consequently, the atmospheric mixing, but also influences photochemical reactions involved in secondary aerosols formation. The negative correlations observed between solar radiation and nitrate plausibly reflect the thermal instability of ammonium nitrate, which tends to dissociate with higher temperatures. The correlation with the height of the mixing layer shows how it limits pollutant dilution when it's lower, leading to higher concentrations near the surface. Finally, it's important to underline that these pairwise correlation can only partially describe the complexity of atmospheric processes, influenced by multiple interacting factors.

4.5. Implications of machine learning models

The machine learning models applied in this study show, in most of the cases, a satisfactory capability in reproducing PM_{10} variability, using meteorological variables and emission-related proxies as predictors. The choice to use mainly ensemble trees models, was made after a comprehensive literature review; in fact they seem to be the most used in the field of air quality and the most reliable when dealing with non-linear datasets. They performed better when estimating variables influenced by multiple sources and processes, such as organic carbon and secondary inorganic aerosols. In contrast the estimation of highly specific components as levoglucosan was more challenging, likely linked to the episodic nature of biomass burning emissions and the relatively limited number of predictors available. Also their predictive skill proved to be satisfactory, even if it was not the best, in fact the linear regression model achieved slightly better performances, plausibly due to the high linearity and auto-regressiveness of the dataset (containing the lag of one day of PM10). The best predictive performances were achieved through the inclusion of lagged variables, indicating that PM_{10} concentrations exhibit strong temporal persistence in the Po Valley. These results are consistent with those reported in other studies applying machine learning models to air quality. (Masih, 2019; Rybarczyk & Zalakeviciute, 2018)

4.6. Limitations of the study

Despite the results achieved by this study, some limitations should be highlighted. The most important is likely the relatively limited dataset, especially in terms of temporal coverage. The analysis mainly focuses on data from 2018-2019, which may not fully capture long term variability of PM_{10} and its components. Another limitations that has been encounter is the data-driven nature of machine learning models, although they can capture statistical relationships between variables, they do not explicitly explain and represent the physical and chemical mechanisms governing particulate matter variability.

Chapter 5

Conclusions

This Thesis investigated the variability and chemical composition of PM_{10} in the Po Valley using measurements collected within the LIFE PrepAIR monitoring network. The integration of different methodologies such as chemical characterization, meteorological analysis and data-driven modelling provided a comprehensive framework for interpreting particulate matter dynamics and variability in a region characterized by complex atmospheric conditions. The chemical analysis confirmed the dominant contribution of organic carbon ad secondary inorganic aerosols to PM_{10} mass, indicating the importance of secondary formation processes and regional atmospheric interactions. The mass closure results showed that the measured species account for the majority of the particulate mass, confirming the reliability of the chemical dataset used. Meteorological factors were found to influence PM_{10} variability. However, the moderate strength of the observed correlation highlights the complexity of the system, where PM_{10} concentration is controlled by multiple interacting drivers. The machine learning analysis showed that data-driven models can reproduce substantially PM_{10} variability, in particular when persistence effects are included through lagged variables. Ensemble tree methods proved effective in capturing the variability of integrated atmospheric variables influenced by multiple sources and processes. These results, consistently with the previous studies consulted, confirmed the potential of machine learning techniques not only as predictive tools, but also as complementary methods for a better understanding of complex atmospheric relationships that traditional statistical analysis has more difficulty in capturing. Despite these results, the study is constrained by the relatively limited temporal coverage of the dataset and by the restricted number of predictors available for modelling. Future research should therefore focus on using larger monitoring time series and integrating additional sources of information, as emissions inventories, satellite observations and chemical transport models. The combination of these approaches could significantly improve the understanding and prediction of particulate matter variability in the Po Valley. An increased under-

standing of the drivers of PM variability in the Po Valley is essential for effective mitigation strategies in one of the most polluted areas in Europe.

Chapter 6

Acknowledgements

I would like to sincerely thank my supervisors, Prof. Marco Ravina and Prof. Deborah Panepinto, for their constant availability, enthusiasm and valuable guidance throughout the development of this work.

Part of the data used in this study were elaborated within the PREPAIR – LIFE 15 IPE IT013 project – Action D6, whose contribution is gratefully acknowledged.

I would also like to thank my friends and my family for their constant support, encouragement and patience during this journey.

Chapter 7

IA declaration

Supporto editoriale:

Gli strumenti di IA generativa sono stati utilizzati esclusivamente per il supporto editoriale, come la correzione grammaticale, le traduzioni, i suggerimenti per la parafrasi o il miglioramento della leggibilità.

Perfezionamento delle idee:

Gli strumenti di IA generativa sono stati utilizzati per assistere il brainstorming delle idee e la sintesi della letteratura.

Generazione di contenuti e supporto analitico:

Gli strumenti di IA generativa non sono stati utilizzati per la generazione di contenuti o il supporto analitico.

Indipendentemente dal livello di coinvolgimento dell'IA, tutti i contenuti generati dall'IA sono stati rivisti criticamente, verificati e modificati per allinearsi al rigore accademico. Le citazioni e le argomentazioni sono state verificate in modo indipendente. L'autore si assume la piena responsabilità dell'accuratezza, dell'originalità e dell'integrità del lavoro finale.

Bibliography

Aas, W., Tsyro, S., Bieber, E., Bergström, R., Ceburnis, D., Ellermann, T., Fagerli, H., Fröhlich, M., Gehrig, R., Makkonen, U., Nemitz, E., Otjes, R., Perez, N., Perrino, C., Prévôt, A. S. H., Putaud, J.-P., Simpson, D., Spindler, G., Vana, M., & Yttri, K. E. (2012). Lessons learnt from the first EMEP intensive measurement periods. *Atmospheric Chemistry and Physics*, *12*, 8073–8094. https://doi.org/10.5194/acp-12-8073-2012

ARPA Lombardia. (2025,). *Il PM10 nei capoluoghi lombardi*.

ARPA Piemonte, C. M. d. T. –. D. A. e. V. A., Dipartimento Territoriale Piemonte Nord-Ovest, A. P. –, & Ambientali, A. P. – Dipartimento Rischi Naturali e. (2024). *Uno sguardo all'aria. Anno 2023: Relazione annuale sui dati rilevati dalla rete metropolitana di monitoraggio della qualità dell'aria* [Technical Report]. https://www.arpa.piemonte.it/sites/default/files/media/2024-10/Uno%20Sguardo%20all%27Aria%202023.pdf

Caserini, S., Giani, P., Cacciamani, C., Ozgen, S., & Lonati, G. (2017). Influence of climate change on the frequency of daytime temperature inversions and stagnation events in the Po Valley: Historical trend and future projections. *Atmospheric Research*, *184*, 15–23. https://doi.org/https://doi.org/10.1016/j.atmosres.2016.09.018

Direttiva (UE) 2024/2881 del Parlamento europeo e del Consiglio. (2024).

ENVI.info. (2024,). *Inquinamento atmosferico a Torino: il caso del processo smog*. https://www.envi.info/it/2024/07/09/inquinamento-atmosferico-a-torino-il-caso-del-processo-smog/

European Environment Agency. (2025). *Air quality status report 2025: Particulate matter*. https://www.eea.europa.eu/en/analysis/publications/air-quality-status-report-2025/particulate-matter

European Monitoring and Evaluation Programme (EMEP). (2026,). *EMEP – Cooperative Programme for Monitoring and Evaluation of the Long-range Transmission of Air Pollutants in Europe*.

Guo, F., Tang, M., Wang, X., Yu, Z., Wei, F., Zhang, X., Jin, M., Wang, J., Xu, D., Chen, Z., & Chen, K. (2022). Characteristics, sources, and health risks of trace metals in PM2.5. *Atmospheric Environment*, *282*, 119314. https://doi.org/10.1016/j.atmosenv.2022.119314

Jordan, T. B., Seen, A. J., & Jacobsen, G. E. (2006). Levoglucosan as an atmospheric tracer for woodsmoke. *Atmospheric Environment*, *40*, 5316–5321. https://doi.org/10.1016/j.atmosenv.2006.03.023

Julaha, K., Ždímal, V., Holubová Šmejkalová, A., Komínková, K., & Zíková, N. (2024). Boundary layer and mixing layer height: Models vs. ground-based measurements intercomparison. *Atmospheric Research*, *302*, 107897. https://doi.org/10.1016/j.atmosres.2024.107897

Kelly, F. J., & Fussell, J. C. (2012). Size, source and chemical composition as determinants of toxicity attributable to ambient particulate matter. *Atmospheric Environment*, *60*, 504–526. https://doi.org/10.1016/j.atmosenv.2012.06.039

Khan, B., Hays, M. D., Geron, C., & Jetter, J. (2011). Differences in the OC/EC Ratios that Characterize Ambient and Source Aerosols due to Thermal-Optical Analysis. *Aerosol Science and Technology*, *46*(2), 127–137. https://doi.org/10.1080/02786826.2011.609194

Kirešová, S., & Guzan, M. (2022). Determining the Correlation between Particulate Matter PM10 and Meteorological Factors. *Eng*, *3*(3), 343–363. https://doi.org/10.3390/eng3030025

Kulkarni, P., Baron, P. A., & Willeke, K. (Eds.). (2011). *Aerosol Measurement: Principles, Techniques, and Applications* (3rd ed.). John Wiley & Sons.

Lelieveld, J., Pozzer, A., Pöschl, U., Fnais, M., Haines, A., & Münzel, T. (2020). Loss of life expectancy from air pollution compared to other risk factors: a worldwide perspective. *Cardiovascular Research*, *116*(11), 1910–1917. https://doi.org/10.1093/cvr/cvaa025

LIFE Integrated Project PREPAIR. (2023,). *LIFE PREPAIR – Po Regions Engaged to Policies of Air*.

LIFE PREPAIR Project. (2023,). *Air Quality and Emission Evaluation*. https://www.lifeprepair.eu/index.php/azioni/air-quality-and-emission-evaluation/

LIFE PREPAIR Project Consortium - Action D. (2023). *Technical report* [Technical report]. https://www.lifeprepair.eu/

Masih, A. (2019). Machine Learning Algorithms in Air Quality Modeling. *Global Journal of Environmental Science and Management*, *5*(4), 515–534. https://doi.org/10.22034/gjesm.2019.04.10

Matta, E., Facchini, M. C., Decesari, S., Mircea, M., Cavalli, F., Fuzzi, S., Putaud, J.-P., & Dell'Acqua, A. (2002). Chemical mass balance of size-segregated aerosol in an urban area of the Po Valley, Italy. *Atmospheric Chemistry and Physics Discussions*, *2*, 2167–2208. https://acp.copernicus.org/preprints/2/2167/2002/acpd-2-2167-2002.pdf

Pachon, J. E., Weber, R. J., Zhang, X., Mulholland, J. A., & Russell, A. G. (2013). Revising the use of potassium (K) in the source apportionment of PM2.5. *Atmospheric Pollution Research*, *4*(1), 1–7. https://doi.org/10.5094/APR.2013.002

Putaud, J.-P., Van Dingenen, R., Alastuey, A., Bauer, H., Birmili, W., Cyrys, J., Flentje, H., Fuzzi, S., Gehrig, R., Hansson, H., Harrison, R., Herrmann, H., Hitzenberger, R., Hüglin, C., Jones, A., Kasper-Giebl, A., Kiss, G., Kousa, A., Kuhlbusch, T., … Raes, F. (2010). A European Aerosol Phenomenology – 3: Physical and Chemical Characteristics of Particulate Matter from 60 Rural, Urban, and Kerbside Sites across Europe. *Atmospheric Environment*, *44*(10), 1308–1320. https://doi.org/10.1016/j.atmosenv.2009.12.011

Raffaelli, K., Deserti, M., Stortini, M., Amorati, R., Vasconi, M., & Giovannini, G. (2020). Improving Air Quality in the Po Valley, Italy: Some Results by the LIFE-IP-PREPAIR Project. *Atmosphere*, *11*(4), 429. https://doi.org/10.3390/atmos11040429

Rybarczyk, Y., & Zalakeviciute, R. (2018). Machine Learning Approaches for Outdoor Air Quality Modelling: A Systematic Review. *Applied Sciences*, *8*(12), 2570. https://doi.org/10.3390/app8122570

Saarikoski, S., Sillanpää, M., Sofiev, M., Timonen, H., Saarnio, K., Teinilä, K., Karppinen, A., Kukkonen, J., & Hillamo, R. (2008). Sources of organic carbon in fine particulate matter in northern European urban air. *Atmospheric Chemistry and Physics*, *8*(20), 6281–6295. https://doi.org/10.5194/acp-8-6281-2008

Schwarze, P. E., Øvrevik, J., Låg, M., Refsnes, M., Nafstad, P., Hetland, R. B., & Dybing, E. (2006). Particulate matter properties and health effects: Consistency of epidemiological and toxicological studies. *Toxicology*, *228*(2–3), 224–232. https://doi.org/10.1177/096032706072520

Seinfeld, J. H., & Pandis, S. N. (2006). *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change* (2nd ed.). John Wiley & Sons.

Sharma, N., & Taneja, A. (2022). Chemical and biological components of atmospheric particulate matter and their impacts on human health and crops: A review. *Aerobiologia*, *38*(3), —. https://doi.org/10.1007/s10453-022-09749-4

Simpson, C. D., Dills, R. L., Katz, B. S., & Kalman, D. A. (2004). Determination of Levoglucosan in Atmospheric Fine Particulate Matter. *Journal of the Air & Waste Management Association*, *54*(6), 689–694. https://doi.org/10.1080/10473289.2004.10470945

Snider, G., Weagle, C. L., Martin, R. V., Donkelaar, A. van, Conrad, K., Cunningham, D., Gordon, C., Zwicker, M., Akoshile, C., Artaxo, P., Anh, N. X., Brook, J., Dong, J., Garland, R. M., Greenwald, R., Griffith, D., He, K., Holben, B. N., Kahn, R., … Liu, Y. (2016). Variation in global chemical composition of PM_{2.5}: emerging results from SPARTAN. *Atmospheric Chemistry and Physics*, *16*, 9629–9653. https://doi.org/10.5194/acp-16-9629-2016

Terzi, E., Argyropoulos, G., Bougatioti, A., Mihalopoulos, N., Nikolaou, K., & Samara, C. (2010). Chemical composition and mass closure of ambient PM10 at urban sites. *Atmospheric Environment*, *44*, 2231–2239. https://doi.org/10.1016/j.atmosenv.2010.02.019

Wang-Li, L. (2015). Insights to the Formation of Secondary Inorganic PM2.5: Current Knowledge and Future Needs. *International Journal of Agricultural and Biological Engineering*, *8*(2), 1–13. https://doi.org/10.3965/j.ijabe.20150802.1810

WHO Global Air Quality Guidelines: Particulate Matter (PM2.5 and PM10), Ozone, Nitrogen Dioxide, Sulfur Dioxide and Carbon Monoxide. (2021,).

Wilks, D. S. (2011). *Statistical Methods in the Atmospheric Sciences: 100* (3rd ed., Vol. 100). Academic Press. https://sunandclimate.wordpress.com/wp-content/uploads/2009/05/statistical-methods-in-the-atmospheric-sciences-0127519661.pdf

Xing Gao, H. W. H. Z., Can Qi. (2025). A Study of Theoretical Modeling for Scavenging Coefficients of Polydisperse Aerosols Removed by Rainfall. *Atmosphere*, *16*(6), 1–18. https://doi.org/10.3390/atmos16060634

Zhang, R., Wang, G., Guo, S., Zamora, M. L., Ying, Q., Lin, Y., Wang, W., Hu, M., & Wang, Y. (2015). Formation of Urban Fine Particulate Matter. *Chemical Reviews*, *115*(10), 3803–3855. https://doi.org/10.1021/acs.chemrev.5b00067