South African Journal of Plant and Soil

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tjps20

Estimation of thrips (Fumekiola serrata Kobus) density in sugarcane using leaf-level hyperspectral data

Efatih M. Abdel-Rahman a b, Mike Way c, Fethi Ahmed a, Riyad Ismail a & Elhadi Adam a
a School of Environmental Sciences, University of KwaZulu-Natal, Durban, 4041, South Africa
b Department of Agronomy, Faculty of Agriculture, University of Khartoum, Khartoum North, 13314, Sudan
c South African Sugarcane Research Institute, Private Bag X02, Mount Edgecombe, 4300, South Africa

Published online: 05 Jul 2013.

To cite this article: Efatih M. Abdel-Rahman, Mike Way, Fethi Ahmed, Riyad Ismail & Elhadi Adam (2013) Estimation of thrips (Fumekiola serrata Kobus) density in sugarcane using leaf-level hyperspectral data, South African Journal of Plant and Soil, 30:2, 91-96, DOI: 10.1080/02571862.2013.803616

To link to this article: http://dx.doi.org/10.1080/02571862.2013.803616

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the “Content”) contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at http://www.tandfonline.com/page/terms-and-conditions
Estimation of thrips (*Fulmekiola serrata* Kobus) density in sugarcane using leaf-level hyperspectral data

Elfatih M Abdel-Rahman1,2*, Mike Way3, Fethi Ahmed1, Riyad Ismail1 and Elhadi Adam1

1 School of Environmental Sciences, University of KwaZulu-Natal, Durban 4041, South Africa
2 Department of Agronomy, Faculty of Agriculture, University of Khartoum, Khartoum North 13314, Sudan
3 South African Sugarcane Research Institute, Private Bag X02, Mount Edgecombe 4300, South Africa
* Corresponding author, e-mail: elfatihabdelrahman@gmail.com

The aim of this study was to investigate the potential use of leaf-level hyperspectral data to predict the density of thrips *Fulmekiola serrata* (Kobus). A hand-held spectroradiometer was used to make the spectral measurements on spindle leaves of 4- to 5-month-old plants of sugarcane cv. N19 growing in commercial fields near Umfolozi, South Africa. A random forest algorithm followed by partial least squares regression was used for the analysis. Developed models were adequate to predict nymph numbers in December and adult numbers in March, but different models were needed for the thrips life stage assessed and the season when the estimation took place.

**Keywords:** density, *Fulmekiola serrata*, hyperspectral data, sugarcane, thrips

---

**Introduction**

Sugarcane thrips, *Fulmekiola serrata* (Kobus) (Thysanoptera: Thripidae), is a recent pest in the South African sugarcane industry. First identified in South Africa in 2004 (Way et al. 2006), it is now widespread throughout the sugarcane-growing regions of the country (Keeping et al. 2008). Thrips are small insects (2–3 mm long) that feed on the spindle leaves of sugarcane. Damage caused by this thrips includes leaf necrosis due to puncturing of the leaf surface (Way et al. 2006). Younger crops tend to be most vulnerable and appear to depend on numbers of thrips present (Keeping et al. 2008). Monitoring thrips involving leaf sampling and laboratory analysis before treatment is expensive and labour intensive. Therefore, complementary methods that can provide up-to-date information are needed.

Remote sensing offers timely data that has potential for sugarcane thrips monitoring as demonstrated by Mirik et al. (2007) for a similar pest of winter wheat. The use of hyperspectral data for this purpose seems particularly promising. Such data are characterised by light reflectance from many (typically several hundred), narrow, contiguous wavebands across the spectrum. Hyperspectral data are able to detect nuanced differences that could be related to different types of stress, pest infestations or disease incidences (Lillesand and Kiefer 2001). However, in order to analyse such large sets of spectral data for model development, one would need to collect many sample data to avoid overfitting (high variable-to-sample ratio problem). The collection of many such data is often impossible due to logistical and other constraints. Therefore, researchers seek techniques and methods that could be used to reduce the redundancy and colinearity in the hyperspectral data without losing information that is relevant to the features of interest. Random forest, a machine learning algorithm developed by Breiman (2001), is a relatively new method that has been used for such a purpose (Chan and Paelinckx 2008, Ismail 2009). The random forest regression method uses several user-defined parameters and random selection of input variables to predict a feature of interest (Breiman 2001, Maindonald and Braun 2006). The method provides information about the importance of the variables on the performance of the predictive model (Breiman 2001, Archer and Kimes 2008). This can be very useful in the selection of spectral variables when hyperspectral data are analysed.

One drawback of the random forest algorithm in selecting variables from the spectroscopic data is that the selected relevant wavebands could still be autocorrelated (Strobl et al. 2008), especially with those of very high spectral resolutions of handheld hyperspectral sensors.

Partial least squares (PLS) regression (Wood et al. 1996) overcomes the colinearity problem (Huang et al. 2004), but one issue with PLS regression in hyperspectral data analysis is the identification of the most influential spectral region(s) during the models development (Huang et al. 2004). Martin et al. (2008) recommended that the PLS regression coefficients be normalised by the average spectral reflectance at all input wavebands. Spectral regions that show high values of normalised coefficients indicate the influence of such regions on the calibrated PLS regression models.

The use of spectroscopy for monitoring insect pest damage in field crops was reviewed by Abdel-Rahman et al. (2010). For monitoring of insect infestation, Yang et al. (2005) used a multispectral handheld radiometer to predict greenbug density in wheat canopies, and found strong...
correlations ($r > 0.90$) with reflectance from bands centred at 694 and 630 nm as well as a ratio vegetation index (RVI). Mirik et al. (2006) generated a damage sensitive spectral index (DSSI) from hyperspectral data and found a high correlation ($R^2 = 0.76$) with greenbug density in wheat canopies, whereas a slightly lower correlation ($R^2 = 0.67$) was obtained when a similar index was derived from multispectral data. On the other hand, Mirik et al. (2007) investigated 20 spectral vegetation indices to explain the variance in Russian wheat aphid density and found coefficients of determination ($R^2$) ranging from 0.91 to 0.01, with a value of 0.31 for DSSI. Hence, although these and additional studies, e.g. on the sunn pest in wheat (Genc et al. 2008) and beet armyworm in cotton (Sudbrink et al. 2003), demonstrated the potential of the technique, they did not identify spectral indicators that consistently performed well for insect infestations in crops.

With regard to sugarcane thrips, Abdel-Rahman et al. (2010) found that the reflectance from severe and moderate thrips-damaged leaves was significantly lower and could be easily discriminated from healthy leaves in the visible region for sugarcane cultivars N19 and N12 in South Africa. In their study, thrips damage severity was visually categorised, but such categorisation may show some bias. Assessing the number or density of thrips infestation per leaf, plant or unit area can be a more objective measure, which could also be better suited for the development of control strategies for early detection before much damage is done. Hence, the objective of this study was to investigate the potential use of leaf-level hyperspectral data to estimate numbers of *F. serrata*.

**Methods**

**Leaf sample collection and spectral measurements**

Leaf spindles were collected from 4- to 5-month-old plants of sugarcane cv. N19 growing in commercial fields in the Umfolozi mill supply area of KwaZulu-Natal province, South Africa, in December 2007 (summer) and March 2008 (autumn). The sample crop showed no nitrogen deficiency, water stress or visual symptoms of any other pest damage or disease. Cultivar N19 was selected because it is the most common cultivar grown in the study area, representing about 39% of the total crop (South African Sugar Association 2007, unpublished industry database). Samples of five spindle leaves each were randomly collected within areas of about a 5 m radius at 2–3 locations in each field. The samples were placed in plastic bags and stored in a refrigerator (at c. 5 °C) until they were taken to the laboratory for spectral measurements.

Spectral reflectance measurements were taken from the leaf samples in the 350–2 500 nm range of the electromagnetic spectrum using ASD FieldSpec® 3 spectroradiometer. The ASD instrument is a compact, non-imaging, battery-powered portable spectrometer, which uses a fibre optic cable for light collection and a notebook computer for data logging. Its spectral resolution is 1.4 nm in the 350–1 000 nm range and 2 nm in the 1 000–2 500 nm range. Spectral measurements were carried out under room temperature (23–25 °C). A 50 W halogen lamp was used (at 40 cm distance and 45° angle from the sample) as a light source. All spectral measurements were made relative to a Spectralon white reference panel, measured before and after the measurements on sugarcane leaves. For each measurement, 20 scans were processed internally by the ASD spectroradiometer. The bottom, middle, and top upper surface of each leaf on both sides of midrib were measured, thus the spectrum of a sample (each comprising five spindle leaves) was the average of 30 spectra.

**Determination of thrips numbers**

Based on the methods described by Way (2008), insects were removed from folds of the leaf spindles by gently separating the individual leaf spindles and rinsing them, while agitating them in warm salty water. The plastic bags that were used for leaf samples preservation were also rinsed to ensure that any thrips in the bag were included in the count. Thrips were filtered through fine muslin cloth (1–2 mm) that was spread onto a laminated grid. Counts of adult thrips and nymph numbers were performed under a microscope.

**Data analysis**

The reflectance spectra were transformed to their first-order derivative to enhance the absorption features on the spectra and to locate positions of absorption features and inflection points on the spectra (Abdel-Rahman et al. 2008). The spectral portions between 1 355–1 450 nm, 1 800–1 950 nm and 2 420–2 500 nm are known water absorption features and were excluded from the analysis because of excessive noise (Abdel-Rahman et al. 2010).

The random forest regression algorithm (Breiman 1996, Archer and Kimes 2008) was used to select the most relevant wavebands from the spectroscopic data set. A PLS regression was then used to predict thrips numbers on the basis of 2.5%, 5%, 7.5%, 10%, 12.5%, 15%, 17.5%, 20%, 22.5% and 25% of the wavebands ranked as most important by the random forest algorithm.

The optimum number of spectral variables and components (factors) in each PLS regression model was determined based on root mean square errors of prediction (RMSEP), with a leave-one-out cross-validation method. For each model, the first minimum RMSEP value was used to indicate a suitable number of components rather than the absolute minimum RMSEP value to avoid overfitting (Smith et al. 2003). Separate PLS models were developed for the number of nymphs and of adult thrips. Subsequently, the influence of each selected wavelength on the predictions was evaluated as suggested by Martin et al. (2008).

The developed PLS regression models were validated using a leave-one-out cross-validation method. Models developed using the December 2007 data set were validated for the March 2008 data set to test their robustness. The analyses were performed with R software for statistical analysis (R Development Core Team 2008).

**Results**

During December (summer), thrips numbers averaged 2.64 nymphs and 7.31 adults per spindle, whereas during March (autumn) the average numbers were 0.51 nymphs and 1.19 adults per spindle.
Figure 1 shows RMSEP values of the developed PLS regression models based on the percentage of the most important wavelengths for predicting thrips density. The calculated RMSEP values suggested that 2.5% of the most important wavelengths were suitable for prediction of thrips numbers (Figure 1), but not for prediction of nymph numbers from the December sampling, for which 5% of the most important wavelengths yielded the lowest RMSEP. Figure 2 shows the mean first-order derivative of sugarcane leaf reflectances and the wavelengths selected using the random forest algorithm.

Table 1 shows the results of the PLS regression models developed. The relative influence of the selected wavelengths on the PLS regression models calibration as indicated by the normalized regression coefficients is shown in Figure 3.

The results of leave-one-out cross-validation for the PLS regression models are shown in Figure 4. These results suggested that thrips counts can be estimated during the two sampling periods, except for adults during summer for which the RMSEP values were relatively high. Results of validating models that were developed from the data set collected during summer (December 2007) with the data set collected during autumn (March 2008) as independent test data are shown in Figure 5.

Discussion and conclusions

The results of this study suggest that sugarcane thrips numbers can be predicted with leaf-level hyperspectral data during summer and autumn. Most of the wavelengths that have high influence on the PLS predictive models (Figure 3) are in the near infrared (NIR; 700–1300 nm).
Abdel-Rahman, Way, Ahmed, Ismail and Adam

and middle infrared (300–2500 nm) regions of the electromagnetic spectrum, whereas few of them are within the visible region (VIS; 400–700 nm). This contrasts with the results of Abdel-Rahman et al. (2010) who found that the VIS region was the most responsive region for discriminating between thrips-damaged leaves and healthy leaves. However, in this study, most of the leaf samples used could be in the healthy category that was employed by Abdel-Rahman et al. (2010), hence the VIS region had little influence on the models.

The negative correlation between some NIR wavelengths (1058, 1169, 1170, 1190 and 1261 nm) and thrips numbers is expected, because NIR spectral features affect plant water content as well as structure of cells in plants (Kumar et al. 2003), whereas leaf ruptures or lesions caused by thrips might cause water loss.

Wavelengths centred at 1169, 1673, 1674 and 2213 nm have negatively correlated with number of adults. These wavelengths are within ±3 nm from known nitrogen-absorption features centred at 1671 nm (Mutanga et al. 2003) and 2210 nm (Yoder and Pettigrew-Crosby 1995).

The striking differences between Figures 3a and b and the poor performance of the models developed for summer (December) observations when applied to autumn (March) conditions (Figure 5) might be due to numerous factors, such as (1) the difference in the physiological age of the crops between the two sampling periods, (2) the difference in thrips counts between the two sampling periods, and (3) differences in confounding factors (if any).
Figure 5: Observed versus predicted nymph and adult numbers derived from application of the PLS regression models developed from the December 2007 to March 2008 data set

Given that the PLS regression method was used for development of the predictive models, and very few components were selected for each model, we do not expect any overfitting problem. However, transformation of the models developed in the present study should be done with some caution as freezing of the leaf samples might have caused some changes in the leaves’ physiological and biochemical components, which could affect the leaves’ spectral characteristics.

The apparent need to use different models for estimating thrips population during summer and autumn would make the approach less suitable for monitoring purposes. Therefore, further research should look for an optimum model to work as a universal index for estimation of sugarcane thrips abundance over different seasons and different regions in different sugarcane cultivars. The use of canopy-level hyperspectral data should also be sought for sugarcane thrips infestation monitoring in order to make an informed decision about insecticide sprays or other control measures based on hyperspectral data.

Acknowledgements — The authors thank the South African Sugarcane Research Institute (SASRI) and the University of KwaZulu-Natal for funding this study. We thank Dr Maurits van den Berg (formerly with SASRI) for his valuable comments. Thanks to Mr Marius Adendorff, Cyril Cele, Nitish Poona, Innocent Shezi and Tholang Mkhole for their support with field data collection. Gratitude is extended to the R Development Core Team for their very powerful open source packages for statistical analysis.

References


Smith M-L, Martin ME, Plourde L, Ollinger V. 2003. Analysis of hyperspectral data for estimation of temperate forest canopy...


