SHORT COMMUNICATION

MODELLING YELLOW SUGARCANE APHID USING IN SITU INFRARED SPECTROSCOPY: A PRELIMINARY STUDY

Poona N[1], Radebe T[1], Gokul A[1], Kistan C[1]

[1] *Plant Environment Resource Centre, South Africa Sugarcane Research Institute, P/Bag X02, Mount Edgecombe. 4300, South Africa*

Nitesh.Poona@sugar.org.za, Thobani.Radebe@sugar.org.za, Anushka.Gokul@sugar.org.za, Carla.Kistan@sugar.org.za

Abstract

Hyperspectral remote sensing (spectroscopy) provides opportunities for the near real-time detection and monitoring of the Yellow Sugarcane Aphid (YSA). Research has previously demonstrated the utility of spectroscopy for modelling thrips, Eldana and the Yellow Leaf virus. In this study, we initiated a glasshouse experiment to determine whether YSA-infested, YSA-damaged and healthy sugarcane could be classified by using in situ infrared spectroscopy and machine learning. The main hypothesis of this research is that the spectral responses of a healthy and a YSA-stressed sugarcane plant are statistically different and can be discriminated by using high-dimensional spectroscopic data. The experiment consisted of two varieties (N36 and N59), each with three classes: healthy (H), infested (I) and damaged (D). Each class comprised 30 samples, with a total of 180 samples. The time series leaf spectra of the third leaf from the top visible dewlap were acquired by using an ASD-FieldSpec-4 Hi-Res spectroradiometer. Classifications were undertaken by using the Random Forest (RF) and Support Vector Machine (SVM) algorithms. The preliminary results (classification accuracies ranged from 23.33% to 98.33%) indicated that healthy, infested and damaged plants can be readily classified by using in situ spectroscopy. These results provide the impetus for evaluating the methodology in field i.e. acquiring time series in situ data for modelling YSA.

Keywords: spectroscopy, sugarcane, Yellow Sugarcane Aphid (YSA), Random Forest (RF), Support Vector Machine (SVM), classification

Introduction

The Yellow Sugarcane Aphid (YSA) *Sipha flava* Forbes (Hemiptera: Aphididae) is an endemic pest of the South African sugarcane industry. It mostly attacks young sugarcane by feeding on immature leaves, which leads to premature senescence/chlorosis. A high infestation on young plants can lead to significant reductions in yield. As it is a sporadic pest, the detection and monitoring of YSA infestations is inherently difficult. The detection of YSA-induced stress, especially asymptomatic stress, will allow for proactive management prior to chlorosis (crop damage) and yield loss.

Hyperspectral remote sensing (spectroscopy) is a non-destructive technology that is increasingly being employed for modelling plant stress in forestry and agriculture. Several authors have successfully demonstrated the potential of spectroscopy for modelling pests and diseases in sugarcane. For example, Abdel-Rahman *et al*. (2009a; 2009b; 2010) employed in situ spectral data (350-2 500 nm) with RF and Partial Least Squares (PLS) regression to predict thrips (*Fulmekiola serrata* (Kobus)) populations in N19. The RF and PLS models achieved coefficient of determination (R^2) values of 0.71 to 0.75. Similarly, Mokhele *et al*. (2009) modelled Eldana (*Eldana saccharina* Walker (Lepidoptera: Pyralidae)) stalk damage in N37 by using the modified Normalised Difference Vegetation Index (mNDVI) derived from in situ spectral data (350-2 500 nm) and linear regression. The authors showed that Eldana stalk damage was linearly and negatively correlated with mNDVI (R^2=0.69). Grisham *et al*. (2010) employed laboratory spectral data (350-800 nm), as well as a discriminant analysis with re-substitution and cross-validation, to detect the Sugarcane Yellow Leaf Virus (SCYLV, genus *Polerovirus*, family *Luteoviridae*, species *sugarcane yellow leaf virus*) in LCP 85-384 and Ho 95-988 cultivars. The prediction accuracies ranged from 41%-95%.

This study represents the preliminary work that is being undertaken to test the utility of RS and ML to model YSA-induced stress in sugarcane. The main hypothesis of this research is that the spectral response of a healthy and a YSA-stressed sugarcane plant is statistically different and that it can be discriminated by using high dimensional spectroscopic data. The overarching aim is to test the utility of field spectroscopy and machine learning for developing models for the detection and monitoring of YSA in sugarcane. The ultimate goal is to inform the development of custom sensors that can be retro-fitted to a platform, such as a UAV, for near real time YSA monitoring.

Materials and Methods

Experimental design

A glasshouse experiment was set up, with two varieties, N36 and N59, representing the irrigated and rainfed regions respectively. Both varieties have an intermediate susceptibility to YSA. Three classes were established for each variety, namely: healthy (H), infested (I) and damaged (D). Each class comprised 30 sugarcane plants (i.e. samples), which resulted in a total of 180 samples. The setts were hot-water-treated and planted on 8 September 2022. The seedlings were potted and set up in BugDorm cages on 17 October 2022, with 10 plants per cage. The H class plants (n=60) were sprayed with Actara on 3 November 2022 to prevent the YSA from latching and feeding on the plants. YSA aphids (n=10) were introduced to each plant (Madiope *et al*. 2021) in the I and D classes (n=120) on 7 November 2022 (the plants were two months old) following spectral data acquisition. The plants were maintained at 26^{o}C and watered every second day, or when the soil was dry.

Spectral data acquisition and analysis

Spectral data were acquired by using an ASD FieldSpec 4 Hi-Res spectroradiometer, which acquires the data in the 350-2500 nm spectral range. A single spectrum was acquired of the third fully-expanded leaf blade of each plant by using a plant probe and leaf-clip assembly. Data were acquired daily from 7 to 11 November 2022, then weekly from 14 November to 12 December 2022, resulting in a total of 1800 spectra.

The RF and SVM algorithms were employed for building classification models. Default hyperparameters were used for both algorithms, and SVM was run by using a linear kernel. Modelling was undertaken in Python by using the RandomForestClassifier and LinearSVC implementations for RF and SVM, respectively. The performance (accuracy) of the model was computed by using a confusion matrix and cross-validation error, based on 10-fold cross-validation.

Results and Discussion

For this short communication, we focus on the results from the week of 7 to 11 November 2022.

Table 1 shows the results by using the combined data (n=180), i.e. N36 and N59. The classification accuracy, using RF, ranged from 36.11% to 86.52% (the cv error ranged from 0.64 to 0.14), whereas the SVM classification accuracy ranged from 45.00% to 92.42% (the cv error ranged from 0.55 to 0.08). The lowest accuracy was obtained for the combined classes (i.e. H-I-D) and the highest accuracy for the H-D class pair. The results compare favourably with those of Poona and Ismail (2014), who illustrated that class pairs can be more readily classified, compared with the combined classes, when classifying *Fusarium circinatum*-induced stress in *Pinus radiata* seedlings. This will most likely impact attempts to elucidate the signatures of, and to classify, multiple stressors.

Table 1. Random Forest and Support Vector Machine classification accuracy (%) using all the data for the week of 7-11 November 2022. The cross-validation (cv) error is shown in parentheses. The top-ten wavebands for the best-performing model are listed

Random Forest			
Date	**H vs I vs D**	**H vs I**	**H vs D**
07-11-2022	45.85 (0.54)	60.23 (0.40)	68.33 (0.32)
08-11-2022	43.33 (0.57)	58.33 (0.42)	61.67 (0.38)
09-11-2022	68.07 (0.32)	63.33 (0.37)	86.52 (0.14)
10-11-2022	46.67 (0.53)	57.50 (0.43)	67.50 (0.33)
11-11-2022	36.11 (0.64)	55.83 (0.44)	50.00 (0.50)
Support Vector Machine			
Date	**H vs I vs D**	**H vs I**	**H vs D**
07-11-2022	54.80 (0.45)	77.88 (0.22)	59.02 (0.41)
08-11-2022	50.56 (0.49)	70.83 (0.29)	63.33 (0.37)
09-11-2022	71.54 (0.29)	60.83 (0.39)	92.42 (0.08)
10-11-2022	70.00 (0.30)	89.17 (0.11)	78.33 (0.22)
11-11-2022	45.00 (0.55)	58.33 (0.42)	65.00 (0.35)
Waveband selection			
H vs D SVM (92.42%)	2301, 2300, 2302, 2299, 2298, 2305, 2304, 2303, 2306, 2297		

The classification results were similar for the individual varieties (Table 2), with the highest accuracies being obtained for the H-D class pair for both RF (96.67%; cv error=0.03) and SVM (98.33%; cv error=0.02), and the lowest accuracies being obtained for the combined classes, 23.33% (cv error=0.77) and 35.56% (cv error=0.64) for RF and SVM, respectively. For both the combined classes and class pairs, the best classification accuracies were obtained three days post-inoculation, i.e. on 9 November 2022. This result suggests that YSA damage could be detected within three days of it being introduced onto the sugarcane. However, this will require further investigation. Overall, the results are in line with those of Abdel-Rahman *et al*. (2009a; 2009b; 2013), Mokhele *et al.* (2009) and Grisham *et al*. (2010), which highlights the utility of spectroscopy for modelling pests and disease in sugarcane.

Table 2. Random Forest and Support Vector Machine classification accuracy (%) for the individual varieties, for the week of 7-11 November 2022. The cross-validation (cv) error is shown in parentheses. The top-ten wavebands for the best-performing models are listed

Random Forest						
Date	**H vs I vs D**		**H vs I**		**H vs D**	
	N36	**N59**	**N36**	**N59**	**N36**	**N59**
07-11-2022	50.00 (0.50)	42.64 (0.57)	80.00 (0.20)	55.33 (0.45)	73.33 (0.27)	59,67 (0.40)
08-11-2022	32.22 (0.68)	53.33 (0.47)	50.00 (0.50)	60.00 (0.40)	51.67 (0.48)	71.67 (0.28)
09-11-2022	65.56 (0.34)	70.83 (0.29)	60.00 (0.40)	70.00 (0.30)	96.67 (0.03)	79.67 (0.20)
10-11-2022	44.44 (0.56)	52.22 (0.52)	61.67 (0.38)	56.67 (0.43)	56.67 (0.43)	68.33 (0.32)
11-11-2022	23.33 (0.77)	41.11 (0.59)	46.67 (0.47)	45.00 (0.55)	48.33 (0.52)	53.33 (0.47)

Support Vector Machine						
Date	**H vs I vs D**		**H vs I**		**H vs D**	
	N36	**N59**	**N36**	**N59**	**N36**	**N59**
07-11-2022	58.89 (0.41)	34.58 (0.65)	86.67 (0.13)	38.00 (0.62)	66.67 (0.33)	54.67 (0.45)
08-11-2022	38.89 (0.61)	56.67 (0.43)	63.33 (0.37)	70.00 (0.3)	51.67 (0.48)	70.00 (0.30)
09-11-2022	64.44 (0.36)	70.00 (0.30)	51.67 (0.48)	73.33 (0.27)	98.33 (0.02)	78.00 (0.22)
10-11-2022	53.33 (0.47)	63.33 (0.37)	80.00 (0.2)	88.33 (0.12)	65.00 (0.35)	81.67 (0.18)
11-11-2022	35.56 (0.64)	43.33 (0.57)	60.00 (0.4)	55.00 (0.45)	60.00 (0.40)	63.33 (0.37)
Waveband selection						
H vs D N36 RF (96.67%)	1408, 1361, 1331, 1428, 1402, 1483, 1446, 1431, 1435, 1434					
H vs D N36 SVM (98.33%)	1000, 999, 998, 997, 996, 995, 993, 994, 992, 2306					

The waveband importance results for the top-performing RF and SVM models are provided in Tables 1 and 2 above. The top-10 wavebands (from the highest to the lowest importance) are shown for each model. For RF, the wavebands in the shortwave infrared (SWIR; 1200-2400 nm) region dominate. Despite yielding similar results, the SVM models selected wavebands in different regions, compared to RF. Important wavebands are in the near infrared (NIR; 700-1200 nm) and SWIR regions. It is worth noting the contiguity of the wavebands, which requires further investigation.

Conclusions

This study successfully demonstrated the utility of in situ spectroscopy data and machine learning for modelling YSA-induced stress in sugarcane, albeit with variable accuracy. The results obtained in this study require further investigation. The success achieved in this study provides the impetus for evaluating the methodology under field conditions. In addition, there are opportunities for investigating and elucidating the YSA spectral response under conditions of multi-stressor effects; for example, when YSA and rust are present concurrently.

Acknowledgements

The authors thank the South African Sugarcane Research Institute for supporting this research. Much appreciation is afforded to Philani Malunga and the Agrochemical staff for their assistance.

References

Abdel-Rahman E, van den Berg M, Way MJ and Ahmed FB (2009a). Hand-held spectrometry for estimating thrips (Fulmekiola serrata) incidence in sugarcane. *Proc IEEE Int Geosci Remote Sens Symp*. https://doi.org/10.1109/IGARSS.2009.5417322.

Abdel-Rahman EM, van den Berg M, Ahmed FB, Way MJ and Sewpersad C (2009b). Using spectroscopic data sets to predict numbers of thrips (*Fulmekiola serrata*) in sugarcane. *Proc S Afr Sug Technol Assoc* 82: 441-445.

Abdel-Rahman EM, Way M, Ahmed F, Ismail R and Adam E (2013). Estimation of thrips (*Fulmekiola serrata* Kobus) density in sugarcane using leaf-level hyperspectral data. *S Afr J Plant Soil* 30(2): 91-96. https://doi.org/10.1080/02571862.2013.803616.

Grisham MP, Johnson RM and Zimba PV (2010). Detecting *Sugarcane Yellow Leaf virus* infection in asymptomatic leaves with hyperspectral remote sensing and associated leaf pigment changes. *J Virol Methods* 167: 140-145. https://doi.org/10.1016/j.jviromet.2010.03.024.

Madiope KW, Keeping MG and Fourie DV (2021). Quantifying yield loss to yellow sugarcane aphid in potted sugarcane. *Proc S Afr Sug Technol Assoc* 93: 145-148.

Mokhele TA, Ahmed F and Conlong DE (2009). Detection of sugarcane African stalk borer *Eldana saccharina* Walker (Lepidoptera: Pyralidae) using hyperspectral remote sensing (spectroradiometry). *Sugar Cane Int* 27(6): 237-244.

Poona N and Ismail R (2014). Using Boruta-selected spectroscopic wavebands for the asymptomatic detection of *Fusarium circinatum* stress. *IEEE J Sel Top Appl Earth Obs Remote Sens* 7(9): 3764-3772. https://doi.org/10.1109/JSTARS.2014.2329763.