With the contribution of the LIFE Programme of the European Union

LIFE 15 IPE IT 013

prepAIR

Po Regions Engaged to Policies of Air

ACTION D5.

Air Quality Assessment

05/06/2021

With the contribution of the LIFE Programme of the European Union

LIFE 15 IPE IT 013

Po Regions Engaged to Policies of Air

This document was produced within the project PREPAIR (LIFE 15 IPE IT 013) financed by the program LIFE 2014-2020 of the European Union.

Closing date of the report 05/06/2021

Beneficiary responsible for implementation ARPAE Emilia Romagna e ARPA Piemonte

Authors:

Stefano Bande (ARPA Piemonte), Michele Stortini, Roberta Amorati, Giulia Giovannini (ARPAE Emilia Romagna), Giovanni Bonafè (ARPA Friuli Venezia Giulia), Luka Matavz (ARSO, Slovenia), Elisabetta Angelino, Loris Colombo, Giuseppe Fossati, Alessandro Marongiu (ARPA Lombardia)

With the contribution
of the LIFE Programme
of the European Union
LIFE 15 IPE IT 013

prepAIR
Po Regions Engaged to Policies of Air

INDEX

1 INTRODUCTION 1
2 ASSESSMENT METHODOLOGY...2
2.1 DATA FUSION INPUT DATA 3
AIR QUALITY DATA...3
2.2 DATA FUSION TECHNIQUES 17
NINFA and Observations Data Fusion 17
2.1.1
FARM-PI and Observations Data Fusion 18
2.2.1
FARM-LO and Observations Data Fusion 19
2.2.2
CAMx-SLO and Observations Data Fusion 20
2.2.3
2.2.4
2.3 DATA FUSION VALIDATION 22
Cross validation 22
2.3.1
Comparison between observed and predicted values 27
2.3.2
3 ASSESSMENT RESULT...30
3.1 PM10 30
3.2 PM2.5 32
3.3 NO2 34
3.4 O3 36
3.5 ATTAINMENT STATUS/POPULATION EXPOSURE..37
4 DISCUSSION 45
5 GLOSSARY 48
6 BIBLIOGRAPHY...50
APPENDIX A: Air Quality Data...53

Index of Figures

Figure 1 - Spatial distribution of monitoring stations availables in C1 dataset. 3
Figure 2 - C1 dataset: data capture percentage for each pollutant and for each data supplier. 5
Figure 3 -. Emission maps for 2013 and 2017 representing PM10 (top) and N0x (bottom). 6
Figure 4 - NINFA model scheme 7

Figure 5 - NINFA PREPSLO domain nested to ITA7 domain. The area covered by region/country project partners is shown in red. The inner EMR3 domain is also shown....7
Figure 6 -. FARM-PI computational system architecture....9
Figure 7 - FARM-PI computational domains....10
Figure 8.- PREPAIR domain of ARPA-LO modelling system....14
Figure 9.- The architecture of ARPA-LO modelling system....14
Figure 10 - Model domain of ALADIN/SI and CAMx model....16
Figure 11– Input data for CAMx model....16
Figure 12.- PM10 annual mean: cross validation scatter plot for FARM-PI (top) and NINFA (bottom). The dotted lines represent the admitted relative uncertainty (50% for PM10 annual mean), while the red lines indicate the EU limit value (40 µg/m3)....23
Figure 13.- PM10 percentile 90.41: cross validation scatter plot for FARM-PI (top) and NINFA (bottom). The dotted lines represent the 50% relative uncertainty, while the red lines indicate the EU limit value (50 µg/m3)....23
Figure 14 - PM2.5 annual mean: cross validation scatter plot for FARM-PI (top) and NINFA (bottom). The dotted lines represent the admitted relative uncertainty (50% for PM2.5 annual mean), while the red lines indicate the EU limit value (25 µg/m3)....24
Figure 15 -. NO2 annual mean: cross validation scatter plot for FARM-PI (top) and NINFA (bottom). The dotted lines represent the admitted relative uncertainty (30% for NO2 annual mean), while the red lines indicate the EU limit value (40 µg/m3)....25
Figure 16 - O3 percentile 93.1: cross validation scatter plot for FARM-PI (top) and NINFA (bottom). The dotted lines represent the admitted relative uncertainty (50% for O3 93.1 percentile), while the red lines indicate the EU target value (120 µg/m3)....26
Figure 17 -. Left: PM10 annual mean, simulated vs observed values for the four data fusion systems. Right: PM10 90.41 percentile, simulated vs observed values for the NINFA, FARM-PI, FARM-LO data fusion systems. . The dotted lines represent the admitted relative uncertainty, while the red lines indicate the EU limit value (40 and 50µg/m3 respectively)....28
Figure 18 - Left: PM2.5 annual mean, simulated vs observed values for the four data fusion systems. Right: NO2 annual mean, simulated vs observed values for the four data fusion systems. The dotted lines represent the admitted relative uncertainty, while the red lines indicate the EU limit value (25 and 40 µg/m3 for PM2.5 and NO2 respectively)....29

Figure 19 -. O3 93.10 percentile, simulated vs observed values for the NINFA, FARM-PI, FARM-LO data fusion systems. The dotted lines represent the admitted relative uncertainty, while the red lines indicate the EU target value (120 µg/m^3).....29
Figure 20.- Maps of PM10 annual mean produced by the four data fusion systems.....30
Figure 21-. Maps of PM10 90.41 percentile produced by NINFA, FARM-PI and FARM-LO systems. 31
Figure 22 - Boxplots of grid point concentration distributions grouped by model and region. Left: PM10 annual mean; right percentile 90.4 of PM10 daily values. The red lines indicate the EU limit value (40 and 50µg/m3 respectively.....31
Figure 23 - Maps of PM2.5 annual mean produced by the four data fusion systems.....32
Figure 24 - PM2.5, annual mean: boxplots of grid point concentration distributions grouped by model and region. The red lines indicate the EU limit value for stage I (25 µg/m3), while the orange one for stage II (20 µg/m3).....33
Figure 25 - Maps of NO_2 annual mean produced by the four data fusion systems.....34
Figure 26 - NO_2 annual mean: boxplots of grid point concentration distributions grouped by model and region. The red lines indicate the EU limit value (40 µg/m^3).....35
Figure 27 - Maps of O3 93.1 percentile produced by NINFA, FARM-PI, FARM-LO systems.....36
Figure 28 - O3 93.1 percentile: boxplots of grid point concentration distributions grouped by model and region. The red lines indicate the EU target value (120 µg/m3).....36
Figure 29 - Attainment (green) and nonattainment (red) areas for PM10 annual mean (left) and PM10 percentile 90.41 (right). 37
Figure 30 - Attainment (green) and nonattainment (red) areas for PM2.5 annual mean (left) and NO2 annual mean (right). In the PM2.5 maps yellow areas indicate attainment regarding the EU limit of 25 µg/m3 and nonattainment for EU limit of 20 µg/m3. 38
Figure 31 - Population exposure estimate for NO2 annual mean.....40
Figure 32 - Population exposure estimate for percentile 90.41 of PM10.....41
Figure 33 - Population exposure estimate for PM2.5 annual mean.....42
Figure 34.- Population exposure estimate for PM10 annual mean.....43
Figure 35 - Population exposure estimate for ozone percentile 93.1.....44

Index of Tables

Table 1 - C1 dataset: monitoring stations grouped according to data supplier (rows), station type classification, zone type classification and measured pollutant (columns). 4

Table 2 - Main configurations of NINFA modelling system...8

Table 3 - Main configurations of FARM-PI modelling system...11

Table 4 - Main configurations of FARM-LO/ARPA-LO modelling system. 13

Table 5 - Cross-validation results: statistical scores for NINFA and FARM-PI data fusion systems.....27

Index of Figures Appendix A

Figure A 1 - PM10 annual mean: maps of observed data, monitoring stations are grouped by station classification 53

Figure A 2 - PM10 annual mean: boxplots of observed data grouped by station type and region..........53

Figure A 3 - PM10 percentile 90.41: maps of observed data, monitoring stations are grouped by station classification 54

Figure A 4.- PM10 percentile 90.41:boxplots of observed data grouped by station type and region. 54

Figure A 5.- PM2.5 annual mean: maps of observed data, monitoring stations are grouped by station classification 55

Figure A 6 -- PM2.5 annual mean:boxplots of observed data grouped by station type and region. 55

Figure A 7-. NO2 annual mean: maps of observed data, monitoring stations are grouped by station classification 56

Figure A 8-. NO2 annual mean:boxplots of observed data grouped by station type and region.............56

Figure A 9. O3 percentile 90.3: maps of observed data, monitoring stations are grouped by station classification 57

Figure A 10 – O3 percentile 93.1 of 8-hour running average daily maximum distribution....................57

1 INTRODUCTION

The Integrated project "Po Regions Engaged to Policies of Air" LIFE-IP PREPAIR supports the implementation of regional air quality plans (AQPs) and of Po Valley agreements on a larger scale, acting in a synergic way, so to strengthen the sustainability and durability of the results. Although the most critical area studied in the project is the Po Valley, the field of study is extended to Slovenia in order to assess and reduce transboundary pollutants transport. Regarding air quality, in fact, all the Regions located south of the Alps face the same adverse climatic conditions, which require higher technical and financial efforts to settle compliance problems, in comparison with other Regions. The Po Valley, a densely populated and heavily industrialized area, represents a non-attaining zone for PM (Particulate Matter), NO_2 (Nitrogen Dioxide) and O_3 (Ozone). Previous experience demonstrates that coordinated and large-scale actions are necessary in this area. A comprehensive policy, acting on a large scale and on several sources of pollutant precursors of PM and O_3, is essential to further reduce pollution levels. For this purpose, all the Regions have clustered in the so-called Po Basin Board and planned actions with the aim of further reducing the emission of pollutants and their precursors.

This first assessment report of action D5 provides a synthetic view on the state of air quality in the Po Valley and Slovenia for year 2020 and examines PM10, PM2.5, nitrogen dioxide and ozone, which are the pollutants whose concentration values more frequently exceed legislation thresholds. However, this report is not intended to be a formal air quality assessment which is responsibility of the regional authorities. The assessment was carried out with data fusion techniques using model output and monitoring data collected by action C1 of the PREPAIR project. Even though four CTM and data fusion modelling systems with different setup (resolution, boundary condition, meteorological data and data fusion technique) have been used, the model outputs are very similar to each other. In this report the assessment methodology, the data fusion technique and results of the most critical indicators compared to the limit values established by the 2008/50/EC Directive are shown.

2 ASSESSMENT METHODOLOGY

The assessment of air quality status in Po Valley and Slovenia for year 2020 was carried out with an integrated approach that exploits two different types of information:

- the air quality monitoring network data, accurate but available only in a limited number of locations;
- high spatial resolution concentration fields produced by means of a chemical transport model (CTM).

Currently, within the PREPAIR project, several CTM modelling systems running operational and air quality data are shared daily by all partners through action C1. Then, concentration fields and air quality monitoring data have been integrated using different data fusion techniques, one for each modelling system.

The assessment has been carried out taking into account the most critical indicators compared to the limit values established by the 2008/50/EC Directive:

- PM10 annual mean concentration values (the limit value set by EU legislation is 40 μg/m^3);
- PM2.5 annual mean concentration values (the limit value set by EU legislation is 25 μg/m^3 for stage I and 20 μg/m^3 for stage II);
- NO_2 (nitrogen dioxide) annual mean concentration values (the limit value set by EU legislation is 40 μg/m^3);
- 90.4 percentile of PM10 daily mean concentration values corresponding to the 36th highest daily mean of the year (the limit value set by EU legislation is 50 μg/m^3);
- 93.1 percentile of O_3 (ozone) maximum daily 8-hour average concentration values corresponding to the 26th highest daily maximum of the running 8-h mean of the year (the target value set by EU legislation is 120 μg/m^3)

In the following paragraphs, input data (air quality measurements and CTM models) have been first briefly described (paragraph 2.1), then the data fusion techniques (paragraph 2.2) and the results of the validation task (paragraph 2.3) are presented.

2.1 DATA FUSION INPUT DATA

2.1.1 AIR QUALITY DATA

The observational database used in data fusion procedures was built from the dataset implemented in action C1 by the support of all the partner. This dataset collects pollutant concentrations measured by monitoring stations managed by PREPAIR project partners, which are divided into urban, sub-urban and rural categories (zone type classification). Moreover, some stations represent the background level (B) whereas some others represent the industrial (I) or traffic (T) level (station type classification). Table 1 summarizes the main stations classification, while Figure 1 shows the spatial distributions of monitoring stations

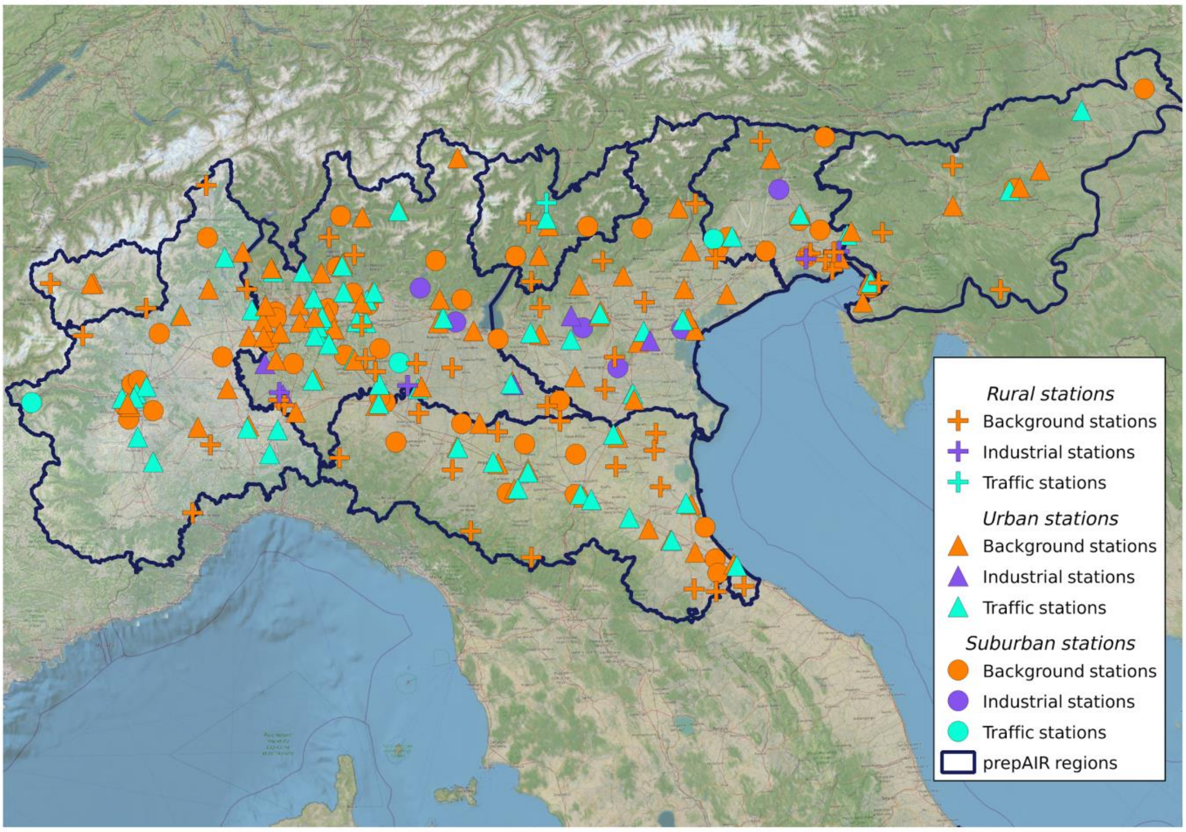


Figure 1 - Spatial distribution of monitoring stations availables in C1 dataset.

The C1 dataset contains hourly measurements of nitrogen dioxide (NO_2), and ozone (O_3), hourly and daily measurements of particulate matter PM10 and PM2.5 (see Table 1). The data were aggregated to obtain the air quality indicators (annual mean and percentiles) used in the assessment.

The data, with different levels of validation depending on data supplier, have been subjected to a quality check before their use (range control, outlier detection, data entry error etc.)

Region	Rural				Sub-urban				Urban				Total	Pollutant			
	B	I	T	Tot	B	I	T	Tot	B	I	T	Tot		NO2	O3	PM10	PM25
Emilia-Romagna	14	-	-	14	9	-	-	9	12	-	12	24	47	44	32	44	25
Friuli-Venezia-Giulia	6	2	-	8	8	3	1	12	6	-	4	10	30	23	20	28	8
Lombardia	10	2	-	12	16	2	1	19	27	3	24	54	85	84	51	64	32
Piemonte	5	-	-	5	9	1	1	11	10	-	12	22	38	37	20	34	21
Trentino	2	-	1	3	2	-	-	2	2	-	1	3	8	8	6	8	3
Valle d'Aosta	2	-	-	2	-	-	-	-	2	-	-	2	4	4	4	3	2
Veneto	7	-	-	7	1	3	-	4	14	2	-6	22	33	32	24	26	9
Slovenia	3	-	-	3	2	-	-	2	5	-	2	7	12	9	10	9	---
Total	**49**	**4**	**1**	**54**	**47**	**9**	**3**	**59**	**78**	**5**	**61**	**144**	**257**	**241**	**167**	**216**	**100**

Table 1 - C1 dataset: monitoring stations grouped according to data supplier (rows), station type classification, zone type classification and measured pollutant (columns).

Among all the stations included in the C1 dataset, the database used in data fusion procedures has been chosen based on the following criteria:

- station type: background stations (urban, suburban or rural) have been chosen; this choice is consistent with the resolution of the modelling systems described in paragraph 2.1.2;
- data capture percentage: stations with data capture percentage not less than 75% have been selected. This value allows to have enough stations in all regions of the domain, as shown in the Figure 2;
- location of monitoring station: for each pollutant, a dataset with homogeneous distribution and sufficient spatial coverage to capture the complexity of different territorial contexts has been built; if multiple stations fall in the same cell of computational domain, the station with the highest data capture percentage has been chosen (indeed this leads to different datasets for each different modelling system described in paragraph 2.1.2).

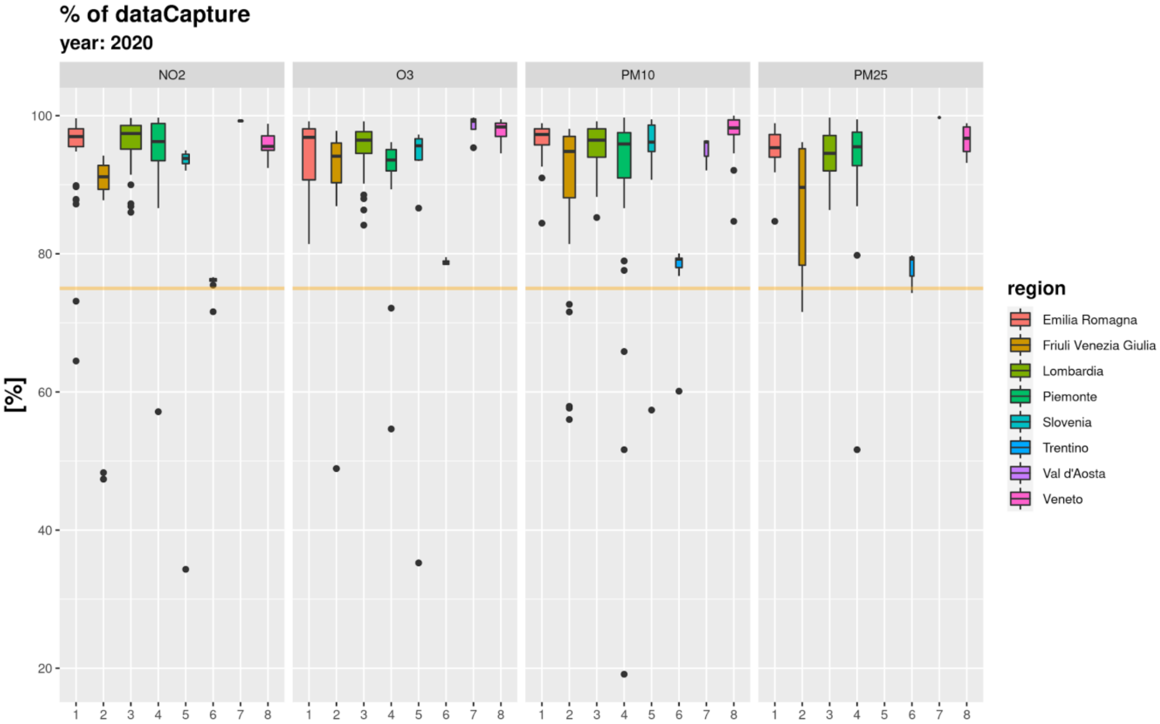


Figure 2 - C1 dataset: data capture percentage for each pollutant and for each data supplier.

Finally, an exploratory analysis on the measured data in 2020 was carried out, with the aim of checking and validating the assessment results obtained by means of data fusion procedures (see paragraph 3). The results of this exploratory analysis are presented in Appendix A.

2.1.1.1 CTM MODELS

Among all the CTM running operational within the PREPAIR project, four modelling systems have been used for the assessment: NINFA (Arpae Emilia-Romagna), FARM-PI (ARPA Piemonte), FARM-LO (ARPA Lombardia), CAMx-SLO (ARSO).

2.1.1.2 Emission data from CTM model

In the PREPAIR Project several activities have been performed for the development of emission datasets also with the aim to support the elaboration of CTM model simulations:

- the emission dataset developed in the Action A1 (dataset of emissions) estimated on reference year 2013 for the overall regions and countries in the model basin domain (figure 3 on the left)
- update on emissions estimates for year 2017 (with a municipal detail) implemented in Action D2 (Figure 3 on the right)

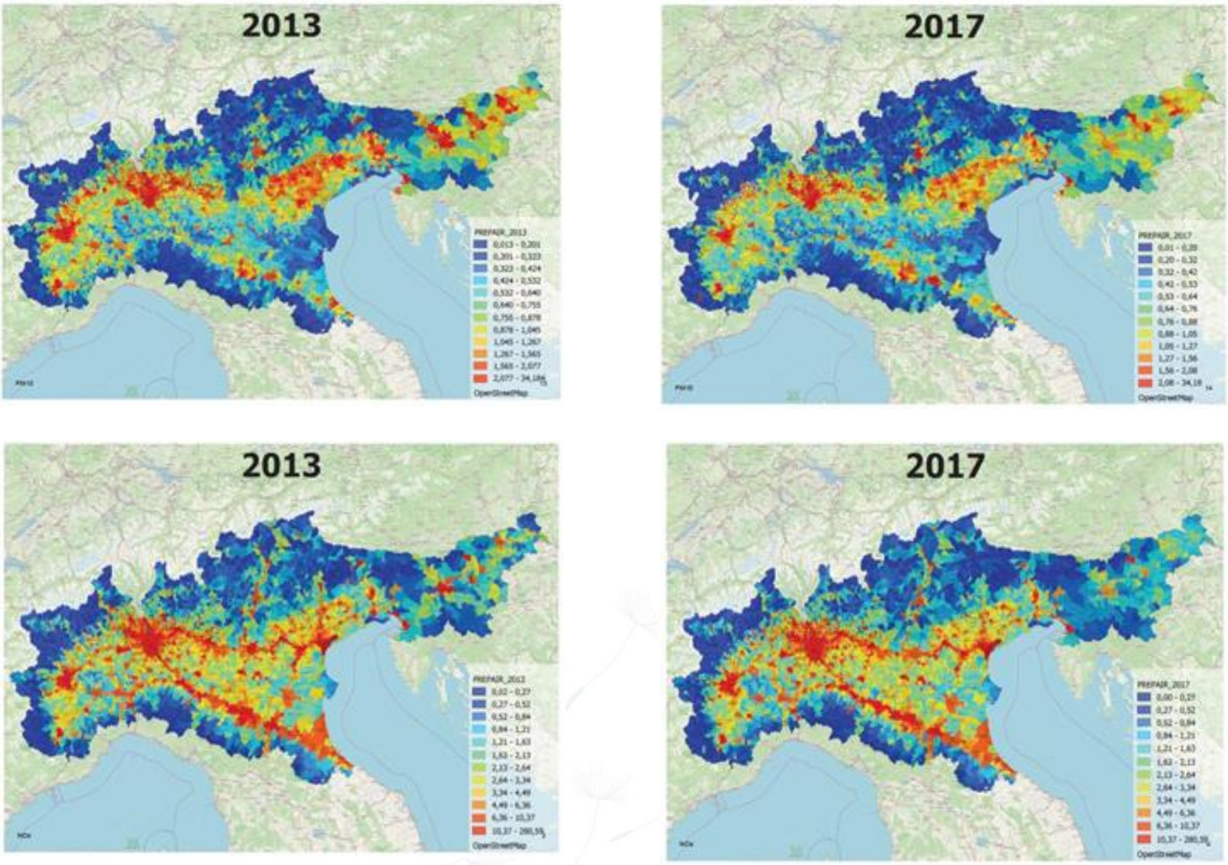


Figure 3 -. Emission maps for 2013 and 2017 representing PM10 (top) and N0x (bottom).

2.1.1.3 Arpae Emilia-Romagna Model (NINFA)

NINFA (Northern Italy Network to Forecast Aerosol pollution) is the operational AQ model of the Environmental Agency of the Emilia-Romagna Region (Arpae). The model suite includes a Chemical Transport Model, a meteorological model and an emissions pre-processing tool. The chemical transport model is CHIMERE, (http://www.lmd.polytechnique.fr/chimere/) an eulerian-type numerical model, which simulates transport, dispersion, chemical transformations and deposition (dry and wet) of air pollutants and aerosols. Starting from the emission data for the Po Valley, Slovenia and the other regions/countries present in the model domain,(http://www.lifeprepair.eu/wp-content/uploads/2017/06/Emissions-dataset_final-report.pdf), the emissions are prescribed to the grid model by using specific proxy variables for each emission activity SNAP3 (i.e. road network for traffic emission, population and urban fabric for domestic heating, and so on). The meteorological hourly input is provided by COSMO, the National NWP model used by the National Civil Protection Department. COSMO is a non-hydrostatic, limited-area atmospheric prediction model, based on the primitive thermo-hydrodynamical equations describing compressible flow in a moist atmosphere, with a variety of physical processes taken into account by dry and moist parameterization schemes. The time-dependent boundary conditions (with hourly frequency) in PREPAIR project are provided by CAMS service (https://doi.org/10.3390/atmos11050447)

The AQF (Air Quality Forecast) modelling system performs simulations over four nested domains:

- a Europe background domain covering with a horizontal resolution of 20 km (MEDL);
- a national background domain covering the whole Italian Peninsula with a horizontal resolution of 7 km (ITA7);
- an inner domain nested to ITA7 with 5 km horizontal resolution, including Northern Italy and Slovenia (PREPSLO). This domain is considered for the present assessment.
- an inner domain nested to ITA7 (EMR3), with 3 km horizontal resolution, centred over Emilia-Romagna region (EMR3);

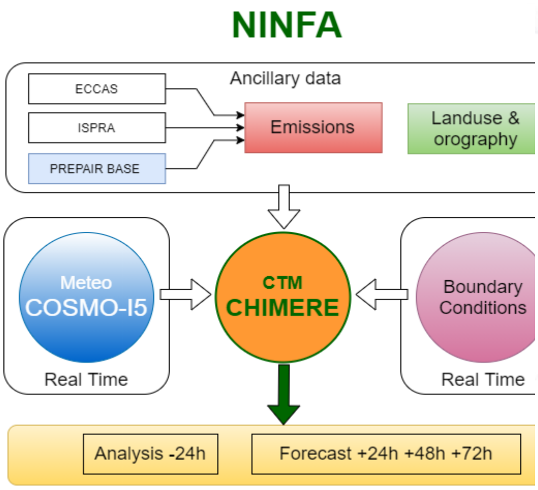


Figure 4 - NINFA model scheme

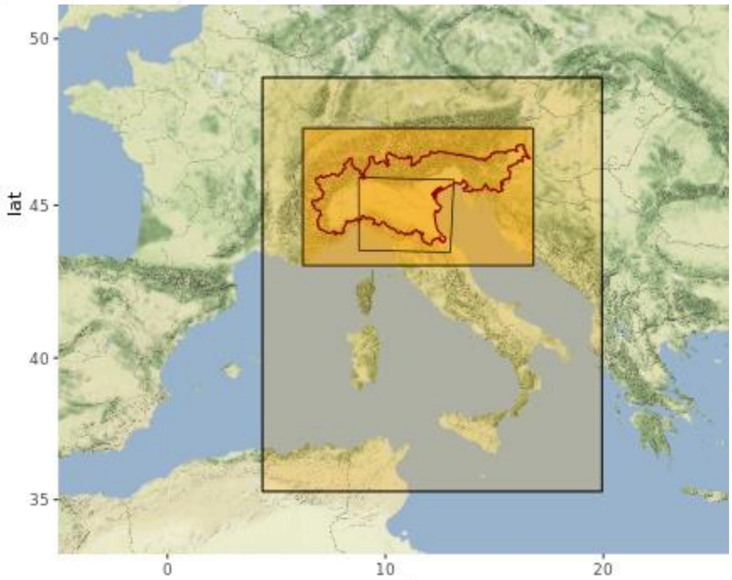


Figure 5 - NINFA PREPSLO domain nested to ITA7 domain. The area covered by region/country project partners is shown in red. The inner EMR3 domain is also shown

Domain	MEDL	ITA7	PREPSLO	EMR3
Bounding Box	Lon: -24.8 - 33.49 Lat: 27.04 - 54.99	Lon: 6.25 - 16.75 Lat: 43.1 - 47.35	Lon: 6.25 - 16.75 Lat: 43.1 - 47.35	xUTM : 482.4 - 821.4 yUTM: 4824.5- 5079.5
Vertical Resolution	9 level up to 500 hPA	9 level up to 500 hPA	9 level up to 500 hPA	15 level up to 500 hPA
Horizontal Resolution	0.18 * 0.17 degree	0.09 * 0.07 degree	0.07 * 0.05 degree	3 * 3 km
CTM Model	CHIMERE2017	CHIMERE2017	CHIMERE2017	CHIMERE2017
BC	CAMS	SNPA CAMS downstream service (MEDL)	SNPA CAMS downstream service (ITA7)	SNPA CAMS downstream service (ITA7)
METEO Model	COSMO5I	COSMO5I	COSMO5I	COSMO5I/COSMO2I
EMISSION	TNO-MACC III	ISPRA, TNO-MACCIII	Prepair, ISPRA TNO-MACCIII	Prepair, ISPRA TNO-MACCIII
OUTPUT	Hindcast, +72houres forecast	Hindcast, +72houres forecast	Hindcast, +72houres forecast	Hindcast, +72houres Forecast

Table 2 - Main configurations of NINFA modelling system.

2.1.1.4 ARPA Piemonte Model (FARM-PI)

The FARM-PI (Giorcelli et al, 2013) model is the operational AQF model of the Environmental Agency of the Piemonte Region (ARPA Piemonte). The forecasting system has been built by using state-of-the-art techniques for atmospheric transport and dispersion modelling. The computational system architecture (Figure 6) is modular, so that the model inter-dependence is limited, in order to facilitate system improvements without modifying the general structure.

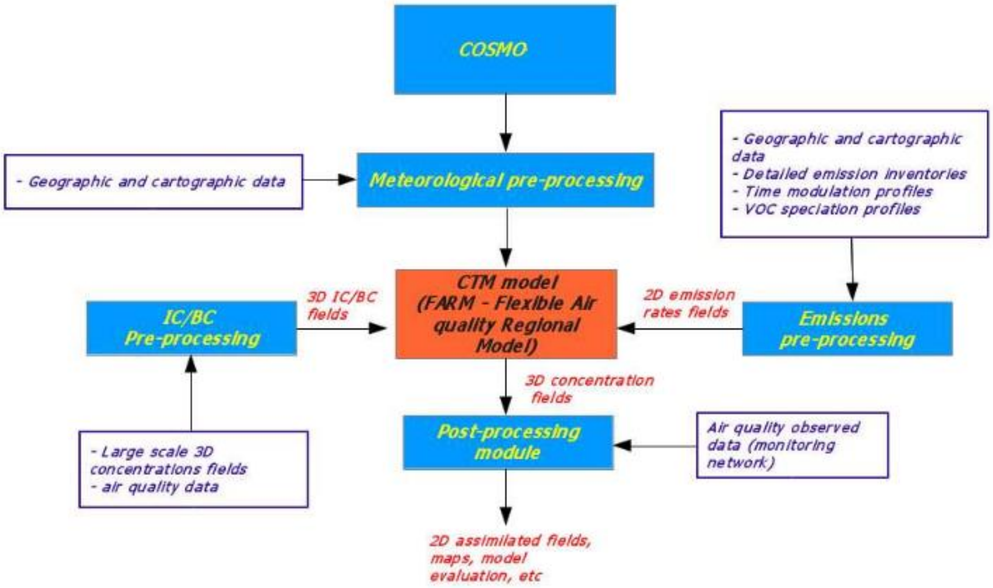
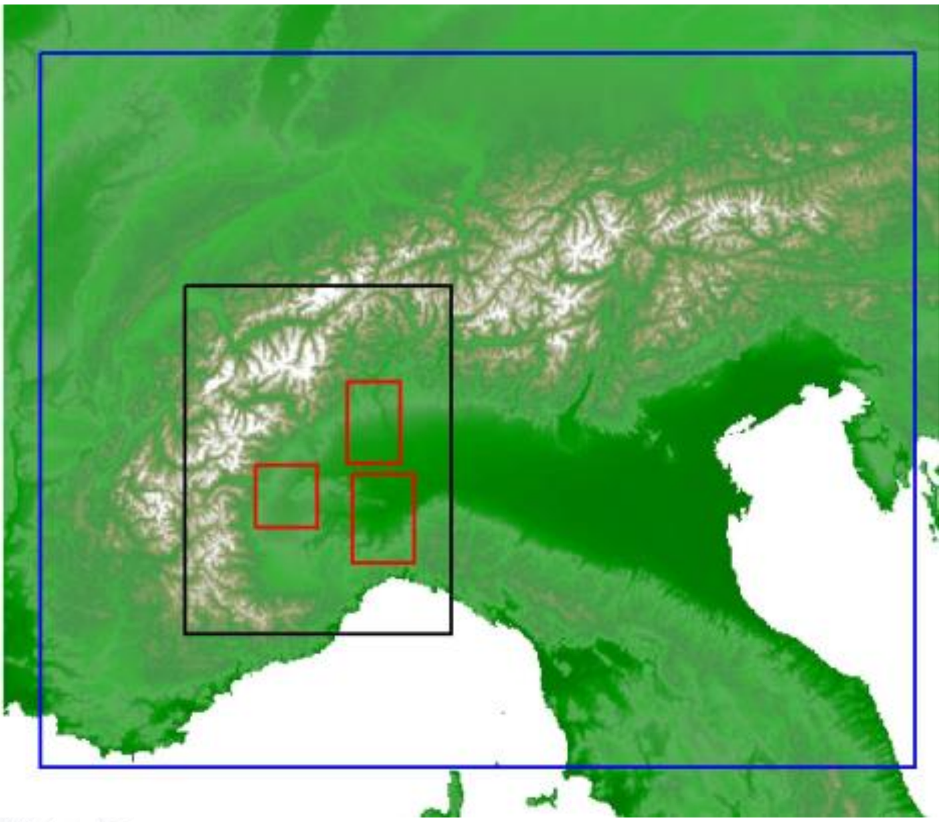


Figure 6 -. FARM-PI computational system architecture

The core of the system is represented by the air quality model FARM (Flexible Air Quality Model, Gariazzo et al, 2007; Silibello et al, 2008), a three-dimensional Eulerian model that accounts for transport, chemical conversion and deposition of atmospheric pollutants. The forecasting system needs a series of detailed input datasets: emission inventories, geographic and physiographic data (to describe topography, surface land cover and urban details), large scale air quality and meteorological forecasts. Some specific modules are needed to process these data in order to produce emissions, meteorological fields and boundary conditions necessary as input to the air quality model. Emission data (point, line and area sources) coming from different resolution inventories available over all computational domains are processed by a specific emission module in order to produce gridded hourly emission rates for all the chemical species considered by the air quality model. This pre-processing system allows non-methanic hydrocarbon speciation and flexible space and time disaggregation, according to cartographic thematic layers and specific time modulation profiles (yearly, weekly and daily). The meteorological fields are provided by 00 UTC runs of COSMO, the National NWP model used by the National Civil Protection Department. The COSMO model levels fields are directly interpolated and adjusted (forced to be non-divergent) over all the computational domains by an interface module Starting from topography and land-use data managed by the modelling system and gridded fields of meteorological variables provided by COSMO, a diagnostic model computes three-dimensional fields of horizontal and vertical diffusivity and two-dimensional fields of deposition velocities for a given set of chemical species. The initial and boundary conditions for the background domain are obtained by continental scale air quality forecasts provided by PrevAir European Scale Air Quality Service (http://www.prevair.org). The AQF modelling system performs simulations over the following three nested domains (two-way nesting), as shown in Figure 7:

- a background domain (g1, blue line), covering Po valley basin and the Alps, with a horizontal resolution of 8 km;
- a regional target domain (g2, black line), covering the whole Piemonte Region with a horizontal resolution of 4 km;
- an inner domain (g3, red lines), with 1 km horizontal resolution, centered over Torino metropolitan area;



. Figure 7 - FARM-PI computational domains.

This multi-scale approach allows to take into account the effect of sources located outside the target areas, and to better describe phenomena characterized by large spatial scales, such as photochemical smog and particulate matter accumulation processes. The forecasting system runs on a daily basis in order to produce air quality forecasts for the current day and the two days after, with one hour time resolution.

	g1	g2	g3
Domain	Lon: 191000-911000 Lat: 4765000-5349000	Lon: 309000-529000 Lat: 4875000-5159000	Lon: 367500-418500 Lat:4961500-5012500
Vertical Resolution	16 level up to 7500 a.g.l	16 level up to 7500 a.g.l	16 level up to 7500 a.g.l
Horizontal resolution	8km x 8km	4km x 4km	1km x 1km
CTM model	FARM v4.13	FARM v4.13	FARM v4.13
BC	PrevAir services	Two-way nesting with g1 grid	One-way nesting with g2 grid
Meteo model	COSMO-I5	COSMO-I5	COSMO-I5
Emission data	Prepair, IREA, ISPRA, EMEP	Prepair, IREA, ISPRA, EMEP	IREA (Piemonte regional inventory)
Output	+72 hours forecast, air quality indicators, air quality maps	+72 hours forecast, air quality indicators, air quality maps	+72 hours forecast, air quality indicators, air quality maps, air quality index

Table 3 - Main configurations of FARM-PI modelling system.

2.1.1.5 ARPA Lombardia Model (FARM-LO/ARPA-LO)

The operational air modelling system of ARPA Lombardia is based on ARIA Regional developed by AriaNET srl. There are two different domain extensions: one for Regione Lombardia (in Figure 8 represented by red line named g3) and one for the PREPAIR project (in Figure 8 represented by blue line named g2) which includes the Po basin extended from western (Piemonte and Valle d'Aosta Regions) to eastern part (Slovenia) and from northern (Trento Province and Friuli Venezia Giulia Regions) to southern (Emilia-Romagna Region). The PREPAIR model domain consists of 210 rows x 105 columns with a cell resolution of 4 km and is vertically discretized into 16 different levels till 4960 m a.s.l.. The main workflow of modelling architecture is composed by (Figure 9):

- WRF suite: the forecasts produced by the deterministic model in a global scale GFS (National Center for Environmental Predictions NCEP) are used as BC (free distributed by National Oceanic and Atmospheric Administration, NOAA; https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/global-forcast-system-gfs)

- SURFPro suite: estimation of micrometeorological fields linked to atmospheric turbulence (i.e., mixing height, atmospheric stability classes, vertical and horizontal diffusivity), dry deposition velocity for several chemical species and natural emissions (from vegetation to winds action).
- EMMA: spatial (i.e., gridding on domain cells) and temporal (i.e., hourly) attribution of the inventory emission data (INEMAR). Furthermore, COV and particulate matter speciation are considered into FARM. Mainly, in order to use the database developed by Action D2, a harmonization procedure of the tables which associate SNAP codes for each inventory to spatial proxy and to contaminants speciation have applied.
- IC/BC: initial condition for chemical species concentration in the model domain and at the beginning of simulation and boundary condition representing the chemical concentration in the border of the domain time-independent during all the simulation process (provided by QualeAria: http://www.qualearia.it).
- FARM: WRF, IC/BC and Emission Inventories are the input for the 3D chemical transport model (CTM) which is a multi-grid Eulerian model for dispersion (wet and dry), transformation and deposition (droplet and gas-phase chemistry) of air pollutants in gas and aerosol phases. This is the core of the modelling system

The main output consists of the estimation of pollutant concentrations (i.e., PM10, NO_2 and O_3). Moreover, these can be corrected based on the observed air quality data provided by the regional monitoring network (i.e., SCM, Successive Correction Method, see the paragraph 2.2.3).These techniques have been applied on hourly simulated concentrations by the modelling system, not on the yearly value, as in other cases. The modelling system with the support of AriaNET srl has been applied over the following two domains, as shown in Figure 8:

- a background domain (g2, blue line), covering Po valley basin and the Alps and Slovenia, with a horizontal resolution of 4 km;
- a regional target domain (g3, red line), covering the whole Lombardia Region with a horizontal resolution of 1 km.

	g2	g3
Domain	Lon: 254506-1112902 Lat: 4808039-5235127	Lon: 452013-699319 Lat:4935490-5170980
Vertical Resolution	16 level up to 4960 a.g.l	16 level up to 4960 a.g.l
Horizontal resolution	4km x 4km	1km x 1km
CTM model	FARM	FARM
BC	QualeAria: http://www.qualearia.it	QualeAria: http://www.qualearia.it
Meteo model	WRF	WRF
Emission data	Prepair, INEMAR, EMEP, ISPRA	Prepair, INEMAR, EMEP
Output	+96 hours forecast, air quality indicators, air quality maps	+96 hours forecast, air quality indicators, air quality maps, air quality index

Table 4 - Main configurations of FARM-LO/ARPA-LO modelling system.

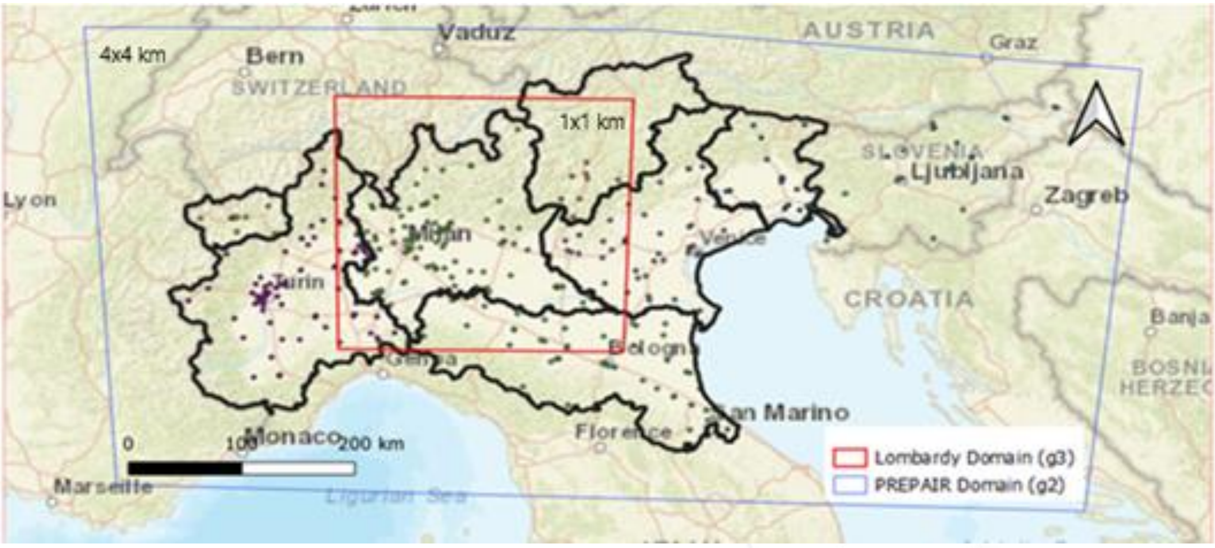


Figure 8.- PREPAIR domain of ARPA-LO modelling system

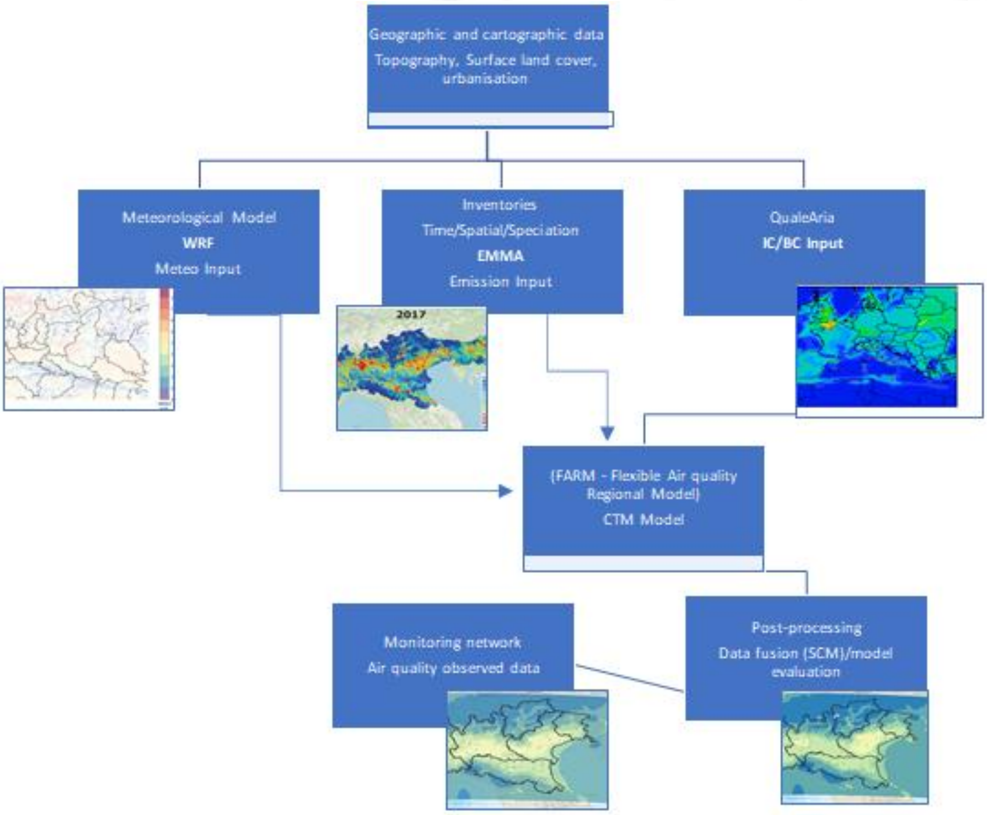


Figure 9.- The architecture of ARPA-LO modelling system.

2.1.1.6 ARSO Model (CAMx-SLO)

ALADIN/SI-CAMx modelling system consists of chemical transport CAMx model (Comprehensive Air Quality Model with Extensions) coupled offline in 1 hour interval with the operational meteorological ALADIN/SI model.

ALADIN/SI model is hydrostatic model, in which the hydrostatic approximation replaces the vertical momentum equation (http://www.umr-cnrm.fr/aladin/). Setup of the model is as follows (Slovenian Environmental Agency, ALADIN/SI Model Products, http://meteo.arso.gov.si/):

- Model with the Central Europe domain (Figure 10). Horizontal resolution: 4.4 km, 421 x 421 model points.
- Vertical resolution: 87 levels (first model level 10 meters above the surface, 19 levels below the pressure surface of 900 hPa, 23 levels below the pressure surface of 850 hPa).
- Meteorological fields for the CAMx input: pressure, temperature, wind, specific humidity, cloud water, rainwater, snow water, falling ice crystal volume, optical cloud thickness, vertical turbulent diffusivity coefficient and the surface temperature field.

CAMx is an Eulerian model, able to simulate transport, dispersion, chemical transformations and deposition (dry and wet) of air pollutants (ENVIRON). The model setup of is as follows:

- Model domain is smaller than the ALADIN/SI domain, but still large enough to cover the entire Po Valley region, Slovenia and the surrounding countries (Figure 10);
- Horizontal resolution: 4.4 km, 270 x 210 model points;
- Vertical resolution: lower 68 levels of the ALADIN/SI's 87 levels;
- Chemical initial conditions: from previous run;
- Chemical boundary conditions: Global model system IFS-TM5 (The European centre for Medium-Range Weather Forecasts, ECMWF). MACC reanalysis, http://pps.ecmwf.int/datasets/data/macc-reanalysis/);
- Different anthropogenic emission databases:
- Emissions over Slovenia: National inventory for year 2013 (resolution: 100 m)
- Emission over Po Valley (i.e., PREPAIR area): PREPAIR emission data base for year 2016
- Emissions outside Slovenia and PREPAIR area: European TNO-MACC-III for year 2011.
- Chemical mechanism used: SAPRC07TC ("Toxics" version of SAPRC07, with additional model species to explicitly represent selected toxics species, https://intra.engr.ucr.edu/~carter/SAPRC/)

Among above listed input data, some additional input data is also required by the CAMx.

These include geographical variables: land use (CORINE database, https://land.copernicus.eu/pan-european/corine-land-cover), Leaf area index (from ALADIN/SI model), and total amount of ozone in the atmosphere (i.e. amount of ozone in each column of air, measured with a spectrometer, TOMS - Total Ozone Monitoring Spectrometer, https://eospso.nasa.gov/missions/total-ozone-mapping-spectrometer-earth-probe).

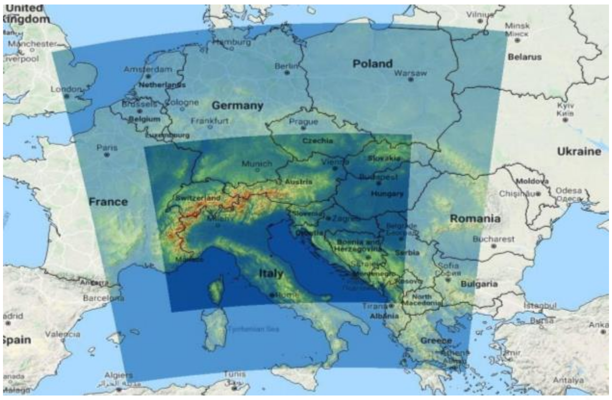


Figure 10 - Model domain of ALADIN/SI and CAMx model.

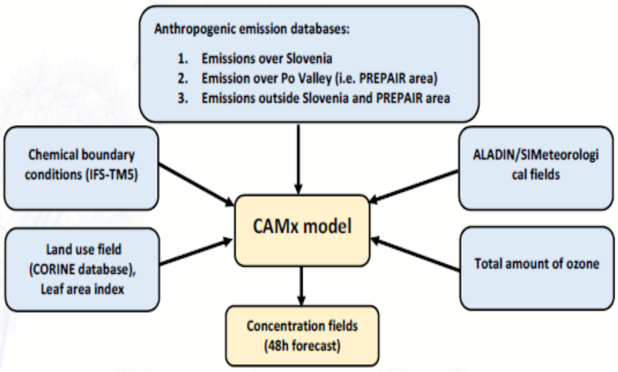


Figure 11– Input data for CAMx model.

2.2 DATA FUSION TECHNIQUES

2.2.1 NINFA and Observations Data Fusion

The pollutant concentration output by the CTM NINFA can well represent the spatial distribution of pollutants while, on the other hand, in situ measurements are more quantitatively accurate. A data fusion post processing is then applied to CTM simulations in order to get the most benefit from both CTM spatial representativeness and observation precision.

A geostatistical algorithm is used in Arpae to merge data from different sources. The pollutant background concentration can be regarded as a phenomenon measured by two variables, one more precise but known at only few locations (the observations) and one less accurate but known in the whole domain (the CTM on a regular grid), so Kriging with External Drift (KED) is a suitable technique to be applied to this dataset.

The considered domain is characterized by a complex orography, so that the elevation above the sea level (h) is considered as a further spatial explanatory variable. A cross validation including or not including elevation was performed to verify the improvement introduced by the second explanatory variable.

Let the statistical process we are estimating (either annual mean concentration or percentile) at X location be Y(X), in KED it is assumed that its expectation E[Y(X)] is equal to a combination of the two explanatory variables, CTM model (m) and elevation (h):

$$E[Z(X)] = a + b \cdot m(X) + c \cdot h(X)$$

(Wackernagel, 2003)

With this assumption on the mean part of the process, the residuals are estimated.

To fulfil the hypothesis of a gaussian process, before fitting the variogram, a Box-Cox transformation with fixed zero lambda parameter is done. Moreover, the covariance function is estimated assuming an exponential variogram.

The KED algorithm has been implemented for the present work by means of the geoR R package (Ribeiro and Diggle, 2001; Diggle and Ribeiro, 2007). For the present assessment, the main indexes are evaluated with the described KED method: PM10 annual mean, PM10 90.41 percentile, PM2.5 annual mean, NO2 annual mean, O3 93.10 percentile.

The KED spatial prediction is performed at the NINFA model grid, i.e. at about 5 X 5 km^2 resolution, on PREPSLO domain.

To test the prediction skill of the used KED method, a cross validation has been carried out and the results are shown in section 2.3.

2.2.2 FARM-PI and Observations Data Fusion

In order to make pollutant model outputs more realistic and their spatial distribution more representative, FARM-PI concentration fields were fused with the observed data through kriging with external drift method (KED, Wackernagel 2003) by employing the geoR package in R (Development Core Team 2010; Ribeiro and Diggle, 2001). Specifically, the kriging was applied on the observations while the external drift was represented by the FARM-PI model output, since KED is a particular case of universal kriging, where the trend component is the CTM output (Ignaccolo et al, 2013; Ghigo et al 2017). To make observed data approximately normally distributed with constant variance, a Box-Cox transformation (Box and Cox 1964) was applied separately per pollutant.

Therefore, transformed observations were interpreted as realizations of a Gaussian spatial process Y (s) at spatial location s, in the domain S, that has the following structure:

$$Y(s) = \mu(s) + w(s) + \varepsilon(s),$$

where:

$\mu(s) = X\beta$ is the spatial trend component, $\beta = \{\beta_0, \beta_1, \beta_2\}$ is the unknown parameter vector, X = [1, FARM-PI(s), H GT (s)] is the deterministic variable including FARM-PI model output as well as orography (HGT): the addition of this variable as auxiliary covariate had the purpose to introduce information about the complex Po basin terrain. w(s) is a zero-mean stationary Gaussian random process with sill σ^2 that takes into account the spatial correlation between observations by means of the spatial correlation function $\rho(\cdot)$ with range φ. Finally, $\varepsilon(s)$ is the error term characterized by the variance τ^2 (nugget). The leave-one-out cross-validation method was performed to choose the spatial covariance function and the best results were obtained with the exponential function, on all pollutants. To fit the model, firstly the parameters of the Box-Cox transformation and then the covariance parameters were estimated by the use of restricted maximum likelihood method.

The KED procedure was applied to the concentration fields of PM10 annual mean, PM10 90.41 percentile, PM2.5 annual mean, NO2 annual mean, O3 93.1 percentile produced by the FARM-PI modelling system on the g1 grid (see paragraph 2.1.3.2).

The model output post-processing performs well. Moreover, we carried out a cross-validation analysis in order to evaluate the KED performance and it showed that kriging results are satisfactory. The results of this analysis are reported shortly in paragraph 2.3

2.2.3 FARM-LO and Observations Data Fusion

ARPMEAS (ARchive Plus MEASurements) combine background 2/3D fields with observed data. Successive correction method (SCM) approach is implemented for the data fusion process. Briefly, production of gridded analysis is based on the Bratseth technique (Bratseth, 1986) that is a successive correction method (SCM, Brewster, 1996 and Daley, 1991) which includes background and observation error statistics. The analysis is initialized with a background field, or first guess, which is then modified by the analysis of local data onto the model grid. The analysis values at observation locations are first obtained using a bilinear interpolation. The analysis of a model variable, s, is then performed at the model grid points:

$$\boxed{s_x(n) = s_x(n-1) + \sum_{j=1}^{nobs} \alpha_{xj} [s_j^0 - s_j(n-1)]}$$

The grid point values are determined using a weighted sum of ‘observation increments’, which are the differences between the observation values s_j^0 and the analysis values at the observation locations s_j(n-1). On the initial pass over the grid, is provided by the background field. The true analysis is performed at the observation locations, which allows additional interpolation to be avoided:

$$s_i(n) = s_i(n-1) + \sum_{j=1}^{nobs} \alpha_{ij} [s_j^0 - s_j(n-1)]$$

Here s_i(n) is the analysis value at the observation location i. On the initial iteration, s_i(n-1) is the background value interpolated to the observation location. The weights are normalized by the observation density around each analysis point:

$$\alpha_{xj} = \frac{\rho_{xj}}{m_j}; \alpha_{ij} = \frac{(\rho_{ij} + \epsilon^2 \delta_{ij})}{m_j}$$

where α_{xj} and α_{ij} are the weights used respectively at the grid point and at the observation locations analysis and δ_{ij} is the Kronecker delta which is zero unless i=j. The correlation coefficients ρ are assumed to be Gaussian functions, allowing the weights to asymptote to zero with increasing observation distance from the analysis point.

$$\rho_{ij} e^{\frac{-|r_{ij}|^2}{R^2}} \cdot e^{\frac{-|\Delta z_{ij}|^2}{R_z^2}}$$

Here rij is the horizontal distance between observations i and j, Δz_{ij} is the vertical distance, R and R_z are the horizontal and vertical scaling distances. The quantity m_j represents the local data density around the analysis point, and includes the error statistics:

$$m_j = \epsilon^2 + \sum_{j=1}^{nobs} \rho_{ij}$$

When combining measurements and model results, it is important to take into account the so-called lack of representativeness errors, which can be defined as “the typical deviances or differences that occur between model calculated and observed concentrations, if their spatial and/or temporal positions, or averaging characteristics, do not match” (Zhan et al., 2006). Observational error variances derive from two different sources: instrumental and those associated with local phenomena (e.g. emissions, local flows and turbulence) at spatial scales not resolved by the underlying model. The second error is denoted “error of representativeness”. The observational error εo is the sum of the instrument and the representativeness errors. According to Elbern et al. (2007) the representativeness error can be expressed by the following formula:

$$\epsilon_{repr} = \epsilon_{abs} \cdot \sqrt{\frac{\Delta x}{L_{repr}}}$$

where Δx is the grid resolution of background field, Lrepr the characteristic length of the observations (e.g. the radius of influence associated with different types of ground based stations), and ε_{abs} is a tuning parameter called “characteristic absolute error”. Pagowski et al. (2010) found experimentally that $\varepsilon_{abs}=1/2\varepsilon_{instr}$ and suggest the following values for L_{repr} : 10, 4 and 2 km respectively for rural, suburban and urban stations. Using the above formula and definition, we obtain the following expression for ε^2:

$$\epsilon^2 = \frac{\sigma_o^2}{\sigma_B^2} = \frac{\langle \epsilon_o^2 \rangle}{\sigma_B^2} = \frac{\langle (\epsilon_{instr} + \epsilon_{repr})^2 \rangle}{\sigma_B^2} = \dots = \frac{\sigma_{instr}^2}{\sigma_B^2}\left(1 + \frac{n \cdot \Delta x}{4 \cdot L_{repr}}\right)$$

The value n=4 is a consistent with the concept of effective model resolution (e.g. 4Δx, see Pielke 2013) and provide a representativeness error that is always greater or equal to the instrument error. Consequently, the spatial features of assimilated fields will depend on the values assumed by the characteristic lengths associated with each monitoring station.

These techniques have been applied in the PREPAIR simulation on hourly concentrations simulated by the modelling system, not on the yearly value, as in other cases.

2.2.4 CAMx-SLO and Observations Data Fusion

Data fusion is considered one of the techniques of data assimilation (Lahoz et al. 2014), where we combine the results of numerical models and the point measurements (Schneider et al, 2015). There are known various statistical and geostatistical approaches to the data fusion (Berrocal et al, 2012). In our case the used statistical method for data fusion was geostatistical approach of kriging with external drift (Cressie, 1993)).

Kriging with external drift is a geostatistical algorithm where the value of a variable (interpolated value) at any grid point is calculated as a linear combination of measurements of the surrounding measuring points. The coefficients of this linear combination are calculated under assumption, that the mean square of the differences between the measured and interpolated values at the measurement points (kriging variance) are the smallest. In addition to this assumption (smallest mean square error), when calculating the coefficients of a linear combination, we also take into account the outcome of the spatial relationship of the variable, which is described by the variogram function (Cressie, 1993). The average of the considered variable may also depend on other explanatory variables, such as the altitude. In such case, we express the average as a linear combination of explanatory variables and look for a spatial correlation only for the residues of this function.

In our case, we performed Kriging with external drift in two stages. In the first stage, we interpolated the results of model concentration fields with a resolution of 4.4 km to the model grid with a resolution of 1 km, taking into account the altitude field and the field of geographical coordinates (latitude and longitude) with 1 km resolution as external variables. In the second stage, we interpolated the measurement points to a model grid with 1 km resolution, taking into account the interpolated field of model values (i.e. the result from the first step) and the field of geographical coordinates (latitude and longitude) at 1 km resolution.

2.3 DATA FUSION VALIDATION

2.3.1 Cross validation

For two data fusion datasets, FARM-PI and NINFA, a kriging methodology cross validation has been carried out. The one-leave-out methodology has been applied to verify the spatial prediction performance. The results are presented both in qualitative terms, by means of scatter plots, and in quantitative terms by means of performance indexes.

The scatter plots of observed/simulated data for each air quality index are shown: PM10 annual mean, PM10 90.41 percentile, PM2.5 annual mean, NO_2 annual mean, O_3 93.1 percentile.

In the following plots the lines defining the admitted model percentage discrepancy (in terms of percentage relative uncertainty) and the EU limit value are depicted for each pollutant index.

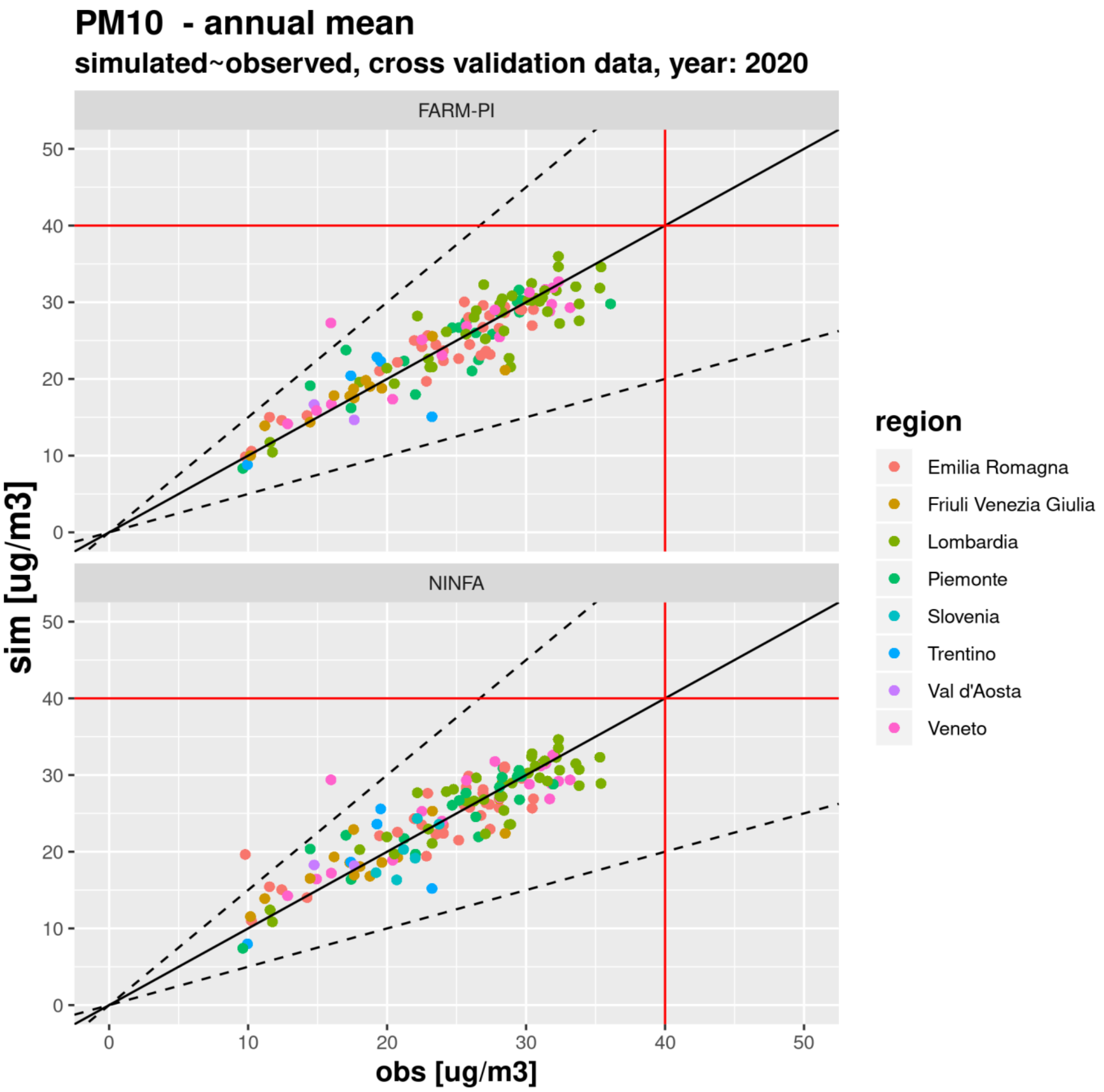


Figure 12.- PM10 annual mean: cross validation scatter plot for FARM-PI (top) and NINFA (bottom). The dotted lines represent the admitted relative uncertainty (50% for PM10 annual mean), while the red lines indicate the EU limit value (40 µg/m3).

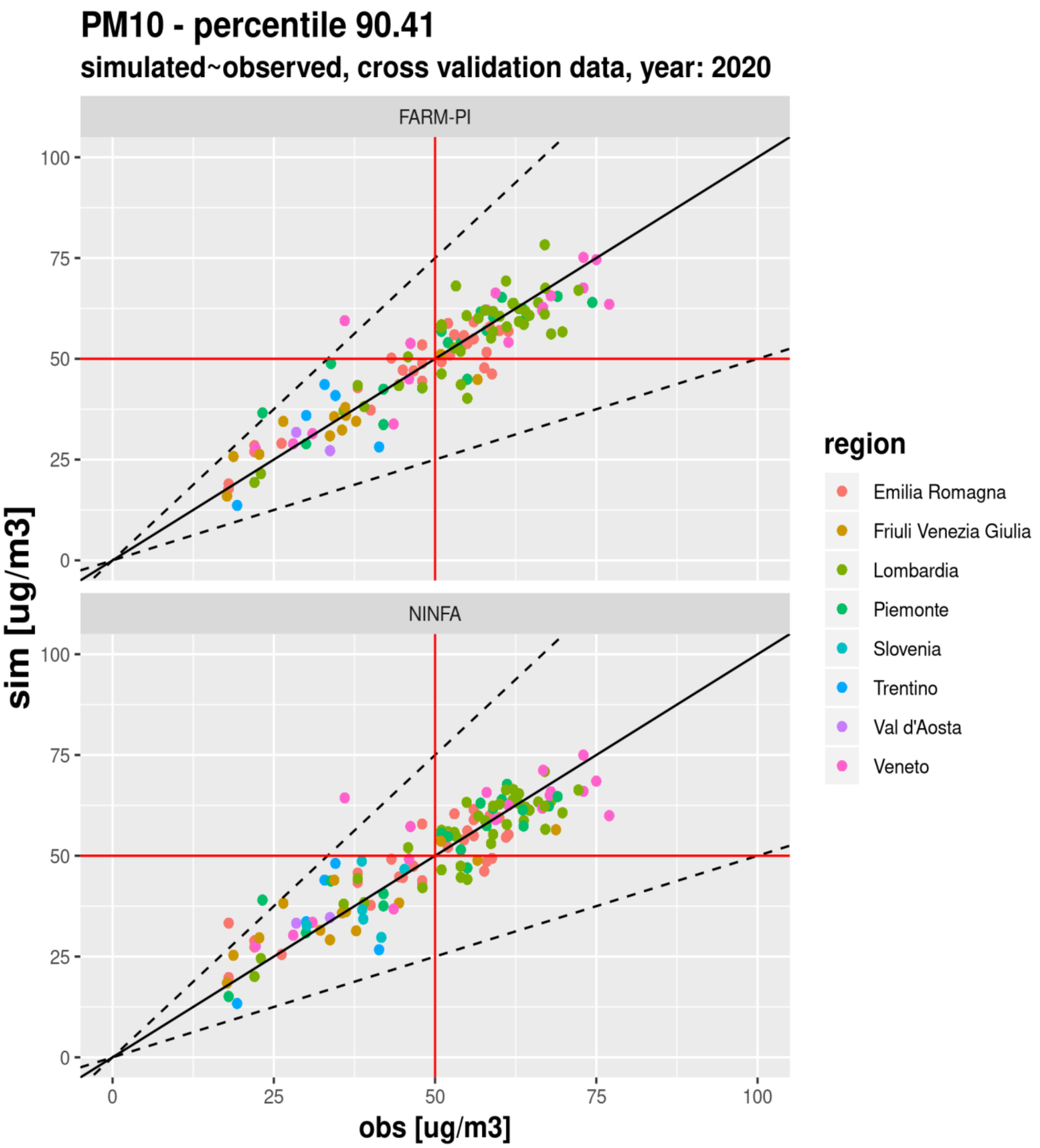


Figure 13.- PM10 percentile 90.41: cross validation scatter plot for FARM-PI (top) and NINFA (bottom). The dotted lines represent the 50% relative uncertainty, while the red lines indicate the EU limit value (50 µg/m3).

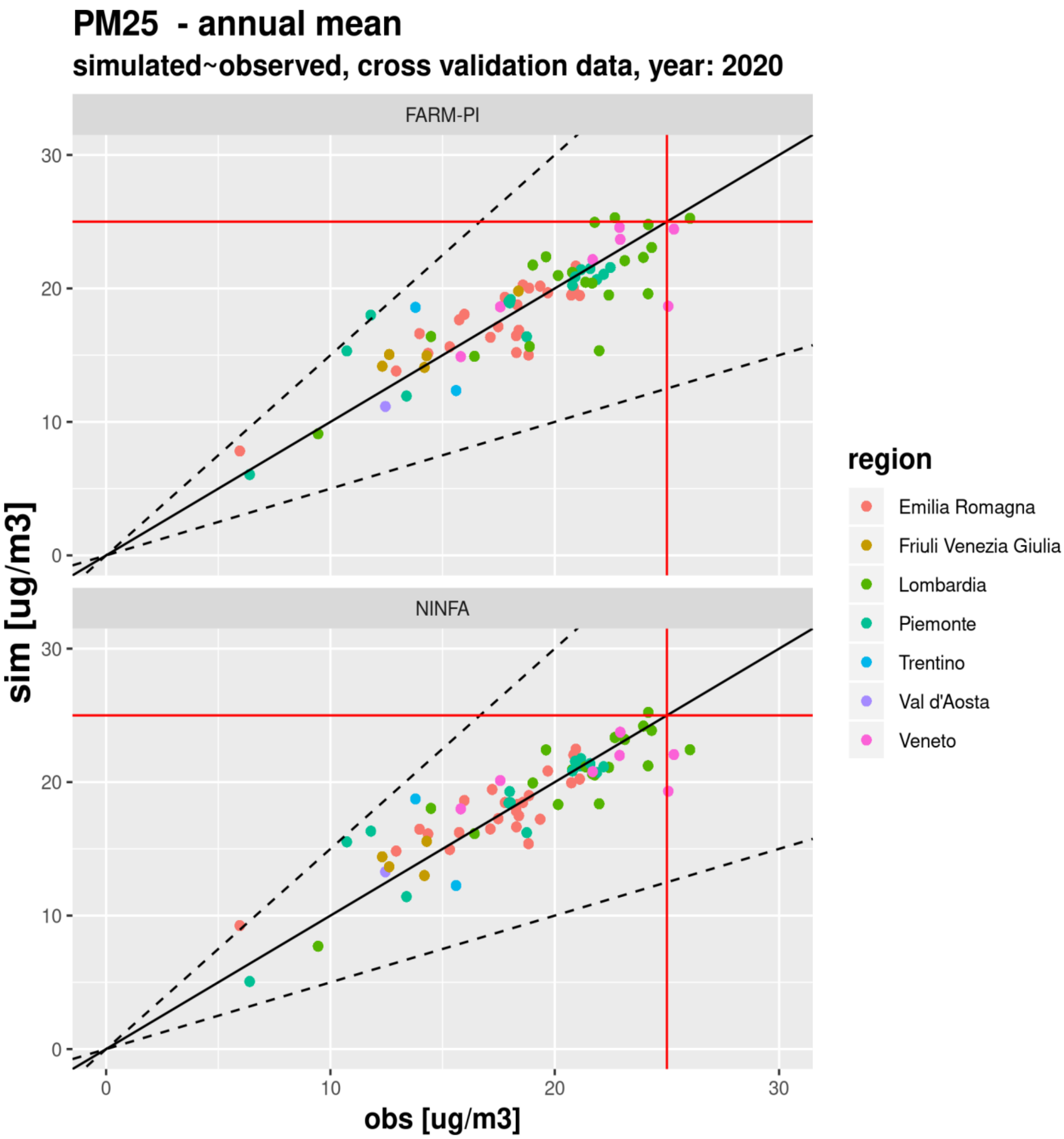


Figure 14 - PM2.5 annual mean: cross validation scatter plot for FARM-PI (top) and NINFA (bottom). The dotted lines represent the admitted relative uncertainty (50% for PM2.5 annual mean), while the red lines indicate the EU limit value (25 µg/m3).

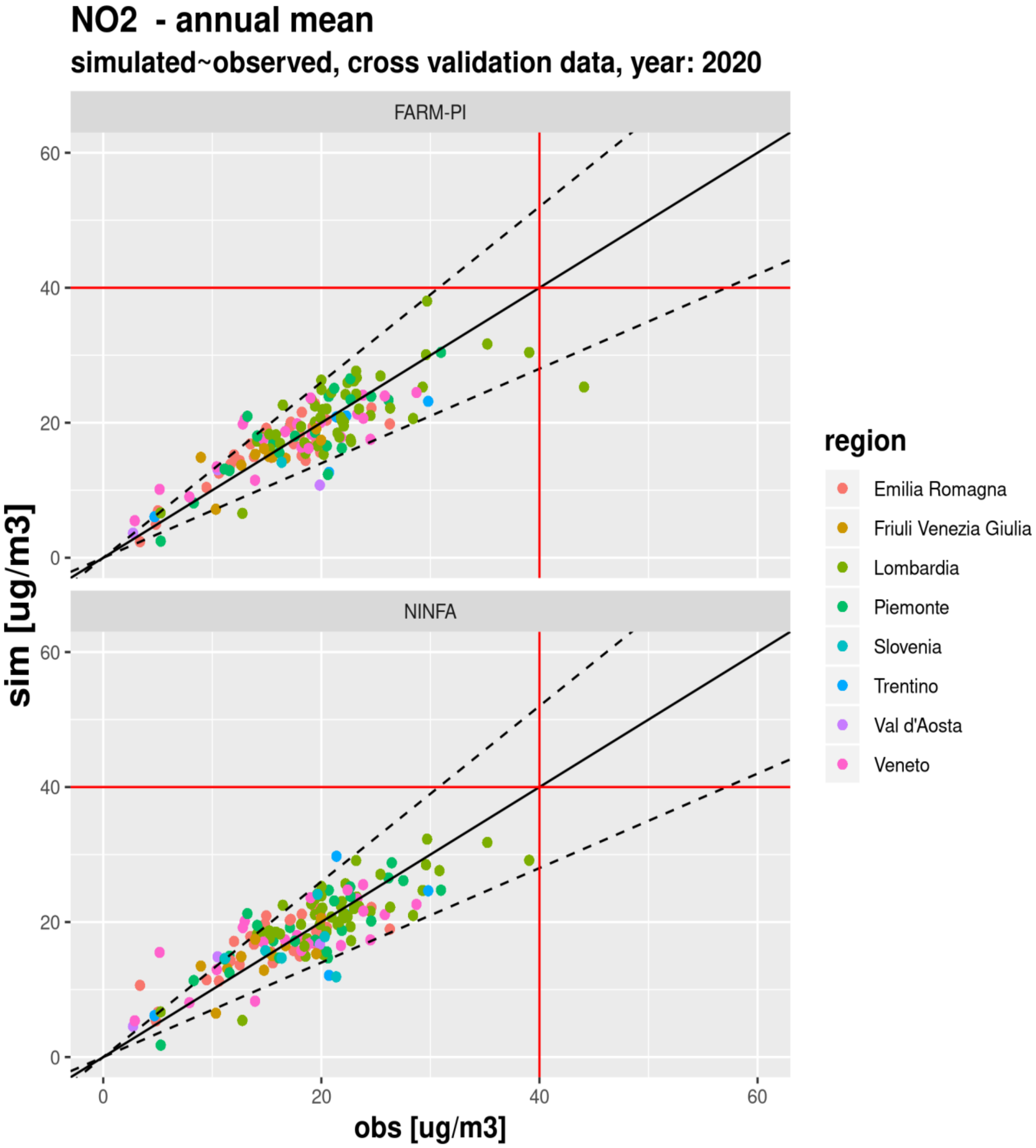


Figure 15 -. NO2 annual mean: cross validation scatter plot for FARM-PI (top) and NINFA (bottom). The dotted lines represent the admitted relative uncertainty (30% for NO2 annual mean), while the red lines indicate the EU limit value (40 µg/m3).

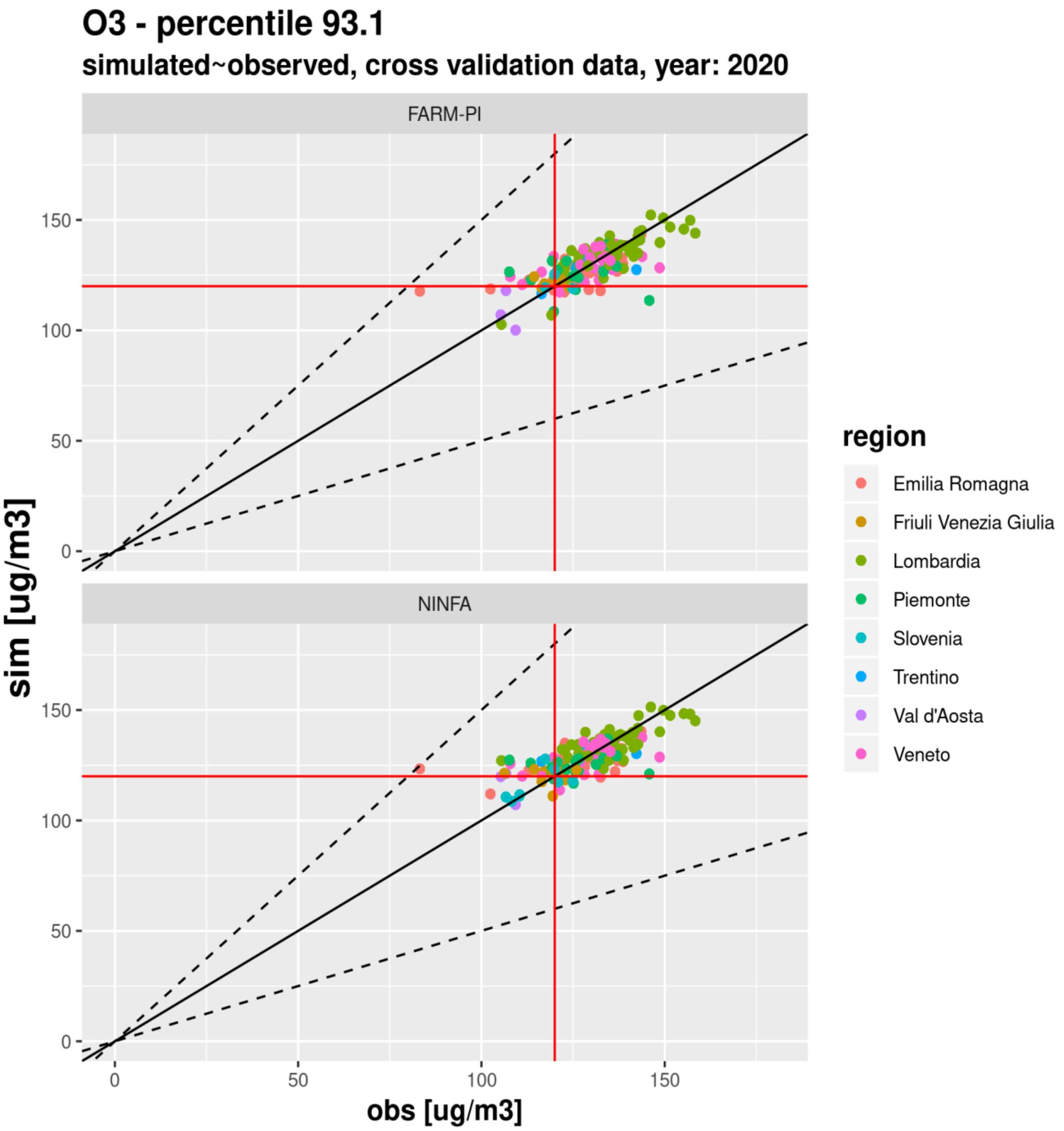


Figure 16 - O3 percentile 93.1: cross validation scatter plot for FARM-PI (top) and NINFA (bottom). The dotted lines represent the admitted relative uncertainty (50% for O3 93.1 percentile), while the red lines indicate the EU target value (120 µg/m3).

Overall, it can be observed a very good agreement between observed and simulated data for both data fusion simulations, in particular for PM10, PM2.5 and ozone.

The bulk of PM10 predictions, either annual mean or 90.41 percentile, lie within the tolerance area; only in very few stations the simulated data exceeds observation beyond the admitted model discrepancy.

All PM2.5 annual mean simulations are within the tolerance, so it is for the O3 93.1 percentile, for both models.

For the NO_2 annual mean the results are good, but the scatter plot shows, for both simulations, some points not included in the tolerance area, with local overestimation or underestimation. This behavior is probably due to high spatial variability of NO2 concentrations and the fact that some stations have local peculiarities which cannot be reproduced at model resolution (about 5 km for NINFA and 8 km for FARM-PI).

In the following table the main performance statistical scores are summarized for NINFA and FARM-PI cross-validation dataset: we considered three typical indexes based on the differences between fused and observed data that provided meaningful information: mean error (ME), unbiased root mean squared error (URMSE) and Pearson correlation (Yu et al, 2006; Denby et al, 2011).

model	index	pollutant	ME	URMSE	PEARSON
NINFA	annualMean	PM10	-0,10	3,08	0,88
FARM-PI	annualMean	PM10	0,08	2,88	0,90
NINFA	perc-90.4	PM10	-0,33	6,55	0,90
FARM-PI	perc-90.4	PM10	0,13	6,20	0,91
NINFA	annualMean	PM25	-0,06	2,02	0,88
FARM-PI	annualMean	PM25	0,01	2,23	0,86
NINFA	annualMean	NO2	-0,12	3,73	0,81
FARM-PI	annualMean	NO2	0,10	3,85	0,82
NINFA	perc-93.1	O3	-0,51	8,00	0,73
FARM-PI	perc-93.1	O3	-0,18	7,96	0,72

Table 5 - Cross-validation results: statistical scores for NINFA and FARM-PI data fusion systems.

The results reported in Table 5, showed very good performances for both kriging methodologies for almost all air quality indexes; nevertheless, there seems to be less accuracy in simulation of ozone levels, with general overestimation for both data fusion systems.

For NO_2, PM10 and PM2.5 NINFA shows a slight tendency to overestimate observed values (negative values of ME index), while FARM-PI to underestimate (positive values of ME index).

2.3.2 Comparison between observed and predicted values

Cross-validation datasets are available only for FARM-PI and NINFA data fusion systems. Therefore, to evaluate and compare the performance of all four data fusion systems, a qualitative analysis between the observed data and the simulated data was carried out (Denby et al, 2011).

The observed dataset was built considering all the stations available in the C1 dataset with data capture percentage not less than 75%; however, the simulated dataset was built extracting, for each system, fused values at station coordinates by means of bilinear interpolation on grid values.

The comparison results are presented in the following graphs in terms of scatter plot for all air quality indexes. All the adopted data fusion techniques provide good results:

- for the PM10 and PM2.5 not only the background stations concentrations, but also the traffic station ones are well reproduced;
- for annual mean of NO_2 background stations are quite well reproduced by all systems, even if it should be noted a clear tendency to overestimate lower values for CAMx-SLO and a slight tendency to underestimate for FARM-LO and FARM-PI; the concentrations of traffic stations are generally underestimate by all systems, except CAMx-SLO. These results are consistent with the spatial resolution of four data fusion systems and confirms that NO_2 levels at traffic station locations come from phenomena occurring at scales that cannot be solved with spatial resolution around 5 km.

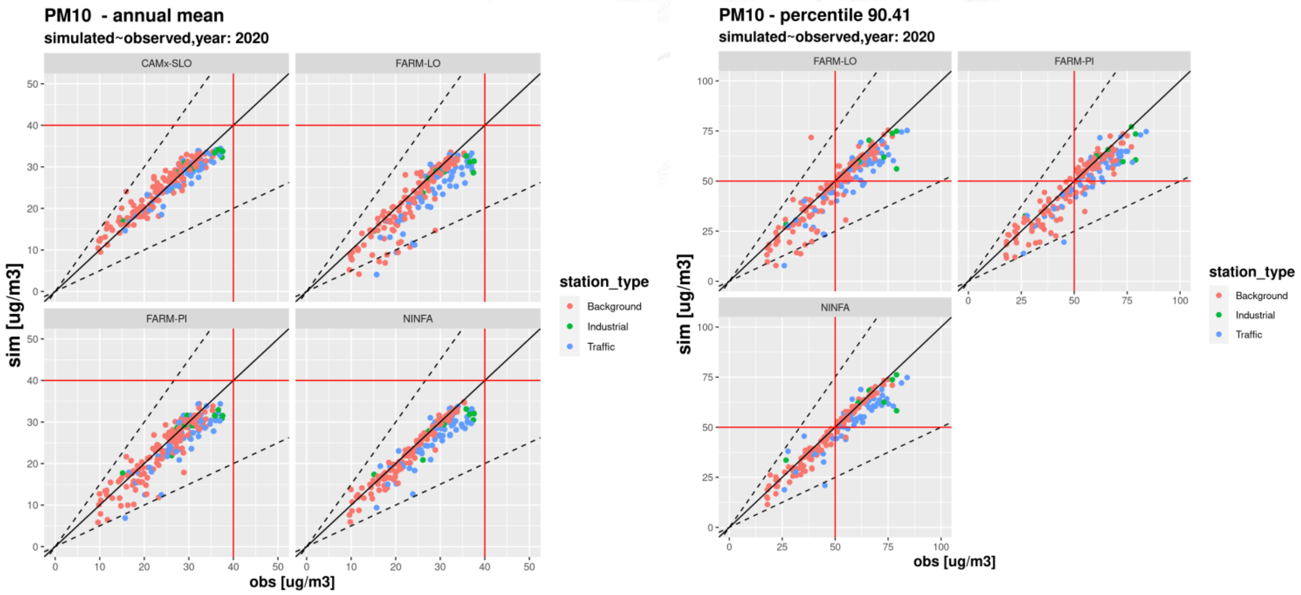


Figure 17 -. Left: PM10 annual mean, simulated vs observed values for the four data fusion systems. Right: PM10 90.41 percentile, simulated vs observed values for the NINFA, FARM-PI, FARM-LO data fusion systems. . The dotted lines represent the admitted relative uncertainty, while the red lines indicate the EU limit value (40 and 50µg/m3 respectively)

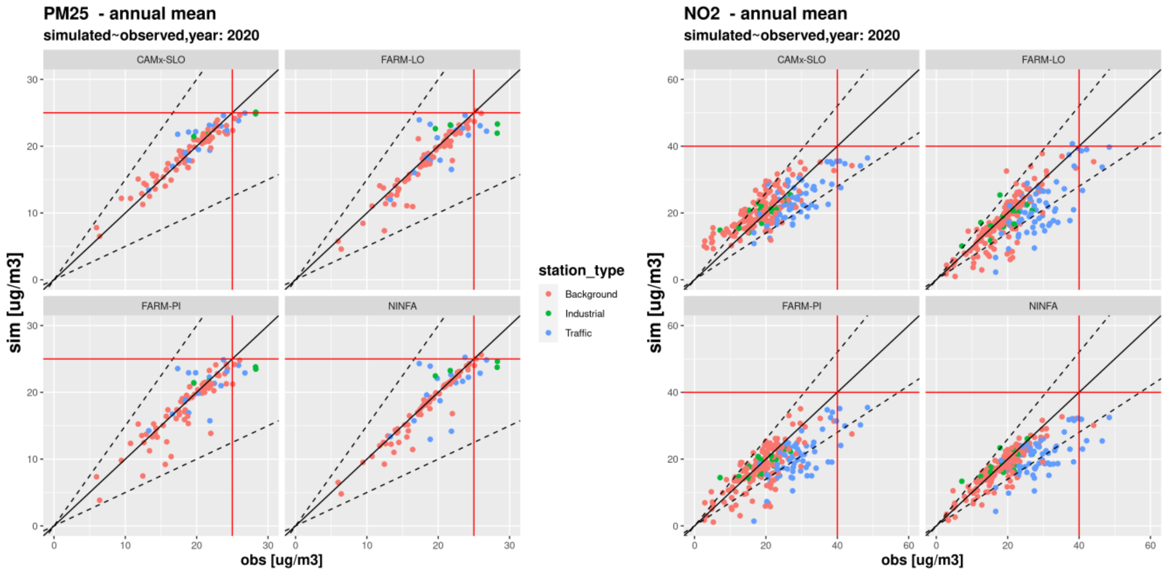


Figure 18 - Left: PM2.5 annual mean, simulated vs observed values for the four data fusion systems. Right: NO2 annual mean, simulated vs observed values for the four data fusion systems. The dotted lines represent the admitted relative uncertainty, while the red lines indicate the EU limit value (25 and 40 µg/m3 for PM2.5 and NO2 respectively)

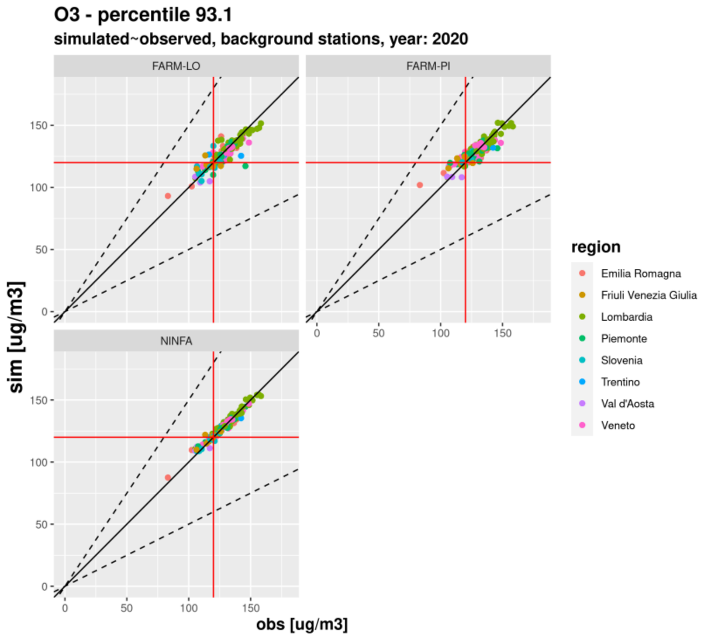


Figure 19 -. O3 93.10 percentile, simulated vs observed values for the NINFA, FARM-PI, FARM-LO data fusion systems. The dotted lines represent the admitted relative uncertainty, while the red lines indicate the EU target value (120 µg/m^3).

3 ASSESSMENT RESULT

3.1 PM10

The spatial distributions of the annual mean and 90.41 percentile of PM10 produced by the four data fusion systems (Figure 20 and Figure 21 respectively) are very similar to each other, showing the same main patterns. The areas with the highest concentrations are located between the Lombardia and Veneto plains and around the metropolitan agglomerations.

No model estimates annual average concentration beyond threshold value of 40 $\mu g/m^3$, while all the models report PM10 concentrations above the EU daily limit value for the flat area of the Po Valley and around Ljubljana agglomeration (only FARM-LO in this case).

Figure 22 shows boxplots of grid point distribution grouped by region for each data fusion system. The distributions are quite similar: NINFA and FARM-PI have very close median values, CAMx-SLO shows the higher median levels while FARM-LO the lowest. The largest differences between NINFA and FARM-PI, on one hand, and CAMx-SLO and FARM-LO on the other, occur in Slovenia and in the Alpine regions of Valle d'Aosta and Trentino. These small differences can be attributed to the fact that FARM-PI, NINFA and CAMx-SLO used very similar data fusion methodologies, but CAMx-SLO achieves a very fine resolution of the fused fields, while FARM-LO has implemented a conceptually different approach.

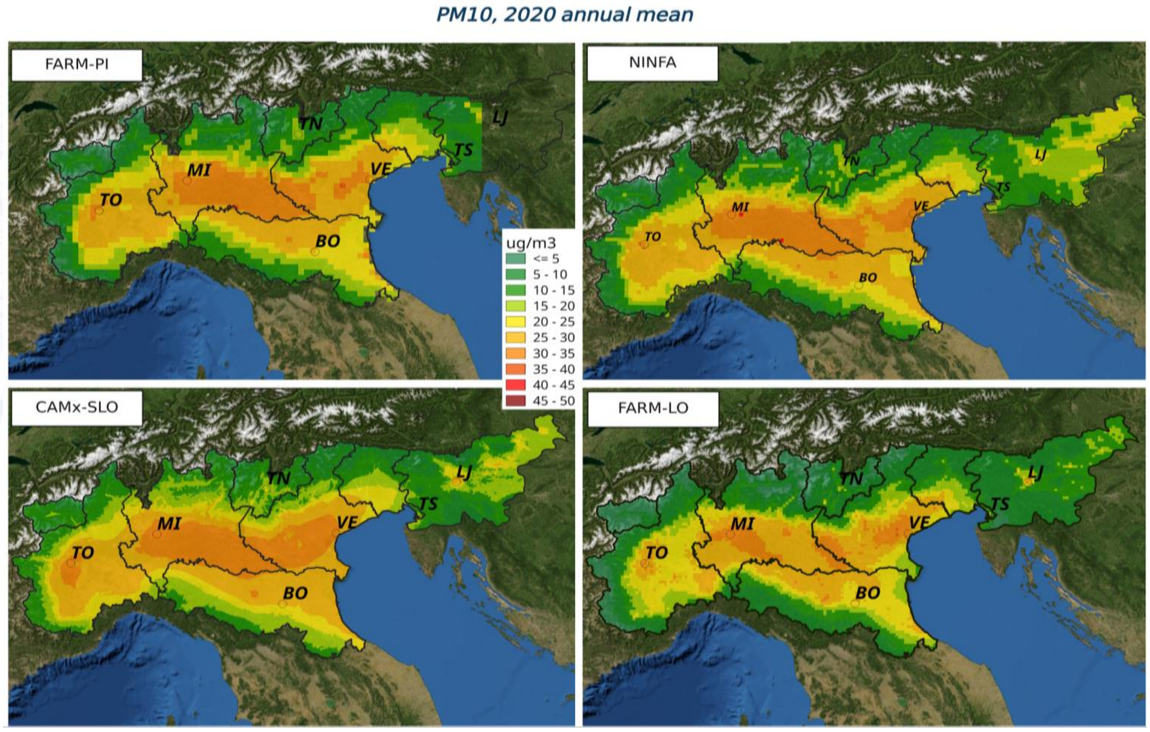


Figure 20.- Maps of PM10 annual mean produced by the four data fusion systems.

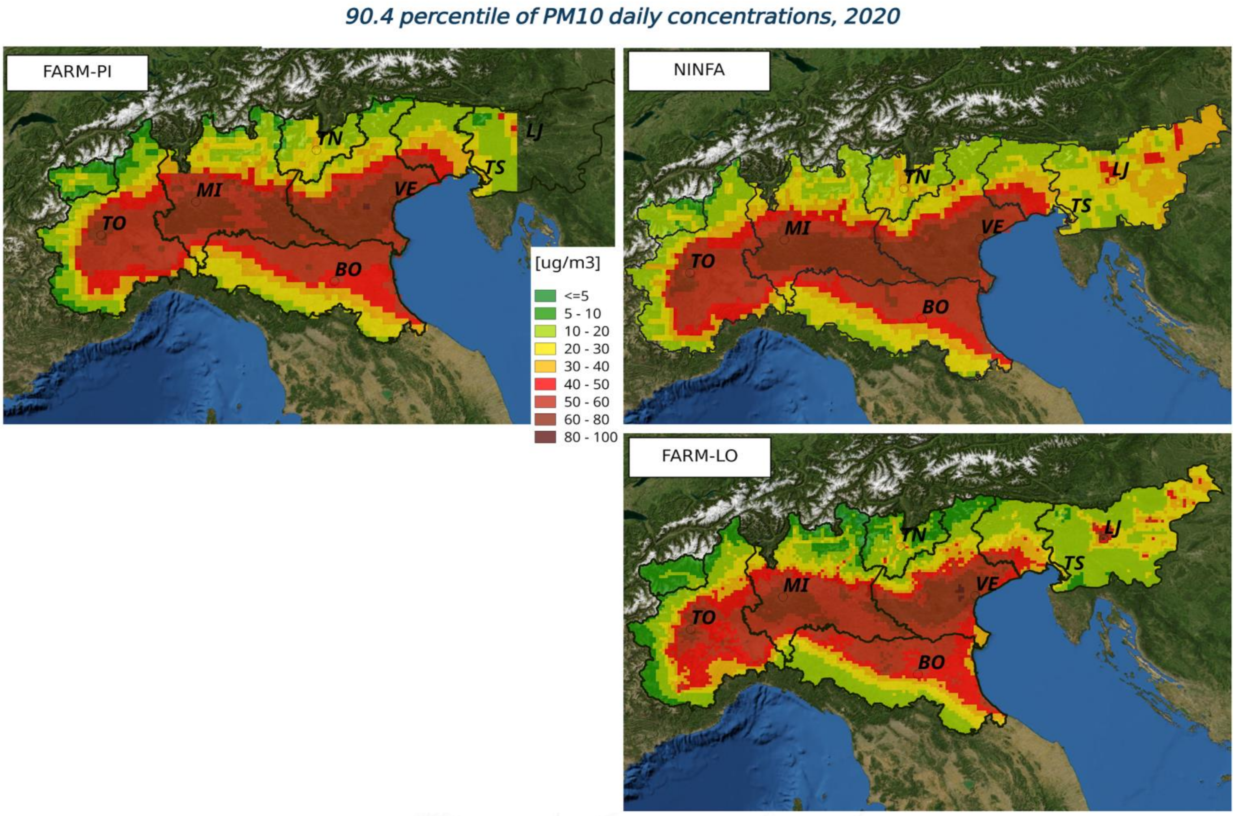


Figure 21- Maps of PM10 90.41 percentile produced by NINFA, FARM-PI and FARM-LO systems.

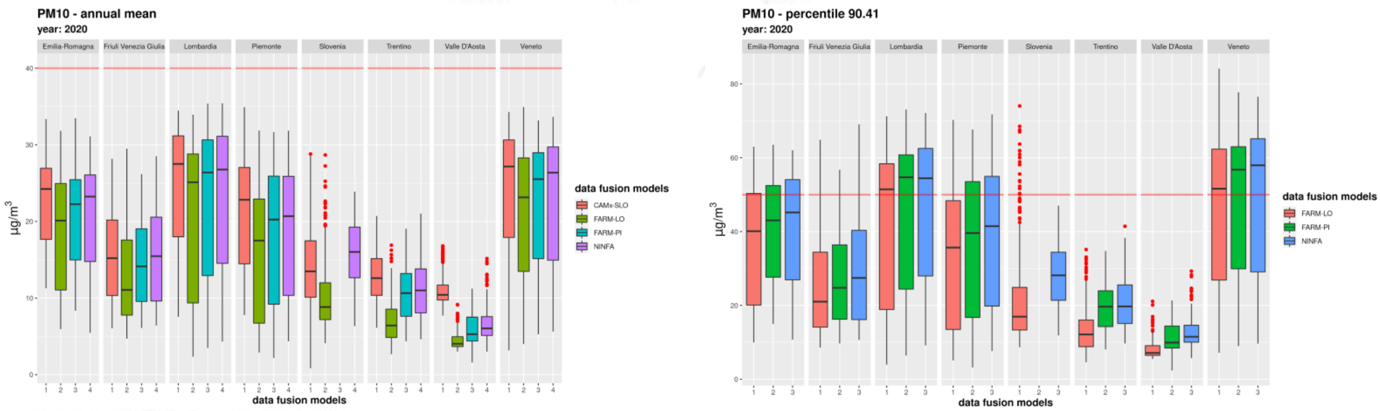


Figure 22 - Boxplots of grid point concentration distributions grouped by model and region. Left: PM10 annual mean; right percentile 90.4 of PM10 daily values. The red lines indicate the EU limit value (40 and 50µg/m3 respectively

3.2 PM2.5

All models agree in estimating average annual values of PM2.5 above 20 µg/m3 in the flat area of Veneto and Lombardia, while CAMx-SLO and FARM-LO also show exceedances around Ljubljana (Figure 23 and Figure 24). The PM2.5 concentration is beyond the EU limit value for the annual mean only in some small areas in Lombardia, Veneto and Slovenia.

The comparison between the spatial structure of the fields confirms what has already been highlighted for PM10. However, in Slovenia, the differences between the three models NINFA, FARM-LO and CAMx-SLO are not negligible (FARM-PI domain does not cover the whole Slovenian territory); this is due to lack of observed data of PM2.5 within the air quality dataset: one of the primary information is missing in the data fusion process, thus the differences between the CTMs are emphasized.

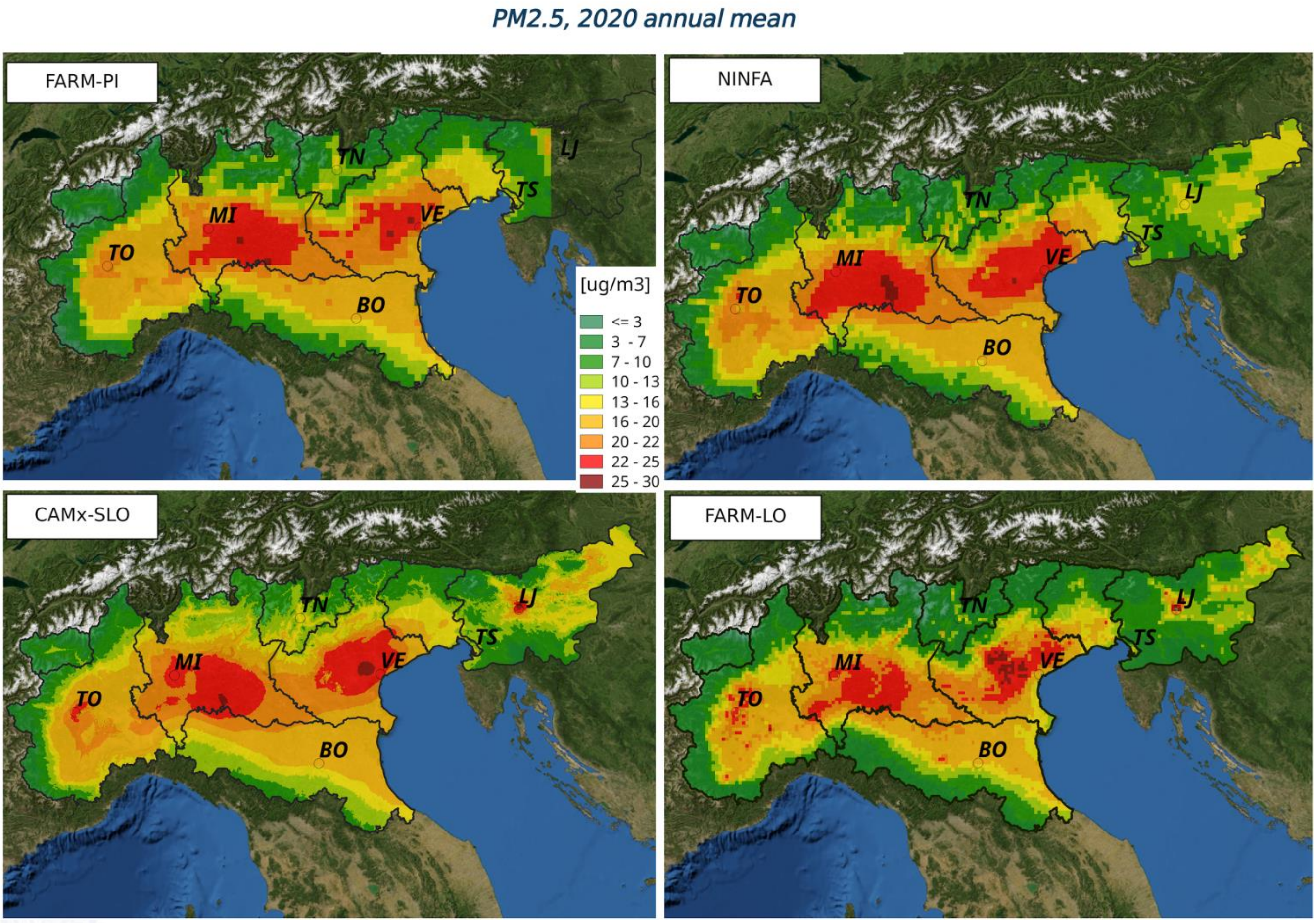


Figure 23 - Maps of PM2.5 annual mean produced by the four data fusion systems.

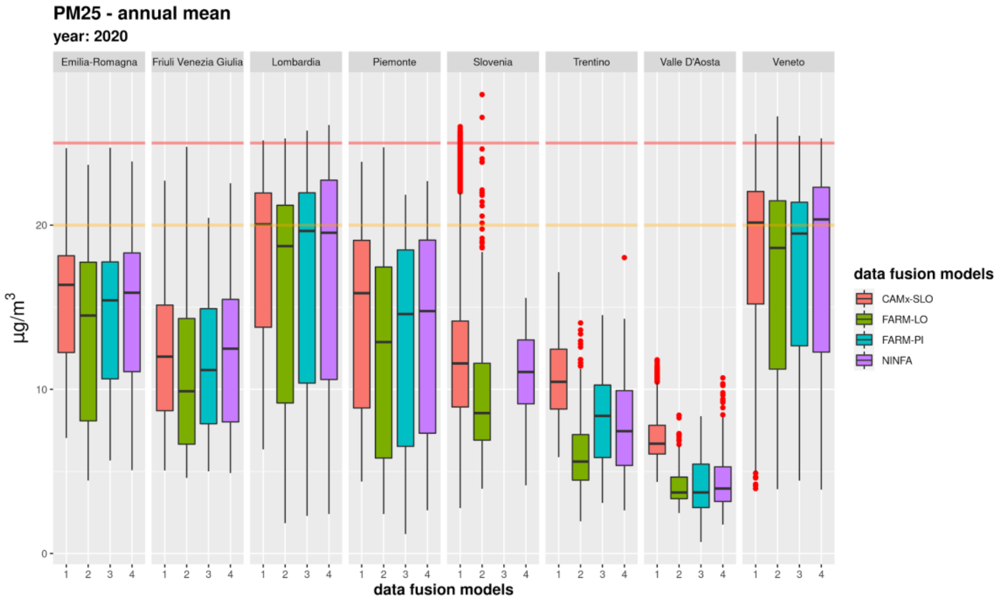


Figure 24 - PM2.5, annual mean: boxplots of grid point concentration distributions grouped by model and region. The red lines indicate the EU limit value for stage I (25 µg/m3), while the orange one for stage II (20 µg/m3).

3.3 NO2

Maps reported in Figure 25 show a quite similar spatial distribution of NO_2 annual mean: all the models identify the main urban agglomerations as areas with the highest values. Only one model out of four (FARM-LO/ARPA-LO) estimates the annual mean of NO_2 concentration above the EU limit value in a very small area in Milan and in Turin.

Figure 25 confirms the considerations expressed in paragraphs 3.1 and 3.2 regarding the differences between the spatial distributions of the various data fusion systems. It is possible to highlight the location of the main highways, in particular from the results of the ARPA LO and CAMx-SLO modelling systems.

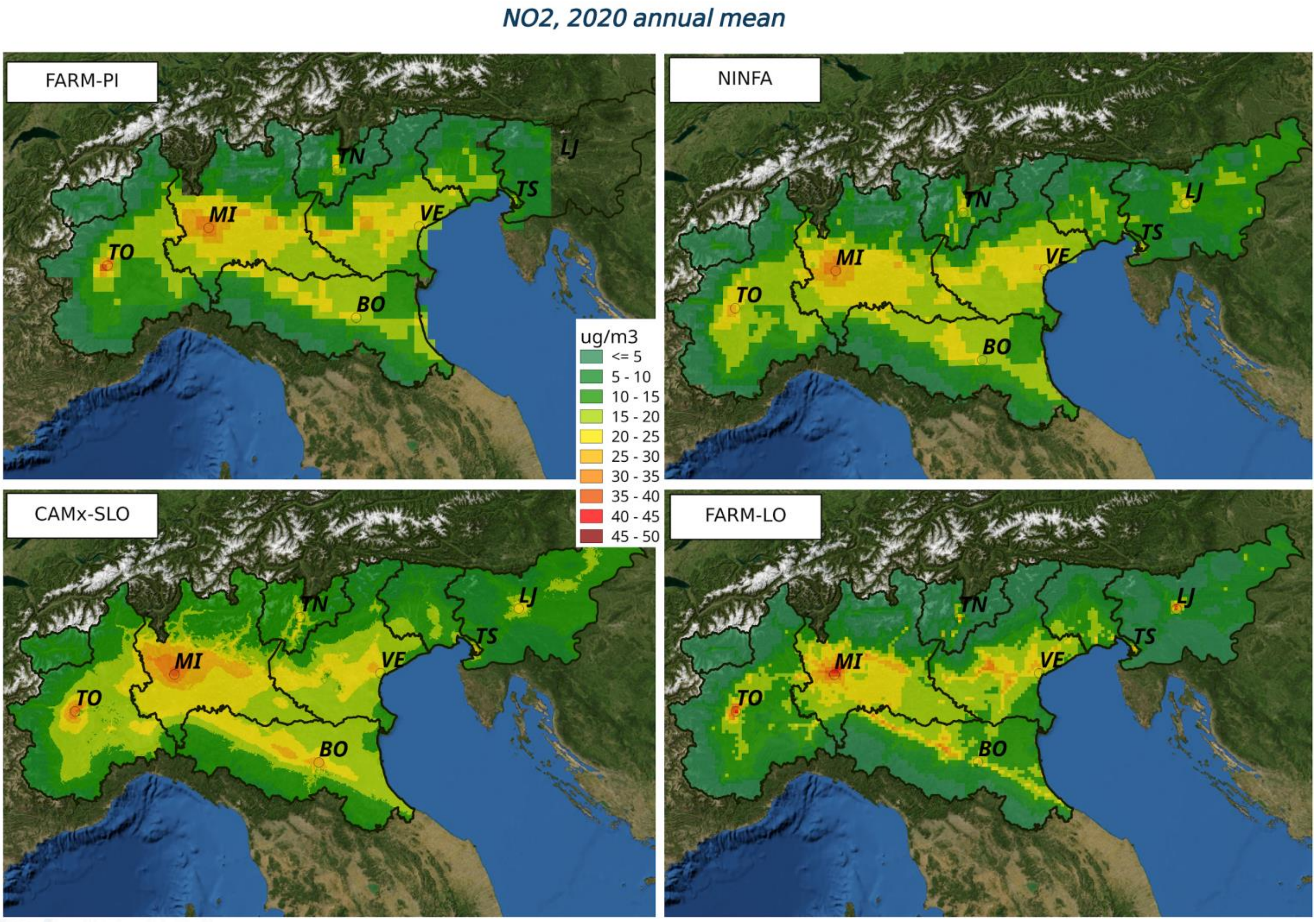


Figure 25 - Maps of NO_2 annual mean produced by the four data fusion systems.

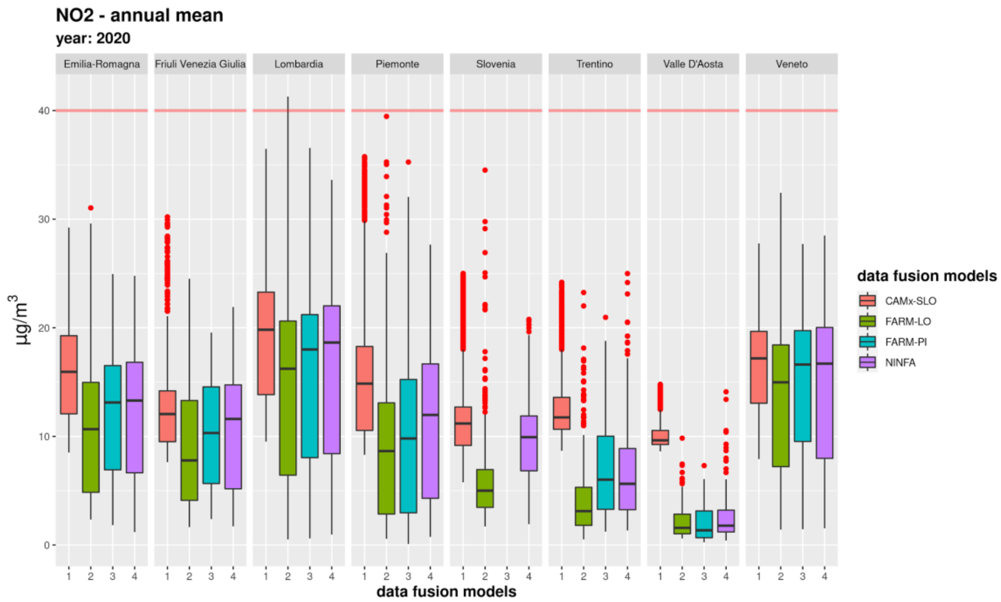


Figure 26 - NO_2 annual mean: boxplots of grid point concentration distributions grouped by model and region. The red lines indicate the EU limit value (40 µg/m³)

3.4 O3

The maps in Figure 27 show the spatial distribution of O_3 maximum daily 8-hour mean concentration values. All the models estimate concentration above the 120 μg/m^3 threshold, implying an exceedance of the target value in almost the entire Po Valley.

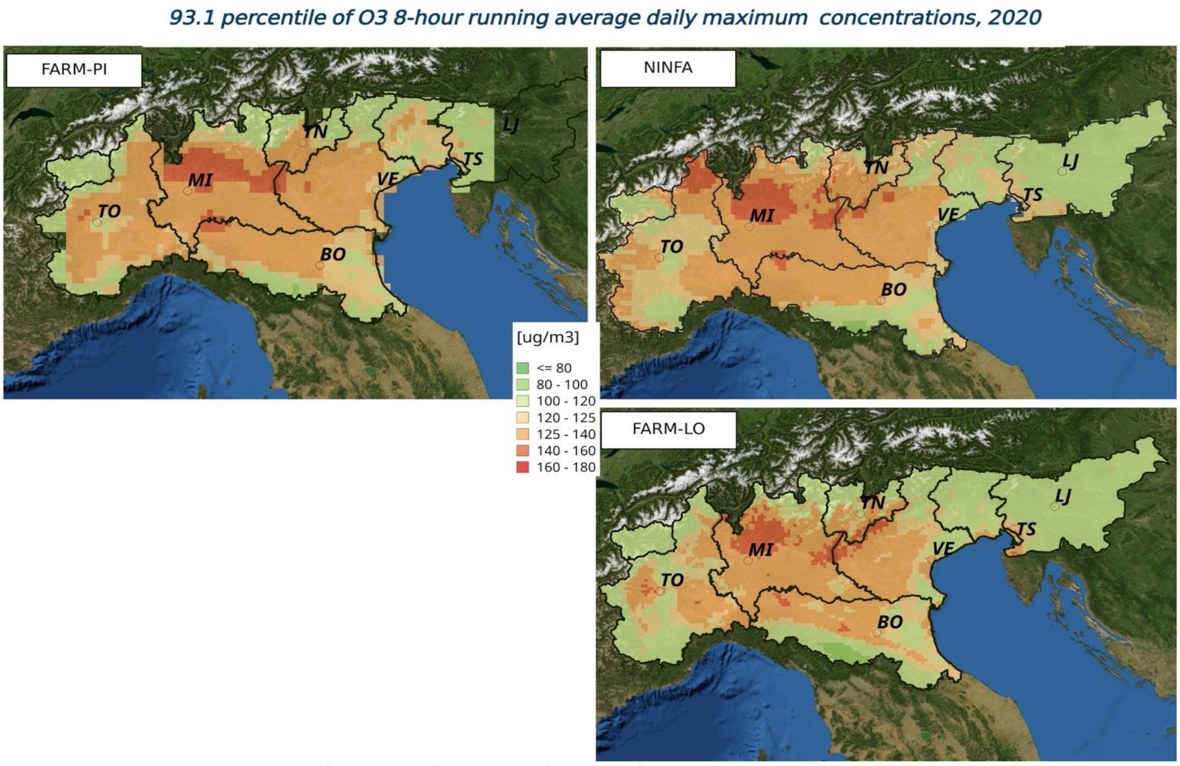


Figure 27 - Maps of O3 93.1 percentile produced by NINFA, FARM-PI, FARM-LO systems.

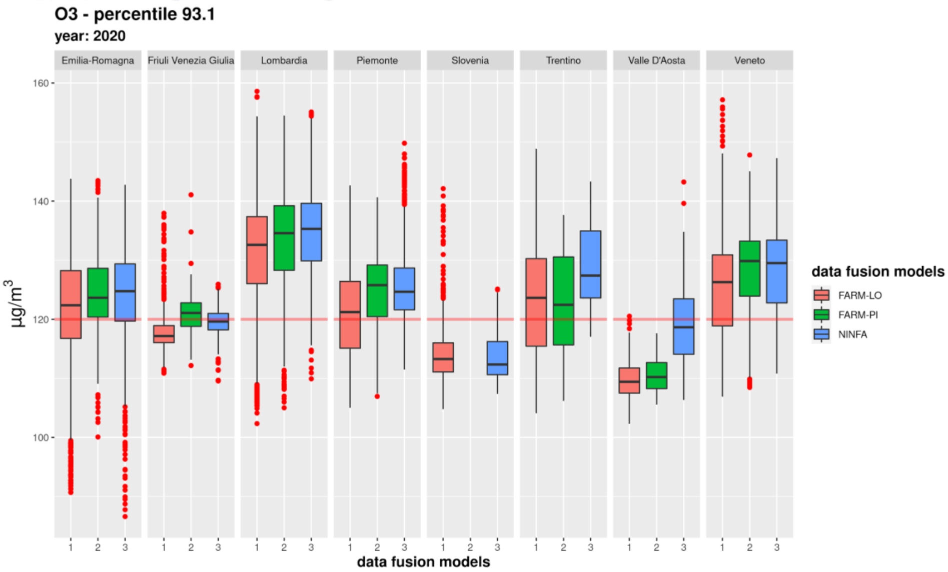


Figure 28 - O3 93.1 percentile: boxplots of grid point concentration distributions grouped by model and region. The red lines indicate the EU target value (120 µg/m3)

3.5 ATTAINMENT STATUS/POPULATION EXPOSURE

The following Figure 29 and Figure 30 show the maps of the five air quality indicators produced by the four data fusion systems with a traffic light classification that highlights the attainment green areas and the nonattainment red areas. In summary, it can be state that:

- there are no nonattainment areas for the annual mean of PM10 (Figure 29 left), as also confirmed by the monitoring data reported in Appendix A;
- there are no nonattainment areas for the annual mean of NO2 (Figure 30, right); only one model predicts two very small nonattainment near Milan and near Turin; the monitoring data, as show in Appendix A, record exceedances only in traffic stations and in one background station;
- for the percentile 90.41 of PM10 the nonattainment area extends across the whole flat area of the Po Valley and, only for one data fusion model, around Ljubljana agglomeration; the monitoring data in Appendix A show exceedances in Piemonte, Lombardia, Emilia-Romagna, Veneto and Friuli Venezia Giulia regions;
- there are few nonattainment red areas located in Lombardia, Veneto and around Ljubljana agglomeration for PM2.5 annual mean regarding EU limit of 25 µg/m3; instead considering the limit of 20 µg/m3 the nonattainment area (yellow areas in Figure 30, left) extends across the large part of Lombardia and Veneto, significant part of Piemonte and minority part of Friuli Venezia Giulia and Emilia-Romagna; this scenario, except for Friuli Venezia Giulia, is described by monitoring data reported in Appendix A;
- for the percentile 93.1 of O3 the nonattainment area extends across almost the whole Po Valley, as also confirmed by the monitoring data reported in Appendix A (Please note that the legal definition of the target value considers not only 1 year but the average over 3 years).

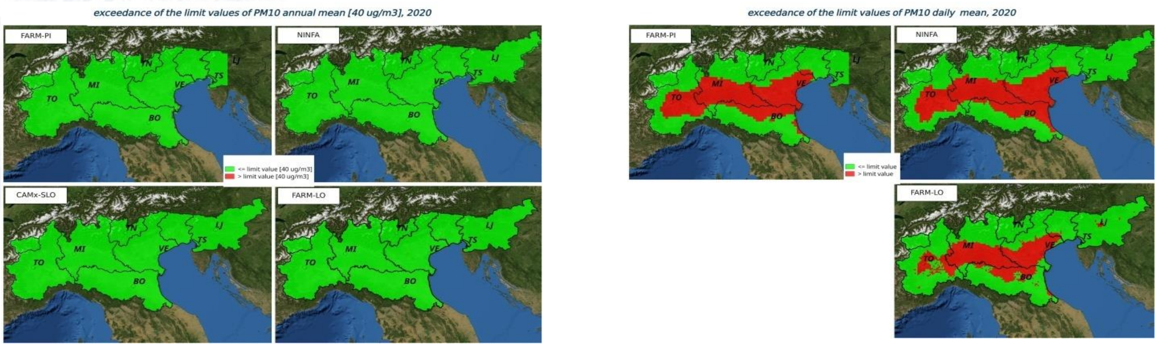


Figure 29 - Attainment (green) and nonattainment (red) areas for PM10 annual mean (left) and PM10 percentile 90.41 (right).

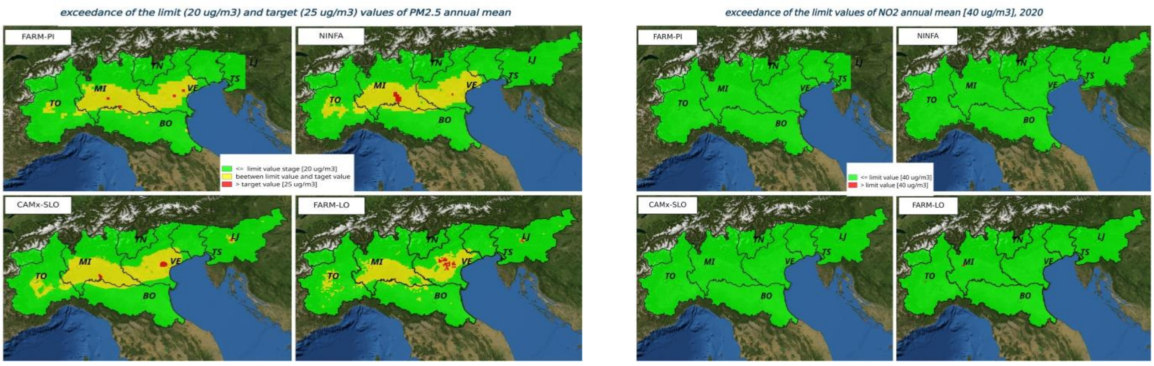


Figure 30 - Attainment (green) and nonattainment (red) areas for PM2.5 annual mean (left) and NO2 annual mean (right). In the PM2.5 maps yellow areas indicate attainment regarding the EU limit of 25 µg/m3 and nonattainment for EU limit of 20 µg/m3.

Annual values of the five air quality indexes considered in this report, as estimated by the four considered chemistry-transport models, are compared with the population data on the same grids, i.e. on the grid of each model, in order to assess the population exposure. Population data have been provided by the Italian Statistical Institute ISTAT for the Italian regions, on the census units, referring to 2011, and for Slovenia by the Statistical Office of the Republic of Slovenia SURS, on a regular grid of 100m resolution, referring to 2019. Population data have been split (in the Italian regions only, given the irregularity of the census units) and reaggregated (both in Italy and in Slovenia), proportionally to the surface, in order to estimate the population residing in each cell of each model.

Finally, for each air quality index, each model and each considered region, the population exposed to different index values was estimated, assuming that each inhabitant is exposed to the concentration that was estimated in the cell in which it resides. In particular, the population exposed to values exceeding the thresholds established by EU legislation has been estimated.

There are some differences in the estimates of the various models, in particular for NO_2 (for which the most marked spatial gradients correspond to the most densely populated areas) and for Slovenia (for which different monitoring stations have been included in the datasets for the data fusion).

According to all models, in year 2020 no citizens were exposed to values beyond the threshold for the PM10 annual average.

Only one model out of four estimates that there were inhabitants exposed to values above the threshold for the NO_2 annual average (about 650,000 in Lombardia and Piemonte together). The other three models remain under the limits across their domain.

The models agree in estimating that a large part of the population of Lombardia and Veneto and a significant part of the population of Piemonte was exposed to average PM2.5 annual values above 20 µg/m^3. Only a minority part of the population of Friuli Venezia Giulia and Emilia-Romagna, and no inhabitants of Valle d'Aosta and Trentino Alto Adige are exposed for this index, while for Slovenia there is little agreement between the three models that cover that area.

About eight million from Lombardia, four million from Veneto, three million from Piemonte, two million from Emilia-Romagna, 200,000 from Friuli Venezia Giulia and - according to one of the models - even 400,000 Slovenians were exposed to more than 35 daily PM10 exceedances in 2020.

Almost ten million Lombards, about four and a half million from Veneto, three and a half million from Piemonte, almost four million from Emilia-Romagna, about half a million from Friuli Venezia Giulia, half a million from Trentino Alto Adige and over one hundred thousand Slovenes and even some thousands of inhabitants of the Valle d’Aosta were exposed to more than 25 daily ozone exceedances in 2020.

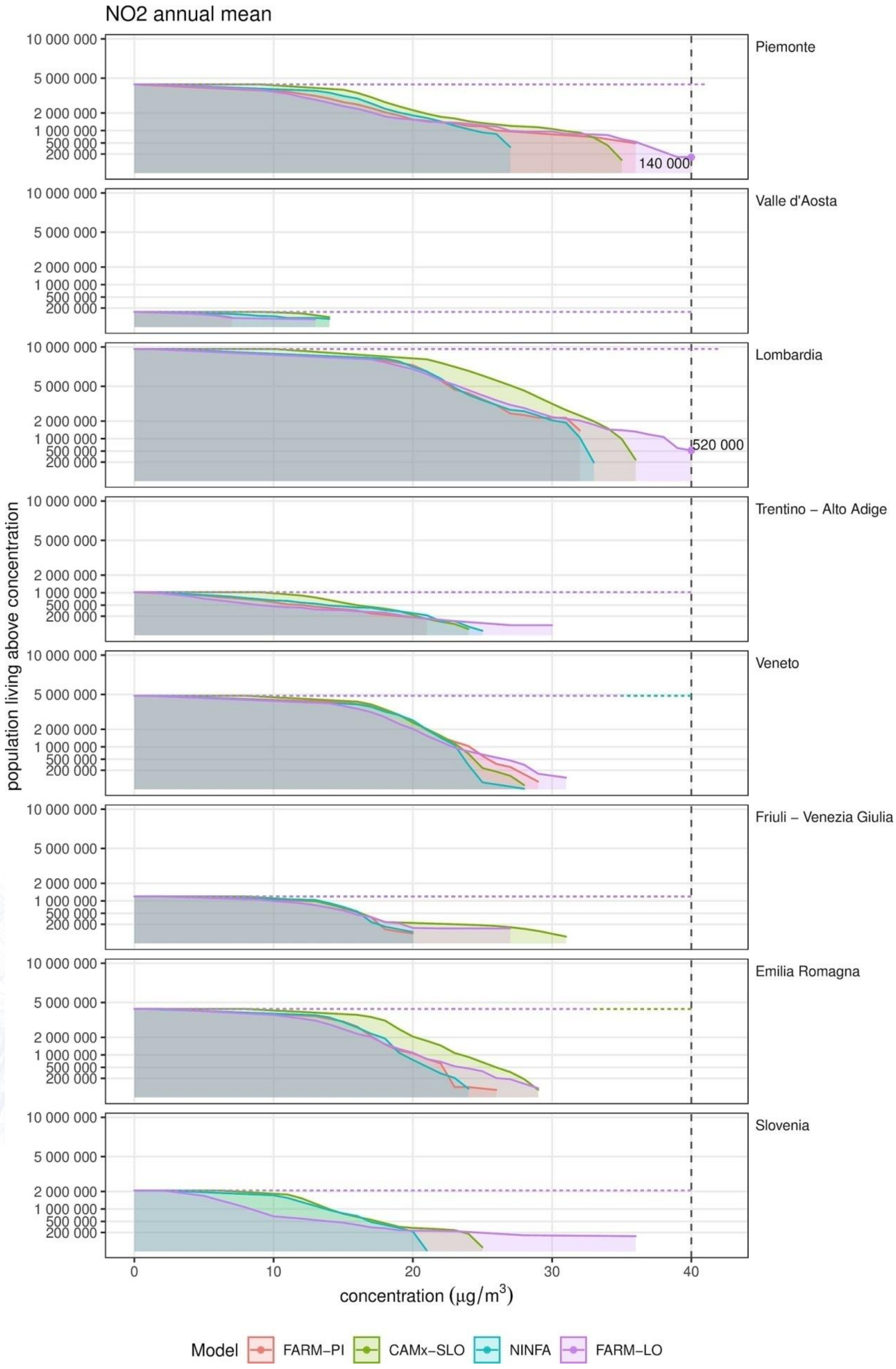


Figure 31 - Population exposure estimate for NO2 annual mean.

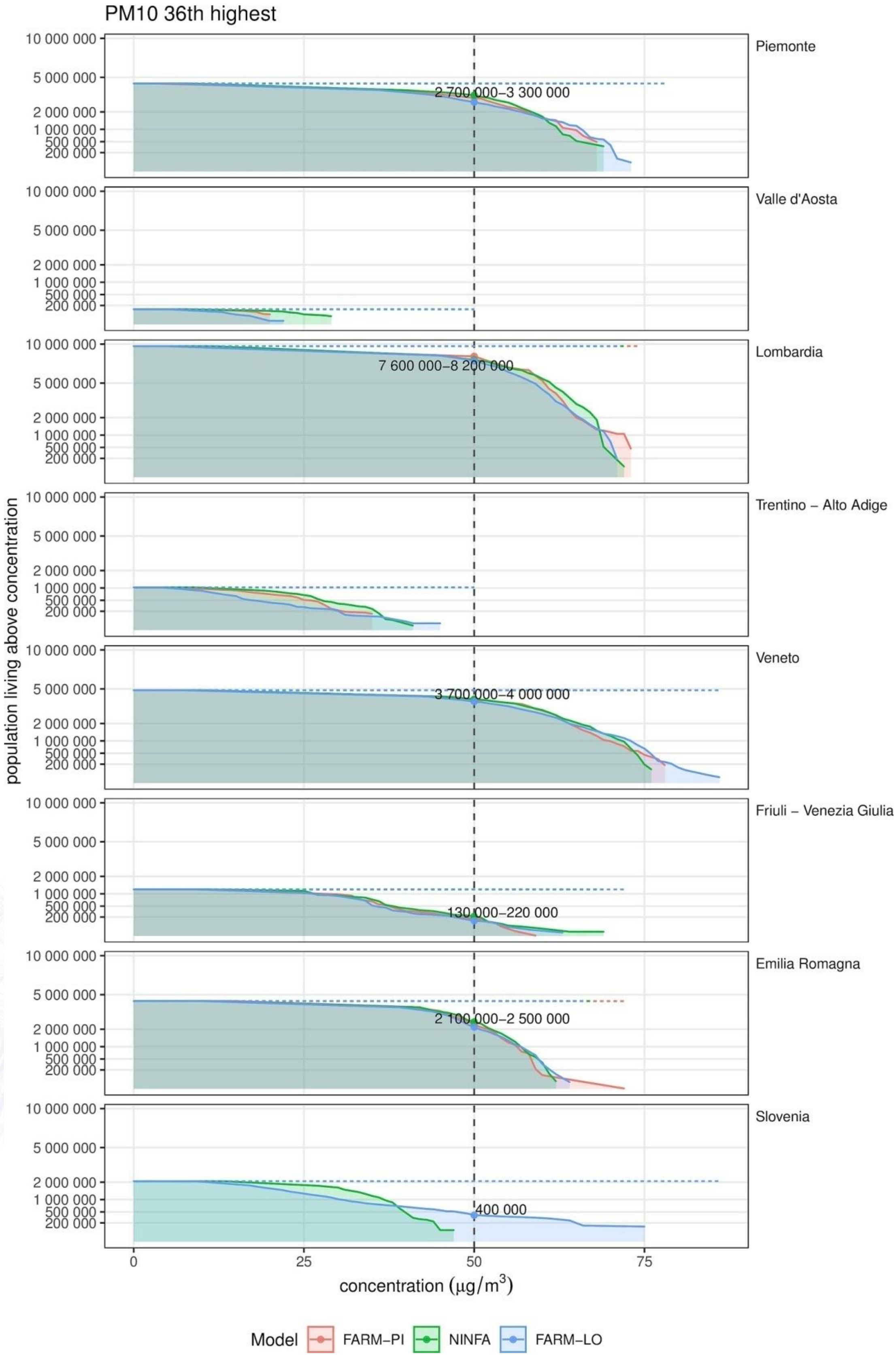


Figure 32 - Population exposure estimate for percentile 90.41 of PM10.

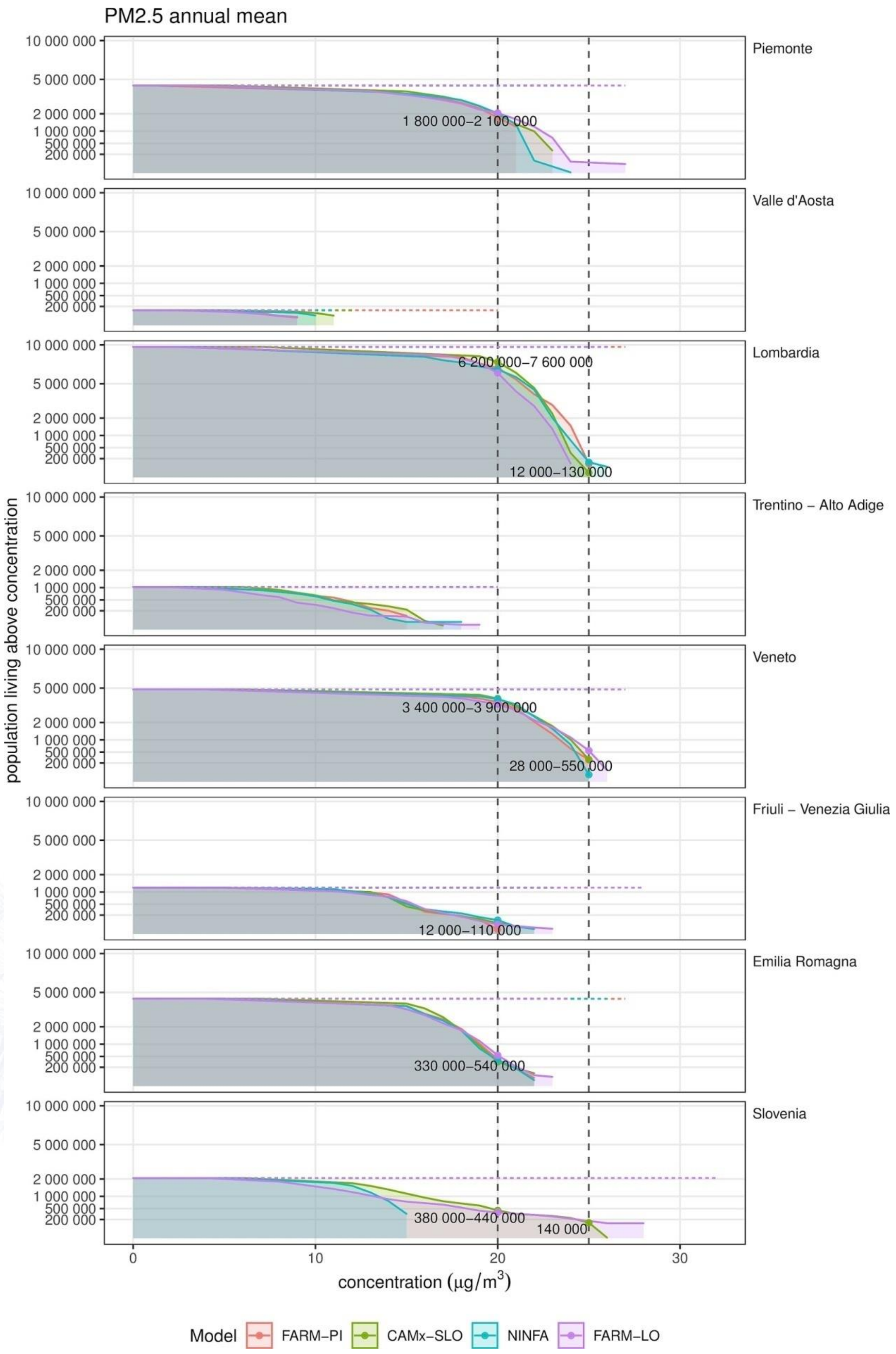


Figure 33 - Population exposure estimate for PM2.5 annual mean.

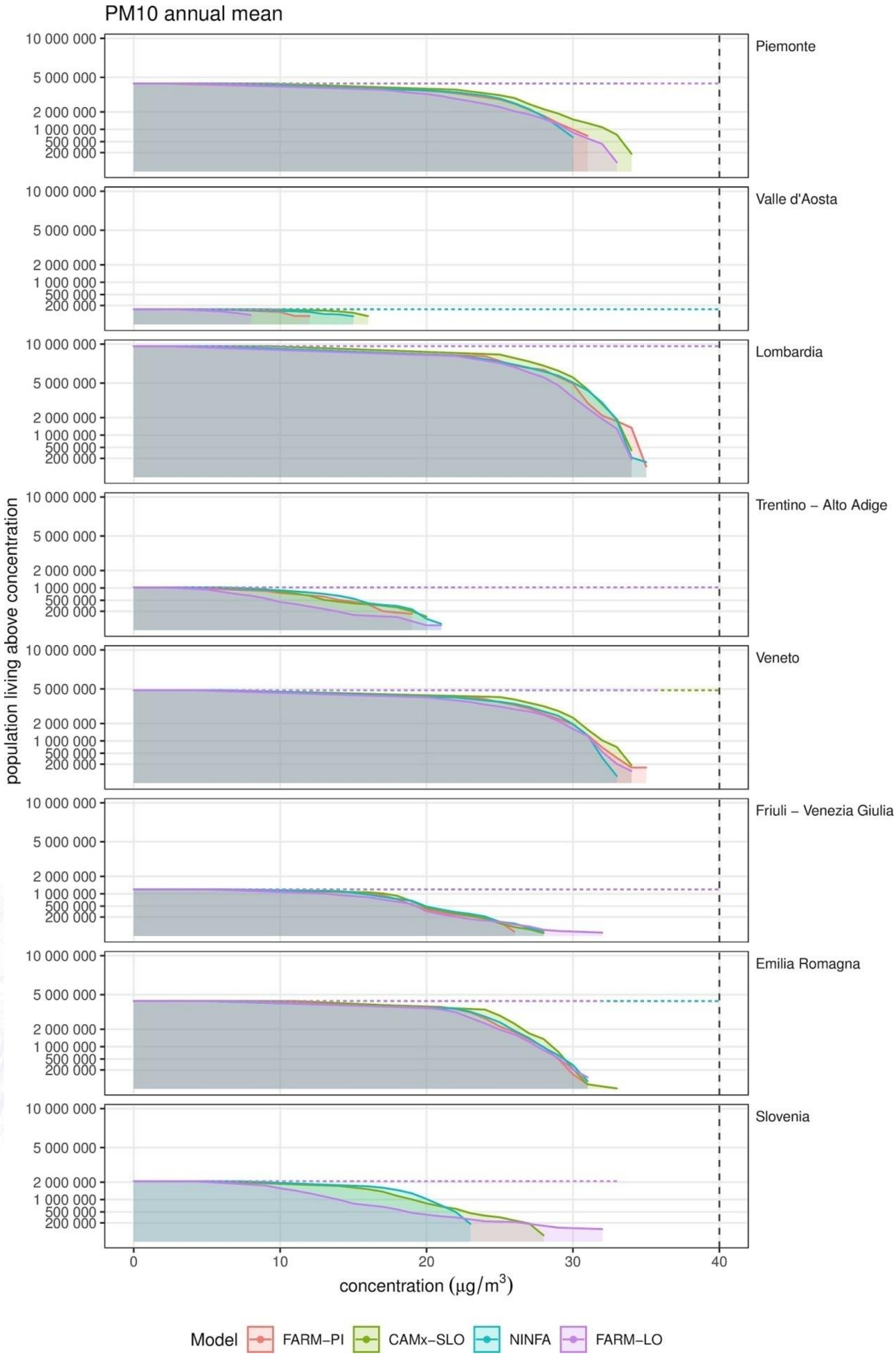


Figure 34.- Population exposure estimate for PM10 annual mean.

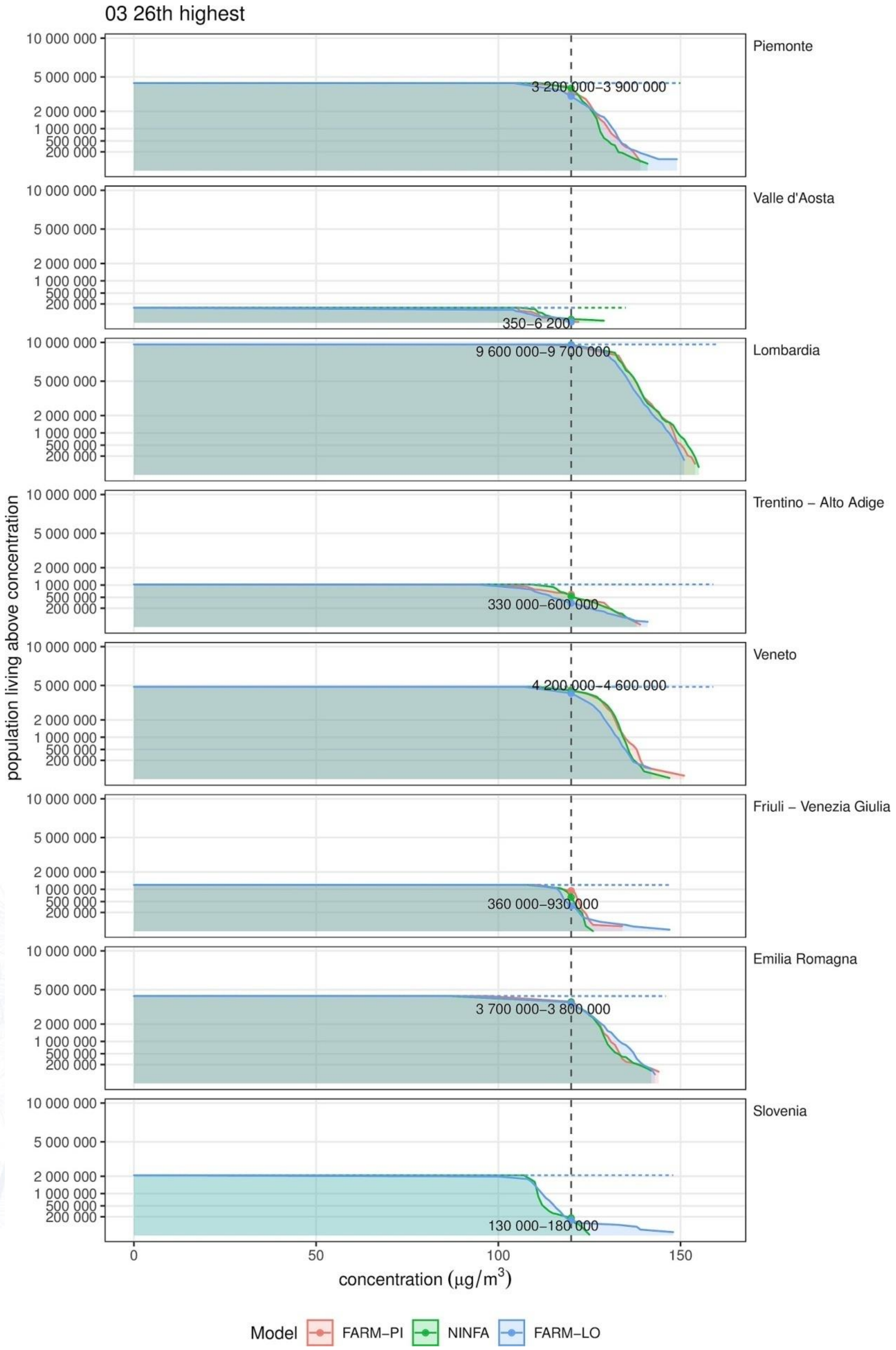


Figure 35 - Population exposure estimate for ozone percentile 93.1

4 DISCUSSION

This first report provides a synthetic view on the state of air quality in Po Valley and Slovenia for year 2020 and examines PM10, PM2.5, nitrogen dioxide and ozone, which are the pollutants whose values more frequently exceed legislation thresholds.

The assessment was carried out with a state-of-art approach that uses data fusion techniques to integrate information coming from air quality monitoring networks and CTM modelling systems. Among all the CTM running operational within the PREPAIR project, four modelling and data fusion systems have been used for the 2020 assessment.

No data fusion system estimates PM10 annual average concentrations beyond the threshold value of 40 $\mu g/m^3$, while all the models report PM10 concentrations above the EU daily limit value for the flat area of the Po Valley, thereby a large percentage of the population is exposed to values beyond the daily limit value.

Only a small percentage of populations in some few areas in Lombardia, Veneto and Slovenia is exposed to values beyond the stage I limit for the annual average of PM2.5 (25 $\mu g/m^3$), while a large percentage, especially in Veneto and Lombardia is exposed to average annual values of PM2.5 above the stage II limit (20 $\mu g/m^3$).

All the data fusion systems identify the main urban agglomerations as areas with the highest values of NO2 concentrations. Only one model out of four estimates the annual mean average of NO2 concentration above the EU limit value in a very small area around Milan and Torino.

All the data fusion systems show ozone concentration above the 120 $\mu g/m^3$ threshold, implying an exceedance of the target value in almost the entire Po Valley and more than 24 million of inhabitants exposed to value beyond EU limit.

However, it should be pointed out that 2020 was a particular year for air quality: the health crisis caused by COVID-19 pandemic and the consequent containment measures adopted have had a considerable impact on air quality. The results of the analyzes on the lockdown period provided an opportunity to verify the validity of these assessments and compare them with the data on the reduction of emissions and concentrations in an unprecedented condition of generalized contraction of human activities.

The assessments of the emission variations relating to the lockdown period can in fact be compared with the target reductions of the plans.

This comparison indicates that:

- NO_X emissions had a decrease comparable to that envisaged by the plans, with a weekly maximum of the order of 40% (variations from week to week and trends are similar in the various regions). This decrease is mainly attributable to the reduction in vehicle traffic which reached 80% for light vehicles and 50 - 60% for heavy commercial vehicles.
- PM10 (primary) emissions had a maximum weekly decrease of the order of 20%, significantly lower than that envisaged by the plans (-40%), with variations from week to week and trends diversified in the various regions. The smaller decrease in PM10 emissions is mainly attributable to space heating; the differences between the regions are mainly due to the different consumption of woody biomass.
- Starting from the first week of May, at the beginning of phase 2 (DPCM April 26, 2020) there is a reversal trend for both pollutants and emissions progressively increase as activities resume.
- Ammonia emissions are not reduced, as agricultural / livestock activities, which emit more than 90% of ammonia, did not change during the lockdown. Small variations (approximately -1%) are due to the reduction in circulating vehicles (catalytic converters). Small variations (approximately -1%) are due to the reduction in circulating vehicles (catalytic converters).

In line with the framework of emissions, gaseous concentrations (NO_2, NO, benzene) in March-May 2020 there have been very significant decreases compared to the average period 2016-2019. PM10 mass, however, shows a less reduction. with concentration values within the variability of previous years (2016-2019), highlighting a decoupling with gaseous pollutants. These data once again highlight the complex dynamics of PM and of the relationships between emissions of precursors and transport, diffusion and physico-chemical processes that determine the formation of the secondary PM, which constitutes a significant part (of the order of 70%) of PM10 in the Po basin. This dynamic, even with reduced emissions, is strongly influenced by weather conditions and can lead to episodes of exceeding of the limit values, although of much lower intensity compared to that which would occur in the usual emission conditions.

Two different chemical models (NINFA and FARM-PI) of transport and dispersion were used which allow to estimate the percentage reductions of the real scenario compared to a hypothetical scenario in which the emissions did not change ("NO-LOCKDOWN" scenario). The results of the two models are consistent with each other and indicate that for nitrogen dioxide (NO_2 the reductions at the end of March reach median values on the Po Valley of about 35-50%, while for PM10 the reductions are smaller, more differentiated by geographical area, more variable in the various weeks, but still reach a median reduction of 15- 30%. In other words, in the absence of the lockdown, in the same weather conditions, the NO_2 concentration was about twice and the concentration of PM would have been higher by about 1/3.

The main hypotheses to explain the causes of the relatively less effective reduction of PM compared to NO_2 are:

- the primary PM10 emissions have not been reduced sufficiently, in particular due to heating emissions;
- some precursors (mainly NH_3) did not decrease. The mixture of precursor gases could have remained such as to maintain a high secondary production potential even in the presence of varied proportions (less NO_X, constant NH_3);
- the high insulation in March increased the production of secondary PM of photochemical origin.

These results seem to confirm the correctness of the strategy of the air quality plans adopted by the Regions and Autonomous Provinces of the Po Basin, as well as the interregional agreements, focused on multi-sectorial and multi-polluting large-scale interventions.

A more in-depth analysis of these effects can be found in the previous three reports “Covid-19 and air quality in the Po Valley" published within the PREPAIR project.

It should be noted that the purpose of this report is informative, it does not replace the annual air quality assessment and reports required by EU directives and decisions (2008/50/EU and 2011/850/EU).

Finally, it must be underlined that although the four CTM systems used have different setup (resolution, boundary condition, meteorological data and data fusion technique), the outputs are very similar to each other showing the reliability of the assessment contained in the report.

5 GLOSSARY

ALADIN	a numerical weather prediction system (Aire Limitée Adaptation dynamique Développement InterNational)
APPA/ARPA/Arpae	Environment protection Agency of one of the Italian regions or autonomous provinces
AQF	Air Quality Forecast
ARSO	Slovenian environment agency
CAMS	Copernicus Atmosphere Monitoring Service
CAMx	Comprehensive Air Quality Model with Extensions
COSMO	Consortium for Small-scale Modelling
CTM	chemistry-transport model
ECMWF	European Centre for Medium-Range Weather Forecasts
EMEP	European Monitoring and Evaluation Programme
FARM	Flexible Air quality Regional Model
IC/BC	initial conditions/boundary conditions
INEMAR	INventario EMissioni ARia
ISPRA	Italian Institute for Environmental Protection and Research (Istituto Superiore per la Protezione e la Ricerca Ambientale)
KED	kriging with external drift
NINFA	Northern Italy Network to Forecast Aerosol pollution
NWP	numerical weather prediction
PREPAIR	Po Regions engaged to Policies of Air
SAPR	chemical mechanism, part of the chemistry-transport models (originally developed by the Statewide Air Pollution Research Center)

SNAP	emitting sources classification (originally defined in the framework of the "Significant New Alternatives Policy" program of US-EPA)
SNPA	the Italian national system for environmental protection (Sistema nazionale per la protezione dell'ambiente)
WRF	Weather Research and Forecasting model

6 BIBLIOGRAPHY

- Berrocal VJ, Gelfand AE, Holland DM (2012). Space-time data fusion under error in computer model output: an application to modeling air quality. Biometrics 68(3): 837–48.
- Box GEP, Cox DR (1964) An analysis of transformations. J R Stat Soc Ser B 26:211–246
- Bratseth, A. M. (1986). Statistical interpolation by means of successive corrections. *Tellus A*, *38*(5),
- Brewster, K. A., 1996: Implementation of a Bratseth analysis scheme including Doppler radar. 15th Conference on WAF, AMS, 92-95
- Brewster K. (1997) 3-D analysis with ARPS Data Assimilation System: ADAS, Supplemental S-1, ARPS User's Guide, CAPS, University of Oklahoma.
- Cressie NAC (1993). Statistics for Spatial Data. 2nd. New York: Wiley
- Daley R (1991). Atmospheric data analysis, Roger Daley, Cambridge Atmospheric and Space Science Series, Cambridge University Press
- 2011/850/EU: Commission Implementing Decision of 12 December 2011 laying down rules for Directives 2004/107/EC and 2008/50/EC of the European Parliament and of the Council as regards the reciprocal exchange of information and reporting on ambient air quality (notified under document C(2011) 9068) GU L 335 del 17.12.2011, pagg. 86–106, http://data.europa.eu/eli/dec_impl/2011/850/oj
- Daley, R., Atmospheric Data Analysis, Cambridge Univer sity Press, New York, 1991
- Denby B, Georgieva E, Lükewille A (2011) The application of models under the European Union’s Air Quality Directive: a technical reference guide. Technical Report 10/2011, European Environmental Agency, Copenhagen
- Development Core Team (2010) A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna
- Diggle, P.J. and Ribeiro Jr. P. J. (2007) Model-based Geostatistics, Springer
- 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 GU L 152 dell' 11.6.2008, pagg. 1–44, http://data.europa.eu/eli/dir/2008/50/oj
- Elbern, H., et al. "Emission rate and chemical state estimation by 4-dimensional variational inversion." Atmospheric Chemistry and Physics 7.14 (2007): 3749-3769.

- ENVIRON International Corporation. CAMx Ozone Particulates TOxics User's Guide, Comprehensive Air Quality Model With Extensions Version 6.2. Novato, California. https://www.camx.com/
- Gariazzo C, Silibello C, Finardi S, Radice P, Piersanti A, Calori G, Cecinato A, Perrino C, Nusio F, Cagnoli M, Pelliccioni A, Gobbi GP, Di Filippo P (2007) A gas/aerosol air pollutants study over the urban area of rome using a comprehensive chemical transport model. Atmos Environ 41:7286–7303
- Ghigo, S.; Bande, S; Ciancarella, L.; Mircea, M.; Piersanti, A.; Righini, G; Baldasano, J.M.; Basagaña,X; Cadum, E. (2017) Mapping air pollutants at municipality level in Italy and Spain. Mapping air pollutants at municipality level in Italy and Spain in support to health impact evaluations. Air Quality, Atmosphere &Health, DOI 10,1007/s 11869-017+0520-x
- Giorcelli M., Bande S., Muraro M., Milelli M. (2013) Operational multiscale modelling system for air quality forecast. COSMO NEWSLETTER, vol. 13, p. 58-63
- Ignaccolo R, Ghigo S, Bande S (2013) Functional zoning for air quality. Environ Ecol Stat 20(3):321–339
- Lahoz W, Schneider P (2014). Data assimilation: making sense of Earth Observation. Front Environ Sci 2(16): 1–28
- Pagowski, Mariusz, et al. "Three-dimensional variational data assimilation of ozone and fine particulate matter observations: some results using the Weather Research and Forecasting—Chemistry model and Grid-point Statistical Interpolation." Quarterly Journal of the Royal Meteorological Society 136.653 (2010): 2013-2024.
- Pielke Sr, Roger A. Mesoscale meteorological modeling. Academic press, 2013.
- Ribeiro JR, Diggle PJ (2001) geoR: a package for geostatistical analysis. R-NEWS 1(2):15–18
- Silibello C, Bolignano A, Sozzi R, Gariazzo C (2014) Application on chemical transport model and optimized data assimilation methods to provide air quality assessment. Air Quality Atmosphere & Health https://doi.org/10.1007/s11869-014-0235-1
- Wackernagel H (2003) Multivariate geostatistics: an introduction with applications. Springer, Berlin
- Yu S, Eder B, Dennis R, Chu S, Schwartz S. 2006. New unbiased symmetric metrics for evaluation of air quality models. Atmospheric Science Letters 7: 26–34.
- Zhan, X., Houser, P. R., Walker, J. P., & Crow, W. T. (2006). A method for retrieving high-resolution surface soil moisture from hydros L-band radiometer and radar observations. IEEE Transactions on Geoscience and Remote Sensing, 44(6), 1534-1544.

- Stortini, M.; Arvani, B.; Deserti, M. Operational Forecast and Daily Assessment of the Air Quality in Italy: A Copernicus-CAMS Downstream Service. Atmosphere 2020, 11, 447. https://doi.org/10.3390/atmos11050447
- Mailler S., L. Menut, D. Khvorostyanov, M. Valari, F. Couvidat, G. Siour, S. Turquety, R. Briant, P. Tuccella, B. Bessagnet, A. Colette, L. Letinois, and F. Meleux, CHIMERE-2017: from urban to hemispheric chemistry-transport modeling , Geosci.ModelDev.,10,2397-2423, https://doi.org/10.5194/gmd-10-2397-2017
- Raffaelli, K.; Deserti, M.; Stortini, M.; Amorati, R.; Vasconi, M.; Giovannini, G. Improving Air Quality in the Po Valley, Italy: Some Results by the LIFE-IP-PREPAIR Project. Atmosphere 2020, 11, 429. https://doi.org/10.3390/atmos11040429

APPENDIX A: Air Quality Data

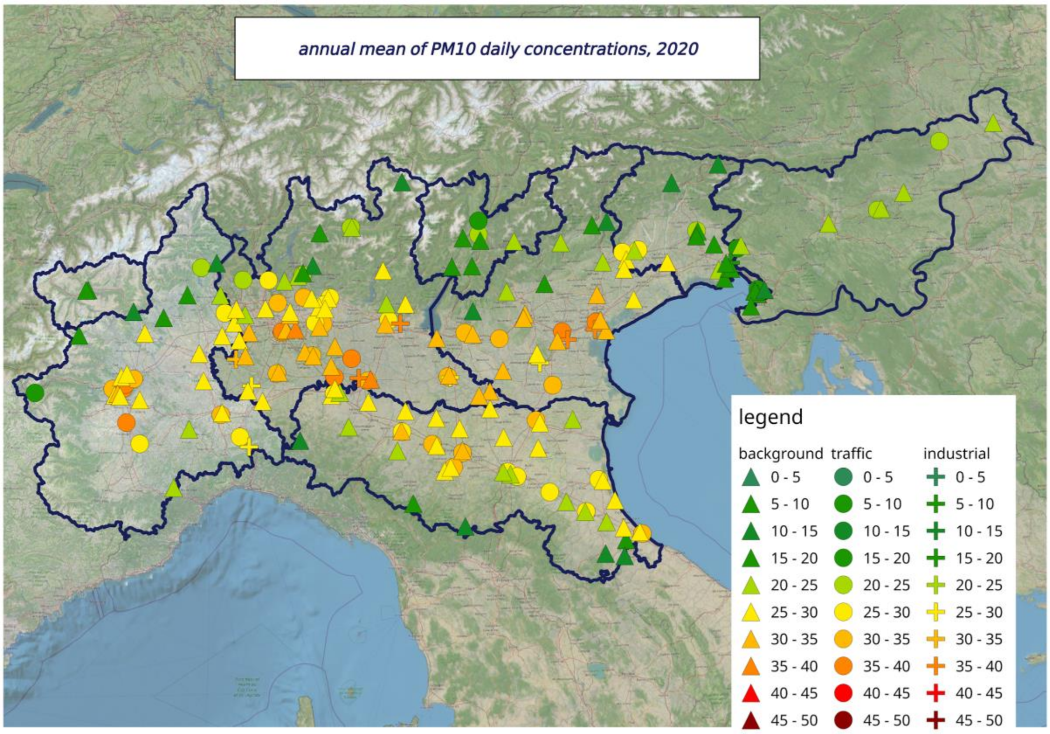


Figure A 1 - PM10 annual mean: maps of observed data, monitoring stations are grouped by station classification

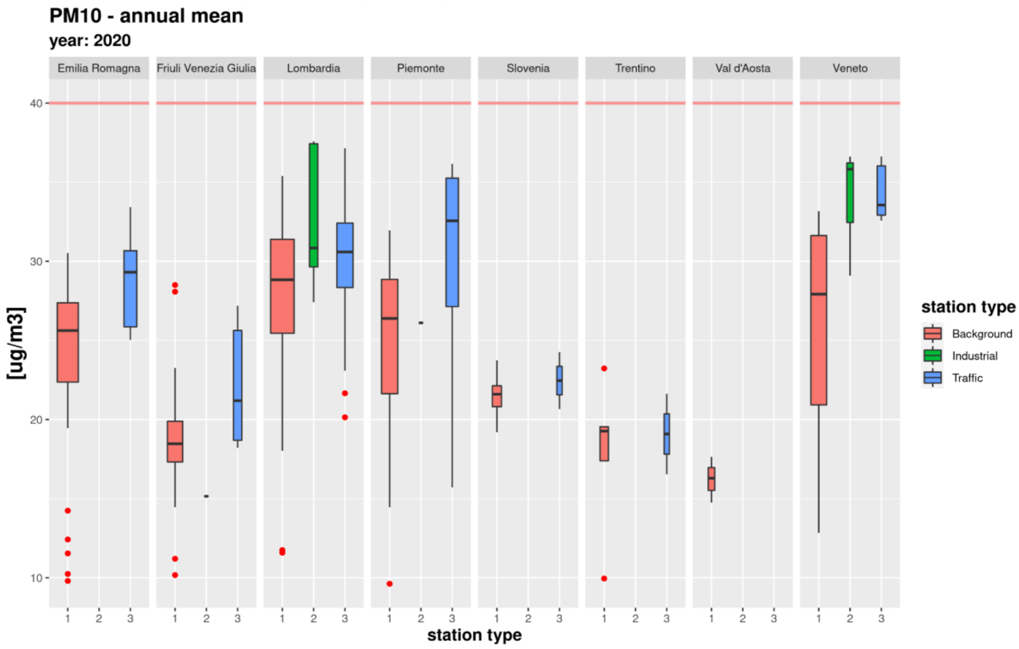


Figure A 2 - PM10 annual mean: boxplots of observed data grouped by station type and region.

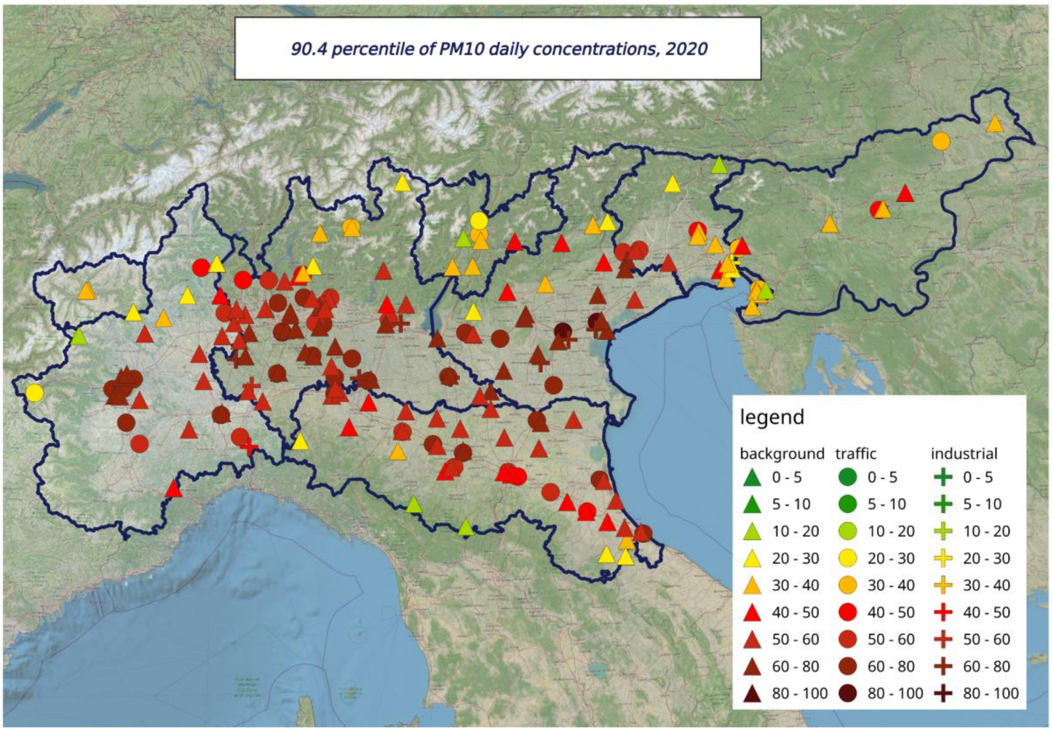


Figure A 3 - PM10 percentile 90.41: maps of observed data, monitoring stations are grouped by station classification

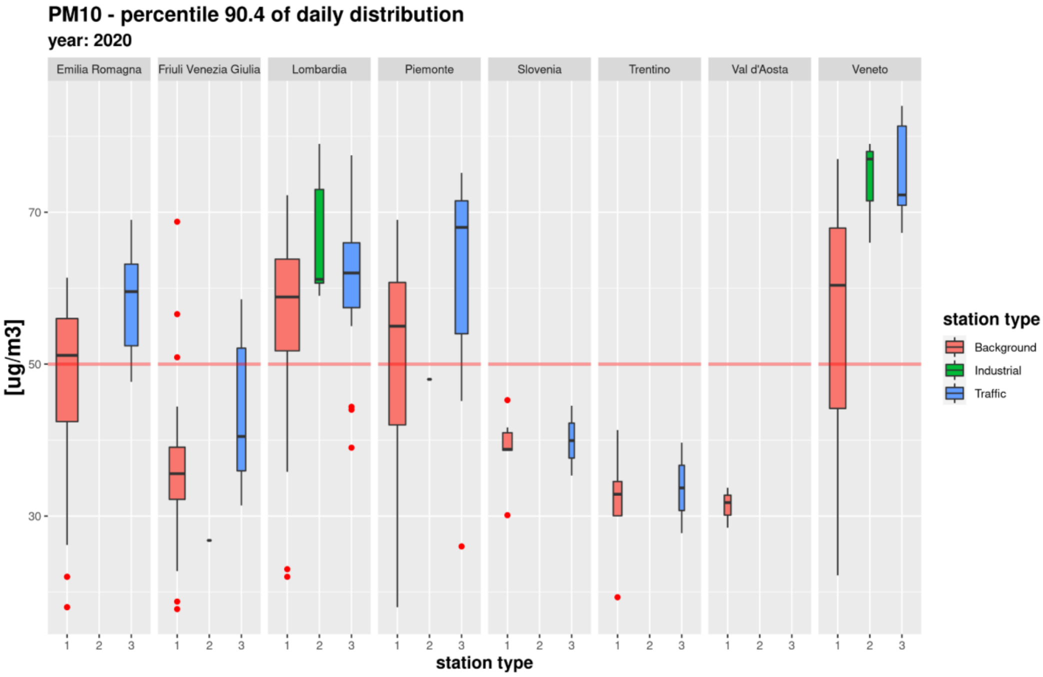


Figure A 4.- PM10 percentile 90.41:boxplots of observed data grouped by station type and region.

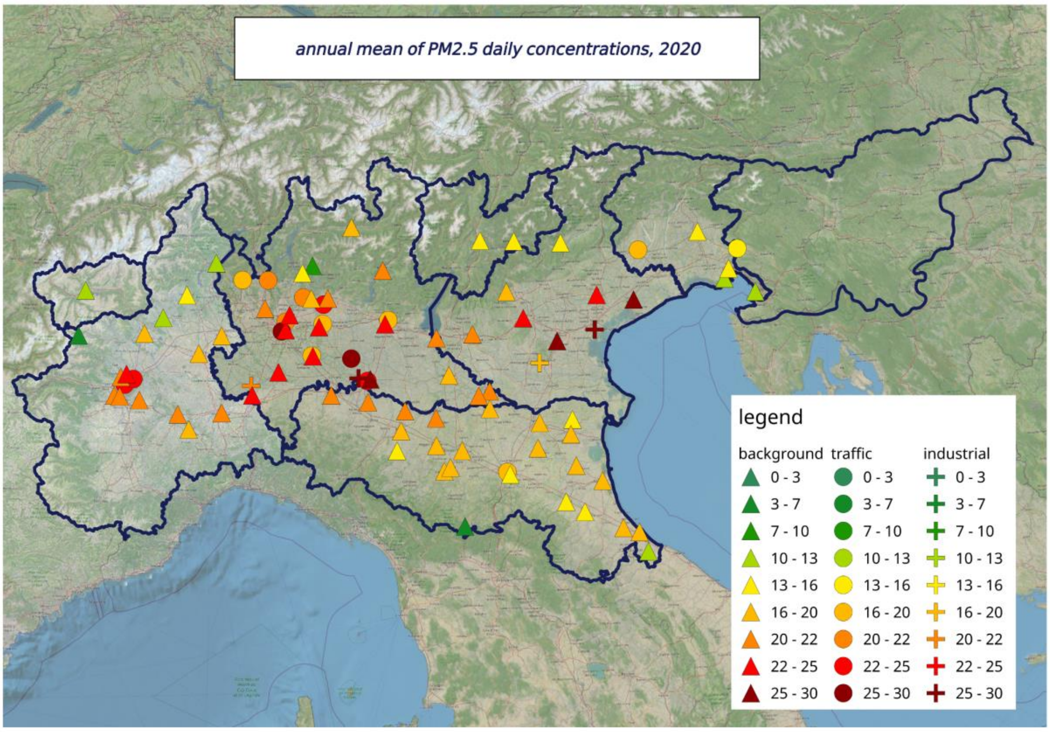


Figure A 5.- PM2.5 annual mean: maps of observed data, monitoring stations are grouped by station classification

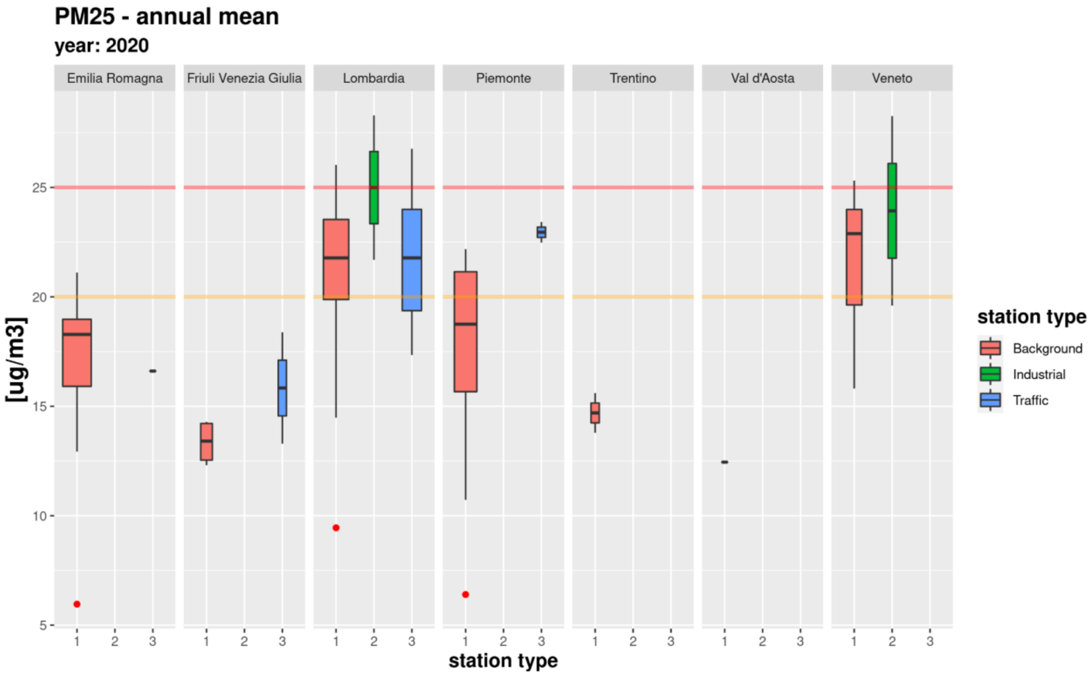


Figure A 6 - PM2.5 annual mean:boxplots of observed data grouped by station type and region.

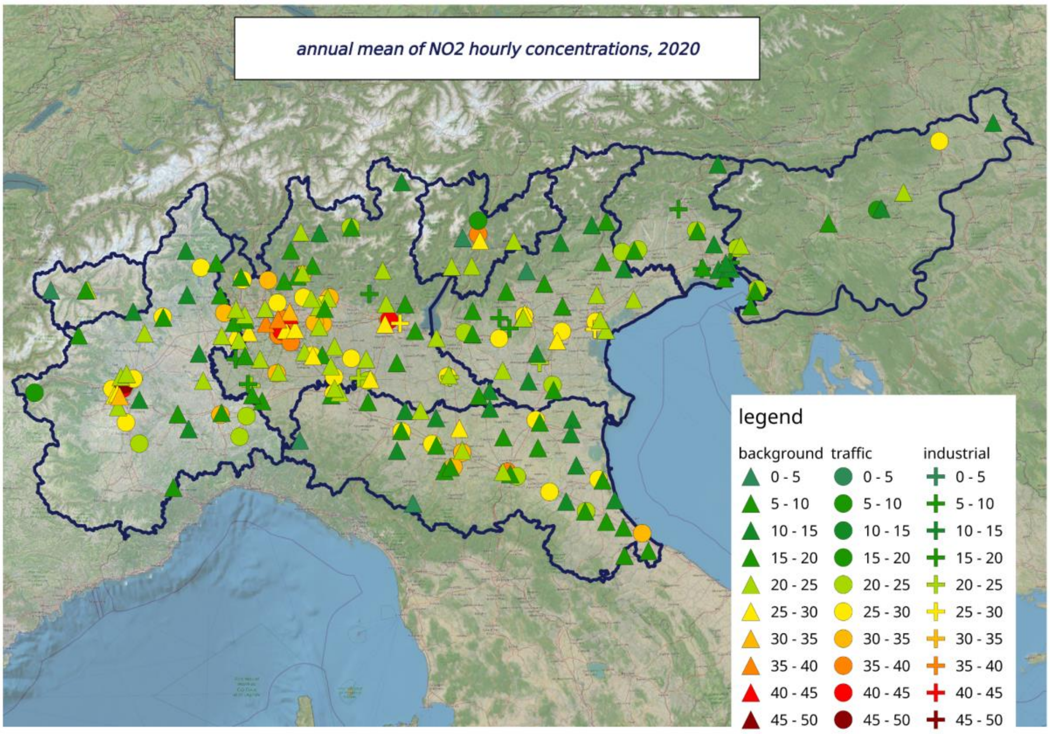


Figure A 7-. NO2 annual mean: maps of observed data, monitoring stations are grouped by station classification

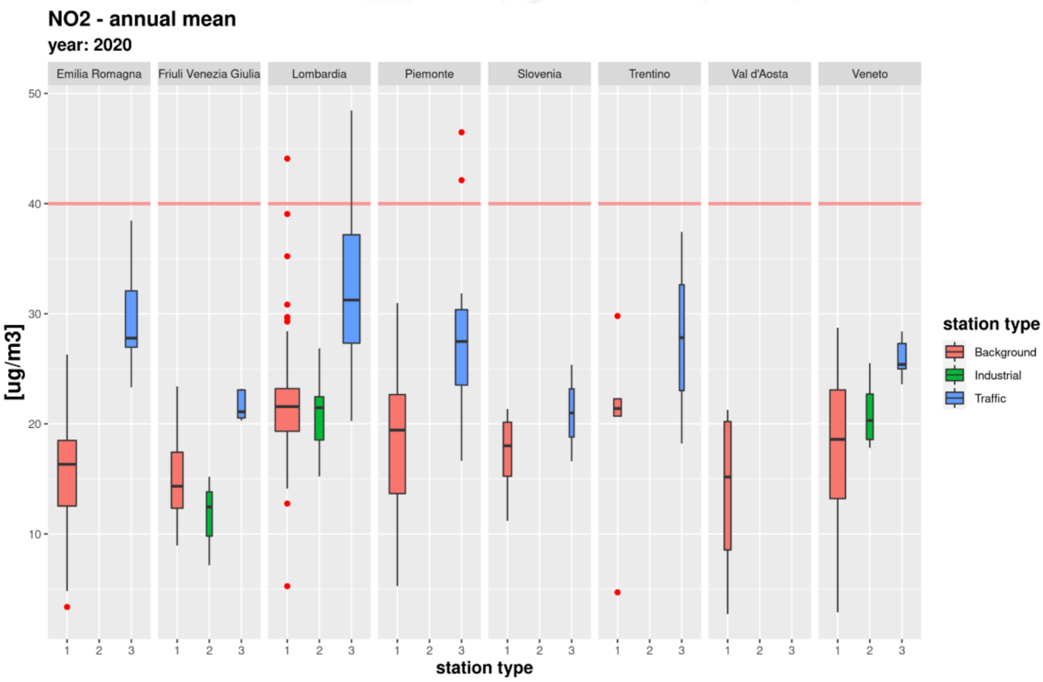


Figure A 8-. NO2 annual mean:boxplots of observed data grouped by station type and region.

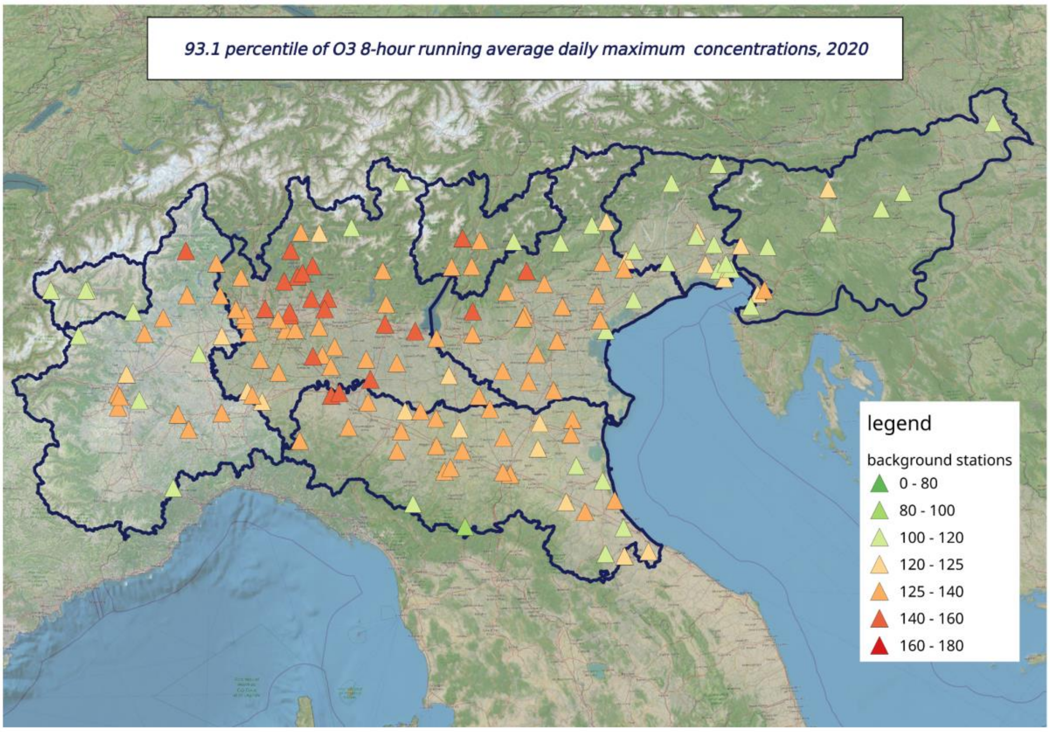


Figure A 9. O3 percentile 90.3: maps of observed data, monitoring stations are grouped by station classification

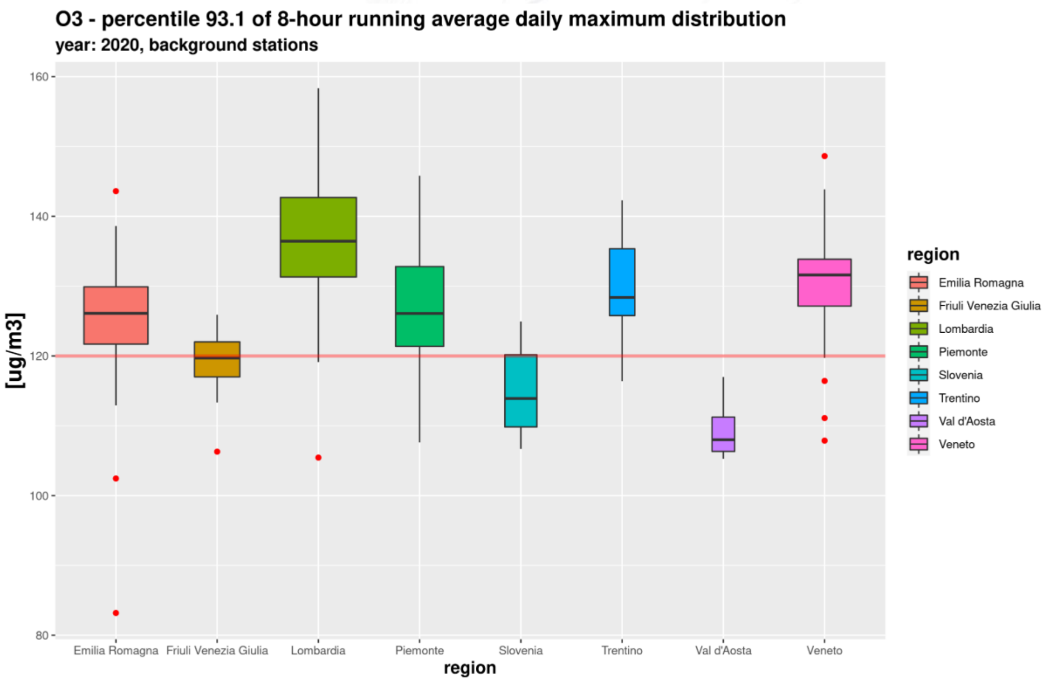


Figure A 10 – O3 percentile 93.1 of 8-hour running average daily maximum distribution

With the contribution of the LIFE Programme of the European Union

LIFE 15 IPE IT 013

prepAIR Po Regions Engaged to Policies of Air

THE PROJECT PREPAIR

The Po Basin represents a critical area for the quality of air, as the limit values of fine powders, nitrogen oxides and ozone set by the European Union are often exceeded. The northern Italian regions re included in this area as well as the metropolitan cities of Milan, Bologna and Turin.

This area is densely populated and highly industrialized. Tons of nitrogen oxides, powders and ammonia are emitted annually into the atmosphere from a wide variety of polluting sources, mainly related to traffic, domestic heating, industry, energy production and agriculture. Ammonia, mainly emitted by agricultural and zootechnical activities, contributes substantially to the formation of secondary powders, which constitute a very significant fraction of total powders in the atmosphere.

Because of the weather conditions and the morphological characteristics of the basin, which prevent the mixing of the atmosphere, the background concentrations of the particulate, in the winter period, are often high.

In order to improve the quality of the air in the Po Valley, since 2005 Regions have signed Program Agreements identifying coordinated and homogeneous actions to limit emissions deriving from the most emissive activities.

The PREPAIR project aims at implementing the measures foreseen by the regional plans and by the 2013 Po Basin Agreement on a wider scale, strengthening the sustainability and durability of the results: in fact, the project involves not only the regions of the Po valley and its main cities, but also Slovenia, for its territorial contiguity along the northern Adriatic basin and for its similar characteristics at an emissive and meteoclimatic level.

The project actions concern the most emissive sectors: agriculture, combustion of biomass for domestic use, transport of goods and people, energy consumption and the development of common tools for monitoring the emissions and for the assessment of air quality over the whole project area.

DURATION

From February 1st, 2017 to January 31, 2024.

With the contribution of the LIFE Programme of the European Union

LIFE 15 IPE IT 013

prepAIR Po Regions Engaged to Policies of Air

TOTAL BUDGET

17 million euros available to invest in 7 years: 10 million of which coming from the European Life Program.

COMPLEMENTARY FUNDS

PREPAIR is an integrated project: over 850 million euros coming from structural funds and from regional and national resources of all partners for complementary actions related to air quality.

PARTNERS

The project involves 17 partners and is coordinated by the Emilia-Romagna Region – General directorate for the territorial and environmental care.

www.lifeprepair.eu – info@lifeprepair.eu