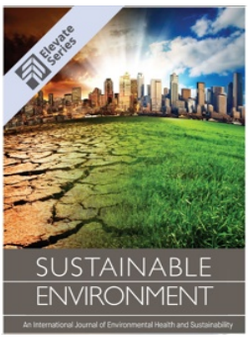


Taylor & Francis
Taylor & Francis Group

Sustainable Environment

An international journal of environmental health and sustainability

ISSN: 2765-8511 (Online) Journal homepage: www.tandfonline.com/journals/oaes21

# Monitoring ryegrass-clover grasslands with multi-spectral UAV imagery will improve the sustainability of small-and medium-sized livestock farmers in the northern Peruvian highlands

**Luis Vallejos-Fernández, Wuesley Alvarez-García, Maycol Abanto-Urbina, Felipe Gutiérrez-Arce , Eduardo Tapia-Acosta , Samuel Pizarro , Cesar Ciprian & Javier Naupari**

**To cite this article:** Luis Vallejos-Fernández, Wuesley Alvarez-García, Maycol Abanto-Urbina, Felipe Gutiérrez-Arce , Eduardo Tapia-Acosta , Samuel Pizarro , Cesar Ciprian & Javier Naupari (2026) Monitoring ryegrass-clover grasslands with multi-spectral UAV imagery will improve the sustainability of small-and medium-sized livestock farmers in the northern Peruvian highlands, Sustainable Environment, 12:1, 2623335, DOI: 10.1080/27658511.2026.2623335

**To link to this article:** https://doi.org/10.1080/27658511.2026.2623335



Published online: 04 Feb 2026.

SUSTAINABLE ENVIRONMENT
2026, VOL. 12, NO. 1, 2623335
https://doi.org/10.1080/27658511.2026.2623335

EARTH SCIENCE | RESEARCH ARTICLE

OPEN ACCESS Check for updates

# Monitoring ryegrass-clover grasslands with multi-spectral UAV imagery will improve the sustainability of small-and medium-sized livestock farmers in the northern Peruvian highlands

Luis Vallejos-Fernández[a,1], Wuesley Alvarez-García[b,1], Maycol Abanto-Urbina[a,1], Felipe Gutiérrez-Arce[c], Eduardo Tapia-Acosta[a], Samuel Pizarro[d], Cesar Ciprian[e] and Javier Naupari[e]

[a]Postgraduate Unit of Livestock Science Engineering, National University of Cajamarca, Cajamarca, Peru; [b]Directorate for Agrarian Technological Development, National Institute for Agrarian Innovation (INIA), La Molina, Lima, Peru; [c]Professional School of Veterinary Medicine, Faculty of Agrarian Sciences, National University of San Martin, Tarapoto, San Martin, Peru; [d]Research Institute for the Sustainable Development of Ceja de Selva (INDES-CES), National University Toribio Rodríguez de Mendoza. Chachapoyas, Amazonas, Peru; [e]Rangeland Ecology and Utilization Laboratory, Department of Animal Production, Universidad Nacional Agraria La Molina, La Molina, Lima, Peru

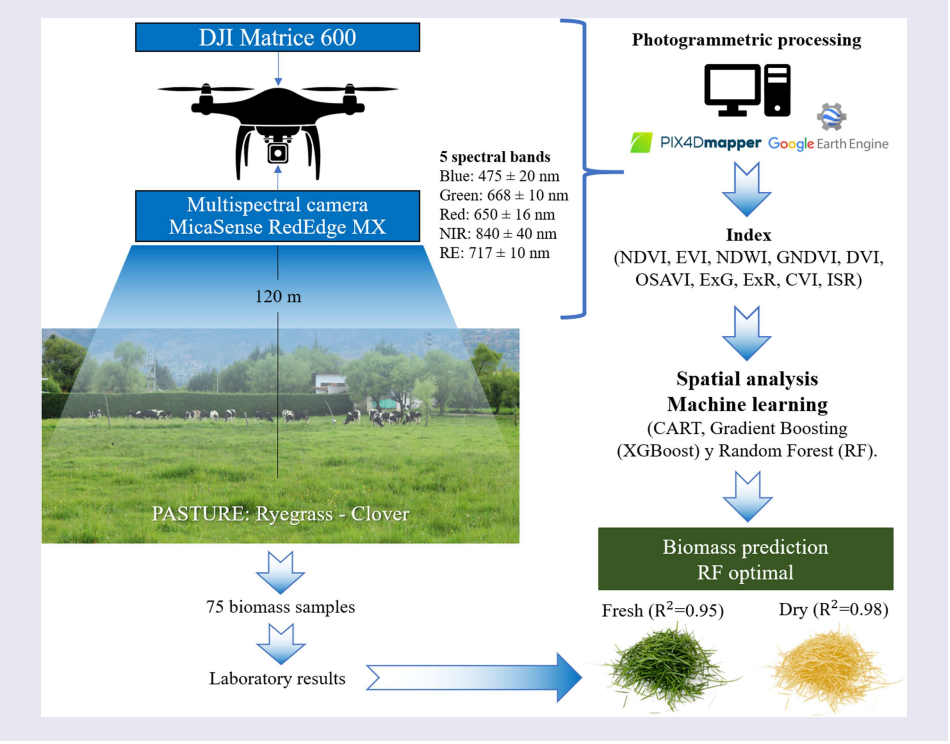
**ABSTRACT**

The underutilization of remote sensing technology has compromised sustainable forage resource management, impeding the progress of livestock farmers in the northern Peruvian highlands. To accurately predict forage biomass in six high-altitude (2600-2800 m) ryegrass (*Lolium multiflorum* Lam) -clover (*Trifolium repens*) paddocks, we applied machine learning models implemented in Google Earth Engine using spectral indices derived from UAV-based multispectral imagery captured by a Micasense RedEdge MX camera mounted on a DJI Matrice 600. A total of 75 forage samples were collected from precisely geo-referenced plots to train and validate machine learning models based on 13 spectral indices. The Random Forest (RF) model, comprising 500 trees for green forage and dry matter, demonstrated high accuracy and efficiency. UAV-based biomass prediction using GEE and ML techniques was validated, achieving $R^2$ values of 0.671 and 0.747 and low errors. By integrating UAVs, sensors, and cloud-based ML, we can decision-support potential in the inter-Andean valley. This innovative approach reduces costs, ensures high-resolution snapshot biomass assessment, and empowers producers to make data-driven decisions.

**ARTICLE HISTORY**
Received 19 February 2025
Accepted 23 January 2026

**KEYWORDS**
Aboveground biomass; ryegrass-clover; UAVs; machine learning; multispectral imaging

**CONTACT** Luis Vallejos-Fernández lvallejos@unc.edu.pe Postgraduate Unit of Livestock Science Engineering, National University of Cajamarca, Cajamarca, Peru; Javier Naupari jnaupariv@lamolina.edu.pe Rangeland Ecology and Utilization Laboratory, Department of Animal Production, Universidad Nacional Agraria La Molina, La Molina, Lima, Peru
[1]These authors contributed equally to this work and shared the first authorship.




## 1. Introduction

The dairy cattle grazing in the northern highlands of Peru, which were established in the 1950s, have been characterised by the maintenance of soil and pasture management practices that have remained unvaried. This has probably had a negative effect on the yield and morphogenetic characteristics of the ryegrass-clover association, which forms the basis of animal feed (Vallejos Fernández et al., 2020; Vallejos-Cacho et al., 2024). The research conducted in this region, which exhibits a range of altitudes (masl), necessitates a considerable investment of physical effort, time, and financial resources. Consequently, the pursuit of solutions is inherently arduous and protracted. In light of these circumstances, the deployment of technological tools is imperative for the monitoring of the 127,120 production units (Cenagro, 2012), with the objective of discerning the forage biomass throughout its growth cycle. This approach facilitates the advancement of sustainable management strategies that integrate the demand for food with the availability of forage, which ultimately translates into efficacious grazing practices (Théau et al. 2021; Villalobos-Villalobos and WingChing-Jones, 2023) and, consequently, an augmentation in production.

The term 'biomass' is defined as the production of primary energy by plants within an ecosystem (De Rosa et al., 2021). It can be assessed by various indirect techniques, such as the use of multispectral microsensors on board unmanned aerial vehicles, which are flexible tools for biomass monitoring (Batistoti et al., 2019; Goebel and Iwaszczuk, 2023; Insua et al., 2019; Villalobos-Villalobos and WingChing-Jones, 2023). The sensors capture the reflectance of different types of vegetation in the form of images, which exhibit unique spectral behaviours (Qamar and Dobler, 2023).

For example, near-infra-red radiation (760-900 nm) is highly reflected by the leaf's cellular structures, whereas visible red radiation (630-690 nm) is absorbed by chlorophyll (Kume, 2017; Van Der Tol et al., 2019). The integration of these two domains facilitates the determination of the photosynthetically active biomass across the vegetation canopy, with healthy vegetation reflecting a greater proportion of near-infra-red wavelengths than senescent vegetation (Butterfield and Malmström, 2009). Vegetation indices employ a range of band combinations to represent or measure existing biomass (Morais et al., 2021).

The combination of machine learning (ML) and cloud computing, exemplified by the Google Earth Engine (GEE), facilitates the examination of extensive datasets with high temporal resolution and spatial density, derived from satellite platforms and from UAV-derived raster products (e.g. orthomosaics and vegetation indices) (M et al., 2023; Ouchra and Belangour, 2024). This approach minimises the impact of

noise and overfitting and is more effective than traditional computer systems and classical analysis methods. Furthermore, it allows for the exploration of a wide range of parameter values, guided by training data, to identify solutions that optimise specific performance metrics (La Salandra et al., 2024). This free platform is capable of storing and analysing high-resolution images as raster data (Javidan et al., 2024). Furthermore, it permits the execution of numerous machine learning algorithms to predict biomass models utilising spectral bands and vegetation indices across multiple processors in a simultaneous manner, thereby reducing processing time and resource usage while delivering accurate results (Bazzo et al., 2023).

In this context, the objective of the research was to predict biomass through the application of digital mapping using Google Earth Engine (GEE), UAV imagery and machine learning (ML) on six farms with predominantly ryegrass (*Lolium multiflorum* Lam)—clover (*Trifolium repens*) grassland located between 2600 and 2800 masl. This approach employed high-resolution reflectance imagery derived from a multispectral camera (Red Edge MX model; MicaSense Inc.) mounted on a UAV, and the data were compared to in situ measurements of wet and dry biomass.

## 2. Materials and methods

### 2.1. Study area

The Cajamarca Valley is situated at an elevation of 2,718 metres above sea level, with average temperatures ranging from 5 to 20 degrees Celsius. The valley is characterised by deep to very deep soils and gentle slopes of 0-8 percent, which combine to create an environment conducive to intensive agriculture and livestock farming. Furthermore, the region displays significant development potential. The annual rainfall is divided into two distinct periods: a high rainfall period from January to March and a low rainfall period from May to October. The precipitation ranges from 530 mm to 849 mm per year (SENAMHI, 2024; Vallejos-Cacho et al., 2024). The research was conducted throughout the Cajamarca Valley in the northern region of Peru, with a particular focus on six grazing paddocks (Figure 1). These paddocks consisted of pastures with a rye grass-clover (*Lolium multiflorum Lam., Fl. Franç. (Lamarck) 3: 621 (1779) - Trifolium repens L., Sp. Pl. 2: 767 (1753).*) association.

### 2.2. Ethics and permissions statement

The investigation complied with all RPAS flight permits issued to the pilot, bearing Accreditation N°02287 from the Directorate General of Civil Aeronautics, for the operation of the imaging equipment. Since the areas under study were cultivated lands, no specific authorisations from the competent environmental authority were required. Consequently, the work proceeded with the sole authorisation from the Ministry of Transport and Communications.

### 2.3. Methodological framework

Figure 2 illustrates the methodological framework used in this study, which is further elaborated in the following three subsections:

### 2.4. Field sampling

In total, 75 biomass samples were collected in six grazing paddocks of the Cajamarca Valley during the second week of November 2021 using a $0.5 \times 0.5$ quadrat. The sample plots were georeferenced using a D-RTK 2-DJI GNSS. This approach offers straightforward implementation, enabling the uncomplicated estimation of quality measures and their precision. This results in relatively precise estimates, eliminating the need for assumptions about the standard error of the estimated quality measures.

During the second week of November 2021, we collected 75 biomass samples from six grazing paddocks in the Cajamarca Valley. Each sample was taken from a $0.5 \times 0.5$-metre square area. The exact location of

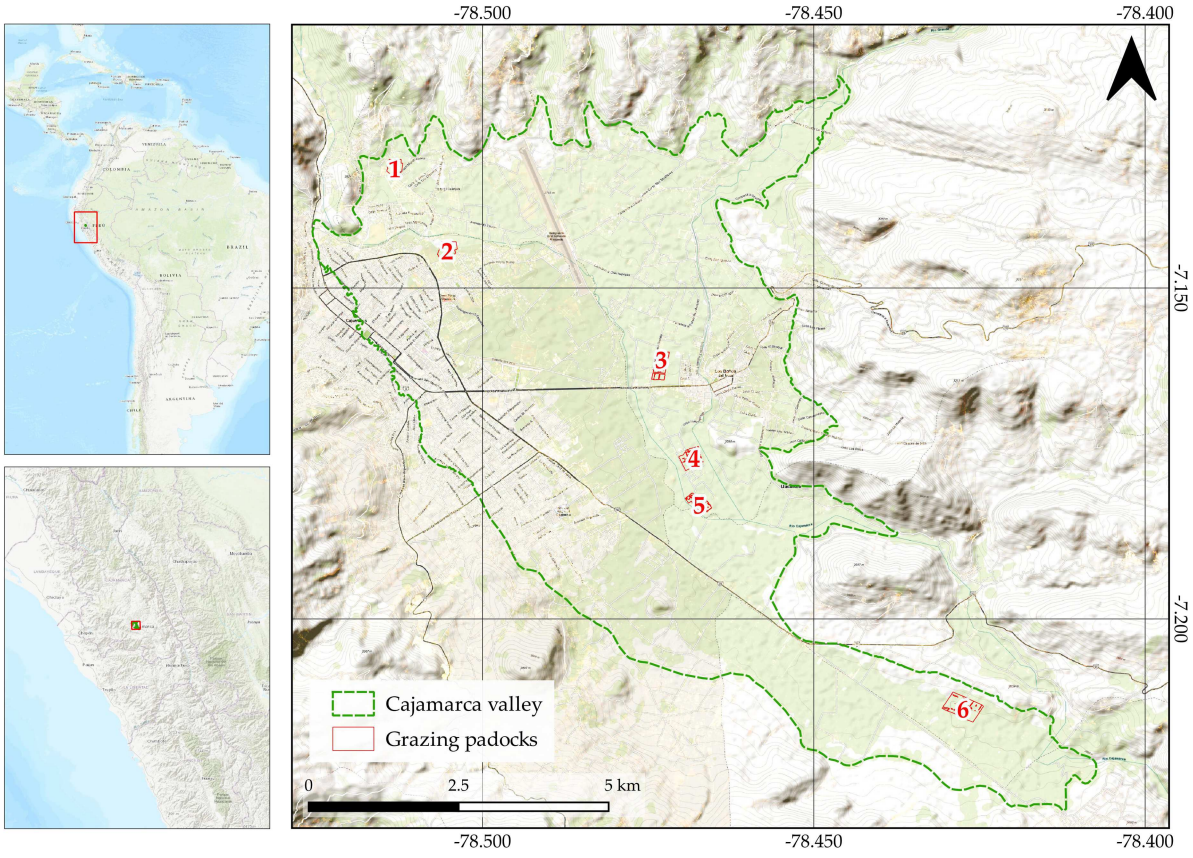



**Figure 1.** Location of the study area, Cajamarca valley (Peru). Prepared using Google Earth.

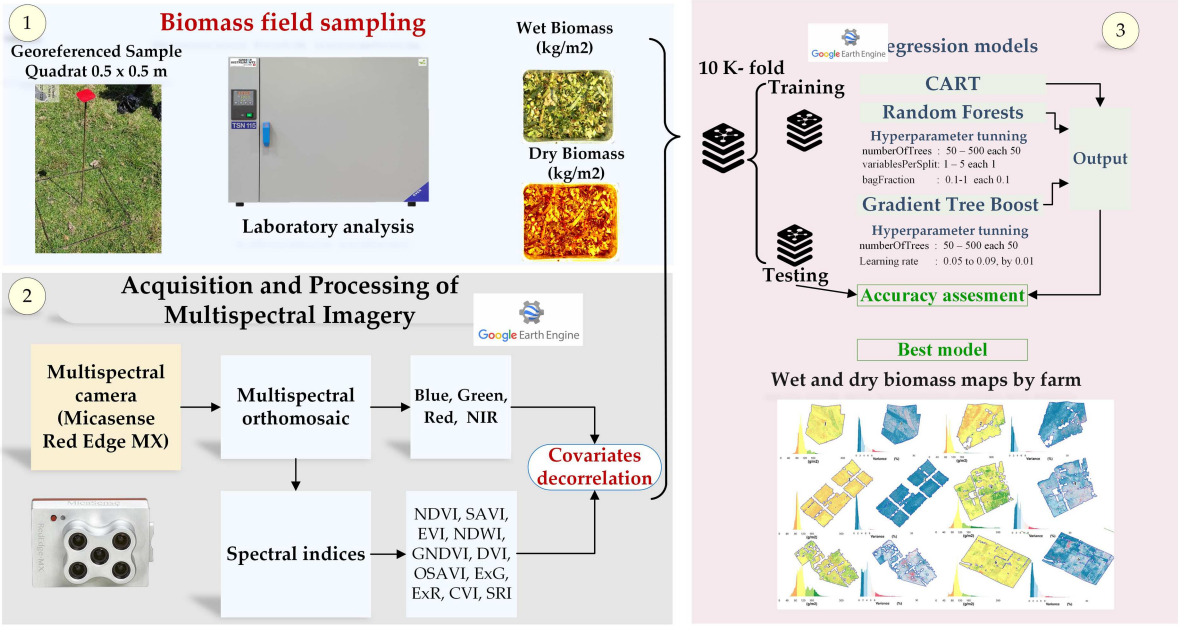



**Figure 2.** Representation of the methodological framework used in this study.

each sample was recorded using a D-RTK 2-DJI GNSS. This method is simple to use and provides accurate estimates of biomass quality without requiring complex calculations or assumptions.

#### 2.4.1. Acquisition and processing of multispectral imagery

High-resolution (3.2 megapixels) multispectral imagery was acquired using a MicaSense RedEdge MX camera mounted on a DJI Matrice 600 drone (Figure 3). The camera captured data in five spectral bands: blue (475 ± 2 nm), green (560 ± 20 nm), red (668 ± 10 nm), near-infra-red (NIR) (840 ± 40 nm), and red-edge (RE) (717 ± 10 nm). Flight planning and camera settings were optimised for data collection.

The aerial survey was carried out around midday, flying at an altitude of 120 m. Images were taken every two seconds, resulting in a consistent 75% frontal overlap and 75% lateral overlap between consecutive photographs. Before and after each flight, images of a calibrated reflectance panel were acquired, and the multispectral camera's irradiance sensor was used during flights to correct for changes in ambient light conditions. These calibration steps allowed the generation of surface reflectance products. All images were saved as high-quality 16-bit TIFF files. The Pix4D Pro Mapper (Prilly, Switzerland) software was used to process the images and generate a precise 3D model, which served to ensure accurate georeferencing. Vegetation indices were then calculated from the orthomosaics for biomass estimation. The minimal discrepancy (less than 0.5%) between the original camera settings and those refined during processing indicated that the initial camera information was accurate enough for producing a reliable orthomosaic.

Within each paddock, four ground control points (GCPs) were collected using a D-RTK 2-DJI GNSS, which provides horizontal and vertical accuracies of 1 cm + 1 ppm (RMS) and 2 cm + 1 ppm (RMS), respectively. These GCPs were integrated to enhance the topographic accuracy of the point cloud and the reflectance bands of the orthomosaic. The final orthomosaic achieved a root mean square error (RMSE) of 5.12 cm, lower than the ground sampling distance (15.42 cm). This ensured reliable georeferencing for vegetation index and biomass analyses and allowed accurate spatial correspondence with the clipped biomass quadrats.

#### 2.4.2. Model development and statistical analysis

**2.4.2.1. Variable extraction.** Thirteen commonly used spectral indices (Table 1) were employed to develop spatial biomass models, as they capture vegetation, soil, and water characteristics. For each 0.5 × 0.5 m quadrat, spectral indices were extracted as mean values from the corresponding orthomosaic area, ensuring a one-to-one correspondence between field biomass measurements and predictor variables.

**2.4.2.2. Covariates selection.** To streamline our analysis and improve computational efficiency, we conducted a de-correlation analysis on the 18 spectral covariates selected as potential predictors. We

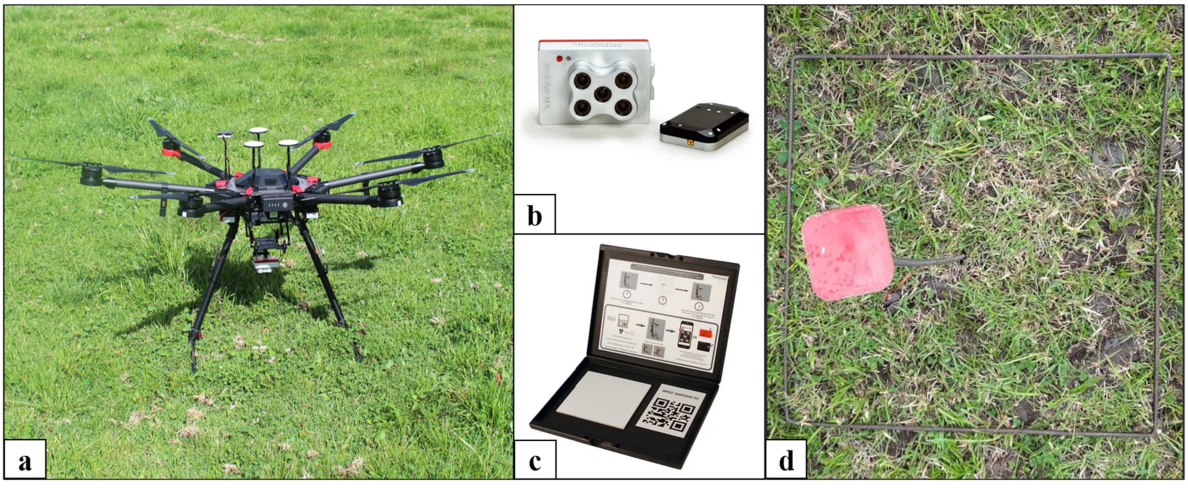



**Figure 3.** (a) Matrice 600 UAV integrated with the Micasense Rededge MX sensor serving as the imaging platform used in this study, (b) Micasense Rededge MX, (c) Reflectance calibration panel, and (d) Georeferenced quadrat plot.

**Table 1.** Spectral indices computed from the Micasense Rededge MX.

| Bands | Wavelength (nm) |
|---|---|
| Normalised Difference Vegetation Index (NDVI) (Rouse and Haas, 1974) | $\frac{(NIR-RED)}{(NIR+RED)}$ |
| Normalised Difference Water Index (NDWI) (McFeeters, 1996) | $NDWI = \frac{(GREEN-NIR)}{(GREEN+NIR)}$ |
| Soil Adjusted Vegetation Index (SAVI) (Qi et al., 1994) | $L = 0.6$ $\left(\frac{(NIR\ -\ RED)}{(NIR\ +\ RED+1)}\right)(1\ +\ L)$ |
| Green Normalised Difference Vegetation Index (GNDVI) (Gitelson et al., 1996) | $\frac{(NIR\ -\ GREEN)}{(NIR\ +\ GREEN)}$ |
| Difference Vegetation Index (DVI) (Richardson and Everitt, 1992) | $(NIR\ -\ RED)$ |
| Optimised Soil Adjusted Vegetation Index (OSAVI) (Rondeaux et al., 1996) | $(1\ +\ 0.16)\left(\frac{(NIR\ -\ RED)}{(NIR\ +\ RED+0.16)}\right)$ |
| Excess Green index (ExG) (Woebbecke and Meyer, 1995) | $2 \times GREEN - RED - BLUE$ |
| Excess Red index (ExR) (Meyer et al., 1999) | $2 \times RED - GREEN$ |
| ExG–ExR (Meyer and Neto, 2008) | $ExG - ExR$ |
| Normalised Difference Index (NDI) (Bannari et al., 1995) | $\frac{(GREEN\ -\ RE)}{(GREEN+\ RE)}$ |
| Red-edge Normalised Difference Vegetation Index (NDRE) (Gitelson and Merzlyak, 1994) | $\frac{(NIR\ -\ RED)}{(NIR\ +\ RED)}$ |
| Chlorophyll vegetation index (CVI) (Vincini et al., 2008) | $NIR \times (\frac{RED}{GREEN^2})$ |
| Simple Ratio Red/Blue Iron Oxide (SRI) (Hewson and Cudahy, 2001) | $RED/BLUE$ |

Micasense RedEdge MX multispectral central wavelengths: B, G, R, RE and NIR: 474, 560, 668, 717 and 840 nm.

retained only those covariates with a pairwise correlation coefficient below 0.85. In cases where multiple covariates were highly correlated, we prioritised the alphabetically first one. This process resulted in a reduced set of five key predictors for our modelling phase.

***2.4.2.3. Spatial analysis.*** **Machine Learning Models.** Three machine learning regression techniques were employed to model fresh and dry biomass: Classification and Regression Trees (CART), Random Forest (RF), and Gradient Boosting (XGBoost).

CART is a non-parametric algorithm that constructs decision tree models by recursively partitioning the data into smaller subsets based on predictor values, allowing it to capture nonlinear relationships and reduce the influence of noise (Jain et al., 2020). However, single CART models can be unstable and sensitive to small data perturbations.

To overcome these limitations, ensemble approaches such as RF and XGBoost use multiple trees to improve predictive performance and stability (Gholizadeh et al., 2018). RF constructs numerous decision trees, each trained on a random subset of data and predictor variables, and produces regression predictions by averaging across trees, thereby reducing variance and enhancing generalisation (Luo et al., 2016). RF is robust to different data distributions, capable of modelling both discrete and continuous variables, and has shown excellent ability to uncover complex nonlinear patterns with strong generalisation performance (Phinzi et al., 2021).

XGBoost, a leading gradient boosting method, trains decision trees sequentially, with each iteration correcting errors from the previous trees. This iterative boosting process effectively manages complex datasets, reduces overfitting, and achieves high predictive accuracy in both classification and regression tasks; Wang and Zhang, 2021).

**Leave-One-Farm-Out Cross-Validation.** To ensure robust and unbiased model evaluation, we implemented a Leave-One-Group-Out Cross-Validation Cross-Validation (LOGO-CV) approach to address spatial autocorrelation within farms and provide realistic estimates of model generalisation to new locations (Meyer and Pebesma, 2021; Ploton et al., 2020; Roberts et al., 2017). The validation consisted of six iterations, each using five farms for training and hyperparameter tuning while holding out one farm for independent validation (Pohjankukka et al., 2017). This spatial validation strategy ensures performance metrics reflect the model's ability to predict biomass at unseen locations, avoiding overly optimistic estimates from random cross-validation in spatially structured data (Schratz et al., 2019; Wadoux and Heuvelink, 2021).

**Hyperparameter Optimisation.** Optimal hyperparameters were determined through grid search on training data before evaluation on each held-out farm (Bergstra and Bengio, 2012; Probst et al., 2019). For XGBoost, we varied the number of trees (NT: 50–500 in steps of 50) and sampling rate (SR: 0.05–0.09 in

steps of 0.01), while holding shrinkage (0.005), max_nodes (unrestricted), loss function (Least Absolute Deviation), and random seed (0) constant (Chen and Guestrin 2016), resulting in 50 combinations. For Random Forest, we varied NT (50–500 in steps of 50), variables per split (VPS: 1–5), and bag fraction (BF: 0.1–1.0 in steps of 0.1) (Breiman, 2001; Probst and Boulesteix, 2018), generating 500 combinations per fold.

**Model Evaluation.** Model performance was assessed using coefficient of determination ($R^2$), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE) (Chai and Draxler, 2014; Hodson, 2022; Legates and McCabe, 1999). Metrics were calculated for each validation fold and averaged to obtain overall performance, with higher $R^2$ and lower RMSE and MAE indicating superior predictive accuracy.

Predictor importance was ranked by calculating normalised percentage contributions to model accuracy, identifying variables most frequently used at decision nodes (Strobl et al., 2007). Spatial uncertainty was assessed by calculating standard deviation across the six validation maps, identifying areas with lower prediction reliability (Hengl and Nussbaum, 2018; Meyer and Pebesma, 2021).

## 3. Results

### 3.1. Descriptive statistics

Figure 4 illustrates the distribution of fresh and dry biomass across farms, with mean values indicated above each boxplot. Fresh biomass ranged from 52.64 to 3,985.40 g/m², while dry biomass varied between 34.74 and 392.58 g/m². In addition to the wide dispersion, the mean values provide a clear indication of central tendencies across farms. Fresh biomass generally exhibited higher mean values and greater variability, particularly in farms 1 and 4, where extreme values and broad interquartile ranges were observed. By contrast, dry biomass showed lower means and narrower interquartile ranges, reflecting a more clustered distribution. These results emphasise both the heterogeneity of biomass production among farms and the stronger variability of fresh compared to dry biomass.

### 3.2. Correlation analysis between predictors and biomass

Figure 5 shows the Pearson correlation coefficients (r) between fresh and dry biomass and the covariables (spectral bands and vegetation indices) derived from Micasense RedEdge-MX imagery, visualised using the corrplot package in R (Wei and Simko, 2017).

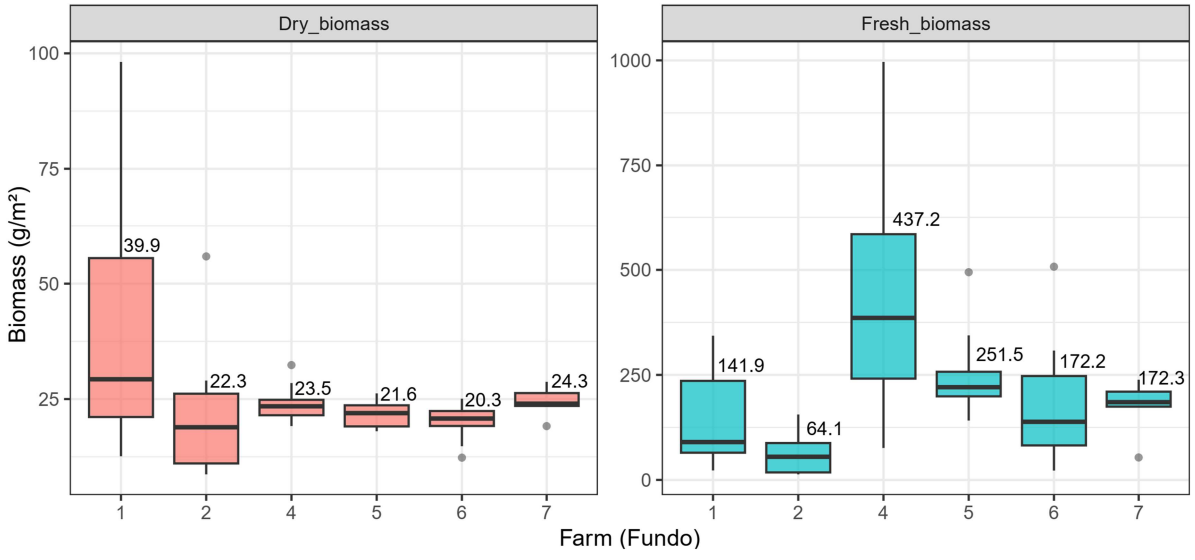



**Figure 4.** Boxplots of fresh and dry biomass (g/m²) across farms (Fundo). Mean values are indicated above each boxplot.

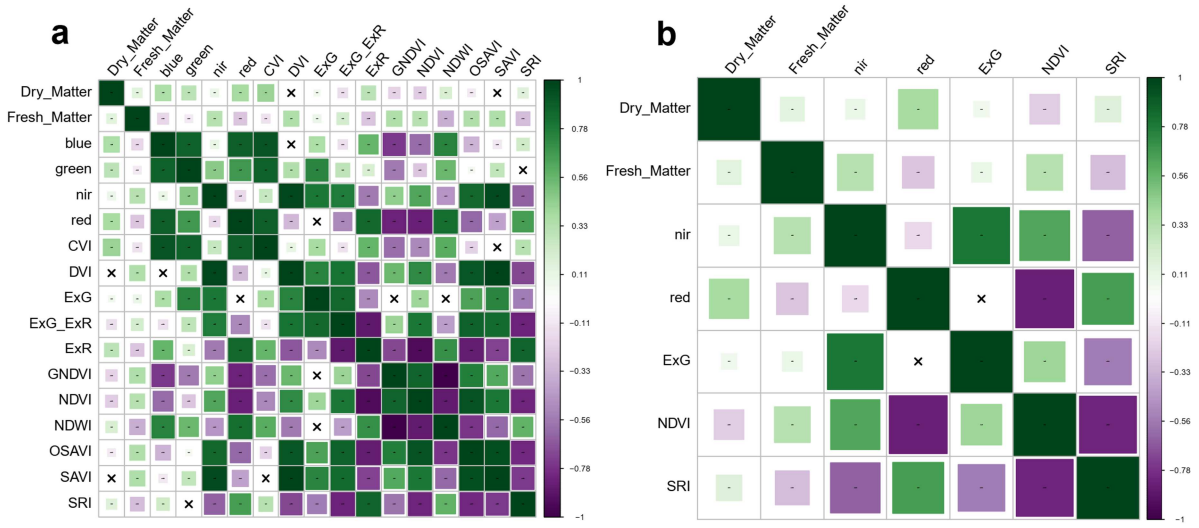



**Figure 5.** Pearson correlation coefficients (r) between measured fresh and dry biomass and spectral predictors (bands and vegetation indices) derived from Micasense RedEdge-MX imagery. Coefficients are significant at the 5% probability level; cells marked with 'X' indicate non-significant correlations.

**Table 2.** Optimal hyperparameters and predictive performance of CART, random forest (RF), and XGBoost models for dry biomass and fresh biomass estimation.

| Variable | Algorithm | Optimal | $R^2$ | RMSE | MAE | RPD |
|---|---|---|---|---|---|---|
| **Dry biomass** | CART | cp = 0.001; maxdepth = 15; minsplit = 10 | 0.578 | 0.364 | 0.180 | 1.519 |
| | RF | ntrees = 500; mtry = 3; bag = 0.9 | 0.747 | 0.279 | 0.130 | 1.982 |
| | XGBoost | nrounds = 600; eta = 0.05 | 0.708 | 0.299 | 0.159 | 1.848 |
| **Fresh biomass** | CART | cp = 0.001; maxdepth = 15; minsplit = 20 | 0.521 | 5.366 | 3.337 | 1.437 |
| | RF | ntrees = 500; mtry = 4; bag = 0.9 | 0.671 | 4.415 | 2.521 | 1.750 |
| | XGBoost | nrounds = 600; eta = 0.10 | 0.616 | 4.768 | 3.034 | 1.6177 |

Panel (a) presents the full correlation matrix including all variables, while panel (b) displays the reduced set of predictors selected through a decorrelation procedure. Because fresh and dry biomass share the same set of spectral predictors, a combined correlation matrix provides a clearer overview of their associations with the covariates. The analysis revealed distinct patterns: dry biomass was positively associated with the red spectral band and the Structural Ratio Index (SRI) (r = 0.16–0.39), whereas fresh biomass was more strongly correlated with the near-infra-red (NIR) band, NDVI, and ExG index (r = 0.10–0.33).

### 3.3. Analysis of modelling outcomes

Table 2 presents the performance of the regression models with the highest predictive accuracy, evaluated using the validation dataset. For fresh biomass estimation, the Random Forest model configured with 500 trees, four variables per split, and a bag fraction of 0.9 yielded the best performance ($R^2 = 0.671$, RMSE = 4.415, MAE = 2.521, RPD = 1.750). For dry biomass, the Random Forest model with 500 trees, three variables per split, and a bag fraction of 0.9 achieved the highest accuracy ($R^2 = 0.747$, RMSE = 0.279, MAE = 0.130, RPD = 1.982). In comparison, the XGBoost model demonstrated moderate predictive capacity for both fresh biomass ($R^2 = 0.616$, RMSE = 4.768) and dry biomass ($R^2 = 0.708$, RMSE = 0.299), while the CART model exhibited the lowest performance across the assessed variables ($R^2 = 0.521$ and 0.578 for fresh and dry biomass, respectively). These results confirm that Random Forest consistently outperformed other algorithms when evaluated on independent validation data, with dry biomass estimation showing superior predictive performance compared to fresh biomass.

### 3.4. Prediction results and relative importance of the predictors

In light of the preceding outcomes, the most efficacious regression models were selected and maps illustrating the spatial quantitative distribution of fresh and dry biomass were created (see Figures 6–8).

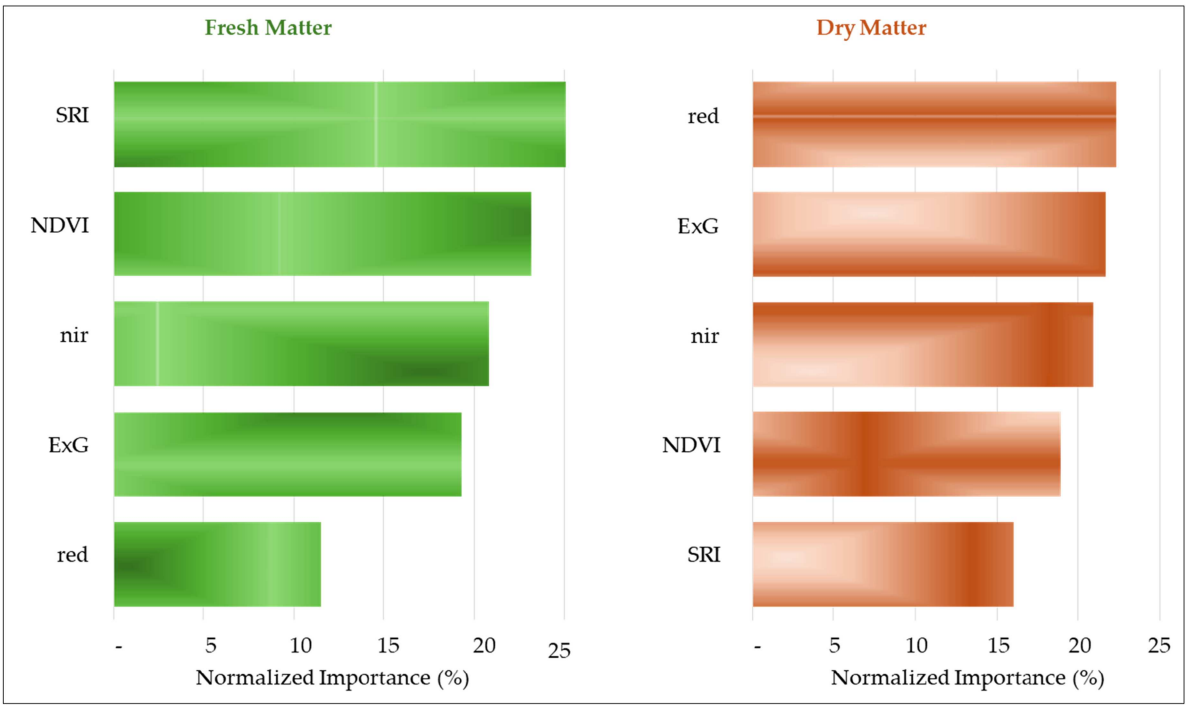



**Figure 6.** Normalised importance for both dry and fresh matter.

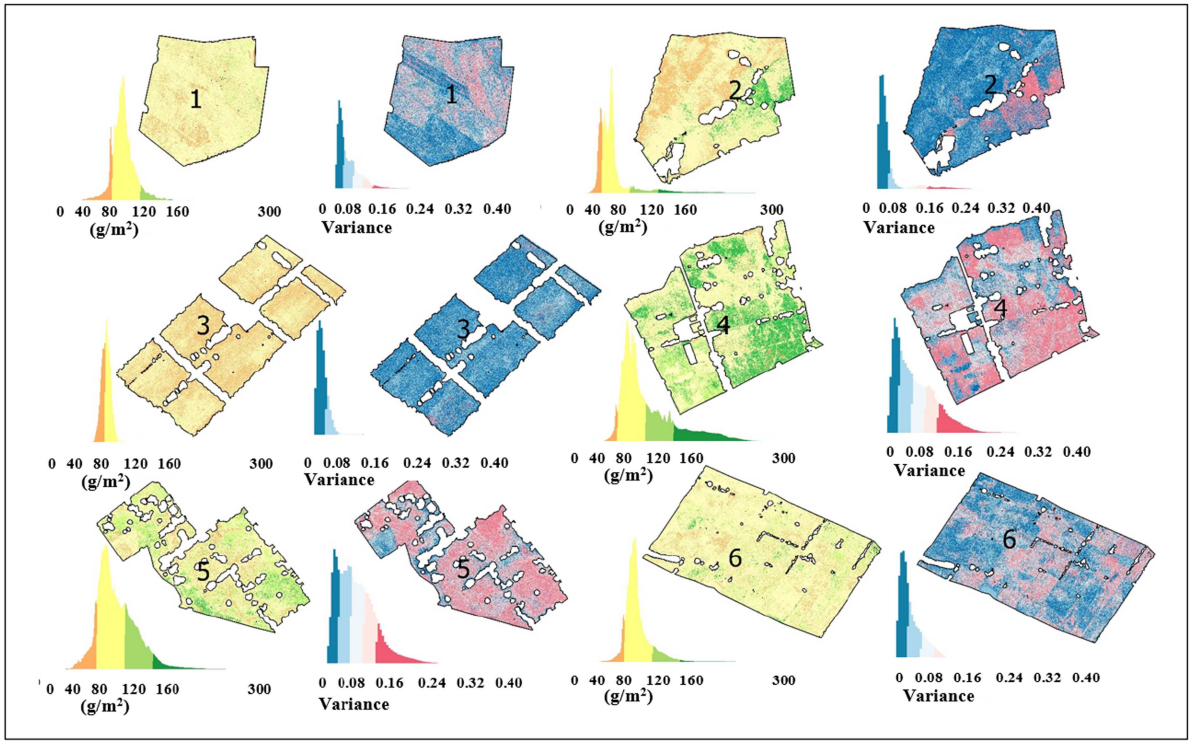



**Figure 7.** Prediction maps for dry biomass with uncertainty maps.

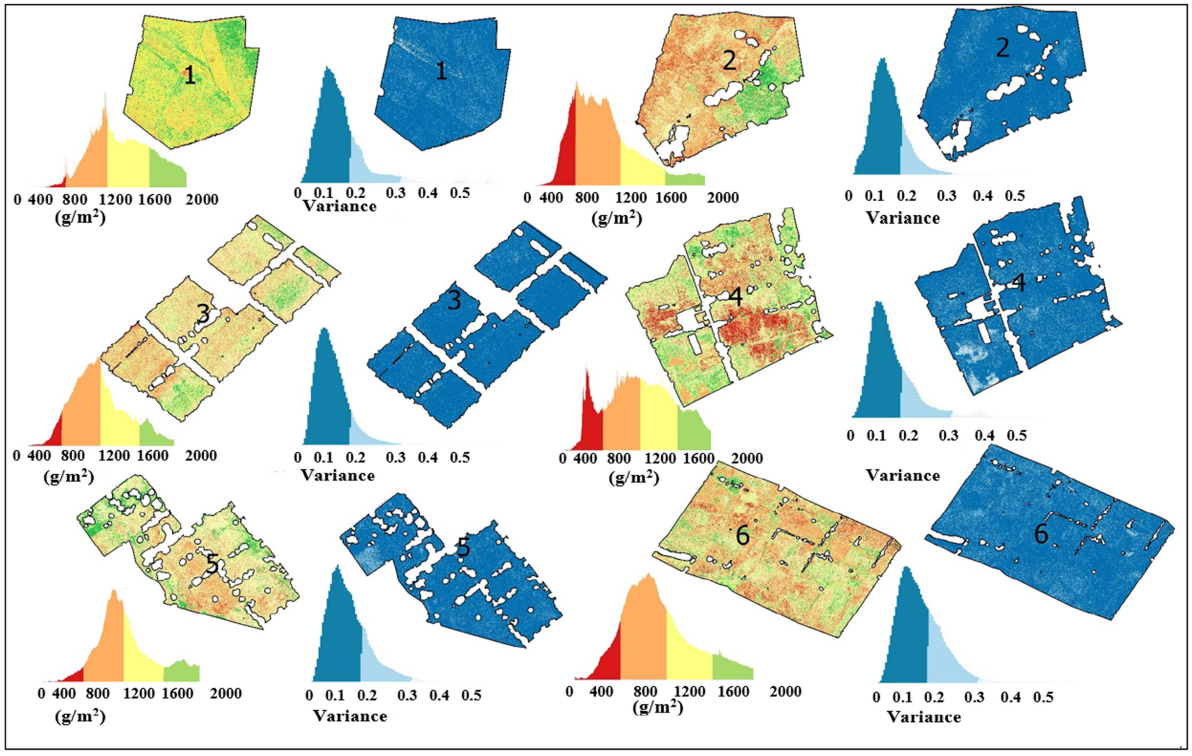



**Figure 8.** Prediction maps for fresh biomass with uncertainty maps.

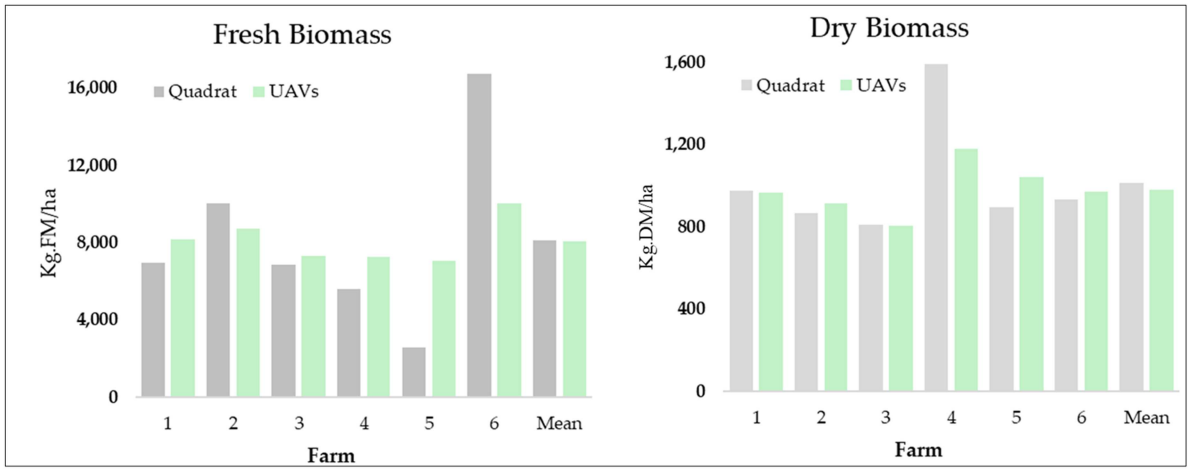



**Figure 9.** Comparison of biomass estimation by quadrats and UAV.

The top-performing models, characterised by high accuracy and low error rates, rely heavily on spectral indices and bands. These models reveal distinct sets of predictors for accurately estimating dry and fresh biomass.

Figure 9 presents a comparison of biomass estimation by quadrats and UAV, derived from the optimal model utilising UAV imagery. The results indicate significant variation in biomass estimations for fresh matter between different farms, while the differences are less pronounced for dry matter estimates. Additionally, the mean biomass values are similar between the two methods for both fresh and dry matter. A key advantage of using UAV estimations is the ability to visualise the spatial variation of biomass within each paddock and among different paddocks, which is more pronounced when estimating fresh matter.

## 4. Discussion

The generation of biomass models using machine learning (ML) represents a viable approach for the prediction of biomass. In this instance, the $R^2$ value for fresh biomass was 0.671, while the $R^2$ value for dry biomass was 0.747, as detailed in Table 2. These results were obtained using random forest (RF). These values are noteworthy in comparison to previous research (Insua et al., 2019; Monteith, 1972). The RF method exhibited superior performance compared to the PLSR ($R^2 = 0.899$) and SVM ($R^2 = 0.837$) methods obtained by Amarsaikhan and Erdenebaatar, 2023) (Amarsaikhan and Erdenebaatar, 2023), potentially due to the height and speed of flight performed. The results obtained are comparable when multispectral images are combined with spectral indices as predictors, with an enhanced correlation with biomass. This improvement can be attributed primarily to the typical absorption characteristics of vegetation in the visible and near-infra-red spectral regions. In contrast, Théau et al. (2021) employed a regression-based biomass prediction model for a grass-based system utilising the green normalised difference vegetation index (GNDVI), attaining an $R^2$ of 0.80 for fresh matter and 0.66 for dry matter (Théau et al., 2021). Such precise estimations are only possible when fresh matter levels are below 3 kg/m².

The productivity of vegetation is strongly associated with the manner in which solar radiation interacts with the plant canopy. This concept was first proposed by Monteith (1972) (Hengl and Hengl and MacMillan, 2019), who suggested that the productivity of stress-free annual crops is linearly related to the photosynthetically active radiation (PAR) absorbed by the vegetation. Consequently, in our study, spectral vegetation indices (VI) such as the normalised difference vegetation index (NDVI) were developed for application to cropland and grassland.

Given that machine learning models are not constrained by the assumption of normality (Näsi et al., 2018), no data transformations were applied to the model used in the present study. Random forest (RF) is regarded as a technique that mitigates the uncertainty associated with classification and regression tree (CART) because RF takes the mean of a set of fully developed trees. The construction of a substantial number of uncorrelated, noisy and approximately unbiased trees serves to reduce the variance and uncertainty of the model, as evidenced by Bazzo et al. (Bazzo et al., 2023). Additionally, RF is capable of running efficiently on large datasets, is relatively robust to outliers, and is an effective method for processing complex training data (Rodriguez-Galiano et al., 2012).

Random Forest (RF) has proven to be highly accurate in estimating biomass in agriculture, comparable to other methods. This aligns with the findings of several studies. Morais et al. (2021) highlight RF as the most common machine learning approach for biomass estimation in grasslands (Zeng et al., 2021). successfully applied RF to estimate Aboveground Biomass (AGB) in Chinese grasslands using Remote Sensing (RS) data and other relevant information. Furthermore Edirisinghe et al. (2011) demonstrated the effectiveness of a similar approach in estimating biomass in Australian grazed grasslands using Normalised Difference Vegetation Index (NDVI) data from RS and a regression model. These successful applications by other researchers further validate the reliability of our methodology.

The overall prevalence of spectral indices such as NDVI, ExG and SRI in fresh and dry biomass is related to the abundance and productivity of green vegetation (Ardiansyah et al., 2018; Bazzo et al., 2023). These inputs are similar to those obtained for biomass prediction based on multiple spectral indices rather than a single one (Li et al., 2021; Théau et al., 2021). Furthermore, it should be noted that mapping the spatial structure of vegetation is essential to improve biomass predictions, which cannot be achieved solely with 3D models based on photogrammetric restitution (Pranga et al., 2021). A more suitable alternative is the use of dense point clouds obtained by LIDAR scanning (Schulze-Brüninghoff et al., 2019). In fact, a combination of both products is required.

The results may be constrained by the routine rotational grazing in each paddock, which precludes the possibility of making predictions regarding stocking rates over an extended period. Other limitations may be attributed to the intrinsic complexity of manuring and fertilisation practices. Nevertheless, this limitation could be offset in future studies by integrating continuous monitoring with satellite data, thereby enhancing the temporal resolution of observations. Conversely, Random Forest (RF) regression is regarded as an appropriate approach for small sample sizes (Luan et al., 2020), and cross-validation is useful for assessing model over-fitting. Increasing the number of samples would be beneficial for future research, despite the additional burden of sampling and in situ measurements. A larger data set would enhance the robustness of the model (Zhao and Wang, 2022).

The findings of our study and methodology reinforce the efficacy of certain management strategies that could be deployed in the northern highlands of Peru. These include determining the optimal stocking rate, estimating milk yield and potentially conserving surplus forage during the rainy season or acquiring forage supplements for the dry season. This approach may support sustainable pasture management strategies.

## 5. Conclusions

This research advances precision agriculture by demonstrating the value of high-resolution multispectral imagery captured by microsensors mounted on unmanned aerial vehicles (UAVs) for creating highly accurate spatial maps of forage biomass in the short time. Our findings show that analysing high resolution imagery, trough machine learning techniques on cloud platforms like Google Earth Engine provides a cost-effective solution with no need of use physical servers. This approach allows for accurate and detailed predictions of forage biomass with satisfactory results. The use of derived multispectral indices, specifically SRI, NDVI, red and NIR bands, significantly improved the accuracy of biomass predictions. Among three machine learning models evaluated (CART, RF, and XGBoost), Random Forest (RF) models demonstrated the highest predictive power and provided more reliable biomass estimates across different geographical areas. Finally, our results indicate that combining machine learning models with augmented real-world data collected by UAVs and multispectral micro sensors is a decision-support potential for estimating aerial biomass. While this study does not directly measure sustainability outcomes, it provides spatial information that may support future sustainable grazing decisions.

## Public interest statement

Traditional pasture measurement in the northern highlands of Peru is time-consuming and labor-intensive. Our research overcomes this barrier by adapting the use of multispectral drones and cloud processing (Google Earth Engine) to the conditions of the Andes. We identified the Random Forest machine learning model as the most effective for this region, enabling the creation of fast and accurate biomass maps. This democratizes access to critical data, enabling livestock farmers in the region to make informed decisions about grazing and improve their productivity.

## Acknowledgements

We would like to express our gratitude to Luis Vilela, Jorge Piedra, Raul Caceres, Fidel Chunque, Luis Rojas Quispe, Roy Florián, Julio Cárdenas, Ezequiel García and Sylvia Saldanha for their invaluable assistance in the conceptualisation and technical-administrative support throughout the research process.

## Author contributions

CRediT: **Luis Vallejos-Fernández:** Conceptualization, Funding acquisition, Methodology, Software, Supervision, Writing – original draft, Writing – review & editing; **Wuesley Alvarez-García:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing; **Maycol Abanto-Urbina:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing; **Felipe Gutiérrez-Arce:** Formal analysis, Investigation, Methodology, Writing – original draft; **Eduardo Tapia-Acosta:** Formal analysis, Funding acquisition, Investigation, Resources, Validation; **Samuel Pizarro:** Data curation, Software, Validation, Writing – original draft; **Cesar Ciprian:** Data curation, Software, Validation, Writing – original draft; **Javier Naupari:** Software, Supervision, Validation, Visualization, Writing – original draft.

## Disclosure statement

The authors declare no conflict of interest.

## Funding

This research was funded by: 'Application of multispectral images in the productive characterisation of ryegrass-clover pastures in Cajamarca.' Project of the Universidad Nacional de Cajamarca - Code PCM-ZOT-007-18.

## ORCID

Luis Vallejos-Fernández 0000-0002-9858-642X
Wuesley Alvarez-García 0000-0002-9655-3149
Maycol Abanto-Urbina 0000-0003-0282-417X

## Data availability statement

The data presented in this study are available on request from the corresponding author.

## References

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. https://doi.org/10.1080/10106049.2023.2195824

Ardiansyah, A., Chusnul, W., Krissandi, M., & Asna (2018). Biomass development in SRI field under unmaintained alternate wetting-drying irrigation. *IOP Conference Series: Earth and Environmental Science*, *147*(May), 012041. https://doi.org/10.1088/1755-1315/147/1/012041

Bannari, A., Morin, D., Bonn, F., & Huete, A. R. (1995). A review of vegetation indices. *Remote Sensing Reviews*, *13*(1–2), 95–120. https://doi.org/10.1080/02757259509532298

Batistoti, J., Marcato Junior, J., Ítavo, L., Matsubara, E., Gomes, E., Oliveira, B., Souza, M., Siqueira, H., Salgado Filho, G., Akiyama, T., Gonçalves, W., Liesenberg, V., Li, J., & Dias, A. (2019). Estimating pasture biomass and canopy height in Brazilian savanna using UAV photogrammetry. *Remote Sensing*, *11*(20), 2447. https://doi.org/10.3390/rs11202447

Bazzo, C. O. G., Kamali, B., Hütt, C., Bareth, G., & Gaiser, T. (2023). A review of estimation methods for aboveground biomass in grasslands using UAV. *Remote Sensing*, *15*(3), 639. https://doi.org/10.3390/rs15030639

Bergstra, J., & Bengio, Y. (2012). 'Random search for hyper-parameter optimization'. Journal of machine learning research: JMLR, *13*(null), 281–305.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Butterfield, H. S., & Malmström, C. M. (2009). The effects of phenology on indirect measures of aboveground biomass in annual grasses. *International Journal of Remote Sensing*, *30*(12), 3133–3146. https://doi.org/10.1080/01431160802558774

Cenagro, T. (2012). 'Censo Nacional Agropecuario'. *Boletín02, MINAGRI*.

Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? – arguments against avoiding RMSE in the literature. *Geoscientific Model Development*, *7*(3), 1247–1250. https://doi.org/10.5194/gmd-7-1247-2014

De Rosa, D., Basso, B., Fasiolo, M., Friedl, J., Fulkerson, B., Grace, P. R., & Rowlings, D. W. (2021). Predicting pasture biomass using a statistical model and machine learning algorithm implemented with remotely sensed imagery. *Computers and Electronics in Agriculture*, *180*(January). 105880. https://doi.org/10.1016/j.compag.2020.105880

Edirisinghe, A., Hill, M. J., Donald, G. E., & Hyder, M. (2011). Quantitative mapping of pasture biomass using satellite imagery. *International Journal of Remote Sensing*, *32*(10), 2699–2724. https://doi.org/10.1080/01431161003743181

Gholizadeh, A., Žižala, D., Saberioon, M., & Borůvka, L. (2018). Soil organic carbon and texture retrieving and mapping using proximal, airborne and Sentinel-2 spectral imaging. *Remote Sensing of Environment*, *218*(December), 89–103. https://doi.org/10.1016/j.rse.2018.09.015

Gitelson, A., & Merzlyak, M. N. (1994). Spectral reflectance changes associated with autumn senescence of Aesculus Hippocastanum L. and Acer Platanoides L. Leaves. Spectral features and relation to chlorophyll estimation. *Journal of Plant Physiology*, *143*(3), 286–292. https://doi.org/10.1016/S0176-1617(11)81633-0

Gitelson, A. A., Kaufman, Y. J., & Merzlyak, M. N. (1996). Use of a green channel in remote sensing of global vegetation from EOS- MODIS'. *Remote Sensing of Environment*, *58*(3), 289–298. https://doi.org/10.1016/S0034-4257(96)00072-7

Goebel, M., & Iwaszczuk, D. (2023). Spectral analysis of images of plants under stress using a close-range camera. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XLVIII-1/W3-2023*(October), 63–69. https://doi.org/10.5194/isprs-archives-XLVIII-1-W3-2023-63-2023

Hengl, T., & MacMillan, R. A. (2019). *Predictive Soil Mapping with R* (p. 370). Wageningen, The Netherlands: OpenGeoHub Foundation.

Hengl, T., Nussbaum, M., Wright, M. N., Heuvelink, G. B., & Gräler, B. (2018). Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. *PeerJ*, *6*(August). e5518. https://doi.org/10.7717/peerj.5518

Hewson, R. D., Cudahy, T. J., & Huntington, J. F. (2001). Geologic and alteration mapping at Mt Fitton, South Australia, using ASTER satellite-borne data. *International Geoscience and Remote Sensing Symposium (IGARSS)*, *2*(C), 724–726. https://doi.org/10.1109/igarss.2001.976615

Hodson, T. O. (2022). Root-mean-square error (RMSE) or mean absolute error (MAE): When to use them or not. *Geoscientific Model Development*, *15*(14), 5481–5487. https://doi.org/10.5194/gmd-15-5481-2022

Insua, J. R., Utsumi, S. A., & Basso, B. (2019). Estimation of spatial and temporal variability of pasture growth and digestibility in grazing rotations coupling unmanned aerial vehicle (UAV) with crop simulation models. *PLoS One*, *14*(3), e0212773. https://doi.org/10.1371/journal.pone.0212773

Jain, P., Coogan, S. C., Subramanian, S. G., Crowley, M., Taylor, S., & Flannigan, M. D. (2020). A review of machine learning applications in wildfire science and management. *Environmental Reviews*, *28*(4), 478–505. https://doi.org/10.1139/er-2020-0019

Javidan, R., Rahmati, O., Soleimanpour, S. M., & Mohammadi, F. (2024). Soil properties mapping using the google earth engine platform, *Remote Sensing of Soil and Land Surface Processes*. 385–398. Elsevier. https://doi.org/10.1016/B978-0-443-15341-9.00022-8

Kume, A. (2017). Importance of the green color, absorption gradient, and spectral absorption of chloroplasts for the radiative energy balance of leaves. *Journal of Plant Research*, *130*(3), 501–514. https://doi.org/10.1007/s10265-017-0910-z

La Salandra, M., Nicotri, S., Donvito, G., Italiano, A., Colacicco, R., Miniello, G., Lapietra, I., Roseto, R., Dellino, P., & Capolongo, D. (2024). A paradigm shift in processing large UAV image datasets for emergency management of natural hazards. *International Journal of Applied Earth Observation and Geoinformation*, *132*(August). 103996. https://doi.org/10.1016/j.jag.2024.103996

Legates, D. R., & McCabe, G. J. (1999). Evaluating the use of "Goodness-of-fit" measures in hydrologic and hydroclimatic model validation. *Water Resources Research*, *35*(1), 233–241. https://doi.org/10.1029/1998WR900018

Li, K.-Y., Burnside, N. G., Sampaio De Lima, R., Villoslada Peciña, M., Sepp, K., Yang, M., Raet, J., Vain, A., & Selge, A. (2021). The application of an unmanned aerial system and machine learning techniques for red clover-grass mixture yield estimation under variety performance trials. *Remote Sensing*, *13*(10), 1994. https://doi.org/10.3390/rs13101994

Luan, J., Zhang, C., Xu, B., Xue, Y., & Ren, Y. (2020). The predictive performances of random forest models with limited sample size and different species traits. *Fisheries Research*, *227*(July). 105534. https://doi.org/10.1016/j.fishres.2020.105534

Luo, Y.-M., Huang, D.-T., Liu, P.-Z., & Feng, H.-M. (2016). An novel random forests and its application to the classification of mangroves remote sensing image. Multimedia Tools and Applications, 75(16), 9707–9722. https://doi.org/10.1007/s11042-015-2906-9

M, A., Ahmed, S. A., & Harishnaika, N. (2023). Correction to: Land use and land cover classification using machine learning algorithms in google earth engine. *Earth Science Informatics*, *16*(4), 3075–3075. https://doi.org/10.1007/s12145-023-01113-5

McFeeters, S. K. (1996). The use of the normalized difference water index (NDWI) in the Delineation of open water features. *International Journal of Remote Sensing*, *17*(7), 1425–1432. https://doi.org/10.1080/01431169608948714

Meyer, G. E., Hindman, T. W., & Laksmi, K. (1999). Machine vision detection parameters for plant species identification, Precision Agriculture and Biological Quality (Vol. 3543, pp. 327–335). SPIE.

Meyer, G. E., & Neto, J. C. (2008). Verification of color vegetation indices for automated crop imaging applications. *Computers and Electronics in Agriculture*, *63*(2), 282–293. https://doi.org/10.1016/j.compag.2008.03.009

Meyer, H., & Pebesma, E. (2021). Predicting into unknown space? Estimating the area of applicability of spatial prediction models. *Methods in Ecology and Evolution*, *12*(9), 1620–1633. https://doi.org/10.1111/2041-210X.13650

Monteith, J. L. (1972). Solar radiation and productivity in tropical ecosystems. *The Journal of Applied Ecology*, *9*(3), 747. https://doi.org/10.2307/2401901

Morais, T. G., Teixeira, R. F., Figueiredo, M., & Domingos, T. (2021). The use of machine learning methods to estimate aboveground biomass of grasslands: A review. *Ecological Indicators*, *130*(November), 108081. https://doi.org/10.1016/j.ecolind.2021.108081

Näsi, R., Viljanen, N., Kaivosoja, J., Alhonoja, K., Hakala, T., Markelin, L., & Honkavaara, E. (2018). Estimating biomass and nitrogen amount of barley and grass using UAV and aircraft based spectral and photogrammetric 3D features. *Remote Sensing*, *10*(7), 1082. https://doi.org/10.3390/rs10071082

Ouchra, H., Belangour, A., Erraissi, A., & Banane, M. (2024). Assessing machine learning algorithms for land use and land cover classification in Morocco using google earth engine. In Foresti, G. L., Fusiello, A., & Hancock, E. (Eds.), *Image Analysis and Processing - ICIAP 2023 Workshops* (Vol. *14365*). Switzerland: Springer Nature. https://doi.org/10.1007/978-3-031-51023-6_33

Phinzi, K., Abriha, D., & Szabó, S. (2021). Classification efficacy using K-Fold cross-validation and bootstrapping resampling techniques on the example of mapping complex gully systems. *Remote Sensing*, *13*(15), 2980. https://doi.org/10.3390/rs13152980

Ploton, P., Mortier, F., Réjou-Méchain, M., Barbier, N., Picard, N., Rossi, V., Dormann, C., Cornu, G., Viennois, G., Bayol, N., Lyapustin, A., Gourlet-Fleury, S., & Pélissier, R. (2020). Spatial validation reveals poor predictive performance of large-scale ecological mapping models. *Nature Communications*, *11*(1), 4540. https://doi.org/10.1038/s41467-020-18321-y

Pohjankukka, J., Pahikkala, T., Nevalainen, P., & Heikkonen, J. (2017). Estimating the prediction performance of spatial models via spatial k-fold cross validation. *International Journal of Geographical Information Science*, *31*(10), 2001–2019. https://doi.org/10.1080/13658816.2017.1346255

Pranga, J., Borra-Serrano, I., Aper, J., De Swaef, T., Ghesquiere, A., Quataert, P., Roldán-Ruiz, I., Janssens, I. A., Ruysschaert, G., & Lootens, P. (2021). Improving accuracy of herbage yield predictions in perennial ryegrass with

UAV-based structural and spectral data fusion and machine learning. *Remote Sensing*, *13*(17), 3459. https://doi.org/10.3390/rs13173459

Probst, P., & Boulesteix, A.-L. (2018). To tune or not to tune the number of trees in random forest. *Journal of Machine Learning Research*, *18*(181), 1–18.

Probst, P., Wright, M. N., & Boulesteix, A.-Laure (2019). Hyperparameters and tuning strategies for random forest. *WIREs Data Mining and Knowledge Discovery*, *9*(3. e1301. https://doi.org/10.1002/widm.1301

Qamar, F., & Dobler, G. (2023). Atmospheric correction of vegetation reflectance with simulation-trained deep learning for ground-based hyperspectral remote sensing. *Plant Methods*, *19*(1), 74. https://doi.org/10.1186/s13007-023-01046-6

Qi, J., Chehbouni, A., Huete, A. R., Kerr, Y. H., & Sorooshian, S. (1994). A modified soil adjusted vegetation index. *Remote Sensing of Environment*, *48*(2), 119–126. https://doi.org/10.1016/0034-4257(94)90134-1

Richardson, A. J., & Everitt, J. H. (1992). Using spectral vegetation indices to estimate rangeland productivity. *Geocarto International*, *7*(1), 63–69. https://doi.org/10.1080/10106049209354353

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., Warton, D. I., Wintle, B. A., Hartig, F., & Dormann, C. F. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, *40*(8), 913–929. https://doi.org/10.1111/ecog.02881

Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., & Rigol-Sanchez, J. P. (2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, *67*(January), 93–104. https://doi.org/10.1016/j.isprsjprs.2011.11.002

Rondeaux, G., Steven, M., & Baret, F. (1996). Optimization of soil-adjusted vegetation indices. *Remote Sensing of Environment*, *55*(2), 95–107. https://doi.org/10.1016/0034-4257(95)00186-7

Rouse, J. W., Haas, R. H., Schell, J. A., & Deering, D. W. (1974). 'Monitoring vegetation systems in the great plains with ERTS, *Proceedings of Third Earth Resources Technology Satellite Symposium* (Vol. *351* 1 p. 309). Washington, DC.

Schratz, P., Muenchow, J., Iturritxa, E., Richter, J., & Brenning, A. (2019). Hyperparameter tuning and performance assessment of statistical and machine-learning algorithms using spatial data. *Ecological Modelling*, *406*(August), 109–120. https://doi.org/10.1016/j.ecolmodel.2019.06.002

Schulze-Brüninghoff, D., Hensgen, F., Wachendorf, M., & Astor, T. (2019). Methods for LiDAR-based estimation of extensive grassland biomass. *Computers and Electronics in Agriculture*, *156*(January), 693–699. https://doi.org/10.1016/j.compag.2018.11.041

SENAMHI. (2024). Datos Hidrometeorológicos a Nivel Nacional. *Servicio Nacional de Meteorología e Hidrología del Perú. Agro/Monitoreo Agroclimático* https://www.senamhi.gob.pe/?p=estaciones

Strobl, C., Boulesteix, A.-L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics*, *8*(1), 25. https://doi.org/10.1186/1471-2105-8-25

Théau, J., Lauzier-Hudon, E., Aubé, L., & Devillers, N. (2021). Estimation of forage biomass and vegetation cover in grasslands using UAV imagery. *PLoS One*, *16*(1), e0245784. https://doi.org/10.1371/journal.pone.0245784

Vallejos Fernández, L. A., Rojas Guevara, I. B., Perinango Gaitán, J. A., & Alcántara Mendoza, J. (2020). Vacas Pastoreadas a Estaca y Su Efecto Sobre El Consumo y Condición de La Pastura. *UCV-Scientia.*, *11*(1), 28–31. https://doi.org/10.18050/ucv-scientia.v11i1.2400

Vallejos-Cacho, R., Vallejos-Fernández, L. A., Alvarez-García, W. Y., Tapia-Acosta, E. A., Saldanha-Odriozola, S., & Quilcate-Pairazaman, C. E. (2024). Sustainability of Lolium Multiflorum L. "Cajamarquino Ecotype", associated with Trifolium Repens L., at three cutting frequencies in the northern highlands of Peru. *Sustainability*, *16*(16), 6927. https://doi.org/10.3390/su16166927

Van Der Tol, C., Vilfan, N., Dauwe, D., Cendrero-Mateo, M. P., & Yang, P. (2019). The scattering and re-absorption of red and near-infrared chlorophyll fluorescence in the models fluspect and scope. *Remote Sensing of Environment*, *232*(October. 111292. https://doi.org/10.1016/j.rse.2019.111292

Villalobos-Villalobos, L. A., & WingChing-Jones, R. (2023). Forage biomass estimated with a pre-calibrated equation of a rising platemeter in pastures grown in tropical conditions. *Grasses*, *2*(2), 127–141. https://doi.org/10.3390/grasses2020011

Vincini, M., Frazzi, E., & D'Alessio, P. (2008). A broad-band leaf chlorophyll vegetation index at the canopy scale. *Precision Agriculture*, *9*(5), 303–319. https://doi.org/10.1007/s11119-008-9075-z

Wadoux, A. M.-C., Heuvelink, G. B., De Bruin, S., & Brus, D. J. (2021). Spatial cross-validation is not the right way to evaluate map accuracy. *Ecological Modelling*, *457*(October). 109692. https://doi.org/10.1016/j.ecolmodel.2021.109692

Wang, J., Zhang, W., Wang, L., & Yang, H. (2021). Investigating the evolution of tree boosting models with visual analytics. *2021 IEEE 14th Pacific Visualization Symposium (PacificVis)*, 186–195. https://doi.org/10.1109/PacificVis52677.2021.00032 April.

Wei, T., Simko, V., Levy, M., Xie, Y., Jin, Y., & Zemla, J. (2017). Package "Corrplot". *The Statistician*, *56*(316), e24.

Woebbecke, D. M., Meyer, G. E., Von Bargen, K., & Mortensen, D. A. (1995). Color Indices for Weed Identification under Various Soil, Residue, and Lighting Conditions. Transactions of the American Society of Agricultural Engineers, 38(1), 259–269. https://doi.org/10.13031/2013.27838

Zeng, N., Xiaoli Ren, H., He, L., Zhang, P., Li, Z., & Niu (2021). Estimating the grassland aboveground biomass in the three-river headwater region of china using machine learning and Bayesian model averaging. *Environmental Research Letters*, *16*(11), 114020. https://doi.org/10.1088/1748-9326/ac2e85

Zhao, Y., Wang, M., Zhao, T., Luo, Y., Li, Y., Yan, K., Lu, L., Tran, N. N., Wu, X., & Ma, X. (2022). Evaluating the potential of H8/AHI geostationary observations for monitoring vegetation phenology over different ecosystem types in northern China. *International Journal of Applied Earth Observation and Geoinformation*, *112*(August). 102933. https://doi.org/10.1016/j.jag.2022.102933