Plant Pathology (2019) Doi: 10.1111/ppa.12996

Developing a spectral disease index for myrtle rust (*Austropuccinia psidii*)

R. H. J. Heim[ab]*, I. J. Wright[a], A. P. Allen[a], I. Geedicke[ab] and J. Oldeland[b]

[a]*Department of Biological Sciences, Macquarie University, Sydney, NSW 2109, Australia;* [b]*Department of Biology, Biocenter Klein Flottbek and Botanical Garden, University of Hamburg Hamburg 22609, Germany*

Since 2010 Australian ecosystems and managed landscapes have been severely threatened by the invasive fungal pathogen *Austropuccinia psidii*. Detecting and monitoring disease outbreaks is currently only possible by human assessors, which is slow and labour intensive. Over the last 25 years, spectral vegetation indices (SVIs) have been designed to assess variation in biochemical or biophysical traits of vegetation. However, diagnosis of individual diseases based on classical SVIs is currently not possible because they lack disease specificity. Here, a novel spectral disease index (SDI), the lemon myrtle–myrtle rust index (LMMR), has been developed. The index was designed from hyperspectral leaf-clip data collected at a lemon myrtle plantation in New South Wales, Australia. A total of 236 fungicide-treated (disease free) and 228 untreated (diseased) lemon myrtle leaves were sampled and a random forest classifier was used to show that the LMMR discriminates those classes with an overall accuracy of 90%. Compared to three classical SVIs (PRI, MCARI, NBNDVI), commonly applied for stress detection, the LMMR clearly improved classification accuracies (58%, 67%, 60%, respectively). If the LMMR can be validated on independent datasets from similar and different host species, it could enable land managers to reduce disease impact by earlier control. There might also be potential to collect useful data for epidemiology models. Calculating the LMMR based on hyperspectral data collected from aerial platforms (e.g. drones) would allow for rapid and high-capacity screening for disease outbreaks.

Keywords: hyperspectral, Myrtaceae, phytopathometry, plant disease detection, rust fungus, spectral vegetation index

Introduction

Plant pathogens, such as rust fungi, play a versatile role in ecology and economy. They affect community dynamics and diversification through coevolution with their host plants (Helfer, 2014) but also cause extensive damage to agricultural and forestry crops. This was recently demonstrated by a new, highly virulent strain of *Puccinia graminis*, a rust fungus that destroyed tens of thousands of hectares of wheat crops in southern Europe (Bhattacharya, 2017). Here, the focus is on the rust fungus *Austropuccinia psidii* (Sphaerophragmiaceae, Pucciniales), an obligate biotrophic plant pathogen in the highly diverse phylum Basidiomycota (Beenken, 2017). In Australia, *A. psidii* is invasive and causes a disease commonly known as myrtle rust that exclusively affects one of Australia's dominant plant families, the Myrtaceae (Carnegie & Pegg, 2018). In contrast to other rust diseases, which are mostly restricted to few host species, myrtle rust infects hundreds of species, escalating the potential consequences for Australia's natural landscapes (Carnegie & Pegg, 2018). Australian native species have already been severely damaged by myrtle rust in the wild (Carnegie & Pegg, 2018).

Industries that rely on species within the Myrtaceae, such as the nursery and garden industry, have also been affected by myrtle rust through losses of commercial varieties, trade restrictions and increased dependency on fungicides (Carnegie & Pegg, 2018). In Australia the expanding lemon myrtle (*Backhousia citriodora*) industry has been particularly affected, as cultivars of *B. citriodora* currently in use are moderately or highly susceptible to myrtle rust (Doran *et al.*, 2012). Leaves of lemon myrtle are commercially harvested to produce lemon-flavoured herbal teas, culinary herbs and lemon-scented essential oils used for food flavouring and personal care products (Clarke, 2012). The farm-gate value of this market has been estimated to be AU$7–23 million annually (Clarke, 2012). Rust-affected leaves of *B. citriodora* are unsuitable for use and cause yield losses up to 70%. The application of fungicides to control the disease is undesirable because the market demands a clean, organic product (Carnegie & Pegg, 2018). Therefore, the industries reliant on lemon myrtle are in urgent need of rust-resistant cultivars or measures to reduce the use of fungicides.

The detection of myrtle rust symptoms and those of other pathogens has traditionally relied on visual

*E-mail: rene.heim@hdr.mq.edu.au

Plant Pathology

assessment. However, visual assessment is somewhat subjective due to the assessor's individual experience, and the visual cues humans use vary through time within individuals (Bock *et al.*, 2010). Automated methods using optical remote sensing have the potential to detect diseases with greater sensitivity, specificity and reliability than what is possible by humans using visual estimation (Mahlein, 2016).

Over the last 30 years, the field of 'precision agriculture' has adopted optical remote sensing to optimize all production-related materials such as fertilizers and agrochemicals (Mulla, 2013). A subfield in precision agriculture is efficient disease detection and management. Acquiring disease-related spectral data is information-intense and often requires a reduction to the most relevant wavebands to reflect the pathosystem under investigation (Stafford, 2000). In many cases, the visual region (VIS, 400–700 nm) has been found most useful for indicating visible disease symptoms (e.g. discolourations), while the near-infrared region (NIR, 700–1300 nm) has indicated changes in structural leaf traits (Jacquemoud & Ustin, 2001; Mulla, 2013). Parallel to the progress in precision agriculture, spectral vegetation indices (SVIs) have been developed to simplify the prediction of biochemical, structural or physiological changes in plants. For instance, the photochemical reflectance index (PRI) was developed as an indicator for the efficiency of carbon fixation using photosynthetic radiation (Gamon *et al.*, 1997). Ashourloo *et al.* (2014) evaluated the effect of wheat rust symptoms on a set of SVIs (e.g. normalized difference vegetation index (NDVI), narrow-band normalized difference vegetation index (NBNDVI) and PRI) and they were found to be effective when disease severity was high, while being less effective in discriminating different symptoms. Mahlein *et al.* (2013) already stated a year earlier that SVIs would not be suitable for disease detection as they were originally designed for other purposes. Therefore, they developed spectral disease indices (SDIs) that could successfully discriminate among different sugar beet diseases.

The primary aim of the present study was to develop an SDI for myrtle rust detection on lemon myrtle plants (*B. citriodora*). The analyses were based on data recorded on a myrtle rust-infested plantation, also used in a previous study (Heim *et al.*, 2018). In that previous work it was shown that fungicide-treated and untreated *B. citriodora* leaves could be classified with high accuracy based on a broad set of 202 wavebands (i.e. predictor variables). In the present study, these 202 wavebands were first refined to provide the minimum number of wavebands required to accurately classify the two classes (treated/untreated). Next, the SDI was designed by using the refined wavebands. Finally, the classification accuracy of the SDI was compared to that of three SVIs widely used in plant disease detection. An additional aim was to provide the first coded framework presented in the R statistical programming environment (R Core Team, 2017) for developing SDIs.

Materials and methods

Data collection

Leaf spectral data were collected on a lemon myrtle plantation in northern New South Wales, Australia (lat −28.6911, long 153.2955). For more information refer to Heim *et al.* (2018). A proportion of trees had been treated with fungicide to control *A. psidii*, while a proportion was untreated and thus diseased. Measurements were made on leaves affected by *A. psidii* (untreated leaves; Fig. 1b,c,d) and on leaves that had been treated with fungicides and therefore showed negligible signs of *A. psidii* infection (Fig. 1a). Leaves from untreated trees had varying levels of disease, including small purple spots through to large necrotic lesions and yellow pustules. Leaves from treated trees showed mostly no signs of *A. psidii* infection, although some had small purple spots, probably due to infection occurring prior to application of fungicides which have been shown effective in halting the infection process (Horwood *et al.*, 2013). The influence of other biotic agents was excluded, as no other serious pathogen on lemon myrtle was known prior to *A. psidii* (Dr Angus Carnegie, Department of Primary Industries, Parramatta, Australia and Gary Mazzorana, Australian Rainforest Products, Lismore, Australia, personal communication). Spectral reflectance signatures between 350 and 2500 nm were recorded with a portable non-imaging spectroradiometer (Spectral Evolution PSR+ 3500) with a spectral resolution of 3 nm steps between 450 and 700 nm, 8 nm steps between 700 and 1500 nm and 6 nm steps between 1500 and 2100 nm. Measurements were made from the adaxial leaf surface using a leaf clip holder with a 3-mm sample area, a built-in reflectance standard and a separate 5 W light source (ILM-105; see Fig. S1 for illumination spectrum). A total of 236 fungicide-treated and 228 untreated lemon myrtle leaf samples were measured, with three leaves sampled per tree ($n = 464$). Further details on sampling design are given by Heim *et al.* (2018).

Data preparation

The original dataset (Heim *et al.*, 2018) contained 2151 spectral wavebands (i.e. predictor variables), thus more predictor variables than observations, a situation referred to as 'high dimensionality' (Hastie *et al.*, 2009). High-dimensional data can contain unknown groups of highly correlated predictors (Genuer *et al.*, 2015). Correlated predictor variables may lead to inaccurate selection of relevant wavebands. To counter this, spectral resampling was used. This reduced the original spectral resolution of 3–8 nm (2151 wavebands) to a resolution of 10 nm (202 wavebands; Heim *et al.*, 2018). The spectral data used in this study is no longer high-dimensional (Mahlein *et al.*, 2013) and still contains 464 spectral reflectance profiles, including 236 fungicide-treated and 228 untreated lemon myrtle leaves.

All analyses were conducted using the R statistical platform (R Core Team, 2017). The full analysis (Fig. S1) can be reproduced using code and data archived at https://github.com/ReneHeim/RustIndex. The provided code has potential to serve as a framework to develop SDIs for other host–pathogen combinations.

Raw data to linear model

Spectral vegetation indices commonly use two to four wavebands, and ratios thereof (Mahlein *et al.*, 2013). Similarly, it

Figure 1 Fungicide-treated (a) and untreated (b, c, d) *Backhousia citriodora* leaves that have been assessed at a lemon myrtle plantation in New South Wales, Australia. Fungicide-treated trees were free of active disease but could show stray, necrotic lesions or purple spots, probably due to infection occurring prior to fungicide application. Leaves that were not treated were largely covered with dark necrotic lesions, purple lesions and yellow spores (d) because *Austropuccinia psidii* was not contained (b). Images: Ina Geedicke

was decided to reduce the 202 wavebands to four wavebands, allowing the design of an easily interpretable index in the form of SVIs like the NBNDVI (Equation 7; Thenkabail *et al.*, 2000).

To create ratio indices from linear models, as attempted here, the original reflectance data (band-wavelength 1, B1; band-wavelength 2, B2) were log-transformed, yielding logged reflectance values (log(B1), log(B2)). By introducing a log-term into a linear equation, the following basic algebra rules (Equations 1 and 2) apply and ratios and products of reflectance adhere to:

$$\log(\mathrm{B1}^{a} * \mathrm{B2}^{b}) = a * \log(\mathrm{B1}) + b * \log(\mathrm{B2}) \quad (1)$$

$$\log(\frac{\mathrm{B1}^{a}}{\mathrm{B2}^{b}}) = a * \log(\mathrm{B1}) - b * \log(\mathrm{B2}) \quad (2)$$

Before these rules became relevant, a random-forest-based feature selection was applied on the original data (Fig. S1a), repeating it 10 times to account for variability in the selection process. This resulted in a set of 27 wavebands that retained predictive power while avoiding redundancy. Here, the R package VSURF (Genuer *et al.*, 2015) was used, as it is suitable for regular and high-dimensional data. This was necessary because the computational effort of a direct exhaustive model selection, as applied in the following, would have been too high and time-consuming using 202 predictor variables. This refined set of 12 wavebands was submitted as a candidate set of predictor variables to an exhaustive model search using a binominal generalized linear model. This allowed identification of a linear model containing the four most relevant wavebands to discriminate the binary response (treated and untreated). The best model was indicated by the small-sample-corrected Akaike information criterion (AIC_c).

At this stage, an intermediate step had to be included to compute the coefficients for the best model, as wavelength (parameters) were provided without the corresponding numerical coefficients. By submitting the linear combination of the best four wavebands (e.g. Response~1 + B545 + B555 + B1505 + B2195) from the previous step to another binominal generalized linear model, the required model coefficients were yielded (Equation 3).

Linear model to classification report

The best four parameters including their coefficients from the above-described binominal generalized linear model characterized the predicted probability, p, that a given leaf was infected (Equation 3). As mentioned in the beginning, the model (Equation 3) contains log-transformed reflectance values to make use of the algebraic rules (Equations 1 and 2):

$$\log\left(\frac{p}{1-p}\right) = 18.39 + 75.38 * \log(\text{B545}) - 78.81 * \log(\text{B555}) + 45.99 * \log(\text{B1505}) - 46.83 * \log(\text{B2195}) \quad (3)$$

Wavelengths 545 and 555 nm straddle the VIS spectrum; 1505 and 2195 nm are both in the short-wavelength infrared (SWIR) spectrum. In this model (Equation 3), the coefficients for log(B545) and log(B555) are of approximately equal magnitude and opposite in sign, as are the coefficients for log(B1505) and log(B2195). This observation also indicates that both pairs of variables can be treated as ratios for the construction of the specific disease index and is further supported by overlapping 95% confidence intervals found during the analysis (95% CI B545 [57.01, 96.09], B555 [99.93, 60.08], B1505 [37.32, 56.03], B2195 [57.07, 38.05]). The magnitudes of the coefficients for log(B545) and log(B555) are approximately 1.66-times greater than those for log(B1505) and log(B2195). Thus, to transpose Eqn 3 into the form of a ratio SDI, giving the lemon myrtle–myrtle rust index (LMMR; Equation 6), the following steps were applied (see also Fig. S1):

Summarize coefficients of approximately equal magnitude and opposite in sign:

$$\log\left(\frac{p}{1-p}\right) = 18.39 + 76.50 * \log\left(\frac{\text{B545}}{\text{B555}}\right) + 46.50 * \log\left(\frac{\text{B1505}}{\text{B2195}}\right) \quad (4)$$

Drop constant coefficient (18.387 = const.) and transpose further ($\frac{76.5}{46.5} \approx \frac{5}{3}$):

$$\log\left(\frac{p}{1-p}\right) = \frac{5}{3} * \log\left(\frac{\text{B545}}{\text{B555}}\right) + \log\left(\frac{\text{B1505}}{\text{B2195}}\right) \quad (5)$$

Take exponential of both sides:

$$\frac{p}{1-p} = \left(\frac{\text{B545}}{\text{B555}}\right)^{\frac{5}{3}} * \frac{\text{B1505}}{\text{B2195}} = \text{LMMR} \quad (6)$$

To assess the performance of the LMMR (Equation 6), its accuracy to discriminate untreated and treated lemon myrtle leaves was compared to the accuracy of spectral vegetation indices commonly applied to detect plant pathogens (Mahlein *et al.*, 2013; Ashourloo *et al.*, 2014). These indices were selected according to the biological processes they indicate and whether these processes could be linked with physiological changes caused by myrtle rust. For example, urediniospores of rust fungi contain carotenoids and melanin-like pigments, hence their brown-orange-yellow colour (Mahlein *et al.*, 2013). Changes in plant pigments can be detected, amongst others, by applying either the photochemical reflectance index (PRI; Equation 7; Gamon *et al.*, 1997) or the modified chlorophyll absorption in reflectance index (MCARI; Equation 8; Daughtry, 2000). Also, the structural integrity of the mesophyll cells is reduced when hyphae of *A. psidii* enter this cell layer (Morin *et al.*, 2014). Processes that interfere with the cellular integrity, and therefore cause stress, are usually reflected in the near-infrared region (Peñuelas & Filella, 1998). Therefore, the narrow-band normalized difference vegetation index (NBNDVI; Equation 9, Thenkabail *et al.*, 2000) could mirror this variation as it measures the ratio between the near-infrared and visual region.

$$\text{PRI} = \frac{\text{B531} - \text{B570}}{\text{B531} + \text{B570}} \quad (7)$$

$$\text{MCARI} = ((\text{B700} - \text{B670}) - 0.2 * (\text{B700} - \text{B550})) * \left(\frac{\text{B700}}{\text{B670}}\right) \quad (8)$$

$$\text{NBNDVI} = \frac{\text{B850} - \text{B680}}{\text{B850} + \text{B680}} \quad (9)$$

Values were calculated for each index from the original reflectance data and yielded a new dataset (n = 464) containing two response classes (treated and untreated) and four predictor variables (PRI, MCARI, NBNDVI and LMMR). This data was randomly split (75/25) into a training set (n = 348) and a test set (n = 116). As the LMMR was developed on the log-scale, it should only receive log-transformed data when compared to other indices.

A logistic regression classifier was used to evaluate which index was the most accurate predictor variable for the classification problem. To increase model accuracy, data were resampled (drawing random samples with replacement) using the '0.632+ bootstrap' method (Efron & Tibshirani, 1997); this approach estimates prediction error with less variability than cross-validation (Efron & Tibshirani,1997). The training models of all four indices were used to predict the probability, using a threshold of 0.5, that a leaf/tree in the test data was either fungicide-treated or untreated. The test dataset was not seen by the classifier before and could therefore be used to validate the models. To evaluate the prediction performance, an error matrix was produced containing the following metrics: overall accuracy (OA), producer accuracy (PA) and user accuracy (UA).

By default, the accuracy of the training and testing process was evaluated using OA as a metric. OA reflects the agreement between the reference and predicted classes and has the most direct interpretation. However, it does not provide information about the origin of an error (Kuhn & Johnson, 2013). Here, PA and UA can indicate class-specific errors (Congalton & Green, 2009). PA is the number of correctly classified references for a class divided by the total number of references of that class and, thus, represents the accuracy of the classification for a specific class. UA divides the number of correct classifications (predictions) for a class by the total number of classifications (predictions) for that class. A high UA means that spectra within that class can be reliably classified as belonging to that class. User accuracy is often termed to be a measure of reliability, which can be also interpreted as the agreement between repeated measurements within a class.

Results

The four most import spectral wavebands, 545, 555, 1505 and 2195 nm, were selected (Fig. 2e; vertical, dashed lines) from a dataset originally containing 202 wavebands for each spectral signature. The binominal generalized linear model containing these wavebands as parameters was more successful in predicting whether a lemon myrtle tree was treated with fungicides or untreated than models containing other wavebands between 500 and 2500 nm. While the wavebands at 545 and 555 nm are situated in the visual region (VIS 400–700 nm) of the electromagnetic spectrum, the wavebands at 1505 and 2195 nm can be found in the short-wave

infrared region (SWIR 1300–2500 nm). Based on these wavebands, a new disease-specific spectral index, the LMMR was derived (Equation 6).

LMMR classification performance

The training process of the classifier was assessed graphically (Fig. 2a–d). The PRI and the MCARI (Fig. 2a,b) could discriminate between treated (red circles) and untreated (blue triangles) lemon myrtle trees only marginally (OA: PRI = 66.7%, MCARI = 66.3%). The NBNDVI (Fig. 2c) does not improve disease detection over random guessing (OA: NBNDVI = 52.9%). By contrast, the LMMR (Fig. 2d) could clearly discriminate treated and untreated trees in the training process (OA: LMMR = 86.5%).

Twenty-five percent of the data was isolated before running the training procedure so as to validate the classifier on data not yet seen by the classifier. For the validation, LMMR substantially outperformed other indices in predicting the disease. The LMMR classified untreated and treated trees with an overall accuracy of 90% (Table 1a–d, lower right cells). Other indices ranged from OA 58% to 67%. Evaluating producer accuracies (PA) and user accuracies (UA) yielded the same overall trend. The MCARI had similar UA (treated = 68%, untreated = 66%) and PA (treated = 66%, untreated = 68%). Also, the UA for both indices (Table 1a,c), the PRI (treated = 58%, untreated = 58%) and the NBNDVI (treated = 59%, untreated = 63%) are balanced. For the PA, the probability that a certain class found on the plantation is classified as such, it seems that treated trees can be detected slightly better (PRI, treated = 63%, untreated = 53%; NBNDVI, treated = 75%, untreated = 46%). Overall, the LMMR delivers high user accuracies (treated = 89%, untreated = 91%) and high producer accuracies (PA) for both classes (treated = 92%, untreated = 88%).

Discussion

This study derived a new potential spectral disease index (SDI) that allowed detection of symptoms caused by the invasive fungal pathogen *A. psidii* on lemon myrtle trees (*B. citriodora*). The LMMR (lemon myrtle–myrtle rust) index discriminated between fungicide-treated and untreated lemon myrtle plants with notably higher accuracy (90%) than classical spectral vegetation indices (SVIs; 58–67%).

The increased classification accuracy was achieved by selecting the four most relevant wavebands from initially 202 wavebands. They are specific to this pathosystem and were able to perform better than indices developed for other situations. The aim was to drop as many wavebands as possible while sustaining substantial prediction accuracies, and was guided by the common principle to use three-band indices at the leaf-scale and four band indices at canopy-scale (Thenkabail *et al.*, 2000).

The waveband selection process resulted in two wavebands (545 and 555 nm) with high predictive power for myrtle rust disease in the VIS region of the electromagnetic spectrum, and two wavebands (1505 and 2195 nm)

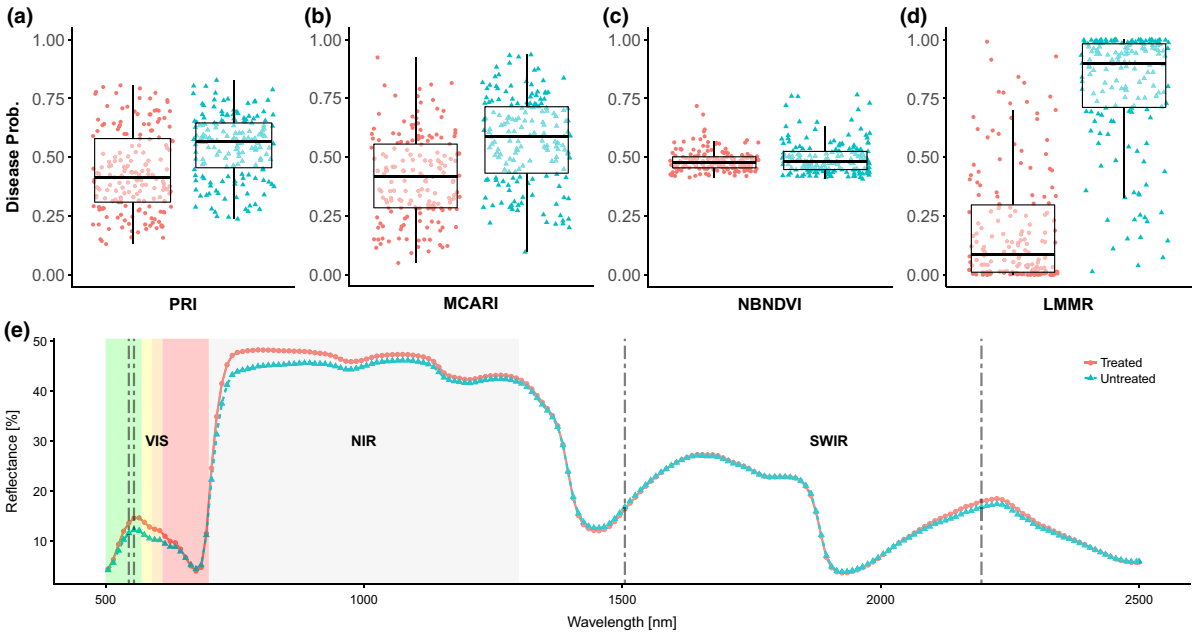

Figure 2 The potential of the applied classifier to discriminate treated (red circle) and untreated (blue triangle) lemon myrtle leaves (*Backhousia citriodora*) after the training process. Plot (e) shows a classical subdivision of the electromagnetic spectrum (VIS, NIR, SWIR) and the locations of the four most important wavebands to successfully discriminate the spectral signatures.

Table 1 Accuracy assessment for the logistic regression classification using validation data (116 observations) that was isolated before the classifier was trained.

(a) PRI		Reference TR	UN	Total	UA (%)
Pred	TR	37	27	64	58
	UN	22	30	52	58
	Total	59	57	116	
PA (%)		63	53		58

(b) MCARI		Reference TR	UN	Total	UA (%)
Pred	TR	39	18	57	68
	UN	20	39	59	66
	Total	59	57	116	
PA (%)		66	68		67

(c) NBNDVI		Reference TR	UN	Total	UA (%)
Pred	TR	44	31	75	59
	UN	15	26	41	63
	Total	59	57	116	
PA (%)		75	46		60

(d) LMMR		Reference TR	UN	Total	UA (%)
Pred	TR	54	7	61	89
	UN	5	50	55	91
	Total	59	57	116	
PA (%)		92	88		90

Classification was performed using the index values derived by applying the indices PRI (a), MCARI (b), NBNDVI (c) and LMMR (d) on spectral reflectance data from fungicide-treated (TR) and untreated (UN) lemon myrtle trees. Accuracy can be evaluated comparing the overall accuracy in every lower right corner of each table and user accuracies (UA) and producer accuracies (PA). The error matrix also shows class totals for the reference columns and prediction rows. The number of correctly classified trees is highlighted in each grey-shaded cell.

in the SWIR region. Variation in reflectance in the VIS region between treated and untreated leaves is mainly caused by changing contents of leaf pigments, while reflectance variation in the SWIR region is often influenced by the composition of leaf chemicals and water content (Jacquemoud & Ustin, 2001).

This study found variation in spectral reflectance around 550 nm, and it is known that between 510 and 550 nm spectral variation is closely related to the total carotenoid pigment content of leaves (Gitelson *et al.*, 2002). Carotenoids are presumably the pigments giving the yellow colour to urediniospores of some rusts (Wang *et al.*, 2018). On *B. citriodora*, yellow pigmented pustules were found on the adaxial (Fig. 3a–c) and abaxial (Fig. 3d–f) leaf surfaces of infected leaves. It is likely that the same pigments also occur within the leaves, as during the infection and penetration process of *A. psidii*, the orange-yellow pigmented contents are transferred into the leaf by the infection hyphae (Hunt, 1968). There are no studies describing the exact biochemical composition of *A. psidii* pigments (Dr Robert Park, Park Plant Breeding Institute, The University of Sydney, Australia, personal communication).

Red discolourations were also observed around lesions caused by *A. psidii* (Fig. 3a–f). Anthocyanins are the basis for most orange, pink, red, magenta, purple, blue and blue-black colours in plants (Davies, 2004) and might be responsible for the red colouration around lesions, as they are often found at later stages of an infection (Glen *et al.*, 2007). Anthocyanins are water-soluble vacuolar pigments of higher plants that are abundant in juvenile and senescing plants and are represented by a spectral reflectance peak around 550 nm (Gitelson *et al.*, 2007). Thus, anthocyanins might be responsible for the observed spectral shift around 550 nm.

Red discolourations were observed on young leaves (Fig. 3h, i) of *B. citriodora* plants, and this might be regarded as a confounding factor. However, red young leaves were present on treated as well as on untreated plants. As the spectral feature linked to anthocyanin content was still selected, it should represent a difference in pigment content. Both carotenoids and anthocyanins absorb light between 500 and 550 nm (Ustin *et al.*, 2009). Nearby wavebands are usually highly correlated, and these have been selected, although methods were applied that were designed to avoid choosing correlated bands. Overlapping signals often result in inconsistencies in separating and quantifying different pigments (Ustin *et al.*, 2009). However, as the wavebands at 545 and 555 nm were chosen consistently in the present study, this may indicate that carotenoids as well as anthocyanins are both independently important for indicating the presence of *A. psidii* urediniospores.

For the two important wavebands at 1505 and 2195 nm found in this study, it is assumed that these might be caused by lack of water, caused by necrotic lesions occurring on leaves during *A. psidii* infection (Glen *et al.*, 2007). Within the SWIR region (1300–2500 nm), light is primarily absorbed by water in a fresh leaf, but also by dry matter. Therefore, this region is linked to changes in water content (Peñuelas & Filella, 1998). It has been shown that water loss in leaves can be caused by the destruction of the leaf cuticle (Lindenthal *et al.*, 2005) that, in this case, was damaged by many necrotic lesions on untreated leaves. Leaves from fungicide-treated trees did have some evidence of *A. psidii* infection (purple spots), probably due to infection occurring prior to fungicide application. However, the fungicides used (e.g. triadimenol and azoxystrobin) have been shown to work effectively as eradicants (i.e. kill the rust) (Horwood *et al.*, 2013) such that these purple spots did not develop further into yellow pustules and necrotic spots as they did on untreated trees. Furthermore, prior to myrtle rust, there were no other significant biotic

Figure 3 Lemon myrtle (*Backhousia citriodora*) leaves as they were assessed at the plantation in New South Wales, Australia. (a–c) adaxial leaf surface with yellow urediniospores (*Austropuccinia psidii*) present; (d–f) yellow urediniospores on the abaxial leaf surface. Red discolourations around lesions are visible from both sides. Images (g–i) show reddish young leaves with various hue intensities. Images: Ina Geedicke.

agents that caused damage to lemon myrtle trees at the experimental site (Gary Mazzorana, Australian Rainforest Products, personal communication). Of course, it needs further testing whether the LMMR index can specifically detect myrtle rust against other stress-causing agents. However, differentiation among diseases using optical sensors has already been proven feasible for sugar beet pathogens (Mahlein *et al.*, 2013).

In conclusion, this study presents a newly developed spectral disease index (SDI) that performs better (90% OA) than common spectral vegetation indices (SVIs, 58–67% OA). By publishing the code of this analysis, a framework is provided to generate new SDIs for other pathosystems. While further testing and validation for the LMMR is required, the concept of specific disease indices is a promising tool in plant disease detection (Mahlein *et al.*, 2013). This study was conducted in a plantation setting where leaves on untreated trees had varying levels of severity of *A. psidii*, from small purple spots through to necrotic lesions. Thus, the LMMR index is specific to physiological and phenotypic changes caused by *A. psidii*. Confounding stress-causing agents could be excluded during the study, as effective fungicides were applied and no stress-causing agent prior to *A. psidii* was known at this site. Future research could focus on the development of specific disease indices for certain infection stages (e.g. early). Additionally, it would be interesting to test the LMMR index on infected lemon myrtle plants at different locations and against other abiotic and biotic stress-causing agents. Moreover, it should be tested if the index correlates with disease severity. Similar goals for the development of specific disease indices have already been postulated (Mahlein, 2016). For the lemon myrtle industry, which seeks to meet organic standards to be able to compete economically (Doran *et al.*, 2012), a validated LMMR could enable land managers to assess highly infected areas of their arable land and make decisions on fungicide applications.

Acknowledgements

The authors thank Gary Mazzorana from Australian Rainforest Products for access to his plantation, Dr Gerhard Muche for invaluable advice on the mathematical development of the LMMR and Dr Daniel S. Falster for his input on the reproducible publication of the code. Special thanks to Drs Anne-Katrin Mahlein, Robert Park and Angus Carnegie for providing invaluable feedback on the plant pathological aspects of the manuscript. Finally, the authors thank Emily K. Lancaster and her

supervisors, Drs Geoff Pegg and Andre Drenth for allowing them to use their experimental setup to collect the spectral signatures. R.H.J.H. was supported by a Macquarie University Research Excellence Scholarship. The authors declare no conflict of interest.

References

Ashourloo D, Mobasheri MR, Huete A, 2014. Evaluating the effect of different wheat rust disease symptoms on vegetation indices using hyperspectral measurements. *Remote Sensing* **6**, 5107–23.

Beenken L, 2017. *Austropuccinia*: a new genus name for the myrtle rust *Puccinia psidii* placed within the redefined family Sphaerophragmiaceae (Pucciniales). *Phytotaxa* **297**, 53–61.

Bhattacharya S, 2017. Deadly new wheat disease threatens Europe's crops. *Nature* **542**, 145–6.

Bock CH, Poole GH, Parker PE, Gottwald TR, 2010. Plant disease severity estimated visually, by digital photography and image analysis, and by hyperspectral imaging. *Critical Reviews in Plant Sciences* **29**, 59–107.

Carnegie A, Pegg G, 2018. Lessons from the incursion of myrtle rust in Australia. *Annual Review of Phytopathology* **56**, 1–22.

Clarke M, 2012. Australian Native Food Industry Stocktake. Rural Industries Research and Development Corporation Publication No. 12/066. Barton, Australia: Rural Industries Research and Development Corporation.

Congalton R, Green K, 2009. *Assessing the Accuracy of Remotely Sensed Data, Principles and Practices*. Boca Raton, FL, USA: CRC Press.

Daughtry C, 2000. Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance. *Remote Sensing of Environment* **74**, 229–39.

Davies K, 2004. *Plant Pigments and their Manipulation*. Oxford, UK: Blackwell.

Doran J, Lea D, Bush D, 2012. *Assessing Myrtle Rust in a Lemon Myrtle Provenance Trial*. Rural Industries Research and Development Corporation. Publication No. 12/098. Barton, Australia: Rural Industries Research and Development Corporation.

Efron B, Tibshirani R, 1997. Improvements on cross-validation: the 0.632+ bootstrap method. *Journal of the American Statistical Association* **92**, 548–60.

Gamon JA, Serrano L, Surfus JS, 1997. The photochemical reflectance index: an optical indicator of photosynthetic radiation use efficiency across species, functional types, and nutrient levels. *Oecologia* **112**, 492–501.

Genuer R, Poggi J-M, Tuleau-Malot C, 2015. VSURF: an R package for variable selection using random forests. *The R Journal* **7**, 19–33.

Gitelson AA, Zur Y, Chivkunova OB, Merzlyak MN, 2002. Assessing carotenoid content in plant leaves with reflectance spectroscopy. *Photochemistry and Photobiology* **75**, 272–81.

Gitelson AA, Merzlyak MN, Chivkunova OB, 2007. Optical properties and nondestructive estimation of anthocyanin content in plant leaves. *Photochemistry and Photobiology* **74**, 38–45.

Glen M, Alfenas AC, Zauza EAV, Wingfield MJ, Mohammed C, 2007. *Puccinia psidii*: a threat to the Australian environment and economy – A review. *Australasian Plant Pathology* **36**, 1–16.

Hastie T, Tibshirani R, Friedman J, 2009. *The Elements of Statistical Learning Data Mining, Inference, and Prediction*. New York, NY, USA: Springer.

Heim RHJ, Wright IJ, Chang H-C *et al.*, 2018. Detecting myrtle rust (*Austropuccinia psidii*) on lemon myrtle trees using spectral signatures and machine learning. *Plant Pathology* **67**, 1114–21.

Helfer S, 2014. Rust fungi and global change. *New Phytologist* **201**, 770–80.

Horwood M, Carnegie A, Park R, 2013. Gathering efficacy data to indentify the most effective chemicals for controlling myrtle rust (*Uredo rangelii*). Plant Health Australia Research Project P219 Final Report. Canberra, Australia: Plant Health Australia. [http://www.plantnethealthaustralia.com.au/wp-content/uploads/2018/10/Gathering-efficacy-data-to-indentify-the-most-effective-chemicals-for-controlling-myrtle-rust.pdf]. Accessed 23 January 2019.

Hunt P, 1968. Cuticular penetration by germinating uredospores. *Transactions of the British Mycological Society* **51**, 103–12.

Jacquemoud S, Ustin SL, 2001. Leaf optical properties: a state of the art. In: *Proceedings of the 8th International Symposium Physical Measurements & Signatures in Remote Sensing*. Aussois, France: CNES, 223–32.

Kuhn M, Johnson K, 2013. *Applied Predictive Modeling*. New York, NY, USA: Springer.

Lindenthal M, Steiner U, Dehne H-W, Oerke E-C, 2005. Effect of downy mildew development on transpiration of cucumber leaves visualized by digital infrared thermography. *Phytopathology* **95**, 233–40.

Mahlein A-K, 2016. Plant disease detection by imaging sensors – parallels and specific demands for precision agriculture and plant phenotyping. *Plant Disease* **100**, 241–51.

Mahlein AK, Rumpf T, Welke P *et al.*, 2013. Development of spectral indices for detecting and identifying plant diseases. *Remote Sensing of Environment* **128**, 21–30.

Morin L, Talbot MJ, Glen M, 2014. Quest to elucidate the life cycle of *Puccinia psidii sensu lato*. *Fungal Biology* **118**, 253–63.

Mulla DJ, 2013. Twenty-five years of remote sensing in precision agriculture: key advances and remaining knowledge gaps. *Biosystems Engineering* **114**, 358–71.

Peñuelas J, Filella L, 1998. Visible and near-infrared reflectance techniques for diagnosing plant physiological status. *Trends in Plant Science* **3**, 151–6.

R Core Team, 2017. R: a language and environment for statistical computing. Vienna, Austria: Foundation for Statistical Computing. [https://www.R-project.org/]. Accessed 21 January 2019.

Stafford JV, 2000. Implementing precision agriculture in the 21st century. *Journal of Agricultural and Engineering Research* **76**, 267–75.

Thenkabail PS, Smith RB, De Pauw E, 2000. Hyperspectral vegetation indices and their relationships with agricultural crop characteristics. *Remote Sensing of Environment* **71**, 158–82.

Ustin SL, Gitelson AA, Jacquemoud S *et al.*, 2009. Retrieval of foliar information about plant pigment systems from high resolution spectroscopy. *Remote Sensing of Environment* **113**, 67–77.

Wang E, Dong C, Park RF *et al.*, 2018. Carotenoid pigments in rust fungi: extraction, separation, quantification and characterisation. *Fungal Biology Reviews* **32**, 166–80.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site.

Figure S1. Workflow summarizing each step from original raw data to the final classification report. (a) This section produces the linear, parsimonious model including the four most relevant wavebands and their coefficients. (b) This section takes the parsimonious model from (a) which is transformed and simplified to yield the new spectral index specific to the pathosystem lemon myrtle–myrtle rust (LMMR). The performance of the LMMR, to discriminate treated and untreated lemon myrtle trees, is compared against common spectral vegetation indices PRI, MCARI and NBNDVI.