
**Matias Lucero[1], Carlos M. Hernández[1], Ana Carcedo[1], Planet Labs Inc[2], AGCO[3], and Ignacio Ciampitti[1].**

[1] Department of Agronomy, Kansas State University, Manhattan, KS, 66506 USA; mflucero@ksu.edu (M.L.); carlosh92@ksu.edu (C.H.); carcedo@ksu.edu (A.C.); ciampitti@ksu.edu (I.C.).

[2] Planet Labs Inc., San Francisco, CA 94107, USA; rasmus.houborg@planet.com (R.H.); ariel.zajdband@planet.com (A.Z.).

[3] AGCO Corporation., Duluth, GA 30096, USA.

**A paper from the Proceedings of the**
**16th International Conference on Precision Agriculture**
**21-24 July 2024**
**Manhattan, Kansas, United States**

***Abstract.***

Satellite imagery is an invaluable decision-support instrument for enhancing on-farm management practices. Creating predictive tools for estimating pasture biomass has emerged as a promising method to assist farmers in determining biomass and maximizing yield. This research aims to develop a new and innovative framework that combines field data and satellite images to estimate aboveground biomass in alfalfa (Medicago sativa L) at on-farm scale. Throughout the 2022 growing season, alfalfa biomass samples were collected, dried, and georeferenced on various mowing dates across different fields in Kansas, USA. The satellite data employed included four sources: Sentinel-2, Planet Scope, Planet Fusion, and Biomass Proxy. To build the predictive model, a Bayesian additive regression tree (BART) and nested cross-validation were implemented. Permutation cross-validation was utilized to identify key variables, thereby avoiding correlation and collinearity among variables. The NIR band from Planet Scope and the version 100% S2 of Biomass Proxy were included due to their significant correlation with alfalfa biomass. The model's performance was assessed using statistical metrics, showing an $R^2$ of 0.61 and an RMSE of 0.15 $kg/m^{-2}$. Additionally, this framework provided uncertainty quantifications to evaluate prediction uncertainty ranges. In conclusion, integrating remote sensing, field data, and a nonparametric approach resulted in an effective predictive tool that could potentially optimize farmers' management decisions.

***Keywords.***
*Medicago sativa, Precision agriculture, Regression model, Machine learning, BART.*

## Defining a framework for spatial mapping alfalfa biomass using remote sensing.

### Introduction:

Alfalfa (Medicago sativa L.) is a perennial forage with high productivity, rich nutrient content (Elfanssi et al., 2018). Alfalfa serves various purposes, including hay, silage, pasture, and biofuel. However, poor mowing or grazing schedules can result in inconsistent production (Lu, 2006). To address this, satellite remote sensing models have been developed to estimate forage quantity and quality, aiding in better management strategies (Reinermann et al., 2020). One promising method is Bayesian additive regression trees (BART), which accounts for non-linear relationships and prediction uncertainties, enhancing decision-making in forage management (Chipman et al., 2010).

### Materials and Methods:

The study was conducted in Wichita, Kansas, during the summer of 2022. Biomass samples were collected within a one-square-meter area. Four data sources were used: Sentinel-2, Planet Fusion, Planet Scope, and Biomass Proxy. Using R, datasets from each satellite source were processed. The "var_selection_by_permute_cv" function from the bartMachine R package was used for feature selection to identify the most relevant spectral bands for estimating alfalfa biomass. Hyperparameter tuning was conducted using a grid search to explore combinations of hyperparameters. Nested cross-validation was employed to enhance model fitting and mitigate overfitting. The dataset was divided into k-folds (outer loop) based on the field, with further partitioning (inner loop) to isolate field and sampling date combinations. This process estimated tuning parameters and built three models, one per field. The regression model for forecasting biomass yield was constructed using the. The model output represented dry biomass yield from the 95% credible intervals.

### Results:

The feature selection analysis revealed significant variations in variable importance among remote sensing bands. NIR (6%) from Planet Scope and Biomass Proxy version 100S2 (6%) exhibited the highest correlation with alfalfa yield estimation. Performance metrics included R2: 0.61, and RMSE: 0.29 kg.m$^{-2}$. Fig 1 shows the result between values observed and predictive.

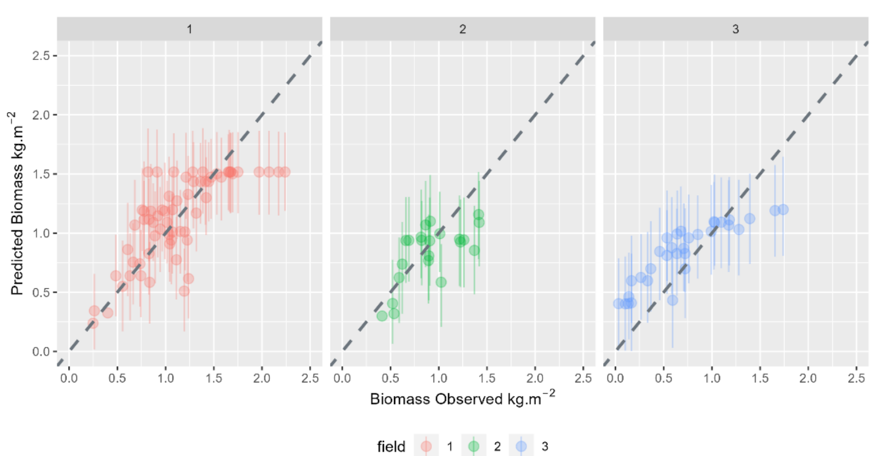


**Fig1 1. Accuracy evaluation of alfalfa regression model. Relation between observed and predicted biomass. The vertical lines represent the interval of 95 of the probability. Different blocks and colors represent the fields (red = field 1, green= field 2, and blue= field 3).**

## Discussion:

A new framework for estimating biomass yield eas presented in this study, integrating field data and satellite data, within a robust cross-validation, and hyperparameter tuning to develop a predictive model with a small dataset. Using an ensemble nested method (Dinh & Aires, 2022), it addresses model overfitting and spatial autocorrelation (Roberts et al., 2017). Bayesian Additive Regression Trees (BART) offer richer information than traditional regression methods, incorporating uncertainty quantification (Williams et al., 2011)and credible intervals for better estimation accuracy (Shirley et al., 2020). The study's findings provide a promising solution to remote sensing and pasture modeling challenges with small datasets.The study highlights the importance of feature selection in remote sensing, emphasizing the need to identify the most relevant spectral bands to minimize classification errors (Kiala et al., 2019). Using permutation importance cross-validation (Zhong et al., 2023), it successfully selects stable features, reducing multicollinearity and enhancing prediction accuracy. A key limitation of previous studies is the lack of spatial uncertainty measurements. This study addresses this by using a BART approach, which has outperformed other methods like gradient-boosting machines and random forests (Hill et al., 2020).

## Conclusion:

An innovative framework was introduced in this study for developing predictive, achieving promising results (R2: 0.61; RMSE: 0.15 kg.m-2. It used nested cross-validation and hyperparameter tuning to effectively quantify prediction uncertainty. This method is a significant step towards creating a nondestructive tool using satellite imagery.

## References:


Chipman, H. A., George, E. I., & McCulloch, R. E. (2010). BART: Bayesian additive regression trees. *The Annals of Applied Statistics*, *4*(1). https://doi.org/10.1214/09-AOAS285

Dinh, T. L. A., & Aires, F. (2022). Nested leave-two-out cross-validation for the optimal crop yield model selection. *Geoscientific Model Development*, *15*(9), 3519–3535. https://doi.org/10.5194/gmd-15-3519-2022

Elfanssi, S., Ouazzani, N., & Mandi, L. (2018). Soil properties and agro-physiological responses of alfalfa (Medicago sativa L.) irrigated by treated domestic wastewater. *Agricultural Water Management*, *202*, 231–240. https://doi.org/10.1016/j.agwat.2018.02.003

Hill, J., Linero, A., & Murray, J. (2020). Bayesian Additive Regression Trees: A Review and Look Forward. *Annual Review of Statistics and Its Application*, *7*(1), 251–278. https://doi.org/10.1146/annurev-statistics-031219-041110

Kiala, Z., Mutanga, O., Odindi, J., & Peerbhay, K. (2019). Feature Selection on Sentinel-2 Multispectral Imagery for Mapping a Landscape Infested by Parthenium Weed. *Remote Sensing*, *11*(16), 1892. https://doi.org/10.3390/rs11161892

Lu, D. (2006). The potential and challenge of remote sensing‐based biomass estimation. *International Journal of Remote Sensing*, *27*(7), 1297–1328. https://doi.org/10.1080/01431160500486732

Reinermann, S., Asam, S., & Kuenzer, C. (2020). Remote Sensing of Grassland Production and Management—A Review. *Remote Sensing*, *12*(12), 1949. https://doi.org/10.3390/rs12121949

Roberts, J., Curran, M., Poynter, S., Moy, A., Ommen, T. V., Vance, T., Tozer, C., Graham, F. S., Young, D. A., Plummer, C., Pedro, J., Blankenship, D., & Siegert, M. (2017). Correlation confidence limits for unevenly sampled data. *Computers & Geosciences*, *104*, 120–124. https://doi.org/10.1016/j.cageo.2016.09.011

Shirley, R., Pope, E., Bartlett, M., Oliver, S., Quadrianto, N., Hurley, P., Duivenvoorden, S., Rooney, P., Barrett, A. B., Kent, C., & Bacon, J. (2020). An empirical, Bayesian approach to modelling crop yield: Maize in USA. *Environmental Research Communications*, *2*(2), 025002. https://doi.org/10.1088/2515-7620/ab67f0

Williams, E., Piepho, H.-P., & Whitaker, D. (2011). Augmented p‐rep designs. *Biometrical Journal*, *53*(1), 19–27. https://doi.org/10.1002/bimj.201000102

Zhong, Y., Chalise, P., & He, J. (2023). Nested cross-validation with ensemble feature selection and classification model for high-dimensional biological data. *Communications in Statistics - Simulation and Computation*, *52*(1), 110–125. https://doi.org/10.1080/03610918.2020.1850790