Environmental Pollution 368 (2025) 125682

Contents lists available at ScienceDirect

Environmental Pollution

journal homepage: www.elsevier.com/locate/envpol

Tracheal, bronchus, and lung cancer mortality and air pollution exposure in Tuscany, Italy: Bayesian Health Impact Assessment and Global Sensitivity Analysis on a sub-regional scale

Michela Baccini [a,b,*], Federico Pirona [c], Laura Grisotto [a,b], Giulia Cereda [a,b], Alessio Lachi [d], Miriam Levi [e], Giulia Carreras [c]

[a] Department of Statistics, Computer Science, Applications "Giuseppe Parenti" (DiSIA) - University of Florence, Viale Giovanni Battista Morgagni 59, Florence, 50134, Italy
[b] Florence Center for Data Science - University of Florence, Viale Giovanni Battista Morgagni 59, Florence, 50134, Italy
[c] Institute for Cancer Research, Prevention and Clinical Network (ISPRO), Via Cosimo il Vecchio 2, Florence, 50139, Italy
[d] Department of Medicine, Saint Camillus International University of Health and Medical Sciences (UniCamillus), Via di Sant'Alessandro 8, Rome, 00131, Italy
[e] Epidemiology Unit of the Department of Prevention - Central Tuscany Local Health Authority, Via di San Salvi 12, Florence, 50135, Italy

ARTICLE INFO

Keywords:
Air pollution
Tracheal, bronchus, and lung cancer mortality
Attributable deaths
Years of life lost
Bayesian analysis
Global Sensitivity Analysis (GSA)

ABSTRACT

Outdoor air pollution is a significant risk factor for tracheal, bronchus, and lung (TBL) cancer. This study employs a Bayesian approach to evaluate TBL cancer mortality due to air pollution in Tuscany, Central Italy, in 2023. Using locally validated data, we assessed the impact of fine particulate matter (PM_{10} and $PM_{2.5}$) and nitrogen dioxide (NO_2) in terms of attributable deaths and years of life lost (YLL). Our three-step methodology included: (1) Bayesian modeling to derive posterior distributions for life expectancy, pollution levels, mortality rates, and exposure-response functions (inputs); (2) Monte Carlo simulations to propagate uncertainty from the inputs to the impact metrics (outputs); and (3) Global Sensitivity Analysis (GSA) to quantify the influence of each input on the outputs. The largest impact was estimated for $PM_{2.5}$, with 432 deaths (50% CrI: 174;705) and 6,500 YLL (50% CrI: 2,624;10,613) in the region due to annual average concentrations exceeding the WHO threshold of 5 $\mu g/m^3$. Central districts, with higher exposure levels, were particularly affected, reporting 14 attributable deaths and 207 attributable YLL per 100,000 inhabitants. The GSA indicated that uncertainty in exposure-response functions and annual average concentrations of air pollutants significantly affected outcomes, highlighting the need to strengthen the regional air quality network and conduct local studies to address effects heterogeneity. Our findings highlight the value of high-quality local health assessments for identifying critical areas, setting intervention priorities, and informing context-specific action plans.

1. Introduction

Among the various risk factors for tracheal, bronchus, and lung (TBL) cancer, outdoor air pollution, defined as the complex mixture of gaseous pollutants and particles originating from natural and anthropogenic sources, is a significant contributor (Malhotra et al., 2016). The evidence of a causal relationship between exposure to air pollutants and lung cancer incidence is clear since 2013, when the International Agency for Research on Cancer (IARC) classified air pollution as carcinogenic to humans. This awareness has strengthened over the years, thanks to numerous large cohort studies conducted in Europe, North America, and Asia, that have reported consistent results (Hamra et al., 2014; Brauer et al., 2022; Brunekreef et al., 2021).

Within the Global Burden of Diseases, Injuries, and Risk Factors Study (GBD) (GBD 2019 Risk Factors Collaborators, 2020; GBD 2021 Diseases and Injuries Collaborators, 2024; Murray, 2022), it was estimated that in Italy in 2019 exposure to ambient particulate matter and ozone pollution was responsible for 6.94 (95% Uncertainty Interval: 4.92;9.42) deaths from TBL cancer and 128.51 (95% Uncertainty Interval: 91.21;173.91) years of life lost per 100,000 inhabitants (Conti et al., 2023). These estimates were approximately confirmed for 2021: out of the 36,685 deaths from TBL, 3,826 were caused by environmental pollution (https://vizhub.healthdata.org/gbd-results/).

In this paper, we propose a complete Bayesian approach to assess, at a sub-regional level, the TBL cancer mortality caused by air pollution

* Corresponding author at: Department of Statistics, Computer Science, Applications "Giuseppe Parenti" (DiSIA) - University of Florence, Viale Giovanni Battista Morgagni 59, Florence, 50134, Italy.
E-mail address: michela.baccini@unifi.it (M. Baccini).

https://doi.org/10.1016/j.envpol.2025.125682
Received 20 August 2024; Received in revised form 13 December 2024; Accepted 10 January 2025
Available online 17 January 2025

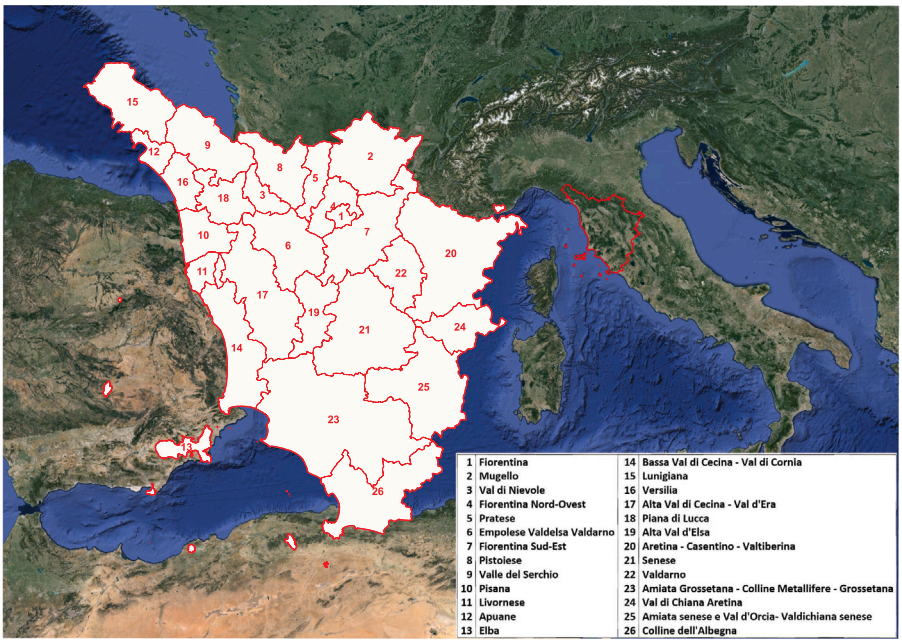

Fig. 1. The Tuscany region with the 26 health districts belonging to the three Local Health Authorities (LHA) (Center LHA: districts 1–8; North-West LHA: districts 9–18; South-East LHA: districts 19–26).

exposure in Tuscany, a region of Central Italy, in the year 2023. Tuscany is the fifth largest Italian region, covering an area of 22,990 km^2 and hosting a population of 3,737,000 inhabitants (data for 2018). It is characterized by hilly terrain, with mountain peaks in the inner and Northern areas and the few plains situated along the coast and to the South. The region shows important disparities in air pollution levels. According to the Regional Agency for Environmental Protection of Tuscany (ARPAT), the most polluted area includes the densely populated provinces of Lucca, Florence, Pistoia, and Prato, in the Northwest of the region. This heavily urbanized area, characterized by the presence of an airport, highway, waste collection facilities, high concentration of industrial settlements, and agriculturally intensive activities, ranks among the most critical in Italy in terms of air pollutant concentrations.

We focused on exposure to fine particles, i.e. particles with diameter less than 2.5 μm ($PM_{2.5}$) and less than 10 μm (PM_{10}), and nitrogen dioxide (NO_2), and quantified their impact in terms of attributable TBL cancer deaths and years of life lost (YLL), adopting a pipeline that develops through three steps. After obtaining via Bayesian modeling the posterior distributions of inputs quantities including air pollution levels, mortality rates, mortality relative risk and life expectancy (step 1), we propagated the uncertainty from them to the attributable impact metrics (outputs) through Monte Carlo (MC) simulations (step 2). Finally, with the aim to enhance comprehension of the mechanisms underlying results construction, we conducted a probabilistic Global Sensitivity Analysis (GSA) quantifying the relative influence of each input on the outputs by computing variance-based indexes (step 3) (Saltelli et al., 2008).

It should be noted that, although seldom used in disease burden assessment, the Bayesian approach provides a natural framework for quantifying and projecting impacts while taking into account uncertainties of various kinds, such as epistemic and sampling uncertainty, heterogeneity, and inherent variability (Spiegelhalter and Best, 2003). Similarly, GSA can help identify the most relevant sources of outcome uncertainty so that future research and actions can be prioritized through targeted knowledge improvement.

2. Data

Tuscany's health services are organized into three areas, each managed by a Local Health Authority (LHA): Center, North-West, and South-East. Each LHA area is further divided into sanitary districts: 8 in the Center, 10 in the North-West, 8 in the South-East. In this paper, we performed analyses at the health district level, thus considering in total 26 sub-regional areas (Fig. 1).

2.1. Mortality and demographic data

Since cause-specific mortality data for 2023 were unavailable at the time of writing this paper, we collected from the Italian National Institute of Statistics (ISTAT) mortality data by sex, age, cause of death and district of residence for the period 2016–2018. Then we used this data to make predictions of the number of deaths from TBL cancer for 2023 (International Classification of Diseases [ICD] 10 codes: C33.0-C34.9, D02.1-D02.3, D14.2-D14.3, D38.1) (Fig. B.1). When the cause of death in the mortality records was ill-defined, a process of redistribution of the events was applied (Monasta et al., 2022). Additionally, we used mortality from all causes among females in the period 2016–2018 for life expectancy estimation. We collected also population data by sex, age and district of residence for the years 2016–2018 and 2023 (www.istat.it), as well as population size on a 4 × 4 km grid in 2015, obtained from the Population Grid of the World, Version 4 (Center for International Earth Science Information Network, 2021).

2.2. Environmental data

In order to account for the latency time between exposure and mortality (Berg et al., 2023), we estimated the health impact in 2023 with reference to regional exposures in 2015. In particular, we collected air pollution data for the year 2015 from two different sources:

- Annual average concentrations of $PM_{2.5}$, PM_{10} and NO_2 for the Tuscany region, defined on a 2 × 2 km grid, estimated through deterministic modeling by the Environmental Monitoring and Modeling Laboratory for the Sustainable Development (LaMMA). The model accounts for meteorological and atmospheric factors, land use, and pollutant emissions derived from the regional inventory;
- Annual average concentrations of $PM_{2.5}$, PM_{10} and NO_2 measured by the monitoring stations belonging to the regional air quality monitoring network of ARPAT. Specifically, we obtained data

from 37 monitors, for a total of 13, 30 and 32 sensors for $PM_{2.5}$, PM_{10} and NO_2, respectively.

3. Methods

3.1. Metrics definition

The aim of this study was separately quantifying the impact of $PM_{2.5}$, PM_{10} and NO_2 on mortality from TBL cancer in terms of attributable deaths (AD) and attributable YLL (AYLL). YLL are a measure of premature mortality, which is computed assigning a weight to each death based on the age at which it occurred. Usually the weight corresponds to the life expectancy at the age of death, derived from local or national life tables, or, sometimes, from an ideal life table, such as the theoretical minimum risk life table adopted by GBD (GBD 2021 Diseases and Injuries Collaborators, 2024). We computed YLL by using the life expectancy of the Tuscany female population as reflecting a regional ideal condition to be applied also to males. For each combination of sex s and district of residence r, YLL from TBL cancer in 2023 are thus defined as:

$$YLL(s,r) = \sum_a D(s,r,a) \times LE(a), \tag{1}$$

where $D(s,r,a)$ is the number of TBL deaths of age a and sex s in district r in 2023, and $LE(a)$ is the age-specific regional female life expectancy.

Under the assumption that the dose–response function describing the long-term relationship between air pollutant concentration and mortality from TBL cancer is linear on a log scale, the number of TBL cancer deaths in district r, attributable to values of the air pollutant exceeding a threshold x_0, can be calculated as follows:

$$AD(r) = \sum_{s,a} D(s,r,a)\Big(1 - \frac{1}{\exp(\beta(x(r) - x_0)I(x(r) > x_0))}\Big), \tag{2}$$

where $x(r)$ is the average level of the air pollutant in 2015 in health district r, β is the logarithm of the relative risk when the air pollutant concentration changes from a generic level x to $x+1$, and $I(x(r) > x_0)$ is a indicator function that is equal to one if $x(r) > x_0$, and zero otherwise. Similarly, the attributable YLL are:

$$AYLL(r) = \sum_s YLL(s,r)\Big(1 - \frac{1}{\exp(\beta(x(r) - x_0)I(x(r) > x_0))}\Big). \tag{3}$$

In our analysis, we fixed the threshold x_0 to the WHO recommended limit for the annual average concentration of the air pollutant, corresponding to 5 $\mu g/m^3$ for $PM_{2.5}$, 15 $\mu g/m^3$ for PM_{10} and 10 $\mu g/m^3$ for NO_2 (World Health Organization, 2021), and, in a second analysis, to a minimum annual concentration across districts estimated for the year 2015. As relative metrics allowing the comparison between areas with different population sizes, we computed the Attributable Fraction (AF), as the ratio between AD and the total number of TBL cancer deaths, the number of AD per 100,000 inhabitants, called Attributable Community Rate (ACR), and the number of AYLL per 100,000 inhabitants.

3.2. Bayesian analysis on the inputs (step 1) and uncertainty propagation (step 2)

The input quantities involved in the computation of the outputs AD and AYLL are the following: annual average level of the air pollutant in 2015 by health district, relative risk of mortality from TBL cancer by level of pollutant, age-specific life expectancy, number of deaths from TBL cancer in 2023 by health district. For each of these inputs, we obtained via Bayesian methods a posterior distribution (step 1). Then, we repeatedly and independently sampled from the posterior distributions of the inputs and, for each combination of the inputs, we used Eqs. (1)–(3), to compute the outputs (step 2). In this way, we obtained for each air pollutant an entire sample from the posterior distribution of the impacts by district, LHA, and for the region as a whole. The study workflow is represented in Fig. A.1. Note that we set AD and AYLL to 0 when a relative risk lower than 0 was sampled, in order to avoid paradoxical negative numbers of attributable events. Therefore, the posterior distributions of the outputs exhibit a probability mass in the null (Baccini et al., 2015).

Below, we briefly describe the Bayesian procedures used to obtain the posterior distributions of the inputs.

3.2.1. Life expectancy

We estimated in a flexible way the age-specific life expectancy $LE(a)$, $a = 0, 1, \ldots, 100$, by specifying a Bayesian generalized additive model on the age-specific mortality rates for females in the period 2016–2018. The sample from the posterior distribution of the age-specific rates, obtained via MCMC through the `rjags` package of R, was then used to calculate the age-specific life expectancies $LE(a)$ according to the usual life table procedure (Cutler and Ederer, 1958). Note that, in this way, we obtained an entire sample from the posterior distribution of $LE(a)$. See Section Life expectancy, Appendix A for details.

3.2.2. Number of deaths

As data on mortality in 2023 were still unavailable at the time we performed the analysis, we obtained a sample of $D(s,r,a)$ adopting, separately by LHA and sex, the following procedure:

- For each health district r in the LHA, we modeled the age-specific mortality rates for the period 2016–2018 with a Bayesian penalized regression spline;
- We combined these first-stage district-specific rates in a Bayesian multivariate random effects meta-analysis and applied a shrinkage procedure to derive second-stage posterior distributions of the district-specific rates (Röver, 2020). The second-stage estimates take advantage of the information from all districts in the LHA, preserving at the same time between-district heterogeneity;
- We combined the rates obtained at the previous point with the observed population in 2023 to derive a posterior distribution of the expected number of TBL cancer deaths at the district level for the year 2023;
- Following a Poisson distribution, we sampled around the expected number of deaths, obtaining a sample from the posterior predictive distribution of the terms $D(s,r,a)$.

The Bayesian analyses were performed via MCMC by using `rjags` and the `mvma.bayesian` command of the `altmeta` package of R software. A detailed description of the method is reported in Section Number of deaths, Appendix A.

3.2.3. Effect of air pollution on tracheal, bronchus, and lung cancer mortality

The effect estimates for the three air pollutants were obtained by combining in separate Bayesian random-effects meta-analyses results from the literature. Specifically, for $PM_{2.5}$ and NO_2 we considered the relative risks of TBL cancer mortality arising from the seven cohorts included in the European multicenter study conducted by Stafoggia et al. (2022). These cohorts were enrolled between 2000 and 2011 and followed up until 2011–17. On the contrary, for PM_{10}, we collected and combined the studies included in three recent meta-analyses (Kim et al., 2018; Chen and Hoek, 2020; Yu et al., 2021). In both cases, the relative risks have been estimated under the assumption of a constant effect of the air pollutant on a logarithmic scale. Through Bayesian meta-analysis, we obtained for each of the three air pollutants a sample of the posterior predictive distribution of the overall log-relative risk β to be used in Eqs. (2) and (3). This posterior distribution incorporates both sampling variability and heterogeneity between cohorts or studies (Riley et al., 2011). The amount of heterogeneity was quantified using the I^2 index, that measures the percentage of total variance due to heterogeneity. Details on data collection and meta-analysis techniques are reported in Section Air pollutant effect, Appendix A.

3.2.4. Exposure levels

To assess exposure in the region, we adopted an approach that integrated information from monitors and models on $PM_{2.5}$, PM_{10}, and NO_2 concentrations. In particular, we jointly estimated the average annual concentration surfaces of the three air pollutants along with their associated uncertainties on a 4 × 4 km grid covering the whole region (note that, for computational reasons, we produced predictions at a lower resolution than the original one of 2 × 2 km). This was achieved through a Bayesian multivariate geostatistical model, integrating sensor measurements with the results of the deterministic model developed by LaMMA. Specifically, we defined a multivariate Normal model on the vector of the annual NO_2, PM_{10} and $PM_{2.5}$ means measured by the sensors, considering as covariates the values of the three pollutants estimated by the deterministic model in the cells where the monitors were located. The model accounts for correlation between pollutants and spatial correlation, defined as a function of the distance between the monitors. It also includes an unstructured heterogeneity component. Informative priors have been defined on the parameters of the distance function, while uninformative priors have been elicited on the other parameters of the model. Inference was conducted via MCMC with the software WinBUGS. After obtaining a sample from the joint posterior distribution of the model parameters, we used the air pollutants levels from the LaMMA deterministic model to obtain the posterior distribution of the predictions at the centroid of 4 × 4 km cells. Finally, for each air pollutant, we averaged the cell predictions within each health district, weighting each cell according to the size of the resident population. In the end, after marginalizing over the posterior distribution of the model parameters, this procedure produced a sample from the joint posterior predictive distribution of the average concentrations of the three pollutants in each health district. Details on the Bayesian multivariate geostatistical model are reported in Section Exposure model, Appendix A.

3.3. Probabilistic global sensitivity analysis (step 3)

The contribution of each input to the outputs variance was quantified by calculating the first order variance indexes and the total variance indexes.

Let $(X_1, X_2, \dots, X_{K_X})$ be K_X mutually independent inputs and $f(\cdot)$ a function which, given the inputs, returns one output Y. The first-order variance index S_j for the input X_j is defined as:

$$S_j = \frac{V_{X_j}(E_{\boldsymbol{X}_{\sim j}}(Y|X_j))}{V(Y)},$$

where $\boldsymbol{X}_{\sim j} = (X_1, X_2, \dots, X_{j-1}, X_{j+1}, \dots, X_{K_X})$. The index S_j represents the main effect contribution of X_j to the variance of the output, and it measures the expected reduction of the variance that would be achieved if the factor X_j could be fixed (Sobol, 1994). S_j does not account for the variance explained by the interaction of X_j with the other inputs. On the contrary, the total variance index S_j^{tot} measures the overall effect of X_j on Y, including all the interactions of X_j with the other inputs. This index corresponds to the expected variance of Y that would be left on average when all the parameters but X_j, $X_{\sim j}$, were fixed:

$$S_j^{tot} = \frac{E_{\boldsymbol{X}_{\sim j}}(V_{X_j}(Y|\boldsymbol{X}_{\sim j}))}{V(Y)}.$$

A total variance index close to zero indicates that the parameter X_j does not influence Y. Conversely, a large total variance index indicates that the parameter does impact on the output. If S_j and S_j^{tot} are similar there is no interaction effect, while if S_j^{tot} is much larger that S_j, there is strong interaction between X_j and the other inputs. Note that, while $\sum_j S_j \leq 1$, the sum of the total variance indexes over all inputs can be larger than 1.

In our analysis, the outputs AD and AYLL are related to the inputs via Eqs. (2) and (3). We computed the variance indices for the whole region, by LHA and district, following the design matrix-based method proposed in Saltelli et al. (2008). According to this design, starting from the simulation of 20,000 samples from each input, we calculated the inputs relying on the results from 120,000 procedures of uncertainty propagation.

4. Results

Applying the shrinkage mortality rates for the years 2016–2018 (Figs. B.2–B.4, Appendix B) to the population size in 2023, we obtained the posterior predictive distributions of the number of deaths from TBL cancer in 2023, by district. Their means range from a few dozen (Elba Island, 31,351 inhabitants) to more than 300 deaths (Florence district, 362,742 inhabitants) (Fig. B.5). Overall, TBL cancer is expected to be responsible in the region of 2'660 (90% CrI: 2,500;2,824) deaths in 2023, that, combined with the life expectancy estimated for women for the period 2016–2018 in the region (left panel of Fig. B.6), generate 40'064 YLL (90% CrI: 37,306;42,950). The largest amount of YLL from TBL cancer is expected in the Center, with 17,691 (90% CrI: 16,194;19,263) years of life lost over 1,604,409 inhabitants, the lowest in the South-East LHA, with 8'268 (90% CrI: 7,242;9,357) years of life lost over 811,242 inhabitants (right panel of Fig. B.6, and Table B.1). The TBL cancer burden is higher among men than among women (Table B.1).

Fig. 2 shows, at a 2 × 2 km resolution, the annual average levels of the three air pollutants estimated for 2015 by the deterministic model developed by LaMMA, together with the locations of the sensors belonging to the regional air quality monitoring network. Note that the sensors are concentrated in the Northern-Central part of the region that, according to the LaMMa model, is the most polluted, leaving the Southern part mostly uncovered. Combining these two sources of information, we obtained for each air pollutant a predictive posterior distribution for the concentration surface over a 4 × 4 km grid (Fig. B.7), then a posterior predictive distribution of the density-weighted average concentration for each health district (Fig. B.8 shows the density levels in 2015 used in the weighting). Fig. 3 summarizes these distributions in terms of mean and coefficient of variation. The largest exposures are in districts belonging to the Center and North-West LHAs: the surrounding of the provincial capitals of Florence, Prato, Lucca, Pistoia, Livorno and Pisa, the Val di Nievole area, and the Apuan Alps area. As expected, the uncertainty is larger in those districts where in 2015 there were few or no monitors: the South-East areas including Siena, Val d'Orcia and Colline di Albegna (Albegna Hills), and the North-West areas of Lunigiana and Serchio Valley.

The posterior predictive distributions of the relative risks expressing the effect of the three air pollutants on TBL cancer mortality exhibit a quite large variability (Fig. 4), that directly derives from the fact that we are predicting effects in a new area in the presence of strong heterogeneity among studies (posterior median of the I^2 equal to 94.0%, 76.8% and 95.2% for $PM_{2.5}$, PM_{10} and NO_2, respectively). The posterior distributions of the relative risks arising from the Bayesian meta-analyses are summarized in Table B.2.

The upper panel of Fig. 5 shows, for each air pollutant and by health district, the expected number of deaths from TBL cancer in 2023 that are attributable to annual average concentrations that exceeded in 2015 the WHO recommended thresholds. To obtain conservative summaries that are not affected by setting to 0 the negative AD values arising from Eqs. (2) and (3), we summarized the posterior distributions using the median instead of the mean. We also computed 50% and 80% CrI, and the posterior probability of a positive number of AD. As detailed in Table 1, where for brevity we have reported only 50% CrI, the highest impact in both absolute and relative terms is expected in the LHA of the Center and for $PM_{2.5}$. In the region as a whole, we estimated that annual average values of this air pollutant exceeding 5 $\mu g/m^3$ are responsible of 432 TBL cancer deaths (16.3% of the total number of TBL cancer deaths). This corresponds to 12 deaths every 100,000 inhabitants and 176 years of life lost every 100,000 inhabitants (1

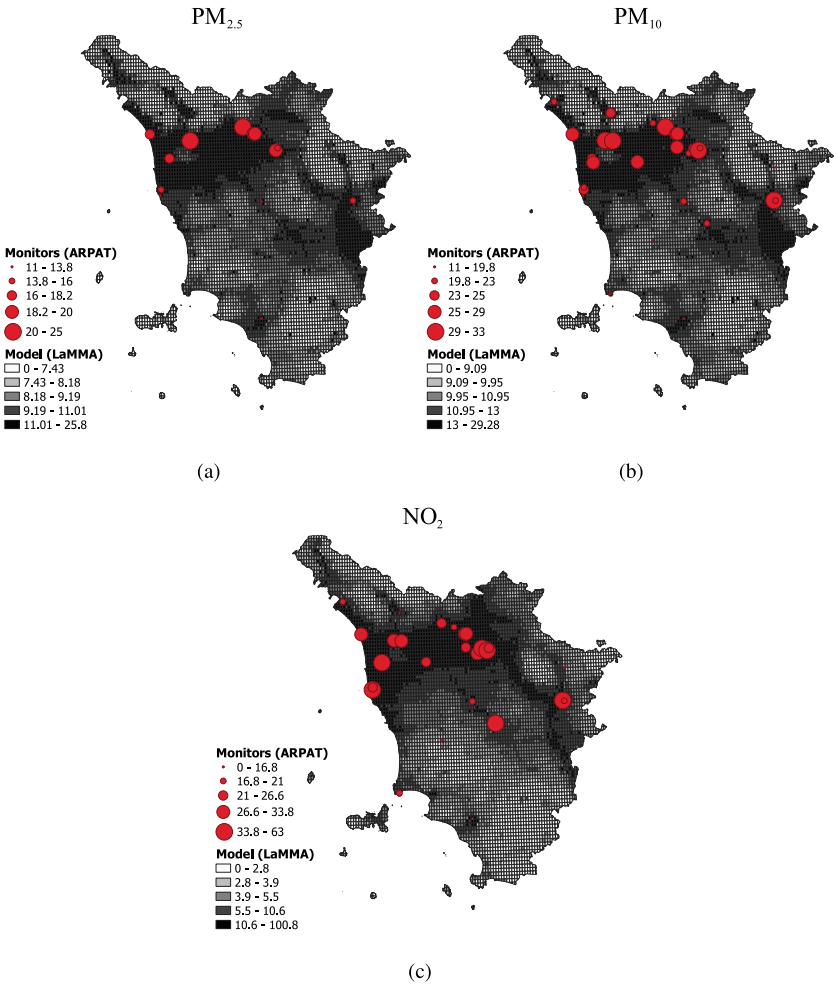

Fig. 2. Average concentrations of $PM_{2.5}$ (a), PM_{10} (b) and NO_2 (c) in the year 2015, measured by the stations belonging to the regional air quality monitoring network (red dots), and estimated at a 2 × 2 km resolution by the deterministic model. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

year lost every 568 inhabitants). The probability that $PM_{2.5}$ above the WHO threshold causes at least one death is higher than 80% in most health districts (lower panel of Fig. 5). Limiting to the whole region, Table 1 reports also the impact when the counterfactual threshold is set, for each air pollutant, to the minimum exposure predicted over the 26 health districts. With this threshold, the impact for $PM_{2.5}$ is lower (the regional minimum is, in fact, higher than 5 $\mu g/m^3$), but still important. On the contrary, the impact is higher for PM_{10} and NO_2, being the regional minimum lower than the WHO limits. The impact of air pollution among males is twice as high as among women, as showed by the results reported in Table B.3.

In order to enhance comparisons among health districts, maps in Fig. 6 illustrate the posterior medians of the district-specific Attributable Community Rates (ACR) with the corresponding posterior coefficient of quartile variation. The highest ACRs are predicted for the areas of Florence, Prato, Pistoia, Lucca and Nievole Valley, the lowest for the districts of Upper Elsa Valley, Siena, Albegna Hills, and Elba Island. The uncertainty around the predictions was larger in the districts where the estimated impact was relatively low.

To investigate the relative importance of different inputs in impact computation, we calculated the Sobol's variance indices. The first-order and total variance indexes for the outcomes AD and AYLL, referred to the region as a whole, are reported in Table 2, limited to $PM_{2.5}$. Results are very similar regardless of the counterfactual threshold used. The most influential input is the air pollutant effect estimate, with S_{tot} greater than 93%, followed by air pollution exposure, with S_{tot} around 14% when fixing the threshold to the WHO limit and 23% when the threshold is sampled from the marginal distribution of the regional minimum. Life expectancy and the number of deaths explain a negligible portion of the outputs variability. Comparing first-order and total variance indexes reveals an interaction between air pollution exposure and both the threshold and relative risk. When we look at the total variance indexes calculated by health district (Fig. 7), the importance of the number of deaths is confirmed. However, a strong impact of air pollution exposure also emerges for districts in the South-East LHA and, more generally, for districts not covered by the regional air quality monitoring network.

5. Discussion and conclusions

We provided an assessment of the impact of air pollution on mortality from TBL cancer in Tuscany at a sub-regional level, mainly based on local validated data. Compared to studies that focus on results at the country level using estimates from the GBD initiative (Conti et al., 2023), we obtained results at the district level, which are particularly useful for detecting critical situations in the region, establishing intervention priorities, and suggesting context-specific action plans.

The Bayesian approach allowed us to account for all sources of variability by obtaining samples from the predictive posterior distributions of the metrics of interest, which completely describe the studied

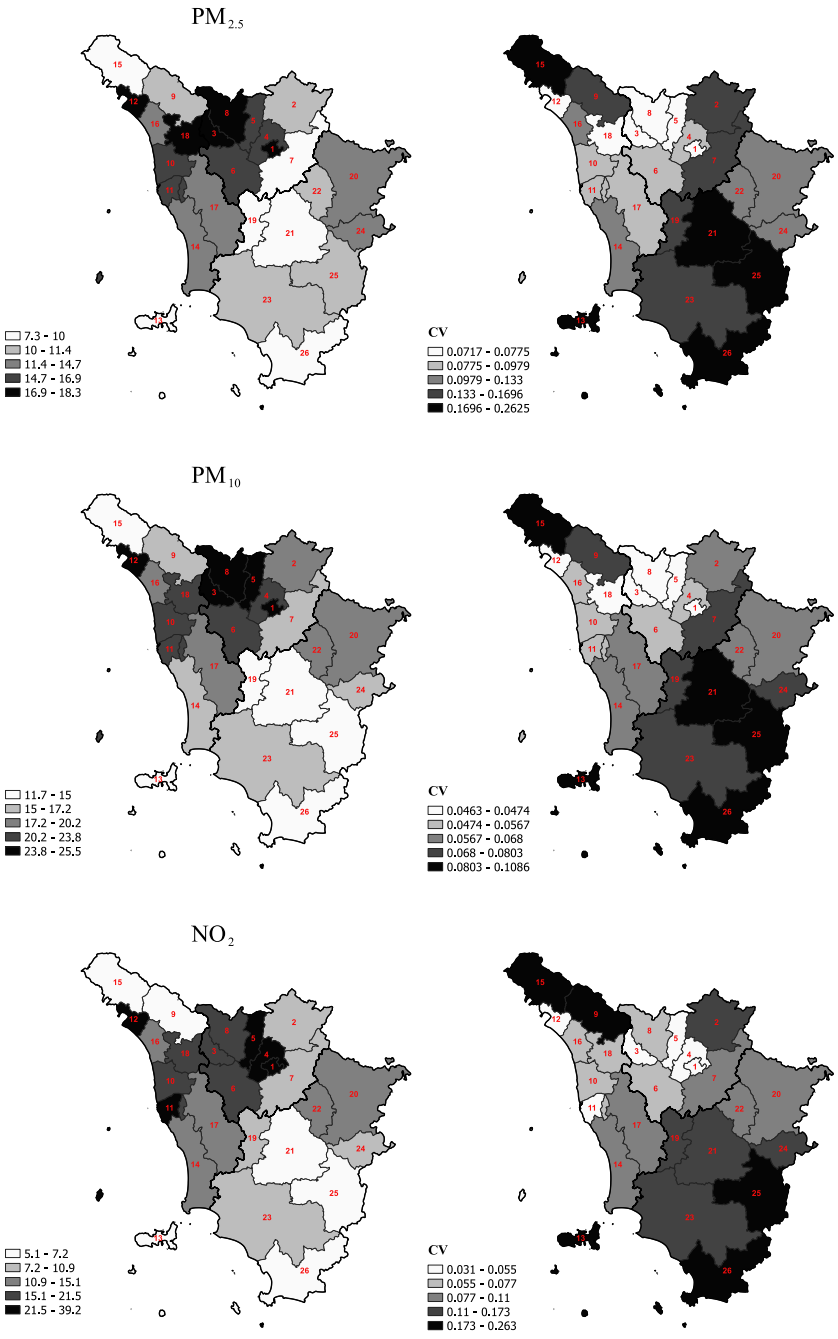

Fig. 3. Means (left maps) and coefficients of variation (right maps) of the posterior predictive distributions of $PM_{2.5}$, PM_{10} and NO_2 average concentrations for the year 2015 in the 26 districts of Tuscany, obtained weighting the exposure predicted over the 4 × 4 km cells by the population density. The bold outlines indicate the Local Health Authorities: Center (1–8), North-West (9–18), South-East (19–26).

phenomenon along with its uncertainty. Additionally, by sampling from the joint posterior distributions of exposure and number of deaths from TBL cancer across areas in the region, we generated reliable credibility intervals that account for correlation among districts.

Despite having the joint distribution of the annual average concentrations of $PM_{2.5}$, PM_{10}, and NO_2 from the Bayesian geostatistical model, we chose to quantify the impacts of $PM_{2.5}$, PM_{10}, and NO_2 separately. Determining an overall comprehensive estimate of the number of deaths attributable to particulate matter and NO_2 levels exceeding recommended thresholds would require strong assumptions about their potential synergistic effects on TBL cancer mortality, or reliable estimates of effect from the literature that account for the interaction between the two exposures, estimates that are not available. For this reason, we have kept the assessments separate. However, it can be argued that, unless unlikely negative interaction effects are present, the overall impact cannot be lower than the highest observed, which corresponds to the impact of $PM_{2.5}$. For this air pollutant, we estimated for 2023 an important impact both in terms of deaths and years of life lost attributable to average annual level concentrations exceeding the WHO limit of 5 $\mu g/m^3$: 432 AD (50% CrI: 174;705) and 6500 AYLL (50%CrI: 2,624;10,613), that corresponds to about 15 years of life lost for each deceased. Reflecting a level of exposure to particles well above that of the rest of the region, districts belonging to the LHA of the Center were the most affected by air pollution, with 14 AD (50% CrI:

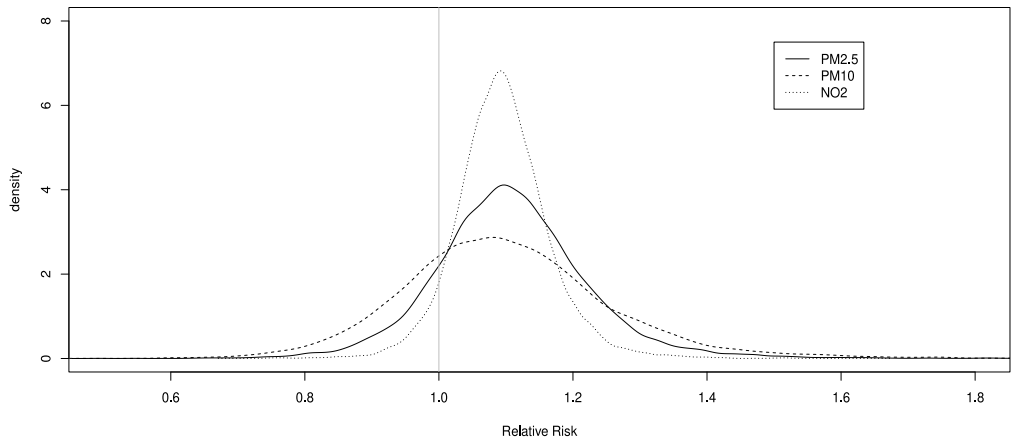

Fig. 4. Predictive posterior distributions of the relative risks of mortality from tracheal, bronchus, and lung cancer associate with an increase of 5 μg/m^3 for $PM_{2.5}$, and with an increase of 10 μg/m^3 for PM_{10} and NO_2.

Table 1
Posterior median and 50% credibility interval (CrI) of the number of Attributable Deaths (AD) and Attributable Years of Life Lost (AYLL) from tracheal, bronchus, and lung cancer expected in 2023 in Tuscany, due to annual levels of air pollutant ($PM_{2.5}$, PM_{10}, NO_2) exceeding the WHO recommended threshold, by Local Health Authority (LHA). The impact is also reported in terms of Attributable Fraction (AF), Attributable Community Rate (ACR), and AYLL per 100,000 inhabitants. For the entire region, the impact is computed considering both the WHO recommended thresholds and the minimum annual average concentration in 2015 predicted across the districts.

	Air pollutant	Threshold	AD		AF (×100)		ACR (×100,000)		AYLL		AYLL (×100,000)	
			Median	(50% CrI)	Median	(50% CrI)	Median	(50% CrI)	Median	(50% CrI)	Median	(50% CrI)
Center LHA	$PM_{2.5}$	5 μg/m^3	224	(90;351)	19.1	(7.7;29.7)	14	(6;22)	3,347	(1,349;5,259)	207	(84;326)
	PM_{10}	15 μg/m^3	76	(0;150)	6.5	(0;12.8)	5	(0;9)	1,141	(0;2,241)	71	(0;139)
	NO_2	10 μg/m^3	133	(79;188)	11.3	(6.7;15.9)	8	(5;12)	1,961	(1,165;2,783)	121	(72;172)
North-West LHA	$PM_{2.5}$	5 μg/m^3	147	(57;242)	15.7	(6.2;25.9)	12	(5;19)	2,221	(869;3,655)	176	(69;290)
	PM_{10}	15 μg/m^3	46	(0;93)	5.0	(0;9.9)	4	(0;7)	696	(0;1,399)	55	(0;111)
	NO_2	10 μg/m^3	68	(40;97)	7.3	(4.2;10.4)	5	(3;8)	1,021	(593;1,449)	81	(47;115)
South-East LHA	$PM_{2.5}$	5 μg/m^3	51	(12;113)	9.3	(2.3;20.8)	6	(1;14)	773	(189;1,720)	93	(23;208)
	PM_{10}	15 μg/m^3	7	(0;18)	1.2	(0;3.4)	1	(0;2)	102	(0;282)	12	(0;34)
	NO_2	10 μg/m^3	4	(1;7)	0.7	(0.3;1.4)	0.4	(0;1)	56	(21;111)	7	(2;13)
All region	$PM_{2.5}$	5 μg/m^3	432	(174;705)	16.3	(6.6;26.5)	12	(5;19)	6,500	(2,624;10,613)	176	(71;287)
	PM_{10}	15 μg/m^3	134	(0;264)	5.0	(0;9.9)	4	(0;7)	2,010	(0;3,972)	54	(0;107)
	NO_2	10 μg/m^3	207	(123;291)	7.8	(4.6;10.9)	6	(3;8)	3,084	(1,821;4,348)	83	(49;117)
	$PM_{2.5}$	minimum	368	(144;588)	13.9	(5.4;22.1)	10	(4;16)	5,521	(2,173;8,846)	149	(59;239)
	PM_{10}	minimum	204	(0;402)	7.7	(0;15.1)	5	(0;11)	3,078	(0;6,049)	83	(0;163)
	NO_2	minimum	304	(181;422)	11.5	(6.8;15.9)	8	(5;11)	4,546	(2,715;6,316)	123	(73;171)

Table 2
First-order (S_i) and total variance indexes (S_{tot}) describing the impact of each model input[a] on attributable deaths (AD) and years of life lost (AYLL) due to annual levels of $PM_{2.5}$ exceeding 5 μg/m^3 or exceeding the minimum regional concentration, for the region as a whole. Negative indexes are due to poor approximation.

Reference scenario	Input	Output			
		AD		AYLL	
		S_i (x100)	$S_{tot}(x100)$	S_i (x100)	$S_{tot}(x100)$
WHO	Deaths	0.3	0.4	0.5	0.5
	RR	83.8	93.9	83.6	93.9
	$PM_{2.5}$	< 0.01	13.8	< 0.01	13.9
	LE	–	–	0.0	0.0
Minimum	Deaths	0.3	0.4	0.5	0.6
	RR	87.6	93.4	87.5	93.5
	$PM_{2.5}$	< 0.01	22.7	< 0.01	23.0
	x_0	18.0	17.6	17.9	17.5
	LE	–	–	0.0	0.0

[a] Deaths: Number of tracheal, bronchus and lung cancer deaths; RR: de-response function; LE: Life expectancy.

6;22) and 207 attributable years of life lost (50% CrI: 84;326) every 100,000 inhabitants. On the other hand, it should be noted that the impact estimates in this LHA are more precise than in the others, with a probability of a positive number of attributable events always larger than 80% in every district, due to the majority of monitoring stations being located there.

The formulation of a multivariate Bayesian geostatistical model on the measured concentrations of $PM_{2.5}$, PM_{10}, and NO_2 allowed us to utilize all available information about these air pollutants to infer the posterior predictive distribution of each. This multivariate approach, sometimes called co-kriging, is not new in the literature (Brown et al., 1994; Furrer and Genton, 2011), but it is not commonly used for jointly

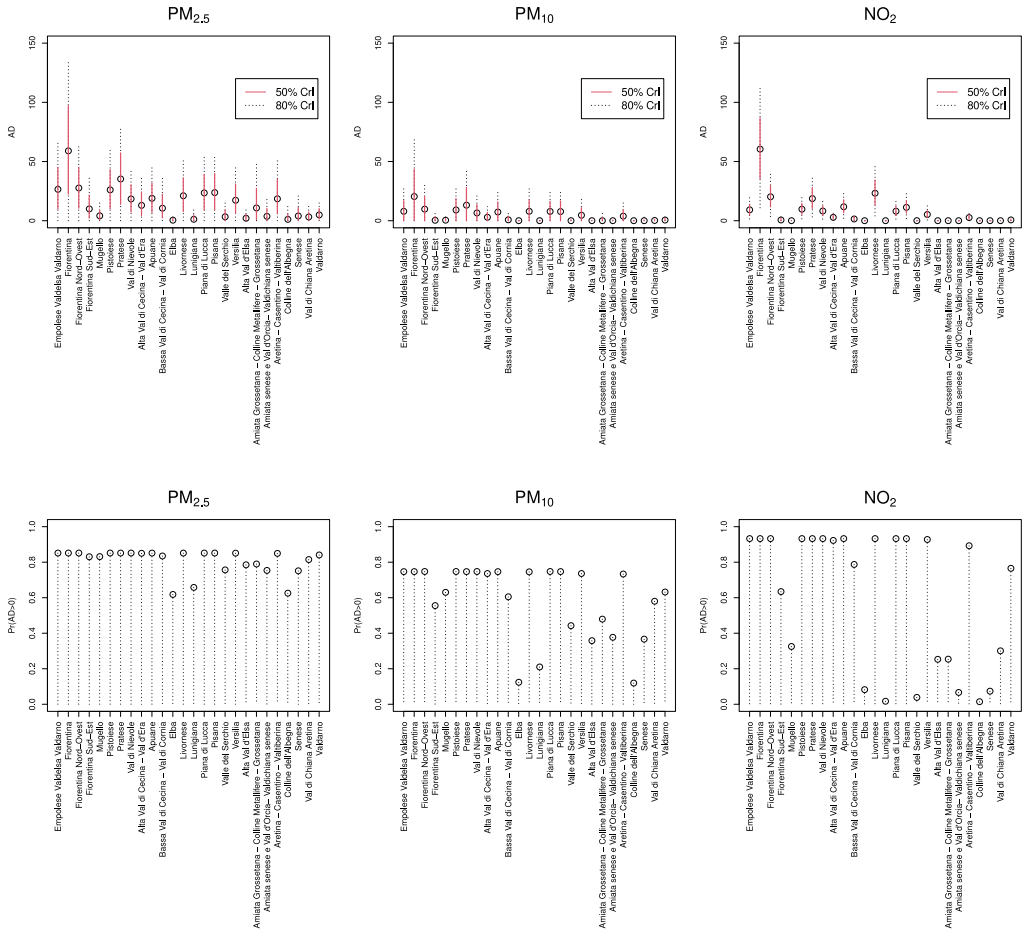

Fig. 5. Posterior median (black circle) and credibility intervals (80% CrI: black dashed interval; 50% CrI: red solid interval) of the number of deaths from tracheal, bronchus and lung cancer in 2023 in Tuscany attributable to annual levels of air pollutant ($PM_{2.5}$, PM_{10}, NO_2) exceeding the WHO recommended threshold (upper panel), and posterior probability of a positive number of attributable deaths (lower panel), by district. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

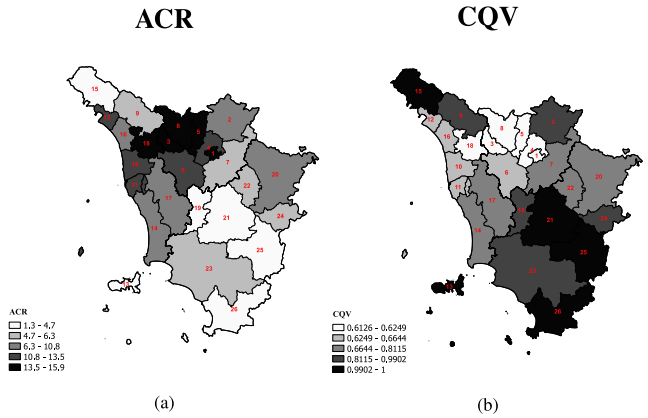

Fig. 6. Posterior medians of the Attributable Community Rates expected for the year 2023 in the 26 districts of Tuscany, setting to 5 μg/m^3 the counterfactual threshold for the annual average concentration of $PM_{2.5}$ (a), and the corresponding posterior coefficient of quartile variation (CQV) (b). The bold outlines indicate the Local Health Authorities: Center (1–8), North-West (9–18), South-East (19–26).

modeling multiple air pollutant exposures. A drawback of co-kriging is the need to specify cross-covariances between model outputs – air pollutants' annual average concentrations in our case – at different locations, which may lead to a complex model with many unknown parameters. In our application, we introduced constraints in the model to reduce the number of parameters to be estimated.

The specification of a multivariate kriging model partially addressed the issue of the limited number of sensors, particularly those measuring $PM_{2.5}$. However, the problem of the non-homogeneous distribution of monitoring stations across the territory remains, with a prevalence of sensors in areas expected to be the most polluted in the region based on modeling. An approach that accounts for this preferential sampling could be employed to obtain more reliable spatial distributions of air pollutant concentrations (Diggle et al., 2010; Cecconi et al., 2016).

The GSA suggests that the uncertainty around the predicted impacts could be significantly reduced by improving our knowledge of the exposure-response function describing the relationship between air pollution and mortality from TBL cancer in the study area. In the absence of local results on the effect of air pollution, we derived the relative risks from the literature and accounted for both sampling variability and between-study heterogeneity of the effects by sampling from the posterior predictive distributions. It is worth noting that the between-study heterogeneity was very high for all air pollutants, possibly reflecting both artefactual differences, such as those arising from discrepancies in study design and statistical analysis, and biological differences, such as variations in particle composition or population frailty. For $PM_{2.5}$ and NO_2, it was possible to rely on recent estimates of long-term effects on mortality derived from seven European cohorts, including one Italian cohort (Stafoggia et al., 2022). The same was not possible for PM_{10}, for which we performed an umbrella meta-analysis. Note that the choice to use, when possible, estimates obtained in

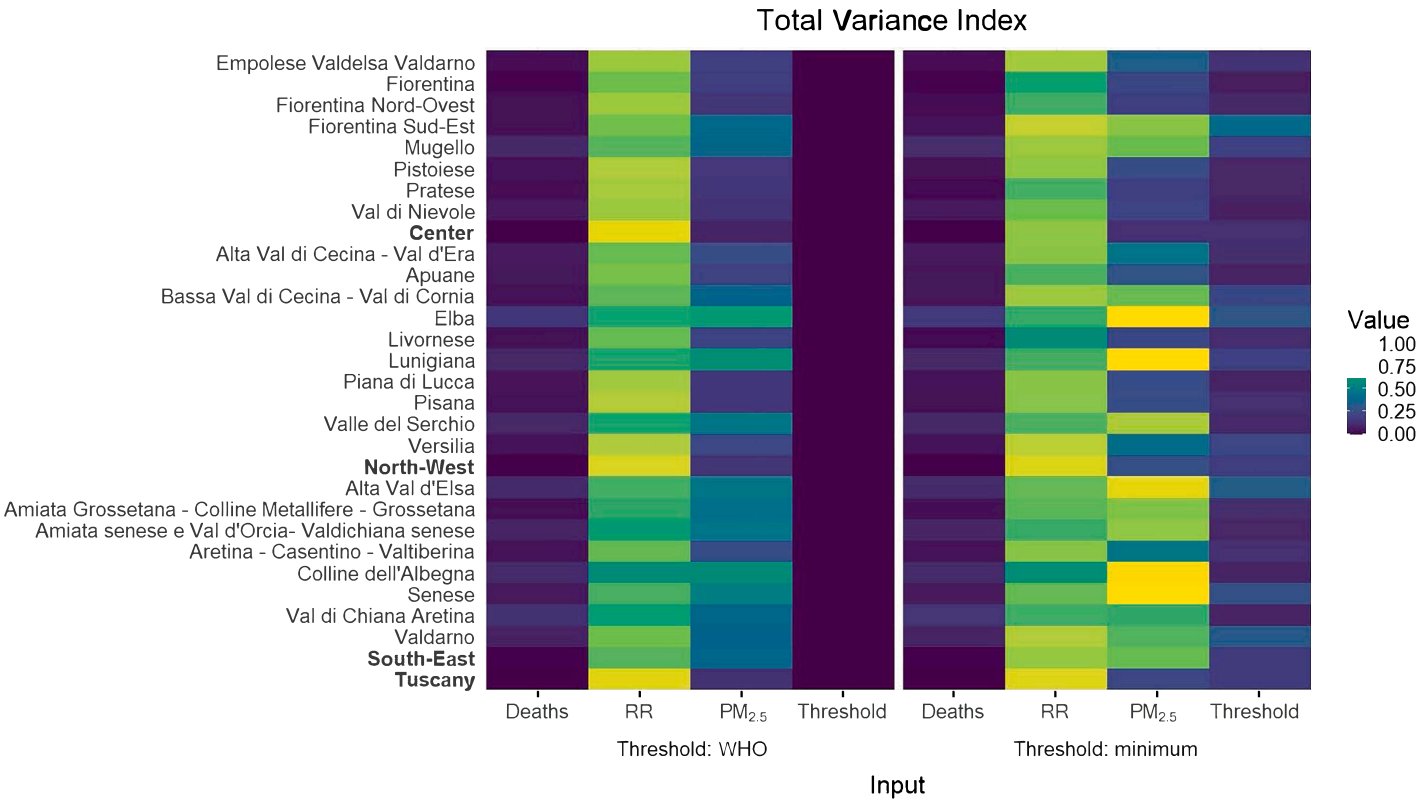

Fig. 7. Total variance indexes describing the impact of each model input on deaths attributable to annual levels of $PM_{2.5}$ exceeding 5 μg/m^3 or exceeding the minimum regional concentration, by district.

contexts similar to the one of interest – rather than extending the meta-analysis to different locations around the world – was specifically aimed at addressing the complex issue of effect heterogeneity. In principle, the optimum would be to obtain local effect estimates, but this would require individual georeferenced longitudinal data that, unfortunately, at the moment are not available for Tuscany.

A possible source of heterogeneity in RR meta-analyses was the air pollution levels observed in the study areas. However, it can be very difficult to disentangle the contribution of this factor from that of socio-economic conditions or demographics, especially when combining results from a small number of studies, as in our case. For this reason, we preferred to use the overall meta-analytic effect, correctly accounting for heterogeneity between cohorts (or studies), instead of selecting RRs estimated in areas where the air pollution levels were similar to those observed in Tuscany, or adopting meta-regression approaches.

The GSA suggests that the uncertainty around the exposure also strongly affects the output variance, especially in the Southern districts. This is another direct consequence of the non-homogeneous distribution of air quality monitoring stations in the region. Implementing a more homogeneous monitoring design in the region, based on a larger number of monitoring stations, would allow for more precise predictions of air pollutant levels and more accurate impact assessments, in addition to providing data for calibrating models such as the one developed by LaMMA. It should be noted that the issue of inadequate monitoring in the Southern areas of the region has not been resolved over the years. In fact, the number and location of monitoring stations in the region have remained almost the same from 2015 to today.

An interesting aspect of the proposed approach is that it can be used, as in this application, to perform medium-term predictions of the impact of air pollution on cause-specific mortality when mortality data for the year of interest are not yet available. However, it should be stressed that the approach relies on the assumption that mortality rates are heterogeneous among areas but constant over time. Age-period-cohort models should be applied to mortality data to obtain valid longer-term predictions of the number of cause-specific deaths (Su, 2023).

A limitation of our study is that we did not account for residence changes in the individual histories of the population and assigned the same average level of exposure to all subjects residing in the same district. Considering finer spatial aggregates might lead to more specific exposure assessments, but at the cost of unstable predictions of deaths from TBL cancer. Furthermore, we assumed a latency time of 8 years – defined as a compromise between data availability and epidemiological considerations – without accounting for cumulative exposure from 2015 to 2023. It is worth underlying that the definition of exposure in our analysis derives from the consideration that in health impact assessment it is important that there is consistency between effect measure and exposure indicator. Since we used published effect estimates, we preferred to be in line and coherent with the design of the studies from which they were derived, rather than considering more complex exposure definitions. For example, the ELAPSE project (Stafoggia et al., 2022) did not consider cumulative exposure, but calculated the effect on mortality of average exposures measured during a reference year. Similarly, we did not consider cumulative exposure, but the level of air pollution in 2015 as a proxy for the exposure of interest. We chose 2015 because for this year both predictions from the deterministic model developed by LaMMA and air pollution measurements from a sufficient number of air quality monitors were available. Note that the yearly average exposure at the health district level has not changed that rapidly over time. For example, the average level of $PM_{2.5}$ calculated at the available monitoring stations was 15.8 μg/m^3 in 2010 and 17.2 μg/m^3

in 2015, with a correlation between years equal to 0.965.

Another potential limitation concerns our assumption of a log-linear exposure-response function. If this assumption is strongly violated, our impact estimates could be underestimated or overestimated, depending on the actual shape of the curve. The log-linearity could be relaxed by using meta-analytic methods that allow for the estimation of a nonlinear exposure-response relationship. For example, the meta-regression Bayesian, regularized, and trimmed (MR-BRT) method combines estimates from published studies conducted in different locations around the world into a flexible exposure-response relationship while imposing constraints on the overall shape of the curve (Yang et al., 2022; Zheng et al., 2021). However, the MR-BRT model requires including a large number of studies in the meta-regression, which contrasts with our choice to focus, when possible, on recent European results.

When interpreting our YLL estimates, it should be considered that we used female life expectancy for both sexes. This choice was guided by the goal of reflecting the maximum potential for years of life lost, following practices used in global disease burden studies, where a benchmark of aspirational life expectancy is often used to ensure comparability across populations (von der Lippe et al., 2020). However, this may have led to an overestimation of YLL among males, especially because the gap in life expectancy between the sexes is mainly driven by factors that are not related to air pollution. For a more nuanced perspective, sex-specific life expectancy benchmarks should be employed in YLL computation.

Finally, although this study focuses on the impact of air pollution on TBL cancer mortality, it is important to recognize that smoking is a well-established major risk factor for TBL occurrence. The interaction between smoking and air pollution exposure may compound the risk, suggesting that individuals who are current or former smokers and live in highly polluted areas could be at even greater risk (Li et al., 2023). This potential synergistic effect underscores the importance of considering both factors in health impact assessment and planning of public health strategies.

In conclusion, the study suggests that the impact of previous exposure to air pollution on mortality from TBL cancer primarily affects people residing in densely populated and industrialized urban districts of the Center and North-West LHAs. Possible interventions in these areas could include actions aimed at reducing air pollution (e.g., promoting public transport, educating the population, providing incentives for renewable energy, implementing traffic restrictions, and urban green plans) to reduce exposure load going forward. The introduction of lung cancer screening programs for early tumor detection, such as those based on low-dose computed tomography proposed within the ITALUNG clinical trial (Mascalchi et al., 2023), could also be evaluated, targeting high-risk groups such as current or former smokers, especially in areas with high levels of air pollution.

Another important message from our analysis is the need to strengthen the air quality monitoring network in the region, with special attention to areas where coverage is currently poor. There is also a need for further investigation into the effects of air pollution through local cohort studies and exploration of the reasons for the high between-area heterogeneity reported in the literature. Finally, it is important to encourage high-quality health impact assessments at the local level as an essential complement to global-level assessments.

CRediT authorship contribution statement

Michela Baccini: Writing – original draft, Supervision, Methodology, Visualization, Investigation, Funding acquisition, Conceptualization. **Federico Pirona:** Writing – original draft, Visualization, Software, Formal analysis. **Laura Grisotto:** Writing – review & editing, Methodology, Visualization, Software, Formal analysis. **Giulia Cereda:** Writing – review & editing, Writing – original draft, Methodology. **Alessio Lachi:** Writing – review & editing, Methodology. **Miriam Levi:** Writing – review & editing, Investigation, Funding acquisition, Data curation. **Giulia Carreras:** Writing – review & editing, Investigation, Funding acquisition, Data curation, Conceptualization.

Funding

The present work is part of the Attributable Cancer Burden in Tuscany (ACAB) project, funded by Regione Toscana under the grant "Bando Ricerca Salute 2018" (www.acab-toscana.it). The funding body did not play a role in the design of the study and in the collection, analysis, and interpretation of the data and in the writing of the manuscript.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank the LaMMA Consortium (Environmental Monitoring and Modeling Laboratory for the Sustainable Development) for the contribution to data collection and validation, and all ACAB project participants for their support.

Appendix A. Supplemental methods

A.1. Life expectancy

We estimated in a flexible way the age-specific life expectancy for females $LE(a)$, $a = 0, 1, \dots, 100$, based on the mortality and population data in the period 2016–2018, by using a Bayesian generalized additive model (Wood, 2016). We assumed that the total number of deaths at age a observed during the three years 2016–2018 among women in Tuscany, d_a^F, followed a Poisson distribution with mean $\lambda^F(a)$ and offset given by the female population during the three years, N_a^F:

$$d_a^F \sim Poisson(\lambda^F(a)) \qquad \log \lambda^F(a) = s^F(a) + \log N_a^F, \qquad \text{(A.1)}$$

where $s^F(a)$ is a penalized thin plate regression spline on age (Wood, 2006, 2016). The spline can be written as a linear combination of K basis functions $B_k^F(a)$:

$$s(a) = \sum_{k=1}^{K} \alpha_k^F B_k^F(a),$$

where $\boldsymbol{\alpha}^F = \{\alpha_1^F, \dots \alpha_K^F\}$ is a vector of unknown spline coefficients on which a constraint is defined according to the parametrization proposed in Wood (2016). In our analysis we set $K = 10$. We assumed an uninformative multivariate Normal prior distribution on $\boldsymbol{\alpha}^F$, then we obtained a sample from its posterior distribution via the algorithm implemented in the `rjags` library of R software, after specifying the model through the `jagam` function of the `mgcv` library. For each vector of coefficients sampled from the posterior distribution of $\boldsymbol{\alpha}^F$, a spline was computed, as well as the corresponding age-specific rates $\exp(s^F(a))$, successively used to calculate the life expectancies $LE(a)$ (Cutler and Ederer, 1958). Note that, in this way, for each sample from the posterior distribution of $\boldsymbol{\alpha}^F$, we obtained a sample from the posterior distribution of the life expectancies vector.

A.2. Number of deaths

In order to compute YLL, one can plug in the number of deaths from trachea, bronchus, and lung (TBL) cancer, into Eq. (1). However, being mortality data for the year 2023 still unavailable at the time we carried out the analyses, we sampled the values $D(s, r, a)$ from their posterior predictive distribution by adopting, separately by sex and LHA the following procedure:

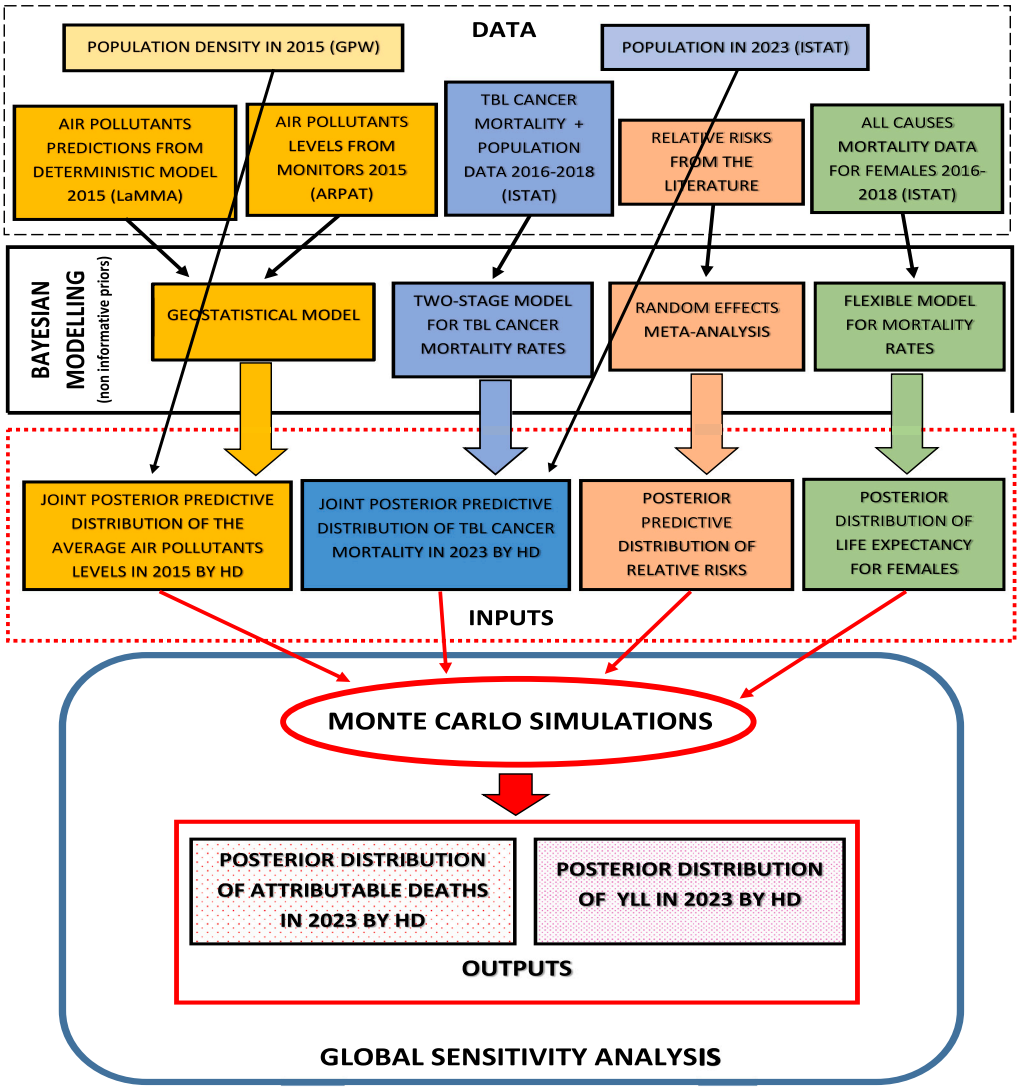

Fig. A.1. Workflow of the study. TBL: Tracheal, broncus and lung; HD: Health district; YLL: Years of Life Lost.

Table A.1
Estimates of the Relative Risks (RR) of lung cancer mortality associated with an increase of 5 μg/m^3 of $PM_{2.5}$ and to an increase of 10 μg/m^3 NO_2 with their 95% Confidence Intervals (95% CI), in the seven cohorts of the European multicenter study by Stafoggia et al. (2022).

Cohort	$PM_{2.5}$			NO_2		
	RR	95% CI	se	RR	95% CI	se
Belgian cohort	0.983	(0.950;1.017)	0.017386	1.060	(1.042;1.078)	0.008665
Danish cohort	1.214	(1.165;1.265)	0.021008	1.178	(1.153;1.203)	0.010830
Dutch cohort	1.169	(1.121;1.218)	0.021171	1.091	(1.069;1.114)	0.010519
English cohort	1.043	(0.960;1.133)	0.042269	1.039	(1.001;1.078)	0.018905
Norwegian cohort	1.180	(1.143;1.219)	0.016422	1.121	(1.094;1.148)	0.012291
Rome cohort	1.013	(0.940;1.091)	0.038003	1.027	(0.999;1.056)	0.014156
Swiss cohort	1.113	(1.078;1.150)	0.016494	1.138	(1.113;1.164)	0.011430

- We defined for each health district r in the LHA a Bayesian Poisson additive model as the one in Eq. (A.1) on the total number of deaths from TBL cancer by age, d_a, observed during the period 2016-2018:

$$d_a \sim Poisson(\lambda(a)) \qquad \log \lambda(a) = s(a) + \log N_a,$$

where $s(a) = \sum_{k=1}^{K} \alpha_k B_k(a)$ is a thin plate regression spline on age, with constrained parameters $\boldsymbol{\alpha} = \{\alpha_1, \dots \alpha_K\}$ (Wood, 2016);
- We combined in a Bayesian multivariate random effects meta-analysis the results of the inference obtained at the previous step. Let $\hat{\boldsymbol{\alpha}}_r$ be the posterior mean of the vector of the 10 spline coefficients for the district r and Σ_r the corresponding variance–covariance matrix. We specified the following meta-analysis model:

$$\hat{\boldsymbol{\alpha}}_r \mid \theta_r, \Sigma_r \sim N(\theta_r, \Sigma_r) \qquad \theta_r \mid \boldsymbol{\mu}, \mathrm{T} \sim N(\boldsymbol{\mu}, \mathrm{T}),$$

where $\boldsymbol{\mu}$ and T are unknown hyperparameters representing respectively the overall meta-analytic vector of the spline parameters and the between-district variance–covariance matrix. After defining uninformative distribution on the model hyperparameters, we obtained a sample from the posterior distribution of $\boldsymbol{\mu}$ and T via Markov Chain Monte Carlo (MCMC), by using the

Table A.2
Studies included in the Bayesian random effects meta-analysis on the long-term effect of PM_{10} on lung cancer mortality. For each study, the table reports: relative risk (RR) associated to a 10 μg/m^3 increase of air pollutant concentration, 95% Confidence Interval (95% CI), inclusion in the meta-analyses by Kim et al. (2018), Chen and Hoek (2020), Yu et al. (2021).

Study	RR	95% CI	Included in
Abbey 1999[a]	1.377	(1.095;1.731)	Kim et al. (2018), Chen and Hoek (2020), Yu et al. (2021)
Carey 2013	1.034	(0.876;1.219)	Kim et al. (2018), Chen and Hoek (2020), Yu et al. (2021)
Chen 2016	1.047	(1.034;1.060)	Kim et al. (2018), Chen and Hoek (2020), Yu et al. (2021)
Fischer 2015	1.260	(1.216;1.306)	Kim et al. (2018), Chen and Hoek (2020), Yu et al. (2021)
Hales 2012	1.160	(1.041;1.293)	Kim et al. (2018), Yu et al. (2021)
Hart 2011	0.998	(0.873;1.142)	Kim et al. (2018), Chen and Hoek (2020), Yu et al. (2021)
Heinrich 2013	2.389	(1.348;4.234)	Kim et al. (2018), Chen and Hoek (2020), Yu et al. (2021)
Hansell 2016	1.600	(1.288;1.987)	Chen and Hoek (2020)
Huss 2010	1.050	(1.035;1.065)	Chen and Hoek (2020)
Katanoda 2011	1.160	(1.068;1.260)	Chen and Hoek (2020)
Kim 2017	0.960	(0.818;1.127)	Chen and Hoek (2020)
Lipsett 2011	0.930	(0.809;1.069)	Kim et al. (2018), Chen and Hoek (2020), Yu et al. (2021)
Nishiwaki 2013	0.910	(0.773;1.071)	Chen and Hoek (2020)
Pope 2002[b]	0.980	(0.950;1.010)	Yu et al. (2021)
Zhou 2014	1.010	(0.991;1.030)	Chen and Hoek (2020)

[a] Result of a weighted average between RRs for males and females.
[b] The study by Pope et al. (2002) refers to the effect of $PM_{2.5-15}$.

`mvma.bayesian` command of the `altmeta` package of R software. Note that we modified the output of the original function in order to obtain the whole sample of the posterior distribution of the spline coefficients;

- We derived the posterior distribution of the second-stage district-specific spline coefficients $\boldsymbol{\theta}_r$ through a shrinkage formula that combines the posterior mean of overall meta-analytic vector $\hat{\boldsymbol{\mu}}$ with the first-stage vectors $\hat{\boldsymbol{\alpha}}_r$, giving them different weights, defined according to the relative importance of sampling variability and between-district heterogeneity (Röver, 2020).[1] Note that the second-stage district-specific splines, having coefficients $\boldsymbol{\theta}_r$, update the first-stage ones taking advantage of the information from all districts in the LHA;
- We used the joint posterior distribution of the district-specific parameters $\boldsymbol{\theta}_r$, combined with the observed population, to derive a posterior distribution of the expected number of deaths from TBL cancer at the district level for the year t;
- Following a Poisson distribution, we sampled around the values from the district-specific posterior distributions obtained at the previous step. This process resulted in a sample from the posterior predictive distribution of the number of deaths $D(s,r,t,a)$.

A.3. Air pollutant effect

We conducted separate Bayesian random effects meta-analysis for each pollutant, after collecting information on the log relative risks from the literature. Specifically, for $PM_{2.5}$ and NO_2, we combined the effects estimated in the seven European cohorts included in a recent two-stage multicenter study (Stafoggia et al., 2022) (Table A.1). On the contrary, for PM_{10} we conducted a rapid umbrella review, detecting three main recent meta-analyses that included published studies on the effect of PM_{10} on lung cancer mortality: the meta-analysis of cohort and case-control studies in Chen and Hoek (2020), and the meta-analyses of cohort studies in Kim et al. (2018) and Yu et al. (2021). We included in our Bayesian meta-analysis all those studies that were present in at least one of the three meta-analyses and that focused on PM_{10} exposure and lung cancer mortality, excluding those that investigated survival in cancer patients (Table A.2). Note that, while Stafoggia et al. (2022) focus on TBL cancer (ICD-9: 162) and Chen and Hoek (2020) on all malignant neoplasms of the respiratory and intrathoracic organs (ICD 10: C30–C39), Kim et al. (2018) and Yu et al. (2021) refer generically to lung cancer, without reporting specific ICD codes.

Let $(b_1,\dots,b_J)$ and $(se_1,\dots,se_J)$ be the effect estimates (log RRs) and the estimated standard errors of the J studies that we decided to include in the meta-analysis for a specific air pollutant. We assumed the following Normal-Normal meta-analytic model:

$$b_j = \beta + \zeta_j + e_j \qquad \zeta_j \sim N(0,\tau^2) \qquad e_j \sim N(0, se_j{}^2), \tag{A.2}$$

In Eq. (A.2), β and τ^2 are model hyperparameters representing the overall meta-analytic effect and the heterogeneity variance, respectively; ζ_j, $j=1,\dots,J$ are independent study-specific random effects, and e_j, $j=1,\dots,J$ independent error terms, such that $\zeta_j \perp e_k \ \forall j,k$. After specifying uninformative distributions on β and τ^2, $\beta \sim N(0,10^3)$ and $\tau^2 \sim IG(10^{-3},10^{-3})$, we obtained a sample of size 20,000 from their joint posterior distribution via MCMC, after a burnin of 5,000 iterations, by using the *bayesmh* function of STATA (version 17.0). Finally, a sample from the posterior predictive distribution of β was achieved, to account for the uncertainty in predicting the air pollutant effect in a new study or area (Riley et al., 2011).

A.4. Exposure model

We label with s the monitors in the study area, $s=1,2,3,\dots,S$, and with i the air pollutants of interest, $i=1,2,3$ (PM_{10}, $PM_{2.5}$ e NO_2, respectively).

Let x_{si} be the level of the air pollutant i measured by monitor s, and z_{si} the logarithm of the average level of the air pollutant i as estimated by the deterministic model developed by LaMMA in the cell which the monitor s belongs to. For each i, the following linear regression model is defined:

$$\log(x_{si}) = \gamma_i^0 + \gamma_i^{(1)} z_{s1} + \gamma_i^{(2)} z_{s2} + \gamma_i^{(3)} z_{s3} + u_{si} + \phi_s + \epsilon_{si}, \tag{A.3}$$

where $\gamma_i^{(0)}$, $\gamma_i^{(1)}$, $\gamma_i^{(2)}$, $\gamma_i^{(3)}$ are unknown coefficients; $\epsilon_{si} \sim N(0,\xi_i^2)$ are error terms, such that $\epsilon_{si} \perp \epsilon_{tj}$ if $i \neq j$ or $s \neq t$; $\phi_s \sim N(0,\xi_\phi{}^2)$ are monitor-specific random effects describing the unstructured heterogeneity among monitors, such that $\phi_s \perp \phi_t$ if $s \neq t$; u_{si} are random effects describing the structured spatial heterogeneity. In particular, indicating with $\mathbf{u}_i$ the vector of the random terms $u_{1i},\dots,u_{Si}$ for the pollutant i, we assume the following multivariate Normal distribution:

$$\mathbf{u} = \begin{bmatrix}\mathbf{u_1}\\ \mathbf{u_2}\\ \mathbf{u_3}\end{bmatrix} = N\left(\begin{bmatrix}\mathbf{0}\\ \mathbf{0}\\ \mathbf{0}\end{bmatrix}, \begin{bmatrix}\Sigma_{11} & \Sigma_{12} & \Sigma_{13}\\ \Sigma_{21} & \Sigma_{22} & \Sigma_{23}\\ \Sigma_{31} & \Sigma_{32} & \Sigma_{33}\end{bmatrix}\right),$$

[1] The $\boldsymbol{\theta}_r$ are Normally distributed with the following mean and variance–covariance matrix:

$$E[\theta_r \mid \mathrm{T}, \boldsymbol{\alpha}_r, \Sigma_r] = \left(\Sigma_r^{-1}\boldsymbol{\alpha}_r + \mathrm{T}^{-1}\hat{\boldsymbol{\mu}}\right)A_r^{-1}$$

$$V[\theta_r \mid \mathrm{T}, \boldsymbol{\alpha}_r, \Sigma_r] = A_r^{-1} + \left(\Sigma_r^{\frac{1}{2}}\mathrm{T}^{-1}A_r^{-1}\right)^T\left(\Sigma_i^{\frac{1}{2}}\mathrm{T}^{-1}A_r^{-1}\right)$$

where $A_r = \left(\Sigma_r + \mathrm{T}\right)$.

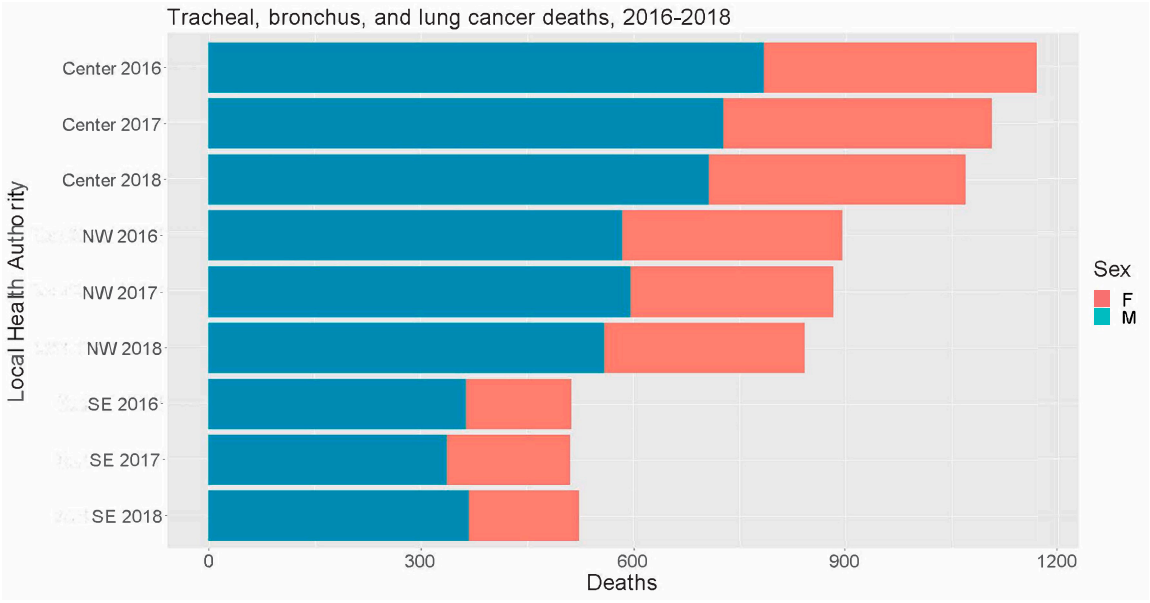

Fig. B.1. Number of deaths from tracheal, bronchus and lung cancer observed in the three Health Local Authorities in the period 2016–2018, by sex.

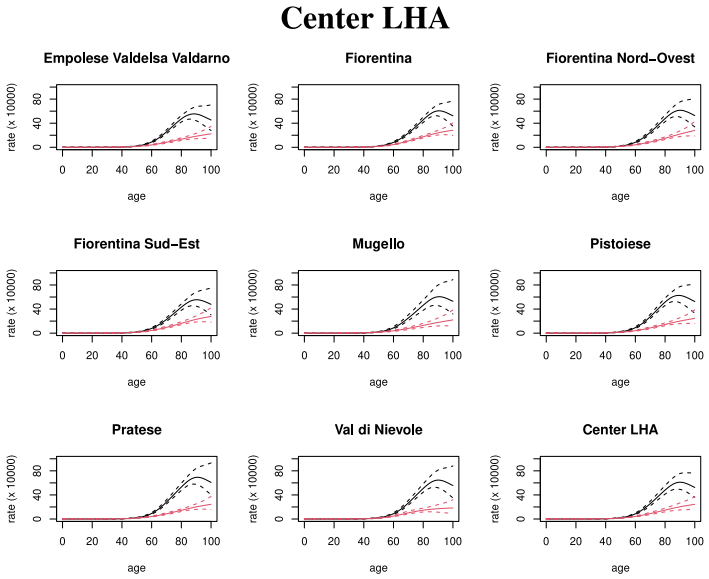

Fig. B.2. Shrinkage estimates of the age-specific mortality rates (posterior mean and 90% CrI bands) in the 8 health districts belonging to the Central Local Health Authority and in the area as a whole for the years 2016–2018, by sex (red lines for females, black lines for males). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

where Σ_{ii} is the $S \times S$ variance–covariance matrix of the levels of the air pollutant i measured by different stations, and Σ_{ij} with $i \neq j$ is the $S \times S$ variance–covariance matrix of the levels of the air pollutants i and j measured by different stations. The sth element of the diagonal of Σ_{ii}, $\Sigma_{ii}(s,s) = \sigma^2_{i_{(s)}}$, is the variance of the value of the air pollutant i in the monitor s, while the element outside the diagonal $\Sigma_{ii}(s,t) = \sigma_{i_{(s)}i_{(t)}}$ is the covariance between the levels of i measured by the monitors s and t. The sth element of the diagonal of Σ_{ij}, $\Sigma_{ij}(s,s) = \sigma_{i_{(s)}j_{(s)}}$, is the covariance between the levels of i and j measured by the monitor s, while the element $\Sigma_{ij}(s,t) = \sigma_{i_{(s)}j_{(t)}}$ is the covariance between the level of i measured by the monitor s and the level of j measured by the monitors t. Under the hypothesis that the correlation between monitors decreases with the distance, we assume:

$$\sigma_{i_{(s)}j_{(t)}} = \rho_{ij}\sigma_{i_{(s)}}\sigma_{j_{(t)}} f(d(s,t),k,\psi_i),$$

where ρ_{ij} is the correlation coefficient between i and j when measured by the same monitor ($\rho_{ii} = 1$), and $f(d(s,t),k,\psi_i)$ is a function that describes how the correlation between values of the same pollutant or between values of different pollutants measured by different monitors declines with the distance between monitors, $d(s,t)$. In particular, we assume that the distance declines according to a spatial exponential model:

$$f(d(s,t),k,\psi) = \exp(-(d(s,t)\psi_i)^k),$$

where ψ_i is the parameter that tunes the velocity of decline of the correlation as the distance increases for the pollutant i, and $k \in (0,2]$ is the parameter, shared by all pollutant, that controls the amount of spatial smoothing. We set $k = 1$ and fix ψ_i to a value calculated based on the max and min distances in the region. Notice that $f(d(s,t),k,\psi_i) = 1$ if $s = t$ and that $f(d(s,t),k,\psi_i)$ tends to 0 for large $d(s,t)$.

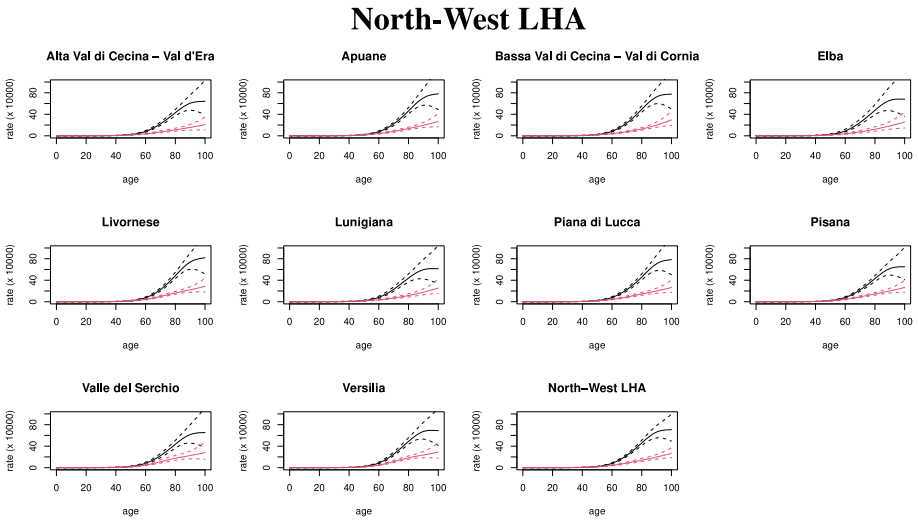

Fig. B.3. Shrinkage estimates of the age-specific mortality rates (posterior mean and 90% CrI bands) in the 10 health districts belonging to the North-West Local Health Authority and in the area as a whole for the years 2016–2018, by sex (red lines for females, black lines for males). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

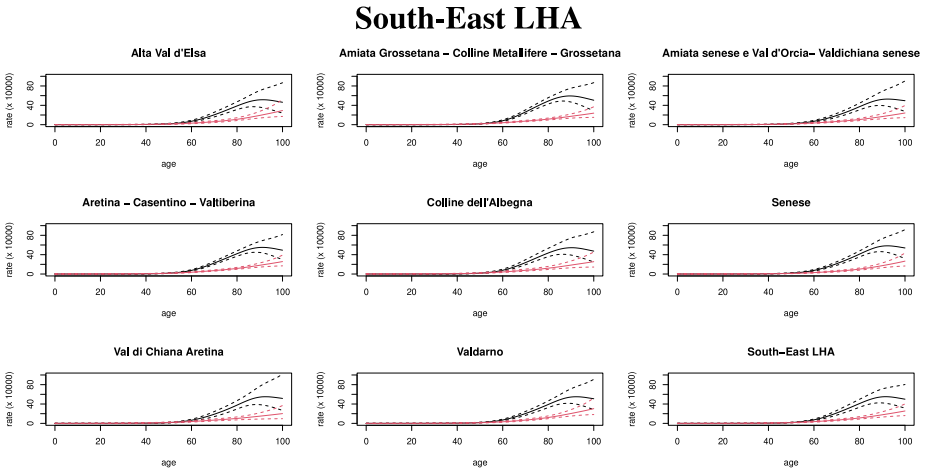

Fig. B.4. Shrinkage estimates of the age-specific mortality rates (posterior mean and 90% CrI bands) in the 8 health districts belonging to the South-East Local Health Authority for the years 2016–2018, by sex (red lines for females, black lines for males). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

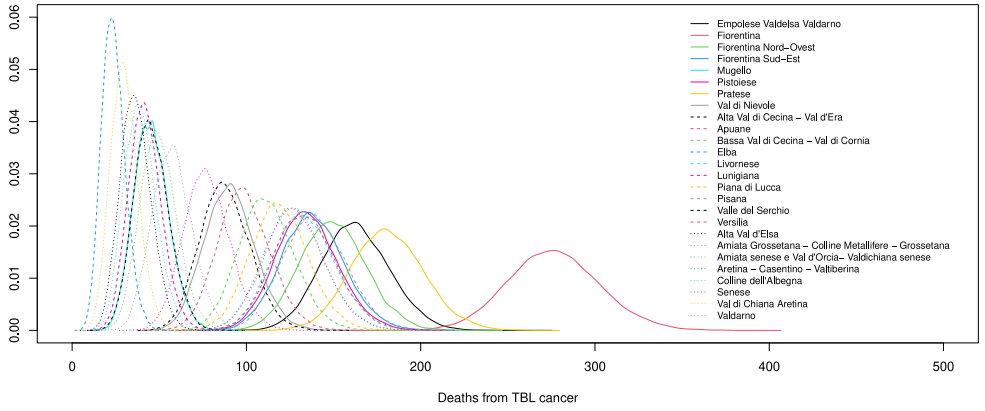

Fig. B.5. Posterior predictive distributions of the number of deaths in 2023, by health district. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

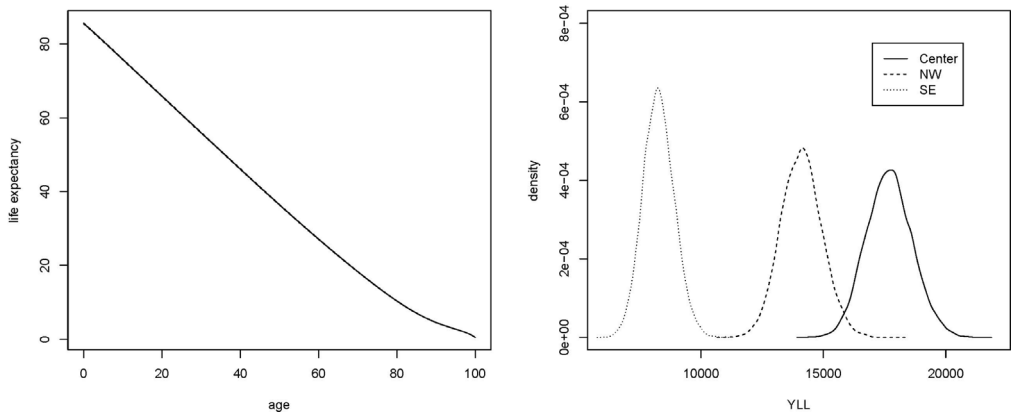

Fig. B.6. Posterior mean and 99% point-wise credibility bands of the age-specific life expectancy for female in Tuscany (2016–2018) (left panel) and posterior predictive distribution of the Years of Life Lost due to tracheal, bronchus, and lung cancer in Tuscany in 2023, by Local Health Authority (LHA) (right panel).

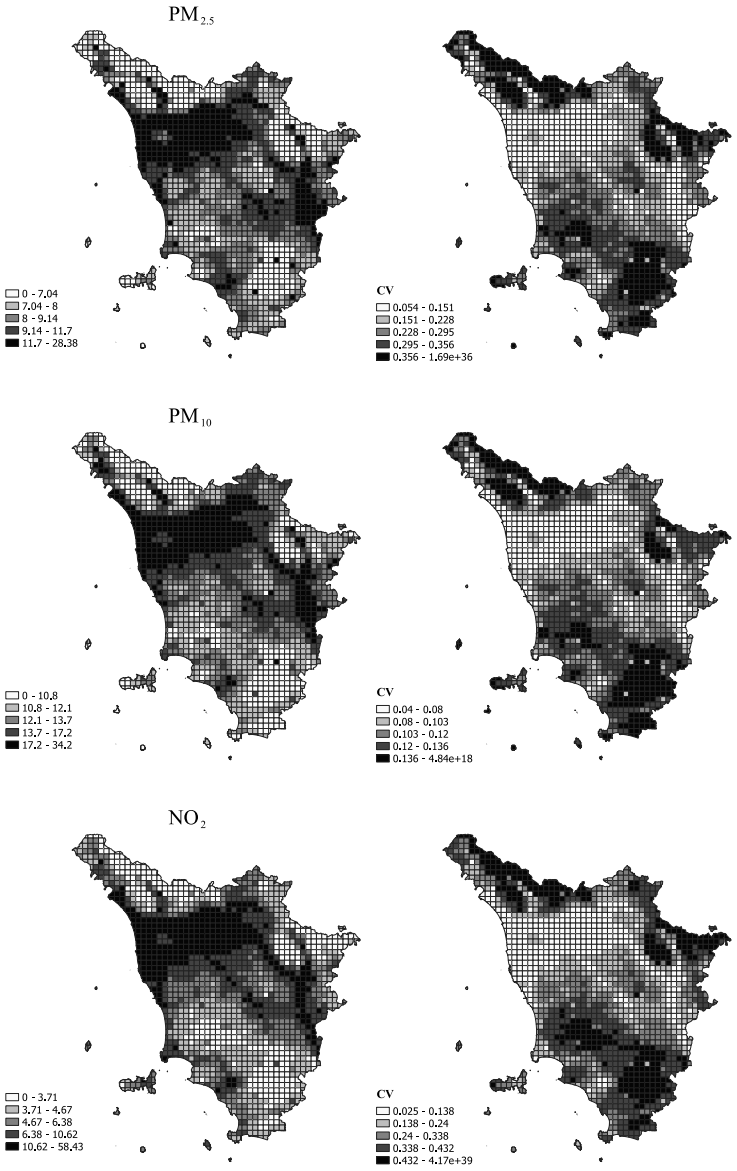

Fig. B.7. Means (left maps) and coefficients of variation (right maps) of the posterior predictive distributions of $PM_{2.5}$, PM_{10} and NO_2 average concentrations for the year 2015 by 4 × 4 km cells.

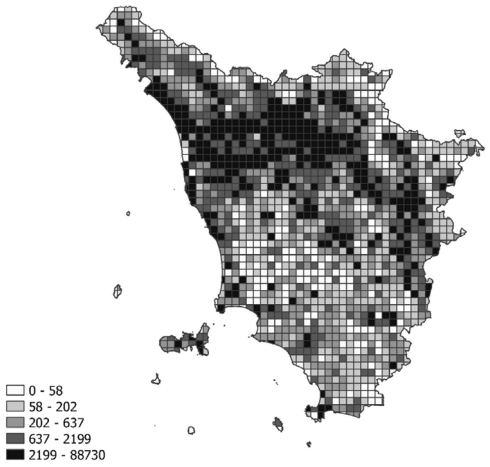

Fig. B.8. Population size in Tuscany in 2015 by 4 × 4 km cells.

Table B.1
Posterior mean and 90% credibility interval (CrI) of the number of deaths and YLL from tracheal, bronchus, and lung cancer expected in 2023 in Tuscany, by sex and Local Health Authority (LHA) (population in 2013 in the first columns).

	Sex	Population	Deaths		YLL	
			Mean	90% CrI	Mean	90% CrI
Center LHA	Males	777,338	788	(735;842)	11,714	(10,807;12,658)
	Females	827,071	392	(354;430)	5,977	(5,291;6,707)
	All	1,604,409	1,180	(1,093;1,269)	17,691	(16,194;19,263)
North-West LHA	Males	604,577	626	(578;675)	9,250	(8,442;10,094)
	Females	641,753	309	(276;343)	4,855	(4,221;5,528)
	All	1,246,330	935	(858;1,013)	14,104	(12,757;15,516)
South-East LHA	Males	394,983	379	(343;418)	5,608	(4,977;6,274)
	Females	416,259	166	(142;192)	2,661	(2,184;3,176)
	All	811,242	546	(487;605)	8,268	(7,242;9,357)
All region	Males	1,776,898	1,793	(1,695;1,893)	26,572	(24,913;28,310)
	Females	1,885,083	867	(800;936)	13,492	(12,224;14,818)
	All	3,661,981	2,660	(2,500;2,824)	40,064	(37,306;42,950)

Table B.2
Mean, 5th and 95th percentiles of the posterior distribution and of the posterior predictive distribution of relative risks of mortality from tracheal, bronchus, and lung cancer associated with an increase of 5 $\mu g/m^3$ in the annual average concentration of $PM_{2.5}$, and with an increase of 10 $\mu g/m^3$ in the annual average concentrations of PM_{10} and NO_2, with the corresponding I^2 indices (posterior median and 90% Credibility Interval), arising from Bayesian random-effects meta-analyses.

Air pollutant	Relative Risk			I^2 index	
	Mean	(90% CrI)	(predictive 90% CrI)	Median	(90% CrI)
$PM_{2.5}$	1.103	(1.037;1.169)	(0.930;1.301)	94.0	(85.7;98.2)
PM_{10}	1.090	(1.026;1.166)	(0.868;1.370)	76.8	(55.9;90.1)
NO_2	1.093	(1.054;1.133)	(0.988;1.210)	95.2	(88.5;98.6)

Table B.3
Posterior median and 50% credibility interval (CrI) of the number of deaths and Years of Life Lost (YLL) from tracheal, bronchus, and lung cancer expected in 2023 in Tuscany, attributable to annual average levels of $PM_{2.5}$ exceeding the WHO recommended threshold, by Local Health Authority (LHA) and sex. The impact is also reported in terms of Attributable Fraction (AF), Attributable Community Rate (ACR), and Attributable YLL (AYLL) per 100,000 inhabitants. For the entire region, the impact is computed also setting the threshold to the minimum annual average concentration in 2015 predicted across the districts.

	Sex	Threshold	AD		AF (×100)		ACR (×100,000)		AYLL		ACRYLL (×100,000)	
			Median	(90% CrI)	Median	(50% CrI)	Median	(50% CrI)	Median	(50% CrI)	Median	(50% CrI)
Center	Males	5 $\mu g/m^3$	149	(60;234)	19.0	(7.7;29.7)	19	(8;30)	2,213	(894;3,474)	285	(115;447)
	Females	5 $\mu g/m^3$	75	(30;117)	19.1	(7.7;29.9)	9	(4;14)	1135	(456;1,784)	136	(55;213)
NW	Males	5 $\mu g/m^3$	98	(38;161)	15.7	(6.2;25.8)	16	(6;26)	1448	(566,2,392)	237	(93,392)
	Females	5 $\mu g/m^3$	49	(19;81)	15.8	(6.2;26.1)	8	(3;12)	770	(301;1,264)	118	(46;194)
SE	Males	5 $\mu g/m^3$	35	(9;79)	9.3	(2.3;20.8)	9	(2;20)	523	(128;1,161)	130	(32;290)
	Females	5 $\mu g/m^3$	15	(4;34)	9.4	(2.3;20.8)	4	(1;8)	249	(61;552)	58	(14;130)
All region	Males	5 $\mu g/m^3$	290	(117;473)	16.2	(6.5;26.4)	16	(7;26)	4,287	(1,732;7,017)	240	(97;392)
	Females	5 $\mu g/m^3$	143	(58;232)	16.5	(6.7;26.7)	7	(3;12)	2,207	(891;3,599)	115	(47;188)
	Males	minimum	179	(68;316)	10.0	(3.8;17.6)	10	(4;18)	2,659	(1,010;4,687)	149	(56;262)
	Females	minimum	89	(34;156)	10.3	(3.9;18.0)	5	(2;8)	1,378	(527;2,412)	72	(28;126)

It should be noticed that in our model the correlation coefficient ρ_{ij} does not vary with the monitor. Similarly, the rate of decline of the correlation does not vary with the air pollutants. If additionally we set $\sigma^2_{i_{(s)}} = \sigma^2_i$, the following parametrization applies:

$$\Sigma_{ii} = \sigma_i^2 \Delta \qquad \Sigma_{ij} = \rho_{ij}\sigma_i\sigma_j\Delta,$$

where Δ is a $S \times S$ distance matrix with elements $\delta(s,t) = \exp(-(d(s,t)\psi_i)^k)$.

Priors on the hyperparameters and Bayesian inference

We specified the following priors on the model hyper-parameters:

- $\xi_\phi{}^2 \sim \text{InvGamma}(0.1, 0.1)$
- $\xi_i^2 \sim \text{InvGamma}(0.1, 0.1) \quad \forall i \in \{1,2,3\}$
- $\gamma_i^0 \sim \text{N}(0, 10^3) \quad \forall i \in \{1,2,3\}$
- $\gamma_i^{(j)} \sim N(0, 1000) \quad \forall i, j \in \{1,2,3\}$
- $\sigma_i^2 \sim \text{InvGamma}(0.1, 0.1) \quad \forall i \in \{1,2,3\}$
- $\rho_{ij} \sim U(a,b) \quad \forall j > i \quad i \in \{1,2,3\}.$

We obtained a sample from the joint posterior distribution of the model parameters via MCMC, by using WinBUGS 1.4 and the R software interface R2WinBUGS. Two chains of 100,000 iterations were generated with a burnin of 50,000 samples. One sample over 5 was selected, in order to reduce the autocorrelation among samples. Visual inspection of chains convergence was performed to assess the stability and reliability of the MCMC simulations, ensuring that the chains had sufficiently mixed and reached stationarity.

Air pollutants predictions

After reducing the resolution of the LaMMA deterministic model to a 4 × 4 km grid by averaging the original values of the 2 × 2 km cells, the joint posterior predictive distribution of the annual average concentrations of the three air pollutants at the centroid of each 4 × 4 km cell was derived by using these values as predictors in Eq. (A.3), and marginalizing over the posterior distribution of the model parameters. For each cell and air pollutant, the posterior distribution of the predicted values has been summarized in terms of mean and coefficient of variation. Note that reducing the resolution of the predictions was necessary for computational reasons, but at the same time produced a grid consistent with that of the population size (Center for International Earth Science Information Network, 2021), which was subsequently used to calculate the weighted average of air pollution levels at health district level.

Appendix B. Supplemental results

See Figs. B.1–B.8 and Tables B.1–B.3

Data availability

Data will be made available on request.

References

Baccini, M., Grisotto, L., Catelan, D., Bertazzi, P., Biggeri, A., 2015. Commuting-adjusted short-term health impact assessment of airborne fine particles with uncertainty quantification via monte carlo simulation. Environ. Health Perspect. 123, 27–33.

Berg, C., Schiller, J., Boffetta, P., Cai, J., Connolly, C., Kerpel-Fronius, A., Kitts, A., Lam, D., Mohan, A., Myers, R., Suri, T., Tammemagi, M., Yang, D., Lam, S., 2023. Air pollution and lung cancer: A review by international association for the study of lung cancer early detection and screening committee. J. Thorac. Oncol. 18, 1277–1289.

Brauer, M., Brook, J.R., Christidis, T., Chu, Y., Crouse, D.L., Erickson, A., Hystad, P., Li, C., Martin, R.V., Meng, J., et al., 2022. Mortality–Air Pollution Associations in Low Exposure Environments (Maple): Phase 2. Research Reports: Health Effects Institute 2022.

Brown, P., Le, N., Zidek, J., 1994. Multivariate spatial interpolation and exposure to air pollutants. Canad. J. Statist. 22, 489–509.

Brunekreef, B., Strak, M., Chen, J., Andersen, Z.J., Atkinson, R., Bauwelinck, M., Bellander, T., Boutron, M.C., Brandt, J., Carey, I., et al., 2021. Mortality and Morbidity Effects of Long-Term Exposure to Low-Level Pm2. 5, Bc, No2, and O3: an Analysis of European Cohorts in the Elapse Project. Research Reports: Health Effects Institute 2021.

Cecconi, L., Grisotto, L., Catelan, D., Lagazio, C., Berrocal, V., Biggeri, A., 2016. Preferential sampling and bayesian geostatistics: Statistical modeling and examples. Stat. Methods Med. Res. 25, 1224–1243.

Center for International Earth Science Information Network, 2021. Gridded Population of the World, Version 4 (GPWv4): Population Density, Revision 11. NASA Socioeconomic Data and Application Center (SEDAC), Palisades, New York.

Chen, J., Hoek, G., 2020. Long-term exposure to pm and all-cause and cause-specific mortality: A systematic review and meta-analysis. Environ. Int. 143, 105974.

Conti, S., Fornari, C., Ferrara, P., Antonazzo, I.C., Madotto, F., Traini, E., Levi, M., Cernigliaro, A., Armocida, B., Bragazzi, N.L., Cadum, E., Carugno, M., Crotti, G., Deandrea, S., Cortesi, P.A., Guido, D., Iavicoli, I., Iavicoli, S., La Vecchia, C., Lauriola, P., Michelozzi, P., Scondotto, S., Stafoggia, M., Violante, F.S., Abbafati, C., Albano, L., Barone-Adesi, F., Biondi, A., Bosetti, C., Buonsenso, D., Carreras, G., Castelpietra, G., Catapano, A., Cattaruzza, M.S., Corso, B., Damiani, G., Esposito, F., Gallus, S., Golinelli, D., Hay, S.I., Isola, G., Ledda, C., Mondello, S., Pedersini, P., Pensato, U., Perico, N., Remuzzi, G., Sanmarchi, F., Santoro, R., Simonetti, B., Unim, B., Vacante, M., Veroux, M., Villafañe, J.H., Monasta, L., Mantovani, L.G., 2023. Time-trends in air pollution impact on health in italy, 1990–2019: An analysis from the global burden of disease study 2019. Int. J. Public Heal. 68, 1605959.

Cutler, S.J., Ederer, F., 1958. Maximum utilization of the life table method in analyzing survival. J. Chronic Dis. 8, 699–712.

Diggle, P., Menezes, R., Su, T., 2010. Geostatistical inference under preferential sampling (with discussion). J. R. Stat. Soc. Ser. C. Appl. Stat. 59, 191–232.

Furrer, R., Genton, M., 2011. Aggregation-cokriging for highly multivariate spatial data. Biometrika 98, 615–631.

GBD 2019 Risk Factors Collaborators, 2020. Global burden of 87 risk factors in 204 countries and territories, 1990–2019: a systematic analysis for the global burden of disease study 2019. Lancet 396, 1223–1249.

GBD 2021 Diseases and Injuries Collaborators, 2024. Global incidence, prevalence, years lived with disability (ylds), disability-adjusted life-years (dalys), and healthy life expectancy (hale) for 371 diseases and injuries in 204 countries and territories and 811 subnational locations, 1990–2021: a systematic analysis for the global burden of disease study 2021. Lancet 403, 2133–2161.

Hamra, G.B., Guha, N., Cohen, A., Laden, F., Raaschou-Nielsen, O., Samet, J.M., Vineis, P., Forastiere, F., Saldiva, P., Yorifuji, T., et al., 2014. Outdoor particulate matter exposure and lung cancer: a systematic review and meta-analysis. Environ. Health Perspect..

Kim, H.B., Shim, J.Y., Park, B., Lee, Y.J., 2018. Long-term exposure to air pollutants and cancer mortality: A meta-analysis of cohort studies. Int. J. Env. Res. Public Heal. 15, 2608.

Li, D., Shi, J., Liang, D., Ren, M., He, Y., 2023. Lung cancer risk and exposure to air pollution: a multicenter North China case-control study involving 14604 subjects. BMC Pulm. Med. 23, 182.

Malhotra, J., Malvezzi, M., Negri, E., La Vecchia, C., Boffetta, P., 2016. Risk factors for lung cancer worldwide. Eur. Respir. J. 48, 889–902.

Mascalchi, M., Picozzi, G., Puliti, D., Diciotti, S., Deliperi, A., Romei, C., Falaschi, F., Pistelli, F., Grazzini, M., Vannucchi, L., Bisanzi, S., Zappa, M., Gorini, G., Carozzi, F., Carrozzi, L., Paci, E., 2023. Lung cancer screening with low-dose ct: What we have learned in two decades of italung and what is yet to be addressed. Diagn. (Basel) 13, 2197.

Monasta, L., Alicandro, G., Pasovic, M., Cunningham, M., Armocida, B., Murray, C.J.L., Ronfani, L., Naghavi, M., 2022. Redistribution of garbage codes to underlying causes of death: a systematic analysis on italy and a comparison with most populous western european countries based on the global burden of disease study 2019. Eur. J. Pub. Health 32, 456–462.

Murray, C.J., 2022. The global burden of disease study at 30 years. Nature Med. 28, 2019–2026.

Pope, I., Burnett, R., Thun, M., et al., 2002. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. JAMA 287, 1132–1141.

Riley, R., Higgins, J., JJ, D., 2011. Interpretation of random effects meta-analyses. BMJ 342, d549.

Röver, C., 2020. Bayesian random-effects meta-analysis using the bayesmeta R package. J. Stat. Softw. 93, 1–51.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., Tarantola, S., 2008. Global Sensitivity Analysis: The Primer. John Wiley and Sons, Ltd, Italia.

Sobol, I.M., 1994. A Primer for the Monte Carlo Method. CRC Press.

Spiegelhalter, D.J., Best, N.G., 2003. Bayesian approaches to multiple sources of evidence and uncertainty in complex cost-e ectiveness modelling. Statist. Med. 22, 3687–3709.

Stafoggia, M., Oftedal, B., Chen, J., Rodopoulou, S., Renzi, M., Atkinson, R.W., Bauwelinck, M., Klompmaker, J.O., Mehta, A., Vienneau, D., Andersen, Z.J., Bellander, T., Brandt, J., Cesaroni, G., de Hoogh, K., Fecht, D., et al., 2022. Long-term exposure to low ambient air pollution concentrations and mortality among 28 million people: results from seven large european cohorts within the elapse project. Lancet Planet Heal. 6, e9–e18.

Su, S., 2023. Synthesized age-period-cohort prediction method: Application to lung cancer mortality in Taiwan. Am. J. Epidemiol. 192, 1712–1719.

von der Lippe, E., Devleesschauwer, B., Gourley, M., Haagsma, J., Hilderink, H., Porst, M., Wengler, A., Wyper, G., Grant, I., 2020. Reflections on key methodological decisions in national burden of disease assessments. Arch. Public Heal. 78, 137.

Wood, S.N., 2006. Generalized Additive Models : An Introduction with R. Chapman and Hall/CRC.

Wood, S.N., 2016. Just Another Gibbs Additive Modeller: Interfacing JAGS and mgcv. J. Stat. Softw. 75, 7.

World Health Organization, 2021. WHO Global Air Quality Guidelines. Particulate Matter (PM2.5 and PM10), Ozone, Nitrogen Dioxide, Sulfur Dioxide and Carbon Monoxide. World Health Organization, Geneva.

Yang, X., Zhang, T., Zhang, X., Chu, C., Sang, S., 2022. Global burden of lung cancer attributable to ambient fine particulate matter pollution in 204 countries and territories, 1990–2019. Environ. Res. 204, 112023.

Yu, P., Guo, S., Xu, R., Ye, T., Li, S., Sim, M.R., Abramson, M.J., Guo, Y., 2021. Cohort studies of long-term exposure to outdoor particulate matter and risks of cancer: A systematic review and meta-analysis. The Innovation 2, 100143.

Zheng, P., Barber, R., Sorensen, R.J.D., Murray, C.J.L., Aravkin, A.Y., 2021. Trimmed constrained mixed effects models: formulations and algorithms. J. Comput. Graph. Stat. 30, 544–556.