Ecological Informatics 37 (2017) 66–76

Contents lists available at ScienceDirect

Ecological Informatics

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

Performance of one-class classifiers for invasive species mapping using airborne imaging spectroscopy

Sandra Skowronek [a,b,*], Gregory P. Asner [b], Hannes Feilhauer [a]

[a] FAU Erlangen-Nürnberg, Wetterkreuz 15, 91058, Erlangen, Germany
[b] Carnegie Institution for Science, 260 Panama St, Stanford, CA 94305, USA

ARTICLE INFO

Article history:
Received 29 August 2016
Received in revised form 17 November 2016
Accepted 17 November 2016
Available online 25 November 2016

Keywords:
Maxent
Biased support vector machine
Boosted regression trees
Imaging spectroscopy
Vegetation
Species distribution model
Yellow starthistle
Centaurea solstitialis
Phalaris aquatica
Carnegie airborne observatory
Hyperspectral

ABSTRACT

Most remote sensing approaches for mapping invasive plant species focus on species in a prominent phenological stage, such as during flowering, and do not systematically evaluate the performance for mapping lower cover fractions. In this study, we used airborne imaging spectroscopy (also known as hyperspectral imaging) to detect the invasive grass *Phalaris aquatica* and the invasive herb *Centaurea solstitialis* in a pre-flowering stage in the Jasper Ridge Biological Preserve, California, and compared the performance of three different one-class classifiers: Maxent, biased support vector machines and boosted regression trees.

We collected presence data for *C. solstitialis* and *P. aquatica* to calibrate each approach and additional presence-absence data to validate model performance on 3 m × 3 m plots. The imaging spectroscopy data were acquired using the Carnegie Airborne Observatory Visible-to-Shortwave Infrared (VSWIR) imaging spectrometer (400–2500 nm range) with a pixel size of 1 m × 1 m.

The resulting overall accuracies were 72–74% for *C. solstitialis*, and 83–88% for *P. aquatica*. For both species, the overall performance was slightly better for Maxent and BRT than for biased SVM. The detection rates for low cover plots were considerably higher for *C. solstitialis* than for *P. aquatica*. The models relied on different areas of the reflectance spectrum, but still produced the same general pattern of predicted species occurrences. We conclude that the different one-class classifiers allow for the detection and monitoring of target species with similar success rates.

1. Introduction

1.1. Remote sensing of invasive plant species using one-class classifiers

Invasive species are among the most important drivers of biodiversity loss, and their impact on Mediterranean drylands is rapidly increasing (Millenium Ecosystem Assessment, 2005). To efficiently manage invasive species, early detection and monitoring are crucial. Remote sensing has the potential to contribute to biological invasion monitoring and forecasting, and allows for the coverage of large areas (He et al., 2011; Rocchini et al., 2015). We may use remote sensing to directly or indirectly detect invasive species through distribution modeling. While multispectral data can be used in some cases to map larger stands of species with characteristic features, most studies have employed imaging spectroscopy datasets acquired with airborne sensor systems for invasive plant species mapping (e.g., Ishii and Washitani, 2013; Atkinson et al., 2014; Chance et al., 2016), as these data provide a high spatial and spectral resolution (Bradley, 2013). He et al. (2015) conclude that while many studies using species distribution modeling have relied on environmental variables with a relatively coarse resolution such as climatic and topographic variables, more fine-grained remote sensing products will significantly contribute to shaping species distribution modeling in the future. For the management of invasive plant species, detecting the species before it becomes dominant is one of the most important challenges, but few studies assess how well the models perform in detecting smaller stands or low cover fractions of the target species or in detecting species in less conspicuous phenological stages, for example before flowering (but see Mirik et al., 2013; Barbosa et al., 2016).

For mapping invasive plant species, one-class classifiers are efficient tools, as only the distribution of the target species is of interest. One-class approaches greatly reduce field work, as compared to approaches attempting to map all land-cover classes (or species) present in a study area. With this method, only a sample of species presence is required to calibrate the classifier. In addition, an independent dataset containing presence and absence information is needed to validate the predictions. Elith et al. (2006) compared sixteen different presence-only methods over 226 species, and found that the performance highly depends on the chosen method. Among the highest performing models were MARS (multivariate adaptive regression splines) community, boosted regression trees (BRT), generalized dissimilarity, and maximum entropy models

* Corresponding author at: FAU Erlangen-Nürnberg, Wetterkreuz 15, 91058, Erlangen, Germany.
E-mail address: Sandra.Skowronek@fau.de (S. Skowronek).

http://dx.doi.org/10.1016/j.ecoinf.2016.11.005
1574-9541/

(Maxent). Elith and Graham (2009) compared different presence-only and presence-absence methods and concluded that the literature on species distribution modeling "has not yet matured to the point that it provides clear guidance for selecting relevant models", and that all studies should start asking the question of why certain methods perform better than others in order to advance. Another technique that can be used for one-class classification, and that was frequently applied in combination with remote sensing data in recent publications, is biased support vector machine classification (biased SVM). Baldeck and Asner (2014) found that biased SVM shows higher performance than a simple one-class SVM for differentiating savanna trees. For our study, we choose to compare three of the algorithms that have recently been successfully applied with remote sensing data: Maxent, biased SVM and BRT.

1.2. Maxent, biased SVM and BRT

Maxent (Phillips et al., 2004) is a one-class classifier frequently used in ecological modeling. It separates the target species from the background by applying a maximum entropy approach, which compares probability densities. Several model parameters can be customized (regularization multiplier β and feature class), but the user may also work with the default parameters ($\beta = 1$, all feature classes allowed depending on the number of calibration plots, see Phillips et al., 2004; Elith et al., 2011 for detailed explication). While Maxent was traditionally used with predictor variables such as climate, topography and soil composition, recent studies have shown that it can be combined with spectral data: Stenzel et al. (2014) successfully classified European Natura 2000 habitat types; Young et al. (2013) mapped four different invasive plant species in Texas with Maxent; Jones et al. (2015) used Maxent to map an invasive plant pest.

SVMs (Drake et al., 2006) create a hyperplane in a multi-dimensional feature space to separate two classes and maximize the distance of the class samples to that hyperplane. The user has to select a kernel and perform a grid-search to tune the parameter combination (C_{neg}, C_{mult} and sigma) usually using cross-validation (Hsu et al., 2008). While the C parameters relate to the cost function and to weighting the data, sigma relates to the kernel. For the biased SVM, one of the two classes represents our target species and the other one our background data. A different misclassification cost term is used for each class, and the misclassification of the background data is penalized less strongly, see Mack et al. (2014) for details. Barbosa et al. (2016) employed a biased SVM to map an invasive tree on Hawaii in the subcanopy; Mirik et al. (2013) used SVM to detect an invasive herb in flowering and pre-flowering stage; and Atkinson et al. (2014) used it to map an invasive shrub/tree.

BRTs (Elith et al., 2008) combine decision trees with boosting, building an ensemble of multiple tree models. At each step, a new tree that best reduces the loss function is added. The user must set three parameters, the learning rate (lr), the tree complexity (tc) and the bag fraction, which determine the final number of trees (Elith et al., 2008). The model with the lowest deviance is then selected as best model. Originally developed for separating two classes, it is equally suitable to discriminating a target species from a background class. Combining remote sensing data with BRT was successfully applied to predict species cover of marine macrophyte and invertebrate species in the Baltic Sea (Kotta et al., 2013), and to predict bird occurrences (Shirley et al., 2013). Van Ewijk et al. (2014) used BRT to model tree species, as BRT outperformed other algorithms in direct comparison.

1.3. Research questions

We chose to apply these three different state-of-the-art one-class-classification methods, and evaluate their performance in detecting two invasive plant species, *Centaurea solstitialis* and *Phalaris aquatica*, in a biological preserve in California using both field data and a remotely sensed imaging spectroscopy dataset. Our research questions are: (1) How does the performance of the three classifiers differ in detecting our two invasive target species? (2) How successful are our models at detecting small cover fractions of these species?

2. Methods

2.1. Study area

Jasper Ridge Biological Preserve (JRBP) is located near the city of Palo Alto in the U. S State of California, and covers a total area of approximately 485 ha. It has a Mediterranean climate and receives about 638 mm of precipitation per year (average of 1975–2015), and spans 66–207 m elevation above sea level. Our study area (Fig. 1) includes all grassland,

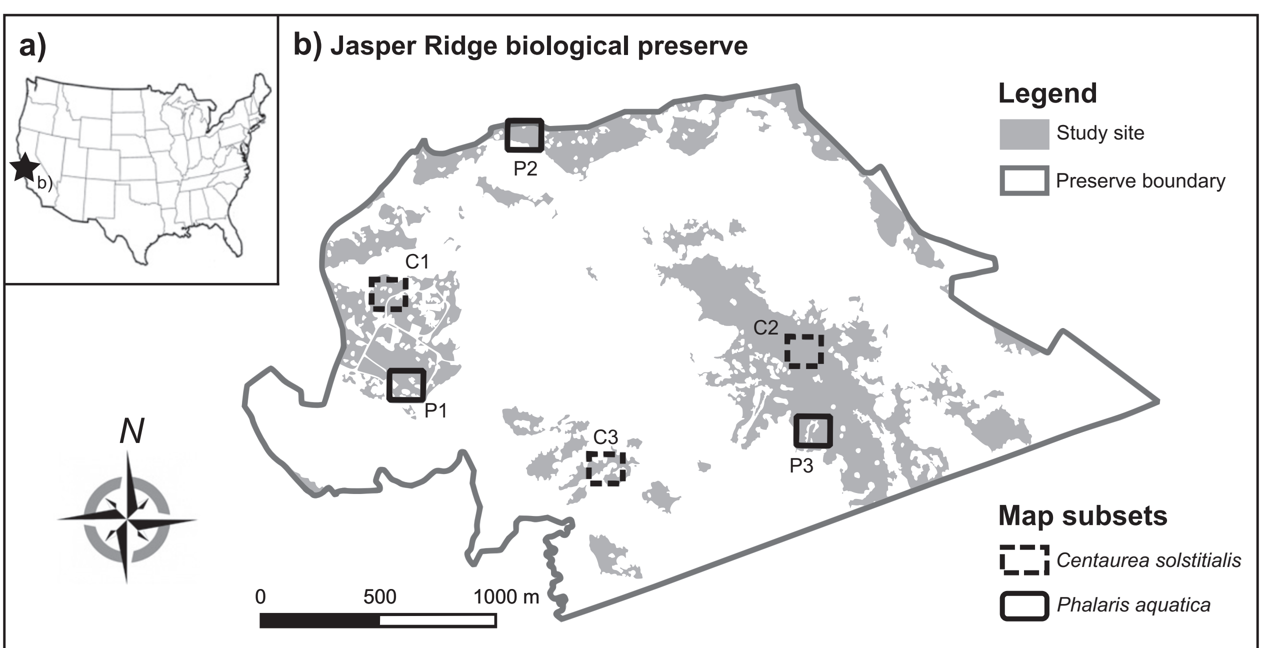

Fig. 1. a) Location of the study site within the United States, b) map of the study area and location of the three subsets for each study species: C1–C3 for *Centaurea solstitialis* and *P1–P3* for *Phalaris aquatica*.