

	<i>Project:</i> <i>APEMAIA</i> Contract n. 2023-39-HH.0	Doc. N.: DEL_012_ARPA_CNR- ITIAm_RF_EN-001 Issue/Rev.:02/01 Date: 25/09/2026 (T0+24) Pages: 27
---	---	---

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

WP4103 – Output products validation

DEL012 – Validation report

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-ITIAm)	DATE SUBMITTED	N/A
AUTHOR	Morabito A., Tanzarella A. (ARPA) Aquilino M., De Lucia M., M. Adamo (CNR-ITIAm)	CONTRACTOR'S REFERENCE	Issue:01 Rev.:01 Date:25/09/26
PROJECT BEGIN DATE	26/09/2023	PROJECT END DATE	25/09/2026



Project:
APEMAIA

Contract n.
2023-39-HH.0

Doc. N.: DEL_012_ARPA_CNR- ITIAm_RF_EN-001

Issue/Rev.:02/01

Date: 25/09/2026 (T0+24)

Pages: 27

EXECUTIVE SUMMARY

This document is the **D012** deliverable 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 activities scheduled for Work Package WP4103 – Output products validation. The release adheres to the planned schedule (T0+24M) in accordance to the timeline established in accordance with the ASI commitment.

The document represents the **D012 Deliverable** to be submitted in its final version at Final Review of Advancements (**RF**) of the project.

NAME OF ASI PROJECT
MANAGER

Name Surname (PM)

Name Surname (RUP)

Name Surname (Project Scientist)

STAR CODE

No OF VOLUMES:



Project:
APEMAIA

Contract n.
2023-39-HH.0

Doc. N.: DEL_012_ARPA_CNR- ITIAm_RF_EN-001

Issue/Rev.:02/01

Date: 25/09/2026 (T0+24)

Pages: 27

TABLE OF CONTENTS

1.	AIM OF THE DOCUMENT	5
1.1	Applicable and reference documents	5
1.2	Applicable documents	5
1.3	Acronyms and abbreviations.....	5
2.	Resume of the preliminary validation plan	7
3.	Validation.....	8
3.1	LOYO and LOSO Validation	8
3.2	Validation of the daily mean PM concentration time series with DELTA TOOL software ..	10
3.3	Validation of spatial PM distribution: a map-to-map comparison	20
3.3.1	Mean bias scatterplot	21

	<i>Project:</i> APEMAIA Contract n. 2023-39-HH.0	Doc. N.: DEL_012_ARPA_CNR- ITIAm_RF_EN-001 Issue/Rev.:02/01 Date: 25/09/2026 (T0+24) Pages: 27
---	---	---

1. AIM OF THE DOCUMENT

This document is the final version of a document released at preliminary stage at RA3 (T0+18) as a preliminary validation plan.

It includes a complete description of all the validation activities.

ARPA is tasked with this activity as it is the provider of in situ data measurements for PM10 and PM2.5. These data sets will serve the purpose of calibrating and validating Machine Learning (ML) models.

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

[AD 2] DEL_004 – Technical Report “User Requirement Document (preliminary version)”

[AD 3] DEL_008 – 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)

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
PM2.5	Particulate matter with a diameter of 2.5 micrometres or less
PM10	Particulate matter with a diameter of 10 micrometres or less
CV	Cross-Validation technique
MAE	Mean Absolute Error
RMSE	Root Mean Squared Error
R^2	R squared coefficient



Project:
APEMAIA

Contract n.
2023-39-HH.0

Doc. N.: DEL_012_ARPA_CNR- ITIAm_RF_EN-001

Issue/Rev.:02/01

Date: 25/09/2026 (T0+24)

Pages: 27

	<i>Project:</i> <i>APEMAIA</i> Contract n. 2023-39-HH.0	Doc. N.: DEL_012_ARPA_CNR- ITIAm_RF_EN-001 Issue/Rev.:02/01 Date: 25/09/2026 (T0+24) Pages: 27
---	---	---

2. Resume of the preliminary validation plan

Below are some preliminary considerations concerning the validation plan:

1. It should be noted that of all the air quality monitoring stations distributed over the sites of the cities of interest in the cities of Taranto and Bari, not all of them provide a measurement of PM_{2.5} concentration. Therefore, the quality of the PM_{2.5} products that will be extracted during the project may suffer from an insufficient number of in-situ reference measurements for the effective calibration of the algorithm.
2. Most ARPA Puglia sensors are configured for 24-hour PM data acquisition, with only the daily average value being reported. A limited number of stations provides the hourly or bi-hourly measurements. It is assumed that the daily average PM value is taken as the reference for all stations. To standardize measurements across all stations, ARPA Puglia will compute daily averages from the hourly and bi-hourly data, employing statistical significance criteria and specific tools.
3. The validation plan will involve the use of the k-fold Cross Validation. A single iteration of cross-validation entails dividing a data sample into complementary subsets, conducting the analysis on one subset (referred to as the training set), and validating the analysis on the other subset (referred to as the validation set or testing set). To minimize variability, many methods involve multiple (k) rounds of cross-validation with different partitions, and the validation results are aggregated (e.g., averaged) across these rounds to provide an estimate of the model's predictive performance. A particular case is the leave-one-group-out (LOGO) cross-validation, where each time one sample is used as a test set while the rest are used to train the ML model. In our case we will perform the Leave-One-station-Out (LOSO) and the Leave-One-Year Out (LOYO) with the aim of assessing the performances of the model in generalizing different spatial and temporal conditions.
4. The validation of the annual time series of daily mean PM concentrations extracted from satellite data in correspondence of the measurement stations will be carried out using the DELTA TOOL software.
5. A qualitative 'map to map' comparison will be made with the maps produced by ARPA's PM modelling system to verify the capability of the implemented system to identify specific patterns of PM accumulation/dispersion

	<i>Project:</i> <i>APEMAIA</i> Contract n. 2023-39-HH.0	Doc. N.: DEL_012_ARPA_CNR- ITIAm_RF_EN-001 Issue/Rev.:02/01 Date: 25/09/2026 (T0+24) Pages: 27
---	---	--

3. Validation

The performance of the developed PM concentration products was assessed through three complementary validation approaches, designed to evaluate the robustness of the results under different spatial and temporal conditions. First, a cross-validation strategy based on Leave-One-Station-Out (LOSO) and Leave-One-Year-Out (LOYO) approaches were adopted to assess the ability of the Machine Learning models to generalize across different monitoring stations and years. Second, the temporal consistency of the daily mean PM concentration time series derived from satellite data was evaluated against ground-based measurements at the monitoring stations using the DELTA TOOL software. Finally, a quantitative/qualitative map-to-map comparison was performed with the PM concentration fields generated by the ARPA modelling system, in order to assess the capability of the developed approach to reproduce the main spatial patterns of PM accumulation and dispersion. Together, these three approaches provide complementary information on the predictive performance, temporal behavior and spatial consistency of the estimated PM concentrations.

3.1 LOYO and LOSO Validation

To assess the generalisation capability of the Machine Learning models under independent spatial and temporal conditions, two cross-validation strategies were applied: Leave-One-Station-Out (LOSO) and Leave-One-Year-Out (LOYO). These approaches were adopted to complement the conventional k-fold cross-validation, providing a more stringent assessment of the model transferability in space and time.

In the LOSO validation, the model was trained using observations from all but one monitoring station and subsequently tested on the excluded station. The procedure was repeated iteratively until each monitoring station had been used as the independent test set. This approach evaluates the capability of the model to predict PM concentrations at locations that were not represented during the training phase, thereby providing an assessment of its spatial generalisation and its potential applicability to areas not covered by monitoring stations.

The LOYO validation followed an analogous procedure in the temporal domain. For each iteration, observations from all but one year were used to train the model, while the excluded year was used exclusively for testing. The procedure was repeated for each year in the dataset. This strategy was

	<i>Project:</i> <i>APEMAIA</i> Contract n. 2023-39-HH.0	Doc. N.: DEL_012_ARPA_CNR- ITIAm_RF_EN-001 Issue/Rev.:02/01 Date: 25/09/2026 (T0+24) Pages: 27
---	---	---

designed to assess the temporal transferability of the models and their ability to reproduce PM concentrations under interannual conditions that were not included in the training data.

Both validation strategies were applied to the XGBoost models selected for PM_{2.5} and PM₁₀ prediction in the Bari and Taranto study areas. Model performance was evaluated using the coefficient of determination (R^2), Root Mean Square Error (RMSE) and Mean Squared Error (MSE). The comparison between LOSO and LOYO results provides complementary information on the spatial and temporal robustness of the models, highlighting the extent to which their predictive performance is maintained when applied to previously unseen monitoring locations or years.

The comparison between the LOSO and LOYO validation results provides useful insights into the spatial and temporal generalisation capability of the XGBoost models. In Bari, the LOSO validation showed relatively robust performance for both PM₁₀ and PM_{2.5}, with R^2 values of 0.71 and 0.70 and RMSE values of 5.64 and 3.44 $\mu\text{g}/\text{m}^3$, respectively. These results are comparable to those obtained with k-fold cross-validation and indicate that the models retain a substantial part of their predictive capability when applied to monitoring stations excluded from the training dataset. This suggests a relatively good spatial transferability within the Bari study area.

The LOYO validation resulted in lower performance, with R^2 values of 0.34 for PM₁₀ and approximately 0.44–0.46 for PM_{2.5}, and RMSE values of 8.76 and approximately 4.8–4.9 $\mu\text{g}/\text{m}^3$, respectively. The reduction in performance compared with LOSO indicates that temporal transferability represents a greater challenge for the models. In particular, the results suggest that interannual changes in meteorological conditions, atmospheric circulation and emission patterns may introduce variability that is not fully represented when the model is trained on the remaining years.

A different behaviour was observed in Taranto. The available results for PM_{2.5} show a k-fold R^2 of 0.67, while the LOYO validation resulted in an R^2 of 0.40, confirming the reduction in predictive performance when the model is applied to an independent year. The LOSO validation produced a substantially lower mean R^2 of 0.08, together with a high standard deviation (0.66), indicating marked differences in model performance among the monitoring stations. The station-level analysis showed that some stations were reasonably reproduced, whereas the model performed poorly at stations characterised by local conditions that were not adequately represented by the remaining monitoring network. This effect is particularly evident at the TAMBURI station, which is strongly influenced by the nearby industrial area.

	<i>Project:</i> <i>APEMAIA</i> Contract n. 2023-39-HH.0	Doc. N.: DEL_012_ARPA_CNR- ITIAm_RF_EN-001 Issue/Rev.:02/01 Date: 25/09/2026 (T0+24) Pages: 27
---	---	--

Overall, the validation results indicate that the models show different degrees of spatial and temporal transferability depending on the study area and PM fraction. LOSO validation provides evidence of reasonable spatial generalisation in Bari, while the lower LOYO performance highlights the challenge of reproducing interannual variability. In Taranto, the lower LOSO performance for PM_{2.5} emphasises the importance of local emission conditions and spatial heterogeneity, particularly in areas affected by strong industrial sources. These results underline the importance of evaluating both spatial and temporal generalisation when applying Machine Learning models to intra-urban PM estimation.

3.2 Validation of the daily mean PM concentration time series with DELTA TOOL software

The evaluation of air quality model performance is an essential step in determining whether a modelling system is suitable for its intended application, according to the *fitness-for-purpose* principle, and in ensuring the reliability of the results used to support air quality assessment and management activities. Conventional statistical indicators, such as root mean square error, bias, correlation coefficient and standard deviation, describe different aspects of the agreement between modelled and observed concentrations. However, when considered individually, they do not establish whether model performance has reached a sufficient quality level for a specific policy or regulatory application (Thunis et al., 2012a; Janssen and Thunis, 2022). To harmonise and increase the transparency of model performance assessment across Europe, the Forum for Air Quality Modelling in Europe (FAIRMODE), coordinated by the Joint Research Centre (JRC) of the European Commission, developed a common quality assurance and quality control (QA/QC) methodology.

This methodology is implemented in the DELTA Tool, a diagnostic software package developed by the JRC, for the comparison of modelled and observed concentrations and the harmonised evaluation of air quality model performance (Thunis et al., 2012b). The DELTA Tool defines the Modelling Quality Objective (MQO) as the minimum level of accuracy that a model is required to achieve. The MQO is based on the principle of uncertainty equivalence: a model is considered fit for purpose when its predictive uncertainty is comparable to the inherent uncertainty of the measurement instruments. The DELTA Tool systematically compares modelled data with observations collected at air quality monitoring stations, providing a direct assessment of the acceptability of model performance through a set of indicators and graphical representations.

	<i>Project:</i> <i>APEMAIA</i> Contract n. 2023-39-HH.0	Doc. N.: DEL_012_ARPA_CNR- ITIAm_RF_EN-001 Issue/Rev.:02/01 Date: 25/09/2026 (T0+24) Pages: 27
---	---	--

The reference mathematical indicator is the Modelling Quality Indicator (MQI). The MQI is a dimensionless indicator expressing the overall discrepancy between modelled and observed concentrations, normalised by the uncertainty associated with the measurements.

For a time-series comprising N valid model–observation pairs, the MQI is defined as follows (Janssen and Thunis, 2022):

$$MQI = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^N (O_i - M_i)^2}}{\beta \sqrt{\frac{1}{N} \sum_{i=1}^N U(O_i)^2}} = \frac{RMSE}{\beta RMS_U}$$

where O_i is the observed concentration at time i ; M_i is the corresponding modelled concentration; N is the number of valid model–observation pairs; $RMSE$ is the root mean square error between modelled and observed concentrations; $U(O_i)$ is the expanded measurement uncertainty associated with the observed concentration O_i , corresponding to a confidence level of approximately 95%; RMS_U is the root mean square of the measurement uncertainties; β is a proportionality coefficient set to 2 within the FAIRMODE methodology.

The expanded uncertainty associated with each observed concentration is concentration-dependent and is calculated as follows (Pernigotti et al., 2013; Janssen and Thunis, 2022):

$$U(O_i) = U_r(RV) \sqrt{(1 - \alpha^2) O_i^2 + \alpha^2 RV^2}$$

where RV is the pollutant-specific reference concentration, $U_r(RV)$ is the relative expanded uncertainty at the reference concentration, α represents the non-proportional fraction of measurement uncertainty.

The parameters adopted within the FAIRMODE methodology for PM_{10} and $PM_{2.5}$ are reported in the following table (Janssen and Thunis, 2022):

Table 3-1- List of parameters used to calculate the uncertainty

Pollutant	$U_r(RV)$	RV	α
PM_{10}	0.28	$50 \mu\text{g m}^{-3}$	0.13
$PM_{2.5}$	0.36	$25 \mu\text{g m}^{-3}$	0.30



Project:
APEMAIA

Contract n.
2023-39-HH.0

Doc. N.: DEL_012_ARPA_CNR- ITIAm_RF_EN-001

Issue/Rev.:02/01

Date: 25/09/2026 (T0+24)

Pages: 27

Bari – Cavour	Xutm= 657.2 Yutm=4554.02	PM10, PM2.5	Traffic-Urban
Bari – CUS	Xutm= 654.88 Yutm=4555.35	PM10	Traffic-Suburban
Bari - Kennedy	Xutm= 656.11 Yutm=4551.48	PM10	Background-Urban
Bitonto – EN1	Xutm= 646.61 Yutm=4549.01	PM10, PM2.5	Industrial-Suburban
Modugno – EN2	Xutm= 648.31 Yutm=4555.52	PM10, PM2.5	Industrial-Suburban
Modugno- EN3	Xutm= 649.65 Yutm=4549.97	PM10	Industrial-Urban
Modugno – EN4	Xutm= 650.12 Yutm=4553.06	PM10	Industrial-Suburban



Figure 3-2: Location of the ARPA Puglia regional air quality monitoring network stations in the Bari area.

Table 3 3-Coordinates and description of the ARPA Puglia regional air quality monitoring network stations in theTarantoi area

Monitoring stations	WGS84/UTM zone 33N (km)	Pollutants	Type
Statte - via delle Sorgenti	Xutm= 686.54 Yutm=4492.53	PM10	Industrial-Suburban
Taranto – San Vito	Xutm=688.79 Yutm=4477.12	PM10	Background-Suburban
Taranto – Talsano	Xutm= 693.78 Yutm=4475.97	PM10	Background-Suburban
Statte – Ponte Wind	Xutm= 684.12 Yutm=4488.43	PM10	Industrial-Rural
Taranto – Paolo VI	Xutm= 690.90 Yutm=4488.03	PM10, PM2.5	Industrial-Rural
Taranto – via Archimede	Xutm= 689.24 Yutm=4485.04	PM10, PM2.5	Industrial-Suburban
Taranto – via Machiavelli	Xutm= 688.65 Yutm=4484.36	PM10, PM2.5	Industrial-Suburban
Taranto – via Alto Adige	Xutm= 691.92 Yutm=4481.35	PM10, PM2.5	Traffic-Urban
Taranto – Orsini (Tamburi)	Xutm= 688.64 Yutm=4485.01	PM10, PM2.5	Industrial-Suburban

	<i>Project:</i> <i>APEMAIA</i> Contract n. 2023-39-HH.0	Doc. N.: DEL_012_ARPA_CNR- ITIAm_RF_EN-001 Issue/Rev.:02/01 Date: 25/09/2026 (T0+24) Pages: 27
---	---	---

The plots generated by the software, summarising the model performance indicators, are presented below: the Assessment Target Plot and the Scatter Plot.

The Assessment Target Plot provides a synthetic graphical representation of model performance at the selected monitoring stations. Each symbol represents one monitoring station, while its position describes the magnitude and main components of the discrepancy between modelled and observed concentrations. The distance of each point from the origin corresponds to the Modelling Quality Indicator (MQI). Therefore, points located closer to the origin indicate better agreement between modelled and observed concentrations.

The value displayed in the upper-left corner MQI_{HD} (daily Modelling Quality Indicator) represents the 90th-percentile MQI across the selected monitoring stations and is calculated from the daily mean concentrations of paired modelled and observed concentrations. MQI_{YR} (yearly Modelling Quality Indicator) evaluates the agreement between modelled and observed annual mean concentrations. This value also refers to the 90th percentile across the monitoring stations. For both indicators, the Modelling Quality Objective is fulfilled when $MQI \leq 1$.

The vertical axis represents the uncertainty-normalised BIAS, while the horizontal coordinate represents the uncertainty-normalised centred root mean square error (CRMSE). Since CRMSE is always non-negative, the two sides of the horizontal axis are used to indicate which component predominantly contributes to the unsystematic error:

- R, on the left-hand side, indicates that the error is mainly related to the temporal correlation between modelled and observed concentrations;
- SD, on the right-hand side, indicates that the error is mainly related to differences between the modelled and observed standard deviations.

Finally, the values reported in the upper-right corner of the diagram represent the coefficients used in the calculations, as listed in the table x. The value ST U_{mod} (RV) is the estimated short-term model uncertainty at the reference concentration, derived from the comparison between the RMSE of the daily series and the corresponding measurement uncertainty. LT U_{mod} (RV) is the estimated long-term model uncertainty at the reference concentration, based on the difference between modelled and observed annual mean concentrations.

At the Bari monitoring sites, the results show that the MQO criterion was fulfilled for both PM₁₀ and PM_{2.5} at both the daily and annual time scales. For PM₁₀, a positive bias was observed at all monitoring stations, indicating an overall tendency of the model to overestimate the observed concentrations. For PM_{2.5}, ST U_{mod} and LT U_{mod} values of 0% indicate that the model–observation



Project:
APEMAIA

Contract n.
2023-39-HH.0

Doc. N.: DEL_012_ARPA_CNR- ITIAm_RF_EN-001

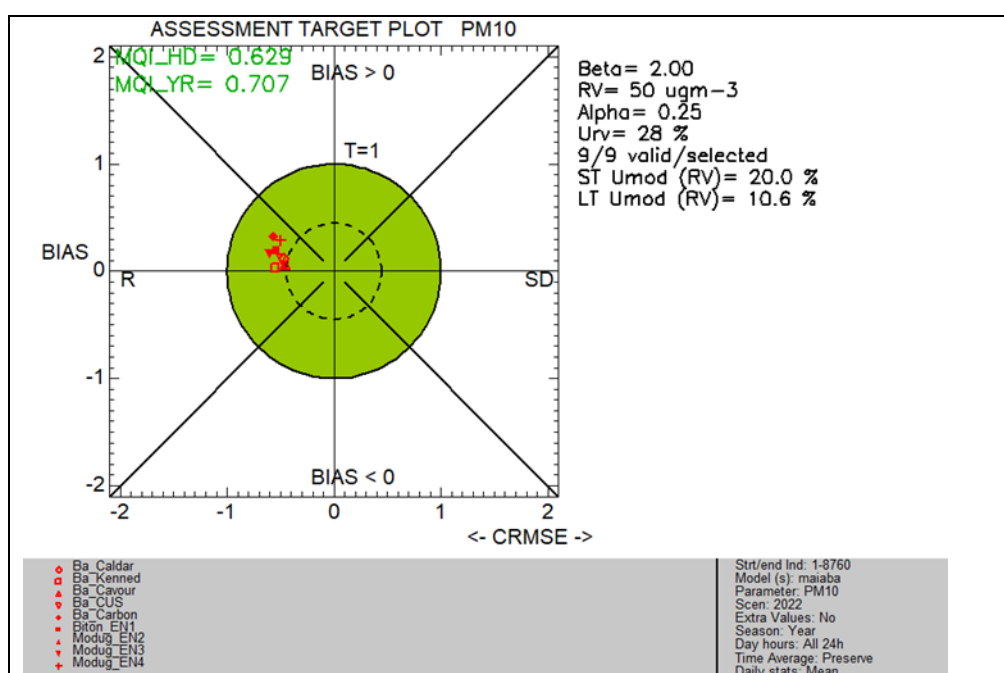
Issue/Rev.:02/01

Date: 25/09/2026 (T0+24)

Pages: 27

discrepancies were within the measurement uncertainty and that no additional model uncertainty distinguishable from the observational uncertainty could be identified.

The MQO criterion was also fully met at the monitoring stations in the Taranto area for both PM₁₀ and PM_{2.5}. The PM_{2.5} ST Umod value reported as “unknown” indicates that the short-term model uncertainty could not be meaningfully quantified, since the differences between modelled and observed daily concentrations were comparable to, or smaller than, the measurement uncertainty. This result therefore does not indicate poor model performance.





Project:
APEMAIA

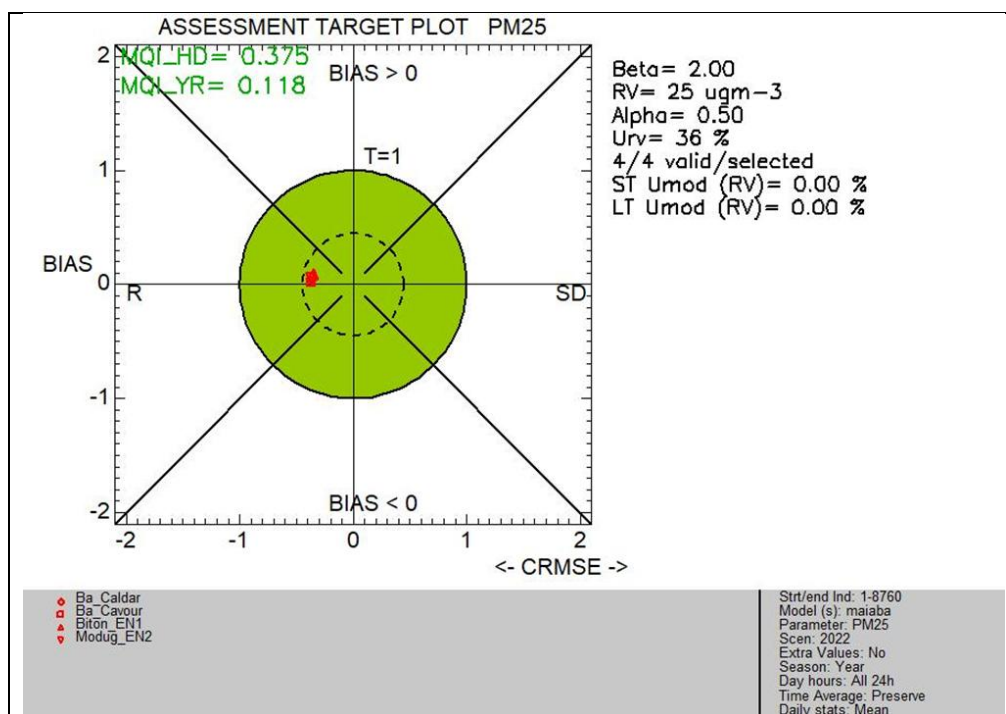
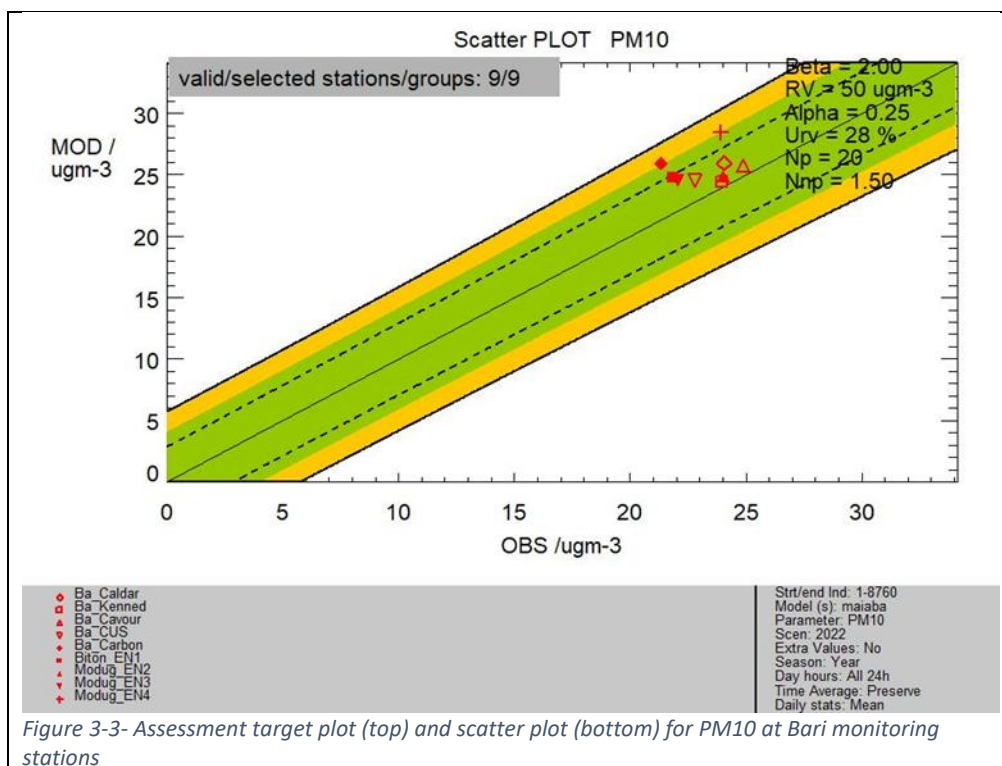
Contract n.
2023-39-HH.0

Doc. N.: DEL_012_ARPA_CNR- ITIAm_RF_EN-001

Issue/Rev.:02/01

Date: 25/09/2026 (T0+24)

Pages: 27





Project:
APEMAIA

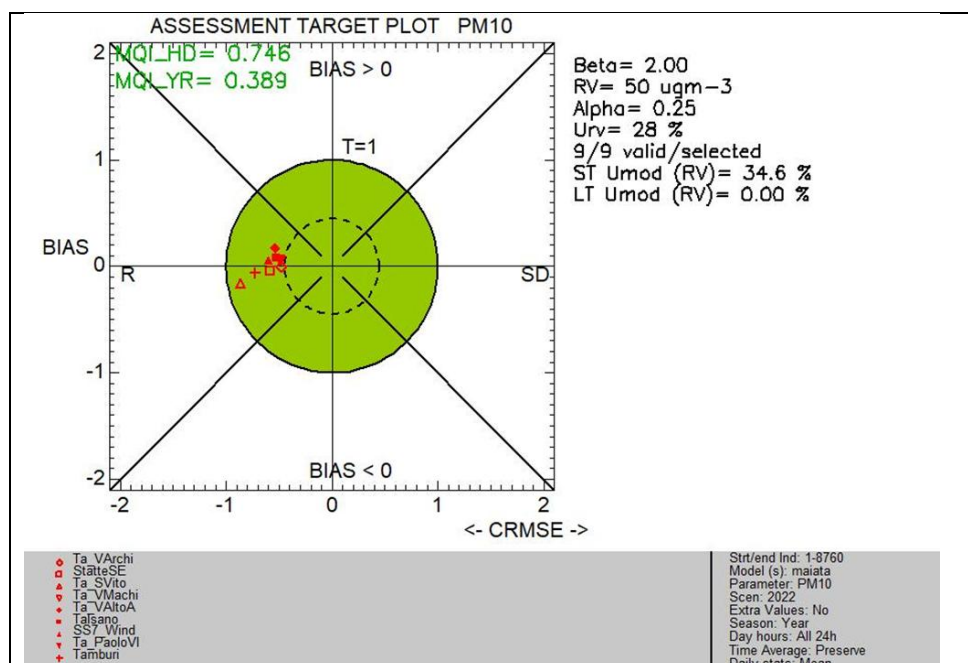
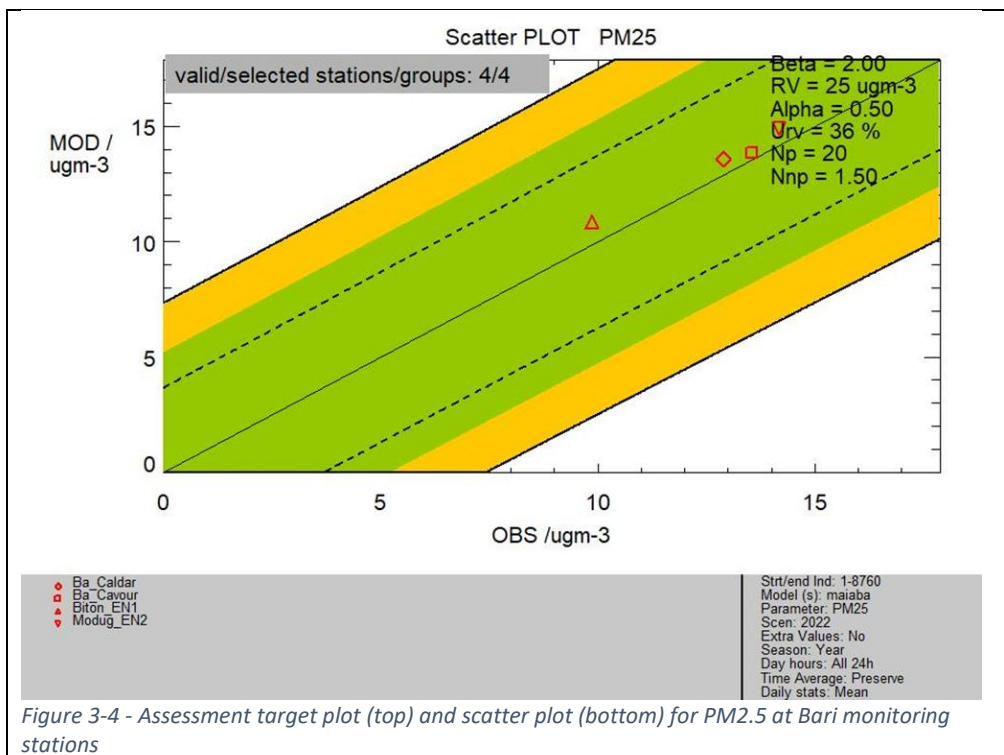
Contract n.
2023-39-HH.0

Doc. N.: DEL_012_ARPA_CNR- ITIAm_RF_EN-001

Issue/Rev.:02/01

Date: 25/09/2026 (T0+24)

Pages: 27





Project:
APEMAIA

Contract n.
2023-39-HH.0

Doc. N.: DEL_012_ARPA_CNR- ITIAm_RF_EN-001

Issue/Rev.:02/01

Date: 25/09/2026 (T0+24)

Pages: 27

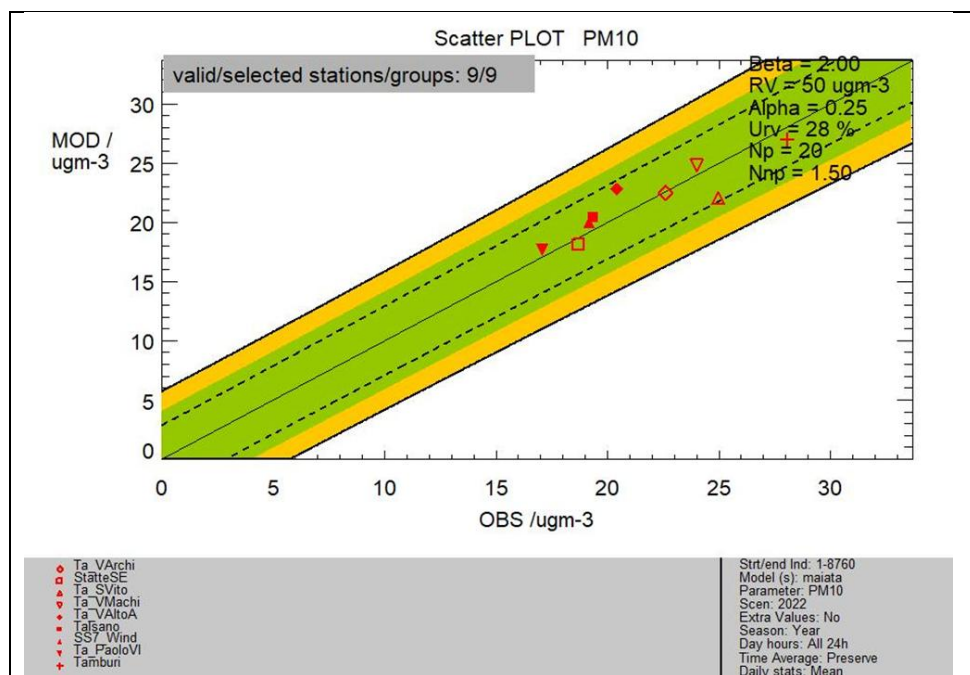
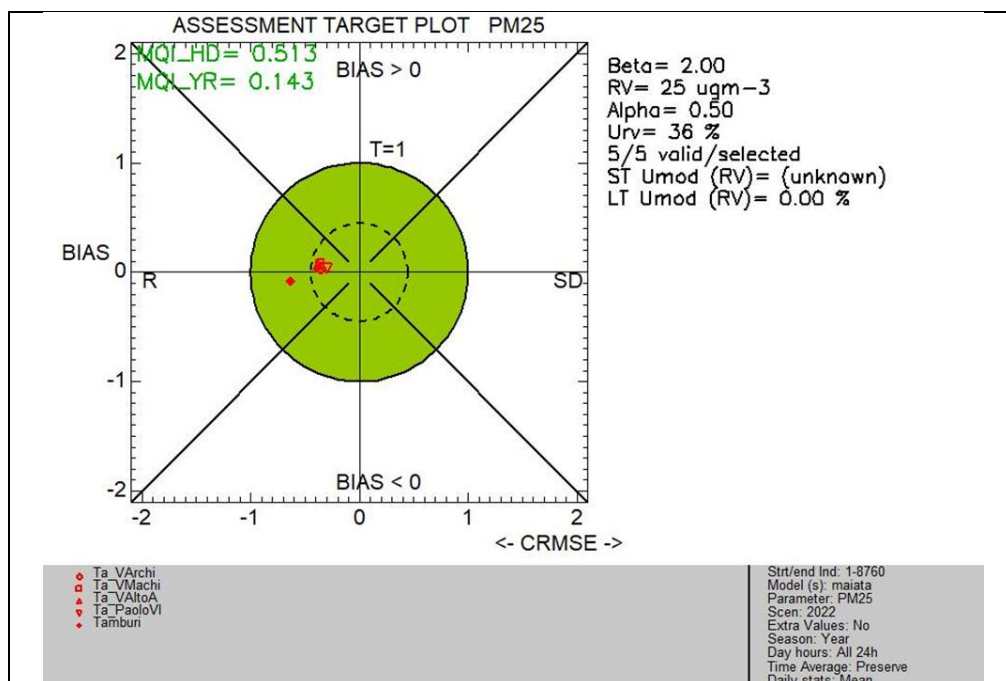


Figure 3-5-Assessment target plot (top) and scatter plot (bottom) for PM2.5 at Bari monitoring stations





Project:
APEMAIA

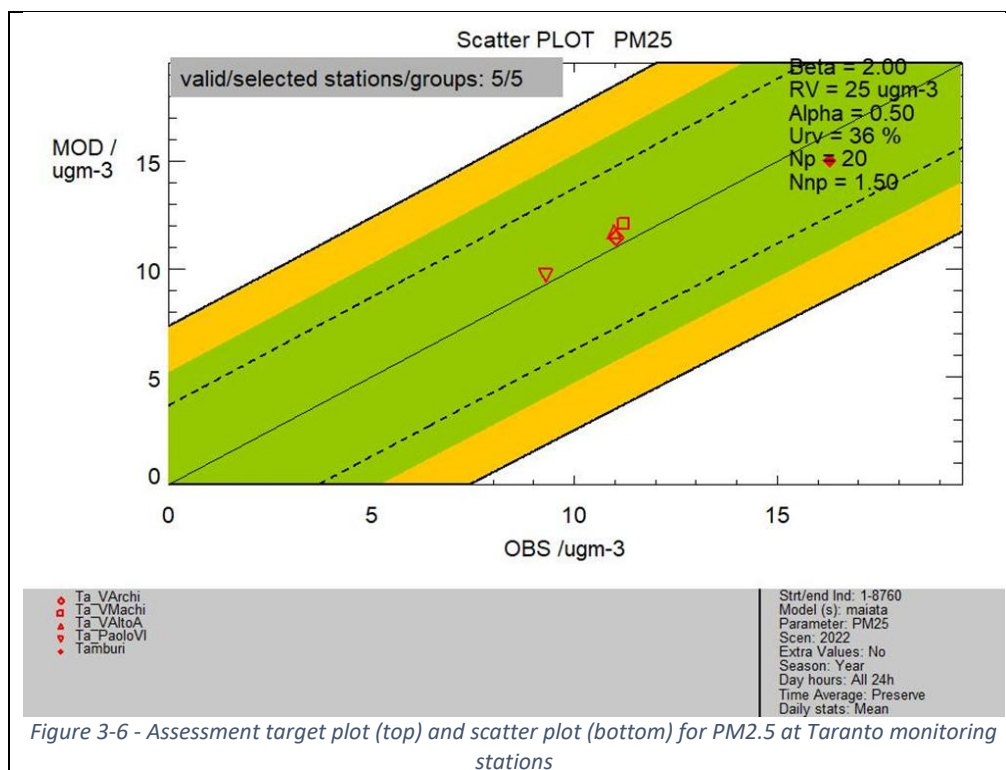
Contract n.
2023-39-HH.0

Doc. N.: DEL_012_ARPA_CNR- ITIAm_RF_EN-001

Issue/Rev.:02/01

Date: 25/09/2026 (T0+24)

Pages: 27



Thunis, P., Pederzoli, A. and Pernigotti, D. (2012a). Performance criteria to evaluate air quality modelling applications. *Atmospheric Environment*, 59, 476–482. <https://doi.org/10.1016/j.atmosenv.2012.05.043>.

Janssen, S. and Thunis, P. (eds.) (2022). FAIRMODE Guidance Document on Modelling Quality Objectives and Benchmarking. Version 3.3. EUR 31068 EN, Publications Office of the European Union, Luxembourg, JRC129254. <https://doi.org/10.2760/41988>.

Thunis, P., Georgieva, E. and Pederzoli, A. (2012b). A tool to evaluate air quality model performances in regulatory applications. *Environmental Modelling & Software*, 38, 220–230. <https://doi.org/10.1016/j.envsoft.2012.06.005>.

Pernigotti, D., Thunis, P. and Gerboles, M. (2013). Model quality objectives based on measurement uncertainty. Part II: PM10 and NO₂. *Atmospheric Environment*, 79, 869–878. <https://doi.org/10.1016/j.atmosenv.2013.07.012>.

	<i>Project:</i> <i>APEMAIA</i> Contract n. 2023-39-HH.0	Doc. N.: DEL_012_ARPA_CNR- ITIAm_RF_EN-001 Issue/Rev.:02/01 Date: 25/09/2026 (T0+24) Pages: 27
---	---	--

3.3 Validation of spatial PM distribution: a map-to-map comparison

Spatial agreement between the high-resolution PM maps and the coarse-resolution reference products provided by the FARM model (SW012 delivered by ARPA Puglia) was assessed through a map-to-map comparison.

To enable a consistent comparison, the 300 m resolution PM outputs were spatially aggregated to match the grid of the coarse FARM reference raster, which has a resolution of 4 km for the 2019–2020 period and 1 km for 2021–2022. For each pollutant, site, and year, the aggregated product was generated on the coarse reference grid and comprises four descriptive layers and one diagnostic bias layer:

1. Layer 1 contains the mean concentration predicted by the 300 m model after aggregation within each coarse reference cell,
2. Layer 2 reports the within-cell standard deviation of the 300 m predictions, describing the sub-grid spatial variability that is not represented by the coarse product,
3. Layer 3 provides the number of valid 300 m pixels contributing to each aggregated estimate,
4. Layer 4 contains the original concentration value of the coarse reference map. Only coarse cells with at least 50% valid fine-resolution pixels were retained for comparison,
5. Layer 5 represents the signed cell-wise difference between the aggregated fine-resolution estimate and the corresponding coarse reference value:

$$Bias_i = PM_{MLmodel,fine\ aggregated,i} - PM_{ARPAmodel,coarse,i}$$

For the latter, positive values indicate that the fine-resolution product is higher than the coarse reference in a given cell, whereas negative values indicate lower estimates.



Project:
APEMAIA

Contract n.
2023-39-HH.0

Doc. N.: DEL_012_ARPA_CNR- ITIAm_RF_EN-001

Issue/Rev.:02/01

Date: 25/09/2026 (T0+24)

Pages: 27

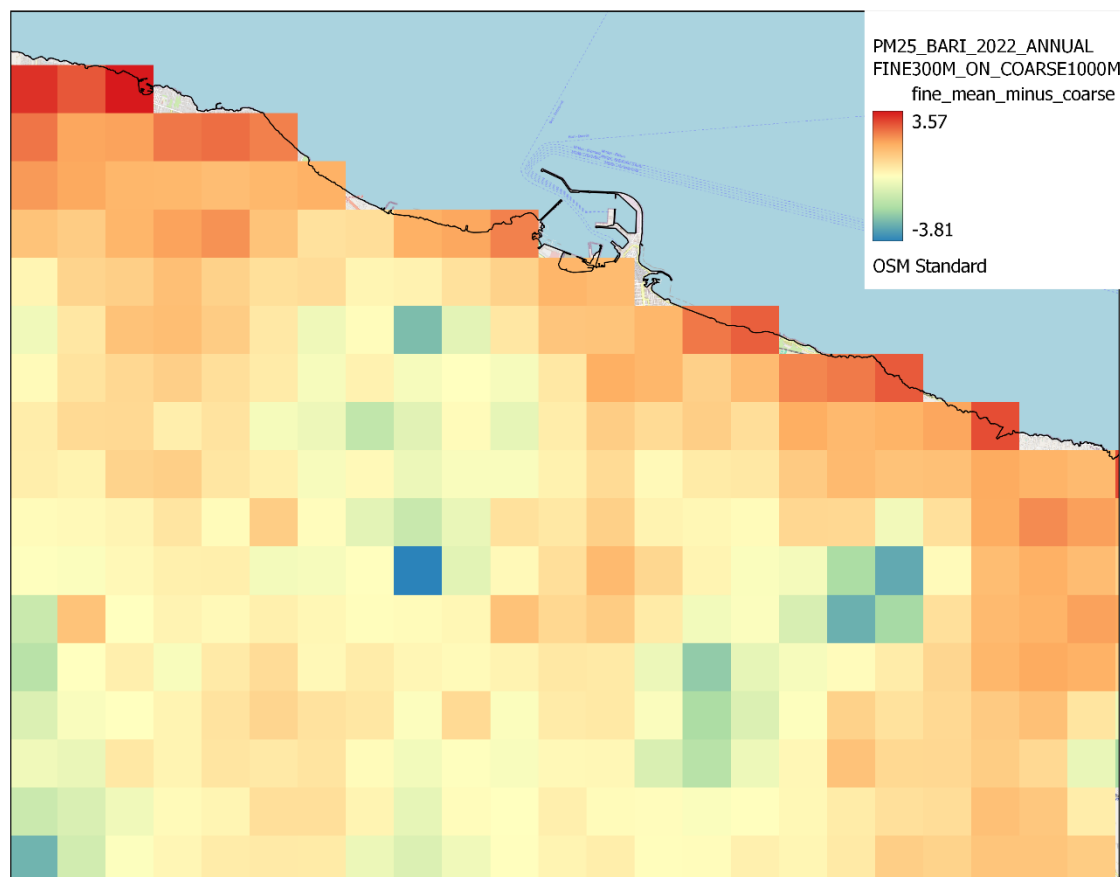


Figure 3-7: Spatial distribution of the cell-wise bias between the aggregated 300 m PM map and the corresponding coarse-resolution reference product. Positive values (red) identify coarse cells in which the fine-resolution model yields higher concentrations than the reference map, while negative values (blue) identify lower concentrations. The mean bias, calculated as the spatial average of valid cell-wise differences, summarizes the overall directional agreement between the two products.

3.3.1 Mean bias scatterplot

To complement the bias maps, scatterplots were produced by comparing Layer 4 (coarse reference concentration) against Layer 1 (mean fine-resolution concentration aggregated to the coarse grid). Each point represents one valid coarse cell in the overlapping domain. The 1:1 line denotes perfect agreement, while the fitted OLS regression line illustrates systematic departures from equality. Points consistently above the 1:1 line indicate a positive bias of the fine-resolution product relative to the coarse reference, whereas a wider spread around the line reflects lower spatial agreement. The regression equation and Pearson's correlation coefficient are reported in each plot.



Project:
APEMAIA

Contract n.
2023-39-HH.0

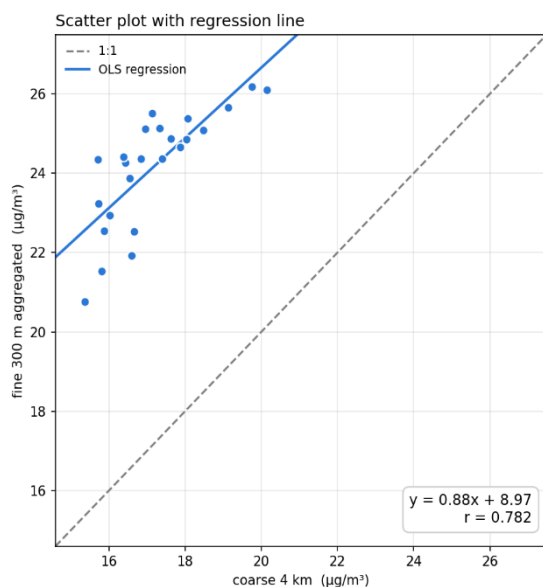
Doc. N.: DEL_012_ARPA_CNR- ITIAm_RF_EN-001

Issue/Rev.:02/01

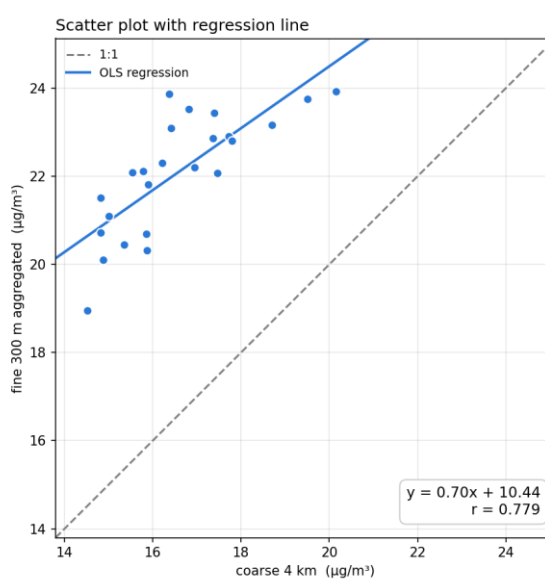
Date: 25/09/2026 (T0+24)

Pages: 27

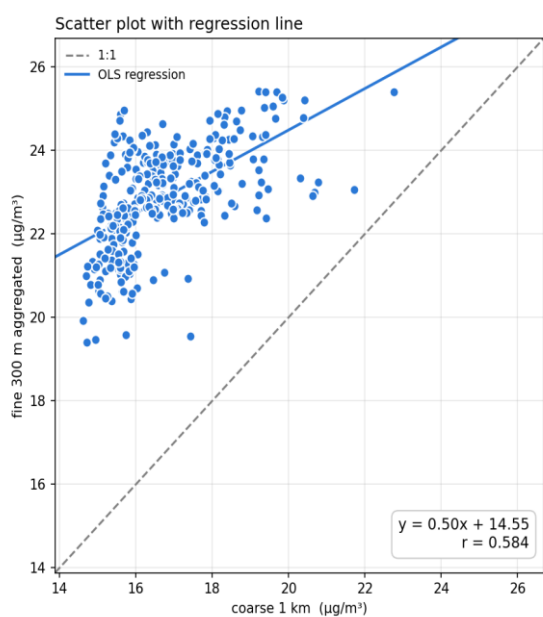
PM10 - 2019



PM10 - 2020



PM10 - 2021



PM10 - 2022

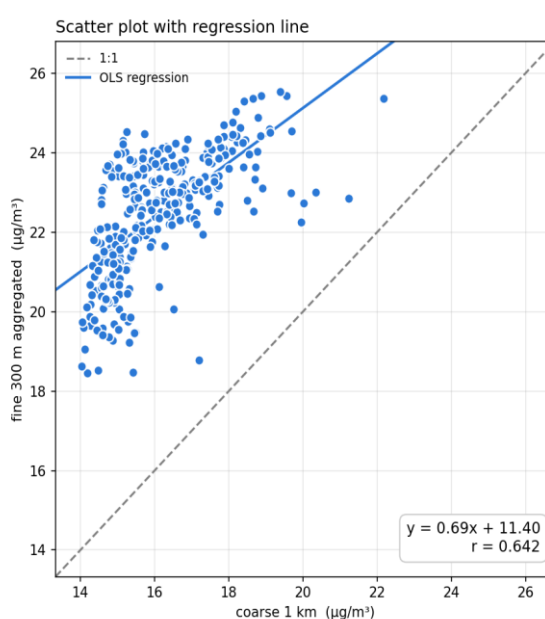


Figure 3-8: Scatterplots comparing annual PM_{10} concentrations from the 300 m model, aggregated to the reference grid, with the corresponding coarse-resolution PM_{10} maps for Bari in 2019–2022. Each point represents one valid overlapping coarse cell. The horizontal axis shows the coarse reference concentration (4 km in 2019–2020; 1 km in 2021–2022), while the vertical axis shows the mean PM_{10} concentration obtained by aggregating the 300 m predictions within the same cell.



Project:
APEMAIA

Contract n.
2023-39-HH.0

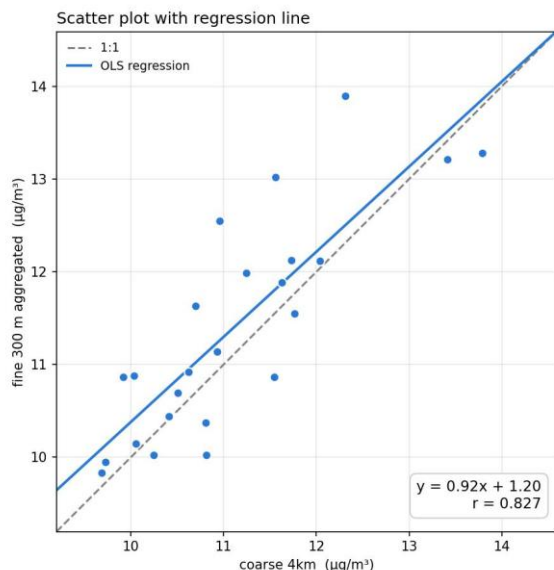
Doc. N.: DEL_012_ARPA_CNR- ITIAm_RF_EN-001

Issue/Rev.:02/01

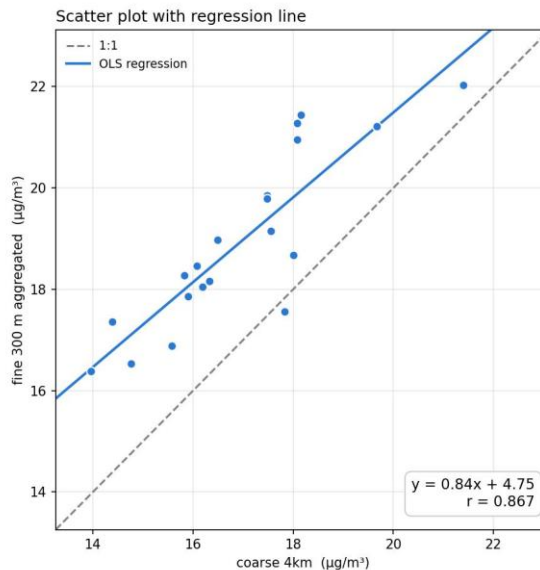
Date: 25/09/2026 (T0+24)

Pages: 27

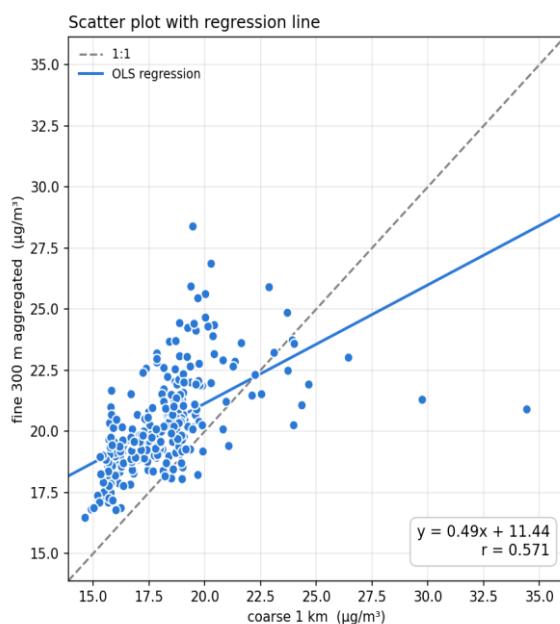
PM10 - 2019



PM10 - 2020



PM10 - 2021



PM10 - 2022

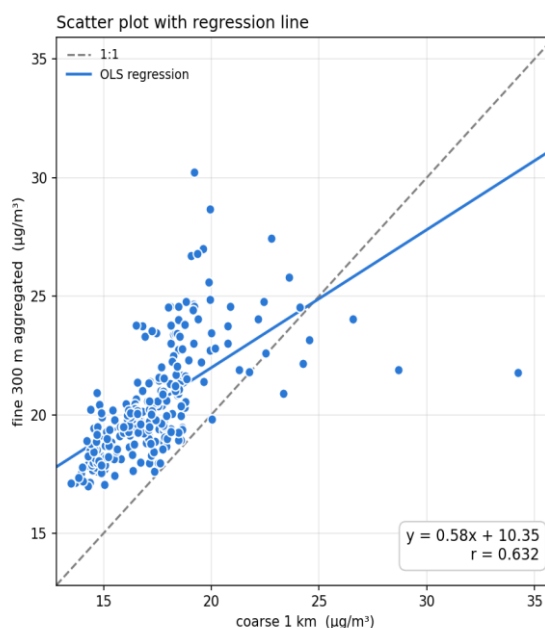


Figure 3-9: Scatterplots comparing annual PM_{10} concentrations from the 300 m model, aggregated to the reference grid, with the corresponding coarse-resolution PM_{10} maps for Taranto in 2019–2022. Each point represents one valid overlapping coarse cell. The horizontal axis shows the coarse reference concentration (4 km in 2019–2020; 1 km in 2021–2022), while the vertical axis shows the mean PM_{10} concentration obtained by aggregating the 300 m predictions within the same cell.



Project:
APEMAIA

Contract n.
2023-39-HH.0

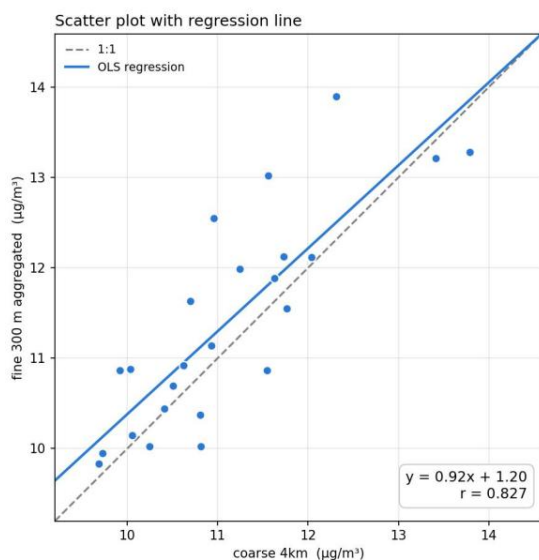
Doc. N.: DEL_012_ARPA_CNR- ITIAm_RF_EN-001

Issue/Rev.:02/01

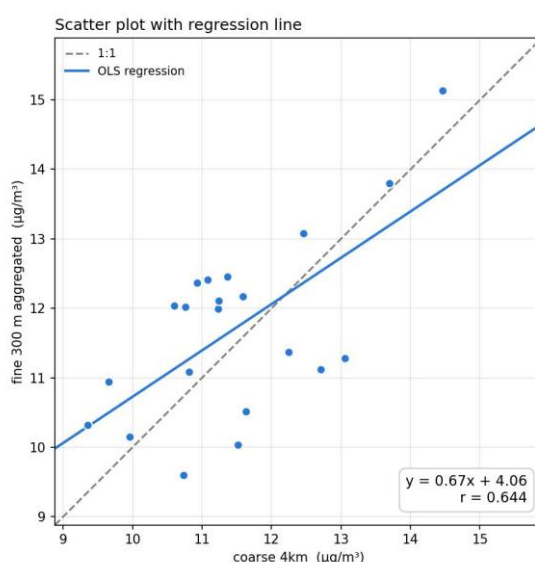
Date: 25/09/2026 (T0+24)

Pages: 27

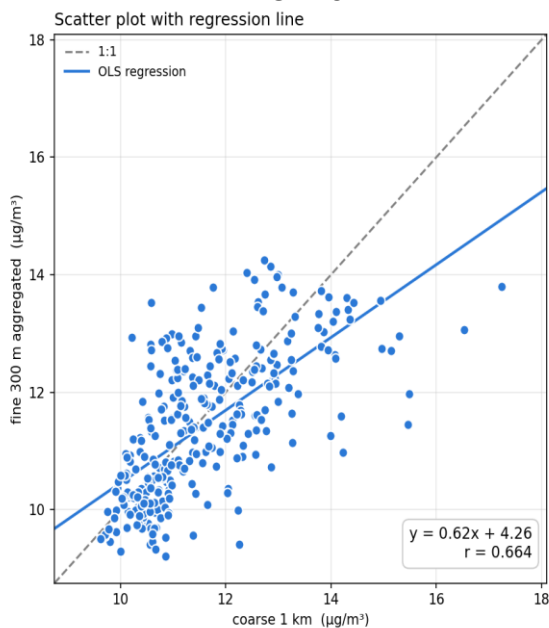
PM2.5 - 2019



PM2.5 - 2020



PM2.5 - 2021



PM2.5 - 2022

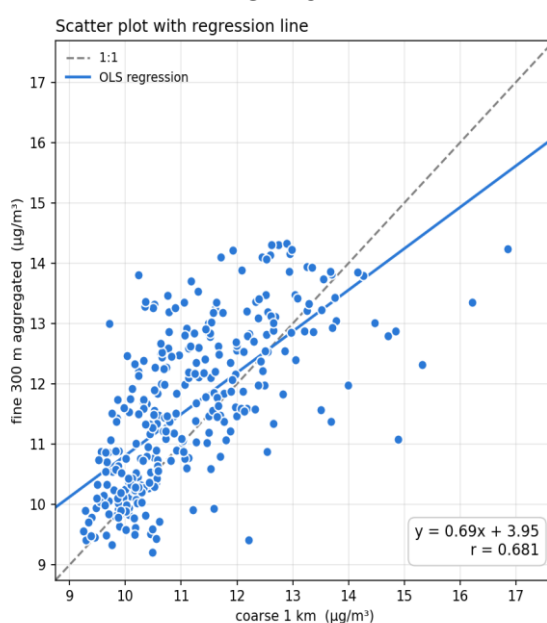


Figure 3-10: Scatterplots comparing annual PM_{25} concentrations from the 300 m model, aggregated to the reference grid, with the corresponding coarse-resolution PM_{25} maps for Bari in 2019–2022. Each point represents one valid overlapping coarse cell. The horizontal axis shows the coarse reference concentration (4 km in 2019–2020; 1 km in 2021–2022), while the vertical axis shows the mean PM_{25} concentration obtained by aggregating the 300 m predictions within the same cell.



Project:
APEMAIA

Contract n.
2023-39-HH.0

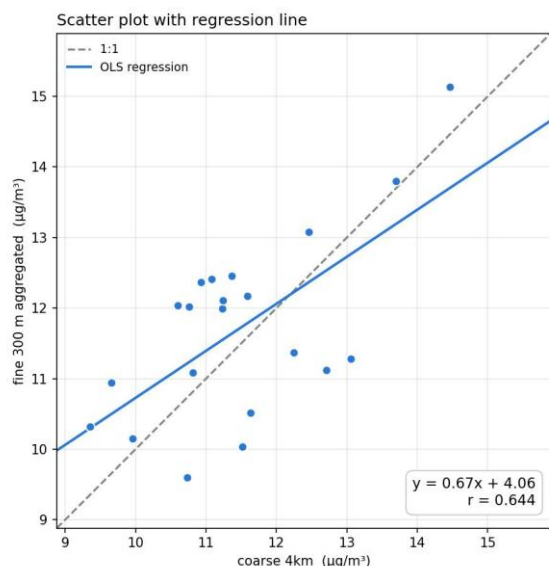
Doc. N.: DEL_012_ARPA_CNR- ITIAm_RF_EN-001

Issue/Rev.:02/01

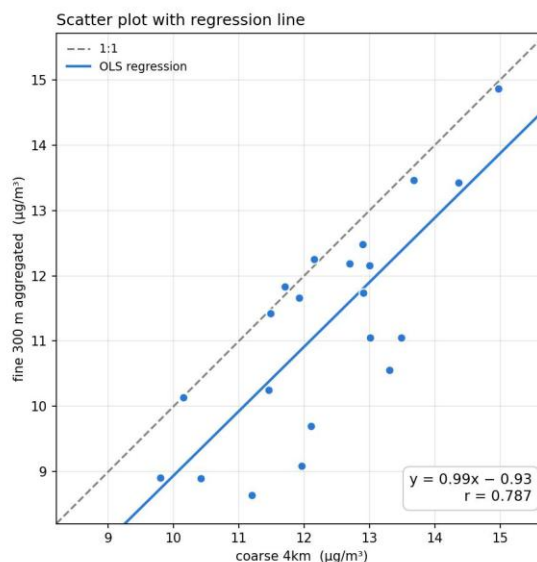
Date: 25/09/2026 (T0+24)

Pages: 27

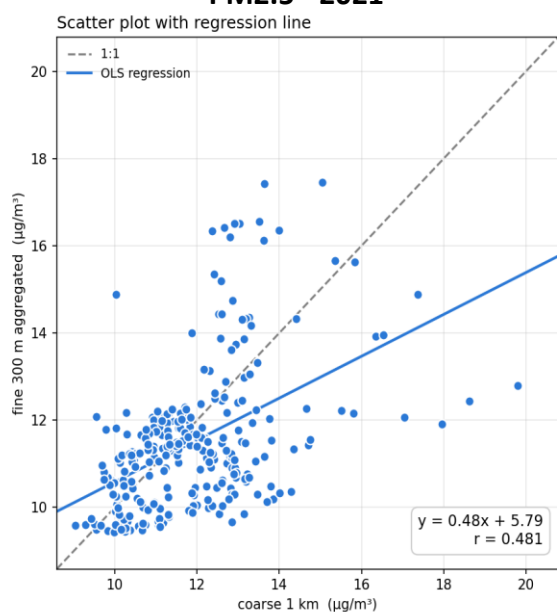
PM2.5 - 2019



PM2.5



PM2.5 - 2021



PM2.5 - 2022

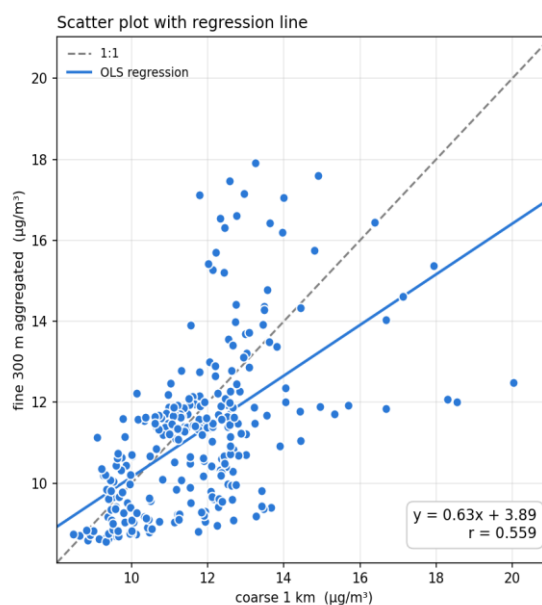


Figure 3-11: Scatterplots comparing annual PM_{25} concentrations from the 300 m model, aggregated to the reference grid, with the corresponding coarse-resolution PM_{25} maps for Taranto in 2019–2022. Each point represents one valid overlapping coarse cell. The horizontal axis shows the coarse reference concentration (4 km in 2019–2020; 1 km in 2021–2022), while the vertical axis shows the mean PM_{25} concentration obtained by aggregating the 300 m predictions within the same cell.



Project:
APEMAIA

Contract n.
2023-39-HH.0

Doc. N.: DEL_012_ARPA_CNR- ITIAm_RF_EN-001

Issue/Rev.:02/01

Date: 25/09/2026 (T0+24)

Pages: 27

The cell-level aggregated concentrations (Layer 1) were compared with the original coarse reference map (Layer 4) using three metrics. The **Mean Bias Error (MBE)** is the arithmetic mean of Layer 5 across all valid overlapping cells:

$$MBE = \frac{1}{N} (PM_{MLmodel, fine\ aggregated, i} - PM_{ARPAmodel, coarse, i})$$

and quantifies the overall directional offset between the two map products. The **Root Mean Square Error (RMSE)** measures the magnitude of the cell-wise differences. The **Mean Fractional Bias (MFB)** provides a dimensionless and symmetric measure of the relative systematic difference between the two products:

$$MFB = \frac{2}{N} \sum_i \frac{(PM_{MLmodel, fine\ aggregated, i} - PM_{ARPAmodel, coarse, i})}{(PM_{MLmodel, fine\ aggregated, i} + PM_{ARPAmodel, coarse, i})}$$

Summary results for all pollutant–site–year combinations are reported in **Errore. L'origine riferimento non è stata trovata.**

Table 3-3: Map-to-map comparison metrics assessing spatial consistency between the fine-resolution and coarse-resolution PM products.

Pollutant	Site	Year	Valid coarse cells	RMSE (µg/m³)	Mean Bias Error (MBE) (µg/m³)	Mean fractional bias (MFE) %
PM10	Bari	2019	24	7.044	6.987	33.88%
		2020	24	5.601	5.522	28.76%
		2021	311	6.268	6.151	31.27%
		2022	312	6.474	6.353	33.07%
	Taranto	2019	20	1.528	0.940	5.07%
		2020	20	2.172	1.984	11.17%
		2021	262	2.896	2.088	11.11%
		2022	262	3.736	3.132	16.92%



Project:
APEMAIA

Contract n.
2023-39-HH.0

Doc. N.: DEL_012_ARPA_CNR- ITIAm_RF_EN-001

Issue/Rev.:02/01

Date: 25/09/2026 (T0+24)

Pages: 27

PM2.5	Bari	2019	24	0.718	0.290	2.49%
		2020	24	0.829	-0.198	-1.75%
		2021	311	1.042	-0.136	-1.14%
		2022	312	1.141	0.458	4.02%
	Taranto	2019	21	1.093	0.234	2.01%
		2020	21	1.480	-1.090	-9.72%
		2021	270	1.708	-0.393	-3.30%
		2022	272	1.836	-0.465	-4.41%

PM₁₀ Bari shows the largest and most consistent bias: MBE ranges between 5.5 and 7.0 $\mu\text{g}/\text{m}^3$ and MFB between 29% and 34% across all four years, indicating a systematic overestimation of the fine-resolution product relative to the FARM reference in this area. PM₁₀ in Taranto exhibits considerably better spatial agreement, with MBE below 3.2 $\mu\text{g}/\text{m}^3$ and MFB not exceeding 17%, suggesting greater consistency between the two products. PM_{2.5} shows overall much closer agreement at both sites: RMSE does not exceed 1.84 $\mu\text{g}/\text{m}^3$ and MFB remains within $\pm 10\%$ in nearly all cases. The negative MBE values observed at Bari (2020–2021) and Taranto (2020–2022) indicate a slight underestimation by the fine-resolution product relative to the FARM reference.

In conclusion, PM_{2.5} shows very high spatial agreement, with RMSE ranging from 0.72 to 1.84 $\mu\text{g}/\text{m}^3$ and low bias across all site-year combinations. For PM₁₀, agreement is markedly better in Taranto, while Bari exhibits a persistent positive bias of approximately 5.5–7.0 $\mu\text{g}/\text{m}^3$.

This map-to-map comparison should not be interpreted as a formal validation, since it compares two independently produced modelling products rather than model estimates against ground-based observations. Instead, it was conducted as a complementary spatial-consistency assessment alongside the formal validation undertaken by ARPA. The observed differences warrant further investigation, particularly regarding the possible presence of systematic bias and the persistence of local discrepancies in specific grid cells. Future work should assess whether such patterns are associated with the location and representation of emission sources, the spatial distribution of monitoring stations, and differences in the underlying modelling frameworks.