

	<i>Project:</i> APEMAIA Contract n. 2023-39-HH.0	Doc. N.: DEL_008_DEF_CNR-IIA_RA3_EN-001 Issue/Rev.:01/01 Date: 25/03/2025 (T0+18) Pages: 64
---	---	---

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

WP2102 – HR AOD extraction

DEL008 – Description of the first system module (WP2) for the extraction of high-resolution AOD maps (SW004); description of the Data-driven and Knowledge-Driven procedures for the extraction of LCLU maps in urban areas (SW005) (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, Marica De Lucia, Cristina Tarantino (CNR-IIA) A. Monaco, R. Cilli, A. Fania (DIF) A. Morabito (ARPA Puglia)	CONTRACTOR'S REFERENCE	Issue:01 Rev.:01 Date:25/04/24
PROJECT BEGIN DATE	26/09/2023	PROJECT END DATE	25/09/2025

	<i>Project:</i> <i>APEMAIA</i> Contract n. 2023-39-HH.0	Doc. N.: DEL_008_DEF_CNR-IIA_RA3_EN-001 Issue/Rev.:01/01 Date: 25/03/2025 (T0+18) Pages: 64
---	---	---

	<i>Project:</i> <i>APEMAIA</i> Contract n. 2023-39-HH.0	Doc. N.: DEL_008_DEF_CNR-IIA_RA3_EN-001 Issue/Rev.:01/01 Date: 25/03/2025 (T0+18) Pages: 64
---	---	---

EXECUTIVE SUMMARY			
This document is the finale release of DEL008 – Description of the first system module (WP2) for the extraction of high-resolution AOD maps (SW004); description of the Data-driven and Knowledge-Driven procedures for the extraction of LCLU maps in urban areas (SW005) (preliminary version) – of the Assessment of PM Exposure at an intra-urban scale in preparation for the MAIA mission (APEMAIA) project. It is released at the conclusion of Work Package WP2102 – HR AOD extraction. The release adheres to the planned schedule: (T0+18M). The final version has been scheduled at Third Release of Advancements (RA3) of the project.			
NAME OF ASI PROJECT MANAGER	Name Surname (PM)		
	Name Surname (RUP)		
	Name Surname (Project Scientist)		
STAR CODE		No OF VOLUMES:	

	<i>Project:</i> <i>APEMAIA</i> Contract n. 2023-39-HH.0	Doc. N.: DEL_008_DEF_CNR-IIA_RA3_EN-001 Issue/Rev.:01/01 Date: 25/03/2025 (T0+18) Pages: 64
---	---	---

	<i>Project:</i> <i>APEMAIA</i> Contract n. 2023-39-HH.0	Doc. N.: DEL_008_DEF_CNR-IIA_RA3_EN-001 Issue/Rev.:01/01 Date: 25/03/2025 (T0+18) Pages: 64
---	---	---

TABLE OF CONTENTS

1	AIM OF THE DOCUMENT	6
1.1	Applicable and reference documents	6
1.2	Applicable documents	6
1.3	Acronyms and abbreviations.....	7
2	FIRST SYSTEM MODULE (WP2) FOR THE EXTRACTION OF HIGH-RESOLUTION AOD MAPS.....	8
2.1	Introduction to the AOD satellite-derived measurements	8
2.1.1	Trade-off analysis on the existing products.....	10
2.1.2	AERONET network reference data.....	11
2.1.3	Preprocessing and Organization of Sentinel-3 AOD Dataset.....	13
2.1.4	Preprocessing and Organization of MODIS MAIAC Dataset	16
2.1.5	Preliminary assessment of AOD satellite estimates in comparison with AERONET measurements.....	17
2.2	AOD improvement methodology	21
2.3	Software environment and resources.....	24
3	EXTRACTION OF LC MAPPINGS IN URBAN AREAS	25
3.1	Case study and materials	25
3.2	Methodology	27
4	Reference.....	31

	<i>Project:</i> APEMAIA <i>Contract n.</i> 2023-39-HH.0	<i>Doc. N.:</i> DEL_008_DEF_CNR-IIA_RA3_EN-001 <i>Issue/Rev.:</i>01/01 <i>Date:</i> 25/03/2025 (T0+18) <i>Pages:</i> 64
---	---	---

1 AIM OF THE DOCUMENT

This document represents the final version of the “DEL008 – Description of the first system module (WP2) for the extraction of high-resolution AOD maps (SW004); description of the Data-driven and Knowledge-Driven procedures for the extraction of LCLU maps in urban areas (SW005)” for the **Assessment of PM Exposure** 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 end of these activities adheres the scheduled program: T0+18M.

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

	<i>Project:</i> <i>APEMAIA</i> Contract n. 2023-39-HH.0	Doc. N.: DEL_008_DEF_CNR-IIA_RA3_EN-001 Issue/Rev.:01/01 Date: 25/03/2025 (T0+18) Pages: 64
---	---	---

1.3 Acronyms and abbreviations

Table 1-1 provides the acronym list.

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
DIF	Dipartimento Interateneo di Fisica

	<i>Project:</i> APEMAIA <i>Contract n.</i> 2023-39-HH.0	<i>Doc. N.:</i> DEL_008_DEF_CNR-IIA_RA3_EN-001 <i>Issue/Rev.:</i> 01/01 <i>Date:</i> 25/03/2025 (T0+18) <i>Pages:</i> 64
---	---	--

2 FIRST SYSTEM MODULE (WP2) FOR THE EXTRACTION OF HIGH-RESOLUTION AOD MAPS

2.1 Introduction to the AOD satellite-derived measurements

Aerosol optical depth (AOD) is a dimensionless physical quantity representing the degree to which atmospheric aerosols impede the transmission of light through the atmosphere. AOD is dimensionless because mathematically, at a given wavelength it is defined as:

$$AOD_{\lambda} = \ln\left(\frac{\Phi_{e,\lambda}^i}{\Phi_{e,\lambda}^t}\right) = -\ln(T_{\lambda})$$

Where

- Φ^i is the spectral radiant flux in wavelength received by the atmosphere,
- Φ^t is the spectral radiant flux in wavelength transmitted by the atmosphere
- T is the spectral transmittance in wavelength of the atmosphere

This optical property quantifies the extinction of light due to scattering and absorption by aerosol particles suspended in the atmospheric column. AOD, therefore, measures the amount of light lost due to the presence of aerosols along a vertical path through the atmosphere. AOD values can take any positive value, but typically range from 0.01, indicating extremely clear conditions, to much higher values, such as 5.0, indicating extremely cloudy atmospheric conditions.

The wavelength dependence of AOD provides information on the size distribution of aerosols, making it a valuable parameter for characterizing atmospheric particulate matter and its relationship to ground-level PM concentrations.

Despite a correlation between AOD and PM, this relationship is considerably complex. Several meteorological factors significantly influence this relationship, affecting the conversion from columnar AOD to surface PM concentrations. For a given AOD value, variations in relative humidity (RH) can substantially alter surface PM concentrations through changes in aerosol extinction efficiency. High RH increases aerosol extinction efficiency as hygroscopic particles grow through water absorption, meaning that the same AOD value can correspond to lower dry particle mass concentrations at the surface. Similarly, the planetary boundary layer height (PBLH) plays a crucial role: for the same columnar AOD, a lower PBLH concentrates aerosols into a thinner atmospheric layer, resulting in higher surface PM concentrations, while a higher PBLH leads to more dilute surface concentrations.

	<i>Project:</i> APEMAIA Contract n. 2023-39-HH.0	Doc. N.: DEL_008_DEF_CNR-IIA_RA3_EN-001 Issue/Rev.:01/01 Date: 25/03/2025 (T0+18) Pages: 64
---	--	--

Aerosol Optical Depth (AOD) can be measured through various observational techniques, including ground-based instrumentation and satellite remote sensing, each employing distinct physical principles and offering different trade-offs in terms of spatial coverage, temporal resolution, and measurement accuracy. These complementary methods enable both localized and large-scale assessments of aerosol load in the atmosphere, making AOD a key variable in atmospheric monitoring and air quality studies.

While in situ AOD measurements, such as those from AERONET stations (subsection 2.1.2), are based on direct sun photometry at fixed ground locations and provide high-accuracy, vertically integrated optical properties under cloud-free conditions, satellite-based AOD estimates rely on remote sensing techniques that infer aerosol presence by analyzing the reflectance of the Earth's surface and atmosphere in different spectral bands.

The Sentinel-3 SYNERGY Level-2 product estimates AOD at a native resolution of 300 meters, combining data from the OLCI (Ocean and Land Colour Instrument) and SLSTR (Sea and Land Surface Temperature Radiometer) sensors. These estimates are derived using radiative transfer models and surface-atmosphere separation algorithms, with additional quality control provided by pixel-level flags to account for clouds, invalid geometry, or saturation.

On the other hand, the MODIS MAIAC (Multi-Angle Implementation of Atmospheric Correction) algorithm delivers AOD at 1 km resolution (nominal) by leveraging multi-temporal and multi-angle observations, along with dynamic land classification and spectral reflectance modeling. MAIAC is particularly effective over bright surfaces and urban areas and includes comprehensive quality assurance data to characterize retrieval reliability.

In summary, while both satellite products enable spatially extensive AOD mapping, their estimates differ from in-situ observations due to:

- the use of indirect, model-based retrievals,
- potential cloud and surface reflectance contamination,
- differences in spatial and temporal averaging,
- and the vertical integration methods applied in the radiative inversion.

Given the intrinsic limitations of satellite-based AOD retrievals compared to in situ measurements, in the following section we performed a trade-off analysis across multiple satellite-derived AOD products. The objective was to assess which datasets offer the most effective balance between

	Project: <i>APEMAIA</i> Contract n. 2023-39-HH.0	Doc. N.: DEL_008_DEF_CNR-IIA_RA3_EN-001 Issue/Rev.: 01/01 Date: 25/03/2025 (T0+18) Pages: 64
---	---	---

accuracy, spatial resolution, and temporal revisit frequency, particularly in the context of urban air quality monitoring.

2.1.1 Trade-off analysis on the existing products

In the Table 2-1 as follows, we have reported the most exhaustive list of AOD products provided from the current operational service, at the best of our knowledge.

Table 2-1: List of AOD products estimated from satellite onboard sensors.

PLATFORM	PRODUCT	PROC. LEVEL	SPATIAL RESOLUTION	TEMPORAL RESOLUTION	ACQUISITION TIME	BAND
Sentinel3 Land Product	S3A_SY_2_SYN,	Level 2	300m	5-6 images per week	8 - 10 UTC	AOD @ 550 nm
	S3B_SY_2_SYN					
Sentinel3 AOD Product	S3A_SY_2_AOD, S3B_SY_2_AOD,	Level 2	4.5 km	3-4 images per month	8 - 10 UTC	AOD @ multiple spectral channels (6)
Modis MaiaC	MCD19A2	Level2	1 km	daily	8-11 UTC	AOD @470 nm & @550 nm
Landsat8	LC08_L2SP	Level2	30 m	16 days	8-10 UTC	SR_QA_AEROSOL
Sentinel2	S2A_MSIL2A	Level2	60 m ricampionat a a 20 m	3 - 5 days	8-10 UTC	Quality_AOT
VIIRS	AERDB_L2_VIIRS_SNPP	Level2	6 km	2 per day	10-13 UTC	AOD @550 nm
Sentinel2 MAJA	SENTINEL2A	Level2	10 m e 20 m	3 - 5 days	8-10 UTC	AOD @550 nm
NASA's TERRA	MISR	Level2	10 km - < 50 km	daily- <weekly	10:30 local time	AOD product from 4 spectral bands: 446, 558, 672, 866 nm

	<i>Project:</i> APEMAIA Contract n. 2023-39-HH.0	Doc. N.: DEL_008_DEF_CNR-IIA_RA3_EN-001 Issue/Rev.:01/01 Date: 25/03/2025 (T0+18) Pages: 64
---	---	---

Sentinel5/5 P TROPOMI	L2__AER_AI, L2__AER_LH	Level2	3.5 x 5.5 Km	varying	Varying (multiple times per day)	Aerosol UV absorption index, layer height, optical depth (TBC)
---	-------------------------------	--------	--------------	---------	----------------------------------	--

At the beginning stage of the project, we collected and analysed a few of the products in Table 2-1 that exhibits specifications in line with the project requirements including Sentinel2 with Maja atmospheric correction algorithm; VIIRS data and Sentinel3 AOD Product at coarse resolution 4.5 km.

However, after a careful evaluation of AOD product benchmark, we considered only Sentinel3 Land Product (300m) and MODIS MAIAC (1 km) suitable for the implementation of “AOD improvement” module.

First, we collected AOD data from the MODIS MAIAC product (MCD19A2) provided by the NASA Terra/Aqua platforms. Then, we downloaded AOD measurements at 550 nm (SY_2_SYN product) from the Sentinel 3 Synergy Land products, which are released at a higher spatial resolution of 300 m. MODIS MAIAC is the current state of the art in AOD retrieval. The Sen3 SYN products are an experimental product which deserves to be investigated due to its high spatial resolution and high revisit frequency.

To systematically analyse Aerosol Optical Depth (AOD) over Italian urban areas such as Bari, Taranto and Rome, a complete preprocessing and data management workflow was developed. This pipeline combines automatic data retrieval, spatial filtering, reprojection, quality control via flag filtering, and time-series extraction for correlation with ground-based AERONET stations.

2.1.2 AERONET network reference data

We have downloaded the daily average AOD measurements from the AERONET network for all available stations in Italy for the reference period from January 2019 to December 2022. The data is freely accessible on the website <https://aeronet.gsfc.nasa.gov/>. For Italy, we were able to collect data from a total of 27 stations shown in Figure 1, amounting to approximately 8,400 daily average measurements over the reference period. The lower-than-expected number of daily average measurements ($4 \times 365 \times 27$) is due to missing data, sometimes over extended periods.



Project:
APEMAIA

Contract n.
2023-39-HH.0

Doc. N.: DEL_008_DEF_CNR-IIA_RA3_EN-001

Issue/Rev.:01/01

Date: 25/03/2025 (T0+18)

Pages: 64



Figure 1 - Spatial Distribution of the AERONET monitoring stations in Italy.

In the Apulia region, only two AERONET stations are currently active. Among them, the Apulia Tavoliere station has been acquiring data only since 2022 and therefore does not cover the entire study period (2019–2023). Lecce station provides a long time series to be compared with satellite AOD measurements.



Project:
APEMAIA

Contract n.
2023-39-HH.0

Doc. N.: DEL_008_DEF_CNR-IIA_RA3_EN-001

Issue/Rev.:01/01

Date: 25/03/2025 (T0+18)

Pages: 64

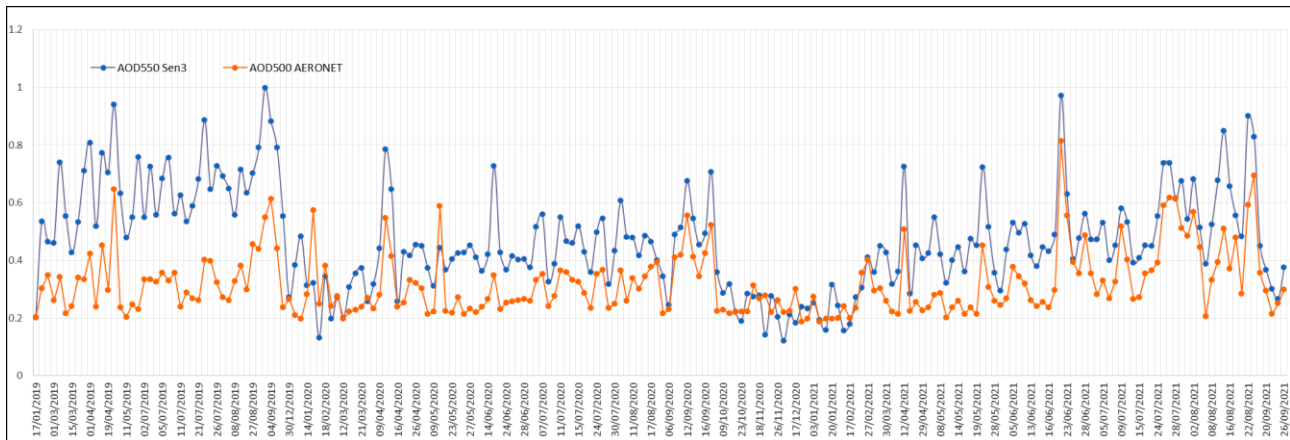


Figure 2 –2020–2021 AOD time series for the LECCE AERONET station and cloud-free S3A_SY_2_SYN products.

For our analysis, we selected AERONET Level 2.0 products (Quality-Assured), which have undergone rigorous validation and quality control, making them suitable for scientific analysis and for fine-tuning AOD maps produced during the project.

2.1.3 Preprocessing and Organization of Sentinel-3 AOD Dataset

Sentinel3 data acquisition was performed using the cdsetool utility (Python environment), enabling automated querying and downloading of Sentinel-3 SYNERGY Level-2 (SY_2_SYN___) products. The resulting .SEN3.zip archives were organized in dedicated ReCas folder by area and year, and each archive was decompressed into a dedicated folder bearing the same filename.

In Linux, this was achieved through a simple bash loop that:

- Lists all .zip files,
- Creates a target directory,
- Extracts the contents, and
- Moves the archive to an internal folder (zipdir) for backup.

Each product includes critical metadata embedded in its filename and folder structure, as specified in Tabe2-1.

	<i>Project:</i> APEMAIA <i>Contract n.</i> 2023-39-HH.0	Doc. N.: DEL_008_DEF_CNR-IIA_RA3_EN-001 Issue/Rev.:01/01 Date: 25/03/2025 (T0+18) Pages: 64
---	---	---

Table 2-2:File name specifications

FIELD	DESCRIPTION
MISSION	S3A or S3B (Sentinel-3 platform)
START/STOP TIME	UTC acquisition window
PROCESSING LEVEL	SY_2_SYN___ for Level-2 SYNERGY
PROCESSING BASELINE	Version of the processor used
ORBIT / SEGMENT	Orbital info for spatial referencing
TIMELINESS	NTC, STC, or NRT
EXTENSION	.SEN3 indicates a product directory

We specifically selected **SY_2_SYN___** products with the **NT** suffix, which refers to **NTC (Non-Time Critical)** timeliness. While not strictly categorized as "Reprocessing", NTC products are the most refined and stable version available within the Copernicus Sentinel-3 operational framework.

2.1.3.1 Reprojection to UTM 33N (EPSG:32633)

To ensure spatial consistency across all Sentinel-3 products, each **SY_2_SYN___** scene was reprojected to the Universal Transverse Mercator (UTM) Zone 33N coordinate system (EPSG:32633). This transformation is critical for geospatial analyses over the Italian peninsula, including aerosol monitoring and pixel-wise comparison with ground truth data.

A SNAP Graph (`snap_reproject.xml`) was designed using the Graph Builder to automate the reprojection workflow and export the results into GeoTIFF format (Figure 2). The graph includes the following main operations:

Read: Ingests the `xfdumanifest.xml` file from each **.SEN3** product folder.

Reproject: Applies the transformation to EPSG:32633 using nearest-neighbour resampling.

Write: Outputs the reprojected raster (AOD and uncertainty) in GeoTIFF format.

The reprojection was executed iteratively on the ReCaS HTCondor infrastructure, leveraging the SNAP Graph Processing Tool (`gpt`) via a custom bash script. This automated batch script loops over all **.SEN3** directories, locates the corresponding `xfdumanifest.xml`, and runs the reprojection graph for both the AOD (T550) and its associated uncertainty layer 'T550_err'.

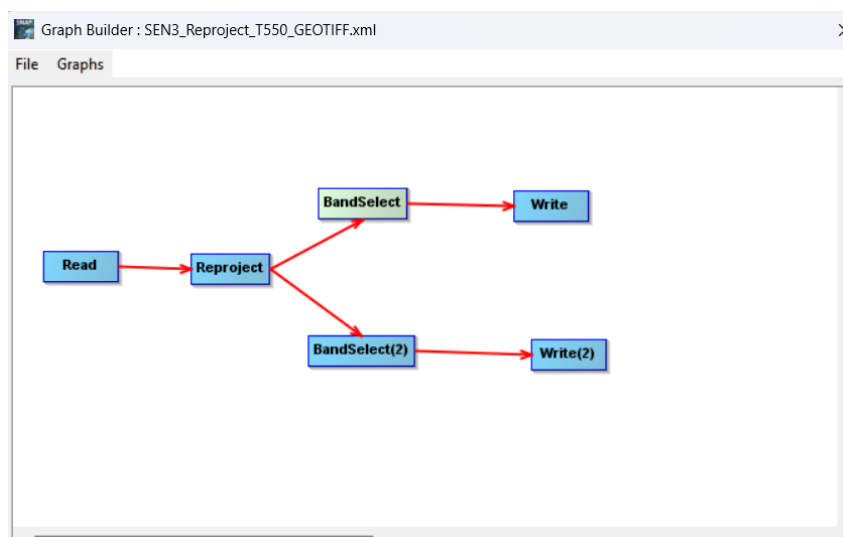


Figure 3 – Reprojecting graph model implemented in SNAP ESA software.

2.1.3.2 Quality Flag Filtering on AOD Band

To ensure reliable AOD estimates, a **BandMaths** operation was used (T550_flagging.xml). It applies a conditional mask on the AOD at 550 nm (T550) using a combination of quality flags as explained in Table

Table 2-3: Applied flag and relative meaning.

Flag Name	Condition Applied	Description
CLOUD_flags_CLOUD	= FALSE	Excludes pixels classified as cloudy by the cloud mask.
SYN_flags_SYN_AOT_climato	= FALSE	Excludes AOD values derived from climatology rather than observation.
SYN_flags_SYN_SDR_OOR	= FALSE	Filters out pixels where the Surface Directional Reflectance is out of range.
SYN_flags_SYN_high_error	= FALSE	Masks AOD pixels with high associated error.
OLC_flags_OLC_invalid	= FALSE	Removes invalid pixels from the OLCI processing chain.

	<i>Project:</i> APEMAIA Contract n. 2023-39-HH.0	Doc. N.: DEL_008_DEF_CNR-IIA_RA3_EN-001 Issue/Rev.:01/01 Date: 25/03/2025 (T0+18) Pages: 64
---	---	---

OLC_flags_OLC_summary_saturated = FALSE	Excludes pixels where any band is saturated (radiometrically unusable).
--	---

The flagging step was executed via a shell script that loops over all reprojected T550 GeoTIFF files and applies the SNAP graph using the **Graph Processing Tool (gpt)**.

Each output file contains a filtered AOD layer where invalid or uncertain pixels are set to NaN. These products represent the final input for time-series extraction and correlation with AERONET stations.

This step ensures that only high-confidence observations are included in the final analysis, improving robustness for scientific validation and comparison with AERONET ground-based measurements.:

- Removing low-quality AOD values based on internal product quality indicators.
- Producing a clean, georeferenced raster containing only valid AOD values.
- Facilitating downstream statistical analysis with a consistent and trustworthy dataset.

2.1.4 Preprocessing and Organization of MODIS MAIAC Dataset

As part of the APEMAIA project's effort to generate high-resolution and validated aerosol products, a dedicated workflow was implemented to process NASA's MODIS MAIAC AOD (MCD19A2) data. The procedure encompassed automated data download, resampling to 300 m resolution, and statistical extraction of AOD values around given locations of interest.

The official MAIAC AOD product (MCD19A2, Version 6) provides daily gridded aerosol optical depth at 1 km resolution and includes quality assurance (QA) flags. A complete description of the product can be found at: <https://lpdaac.usgs.gov/products/mcd19a2v006/>. The data were retrieved from NASA's LAADS DAAC portal: <https://ladsweb.modaps.eosdis.nasa.gov/>.

To download the data efficiently in batch mode, a wget command was customized in a Linux environment. Files were downloaded and stored on ReCas environment in structured folders and named by date and tile.

Subsequently, a custom Python script was developed using GDAL, PyProj, and Pandas to automate the following key preprocessing tasks:

a Resampling to 300 m

- Although MAIAC provides 1 km resolution, the dataset was resampled to 300 meters using the `gdal.Warp()` function with cubic interpolation, in memory. This downscaling supports

	<i>Project:</i> <i>APEMAIA</i> Contract n. 2023-39-HH.0	Doc. N.: DEL_008_DEF_CNR-IIA_RA3_EN-001 Issue/Rev.:01/01 Date: 25/03/2025 (T0+18) Pages: 64
---	---	---

higher-resolution modeling and alignment with Sentinel-3 grids. Invalid or missing values (32767, -9999, etc.) were excluded. The mean AOD coming from valid pixels of multiple bands in MAIAC products was computed and stored.

c. Matching by Date. To ensure temporal consistency:

- Each input filename was parsed to extract the date.
- The script searched for the corresponding HDF file (A<YYYY><DOY>.hdf) in the MODIS archive.

d. AOD and QA Band Access. The script opened two specific subdatasets from each HDF:

- Optical_Depth_055: the core AOD band.
- AOD_QA: bitmask flags describing the pixel's quality, snow presence, and water cover.

A dedicated function translated these flags into readable categories (e.g., “Good / Snow: No / Water: Yes”).

For comparison with AERONET station measurements, we developed automated scripts that generate output files containing the latitude and longitude of each extracted pixel, the corresponding AOD values from the resampled MODIS MAIAC layer, and all associated quality assurance (QA) parameters, including pixel-level flags indicating retrieval quality, cloud presence, snow cover, and water classification.

2.1.5 Preliminary assessment of AOD satellite estimates in comparison with AERONET measurements

A preliminary evaluation of satellite-derived AOD estimates was conducted to assess their accuracy and reliability compared to ground-based measurements. The analysis used AERONET Level 2 data that includes data subjected to both cloud masking and rigorous quality control from seven stations across Italy over the period 2019–2022 as detailed in Table below.

	<i>Project:</i> <i>APEMAIA</i> Contract n. 2023-39-HH.0	Doc. N.: DEL_008_DEF_CNR-IIA_RA3_EN-001 Issue/Rev.:01/01 Date: 25/03/2025 (T0+18) Pages: 64
---	---	---

STATION	COORDINATE	WAVELENGTH	ANGSTROM EXPONENT
Lecce	(40.33,18.11)	500	500–870_Angstrom_Exponent
Foggia	(41.46,15.50)	500	500–870_Angstrom_Exponent
Potenza	(40.60,15.72)	500	500–870_Angstrom_Exponent
Roma La Sapienza	(41.90,12.51)	500	500–870_Angstrom_Exponent
Roma Tor Vergata	(41.83, 12.64)	500	500–870_Angstrom_Exponent
Ispira	(45.80,8.62)	490	440–675_Angstrom_Exponent
Modena	(44.63,10.94)	500	500–870_Angstrom_Exponent

Tab. AERONET stations

For almost all stations, the AOD value at = 500nm was available, so the AOD at 550nm was calculated using the 500 – 870 Angstrom Exponent value reported in the downloaded dataset. The only exception concerns the Ispira station, where the AOD value at 490nm was used and the Angstrom exponent was in the range 440 – 675nm, and the conversion formula was adjusted accordingly to obtain the AOD at 550nm. This conversion was necessary to standardise the AERONET measurements to the 550nm wavelength commonly used in satellite AOD products, thus ensuring a direct and accurate comparison between ground-based and satellite data.

$$AOD_{550nm} = AOD_{500nm} \left(\frac{550nm}{500nm} \right)^{-\alpha}$$

Fig. Angstrom equations

This process identified and removed extreme AOD values that deviated significantly from the typical distribution as potential outliers. These outlier AOD measurements are likely indicative of uncorrected instrumental errors rather than actual atmospheric conditions.

	<i>Project:</i> APEMAIA <i>Contract n.</i> 2023-39-HH.0	<i>Doc. N.:</i> DEL_008_DEF_CNR-IIA_RA3_EN-001 <i>Issue/Rev.:</i> 01/01 <i>Date:</i> 25/03/2025 (T0+18) <i>Pages:</i> 64
---	---	--

Figure presents the correlation analysis between the MAIAC AOD readings and the corresponding AERONET measurements of 0.75 based on 2561 coincident observations. This result is consistent with the established accuracy range for MAIAC AOD retrievals reported in the literature.

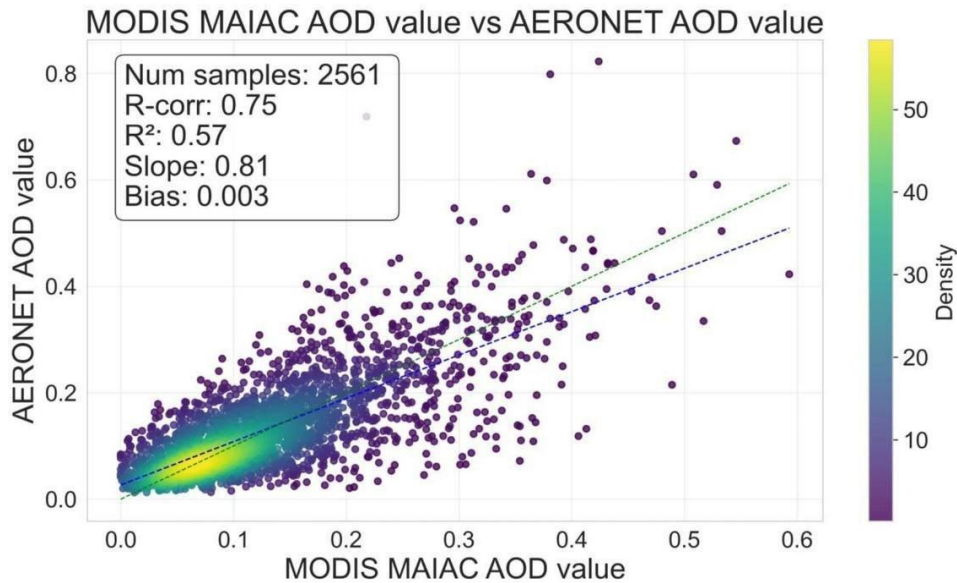


Fig. Correlation between MODIS MAIAC and AERONET data

In the context of preprocessing AOD data, we evaluated how to handle the overlapping products (synergy-syn) from Sentinel-3 Synergy A and B by examining their correlation with AERONET reference data. As illustrated in figure 4.3, both products showed a Pearson correlation coefficient with AERONET of approximately 0.49, based on 2550 and 2088 observations, respectively.

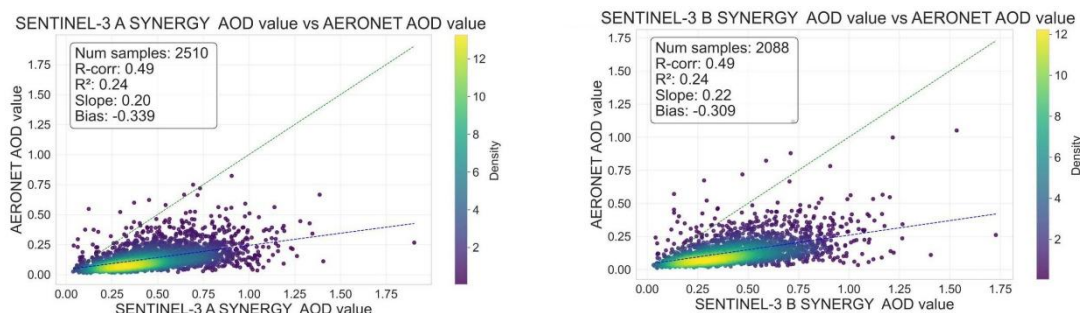


Fig. Correlation analysis between satellite Sentinel-3 retrievals and AERONET ground-based measurements for (A) Synergy A and (b) Synergy B

	<i>Project:</i> APEMAIA <i>Contract n.</i> 2023-39-HH.0	<i>Doc. N.:</i> DEL_008_DEF_CNR-IIA_RA3_EN-001 <i>Issue/Rev.:</i> 01/01 <i>Date:</i> 25/03/2025 (T0+18) <i>Pages:</i> 64
---	---	--

On days when both products were present, the average of the two values was considered, and the time series shows a correlation of 0.50 with AERONET data.

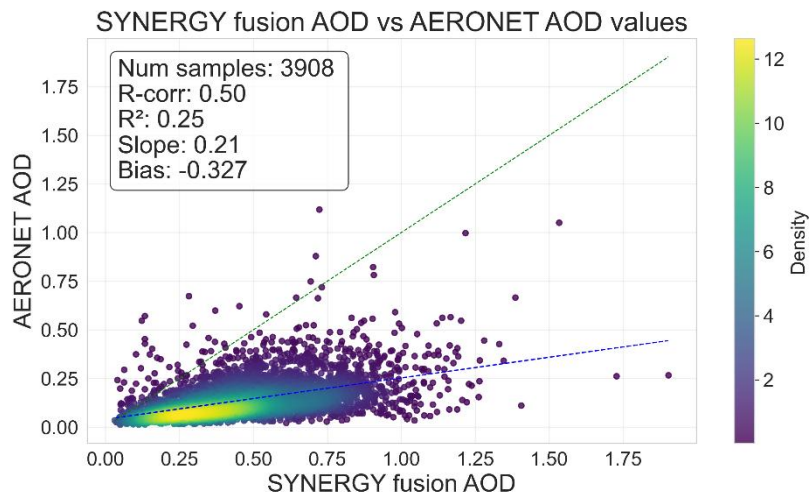


Fig. Correlation analysis between satellite Sentinel-3 Synergy fusion and AERONET ground-based measurements

An example of a detailed temporal analysis was conducted at the Rome La Sapienza station, demonstrating clear seasonal trends in AOD variability and highlighting systematic differences between products. In particular, Sentinel-3 measurements show a tendency to systematically overestimate AOD values compared to AERONET, particularly during episodes with high aerosol loading, suggesting the need for error correction procedures in subsequent modelling applications.

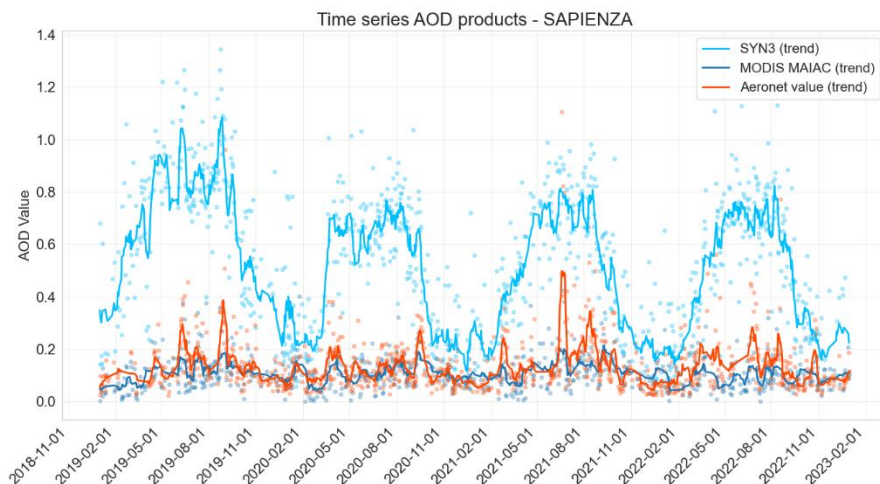


Fig. Time series of AOD products (SYN3, MODIS MAIAC, and AERONET) at the "Roma la Sapienza" station (2019-2022)

	<i>Project:</i> APEMAIA Contract n. 2023-39-HH.0	Doc. N.: DEL_008_DEF_CNR-IIA_RA3_EN-001 Issue/Rev.:01/01 Date: 25/03/2025 (T0+18) Pages: 64
---	---	---

2.2 AOD improvement methodology

For improving and make more robust AOD high resolution daily estimates, we proposed a two-step procedure based on 1) AOD data fusion with ensemble models and, as second, 2) Gap-filling and downscaling of existing coarse resolution AOD products. These include the need to additionally acquire MERRA2 and CAMS reanalysis products.

A **statistical comparison** between ground-based AERONET observations and several satellite and reanalysis products has been performed as a preliminary step:

- **CAMS (ECMWF) and NASA MERRA2:** Global reanalysis datasets with 50 km resolution.
- **MODIS MAIAC:** 1 km AOD derived from MODIS satellites using the MAIAC algorithm.
- **Sentinel-3 SYNERGY (SY_2_SYN___):** 300 m native resolution, processed with rigorous flagging and reprojection.

Table 2-4:Key Metrics (Satellite/Model AOD vs AERONET).

Product	Pearson's R	RMSE	BIAS	EE (%)	N
CAMS AOD	67.6%	0.105	-0.014	72.5%	7240
MERRA2 AOD	73.7%	0.10	-0.04	74.0%	7240
MODIS MAIAC	76%	0.08	-0.001	75.4%	4486
SEN3_SYN	46.9%	0.08	0.275	13.1%	3272

Sentinel-3 performed less effectively in terms of raw metrics due to missing data and higher relative errors in single pixel measurements. Nevertheless, it was considered valuable for its native high spatial resolution

To maximize the strengths of all aforementioned data we proposed to implement an **ENSEMBLE Model** that combines multiple sources of AOD data into a single high-resolution predictor. The Ensemble model significantly outperformed each individual source.

We chose two benchmark models: the Ordinary least squared - linear regression and Random Forest. These models were trained on about 4000 observations of the AERONET network in Italy, period 2019-2023.



Project:
APEMAIA

Contract n.
2023-39-HH.0

Doc. N.: DEL_008_DEF_CNR-IIA_RA3_EN-001

Issue/Rev.:01/01

Date: 25/03/2025 (T0+18)

Pages: 64

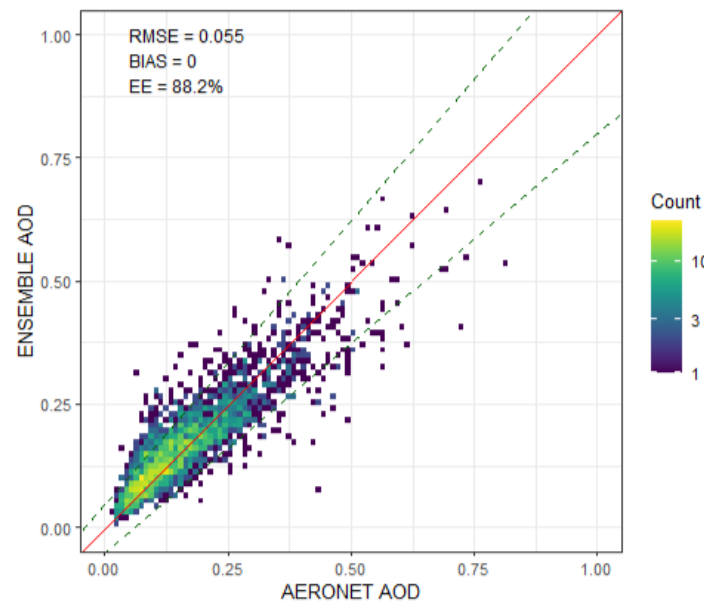


Figure 4 – Ensemble Model performances.

A permutation feature importance analysis was performed on the Random Forest model, computing the mean decrease in accuracy to quantify the relevance of each predictor to the final AOD estimates (Figure 3). Notably, the S3_SYN_product exhibited a lower importance than initially expected.

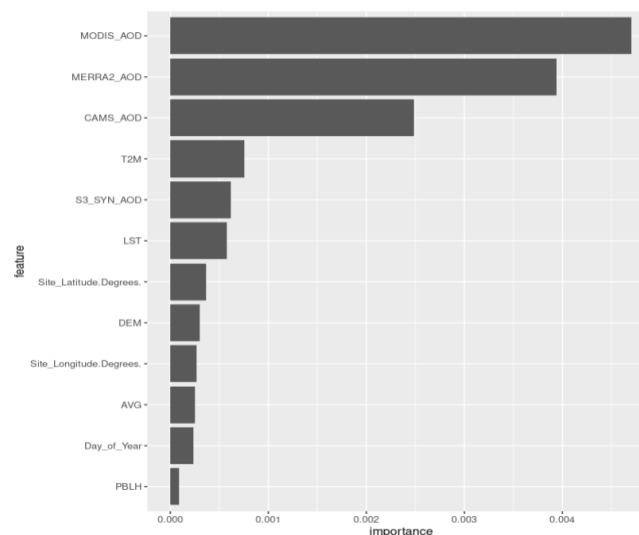
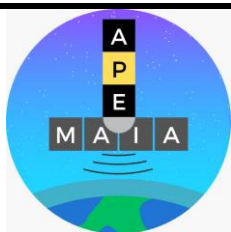


Figure 5 - Random Forest Permutation Feature Importance



Project:
APEMAIA

Contract n.
2023-39-HH.0

Doc. N.: DEL_008_DEF_CNR-IIA_RA3_EN-001

Issue/Rev.:01/01

Date: 25/03/2025 (T0+18)

Pages: 64

A major challenge was the **high frequency of missing data**, particularly in Sentinel-3 and MODIS due to cloud cover. For example, only a small fraction of pixels had valid AOD values across all dates in Puglia from 2019–2023. To address this issue, a **Gap-Filling procedure** was developed using both **Linear Regression (LM)** and **Random Forest (RF)** models. Predictors included are X, Y, Elevation, MODIS AOD (mean), MERRA2, CAMS.

We set up the AOD time series gap-filling problem as a regression problem whose predictors are the AOD estimates from reanalysis products and other auxiliary data while the target is the AOD from the linear ensemble model. Figure 6 shows a scatterplot between gap-filled AOD from the linear model and the AOD measurements from the AERONET in Italy.

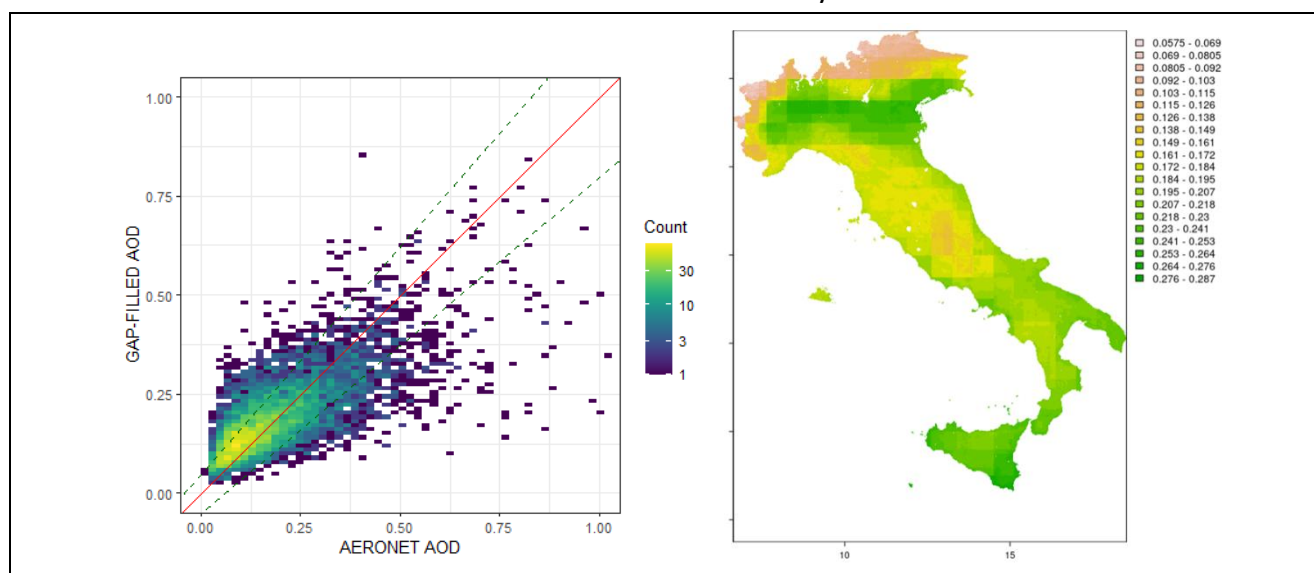


Figure 6 -Scatterplot between gap-filled AOD from the linear model and the AOD measurements from the AERONET in Italy; on the right Mean Gap-filled AOD for year 2022.

The models achieved strong performance even under **leave-one-year-out (LOYO)** and **leave-one-station-out (LOSO)** cross-validation scenarios (Table 2-4).

Table 2-5: Leave-one-year and Leave-one-Station out CV of ML models

Model	R	RMSE	EE (%)
-------	---	------	--------

	<i>Project:</i> <i>APEMAIA</i> Contract n. 2023-39-HH.0	Doc. N.: DEL_008_DEF_CNR-IIA_RA3_EN-001 Issue/Rev.:01/01 Date: 25/03/2025 (T0+18) Pages: 64
---	---	---

Gap-filling LM (test 2022)	79.7%	0.083	80.3%
LM Cross-validation (LOYO)	88.2%	0.054	89.0%
LM Cross-validation (LOSO)	86.0%	0.057	86.7%

These findings demonstrate that there is room for further improvement of AOD products by imputing missing AOD values and by further calibrating AOD using regression models fed with available AOD estimates and auxiliary data.

Finally, we are also testing deterministic spatial interpolation algorithms. Currently, the gap-filling of AOD data is being evaluated using inverse distance weighting (IDW) interpolation. This technique estimates missing values based on the proximity of neighbouring points, assigning greater influence on those that are closer. The goal of this approach is to ensure spatial continuity in the dataset, particularly in regions with sparse or incomplete measurements, improving the accuracy of the final AOD maps.

2.3 Software environment and resources

The development of software for synthesizing recalibrated AOD maps was implemented using R programming language, relying on functions for geospatial data manipulation and data-driven modeling, such as terra, sf, stat, and ranger. Currently, all the processing chains are deployed and distributed on the HTC clusters and rely on CPUs only. All AOD processing routines and the AOD products are available on the shared space of the ReCaS platform at the path /lustre/apemaia_project/Output/DIF/AOD_24_mag_2024/.

	<i>Project:</i> APEMAIA <i>Contract n.</i> 2023-39-HH.0	<i>Doc. N.:</i> DEL_008_DEF_CNR-IIA_RA3_EN-001 <i>Issue/Rev.:</i> 01/01 <i>Date:</i> 25/03/2025 (T0+18) <i>Pages:</i> 64
---	---	--

3 EXTRACTION OF LC MAPPINGS IN URBAN AREAS

3.1 Case study and materials

The first case study analysed concerns the city of Bari in the Apulia region, southern Italy. Although Bari municipality covers an area of about 117 km², a rectangular study area of 455 km² was considered for computational purposes, including other nearby municipalities such as Modugno, Bitonto, Bitetto, Triggiano, Bitritto, Valenzano, Capurso and Noicattaro.

Figure 3 shows the study area with the boundaries of the municipalities overlaid in blue and Bari municipality boundary in red.

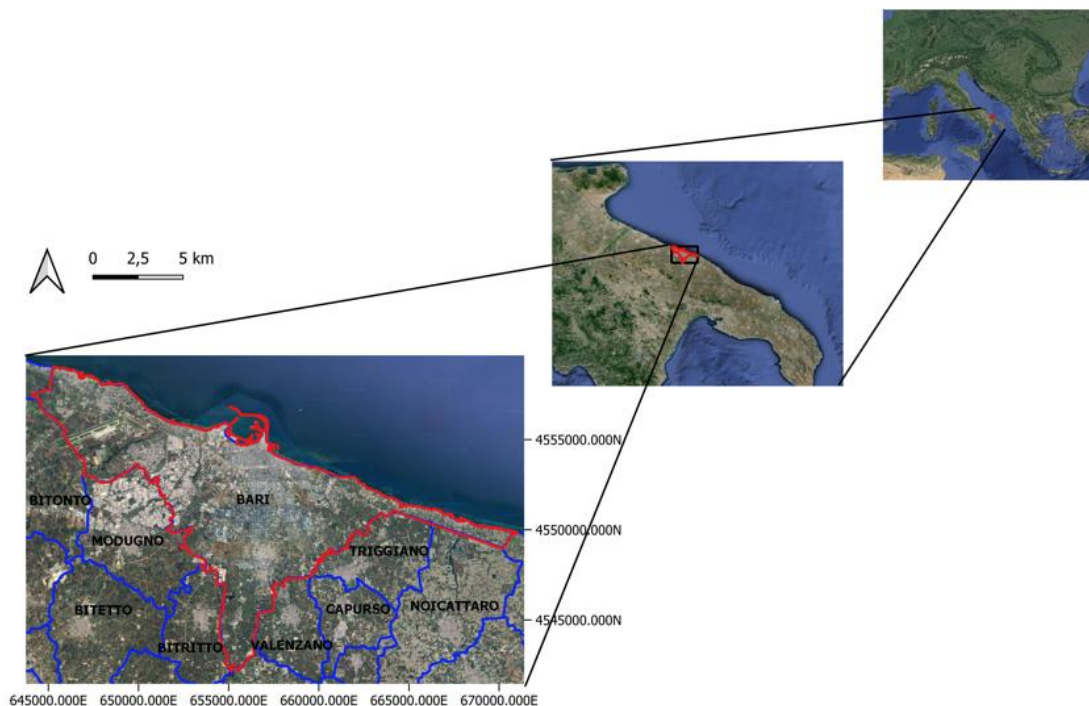


Figure 3. The study area with its geographic location: the red line shows the boundary of Bari municipality and the blue line shows the boundary of the surrounding municipalities. Image source: Google Satellite.

A multi-temporal Sentinel-2A/B dataset composed by 4 images, one for each season, for each year from 2019 to 2022 was considered. The images were downloaded from the Copernicus Data Space Ecosystem (<https://dataspace.copernicus.eu/>) - also known as Copernicus Open Access Hub - and were selected based on having less than 10% cloud cover, atmospherically corrected as surface

	<i>Project:</i> APEMAIA <i>Contract n.</i> 2023-39-HH.0	<i>Doc. N.:</i> DEL_008_DEF_CNR-IIA_RA3_EN-001 <i>Issue/Rev.:</i>01/01 <i>Date:</i> 25/03/2025 (T0+18) <i>Pages:</i> 64
---	---	---

reflectance data and geometrically orthorectified (level 2A). The preferred acquisition dates remained the same from year to year. Tile 33TXF was selected for the area of interest. Table 1 reports the list of the images considered.

Table 1: List of Sentinel-2A/B images considered.

Year	Acquisition date
2019	February 19 th
	June 7 th
	July 22 nd
	October 2 nd
2020	February 17 th
	May 22 nd
	July 23 rd
	October 11 th
2021	February 26 th
	May 24 th
	July 23 rd
	October 4 th
2022	February 21 st
	May 22 nd
	July 23 rd
	October 4 th

The images were cropped according to the area of interest and stacked in a multi-temporal file for each year.

	<i>Project:</i> APEMAIA <i>Contract n.</i> 2023-39-HH.0	<i>Doc. N.:</i> DEL_008_DEF_CNR-IIA_RA3_EN-001 <i>Issue/Rev.:</i> 01/01 <i>Date:</i> 25/03/2025 (T0+18) <i>Pages:</i> 64
---	---	--

3.2 Methodology

For each year, LC mappings at a spatial resolution of 10 meters were produced by considering the Sentinel-2A/B data listed in Table 1 as input for a semi-automatic, data-driven, object-based classification algorithm. According to (Blaschke, 2010), the Geographic Object-Based Image Analysis (GEOBIA) (Chen et al., 2018) can improve classification accuracy in areas with complex structures such as cities. Hence, this approach was preferred due to the urban nature of the study area. The reference data used to train the classifier were extracted from the visual interpretation of the satellite images identifying the 10 LC classes reported in Table 2. These classes were expressed according to the Food and Agriculture Organization-Land Cover Classification System 2 (FAO-LCCS2) taxonomy. This latter was preferred because it is based on a set of hierarchical rules for integrating EO data with in-situ and ancillary data (Tarantino et al., 2023; Adamo et al., 2020; Tomaselli et al., 2013).

Table 2: List of LC classes mapped.

No.	FAO-LCCS 2 Code	FAO-LCCS 2 Description	Generic Description
1	-----	-----	NO DATA
2	B15/A1.A13	BUILT UP.URBAN AREAS	BUILDINGS
3	A12/A3orA4.E2	NATURAL VEGETATION/TREESorSHRUBS.DECIDUOUS	NATURAL VEGETATION DECIDUOUS
4	B15/A1.A3.A7	ROADS	ROADS
5	B28/A1	WATER	WATER
6	B15/A2.A6	EXTRACTION SITES	QUARRY
7	A12/A3orA4.E1	NATURAL VEGETATION/TREESorSHRUBS.EVERGREEN	NATURAL VEGETATION EVERGREEN
8	A12/A2.A6	NATURAL VEGETATION/HERBACEOUS.GRAMINOID	GRASSLAND
9	A11	CULTIVATED LAND	ARABLE LAND
10	-----	-----	SHADOW

According to the GEOBIA approach, the process involved two steps: a segmentation followed by a labelling. The full stack of 4 seasonal images for 2022 was segmented by applying the multi-resolution algorithm in the eCognition 9.0 framework (Trimble, 2019). The configuration parameters were set using a trial-and-error process to determine the cluster size that best matched the different patterns in the scene. Figure 4 shows a close-up of the scene with the segmentation layer imposed.

	<i>Project:</i> <i>APEMAIA</i> Contract n. 2023-39-HH.0	Doc. N.: DEL_008_DEF_CNR-IIA_RA3_EN-001 Issue/Rev.:01/01 Date: 25/03/2025 (T0+18) Pages: 64
---	---	---

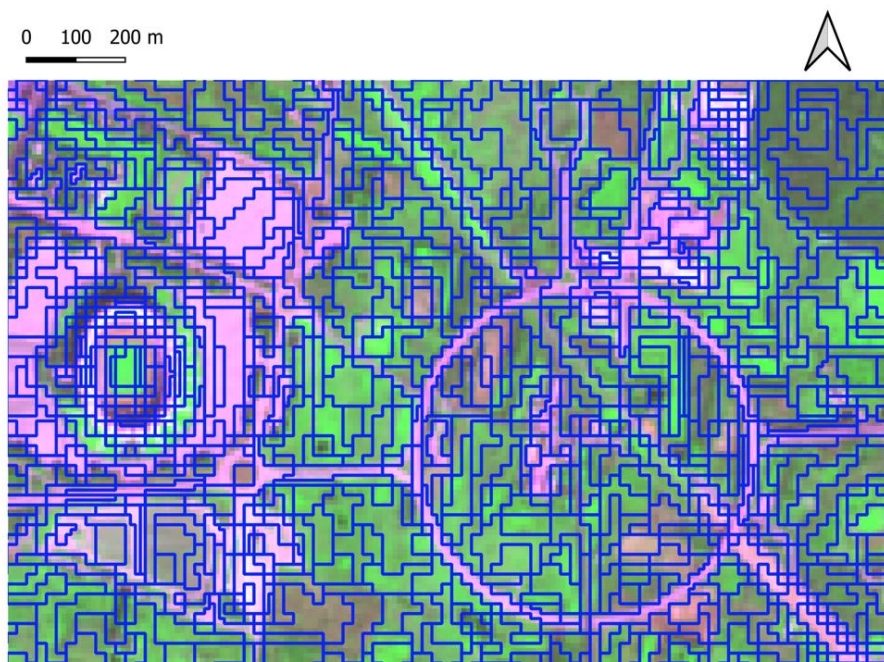


Figure 4. Close-up of a Sentinel-2 image (False Colour Composite RGB=Red-NIR-Blue)
with segmentation layer imposed in blue.

The LC mapping was obtained using Random Forest (RF) classifiers (Breiman, 2001). They were preferred, after comparing them with Support Vector Machine (SVM) classifiers, for their better performances probably due to the use of a large training dataset (Liaw & Wiener, 2002).

The reference dataset was composed of clusters from the segmentation layer labelled using a visual interpretation process on multi-temporal images from 2022. This dataset was used to train the RF classifier which was fed in input with the stack of 4 seasonal images within Orfeo ToolBox environment (OTB). The configuration of the classifier parameters was determined through a trial-and-error process to achieve the highest possible Overall Accuracy (OA) in the classified maps. This latter was computed using a validation dataset composed by polygons from visual interpretation of high-resolution images and in-field campaigns in the framework of previous projects.

The same procedure was adopted for the classification of 2019, 2020 and 2021 images: the reference data were checked on the images related to these years deleting those clusters changed in label. The 4 mappings obtained are reported in Figure 5.

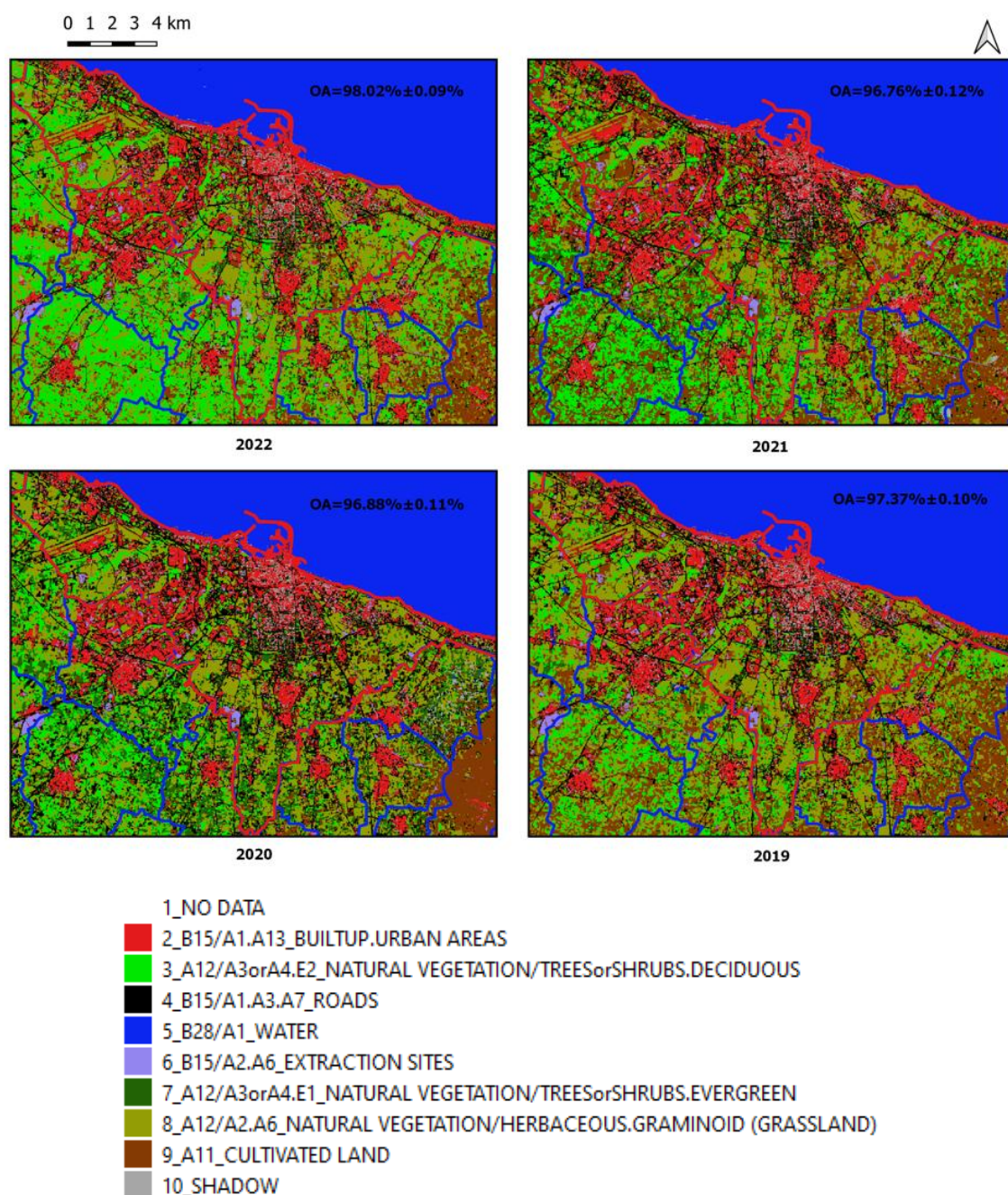


Figure 5. LC mappings for 2019, 2020, 2021, 2022.

	<i>Project:</i> APEMAIA <i>Contract n.</i> 2023-39-HH.0	<i>Doc. N.:</i> DEL_008_DEF_CNR-IIA_RA3_EN-001 <i>Issue/Rev.:</i>01/01 <i>Date:</i> 25/03/2025 (T0+18) <i>Pages:</i> 64
---	---	---

Time series analysis of the LC mapping reveals a clear upward trend in built-up areas as can be observed in Figure 6. This is at the expense of crops and natural ecosystems. The trend results in agreement with land consumption trend reported by ISPRA (Figure 7).

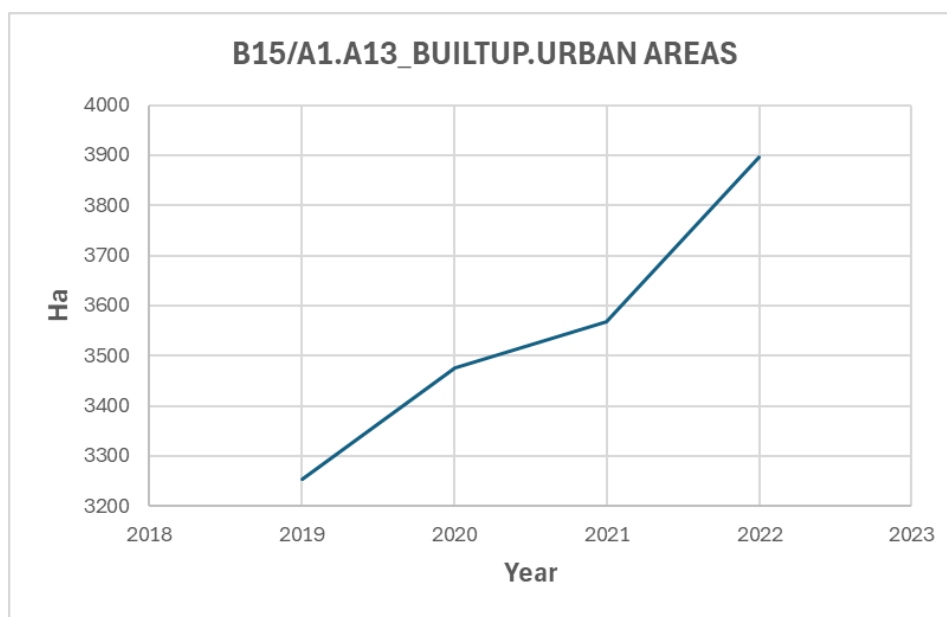


Figure 6. Temporal trend for the B15/A1.A13_BUILTUP.URBAN AREAS class.

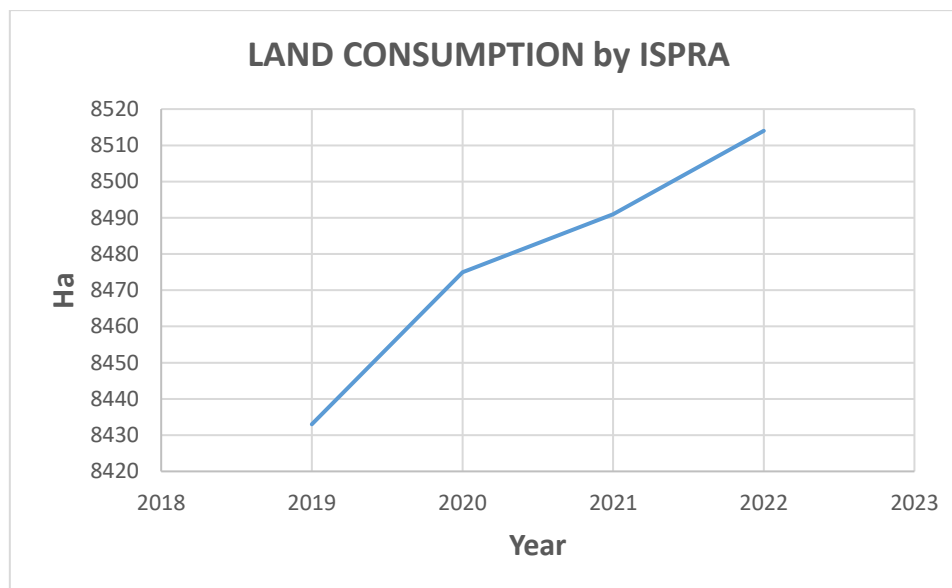


Figure 7. Temporal trend for land consumption reported by ISPRA.

	<i>Project:</i> APEMAIA <i>Contract n.</i> 2023-39-HH.0	<i>Doc. N.:</i> DEL_008_DEF_CNR-IIA_RA3_EN-001 <i>Issue/Rev.:</i> 01/01 <i>Date:</i> 25/03/2025 (T0+18) <i>Pages:</i> 64
---	---	--

4 Reference

Adamo, M.; Tomaselli, V.; Tarantino, C.; Vicario, S.; Veronico, G.; Lucas, R.; Blonda, P. (2020). "Knowledge-Based Classification of Grassland Ecosystem Based on Multi-Temporal Worldview-2 data and FAO-LCCS Taxonomy", *Remote Sensing*, Special Issue "Remote Sensing for Habitat Mapping", 12, 1447, <https://doi.org/10.3390/rs12091447>

Blaschke, T. (2010). Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65(1), 2-16.

Breiman, L. Random Forests. *Machine learning*, 2001, 45: 5-32. <https://link.springer.com/article/10.1023/A:1010950718922>

Chen, G.; Weng, Q.; Hay, G.J.; He, Y. (2018). Geographic Object-Based Image Analysis (GEOBIA): Emerging Trends and Future Opportunities. *GIScience Remote Sens.*, 55, 159–182. <https://doi.org/10.1080/15481603.2018.1426092>

Liaw, A. & Wiener, M. (2002). Classification and regression by randomForest. *R news*, 2/3, 18-22. <https://journal.r-project.org/articles/RN-2002-022/RN-2002-022.pdf>

OTB CookBook 6.4.0, 2018. <https://www.orfeo-toolbox.org/CookBook-6.4/index.html#>

Tarantino, C.; Aquilino, M.; Labadessa, R.; Adamo, M. (2023). "Time Series of Land Cover Mappings Can Allow the Evaluation of Grassland Protection Actions Estimated by Sustainable Development Goal 15.1.2 Indicator: The Case of Murgia Alta Protected Area", *Remote Sensing*, Special Issue "Remote Sensing Measurements for Monitoring Achievement of the Sustainable Development Goals (SDGs)", 15(2), 505, <https://doi.org/10.3390/rs15020505>

Tomaselli, V.; Dimopoulos, P.; Marangi, C.; Kallimanis, A. S.; Adamo, M.; Tarantino, C.; Panitsa, M.; Terzi, M.; Veronico, V.; Lovergine, F.; Nagendra, H.; Lucas, R.; Mairota, P.; Mucher, C.A.; Blonda, P. (2013). Translating land cover/land use classifications to habitat taxonomies for landscape monitoring: a Mediterranean assessment, *Landscape Ecology*, 20, 905-930, doi: 10.1007/s10980-013-98633

Trimble. ECognition Developer 9.5. Reference Book. (2019). <https://geospatial.trimble.com/en/products/software/trimble-ecognition>