



# Statistical correlation analysis between multispectral survey and agricultural soil physics variables

Diego Peláez Carrillo<sup>1</sup>, Oscar Gualdron Guerrero<sup>2,4</sup>, Marisol Maestre Delgado.<sup>3</sup>

<sup>1</sup>Department of Mechatronics Engineering, University of Pamplona,

<sup>2</sup>Department of Electronic Engineering, University of Pamplona - Colombia.

<sup>3</sup>Faculty of Economics and Business Administration, University of Pamplona, Colombia.

<sup>4</sup>Agroinnova Research Group, Ernesto Schiefelbein Foundation, Bogotá, Colombia

Email: diego.pelaez@unipamplona.edu.co, oscar.gualdron@unipamplona.edu.co,  
marisol.maestre@unipamplona.edu.co

**Abstract—** This study explores the use of multispectral surveys as a tool to acquire information on productive soils through flights with unmanned aerial vehicles (UAVs) equipped with cameras that capture images in specific ranges of visible and invisible light. Its objective is to observe the correlations between the Normalized Difference Vegetation Index (NDVI) and variables obtained in soil analysis, such as nitrogen, phosphorus, potassium and iron. The analysis seeks to identify the behavior of these physical variables by means of statistical models and to establish a rapid inspection methodology to characterize the soil and optimize agricultural management. Sampling was carried out on 40 properties in the municipality of Corozal, Sucre, Colombia, using two methods: autonomous overflight with UAVs to capture NDVI and manual collection of soil samples for nutrient analysis. The results include linear and polynomial correlation models between NDVI and physical variables, thus facilitating the characterization of the plots in the region. The main conclusion indicates that the third order polynomial model is the one that best fits the real behavior of the properties evaluated in the study area, with coefficients of determination higher than 0.4. This allows ruling out the nullity of the model and confirming its usefulness in the characterization of the soil conditions of the sampled properties.

**Keywords—** NDVI, Polynomial statistical analysis, multispectral imaging, soil characterization, image processing.

## I. INTRODUCTION

The implementation of new technologies in the agricultural activity is one of the most important challenges that has been accepted by science due to the fact that the world population is required to increase food production due to its exponential increase [1] and has achieved this through the

inclusion of remote monitoring tools that are a valuable source of information that can help motivate increased production and decreased investment of resources such as fertilizers, pesticides and, more importantly, the efficient use of available water resources [2].

In recent years, there has been an increase in the demand for natural resources, leading to overexploitation of ecosystems, particularly of the nutritional characteristics of productive land [3]. All this underestimating the risk and intensity of damage, degrading the land and thus increasing the desertification process in the long term, which is why it is necessary to implement planned techniques that contribute to appropriate land use [4].

Precision agriculture allows, through different inspection systems, to acquire precise information on the eternal conditions in which a crop is grown so that decisions can be made on the basis of the data collected [5]. Thanks to the intention of developing new acquisition tools, research is being carried out using multispectral cameras [6]. The cameras allow the acquisition of monochromatic images that capture the reflectance of different wavelengths, positioning itself as one of the most powerful tools in the identification of vegetation and how it is affected by its environment [7]. It works through the calculation of different indices, including one of the most common, the calculation of the normalized vegetation index (NDVI), with which the quality and quantity of biomass of the sampled habitat can be approximated [8]. It collects data on the different plant species of the sampled area, also identifying the amount of bare ground, thus being able to identify and quantify organic matter (Formica et al., 2017).

The NDVI allows to extract accurate data on the amount of plant material and the amount of bare soil [9] allowing, characterizing by means of percentages of cover, different levels of presence of plants with variable foliage, as well as possible presence of water footprint and the amount of bare soil present in the sampled area [10]. With this index, a comparative

process can be carried out to allow appropriate monitoring of possible nutritional needs and strategic decisions to be made [11].

Plants and any type of vegetation need certain nutrients that are essential for their full growth and the achievement of optimal yields [12],[13]. The consequences of a lack of these nutrients can range from impaired growth and discoloration of leaves to loss of fruiting bodies [14]. Considering the inspection problem and the identification of nutritional deficiencies, correlations of soil physical variables with NDVI values are proposed, taking into account those nutrients that have a direct impact on the characteristics of chlorophyll in the plant, based on the fact that this substance is what gives the plant its color [15].

Unmanned aerial vehicles (UAVs) enable non-invasive observation and data collection that facilitates phytosanitary care and crop data collection that allows early identification of diseases and pests [16]. By means of high-resolution cameras, UAVs acquire images with global position metadata in which the capture is performed [17], [18], thus they are widely used thanks to their features such as resolutions of 1.5 cm per pixel and the possibility of acquiring images at different heights [19].

The implementation of statistical methods allows comparing the information and identifying patterns in order to generate the classification of the sampled soils. Depending on the normality of the collected data it may be necessary to implement complex data analysis methods. Similarly, algorithms can be developed to make predictions in real-time [20]. These tools promise very accurate results that need to be thoroughly calibrated, for which multiple investigations are developed [21],[22].

## II. MATERIALS AND METHODOLOGY

The research was carried out in the municipality of Corozal, located in the department of Sucre, Colombia, at an altitude of 142 m.a.s.l. with temperatures that vary between 22 and 35°C. It is a population whose economy is largely based on agriculture, concentrating on the implementation of crops using traditional methods with little technology and lacking the use of advanced technologies.

The main objective of soil sampling is to obtain information that will allow fertilisation recommendations to be proposed thanks to the different samples that accurately represent whether the field has an optimal fertility status or nutrient deficiency for the growth of healthy vegetation. Initially [23], the purpose is to determine a measure of the field fertility level and a measure of the variability of this fertility [24]. Due to the cost of implementing survey technologies, observing soil fertility over time is not considered in the research as it would be

necessary to repeat the methodology with a high frequency [25].

The soil is generally heterogeneous with different nutrient levels in very close proximity, which is why it was decided to have a number of 40 plots of one hectare each [26]. Soil properties, including fertility, vary from one site to another in the field. As it is not easy to sample all the extensions of each plot, it was decided to extract sub-samples in order to estimate the level of phosphorus, potassium, nitrogen, and iron of the whole plot.

Taking into account the characteristics of the area, it was decided to involve data acquisition tools that allow more precise observation of the state of productive soils in the region [27]. One of the most powerful tools in the identification of vegetation and how it is affected by its environment is the calculation of the normalised vegetation index (NDVI), with which the biomass quality and quantity of the sampled habitat can be approximated by collecting data on the different plant species typical of the sampled area and also identifying the amount of bare ground, thus being able to identify and quantify organic matter [28], [29]. For the multispectral surveys, a DJI Matrix 100 UAV was used with a battery that provides a flight autonomy of 30 minutes and a power that allows the multispectral sensor to be integrated into its system [30],[31]. The camera used for image acquisition was the Micasense RedEdge-M, which allows information to be obtained from 5 wavelength ranges between 400 nm and 950 nm, and resolutions of 0.08 m per pixel at a height of 120 m [32].

The image acquisition process began by establishing the points of the polygons that correspond to the terrain boundaries that were selected for analysis, first performing a visual inspection, defining the physical properties of the terrain, and identifying possible obstacles that the UAV may face in order to avoid accidents and flight delay. Obtaining the global positioning system points is done by means of satellite equipment that enables obtaining points in real-time with high accuracy of the acquired positioning [33].

Once the polygons have been defined, the flight plans are created in which the aerial shots are configured, including different options such as the altitude at which the image acquisition will be performed, setting the parameter between 100 and 120 m. It is important to note that the lower the altitude, the higher the number of images and the memory consumption of the camera. The flight speed and percentage of horizontal and vertical overlap of the images are also configured. The flight plan is executed by the vehicle under the supervision of the pilot in charge of the aircraft, in case it is necessary to intervene in any manoeuvre, which is called autonomous flight.

For the analysis of the physical variables of each plot, 7 samples were taken from different parts of each of the plots in such a way that, by means of different

laboratory tests, the amounts of phosphorus, potassium, and iron were determined by means of the experimental technique called spectrophotometry. In the same way, the amount of available nitrogen was determined by means of volumetry.

Image processing requires the use of photometry software that allows the manipulation of RGB and multispectral images. On this occasion, Pix4D software was used, which allows: capturing images with any camera; transforming images into digital models; managing quality reports and calibration details; measuring distances, areas, and volumes; and extracting elevation profile data. Moreover, it allows the calculation of different vegetation indices, including NDVI, which was the index selected for the current project.

Once the multispectral orthophotomosaics are obtained for each location, the images are presented with their proper georeferencing, and the results of the percentages of covered areas are grouped into seven intervals in which the normalised difference vegetation indices are categorised and condensed in a table shown by property. The respective analysis of each image was carried out, presenting the information in a disaggregated and understandable way, allowing the farmer and his advisors to make timely decisions regarding the measures to be implemented within their property and characterising the soils of the municipality in general.

The main objective of the comparative activity carried out in the municipality of Corozal is to allow a global observation of the behaviour of the study variable. For the current case, the variables of the soil study of physical samples are related to the NDVI data acquired by means of multispectral surveys. Of the variables acquired from the physical samples, the most important were identified as those related to the chemical constitution of chlorophyll within the plants in such a way as to allow a characterisation of the soil with the help of multispectral images and the physical samples acquired from the different plots.

Statistical analysis is used to identify the possible pattern of trends in the data, giving the possibility of predicting the emergence of new values [28]. In order to fulfil this possibility, it was necessary to involve a certain number of standard components in this type of calculations. The data used in the statistical analysis were the coverage area percentages of a characteristic NDVI value for each of the 40 plots of the municipality; in addition, the data of potassium, phosphorus, iron, and nitrogen were included and crossed with the vegetation indices. The study is performed by grouping the NDVI generating models in order to compare the data and determine a representative behaviour [8].

For the data processing, an algorithm was developed within the mathematical software, which firstly allows the identification of the average NDVI vegetation index data and the percentage of presence for each

property. At the same time, a database is generated, where the results of the physical test are included for each property. This database is entered into the algorithm as a matrix to be filtered and smoothed by eliminating repeated data and outliers that may occur.

### III. RESULTS

The current items present the results of the data processing in which third-order polynomial models were generated for each of the correlations between the NDVI vegetation indices and the different physical variables acquired at the first sampling moment. Firstly, the correlation data between the most representative vegetation indices and the percentages of area covered by these indices and the model that represents their behaviour are shown, followed by the correlations between NDVI and the four most important physical variables present in the soil that are related to the production of chlorophyll in the plants and their respective models. In addition, some general considerations are made according to the results acquired, and trends are identified from the information shown, providing a global point of comparison for future sampling of the evolution of the vegetation cover.

The graphs resulting from the generation of the polynomial modelling generated are listed below:

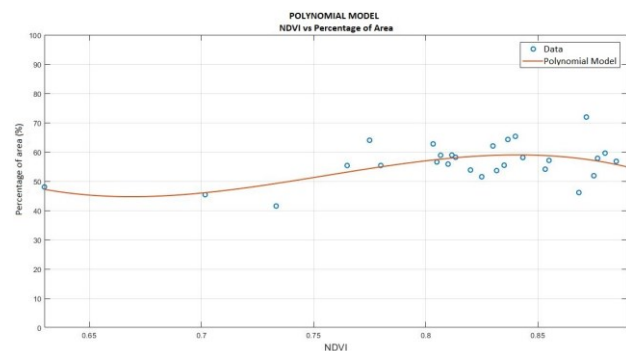


Figure 1. Polynomial NDVI vs Percentage Area Model

The data provided by the statistical analysis methodology implemented to obtain trend models allowed the acquisition of the results shown in figure 1. Initially, you can visualize the data used to feed the algorithm with which the information was processed and where all data were concentrated in NDVI vegetation indices between 0.701 and 0.89. These indices gave as first statement that all properties are highly represented by the NDVI interval called “high coverage” whose indices are concentrated between 0.75 and 1 and are related to characteristics of vegetation with high density, soil covered with plants with high foliage, or tall bushes with a hydration that allows adequate growth. On the other hand, the area percentage values are concentrated in data between 40% as a minimum value and 73% as a maximum

value, which indicates a medium and high presence of values close to 1 NDVI in most of the plots. The polynomial model represented by the red curve in figure 1 shows us an approach to the behavior of the variables. This curve develops within the values of 45% and 60%, indicating that within this range, we can find future values of plots with similar characteristics to those that were the subject of the study.

The algorithm implemented was able to correlate the physical variables with the NDVI indices of the different plots. These models are presented in figure 2, where each graph represents the relationship of each element as a function of the vegetation index. In general, the third-order polynomial model was the best adapted to the real behavior of the sampled plots in the municipality of Corozal. For the element Phosphorus in the soil in function of the NDVI, the model presents a correlation coefficient of 0,63, which indicates that the variables present a medium correlation, and the model is developed between the data of the variable between 2,55 and 5 mg/kg. In addition, having the goodness of the adjustment made, it can be affirmed that the null hypothesis is discarded, for which it is concluded that the variables are related, and the model is valid and replicable.

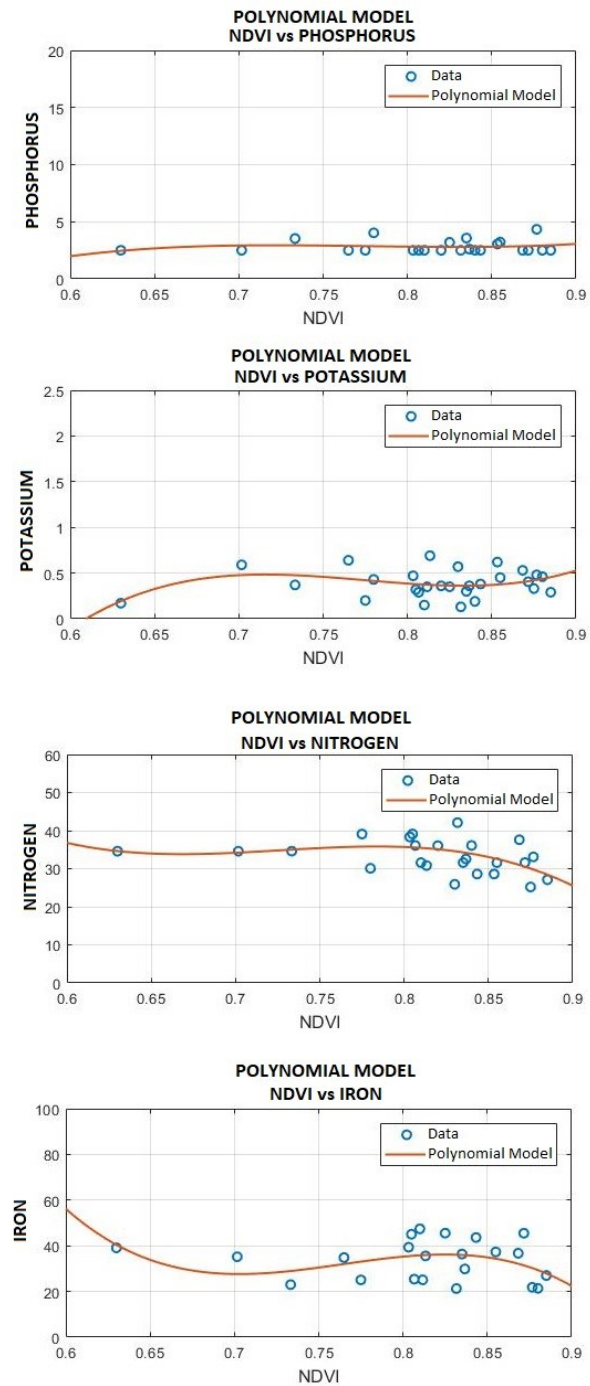


Figure 2. Polynomial model of correlation of NDVI with soil physical variables.

For the element Iron in the soil as a function of NDVI, the variables can be related under a linear or second order polynomial model. It is necessary to highlight that the models considered the data after being filtered, obtaining a determination coefficient calculated from the regression equal to 0.45, concluding that the variables are related. For the element in question, the third-order polynomial model shown in figure 2 is adopted, given that it provides better behavior considering the totality of the database. The element Nitrogen as a function of

NDVI is presented with a Pearson coefficient or correlation coefficient of 0.4, which shows that the variables have a medium correlation. Finally, the Potassium element in the soil as a function of NDVI, the variables can be related under a linear and a polynomial model of third order, adopting the polynomial model because its coefficient of determination was the highest, presenting a value of 0.53. The correlations of the models are medium, but still show an appropriate correlation between the variables.

#### IV. CONCLUSION

The statistical analysis implemented by means of the algorithm developed allowed the generation of correlation models, both linear and polynomial, generating spaces of analysis and comparison of data according to the behavior of the tendency curves. Allowing to conclude that the polynomial model of third order adapts to the real behavior of the sampled properties in the municipality of Corozal, this is verified thanks to the fact that the totality of the coefficients of determination calculated for each one of the regressions obtained data is superior to 0.4, demonstrating the low quality of the model to be able to be replicated later but discarding its nullity. The data obtained from the data processing allows us to ensure that the municipality of Corozal presents a behavior related to biological material of high quality and coverage, establishing the most representative vegetation index data between 0.63 and 0.8 with a calculated average of 0.81, also ensuring an absence of hydric stress and easy conditions to implement agro-ecological models with the correlations shown.

#### V. REFERENCES

- [1] A. Garbero and L. Jäckering, "The potential of agricultural programs for improving food security: A multi-country perspective," *Glob. Food Sec.*, vol. 29, p. 100529, 2021, doi: <https://doi.org/10.1016/j.gfs.2021.100529>.
- [2] E. Bwambale, F. K. Abagale, and G. K. Anornu, "Smart irrigation monitoring and control strategies for improving water use efficiency in precision agriculture: A review," *Agric. Water Manag.*, vol. 260, p. 107324, 2022, doi: <https://doi.org/10.1016/j.agwat.2021.107324>.
- [3] C. Griesche and A. J. Baeumner, "Biosensors to support sustainable agriculture and food safety," *TrAC Trends Anal. Chem.*, vol. 128, p. 115906, 2020, doi: <https://doi.org/10.1016/j.trac.2020.115906>.
- [4] M. L. Hossain and J. Li, "NDVI-based vegetation dynamics and its resistance and resilience to different intensities of climatic events," *Glob. Ecol. Conserv.*, vol. 30, p. e01768, 2021, doi: <https://doi.org/10.1016/j.gecco.2021.e01768>.
- [5] Y. Vecchio, M. De Rosa, F. Adinolfi, L. Bartoli, and M. Masi, "Adoption of precision farming tools: A context-related analysis," *Land use policy*, vol. 94, p. 104481, 2020, doi: <https://doi.org/10.1016/j.landusepol.2020.104481>.
- [6] J. Tang, P. Petrie, and M. Whitty, "Low-Cost Filter Selection from Spectrometer Data for Multispectral Imaging Applications\*\*This work was supported by Wine Australia," *IFAC-PapersOnLine*, vol. 52, no. 30, pp. 277–282, 2019, doi: <https://doi.org/10.1016/j.ifacol.2019.12.534>.
- [7] G. Modica, G. Messina, G. De Luca, V. Fiozzo, and S. Praticò, "Monitoring the vegetation vigor in heterogeneous citrus and olive orchards. A multiscale object-based approach to extract trees' crowns from UAV multispectral imagery," *Comput. Electron. Agric.*, vol. 175, p. 105500, 2020, doi: <https://doi.org/10.1016/j.compag.2020.105500>.
- [8] R. R. Fern, E. A. Foxley, A. Bruno, and M. L. Morrison, "Suitability of NDVI and OSAVI as estimators of green biomass and coverage in a semi-arid rangeland," *Ecol. Indic.*, vol. 94, pp. 16–21, 2018, doi: <https://doi.org/10.1016/j.ecolind.2018.06.029>.
- [9] C. Piedallu *et al.*, "Soil and climate differently impact NDVI patterns according to the season and the stand type," *Sci. Total Environ.*, vol. 651, pp. 2874–2885, 2019, doi: <https://doi.org/10.1016/j.scitotenv.2018.10.052>.
- [10] S. Kharuf-Gutierrez, L. Hernández-Santana, R. Orozco-Morales, O. de la Aday Díaz, and I. Delgado Mora, "Análisis de imágenes multiespectrales adquiridas con vehículos aéreos no tripulados," *Análisis imágenes multiespectrales Adquir. con vehículos aéreos no Tripul.*, vol. 39, no. 2, 2018.
- [11] X. Yang, Y. Liu, H. Tian, and Y. Shen, "Short-term nitrogen and phosphorus additions rather than mycorrhizal suppression determine plant community composition and productivity in desert steppe," *Appl. Soil Ecol.*, vol. 168, p. 104144, 2021, doi: <https://doi.org/10.1016/j.apsoil.2021.104144>.
- [12] J. K. Carrillo Gómez, C. M. Duran Acevedo, and R. O. Garcia Rico, "Discriminación de bacterias en agua potable a través de una nariz electrónica y un equipo de extracción de volátiles," *revista colombiana de tecnologías de avanzada (RCTA)*, vol. 1, no. 33, pp. 155–165, Jan. 2019, doi: [10.24054/rcta.v1i33.99](https://doi.org/10.24054/rcta.v1i33.99).
- [13] G. Illarze, A. del Pino, S. Riccetto, and P. Irisarri, "Emisión de óxido nitroso, nitrificación, desnitrificación y mineralización de nitrógeno durante el cultivo del arroz en 2 suelos de Uruguay," *Rev. Argent. Microbiol.*, vol. 50, no. 1, pp. 97–104, 2018, doi: <https://doi.org/10.1016/j.ram.2017.05.004>.
- [14] S. M. Zaigham Abbas Naqvi, M. Awais, F. S. Khan, U. Afzal, N. Naz, and M. I. Khan, "Unmanned air vehicle based high resolution imagery for chlorophyll estimation using spectrally modified vegetation indices in vertical hierarchy of citrus grove," *Remote Sens. Appl. Soc. Environ.*, vol. 23, p. 100596, 2021, doi: <https://doi.org/10.1016/j.rsase.2021.100596>.
- [15] L. Deng, Z. Mao, X. Li, Z. Hu, F. Duan, and Y. Yan, "UAV-based multispectral remote sensing for precision agriculture: A comparison between different cameras," *ISPRS J. Photogramm. Remote Sens.*, vol. 146, pp. 124–136, 2018, doi: <https://doi.org/10.1016/j.isprsjprs.2018.09.008>.
- [16] M. A. Kaljahi *et al.*, "An automatic zone detection system for safe landing of UAVs," *Expert Syst. Appl.*, vol. 122, pp. 319–333, 2019, doi: <https://doi.org/10.1016/j.eswa.2019.01.024>.
- [17] J. A. Gómez, H. Yulady Jaramillo, and L. A. Coronel Rojas, "Sistema para detección de fallos críticos en tuberías horizontales," *revista colombiana de tecnologías de avanzada (RCTA)*, vol. 1, no. 35, pp. 44–51, Feb. 2020, doi: [10.24054/rcta.v1i35.41](https://doi.org/10.24054/rcta.v1i35.41).
- [18] W. H. Maes and K. Steppe, "Perspectives for Remote Sensing with Unmanned Aerial Vehicles in Precision Agriculture," *Trends Plant Sci.*, vol. 24, no. 2, pp. 152–164, 2019, doi: <https://doi.org/10.1016/j.tplants.2018.11.007>.
- [19] T. Talaviya, D. Shah, N. Patel, H. Yagnik, and M. Shah, "Implementation of artificial intelligence in agriculture for optimisation of irrigation and application of pesticides and herbicides," *Artif. Intell. Agric.*, vol. 4, pp. 58–73, 2020, doi: <https://doi.org/10.1016/j.aiaa.2020.04.002>.
- [20] D. I. Patrício and R. Rieder, "Computer vision and artificial intelligence in precision agriculture for grain crops: A systematic review," *Comput. Electron. Agric.*, vol. 153, pp. 69–81, 2018, doi: <https://doi.org/10.1016/j.compag.2018.06.029>.

- <https://doi.org/10.1016/j.compag.2018.08.001>.
- [21] F. R. Cuesta Quintero, L. A. Coronel Rojas, D. Rico Bautista, E. Barrientos Avendaño, O. J. Montañez vergel, and C. M. Páez Noriega, "Sistema de detección de intrusos a través de una red honeynet para entornos de red cableada sobre IPv6," *revista colombiana de tecnologías de avanzada (RCTA)*, vol. 1, no. 33, pp. 104–113, Jan. 2019, doi: 10.24054/rcta.v1i33.92.
- [22] T. L. Roberts and J. . Henry, "El muestreo de suelos: Los beneficios de un buen trabajo," *Inf. Agronómicas del Cono Sur*, vol. 8, 2000.
- [23] E. Cordero, L. Longchamps, R. Khosla, and D. Sacco, "Joint measurements of NDVI and crop production data-set related to combination of management zones delineation and nitrogen fertilisation levels," *Data Br.*, vol. 28, p. 104968, 2020, doi: <https://doi.org/10.1016/j.dib.2019.104968>.
- [24] I. M. S. Eddy *et al.*, "Integrating remote sensing and local ecological knowledge to monitor rangeland dynamics," *Ecol. Indic.*, vol. 82, pp. 106–116, 2017, doi: <https://doi.org/10.1016/j.ecolind.2017.06.033>.
- [25] J. R. Saura, A. Reyes-Menendez, and P. Palos-Sanchez, "Mapping multispectral Digital Images using a Cloud Computing software: applications from UAV images," *Heliyon*, vol. 5, no. 2, p. e01277, 2019, doi: <https://doi.org/10.1016/j.heliyon.2019.e01277>.
- [26] L. Bonnaire Rivera, B. Montoya Bonilla, and F. Obando-Vidal, "Procesamiento de imágenes multiespectrales captadas con drones para evaluar el índice de vegetación de diferencia normalizada en plantaciones de café variedad Castillo," *Cienc. Tecnol. Agropecu.*, vol. 22, no. 1, 2021, doi: 10.21930/rcta.vol22\_num1\_art:1578.
- [27] D. A. Pelaez, O. E. Gualdron, and I. Torres, "Soil characterization through remote acquisition of electromagnetic radiation," *J. Phys. Conf. Ser.*, vol. 1587, p. 12033, Jul. 2020, doi: 10.1088/1742-6596/1587/1/012033.
- [28] D. Pérez García, F. García Reina, and D. Hernández Eduardo, "Disminución de las pérdidas de energía eléctrica por distribución usando una tecnología novedosa de mediciones y control para la toma de decisiones," *revista colombiana de tecnologías de avanzada (RCTA)*, vol. 2, no. 34, pp. 144–150, Jul. 2019, doi: 10.24054/rcta.v2i34.75.
- [29] F. Librán-Embid, F. Klaus, T. Tschardtke, and I. Grass, "Unmanned aerial vehicles for biodiversity-friendly agricultural landscapes - A systematic review," *Sci. Total Environ.*, vol. 732, p. 139204, 2020, doi: <https://doi.org/10.1016/j.scitotenv.2020.139204>.
- [30] M. Balbis Morejón, F. García Reina, J. J. Cabello Eras, and V. Sousa Santos, "Caracterización energética del funcionamiento de un equipo de aire acondicionado en un local dado," *revista colombiana de tecnologías de avanzada (RCTA)*, vol. 2, no. 34, pp. 71–76, Jul. 2019, doi: 10.24054/rcta.v2i34.65.
- [31] P. J. Singh and R. de Silva, "Design and implementation of an experimental UAV network," in *2018 International Conference on Information and Communications Technology (ICOIACT)*, 2018, pp. 168–173. doi: 10.1109/ICOIACT.2018.8350739.
- [32] H. A. Kurdi *et al.*, "Autonomous task allocation for multi-UAV systems based on the locust elastic behavior," *Appl. Soft Comput.*, vol. 71, pp. 110–126, 2018, doi: <https://doi.org/10.1016/j.asoc.2018.06.006>.
- [33] B. Jiang *et al.*, "Fusion of machine vision technology and AlexNet-CNNs deep learning network for the detection of postharvest apple pesticide residues," *Artif. Intell. Agric.*, vol. 1, pp. 1–8, 2019, doi: <https://doi.org/10.1016/j.iaia.2019.02.001>.