Spatial Patterns of Urban Air Pollution Exposure in Cultural Heritage Areas Derived from Downscaled Satellite Observations

Maria Danese[1], Felice Perciante[2(✉)], and Rosa Lasaponara[2]

[1] ISPC-CNR, Cont. S. Loya, 85050 Tito Scalo, PZ, Italy
[2] IMAA-CNR; Cont. S. Loya, 85050 Tito Scalo, PZ, Italy
felice.perciante@cnr.it

Abstract. Satellite data on pollutants have too coarse a spatial resolution to be suitable for use in studying the exposure of cultural heritage. The integrated use of environmental and anthropogenic data with pollution data derived from Sentinel-5p using machine learning methods makes it possible to obtain high-resolution data that can be used to map urban pollution, which is useful for assessing the potential exposure of cultural heritage to atmospheric degradation.

Keywords: Sentinel-5p · Pollution · Cultural heritage · machine learning · spatial analysis · Rome

1 Introduction

Among the effects of the environmental emergency that characterises our times, one in particular is the cause of the acceleration of the material degradation of cultural heritage (CH): the increase in atmospheric pollution and in particular of gaseous compounds such as carbon monoxide (CO), sulphur dioxide (SO_2) and nitrogen dioxide (NO_2) [1]. This occurs mainly in urban contexts characterised by high anthropic pressure, resulting from high building density, high traffic volumes and the presence of large impermeable surfaces at the expense of green spaces [2, 3].

In general, these pollutants, even in combination with other environmental variables such as humidity, temperature and wind, lead to an increase in the phenomena of alteration [4] of stone materials and decorated surfaces, as well as acidification and corrosion of metals [5]. In the long term, this can compromise the physical integrity of cultural heritage.

More specifically, each of the aforementioned compounds acts differently on materials. For example, SO_2, when reacting with atmospheric humidity, generates sulphuric acid, which is particularly harmful to carbonate materials (marble, limestone). Within these materials, it causes an increase in porosity or even surface defects, which overall leads to an increase in the fragility of the material itself [6]. NO_2 acts in a similar way, causing acid rain when present in excess, as with SO_2, and the deposition of reactive compounds, as it acts as a catalyst for the oxidation of SO_2 and accelerates the formation

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2027
O. Gervasi et al. (Eds.): ICCSA 2026 Workshops, LNCS 16763, pp. 385–400, 2027.
https://doi.org/10.1007/978-3-032-30530-5_24

of crusts [7]. These react both on the surface materials of heritage sites, especially on natural stone, plaster and decorated surfaces [8]. CO is a less reactive pollutant, although, in the presence of other pollutants and other environmental parameters, including, once again, humidity, it contributes to an increase in degradation phenomena [9].

The distribution of CO, SO_2 and NO_2 values is strongly dependent on the spatial distribution of environmental parameters and anthropogenic variables, which can change even over relatively short distances. Therefore, effective mapping of these compounds on a scale appropriate to cultural heritage is very important for the assessment and prevention of cultural heritage degradation processes.

One of the primary sources useful for mapping compounds present in the air are ground-based monitoring networks such as the ARPA network in Italy [10]. However, these have the limitation of being point data, distributed in a limited and uneven spatial manner, despite attempts to make the data continuous through spatial interpolation methods [11, 12].

Another useful method for studying the distribution of atmospheric pollutants is the use of IoT networks, with distributed sensors [13]. However, these still have the limitation of being rather expensive in terms of infrastructure and management costs.

This is where Earth observation data from satellite missions come in, opening up new perspectives for atmospheric monitoring. More specifically, since 2018, as part of the Sentinel-5P TROPOMI mission of the Copernicus programme led by the European Commission in collaboration with the European Space Agency (ESA) [14], has been collecting data useful for assessing air quality. Although the data obtained from this mission has the advantage of offering global coverage and temporal continuity (with daily resolution) that other instruments do not have, it nevertheless has the limitation of having a spatial resolution, on the order of kilometres, which is not sufficient for applications on the scale of cultural heritage or, more generally, even on an urban scale. Furthermore, the data sometimes presents spatial gaps due to the presence of clouds or reflections from certain surfaces.

This work therefore proposes a method based on the use of machine learning techniques and the integration of heterogeneous and multi-source data for the reconstruction of data relating to the aforementioned pollutants, capable of reconstructing it in missing areas and increasing its spatial resolution thanks to its dependence on environmental, morphological and anthropogenic variables. In addition, an index will be introduced to assess the exposure of cultural heritage by evaluating its hot spots, thus creating a decision-making tool for monitoring and strategic thinking, as well as heritage protection. The case study is a complex urban context rich in cultural heritage, the Municipality of Rome, characterised by environmental conditions that are potentially critical for it.

2 Case Study

The area selected as the case study is the municipal territory of the city of Rome (Italy), which covers an area of approximately 1,285 km^2 and hosts over 2.8 million inhabitants. It represents an extremely complex and heterogeneous urban environment, characterized by the coexistence of a high density of cultural heritage assets—including globally significant sites such as the Colosseum, the Imperial Fora, and numerous basilicas—alongside a substantial building density and vehicular traffic. These elements alternate

with green spaces and industrial or commercial areas (Fig. 1). Within this context, urban morphology plays a key role in influencing the distribution of pollutants, thereby justifying the selection of the city as a case study.

In addition to urban morphology, the distribution of environmental parameters—and consequently of pollutants—is closely related to the city's geographical configuration, as it lies within the Tiber River basin and is surrounded by volcanic structures and a system of hilly reliefs.

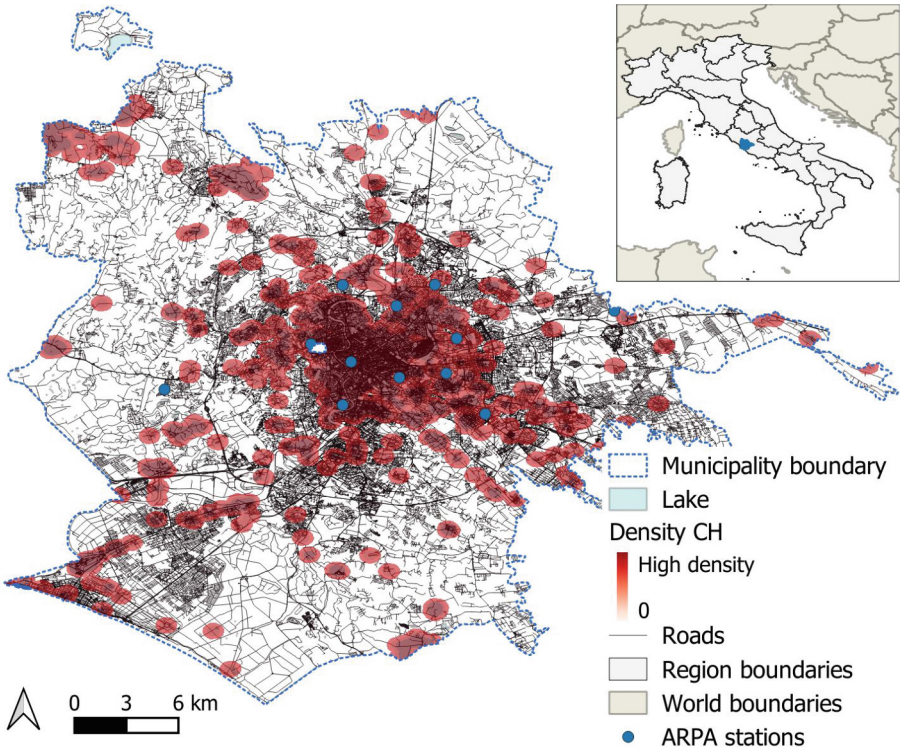

Fig. 1. The case study: the Rome Municipality with the Kernel Density Estimation of its cultural heritage. The first-order expected Nearest Neighbor was used as bandwidth in the estimation.

3 Materials and Methods

3.1 Input Datasets

In this study, the three S-5P datasets relating to the pollutants CO, NO_2 and SO_2 have a spatial resolution of 3.5 × 7 km and a daily temporal resolution, with a pass between 11:00 and 12:00 UTC (Coordinated Universal Time), corresponding to 13:00–14:00 local Italian time over the Rome area. These data consists of the instantaneous measurement of the vertical column density of the pollutant in the atmosphere, expressed in mol/m^2 [14].

As for the explanatory variables, both static and dynamic variables were included in the model. The static variables comprise groups of static morphological and anthropogenic parameters; the dynamic variables comprise environmental parameters.

With regard to the morphological parameters, the TINITALY DEM [15], was used, with an initial resolution of 10 m, which was downscaled to a resolution of 500 m using a mean focal function.

With regard to the anthropogenic data, the following variables were used:

- population and built-up density for 2025, taken from the GHSL (Global Human Settlement Layer) dataset of the Copernicus Emergency Management Service, developed by the European Commission's Joint Research Centre, with a spatial resolution of 100 m, to map the population and human settlements.
- road and industrial density in 2025. Roads and industrial sites were derived from OpenStreetMap (OSM) [16] via the website https://overpass-turbo.eu/ (tags: roads, industry), and the densities were subsequently calculated as road length in km per pixel and as the number of industrial buildings per pixel, respectively.

For what concerns the environmental parameters, the following have been included in the model:

- climatic data derived from ERA5-Land Daily, developed by the European Centre for Medium-Range Weather Forecasts (ECMWF) as part of the Copernicus Climate Change Service (C3S) programme [17]. More specifically, the following variables were considered: average, minimum and maximum air temperature at 2 m, surface pressure (measured in Pa), relative humidity (%), wind speed at 10 m (in m/s), and average solar radiation at ground level (W/m^2). These data have a spatial resolution of 9 km and a daily update frequency.
- daily cumulative precipitation derived from CHIRPS (Climate Hazards Group InfraRed Precipitation with Station data) [18], with a spatial resolution of 5 km.
- Dynamic World V1 land cover, derived from Sentinel products by the World Resources Institute (WRI) [19], with a spatial resolution of 10 m and daily temporal resolution, containing the following land cover classes: water, trees, grass, flooded vegetation, crops, shrubs and scrub, built-up areas, bare ground, snow and ice.
- NDVI index, calculated as the normalised combination of the red (RED) and near-infrared (NIR) bands from the Sentinel-2 (S2) dataset, it belongs to the mission of the European Copernicus programme [20]. It ranges from 1 to -1, and its values indicate changes in the health and density of vegetation. It has a temporal resolution of approximately 5 days and a spatial resolution of 10 m.

The model utilised environmental parameters from the day prior to the passage of S-5P, as the formation and changes in pollutant data are not instantaneous but depend on the weather conditions (temperature, rainfall, wind) of the preceding hours.

Finally, the data on Rome's CH were also derived from OSM.

3.2 Processing Environment and Modelling Approach

Part of the data processing and modelling workflow mentioned in the previous section was carried out within Google Earth Engine (GEE). This is a cloud-based platform for

processing and analyzing large-scale geospatial data, which has the following advantages: high-speed parallel processing without the need for large computing resources on one's own computer, easy access to data and methods for different types of users [21, 22]. It is therefore particularly suitable for processing satellite images and analyzing climate datasets [23, 24]. In fact, it provides free access to a vast dataset [25], including all the data mentioned in the previous section.

Within GEE there are several algorithms that can be used to analyze spatial and satellite data, and among these are some machine learning methods such as supervised methods like the support vector machine or the random forest (RF) [26, 27], used in this work for the prediction of missing values (nodata) in air pollutant data derived from the Sentinel-5P satellite.

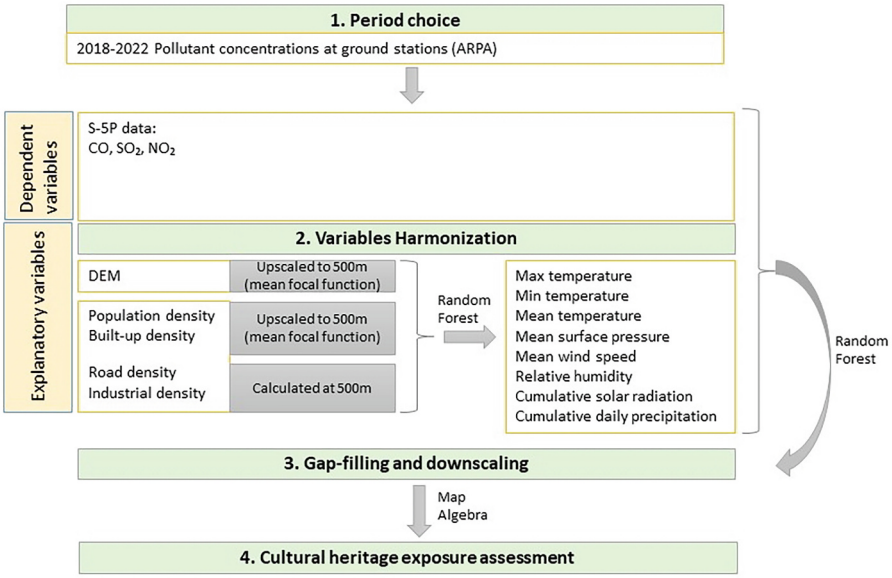

Fig. 2. Flow chart resuming the process.

The RF method is an ensemble-type supervised learning algorithm, based on the optimization of a set of decision trees following recursive execution of the algorithm several times; this improves the predictive ability of the model by reducing data overfitting [28, 29]. A training set must be chosen within the population dataset, then the decision tree classification is performed. Finally, as anticipated, the process is repeated to improve the learning process and return the best classification.

Instead, the calculation of statistics on gap-filled and downscaled data and the estimation of exposure were carried out using map algebra methods [30], a set of mathematical and logical operations applied to raster data, capable of performing calculations on a cell-by-cell basis (local functions), within predefined zones (zonal functions) or within the neighborhoods of each pixel (focal functions).

Figure 2 shows a flowchart summarising the data and the method followed, the steps of which will be explained in the following section on the results.

4 Results

4.1 Period Choice

For the aim of the analysis presented in this study, the first step involved examining the temporal trends in pollutant concentrations, derived from data collected during the ARPA Lazio air quality monitoring campaigns between 2018 and 2025. In the Rome area, specifically, there are 11 stations, whose pollutant data are recorded hourly, daily and monthly. Figure 3 illustrates the trends for each station measuring the concentrations of each pollutant under consideration; the figure reveals the presence of seasonal patterns. More specifically, the maximum peaks for NO_2 and CO occur during the winter months. This may be linked to the combined conditions arising from the increase in emissions during this period due to heating and road traffic, as well as meteorological conditions that favour the stagnation of pollutants in the lower atmospheric layers. In contrast, higher concentrations of SO_2 are observed in summer.

Consequently, this study focused on the month of January for NO_2 and CO data, whilst July was selected for SO_2 data, with the aim of analyzing conditions that are more representative of high environmental pressure, in order to identify the most significant spatial patterns of the pollutants and, consequently, the areas where the heritage is actually exposed.

4.2 Variable Harmonization

Once the study period has been selected, it is necessary to spatially harmonize the explanatory variables in order to ensure the spatial consistency of the predictor stack. The aim is to standardize all predictive variables to a spatial resolution of 500m, a value suitable for urban-scale analyses of CH, in line with the objectives of this study.

Initially, all static information layers with high spatial resolution, along with the NDVI – namely the DEM, the NDVI, population density and built-up density – were upscaled using a mean focal function from map algebra. The exception is the land cover derived from Dynamic World, as it is a categorical variable, for which a majority focal function was used.

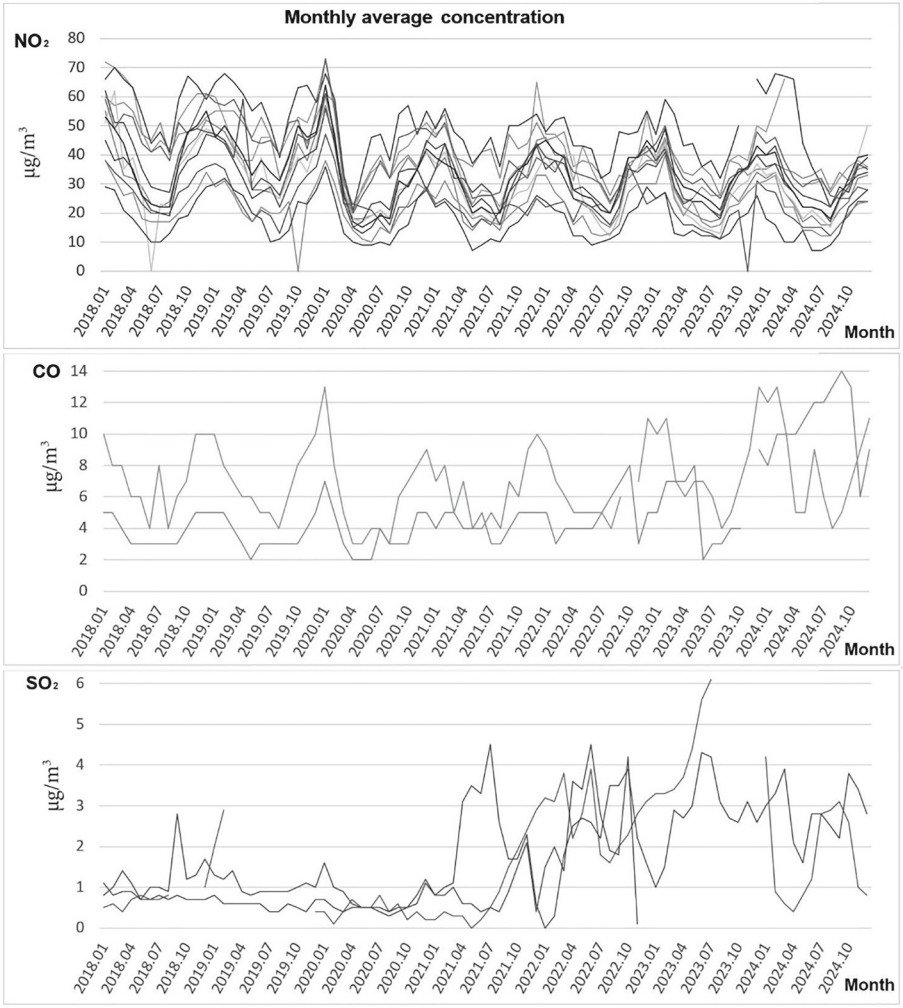

Fig. 3. Monthly pollution trends. Each line represents a meteorological station. Overall, the seasonal trends are evident.

These elements constitute the morphological, environmental and anthropogenic bases that characterize the territory and influence meteorological conditions; therefore, they were used as input datasets for the downscaling of environmental parameters derived from ERA5 and CHIRPS data. The method used was RF, with 100 trees. The result is a dataset of variables with a resolution of 500 m which nevertheless retains the information content of the input parameters, whilst improving the representation of local gradients (Fig. 4).

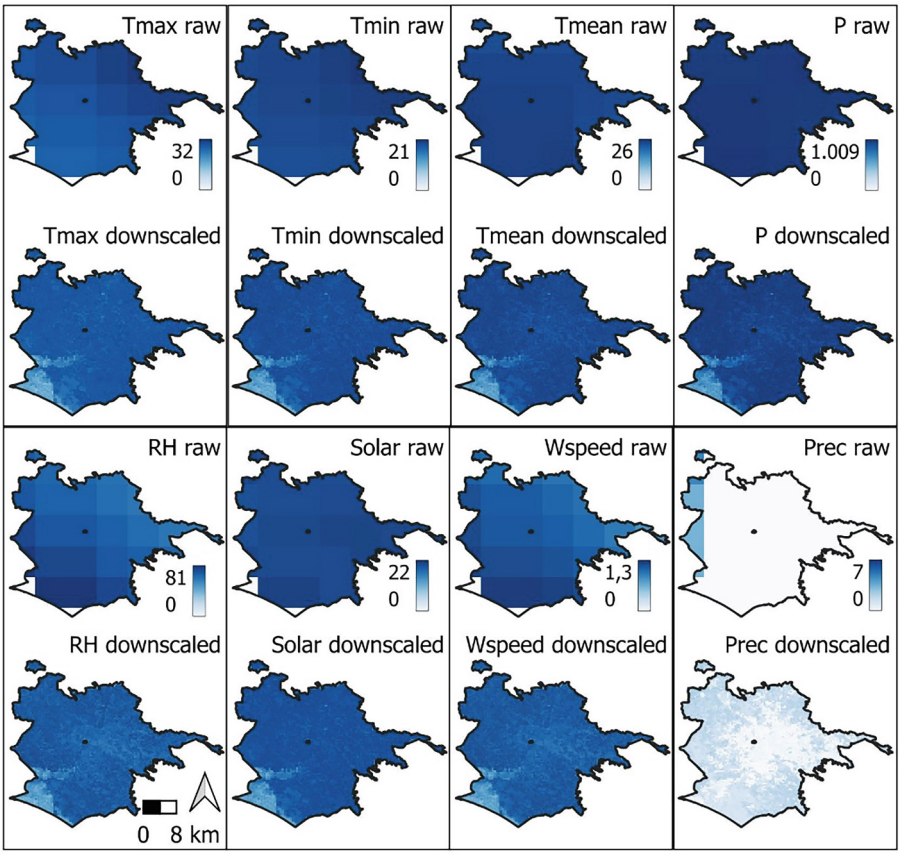

Fig. 4. Example of environmental parameters raw and downscaled.

4.3 Pollution Gap-Filling and Downscaling

Once the dataset had been prepared with the dependent and independent variables, the missing values in the atmospheric column for the three selected pollutants were imputed using the Random Forest algorithm within GEE, thereby producing a continuous map. Furthermore, this process improved the spatial resolution of the pollution data to 500 m (Fig. 5).

The results were validated by calculating the root mean square error (RMSE) and Pearson's correlation coefficient (r). The low values of the former and high values of the latter (Table 1) indicate that the model has good predictive capability, ensuring that the downscaled data remain consistent with the original satellite data. This confirms the robustness of the method.

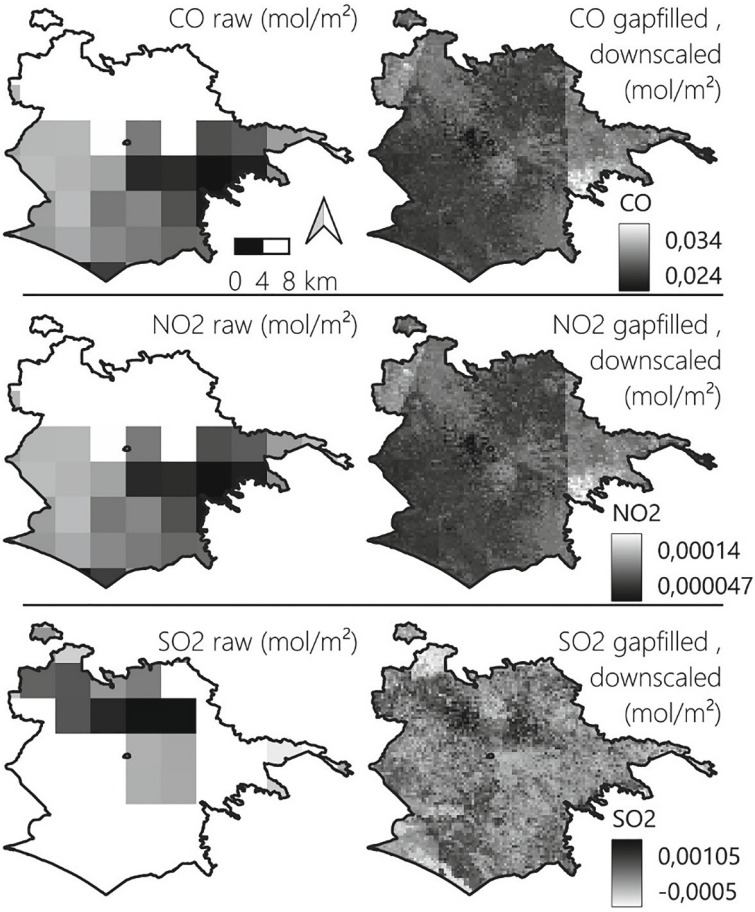

Fig. 5. Pollutants raw and gap filled-downscaled rasters.

Table 1. Range found for RMSE and Pearson's correlation (r).

Pollutant	RMSE range (mol/m^2)	r range
CO	0.0001–0.0014	0.86–0.96
NO_2	$10^{-6} - 10^{-5}$	0.85–0.95
SO_2	0.0001– 0.00045	0.90–0.94

4.4 Cultural Heritage Exposure Assessment

Once the continuous, high-resolution rasters of the pollutants had been obtained, a raster of average values was calculated for each pollutant, using the local statistics of map

algebra, for each pixel within the month in question. The same was done with the maximum values. The first result provides information on the background concentration of the pollutant, as it describes an average monthly level of pollution. It can therefore be associated with slow but cumulative and long-term degradation processes affecting CH. The second result, on the other hand, provides information on peaks occurring during the period and thus on locations of episodic but intense exposure. It therefore implies phenomena of accelerated degradation processes, indicating areas with a higher risk, albeit for shorter and rarer periods. The combined use of the two indicators thus helps to provide a more complete picture of the heritage's exposure.

Subsequently, for the purpose of calculating a CH exposure index, a Euclidean distance was calculated from the CH sites, which was then normalised.

The following Heritage Exposure Index (HEI) was calculated from the combination of the normalised pollutants and the normalised distance from the heritage sites, Eq. 1:

$$HEI = pollutant\ concentration\ x\ (1 - proximity\ to\ heritage\ sites) \quad (1)$$

The result of this equation returns: (1) a value approaching 1 (worst-case scenario), if there is a high concentration in the vicinity of assets; (2) a value approaching 0 (best-case scenario), if there is a low concentration at a distance from assets; (3) intermediate values that occur in the case of low concentration near assets or in the case of high concentration at a distance.

The HEI values were classified into five categories using the quantile method (Figs. 6, 7 and 8), in order to highlight the relative distribution of exposure within the study area and identify the areas characterised by higher levels of severity.

Analysis of the findings from the HEI index shows that the pollutants analysed exhibit different spatial patterns.

As regards CO, the concentrations in the 'Very high' and 'High' classes of average values are found in the central area of Rome, which accounts for 60% and 33% of the CH respectively. This is undoubtedly due to the widespread presence of these pollutants in the city's most densely populated urban areas. Incidents of maximum CO levels are also found predominantly in areas classified as 'High'.

NO_2 levels show a similar pattern, with different percentage values but a comparable trend, again linked to road traffic and urban density.

SO_2 behaves differently, however, both in terms of average values – which are concentrated mainly in the Medium category – and maximum values, which are found predominantly in the Low and Very Low categories. This is likely due to the spatial dispersion of emission sources and an overall lower concentration of the pollutant (Table 2).

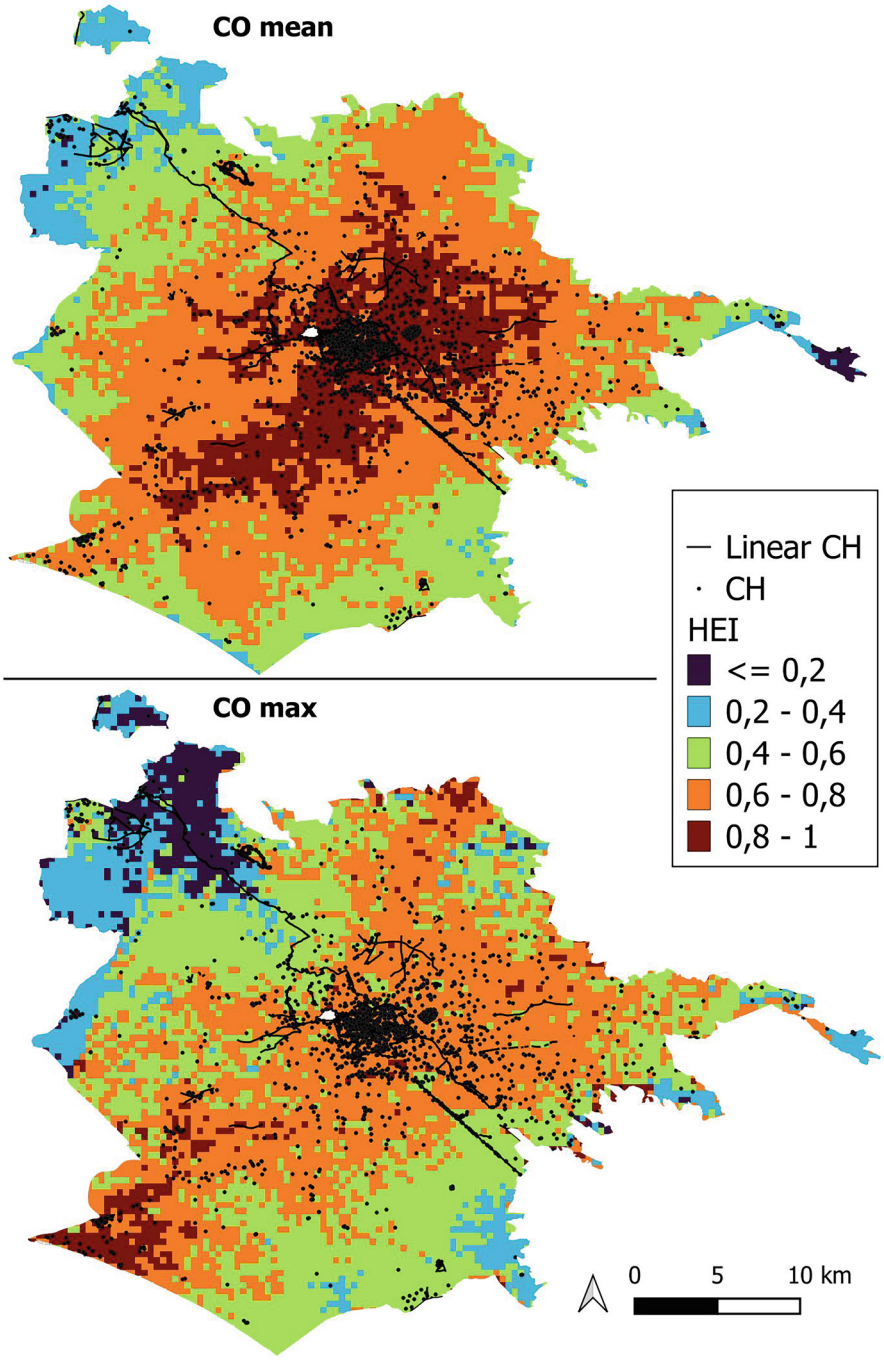

Fig. 6. HEI index calculated for CO mean and maximum.

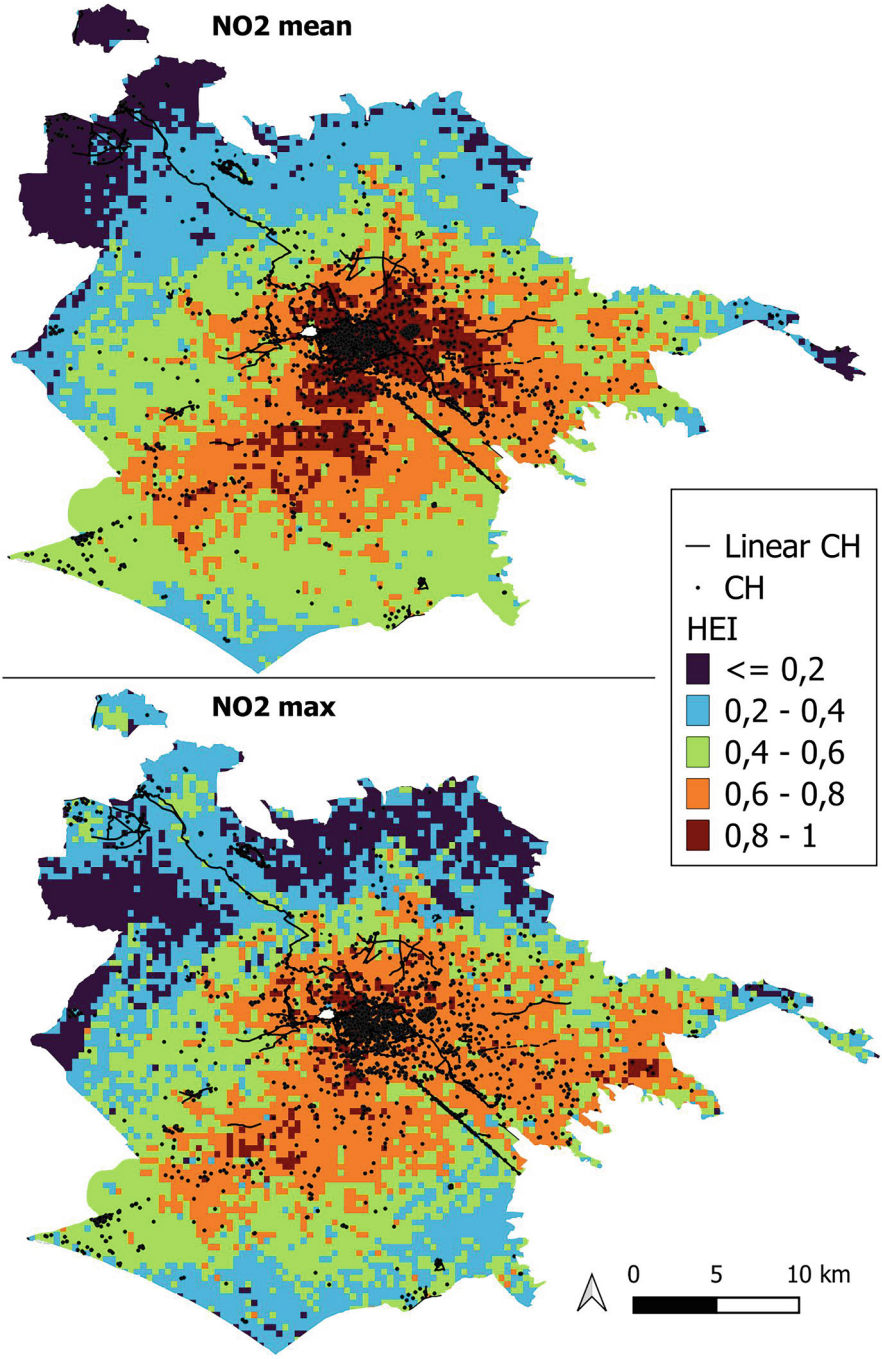

Fig. 7. HEI index calculated for NO_2 mean and maximum.

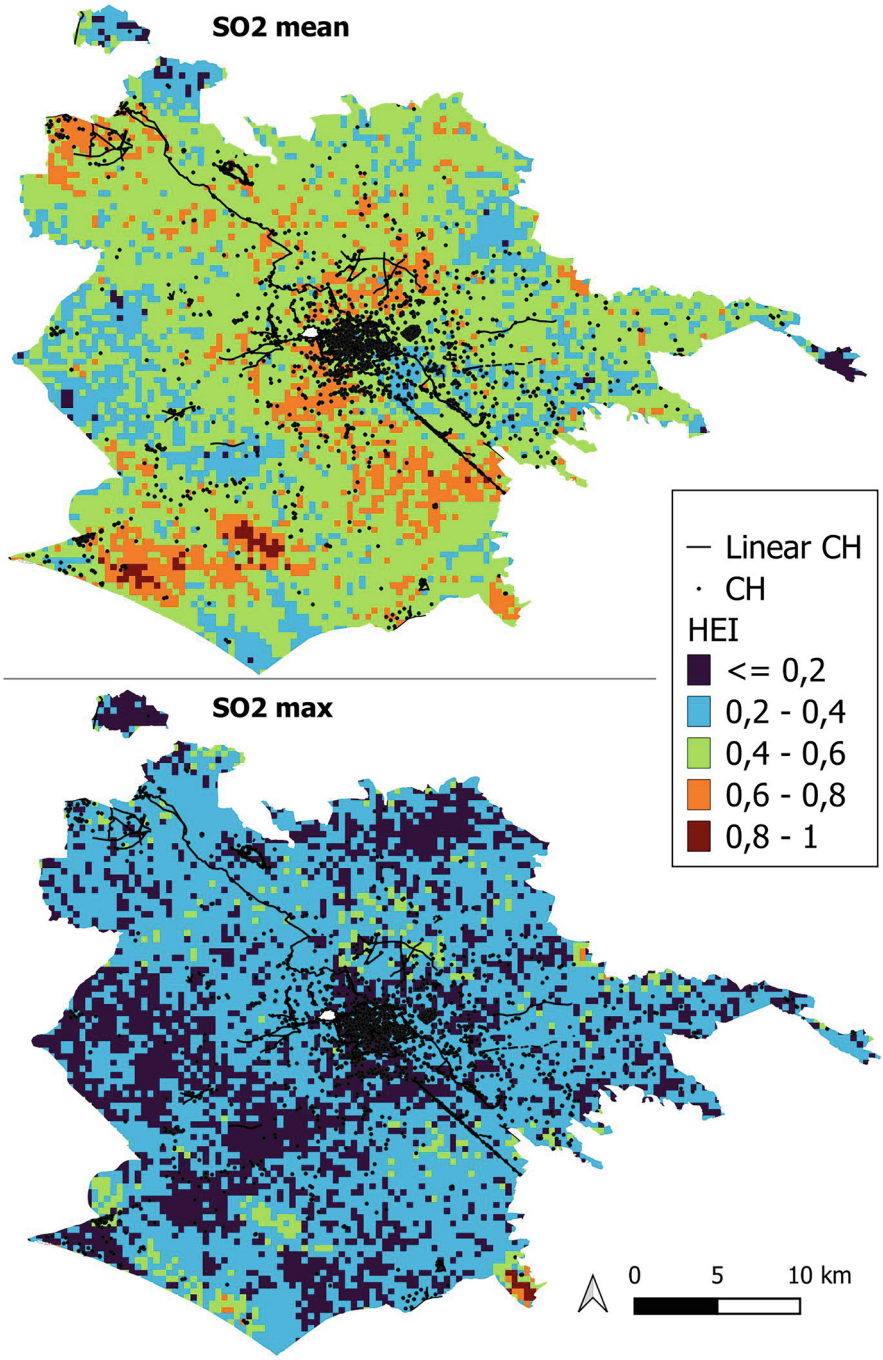

Fig. 8. HEI index calculated for SO_2 mean and maximum.

Table 2. CH percentage (%) following in HEI classes for each pollutant.

HEI	CO mean	CO max	NO_2 mean	NO_2 max	SO_2 mean	SO_2 max
Very low	0.03	0.46	0.68	2.2	0.11	40.4
Low	0.36	2.66	4.8	7.37	12.93	54.44
Medium	5.79	19.47	16.09	18.05	72.19	3.40
High	33.42	74.59	29.30	61.99	14.69	0.72
Very high	60.4	2.81	49.14	10.39	0.07	1.03

5 Discussion and Conclusions

This study presents an approach aimed at assessing the exposure of the CH to some of the main atmospheric pollutants, based on the integration of remote sensing, machine learning and spatial analysis. Overall, the approach adopted has demonstrated how the integration of technologies and methods can help to improve the available data and carry out a more effective analysis at an urban scale.

The introduction of the HEI index has made it possible to combine information on pollutants with the location of properties, providing a representation and quantification of exposure levels. This analysis revealed that, in the context under investigation, the main exposure factors are attributable to CO and NO_2, whilst the contribution of SO_2 is more limited. The indicator provides a useful decision-support tool for programming and planning, helping to identify priority areas for monitoring and preventive conservation.

The method could be further developed by incorporating the contribution of other pollutants, extending the analysis to longer time series, and including the specific vulnerability of different materials in the analysis.

However, the proposed method constitutes an easily replicable and scalable protocol to support the protection of cultural heritage in a context of increasing environmental pressure.

Acknowledgments. IR0000032 – ITINERIS, Italian Integrated Environmental Research Infrastructures System (D.D. n. 130/2022 - CUP B53C22002150006) Funded by EU - Next Generation EU PNRR- Mission 4 “Education and Research” - Component 2: “From research to business” - Investment 3.1: “Fund for the realization of an integrated system of research and innovation infrastructures”. Project “Study of hazard elements and their variation as a function of climate change, on CH using spatial ML methods” inside “CH and creativity in green and digital transitions for inclusive societies” (FOE2021 - Ordinary fund for research organizations and institutions).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Tidblad, J., et al.: Effects of air pollution on materials and cultural heritage: ICP materials celebrates 25 years of research. Int. J. Corrosion **2012**, 496321 (2012)

2. Vardoulakis, S., Fisher, B.E., Pericleous, K., Gonzalez-Flesca, N.: Modelling air quality in street canyons: a review. Atmos. Environ. **37**, 155–182 (2003)
3. Nowak, D.J., Hirabayashi, S., Bodine, A., Greenfield, E.: Tree and forest effects on air quality and human health in the United States. Environ. Pollut. **193**, 119–129 (2014)
4. Esteban-Cantillo, O.J., Menendez, B., Quesada, B.: Climate change and air pollution impacts on cultural heritage building materials in Europe and Mexico. Sci. Total. Environ. **921**, 170945 (2024)
5. Brimblecombe, P., Grossi, C.M.: Damage to buildings from future climate and pollution. APT Bull. J. Preserv. Technol. **38**, 13–18 (2007)
6. Charola, A.E.: Acid rain effects on stone monuments. J. Chem. Educ. **64**, 436 (1987)
7. Bonazza, A., Messina, P., Sabbioni, C., Grossi, C.M., Brimblecombe, P.: Mapping the impact of climate change on surface recession of carbonate buildings in Europe. Sci. Total. Environ. **407**, 2039–2050 (2009)
8. Ausset, P., et al.: Experimental study of limestone and sandstone sulphation in polluted realistic conditions: the Lausanne Atmospheric Simulation Chamber (LASC). Atmos. Environ. **30**, 3197–3207 (1996)
9. Graedel, T.E., Hawkins, D.T., Claxton, L.D.: Atmospheric Chemical Compounds: Sources, Occurrence and Bioassay. Elsevier (2012)
10. Zavatti, A., Lauriola, P.: The contribution of environmental monitoring in the epidemiological assessment of exogenous risk: the experience of ARPA in the Emilia-Romagna Region of Italy. Ann. N. Y. Acad. Sci. **895**, 173–187 (1999)
11. Bechle, M.J., Millet, D.B., Marshall, J.D.: Remote sensing of exposure to NO2: satellite versus ground-based measurement in a large urban area. Atmos. Environ. **69**, 345–353 (2013)
12. Jerrett, M., et al.: A review and evaluation of intraurban air pollution exposure models. J. Eposure Sci. Environ. Epidemiol. **15**, 185–204 (2005)
13. Channi, H.K., Kumar, R.: The role of smart sensors in smart city. In: Singh, U., Abraham, A., Kaklauskas, A., Hong, T.P. (eds.) Smart Sensor Networks. Studies in Big Data, vol. 92. Springer, Cham (2022). https://doi.org/10.1007/978-3-030-77214-7_2
14. Veefkind, J.P., et al.: TROPOMI on the ESA Sentinel-5 Precursor: a GMES mission for global observations of the atmospheric composition for climate, air quality and ozone layer applications. Remote Sens. Environ. **120**, 70–83 (2012)
15. Tarquini, S., et al.: TINITALY/01: a new triangular irregular network of Italy. Ann. Geophys. **50**, 407–425 (2007)
16. Nelson, A., de Sherbinin, A., Pozzi, F.: Towards development of a high quality public domain global roads database. Data Sci. J. **5**, 223–265 (2006)
17. Hersbach, H., et al.: The ERA5 global reanalysis. Q. J. R. Meteorol. Soc. **146**, 1999–2049 (2020)
18. Funk, C., et al.: The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. Sci. Data **2**, 1–21 (2015)
19. Brown, C.F., et al.: Dynamic World, Near real-time global 10 m land use land cover mapping. Scientific Data **9**, 251 (2022)
20. Drusch, M., et al.: Sentinel-2: ESA's optical high-resolution mission for GMES operational services. Remote Sens. Environ. **120**, 25–36 (2012)
21. Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R.: Google earth engine: planetary-scale geospatial analysis for everyone. Remote Sens. Environ. **202**, 18–27 (2017)
22. Sangeetham, R.: a review on google earth engine: an open access cloud analysis platform for planetary scale satellite data. In: E3S Web of Conferences, p. 09011. EDP Sciences
23. Abijith, D., et al.: Assessing the impact of climate and land use change on flood vulnerability: a machine learning approach in coastal region of Tamil Nadu India. Geosci. Lett. **12**, 1 (2025)

24. Khalil, M., Kumar, J.S.: Predictive modeling of land surface temperature dynamics in Damascus, Syria using google earth engine: a remote sensing and random forest approach. Earth Sci. Inf. **18**, 1–20 (2025)
25. Tamiminia, H., Salehi, B., Mahdianpari, M., Quackenbush, L., Adeli, S., Brisco, B.: Google Earth Engine for geo-big data applications: a meta-analysis and systematic review. ISPRS J. Photogramm. Remote Sens. **164**, 152–170 (2020)
26. Ouchra, H., Belangour, A., Erraissi, A.: Exploring google earth engine platform for satellite image classification using machine learning algorithms. In: Ben Ahmed, M., Boudhir, A.A., El Meouche, R., Karaş, İ.R. (eds.) Innovations in Smart Cities Applications Volume 7. SCA 2023. Lecture Notes in Networks and Systems, vol. 938. Springer, Cham (2024). https://doi.org/10.1007/978-3-031-54376-0_24
27. Zhao, Z., et al.: Comparison of three machine learning algorithms using google earth engine for land use land cover classification. Rangel. Ecol. Manage. **92**, 129–137 (2024)
28. Kleinberg, E.M.: On the algorithmic implementation of stochastic discrimination. IEEE Trans. Pattern Anal. Mach. Intell. **22**, 473–490 (2002)
29. Kleinberg, E.: An overtraining-resistant stochastic modeling method for pattern recognition. Ann. Stat. **24**, 2319–2349 (1996)
30. Tomlin, C.D.: Geographic Information Systems and Cartographic Modeling. Prentice Hall Englewood Cliffs, NJ (1990)