Geospatial analysis with Python for sustainable management and optimization of land productivity. Case study: La Union - Valle del Cauca

Cristian Fabian Gallego-Martinez, Jorge Andrés Jola-Hernández, Mario Alejandro Celis-Torres

1. Introduction

The municipality of La Unión is located in the department of Valle del Cauca, Colombia. It borders Toro to the north, Roldanillo to the south, the Cauca River and the municipalities of La Victoria and Obando to the east, and El Dovio and Versalles to the west. Its total area is 125 km², of which 2.81 km² corresponds to the urban area and 122.19 km² to the rural area. Agriculture is the main economic activity in the municipality, with notable cultivation of grapes, passion fruit, various fruit crops, and sugarcane (Quillas, 2024).

Land-use conflict is a significant issue in La Unión, where the current use of land does not align with its environmental, ecological, cultural, social, and economic potential and constraints. The Regional Autonomous Corporation of Valle del Cauca (CVC), in collaboration with the municipal government (Villaquiran G, 2024), conducted a study to identify areas where vegetation cover or land use diverges from the physical conditions of the soil. This study revealed that 5,700 hectares, equivalent to 47% of the municipal area, show a high degree of land-use conflict, while 2,048 hectares, representing 17% of the municipal area, exhibit moderate conflict. In total, over 50% of the municipality faces land-use issues, with cases where conservation-oriented lands are being used for agricultural or forestry activities, increasing the risk of contamination due to land-use conflicts (IGAC, 2012).

Geographic Information Systems (GIS) using python has become an essential tool for managing and analyzing spatial information (Bolstad P, 2016). The rapid development of GIS, along with the emergence of methodologies based on artificial intelligence and deep learning, as well as increased access to spatial data, has positively impacted problems involving land use and land cover. In a Bayesian characterization study of urban land-use configurations using very high-resolution (VHR) remote sensing images, Li et al. (2020) integrated spatial disposition and composition variables for urban land-use extraction. The results showed that the proposed method produced urban land-use extractions with comparable or better accuracy than existing methods, achieving 86% and 93% precision, surpassing existing methods' 83% and 88%.

Zoungrana et al. (2023) developed a methodology to leverage optical and radar time-series images for estimating wheat cultivated areas before the harvest period with an accuracy of 84%. In coastal areas, the application of convolutional neural networks combined with object-based image analysis improved classification accuracy, producing final maps for regional and national decision-making with overall accuracy values of 93.5% using Pléiades satellite images (Zaabar N et al., 2022).

Additionally, Pachón et al. (2018) developed a raster geographic viewer and profile for the Regional Autonomous Corporation of Valle del Cauca (CVC). Using digital elevation models and other GIS tools, the viewer facilitates geospatial data visualization and analysis, supporting environmental planning and management. The implementation used Python, ArcGIS, and geographic web services to enhance accessibility and utility for decision-making. Jaramillo, F. (2024) created a geographic web portal to support water resource planning in La Unión by combining geospatial information from various sources with a multi-criteria analysis methodology. This initiative identified strategic areas for water conservation. Salazar, A. (2021) developed a monitoring and management tool for soils and zoning for the Roldanillo, La Unión, and Toro irrigation district in Valle del Cauca, integrating 208 spatial entities into a spatial database.

Spatial analysis within GIS enables the collection, management, analysis, and presentation of geospatial data, including sampling methods, visualization, representation, and spatial component analysis (Pu Hao, 2019). The location and attributes of spatial objects are critical in such analyses, as results depend on the positions and relationships of the analyzed objects. Data serve as the starting point for spatial analysis, and their quality depends on the tools and methods used for collection. In the past, obtaining data was time-consuming and costly, involving extensive fieldwork. Today, satellite images from platforms like Sentinel 1 and 2, processed in the Google Earth Engine geospatial analysis platform with 10 m spatial resolution, are readily available online.

Final data analyses have been conducted using programming languages like Python due to its versatility and libraries such as *sklearn*, *matplotlib*, *pandas*, *geopandas*, *keras, rasterio*, *pyidw*, and *geometry*. Panyadee P. (2004) used these tools to generate a flood risk map in northern Thailand.

This project aims to conduct an advanced spatial analysis using technological tools, including specialized programming languages like Python. The primary objective is to develop well-grounded recommendations to promote sustainable land management, contributing to the optimization of land productivity based on variables derived from diverse data sources.

2. Area of interest

The eastern sector of La Unión lies in the alluvial plain of the lower Cauca River valley, with soils primarily used for agriculture, supported by the RUT irrigation district (1958-1966), covering 10,300 hectares (Figure 1). Moving westward towards the Pacific, the landscape transitions into foothills with a dry warm climate (IGAC, 2012), progressively increasing in slope and reaching mountainous terrain (1,600-1,700 m.a.s.l.). The climate becomes moderately humid, and the steeper slopes (70%-100%), along with a greater presence of forest trees, shift the land's suitability towards conservation (Villaquirán G., 2024).

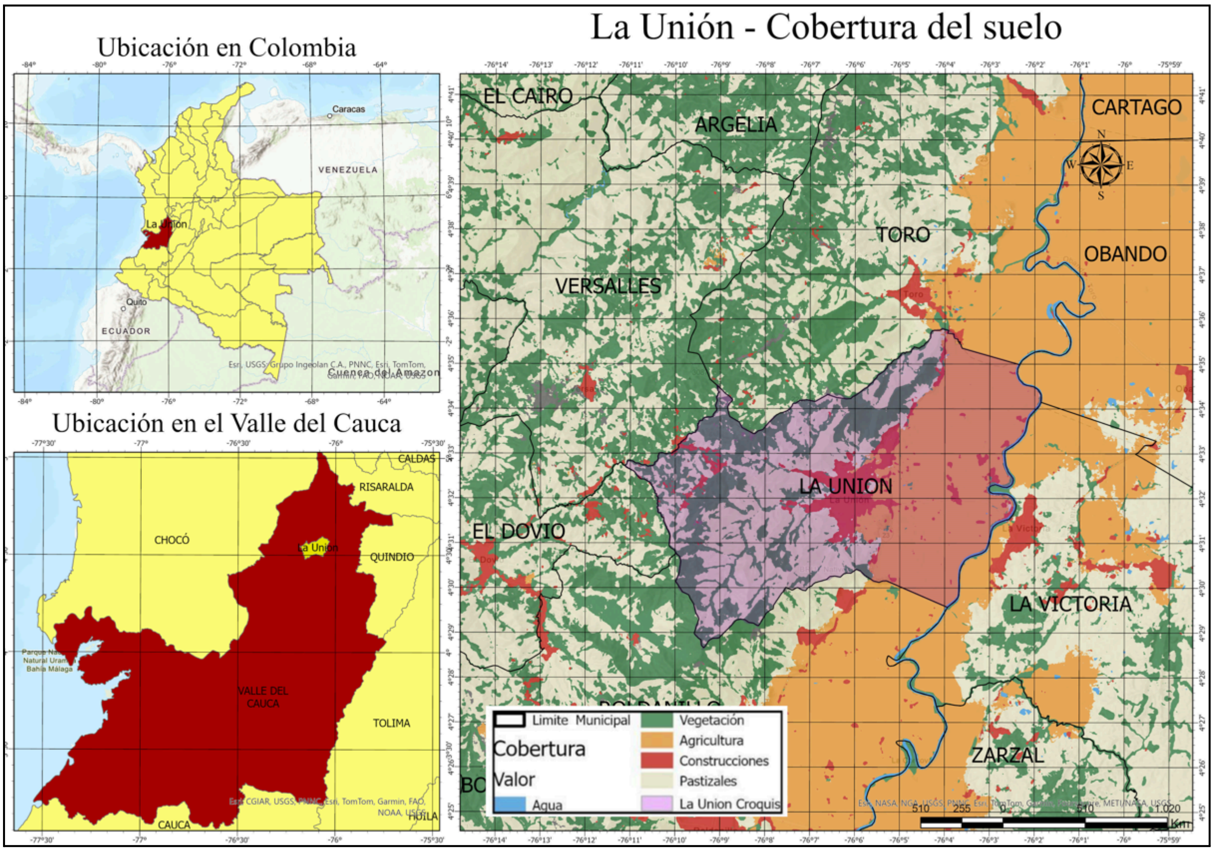


Figure 1. Research area

3. **Objectives**

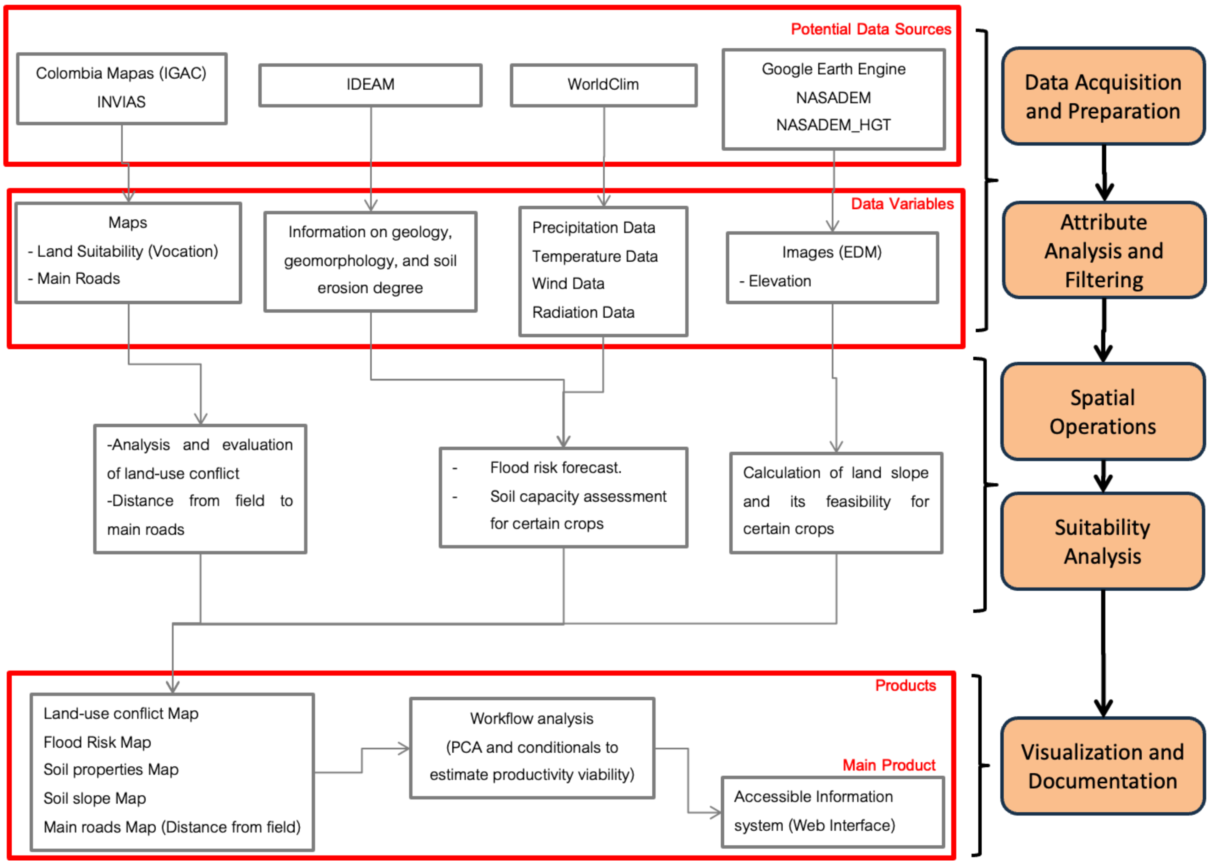
   General Objective:

- Implement operations and algorithms using Python programming for geospatial and edaphoclimatic analysis to estimate the productivity of properties in the municipality of La Unión.

   Specific Objectives:

- Execute an algorithm to identify conflict areas between land use and land suitability through vocation and coverage maps.
- Develop and implement processes in Python to estimate the productivity of agricultural plots, considering factors such as edaphoclimatic information, slopes, water access, and proximity to main roads, using GIS programming techniques.
- Generate code to identify properties affected by mining areas and calculate the distance of the properties to the main roads.
- Automate the estimation of agricultural activity viability on properties in the municipality of La Unión using a Python workflow, integrating it into an accessible information system (Web Interface).
-

4. **Methodology**



4.1 Data Sources

IGAC: Land use zoning maps from the national level in the "Colombia Mapas" web geographic service.

INVIAS: Cartographic information of the National Road Network of Non-Concessioned Highways under INVIAS, concessioned roads under ANI administration and secondary roads (Regional Road Plan Program - PVR)

IDEAM: Data from monitoring and tracking of soils and lands, with coverage of the National Land Cover Legend and CORINE Land Cover methodology adapted for Colombia, scale 1:100.000.

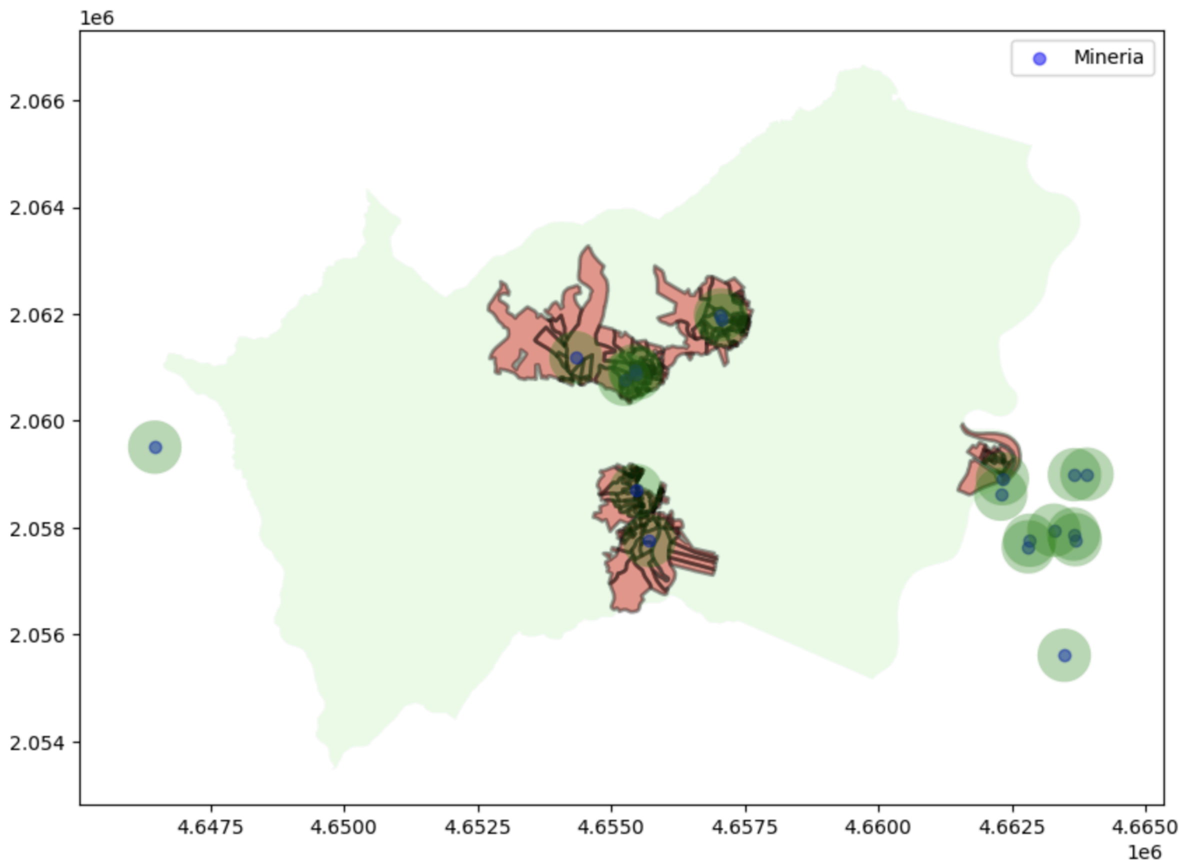
WorldClim: Climatic information with a spatial resolution of 30 meters for the year 2020.

Google Earth Engine: Extraction of a digital elevation model with a spatial resolution of 30 meters.

4.2 Operations

4.2.1 Mining impact of each plot.

To estimate properties affected by mining, a vector file containing points of mining activity areas was used. A 500-meter buffer was generated ('miner.buffer(500)'), and this information was saved as a GeoPandas object. Then, properties intersecting with these "mining areas" were selected and saved in a new .geojson file.

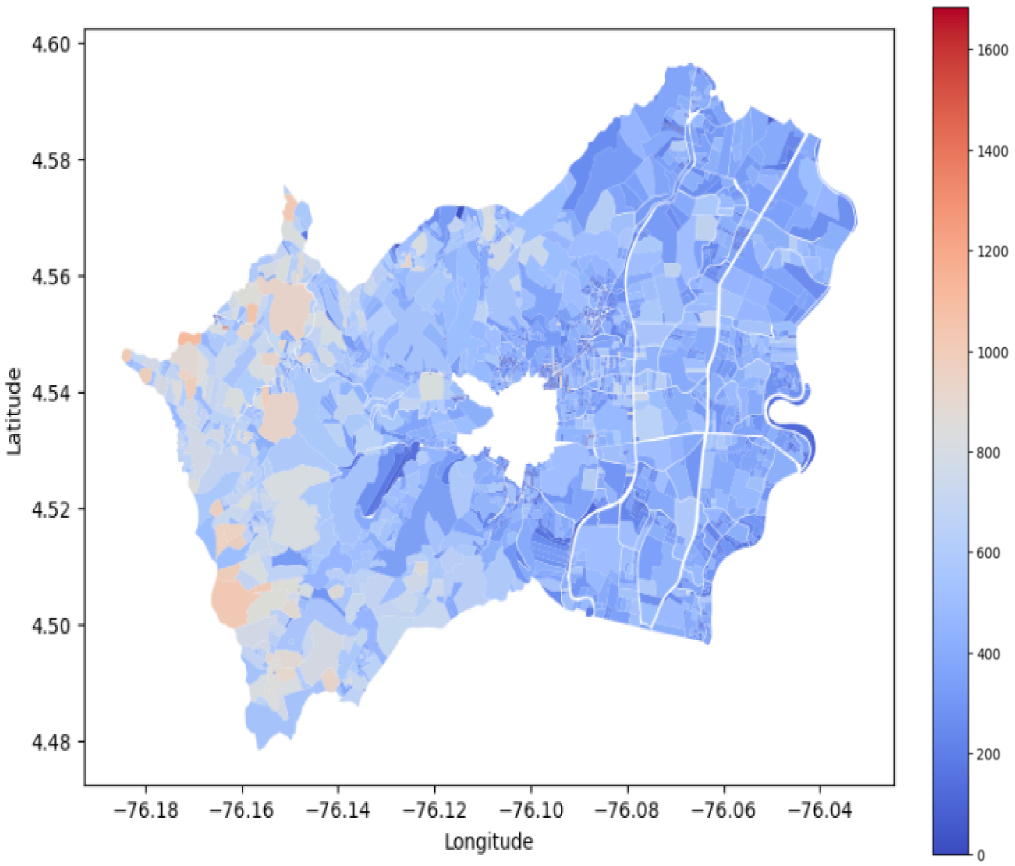


4.2.2 Slope of each plot.

To estimate the slope of each property, the elevation raster file was opened using rasterio. The elevation data was read and the slope was calculated by applying the gradient to the elevation values. The slope was then computed using the arctangent of the gradient and converting it to degrees.

Next, the properties were checked to ensure their coordinate reference system (CRS) matched the CRS of the raster file. If necessary, the properties were transformed to the raster's CRS. For each property in the dataset, the geometry was extracted, and a mask was applied to the elevation raster using a mask to focus on the area corresponding to the property. The valid (non-NaN) values within the masked area were used to calculate the average slope. This value was appended to a list.

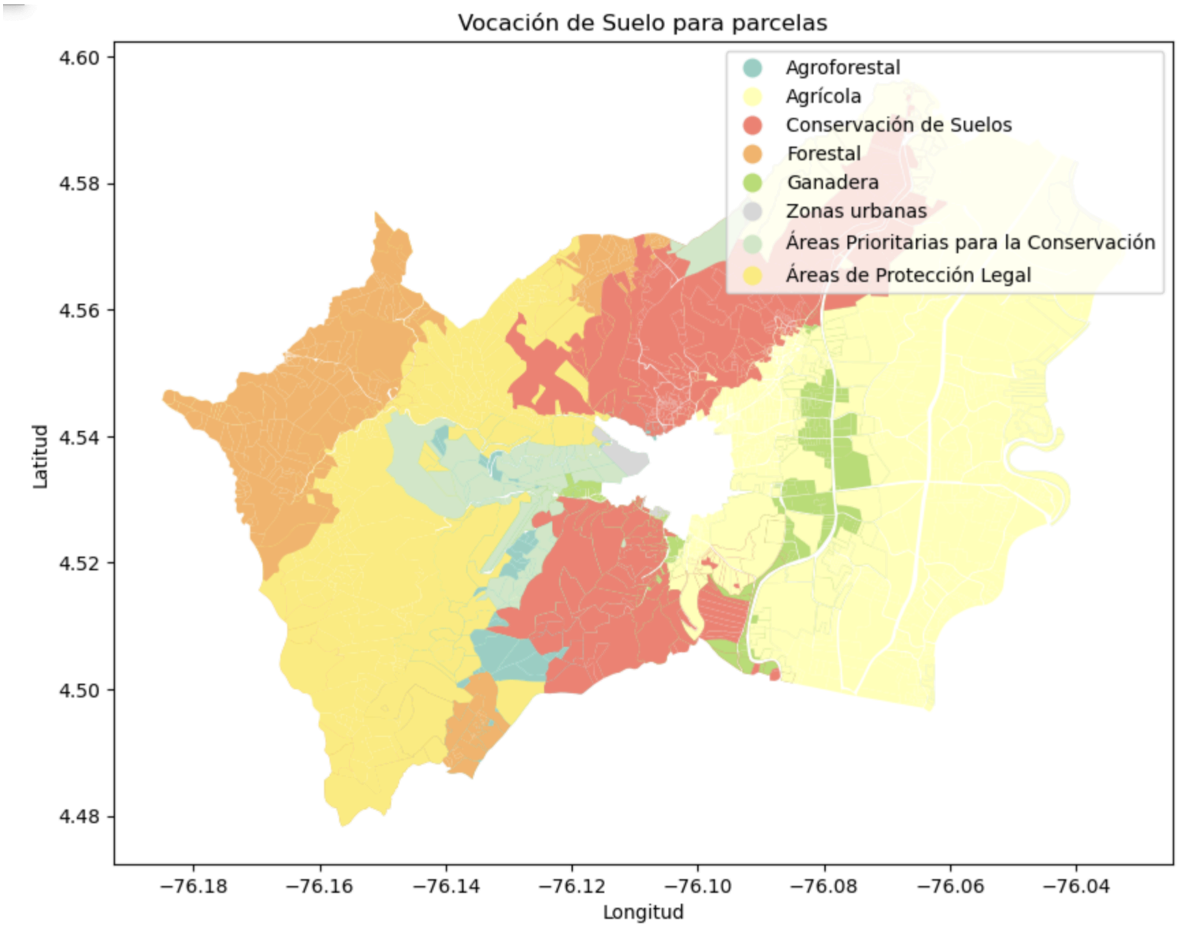
Finally, the calculated slopes were added to the properties dataset, and the results were saved to a new .geojson file.



4.2.3 Land vocation of each plot

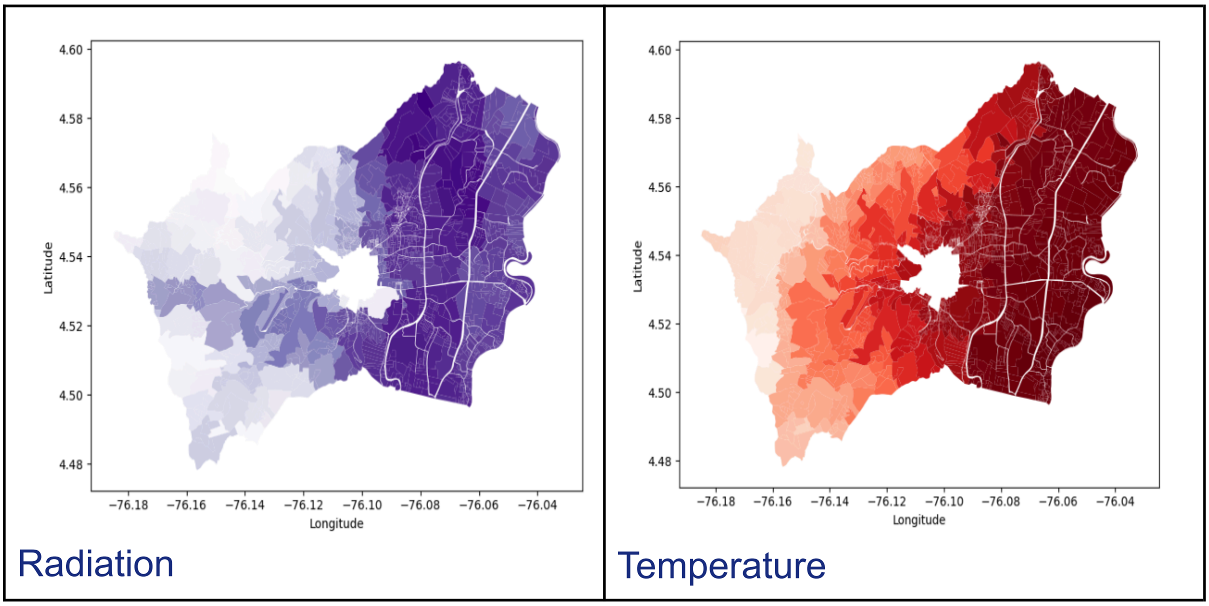
To estimate the land vocation for each property, the coordinate reference systems (CRS) of the property dataset and the land vocation dataset were checked. If the CRS of the two datasets didn't match, the vocation dataset was transformed to match the CRS of the property dataset.

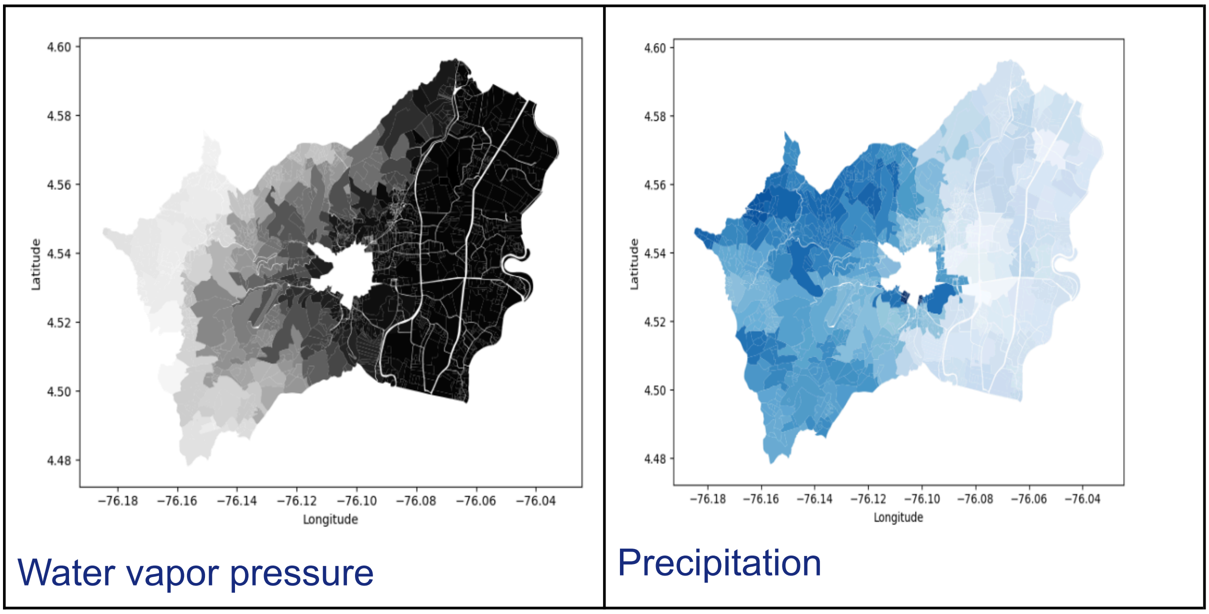
Then, a spatial join was performed to assign the land vocation value from the vocation dataset to the properties in the property dataset that intersected with the land vocation areas. The result was saved in a new .geojson file.



4.2.4 Edaphoclimatic properties of each plot.

The centroids of the properties were used to extract the values of climatic variables (temperature, precipitation, wind speed, vapor pressure, and radiation) from the corresponding rasters. These values were obtained at the centroid coordinates and added to a new field in the properties' GeoDataFrame. The result was saved in a GeoJSON file for further analysis and visualization.

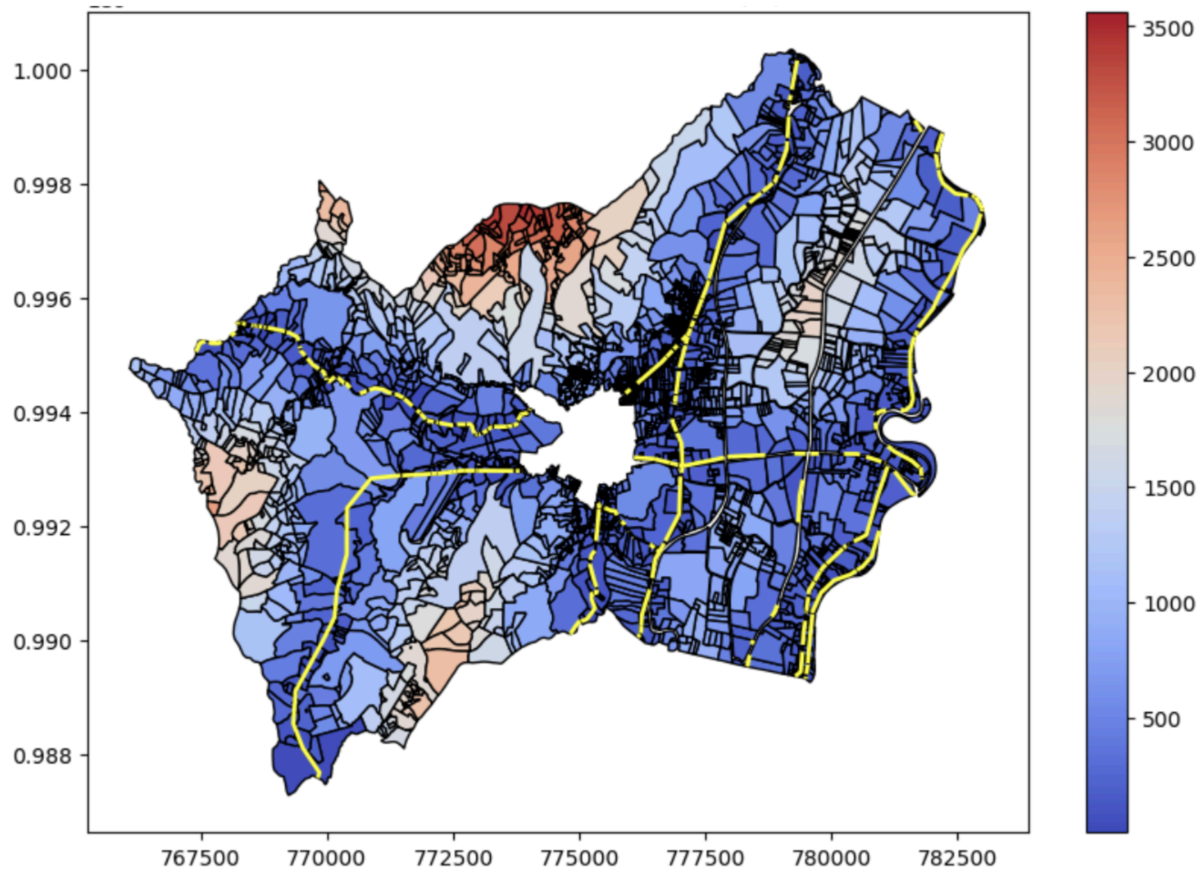




4.2.5 Distance to main roads of each plot.

Determining the distance from each parcel to the nearest road is essential in agriculture to optimize product transport, reduce logistics costs, improve access to inputs and machinery, and improve connectivity to markets and distribution centers. To calculate these distances, parcel and road datasets were loaded as GeoDataFrames using the Python GeoPandas library, both datasets were projected to the Magna-SIRGAS / UTM zone 18N (EPSG:3116) coordinate reference system to ensure accuracy of the distance measurements.

The centroid of each plot was calculated and the minimum distance from each centroid to the nearest road was determined. These distance values were added to the parcel dataset and a new shapefile was saved with the updated properties. Finally, a map was generated to visualize the distance distribution, with plots color-coded according to their proximity to the nearest road and roads highlighted in yellow.



5. References

- Bolstad, P. (2016). *GIS Fundamentals: A First Text on Geographic Information Systems*. Eider (Press Minnesota).
- IGAC. (2012). *Estudio suelos del Territorio Colombiano a escala 1:100.000. Departamento: Valle Del Cauca.* Departamento de agrología. https://geoportal.igac.gov.co/contenido/datos-abiertos-agrologia
- IGAC. (2012). *Estudio de los conflictos de uso del territorio Colombiano a escala 1:100.000*. https://geoportal.igac.gov.co/contenido/datos-abiertos-agrologia
- Jaramillo Urdinola, F. (2024). *Portal Web geográfico para apoyar la planificación del recurso hídrico en La Unión, Valle del Cauca*.
- Li, Mengmeng & Stein, Alfred & de Beurs, Kirsten. (2020). *A Bayesian characterization of urban land use configurations from VHR remote sensing images*. International Journal of Applied Earth Observation and Geoinformation. 92. 102175. 10.1016/j.jag.2020.102175.
- Pachón, L. E., Enrique, J., & Ojeda, H. (2018). *VISOR GEOGRÁFICO RASTER Y PERFIL PARA LA CORPORACIÓN AUTÓNOMA REGIONAL DEL VALLE DEL CAUCA*.
- Panyadee, P., & Champrasert, P. (2024). *Spatiotemporal Flood Hazard Map Prediction Using Machine Learning for a Flood Early Warning Case Study: Chiang Mai Province, Thailand*. Sustainability, 16(11), 4433. https://doi.org/10.3390/su16114433

- Hao, P. (2019). Spatial Analysis. The Wiley Blackwell Encyclopedia of Urban and Regional Studies.
- Quillas, C. I. L., Rodríguez, H. F. R., & Sabogal, J. R. (2024). Instituciones y desempeño institucional en la región del centro del Valle del Cauca. https://doi.org/10.25100/peu.913
- Salazar, A. (2021). *IMPLEMENTACIÓN DE CUADRO DE MANDO PARA LA GESTIÓN DEL DISTRITO DE RIEGO ROLDANILLO, LA UNIÓN Y TORO, VALLE DEL CAUCA*.
- Villaquiran G. (2024). Plan de desarrollo 2024-2027 municipio La Unión- Valle. Alcaldía municipal.
- Zoungrana, L.E., Barbouchi, M., Toukabri, W. *et al.* Sentinel SAR-optical fusion for improving in-season wheat crop mapping at a large scale using machine learning and the Google Earth engine platform. *Appl Geomat* **16**, 147–160 (2024). https://doi.org/10.1007/s12518-023-00545-4