❏   732

Remote sensing in the analysis between forest cover and COVID-19 cases in Colombia

Vladimir Henao-Céspedes, Yeison Alberto Garcés-Gómez
Faculty of Engineering and Architecture, Universidad Católica de Manizales, Manizales, Colombia

Article Info

ABSTRACT

This article explores the relationship between forest cover and coronavirus disease 2019 (COVID-19) cases in Colombia using remote sensing techniques and data analysis. The study focuses on the CORINE land cover methodology's five main land cover categories: artificial territory, agricultural territories, forests and semi-natural areas, humid areas, and water surfaces. The research methodology involves several phases of the unified method of analytical solutions for data mining (ASUM-DM). Data on COVID-19 cases and forest cover are collected from the Colombian National Institute of Health and Advanced Land Observation Satellite (ALOS PALSAR), respectively. Land cover data is processed using QGIS software. The results indicate an inverse relationship between forest cover and COVID-19 cases, as evidenced by Pearson's index ρ of -0.439 (p-value <0.012). In addition, a negative correlation is observed between case density and forests and semi-natural areas, one of the land cover categories. The findings of this study suggest that higher forest cover is associated with lower numbers of COVID-19 cases in Colombia. The results could potentially inform government organizations and policymakers in implementing strategies and policies for forest conservation and the inclusion of green areas in densely populated urban areas.

Corresponding Author:

Vladimir Henao-Céspedes
Faculty of Engineering and Architecture, Universidad Católica de Manizales
Manizales, Colombia
Email: vhenao@ucm.edu.co

1. INTRODUCTION

In 2019, after several years, the world again experienced a pandemic caused by the coronavirus disease 2019 (COVID-19) disease, caused by the SAR-CoV-2 virus [1]. To mitigate the spread of the virus and reduce the risk of the population suffering from COVID-19, a large portion of the world´s governments implemented mitigation strategies [2], [3]. On the other hand, the research and academic sectors began to develop research to understand the relationships between the spread of the virus and different factors, such as environmental [4]–[6], demographic and economic [7]–[9], age-related, climatological, people with the Bacillus Calmette-Guérin (BGC) vaccine, and people under treatment for malaria [10]. Likewise, different authors have worked on the development of mathematical models that allow modeling and estimating the long-term spread of COVID-19 [11].

Spatial data mining techniques allow the extraction of information from geographic data to perform correlation analysis [12] between different variables, such as: land use and forest fire risk [13], the spatial distribution of forest cover and environmental variables [14], and the quantification of vegetation types using multitemporal analysis [15]. Thanks to the above and considering that, remote sensing, it is possible to carry out vegetation cover analysis [16]–[18], this is another factor studied concerning the spread of COVID-19.

Thus, another factor that has been studied with regard to COVID-19 are the green areas present in cities, in such a way that it can be defined if the existence of these spaces mitigates or not the spread of the virus. Studies focused on this factor have managed to establish that there is a negative correlation between the existence of green spaces and the infection rate, that is, the number of infections is lower in areas where there are more green spaces [19]–[21]. If the percentage of vegetation coverage increases, the spread rate of the virus decreases, according to numerical data. If the coverage increases by 1%, the number of accumulated cases of COVID-19 decreases by 2.6%, according to [22]. The preceding is consistent with studies carried out on the relationship of some respiratory diseases with green areas, such as the case presented in [23], where it is suggested that growing up in urban areas with large green areas reduces the risk of developing asthma thanks to the macrobiota present. Considering the importance of green regions to mitigate the chances of developing some diseases, some authors propose incorporating biophilic elements into urban design, such as green roofs, which could help improve air quality [24].

In this article, using remote sensing data associated with forest cover, the results of a study on its relationship with COVID-19 are presented, applying various phases of the analytics solutions unified method for data mining (ASUM-DM), methodology for data analysis projects. Remote sensing data have allowed modeling the spread of diseases from a spatial and temporal dynamic [7], [25]. Finally, this article analyzes the association between COVID-19 cases recovered, deceased, and forest cover. The research hypothesis that "greater forest cover is associated with a lower number of COVID-19 cases" was verified in a case study in Colombia.

2. METHOD
2.1. Study area
The study area is Colombia, a country comprising 32 departments according to its internal political division. Colombia is located in the northwest of South America. The country has a Gini index of 0.53 [26], indicating a state of neither perfect equality nor perfect inequality. However, based on this, it can be inferred that a significant portion of the population lacks sufficient economic income. Colombia exhibits a diverse range of land cover types across its five regions: the Caribbean region, characterized by grasslands, wetlands, and bare soils; the Pacific region, known for its high rainfall and lush rainforests; the Orinoco region, dominated by vast savannahs on a flat landscape; the Amazon rainforest area, featuring dense tropical jungles; and the Andes Mountain range, comprising heavily modified natural ecosystems, extensive agricultural cultivation, road networks, and urban areas [27].

2.2. Data sources
The data that was used in this research and the respective sources are presented in Table 1. The population data for the study area was obtained from the 2018 population census report carried out by the national administrative department of statistics [28]. This makes it clear that the most recent census in Colombia is the one mentioned. The number of cases of COVID-19 is obtained from the National Institute of Health of Colombia, also known as Instituto Nacional de Salud (INS) by its name in Spanish [29]. The land cover data for Colombia were obtained from the National Institute of Hydrology, Meteorology and Environmental Studies (IDEAM) [30] and geographically processed using QGIS software [31], [32]. The IDEAM employs the CORINE Land Cover methodology [33], [34] to generate the land cover classification. Finally, the forest cover data was obtained from the global forest/non-forest map (FNF) dataset [35] implemented in google earth engine (GEE).

Table 1. Data sources

Data	Acquisition date	Source
Population	01/10/2020 - 31/12/2021	[28]
Covid–19 (amount cases)	01/10/2020 - 31/12/2021	[29]
Forest	01/10/2020 - 31/12/2021	[36]
Land Cover		[30]

2.3. Methodology
To start, the study area is defined, in this case, Colombia, and a time window is established from October 1, 2020 to December 31, 2021, the window for which the primary data of the study is covered. Subsequently, data from the study area on the behavior of the COVID-19 pandemic, in which the number of cases is identified, are obtained from the information system of the National Institute of Health in Colombia. Forest data is also obtained for the study area for the established time window, and this study focused on

analyzing the five main categories of land cover: i) artificialized territory (CLC1), ii) agricultural territories (CLC2), iii) forests and semi-natural areas (CLC3), iv) wet areas (CLC4), and v) water surfaces (CLC5). For the execution of the study, some stages of the analytics solutions unified method for data mining/predictive analytics (ASUM–DM) methodology created by IBM in 2015 [37] were applied in an adjusted way for a better understanding. This can be observed in Figure 1.
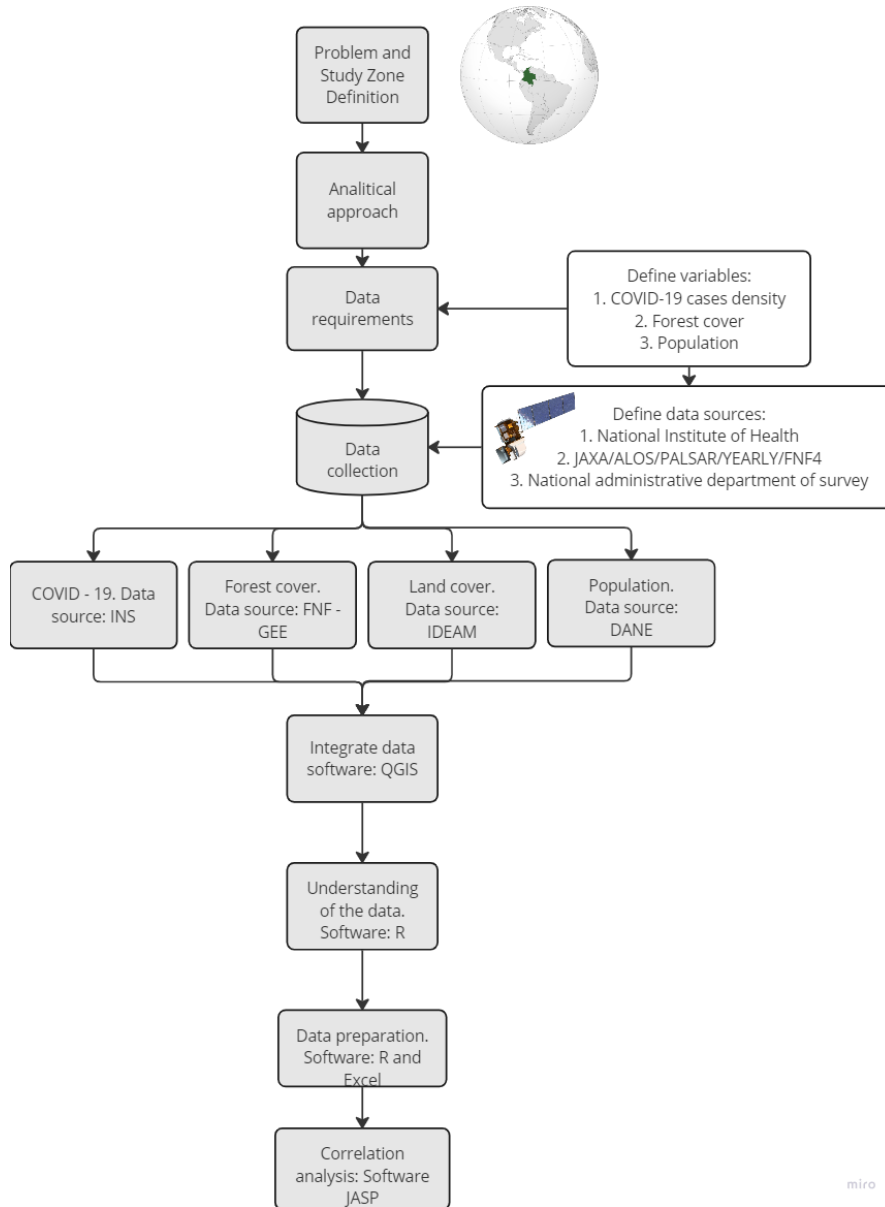

Figure 1. Methodology flowchart

The analytical approach of the problem phase is carried out to identify the variables that are required in the study. Subsequently, the data requirement phase establishes these variables with their respective characteristics. Once the variables necessary for the study with their respective characteristics were defined, the data collection phase was carried out, during which the different sources where the variables were found were accessed. After obtaining the data from the variables, an initial exploratory analysis is carried out in the data comprehension phase, to observe the variables' behavior. Once the behavior of the variables was known, the data preparation phase was executed, where processes such as cleaning and standardization of erroneous and atypical data were carried out. Finally, an analysis of the data is generated in the phase of the construction of the model-analysis.

3. RESULTS AND DISCUSSION

The analytical approach to the problem suggests the need to understand the relationship between the variable "case density" and the variables "forest density" and "land cover" in order to identify the impact of forest coverage on the number of COVID-19 infections. Accordingly, the COVID-19 dataset was filtered and consolidated, specifically by departments, using the R software. The land cover data were processed in QGIS using the "Join attributes by location" data management tool with the geometric predicate "contains." This methodology allows for associating each geographical unit with the corresponding land cover categories within its boundaries. Subsequently, the land cover data were merged into the preexisting COVID database to create an integrated dataset. Following the data cleaning process performed with R and Excel, the data comprehension phase commenced, including an exploratory analysis that revealed the findings in Table 2. As observed in the table as mentioned earlier, the Shapiro-Wilk test indicates that the analyzed dataset is not normally distributed, as all the p-values from the Shapiro-Wilk test are less than 0.05. However, the forest density among the variables shows the closest approximation to a normal distribution, with a Shapiro-Wilk p-value of 0.102.

Table 2. Descriptive statistic

	Cases density	Forest density	CLC1	CLC2	CLC3	CLC4	CLC5
Valid	32	32	32	32	32	32	32
Missing	0	0	0	0	0	0	0
Mean	0.076	0.511	18994.868	828732.63	1.27E+06	27878.697	20090.187
Std. Deviation	0.112	0.277	15999.318	681897.086	1.60E+06	41316.012	26982.455
Shapiro-Wilk	0.692	0.945	0.826	0.906	0.67	0.713	0.690
P-value of Shapiro-Wilk	< .001	0.102	< .001	0.009	< .001	< .001	< .001
Minimum	3.05E-04	1.92E-04	956.148	1398.354	1041.032	0	0
Maximum	0.432	0.969	61877.915	2.90E+06	7.80E+06	182599.677	129055.454

Given the purpose of this study and considering the rejection of the normality assumption of the data, Pearson's correlation coefficient was applied, obtaining the values indicated in Table 3. As can be seen, there is a Pearson index ρ of -0.439 (p-value<0.012) between the variables forest density and cases density. On the other hand, between the variables case density and CLC3, a Pearson index ρ of -0.383 (p-value<0.031) was obtained.

Table 3. Pearson's Correlations of the variables case density, forest density, and the 5 cover categories

Variable		Cases density	Forest density	CLC1	CLC2	CLC3	CLC4	CLC5
1. Cases density	Pearson's r	—						
	p-value	—						
2. Forest density	Pearson's r	-0.439*	—					
	p-value	0.012	—					
3. CLC1	Pearson's r	-0.091	-0.146	—				
	p-value	0.621	0.425	—				
4. CLC2	Pearson's r	-0.253	-0.135	0.251	—			
	p-value	0.162	0.461	0.165	—			
5. CLC3	Pearson's r	-0.383*	0.109	-0.22	0.266	—		
	p-value	0.031	0.554	0.226	0.142	—		
6. CLC4	Pearson's r	-0.264	-0.347	0.142	0.288	0.068	—	
	p-value	0.145	0.052	0.438	0.109	0.713	—	
7. CLC5	Pearson's r	-0.252	-0.2	0.139	0.316	0.006	0.900***	—
	p-value	0.164	0.273	0.448	0.079	0.976	< .001	—

* p < .05, ** p < .01, *** p < .001

Results in Table 3 indicate an inverse trend between forest density and the variables cases density and CLC3, with departments with high forest cover presenting a low number of COVID-19 cases, thus validating the hypothesis raised for the development of this article. This is consistent with what was proposed in [19], [20] where green areas are associated with the low severity of COVID-19 infections and the low transmission rate, thanks to optimum air quality. Considering the correlation obtained for the CLC3 variables, an analysis was conducted that encompassed the vegetation covers within category 3. These covers include, first, those of a forest type (code: 31); second, those of a shrubby and herbaceous type (code: 32); and third, territories composed of bare soils, rocky outcrops, and sandy areas (code: 33). The analysis reveals a significant correlation (ρ=-0.368, p-value=0.038) between the density of COVID-19 cases and coverage 31, which corresponds to forested areas as shown in Table 4. This correlation value aligns with the findings of

the previous analysis conducted using forest density data obtained from the FNF map, further confirming the inverse relationship between the number of COVID-19 cases and the extent of forest presence in the study areas.

Table 4. Pearson's Correlations of the case density variables and the three categories of the CLC3 level of coverage

Variable		Cases density	31 - Forest	32 - Areas with herbaceous and/or shrubby vegetation	33 - Open areas, with little or no vegetation
Cases density	Pearson's r	—			
	p-value	—			
31 - Forest	Pearson's r	-0.368*	—		
	p-value	0.038	—		
32 - Areas with herbaceous and/or shrubby vegetation	Pearson's r	-0.328	0.68***	—	
	p-value	0.067	< .001	—	
33 - Open areas, with little or no vegetation	Pearson's r	-0.232	0.194	0.419*	—
	p-value	0.201	0.279	0.015	—
* $p < .05$, ** $p < .01$, *** $p < .001$					

The correlation between the cases density of COVID-19 and the CLC3 coverage is visually presented in Figure 2 and Figure 3. The graphical representation reveals that certain departments with a high case density (depicted in an intense red color in Figure 3 exhibit low CLC3 coverage (depicted in a very light green color in Figure 2). The foregoing can also be verified in Figure 4, where the trend of a decrease in cases is observed as the area of CLC3 coverage by department increases.
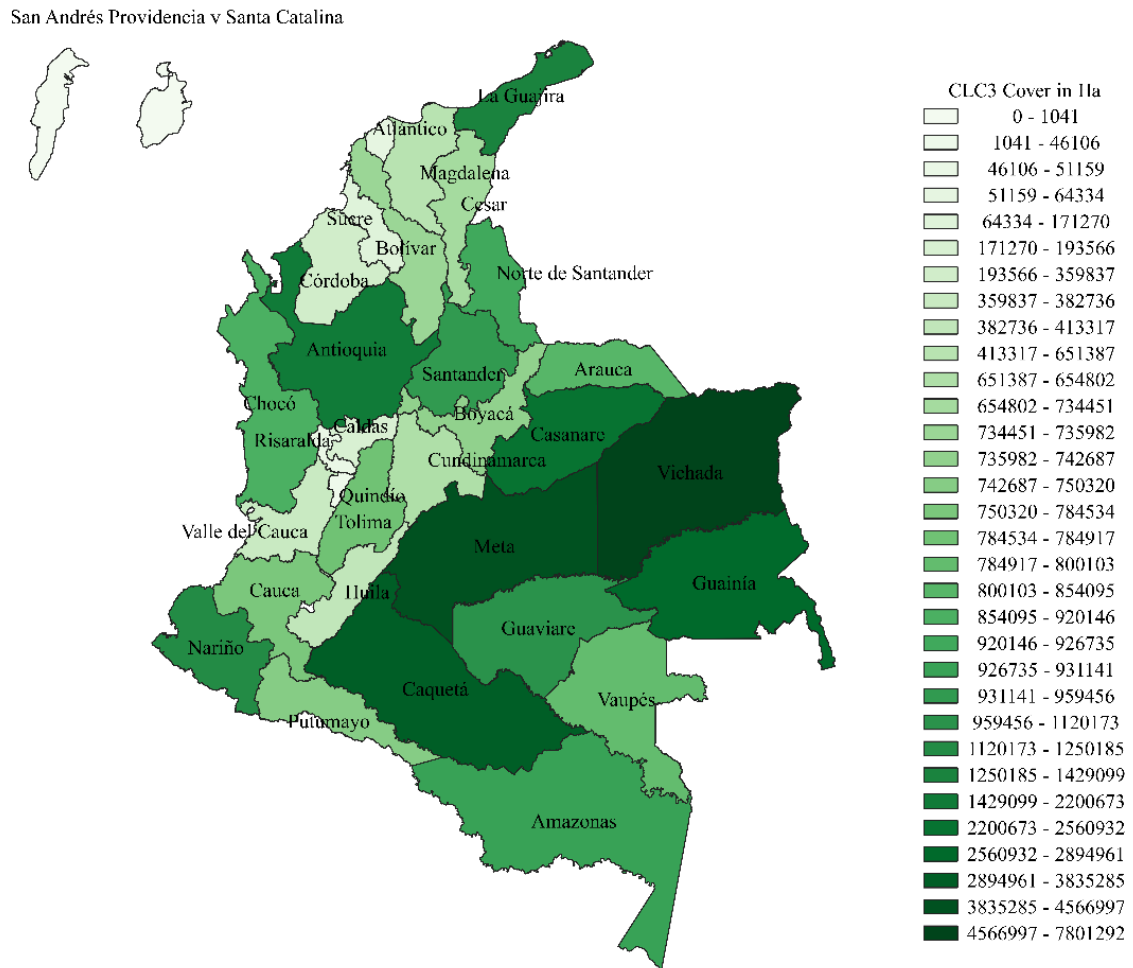

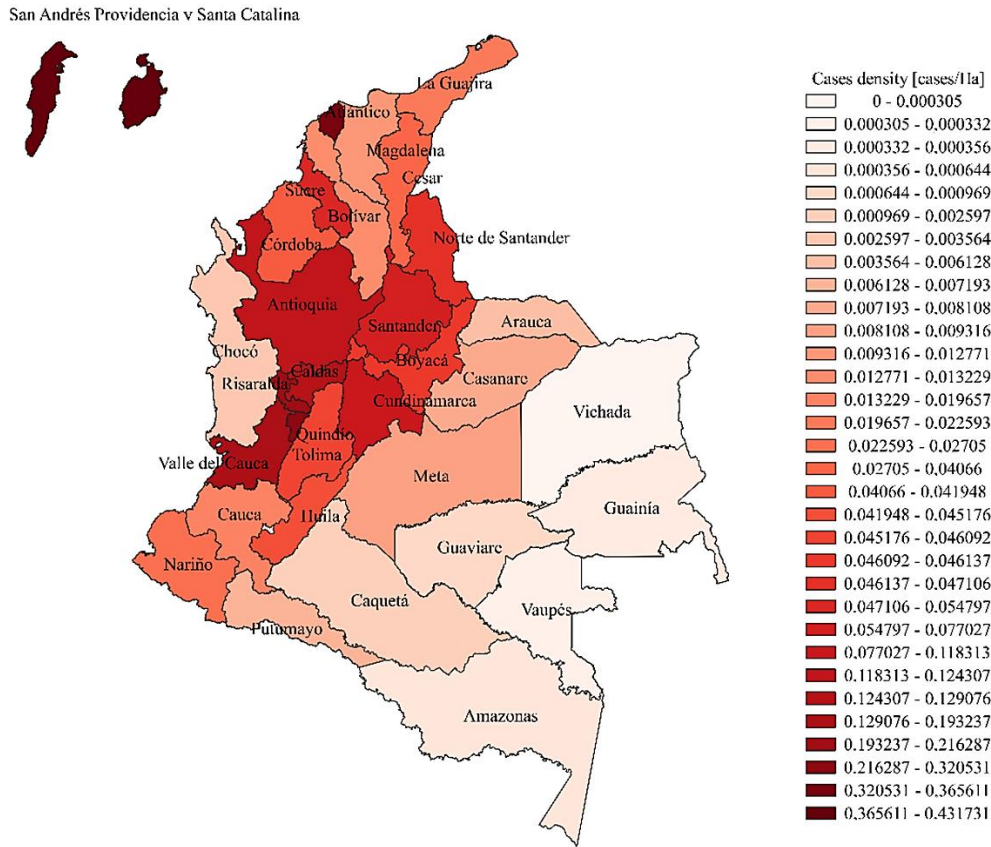Figure 2. Distribution of CLC3 coverage in Colombia

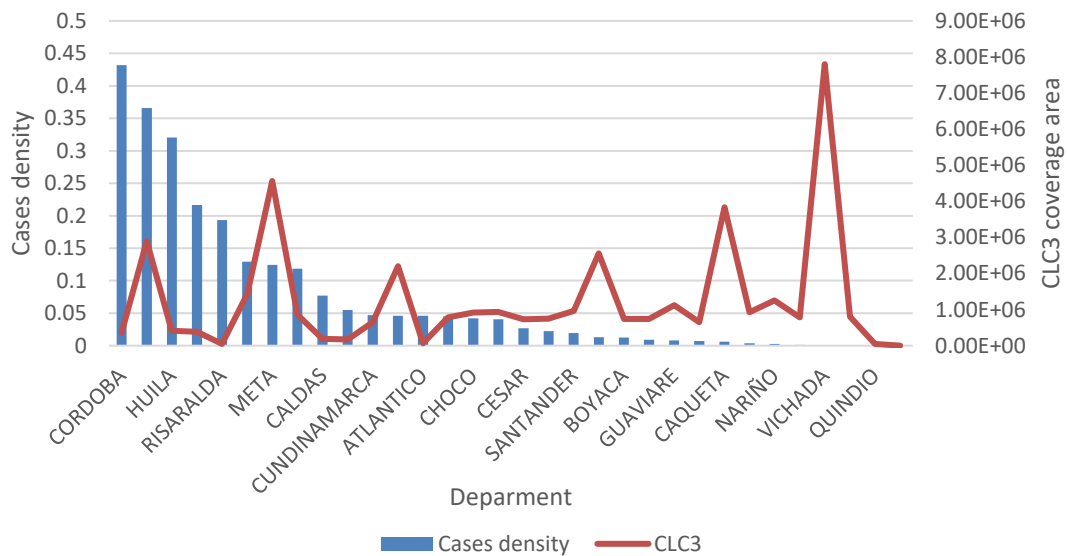Figure 3. Distribution of COVID - 19 cases in Colombia

Figure 4. Relation between area of CLC 3 and cases density

In a more detailed way, Figures 5, 6, and 7 graphically present the relationship between the three coverages analyzed at the CLC3 level. It is possible to observe the negative relationship that exists between each of the coverages and the density of cases, as shown in Table 4. This supports the idea that in places where there is a reduced surface extension of the coverages under analysis, there were quite a few cases of COVID-19.
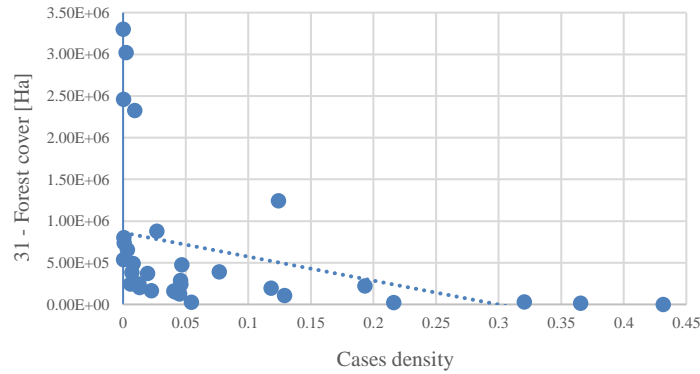
Figure 5. Cover with code 31 vs cases density
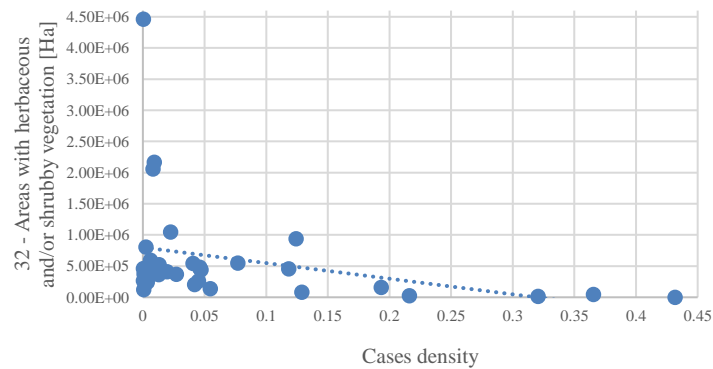
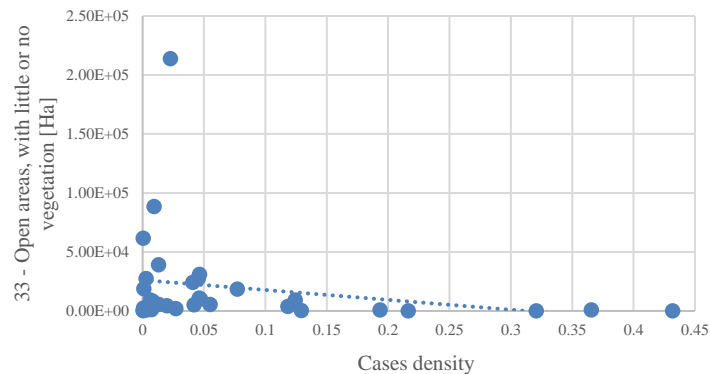Figure 6. Cover with code 32 vs cases density

Figure 7. Cover with code 33 vs cases density

Finally, unlike other studies, the present study was able to present results of the correlation between five categories of land cover and the density of cases of COVID-19. The cover that presents the highest negative correlation is CLC3 associated with forest; although this has been reported in other articles, in the present article, it is evident that depending on the type of forest, the correlation with the cases of COVID-19 will be the same. Thus, in areas with forest cover (code 31), the highest negative correlation is presented; this cover includes natural forests and plantations, and at the local level in Colombia, it also includes natural biological forms such as palm and guadua. Then, a lower negative correlation is found, associated with herbaceous and/or shrub vegetation cover (code 32), including low vegetation and tangle vegetation for Colombia. Finally, the lowest negative correlation is found with the cover of open areas with little or no vegetation (code 33), particularly bare and burned soils or soils covered by ice and snow.

4. CONCLUSION

The present study allowed us to explore the relationship between forest cover and the number of positive cases of COVID-19 in Colombia. It is possible to observe that in the areas where there is a greater forest cover, the number of cases is lower. This may be due, on the one hand, to what has been stated by different authors who indicate that the vegetation cover allows to improve air quality indices and therefore reduce the spread of COVID-19. On the other hand, this relationship can also occur given that in the areas with the greatest forest cover in Colombia, the population density is much lower than in the areas where there is less forest cover, for both in the areas with the greatest forest cover, the transmissibility of the virus may be lower. The results obtained have some limitations associated with the amount of data available to carry out the study and their variability in Colombia. This country presents substantial differences between its departments due to its geographic location, population, and distribution of forests. It is hoped that a future study can be carried out with continental scale data to have a larger dataset to validate the hypothesis, or at the national level as in the present study but with forest cover data at the urban area level, requiring very high-resolution images. It is hoped that studies of this type can provide decision-making tools to government entities to support forest conservation and expansion processes and the generation of incentives for constructions that include biophilic elements to increase vegetation cover within urban areas.

REFERENCES

[1] J. Piret and G. Boivin, "Pandemics throughout history," *Frontiers in Microbiology*, vol. 11, no. January, 2021, doi: 10.3389/fmicb.2020.631736.

[2] Y. Wang *et al.*, "Monitoring crop growth during the period of the rapid spread of COVID-19 in China by remote sensing," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 6195–6205, 2020, doi: 10.1109/JSTARS.2020.3029434.

[3] L. Zhang, F. M. Shen, F. Chen, and Z. Lin, "Origin and evolution of the 2019 novel coronavirus," *Clinical Infectious Diseases*, vol. 71, no. 15, pp. 882–883, 2020, doi: 10.1093/cid/ciaa112.

[4] S. Singh *et al.*, "Impact of forest fires on air quality in Wolgan Valley, New South Wales, Australia—a mapping and monitoring study using Google Earth engine," *Forests*, vol. 13, no. 1, 2022, doi: 10.3390/f13010004.

[5] R. Khan, K. R. Kumar, and T. Zhao, "The impact of lockdown on air quality in Pakistan during the covid-19 pandemic inferred from the multi-sensor remote sensed data," *Aerosol and Air Quality Research*, vol. 21, no. 6, pp. 1–18, 2021, doi: 10.4209/aaqr.200597.

[6] S. Sannigrahi *et al.*, "Examining the status of improved air quality in world cities due to COVID-19 led temporary reduction in anthropogenic emissions," *Environmental Research*, vol. 196, no. February, 2021, doi: 10.1016/j.envres.2021.110927.

[7] S. Saran, P. Singh, V. Kumar, and P. Chauhan, "Review of geospatial technology for infectious disease surveillance: Use case on COVID-19," *Journal of the Indian Society of Remote Sensing*, vol. 48, no. 8, pp. 1121–1138, 2020, doi: 10.1007/s12524-020-01140-5.

[8] L. M. Pereira *et al.*, "Social vulnerability and COVID-19 in Maringá, Brazil," *Spatial Information Research volume*, vol. 31, pp. 51–59, 2022, doi: 10.1007/s41324-022-00479-w.

[9] V. Henao-Cespedes, Y. A. Garcés-Gómez, S. Ruggeri, and T. M. Henao-Cespedes, "Relationship analysis between the spread of COVID-19 and the multidimensional poverty index in the city of Manizales, Colombia," *Egyptian Journal of Remote Sensing and Space Science*, vol. 25, no. 1, pp. 197–204, 2022, doi: 10.1016/j.ejrs.2021.04.002.

[10] H. M. Zawbaa *et al.*, "A study of the possible factors affecting COVID-19 spread, severity and mortality and the effect of social distancing on these factors: Machine learning forecasting model," *International Journal of Clinical Practice*, vol. 75, no. 6, pp. 1–14, 2021, doi: 10.1111/ijcp.14116.

[11] M. A. Nanda *et al.*, "The susceptible-infected-recovered-dead model for long-term identification of key epidemiological parameters of COVID-19 in Indonesia," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 12, no. 3, pp. 2900–2910, 2022, doi: 10.11591/ijece.v12i3.pp2900-2910.

[12] H. Ouchra, A. Belangour, and A. Erraissi, "Spatial data mining technology for GIS: a review," *2022 International Conference on Data Analytics for Business and Industry, ICDABI 2022*, pp. 655–659, 2022, doi: 10.1109/ICDABI56818.2022.10041574.

[13] S. Eskandari, H. R. Pourghasemi, and J. P. Tiefenbacher, "Relations of land cover, topography, and climate to fire occurrence in natural regions of Iran: Applying new data mining techniques for modeling and mapping fire danger," *Forest Ecology and Management*, vol. 473, no. June, p. 118338, 2020, doi: 10.1016/j.foreco.2020.118338.

[14] A. Abdollahnejad, D. Panagiotidis, S. S. Joybari, and P. Surovỳ, "Prediction of dominant forest tree species using quickbird and environmental data," *Forests*, vol. 8, no. 2, 2017, doi: 10.3390/f8020042.

[15] N. S. Saito, F. V. P. Arguello, M. A. Moreira, A. R. dos Santos, F. C. Eugenio, and A. C. Figueiredo, "Use of geotechnology for temporal analysis of forest cover," (in Portuguese) *Cerne*, vol. 22, no. 1, pp. 11–18, 2016, doi: 10.1590/01047760201622011935.

[16] S. Ruggeri, V. Henao-Cespedes, Y. A. Garcés-Gómez, and A. Parra Uzcátegui, "Optimized unsupervised CORINE Land Cover mapping using linear spectral mixture analysis and object-based image analysis," *Egyptian Journal of Remote Sensing and Space Science*, vol. 24, no. 3, pp. 1061–1069, 2021, doi: 10.1016/j.ejrs.2021.10.009.

[17] J. D. Colby and P. L. Keating, "Land cover classification using Landsat TM imagery in the tropical highlands: The influence of anisotropic reflectance," *International Journal of Remote Sensing*, vol. 19, no. 8, pp. 1479–1500, 1998, doi: 10.1080/014311698215306.

[18] M. Gábor, L. Karlík, V. Faltan, and F. Petrovič, "Implementation of semi-automated object-based image land cover classification

methods: A case study of the Male Karpaty Mts. (Slovakia)," *Geograficky Casopis*, vol. 69, no. 2, pp. 113–128, 2017.

[19] B. Jiang *et al.*, "Green spaces, especially forest, linked to lower SARS-CoV-2 infection rates: A one-year nationwide study," *medRxiv*, vol. 228, 2021, doi: 10.1016/j.landurbplan.2022.104583.

[20] A. Falco *et al.*, "COVID-19 epidemic spread and green areas Italy and Spain between 2020 and 2021: An observational multi-country retrospective study," *Environmental Research*, vol. 216, no. August 2022, 2023, doi: 10.1016/j.envres.2022.114089.

[21] R. Suligowski and T. Ciupa, "Five waves of the COVID-19 pandemic and green–blue spaces in urban and rural areas in Poland," *Environmental Research*, vol. 216, no. August 2022, 2023, doi: 10.1016/j.envres.2022.114662.

[22] Y. You and S. Pan, "Urban vegetation slows down the spread of coronavirus disease (COVID-19) in the United States," *Geophysical Research Letters*, vol. 47, no. 18, pp. 1–9, 2020, doi: 10.1029/2020GL089286.

[23] M. H. Winnicki, R. R. Dunn, M. Winther-Jensen, T. Jess, K. H. Allin, and H. H. Bruun, "Does childhood exposure to biodiverse greenspace reduce the risk of developing asthma?," *Science of the Total Environment*, vol. 850, no. July, 2022, doi: 10.1016/j.scitotenv.2022.157853.

[24] N. Mohamed and F. A. Hemeida, "The city lung and COVID-19: Effect of air quality on infection control and human health," *Future Cities and Environment*, vol. 8, no. 1, pp. 1–17, 2022, doi: 10.5334/fce.145.

[25] A. Tran and M. Raffy, "On the dynamics of dengue epidemics from large-scale information," *Theoretical Population Biology*, vol. 69, no. 1, pp. 3–12, 2006, doi: 10.1016/j.tpb.2005.06.008.

[26] United Nations CEPALSTAT, "Colombia: national socio-demographic profile," United Nations CEPALSTAT, 2023.

[27] J. A. Anaya, R. R. Colditz, and G. M. Valencia, "Land cover mapping of a tropical region by integrating multi-year data into an annual time series," *Remote Sensing*, vol. 7, no. 12, pp. 16274–16292, 2015, doi: 10.3390/rs71215833.

[28] DANE, "National Population and Housing Census-CNPV-2018," (in Spanish), National Administrative Department of Statistics, 2022. https://www.dane.gov.co/index.php/estadisticas-por-tema/demografia-y-poblacion/proyecciones-de-poblacion (accessed Nov. 28, 2022).

[29] INS, "COVID-19 in Colombia," Instituto Nacional de Salud, 2022. https://www.ins.gov.co/Noticias/Paginas/coronavirus-casos.aspx (accessed Nov. 28, 2022).

[30] IDEAM, "Land cover map. CORINE land cover adaptation. Republic of Colombia. Scale 1:100,000. 2018 period," (in Spanish), Colombia en mapas, 2021. https://www.colombiaenmapas.gov.co/?e=-76.72593182729393,3.5648638999446645,-76.37711590932565,3.758445174465464,4686&b=igac&l=880&u=76377&t=32&servicio=880# (accessed May 26, 2023).

[31] QGIS, "Developers guide for QGIS," *QGIS Documentation*, 2023. https://docs.qgis.org/3.28/en/docs/developers_guide/index.html (accessed Jun. 02, 2023).

[32] QGIS, "A gentle introduction to GIS, spatial analysis (Interpolation)" *QGIS Documentation*, 2020. https://docs.qgis.org/testing/en/docs/gentle_gis_introduction/spatial_analysis_interpolation.html (accessed Jun. 18, 2020).

[33] European Environmental Agency, *CORINE land cover - technical guide*. European Environmental Agency, 1997.

[34] IDEAM, *Leyenda nacional de coberturas de la tierra. Metodología CORINE Land Cover Adaptada para-Colombia Escala 1:100.000*. Bogotá, D. C., 2010.

[35] M. Shimada *et al.*, "New global forest/non-forest maps from ALOS PALSAR data (2007-2010)," *Remote Sensing of Environment*, vol. 155, pp. 13–31, 2014, doi: 10.1016/j.rse.2014.04.014.

[36] ALOS, "Global PALSAR-2/PALSAR/JERS-1 mosaics and forest/non-forest maps," ALOS Research and Application Project, 2023, https://www.eorc.jaxa.jp/ALOS/en/dataset/fnf_e.htm (accessed Jul. 04, 2023).

[37] S. Angée, S. I. Lozano-Argel, E. N. Montoya-Munera, J. D. Ospina-Arango, and M. S. Tabares-Betancur, "Towards an improved ASUM-DM process methodology for cross-disciplinary multi-organization big data & analytics projects," *Communications in Computer and Information Science*, vol. 877, pp. 613–624, 2018, doi: 10.1007/978-3-319-95204-8_51.

BIOGRAPHIES OF AUTHORS

Vladimir Henao-Céspedes 🔶 📇 SC 🔷 received the B.S degree in electronic engineering, the M.Sc. degree and the PhD in engineering from Universidad Nacional de Colombia, Manizales. He currently an associate professor at the Academic Unit for Training in Natural Sciences and Mathematics, Universidad Católica de Manizales, Manizales. His research interests include internet of things, industry 4.0, remote sensing, electromagnetic compatibility, electromagnetic pollution, lightning discharges. He is member of the Research Group on Technological and Environmental Development GIDTA. He can be contacted at email: vhenao@ucm.edu.co.

Yeison Alberto Garcés-Gómez 🔶 📇 SC 🔷 Received bachelor's degree in electronic engineering, and master's degrees and PhD in engineering from Electrical, Electronic and Computer Engineering Department, Universidad Nacional de Colombia, Manizales, Colombia, in 2009, 2011 and 2015, respectively. He is full professor at the Academic Unit for Training in Natural Sciences and Mathematics, Universidad Católica de Manizales, and teaches several courses such as experimental design, statistics, physics. His main research focus is on applied technologies, embedded system, power electronics, power quality, but also many other areas of electronics, signal processing and didactics. He published more than 30 scientific and research publications, among them more than 10 journal papers. He worked as principal researcher on commercial projects and projects by the Ministry of Science, Tech and Innovation, Republic of Colombia. He can be contacted at email: ygarcesg@ucm.edu.co.