

A remote sensing approach to mapping fire severity in south-eastern Australia using sentinel 2 and random forest



Rebecca Gibson^{a,*}, Tim Danaher^{a,b}, Warwick Hehir^c, Luke Collins^{d,e,f}

^a Remote Sensing and Landscape Science, Department of Planning, Industry and Environment, Alstonville, NSW 2477, Australia

^b Joint Remote Sensing Research Program, School of Geography, Planning and Environmental Management, University of Queensland, Brisbane, QLD 4072, Australia

^c New South Wales Rural Fire Service, Sydney Olympic Park, NSW 2127, Australia

^d Department of Ecology, Environment & Evolution, La Trobe University, Bundoora, Victoria 3086, Australia

^e Arthur Rylah Institute for Environmental Research, Department of Environment, Land, Water and Planning, PO Box 137, Heidelberg, Victoria 3084, Australia

^f Research Centre for Future Landscapes, La Trobe University, Bundoora, Victoria 3086, Australia

ARTICLE INFO

Edited by Marie Weiss

Keywords:

Fire severity

Sentinel 2

Aerial photograph interpretation

Random forest

Machine learning

Fractional cover

Normalised burn ratio

Supervised classification

Canopy density

Topographic complexity

Regional scale

Australia

ABSTRACT

Accurate and consistent broad-scale mapping of fire severity is an important resource for fire management as well as fire-related ecological and climate change research. Remote sensing and machine learning approaches present an opportunity to enhance accuracy and efficiency of current practices. Quantitative biophysical models of photosynthetic, non-photosynthetic and bare cover fractions have not been widely applied to fire severity studies but may provide greater consistency in comparisons of different fires across the landscape compared to reflectance-based indices. We systematically tested and compared reflectance and fractional cover candidate severity indices derived from Sentinel 2 satellite imagery using a random forest (RF) machine learning framework. Assessment of predictive power (cross-validation) was undertaken to quantify the accuracy of mapping severity of new fires. The effect of environmental variables on the accuracy of the RF predicted severity classification was examined to assess the stability of the mapping across the landscape. The results indicate that fire severity can be mapped with very high accuracy using Sentinel 2 imagery and RF supervised classification. The mean accuracy was > 95% for the unburnt and extreme severity class (complete crown consumption), > 85% for high severity class (full crown scorch), > 80% for low severity (burnt understorey, unburnt canopy) and > 70% for the moderate severity class (partial canopy scorch). Higher canopy cover and higher topographic complexity was associated with a higher rate of under-prediction, due to the limitations of optical sensors in viewing the burnt understorey of low severity classes under these conditions. Further research is aimed at improving classification accuracy of low and moderate severity classes and applying the RF algorithm to hazard reduction fires.

1. Introduction

Fire severity is a metric of the loss or change in organic matter caused by fire (Keeley, 2009). Although fire severity may be correlated with fire intensity (i.e. the energy output of a fire, Hammill and Bradstock, 2006; Keeley, 2009), factors such as climate, weather conditions, topography and vegetation community composition strongly influence the way fire intensity is translated into fire severity (Turner et al., 1999; Hammill and Bradstock, 2006; Bradstock et al., 2010; Cansler and McKenzie, 2014; Keeley and Syphard, 2016; Zylstra et al., 2016). Fire severity is typically heterogeneous within the extent of a fire-affected area, particularly in forested landscapes, with spatial heterogeneity increasing as fire severity decreases (Turner et al., 1999;

Hudak et al., 2004; Cansler and McKenzie, 2014). Knowledge of the spatial patterns of fire severity is critical for effective management in fire-prone landscapes, including prioritisation of post-fire emergency response (Keeley, 2009) and understanding the efficacy of fuel treatments in reducing wildfire risk (Cary et al., 2017; Tubbessing et al., 2019). Furthermore, fire severity mapping can facilitate a better understanding of the ecological effects of fire and provide insight into the drivers of wildfire behaviour (Collins et al., 2014; Coppoletta et al., 2016). Accurate and consistent fire severity mapping at suitable temporal and spatial scales is a vital resource for ecological and climate change research as well as operational fire management and future planning (Flannigan et al., 2009; Keeley and Syphard, 2016).

Broad scale fire severity mapping is predominantly achieved using

* Corresponding author.

E-mail address: Rebecca.Gibson@environment.nsw.gov.au (R. Gibson).

remotely sensed imagery (Keeley, 2009; Pettorelli et al., 2014). Image differencing techniques using multi-date change detection (e.g. between pre- and post-fire images) have been extensively used in remote sensing fire severity mapping studies (Chafer et al., 2004; Lentile et al., 2006; Miller and Thode, 2007; Hall et al., 2008; Cansler and McKenzie, 2014). Numerous reflectance indices have been derived and compared, including the differenced Normalised Burn Ratio (dNBR), the soil-adjusted vegetation index (Hoy et al., 2008; Smith et al., 2010), burned area index (Marino et al., 2014), and Tasseled-cap brightness and greenness transformations (Epting et al., 2005). The most commonly used index is the differenced Normalised Burn Ratio (dNBR), which has been shown to produce reasonable mapping of the spatial variation in severity within a single fire, across a range of vegetation communities (e.g. approximately 60–70% accuracy compared to field validation, Miller and Thode, 2007; Soverel et al., 2010). Relativising the dNBR with the pre-fire NBR (RdNBR) has been shown to improve accuracy, particularly for higher severity classes in heterogeneous landscapes (Miller et al., 2009). Thresholding the NBR into severity classes has formed the basis for the small number of national fire severity mapping programs (e.g. Monitoring Trends in Burn Severity program, (Eidenshink et al., 2007)). However, classification of dNBR or RdNBR into standardised severity classes that are consistent between fires across the landscape may be problematic due to the influence of pre-fire vegetation structure, soil type and vegetation moisture on the NBR (Miller and Thode, 2007; Kolden et al., 2015). A combination of multiple indices may provide more complete and accurate information than any one alone (Miller and Thode, 2007).

Indices measuring the relative cover of photosynthetic, non-photosynthetic material and bare ground may provide a more useful 2-dimensional remote sensing surrogate to estimate the quantity of organic matter consumed by fire than traditional reflectance-based estimates (e.g. NBR). Spectral un-mixing estimates the relative sub-pixel fractions of photosynthetic and non-photosynthetic vegetation as well as bare soil cover, using a calibrated relationship with high quality, quantitative field data (Scarth et al., 2010; Guerschman et al., 2015). A fractional cover-derived fire severity index may be more directly analogous to traditional field-based severity assessment, with potential for consistent thresholding of severity classes between vegetation types across the landscape based on the loss of cover after fire (Lentile et al., 2006; Morgan et al., 2014; Meddens et al., 2016).

Areas with dense canopy and high topographic relief associated with topographic shadows are known to create challenges for remote sensing change detected techniques, including fire severity mapping (Hoy et al., 2008; Lyndersen et al., 2016). This may be particularly problematic for mapping the burnt understorey of low severity classes, due to the reflectance signal being obscured by canopy or topographic shadows (Lyndersen et al., 2016). Other topographic effects on dNBR have been reported where sites with steep north-facing aspect had consistently lower pre-fire NBR compared to other aspects, due to the solar angle orientation (i.e. a steep northern aspect faces away from incoming solar radiation in the far northern hemisphere; (Verbyla et al., 2008)). However, influences of atmospheric attenuation and the scattering of electromagnetic radiation (e.g. bidirectional surface reflectance effects) are markedly reduced in topographically complex areas with appropriate imagery pre-processing corrections for normalised surface reflectance (Ediriweera et al., 2013; Flood et al., 2013b). Furthermore, spectral un-mixing analyses have been shown to minimise effects of topography on fire severity classification accuracy (Rogan and Franklin, 2001). Understanding the performance of remotely sensed fire severity indices across landscape factors is important for appropriate interpretation and application in land management decisions.

Supervised classification of remote sensing data through machine learning approaches such as random forest (RF) has the potential to overcome limitations of using a single index or simple linear regression models within complex heterogeneous landscapes that tend to generate low inter-class discrimination and high intra-class variation (Cutler

et al., 2007; Ghimire et al., 2012; Collins et al., 2018). Random forest is an ensemble learning algorithm, with an iterative nature ensuring a convergence approach to classification. These algorithms can improve accuracy considerably because the error of a single classification is outweighed across the ensemble of multiple classifications through a rule-based approach (i.e. majority voting; Breiman, 2001a; Ghimire et al., 2012; Rodriguez-Galiano et al., 2012). Random forest algorithms have strong predictive accuracy because the majority voting method is robust against over-fitting (Breiman, 2001b; Belgiu and Dragut, 2016). Furthermore, non-parametric supervised classifiers, such as RF, do not make any assumptions about the data and, as such, are increasingly used to classify remote sensing data, which rarely have normal distributions (Belgiu and Dragut, 2016). The use of RF algorithms have been shown to improve fire severity classification compared to traditional parametric approaches (e.g. Meddens et al., 2016; Meng et al., 2017; Collins et al., 2018).

The Sentinel 2 constellation was launched in late 2015, with two twin polar-orbiting satellites operational by mid-2016 capable of repeat coverage every 5 days at the equator, which results in 2–3 days at mid-latitudes. Operationally, currently site revisit time is approximately 5 days at mid-latitudes. Sentinel 2 has systematic coverage limited to between latitudes 56° south and 84° north (ESA, 2019). The multispectral instrument onboard the Sentinel 2 satellites captures imagery across 13 spectral bands, including the visible and near infra-red (NIR) bands captured at 10 m resolution and the short wave infra-red (SWIR) bands captured at 20 m resolution (ESA, 2019). These bandwidths are commonly used in remote sensing studies of vegetation, soil and water cover (Summers et al., 2011; Zhang et al., 2011) as well as in fire severity indices including the NBR (Epting et al., 2005). The higher frequency and resolution of Sentinel 2 imagery offers potential improvements in accuracy and cloud contamination issues for application in broad-scale fire severity mapping, relative to other moderate resolution sensors such as Landsat TM (30 m resolution).

The primary aims of the study were to (i) assess the utility of high-resolution Sentinel imagery for fire severity mapping, (ii) evaluate the performance of fire severity classification using traditional spectral indices and fractional cover products, and (iii) identify landscape factors associated with fire severity misclassification. The study focused on eight mixed severity fires occurring across a broad environmental gradient in southern Australia. A random forest modelling approach was used for fire severity classification.

2. Methods

2.1. Study area

The study took place across the state of New South Wales, eastern Australia (Fig. 1). The study included eight wildfires, which were selected based on the availability of high resolution (< 50 cm) 4-band (blue, green, red and NIR) post-fire digital aerial photography (provided by NSW Rural Fire Service, RFS) overlapping with Sentinel 2 satellite imagery (post-2016) (Fig. 1, Table 1).

The vegetation type across the study fires was dominated by dry sclerophyll woodland and open-forest, characterised by a ground layer of graminoids and herbaceous plants, a diverse shrub understorey (< 5 m height), and a tree canopy dominated by *Eucalyptus* species of variable height. Low trees (10 m) are more common on upper slopes and ridges while taller trees (20–30 m) are more common on lower slopes and in gullies. Shrubland (i.e. heath), grassland, grassy woodland, tall wet forest and rainforest were major components of several fires, though were not consistent across all fires (Supplementary material, Table 2.1). The study fires occurred across several climatic zones including the temperate (Sir Bertram, Holsworthy, Wollemi and Tathra fires), Mediterranean-type (i.e. warm dry summer/cool winter, Sir Ivan and Mt. Canobolas fires), sub-tropical (Frenchman's West fire) and semi-arid zones (Pilliga fire, Hutchinson et al., 2005). Soil type varied

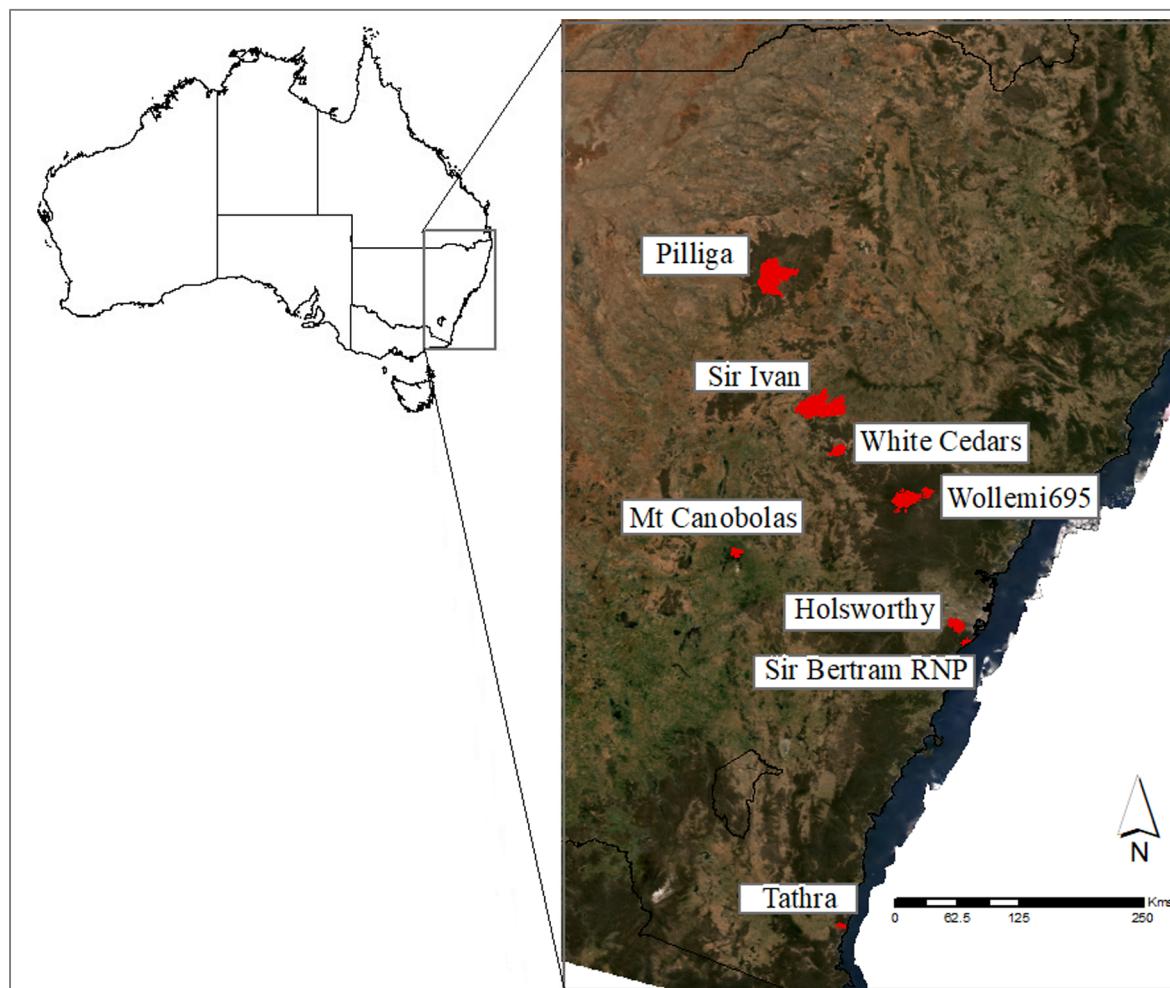


Fig. 1. Location of the study fires in NSW, Australia.

across the study area, governed by complex interactions with parent material, which included predominantly sandstones (e.g. Pilliga, Sir Bertram, Holsworthy, Wollemi) and volcanic (e.g. Mt. Canobolas fire) material.

2.2. Training and validation dataset

Aerial photograph interpretation (API) was undertaken to generate hand digitised spatial polygons of homogenous sample areas of fire severity classes in ArcMap v10.4. Severity classification through API followed clearly defined rules and interpretation cues ([Table 2](#)) to discriminate five severity classes based on varying levels of vegetation scorch and consumption of canopy and understorey layers ([Hudak](#)

[et al., 2004; McCarthy et al., 2017; Collins et al., 2018](#)). These classes have been used to distinguish different degrees of post-fire change to foliage cover and biomass in Australian forests, woodlands and shrublands, with strong correlations to field-based measures of fire severity ([Hamill and Bradstock, 2006; McCarthy et al., 2017](#)).

Random sampling points were generated within the API severity class polygons, such that points were > 15 m from the polygon edge to ensure sampling of Sentinel 2 pixels that fell entirely within the target severity class. The number of sampling points generated was as large as possible, while being randomly distributed and representative of the real-world occurrence of fire severity classes in each fire ([Breiman, 2001a; Millard and Richardson, 2015](#)). Although the numbers of sampling points were unbalanced between severity classes and fires,

Table 1

Summary of fires with start and end dates, details of high-resolution photography (airborne digital sensor, ADS, or unmanned aerial vehicle, UAV) and the use of data for aerial photo interpretation (API). The number of days the photography was captured after the fire end date is included in parentheses. Sentinel 2 (Sen2) satellite imagery dates that were manually selected. The number of days before the fire start date or after the fire end date is included in parentheses.

RFS fire name	Start date	End date	Photography	Photography date	Pre-fire Sen2 date	Post-fire Sen2 date	Fire size (Ha)
Sir Ivan	20170211	20170217	50 cm 4-band ADS	20170218(1)	20170207(4)	20170224(7)	47,105
White Cedars	20170212	20170217	20 cm 4-band ADS	20170216(-1)	20161126(68)	20170224(7)	5217
695 (Wollemi National Park)	20180128	20180215	50 cm 4-band ADS	20180317(30)	20180105(23)	20180301(14)	14,178
Mt Canobolas	20180210	20180216	30 cm 4-band ADS	20180309(21)	20180118(24)	20180227(11)	1891
Sir Bertram Royal National Park	20180120	20180125	50 cm 4-band ADS	20180311(38)	20171222(29)	20180130(5)	2241
Pilliga	20180119	20180125	50 cm 4-band ADS	20180311(38)	20171214(38)	20180212(18)	57,822
Tathra Reedy Swamp	20180318	20180319	10 cm 4-band ADS	20180320(1)	20180224(23)	20180326(7)	1258
Holsworthy	20180413	20180418	50 cm 4-band ADS	20180424(6)	20180311(34)	20180420(2)	3955

Table 2
Aerial photo interpretation classification of fire severity, adapted from Hudak et al. (2004) and McCarthy et al. (2017) and Collins et al. (2018), including number and proportion of sampling points and the number of independent polygons for each severity class.

Severity class	Description	Interpretation cues (false colour infra-red aerial photos)	% foliage fire affected	N sampling points	Prop. of total points	N polygons
Unburnt Low	Unburnt surface with green canopy	Dark red (live understorey) between the dark red tree crowns	0% canopy and understory burnt	40,444	38%	272
	Burnt surface with unburnt canopy	Dark grey (burnt understorey) between the dark red tree crowns	> 10% burnt understory	11,376	11%	548
Moderate High	Partial canopy scorch	A mixture of green, orange and brown colours in tree canopies	> 90% green canopy	9459	9%	817
	Full canopy scorch (\pm partial canopy consumption)	No green or orange, but an even brown colour in tree canopies	20–90% canopy scorched	23,231	22%	1006
Extreme	Full canopy consumption	Mostly black and dark grey, largely no canopy cover	< 50% canopy biomass consumed	20,878	20%	1106
			> 50% canopy biomass consumed			

previous studies using RF algorithms for image classification recommend against boosted sampling in naturally rare classes (Millard and Richardson, 2015). The average distance between sampling points was 28 m, ensuring a single Sentinel 2 pixel was not sampled twice. Sampling points were attributed through spatial joins with the corresponding API severity class. For each sampling point in each fire, corresponding pixel values were extracted for each candidate fire severity index using the raster and shapefile packages in R (v3.5.0) to create a data-frame of training and validation data used as input into the RF models.

2.3. Imagery selection and pre-processing

Sentinel 2 satellite imagery with low cloud cover (< 5%), as close as possible to the start and end date of each fire were manually selected (see Table 1 for the number of days). Sentinel 2 tiles (100 km \times 100 km) were downloaded from the Copernicus Hub (<http://nci.org.au>) as level 1C products, which represent orthorectified, top-of-atmosphere reflectance. The data are packaged into images according to the spatial resolution of the bands; 10 m (bands 2, 3, 4 and 8), 20 m (bands 5, 6, 7, 8b, 11 and 12) and 60 m (bands 1, 9 and 10). The images were processed to represent standardised surface reflectance with a nadir view angle and incidence angle of 45° (Flood et al., 2013a, b). This corrected for variations due to atmospheric conditions and the bi-directional reflectance distribution function (BRDF), which also accounted for topographic variations using a 30 m digital surface model (DSM) derived from the Shuttle Radar Topography Mission digital elevation models (Farr et al., 2007; Gallant and Read, 2009). These corrections minimise the differences between scenes caused by different sun and view angles.

The Sentinel 2 SWIR bands (11 and 12) were pan-sharpened from 20 m to 10 m resolution. The Theil-Sen Estimator, a robust regression technique (Sen, 1968), was used to fit a local linear relationships between each low resolution SWIR band and the high resolution red band using the pixels in a 7 by 7 window (based on the high-resolution pixels, i.e. 70 m by 70 m). The low-resolution band pixels within each window were then recalculated at the higher resolution using the specific local regression model. The 10 m surface reflectance images were recombined into a single stack of Sentinel 2 bands at 10 m resolution that correspond to the equivalent Landsat TM bands (blue, green, red, NIR, SWIR1, SWIR2). Fractional cover products were generated for each image using a fractional cover model, which calculates for each pixel, the proportion of photosynthetic (green) vegetation, non-photosynthetic ('non-green', dead or senescent vegetation), and bare ground cover (Guerschman et al., 2015). This fractional cover model was developed for Landsat imagery so a surface reflectance adjustment (Flood, 2017) is applied to the Sentinel 2 images before applying the Landsat fractional cover model. Field validation of sentinel 2 fractional cover done for pre and post-fire images at Sphinx HR and Frenchman's West HR using the star transect method which captures the relative bare, photosynthetic and non-photosynthetic cover (Muir et al., 2011). The field measurements had robust agreement with the Sentinel 2 satellite derived estimates for each fraction (Supplementary material, Fig. 1.1), with comparable accuracy to other fractional cover field validation (Guerschman et al., 2015).

2.4. Candidate fire severity indices

Timeseries plots were used to examine the variation in the NBR and fractional cover values to assess the sensitivity of indices to fire severity classes. Random sampling points within the hand digitised fire severity classes in each study fire were used to extract the NBR value from the Sentinel 2 reflectance image and each band (bare, green and non-green) of the Sentinel 2 fractional cover image for every cloud free image date before and after each fire within the available Sentinel 2 archive (2016–2018). The NBR and fractional cover band values were displayed

within each severity class for each fire separately, due to different fire start dates.

The following candidate fire severity indices were generated for each pair of pre- and post-fire images: differenced normalised burn ratio (dNBR); relativised dNBR (RdNBR); relativised change in total fractional cover (RdFCT); and change in bare fractional cover (dFCB), the differenced normalised differenced vegetation index (dNDVI) and the differenced Bare Soil Index (dBSI), given by the following formulas:

$$\text{preNBR} = \left(\frac{\text{pre NIR} - \text{pre SWIR}}{\text{pre NIR} + \text{pre SWIR}} \right) \quad (1)$$

$$\text{postNBR} = \left(\frac{\text{post NIR} - \text{post SWIR}}{\text{post NIR} + \text{post SWIR}} \right) \quad (2)$$

$$\text{dNBR} = \text{preNBR} - \text{postNBR} \quad (3)$$

$$\text{ABS pre NBR} = |\text{preNBR}| \quad (4)$$

$$\text{RdNBR} = \left(\frac{\text{dNBR}}{\sqrt{(\text{ABS pre NBR})}} \right) \quad (5)$$

$$\text{pre Total} = \text{pre Green} + \text{pre Non Green} \quad (6)$$

$$\text{post Total} = \text{post Green} + \text{post Non Green} \quad (7)$$

$$\text{RdFCT} = \frac{\text{pre Total} - \text{post Total}}{\sqrt{(\text{pre Total})}} \quad (8)$$

$$\text{dFCB} = \text{post Bare} - \text{pre Bare} \quad (9)$$

$$\text{preNDVI} = \left(\frac{\text{pre NIR} - \text{pre Red}}{\text{pre NIR} + \text{pre Red}} \right) \quad (10)$$

$$\text{postNDVI} = \left(\frac{\text{post NIR} - \text{post Red}}{\text{post NIR} + \text{post Red}} \right) \quad (11)$$

$$\text{dNDVI} = \text{post NDVI} - \text{pre NDVI} \quad (12)$$

$$\text{preBSI} = \left(\frac{(\text{pre SWIR} + \text{pre Red}) - (\text{pre NIR} - \text{pre Blue})}{(\text{pre SWIR} + \text{pre Red}) - (\text{pre NIR} - \text{pre Blue})} \right) \quad (13)$$

$$\text{postBSI} = \left(\frac{(\text{post SWIR} + \text{post Red}) - (\text{post NIR} - \text{post Blue})}{(\text{post SWIR} + \text{post Red}) - (\text{post NIR} - \text{post Blue})} \right) \quad (14)$$

$$\text{dBSI} = \text{post BSI} - \text{pre BSI} \quad (15)$$

The BSI is used in soil and landuse type classifications and is not widely recognised as useful for fire severity classification. Similarly, previous studies indicate dNBR outperforms dNDVI in fire severity classification (Veraverbeke et al., 2010; Collins et al., 2018). However, these reflectance-based indices were included as they are the closest analogous reflectance-based indices to the fractional cover indices of change in bare cover and change in total vegetative cover. Therefore, the relative value of fractional cover compared to reflectance-based indices will be comprehensively evaluated.

2.5. Random forest modelling

Random forest generates an ensemble of regression trees, through a bootstrap aggregated (bagging) sampling of training data. At each node in a decision tree, a random selection (parameter *mtry*) of predictor variables is evaluated for their ability to split training data into response classes, with the variable leading to the most homogeneous classification being selected. The Gini Index (Gini impurity criterion) measures the similarity of a given element with respect to the rest of the classes and is used to find the best split selection at each node of the decision tree (Breiman, 2001a). The RF algorithm uses an out-of-bag (OOB) estimate of accuracy, whereby for each tree (of *n* trees in the 'forest'), a random subset of 70% of the data is used to build the model

with the remaining 30% used to test the model (i.e. bootstrap with replacement). The final OOB estimate is the average accuracy across all individual trees (Breiman, 2001a). Model training and predictions were undertaken using the caret package in R (Kuhn, 2019). The number of trees was 500 and number of predictor variables at each node was the square root of the number of variables used in the model (i.e. default values; Breiman and Cutler, 2018).

Separate models were systematically compared for each candidate severity index, alone and in combination, for each fire and for all fires combined. All models were trained on a subset of 70% of the full dataset, while the remaining 30% was used to test the performance of the model. Models for each fire were also tested separately in a cross-validation assessment, with true independent data (Millard and Richardson, 2015). This showed the predictive power of the model in mapping severity of novel fires (not used to train the model), to indicate the accuracy expected in an operational application. Analysis of variance and Tukey's Honestly Significant Difference post-hoc tests were used to assess the difference in balanced accuracy statistics between trained and predictive models and across severity classes.

Spurious results occurred when the training data for the unburnt class from all fires was included in the cross-validation models. The variation in vegetation and soil properties at the locations of each fire across the region is expected to be high, given the variation in climate. This would have a strong influence on the fire severity indices, such that 'unburnt' could not be classified in a standardised way across all fires. To improve the classification accuracy, unburnt training data was selected from within 2–4 km of the fire affected area around the target fire only. A similar approach is taken under the 'Monitoring Trends in Burn Severity' (MTBS) program in the US, where a phenology offset is manually selected of a homogenous unburnt area outside the fire perimeter and applied to the dNBR (Key, 2006). This is particularly important for comparing severity thresholds among different fires (Park et al., 2014). Training data for the low to extreme severity classes from all other fires (excluding the target fire) were combined with unburnt training data from the target fire. Validation data was the full sampling data of the target fire, which was completely independent from the training data.

Classification accuracy was primarily assessed using the balanced accuracy metric for all classes, which is more appropriate than overall accuracy for multi-class problems and data with unbalanced sampling (Brodersen et al., 2010). Balanced accuracy for each class is the average of the sensitivity (true positive rate) and specificity (true negative rate), based on the confusion matrix. The Kappa statistic was also used to determine the statistical agreement between the model and the validation data and to compare performance between models (Story and Congalton, 1986). The Kappa statistic compares observed accuracy against expected accuracy given by chance. Kappa statistic values range from -1 to +1, where +1 represents perfect agreement, and values < 0 represent a performance no better than random. Models with Kappa values > 0.75 are generally considered to be excellent, while models with Kappa values < 0.4 are considered to be poor (Allouche et al., 2006).

The mean decrease in the Gini Index was calculated for the full model trained on all fires combined, to assess variable importance on classification accuracy. An additional raster was generated with random numbers in a uniform distribution between 0 and 1 and included in calculating the mean decrease in the Gini Index to evaluate the relative importance of each fire severity index on the RF model performance above the level of chance.

2.6. Effect of landscape variables on RF classification accuracy

Some limitations are expected of a satellite-based approach to mapping fire effects obscured by dense canopy or shadow cast by steep slopes or rugged terrain. We aimed to quantify the nature of these limitations by assessing the effect of canopy density and topographic

ruggedness on the misclassification of fire severity in our study fires. Foliage projective cover (FPC), an indicator of long-term canopy density (robust to seasonal variation), was developed from a model, calibrated to ground-based measurements using fractional cover transects and validated against LiDAR data (Armston et al., 2009). A high resolution version of FPC developed by Fisher et al. (2016) was used, which is calculated from SPOT5 satellite imagery across four consecutive summers (2008–2011). The Terrain Ruggedness Index (TRI) is a standard index defined as the mean difference between a central pixel and its surrounding cells (Wilson et al., 2007); a measure of topographic complexity indicating relative steep or flat areas. The TRI was derived from the 1 s Shuttle Radar Topography Mission (SRTM) Digital Elevation Model (DEM, 30 m pixel size).

Differences in community structure may also affect remotely sensed fire severity indices between different vegetation type (Hammill and Bradstock, 2006; Miller and Thode, 2007; Tran et al., 2018). For example, dense canopy and mid-story layers may obscure understory layers from satellite view. Vegetation type (OEH, 2017) across the case study fires was dominated by dry sclerophyll, comprising over 70% of the sampling points (Supplementary material, Fig. 2.1), with other vegetation types having low representation across the study fires. Therefore, we restricted our analysis to dry sclerophyll forest to avoid confounding the effects of vegetation community with discrete fire events. As such, the effect of vegetation type on misclassification could not be systematically tested.

The association between the landscape factors (TRI and FPC) and the accuracy of the RF predicted severity class were analysed separately using random effects-mixed models, with fire and API polygon as nested random effects. In each model, the landscape factor was the response variable and the RF predicted severity class accuracy was the explanatory variable with 3 levels:

- under-predicted, where the RF model predicted a lower severity class compared to the API validation class;
- correct, where the RF model predicted the same class as the API validation class, and
- over-predicted, where the RF model predicted a higher severity class compared to the API validation class.

3. Results

3.1. Reflectance and fractional cover model comparisons

A total of 105,388 data points were generated across the eight study fires for training and validation of the random forest models. The distribution of data points across the severity classes was approximately 10% in each of low and moderate severity classes, approximately 20% in each of high and extreme severity classes, and approximately 40% in unburnt (Table 2). The high proportion of unburnt sample points for each fire was required to balance the predictive models, where the burnt data points from all other fires were combined with the unburnt data points from the target fire. The low sample numbers in the low and moderate classes reflect the natural occurrence of these classes across the wildfires, which often occurred as narrow transitional zones between more homogenous zones of unburnt and complete canopy scorch (high severity). Each fire contributed approximately 10,000 data points, while the larger fires (Pilliga and Wollemi) contributed around 20,000 data points (Supplementary material, Fig. 2.1).

Timeseries plots used to examine the variation in the NBR and fractional cover values before and after each fire demonstrated greater separation between pre- and post-fire pixels values with increasing severity for NBR, bare and green cover fractions for the Sir Ivan fire (Fig. 2a and b). The other study fires showed the same trends in NBR and fractional cover pixel values. The very distinct change in bare cover due to the fire was well above the variation unrelated to fire, which justified creating a severity index based on the change in bare cover

alone. By contrast, there was larger variation in both green and non-green cover that was unrelated to fire, but a distinctive drop in green cover in each severity class. The drop in non-green cover was observed only in the extreme severity class because fire will convert a proportion of green fractions to non-green fractions in addition to partial consumption of non-green fractions in moderate and high severity classes. Green and non-green cover types were combined for a severity index based on the change in total cover.

The comparison of each fire severity index, alone and in combination with all other indices, demonstrates the comparable accuracy of each fractional cover index to each reflectance-based index (Table 3). There was a consistent trend of increasing Kappa scores and balanced accuracy with models that included at least one of the NBR-based indices with at least one of the fractional cover-based indices. The RF model with the highest Kappa scores and highest balanced accuracy values was given by the model that included both fractional cover indices with both NBR-reflectance indices ($Class \sim dFCB + RdFCT + dNBR + RdNBR$; Table 3). There were no significant differences in the Kappa statistics and balanced accuracy values between the trained and predictive versions (Fig. 3). Balanced accuracy statistics were significantly different between fire severity classes (Fig. 3). There were larger confidence intervals for the balanced accuracy statistic for the low severity class. This may have been largely driven by one fire, as Sir Ivan had a balanced accuracy of 0.58 for the low severity class, while the average of all other fires was 0.80. The relatively lower classification accuracy of the moderate severity class may have been related to limitations of the validation data rather than accuracy per se.

The accuracy statistics results for the random forest model trained on all 8 study fires (Table 4) were consistent with the results of the cross-validation analysis (Supplementary material Tables 4.1–4.8b). In both cases, the unburnt and extreme classes were consistently predicted with very high accuracy (mean > 0.95), the low and high classes were predicted with high accuracy (mean > 0.80 and > 0.85 respectively), while the moderate class was predicted with relatively lower accuracy (mean > 0.70).

The mean decrease in the Gini index for the full model (training and validation data combined) ranked the fire severity indices in decreasing importance as $RdNBR > dFCB > dNBR > RdFCT > \text{random}$. $RdNBR$ and $dFCB$ had a gini index values ~5 times greater than the random values, whereas $RdFCT$ was only slightly better than random (Fig. 4).

3.2. Effect of landscape factors on RF classification accuracy

A total of 41,437 sampling points within dry sclerophyll vegetation type were analysed for the influence of foliage projective cover and topographic roughness on RF classification accuracy. There was large variation in foliage projective cover (10–95%) and topographic roughness values (0–47) across the study fires (Supplementary material, Fig. 2.3) providing a diverse range of conditions to test misclassification. There were 5627 points under-predicted (14%), 33,175 points correctly predicted (80%) and 2635 points over-predicted (6%). Of the 14% of sampling points that were under-predicted, the low severity class represented the highest proportion; 10–20% higher than moderate, high and extreme severity classes (Fig. 5). Of the 6% of sampling points that were over-predicted, the moderate severity class represented the highest proportion; > 50% higher than the unburnt, low and high severity classes (Fig. 5).

Significantly higher foliage projective cover and topographic roughness values were associated with the RF model under-predicting the severity class compared to correctly predicted classifications (Type II Wald chi square tests $p = 0.04$ and 0.0006 , respectively, Figs. 6, 7 and Table 5). Higher foliage projective cover was also related to RF model over-predicting the severity class (Fig. 6).

There was improvement in visual assessment of commission error when additional unburnt training data was taken from topographic shadows (Supplementary material, Fig. 5.5). However, the accuracy of

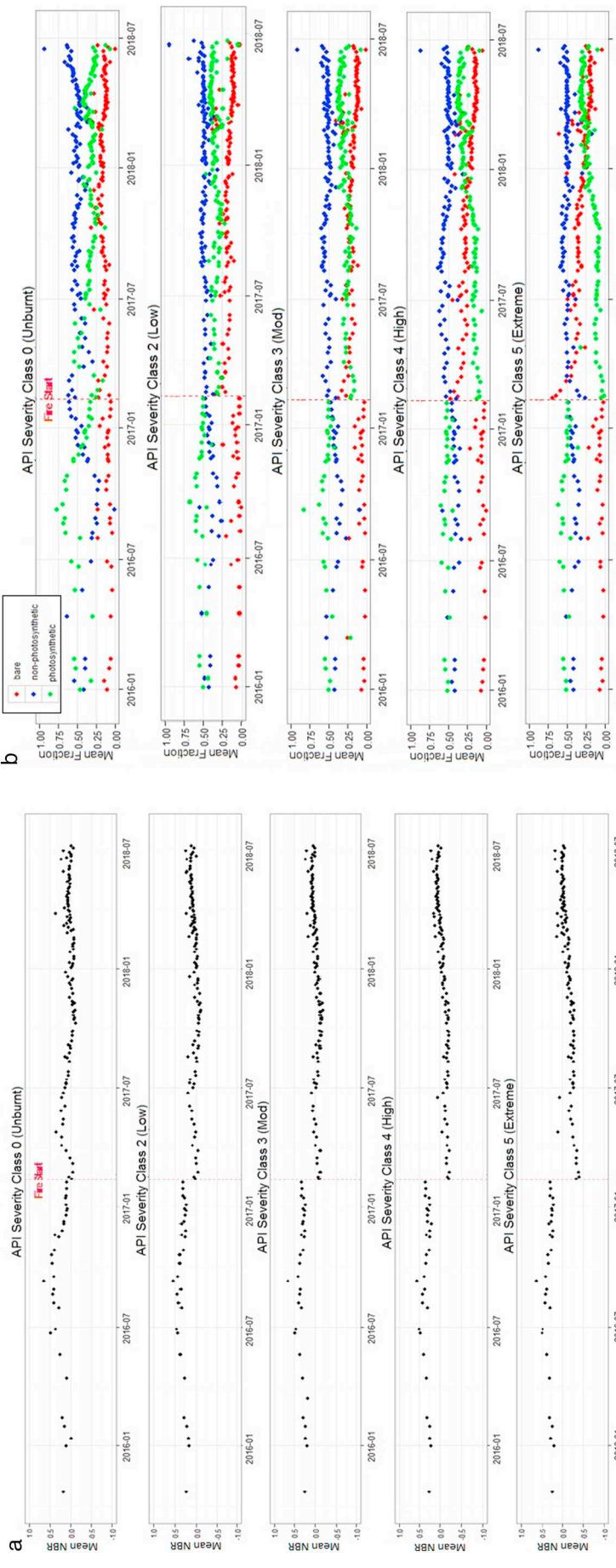


Fig. 2. Timeseries plots of Sentinel 2 values across aerial photo interpretation (API) validation fire severity classes at the Sir Ivan fire for a) mean Normalised Burn Ratio (NBR) and b) mean fractional cover values (red = bare, blue = non-photosynthetic, green = photosynthetic). Consistent patterns were observed for the other study fires. Severity classes include unburnt, low severity (burnt understorey/unburnt canopy), moderate severity (partial canopy scorch), high severity (full canopy scorch/partial consumption), extreme severity (full canopy consumption). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 3

Comparison of the kappa and balanced accuracy statistics of the differenced and relativised normalised burn ratios (dNBR, RdNBR), differenced bare fractional cover (dFCB), relativised differenced total fractional cover (RdFCT) and the combined models examined in the study.

Predictor variables	Kappa	Balanced accuracy statistics				
		Unburnt	Low	Mod	High	Extreme
1. dBSI	0.424	0.599	0.514	0.504	0.541	0.617
2. dNBR + dNDVI	0.502	0.799	0.712	0.576	0.711	0.833
3. dFCB + dNDVI	0.579	0.913	0.705	0.602	0.751	0.784
4. dNBR	0.583	0.89	0.693	0.564	0.773	0.834
5. RdNBR + dNDVI	0.592	0.823	0.697	0.587	0.767	0.945
6. RdNBR + dBSI	0.6	0.848	0.671	0.591	0.76	0.941
7. dNDVI	0.603	0.807	0.554	0.529	0.773	0.743
8. RdFCT	0.618	0.929	0.647	0.5	0.842	0.807
9. dNBR + dFCB	0.634	0.936	0.719	0.631	0.772	0.831
10. dFCB	0.636	0.933	0.676	0.6	0.844	0.792
11. RdNBR	0.641	0.859	0.67	0.572	0.836	0.959
12. dNBR + RdFCT	0.641	0.933	0.715	0.6	0.802	0.83
13. dFCB + RdFCT	0.643	0.933	0.675	0.599	0.839	0.813
14. dFCB + dBSI	0.653	0.911	0.699	0.624	0.798	0.905
15. dNDVI + dBSI	0.656	0.812	0.679	0.593	0.757	0.858
16. dNBR + RdNBR	0.668	0.881	0.717	0.625	0.803	0.969
17. dNDVI + dNBR + RdNBR	0.668	0.862	0.743	0.619	0.817	0.946
18. dBSI + dNBR + RdNBR	0.672	0.881	0.741	0.636	0.815	0.944
19. dNDVI + dBSI + dNBR	0.681	0.873	0.76	0.635	0.831	0.94
20. dFCB + RdFCT + dNBR	0.682	0.945	0.741	0.613	0.823	0.858
21. dNBR + dBSI	0.697	0.831	0.706	0.592	0.767	0.929
22. dNDVI + dBSI + RdNBR	0.697	0.885	0.751	0.64	0.833	0.9568
23. RdNBR + RdFCT	0.705	0.915	0.666	0.604	0.869	0.959
24. RdNBR + dFCB	0.716	0.923	0.711	0.645	0.851	0.959
25. dNDVI + dBSI + dNBR + RdNBR	0.743	0.906	0.788	0.665	0.869	0.962
26. dFCB + RdFCT + RdNBR	0.754	0.936	0.743	0.619	0.889	0.969
27. RdFCT + dNBR + RdNBR	0.759	0.942	0.745	0.647	0.876	0.975
28. dFCB + dNBR + RdNBR	0.791	0.959	0.805	0.657	0.886	0.976
29. dFCB + RdFCT + dNBR + RdNBR	0.799	0.962	0.803	0.652	0.895	0.976

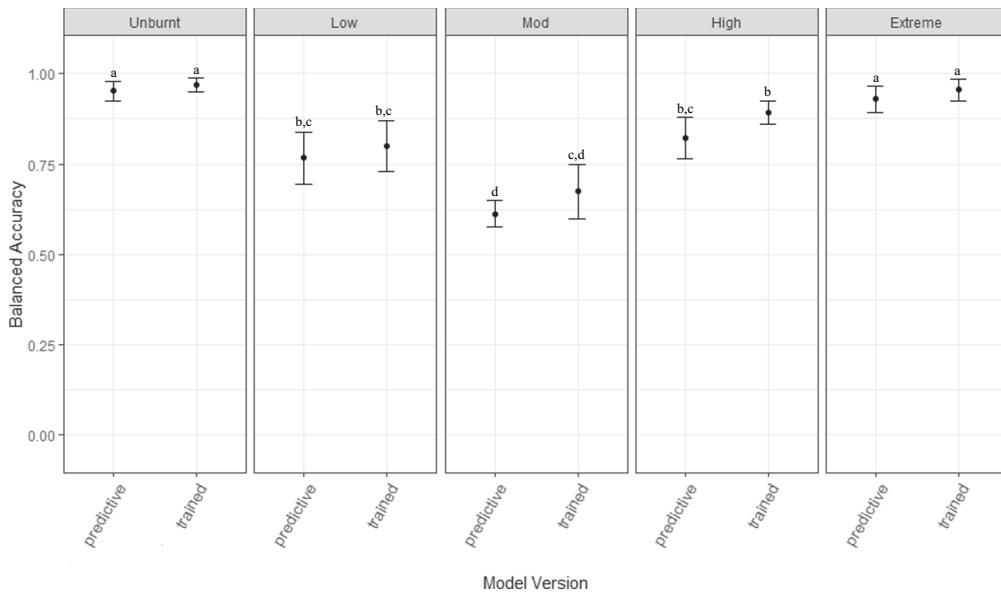


Fig. 3. Comparison of balanced accuracy statistics, with mean and 95% confidence intervals, for predictive (excludes training data from target fire) and trained (includes training data from target fire) models. Different superscript letters indicate a statistical significance at $p > 0.05$.

the RF predicted severity classification into areas of topographic shadow was not statistically validated, due to the topographic shadow also being present in the high-resolution ADS photography used to generate the validation data.

4. Discussion

Our study demonstrates that fire severity can be mapped with high accuracy (mean accuracy of 88% across all classes) using indices

derived from Sentinel 2 imagery with random forest supervised classification. Classification accuracy was very high (> 85%) for unburnt and the higher severity classes (i.e. canopy scorching and consuming fires), but slightly lower for the lower severity classes (70–80%). These findings agree with previous work using Landsat imagery and high-resolution unmanned aerial vehicle imagery, which have shown that machine learning classifiers are well suited to broad-scale mapping of fire severity with remotely sensed imagery (Meddens et al., 2016; Meng et al., 2017; Collins et al., 2018).

Table 4

Confusion matrix and balanced accuracy statistic for the random forest model trained on all 8 case study fires.

		Reference – aerial photo class					Balanced accuracy
		Unburnt	Low	Mod	High	Extreme	
Predicted – RF model	Unburnt	32,297	886	166	112	9	0.950
	Low	2385	7341	1173	473	4	0.848
	Moderate	280	1508	3613	4011	47	0.749
	High	218	170	1464	20,538	840	0.878
	Extreme	13	1	19	747	20,098	0.968
Total n samples		35,193	9906	6435	25,881	20,998	98,413

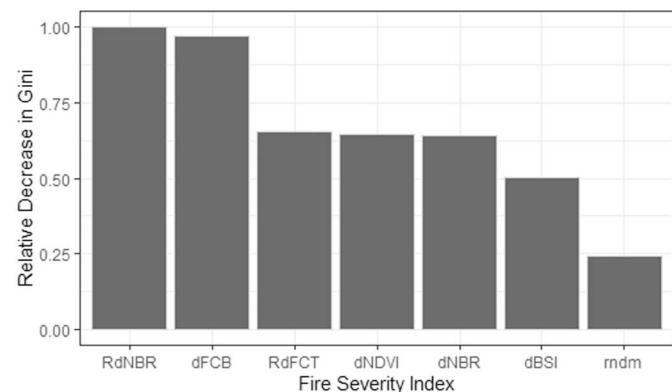


Fig. 4. Ranked Gini index indicating the relative importance of each fire severity index on random forest classification accuracy, with 1 representing the most important variable. The fire severity indices are the differenced and relativised normalised burn ratios (dNBR, RdNBR), differenced bare fractional cover (dFCB), relativised change in total fractional cover (RdFCT). Random was a raster created from randomly generated values, to indicate the relative importance of each index above the level of random chance.

Fractional cover indices (Table 3, models 8 and 10) each performed better in classifying unburnt compared to each of the NBR indices (Table 3, models 4 and 11). Likewise, the fractional cover indices consistently out-performed the reflectance analogues (dNDVI and dBsi) in classifying unburnt in each equivalent model (e.g. model 2 vs 12, models 9 vs 21, models 18 vs 20, models 25 vs 29). The combination of dNBR and RdNBR (Table 3, model 5) performed better than either index alone, which has been suggested by other studies (Miller and Thode, 2007). The addition of fractional cover indices to the combined dNBR and RdNBR model (models 25, 26 and 27) improved the Kappa score by 9–13% and improved the classification accuracy of each of the fire severity classes between 2 and 8%. The addition of the reflectance-based analogues of the fractional cover variables (dNDVI and dBsi) to the combined dNBR and RdNBR model also improved the Kappa and

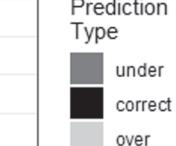
balanced accuracy statistics, but to a lesser degree than the fractional cover variables. This demonstrates the value of the fractional cover data for mapping fire severity.

The practical application of the severity maps in fuel hazard assessment and fire behaviour analysis require clear understanding of how mapping approaches perform across different severity classes and landscapes. Our mapping approach, using random forest classification of Sentinel 2-based reflectance and fractional cover indices, produced highly accurate mapping of unburnt and high severity wildfire in landscapes with moderate canopy density and low topographic ruggedness. Higher rates of misclassification occurred for the low and moderate fire severity classes and in areas of dense canopy cover and rugged topography.

The lower accuracy statistics of the moderate fire severity class likely reflects a difference in the scale of the training data and the resolution of the satellite imagery, which is a common challenge in remote sensing validation studies (Wu and Li, 2009; Walker et al., 2016). Moderate severity is defined as partial canopy scorch, which appears as a mixture of green, orange and brown in the tree canopies in high resolution ADS photography (Table 2). The Sentinel 2-pixel size (10 m) in some cases will be smaller than a single mature tree canopy. In many instances the RF modelling tended to separate out pixels representing low (unburnt canopy), moderate (partial canopy scorch) and high (full canopy scorch) severity (Supplementary material, Fig. 5.1) within an area hand digitised as a moderate severity sample. Although the RF modelling adhered to the API severity classification rules at the pixel-scale, the statistical analysis against validation data determined many of these points were misclassified. Severity classification using Landsat imagery (30 m resolution, 16 day site revisit time) similarly had lower accuracy for the moderate severity class (partial canopy scorch) compared to the other severity classes (Collins et al., 2018), which suggests discriminating the moderate class is challenging. The moderate class had the lowest representation of sampling points of all the severity classes in the training and validation datasets, due to the naturally lower occurrence of this class within the study fires (Table 4). An increased volume of training data for the moderate class may improve accuracy given the power of machine learning. However, a secondary classification approach may also be required, such as object-based image analysis of the RF predicted severity classification, to amalgamate highly variable class areas into the single moderate class with greater consideration of neighbourhood context patterns (Lebourgues et al., 2017).

In open canopy vegetation types such as grassy woodlands within the low severity class, the RF modelling tended to isolate unburnt tree canopies within a background of burnt understory (Supplementary material, Fig. 5.2). The Sir Ivan fire contained a high proportion of grassland and open grassy woodland (Supplementary material Table 2.1). Mapping severity of burnt grasslands is unlikely to be meaningful from an ecological or fuel analysis viewpoint because in

Fig. 5. The proportion of sampling points within dry sclerophyll vegetation across severity classes in each classification accuracy category; under-predicted (random forest, RF, algorithm predicted a lower severity class compared to the validation data), correctly predicted and over-predicted (RF predicted a higher severity class compared to the validation data). Unburnt could not be under-predicted and extreme severity could not be over-predicted (i.e. NA).



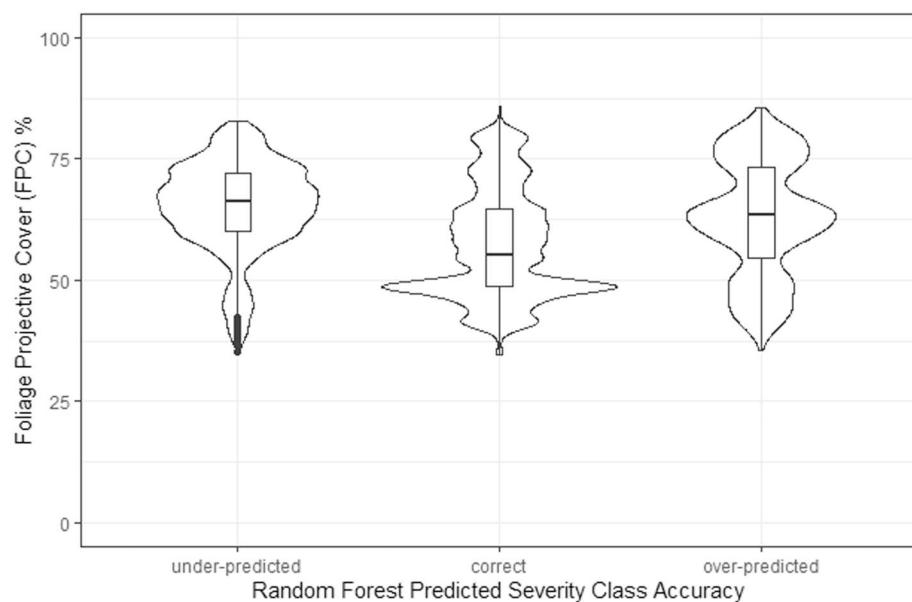


Fig. 6. The accuracy of the random forest (RF) predicted severity class as a function of percent foliage projective cover (FPC), within dry sclerophyll vegetation type. The violin density plots display the probability density of the data, with a boxplot of the median (horizontal line), interquartile range (black rectangle), and the first and third quartile ± 1.5 of the interquartile range (black vertical lines). Under-predicted represents where the RF model predicted a lower severity class compared to the aerial photo-derived validation severity class. Over-predicted represents where the RF model predicted a higher severity class compared to the aerial photo-derived validation severity class. Severity classes include unburnt, low severity (burnt understory/unburnt canopy), moderate severity (partial canopy scorch), high severity (full canopy scorch/partial consumption), extreme severity (full canopy consumption).

general, grasslands will either burn or not burn. This highlights the importance of interpreting the RF severity maps alongside local knowledge of vegetation types to derive information for use in other applications such as fuel hazard assessment and fire behaviour modelling. Further research is required for more appropriate classification of burnt grasslands.

Higher FPC was associated with under and over-prediction of the RF predicted severity class compared to the API validation class (Fig. 6). High canopy density may obscure the burnt understorey of low severity classes, resulting in under-prediction errors. Greater canopy gap fractions under lower FPC would allow more of the mid-story and understorey layers to be visible to the sentinel 2 satellites, resulting in higher classification accuracy. Over-prediction errors under higher FPC were predominately caused by moderate severity being incorrectly predicted as high severity (Fig. 5, Supplementary material, Tables 4.1a–4.8a). Therefore, further research to improve classification of the moderate severity class may reduce these over-prediction errors.

Higher terrain roughness was associated with under-prediction of fire severity (Fig. 7). Over-prediction of unburnt as low severity was

Table 5

Results of separate generalised linear mixed model (GLMM) analysis, with random (study fire and aerial photo interpretation, API, polygon) and fixed effects (random forest, RF, accuracy factor, with 3 levels, under-predicted, correctly predicted and over-predicted) for response variable 1. Foliage Projective Cover (FPC) and 2. Terrain roughness index (TRI).

Model	Fixed effects	Estimate	Std. Error	t value
FPC	(Intercept)	59.817	3.686	16.228
	Over-predicted	0.206	0.118	1.740
	Under-predicted	-0.087	0.077	-1.125
TRI	(Intercept)	0.913	0.115	7.961
	Over-predicted	-0.014	0.004	-3.737
	Under-predicted	-0.005	0.002	-2.153

also observed in areas with a large amount of topographic shadow, such as the Wollemi 695 fire. Application of the mapping to future fires in areas of high topographic shadow may require unburnt training data to be manually boosted from topographic shadow areas well outside the fire-affected area.

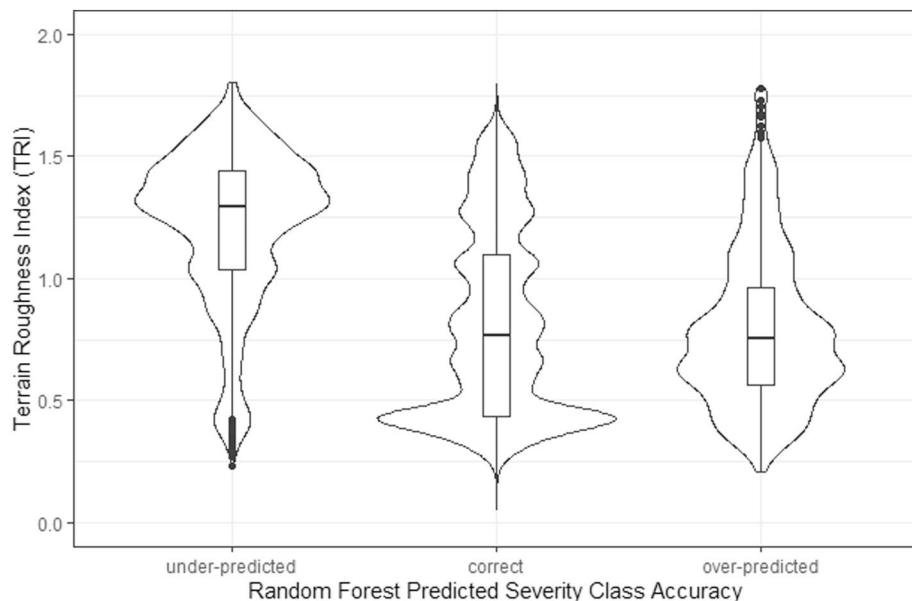


Fig. 7. The accuracy of the random forest (RF) predicted severity class as a function of terrain roughness index (TRI), within dry sclerophyll vegetation type. The violin density plots display the probability density of the data, with a boxplot of the median (horizontal line), interquartile range (black rectangle), and the first and third quartile ± 1.5 of the interquartile range (black vertical lines). Under-predicted represents where the RF model predicted a lower severity class compared to the aerial photo-derived validation severity class. Over-predicted represents where the RF model predicted a higher severity class compared to the aerial photo-derived validation severity class. Severity classes include unburnt, low severity (burnt understory/unburnt canopy), moderate severity (partial canopy scorch), high severity (full canopy scorch/partial consumption), extreme severity (full canopy consumption).

Spatial auto-correlation is inherently present in fire severity data (Taylor et al., 2014; Collins et al., 2018). Assumptions of independence are violated when training and validation data are taken from the same homogenous areas, as used in the sampling design of this study. This may result in optimistic bias in regression models, where accuracy statistics are inflated (Hammond and Verbyla, 1996). However, the balanced accuracy statistics reported were from the cross-validation models, which ensured validation points were drawn from a sample that was independent of the training data to avoid optimistic bias (Millard and Richardson, 2015). Therefore, spatial autocorrelation is unlikely to have had any effect on the predictive performance of the RF models.

The seasonal effect of solar elevation angle was not examined in this study. The lower solar elevation angle through winter is known to result in greater topographic and tree shadows in optical satellite imagery (Rogan and Franklin, 2001; Lyndersen et al., 2016). As such, the accuracy of these RF algorithms trained on wildfires require further examination for application to HR fires, which typically occur in late autumn and winter months. Incorporating training data from HR fires may be necessary to increase accuracy of the RF algorithm application to HR fires (authors' unpublished data).

Sentinel 2 satellite imagery offers a progression from Landsat for future mapping of fire severity. The higher spatial resolution of Sentinel 2 allows for greater detailed mapping, while the higher temporal resolution allows for greater likelihood of obtaining imagery free of cloud contamination. The comparative accuracy of Sentinel 2 and Landsat-based severity mapping with RF algorithms suggests fire severity mapping from these different image sources can be used complimentarily, for example in creating fire severity mapping archives. Ongoing technology developments in the spatial, temporal and spectral resolution of satellite imagery have the potential to further improve the accuracy of remote sensing of fire severity. Very high-resolution satellite technology in the future that maintains the spectral resolution required for remote sensing of fire (i.e. NIR and SWIR wavelengths) may improve the classification of low severity fire in dense canopy forests. Future development in machine learning techniques that incorporates computer vision approaches, such as U-Net (Ronnenberger et al., 2015; Yao et al., 2019) may also allow for greater sophistication in context-based machine learning. This may help to overcome the limitations of the pixel-based approach used in this study, particularly for improving classification of low severity in grassy woodlands, and the moderate severity class. Integration of the modelling developed in this study with other quantitative remote sensing data types, such as radar (Addison and Oommen, 2018) and LiDAR (Hu et al., 2019), may be valuable for providing insights into the effect of severity on fuel hazard in different vegetation types.

Operationalisation of the sentinel 2 based fire severity mapping is currently being developed and tested at a state-wide scale for NSW, which involves NSW Rural Fire Service supplying spatial data of fire event date and locations to the NSW Department of Planning Industry and Environment (DPIE) to conduct the automated imagery processing and severity classification. Application of the method presented here is suited to mapping the extent and severity of fires of known date and locations. Remote sensing detection of fires of unknown date and location is a separate research question that requires very different practical solutions compared to mapping fire severity and was not the focus of this project. Fire event records held by fire management agencies commonly have incorrect geolocation data. Rather than exclude a large proportion of these databases, Benali et al. (2016) found that 2 km was an optimal threshold for including records with relatively small spatial inaccuracy. A suitable buffer distance threshold around the recorded fire location must be set to enable the RF predicted severity model to map the boundary of fires with a spatial inaccuracy tolerance (e.g. ~2 km). Further research is ongoing, to continue to improve the method and state-wide operational processing system. While the method described is not appropriate for detecting fire events

at unknown locations across large geographic regions, there is future potential to integrate multiple automated systems for a more complete remote sensing fire mapping system.

CRediT authorship contribution statement

Rebecca Gibson:Conceptualization, Methodology, Investigation, Formal analysis, Writing - original draft.**Tim Danaher:**Resources, Supervision, Writing - review & editing.**Warwick Hehir:**Resources, Funding acquisition, Writing - review & editing.**Luke Collins:**Conceptualization, Methodology, Writing - review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was supported by funding provided by NSW RFS, as part of the Fire Extent and Severity Mapping project, with in-kind contributions from the Remote Sensing & Regulatory Mapping Team, NSW DPIE. Specialist remote sensing advice was provided by colleagues within the Joint Remote Sensing Research Program (JRSRP; a partnership between DPIE, QLD Government Remote Sensing Centre, University of NSW and the University of Queensland) including Adrian Fisher, Jim Watson, Lisa Collett, Leo Hardtke and Peter Scarth. Field assistance was provided by Adrian Fisher, Geoff Horn (Remote Sensing & Regulatory Mapping, DPIE), Warwick Hehir (NSW RFS), colleagues from Macquarie University, NPWS and other DPIE staff. Liz Tasker (Conservation Restoration Science, DPIE) provided specialist fire ecology advice and valuable comments on an early draft of the manuscript.

Data accessibility

All images from Sentinel 2 satellites and derived products are routinely pre-processed with standardised surface reflectance corrections suited to eastern Australia (Flood et al., 2013a and stored by the OEH Remote Sensing & Analysis Team. The fire severity mapping algorithms and training data used in this study is held by the OEH Remote Sensing & Analysis Team. Automated systems are in development by OEH in collaboration with RFS, who will be the primary data custodian for NSW fire extent and severity data. In the interim, fire severity mapping for specific fire events in NSW since mid-2016 (Sentinel 2 imagery availability) may be requested by email to Spatial.imagery@environment.nsw.gov.au.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.rse.2020.111702>.

References

- Addison, P., Oommen, T., 2018. Utilizing satellite radar remote sensing for burn severity estimation. *Int. J. Appl. Earth Obs. Geoinf.* 73, 292–299.
- Allouche, O., Tsoar, A., Kadmon, R., 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *J. Appl. Ecol.* 43, 1223–1232.
- Armston, J.D., Denham, R., Danaher, T., Scarth, P., Moffet, T.N., 2009. Prediction and validation of foliage projective cover from Landsat-5 TM and Landsat-7 ETM+ imagery. *J. Appl. Remote. Sens.* 3, 033540.
- Belgiu, M., Dragut, L., 2016. Random forest in remote sensing: a review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* 114, 24–31.
- Benali, A., Russo, A., Sa, A.C.L., Pinto, R.M., Price, O., Koutsias, N., Pereira, J.M.C., 2016.

- Determining fire dates and locating ignition points with satellite data. *Remote Sens.* 8, rs8040326.
- Bradstock, R.A., Hammill, K.A., Collins, L., Price, O., 2010. Effects of weather, fuel and terrain on fire severity in topographically diverse landscapes of south-eastern Australia. *Landscape Ecol.* 25, 607–619.
- Breiman, L., 2001a. Random forests. *Mach. Learn.* 5, 5–32.
- Breiman, L., 2001b. Statistical modeling: the two cultures. *Stat. Sci.* 16, 199–231.
- Breiman, L., Cutler, A., 2018. Breiman and Cutler's random forest for classification and regression (package 'randomForest'). Available at: <https://CRAN.R-project.org/package=randomForest>.
- Brodersen, K.H., Ong, C.S., Stephan, K.E., Buhmann, J.M., 2010. The balanced accuracy and its posterior distribution. In: International Conference on Pattern Recognition. IEEE Computer Society, Istanbul, Turkey, pp. 1051–1051.
- Cansler, C.A., McKenzie, D., 2014. Climate, fire size and biophysical setting control fire severity and spatial pattern in the northern Cascade Range, USA. *Ecol. Appl.* 24, 1037–1056.
- Cary, G.J., Davies, I.D., Bradstock, R.A., Keane, R.E., Flannigan, M.D., 2017. Importance of fuel treatment for limiting moderate-to-high intensity fire: findings from comparative modelling. *Landscape Ecol.* 32, 1473–1483.
- Chafner, C.J., Noonan, M., Macnaughton, E., 2004. The post-fire measurement of fire severity and intensity in the Christmas 2001 Sydney wildfires. *Int. J. Wildland Fire* 13.
- Collins, L., Bradstock, R.A., Penman, T.D., 2014. Can precipitation influence landscape controls on wildfire severity? A case study within temperate eucalypt forests of south-eastern Australia. *Int. J. Wildland Fire* 23, 9–20.
- Collins, L., Griffioen, P., Newell, G., Mellor, A., 2018. The utility of random forests in Google Earth Engine to improve wildfire severity mapping. *Remote Sens. Environ.* 216, 374–384.
- Coppoletta, M., Merriam, K.E., Collins, B.M., 2016. Post-fire vegetation and fuel development influences fire severity patterns in reburns. *Ecol. Appl.* 26, 686–699.
- 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.
- Edirweera, S., Pathirana, S., Danaher, T., Nichols, D., Moffett, T., 2013. Evaluation of different topographic corrections for Landsat TM data by prediction of foliage projective cover (FPC) in topographically complex landscapes. *Remote Sens.* 5, 6767–6789.
- Eidenshink, J., Schwind, B., Brewer, K., Zhu, Z., Quayle, B., Howard, S., 2007. A project for monitoring trends in burn severity. *Fire Ecol. Spec. Issue* 3, 3–21.
- Epting, J., Verbyla, D., Sorbel, B., 2005. Evaluation of remotely sensed indices for assessing burn severity in interior Alaska using Landsat TM and ETM+. *Remote Sens. Environ.* 96, 328–339.
- ESA, 2019. SENTINEL-2 MSI user guide. Available at: <https://earth.esa.int/web/sentinel/user-guides/sentinel-2-msi>.
- Farr, T.G., Rosen, P.A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Wener, M., Oskin, M., Burbank, D., Alsdorf, D., 2007. The Shuttle Radar Topography Mission. *Rev. Geophys.* 45 (RG2004).
- Fisher, A., Day, M., Gill, T., Roff, A., Danaher, T., Flood, N., 2016. Large-area, high-resolution tree cover mapping with multi-temporal SPOT5 imagery, New South Wales, Australia. *Remote Sens.* 8.
- Flannigan, M.D., Krawchuk, M.A., de Groot, W.J., Wotton, B.M., Gowman, L.M., 2009. Implications of changing climate for global wildland fire. *Int. J. Wildland Fire* 18, 483–507.
- Flood, N., 2017. Comparing Sentinel-2A and Landsat 7 and 8 using surface reflectance over Australia. *Remote Sens.* 9, 659.
- Flood, N., Danaher, T., Gill, T., Gillingham, S., 2013a. An operational scheme for deriving standardised surface reflectance from Landsat TM/ETM and SPOT HRG imagery for Eastern Australia. *Remote Sens.* 5, 83–109.
- Flood, N., Danaher, T., Gill, T., Gillingham, S., 2013b. An operational scheme for deriving standardised surface reflectance from Landsat TM/ETM and SPOT HRG imagery for Eastern Australia. *Remote Sens.* 5, 83–109.
- Gallant, J., Read, A., 2009. Enhancing the SRTM data for Australia. In: Proceedings of Geomorphometry, (ed. by Zurich, Switzerland).
- Ghimire, B., Rogan, J., Galiano, V.R., Panday, P., Neeti, N., 2012. An evaluation of bagging, boosting and random forests for land-cover classification in Cape Cod, Massachusetts, USA. *GI Sci. Remote Sens.* 49, 623–643.
- Guerschman, J.P., Scarth, P.F., McVicar, T.R., Renzullo, L.J., Malthus, T.J., Stewart, J.B., Rickards, J.E., Trevithick, R., 2015. Assessing the effects of site heterogeneity and soil properties when unmixing photosynthetic vegetation, non-photosynthetic vegetation and bare soil fractions for Landsat and MODIS data. *Remote Sens. Environ.* 161, 12–26.
- Hall, R.J., Freeburn, J.T., de Groot, W.J., Pritchard, J.M., Lynham, T.J., Landry, R., 2008. Remote sensing of burn severity: experience from western Canada boreal fires. *Int. J. Wildland Fire* 17, 476–489.
- Hammill, K.A., Bradstock, R.A., 2006. Remote sensing of fire severity in the Blue Mountains: influence of vegetation type and inferring fire intensity. *Int. J. Wildland Fire* 15, 213–226.
- Hammond, T., Verbyla, D., 1996. Optimistic bias in classification accuracy assessment. *Int. J. Remote Sens.* 7, 1261–1266.
- Hoy, E.E., French, N.H.F., Turetsky, M.R., Trigg, S.N., Kasischke, E.S., 2008. Evaluating the potential of Landsat TM/ETM+ imagery for assessing fire severity in Alaskan black spruce forests. *Int. J. Wildland Fire* 17, 500–514.
- Hu, T., Ma, Q., Su, Y., Battles, J.J., Collins, B.M., Stephens, S.L., Kelly, M., Guo, Q., 2019. A simple and integrated approach for fire severity assessment using bi-temporal airborne LiDAR data. *Int. J. Appl. Earth Obs. Geoinf.* 78, 25–38.
- Hudak, A.T., Robichaud, P.R., Evans, J.S., Clark, J., Lannom, K., Morgan, P., Stone, C., 2004. Field validation of burned area reflectance classification (BARC) products for post fire assessment. In: Remote Sensing for Field Users: Proceedings of the Tenth Forest Service Remote Sensing Applications Conference, pp. 13 ed by. (Salt Lake City, Utah).
- Hutchinson, M.F., McIntyre, S., Hobbs, R.J., Stein, J.L., Garnett, S., Kinloch, J., 2005. Integrating a global agro-climatic classification with bioregional boundaries in Australia. *Glob. Ecol. Biogeogr.* 14, 197–212.
- Keeley, J.E., 2009. Fire intensity, fire severity and burn severity - a brief review and suggested usage. *Int. J. Wildland Fire* 18, 116–126.
- Keeley, J.E., Syphard, A.D., 2016. Climate change and future fire regimes: examples from California. *Geosciences* 6.
- Key, C.H., 2006. Ecological and sampling constraints on defining landscape fire severity. *Fire Ecol.* 2, 34–59.
- Kolden, C.A., Abatzoglou, J.T., Smith, A.M.S., 2015. Limitations and utilisation of monitoring trends in burn severity products for assessing wildfire severity in the USA. *Int. J. Wildland Fire* 24, 1023–1028.
- Kuhn, M., 2019. Classification and regression training (package 'caret'). Available at: <https://CRAN.R-project.org/package=caret>.
- Lebourgeois, V., Dupuy, S., Vintrou, E., Ameline, M., Butler, S., Begue, A., 2017. A combined random forest and OBIA classification scheme for mapping smallholder agriculture at difference nomenclature levels using multisource data (simulated Sentinel-2 time series, VHRS and DEM). *Remote Sens.* 9 Article rs9030259.
- Lentile, L.B., Holden, Z.A., Smith, A.M.S., Falkowski, M.J., Hudak, A.T., 2006. Remote sensing techniques to assess active fire characteristics and post-fire effects. *Int. J. Wildland Fire* 15, 319–345.
- Lyndersen, J.M., Collins, B.M., Miller, J.D., Fry, D.L., Stephens, S.L., 2016. Relating fire-caused change in forest structure to remotely sensed estimates of fire severity. *Fire Ecol.* 12, 99–116.
- Marino, E., Guillen-Climent, M., Ranz, P., Tome, J.L., 2014. Fire severity mapping in Garajonay National Park: comparison between spectral indices. *FLAMMA* 7, 22–28.
- McCarthy, G., Moon, K., Smith, L., 2017. Mapping fire severity and fire extent in forest in Victoria for ecological and fuel outcomes. *Ecol. Manag. Restor.* 18, 54–64.
- Meddens, A.J.H., Kolden, C.A., Lutz, J.A., 2016. Detecting unburned areas within wildfire perimeters using Landsat and ancillary data across the northwestern United States. *Remote Sens. Environ.* 186, 275–285.
- Meng, R., Wu, J., Schwager, K.L., Zhao, F., Dennison, P.E., Cook, B.D., Brewster, K., Green, T.M., Serbin, S.P., 2017. Using high spatial resolution satellite imagery to map forest burn severity across spatial scales in a Pine Barrens ecosystem. *Remote Sens. Environ.* 191, 95–109.
- Millard, K., Richardson, M., 2015. On the importance of training data sample selection in random forest image classification: a case study in peatland ecosystem mapping. *Remote Sens.* 7, 8489–8515.
- Miller, J.D., Thode, A.E., 2007. Quantifying burn severity in a heterogeneous landscape with a relative version of the delta Normalized Burn Ratio (dNBR). *Remote Sens. Environ.* 109, 66–80.
- Miller, J.D., Knapp, E.E., Key, C.H., Skinner, C.N., Isbell, C.J., Creasy, R.M., Sherlock, J.W., 2009. Calibration and validation of the relative differenced Normalised Burn Ratio (RdNBR) to three measures of fire severity in the Sierra Nevada and Klamath Mountains, California, USA. *Remote Sens. Environ.* 113, 645–656.
- Morgan, P., Keane, R.E., Dillon, G.K., Jain, T.B., Hudak, A.T., Karau, E.C., Sikkink, P.G., Holden, Z.A., Strand, E.K., 2014. Challenges of assessing fire and burn severity using field measures, remote sensing and modelling. *Int. J. Wildland Fire* 23, 1045–1060.
- Muir, J., Schmidt, M., Tindall, D., Trevithick, R., Scarth, P., Stewart, J.B., 2011. In: Q.D.O.E.a.R. Management (Ed.), *Field Measurements of Fractional Ground Cover: A Technical Handbook Supporting Ground Cover Monitoring for Australia*. Australian Bureau of Agricultural and Resource Economics and Sciences, Canberra.
- OEH, 2017. The NSW State Vegetation Type Map: Methodology for a Regional Scale Map of NSW Plant Community Types. NSW Office of Environment and Heritage, Sydney, Australia.
- Park, S.E., Dillon, G.K., Miller, C., 2014. A new metric for quantifying burn severity: the relativised burn ratio. *Remote Sens.* 6, 1827–1844.
- Pettorelli, N., Laurance, W.F., O'Brien, T.G., Wegmann, M., Nagendra, H., Turner, W., 2014. Satellite remote sensing for applied ecologists: opportunities and challenges. *J. Appl. Ecol.* 51, 839–848.
- Rodriguez-Galiano, V.F., Ghimire, B., Rogan, J., Chica-Olmo, M., Rigol-Sanchez, J.P., 2012. An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS J. Photogramm. Remote Sens.* 67, 93–104.
- Rogan, J., Franklin, J., 2001. Mapping burn severity in southern California using spectral mixture analysis. In: International Geoscience and Remote Sensing Symposium, (ed. by Sydney, Australia).
- Ronnenberger, O., Fischer, P., Brox, T., 2015. U-Net: convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, (ed. by Cham, Switzerland).
- Scarth, P., Roder, A., Schmidt, M., 2010. Tracking grazing pressure and climate interaction - the role of Landsat fractional cover in time series analysis. In: Proceedings of the 15th Australasian Remote Sensing and Photogrammetry Conference, (ed. by Alice Springs, Australia).
- Sen, P.K., 1968. Estimates of the regression coefficient based on Kendall's tau. *J. Am. Stat. Assoc.* 63, 1379–1389.
- Smith, A.M.S., Eitel, J.U.H., Hudak, A.T., 2010. Spectral analysis of charcoal on soils: implications for wildland fire severity mapping methods. *Int. J. Wildland Fire* 19, 976–983.
- Soverel, N.O., Perrakis, D.D.B., Coops, N.C., 2010. Estimating burn severity from Landsat dNBR and RdNBR indices across western Canada. *Remote Sens. Environ.* 114, 1896–1909.
- Story, M., Congalton, R.G., 1986. Accuracy assessment: a user's perspective. *Photogramm. Eng. Remote. Sens.* 52, 397–399.

- Summers, D., Lewis, M., Ostendorf, B., Chittleborough, D., 2011. Visible near-infrared reflectance spectroscopy as a predictive indicator of soil properties. *Ecol. Indic.* 11, 123–131.
- Taylor, C., McCarthy, M.A., Lindenmayer, D.B., 2014. Nonlinear effects of stand age on fire severity. *Conserv. Lett.* 7, 355–370.
- Tran, B.N., Tanase, M.A., Bennett, L.T., Aponte, C., 2018. Evaluation of spectral indices for assessing fire severity in Australian Temperate Forests. *Remote Sens.* 10, 1680.
- Tubbesing, C.L., Fry, D.L., Roller, G.B., Collins, B.M., Fedorova, V.A., Stephens, S.L., Battles, J.J., 2019. Strategically placed landscape fuel treatments decrease fire severity and promote recovery in the northern Sierra Nevada. *For. Ecol. Manag.* 436, 45–55.
- Turner, M.G., Romme, W.H., Gardner, R.H., 1999. Prefire heterogeneity, fire severity and early postfire plant reestablishment in subalpine forests of Yellowstone National Park, Wyoming. *Int. J. Wildland Fire* 9, 21–36.
- Veraverbeke, S., Verstraeten, W., Lhermite, S., Goossens, R., 2010. Evaluating Landsat Thematic Mapper spectral indices for estimating burn severity of the 2007 Peloponnese wildfires in Greece. *Int. J. Wildland Fire* 19, 558–569.
- Verbyla, D.L., Kasischke, E.S., Hoy, E.E., 2008. Seasonal and topographic effects on estimating fire severity from Landsat TM/ETM+ data. *Int. J. Wildland Fire* 17, 527–534.
- Walker, D.A., Daniels, F.J.A., Alsos, I., Bhatt, U.S., Breen, A.L., Buchhorn, M., Bultmann, H., Druckenmiller, L.A., Edwards, M.E., Ehrich, D., Epstein, H.E., Gould, W.A., Ims, R.A., Meltofte, H., Raynolds, M.K., Sibik, J., Talbot, S.S., Webber, P.J., 2016. Circumpolar arctic vegetation: a heirarchic review and roadmap toward an internationally consistent approach to survey, archive and classify tundra plot data. *Environ. Res. Lett.* 11, 055005.
- Wilson, M.F.J., O'Connell, B., Brown, C., Guinan, J.C., Grehan, A.J., 2007. Multiscale terrain analysis of multibeam bathymetry data for habitat mapping on the continental slope. *Mar. Geod.* 30, 3–35.
- Wu, H., Li, Z.L., 2009. Scale issues in remote sensing: a review on analysis, processing and modeling. *Sensors* 9.
- Yao, H., Qin, R., Chen, X., 2019. Unmanned aerial vehicle for remote sensing applications - a review. *Remote Sens.* 11 Article rs11121443.
- Zhang, L., Ji, L., Wylie, B.K., 2011. Response of spectral vegetation indices to soil moisture in grasslands and shrublands. *Int. J. Remote Sens.* 32, 5267–5286.
- Zylstra, P., Bradstock, R.A., Bedward, M., Penman, T.D., Doherty, M.D., Weber, R.O., Gill, A.M., Cary, G.J., 2016. Biophysical mechanistic modelling quantifies the effects of plant traits on fire severity: species, not surface fuel loads, determine flame dimensions in Eucalypt forests. *PLoS One* 11, e0160715.