MACHINE LEARNING APPROACH FOR CORN NITROGEN RECOMMENDATION

Data de aceite: 03/07/2023

Franciele Morlin Carneiro
Federal Technological University of Paraná (UTFPR), Santa Helena 85892-000, PR, Brazil

Armando Lopes de Brito Filho
São Paulo State University (UNESP), School of Agricultural and Veterinarian Sciences, Jaboticabal 14884-900, SP, Brasil

Murilo de Santana Martins
School of Plant, Environmental and Soil Sciences, Louisiana State University (LSU), Baton Rouge 70803, LA, USA

Ziany Neiva Brandão
Brazilian Agricultural Research Corporation (EMBRAPA Cotton), Campina Grande 58428-095, PB, Brasil

Luciano Shozo Shiratsuchi
School of Plant, Environmental and Soil Sciences, Louisiana State University (LSU), Baton Rouge 70803, LA, USA

ABSTRACT: Nitrogen (N) fertilizer recommendation tools are vital to precise agricultural management. The objectives of this research were to determine how many variables and remote sensor data are needed to prescribe N fertilizer in corn (*Zea mays* L.), PFP (partial factor productivity), and yield integrating remote sensing and soil sensor technologies. The variables of this work were NIR, Red, Red-Edge wavelengths, plant height, canopy temperature, LAI (leaf area index), and apparent *soil* electrical conductivity (ECa). Random Forest Classifier was used to select the best input to estimate N rates, PFP, and corn yield. A confusion matrix was used to identify the accuracy of the Random Forest Classifier to detect the best inputs to estimate for which input we evaluated in this work. According to Random Forest, the best inputs to estimate the N rate and PFP were Red-Edge, Red, and NIR wavelengths, plant height, and canopy temperature. For estimate corn yield were: NIR wavelengths, N rates, plant height, Red-Edge, and canopy temperature.

KEYWORDS: Active sensor, Random Forest, remote sensing, corn, yield estimate.

1 I INTRODUCTION

By the year 2050, it is estimated that agricultural production levels will have to double to meet the rising level of population growth (FOLEY et al., 2011; THE ROYAL

SOCIETY, 2016; NARVAEZ et al., 2017). This way, strategies must be created to meet sustainability demands, food security (produce food for everybody), and governance. Thus, the application of tools that support agricultural management has been gaining more and more prominence. That said, developing remote sensing technologies (e.g., sensors) is now considered one of the most effective tools for crop monitoring.

Several studies have applied remote sensing as a data acquisition tool for fast, profitable, and economically elaborating solutions in this context. Determination of crop yields is an essential information for crop field management. Measure the variability in greenness is one method to use in the field, to estimate crop yield, by using the greenest area within the field as the N non-limiting standard (HOLLAND and SCHEPERS, 2010). Considering this way, the integration of machine learning techniques (e.g., random forest (RF), artificial neural network(ANN)) are generally used for estimating crop yield out of remote sensing data as data-driven models.

The main objective of this experiment was to determine how many variables and how many remote sensor data are needed to prescribe N fertilizer in corn, PFP (partial factor productivity), and yield integrating remote sensing and soil sensor technologies.

2 I MATERIAL AND METHODS

The experiment was conducted during 2019-2021 continuous corn growing seasons at the Louisiana State University Doyle Chambers Central Research Station, Baton Rouge, LA, 30.365°N, -91.166°W. The soil type are Cancienne silt loam and Thibaut silt clay. The experimental design was a latin square with 4 replications (0, 45, 90, 180 kgNha^{-1}). Proximal sensing data were collected with a Phenom (ACS430®plus DAS43X® sensors) active crop canopy sensor of Holland Scientific®.

This sensor collects reflectance data in Red (670 nm), Red-Edge (730 nm), and NIR (near-infrared , 780 nm) wavelengths as well as automatically calculated NDVI and NDRE. Additionally, the Phenom sensor system also calculates LAI (leaf area index), and CCC (Canopy Chlorophyll Content) using empirical relationships with spectral bands(CUMMINGS, et. al., 2021), as presented on Table 1.

Data	Abbreviation	Formula	Reference
Normalized Difference Vegetation Index	NDVI	$(NIR - RED)/(NIR + RED)$	(ROUSE, 1974)
Normalized Difference Red-Edge Index	NDRE	$\frac{(NIR - (\mathrm{Re}\,d - Edge))}{(NIR + (\mathrm{Re}\,d - Edge))}$	(BARNES *et al.*, 2000)
Estimated Leaf Area Index	LAI	$(k \cdot \ln(1 - NDVI))$	(BASTIAANSSEN *et al.*, 1998)
Canopy Chlorophyll Content	CCC	$\frac{(a * NIR - (b * \mathrm{Re}\,d - Edge))}{(c * \mathrm{Re}\,d - Edge - (d * \mathrm{Re}\,d))}$	(CUMMINGS *et al.*, 2021)

Table 1: Parameters calculated by Phenom Sensor®

*Where a, b, c, d and k are scaling constants.

Biophysical characteristics as plant height, canopy temperature, were also obtained from V6 to tasselling growth stages. Apparent soil electrical conductivity was obtained with a GSSI EMP 400 Profiler®sensor using 5, 10, and 15 kHz main frequency as a proxy of soil fertility status. During several corn growth stages (from V6 to tasselling), this experiment was mapped using Profiler® and Phenom® sensors. Random forest analysis using the *R* package (caret) was performed to classify the importance of each variable plays in estimating the N rates. In addition, Table 2 details the hyperparameters used for Random Forest Classifier.

Classification model	Hyperparameters	Candidate values	Variables estimates
	ntree	300	
	mtry	8	
RFC	proximaty	True	For N rates, PFP, and yield
	importance	True	
	Type of random forest	Classification	
	Random state	0	
	Max_features	sqrt	
	N_estimators	7	
	Max_depths	6	
	Criterion	squared_error	N rates (Top 12)
	Min_samples_leaf	4	
	Min_samples_split	2	
	Verbose	0	
RFR	Bootstrap	False	
	Random state	0	
	Max_features	sqrt	
	N_estimators	9	
	Max_depths	4	
	Criterion	squared_error	N rates (Top 5)
	Min_samples_leaf	6	
	Min_samples_split	2	
	Verbose	0	
	Bootstrap	False	
	Random state	0	
	Max_features	sqrt	
	N_estimators	9	
	Max_depths	5	
	Criterion	squared_error	PFP
	Min_samples_leaf	2	
	Min_samples_split	5	
	Verbose	0	
	Bootstrap	False	

Random state	0	
Max_features	sqrt	
N_estimators	10	
Max_depths	6	
Criterion	squared_error	Yield
Min_samples_leaf	24	
Min_samples_split	2	
Verbose	0	
Bootstrap	False	

Table 2. Hyperparameters using Random Forest Classifier (RFC) and Regressor (RFR).

3 I RESULTS

3.1 Machine learning to estimate N rates, PFP, and yield

Random Forest Classifier was used to select the best input to estimate N rates (Figures 1 a and 1 b), PFP (Figure 1 c), and corn yield (Figure 1 d). The inputs used were:GSSI Profiler EMP400® (soil electromagnetic induction sensor) at 5, 10, and15 kHz frequencies, NDVI, NDRE, NIR, Red and Red-Edge wavelengths, LAI, CCC, AIR_TMP (air temperature), RH (relative humidity), CAN_TMP (canopy temperature), I_PAR and R_PAR (incident and reflected photosynthetically active radiation), PRES (pressure),CH1 (chlorophyll a), and CH2 (chlorophyll b).

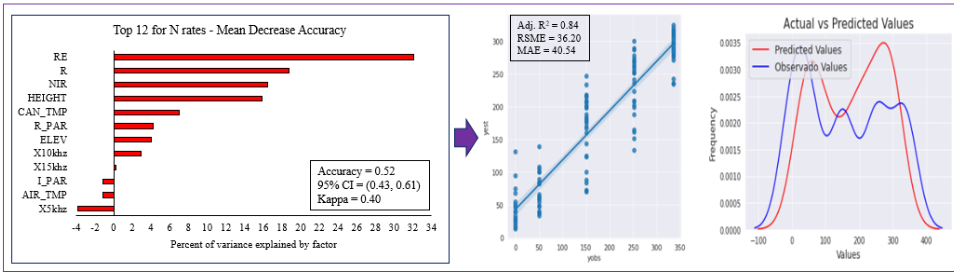
According to RFR (Random Forest Regressor), were selected the twelve (Figure 1a) and five (Figure 1b) inputs to estimate N rate. It was observed that the coefficient of determination (R^2) showed adifference of 0.15, indicating that to determine the N rate to be used, the producer do not need several inputs for your fertilizer application. In this case, it wasrequired to use just parameters as Red-edge, Red, and NIR wavelengths, plant height, and canopy temperature. In addition, we can see the accuracy from RFC the difference was very low (0.03), these results were greater for farmer because to facilitate to their to collection data and decision making.

PFP (Partial Factor Productivity)estimate, the top five inputs select for these inputs were: red-edge, red, nir, canopy temperature, plant height. For and corn yield estimate, the best five variables were: nir, N rate, plant height, red-edge, and canopy temperature.

The wavelengths got greater results than other inputs mainly red-ege, red, and nir to estimate N rate, PFP, and corn yield.

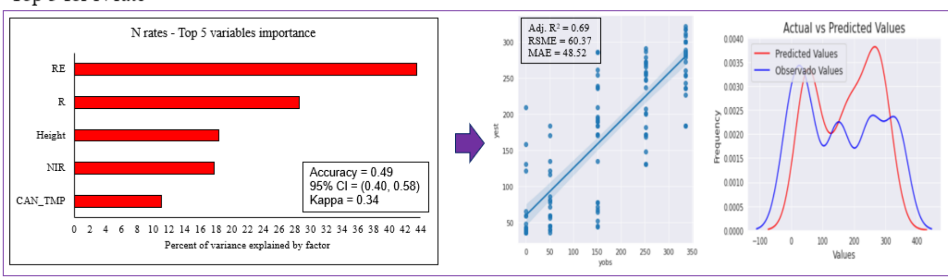
(a)

Top 12 for N rate



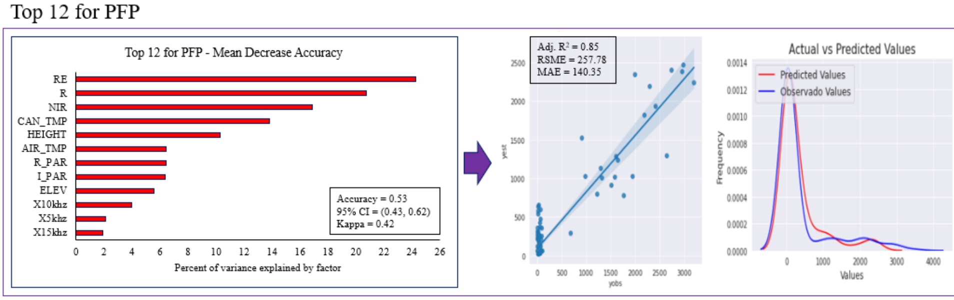
(b)

Top 5 for N rate



(c)

Top 12 for PFP



(d)

Top 13 for Yield

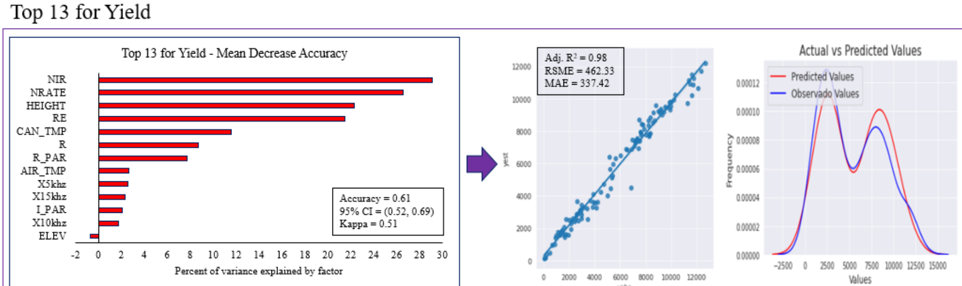


Figure 1. Random Forest Classifier to select the best inputs to estimate N rates (a anb b), PFP - partial factor productivity (c), and corn yield (d).

3.2 Random Forest Model Accuracy Validation

Confusion matrix was used to identify the accuracy to Random Forest Classifier to detect what the best inputs to estimate for which input that we evaluated in this work. For the best accuracy was yield estimate.

Predicted	Validation Data (Number)					Accuracy (%)	Overall Statistic	
	A	B	C	D	E			
Top 12 for Nitrogen rates								
A	**19**	6	3	3	1	72.52	Accuracy	0.5191
B	5	**18**	4	1	2	68.70	95% CI	(0.4301, 0.6072)
C	0	4	**8**	5	3	30.53	No Information Rate	0.2214
D	2	0	5	**11**	8	41.98	P-Value [Acc> NIR]	$9.718e^{-14}$
E	0	1	3	7	**12**	45.80	Kappa	0.3975
Top 5 for Nitrogen rates								
A	**18**	7	3	3	2	0.6870	Accuracy	0.4885
B	5	**17**	3	0	3	0.6489	95% CI	(0.4003, 0.5774)
C	0	2	**7**	7	3	0.2672	No Information Rate	0.2214
D	2	1	2	**11**	7	0.4198	P-Value [Acc> NIR]	$1.653e^{-11}$
E	1	2	8	6	**11**	0.4198	Kappa	0.3596
Top 12 for Partial Factor Productivity (PFP)								
A	**10**	0	4	4	0	38.17	Accuracy	0.5344
B	6	**13**	6	3	2	49.62	95% CI	(0.4452, 0.6219)
C	1	7	**11**	2	0	41.98	No Information Rate	0.2214
D	3	2	5	**15**	3	57.25	P-Value [Acc> NIR]	$6.227e^{-15}$
E	8	0	3	2	**21**	80.15	Kappa	0.4199
Top 13 for Yield								
A	**23**	7	1	0	0	87.79	Accuracy	0.6107
B	7	**14**	7	0	1	53.44	95% CI	(0.5216, 0.6946)
C	0	5	**12**	4	1	45.80	No Information Rate	0.2366
D	1	0	4	**16**	7	61.07	P-Value [Acc> NIR]	$< 2.2e^{-16}$
E	0	0	0	6	**15**	57.25	Kappa	0.5118

Table 2. Confusion matrix parameters from Random Forest Classifier to estimate N rates, PFP, and N rates.

3.2.1 Comparison of metric parameters among variables estimated as N rates, PFP, and yield

Yield estimate (Figure 2) had more range variable than N rates and PFP due to yield was affect many factors as harvest machine, labor, weather conditions, crop, soil conditions, area topography and other.

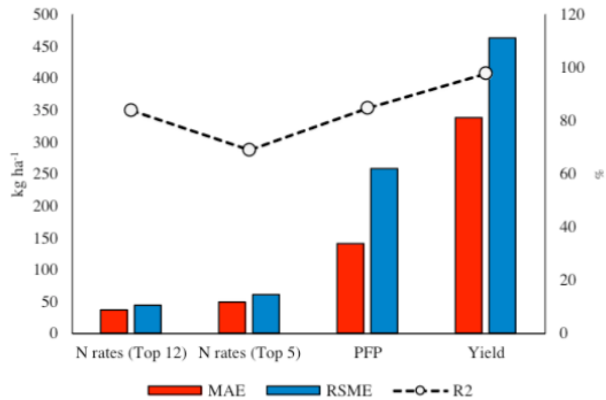


Figure 2. Metrics parameters from Random Forest Regressor using MAPE, RSME, and R^2 to estimate N rates, PFP, and yield.

4 I DISCUSSION

The main challenge nowadays is to produce food the sustainable ways. To reduce excess nitrogen application, we can use remote sensing tools to verify the variables present within the field to allow applying the right rate and place according to the crop demand. Furthermore, remote sensing is increasingly used for more sustainable production in agriculture, in addition to helping the farmer to support decision-making quickly and assertively.

PFP offered a betterway to monitor how much the farmer has increased kg grain per N applied. This information allows farmers to apply the N rate level precisely according to crop needs and consequently have a low environmental impact, reduce cost, and increase yield.

5 I CONCLUSIONS

The use of Random Forest established that the best inputs to estimate N rate and PFP were Red-Edge, Red, and NIR wavelengths, plant height, and canopy temperature. To estimate corn yield, the best inputs were: NIR wavelengths, N rates, plant height, Red-Edge, and canopy temperature.

REFERENCES

BARNES, E.M.; CLARKE, T.R.; RICHARDS, S.E.; COLAIZZI, P.D.; HABERLAND, J.; KOSTRZEWSKI, M.; WALLER, P.; CHOI, C.; RILEY, E.; THOMPSON, T.; LASCANO, R. J.; LI, H.; MORAN, M.S. **Coincident detection of crop water stress, nitrogen status and canopy density using ground-based multispectral data.** Proceedings of the 5th International Conference on Precision Agriculture. July 16-19, 2000, Bloomington, MN. USA.

BASTIAANSSEN, W.G.M., MENENTI, M., FEDDES, R.A., HOLTSLAG, A.A.M. A remote sensing surface energy balance algorithm for land (SEBAL). 1. Formulation. **Journal of Hydrology**. v. 213, p. 198–212. 1998.

CUMMINGS, C.; MIAO, Y.; PAIAO, G.D.; KANG, S.; FERNÁNDEZ, F.G. Corn Nitrogen Status Diagnosis with an Innovative Multi-Parameter Crop Circle Phenom Sensing System. Remote Sensing, v.13, 401, p. 1-18. 2021. https:// doi.org/10.3390/rs13030401

FOLEY, J. A.;RAMANKUTTY, N.; BRAUMAN, K. A;CASSIDY, E. S.;GERBER, J. S.; JOHNSTON, M.; MUELLER, N. D.; O'CONNELL, C.; RAY, D. K.; West, P. C.; BALZER, C., BENNET, E. M.; CARPENTER, S. R.; Hill, J.; MONFREDA, C.; POLASKY, S.; ROCKSTRÖM, J.; SHEEHAN, J.; SIEBERT, S.; TILMAN, D.; ZAKS, D. P. M. Solutions for a cultivated **PlanetNature***,* 478.p. 337-342,2011. https://doi.org/10.1038/nature10452

HOLLAND, K.H.; SCHEPERS, J.S. Derivation of a variable rate nitrogen application model for in-season fertilization of corn. **Agronomy Journal.** v.102, p.1415–1424. 2010.

NARVAEZ, F. Y.; REINA, G.; TORRES-TORRITI, M.; KANTOR, G.; CHEEIN, F. A. A Survey of Ranging and Imaging Techniques for Precision Agriculture Phenotyping. **IEEE/ASMETransactions on Mechatronics**, v. 22, p. 2428-2439. 2017.

ROUSE, J. W., HAAS, R. H., SCHELL, J. A., DEERING, D. W., HARLAN, J. C. Monitoring the vernal advancement of retrogradation (greenwave effect) of natural vegetation. **NASA/GSFC, Final Report,** Greenbelt, MD, 1974. 371 p.

THE ROYAL SOCIETY. **Reaping the benefits: science and the sustainable intensification of global agriculture**. Disponívelem: http://royalsociety.org/Reapingthebenefits. Accessed in: December 18th, 2019.