The International Society of Precision Agriculture presents the
15th International Conference on Precision Agriculture
26–29 JUNE 2022
Minneapolis Marriott City Center | Minneapolis, Minnesota USA

PREDICTING BELOW AND ABOVE GROUND PEANUT BIOMASS AND MATURITY USING MULTI-TARGET REGRESSION

Mailson F. Oliveira[*1], Franciele M Carneiro[1], Megan Thurmond[1], Marina D del Val[1], Luan P Oliveira[2], Brenda V. Ortiz[1], Alvaro Sanz-Saez[1], Danilo Tedesco[3]

Auburn University, Department of Crop Soil and Environmental Sciences, Auburn, AL, U.S.

University of Nebraska–Lincoln, Water, and Integrated Cropping Systems Asst. Extension Educator

São Paulo State University, Department of Engineering and Mathematical Sciences, Jaboticabal, SP, BR

A paper from the Proceedings of the
15th International Conference on Precision Agriculture
June 26-29, 2022
Minneapolis, Minnesota, United States

Abstract. *Peanut growth and maturity prediction can help farmers and breeding programs improve crop management. Remote sensing images collected by satellites and drones make possible and accurate in- season crop monitoring. Today, empirical relationships between crop biomass and spectral reflectance could be used to predict single variables such as aboveground crop biomass, pod weight (PW), or peanut maturity. Robust algorithms such as multi-output regression (MTR) implemented through multioutput random forest (RF) and K-nearest neighbor regression algorithms capable of predicting multioutput variables have not been proposed for peanut management. We developed experiments to predict multiple peanut growth variables using the MTR approach. The experiment was conducted in 2021 on an 8.5 hectare irrigated commercial peanut field located near Auburn, Alabama. The field was divided into square grids (0.01 hectare size), and 20 grids of contrasting soil characteristics were selected for data collection. Starting 92 days after planning, peanut biomass samples were collected weekly from 1.5 m row length inside each grid. Assessment of peanut maturity was done manually on 200-pod sample using the hull-scrape method and the peanut profile board. Peanut maturity indices (PMI) were calculated using two equations, one considering pods from Brown to Black class and the other considering Orange to Black classes. Aboveground biomass was also estimated from each sampling location. Multi-output regression (MTR) models were built to establish a functional relationship between peanut above ground biomass, maturity, and spectral reflectance changes of the canopy over time. Reflectance from individual specific spectral bands and also vegetation indices (VI) of the study field were extracted from Planet Labs' satellite images. The indices NDVI, GNDVI, NLI, MNLI, SAVI, and spectral bands were used as explanatory variables. Training (80% of the original data set) and cross-validation (20% of data) of algorithms were developed using toolkits available in the Scikit-learn python library. The metric to analyze the performance of the algorithms was the mean absolute error MAE. The RF algorithm outputted multiple numeric values of PMI upon VIs and spectral bands, supporting our hypothesis that MTR can predict the maturity of peanut at the field level. The use of spectral reflectance from the Planet Labs imagery*

The authors are solely responsible for the content of this paper, which is not a refereed publication. Citation of this work should state that it is from the Proceedings of the 15th International Conference on Precision Agriculture. EXAMPLE: Last Name, A. B. & Coauthor, C. D. (2018). Title of paper. In Proceedings of the 15th International Conference on Precision Agriculture (unpaginated, online). Monticello, IL: International Society of Precision Agriculture.

to assess peanut maturity resulted on a prediction error of 0.09 % for PMI using the brown to black pods and 0.13 % when predicting PMI using orange to black pods. When the MTR model was evaluated for its accuracy in predicting PW and biomass, a small prediction error was observed for aboveground biomass MAE = 892.62) compared to PW (MAE = 1039.19). Our findings demonstrated a promising method to assess within-field variability of peanut maturity using remote sensing images which could reduce the subjectivity of the manual method. Another promising outcome is that the spatial and temporal prediction of peanut aboveground and belowground biomass could support farmers and researchers decisions not only with respect to harvest but also market and even plant breeding. Future research should focus on integrating other explanatory variables, mainly related to topography and soil conditions like temperature. These variables could help understand the driving factors of peanut growing and maturation at the field level.

Keywords.
biomass, crop maturity, machine learning, multi-task learning, remote sensing, peanut.

Introduction

Assessment of peanut pod maturity is not only linked to productivity (quantity and quality) but also yield losses. Determining the ideal time to start the harvest is one of the most challenging decisions during peanut cultivation (Colvin et al., 2018). Early or late harvest reduces productivity and product quality (Ab-El Monsef et al., 2019).

One way to determine pod maturation is to use the maturation table (Peanut Profile Board, Williams and Drexler 1981) in conjunction with the peanut maturation index (Peanut Maturation Index - PMI, Rowland et al., 2006) which considers optimal maturation when the peanut sample has a PMI of 0.7. Overall, this manual method is laborious, subjective, and requires dense sampling, as peanuts show different maturation rates at different locations within a peanut field (Vellidis and Beasley 2013). Therefore, alternative non-destructive methods that can estimate the within-field maturation across production fields, instead of the maturity of peanut from randomly selected plants, are necessary.

Alternatives to non-destructively peanut maturation estimation were studied by Rowland et al. (2008). They showed a correlation between the infrared spectral bands and the maturation of peanut pods using proximal remote sensing. On a plot scale, Vellidis and Beasley (2013) demonstrated that the Non-Linear Vegetation Index (NLI) could be used to estimate the maturation of the peanut crop. Santos (2019) using data from production peanut fields confirmed the existence of a linear relationship between peanut maturation and modified vegetation indices, with the Modified Non-linear Index (MNLI) being the most promising index next to the NLI.

High-resolution satellite imagery and drone images were used to develop non-linear and neural network models to predict peanut maturity (Santos et al., 2021; Santos et al., 2022). These researches highlighted the use of remote sensing to predict peanut biophysical variables. Therefore, there is no model capable of predicting multi-output peanut biophysical variables to our knowledge. An alternative to output multiple numeric values of peanut biophysical variables, we think, would be multi-target regression (MTR). MTR can predict multiple output processing numeric input variables (Melki et al., 2017). MTR can have advantages relative to single-target regression (STR), with better predictive performance and learning from several tasks (Borchani et al., 2015). Based on this rationale, the objective of this research was to predict multiple peanut variables using the MTR approach.

Material and Methods

A commercial irrigated field located in Society Hill, AL, U.S, was used in this experiment during the 2021 growing season (Figure 1). The irrigated field was planted on May 26, 2021 using the peanut runner-type cultivar ACI 3321, which has a growing cycle of approximately 145 days.

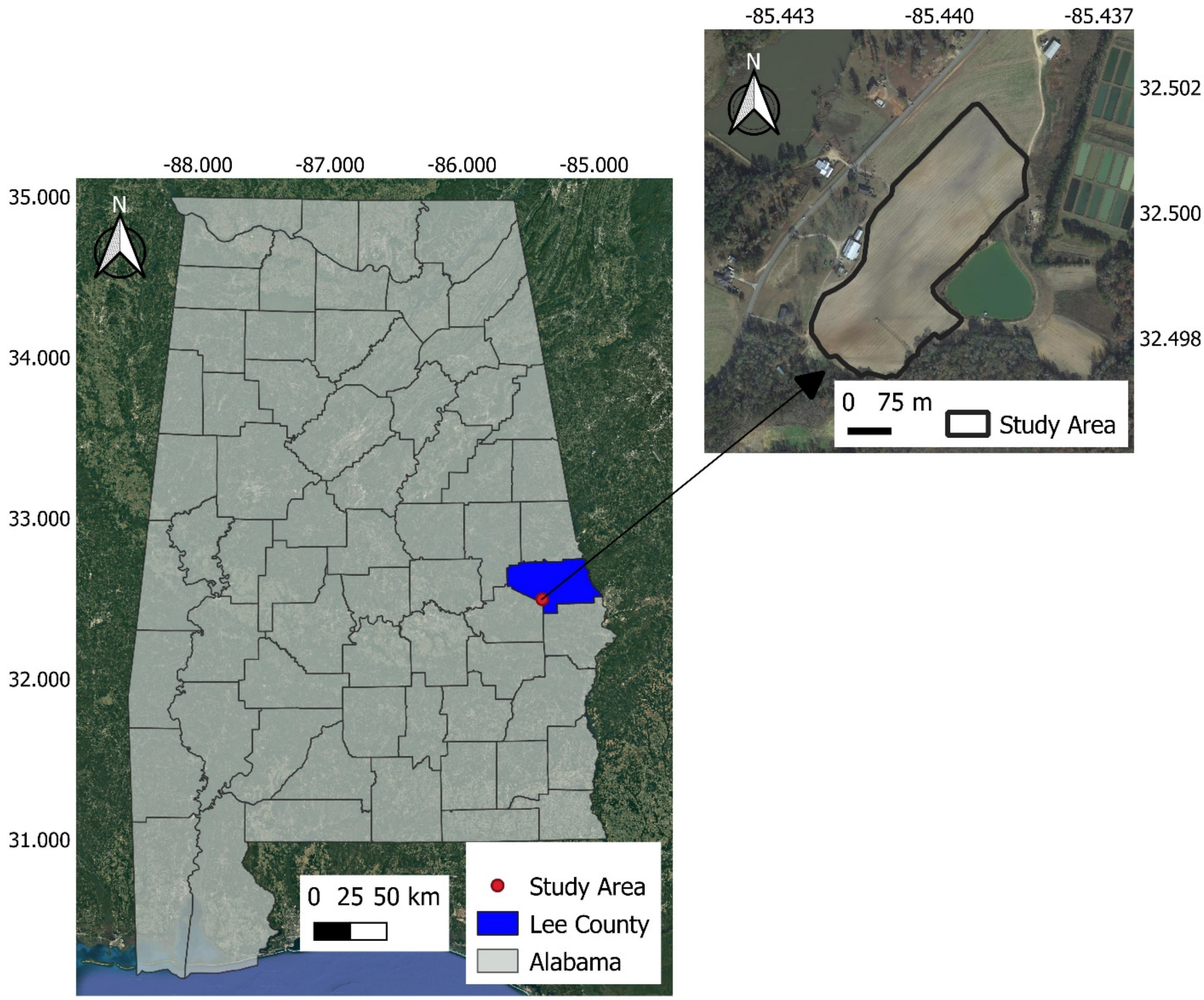

Figure 1. Location of the study area.

Ground data sampling

The field was divided into square grids (0.01 hectare size), and 20 grids of contrasting soil characteristics were selected for data collection. Peanut biomass samples were collected weekly from 1.5 m row length inside each grid, starting 92 days after planting. This study divided the biomass into above-ground biomass (AGB) and below-ground biomass (Pod weight, Figure2). Assessment of peanut maturity was done manually on 200-pod subsample sample, extracted from the 1.5 m biomass sample, using the hull-scrape method and the peanut profile board (Wiliams and Drexler, 1981; Figure 3).

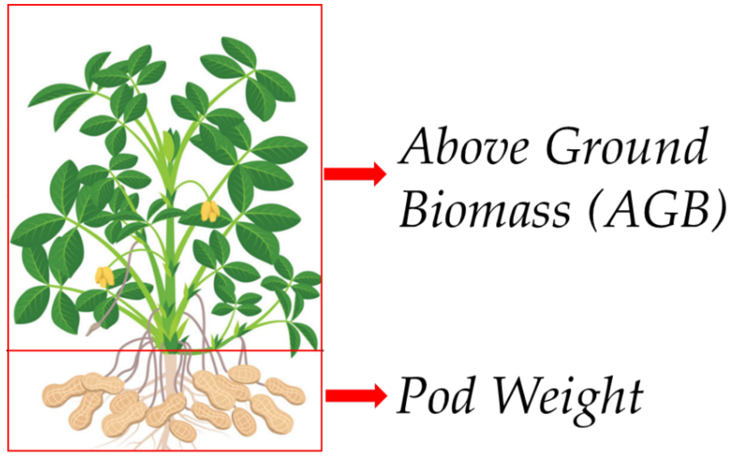

Figure 2. Example of what was considered AGB and pod weight.

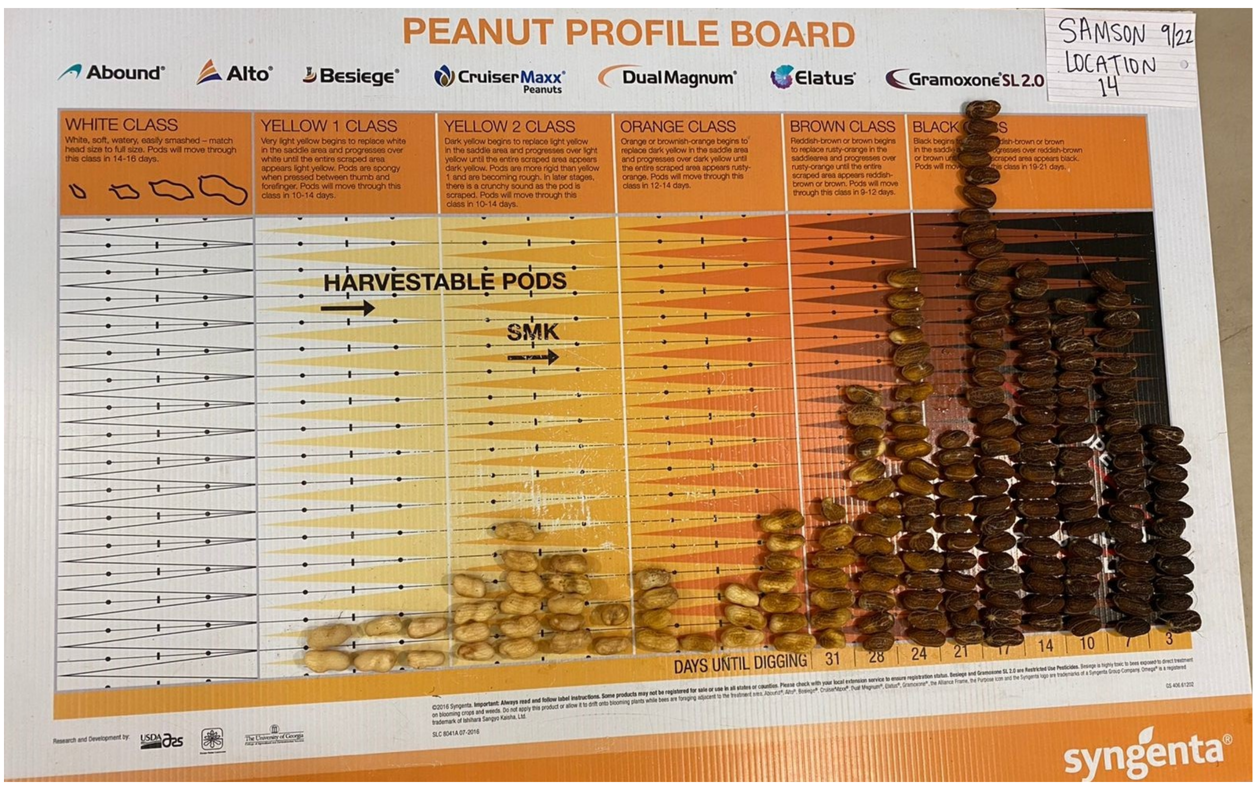

Figure 3. Peanut profile board with pods sorted by color class.

Peanut maturity indices (PMI) were calculated using two equations, one considering pods from Brown to Black class and other considering Orange to Black classes (Figure 3) (Equations 1 and 2).

$$PMI_{BB} = \frac{N\_bbp}{T\_p} \tag{1}$$

$$PMIob = \frac{N\_obp}{T\_p} \tag{2}$$

Where, PMI_{BB} is the peanut maturity index considering brown to black pods class
PMI_{ob} is the peanut maturity index considering orange to black pods class
N_bbp is the number of pods in the brown and black class
N_obp is the number of pods in the orange, brown and black class

Sattelite Imagery

To establish a functional relationship between peanut biophysical variables and spectral

reflectance changes of the canopy over time, Planet Labs imagery data was used to extract reflectance from specific spectral bands and calculate several vegetation indices (VIs).
The surface reflectance Ortho Scene product was acquired from PlanetScope, Planet Labs, Inc., San Francisco, USA (Planet, 2020) under a student license. Cloud Planet Scope satellite data provide 3 m spatial resolution images. The PlanetScope satellite data used had four spectral bands: blue (455–515 nm), green (500–590 nm), red (590–670 nm), and near-infrared (NIR, 780 – 860 nm) in a 16-bit GeoTiff format. The spectral band images were carefully selected for days with 0% cloud over the study area.

Vegetation Indices

Five VIs used in previous studies to predict peanut maturity were selected as input for developing the machine learning models (Table 1).

Vegetation Index	Equation	Reference
NDVI	(NIR−RED)/(NIR + RED)	Rouse et al. (1974)
NLI	(NIR^2−RED)/(NIR^2 + RED)	Goel and Qin (1994)
GNDVI	(NIR−Green)/(NIR + Green)	Gitelson and Merzlyak (1996)
MNLI	(NIR^2−RED) × (1 + L)/(NIR^2 + RED + L)	Gong et al. (2003)
SAVI	(NIR−RED)/(NIR + RED + L) * (1 + L)	Huete (1988)

The geoprocessing steps of extracting reflectance data and calculating the five vegetation indices were performed using QGIS software (Free software Inc, Boston, United States).

Multi-target regression MTR

Since peanut produces above and below-ground biomass and presents different possibilities for calculating pod maturity, MTR was implemented to predict above and below-ground biomass and other types of peanut maturity indices (PMI_bb and PMI_OB). We tested two algorithms (random forest RF, Belgiu and Drăguţ, 2016; and K-Nearst Neighbour, Ali et al., 2019) to output multiple numeric values for the independent variables using VIs and spectral bands as inputs. Training (80% of the original data set) and cross-validation (20% of data) of algorithms were developed using toolkits available in the Scikit-learn python library (Pedregosa et al., 2011). Cross-validation process it's an interactive epochs-based method to prevent overfitting (Duan et al., 2014). The datasets were scaled using the StandardScaler method during the training process, and hyper-parameters were optimized using GridSearchCV methods.

Data analysis

The indices NDVI, GNDVI, NLI, MNLI, SAVI and spectral bands were used as explanatory variables to predict above and below-ground biomass, and PMI_bb and PMI_OB. The metric to analyze the performance of the algorithms was the mean absolute error MAE (Equation 2).

$$MAE = \frac{\sum_{i-1}^{n}(Yesti_i - Yobs_i)}{n}$$

where, n is the number of data, $Yest_i$ is the value of the variable estimated by algorithm, $Yobs_i$ is the value of the observed variable.

Results and Discussion

The RF and KNN algorithms outputted multiple numeric values of PMI and peanut biomass using VIs and spectral bands as explanatory variables, supporting our hypothesis that MTR can predict the maturity of peanut at the field level. The Sanken diagram (Figures 4 and 5) adequately illustrated the most accurate combination to transform the reflectance into the estimated peanut biomass and maturity. The KNN outperformed the RF algorithm in predicting peanut biomass and maturity, demonstrating that the prediction of peanut biophysical variables is possible using spectral bands and vegetation indices.

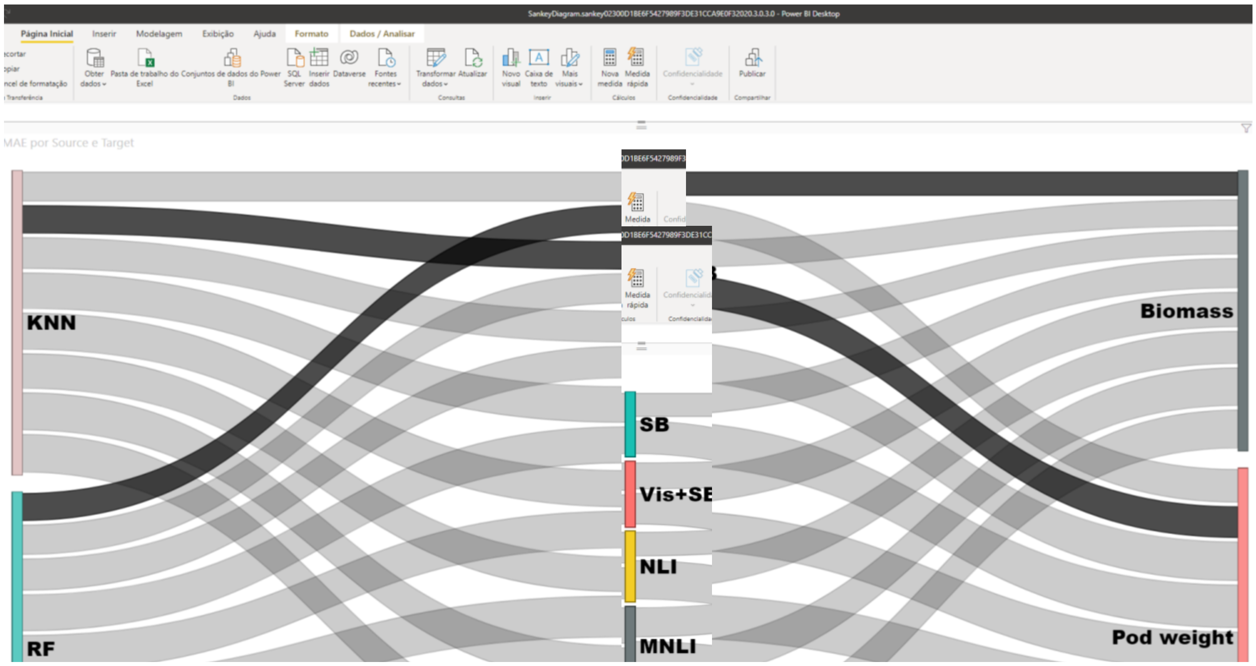

Figure 3. Sanken Diagram for the combination of KNN and RF algorithms and features to predict biomass and pod weight. Darkness paths visually assign the best solutions possible to MTR.

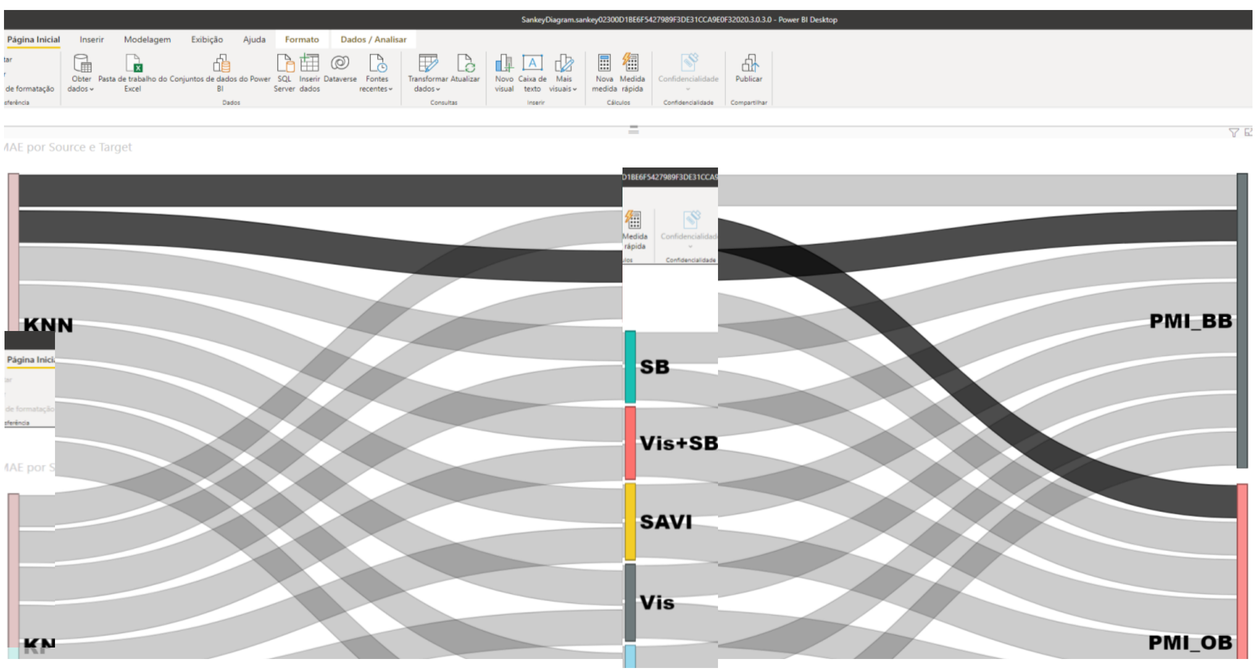

Figure 4. Sanken Diagram for the combination of KNN and RF algorithms and features to predict peanut maturity indices. Darkness paths visually assign the best solutions possible to MTR.

The use of spectral data from high-resolution satellite as explanatory variables on the MTR algorithms resulted on a predictive mean average error (MAE) of 1481 – 975 kg ha for pod weight, 779 – 1241 kg ha for AGB, 8 – 13 % for PMI_BB, and 10 – 17 % for PMI_OB. These results suggest that MTR could be used to predict both, peanut biomass and maturity at the same time. These results demonstrates the robustness of MTR algorithms in capturing spatial-temporal variability existing across crop fields. The model with the lowest MAE was the one including the KNN algorithm and used spectral bands and vegetation indices as features independently of the variable predicted.

Table 1. Mean absolute error for different input combinations and algorithms for peanut biomass and maturity prediction.

Combination	Algorithm	Pod weight	AGB	PMI_BB	PMI_OB
			Accuracy		
SB+Vis	KNN	975.416509	845.0708	0.07464	0.116308
	RF	1039.18735	892.6217	0.086901	0.120839
Vis	KNN	1291.922	951.2966	0.100696	0.154815
	RF	1480.93171	1143.046	0.111972	0.177225
SB	KNN	1119.42516	786.1444	0.077318	0.102177
	RF	1018.06169	778.8137	0.082414	0.110811
NDVI	KNN	1203.26355	1134.416	0.120178	0.169968
	RF	1257.2143	1107.873	0.121517	0.180747
GNDVI	KNN	1217.80027	1208.026	0.121032	0.175215
	RF	1247.72081	1241.112	0.122085	0.180388
NLI	KNN	1221.0988	798.8786	0.119734	0.170339
	RF	1255.31804	794.854	0.130565	0.191255
MNLI	KNN	1331.62573	971.1612	0.115103	0.167495
	RF	1388.23412	945.1607	0.127183	0.167197
SAVI	KNN	1189.25954	1209.084	0.111987	0.148624
	RF	1261.44942	1212.852	0.119148	0.162855

Graphical analysis was performed between the in-field measured variable and the predicted variable using the KNN algorithm and spectral bands and vegetation indices as input to evaluate the performance of the selected algorithm with the lowest MAE (Figure 5). We inferred from the graphs that the KNN algorithm underestimates AGB, mainly when the observed values are higher than 6000 kg ha. In contrast, the model overestimates pod weight when the observed data are less than 5000 kg ha. This pattern was not observed for peanut maturity prediction, demonstrating that MTR learned accurately from the dataset. The KNN algorithm had better accuracy in predicting PMI_BB.

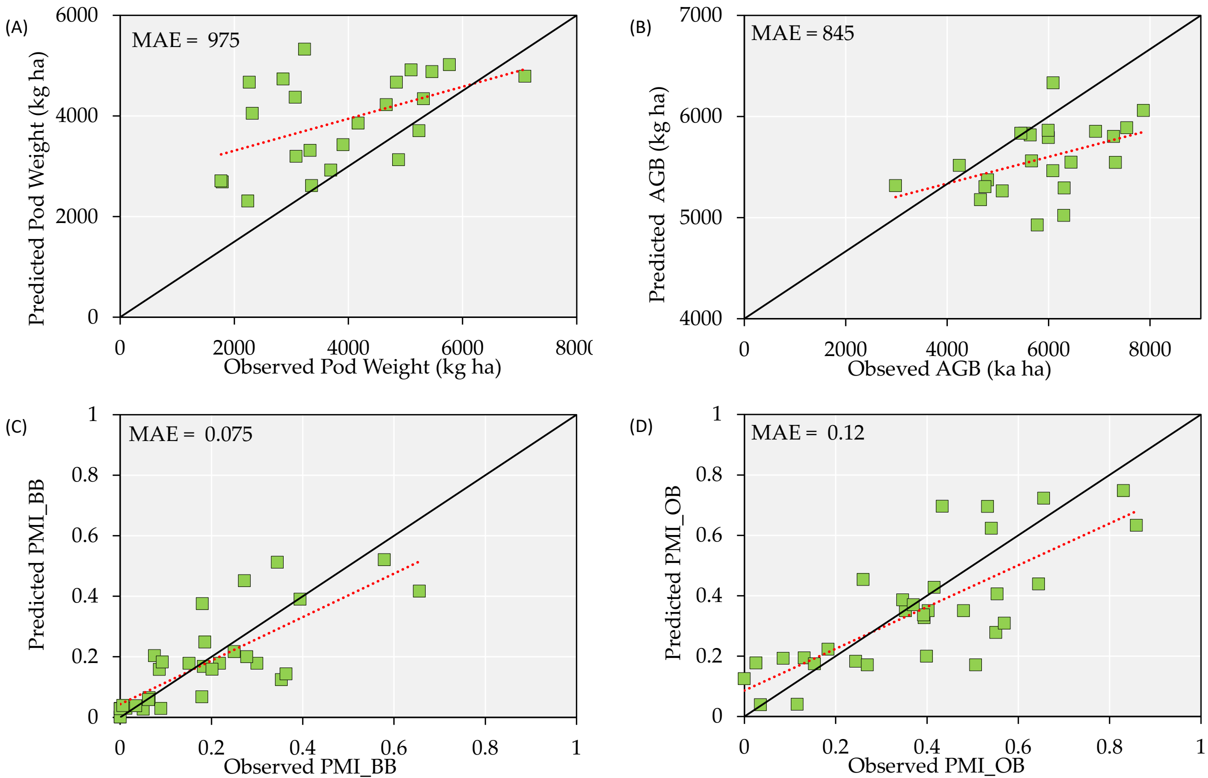

Figure 5. Performance of the best models used to predict peanut biomass and maturity using spectral bands + vegetation indices as features.

Our study is the first to mention the use of MTR in the prediction of peanut biomass and maturity upon spectral data. The RF and KNN algorithm can successfully learn the pattern between spectral bands vegetation indices and biophysical variables. Using spectral bands and vegetation indices as input, the KNN can predict the peanut variables more accurately than RF. The outperformance of KNN is attributable to the way that the algorithm works. The inputs consist of the closest training examples in the dataset. The output is the property value for the object, or equivalently, the average k-nearest neighbors' values (Tedesco et al., 2022). Contrasting, according to the authors above, RF can fail when processing data with collinearity or permutation, thus favoring features with more levels and smaller groups over more prominent groups. From a practical standpoint, producers and researchers can use MTR associated with remote sensing to determine the spatial within-field variability of peanut biomass and maturity to make agronomic decisions based on the estimated field variability. It is needed to create a robust model based on data from different varieties and crop systems to cover the main aspects of the southeastern peanut production area.

Conclusion

Our findings demonstrated a promising alternative to predict multiple PMI at a field scale using remote sensing, which may reduce the subjectivity of determining peanut maturity. Another promising outcome is that by predicting peanut biomass above ground and below ground, farmers and researchers can have quantitative values of those variables allowing characterize the peanut variability throughout the space and time. Future research should focus on integrating other explanatory variables, mainly related to topography and soil conditions like temperature. These variables could help understand the driving factors of peanut growing and maturation at the field level.

References

Abd-El Monsef, H., Smith, S. E., Rowland, D. L., & Abd El Rasol, N. (2019). Using multispectral imagery to extract a pure spectral canopy signature for predicting peanut maturity. Computers and Electronics in Agriculture, 162, 561-572.

Ali, N., Neagu, D., & Trundle, P. (2019). Evaluation of k-nearest neighbour classifier performance for heterogeneous data sets. SN Applied Sciences, 1(12), 1-15.

Belgiu, M., & Drăguţ, L. (2016). Random forest in remote sensing: A review of applications and future directions. ISPRS journal of photogrammetry and remote sensing, 114, 24-31.

Borchani, H., Varando, G., Bielza, C., & Larranaga, P. (2015). A survey on multi-output regression. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 5(5), 216-233.

Colvin, B. C., Tseng, Y. C., Tillman, B. L., Rowland, D. L., Erickson, J. E., Culbreath, A. K., & Ferrell, J. A. (2018). Consideration of peg strength and disease severity in the decision to harvest peanut in southeastern USA. Journal of crop improvement, 32(3), 287-304.

Duan, S. B., Li, Z. L., Wu, H., Tang, B. H., Ma, L., & Zhao, E. (2014). PROSAIL Model to Estimate Leaf Area Index of Maize, Potato, and Sunflower Fields from Unmanned Aerial Vehicle Hyperspectral Data. Inversion of the, 26, 12-20.

Gitelson, A. A., & Merzlyak, M. N. (1996). Signature analysis of leaf reflectance spectra: algorithm development for remote sensing of chlorophyll. Journal of plant physiology, 148(3-4), 494-500.

Goel, N. S., & Qin, W. (1994). Influences of canopy architecture on relationships between various vegetation indices and LAI and FPAR: A computer simulation. Remote Sensing Reviews, 10(4), 309-347.

Gong, P., Pu, R., Biging, G. S., & Larrieu, M. R. (2003). Estimation of forest leaf area index using vegetation indices derived from Hyperion hyperspectral data. IEEE transactions on geoscience and remote sensing, 41(6), 1355-1362.

Huete, A. R. (1988). A soil-adjusted vegetation index (SAVI). Remote sensing of environment, 25(3), 295-309.

Melki, G., Cano, A., Kecman, V., & Ventura, S. (2017). Multi-target support vector regression via correlation regressor chains. Information Sciences, 415, 53-69.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. (2019). Scikit-learn: machine learning in Python. 2011. Moon data set: https://scikit-learn. org/stable/modules/generated/sklearn. datasets. make_moons. html.

Rouse, J. W., Haas, R. H., Schell, J. A., & Deering, D. W. (1974). Monitoring vegetation systems in the great plains with ERTS proceeding. In Third Earth Reserves Technology Satellite Symposium, Greenbelt: NASA SP-351 (Vol. 30103017).

Rowland, D. L., Sorensen, R. B., Butts, C. L., Faircloth, W. H., & Sullivan, D. G. (2008). Canopy characteristics and their ability to predict peanut maturity. Peanut Science, 35(1), 43-54.

Santos, A. F., Corrêa, L. N., Lacerda, L. N., Tedesco-Oliveira, D., Pilon, C., Vellidis, G., & da Silva, R. P. (2021). High-resolution satellite image to predict peanut maturity variability in commercial fields. Precision Agriculture, 22(5), 1464-1478.

Santos, A. F., Lacerda, L. N., Gobbo, S., Tofannin, A., Silva, R. P., & Vellidis, G. (2019). Using remote sensing to map in-field variability of peanut maturity. In Precision agriculture'19 (pp. 91-101). Wageningen Academic Publishers.

Santos, A. F., Lacerda, L. N., Rossi, C., Moreno, L. D. A., Oliveira, M. F., Pilon, C., ... & Vellidis, G. (2021). Using UAV and Multispectral Images to Estimate Peanut Maturity Variability on Irrigated and Rainfed Fields Applying Linear Models and Artificial Neural Networks. Remote Sensing, 14(1), 93.

Vellidis, G., & Beasley, J. (2013). Using vegetation indices to determine peanut maturity. Report to the Georgia Agricultural Commodity Commission for Peanuts.

Williams, E. J., & Drexler, J. S. (1981). A non-destructive method for determining peanut pod maturity. Peanut Science, 8(2), 134-141.