A comprehensive environmental exposure indicator and respiratory health in asthmatic children: a case study

Giovanna Cilluffo[1,2] · Gianluca Sottile[3] · Giuliana Ferrante[4] · Salvatore Fasola[5] · Velia Malizia[5] · Laura Montalbano[5] · Andrea Ranzi[6] · Chiara Badaloni[7] · Giovanni Viegi[5,8] · Stefania La Grutta[5]

Abstract

The primary goal of asthma management is to achieve and maintain asthma control, which can be influenced by environmental factors. This longitudinal study aimed to construct a comprehensive environmental indicator to predict asthma control in children with asthma in Palermo, Italy. The study included 179 asthmatic children aged 5–16 years. The Normalized Difference Vegetation Index (NDVI) was used to measure green cover, and the Coordination of Information on the Environment (CORINE) framework was used to assess land use based on each home address. A land use regression (LUR) model centered on the home address estimated NO_2 exposure for each child using GIS. An environmental indicator, including environmental and personal exposure, was formulated using an additive value model approach. A logistic regression mixed model assessed the association between the environmental indicator and uncontrolled asthma. A probability map of uncontrolled asthma was constructed. In conclusion, a comprehensive environmental indicator proved effective in identifying areas at higher and lower risk of uncontrolled asthma.

Keywords Asthma control · Children · CORINE land cover · Greenness · Land Use Regression · NDVI

1 Introduction

Asthma is a non-communicable disease that affects both children and adults (Dharmage et al. 2019). The Global Initiative for Asthma (GINA) guidelines (Bateman et al. 2008), state that the primary objectives of asthma management are to attain and maintain disease control, as poor control is associated with adverse disease outcomes and reduced lung function (Montalbano et al. 2016). Environmental

---

exposures have been identified as potential factors that can affect asthma control and exacerbations (Ferrante et al. 2020). Obtaining conclusive evidence on the impact of green spaces on asthma in children and adolescents has been challenging to date (Andrusaityte et al. 2016).

The role of green spaces in relation to asthma has been investigated in various studies with mixed results. Several cross-sectional studies have reported positive associations between exposure to green spaces and asthma morbidity across various age groups (Dadvand et al. 2014; Lovasi et al. 2008; Wu et al. 2021). However, some studies have failed to replicate these findings (Feng and Astell-Burt 2017; Lovasi et al. 2013). Longitudinal studies have shown conflicting results regarding the effects of green spaces on asthma, with some studies demonstrating detrimental effects (Aerts et al. 2020; Tischer et al. 2017) and others showing protective effects (Cavaleiro Rufo et al. 2021). A single cross-sectional study investigating the impact of urban green areas on asthma control reported no significant effect (Chen et al. 2017). Very few research has examined the relationship between urban green spaces and outdoor/indoor pollutants and their impact on asthma symptoms and severity. A cross-sectional study conducted in an Italian metropolitan area found associations between multiple exposures to low levels of 'greenery' (measured by the Normalized Difference Vegetation Index (NDVI)), 'grayness' (concrete urban areas), NO_2 exceeding World Health Organization limits, and nasal, ocular, and general symptoms in 244 schoolchildren (Cilluffo et al. 2018). A cross-sectional study conducted in Italy involving 187 schoolchildren found that living in close proximity to green areas was associated with a higher likelihood of asthma compared to living further away from such areas (Squillacioti et al. 2020). In a more recent study, the effects of multiple exposure factors on lung function were investigated in a sample of 2082 children and adolescents (Cilluffo et al. 2022).

The aforementioned studies investigated the association between environmental factors and asthma symptoms or lung function using conventional regression techniques, such as logistic regression models (Andrusaityte et al. 2016; Cavaleiro Rufo et al. 2021; Dadvand et al. 2014; Feng and Astell-Burt 2017; Squillacioti et al. 2020; Tischer et al. 2017), penalized regression models (Cilluffo et al. 2018), quantile regression models (Cilluffo et al. 2022) and generalized mixed effects models with repeated measures (Aerts et al. 2020).

To date, no longitudinal study has thoroughly evaluated the collective impact of various environmental factors on asthma management in children. This study follows a comprehensive approach to enhance the prediction of uncontrolled asthma by integrating diverse environmental factors. To accomplish this, a composite indicator is proposed as a versatile tool to account for multiple environmental factors that may impact asthma control. The benefits of using a composite indicator are clear and can be summarized as: unidimensional measurement of the phenomenon, ease of interpretation compared to a set of many individual indicators, and simplification of data analysis. This approach combines sources such as air quality monitoring, exposure to greenery, smoke, and mold to provide a more nuanced and accurate assessment of the environmental factors that influence asthma.

Various statistical approaches are used in environmental health research to assess the cumulative effects of multiple exposures. These include additive value models (AVM)

(Hanushek and Rivkin 2006; McCaffrey et al. 2004), principal component analysis (PCA) (Sendhil et al. 2018), and weighted quantile sum (WQS) regression (Carrico et al. 2015).

The first method focuses on combining individual exposures by adding their values, assuming a known contribution from each exposure. It has traditionally been used for various purposes, such as assessing the overall performance of countries (Siskos et al. 2014), evaluating sustainability (Dias et al. 2015), and measuring quality of life (Rădulescu et al. 2019, Huang et al. 2011; Mourmouris and Potolias 2013). The second assigns weights to the variables by performing PCA on the normalized data and selecting those with an eigenvalue greater than one, which corresponds to the maximum variation in the data. This method has been applied in ecology to consider climate change adaptation (Wu 2021). The latter method assigns weights to each exposure based on its one-quantile incremental contribution to a composite score. It has been applied to study the association between chemical exposures and child intelligence quotient scores (Tanner et al. 2019), leukemia cancer risk (Czarnota et al. 2015), and other health outcomes (Renzetti et al. 2023). Gennings et al. (2020) proposed the use of lagged WQS regression to account for longitudinal exposures. The weights are estimated for each time point, making it a useful method for analyzing such data.

This study presents a method for determining optimal weights in the AVM. The model is compared to AVM with equal weights, assuming that each exposure contributes equally to the overall effect. Additionally, it is compared to PCA and WQS, including its longitudinal version. In both the simulation study and the real data analysis, we provide new insights into the use of AVM as a simple and transparent approach to understanding the combined impact of multiple factors on uncontrolled asthma. This may lead to more effective strategies for preventing and managing such condition.

The research provides four main contributions. Firstly, it uses the Geographical Information System (GIS) to extract environmental information on children's personal exposures. Secondly, it employs the AVM to combine environmental indicators and provide a comprehensive assessment of personal exposure levels. Thirdly, it uses the logistic regression mixed model to assess the association between uncontrolled asthma and the composite environmental indicator obtained. Finally, it employs the Inverse Distance Weighted (IDW) interpolation method to obtain the probability map for uncontrolled asthma in the city of Palermo, Italy. Specifically, Sect. 2 discusses the study design, as well as the extrapolation and integration of environmental data. Section 3 covers the statistical methods used. Sections 4 and 5 are dedicated to the simulation study and the analysis of real data, respectively. Finally, Sects. 6 and 7 contain the discussion and conclusions.

2 Materials and methods

2.1 Study design and clinical assessments

The Childhood Asthma and Environment Research Study (CHASER Study) systematically enrolled a cohort of asthmatic children from September 2015 to December

2018. The study was conducted at the Pediatric Allergology and Pulmonology Outpatient Clinic at the IRIB-CNR in Palermo, Italy. A total of 179 asthmatic children underwent multiple assessments, with up to four visits per participant. Specifically, 38 participants (21.2%) completed four visits, 17 (9.5%) completed three visits, 36 (20.1%) completed two visits, and the remaining participants only completed the first visit. The average interval between consecutive visits was 4.53 ± 2.86 months.

Palermo is a Mediterranean region located in the northwest of the island of Sicily in southern Italy. It is bordered by the Tyrrhenian Sea (38°06′56″N 13°21′41″E). The city has a population of approximately 678,492 inhabitants according to the 2011 census. Palermo has a Mediterranean climate characterized by hot and dry summers, while the remaining months maintain mild temperatures. The study (Protocol N. 08/2014) was approved by the local ethics committee and registered on Clinical-Trials.gov with ID NCT02433275. Written informed consent was obtained from all parents. The inclusion criteria for participants were: (1) both males and females, (2) aged between 5 and 16 years, and (3) diagnosed with asthma according to the GINA recommendations (https://ginasthma.org, 2020). This diagnosis included assessing the history of respiratory symptoms (wheezing, shortness of breath, chest tightness, and cough), environmental exposures (such as smoke, mold, and traffic), spirometry, and skin prick testing.

The study's exclusion criteria consisted of individuals with immunologic, metabolic, cardiac, or neurologic diseases, significant malformations of the respiratory system, and active smoking. During the first visit, experienced physicians (VM and SLG) collected anamnestic data, including the age of asthma onset, frequency of severe exacerbations, and emergency visits in the previous year, and assessed individual characteristics and environmental exposures. The children underwent a clinical examination and skin prick testing. Asthma control was assessed at each visit using the Childhood Asthma Control Test (C-ACT). As per the GINA guidelines, asthma severity was assessed retrospectively based on the minimum effective treatment required to control symptoms and exacerbations. Parents were asked about changes in home address or parental lifestyle habits related to environmental tobacco smoke (ETS) at each visit. Skin prick testing was performed following the European Academy of Allergy and Clinical Immunology (EAACI) recommendations (ALK-Abellò, Milan, Italy; www.alk.it) with a standard panel of inhalant allergens, which included a positive control (histamine 1%) and a negative control (saline).

2.2 Environmental data extrapolation and integration

2.2.1 CORINE land-cover classes

The CORINE framework is a Europe-wide satellite-based land cover inventory developed by the European Environment Agency (EEA). Its primary purpose is to provide a Geographic Information System (GIS) to disseminate environmental information. The CORINE program classifies land cover into categories at a scale of 1:100,000, a classification that was lastly updated in 2006. The CORINE land cover classes (CLC) are organized into three hierarchical levels: level 1 has five categories,

level 2 has fifteen categories, and level 3 has forty-four categories, based on the unit area definition. Each home address in our study was assigned to one of the forty-four Level 3 categories.

Our research defined two overarching macro-classes. The first class/category consists in an artificial area defined by 'continuous urban fabric', where buildings and transportation infrastructure cover most of the land. The second class/category consists of 'natural and semi-natural areas', which include 'discontinuous urban fabric', 'sports and leisure facilities', 'fruit and berry plantations', 'complex cropping patterns', and 'sclerophyllous vegetation'. In this category, buildings, roads, and artificial surfaces coexist with vegetated areas.

2.2.2 Nitrogen dioxide (NO_2) concentrations

A Land Use Regression (LUR) model was used to estimate each child's exposure to NO_2 using a GIS based on their home address. The LUR method predicts pollution concentration at a given location based on various characteristics of the surrounding area, including land use, traffic intensity, proximity to emission sources, and meteorological factors. This model incorporated GIS variables that measured the length (in meters) of High Traffic Roads (HTRs) (roads with more than 10,000 vehicles per day) within 200 m of the residence. The ESCAPE project (European Study of Cohorts for Air Pollution Effects, www.escapeproj ect.eu) developed a standardized procedure for implementing LUR. This procedure includes criteria for selecting sites, defining GIS predictors, and developing multiple regression models. Linear regression models were constructed using a supervised stepwise selection procedure. The procedure began with univariate regressions of corrected annual mean concentrations and included all available potential predictors according to established procedures. The predictor with the highest adjusted R^2 was included in the model, provided that its direction of effect was consistent with the a priori definition. A search for additional predictors was conducted to increase the adjusted R^2. Only predictors that contributed the highest gain in adjusted R^2 and were consistent with the expected effect direction were selected. Any subsequent variables that changed the direction of the effect of previously selected variables were excluded. The iterative process was continued until no further variables met the criterion of adding at least 0.01 (1%) to the adjusted R^2 of the previous model in terms of the expected direction of effect. As a final step, any variable with a p value greater than 0.10 was dropped from the LUR model. If the Variance Inflation Factor (VIF) exceeded 3, indicating collinearity, the variable with the highest VIF was removed, and the model was re-evaluated. Cook's D statistic was utilized to identify influential observations. Values greater than one were subjected to further scrutiny by assessing the changes in model coefficients when the identified influential site was removed. If removal resulted in significant changes in the coefficient of a particular variable, the modeling procedure was repeated, including all sites but excluding that variable to ensure model robustness. The model's global performance was evaluated through leave-one-out cross-validation (LOOCV), sequentially excluding each site and assessing the model's performance using the remaining data. Consistent with

the ESCAPE study protocol, three one-week monitoring periods (winter, summer, and mid-season) were conducted in 2010 using 30 passive samplers for NO_2 measurements. Each passive sampler was composed of three Palms-type tubes that contained a triethanolamine solution on stainless steel mesh to facilitate NO_2 uptake. The final LUR model (model $R^2 = 0.73$; cross-validation $R^2 = 0.82$) enabled the prediction of NO_2 concentrations at the home address of each child.

2.2.3 The normalized difference vegetation index

The normalized difference vegetation index (NDVI), introduced by Weier and Herring (2011), quantifies the reflectance of the land surface to indicate greenness. Its values range from 0 to 1, with 0 indicating the absence of vegetation and values close to 1 (typically 0.8–0.9) indicating the presence of dense green foliage. The NDVI is calculated using the visible infrared (RED, 0.63–0.69 µm) and near-infrared (NIR, 0.76–0.86 µm) bands of the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) multispectral images. These images have a spatial resolution of 15 m × 15 m. The NDVI is derived using the following equation:

$$NDVI = \frac{NIR - RED}{NIR + RED}$$

The ASTER sensor is deployed on the Terra satellite, which is a component of NASA's Earth Observing System that was launched in December 1999. The satellite operates in a sun-synchronous, near-polar orbit and crosses the equator at approximately 10:30 am/pm. The ASTER instrument has a swath width of 60 km, allowing it to image every point on the Earth's surface at least once every 16 days. The sensor has 14 spectral bands, including three in the Visible and Near Infrared (VNIR) range with a spatial resolution of 15 m, six in the Shortwave Infrared (SWIR) range with a spatial resolution of 30 m, and five in the Thermal Infrared (TIR) range with a spatial resolution of 90 m (Kahle et al. 1991).

NDVI maps were generated using images extracted from the ASTER VNIR surface reflectance level 2 product (NASA LP DAAC, 2006). The ASTER VNIR data sets include reflectance values that have been corrected for atmospheric and topographic factors, such as slope and aspect. These corrections are based on available climatological data and global digital elevation data and do not have specific units.

3 Statistical analyses

3.1 The additive value model

Composite indicators are a means of combining multiple measures or scores into a single representative value. One approach that can be used for this purpose is the AVM. To create a composite indicator, follow these steps:

1. *Identify the dimensions or attributes* to be included in the composite indicator. To evaluate the overall well-being of a community, dimensions such as health, education, income, and environmental quality can be considered.
2. *Assign relative weights* to each dimension or attribute to reflect their importance in the composite indicator. These weights should sum up to 1 or 100% to represent the distribution of importance among the dimensions.
3. *Evaluate Dimensions*: Calculate a numerical value based on data or scores. For example, health metrics can be used for the health dimension and educational attainment for the education dimension.
4. Multiply each evaluated value for a dimension by its relative weight and sum up these products to *calculate the composite indicator*. The result is the value of the composite indicator, representing the overall assessment based on the different dimensions.
5. *Interpret and Communicate*: The composite indicator obtained provides a concise and informative representation that can be used for communication and overall assessment.

The AVM is a linear method used to evaluate and quantify the joint effects or values of multiple factors or attributes. To ensure comparability among different factors and their corresponding weights, the standardized data values must be summed up, as shown in the following equation:

$$v(x) = \sum_{j=1}^{k} w_j v_j(x),$$

where $v_j(x)$ and w_j are the standardized value and the corresponding weight, respectively, for the jth attribute, where $j = 1, \ldots, k$, and k representing the total number of the factors. The weights are such that $\sum_{j=1}^{k} w_j = 1$.

Our study identifies NDVI, CLC, NO2, Crowding Index (CIx), Current Mold Exposure (CME), and Maternal Smoke during Pregnancy (MSP) as useful attributes for constructing the composite indicator. NDVI and CLC were used with opposite signs to ensure that all variables have the same direction. A higher value indicates the worst situation, while a lower value indicates the best situation. In this case, a scoring algorithm generates scores for each attribute on a scale from 0.00 to 1.00. Standardized values are then calculated for each attribute:

$$v_j(x) = \frac{1 - e^{\frac{-(x_{high}-x)}{\rho}}}{1 - e^{\frac{-(x_{high}-x_{low})}{\rho}}}, j = 1, \ldots, k,$$

where $\rho = -\hat{r}(x_{high} - x_{low})$ and $\hat{r}$ is the root of the function $f(r) = -0.5 + \frac{1-e^{\frac{-z}{r}}}{1-e^{\frac{-1}{r}}}$ in which $z = \frac{(x_{high}-x_{mid})}{(x_{high}-x_{low})}$. Then, each standardized attribute should be assigned a weight to evaluate its individual contribution to the composite score. Two strategies were compared: the first assigned equal weights ($w_j = 0.167$) without any prior knowledge, and the second optimized the weights to maximize the area under the receiver

operating curve. This curve was constructed to test the association between our outcome and the composite indicator constructed. The process for determining the optimal weights is described in Sect. 1 of the Supplementary Material. Finally, the score is calculated by a linear combination of the standardized attributes and the weights.

3.2 Logistic regression mixed model

A logistic regression mixed model was used to assess the association between uncontrolled asthma and the environmental indicator obtained using the AVM approach. The model included individual-level random effects to account for repeated measures and took into account the variability between patients. The formula used to estimate the logistic regression mixed model is as follows:

$$logit(\theta) = \beta_0 + \beta_1 CI + \beta_2 Age + \beta_3 Sex + \beta_4 BMI + \gamma,$$

where $\theta = Pr(UA = 1)$ and *UA* (uncontrolled asthma) is the response variable (Yes = 1 vs. No = 0), *CI* is the composite indicator obtained using an AVM, age, sex and Body Mass Index (*BMI)* are considered as confounders to adjust for the effects between *CI* and the probability of *UA*, and $\gamma \sim N(0, \sigma^2 D)$ is the random intercept parameter, where D is a structured matrix depending on the patient identifier.

3.3 Inverse distance weighted interpolation

Inverse Distance Weighted (IDW) interpolation is a spatial interpolation technique commonly used in GIS and geostatistics to estimate values at unmeasured locations based on known values at nearby locations. The IDW interpolation method assumes that the interpolating surface should be influenced primarily by nearby scatter points, with decreasing influence attributed to more distant points. The interpolating surface is generated by calculating a weighted average of point data. The weight assigned to each scatter point decreases proportionally with the distance from the interpolation point. The steps of IDW are as follows:

1. *Data Collection*: Assign values or measurements of the variable of interest to GIS 'sample points" or "data points".
2. *Select a Location for Estimation*: Select a grid to interpolate the variable of interest.
3. *Define a Distance Metric*: Select the distance matrix to use. For example, Euclidean distance (straight-line distance), Manhattan distance (sum of absolute differences in coordinates), or other distance measures, depending on the specific problem.
4. *Calculate Distances*: Calculate the distance between the location to be estimated and each known data point using the selected distance metric.
5. *Assign Weights*: Assign weights to the data points based on their distances. The basic idea is that closer data points should affect the estimated value more than those farther away.

6. *Weighted Sum*: Calculate a weighted sum of the values associated with the data points.
7. *Estimate the Value*: The weighted sum represents the estimated value at the location of interest.

The quality of the interpolation was evaluated using the Root Mean Square Error (RMSE), which should be minimized. The RMSE was calculated using the LOOCV resampling method. For each data point, the spatial interpolation model was trained using all other data points except the one to be predicted. This process was repeated for each point until all points were predicted. The left-out point was then predicted, and subsequently added back into the data set. This process was repeated for each input point i and for each variable j,

$$RMSE_j = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(\hat{x}_j(s_i) - x_j(s_i)\right)^2},$$

where n is the number of points, s_i is the location of the hidden point, $x_j(s_i)$ is the measured value at the location and $\hat{x}_j(s_i)$ is the predicted value. To compare the RMSE results, we calculated a normalization procedure that divides the RMSE by the mean of each variable, i.e. the coefficient of variation of the RMSE:

$$NRMSE_j = \frac{RMSE_j}{\overline{x}_j},$$

where lower values indicate less residual variance.

The data were presented as absolute and percentage frequencies or as median and interquartile range. Categorical variables were compared using the X^2 test, while quantitative variables were compared using the Mann-Whitney test to avoid distributional assumptions. The analyses were performed using R version 4.2.0. Logistic regression mixed models were performed using the R package lme4. Maps were created using QGIS. A p value < 0.05 was considered statistically significant.

4 Simulation study

We compared AVM with optimal weighting, PCA and WQS regression. The comparison was based on (1) the mean $\left(\overline{w}\right)$ and the root mean squared error (RMSE) of the estimated weights, and (2) the area under the receiver operating curve (AUC) as a proxy for the association between the outcome and the constructed composite score.

For the sake of simplicity, the simulation is based on a binary logistic regression model where the outcome is $Y_1, \ldots, Y_n$ with $Y_i \sim Bern(\theta_i)$. We model $logit(\theta) = CI = X^T w$ where $X \sim N_5(0, \sigma^2 I_5)$ and w is the known vector of weights used to construct the composite indicator $CI = 0.30X_1 + 0.20X_2 + 0.50X_5$ and we assume that X_3 and X_4 are not associated with Y.

We define different settings using three sample sizes $n = \{100, 250, 500\}$ and three values of the signal-to-noise ratio (SNR), defined as the variance of the composite indicator divided by the dispersion parameter of the binomial family ($\varphi = 1$), i.e., $SNR = \frac{Var(CI)}{\varphi} = \{0.5, 1, 5\}$. In each scenario, the variance of the covariates can be derived by the SNR formula, which is used to maintain a consistent level of association between our composite score and the outcome. Additionally, a higher SNR value indicates a stronger association (see Table 1).

The effectiveness of the AVM method in constructing a composite indicator is demonstrated by the results when compared to its competitors. The AVM method exhibits high classification ability in all settings, consistently achieving a higher area under the curve than its competitors. Furthermore, there are only slight variations between AVM and WQS regarding the mean and RMSE of the estimated weights. This suggests that both methods perform well in comparison to the PCA method.

5 Results from real data

Table 2 summarizes the characteristics of the study sample based on baseline asthma control status. The children in the *UA* group had a higher frequency of persistent asthma, greater exposure to MSP, and a higher crowding index than those in the *C* group. Supplementary Fig. 1 shows the correlation structure of the six environmental indicators considered. In particular, there were strong significant correlations between NO_2 and CLC ($\rho = 0.448, p < 0.001$) and between CME and MSP ($\rho = 0.193, p < 0.001$), as indicated by Spearman's ρ statistic.

Figure 1 displays the spatial distribution of the children's homes. Interpolated maps were created for each environmental factor in Palermo, and, at such regard, Fig. 2 shows the CLC classes (panel a), NO_2 concentrations (panel b), and NDVI values (panel c) of the city. The average NO_2 concentration was 63 μg/m^3, exceeding the 2005 WHO Air Quality Guidelines limit of 40 μg/m^3. The mean NDVI value was 0.49, indicating an average presence of greenness.

The study compared three methods for constructing a composite score that included multiple environmental factors: the AVM (with equal and optimal weighting), PCA, and WQS regression (including the lagged version to account for repeated measures). The environmental factors included NDVI, CLC, NO_2, CIx, CME, and MSP. For each composite indicator constructed, we calculated a logistic regression mixed model that included individual-level random effects to account for repeated measures. This model was used to assess the association between the environmental indicator and asthma control, while adjusting for age, sex, and BMI. For each model, we assessed the uncorrelated errors over time and space assumption by analyzing the partial autocorrelation function (Supplementary Material, Fig. 2) and the Ljung-Box test statistic (Supplementary Material, Table 1) for different lag values ranging from 1 to 3, and the Moran's I statistic (Supplementary Material, Table 2), respectively. No significant violations were detected, suggesting the appropriate use of the mixed models. We selected the best model based on the AUC of the ROC, as shown in Table 3.

Table 1 Simulation results

	$w_1 = 0.30$		$w_2 = 0.20$		$w_3 = 0.00$		$w_4 = 0.00$		$w_5 = 0.50$		AUC
	$\overline{w_1}$	RMSE	$\overline{w_2}$	RMSE	$\overline{w_3}$	RMSE	$\overline{w_4}$	RMSE	$\overline{w_5}$	RMSE	
					$n = 100$						
$SNR = 0.5$											
PCA	0.198	0.106	0.202	0.027	0.200	0.201	0.200	0.202	0.201	0.300	0.647
WQS	0.254	0.127	0.183	0.105	0.079	0.104	0.078	0.108	0.406	0.161	0.716
AVM	0.257	0.144	0.195	0.144	0.060	0.102	0.060	0.109	0.428	0.162	0.735
$SNR = 1.0$											
PCA	0.198	0.106	0.203	0.027	0.199	0.200	0.197	0.200	0.203	0.298	0.692
WQS	0.247	0.117	0.183	0.098	0.066	0.086	0.061	0.086	0.444	0.126	0.773
AVM	0.256	0.119	0.186	0.113	0.045	0.080	0.046	0.087	0.466	0.129	0.794
$SNR = 5.0$											
PCA	0.195	0.109	0.207	0.026	0.196	0.198	0.198	0.199	0.204	0.297	0.780
WQS	0.272	0.076	0.178	0.069	0.043	0.057	0.045	0.064	0.462	0.070	0.888
AVM	0.280	0.080	0.181	0.073	0.028	0.048	0.029	0.056	0.483	0.077	0.910
					$n = 250$						
$SNR = 0.5$											
PCA	0.199	0.105	0.203	0.029	0.203	0.205	0.201	0.202	0.193	0.308	0.642
WQS	0.292	0.090	0.160	0.080	0.054	0.076	0.053	0.072	0.442	0.110	0.702
AVM	0.308	0.099	0.161	0.096	0.032	0.064	0.033	0.060	0.466	0.108	0.716
$SNR = 1.0$											
PCA	0.196	0.108	0.201	0.028	0.208	0.209	0.201	0.203	0.194	0.308	0.675
WQS	0.287	0.071	0.169	0.067	0.033	0.046	0.047	0.066	0.465	0.081	0.753
AVM	0.295	0.077	0.173	0.076	0.019	0.041	0.030	0.056	0.484	0.080	0.769
$SNR = 5.0$											
PCA	0.198	0.105	0.200	0.029	0.205	0.207	0.200	0.202	0.197	0.305	0.763
WQS	0.294	0.043	0.190	0.041	0.025	0.037	0.029	0.039	0.462	0.058	0.875
AVM	0.292	0.050	0.188	0.044	0.015	0.030	0.020	0.037	0.486	0.058	0.897
					$n = 500$						
$SNR = 0.5$											
PCA	0.202	0.103	0.198	0.031	0.198	0.201	0.204	0.206	0.198	0.303	0.647
WQS	0.275	0.061	0.194	0.064	0.043	0.057	0.036	0.050	0.452	0.076	0.698
AVM	0.280	0.063	0.200	0.064	0.027	0.049	0.024	0.044	0.470	0.084	0.710
$SNR = 1.0$											
PCA	0.197	0.108	0.202	0.029	0.203	0.205	0.200	0.202	0.198	0.303	0.686
WQS	0.277	0.047	0.192	0.048	0.030	0.042	0.034	0.046	0.467	0.052	0.752
AVM	0.283	0.052	0.195	0.052	0.019	0.040	0.022	0.041	0.481	0.054	0.768
$SNR = 5.0$											
PCA	0.197	0.106	0.201	0.029	0.201	0.203	0.203	0.205	0.198	0.304	0.771
WQS	0.288	0.028	0.201	0.027	0.017	0.025	0.021	0.028	0.473	0.038	0.877
AVM	0.292	0.034	0.196	0.037	0.009	0.023	0.011	0.024	0.494	0.040	0.898

AVM, PCA and WQS were compared at different settings in terms of the mean ($\bar{w}$) and RMSE of the estimated weights, and in terms of the AUC of the receiver operating curve, which assesses the association between the outcome and the corresponding constructed composite indicator

Table 1 (continued)

Specifically, varying the sample size $n = \{100, 250, 500\}$ and the signal-to-noise ratio $SNR = \{0.5, 1, 5\}$.

Table 2 Demographic characteristics by asthma control status at baseline

	All	C	UA	*p* value
n	179	129	50	
Age, years, median (IQR)	8.00 (3.00)	8.00 (3.00)	9.00 (4.00)	0.161
Gender: Female, n (%)	72 (40.22)	47 (36.43)	25 (50.00)	0.136
BMI, kg/m², median (IQR)	18.6 (5.80)	18.1 (4.40)	20.0 (6.10)	0.099
Persistent asthma, n (%)	99 (55.31)	62 (48.06)	37 (74.00)	0.003
CME, n (%)	38 (21.59)	22 (17.46)	16 (32.00)	0.056
CIx, median (IQR)	1.00 (0.53)	1.00 (0.53)	1.10 (0.36)	0.020
MSP, n (%)	16 (9.04)	7 (5.51)	9 (18.00)	0.020

UA: ACT/C-ACT ≤ 19, C: ACT/C-ACT > 19

CME Current Mold Exposure, *CIx* Crowding index, *MSP* Maternal Smoke during Pregnancy
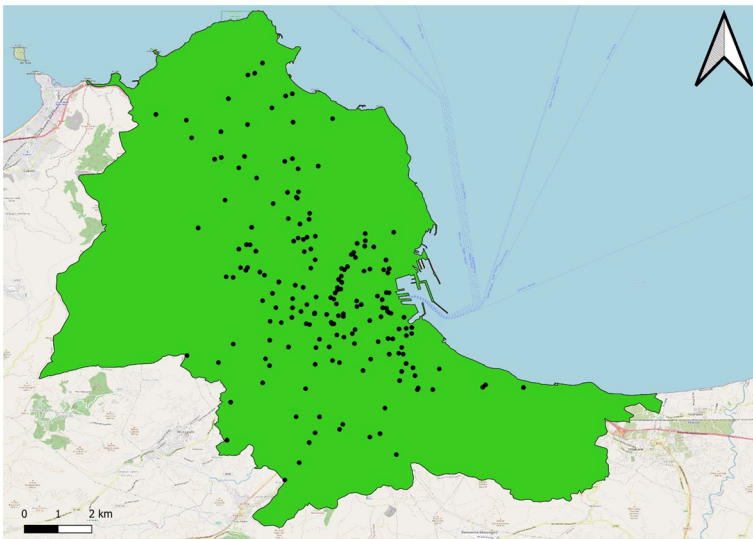
Fig. 1 Spatial distribution of children's house

The study results suggest that the AVM with optimal weighting, as recommended by the simulation study, outperformed the competitors in terms of AUC. Therefore, we first consider the environmental indicator obtained by the AVM, which was significantly associated with the outcome (log(OR) = 1.982, standard error = 0.994, *p* value = 0.046), to create an interpolated score map. We then generate a probability map for uncontrolled asthma. Figures 3 and 4 display the interpolated score, which ranges from 0 to 1. A score of 0 represents the best environmental conditions, while
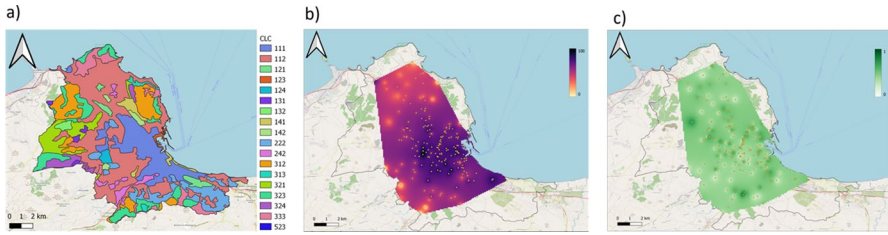
Fig. 2 panel (**a**), CLC classes, panel (**b**), NO_2 interpolated concentrations, and panel (**c**) NDVI interpolated level maps of Palermo city

a score of 1 represents the worst. The probability map indicates areas with a high probability of uncontrolled asthma. 'Red' and 'yellow' areas indicate a moderate probability of uncontrolled asthma.

Table 4 presents the LOOCV NRMSE results for the quality of the IDW interpolation, indicating reasonable performance for the interpolated maps.

6 Discussion

Our study compares different approaches to constructing composite indicators and proposes the use of the Additive Value Model (AVM), with the double aim to combine environmental indicators and provide a comprehensive assessment of personal exposure. The AVM composite score provided the best ability to predict outcomes, as represented by the AUC of the receiver operating curves, in both the simulation study and the real data analysis. A significant association was found

Table 3 Estimated weights for the construction of the environmental indicators and AUC of the investigated methods, i.e., principal components analysis PCA, weighted quantile sum WQS regression (including the lagged version to account for repeated measures) and additive value model AVM (with equal and optimal weights)

Method		Weights						AUC
		NDVI	CIx	NO_2	CLC	CME	MSP	
PCA		0.175	0.089	0.101	0.086	0.323	0.226	0.912
WQS		0.118	0.339	0.001	0.173	0.184	0.184	0.902
WQS (lagged)	T0	0.024	0.444	0.016	0.241	0.136	0.139	0.891
	T1	0.301	0.159	0.028	0.112	0.200	0.200	
	T2	0.134	0.238	0.006	0.394	0.114	0.114	
	T3	0.100	0.205	0.004	0.516	0.088	0.087	
AVM (equal weights)		0.167	0.167	0.167	0.167	0.167	0.167	0.899
AVM (optimal weights)		0.550	0.100	0.050	0.050	0.200	0.050	0.922

NDVI Normalized Difference Vegetation Index, *CIx* Crowding Index, *CLC* Corine Land Cover, *CME* Current Mold Exposure, *MSP* Maternal Smoke during Pregnancy
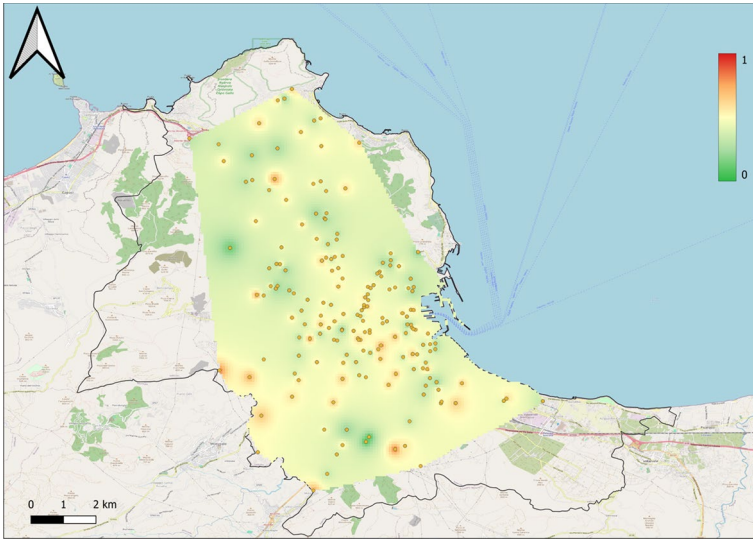
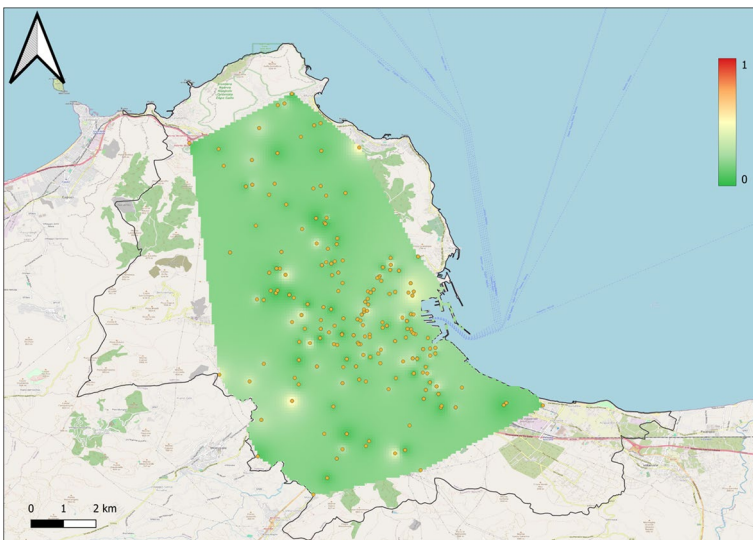**Fig. 3** Interpolated environmental indicator score

Fig. 4 Interpolated probability map

between the composite indicator and asthma control, as well as the higher importance of the NDVI in predicting asthma control. Since no study to date has evaluated the effect of a composite indicator on asthma control, the results of the present study cannot be directly compared with those of previous studies that have examined asthma control and environmental measures using standard approaches such as linear regression.

Table 4 Leave-one-out cross-validation normalized RMSE

	NRMSE
NO$_2$	0.136
NDVI	0.380
Environmental CI	0.331
Probability map	0.432

Univariate analyses revealed borderline significant associations between asthma control and maternal smoking during pregnancy, crowding index, and current mold exposure. Maternal smoking during pregnancy may increase the risk of developing asthma due to changes in placental cytokine production (Macaubas et al. 2003). Previous studies have also suggested that maternal smoking during pregnancy can have long-lasting effects on lung development, function, and respiratory health later in life (Li et al. 2000; Zacharasiewicz 2016). Housing crowding can be considered a marker of socioeconomic status, as a lower score is associated with better social and financial status (Burr et al. 2010). A higher crowding index increases the risk of hypertension, respiratory disease, and infection, which can lead to more stress, impaired social relationships, and poor sleep (Gray 2001; Illi et al. 2012). Consistent with previous research, the crowding index was found to be associated with poorer control in children with asthma (Illi et al. 2012; Kopel et al. 2014). While a borderline association with current mold exposure was observed, the results are in line with a study that reported increased mold levels to be linked with higher rates of emergency department visits, hospitalizations, and inpatient stays for asthma (Lewis et al. 2020). A recent study conducted in Puerto Rico found that children living in homes with mold were more likely to have uncontrolled asthma (Cowan et al. 2022). The importance of NDVI may be attributed to its potential as an indicator of improved air quality, reduced stress, and increased physical activity.

Our study has several methodological strengths. Specifically, AVMs are easier to understand and implement than more complex models. The linear nature of the model provides transparency in understanding how each factor contributes to the overall value. Additionally, the results of the model are easily interpretable, making it accessible to a wide audience, such as doctors. However, it is important to note that the methodology has some limitations. For example, it assumes linearity, which may not hold in all real-world situations. Similarly, other methods such as WQS have reported limitations such as independence of exposure measures and model complexity, which can be computationally demanding depending on the size of the dataset.

7 Conclusions

This paper proposes an innovative application of the additive value model to construct a composite environmental indicator. The study comprehensively assesses the influence of urban environmental factors on childhood asthma control using

multiple indicators measuring green spaces and indoor/outdoor environments. Composite indicators offer several advantages in different areas. The tool provides a comprehensive view by integrating various environmental factors, both indoor and outdoor, into a single measure. It simplifies complex information by condensing it into a single score, making it easier for pediatricians to interpret and understand. Additionally, it acts as a communication tool, simplifying complex information for parents and raising awareness and understanding of asthma, that is indeed a complex issue. The study provides insights into the relationship between exposure to green spaces and indoor environments and asthma control in children. These findings could inform the implementation of smoke-free policies and nature-based solutions in urban areas.

Declarations

Conflict of interest The authors have no conflicts of interest to declare.

References

Aerts R, Dujardin S, Nemery B et al (2020) Residential green space and medication sales for childhood asthma: a longitudinal ecological study in Belgium. Environ Res 189:109914

Andrusaityte S, Grazuleviciene R, Kudzyte J et al (2016) Associations between neighbourhood greenness and asthma in preschool children in Kaunas, Lithuania: a case–control study. BMJ Open 6:e010341

Bateman ED, Hurd SS, Barnes PJ et al (2008) Global strategy for asthma management and prevention: GINA executive summary. Eur Respir J 31:143–178

Burr JA, Mutchler JE, Gerst K (2010) Patterns of residential crowding among hispanics in later life: immigration, assimilation, and housing market factors. J Gerontol B Psychol Sci Soc Sci 65:772–782

Carrico C, Gennings C, Wheeler DC, Factor-Litvak P (2015) Characterization of weighted quantile sum regression for highly correlated data in a risk analysis setting. J Agric Biol Environ Stat 20:100–120. https://doi.org/10.1007/s13253-014-0180-3

Cavaleiro Rufo J, Paciencia I, Hoffimann E et al (2021) The neighbourhood natural environment is associated with asthma in children: a birth cohort study. Allergy 76:348–358

Chen E, Miller GE, Shalowitz MU et al (2017) Difficult family relationships, residential greenspace, and childhood asthma. Pediatrics. https://doi.org/10.1542/peds.2016-3056

Cilluffo G, Ferrante G, Fasola S et al (2018) Associations of greenness, greyness and air pollution exposure with children's health: a cross-sectional study in Southern Italy. Environ Health 17:1–12

Cilluffo G, Ferrante G, Fasola S et al (2022) Association between greenspace and lung function in Italian children-adolescents. Int J Hyg Environ Health 242:113947

Cowan KN, Qin X, Ruiz Serrano K et al (2022) Uncontrolled asthma and household environmental exposures in Puerto Rico. J Asthma 59:427–433

Czarnota J, Gennings C, Wheeler DC (2015) Assessment of weighted quantile sum regression for modeling chemical mixtures and cancer risk. Cancer Inform https://doi.org/10.4137/CIN.S17295

Dadvand P, Villanueva CM, Font-Ribera L et al (2014) Risks and benefits of green spaces for children: a cross-sectional study of associations with sedentary behavior, obesity, asthma, and allergy. Environ Health Perspect 122:1329–1335

Dharmage SC, Perret JL, Custovic A (2019) Epidemiology of asthma in children and adults. Front Pediatr 7:246

Dias LC, Silva S, Alçada-Almeida L (2015) Multi-criteria environmental performance assessment with an additive model. Handb Methods Appl Environ Stud. https://doi.org/10.4337/9781783474646.00027

Feng X, Astell-Burt T (2017) Is neighborhood green space protective against associations between child asthma, neighborhood traffic volume and area safety? Multilevel analysis of 4447 Australian children. J Transp Health 5:S40–S41

Ferrante G, Asta F, Cilluffo G et al (2020) The effect of residential urban greenness on allergic respiratory diseases in youth: a narrative review. World Allergy Organ J 13:100096

Gennings C, Curtin P, Bello G et al (2020) Lagged WQS regression for mixtures with many components. Environ Res 186:109529. https://doi.org/10.1016/j.envres.2020.109529

Gray A (2001) Definitions of crowding and the effects of crowding on health: A literature review. Prepared for the ministry of social policy. Ministry of Social Development (MSD). Wellington, New Zealand, pp 1–40

Hanushek EA, Rivkin SG (2006) Chap. 18 teacher qualit. Handbook Econ Educ 2:1051–1078

Huang IB, Keisler J, Linkov I (2011) Multi-criteria decision analysis in environmental sciences: ten years of applications and trends. Sci Total Environ 409:3578–3594

Illi S, Depner M, Genuneit J et al (2012) Protection from childhood asthma and allergy in alpine farm environments—the GABRIEL advanced studies. J Allergy Clin Immunol 129:1470–1477

Kahle AB, Palluconi FD, Hook SJ et al (1991) The advanced spaceborne thermal emission and reflectance radiometer (Aster). Int J Imaging Syst Technol 3:144–156. https://doi.org/10.1002/ima.1850030210

Kopel LS, Phipatanakul W, Gaffin JM (2014) Social disadvantage and asthma control in children. Paediatr Respir Rev 15:256–263

Lewis LM, Mirabelli MC, Beavers SF et al (2020) Characterizing environmental asthma triggers and healthcare use patterns in puerto rico. J Asthma 57:886–897

Li Y-F, Gilliland FD, Berhane K et al (2000) Effects of in utero and environmental tobacco smoke exposure on lung function in boys and girls with and without asthma. Am J Respir Crit Care Med 162:2097–2104

Lovasi GS, Quinn JW, Neckerman KM et al (2008) Children living in areas with more street trees have lower prevalence of asthma. J Epidemiol Community Health 62:647–649

Lovasi GS, O'Neil-Dunne JP, Lu JW et al (2013) Urban tree canopy and asthma, wheeze, rhinitis, and allergic sensitization to tree pollen in a New York city birth cohort. Environ Health Perspect 121:494–500

Macaubas C, De Klerk N, Holt B et al (2003) Association between antenatal cytokine production and the development of atopy and asthma at age 6 years. Lancet 362:1192–1197

McCaffrey DF, Lockwood JR, Koretz D et al (2004) Models for value-added modeling of teacher effects. J Educ Behav Stat Q Publ Spons Am Educ Res Assoc Am Stat Assoc 29:67–101. https://doi.org/10.3102/10769986029001067

Montalbano L, Cilluffo G, Gentile M et al (2016) Development of a nomogram to estimate the quality of life in asthmatic children using the childhood asthma control test. Pediatr Allergy Immunol 27:514–520

Mourmouris JC, Potolias C (2013) A multi-criteria methodology for energy planning and developing renewable energy sources at a regional level: a case study thassos, greece. Energy Policy 52:522–530

Rădulescu CZ, Rădulescu M, Alexandru A et al (2019) A multi-criteria weighting approach for quality of life evaluation. Procedia Comput Sci 162:532–538

Renzetti S, Gennings C, Calza S (2023) A weighted quantile sum regression with penalized weights and two indices. Front Public Health 11:1151821. https://doi.org/10.3389/fpubh.2023.1151821

Sendhil R, Jha A, Kumar A, Singh S (2018) Extent of vulnerability in wheat producing agro-ecologies of India: tracking from indicators of cross-section and multi-dimension data. Ecol Indic 89:771–780. https://doi.org/10.1016/j.ecolind.2018.02.053

Siskos E, Askounis D, Psarras J (2014) Multicriteria decision support for global e-government evaluation. Omega 46:51–63

Squillacioti G, Bellisario V, Levra S et al (2020) Greenness availability and respiratory health in a population of urbanised children in North-Western Italy. Int J Environ Res Public Health 17:108

Tanner EM, Bornehag C-G, Gennings C (2019) Repeated holdout validation for weighted quantile sum regression. MethodsX 6:2855–2860. https://doi.org/10.1016/j.mex.2019.11.008

Tischer C, Gascon M, Fernández-Somoano A et al (2017) Urban green and grey space in relation to respiratory health in children. Eur Respir J 49:1502112

Weier J, Herring D (2011) Measuring vegetation (ndvi & evi). In: earth obs. NASA. https://earthobservatory.nasa.gov/Features/MeasuringVegetation/

Wu T (2021) Quantifying coastal flood vulnerability for climate adaptation policy using principal component analysis. Ecol Indic 129:108006. https://doi.org/10.1016/j.ecolind.2021.108006

Wu M, Xie J, Wang Y, Tian Y (2021) Greenness and eosinophilic asthma: findings from the UK biobank. Eur Respir J 58:2101597

Zacharasiewicz A (2016) Maternal smoking in pregnancy and its influence on childhood asthma. ERJ Open Res. https://doi.org/10.1183/23120541.00042-2016

Authors and Affiliations

Giovanna Cilluffo[1,2] · Gianluca Sottile[3] · Giuliana Ferrante[4] · Salvatore Fasola[5] · Velia Malizia[5] · Laura Montalbano[5] · Andrea Ranzi[6] · Chiara Badaloni[7] · Giovanni Viegi[5,8] · Stefania La Grutta[5]

✉ Giovanna Cilluffo
giovanna.cilluffo@unipa.it

Gianluca Sottile
gianluca.sottile@unipa.it

Giuliana Ferrante
giuliana.ferrante@univr.it

Salvatore Fasola
salvatore.fasola@ift.cnr.it

Velia Malizia
velia.malizia@ift.cnr.it

Laura Montalbano
laura.montalbano@ift.cnr.it

Andrea Ranzi
aranzi@arpae.it

Chiara Badaloni
c.badaloni@deplazio.it

Giovanni Viegi
viegig@ifc.cnr.it

Stefania La Grutta
stefania.lagrutta@ift.cnr.it

¹   Department of Earth and Marine Sciences, University of Palermo, Palermo, Italy

²   National Inter-University Consortium for Marine Sciences CoNISMa, Rome, Italy

³   Department of Economics, Business and Statistics, University of Palermo, Palermo, Italy

⁴   Pediatric Division, Department of Surgical Sciences, Dentistry, Gynecology and Pediatrics,
     University of Verona, 37134 Verona, Italy

⁵   Institute of Translational Pharmacology (IFT), National Research Council, 90146 Palermo, Italy

⁶   Environmental Health Reference Centre, Regional Agency for Environmental Prevention
     of Emilia-Romagna, Bologna, Italy

⁷   Department of Epidemiology, Lazio Regional Health Service, ASL Roma 1, Rome, Italy

⁸   Institute of Clinical Physiology (IFC), National Research Council of Italy, Pisa, Italy