Green space exposure and mortality in the Florentine plain: preventable deaths at the census tract level and uncertainty analysis – Supplementary Material

Michela Baccini[1], Giorgia Burbui[2], Daniela Nuvolone[3], Costanza Borghi[4], Alessio Lachi[5], Stefano Mancuso[4,6], Gherardo Chirici[4,6]

Corresponding author: Michela Baccini, michela.baccini@unifi.it

[1] Department of Statistics, Computer Science, and Applications, University of Florence, Florence, Italy

[2] Department of Architecture, University of Florence, Florence, Italy

[3] Epidemiology Observatory, Regional Health Agency of Tuscany, Florence, Italy

[4] Department of Agricultural, Food, Environmental and Forestry Sciences and Technologies, University of Florence, Florence, Italy

[5] Departmental Faculty of Medicine, Saint Camillus International University of Health and Medical Sciences, Rome, Italy

[6] Fondazione per il Futuro delle Città, Florence, Italy

This file provides additional methodological and computational details supporting the analysis presented in the main articol. Specifically, it includes:

i. a detailed description of the Normalized Difference Vegetation Index (NDVI);
ii. methodological and computational specifications for the random effect meta-analysis;
iii. the definition and implementation details of Total Variance indeces used in the Global Sensitivity Analysis (GSA).

The document also contains supplementary figures (Fig. S1-S8) and tables (Tables S1-S5), which illustrate and expand upon the results reported in the main text.

S1. NDVI definition

The NDVI is defined as the ratio between the difference and the sum of near-infrared (NIR), which vegetation strongly reflects and transmits, and visible red (RED), which vegetation absorbs:

NDVI = $\frac{(NIR - RED)}{(NIR + RED)}$.

The NDVI ranges from -1 to +1. There are not exact thresholds for specific land cover types, but generally, negative NDVI values correspond to water; values near zero (approximately -0.1 to 0.1) correspond to urbanized or arid areas with rock, sand, or snow; positive but low values represent sparse vegetation such as shrubs and grasslands (between 0.2 and 0.4); higher values (approximately 0.6 to 1) indicate denser green covers, such as forests or woodlands (Rouse et al. 1973).

S2. Random effects meta-analysis

Let $(b_1, \dots, b_J)$ and $(se_1, \dots, se_J)$ be the effect estimates (log HR or log RR) and the estimated standard errors of the J studies included in the meta-analysis. We assumed the following Normal-Normal model:

$$b_j = \beta + u_j + e_j \qquad u_j \sim N(0, \tau^2) \qquad e_j \sim N\left(0, se_j^2\right), \tag{4}$$

where β and τ^2 are hyperparameters representing the overall meta-analytic effect and the heterogeneity variance, respectively; $u_j, j = 1,2, \dots J$ are independent study-specific random effects, and $e_j, j = 1,2, \dots J$ are independent error terms; u_j and e_k are assumed to be mutually independent $\forall j, k$ (Riley et al. 2011). After specifying uninformative prior distributions on β and τ^2, three independent Markov Chain Monte Carlo (MCMC) chains of 20,000 iterations each were generated. For each chain, the first 4,000 iterations were discarded as burn-in, and a thinning interval of 5 was applied. This resulted in a total of 9,600 posterior samples used to derive the posterior predictive distribution of β.

S3. Total variance index definition and computation

The total variance index S_k^{tot} for the kth input X_k on the output Y quantifies the contribution of X_k to the variance of Y, taking into account both the individual effect of X_k and all its interactions with the other inputs (Homma and Saltelli 1996; Sobol 2001). The index can be seen as the ratio between the expected value of the conditional variance of Y given all inputs but X_k ($\mathbf{X}_{\sim k}$), and the unconditional variance of Y:

$$S_k^{tot} = \frac{E_{X \sim j}\left(V_{X_k}(Y|\mathbf{X}_{\sim k})\right)}{V(Y)}.$$

According to Sobol (2001) and Saltelli (2002, 2008), the computation of the total variance indexes was performed by calculating the model output on specific sampling matrices: the base matrices **A** and **B** of dimension $N \times K$, where N is the number of simulations from the inputs distributions and K is the number of inputs, each containing independent samples of the input parameters, and the matrices $\boldsymbol{C_j}$, j=1,2,..K, one for each input, with $\boldsymbol{C_j}$ constructed by replacing the j-th column of **A** with the corresponding column from **B**. Calculating the output for each row of the three matrices, it is possible to isolate the contribution of each input variable to the output(s) variance and this allows for the estimation of the variance-based indices (Saltelli 2002). In our analysis, we defined the matrices **A**, **B**, and $\boldsymbol{C_j}$ of dimension 4,800×3, or 4,800×4 for the WHO scenario, with a final sample size for the MC simulations respectively equal to 4,800×5=24,000 or 4,800×6=28,800. For details on the calculation of the total variance indexes from the simulated outputs arising from **A**, **B**, and $\boldsymbol{C_j}$, see Sobol (2001) and Saltelli (2002, 2008).

In our framework, a particular challenge for the computation of S_k^{tot} involves the definition in the matrices **A** and **B** of the inputs X_3, i.e. the Multinomial allocation of deaths at the census tract level, and X_4, i.e. the definition in terms of NDVI of the WHO threshold. In order to introduce the stochastic nature of the Multinomial distribution in the GSA, the input X_3 was defined as a quasi-random number in [0,1] (Sobol 1994). This value served as a seed for generating a Uniform quantile sequence, which is then used to obtain a sample from the Multinomial distribution through the inverse transform sampling method. This approach allowed us to identify the Multinomial sampling as a source of uncertainty within the GSA framework, thus, to calculate the corresponding total variance index. Following a similar approach, the uncertainty surrounding the NDVI cut-off was handled by defining the input X_4 as a quasi-random number in [0,1], then using the inverse transform sampling to generate a value from the Normal distribution.

References

Homma T, Saltelli A (1996) Importance measures in global sensitivity analysis of nonlinear models. Reliab Eng Syst Saf 52(1):1–17. https://doi.org/10.1016/0951-8320(96)00002-6

Riley RD, Higgins JPT, Deeks JJ (2011) Interpretation of random effects meta-analyses. BMJ 342:d549. https://doi.org/10.1136/bmj.d549

Rouse JW, Haas RH, Schell JA, Deering DW (1973) Monitoring vegetation systems in the Great Plains with ERTS. In: Third Earth Resources Technology Satellite-1 Symposium. NASA SP-351. National Aeronautics and Space Administration, Washington (DC), pp 309–317

Saltelli A (2002) Sensitivity analysis for importance assessment. Risk Anal 22(3):579–590. https://doi.org/10.1111/0272-4332.00040

Saltelli A, Ratto M, Andres T, Campolongo F, Cariboni J, Gatelli D, et al (2008) Global Sensitivity Analysis: The Primer. John Wiley & Sons, Chichester (UK). https://doi.org/10.1002/9780470725184.ch6

Sobol IM (1994) A Primer for the Monte Carlo Method. CRC Press, Boca Raton (FL). https://doi.org/10.1201/9781315136448

Sobol IM (2001) Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. Math Comput Simul 55(1–3):271–280. https://doi.org/10.1016/S0378-4754(00)00270-6

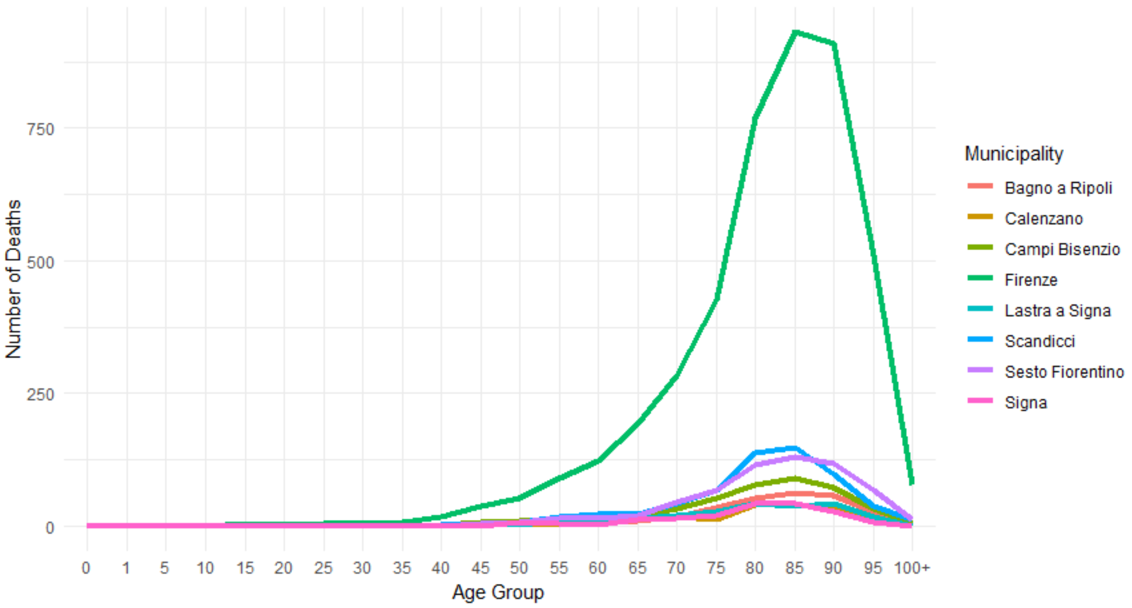

Fig. S1 Number of deaths for each municipality in the study area, by age. Year 2021

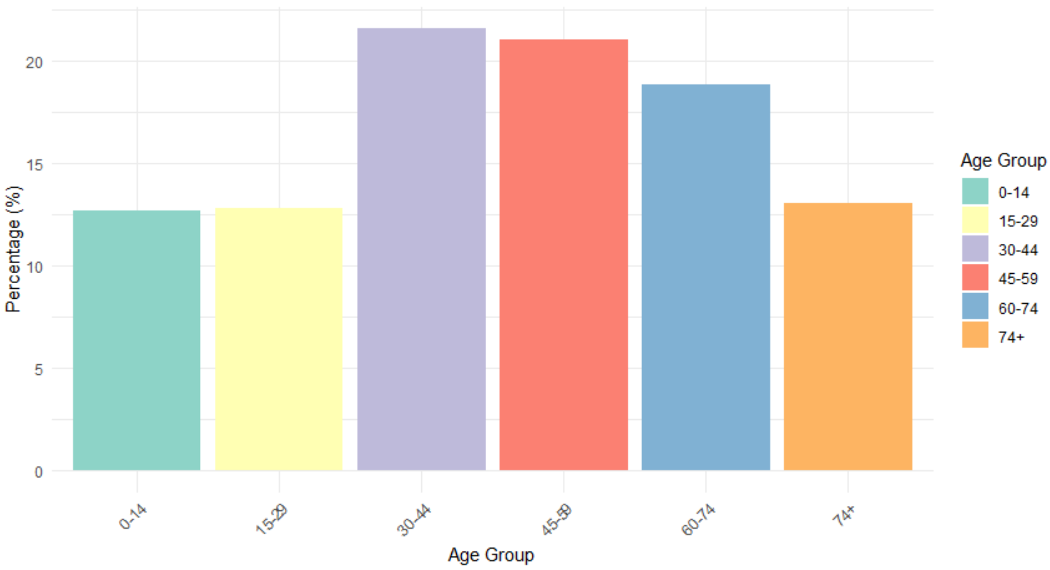

Fig. S2 Distribution by age of the population residing in the study area. Year 2011

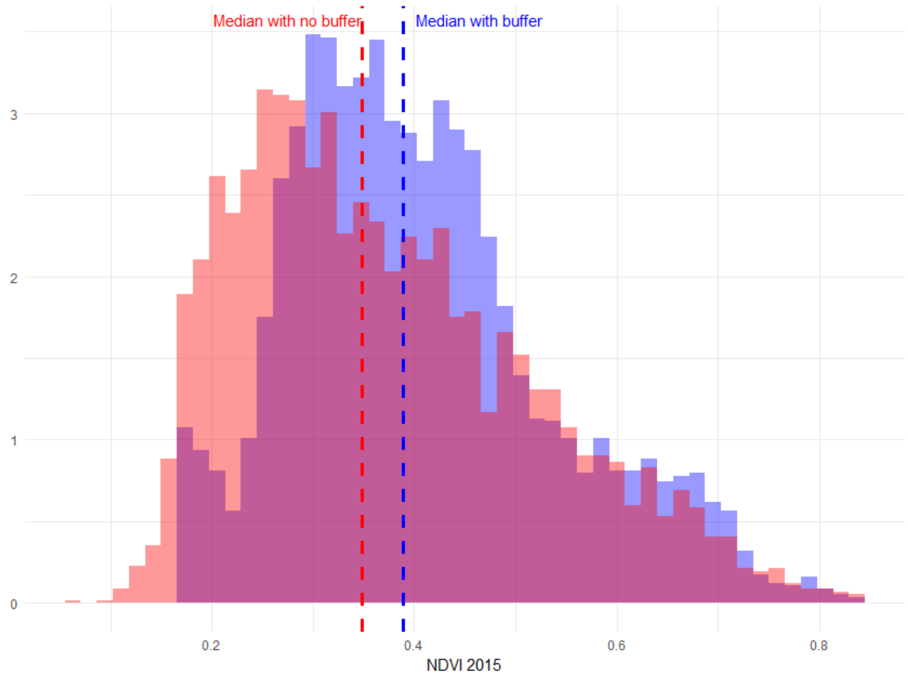

Fig. S3 Distribution of the average NDVI measured in the census tracts belonging to the study area, with and without considering the buffer of 300 m (blue: with buffer; red: without buffer). Dashed vertical lines represent the medians. Year 2015

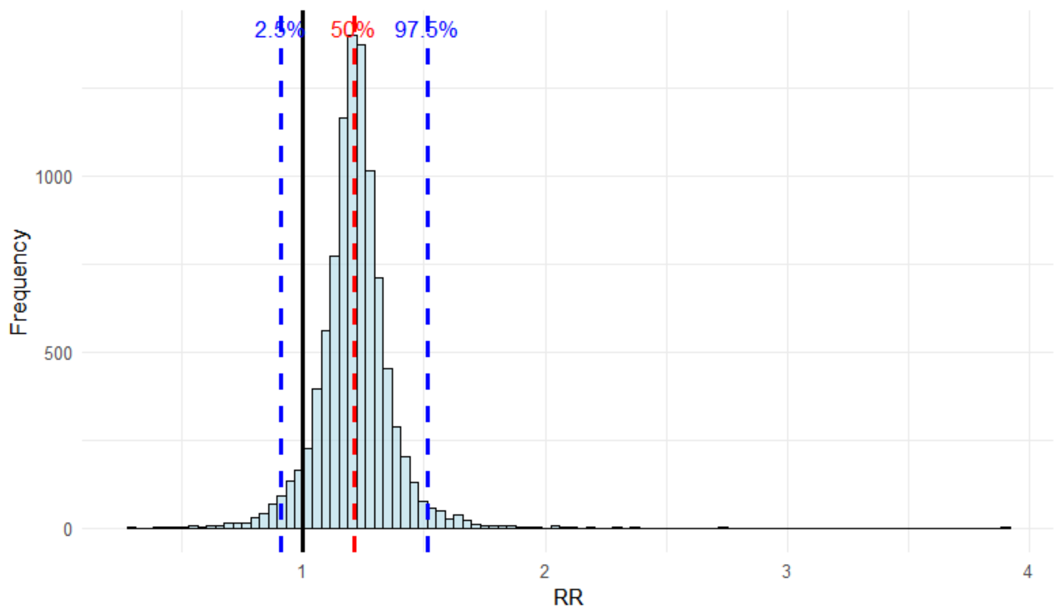

Fig. S4 Posterior predictive distribution of the all-cause mortality RR comparing the most deprived (fifth) with the least deprived (first) quintile of deprivation. Dashed red line: median. Dashed blue lines: 90% Credibility Interval. Black line: null hypothesis, RR=1

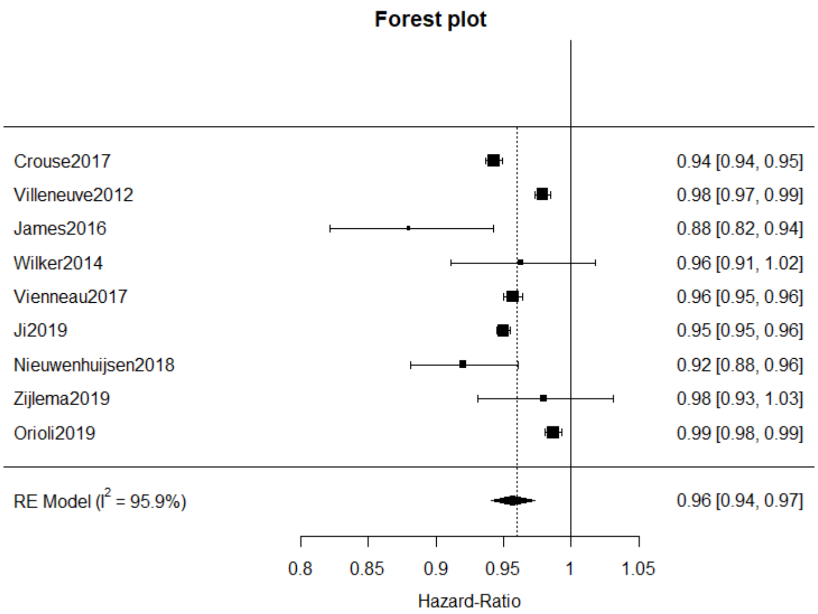

Fig. S5 Forest plot of the meta-analysis by Rojas-Rueda et al. showing the association between residential greenness (NDVI) and all-cause mortality per 0.1 increase in NDVI within a 500 m buffer

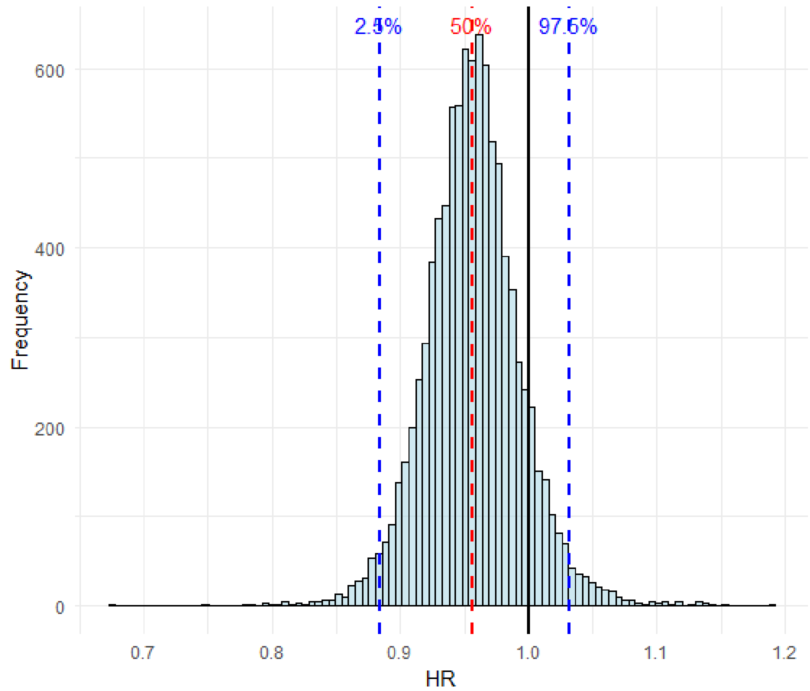

Fig. S6 Posterior predictive distribution of the HR expressing the effect of a 0.1 increase in NDVI on mortality. Dashed red line: median. Dashed blue lines: 90% Credibility Interval. Black line: null hypothesis, HR=1

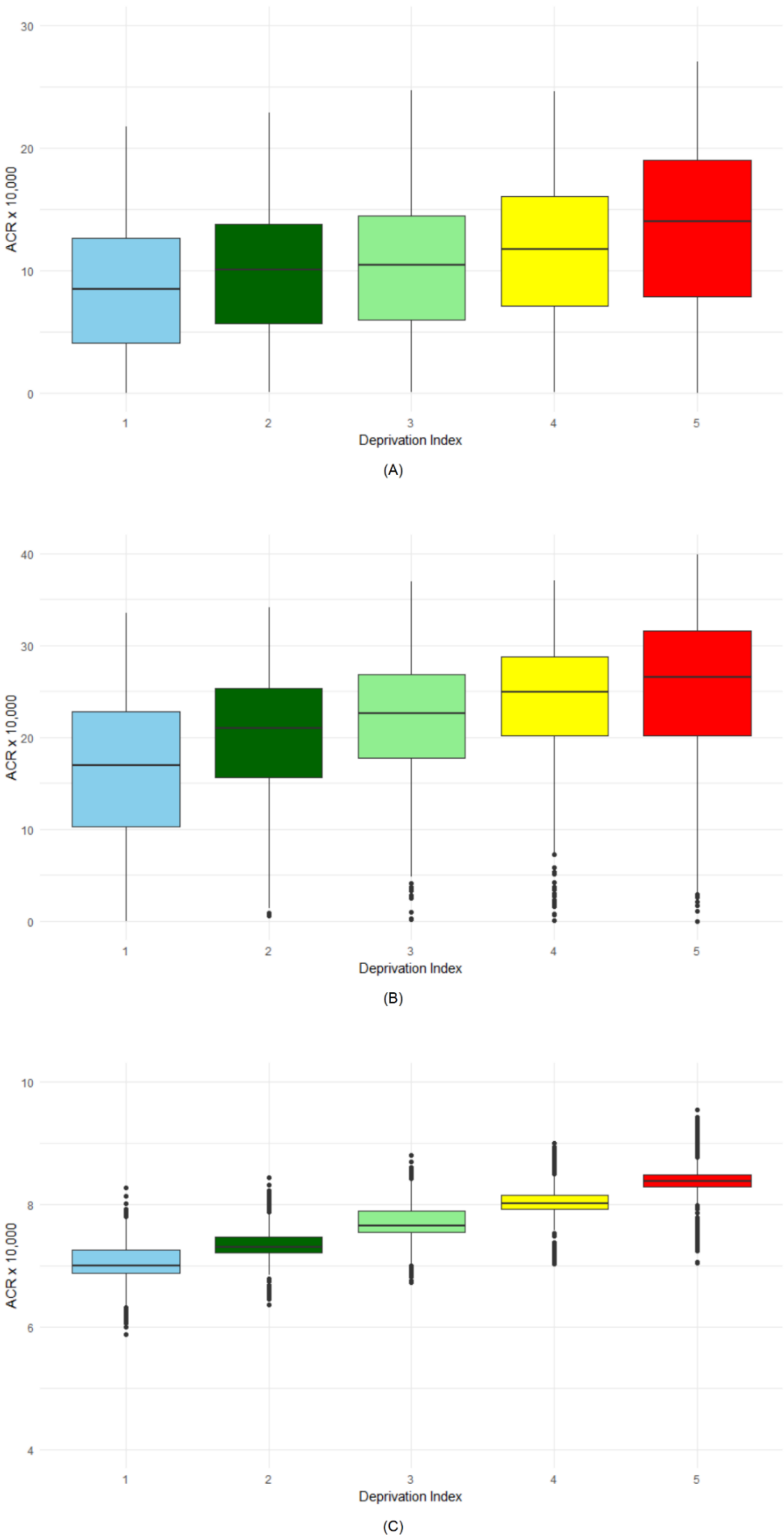

Fig. S7 Boxplots of the Attributable Community Rate × 10,000 residents by quintiles of the Deprivation Index for the entire area, fixing the NDVI counterfactual threshold to 0.5 (A); to 0.7 (B) and assuming a 0.1 NDVI units increase (C)

Table S1 Variance-based sensitivity analysis of attributable deaths under scenario S1. Total variance indices (S_j^{tot}) and Mean Dimension (MD) for each input parameter and for each municipality under the scenario with NDVI threshold fixed to 0.5

	Total Variance index			
Municipality	**X1 (HR)**	**X2 (RR)**	**X3 (q)**	***MD***
Bagno a Ripoli	0.8652	0.0016	0.2037	1.0705
Calenzano	0.9638	0.0004	0.0256	≈1
Campi Bisenzio	0.9815	0.0002	0.0088	≈1
Firenze	0.9882	<0.0001	0.0002	≈1
Lastra a Signa	0.9576	0.0034	0.0384	≈1
Scandicci	0.9894	<0.0001	0.0010	≈1
Sesto Fiorentino	0.9876	<0.0001	0.0009	≈1
Signa	0.9744	0.0005	0.0254	1.0002

Table S2 Variance-based sensitivity analysis of attributable deaths under scenario S2. Total variance indices (S_j^{tot}) and Mean Dimension (MD) for each input parameter and for each municipality under the scenario with NDVI threshold fixed to 0.7

Municipality	**X1 (HR)**	**X2 (RR)**	**X3 (q)**	***MD***
Bagno a Ripoli	0.9889	<0.0001	0.0015	≈1
Calenzano	0.9836	<0.0001	0.0024	≈1
Campi Bisenzio	0.9875	<0.0001	0.0005	≈1
Firenze	0.9888	<0.0001	<0.0001	≈1
Lastra a Signa	0.9745	0.0009	0.0129	≈1
Scandicci	0.9891	<0.0001	0.0002	≈1
Sesto Fiorentino	0.9883	<0.0001	0.0002	≈1
Signa	0.9882	<0.0001	0.0006	≈1

Table S3 Variance-based sensitivity analysis of attributable deaths under scenario S3. Total variance indices (S_j^{tot}) and Mean Dimension (MD) for each input parameter and for each municipality under the scenario with an increase of NDVI of 0.1

Municipality	**X1 (HR)**	**X2 (RR)**	**X3 (q)**	$\boldsymbol{MD}$
Bagno a Ripoli	0.9882	<0.0001	<0.0001	≈1
Calenzano	0.9882	<0.0001	<0.0001	≈1
Campi Bisenzio	0.9882	<0.0001	<0.0001	≈1
Firenze	0.9882	<0.0001	<0.0001	≈1
Lastra a Signa	0.9882	<0.0001	<0.0001	≈1
Scandicci	0.9882	<0.0001	<0.0001	≈1
Sesto Fiorentino	0.9882	<0.0001	<0.0001	≈1
Signa	0.9882	<0.0001	<0.0001	≈1

Table S4 Variance-based sensitivity analysis of attributable deaths under scenario S4. Total variance indices (S_j^{tot}) and Mean Dimension (MD) for each input parameter and for each municipality under the scenario with an increase of NDVI of 20%

Municipality	**X1 (HR)**	**X2 (RR)**	**X3 (q)**	$\boldsymbol{MD}$
Bagno a Ripoli	0.9877	<0.0001	<0.0001	≈1
Calenzano	0.9880	<0.0001	0.0006	≈1
Campi Bisenzio	0.9881	<0.0001	0.0002	≈1
Firenze	0.9885	<0.0001	<0.0001	≈1
Lastra a Signa	0.9892	<0.0001	0.0006	≈1
Scandicci	0.9882	<0.0001	0.0002	≈1
Sesto Fiorentino	0.9873	<0.0001	0.0003	≈1
Signa	0.9880	<0.0001	0.0001	≈1

Table S5 Variance-based sensitivity analysis of attributable deaths under scenario S5. Total variance indices (S_j^{tot}) and Mean Dimension (MD) for each input parameter and for each municipality under the scenario with NDVI threshold fixed to WHO target

Municipality	**X1 (HR)**	**X2 (RR)**	**X3 (q)**	**X4 (WHO target)**	***MD***
Bagno a Ripoli	NA	NA	NA	NA	NA
Calenzano	0.4815	0.0034	0.1308	0.6686	1.2842
Campi Bisenzio	0.3871	0.0004	0.0147	0.9029	1.3051
Firenze	0.7483	<0.0001	0.0004	0.3567	1.1054
Lastra a Signa	0.3908	0.0014	0.0343	0.8716	1.2980
Scandicci	0.4264	<0.0001	0.0024	0.8206	1.2494
Sesto Fiorentino	0.5926	<0.0001	0.0022	0.5759	1.1707
Signa	0.3298	0.0012	0.0122	1.0803	1.4236

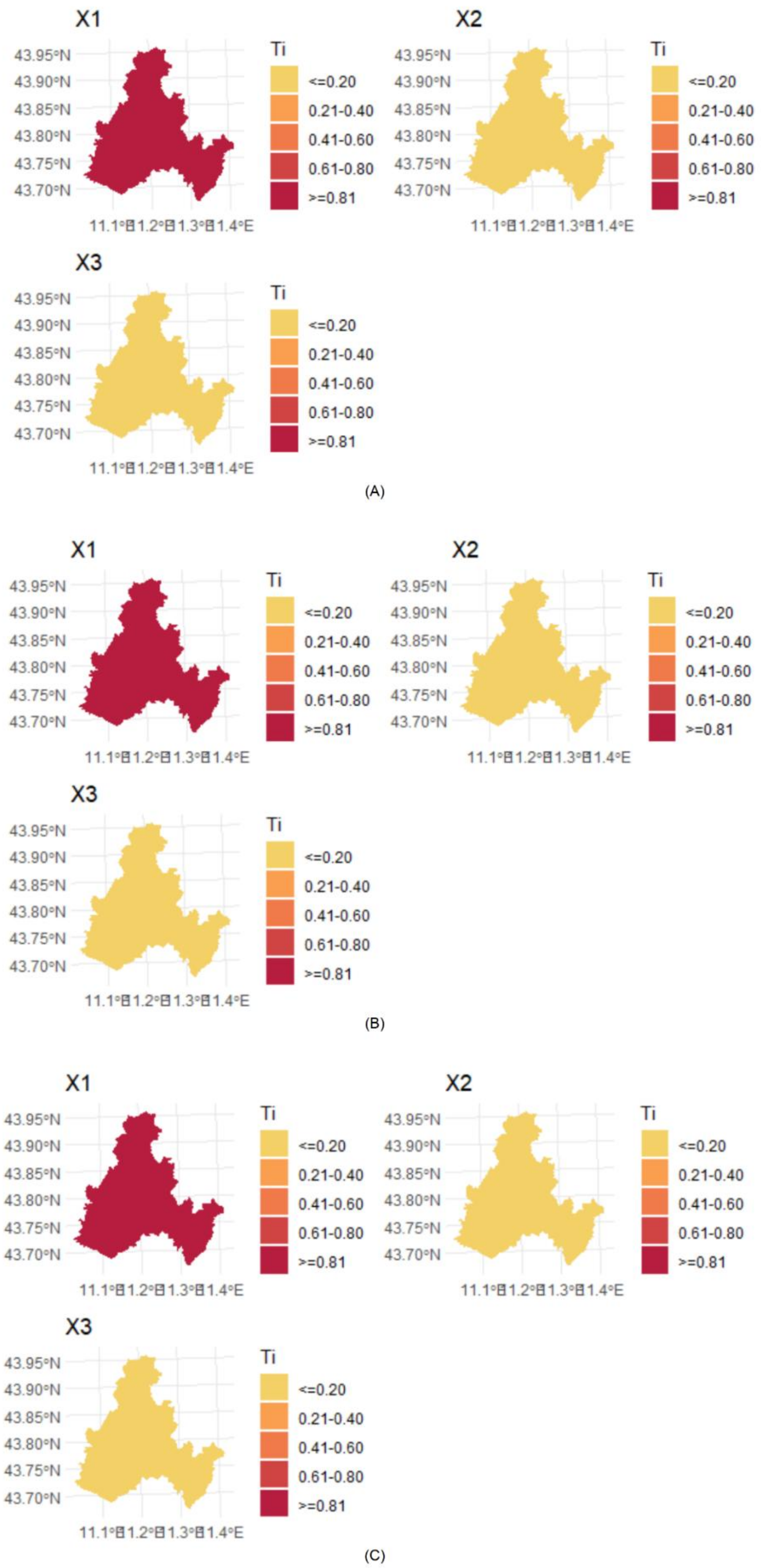

Fig. S8 Total variance indices for each input parameter across municipalities in the study area, fixing the NDVI counterfactual threshold to 0.7 (A); assuming a 0.1 NDVI units increase (B); assuming a 20% NDVI increase (C)