Biological Conservation 153 (2012) 257–264

Contents lists available at SciVerse ScienceDirect

Biological Conservation

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

Using species distribution modeling to improve conservation and land use planning of Yunnan, China

Ming-Gang Zhang [a,b], Zhe-Kun Zhou [a,c], Wen-Yun Chen [c], J.W. Ferry Slik [a,*], Charles H. Cannon [a,d], Niels Raes [e]

[a] *Plant Geography Lab, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Menglun, Yunnan 666303, China*
[b] *Graduate University of Chinese Academy of Sciences, 19 Yuquan Road, Beijing, China*
[c] *Key Laboratory of Biogeography and Biodiversity, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, Yunnan 650204, China*
[d] *Department of Biological Sciences, Texas Tech University, Lubbock, TX 79409, USA*
[e] *National Herbarium of the Netherlands, Leiden University Branch, Einsteinweg 2, NL-2300 RA Leiden, The Netherlands*

ARTICLE INFO

Article history:
Received 27 January 2012
Received in revised form 21 April 2012
Accepted 21 April 2012
Available online 29 June 2012

Keywords:
Yunnan
Species distribution model
Species diversity
Floristic regions
Land use
Systematic conservation planning

ABSTRACT

Part of the Himalayan biodiversity hotspot, Yunnan province in China is a highly diverse terrestrial region, particularly in the wide range of natural forest types. These forests are under considerable conversion pressure as land use intensifies with expanding human population and economic development. Conservation strategies based on the geographic patterns of botanical species richness, including the identification of meaningful floristic regions and priority areas for conservation, could improve the effectiveness of forest policy and management. These strategies should also include current threats of loss due to forest conversion to address the more urgent challenges for sustainable development. Here, we produce distribution models at ~10 km^2 resolution for 2319 plant species, using geo-referenced herbarium collections, corrected for spatial bias using a null model, and detailed environmental variables. Based on 1996 species with significant non-random habitat preferences, we identify four important aspects of plant species distribution in Yunnan: (1) species diversity hotspots; (2) seven major floristic regions, using a cluster analysis of species presence/absence; (3) priority areas for conservation based on the concept of the 'irreplaceability' value of planning units and (4) the percentage remaining natural forest among the species rich and conservation priority areas, to assess the level of endangerment. Our maps provide clear priorities for the development of a sustainable and feasible biodiversity conservation strategy for Yunnan.

1. Introduction

Yunnan province, SW China (Fig. 1), is one of the botanically most diverse terrestrial regions on Earth and part of the Himalaya biodiversity hotspot (Myers et al., 2000). This region is located at a transitional zone, characterized by strong environmental gradients, between tropical, sub-tropical, temperate and alpine vegetation types, with the flora of tropical Indochina, mixing with the subtropical East Asian flora, and between major floristic regions, with the Sino-Japanese floristic region in the east and the Sino-Himalayan floristic region in the west (Li and Li, 1997; Zhu et al., 2006). The region also possesses a rich diversity of forest types, including tropical rain forest, seasonal rain forest, broadleaved evergreen forest, broadleaved deciduous forest, needle leaved forest, alpine and sub-alpine meadow, bush and mixed forest (Wu, 1987). Some parts of Yunnan have been identified as refugia during the Pleistocene (López-Pujol et al., 2011). The region has a disproportionate amount of China's overall floristic diversity (51.6%), with over 18,000 plant species (Yang et al., 2004), which includes high levels of endemism.

Yunnan's biodiversity is under considerable pressure due to the intensification of land use and an expanding human population (Li et al., 2011; Willson, 2006; Yang et al., 2004; Zhou and Grumbine, 2011). Forests have become increasingly fragmented through agriculture, logging, the planting of economic plants, mining activities and changing environment (Li et al., 2006; Xu et al., 2005). With the combined impact of anthropogenic and climate change, describing and understanding species compositional patterns and identifying areas of high species richness and priority areas for conservation has become critical for the development of sound conservation policies and their integration into a sustainable land development strategy for Yunnan.

Species distribution modeling is a widely used method to determine species diversity and compositional patterns at large spatial scales (Araújo et al., 2011; Guisan and Zimmermann, 2000;

---

[*] Corresponding author.
E-mail address: FerrySlik@hotmail.com (J.W.F. Slik).

0006-3207/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.biocon.2012.04.023

Pearson et al., 2006) and it produces spatially explicit and comprehensive maps that are particularly useful for identifying areas where conservation efforts are most needed or effective. The distribution models are based on data collected during botanical surveys and primarily exist as simple geo-referenced presence–absence records in herbarium collection databases, with little reliable abundance measurements. Fortunately, this particular limitation is a common feature of plant species distribution data and several modeling applications were explicitly designed to exploit presence only data (Elith et al., 2011; Pearce and Boyce, 2006; Pearson et al., 2006; Tsoar et al., 2007). These approaches can now be used effectively for a wide number of applications, greatly enhancing the power and usefulness of the large existing databases of herbarium collections.

Generally, nature reserves and protected areas are selected primarily because of their inaccessibility or unsuitable nature for other purposes, such as intensive agriculture or urban development and not to meet specific conservation objectives. This approach for designing conservation strategies exposes many species found in productive or potentially productive landscapes to conversion of habitat (Margules and Pressey, 2000). Given our current ability to model species distributions based on large databases and robust algorithms, more scientific and rigorous conservation strategies can be designed, utilizing objective predictions of species composition and diversity across the entire landscape, even in areas with little history of specimen collection. Here, we apply the theory of systematic conservation planning (Margules and Pressey, 2000) and decision making tools (Possingham et al., 2000) to identify priority areas for conservation (Carvalho et al., 2011, 2010; Wilson et al., 2005), based upon plant species distribution models. We address the following major objectives: (1) model the distribution of the woody plant flora of Yunnan to map diversity and floristic patterns across the entire province; (2) combine the current forest cover and land use maps with our species distribution models to estimate how much of Yunnan's flora has already disappeared and which of the remaining high diversity areas are most vulnerable, especially in relation to the current protected forest network in Yunnan; and (3) use these results to design a spatially explicit conservation network that would protect as many of Yunnan's woody plant species as possible in the most spatially efficient way.

2. Methods

2.1. Species data

The basic herbarium collection data for all woody species, except Fagaceae, for Yunnan province was provided by the Kunming Institute of Botany, Chinese Academy of Sciences, totaling 85,289 records. Most of these collections have been examined for the flora of China, and although some identification errors are likely, overall the identifications can be considered accurate. Although many of these records did not have latitude and longitude data, we were able to geo-reference 60,552 collections within Yunnan, using the location descriptions in the label information. These resolution of the label locations are roughly equivalent to the Chinese village level. Subsequently, species presences were scored in 5 arc min grid cells (ca. 10 × 10 km), avoiding duplicate species records in each grid cell. We used the 5 arc min spatial resolution because this corresponded to the environmental data resolution (WorldClim and FAO soil properties) but also because a higher resolution was not possible due to the spatial error in the geo-referenced specimen data. Species that were present in fewer than five grid cells were removed from the analysis because the statistical power of the species distribution would be too weak. Of the 60,552 geo-referenced specimens, 42,114 records, belonging to 118 plant families representing 2319 species, remained for species distribution modeling.

2.2. Environmental predictors

We initially selected 35 environmental predictors to model the species distributions. These included 19 bioclimatic predictors (1950–2000) plus elevation of the WORLDCLIM dataset (<www.worldclim.org>) and 15 soil variables selected from the FAO database for poverty and insecurity mapping (FAO, 2002) for Yunnan at 5 arc min resolution, resulting in 4936 grid cells covering the entire Yunnan province. A serious problem in species distribution modeling is formed by multi-colinearity of variables which can result in model over-fitting (Graham, 2003; Pearson et al., 2006). To avoid this problem we removed highly correlated environmental predictors. For both bio-climate and soil predictors, we used spearman's rank correlation to select the least correlated variables (spearman's <0.75). Of variables that showed correlations higher than 0.75 only the ecologically most meaningful factors were kept (Tables S1 and S2). For the bio-climate predictors the following variables were included in the analyses: (1) BIO1: Annual Mean Temperature; (2) BIO2: Mean Diurnal Temperature Range; (3) BIO4: Temperature Seasonality; (4) BIO7: Temperature Annual Range; (5) BIO12: Annual Precipitation; (6) BIO14; Precipitation of Driest Month; (7) BIO15: Precipitation Seasonality. Of the soil predictors the following variables were included in the analysis: (1) CE-T: CEC clay topsoil (CEC = cation exchange capacity); (2) CN-T: C:N ratio class topsoil; (3) CP-T: organic carbon pool topsoil; (4) DEPTH: effective soil depth; (5) DRAIN: soil drainage class; (6) NN-T: nitrogen% topsoil; (7) PH-T: pH top soil; (8) SOIL-PROD: soil production index; (9) textural class subsoil. In total 16 of the 35 predictors were kept as environmental layers for the species distribution models.

2.3. Species distribution model building and collection bias correction

To model species distributions we used the modeling application Maxent (ver. 3.3.1; <www.cs.princeton.edu/~schapire/maxent/>) (Phillips et al., 2006). Maxent was specifically developed to model species distributions with presence-only data. Of available species distribution modeling algorithms, Maxent has been shown to perform best, especially when few presence records are available, while it is also the least affected by location errors in occurrences (Graham et al., 2007). Maxent was run with the following modeling rules: (1) for species with 5–10 collection records linear features were applied, (2) for species with 10–14 records quadratic features were applied, while (3) for species with >15 records hinge features were applied (Raes and ter Steege, 2007). These rules were included in Maxent to counteract the tendency of species distribution models (SDMs) to over-fit, especially when few presence records are available. For each of the 2319 species a SDM was developed based on its unique presence records and the 16 environmental predictors.

As a measure of the accuracy of the SDMs, we used the threshold independent and prevalence insensitive area under the curve (AUC) of the receiver operating characteristic (ROC) plot produced by Maxent. All measures of SDM accuracy require absences, when these are lacking, as is the case here, they are replaced by pseudo-absences or sites randomly selected at localities where no species presence was recorded (Phillips et al., 2006). However, when SDM accuracy measures are based on presence-only data and pseudo-absences, the standard measures of accuracy (e.g. the often used measure AUC > 0.7) do not apply (Raes et al., 2009; Raes and ter Steege, 2007). Therefore, we applied the bias corrected null-model developed by Raes and ter Steege (2007) to test the AUC value of an SDM developed with all presence records against the AUC values