CHILEAN JOURNAL
of AGRICULTURAL RESEARCH

RESEARCH ARTICLE

Comparison of models estimating biomass in *Cenchrus clandestinus* (Hochst. ex Chiov.) Morrone pastures at high-altitude tropical farms

Mayerling Sanabria-Buitrago[1*], Martha Patricia Valbuena-Gaona[2], Jorge Fernando Triana-Valenzuela[3], Juan Carlos Velásquez-Mosquera[3], and Joao Alveiro Alvarado-Rincón[4]

[1]Universidad de La Salle, Facultad de Ingeniería, Bogotá, Colombia.
[2]Procálculo Prosis SAS, Departamento de Investigación, Desarrollo e Innovación, Bogotá, Colombia.
[3]Universidad de La Salle, Facultad de Ciencias Agropecuarias, Bogotá, Colombia.
[4]Universidad de La Salle, Facultad de Ciencias Agropecuarias, Yopal, Colombia.
*Corresponding author (msanabria@unisalle.edu.co)
Received: 5 February 2026; Accepted: 21 May 2026, Available online: 03 August 2026. doi:10.4067/S0718-58392026000500668

ABSTRACT

The use of remote sensing technologies has been increasingly recognized as a reliable tool for biomass estimation in tropical pastures. This study explores the application of predictive models that integrate agroclimatic variables, complementary field data, and satellite image processing to estimate biomass in rotational grazing systems of Kikuyu grass (*Cenchrus clandestinus* (Hochst. ex Chiov.) Morrone) in the Colombian high tropics. For this purpose, PlanetScope satellite imagery was used in combination with field-collected data, such as forage type and age, temperature, humidity, precipitation, and vegetation spectral index. Four regression models were compared: Generalized Linear Model with Poisson distribution, Multiple Linear Regression, Multilayer Perceptron Artificial Neural Network (MLP), and Random Forest Regression, in order to identify the most accurate statistical model for estimating forage biomass. Among the evaluated models, the MLP demonstrated the best performance (R^2 = 0.7). Variable importance analysis highlighted the Green Normalized Difference Vegetation index (NDVI), forage age, and the Green Chlorophyl Index (CIGreen) as the most relevant predictors. The results of this study suggest that remote sensing techniques, when complemented with forage sampling, offer an effective strategy for large-scale biomass monitoring and pasture management. The implementation of these models can enhance decision-making in livestock systems in the high tropics, improving grazing planning and forage resource management.

Key words: Grazing, Kikuyu, NDVI, predictive models, remote sensing, satellite images.

INTRODUCTION

Kikuyu grass (*Cenchrus clandestinus* (Hochst. ex Chiov.) Morrone) is the most widespread species in livestock systems in the high-altitude tropics of Colombia due to its adaptability and persistence (Vargas-Martínez et al., 2018). Dairy production requires a consistent supply of high-quality and quantity forage to ensure animal productivity, making efficient biomass quantification crucial. Traditional estimation methods (cut-and-weight) are accurate but labor-intensive and have low spatial coverage (Pereira et al., 2022).

To overcome these limitations, remote sensing and satellite image interpretation emerge as a feasible option, since the systematic collection of Earth's images enables land cover characterization and the extraction of valuable information for estimating biomass production. In this regard, the use of remote technologies is a viable alternative to obtain quick and accurate estimates of above-ground biomass at a landscape scale (Frank et al., 2025). On the other hand, climate variability directly influences forage availability. Hence, accessing information from weather stations allows for real-time monitoring of variations in temperature, humidity, and precipitation that affect the growth and availability of pasture biomass. By incorporating agroclimatic variables into biomass estimation models, the prediction accuracy increases (Nuñez et al., 2021). Frank et al. (2025) note that incorporating variables such as annual mean temperature and the normalized difference vegetation index (NDVI) into models for predicting grass biomass production enhances their predictive power, with improvements spanning 5.47% to 20.19%.

Alternatively, the growing data production in several disciplines has generated a pressing need for better analytical tools. Machine learning algorithms have gained wide acceptance in biomass estimation due to their ability to handle both complex systems and nonlinearity, typically in environmental variables.

That is why statistical and machine learning models have become highly viable options for conducting analyses, generating predictions, and thereby improving decision-making (Benos et al., 2021). This study focuses on four widely used models: Generalized Linear Models (GLM), Multiple Linear Regression (MLR), Multilayer Perceptron (MLP), and Random Forest Regression (RFR). These models were selected for their frequent use in scientific literature and their proven adaptability across different thematic contexts (Benos et al., 2021; Chen et al., 2022; El-Khalafy et al., 2025).

Traditional linear regression assumes a response variable that often follows a normal distribution; however, not all real-world variables conform to this. A key advantage of GLMs lies in their flexibility to accommodate response variables that do not follow a normal distribution, making them applicable to a broad class of response variables beyond normality, as commonly encountered in biological and agricultural research. The GLMs employ a link function to connect the linear combination of predictors to the expected value of the response variable, enabling the representation of non-linear relationships within a linear framework (Wüthrich and Merz, 2023). This link function provides an effective way to adapt the linear predictor to the distributional nature of the response variable, extending the scope of linear modeling to capture relationships that ordinary least-squares regression cannot adequately represent.

The MLR serves a dual purpose: Predicting the dependent variable from a set of independent variables and describing the relationships among them, making it a valuable tool for both applied and theoretical research. However, MRL presents well-known limitations, including multicollinearity when there are more variables than observations, or when the dataset contains outliers that can reduce predictive reliability (Genç and Mendeş, 2024).

The MLP is a type of artificial neural network designed to learn complex nonlinear relationships between input and output variables through the combined action of hidden layers and nonlinear activation functions. The MLPs are widely employed in pattern recognition tasks and constitute a fundamental component of deep learning architectures, which have demonstrated remarkable performance across diverse and complex tasks (Boniecki et al., 2020; Nosratabadi et al., 2021).

Finally, RFR is a machine learning method that combines the predictions of multiple decision trees to achieve greater accuracy and stability than a single decision tree. By averaging predictions of many correlated decision trees, RFR reduces variance and mitigates the risk of overfitting, yielding models that are both robust to outliers and tolerant to missing values (Iranzad and Liu, 2024). Additionally, it can provide measures of the relative importance of each feature, helping to identify the most influential predictors.

Based on the complementary strengths of these algorithms and the demonstrated potential of integrating remote sensing with field-based data, the present study combined pasture variables, vegetation indices derived from satellite imagery, and agroclimatic data to predict forage biomass in Kikuyu-dominated grasslands within a specialized dairy farm in the high-altitude tropics of Colombia.

MATERIALS AND METHODS

The study was conducted at the Santa María Research and Training Centre (RTC) of Universidad de La Salle, an 80 ha specialized dairy production system located in the Aposentos district, municipality of Sopó-Briceño, Cundinamarca, Colombia. (Figure 1). The farm is situated at 4°57'44.71" N, 73°59'37.49" W, at 2630 m a.s.l., with a mean annual temperature of 12 °C and a mean relative humidity of 67% (Climate-Data.org, 2025).

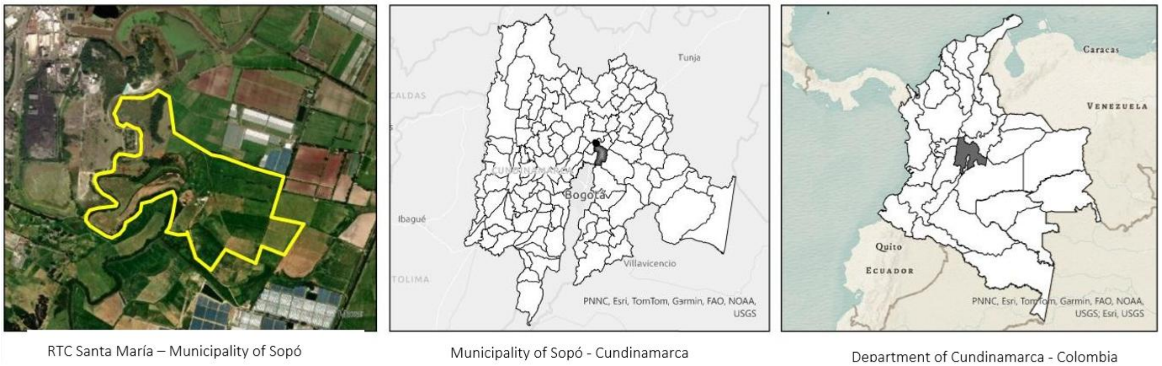

Figure 1. From the first to the last panel of the figure, the localization scale of the study area is progressively expanded. The first panel shows the study site, the second panel represents the regional-scale unit, and the final panel corresponds to the national-scale unit.

Forage samples were collected from a 25 ha area comprising nine paddocks dominated by 90% Kikuyu grass (*Cenchrus clandestinus* (Hochst. ex Chiov.) Morrone), along with *Lolium* spp. and *Trifolium* spp. Within each paddock spatial variability in biomass was captured by placing 25 × 25 cm sampling frames at representative locations; forage was cut 5 cm above ground level, weighed fresh in the field and subsequently dried at 64 °C for 72 h to determine DM biomass (kg DM m^{-2}). Agroclimatic data (precipitation, air temperature, and humidity) were recorded at 15 min intervals using a Vantage Pro2 automatic weather station (Davis Instruments, Hayward, California, USA) from August 2023 to July 2024, capturing wet and dry season periods and capturing inter-annual climatic variability. The resulting field measurements served as the reference dataset for calibrating and validating the satellite-based biomass estimation models, establishing empirical linkages between observed DM values and the corresponding spectral and agroclimatic predictors.

PlanetScope satellite images

Images downloaded from the Planet constellation (Planet Labs, San Francisco, California, USA) were used and offer a spatial resolution of 3 m, a spectral resolution of 8 bands, and a daily temporal resolution. This temporal frequency is beneficial in areas with variable cloudiness, as it increases the likelihood of obtaining cloud-free images.

The availability of 8 spectral bands, spanning from coastal blue to near-infrared (NIR), provides a robust foundation for computing spectral indices with greater accuracy and sensitivity to vegetation biophysical properties. This spectral variety (Figure 2), supported by high spatial resolution, enables finer discrimination among different land cover types and the assessment of specific biophysical traits, such as chlorophyll content, canopy structure, and soil moisture. These characteristics make PlanetScope imagery particularly well suited for detecting subtle variations in vegetation cover, supporting detailed landscape characterization in heterogeneous pastoral environments (Frazier and Hemingway, 2021).

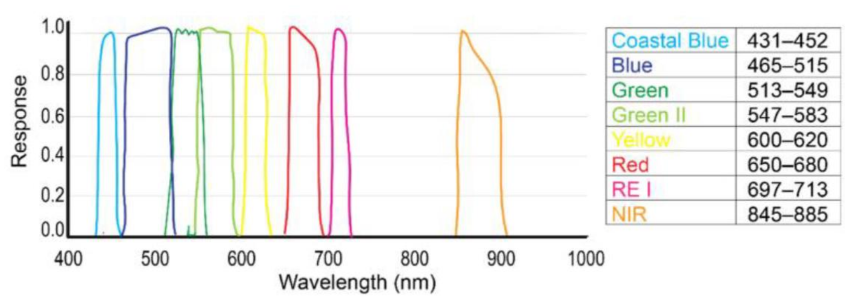

Figure 2. Spectral resolution of the SuperDove sensors from the PlanetScope constellation. Each spectral band is distinguished by a specific color, and the accompanying legend indicates the corresponding wavelength range for each band. Source: Planet imagery product specifications.

Cloud cover is a persistent constraint in the study area, resulting in the recovery of 153 cloud-free images from a total of 285 field sampling dates, representing a 54% data retention rate. While the remaining dates 132 dates were excluded from analysis due to cloud cover. Thus, the final dataset comprised 153 observations, each consisting of a measure DM biomass value (response variable) paired with its corresponding set of predictors (spectral indices derived from PlanetScope imagery, agroclimatic variables, and direct field measurements of forage characteristics).

Spectral index obtained from satellite images

The analysis of satellite imagery is based on the surface reflectance of different land covers at various wavelengths. The relationship between photosynthetic activity and vegetation's spectral response has been extensively studied, resulting in vegetation indices that emphasize specific attributes such as vigor, growth, biomass, and chlorophyll content (Chuvieco, 2020).

Each surface reflects and absorbs light uniquely, enabling the generation of indices that express biomass status and provide insights into vegetation's physiological and structural conditions (Jota et al., 2024). Plants strongly absorb radiation in the red region (600-700 nm) due to chlorophyll and reflect intensely in the near-infrared (700-1300 nm) because of their cellular structure; thus, characteristics such as phenological stage, leaf anatomy, and chemical composition influence reflectance patterns that define vegetation's spectral behavior (Chuvieco, 2020).

In precision pasture management, several greenness spectral indices have been developed to evaluate the spatial and temporal dynamics of forage availability in livestock systems, combining multiple spectral bands to reflect vegetation properties such as biomass, absorbed radiation, and chlorophyll content (Díaz Giraldo et al., 2024). The normalized difference vegetation index (NDVI) is perhaps the most widely used indicator to measure biomass health and density. This index is based on the difference between light reflectance in the red and NIR spectra, expressed as:

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

Healthy vegetation reflects strongly in the NIR and absorbs in the red region; therefore, NDVI values close to 1 indicate dense and vigorous foliage, while negative values suggest stress or bare soil. However, this index is sensitive to soil moisture, which may affect biomass estimation accuracy, and it tends to saturate under high vegetation density due to strong chlorophyll absorption. To overcome these limitations, alternative indices such as the enhanced vegetation index (EVI) have been developed, incorporating the blue band to correct atmospheric effects and reduce saturation in dense vegetation areas, thus improving biomass detection. This index has been applied, for example, in the Mongolian steppes to estimate forage availability (Amarsaikhan et al., 2023).

Another proposed index is the Green NDVI, which is more sensitive to chlorophyll variation and useful in advanced vegetation stages, as it exhibits lower saturation than NDVI (Chuvieco, 2020). The calculation of Green NDVI involves the NIR band and the green band of the electromagnetic spectrum, as shown in the following equation.

$$\text{Green NDVI} = \frac{\text{NIR} - \text{Green}}{\text{NIR} + \text{Green}}$$

Thus, values below 0 represent a loss of plant vigor, likely due to water stress, nutrient deficiency, or the presence of pests. In contrast, values close to 1 indicate high photosynthetic activity and good chlorophyll content in the plant, and therefore a vigorous condition of the vegetation cover.

Estimating pasture productivity through remote sensing is a well-studied approach for assessing spatial and temporal variations in grassland cover using spectral imagery (Théau et al., 2021). In the Peruvian tropics, NDVI- and EVI-based models have achieved determination coefficients above 60% (Nuñez et al., 2021). In Colombia, research on this topic remains limited. Posada-Asprilla et al. (2019) reported using a multispectral sensor to estimate both yield and chemical composition of Kikuyu grass, with coefficients of determination above 0.90, while Díaz Giraldo et al. (2021) demonstrated the effectiveness of remote sensing in monitoring Brachiaria pastures via NDVI-forage regression.

Predictive models

Regression modeling provides a mathematical framework for quantifying the relationship between a set of predictor variables (independent variables, X) and a response variable (dependent variable, Y), and constitutes the analytical foundation of the four approaches evaluated in this study (Saputro et al., 2021).

A dataset of 153 observations was assembled by integrating field-collected forage measurements with spectral indices derived from PlanetScope imagery. Each observation included ten (10) independent variables: (1) Forage age (days since last grazing event); (2) air temperature (°C); (3) relative humidity (%); (4) precipitation (mm); (5) temperature-humidity-sun-wind index (THSW); (6) Normalized Difference Vegetation Index (NDVI); (7) Normalized Difference Red Edge Index (NDRE); (8) Green NDVI; (9) Ratio Vegetation Index (RVI); and (10) Green Chlorophyll Index (CIGreen).

A hyperparameter optimization process was implemented to identify the most suitable configuration for each of the evaluated regression models. An automated iterative procedure was developed using a while loop, in which hyperparameters were randomly selected within predefined ranges. Each generated combination was subsequently used to train the model and evaluate its predictive performance. Model acceptance was based on a performance threshold greater than 70%, which was established as the minimum criterion for precision and reliability.

Only those configurations that exceeded the defined threshold and demonstrated adequate statistical robustness were retained for subsequent comparative analysis. This independent optimization strategy for each algorithm ensured that every regression architecture—applied separately across the four selected methods—reached its optimal parameter configuration, allowing the models to adapt effectively to the dimensionality and intrinsic variability of the training dataset.

The optimal hyperparameter combination for each model was selected according to overall predictive performance and stability across iterations, thereby ensuring the reproducibility, consistency, and robustness of the inferred results.

The dataset was partitioned into two subsets for model development and evaluation. The training set (70% of observations, $n = 107$) was used to fit each by estimating the relationship between the independent variables and forage biomass response variable, while the validation set (30%, $n = 46$) comprised observations withheld from training to provide an unbiased estimate of predictive performance on unseen data. The four regression models evaluated in this study were:

Generalized linear model based on Poisson family (GLM Poisson). When the data of the response variable is not normally distributed and does not have linear parameters, nonlinear regression models are used. To model nonlinear data, the GLM is applied. The GLM is very helpful for modeling the relationship between a response variable and predictor variables when the response variable does not follow a normal distribution but instead belongs to the exponential family of distributions. The three parts of GLM are the random component, the systematic component, and the link function. One type of nonlinear regression modeled with GLM is Poisson regression (Saputro et al., 2021). The family of Poisson distributions is more adaptable than other models, as it can handle a wide variety of skewed data.

In GLM Poisson, overdispersion leads to an underestimated standard error and, consequently, invalid conclusions. When analyzing data with overdispersion using Poisson regression models, information is missing because the dispersion parameters are not modeled in the formulated regression model. Dispersion parameters refer to those that appear as a result of the absence of equidispersion (Saputro et al., 2021).

Multiple linear regression (MLR). Linear regression models have linear parameters and are typically distributed. The MLR is a statistical technique used to predict the outcome of a response variable using multiple explanatory variables. The objective of MLR is to model the linear relationship between the independent variables (x) and the dependent variable (y) that will be analyzed (Saputro et al., 2021).

Artificial neural networks: Multilayer perceptron (MLP). The MLP is an artificial neural network implemented for modeling nonlinear data and complex problems. The MLP consists of input, output, and hidden layers. Each layer is composed of elementary units called neurons or nodes. Training based on backpropagation is selected due to its simplicity, wide application, and effectiveness for nonlinear data modeling (Shams et al., 2021). The network was parameterized with 100 hidden layers and a maximum of 10 000 iterations.

Random forest regression (RFR). The RFR is a machine learning algorithm that uses an ensemble of decision trees to predict outcomes based on a set of predictors. It involves constructing decision trees from a training dataset to generate the average prediction results (in a regression framework) of individual trees. In RFR, many decision trees are created, with each tree grown using a unique bootstrap sample of the training data (Wang et al., 2021).

All statistical analyses and predictive models were implemented using Python (version 3.11). using the following libraries: Scikit-learn for MLR, RFR, and MLP models; Statsmodels for the GLM; Pandas and NumPy for data processing; and Matplotlib and Seaborn for data visualization. Satellite image preprocessing and spectral index calculation were performed using ArcGIS Pro (Esri, Redlands, California, USA) and Google Earth Engine.

RESULTS AND DISCUSSION

To evaluate model performance, two analytical scenarios were designed: The first included only spectral indices derived from satellite imagery as dependent variables, while the second incorporated supplementary field-collected variables related to forage characteristics, as well as agroclimatic data recorded by the on-site weather station. In the first phase, models were created based on spectral data, where the independent variables were the calculated vegetation indices and the dependent variable was the above-ground DM forage biomass.

The results (Table 1) showed a low correlation between spectral variables and biomass in the first scenario (R^2 = 0.44 for multilayer perceptron artificial neural network [MLP], mean squared error [MSE] = 0.004), with MLP outperforming multiple linear regression (MLR), random forest regression (RFR), and generalized linear models (GLM) Poisson (Table 1). When incorporating additional variables such as forage age and agroclimatic data in the second scenario, predictive performance improved notably across all models, with MLP achieving the highest accuracy (R^2 = 0.7, MSE = 0.003). The limited predictive capacity observed in the first scenario, highlights the limitations of optical remote sensing in high-productivity tropical canopies, since spectral indices such as normalized difference vegetation index (NDVI) tend to saturate under dense vegetation cover, reducing their sensitivity to biomass variation beyond certain threshold (Chuvieco, 2020). The incorporation of forage age and agroclimatic variables in the second scenario addressed this limitation by capturing the influence contrasting seasonal conditions as rainy and dry periods, as well as inter-annual variability associated with El Niño and La Niña events, providing a more accurate representation of environmental and management drivers of biomass dynamics throughout the 12 mo monitoring period.

Table 1. Comparison of predictive models for estimating the amount of forage biomass according to the mean squared error (MSE) and the coefficient of determination (R^2). GLM Poisson: Generalized Linear Model based on Poisson family.

	Spectral index		Spectral index, forage age, and meteorological data	
Model	MSE	R^2	MSE	R^2
GLM Poisson	0.005	0.19	0.035	0.40
Multiple Linear Regression	0.004	0.36	0.003	0.61
Multilayer Perceptron	0.004	0.44	0.003	0.70
Random Forest Regression	0.004	0.29	0.003	0.55

Among individual model results in the second scenario, the GLM Poisson (Figure 3) recorded the lowest performance (R^2 = 0.4, MSE = 0.035) and, even when integrating spectral indices, agroclimatic variables, and forage age. The MLR showed a notable improvement recording an MSE = 0.003 with an R^2 = 0.61 (Figure 4). The MLP achieved the highest predictive accuracy (R^2 = 0.7, MSE = 0.003; Figure 5).

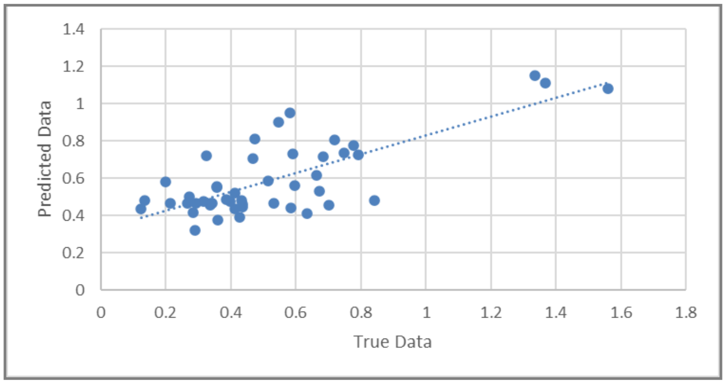

Figure 3. Scatter plot comparing observed (true) data and predicted values obtained with the Poisson Generalized Linear Model (GLM). Each point represents an individual observation, where the x-axis corresponds to the observed values and the y-axis to the model predictions. The dotted line represents the fitted linear trend, illustrating the overall agreement between predicted and observed data across the range of values.

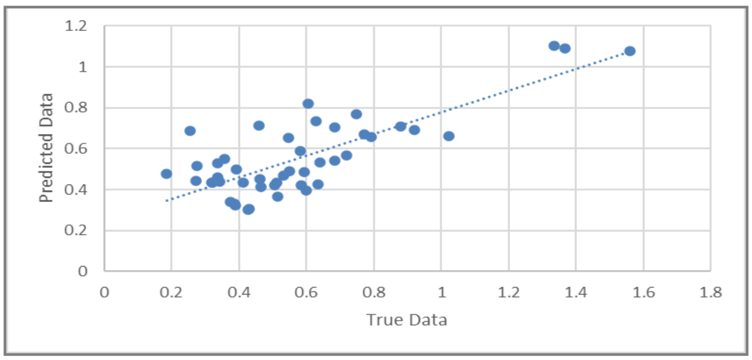

Figure 4. Scatter plot comparing observed values and values predicted by the Multiple Linear Regression (MLR) model. Each point represents an individual observation, with the x-axis corresponding to observed values and the y-axis to model predictions. The dotted line represents the fitted linear trend, allowing assessment of the agreement between observed and predicted data.

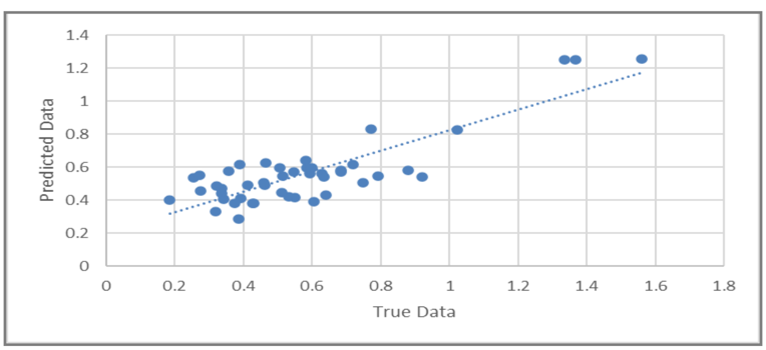

Figure 5. Scatter plot comparing observed values and values predicted by the Multilayer Perceptron (MLP) model. Each point represents an individual observation, with the x-axis corresponding to observed values and the y-axis to model predictions. The dotted line represents the fitted linear trend, allowing assessment of the agreement between observed and predicted data.

The poor performance of the GLM Poisson model (R^2 = 0.40; MSE = 0.035) suggests that the intrinsic assumptions of the Poisson distribution do not align with the nature of the evaluated biomass data. This limitation is primarily attributed to the dispersion of the field measurements.

Unlike MLP, the GLM relies on a fixed parametric structure that fails to adequately capture the complex and nonlinear interactions between agroclimatic variables and spectral indices, particularly in the presence of multicollinearity among vegetation descriptors (e.g., NDVI and Green NDVI). Consequently, the inability of the GLM to account for the heterogeneity of high-altitude tropical grasslands confirms that generalized linear approaches are insufficient to represent the dynamics of complex biological systems when compared with deep learning algorithms.

The consistent superiority of MLP across both scenarios is attributed to its capacity to capture complex, nonlinear patterns among predictors in the data RFR yielded intermediate performance (R^2 = 0.55, MSE = 0.003; Figure 6), consistent with the known sensitivity of tree-based methods to correlate predictors (Chen et al., 2022).

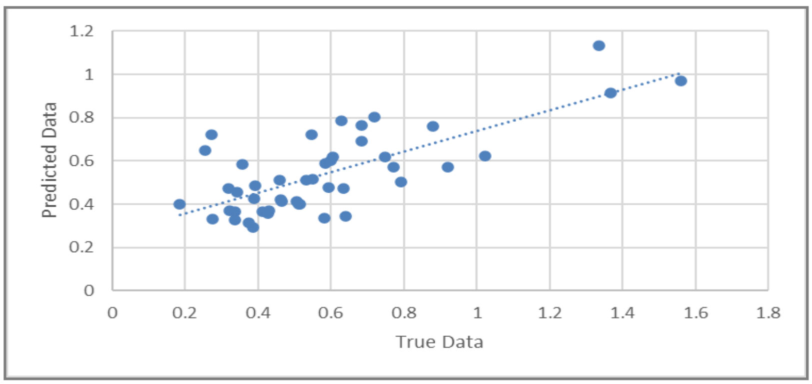

Figure 6. Scatter plot comparing observed values and values predicted by the Random Forest Regression (RFR) model. Each point represents an individual observation, with the x-axis corresponding to observed values and the y-axis to model predictions. The dotted line represents the fitted linear trend, allowing assessment of the agreement between observed and predicted data.

Therefore, the inherent complexity of biological systems, such as forage growth dynamics influenced by multiple interacting factors, may favor nonlinear models like MLP over linear approaches. However, the specific characteristics of the dataset, including the correlation between the attributes used and the target variable (biomass estimation), may lead to different models achieving superior performance in various contexts.

Building on the performance of the second scenario models, a variable importance analysis was conducted to rank each predictor's contribution to biomass estimation (Figure 7). The three most relevant variables were the Green NDVI, forage age, and the Green Chlorophyl Index (CIGreen), while meteorological variables ranked lowest in terms of relative importance. This outcome is likely attributable to the limited representation of climatic data form a single weather station, which cannot capture paddock-level variability. Although agroclimatic variables improved overall model performance, their marginal contribution within the MLP was limited, suggesting that spatially distributed climatic inputs could further enhance predictive accuracy.

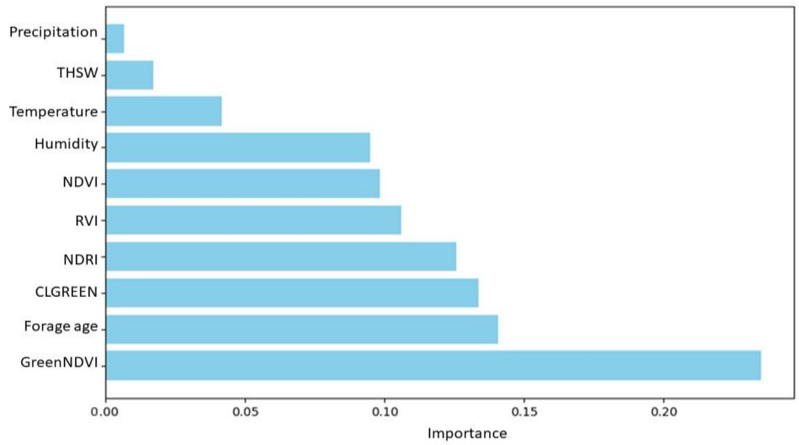

Figure 7. Relative importance of the predictor variables used in the biomass estimation model, arranged in ascending order according to their contribution to predictive performance. The figure illustrates the influence of each variable on model estimation, enabling the identification of the most relevant factors driving biomass prediction accuracy. GreenNDVI exhibited the highest relative importance, followed by forage age and the Chlorophyll Green Index (CIGREEN), highlighting the strong influence of vegetation vigor and chlorophyll-related spectral responses on biomass estimation. In contrast, meteorological variables such as precipitation and temperature showed comparatively lower contributions. THSW: Temperature-humidity-sun-wind index; NDVI: Normalized Difference Vegetation Index; RVI: Ratio Vegetation Index; NDRI: Normalized Difference Red Edge Index; CIGREEN: Chlorophyll Green Index.

Images and vegetation index

The results of this study confirm the general agreement with previous research on the usefulness of the spectral indices for estimating biomass in pastures intended for livestock production (Jota et al., 2024). For example, Posada-Asprilla et al. (2019) identified NDVI as the most effective index for estimating biomass in Kikuyu grasslands in Colombia. While, Netsianda and Mhangara (2025) reported strong NDVI performance in South African grasslands, supporting its reliability across different ecological settings. However, Bernal et al. (2023) found no strong NDVI-biomass relationship in Kikuyu pastures of the Bogotá savanna, Colombia, a finding consistent with the saturation behavior of this index under dense canopy conditions (Chuvieco, 2020). These contrasting results highlight the need for multi-variable, site specific calibration approaches, that integrate additional data sources beyond broadband spectral indices to achieve reliable biomass estimation.

In this regard, the Green NDVI index emerges as a more sensitive alternative. Unlike NDVI, Green NDVI replaces the red band with the green band, improving its sensitivity to chlorophyll variations and reducing saturation under high vegetation productivity conditions (Théau et al., 2021).

Studies conducted in Mongolia (Amarsaikhan et al., 2023), and using unmanned aerial vehicles (UAV) imagery (Pereira et al., 2022) have reported strong correlation for biomass estimation using Green NDVI, supporting its selection as the most influential predictor in the present model. Additionally, Pereira et al. (2022) reported that the prediction of aerial forage biomass achieved better accuracy when using combined PlanetScope and Sentinel-2A imagery than when using images from a single platform. This could enhance the precision of future studies.

This finding represents a relevant contribution of the current study. While most biomass estimation research in Colombian Kikuyu pastures has relied primarily on NDVI (Posada-Asprilla et al., 2019; Bernal et al., 2023), the present results suggest that CIGreen and Green NDVI offer superior discriminative capacity in dense, high-altitude tropical canopies, and should be prioritized in future monitoring frameworks. In this context, it is also worth noting the growing interest in UAV-based approaches using RGB cameras as low-cost alternatives for forage estimation. However, visible-spectrum indices derived from RGB imagery (Red-Green-Blue Vegetation Index [RGBVI]) have shown very low predictive capacity for biomass in Kikuyu pastures (R^2 = 0.02;

Bernal et al., 2023), reflecting the spectral limitations of RGB sensors in capturing the near-infrared reflectance patterns that are most informative for dense-canopy biomass estimation. Complementary approaches based on UAV-mounted RGB cameras and visible-spectrum algorithms have been proposed in Colombia for farm-level forage assessment (Ospina-Rivera et al., 2020, 2021). However, further validation under controlled and replicated experimental conditions is needed to establish their predictive reliability across diverse production environments. The multispectral indices identified in the present study such as Green NDVI and CIGreen can be computed from PlanetScope imagery using open-source GIS platforms (QGIS or Google Earth Engine), offering a reproducible and scalable alternative for operational pasture monitoring.

Most influential attributes in the models

By ranking the contribution of each variable in the prediction process, the model positioned Green NDVI as the most relevant index for estimating biomass in high-altitude tropical pastures. and also, as the attribute with the most significant influence on the response variable according to the variable importance analysis. This trend is consistent with Théau et al. (2021), who identified Green NDVI as the primary variable in biomass prediction models, achieving high coefficients of determination (R^2 = 0.80 for fresh biomass and R^2 = 0.66 for dry biomass), particularly in scenarios where biomass did not exceed 3 kg m^{-2} fresh matter. The relevance of Green NDVI over broadband NDVI in the present study aligns with the recognition that chlorophyll-sensitive indices offer greater discriminative capacity in dense, high-productivity canopies where traditional NDVI tends to saturate (Chuvieco, 2020). The importance of combining spectral indices with field-based measurements is further supported by Cândido et al. (2025), who found that integrating proximal sensing variables (including vegetation height) with NDVI and Modified Soil Adjusted Vegetation Index 2 (MSAVI2) from Landsat 7 and Sentinel-2 imagery improved biomass prediction to R^2 = 0.86. This is consistent with the prominent role of forage age observed in the present study, reinforcing the value of incorporating field-collected management variables alongside spectral data in biomass estimation models.

Forage age was the second most relevant predictor. consistent with Khabbazan et al. (2019), who estimated ryegrass biomass using Sentinel-1 imagery and attributed the effect to pasture growth dynamics and age, which is particularly sensitive to the increase in fresh biomass during vegetative stages and decreases during senescence as the vegetation's water content declines. The relevance of CIGreen as the third-ranked predictor further confirms the utility of chlorophyll-based indices for biomass estimation in Kikuyu dominated systems, where canopy greenness and chlorophyll content vary substantially across the rotational grazing cycle.

Meteorological variables ranked lowest in the model. Humidity has a greater influence than rainfall in the model, meaning that the relative contribution of each variable is critical for determining model sensitivity. This pattern is consistent with findings from Australia reported by Aditya (2022), who used MLP neural networks to estimate biomass in *Cenchrus ciliaris* pastures and showed that the importance of rainfall and humidity varies with the climatic season evaluated (wet or dry).

Model accuracy

Most machine learning algorithms provide some form of feature contribution to the results, but their performance can be strongly affected by the structure of the input data. In particular RFR, may perform sub optimally l when predictors are highly correlated (Chen et al., 2022).

Temporal smoothing methods and gap filling strategies offer another way to improve predictive performance and stability, especially in cloud-prone environments. weighting-based algorithms, can achieve higher accuracy by assigning different weights to data with varying levels of uncertainty (Yang et al., 2019; Chu et al., 2021; Li et al., 2021) and Bayesian approaches has been successfully used to reconstruct vegetation index time-series and quantify associated uncertainty for variables such as leaf area index (Pipia et al., 2019). Recent work using Sentinel-2 time series and machine learning for pasture biomass estimation also emphasizes the importance of careful temporal pre-processing to mitigate noise and missing data, particularly in humid and highly dynamic grazing systems (Chen et al., 2022; Furnitto et al., 2025).

Comparative studies in other production systems confirm that there is no single universally superior algorithm for pasture biomass estimation. In Brazilian maize, Filgueiras et al. (2019) reported that the RFR model outperformed several alternative machine learning methods (e.g., support vector regression, Bayesian regularized neural networks, and gradient boosting machines) for NDVI-based biomass prediction, whereas in tropical pastures Bonini-Neto et al. (2023) found that artificial neural networks (ANN), specifically MLP were particularly effective for estimating Marandu grass yield. Evidence from temperate and subtropical dairy systems also indicates that ANN can achieve very high accuracy when combining multisensor optical and synthetic aperture radar (SAR) data. For example, Vahidi et al. (2023) reported R^2 values around 0.83 for pasture biomass estimation using Sentinel-1 (SAR) and Sentinel-2 time series, with ANN outperforming Random Forest (RF) and support vector regression in most cases. In contrast, Irisarri et al. (2025) showed that RF models calibrated with Sentinel-2 imagery and proximal near-infrared measurements achieved R^2 between 0.77 and 0.86 for forage quality traits in temperate grasslands, highlighting that tree-based methods can be highly competitive when rich, well-structured training datasets are available.

The results obtained from the most predictive models in the present study are therefore only partially comparable to those reported by Zwick et al. (2024), who found in Colombian tropical grasslands that the best performing algorithm varied according to site and response variable, with Regularized Random Forest, Partial Least Squares, Random Forests, Bagged Multivariate Adaptive Regression, and Bayesian Regularized Neural Networks identified as top performing algorithms. Collectively, these studies reinforce that model performance is strongly context-dependent and that the relative advantage of MLP observed here is consistent with scenarios characterized by complex, nonlinear interactions among spectral, structural, and climatic predictors.

CONCLUSIONS

This study demonstrated that spectral indices derived from PlanetScope imagery alone have limited capacity to predict forage biomass in Kikuyu grass (*Cenchrus clandestinus* (Hochst. ex Chiov.) Morrone) pastures ($R^2 \leq 0.44$ across all models tested), reflecting the inherent limitations of spectral data in capturing the full complexity of pasture growth dynamics in high-altitude tropical environments subject to variable cloud cover. The integration of forage age and agroclimatic variables substantially improved predictive performance across all models, with the Multilayer Perceptron (MLP) achieving the highest accuracy (R^2 = 0.7), followed by Multiple Linear Regression (R^2 = 0.61), Random Forest Regression (R^2 = 0.55), and Generalized Linear Models Poisson (R^2 = 0.40). Variable importance analysis identified Green Normalized Difference Vegetation Index (Green NDVI), forage age and Green Chlorophyl Index as the dominant predictors, confirming that chlorophyll-sensitive spectral indices and temporal management variables are more informative than broadband indices and meteorological data for biomass estimation in dense Kikuyu canopies. These findings establish a reproducible methodological baseline for operational pastures monitoring in Colombian high-altitude dairy systems, where traditional cut and weight sampling remains the predominant estimation method.

The practical value of MLP model lies in its potential integration into pasture management decisions. By combining daily PlanetScope imagery with paddock-level forage age records and open-source geographic information system tools, farm managers can generate spatially explicit biomass maps to supports rotational grazing decisions. Furthermore, incorporating additional predictors such as soil properties, terrain attributes, and solar radiation should be explored to increase predictive accuracy. Finally, validating the models across different locations and forage species would be valuable to assess their generalization and robustness under diverse environmental and ecological conditions. In this context, future applications should prioritize the development of operational monitoring platforms integrating remote sensing, machine learning, and automated weather data to support sustainable pasture-based dairy systems in tropical regions.

Author contribution

Conceptualization: M.S.B., M.V.G. Methodology: J.T.V., M.V.G. Software: M.V.G. Validation: M.S.B., M.V.G., J.V.M. Formal analysis: M.S.B., M.V.G., J.V.M. Investigation: M.S.B., M.V.G., J.V.M. Resources: J.T.V., J.A.R. Data curation: J.T.V., J.A.R. Writing-original draft: M.S.B., M.V.G., J.V.M. Writing-review & editing: J.A.R., J.V.M., J.T.V. Supervision: J.V.M. Project administration: J.V.M. All coauthors reviewed the final version and approved the manuscript before submission.

Acknowledgement

This research was funded by the Vice-Provost Office for Research and Transfer of Universidad de la Salle, Colombia, Procalculo Prosis SAS, and the Australian Embassy in Colombia, Australian-AID.

References

Aditya, M.I. 2022. Analyzing the influence of climate variables on native pasture phenology in the Fitzroy Basin, Australia. Papers in Applied Geography 8(4):434-451. doi:10.1080/23754931.2022.2081815.

Amarsaikhan, E., Erdenebaatar, N., Amarsaikhan, D., Otgonbayar, M., Bayaraa, B. 2023. Estimation and mapping of pasture biomass in Mongolia using machine learning methods. Geocarto International 38(1):2195824. doi:10.1080/10106049.2023.2195824.

Benos, L., Tagarakis, A., Dolias, G., Remigio, B., Kateris, D., Bochtis, D. 2021. Machine learning in agriculture: A comprehensive updated review. Sensors 21(11):3758. doi:10.3390/S21113758.

Bernal, A., Acero, A., Gutiérrez, A.F. 2023. Estimación de la biomasa en una pradera de kikuyo mediante el uso de sensores remotos. Revista de Investigación Agraria y Ambiental 15(1):175-190. doi:10.22490/21456453.6657.

Boniecki, P., Zaborowicz, M., Pilarska, A.A., Piekarska-Boniecka, H. 2020. Identification process of selected graphic features apple tree pests by neural models type MLP, RBF and DNN. Agriculture 10(6):218. doi:10.3390/agriculture10060218.

Bonini-Neto, A., Moreira, A., Bonini, C.S.B., Campos, M., Andrighetto, C. 2023. Fuzzy logic and artificial neural network perceptron multi-layer and radial basis in estimating Marandu grass yield in integrated systems. Communications in Soil Science and Plant Analysis 54(21):2965-2976. doi:10.1080/00103624.2023.2252839.

Cândido, B., Mindala, U., Ebrahimy, H., Zhang, Z., Kallenbach, R. 2025. Integrating proximal and remote sensing with machine learning for pasture biomass estimation. Sensors 25(7):1987. doi:10.3390/s25071987.

Chen, X., Zheng, H., Wang, H., Yan, T. 2022. Can machine learning algorithms perform better than multiple linear regression in predicting nitrogen excretion from lactating dairy cows. Scientific Reports 12(1):12478. doi:10.1038/s41598-022-16490-y.

Chu, D., Shen, H., Guan, X., Chen, J.M., Li, X., Li, J., et al. 2021. Long time-series NDVI reconstruction in cloud-prone regions via spatio-temporal tensor completion. Remote Sensing of Environment 264:12632. doi:10.1016/J.RSE.2021.112632.

Chuvieco, E. 2020. Fundamentals of satellite remote sensing: An environmental approach. 3rd ed. CRC Press, Boca Raton, Florida, USA. doi:10.1201/9780429506482.

Climate-Data.org. 2025. Clima Sopó: Temperatura, climograma y tabla climática para Sopó. Available at https://es.climate-data.org/america-del-sur/colombia/cundinamarca/sopo-49869/ (accessed 6 June 2025).

Díaz Giraldo, R.A., Álvarez de León, M., Pérez, O. 2021. Uso de sensores remotos en la determinación del forraje disponible de *Urochloa humidicola* cv. Llanero bajo pastoreo en la Altillanura colombiana. Tropical Grasslands-Forrajes Tropicales 9(3):376-382. doi:10.17138/TGFT(9)376-382.

Díaz Giraldo, R.A., Álvarez de León, M., Pérez, O., Parrado, S.L., Arango, M. 2024. Aplicación de 2 modelos basados en teledetección para estimar la producción primaria neta aérea en un sistema pastoril en la Altillanura colombiana. Tropical Grasslands-Forrajes Tropicales 12(1):24-35. doi:10.17138/TGFT(12)24-35.

El-Khalafy, M.M., El-Kenany, E.T., Al-Mokadem, A.Z., Shaltout, S.K., Mahmoud, A.R. 2025. Habitat suitability modeling to improve conservation strategy of two highly-grazed endemic plant species in Saint Catherine Protectorate, Egypt. BMC Plant Biology 25(1):485. doi:10.1186/S12870-025-06401-4.

Furnitto, N., Ramírez-Cuesta, J.M., Intrigliolo, D.S., Todde, G., Failla, S. 2025. Remote sensing for pasture biomass quantity and quality assessment: Challenges and future prospects. Smart Agricultural Technology 12:101057. doi:10.1016/j.atech.2025.101057.

Filgueiras, R., Mantovani, E.C., Althoff, D., Fernandes Filho, E.I., Franca da Cunha, F. 2019. Crop NDVI monitoring based on Sentinel 1. Remote Sensing 11(12):1441. doi:10.3390/RS11121441.

Frank, T., Da Silva Junior, C.A., Teodoro, P.E., Oliveira-Júnior, J.F., Bennett, J., Guo, X. 2025. Using the dry matter productivity model as an estimator of biomass production in native grassland communities. Remote Sensing Applications: Society and Environment 37:101441. doi:10.1016/J.RSASE.2024.101441.

Frazier, A.E., Hemingway, B.L. 2021. A technical review of Planet Smallsat data: Practical considerations for processing and using PlanetScope imagery. Remote Sensing 13(19):3930. doi:10.3390/RS13193930.

Genç, S., Mendeş, M. 2024. Multiple linear regression versus automatic linear modelling. Arquivo Brasileiro de Medicina Veterinária e Zootecnia 76(1):131-136. doi:10.1590/1678-4162-13071.

Iranzad, R., Liu, X. 2024. A review of random forest-based feature selection methods for data science education and applications. International Journal of Data Science and Analytics 20:197-211. doi:10.1007/S41060-024-00509-W.

Irisarri, J.G.N., Texeira, M.A., Harris, P., Pembleton, K.G. 2025. Satellite-based monitoring of forage quality in grasslands of the United Kingdom using Sentinel-2 data and random forest regression. Frontiers in Veterinary Science 12:1678123. doi:10.3389/fvets.2025.1678123.

Jota, P.H., Xiong, Y., Davies-Jenkins, S.L., Zhao, B., Drewnoski, M.E. 2024. Comparison between vegetation indices to estimate biomass in cereal rye. Journal of Animal Science 102(Supplement_3):183-184. doi:10.1093/JAS/SKAE234.215.

Khabbazan, S., Vermunt, P., Steele-Dunne, S., Arntz, L.R., Marinetti, C., van der Valk, D., et al. 2019. Crop monitoring using Sentinel-1 data: A case study from The Netherlands. Remote Sensing 11(16):1887. doi:10.3390/RS11161887.

Li, S., Xu, L., Jing, Y., Yin, H., Li, X., Guan, X. 2021. High-quality vegetation index product generation: A review of NDVI time series reconstruction techniques. International Journal of Applied Earth Observation and Geoinformation 105:102640. doi:10.1016/J.JAG.2021.102640.

Netsianda, A., Mhangara, P. 2025. Aboveground biomass estimation in a grassland ecosystem using Sentinel-2 satellite imagery and machine learning algorithms. Environmental Monitoring and Assessment 197:138. doi:10.1007/S10661-024-13610-1.

Nosratabadi, S., Ardabili, S., Lakner, Z., Makó, C., Mosavi, A. 2021. Prediction of food production using machine learning algorithms of multilayer perceptron and ANFIS. Agriculture 11(5):408. doi:10.3390/AGRICULTURE11050408.

Nuñez, J., Pizarro-Carcausto, S.E., Gutiérrez, M., Naupari, J.A. 2021. Dinámica espacio temporal de la biomasa aérea en pastizales altoandinos basado en NDVI-MODIS validado por espectrometría in situ. Revista de Investigaciones Veterinarias del Perú 32(3):e20392. doi:10.15381/rivep.v32i3.20392.

Ospina-Rivera, O.F., Anzola-Vásquez, H.J., Ayala-Duarte, O.A., Baracaldo-Martínez, A. 2020. Validación de un algoritmo de procesamiento de imágenes Red Green Blue (RGB), para la estimación de proteína cruda en gramíneas vs la tecnología de espectroscopía de infrarrojo cercano (NIRS). Revista de Investigaciones Veterinarias del Perú 31(2):e17940. doi:10.15381/rivep.v31i2.17940.

Ospina-Rivera, O.F., Anzola-Vásquez, H.J., Ayala-Duarte, O.A., Baracaldo-Martínez, A. 2021. Comparación de la fibra detergente neutra en gramíneas calculada mediante algoritmo de análisis de imágenes rojo, verde y azul vs espectroscopia del infrarrojo cercano. Revista de Investigaciones Veterinarias del Perú 32(1):e17498. doi:10.15381/rivep.v32i1.17498.

Pereira, F.R., Lima, J., Freitas, R., Dos Reis, A.A., do Amaral, L.R., Figueiredo, G., et al. 2022. Nitrogen variability assessment of pasture fields under an integrated crop-livestock system using UAV, PlanetScope, and Sentinel-2 data. Computers and Electronics in Agriculture 193:106645. doi:10.1016/J.COMPAG.2021.106645.

Pipia, L., Muñoz, J., Amin, E., Belda, S., Camps-Valls, G., Verrelst, J. 2019. Fusing optical and SAR time series for LAI gap filling with multioutput Gaussian processes. Remote Sensing of Environment 235:111452. doi:10.1016/J.RSE.2019.111452.

Posada-Asprilla, W., Medina-Sierra, M., Cerón-Muñoz, M. 2019. Estimación de la calidad y cantidad de pasto kikuyo (*Cenchrus clandestinum* (Hochst. ex Chiov.) Morrone) usando imágenes multiespectrales. Revista UDCA Actualidad & Divulgación Científica 22(1):e1195. doi:10.31910/RUDCA.V22.N1.2019.1195.

Saputro, D.R., Susanti, A., Pratiwi, N.B. 2021. The handling of overdispersion on Poisson regression model with the generalized Poisson regression model. AIP Conference Proceedings 2326:020026. doi:10.1063/5.0040330.

Shams, R., Jahani, A., Kalantary, S., Moeinaddini, M., Khorasani, N. 2021. The evaluation on artificial neural networks (ANN) and multiple linear regressions (MLR) models for predicting SO_2 concentration. Urban Climate 37:100837. doi:10.1016/J.UCLIM.2021.100837.

Théau, J., Lauzier-Hudon, É., Aubé, L., Devillers, N. 2021. Estimation of forage biomass and vegetation cover in grasslands using UAV imagery. PLOS ONE 16(1):e0245784. doi:10.1371/journal.pone.0245784.

Vahidi, M., Shafian, S., Thomas, S., Maguire, R. 2023. Estimation of bale grazing and sacrificed pasture biomass through the integration of Sentinel satellite images and machine learning techniques. Remote Sensing 15(20):5014. doi:10.3390/rs15205014.

Vargas-Martínez, J., Sierra-Alarcón, A.M., Mancipe-Muñoz, E.A., Avellaneda-Avellaneda, Y. 2018. El kikuyo, una gramínea presente en los sistemas de rumiantes en trópico alto colombiano. CES Medicina Veterinaria y Zootecnia 13(2):137-156. doi:10.21615/cesmvz.13.2.4.

Wang, F., Wang, Y., Zhang, K., Hu, M., Weng, Q., Zhang, H. 2021. Spatial heterogeneity modeling of water quality based on random forest regression and model interpretation. Environmental Research 202:111660. doi:10.1016/j.envres.2021.111660.

Wüthrich, M.V., Merz, M. 2023. Generalized linear models. p. 111-205. In Wüthrich, M.V., and Merz, M. (eds.) Statistical foundations of actuarial learning and its applications. Springer International Publishing, Cham, Switzerland.

Yang, Y., Luo, C., Huang, R., Wu, W., Sun, L. 2019. Weighted double-logistic function fitting method for reconstructing the high-quality Sentinel-2 NDVI time series data set. Remote Sensing 11(20):2342. doi:10.3390/RS11202342.

Zwick, M., Cardoso, J.A., Gutiérrez-Zapata, D.M., Cerón-Muñoz, M., Gutiérrez, J.F., Raab, C., et al. 2024. Pixels to pasture: Using machine learning and multispectral remote sensing to predict biomass and nutrient quality in tropical grasslands. Remote Sensing Applications Society and Environment 36:101282. doi:10.1016/j.rsase.2024.101282.