Ecological Modelling 272 (2014) 129–140

Contents lists available at ScienceDirect

Ecological Modelling

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

Comparing the suitability of classified land cover data and remote sensing variables for modeling distribution patterns of plants

Anna F. Cord[a,*], Doris Klein[b], Franz Mora[c], Stefan Dech[b,d]

[a] Helmholtz Centre for Environment Research – UFZ, Department of Computational Landscape Ecology, Permoserstraße 15, 04318 Leipzig, Germany
[b] German Aerospace Center (DLR), German Remote Sensing Data Center (DFD), Oberpfaffenhofen, Münchner Straße 20, 82234 Wessling, Germany
[c] National Commission for the Knowledge and Use of Biodiversity (CONABIO), Avenida Liga Periférico-Insurgentes Sur 4903, Col. Parques del Pedregal, Del. Tlapan, 14010 Mexico-City, Mexico
[d] University of Würzburg, Institute of Geography and Geology, Department of Remote Sensing, Oswald-Külpe-Weg 86, 97074 Würzburg, Germany

ARTICLE INFO

Article history:
Received 12 April 2013
Received in revised form
20 September 2013
Accepted 22 September 2013

Keywords:
Land cover classification
Mexico
MODIS
Species distribution model
Time series
Vegetation phenology

ABSTRACT

Given the rapid loss of biodiversity worldwide and the resulting impacts on ecosystem functions and services, we more than ever rely on current and spatially continuous assessments of species distributions for biodiversity conservation and sustainable land management. Over the last decade, the usefulness of categorical land cover data to account for the human-induced degradation, transformation and loss of natural habitat in species distribution models (SDMs) has been questioned and the number of studies directly analyzing remotely sensed variables has lately multiplied. While several assumptions support the advantages of remote sensing data, an empirical comparison is still lacking. The objective of this study was to bridge this gap and compare the suitability of an existing categorical land cover classification and of continuous remote sensing variables for modeling the distribution patterns of 30 Mexican tree species. We applied the *Maximum Entropy* algorithm to predict species distributions based on both data types independently, quantified model performance and analyzed species–land cover relationships in detail. As part of this comparison, we focused on two particular aspects, namely the effects of (1) thematic detail and (2) spatial resolution of the land cover data on model performance. Our analysis revealed that remote sensing data were significantly better model predictors and that the main obstacle of the land cover-based SDMs were their bolder predictions, together with their overall overestimation of suitability. Among the land cover-based models, we found that thematic detail was more important than spatial resolution for SDM performance. However, our results also suggest that the suitability of land cover data differs largely among species and is dependent on their habitat distinctiveness. Our findings have relevant implications for future species distribution modeling studies which aim at complementing their set of topo-climatic predictors by data on land surface characteristics.

1. Introduction

Biodiversity is rapidly declining worldwide (Butchart et al., 2010), and there is considerable evidence that biodiversity loss strongly affects ecosystem functioning and stability as well as the provision of ecosystem services (Millennium Ecosystem Assessment, 2005). Spatial decision support systems for biodiversity conservation and land management hence more than ever rely on current and spatially continuous assessments of species distribution patterns. These are affected by a variety of environmental factors at varying scales (Luoto et al., 2007; Pearson and Dawson, 2003; Pearson et al., 2004). While climatic gradients govern species distributions across large biogeographical scales, land use/land cover (LULC) affects species occupancy patterns at the landscape scale. Soil type and other more specific habitat conditions are particularly relevant at the local scale (Pearson and Dawson, 2003).

An effective way to maximize the information content of species locality data is to apply species distribution models (SDMs) which predict the distribution of a species as a function of selected environmental variables (Guisan and Zimmermann, 2000). Several studies have used SDMs to examine the potential impacts of changing climate on species ranges over the last decade (Araújo et al., 2011; Guisan and Theurillat, 2000; Jeschke and Strayer, 2008; Pearson et al., 2002; Sinclair et al., 2010; Yates et al., 2010). However, given the accelerating anthropogenic LULC changes which lead to habitat loss, degradation and transformation (Millennium Ecosystem Assessment, 2005), these climate-based SDMs are expected to provide increasingly inefficient predictions of actual species distributions (Araújo et al., 2011; Thuiller et al., 2004). The

* Corresponding author. Tel.: +49 341 235 1070; fax: +49 341 235 1939.
E-mail address: anna.cord@ufz.de (A.F. Cord).

0304-3800/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ecolmodel.2013.09.011

incorporation of land cover data in SDMs, which enables regions with suitable climate but otherwise unsuitable environmental conditions to be identified, however, improves climate-driven model predictions in particular at landscape to national scales (Pearson et al., 2004; Sánchez-Cordero et al., 2005). Besides species distribution modeling, a number of other disciplines refer to classified land cover (change) data as the most obvious indicator of land surface characteristics and human impact (Herold, 2009). A growing number of global, continental, and national mapping activities have been recently initiated, including efforts with very high spatial resolution (Gong et al., 2013). Many of the land cover products derived from remote sensing data are freely distributed in 'ready-to-use' formats including metadata information, e.g. IGBP DISCover, GLC2000 or GLOBCOVER (for a more detailed description see Herold et al., 2008). The discrete representation of land surface characteristics in these products has the advantages of conciseness and clarity for many applications and requires low data volumes (Lambin, 1999). Land cover data is hence still the most commonly used categorical predictor in species distribution modeling (Franklin, 2009).

Even though many studies have reported statistically significant relationships between land cover data and the regional or nationwide distribution of species (Eyre et al., 2004; Luoto et al., 2006; Heikkinen et al., 2004; Maes et al., 2003; Siriwardena et al., 2000), the causal relationship between classified land cover and species distributions is often indirect. This is in particular the case for plants, for which land cover is rather a generally limiting factor without having direct physiological impact (Thuiller et al., 2004). Bearing in mind that classified land cover data cannot capture inherent environmental variation (Goodchild et al., 1992), the use of continuous remotely sensed predictors in SDMs has considerably increased in recent years (Buermann et al., 2008; Cord and Rödder, 2011; Prates-Clark et al., 2008; Saatchi et al., 2008; Tuanmu et al., 2010). Three main issues highlight the benefits of such continuous remote sensing data in comparison to categorical land cover data: (1) Land cover products are always designed for the purpose of a specific study or mapping campaign and adapted to the corresponding legend specifications. They may therefore not be representative and thematically detailed enough for the focal species (Bradley and Fleishman, 2008). (2) While conventional land cover classifications are based on discrete arbitrary classes which do not capture gradual changes in the landscape (Krishnaswamy et al., 2009), the use of remote sensing data preserves continuous geographical variation over the study area (Goodchild et al., 1992). (3) No additional error, as it is inherent in any land cover classification procedure, is introduced when remote sensing data are directly used as model predictors. While the usefulness of both data sources has been discussed (Bradley and Fleishman, 2008) and the three above-mentioned reasons support the advantages of remote sensing data, an empirical comparison is still lacking.

In this study, our objective was to bridge this gap and compare the suitability of an existing categorical land cover classification and of continuous remote sensing variables for modeling the distribution patterns of 30 Mexican tree species based on presence/absence data from the Mexican National Forest Inventory. Both land cover data and remote sensing variables used here reflect the current state of research and represent the typical type of independent variables that researchers can choose from in order to use them in SDMs. As part of this comparison, we focused on two particular aspects, namely the effects of (1) thematic detail and (2) spatial resolution of the land cover data on model performance. We implemented SDMs using the *Maximum Entropy* algorithm (Maxent; Phillips et al., 2006) which has been previously applied in combination with both land cover data (Kuemmerle et al., 2010; Wilting et al., 2010) and remote sensing data (Buermann et al., 2008; Cord and Rödder, 2011; Saatchi et al., 2008; Tuanmu et al., 2010). To disentangle the suitability of both data sets while taking into account species characteristics, we estimated model performance by the threshold-independent area under curve (AUC) and the threshold-dependent percent correct classification (PCC) and related it to the characteristics of our focal species. To our knowledge, this is the first direct comparison of categorical land cover data and continuous remote sensing variables in SDMs for a shared set of focal species.

2. Materials and methods

2.1. Study area and species records

Mexico covers a latitudinal gradient from 32° N to 14° N which promotes a great diversity of climatic conditions (Cavazos and Hastenrath, 1990). Beyond topo-climatic variability, Mexico's complex geological history has been one of the major evolutionary forces (Miranda and Hernández, 1963; Ramamoorthy et al., 1993) and is one of the main reasons for its remarkable phytodiversity (Cevallos-Ferriz and González-Torres, 2005). Mexico holds 10–12% of the world's total flowering plant species (Toledo and Ordóñez, 1993) and has at least 2000 and possibly over 4000 native tree species (Ricker et al., 2007). The major forest types in the country are (Ricker et al., 2007): dry forest (10.9%), oak-pine forest (7.0%), oak forest (5.1%), tropical rain forest (5.1%), coniferous forest (3.9%), moist montane forest (0.9%), mangrove (0.4%), palm forest (0.06%), and gallery forest (<0.01%). As a result of land consumption for agriculture as well as industrial and infrastructural purposes, Mexico has experienced extensive LULC changes during the last decades (Ricker et al., 2007; Sarukhán et al., 2010).

We selected 30 focal tree species which are typical of various Mexican vegetation types and have different prevalence, i.e. different proportions of sampled sites where the species was present (Table 1). Six of the study species are listed as subject to special protection in the NORMA Oficial Mexicana NOM-059 (SEMARNAT, 2010). For the focal species, we assembled presence-absence records from a total of 19,319 1 ha survey sites of the Mexican National Forest Inventory (INF; CONAFOR, 2009). Species locality information in this data set was georeferenced to the center of each survey site. The forest inventory data was collected between 2004 and 2007 and is the latest countrywide estimate of tree species distributions in Mexico. Further, the inventory is temporarily consistent with the acquisition dates of both the remote sensing variables and the land cover data, which is a crucial assumption for reliable species distribution modeling (Phillips et al., 2006). The INF sampling scheme covers natural vegetation complemented by forest plantations whereas distances between survey sites range from 5 km (for forests), through 10 km (dry forests, mangroves, wetlands) to 20 km (matorral).

2.2. Remote sensing data

The MODIS sensor is the most widely used source of remote sensing data for modeling species distributions at regional to continental scales (e.g. in Buermann et al., 2008; Saatchi et al., 2008; Tuanmu et al., 2010). We selected the Terra-MODIS *Enhanced Vegetation Index* (EVI, MOD13A2), *Surface Reflectance* (SR, MOD13A2) – which includes blue, red, near-infrared (NIR), and middle-infrared (MIR) wavelengths – and *Land Surface Temperature* (LST, MOD11A2) products. These have been previously used to model plant species distributions (Buermann et al., 2008; Saatchi et al., 2008; Tuanmu et al., 2010) and are well-suited to characterize ecosystem functioning (Pettorelli et al., 2005; Quattrochi and Luvall, 1999). The specific products used here provide different levels of information: Reflectance data quantify surface reflectance at ground level in the absence of atmospheric scattering or absorption. They indicate, among others, the proportions of broadleaf and needleleaf plants in the canopy (Hall et al., 1992) and the overall density of the