



RESEARCH

How does air quality reflect the land cover changes: remote sensing approach using Sentinel data

Pamela Guamán-Pintado · Evelyn Uuemaa ·
Danilo Mejia · Szilárd Szabó

Received: 12 July 2024 / Accepted: 26 November 2024
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2024

Abstract Significant environmental challenges, such as urban and industrial expansion, alongside vegetation preservation, directly influence the concentrations of critical air pollutants and greenhouse gases in cities and their surroundings. The urban development and expansion process is aptly captured by classifying land use and land cover (LULC). We aimed to analyze LULC changes in an Andean area, Ecuador, and to reveal the relations of LULC classes with three air pollutants ozone (O_3), nitrogen dioxide (NO_2), and sulfur dioxide (SO_2), using remote sensing datasets (Sentinel-5P - Sentinel 1 - Sentinel-2) across different periods. Results showed that SO_2 is not a reliable indicator for assessing its behavior based on LULC classes, as it was difficult to distinguish between different land cover types using this pollutant. For NO_2 , the analysis showed a

moderate distinction among LULC classes, suggesting some variability in its distribution across different land cover classes. On the other hand, O_3 analysis shows that all land cover classes are statistically distinguishable, demonstrating that urban, shrubland, green areas, and forest classes influenced ozone distribution. These findings emphasize the importance of accurate land cover classification in understanding air pollutants' spatial distribution and dynamics. This analysis is crucial for understanding the impacts of land use and land cover changes on urban health and well-being and the effects of rapid urban expansion.

Keywords Urban expansion · Land cover changes · Machine learning · Environmental monitoring · Google Earth Engine · Sentinel 5-P · Statistical significance

P. Guamán-Pintado (✉) · E. Uuemaa
Department of Geography, University of Tartu, Vanemuise 46,
Tartu 51003, Tartumaa, Estonia
e-mail: pamela.maricela.guaman.pintado@ut.ee

E. Uuemaa
e-mail: evelyn.uuemaa@ut.ee

P. Guamán-Pintado · S. Szabó
Department of Physical Geography and Geoinformatics, University of Debrecen, Egyetem tér 1, Debrecen 4032, Hajdú-Bihar,
Hungary
e-mail: szabo.szilard@science.unideb.hu

D. Mejia
CATOx group, University of Cuenca, Campus Balzay, 010207
Cuenca, Ecuador
e-mail: danilo.mejia@ucuenca.edu.ec

1 Introduction

The rapid expansion of urban areas and changes in land use patterns have led to a significant increase in air pollution, representing a global challenge impacting human health and the environment (Cohen et al., 2017). It is linked to numerous health issues, including respiratory and cardiovascular ailments, affecting the most susceptible groups children, elderly, and expectant mothers (Salvi, 2007). The World Health Organization (WHO) stated that polluted air is inhaled by 90% of the world's population, contributing to approximately 7 million premature deaths each year (WHO,

2022a). Additionally, more than 3 billion individuals globally depend on solid fuels such as wood, charcoal, and coal for cooking and heating, resulting in elevated indoor air pollution levels contributing to roughly 4 million early fatalities annually (Riojas-Rodríguez et al., 2016; WHO, 2022b). Most of these deaths occurred in low and middle-income countries, including South America (WHO, 2022a), where air pollution levels are often high due to the interaction of anthropogenic-natural activities, such as transportation, biomass burning, industrial processes, wildfires, and dust storms (Lucila et al., 2016; Manosalidis et al., 2020).

Furthermore, the economic costs of air pollution include healthcare expenses, productivity loss, and damage to infrastructure and crops (Ali et al., 2023; Sapkota & Bastola, 2017). The environmental consequences of air pollution include climate change, acid rain, and damage to ecosystems and biodiversity (Manosalidis et al., 2020; Syuhada et al., 2023). Urban regions, notably cities such as Mexico City, São Paulo, and Santiago, consistently register elevated air pollutant concentrations (Bell et al., 2006). Urban air quality in Ecuador is significantly impacted by vehicular traffic and industrial sources, which are major contributors to air pollutant emissions and consequently diminished air quality (Parra & Espinoza, 2020); for instance, larger cities such as Quito and Cuenca record some of the most elevated PM_{2.5} pollutant concentrations (C40-Cities, 2022; Empresa Pública Municipal de Movilidad, Tránsito y Transporte de Cuenca - EMOV EP, 2022).

To effectively combat air pollution and its adverse effects, early detection and monitoring are crucial for shaping effective strategies; in that way, remote sensing stands out as a potent tool, especially in areas lacking ground-based monitoring or with concentrated pollution, offering insights into pollutant types, concentrations, and spatiotemporal patterns (Ialongo et al., 2020; Zhang et al., 2022). Among various techniques, satellite imagery is predominant, capturing extensive atmospheric views and aiding in tracking global pollution trends and cross-border pollutant movements. While each remote sensing method has its merits and constraints, the choice hinges on the specific pollutant and application (Vîrghileanu et al., 2020). Monitoring data guides policymakers on pollution's sources and impacts. As global concerns about air quality intensify, it becomes increasingly critical to understand its scope, underscoring the importance of remote sensing technology (Huang et al., 2022; Zhang et al., 2022).

On the other hand, it is important to understand the factors such as vegetation, weather conditions, socio-economics, and its spatial patterns and how they are related to the increase of air pollutants (Luo et al., 2017). Urban expansion, driven by population growth, often replaces forests and affects land cover, e.g., vegetation helps remove pollutants and reduces greenhouse gases, but urbanization and human activities amplify emissions (Bonilla-Bedoya et al., 2021; Frimpong et al., 2022; Jiang et al., 2021). Furthermore, landfills and agricultural activities can release pollutants, such as methane and ammonia, into the atmosphere (Bechle et al., 2011; Escobedo et al., 2011). Therefore, we can say land cover influences the distribution of air pollution recognizing the balance between urban growth, green area preservation (Hawash et al., 2021). Despite extensive studies on changes in LULC, most of them are focused on biodiversity impacts, but studies on air quality effects remain particularly limited (Frimpong et al., 2022). Many studies link land cover with air pollutants using ground data; however, the role of geospatial technologies in bolstering these studies needs further exploration (Ialongo et al., 2020).

Remote sensing data studies have significantly advanced our understanding of the relationship between land use land cover changes, and air pollution. For instance, Jodhani et al. (2024) analyzed the impact of urban and industrial expansions in Gujarat, India, finding that these factors led to increased concentrations of key pollutants such as carbon monoxide and nitrogen dioxide. In contrast, Frimpong et al. (2022) documented air quality improvements in East Baton Rouge, attributing these to stringent environmental regulations despite urban growth. Similarly, Bonilla-Bedoya et al. (2021) demonstrated in Andean cities that increased urban forestry significantly correlated with reductions in air pollutants, suggesting that green infrastructure could mitigate pollution effects. Vadrevu et al. (2017) also noted that in South and Southeast Asia, urbanization and deforestation exacerbated air pollution through heightened energy demands and biomass burning. In this study, we introduced an approach by integrating machine learning algorithms to refine LULC data classification, which guarantees more precise data and enables a detailed examination of how LULC changes influence the distribution and concentration of gaseous pollutants in urban environments. This research aims to identify the connection between LULC types with gaseous pollutants data (O_3 , NO_2 , SO_2), concentrat-

ing on how these pollutants are distributed across urban spaces and identifying contributing factors.

2 Materials and methods

2.1 Study area

This study focuses on Cuenca city, the third largest city in Ecuador, and its surrounding settlements, Paute and Azogues (Fig. 1). Cities (Cuenca, Paute, and Azogues) have an average height of 2500 m.a.s.l., and have a population of around 600k, 26k, and 40k inhabitants respectively (National Institute of Statistics and Census of Ecuador, 2022). The study area is located in the Sierra region of the Andean highlands, within the provinces of Azuay and Cañar. The study area

encompasses 1070 km², with a complex topography of inter-Andean valleys and mountains. The climate is classified as subtropical highland, characterized by mild temperatures, frequent rainfall, and relatively high humidity. The annual rainfall ranges between 700 and 1000 mm/year (Empresa Pública Municipal de Movilidad, Tránsito y Transporte de Cuenca - EMOV EP, 2022).

This region's climate exhibits dynamic seasonal patterns (Parra & Espinoza, 2020), with a pronounced wet season from March to May and a dry season from June to September. During the wet season, temperatures stay around the lower end of the annual average (15 – 16°C), and vegetation growth increases, which may help absorb pollutants. In the dry season, temperatures reach between 17.6°C and 22°C at their peak, with less atmospheric moisture, often leading to higher

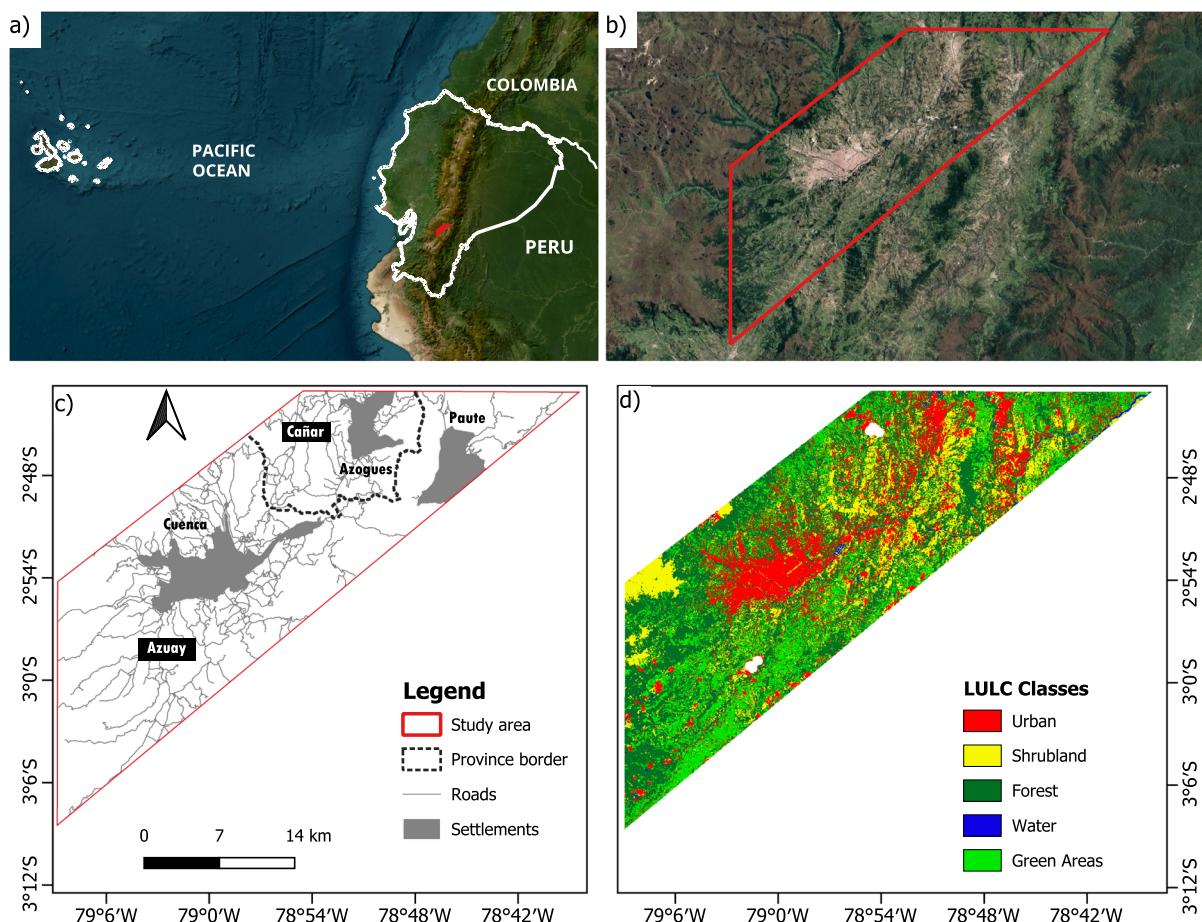


Fig. 1 Overview of the study area; **a** location of the study area in Ecuador; **b** zoom in of the location of the study area; **c** settlements boundary in Azuay and Cañar provinces; and **d** LULC map of the study area)

pollutant levels due to limited natural dispersion. Tobar and Wyseure (2018). Northeast winds dominate year-round, averaging 1.95 m/s, with stronger winds during the dry season that aid pollutant dispersion within the valley structure. Additionally, interannual climate phenomena, particularly El Niño and La Niña events, significantly affect rainfall and temperature variability, leading to year-to-year differences in vegetation density and pollutant distribution across the study area (Parra & Espinoza, 2020).

2.2 Data collection

We implemented an integrated approach for land cover classification using data from Sentinel-2 Level-2A and Sentinel-1 satellites (Table 1), both part of the Copernicus program designed to monitor the Earth. Sentinel-2 Level-2A data provides atmospherically-corrected surface reflectance across 13 spectral bands ranging from visible to shortwave infrared, essential for detailed analyses in agriculture, forestry, and land cover mapping applications (Page et al., 2020).

To complement Sentinel-2 optical data, we employed Sentinel-1 SAR GRD, derived from the C-band Synthetic Aperture Radar (SAR). This SAR technology captures the Earth's surface under all weather conditions and at any time of day, making it especially valuable in regions with high cloud cover or challenging weather. Sentinel-1 provides dual polarization data

(VV and VH), which enhances land cover classification accuracy by distinguishing surface textures and structures, particularly useful for separating urban areas from forest cover.

The Ground Range Detected (GRD) data, pre-processed and provided in log scaling, represents radar backscatter values. Additionally, the dual polarization capability of Sentinel-1's SAR enables its application in various fields, including land cover mapping, disaster monitoring, and maritime surveillance (Page et al., 2020). Combining Sentinel-1 Synthetic Aperture Radar (SAR) data with Sentinel-2 multispectral data can improve classification accuracy by exploiting the complementary information provided by each sensor (Vizzari, 2022).

We also included atmospheric data of NO_2 , SO_2 , and O_3 collected from Google Earth Engine (GEE). This data comes from the Sentinel-5P satellite, specifically designed for atmospheric monitoring. Its initial Level 2 (L2) data are organized by time, not latitude and longitude. To facilitate data integration in GEE, the data was converted to Level 3 (L3) using the “hard-cover” tool, where data is filtered to remove pixels with Quality Control (QC) values below 75% for the vertical column band tropospheric NO_2 , while for O_3 and SO_2 were ingested directly as L3 products without additional filtering (Kaplan & Avdan, 2020).

The data from Sentinel-1 and Sentinel-2 were maintained at a 10m resolution. Sentinel-5P, originally at 1113.2 ms, was resampled to 10 ms to align with this

Table 1 Characteristics of satellites and sensors used in the study

Satellite	Instrument	Instrument type	Spatial resolution	Bands	Purpose	Study period
Sentinel-1	C-SAR (C-band Synthetic Aperture Radar)	Imaging microwave radar	10m	VV, VH	ML inputs for LULC model	1-31 May 2019 1-31 May 2022
Sentinel-2	MSI (Multi-Spectral Instrument)	High resolution optical images	10m, 20m, 60m	B1, B2, B3, B4, B5, B6, B7, B8, B8A, B9, B11, B12	ML inputs for LULC model	15-May 2019 12-May 2022
Sentinel-5P	Tropospheric Monitoring Instrument (Column number density [mol/m^2])	Optical Imaging spectrometer	1113.2m	UVNS (UV NIR SWIR) Spectrometer	air quality analysis	March-April May 2019 March-April May 2022

resolution. The land cover classification was conducted for two periods, May 2019 and May 2022, with the following land cover classes: 1) urban, 2) shrubland, 3) forest, 4) water, and 5) green areas (Table 2).

2.3 Land cover classification using random forest

We used random forest (RF) algorithm, recognized for its robust performance in complex land cover classification tasks (Abida et al., 2022; Talukdar et al., 2020). The land cover model was trained using a combination of the bands from Sentinel-1 and Sentinel-2, 14 input bands in total. Given the need for finer spatial detail to capture smaller land features, we chose a resolution of 10 m. The training samples were manually selected from periods shown in Table 1. The images underwent extensive preprocessing, including noise reduction with the refined Lee filter, radiometric calibration, and data fusion, ensuring high data quality and comparability between the different satellite datasets.

The study produced two LULC maps for 2019 and 2022, processed and classified in Google Earth Engine using the ee.Classifier.smileRandomForest method. The RF model was configured with 10 decision trees to optimize accuracy and computational efficiency. It was trained on a dataset of 6038 labeled samples, representing five distinct land cover classes (Table 2). The division of the dataset into 80% training and 20% validation sets (Feng et al., 2022) ensured rigorous assessment and minimized overfitting, underpinning the reliability of our classification results. Each land cover class was represented by at least 50 samples (Basheer et al., 2022), ensuring adequate representation across classes. Specifically, the sample distribution included 318 samples for urban, 716 for water, 2260 for forest, 2417 for shrubland, and 327 for green areas.

Each tree of random forest used approximately 2408 data points for classification, which were strategically

developed based on optimal splits that minimized the classification error at each node. Terminal nodes, or leaves, were defined for cases that could no longer be split, effectively categorizing them into one of the pre-defined land use classes. This process allowed the RF to learn from complex patterns in the data, enhancing the classification accuracy of the observed land cover types.

The LULC classification accuracy was assessed using the validation dataset, resulting in an overall accuracy of 99% and a Kappa coefficient of 0.9857. Class-specific user's and producer's accuracies were also calculated to evaluate classification robustness, showing the model's effectiveness in distinguishing between the five land cover classes. These accuracy assessments confirm the reliability of the classification in capturing the land cover variability within the study area.

2.4 Analysis of land cover classification and air pollutants

Sentinel-5P (S-5P) data for O_3 , SO_2 , and NO_2 were collected during the annual wet season from March to May, aligning this collection with land cover data acquisitions for 2019 and 2022. Specifically, May was selected to enable meaningful comparisons between air pollutant concentrations and land cover classifications, ensuring temporal consistency for our multi-year environmental analysis.

To effectively manage the extensive dataset of 10,358,000 pixels, a stratified sampling method was employed. By utilizing a proportional allocation strategy, we sampled only 0.01% of the most significant points from each land cover class, ensuring that the sampled subset accurately represented the overall dataset and maintained the same statistical characteristics across each defined subgroup.

Table 2 Name and description of land cover classes scheme

Class name	Class description
Urban	Buildings, residential, commercial, industrial, roads and transportation, communications, and utilities.
Shrubland	Shrub, shrub-grass, cropland, herb, bare land, pen fields with little vegetation, mixed barren land.
Forest	Natural forest, mixed forest
Water	Lakes, rivers, and ditches
Green areas	Parks, home grass yards, garden, natural or induced vegetation located in urban areas.

According to the statistical Shapiro-Wilk test, our dataset sample did not follow a normal distribution; therefore, a non-parametric statistical test is ideal for analyzing the variance in heterogeneity across classes. The Kruskal-Wallis test was applied to analyze the relationship between LULC classes and atmospheric pollutants. To provide a numeric representation of LULC, various spectral indices were tested in this study; however, the Normalized Difference Vegetation Index (NDVI) had the best performance representing not only vegetative areas like the “green areas” and “forest” classes but also in effectively capturing characteristics of urban areas, making it versatile across land cover types. NDVI provides continuous values from 1 to -1, allowing for a more precise numeric analysis compared to categorical LULC classes, making it particularly suitable for non-parametric tests like Kruskal-Wallis (da Silva et al., 2020; Szabo et al., 2016).

Furthermore, the Bonferroni correction was applied during pairwise comparisons of LULC classes to adjust the significance levels appropriately. The Kruskal-Wallis null hypothesis testing and correlation analysis were conducted in R 3.4.2 software (R Core Team 2022) with the help of the “coin” and “ggplot2” packages (Hothorn et al., 2006; Wickham, 2009), which provided the necessary tools for detailed statistical evaluation.

3 Results

3.1 Air quality data analysis

This section analyzes the statistics of the data samples of each air pollutant: ozone O_3 , sulfur dioxide SO_2 , and nitrogen dioxide NO_2 , in addition to their spatiotemporal distribution in the study area.

3.1.1 Air quality dynamics and distribution

Figure 2 shows the box plots for these three pollutants, and Fig. 3 illustrates the dataset’s frequency histogram, density curve, and theoretical normal distribution. For O_3 , a slight asymmetry to the left (negatively skewed) is observed. This is supported by the mean (0.1127 mol/m^2) being smaller than the median (0.1129 mol/m^2). Additionally, the distribution appears to have two dominant modes, which disqualifies the assumption of normality for O_3 . In the case of SO_2 , the distribution is clearly skewed to the right, as shown in the box plot where most values are concentrated near the minimum value $0.08 \mu\text{mol/m}^2$, with a median of $28 \mu\text{mol/m}^2$. For NO_2 , the mean is $33.2 \mu\text{mol/m}^2$. The box plot displays a range from the first quartile $32.2 \mu\text{mol/m}^2$ to the third quartile

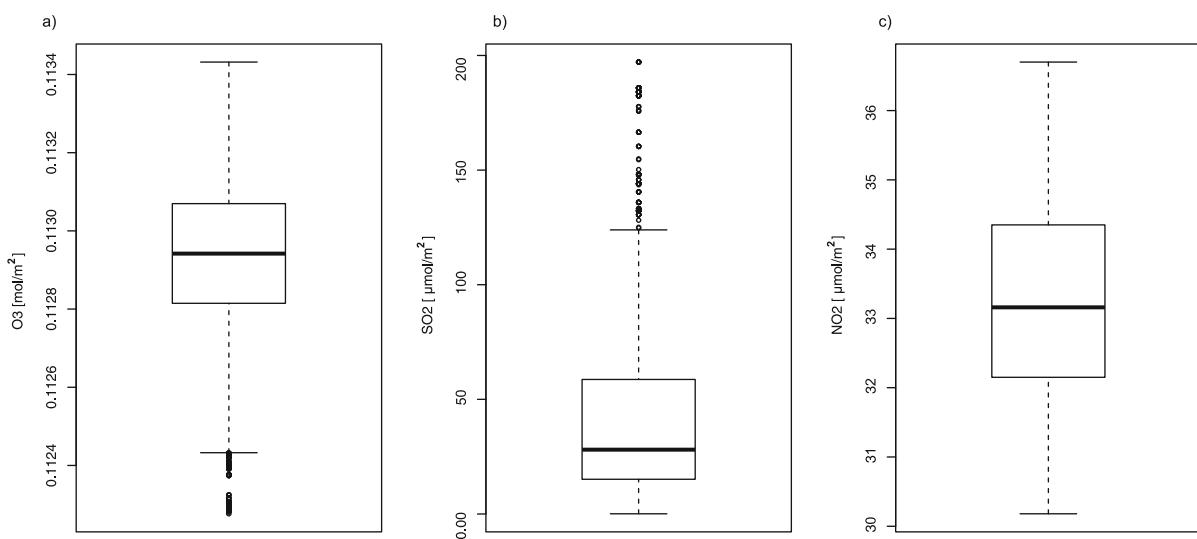


Fig. 2 Box plot representing air pollutant levels during the period of March to May 2019; **a** ozone O_3 ; **b** sulfur dioxide SO_2 ; and **c** nitrogen dioxide NO_2

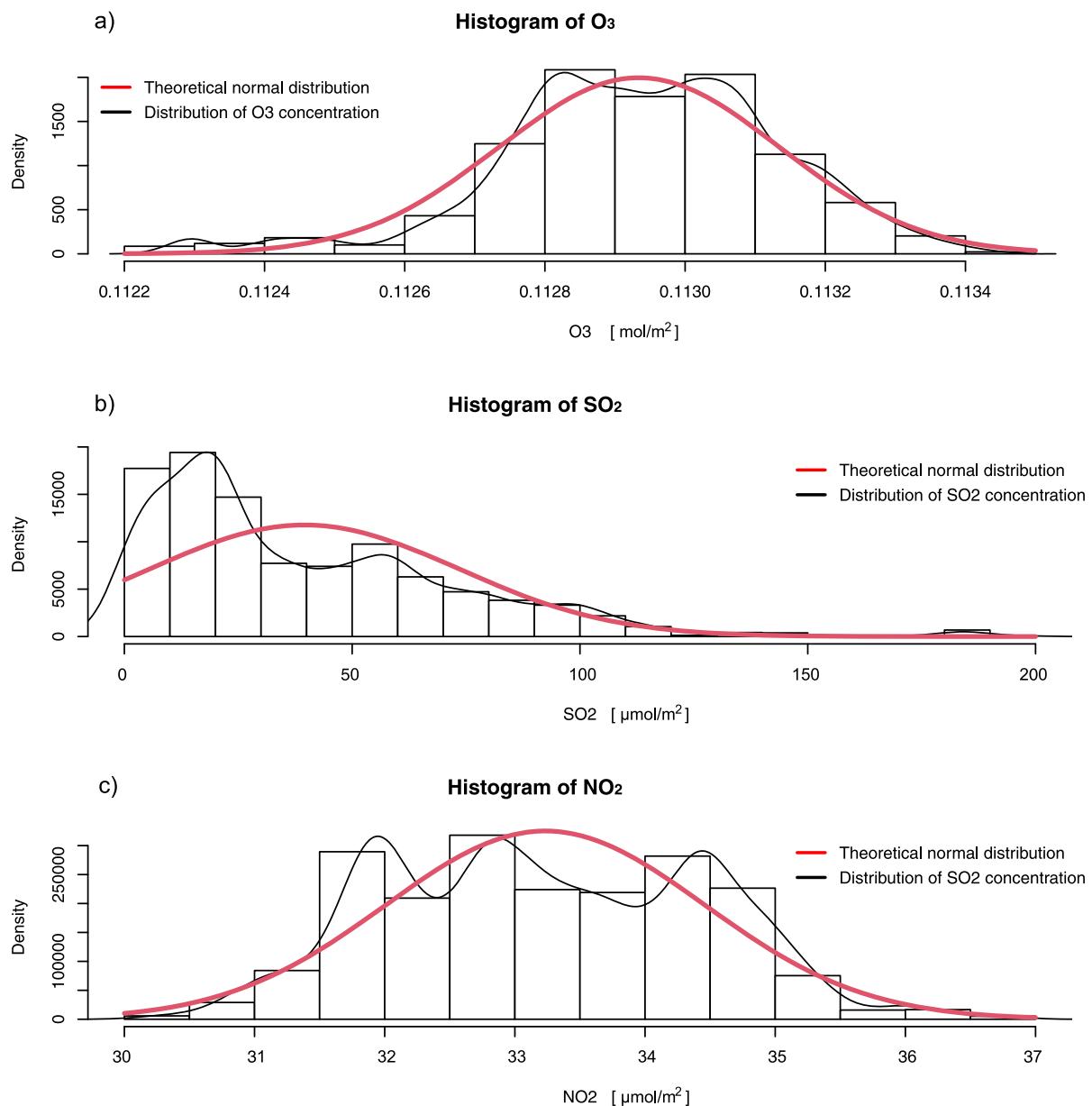


Fig. 3 Histograms of air pollutants with density and normal distribution curves; **a** ozone O_3 ; **b** sulfur dioxide SO_2 ; and **c** nitrogen dioxide NO_2)

34.4 $\mu\text{mol}/\text{m}^2$. The histogram indicates a multi-modal distribution around the mean, differing significantly from a theoretical normal distribution.

Goodness-of-fit tests, including the Shapiro-Wilk, Kolmogorov-Smirnov, Anderson-Darling, and Jarque-Bera tests, were applied to assess normality. For all tests $p > 0.05$, the null hypothesis of normality was rejected, indicating that the distributions of O_3 , SO_2 ,

and NO_2 are asymmetric and significantly different from a normal distribution.

3.1.2 Analysis of spatiotemporal patterns

Monitoring and analyzing the spatiotemporal patterns of air pollutants can provide valuable information for formulating effective pollution control strategies. In

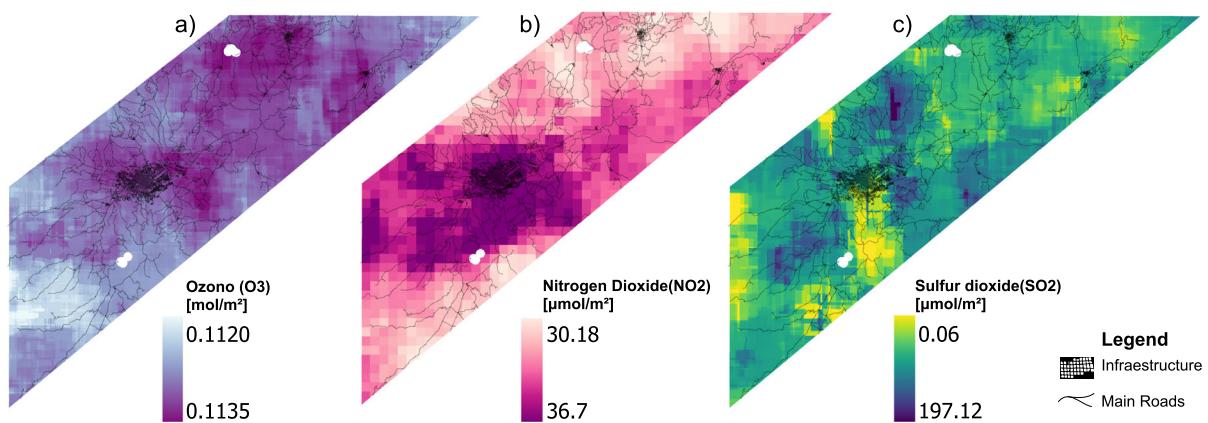


Fig. 4 Air pollutants spatial distribution 2019 period; **a** ozone; **b** nitrogen dioxide; and **c** sulfur dioxide

this section, we analyze the spatiotemporal patterns of the air pollutants ozone, sulfur dioxide, and nitrogen dioxide over the study area. By comparing the data from 2019 to 2022, we aim to understand the changes in air quality and pollution patterns over time. Some spatiotemporal statistical methods and Geographic Information Systems (GIS) techniques were employed to identify trends, hotspots, and dispersion patterns.

Figures 4 and 5 show three different maps of the spatio-temporal distribution of pollutants O_3 , SO_2 , and NO_2 for periods 2019 and 2022; the region that is shown in darker color represents a higher concentration of the pollutant in that area. The ozone concentrations can vary according to season and weather conditions. In general, ozone levels tend to be higher during the drier and warmer months because photochemical reac-

tions that generate ozone are favored by intense solar radiation and higher temperatures (Monks et al., 2015).

On the other hand, during the wet and cold months, atmospheric conditions are not as conducive to ozone formation. Additionally, rainfall can help remove some ozone precursors, such as nitrogen oxides (NOx) and volatile organic compounds (VOCs), by transporting them to the ground (Clapp & Jenkin, 2001). This behavior can explain why the average concentration of O_3 decreased slightly between 2019 and 2022 (Table 3), considering that the study corresponds to the humidity period of the region. For instance, In May 2019, Azuay province recorded 77.2 mm of rainfall, while Cañar province had 45.9 mm, supporting wetter conditions that likely reduced ozone levels. In contrast, May 2022 saw lower rainfall in both regions, with Azuay receiving 72.2 mm and Cañar 26.7 mm, which contributed

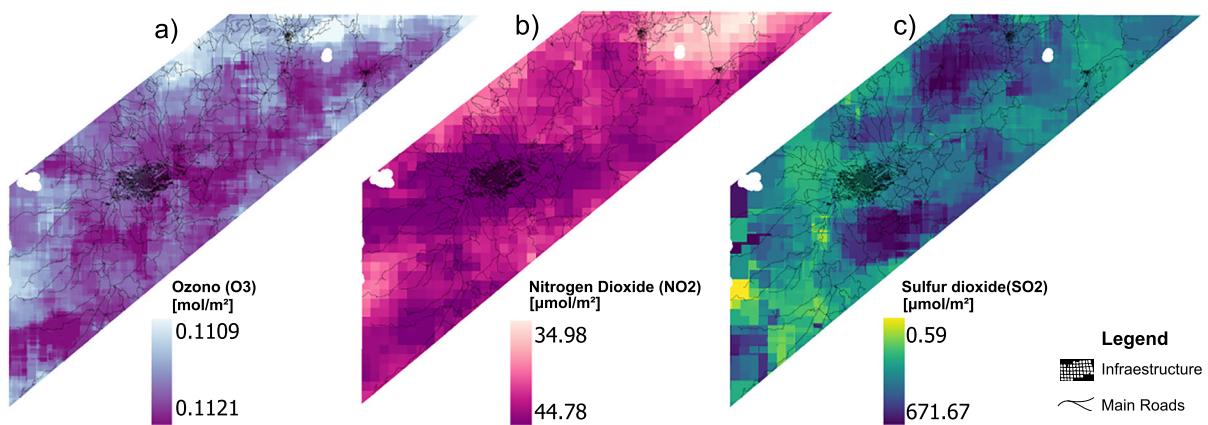


Fig. 5 Air pollutants spatial distribution 2022 period; **a** ozone; **b** nitrogen dioxide; and **c** sulfur dioxide

Table 3 Variation percent of the average values of air pollutants concentration for 2019 and 2022 period

Air pollutants	2019 [mol/m ²]	2022 [mol/m ²]	Variation [%]
Ozone (O_3)	0.112935	0.111630	-1.16
Sulfur dioxide (SO_2)	0.000040	0.000142	258.93
Nitrogen dioxide (NO_2)	0.000033	0.000041	22.62

to higher ozone concentrations due to reduced atmospheric cleansing (Instituto Nacional de Meteorología e Hidrología (INAMHI), 2019, 2022).

Slightly high concentrations of O_3 were observed in urban areas, particularly in Cuenca city, due to increased traffic and industrial emissions. NO_2 concentrations were highest in urban areas and along major transportation corridors. Rural areas, such as Paute and Azogues, also exhibited a few high levels of O_3 . Figure 4 identified NO_2 critical points near the main highways, industrial areas, and densely populated urban centers in 2019.

It is important to note that trends in air pollutant concentrations may vary between regions and urban-rural areas. Unfavorable weather conditions could cause an accumulation of pollutants such as SO_2 and NO_2 in certain areas. Therefore, analyzing the specific sources of emissions and pollution control policies in the study area is critical for an understanding of concentration changes. The increases in SO_2 and NO_2 , seen in Table 3, can be attributed to various factors, such as economic and industrial growth; in this case, the region under study depends to a large extent on the burning of fossil fuels, which generates more pollutant concentration. Furthermore, these settlements have shown an increment in energy demand due to the population increase. Also, the growth of vehicular traffic, added to the lack of efficient public transport systems, can increase NO_2 emissions. On the other hand, variations in weather conditions, such as temperature, wind speed, direction, and humidity, can affect the dispersion of pollutants in the atmosphere.

Cuenca City is a particular case of study, which has had improvements in public transportation, and the dispersion patterns of the three pollutants showed slight and strong changes between 2019 and 2022. The improvement in air quality can be attributed to several factors, including emission control policies, stricter regulations, cleaner energy sources, and better public transportation systems. According to Parra and Espinoza (2020), in his study highlights the benefits

that Cuenca City has initiated the operation of an electric tram system, which is accompanied by the introduction of new urban bus routes. These changes alter the spatial distribution of the city's emissions and reduce the historic center's environmental impact. The electric tram helps to reduce air pollutant emissions along its path, encompassing the historic center. The relocated buses will transfer their emissions to the newly established routes. Therefore, it is important to do a more in-depth analysis, considering a longer study period and the climatic seasons, to know the pollutant distribution along the area.

3.2 Analysis of relationship between land cover classification and air quality

This section illustrates the results obtained after comparing the five land cover classes and the three air pollutants, O_3 , NO_2 , and SO_2 .

3.2.1 Air pollutants dynamics based on LULC classes

This section analyzes the correlation between air pollutants and land cover classification to assess any relationship that might be present. Three independent datasets (O_3 , SO_2 , NO_2) are constructed for each of the five LULC classes under analysis. Figure 6 shows the box plot comparing all five LULC classes with a specific air pollutant.

One of the important characteristics is to assume equal variances across groups. As we can see in Fig. 6, the variance does not differ significantly between classes, except for the water variance with not enough data for O_3 .

The analysis of O_3 can be explained based on Fig. 7, where we can see the distribution function of the five LULC classes. All classes are asymmetric (negative asymmetric), which leads us to consider the dataset suitable for statistical analysis using the Kruskal-Wallis

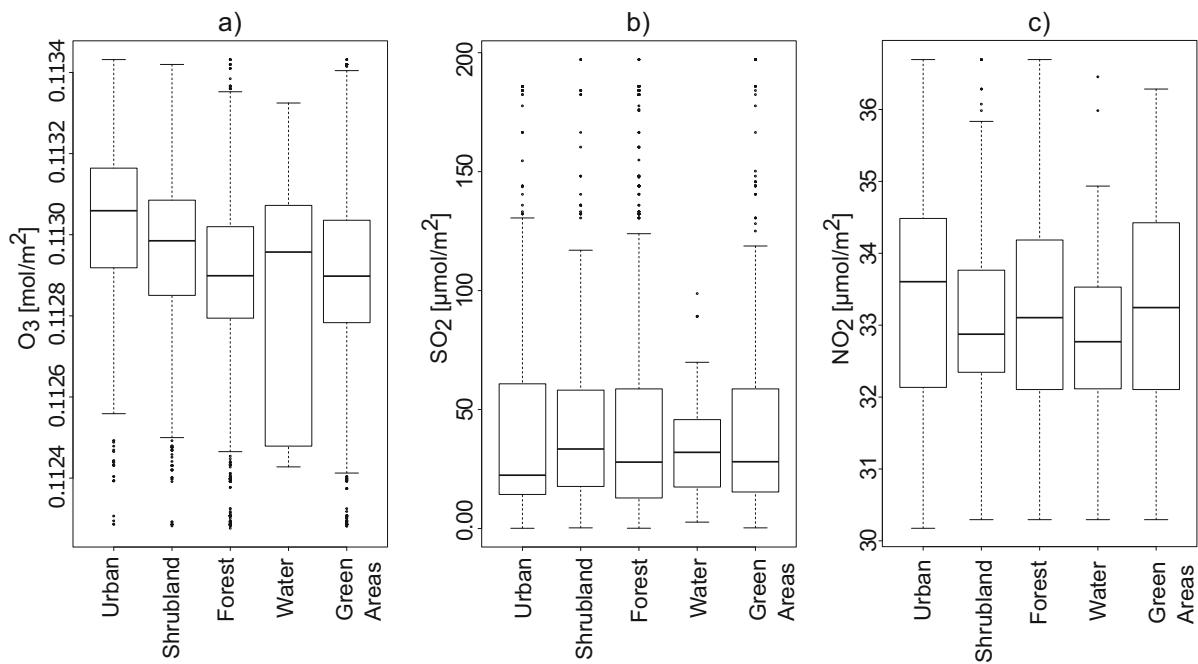


Fig. 6 Box plot of land cover classes compared to **a** O_3 ; **b** SO_2 ; and **c** NO_2 for 2019 period

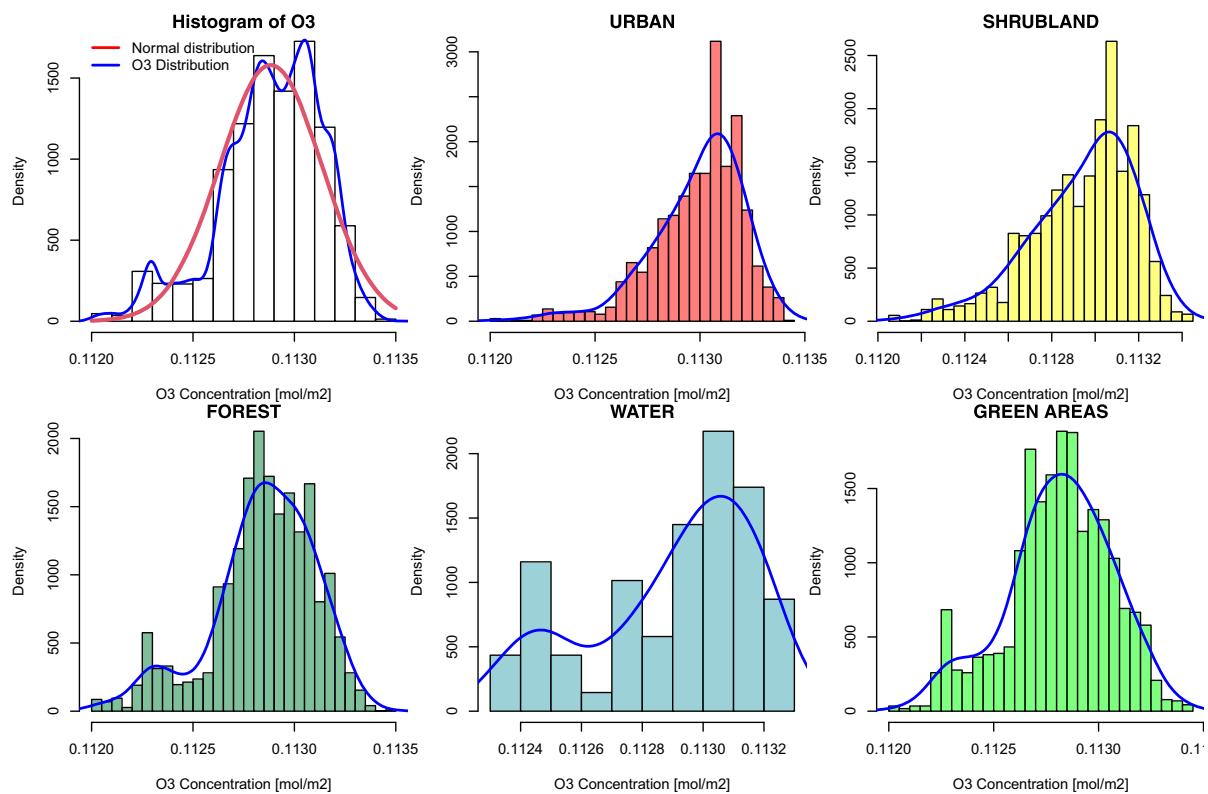


Fig. 7 Distribution function of ozone O_3 by each land cover class for 2019 period

Table 4 Differences among land cover classes for O_3 (p -values, $p < 0.05$ Kruskal-Wallis test with Bonferroni Correction)

O_3	Urban	Forest	Shrubland	Water
Green areas	< 0.0001	< 0.0001	< 0.0001	0.07524
Urban		< 0.0001	< 0.0001	0.01598
Forest			< 0.0001	0.67
Shrubland				1

test. The test is justified because we can clearly see that (1) the independent variable is categorical (more than three categories); (2) the observation in each group of land cover classes is independent; (3) the data do not follow the normal distribution for the complete dataset O_3 (first plot in Fig. 7) and all five independent distribution classes; and (4) The variance between classes is not extremely heterogeneous. The same analysis is done for NO_2 and SO_2 , as shown in Fig. 3; both air pollutants are positively asymmetric, and also, their classes are expected to follow the same distribution form. With the same criteria made for O_3 , these two air pollutants are also suitable for applying a Kruskal-Wallis analysis.

3.2.2 Non-parametric statistical analysis

Tables 4, 5, and 6 show the Kruskal-Wallis test results comparing the distribution of O_3 , SO_2 , and NO_2 across the five land cover types. For O_3 , the p -value between classes is shown to be very weak, which means that the null hypothesis is rejected. For that, we can say that each land cover class is significantly different from the others. Furthermore, we can analyze Fig. 8, which displays the scattered points for each class distributed and grouped in an elliptical shape. The water analysis in this study can be set aside from the analysis of other classes. Firstly, the scattered points for water are fewer compared to other classes, indicating that the study area contains limited water bodies and surfaces. Another notable characteristic is that the water class, when compared to green areas, forests, and shrubland,

shows a $p > 0.05$, indicating no significant difference between water and these natural classes. This result can be explained by the misclassification of riverbanks as forest or natural areas.

On the other hand, we can compare one pair of these classes, forest and green areas, that seem to share certain properties. The ellipses for these classes (Fig. 8) are around $> 90\%$ covered by each other, which is logical because these classes have a direct relationship with the NDVI; however, O_3 seems to be able to separate these two classes as $p < 0.05$ meaning a rejected null hypothesis making these two classes statistical distinguishable.

In the case of SO_2 , we can see that Kruskal-Wallis analysis shows that the null hypothesis can be accepted; the conclusion is that almost all classes are not significantly different. None of the natural classes (shrubland, forest and green areas) can be differentiated. In the Fig. 8, we can see that the forest class and the green areas overlap each other almost 100%. The figure also shows that urban areas are slightly differentiated from others, and water pixels arbitrarily appear in all other classes, making $p > 0.05$. Only the urban class significantly differs from shrubland and green areas with $p = 0.0127$ and $p = 0.0009$, respectively.

Table 6, shows the results of the Kruskal-Wallis analysis in NO_2 . For NO_2 , it is also shown that for the pair urban-green areas, the p -value accepts the null hypothesis $p > 0.05$, it can not be clearly represented in the Fig. 8; however, this result shows that many pixels of urban areas share features with green areas which can

Table 5 Differences among the land cover classes for SO_2 (p -values, $p < 0.05$ Kruskal-Wallis test with Bonferroni Correction)

SO_2	Urban	Forest	Shrubland	Water
Green areas	0.0127	1	1	1
Urban		0.1595	0.0009	1
Forest			0.3002	1
Shrubland				1

Table 6 Differences among the land cover classes for NO_2 (p -values, $p < 0.05$ Kruskal-Wallis test with Bonferroni Correction)

NO_2	Urban	Forest	Shrubland	Water
Green areas	0.0923	0.0006	< 0.0001	0.7753
Urban		< 0.0001	< 0.0001	0.2026
Forest			0.0006	1
Shrubland				1

decrease the chance to separate both classes without errors. The NO_2 figure also shows small differences between the ellipses of natural classes (forest, green areas, and shrubland), and it is validated by Kruskal-Wallis, which demonstrated that these three classes are statistically distinguishable, so they could be distinguished from each other using NO_2 correlated to land cover classes.

Another analysis was done by taking only the shrubland class. This class is a surface containing small scrubs, brushes, and even soil with few herbs. This surface can either occur naturally or be the result of human activity. For that, we can see that the shrubland ellipse is between the natural classes (forest and green areas) and the artificial class (urban). Urban type is a class that stays different from the others in the three gases O_3 , NO_2 , and SO_2 ; in some way, urban is related to shrubland, but it has around 60% of the data points out of the rest of the groups.

4 Discussion

Numerous studies have explored the connections between land cover types and air pollutant levels, concentrat-

ing on the influence of vegetation, urbanization, and industrialization on the spatial dispersion of contaminants (Nowak et al., 2006; Superczynski & Christopher, 2011; Weng & Yang, 2006; Zhang et al., 2022). It has been found that vegetation can help mitigate air pollution by absorbing and filtering pollutants, while urban and industrial areas can contribute to higher pollutant levels (Bechle et al., 2011; Escobedo et al., 2011). In addition, land cover changes can influence the dispersion and deposition of pollutants, affecting their concentrations and distribution patterns (Knibbs et al., 2014). However, South America has seen limited research into how LULC types influence air pollution levels and their dispersion. This research gap is compounded by the lower prioritization of air pollution management policies in many South American countries (UNEP, 2018).

This study shows that vegetation land cover classes such as forest, shrubland, and green areas are statistically different when their spatio-temporal distribution is correlated with the spatio-temporal distribution of ozone. The study does not show which of these classes have the main contribution of O_3 concentration. However, rejecting the null hypothesis is the same as saying that these four classes, including urban areas, are distin-

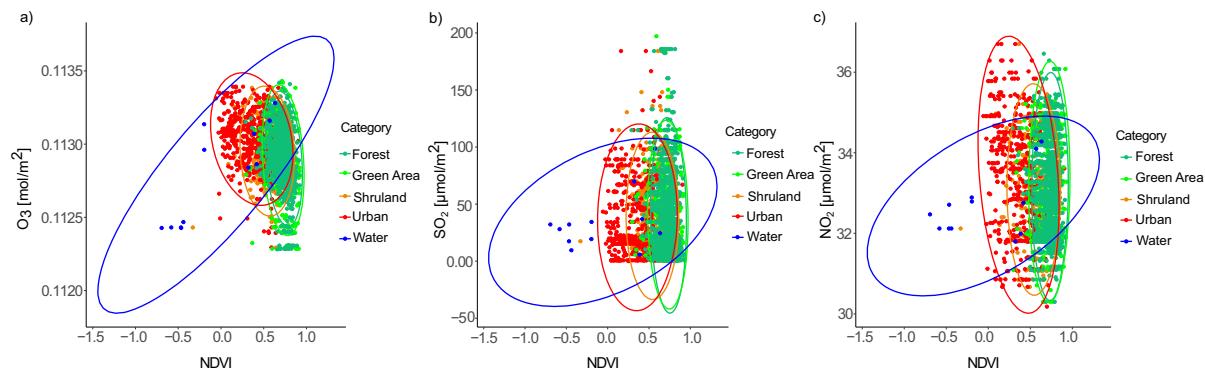


Fig. 8 Scatter plot of pairs (gaseous air pollutant - spectral index); **a** O_3 -NDVI; **b** SO_2 -NDVI; and **c** NO_2 -NDVI

guishable on an analysis; thus, these can directly affect O_3 distribution around the land cover areas. Other research has also shown the same and concluded that remote sensing data can be used to monitor changes in land cover and their relationship with gaseous air pollutants (Larkin et al., 2017). O_3 concentrations can be affected by changes in land cover, particularly changes in vegetation cover (Li et al., 2023; Monks et al., 2015).

NO_2 and SO_2 are other gaseous air pollutants that are affected by land cover, particularly an increase in impervious surfaces, which can contribute to increased concentration levels (Bechle et al., 2011; Escobedo et al., 2011). Kaplan and Avdan (2020) used Sentinel-5P data to analyze air pollutant distribution and its relationship with demographic and geographic factors in Paris, Istanbul, and St. Petersburg. Their results showed significant spatial variations in pollutant concentrations, with higher levels in urban and industrial areas. A positive correlation between population density and NO_2 concentrations suggests that increased human activity contributes to pollution. Our study considers urban land use class as the main source of industrial activities and human-related impact. The Kruskal-Wallis analysis shows that for SO_2 , urban areas are statistically different from green areas and shrublands. For NO_2 , it is also shown that urban classes are statistically distinguishable from forest and shrubland. This suggests that urban expansion can be a significant estimator or explainer in the distribution of NO_2 and SO_2 .

This research also highlights the importance of considering demographic and geographic factors when addressing air pollution, demonstrating the potential of Sentinel-5P data for monitoring pollution. Furthermore, this research contributes to filling the gap in studies from some regions of Ecuador, particularly the Andean ecosystems, and provides valuable data for future environmental management and policy-making.

As mentioned before, we do not use a separate land cover class for industrial zones; we only use urban areas that encompass all impervious surfaces. NO_2 and SO_2 are primarily emitted from industrial sources such as power plants and refineries (U.S. Environmental Protection Agency (EPA), 2021). Our analysis does not conclude a direct relationship between certain vegetation classes and urban areas over SO_2 and NO_2 , as it does for O_3 . Future research can include more specific data about land cover for urban areas and redo this process to see if different urban classes reflect an effect on

the distribution of NO_2 and SO_2 across various types of urban areas.

5 Conclusion

The study aimed to assess the significance of accurate land cover classification for understanding its potential relationship with O_3 , NO_2 , and SO_2 gaseous air pollutants' spatial distribution and dynamics.

Analyzing the correlation of gaseous pollutants and land cover approach represented by NDVI index. We can conclude that SO_2 is not suitable for assessing its behavior based on LULC classes. Almost all pairs for the LULC classes accept the null hypothesis, making it impossible to distinguish classes based only on this pollutant. We concluded that land cover changes are not significantly associated with variations in SO_2 concentrations, suggesting that other factors may influence its distribution. For NO_2 , the analysis presents a moderate distinction of LULC classes based on the distribution of this pollutant. For O_3 , the conclusion is clear, and all classes reject the null hypothesis, making LULC classes significantly different from each other; the only exception is water, which is not well represented because the study area contains few water bodies.

The study identified correlations between specific land cover types and air pollutant concentrations. Urban areas were found to have the highest levels of NO_2 and SO_2 , primarily due to traffic emissions and industrial activities. In contrast, forested regions exhibited lower concentrations of these pollutants and higher levels of O_3 , which can be attributed to the natural production of O_3 through the interaction of sunlight with volatile organic compounds emitted by vegetation. These findings underscore the importance of accurate land cover classification for air quality assessment and management.

The present work has shown the importance of accurate land cover classification for air quality assessment and management, providing valuable insights for policymakers and stakeholders to implement targeted measures for reducing air pollution and mitigating its adverse effects on public health and the environment. Future research could focus on expanding the study area, incorporating more recent data, and exploring the potential of other machine learning techniques for land cover classification. Additionally, further investigation

into the impact of meteorological factors on the relationship between land cover and air pollutant concentrations could provide a more comprehensive understanding of air pollution dynamics.

This study's analysis can be improved by considering more variables, such as wind direction, temperature, elevation, slope, and solar radiation, to better understand the relationship between soil covers and air pollutants. The results of such an analysis depend on factors like data quality and quantity, study area type, data sources, classification methods, spatial and temporal scales, meteorological conditions, emission sources, topography, and vegetation type and density. Furthermore, incorporating socioeconomic factors, population density, urbanization, and industrial activities can enhance the research by accounting for their impact on land cover patterns and air pollutant emissions.

Author contributions PG conducted the main design of the study, analysis, methodology, and results; produced the figures and tables; wrote the main manuscript; and performed all revisions, modifications, and reviews of the manuscript. SS and DM assisted in the methodology design and reviewed the manuscript. EU provided extensive support throughout the manuscript development, including reviewing and tracking changes, providing feedback on the structure and clarity, and advising on responses to reviewer comments. All authors have read and agreed to the final version of the manuscript.

Funding This work was funded by the Estonian Research Agency (grant number PRG1764), Estonian Ministry of Education and Research, Centre of Excellence for Sustainable Land Use (TK232). S. Szabo was supported by the K138079 NKFH project.

Data Availability The code for the study, “How does air quality reflect land cover changes: A remote sensing approach using Sentinel data,” is available on GitHub to support reproducibility and further research. The repository includes scripts for data processing, land use and land cover (LULC) classification, and air quality analysis. Access the repository at github.com/pamelaguamamp/lulc-air-quality-sentinel.

Declarations

Ethics Approval All authors have read, understood, and have complied as applicable with the statement on “Ethical responsibilities of Authors” as found in the Instructions for Authors.

Conflict of Interest The authors declare no competing interests. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abida, K., Barbouchi, M., Boudabbous, K., et al. (2022). Sentinel-2 data for land use mapping: Comparing different supervised classifications in semi-arid areas. *Agriculture, 12*(9), 1429.
- Ali, E. B., Shayamnahr, S., Radmehr, R., et al. (2023). Exploring the impact of economic growth on environmental pollution in South American countries: How does renewable energy and globalization matter? *Environmental Science and Pollution Research, 30*(6), 15505–15522.
- Basheer, S., Wang, X., Farooque, A. A., et al. (2022). Comparison of land use land cover classifiers using different satellite imagery and machine learning techniques. *Remote Sensing, 14*(19), 4978.
- Bechle, M. J., Millet, D. B., & Marshall, J. D. (2011). Effects of income and urban form on urban no₂: Global evidence from satellites. *Environmental science & technology, 45*(11), 4914–4919.
- Bell, M. L., Davis, D. L., Gouveia, N., et al. (2006). The avoidable health effects of air pollution in three Latin American cities: Santiago, Sao Paulo, and Mexico City. *Environmental Research, 100*(3), 431–440.
- Bonilla-Bedoya, S., Zalakeviciute, R., Coronel, D. M., et al. (2021). Spatiotemporal variation of forest cover and its relation to air quality in urban Andean socio-ecological systems. *Urban Forestry & Urban Greening, 59*,
- C40-Cities (2022) C40 cities technical assistance report - Quito. Tech. rep., C40 Climate Leadership Group
- Clapp, L. J., & Jenkin, M. E. (2001). Analysis of the relationship between ambient levels of O₃, NO₂ and NO as a function of NO_x in the UK. *Atmospheric Environment, 35*(36), 6391–6405.
- Cohen, A. J., Brauer, M., Burnett, R., et al. (2017). Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: An analysis of data from the global burden of diseases study 2015. *The lancet, 389*(10082), 1907–1918.
- da Silva, V. S., Salami, G., da Silva, M. I. O., et al. (2020). Methodological evaluation of vegetation indexes in land use and land cover (LULC) classification. *Geology, Ecology, and Landscapes, 4*(2), 159–169.
- Dore, A., Carslaw, D., Braban, C., et al. (2015). Evaluation of the performance of different atmospheric chemical transport models and inter-comparison of nitrogen and sulphur deposition estimates for the UK. *Atmospheric Environment, 119*, 131–143.
- Empresa Pública Municipal de Movilidad, Tránsito y Transporte de Cuenca - EMOV EP (2022) Informe de Calidad del Aire de Cuenca 2022. URL https://ierse.uazuay.edu.ec/proyectos/links_doc_contaminantes/Informes-Calidad-Aire/Informe_Calidad_Aire_Cuenca_2022.pdf. Accessed:2024-07-12
- Escobedo, F. J., Kroeger, T., & Wagner, J. E. (2011). Urban forests and pollution mitigation: Analyzing ecosystem services and disservices. *Environmental Pollution, 159*(8–9), 2078–2087.
- Feng, S., Li, W., Xu, J., et al. (2022). Land use/land cover mapping based on gree for the monitoring of changes in ecosystems

- tem types in the upper yellow river basin over the Tibetan plateau. *Remote Sensing*, 14(21), 5361.
- Frimpong, D. B., Twumasi, Y. A., Ning, Z. H., et al. (2022). Assessing the impact of land use and land cover change on air quality in east baton rouge—Louisiana using earth observation techniques. *Advances in Remote Sensing*, 11(3), 106.
- Hawash, E., El-Hassanin, A., Amer, W., et al. (2021). Change detection and urban expansion of Port Sudan, Red Sea, using remote sensing and GIS. *Environmental Monitoring and Assessment*, 193, 1–22.
- Hothorn, T., Hornik, K., Van De Wiel, M. A., et al. (2006). A Lego system for conditional inference. *The American Statistician*, 60(3), 257–263.
- Huang Y, Lee CK, Yam YS, et al (2022) Rapid detection of high-emitting vehicles by on-road remote sensing technology improves urban air quality. *Science Advances* 8(5):eabl7575
- Ialongo, I., Virta, H., Eskes, H., et al. (2020). Comparison of TROPOMI/Sentinel-5 Precursor NO₂ observations with ground-based measurements in Helsinki. *Atmospheric Measurement Techniques*, 13(1), 205–218.
- Instituto Nacional de Meteorología e Hidrología (INAMHI) (2019) Boletín Agroclimático Nacional - Mayo 2019. URL <http://sipa.agricultura.gob.ec>. Accessed: 28 October 2024
- Instituto Nacional de Meteorología e Hidrología (INAMHI) (2022) Boletín Agroclimático Nacional - Mayo 2022. URL <http://sipa.agricultura.gob.ec>, accessed: 28 October 2024
- Jiang, Z., Cheng, H., Zhang, P., et al. (2021). Influence of urban morphological parameters on the distribution and diffusion of air pollutants: A case study in China. *Journal of Environmental Sciences*, 105, 163–172.
- Jodhani, K. H., Gupta, N., Parmar, A. D., et al. (2024). Synergizing google earth engine and earth observations for potential impact of land use/land cover on air quality. *Results in Engineering*, 22, 102039.
- Kaplan, G., & Avdan, Z. Y. (2020). Space-borne air pollution observation from Sentinel-5P TROPOMI: Relationship between pollutants, geographical and demographic data. *International Journal of Engineering and Geosciences*, 5(3), 130–137.
- Knibbs, L. D., Hewson, M. G., Bechle, M. J., et al. (2014). A national satellite-based land-use regression model for air pollution exposure assessment in Australia. *Environmental Research*, 135, 204–211.
- Larkin, A., Geddes, J. A., Martin, R. V., et al. (2017). Global land use regression model for nitrogen dioxide air pollution. *Environmental science & technology*, 51(12), 6957–6964.
- Li Y, Hong T, Gu Y, et al (2023) Assessing the spatiotemporal characteristics, factor importance, and health impacts of air pollution in Seoul by integrating machine learning into land-use regression modeling at high spatiotemporal resolutions. *Environmental Science & Technology*
- Lucila, R. A., Edit, I. V., Fernando, A., et al. (2016). Lower respiratory tract infection in children younger than five years of age and adverse pregnancy outcomes related to household air pollution in Bariloche (Argentina) and Temuco (Chile). *Indoor Air*, 26(6), 964.
- Luo J, Du P, Samat A, et al (2017) Spatiotemporal pattern of pm_{2.5} concentrations in mainland China and analysis of its influencing factors using geographically weighted regression. *Scientific reports* 7(1):40607
- Manosalidis I, Stavropoulou E, Stavropoulos A, et al (2020) Environmental and health impacts of air pollution: A review. *Frontiers in public health* p 14
- Monks, P. S., Archibald, A., Colette, A., et al. (2015). Tropospheric ozone and its precursors from the urban to the global scale from air quality to short-lived climate forcer. *Atmospheric Chemistry and Physics*, 15(15), 8889–8973.
- National Institute of Statistics and Census of Ecuador. (2022). Population data analytics. Retrieved from <https://censoecuador.ecudatanalytics.com/>
- Nowak, D. J., Crane, D. E., & Stevens, J. C. (2006). Air pollution removal by urban trees and shrubs in the united states. *Urban forestry & urban greening*, 4(3–4), 115–123.
- Page, R., Lavender, S., Thomas, D., et al. (2020). Identification of tyre and plastic waste from combined Copernicus Sentinel-1 and-2 data. *Remote Sensing*, 12(17), 2824.
- Parra, R., & Espinoza, C. (2020). Insights for air quality management from modeling and record studies in Cuenca, Ecuador. *Atmosphere*, 11(9), 998.
- Rojas-Rodríguez, H., da Silva, A. S., Texcalac-Sangrador, J. L., et al. (2016). Air pollution management and control in Latin America and the Caribbean: Implications for climate change. *Revista Panamericana de Salud Pública*, 40, 150–159.
- Salvi, S. (2007). Health effects of ambient air pollution in children. *Paediatric Respiratory Reviews*, 8(4), 275–280.
- Sapkota, P., & Bastola, U. (2017). Foreign direct investment, income, and environmental pollution in developing countries: Panel data analysis of Latin America. *Energy Economics*, 64, 206–212.
- Superczynski, S. D., & Christopher, S. A. (2011). Exploring land use and land cover effects on air quality in central Alabama using GIS and remote sensing. *Remote sensing*, 3(12), 2552–2567.
- Syuhada, G., Akbar, A., Hardiawan, D., et al. (2023). Impacts of air pollution on health and cost of illness in Jakarta, Indonesia. *International Journal of Environmental Research and Public Health*, 20(4), 2916.
- Szabo, S., Gácsi, Z., & Balazs, B. (2016). Specific features of NDVI, NDWI and MNDWI as reflected in land cover categories. *Acta Geographica Debrecina Landscape & Environment Series*, 10(3/4), 194.
- Talukdar, S., Singha, P., Mahato, S., et al. (2020). Land-use land-cover classification by machine learning classifiers for satellite observations—A review. *Remote Sensing*, 12(7), 1135.
- Tobar, V., & Wyseure, G. (2018). Seasonal rainfall patterns classification, relationship to ENSO and rainfall trends in Ecuador. *International Journal of Climatology*, 38(4), 1808–1819.
- UNEP (2018) Efforts to reduce air and climate pollutants in Latin America could reap immediate health benefits. URL <https://www.unep.org/news-and-stories/story/efforts-reduce-air-and-climate-pollutants-latin-america-could-reap-immediate>, retrieved from UNEP
- U.S. Environmental Protection Agency (EPA). (2021). Air quality and emissions trends report. URL <https://www.epa.gov/air-trends>. Retrieved from EPA
- Vadrevu, K., Ohara, T., & Justice, C. (2017). Land cover, land use changes and air pollution in Asia: A synthesis. *Environmental Research Letters*, 12(12)

- Vîrghileanu, M., Săvulescu, I., Mihai, B. A., et al. (2020). Nitrogen dioxide (NO₂) pollution monitoring with Sentinel-5P satellite imagery over Europe during the coronavirus pandemic outbreak. *Remote Sensing*, 12(21), 3575.
- Weng, Q., & Yang, S. (2006). Urban air pollution patterns, land use, and thermal landscape: An examination of the linkage using GIS. *Environmental Monitoring and Assessment*, 117, 463–489.
- WHO WHO. (2022a). Ambient (outdoor) air pollution. URL [https://www.who.int/news-room/fact-sheets/detail/ambient-\(outdoor\)-air-quality-and-health](https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health)
- WHO WHO. (2022b). Household air pollution. URL <https://www.who.int/news-room/fact-sheets/detail/household-air-pollution-and-health>
- Wickham, H. (2009). *ggplot2: Elegant graphics for data analysis*. New York. NY: Springer.
- Zhang, J., Yang, G., Yang, L., et al. (2022). Dynamic monitoring of environmental quality in the loess plateau from 2000 to 2020 using the google earth engine platform and the remote sensing ecological index. *Remote Sensing*, 14(20), 5094.

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.