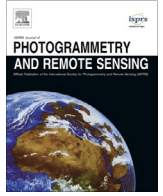




Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Assessing very high resolution UAV imagery for monitoring forest health during a simulated disease outbreak

Jonathan P. Dash^{a,*}, Michael S. Watt^b, Grant D. Pearse^a, Marie Heaphy^a, Heidi S. Dungey^a^a Scion, 49 Sala Street, Rotorua, New Zealand^b Scion, 10 Kyle St, Riccarton, Christchurch, New Zealand

ARTICLE INFO

Article history:

Received 4 May 2017

Received in revised form 10 July 2017

Accepted 15 July 2017

Available online 25 July 2017

Keywords:

UAV

Plantation forest

Multi-spectral

RPAS

Tree health

Physiological stress

ABSTRACT

Research into remote sensing tools for monitoring physiological stress caused by biotic and abiotic factors is critical for maintaining healthy and highly-productive plantation forests. Significant research has focussed on assessing forest health using remotely sensed data from satellites and manned aircraft. Unmanned aerial vehicles (UAVs) may provide new tools for improved forest health monitoring by providing data with very high temporal and spatial resolutions. These platforms also pose unique challenges and methods for health assessments must be validated before use. In this research, we simulated a disease outbreak in mature *Pinus radiata* D. Don trees using targeted application of herbicide. The objective was to acquire a time-series simulated disease expression dataset to develop methods for monitoring physiological stress from a UAV platform. Time-series multi-spectral imagery was acquired using a UAV flown over a trial at regular intervals. Traditional field-based health assessments of crown health (density) and needle health (discolouration) were carried out simultaneously by experienced forest health experts. Our results showed that multi-spectral imagery collected from a UAV is useful for identifying physiological stress in mature plantation trees even during the early stages of tree stress. We found that physiological stress could be detected earliest in data from the red edge and near infra-red bands. In contrast to previous findings, red edge data did not offer earlier detection of physiological stress than the near infra-red data. A non-parametric approach was used to model physiological stress based on spectral indices and was found to provide good classification accuracy (weighted kappa = 0.694). This model can be used to map physiological stress based on high-resolution multi-spectral data.

© 2017 Scion (New Zealand Forest Research Institute). Published by Elsevier B.V. on behalf of International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS). This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Forest plantations provide a wide range of economic, social and environmental benefits in many parts of the world (Yao et al., 2014). In New Zealand, forestry exports provide the third largest source of export earnings, worth over NZ\$5 billion per annum and forest industry directly employs 18,000 people (NZFOA, 2013). New Zealand's commercial plantation forest sector is dominated by the fast-growing conifer species *Pinus radiata* D. Don (*P. radiata*) which occupies 90% of the total plantation area (Watt et al., 2017). Such reliance on a single species means that the forest industry is particularly vulnerable to biosecurity incursions by plant pests and pathogens that pose a threat to *P. radiata*. Despite being an island nation with rigorous biosecurity protocols and

monitoring programmes, incursions remain a serious threat with recent introductions estimated to have cost New Zealand's primary sector upwards of NZ\$ 400 million (Hulme, 2014). In the context of forestry, *P. radiata* in New Zealand is currently affected by a number of pathogens including *Dothistroma septosporum*, *Cyclaneusma minus* and *Phytophthora pluvialis*. Infection by these organisms results in a significant loss in forest productivity. It is estimated that *Dothistroma* Needle Blight alone resulted in a cost of NZ \$19.8 million per year to the New Zealand economy throughout the 2000s (Watt et al., 2011).

Forest health surveillance forms a significant part of biosecurity monitoring and effective pathogen management efforts (Bulman et al., 2016). Surveillance methods are based on the detection of infection symptoms; these are specific to each pathogen but for plantation trees these commonly include foliar discolouration followed by some degree of defoliation. Traditionally, regular ground-based surveys carried out by highly trained field technicians have

* Corresponding author.

E-mail address: jonathan.dash@scionresearch.com (J.P. Dash).

formed the basis of monitoring efforts in many countries (Brown and Webber, 2008; Bulman et al., 2004; Smith et al., 2008). These surveys are commonly carried out on an annual basis (Coops et al., 2006) and are used both to identify new incursions and to plan control or damage mitigation measures (Bulman et al., 2016). Furthermore, information from these surveys contributes towards policy development, international reporting obligations, forest planning and investment decision making (Stone and Coops, 2004). The utility of traditional survey techniques are limited by the small spatial coverage that can be achieved and the inevitable subjectivity between assessors conducting ground surveys.

Modern remote sensing tools have the potential to complement ground-based surveys by expanding spatial coverage and offering objective assessments of tree and forest health. To realise these benefits it is important to correctly identify the survey techniques and data sources required to meet the surveillance objectives (Wulder et al., 2006a). An appropriate sensor choice and resolution can be defined by the desired sensitivity of the surveillance and the physiological impacts of the stressor on observable properties such as the spectral characteristics of foliage. With an appropriate sensor choice and resolution defined, the increased area that can be covered and assessed using remote sensing is typically large and complete coverage may even be achievable, greatly improving the accuracy and completeness of surveillance (Wulder et al., 2006b). Aerial survey (also referred to as aerial sketch mapping), which involves manual identification of disease outbreaks by a skilled observer on-board an aircraft, is a commonly used, affordable and flexible approach for large-scale mapping of disease outbreaks. This method provides an accurate means of detection of a wide-range of forest health symptoms but offers limited ability to resolve different levels of physiological damage (Stone et al., 2013) and often lacks spatial specificity for damage that occurs at a fine scale over a relatively narrow time-frame (Johnson and Ross, 2008).

The acquisition of digital spatial data has many, often complementary, advantages compared with manual aerial surveys, including a high level of consistency, spatial accuracy and automation of subsequent analysis. Most research has used satellite imagery to quantify the spatial extent of insect outbreaks and tree mortality at regional and landscape scales using imagery of a moderate (5–30 m) resolution (Meigs et al., 2011; Fraser and Latifovic, 2005). Over the last decade a growing number of studies have used high resolution (<5 m) satellite imagery to characterise tree mortality at finer spatial scales on individual trees, or clusters of trees, within a stand (Stone et al., 2012; Dennison et al., 2010; Hicke and Logan, 2009; Guo et al., 2007; Coops et al., 2006). Generally, the use of satellite imagery with a finer spatial resolution (Coops et al., 2006; Hicke and Logan, 2009; White et al., 2005; Wulder et al., 2008) has been shown to more accurately classify mortality, resulting from insect outbreaks, than imagery with medium resolution (Franklin et al., 2003; Skakun et al., 2003; Wulder et al., 2006).

Pest and disease detection remains a key target for remote sensing technologies. To date, only a few studies have demonstrated successful disease detection in forests using high resolution multi-spectral imagery from aircraft (Leckie et al., 2004), satellites (Poona and Ismail, 2013) and hyper-spectral imagery from manned aerial platforms (Coops et al., 2003; Calderón et al., 2015; Pu et al., 2008). This is in contrast to agricultural research where a range of studies have examined the use of remote sensing data for detection of foliar pathogens in annual agricultural crops. This may be because disease symptoms in these crops are often expressed in the upper parts of the plant and agricultural cropping systems are structurally simple and small in scale by comparison to forests (Barton, 2012; West et al., 2003; Sankaran et al., 2010).

Change detection techniques provide a useful starting point for the identification of subtle changes in forest health based on spec-

tral information. The aggregation of multi-temporal composite images can greatly improve the signal to noise ratio (Rullan-Silva et al., 2013) and using longer time periods can mitigate the detrimental impacts of environmental factors, such as cloud cover, on detection. Using this approach, acceptable accuracies have been demonstrated for detection of bark beetle attack on conifer species (Garrity et al., 2013; Meddens and Hicke, 2014; Havašová et al., 2015; Goodwin et al., 2008) and needle discoloration resulting from *Dothistroma pini* (Coops et al., 2003). Analysis of time-series data was also used by Eitel et al. (2011) who girdled a number of trees in a piñon-juniper woodland and then successfully detected tree stress using a dense sequence of multi-spectral images from the RapidEye satellite constellation (Eitel et al., 2011). Invoking stress symptoms in this manner provides superior experimental control and is well suited to testing and calibration of methods for the early detection of symptoms such as pathogen induced physiological changes in foliage.

Imagery from manned and satellite platforms provides coverage over large areas but is typically time consuming and relatively costly to acquire on a regular basis. This makes data from these platforms poorly suited as means for early detection of outbreaks, near-continuous monitoring of high-risk sites such as those frequently accessed for public recreation, or for identification of small, isolated outbreaks that could easily be missed in medium resolution imagery. The development of unmanned aerial vehicles (UAVs) may offer new platforms for the collection of very high resolution imagery, while also offering the ability to collect data at short intervals in a cost-effective manner.

Despite their potential advantages, studies using UAVs to detect biotic damage in forests are scarce. Hyper-spectral data were acquired from a UAV over a stand of Norway spruce (*Picea abies* L. Karst.) infested with the European spruce bark beetle (*Ips typographus* L.) (Näsi et al., 2015). Using this data, Näsi et al. (2015) were able to classify individual trees into classes of healthy, infested and dead with a reasonable accuracy (Cohen's kappa = 0.6). Lehmann et al. (2015) used a UAV equipped with compact digital camera to characterise defoliation of oak trees by the oak splendor beetle (*Agrilus biguttatus*). A modified normalized difference vegetation index (mNDVI) derived classification was used to distinguish between five vegetation health classes with Kappa index of agreement ranging from 0.77 to 0.81 for the two study sites (Lehmann et al., 2015).

In this study, disease symptoms in a stand of *P. radiata* were simulated through careful application of herbicide to groups of trees of varying sizes. A UAV equipped with a multi-spectral sensor was used to regularly monitor changes in needle colour prior to treatment and for three months after herbicide application with the aims of (i) developing suitable data collection methods and processing procedures for UAV data acquisition and analysis to detect physiological stress, (ii) to test the sensitivity of specific spectral indices to provide an indication of the early onset of physiological stress and (iii) to determine the optimal spatial resolution of imagery for detection of a simulated disease outbreak across tree clusters of varying size.

2. Materials and methods

2.1. Study site

The study site was located in Kinleith Forest in New Zealand's Central North Island (Fig. 1) (latitude 38° 24'18.74S, longitude 176° 0'59.28E), approximately 28 km southeast of the township of Tokoroa. The site is around 230 m above sea level and slopes gradually up towards the south-eastern corner of the trial. The site experiences a temperate climate (total annual rainfall = 1238 mm, mean annual temperature = 13.4 °C) and the soils are loam

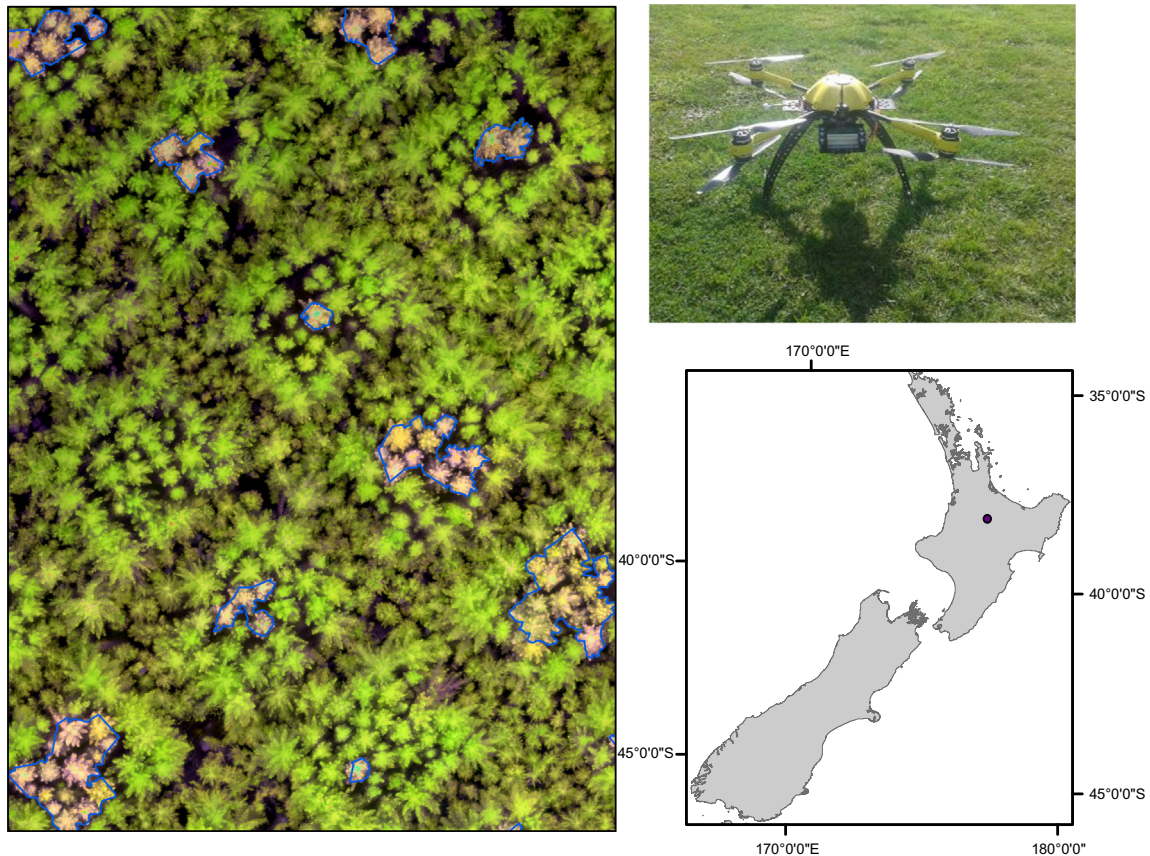


Fig. 1. Location of the study site within New Zealand and an image of the coaxial quad-copter UAV used during data collection. The main panel shows an image of the multi-spectral data collected with treated trees evident and outlined in blue (image date = 03-03-2016). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

belonging to the Kairanga series (Hewitt, 2010). The trial area of interest was 2.7 ha and was located within a 7.3 ha stand of *P. radiata* planted in July 1993. Site preparation was via V-blading and the previous land use at the site was also a forest plantation. Planting stand density was 635 sph ($4.5 \text{ m} \times 3.5 \text{ m}$), no thinning was undertaken and trees that suffered mortality early in the rotation were replaced. The site was organised into 35 plots as part of a fungicide spray trial into genetic disease resistance in immature trees. The fungicide spray trial was completed before age 6 (1999). At the initiation of this study a comprehensive survey of the site revealed that there was no evidence of disease expression and trees were in good general health. Study trees were mature and large (mean diameter at breast height at 1.4 m above ground = 462 mm, mean top height = 43.5 m) during data collection and were harvested soon after the research concluded.

2.2. Experimental treatments

The original field plots were used to distribute experimental treatments throughout the study area. These were used because they were clearly demarcated and accurately mapped making access and navigation simple. Five plots were randomly assigned to each of 6 treatments (Table 1) and the remaining 5 plots were excluded from the trial. All plots were rectangular and contained 35 trees in seven rows planted with a spacing of $3.5 \times 4.5 \text{ m}$. Trees that were dead or severely suppressed were not included in the trial. The experimental treatment defined the size of the cluster (groups of treated trees) that were poisoned to cause physiological stress and simulate the onset of colour change and changes in needle retention as might be expected during a disease outbreak. The

Table 1

Experimental treatments used in the study. Cluster size refers to the number of trees subjected to herbicide application in the treatment.

Treatment	Description	Cluster size
0	Control no trees poisoned	0
1	Single tree closest to plot centre is poisoned	1
2	Two trees closest to the plot centre poisoned	2
3	Four trees closest to the plot centre poisoned	4
4	Eight trees closest to the plot centre poisoned	8
5	Sixteen trees closest to the plot centre poisoned	16

objective of using different sized clusters was to investigate the utility of different remotely sensed datasets for detecting different levels of physiological stress as might be expected from symptom expression from certain diseases. Subject trees were treated with Metsulfuron methyl in the form of water dispersible granules at a rate of 200 g/l. This mixture was injected through holes drilled into the tree stem. Holes were drilled 25 cm apart around the circumference of each stem and 15 ml of herbicide was deposited into each hole. In total 13.32 l of Metsulfuron mixture were applied to the study site. This herbicide application technique was selected as it allowed for highly targeted application and because expert advice indicated it would invoke changes in foliar colour and retention over the duration of the study. Herbicide application was completed for all trees in a single day on the 3rd December 2015.

2.3. Tree health data

A ground reference dataset was recorded on all study trees. Tree canopies were assessed for needle discolouration and canopy den-

sity from the ground by an experienced assessor who is regularly engaged in routine forest health monitoring. Assessors undergo regular training, auditing, and calibration and carry photographic reference material to improve consistency. To reduce the variance associated with assessment subjectivity the same expert assessor was used for all assessments for the duration of the trial. Each tree was assessed from two separate angles, approximately 180° from each other, so that a comprehensive assessment of the tree canopy could be made. Assessment locations were permanently marked with paint on the ground and trees were assessed from the same positions, and at approximately the same time of day, each time. Wherever possible, ground assessments were made on the same day remotely sensed data were collected. A commonly used approach to classifying stress symptom expression using data collected in the field is to categorise discolouration scores into symptom severity classes. The classes used, based on the forest manager's standard procedures, were: no symptom expression (None = 0–10% canopy discolouration), slight symptom expression (Slight = 11–40% canopy discolouration), moderate symptom expression (Moderate = 41–80% canopy discolouration) and severe symptom expression (Severe = 80–100% canopy discolouration).

2.4. Remotely sensed data collection

A coaxial quad-copter UAV was used for all data collection. The craft was a modified version of the Aeronavics (www.aeronavics.co.nz) Skyjib heavy-lift airframe (Fig. 1). Lift was provided by eight kdc 5215 motors which were used to power 38 cm propellers. This craft had a flying time of around 15 min with a payload of approximately 5 kg and provided a stable platform even in strong and turbulent wind conditions. A single flight plan was used for all data collections to minimise variance associated with campaign settings in subsequent acquisitions. The flight plan used 18 flight lines to cover the area of interest and ensured that all craft manoeuvres and altitude adjustments occurred outside of the study area to minimise flight-related artefacts in the data collected. The UAV launch site was located in an adjacent elevated, recently harvested stand and afforded the pilots and ground control equipment uninterrupted view and communication with the craft.

A series of nine ground control points (GCPs) were established around the study area and fixed using a Trimble Geo7X GNSS (Trimble Navigation Ltd., Sunnyvale, CA, USA). These fixes were subsequently differentially corrected using a local base station network maintained by Land Information New Zealand (LINZ). The targets established at the ground control points had reflective properties that could be easily manually identified in the multi-spectral imagery. The GCPs were used to geo-rectify all imagery with a high degree of accuracy.

2.4.1. Multi-spectral imagery

Narrow band multi-spectral imagery was collected using a MicaSense RedEdge 3 camera (Micasense, Seattle, USA). The camera provides imagery in five narrow bands (blue = 455–495 nm, green = 540–580 nm, red = 658–678 nm, red edge = 707–727 nm and near-infrared = 800–880 nm) via five separate imaging sensors that operate nearly simultaneously. The camera has a lens focal length of 5.5 mm and field of view of 87.4°. The camera was housed in a gimbal to ensure nadir orientation during image acquisition. The flight plan ensured cross-track and along-track overlap of 85% and a calibrated reflectance panel was imaged directly before and after each flight and used for reflectance calibration using the empirical line method (Smith and Milton, 1999). The multi-spectral imagery was collected from a flying altitude of approximately 90 m above the local terrain and resulted in a ground sample distance (GSD) of 6 cm. Imagery acquisition was limited to

within two hours either side of local solar noon to minimise shadows in the final imagery.

Several spectral indices were chosen and calculated from each set of images based on a review of previous literature (Table 2). Spectral indices were selected for inclusion based on the spectral bands available in the multi-spectral dataset and evidence of correlation with plant physiological stress from previous research. Normalised difference vegetation index (NDVI) is a very well known and widely used ratio that has been shown many times to be correlated with plant condition, physiological stress and photosynthetic activity in a wide range of environments (Wang et al., 2004; Goetz et al., 2003; Verbesselt et al., 2009; Garcia-Ruiz et al., 2013; Cunningham et al., 2007). The green normalised difference vegetation index (GNDVI) and the red edge normalised difference vegetation index (RENDVI) were included specifically to examine the sensitivity of indices including the red edge and green bands and compare these directly to an index based on near-infrared band (NDVI). The non-linear index (NLI) was selected because some previous research findings suggest that the relationship between spectral indices and some tree bio-physical properties are non-linear (Goel and Qin, 1994). Therefore, it is possible that NLI might be valuable as an indicator of various stages of physiological stress as it expresses the non-linear relationships in a linear manner (see Table 3).

2.4.2. Multi-spectral data processing

Multi-spectral images were exported from the MicaSense RedEdge 3 in 12-bit RAW format. These images were mosaicked into a single, multi-band image covering the entire study area. The raw digital numbers (DNs) were converted into reflectance values using the calibrated reflectance panel. The relationship between DN and the natural logarithm of the image surface reflectance is linear and the y-intercept can be interpreted as the minimum surface reflectance that can be detected for each wavelength band (Smith and Milton, 1999). This can then be used to develop a regression equation to convert DN to reflectance values using a calibration equation (Furby and Campbell, 2001). Images were geo-rectified using the GCPs recorded within the study site and visible in the imagery.

Data collection spanned several months and so methods were needed to account for background changes in the target vegetation that were not caused by the experimental herbicide application. These changes include physiological response to the growing season, changes caused by climatic events and atmospheric effects that could not be controlled for using other methods. Without a suitable method for control these influences would confound the physiological stress signal induced by the application of the herbicide. Following the method of Healey et al. (2005), spectral indices were detrended to remove the extraneous effects. The approach was based on the assumption that changes in non-treated parts of the study site could be assumed to be caused by factors unrelated to the herbicide-induced stress. By characterising these unre-

Table 2
Spectral indices calculated from the multi-spectral imagery.

Spectral index	Equation	Source
Normalised Difference Vegetation Index (NDVI)	$\frac{(NIR-RED)}{(NIR+RED)}$	Rouse et al. (1974)
Green Normalised Difference Vegetation Index (GNDVI)	$\frac{(NIR-GREEN)}{(NIR+GREEN)}$	Gitelson and Merzlyak (1998)
Red Edge Normalised Difference Vegetation Index (RENDVI)	$\frac{(NIR-REDEGE)}{(NIR+REDEGE)}$	Gitelson and Merzlyak (1994) and Sims and Gamon (2002)
Non Linear Index (NLI)	$\frac{(NIR^2+RED)}{(NIR^2-RED)}$	Goel and Qin (1994)

Table 3

Overall classification accuracy expressed through Cohen's Kappa and associated 95% confidence interval (KappaCI), for all 13 models examined calculated using an independent 10% resampling of the training dataset. Predictor variables denoted with an * were included in the model.

Model Identifier	Predictor variables				Accuracy statistics	
	NDVI	GNDVI	RENDVI	NLI	Kappa	KappaCI
1	*	*	*	*	0.644	0.621–0.667
2	*	*	*		0.628	0.605–0.651
3	*	*		*	0.631	0.609–0.654
4	*		*	*	0.694	0.672–0.717
5		*	*	*	0.583	0.559–0.607
6			*	*	0.635	0.612–0.659
7		*	*		0.542	0.518–0.566
8	*	*			0.574	0.551–0.598
9	*			*	0.610	0.588–0.633
10	*				0.570	0.546–0.594
11		*			0.484	0.459–0.501
12			*		0.528	0.504–0.553
13				*	0.390	0.364–0.415

lated trends we could minimise the potential impact on the any signal from the stressed study trees (Eitel et al., 2011).

Spectral indices were detrended by identifying 25 control areas within the study area, based on the final multi-spectral image following review of the full time-series of mosaicked images to ensure that the control regions were representative. Care was taken to ensure that these areas did not overlap the experimental tree clusters. Control areas were 5 m squares and were carefully selected to cover conditions that were representative of the conditions in the study plots. The spectral indices extracted from the control areas were deemed to represent the average conditions of the study forest within a specific time step. Spectral indices extracted from within the experimentally treated areas were rescaled to the standard deviation above or below the scene's mean forest value for the spectral indices as extracted from the control areas (Healey et al., 2005) as follows:

$$S_i = (S - S_\mu) / S_\delta \quad (1)$$

where S is the original spectral index value, S_i is the rescaled spectral index, S_μ is the mean value for the control areas within a time step and S_δ is the standard deviation of values from the control areas. Using this method, the time-series spectral indices can be detrended in a robust and easy to implement manner provided that the control areas selected are representative of the average forest conditions within a time step (Healey et al., 2005).

2.5. Image resolution

The multi-spectral imagery collected using the UAV were re-sampled to coarser resolutions. This allowed us to examine the effect of image resolution on the sensitivity of detrended spectral indices to physiological stress. We chose to resample from the original GSD of 0.06 m to generate imagery with 0.3 m, 1 m, 3 m, and 5 m resolutions. Image resolution is closely related to acquisition costs and these resolutions were selected because they were similar to the specifications for several cost-effective imagery products routinely acquired for forest monitoring and management applications in New Zealand. This is important because new tools must remain cost-effective if they are to achieve uptake for forest monitoring applications. All image re-sampling was completed using the open-source gdalwarp function of the GDAL software package (GDAL Development Team, 2016).

2.5.1. Random forest

The ensemble decision tree classifier random forest (RF) uses bootstrap aggregated sampling (bagging) to construct many individual decision trees, from which a final class assignment is determined (Breiman, 2001). RF is now regularly applied to natural

resource assessment (Mellor et al., 2013) and has previously been used, in combination with remotely sensed data, to successfully model several variables of interest in this forest type (Dash et al., 2015, 2016; Watt et al., 2015, 2016). Decision trees are constructed using a sample from the available training data, with the remaining assigned as out-of-bag (OOB) samples. At each node, a random subset of predictor variables are tested to partition the observation data into increasingly homogeneous subsets. The node-splitting variable selected from the variable subset is that which resulted in the greatest increase in data purity (variance or Gini) before and after the tree node split (Cutler et al., 2007). This process ends when there are no further gains in purity. Response variables can be continuous, calculated by averaging, or categorical where predictions are derived from a model vote among decision trees. Computational efficiency of the algorithm is enhanced, compared with alternative approaches, as only a sample of variables are used at each node split. This also reduces correlation between trees, improving both predictive power and classification accuracy. The OOB sample data are used to compute accuracies and error rates, averaged over all predictions, and estimate variable importance (Cutler et al., 2007; Mellor et al., 2013). RF provides two methods to estimate the importance of each predictor variable in the model. The mean decrease in accuracy (MDA) importance measure is calculated as the normalised difference between OOB accuracy of the original observations to randomly permuted variables (Cutler et al., 2007; Mellor et al., 2013). An alternative variable importance measure is calculated by summing all of the decreases in Gini impurity at each tree node split, normalised by the number of trees (Criminisi et al., 2012; Mellor et al., 2013). RF is a well-regarded machine learning tool that can identify complex and non-linear relationships in fitting datasets and has been shown to offer high classification accuracy (Cutler et al., 2007; Criminisi et al., 2012; Dash et al., 2016).

Several RF models were used to examine the data using the randomForest R package (Liaw and Wiener, 2002) and the R statistical computing environment (R Core Team, 2016). RF models were fitted using the stress classes calculated from the tree health assessment data as response variables and the detrended spectral indices serving as candidate predictor variables. The purpose of model development was to compare the performance of models using various combinations of spectral indices. Once the best performing model was identified this was used to map the progression of stress symptoms across the study site.

2.5.2. Model accuracy and statistical analysis

Classification performance for each model (Table 3) was assessed using Cohens Kappa (kappa) coefficient based on a 10% resampling of the training dataset repeated for 1000 iterations.

During the re-sampling process, 10% of samples were excluded from the training dataset and the remaining samples were used to train a RF model to predict the stress classes of the excluded samples. Many iterations were used to eliminate the possibility that the results observed were associated with the randomly drawn exclusion sample. This provided a large (27,000 samples) independent dataset to assess classification accuracy. Cohen's kappa is a robust and widely used metric for assessing the agreement between two sets of observations. Kappa values and associated confidence intervals were calculated using the 'psych' R package (Revelle, 2016) and the weighted kappa statistics were reported. Confidence intervals for kappa values were calculated using the methods proposed by Fleiss et al. (1969) available through the 'psych' R package. The kappa statistic is generally deemed to be robust because it accounts for agreements occurring through chance. Weighted kappa values were used so that misclassification by more than one category was penalised more heavily, ensuring that the kappa values were conservative. Several authors have suggested that the agreement expressed through kappa, which varies between 0 and 1, can be broadly classified as slight (0–0.20), fair (0.21–0.40), moderate (0.41–0.60) and substantial (0.61–1) (Richard Landis, 1977; Hauglin and Orka, 2016; Dash et al., 2017).

Time series data were assessed using the cross correlation factor (CCF) available from the 'stats' R package. The CCF was calculated between field discolouration measures and all spectral indices examined. This analysis was designed to examine whether the spectral indices and the field discolouration scores varied together or whether there was a temporal lag that would indicate that a spectral index was sensitive earlier or later to the field measured scores. Linear models were used to examine the effect of cluster size and image resolution on classification accuracy. The data were visually inspected and a linear model using ordinary least squares (OLS) including the main effects and the interaction was fit to the data. Diagnostic plots were examined in all cases to check for compliance with the standard assumptions of the linear models fitted. Where evidence of a failure to comply were noted, an appropriate alternative modelling approach was selected based on generalised least squares. An analysis of variance (ANOVA) procedure was used to examine the influence of the main effects, cluster size and image resolution, and the interaction between the two main effects on various response variables. Post-hoc Tukey multiple comparison tests were used to test for significant differences between groups. The interaction between main effects was eliminated from the model if its influence on the response variable was not significant ($p > 0.05$). OLS models were fitted using the 'stats' R package (R Core Team, 2016) and GLS models were developed using the "nlme" R package (Pinheiro et al., 2017). Multiple comparison tests and analysis of deviance tables were generated using the "car" R package (Fox and Weisberg, 2011).

3. Results

3.1. Summary of study datasets

In total, 21 separate datasets were collected, processed and available for inclusion in the analysis. In most cases, the field data and the multi-spectral imagery were recorded on the same date. All remotely sensed data from the 27th November was abandoned as unfavourable weather conditions made the data unusable. After March 2016 the study trees were harvested and the study was concluded as no further data could be collected.

3.2. Field data and simulated disease expression

Within the study period, trees treated with herbicide exhibited symptoms including discolouration of previously green needles in

the tree canopy and, to a lesser extent, defoliation (Fig. 2). Discolouration became apparent to the field assessor in herbicide treated trees around one month after herbicide application. At this date, the field assessments revealed needle discolouration scores increasing from an average of 0% to an average of 80% over the course of a month. During the entire study period there was no change in the status of the trees allocated to the control treatment (0–10%) where no herbicide was applied. This indicates that there was no discernible needle discolouration in the portion of the canopy visible from the ground. Nine trees that were treated with herbicide application did not show any needle discolouration over the study period. We are confident that this was the result of error in treatment application.

In general, herbicide application induced more discolouration than defoliation by the end of the study. The defoliation that was observed was less widespread across the study, whereas discolouration occurred across nearly all treated blocks. A significant number of trees exposed to herbicide application exhibited no changes in canopy density over the course of the study. For both discolouration and canopy density the curve fitted to the trees with herbicide application can be interpreted as the "simulated disease expression curve" for the experimental treatments in this study.

3.3. Spectral indices

Raster datasets detailing four spectral indices (Table 2) related to plant stress were calculated from each image mosaic. These were clipped to the boundaries of the study tree clusters that were manually delineated using the final image - where clusters were easily and precisely identified (3rd March 2016). The values for all pixels within a cluster were averaged to provide a mean value for each cluster at each date. The spectral indices for each date were detrended using the method described in Section 2.5.2. Examination of time-series spectral indices from the control treatment indicated that there was a significant trend in the data that was not due to the herbicide induced stress. Once the background changes were accounted for, the reduction in spectral indices in the treated tree clusters was more pronounced, this suggests that the background noise which could hide the spectral stress signal had been adequately controlled for. The changes observed in the control plots were likely caused by the annual needle flush of *P. radiata* occurring in the early summer and from the greening of trees and understorey vegetation associated with weather patterns (Fig. 3). Meteorological data from a nearby station showed that after a dry spring in the region, there was low rainfall at the site through November (93 mm) and December (32 mm) 2015. This was followed by a return to higher rainfall levels in January (129 mm), February (163 mm) and March (162 mm) 2016 and this pattern probably induced a late flush in the study trees and understorey vegetation.

Spectral indices in the control and treated areas fluctuated throughout the experiment (Fig. 3a, b, e, f, i and j). Once detrending was applied, the effects of physiological stress became evident. Previous studies (Eitel et al., 2011) have used the 25th and 75th percentiles of the indices extracted from healthy trees to define a "no-change" region. Plots could be considered stressed if the 75th percentile detrended spectral index values dropped below the lower bound of the no change region. Following detrending, all spectral indices remained within the no change region for the duration of the experiment (Fig. 3c, g and i.). By contrast, the detrended spectral indices in treated plots dropped below the no change region following herbicide application (Fig. 3d, h and l). The time elapsed before spectral index values moved out of the no-stress zone provides valuable insight into the utility of the spectral index and the physiological changes occurring in the subject trees. In our study the NDVI, GNDVI, RENDVI and NLI all dropped

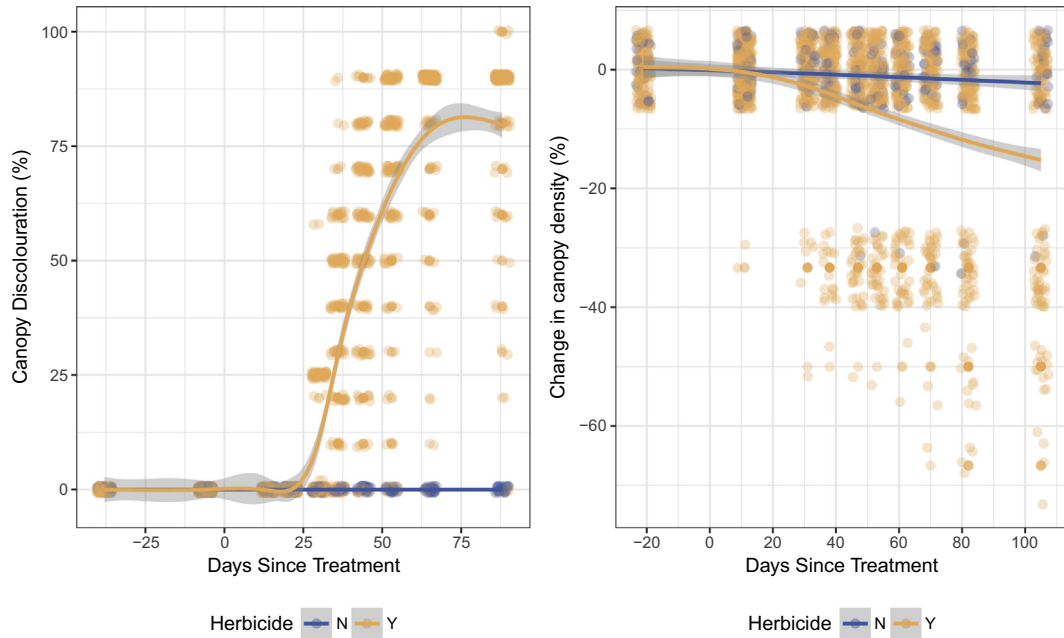


Fig. 2. The study tree health data showing changes in canopy needle colour and canopy density as scored from the ground by an experienced forest health assessor. Each datum shows data from a single assessment of a single tree and a smoothed line of best fit with 95% confidence interval for trees with (blue) and without (yellow) herbicide application is shown. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

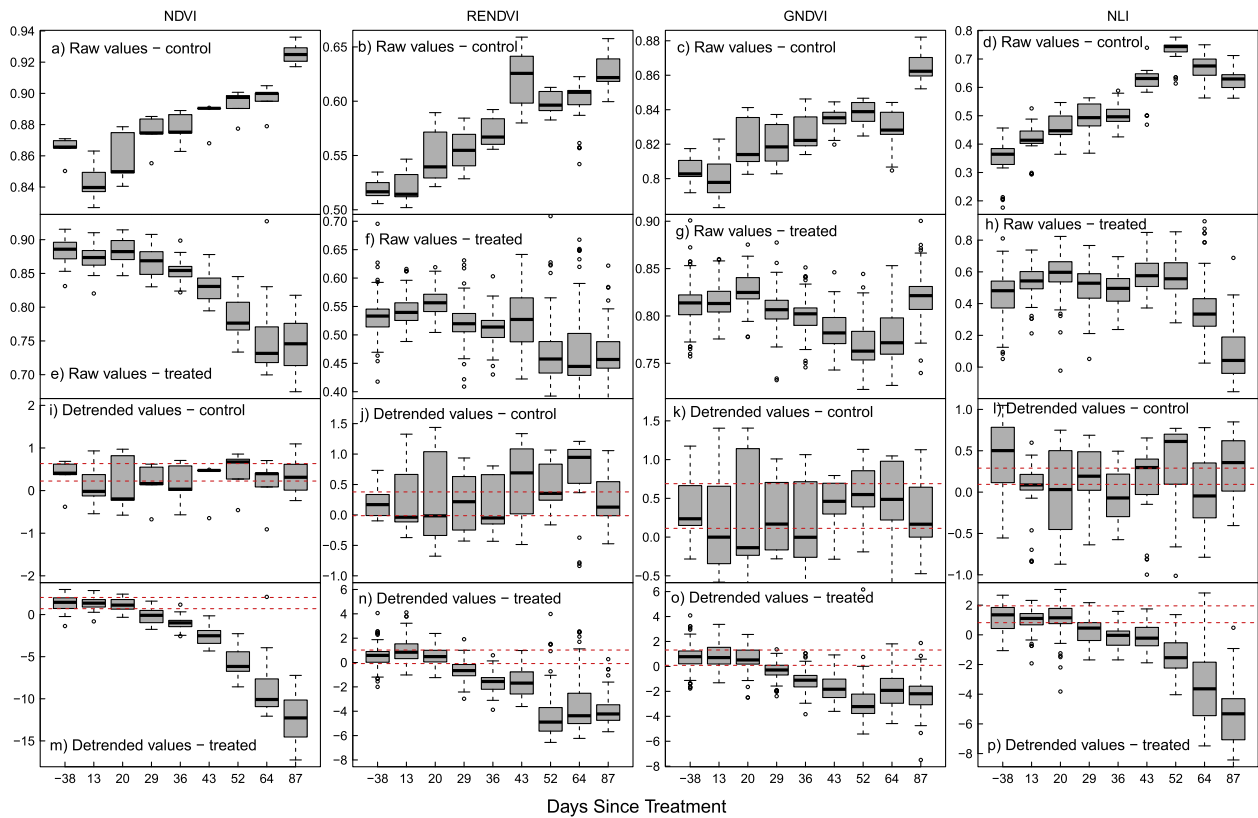


Fig. 3. Box and whisker plots showing the raw and detrended values for selected spectral indices observed in the control plots ($n = 5$) and the treated plots $n = 30$. Upper and lower red dotted lines indicate the extent of the “no-change” region. The no change region was defined as the 25th and 75th percentiles of the spectral value prior to herbicide treatment (30th October 2015). An area was classified as exhibiting significant stress if the 75th percentile of the spectral index value dropped below the no change region. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

below the no change region 29 days after herbicide application. This coincided with the first detection of discolouration by field crews.

The field measurements of tree condition for the herbicide treated trees were plotted on the same x-axis as the spectral indices (Fig. 4). The changes in the spectral indices were very closely cor-

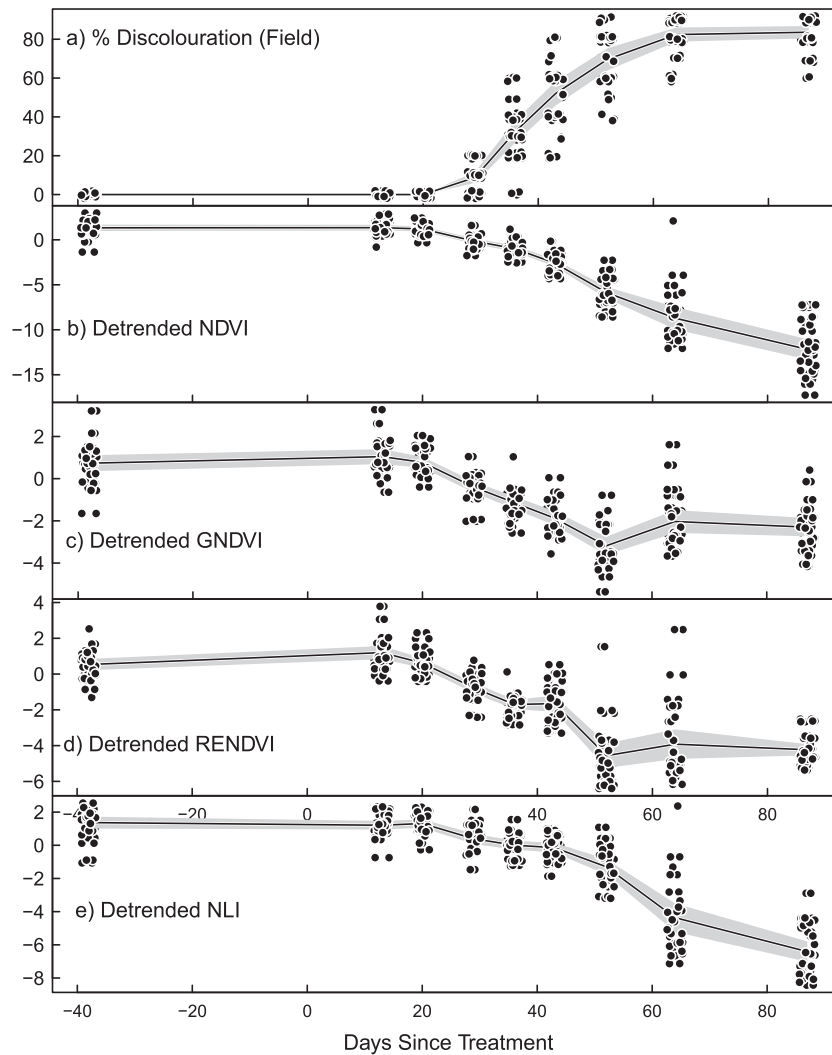


Fig. 4. The percentage discolouration within treated trees recorded by field crews (a) and the detrended values for spectral indices NLI (b), NDVI (c), GNDVI (d) and RENDVI (e). Each datum represents a single observation of a treated tree cluster. The solid black line shows the mean value and the grey shaded area shows the 95% confidence interval about the mean.

related with changes in canopy colour identified by traditional field measurement. Cross-correlation analysis between the response variables (canopy discolouration) and the spectral indices showed that there was no lag in changes between field observed canopy discolouration and the spectral indices. The cross-correlations were largest when the lag was equal to zero. This indicates that the spectral indices provided an accurate indicator of field measured values but that in this dataset they did not provide an earlier indication of stress than the field observations.

3.4. Classification models

Random forest models were fitted to the discolouration classes calculated from the field measurements using all four spectral indices (Table 2) as predictors. The importance scores extracted from this model offered some insight into the relationship between various spectral indices and tree health scores. The importance scores showed that NDVI was the most important predictor of discolouration (MDA = 55.36) followed by RENDVI (MDA = 47.98), GNDVI (MDA = 23.37) and NLI (MDA = 15.01). Thirteen separate random forest models were subsequently fitted to examine the influence of the number and choice of spectral indices. Independent validation of these models using a 10% resampling approach

indicated that the best performing model was a three variable model (Model 4) that included NDVI, RENDVI, and NLI as predictors (kappa = 0.694). The second most accurate model was the four variable model included all four spectral indices trialled (Model 1).

Graphical analysis of the best performing model (Model 4 in Table 3) provides a means of examining classification accuracy throughout the course of the study (Fig. 5). This analysis indicated that classification accuracy varied throughout the course of the study. For some time steps (13, 20, 64, 87 days), the agreement between observed and predicted levels of physiological stress were in perfect agreement with the field observations. For other time steps the correspondence was less accurate. For all time steps, the general pattern of classification was highly accurate, with a small amount of noise associated with model mis-specification that was most apparent during the middle stages of the appearance of symptoms of physiological stress expression. This result suggests that the RF classification model would likely be a useful predictive model for practical applications.

3.5. Image resolution

The original imagery was re-sampled to four coarser resolution datasets. These were processed using a custom R programme

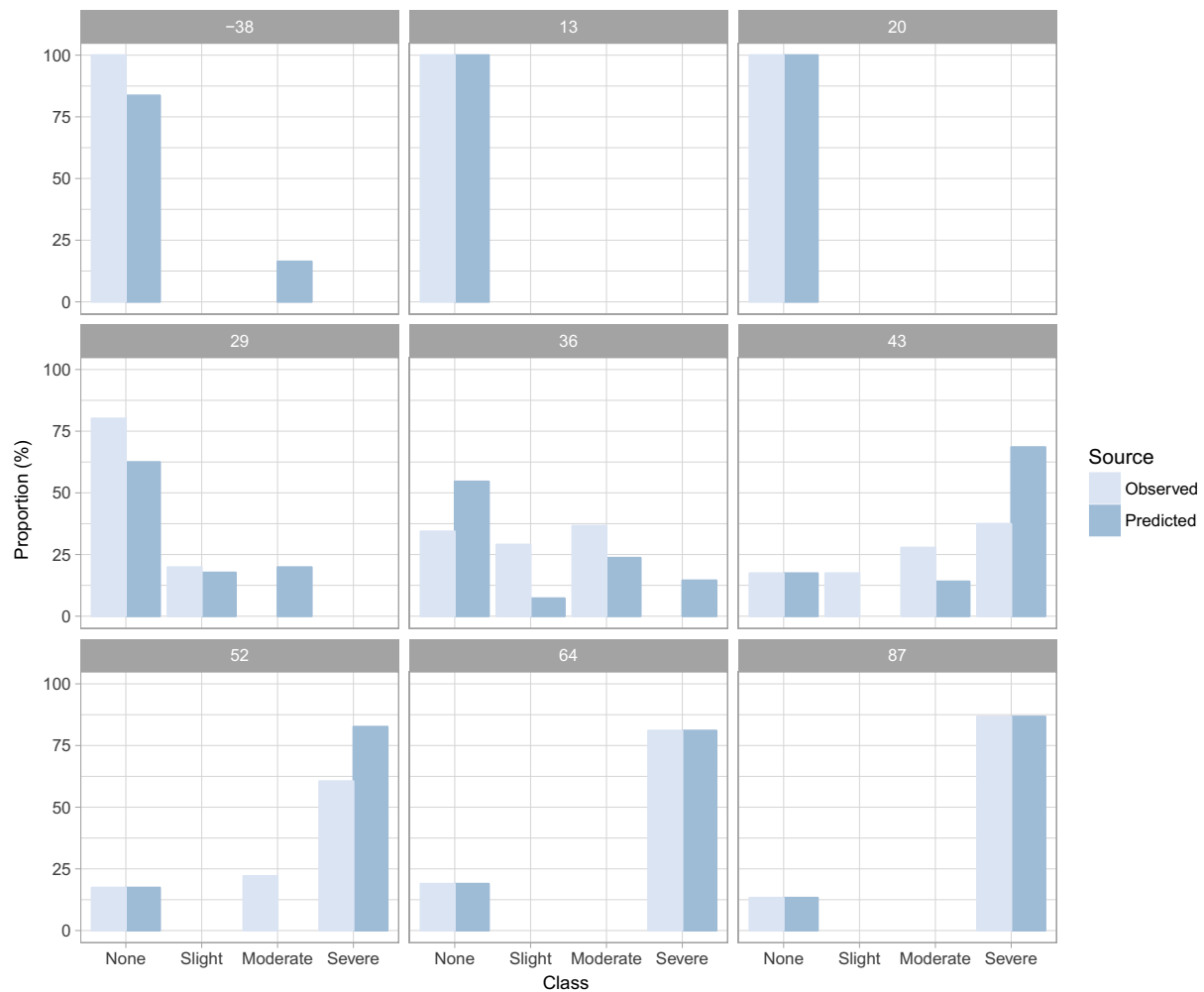


Fig. 5. Observed (field) and predicted (random forest) proportion of trees assigned to each stress class. Results are based on a 10% random resampling of the training dataset, this provides total separation of training and testing data and so is an independent validation. Each panel shows the results for a different time step, the number in the header refers to the number of days since treatment.

developed by the authors for the original imagery to extract and detrend spectral indices for the study tree clusters. As before, the changes in selected spectral indices over time were plotted to assess their utility for indicating physiological stress. As the threshold for detection of stress was of interest, separate lines were plotted for each cluster size (Fig. 6). For all indices, noise levels increased as the image resolution became coarser. For all resolution imagery, NDVI was most sensitive to physiological stress. NDVI moved further below the lower threshold of the no change zone than the other indices. For finer resolution imagery it was also evident that NDVI was considerably less noisy than the other indices examined (Fig. 6). At a 5 m image resolution, the mean NDVI values for all clusters showed substantially less change than at finer resolutions. Many of the treatments did not move outside of the no-change zone and those that did (4 and 2) were only marginally outside of the no-change zone. Values for the smallest cluster size (single tree) displayed considerably more variation and showed the least change between their starting and ending values during the course of the study.

The effect of image resolution and the cluster size of treated trees on model accuracy were examined using an iterative 10% re-sampling of the training dataset. For each image resolution 10% of the clusters were excluded at random from the training dataset. The remaining clusters were used to fit the best performing classification model (Model 4) which was used to predict

symptom class at each time step for the excluded clusters. To account for random variation associated with the sample, re-sampling was repeated 1000 times. Kappa was then calculated for each cluster size at each image resolution. This analysis revealed that model accuracy was greatest for all cluster sizes at a 1 m image resolution (Fig. 7). When we examined the effect of image resolution and cluster size, model accuracy was poorest for the smallest clusters (1 and 2 trees) and decreased with image resolution after increasing between 0.06 and 1 m. For larger clusters (4, 8 and 16 trees), model accuracy was relatively invariant to changes in image resolution within the range tested.

4. Discussion

Our results showed NDVI to be the most useful vegetation index for detecting the discolouration caused by physiological stress over time. Many indices have been used within remote sensing research over the last 40 years (Bannari et al., 1995), but NDVI remains the most popular (Thorp and Tian, 2004) and has been widely used to discriminate healthy from senescent foliage in forest stands affected by insects attack (Babst et al., 2010; Jepsen et al., 2009; Kharuk et al., 2009; Spruce et al., 2011; Lottering and Mutanga, 2016). Our results, based on simulated disease expression, support the broad utility of NDVI for forest health assessments where mod-

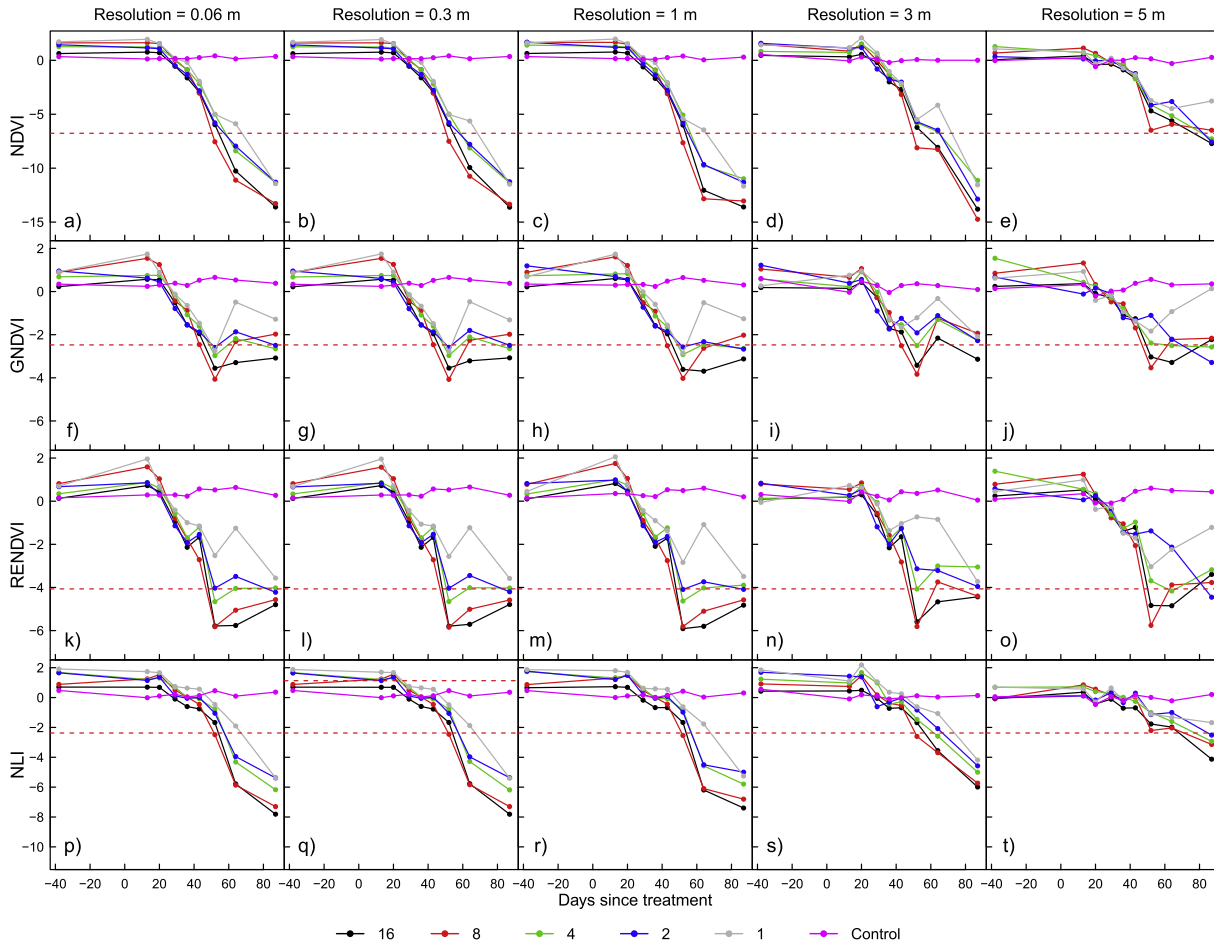


Fig. 6. Mean detrended spectral indices for clusters within the study area. Red dashed line show the 25th and 75th percentile of the pre-treatment spectral index. The area between the red-dashed line can be deemed a “no-change” region, a spectral index can be considered to have changed over time when its mean value moves outside of the no-change zone. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

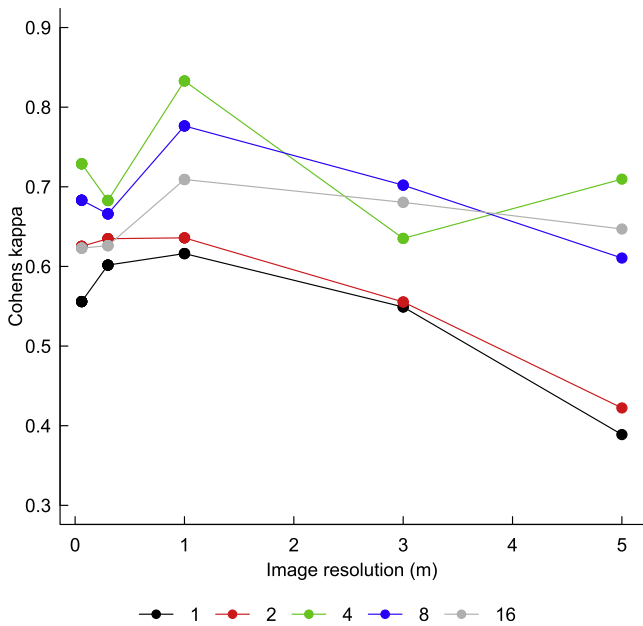


Fig. 7. Cohen's kappa coefficient plotted against image resolution for the different cluster sizes in each experimental treatment.

erate levels of foliar discoloration are expected. High values of NDVI in healthy foliage are associated with photosynthetic activity due to contrasting low leaf reflectance in the red band that occurs with increased chlorophyll absorption and high infrared reflection within the spongy mesophyll layer (Rullan-Silva et al., 2013). Values of NDVI diminish with foliage senescence which is associated with increased reflectance in the red band (Asner, 1998; Rullan-Silva et al., 2013).

As an individual predictive variable, RENDVI was the second most important vegetation index of those tested, and the combination of NDVI, RENDVI and NLI produced the most accurate model. Several studies have reported that reflectance in the red edge portion of the spectrum (690–740 nm) is one of the most informative descriptors of foliar chlorophyll concentration. This portion of the spectrum covers the region where reflectance increases sharply from the chlorophyll-absorbed red portion of the spectrum (near 680 nm) to the strongly reflected near-infrared region of the spectrum (Curran et al., 1990). The red edge has been found to be very useful for early detection of stress in conifers (Eitel et al., 2011) as this portion of the spectrum is where the first stress-induced changes in needle reflectance are likely to occur (Rullan-Silva et al., 2013). An advantage of RENDVI in detection of early stress is that this index is reportedly more sensitive to changes in high chlorophyll levels than indices such as NDVI

(Eitel et al., 2011). Although we did not note this greater sensitivity in our study this may have been attributable to measurements not being at a fine enough temporal scale during early onset stress expression. Given the disparity between our findings and those previously reported a logical approach may be to include both indices within predictive models of discolouration for practical deployment.

The optimal spatial resolution of imagery for maximising predictive power of the models was found to be 1 m for both small and large clusters. Previous research has shown optimal spatial resolution of multi-spectral imagery for detecting insect damage of 2.4 m (Meddens and Hicke, 2014) for *Pinus contorta* dominated stands, 1.25–2.25 m for *Eucalyptus* spp. (Lottering and Mutanga, 2016) and 1.75–2.3 m for *P. radiata* (Ismail et al., 2008). Our results generally agree with the findings of this research, with our results suggesting that the optimal resolution should be approximately equal to the crown radius (approx. 1.5–2 m in this study).

Airborne imagery and data from the modern, sub-metre resolution satellites (e.g. GeoEye-1 and WorldView-3) provide imagery of a spatial resolution that is equivalent to or lower than the optimal resolution identified here. New opportunities for collection of high resolution data are now also emerging from constellations of cubesats. These are simple, small satellites that are relatively low cost, operate in low orbits and can be deployed in large numbers. Several planned constellations of these satellites aim to collect 3–10 m multispectral imagery covering large portions of the Earth with high frequency. These new data streams may have the potential to inform forest health assessments using the approach demonstrated in this study. However, future work will be required to assess whether the radiometric and geometric corrections that must be applied to these data will impact the utility of the imagery.

Clusters of four and eight trees were detected with the highest levels of accuracy at the initial resolution. These cluster sizes were more reliably detected than smaller clusters at all resolutions tested. When dealing with larger clusters (4, 8, 16) the accuracy was less affected by decreasing resolution. Interestingly, at the original resolution, the largest clusters of 16 trees were poorly identified by the best model and only the 1 tree clusters had a lower kappa. This was the result of two different factors. In the case of the 1-tree clusters, the delineated crowns covered a relatively small number of pixels and suffered from a high noise to signal ratio where unavoidable changes in image quality and the accuracy of geo-registration between scenes nearly overwhelmed the signal from the relatively small number of pixels representing poisoned trees. However, in the case of the largest clusters, our analysis faced the opposite problem where the outline of the poisoned trees formed a large, contiguous polygon that included many healthy boughs and pixels representing gaps and understory. In theory, it would be possible to apply more advanced delineation of tree canopies using, for example, object-based analysis to improve this result. However, any fine-scale identification of affected trees would then suffer from the inevitable fluctuations in geo-registration from scene to scene. Future work might explore the use of airborne laser scanning to accurately delineate crown shapes and reduce the sources of error we encountered with the largest and smallest cluster sizes. Recent research has suggested that this is practical in regularly spaced plantations where airborne laser scanning has been used to model individual tree size and shape (Pont et al., submitted for publication). Fusing this data with multi-spectral data, as in Dash et al. (2017), may provide insight into both tree growth and tree health. This has many potential applications including offering powerful insights for genetic improvement programmes (e.g. for *P. radiata* (Dungey et al., 2009) *Pseudotsuga menziesii* (Dungey et al., 2012)) that have been

shown to add significant value for industrial forest plantations (Moore et al., 2017).

Although small clusters were detected with lower accuracy, the model was still able to identify these with considerable certainty. This suggests that the methods demonstrated in this study might be able to detect small, isolated outbreaks during the early stages of a disease associated with needle discolouration. We are unaware of previous research that has investigated both the influence of the size of tree clusters affected by disease and the interaction between cluster size and spatial resolution on the ability to detect needle discolouration from UAV imagery. Detection systems of this type may be extremely useful for studying disease dynamics in forest systems, this will enhance forest health surveillance efforts.

Model accuracy was greatest when predicting extreme classes of physiological stress in the foliage rather than intermediate levels. Using hyper-spectral imagery and photogrammetry collected from a UAV, Näsi et al. (2015) predicted defoliation of *Picea abies* from European spruce bark beetle. When three colour classes, that included healthy, infested and dead, were used the kappa was 0.6, but this accuracy increased to 0.8 with only healthy and dead colour classes (Näsi et al., 2015). Our results are consistent with this finding and clearly show high, and sometimes perfect, classification when prediction was made of either the healthy or severe discolouration classes, but lower accuracy for the intermediate discolouration classes.

Although physiological modelling approaches have been used in previous research for modelling forest health (Zarco-Tejada et al., 2004; Ligot et al., 2014), empirical approaches are far more common. Most empirical approaches that have been used to date are parametric models that include approaches such as generalised linear regression models, generalised additive models (Meddens and Hicke, 2014), mixed-effects models (Rullán-Silva et al., 2015; Townsend et al., 2012), logistic regression (Townsend et al., 2012; Kantola et al., 2010) and partial least squares regression models (Oumar and Mutanga, 2014). Non-parametric modelling approaches are less common in the literature and include classification and regression tree analysis (CART) (Stone and Coops, 2007), artificial neural networks (ANN) (Lottering and Mutanga, 2016) and support vector algorithms (SVM) (Fassnacht et al., 2014). The machine learning algorithm random forest (RF) used here is a non-parametric method that has been successfully used elsewhere to predict forest health (Abdel-Rahman et al., 2014; Wang et al., 2015). In contrast to parametric approaches, which often rely on assumptions about the underlying data, non-parametric models have the advantage of freedom from distributional assumptions about the underlying data. Furthermore, RF can accommodate noisy data, missing data values, both numerical and categorical data and can be configured to incorporate other sources of auxiliary spatial data. Importantly, these models have low susceptibility to over-fitting which allows development of more complex models with correlated variables such as RENDVI and NDVI.

Our results support evidence for the utility of UAVs as a platform for forest health monitoring. The high spatial and temporal resolution of these datasets enabled detection of a simulated disease outbreak, showing no lag from traditional forest health assessments. However, classification accuracy was lower during the very early stages of simulated disease expression in our study. This means that detection accuracy during the very early stage of an outbreak using this method may be limited using the multi-spectral sensor deployed in our research. Previous research in a similar context has shown that physiological stress can be accurately detected at earlier stages when hyper-spectral sensors are used (Abdel-Rahman et al., 2014). This may be attributed to the increased spectral resolution of these sensors that can observe physiological changes resulting in altered reflectance outside of the wavelengths observed by sensors such as the MicaSense

RedEdge 3. For example, previous work has shown that bands in the short wave infrared (SWIR) can be used to detect the effects of early stages of *Sirex noctilio* infestation (Ismail and Mutanga, 2011). While our work demonstrates that the choice of spectral index is important, it is likely that the narrow spectral range limits the detection of changes in the very early stages. This limitation is likely to be related to the nature of the physiological response during the early stages of insect or pathogen attack. The greater accuracy achieved during the early phases of symptom expression in previous studies, using more sophisticated sensors, shows that carefully matching sensor capability to the required objective is important for developing effective remote sensing solutions. Based on our results, UAV-based monitoring can be considered a valuable tool for monitoring forest health but the delayed detection relative to more sophisticated sensors suggests that careful matching between sensor and application is necessary. Emerging platforms and sensors are currently best deployed to augment existing technologies rather than to supersede more sophisticated sensors on more traditional platforms.

The success of forest surveillance is largely governed by being able to match the temporal, spatial and spectral characteristics of symptoms with an appropriate platform, sensor and image analysis method. Over the past two decades high resolution satellite imagery has not had an adequate temporal resolution for routine monitoring. As a result, airborne platforms have been more widely used as they are more flexible and responsive (McConnell et al., 2000). The advent of low-cost, easy to use UAVs and miniaturised multi-spectral sensors may change the operational strategy for forest health monitoring. Our results suggest a role for UAVs as a means of filling the gap between the large-scale capabilities of airborne surveillance and the need for targeted, regular surveillance of smaller areas in a cost-effective manner. In contrast to orbital satellite platforms, UAVs can also supply imagery on cloudy days which can be an important consideration in areas where overcast conditions are common during periods of peak symptom expression. Compared to both satellite and manned aircraft platforms, the time and financial resources needed to prepare and initiate a UAV flight is greatly reduced, allowing for greater flexibility in scheduling imagery acquisition.

5. Conclusions

Results from this study clearly show that collection of multi-spectral imagery from a UAV provides a useful method to detect and monitor the symptoms resulting from a simulated forest disease outbreak resulting in foliar discoloration. The non-parametric models created from this data predicted the onset of needle discoloration well and had a high overall accuracy (κ range 0.39–0.64). Analyses showed the optimal spatial resolution of imagery to range from 1 m for symptoms expressed on both small and large tree clusters. Further research should compare the accuracy of models created from high resolution satellite imagery with those from UAVs in detection of physiological stress resulting from disease or pest outbreaks. We conclude that UAVs may have an important role to play in forest health monitoring, filling the spatial and temporal gap between large-scale airborne assessment and infrequent small-scale ground surveys.

Acknowledgments

This study was funded as part of the Growing Confidence in Forestry's Future (GCF) programme (Contract number: CO4X1306). GCF was co-funded by the Forest Growers Levy Trust and the New Zealand Ministry for Business, Innovation and Employment (MBIE). The authors recognise the contributions of Toby Stovold,

Kane Fleet, Mark Millar and Rod Brownlie who completed all field work and helped develop the study methods. Scion colleague Stefan Gous is acknowledged for offering advice on herbicide application rates. Graham Coker (Scion) and Mike Baker (Hancock Forest Management) provided valuable reviews of the document. We are grateful to the forest managers for providing access and allowing us to complete this study in their forest.

Author contributions

J.D. M.W. and H.D. conceived and designed the trial. J.D. developed and completed the analysis with assistance from G.P. M.H. Decapitalise Managed data collection and incoming data management. J. D. and M.W. wrote the initial draft of this manuscript which was then reviewed and edited by all authors.

Conflict of interests

The authors declare no conflict of interest.

References

- Abdel-Rahman, E.M., Mutanga, O., Adam, E., Ismail, R., 2014. Detecting *Sirex noctilio* grey-attacked and lightning-struck pine trees using airborne hyperspectral data, random forest and support vector machines classifiers. ISPRS J. Photogram. Remote Sens. 88, 48–59.
- Abdel-Rahman, E.M., Mutanga, O., Adam, E., Ismail, R., 2014. Detecting *Sirex noctilio* grey-attacked and lightning-struck pine trees using airborne hyperspectral data, random forest and support vector machines classifiers. ISPRS J. Photogram. Remote Sens. 88, 48–59.
- Asner, G.P., 1998. Biophysical and biochemical sources of variability in canopy reflectance. Remote Sens. Environ. 64, 234–253.
- N.Z.F.O. Association, 2013. New Zealand Plantation Forest Industry Facts and Figure. New Zealand Forest Owners Association, New Zealand.
- Babst, F., Esper, J., Parlow, E., 2010. Landsat TM/ETM+ and tree-ring based assessment of spatiotemporal patterns of the autumnal moth (*Epirrita autumnata*) in northernmost Fennoscandia. Remote Sens. Environ. 114, 637–646.
- Bannari, A., Morin, D., Bonn, F., Huete, A., 1995. A review of vegetation indices. Remote Sens. Rev. 13, 95–120.
- Barton, C.V., 2012. Advances in remote sensing of plant stress. Plant Soil 354, 41–44.
- Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.
- Brown, A.V., Webber, J.F., 2008. Red Band Needle Blight of Conifers in Britain: Research Note 002. Forestry Commission, Edinburgh, UK. [https://www.forestry.gov.uk/PDF/fcrn002.pdf/\\$FILE/fcrn002.pdf](https://www.forestry.gov.uk/PDF/fcrn002.pdf/$FILE/fcrn002.pdf).
- Bulman, L.S., Gadjil, P.D., Kershaw, D.J., Ray, J.W., 2004. Assessment and Control of Dothistroma Needle-blight. Forest Research Bulletin No. 229, 48. New Zealand Forest Research Institute, Rotorua, New Zealand.
- Bulman, L.S., Bradshaw, R.E., Fraser, S., Martn-Garca, J., Barnes, I., Musolin, D.L., La Porta, N., Woods, A.J., Diez, J.J., Koltay, A., Drenkhan, R., Ahumada, R., Poljakovic-Pajnik, L., Queloz, V., Pikur, B., Domu-Lehtijarvi, H.T., Chira, D., Tomeov-Haataja, V., Georgieva, M., Jankovsk, L., Anselmi, N., Markovskaja, S., Papazova-Anakieva, I., Sotirovski, K., Lazarevi, J., Adamkov, K., Boro, P., Bragana, H., Vettraino, A.M., Selikhovkin, A.V., Bulgakov, T.S., Tubby, K., 2016. A worldwide perspective on the management and control of Dothistroma needle blight. Forest Pathol. 46, 472–488.
- Calderón, R., Navas-Cortés, J.A., Zarco-Tejada, P.J., 2015. Early detection and quantification of verticillium wilt in olive using hyperspectral and thermal imagery over large areas. Remote Sens. 7, 5584–5610.
- 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.
- Coops, N.C., Johnson, M., Wulder, M.A., White, J.C., 2006. Assessment of QuickBird high spatial resolution imagery to detect red attack damage due to mountain pine beetle infestation. Remote Sens. Environ. 103, 67–80.
- Coops, N.C., Johnson, M., Wulder, M.A., White, J.C., 2006. Assessment of quickbird high spatial resolution imagery to detect red attack damage due to mountain pine beetle infestation. Remote Sens. Environ. 103, 67–80.
- Criminisi, Antonio, Shotton, Jamie, Konukoglu, Ender, 2012. Decision forests: a unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning. Found. Trends Comput. Graph. Vis. 7 (2), 81–227. NOW Publishers.
- Cunningham, S.C., Read, J., Baker, P.J., Nally, R.M., 2007. Quantitative assessment of stand condition and its relationship to physiological stress in stands of *Eucalyptus camaldulensis* (Myrtaceae). Aust. J. Bot. 55, 692–699.
- Curran, P.J., Dungan, J., Gholz, H., 1990. Exploring the relationship between reflectance red edge and chlorophyll content in slash pine. Tree Physiol. 7, 33–48.

- Cutler, D.R., Edwards, T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J., Lawler, J.J., 2007. Random forests for classification in ecology. *Ecology* 88, 2783–2792.
- Dash, J.P., Marshall, H.M., Rawley, B., 2015. Methods for estimating multivariate stand yields and errors using k-NN and aerial laser scanning. *Forestry* 88, 237–247.
- Dash, J.P., Watt, M.S., Bhandari, S., Watt, P., 2016. Characterising forest structure using combinations of airborne laser scanning data, RapidEye satellite imagery and environmental variables. *Forestry* 89, 159–169.
- Dash, J.P., Pearse, G.D., Watt, M.S., Paul, T., 2017. Combining airborne laser scanning and aerial imagery enhances echo classification for invasive conifer detection. *Remote Sens.* 9, 156.
- Dennison, P.E., Brunelle, A.R., Carter, V.A., 2010. Assessing canopy mortality during a mountain pine beetle outbreak using GeoEye-1 high spatial resolution satellite data. *Remote Sens. Environ.* 114, 2431–2435.
- Dungey, H., Brawner, J., Burger, F., Carson, M., Henson, M., Jefferson, P., Matheson, A., 2009. A new breeding strategy for *Pinus radiata* in New Zealand and New South Wales. *Silvae Genetica* 58, 28–38.
- Dungey, H., Low, C., Lee, J., Millar, M., Fleet, K., Yanchuk, A., 2012. Developing breeding and deployment options for douglas-fir in New Zealand: breeding for future forest conditions. *Silvae Genetica* 61, 104–115.
- Eitel, J.U.H., Vierling, L.A., Litvak, M.E., Long, D.S., Schulthess, U., Ager, A.A., Krofcheck, D.J., Stoscheck, L., 2011. Broadband, red-edge information from satellites improves early stress detection in a New Mexico conifer woodland. *Remote Sens. Environ.* 115, 3640–3646.
- Fassnacht, F.E., Latifi, H., Ghosh, A., Joshi, P.K., Koch, B., 2014. Assessing the potential of hyperspectral imagery to map bark beetle-induced tree mortality. *Remote Sens. Environ.* 140, 533–548.
- Fleiss, J., Cohen, J., Everitt, B., 1969. Large sample standard errors of kappa and weighted kappa. *Psychol. Bull.* 72, 323–327 (Cited By 72).
- Fox, J., Weisberg, S., 2011. *An R Companion to Applied Regression*, second ed. Sage, Thousand Oaks CA.
- Franklin, S., Wulder, M., Skakun, R., Carroll, A., 2003. Mountain pine beetle red-attack forest damage classification using stratified landsat TM data in British Columbia, Canada. *Photogram. Eng. Remote Sens.* 69, 283–288.
- Fraser, R., Latifovic, R., 2005. Mapping insect-induced tree defoliation and mortality using coarse spatial resolution satellite imagery. *Int. J. Remote Sens.* 26, 193–200.
- Furby, S.L., Campbell, N.A., 2001. Calibrating images from different dates to like-valued digital counts. *Remote Sens. Environ.* 77, 186–196.
- García-Ruiz, F., Sankaran, S., Maja, J.M., Lee, W.S., Rasmussen, J., Ehsani, R., 2013. Comparison of two aerial imaging platforms for identification of Huanglongbing-infected citrus trees. *Comput. Electron. Agric.* 91, 106–115.
- Garrity, S.R., Allen, C.D., Brumby, S.P., Gangodagamage, C., McDowell, N.G., Cai, D.M., 2013. Quantifying tree mortality in a mixed species woodland using multitemporal high spatial resolution satellite imagery. *Remote Sens. Environ.* 129, 54–65.
- GDAL Development Team, GDAL, 2016. Geospatial Data Abstraction Library. Version 2.02. Open Source Geospatial Foundation.
- Gitelson, A., Merzlyak, M.N., 1994. Spectral reflectance changes associated with autumn senescence of *Aesculus hippocastanum* L. and *Acer platanoides* L. Leaves. Spectral features and relation to chlorophyll estimation. *J. Plant Physiol.* 143, 286–292.
- Gitelson, A.A., Merzlyak, M.N., 1998. Remote sensing of chlorophyll concentration in higher plant leaves. *Adv. Space Res.* 22, 689–692.
- Goel, N.S., Qin, W., 1994. Influences of canopy architecture on relationships between various vegetation indices and LAI and Fpar: a computer simulation. *Remote Sens. Rev.* 10, 309–347.
- Goetz, S.J., Wright, R.K., Smith, A.J., Zinecker, E., Schaub, E., 2003. IKONOS imagery for resource management: tree cover, impervious surfaces, and riparian buffer analyses in the mid-Atlantic region. *Remote Sens. Environ.* 88, 195–208.
- Goodwin, N.R., Coops, N.C., Wulder, M.A., Gillanders, S., Schroeder, T.A., Nelson, T., 2008. Estimation of insect infestation dynamics using a temporal sequence of landsat data. *Remote Sens. Environ.* 112, 3680–3689.
- Guo, Q., Kelly, M., Gong, P., Liu, D., 2007. An object-based classification approach in mapping tree mortality using high spatial resolution imagery. *GISci. Remote Sens.* 44, 24–47.
- Hauglin, M., Orka, H.O., 2016. Discriminating between native Norway Spruce and invasive sitka spruce: a comparison of multitemporal landsat 8 imagery. *Remote Sens.* 8, 363.
- Havašová, M., Bucha, T., Ferenčík, J., Jakuš, R., 2015. Applicability of a vegetation indices-based method to map bark beetle outbreaks in the High Tatra Mountains. *Ann. For. Res.* 58, 295–310.
- Healey, S.P., Cohen, W.B., Zhiqiang, Y., Krankina, O.N., 2005. Comparison of Tasseled Cap-based Landsat data structures for use in forest disturbance detection. *Remote Sens. Environ.* 97, 301–310.
- Hewitt, A.E., 2010. *New Zealand Soil Classification*, third ed. Landcare Research Science Series No. 1.
- Hicke, J.A., Logan, J., 2009. Mapping whitebark pine mortality caused by a mountain pine beetle outbreak with high spatial resolution satellite imagery. *Int. J. Remote Sens.* 30, 4427–4441.
- Hulme, P., 2014. Future proofing the biosecurity of New Zealand. *Primary Indust. Manage.* 18, 22–24.
- Ismail, R., Mutanga, O., 2011. Discriminating the early stages of *Sirex noctilio* infestation using classification tree ensembles and shortwave infrared bands. *Int. J. Remote Sens.* 32, 4249–4266.
- Ismail, R., Mutanga, O., Kumar, L., Bob, U., 2008. Determining the optimal spatial resolution of remotely sensed data for the detection of *Sirex noctilio* infestations in pine plantations in KwaZulu-Natal, South Africa. *S. Afr. Geogr. J.* 90, 22–31.
- Jepsen, J., Hagen, S., Høgda, K., Ims, R., Karlsen, S., Tømmervik, H., Yoccoz, N., 2009. Monitoring the spatio-temporal dynamics of geometrid moth outbreaks in birch forest using MODIS-NDVI data. *Remote Sens. Environ.* 113, 1939–1947.
- Johnson, E.W., Ross, J., 2008. Quantifying error in aerial survey data. *Austr. Forest.* 71, 216–222.
- Kantola, T., Vastaranta, M., Yu, X., Lyytikäinen-Saarenmaa, P., Holopainen, M., Talvitie, M., Kaasalainen, S., Solberg, S., Hyyppä, J., 2010. Classification of defoliated trees using tree-level airborne laser scanning data combined with aerial images. *Remote Sens.* 2, 2665–2679.
- Kharuk, V., Ranson, K., Im, S., 2009. Siberian silkmouth outbreak pattern analysis based on spot vegetation data. *Int. J. Remote Sens.* 30, 2377–2388.
- Leckie, D., Jay, C., Gougeon, F., Sturrock, R., Paradine, D., 2004. Detection and assessment of trees with *Phellinus weirii* (laminated root rot) using high resolution multi-spectral imagery. *Int. J. Remote Sens.* 25, 793–818.
- Lehmann, J.R.K., Nieberding, F., Prinz, T., Knoth, C., 2015. Analysis of unmanned aerial system-based CIR images in forestry new perspective to monitor pest infestation levels. *Forests* 6, 594–612.
- Liaw, A., Wiener, M., 2002. Classification and regression by randomforest. *R News* 2, 18–22.
- Ligot, G., Balandier, P., Courbaud, B., Claessens, H., 2014. Forest radiative transfer models: which approach for which application? *Can. J. For. Res.* 44, 391–403.
- Lottering, R., Mutanga, O., 2016. Optimising the spatial resolution of WorldView-2 pan-sharpened imagery for predicting levels of *Gonipterus scutellatus* defoliation in KwaZulu-Natal, South Africa. *ISPRS J. Photogram. Remote Sens.* 112, 13–22.
- McConnell, T., Johnson, E., Burns, B., 2000. *A Guide to Conducting Aerial Sketchmap Surveys*. US Forest Service, fhrtet 00-01. Forest Health Technology Enterprise Team, Fort Collins, CO. 88 p.
- Meddens, A.J., Hicke, J.A., 2014. Spatial and temporal patterns of landsat-based detection of tree mortality caused by a mountain pine beetle outbreak in Colorado, USA. *For. Ecol. Manage.* 322, 78–88.
- Meigs, G.W., Kennedy, R.E., Cohen, W.B., 2011. A landsat time series approach to characterize bark beetle and defoliator impacts on tree mortality and surface fuels in conifer forests. *Remote Sens. Environ.* 115, 3707–3718.
- Mellor, A., Haywood, A., Stone, C., Jones, S., 2013. The performance of random forests in an operational setting for large area sclerophyll forest classification. *Remote Sens.* 5, 2838.
- Moore, J.R., Dash, J.P., Lee, J.R., McKinley, R.B., Dungey, H.S., 2017. Quantifying the influence of seedlot and stand density on growth, wood properties and the economics of growing radiata pine. *Forest: Int. J. For. Res.* 00, 1–14.
- Näsi, R., Honkavaara, E., Lyytikäinen-Saarenmaa, P., Blomqvist, M., Litke, P., Hakala, T., Viljanen, N., Kantola, T., Tanhuanpää, T., Holopainen, M., 2015. Using UAV-based photogrammetry and hyperspectral imaging for mapping bark beetle damage at tree-level. *Remote Sens.* 7, 15467–15493.
- Oumar, Z., Mutanga, O., 2014. Integrating environmental variables and WorldView-2 image data to improve the prediction and mapping of *Thaumastocoris peregrinus* (bronze bug) damage in plantation forests. *ISPRS J. Photogram. Remote Sens.* 87, 39–46.
- Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., R Core Team, nlme: Linear and Nonlinear Mixed Effects Models. R Package Version 3.1-130.
- Pont, D., Watt, M.S., Morgenroth, J., Dungey, H.S., 2017. Correlating size, form, wood quality, and disease attributes with crown metrics from tree-based aerial lidar (submitted for publication).
- Poona, N.K., Ismail, R., 2013. Discriminating the occurrence of pitch canker fungus in *Pinus radiata* trees using quickbird imagery and artificial neural networks. *Southern For.: J. For. Sci.* 75, 29–40.
- Pu, R., Kelly, M., Anderson, G.L., Gong, P., 2008. Using CASI hyperspectral imagery to detect mortality and vegetation stress associated with a new hardwood forest disease. *Photogram. Eng. Remote Sens.* 74, 65–75.
- R Core Team, 2016. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Revelle, W., 2016. psych: Procedures for Psychological, Psychometric, and Personality Research. Northwestern University, Evanston, Illinois. R package version 1.6.12.
- Richard Landis, G.G.K.J., 1977. The measurement of observer agreement for categorical data. *Biometrics* 33, 159–174.
- Rouse Jr., J.W., Haas, R.H., Schell, J.A., Deering, D.W., 1974. Monitoring vegetation systems in the great plains with ERTS. *NASA Spec. Publ.* 351, 309.
- Rullan-Silva, C., Olthoff, A., de la Mata, J.D., Pajares-Alonso, J., 2013. Remote monitoring of forest insect defoliation—a review. *For. Syst.* 22, 377–391.
- Rullan-Silva, C., Olthoff, A.E., Pando, V., Pajares, J.A., Delgado, J.A., 2015. Remote monitoring of defoliation by the beech leaf-mining weevil *Rhynchaenus fagi* in Northern Spain. *For. Ecol. Manage.* 347, 200–208.
- Sankaran, S., Mishra, A., Ehsani, R., Davis, C., 2010. A review of advanced techniques for detecting plant diseases. *Comput. Electron. Agric.* 72, 1–13.
- Sims, D.A., Gamon, J.A., 2002. Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. *Remote Sens. Environ.* 81, 337–354.
- Skakun, R.S., Wulder, M.A., Franklin, S.E., 2003. Sensitivity of the thematic mapper enhanced wetness difference index to detect mountain pine beetle red-attack damage. *Remote Sens. Environ.* 86, 433–443.
- Smith, G.M., Milton, E.J., 1999. The use of the empirical line method to calibrate remotely sensed data to reflectance. *Int. J. Remote Sens.* 20, 2653–2662.

- Smith, D., Smith, I., Collett, N., Elms, S., 2008. Forest health surveillance in victoria. *Austr. Forest.* 71, 188–195.
- Spruce, J.P., Sader, S., Ryan, R.E., Smoot, J., Kuper, P., Ross, K., Prados, D., Russell, J., Gasser, G., McKellip, R., et al., 2011. Assessment of modis NDVI time series data products for detecting forest defoliation by gypsy moth outbreaks. *Remote Sens. Environ.* 115, 427–437.
- Stone, C., Coops, N.C., 2004. Assessment and monitoring of damage from insects in Australian eucalypt forests and commercial plantations. *Austr. J. Entomol.* 43, 283–292.
- Stone, C., Coops, N.C., 2007. Assessing the health of *Pinus radiata* plantations using remote sensing data and decision tree analysis. *NZ J. Forest. Sci.* 37, 000–000.
- Stone, C., Penman, T., Turner, R., 2012. Managing drought-induced mortality in *Pinus radiata* plantations under climate change conditions: a local approach using digital camera data. *For. Ecol. Manage.* 265, 94–101.
- Stone, C., Carnegie, A., Melville, G., Smith, D., Nagel, M., 2013. Aerial mapping canopy damage by the aphid *Essigella californica* in a *Pinus radiata* plantation in southern New South Wales: what are the challenges? *Austr. Forest.* 76, 101–109.
- Thorp, K., Tian, L., 2004. A review on remote sensing of weeds in agriculture. *Precision Agric.* 5, 477–508.
- Townsend, P.A., Singh, A., Foster, J.R., Rehberg, N.J., Kingdon, C.C., Eshleman, K.N., Seagle, S.W., 2012. A general landsat model to predict canopy defoliation in broadleaf deciduous forests. *Remote Sens. Environ.* 119, 255–265.
- Verbesselt, J., Robinson, A., Stone, C., Culvenor, D., 2009. Forecasting tree mortality using change metrics derived from MODIS satellite data. *For. Ecol. Manage.* 258, 1166–1173.
- Wang, J., Rich, P.M., Price, K.P., Kettle, W.D., 2004. Relations between NDVI and tree productivity in the central Great Plains. *Int. J. Remote Sens.* 25, 3127–3138.
- Wang, H., Pu, R., Zhu, Q., Ren, L., Zhang, Z., 2015. Mapping health levels of *Robinia pseudoacacia* forests in the Yellow River delta, China, using IKONOS and landsat 8 OLI imagery. *Int. J. Remote Sens.* 36, 1114–1135.
- Watt, M., Palmer, D., Bulman, L., 2011. Predicting the severity of *Dothistroma* on *Pinus radiata* under future climate in New Zealand. *NZ J. Forest. Sci.* 41, 207–215.
- Watt, M.S., Dash, J.P., Bhandari, S., Watt, P., 2015. Comparing parametric and non-parametric methods of predicting site index for radiata pine using combinations of data derived from environmental surfaces, satellite imagery and airborne laser scanning. *For. Ecol. Manage.* 357, 1–9.
- Watt, M.S., Dash, J.P., Watt, P., Bhandari, S., 2016. Multi-sensor modelling of a forest productivity index for radiata pine plantations. *NZ J. Forest. Sci.* 46, 1–14.
- Watt, M.S., Kimberley, M.O., Dash, J.P., Harrison, D., 2017. Spatial prediction of optimal final stand density for even-aged plantation forests using productivity indices. *Can. J. For. Res.* 47, 527–535.
- West, J.S., Bravo, C., Oberti, R., Lemaire, D., Moshou, D., McCartney, H.A., 2003. The potential of optical canopy measurement for targeted control of field crop diseases. *Annu. Rev. Phytopathol.* 41, 593–614.
- White, J.C., Wulder, M.A., Brooks, D., Reich, R., Wheate, R.D., 2005. Detection of red attack stage mountain pine beetle infestation with high spatial resolution satellite imagery. *Remote Sens. Environ.* 96, 340–351.
- Wulder, M.A., Dymond, C.C., White, J.C., Leckie, D.G., Carroll, A.L., 2006. Surveying mountain pine beetle damage of forests: a review of remote sensing opportunities. *For. Ecol. Manage.* 221, 27–41.
- Wulder, M.A., White, J.C., Bentz, B., Alvarez, M.F., Coops, N.C., 2006a. Estimating the probability of mountain pine beetle red-attack damage. *Remote Sens. Environ.* 101, 150–166.
- Wulder, M.A., White, J.C., Bentz, B.J., Ebata, T., 2006b. Augmenting the existing survey hierarchy for mountain pine beetle red-attack damage with satellite remotely sensed data. *Forest. Chron.* 82, 187–202.
- Wulder, M.A., White, J.C., Coops, N.C., Butson, C.R., 2008. Multi-temporal analysis of high spatial resolution imagery for disturbance monitoring. *Remote Sens. Environ.* 112, 2729–2740.
- Yao, R.T., Scarpa, R., Turner, J.A., Barnard, T.D., Rose, J.M., Palma, J.H., Harrison, D.R., 2014. Valuing biodiversity enhancement in New Zealand's planted forests: socioeconomic and spatial determinants of willingness-to-pay. *Ecol. Econ.* 98, 90–101.
- Zarco-Tejada, P.J., Miller, J.R., Harron, J., Hu, B., Noland, T.L., Goel, N., Mohammed, G. H., Sampson, P., 2004. Needle chlorophyll content estimation through model inversion using hyperspectral data from boreal conifer forest canopies. *Remote Sens. Environ.* 89, 189–199.