Efficacy of multi-season Sentinel-2 imagery for compositional vegetation classification

Paul Macintyre\textsuperscript{a,*}, Adriaan van Niekerk\textsuperscript{b}, Ladislav Mucina\textsuperscript{b,c}

\textsuperscript{a} School of Biological Sciences, The University of Western Australia, 35 Stirling Highway, Crawley WA 6009, Perth, Australia
\textsuperscript{b} Centre for Geographic Analysis, Department of Geography and Environmental Studies, Stellenbosch University, Private Bag X1, Matieland, 7602, Stellenbosch, South Africa
\textsuperscript{c} Iuka Chair in Vegetation Science and Biogeography, Harry Butler Institute, Murdoch University, 90 South Street, Murdoch, WA 6150, Perth, Australia

\textbf{ARTICLE INFO}

\textbf{Keywords:}
Mapping
Multispectral
Phenology
Tasselled cap
Vegetation classification

\textbf{ABSTRACT}

Vegetation maps are essential tools for the conservation and management of landscapes as they contain essential information for informing conservation decisions. Traditionally, maps have been created using field-based approaches which, due to limitations in costs and time, restrict the size of the area for which they can be created and frequency at which they can be updated. With the increasing availability of satellite sensors providing multispectral imagery with high temporal frequency, new methods for efficient and accurate vegetation mapping have been developed. The objective of this study was to investigate to what extent multi-seasonal Sentinel-2 imagery can assist in mapping complex compositional classifications at fine spatial scales. We deliberately chose a challenging case study, namely a visually and structurally homogenous scrub vegetation (known as kwongan) of Western Australia. The classification scheme consists of 24 target classes and a random 60/40 split was used for model building and validation. We compared several multi-temporal (seasonal) feature sets, consisting of numerous combinations of spectral bands, vegetation indices as well as principal component and tasselled cap transformations, as input to four machine learning classifiers (Support Vector Machines; SVM, Nearest Neighbour; NN, Random Forests; RF, and Classification Trees; CT) to separate target classes. The results show that a multi-temporal feature set combining autumn and spring images sufficiently captured the phenological differences between the classes and produced the best results, with SVM (74%) and NN (72%) classifiers returning statistically superior results compared to RF (65%) and CT (50%). The SWIR spectral bands captured during spring, the greenness indices captured during spring and the tasselled cap transformations derived from the autumn image emerged as most informative, which suggests that ecological factors (e.g. shared species, patch dynamics) occurring at a sub-pixel level likely had the biggest impact on class confusion. However, despite these challenges, the results are auspicious and suggest that seasonal Sentinel-2 imagery has the potential to predict compositional vegetation classes with high accuracy. Further work is needed to determine whether these results are replicable in other vegetation types and regions.

1. Introduction

Accurate maps of the distribution and extent of vegetation across a landscape are invaluable for conservation planning and for assessing landscape changes (Falcucci et al., 2007; Ferrier, 2002). Traditionally, such maps were produced using a combination of field work and aerial image interpretation. The time and cost associated with such approaches effectively constrain their application to relatively small areas (Lee and Lunetta, 1996). In contrast, maps generated from existing geospatial data, a small set of reference data (training sites), and machine learning (ML) are becoming increasingly popular. Two mapping approaches stand out, namely (1) predicting the potential natural vegetation (PNV; e.g. Mucina, 2010) patterns using abiotic variables (e.g. climate and soils) as input, or (2) classifying remotely sensed imagery into discrete thematic classes representing natural vegetation. The PNV maps have additional value as they can predict vegetation patterns in disturbed and previously vegetated areas. However, their implementation is often impractical due to the time and cost involved in collecting the necessary field information (e.g. soil texture or chemistry) to inform the relationship between the vegetation and the environment (Xie et al., 2008), particularly at fine spatial scales. The complexity of a region’s ecology and past dynamics may also reduce the

\textsuperscript{*} Corresponding author.
\textit{E-mail address:} paul.macintyre@research.uwa.edu.au (P. Macintyre).

https://doi.org/10.1016/j.jag.2019.101980

Received 31 January 2019; Received in revised form 25 September 2019; Accepted 27 September 2019
Available online 28 October 2019

0303-2434/ © 2019 Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/BY-NC-ND/4.0/).
accuracy of these predictions (e.g. Macintyre et al., 2018).

Supervised classification is commonly applied to remotely sensed imagery and involves the use of a reference set of labelled training sites (of known classification) to train ML algorithms to label unknown samples (Muñoz-Mari et al., 2007). With the increasing availability of remotely sensed data, the creation of such maps is often less costly than using extensive ground-based surveys (Foody, 2009). Supervised classifications have been carried out for many regions and vegetation types around the world (e.g. Carreiras et al., 2017; Fritz et al., 2017; Li et al., 2017), with the success of such approaches dependent on having representative training sites (Kavzoglu, 2009) and target classes separable in spectral space (Gamaroba et al., 2010). Target classes are typically defined by vegetation structure or terrain (structural classification) (Davis et al., 1994), but statistical (compositional; species-based) classifications have also been used (Cingolani et al., 2004; Zak and Cabido, 2002). However, compositional classifications may be challenging to perform with remotely sensed multispectral imagery as its spectral resolution is limited and may be inadequate to separate the target classes (McIver and Friedl, 2002) or where the imagery is not at a fine-enough spatial resolution to examine local patterns (Woodcock and Strahler, 1987). Class separability can be increased by generalizing compositional classes into coarser units (merging classes together) or alternatively, or by using imagery with wider spectral ranges and narrower bandwidths. Such hyperspectral data have been employed for various fine-scale applications, including forestry resource inventories (Goverdner et al., 2008), invasive species detection (Pengra et al., 2007) and mapping vegetation distribution patterns (Statoulias et al., 2018). Hyperspectral data are, however, often unavailable or prohibitively expensive, especially across large areas. Until such data become mainstream, efforts should focus on improving vegetation classification accuracies using available multispectral imagery.

The performance of multispectral image classifications can be improved by including additional input variables (Khatani et al., 2016), with the use of image transformations (e.g. Lu & Wong, 2007; Otukei and Blaschke, 2010) and vegetation indices (Peña-Barragán et al., 2011; Sahebjalal and Dashtekian, 2013) being popular. Creating multi-temporal datasets that highlight phenomenological differences have also shown promise (Verhulp and Van Niekirk, 2016). Logically, capturing any phenomenological difference between target classes with multi-temporal data should improve the classification of compositional vegetation and has been shown to improve classification accuracy in vegetation with distinct phenological stages (i.e. deciduous woodlands; Hill et al., 2010). However, although the value of multi-temporal imagery has been examined for agriculture (Gilbertson and Van Niekerk, 2017) and land-use mapping (Kübert et al., 2016), its application for vegetation mapping has been mainly limited to structural classifications targeting units such as cropland, mixed forest or woodland (e.g. Boles et al., 2004– in Temperate East Asia; Silva et al., 2011– in Brazilian Cerrado) and species-based compositional classifications within small geographic areas (Dechka et al., 2002– in Canadian Prairie; Förster et al., 2012– German Grasslands). While studies demonstrate the potential of multi-temporal imagery, it is not known whether they can be scaled to regional level or whether the results can be replicated in the Australian vegetation.

The recent availability of Sentinel-2 imagery (Berger et al., 2012) – offering 10–60 m spatial, 13-band spectral and 5-day temporal resolutions – opens new opportunities for investigating the potential of remote sensing for fine-scale compositional classifications, particularly in landscapes supporting complex vegetation patterns. Pre-operational studies using simulated Sentinel-2 imagery examining fine-scale classifications in complex vegetation types, including savanna (Hill, 2013), wetlands (Statoulias et al., 2015) and forests (Laurin et al., 2016), concluded that the spectral and spatial resolution of the Sentinel-2 imagery could potentially produce accurate classifications. More recently, Erinjeri et al. (2018) demonstrated how Sentinel-2 imagery could be used to produce highly accurate classifications of vegetation structure and land use within a rainforest landscape. These results are promising, but the efficacy of Sentinel-2 imagery for producing maps of compositional vegetation in landscapes where the vegetation structure and land use is predominately uniform, remains untested.

The Geraldton Sandplains of Western Australia is a landscape that supports multifaceted vegetation, known as kwongan (Beard, 1984; Mucina et al., 2014). Macintyre et al. (2018) attempted to model the PNV of the kwongan by using ML and environmental variables, but the complexity of the vegetation–environment relationship in the region limited predictive accuracies. The study concluded that spectral data, recorded by satellite sensors, may improve modelling accuracies. Although the kwongan vegetation is visually homogenous (Hopkins and Griffin, 1984), the authors postulated that spectral variations outside of the visible spectrum may help differentiate some vegetation communities.

This study investigates the value of multi-temporal (seasonal) imagery for vegetation mapping, specifically when applied to compositional classifications. To achieve this, we assessed the classification results of four ML classifiers (classification trees, random forests, nearest-neighbour, and support vector machines) – applied to several single- and multi-temporal Sentinel-2 datasets (scenarios) – against a limited number of field-collected reference sites representing the compositional vegetation of the kwongan. These results were statistically compared to identify the combinations of season and data type that produced the highest level of classification accuracy. From this, we aim to determine whether multi-seasonal approaches using Sentinel-2 data are usable to create real-world maps of compositional vegetation, and what limitations exist to their implementation.

2. Materials and methods

2.1. Study region

The study area is located within the Geraldton Sandplains, about 270 km north of Perth in Western Australia (Fig. 1). The area covers 1210 km², extending from 29°32′59″ to 29°45′37″ S and 115°04′44″ to 115°16′57″ E. The region is characterised by Mediterranean-like climate (cool and wet winters, and hot and dry summers). The winter to spring wet period (June to October) on average provides 300–500 mm of rainfall per year (Beard, 1990). The annual rainfall during the year of this study (2017) was 90 mm below the long-term average, with higher than average rainfall recorded during spring (Fig. 2). The kwongan contains higher species diversity than any other sclerophyllous vegetation in Western Australia (Lamont et al., 1982) and has high levels of endemism (Mucina et al., 2014). Predominant land use (69%) is agriculture (Desmond and Chant, 2002), with remnant vegetation significantly fragmented.

2.2. Satellite data

Sentinel-2 images representing the main seasons of the year (summer, winter, autumn and spring) were acquired from the European Space Agency (ESA) Copernicus Open Access Hub. The acquired image dates were 7 January (summer), 17 May (autumn), 26 June (winter) and 4 October (spring) of 2017. The images of the spring and autumn seasons align with the known phenology of peak and lowest flowering periods respectively (Bell and Stephens, 1984); while the summer and winter seasons lack a distinct phenotypic stage they represent the major climate differences (peak drought vs peak wet). All images were corrected to base of atmosphere reflectance using the Sen2Cor v2.3.1 Python module (Main-Knorn et al., 2017), then divided by 10 000 to account for the scaling factor. Bands uninformative for vegetation analysis (coastal aerosol, water vapour, SWIR-cirrus) were excluded.
assessed, and sites disturbed by fire or land clearing since establishment were removed. This reduced the total number of plots to 301. The previous maps created by field-survey (Woodman Environmental Consulting, 2009) were used in conjunction with the Sentinel-2 imagery to identify areas of homogenous vegetation around each of the remaining vegetation plots. These areas ranged from one (10 m × 10 m) to a maximum of four (20 m × 20 m) pixels and used as samples for classifier training and accuracy assessment. Due to the nature of the species-based classification and the elimination of disturbed plots, the training data were unbalanced with some classes having small numbers of representatives (Table 1).

2.4. Feature-set generation

The spectral bands were supplemented with four spectral indices, the enhanced vegetation index (EVI), green normalised difference vegetation index (GNDVI), modified soil adjusted vegetation index 2 (MSAVI2) and normalised difference vegetation index (NDVI), all of which are commonly used in vegetation analyses. The normalised difference red-edge index (NDRE) was also added. Two image transformations, namely principal component analysis (PCA) and tasselled cap transformation (TCT) (Kauth & Thomas, 1976), were implemented. Coefficients by Nedkov (2017) were used to create the TCT brightness, greenness and wetness transformations within ArcMap v 10.3. PCA was implemented using the PCA tool within ArcMap v10.3 to extract the first five principal components (PCs). These PCs explained > 95% of the data variation and were generated using the spectral information for each image individually. An additional PCA (five axes) was performed on all bands of all images combined, effectively condensing the variation of all spectral data to five PCs. The entire feature-set thus consisted of 97 variables, including ten spectral bands, five indices, eight image transformations per image date, and the five PCs of all images combined.

2.5. Classification scenarios

Table 2 details all combinations of input variables examined to investigate the effects of seasonality and feature type on classification accuracy. Variable importance was generated using the random forest (RF) classifier within Salford Systems Predictive Modeller V 8.0 (Salford Systems, 2017). The single most and the 10, 25 and 50 most important variables, as inherently determined by the classifier, were used for Scenarios 23, 22, 21 and 20 respectively. Although such feature filter approaches are not always optimal (because they do not take the learning algorithm’s performance into consideration) (Saeyes et al., 2007), their simplicity and ease of implementation outweighed the application of a wrapper approach. RF-based filter feature selection has also been successfully applied in a wide range of studies (Chan and Paolini, 2008; Chehata et al., 2009; Duro et al., 2012a; Eisavi et al., 2015) and thus allows direct comparison.

The supervised learning image classification environment (SLICE) software, developed by Myburgh and Van Niekerk (2013), was used to examine the classification scenarios. Based on the GDAL (GDAL Development Team, 2010), OpenCV (Bradski and Kaehler, 2000) and LibSVM (Chang and Lin, 2011) libraries, SLICE includes four classifiers, namely support vector machines (SVM; Mountrakis et al., 2011), nearest-neighbour (NN; Myint et al., 2011), classification trees (CT; Breiman et al., 1984), and random forests (RF; Breiman, 2001). A 60/40 random split is used to create separate training and validation datasets. The latter are used to create confusion matrices and to calculate the mean overall accuracy (MOA) and mean kappa coefficients for each scenario. Each scenario was iterated 30 times to reduce the stochastic effect of training sample selection on classification accuracy.

Significant differences in accuracy were assessed using the aligned rank transform (Wobbrock et al., 2011), which was implemented in the R package ARTool (Kay and Wobbrock, 2016). Multiple comparisons for...
significance were examined using the function testInteractions in the R package PhDIA (Rosario-Martinez, 2015). The Holm method was applied for p-value adjustment. Unless specified all significant differences are considered at the level of p < 0.001.

3. Results

3.1. Overall classification accuracy

The results of all scenarios (Table 3) show that the classification performance of single season datasets (Scenarios 1–4) was mixed. The classifications that made use of the summer and winter image sets returned unusable (< ??%) accuracies. The spring feature set returned accuracies higher than all other scenarios, with the exception of a single multi-temporal feature set comprised of autumn and spring images (Scenario 9), which returned the highest accuracy across all scenarios. The addition of the autumn images to the spring dataset set led to accuracy improvements ranging from 4% (CT) to 15% (SVM and NN). The performance of individual classifiers in Scenario 9 showed that SVM returned the highest classification accuracy (68%), with RF and NN returning comparable accuracies (63% and 62% respectively). The

Table 2

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Description</th>
<th>Scenario</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1–4</td>
<td>Each image individually</td>
<td>14</td>
<td>Spring, autumn &amp; winter</td>
</tr>
<tr>
<td>5</td>
<td>Summer &amp; winter</td>
<td>15</td>
<td>All images</td>
</tr>
<tr>
<td>6</td>
<td>Summer &amp; autumn</td>
<td>16</td>
<td>All images: spectral bands only</td>
</tr>
<tr>
<td>7</td>
<td>Summer &amp; spring</td>
<td>17</td>
<td>PCA of all images</td>
</tr>
<tr>
<td>8</td>
<td>Autumn &amp; winter</td>
<td>18</td>
<td>PCA of each image &amp; PCA of all</td>
</tr>
<tr>
<td>9</td>
<td>Autumn &amp; spring</td>
<td>19</td>
<td>All images: TCT only</td>
</tr>
<tr>
<td>10</td>
<td>Spring &amp; winter</td>
<td>20</td>
<td>Feature selection (50)</td>
</tr>
<tr>
<td>11</td>
<td>Summer, winter &amp; spring</td>
<td>21</td>
<td>Feature selection (25)</td>
</tr>
<tr>
<td>12</td>
<td>Summer, winter &amp; autumn</td>
<td>22</td>
<td>Feature selection (10)</td>
</tr>
<tr>
<td>13</td>
<td>Summer, spring &amp; autumn</td>
<td>23</td>
<td>Feature selection (1)</td>
</tr>
</tbody>
</table>

Table 1

<table>
<thead>
<tr>
<th>Target class</th>
<th>Broad vegetation description</th>
<th>Major environment type</th>
<th># Training sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Mixed Banksia shrubland over sedge-like herb</td>
<td>Sand</td>
<td>88</td>
</tr>
<tr>
<td>2</td>
<td>Banksia shrubland</td>
<td>Sand</td>
<td>169</td>
</tr>
<tr>
<td>3</td>
<td>Semi-open shrubland</td>
<td>Sand</td>
<td>20</td>
</tr>
<tr>
<td>4</td>
<td>Mixed low shrubland over herbs</td>
<td>Sand</td>
<td>23</td>
</tr>
<tr>
<td>5</td>
<td>Mixed low shrubland Hakea dominant</td>
<td>Sand</td>
<td>25</td>
</tr>
<tr>
<td>6</td>
<td>Banksia over mixed shrubs</td>
<td>Laterite</td>
<td>31</td>
</tr>
<tr>
<td>7</td>
<td>Open shrubland over sedge-like herbs</td>
<td>Sand</td>
<td>42</td>
</tr>
<tr>
<td>8</td>
<td>Banksia shrubland over sedge-like herbs</td>
<td>Sand</td>
<td>88</td>
</tr>
<tr>
<td>9</td>
<td>Shrubland of Acacia and Banksia</td>
<td>Laterite</td>
<td>14</td>
</tr>
<tr>
<td>10</td>
<td>Shrubland of Banksia and Melaleuca</td>
<td>Laterite</td>
<td>51</td>
</tr>
<tr>
<td>11</td>
<td>Low closed shrubland</td>
<td>Laterite</td>
<td>18</td>
</tr>
<tr>
<td>12</td>
<td>Shrubland of Acacia and Allocasuarina</td>
<td>Laterite</td>
<td>24</td>
</tr>
<tr>
<td>13</td>
<td>Woodland over shrubs</td>
<td>Laterite</td>
<td>26</td>
</tr>
<tr>
<td>14</td>
<td>Mixed shrubland</td>
<td>Laterite</td>
<td>14</td>
</tr>
<tr>
<td>15</td>
<td>Shrubland of Banksia and Callitris</td>
<td>Laterite</td>
<td>52</td>
</tr>
<tr>
<td>16</td>
<td>Mixed Shrubland over sedge-like herbs</td>
<td>Laterite</td>
<td>47</td>
</tr>
<tr>
<td>17</td>
<td>Trees over mixed shrubland</td>
<td>Laterite</td>
<td>51</td>
</tr>
<tr>
<td>18</td>
<td>Mixed shrubland over mixed herbs</td>
<td>Laterite</td>
<td>10</td>
</tr>
<tr>
<td>19</td>
<td>Woodland over shrubs</td>
<td>Wet</td>
<td>20</td>
</tr>
<tr>
<td>20</td>
<td>Woodland swamp</td>
<td>Wet</td>
<td>12</td>
</tr>
<tr>
<td>21</td>
<td>Open shrubland over mixed sedges</td>
<td>Wet</td>
<td>15</td>
</tr>
<tr>
<td>22</td>
<td>Wet saline woodland</td>
<td>Wet</td>
<td>15</td>
</tr>
<tr>
<td>23</td>
<td>Mixed shrubland, emergent trees</td>
<td>Wet</td>
<td>49</td>
</tr>
<tr>
<td>24</td>
<td>Melaleuca shrubland</td>
<td>Wet</td>
<td>28</td>
</tr>
</tbody>
</table>
Reducing the dimensionality, by using only image transformation features as input (Scenarios 17–19), had a detrimental effect on the accuracies of all classifiers. Of these, the TCT of all images (Scenario 19) returned the best accuracies. However, the accuracies in Scenario 19 was considerably lower (43% accuracy reduction for SVM) than the single spring season feature set (Scenario?). Dimensionality reduction through feature selection (Scenarios 20–23) typically returned lower accuracies for all classifiers, with RF being the only classifier to maintain accuracies using the 50 most (32%) and 10 most (30%) important variables. Using the single most important variable (EVI derived from the spring season imagery; Supplementary Material 3) returned identical performance with all classifiers (16%).

### 3.2. Refinement of the autumn-spring combined feature set

An additional set of experiments (not included in the experimental design) was carried out on five feature subsets to understand why Scenario 9 outperformed the other scenarios. These subsets were formed by grouping the features used in Scenario 9 into themes: a) spectral bands; b) PCs of individual images; c) PCs of both images combined; d) TCT on individual images; and e) all image transforms combined. The results of these experiments are summarised in Table 4 and graphically compared in Fig. 3. The accuracy of the subsets featuring only the spectral bands (65%; Scenario 9a) and the TCT components (65%; Scenario 9d) were significantly higher than those of the other subsets, but not statistically different from each other. In these scenarios, the difference in performance between SVM and NN (Scenario 9a SVM = 72% and NN = 71%, Scenario 9d SVM = 74% and NN = 71%) were statistically insignificant (Fig. 4). These two classifiers provided significantly higher accuracies than RF and CT in both scenarios (See Supplementary Material 4 for example classification maps). RF variable importance for these scenarios (Supplementary material 5) identified the SWIR bands (Bands 11 and 12) and the first red-edge band (Band 5) of the spring image from Scenario 9a to be the most informative. The TCT greenness index from both seasons, followed by the wetness index for spring, was the most informative variables in Scenario 9d.

### 3.3. Inter-class confusion

The mean spectral response of each training class across the combined autumn and spring feature sets is shown in Fig. 5. While the accuracies are statistically comparable, the confusion matrices of combined features are presented in Table 4 and graphically compared in Fig. 3. The accuracy of the subsets featuring only the spectral bands (65%; Scenario 9a) and the TCT components (65%; Scenario 9d) were significantly higher than those of the other subsets, but not statistically different from each other. In these scenarios, the difference in performance between SVM and NN (Scenario 9a SVM = 72% and NN = 71%, Scenario 9d SVM = 74% and NN = 71%) were statistically insignificant (Fig. 4). These two classifiers provided significantly higher accuracies than RF and CT in both scenarios (See Supplementary Material 4 for example classification maps). RF variable importance for these scenarios (Supplementary material 5) identified the SWIR bands (Bands 11 and 12) and the first red-edge band (Band 5) of the spring image from Scenario 9a to be the most informative. The TCT greenness index from both seasons, followed by the wetness index for spring, was the most informative variables in Scenario 9d.

### Table 3

Mean overall accuracy (MOA) and mean kappa coefficients (MKC) values for all scenarios. SU = summer season, AU = autumn season, WI = winter season, SP = spring season. PCA_IND = PCA of individual seasons, PCA_All = single PCA of all spectral features, TCT = tasselled cap transformation. Colours range from green (high) to red (low). This is scaled across all MOA and K columns and as such it visualises accuracies per scenario and classifier.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>SVM</th>
<th>NN</th>
<th>CT</th>
<th>RF</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 SU</td>
<td>5.00 0.00 0.01 0.04 0.05</td>
<td>4.00 0.00 0.01 0.04 0.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 AU</td>
<td>5.00 0.00 0.01 0.04 0.05</td>
<td>4.00 0.00 0.01 0.04 0.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 WI</td>
<td>5.00 0.00 0.01 0.04 0.05</td>
<td>4.00 0.00 0.01 0.04 0.05</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 4

MOA and MKC values for additional classification experiments of scenario 9. AU = autumn season, SP = spring season. PCA_IND = PCA of individual seasons, PCA_All = single PCA of all spectral features, TCT = tasselled cap transformation. Colours range from green (high) to red (low). This is scaled across all OA and K columns and as such it visualises accuracies per scenario and classifier.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>SVM</th>
<th>NN</th>
<th>CT</th>
<th>RF</th>
</tr>
</thead>
<tbody>
<tr>
<td>9 AU &amp; SP</td>
<td>5.00 0.00 0.01 0.04 0.05</td>
<td>4.00 0.00 0.01 0.04 0.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9a AU, SP - Bands only</td>
<td>5.00 0.00 0.01 0.04 0.05</td>
<td>4.00 0.00 0.01 0.04 0.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9b AU, SP - PCA_IND</td>
<td>5.00 0.00 0.01 0.04 0.05</td>
<td>4.00 0.00 0.01 0.04 0.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9c AU, SP - PCA_All</td>
<td>5.00 0.00 0.01 0.04 0.05</td>
<td>4.00 0.00 0.01 0.04 0.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9d AU, SP - TCT</td>
<td>5.00 0.00 0.01 0.04 0.05</td>
<td>4.00 0.00 0.01 0.04 0.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9e AU, SP - Transforms</td>
<td>5.00 0.00 0.01 0.04 0.05</td>
<td>4.00 0.00 0.01 0.04 0.05</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Scenarios 9a and 9d (aggregated across all iterations; Tables 5 and 6) highlight that the major difference is in the producer’s and consumer’s accuracies; with Scenario 9a (Table 5) returning higher producer’s accuracies compared to Scenario 9d (Table 6), which has a higher consumer’s accuracy. Despite this difference, both agree in two ways: the poor accuracy and reliability of Class 14, and low confusion between classes 19–24. Class 14 broadly represents mixed species shrubland over laterite and has the highest levels of omission and commission errors. It was frequently misclassified as Classes 8, 10, 12, 13 and 17 — all shrubland vegetation types dominated by Banksia and Acacia species (Table 1). All these classes, except for Class 8, are representatives of the laterite group. Classes 19–24 are wet vegetation types and were rarely misclassified (low errors of omission). They were mostly only confused with other wet vegetation types (e.g. Class 23 with Class 20). All the wet vegetation types showed very low errors of commission, suggesting good separation from other classes.

4. Discussion

4.1. Accuracy and vegetation phenology

Accurate and reliable maps that depict the patterning of vegetation at fine scales are of increasing importance for effective conservation and land-use planning. Although maps created using traditional methods are preferable, time and costs associated with producing such maps reduce their viability, especially across large areas. This study found that classifications made using multi-temporal Sentinel-2 imagery of two contrasting seasons (spring and autumn) were the most accurate, while classifications based on the single-season spring imagery were the second most accurate. Classifications made using other combinations of seasons and feature types varied, but returned significantly lower accuracies. In the kwongan, spring and autumn represent the periods of peak (spring) and lowest (autumn) flowering (Bell and Stephens, 1984). Although the good performance of the spring season Sentinel-2 imagery – by itself – suggests that the differences in flowering phenology between the target classes (floristic communities) are detectable, the addition of the contrasting autumn season is needed to improve classifications.

Several studies have shown that the accuracy of vegetation-focused classifications can be increased using TCT components, both in natural systems such as forests (Dymond et al., 2002) and human-made systems such as agriculture (Oetter et al., 2001). The greenness component is particularly useful for vegetation as it depicts variations in photosynthetically active vegetation (Lobser and Cohen, 2007), potentially emphasising the contrast between the spring and autumn seasons. Similarly, the red-edge band has a strong relationship with the...
chlorophyll content of leaves (Delegido et al., 2011) and likely contributed to the separation of the vegetation classes. The prominence of the SWIR bands in the variable importance list is attributed to differences in moisture content of the kwongan vegetation and soil (Ceccato et al., 2001). This is supported by the importance of the spring wetness index (see also Wang et al., 2008). The natural checked-board patchiness in kwongan vegetation (allowing for the occurrence of bare earth reflectance) and the above-average rainfall in spring may have improved vegetation class separability by emphasising differences in soil conditions. These findings suggest that climate and the physical environment, particularly soil characteristics, could have a substantial impact on the classification accuracies.

### 4.2. Sources of the inter-class confusion

The compositional classification scheme used for training is based on the premise that each class is the result of a distinct combination of diagnostic species (see Chytrý et al., 2002 for a definition). These species are the ecological and compositional factors used to differentiate the vegetation classes. In reality, a species may be present as a diagnostic within multiple classes. Further increasing the complexity is that each plot may contain multiple ancillary species considered non-
diagnostic, i.e. they are not used to differentiate the target classes. Given that the spectral signatures of each sampled plot, represented by an individual or group of pixels, are a combination of all underlying species, those sites with more species in common are likely to have more similar signatures. However, the spectral signatures are also affected by the percentage cover of each species per plot, which naturally varies even within representatives of a single target class.

The results suggest that soil variations may also have affected classification accuracies. In the ecological context (as manifested at sub-pixel level) the patch dynamics of the kwongan vegetation is likely a driver of this confusion. The kwongan vegetation is characterised by incomplete vegetation cover with areas of bare soil frequently visible between plants ranging from a few square centimetres to several square meters (Enright et al., 2007). The resolution of the Sentinel-2 images is too coarse to determine whether the size or occurrence of patches had any relationship to the target classes, but the natural (stochastic) occurrence of the patches may have had a significant effect on intra-class spectral variation as the bare earth reflectance would have influenced the (mixed) reflectance measured by the sensor.

The variable importance list for Scenario 9a highlighted the value of spectral bands associated with moisture (e.g. B11 and B12). This likely reflects the differences in water stress and water use by the species within each class, but the topographic positions of the training sites are equally important. For the most part, the study region is relatively homogenous (topographically), but changes in topography such as drainage channels and dune features are present at a local scale. Except for a small number of classes specifically describing wetland vegetation, the rest are not associated with specific topographic positions. At close scrutiny of individual sites, it seems that presence (or absence) of drainage channels may have affected water availability at specific sites hosting the same target class. These differences in water availability could have increased spectral variation among training sites and contributed to misclassifications.

It is well known that the performance of classifiers is affected by the classification problem, training sites and feature sets (Fernández-Delgado et al., 2014; Kattan and Cooper, 2000; Wolpert and Macready, 1997), however, the variation in performance observed in this study was unexpected. Given the findings of Duro et al. (2012b) and Shao and Lunetta (2012) it was expected that both the SVM and RF classifiers would return the best classification accuracies. However, this was not the case for most of the scenarios in our study. There was a consistent level of accuracy by the RF classifier but overall this was poor. In contrast, the SVM and NN classifiers failed to return useable accuracy with most feature sets except those that made use of the spring and combined autumn/spring features. Given the black-box nature of these classifiers it is difficult to explain this finding with a high level of confidence. However, some suppositions can be made by taking the design of the algorithms into consideration. For example, the tree-based classifiers (RF, CT) rely on recursive sub-setting, which organizes the training data into groups of increasing homogeneity (Breiman et al., 1984; Breiman, 2001). In contrast, the SVM and NN classifiers rely on splitting the training data mathematically, using either hyperplanes (SVM; Mountrakis et al., 2011) or by assigning due to position in multivariate space (NN; Miynt et al., 2011). The relatively lower performance of the tree-based classifiers suggests that there is a level of intra-class variability that makes recursive sub-setting inefficient.

The classification accuracies obtained in this study (72–74% for the best scenarios) show that the Sentinel-2 data were able to produce reasonably accurate classifications of fine-scale compositional vegetation. Whether this accuracy is high enough for it to be recommended for use in practical applications is, however, dependent on the intended use of the final product. For example, when exact delineations at large mapping scales are required, a higher accuracy may be more sensible, but for regional mapping exercises of vegetation types, the accuracies obtained in this study are likely sufficient. Based on qualitative interpretations, it seems that maps generated from the Sentinel-2 imagery are useful for the regional-scale identification of compositional patterns and planning. However, the map likely has limited value for delineating units of conservation significance, which agrees with Feilhauer et al. (2013) who, based on pre-operational examinations, concluded that the spatial resolution of Sentinel-2 imagery may be too low for conservation applications. Of importance, however, is that our results highlight the importance of accounting for how the underlying ecology of the vegetation may affect class confusion and separability.

Direct comparison of this study to others is difficult, mainly because of the limited studies examining compositional vegetation and the absence of studies in which Sentinel-2 imagery was utilised. The study by Erinjey et al. (2018) is the most comparable. They recorded overall accuracies of 82–86%, substantially higher than the 72–74% (on average) and 78% (best) recorded in our study. However, their classification scheme consisted of a mixture of land covers/uses that are spectrally and structurally distinct. In contrast, we targeted 24 vegetation classes (groupings of species) in a spectrally and structurally homogenous land-cover (shrubland) within a conservation area (i.e. one land cover/use). Within this context, our results compare favourably.

5. Conclusion

The increasing availability of earth observation satellite missions offering free imagery with relatively high spatial and spectral resolutions provide opportunities for the efficient assessment of vegetation across large areas. While the utility of remotely sensed data has been proven for studies focussing on structural (homogenous land cover) target classes (e.g. paddock, woodland, forest) there is a gap in the knowledge as to whether this new source of satellite imagery is suitable for accurately and reliably mapping target classes representing the compositional (species-based) patterns. This level of mapping is of interest as the high costs associated with field-based methods impedes their production at regional scales. This study examined whether the spatial and spectral resolution of a multi-temporal dataset derived from Sentinel-2 imagery were sufficient to accurately classify compositional vegetation patterns across a homogeneous shrubland at fine scales. The results show that feature sets comprised of a single season were not sufficient to accurately classify the target classes, but the inclusion of additional seasons was able to significantly increase the classification accuracies. Of all combinations, a multi-temporal feature set consisting of imagery from the Autumn and Spring seasons – representing two contrasting periods of phenology (peak vs lowest flowering) – returned the highest classification accuracies, with the SVM and NN classifiers being the most accurate. Examination of the target class confusions highlighted that ecological features occurring at scales below that of the pixel size, such as bare earth presence, shared species and fine-scale topographic variation, likely reduced classification accuracies. These results show that the Sentinel-2 platform can be used for classifications of target units at much finer levels than currently conducted, however there are a number of aspects that need further investigation. More field verification (in addition to the plots used for accuracy assessment) is needed. Specifically, an assessment of the effects of topographic variability on class separability is recommended. These field verifications should preferably be compared to land surface parameters derived from a high (e.g. 5 m) resolution DEM. More research is also needed to examine whether the methodology employed in this study can be replicated in different regions and vegetation types, particularly where vegetation structure exhibits a large amount of diversity.

Authors’ contribution statement

PM conceived the original study with AV providing technical guidance for the design of the methods. PM led the writing of the manuscript with AV and LM providing conceptual and editing guidance. All authors contributed critically to the drafts and gave final approval for
the publication.

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 study was funded by benefiting from the ARC Linkage grant LP150100339. The authors would also like to acknowledge Timothy Hammer for his insight into ecological interpretations and James Tsakalos for providing the vegetation classification scheme. The Authors would like to thank two anonymous reviewers for their valuable suggestions.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.jag.2019.101980.

References


