Using single- and multi-target regression trees and ensembles to model a compound index of vegetation condition

Dragi Kocev [a,*], Sašo Džeroski [a], Matt D. White [b], Graeme R. Newell [b], Peter Griffioen [c]

[a] Dept. of Knowledge Technologies, Jožef Stefan Institute, Jamova 39, 1000 Ljubljana, Slovenia
[b] Dept. of Sustainability and Environment, Arthur Rylah Institute for Environmental Research, 123 Brown Street, Heidelberg, Victoria 3084, Australia
[c] Acromap, Pty. Ltd., 37 Gloucester Drive, Heidelberg, Victoria 3084, Australia

ABSTRACT

An important consideration in conservation and biodiversity planning is an appreciation of the condition or integrity of ecosystems. In this study, we have applied various machine learning methods to the problem of predicting the condition or quality of the remnant indigenous vegetation across an extensive area of south-eastern Australia—the state of Victoria. The field data were obtained using the 'habitat hectares' approach. This rapid assessment technique produces multiple scores that describe the condition of various attributes of the vegetation at a given site. Multiple sites were assessed and subsequently circumscribed with GIS and remote-sensed data.

We explore and compare two approaches for modelling this type of data: to learn a model for each score separately (single-target approach, a regression tree), or to learn one model for all scores simultaneously (multi-target approach, a multi-target regression tree). In order to lift the predictive performance, we also employ ensembles (bagging and random forests) of regression trees and multi-target regression trees. Our results demonstrate the advantages of a multi-target over a single-target modelling approach. While there is no statistically significant difference between the multi-target and single-target models in terms of model performance, the multi-target models are smaller and faster to learn than the single-target ones. Ensembles of multi-target models, also, improve the spatial prediction of condition.

The usefulness of models of vegetation condition is twofold. First, they provide an enhanced knowledge and understanding of the condition of different indigenous vegetation types, and identify possible bio-physical and landscape attributes that may contribute to vegetation decline. Second, these models may be used to map the condition of indigenous vegetation, in support of biodiversity planning, management and investment decisions.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

Governments and other agencies worldwide are increasingly required to demonstrate their compliance with the policies and legislation relevant to the protection and management of remnant indigenous vegetation (Parkes and Lyon, 2006). To this end, government agencies are seeking to extend the requisite knowledge base and representation of vegetation beyond just 'extent' and 'type', to incorporate the notion of 'condition' or 'quality'. The concept of vegetation condition is typically idiosyncratic and/or context-specific. For example, the performance or quality of native vegetation could be evaluated in terms of its capacity to deliver services such as energy storage (including carbon sequestration), nutrient cycling,

landscape stability, fodder production for domestic stock or habitat for species. A key challenge has been to develop metrics that facilitate comparisons of condition both within and between disparate ecosystem types. Recent attempts have been made to clarify these concepts (Andreasen et al., 2001; Gibbons et al., 2006), and develop general and widely applicable metrics and indices for assessing vegetation or ecosystem condition from a biodiversity perspective (Parkes et al., 2003; Scholes and Biggs, 2005; Oliver, 2004; Eyre et al., 2006; Gibbons et al., 2009).

With an increasing emphasis on landscape scale planning for biodiversity investment (Margules and Pressey, 2000; Rouget et al., 2006; Knight et al., 2006; Moilanen, 2007) and widespread access to Geographic Information Systems (GIS) and associated data and software, the production of maps or spatially explicit models of landscape indices, species distributions and other ecological phenomena has become commonplace (see Li and Wu, 2004; Guisan and Thuiller, 2005). The apparent utility of compound indices, such as vegetation condition or ecosystem integrity presents a generic problem for the land management agencies which employ them:

* Corresponding author.
E-mail addresses: Dragi.Kocev@ijs.si (D. Kocev), Saso.Dzeroski@ijs.si (S. Džeroski), Matt.White@dse.vic.gov.au (M.D. White), Graeme.Newell@dse.vic.gov.au (G.R. Newell), pgriffioen@acromap.com (P. Griffioen).

can we usefully predict such attributes from site data across extensive geographic regions, from a vector of covariate remote sensed and ancillary environmental data?

The focus of this study is to take data from site assessments employing a multi-component index of vegetation condition and attempt to fit a generalized view of this index over an extensive area—in this case the State of Victoria, Australia, an area of some 227,000 km^2. So, the problem that we are addressing is how to predict multiple target variables (responses) from a vector of ecological/remote-sensed data. We employed two modelling scenarios: (1) learn a model for each component of the overall index separately and (2) learn a model for all component scores simultaneously. For the first scenario, we applied regression trees (RTs) (Breiman et al., 1984) and ensembles of RTs (Breiman, 1996, 2001) to the problem, while for the second, we applied multi-target regression trees (MTRTs) (Struyf and Džeroski, 2006) and ensembles of MTRTs (Kocev et al., 2007).

Regression trees are decision trees that predict the value of a single numeric target variable. The multi-target regression trees are a generalization of RTs. They are able to predict the value of multiple numeric target variables. Their main advantages (over building a separate model for each target attribute) are: (1) a multi-target model is smaller than the total size of the individual models for all target attributes and (2) a multi-target model explains dependencies between different target attributes (Blockeel et al., 1998; Struyf and Džeroski, 2006). We selected regression trees and multi-target regression trees because they are easy to understand and interpret and yet offer satisfactory predictive power.

To obtain models that have improved predictive performance we used ensembles. Ensemble learning combines the predictions of multiple models and lifts the predictive performance of their base classifiers, both in the single-target (Breiman, 1996) and the multi-target setting (Kocev et al., 2007). We focus on the two most widely used ensemble learning methods that use tree models as base classifiers: bagging (Breiman, 1996) and random forests (Breiman, 2001).

We perform the analysis using two scenarios: (1) we learn pruned tree models (smaller tree models) to obtain some knowledge and understanding about the condition of the indigenous vegetation and (2) we learn ensembles of trees opting for better predictive performance that will yield more precise and reliable maps of the vegetation condition.

The development of predictive models of condition for remnant indigenous vegetation may assist in identifying the relative importance of associated biophysical and landscape attributes in explaining observed condition states, across vegetation types and landscape scales. In addition, spatially explicit models of condition, could, when used in conjunction with other data, inform natural resource investment decisions, statutory protection and reserve design, while providing a basis for new forms of environmental accounting and potentially monitoring landscape change.

The remainder of this paper is organized as follows: In Section 2, we describe our modelling methodology, and in Section 3 the data. The experimental setup for data analysis is presented in Section 4. In Section 5, we present, discuss and compare the models that we obtained. Finally, we outline our conclusions in Section 6.

2. Machine learning methodology

2.1. Regression trees

Regression trees are decision trees that predict the value of a numeric target variable (Breiman et al., 1984). Regression trees are hierarchical structures, where the internal nodes contain tests on the input attributes. Each branch of an internal test corresponds to an outcome of the test, and the prediction for the value of the target attribute is stored in a leaf. Each leaf of a regression tree contains a constant value as a prediction for the target variable (regression trees represent piece-wise constant functions).

To obtain the prediction for a new data record, the record is sorted down the tree, starting from the root (the top-most node of the tree). For each internal node that is encountered on the path, the test that is stored in the node is applied. Depending on the outcome of the test, the path continues along the corresponding branch (to the corresponding subtree). The resulting prediction of the tree is taken from the leaf at the end of the path. The tests in the internal nodes can have more than two outcomes (this is usually the case when the test is on discrete-valued attributes where a separate branch/subtree is created for each value). Typically each test has two outcomes: the test has succeeded or the test has failed. The trees in this case are also called binary trees.

2.2. Multi-target regression trees

Multi-target regression trees (Blockeel et al., 1998; Struyf and Džeroski, 2006) generalize regression trees to the prediction of several numeric target attributes simultaneously. The leaves of a multi-target regression tree store a vector, instead of storing a single numeric value. Each component of this vector is a prediction for one of the target attributes. An example of a multi-target regression tree is shown in Fig. 3.

A multi-target regression tree (of which a regression tree is the special case with a single response variable) is usually constructed with a recursive partitioning algorithm from a training set of records. The algorithm is known as Top-Down Induction of Decision Trees (TDIDT). The records include measured values of the descriptive and the target attributes. The tests in the internal nodes of the tree refer to the descriptive, while the predicted values in the leaves refer to the target attributes.

The TDIDT algorithm starts by selecting a test for the root node. Based on this test the training set is partitioned into subsets according to the test outcome. In the case of binary trees, the training set is split into two subsets: one containing the records for which the test succeeds (typically the left subtree) and the other contains the records for which the test fails (typically the right subtree). This procedure is recursively repeated to construct the subtrees.

The partitioning process stops when a stopping criterion is satisfied (e.g., the number of records in the induced subsets is smaller than some predefined value; the length of the path from the root to the current subset exceeds some predefined value, etc.). In that case, the prediction vector is calculated and stored in a leaf. The components of the prediction vector are the mean values of the target attributes calculated over the records that are sorted into the leaf.

One of the most important steps in the tree induction algorithm is the test selection procedure. For each node a test is selected by using a heuristic function computed on the training data. The goal of the heuristic is to guide the algorithm towards smaller trees with good predictive performance.

In this paper, we use the CLUS (Blockeel and Struyf, 2002) system for constructing (multi-target) regression trees (the system is available at http://www.cs.kuleuven.be/~dtai/clus/). The heuristic used for selecting the attribute tests (that define the internal nodes) in this algorithm is the intra-cluster variance summed over the subsets induced by the test. Intra-cluster variance is defined as $N \cdot \sum_{t=1}^{T} Var[y_t]$ with N the number of examples in the cluster, T the number of target variables, and $Var[y_t]$ the variance of target variable y_t in the cluster. The variance function is standardized so that the relative contribution of the different targets to the heuris-

tic score is equal. Lower intra-subset variance results in predictions that are more accurate.

The multi-target regression trees are an instantiation of the predictive clustering trees (PCTs) framework proposed in (Blockeel et al., 1998). In the PCTs framework, a tree is viewed as a hierarchy of clusters: the top node corresponds to one cluster containing all data, which is recursively partitioned into smaller clusters while moving down the tree. The PCTs can be instantiated for different tasks using adequate variance and prototype functions. So far, PCTs have been used to handle multiple targets (Struyf and Džeroski, 2006), time series (Džeroski et al., 2007) and hierarchical multi-label classification (Vens et al., 2008).

2.3. Ensembles

An ensemble method constructs a set of predictive models (called an ensemble) (Dietterich, 2000). An ensemble gives a prediction for a new data instance by combining the predictions of its models for that instance. For regression tasks, the predictions can be combined by averaging the outputs of the models.

In order for an ensemble to be more accurate than any of its individual members, the individual models need to be accurate and diverse (Hansen and Salamon, 1990). An accurate model is one that performs better than random guessing on new examples. A set of models is diverse if the models make different errors on new examples. The diversity in an ensemble can be introduced in various ways: by manipulating the training set (changing the weight of examples or changing the weight of attributes) or by manipulating the learning algorithm used to obtain the models.

Ensembles of MTRTs are sets of MTRTs, obtained by applying the same TDIDT algorithm. A prediction of an ensemble of MTRTs is obtained by averaging the predictions of its models. They are able to lift the predictive performance of a single MTRT (also in the case of a single target) (Breiman, 1996; Kocev et al., 2007). In this work, we use bagging and random forests, the two most widely used ensemble methods to produce ensembles of RTs and MTRTs. An illustration of these two methods is presented in Fig. 2.

2.3.1. Bagging

Bagging (Breiman, 1996) is an ensemble method that constructs the different models in the ensemble by making bootstrap replicates of the training set; these are used to construct individual models (Fig. 2). Each bootstrap sample is obtained by randomly sampling training instances, with replacement, from the original training set. The bootstrap sample and the training set have the same number of instances. Bagging can yield substantial gains in predictive performance, when applied to unstable learners (i.e., a learner for which small changes in the training set can result in large changes in the predictions), such as classification and regression tree learners (Breiman, 1996). The diversity in bagging comes from the variation in the training sets used to construct the individual models in the ensemble.

2.3.2. Random forests

A random forest (Breiman, 2001) is an ensemble of trees, where the diversity of the trees is obtained from two sources: (1) by using bootstrap sampling and (2) by changing the feature set during learning (this is done by a randomized decision tree algorithm, see Fig. 2). At each node in the decision tree, a random subset of the input features is taken and the best split is selected from this subset. The size of the random subset is given by a function F of the number of descriptive attributes M $\left(\text{e.g., } F = 1, F = \lfloor\sqrt{M}\rfloor, F = \lfloor\log_2 M\rfloor + 1, F = \lfloor M/2 \rfloor, \ldots\right)$. If $F = M$, then the random forests algorithm is equal to the bagging algorithm.

3. Data description

In this study, we use field data acquired using the habitat hectares approach (Parkes et al., 2003), a technique for the rapid assessment of vegetation condition, developed primarily for biodiversity conservation planning. 'Vegetation quality' in the 'habitat hectares' approach is defined as the degree to which the current vegetation differs from a 'benchmark' that represents the average characteristics of a mature and long-undisturbed stand of the same plant community. Against the benchmark, the decline in quality can be estimated for each vegetation type and dissimilar community assemblages, such as rainforests and savannahs can be compared by employing the same general index. This general approach has become a standard method used to quantify the condition of habitat within the state of Victoria (www.dse.vic.gov.au) and has been emulated to some degree by other jurisdictions within Australia (see Eyre et al., 2006; Gibbons et al., 2009).

The 'habitat hectares' score is the weighted sum of 7 site and 3 landscape scale metrics. The landscape components of the 'habitat hectares' score can be readily rendered spatially within a GIS using tools such as FRAGSTATS (McGarigal et al., 2002) and have not been further considered in this study. The objective was to make spatially explicit predictions of the 7 site scale components of the 'habitat hectares' score (hereafter referred to as the 'habitat hectares' site score or HHSS).

Employing the 'habitat hectares' approach, 16,967 'homogenous' sites were sampled within the State of Victoria, Australia (see Fig. 1) between the years 2001 and 2005. Each sampling point is described by 40 independent (or feature) variables (GIS and remote-sensed data with a pixel resolution of $30\,\text{m} \times 30\,\text{m}$) and 7 dependent (or target) variables (the HHSS). The HHSS is a numeric variable composed as a weighted average of the following components: *Large Trees*; *Tree (canopy) Cover*; *Understorey (non-tree) Strata*; *Lack of Weeds*; *Recruitment*; *Organic Litter*; and, *Logs*. Apart from *Lack of Weeds*, each component score was calculated comparing the current status of the vegetation with a benchmark. For a basic statistic of the target attributes see Table A2 in Appendix.

The *Large trees score* represents the number of large trees (both living and dead) that are present at the measuring site (compared to the 'benchmark' archetype). The *Tree Canopy* score assesses the
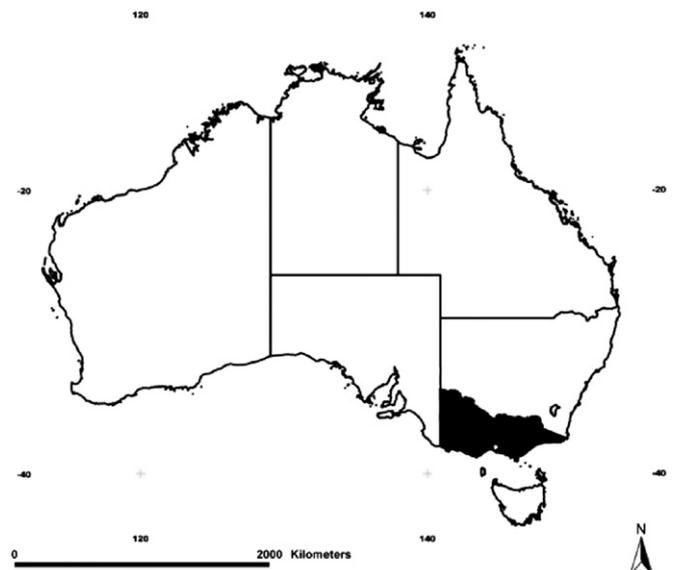
Fig. 1. Map of Australia with latitude and longitude shown. The State of Victoria in the south east of mainland Australia (our study area) is shaded.
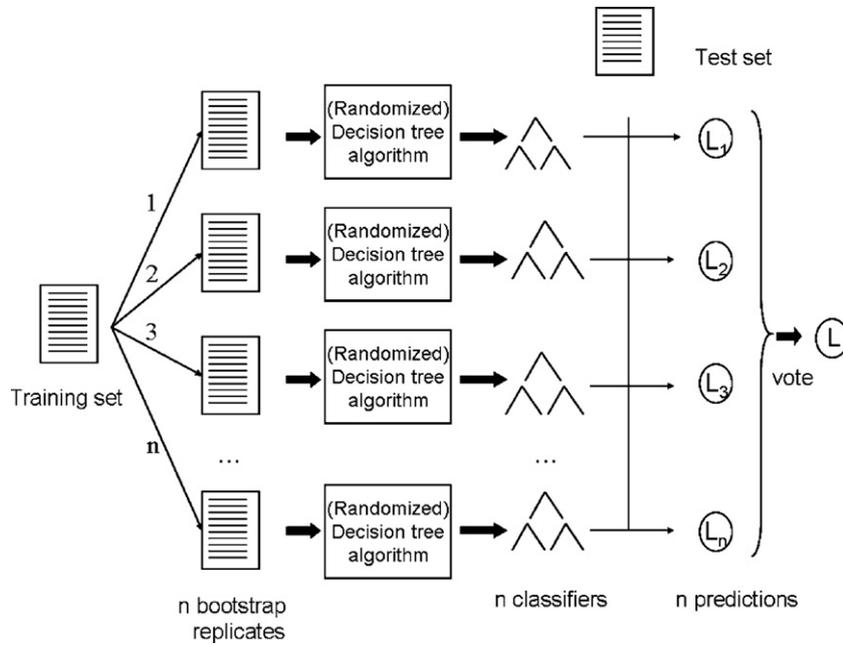
Fig. 2. A generic algorithm for learning ensembles of decision trees. Bagging uses a standard decision tree algorithm, while random forests use a randomized decision tree algorithm.
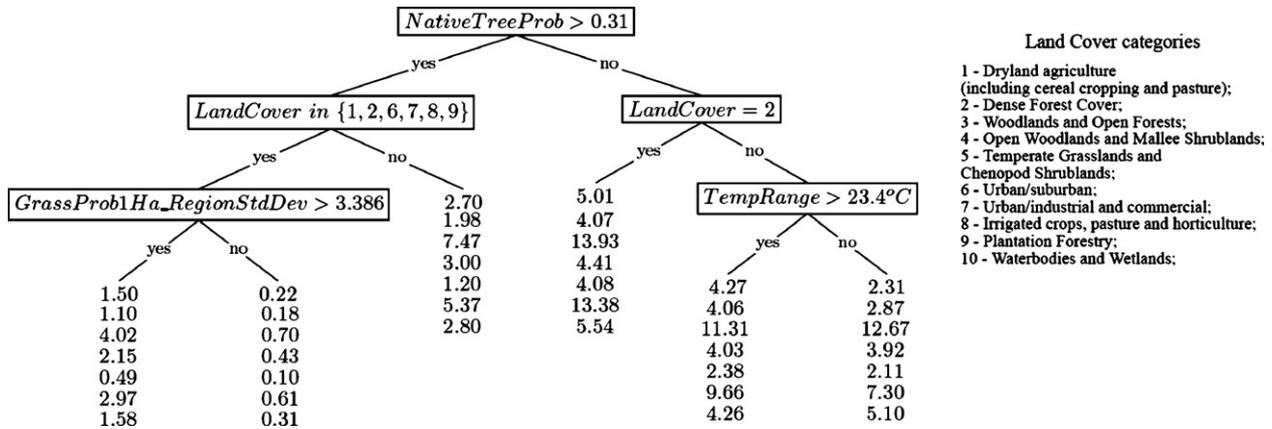
Fig. 3. Pruned multi-target regression tree (the target attributes in the leaves are ordered as per Table 1).

projective foliage cover of canopy trees in the stand, while the *Understorey score* assesses the abundance and diversity of various shrubs and forb/herb strata of a community. The understorey assessment includes only indigenous plant species. The *Lack of weeds score* is calculated from the cover of non-indigenous weed species.

The *Recruitment score* provides an indication of the level of regeneration of woody plant species and could be seen as a surrogate measure of the long-term viability of the site's structural characteristics. *Litter* represents both fine and coarse plant debris less than 10 cm diameter, while *Logs* represent the fallen timber or branches of trees that are substantially detached from the parent tree. An unabridged description of the 'habitat hectares' scores and methods can be found in (Parkes et al., 2003, 2004) and at www.dse.vic.gov.au.

The 40 independent variables include 39 continuous variables and one categorical variable (see Appendix Table A1). The categorical variable *LandCover* surface was derived from Landsat 7 TM spectral data. Classes were obtained by applying a k-means clustering procedure to a stack of median values for all Landsat 7 TM spectral bands and the Normalised Difference Vegetation Index (see Tucker, 1979) across the years spanning 1989–2005. The 50 classes that emerged from the unsupervised classification were 'lumped' into 10 bins that were partially informed by a landuse model similarly derived using an ANN process. This procedure allowed for temporal states consequent of clearing, wildfire and forest harvesting to remain evident within broad landuse classes. The 10 categories approximate to the descriptions in Fig. 3.

4. Experimental setup for data analysis

From the description of the data, we can define a multi-target regression problem, to be solved either by the single-target or the multi-target regression approach. The goal is to predict multiple continuous targets (responses, outputs) from a vector of descriptive (independent) variables. When applying the single-target approach, we learn a regression tree (or an ensemble of regression trees) for each target attribute separately (in our case, this means that we will have seven models or ensembles). With the multi-target approach we learn a multi-target regression tree (or ensemble of multi-target regression trees) for all target attributes (meaning that the output is a single model or ensemble).

We define two experimental scenarios. In the first scenario, the purpose of the modelling is to learn about the condition of the indigenous vegetation, and the relative importance of different biophysical and landscape attributes for that condition. We focus on interpretability to obtain such knowledge: the models need to have reasonable size and predictive power. We prune our models by setting the minimal number of instances in a leaf to 2048 (for both the single-target and multi-target approach). We varied this pruning parameter starting from 4 up till 4096 (taking numbers that are power of 2). We selected 2048, because it offered the best trade-off between the size and the performance of the model.

In the second scenario, we are not interested in the size of the models, but in their predictive power. To improve predictive performance, we use ensembles of unpruned single- and multi-target regression trees. We constructed ensembles consisting of 100 unpruned trees as recommended in (Bauer and Kohavi, 1999; Breiman, 1996, 2001). To combine the predictions of the trees we averaged the predictions from each tree. The size of the feature subsets for the random forests (F) was set to $F = \lfloor \log_2 M \rfloor + 1$ as suggested in (Breiman, 2001).

The learned models, from both scenarios, were then used to derive maps of remnant indigenous vegetation condition. Combined with other data, these maps will contribute to investment decisions in natural resource management, statutory protection and reserve design.

We compare the single-target and multi-target regression trees and ensembles. For baseline comparison, we use linear regression (as implemented in the WEKA system, Witten and Frank, 2005). We compared the methods in terms of their predictive performance (correlation coefficient between predictions and observed values, and root mean squared error—RMSE), time efficiency and model size. To estimate the predictive performance of the models on unseen data, we employed 10 times 10-fold cross-validation, thus we present the performance results with respective confidence intervals.

To assess whether the differences in performance are statistically significant, we employed the corrected Friedman test (Friedman, 1940) and the post hoc Nemenyi test (Nemenyi, 1963) as recommended by Demšar (2006). The Friedman test is a non-parametric test for multiple hypotheses testing. It ranks the algorithms according to their performance for each dataset separately, thus the best performing algorithm gets the rank of 1, second best the rank of 2..., and in case of ties it assigns average ranks (see Tables A2 and A3 in Appendix). Then, the Friedman test compares the average ranks of the algorithms and calculates the Friedman statistic χ_F^2, distributed according to the χ_F^2 distribution with $k - 1$ degrees of freedom (k being the number of algorithms). Iman and Davenport (1980) show that the Friedman statistic is undesirably conservative and derive a corrected F-statistic that is distributed according to the F-distribution with $k - 1$ and $(k - 1) \times (N - 1)$ degrees of freedom (k being the number of algorithms and N being the number of datasets).

If there is a statistically significant difference in the performance, than we can proceed with a post hoc test. The Nemenyi test is used to compare all the classifiers to each other. In this procedure, the performance of two classifiers is significantly different if their average ranks differ more than some critical distance. The critical distance depends on the number of algorithms, number of datasets and critical value (for a given significance level) that is based on the Studentized range statistic and can be found in statistical textbooks.

We present the result from the Nemenyi post hoc test with an average ranks diagram as suggested by Demšar (2006). An average ranks diagram can be seen in Fig. 6 (and Figure A1 in Appendix). The ranks are depicted on the axis, in such a manner that the best ranking algorithms are at the right-most side of the diagram. The algorithms that do not differ significantly are connected with a line.

5. Interpretation and evaluation of the vegetation condition models

We followed the analysis scenarios, described in the previous section and obtained two sets of models. The first set consists of single models (single-target regression trees and multi-target regression trees) and is concerned with the process of knowledge extraction (the first scenario). The second set consists of ensembles (of single-target and multi-target regression trees) and is concerned with better predictive power (the second scenario). All models are presented and discussed in the next subsections.

5.1. Models for knowledge extraction

In this sub-section, we present and discuss the models that were obtained with the first scenario described in Section 4. This set of models contains single-target regression trees for each target and one multi-target regression tree for all targets. We compared the performance of the models (Table 1), with both approaches yielding models of comparable predictive performance. The difference is in the interpretability and the time and size efficiency. The time needed for learning the MTRT was 2.33 s, while learning all regression trees takes 13.77 s (a speed-up of factor 5.9). The speed can be very important in real-time applications. Also, the MTRT is of size 11 (total number of nodes), while all single-target regression trees taken together have size 81 (a ratio of 7.4). These models are depicted in Figs. 3 (MTRT) and 4 (single-target trees).

One of the most important differences between the two approaches is in their interpretability. It is much easier to interpret one tree that describes all target variables, than interpreting each regression tree separately and trying to find some connection between the different models. The multi-target regression tree gives us a more general overview of the knowledge that is hidden in the data.

The pruned multi-target regression tree shown in Fig. 3 is readily interpreted, grouping the data into six clusters. The clusters that are in the right-hand side have (on average) a higher HHSS, indicating that such sites are likely to support indigenous vegetation close to its benchmark state. An intuitively robust, if somewhat simplified overview of vegetation condition across the State of Victoria is provided by a map generated from the multi-target solution and applied to the spatial covariates (Fig. 5).

The key variable at the initial node of the tree is *NativeTreeProb* which is the prediction of a Neural Network model (ANN) of the probability of a lack of native tree cover for Victoria, informed by a chronosequence of Landsat imagery from 1989 to 2005. A *NativeTreeProb* > 0.31 is equivalent to a predicted probability of greater than 0.31 of the subject pixel supporting tree cover. Given that three of the sub-components of the HHSS depend directly on the presence of tree cover (*Large tree score*, *Canopy cover score* and *Logs score*), its central role in partitioning the data is logical.

Table 1
Comparison of the performance of the pruned multi-target regression tree for all scores with the regression trees for each score (MTRT—multi-target regression tree, RT—regression tree).

Target	Correlation		RMSE	
	MTRT	RT	MTRT	RT
Large tree score	0.52 ± 0.02	0.53 ± 0.02	2.88 ± 0.06	2.86 ± 0.06
Tree canopy score	0.68 ± 0.02	0.68 ± 0.01	1.63 ± 0.04	1.64 ± 0.03
Understorey score	0.70 ± 0.02	0.71 ± 0.02	5.11 ± 0.13	5.05 ± 0.13
Litter score	0.72 ± 0.02	0.69 ± 0.02	1.43 ± 0.03	1.47 ± 0.04
Logs score	0.70 ± 0.02	0.71 ± 0.02	1.48 ± 0.03	1.47 ± 0.03
Weeds score	0.75 ± 0.01	0.78 ± 0.01	4.04 ± 0.09	3.83 ± 0.10
Recruitment score	0.61 ± 0.02	0.62 ± 0.02	2.59 ± 0.07	2.57 ± 0.06
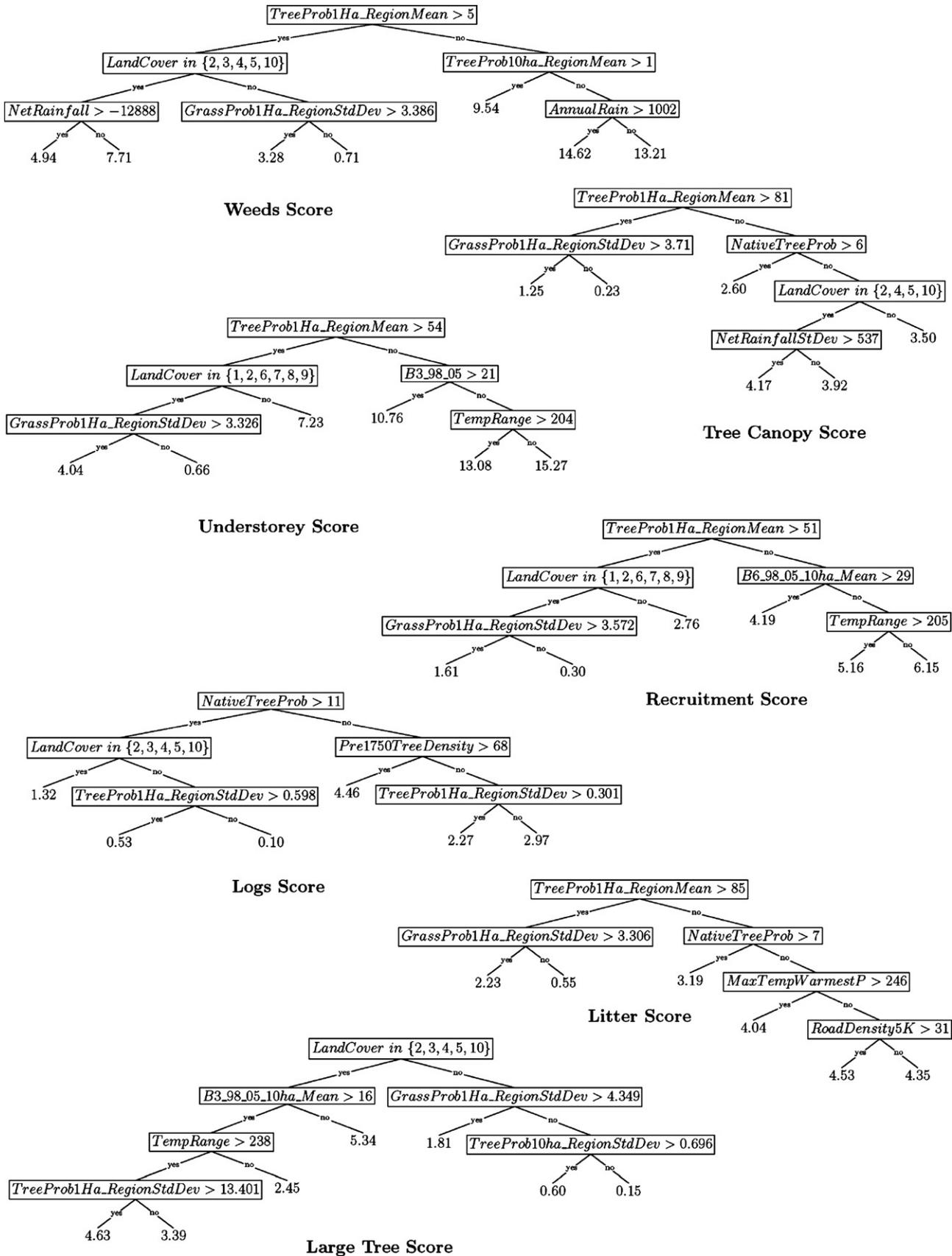
Fig. 4. Regression trees for each *Habitat Hectares* site score. The sum of these attributes comprises the overall *Habitat Hectares* site score.
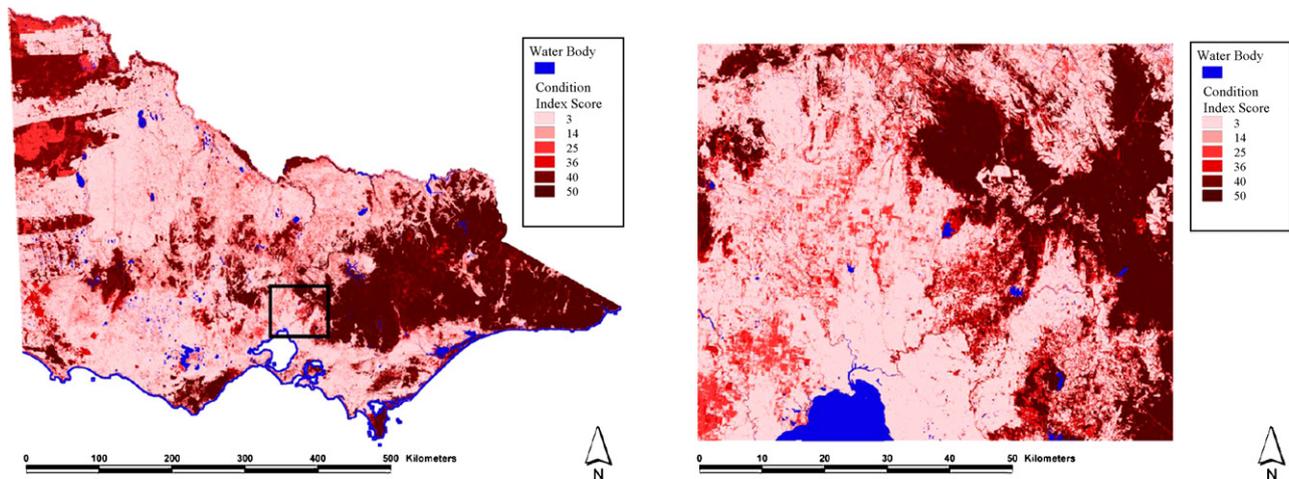
Fig. 5. Map of the condition of indigenous remnant vegetation in Victoria derived from the application of the pruned multi-target regression tree model to the explanatory features (left-hand side figure). The dark bordered rectangular inset refers to the area represented at higher resolution in the right-hand side figure.

Focusing on the 'no' branch of the tree (with the higher HHSS) the next decision node pivots on the membership of data to the *LandCover* category 2. *LandCover* category 2 corresponds with dense comparatively undisturbed forest cover and has the highest overall site score (of 50). All other *LandCover* categories proceed to the next node that partitions further, employing the variable *TempRange*. *TempRange* is one of many climate variables or features created using the ANUCLIM software package (see Houlder et al., 2000). This variable describes the annual range in temperature at a site by subtracting the climate model for the minimum temperature of coldest period of the year from the maximum temperature of the warmest period of the year. A *TempRange* of greater than 23.4 °C can be found in the semi-arid North West of Victoria where plant growth rates and consequently recovery from perturbation is generally slow.

The left-hand side of the tree, where the probability of tree cover is smaller than 0.31, is further partitioned by membership or otherwise of the *LandCover* categories 1, 2, 6, 7, 8, and 9. Apart from *LandCover* category 2 (i.e., Dense Forest Cover) these land cover types are all highly modified land use settings with correspondingly low habitat hectare scores. The small areas with *LandCover* category 2 that have a high probability of not finding tree cover (i.e. greater than 0.31) are likely to be feature data errors carried over from the land use mapping employed.

All these categories when *NativeTreeProb* is greater than 0.31 are assigned moderate condition scores (mean 25) by the pruned regression tree model. These are predominantly areas where tree cover is either absent, partially cleared or tree cover has been removed by recent wildfires. Fire scars are apparent in the North West region of the map. The incidence of fire has not been explicitly addressed in this study, however, future modelling will investigate the impact of fire on the HHSS and other condition indices through the inclusion of mapped fire boundaries derived from satellite imagery and historic cartographic sources.

The final node in the multi-target regression tree to be discussed here is regulated by the variable *Grass1ha RegionStdDev*. This variable is derived from an ANN model of the probability of native grass cover for every pixel in Victoria, informed by aforementioned chronosequence of Landsat imagery. A neighborhood of 1 ha around each pixel was interrogated and the standard deviation of the probability of indigenous grass cover across that area was obtained. Although speculative, this variable identifies spectrally uniform areas—regions that if they support treeless native vegetation could be relatively free of the degrading edge effects such as weed invasion that may emanate from surrounding land uses. The variable may be interpreted as a surrogate for the core area concept in landscape ecology (*sensu* Botequila Leitão et al., 2006) seen here as a useful predictor of grassland vegetation condition in Victoria.

The regression trees for each target attribute are shown in Fig. 4. If we compare Figs. 3 and 4, each of the components of the habitat hectare site score use different features and sequences of features to that of the tree that predicts the site score alone. This adds complexity and removes ecological naivety from the model. As with the single-target solution, we can closely examine the internal logic of each regression tree for the component scores. Prima-facie, each of the single-target regression trees is ecologically interpretable.

For example, if we just follow the positive or far left-hand side of the tree predicting Weed Score, it initially partitions the data on the basis of *TreeProb1HaRegionMean*: mean probability of detecting no tree cover within a 1 ha area around the subject pixel. This variable effectively divides the landscape into forests and treeless areas or areas with only scattered trees. Following the positive or left-hand side of the tree the data is further partitioned by the land cover classes. Classes 2, 3, 4, 5, and 10—all of these classes are natural or semi-natural areas and we should expect these areas to have a higher weed score (a high positive score reflects the absence of weeds rather than weed infestation) relative to other thinly treed areas. This is borne out by the regression tree. The final node is controlled by *NetRainfall*. *NetRainfall* is a variable that is derived from both mean monthly rainfall and mean evaporation rates. In essence it reflects the amount of effective rainfall (rainfall less evaporation) over an entire year. Once we have reached this node the model predicts that the drier and hotter a place is, the higher the weed score (provided we have satisfied earlier criteria). This reflects the current on-ground ecological reality in south-eastern Australia where there have been few deliberate introductions of exotic plant species into specialist habitat types, such as semi-arid regions, in comparison with temperate and sub-humid climatic regions that have been favoured by human settlement and intensive agriculture.

A further advantage of the multi-target approach is that is can reveal relationships between response variables. It is apparent that *Recruitment score* and *Understorey score* are positively related (see Fig. 4). The single-target regression trees of these scores are structurally identical and both employ very similar explanatory variables at similar junctures. Again, this is consistent with both field observation and ecological theory: a diverse and structurally intact understorey implies an adequate level of shrub and tree regeneration. The reverse is also likely. Within defined ecosystem types and states, a positive relationship between ecosystem function and structure is generally accepted by ecologists (Cortina et al., 2006; Bradshaw, 1984). Overall, the most important variables influ-

Table 2
Correlation coefficients of the obtained models (LR—linear regression, MTRT—multi-target regression tree, RT—regression tree, Bag—bagging, RF—random forests).

Target	LR	MTRT	RT	BagMTRT	Bag RT	RF MTRT	RF RT
Large tree score	0.61 ± 0.02	0.63 ± 0.02	0.60 ± 0.02	0.69 ± 0.01	0.69 ± 0.02	0.69 ± 0.01	0.69 ± 0.01
Tree canopy score	0.76 ± 0.01	0.76 ± 0.01	0.74 ± 0.02	0.80 ± 0.01	0.81 ± 0.01	0.81 ± 0.01	0.81 ± 0.01
Understorey score	0.77 ± 0.02	0.78 ± 0.01	0.77 ± 0.01	0.83 ± 0.01	0.83 ± 0.01	0.83 ± 0.01	0.83 ± 0.01
Litter score	0.76 ± 0.01	0.77 ± 0.01	0.76 ± 0.01	0.81 ± 0.01	0.82 ± 0.01	0.82 ± 0.01	0.82 ± 0.01
Logs score	0.75 ± 0.01	0.76 ± 0.01	0.75 ± 0.02	0.80 ± 0.01	0.80 ± 0.01	0.80 ± 0.01	0.80 ± 0.01
Weeds score	0.82 ± 0.01	0.83 ± 0.01	0.83 ± 0.01	0.87 ± 0.01	0.87 ± 0.01	0.87 ± 0.01	0.87 ± 0.01
Recruitment score	0.67 ± 0.02	0.69 ± 0.02	0.67 ± 0.02	0.74 ± 0.02	0.74 ± 0.01	0.74 ± 0.01	0.75 ± 0.01

Table 3
Root mean squared error of the obtained models (LR—linear regression, MTRT—multi-target regression tree, RT—regression tree, Bag—bagging, RF—random forests).

Target	LR	MTRT	RT	BagMTRT	Bag RT	RF MTRT	RF RT
Large tree score	2.66 ± 0.05	2.62 ± 0.05	2.72 ± 0.06	2.43 ± 0.05	2.44 ± 0.06	2.44 ± 0.05	2.43 ± 0.05
Tree canopy score	1.46 ± 0.03	1.45 ± 0.03	1.52 ± 0.04	1.33 ± 0.03	1.32 ± 0.03	1.32 ± 0.03	1.32 ± 0.03
Understorey score	4.59 ± 0.16	4.47 ± 0.13	4.58 ± 0.15	4.04 ± 0.12	4.04 ± 0.12	4.05 ± 0.11	4.03 ± 0.11
Litter score	1.34 ± 0.03	1.30 ± 0.03	1.34 ± 0.03	1.19 ± 0.03	1.18 ± 0.03	1.18 ± 0.03	1.18 ± 0.03
Logs score	1.37 ± 0.03	1.35 ± 0.03	1.39 ± 0.04	1.25 ± 0.03	1.26 ± 0.03	1.25 ± 0.03	1.25 ± 0.03
Weeds score	3.48 ± 0.09	3.41 ± 0.09	3.49 ± 0.10	3.01 ± 0.08	3.01 ± 0.08	3.02 ± 0.08	3.01 ± 0.08
Recruitment score	2.41 ± 0.08	2.35 ± 0.07	2.43 ± 0.08	2.18 ± 0.07	2.18 ± 0.06	2.18 ± 0.06	2.18 ± 0.06

encing all components of the HHSS are those immediately related to (the probability) of (indigenous and non-native) tree cover (such as *NativeTreeProb* that appears in the root of the multi-target tree, and *TreeProb1HaRegionMean*, which appears in the roots of 5/7 single-target trees). It is interesting to note that this is the case also for the sub-components that do not depend directly on the presence of tree cover, e.g. *Weeds Score*. Following closely is *LandCover* (as modelled from satellite images), with dense forest cover (category 2) yielding high HHSS scores. Finally, climate plays an important role, with variables describing temperatures, rainfall and their variability appearing in most of the models.

5.2. Models with superior predictive performance

This sub-section presents and discusses the models obtained with the second scenario (see Section 4). Here, we compare linear regression, multi-target regression trees, regression trees, ensembles of multi-target regression trees and ensembles of regression trees to investigate the possible improvements in prediction performance (Tables 2 and 3) and computational efficiency (Table 4) that can be achieved by ensemble methods.

We present the predictive performance of the obtained models in terms of their correlation coefficient s and RMSEs. The results are presented with the corresponding confidence intervals, to show the stability of the used algorithms. Recall that 10 times 10-fold cross-

validation was used to estimate the performance on unseen data. We can note that the confidence intervals are small. This is due to the size of the dataset (16,967 samples).

To check whether the differences in performance are of statistical significance, we used the corrected Friedman test for multiple hypothesis testing. To detect which algorithms perform significantly better or worse than the others we used the Nemenyi post hoc test. The result of the corrected Friedman test is that the difference in performance of these algorithms is statistically significant with a *p-value* smaller than 0.01. The results of the Nemenyi post hoc test for the RMSE comparison are presented in Fig. 6 with an average ranks diagram. On the axis the algorithms are plotted according to their average rank. The best performing algorithm is random forests with single-target regression trees, while the worst performing algorithm is the single-target regression tree. The critical distance is calculated for the significance level of 0.05.

The Nemenyi test shows that the performance of the ensemble methods (in terms of RMSE) is significantly better than the one of individual trees. The ensembles from both MTRTs and RTs are not significantly better than the single MTRT (at $p = 0.05$). However, the ensembles of MTRTs (both bagging and random forests) and the random forests of RTs are significantly better than linear regression and single-target regression trees. The difference in performance between MTRTs, RTs and linear regression is not statistically sig-
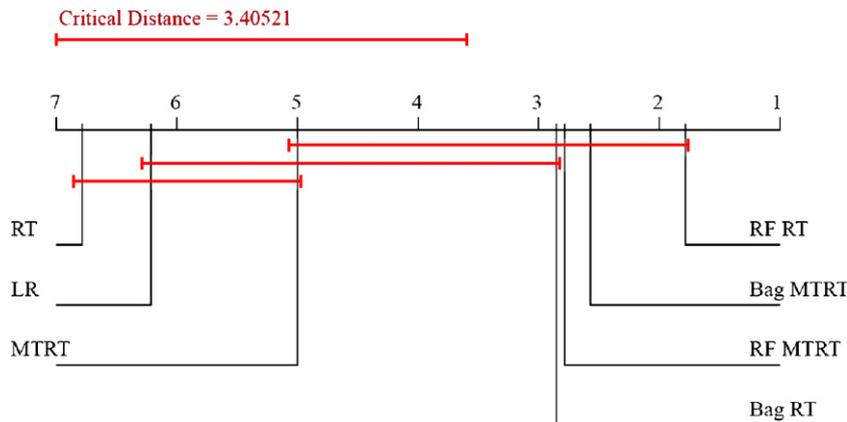
Fig. 6. Average ranks diagram for the applied algorithms (comparing by RMSE). Algorithms that do not differ significantly ($p = 0.05$) are connected with a line.

Table 4
Comparison of the time and size efficiency of the algorithms (LR—linear regression, MTRT—multi-target regression tree, RT—regression tree, Bag—bagging, RF—random forests).

	LR	MTRT	RT	Bag MTRT	Bag RT	RF MTRT	RF RT
Time (s)	8.06	7.18	36.18	430.94	2053.50	87.69	385.38
Size	332	345	4729	10,639.94	35,145.02	10,907.66	43,030.76
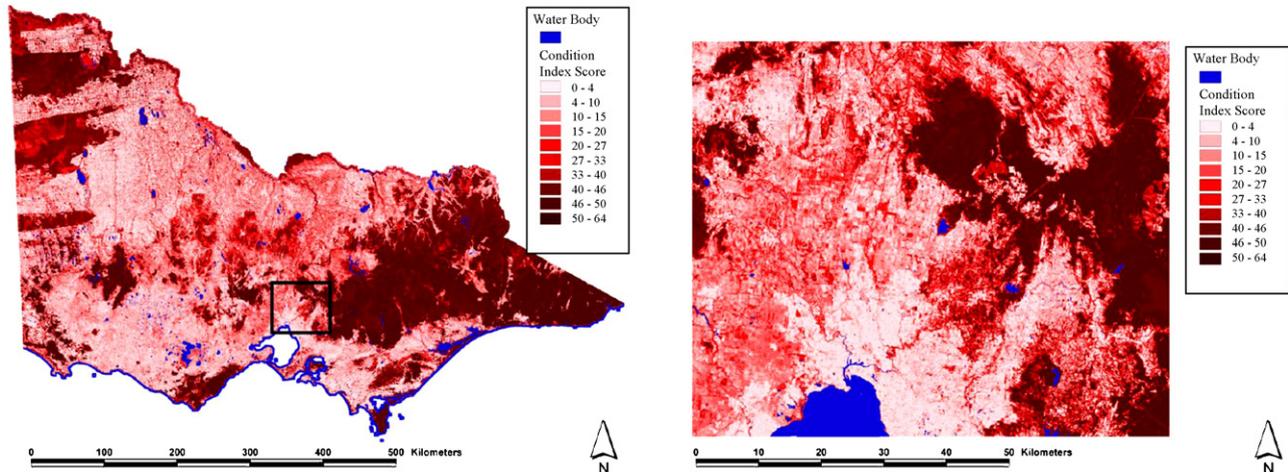
Fig. 7. Map of the condition of indigenous remnant vegetation in Victoria derived from the application of the random forests of MTRTs (left-hand side figure). The dark bordered rectangular inset refers to the area represented at higher resolution at the right-hand side figure.

nificant. Similar conclusions can be drawn if instead of the results for RMSE we consider the results for the correlation coefficient (Figure A1 from Appendix).

In addition, we compared the approaches by their time and size efficiency (Table 4). For the single-target scenarios (linear regression, regression trees, bagging of regression trees and random forests of regression trees) the time efficiency is calculated as the sum of the times used to learn a model for each target separately. The size of a linear regression model is the number of terms in the equation. The size of a MTRT is calculated as the number of nodes in the tree, while the size of a regression tree is the sum of the number of nodes in the trees over all targets. For bagging and random forests of multi-target regression trees, the size efficiency is the sum of size of the trees in the ensemble, while for the bagging and random forests of regression trees the size is the sum of the sizes of the ensembles for each target.

When comparing ensemble methods, the speed-up ratio of multi-target over single-target tree models remains high (4.5 on average), while the size of the multi-target tree models is about 0.25 of the size of single-target tree models. Multi-target regression consistently delivers models that have equally good predictive power, but are smaller and faster to learn (and apply). Linear regression has comparable time and size efficiency with multi-target regression models.

Overall, random forests of multi-target regression trees should be preferred, given that they improve the predictive performance and stability of multi-target trees in general, and are not as computationally expensive as bagging.

The spatially explicit map produced by the MTRT random forest ensemble, provides a subtle and accurate reflection of the condition of indigenous vegetation across the State of Victoria (Fig. 7). As we can see in the detailed inset, the modelled condition is finely resolved and nuanced, responding appropriately to local conditions, land use and land tenure. Application of the models allows for their further evaluation by experts familiar with local study areas. Such an evaluation is an ongoing process—but preliminary assessment indicates that the random forest MTRT is a robust

model across a wide range of landscape, landuse and historical contexts.

6. Conclusions

In this work, we model the condition of remnant indigenous vegetation with machine learning techniques. The condition of the vegetation is described by multiple (habitat hectares) scores that reflect the structural and compositional attributes of a wide variety of plant communities. To model the multiple scores, we used two approaches: single-target and multi-target regression. With single-target regression we learn a model for each score separately, while with the multi-target regression we learn one model for all scores. The results show the advantages of multi-target over single-target regression: multi-target models have a smaller size and are faster to learn and apply. Also, there is no statistically significant difference in their predictive power.

We performed two sets of experiments. With the first set we were interested in knowledge extraction, and with the other we opted for models that have better predictive power. For knowledge extraction, we used pruned regression trees and pruned multi-target regression trees. The goal was to better understand the resilience of some indigenous vegetation types and the relative importance of biophysical and landscape attributes that influence their condition. From the learned models, we can conclude that the most important variables influencing all scores are those related to tree cover. This holds also for scores that do not depend directly on the presence of tree cover. Land cover is also of high importance, with dense forest cover yielding high scores. Finally, climate (including the variability of weather conditions) also plays an important role.

Predictive power and efficiency was an imperative for the selection of the preferred model from the second set of experiments. In order to obtain models that have high predictive power we used unpruned regression trees, ensembles of regression trees, unpruned multi-target regression trees and ensembles of multi-target regression trees. Given the results of the statistical tests for the predictive

power, and the time and size efficiency, the random forests of multi-target regression trees should be preferred.

An important consideration of model utility is the spatial aspect at which the models are to be used and the specific purpose for which the model has been developed. The development of both single trees and ensembles of trees has highlighted the trade-off in model selection between complexity and predictive power on one hand and interpretability on the other. The pruned single tree based solutions to the prediction problem are transparent and facilitate almost immediate interpretation and qualitative evaluation by a range of users with varying degrees of understanding of the underlying learning algorithm. However, due to their simplicity, the predictions of single (pruned) trees as rendered by mapping produce generalized surfaces apparently devoid of the heterogeneity and subtlety of the real world. This may be a useful outcome if the objective is to produce a simple model. Conversely, due to the high predictive power, the ensemble models provide for the complexity and fine scale accuracy absent from the single trees, but are not readily interpretable to users.

It is apparent from this study that complex weighted metrics such as the habitat hectare index of vegetation condition can be modelled across extensive areas with some predictive confidence, using easily obtained remotely acquired data and provided adequate field data is collected. Such products can provide a 'snapshot' of the prevailing conditions and provide investment and decision support for natural resource managers.

We intend to extend out work in several directions. We hope to use new features that summarise relevant past and prevailing environmental disturbances and land uses, with a view to improving spatial models of vegetation condition, while realising some view of condition trajectory. In addition, we intend to develop spatially explicit models of both the untransformed and unweighted field measures that inform each of the components of the HHSS and the benchmark or reference values for these measures. Finally, we are interested in investigating the potential for implementing cost-sensitive learning to reflect heightened regulatory, planning or investment interest in particular geographic regions or particular index value ranges.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ecolmodel.2009.01.037.

References

Andreasen, J.K., O'Neill, R.V., Noss, R., Slosser, N.C., 2001. Considerations for the development of a terrestrial index of ecological integrity. Ecological Indicators 1 (1), 21–35.
Bauer, E., Kohavi, R., 1999. An empirical comparison of voting classification algorithms: bagging, boosting, and variants. Machine Learning 36 (1–2), 105–139.
Botequila Leitão, A., Miller, J., Ahern, J., McGarigal, K., 2006. Measuring Landscapes: A Planner's Handbook. Island Press, Washington, DC.
Blockeel, H., De Raedt, L., Ramon, J., 1998. Top-down induction of clustering trees. In: 15th Int'l Conf. on Machine Learning, pp. 55–63.
Blockeel, H., Struyf, J., 2002. Efficient algorithms for decision tree cross-validation. Journal of Machine Learning Research 3, 621–650.
Bradshaw, A., 1984. Ecological principles and land reclamation practice. Landscape Planning 11, 35–48.
Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. Classification and Regression Trees. Wadsworth, Belmont.
Breiman, L., 1996. Bagging predictors. Machine Learning 24 (2), 123–140.
Breiman, L., 2001. Random forests. Machine Learning 45 (1), 5–32.
Cortina, J., Maestre, F.T., Vallejo, R., Baeza, M.J., Valdecantos, A., Pérez-Devesa, M., 2006. Ecosystem structure, function, and restoration success: are they related? Journal for Nature Conservation 14 (3–4), 152–160.
Demšar, J., 2006. Statistical comparisons of classifiers over multiple data sets. Journal of Machine Learning Research 7, 1–30.
Dietterich, T., 2000. Ensemble methods in machine learning. In: Proc. of the 1st Int'l Workshop on Multiple Classifier Systems, vol. 1857 of LNCS. Springer, Berlin, pp. 1–15.
Džeroski, S., Gjorgjioski, V., Slavkov, I., Struyf, J., 2007. Analysis of time series data with predictive clustering trees. LNCS 4747, 63–80.
Eyre, T.J., Kelly, A.L., Neldner, V.J., 2006. BioCondition: A Terrestrial Vegetation Condition Assessment Tool for Biodiversity in Queensland. Field Assessment Manual. Version 1. 5. B. S. u. Environmental Protection Agency, Brisbane.
Friedman, M., 1940. A comparison of alternative tests of significance for the problem of m rankings. Annals of Mathematical Statistics 11, 86–92.
Gibbons, P., Briggs, S.V., Ayers, D., Seddon, J., Doyle, S., Cosier, P., McElhinny, C., Pelly, V., Roberts, K., 2009. An operational method to assess impacts of land clearing on terrestrial biodiversity. Ecological Indicators 9, 26–40.
Gibbons, P., Zerger, A., Jones, S., Ryan, P., 2006. Mapping vegetation condition in the context of biodiversity conservation. Ecological Management and Restoration 7, S1–S2.
Guisan, A., Thuiller, W., 2005. Predicting species distribution: offering more than simple habitat models. Ecology Letters 8 (9), 993–1009.
Hansen, L., Salamon, P., 1990. Neural network ensembles. IEEE Transactions on Pattern Analysis and Machine Intelligence 12, 993–1001.
Houlder, D.J., Hutchinson, M.F., Nix, H.A., McMahon, J.P., 2000. NUCLIM User Guide, Version 5.1. Centre for Resource and Environmental Studies, Australian National University, Canberra, Australia.
Iman, R.L., Davenport, J.M., 1980. Approximations of the critical region of the Friedman statistic. Communications in Statistics, 571–595.
Knight, A.T., Cowling, R.M., Campbell, B.M., 2006. An operational model for implementing conservation action. Conservation Biology 20 (2), 408–419.
Kocev, D., Vens, C., Struyf, J., Džeroski, S., 2007. Ensembles of multi-objective decision trees. In: Machine Learning: ECML 2007, 18th European Conference on Machine Learning, Proceedings, volume 4701 of LNCS, pp. 624–631.
Li, H., Wu, J., 2004. Use and misuse of landscape indices. Landscape Ecology 19 (4), 389–399.
McGarigal, K., Cushman, S.A., Neel, M.C., 2002. FRAGSTATS: Spatial Pattern Analysis Program for Categorical Maps. University of Massachusetts, Amherst (Computer software program produced by the authors at the University of Massachusetts, Amherst).
Margules, C., Pressey, B., 2000. Systematic conservation planning. Nature 405, 243–253.
Moilanen, A., 2007. Landscape zonation, benefit functions and target-based planning: unifying reserve selection strategies. Biological Conservation 134 (4), 571–579.
Nemenyi, P.B., 1963. Distribution-free multiple comparisons. PhD Thesis. Princeton University, Princeton, NY, USA.
Oliver, I., 2004. A framework and toolkit for scoring the biodiversity value of habitat, and the biodiversity benefits of land use change. Ecological Management & Restoration 5 (1), 75–78.
Parkes, D., Lyon, P., 2006. Towards a national approach to vegetation condition assessment that meets government investors needs: a policy perspective. Ecological Management and Restoration 7, S3–S5.
Parkes, D., Newell, G., Cheal, D., 2003. Assessing the quality of native vegetation: the habitat hectares approach. Ecological Management and Restoration 4, S29–S38.
Parkes, D., Newell, G., Cheal, D., 2004. The development and raison d'etre of 'habitat hectares': a response to McCarthy et al. (2004). Ecological Management and Restoration 5 (1), 28–29.
Rouget, M., Cowling, R.M., Lombard, A.T., Knight, A.T., Kerley, G.I.H., 2006. Designing large-scale conservation corridors for pattern and process. Conservation Biology 20 (2), 549–561.
Scholes, R.J., Biggs, R., 2005. A biodiversity intactness index. Nature 434, 45–49.
Struyf, J., Džeroski, S., 2006. Constraint based induction of multi-objective regression trees. In: Knowledge Discovery in Inductive Databases, 4th Int'l Workshop, KDID'05, Revised, Selected and Invited Papers, vol. 3933 of LNCS, pp. 222–233.
Tucker, C.J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. Remote Sensing of Environment 8 (2), 127–150.
Vens, C., Struyf, J., Schietgat, L., Džeroski, S., Blockeel, H., 2008. Decision trees for hierarchical multi-label classification. Machine Learning 73 (2), 185–214.
Witten, I.H., Frank, E., 2005. Data Mining: Practical Machine Learning Tools and Techniques, 2nd edition. Morgan Kaufmann, San Francisco.