Politecnico di Torino

Master's Degree Thesis
in Physics of Complex Systems

Analysis of Air Quality in the Urban Area of Turin: Focus on Nitrogen Oxides and Particulate Matter

Supervisors
Prof. Sofia FELLINI
Prof. Luca RIDOLFI

Candidate
Pierpaolo CAREDDU

October 2024
A.Y. 2023/2024

Abstract

This thesis uses information collected from five monitoring stations between 2000 and 2022 to present a comprehensive database analysis of the air quality in the urban area of Turin, with a focus on nitrogen oxides and particulate matter. The goal is to demonstrate a methodology for approaching pollutant concentration databases, and assess the impact of regulatory measures on air quality, determining whether introduced environmental policies have influenced the concentrations of nitrogen oxides (NO, NO_2, NO_x) and particulate matter (PM_{10} and $PM_{2.5}$). The study uses publicly available datasets, that provide detailed daily and hourly measurements of pollutant concentrations from the Sistema Regionale di Rilevamento di ARPA Piemonte (database accessible at https://aria.ambiente.piemonte.it/#/qualita-aria/dati). The work on the database begins with characterizing the different measurement methods employed at each station, assessing their reliability and comparability. A geographical map illustrates the spatial distribution of the stations, complemented by an evaluation of data coverage that highlights measurement variability over time. This analysis gives an overview on when each station began operating and highlights key periods of reliable data collection, providing insights into the functionality and data quality of each station. The raw data were reorganized and cleaned to ensure accuracy, underling inconsistencies and filling data gaps where possible. This process helped to highlight significant trends, gaps, and possible errors in the measurements over the years. In the main part of the thesis, Long-term trends in pollutant levels are analyzed using both simple linear trend analysis and advanced techniques such as STL (Seasonal and Trend decomposition using Loess), which helps firstly to provide a general view of the pollution levels, and secondly to identify trends, seasonal patterns, and irregular changes in the data that are not immediately visible. Additionally, a preliminary exploration of the potential use of the cleaned dataset in the Urban Atmospheric Pollutant Dispersion Model SIRANE, is provided. This stresses its possible future application in studies focused on the area of Turin. The results of this thesis present a comprehensive view of air quality database, providing insight on data reliability and measurement trends and variability, across Turin over the first two decades of the 21st century. The analysis evidences an improvement in air quality condition, with decreasing trends in nitrogen oxides and particulate matter concentrations. This reduction is mainly attributed to the implementation of stringent environmental policies, demonstrating the effectiveness of the regulations introduced by the European Parliament and concretely adopted by the city of Turin. This work not only provides a clear perspective on the historical and current state of urban air quality in Turin but also provides a valuable reference for future air quality comparisons and air pollution management in the metropolitan area.

Table of Contents

Chapter 1

Introduction

1.1 Context

Air pollution is becoming a severe issue for public health and environmental sustainability, especially in urban areas worldwide. The city of Turin, with approximately 850,000 residents in its urban area, is the fourth largest city in Italy and one of the most important economic and industrial center, especially until the first decade of the 21st century with the FIAT factory. Despite being a green hub of culture and history, the city faces persistent pollution problems, partly due to its geographical layout. The surrounding Alps and Appenine mountains, from the west to the north, contribute to poor air circulation, especially in winter. During this season indeed, stable atmospheric conditions occur for example when warm air traps colder air near the ground, therefore preventing pollutants from dispersing upward and causing persistent smog. This, combined with the city's traffic and industrial activities, worsens air quality, significantly affecting the environment and living condition. Nitrogen oxides (NO, NO_2, NO_X) and particulate matter (PM_{10}, $PM_{2.5}$) are among the most critical air pollutants, primarily emitted from vehicular traffic, industrial processes, and domestic heating. Nitrogen oxides are particularly harmful because, beyond their direct toxicity, they contribute significantly to the formation of photochemical smog and secondary pollutants, such as acids and organic derivatives. PM_{10} and $PM_{2.5}$ refer to particulate matter with diameters of 10 microns and 2.5 microns, respectively. These fine particles consist of dust, smoke, and various other substances that due to their small size, can penetrate deeply into the respiratory system, causing significant health risks such as respiratory diseases and cardiovascular problems which contribute to decrease the life expectancy. Not surprisingly their concentration are regulated by appropriate national and European laws and standards (there is an upper limit of "overruns", in days in a year, of these thresholds, beyond which European sanctions are triggered).

The constant monitoring of these pollutants is crucial for understanding air quality and evaluating the impact of regulatory measures. In addition, it should be recalled that the interplay between pollution and weather conditions makes managing urban air quality particularly challenging, since pollutants can alter local weather, while weather affects pollutant dispersion. The SIRANE model briefly introduced in this thesis, which simulates pollution dispersion in urban areas by considering various factors including meteorological conditions, offers a way to better understand how pollutants spread across the city. By integrating weather data, the model provides predictions that can guide urban planning and policy decisions aimed at mitigating pollution.
This thesis goes on to investigate at Turin's air quality by analyzing publicly available datasets to understand pollution trends and patterns and this introductory chapter sets the stage by presenting the broader context of the study, highlighting the complexities of working with large datasets, and introducing the main topics that will be examined in detail throughout the thesis.

1.2 Objectives

The main objective of this thesis is to answer the question, *how has pollution in Turin changed in recent years?* In particular, by analyzing long-term data on the most important pollutants present at the urban level, the study primarily seeks to uncover trends and irregular changes in air quality, which could provide insights into the effectiveness of current regulations or correlate changes in air quality with climate change. This will also provide a comprehensive overview of the database that collect all measurement data from the monitoring stations within the city, allowing to study the behavior of the monitoring stations during their working period, highlighting possible issues in detection. This research aims also to evaluate the impact of environmental policies and create a starting point for further analysis on understanding how weather conditions influence pollution levels.

1.3 Implemented Methodology

The methodology used involves mainly techniques of data cleaning, statistical analysis methods and modeling, to explore the air quality database in order to obtain several information, including trends and patterns. Data preprocessing was performed using Python Pandas and Julia DataFrames to clean and organize the big datasets, addressing missing values and making data format simpler and organized. The geographic information system tool QGIS was used to visualize the spatial distribution of the monitoring stations in the urban area, showing their proximity to industrial or high trafficked area. It also facilitated the accurate

representation of the vector layer used by the SIRANE model to define the gas dispersion network and allowed visualization of pollutant concentrations calculated by the model, directly on the streets using heatmap layers. A simple trial was conducted using SIRANE to assess pollutant dispersion, taking into account factors such as meteorological conditions and urban layout, with the aim of evaluating its reliability within Turin's urban environment by comparing predicted concentrations with measured data. Advanced statistical techniques, such as *Seasonal and Trend decomposition using LOESS (STL)*, were employed to separate long-term trends from seasonal variations, particularly suited to recurring time patterns. This approach made it possible to emphasize certain aspects of the data and make the analysis more comprehensive, helping to evaluate the performance of monitoring stations and highlighting significant trends or particularly pollution period.
This thesis is organized as follows. In the "Data Source" chapter, a complete explanation and analysis of the databases public available for nitrogen oxides and particulate measures is reported. In the "Measure Comparison" chapter, the analysis of the correlation between different measurements is presented. The analysis of the air quality and trend of concentration of pollutant is reported in the "Temporal Analysis" chapter. Results are presented at the end of each chapter and discussed in the "Conclusion" chapter.

Chapter 2

Data Source

2.1 Data Overview

In this section is provided a compressive view of the publicly available database of air quality data collected by different monitoring stations, located in the municipality of Turin, and operating over the past two decades. The database offers the detection in a large spectrum of a multitude of pollutant gases and particulate concentrations. However the discussion focuses on the data collected in the urban area of Turin, in particular on the historical series of NO, NO_2, NO_X, and PM measure, which based on an initial analysis, these are the most standard measures, mostly detected since the start of the 21st century. The data span 23 years, from 2000 to 2022, providing a very long time series data which could be helpfully to understand and give a clear vision of how has air quality changed over time.

2.2 Data Acquisition and Preprocessing

The data used in this study were provided by the *Sistema Regionale di Rilevamento di ARPA Piemonte*, accessible through the air quality data portal [9]. In the site the pollutant measurement data can be downloaded as a comma-separated values (.csv) file, after selecting the pollutant type, the monitoring station, the year of the measurements and other parameters. However the downloading process of all files from the portal can be intricate and slow, so a formal request can be presented in order to gain access to the whole database, simplifying the collection process.
For this thesis, a formal request was submitted, and a large portion of the full database was provided, covering the period 2000-2022. Years 2023 and 2024 were not included, as they had not yet been certified by ARPA at the time data were obtained for this analysis. The data for $PM_{2.5}$ does not cover the period from 2000 to 2004, as records for those years were not made available.

2.2.1 Structure of the Data Files

The database is exported into multiple files, each organized following a specific naming convention, which reflects the pollutant type and year of measurement:

- `Export_2000_NO_torino.csv`: Contains NO measurements for the year 2000.
- `Export_2001_NO2_torino.csv`: Contains NO_2 measurements for the year 2001.
- `Export_2022_PM10_torino.csv`: Contains PM_{10} measurements for the year 2022.

Each of these files contains pollutant concentration measurements collected from several stations within the municipality of Turin.

In addition, an extra file `stazioni.csv` containing a description of each monitoring station was provided. The information in this file is particularly useful for understanding and classifying the measurements collected in each data file, and the key columns are reported in the Table 2.1.

Table 2.1: Key Columns and Description from the `stazioni.csv` File

Column Name	Description
codice_istat_comune	ISTAT code identifying the municipality to which the station belongs.
progr_punto_com	Identifier of the monitoring point station.
denominazione	Name or code of the station.
data_inizio	Start date of the station's detection.
data_fine	End date of the station's operation, when present.
indirizzo_localita	Address where the station is situated.
utm_x	UTM coordinate (x) for geographic location.
utm_y	UTM coordinate (y) for geographic location.
quota_stazione	Altitude of the station (if available).

The CSV files containing the measures of concentration are organized in different lines, and the key columns are reported in Table 2.2.

Table 2.2: Description of the Most Important Columns in the Measurements Files

Column Name	**Description**	**Possible Values**
rete_monitoraggio	Monitoring network.	13 (Turin)
codice_istat_comune	Municipality where the station is located.	1272 (Turin area), 2121
progr_punto_com	Specific monitoring station within the municipality.	822, 803(Torino-Consolata)
id_parametro	Identifier for the pollutant being measured.	21 (NO), 4 (NO_2), 22
valore_validato	Measured concentration value	[45.6, 32.8, 380.2]
flag_gestore_sistema	Validation status by ARPA.	0 (validated), 1 (invalidated)
datetime	Date and hour of the measurement (YYYY-MM-DD HH:MM).	2000-01-01 00:00

2.2.2 Data Organization

Before preprocessing the data, a pre-filtering step was necessary to reduce the dataset's size and make it more manageable. The raw dataset, as stated before, contains measurements from various monitoring stations distributed across different municipalities; so a previous step is needed in order to keep only the data related to monitoring stations within the municipality of Turin.
This shifts the focus only on the data relevant to this thesis analysis. In particular, five monitoring stations were identified as being within Turin's urban area: Torino-Consolata, Torino-Rebaudengo, Torino-Lingotto, Torino-Grassi, and Torino-Rubino. An explicit overview of these stations will be provided later in the chapter (subsection 2.7).
The organization process also involves aggregating each pollutant—NO, NO_2, NO_X, PM_{10}, and $PM_{2.5}$—all measurement from 2000 to 2022 into a single dataset for each pollutant, maintaining the original structure.

2.2.3 Data Preprocessing

The aim of preprocessing is to have easily accessible data and ensure quality and reliability of the measurements all their parameters, before conducting further analysis.

Methodology

The preprocessing step involved the exploration of each datafile for each pollutant type, the understanding of completeness of the time series (to ensure the `datetime` column was present and continuous), the verification of the correct labeling, and the identification of all types of measures present. Of critical importance was the validity check conducted on each measurement, identifying missing and unrealistic range values. This were classified based on a review of the `flag_gestione_sistema` status (na value introduced) as follows:

- **Valid Measurements**: Correctly recorded measurements, flagged as valid (`flag_gestore_sistema = 0`).
- **Invalid Measurements**: `flag_gestione_sistema` = 1, indicates measure present in the time series but not valid (e.g., negative or error values).
- **Missing Measurements**: Values (`valore_validato = na`) flagged as missing (`flag_gestore_sistema = na`), this ensure a clear distinction from invalid data.

Summary

The table 2.3 below compares the datasets of NO, NO_2, NO_X before and after preprocessing, showing the improvements in flag accuracy:

Table 2.3: Comparison of Flags Before and After Preprocessing

Dataset	Measurement Type	Before Correction	After Correction
NO	Correct Invalid Flags (set to 1)	0	8
	Correct Missing Flags (set to NaN)	1,909	41,435
	Correct Valid Measurements (set to 0)	702,269	702,269
	Incorrect Valid Flags	8	0
	Incorrect Missing Flags	39,526	0
	Incorrect Invalid Flags	0	0
NO_2	Correct Invalid Flags (set to 1)	0	0
	Correct Missing Flags (set to NaN)	1,948	48,295
	Correct Valid Measurements (set to 0)	695,417	695,417
	Incorrect Valid Flags	0	0
	Incorrect Missing Flags	46,347	0
	Incorrect Invalid Flags	0	0
NO_X	Correct Invalid Flags (set to 1)	0	1,062
	Correct Missing Flags (set to NaN)	1,906	49,649
	Correct Valid Measurements (set to 0)	690,769	690,769
	Incorrect Valid Flags	1,062	0
	Incorrect Missing Flags	47,743	0
	Incorrect Invalid Flags	0	0

The *Correct Invalid Flags (set to 1)* and *Incorrect Missing Flags (set to 1)* highlight the most important quantities, providing an idea of how was the quality of the data before and after the preprocessing, showing respectively how many invalid measurements are correctly flagged, and how many missing values are not correctly flagged. This methodology significantly improved the accuracy of the flags, and it solves the problem of misclassification of missing and invalid measures. The adjustments made a solid data base, ensuring that all measurements are correctly categorized, and can be safely used for subsequent analysis.

The preprocessing also facilitated the identification of the sampling frequency for each pollutant concentration measurement (directly inferred from the data), distinguishing between measurements collected on an hourly or daily basis.
Additionally, this preprocessing phase uncovered that particulate matter (PM_{10} and $PM_{2.5}$) data contained multiple types of measurements. These varieties were identified by different `id_parametro` values, corresponding to different instruments or methods for measuring PM. The discovery led to further investigation which will be detailed in the Chapter 3 Measure Comparison, which report the analysis of the consistency between different types of particulate matter measurements.

2.3 Sampling Frequency of Measurements

Upon reviewing the data, it was observed that NO, NO_2 and NO_X measurements were consistently collected on an hourly basis, with the concentrations expressed in $\mu g/m^3$. As mentioned earlier, it was noticed that for the same monitoring station the presence of multiple particulate measurements (PM10, PM2.5) that differ based on the type and instruments used for the measurement. In particular, certain monitoring stations (e.g., Torino-Lingotto) offer more specific detection and also report concentrations of elements dispersed into the atmosphere, nevertheless these measurements will not be considered during the subsequent analysis.
This diversity in instrumentation has a direct impact on the sampling frequency, which can vary from hourly to daily. Tables 2.4 and 2.5 report all types of measurements present in the PM_{10} and $PM_{2.5}$ datasets, grouped by sampling frequency, and with a brief description.

Table 2.4: Summary table of PM10 measurements.

PM10 Measurements		
id_parameter	**name**	**u.m.**
	Sampling Frequency: Daily	
PM10_GAV	PM10 - High Volume	µg/m³
PM10_GBV	PM10 - Low Volume	µg/m³
PM10_B	PM10 - Beta	µg/m³
PM10_B_p	PM10 - Beta - n.2	µg/m³
PM10_BD	PM10 - Beta (daily average)	µg/m³
PM10GAV1	PM10 - High Volume - n.2	µg/m³
	Sampling Frequency: Hourly	
PM10_TCON	PM10 - Teom (correction factor + 50% -15)	µg/m³
PM10_CP	PM10 - from particle counter	µg/m³
PM10_BH	PM10 - Beta (hourly average)	µg/m³
PM10_N	PM10 - Nephelometer	µg/m³
PM10_T	PM10 - Teom	µg/m³

Table 2.5: Summary table of PM2.5 measurements.

PM2.5 Measurements		
id_parameter	**name**	**u.m.**
	Sampling Frequency: Daily	
PM2.5	PM2.5 - Low Volume	µg/m³
AS_PM2.5	Arsenic in PM2.5	ng/m³
BAAPM2.5	Benzo(a)anthracene in PM2.5	ng/m³
BP_PM2.5	Benzo(a)pyrene in PM2.5	ng/m³
BJKPM2.5	Benzo(b+j+k)fluoranthene in PM2.5	ng/m³
CD_PM2.5	Cadmium in PM2.5	ng/m³
NI_PM2.5	Nickel in PM2.5	ng/m³
PB_PM2.5	Lead in PM2.5	µg/m³
PM2.5_BD	PM2.5 - Beta (daily average)	µg/m³
PM2.5_B	PM2.5 - Beta	µg/m³
	Sampling Frequency: Hourly	
PM2.5_CP	PM2.5 - from particle counter	µg/m³
PM2.5_BH	PM2.5 - Beta (hourly average)	µg/m³

For each type of measurement, it was verified that sampling frequency continuity was maintained throughout the period 2000-2022.

2.4 Coverage Analysis

An important part of data preparation for analysis is evaluating the completeness and continuity of the measurements over time. The coverage analysis assesses the extent in which each pollutant was monitored across the selected stations in Turin. This ensures that sufficient data is available to support a reliable trend analysis and identifies any gaps that may need to be taken into account in a second phase of the analysis. The following sections present a detailed review of the data coverage for nitrogen oxides and particulate matter. For each pollutant group, valid and invalid measurements are visualized across all selected monitoring stations, providing a simple and effective understanding of the robustness and reliability of the dataset.

2.4.1 Methodology

To determine the temporal coverage of the measurements, the following steps were taken:

- Each pollutant data was filtered and sorted by `datetime` for each available detection station, and classified by measurement type.
- The presence and validity of data were assessed daily. For daily measurements, a day was considered valid if at least one valid measurement (flag_gestore_sistema = 0) was recorded. For hourly measurements, a day was considered valid if at least **min_valid_hours** = 12 valid measurements were recorded in that day.

The parameter **min_valid_hours** is introduced to ensure a consistent methodology for evaluating the temporal coverage of hourly measurements by defining a minimum threshold of valid data needed to consider a day reliable. The choice of a 12-hour threshold reflects a balance between maintaining a sufficient number of valid days and ensuring that the daily averages accurately represent true daily conditions.
Since this threshold impacts not only the temporal coverage but also the reliability of subsequent analyses, it is crucial to further evaluate its appropriateness. A complete reasoning about the threshold is provided in subsection 2.6, to confirm that this choice offers an optimal trade-off between data availability and representativeness.

2.4.2 Nitrogen Oxides (NO, NO_2, NO_X) Measurements

Figure 2.1 highlights the time series during which the measurements were available (in grey) and valid (colored) for each station for the entire period between 2000 to 2022.

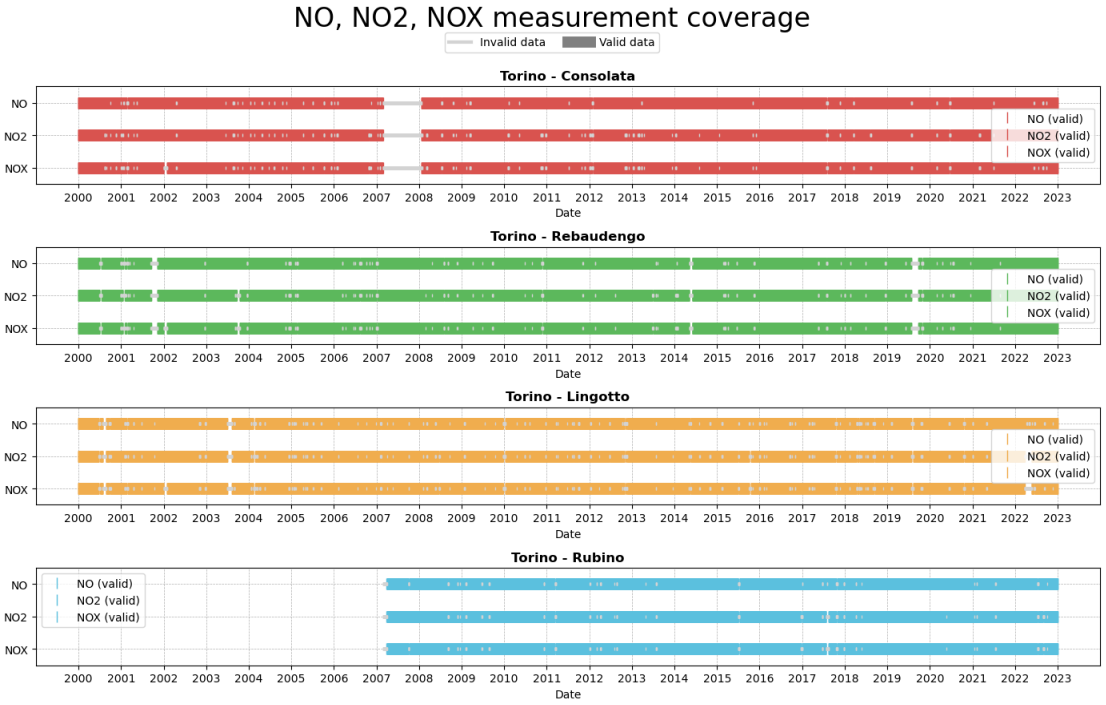

Figure 2.1: Nitrogen Oxides Measurement Daily Coverage for All Available Stations.

From an initial crude graphical view, it is noticeable that most stations exhibit a consistent and substantial amount of valid data. The station Torino-Rubino only started its measurements in 2007, and particularly Torino-Rebaudengo and Torino-Lingotto have good overall coverage throughout all measured years. However, some gaps in the data are also evident, e.g, for the station Torino-Consolata between 2007 and 2008, which require further investigation that will be addressed in the Gap Analysis section.

Important Note: One important observation is the complete absence of NO, NO_2, and NO_X measurements for the Torino-Grassi station. As a result, the following analysis for nitrogen oxides will not include this station.

2.4.2.1 Summary of Coverage Analysis of Nitrogen Oxides

In the Table 2.6 are reported the percentage of the valid measurement days relative to both the number of days with any measurement and over the total period between 2000 and 2022.

Table 2.6: Nitrogen Oxides Measurement Daily Coverage Summary.

Coverage percentage of NO, NO_2, and NO_X measurements		
Gas	**On present days (%)**	**On total period (%)**
Torino-Consolata		
NO	99.40	94.23
NO_2	99.10	92.95
NO_X	99.11	92.44
Torino-Rebaudengo		
NO	99.32	95.29
NO_2	99.24	94.57
NO_X	99.25	94.08
Torino-Lingotto		
NO	98.63	94.08
NO_2	98.60	93.20
NO_X	98.57	92.56
Torino-Rubino		
NO	99.44	67.09
NO_2	99.43	66.66
NO_X	99.41	66.58

As anticipated from the graphical result, the overall coverage is very high across all stations, with percentages above 98% for present days. The station Torino-Rubino exhibit a lower coverage over the total period, and this is expected due to the later start of data collection at that site.

2.4.3 Particulate Matter (PM_{10}, $PM_{2.5}$) Measurements

The results, illustrated in Figures 2.2 and 2.3, highlight the periods during which PM_{10} and $PM_{2.5}$ measurements were available (in grey) and valid (colored) for each station. The data for each station include various measurement types with different sampling frequencies. An initial view shows that data are present for all five stations in the urban area of Turin.

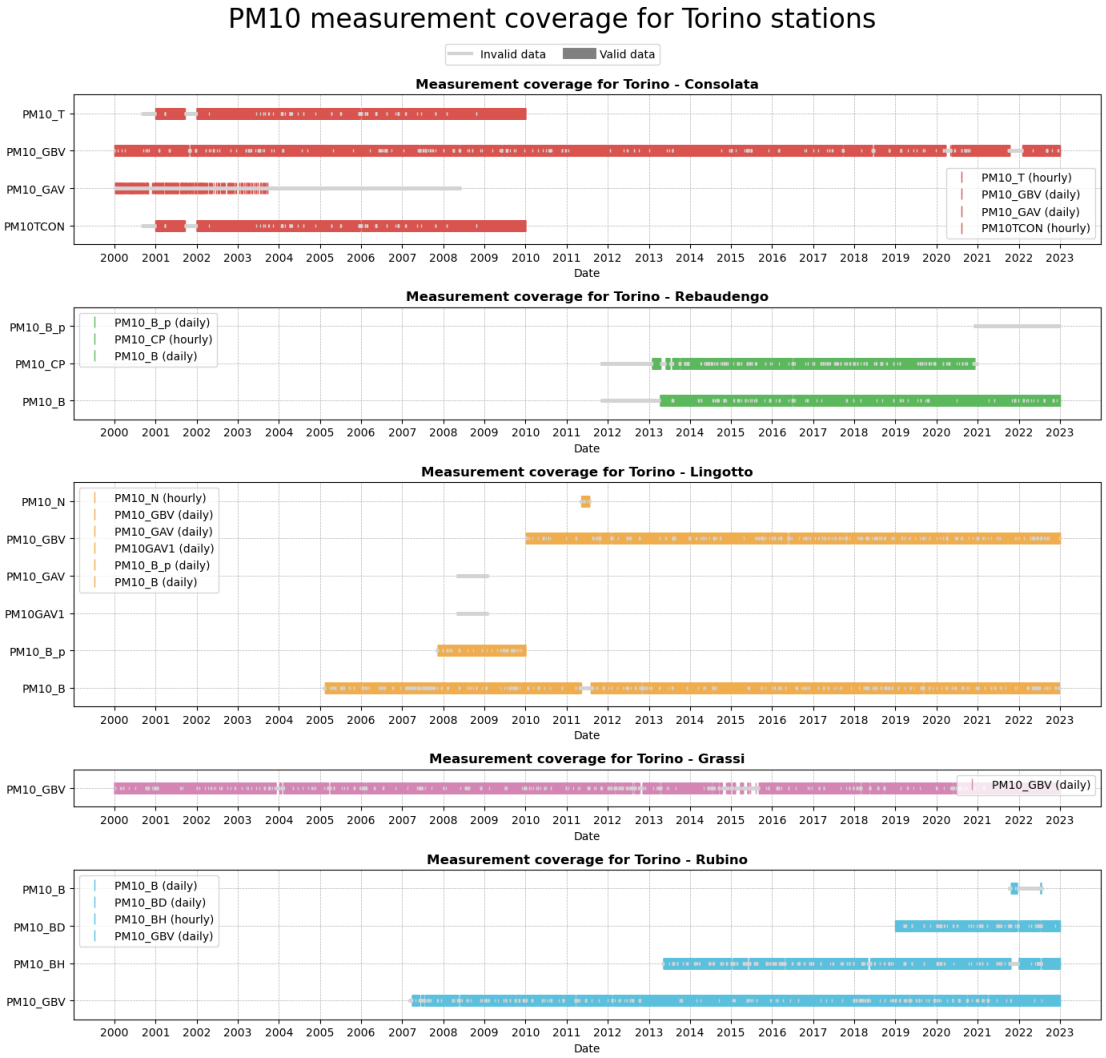

Figure 2.2: PM10 Measurement Daily Coverage for All Stations

When examining the PM10 measurements Torino-Consolata has extensive data for PM10_GBV from 2000 to 2022, despite some gaps in the last years. PM10_T and PM10_TCON are present for nearly the same period (2001-2010), instead PM10_GAV data are present only up to the end of 2003. Torino-Rebaudengo shows consistent coverage for PM10_B from 2013 to 2022, and for PM10_CP from 2013 to 2022, whereas the measures of PM10_B_p are all shown to be invalid.

At Torino-Lingotto, PM10_B data span from 2005 to 2022, though there is a small evident gap at the middle of 2011. The time period covered by the combined PM10_GBV and PM10_B_p datasets spans from 2008 to 2022. Torino-Grassi maintains almost continuous data for PM10_GBV from 2000 to 2022. Torino-Rubino has a coverage for PM10 data from 2007 onwards, with valid data especially for PM10_GBV, PM10_BH, and PM10_BD. PM10_B has valid data for a small period mostly concentrated around 2022.

For PM10 measurements, the Torino-Consolata and Torino-Grassi stations are the most reliable, since they provide the most complete records from 2000 to 2022. For detailed and long-term studies, it is best to focus on PM10_T and PM10_GBV at Torino-Consolata, PM10_B at Torino-Rebaudengo, PM10_B and PM10_GBV at Torino-Lingotto, PM10_GBV at Torino-Grassi, and PM10_B and PM10_BH at Torino-Rubino. These measurements offer the most consistent data coverage and can help in a robust analysis.

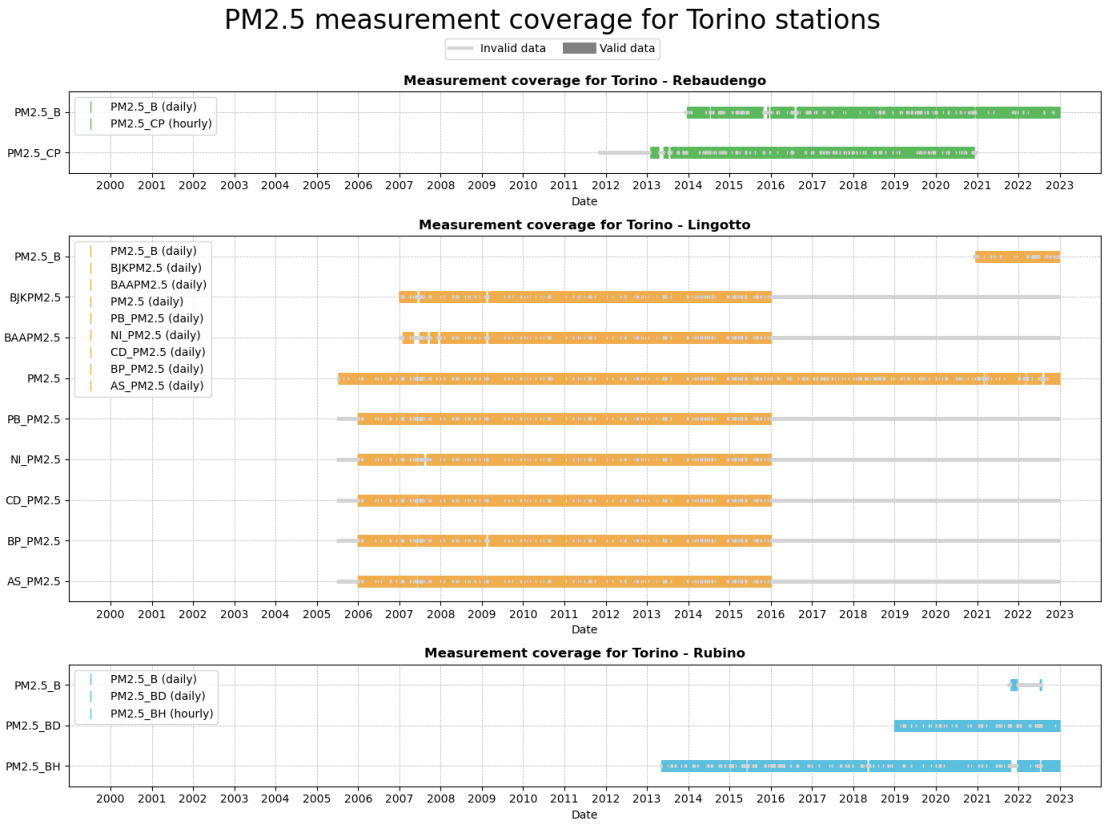

Figure 2.3: PM2.5 Measurement Daily Coverage for All Stations

When examining the PM2.5 fine particulate measurements, at the Torino-Rebaudengo station data collection began at the end of 2011 and has been ongoing since then, but valid measurements are present only from 2013. The Torino-Lingotto station stands out with a continuous record of PM2.5 measurements spanning from 2005 to 2022 with several types of measurements (including atmosphere polluting elements). Similarly, the Torino-Rubino station started gathering PM2.5 data in 2013, maintaining consistency over the years.

The results show that different types of PM measurements for the same station have similar data coverage intervals. This observation suggests that it might be possible to integrate these different measurements temporally, and extend the temporal coverage of particulate matter data for a single station. Further analysis will focus on understanding the similarity between these different measurements to facilitate their integration (chapter 3).

2.4.3.1 Summary of Coverage Analysis of PM

To better identifying the measurement types that are most frequently recorded across all stations, in Table 2.7, Table 2.8 are reported the percentage coverage of PM10, PM2.5 measurements. (For completeness PM2.5 element dispersion measurements coverage is reported in Appendix in Table B.1).
Each table includes two percentages for each measurement: the percentage of valid data when the measurement is present, and the percentage of valid data over the entire 23-year period from 2000 to 2022 (8401 days) which reflects the overall presence and completeness of the measurement across the study period.

Table 2.7: Coverage percentage of PM10 measurements.

Coverage percentage of PM10 measurements		
Measure	**On present days (%)**	**On total period (%)**
Torino-Consolata		
PM10TCON	90.14	36.58
PM10_GAV	9.04	3.31
PM10_GBV	93.66	93.66
PM10_T	90.17	36.59
Torino-Rebaudengo		
PM10_B	83.11	40.35
PM10_CP	74.89	29.85
Torino-Lingotto		
PM10_N	64.13	0.70
PM10_GBV	92.25	52.14
PM10_B_p	91.92	8.67
PM10_B	89.41	69.63
Torino-Grassi		
PM10_GBV	87.50	87.50
Torino-Rubino		
PM10_B	46.34	0.68
PM10_BD	89.39	15.55
PM10_BH	85.87	36.10
PM10_GBV	91.44	62.97

Table 2.8: Coverage percentage of PM2.5 fine particulate measurements.

Coverage percentage of PM2.5 fine particulate		
Measure	**On present days (%)**	**On total period (%)**
Torino-Rebaudengo		
PM2.5_CP	75.04	29.91
PM2.5_B	86.65	34.22
Torino-Lingotto		
PM2.5	90.97	69.23
PM2.5_B	86.99	7.88
Torino-Rubino		
PM2.5_BH	86.75	36.47
PM2.5_BD	90.35	15.71
PM2.5_B	47.15	0.69

These results align with the coverage analysis graphical results, highlighting that PM10_GBV and PM10_B are the most consistently recorded measurements for PM10 across the various stations. Their frequent recording and high percentages of valid data suggest they are good indicators of PM10 levels for our study case. Similarly, for PM2.5 the measurements PM2.5_BH and PM2.5_B emerge as the most reliable and offer substantial coverage across the observation period at their respective stations.

2.4.4 Final Comment on the Coverage Analysis

The coverage analysis reveals that most stations provide reliable and continuous data, particularly with regard to measurements of nitrogen oxides, offering strong coverage across the full study period. For the particulate matter, the most reliable measurements come from Gravimetric and Beta methods, which provide the longest coverage over the years. Torino-Lingotto distinguishes as the station with the most diverse range of PM measurements, capturing data from multiple measurement methods. A key finding of the analysis is the complete lack of NO, NO_2 and NO_X data for Torino-Grassi, meaning this station will not be considered in the nitrogen oxides analysis. Additionally, as shown in the coverage plot, for $PM_{2.5}$ the data is only available for Torino-Rebaudengo, Torino-Lingotto, and Torino-Rubino stations.

Important Note: After conducting the coverage analysis, it was necessary to select the measurements with the most reliable and complete time series data for each station to ensure a cleaner and more straightforward analysis of particulate matter, primarily aimed at obtaining a comprehensive trend analysis. Only measurements that demonstrated adequate data coverage across the study period were maintained for further analysis, including the gap analysis.
The following measurements were chosen for each station:

- **Torino-Consolata**: PM10_T, PM10_GBV
- **Torino-Rebaudengo**: PM10_B, PM10_CP, PM2.5_B, PM2.5_CP
- **Torino-Lingotto**: PM10_B, PM10_GBV, PM2.5, PM2.5_B
- **Torino-Grassi**: PM10_GBV
- **Torino-Rubino**: PM10_B, PM10_GBV, PM2.5_B

2.5 Gap Analysis

This section focuses on quantifying missing periods in the measurement data for nitrogen oxides and particular matter, already partially identified by the coverage analysis. A complete analysis is reported, identifying critical periods, period with longest gaps and the distribution of all continuous missing measurements for each stations and type.
This analysis is crucial for assessing the quality and consistency of the data and provides an important contribution to the coverage analysis. It is particularly useful to identify whether the dataset contains many short gaps that could potentially be addressed, thereby increasing its robustness, particularly for further trend analysis.

2.5.1 Nitrogen Oxides (NO, NO_2, NO_X) Measurements

The gap analysis for nitrogen oxides across the four available stations is summarized in Figure 2.4. An interesting view of the distribution of the gaps in the data is given through the gap classification into six duration bins: 1 hour - 24 hours, 1 day - 7 days, 1 week - 4 weeks, 1 month - 3 months, 3 months - 9 months, and gaps greater than 9 months. This helps quantify short-term incomplete as well as extended periods of data loss. A more detailed view of the 1-24 hour bin is provided on the right side of the figure, which provides a more in-depth view of the duration and frequency of these shorter gaps.

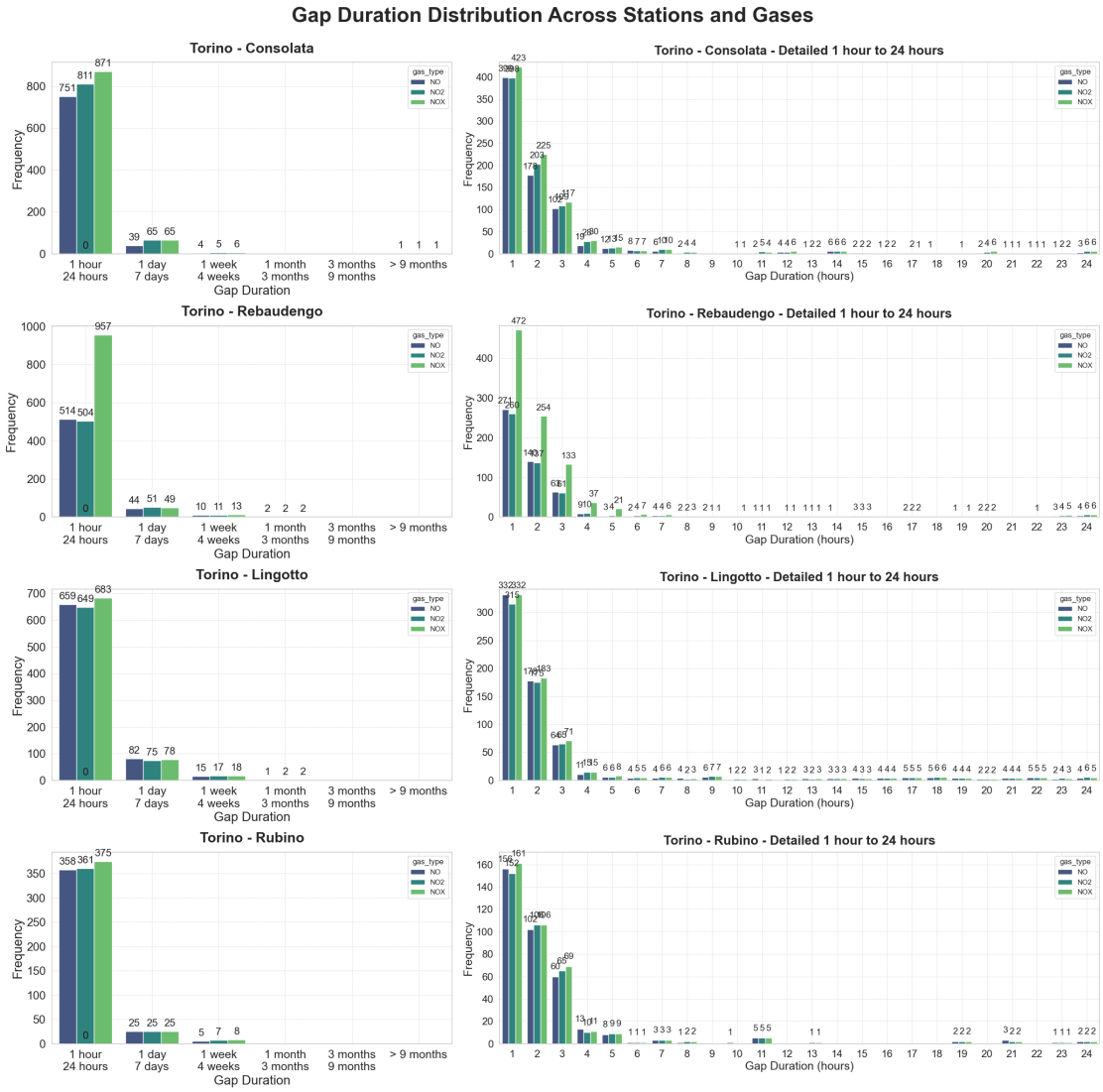

Figure 2.4: Gap Duration Distribution for Nitrogen Oxides Across Stations

Important Observations:

- Short-term gaps (1 to 24 hours) dominate, indicating frequent brief disruptions in data collection.
- In the 1 to 24-hour range, most gaps last between 1 and 3 hours.
- Torino-Consolata and Torino-Rebaudengo have slightly higher occurrences of gaps, especially for NO_X.
- Longer gaps (more than 1 week) are rare but do occur occasionally.

- For a single station the plots shows that gaps are very similar across all gases, with slight variations in frequency.

This visualization helps understand the extent and pattern of missing data, which is important for ensuring the accuracy of trend analyses in the study.

2.5.1.1 Summary of Gap Analysis of Nitrogen Oxides

The gap analysis for nitrogen oxides across the four available stations is summarized in Tables 2.9, 2.10, and 2.11.

Table 2.9: Gap Duration Summary for NO

Station	Total	Most Freq.	Longest Gap	% Gaps
Torino-Consolata	795	1	ca. 10 mo (2007-02 to 2008-01)	1h-24h: 94.47%, >3mo: 0.13%
Torino-Rebaudengo	570	1	ca. 2 mo (2019-07 to 2019-09)	1h-24h: 90.18%, >3mo: 0.35%
Torino-Lingotto	757	1	ca. 1 mo (2003-06 to 2003-08)	1h-24h: 87.05%, >3mo: 0.13%
Torino-Rubino	388	1	ca. 4 wk (2007-03 to 2007-03)	1h-24h: 92.27%, >3mo: 0.0%

Table 2.10: Gap Duration Summary for NO_2

Station	Total	Most Freq.	Longest Gap	% Gaps
Torino-Consolata	882	1	ca. 10 mo (2007-02 to 2008-01)	1h-24h: 91.95%, >3mo: 0.11%
Torino-Rebaudengo	568	1	ca. 2 mo (2019-07 to 2019-09)	1h-24h: 88.73%, >3mo: 0.35%
Torino-Lingotto	743	1	ca. 2 mo (2022-03 to 2022-05)	1h-24h: 87.35%, >3mo: 0.27%
Torino-Rubino	393	1	ca. 4 wk (2007-03 to 2007-03)	1h-24h: 91.84%, >3mo: 0.0%

Table 2.11: Gap Duration Summary for NO_X

Station	Total	Most Freq.	Longest Gap	% Gaps
Torino-Consolata	943	1	ca. 10 mo (2007-02 to 2008-01)	1h-24h: 92.36%, >3mo: 0.11%
Torino-Rebaudengo	1021	1	ca. 2 mo (2019-07 to 2019-09)	1h-24h: 93.73%, >3mo: 0.35%
Torino-Lingotto	781	1	ca. 2 mo (2022-03 to 2022-05)	1h-24h: 87.45%, >3mo: 0.23%
Torino-Rubino	408	1	ca. 4 wk (2007-03 to 2007-03)	1h-24h: 91.96%, >3mo: 0.0%

Figure 2.5: Gap Duration Summary by Station and Gas Type: NO, NO_2, and NO_X

Of particular importance are the percentage of gaps falling within the 1-24 hour range and those exceeding 3 months. The majority of gaps for all gases are short-term (lasting between 1 and 24 hours), making up over 85% of gaps at each station, with the most frequent being of 1 hour. These type of gaps do not represent a big challenges in the further data analysis since, frequent short missing data can often be addressed through standard data-cleaning techniques such as linear interpolation.

A key observation that results from the summary tables is that gaps in the data for different measurements at the same station often happen during the same periods.

These gaps could be due to temporary issues or malfunctions at the monitoring stations during those times. For example, the longest gap recorded occurred at Torino-Consolata, where NO, NO_2, and NO_X data were missing for approximately 10 months between February 2007 and January 2008.

2.5.2 Particulate Matter (PM_{10}, $PM_{2.5}$) Measurements

For particulate matter, a similar gap analysis was performed as for nitrogen oxides, restricting the analysis only at the measurement type with the longest time series as stated at the end of coverage analysis. This ensure the analysis to be consistent and reliable with the subsequent trend analysis.

The selected measurement types for both PM_{10} and $PM_{2.5}$ are: PM10_T, PM10_GBV, PM10_B, PM10_CP, PM2.5_B, PM2.5_CP, and PM2.5.

In order to maintain uniformity in the analysis, the gap analysis for particulate measurements was performed on a daily basis. For measures with hourly sampling, daily averages were calculated using a threshold of **min_valid_hours** = 12, as done previously during the coverage analysis.

The distribution of daily gaps for particulate matter measurements is summarized in Figure 2.6. The figure is divided into two sections: on the left side, the main gap distribution is displayed by five bins ranging from 1 day to more than 9 months, while on the right side a more detailed breakdown of a week bin is provided.

A key observation is that, similar to the nitrogen oxides analysis, short-term gaps are the most frequent at all stations for both PM10 and PM2.5 measurements. Some longer gaps (lasting more than a week) are evident at certain stations, with Torino-Rebaudengo and Torino-Grassi standing out for having a higher number of long gaps in PM10 measurements. The longest gaps are no longer than 9 months across any of the stations. However, Torino-Consolata exhibits some of the largest gaps among all stations on the PM10 data. Similarly, Torino-Grassi also shows extended gaps in the PM10 measurements, particularly for PM10_GBV. Although less frequent, gaps of a few weeks are present for each stations.

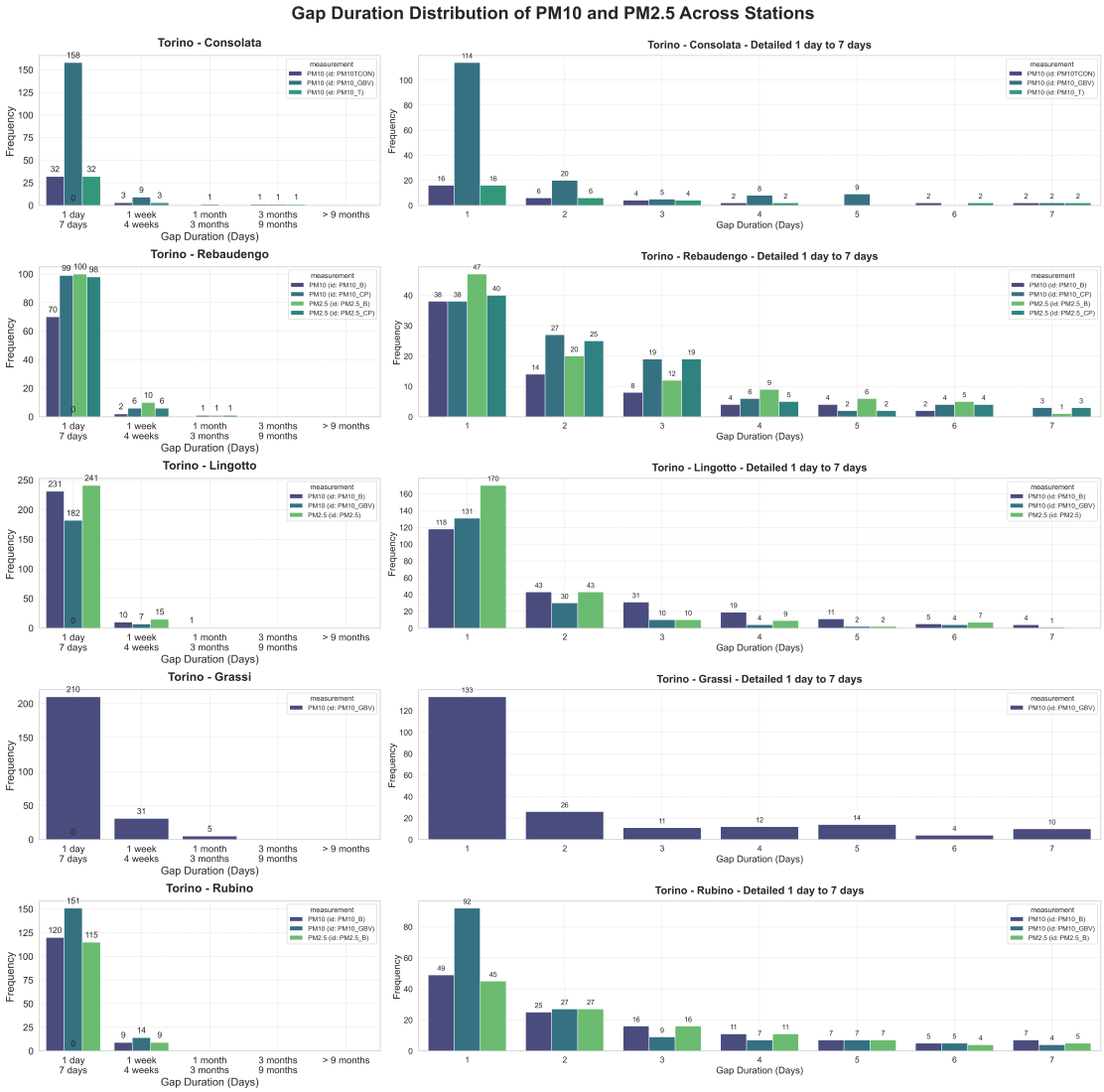

Figure 2.6: Gap Duration Distribution for PM_{10} and $PM_{2.5}$ Across Stations

2.5.2.1 Summary of Gap Analysis of PM

Here is reported a summary of the gap duration's for PM10 and PM2.5 measurements across all the stations, for the type of measurements available and with the longest time series data. For each measurement in each station is explicitly reported the length of the longest gap (with the explicit period) and the missing day percentage over the total number of present days.

Table 2.12: PM10 Missing Data Summary

PM10 Missing Data Summary				
Measurement	**Missing Days**	**Longest Gap**	**Most Frequent Gap**	**Missing %**
		Torino-Consolata		
PM10TCON	214	107 days (2001-09-16 to 2001-12-31)	1	6.51
PM10_GBV	531	116 days (2021-10-12 to 2022-02-04)	1	6.32
PM10_T	213	107 days (2001-09-16 to 2001-12-31)	1	6.48
		Torino-Rebaudengo		
PM10_B	162	14 days (2016-06-26 to 2016-07-09)	1	4.56
PM10_CP	355	48 days (2013-04-13 to 2013-05-30)	1	12.40
		Torino-Lingotto		
PM10_B	676	89 days (2011-05-09 to 2011-08-05)	1	10.36
PM10_GBV	362	14 days (2016-05-16 to 2016-05-29)	1	7.63
		Torino-Grassi		
PM10_GBV	1050	54 days (2015-06-19 to 2015-08-11)	1	12.50
		Torino-Rubino		
PM10_B	432	28 days (2018-04-27 to 2018-05-24)	1	12.26
PM10_GBV	466	17 days (2007-06-01 to 2007-06-17)	1	8.10

Table 2.13: PM2.5 Missing Data Summary

PM2.5 Missing Data Summary				
Measurement	**Missing Days**	**Longest Gap**	**Most Frequent Gap**	**Missing %**
		Torino-Rebaudengo		
PM2.5_B	421	43 days (2015-10-21 to 2015-12-02)	1	12.77
PM2.5_CP	349	48 days (2013-04-13 to 2013-05-30)	1	12.19
		Torino-Lingotto		
PM2.5	565	28 days (2022-07-26 to 2022-08-22)	1	8.85
PM2.5_B	80	9 days (2022-12-20 to 2022-12-28)	1	10.78
		Torino-Rubino		
PM2.5_B	400	28 days (2018-04-27 to 2018-05-24)	1	11.35

The key patterns in missing data highlighted by the Tables 2.12 and 2.13 reveal several important insights. As anticipated from the graphical representation, for PM10 the Torino-Consolata station has three notable gaps, with two of them for the same period both lasting approximately 107 days. The Torino-Grassi station exhibits the highest percentage of missing data for PM10, with 12.5% missing and a longest gap of 54 days, also due to the smaller number of measurements for this specific measurement . Torino-Lingotto and Torino-Rubino also show significant data gaps, with missing percentages of 10.3% and 12.2%, respectively.

For PM2.5, Torino-Rebaudengo exhibits the highest percentage of missing data, with longest gap lasting about 43 days. Torino-Lingotto shows the highest number of missing days (565), but with a lower missing percentage of 8.8%, which is possible and make sense given its longer time series data.

In general, despite short-term gaps dominate across all stations, the presence of these extended gaps, presents potential challenges for the robustness and validity of the trend analysis.

2.5.3 Final Comment on the Gap Analysis

The gap analysis led to the discovery of two important insights regarding the dataset provided by (ARPA Piemonte). First, while each measurement exhibited different frequency of missing data, the majority of the gaps were short-term, both for hourly sampling data (with most of the gaps lasting between 1 and 3 hours) and daily sampling data. These shorter gaps can be addressed using standard data-cleaning techniques, such as interpolation, and do not pose a real problem especially if the aim is the analysis of long term trend.

However, more significant issues are present, such as long-term gaps (over a month), which are probably caused by monitoring station malfunctions. These extended gaps present a substantial challenge to annual trend analysis, as they may compromise the accuracy even of long-term evaluations like yearly averages.

In this section, the analysis presented is essential to identify when and where these gaps occur, making them to be properly considered when interpreting the results of the trend analysis. A proposal to mitigate the impact of these larger gaps is to segment the trend analysis into shorter, more consistent periods or focus on shorter-term evaluations that minimize the influence of these interruptions. More attention to this will be reserved later in the thesis.

Overall, the gap analysis highlights the need to be careful when interpreting trends from this datasets, especially for long-term air quality assessments. Conclusions must account for the gaps, both short and long-term, to ensure accurate and meaningful results.

2.6 Sensitivity Analysis of Hourly Data Threshold

The goal of this analysis is to see what could be an appropriate threshold **min_valid_hours** for the minimum number of valid hourly measurements required to calculate reliable daily averages of Nitrogen oxides and particulate concentrations. Setting an optimal threshold ensures that the daily averages are representative of true daily conditions.
Indeed a threshold that is too low may include days with insufficient data, leading to biased or unreliable averages that do not capture the true variability of pollutant concentrations. Conversely, a threshold that is too high may exclude too many days, potentially reducing the dataset and limiting the temporal coverage of the analysis.

2.6.1 Sensitivity Analysis for Nitrogen Oxides

In this section is conducted the sensitivity analysis for the hourly measures of NO, NO_2, and NO_X pollutant, by the four monitoring stations in Turin with available data.

Each station's data is evaluated through four key visual components:

- **Distribution of Valid Hours Per Day:** A plot that shows the frequency distribution of valid hourly measurements per day, highlighting the proposed 12-hour threshold with a dashed red line.
- **Missing Hours Split by Time of Day:** A bar plot that presents the missing data patterns, grouped by different periods of the day (Morning, Afternoon, Evening, Night) to identify temporal biases in the measurements.
- **Daily Hourly Coverage:** Displays the hourly coverage across the entire study period, which provide a general overview of how many days are over the threshold.
- **Distribution of Hourly Measurements for Different Completeness Levels:** A histogram that highlights the completeness of hourly measurements on daily basis, divided into four categories: less than 6 hours, 6-12 hours, 12-18 hours, and 18-24 hours of valid measurements.

2.6.1.1 Representative Results from Torino-Consolata

Figure 2.7 shows the results of the sensitivity analysis for the Torino-Consolata station. The analysis indicates that the majority of days have a high level of completeness, with a significant portion of the measurements meeting the 12-hour threshold. Similar patterns were observed for the other stations (Torino-Lingotto, Torino-Rebaudengo, and Torino-Rubino), thus confirming that the 12-hour threshold for the averaging of hourly data, will not discard many measurements across all stations.

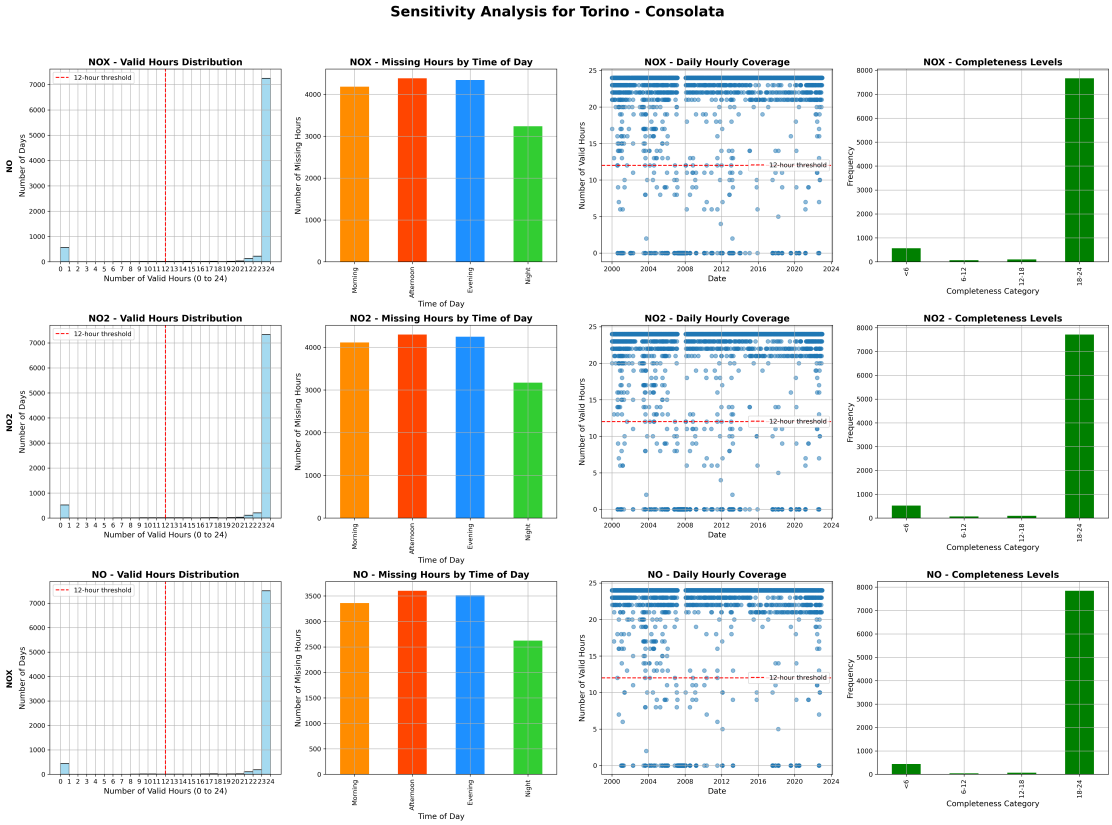

Figure 2.7: Sensitivity Analysis of NO, NO_2, and NO_X for Torino-Consolata.

2.6.2 Sensitivity Analysis for PM_{10} and $PM_{2.5}$

In order to evaluate the optimal threshold, a visual analysis is conducted, focusing on the hourly measures for both particulate measurements available in each station. It is recalled that PM measurement with hourly sampling frequency, among the selected with a long-term temporal coverage are:

For PM_{10}: PM10_T, PM10_CP.
For $PM_{2.5}$: PM2.5_B, PM2.5_BH, PM2.5_CP, and PM2.5.

The analysis is based on four key visual components represented in the figure:

- **Distribution of Valid Hours Per Day:** Through bar plot is reported the number of days with varying numbers of valid hourly measurements. A red dashed line is drawn at the 12-hour mark, highlighting the threshold under consideration. This distribution helps us understand how frequently days fall below or above specific completeness levels.
- **Missing Hours Distribution Between Day and Night:** A plot that presents the division of missing measurements between daytime (06:00 to 18:00) and nighttime (18:00 to 06:00), a common division in air quality analysis. This give a general overview of the hourly measurements distribution over a day, highlighting in a uneven case, the presence of biases in the calculus of the averages.
- **Daily Hourly Coverage:** A scatter plot that provides a temporal view of hourly coverage for the entire data period. It allows showing how the number of valid measurements varies day-to-day, providing a vision on how consistently the 12-hour threshold is met across the monitoring period and whether the coverage is evenly distributed.
- **Distribution of Hourly Measurements for Different Completeness Levels:** This set of histograms is the most representative plot of the analysis as it directly reveals the biases in the distribution of hourly measurements across a day. Four ranges of completeness are represented: less than 6 hours, 6-12 hours, 12-18 hours, and 18-24 hours of valid measurements per day. The histograms show how hourly measurements are spread throughout the day, with the associated density profile.

In this section only the results for $PM_{2.5}$ are reported, as the PM_{10} data exhibits the same pattern, supporting the decision to use the same threshold for both pollutants.

For completeness, the analogous analysis of PM_{10} data including the same visual components reported here, is provided in the Appendix (Fig. B.2, B.3).

Figure 2.8 shows the results of the sensitivity analysis for $PM_{2.5}$ hourly data, specifically for the PM2.5_CP measurement from Torino-Rebaudengo and PM2.5_BH measurement from Torino-Rubino.

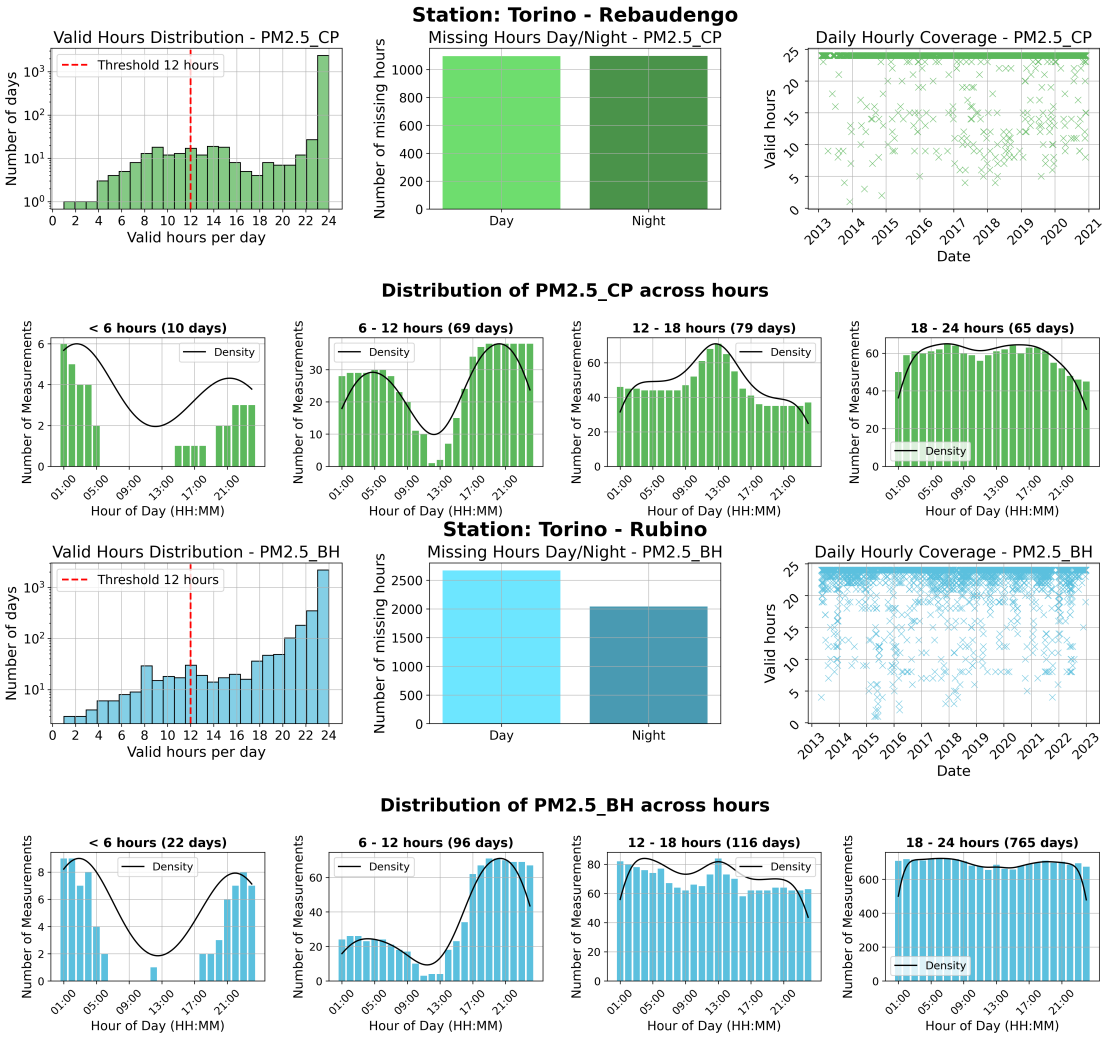

Figure 2.8: Analysis of $PM_{2.5}$ hourly data sensitiveness

The plots provide clear evidence that days with more than 12 valid hours tend to have balanced measurements across the entire day, with a nearly uniform distribution, ensuring comprehensive daily averages. In contrast, days with fewer than 12 hours of valid data show gaps that often coincide with key daytime hours (during middle hours), leading to potential biases in daily averages.

Additionally, as shows the Valid Hours Distribution histogram, for days that do not have the full 24 valid measurements, most still have significantly more than the chosen threshold of 12 hours. Moreover, there is no substantial difference between the number of missing hours during the day and night.

In conclusion, the analysis supports the use of the 12-hour threshold. This choice maintains data integrity and allows to minimize potential biases introduced by the averaging process, without needing to discard many days, thus preserving broad data coverage.

2.7 Monitoring Stations in Turin urban area

As mentioned earlier, the dataset provided by ARPA Piemonte includes a file, `stazioni.csv`, which contains detailed information about all monitoring stations, having data registered in the database. This file provides essential details such as geographical coordinates, operational periods, the types of pollutants measured at each station, and the unique station identifiers, especially crucial for filtering all the measurements. Indeed the air quality database includes measurements from across the entire municipality of Turin, so the filtering step was necessary to focus on the stations located specifically within the urban area.

This section presents an overview of the five monitoring stations situated in Turin's urban area, emphasizing the types of pollutants each station measures and the environmental context surrounding each location. This focus ensures a clear understanding and allows for concrete reasoning of how the stations' placement reflects the different pollution sources and urban dynamics of the surrounding environment. Further information on these five stations is available on the official website of the Comune di Torino [4].

2.7.1 Selection of Urban Stations

To ensure that the data analyzed through this thesis accurately represented urban air quality of Turin, only the measures from certain stations were maintained from the original database provided.
The filtering criteria included two key columns from the `stazioni.csv` file:

`rete_monitoraggio` = 13, which identifies the network monitoring stations of Turin.

`codice_istat_comune` = 001272, which identifies the stations belonging to the administrative area of Turin.

Applying these filters, the resultant unique stations in the datasets are associated with the following five `progr_punto_com` values: 803, 805, 806, 819, and 822.

These `progr_punto_com` values correspond to the following monitoring stations:

- Torino-Consolata(803)
- Torino-Rebaudengo(805)
- Torino-Lingotto(806)
- Torino-Grassi(819)
- Torino-Rubino(822)

These finding agree with the five monitoring station listed in the Comune di Torino website, previusly cited.

2.7.2 Urban Distribution of Monitoring Stations

Using QGIS, the coordinates from the `stazioni.csv` file were used to create a map that visualizes the spatial distribution of the monitoring stations within Turin urban area. The results is reported in Figure 2.9.
This map highlights the station locations relative to key urban features, such as major roads, residential and central areas, providing context for the spatial analysis of air quality. The five stations are uniformly distributed in the Turin urban area

Figure 2.9: Map of monitoring stations in Turin, visualized with QGIS.

The distances between the stations can be of can be of relevant importance in understanding potential spatial dispersion of pollutants and the variations in air quality across different zones of the city. Using the Measure Line tool in QGIS, and after correcting the map's geo-reference, the distances between each station were calculated.
The stations closest to each other are Torino-Consolata and Torino-Rebaudengo, with a distance of approximately 2.65 km. On the other hand, more distant stations are Torino-Lingotto and Torino-Rubino, which are 6.10 km apart from Torino-Consolata.

These variations in distance can provide a starting point on further investigation that focus on correlations between air quality measurements at different stations, considering their spatial positioning. Specifically, stations located farther apart may capture distinct pollution profiles, while closer stations may show similar trends. This can potentially highlight the possibility of shared sources of pollution, similar urban environments, and provide insight into the range of spacial variability for a pollutant gas or PM.

2.7.3 Station Descriptions

The monitoring stations in Turin are categorized based on two combined factors: the nearest sources of pollution and their geographical setting. These categories combine to provide a more comprehensive classification of monitoring stations. An overview of typical station types based on their combined characteristics is reported below:

Suburban Background (Fondo Suburbana): Mainly define urban areas with residential or commercial characteristics, and generally far from direct pollution sources, such as heavy traffic or industrial zones. The primarily measures that provide insight into the air quality here are nitrogen oxides (NO_X).

Urban Traffic (Traffico Urbana): Stations placed mainly in high-traffic areas of the city, where vehicular emissions are the primary source of pollution. These stations are often equipped with instruments that measure carbon monoxide (CO) and nitrogen oxides (NO_X) to monitor the direct impact of traffic.

Rural Background (Fondo Rurale): Stations located away from urban centers, in areas with residential and agricultural features. The instrumentation at these stations typically measures nitrogen oxides (NO_X), ozone (O_3), sulfur dioxide (SO_2), and particulate matter (PM_{10}).

The classification on the Turin municipality website reports the categorization of each of the five monitoring stations in the Turin urban area as followings:

- **Torino-Consolata:** Urban Traffic, with residential/commercial feature. This station, indeed is located near the city center. It is one of the most well-equipped stations for monitoring both gaseous pollutants and particulate matter across a wide range of sources, including the detection of various organic compounds and heavy metals.

- **Torino-Rebaudengo:** Urban Traffic as the previous. Situated near a very busy intersection, for both local traffic and vehicles entering or exiting the city, and still in a industrial/commercial zone.
- **Torino-Lingotto:** Urban Background. This station is located near a mix of residential, commercial, and industrial areas, monitors a wide range of general pollutants. Its instrumentation includes sensors for Total Suspended Particles (PTS), Carbon Monoxide (CO), Nitric Oxide (NO), Nitrogen Dioxide (NO_2), and Ozone (O_3), making it one of the most reliable stations for air quality monitoring.
- **Torino-Grassi:** Suburban Traffic. Situated in prevalent residential and slightly industrial area, the station monitors mainly Total Suspended Particles (PTS) and particulate matter (PM_{10}).
- **Torino-Rubino:** Urban Background, as Lingotto one. However the area is reported to be mainly residential, without the presence of industrial activities. A quieter populated area with the pollution source mostly due to vehicles activity, this station monitors pollutants such as CO and NO_X.

As it is possible to understand, the five air quality monitoring stations considered for analysis in this thesis provide a comprehensive representation of Turin's urban environment. Thanks to the slight differences in the surrounding environments of each station, ranging from residential to high-traffic or semi-industrial areas, it is possible to obtain a general overview of how pollutant levels varies across different urban settings.

By understanding the environment and pollutant sources surrounding each station, this section forms the basis for the subsequent analysis, which will allow more in-depth consideration over the air quality trends, pollutant levels, and their potential impacts on public health and urban life.

Chapter 3

Measure Comparison

3.1 Correlation Analysis of Different Types of Particulate Matter (PM_{10}, $PM_{2.5}$) Measurements

As pointed out in Chapter 2, the dataset for particulate matter contains various measurements that differ based on the instrumentation used for the detection or the type of the particle. In this section is reported the comparison between the different types of PM measurements. The primary goal is to assess the similarity between measures by different detection systems and determine the possibility of integrating the data to improve temporal coverage.

3.1.1 Methodology

In order to compare measurements with different sampling frequencies (daily and hourly data), the data was processed to match the lowest common frequency. Hourly measurements were averaged, over the 24-hour, to create daily averages allowing for consistent comparisons on a daily basis. The hourly measurements averaged into daily values are denoted in the scatter plot with the suffix '_mean'. Creating the scatter plots for the comparison of each measure involved the calculation of the common period of validity for each couple of measures. Only paired measurements, whether belonging to the same station or not, with both data points valid for the same period were compared daily.

In order to quantify the similarity between different measurements, the Pearson correlation coefficient, mean absolute error (MAE), and root mean square error (RMSE) were calculated for each measure pairs. Below are reported the metrics used in this analysis, with their mathematical definitions.

Pearson correlation coefficient: The Pearson correlation coefficient measures the linear correlation between two time series datasets, X and Y (e.g. daily PM measurements recorded by different instruments) and it is defined as:

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}} \tag{3.1}$$

where X_i and Y_i represent the PM concentrations measured on each day, and $\bar{X}$ and $\bar{Y}$ are the means of the time series. The Pearson coefficient can range from -1 to 1, with values closer to 1 indicating a strong positive correlation, values closer to -1 indicating a strong negative correlation, and values around 0 indicating no linear correlation.

Mean Absolute Error (MAE) The Mean Absolute Error (MAE) is a measure of errors between paired observations and is calculated as:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|X_i - Y_i| \tag{3.2}$$

where X_i and Y_i represent the concentrations measured on the same day, for instance by two different instruments. MAE is essentially the average absolute difference between these values and quantify whether the two measurements are compatible. It is also effective for highlighting differences in measurements of the same type taken (at the same time) but from different locations, thereby emphasizing spatial variability in particulate concentrations.

Root Mean Square Error (RMSE) The Root Mean Square Error provides a measure of the differences between two values. The formula for RMSE is:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - Y_i)^2} \tag{3.3}$$

MAE and RMSE are two measure of the average magnitude of the error, however Root Mean Square Error gives a relatively stronger weight to large errors, as it squares the differences before averaging, making it particularly useful in detecting large errors.

Important Note: In this chapter two main central aspect on the comparison are reported: the first relative to the comparison of different measures from the same station, the second relative to the comparison of measures from the different station. The first methodology aimed to compare spatially compatible measures, to see if they could be combined to extend data coverage. The second help better to understand how similar or different the monitoring stations are, or address particular behaviour to fast or slow dispersion process in the pollutant. The results for the two aspect are reported separately.

3.1.2 Correlation between measurements from the same station

Comparing measurements from the same station helps to identify how different methods can complement each other to enhance temporal coverage, uncover discrepancies between measurement types or detect potential errors in data labeling for identical measures.

To make the results more readable and organized, particulate measurements and element dispersed in atmosphere measurements (detected only by Torino-Lingotto station) are separated in the analysis. Only the scatter plots related to particulate measurements (PM_{10} and $PM_{2.5}$) are presented in this section. Scatter plots for atmospheric elements highly correlated are included in the appendix, Figure B.1, so that these results can offer interesting insights that could be explored in future research. In addition, scatter plots with highly correlated measurements (Pearson correlation coefficient greater than 0.9) are colored with a red border to facilitate the identification of measurement pairs that can be considered for integration. Pairs of measurements shared by multiple stations are presented in the same subplot.

Scatter plots of PM_{10}

In Figures 3.1, 3.2, and 3.3, the scatter plots for PM_{10} measurements at various stations, for which multiple measures are present, are shown.

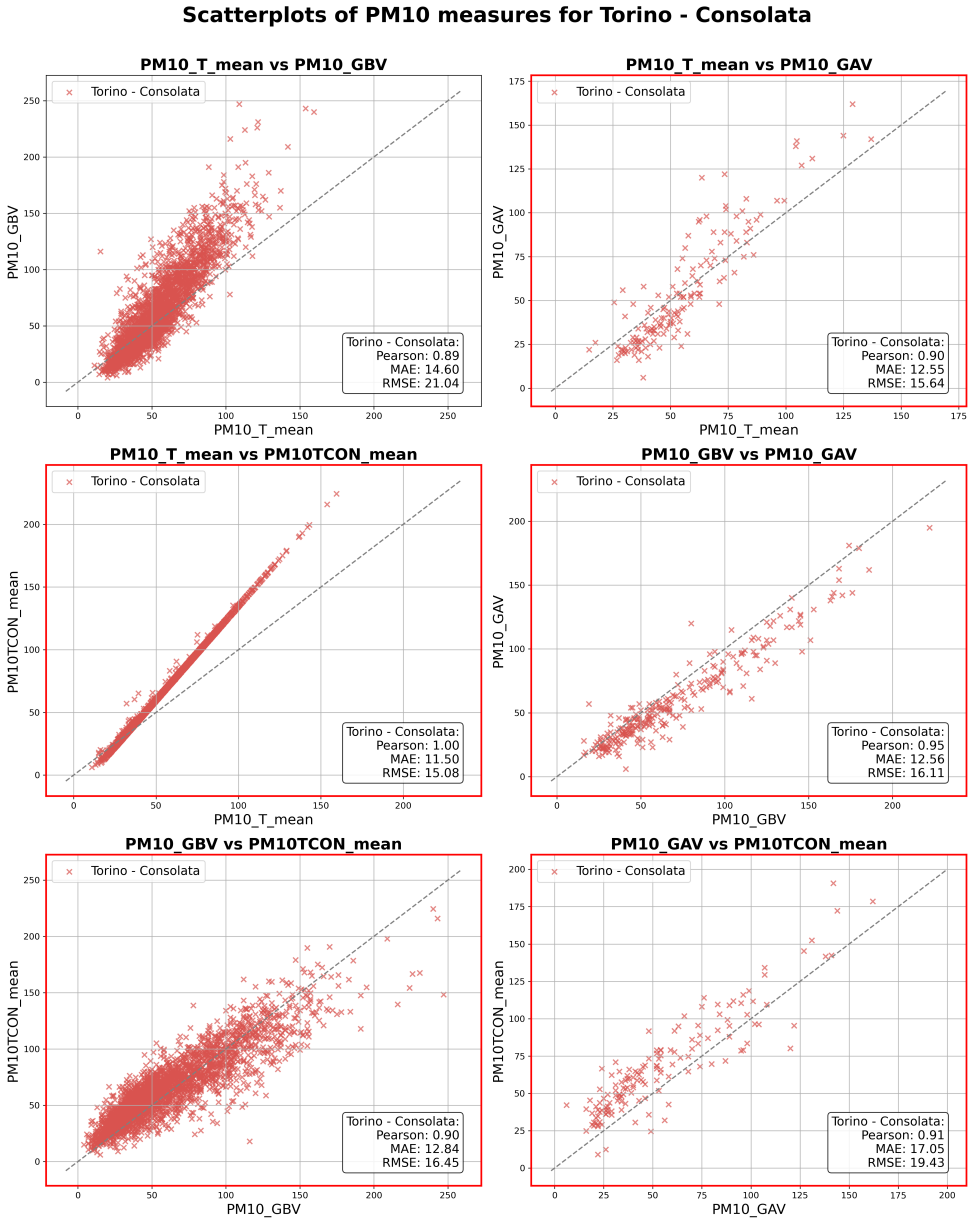

Figure 3.1: Scatter plots of PM_{10} measurements for Torino-Consolata

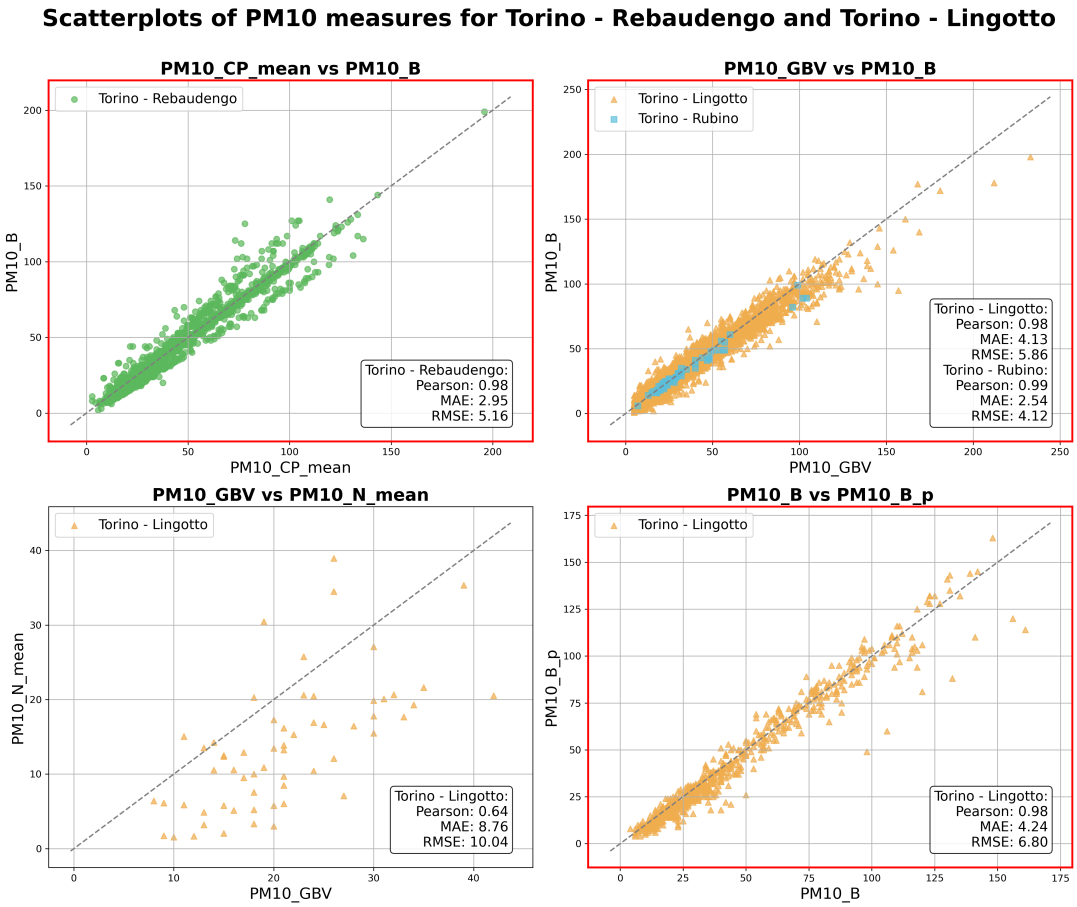

Figure 3.2: Scatter plots of PM_{10} measurements for Torino-Rebaudengo and Torino-Lingotto

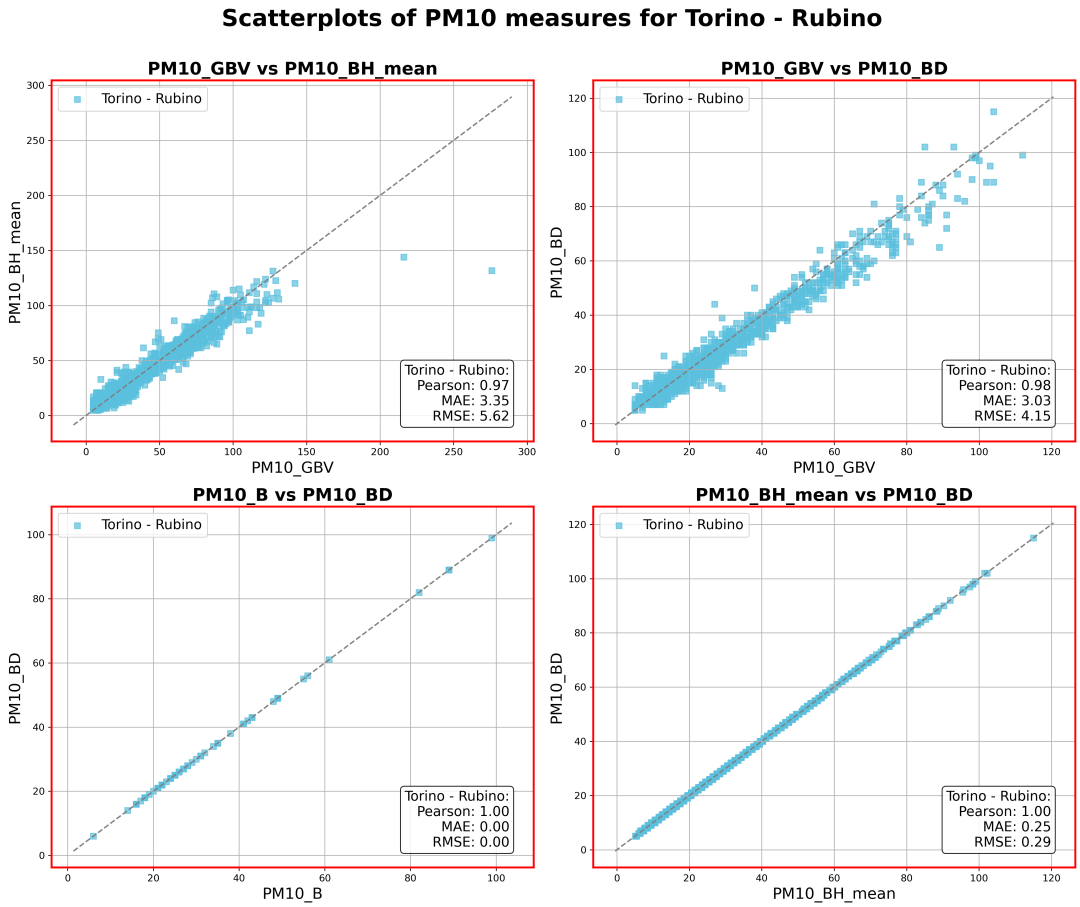

Figure 3.3: Scatter plots of PM_{10} measurements for Torino-Rubino

The previous figures reveals several important information about correlations, which are summarized for each station as following:

- At **Torino-Consolata**, PM10_GBV and PM10_GAV show a good correlation (Pearson 0.95). This two measures also show a correlation above 0.8 with PM10_T and PM10_TCON measures. Additionally, PM10_T (Hourly) measurements align perfectly with PM10TCON values, which are adjusted TEOM (Tapered Element Oscillation Microbalance) measurements (raw values reduced by 50% and subtracting 15) to match actual PM_{10} concentrations. From the literature it emerges that the TEOM tends to underestimate the PM compared to the gravimetric samplers and require adjustments [2].

- **Torino-Rebaudengo** present a very high correlation (Pearson = 0.98) between PM10_CP and PM10_B, indicating an agreement between measures from the particle counter and measures with Beta method.
- At **Torino-Lingotto**, there is a high correlation (Pearson = 0.95) between PM10_GBV and PM10_B and also between PM10_B and PM10_B_p, indicating these measures are possibly suitable for integration. In contrast, PM10_GBV and PM10_N do not show a high correlation, suggesting some differences between filter-derived and Nephelometer measures of concentration.
- **Torino-Rubino** show a perfect correlation (Pearson = 1.00) for PM10_B and PM10_BD indicating the same type of measure registered with different names (most likely due to error in the registration). Also PM10_BH and PM10_BD present a perfect correlation which confirming that hourly averaged measurements can reliably substitute for daily measurements. Also in this station

Scatter plots of $PM_{2.5}$

Similarly to the approach used for PM_{10}, the comparison represented in each subplot in Figure 3.4 shows the correlation between measurements from Torino-Rebaudengo, Torino-Rubino, and Torino-Lingotto stations. These stations have different types of fine particulate measurements available, which can be compared.

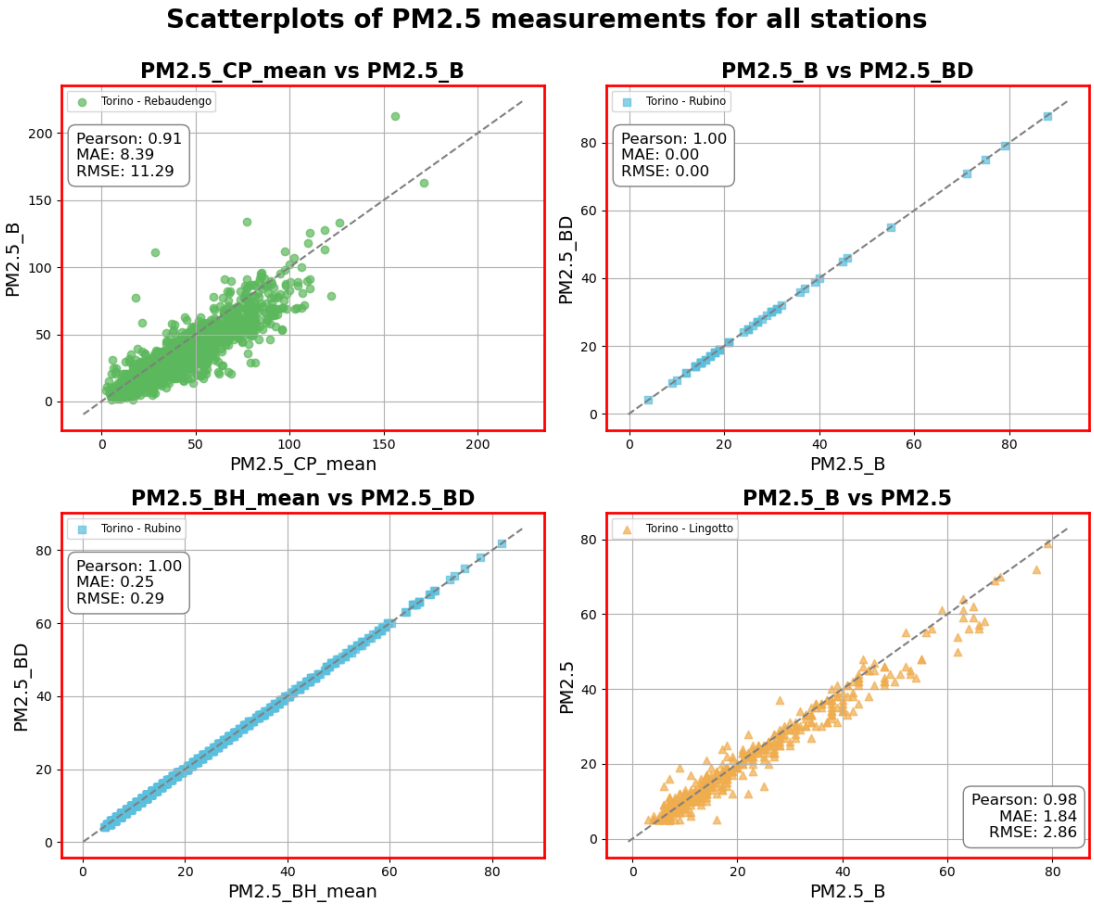

Figure 3.4: Scatter plots of $PM_{2.5}$ measurements for all stations

The key findings for each station and measures of PM2.5 summarizes in:

- **Torino-Rebaudengo** show a high correlation for PM2.5_CP_mean and PM2.5_B (Pearson = 0.91) measurements, and this again suggests a close relationship between particle counter and Beta detection methodology.

- **Torino-Rubino** shows two couples of measures with perfect correlation. PM2.5_B and PM2.5_BD with Pearson = 1.00 indicating that these measurements are effectively identical; and PM2.5_BH and PM2.5_BD indicating that the hourly average concentration corresponds to the daily measurement.
- At **Torino-Lingotto**, PM2.5_B (Beta) and PM2.5 (Low Volume gravimetric) measurements show strong agreement.

Summary and Comparison of Different Methods of Measure

The representation with scatter plots of the different concentration measurements, for common intervals belonging to the same station, provided a clearer view of the particulate data, highlighting certain similarity between them.
Possible nomenclature registration error (or changes in measurement instrumentation) were discovered in the data, whereby identical measurements were named differently (e.g. PM10_B and PM10_BD; PM2.5_B and PM2.5_BD).
Furthermore, several pairs of PM_{10} and $PM_{2.5}$ measurements showed high correlations across different stations, which made them suitable for temporal integration, which enhance data robustness and addresses missing values.

In the continuation of this thesis, temporal integration is applied exclusively to measurement pairs with a Pearson correlation coefficient of 1, specifically to the _B, _BH and _BD measurements for both PM_{10} and $PM_{2.5}$.
The trend analysis in the following chapter continue using the cleaned raw data, with adjustments based on the integrated values.

From the comparison interesting differences between various measurement methods are found:

- Gravimetric analysis (used in Low Volume measurement) and Beta-ray attenuation methods (used for Beta measurement) show a good relationship for both PM_{10} and $PM_{2.5}$ measurements, with Pearson correlation above 98% . This finding aligns with result present in the literature, which reports a significant correlation between these two methods [12].
- Filter-derived and Nephelometer data (PM10_GBV and PM10_N) do not exhibit a strong correlation, which highlight a notable difference in the two types of concentration measurements. This observation is supported by the literature [14].

3.2 Scatter plots of general correlation (between measures from different stations)

The scatter plots reported in this section allow for the evaluation of correlations between different types of particulate matter measurements, despite the fact that the measurements come from different locations, where pollution levels can typically vary due to local factors.

3.2.1 PM_{10}

In the original data, are identified 10 different types of PM_{10} measurements. In principle this allows for the comparison of $\binom{10}{2} = 45$ pairs. However, only 23 of these pairs have overlapping data for the same periods (same days), making them the only ones meaningful to represent in the scatter plots.

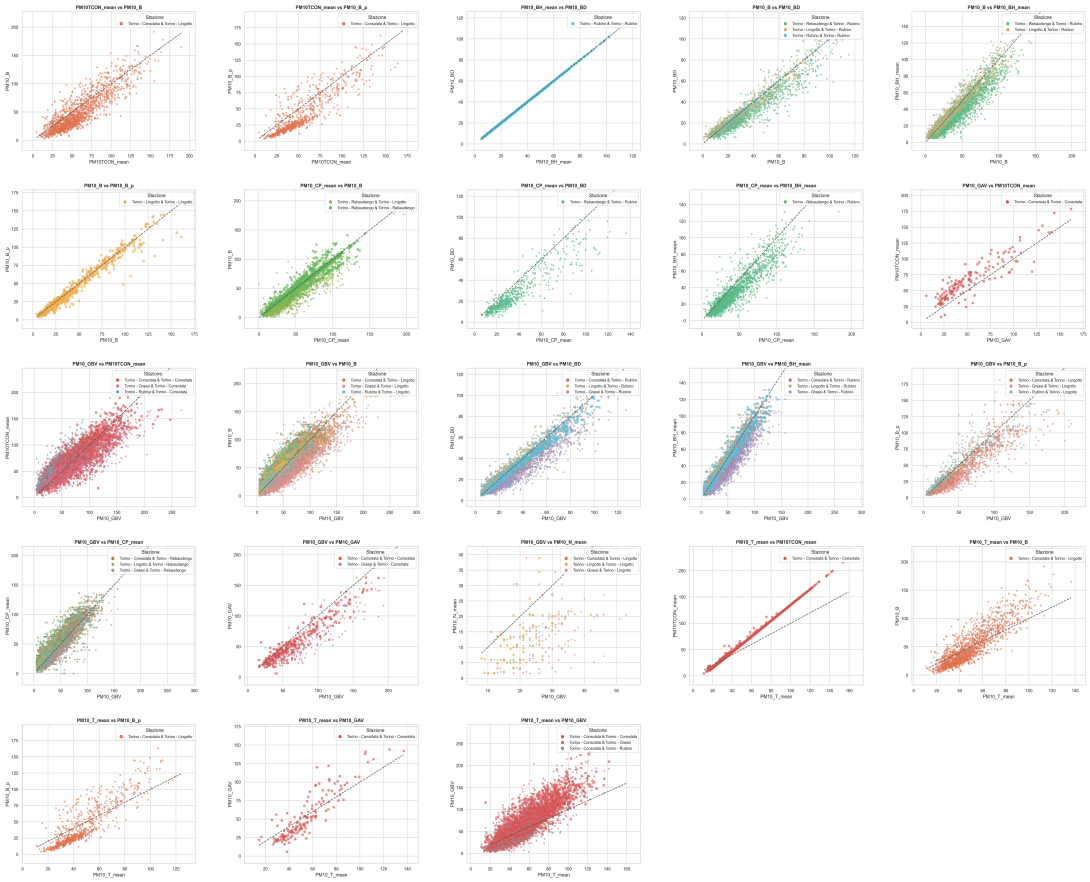

Figure 3.5: Scatter plots of PM_{10} measurements pairs across all stations

3.2.2 $PM_{2.5}$

For the $PM_{2.5}$, there are only ten pair of overlapping measurements that have overlapping data for the same days.

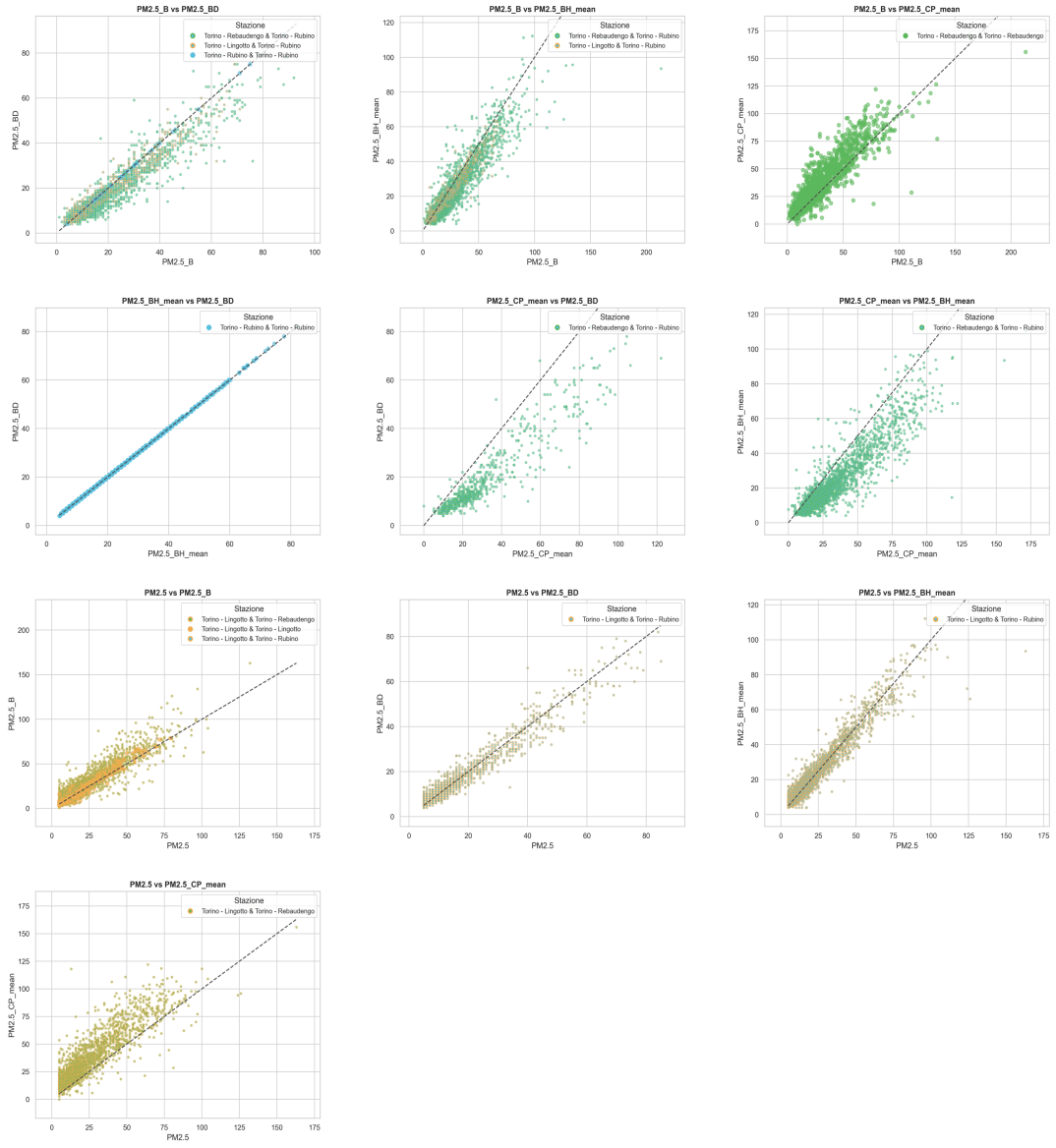

Figure 3.6: Scatter plots of $PM_{2.5}$ measurements among all stations

Summary of Correlation Findings

Overall, from Figures 3.5 and 3.6, it is observed that a strong correlation exists between the various types of PM_{10} and $PM_{2.5}$ measurements across different stations. This confirms the results from the earlier analysis, which highlighted that the different measurement pairs showing strong correlation are consistent with those previously identified.

Chapter 4

Temporal Analysis

This chapter is the core and the turning point that encloses the central topic of the thesis. In order to provide a clear view of long-term time series two methodologies are combined: first, through the use of summary and advanced statistical methods and second, with comprehensive graphical representation which are crucial to highlight environmental air quality condition.

4.1 Linear Trend

4.1.1 Methodology

The analysis mainly focuses on using annual averages to obtain a smooth representation of the temporal evolution of gases concentration. This approach effectively reduce the influence of short-term fluctuations, and as a result the trend lines can better capture the underlying continuous changes in pollutant levels. The averages were computed over the solar year, from January to December, instead of considering seasonal years (e.g., from autumn to autumn). The second choice, would seem the most appropriate for environmental analysis issues, as seasonality and seasonal variation are more enhanced. Nevertheless, most scientific papers regarding environmental monitoring and in regulatory environments (e.g European Environment Agency) rely on annual averages to define standard levels. The approach addressed in this thesis entail a straightforward and simple calculation that ensure the results are standardized and can be used for comparisons with legislative parameters.

For the graphical representation of the linear regression line, the `regplot` function from the Python library `Seaborn` was used [15] and as a complement, to statistically quantify the trends, the Ordinary Least Squares (OLS) method present in

`statsmodels` was utilized[11]. The OLS method works by minimizing the sum of the squared differences residuals between the observed data points and the estimated regression line. This tool provides different detailed statistical parameters such as slope, intercept, R-squared values, and sensitivity like p-values.

4.1.2 Handling of Data Gaps

A significant aspect of the data analysis involved dealing with large gaps in the temporal series for the concentration measurements. The presence and identification of data gaps, was done during the data preprocessing stage (detailed in Chapter 2) where results were saved and accessible for later and further different task.

Important Note: For this initial trend analysis, gaps in the data were not specifically accounted for, and annual averages were calculated based on the available data points for each year, without exclusions. Gaps in the data were anyway highlighted in trend representation, to better emphasize where the calculation may be unreliable due to missing information.

The decision of using raw data for the initial trend analysis provides a primary view of trends, but some statistical detailed results are to be considered distorted. A more refined approach would involve considering segmentation of the long term time series before and after significant continue missing measurements (e.g., those exceeding two months). In addition boundaries problem can occur, when in a time series, the presence of partially covered year, due for not perfect alignment with the start or ending point of monitoring period for stations, might affect averages. In this way, the segments are therefore treated independently and a more correct estimation of the trend can be obtained.

4.1.3 Linear Trend Analysis of NO, NO_2, NO_X

This section summarize the trend analysis results both as graphical and tabular forms, illustrating the temporal evolution of nitrogen oxides (NO, NO_2, and NO_X) concentrations across the considered monitoring stations.

4.1.3.1 Station-specific plots

The station-specific plots in Figure 4.1 collect the the annual average temporal series of NO, NO_2, and NO_X concentrations. Each station plot reveal critical points, such as the sharp NO peak at Torino-Consolata in 2005, which may be attributed to specific data anomalies (as expected in correspondence of the underlined large gap). Also for Torino-Rubino stations is it possible to see the boundary effect of the

not perfect alignment between the start of 2007 and the start of monitoring period for the station. In fact the partial initial uncover for Torino-Rubino, especially with missing data in Winter month lead to an underestimation of the average concentration for that year.

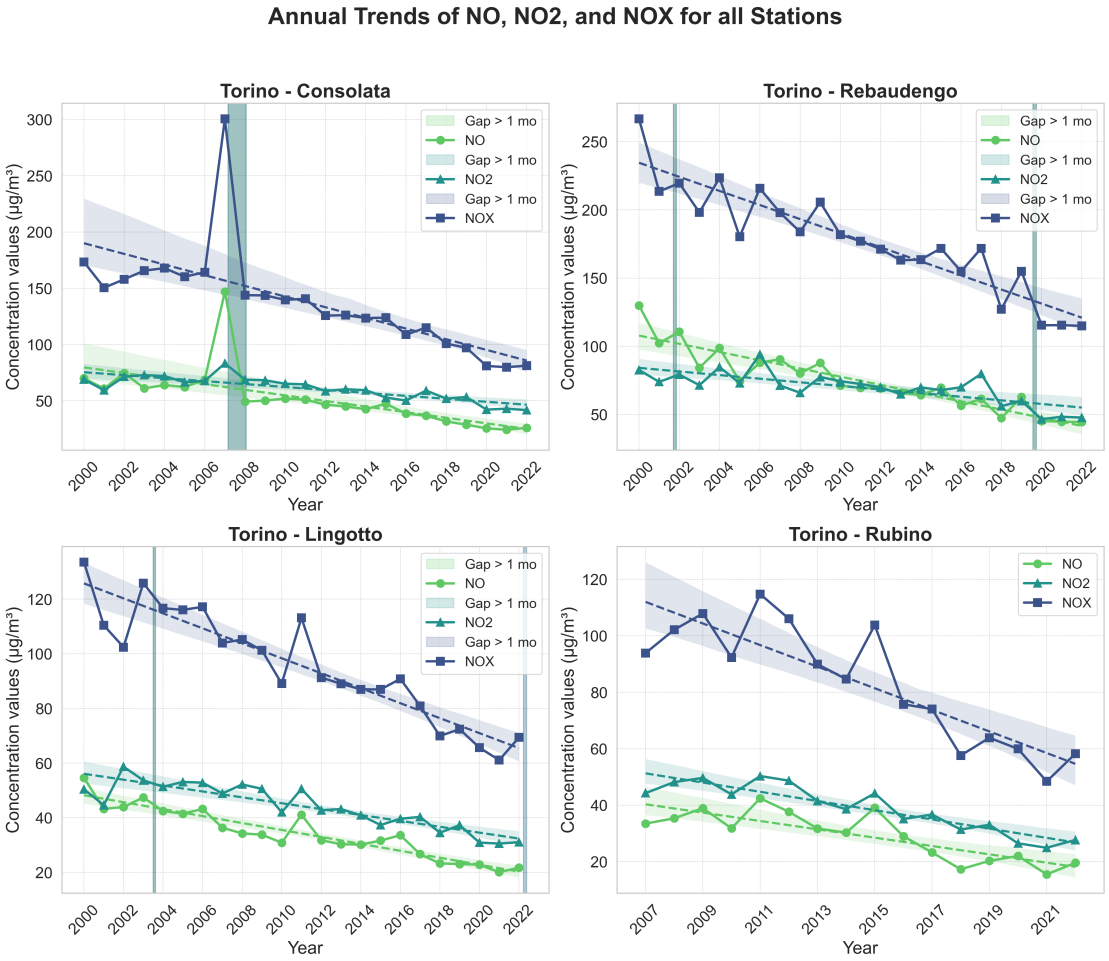

Figure 4.1: Annual Trends of Nitrogen Oxides with gap highlighted

Specific observations at each station further clarify these trends:

- **Torino-Consolata**: This station exhibits a large and constant decline in all pollutants levels, may related to the effectiveness of urban traffic reduction measures and improved vehicle emissions standards, especially considering its location in the residential historic city center.

- **Torino-Rebaudengo**: Consistently high pollutant levels are observed throughout the analysis period, emphasizing the impact of dense traffic (station located near major roads of Turin, constantly affected by high traffic). Nevertheless a gradual decline is visible.
- **Torino-Lingotto**: This station shows a consistent and moderate decrease in pollutant concentrations, with a small variability. This might be due to more stable and reliable data togheder with the presence of more stable pollution sources in the surrounding area, residential commercial with and light industrial activities.
- **Torino-Rubino**: This station presents lower overall concentrations and a gradual decline, that align with the quieter, suburban-like residential area with low traffic and industrial impact.

Despite the lack of consideration for data gaps, the figure 4.1 and the observations clearly highlight a decreasing trend in the concentration of all gases across all stations over the years.

4.1.3.2 Summary of Trend Analysis of NO, NO_2, NO_X

The analysis of station-specific plots and linear trend regressions illustrate a significant reduction in NO, NO_2, and NO_X concentrations across all monitoring stations under analysis, confirming and making more evident the overall downward trends.

In Table 4.1, as support for the previous graphical representation, results from the of the Ordinary Least Squares analysis are summarized. In each columns the table reports key regression statistics, including slope, p-values, and R-squared values, for each station and pollutant, highlighting in bold significant trends. The significance level indicates how statistically reliable the results are, with symbols representing the strength of the evidence against the assumption that the observed trend is due to random chance: triple XXX indicates highly significant results (p-value < 0.001), double XX still very significant results (p-value between 0.001 and 0.01), a single X indicates significant results (p-value between 0.01 and 0.05) instead ns stands for not significant (p-value ≤ 0.05). This notation is commonly used to provide a straightforward understanding of the p-value's magnitude.

The negative slopes confirm the observed downward trends.

Table 4.1: Regression Statistics on Annual Average by Station and Gas

Regression Statistics by Station and Gas

Station	Gas	Slope	P-Value	R-Squared	Significance Level	Trend Direction
Torino - Consolata	NO	-2.4845	0.0006	0.4386	X X X	Decreasing
Torino - Consolata	NO2	-1.3126	0.0000	0.6865	X X X	Decreasing
Torino - Consolata	NOX	-4.7397	0.0002	0.4929	X X X	Decreasing
Torino - Lingotto	NO	-1.2713	0.0000	0.8777	X X X	Decreasing
Torino - Lingotto	NO2	-1.0775	0.0000	0.7877	X X X	Decreasing
Torino - Lingotto	NOX	-2.7408	0.0000	0.8520	X X X	Decreasing
Torino - Rebaudengo	NO	-3.0108	0.0000	0.8536	X X X	Decreasing
Torino - Rebaudengo	NO2	-1.3217	0.0000	0.5702	X X X	Decreasing
Torino - Rebaudengo	NOX	-5.1630	0.0000	0.8432	X X X	Decreasing
Torino - Rubino	NO	-1.4775	0.0001	0.6749	X X X	Decreasing
Torino - Rubino	NO2	-1.6321	0.0000	0.8214	X X X	Decreasing
Torino - Rubino	NOX	-3.8254	0.0000	0.7421	X X X	Decreasing

As stated before, the analysis of linear trends included all available data points, so it is important to acknowledge that it is heedless of the missing values. This may affect the precision of the calculated statistics. Despite this limitation, the findings offer meaningful insights into the overall trends about pollutant concentrations and underline a notable decline across the study period. The observed regression slopes provide a general understanding of how pollution levels have changed over time and highlight the relative success of emission reduction measures implemented at different monitoring stations.

4.1.3.3 Analysis of Variability

Understanding how pollutant concentrations fluctuate throughout the year gives a deeper insight into air quality dynamics during the whole period of the analysis. Indeed, it's important to not only consider annual average levels but also higher-order moments such as the standard deviation (second moment) and kurtosis (fourth moment), which provide a deeper layer of information.

The standard deviation reflects the degree of fluctuations in pollutant levels, and given that all measurements over the course of a year are considered, a reasonably large level of variability is expected. However, the evolution of standard deviation over time is particularly revealing since it can highlight shifts either toward greater

stability, or increased variability, in pollutant levels. For instance, a decreasing trend in standard deviation, might indicate that pollutant levels are becoming more stable. This can reflect the effective pollution control measures or changes in emission sources.
On the other hand, kurtosis looks at the tail of pollutant concentrations distribution and can reveals the presence of extreme events. High kurtosis values suggest that there are more frequent or major pollution spikes, even if overall average levels appear stable.

Figures 4.2 and 4.3 illustrate the annual standard deviation and kurtosis of NO, NO_2, and NO_X concentrations across the considered monitoring stations.

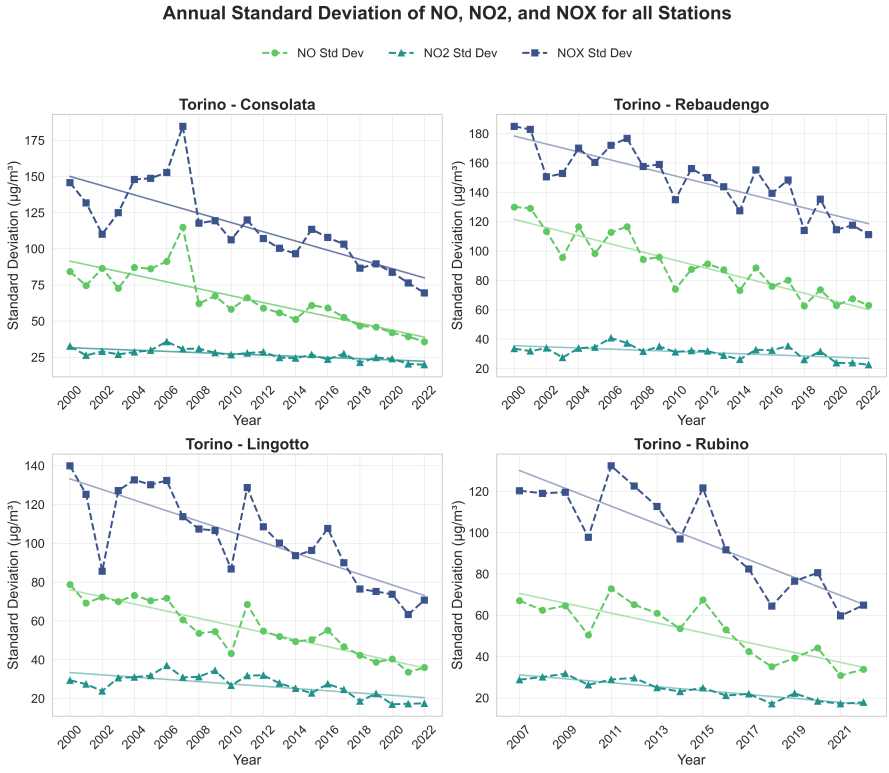

Figure 4.2: Annual Standard Deviation for Nitrogen Oxides Across all Stations

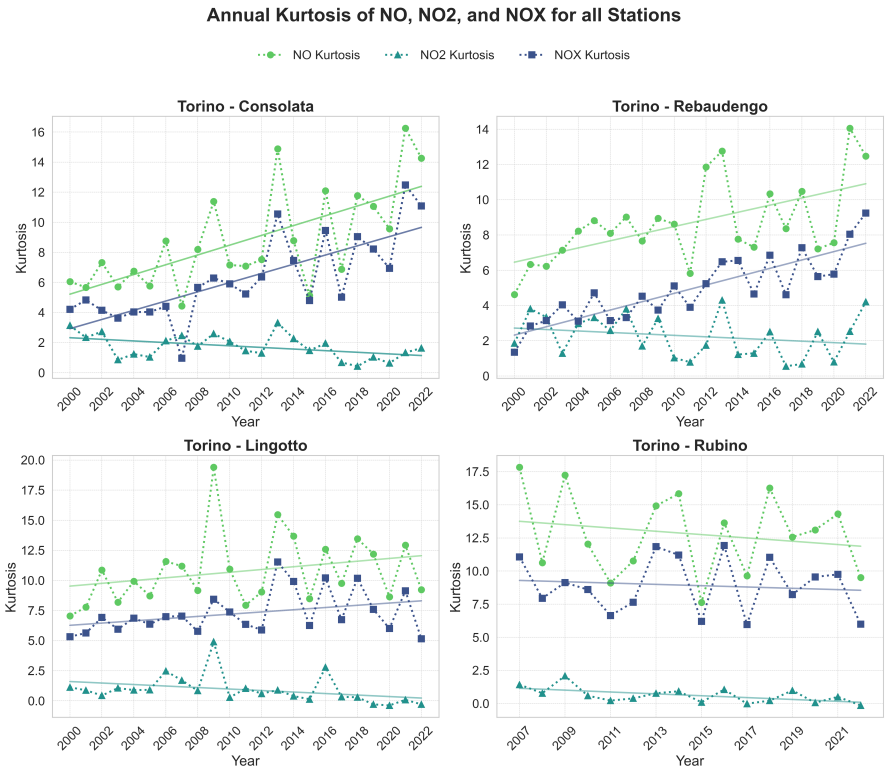

Figure 4.3: Annual Kurtosis for Nitrogen Oxides Across all Stations

4.1.3.4 Comparative Trends Across Stations

Figures 4.4, 4.5, and 4.6 illustrate the annual trends of NO, NO_2, and NO_X concentrations for the considered monitoring stations in the same plot. This representation better emphasizes the previous results, by allowing a direct comparison between stations.

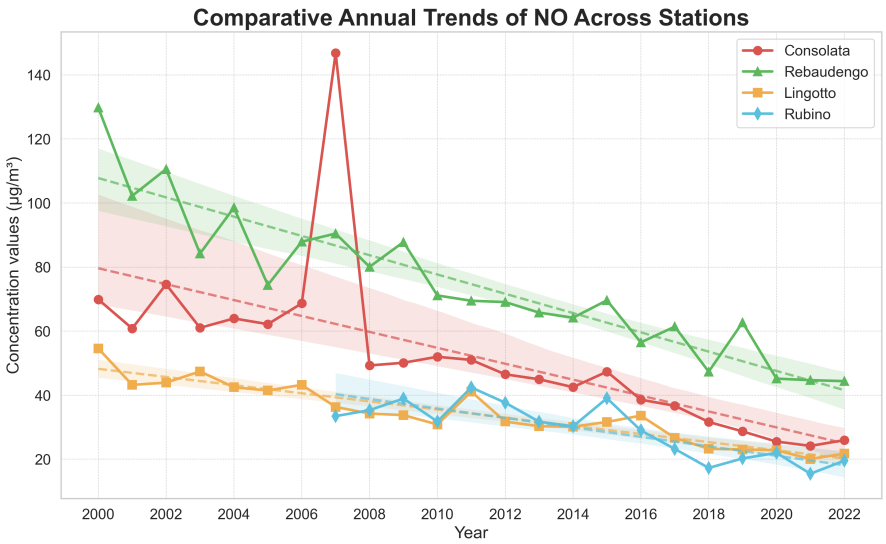

Figure 4.4: Comparative Annual Trends of NO Across Stations

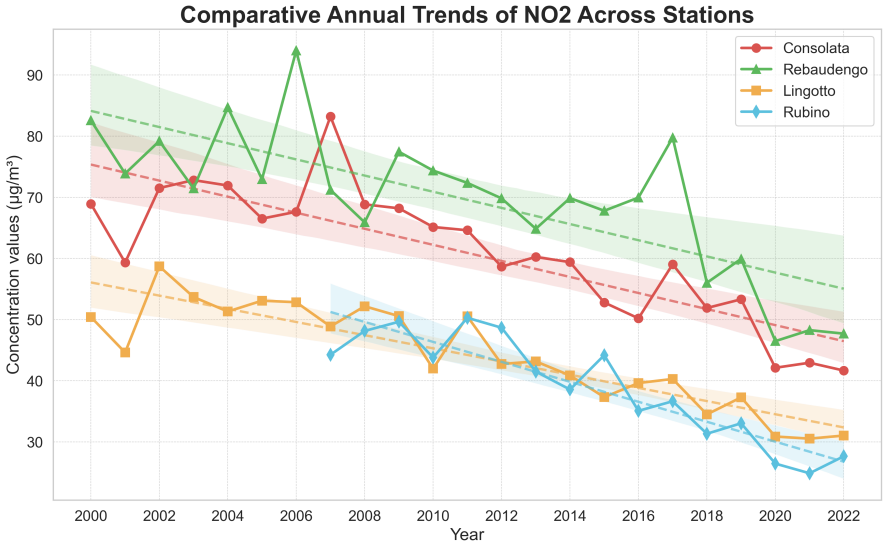

Figure 4.5: Comparative Annual Trends of NO_2 Across Stations

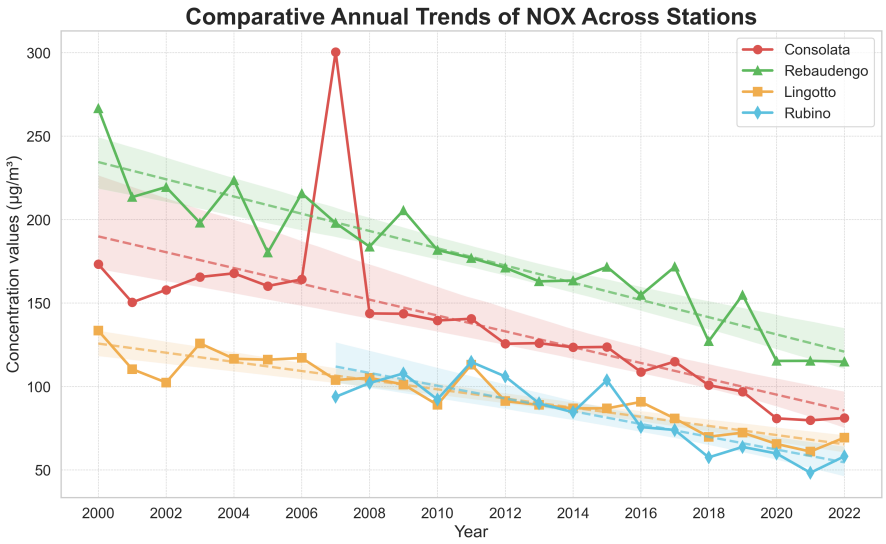

Figure 4.6: Comparative Annual Trends of NO_X Across Stations

From the comparative analysis, information of the temporal distribution of pollutant levels across different monitoring stations are obtained. The results aligns with the expectation given the geographical locations of each station. As previously anticipated, the highest concentration levels are observed at Torino-Rebaudengo probably related to its environmental context, while the lowest levels are recorded at Torino-Rubino, compatible with its quieter residential area. This highlights how the station locations significantly influence the observed data.

4.1.3.5 Heatmap Analysis of Temporal Changes

In this paragraph, another graphical representation is presented, which can be adopted to emphasize certain aspects of the linear trend analysis.
Figure 4.7 presents the monthly concentration heatmaps of NO, NO_2, and NO_X across the four monitoring stations. These heatmaps report a granular visualization of the temporal dynamics of pollutant levels, highlighting both seasonal patterns, longer-term trends and a straightforward visualization of month gaps. By showing pollutant levels month-by-month, the heatmaps allow to closely examine how local emissions, seasonal weather conditions, and changes over time affect air quality. This particular representation is suitable also for further underlining with different colors the months where the average concentration exceeded a fixed threshold, appropriately based on the EU air quality directive [7].

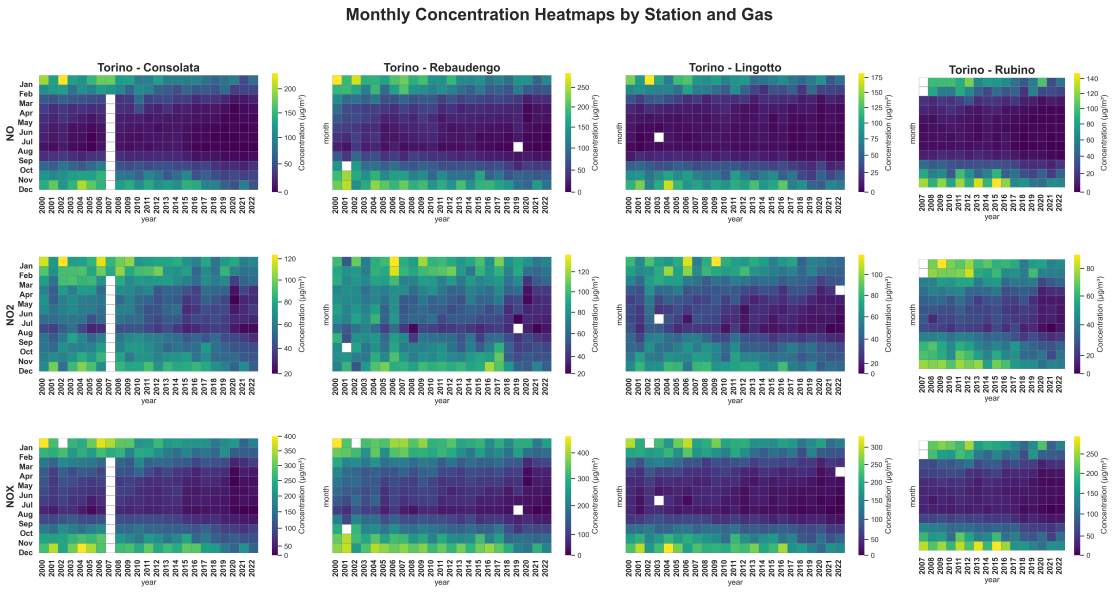

Figure 4.7: Monthly Concentration Heatmaps by Station and Gas

A quick guide to correctly interpret the heatmaps is reported.
The x-axis represents the years, while the y-axis shows the months of the year from January to December. The color scale on the right of each subplot indicates the concentration levels of the pollutants, with lighter colors (yellow) representing higher concentrations and darker colors (purple) representing lower concentrations. White squares indicate missing data for the corresponding month.

Specific observations from the heatmaps further clear up the key aspects of pollutant dynamics:

- **Seasonal Variability**: The heatmaps show a clear pattern of higher pollutant concentrations during the colder months (October to February), associated with increased emissions from domestic heating and stable weather conditions that trap pollutants close to the ground. This seasonal effect is consistent across all stations and pollutants.
- **Long-Term Declining Trends**: Despite seasonal peaks, the heatmaps also reveal a gradual overall decline in concentrations over the years (shadowing of plots from left to right), supporting the results obtained with Linear trend analysis, may due to the effectiveness of implemented air quality control measures. This trend is most visible at Torino-Rubino and Torino-Lingotto.
- **Comparison Across Pollutants**: The NO gas often exhibits sharper seasonal peaks compared to NO_2 and NO_X, that might reflect possible differences in emission sources or atmospheric lifetimes and dispersion.
- **Data Gaps**: Noticeable gaps in the data, especially at Torino-Consolata and Torino-Lingotto (reflect known periods of missing measurements) and Torino-Rubino that started the detection in March 2007.

4.1.4 Linear Trend Analysis of PM_{10} and $PM_{2.5}$

The trend analysis for PM10 and $PM_{2.5}$ was conducted using an approach similar to that applied to nitrogen oxides. The analysis was started by creating the `daily_aggregated` dataset for PM10 and PM2.5 data. In the `daily_aggregated` data, hourly frequency measurements are included but are aggregated into daily averages, computed only for days respecting the **min_valid_hours = 12** threshold discussed in chapter 2.
This choice was motivated by several factors:

- **Stability of aggregated data:** Aggregating the data on a daily level provides a more stable and uniformed approximation of pollutant levels, between hourly and daily sampled measurements, which reduces variability.
- **Data continuity and noise reduction:** Daily aggregation mitigates the frequent gaps present in hourly data, improving overall data continuity and reducing noise.

Similar reasoning could be applied to create `monthly_aggregated` and `seasonal_aggregated` datasets, which would provide a smoother analysis of long-term trends by further reducing short-term variability. However, these datasets are not included in the scope of this thesis.

A linear regression model was used to fit annual averages of PM concentrations (calculated for each year using daily measurement) from the year 2000 to 2022, focusing only on the measurements that provided the best temporal coverage, as outlined earlier.
Important Note: The selected measurements for PM_{10} were PM10_T, PM10_GBV, PM10TCON, PM10_B, and PM10_CP, while for $PM_{2.5}$ the analysis focused on the measurements PM2.5, PM2.5_B, and PM2.5_CP. This approach provides a more reliable and general overview of how particulate concentrations have changed over the period in analysis.

4.1.4.1 Station-specific Plots

The station-specific plots in Figures 4.8 and 4.9 allow to examine the general trend of particulate matter concentrations over each period where the measurements time series was present. As for Nitrogen Oxides, periods of missing data longer than one month are highlighted in the plots to indicate potential uncertainties. In the plots localized peaks may appear, potentially reflecting either episodic high pollution events or the influence of data gaps.

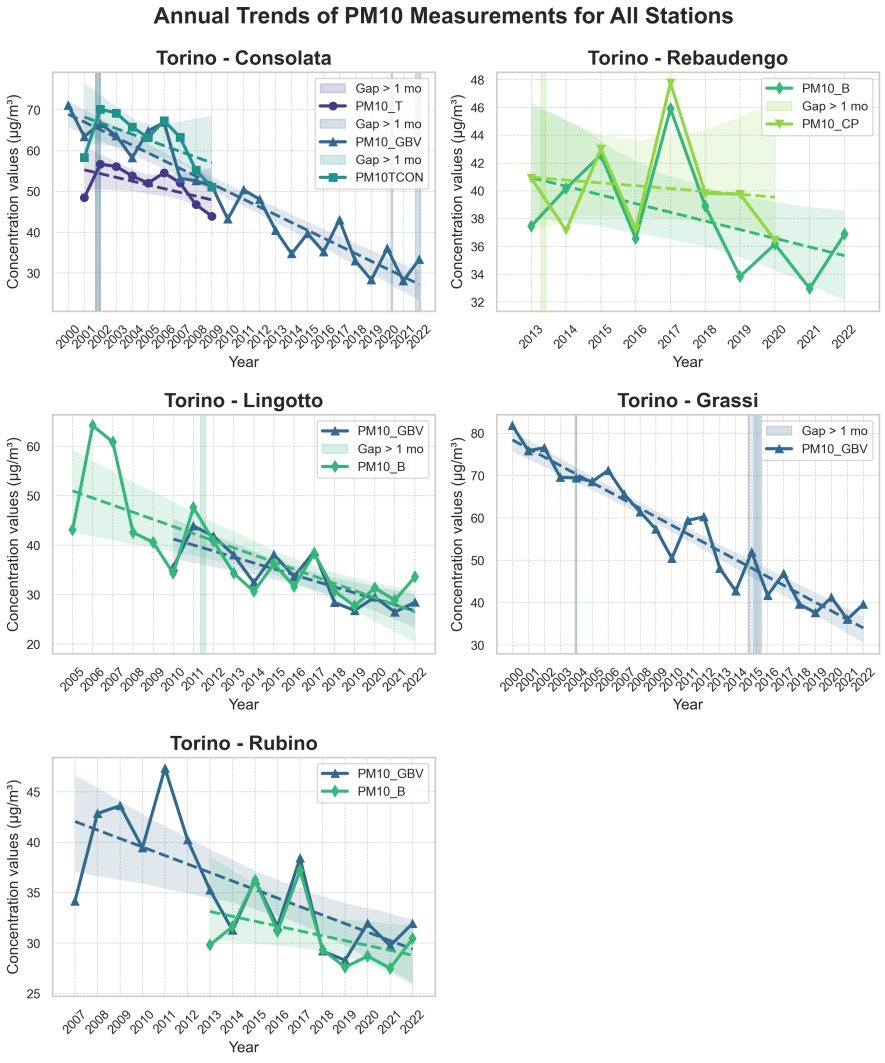

Figure 4.8: Annual Trends of PM_{10} Measurements for All Stations

Specific observations evidenced for PM_{10} for each station by Figure 4.8 are:

- **Torino-Consolata**: A decreasing trend is observed for all measure of PM_{10} concentrations, though some peaks are notable, especially at the edges of the periods. Large data gaps are also visible (e.g. 2001-2002, 2021-2022), which indicate potential uncertainties.
- **Torino-Rebaudengo**: The PM_{10} concentration levels are very fluctuating, particularly with a high peaks around 2017, but no presence of gap is highlighted for that year. This specific year's spike, confirmed by historical report from ARPA Piemontes, highlights a significant pollution event majorly attributed to unfavorable weather conditions [8].
- **Torino-Lingotto**: The trend displays a decrease in concentrations from 2005 onwards, with occasional spikes, mainly in the first period.
- **Torino-Grassi**: A consistent decrease is observed in PM_{10} levels, with more stable readings in periods 2000-2005 and 2016-2022. Significant several gaps appear in 2014–2015.
- **Torino-Rubino**: Both PM10_GBV and PM10_B concentrations show declining trends, however PM10_GBV displays significant fluctuations, particularly between 2007 and 2013. A more careful assessment may be needed for this period in order to address this behavior to possible incompleteness of the data, although no long large gaps are highlighted.

Figure 4.9: Annual Trends of $PM_{2.5}$ Measurements for All Stations

Specific observations evidenced for $PM_{2.5}$ for each station by Figure 4.9 are:

- **Torino-Rebaudengo**: The particle count measure does not provide a very effective visualization of a decreasing trend, however as for PM_{10} a peak in 2017 is highlighted. PM2.5_B shows a decline for the period 2013-2022.
- **Torino-Lingotto**: A clear decline is visible in $PM_{2.5}$ levels from 2005 to 2022, with several fluctuations principally between 2010 and 2015, probably related to gaps in the data, although smaller than one month.
- **Torino-Rubino**: This station show a slight decline in $PM_{2.5}$ concentration, though variability and occasional peaks are present.

4.1.4.2 Summary of Trend Analysis for PM_{10} and $PM_{2.5}$

To support the graphical annual trends, the regression statistics based on the OLS method are summarized in Table 4.2 and Table 4.3 for PM_{10} and $PM_{2.5}$ respectively. These results confirm the overall decrease particulate matter concentrations across the stations.

Table 4.2: Regression Statistics for PM_{10} by Station, for PM_{10}

Regression Statistics for PM10 by Station and Measurement

Station	Measurement	Slope	P-Value	R-Squared	Significance Level	Trend Direction
Torino - Consolata	PM10_T	-0.9204	0.1040	0.3327	ns	Decreasing
Torino - Consolata	**PM10_GBV**	-1.8972	0.0000	0.8959	**X X X**	**Decreasing**
Torino - Consolata	PM10TCON	-1.4046	0.0954	0.3465	ns	Decreasing
Torino - Grassi	**PM10_GBV**	-2.0187	0.0000	0.9249	**X X X**	**Decreasing**
Torino - Lingotto	**PM10_GBV**	-1.2095	0.0009	0.6478	**X X X**	**Decreasing**
Torino - Lingotto	**PM10_B**	-1.4411	0.0003	0.5638	**X X X**	**Decreasing**
Torino - Rebaudengo	PM10_B	-0.6244	0.1586	0.2320	ns	Decreasing
Torino - Rebaudengo	PM10_CP	-0.2047	0.7516	0.0180	ns	Decreasing
Torino - Rubino	**PM10_GBV**	-0.8440	0.0024	0.4933	**X X**	**Decreasing**
Torino - Rubino	PM10_B	-0.4813	0.2053	0.1920	ns	Decreasing

Table 4.3: Regression Statistics for $PM_{2.5}$ by Station, for $PM_{2.5}$

Regression Statistics for PM2.5 by Station and Measurement

Station	Measurement	Slope	P-Value	R-Squared	Significance Level	Trend Direction
Torino - Lingotto	**PM2.5**	-1.2667	0.0000	0.8840	**X X X**	**Decreasing**
Torino - Rebaudengo	**PM2.5_B**	-0.9031	0.0340	0.4490	**X**	**Decreasing**
Torino - Rebaudengo	PM2.5_CP	-0.1847	0.7371	0.0202	ns	Decreasing
Torino - Rubino	**PM2.5_B**	-0.6426	0.0356	0.4432	**X**	**Decreasing**

As in previous analyses, significance levels are based on p-values, with stronger significance indicating greater confidence in the observed trends, and in the tables, the measurement types that showed significant decreasing trends are highlighted in bold.

The overall trends for both PM_{10} and $PM_{2.5}$ are predominantly downward for almost all measurement type across the stations. However, it is worth noting that the regression for PM_{10} at the Torino-Rebaudengo station did not pass the p-value threshold. Therefore, the decreasing trend at this station cannot be confirmed

as statistically significant. This may be related to the fact that annual averages were used, but the time series for measurements at this station is relatively short, and the data exhibit considerable fluctuations. From the coverage analysis in Chapter 2, it is recalled that for this station, data coverage on valid days for PM_{10} measurements was 83.11% for PM10_B and 74.89% for PM10_CP, both among the lowest percentages of valid days for PM measurements. The presence of gaps, although smaller than one month, could still impact the robustness of the trend analysis and contribute to the reduced significance of the results. A special focus is addressed to this station later in the thesis through STL analysis.

4.1.4.3 Comparative Trends Across Stations

In order to give a clear, visual comparison overview of how particulate matter measurements have evolved across different stations, in Figures 4.10 and 4.11 are reported the time evolution of the annual averages of PM_{10} and $PM_{2.5}$. These figures offer a way to observe differences in pollutant levels and trends for the same measurements across stations, highlighting variations that can be associated to the spatial distribution of the monitoring points.

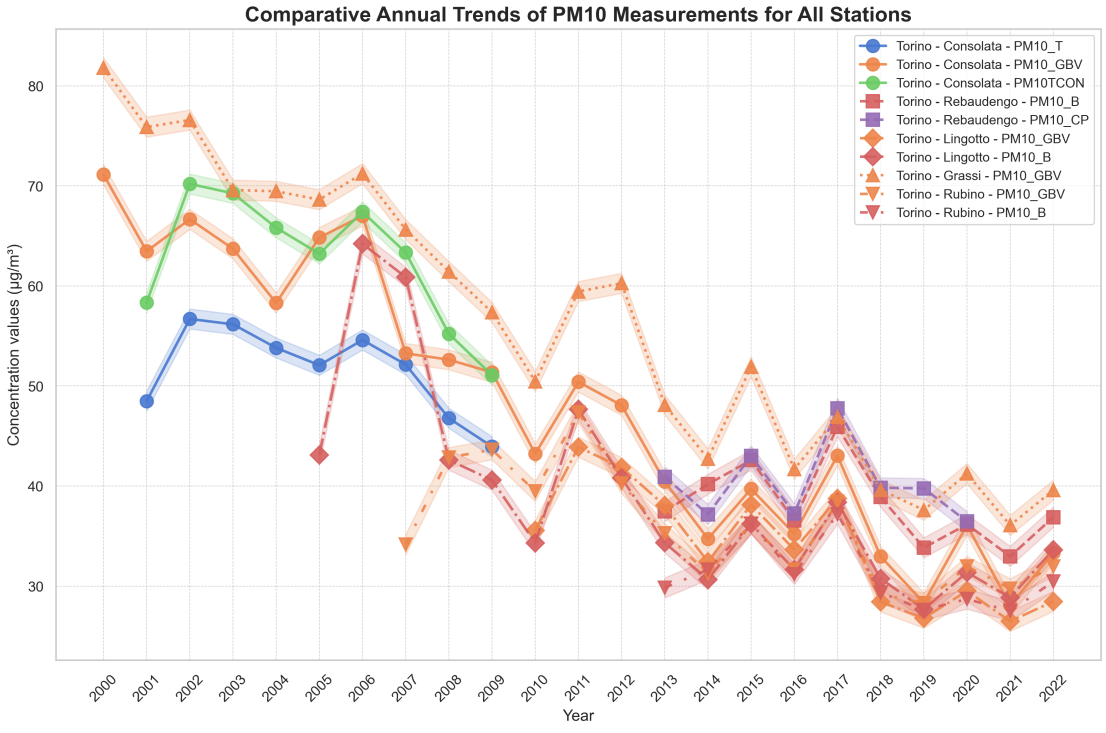

Figure 4.10: Comparative Annual Trends of PM_{10} Measurements

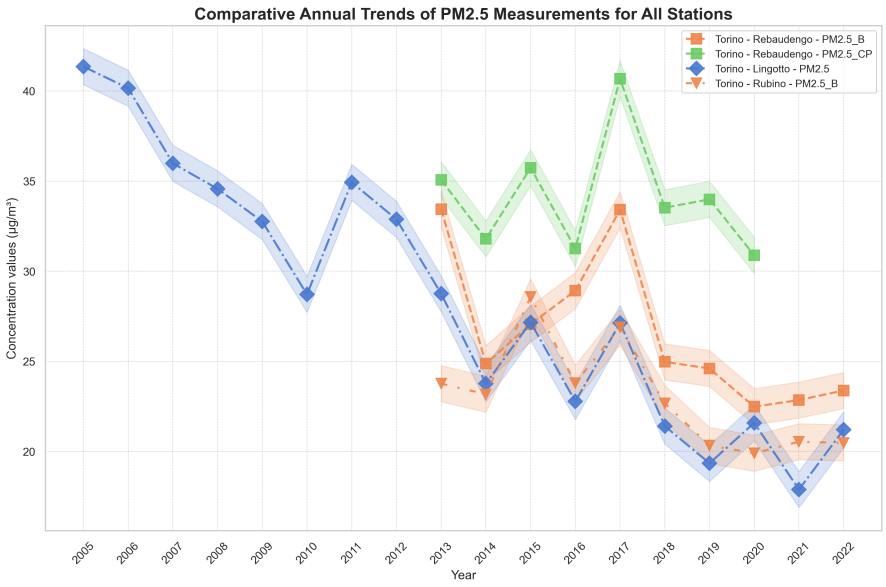

Figure 4.11: Comparative Annual Trends of $PM_{2.5}$ Measurements

Overall, as observed in the nitrogen oxides analysis, stations like Torino-Rebaudengo show higher pollution levels, while Torino-Rubino consistently reports the lowest. It is notable that Torino-Grassi consistently shows slightly higher concentrations of PM10_GBV compared to the Torino-Consolata measurement. This difference is particularly evident during the earlier years of the analyzed period, though the gap narrows in recent years since stations show a similar downward trend. For $PM_{2.5}$, Torino-Lingotto consistently shows the lowest and most reliable concentration levels. In contrast, Torino-Rebaudengo exhibits much more variability, as evidenced previously particularly in 2017, where both PM2.5_B and PM2.5_CP measurements show sharp peaks. Torino-Rubino shows $PM_{2.5}$ levels similar to those at Torino-Lingotto, with smaller fluctuations, especially in the most recent years.

4.2 Seasonal Trend Decomposition

The *Seasonal and Trend decomposition using LOESS (STL)* is a reliable and powerful method for time-series analysis [3], that allows to separate a data into three key components: trend, seasonal, and residual using *LOESS (Locally Weighted Scatterplot Smoothing)*, which works applying local regression. This methodology is particularly useful in this case study, especially in understanding long-term trends, identifying recurring seasonal patterns, and detecting irregular fluctuations that may indicate anomalies in the monitoring stations or extreme events that impacted air quality.

In is notable an analogy with the well known *Fourier Decomposition*, which breaks down the time-series in sinusoidal functions and assumes the seasonal component to be of fixed length. STL decomposition instead offers a better flexibility, as it allows the seasonal component to vary locally, making it well-suited for air quality data where periodic patterns may change over time, influenced by seasonal factors such as weather, traffic and human activities.

The STL decomposition used in this thesis is based on Python's *Statsmodels* library, which provides flexible and clear implementation of STL [5, 11].
A direct explanation of each component in the STL decomposition is as follows:

- **Trend Component**: reflects the underlying long-term variation of pollutant levels, indicating whether concentrations are generally increasing, decreasing, or stable over time.
- **Seasonal Component**: highlights recurrent seasonal patterns especially useful in air quality analysis.
- **Residual Component**: underlines irregular fluctuations or unexpected pollution spikes relative to anomalies or extreme events.

The following sections present the results of applying STL to analyze the time-series data, which was previously examined using linear trend analysis. Different graphical representations are used to further highlight the findings for nitrogen oxides and particulate matter measurements.

4.2.1 STL Results of NO, NO_2, NO_X

The STL decomposition for nitrogen oxides reproduced results similar to those obtained from the linear trend analysis, providing a clearer understanding of the trends across different monitoring stations.

As a general visual overview, Figure 4.12 shows only the trend component for NO, NO_2, and NO_X across the four stations, where also the average concentrations for the first and last year of the time period are highlighted.

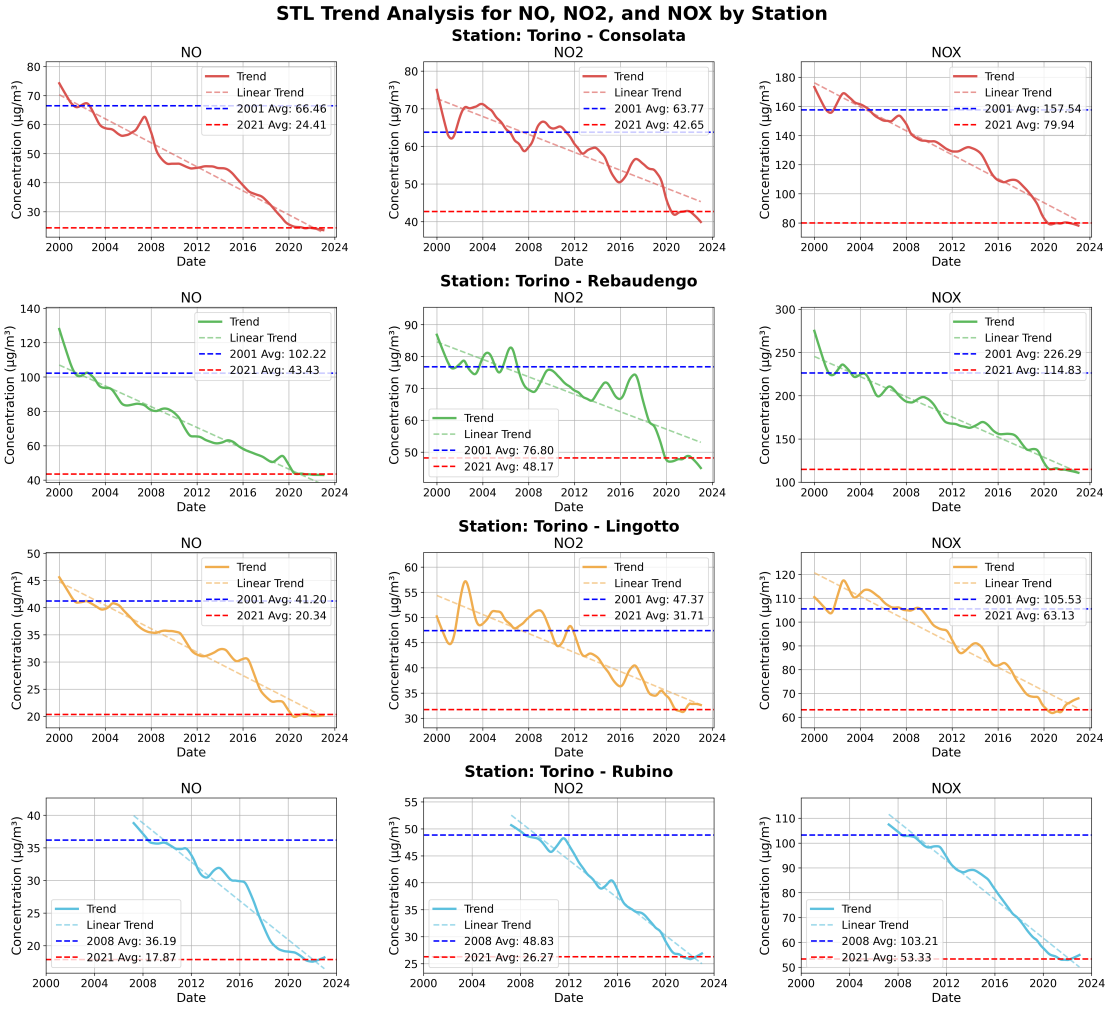

Figure 4.12: STL Trend Analysis for NO, NO_2, and NO_X by Station

As support of the previous graphical representation, the average concentrations at the start and end of the analysis period are reported in the Table 4.4.

Table 4.4: Summary of Average Concentrations at Start and End of Analysis Period

Summary of Average Concentrations at Start and End of Analysis Period

Station	Gas	First Year	First Year Avg (µg/m³)	Last Year	Last Year Avg (µg/m³)
Torino - Consolata	NO	2000	70.81	2022	24.03
Torino - Consolata	NO2	2000	68.60	2022	41.46
Torino - Consolata	NOX	2000	165.29	2022	79.36
Torino - Lingotto	NO	2000	43.90	2022	20.08
Torino - Lingotto	NO2	2000	47.35	2022	32.80
Torino - Lingotto	NOX	2000	108.03	2022	66.84
Torino - Rebaudengo	NO	2000	117.46	2022	43.02
Torino - Rebaudengo	NO2	2000	82.25	2022	47.05
Torino - Rebaudengo	NOX	2000	253.71	2022	112.46
Torino - Rubino	NO	2007	37.99	2022	17.83
Torino - Rubino	NO2	2007	50.14	2022	26.29
Torino - Rubino	NOX	2007	106.01	2022	53.81

Looking carefully at Table 4.4, it is evident that the average concentration of pollutant in the final year are significantly lower compared to the first year of analysis. It is important to note, however that the potential influence of missing data, may introduce some degree of error into the results. These values should be considered as reasonable estimates rather than definitive measurements of the actual year's average concentration.

The full STL decomposition results, including the seasonal and residual components, as long as the observed time series, is reported in Figure 4.13 for Torino-Rebaudengo. From the figure, the decreasing trends of nitrogen oxides are evident, and spikes and potential errors that may have affected the data are visible for both seasonal and residual component, especially for NO in the first period.

The STL results for the other stations are reported in the Appendix (Figures A.1, A.2, A.3), as they show similar patterns.

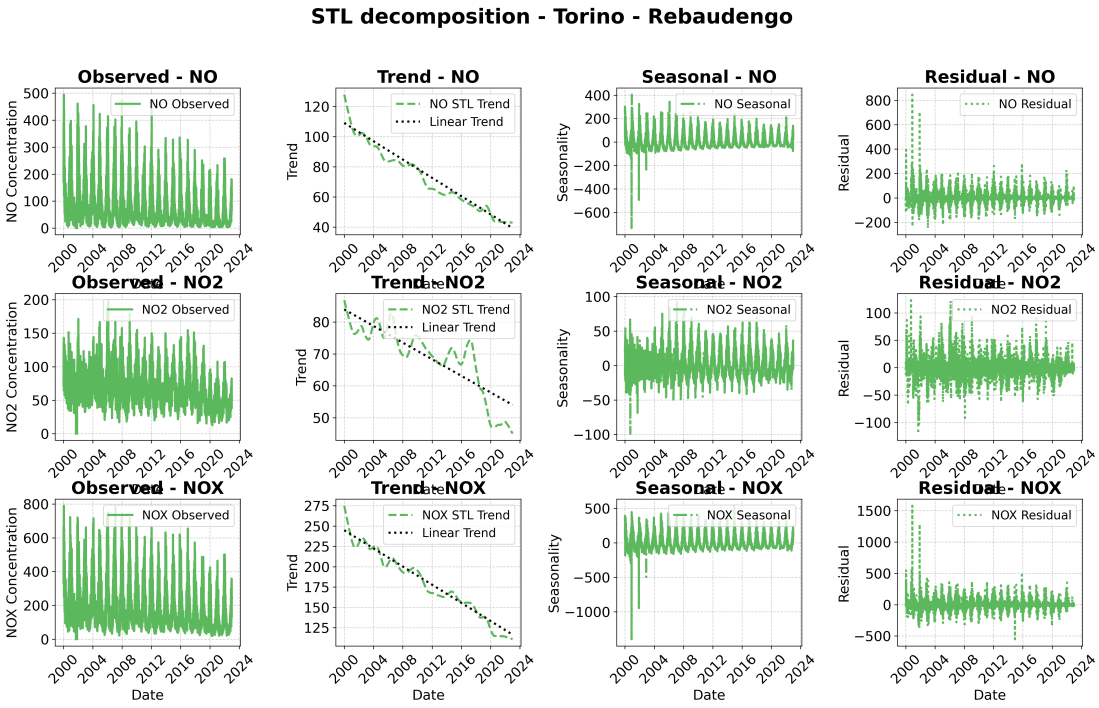

Figure 4.13: STL decomposition for Torino-Rebaudengo Station

Table 4.5, as support of the complete graphical representation, shows the summary of the key STL statistics for each pollutant and station. This table highlights interesting values such as the trend mean, seasonal amplitude, and the statistical significance level (p-values), which by a quick observation it appears ≈ 0.0. This finding underscore the reliability of the STL trend estimates, indicating the trends to be highly significant across all stations.

Table 4.5: STL Analysis Summary

STL Analysis Summary for NO, NO2, and NOX by Station

Station	Gas	Trend Mean	Trend Std Dev	Seasonal Amplitude	Residual Mean	Residual Std Dev	Trend Change Rate (Slope)	Trend Significance (p-value)	Significance Level	Trend Direction
Torino - Consolata	NO	46.50	13.98	388.31	1.32	29.03	-0.01	0.0e+00	X X X	Decreasing
Torino - Consolata	NO2	59.00	8.72	133.04	0.79	13.74	-0.00	0.0e+00	X X X	Decreasing
Torino - Consolata	NOX	128.86	28.00	617.37	2.94	53.08	-0.01	0.0e+00	X X X	Decreasing
Torino - Lingotto	NO	32.35	7.32	612.77	1.29	25.98	-0.00	0.0e+00	X X X	Decreasing
Torino - Lingotto	NO2	43.51	6.92	106.33	0.79	12.57	-0.00	0.0e+00	X X X	Decreasing
Torino - Lingotto	NOX	92.18	17.25	584.21	3.04	48.57	-0.01	0.0e+00	X X X	Decreasing
Torino - Rebaudengo	NO	72.12	20.51	1138.82	1.92	42.20	-0.01	0.0e+00	X X X	Decreasing
Torino - Rebaudengo	NO2	68.84	10.54	193.65	0.16	16.72	-0.00	0.0e+00	X X X	Decreasing
Torino - Rebaudengo	NOX	178.33	39.32	1937.69	2.29	75.69	-0.02	0.0e+00	X X X	Decreasing
Torino - Rubino	NO	28.20	6.98	379.36	1.27	25.90	-0.00	0.0e+00	X X X	Decreasing
Torino - Rubino	NO2	38.76	8.06	128.36	0.35	10.76	-0.00	0.0e+00	X X X	Decreasing
Torino - Rubino	NOX	80.81	18.02	622.12	3.05	49.18	-0.01	0.0e+00	X X X	Decreasing

As already shown by the linear trend analysis, the STL decomposition confirms that trends for NO, NO_2, and NO_X are decreasing in all monitoring stations from 2000 to 2022.

4.2.2 STL Results of PM_{10} and $PM_{2.5}$

This section presents the results of the STL decomposition applied to `daily_aggregated` PM10 and PM2.5 concentration data using a 365-day cycle, in analogy to the nitrogen oxides analysis, and consistent with the methodology used in the particulate matter linear trend analysis.
Applying STL decomposition to monthly or seasonal data could offer additional insights by smoothing out daily noise; however, this would result in fewer data points, potentially diminishing the statistical reliability of the residual analysis and reducing sensitivity to short-term anomalies.

For each type of measurement and station, the analysis period was determined by identifying the first and last valid data points within the time series.
As a initial approach to the STL decomposition for PM, the analysis was conducted on the raw data, for which not yet the temporal integration of perfect correlated measurement was performed (as discussed in Chapter 3). However, only type of measurements with time-series of at least 365 valid data points were included in the analysis. This criterion ensures that the seasonal analysis is robust and reliable, and does take into account also measurements with not very extended data.

The measurements that did not meet this threshold and were excluded from the STL representation due to insufficient data were: for Torino-Consolata PM10_GAV with 278 valid data points, for Torino-Lingotto PM10_N with 59 valid data points, for Torino-Rubino PM10_B with 57 valid data points and for $PM_{2.5}$ Torino-Rubino PM2.5_B with only 58 valid data.

Below, the full STL decomposition results for Torino-Rebaudengo station are shown in Figures 4.14 and 4.15 for PM_{10} and $PM_{2.5}$ respectively. Particular attention should be paid to this station, as the linear trend analysis previously indicated that the station did not exhibit a significant decreasing trend for the PM_{10} measurements.

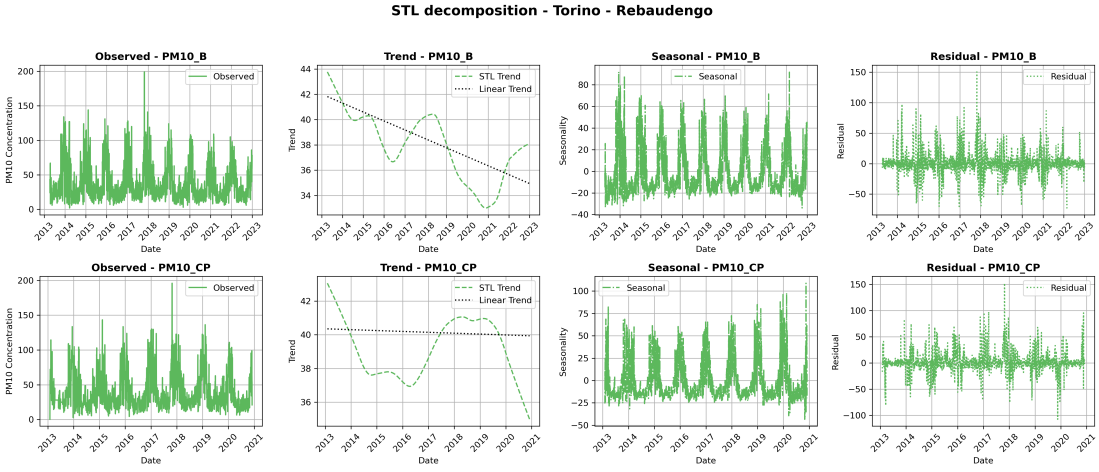

Figure 4.14: STL decomposition for PM_{10} at Torino-Rebaudengo Station

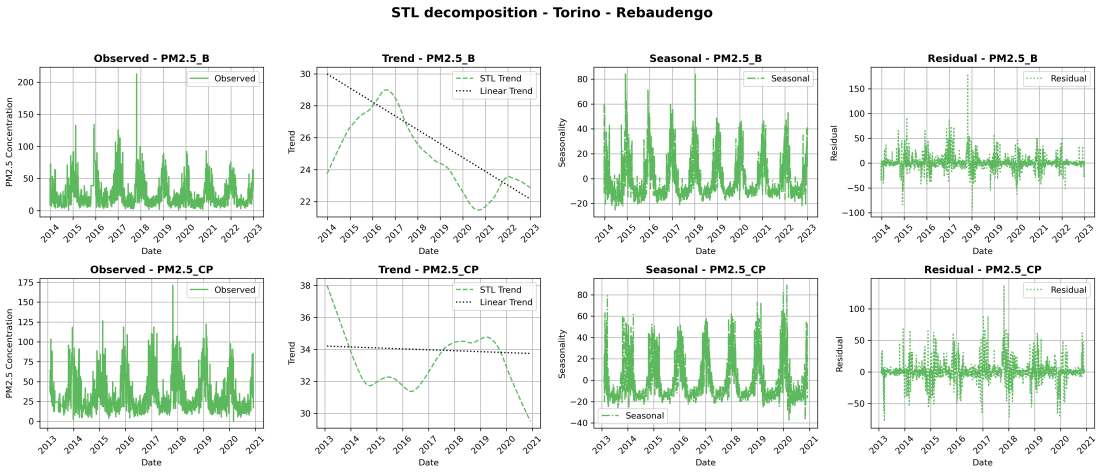

Figure 4.15: STL decomposition for $PM_{2.5}$ at Torino-Rebaudengo Station

These figures shown above highlight distinct seasonal fluctuations with sharp peaks during the colder months and lower values in the warmer periods for Torino-Rebaudengo station. Looking at the trend component, there is lack of clarity especially at the edges of the considered period, that might reflect some possible missing data which might play a more significant role in influencing the PM levels. Relying only on the graphical trend line for conclusions on the variation of the pollution levels might be insufficient, so as support the statistics from the STL analysis are reported later in a table.

A particularly notable observation is the great similarity in the trend and seasonal components between PM10_CP and PM2.5_CP measurements. Both exhibit nearly identical patterns across all STL components, which suggest a strong reliability in the measurement process at this station. The main difference lies in the magnitude of the particulate levels, with PM_{10} consistently having higher concentrations than $PM_{2.5}$, as expected given the larger particle size.

For further observations and graphical representations of the complete STL analysis for measurements by other monitoring stations, refer to Appendix (B.4).

As summary of the STL analysis for particulate matter measurements, in Table 4.6 and Table 4.7, the complete STL decomposition results including key quantities such as trend direction and time series length (in days), are reported for all stations. In addition, the average trend of PM_{10} and $PM_{2.5}$ pollutant in each station were calculated. For each station, the main averages quantities are reported and the general direction of the trend calculated by weighting the trend values according to the length of the time series for each measurement.
This approach was choose to ensure that longer datasets have a greater influence on the station's overall trend, with a more accurate reflection of particulate matter levels variations.

Table 4.6: STL Analysis Summary for PM_{10}

Station	Measure	Trend Mean	Trend Std Dev	Seasonal Amplitude	Residual Mean	Residual Std Dev	Trend Change Rate (Slope)	Trend Significance (p-value)	Significance Level	Trend Direction	Series Length
Torino - Consolata	PM10TCON	59.603649	4.945663	161.319513	2.137718	23.284354	-0.001142	0.000000	X X X	Decreasing	3287
Torino - Consolata	PM10_GBV	46.354968	13.007349	250.858165	1.765791	22.924017	-0.005212	0.000000	X X X	Decreasing	8399
Torino - Consolata	PM10_T	49.643997	3.347644	109.316436	1.300418	15.644986	-0.000723	0.000000	X X X	Decreasing	3287
Torino - Consolata Average		51.867538	7.100219	173.831372	1.734642	20.617786	-0.002359	0.000000		Decreasing	
Torino - Rebaudengo	PM10_B	37.918098	2.653434	125.850828	1.068823	16.222361	-0.001894	0.000000	X X X	Decreasing	3552
Torino - Rebaudengo	PM10_CP	39.150065	1.782227	152.022604	1.637483	17.790283	-0.000270	0.000000	X X X	Decreasing	2863
Torino - Rebaudengo Average		38.534082	2.217830	138.936716	1.353153	17.006322	-0.001082	0.000000		Decreasing	
Torino - Lingotto	PM10_B	37.631063	8.859841	161.068053	1.076811	18.928232	-0.003985	0.000000	X X X	Decreasing	6526
Torino - Lingotto	PM10_B_p	43.587935	1.483364	180.748057	-0.272303	1.611091	0.004172	0.000000	X X X	Increasing	779
Torino - Lingotto	PM10_GBV	31.983614	4.424287	166.308287	1.869157	18.770775	-0.003023	0.000000	X X X	Decreasing	4742
Torino - Lingotto Average		37.734204	4.922498	169.374799	0.891221	13.103366	-0.000945	0.000000		Decreasing	
Torino - Grassi	PM10_GBV	54.525176	13.368594	194.819620	1.228701	23.167922	-0.005383	0.000000	X X X	Decreasing	8401
Torino - Grassi Average		54.525176	13.368594	194.819620	1.228701	23.167922	-0.005383	0.000000		Decreasing	
Torino - Rubino	PM10_BD	28.184026	1.624858	164.074367	0.980014	14.106430	0.003080	0.000000	X X X	Increasing	1461
Torino - Rubino	PM10_BH	30.233817	1.798076	102.288329	1.509379	14.875609	-0.001018	0.000000	X X X	Decreasing	3523
Torino - Rubino	PM10_GBV	34.187457	4.067481	173.085720	1.757313	19.323932	-0.002122	0.000000	X X X	Decreasing	5756
Torino - Rubino Average		30.868434	2.496805	146.482805	1.415569	16.101990	-0.000020	0.000000		Decreasing	

Table 4.7: STL Analysis Summary for $PM_{2.5}$

Station	Measure	Trend Mean	Trend Std Dev	Seasonal Amplitude	Residual Mean	Residual Std Dev	Trend Change Rate (Slope)	Trend Significance (p-value)	Significance Level	Trend Direction	Series Length
Torino - Rebaudengo	PM2.5_B	24.943197	2.265435	109.372493	1.346885	13.651931	-0.001758	0.000000	X X X	Decreasing	3296
Torino - Rebaudengo	PM2.5_CP	33.215073	1.648411	127.332939	1.321565	15.460621	-0.000447	0.000006	X X X	Decreasing	2862
Torino - Rebaudengo Average		29.079135	1.956923	118.352716	1.334225	14.556276	-0.001102	0.000003		Decreasing	
Torino - Lingotto	PM2.5	26.382080	6.145723	126.119491	1.699607	16.215273	-0.003126	0.000000	X X X	Decreasing	6381
Torino - Lingotto	PM2.5_B	22.269302	1.759350	77.505780	0.027865	0.130093	0.008188	0.000000	X X X	Increasing	742
Torino - Lingotto Average		24.325691	3.952536	101.812635	0.863736	8.172683	0.002531	0.000000		Decreasing	
Torino - Rubino	PM2.5_BD	19.344459	0.818078	187.712757	0.634308	10.646136	0.001346	0.000000	X X X	Increasing	1461
Torino - Rubino	PM2.5_BH	22.365686	2.035756	96.391583	0.810723	11.502527	-0.001839	0.000000	X X X	Decreasing	3523
Torino - Rubino Average		20.855073	1.426917	142.052170	0.722515	11.074331	-0.000246	0.000000		Decreasing	

The summary tables confirm the overall trend, previously obtained from the Linear Trend analysis, to be decreasing for all monitoring stations in Turin. Even for Torino-Rebaudengo, where the linear trend was initially unclear, the STL decomposition allowed for a clearer interpretation, showing a consistent decreasing trend for all PM_{10} measurements. Noteworthy is the fact that for most measurement types with good data coverage, the analysis was reliable and returned significantly trends with very high significance levels.
Although gaps were not taken into account, still aware of their presence (some of considerable duration), the methodology used was still able to provide a good approximation of long-term trends.

Chapter 5

Model-Data comparison

The goal of this part of the thesis is to explore the potential of the SIRANE model in simulating pollutant dispersion, particularly in the urban area of Turin. This analysis provides an initial evaluation of the model's ability to represent the spatial and temporal variations of key pollutants, including NO, NO_2, NO_X, PM_{10}, and $PM_{2.5}$. The focus is on understanding how the model performs under the specific conditions of Turin. This chapter presents a summary of the approach taken with the SIRANE, along with the key results obtained, setting up a methodology that makes use of the air quality measurements discussed earlier in the thesis, crucial for validating and calibrating the model, but not necessarily direct inputs for running the model itself.

5.1 Introduction to SIRANE model

The SIRANE model [13] was developed by the École Centrale de Lyon to calculate the concentrations of pollutants in the street network of a city. It is one of the few air quality (AQ) tools specifically designed to account for urban canopy effects, making it highly responsive to road traffic emissions. The model provides pollutant concentration estimates with an hourly time resolution and a spatial resolution down to meters.
SIRANE model has been validated both in wind tunnels [10] and either in real context and, in the last decade, SIRANE has been applied continuously in various European urban areas, with some pilot studies developed in Italy mainly as master degree reports. It is currently used by many French's public authorities (Lyon, Le Havre, Paris, Saint-Étienne) and within research projects.

SIRANE's main application are: evaluate population exposure; build cartographies at the district scale; determine the representativeness of monitoring stations, predict

pollutants temporal and spatial pathways, including peaks and evaluate the impact of new urban policy. The model is optimised to calculate UAQ on target domains with length scales up to a few kilometres and a 10-metre spatial resolution. The main innovation of SIRANE is that it is based on the concept of the street network, which is dependent on the buildings' volume, the spacing and is able to capture the main characteristics of the city. The basic element of SIRANE geometry relies on the concept of the open street and the street canyon, i.e. a cavity between buildings with a specific ratio among width (W) and height (H), where W is the width of the street and H is the height of the surrounding buildings. A street is classified as a canyon when the W/H ratio is 3 or less, while a ratio greater than 3 defines an open street. This distinction allows the model to adjust for different dispersion behaviors based on the urban layout.

Despite its strengths, the SIRANE model holds some limitations. For each time step pollutant dispersion is computed assuming steady conditions and concentration are estimated independently from that in the previous periods. In particular, in case of calm wind conditions, the model fails to adequately account for the accumulation of pollutants over the urban area. Moreover, SIRANE needs a large input data set and, in order to build this data set, three main problems must be taken into account: characterisation of the complexity of the geometry of the domain, estimate of the intensity and spatial distribution of the pollutant sources and adoption of the crucial parameters to describe the meteorological conditions [13].

The setup for this study benefited from existing previous pilot studies that had already adapted the model to Turin's urban environment (Bo M., 2020) [1]. These included SIRANE versions v.1, v.2, and v.3, which used emissions data based on 2004 fluxes combined with 2014 vehicular fleet emission factors. Additionally, the meteorological data, receptor datasets, and uniform background concentrations (using data from the ARPA station of Vinovo) all corresponded to the year 2014. A further refinement of the model can be made, using more recent air quality data.

5.2 Analysis Overview

This study utilized the SIRANE model to simulate pollutant concentrations over three distinct periods: January 1st, April 1st, and August 1st, 2014, each covering a 24-hour time frame. PM_{10}, and $PM_{2.5}$. The work focused on three areas in Turin, referred to as A1, A2, and A3, respectively. Area A1 corresponds to the zone around Torino-Lingotto, A2 to the zone around Torino-Consolata, and A3 to the zone around Torino-Rebaudengo. Each of these areas defines a circular zone with a 1 km radius centered around the monitoring stations. The reason for this choice was to include areas that differ in urban structure, as previously discussed, and to compare the average pollution levels across the defined street network with the measurements at the specific monitoring stations. For this case study , the areas were further classified by counting the number of street canyons and open streets in each zone, as detailed later in the analysis.

The selection of time periods was made based on the exploratory nature of the study, and the significant computational resources required to run the SIRANE model for extended periods. Simulating pollutant dispersion over longer periods, such as several weeks or months, requires running large data sets, multiple input parameters, and performing complex calculations to reflect changes in pollutant dispersion. These requirements have made it difficult to run longer simulations. Focusing on shorter periods, lasting 24 hours, provided an initial understanding of the model's capabilities.

5.3 Methodology and Data Processing

The whole urban graph network representation of Turin, present in SIRANE (`/RESULTS/reseau-rues.shp` file) was imported into QGIS software to facilitate the selection of specific areas for analysis. The streets in the model are identified by an ID that groups together streets with identical characteristics. This means that the ID is not unique; however, by visualizing the data in QGIS, it is possible to verify that duplicate IDs represent segments of the same road section, ensuring coherent geographic reference.

A summary of the selected areas, including the number of streets and their classification as either street canyons or open streets, is provided in Table 5.1.

Table 5.1: Summary Selected Study Areas Street Characteristics in SIRANE

Area	**Total Number of Streets**	**Number of Street Canyons (Type=0)**	**Number of Open Streets (Type=1)**	**Percentage of Street Canyons**
reseau_rue (Full urban area)	19152	7477	11675	0.39
A1 (Torino-Lingotto)	389	162	227	0.42
A2 (Torino-Consolata)	842	642	200	0.76
A3 (Torino-Rebaudengo)	405	166	239	0.41

It is observed that the Areas A1 (Torino-Lingotto) and A3 (Torino-Rebaudengo) have a similar proportion of street canyons, with around 40% of their streets classified as canyons. Area A2 (Torino-Consolata), on the other hand, has a significantly higher ratio of street canyons (0.76%), indicating a dense urban environment with many narrow streets, compatible with Turin's city center. In such areas, pollutants tend to become trapped for extended period due to reduced airflow and limited dispersion pathways, which can result in higher pollutant concentration levels.
The overall street network of Turin has a lower ratio of street canyons (0.39%), suggesting a more mixed urban layout considering the entire city, often characterized by wider roads and green spaces.

The pollutant concentrations simulated by the model for each street are obtained from files stored in the `/RESULTS/RUES_PAR_RUE` directory, with files named using the format `Rues_id2_id1.dat`. Each of this files refers to a street, and in the naming convention scheme, `id2` represents the road section, while `id1` distinguishes different segments within the same section. Also, each file contains hourly concentration data and deposition rates for a range of pollutants, which can be selected based on the analysis requirements. For this study, the focus was limited to nitrogen oxides and particulate matter simulated concentration.

5.4 Analysis Results and Model Evaluation

As previously discussed in the analysis overview, the SIRANE software was used to run simulations of 24-hour time frame, for three specific days in 2014: January 1st, April 1st, and August 1st. These days were selected as an initial and general approach to capture seasonal variation, representing winter, spring, and summer conditions. It is important to note that the results from these simulations are intended to give insight for a preliminary descriptive analysis, rather than providing definitive conclusions. The limited time periods means that the findings offer insight into the model's behavior under varied seasonal conditions, but should be interpreted with caution since not fully representative of long-term seasonal trends. For instance, while January 1st is characterized by typical winter conditions, it also coincides with a public holiday, which may influence traffic and other sources of pollution, making it less representative of typical winter pollution levels.

The results from the SIRANE model are compared with the corresponding measurements from the monitoring stations located within the study areas A1 (Torino-Lingotto), A2 (Torino-Consolata), and A3 (Torino-Rebaudengo). Particularly, the comparison is performed by taking the simulated concentrations on each specific point of the receptor within the selected areas for each hour, and comparing these hourly simulated concentrations to the corresponding hourly measurements recorded at the correspondent monitoring stations.

The following analysis is organized into two main parts: first, the evaluation of mean concentrations of NO, NO_2, NO_X, PM_{10}, and $PM_{2.5}$ across each study area for the selected days is reported; and second, the comparison of the simulated concentrations with the hourly measurements recorded at the monitoring stations within each area. This approach aims to provide a methodology approach to test how well the model captures the spatial patterns of pollutant distribution in the study areas, as well as its accuracy in replicating observed temporal variations and peaks at specific monitoring points.

5.4.1 Area-Averaged Concentrations Results

Figure 5.1, Figure 5.2, and Figure 5.3 present the daily trend of the concentration simulated by SIRANE in each street, and averaged for the three study areas: A1 (Torino-Lingotto), A2 (Torino-Consolata), and A3 (Torino-Rebaudengo). In the plots, different colors are used to represent each of the three areas, and the legend displays the average concentration level for the entire day for the pollutant being analyzed in each respective area. This results provides a general overview of the average air pollutant concentrations simulated by the model in the selected areas, highlighting the model's ability to reproduce the characteristic pollutant trends during the day.

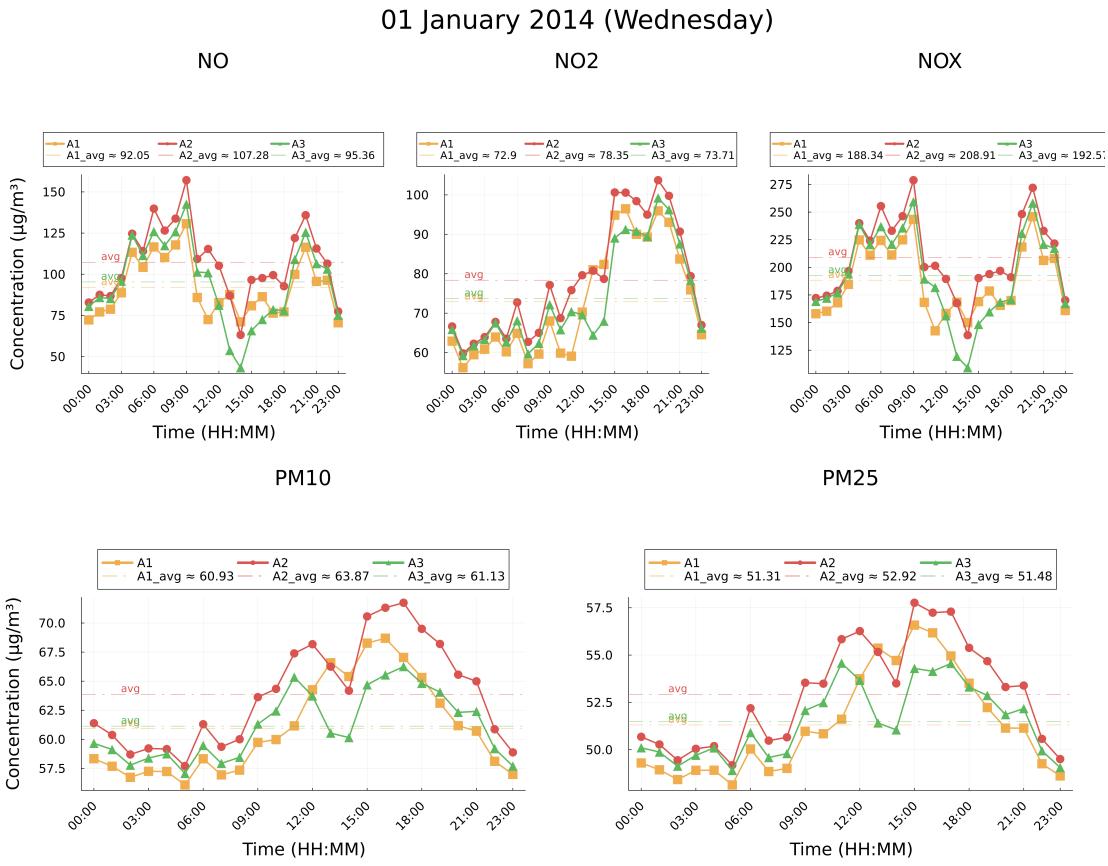

Figure 5.1: Average pollutant concentrations for January 1st, 2014.

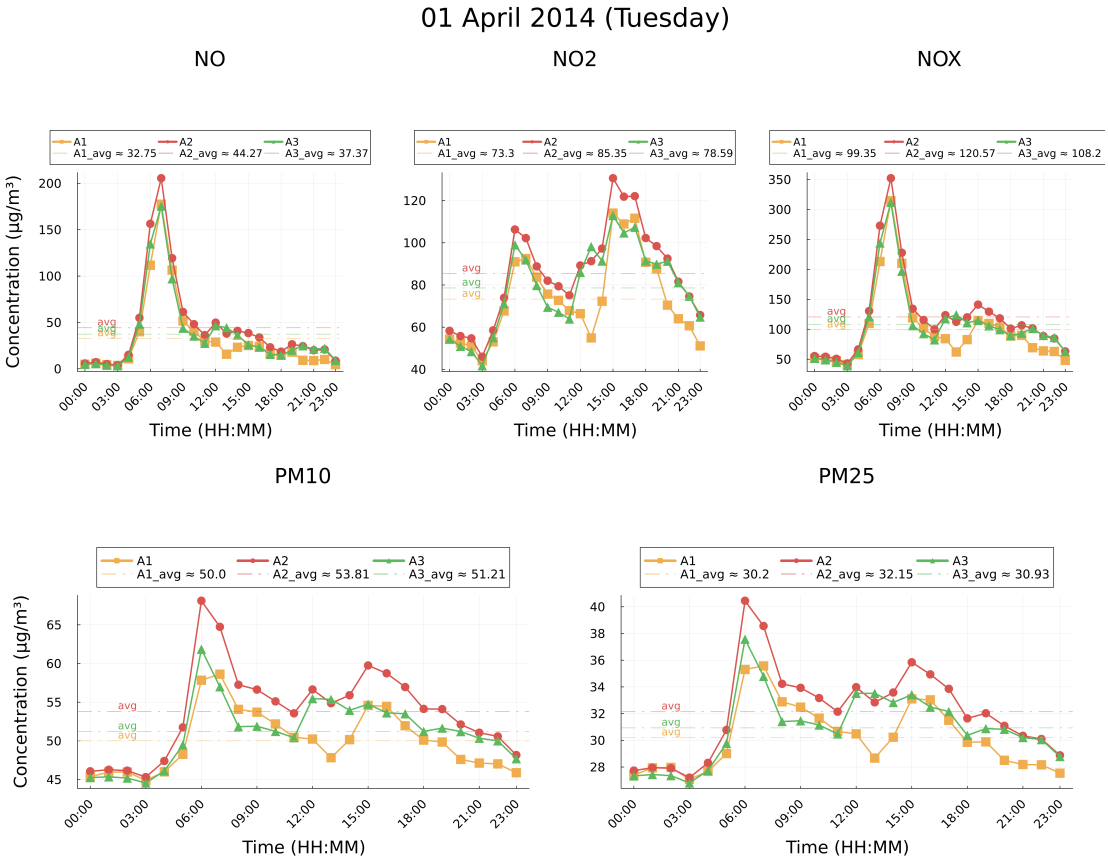

Figure 5.2: Average pollutant concentrations for April 1st, 2014.

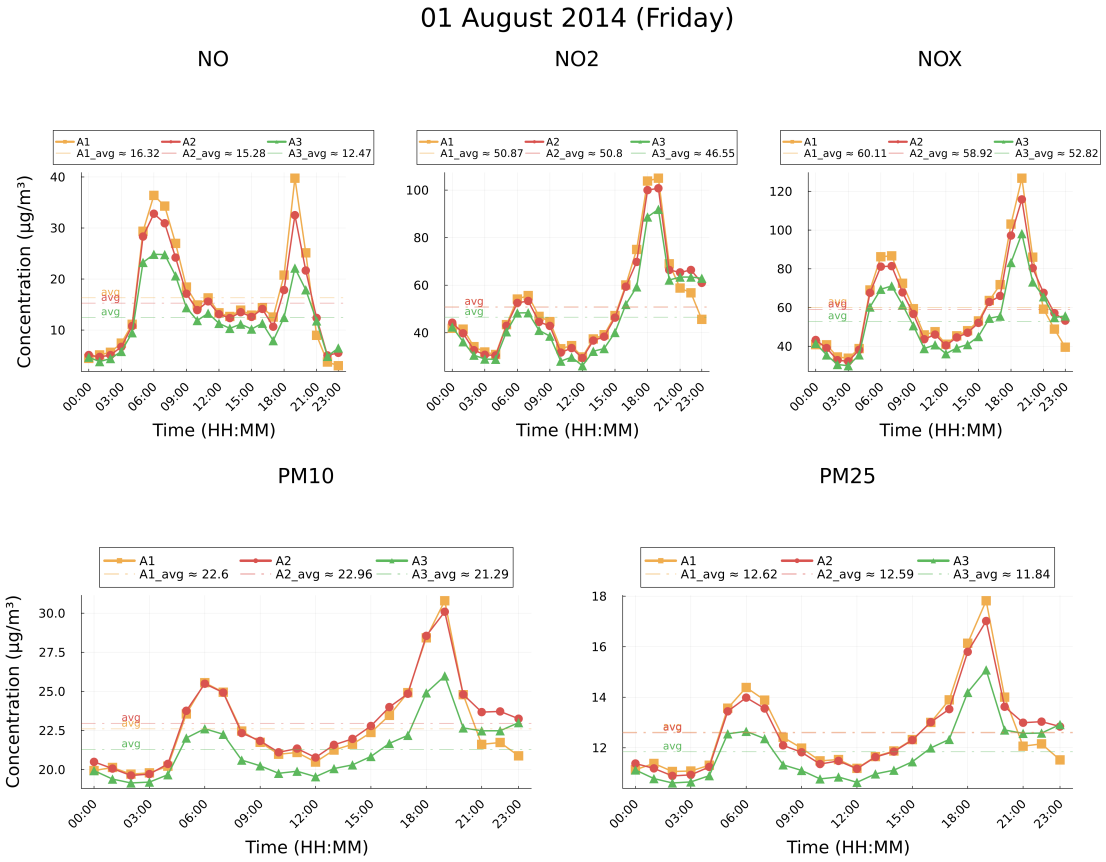

Figure 5.3: Average pollutant concentrations for August 1st, 2014.

Despite the limited period examined for the simulations, a clear idea of how pollutant gas concentrations change over the course of the day can already be obtained from the plots. In particular, no major differences between different areas are identified, yet expected trends are well present, occurring with some minor differences for each of the three chosen periods. With regard to concentrations of nitrogen oxides, of which one of the main emission sources is high-temperature combustion from on-road vehicles, peaks are observed during peak traffic hours, especially in the morning (between 06:00 and 09:00) and in the evening (between 18:00 and 21:00).
When observing particulate matter, PM_{10} and $PM_{2.5}$ show quite similar trends across the selected days, possibly due to the fact that both particulate matter fractions share common sources or are influenced by similar environmental factors.

Table 5.2 summarizes the average pollutant levels across the three areas for each selected day in 2014, showing a clear decrease in concentrations from winter (January) to summer (August).

Table 5.2: Mean Pollutant Levels (µg/m^3) in Three Areas on Selected Days in 2014

Day	**Area**	**NO**	**NO_2**	**NO_X**	**PM_{10}**	**$PM_{2.5}$**
January 1st	**A1 (Torino-Lingotto)**	92.05	72.9	188.34	60.93	51.31
	A2 (Torino-Consolata)	107.28	78.35	208.91	63.87	52.92
	A3 (Torino-Rebaudengo)	95.36	73.71	192.57	61.13	51.48
April 1st	**A1 (Torino-Lingotto)**	32.75	73.3	99.35	50.0	30.2
	A2 (Torino-Consolata)	44.27	85.35	120.57	53.81	32.15
	A3 (Torino-Rebaudengo)	37.37	78.59	108.2	51.21	30.93
August 1st	**A1 (Torino-Lingotto)**	16.32	50.87	60.11	22.6	12.62
	A2 (Torino-Consolata)	15.28	50.8	58.92	22.96	12.59
	A3 (Torino-Rebaudengo)	12.47	46.55	52.82	21.29	11.84

Overall, two main aspects can be observed from the area-averaged concentration plots:

Area Differences: Across all periods, Area A2 (Torino-Consolata) consistently shows higher pollutant concentrations compared to the other areas. Across all periods, Area A2 (Torino-Consolata) consistently exhibits higher pollutant concentrations compared to the other areas. Recalling that A2 has the highest density of street canyons, this indicates that the model is accurately capturing the reduced dispersion of pollutants in such dense urban environments. Area A3 (Torino-Rebaudengo) generally exhibits lower pollutant levels, reflecting better dispersion due to its more open street layout.
However, as previously discussed in the thesis, the Torino-Rebaudengo monitoring station is located in an area characterized by significant traffic density. The result suggest that the model may not fully capture the impact of localized traffic emissions in this area. Further and longer simulations are needed to clarify this aspect, focusing more accurately on the emission sources specified as inputs to the model.

Trend Variations: There is a clear seasonal trend for all pollutants, with the

highest concentrations occurring on January 1st, likely due to increased heating. Pollutant levels decrease notably in April. The lowest concentrations are observed on August 1st, especially for nitrogen oxides, compatible with reduced traffic during the holidays and better atmospheric conditions for pollutant dispersion. Overall the daily trend suggest that the model is capable of representing the expected diurnal variations in pollutant levels.

5.4.2 Comparison of Simulated and Measured Data

To evaluate the performance of the SIRANE model, simulated hourly concentrations were compared with measurements from the monitoring stations in each study area. The analysis is not taking in consideration the PM_{10} and $PM_{2.5}$ pollutant concentration, since these are only available as daily averages at the stations.
The comparison is conducted for January 1st, April 1st, and August 1st, 2014. By using a graphical representation, measured data (denoted as "Mes") and simulated data (denoted as "Mod") of concentration levels for NO, NO_2 and NO_X are shown for each area. Receptors are represented as R1, R2, and R3, corresponding to the monitoring stations in Area A1 (Torino-Lingotto), Area A2 (Torino-Consolata), and Area A3 (Torino-Rebaudengo), respectively. The legend also reports the daily average for both measured and simulated concentrations. In addition, error calculations, such as normalized absolute differences, are used to assess the model's performance in replicating measured concentrations. This approach gives insight of how well the model aligns with real-world pollution levels.

Figure 5.4, Figure 5.5, and Figure 5.6 show the comparison for January 1st, where both the measured and modelled concentration levels are reported along with the normalized difference between the two. The plots for April 1st and August 1st, 2014, are not included since they exhibit similar behavior between the model and measurements to those observed on January 1st.

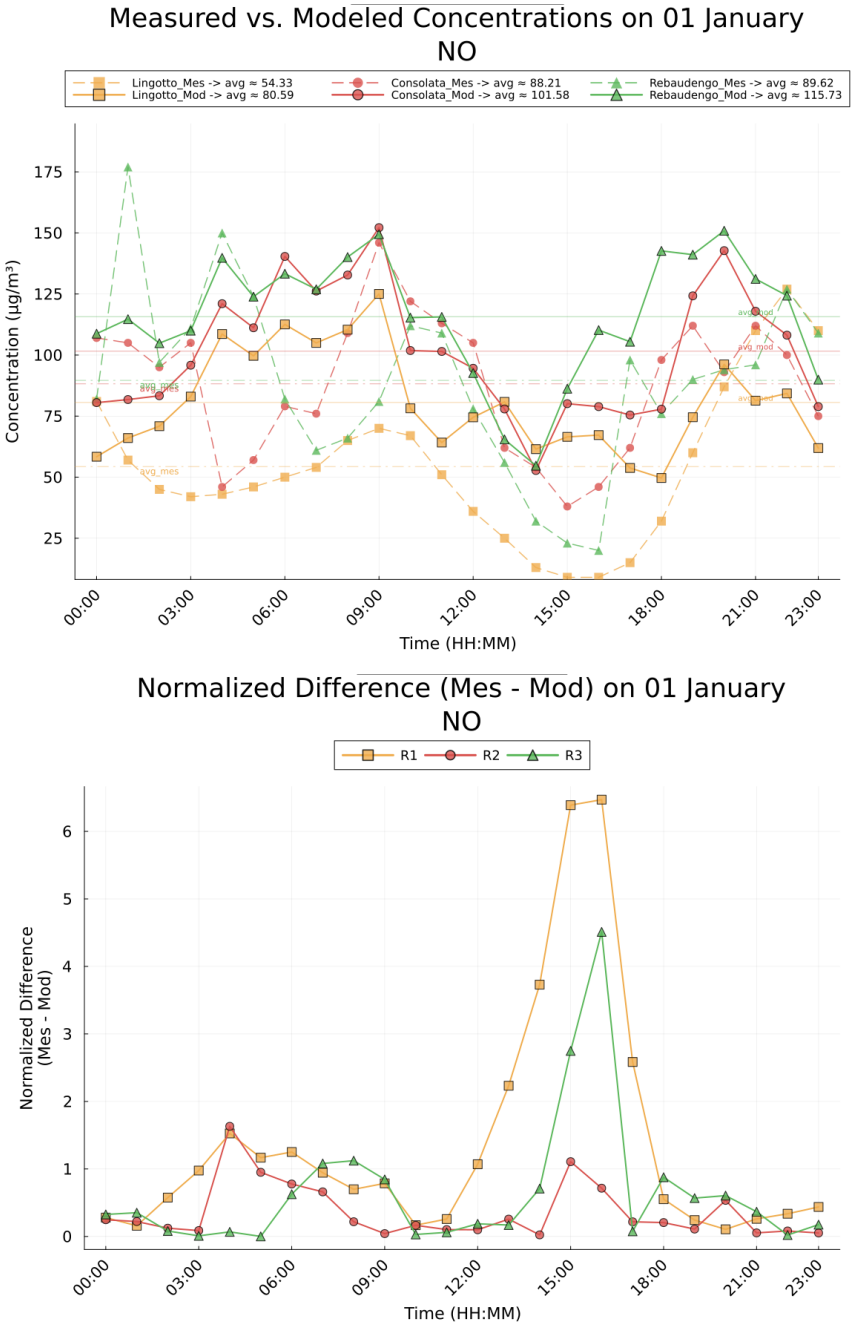

Figure 5.4: Measured and modelled NO concentrations, with the normalized difference between them below.

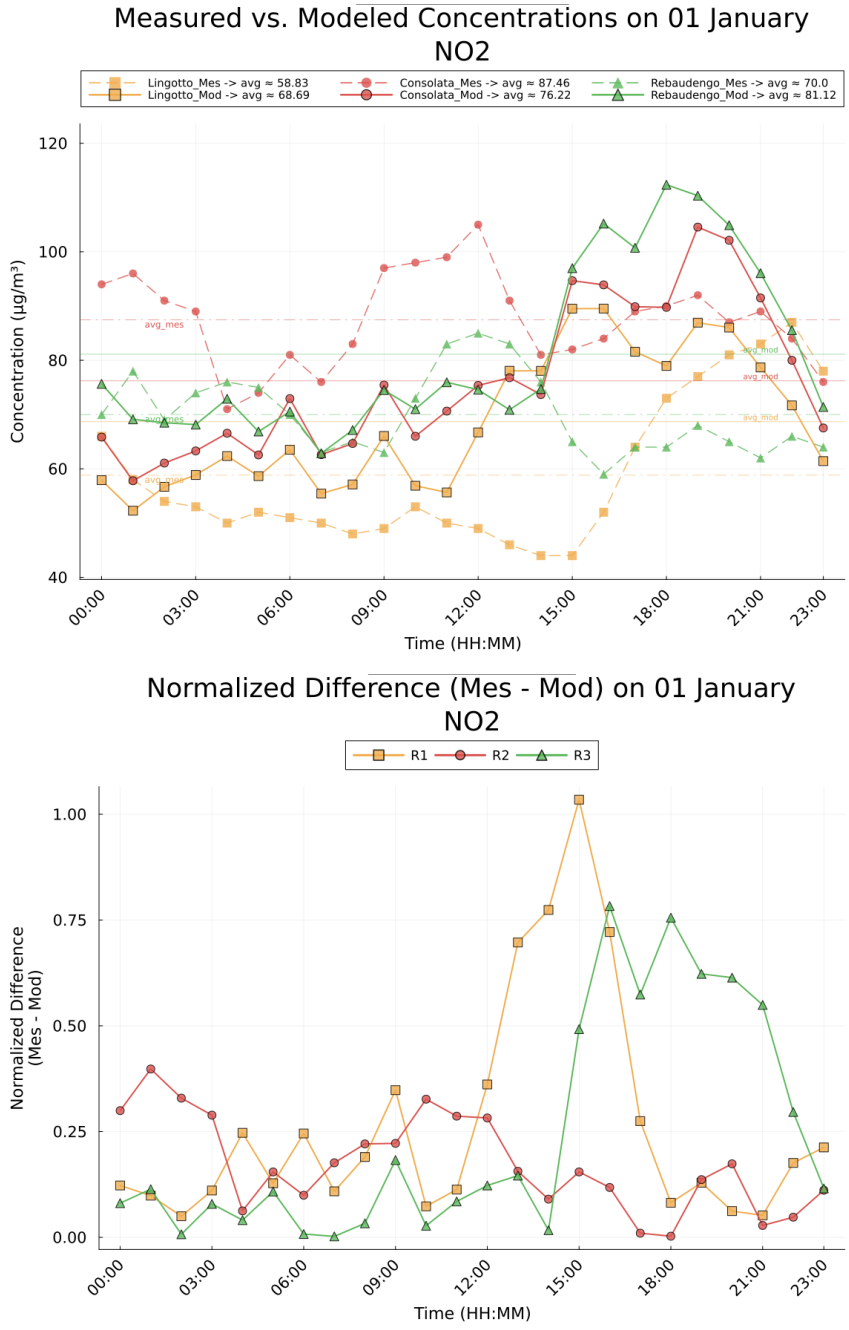

Figure 5.5: Measured and modelled NO_2 concentrations, with the normalized difference between them below.

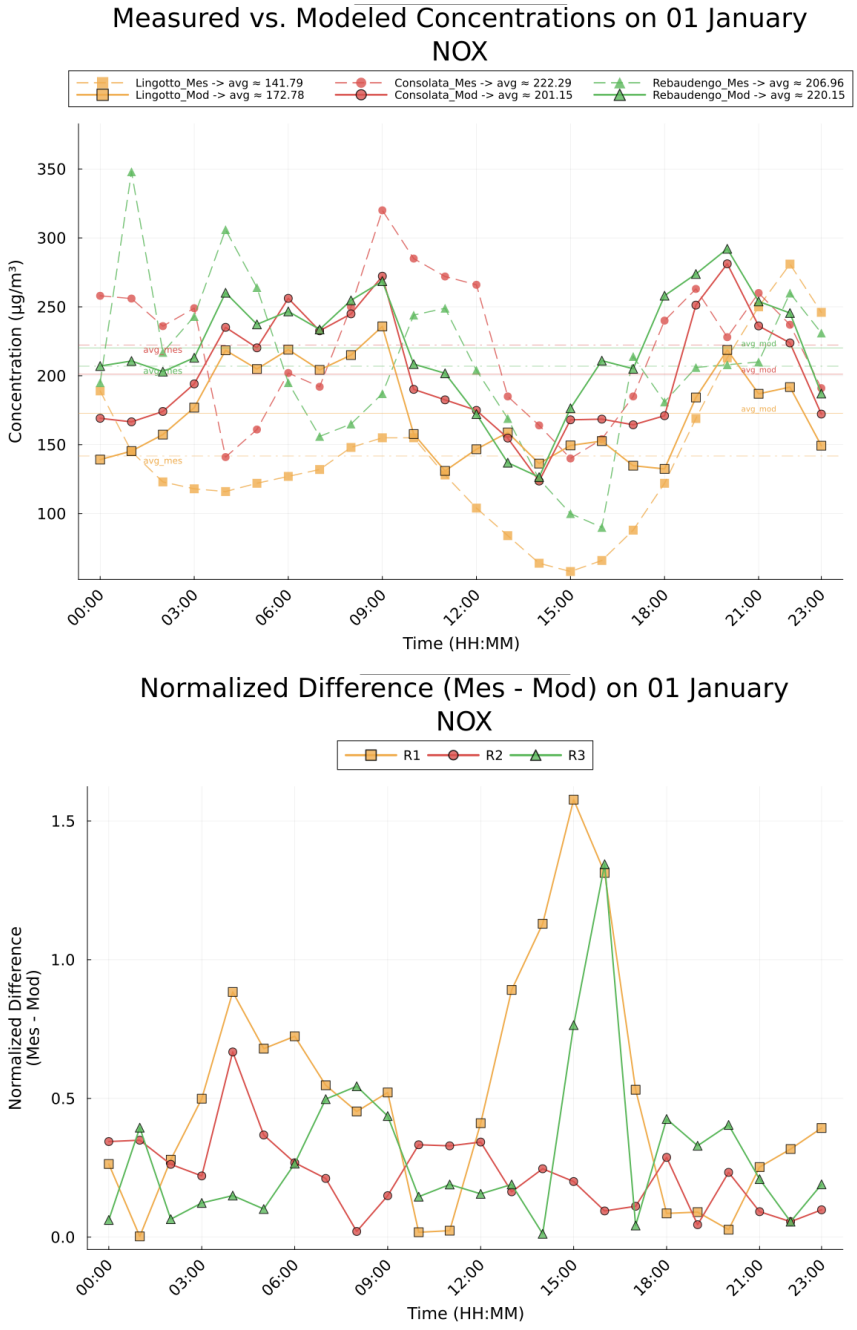

Figure 5.6: Measured and modelled NO_X concentrations, with the normalized difference between them below.

The key observations that emerge from the above graphical representation are as follows:

- A general overestimation of simulated concentrations of NO, NO_2, and NO_X by SIRANE, compared to measured values, is observed across all three receptors (R1, R2, and R3) when compared to measured values. This discrepancy may be due to several factors. As suggested by different studies on the model [13] the better tuning of input parameters, such as emission sources, traffic activity patterns and meteorological data, can significantly improve the model's accuracy.
  Despite the hourly differences, the daily average values do not present significant disagreement, indicating that while the model may struggle to represent hourly variations, it provides a closer approximation to the observed daily pollution trends when averaged over time.

- The normalized difference plots highlight that the largest discrepancies occur during specific periods, particularly in the afternoon (around 15:00 to 18:00). This may be addressed to specific behaviour of the model, which require a deeper investigation.

Chapter 6

Conclusions

In this thesis, various analyses are discussed, and the key findings are summarized below:

Structure of Air Quality Database The air quality database for Turin monitoring stations integrates data from five points in the urban area, each equipped with different instruments and methods for measuring pollutants. Particular attention should be paid to the sampling frequency of each measurement, that might be different. The dataset appears overall complete and able to provide a very reliable analysis for the nitrogen oxides, while for particulate matter additional research work may be done to fill some periods of data unavailability. A general aspect of the air quality database is the presence of missing data and potential errors in data recording, which is crucial to address, for conducting a subsequent reliable and accurate analysis.

Different Measurement Types Different instruments and methods for mainly particulate matter measurements, were discovered to be present in the data, across multiple stations. A correlation analysis between these measures was then conducted. The analysis firstly highlighted errors in labeling within the dataset, where differently named entries showed a perfect correlation, indicating they were likely the same measurement. In addition, highly similar measurements were identified, for which temporal integration could be used to improve the overall data coverage. Through the analysis it was shown that Gravimetric and Beta Ray detection methods presented a high good correlation, confirming the consistency of these two approaches. Finally, strong correlations were observed between specific atmospheric elements, such as BP_PM2.5 and BAAPM2.5, pointing out potential relations that could be worth investigating in future.

Trend Analysis The trend analysis revealed a consistent decrease in pollutant levels across all monitoring stations in Turin from the start of the 21st century until 2022. Both nitrogen oxides (NO, NO_2, NO_X) and particulate matter (PM_{10}, $PM_{2.5}$) exhibited a clear downward trend over the study period. Although the linear trend analysis faced challenges due to frequent data gaps, it still provided a useful overview of the overall decline in pollution levels through annual averages. The STL analysis, on the other hand, proved to be robust and reliable, handling even shorter time series effectively. Both analyses produced consistent results, reinforcing the observed long-term decreasing trend in pollution. The comparison between trends across different stations highlighted variations in pollution levels, mainly attributed to the environmental context of each monitoring point. Results reflecting the classification of each station's environment were obtained, for instance, stations located near traffic-heavy areas, like Torino-Rebaudengo, recorded higher pollutant concentrations, while stations in more residential or suburban areas, such as Torino-Rubino, showed lower levels. This consistency across stations further confirm the reliability of the analysis conducted and demonstrate how that the observed trends reflect the local environmental conditions.

Sirane Model From this study is provided a picture of the SIRANE model, and its potential for simulating pollutant dispersion in Turin. The simulations, carried out for only a restricted period offered a preliminary evaluation of the model's ability to capture general pollutant variation, especially during the day. For the case study basic settings were used, to the scope of evaluating how the model was suitable to provide a description of the dispersion of pollutant in the urban area. The reported analysis, albeit general and straightforward, still involved a complete and deep understanding of the model, such that it could be considered reliable for future, more specific analysis. Limitations were identified, particularly in the choice and definition of the model's input parameters, which could be improved with more recent air quality and weather data.

Overall conclusions This thesis focused specifically on the analysis of a large part of the database consisting of measurements of environmental pollutant concentrations on the municipality of Turin, first providing a clear view of its structure and consequently providing a methodology for analyzing these data. The research stresses the importance of careful processing techniques in order to obtain relevant results, such as providing a comprehensive view of how air pollution has changed from the beginning of the century to the present, and crucially, quantifying these changes with accurate numerical data. I personally believe this is particularly important since, according to the European Environment Agency, air pollution is the largest environmental health risk in Europe, causing a significant number of premature deaths and diseases each year [6].

Regarding the specific analysis done, on the urban area of Turin, although the concentration levels of pollutant are decreasing, and there is a consequent improvement in air quality, the values still remain very high. The analysis was able to reveal how the city of Turin, has been committed to fulfilling and complying with European policies over the past two decades, in order to reduce general pollutant emissions. Further analysis could focus particularly on the periods when pollutant levels significantly exceeded regulatory thresholds.

It is hoped that the results discussed in this work, that lays a foundational understanding of the database of pollution concentration in Turin during the first two decades of the 21st century, can serve as a reference for future studies and policy-making. By continuing to provide data analysis of this type, it is possible to better guides public health strategies and support adapted interventions to reduce the air pollution in urban areas. Further research is encouraged to refine the methods presented, improving data completeness, and exploring closely the relationships between pollutant and weather condition, recalling SIRANE as a valid tool for this purpose.

Appendix A

Nitrogen oxides

A.1 STL analysis

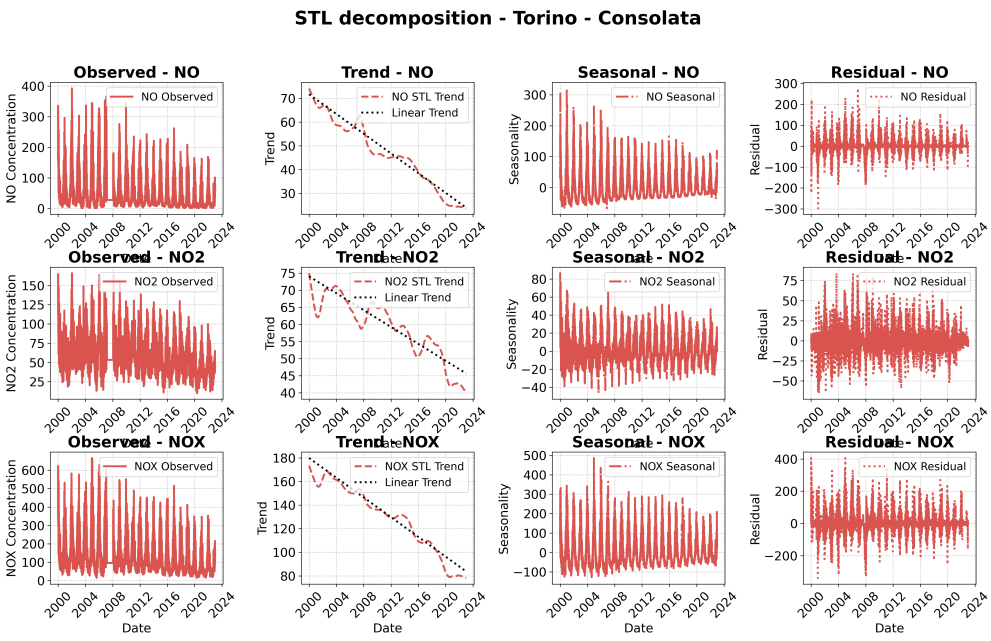

Figure A.1: STL decomposition for Torino-Consolata Station

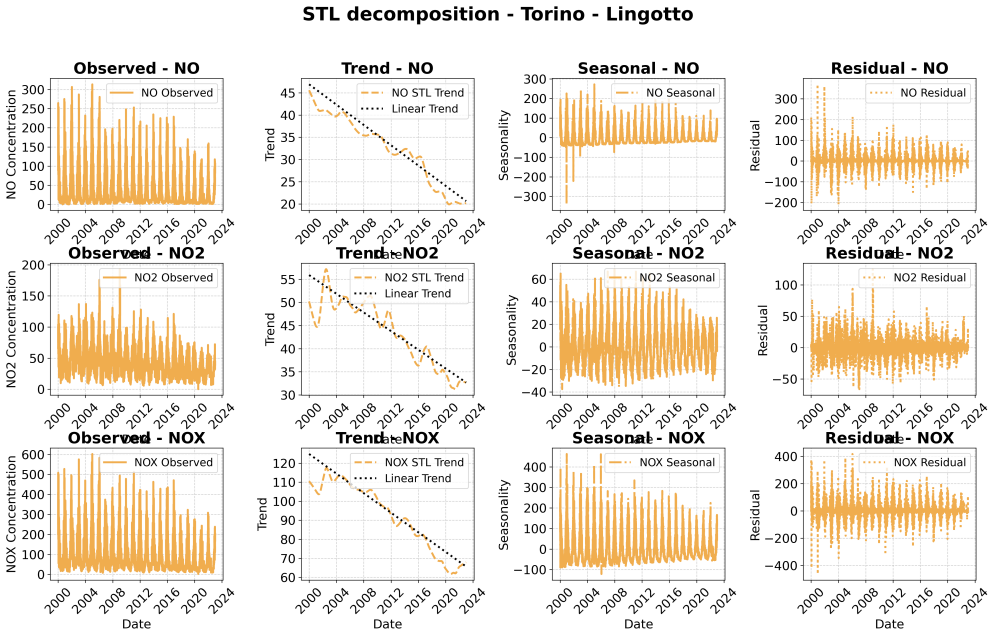

Figure A.2: STL decomposition for for Torino-Lingotto Station

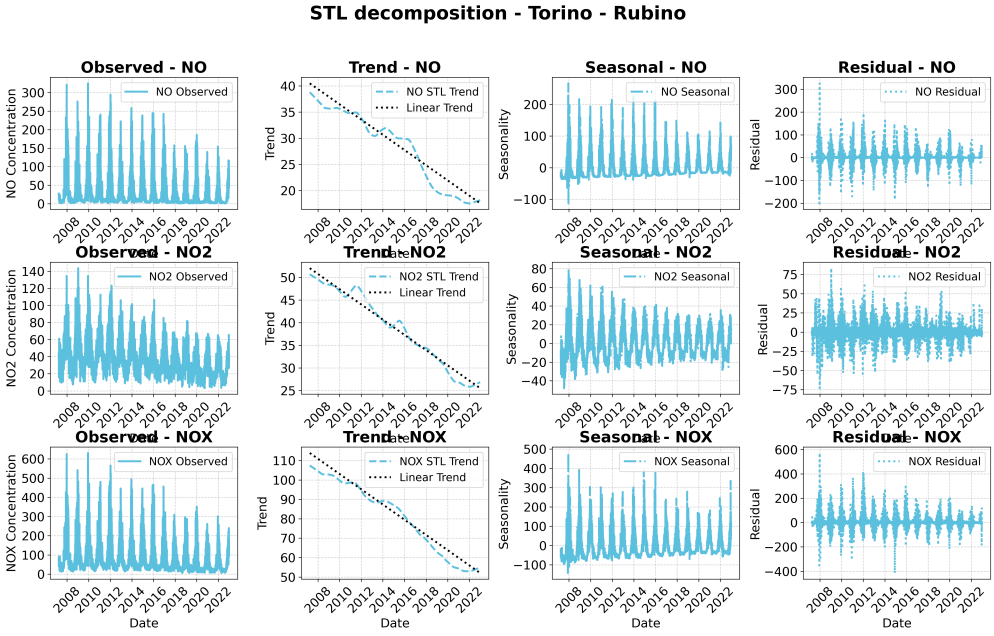

Figure A.3: STL decomposition for Torino-Rubino Station

Appendix B

Particulate matter

B.1 Scatter Plots of Highly Correlated Atmospheric Element Measurements

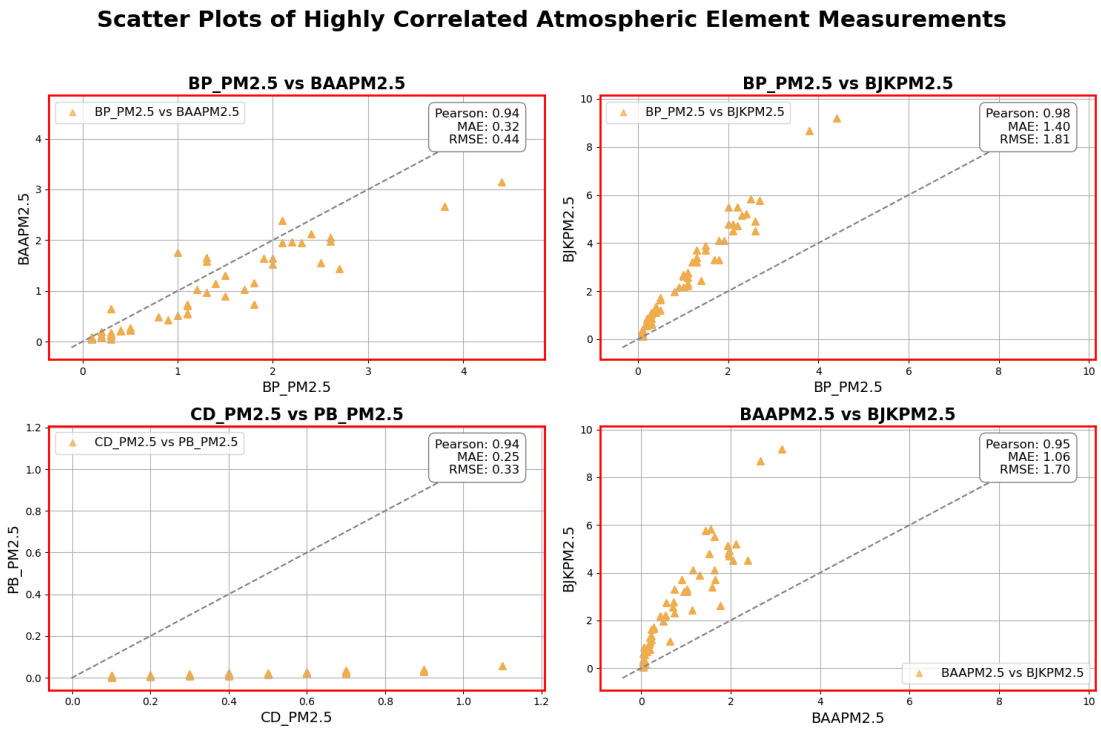

Figure B.1: Scatter Plots of Highly Correlated Atmospheric Element Measurements at Torino-Lingotto

B.2 Coverage Percentage of PM2.5 Elements

Coverage percentage of PM2.5 elements		
Measure	**On present days (%)**	**On total period (%)**
	Torino - Lingotto	
AS_PM2.5	52.90	40.26
BP_PM2.5	52.67	40.08
CD_PM2.5	52.90	40.26
NI_PM2.5	52.43	39.90
PB_PM2.5	52.90	40.26
BAAPM2.5	49.57	34.48
BJKPM2.5	51.51	35.83

Table B.1: Coverage percentage of PM2.5 elements dispersion measurements.

B.3 Analysis of PM_{10} hourly data sensitiveness

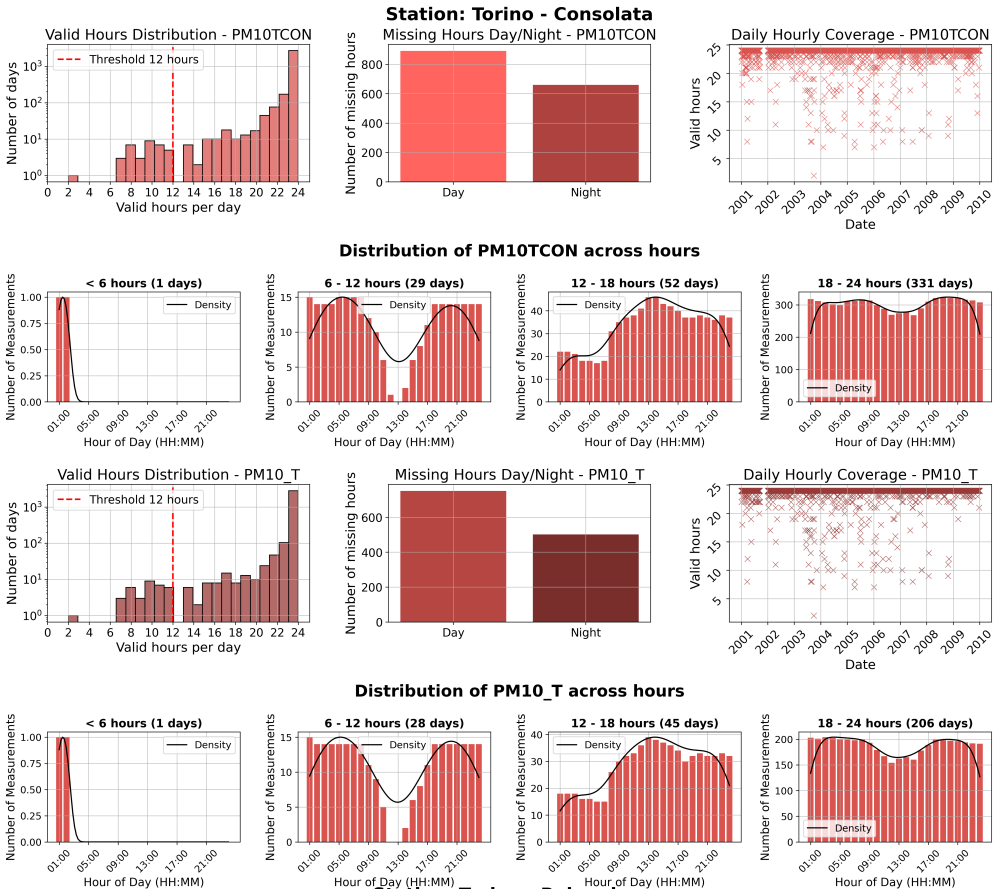

Figure B.2: PM_{10} hourly data summary

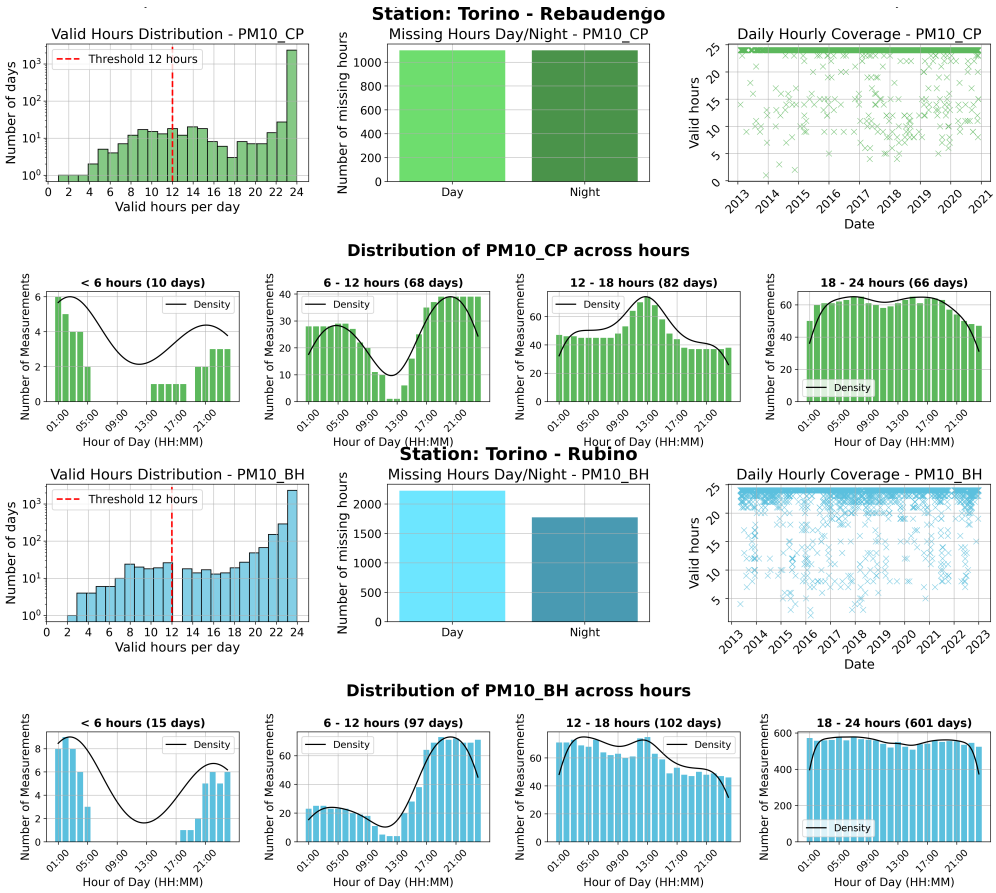

Figure B.3: PM_{10} hourly data summary

B.4 STL analysis

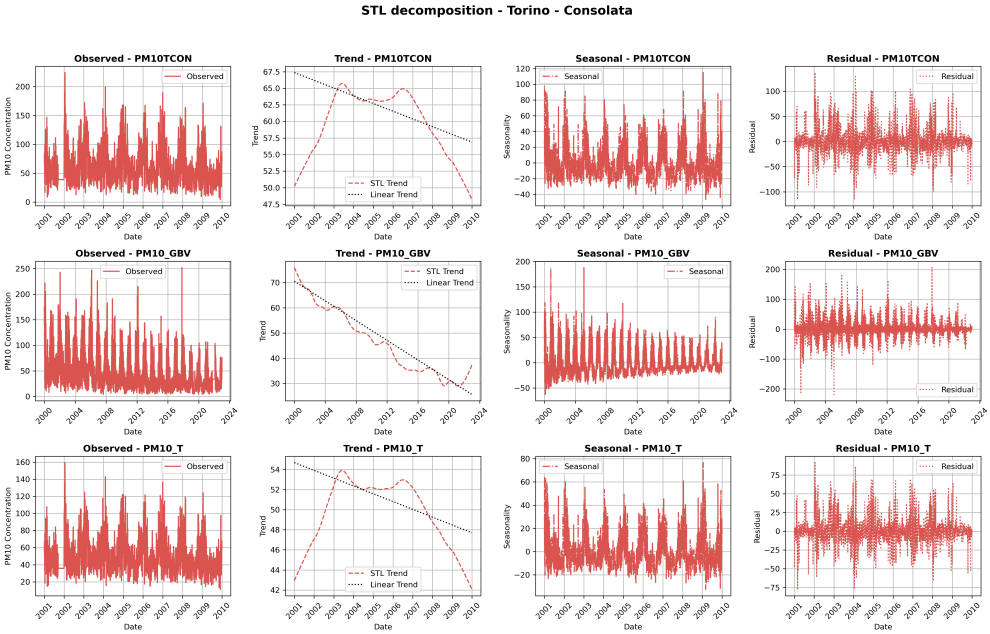

Figure B.4: STL decomposition for PM_{10} at Torino-Consolata Station

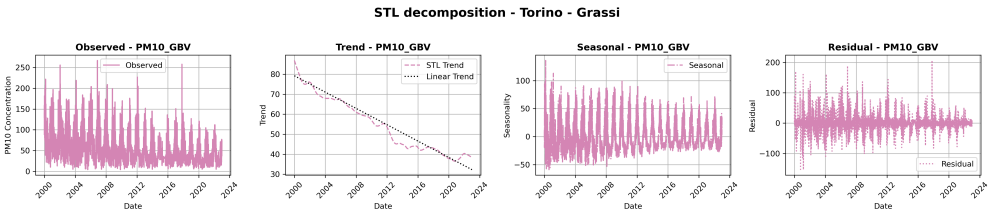

Figure B.5: STL decomposition for PM_{10} at Torino-Grassi Station

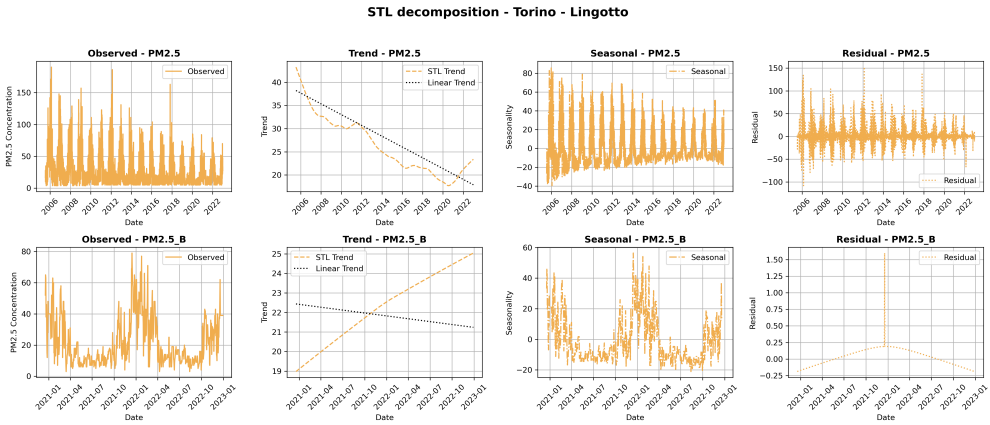

Figure B.6: STL decomposition for $PM_{2.5}$ at Torino-Lingotto Station

References

[1] Matteo Bo. «Study of aerosols air pollution assessments in indoor and outdoor environments based on measuring and modelling approaches». Doctoral Dissertation. PhD thesis. Politecnico di Torino, Ecole Centrale de Lyon, 2020. URL: http://hdl.handle.net/10589/176717 (cit. on p. 78).

[2] Aurelie Charron, Roy M Harrison, Steve Moorcroft, and Jeff Booker. «Quantitative interpretation of divergence between PM10 and PM2.5 mass measurement by TEOM and gravimetric (Partisol) instruments». In: *Atmospheric Environment* 38.3 (2004), pp. 415–423. ISSN: 1352-2310. DOI: https://doi.org/10.1016/j.atmosenv.2003.09.072. URL: https://www.sciencedirect.com/science/article/pii/S1352231003008616 (cit. on p. 43).

[3] Robert B Cleveland, William S Cleveland, Jean E McRae, and Irma Terpenning. «STL: A Seasonal-Trend Decomposition Procedure Based on LOESS». In: *Journal of Official Statistics* 6.1 (1990), pp. 3–73 (cit. on p. 68).

[4] Comune di Torino. *Air Quality Monitoring Stations*. Accessed: 2024-10-02. 2024. URL: http://www.comune.torino.it/ambiente/aria/aria_torino/stazioni_aria (cit. on p. 34).

[5] Statsmodels developers. *STL decomposition using LOESS in Python*. 2022. URL: https://www.statsmodels.org/stable/examples/notebooks/generated/stl_decomposition.html (cit. on p. 68).

[6] European Environment Agency. *Health Impacts of Air Pollution in Europe*. Accessed: 2024-09-14. European Environment Agency, 2022. URL: https://www.eea.europa.eu/publications/health-impacts-of-air-pollution (cit. on p. 94).

[7] European Parliament and Council of the European Union. *Directive 2008/50/EC of the European Parliament and of the Council of 21 May 2008 on ambient air quality and cleaner air for Europe*. Official Journal of the European Union. 2008. URL: https://eur-lex.europa.eu (cit. on p. 59).

[8] ARPA Piemonte. *L'aria in Piemonte nel 2017*. 2017. URL: https://www.snpambiente.it/snpa/arpa-piemonte/laria-piemonte-nel-2017/ (visited on 10/15/2024) (cit. on p. 63).

[9] ARPA Piemonte. *Sistema Regionale di Rilevamento della Qualità dell'Aria*. Accessed: 2024-10-14. URL: https://aria.ambiente.piemonte.it/%5C#/qualita-aria/dati (cit. on p. 4).

[10] Nabil Ben Salem, Lionel Soulhac, Pietro Salizzoni, and Massimo Marro. «Pollutant source identification in a city district by means of a street network inverse model». In: *International Journal of Environment and Pollution* 55.1-2-3-4 (2014), pp. 50–57. DOI: 10.1504/IJEP.2014.065904. URL: https://hal.science/hal-01296878 (cit. on p. 77).

[11] Skipper Seabold and Josef Perktold. «statsmodels: Econometric and statistical modeling with python». In: *9th Python in Science Conference*. 2010. URL: https://conference.scipy.org/proceedings/scipy2010/pdfs/seabold.pdf (cit. on pp. 51, 68).

[12] Kritika Shukla and Shankar G. Aggarwal. «Performance check of beta gauge method under high PM2.5 mass loading and varying meteorological conditions in an urban atmosphere». In: *Atmospheric Pollution Research* 12.11 (2021), p. 101215. ISSN: 1309-1042. DOI: https://doi.org/10.1016/j.apr.2021.101215. URL: https://www.sciencedirect.com/science/article/pii/S1309104221002798 (cit. on p. 46).

[13] Lionel Soulhac, Pietro Salizzoni, F.-X. Cierco, and Richard Perkins. «The model SIRANE for atmospheric urban pollutant dispersion; part I, presentation of the model». English. In: *Atmospheric Environment* 45.39 (2011), pp. 7379–7395. ISSN: 1352-2310. DOI: https://doi.org/10.1016/j.atmosenv.2011.07.008. URL: https://www.sciencedirect.com/science/article/pii/S1352231011007096 (cit. on pp. 77, 78, 92).

[14] Jessica Tryner, Nicholas Good, Ander Wilson, Maggie L Clark, Jennifer L Peel, and John Volckens. «Variation in gravimetric correction factors for nephelometer-derived estimates of personal exposure to PM2. 5». In: *Environmental Pollution* 250 (2019), pp. 251–261 (cit. on p. 46).

[15] Michael Waskom. *seaborn: statistical data visualization*. Accessed: October 15, 2024. 2021. URL: https://seaborn.pydata.org/ (cit. on p. 50).