NITROGEN RECOMMENDATION BASED ON MACHINE LEARNING APPROACH AND ACTIVE REMOTE SENSING

Franciele Morlin Carneiro[1], Armando Lopes de Brito Filho[2], Murilo de Santana Martins[3], Ziany Neiva Brandão[4], Luciano Shozo Shiratsuchi[3]

[1] Federal Technological University of Paraná (UTFPR), Santa Helena 85892-000, PR, Brazil; fmcarneiro@utfpr.edu.br
[2] São Paulo State University (UNESP), School of Agricultural and Veterinarian Sciences, Jaboticabal 14884-900, SP, Brasil; armando.brito@unesp.br
[3] School of Plant, Environmental and Soil Sciences, Louisiana State University (LSU), Baton Rouge 70803, LA, USA; lshiratsuchi@agcenter.lsu.edu
[4] Brazilian Agricultural Research Corporation (EMBRAPA Cotton), Campina Grande 58428-095, PB, Brasil; ziany.brandao@embrapa.br

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, 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, and apparent soil electrical. 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.

Key words — active sensor, Random Forest, remote sensing, corn, yield estimate.

1. INTRODUCTION

By the year 2050, it is estimated that agricultural production levels will have to double to meet the rising level of population growth [1-3]. 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 essential information for crop field management. This way, the integrating machine learning techniques (e.g., random forest (RF), artificial neural network) 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. 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 proximately are Canciene silt loam and Thibalt silty clay. The experimental design was a latin square with 4 replications (0, 45, 90, 180 kg N ha^{-1}). Active crop canopy sensor reflectance (NIR, Red and Red Edge wavelengths), plant height, canopy temperature, LAI were obtained using a Holland Scientific sensor called Phenom (ACS430 plus DAS43X sensors). Apparent soil electrical conductivity was obtained with a GSSI EMP 400 Profiler sensor using 5, 10, and 15 kHz as the main frequency as a proxy of soil fertility status. During several growth stages of corn this experiment was mapped using profiler and phenom sensors. Random forest analysis using the R package (caret) was performed to rank the importance of each variable to estimate N rates. In addition, Table 1 details the hyperparameters used for Random Forest Classifier.

Classifi-cation model	Hyperparame-ters	Candidate values	Variables estimates
RFC	ntree mtry proximaty importance	300 8 True True	For N rates, PFP, and yield

	Type of random forest	Classification	
RFR	Random state	0	N rates (Top 12)
	Max_features	sqrt	
	N_estimators	7	
	Max_depths	6	
	Criterion	squared_error	
	Min_samples_leaf	4	
	Min_samples_split	2	
	Verbose	0	
	Bootstrap	False	
	Random state	0	N rates (Top 5)
	Max_features	sqrt	
	N_estimators	9	
	Max_depths	4	
	Criterion	squared_error	
	Min_samples_leaf	6	
	Min_samples_split	2	
	Verbose	0	
	Bootstrap	False	
	Random state	0	PFP
	Max_features	sqrt	
	N_estimators	9	
	Max_depths	5	
	Criterion	squared_error	
	Min_samples_leaf	2	
	Min_samples_split	5	
	Verbose	0	
	Bootstrap	False	
	Random state	0	Yield
	Max_features	sqrt	
	N_estimators	10	
	Max_depths	6	
	Criterion	squared_error	
	Min_samples_leaf	24	
	Min_samples_split	2	
	Verbose	0	
	Bootstrap	False	

Table 1. Hyperparamenters using Random Forest Classifier (RFC) and Regressor (RFR).

3. 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 (Figure 1), PFP (Figure 2), and corn yield (Figure 3). 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 (leaf area index), 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), we selected the twelve (Figure 1 a) and five (Figure 1 b) inputs to determine estimating N rate, we can observed that the coefficient determination (R^2) had 0.15 difference, that we can concluded that farmer do not need several inputs to determine the N rate for their fertilizer application. For this case, they just need to use 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.

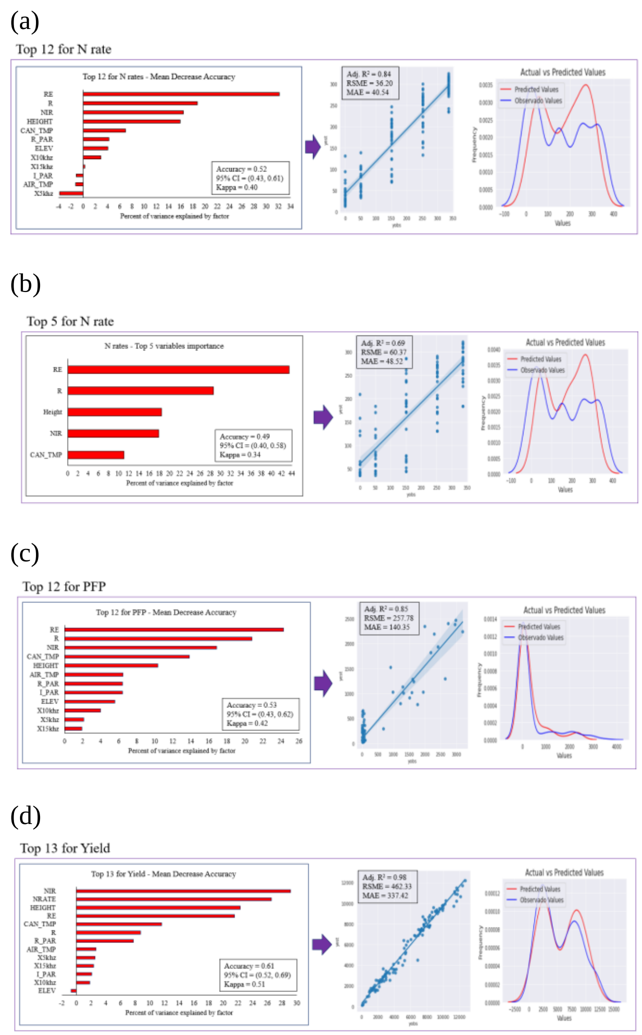

Figure 1. Random Forest Classifier to select the best inputs to estimate N rates, PFP, and corn yield.

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

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

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) A	B	C	D	E	Accuracy (%)	Overall Statistic	
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 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.

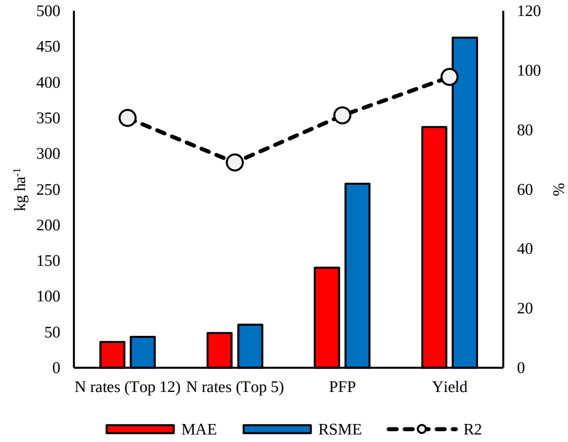

Figure 4. Metrics parameters from Random Forest Regressor using MAPE, RSME, and R^2 to estimate N rates, PFP, and yield.

4. 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 is a greater output to monitor how much the farmer has gained 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. CONCLUSIONS

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.

6. REFERENCES

[1] 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., and Zaks, D. P. M. (2011). Solutions for a cultivated Planet. *Nature* 478: 337-342. https://doi.org/10.1038/nature10452

[2] The Royal Society. *Reaping the benefits: science and the sustainable intensification of global agriculture*. Disponível

em:<http://royalsociety.org/Reapingthebenefits>. Accessed in: December 12nd, 2016.

[3] Narvaez, F. Y., Reina, G., Torres-Torriti, M., Kantor, G., Cheein, F. A. (2017). A Survey of Ranging and Imaging Techniques for Precision Agriculture Phenotyping. *IEEE/ASME Transactions on Mechatronics* 22: 2428-2439.