


Machine learning prediction of coffee rust severity on leaves using spectroradiometer data

Abel Chemura^{1,2}  · Onesimo Mutanga¹ · Mbulisi Sibanda¹ · Pardon Chidoko³

Received: 13 March 2017 / Accepted: 11 September 2017 / Published online: 27 September 2017
© Sociedade Brasileira de Fitopatologia 2017

Abstract The interest in using remote sensing data in agriculture, including plant disease assessments, has increased considerably in the last years. The satellite-based Sentinel-2 MultiSpectral Imager (MSI) sensor has been launched recently for multispectral vegetation condition assessment for agricultural and ecosystem applications. The aim of this pilot study conducted in the greenhouse using a hand-held spectroradiometer was to assess the utility of the same wavebands as used in the Sentinel-2 MSI in assessing and modeling coffee leaf rust (CLR) based on the non-linear radial basis function-partial least squares regression (RBF-PLS) machine learning algorithm, compared with ordinary partial least squared regression (PLSR). The RBF-PLS derived models satisfactorily described CLR severity ($R^2=0.92$ and $RMSE=6.1\%$ with all bands and $R^2=0.78$ and $RMSE=10.2\%$ with selected bands) when compared with PLSR ($R^2 = 0.27$ and $RMSE = 18.7\%$ with all bands and $R^2 = 0.17$ and $RMSE = 19.8\%$ with selected bands). Specifically, four bands, B2 (490 nm), B4 (665 nm), B5 (705 nm) and B7 (783 nm) were identified as the most important spectral bands in assessing and modeling CLR severity. Better accuracy was obtained for most severe levels of CLR ($R^2=0.71$ using all variables) than for moderate levels

($R^2=0.38$ using all variables). Overall, the findings of this study showed that the use of RBF-PLS and the four Sentinel-2 MSI bands could enhance CLR severity estimation at the leaf level. Further work will be needed to extrapolate these findings to the crop level using the Sentinel-2 platform.

Keywords *Hemileia vastatrix* · Multispectral imaging · Coffee rust · Machine learning

Introduction

Coffee leaf rust (CLR), caused by the fungus *Hemileia vastatrix*, poses the greatest threat to the global coffee industry (Cristancho et al. 2012; Cressey 2013). Of the two commercially produced coffee types, CLR is the most severe on *Coffea arabica*, which contributes over 70% of all the coffee produced and consumed in the world (Eskes 2005; Dinesh et al. 2011). *Hemileia vastatrix* is an obligate biotrophic fungus capable of long-distance dispersal, present across major coffee producing regions (Brown and Hovmøller 2002). During the early stages of colonization, *H. vastatrix* causes small chlorotic spots from which the fungus quickly produces yellow to orange uredia that grow and coalesce (Haddad et al. 2009). The features of CLR that differentiate it from other plant diseases are that symptoms and signs appear only on the abaxial leaf side and are not usually necrotic (Belan et al. 2015). CLR can result in up to 50% loss of leaves and 70% coffee yield reduction if not properly managed (Avelino et al. 2004). These losses mainly occur through premature leaf drop, primary branch dieback and general debilitation of trees, which eventually leads to the death of coffee plants and beans of poor quality (Melo et al. 2006; Silva et al. 2006). To avoid these losses, there is an urgent need for novel techniques for predicting CLR epidemics. Monitoring agricultural crop

Section Editor: Harald Scherm

✉ Abel Chemura
achemura@gmail.com

¹ School of Agricultural, Earth & Environmental Sciences, Geography Department, University of KwaZulu-Natal, P/Bag X01, Scottsville, Pietermaritzburg 3209, South Africa

² Environmental Science & Technology Dept., Chinhoyi University of Technology, P.Bag, 7724 Chinhoyi, Zimbabwe

³ Coffee Research Institute, DR&SS, P.O.Box 61, Chipinge, Zimbabwe

health is a critical step in managing insects as well as diseases, which often result in high yield losses, poor quality beans, and economic uncertainties.

In the major coffee-producing countries, quantification of CLR incidence, its spatial extent as well as severity largely hinges on visual assessments by trained and experienced personnel (Zambolim 2016). However, this procedure is time-consuming, tedious and subjective. Consequently, there is a growing interest in using earth observation data for a timely, reliable and spatially explicit procedure for detecting crop diseases (Price et al. 1993; Sankaran et al. 2010; Mahlein et al. 2012a). Remote sensing data have great potential for continuous remote monitoring of the condition of coffee and other agricultural crops, offering accurate and up-to-date crop condition information required for improving crop protection and crop productivity (Baret and Buis 2008; Mulla 2013). In addition, and perhaps most importantly, sensors in precision agriculture can reduce costs of crop protection as disease control can be done early and in a more targeted way (Laudien et al. 2004; Barbedo 2013). This could provide useful information for decision-making on the necessity and appropriate timing of fungicide applications. Therefore, affordable, rapid and consistent methods for agricultural crop monitoring and disease forecasting are urgently needed, especially in Africa where diseases such as CLR are frequent and severe.

Remotely sensed data from earth observation platforms has been widely used in monitoring agricultural crops (Bauriegel and Herppich 2014; Yuan et al. 2014; Devadas et al. 2015; Yuan et al. 2017). It has been established that remotely sensed wavebands and vegetation indices can be used in detection and discrimination of plant diseases. For instance, Hillnhütter et al. (2011) achieved an overall accuracy of 72% in classifying nematode and *Rhizoctonia* infested and non-infested field patches with Airborne Imaging Spectroradiometer for Applications (AISA) eagle data with the spectral angle mapper algorithm (SAM). Similarly, Mahlein et al. (2012b) reported accuracies of over 90% in the identification of *Cercospora* leaf spot, powdery mildew and sugar beet rust infected leaves, concluding that proximal sensing has the potential to identify and differentiate disease incidence. It was also reported that disease-specific vegetation indices result in high accuracy (over 90%) in discriminating levels of infection of yellow mosaic disease in black gram (*Vigna mungo*) when used with multinomial logistic models (Prabhakar et al. 2013). Other studies have also demonstrated the applicability of remote sensing in disease detection, such as oil palm infection by *Ganoderma orbiforme* (Lelong et al. 2010; Shafri et al. 2011), olive fruit infested by the fruit fly *Bactrocera oleae* (Moscetti et al. 2015) and coffee infected with *H. vastatrix* (Chemura et al. 2016).

The majority of the studies on crop condition assessment have been done using narrow hyperspectral bands and vegetation indices (Thenkabail et al. 2000; Coops et al. 2003;

Zarco-Tejada et al. 2005; Huang et al. 2007; Larsolle and Muhammed 2007; Li et al. 2014; Devadas et al. 2015). However, these hyperspectral data are not easily available in coffee-producing countries, have high dimensionality and are very expensive. Newly developed sensors, such as Sentinel-2 MultiSpectral Imager (MSI) and WorldView 2 have incorporated the features of narrow bands, particularly in the red-edge and near infrared (NIR) portions of the spectrum for vegetation condition assessment. Sentinel-2 is particularly attractive because it has a very large swath-width of 290 km and has unique spectral bands designed for vegetation condition assessment. These are available at high spatial resolutions with four bands at 10 m spatial resolution and six bands at 20 m spatial resolution (Clevers and Gitelson 2013). It also has a temporal resolution of 5 days that is important in providing frequent data for monitoring. There is already evidence that the Sentinel-2 MSI is capable of estimating grassland biomass (Sibanda et al. 2016), nitrogen content (Ramoelo et al. 2015), chlorophyll (Vincini et al. 2014) and other biophysical variables (Herrmann et al. 2011; Frampton et al. 2013). All these studies showed that Sentinel-2 MSI spectral settings surpass the limits of predecessor multispectral sensors.

Remote sensing disease stress in plants is made possible due to internal leaf structure and leaf content-mediated absorption patterns of radiation (Mutanga and Skidmore 2007; Eitel et al. 2011). This ability to detect and discriminate plant diseases with remote sensing is important for many applications at different scales, ranging from understanding cellular disease dynamics to decision support in large agricultural fields. Although the ability to detect and discriminate disease infection levels is notable, not much work has been reported in modeling disease severity with remote sensing bands and vegetation indices. This is mainly because there is a complex, nonlinear relationship between remotely sensed indices and disease severity. For instance, Reynolds et al. (2012) showed that when studying sugar beets, the best performing linear model was between *Rhizoctonia* root rot and leaf water index but had an R^2 of 0.52, while all spectral bands had R^2 values less than 0.1. Given this problem, nonlinear models, such as higher-order polynomials have been used in improving accuracy of modeling disease severity with remotely sensed variables (Reynolds et al. 2012; Yu et al. 2014; Feng et al. 2016). This, however, poses challenges of transferability of the models for field application, given confounding factors such as plant age and crop variety.

Robust statistical models, such as partial least square regression (PLSR) and artificial neural networks (ANN), have been widely used to model disease severity with remote sensing variables. For example, Yuan et al. (2013) showed that models based on the PLSR performed well in modeling yellow stripe rust severity in winter wheat up to grain filling stage ($R^2 = 0.85$ and RMSE = 0.10). In another study, Zhang et al. (2012) showed that PLSR consistently performed better than

multiple linear regression in the prediction of powdery mildew severity in winter wheat ($R^2 = 0.80$ and $RMSE = 0.23$). However, Zhang et al. (2014) observed that the PLSR can lead to a serious overestimation of disease severity and result in a high level of uncertainty. This in part is because the conventional PLSR also makes a normality assumption about the distribution of the response variable, which is often not met by remotely sensed data (Ramoelo 2013).

The ability of robust algorithms to capture non-linear relationships between remotely sensed variables and crop disease severity is more complicated when they are applied on multi-spectral scanners. Given the outstanding performance of Sentinel-2 MSI band settings, we hypothesize that its application in disease severity modeling will produce favorable results when compared with other broadband multispectral scanners. This is important as it reduces the challenges related to cost and dimensionality associated with hyperspectral data. The aim of this pilot study was therefore to (i) assess CLR severity levels with the same wavebands used in the Sentinel-2 platform but using a hand-held scanner; (ii) evaluate the robustness of ordinary PLSR and non-linear radial basis function-partial least squares regression (RBF-PLS) algorithms in modeling CLR severity based on these data; and (iii) determine if modeling accuracy was affected by disease severity levels.

Materials and methods

Coffee leaf rust inoculation

The study was carried out using greenhouse potted plants at the Coffee Research Institute, Chipinge, Zimbabwe (32°37.523'E, 20°12.474'S). All nursery plant management activities were done according to the Coffee Handbook (Logan and Biscoe 1987). The seedlings were left to grow under a nursery shade (70%) until they were 6 months old (May 2014) and then transferred to the greenhouse where pathogen inoculations were done after acclimatization to the greenhouse conditions. When the coffee plants were 8 months old (with an average of 12 true leaves) inoculation with the fungus *H. vastatrix* was done. The inoculation procedure followed Chemura et al. (2016). A total of 63 leaf samples were used. These were further classified into asymptomatic (no symptoms or signs), moderate (1–10% leaf area covered) and severe (>10% leaf area covered). Diseased area was measured using a graduated transparent polythene sheet and converted to proportion of infected area (%) by dividing over leaf area. The distribution of the leaf area of the leaf samples is shown in Fig. 1a and the distribution of area diseased for the asymptomatic, moderate and severe leaf samples are shown in Fig. 1b.

Reflectance measurements and resampling

Reflectance was measured using a hand-held Apogee VIS-NIR spectroradiometer (Apogee Instruments, Inc.) with an effective spectral range of 400–900 nm and a spectral resolution of 0.5 nm. Three scans were done 15 cm above the coffee leaf of interest at 30° field of view. To determine reflectance, a white polytetrafluoroethylene (PTFE) reflectance standard was used as a reference. Reflectance was therefore determined as the ratio of scene reflectance to the reflectance of the standard reference. To smoothen the spectra, a Savitzky–Golay filter with a frame-size of 3 data points and a 2nd order polynomial was used on the data (Savitzky and Golay 1964). The reflectance was further averaged to 5 nm to reduce dimensionality.

The smoothened and averaged reflectance measurements were resampled in ENVI 4.7 software (Exelis Visual Information Solutions) to simulate the Sentinel-2 MSI satellite sensor's reflectance (Table 1). The Full Width at Half-Maximum (FWHM) method was used to resample the spectra. The technique uses the field spectral data from the spectroradiometer and resamples it to the spectral width of the sensor being simulated. The data were resampled to seven Sentinel-2 MSI land management bands. This was because the other Sentinel-2 MSI bands were considered unnecessary for plant condition assessment. Furthermore, these omitted bands had insufficient spatial resolution for application in coffee or were outside the range of the spectroradiometer.

CLR severity modeling approaches

Two modeling approaches, PLSR and RBF-PLS regression were used to model CLR severity levels using reflectance data. PLSR uses the response variable information to perform decomposition on the spectral variables by finding the multidimensional direction in the space of predictor variables that explains the maximum variance in the response variable (Atzberger et al. 2010). PLSR is commonly used when the predictor matrix is poorly conditioned, typical of reflectance of stressed plants (Darvishzadeh et al. 2011). The PLSR algorithm has been widely used and is well described in literature (Martens and Naes 2001; Wold et al. 2001; Hansen and Schjoerring 2003). PLSR was implemented using the mixOmics library (Cao et al. 2015) in R (R Core Team 2013). The RBF-PLS is part of the kernel learning family of algorithms. These learning methods are based on mapping the originally observed data into a high-dimensional feature space where simple linear models are then constructed (Rosipal 2010). This transformation usually results in better accuracy than untransformed data, explaining why this method was chosen for this study. The concept of the RBF-PLS is described in detail in the literature (Orr 1996; Yan et al. 2004; Jia et al. 2010; Jiang et al. 2013). RBF-PLS was implemented in Matlab (MathWorks) using the TOMCAT toolbox (Daszykowski et al. 2007).

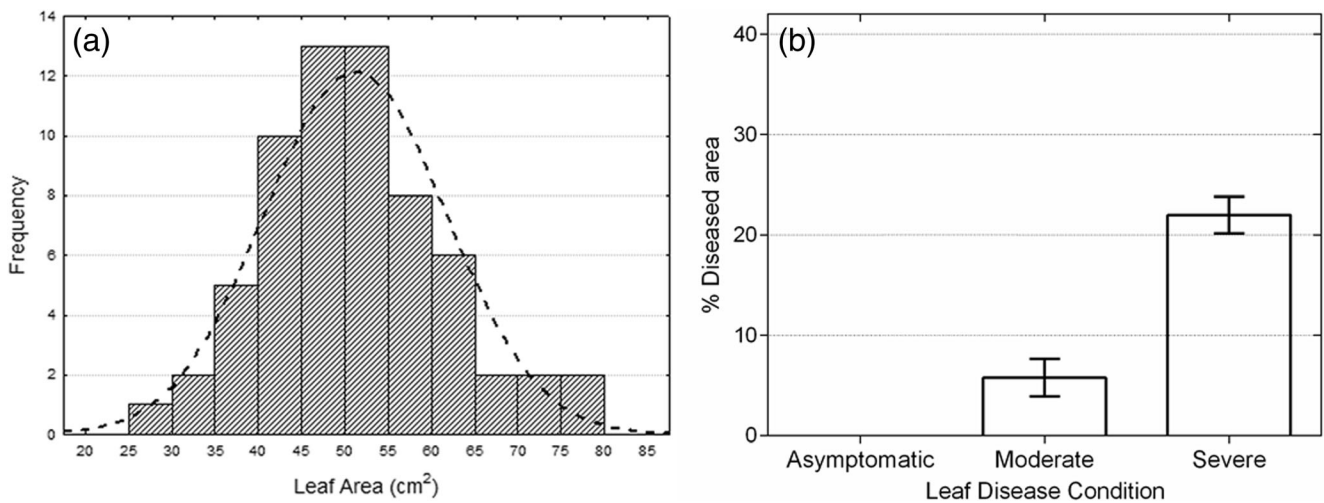


Fig. 1 **a** Histogram showing distribution of leaf area of coffee leaf samples and **(b)** means and standard errors of percent diseased area with coffee leaf rust (asymptomatic = 0%, moderate = 1–10%, severe \geq 10%) as measured on the day of reflectance measurements ($n = 63$)

Determining contribution of variables

Variable importance in projection (VIP) scores was used to weight the importance of the variables used in CLR severity modeling. The VIP is a quantitative estimate of the importance of each individual band used in the model. The VIP is a weighted sum of squares of the PLS loadings that considers the amount of explained Y-variance of each component. The weights are a function of the reduction of the sums of squares across the number of PLS components. A variable with a VIP Score greater than 1 is considered important in the model while those with VIP scores less than 1 are less important (de Almeida et al. 2013).

Accuracy assessment

In order to assess the performance of CLR severity predictions, k-fold cross-validation with 100 folds was used since the sample number was relatively small ($n = 63$) for sub-

setting the data into training and test data. The correlation coefficient (r) and coefficient of determination (R^2) were used to assess the goodness of fit of the predicted and observed CLR severity values. In addition, mean absolute error (MAE, Eq. 1), RMSE (Eq. 2) and percent bias (pBias, Eq. 3) was used to determine the errors of the model in predicting CLR severity from variables.

$$MAE = \left(\frac{1}{n} \sum |y_i - \hat{y}_i| \right) \quad (1)$$

$$RMSE = \sqrt{\frac{1}{n} \sum (y_i - \hat{y}_i)^2} \quad (2)$$

$$pBias = \left(\frac{\sum (y_i - \hat{y}_i) * 100}{\sum y_i} \right) \quad (3)$$

In the above cases, n is the number of data points, y_i is the measured CLR severity at that data point and \hat{y}_i is the model-predicted CLR severity at that data point (Moriassi et al. 2007; DeJonge et al. 2016).

Results

Linear relationship between measured wavebands and CLR severity

The results showed that there were weak correlations between CLR severity and the seven measured wavebands (Table 2). Only three out of the seven bands were significantly ($P < 0.05$) correlated with CLR severity (Blue, Red and RE1). In both cases, the correlations were positive. Results show very poor

Table 1 Specifications of the hand-held spectroradiometer band settings used to assess coffee leaf rust severity on individual leaves, showing center wavelengths, bandwidth and spatial resolution^a

Spectral band	Center wavelength (nm)	Band width (nm)	Spatial resolution ^a (m)
B2 (blue)	490	65	10
B3 (green)	560	35	10
B4 (red)	665	30	10
B5 (RE1)	705	15	20
B6 (RE2)	740	15	20
B7 (RE3)	783	20	20
B8 (NIR)	842	115	10

^a Spectral bands selected to coincide with the Sentinel-2 MSI remote sensing platform, with the theoretical field-level spatial resolution indicated for each band

Table 2 Correlation coefficients (above the diagonal) and significance of correlation (below the diagonal) between hand-held spectroradiometer wave bands (B2 to B8) and coffee leaf rust severity on individual leaves

	Severity (%)	Leaf area	B2	B3	B4	B5	B6	B7	B8
Severity (%)	1	-0.127	0.264	-0.102	0.331	0.254	-0.049	-0.101	-0.106
Leaf Area	0.320	1	0.026	0.118	0.047	0.116	0.145	0.126	0.096
B2	0.036	0.839	1	0.835	0.954	0.749	0.550	0.468	0.517
B3	0.426	0.355	<0.001	1	0.755	0.858	0.760	0.632	0.616
B4	0.008	0.711	<0.001	<0.001	1	0.751	0.497	0.422	0.473
B5	0.044	0.363	<0.001	<0.001	<0.001	1	0.607	0.453	0.474
B6	0.697	0.256	<0.001	<0.001	<0.001	<0.001	1	0.974	0.931
B7	0.432	0.323	<0.001	<0.001	<0.001	<0.001	<0.001	1	0.973
B8	0.409	0.454	<0.001	<0.001	<0.001	<0.001	<0.001	0 < 0.001	1

*Bold numbers indicate significant correlations ($p < 0.05$)

linear relationships with CLR severity ($r < 0.5$). It is therefore difficult to use linear modeling to relate CLR severity with the measured spectral variables. Leaf size did not have a significant influence on reflectance as leaf area was not significantly correlated ($P > 0.05$) with any of the bands. Interestingly, all spectral bands were significantly correlated with each other, presenting potential challenges of multicollinearity in the model if linear methods are applied (Table 2).

Modeling CLR severity with ordinary PLSR

Four factors were identified as important when using all spectral bands (Fig. 2a) while only three were identified when selected variables (Fig. 2b) were used in modeling CLR severity using ordinary PLSR. The VIP scores confirmed that the three CLR severity-correlated bands were important in the PLSR models together with an uncorrelated RE3 (Fig. 3). The NIR, green and Red-

edge 2 did not significantly contribute to CLR severity modeling with ordinary PLSR.

Results showed that using all variables yielded better results than the use of a few model-selected variables (Fig. 4). However, CLR severity prediction accuracy was low for all variables ($R^2 = 0.26$, RMSE = 18.8, Fig. 5a) and for selected variables ($R^2 = 0.17$, RMSE = 19.7, Fig. 5b). Using all spectral bands managed to explain only 26% of the variance in CLR severity. This value further decreased when selected variables were used (17% of variability explained). Covariance analysis showed that there were no significant ($P = 0.283$) differences in the coefficients of the relationship between CLR severity and modeled CLR severity with all variables and with selected variables. In both cases, larger values (those exceeding 45% observed diseased area) were poorly predicted with the ordinary PLSR analysis. However, there was evidence that the performance was slightly higher for severe than for moderate levels of CLR (Table 3).

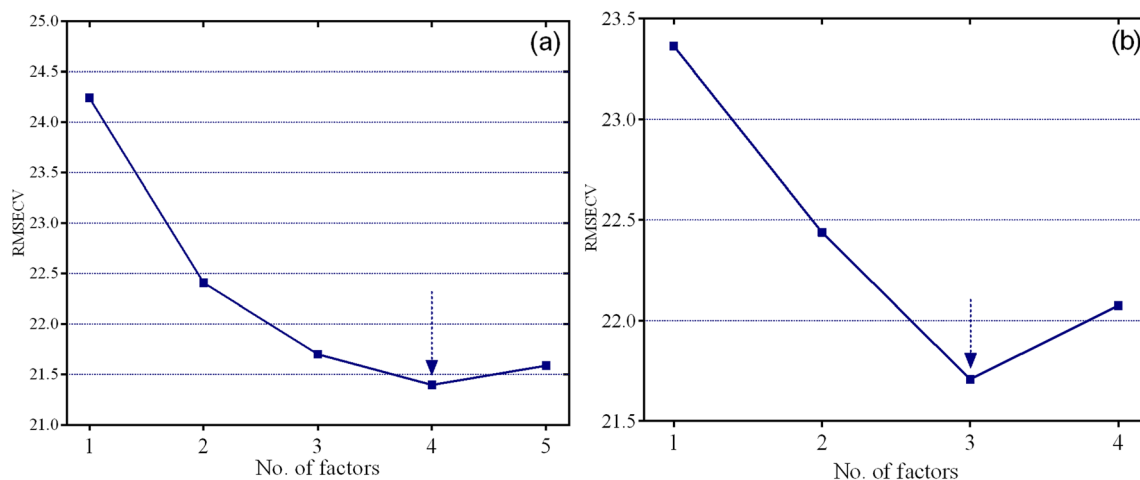


Fig. 2 Selection of number of factors for use in ordinary partial least squares regression with (a) all spectroradiometer bands and (b) selected bands. The arrows show the cut-off points for the best number of factors used in selection of variables. RMSE = root mean square error

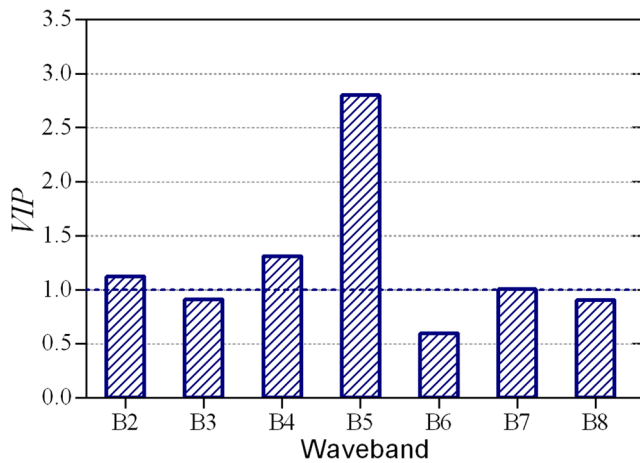


Fig. 3 Variance in projection of spectroradiometer bands in variable weighting of modeling coffee leaf rust severity with partial least squares regression models. VIP = variable importance in projection

Modeling CLR severity with non-linear RBF-PLS

Determining Gaussian widths

Figure 5 shows the relationship between cross-validated RMSE and σ representing Gaussian widths. The results show that the best Gaussian width for the RBF-PLS model is 0.3 as it produces the least cross-validated training RMSE of 6.7 for all variables (Fig. 5a). The model therefore used this as the σ for the model. For selected variables, however, the best Gaussian width was lower at 0.2 but with a minimum cross-validated RMSE of 6.8 (Fig. 5b). The learning process thus produced different results for developing the model with all variables vs. selected variables.

RBF-PLS model performance

Compared with ordinary PLSR, there was a significant improvement in CLR severity modeling accuracy with RBF-PLS regression. Using all bands as variables explained 92% of the variance in CLR severity using RBF-PLS (Fig. 6a, Table 3). On the other hand, the use of model-selected variables reduced the accuracy of model ($R^2 = 0.78$, RMSE = 10.2, Fig. 6b, Table 4). A comparison shows that using all bands outperforms the use of a few model-selected variables in modeling CLR ($P < 0.05$). Although the results show that both models are good, there is a general indication that moderate levels of infection are more difficult to predict than severe levels using all bands and only significantly correlated bands ($R^2 = 0.71$ for severe and $R^2 = 0.38$ for moderate bands, Table 4).

Discussion

Given the importance of coffee leaf rust worldwide, improved detection and monitoring tools to safeguard investments and to increase productivity are needed. This study sought to model CLR severity levels with spectral wavebands similar to those used by the Sentinel-2 MSI, and to evaluate the robustness of ordinary PLS regression and RBF-PLS regression algorithms in estimating CLR severity at different disease severity levels.

Using only spectral bands together with the RBF-PLS showed a high accuracy ($R^2 = 0.92$) in predicting CLR, which is an encouraging achievement. This is so because much of the reported favorable results in vegetation condition modeling

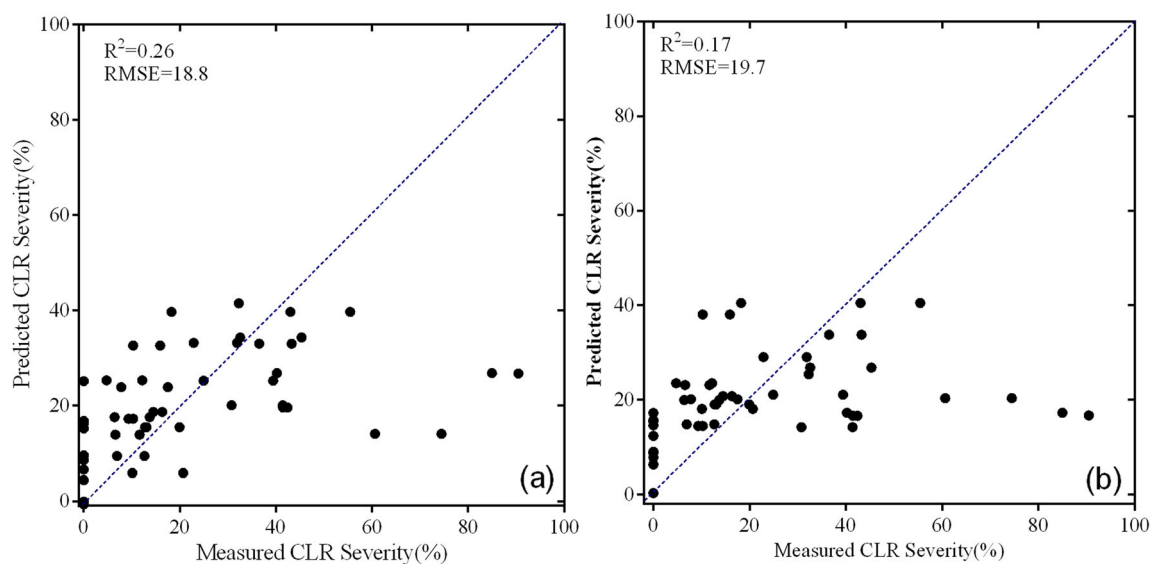


Fig. 4 Relationship between observed and modeled coffee leaf rust (CLR) severity with ordinary partial least squares regression with (a) all spectroradiometer bands and (b) selected bands to determine model performance

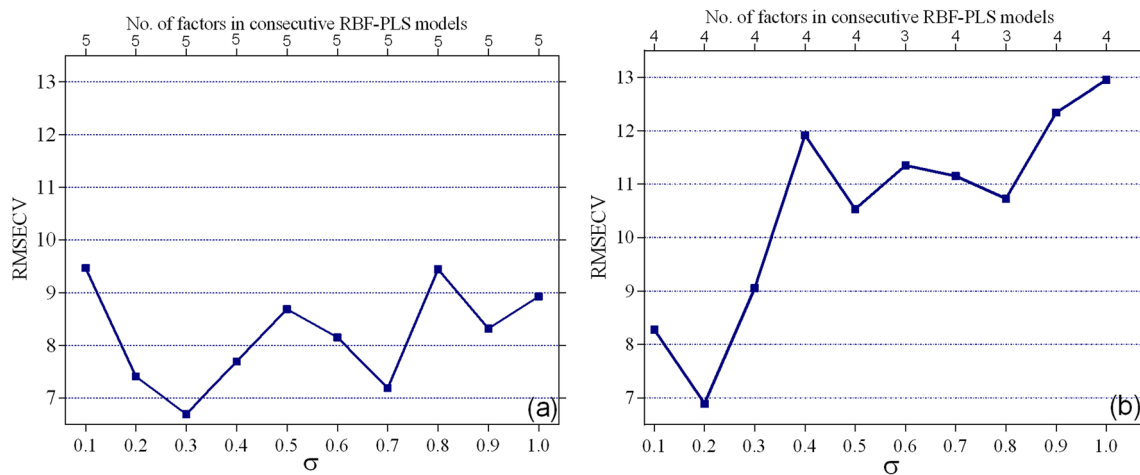


Fig. 5 Determination of σ through cross-validated root mean square error (RMSE) for use as Gaussian widths in the radial basis function-partial least squares models for (a) all spectroradiometer bands and (b) selected bands

have been from vegetation indices or at least a combination of vegetation indices and spectral bands, even with hyperspectral data. Vegetation indices use ratio or derivative transformation of spectral bands that yield more information than raw bands (Baret and Guyot 1991; Kanke et al. 2016). However, most of the vegetation indices use only two spectral bands, neglecting often small but additive values of other bands. It was clear in this study that using just selected bands had lower accuracy because the omitted bands contribute to model performance, resulting in higher accuracies where all bands are used.

Many vegetation indices are derivatives of the NIR or the red-edge bands. Entering more than one index in a model results in serious over-fitting, which eliminates transferability of the model. For example, the normalized difference vegetation index, simple ratio, renormalized normalized difference vegetation index and simplified canopy chlorophyll index all use B8 and B4 bands. Thus, developing a predictor with all these vegetation indices will likely produce an unstable model. Therefore, having a band-based model producing this level of accuracy is important, as each individual variable may be

unique in their contribution to the outcome. Ramoelo et al. (2015) also used only spectral bands to model nitrogen in rangelands and obtained a high accuracy ($R^2 = 0.90$, $RMSE = 0.04$), confirming that the spectral settings used here are well suited for vegetation condition assessments. Notwithstanding the contribution of each band to overall accuracy, the VIP scores confirmed the spectral bands that are most important in modeling CLR severity. These similar bands (except for B2) were identified as important in CLR discrimination previously using different modeling approaches (Chemura et al. 2016). This confirms that these bands can be used to perform both discrimination and modeling of CLR severity with the wavebands used here.

The finding that CLR severity was not significantly correlated with most spectral bands and weakly so for bands where results were significant was not surprising. This could be because of the non-linear influence of infection on spectral reflectance. The influence of the disease severity level on the leaf structure that influences the reflectance may affect specific wavelengths of the spectrum (Zhang et al. 2012; Mahlein et al. 2013). Thus, when reflectance is averaged across a spectral width, this direct influence is lost, resulting in non-linear relationships. Even for the specific narrow-band indices obtained from hyperspectral indices, linear relationships are also not always obvious because of other confounding factors that may influence reflectance. For example CLR infection has a significant influence on nutrient distribution within the leaf (Belan et al. 2015). The effect of the disease on nutrient distribution will then result in nutritional composition having direct influences on reflectance. Other studies have also shown that disease and pest incidence have significant influences on leaf water content (Mutanga and Ismail 2010; Oumar and Mutanga 2014), which in turn influences water absorption features. The influence of CLR severity on water content may

Table 3 Error metrics^a for all models in predicting coffee leaf rust (CLR) severity on individual leaves from hand-held spectroradiometer variables with ordinary partial least squares regression (PLSR)

Method	Disease severity	MAE	RMSE	pBias	R ²
PLSR: All	Moderate CLR	8.38	10.47	48.4	0.01
	Severe CLR	20.05	27.89	-35.6	0.12
	All levels	13.1	18.77	-	0.26
PLSR: Selected	Moderate CLR	9.38	11.65	75.0	0.00
	Severe CLR	22.25	30.1	-44.0	0.10
	All levels	14.04	19.8	-	0.17

^a MAE Mean absolute error, RMSE Root mean square error, pBias Percent bias

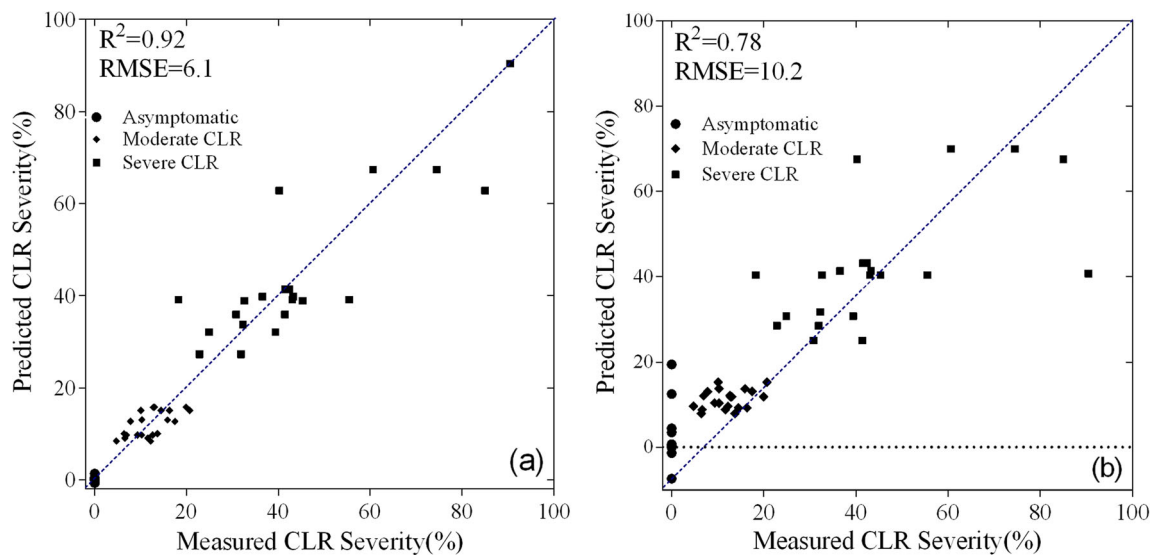


Fig. 6 Relationship between observed and modeled coffee leaf rust (CLR) severity with radial basis function-partial least squares regression with (a) all spectroradiometer wave bands and (b) selected wavebands used to determine model performance

explain the significance of the blue band in this study. More research is required to determine the exact relationship between CLR severity and leaf water content, as this relationship is disease-specific. The role of the red and red-edge bands found important in this study in stress detection has been reported widely (Rumpf et al. 2010; Eitel et al. 2011).

Results showed that the correlation between modeled and observed disease severity values was higher for severe levels of CLR compared with moderate levels when RBF-PLS regression was used. These observations may be explained by the fact that on more serious levels of infection, there is little spectral confusion as effects will be distinct compared with moderate levels (Chemura et al. 2016). However, from a practical application viewpoint, it is more useful to be able to assess CLR at moderate levels because control measures can be implemented more successfully (Carter and Miller 1994; Rumpf et al. 2010; Zambolim 2016).

Results showed that nonlinear RBF-PLS regression resulted in higher accuracy than ordinary PLSR for all spectral

bands and also specifically selected bands. The nonlinear RBF-PLS, while maintaining advantages of PLSR over linear models, has the added advantage of the kernel learning neural network. Ordinary PLSR maximizes covariance between data sets, while minimizing the variance of the prediction, which reduces the dimensionality of the data through decomposition of the independent variables into uncorrelated latent variables (Höskuldsson 1988; Wold et al. 2001). The inclusion of the neural network through the RBF kernel further improves on ordinary PLSR by making it a nonparametric model. Nonparametric models are advantageous in terms of application because they are not restricted by the nature of the statistical distribution of the data (Orr 1996; Martens and Naes 2001; Rosipal 2010). Thus, nonparametric RBF-PLS regression may be better in dealing with potential problems of model over-fitting associated with collinear variables and may result in more stable and robust models with better accuracy and transferability. There is thus a need for broader application of methods like RBF-PLS in remote sensing-based biophysical and biochemical modeling.

Although our results are encouraging in indicating potential application of sensor-based disease assessment and modeling, factors such as canopy structure and distribution of the disease across the canopy need to be considered for practical application at the field scale. There is need for more field studies to apply RBF-PLS in modeling CLR severity and other biophysical and biochemical variables. Leaf-level CLR modeling may not translate to field applications because of the spatial resolution associated with field-level remote sensing platforms that may limit measured data, especially considering that the Sentinel-2 red-edge bands are at lower resolutions compared with VIS/NIR. Similarly, there is a need for more detailed studies of the general spread and distribution of

Table 4 Error metrics^a for all models in predicting coffee leaf rust (CLR) severity from hand-held spectroradiometer variables with radial basis function-partial least squares (RBF-PLS) regression

Method	Disease severity	MAE	RMSE	pBias	R ²
RBF-PLS: All	Moderate CLR	3.05	3.37	-0.1	0.38
	Severe CLR	7.43	10.06	0.0	0.71
	All levels	3.58	6.13	-	0.92
RBF-PLS: Selected	Moderate CLR	3.56	4.2	-6.8	0.11
	Severe CLR	10.3	15.31	-4.9	0.37
	All levels	6.18	10.18	-	0.78

^aMAE Mean absolute error, RMSE Root mean square error, pBias Percent bias

CLR in the coffee plant canopy that would allow converting leaf-level assessments to canopy-level assessments. Notwithstanding these limitations, the findings of this study underscore the potential importance of sensors in monitoring plant disease epidemics, which is critical in minimizing economic costs. The findings of this study are an important first step towards the development and application of hand-held, airborne or satellite sensors for predicting CLR severity levels in the field.

Acknowledgements We are very grateful to the Coffee Research Institute for providing facilities and staff to support this research. This research was also partly funded by IFS grant D/5441. We are also thankful to anonymous reviewers whose comments improved this paper.

References

- Atzberger C, Guerif M, Baret F, Werner W (2010) Comparative assessment of three chemometric techniques for the spectroradiometric assessment of canopy chlorophyll content in winter wheat. *Computers and Electronics in Agriculture* 73:165–173
- Avelino J, Willcoquet L, Savary S (2004) Effects of crop management patterns on coffee rust epidemics. *Plant Pathology* 53:541–547
- Barbedo JGA (2013) Digital image processing techniques for detecting, quantifying and classifying plant diseases. *Springer Plus* 2(660):1–12
- Baret F, Buis S (2008) Estimating canopy characteristics from remote sensing observations: review of methods and associated problems advances in land remote sensing. Springer, pp. 173–201
- Baret F, Guyot G (1991) Potentials and limits of vegetation indices for LAI and APAR assessment. *Remote Sensing of Environment* 35:161–173
- Bauriegel E, Herppich WB (2014) Hyperspectral and chlorophyll fluorescence imaging for early detection of plant diseases, with special reference to *Fusarium* spec. infections on wheat. *Agriculture* 4:32–57
- Belan LL, Pozza EA, de Oliveira Freitas ML, Pozza AAA, de Abreu MS, Alves E (2015) Nutrients distribution in diseased coffee leaf tissue. *Australasian Plant Pathology* 44:105–111
- Brown JKM, Hovmöller MS (2002) Aerial dispersal of pathogens on the global and continental scales and its impact on plant disease. *Science* 297:537–541
- Cao K.-AL, Gonzalez I, Dejean S (2015) Package ‘mixOmics’, University of Queensland, Australia
- Carter GA, Miller RL (1994) Early detection of plant stress by digital imaging within narrow stress-sensitive wavebands. *Remote Sensing of Environment* 50:295–302
- Chemura A, Mutanga O, Dube T (2016) Separability of coffee leaf rust infection levels with machine learning methods at Sentinel-2 MSI spectral resolutions. *Precision Agriculture* 18:859–881
- Clevers JGPW, Gitelson AA (2013) Remote estimation of crop and grass chlorophyll and nitrogen content using red-edge bands on Sentinel-2 and -3. *International Journal of Applied Earth Observation and Geoinformation* 23:344–351
- Coops N, Stanford M, Old K, Dudzinski M, Culvenor D, Stone C (2003) Assessment of Dothistroma needle blight of *Pinus radiata* using airborne hyperspectral imagery. *Phytopathology* 93:1524–1532
- Core Team R (2013) R: a language and environment for statistical computing. Austria, Vienna Retrieved from <http://www.R-project.org/>
- Cressey D (2013) Coffee rust regains foothold. *Nature* 493:587
- Cristancho M, Rozo Y, Escobar C, Rivillas C, Gaitán A (2012) Outbreak of coffee leaf rust (*Hemileia vastatrix*) in Colombia. *New Disease Reports*, 25 <https://doi.org/10.5197/j.2044-0588.2012.025.019>
- Darvishzadeh R, Atzberger C, Skidmore A, Schlerf M (2011) Mapping grassland leaf area index with airborne hyperspectral imagery: a comparison study of statistical approaches and inversion of radiative transfer models. *SPRS Journal of Photogrammetry and Remote Sensing* 66:894–906
- Daszykowski M, Serneels S, Kaczmarek K, Van Espen P, Croux C, Walczak B (2007) TOMCAT: a MATLAB toolbox for multivariate calibration techniques. *Chemometrics and Intelligent Laboratory Systems* 85:269–277
- de Almeida MR, Correa DN, Rocha WF, Scafi FJ, Poppi RJ (2013) Discrimination between authentic and counterfeit banknotes using Raman spectroscopy and PLS-DA with uncertainty estimation. *Microchemical Journal* 109:170–177
- DeJonge KC, Mefford n S, Chávez J (2016) Assessing corn water stress using spectral reflectance. *International Journal of Remote Sensing* 37:2294–2312
- Devadas R, Lamb D, Backhouse D, Simpfendorfer S (2015) Sequential application of hyperspectral indices for delineation of stripe rust infection and nitrogen deficiency in wheat. *Precision Agriculture* 16:477–491
- Dinesh KP, Shivanna P, Santa Ram A (2011) Identification of RAPD (random amplified polymorphic DNA) markers for Ethiopian wild *Coffea arabica* L genetic resources in the tropics. *Research: Plant Genomics* 2:1–7
- Eitel JUH, Vierling LA, Litvak ME, Long DS, Schulthess U, Ager AA, Krofcheck DJ, Stoscheck L (2011) Broadband, red-edge information from satellites improves early stress detection in a New Mexico conifer woodland. *Remote Sensing of Environment* 115:3640–3646
- Eskes A (2005) Phenotypic expression of resistance to coffee leaf rust and its possible relationship with durability. *Durable Resistance to Coffee Leaf Rust*. Viçosa MG. Universidade Federal de Viçosa, 305–332
- Feng W, Shen W, He L, Duan J, Guo B, Li Y, Wang C, Guo T (2016) Improved remote sensing detection of wheat powdery mildew using dual-green vegetation indices. *Precision Agriculture* 17:608–627
- Frampton WJ, Dash J, Watmough G, Milton EJ (2013) Evaluating the capabilities of Sentinel-2 for quantitative estimation of biophysical variables in vegetation. *ISPRS Journal of Photogrammetry and Remote Sensing* 82:83–92
- Haddad F, Maffia LA, Mizubuti ES, Teixeira H (2009) Biological control of coffee rust by antagonistic bacteria under field conditions in Brazil. *Biological Control* 49:114–119
- Hansen PM, Schjoerring JK (2003) Reflectance measurement of canopy biomass and nitrogen status in wheat crops using normalized difference vegetation indices and partial least squares regression. *Remote Sensing of Environment* 86:542–553
- Herrmann I, Pimstein A, Karnieli A, Cohen Y, Alchanatis V, Bonfil DJ (2011) LAI assessment of wheat and potato crops by VEN μ S and Sentinel-2 bands. *Remote Sensing of Environment* 115:2141–2151
- Hillnhütter C, Mahlein A-K, Sikora R, Oerke E-C (2011) Remote sensing to detect plant stress induced by *Heterodera schachtii* and *Rhizoctonia solani* in sugar beet fields. *Field Crops Research* 122:70–77
- Höskuldsson A (1988) PLS regression methods. *Journal of Chemometrics* 2:211–228
- Huang W, Lamb D, Niu Z, Zhang Y, Liu L, Wang J (2007) Identification of yellow rust in wheat using in-situ spectral reflectance measurements and airborne hyperspectral imaging. *Precision Agriculture* 8:187–197

- Jia R, Mao Z, Chang Y (2010). A nonlinear robust partial least squares method with application. Paper presented at the 2010 Chinese control and decision conference, Xuzhou, China, 2334–2339
- Jiang J, Hu R, Han Z, Wang Z, Chen J (2013) Two-step superresolution approach for surveillance face image through radial basis function-partial least squares regression and locality-induced sparse representation. *Journal of Electronic Imaging* 22:041120–041120
- Kanke Y, Tuban B, Dalen M, Harrell D (2016) Evaluation of red and red-edge reflectance-based vegetation indices for rice biomass and grain yield prediction models in paddy fields. *Precision Agriculture* 17: 507–530
- Larsolle A, Muhammed HH (2007) Measuring crop status using multivariate analysis of hyperspectral field reflectance with application to disease severity and plant density. *Precision Agriculture* 8:37–47
- Laudien R, Bareth G, Doluschitz R (2004) Comparison of remote sensing based analysis of crop diseases by using high resolution multispectral and hyperspectral data: case study: *Rhizoctonia solani* in sugar beet. Paper presented at the 12th international conference on Geoinformatics – geospatial information research: bridging the Pacific and Atlantic, University of Gävle, Sweden
- Lelong CCD, Roger JM, Brégand S, Dubertret F, Lanore M, Sitorus NA, Raharjo DA, Caliman JP (2010) Evaluation of oil-palm fungal disease infestation with canopy hyperspectral reflectance data. *Sensors* 10:734–747
- Li H, Lee W, Wang K, Ehsani R, Yang C (2014) ‘Extended spectral angle mapping (ESAM)’ for citrus greening disease detection using airborne hyperspectral imaging. *Precision Agriculture* 15:162–183
- Logan WJC, Biscoe J (1987) *Coffee handbook*. Zimbabwe Coffee Growers' Association, Harare
- Mahlein AK, Oerke E-C, Steiner U, Dehne HW (2012a) Recent advances in sensing plant diseases for precision crop protection. *European Journal of Plant Pathology* 133:197–209
- Mahlein AK, Steiner U, Hillnhütter C, Dehne HW, Oerke E-C (2012b) Hyperspectral imaging for small-scale analysis of symptoms caused by different sugar beet diseases. *Plant Methods* 8:3
- Mahlein AK, Rumpf T, Welke P, Dehne HW, Plümer L, Steiner U, Oerke EC (2013) Development of spectral indices for detecting and identifying plant diseases. *Remote Sensing of Environment* 128:21–30
- Martens H, Naes T (2001) Multivariate calibration by data compression. In P. Williams, Norris, K. (Ed.), *Near infrared Technology in the agricultural and food industries* (2nd ed., pp. 59–100). Minnesota: American Association of Cereal Chemists
- Melo GA, Shimizu MM, Mazzafera P (2006) Polyphenoloxidase activity in coffee leaves and its role in resistance against the coffee leaf miner and coffee leaf rust. *Phytochemistry* 67:277–285
- Moriasi DN, Arnold JG, Van Liew MW, Bingner RL, Harmel RD, Veith TL (2007) Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE* 50:885–900
- Moscetti R, Haff RP, Stella E, Contini M, Monarca D, Cecchini M, Massantini R (2015) Feasibility of NIR spectroscopy to detect olive fruit infested by *Bactrocera oleae*. *Postharvest Biology and Technology* 99:58–62
- Mulla DJ (2013) Twenty five years of remote sensing in precision agriculture: key advances and remaining knowledge gaps. *Biosystems Engineering* 114:358–371
- Mutanga O, Ismail R (2010) Variation in foliar water content and hyperspectral reflectance of *Pinus patula* trees infested by *Sirex noctilio*. *Southern Forests: A Journal of Forest Science* 72:1–7
- Mutanga O, Skidmore AK (2007) Red-edge shift and biochemical content in grass canopies. *ISPRS Journal of Photogrammetry and Remote Sensing* 62:34–42
- Orr MJ (1996) Introduction to radial basis function networks: technical report. University of Edinburgh, Center for Cognitive Science
- Oumar Z, Mutanga O (2014) Predicting water stress induced by *Thaumastocoris peregrinus* infestations in plantation forests using field spectroscopy and neural networks. *Journal of Spatial Science* 59:79–90
- Prabhakar M, Prasad YG, Desai S, Thirupathi M, Gopika K, Rao GR, Venkateswarlu B (2013) Hyperspectral remote sensing of yellow mosaic severity and associated pigment losses in *Vigna mungo* using multinomial logistic regression models. *Crop Protection* 45: 132–140
- Price T, Gross R, Wey J, Osborne C (1993) A comparison of visual and digital image-processing methods in quantifying the severity of coffee leaf rust (*Hemileia vastatrix*). *Australian Journal of Experimental Agriculture* 33:97–101
- Ramoelo A, Cho M, Mathieu R, Skidmore AK (2015) Potential of Sentinel-2 spectral configuration to assess rangeland quality. *Journal of Applied Remote Sensing* 9:09409611–094096112
- Ramoelo A, Skidmore AK, Cho MA, Mathieu R, Heitkönig IMA, Dudeni-Tlhone N, Schlerf M, Prins HHT (2013) Non-linear partial least square regression increases the estimation accuracy of grass nitrogen and phosphorus using in situ hyperspectral and environmental data. *ISPRS Journal of Photogrammetry and Remote Sensing* 82:27–40
- Reynolds GJ, Windels CE, MacRae IV, Laguette S (2012) Remote sensing for assessing *Rhizoctonia* crown and root rot severity in sugar beet. *Plant Disease* 96:497–505
- Rosipal R (2010) Nonlinear partial least squares: an overview. *Chemoinformatics and advanced machine learning perspectives: complex computational methods and collaborative techniques*, 169–189
- Rumpf T, Mahlein AK, Steiner U, Oerke EC, Dehne HW, Plümer L (2010) Early detection and classification of plant diseases with support vector machines based on hyperspectral reflectance. *Computers and Electronics in Agriculture* 74:91–99
- Sankaran S, Mishra A, Ehsani R, Davis C (2010) A review of advanced techniques for detecting plant diseases. *Computers and Electronics in Agriculture* 72:1–13
- Savitzky A, Golay M (1964) Smoothing and differentiation of data by simplified least square procedure. *Analytical Chemistry* 36:1627–1638
- Shafri HZM, Anuar M, Seman IA, Noor M (2011) Spectral discrimination of healthy and *Ganoderma*-infected oil palms from hyperspectral data. *International Journal of Remote Sensing* 32: 7111–7129
- Sibanda M, Mutanga O, Rouget M (2016) Comparing the spectral settings of the new generation broad and narrow band sensors in estimating biomass of native grasses grown under different management practices. *GIScience & Remote Sensing* 53:614–633
- Silva M d C, Várzea V, Guerra-Guimarães L, Azinheira HG, Fernandez D, Petitot A-S, Bertrand B, Lashermes P, Nicole M (2006) Coffee resistance to the main diseases: leaf rust and coffee berry disease. *Brazilian Journal of Plant Physiology* 18:119–147
- Thenkabail PS, Smith RB, De Pauw E (2000) Hyperspectral vegetation indices and their relationships with agricultural crop characteristics. *Remote Sensing of Environment* 71:158–182
- Vincini M, Amaducci S, Frazzi E (2014) Empirical estimation of leaf chlorophyll density in winter wheat canopies using Sentinel-2 spectral resolution. *IEEE Transactions of GeoScience & Remote Sensing* 52:3220–3235
- Wold S, Sjöström M, Eriksson L (2001) PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems* 58:109–130
- Yan X, Du W, Qian F (2004) Development of a kinetic model for industrial oxidation of p-xylene by RBF-PLS and CCA. *AICHE Journal* 50:1169–1176
- Yu K, Leufen G, Hunsche M, Noga G, Chen X, Bareth G (2014) Investigation of leaf diseases and estimation of chlorophyll concentration in seven barley varieties using fluorescence and hyperspectral indices. *Remote Sensing* 6:64–86

- Yuan L, Zhang JC, Wang K, Loraamm RW, Huang WJ, Wang JH, Zhao JL (2013) Analysis of spectral difference between the foreside and backside of leaves in yellow rust disease detection for winter wheat. *Precision Agriculture* 14:495–511
- Yuan L, Huang Y, Loraamm RW, Nie C, Wang J, Zhang J (2014) Spectral analysis of winter wheat leaves for detection and differentiation of diseases and insects. *Field Crops Research* 156:199–207
- Yuan L, Zhang H, Zhang Y, Xing C, Bao Z (2017) Feasibility assessment of multi-spectral satellite sensors in monitoring and discriminating wheat diseases and insects. *Optik-International Journal for Light and Electron Optics* 131:598–608
- Zambolim L (2016) Current status and management of coffee leaf rust in Brazil. *Tropical Plant Pathology* 41:1–8
- Zarco-Tejada PJ, Ustin SL, Whiting ML (2005) Temporal and spatial relationships between within-field yield variability in cotton and high-spatial hyperspectral remote sensing imagery. *Agronomy Journal* 97:641–653
- Zhang JC, Pu R-L, Wang J-H, Huang W-J, Yuan L, Luo J-H (2012) Detecting powdery mildew of winter wheat using leaf level hyperspectral measurements. *Computers and Electronics in Agriculture* 85:13–23
- Zhang JC, Pu R, Yuan L, Wang J, Huang W, Yang G (2014) Monitoring powdery mildew of winter wheat by using moderate resolution multi-temporal satellite imagery. *PLoS One* 9(4):e93107