

Ammonia emissions and fine particulate matter: some evidence in Lombardy

Alessandro Fusta Moro, Matteo Salis, Andrea Zucchi, Michela Cameletti,
Natalia Golini, Rosaria Ignaccolo

1. Introduction

Air quality in the Lombardy region (northern Italy) is affected by high concentrations of pollutants. One of the reasons is that Lombardy is localised in the Po Valley where air circulation is very weak because of the mountains surrounding the area. The peculiar weather conditions, the industrial development and the population density make Lombardy one of the worst European region in terms of air quality¹. As a result, epidemiological studies have found that Lombardy is characterized by an elevated mortality rate related to fine particulate matter (PM_{2.5}) exposure. It is well known that a considerable part (from 10% up to 50%) of the PM_{2.5} is formed by the chemical reactions of the ammonia (NH₃) with other precursors. In the Lombardy region, 97% of the total NH₃ gaseous emissions are linked to the agriculture sector (INEMAR - ARPA Lombardia, 2022). Considering that Lombardy is the leading region in Italy for agriculture production, with the highest regional density of swine and bovines, it is clear that the agriculture section has a considerable impact on air quality.

The project *Agriculture Impact On Italian Air Quality*, hereafter *AgrImOnIA*, aims to estimate the local impact of ammonia emissions on particulate matters (PM₁₀ and PM_{2.5}). This information can be crucial for the policy-makers who have to prioritise interventions. *AgrImOnIA* is an ongoing research project, promoted and financed by Fondazione Cariplo within the framework of *Data Science for science and society*. More information on the project are available on <https://agrimonia.net/>.

In this work, we present preliminary results providing continuous spatial maps of PM_{2.5} concentrations (with daily temporal scale) in the Lombardy region using the *AgrImOnIA dataset*², which contains harmonised data on meteorology, emissions and land use. We implement three spatial prediction methods whose performance will be compared by using standard indexes computed with the Leave-One-Out Cross-Validation strategy (LOOCV). In particular, we consider a spatio-temporal Kriging model with external drift, and two random forest algorithms which include spatial and temporal components.

2. Data

The *AgrImOnIA dataset* is an open access data set containing Air Quality (AQ), Weather (WE), Land cover (LA), Emission (EM) and Livestock (LI) data with daily temporal resolution. The data are available for all the air quality monitoring stations after a pre-processing step to change the support of spatial data from area to point, when necessary. We consider the period from 2017 to 2020. The area covered by the *AgrImOnIA dataset* includes the Lombardy region

¹<https://www.eea.europa.eu/publications/air-quality-in-europe-2021>

²The *AgrImOnIA dataset* will soon be available on Zenodo, which is an open repository operated by CERN (<https://zenodo.org/>).

Alessandro Fusta Moro, University of Turin, Italy, alessandro.fustamoro@unito.it, 0000-0003-1129-5038

Matteo Salis, University of Turin, Italy, matteo.salis@edu.unito.it

Andrea Zucchi, University of Turin, Italy, andrea.zucchi925@edu.unito.it

Michela Cameletti, University of Bergamo, Italy, michela.cameletti@unibg.it, 0000-0002-6502-7779

Natalia Golini, University of Turin, Italy, natalia.golini@unito.it, 0000-0003-4457-5781

Rosaria Ignaccolo, University of Turin, Italy, rosaria.ignaccolo@unito.it, 0000-0003-2998-1714

Referee List (DOI 10.36253/fup_referee_list)

FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup_best_practice)

Alessandro Fusta Moro, Matteo Salis, Andrea Zucchi, Michela Cameletti, Natalia Golini, Rosaria Ignaccolo, *Ammonia emissions and fine particulate matter: some evidence in Lombardy*, © Author(s), CC BY 4.0, DOI 10.36253/979-12-215-0106-3.40, in Enrico di Bella, Luigi Fabbris, Corrado Lagazio (edited by), *ASA 2022 Data-Driven Decision Making. Book of short papers*, pp. 227-232, 2023, published by Firenze University Press and Genova University Press, ISBN 979-12-215-0106-3, DOI 10.36253/979-12-215-0106-3

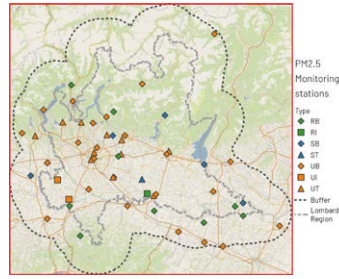


Figure 1: Area of interest with $PM_{2.5}$ monitoring stations, classified by type: rural background (RB); rural industrial (RI); suburban background (SB); suburban traffic (ST); urban background (UB); urban industrial (UI); urban traffic (UT).

and a neighbouring area defined by applying a 0.3-degree buffer to the regional boundaries. The area and the $PM_{2.5}$ monitoring network considered by the *AgrImOnIA* project can be seen in Figure 1.

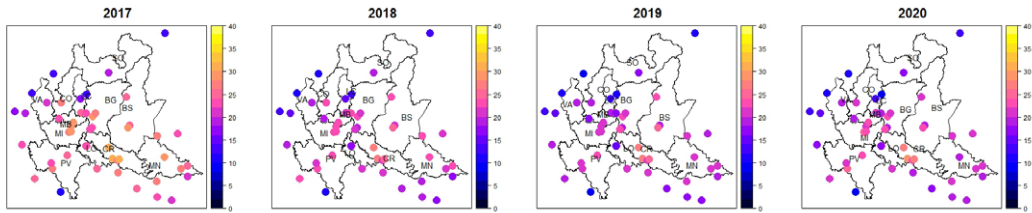


Figure 2: $PM_{2.5}$ annual mean concentrations ($\mu g/m^3$) for each station from 2017 to 2020.

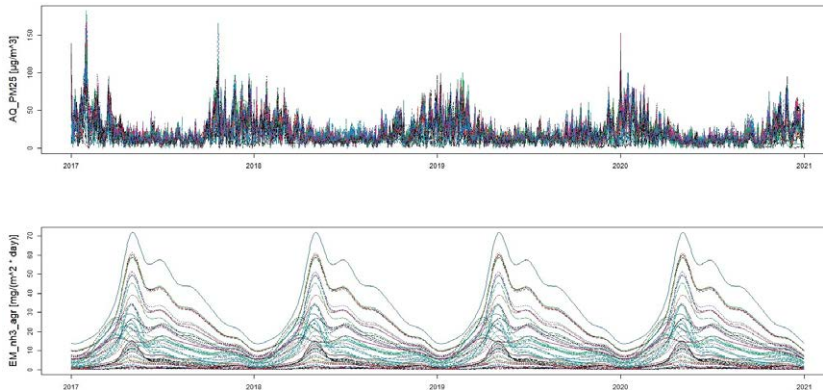


Figure 3: Time series of $PM_{2.5}$ concentrations (top) and NH_3 agriculture emissions (bottom) for all the monitoring sites

From the AQ available variables, we selected $PM_{2.5}$ (AQ_pm25, in $\mu g/m^3$) as the response variable in logarithmic scale. Figure 2 shows the annual mean of $PM_{2.5}$ concentrations in each monitoring station: higher values are located in the lower Po Valley, particularly in the provinces of Milan, Cremona, Lodi and Brescia. The other selected variables are described in Table 1. The overall NH_3 emissions from the agriculture sector (nh3_agr) are calculated by summing up NH_3 emissions from manure management, agriculture soil and agriculture waste burning ($EM.nh3.livestock.mm + EM.nh3.agr.soils + EM.nh3.waste.burn$). To generate

continuous maps as output, the covariates are also obtained on a regular grid $0.1^\circ \times 0.1^\circ$. Figure 3 displays the $PM_{2.5}$ and NH_3 daily time series for all the monitoring stations. As it can be seen, $PM_{2.5}$ concentrations follow a seasonality with peaks during the winter, while ammonia shows higher values in summer, likely because of a higher uses of fertilisers.

Table 1: Description of the selected variables

Variable description	Variable name	Unit	Source
Longitude of the monitoring stations	Longitude	<i>degree</i>	AgrImOnIA dataset
Latitude of the monitoring stations	Latitude	<i>degree</i>	AgrImOnIA dataset
Date	Time	Date	AgrImOnIA dataset
Altitude of the monitoring stations	Altitude	<i>m</i>	AgrImOnIA dataset
$PM_{2.5}$ concentrations	AQ_pm25	$\mu g/m^3$	AgrImOnIA dataset
Temperature of air at 2 m	WE_2m_temperature	$^\circ C$	AgrImOnIA dataset
Mean horizontal wind speed at 10 m	WE_wind_speed_10m_mean	<i>m/s</i>	AgrImOnIA dataset
The accumulated water fallen	WE_tot_precipitation	<i>m</i>	AgrImOnIA dataset
The pressure of the atmosphere	WE_surface_pressure	<i>Pa</i>	AgrImOnIA dataset
Net solar radiation	WE_solar_radiation	J/m^2	AgrImOnIA dataset
High vegetation index	WE_hvi	m^2/m^2	AgrImOnIA dataset
Low vegetation index	WE_lvi	m^2/m^2	AgrImOnIA dataset
Mean horizontal wind speed at 100 m	WE_wind_speed_100m_mean	<i>m/s</i>	AgrImOnIA dataset
Maximum of boundary layer height	WE_blh_layer_max	<i>m</i>	AgrImOnIA dataset
Minimum of boundary layer height	WE_blh_layer_min	<i>m</i>	AgrImOnIA dataset
Maximum of relative humidity	WE_rh_max	%	AgrImOnIA dataset
NH_3 emissions - manure management	EM_nh3_livestock_mm	mg/m^2	AgrImOnIA dataset
NH_3 emissions - agriculture soil	EM_nh3_agr_soils	mg/m^2	AgrImOnIA dataset
NH_3 emissions - agriculture waste burning	EM_nh3_waste_burn	mg/m^2	AgrImOnIA dataset
NH_3 total emissions	EM_nh3_sum	mg/m^2	AgrImOnIA dataset
SO_2 total emissions	EM_so2_sum	mg/m^2	AgrImOnIA dataset
NO_x total emissions	EM_nox_sum	mg/m^2	AgrImOnIA dataset
NH_3 emissions - agriculture sector	nh3_agr	mg/m^2	Own elaboration
Day of the week	day_week	Categorical	Own elaboration
Type of season	season	Categorical	Own elaboration
Type of station	type_station	Categorical	Own elaboration

(see Figure 1)

3. Spatial prediction techniques

In order to perform spatial prediction and to produce continuous spatial maps of $PM_{2.5}$ concentrations, we consider two approaches: 1) spatio-temporal kriging with external drift (STKED), a classical approach in geostatistics framework; 2) a well-known machine learning method - random forest (RF) - extended to the case of data correlated in space and time.

Spatio-temporal kriging with external drift

Spatio-temporal kriging is a supervised parametric model which assumes that the observed $PM_{2.5}$ data are generated by a given stochastic spatio-temporal model. In particular, we suppose that the response variable $\log(AQ_pm25)$ is Normally distributed with a mean changing in space and time and a variance given by the measurement error variance (i.e. the nugget). The mean of the response field is in turn defined as the sum of a large-scale trend (or external drift, which includes the linear effect of the covariates described in Table 1), and a residual spatio-temporal process with separable space-time covariance function. For the implementation of this method we use the `gstat` R-package (Gräler et al., 2016), which requires the estimation of a

spatio-temporal variogram. Once all the models parameters are estimated, spatial prediction of the expected $\log(\text{AQ}_{\text{pm}25})$ value at any location in Lombardy region is straightforward: it is given basically by a local weighted mean of the spatio-temporal residuals plus the large scale component (evaluated using the covariate values at the new sites).

Random forest for spatio-temporal data

Random forest is a data-driven non-parametric machine learning technique given by an ensemble of regression trees fitted using several bootstrapped version of the original data and subsets of the considered covariates. The main limitation of random forest is that it is not able to take directly into account the temporal and spatial correlation, as kriging does. In order to include in the fitting algorithm some information about the data spatial structure, we propose here two different implementations of the method: 1) the standard RF algorithm (RFbase) which includes in the covariate set, besides the variables described in Section 2, also the spatial coordinates (longitude and latitude) of each observation; 2) the spatial RF (RFsp) method proposed by Hengl et al. (2018). This method expands the set of covariates by including the buffer distances from the observation sites (i.e., if we have n monitoring stations we will have n additional columns in the training set each referring to a given station and including the distances from the remaining locations). To take into account the temporal component we consider as covariates the date of the day, the day of the week and the type of season. The two RF algorithms are implemented using the *Ranger* R-package. Prediction in a new spatial location is usually given by the averages of the, say, B predictions computed using the single trees in the forest. Indeed, we consider an ordinary spatio-temporal kriging model for the differences between observed and predicted data in order to include a term taking into account spatio-temporal correlation and predict a term for the small scale component.

4. Preliminary results

Starting considering the STKED technique, we subset covariates through stepwise strategy and we estimate the coefficients shown in Table 2 also referred to interaction terms between season and emissions. We can see that, among the emissions, `nh3_agr` has a larger impact on $\log(\text{AQ}_{\text{pm}25})$ during the winter, while `EM_nox_sum` shows larger effects in the remaining seasons; this is consistent with the results in Thunis et al. (2021). The sample variogram of the residuals of the large-scale component is used to choose the exponential variogram model.

Figure 4 (right) shows the 2020 mean of the daily predicted $\text{PM}_{2.5}$ concentrations in the area of interest. It is worth to note that higher $\text{PM}_{2.5}$ concentrations are predicted where we observe higher NH_3 emissions from the agriculture sector, as shown in Figure 4 (left).

As for the ML approach, the variable importance analysis returns similar results for RFbase and RFsp. Figure 5 shows the weather components as the most important, in accordance with the literature (Cameletti et al., 2011) together with the temporal components. Moreover, it turns out that the euclidean distances between sites (`dist_from_`) is not very important.

The comparison between the prediction capability of the three models is performed through LOOCV and the results are shown in Table 3. STKED shows higher performance compared to the two versions of RF, although it is worth to note that these results are based on a preliminary version of the *AgrImOnIA* dataset.

5. Discussion and further development

Further developments of this work will consider the forthcoming versions of the *AgrImOnIA* dataset and extensions of the considered techniques, always in the framework of the comparison

Table 2: Large-scale component coefficient estimates of STKED

Covariate	Dependent variable: log(AQ_pm25)	
	Estimate	SE
Altitude	-0.152***	0.003
Latitude	-0.032***	0.003
Longitude	-0.022***	0.003
WE_sum_total_precipitation	-0.119***	0.003
WE_mean_wind_speed_10m	-0.142***	0.003
WE_min_boundary_layer_height	-0.082***	0.003
seasonautumm	-0.149***	0.008
seasonspring	-0.411***	0.008
seasonwinter	0.600***	0.010
EM_so2_sum	-0.038***	0.007
EM_nox_sum	0.167***	0.007
nh3_agr	0.111***	0.008
seasonautumm:EM_so2_sum	-0.017*	0.010
seasonspring:EM_so2_sum	0.003	0.010
seasonwinter:EM_so2_sum	0.022**	0.009
seasonautumm:EM_nox_sum	0.027**	0.013
seasonspring:EM_nox_sum	-0.116***	0.016
seasonwinter:EM_nox_sum	-0.077***	0.008
seasonautumm:nh3_agr	-0.130***	0.009
seasonspring:nh3_agr	-0.078***	0.009
seasonwinter:nh3_agr	0.195***	0.016
Constant	2.774***	0.005
Observations	70,119	
R ²	0.325	
Adjusted R ²	0.325	
Residual Std. Error	0.664 (df = 70097)	
F Statistic	1,607.042*** (df = 21; 70097)	

Note: *p<0.1; **p<0.05; ***p<0.01

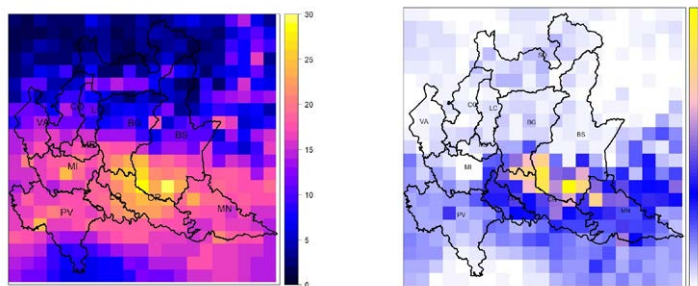


Figure 4: Mean of daily NH₃ emissions from agriculture over the period 2017-2020 (left); mean of daily PM_{2.5} concentrations predicted by STKED for 2020 (right).

Table 3: LOOCV comparison of the models described in Section 3.

	MAE	RMSE	BIAS	COR
STKED	0.316	0.510	0.021	0.782
RFbase	0.352	0.543	-0.022	0.741
RFsp	0.361	0.547	-0.041	0.738

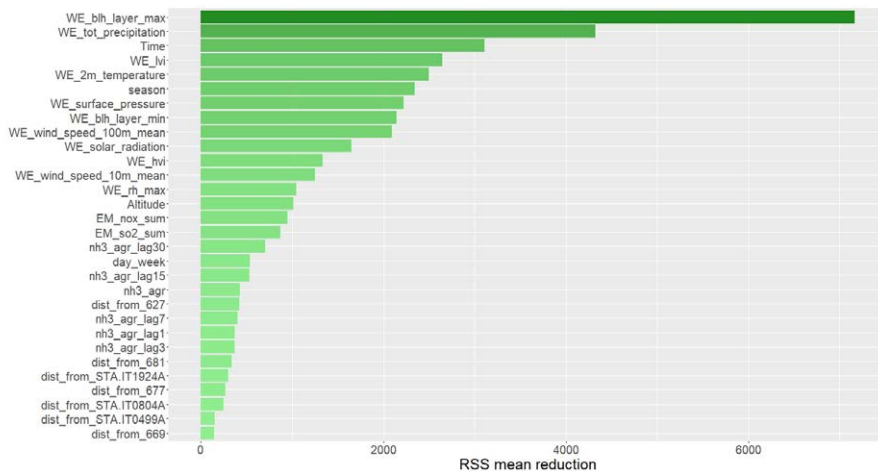


Figure 5: Variable importance of the spatial random forest (RFsp) measured by the RSS mean reduction for each variable.

between classical geostatistics and machine learning approaches.

Recent studies found that NH₃ emissions reductions are the most cost-effective way to reduce PM_{2.5} concentrations (Gu et al. 2021). Scenario analysis based on spatial prediction techniques, in compliance with the Regional Plane for emissions reductions³, will allow to assess the expected impact of NH₃ emissions reduction’s policies before their implementation.

References

- Cameletti, M., Ignaccolo, R., and Bande, S. (2011). Comparing spatio-temporal models for particulate matter in Piemonte. *Environmetrics*, 22(8):985–996.
- Gräler, B., Pebesma, E., and Heuvelink, G. (2016). Spatio-Temporal Interpolation using gstat. *The R Journal*, 8(1):204–218.
- Gu, B., Zhang, L., Dingenen, R. V., Vieno, M., Grinsven, H. J. V., Zhang, X., Zhang, S., Chen, Y., Wang, S., Ren, C., Rao, S., Holland, M., Winiwarter, W., Chen, D., Xu, J., and Sutton, M. A. (2021). Abating ammonia is more cost-effective than nitrogen oxides for mitigating PM_{2.5} air pollution. *Science*, 374(6568):758–762.
- Hengl, T., Nussbaum, M., Wright, M. N., Heuvelink, G. B., and Gräler, B. (2018). Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. *PeerJ*, 6:e5518.
- INEMAR - ARPA Lombardia (2022). *INEMAR, Inventario Emissioni in Atmosfera: emissioni in Regione Lombardia nell’anno 2019 - versione in revisione pubblica*. ARPA Lombardia Settore Monitoraggi Ambientali. <https://www.inemar.eu/xwiki/bin/view/InemarDatiWeb/Inventario+delle+emissioni+in+atmosfera>.
- Thunis, P., Clappier, A., Beekmann, M., Putaud, J. P., Cuvelier, C., Madrazo, J., and de Meij, A. (2021). Non-linear response of PM_{2.5} to changes in NO_x and NH₃ emissions in the Po basin (Italy): consequences for air quality plans. *Atmospheric Chemistry and Physics*, 21(12):9309–9327.

³<https://www.regione.lombardia.it/wps/portal/istituzionale/HP/DettaglioRedazionale/istituzione/direzioni-generalidi/direzione-generale-ambiente-e-clima/piano-regionale-interventi-qualita-aria-pria>