
# Development of a predictive machine learning model for pasture biomass using satellite vegetation indices and climate data in Spanish Dehesa systems

**Ailí A. Tamayo[1], Laura Moreno[1], Gerardo Moreno[2], Carlos Ferraz[1], Fernando Estelles[3]**

[1]HEMAV Technology SL, Fontsanta 46, 08970 Sant Joan Despí, Barcelona, Spain

[2]Grupo de Investigación Forestal, INDEHESA, Universidad de Extremadura, Plasencia, Spain

[3]Instituto de Ciencia y Tecnología Animal, Universitat Politècnica de València, Camino de Vera s/n, 46022 Valencia, Spain

**A paper from the Proceedings of the**

**17th International Conference on Precision Agriculture and 11th Brazilian Congress on Precision and Digital Agriculture**

**13-15 July 2026**

**Porto Alegre, Brazil**

***Abstract.***

*Extensive livestock systems are fundamental to the ecological, economic, and cultural sustainability of Mediterranean agroecosystems such as the Spanish dehesa. These silvopastoral landscapes support biodiversity, prevent land abandonment, and sustain rural livelihoods, but their productivity is highly dependent on pasture availability. Efficient management therefore requires accurate and timely information on pasture biomass availability, which is strongly influenced by climatic variability, soil properties, and seasonal dynamics. However, traditional field-based biomass measurements are labor-intensive, costly, and spatially limited, making them impractical for continuous monitoring across large and heterogeneous grazing areas.*

*This study develops a scalable and cost-effective predictive framework for estimating pasture biomass using exclusively remotely sensed and environmental data. A dataset of 700 georeferenced ground-truth biomass samples collected across three representative dehesa regions (Salamanca, Cáceres and Badajoz) between 2022 and 2023, was integrated with vegetation indices derived from Sentinel-2 imagery, historical climate variables, and soil characteristics. The samples span multiple phenological stages of the ecosystem, ensuring robust representation of temporal variability. The selected predictors were designed to capture the primary biophysical drivers of pasture growth, including water availability, temperature accumulation, solar radiation, and edaphic conditions.*

*Four machine learning algorithms: Multiple Linear Regression, k-Nearest Neighbors, Random Forest, and XGBoost, were trained and evaluated to assess their predictive performance and robustness. Ensemble-based methods outperformed simpler approaches, with Random Forest (R2 = 0.79, RMSE = 255.49) and XGBoost (R2 = 0.88, RMSE = 238.8) achieving the highest accuracy and generalization capacity. Variable importance analyses revealed that accumulated solar radiation, growing degree days, vegetation indices, and soil properties were the most influential predictors, highlighting the combined role of phenological development and site conditions in determining pasture productivity.*

*The results demonstrate that integrating multi-source environmental data with machine learning*

Tamayo, A. A., Moreno, L., & Moreno, G. (2026). Development of a predictive machine learning model for pasture biomass using satellite vegetation indices and climate data in Spanish Dehesa systems. In Proceedings of the 17th International Conference on Precision Agriculture and 11th Brazilian Congress on Precision and Digital Agriculture (unpaginated, online). Monticello, IL: International Society of Precision Agriculture and Brazilian Congress on Precision and Digital Agriculture.

*techniques enables accurate and operationally scalable estimation of pasture biomass in extensive grazing systems. This approach provides a practical precision livestock farming tool that supports informed decision-making on stocking rates, grazing rotation, and feed supplementation. Moreover, it facilitates adaptive management under increasing climatic variability and supports the long-term sustainability of Mediterranean pastoral systems by enhancing resource efficiency, reducing monitoring costs, and enabling proactive responses to environmental change.*

***Keywords.***
*Pasture biomass, predictive model, vegetation indices, Sentinel-2, machine learning, precision livestock farming, sustainability.*

## Introduction

Natural pastures cover approximately 19% of Spain's total area, equivalent to 9.45 million ha, serving as the productive foundation of extensive livestock systems and delivering significant ecological, landscape, and socioeconomic value across rural territories (INE 2023; San Miguel et al. 2001). In silvopastoral landscapes such as the dehesa, characterised by scattered evergreen oaks over an annual herbaceous understorey, pasture productivity is determined by complex interactions among climatic variability, soil properties, botanical composition, and seasonal phenological dynamics.

Herbaceous communities in Mediterranean climates exhibit a distinctive bimodal growth pattern: a dominant spring flush driven by warming temperatures and soil moisture recharge, and a secondary, lower-amplitude autumn peak; production virtually ceases during the summer drought (San Miguel 2001; Olea et al. 1990). Within this general pattern, considerable inter-annual and spatial variability arises from differences in autumn rainfall, which exerts greater influence on total annual production than spring rainfall and from the buffering effect of soil water-holding capacity (Olea et al. 1990; Colabelli et al. 1998). This dynamic behaviour complicates livestock planning: overgrazing degrades ecosystem structure and productive capacity, while undergrazing promotes fire risk, reduces plant-species diversity, and diminishes landscape value. Reliable, spatially explicit, and temporally continuous information on biomass availability is therefore essential for adaptive grazing management.

Traditional field-based estimation methods, such as destructive cutting plots, rising-plate meters, visual scoring provide accurate measurements but are labour-intensive, costly, and impractical for continuous monitoring across the large, heterogeneous areas characteristic of dehesa management units (Jaurena et al. 2018; Díaz Gaona et al. 2013). Remote sensing offers a scalable alternative: vegetation indices (VIs) derived from multispectral satellite imagery show strong empirical relationships with pasture biophysical parameters, with NDVI alone explaining up to 83% of variance in biomass estimation studies (Ogungbuyi et al. 2023; Furnitto et al. 2025). Sentinel-2 with its 10 m spatial resolution, 13 spectral bands, and 5-day revisit time is particularly suited to pasture monitoring at the farm and landscape scale (European Space Agency n.d.).

Machine learning (ML) algorithms are well-suited to exploit the nonlinear relationships among multi-source predictors and deliver robust, generalisable biomass models. A growing body of literature (2015–2024) has demonstrated the feasibility of satellite-based pasture biomass estimation in diverse regions, including Australia, the United Kingdom, Belgium, Argentina, Colombia, Uruguay, and Spain; using algorithms such as Random Forest (RF), Support Vector Machines (SVM), and gradient boosting, with reported $R^2$ of 0.62–0.92 and average RMSE around 405.8 kg DM $ha^{-1}$ (Furnitto et al. 2025; Shahi et al. 2025; De Rosa et al. 2021). However, studies specifically targeting the Spanish dehesa that integrate multi-source climate and soil covariates alongside satellite spectral data remain scarce, limiting the operational applicability of existing frameworks to this biome.

## Materials and methods

### Study area

The study was conducted in three dehesa areas of northwestern Spain representing a gradient of climatic and edaphic conditions (Figure 1): Villar del Rey (Badajoz), Majadas (Cáceres), and Matilla de los Caños del Río (Salamanca). Mean annual temperatures are 17.8, 16.3, and 13.3 °C, and mean annual precipitation is 436, 572, and 497 mm, respectively (Weatherbit 2025). The terrain is predominantly flat, with average altitudes of 241, 256, and 818 m above sea level.

The vegetation comprises two contrasting strata. The tree stratum (~20% cover) is dominated by holm oak (*Quercus ilex* L. subsp. *ballota*), with scattered cork oak (*Quercus suber* L.) and gall oak (*Quercus faginea* Lam.). The herbaceous stratum covers virtually the entire area and is composed predominantly of annual species including *Tolpis barbata*, *Chamaemelum mixtum*, *Plantago lagopus*, *Echium plantagineum*, and *Cynodon dactylon*. This stratum exhibits a strongly marked phenological cycle with peak green biomass in spring and a secondary flush in autumn driven by the annual rainfall regime and the shallow root systems characteristic of annual Mediterranean grasses (Martín et al. 2020).


**Figure 1 Location of the sampling points (marked in yellow) of the dataset used.**

### Ground-truth biomass data

Reference biophysical data were obtained through destructive sampling in 19 plots of 25×25 m distributed across the three study areas. Within each plot, vegetation was collected from 3–4 quadrants of 25×25 cm, located by non-probabilistic selection in zones visually identified as representative of the whole plot. All rooted vegetation (green and dry) was cut at ground level, weighed fresh in the field using a precision scale, stored in portable refrigerators, and transported to the laboratory. Samples were oven-dried at 60 °C for 48 h and re-weighed to determine dry matter content (Martín et al. 2020). Data collection covered multiple phenological phases between 2022 and 2023, yielding 1020 initial observations, reduced to 722 after satellite image filtering and to 675 after outlier removal.

### Satellite data

Spectral reflectance data were acquired from Sentinel-2 (S2) MultiSpectral Instrument (MSI) imagery, comprising 13 spectral bands from 443 to 2190 nm with spatial resolution of 10–60 m and a 5-day revisit frequency (European Space Agency n.d.). Acquisitions were filtered for cloud contamination over sampling locations; ground-truth records without a valid S2 image within ±5 days of the sampling date were excluded. The 13 vegetation indices listed in Table 1 were computed from the filtered imagery using the rgee package (Aybar 2023) in R via Google Earth Engine.

**Table 1 Vegetation indices derived from Sentinel-2 imagery and used as candidate predictors.**

| VI | Definition | Formula | Reference |
|---|---|---|---|
| NDVI | Normalized Difference Vegetation Index | (NIR−R)/(NIR+R) | Insúa et al. 2024 |
| EVI | Enhanced Vegetation Index | 2.5×(NIR−R)/(NIR+6R−7.5B+1) | Hansen et al. 2016 |
| SAVI | Soil-Adjusted Vegetation Index | 1.5(NIR−R)/(NIR+R+0.5) | Da Silva et al. 2020 |
| OSAVI | Optimized Soil-Adjusted VI | (NIR−R)/(NIR+R+0.16) | Rondeaux et al. 1996 |
| MSAVI | Modified Soil-Adjusted VI | (2NIR+1−√[(2NIR+1)²−8(NIR−R)])/2 | Hansen et al. 2016 |
| NDRE | Red Edge Normalized Difference Index | (NIR−RE)/(NIR+RE) | Insúa et al. 2024 |
| GNDVI | Green Normalized Difference VI | (NIR−G)/(NIR+G) | Insúa et al. 2024 |
| RVI | Ratio Vegetation Index | R/NIR | Richardson & Wiegand 1977 |
| SIPI | Structure-Intensive Pigment Index | (NIR−B)/(NIR+R) | Penuelas et al. 1995 |
| TVI | Transformed Vegetation Index | (NDVI+0.5)^0.5 | Tucker 1979 |
| NDPI | Normalized Difference Phenology Index | (NIR−(0.74R+0.26SWIR))/(NIR+(0.74R+0.26SWIR)) | Wang et al. 2017 |
| SATVI | Soil-Adjusted Total Vegetation Index | (SWIR1−R)/(SWIR1+R+0.1)−SWIR2/2 | Marsett et al. 2006 |

* R = Red; G = Green; B = Blue; RE = Red Edge; NIR = Near Infrared; SWIR = Short-Wave Infrared.

### Climate data

Historical meteorological variables were retrieved through the WeatherbitAPI, which provides climate records compiled from GLDAS and ERA5 reanalyses, complemented by satellite-derived precipitation data (Weatherbit 2025). Variables included minimum, mean, and maximum air temperature; soil temperature and moisture at multiple depths; reference evapotranspiration ($ET_0$); accumulated precipitation; specific air humidity; atmospheric pressure; wind speed at 10 m; and solar radiation. Accumulated versions of temperature, precipitation, $ET_0$, and degree days over 2- and 4-week windows were computed to capture lagged effects on plant growth.

### Soil data

Soil properties were obtained from SoilGrids (Poggio et al. 2021), a global digital soil mapping product that predicts soil properties at 250 m spatial resolution using ML models trained on ~240,000 soil profiles and more than 400 environmental covariates. Variables extracted included soil organic carbon content, total nitrogen, pH (water), cation exchange capacity, bulk density, coarse fragment fraction, and texture fractions (sand, silt, clay) at six standard depths from 0 to 200 cm. All variables, units, and Google Earth Engine API codes are listed in Annex 1 of the full thesis report.

### Modelling methodology

All analyses were performed in R (version 4.5.0). The modelling workflow comprised four stages (Figure 2): (1) dataset construction; (2) data cleaning and variable selection; (3) model development; and (4) performance evaluation.

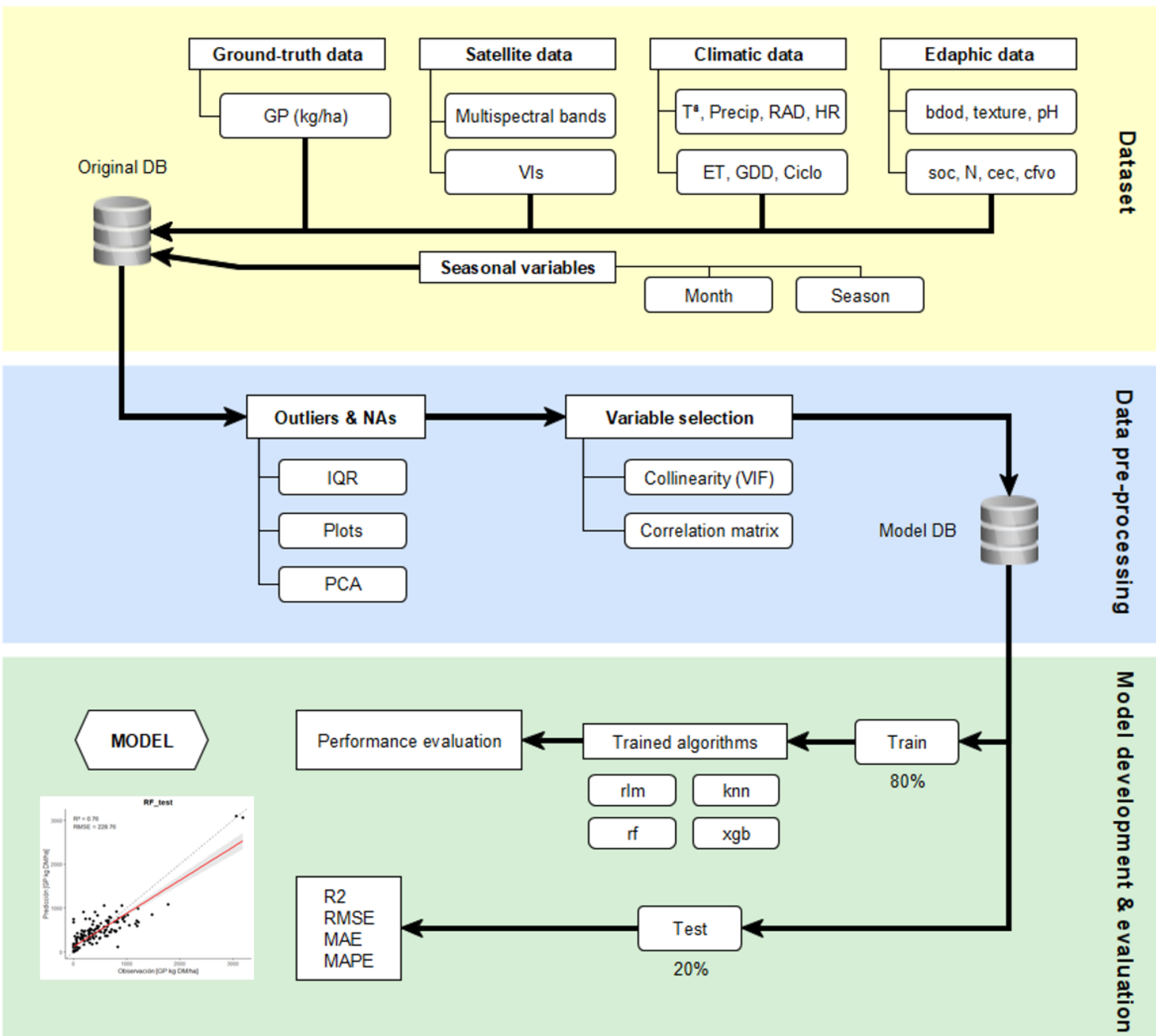


**Figure 2. Diagram of the predictive model development workflow.**

## Dataset construction

A harmonised dataset was assembled by aligning coordinate reference systems (CRS) between ground-truth sample points and Sentinel-2 imagery. Spectral values for all 13 S2 bands were extracted at each sample location, and the 13 VIs in Table 1 were calculated. Soil data were added using sample coordinates as reference points (package rgee, SoilGrids API). Climate variables were added manually from WeatherbitAPI exports in Excel, resulting in a dataset of 63 candidate predictor variables.

## Data pre-processing

Outliers in the target variable (green dry-matter biomass, GP) were removed using the interquartile range (IQR) method, flagging values below Q1 − 1.5×IQR or above Q3 + 1.5×IQR, which eliminated 47 records. Variable relevance was assessed by three complementary approaches. First, a Pearson correlation matrix was computed between all predictors and GP, retaining variables with statistically significant correlations. Second, PCA (package FactoMineR) was used to explore data structure and identify redundant information among collinear predictors. Third, the Variance Inflation Factor (VIF) was calculated for all candidate predictors to detect and eliminate multicollinear variables, following the methodology of Tierney (2003).

The final predictor set of 23 variables was selected by manually removing variables with low significance or high redundancy. It comprised: spectral bands B8, B8A, B9, and B12; seven VIs (MSAVI, NDRE, GNDVI, SATVI, SIPI, NDPI, RVI); six climate accumulators (solar radiation RAD, mean temperature Temp_med_acum_4S, precipitation Precip_acum_2S, relative humidity RH_acum_4S, growing degree days GDD_acum_4S, and evapotranspiration ET_acum_4S); three soil properties at 15 cm depth (bulk density BDOD, pH, and sand fraction Sand); and two temporal variables (Day_Cycle and Month).

## Model development

The dataset was randomly split 80:20 (training:test), stratified to preserve the distribution of the target variable. Four algorithms were trained and hyperparameter-tuned (Table 2): (i) MLR with Lasso regularisation, which penalises the regression coefficients to reduce overfitting and automatically selects relevant predictors; (ii) KNN (k = 5), which classifies or predicts a new observation based on the majority or average of its k nearest neighbours in feature space; (iii) RF (mtry = 8, ntree = 478), an ensemble of independently grown decision trees that aggregates predictions by averaging, reducing variance while maintaining low bias; and (iv) XGBoost (max_depth = 3, η = 0.05, nrounds = 300), a gradient-boosted tree ensemble that sequentially corrects prediction errors, incorporating regularisation terms (gamma, colsample_bytree, min_child_weight, subsample) to prevent overfitting.

**Table 2 Hyperparameters and selection criteria used to validate the best prediction models.**

| Model | Algorithm | Key Parameters | Criteria | R Package |
|---|---|---|---|---|
| LMR | Linear Multiple Regression | lambda.min, intercept | Residual error | lm |
| KNN | k-Nearest Neighbors | k = 5 | RMSE | class |
| RF | Random Forest | mtry = 8, ntree = 478 | OOB $R^2$ | randomForest |
| XGB | eXtreme Gradient Boosting | max_depth=3, η=0.05, nrounds=300, gamma=0, colsample=0.6, min_child=5, subsample=0.6 | Test RMSE (CV) | xgboost |

## Performance evaluation

Predictive performance was assessed on the test subset using the coefficient of determination ($R^2$) and the root mean square error (RMSE), calculated as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Obs_i - Est_i)^2}{\sum_{i=1}^{n}(Obs_i - \overline{Obs})^2} \qquad (1)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Est_i - Obs_i)^2} \qquad (2)$$

Variable importance was extracted from each model using algorithm-specific metrics: Lasso coefficient magnitude; KNN sensitivity analysis; RF permutation importance (mean decrease in accuracy); and XGBoost gain-based importance. This allowed identification of the most influential predictor types across models.

## Results and Discussion

### Dataset characteristics and seasonal patterns

After Sentinel-2 filtering and outlier removal, the final dataset comprised 675 georeferenced observations distributed across the three study areas. Green biomass ranged from 0 to 4816 kg DM ha$^{-1}$, with a mean value of 444.3 kg DM ha$^{-1}$. The distribution was positively skewed, reflecting the prevalence of low-to-moderate values and sporadic high-production events. The number of observations per month ranged from 32 to 118, reflecting the uneven distribution of sampling effort and cloud-free satellite acquisitions across the two years.

Seasonal patterns were consistent with the literature for annual Mediterranean pastures (San Miguel 2001; Olea et al. 1990): highest biomass was recorded in spring (particularly April–May), a secondary peak was observed in autumn, and near-zero values occurred in summer drought period (Figure 3). Cáceres showed the highest spring medians, likely reflecting the combined effect of antecedent autumn rainfall and rising temperatures, yet also exhibited the greatest inter-annual variability, which can challenge model generalisation. Fewer cloud-free acquisitions were available in 2022 due to higher rainfall, resulting in a temporal imbalance that may limit the model's ability to capture inter-annual dynamics fully.

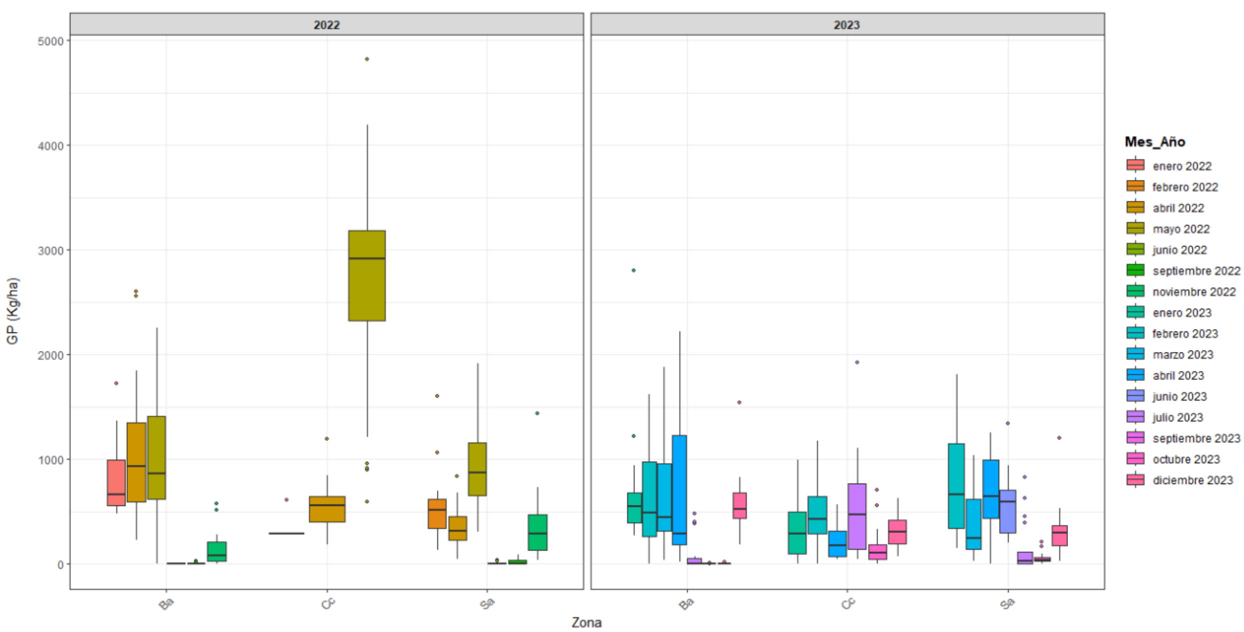


**Figure 3. Distribution of green grass biomass by study area and month. Ba = Badajoz, Cc = Cáceres and Sa = Salamanca.**

The PCA analysis showed that two principal components explained 56.6% of total variance, reaching 85% with six components. The first component was dominated by VIs and accumulated climate variables, and the second by spectral bands and soil data. Although some separation among study areas was observed, the overlap among sites suggested that a common predictive framework could be developed across the three dehesa systems.

### Variable selection

The Pearson correlation matrix identified significant relationships between GP and: (i) spectral bands B6, B7, B8, B8A, B9, B11, and B12 (r = 0.25–0.39, highest for B8 and B9); (ii) all 13 VIs (r = 0.37–0.39), with the exception of SATVI, which showed lower correlation due to its sensitivity to bare-soil exposure rather than green biomass per se; (iii) accumulated climate variables, two-week precipitation, Day_Cycle, four-week mean temperature, growing degree days, radiation, and relative humidity (r = 0.13–0.32); and (iv) soil bulk density, pH, and sand content at 15 cm depth. The month variable was the only temporal predictor significantly correlated with GP.

VIF analysis was used to detect severe multicollinearity among VIs (particularly NDPI, TVI, MSAVI, NDVI, OSAVI, SAVI, and NDRE), and among cumulative climate variables (all accumulation windows derived from the same base series). Variables with VIF > 5 were progressively removed while preserving one representative from each correlated group.

### Model performance and variable importance

The predictive performance of the four evaluated models is summarized in Table 3, whereas Figure 4 visually depicts the agreement between observed and predicted pasture biomass values for the best-performing models.

**Table 3 Performance metrics ($R^2$ and RMSE) for each prediction model.**

| | Calibration (n= 540) | | Validation (n=135) | |
|---|---|---|---|---|
| Model | $R^2$ | RMSE (kg DM ha$^{-1}$) | $R^2$ | RMSE (kg DM ha$^{-1}$) |
| LMR (Lasso) | | | 0.49 | 422.59 |
| KNN | 0.87 | 213.15 | 0.85 | 267.74 |
| RF | 0.94 | 132.80 | 0.79 | 255.49 |
| XGBoost | 0.95 | 136.67 | 0.88 | 238.81 |

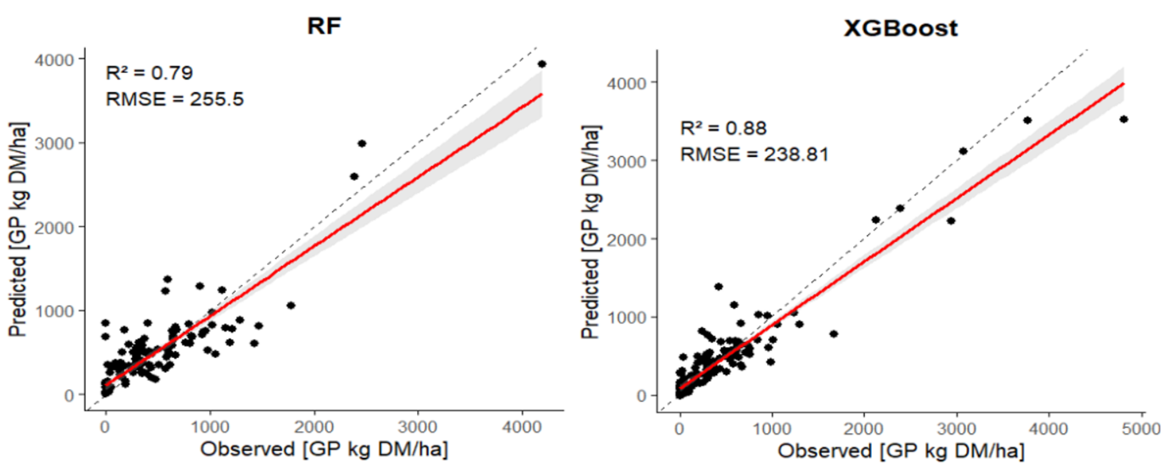


**Figure 4. Scatter plots (test subset) of the model's predictions with better performance versus observations after linear correction that are displayed with performance metrics (R2 and RMSE).**

The MLR Lasso model showed the weakest predictive performance ($R^2$ = 0.49, RMSE = 422.59 kg DM ha$^{-1}$), with a systematic tendency to underestimate high biomass values. The poor performance confirms that linear relationships are insufficient to capture the complex, nonlinear interactions between multi-source predictors and pasture productivity in heterogeneous dehesa environments.

The KNN model achieved a high coefficient of determination ($R^2$ = 0.85), outperforming UAV-based approaches that use only VIs ($R^2$ = 0.62; Álvarez-Mendoza et al. 2022). However, its RMSE

(267.74 kg DM ha⁻¹) was the highest among nonlinear models, which indicates that may be sensitive to the distribution of the training data for extreme values.

RF achieved $R^2 = 0.79$ and the lowest RMSE (255.49 kg DM ha⁻¹), placing it within the 0.73–0.92 range reported for comparable studies (Furnitto et al. 2025). The slight underestimation at low biomass values is likely attributable to the high variability of ground-truth data at the lower end of the distribution.

XGBoost delivered the best overall performance ($R^2 = 0.88$, RMSE = 238.81 kg DM ha⁻¹), exceeding the highest $R^2$ previously reported for dehesa-type systems (0.78; Furnitto et al. 2025) and substantially outperforming UAV-only VI models (Álvarez-Mendoza et al. 2022). The predicted values were closely aligned with observations across the full biomass range, with the regression line nearly coinciding with the ideal 1:1 reference.

Based on the importance analysis shown in Figure 5. Across all nonlinear models, cumulative climate variables: solar radiation, growing degree days, and precipitation, emerged consistently as the most influential predictors. This is consistent with the physiology of Mediterranean annual grasses, in which thermal accumulation drives leaf appearance rate, radiation determines photosynthetic potential, and water availability constrains leaf area expansion and tillering (Colabelli et al. 1998). Edaphic variables (bulk density, sand fraction, pH) provided secondary but meaningful discriminative power, reflecting the well-documented relationship between soil water-holding capacity and annual pasture production in dehesa systems (Olea et al. 1990).

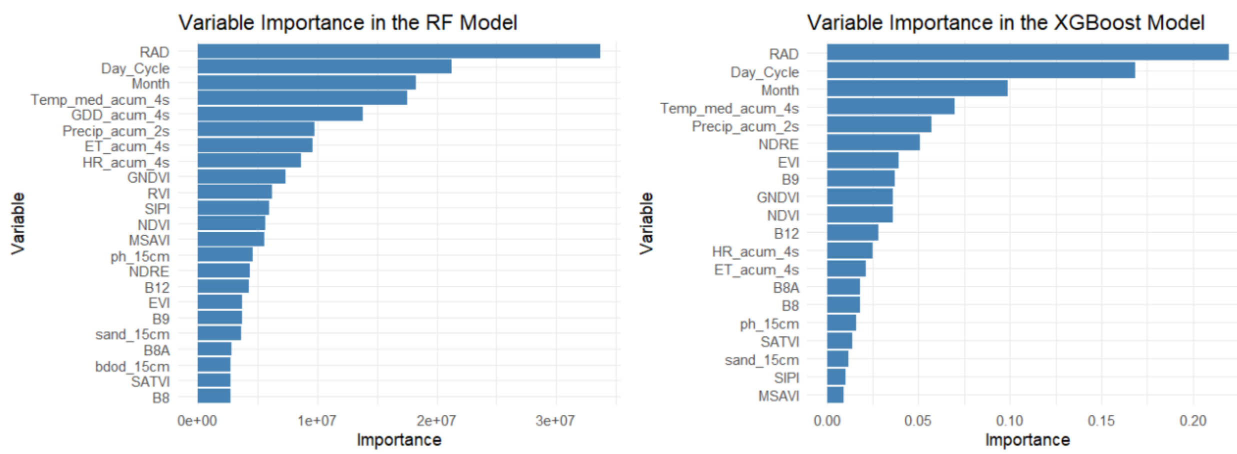


**Figure 5. Relative importance of predictor variables in Random Forest, and XGBoost models developed for pasture green biomass prediction. Higher importance values indicate a greater contribution of the predictor to model performance.**

## Conclusion

This study demonstrated the feasibility of developing accurate predictive models of natural pasture biomass in Spanish dehesa systems using exclusively remotely sensed, climatic, and edaphic variables, without reliance on real-time in-situ measurements.

Among the four machine learning algorithms evaluated, XGBoost and Random Forest achieved the highest predictive performance, followed by KNN. In contrast, the multiple linear regression model performed substantially worse, highlighting the predominantly nonlinear relationships governing pasture productivity. Across the nonlinear models, temporal and cumulative climatic variables, namely solar radiation, growing degree days, cycle day, and precipitation accumulators, emerged as the most influential predictors, underscoring the central role of water–energy balance and phenological development in Mediterranean annual grasslands. Soil properties, particularly bulk density, texture, and pH, provided complementary explanatory power at the site level, while spectral vegetation indices contributed additional information related to vegetation greenness and condition.

The proposed framework provides a foundation for a scalable precision livestock farming tool capable of supporting management decisions such as stocking rate adjustment, grazing rotation planning, supplementary feeding strategies, and adaptive responses to climatic variability, while minimizing the need for continuous field sampling. Future research should focus on expanding the dataset across a broader range of years, locations, and environmental conditions, incorporating management-related variables, and validating model performance in additional agro-climatic regions of Spain to enhance operational applicability. Furthermore, the integration of advanced temporal machine learning architectures and higher-temporal-resolution satellite observations may further improve prediction accuracy, robustness, and transferability under increasingly variable climatic conditions.

## References


Álvarez-Mendoza CI, Guzman D, Casas J, et al. (2022) Predictive modeling of above-ground biomass in Brachiaria pastures from satellite and UAV imagery using machine learning approaches. Remote Sensing 14(22): 5870.

Aybar C (2023) R Bindings for Calling the Earth Engine API [R package rgee version 1.1.7]. CRAN: Contributed Packages. doi:10.32614/CRAN.PACKAGE.RGEE

Colabelli M, Agnusdei MG, Mazzanti A, Labreveux M (1998) The process of growth and development of forage grasses as a basis for defoliation management. INTA Balcarce Experimental Station, Balcarce, AR.

Da Silva EE, Rojo Baio FH, Ribeiro Teodoro LP, et al. (2020) UAV-multispectral and vegetation indices in soybean grain yield prediction. Remote Sensing Applications: Society and Environment 18: 100318.

De Rosa D, Basso B, Fasiolo M, et al. (2021) Predicting pasture biomass using a statistical model and machine learning algorithm implemented with remotely sensed imagery. Computers and Electronics in Agriculture 180: 105880.

Díaz Gaona C, Rodríguez Estévez V, Sánchez Rodríguez M, et al. (2013) Estudio de los pastos en Andalucía y Castilla la Mancha y su uso racional con ganadería ecológica. Ecovalia, Seville, ES.

European Space Agency (n.d.) SENTINEL 2. https://www.esa.int/Space_in_Member_States/Spain/SENTINEL_2. Accessed 1 March 2025.

Furnitto N, Ramírez-Cuesta JM, Intrigliolo DS, Todde G, Failla S (2025) Remote sensing for pasture biomass quantity and quality assessment: challenges and future prospects. Smart Agricultural Technology 12: 101057.

Hansen MC, Potapov PV, Goetz SJ, et al. (2016) Mapping tree height distributions in Sub-Saharan Africa using Landsat 7 and 8 data. Remote Sensing of Environment 185: 221–232.

INE (2023) Survey on the Structure of Agricultural Holdings (EEA). Year 2023. National Institute of Statistics, Madrid, ES.

Insúa J, Laplacette C, Utsumi S, et al. (2024) Innovation for pasture management, ATN/RF-18077-RG. FONTAGRO, Washington, DC, USA.

Jaurena M, Porcile V, Baptista R, Carriquiry E, Díaz S (2018) La Regla Verde: una herramienta para el manejo del campo natural. Revista INIA Uruguay 54: 24–27.

Marsett RC, Qi J, Heilman P, et al. (2006) Remote sensing for grassland management in the arid Southwest. Rangeland Ecology and Management 59(5): 530–536.

Martín MP, Pacheco-Labrador J, González-Cascón R, et al. (2020) Estimation of essential vegetation variables in a dehesa ecosystem using seasonally simulated reflectivity factors. Revista de Teledetección 55: 31.

Ogungbuyi MG, Mohammed C, Ara I, Fischer AM, Harrison MT (2023) Advancing skyborne technologies and high-resolution satellites for pasture monitoring and improved management: a review. Remote Sensing 15(19): 4866.

Olea Márquez de Prado L, Verdasco Giralt P, Paredes Galán J (1990) Características y producción de los pastos de los dehesares del suroeste de la Península Ibérica. Pastos 20–21: 131–156.

Penuelas J, Baret F, Filella I (1995) Semi-empirical indices to assess carotenoids/chlorophyll ratio from leaf spectral reflectance. Photosynthetica 31(2): 221–230.

Poggio L, De Sousa LM, Batjes NH, et al. (2021) SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty. SOIL 7(1): 217–240.

Richardson AJ, Wiegand CL (1977) Distinguishing vegetation from soil background information. Photogrammetric Engineering and Remote Sensing 43: 1541–1552.

Rondeaux G, Steven M, Baret F (1996) Optimization of soil-adjusted vegetation indices. Remote Sensing of Environment 55(2): 95–107.

San Miguel A (2001) Pastos Naturales Españoles, 1st edn. Ediciones Mundi-prensa, Madrid, ES.

San Miguel Ayanz A, Olea Márquez de Prado L, Ferrer Benimeli C (2001) Nomenclátor básico de pastos en España. Pastos 31(1): 7–44.

Shahi TB, Balasubramaniam T, Sabir K, Nayak R (2025) Pasture monitoring using remote sensing and machine learning: a review of methods and applications. Remote Sensing Applications: Society and Environment 37: 101459.

Tierney L (2003) Name Space Management for R. R Project, Vienna, AT.

Tucker CJ (1979) Red and photographic infrared linear combinations for monitoring vegetation. Remote Sensing of Environment 8(2): 127–150.

Wang C, Chen J, Wu J, et al. (2017) A snow-free vegetation index for improved monitoring of vegetation spring green-up date in deciduous ecosystems. Remote Sensing of Environment 196: 1–12.

Weatherbit (2025) Agriculture Historical Weather Data API Documentation. Weatherbit. https://www.weatherbit.io/api/ag-weather-api. Accessed 1 March 2025.

The authors used Claude (Anthropic, San Francisco, CA, USA) as an AI writing assistance tool to support language editing and manuscript formatting. All scientific content, data analysis, and conclusions are the sole responsibility of the authors.