https://doi.org/10.1038/s44407-026-00094-1

From data to decisions: how neural network input structures propagate to air quality policies

Check for updates

Laura Zecchi[1], Michele F. Arrighini[1] ✉, Claudio Marchesi[1], Giorgio Guariso[2] & Marialuisa Volta[1]

In air-quality planning, Artificial Neural Networks (ANNs) have emerged as surrogate models capable of capturing complex nonlinear relationships between emissions and pollutant concentrations with reduced computational time required by chemical and transport models. However, the extent to which surrogate model design influences policy optimization outcomes remains poorly investigated. This study analyzes how alternative ANN architectures affect both predictive performance and the identification of cost-effective air-quality policies at the regional scale. A set of ANN architectures differing in their spatial aggregation and weighting strategies is identified and applied within an Integrated Assessment Model aiming at designing efficient Air Quality policies for the Po Valley basin. The ANNs are trained by processing a set of scenarios simulated by a deterministic chemical-transport model and validated against a real scenario. Such ANNs have been implemented within a multi-objective optimization framework to identify cost-effective emission-reduction policies. The results show differences in ANN performance across different configurations. The largest difference appears between the single basin ANN, which achieves a relative mean absolute error of 6.4%, and regionalized ANNs, whose errors range from 2.4–3.3%. This difference propagates in the optimization results. At low implementation costs (50 M€–150 M€ per year), air-quality outcomes are comparable across architectures, with policies targeting three key macrosectors: domestic heating, transport, and agriculture, though with different resource allocation. At higher cost levels, regionalized architectures outperform the basin-wide model, highlighting how surrogate model design can influence both the fidelity of the simulation and the structure of optimal policy portfolios.

Air pollution is one of the most critical environmental and public health challenges worldwide, contributing to premature mortality, ecosystem degradation, and substantial economic losses[1,2]. Over the past decades, deterministic approaches such as chemical and transport models (CTMs) have represented the backbone of air-quality assessment and regulatory support[3,4]. While physically grounded and scientifically robust, these models are computationally demanding, which limits their flexibility for rapid scenario analysis and integration into optimization-based policy frameworks.

In recent years, artificial intelligence (AI) techniques, such as artificial neural networks (ANNs), have emerged as valuable complementary tools for air-quality modelling[5–7]. ANNs are well-suited to capture complex, nonlinear relationships between pollutant concentrations, meteorological variables, and emission drivers, without the need for explicit representation of atmospheric chemistry. For example, Cabaneros et al[8]. reviewed the use of feedforward and hybrid ANN models and found that these approaches achieve strong predictive performance for ambient air pollutant concentrations in urban environments. Their review also highlights that the specific design of the network architecture and the configuration of the training process are critical factors influencing forecasting accuracy, emphasizing the need for systematic evaluation of model design. Building on these findings, Yadav et al[9]. provide an extensive overview of ANN applications for predicting environmental pollutants and conclude that neural networks often outperform traditional statistical approaches. Their analysis further demonstrates that the selection of input variables can affect the model's ability to make accurate predictions for new scenarios.

In addition to these individual studies, comparative assessments of different machine learning configurations have further demonstrated that

[1]Department of Mechanical and Industrial Engineering, University of Brescia, Brescia, Italy. [2]Department of Electronics, Information, and Bioengineering, Politecnico di Milano, Milan, Italy. ✉e-mail: michele.arrighini@unibs.it

Table 1 | Statistical indices of the ANNs validation phase

	ME [μg/m³]	MAE [μg/m³]	Norm. MEA	St. dev. [μg/m³]	Correlation
Basin	0.006	0.626	0.064	0.826	0.982
Regional	0.006	0.342	0.033	0.462	0.995
Regional-W	0.006	0.258	0.025	0.361	0.997
Regional-WE	0.003	0.247	0.024	0.352	0.997

model structure plays a decisive role in spatial and temporal air pollution estimation. Cerezuela-Escudero et al. [10] and Garbagna et al. [11] compared multiple machine learning models, including different neural network architectures, for spatial estimation of air pollutants and showed that architectural choices and feature engineering strategies impact predictive skill. Their work reinforces the need for structured comparisons rather than single-model implementations, particularly when models are intended to support environmental decision-making.

Despite this growing body of literature, limited attention has been devoted to investigating the impact of different surrogate models in policy evaluation[12]. ANNs are, in fact, well-suited to be implemented within Integrated Assessment Models (IAMs). In the context of air-quality planning, IAMs are a decision-support framework that links emission sources, atmospheric response relationships, economic cost and health modules, and optimization algorithms to identify cost-effective policy strategies in terms of environmental and health objectives.tt Within such a framework, surrogate models are not used solely for predictive reconstruction of pollutant concentrations; rather, they become functional components of a broader optimization process in which emission reductions, economic constraints, and air-quality targets are evaluated at the same time[13,14].

In this study, we compare alternative ANN architectures and input configurations to investigate how surrogate model design affects predictive performance and, when embedded within an IAM, how structural differences propagate into policy optimization outcomes.

The evaluation proceeds in three stages: first, we assess predictive performance using a validation set; second, we compare the base-case scenario with reference data; and third, we implement the different ANN architectures to solve a decision problem, analyzing how structural differences influence the outcomes. The decision problem is formulated as a multi-objective optimization aimed at minimizing the population-weighted $PM_{2.5}$ concentration and the total implementation cost of emission-reduction measures. These are defined by activity sector and pollutant precursor, allowing for the identification of cost-effective emission-reduction strategies

The Po Valley (Northern Italy) represents a suitable test case for this research. Due to its high population density, intense industrial and agricultural activity, and frequent atmospheric stagnation events, the region experiences some of the highest fine particulate matter ($PM_{2.5}$) concentrations in Europe[15,16]. The strong nonlinear interplay between precursor emissions, secondary aerosol formation, and meteorological conditions makes the Po Valley a complex and policy-relevant case for AI-based air-quality modeling[17]. In such a setting, understanding how ANN structural choices affect simulation reliability and optimization robustness is crucial to supporting effective emission-reduction strategies.

By disentangling the role of ANN design in both simulation and optimization contexts, this work contributes to the methodological advancement of AI-based air-quality modeling and aligns with the broader objective of leveraging artificial intelligence to enhance environmental decision support.

Results

Before presenting the optimization outcomes, four ANNs were evaluated for their ability to reproduce annual mean $PM_{2.5}$ concentrations over the Po Valley domain. This preliminary assessment is necessary because, within the IAM, the ANN not only provides concentration estimates but also determines how emission reductions are translated into air-quality improvements during the optimization process. Therefore, differences in ANN structure may affect both predictive accuracy and the resulting set of cost-effective policies.

The 4 ANN configurations are:

- The *Basin* configuration consists of a single ANN trained over the entire Po Valley domain. The inputs include grid-cell coordinates and the aggregated emissions within the cell and its surrounding boundary cells in four spatial slices.
- The *Regional* extends this framework by partitioning the Po Valley into its four administrative regions, each represented by a dedicated neural network, allowing the model to capture region-specific relationships between emissions and pollutant concentrations.
- The *Regional-W* configuration maintains the regional structure but replaces the simple emission aggregation with a distance-based weighting scheme to better represent the spatial influence of nearby sources.
- Finally, *Regional-WE* further enhances this architecture by including average grid-cell elevation as an additional input variable.

Before embedding these four ANN configurations within the optimization framework, a performance assessment was conducted at two different levels. First, the ANNs were evaluated against the CAMx simulations, which generated the training dataset *to* assess their ability to reproduce CTM behavior. Second, an external validation was performed against the EEA 2019 reference scenario to evaluate its capability to reproduce real-world concentration patterns. This step is essential to ensure that the surrogate models provide a reliable representation of the nonlinear relationships between precursor emissions and $PM_{2.5}$ concentrations and can generalize to unseen emission scenarios.

ANNs validation performance

All ANNs are identified and validated using the same dataset, generated with the CAMx model[18] (www.CAMx.com). The dataset is developed from a base-case scenario, referred to as 2019, an average meteorological year without extreme episodes for which a calibrated implementation of CAMx was already available. The complete dataset definition is described below in the "Training dataset" section. The dataset is divided into training and validation sets; the latter is defined as a randomly selected 20% of the available cells across all scenarios.

Table 1 reports the Mean Error (ME), Mean Absolute Error (MAE), Normalized MAE, standard deviation of residuals, and correlation coefficient. All ANN structures exhibit very high predictive capability, with correlation values ranging from 0.982 to 0.997, indicating a strong agreement with reference model outputs. The Basin configuration shows the largest dispersion (MAE = 0.626; normalized MAE = 0.064), suggesting that purely aggregated emission inputs may limit the model's ability to capture spatial heterogeneity. In contrast, regionalized formulations improve performance. The Regional-W and Regional-WE configurations achieve the lowest prediction errors (MAE = 0.258 and 0.247, respectively) and the highest correlation (0.997), indicating that incorporating spatial differentiation and weighted aggregation enhances the surrogate model's representativeness.

The validation results indicate that regionalized architectures reduce prediction errors by 50-60% compared with the Basin ANN. The following sections investigate whether these gains in predictive accuracy translate into different policy recommendations.

ANN's performance on a real-world scenario

To assess the performance of the different ANNs, the results obtained for the base-case scenario are compared with an external scenario, which was not included in the ANN training dataset, developed by the European Environment Agency (EEA)[19] for the same year as the reference base-case scenario. The scenario considered is the year 2019, which corresponds to the meteorological and emission base-case study year. Such out-of-sample

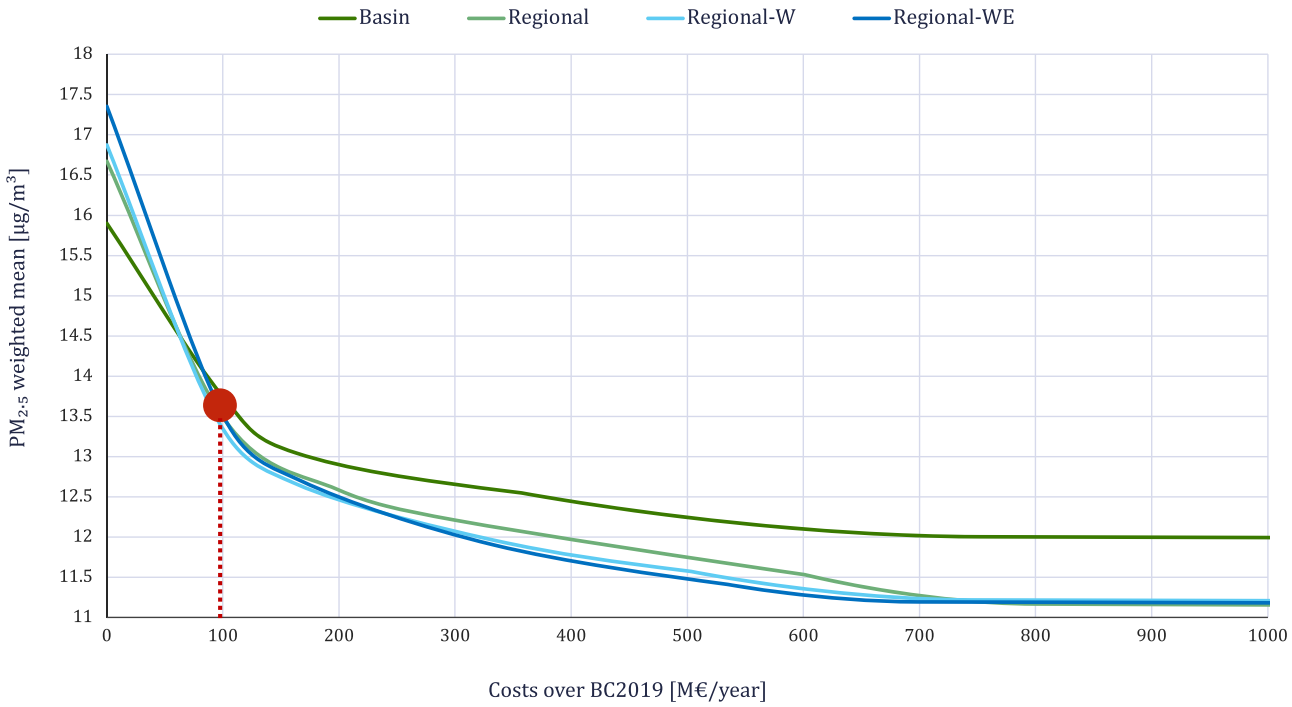


Fig. 1 | Pareto fronts. Solution of the decision problem in the objective spaces: total cost of measure implementation over base-case scenario (horizontal-axis) in M€/year and population-weighted yearly average $PM_{2.5}$ in µg/m^3 (vertical-axis) implementing the four different ANNs.

Table 2 | Statistical indices of EEA 2019 vs ANNs 2019 simulated scenario

	ME [µg/m^3]	MAE [µg/m^3]	Norm. MEA	St. dev. [µg/m^3]	Correlation
Basin	0.823	2.470	0.198	2.767	0.932
Regional	1.270	2.504	0.201	2.621	0.934
Regional-W	1.088	2.327	0.186	2.508	0.935
Regional-WE	1.326	2.312	0.185	2.359	0.945

validation is especially important in Integrated Assessment frameworks, where surrogate models are expected to approximate real conditions and support policy-relevant decision-making.

Table 2 reports the statistical comparison results. As expected, error magnitudes increase compared to the validation phase (Table 1). The ME values are positive for all models, indicating a systematic overestimation of $PM_{2.5}$ concentrations. However, the magnitude of this bias remains below 1.33 µg/m^3 across all configurations, indicating that global concentration levels are well preserved. MAE values increase to approximately 2.3–2.5 µg/m^3, while standard deviations rise to the range 2.3–2.6 µg/m^3.

The limited increase in both bias and dispersion reflects the intrinsic difficulty of extrapolating beyond the training dataset. Differences among ANN configurations become less pronounced under this external scenario. Regionalized models still show marginally better performance, particularly Regional-WE, which exhibits the lowest MAE (2.311 µg/m^3). The relative MAE, which remains below 7% for the validation set, increases under the EEA scenario to 18-20%.

This suggests that when operating outside the calibration manifold, extrapolation uncertainty dominates over structural refinements in network architecture.

In the following sections, the different ANN configurations are embedded within the IAM to investigate how structural variations influence the resulting Pareto fronts and policy-relevant trade-offs.

Impact of ANN structure on multi-objective optimization problem

The multi-objective optimization problem implemented in this study aims to simultaneously minimize the population-weighted yearly average $PM_{2.5}$ concentration and the total implementation costs of emission-reduction measures. The control variables are the application rates of emission-reduction measures, which are detailed for each activity source and the removal efficiency of precursor emissions. The solution to the decision problem is represented by a set of Pareto frontiers, shown in Fig. 1 for the different ANN configurations.

The intercept on the vertical axis corresponds to the population-weighted annual average $PM_{2.5}$ concentration under the base case scenario. As illustrated in the figure, Basin ANN predicts the lowest baseline concentration, whereas the highest baseline value is obtained with the Regional-WE configuration. Despite starting from a lower initial concentration, the Pareto frontier generated by the Basin ANN is dominated by that of the regionalized configurations. In other words, for similar cost levels, the regional ANNs achieve lower $PM_{2.5}$ concentrations (e.g., for an implementation cost of 300 M€/year, the basin ANN achieves 12.6 µg/m^3, compared to 12–12.2 µg/m^3 of the regional architectures). This result indicates that using a single ANN trained over the entire domain limits the model's ability to capture region-specific nonlinear emission-concentration relationships, and that these structural limitations become particularly relevant within the optimization process.

In contrast, the regionalized configurations (Regional, Regional-W, and Regional-WE) achieve similar reductions at comparable cost levels.

Although minor differences appear at lower cost ranges, where some configurations predict slightly higher baseline concentrations, the maximum achievable reduction is consistent across models. For all ANN structures, the "elbow" of the Pareto frontier is observed approximately between 100 M€/year and 250 M€/year, indicating a common cost range in which marginal air-quality improvements begin to decline relative to expenditure.

To further analyze the influence of ANN structure on optimal policy selection, four scenarios corresponding to an implementation cost of 100 M €/year were selected. At this cost level, the four Pareto curves converge to a

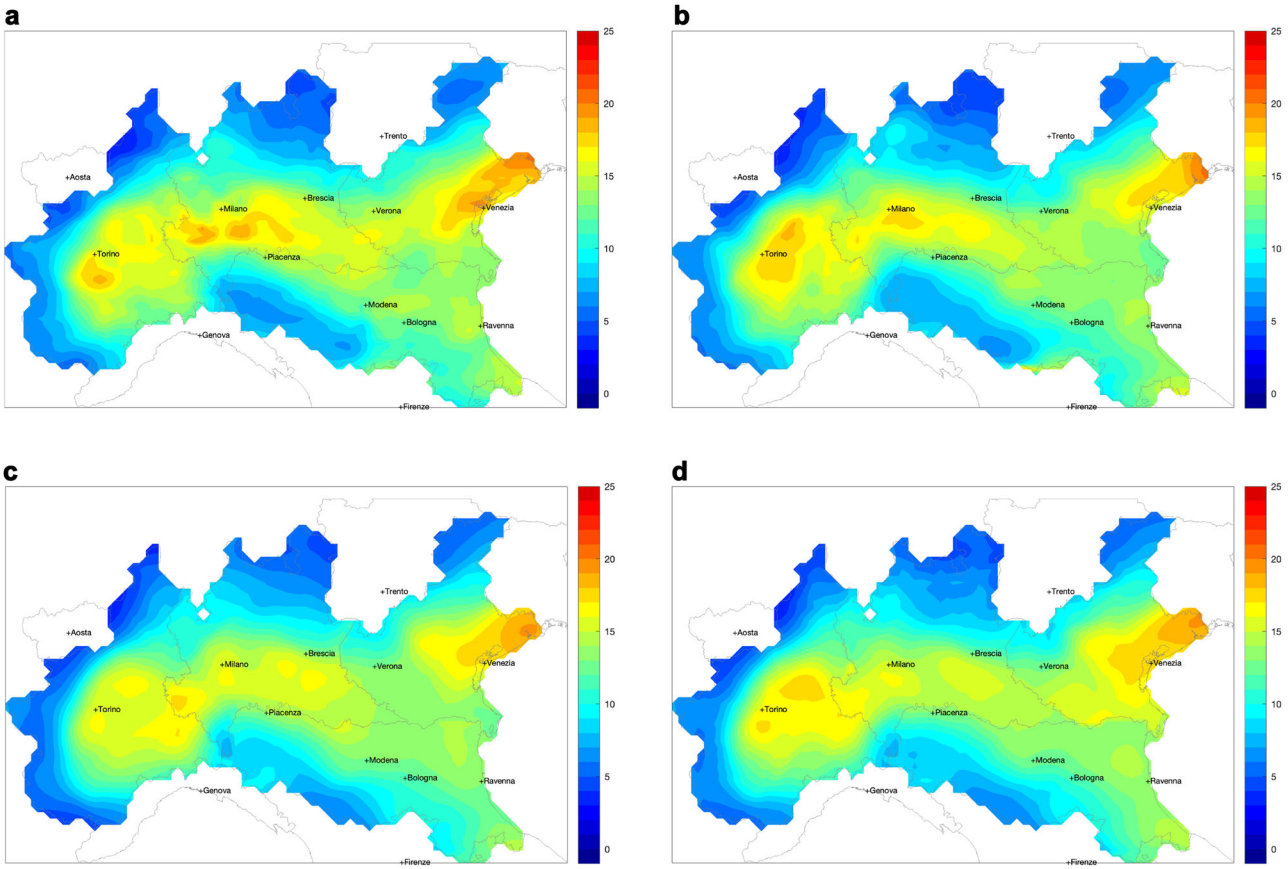


Fig. 2 | Yearly average $PM_{2.5}$ concentration maps. Yearly average PM2.5 concentration in μg/m^3 for a cost of 100 M€/year generated using Basin ANN (**a**), Regional ANN (**b**), Regional-W ANN (**c**), and Regional-WE ANN (**d**).

similar concentration value of ~13.6 μg/m^3, enabling a consistent comparison of the underlying policy portfolios.

The first analysis considers the spatial distribution of $PM_{2.5}$ concentrations in the domain. Although the aggregated indexes are very similar, some differences appear in the concentration distribution, as represented in Fig. 2. Therefore, pairwise t-tests were performed on the gridded yearly average $PM_{2.5}$ concentrations and revealed that all value distributions differed significantly ($p < 0.05$). Maps of yearly average $PM_{2.5}$ concentrations for the base case scenario are reported in Section S4 of Supplementary Materials.

Given the large number of applicable measures evaluated by the optimization framework and the complexity of the resulting policy portfolios, the analysis of the different policies selected focuses on the measures to which the largest share of total costs is allocated. Table 3 reports the measures receiving the highest resource allocation across the four scenarios, identified by implementing the four ANNs (Scenario A with Basin; Scenario B with Regional; Scenario C with Regional-W; and Scenario D with Regional-WE). It is important to note that the sequential quadratic optimization algorithm uses several hundred decision variables. This means there is a large area of indifference among the measures, so the proposed solutions are not structurally linked to the specific ANN used to determine them. Overall, the selected measures are broadly similar across configurations, often with comparable application rates. On the energy side, the transport sector emerges as a priority, through measures such as optimizing urban goods delivery, shifting modal share toward public transport, and electrifying urban buses.

Another sector systematically selected across all configurations is domestic heating, especially interventions targeting residential fireplaces. Emission reductions are achieved either through technological upgrades that improve combustion efficiency or through the installation of thermostatic valves, which enhance energy efficiency when fireplaces are used for domestic heating.

The agricultural sector also plays a central role in optimal portfolios through measures to reduce ammonia emissions. These include improved fertilizer application practices and technological solutions to limit emissions during livestock housing and manure spreading.

The total allocation of costs and the associated emission reductions, aggregated, are reported in Fig. 3. Agriculture, domestic heating, and road traffic account for most of the expenditure in all scenarios; however, their relative shares vary among policies (Fig. 2a).

In policy D, a large proportion of costs is allocated to road traffic measures, with lower investment in agriculture. Conversely, policy B assigns a higher share of resources to agricultural measures and a smaller fraction to transport. Policy C shows increased expenditure on domestic heating, while Policy A presents intermediate allocation patterns.

Analysis of emission reductions (Fig. 2.b) confirms that all policy portfolios prioritize reductions of primary PM and key secondary precursors, notably NH_3 and NO_x. However, differences in cost allocation led to distinct patterns of precursor reduction. For example, the higher investment in traffic measures under the policy D results in greater NO_x reductions and relatively smaller NH_3 reductions compared to configurations that emphasize agriculture (policy B).

ANN sensitivity to precursor reduction

The policies derived from the neural networks show differences in the prioritization of emission-reduction strategies, with some models emphasizing NO_x reductions and others NH_3. This motivates an investigation into whether such differences arise from varying sensitivities of the networks to reductions in precursor emissions. It is important to note that the networks establish relationships between emissions and concentrations and do not

Table 3 | The main measure application and associated costs are identified in the four scenarios. Scenario A corresponds to Basin ANN, Scenario B to Regional, Scenario C to Regional-W, and Scenario D to Regional-WE

Type	Macrosector	Sector	Activity	Measure	Cost			
					A	B	C	D
End-of-pipe	Agriculture	Urea application	No fuel use	Urea substitution	17.7	17.7	17.7	0
End-of-pipe	Domestic heating	Fireplaces	Fuelwood	Improved fireplace	10.8	10.8	10.8	10.8
Energy	Road transport	Light duty vehicles-Urban	Diesel	Optimization of urban goods delivery	9.7	9.7	9.7	9.7
End-of-pipe	Agriculture	Agriculture	Diesel	Stage 3B	8.9	8.9	8.9	8.9
Switch	Road transport	Cars	Natural gas/Diesel	Increase in methane cars instead of diesel cars	18.5	0	0	20.2
End-of-pipe	Agriculture	Other cattle	Other cattle - liquid systems	Combination of covered system and low ammonia emissions	6.3	15.0	15.0	0.00
Switch	Road transport	Cars-Extraurban	LPG/diesel	Increase in GPL cars instead of diesel cars	8.4	0	0	10.2
Energy	Domestic heating	Fireplaces	Fuelwood	Thermostatic valves	0.00	5.6	5.6	5.6
Switch	Road transport	Buses-Extraurban	Diesel	EV vehicles	0.00	0	0	3.7
End-of-pipe	Agriculture	Other cattle	Other cattle - solid systems	Low ammonia application; high efficiency	3.8	3.8	3.8	1.7
End-of-pipe	Agriculture	Dairy cattle	Dairy cows - liquid systems	Low ammonia application; high efficiency	2.3	2.3	3.5	0
End-of-pipe	Agriculture	Pigs	Pigs - liquid systems	Low ammonia application; high efficiency	1.8	1.4	1.4	1.3
End-of-pipe	Agriculture	Pigs	Pigs - liquid systems	Low ammonia application; mean efficiency	1.7	1.3	1.3	1.2
End-of-pipe	Production process	Agglomeration plant-sinter	No fuel use	High efficiency deduster	0.00	2.4	2.4	2.6
End-of-pipe	Industrial combustion	Industrial furnaces	Heavy fuel oil	High efficiency deduster	0.00	2.1	2.1	2.1

explicitly incorporate information on specific mitigation measures. Therefore, rather than emphasizing the order of importance of intervention sectors, this analysis concentrates on their sensitivity to the type of emission reductions.

Figure 4 presents the isolines of the annual average $PM_{2.5}$ concentration simulated by each ANN configuration analyzed in this study. The analysis is conducted by uniformly reducing NO_x and NH_3 emissions across the entire domain within a range from 0% to 90% of the reference level.

The figure highlights the different responses of the ANN-based source-receptor models to variations in precursor emissions. For moderate emission reductions, all configurations exhibit approximately linear behavior, indicating a near-proportional relationship between precursor abatement and changes in PM2.5 concentration. However, as emission reductions increase, nonlinearities emerge, and the response curves progressively deviate from linearity. The extent and shape of this nonlinearity vary across ANN architectures, reflecting structural differences in how emission-concentration sensitivities are represented.

This comparison highlights the different behavior well visible in the Basin ANN, while the other three configurations have similar results.

Health impact assessment

Differences in concentration reductions further translate into heterogeneous health impacts across major urban areas. Although domain-average $PM_{2.5}$ improvements are broadly comparable, city-level attributable deaths vary across configurations, reflecting differences in the spatial distribution of emission reductions. Table 4 reports attributable deaths per 100,000 inhabitants for the main city of the Po Valley, including the affected population (over 30 years of age), and confidence intervals that account for ANN uncertainty and variability in relative risk estimates (see Supplementary Materials).

Overall, the estimated health impacts fall within comparable ranges across configurations. However, regionalized models generally produce lower or more spatially balanced attributable mortality in several cities, suggesting that improved spatial sensitivity enhances the capacity to target densely populated areas more effectively. At the same time, variability across cities confirms that structural modeling assumptions influence the geographic distribution of benefits.

From a policy perspective, this finding is particularly relevant: two surrogate models that deliver similar domain-average $PM_{2.5}$ reductions may imply different local health outcomes and distinct regional equity implications.

Discussion

This study investigated how alternative Artificial Neural Network architectures influence both predictive performance and optimization-driven policy outcomes within an Integrated Assessment Model applied to the Po Valley.

Four ANN configurations (Basin, Regional, Regional-W, and Regional-WE) were compared. All ANN configurations demonstrated strong predictive capability in reproducing deterministic CTM outputs, confirming the suitability of ANN-based surrogates for air-quality assessment. The analysis outlines structural refinements, such as regional disaggregation and distance-weighted aggregation, that improve accuracy and better capture spatial emission–concentration relationships. However, when evaluated with a scenario not included in the training domain, all models exhibited increased discrepancies, highlighting the importance of training data coverage alongside architectural design.

When embedded within the multi-objective optimization framework, structural differences propagated into policy-relevant outcomes. In particular, the optimization results show that at relatively low implementation

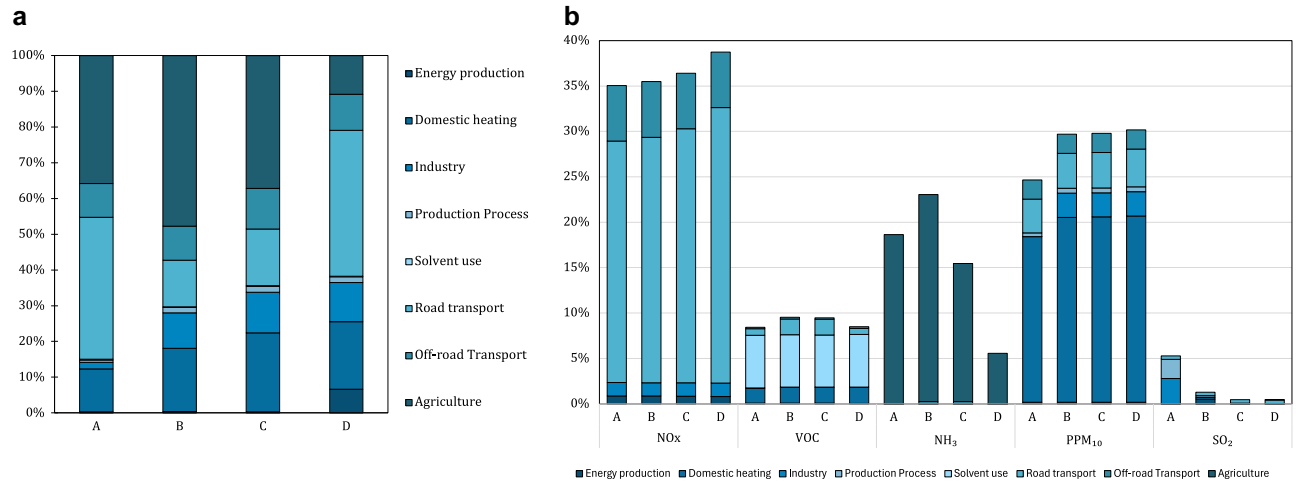


Fig. 3 | Cost and emission data. Cost allocation per macrosector (**a**) and total emission reduction per precursor and macrosector (**b**) in the four optimal scenarios.

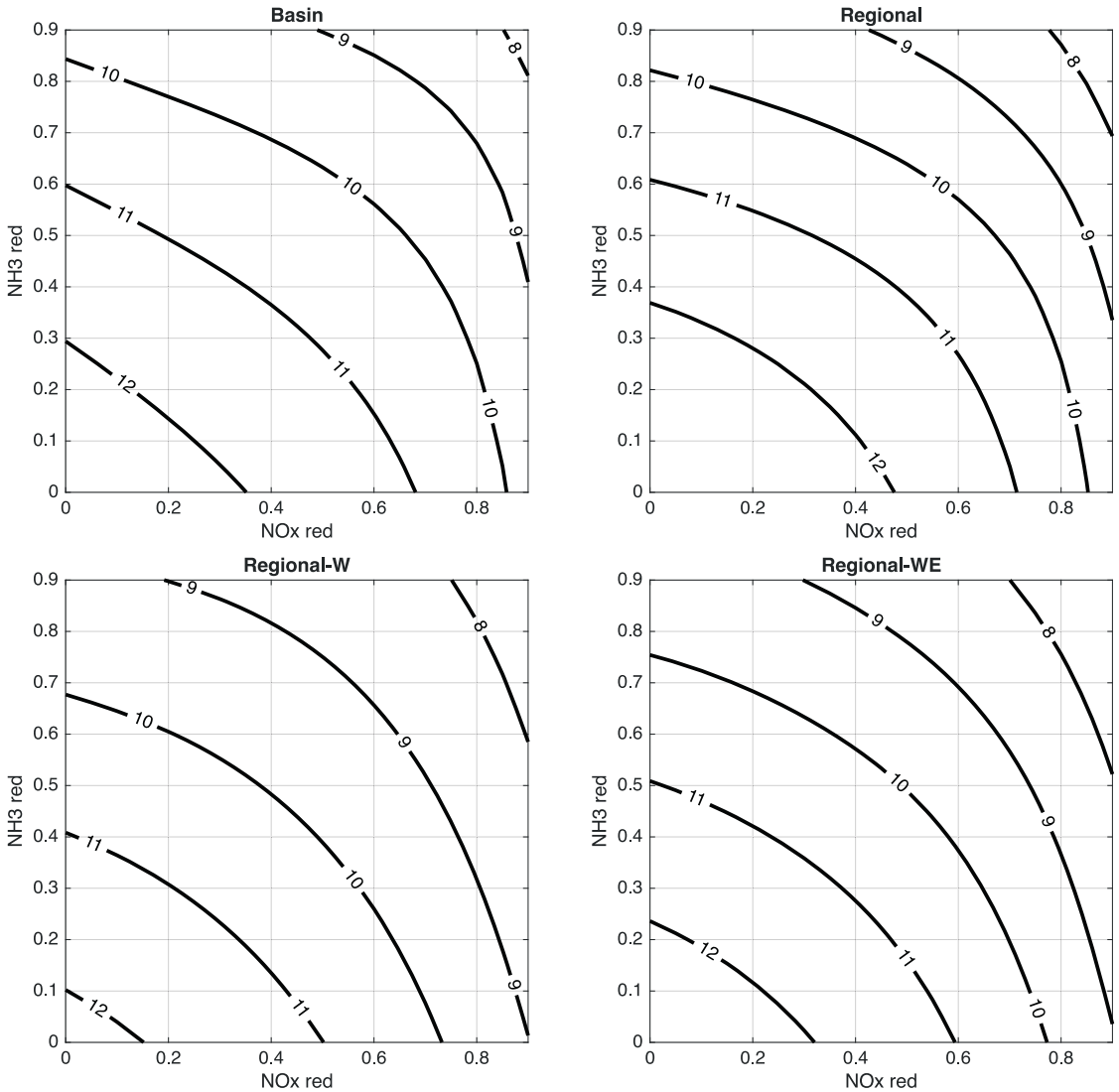


Fig. 4 | $PM_{2.5}$ isolines. $PM_{2.5}$ yearly average concentration [μg/m^3] isolines relative to NO_x (x-axis) and NH_3 (y-axis) reduction for the four ANNs considering base-case emissions.

Table 4 | Health impact expressed as Attributable Deaths over 100k inhabitants in the main cities of the Po Valley

	Population	Basin	Regional	Regional-W	Regional-WE
Milano	1,019,881	103 (69–138)	107 (82–134)	102 (79–125)	99 (77–123)
Torino	640,377	128 (88–172)	130 (100–161)	121 (95–149)	127 (100–153)
Bologna	294,208	89 (51–127)	82 (60–106)	76 (57–96)	74 (57–94)
Venezia	195,168	160 (116–205)	140 (109–174)	147 (117–178)	150 (119–181)
Verona	188,902	95 (59–132)	77 (55–101)	82 (61–102)	78 (59–100)
Padova	155,596	137 (96–179)	120 (92–151)	128 (103–157)	138 (110–168)
Parma	144,062	73 (40–108)	58 (40–79)	71 (55–89)	62 (46–79)
Brescia	141,834	105 (71–145)	86 (62–112)	98 (76–123)	87 (67–110)
Modena	136,113	78 (44–116)	78 (58–99)	83 (64–104)	81 (63–101)
Reggio Emilia	120,264	72 (39–106)	73 (53–93)	78 (60–96)	73 (56–92)
Ravenna	118,554	107 (68–149)	101 (77–126)	95 (74–119)	101 (79–123)
Rimini	108,763	115 (74–160)	107 (82–135)	101 (77–124)	113 (90–139)
Ferrara	102,622	103 (61–146)	104 (79–130)	108 (84–134)	105 (82–130)
Monza	90,166	88 (53–125)	108 (82–137)	105 (83–129)	92 (71–115)
Bergamo	87,605	88 (48–128)	72 (47–98)	85 (64–108)	73 (53–95)
Forli	86,803	84 (45–123)	85 (64–110)	79 (60–100)	90 (69–111)

costs (below ~100 M€/year), all ANN configurations yield very similar population-weighted PM2.5 concentrations (AQI).

However, as the allocated budget increases, the differences among configurations become more evident. The Pareto frontier obtained with the Basin configuration is dominated by those derived from the regionalized ANNs, meaning that, for the same cost levels, regional models achieve lower $PM_{2.5}$ concentrations. In contrast, the regional configurations (Regional, Regional-W, and Regional-WE) exhibit very similar Pareto fronts, indicating a consistent representation of emission–concentration relationships across these architectures.

Despite this convergence in objective space, the corresponding optimal policy portfolios differ in terms of cost allocation across sectors. All configurations consistently prioritize three key macrosectors (transport, agriculture, and domestic heating), but with heterogeneous distributions of investment and emission reductions among them.

To assess whether these differences arise from the ANN representation of the emission–concentration relationship or from the optimization process itself, an additional sensitivity analysis was conducted on precursor emissions. The comparison of ANN responses to NO_x and NH_3 reductions shows that regionalized configurations exhibit very similar behavior, while only minor deviations emerge with the Basin ANN.

These results indicate that differences in the surrogate models do not primarily drive the observed variability in optimal policies, but rather by the optimization process.

It is important to note that the ANN-based source-receptor models operate on aggregated precursor emissions rather than on individual mitigation measures. Therefore, different combinations of measures may generate similar emission reductions and therefore lead to comparable $PM_{2.5}$ concentration outcomes. The existence of multiple optimal policy portfolios observed in this study is therefore primarily related to the structure of the decision problem rather than to differences among ANN architectures.

This study has two main limitations that should be considered when interpreting the results.

First, meteorological variability is not explicitly included among the ANN input variables. This choice is consistent with the policy-oriented purpose of the framework, since future meteorological conditions are not known when emission-reduction scenarios are designed. The models are trained on annual simulations based on 2019 meteorological conditions, and therefore the estimated source-receptor relationships should be interpreted as representative of that specific meteorological year. Future developments should expand the training dataset across multiple meteorological years to assess the robustness of optimized policies under varying atmospheric conditions.

Second, the analysis focuses on $PM_{2.5}$ concentrations. Although $PM_{2.5}$ is a key pollutant in the Po Valley and is associated with substantial health impacts, air-quality management also requires reliable assessment tools for other regulated pollutants. Future developments should therefore apply the same modelling and optimization framework to additional pollutants, such as NO_2, PM_{10}, and ozone, to evaluate whether ANN architecture affects predictive performance and policy optimization outcomes in a similar way.

From a policy perspective, this result provides additional flexibility for decision-makers. Multiple portfolios can achieve similar air-quality improvements at comparable implementation costs, allowing policy selection to account for practical constraints, stakeholder preferences, sector-specific priorities, or implementation feasibility.

Beyond the specific case study of the Po Valley, this work highlights the importance of assessing surrogate model architectures when they are embedded within optimization frameworks. The results show that predictive accuracy alone is not sufficient to qualify model application, as structural choices may propagate into policy-relevant outcomes. Therefore, surrogate model evaluation should consider both simulation skill and decision-support implications.

Methods

Domain

The Po Valley is a geographical area located in northern Italy, roughly between 40° and 50° N latitude and 5° and 15° E longitude. The whole area is partitioned into four administrative regions: Piedmont, Lombardy, Emilia-Romagna, and Veneto.

The area covers approximately 70,000 km². The Po Valley is bound by the Alps to the north and west and the Apennines to the south, forming a semi-enclosed basin that strongly limits atmospheric ventilation. This orographic configuration, combined with frequent thermal inversions and stagnant meteorological conditions, favors the accumulation of air pollutants, making the region one of the European hotspots for particulate matter concentrations. The domain hosts a population of nearly 20 million inhabitants. It is characterized by intensive industrial production, high traffic density, and extensive agricultural activities, particularly livestock farming, which contributes significantly to ammonia emissions. For modeling purposes, the study area is discretized into a regular grid of square cells with a spatial resolution of 6×6 km². This gridded representation allows spatially explicit simulation of emission distributions and pollutant concentrations

Table 5 | Value of coefficient M per scenario and precursor

Scenario	NO_x	NMVOC	NH_3	SO_2	PPM_{10}
1	1	1	1	1	1
2	1.1	1.1	1.1	1.1	1.1
3	0.1	0.1	0.1	0.1	0.1
4	0.1	1.1	1.1	1.1	1.1
5	1.1	0.1	1.1	1.1	1.1
6	1.1	1.1	0.1	1.1	1.1
7	1.1	1.1	1.1	0.1	1.1
8	1.1	1.1	1.1	1.1	0.1
9	0.35	0.35	0.35	0.85	0.35
10	0.35	0.85	0.35	0.85	0.35
11	0.35	0.85	0.85	0.85	0.35
12	0.6	0.6	0.6	0.6	0.6
13	0.85	0.35	0.85	0.85	0.35
14	0.85	0.85	0.35	0.85	0.35
15	0.35	0.35	0.85	0.35	0.85

Table 6 | Summary of the four ANN architectures considered in this study and the reference name

Name	Spatial training	Emission aggregation	Input	*N*	Radius
Basin	Whole domain	Simple sum	Precursors emissions Coordinates	22	Short
Regional	Regional	Simple sum	Precursors emissions Coordinates	22	Short
Regional-W	Regional	Distance-weighted sum	Precursors emissions Coordinates	22	Extended
Regional-WE	Regional	Distance-weighted sum	Precursors emissions Coordinates Average cell elevation	23	Extended

while maintaining computational tractability within the Integrated Assessment framework. The combination of high emission density, complex atmospheric dynamics, and strong population exposure makes the Po Valley a particularly relevant case study for evaluating surrogate modeling approaches in air-quality policy optimization. Data collected for the domain regards the regional emission inventory (INEMAR 2017, carried out in the project: PREPAIR—LIFE15 IPE IT013, https://www.lifeprepair.eu), meteorological data, population, and geographical attributes (National Statistics Institute, www.istat.it), and concentration data (European Environment Agency).

Training data

The source-receptor model is trained using a set of emission scenarios that covers the model's simulation range. The key challenge is data availability: historical data covers a limited range and generally reflects higher past emissions and concentrations. To address this problem, a pre-processing phase is required to generate a limited number of scenarios that effectively cover the space of possible precursor emission variations.

Emission reduction scenarios are defined starting from a base case and simulated with a CTM; the results are subsequently used to train the source-receptor model. The base case scenario is based on emissions described in the previous section and on concentrations computed using 2019 meteorological data.

The emission scenarios were obtained by systematically varying $PM_{2.5}$ precursor emissions within a range spanning from a 90% reduction to a 10% increase.

In this study, the dataset is obtained with the CAMx (Comprehensive Air quality Model with extensions) model, a chemical-transport model that simulates an entire year at hourly resolution. Simulation is necessary to ensure sufficient variability in emissions and concentrations, although it is computationally expensive and limits the number of scenarios. The selected scenario must represent the cause-and-effect relationship between precursor emissions and the various air-quality indices considered. In this case, 15 emission scenarios have been simulated, where the emission value $E_{i,p}$ for the *i-th* scenario and for the precursor p is calculated as:

$$E_{i,p} = M(i,p) \cdot E_{BC,p} \quad (1)$$

where coefficients $M(i,p)$ are presented in Table 5 and $E_{BC,p}$ is the emission of pollutant p in the base-case scenario. Scenarios 3-7 represent the impact of the reductions of a single precursor, and scenarios 8 -14 are obtained through the application of an optimal hyperspace sampling technique based on Sobol's sequences[20]. This technique ensures a uniform distribution of emission variations across the dimensional space defined by all possible emission reductions for the 5 precursors.

The dataset is partitioned into identification and validation sets, with the latter comprising a randomly selected 20% of cells across all situations.

ANNs architecture

The ANNs developed in this study are designed as surrogate source-receptor models that approximate the nonlinear relationship between precursor emissions and annual average $PM_{2.5}$ concentrations at the grid-cell level. Each network estimates pollutant concentration in a receptor cell as a function of spatially processed emissions and geographical descriptors.

In this work, all configurations adopt a feedforward multilayer perceptron (MPL) architecture trained through backpropagation. The network consists of two layers of N nodes (reported in Table 6) with a nonlinear activation function.

Let $\mathbf{x} \in \mathbb{R}^Q$ be the input vector, where Q is the number of input features. The network output can be formalized as follows:

$$\mathbf{y} = f_2\left[W_2 f_1\left(W_1\,\mathbf{x} + \boldsymbol{b}_1\right)\right] + \boldsymbol{b}_2$$

where $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ are the weight matrices of the hidden and output layers, $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$ are the bias vectors. The activation functions are f_1 for the hidden layer (log-sigmoid) and f_2 for the output layer (hyperbolic tangent).

Since pollutant concentrations at a given location are influenced not only by local emissions but also by surrounding sources, a spatial input treatment borrowed from previous studies[12] is implemented. For each cell and each precursor, 4 input values corresponding to the summed contributions from cells belonging to four triangular slices surrounding the receptor cell are added. This approach preserves directional information (from wind) and allows for capturing the short-range transport effect in an aggregate way. The length of the slices (radius of influence) varies with the ANN configuration (shorter for simple sum and longer for weighted sum).

For $PM_{2.5}$, five precursors are considered, resulting in 20 emission inputs (5 precursors×4 slices). PPM_{10} was used as the primary particulate precursor because PM_{10} emission estimates are generally more consolidated in regional emission inventories and are strongly correlated with $PPM_{2.5}$ emissions.

Additionally, all the ANNs considered included geographical UTM coordinates to encode spatial position.

Starting from this configuration, different ANNs are identified that differ by:

- Spatial training domain (entire domain vs regional domains)
- Emission aggregation method (simple sum vs distance-weighted sum)
- Inclusion of elevation as an additional geographical variable.

Table 6 reports the characteristics of the four ANNs identified in the study.

The ANNs are identified using the Levenberg-Marquardt back-propagation algorithm, with mean square error as the performance objective.

IAM framework and multi-objective optimization

The second part of this analysis is developed using an air-quality IAM that integrates emission inventories, socioeconomic activity data, and sector-specific mitigation measures within a unified framework to simulate how changes in precursor emissions propagate through atmospheric processes and ultimately affect annual $PM_{2.5}$ concentrations across the domain.

The decision problem is aimed at identifying the optimal combination of emission-reduction measures that minimize air pollution levels and implementation costs at the same time. The decision variables represent the application rates of a comprehensive set of technological and energy-efficiency measures across multiple emission sectors and activities within the study domain. Each measure affects precursor emissions, which in turn determine annual population-weighted $PM_{2.5}$ concentrations, computed through the ANN-based source-receptor model.

The first objective function corresponds to the minimization of the Air Quality Index (AQI), defined as the yearly average population-weighted $PM_{2.5}$ concentration. In contrast, the second objective minimizes total implementation costs (TC), computed as the aggregate annualized cost of all selected measures based on their unit costs and degrees of application (the formalization of the decision problem is reported in section S1 of Supplementary Materials). The AQI is a function of the emission abatement measures (the decision variables π) as:

$$\frac{\partial AQI}{\partial \pi} = \frac{\partial AQI}{\partial E} \cdot \frac{\partial E}{\partial \pi} \tag{3}$$

where E represents the emissions of the main precursors contributing to $PM_{2.5}$ formation, namely primary particulate matter (PPM_{10}), nitrogen oxides (NO_x), sulphur dioxide (SO_2), ammonia (NH_3), and non-methane volatile organic compounds (NMVOC).

The first term of this equation, describing the relationship between the AQI and the precursor emission, is estimated by the ANNs, while the second term is a function of the application of policy π determined by the decision problem.

The IAM adopted in this study is the MAQ model[21], a bottom-up model that has been extensively applied to air-quality planning in the Po Valley[14,22]. The model integrates multiple interconnected modules, including emission sources, policy cost functions, source-receptor relationships, greenhouse gas emissions, and health impact assessment. Its overall structure is summarized in Section S2 of the Supplementary Material.

In the present study, the IAM incorporates the four different ANNs as computationally efficient source-receptor functions. The model identifies optimal policy in terms of (i) measure selection, (ii) spatial allocation of interventions across the domain, and (iii) associated emission reductions for each precursor. For every candidate efficient policy, which corresponds to a specific set of emission reductions, the embedded ANN computes the resulting annual average $PM_{2.5}$ concentrations at the grid-cell level, allowing both spatially explicit concentration mapping and the calculation of aggregated indicators, such as the population-weighted annual mean concentration.

In addition, the framework quantifies the health impacts associated with each optimized scenario. Health effects are estimated in terms of Attributable Deaths (AD) among the exposed population, based on concentration–response relationships suggested by the World Health Organization[23,24] applied to the modeled $PM_{2.5}$ levels. This integrated structure ensures internal consistency between emission control strategies, air-quality outcomes, economic costs, and public health implications (see Sections S2–S3 of supplementary Materials).

Data availability

Data used for all analyses and analytic code, whenever applicable, are available from the corresponding author upon reasonable request.

Received: 31 March 2026; Accepted: 2 July 2026;
Published online: 15 August 2026

References

1. Boogaard, H., Walker, K. & Cohen, A. J. Air pollution: the emergence of a major global health risk factor. *Int. Health* **11**, 417–421 (2019).
2. Wang, S. et al. The costs, health and economic impact of air pollution control strategies: a systematic review. *Glob. Health Res. Policy* **9**, 30 (2024).
3. Schaap, M. et al. Performance of European chemistry transport models as function of horizontal resolution. *Atmos. Environ.* **112**, 90–105 (2015).
4. Gao, Z. & Zhou, X. A review of the CAMx, CMAQ, WRF-Chem and NAQPMS models: application, evaluation and uncertainty factors. *Environ. Pollut.* **343**, 123183 (2024).
5. Maleki, H. et al. Air pollution prediction by using an artificial neural network model. *Clean. Technol. Environ. Policy* **21**, 1341–1352 (2019).
6. Cordova, C. H. et al. Air quality assessment and pollution forecasting using artificial neural networks in Metropolitan Lima-Peru. *Sci. Rep.* **11**, 24232 (2021).
7. Chae, S. et al. PM10 and PM2.5 real-time prediction models using an interpolated convolutional neural network. *Sci. Rep.* **11**, 11952 (2021).
8. Cabaneros, S. M., Calautit, J. K. & Hughes, B. R. A review of artificial neural network models for ambient air pollution prediction. *Environ. Model. Softw.* **119**, 285–304 (2019).
9. Yadav, V., Yadav, A. K., Singh, V. & Singh, T. Artificial neural network an innovative approach in air pollutant prediction for environmental applications: a review. *Results Eng.* **22**, 102305 (2024).
10. Cerezuela-Escudero, E., Montes-Sanchez, J. M., Dominguez-Morales, J. P., Duran-Lopez, L. & Jimenez-Moreno, G. A systematic comparison of different machine learning models for the spatial estimation of air pollution. *Appl. Intell.* **53**, 29604–29619 (2023).
11. Garbagna, L., Babu Saheer, L. & Maktab Dar Oghaz, M. AI-driven approaches for air pollution modelling: a comprehensive systematic review. *Environ. Pollut.* **373**, 125937 (2025).
12. Carnevale, C., Finzi, G., Guariso, G., Pisoni, E. & Volta, M. Surrogate models to compute optimal air quality planning policies at a regional scale. *Environ. Model. Softw.* **34**, 44–50 (2012).
13. Amann, M. et al. Cost-effective control of air quality and greenhouse gases in Europe: modeling and policy applications. *Environ. Model. Softw.* **26**, 1489–1501 (2011).
14. Zecchi, L., Arrighini, M. F., Guariso, G. & Volta, M. Efficiently supporting decision-making in a multi-pollutant environment. *J. Environ. Manag.* **396**, 128056 (2025).
15. Colombo, L. et al. Assessing the impacts and feasibility of emissions reduction scenarios in the Po Valley. *Front. Environ. Sci*. **11**, 1240816 (2023).
16. Squizzato, S. et al. Factors determining the formation of secondary inorganic aerosol: a case study in the Po Valley (Italy). *Atmos. Chem. Phys.* **13**, 1927–1939 (2013).
17. Ferrari, L. & Guariso, G. Geography-based neural networks for the simulation of air pollution. in *IFAC-PapersOnLine*, Vol. 56, 8296–8301 (Elsevier B.V., 2023).
18. Comprehensive air quality model with extensions (CAMx). Available at https://www.ramboll.com/products/government-and-public/camx (2005).
19. Air Quality Health Risk Assessments. https://www.eea.europa.eu/en/datahub/datahubitem-view/49930245-dc33-4c47-93b8-9512f0622ebc (*EEA*).

20. Sobol, I. M. Uniformly distributed sequences with an additional uniform property. *USSR Comput. Math. Math. Phys.* **16**, 236–242 (1976).
21. Turrini, E., Carnevale, C., Finzi, G. & Volta, M. A non-linear optimization programming model for air quality planning including co-benefits for GHG emissions. *Sci. Total Environ.* **621**, 980–989 (2018).
22. Zecchi, L., Arrighini, M. F., Marchesi, C., Donato, F. & Volta, M. Progress and policies to achieve the zero pollution action plan and EU 2024/2881 PM2.5 targets in Northern Italy. *Sci. Rep.* **15**, 31439 (2025).
23. Soares, J. et al. *Health Risk Assessment of Air Pollution and the Impact of the New WHO Guidelines*. https://www.eionet.europa.eu/etcs/all-etc-reports (2022).
24. WHO. *WHO Global Air Quality Guidelines. Particulate Matter (PM2.5 and PM10), Ozone, Nitrogen Dioxide, Sulfur Dioxide and Carbon Monoxide* (WHO, 2021).

Acknowledgements

This research was partially funded by Fondazione Cariplo under the grant "AgriAir: Data science to reduce agri-food impact on air quality in the Po Valley". Call 2021 "Data Science for science and society", GrantNumber 2020-4295. The authors are grateful to C. Carnevale, A. Tonola, A. Basak (University of Brescia), and L. Ferrari (Politecnico di Milano) for collaborating on the project.

Author contributions

L.Z. and M.A. contributed to data curation, formal analysis, methodology development, software implementation, investigation, validation, visualization, and manuscript review and editing. C.M. contributed to the formal analysis and manuscript review. G.G. contributed to methodology development, supervision, and manuscript review and editing. M.V. contributed to conceptualization, methodology development, project administration, resource provision, supervision, and manuscript review and editing. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s44407-026-00094-1.

Correspondence and requests for materials should be addressed to Michele F. Arrighini.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2026