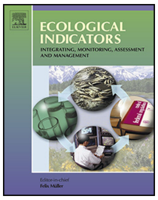
Ecological Indicators 64 (2016) 49–58

Contents lists available at ScienceDirect

Ecological Indicators

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



Mapping tree species diversity of a tropical montane forest by unsupervised clustering of airborne imaging spectroscopy data

Elisa Schäfer*, Janne Heiskanen, Vuokko Heikinheimo, Petri Pellikka

Department of Geosciences and Geography, University of Helsinki, P.O. Box 64, FI-00014 Helsinki, Finland

A R T I C L E   I N F O

Article history:
Received 14 April 2015
Received in revised form 6 November 2015
Accepted 14 December 2015
Available online 7 January 2016

Keywords:
Tree species diversity
Imaging spectroscopy
Clustering
Tropical forest
Object-based image analysis

A B S T R A C T

With the ongoing global biodiversity loss, approaches to measuring and monitoring biodiversity are necessary for effective conservation planning, especially in tropical forests. Remote sensing has much potential for biodiversity mapping, and high spatial resolution imaging spectroscopy (IS) allows for direct prediction of tree species diversity based on spectral reflectance. The objective of this study was to test an approach for mapping tree species alpha diversity that takes advantage of an unsupervised object-based clustering. Tree species diversity of a tropical montane forest in the Taita Hills, Kenya, was mapped based on spectral variation of high spatial resolution IS data.

Airborne IS data and species data from 31 field plots were collected in the study area. Species diversity measures were obtained from the IS data by clustering spectrally similar image segments representing tree crowns. In order to do this, the image was segmented to objects that represented tree crowns. Three measures of species diversity were calculated based on the field data and on the clustering results, and the relationships were statistically analyzed.

According to the results, the approach succeeded well in revealing tree species diversity patterns. Especially, tree species richness was well predicted (RMSE = 3 species; $r^2 = 0.50$) directly based on the clustering results. The optimal number of clusters was found to be close to the estimated number of tree species in the forest. Minimum tree size was an important determinant of the relationships, because only part of the trees are visible to the airborne sensor in the multi-layered closed canopy forest.

In general, the object-based approach proved to be a viable alternative to a pixel-based clustering. The approach takes advantage of the capability of IS to detect spectral differences among tree crowns, but without the need for spectral training data, which is expensive to collect. With further development, the approach could be applied also for estimating beta diversity.

1. Introduction

While global biodiversity is rapidly declining (Butchart et al., 2010), efficient methods are needed for biodiversity assessment at all scales to provide adequate information for biodiversity conservation and management. This is especially true in the tropics, where most biodiversity is found (Gaston, 2000) and where species diversity and its significance to the ecosystems are poorly understood (Milliken et al., 2010). Aspects of biodiversity include diversity within communities (alpha diversity) and the diversity between communities of an area or along an environmental gradient (beta diversity) (Whittaker et al., 2001).

The potential of remote sensing for biodiversity assessment and monitoring has been recognized for a long time (Kuenzer et al., 2014; Nagendra, 2001; Nagendra et al., 2013; Pettorelli et al., 2014; Rocchini et al., 2010; Turner et al., 2003). Nagendra (2001) categorized remote sensing studies on species distribution patterns to (1) direct mapping of individual plants and species, (2) habitat mapping, and (3) modelling the relationship between species distribution patterns and remotely sensed data. Medium resolution multispectral data, such as Landsat imagery, provide means for making habitat maps (Nagendra et al., 2013) and extracting measures of spectral entropy for predicting species diversity (Hernández-Stefanoni et al., 2011; Maeda et al., 2014; Rocchini et al., 2010).

Plant-level assessments, which typically refer to tree species identification, require higher spatial and spectral resolution than offered by medium resolution satellite data (Gillespie et al., 2008; Turner et al., 2003). The high spectral resolution of imaging

---

* Corresponding author at: Department of Geosciences and Geography, University of Helsinki, P.O. Box 64, FI-00014 Helsinki, Finland. Tel.: +358 468110532.
E-mail address: elisaschafer@hotmail.com (E. Schäfer).

http://dx.doi.org/10.1016/j.ecolind.2015.12.026

spectrometers enables detecting subtle differences in reflectance among species, caused by their biochemical and biophysical characteristics. This has led to advancements in tree species identification from airborne imaging spectroscopy data (Asner et al., 2008; Clark et al., 2005; Clark and Roberts, 2012; Cochrane, 2000; Vaglio Laurin et al., 2014). Tree species identification based on their spectral properties requires sufficient spectral dissimilarity among species. Studies in tropical forests in Hawaii (Asner and Martin, 2009), Australia (Asner et al., 2009), and Amazonia (Asner and Martin, 2011) have shown that tree species often have unique spectral signatures, linked to their structural and biochemical properties.

Mapping single tree species based on supervised classification requires spectral training and validation data on each target species (e.g. Baldeck and Asner, 2013). In tropical forests with vast amount of tree species, many of them rare, it is practically impossible to obtain training and validation data for each species. Therefore, unsupervised classification, or clustering, is an appealing alternative for mapping species diversity as it does not require any prior information of the species. This approach has some similarity to the spectral variation hypothesis (SVH) (Palmer et al., 2002) as the clustering result can be interpreted as a measure of spectral heterogeneity (Féret and Asner, 2014; Medina et al., 2013). However, when using high spatial resolution remote sensing data, it is assumed that spectral variation arises from variation in canopy-level tree species composition, not from heterogeneity of habitats as suggested by the SVH (Féret and Asner, 2014). This also means that retrieved information is limited to those species which are visible to the sensor.

So far, unsupervised tree species classification has been rarely applied for diversity studies. Baldeck and Asner (2013) tested unsupervised *k*-means clustering of imaging spectroscopy data in the visible and near-infrared (NIR) spectral range for estimation of beta diversity in an African savanna. According to their results, the beta diversity estimates based on unsupervised clustering were similar to those obtained from supervised classification. Féret and Asner (2014) tested a similar approach using imaging spectroscopy data in the visible to shortwave infrared range to map alpha and beta diversity in the Peruvian Amazon. After a principal component analysis (PCA), they applied a *k*-means clustering to selected principal components. As a result, the Shannon–Wiener index values based on the clustering results were systematically underestimated, but correlations with field measurements were high ($r = 0.86$). Furthermore, Medina et al. (2013) studied alpha diversity of a dry tropical forest in Puerto Rico using imaging spectroscopy data in the visible and NIR spectral range. The correlations between the Shannon–Wiener index obtained from clustering and field data were variable, even negative in some cases, but were considerably improved when spectral unmixing was applied after the clustering.

The aforementioned studies indicate that unsupervised approaches have potential for direct prediction of alpha diversity from high resolution imaging spectroscopy data. However, the previous studies have poorly addressed the two components of alpha diversity, namely richness and evenness. The richness refers simply to the total number of species at a location but evenness accounts for relative species abundance. In the latter case, a community with a relatively equal species abundance is considered more diverse than a community with a few dominating species but the same total number of species. Both Medina et al. (2013) and Féret and Asner (2014) studied only the evenness-based Shannon–Wiener index, which has sometimes been reported to be better predicted by remote sensing data than species richness as rare species get less weighted (Oldeland et al., 2010). However, a single diversity measure is not necessarily appropriate for characterizing alpha diversity (Rocchini et al., 2015).

Scale remains a challenge in the remote sensing studies of tree species diversity, as results are dependent on the pixel resolution (Rocchini et al., 2015). However, the spatial resolution of a few metres or less allows detecting individual tree crowns from the image, and thus treating tree crowns as objects. Object level classification has the advantage that reflectance characteristics are averaged over entire tree crowns, which decreases within-species spectral variation (Blaschke et al., 2014). Delineating individual tree crowns to objects from remotely sensed imagery has been found useful in species discrimination (Féret and Asner, 2012; Lucas et al., 2008), and has therefore potential also for unsupervised classification.

The objective of this study was to test an approach for tree species alpha diversity mapping that takes advantage of an unsupervised object-based clustering. Tree species diversity was predicted for a tropical montane forest in the Taita Hills, Kenya, based on spectral variation of high spatial resolution imaging spectroscopy data. The more specific research questions were:

(1) How accurately can we estimate tree species diversity measures based on spectral differences among tree crowns?
(2) How does the relationship between clustering results and field data depend on the size of the trees considered?
(3) How does tree species richness vary spatially in the Ngangao forest in the Taita Hills, Kenya?

2. Materials and methods

2.1. Study area

The study area was the Ngangao forest fragment in the Taita Hills, in the Taita-Taveta district of Southern Kenya (Fig. 1). It consists of a hilltop covered with moist montane forest at an altitude of 1700–1952 m. The climate is tropical, characterized by a shorter rainy season in November–December, and a longer rainy season in March–May. The forested hilltops of the Taita Hills trap moisture-laden clouds coming from coastal areas, and therefore the forests remain relatively humid throughout the year (Pellikka et al., 2009).

The forests in the Taita Hills are remnants of a larger forest cover, now covering only the highest hilltops (Pellikka et al., 2013). The Taita Hills belong to a globally important biodiversity hotspot together with other mountains of the Eastern Arc (Myers et al., 2000). The hotspots are characterized by a high degree of endemic species and a high threat of extinction, and together contain a large portion of the world's biodiversity while covering only a small area.

The Ngangao forest fragment covers 120 ha, including 18 ha of plantations of exotic pine (*Pinus patula*) and cypress (*Cupressus lusitanica*). The plantations were established in the 1970s mostly on cleared land, so that despite forest loss to agricultural expansion, the total forest area has remained about the same since 1955 (Pellikka et al., 2009). Long before this, the forests in the Taita Hills have been under human influence because of the long history of settlement (Pellikka et al., 2013).

The high proportion of secondary successional species in the forest indicates that the community composition has undergone disturbance (Aerts et al., 2011; Omoro et al., 2010). In a recent survey, 73 woody species were recorded for Ngangao and another forest fragment together (Mbuthia, 2003, cited by Aerts et al., 2011). The total number of tree species in Ngangao is probably somewhat smaller, as the survey included also palms.

2.2. Field data

Field data were collected in January and February of 2013 and 2014. 31 field plots were established, aiming for a spatially representative sampling but ensuring that the pine and cypress plantations were also sampled. The plot centres were positioned