Assessment of PM Exposure at intra-urban scale in preparation of MAIA mission (APEMAIA) Project

WP1102 – User Requirement Analysis

DEL004 – User Requirement Document - USD (preliminary version)

Contract n. 2023-39-HH.0

CUP n. F83D23000210005

ASI CONTRACT N°	2023-39-HH.0		
PROJECT NAME	Assessment of PM Exposure at intra-urban scale in preparation of MAIA mission (APEMAIA)	**CONTRACTOR**	National Research Council – Institute of Atmospheric Pollution Research
PROJECT MANAGER	Maria Adamo (CNR-IIA)	**DATE SUBMITTED**	N/A
AUTHOR	M. Adamo, M. Aquilino, S. Maggi, S. Fuina, S. Cinnirella, M. De Lucia (CNR-IIA) A. Morabito (ARPA-Puglia)	**CONTRACTOR'S REFERENCE**	Issue:01 Rev.:01 Date:25/11/23
PROJECT BEGIN DATE	26/09/2023	**PROJECT END DATE**	25/09/2025

Institute of Atmospheric Pollution Research National Research Council of Italy	*Project: APEMAIA* Contract n. 2023-39-HH.0	Doc. N.: DEL_004_USD_CNR-IIA_RA1_EN-001 Issue/Rev.:01/01 Date: 25/11/2023 (T0+2) Pages: 64

EXECUTIVE SUMMARY			
This document is the first deliverable of the **A**ssessment of **P**M **E**xposure at an intra-urban scale in preparation for the **MAIA** mission (APEMAIA) project. It is released at the conclusion of activities scheduled for Work Package WP1102 – User Requirement Analysis. The release adheres to the planned schedule: (T0+2M). The document represents a preliminary version of the **Us**er requirement **D**ocument (USD) submitted as First Release of Advancements (**RA1**). It is worthy to note that USD may be subject to revision based on any additional user requirements that may emerge throughout the project.			
NAME OF ASI PROJECT MANAGER	Name Surname (PM) Name Surname (RUP) Name Surname (Project Scientist)		
STAR CODE		No OF VOLUMES:	

TABLE OF CONTENTS

1 AIM OF THE DOCUMENT 6
1.1 Applicable and reference documents 6
1.2 Applicable documents 7
1.3 Reference documents 7
1.4 Acronyms and abbreviations 8
2 IDENTIFICATION OF USER NEEDS 10
2.1 User communities 10
2.2 List of users' needs 11
3 COLLECTION AND ANALYSIS OF PRE-EXISTING STUDIES 16
4 ANALYSIS OF STUDY AREAS 21
4.1 Apulian test sites: Bari and Taranto 21
4.1.1 Air quality monitoring network 23
4.1.2 Meteorological and climatic conditions 29
4.2 Primary Target Area: Roma 37
1.1.1 Air Quality Monitoring Network 37
1.1.2 Meteorological and climatic conditions 41
5 TECHNICAL REQUIREMENT SPECIFICATIONS 49
5.1 Requirements Classification 49
5.2 General functional requirements 51
5.3 Data management requirements 53
5.4 Performance requirements 55
5.4.1 Spatial resolution 55
5.4.2 Temporal references 55
5.4.3 Frequency 55
5.5 Storage requirements 56
5.6 Dissemination requirements 56

5.7 Verification requirements 57
6 TRACEABILITY MATRICES USER NEEDS VS USER REQUIREMENTS 58
6.1 User Needs vs. User Requirements 58
6.2 User requirements vs. User Needs 58
7 REFERENCES 60

1. AIM OF THE DOCUMENT

This document represents the preliminary version of the User Requirement Document (USD) submitted as First Release of Advancements (RA1) for the **A**ssessment of **P**M **E**xposure at an intra-urban scale in preparation for the **MAIA** mission (APEMAIA) Project.

The APEMAIA project's activities are carried out by a consortium led by the Institute of Atmospheric Pollution Research of the National Research Council of Italy (CNR-IIA), acting as the prime contractor, and supported by the Apulian Regional Agency for Environmental Protection (ARPA Puglia) and the Physics Department of the University of Bari (DIF-UNIBA).

The document serves as a formal deliverable released at the conclusion of activities scheduled for Work Package WP1102 – User Requirement Analysis.

The WP1102 includes the list of tasks identifies as follows:

- Identification of user needs (in collaboration with ARPA Puglia).
- Analysis of study areas.
- Collection and analysis of pre-existing studies.
- Definition of technical specifications for the monitoring application to be implemented.
- Definition of spatial and temporal criteria for the selection of satellite data.

The end of these activities has been planned for the period T0+2M.

This report is related to WP1102 tasks and describes an overview of user communities and their specific needs/requirements in **Chapter 2**. **Chapter 3** gives a comprehensive review of recent bibliography in the project's domain; while **Chapter 4** conducts an analysis of the study areas to accurately introduce the pre-existing air quality and climatic conditions in the designated locations**. Chapter 5** translates the identified needs into detailed scientific and technical requirement specifications for the project, which are designed to guide the development of APEMAIA algorithms, products, and applications, as well as to facilitate the optimal selection of satellite data for the project's objectives. The latter aspect has been just started in the WP1102 activities and will be further developed during the 'WP2101 – Collection of EO/ancillary data' phase, with a detailed reporting in the following progress report releases. Finally, **Chapter 6** reports traceability matrices of user needs versus user requirements and vice versa.

1.1 Applicable and reference documents

This document incorporates provisions from other dated or undated documents. These references are cited at appropriate places in the text and are listed hereafter. For undated references, the latest edition of the publication referred applies.

In case of conflict between two or more applicable documents, the higher document will prevail.

1.2 Applicable documents

[AD 1] ASI Contract No. 2023-39-HH.0

1.3 Reference documents

[RD 1] ECSS Configuration and information management, Doc Id. ECSS-M-ST-40C Rev. 1

[RD 2] ECSS Technical requirements specification, Doc Id. IECSS-E-ST-10-06C Rev.2 (6 March 2009)

[RD 3] NASA System Engineering Handbook, NASA SP-2016-6105 Rev2

[RD 4] WHO global air quality guidelines: particulate matter (PM2.5 and PM10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide

1.4 Acronyms and abbreviations

Table 1-1 provides the acronym list.

ha eliminato: Table 1-1

Table 1-1: Acronyms and abbreviations

Acronym	Description
AOD	Aerosol Optical Depth
AI	Artificial Intelligence
APEMAIA	Assessment of PM Exposure at intra-urban scale in preparation of MAIA mission
ARPA Puglia	Apulian Regional Agency for Environmental Protection
ASDM	Artificial Neural Network Sequential Downscaling Method
ASI	Italian Space Agency
BTX	Benzene, Toluene, and Xylene
BME	Bayesian Maximum Entropy
CM	Chemical Transport Models
CNR-IIA	Institute of Atmospheric Pollution Research of the National Research Council
CO	Carbon Monoxide
CTMs	Chemical Transport Models
DIF-UNIBA	Physics Department of the University of Bari
DSM/DTM	Digital Surface Model/Digital Terrain Model
EEA	European Environment Agency
EO	Earth Observation
ESA	European Space Agency
FARM1	Flexible Air quality Regional Model
FAIR	Findability, Accessibility, Interoperability, and Reusability
FAIRMODE	Framework for Air Quality Models and Their Interactions in Europe
GAP	Generic Attribution Preprocessor
HR	High-Resolution
INEMAR	Istituto Nazionale di Ecologia Marina
ISPRA	Istituto Superiore per la Protezione e la Ricerca Ambientale
ISO	International Standards Organization
JPL	Jet Propulsion Laboratory
JRC	Joint Research Centre
LR	Linear Regressions
LCLU	Land Cover Land Use
LUR	Land Use Regression
MAIA	Multi-Angle Imager for Aerosols
ML	Machine Learning
MLR	Multiple Linear Regression
NASA	National Aeronautics and Space Administration
NDVI	Normalized Difference Vegetation Index

NO2	Nitrogen Dioxide
O3	Ozone
OGC	Open Geospatial Consortium
OI	Optimal Interpolation
PCA	Principal Component Analysis
PM	Particulate Matter
PWELs	Population-weighted PM exposure levels
RF	Random Forest
RLAs	Regional and Local Authorities
RRQA	Regional Air Quality Monitoring Network
SO2	Sulfur Dioxide
SVMR	Support Vector Machine Regression
SURFPRO	Surface Processor
WebGIS	Web Geographic Information System
WFS	Web Features Services
WMS	Web Mapping Services
WP	Work Package
XGBoost	eXtreme Gradient Boosting

2. IDENTIFICATION OF USER NEEDS

This section outlines the first task reporting the user requirement identification workflow and the list of all the identified needs.

2.1 User communities

The APEMAIA project aims to play a pivotal role in shaping future evidence-based policies, interventions, and mitigation strategies, thereby safeguarding public health in densely populated urban environments.The APEMAIA project has the potential to impact various stakeholder communities, including scientists and researchers from both governmental and industrial/commercial sectors, technical professionals from local and regional authorities (RLAs), such as epidemiologists, air quality specialists, urban planners, and, last but not least, the broader civil society as it addresses health-related concerns affecting everyone (Figure 2-1).

ha eliminato: Figure 2-1

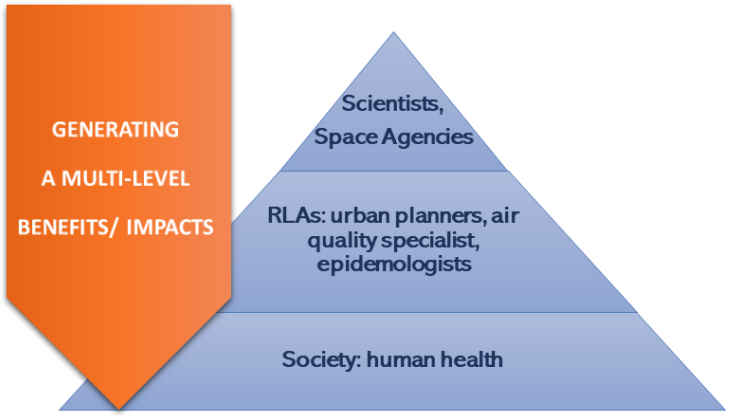

Figure 2-1 – Stakeholders of APEMAIA project.

The scientific community and the industrial research sector in the field of space are highly interested in technological advancements related to the generation of particulate matter (PM) pollution maps from Earth Observation (EO) data. This interest arises from the fact that APEMAIA represents a pilot project to leverage future products from the NASA JPL's Multi-Angle Imager for Aerosols (MAIA) mission, in collaboration with ASI, which is currently in pre-launch studies phase.

The primary stakeholder and, simultaneously, partner of the APEMAIA project is ARPA Puglia, which already provides a 1 km resolution air quality modelling service as part of its institutional and legal obligations. Therefore, there is significant interest in enhancing the service by providing maps that

inform citizens at an intra-urban scale. This effort not only aligns with their statutory responsibilities but also addresses the growing demand for air quality information, particularly in urban areas.

2.2 List of users' needs

The user requirements phase is based on the legacy of projects in, which CNR-IIA was actively involved before to the APEMAIA proposal. EC funded projects such as SMURBS ("SMURBS project," 2017) and e-SHAPE ("e-SHAPE project," 2019) have already addressed the problem of air quality in urban areas. These initiatives have revealed a significant lack of local-scale data to inform citizens about their long-term exposure to PM pollution.

Additional requirements stem from direct engagement with stakeholders, particularly the air quality officers from ARPA Puglia, who are providing valuable input as project collaborators.

The APEMAIA project aims to play a pivotal role in shaping future evidence-based policies, interventions, and mitigation strategies, thereby safeguarding public health in densely populated urban environments.

The table below lists all the Users Problems/Needs accordingly with the type of user.

The 3 key stakeholders (User Segments) are identified with the code:

US1. Space agencies, scientists.

US2. Local and regional authorities/technicians.

Society.

Each need is progressively numbered as follows (Table 2-1).

Table 2-1: User needs

Type of User	User Need ID-code	Rationale Statement	Problems/Needs	Baseline Expectations	
US1	**US1_US1N0100**	NASA's JPL intend to demonstrate the capabilities of MAIA's new technology for future missions. MAIA is a partnership between NASA and ASI.	NASA needs to extend its team in order to conduct scientific investigations that complement those performed by the NASA team.	APEMAIA project, funded by ASI, will support innovative research on PM Exposure estimation at an intra-urban scale, contributing to strengthen MAIA ASI Science Team partnership	

Commentato [PM1]: Sarebbe utile inserire una tabella con la lista degli user identificati e per i quali si fa riferimento nella tabella 2-1.

Commentato [MA2R1]: Inserita la tabella 2-1. La precedente tabella 2-1 è ora la tabella 2-2.

Commentato [PM3]: Sarebbe utile identificare una strategia da seguire per verificare la lista degli user needs ed, eventualmente, aggiornarla.

Commentato [MA4R3]: La lista attuale è stata improntata sulla base di spunti raccolti da progetti precedenti e validata da ARPA Puglia (vedi introduzione al paragrafo). Le agenzia ARPA, in Italia, sono gli enti istituzionalmente incaricati di monitorare la qualità dell'aria e tutelare i cittadini da eventuali rischi della salute legati all'inquinamento atmosferico. Pertanto abbiamo valutato di affidarci alle competenze di ARPA, per la validazione della lista corrente e delle eventuali modifiche/integrazioni sugli user needs. Questo concetto è stato aggiunto all'inizio del paragrafo. La strategia per incrementare la numerazione è indicata immediatamente dopo la Table 2-2.

Commentato [PM5]: Sarebbe utile fornire anche una indicazione più specifica relativamente allo user a cui si riferisce lo user need (in relazione alla tabella richiesta dal commento precedente).

Commentato [MA6R5]: Inserita la colonna (Users reference number)

ha eliminato: Table 2-1

Commentato [PM7]: E' prevista una integrazione per eventuali ulteriori user needs legati ad aspetti epidemiologici?

Commentato [MA8R7]: Non è stata esplicitamente prevista dalle attività del progetto, ciò non toglie che nuovi requisiti possano emergere dall'interlocuzione crescente con gli attori della piattaforma EIRENE-RI.

	US1_N0200	To support ASI research goals, three MAIA target areas have been established in Italy. These cover a number of cities such as Rome, Bologna, Milan, Turin, Bari, and Taranto.	More local data need to validate, establish, and maintain a version of the MAIA data archive in Italy.	The APEMAIA project, through partnership with ARPA Puglia, aims to collaborate to provide ground validation data for some cities included in the target areas.	
	US1_N0300	MAIA has established an Early Adopters Program aimed at helping those interested in MAIA data products understand how to use them.	Before launch, the MAIA team need to collect feedback from the Early Adopters around the accessibility and usability of the MAIA data in their work.	MAIA ASI Science Team will provide to the APEMAIA's team an early access to the MAIA simulated products. By using such data APEMAIA expect to early develop algorithms to facilitate the future usage of MAIA products from various categories of the stakeholders.	
	US1_N0400	The estimation of predictors that may influence the spatial variability at the intra-urban scale of PM concentrations through the use of satellite data/products (e.g., Copernicus) is an open issue of recent research on urban air pollution.	The current literature mainly tests the LUR model, based on multiple linear regressions or other statistical approaches, to identify spatial patterns of PM in cities.	The APEMAIA project intends to investigate a wide variety of ML techniques, including Neural Networks and Random Forest, for better improving the understanding of spatial variability of PM patterns in cities.	
	US1_N0500	AOD has been considered, in recent literature, as an important predictor for PM estimation in cities with a low number of air quality monitoring stations.	Scientists have increasing availability of various remote sensing AOD data (i.e., Sentinel3 Synergy Product and, in future, MAIA's AOD products) but such data are not adequately used in recent research studies.	The APEMAIA project intends to develop procedures for the extraction of time series of AOD with high spatial resolution, through the use of multi-scale AOD data (measured via MISR, MODIS, Sentinel2/3 and simulated MAIA's AOD products).	
	US1_N0600	The Italian scientific community is actively contributing to the construction of the European Research Infrastructure for Environmental Exposure Assessment in Europe (EIRENE-RI). CNR-IIA is leading the Italian node to which ASI	Further studies and comprehensive investigations are needed to deepen our understanding of the complex relationship between high resolution of urban air quality and its potential health impacts.	The APEMAIA project has the potential to catalyze future endeavors specifically focused on health risk assessment utilizing epidemiological data. Some datasets, focusing on the Primary Target Area (Rome) and the two sites in Bari and	

		and the Italian National Institute of Health (ISS) are formally participating. The preparatory phase of this infrastructure was started on October 1, 2022, with the EIRENE PPP project and is expected to span three years.		Taranto, are already available.	
US2	**US12_N0700**	ARPA Puglia provides PM measurements from stations located across the territory of Bari and Taranto.	These measurements are mostly insufficient for monitoring large areas.	By using some predictors derived from satellite data, which has greater spatial capillarity, the APEMAIA project would contribute to improve the spatial spatial/temporal variability estimation of PM concentrations in cities.	
	US2_N0800	ARPA Puglia provides PM (Particulate Matter) maps using Eulerian and Lagrangian modeling models.	The PM maps generated by ARPA Puglia have, currently, a spatial resolution ranging from 1 to 4 km. It's important to note that this resolution is influenced by the utilization of low-resolution meteorological model data as input for the models.	The APEMAIA project, aligned with recent ARPA Puglia initiatives, aims to improve the spatial resolution in PM maps by incorporating artificial intelligence approaches alongside existing models	
	US2_N0900	In alignment with the international framework outlined in the Sustainable Development Goals of the United Nations 2030 Agenda (specifically Target 11.6) and the directives of the World Health Organization [RD 4], it is mandatory that PM concentrations in urban areas be evaluated in correlation with population data.	National and Regional Environmental Protection Agencies (as ARPA Puglia) will soon be required to revise their methodology for estimating critical PM exposure levels, which have direct implications for human health. For instance, as of October 26, 2023, the European Commission has introduced a proposal for a new air quality directive intended to replace and consolidate the existing directives, namely,	One of the primary objectives of the APEMAIA project is to extract Population-Weighted PM Exposure Levels (PWELs). This aligns with the evolving international and regional directives that emphasize the significance of considering population density in assessing air quality, ensuring a more comprehensive understanding of the real impact on human health.	

ha eliminato: US

			2008/50/EC and 2004/107/EC.		
	US2_N1000	The European Environment Agency (EEA) and the Joint Research Centre of the European Commission (JRC) established the Forum for Air Quality Modelling (FAIRMODE) in 2007. This network brings together modelers and air quality experts to promote and support the harmonized use of models by EU Member States, with a particular focus on applying models in accordance with European air quality directives.	There is a need to subject the PM maps derived from modelling approaches (any model which produces Chemical Transport Model comparable output) to the assessment procedures outlined by the Air Quality Directives.	The APEMAIA project strives to subject the generated PM maps to the quality control mechanism recommended by FAIRMODE, incorporating the use of standardized quality objectives. The project will use the freely available DELTATOOL software, developed by JRC researchers, for this purpose.	
	US2_N1100	The modelling system for the prediction and assessment of air quality by ARPA Puglia recognizes the significance of the marine aerosol phenomena and the intrusion of Saharan dust into the Mediterranean basin. The latter can result in an abnormal increase in PM10 concentration values and, in some cases, contribute to exceeding the limits set by regulations.	ARPA Puglia acknowledges the need to employ different models for PM estimation based on the occurrences of these two phenomena, which significantly impact PM measurements.	The APEMAIA project aims to deploy various models and algorithms designed for two distinct operational scenarios: 1) marine aerosol and 2) Saharan dust.	
	US2_N1200	The PM mapping model developed by ARPA Puglia is continually enhancing our understanding of the spatio-temporal variability of PM patterns in urban areas. Currently, certain areas in the Apulian region are covered by a 1km mapping service, and there is an ongoing ambition to extend mapping capabilities to lower resolutions in the near future.	There is a recognized necessity to enhance the representation of spatial variability in PM concentrations within the specified area of interest, by effectively capturing the concentration variations found in urban, industrial, suburban, and port areas. Additionally, the models need to accurately	The APEMAIA project, through the integration of meteorological, land cover, urban morphology, socio-economic, and other ancillary predictors into its models, aims to identify the factors influencing spatiotemporal variability in PM concentration patterns, addressing the needs expressed by users.	

			reproduce temporal variability influenced by changes in emission sources and meteorological conditions, which can notably impact particulate concentrations, including phenomena such as marine aerosol and Saharan intrusions.		
Q_	Q_N1300	Particulate air pollution poses significant health risks (e.g., cardiovascular and respiratory diseases).	The relationship between long-term exposure to air pollution and health in urban areas is currently undergoing thorough examination by the broader society.	The APEMAIA project is aimed at identifying "warning areas" within cities with increasing trends or abrupt changes in PM concentration that may have a medium- to long-term impact on health.	
	Q_N1400	There is a widespread lack of adequate awareness among the public regarding the potential risks associated with long-term exposure to PM in the neighborhoods where individuals commonly reside, work, or study.	This information gap accentuates the necessity to establish an information system on open air quality data that is easily accessible to a broad society.	In order to ensure widespread access and usability of the generated information, the APEMAIA project intends to share all scientific outcomes, including maps and geo-stories, through a prototype of a territorial information system accessible via the web (WebGIS).	

ha eliminato: US3

ha eliminato: US3

ha eliminato: US3

3. COLLECTION AND ANALYSIS OF PRE-EXISTING STUDIES

Commentato [PM9]: Trovo questa review molto interessante ma suggerirei di aggiungere una sezione che sia collegata a degli user requiremnet, derivati dai problemi aperti dal punto di vista scientifico.

Commentato [MA10R9]: Molti degli user needs trovano riscontro nei problemi aperti sul piano scientifico che emergono dalla letteratura. Pertanto la sezione è stata rivista inserendo gli opportuni riferimenti ai requisiti utente ed alcune frasi che meglio chiariscono la logica esplorativa della bibliografia consultata.

The most common approach for the measurement of fine (PM2.5) and coarse (PM10) particulate matter concentrations in ambient air is through scattered ground monitoring stations. Although station-based air quality monitors can provide PM concentration (hereafter we refer also to its concentration if the term is not specified) with high accuracy and temporal resolution, these in situ measurements can only represent the concentration of relatively small areas, which cannot reflect large-scale population exposure. Satellite remote sensing can overcome these drawbacks and offers the possibility to continuously monitor PM with high spatial coverage using Aerosol Optical Depth (AOD). The fact that ground-level PM contributes to satellite-derived AOD provides an important premise and theoretical basis for estimating PM concentrations using satellite-derived AOD. Nevertheless, there are many factors that affect the relationship between AOD and PM, such as humidity, planetary boundary layer height, and size fraction. Specifically, surface PM refers to the mass concentration near the ground, while AOD measures the extinction in the total column of the atmosphere; PM refers to the concentration of both PM2.5 and PM10 particles while in the AOD measures the aerosol extinction is also influenced by water vapour. Aerosol size distribution and chemical composition may also affect the AOD-PM relationship in different spatial and temporal domains. With growing attention to air quality and sufficient ground-level observations, various models for PM estimation have been developed and applied during the last two decades, using different datasets and considering various study regions and time periods. A review can be found in (Bai et al., 2023; Holloway et al., 2021; Rahman, 2023; Ranjan et al., 2021; Shin et al., 2020; Xu et al., 2021; Zhang et al., 2021).

Considering the analysis of these pre-existing studies we can argue that the methods and approaches applied for satellite-based PM concentration mapping can be grouped into the following categories:

Chemical Transport Models (CTMs) based approaches. These approaches capitalize on cutting-edge CTM models to replicate the spatial-temporal variations in scaling factors between Aerosol Optical Depth (AOD) and PM mass concentration across different locations. Subsequently, these factors are applied to interpret the observed AOD from satellites, facilitating the estimation of near-surface PM mass concentration. In contrast to statistical models, simulation models possess robust geophysical mechanisms, providing an alternative for predicting PM without relying on ground-based but involving satellite AOD observations as a constrain for AOD. While several studies highlight the benefits of employing CTMs to reduce errors in satellite-retrieved PM, these techniques demand substantial resources in terms of data, computational capacity, expertise, and time (Liu, Ming, 2020). Furthermore, the performance of simulation results is vulnerable to uncertainties arising from simulated chemical and transport mechanisms, emission inventories,

atmospheric conditions (including meteorology and chemical composition), and other initial parameters.

Statistical approaches. Quantitative associations between satellite-observed Aerosol Optical Depth (AOD) and ground-level PM measurements form the foundation for statistical models. These models encompass various statistical approaches, including but not limited to linear regressions (LR) and Multiple Linear Regression (MLR) (Bakaeva and Le, 2022), Land Use Regressions (LUR) (Alvarez-Mendoza et al., 2019; Bakaeva and Le, 2022; Shi et al., 2021), Geographically and Temporally Weighted Regression (GTWR) models (He and Huang, 2018; Ma et al., 2014; Xue et al., 2020), Mixed Effects Models (MEM) (Pedde et al., 2022), Generalized Additive Models (GAM) (Chen et al., 2021) and Bayesian hierarchical model (BHM) (Murray et al., 2019; Yu et al., 2017).

Machine/Deep Learning approaches. Machine learning stands out as a robust modelling tool, leveraging its potent nonlinear processing and parallel computing capabilities. It has demonstrated considerable success across various domains, including signal processing, predicting pollutant concentrations, pattern recognition, and more. These approaches include Tree-based models, such as Random Forest (RF) and eXtreme Gradient Boosting (XGboost), Kernel-based models such ad Support Vector Machine Regression (SVMR), and Neural Networks-based models.

In the past decade, machine learning techniques have been widely employed for extracting PM information from satellite data. Over the last five years, their use has significantly increased, and there exists a substantial body of literature on this subject. Below, we present some of the most pertinent findings based on our analysis.

Table 3-1: Review on ML techniques employed for extracting PM from satellite data

PM Type	Method	Study Area	Satellite AOD	Reference
2.5	GTWNN	China	MODIS (1km)	(Li et al., 2022)
2.5	RF, GAMM, Empirical	Iraq, Kuwait	MODIS (1km)	(Li et al., 2021)
10	RF	South Africa	MAIAC (1km)	(Arowosegbe et al., 2022)
2.5	RF	Europe	MODIS (1km)	(Ibrahim et al., 2022)
2.5	XGBoost	Mexico City (Mexico)	MAIAC (1km)	(Gutiérrez-Avila et al., 2022)
2.5	BiLSTM	Guangdong Province (China)	MAIAC (1km)	(Miao et al., 2022)
10	MLP	Quito (Equador)	MODIS (1km)	(Alvarez-Mendoza et al., 2019)
2.5	BPNN	Dalian (China)	MAIAC (1km)	(Gu et al., 2022)
2.5	CNN	Taipei (Taiwan)	MAIAC (1km)	(Kibirige et al., 2023)
10-2.5		Baltimore, Chicago, Los Angeles, New York, St. Paul, Winston-Salem (US)	MAIAC (1km)	(Pedde et al., 2022)

2.5	RF	California	MISR fractional AOD (4 km)	(Geng et al., 2020)
2.5	XGBoost	China	MODIS AOD	(Pan, 2018)
2.5	RF XGBoost	China	Himawari-8/AHI AOT (5 km)	(Zang et al., 2021)
2.5	RF SVMR XGBoost	Italy	MAIAC (1km)	(Yu et al., 2022)
2.5	RF	Beijing-Tianjin Hebei (China)	MAIAC (1km)	(Li and Zhang, 2019)
2.5	RF	Guangdong-Hong Kong-Macao Greater Bay Area, China	MAIAC (1km)	(Chen et al., 2020)
2.5	RF GBRT	Yangtze River Delta (China)	Himawari-8 AOD (5 km)	(Karimian et al., 2023)
2.5 10	RF	Italy	MAIAC (1km)	(Stafoggia et al., 2019)
2.5	CNN	China	MODIS AOD	(Xue et al., 2019)
2.5	RF	Sichuan Basin (China)	Himawari-8 and VIIRS AOD	(Y. Liu et al., 2022)
2.5	RF GB SVMR	Tehran (Iran)	MAIAC (1km)	(Nabavi et al., 2019)
2.5	SVR RF XGBoost	Tehran (Iran)	MAIAC (1km)	(Bagheri, 2022)
10	GBRT	Germany	MAIAC (1km)	(Stirnberg et al., 2020)
2.5	CatBoost	Beijing-Tianjin Hebei (China)	MAIAC (1km)	(Ding et al., 2021)

As shown in the literature, AOD plays a pivotal role in estimating PM. However, uncertainties in explaining the spatio-temporal changes of PM persist within AOD. Hence, subsequent investigations have tackled this issue by integrating various supplementary factors to address aspects that AOD alone cannot clarify. This has led to the development of statistical and machine-learning models incorporating the following:

- Meteorological variables, encompassing wind speed and direction, air temperature, ground temperature, relative humidity, Planetary Boundary Layer Height, and precipitation.
- Variables associated with land cover and morphology, such as Land Cover Land Use (LCLU), spectral indices (e.g., NDVI, NDBI), and Digital Surface Model/Digital Terrain Model (DSM/DTM) along with the derived building morphological indicators.
- Socio-economic variables, including population density, the quantity and scope of emissive industrial sources, and energy consumption by residential and industrial users, Light at night.

- Variables derived from auxiliary information layers that indirectly provide insights into vehicle traffic.

The performance of the models adopted in the literature results from differences in the number of variables or in the different time scales of the same variables, denoted as the scale effect of variable contributions. Notably, this scale effect, prompted by the number of variables, doesn't manifest as a straightforward positive correlation between the number of variables and model performance. In more severe instances, it may lead to the issue of dimensionality, posing significant challenges for the model in terms of fitting and potentially resulting in overfitting during the training phase of the models. This phenomenon is attributable to the fact that an excessive number of variables can generate sparse, high-dimensional samples, making it challenging to clearly discern the relationship between dependent and independent variables. Moreover, some independent variables originate from the atmospheric circulation system and exhibit strong correlations.

Efficient feature selection contributes to the improvement of PM prediction models in several ways. First, it helps mitigate the curse of dimensionality, a common challenge in air quality data, where the number of features often exceeds the number of observations. By selecting the most relevant features, models become more interpretable, computationally efficient, and less prone to overfitting. Secondly, feature selection improves the generalisation of the model by reducing noise and irrelevant information. Air quality datasets often include a multitude of variables and not all of them contribute equally to the prediction of PM concentrations. Feature selection ensures that the model focuses on the most influential factors, leading to more robust and accurate predictions. In the literature about PM prediction there are different approaches:

- Correlation-based Feature Selection: by evaluating correlations between variables (es. Spearman correlation coefficient, or sensitive analysis), a subset of independent or weakly correlated variables is selected (Ibrahim et al., 2022; Kapadia and Jariwala, 2022; J. Liu et al., 2022)
- Importance of features from tree-based models: To reduce the complexity of the model, the algorithm excludes input variables according to the importance of the features derived from the tree-based model(Handschuh et al., 2023; Ibrahim et al., 2022)

Backward and forward Feature Elimination: iteratively add and remove the most and least informative features for the predictive model, based on the study of the importance of a feature (J. Liu et al., 2022).

There are other techniques on feature selection and feature extraction that can be applied to optimise the model. These include PCA (principal component analysis)

- Principal component analysis (PCA): it is a dimensionality reduction technique that can be used to project the original variables into a lower dimensional space, while still preserving as much variance as possible. It improves the performance of a model when applied to a set of features that are well correlated with each other (Azid et al., 2014)
- Boruta algorithm: it is to compare the importance of each feature in the original dataset with the importance of randomly shuffled, uninformative features by use of a random forest classifier (Kapadia and Jariwala, 2022)

Many auxiliary variables are characterized by higher resolutions than AOD data, such as land cover type, population density, and road network density. To manage this issue the procedures used essentially involve upscaling, i.e., transitioning from high resolution to low resolution. This operation serves to enhance the information of ground grid cells and mitigate the adverse effects of variable spatial heterogeneity on PM over larger areas. Conversely, a lack of information on auxiliary variables for smaller areas may lead to substantial estimation errors. Moreover, despite the accurate retrieval of AOD data from satellite remote sensing images, the existing aerosol products are constrained by coarse spatial resolution, impeding their ability to meet the growing demand for aerosol retrieval at intra-urban scale. Addressing this challenge has prompted numerous initiatives involving the improvement of AOT extraction from HR data such as Sentinel-2, Landsat and Sentinel-3 (Gitahi and Hahn, 2020; Hau et al., 2020; Yang et al., 2021) and the use of ML for the data fusion of differ sources of AOD data. In (Zhou et al., 2019), a Bayesian maximum entropy (BME) approach was employed to fuse data from eight publicly released aerosol products over China. The goal was to enhance both the quality and resolution of AOD. In (Wang et al., 2022) an Artificial Neural Network Sequential Downscaling Method (ASDM) was applied to translate time-series of AOD values from a coarse to fine scale by fusing GEOS-5 satellite data and modelled outputs, in conjunction with elevation data (with a 1 km resolution). In (Zhang et al., 2023) data from Himawari-8 GEO meteorological satellite and MODIS have been used as input to a hour-stage BME fusion approach.

These studies aim to enhance the temporal resolution of AOD data by employing gap-filling and data fusion methods to maximize data density, thus focusing on the time scale. Nevertheless, it has been found that observations across monthly, seasonal, and annual scales consistently yield better results compared to daily-scale assessments within the same model, irrespective of whether the estimation is for the short term, mid-to-long term, or long term (Xu et al., 2021).

4. ANALYSIS OF STUDY AREAS

This section contains a description of the use cases studies selected for APEMAIA project, reporting the collection of data for evaluating and understanding the environmental, climatic, geographical, and contextual conditions in, which the scientific project activities will take place.

The selected study areas include the metropolitan area of Rome (the primary target area chosen for MAIA) and the urban areas of Taranto and Bari (Figure 4-1: Study areas.Figure 4-1).

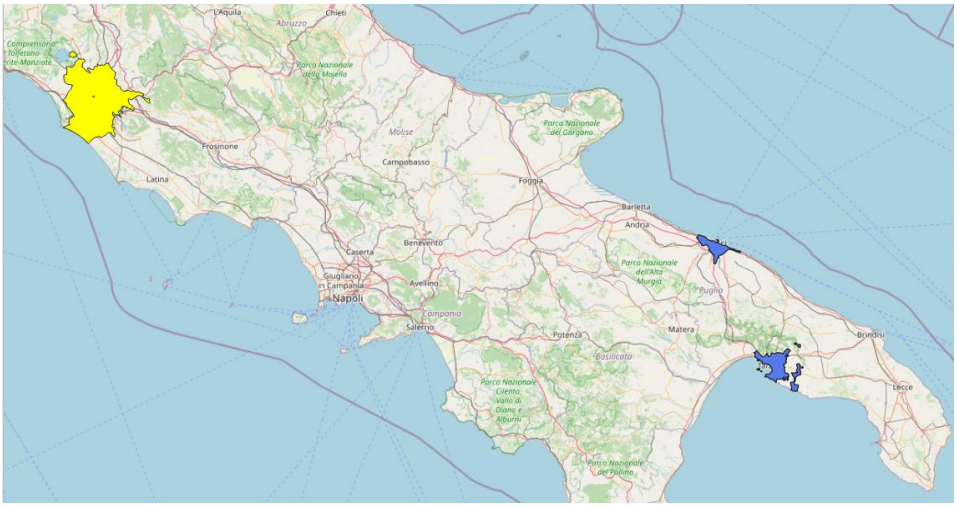

Figure 4-1: Study areas.

As the project is in its early stages, Section 4.1 primarily focuses on the urban areas of Bari and Taranto, with a brief introduction to the Rome study area in Section 4.2.

4.1 Apulian test sites: Bari and Taranto

The results of monitoring campaigns and studies conducted by ARPA Puglia to assess environmental impacts on air quality in cities reveal that the municipality of Taranto faces critical issues related to pollutants known to be harmful to human health.

Bari, on the other hand, generally enjoys good air quality, except for episodes of intense vehicular traffic occurring in the most densely populated areas of the town (Figure 4-2).

Commentato [PM11]: Anche questa sezione del documento è molto ben strutturata, ma sarebbe utili evidenziare le connessioni con gli user requirement. Quali user req. Si possono soddisfare sfruttando lo stato dell'arte del network di monitoraggio delle AoI? Cosa sarebbe necessario migliorare per soddisfare specifici user req.? Cosa verrà eventualmente già sviluppato in APEMAIA?

Commentato [MA12R11]: Le connessioni con gli User Requirements sono state aggiunte inserendo il riferimento agli User needs in Table 2.2. Nel testo sono state anche aggiunte alcune frasi che indicano quali dati siano utili agli scopi di APEMAIA, quali invece andrebbero verificati. Ad esempio, il ridotto numero di stazioni di misura del PM2.5. rispetto a quelle che misurano il PM10 potrebbe riflettersi sugli scopi della mappatura giornaliera. Di questo aspetto si tiene conto nella definizione delle specifiche implementative dell'applicazione. Altre frasi chiariscono come alcune informazioni siano trattate solo a scopo introduttivo, per una migliore comprensione dei siti di studio.

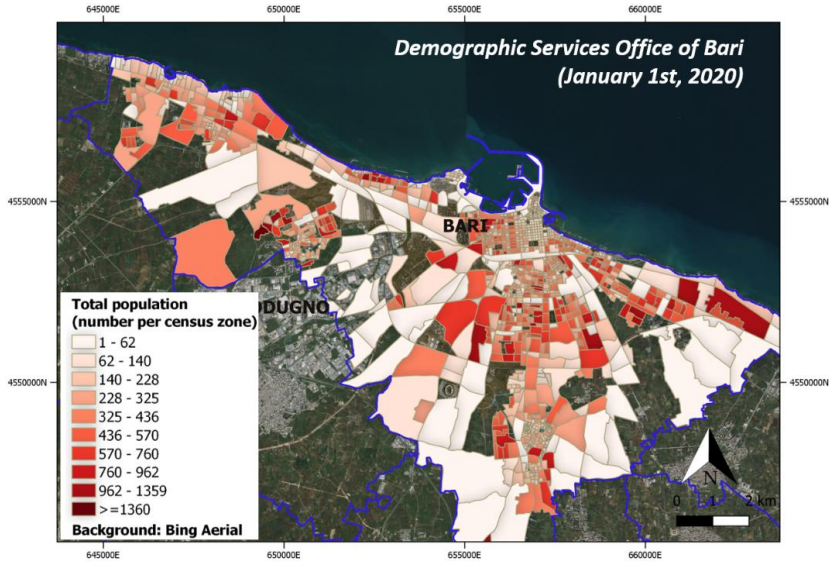

Figure 4-2: Population data per census area in Bari.

Nevertheless, despite Bari's relatively good air quality, it has been selected as the calibration/validation site for the APEMAIA project. This decision is driven by the availability of comprehensive information about the territory, including environmental data, urban morphology (Figure 4-3) and ancillary socio-economic data. Much of this information has been collected through previous projects such as SMURBS ("SMURBS project," 2017) and e-SHAPE ("e-SHAPE project," 2019).

Figure 4-3: Building layer from the"civilario comunale" dataset (on the left); urban 3D model obtained by combining LiDAR and building layers (on the right)(Aquilino et al., 2021).

4.1.1 Air quality monitoring network

The Regional Air Quality Monitoring Network (RRQA) of ARPA Puglia consists of 53 fixed stations (including 41 publicly owned and 12 privately owned), as shown in Figure 4-4.

ha eliminato: Figure 4-4

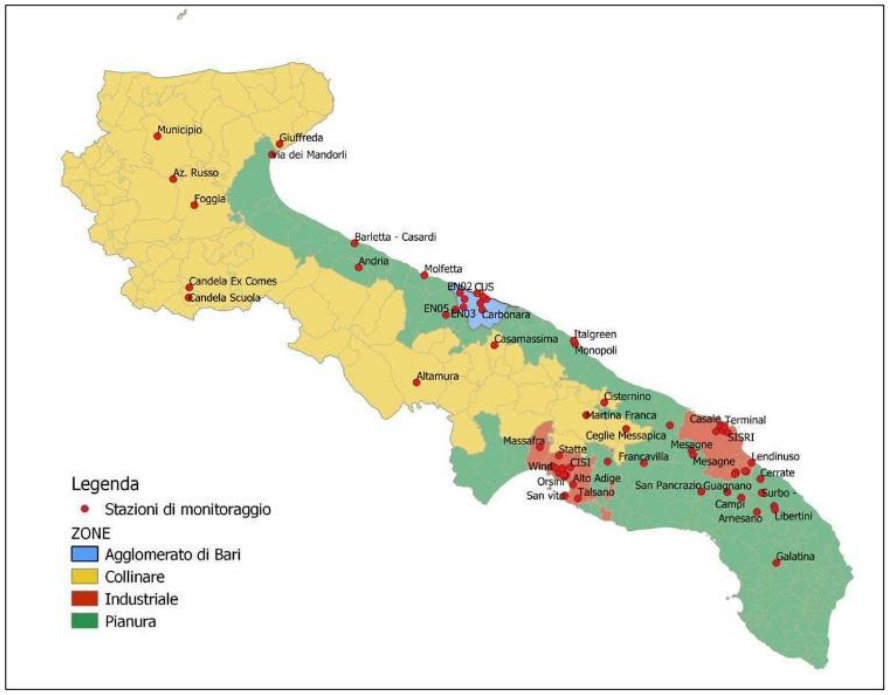

Figure 4-4. Air quality monitoring network of ARPA Puglia. The red dots mark the location of the monitoring stations, the cyan background color marks the metropolitan area of Bari and the green, yellow, and red colors indicate the lowland, hill and industrial areas.

These stations analyse air quality related to traffic in urban and suburban areas, background air quality (in either urban, suburban, or rural areas), and industrial air quality (again, in urban, suburban, and rural areas).

Of the total 53 air quality monitoring stations, 8 are located in the metropolitan area of Bari (5 in urban/suburban areas and 3 in the industrial area of the city) (Figure 4-5), 18 in lowland areas, 9 in hilly areas and 18 in the industrial areas around Brindisi and Taranto.

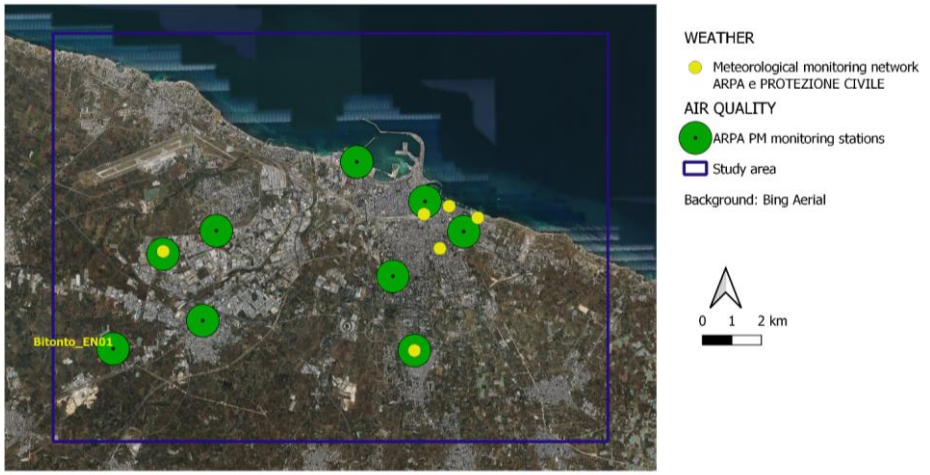

Figure 4-5: Air quality monitoring stations in Bari (green marker). Bitonto_EN01 is a private air quality station not managed by ARPA Puglia.

Taranto, due to its severe air pollution problems, has 3 stations in the urban/suburban area and 6 monitoring the air quality in the large industrial area around the city.

All 53 stations monitor both PM10 and NO_2 (in the latter case, with one exception), while about 30 stations can also monitor PM2.5. The levels of O_3, CO, SO_2 and BTX (i.e., benzene, toluene, and xylene) are measured at between 1/3 and 1/2 of the monitoring sites, most of, which are located in urban traffic or industrial areas.

In compliance with both European and national regulations, specifically Legislative Decree 155/2010(DLGS 155, 2010) and the European Air Quality Directive 2008/50/EC (AQD, 2008), a Flexible Air quality Regional Model (FARM) system has been recognized as a method for assessing air quality, along with the standard measurements conducted at fixed sites. The FARM model relies on a set of equations governing the chemical and transport characteristics, enabling the simulation of concentrations for both gaseous and particulate pollutants across the entire region. This system is anchored in a three-dimensional Eulerian model that addresses the transport and chemistry of the atmosphere. By considering pollutant emissions from various sources (both natural and anthropogenic) and incorporating the physical state of the atmosphere (wind fields, temperature, humidity, turbulence, etc.), the model can generate three-dimensional hourly concentration fields for different pollutants. Such a model implements chemical mechanisms of varying complexity for the treatment of photochemistry and particulate matter, treating aqueous and heterogeneous phase chemistry, as well as dry and wet deposition processes.

The modelling system uses several datasets as input, which are processed appropriately by the EMMA, GAP, and SURFPRO preprocessors to generate all the data necessary for the FARM chemical air quality model on the simulation domains. Specifically, to run the simulation for the year 2021, FARM was fed with:

- the meteorological analysis fields, developed by the Physical Agents Service of ARPA Puglia (www.arpa.puglia.it/web/guest/servizio_meteo) for the year 2022. These fields were generated with the WRF (Weather Research and Forecasting) prognostic meteorological model;
- the emission inventories, INEMAR for the Puglia region and ISPRA for the neighbouring regions, duly integrated and updated with available information;
- boundary conditions for 2022 derived from operational prediction modelling systems such as QualeAria (for homogeneous, gaseous and particulate phase pollutants, on a three-hourly time basis). For heavy metals, dioxins and furans, climatological elaborations produced by EMEP-MSCE models were used.

The modelling system was applied simultaneously to a domain covering the region of Puglia with a spatial resolution of 4 km and to a domain, grafted on the previous one, covering only the provinces of Brindisi, Lecce and Taranto with a resolution of 1 km (Figure 4-6).

ha eliminato: Figure 4-6

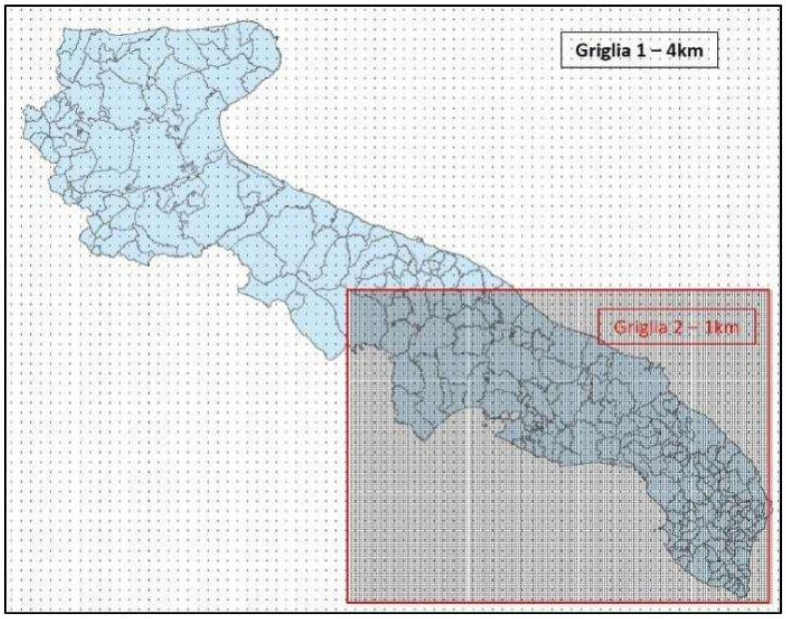

Figure 4-6: Model grid used for the entire region, overlaid with the higher resolution grid used for the three southern provinces of Brindisi, Lecce, and Taranto.

To ensure the proper and appropriate use of models, the European Environment Agency (EEA) and the European Commission's Joint Research Centre (JRC) have established the FAIRMODE working

group to identify and promote the use of best practices in air quality modelling. For the assessment of quality status using models, FAIRMODE recommends the use of standardized quality objectives for models that can be used to assess the quality of a particular modelling application carried out within the framework of European air quality directives.

The quality of the simulation was assessed using DELTA TOOL, a statistical analysis software developed under FAIRMODE to evaluate the performance of the model in specific applications. This software, which deals with the pollutants PM10, PM2.5, NO_2, O_3 and is widely used by the scientific community, is based on the statistical comparison of modelled and measured data from air quality monitoring stations and provides a quick evaluation of the model performance in terms of various indicators and statistical plots.

The statistical evaluation conducted with the DELTA TOOL showed for the pollutants PM10, PM2.5, NO_2 and O_3 the ability of the model to adequately reconstruct the measured concentrations on both the regional and city-scale grids, as the quality criterion appears to be always met.

To further improve the model reconstruction of concentration levels and their spatial distribution, the model simulations were combined with the measured data. Such integration, called assimilation, allows for a realistic assessment of air quality status by making the best use of each type of measurement or estimate, each with its own degree of accuracy and reliability.

In assimilation techniques, the two-dimensional concentration field produced by the chemical model represents the background field with, which the interpolation procedure is initialized; this field is subsequently modified as observations are introduced into the model grid through a specific interpolation scheme.

Specifically, the Optimal Interpolation (OI) method was used, which allows the interpolation of observations located on a domain using, as a first approximation, the background (or background) field generated by the model. Since both the observations and the background field may contain errors, OI constructs the analysis field by minimizing the variances of the errors of the observations and the background field. The resulting field is optimal in the sense that it has the minimum variance.

Assimilation involved the pollutants NO_2, SO_2, PM10, PM2.5, O_3, CO and C_6H_6 monitored by the air quality monitoring network operated by ARPA. Specifically, assimilation occurred on an hourly basis for NO_2, SO_2, O_3, CO and C_6H_6 species, while for PM10 and PM2.5 species it occurred on a daily basis.

The wide model grid used for the urban area of Bari adds little information to the air quality data measured at the 8 stations of the urban area of the city but can be of paramount importance to

better evaluate the air quality conditions in the area of Taranto, using MAIA satellite data, where air pollution due to its industrial environment is still a serious threat to the health of its inhabitants.

The analysis of the PM10 concentration trends between 2010 and 2022 does not show a significant upward or downward trend for the stations under consideration here. In other words, the measured concentration of PM10 has remained basically stable over the last 12 years. No PM2.5 data is available for these stations.

The results of the modelling system for PM10 The model shows no exceedance of the annual concentration limit either on the regional domain at 4 km resolution or on the 1 km detail of the provinces of Brindisi, Lecce and Taranto (Figure 4-7).

ha eliminato: Figure 4-7

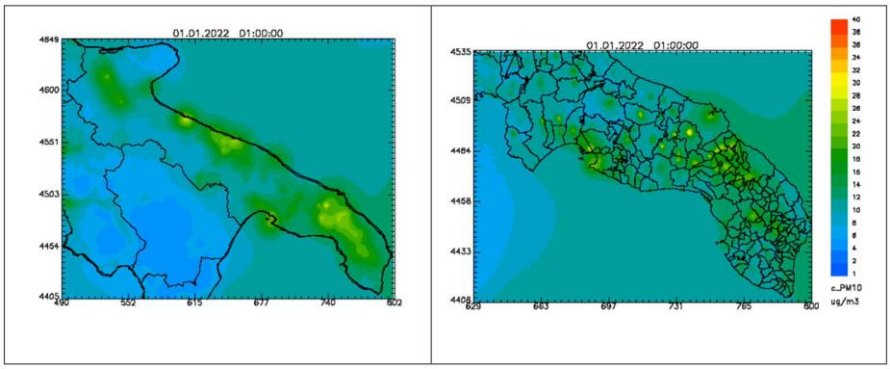

Figure 4-7: Maps of the annual average concentration of PM10 (µg/m³) over the whole region of Apulia (left) and the provinces of Lecce, Brindisi, and Taranto (right).

The highest annual average concentrations are estimated on the regional domain at the urbanized areas of some municipalities in the metropolitan area of Bari, some municipalities in the provinces of Brindisi and Lecce, the municipality of Taranto, and the municipality of Foggia.

The analysis of the NO_2 concentration trends between 2010 and 2022 shows a significant downward trend for the Bari-Carbonara station and a constant level of NO_2 concentration for Taranto-San Vito. All other parameters are not monitored in the 4 stations considered here.

Two scenarios have been identified impacting on PM2.5/10 measurements for the study areas: dust storms and marine aerosols events.

In Mediterranean regions, occurrences of natural dust intrusion from the Sahara are not unusual. Particulate matter ascends from the desert due to updrafts and is transported to Europe by the wind, predominantly impacting the southernmost areas of the continent. Intrusions of Saharan dust

can lead to an abnormal increase in PM10 concentration values, and in some cases, they may contribute to the surpassing of legally established limit values. Consequently, it is essential to discern this contribution, that can also affect the performances of ML algorithms to extract PM.

An approach to addressing this issue could involve identifying events characterized by this scenario and attempting to model the relationships between atmospheric conditions, aerosol optical depth (AOD), and other predictors with PM10 by constructing a specialized model. In 2011, the European Community formulated a guideline (European Commission, 2011) outlining the identification of Saharan advection episodes and the estimation of increases in PM10 concentrations attributed to them. In accordance with the procedure outlined, ARPA Puglia identifies intrusion episodes that lead to daily exceedances of the limit value, ensuring that such exceedances are not included in the annual count. These events are subsequently documented for each year and are typically regarded as the most significant occurrences. Moreover, ARPA Puglia modelling system, designed for forecasting and evaluating air quality, has the capability to simulate the impact of cross-border desert dust advections on an hourly basis and, in this way, to identify the events.

Concerning marine aerosols, in collaboration with ARPA Puglia, we are currently exploring the definition of empirical criteria to identify conditions in which marine aerosols dominate.

4.1.2 Meteorological and climatic conditions

Meteorological conditions have a very significant impact on air quality. Atmospheric parameters such as wind, temperature, precipitation, and solar radiation can affect the dispersion and removal of pollutants or create conditions conducive to the formation of secondary pollutants.

The Region of Apulia has an extensive network of meteorological stations operated by the Civil Protection Service (Figure 4-8) and an additional network of meteorological stations operated by ARPA, consisting of 5 automatic stations located at the provincial headquarters (Bari, Brindisi, Foggia, Lecce and Taranto) and 19 additional stations accompanying the air quality monitoring network (Figure 4-8).

ha eliminato: Figure 4-8

ha eliminato: Figure 4-9

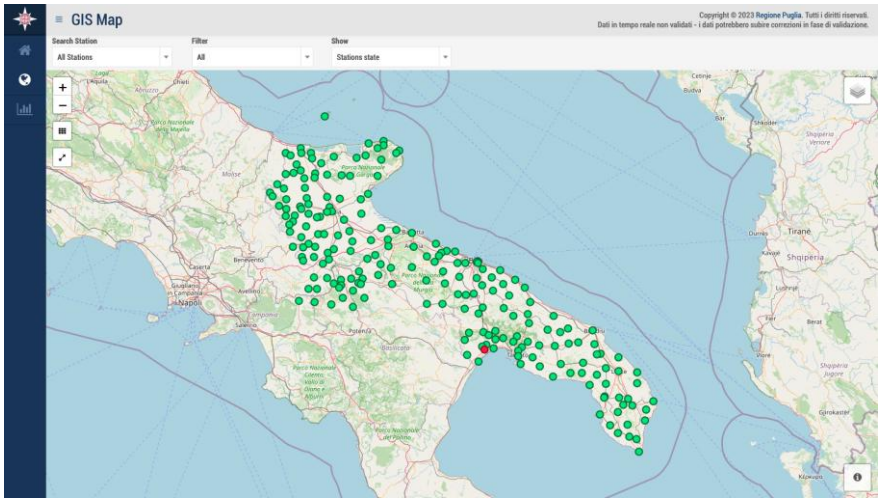

Figure 4-8: Meteorological monitoring network of the Civil Protection Service.

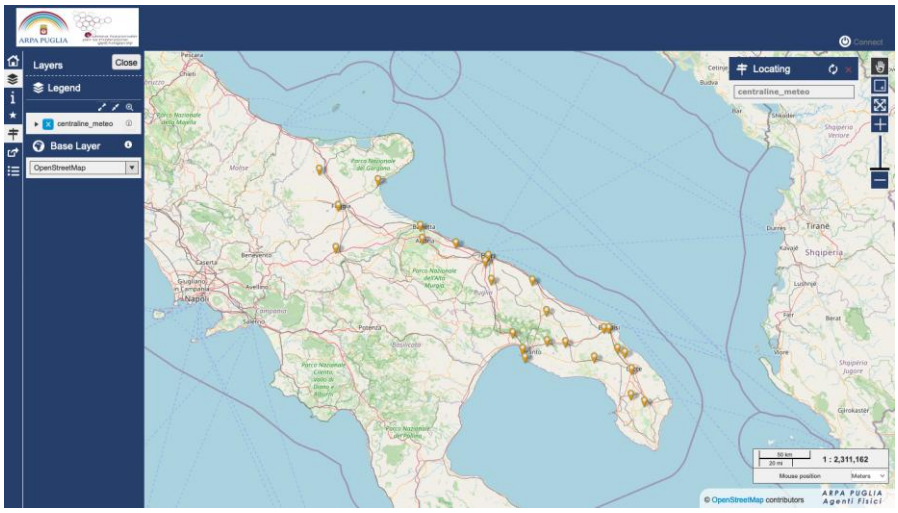

Figure 4-9: Meteorological monitoring network of ARPA Puglia.

For the meteorological characterization of the Bari area, there is no single station with all the meteorological parameters of interest. For this reason, the station "Bari-Idrografico" of the Civil Protection Service of the Apulia Region, located along the coast of Bari at a distance of about 5 km

from the ARPA Apulia station, is used together with the ARPA meteorological station (see Figure 4-5, yellow marker). Wind and precipitation data from this station are used, while the ARPA air quality monitoring station called "Bari-Carbonara" is used for temperature and global solar radiation data.

For the Taranto area (Figure 4-10), ARPA uses the "Meteoflux" meteorological station located in the area of the ENI plant near the northern border of the city a few hundred meters from the coast, representative of the meteorological conditions that characterize the Taranto industrial zones. Since this station does not have a sensor for measuring global radiation and, for the year 2022, valid precipitation data, the station of the regional air quality network, located at Capo San Vito, which is also equipped with meteorological instrumentation, was considered for the aforementioned variables (Figure 4-10).

ha eliminato: Figure 4-9Figure 4-9

ha eliminato: Figure 4-9

ha eliminato: Figure 4-10

ha eliminato: Figure 4-10

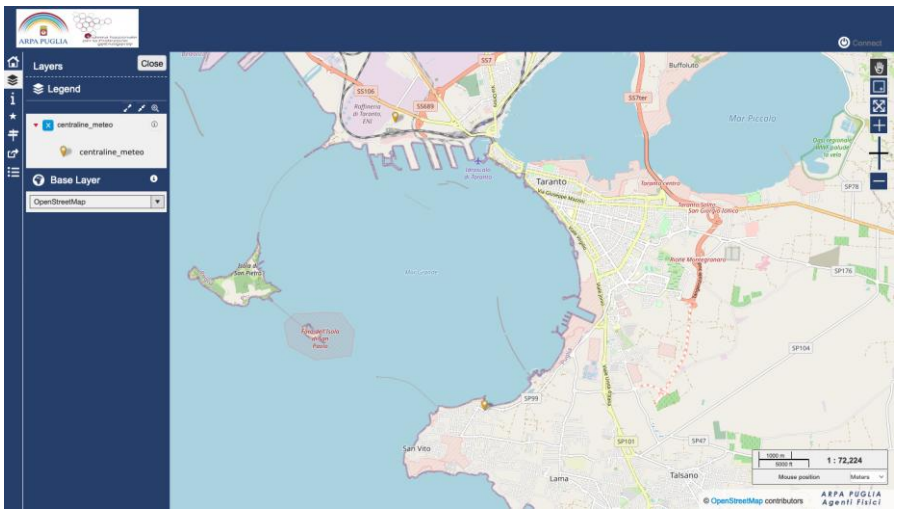

Figure 4-10: The two ARPA meteorological stations in Taranto (orange markers).

The main characteristics of the 4 aforementioned stations are summarized in Table 4-1.

Table 4-1: Main characteristics of the meteorological stations used by ARPA in the urban areas of Bari and Taranto.

Station	Coodinates WGS84-UTM33		Sensor height	Measured parameters
	X-UTM (km)	Y (UTM) (km)	(m)	
Bari Idrografico (Civil Protection)	657.991	4 553.753	30	wind speed and wind direction (10 min), temperature, relative humidity (30 min), precipitation (5 min)
Bari Carbonara (ARPA)	656.837	4 548.966	5	wind speed, wind direction, temperature, relative humidity, atmospheric pressure, global radiation, precipitation (60 min)
Taranto-Meteoflux (ENI)	685.952	4 485.618	6	wind speed, wind direction, temperature, precipitation (60 min)
Taranto-San Vito (ARPA)	688.791	4 477.122	10	wind speed, wind direction, temperature, relative humidity, atmospheric pressure, global radiation, precipitation (60 min)

Considering the percentage of validity of the data measured during 2022 from each station, it can be seen that in most cases the coverage is very close to 100 percent, with the exception of the Taranto-Meteoflux station for the precipitation parameter (57 percent valid data) and the Bari-Carbonara station, where the percentage of valid precipitation data is 91 percent. For technical reasons, wind intensity and direction data could not be acquired at the latter station during 2022. Therefore, for the anemological characterization of Bari, the wind measurements of the Bari-Idrografico weather station, which is part of the regional civil protection monitoring network, were used. In addition, for precipitation, although data measured by the Bari-Carbonara station were available in a percentage of 91%, it was preferred to use those of Bari-Idrografico because they were available with a better coverage (100%).

Air quality is strongly correlated with wind speed and direction. The following plots show the frequency counts by wind direction and speed measured at the two stations of Bari-Idrografico (Figure 4-11) and Taranto-Meteoflux (Figure 4-12) for the years 2022 (left) and 2021 (right). Patterns of wind conditions and air quality might change with the seasons.

ha eliminato: Figure 4-11

ha eliminato: Figure 4-12

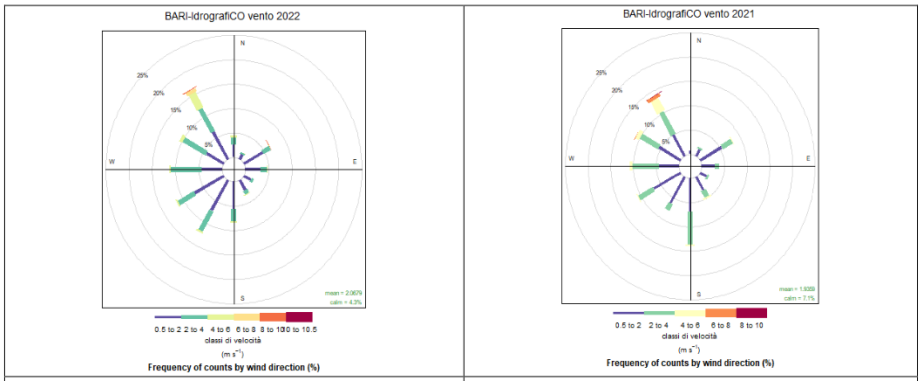

Figure 4-11: Bari-Idrografico station: annual wind rose for the years 2022 (left) and 2021 (right).

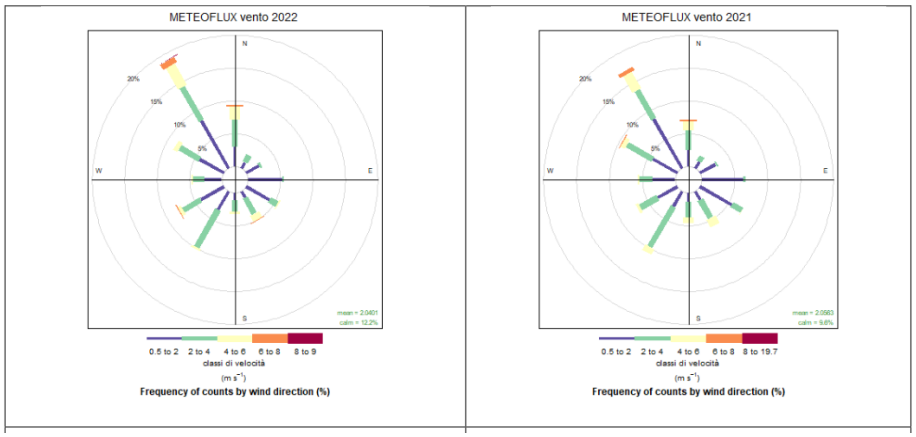

Figure 4-12:Taranto-Meteoflux station: annual wind rose for the years 2022 (left) and 2021 (right).

To take into account the seasonal variations, the seasonal wind roses for the two stations Bari-Idrografico and Taranto-Meteoflux are shown in Figure 4-13 and Figure 4-14, respectively.

ha eliminato: Figure 4-13

ha eliminato: Figure 4-14

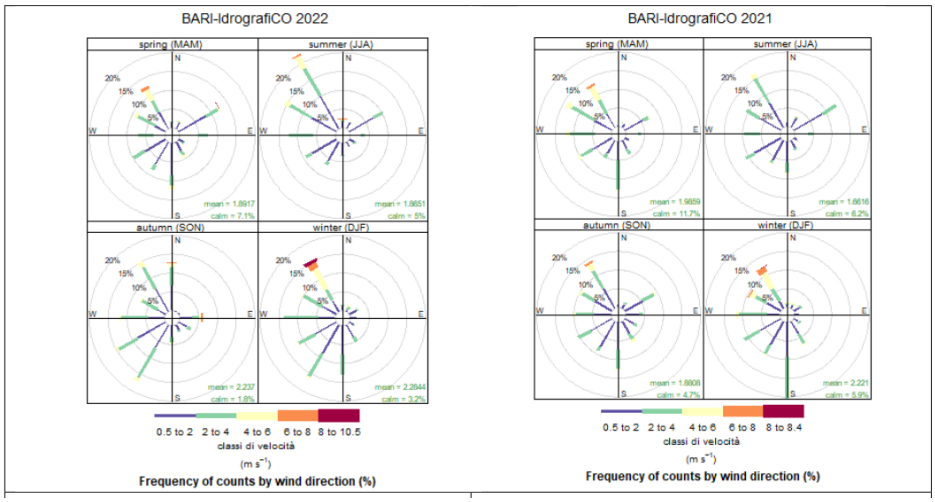

Figure 4-13: Bari-Idrografico station: seasonal wind rose for the years 2022 (left) and 2021 (right).

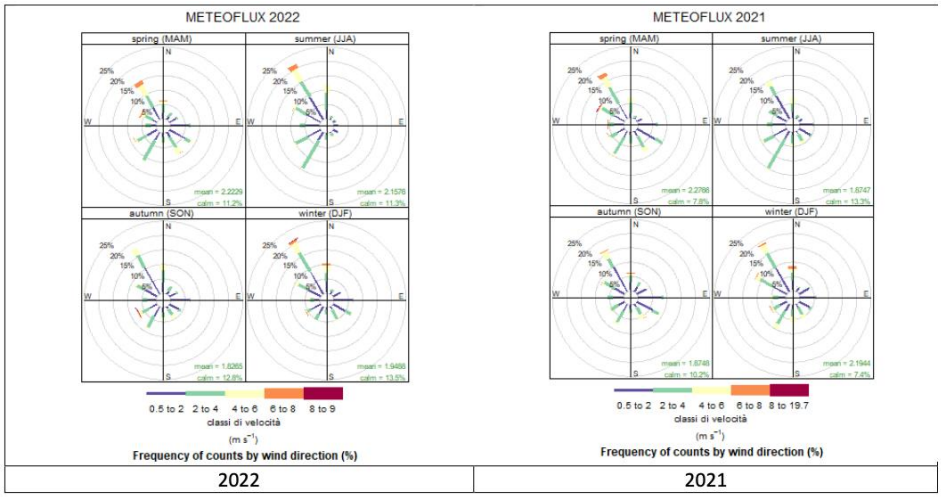

Figure 4-14: Taranto-Meteoflux station: seasonal wind rose for the years 2022 (left) and 2021 (right).

It should be noted that wind speed is a variable that can be strongly influenced by the context in, which the anemometer is placed. In fact, the wind can be affected by the effects of local site characteristics (such as roughness, elevation, exposure, presence of obstacles, etc.). This is

particularly evident when comparing meteorological data between different stations. In this regard, it can be noted that the stations located in the centre of the urbanized area generally have lower wind intensity values than the other stations. Nevertheless, all the plots shown here show a very similar wind speed distribution between 2022 and 2021.

On an annual basis, there is a slight increase in the average wind speed and a decrease in the percentage of calm winds for the Bari-Idrografico station between 2021 and 2022. The wind roses processed on a seasonal basis show no significant changes with respect to the previous year, except for a decrease of the S component in all seasons. In 2022, an increase of the dominant NO component is observed in the summer and winter seasons (Figure 4-11 and Figure 4-13).

For the Taranto-Meteoflux station, located in the industrial area, there is a significant increase in the percentage of calm winds, which increases from 9.6% (in 2021) to 12.2% (in 2022), while the average annual wind intensity value remains basically unchanged (Figure 4-12 and Figure 4-14).

Nearly all stations for, which the climatological dataset is available show monthly mean temperatures in 2022 that are higher than the climatological mean, with the notable exception of March. It should be noticed that the mean temperature is the average of the temperature over the 24-hour day. In addition, all stations in 2022 show a decrease in monthly mean temperatures in March and August and an increase in April, May, June, October and December compared to the previous year. The two stations of Bari-Carbonara and Taranto-Meteoflux are in line with this scenario, with the notable exception of March, whose monthly mean temperature is slightly lower than the climatological mean. (Figure 4-15).

ha eliminato: Figure 4-11

ha eliminato: Figure 4-13

ha eliminato: Figure 4-12

ha eliminato: Figure 4-14

ha eliminato: Figure 4-15

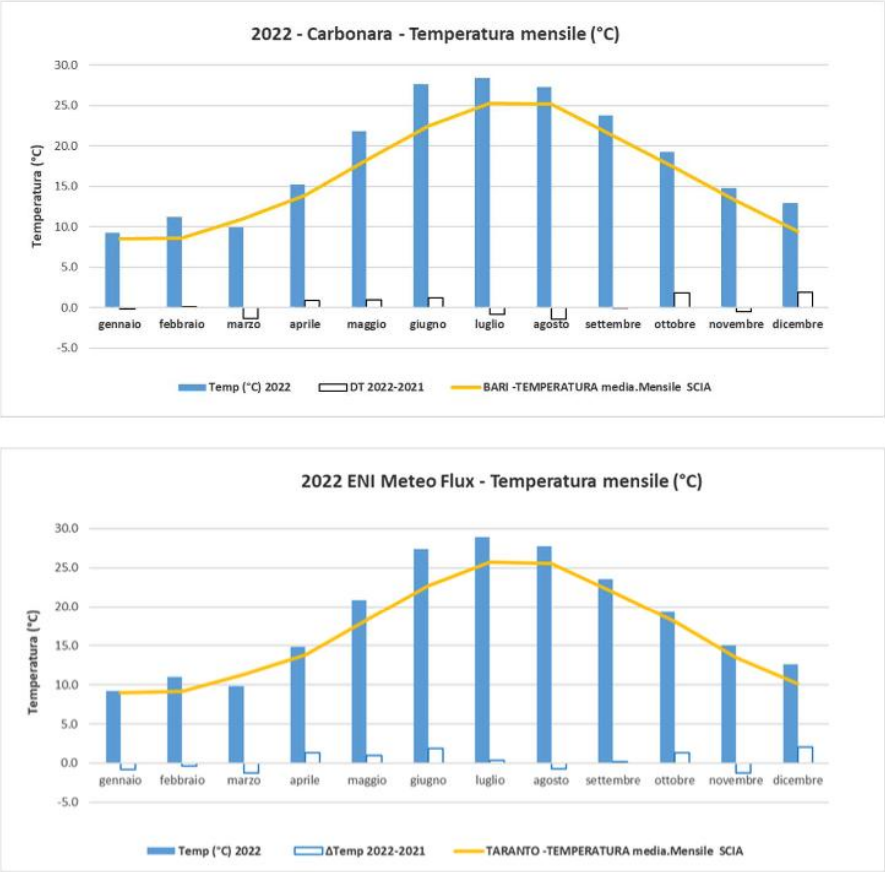

Figure 4-15: Monthly average temperature measured at the meteorological stations of Bari-Carbonara and Taranto-Meteoflux.

The increase in mean temperatures observed in recent years can be linked to an increase in the absorption and retention of solar radiation, contributing to a compounding warming effect. In the meteorological stations located in the urban areas of Bari (Figure 4-16) and Taranto (Figure 4-17), an increase in monthly average global radiation is observed in all months of 2022 with respect to the previous year, with the exception of August for both cities and the month of December for Taranto.

ha eliminato: Figure 4-17

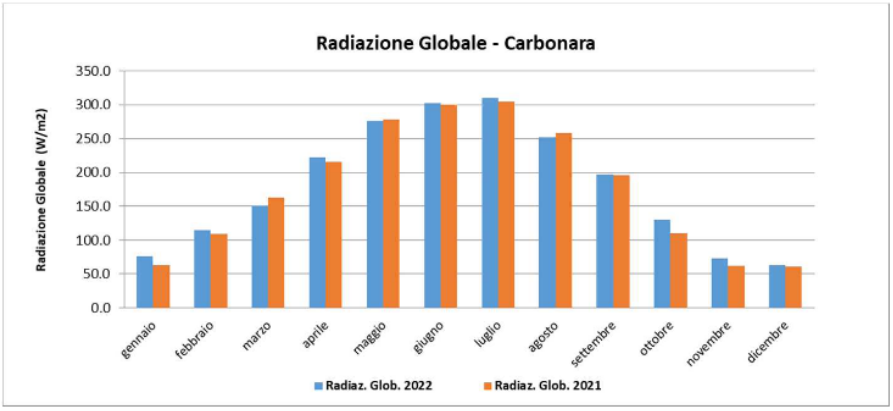

Figure 4-16: Monthly average global radiation in Bari-Carbonara.

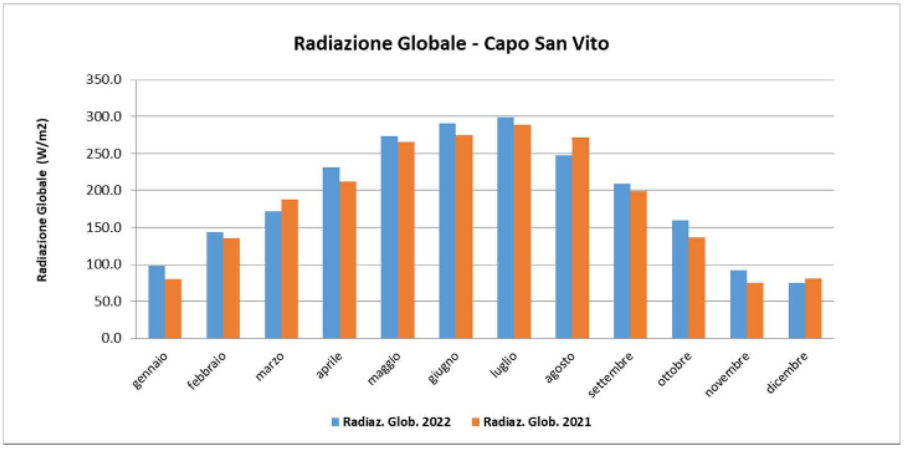

Figure 4-17: Monthly average global radiation in Taranto-San Vito.

Regarding the amount of precipitation assessed on an annual basis, in 2022 there is a general increase compared to the previous year at all stations except Taranto-San Vito, where there is a decrease of about 40 mm in the annual cumulative precipitation compared to the previous year, accompanied by a decrease in the number of rainy days (Figure 4-18). The Bari-Idrografico station, on the other hand, shows an increase in the number of rainy days in 2022, with a concomitant increase in cumulative annual precipitation (Figure 4-18).

ha eliminato: Figure 4-18

ha eliminato: Figure 4-18

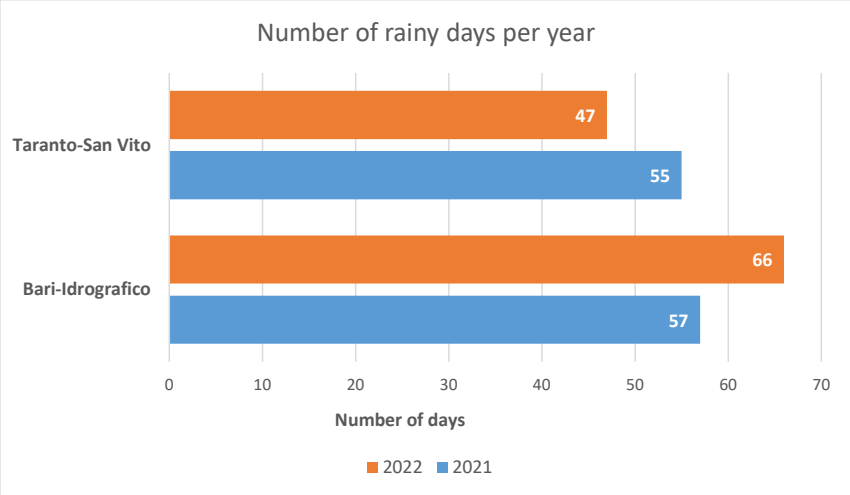

Figure 4-18: Number of rainy days per year in the meteorological stations of Bari-Idrografico and Taranto-San Vito.

On a seasonal basis, during the summer season, there is an increase in both precipitation and the number of rainy days for the Bari-Idrografico station, and conversely a decrease in the same parameters for Taranto-San Vito. During the winter season, an increase in the number of rainy days is observed at Bari-Idrografico with a simultaneous decrease in the measured precipitation, while at Taranto-San Vito both the precipitation and the number of rainy days decrease in 2022 compared to the previous year. Lastly, no major changes are observed for spring and autumn.

4.2 Primary Target Area: Roma

4.2.1 Air Quality Monitoring Network

The Regional Air Quality Monitoring Network (RRQA) of ARPA Lazio consists of 55 fixed monitoring stations, 46 of which are included in the Regional Air Quality Assessment Program. These stations are located throughout the region as follows (Figure 4-19):

- 5 stations in the Apennine area;
- 10 stations in the Sacco Valley area;
- 18 stations in the Rome agglomeration (of which 2 are not included in the Regional Assessment Program);
- 22 stations in the Litoranea area (of which 7 are not included in the Regional Assessment Program.

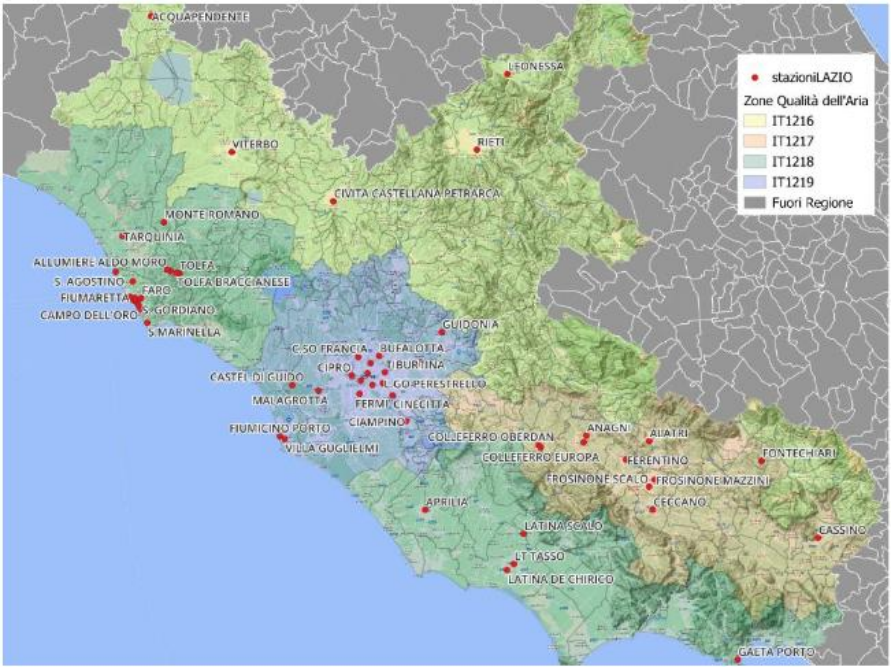

Figure 4-19: Air quality monitoring network of ARPA Lazio. The red dots mark the location of the monitoring stations, while the background color marks the four areas into which the region is divided: (light blue) Rome agglomeration, (light green) Apenine area, (dark green) Litoranea area, and (light brown) Sacco Valley area. shows the location of the monitoring stations on the regional territory.

All of the 55 stations monitor both PM10 and NO2 (in the former case, with three exceptions), while 25 stations can also monitor PM2.5. The levels of O_3, CO, SO_2 and BTX (i.e. benzene, toluene, and xylene) are measured in a small number of the monitoring sites, most of which are located in urban traffic or industrial areas.

A modelling system similar to the one used by ARPA Puglia is used to simulate the spatio-temporal distribution of pollutant concentrations over the entire regional territory, as required by Legislative Decree 155/2010. The system is used in both prediction and reconstruction modes. The system provides daily:

- forecasts of the spatial distribution of the concentration of air pollutants over the regional territory up to 120 hours (5 days) ahead;
- reconstruction of the air quality status of the previous days: information necessary for municipalities to adopt measures to combat pollution;

- near-real time reconstruction that reproduces the most probable snapshot of the regional air quality status and the meteorological and micrometeorological causes that determine it obtained through the acquisition, with a delay of a few hours, of the concentration measurements of the monitoring network integrated with the modelling system using data assimilation techniques.

The modelling system is also used to carry out the Annual Air Quality Assessment over the regional territory. At the end of each calendar year, the system is used to verify compliance with the limits set by the standard over the entire territory by integrating, through data fusion techniques and objective estimation statistics, the concentration fields produced by the modelling chain with the measurements provided by both the fixed monitoring stations and the measurement campaigns carried out with mobile means.

The system architecture consists of the following modules:

- prognostic weather models, RAMS and WRF, for downscaling weather forecasts from the synoptic scale (forecasts produced by the US-NOAA) to the regional and urban scales;
- interface modules for fitting the meteorological fields produced by RAMS and WRF to the computational domains of the air quality model, description of atmospheric turbulence, and definition of dispersive parameters (GAP/SURFpro);
- processor (Emission Manager) for the preparation of the hourly gridded emission input of the to the air quality model, from data of regional (ARPA Lazio) and national (ISPRA) emission inventories;
- Eulerian multigrid model (FARM) for dispersion and chemical reactions of pollutants in the atmosphere;
- post-processing modules for the combination of the results of the air quality model with the values measured on the territory and the calculation of the parameters necessary for the verification of compliance with the legal limits pursuant to Legislative Decree No.155/2010.

The modeling system is applied simultaneously to the Lazio region and to three portions of the territory with greater spatial resolution: the Rome agglomeration area, the entire Sacco Valley, and the Civitavecchia district (Table 4-2).

ha eliminato: Table 4-2

Table 4-2: Domain area and resolutions of the model grids used for the whole Lazio region and for three specific parts of the regional territory.

Area	**Domain (km^2)**	**Resolution (km^2)**
Lazio	240x200	4x4
Rome	60x60	1x1
Sacco Valley	116x70	1x1
Civitavecchia	24x24	1x1

The technique of nesting computational domains thus makes it possible to describe the effects of sources outside the area of interest and processes dominated by spatial scales larger than the urban scale (Figure 4-19).

ha eliminato: Figure 4-20Figure 4-20

ha eliminato: Figure 4-20

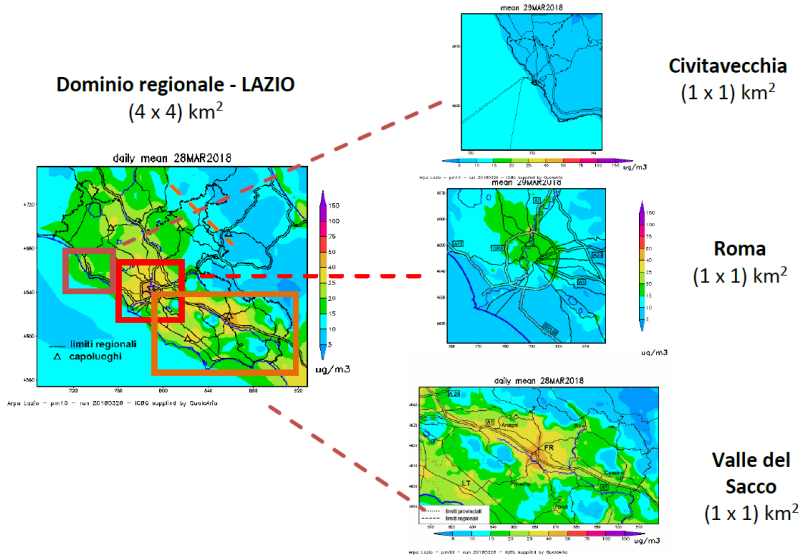

Figure 4-20: Modeling domain of the whole Lazio region (left) and enlargements of the higher resolution domains used for three specific areas of the region (right).

Hourly emissions are calculated by means of a processor (EMMA) that allows spatial disaggregation, temporal modulation, and speciation of volatile organic compounds for point, areal, and linear source inventory data using regional numerical mapping as supporting information.

The preparation of the emission files to be used as input to the FARM code is carried out from different data sources derived from the harmonization of the national inventory, by ISPRA, and regional inventory, by ARPA Lazio:

- diffuse emissions from all sectors except for highway sections and urban and suburban emissions in the municipality of Rome;
- point source emissions;
- traffic estimates on the Rome network;
- AISCAT data for highway emissions.

The Rome agglomeration and the Sacco Valley emerge as the most critical areas (Table 4-3). In fact, in the Sacco Valley exceedances of the limit values for PM10 and PM2.5 were recorded, and in the Rome Agglomeration noncompliance with the limits imposed by the standard concerned PM10 and NO2. Regarding O3, the exceedance of the target value for the protection of vegetation and for the protection of human health concerned the Littoral Zone.

Area	SO2	NO2	PM10	PM2.5	CO	O3	Benzene	B(a)P	Metals
IT1219 Rome agglomeration									
IT1216 Apenine									
IT1218 Litoranea									
IT1217 Sacco Valley									

Table 4-3: Summary overview of exceedances detected by the Lazio air monitoring network in 2021.

4.2.2 Meteorological and climatic conditions

The following is a description of the main meteorological characteristics of the region, focusing on the meteorological phenomena that play a crucial role in the dispersion and reduction of air pollutant concentrations. These include precipitation, winds, and variables associated with atmospheric turbulence. Meteorological conditions have a profound influence on air quality as they affect the dispersion and elimination of pollutants, as well as create an environment that may foster the formation of secondary pollutants. Key atmospheric parameters such as wind, temperature, precipitation, and solar radiation all contribute to these effects.

In addition, the complex orographic structure of the territory of Lazio greatly influences the meteorological and micrometeorological characteristics of the territory, which are the basis of the dispersion processes of pollutants released into the atmosphere.

The Regional Meteorological Network (RMR) is composed by 8 fixed stations that make up the Micrometeorological Network (RMM) and by 3 additional mobile vehicles and by a set of sensors on the air quality stations that make up the Conventional Meteorological Network (RMC).

Four RMM stations are located in the Rome agglomeration and the others in the four provincial capitals of the region (Figure 4-21 and Table 4-4).

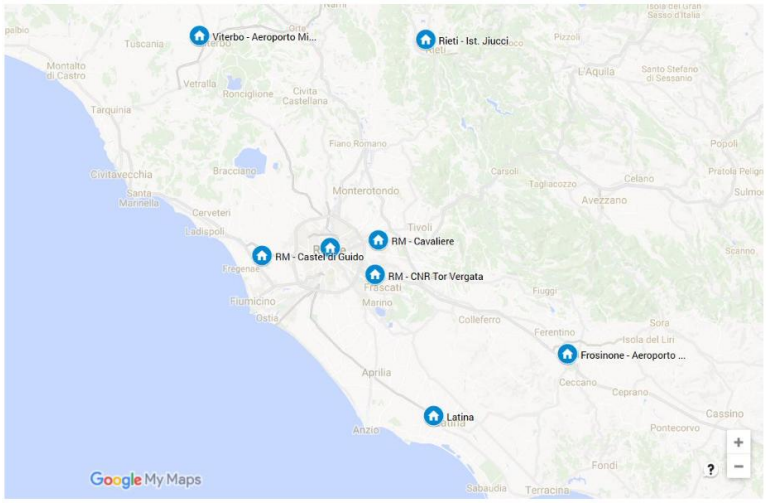

Figure 4-21: Stations of the Micrometeorological Network (RMM) of ARPA Lazio.

The stations are equipped with classical meteorological sensors (temperature, humidity, pressure and precipitation) associated with instruments dedicated to the characterization of the turbulent and radiative properties of the atmosphere, which are important for the dispersion of pollutants and air quality in general (sonic anemometers, pyranometers and pyrgeometers).

Using data from ARPA Lazio's network of micro-meteorological stations, it is possible to show the distributions of wind intensity and direction at 8 points in the region, 4 of which belong to the Rome agglomeration and the other 4 in the remaining provincial capitals of the region.

The following images show the wind roses in the order: Rome Tor Vergata and Latina, Rome Tenuta del Cavaliere and Rome Castel di Guido (Figure 4-22), Rieti and Frosinone, Rome Via Boncompagni and Viterbo (Figure 4-23).

Table 4-4. Location of the 8 stations of ARPA Lazio Micrometeorological Network (RMM).

Zone	Code	Station name	Latitude	Longitude
IT1219 Rome agglomeration	AL001	Rome Tor Vergata	41.8417	12.6476
	AL003	Rome Tenuta del Cavaliere	41.9290	12.6583
	AL004	Rome Castel di Guido	41.8894	12.2664
	AL007	Rome Boncompagni	41.9093	12.4965
IT1217 Sacco Valley	AL006	Frosinone	41.6471	13.2999

IT1218 Litoranea	AL002	Latina	41.4850	12.8457
IT1216 Apenine	AL005	Rieti	42.4294	12.8191
	AL008	Viterbo	42.4308	12.0625

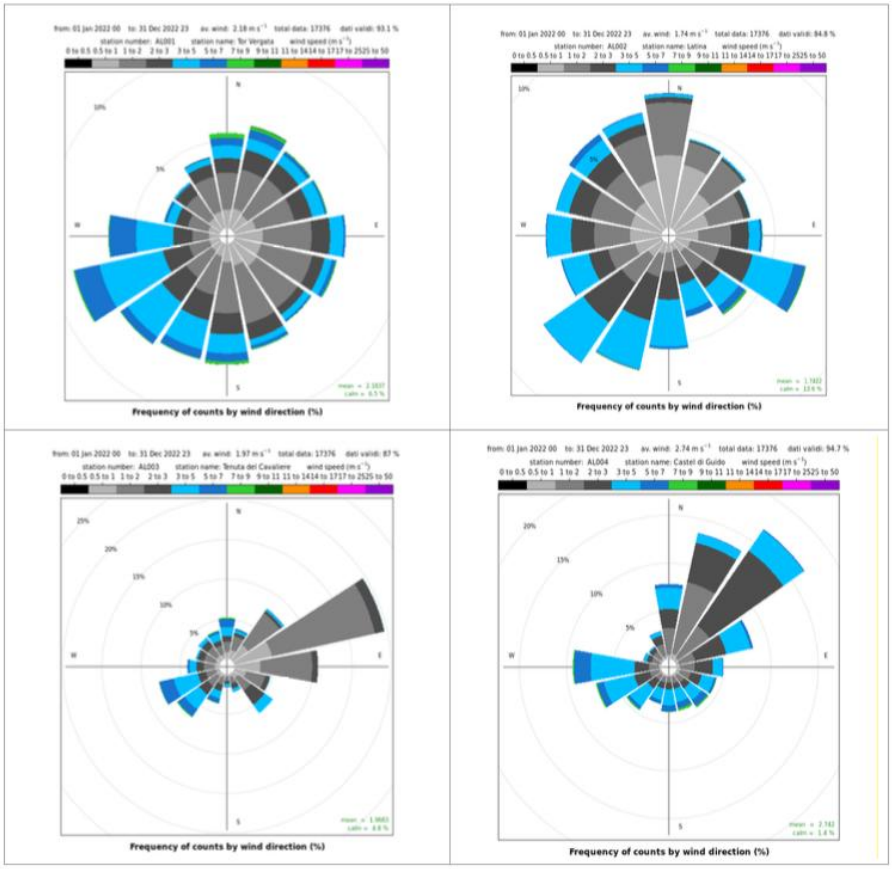

Figure 4-22. Wind roses measured at the ARPA Lazio meteorological stations of (top left) Rome-Tor Vergata, (top right) Latina, (bottom left) Rome-Tenuta del Cavaliere, (bottom right) Rome-Castel di Guido.

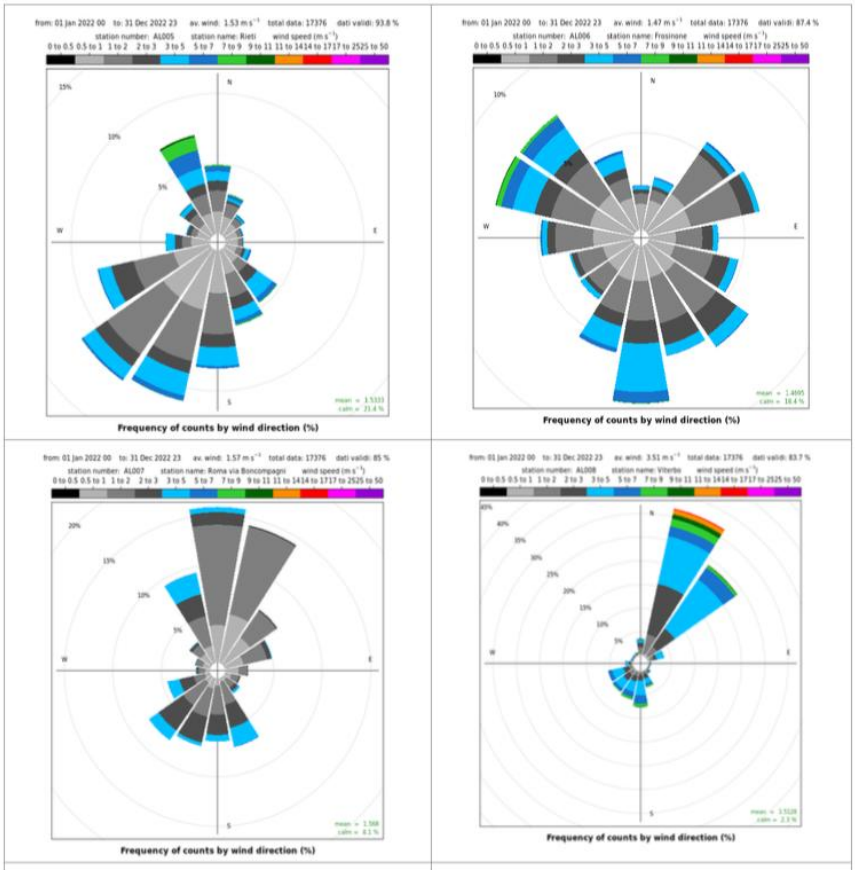

Figure 4-23. Wind roses measured at the ARPA Lazio meteorological stations of (top left) Rieti, (top right) Frosinone, (bottom left) Rome-Via Boncompagni, (bottom right) Viterbo.

The wind roses show that the circulation at the stations of Viterbo, Rome-Tenuta del Cavaliere and Rome-Boncompagni is strongly directional. At Viterbo and Rome-Tenuta del Cavaliere the limits are imposed by the orography, while at Rome-Boncompagni they are imposed by the surrounding urban context. Rome-Tor Vergata and Latina have more distributed roses, with preferential directions from the south and southwest quadrants, while Rieti and Frosinone tend to be more influenced by their geographical position, surrounded by valleys.

The average annual wind speed ranges from 1.47 m/s at Frosinone, protected by the surrounding orography, to 3.51 m/s at Viterbo, where there is probably a channelling effect of the currents. The sea breeze is not very pronounced, but still noticeable in the third quadrants of Rome-Tor Vergata and Rome-Castel di Guido. The katabatic winds descending the Tiber valley, recorded from the Rome-Tenuta del Cavaliere station (I quadrant), are weak but very frequent.

In terms of ventilation, the year 2022 was slightly less windy than 2021 and the average of the previous 10 years (2012-2021). The percentage of wind calm was slightly higher compared to the previous year (2021) and in line with the available climate series (2012-2021).

The year 2022 was overall less rainy than previous years, especially for the first months of the year. The spatial distribution of cumulative precipitation shows maxima over the southern part of the region, between Latina and Frosinone.

The histogram below (Figure 4-25) shows the 2022-2021 comparison of precipitation for the 8 stations of the RMM network (Table 4-4) and for 5 station of an additional regional monitoring network (ARSIAL), selected as reference for each provincial capital (VT-Montefiascone, RI-Rieti, RM-Lanciani, FR-Alatri, LT-Pontinia). Rainfall in 2022 was lower than in the previous year at all stations (Figure 4-25).

ha eliminato: Table 3

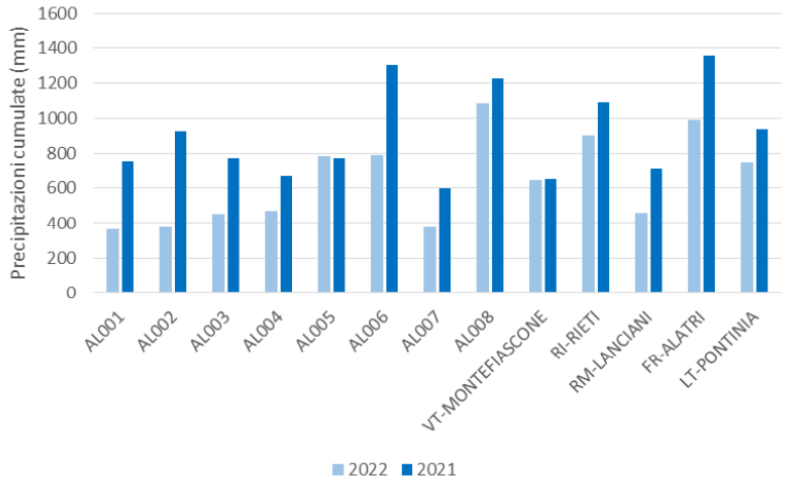

Figure 4-25. Comparison of precipitation: 2022 vs. 2021.

Analysing the data from the ARPA Lazio micrometeorological network and those from the ARSIAL network (Figure 4-25), we see that the year 2022 is warmer overall than the previous one (Figure 4-27), especially for the months of May, June, July and December (Figure 4-28). There is an average temperature increase of 0.82°C in 2022 compared to 2021. (This temperature increase was observed all over Italy: according to the CNR-ISAC results, 2022 was the hottest year in the last 30 years, with an anomaly of +1.15°C).

ha eliminato: Figure 4-26

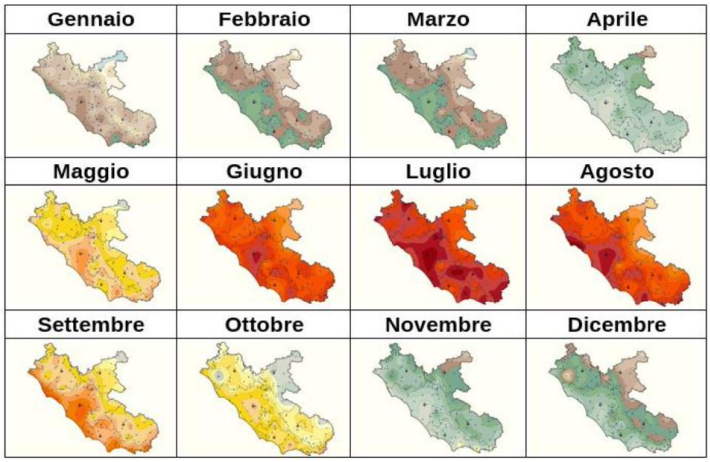

Figure 4-26 ARPA Lazio: monthly temperature maps in 2022.

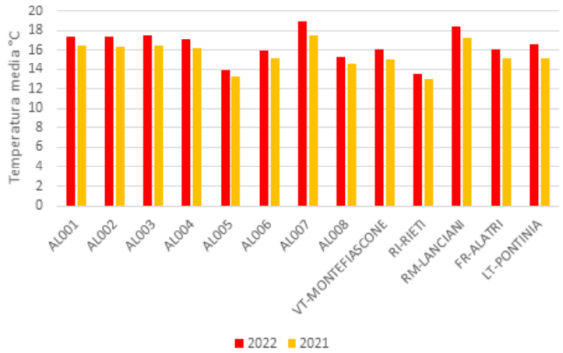

Figure 4-27. Comparison of average temperature: 2022 vs. 2021.

Analysing the monthly trends, the months that showed the greatest increase in temperature compared to the previous year were May, June, July and December. Figure 4-28 compares the mean monthly temperatures between 2022 and 2021 for the station AL001 Rome Tor Vergata, showing a temperature difference of +2.92 °C for May, +2.39 °C for June, +1.99 °C for July and +2.22 °C for December.

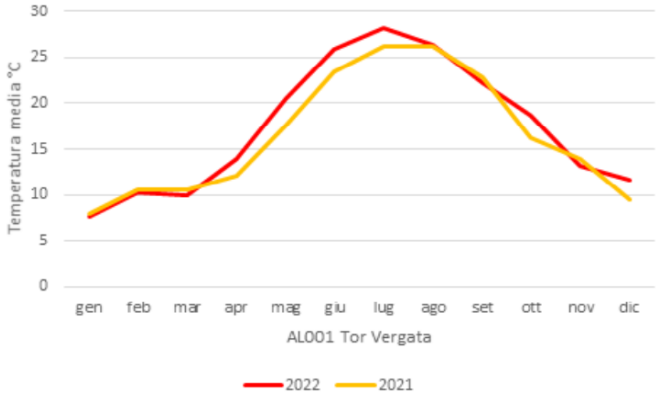

Figure 4-28. Station AL001 Rome Tor Vergata: comparison of the mean temperature trend for the years 2022 and 2021.

5. TECHNICAL REQUIREMENT SPECIFICATIONS

The Technical Requirements Definition Process transforms the stakeholder expectations into a definition of the problem and then into a complete set of validated technical requirements expressed as "shall" statements that can be used for defining a design solution [RD 3].

ha eliminato: [RD-3]

According to [RD-3], the defined requirements should encompass, at a minimum, functional requirements (specifying the functions to be performed), performance requirements (detailing how well these functions are to be executed), and interface requirements (outlining product interaction criteria).

Leveraging the user needs gathered in Section 2, the functional architecture for the APEMAIA project can be preliminarily outlined, as depicted in Figure 5-1.

ha eliminato: Figure 5-1

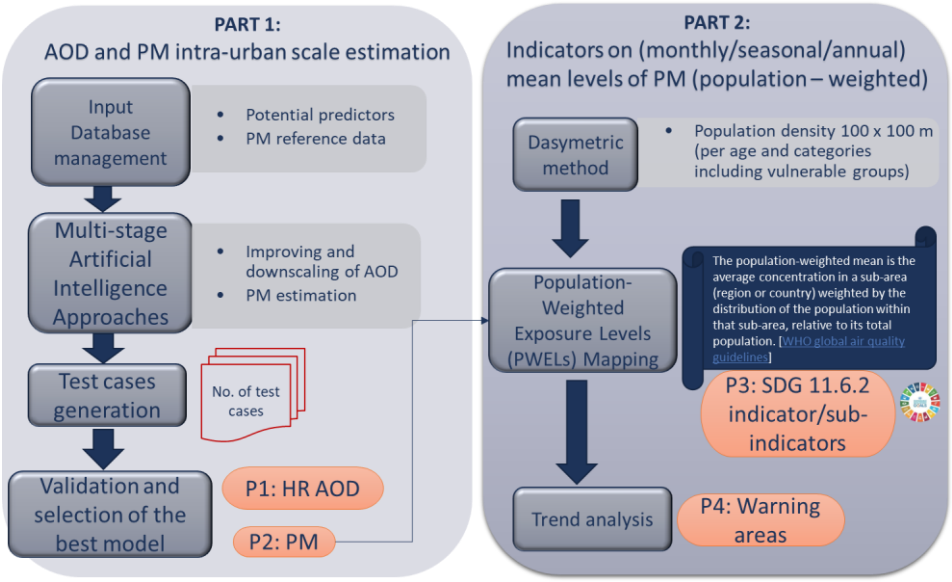

Figure 5-1- Functional architecture of APEMAIA project.

5.1 Requirements Classification

Any requirement shall be defined on the basis of the following rules. In particular, it will be listed in a row of table as presented in Table 5-1.

ha eliminato: Table 5-1

Table 5-1: Requirement classification.

<Requirement ID>	<Requirement Title>	<Method>	<Criticality>
Description			
Dependencies			
Comment			

Where:

- **Requirement Identifier**: univocally identifies the APEMAIA mission requirement throughout the document following the naming convention described hereinafter.
- **Requirement Title**: is a short requirement description.
- **Method**: describes the method to be used for verifying the requirement compliance. The following values can be used:
    - **Analysis testing (A)** - This verification method implies use of analytical techniques (such as system engineering analysis, statistics, mathematical modelling, and simulations) and shall be used to verify such requirements.
    - **Review of Design (D)** - This verification method may be used when approved design, technical descriptions, engineering drawings unambiguously show that the requirement is met.
    - **Inspection (I)** – Verification by inspection is only done when testing is insufficient or inappropriate. This method of verification is for those requirements that are normally performed by some form of visual inspection. This would include workmanship, labelling, environment requirements etc.
    - **Test (T)** – The compliance with the requirement will be directly measured through one or more test procedures or test cases. This method shall be preferred, unless one of the other methods can be established at acceptable level of confidence while being cost and schedule effective.
- **Criticality**: describes the importance of the system compliance with the requirement. The following values are used in this document:
    - **Mandatory (M)** – the requirement implementation is compulsory and is indicated with "shall".
    - **Optional (O)** – the requirement implementation is optional and is indicated with "should".
- **Requirement Description**: describes the requirement.
- **Requirement Dependencies**: contains the references of upper-level requirements. These source requirements are traced through requirement matrices (being provided in separate documents).
- **Requirement Comment**: optional field, which intends to provide additional information, justification, or explanation of the requirement. Comment is for information only and it is not considered part of the requirement.

The APEMAIA mission requirement identifier scheme follows the pattern:

APEMAIA-<CLASS>-XXXX

Where:

- <CLASS> is one of the following requirement's classes:
  - **GEN:** General functional requirements;
  - **DAT:** Data management requirements;
  - **PER**: Performance requirements;
  - **STO**: Storage requirements;
  - **DIS**: Dissemination requirements;
  - **QUA**: Quality requirements;
  - **VER**: Verification requirements.
- XXXX is the progressive number unique for each APEMAIA mission requirement.

5.2 General functional requirements

APEMAIA-GEN-0010	**Study areas**	**I**	**M**
Description	The project shall focus on the following study areas: the Rome metropolitan area (chosen as the Italian Primary Target Area for MAIA) and the urban areas of Taranto and Bari (southern Italy). The latter will serve as the calibration/validation site for the developed models.		
Dependencies			
Comment			

ha eliminato: 0010

APEMAIA-GEN-0020	**Population density maps**	**I**	**M**
Description	The project shall produce an updated density population map at an intra-urban scale for the study areas. This will be achieved by utilizing the CNR-IIA dasymetric method developed in-house or by incorporating ancillary open datasets.		
Dependencies			
Comment			

ha eliminato: 0020

APEMAIA-GEN-0030	**Products of the project: P1 – Improved AOD**	**I**	**M**
Description	The project shall generate Product P1 as one of its outputs, which consists of high-resolution time series of Aerosol Optical Depth (AOD) for the designated study areas (APEMAIA-GEN-0010).		
Dependencies			
Comment			

ha eliminato: 0030

APEMAIA-GEN-0040	**Products of the project: P2a – PM10**	**I**	**M**
Description	The project shall generate Product P2a as one of its outputs, which consists of a time series of PM 10 concentration maps at the intra-urban scale for the designated study areas (APEMAIA-GEN-0010).		
Dependencies	APEMAIA-GEN-0030		
Comment			

APEMAIA-GEN-0050	**Products of the project: P2b – PM2.5**	**I**	**O**
Description	The project shall generate Product P2b as one of its outputs, which consists of a time series of PM 2.5 concentration maps at the intra-urban scale for the designated study areas (APEMAIA-GEN-0010).		
Dependencies	APEMAIA-GEN-0030		
Comment	Optional: based on the availability of PM2.5 in situ reference data.		

APEMAIA-GEN-0060	**Products of the project: P3a – SDG 11.6.2 indicator**	**I**	**M**
Description	The project shall produce Product P3a as one of its outputs, encompassing Population Weighted Exposure Levels to the annual, seasonal, and monthly average PM10 concentrations for the Bari study area.		
Dependencies	APEMAIA-GEN-0020, APEMAIA-GEN-0040		
Comment			

APEMAIA-GEN-0070	**Products of the project: P3b – SDG 11.6.2 indicator**	**I**	**O**
Description	The project shall produce Product P3b as one of its outputs, encompassing Population Weighted Exposure Levels to the annual, seasonal, and monthly average PM 2.5 concentrations for the Bari study area.		
Dependencies	APEMAIA-GEN-0020, APEMAIA-GEN-0050		
Comment			

APEMAIA-GEN-0080	**Products of the project: SDG 11.6.2 sub-indicators**	**I**	**O**
Description	The project will produce sub-indicators for Product P3a,b tailored to address vulnerable categories, such as children or elderly people, in the Bari study area.		
Dependencies	APEMAIA-GEN-0020, APEMAIA-GEN-0030, APEMAIA-GEN-0040		
Comment	Optional: based on the availability of disaggregated population data for Bari study area		

APEMAIA-GEN-0090	**Products of the project: P4 – Warning areas**	**I**	**M**
Description	The project will generate Product P4, involving the delineation of 'warning areas' within the Bari study area. These areas will be identified based on increasing trends or abrupt changes in PM concentration, which may have a medium- to long-term impact on health.		
Dependencies	APEMAIA-GEN-0060, APEMAIA-GEN-0070		
Comment			

APEMAIA-GEN-0100	**Additional output: LCLU maps**	**I**	**M**
Description	As an additional output, the project shall provide an updated land cover and land use (LCLU) map at high spatial resolution (≤ 20m) for the designated study areas, to be used as input of the ML algorithms.		
Dependencies			
Comment			

APEMAIA-GEN-0110	**Additional studies: Urban Heat Island**	**I**	**O**
Description	As an additional output, the project shall provide an exploratory analysis of the relationship between urban heat islands and intra-urban pollution, with a focus on warning areas in Bari study area.		
Dependencies	APEMAIA-GEN-0090		
Comment			

APEMAIA-GEN-0120	Machine Learning algorithms	D	M
Description	The project shall develop and test a minimum of two Machine Learning algorithms for generating P1, P2a,b products, whereas it should consolidate a single algorithm for updating LCLU maps in urban areas.		
Dependencies			
Comment			

ha eliminato: 0120

APEMAIA-GEN-0130	ML custom implementations for different scenarios	D	M
Description	The APEMAIA project shall implement custom tuning of ML algorithms tailored for two distinct operational scenarios: 1) marine aerosol and 2) Saharan dust.		
Dependencies	APEMAIA-GEN-0120		
Comment			

ha eliminato: 0130

APEMAIA-GEN-0140	Lifetime of the project	I	M
Description	The project shall have a duration of two years.		
Dependencies			
Comment			

APEMAIA-GEN-0150	Scientific Readiness Level Advancement	I	M
Description	The project will produce an advancement of Scientific Readiness Level (SRL) from 3 to 6.		
Dependencies			
Comment			

5.3 Data management requirements

APEMAIA-DAT-0160	Dataset of pre-processed EO data	I	M
Description	The project shall create a dataset of pre-processed Earth Observation (EO) data on the study areas to be used as input for the ML algorithms.		
Dependencies			
Comment			

ha eliminato: 0160

APEMAIA-DAT-0170	Dataset of 3-dimentional meteorological variables	I	M
Description	The project shall create a dataset of pre-processed 3-dimensional meteorological variables on the study areas provided by ARPA Puglia and Lazio and re-analysis models, to be used as input for the ML algorithms.		
Dependencies			
Comment			

ha eliminato: 0170

APEMAIA-DAT-0180	Imagery data format specifications (raster data)	I	M
Description	The software environments/equipment utilized for the project must be capable of ingesting, managing, and processing raster data in NetCDF, GRIB, HDF, SAFE, JPEG2000, and GeoTIFF formats.		
Dependencies			
Comment			

ha eliminato: 0180

APEMAIA-DAT-0190	Data Format for MAIA Simulated Data	I	M
Description	In accordance with MAIA's core data products format, APEMAIA project shall be able to ingest input data in NetCDF format, with metadata conforming to ISO-3166-1 & ISO-8601 and ISO Metadata Evolution for NASA Data Systems (MENDS) standards.		
Dependencies			
Comment			

ha eliminato: 0190

APEMAIA-DAT-0200	Database of geographically ancillary data	I	M
Description	A database of ancillary data for the study areas, including socio-economic variables (such as population density), urban morphology (like buildings footprint/volume), urban zoning (percentage of residential, commercial/industrial, rural areas), and information on the number of nodes and street segments in the road network, shall be created to be used as input for ML algorithms.		
Dependencies			
Comment			

ha eliminato: 0200

APEMAIA-DAT-0210	Geographically ancillary data format specification (vector data)	I	M
Description	The software environments/equipment utilized for the project must be capable of ingesting, managing, and organize database of ancillary data with a spatial dimension in Shapefile, Geopackage (.gpkg) or Comma Separated Value (.csv) formats.		
Dependencies			
Comment			

ha eliminato: 0210

APEMAIA-DAT-0220	Common workspace for datasets	I	M
Description	All the datasets of the project will be shared among the APEMAIA partnership on a shared cloud workspace.		
Dependencies			
Comment			

ha eliminato: 0220

APEMAIA-DAT-0230	Metadata standardization	I	M
Description	All the datasets produced as the output of the project (i.e. P1, P2a, P2b, P3a, P3b, P4) shall be accompanied by metadata following ISO standards (ISO 19115/19139) in accordance with OGC standard formats (WMS, WFS).		
Dependencies			
Comment			

ha eliminato: 0230

APEMAIA-DAT-0240	FAIR principles	I	M
Description	All data generated throughout the project shall be managed according to FAIR (findability, accessibility, interoperability, and reusability) principles.		
Dependencies			
Comment			

ha eliminato: 0240

5.4 Performance requirements

5.4.1 Spatial resolution

APEMAIA-PER-0250	Spatial resolution of improved AOD maps (P1)	I	M
Description	The project shall provide a temporal series of AOD maps for the area of studies at a spatial resolution lower than 500 m.		
Dependencies	APEMAIA-GEN-0030		
Comment			

ha eliminato: 0250

APEMAIA-PER-0260	Spatial resolution of PM maps(P2a-b)	I	M
Description	The project shall provide a temporal series of PM maps for the area of studies at a spatial resolution lower than 500 m.		
Dependencies	APEMAIA-GEN-0040, APEMAIA-GEN-0050		
Comment			

ha eliminato: 0260

APEMAIA-PER-0270	Spatial resolution of Exposure maps (P3a-b)	I	O
Description	The project shall provide a temporal series of population weighted exposure maps to PM; for the area of studies at a spatial resolution lower than 500 m.		
Dependencies	APEMAIA-GEN-0060, APEMAIA-GEN-0070		
Comment			

5.4.2 Temporal references

APEMAIA-PER-0280	Temporal reference window	I	M
Description	The project shall consider a temporal reference window from 2019 to the present for collecting data, creating the input datasets and producing the output products.		
Dependencies			
Comment	This temporal slot has been selected based on the availability of satellite and ground reference data.		

ha eliminato: 0280

APEMAIA-PER-0290	Time intervals of input datasets	I	M
Description	The APEMAIA project shall incorporate input datasets spanning a minimum time interval of 3 years for the development and testing of algorithms focusing on Bari city. For datasets related to the other sites, Taranto and Rome, a minimum temporal interval of 3 months should be taken into account.		
Dependencies			
Comment			

ha eliminato: 0290

5.4.3 Frequency

APEMAIA-PER-0300	Frequency of P1 products	I	M
Description	The project shall generate P1 product maps for the specified temporal reference period (APEMAIA-PER-0270, APEMAIA-PER-0280) with a daily frequency.		
Dependencies			
Comment	Condition: if satellite input data for AOD estimation is available and meets high-quality standards (e.g., absence of total cloud cover or absence of payload acquisition errors).		

ha eliminato: 0300

APEMAIA-PER-0310	Frequency of P2a,b products	I	M
Description	The project shall produce daily P2a,b products maps for the temporal reference period (APEMAIA-PER-0270, APEMAIA-PER-0280), when P1 products are available as input.		
Dependencies	APEMAIA-GEN-0040, APEMAIA-GEN-0050		
Comment			

ha eliminato: 0310

APEMAIA-PER-0320	Frequency of P3a,b products	I	M
Description	The project shall provide a monthly/seasonal/annual indicator on mean levels of PM concentration for the temporal reference period (APEMAIA-PER-0270, APEMAIA-PER-0280), weighted with respect to the urban population in Bari.		
Dependencies	APEMAIA-GEN-0060, APEMAIA-GEN-0070		
Comment			

APEMAIA-PER-0330	Frequency of P4 products: warning areas	I	M
Description	The project will release a one-time output map for warning areas in Bari, valid for the entire reference period (APEMAIA-PER-0270, APEMAIA-PER-0280).		
Dependencies	APEMAIA-GEN-090, APEMAIA-PER-320		
Comment			

5.5 Storage requirements

APEMAIA-STO-0340	Input data volume	I	M
Description	The storage capability for ingesting input data, needed for the algorithm developing and testing, and storing the output products shall be not less than 20 TB.		
Dependencies			
Comment			

ha eliminato: 0340

5.6 Dissemination requirements

APEMAIA-DIS-0350	WebGIS	I	M
Description	All products generated throughout the project will be accessible through a dedicated WebGIS portal with visualization and query capabilities.		
Dependencies			
Comment			

ha eliminato: 0350

APEMAIA-DIS-0360	Open-source codes	I	M
Description	For the benefit of the scientific community, the software modules for data management and the ML algorithms should be based on open-source software and programming languages.		
Dependencies	APEMAIA-GEN-0120		
Comment			

5.7 Verification requirements

APEMAIA-VER-0370	Reference data for AOD	I	M
Description	The APEMAIA project shall use in situ AOD measurements obtained from at least 4 stations within the AERONET network, distributed across the central-southern Italy territory, for the accuracy assessment of satellite AOD data.		
Dependencies			
Comment			

ha eliminato: 0370

APEMAIA-VER-0380	Reference data for PM	I	M
Description	The APEMAIA project shall leverage ground validated PM measurements obtained from air quality monitoring stations distributed across the study areas, specifically provided by ARPA Puglia and Lazio, for the validation of P2 products.		
Dependencies			
Comment			

ha eliminato: 0380

APEMAIA-VER-0390	Accuracy for LCLU mapping	T	M
Description	The land cover maps generated for the study areas must have an overall accuracy not less than 85%.		
Dependencies			
Comment			

ha eliminato: 0390

APEMAIA-VER-0400	Accuracy for AOD data	T	M
Description	The project shall assess the uncertainties of satellite AOD data by demonstrating a correlation with in-situ measurements, achieving an R-squared value greater than 0.60.		
Dependencies	APEMAIA-VER-0370		
Comment			

ha eliminato: 0400

APEMAIA-VER-0410	Accuracy for PM mapping	T	M
Description	The project shall assess the uncertainties of P2a,b products derived by ML algorithms by demonstrating a correlation with in-situ data from air quality monitoring stations, achieving an R-squared value greater than 0.60.		
Dependencies	APEMAIA-VER-0380		
Comment			

ha eliminato: 0410

APEMAIA-VER-0420	Validation tools	T	O
Description	The Delta Tool statistical software shall be employed by ARPA Puglia in support to the uncertainties related to PM products. This tool relies on the statistical comparison between modelled data and measurements from air quality monitoring stations, providing a quick assessment of the model's performance in terms of various indicators and statistical diagrams.		
Dependencies	APEMAIA-VER-0380, APEMAIA-VER-0410		
Comment			

ha eliminato: 0420

6. TRACEABILITY MATRICES USER NEEDS VS USER REQUIREMENTS

6.1 User Needs vs. User Requirements

Table 6-1: User Needs vs. User requirements Traceability matrix

User Needs ID	User Req. ID
US1_US1N0100,	APEMAIA-GEN-0020, APEMAIA-GEN-0030, APEMAIA-GEN-0040, APEMAIA-GEN-0050, APEMAIA-GEN-0060, APEMAIA-GEN-0070, APEMAIA-GEN-0080, APEMAIA-GEN-0140, APEMAIA-GEN-0150
US1_N0200,	APEMAIA-GEN-0010, APEMAIA-VER-0380
US1_N0300	APEMAIA-DAT-0190
US1_N0400	APEMAIA-GEN-0120, APEMAIA-DIS-0360
US12_N0700,	APEMAIA-GEN-0030, APEMAIA-PER-0250 APEMAIA-VER-0370, APEMAIA-VER-0400
US1_N0600	APEMAIA-GEN-0080, APEMAIA-GEN-0090
US12_N0700,	APEMAIA-GEN-0100, APEMAIA-DAT-0160, APEMAIA-DAT-0170, APEMAIA-DAT-0200, APEMAIA-PER-0250, APEMAIA-PER-0260, APEMAIA-PER-0270, APEMAIA-PER-0300, APEMAIA-PER-0310
US2_N0800	APEMAIA-PER-0260, APEMAIA-GEN-0120
US2_N0900	APEMAIA-GEN-0020, APEMAIA-GEN-0040, APEMAIA-GEN-0050, APEMAIA-GEN-0060, APEMAIA-GEN-0070, APEMAIA-GEN-0080
US2_N1000	APEMAIA-VER-0410, APEMAIA-VER-0420
US2_N1000,	APEMAIA-GEN-0120, APEMAIA-GEN-0130
US2_N1000,	APEMAIA-GEN-0100, APEMAIA-GEN-0120, APEMAIA-GEN-0130, APEMAIA-DAT-0160, APEMAIA-DAT-0170, APEMAIA-DAT-0200, APEMAIA-PER-0250, APEMAIA-PER-0260, APEMAIA-PER-0300, APEMAIA-PER-0310
US2_N1100,	APEMAIA-GEN-0060, APEMAIA-GEN-0070, APEMAIA-GEN-0080, APEMAIA-GEN-0090, APEMAIA-PER-0320, APEMAIA-PER-0330
US2_N1200,	APEMAIA-DAT-0180, APEMAIA-DAT-0190, APEMAIA-DAT-0210, APEMAIA-DAT-0220, APEMAIA-DAT-0230, APEMAIA-DAT-0240, APEMAIA-PER-0280, APEMAIA-PER-0290, APEMAIA-STO-0340, APEMAIA-DIS-0350, APEMAIA-DIS-0360
~~Errore. L'origine riferimento non è stata trovata.~~	~~APEMAIA-DAT-0160, APEMAIA-GEN-0110, APEMAIA-VER-0390~~

ha eliminato: US1_N0100

ha eliminato: US1_N0200

ha eliminato: US1_N0500US1_N0500

ha eliminato: US1_N0500

ha eliminato: US2_N0700

ha eliminato: US2_N1100

ha eliminato: US2_N1200

ha eliminato: US3_N1300

ha eliminato: US3_N1400

ha eliminato: ~~US3_N1500~~

6.2 User requirements vs. User Needs

Table 6-2: User requirements vs. User Needs Traceability matrix

User Req. ID	User Needs ID
APEMAIA-GEN-0010	US1_N0200,
APEMAIA-GEN-0020	US1_US1N0100, US2_N0900
APEMAIA-GEN-0030	US1_US1N0100, US12_N0700,
APEMAIA-GEN-0040	US1_US1N0100, US2_N0900
APEMAIA-GEN-0050	US1_US1N0100, US2_N0900
APEMAIA-GEN-0060	US1_US1N0100, US2_N0900, US2_N1100,
APEMAIA-GEN-0070	US1_US1N0100, US2_N0900, US2_N1100,
APEMAIA-GEN-0080	US1_US1N0100, US1_N0600, US2_N0900, US2_N1100,
APEMAIA-GEN-0090	US1_N0600, US2_N1100,
APEMAIA-GEN-0100	US1_N0600, US2_N1100,
APEMAIA-GEN-0110	Errore. L'origine riferimento non è stata trovata.
APEMAIA-GEN-0120	US1_N0400, US2_N0800, US2_N1000, US2_N1000,
APEMAIA-GEN-0130	US2_N1000, US2_N1000,
APEMAIA-GEN-0140	US1_US1N0100,
APEMAIA-GEN-0150	US1_US1N0100,
APEMAIA-DAT-0160	US1_N0600, US2_N1000, Errore. L'origine riferimento non è stata trovata.
APEMAIA-DAT-0170	US1_N0600, US2_N1000,
APEMAIA-DAT-0180	US2_N1200,
APEMAIA-DAT-0190	US1_N0300, US2_N1200,
APEMAIA-DAT-0200	US1_N0600, US2_N1000,
APEMAIA-DAT-0210	US3_N1400
APEMAIA-DAT-0220	US3_N1400
APEMAIA-DAT-0230	US2_N1200
APEMAIA-DAT-0240	US2_N1200
APEMAIA-PER-0250	US12_N0700, US12_N0700, US2_N1000,
APEMAIA-PER-0260	US1_N0600, US2_N0800, US2_N1000,
APEMAIA-PER-0260	US1_N0600
APEMAIA-PER-0280	US2_N1200
APEMAIA-PER-0290	US2_N1200
APEMAIA-PER-0300	US1_N0600, US2_N1000,
APEMAIA-PER-0310	US1_N0600, US2_N1000,
APEMAIA-PER-0320	US2_N1100,

ha eliminato: US1_N0200
ha eliminato: US1_N0100
ha eliminato: US1_N0100
ha eliminato: US1_N0500US1_N0500
ha eliminato: US1_N0500
ha eliminato: US1_N0100
ha eliminato: US1_N0100
ha eliminato: US1_N0100
ha eliminato: US3_N1300
ha eliminato: US1_N0100
ha eliminato: US3_N1300
ha eliminato: US1_N0100
ha eliminato: US3_N1300
ha eliminato: US3_N1300
ha eliminato: US2_N0700
ha eliminato: US3_N1200
ha eliminato: US3_N1500
ha eliminato: US2_N1100
ha eliminato: US2_N1200
ha eliminato: US2_N1100
ha eliminato: US2_N1200
ha eliminato: US1_N0100
ha eliminato: US1_N0100
ha eliminato: US2_N0700
ha eliminato: US2_N1200
ha eliminato: US3_N1500
ha eliminato: US2_N0700
ha eliminato: US2_N1200
ha eliminato: US3_N1400
ha eliminato: US3_N1400
ha eliminato: US2_N0700
ha eliminato: US2_N1200
ha eliminato: US1_N0500US1_N0500
ha eliminato: US1_N0500
ha eliminato: US2_N1200
ha eliminato: US2_N0700
ha eliminato: US2_N1200
ha eliminato: US2_N0700
ha eliminato: US2_N1200
ha eliminato: US2_N0700
ha eliminato: US2_N1200
ha eliminato: US3_N1300

APEMAIA-PER-0330	US2_N1100,
APEMAIA-STO-0340	US2_N1200
APEMAIA-DIS-0350	US2_N1200
APEMAIA-DIS-0360	**US1_N0400**, US2_N1200
APEMAIA-VER-0370	US12_N0700,
APEMAIA-VER-0380	US1_N0200,
APEMAIA-VER-0390	Errore. L'origine riferimento non è stata trovata.,
APEMAIA-VER-0400	US12_N0700,
APEMAIA-VER-0410	**US2_N1000**
APEMAIA-VER-0420	**US2_N1000**

ha eliminato: US3_N1300

ha eliminato: US1_N0500US1_N0500

ha eliminato: US1_N0500

ha eliminato: US1_N0200

ha eliminato: US3_N1500

ha eliminato: US1_N0500US1_N0500

ha eliminato: US1_N0500

7. REFERENCES

Alvarez-Mendoza, C.I., Teodoro, A.C., Torres, N., Vivanco, V., 2019. Assessment of Remote Sensing Data to Model PM10 Estimation in Cities with a Low Number of Air Quality Stations: A Case of Study in Quito, Ecuador. Environments 6, 85. https://doi.org/10.3390/environments6070085

AQD, 2008. 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.

Aquilino, M., Adamo, M., Blonda, P., Barbanente, A., Tarantino, C., 2021. Improvement of a Dasymetric Method for Implementing Sustainable Development Goal 11 Indicators at an Intra-Urban Scale. Remote Sens. 13, 2835. https://doi.org/10.3390/rs13142835

Arowosegbe, O.O., Röösli, M., Künzli, N., Saucy, A., Adebayo-Ojo, T.C., Schwartz, J., Kebalepile, M., Jeebhay, M.F., Dalvie, M.A., De Hoogh, K., 2022. Ensemble averaging using remote sensing data to model spatiotemporal PM10 concentrations in sparsely monitored South Africa. Environ. Pollut. 310, 119883. https://doi.org/10.1016/j.envpol.2022.119883

Azid, A., Juahir, H., Toriman, M.E., Kamarudin, M.K.A., Saudi, A.S.M., Hasnam, C.N.C., Aziz, N.A.A., Azaman, F., Latif, M.T., Zainuddin, S.F.M., Osman, M.R., Yamin, M., 2014. Prediction of the Level of Air Pollution Using Principal Component Analysis and Artificial Neural Network Techniques: a Case Study in Malaysia. Water. Air. Soil Pollut. 225, 2063. https://doi.org/10.1007/s11270-014-2063-1

Bagheri, H., 2022. A machine learning-based framework for high resolution mapping of PM2.5 in Tehran, Iran, using MAIAC AOD data. Adv. Space Res. 69, 3333–3349. https://doi.org/10.1016/j.asr.2022.02.032

Bai, K., Li, K., Sun, Y., Wu, L., Zhang, Y., Chang, N.-B., Li, Z., 2023. Global synthesis of two decades of research on improving PM2.5 estimation models from remote sensing and data science perspectives. Earth-Sci. Rev. 241, 104461. https://doi.org/10.1016/j.earscirev.2023.104461

Bakaeva, N., Le, M.T., 2022. Determination of urban pollution islands by using remote sensing technology in Moscow, Russia. Ecol. Inform. 67, 101493. https://doi.org/10.1016/j.ecoinf.2021.101493

Chen, C.-C., Wang, Y.-R., Yeh, H.-Y., Lin, T.-H., Huang, C.-S., Wu, C.-F., 2021. Estimating monthly PM2.5 concentrations from satellite remote sensing data, meteorological variables, and land use data using ensemble statistical modeling and a random forest approach. Environ. Pollut. 291, 118159. https://doi.org/10.1016/j.envpol.2021.118159

Chen, W., Ran, H., Cao, X., Wang, J., Teng, D., Chen, J., Zheng, X., 2020. Estimating PM2.5 with high-resolution 1-km AOD data and an improved machine learning model over Shenzhen, China. Sci. Total Environ. 746, 141093. https://doi.org/10.1016/j.scitotenv.2020.141093

Ding, Y., Chen, Z., Lu, W., Wang, X., 2021. A CatBoost approach with wavelet decomposition to improve satellite-derived high-resolution PM2.5 estimates in Beijing-Tianjin-Hebei. Atmos. Environ. 249, 118212. https://doi.org/10.1016/j.atmosenv.2021.118212

DLGS 155, 2010. DECRETO LEGISLATIVO 13 agosto 2010, n. 155 Attuazione della direttiva 2008/50/CE relativa alla qualita' dell'aria ambiente e per un'aria piu' pulita in Europa.

e-SHAPE project, 2019.

European Commission, 2011. Establishing guidelines for demonstration and subtraction of exceedances attributable to natural sources under the Directive 2008/50/EC on ambient air quality and cleaner air for Europe.

Geng, G., Meng, X., He, K., Liu, Y., 2020. Random forest models for PM $_{2.5}$ speciation concentrations using MISR fractional AODs. Environ. Res. Lett. 15, 034056. https://doi.org/10.1088/1748-9326/ab76df

Gitahi, J., Hahn, M., 2020. High-resolution urban air quality monitoring using sentinel satellite images and low-cost ground-based sensor networks. E3S Web Conf. 171, 02002. https://doi.org/10.1051/e3sconf/202017102002

Gu, J., Wang, Y., Ma, J., Lu, Y., Wang, S., Li, X., 2022. An Estimation Method for PM2.5 Based on Aerosol Optical Depth Obtained from Remote Sensing Image Processing and Meteorological Factors. Remote Sens. 14, 1617. https://doi.org/10.3390/rs14071617

Gutiérrez-Avila, I., Arfer, K.B., Carrión, D., Rush, J., Kloog, I., Naeger, A.R., Grutter, M., Páramo-Figueroa, V.H., Riojas-Rodríguez, H., Just, A.C., 2022. Prediction of daily mean and one-hour maximum PM2.5 concentrations and applications in Central Mexico using satellite-based machine-learning models. J. Expo. Sci. Environ. Epidemiol. 32, 917–925. https://doi.org/10.1038/s41370-022-00471-4

Handschuh, J., Erbertseder, T., Baier, F., 2023. Systematic Evaluation of Four Satellite AOD Datasets for Estimating PM2.5 Using a Random Forest Approach. Remote Sens. 15, 2064. https://doi.org/10.3390/rs15082064

Hau, L., Chen, Yunping, Ma, C., Yang, Y., Chen, Yan, Sun, Y., Gu, X., 2020. A High-Spatial-Resolution Aerosol Retrieval Algorithm for Sentinel-2 Images Over Bright Urban Surfaces, in: IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium. Presented at the IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium, IEEE, Waikoloa, HI, USA, pp. 5592–5595. https://doi.org/10.1109/IGARSS39084.2020.9323605

He, Q., Huang, B., 2018. Satellite-based high-resolution PM2.5 estimation over the Beijing-Tianjin-Hebei region of China using an improved geographically and temporally weighted regression model. Environ. Pollut. 236, 1027–1037. https://doi.org/10.1016/j.envpol.2018.01.053

Holloway, T., Miller, D., Anenberg, S., Diao, M., Duncan, B., Fiore, A.M., Henze, D.K., Hess, J., Kinney, P.L., Liu, Y., Neu, J.L., O'Neill, S.M., Odman, M.T., Pierce, R.B., Russell, A.G., Tong, D., West, J.J., Zondlo, M.A., 2021. Satellite Monitoring for Air Quality and Health. Annu. Rev. Biomed. Data Sci. 4, 417–447. https://doi.org/10.1146/annurev-biodatasci-110920-093120

Ibrahim, S., Landa, M., Pešek, O., Brodský, L., Halounová, L., 2022. Machine Learning-Based Approach Using Open Data to Estimate PM2.5 over Europe. Remote Sens. 14, 3392. https://doi.org/10.3390/rs14143392

Kapadia, D., Jariwala, N., 2022. Prediction of tropospheric ozone using artificial neural network (ANN) and feature selection techniques. Model. Earth Syst. Environ. 8, 2183–2192. https://doi.org/10.1007/s40808-021-01220-6

Karimian, H., Li, Y., Chen, Y., Wang, Z., 2023. Evaluation of different machine learning approaches and aerosol optical depth in PM2.5 prediction. Environ. Res. 216, 114465. https://doi.org/10.1016/j.envres.2022.114465

Kibirige, G.W., Yang, M.-C., Liu, C.-L., Chen, M.C., 2023. Using satellite data on remote transportation of air pollutants for PM2.5 prediction in northern Taiwan. PLOS ONE 18, e0282471. https://doi.org/10.1371/journal.pone.0282471

Li, J., Garshick, E., Hart, J.E., Li, L., Shi, L., Al-Hemoud, A., Huang, S., Koutrakis, P., 2021. Estimation of ambient PM2.5 in Iraq and Kuwait from 2001 to 2018 using machine learning and remote sensing. Environ. Int. 151, 106445. https://doi.org/10.1016/j.envint.2021.106445

Li, T., Shen, H., Yuan, Q., Zhang, L., 2022. A Locally Weighted Neural Network Constrained by Global Training for Remote Sensing Estimation of $PM_{2.5}$. IEEE Trans. Geosci. Remote Sens. 60, 1–13. https://doi.org/10.1109/TGRS.2021.3074569

Li, X., Zhang, X., 2019. Predicting ground-level PM2.5 concentrations in the Beijing-Tianjin-Hebei region: A hybrid remote sensing and machine learning approach. Environ. Pollut. 249, 735–749. https://doi.org/10.1016/j.envpol.2019.03.068

Liu, J., Gao, X., Ruan, Z., Yuan, Y., Dong, S., 2022. Analysis of spatial and temporal distribution and influencing factors of fine particles in Heilongjiang Province. Urban Clim. 41, 101070. https://doi.org/10.1016/j.uclim.2021.101070

Liu, Ming, 2020. Satellite-based PM2.5 Exposure Estimation and Health Impacts over China (PhD Thesis). UWSpace.

Liu, Y., Li, C., Liu, D., Tang, Y., Seyler, B.C., Zhou, Z., Hu, X., Yang, F., Zhan, Y., 2022. Deriving hourly full-coverage PM2.5 concentrations across China's Sichuan Basin by fusing multisource satellite retrievals: A machine-learning approach. Atmos. Environ. 271, 118930. https://doi.org/10.1016/j.atmosenv.2021.118930

Ma, Z., Hu, X., Huang, L., Bi, J., Liu, Y., 2014. Estimating Ground-Level PM $_{2.5}$ in China Using Satellite Remote Sensing. Environ. Sci. Technol. 48, 7436–7444. https://doi.org/10.1021/es5009399

Miao, L., Tang, S., Ren, Y., Kwan, M.-P., Zhang, K., 2022. Estimation of daily ground-level PM2.5 concentrations over the Pearl River Delta using 1 km resolution MODIS AOD based on multi-feature BiLSTM. Atmos. Environ. 290, 119362. https://doi.org/10.1016/j.atmosenv.2022.119362

Murray, N.L., Holmes, H.A., Liu, Y., Chang, H.H., 2019. A Bayesian ensemble approach to combine PM2.5 estimates from statistical models using satellite imagery and numerical model simulation. Environ. Res. 178, 108601. https://doi.org/10.1016/j.envres.2019.108601

Nabavi, S.O., Haimberger, L., Abbasi, E., 2019. Assessing PM2.5 concentrations in Tehran, Iran, from space using MAIAC, deep blue, and dark target AOD and machine learning algorithms. Atmospheric Pollut. Res. 10, 889–903. https://doi.org/10.1016/j.apr.2018.12.017

Pan, B., 2018. Application of XGBoost algorithm in hourly PM2.5 concentration prediction. IOP Conf. Ser. Earth Environ. Sci. 113, 012127. https://doi.org/10.1088/1755-1315/113/1/012127

Pedde, M., Kloog, I., Szpiro, A., Dorman, M., Larson, T.V., Adar, S.D., 2022. Estimating long-term PM10-2.5 concentrations in six US cities using satellite-based aerosol optical depth data. Atmos. Environ. 272, 118945. https://doi.org/10.1016/j.atmosenv.2022.118945

Rahman, M.M., 2023. Recommendations on the measurement techniques of atmospheric pollutants from in situ and satellite observations: a review. Arab. J. Geosci. 16, 326. https://doi.org/10.1007/s12517-023-11410-4

Ranjan, A.K., Patra, A.K., Gorai, A.K., 2021. A Review on Estimation of Particulate Matter from Satellite-Based Aerosol Optical Depth: Data, Methods, and Challenges. Asia-Pac. J. Atmospheric Sci. 57, 679–699. https://doi.org/10.1007/s13143-020-00215-0

Shi, Y., Lau, A.K.-H., Ng, E., Ho, H.-C., Bilal, M., 2021. A Multiscale Land Use Regression Approach for Estimating Intraurban Spatial Variability of PM2.5 Concentration by Integrating Multisource Datasets. Int. J. Environ. Res. Public. Health 19, 321. https://doi.org/10.3390/ijerph19010321

Shin, M., Kang, Y., Park, S., Im, J., Yoo, C., Quackenbush, L.J., 2020. Estimating ground-level particulate matter concentrations using satellite-based data: a review. GIScience Remote Sens. 57, 174–189. https://doi.org/10.1080/15481603.2019.1703288

SMURBS project, 2017.

Stafoggia, M., Bellander, T., Bucci, S., Davoli, M., De Hoogh, K., De' Donato, F., Gariazzo, C., Lyapustin, A., Michelozzi, P., Renzi, M., Scortichini, M., Shtein, A., Viegi, G., Kloog, I., Schwartz, J., 2019. Estimation of daily PM10 and PM2.5 concentrations in Italy, 2013–2015, using a spatiotemporal land-use random-forest model. Environ. Int. 124, 170–179. https://doi.org/10.1016/j.envint.2019.01.016

Stirnberg, R., Cermak, J., Fuchs, J., Andersen, H., 2020. Mapping and Understanding Patterns of Air Quality Using Satellite Data and Machine Learning. J. Geophys. Res. Atmospheres 125, e2019JD031380. https://doi.org/10.1029/2019JD031380

Wang, M., Franklin, M., Li, L., 2022. Generating Fine-Scale Aerosol Data through Downscaling with an Artificial Neural Network Enhanced with Transfer Learning. Atmosphere 13, 255. https://doi.org/10.3390/atmos13020255

Xu, X., Zhang, C., Liang, Y., 2021. Review of satellite-driven statistical models PM 2.5 concentration estimation with comprehensive information. Atmos. Environ. 256, 118302. https://doi.org/10.1016/j.atmosenv.2021.118302

Xue, T., Zheng, Y., Tong, D., Zheng, B., Li, X., Zhu, T., Zhang, Q., 2019. Spatiotemporal continuous estimates of PM2.5 concentrations in China, 2000–2016: A machine learning method with inputs from satellites, chemical transport model, and ground observations. Environ. Int. 123, 345–357. https://doi.org/10.1016/j.envint.2018.11.075

Xue, Y., Li, Y., Guang, J., Tugui, A., She, L., Qin, K., Fan, C., Che, Y., Xie, Y., Wen, Y., Wang, Z., 2020. Hourly PM2.5 Estimation over Central and Eastern China Based on Himawari-8 Data. Remote Sens. 12, 855. https://doi.org/10.3390/rs12050855

Yang, Y., Chen, Yunping, Yang, K., Cermak, J., Chen, Yan, 2021. High-resolution aerosol retrieval over urban areas using sentinel-2 data. Atmospheric Res. 264, 105829. https://doi.org/10.1016/j.atmosres.2021.105829

Yu, W., Li, S., Ye, T., Xu, R., Song, J., Guo, Y., 2022. Deep Ensemble Machine Learning Framework for the Estimation of PM2.5 Concentrations. Environ. Health Perspect. 130, 037004. https://doi.org/10.1289/EHP9752

Yu, W., Liu, Y., Ma, Z., Bi, J., 2017. Improving satellite-based PM2.5 estimates in China using Gaussian processes modeling in a Bayesian hierarchical setting. Sci. Rep. 7, 7048. https://doi.org/10.1038/s41598-017-07478-0

Zang, Z., Li, D., Guo, Y., Shi, W., Yan, X., 2021. Superior PM2.5 Estimation by Integrating Aerosol Fine Mode Data from the Himawari-8 Satellite in Deep and Classical Machine Learning Models. Remote Sens. 13, 2779. https://doi.org/10.3390/rs13142779

Zhang, T., Shen, H., Xia, X., Wang, L., Mao, F., Yuan, Q., Gu, Y., Zhao, B., Zhu, Z., Lin, Y., Bo, Y., Gong, W., 2023. Himawari-8 High Temporal Resolution AOD Products Recovery: Nested Bayesian Maximum Entropy Fusion Blending GEO With SSO Satellite Observations. IEEE Trans. Geosci. Remote Sens. 61, 1–15. https://doi.org/10.1109/TGRS.2023.3262785

Zhang, Ying, Li, Z., Bai, K., Wei, Y., Xie, Y., Zhang, Yuanxun, Ou, Y., Cohen, J., Zhang, Yuhuan, Peng, Z., Zhang, X., Chen, C., Hong, J., Xu, H., Guang, J., Lv, Y., Li, K., Li, D., 2021. Satellite remote sensing of atmospheric particulate matter mass concentration: Advances, challenges, and perspectives. Fundam. Res. 1, 240–258. https://doi.org/10.1016/j.fmre.2021.04.007

Zhou, P., Han, J., Cheng, G., Zhang, B., 2019. Learning Compact and Discriminative Stacked Autoencoder for Hyperspectral Image Classification. IEEE Trans. Geosci. Remote Sens. 57, 4823–4833. https://doi.org/10.1109/TGRS.2019.2893180