BSPP

Plant Pathology (2018) Doi: 10.1111/ppa.12830

Detecting myrtle rust (*Austropuccinia psidii*) on lemon myrtle trees using spectral signatures and machine learning

R. H. J. Heim[ab]*, I. J. Wright[a], H.-C. Chang[c], A. J. Carnegie[d], G. S. Pegg[e], E. K. Lancaster[f], D. S. Falster[g] and J. Oldeland[b]

[a]Department of Biological Sciences, Macquarie University, Sydney, NSW 2109, Australia; [b]Biodiversity, Ecology and Evolution of Plants, Biocentre Klein Flottbek and Botanical Garden, University of Hamburg, Hamburg, 22609, Germany; [c]Department of Environmental Sciences, Macquarie University, Sydney, NSW 2109; [d]NSW Department of Primary Industries, NSW Forest Science, Parramatta, NSW 2150; [e]Department of Agriculture and Fisheries, Horticulture and Forestry Science, Brisbane, Qld; [f]Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, St Lucia, Qld 4067; and [g]School of Biological, Earth and Environmental Sciences, University of New South Wales, Sydney, NSW 2025, Australia

Hundreds of species in one of Australia's dominant plant families, the Myrtaceae, are at risk from the invasive pathogenic fungus *Austropuccinia psidii*. Since its arrival in Australia in 2010, native plant communities have been severely affected, with highly susceptible species likely to become extinct from recurring infections. While severe impact on Australian native and plantation forestry has been predicted, the lemon myrtle industry is already under threat. Commercial cultivars of lemon myrtle (*Backhousia citriodora*) are highly susceptible to *A. psidii*. Detecting and monitoring disease outbreaks is currently only possible by eye, which is costly and subject to human bias. This study aims at developing a proof-of-concept for automated, non-biased classification of healthy (naïve), fungicide-treated and diseased lemon myrtle trees by means of their spectral reflectance signatures. From a lemon myrtle plantation, spectral signatures of fungicide-treated and untreated leaves were collected using a portable field spectrometer. A third class of spectra, from naïve lemon myrtle leaves that had not been exposed to *A. psidii*, was collected from a botanical garden. Reflectance spectra in their primary form and their first-order derivatives were used to train a random forest classifier resulting in an overall accuracy of 78% (kappa = 0.68) for primary spectra and 95% (kappa = 0.92) for first-order derivative-transformed spectra. Thus, an optical sensor-based discrimination, using spectral reflectance signatures of this as yet uninvestigated pathosystem, seems technically feasible. This study provides a foundation for the development of automated, sensor-based detection and monitoring systems for myrtle rust.

Keywords: derivative spectrometry, field spectroscopy, plant pathogens, precision agriculture, random forest, rust fungi

Introduction

Rust fungi and other plant pathogens are affecting humans and their environment by damaging plants and their products on which we depend for clothing, housing and, most importantly, food. Outbreaks of rust fungi may result in extensive damage to agricultural and forestry crops, as seen when a new, highly virulent strain of *Puccinia graminis* destroyed tens of thousands of hectares of wheat crops in southern Europe (Bhattacharya, 2017). This study focuses on the rust fungus *Austropuccinia psidii* (Sphaerophragmiaceae, Pucciniales). In Australia, *A. psidii* causes a disease commonly known as myrtle rust and is an obligate biotroph and pathogenic organism in the highly diverse phylum Basidiomycota (Helfer, 2014). In contrast to most other rust diseases, myrtle rust has the potential to infect hundreds of different species, escalating the potential consequences of infection.

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

Myrtle rust has already caused damage to a multitude of species in South and Central America, its native region (Coutinho *et al.*, 1998), and to native vegetation in various countries between the Americas and Australia (Loope, 2010). *Austropuccinia psidii* was first identified in Australia in 2010, on the central east coast of New South Wales (NSW; Carnegie *et al.*, 2010). Subsequently, a single strain of *A. psidii* (Machado *et al.*, 2015) has undergone a remarkable range expansion, establishing along the Australian east coast from NSW to Queensland, and with localized distributions in Victoria, Tasmania and Northern Territory (Carnegie *et al.*, 2016; Berthon *et al.*, 2018). Myrtle rust has since also invaded New Caledonia (Giblin, 2013), South Africa (Roux *et al.*, 2013), Indonesia (McTaggart *et al.*, 2016), Singapore (du Plessis *et al.*, 2017) and, most recently, New Zealand (www.mpi.govt.nz, accessed 02 December 2017).

Although most rust pathogens are limited to infecting only a few host species (Makinson, 2014), myrtle rust infects many hundred species, meaning the potential impact on the Australian flora is very serious. The

Myrtaceae is one of the dominant plant families in the Australian flora, contributing more than 2000 species (approximately 10% of the total flora), including iconic, widespread genera such as *Eucalyptus* (gum trees) and *Melaleuca* (paper barks). About half of all Australian Myrtaceae occur in climatic zones identified as suitable for the establishment of myrtle rust (Berthon *et al.*, 2018). Currently, over 500 host species in 86 Myrtaceae genera are known worldwide, with 347 of these species occurring in Australia (Carnegie *et al.*, 2016; Soewarto *et al.*, 2017). Of the species studied in Australia, including those from controlled screening (Potts *et al.*, 2016) and field-based assessments (Pegg *et al.*, 2014; Carnegie *et al.*, 2016), 90% have been identified as being susceptible to *A. psidii*. Several species have been identified as being highly susceptible, with severe decline in natural populations recorded. This includes the common species *Rhodamnia rubescens* and *Rhodomyrtus psidioides*, where deaths of mature stands have been reported (Carnegie *et al.*, 2016). Both species, previously listed as 'least concern', have now been provisionally listed as 'critically endangered' in NSW. The NSW scientific committee of the Department for Environment and Heritage has acknowledged myrtle rust as constituting a major threat to the native Australian environment and the Myrtaceae, listing it as a 'key threatening process' (NSW Scientific Committee, 2011).

Unfortunately, the impact of myrtle rust in Australia has not been limited to native ecosystems, with industries reliant on Myrtaceae also affected. Loss of commercial varieties and trade restrictions, in addition to increased reliance on fungicides, have severely affected the nursery and garden industry. The young, expanding lemon myrtle industry has also been significantly impacted (Doran *et al.*, 2012). *Backhousia citriodora* (lemon myrtle) is a small to medium-sized tree (2 to 30 m), occurring naturally in Queensland coastal forests from Brisbane to Mackay. Lemon myrtle leaves are rich in antioxidants, vitamin E, lutein (a carotenoid compound important for eye function) and calcium. Lemon myrtle has antimicrobial and antifungal properties that are superior to tea tree oil (Rural Industries Research and Development Corporation, 2012). Leaves are commercially harvested to produce lemon-flavoured herbal teas, culinary herbs or lemon-scented essential oils used for food flavouring and in personal care products. Cultivars of *B. citriodora* currently in use are moderately to highly susceptible to myrtle rust. Rust-affected leaves of *B. citriodora* are unsuitable for its main uses and the application of fungicides to control the disease is undesirable as the market demands a clean, organic product. A total farm gate value at between AU$7 million and AU$23 million is estimated for dried leaf and essential oil, respectively (Rural Industries Research and Development Corporation, 2012). Therefore, the industry's reliance on lemon myrtle is in urgent need of rust-resistant cultivars or measures to reduce the use of fungicides. Reports of susceptibility within the eucalypts (Pegg *et al.*, 2014; Potts *et al.*, 2016) indicate an escalation of the problem as it suggests the potential of myrtle rust to affect the forestry industry in Australia, both native and plantation. In Brazil, commercial plantations of *Eucalyptus globulus* and *E. viminalis* have suffered reduced growth and yield loss because of myrtle rust incursions (Alfenas *et al.*, 2003).

Myrtle rust forms purplish lesions with abundant bright, orange-yellow urediniospores on young leaves and shoots, which may die-back because of rust attack. Current field identification of these symptoms and disease incidence assessments for myrtle rust and other plant pathogens are reliant on trained experts and are dependent on the experience and performance of individuals that vary considerably, leading to issues with repeatability (Mahlein, 2016). Thus, disease assessment is subject to human bias. Automated, sensor-based disease detection can be performed with high reliability, sensitivity and specificity and improve the assessment of disease incidence and severity beyond the processes of visual disease detection (Mahlein, 2016). To date, no one has explored sensor-based methods that could detect these and less obvious symptoms of myrtle rust.

Currently there is increasing interest in using spectral reflectance measurements (field spectroscopy) to detect and discriminate plant pathogens in precision agriculture (Mahlein, 2016). Spectral reflectance signatures of vegetation can indicate biochemical, physiological and molecular changes caused by abiotic or biotic processes (Mahlein *et al.*, 2010). Disease symptoms often result from such changes brought about by pathogens and can be investigated by analysing spectral reflectance signatures (Bravo *et al.*, 2003; Delalieux *et al.*, 2007; Mahlein *et al.*, 2010). Mahlein *et al.* (2010) used reflectance spectra of sugar beet leaves to show that there was a distinctive differentiation of three sugar beet fungal pathogens, *Cercospora beticola* (cercospora leaf spot), *Erysiphe betae* (powdery mildew) and *Uromyces betae* (beet rust). Bravo *et al.* (2003) built a classification model that could discriminate wheat plants infected with *Puccinia striiformis* (yellow rust) from healthy ones with an overall accuracy of 96%. However, spectral reflectance signatures are very specific to the source of reflection (e.g. specific to the pathogen infecting a certain species or specific to the content of biochemical compounds of a leaf) and more research is required to explore the utility of these approaches for other pathosystems.

The present study builds on the fact that spectral information (including visible light) is reflected by leaf surfaces. This reflection was captured with an optical sensor and portrayed as a waveform (intensity versus wavelength). Waveforms from different plant pathogens are likely to vary in distinct sections of the light spectrum, e.g. in the visible portion (VIS, 400–700 nm), where the bright, orange-yellow urediniospores of myrtle rust would potentially cause variation in reflectance. For myrtle rust, no efforts have yet been made to investigate its specific spectral reflectance signature. Consequently, the aim of this study was to test whether it is possible to spectrally discriminate naïve, fungicide-treated and infected leaves of *B. citriodora* trees. In this study the following questions were addressed:

(i) Can the spectral response of naïve and fungicide-treated *B. citriodora* individuals be distinguished from ones displaying infection symptoms of *A. psidii* (myrtle rust)?
(ii) Amongst all predictor variables (wavebands), what are the most useful wavebands for discriminating spectral responses of these classes?

Materials and methods

Study site

Spectral reflectance signatures of lemon myrtle leaves were measured at two locations in subtropical eastern Australia. The first site was a commercial lemon myrtle plantation in northern NSW (lat. −28.691, long. 153.295) and the second site was the Australian Botanic Garden at Mount Annan (lat. −34.071, long. 150.766), 800 km south of the plantation, also in NSW near Sydney. At the plantation site, the mean annual temperature is 19.4 °C and mean annual rainfall 1343 mm, while a mean annual temperature of 16.7 °C and mean annual rainfall of 792.4 mm have been recorded at the botanical garden (Australian Government Bureau of Meteorology, 2017). The plantation site was selected to take advantage of an existing experiment in which the impact of fungicide was being measured on lemon myrtle trees affected by myrtle rust. The plantation had trees that were free of active disease symptoms, having had fungicide successfully applied to them (‘treated’ trees), and ‘untreated’ trees, showing symptoms of active myrtle rust infection. Treated trees could potentially have been infected previously with myrtle rust (prior to fungicide application) and thus the leaves may have had necrotic lesions even after killing the fungus by fungicide treatment. Consequently, the botanical garden was chosen as an additional field site: it offered plants in a region that suffers only rare episodes of myrtle rust, and were, therefore, free from infection and corresponding symptoms (here deemed ‘naïve’ trees).

Trees sampled at the plantation were approx. 2 m tall and pruned regularly into a pyramid shape to get maximal sunlight and increase foliage production rates. The sampling area was composed of nine rows of trees with the treated and untreated trees separated by buffer rows (Fig. 1a). Plants at the botanical garden were not managed and varied in their habits. In general, they were approximately 2–3 times taller, produced larger, tougher leaves, and were planted in clusters instead of rows. Including these plants from a different location (and provenance, most probably) added an unknown degree of variation to the spectral reflectance measured in this study. However, it gave a valuable comparison, providing spectral signatures from the same species that were not influenced by myrtle rust.

Spectral measurements

Spectral reflectance between 350 and 2500 nm was measured using a portable, non-imaging spectroradiometer (Spectral Evolution PSR+ 3500) with spectral resolutions of 3 nm up to 700 nm, of 8 nm up to 1500 nm, and of 6 nm up to 2100 nm. The field spectrometer was set to 15 internal repetitions, meaning that each spectrum was measured 15 times, in order to reduce measurement variability. A leaf clip holder with a 3 mm sample area, a built-in reflectance standard and a separate 5 W light source (ILM-105) was used to take measurements, while also keeping heat from the light source away from the plant tissue. Heat stress could complicate the selection of relevant wavebands by causing physiological, biochemical or molecular changes in plants that would be represented in spectral reflectance responses and mix with the stress signal caused by myrtle rust infections.

At the plantation, from each of the two untreated and treated rows of trees, leaves from five trees were selected to record spectra at three sampling points: 180, 100 and 50 cm height (Fig. 1b). Plants that appeared disturbed by close proximity to frequently used management trails were avoided. For each sampling point, one terminal shoot was selected and the first two pairs (four leaves) of newly expanding leaves, just large enough to apply the leaf clip accessory, were used to record the spectral responses. The height-stratified sampling points were chosen on both east- and west-facing sides of trees, in order to represent the spectral response of a single tree most effectively. This design resulted in 240 spectra from 240 leaves for each class and 480 spectra in total at the plantation.

Spectra at the botanical garden were collected following the same general procedure as on the plantation (i.e. height stratification; east- and west-sampling). Ten naïve plants located in the Myrtaceae beds of the garden were selected. In total, 240 spectra of 240 naïve leaves were sampled, resulting in an overall dataset of 720 spectra.

Analysis pipeline

After data collection, all analyses were conducted using the R statistical platform (R Core Team, 2016) using several add-on packages (detailed below). For transparency and reproducibility, the full analysis, including figures and tables, can be repeated using code and data archived at https://github.com/ReneHeim/MyrtleRust-LemonMyrtle-Classification (https://doi.org/10.5281/zenodo.1142944).

Preprocessing of spectral reflectance signatures

First, all wavelengths below 500 nm were deleted because they contained intense spectral noise. Detection and removal of outlying spectra followed, using depth measures included in the FUNCTIONAL DATA ANALYSIS AND UTILITIES (FDA) package (Febrero-Bande & Oviedo de la Fuente, 2012). After the outliers had been removed the final dataset consisted of 216 observations for the naïve class, 236 for the treated class and 228 for the untreated class (from the original 240 spectra per class).

Spectral resampling was used to reduce multicollinearity between predictor variables. This reduced the spectral resolution from 3–8 nm (2151 predictor variables) to a resolution of 10 nm (202 predictor variables). Spectral resampling was carried out using the PROSPECTR package (Stevens & Ramirez-Lopez, 2014).

Finally, first-order derivatives (FOD) of each spectral signature were calculated. FOD transformations of the spectral curve are a commonly applied technique used to increase classification quality by enhancing spectral features and minimizing random noise (Demetriades-Shah *et al.*, 1990).

Random forest classification

An ensemble machine learning method was used to assign each spectrum to one of the three classes (i.e. naïve, treated, untreated). Ensemble methods reduce variance by providing an

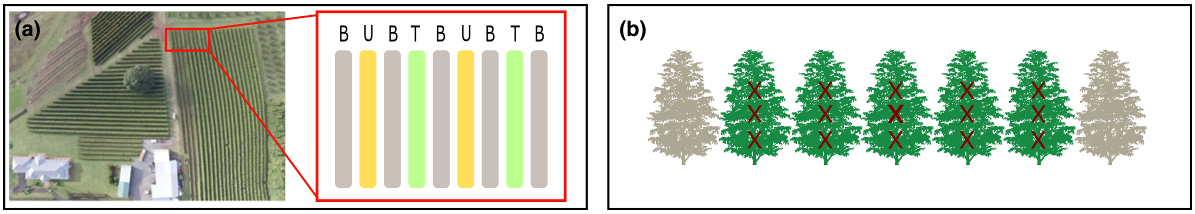

Figure 1 Sampling design. (a) An aerial image of the plantation site; the box highlights nine rows of *Backhousia citriodora* trees that had been exposed to *Austropuccinia psidii*. Spectra were collected at three sampling points from both sides of each tree in rows that had been untreated (U) or treated (T) with fungicide, separated by buffer trees (B). (b) One side of a single row with sampling points highlighted as crosses. At each sampling point, the first two pairs of freshly expanding leaves, just large enough to apply the leaf clip accessory, were used to record the spectra. Plants growing close to management trails (grey) were not used in this study.

outcome that is based on multiple independent classifiers. Here, a random forest classifier (Breiman, 2001) was used. This approach is based on multiple decision trees (Hastie *et al.*, 2009) and is nonparametric, because high-resolution spectral data rarely meet the criteria for standard parametric tests. Several studies, including the original paper by Breiman (2001), have shown that random forest classifiers are a suitable tool for analysing spectral and other high-dimensional, multicollinear data (e.g. Immitzer *et al.*, 2012).

For classification, the CARET package (Kuhn *et al.*, 2017) was used. Two parameters are primarily responsible for the performance of a random forest classifier and must be tuned depending on the dataset to be classified. First, the number of randomly selected predictors to choose from at each split (*mtry*) was optimized. Secondly, the number of trees generated to gain a full ensemble (*n*-tree) was optimized. The dataset was split 80:20 into training and test data subsets, and a 10-fold repeated cross-validation was applied on the training data. This approach breaks the data into 10 equal-sized fractions: nine of them are used to build/train a tree, and then used to predict the values of the 10th fraction, allowing the user to estimate the training accuracy. This process was repeated 100 times and the mean accuracy over these repetitions was calculated.

By default, the accuracy of the training and validation process was evaluated using the overall accuracy (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, an additional metric, the kappa statistic (Cohen, 1960) is useful. Kappa can take on values between −1 and 1; a value of 0 implies no agreement between the observed and predicted classes, while kappa of 1 indicates perfect concordance of the model prediction and the observed classes. Landis & Koch (1977) first defined the following standards for the strength of agreement: kappa of 0 = poor; kappa 0.01–0.20 = slight; 0.21–0.40 = fair; 0.41–0.60 = moderate; 0.61–0.80 = substantial; and 0.81–1 = almost perfect. In addition to kappa and OA, two further metrics were used that can indicate class-specific errors, the producer accuracy (PA) and user accuracy (UA; Story & Congalton, 1986). 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. UA is often termed to be a measure of reliability, which can be also interpreted as the agreement between repeated measurements within a class (Jones & Vaughan, 2010).

Finally, all three classes, naïve (216 observations), treated (236 observation) and untreated (228 observations) were classified based on 202 predictor variables (wavebands). The final model parameters were tuned to *mtry* = 52 and *n*-tree = 2000 after the best classifier was identified using the training data. Eventually, the test data (20% = 135 spectra) were used to validate the classifier, using kappa, OA, PA and UA as accuracy indices.

Waveband selection

Spectral datasets often contain thousands of predictor variables (wavebands). Using all available wavebands at a time to make a prediction is computationally intensive and problems with multicollinearity are very likely. Waveband selection techniques reduce the predictor space and provide a reduced set of wavebands that can be used in the same efficiency to predict the response variable. Here, a first set of the most important wavebands to classify naïve, untreated and treated trees was identified to provide future studies a starting point for validation or further classification tests. While including the naïve class in the waveband selection might be valuable to distinguish truly healthy trees from infected ones, the waveband selection derived only for the classes untreated and treated may be more relevant for detection systems applied on plantations as naïve plants are unlikely to occur there. Waveband selection was performed using the VSURF package in R (Genuer *et al.*, 2015).

Results

Random forest classification

In order to investigate whether it was possible to spectrally discriminate naïve, treated and untreated lemon myrtle trees, 135 primary spectra (i.e. 20% of the dataset) were analysed. The random forest classifier internally compared the prediction to the known class information of each spectral group, achieving a substantial prediction accuracy according to the scheme of Landis & Koch (1977): kappa = 0.68, OA = 79%. The procedure was repeated using 135 FOD spectra, yielding markedly improved accuracy (kappa = 0.92, OA = 95%). According to Landis & Koch (1977) these accuracies can be considered almost perfect.

When evaluating the accuracy assessment in greater detail (Table 1) it was found that the naïve (N) and treated (T) spectral responses received good PA values (N = 79.1%, T = 87.2%) and UA values (N = 97.1%, T = 75.9%). By contrast, spectral response from untreated (U) trees received a slightly lower PA (U = 68.9%) and UA (U = 67.4%), meaning that this class-specific prediction was less accurate. For the FOD-transformed spectra, all three class-specific accuracies were excellent (UA: N = 100%, T = 90.0%, U = 95.3%; PA: N = 97.7%, T = 95.7%, U = 91.1%).

Important wavebands for this classification

Many of the features useful for discriminating all three classes (Table 2a, Fig. 2a,c) were within the shortwave infrared region (SWIR; 1300–2500 nm), and this was true for analyses based on primary spectra or on FOD spectra. Considering other spectral regions, the visible region (VIS; 400–700 nm) was more useful for discriminating primary spectra, and the near-infrared region (NIR; 700–1300 nm) was more useful for discriminating FOD spectra. Figure 2b, d and Table 2b highlight features that were selected when comparing spectra collected only at the plantation. By examining the NIR region (Fig. 2a) and comparing the average spectral signature for naïve and treated, it can be observed that they are more similar to each other than compared to untreated.

Table 1 Assessment of classification accuracy using (a) primary spectra and (b) first-order derivative-transformed spectra

No. of samples	Reference Naïve	Treated	Untreated	Total	User accuracy (%)
(a) Primary spectra					
Prediction					
Naïve	34	0	1	35	97.1
Treated	0	41	13	54	75.9
Untreated	9	6	31	46	67.4
Total	43	47	45	135	
Producer accuracy (%)	79.1	87.2	68.9		78.5
(b) First-order derivative-transformed spectra					
Prediction					
Naïve	42	0	0	42	100.0
Treated	1	45	4	50	90.0
Untreated	0	2	41	43	95.3
Total	43	47	45	135	
Producer accuracy (%)	97.7	95.7	91.1		94.8

Diagonals represent correctly classified groups, off-diagonals were misclassified. The lower right cell contains the overall accuracy (no. of correct classified groups/total no. of groups (135)). User accuracy and producer accuracy are shown to provide class-specific accuracies.
Lemon myrtle trees were in the plantation, treated and untreated with fungicide against myrtle rust, and in the botanical garden unexposed to myrtle rust (naïve).

Discussion

In this study, spectral signatures were compared of lemon myrtle leaves from plants that had not been exposed to *A. psidii* (naïve), from plants that were infected and thus showing myrtle rust symptoms (untreated), and from plants that had been treated with fungicide and had no obvious symptoms of active myrtle rust (treated). Naïve trees were not available at the plantation, so trees from a botanical garden in a separate region were used. The spectral signatures of naïve trees were expected to be more similar to those from treated plants than to untreated plants, and indeed this was the case. High NIR (700–1300 nm) reflectance is generally an indicator for cellular integrity (Jensen, 2009). Thus, the similarly high NIR reflectance in naïve and treated plants in the present study suggest that *A. psidii* was not present (or at least had not caused damage) in fungicide-treated plants; if it had been present, it would have caused damage in mesophyll cells (Morin *et al.*, 2014) and would have been detected in the NIR region. All three groups were correctly classified, either using primary spectral signatures (kappa = 0.68, OA = 79%), or, with far higher accuracy, using FOD spectra (kappa = 0.92, OA = 95%).

It was not surprising that the FOD spectra performed better than primary spectral signatures, as FOD spectra are in general better at resolving overlapping wavebands and at reducing random noise (Demetriades-Shah *et al.*, 1990). Better performance of FOD than primary spectra was also reported by Mutanga *et al.* (2004) in a study focusing on biochemical indices of pasture quality in five grass species. However, this is not the case in every study. In a detailed investigation of spectral classification techniques, Ghiyamat *et al.* (2013) showed that FOD-based approaches showed least improvement (over primary spectra) in very complex datasets, and most improvement in less complex datasets, but also that some classification methods (e.g. Euclidean distance and Jeffreys–Matusita distance) typically performed better than others. Here, it was found that classification accuracy was increased when using a random forest classifier combined with FOD spectra. Classification accuracy using primary spectra and a random forest classifier could still be considered as substantial. Taken together with results from other studies, it seems that both the choice of classification method and the number of classes included in the classification influence whether FOD spectra improve classification accuracy.

In another step in the present study, the number of wavebands was reduced from 2151 to 202 to avoid effects of multicollinearity; nevertheless, high classification accuracies were still achieved. Therefore, it is reasonable to assume that, as in other spectral datasets, some wavebands contained redundant information (Thenkabail *et al.*, 2011). In addition, to avoid multicollinearity,

Table 2 Selected features for the spectra collected from myrtle trees at the plantation, treated and untreated with fungicide against myrtle rust, and from myrtle trees at the botanical garden unexposed to myrtle rust (naïve)

Spectral region	Primary spectra	FOD spectra
(a) Botanical garden and plantation (naïve, treated, untreated)		
VIS	555, 605, 695, 715	—
NIR	725, 735, 755	795, 815, 825, 915
SWIR	1405, 1415, 1425, 1435, 1895, 2025, 2035, 2085, 2095, 2115, 2145, 2165, 2175	1435, 1445, 1455, 1665, 1775, 1805, 1815, 2145, 2225, 2295
(b) Plantation only (treated, untreated)		
VIS	545, 555, 715	555, 625
NIR	725, 735, 745	795, 815, 845, 915
SWIR	1455, 1475, 1485, 2125, 2145, 2175	1645, 1655, 2145, 2225

FOD, first-order derivative; VIS, visible; NIR, near-infrared; SWIR, shortwave infrared.

reducing the number of wavebands can have multiple positive effects, for example (i) reducing long computation times, (ii) identifying critical wavebands specific to host and pathogen, (iii) using these wavebands as input parameters to improve classifiers, or (iv) to design disease-specific vegetation indices based on such a refined set of wavebands (Thenkabail *et al.*, 2011). For the pathosystem in the present investigation (lemon myrtle–myrtle rust), a preliminary set of wavebands was identified that had higher relevance over other wavebands used in this study.

Wavebands in the visible (VIS, 400–700 nm) and near-infrared (NIR, 700–1300 nm) regions were most important for distinguishing the three infection groups. The visible domain (556–660 nm) often corresponds to necrotic or chlorotic lesions, and a reduction in chlorophyll activity, while the red-edge (685–715 nm) can be used to detect general symptoms of plant stress (Delalieux *et al.*, 2007). Some wavebands in the shortwave infrared region (SWIR; 1300–2500 nm) were also found to be important. Variation in this region has been linked to changes in water content caused by air humidity or water loss from lesions (Delalieux *et al.*, 2007). In the present study, many lesions were observed on untreated (infected) leaves, very few, and probably old, on treated leaves, and none on naïve leaves. The few lesions found on the naïve leaves were probably caused by other biotic or abiotic factors. These observations provide a possible explanation why the waveband selection resulted in wavebands within the SWIR regions. Consequently,

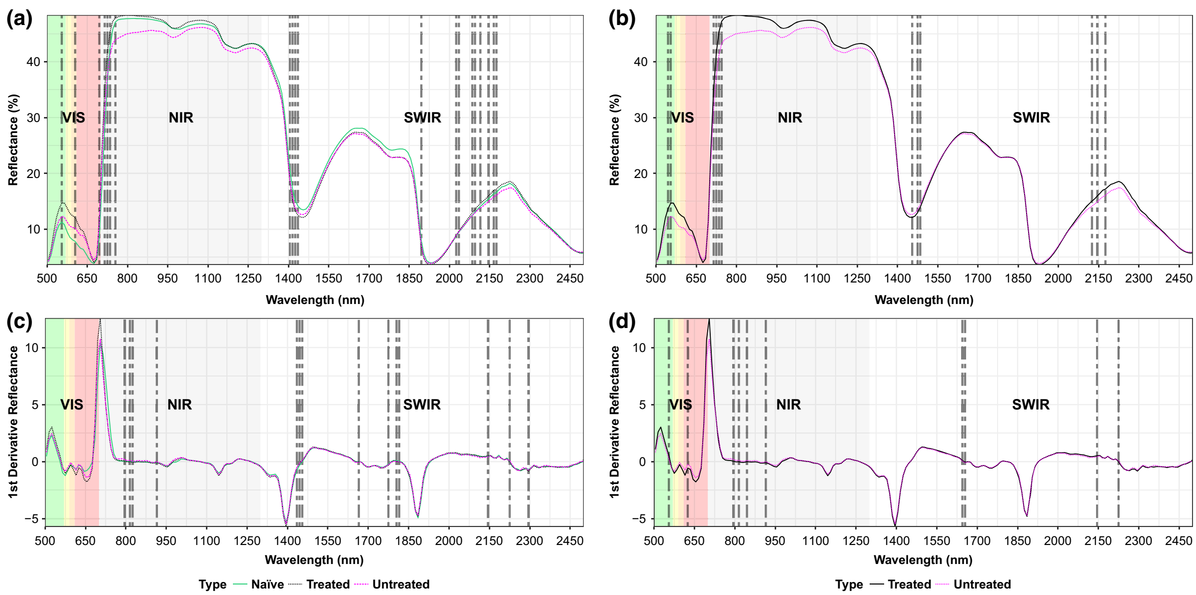

Figure 2 Spectral signatures and selected features for spectra from lemon myrtle trees collected at the plantation and botanical garden (a, c) and at the plantation only (b, d). Trees at the plantation were either treated or untreated with fungicide to eliminate myrtle rust, whereas trees at the botanical garden were untreated but free from infection (naïve). Important wavebands (grey dashed vertical lines) are presented for primary spectra (a, b) and their first-order derivatives (c, d). All plots emphasize a traditional subsetting of the electromagnetic spectrum to better assign the features to a specific region and support interpretation of each feature. VIS, visible; NIR, near-infrared; SWIR, shortwave infrared.

results of this study echoed those of Delalieux *et al.* (2007), not only in the high classification accuracies (80%) but also in finding similar wavebands important for plant disease detection. The refined set of wavebands listed in the present study are a first indication of which wavebands might be unique for this pathosystem but, without further refinement, they cannot be confidently distinguished from general indicators of stress caused by fungal pathogens.

There is increasing demand in precision agriculture for differentiation between pathogen-related effects and other stress-inducing factors (Mahlein, 2016). For example, it may be necessary to refine spectral sets of data down to some level where single wavebands can be considered unique for the system under investigation. Amongst the important regions found in the present study, that could be related to those found by Delalieux *et al.* (2007), the wavebands 545, 555, 625, 745, 755 and 845 nm were also considered by Bravo *et al.* (2003) to be important for successfully classifying spectral signatures from healthy and yellow rust (*P. striiformis*) infected wheat plants. Bravo *et al.* (2003) were limited to using wavebands between 460 and 900 nm but could still achieve classification accuracies of 96% using four wavebands only (543, 630, 750 and 861 nm). As yellow rust is a closely related family to *A. psidii* and the wavebands found in the present investigation are in close proximity to those found by Bravo *et al.* (2003), it seems reasonable to suggest that some of the selected wavebands might be unique for the lemon myrtle–myrtle rust pathosystem. Further investigation would be needed to confirm or disprove this suggestion.

The successful discrimination between spectral signatures is only the first step towards using spectral approaches to detect and monitor myrtle rust in the lemon myrtle industry. A promising next step would be the development of a disease-specific vegetation index (Mahlein *et al.*, 2013). Such indices allow land managers a straightforward disease assessment by indicating, for example, the level of infection. In theory, a disease-specific spectral index for lemon myrtle–myrtle rust pathosystem could be developed by refining the provided set of wavebands. Sensor systems based on those specifications could be built and used in a plantation setting, mounted on terrestrial or aerial vehicles, to detect infection hotspots and enable targeted fungicide application. This would reduce the costs spent on fungicides, the human-caused interassessor bias and also damage caused by fungicides on vegetation in close proximity to the crop of interest.

Austropuccinia psidii, has not previously been subject to investigations using spectral sensor systems. The results of the present study represent a proof-of-concept for incorporating a spectral approach into a precision farming tool used for the lemon myrtle industry as well as other industries. The establishment of spectral libraries of specific plant–pathogen interactions could enable land managers to detect pathogens before symptoms are visible to the naked eye, accurately track the spread of infection, objectively quantify disease severity and differentiate pathogen-related effects from other stress-inducing factors.

Acknowledgements

The authors thank Gary Mazzorana from Australian Rainforest Products for access to his plantation, and the Australian Botanic Garden at Mount Annan for access to their gardens. Anne-Katrin Mahlein, Pablo J. Zarco-Tejada and Andrew Allen gave invaluable advice on various aspects of this study. R.H.J.H. was supported by a Macquarie University Research Excellence Scholarship. Special thanks are given to Michelle Leishman for suggesting the application of a spectral approach to investigating myrtle rust. The authors declare no conflict of interest.

References

Alfenas AC, Zauza EAV, Assis TF, 2003. First record of *Puccinia psidii* on *Eucalyptus globulus* and *E. viminalis* in Brazil. *Australasian Plant Pathology* **32**, 325–6.

Australian Government Bureau of Meteorology, 2017. Climate data online. [http://www.bom.gov.au/climate/data/index.shtml?bookmark=200]. Accessed 17 August 2017.

Berthon K, Esperon-Rodrigues M, Beaumont L, Carnegie AJ, Leishman L, 2018. Assessment and prioritisation of plant species at risk from myrtle rust (*Austropuccinia psidii*) under current and future climates in Australia. *Biological Conservation* **218**, 247–53.

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

Bravo C, Moshou D, West J, McCartney A, Ramon H, 2003. Early disease detection in wheat fields using spectral reflectance. *Biosystems Engineering* **84**, 137–45.

Breiman L, 2001. Random forests. *Machine Learning* **45**, 5–32.

Carnegie AJ, Lidbetter JR, Walker J *et al.*, 2010. *Uredo rangelii*, a taxon in the guava rust complex, newly recorded on Myrtaceae in Australia. *Australasian Plant Pathology* **39**, 463–6.

Carnegie AJ, Kathuria A, Pegg GS, Entwistle P, Nagel M, Giblin FR, 2016. Impact of the invasive rust *Puccinia psidii* (myrtle rust) on native Myrtaceae in natural ecosystems in Australia. *Biological Invasions* **18**, 127–44.

Cohen J, 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* **20**, 37–46.

Coutinho TA, Wingfield MJ, Alfenas AC, Crous PW, 1998. Eucalyptus rust: a disease with the potential for serious international implications. *Plant Disease* **82**, 819–25.

Delalieux S, van Aardt J, Keulemans W, Coppin P, 2007. Detection of biotic stress (*Venturia inaequalis*) in apple trees using hyperspectral data: non-parametric statistical approaches and physiological implications. *European Journal of Agronomy* **27**, 130–43.

Demetriades-Shah TH, Steven MD, Clark JA, 1990. High resolution derivative spectra in remote sensing. *Remote Sensing of Environment* **33**, 55–64.

Doran J, Lea D, Bush D, 2012. *Assessing Myrtle Rust in a Lemon Myrtle Provenance Trial*. Rural Industries Research and Development Corporation, Barton, Australia.

Febrero-Bande M, Oviedo de la Fuente M, 2012. Statistical computing in functional data analysis: The R package FDA.USC. *Journal of Statistical Software* **51**, 1–28.

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.

Ghiyamat A, Shafri HZM, Amouzad Mahdiraji G, Shariff ARM, Mansor S, 2013. Hyperspectral discrimination of tree species with different classifications using single- and multiple-endmember. *International Journal of Applied Earth Observation and Geoinformation* **23**, 177–91.

Giblin F, 2013. *Myrtle Rust Report: New Caledonia*. University of the Sunshine Coast, Maroochydore.

Hastie T, Tibshirani R, Friedman J, 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd edn. Berlin, Germany: Springer Science & Business Media.

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

Immitzer M, Atzberger C, Koukal T, 2012. Tree species classification with random forest using very high spatial resolution 8-band Worldview-2 satellite data. *Remote Sensing* **4**, 2661–93.

Jensen JR, 2009. *Remote Sensing of the Environment: An Earth Resource Perspective*. 2nd edn. New Delhi, India: Pearson Education India.

Jones HG, Vaughan RA, 2010. *Remote Sensing of Vegetation: Principles, Techniques, and Applications*. Oxford, UK: Oxford University Press.

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

Kuhn M, Wing J, Weston S *et al.*, 2017. CARET: Classification and regression training. R package version 6.0-76. [https://CRAN.R-project.org/package=caret.]. Accessed 9 January 2018.

Landis JR, Koch GG, 1977. The measurement of observer agreement for categorical data. *Biometrics* **33**, 159–74.

Loope L, 2010. *A Summary of Information on the Rust* Puccinia psidii *Winter (Guava Rust) with Emphasis on Means to Prevent Introduction of Additional Strains to Hawaii*. Open-File Report 2010–1082. Reston, Virginia: US Geological Survey.

Machado P, Alfenas AC, Alfenas RF, Mohammed CL, Glen M, 2015. Microsatellite analysis indicates that *Puccinia psidii* in Australia is mutating but not recombining. *Australasian Plant Pathology* **44**, 455–62.

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

Mahlein AK, Steiner U, Dehne H-W, Oerke E-C, 2010. Spectral signatures of sugar beet leaves for the detection and differentiation of diseases. *Precision Agriculture* **11**, 413–31.

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.

Makinson B, 2014. Myrtle rust – what's happening? *Australasian Plant Conservation* **23**, 13–5.

McTaggart AR, Roux J, Granados GM *et al.*, 2016. Rust (*Puccinia psidii*) recorded in Indonesia poses a threat to forests and forestry in South-East Asia. *Australasian Plant Pathology* **45**, 83–9.

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

Mutanga O, Skidmore A, Prins HH, 2004. Predicting *in situ* pasture quality in the Kruger National Park, South Africa, using continuum-removed absorption features. *Remote Sensing of Environment* **89**, 393–408.

NSW Scientific Committee, 2011. *Introduction and Establishment of Exotic Rust Fungi of the Order Pucciniales Pathogenic on Plants of the Family Myrtaceae – Key Threatening Process Listing*. Sydney, Australia: NSW Scientific Committee.

Pegg GS, Brawner JT, Lee DJ, 2014. Screening *Corymbia* populations for resistance to *Puccinia psidii*. *Plant Pathology* **63**, 425–36.

du Plessis E, McTaggart AR, Granados GM *et al.*, 2017. First report of myrtle rust caused by *Austropuccinia psidii* on *Rhodomyrtus tomentosa* (Myrtaceae) from Singapore. *Plant Disease* **101**, 1676.

Potts BM, Sandhu KS, Wardlaw T *et al.*, 2016. Evolutionary history shapes the susceptibility of an island tree flora to an exotic pathogen. *Forest Ecology and Management* **368**, 183–93.

R Core Team, 2016. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Roux J, Greyling I, Coutinho TA, Verleur M, Wingfield MJ, 2013. The myrtle rust pathogen, *Puccinia psidii*, discovered in Africa. *IMA Fungus* **4**, 155–9.

Rural Industries Research and Development Corporation, 2012. *Lemon Myrtle*. RIRDC Publication No. 12/066. Canberra, Australia: Australian Native Food Industry Stocktake, RIRDC, 5–7.

Soewarto J, Carriconde F, Hugot N, Bocs S, Hamelin C, Maggia L, 2017. Impact of *Austropuccinia psidii* in New Caledonia, a biodiversity hotspot. *Forest Pathology*. https://doi.org/10.1111/efp.12402.

Stevens A, Ramirez-Lopez L, 2014. An introduction to the PROSPECTR package. R Package Vignette, Report No. 3, R Package version 0.1. [https://cran.r-project.org/web/packages/prospectr/vignettes/prospectr-intro.pdf.] Accessed 9 January 2018.

Story M, Congalton R, 1986. Accuracy assessment – a user's perspective. *Photogrammetric Engineering and Remote Sensing* **52**, 397–9.

Thenkabail PS, Lyon GJ, Huete A, 2011. *Hyperspectral Remote Sensing of Vegetation*. Boca Raton, FL, USA: CRC Press.