Applied Vegetation Science **14** (2011) 388–401

Using climatically based random forests to downscale coarse-grained potential natural vegetation maps in tropical Mexico

Raúl A. Vaca, Duncan J. Golicher & Luis Cayuela

Keywords
Downscaling; Machine learning algorithms; Random forests; Rzedowski's vegetation map; Spatial resolution.

Nomenclature
The International Plant Names Index (2008).

Received 2 November 2009
Accepted 9 March 2011
Co-ordinating Editor: Geoffrey Henebry

Vaca, R.A. (corresponding author, rgenuit@ecosur.mx): Departamento de Ecología y Sistemática Terrestre, División de la Conservación de la Biodiversidad, El Colegio de la Frontera Sur, Carretera Panamericana y Periférico Sur s/n, 29290 San Cristóbal de Las Casas, Chiapas, México
Golicher, D.J. (dgolicher@bournemouth.ac.uk): Departamento de Ecología y Sistemática Terrestre, División de la Conservación de la Biodiversidad, El Colegio de la Frontera Sur, Carretera Panamericana y Periférico Sur s/n, 29290 San Cristóbal de Las Casas, Chiapas, México Centre for Conservation Ecology & Environmental Change, School of Conservation Sciences, Dorset House DG 38B, Bournemouth University, Fern Barrow, Poole (Dorset) BH12 5BB UK
Cayuela, L. (lcayuela@ugr.es): Área de Biodiversidad y Conservación, Universidad Rey Juan Carlos, Departamental 1 DI. 231, c/Tulipán s/n, E-28933 Móstoles (Madrid), España

Abstract

Questions: Can the accuracy of coarse resolution potential vegetation maps be improved through downscaling to finer resolution climatic grids? Can output from random forests produce new insight into the climatic characteristics that are associated with the structural characteristics of the vegetation?

Location: Southern Mexico.

Methods: A potential vegetation map (National Atlas of Mexico) at a 1:4 000 000 scale, was downscaled to a 1 km^2 grid resolution using climatically based random forests models. The resulting map was then evaluated against 256 inventory plots sampled at the transition between different vegetation types in Southern Mexico.

Results: Downscaling increased accuracy up to 0.40, as measured by the Kappa Index of Agreement. Multivariate analysis of the results allowed the association between Rzedowski's classification and climatic variation to be explored. This confirmed that many of the structural aspects of the vegetation that are used by the Rzedowski classification are closely associated with climate, but it also revealed weaknesses in the underlying basis of this classification system.

Conclusions: Rzedowski's scheme for vegetation classification may require further modification in order to be an effective tool for research into vegetation–climate relationships.

Introduction

The potential natural vegetation is the theoretical climax vegetation community that would occupy an area if left undisturbed by humans (Moravec 1998; Zerbe 1998). Potential vegetation maps are widely used in research, conservation and management. Some applications of potential vegetation maps include analysing vegetation dynamics (Hessburg et al. 1999; Hickler et al. 2004), quantifying spatial patterns of deforestation (Trejo & Dirzo 2000), planning land management (Hughes et al. 1986; Felix et al. 2004), selecting sites and species for restoration (Miyawaki 2004), analysing the effect of climate change on vegetation (Yates et al. 2000; Koca et al. 2006; Franke & Köstner 2007), and predicting spatial patterns of species diversity (Golicher et al. 2008).

A typical method for constructing a potential vegetation map involves identifying remnants of vegetation with natural or near-natural character (Zerbe 1998). Natural vegetation includes vegetation that is assumed to approximate to the theoretical climax community. Near-natural vegetation has been influenced by levels of human disturbance of similar frequency and intensity to natural disturbance. It is therefore thought to have a

Applied Vegetation Science
Doi: 10.1111/j.1654-109X.2011.01132.x

similar species composition to the theoretical climax community (White 1979). The vegetation found in remnants of natural or near-natural vegetation may be assumed to potentially extend to a wider geographical area with similar environmental conditions (Cross 1998; Moravec 1998; Zerbe 1998). Potential vegetation maps are therefore typically coarse scale in nature because the delineation of areas with similar environmental conditions has typically involved making simplifying assumptions that ignore some of the environmental heterogeneity (Lüttge 2008). Potential vegetation maps produced using traditional photo-interpretative and cartographic methods may be designed primarily for communication purposes rather than as analytical tools (DeMers 1991; Franklin 1995), although the process of classification may produce insight into the climatic characteristics associated with the potential vegetation (DeMers 1991; Franklin 1995).

At present, one widely recognized source of information on the potential distribution of vegetation types of Southern Mexico is the system proposed by Rzedowski (1978) (Olson et al. 2001; González-Medrano 2003). This has been spatially represented at a scale 1:4 000 000 (Rzedowski 1990), and digitized by the Comisión Nacional para el Conocimiento y el Uso de la Biodiversidad (CONABIO). Rzedowski's terminology forms a fundamental framework for describing the regional vegetation that is widely used and understood within Mexico (Challenger & Soberón 2008). This map, however, cannot be fully relied on for various reasons described by the original author (Rzedowski 1978): (1) it is a coarse representation of the potential vegetation distribution; (2) this coarse resolution has prevented the representation of dispersed fragments of the potential vegetation; and (3) the definition of boundaries is inaccurate in some areas, owing to the scarcity of information on natural and near-natural vegetation that was available to the author.

There is a general concern that the resolution of traditional maps of potential vegetation such as that based on Rzedowski's classification are too coarse for real-world applications (Stoms 1992; Bredenkamp et al. 1998; Hopkinson et al. 2000; Kunin et al. 2000; Hulme 2003; Rouget 2003; Hartley et al. 2004). Map inaccuracies can result from the scarcity of information on natural and near-natural vegetation (Zerbe 1998), and from the coarse resolution in the available maps of predictor variables (van Etten 1998). Coarse-scale maps may overlook variability in mountainous and other areas in which fine-scaled climatic gradients determine the observed vegetation type (Franklin 1995). In order to use vegetation maps effectively, their spatial resolution must be adjusted to the needs of pure and applied biological and ecological research (Araújo et al. 2005; McPherson et al. 2006). For many applications, data at a spatial resolution of, at least, 1 km × 1 km are necessary to capture environmental variability that is lost at coarse resolutions (Hijmans et al. 2005). These data could be useful for regional/national purposes (Rogan & Chen 2004).

To date, fine grained spatial representations of potential vegetation types have not been available for most of Mexico. Examples of studies using Rzedowski's potential vegetation map include a description of changes in seasonally dry tropical forests in Mexico (Trejo & Dirzo 2000), analyses of the vulnerability of Mexican vegetation types to climate change (Villers-Ruiz & Trejo-Vázquez 1997, 2003; Benítez-Badillo et al. 2003), the identification of areas for conservation in tropical deciduous forests (Cué-Bär et al. 2006), and studies modeling ecological niches of Mexican bird species to reconstruct population losses (Peterson et al. 2006). Although there has been progress in downscaling naturally coarse-grained variables, such as temperature or precipitation (Golicher et al. 2006; Kostopoulou et al. 2006; Vrac et al. 2007), the potential of contemporary statistical methods to downscale vegetation maps has been, to date, insufficiently examined.

Statistical downscaling is expected to be an effective method for improving both the spatial resolution and accuracy of potential vegetation maps even in a challenging area such as Southern Mexico. Such approaches could then increase the utility of potential vegetation maps. For example, Villers-Ruiz & Trejo-Vázquez (1997) established a climate–vegetation correspondence between the Köppen climate classification, as modified by García (1988), and the potential vegetation map for Mexico by Rzedowski using a geographic information system. Benítez-Badillo et al. (2003) used regression tree classification analysis and the Rzedowski's potential vegetation map, to establish a correspondence between climatic and topographic variables, with the distribution of the vegetation types for the state of Veracruz. Both approaches were developed in order to identify the climatic conditions that support each type of vegetation and to use this information as a baseline for the projection of alternative scenarios based on different climate change models.

In this study we aimed to downscale a 1:4 000 000 map based on Rzedowski's classification to a 1 km^2 grid resolution using random forests algorithm. The output was aimed to provide the necessary baseline information in the context of a larger research study (Newton 2008) as a first step to investigate patterns of tropical dry forest loss and fragmentation. If downscaling yields more accurate predictions and greater insight into the relationship between climate and vegetation, then it could be more generally extended and used to support land cover change studies as well as others applications related to landscape and conservation planning in Mexico.

More generally, our study aimed to evaluate the linkage between Rzedowski's structurally based vegetation classification and climate. The challenges that arise when downscaling may provide insight into the relationship between vegetation structure and climatic drivers.

Methods

Study area

The study area included the states of Veracruz, Oaxaca, Chiapas, Tabasco, Campeche, Quintana Roo and Yucatán. This area extends over ca. 400 000 km^2, including very diverse climatic and physiographic conditions.

The vegetation types that occur in Mexico are extremely varied as a result of its diverse climate and topography (Challenger 1998a). Mexico has one of the most diverse topographies of the world: 67% of its continental area is over 500 m above sea level (m a.s.l) and 50% is over more than 1000 m a.s.l. (Challenger 1998a). The physiography of the study area includes mountainous systems, coastal plains and hills, plateaus, valleys and high plains. The main climatic types are: tropical humid with different variations (with rain all the year through and with a short dry season), tropical sub humid, temperate humid with different variations (with rain all the year through and with winter dry season) and semiarid climate (Challenger 1998a). The most representative soils in the study area are: regosol, litosol, rendzinas, luvisol, cambisol and gleysol (SEMARNAT 1998).

Rzedowski (1990) defines 10 major vegetation types for Mexico, of which the following nine are represented in the study area: pine–oak forest, montane cloud forest, tropical evergreen forest, tropical subdeciduous forest, tropical deciduous forest, xerophytic shrubland, thorn forest, pastureland and aquatic vegetation. Of these, aquatic vegetation, pastureland and a group of communities of reduced extent in Mexico (e.g. palm forest) are considered 'azonal' by Rzedowski (1978), which means that they are not climatically driven. Aquatic vegetation is the result of naturally flooded soils (Rzedowski 1978). Pastureland is usually the consequence of human activities such as fire and cattle grazing (Miranda 1952; Sarukhán 1968; Puig 1972). Potentially, dry tropical forest is one of the most extensive vegetation types of Mexico (Challenger 1998b), covering over 60% of the total area of tropical vegetation (Trejo & Dirzo 2000).

Data

To obtain categorical values for the dependent variable (classes of potential vegetation), we used Rzedowski's 1:4 000 000 potential vegetation map. We systematically extracted points at 1-km spacing. We did not obtain samples from aquatic vegetation and pastureland because these vegetation types are 'azonal', and therefore it is more difficult to delineate their boundaries using solely climatically based models. Overall, we obtained 499 115 sampling points from the seven non-azonal vegetation types (i.e. excluding aquatic vegetation and pastureland).

At each point, we extracted values from 55 climatic variables obtained from the WorldClim site (http://www.worldclim.org/current). The climatic layers of this dataset were generated through interpolation of average monthly climate data from weather stations on a 30 arc-second resolution grid (equivalent to about 0.86 km^2; often referred to as '1 km' resolution). The dataset consists of 36 grids of monthly mean variables (calculated over the 1950–2000 period): minimum temperature (12 variables), maximum temperature (12 variables) and precipitation (12 variables); and a set of 19 bioclimatic variables. Bioclimatic variables are derived from the monthly temperature and rainfall values in order to generate more biologically meaningful variables. These variables represent annual trends, seasonality, and extreme or limiting environmental factors. These are: (1) annual mean temperature; (2) mean diurnal range; (3) isothermality; (4) temperature seasonality; (5) maximum temperature of warmest month; (6) minimum temperature of coldest month; (7) temperature annual range; (8) mean temperature of wettest quarter; (9) mean temperature of driest quarter; (10) mean temperature of warmest quarter; (11) mean temperature of coldest quarter; (12) annual precipitation; (13) precipitation of wettest month; (14) precipitation of driest month; (15) precipitation seasonality; (16) precipitation of wettest quarter; (17) precipitation of driest quarter; (18) precipitation of warmest quarter; and (19) precipitation of coldest quarter. A detailed description of the climatic dataset can be found in Hijmans et al. (2005).

Although most of the bioclimatic variables associated with temperature rely on its mean annual and seasonal values, these variables may not always or even frequently have a biological meaningful (Box 1995). In contrast, different aspects of the scattering of temperature values, may be more relevant to explain vegetation patterns. Rzedowski (1978), for example, highlighted the importance of diurnal temperature oscillations in explaining the distribution of vegetation types in Mexico. Box (1995) in turn, suggested that maximum and minimum temperatures, among others, may be the most important climate-related limiting factors and mechanisms for terrestrial vegetation types. In order to incorporate additional variables involving the variability and tendencies of temperature values, we finally derived two more variables from the existing dataset which are temperature ranges that underscore the fluctuation in these values, and may be relevant to explain vegetation patterns: (1) annual

Applied Vegetation Science
Doi: 10.1111/j.1654-109X.2011.01132.x

difference in maximum temperature (calculated as: highest maximum temperature – lowest maximum temperature); and (2) maximum daily temperature difference (the maximum value obtained from: monthly maximum temperature – monthly minimum temperature).

Data analysis

Random forest

We used climatically based random forests models to find the explanatory variables that contribute the most to explain patterns of potential vegetation types as defined in Rzedowski's map (1990). Random forest is a non-parametric method used to predict membership of cases or objects in the classes of a categorical dependent variable from their measurements on predictor variables. The algorithm produces multiple independent trees (an ensemble of trees) using a bootstrap sample of the data set, each of which produces a vote (an instance statistic) (Breiman 1996). In the end, the set of votes is used to generate a simple majority vote for prediction, or scores that provide basic probability estimates, which may then be used in weighted voting (Fawcett 2006).

The following attributes of original data required consideration for model building: (1) data were unbalanced; (2) data were noisy (particularly at boundaries of vegetation types); and (3) data had some degree of class overlapping at their boundaries. The following steps were followed in order to deal with such data. First, a finer-grained potential vegetation map was predicted based on majority vote prediction rules. We used all 57 explanatory variables to fit the random forests model and predict membership of potential vegetation types. The analysis was performed with 1000 trees. Random forest's strength lies in its predictive ability, therefore unbalanced data poses less of a problem than it would if the model were used in an explanatory capacity. We therefore did not require a balanced sample size from each class (Rogan et al. 2008). The bootstrap training sample on which each tree is grown randomly selects approximately 63.2% of the cases ($1/e$ of instances are not considered in tree construction). We controlled the size of the individual trees by defining the minimum size node to split (Batista et al. 2004). The minimum size of the nodes was set to 1% of the total data thus avoiding the most specialized and less significant branches.

In the presence of unbalanced data and class overlaps, decision trees may need to create many nodes to distinguish the minority class cases from majority class cases. Pruning the decision trees can be useful to deal with this by removing over-specialized branches. However pruning does not always prevent overfitting (Batista et al. 2004). In such cases, the probability estimates provided by random forests models can be used a posteriori, as threshold for predictions. The main motivation behind such approaches is to remove noisy examples lying on the wrong side of the decision border. The removal of noisy examples might help to find better-defined class clusters, allowing the creation of simpler models with better generalization capabilities (Batista et al. 2004). In order to improve the model first developed based on majority vote prediction rules, we followed some final steps. We validated the model from forest inventory plots (see accuracy assessment section) and used validation results to obtain insights into the overall performance of the model and the performance acquired for the different vegetation classes, especially for those unbalanced minority classes in the original data. We used the probability estimates provided by random forests to find the thresholds that best define the distribution area of classes whose accuracy remained low (see the next section 'ROC curves and decision thresholds for random forest'). We finally used these thresholds to weight votes at the corresponding pixels in the original model and produce a new potential vegetation map.

Random forests models are expected to have a degree of predictive accuracy that cannot be obtained using single-tree models (Breiman 2001; Prasad et al. 2006; Siroky 2009). The disadvantage is that the model cannot be visualized as a single tree (Breiman 2001; Siroky 2009). Nevertheless, one useful property of random forest is that it provides a proximity matrix (a similarity measure between objects) (Liaw & Wiener 2002). The proximity matrix can be used as input for non-metric multidimensional scaling (NMDS), allowing the description of clusters (Liaw & Wiener 2002; Shi & Horvath 2006). Thus, we finally used NMDS in order to visualize the classification and describe the climatic characteristics associated with the potential vegetation (see the Non-metric multidimensional scaling section later).

Receiver operating characteristics (ROC) curves and decision thresholds for random forests

We used ROC curves to select the decision thresholds (cut-off thresholds based on probability estimates) that encompass the distribution limits for classes with low agreement and maximize overall prediction accuracy (Fawcett 2006). We defined the decision threshold as the value in the ROC curve that minimizes false positives while maximizing true positives, that is, the same as maximizing the total accuracy (Zweig & Campbell 1993; Chin et al. 2009). We extracted this cut-off value for the classes of interest, by finding the value that maximizes:

$$\sqrt{\frac{TP}{TP+FP}+\frac{TP}{TP+FN}}$$

Doi: 10.1111/j.1654-109X.2011.01132.x

TP, 'true positive rate'; *FP*, 'false positive rate'; *FN*, 'false negative rate' (Chin et al. 2009).

Once the potential vegetation map was generated using threshold values, the distribution of aquatic vegetation was defined based on a soil map developed for Mexico at a scale 1:1 000 000 by governmental agencies (INIFAP & CONABIO 1995). We did not include pastureland in the final map because this vegetation type is originated by human activities throughout Mexico.

Non-metric multidimensional scaling

We used NMDS in order to visualize the classification through projecting the results onto a lower dimensional space. Non-metric multidimensional scaling is a dimension reduction technique based on proximities between objects (samples), where proximities express their similarity or dissimilarity on a multidimensional space represented by a set of variables (Breiman 2003; Härdle & Simar 2007).

We then plotted specific climatic variables onto the NMDS ordination by fitting thin plate splines using general additive models. This allowed us to produce a climatic description of major vegetation types by looking at where the vegetation classes fall. The variables plotted onto the NMDS ordination were defined by means of the permutation accuracy importance measure. This is the average per cent change in predictive accuracy when a variable is included and then excluded from the model (Strobl et al. 2008).

Accuracy assessment

To test the accuracy of the different vegetation maps, field data was obtained from different projects. From these datasets, we selected plots that were located in areas of low anthropogenic disturbance (remnants of mature vegetation and protected areas), and near the transitions between different vegetation types. We selected sites near the transitional areas, because most of the problems of misclassification in coarse-grained potential vegetation maps are restricted to the boundaries between vegetation types, thus allowing the compilation of data in the areas with the highest rates of misclassification. However, we purposefully avoided plots at ecotones in order to compile data in areas of high vegetation homogeneity, and separated less than 2 km, thus ameliorating scale differences between maps and ground data. Furthermore, maps represent homogeneous areas (vegetation classes) and do not contain ecotones. These selection criteria unavoidably produce a large inequality in the number of plots per vegetation type.

We obtained measures of the floristic composition in 256 floristic circular plots of 1000 m^2 each. Rzedowski (1978) provided an extensive description of the dominant floristic composition of each vegetation type. This allowed us to assign each sampled plot to one of Rzedowski's vegetation types based on species composition. The plots were located at the following sites: 46 plots in the transition between pine–oak forest and tropical deciduous forest in the Highlands of Chiapas, 76 plots in the transition between pine–oak forest and tropical deciduous forest in the Sierra Madre of Chiapas, 86 plots in the transition between pine–oak forest and montane cloud forest in the Sierra Madre of Chiapas, seven plots in the transition between tropical evergreen forest and tropical deciduous forest in the Central Valley, six plots in the transition between tropical evergreen forest and montane cloud forest in the Lacandona Forest region, and 35 plots in the transition between tropical evergreen forest and montane cloud forest in the Sierra Madre of Chiapas. This allowed us to test the accuracy of our downscaled map compared with Rzedowski's original vegetation map in four out of five of the major vegetation types found in Chiapas. Figure 1 shows the location of the sampled plots in the state of Chiapas.

Confusion matrices and the Kappa Index of Agreement with 95% confidence intervals were used to estimate consistency of classifications accuracy (Rosenfield & Fitzpatrick-Lins 1986). Results were compared with those obtained from Rzedowski's potential vegetation map.

Software used

Random forests analysis was performed with the randomForest R package (version 4.5-3. Liaw & Wiener 2002; FORTRAN original by Breiman 2001). The ROC analysis was performed with the ROCR R package (version 1.0-4, Sing T., Sander O., Beerenwinkel N. & Lengauer T. 2009). Spatial data was handled with GRASS (Geographic Resources Analysis and Support System, version 6.3.0. GRASS Development Team, 2008; ITC-irst. Trento, Italy), POSTGIS (version 1.3.3; Refractions Research Inc., Victoria, BC, CA), and Quantum GIS (Quantum GIS Geographic Information System, version 1.0.1-Kore, 2009; Quantum GIS Development Team, Open Source Geospatial Foundation Project, http://qgis.org/).

Results

Model validation

The Kappa Index of Agreement showed an increase in accuracy from 0.40 for the Rzedowski map (95% confidence intervals between 32.5 and 48.7%) to 0.56 for the climatically derived map with majority vote prediction rules (95% confidence intervals between 48.6 and 64.2%). Overall agreement increased from 55.5% to 68.0% (Table 1). Agreement was improved for all classes when using random forests with majority vote predictions. Improvement for the montane cloud forest class, however, remained low and there was a high commission rate to other forest types, particularly pine–oak forest. To improve accuracy of this class we used ROC curves. A threshold of 0.18 maximized overall accuracy while increasing class accuracy. We used this result to weight votes in the original model, and produce a new potential

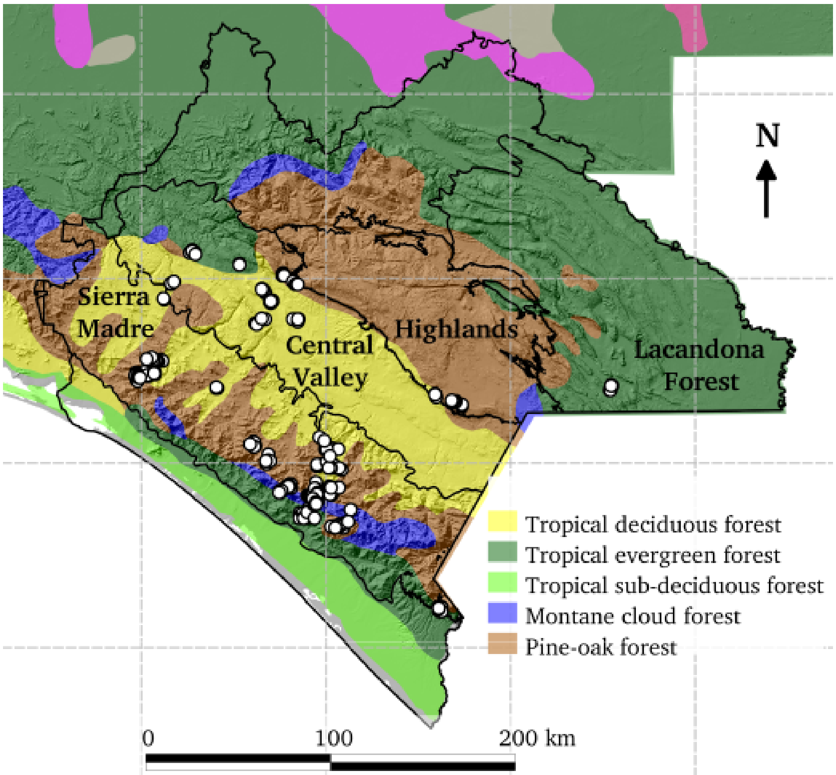

Fig. 1. Location of verification plots within the state of Chiapas (white points). Rzedowski's potential vegetation map (color map) and the physiographic regions of Chiapas mentioned in the text (delineated by dotted black lines) are also shown.

Table 1. Confusion matrices for: (a) Rzedowski's map, (b) climatically derived map with majority vote prediction rules and (c) climatically derived map with probability predictions combined. Rzedowski (1978).

	Verification points							
	Df	Ef	Po	M	Sum	% Commission	% Omission	Estimated Kappa
(a) Rzedowski								
Deciduous forest (Df)	35	7	10	0	52	32.7	36.3	0.58
Evergreen forest (Ef)	0	22	0	0	22	0	53.2	1
Pine-oak forest (Po)	16	0	59	1	76	22.3	53.5	0.55
Montane cloud forest (M)	4	18	58	26	106	75.5	3.7	0.15
Sum	55	47	127	27	256			
(b) Random forests (prediction rule: majority vote)								
Deciduous forest (Df)	49	0	3	0	52	5.8	27.9	0.92
Evergreen forest (Ef)	0	22	0	0	22	0	35.3	1
Pine–oak forest (Po)	19	0	57	0	76	25.0	47.2	0.57
Montane cloud forest (M)	0	12	48	46	106	56.6	0	0.31
Sum	68	34	108	46	256			
(c) Random forests (probability threshold for M)								
Deciduous forest (Df)	49	0	3	0	52	5.8	27.9	0.92
Evergreen forest (Ef)	0	22	0	0	22	0	18.5	1
Pine–oak forest (Po)	19	0	49	8	76	35.5	7.5	0.55
Montane cloud forest (M)	0	5	1	100	106	5.6	7.4	0.90
Sum	68	27	53	108	256			

Doi: 10.1111/j.1654-109X.2011.01132.x

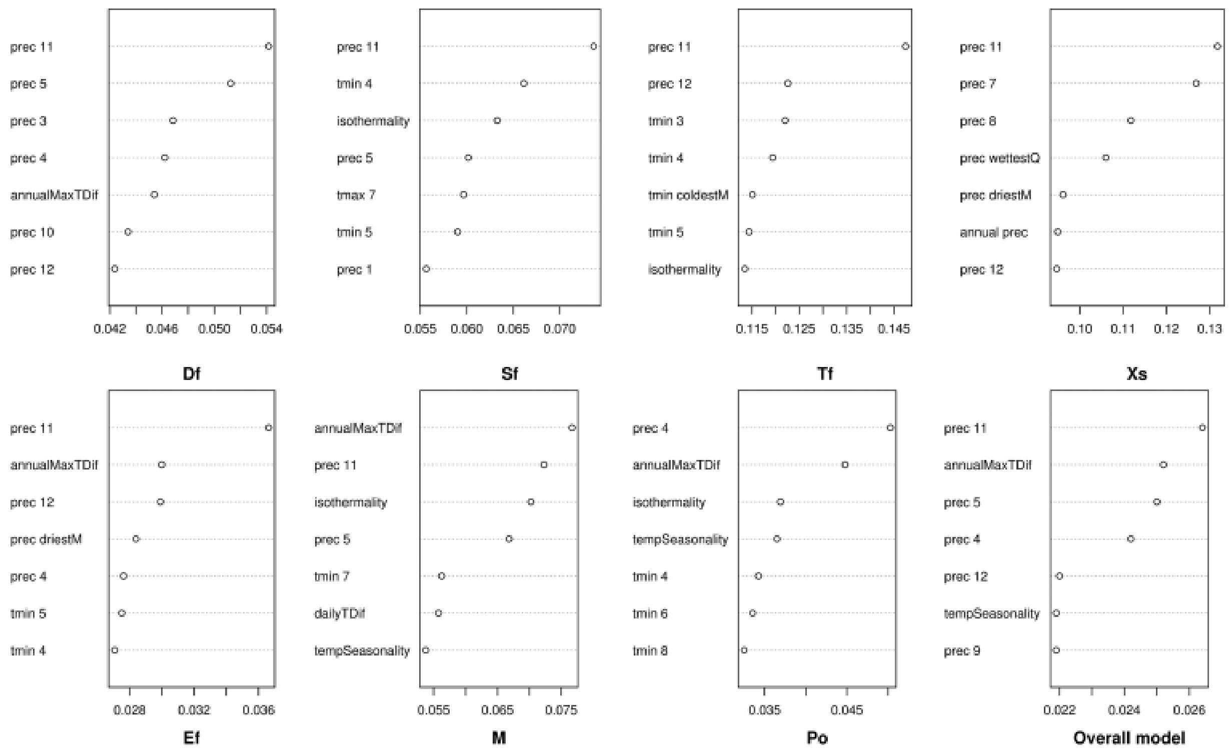

Fig. 2. Variable importance plots for vegetation classes: Tropical deciduous forest (Df); tropical sub-deciduous forest (Sf); tropical evergreen forest (Ef); thorn forest (Tf); xerophytic shrubland (Xs); montane cloud forest (M); and pine–oak forest (Po). *X* axis: importance index (mean decrease accuracy). Predictor variables: numbers are months of the year (e.g. 1 for Jan, 12 for Dec). Precipitation = prec; maximum temperature = *t*max; minimum temperature = *t*min; M = month; and Q = quarter (group of 3 months). Thus, for example: precipitation for Nov = prec 11; maximum temperature of Jul = *t*max 7; minimum temperature of Apr = *t*min 4; precipitation of driest month = prec driestM; precipitation of wettest quarter = prec wettestQ. Annual precipitation = annual prec; annual difference in maximum temperature = annualMaxTDif; maximum daily temperature difference = dailyTDif; and temperature seasonality = tempSeasonality.

vegetation map with the distribution of montane cloud forest redraw. The Kappa Index of Agreement for the climatically derived map with probability predictions combined increased to 0.80 (95% confidence intervals between 73.9 and 86.0%). Overall agreement increased from 55.5% to 85.9% (Table 1). The downscaled map showed montane cloud forest occurrence in some important areas that were not included in the original map. Examples include the Northern Mountains of Chiapas (Challenger 1998c).

Climatic description of major vegetation types

Figure 2 shows the seven most important variables for each vegetation class and for the overall model. The most important variables for the overall model were also shown to be the most important predictors of individual vegetation types. In particular, we considered the following variables for interpolation onto the unconstrained NMDS ordination diagram (Fig. 3): precipitation during Nov, annual difference in maximum temperature, and precipitation during dry months (Apr, May).

The major vegetation gradients are revealed when using the first two ordination axes of the ordination. These gradients are: (1) tropical evergreen forest – montane cloud forest – pine-oak forest; (2) tropical evergreen forest – tropical subdeciduous forest – tropical deciduous forest; (3) tropical deciduous forest – pine–oak forest.

Precipitation during Nov, as a surrogate of the extent of the rainy period, clearly differentiated between humid and dry vegetation types (Fig. 3a). Tropical deciduous forest is the driest vegetation type, with the rainy period finishing by Nov (precipitation regime from 0 to 20 mm). Intermediate precipitation (40–80 mm) is associated with tropical subdeciduous forest, thorn forest, xerophytic shrubland and pine-oak forest. Higher precipitation occurs in tropical evergreen forests (80–100 mm) and montane cloud forest (> 100 mm), indicating that the rainy period extends over Nov for these vegetation types.

Annual difference in maximum temperature (Fig. 2) resulted in one of the most important variables for overall model. This variable is interpreted as a surrogate of the frequency and persistence of extreme temperature events, and was useful in explaining the climate-related

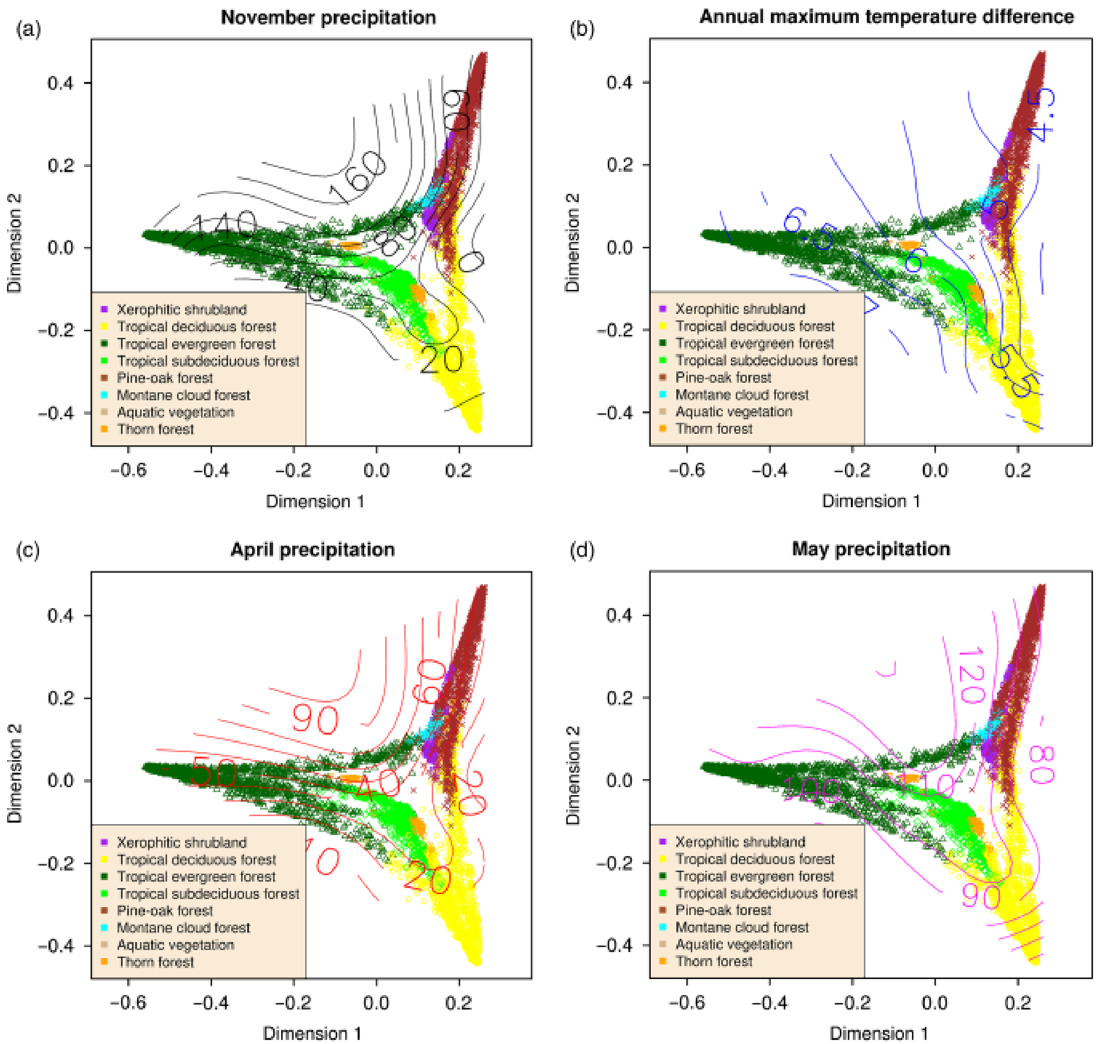

Fig. 3. Multidimensional scaling plot for Rzedowski's vegetation types in Southern Mexico. Fitted contours represent the main predictors of vegetation distribution, as mentioned in the title of each of the figures.

limiting factors for tropical vegetation types that follow elevational gradients, particularly for one of the existing vegetation gradients: tropical evergreen forest – montane cloud forest – pine–oak forest. Temperature extremes may affect vegetation through limitation of water uptake or by causing excessive water loss (Box 1995), and have consequences for productivity and ecological interactions (Strandman et al. 1993). This variable clearly discriminated (Fig. 3b) between more temperate vegetation types (pine–oak forest and montane cloud forest) distributed in mountainous areas, with lower oscillations in maximum temperatures (under 5 degrees), and tropical vegetation types (tropical evergreen forest and tropical deciduous forest), predominantly distributed in plain areas, with higher oscillations in maximum temperature (5.5–7 degrees).

Finally, precipitation rate during Apr and May can be important predictors to discriminate between some forest types, such as pine–oak and tropical deciduous forest (Fig. 3c,d). This is related to the length of the dry season, which extends over Apr (< 20 mm) for deciduous forest, until May (up to 80 mm); while in pine-oak forests it often comes to an end in Apr.

Extent of major potential vegetation types

Figure 4 shows the distribution of potential vegetation types, as described by Rzedowski's map and the resulting climatically derived map. Table 2 shows the total extent of Southern Mexican vegetation types, as estimated by both maps. The area of tropical deciduous forest, tropical evergreen forest, montane cloud forest and aquatic vegetation

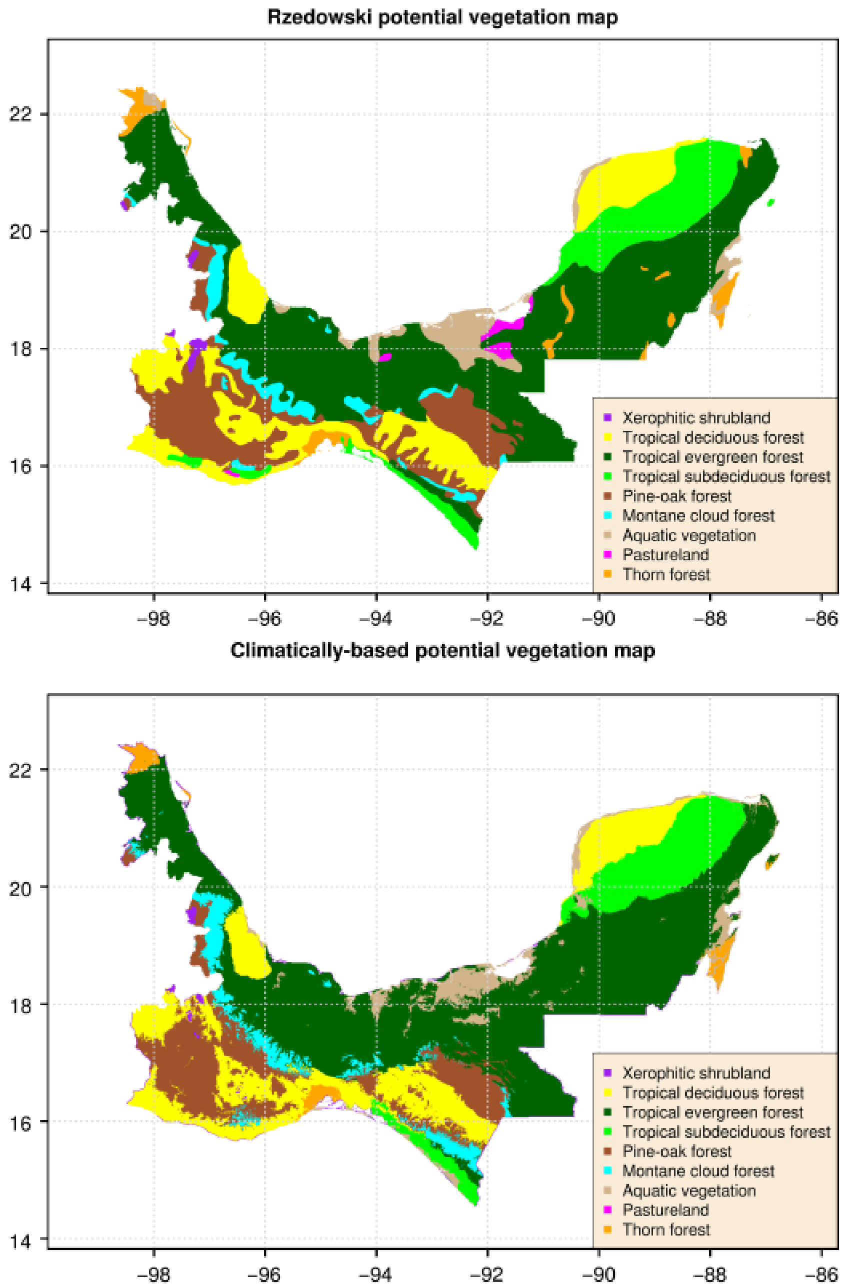

Fig. 4. Distribution of Southern Mexican vegetation types as described by (1) Rzedowski's potential vegetation map, and (2) random forests combining majority vote prediction and probability prediction rules. No post-processing was performed on this map.

increased in the downscaled map. The area of tropical subdeciduous forest, pine–oak forest, thorn forest and xerophytic shrubland was reduced as a result of downscaling. Montane cloud forest, aquatic vegetation, thorn forest, and xerophytic shrubland registered the highest per cent change. Montane cloud forest and tropical deciduous forest showed an increase in their predicted distribution areas, mainly at the expense of a reduction in the distribution area of pine–oak forest (see Table 1). Aquatic vegetation was increased mainly because of its predicted distribution with soil data in the Pacific coast of Chiapas. This area is characterized by the occurrence of mangroves (Breedlove 1973), a vegetation type that is classified as aquatic vegetation by Rzedowski (1978). Nevertheless, this vegetation type was not delineated for the coast of Chiapas by the author. Thorn forest showed a reduction in its total extent. Although it was possible to climatically predict its occurrence at the coast areas of

Applied Vegetation Science
Doi: 10.1111/j.1654-109X.2011.01132.x

Table 2. Total extent of Southern Mexican vegetation types as estimated by Rzedowski's potential vegetation map and by the climatically-based random forests model. Rzedowski (1978).

Vegetation type	Potential vegetation map	Area (km^2)	Differences (km^2)	Differences (%)
Tropical deciduous forest	Rzedowski	66 963		
	Climatically based	69 659	2696	4.03
Tropical evergreen forest	Rzedowski	182 637		
	Climatically based	189 015	6378	3.49
Tropical sub-deciduous forest	Rzedowski	40 682		
	Climatically based	37 965	− 2717	6.68
Pine-oak forest	Rzedowski	58 134		
	Climatically based	52 174	− 5960	10.25
Montane cloud forest	Rzedowski	13 924		
	Climatically based	17 880	3956	28.41
Aquatic vegetation	Rzedowski	17 390		
	Climatically based	20 264	2874	16.53
Thorn forest	Rzedowski	9882		
	Climatically based	6376	− 3506	35.48
Xerophytic shrubland	Rzedowski	1811		
	Climatically based	1203	− 608	33.57
Pastureland	Rzedowski	3162		

Mexico, this was not feasible for the interior areas of the Yucatan Peninsula (see Fig. 4) where its occurrence is associated with flooded soils (Rzedowski 1978).

Discussion

The study showed that contemporary computational tools can be effectively employed to refine existing maps. Other authors have found decision tree algorithms satisfactory to establish relationships between vegetation units and variables for predictive vegetation mapping (Michaelsen et al. 1994; van Etten 1998; Muñoz & Felicísimo 2004).

At the same time, the results demonstrate errors in the original map. Rzedowski's (1978) map underestimated the extent of potential montane cloud forest with a high commission rate to pine–oak forest. The author, pointed out that: (1) montane cloud forest and pine–oak forest often extent and overlap along the same elevational levels, with montane cloud forest being mainly restricted to slopes protected from insolation; and (2) the coarse resolution of the map hindered the accurate representation of the naturally fragmented and dispersed distribution of montane cloud forest, resulting in an underestimation of its extent. Alcántara et al. (2002) in turn, confirmed that montane cloud forest follows complex distribution patterns along elevation, and that its geographic distribution follows an archipelago-like pattern across Mexico that has not been thoroughly investigated. Cayuela et al. (2006) found major discrepancies between coarse-scale maps and fine-scale maps showing the extent and distribution of montane cloud forest. Montane cloud forest is associated with microclimatic conditions (persistent mist or clouds at the vegetation level, resulting in the reduction of direct sunlight and thus of evapotranspiration) that create large local differences in mountainous areas (Brown & Kappelle 2001). The spatial resolution of climatic surfaces do not capture all the variation that may occur at a resolution of 1 km particularly in mountainous areas, owing to the overall low density of available climate stations, and because locally important drivers such as aspect may be ignored (Hijmans et al. 2005). This could lead to a non-definitive solution as occurred in the present study. Although accuracy in classification of montane cloud forest was improved through the use of probability estimates, there was an increased misclassification of pine–oak forest as montane cloud forest. The ROC analysis improved classification, but generalizing this analysis to more than two classes is problematic (Hand & Till 2001; Lachiche & Flach 2003; Fawcett 2006). Future work may include further studies on the feasibility of finding a globally optimal multi-class decision criterion.

Tropical deciduous forest and pine–oak forest were also confused in Rzedowski's map. While tropical deciduous forest classification accuracy was markedly improved with climatically based random forests models, there was still a high commission rate of pine–oak forest as tropical deciduous forest (19 plots) in our final model, and the estimated Kappa value remained low. Rzedowski (1978) pointed out that the lower warm limits of pine–oak forest in different parts of Mexico are occupied by discontinuous oak forests, which are very common in the dry hills of the Central Valley of Chiapas, and can be found at elevations as low as 700 m a.s.l. This vegetation type is predominantly composed of deciduous *Quercus* species with microphilous and sclerophyllous leaves:

Q. peduncularis, *Q. polymorpha* and *Q. conspersa* are the most common species, and this particular physiognomy distinguishes them from pine–oak forests occurring at higher elevations with temperate humid conditions (composed of perennial *Quercus* species, with macrophilous leaves). These characteristics are suggested to be adaptations to drought conditions. In the present study, 12 plots from the group of 20 oak forest plots were classified as tropical deciduous forest, thus reflecting climatic conditions for potential tropical vegetation within the pine–oak zone. This observation suggests that factors in addition to climate may be important in determining the transition between the two vegetation types. Several authors (Sarukhán 1968; Pennington & Sarukhán 2005) have suggested that these oak forests that now occur in comparatively warm climates represent relict vegetation associated with previously cooler climate conditions. Misclassification of tropical forest as pine oak also occurred in mountainous areas with steep terrain where dry forest occurs in close proximity with oak forest (Rzedowski 1978; Challenger 1998b). Trejo & Dirzo (2000) found that up to 75% of the steeper terrain in the tropical dry forest areas of the state of Morelos are still covered by dry forest, most of which can be considered as intact forest, compared with areas of gentle slopes where agricultural activity has resulted in $< 20\%$ of the area being covered by dry forest.

The results suggest that the potential vegetation map in current usage may have insufficient spatial resolution, particularly in areas of vegetation transition in which fine-scaled climatic gradients determine the observed vegetation type such as mountainous areas. However, while this heterogeneity may be overlooked by coarse-scale maps, steep temperature and precipitation gradients in mountainous regions make computer-based predictive mapping easier as the relationship between vegetation and climate may be more apparent.

There are other caveats to this study. Validation of any classification of potential vegetation is complicated by human impact and other non-climatic determinants of the actual vegetation found on the ground. Although the plots chosen were assumed to have been subjected to only a low level of human impact, given the long history of human habitation in the region most of the vegetation has been affected in some manner. There is therefore no guarantee that the present vegetation corresponds to the theoretical potential vegetation. Validation plots were sited close to the assumed boundaries of Rzedowski's vegetation types. Although the thematic map assumes that such boundaries can be drawn, such areas may consist of ecotones with a mixture of species drawn from more than a single vegetation type. Such issues raise important questions regarding the underlying nature of plant communities; these questions have been the topic of ongoing debate in vegetation science since the time of Clements and Gleason (Pickett et al. 2009).

Rzedowski's classification has been used as a baseline map for other widely used classifications. For example Olson et al. (2001) used the scheme in a world ecoregions map and Toledo & Ordóñez (2009) for a new ecoregions map of Mexico. These uses of Rzedowski's map rest on the assumption that visible structural attributes can be used as surrogates for the more illusive functional patterns that are related to climate (Box 1995). Our study suggested that a scheme based only on structural attributes, may not accurately reflect relationships between climate and vegetation type. While analysis of structural characteristics may assume that form follows function (Box 1995), the relationship between vegetation structure and climate is not straightforward. For example, transitions between savannah and dry forest in Southern Mexico are known to be determined by additional non-climatic factors (Pérez-García & Meave 2006). Rzedowski's justified his scheme on pragmatic, rather than theoretical grounds arguing that the criteria used to describe the vegetation types, were selected in order to facilitate recognition on the ground (Rzedowski 1978). Explicitly climate-based classification schemes such as the life zones approach of Holdridge (1947) may provide greater insight into climatically associated vegetation patterns and more studies are required that contrast and compare the spatial patterns suggested by these alternative classification schemes. Multivariate techniques may also be used that allow vegetation composition to define classification schemes in relation to climate (Golicher et al. 2008).

Conclusions

Downscaling increased accuracy as measured by the Kappa Index of Agreement. Multivariate analysis confirmed that many of the structural aspects of the vegetation that are used by the Rzedowski classification are associated with climate, but it also indicated some of the weaknesses in the underlying basis of this classification system. Rzedowski's scheme for vegetation classification may require further modification in order to be an effective tool for research into vegetation–climate relationships. Further research on the relationship between vegetation type and climate is required in order to produce vegetation maps that can inform regional decision making. This may be particularly important in the current context of changing climate.

Acknowledgements

Field data were obtained as part of the Project ReForLan (Restoration of Forest Landscapes for Biodiversity

Conservation and Rural Development in the Drylands of Latin America), and from Biodiversidad: Conservación y Restauración, AC. We are grateful to the three anonymous reviewers whose extensive comments were extremely useful and intellectually stimulating, and helped to improve the clarity, accuracy and coherence of the manuscript.

References

Alcántara, O., Luna, I. & Velázquez, A. 2002. Altitudinal distribution patterns of Mexican cloud forests based upon preferential characteristic genera. *Plant Ecology* 161: 167–174.

Araújo, M.B., Thuiller, W., Williams, P.H. & Reginster, I. 2005. Downscaling European species atlas distributions to a finer resolution: implications for conservation planning. *Global Ecology and Biogeography* 14: 17–30.

Batista, G.E., Prati, R.C. & Monard, M.C. 2004. A study of the behavior of several methods for balancing machine learning training data. *ACM SIGKDD Explorations Newsletter* 6: 20–29.

Benítez Badillo, G., Hernández Huerta, A., Equihua Zamora, M.E., Medina Chena, A., Álvarez Palacios, J.L., Ibáñez Bernal, S. & Delfin Alonso, C. 2003. *Biodiversidad y cambio climático*. Instituto de Ecología, A.C. (INECOL) Departamento de Ecología Aplicada, Xalapa, MX.

Box, E.O. 1995. Factors determining distributions of tree species and plant functional types. *Vegetatio* 121: 101–116.

Bredenkamp, G., Chytrý, M., Fischer, H.S., Neuhäuslová, Z. & Maarel, E.V. 1998. Vegetation mapping: Theory, methods and case studies. *Applied Vegetation Science* 1: 161–266.

Breedlove, D.E. 1973. The phytogeography and vegetation of Chiapas (Mexico). In: Graham, A. (ed.) *Vegetation and vegetational history of northern Latin America*. pp. 145–165. Elsevier, Amsterdam.

Breiman, L. 1996. Bagging predictors. *Machine Learning* 24: 123–140.

Breiman, L. 2001. Random forests. *Machine Learning* 45: 5–32.

Breiman, L. 2003. Manual—setting up, using and understanding random forests V4.0. Available at: ftp://ftp.stat.berkeley.edu/pub/users/breiman/Using random forests v4.0.pdf

Brown, A.D. & Kappelle, M. 2001. Introducción a los bosques nublados del Neotrópico: Una síntesis regional. In: Kappelle, M. & Brown, A.D. (eds.) *Bosques nublados del Neotrópico*. pp. 25–40. Instituto Nacional de Biodiversidad, Costa Rica.

Cayuela, L., Golicher, D.J. & Rey-Benayas, J.M. 2006. The extent, distribution, and fragmentation of vanishing montane cloud forest in the Highlands of Chiapas, Mexico. *Biotropica* 38: 544–554.

Challenger, A. 1998a. *Utilización y conservación de los ecosistemas terrestres de México. Pasado, presente y futuro. Ambiente físico y zonas ecológicas de México*. Comisión Nacional para el conocimiento y uso de la biodiversidad (CONABIO), Mexico pp. 269–292.

Challenger, A. 1998b. *Utilización y conservación de los ecosistemas terrestres de México. Pasado, presente y futuro. La zona ecológica templada húmeda (bosque mesófilo de montaña)*. Comisión Nacional para el conocimiento y uso de la biodiversidad (CONABIO), Mexico pp. 443–508.

Challenger, A. 1998c. *Utilización y conservación de los ecosistemas terrestres de México. Pasado, presente y futuro. La zona ecológica tropical subhúmeda (selva subhúmeda)*. Comisión Nacional para el conocimiento y uso de la biodiversidad (CONABIO), Mexico pp. 375–433.

Challenger, A. & Soberón, J. 2008. *Capital natural de México, vol. I: Conocimiento actual de la biodiversidad. Los ecosistemas terrestres*. Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (CONABIO), Mexico pp. 87–108.

Chin, F.Y., Leung, H.C., Li, W. & Yiu, S. 2009. Finding optimal threshold for correction error reads in DNA assembling. *BMC Bioinformatics* 10: S15.

Cross, J. 1998. An outline and map of the potential natural vegetation of Ireland. *Applied Vegetation Science* 1: 241–252.

Cué-Bär, E.M., Villaseñor, J.L., Morrone, J.J. & Ibarra-Manríquez, G. 2006. Identifying priority areas for conservation in Mexican tropical deciduous forest based tree species. *Interciencia* 31: 712–719.

Demers, M.N. 1991. Classification and purpose in automated vegetation maps. *Geographical Review* 81: 267–280.

Fawcett, T. 2006. An introduction to ROC analysis. *Pattern Recognition Letters* 27: 861–874.

Felix, A.B., Campa, H., Millenbah, K.F., Winterstein, S.R. & Moritz, W.E. 2004. Development of landscape-scale habitat-potential models for forest wildlife planning and management. *Wildlife Society Bulletin* 32: 795–806.

Franke, J. & Köstner, B. 2007. Effects of recent climate trends on the distribution of potential natural vegetation in Central Germany. *International Journal of Biometeorology* 52: 139–147.

Franklin, J. 1995. Predictive vegetation mapping: geogaphic modelling of biospatial patters in relation to environmental gradients. *Progress in Physical Geography* 19: 474–499.

García, E. 1988. *Modificaciones al sistema de clasificación climática de Köppen*. Offset Larios, Mexico.

Golicher, D.J., Ramírez-Marcial, N. & Levy-Tacher, S.I. 2006. Correlations between precipitation patterns in Southern Mexico and the El Niño sea surface temperature index. *Interciencia* 31: 80–86.

Golicher, D.J., Cayuela, L., Alkemade, J.R., González-Espinosa, M. & Ramírez-Marcial, N. 2008. Applying climatically associated species pools to the modelling of compositional change in tropical montane forests. *Global Ecology and Biogeography* 17: 262–273.

González Medrano, F. 2003. *Las comunidades vegetales de México. Propuesta para la unificación de la clasificación y nomenclatura de la vegetación de México*. Instituto Nacional de Ecología (INE-SEMARNAT), Mexico pp. 87.

Hand, D.J. & Till, R.J. 2001. A simple generalisation of the area under the ROC curve for multiple class classification problems. *Machine Learning* 45: 171–186.

Härdle, W. & Simar, L. 2007. *Applied multivariate statistical analysis*. Springer, Berlin, DE.

Hartley, S., Kunin, W.E., Lennon, J.J. & Pocock, M.J. 2004. Coherence and discontinuity in the scaling of specie's distribution patterns. *Proceedings of The Royal Society/ Biological sciences* 271: 81–88.

Hessburg, P.F., Smith, B.G. & Brion Salter, R. 1999. Detecting change in forest spatial patterns from reference conditions. *Ecological Applications* 9: 1232–1252.

Hickler, T., Smith, B., Sykes, M.T., Davis, M.B., Sugita, S. & Walker, K. 2004. Using a generalized vegetation model to simulate vegetation dynamics in Northeastern USA. *Ecology* 85: 519–530.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* 25: 1965–1978.

Holdridge, L.R. 1947. Determinations of world formulations for simple climatic data. *Science* 105: 367–368.

Hopkinson, P., Evans, J. & Gregory, R.D. 2000. National-scale conservation assessments at an appropriate resolution. *Diversity and Distributions* 6: 195–204.

Hughes, R.M., Larsen, D.P. & Omernik, J.M. 1986. Regional reference sites: a method for assessing stream potentials. *Environmental Management* 10: 629–635.

Hulme, P.E. 2003. Biological invasions: winning the science battles but losing the conservation war? *Oryx* 37: 178–193.

Instituto Nacional de Investigaciones Forestales y Agropecuarias (INIFAP) & Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (CONABIO). 1995. *Edafología. Escalas 1:250 000 y 1:1 000 000*. México DF. http://www.inifap.gob.mx/ and http://www.conabio.gob.mx/

The International Plant Names Index. 2008. Published on the Internet http://www.ipni.org

Koca, D., Smith, B. & Sykes, M.T. 2006. Modelling regional climate change effects on potential natural ecosystems in Sweden. *Climatic Change* 78: 381–406.

Kostopoulou, E., Giannakopoulos, C., Anagnostopoulou, C., Tolika, K., Maheras, P., Vafiadis, M. & Founda, D. 2006. Simulating maximum and minimum temperature over Greece: a comparison of three downscaling techniques. *Theoretical and Applied Climatology* 90: 65–82.

Kunin, W.E., Hartley, S. & Lennon, J.J. 2000. Scaling down: on the challenge of estimating abundance from occurrence patterns. *American Naturalist* 156: 560–566.

Lachiche, N. & Flach, P. 2003. Improving accuracy and cost of two-class and multi-class probabilistic classifiers using ROC curves. *Proceedings of the Twentieth International Conference on Machine Learning*, pp. 416–423. AAAI Press.

Liaw, A. & Wiener, M. 2002. Classification and regression by random forest. *R News* 2: 18–22.

Lüttge, U. 2008. *Physiological ecology of tropical plants*. Springer-Verlag, Berlin, DE, 458pp.

Mcpherson, J.M., Jetz, W. & Rogers, D.J. 2006. Using coarse-grained occurrence data to predict species distributions at finer spatial resolutions—possibilities and limitations. *Ecological Modelling* 192: 499–522.

Michaelsen, J., Schimel, D.S., Friedl, M.A., Davis, F.W. & Dubayah, R.C. 1994. Regression tree analysis of satellite and terrain data to guide vegetation sampling and surveys. *Journal of Vegetation Science* 5: 673–686.

Miranda, F. 1952. *Vegetación de Chiapas*. Volumen 1, Ediciones del Gobierno del Estado, Chiapas, MX.

Miyawaki, A. 2004. Restoration of living environment based on vegetation ecology: theory and practice. *Ecological Research* 19: 83–90.

Moravec, J. 1998. Reconstructed natural versus potential natural vegetation in vegetation mapping: a discussion of concepts. *Applied Vegetation Science* 1: 173–176.

Muñoz, J. & Felicísimo, A.M. 2004. Comparison of statistical methods commonly used in predictive modelling. *Journal of Vegetation Science* 15: 285–292.

Newton, A.C. 2008. Restoration of Dryland Forests in Latin America: the ReForLan Project. *Ecological Restoration* 26: 10–13.

Olson, D.M., Dinerstein, E., Wikramanaya, E.D., Burgess, N.D., Powel, G.V.N. & Underwood, E.C., *et al*. 2001. Terrestrial ecoregions of the world: a new map of life on earth. *BioScience* 51: 933–938.

Pennington, T.D. & Sarukhán, J. 2005. Árboles tropicales de México. Manual para la identificación de las principales especies, pp. 523. Universidad Nacional Autónoma de México. Dirección General de Publicaciones y Fomento Editorial. Mexico.

Pérez-García, E.A. & Meave, J.A. 2006. Coexistence and divergence of tropical dry forests and savannas in Southern Mexico. *Journal of Biogeography* 33: 438–447.

Peterson, A.T., Sánchez-Cordero, V., Martínez-Meyer, E. & Navarro-Sigüenza, A.G. 2006. Tracking population extirpations via melding ecological niche modeling with land-cover information. *Ecological Modelling* 195: 229–236.

Pickett, S.T.A., Cadenasso, M.L. & Meiners, S.J. 2009. Ever since Clements: from succession to vegetation dynamics and understanding to intervention. *Applied Vegetation Science* 12: 9–21.

Prasad, A.M., Iverson, L.R. & Liaw, A. 2006. Newer classification and regression tree techniques: bagging and random forests for ecological prediction. *Ecosystems* 9: 181–199.

Puig, H. 1972. *Memorias I Congreso Latinoamericano de Botánica*. La sabana de Huimanguillo, Tabasco, M é xico, M é xico, DF pp. 389–411.

Rogan, J. & Chen, D. 2004. Remote sensing technology for mapping and monitoring land-cover and land-use change. *Progress in Planning* 61: 301–325.

Rogan, J., Franklin, J., Stow, D., Miller, J., Woodcock, C. & Roberts, D. 2008. Mapping land-cover modifications over large areas: a comparison of machine learning algorithms. *Remote Sensing of Environment* 112: 2272–2283.

Rosenfield, G.H. & Fitzpatrick-Lins, K. 1986. A coefficient of agreement as a measure of thematic classification accuracy. *Photogrammetric Engineering and Remote Sensing* 52: 223–227.

Rouget, M. 2003. Measuring conservation value at fine and broad scales: implications for a diverse and fragmented region, the Agulhas Plain. *Biological Conservation* 112: 217–232.

Rzedowski, J. 1978. *Vegetación de México*. Editorial Limusa S.A., México.

Rzedowski, J. 1990. *Vegetación Potencial*. Atlas Nacional de México. Vol. 2. Scale 1:4 000 000. Instituto de Geografía. Universidad Nacional Autónoma de México (UNAM).

Sarukhán, K.J. 1968. Los tipos de vegetación arbórea de la zona cálido-húmeda de México. In: Pennington, T.D. & Sarukhán, J. (eds.) *Manual para la identificación de los principales árboles tropicales de México*. pp. 3–46. Instituto Nacional de Investigaciones Forestales & Food and Agriculture Organization (FAO), Mexico.

Secretaría de Medio Ambiente y Recursos Naturales (SEMARNAT). 1998. *Mapa de suelos dominantes de la República Mexicana*. Scale 1:4 000 000. Mexico.

Shi, T. & Horvath, S. 2006. Unsupervised learning with random forest predictors. *Journal of Computational and Graphical Statistics* 15: 118–138.

Sing, T., Sander, O., Beerenwinkel, N. & Lengauer, T. 2005. ROCR: visualizing classifier performance in R. *Bioinformatics* 21: 3940.

Siroky, D.S. 2009. Navigating Random Forests and related advances in algorithmic modeling. *Statistics Surveys* 3: 147–163.

Stoms, D.M. 1992. Effects of habitat map generalization in biodiversity assessment. *Photogrammetric Engineering and Remote Sensing* 58: 1587–1591.

Strandman, H., Väisänen, H. & Kellomäki, S. 1993. A procedure for generating synthetic weather records in conjunction of climatic scenario for modelling of ecological impacts of changing climate in boreal conditions. *Ecological Modelling* 70: 195–220.

Strobl, C., Boulesteix, A., Kneib, T., Augustin, T. & Zeileis, A. 2008. Conditional variable importance for random forests. *BMC Bioinformatics* 9: 307–317.

Toledo, V.M. & Ordóñez, M.de J. 2009. *Zonas ecológicas de México. Proyectos A006 y E021: "Diagnostico de los escenarios de la biodiversidad en México", fases 1 y 2. Scale 1:1 000 000*. Centro de Ecología, UNAM, Mexico.

Trejo, I. & Dirzo, R. 2000. Deforestation of seasonally dry tropical forest: a national and local analysis in Mexico. *Biological Conservation* 94: 133–142.

van Etten, E.J.B. 1998. Mapping vegetation in an arid, mountainous region of Western Australia. *Applied Vegetation Science* 1: 189–200.

Villers-Ruiz, L. & Trejo-Vázquez, I. 1997. Assessment of the vulnerability of forest ecosystems to climate change in Mexico. *Climate Research* 9: 87–93.

Villers-Ruiz, L. & Trejo-Vázquez, I. 2003. El cambio climático y la vegetación en México. In: Gay Garca, C. (ed.) *México: una visión hacia el siglo XXI. El cambio climático en México*. pp. 57–72. SEMARNAT, UNAM, U.S. Country Studies Program, Mexico, DF.

Vrac, M., Paillard, D. & Naveau, P. 2007. Non-linear statistical downscaling of present and LGM precipitation and temperatures over Europe. *Climate of the Past Discussions* 3: 899–933.

White, P.S. 1979. Pattern, process, and natural disturbance in vegetation. *The Botanical Review* 45: 229–299.

Yates, D.N., Kittel, T.G. & Cannon, R.F. 2000. Comparing the correlative Holdridge model to mechanistic biogeographical models for assessing vegetation distribution response to climate change. *Climatic Change* 44: 59–87.

Zerbe, S. 1998. Potential natural vegetation: validity and applicability in landscape planning and nature conservation. *Applied Vegetation Science* 1: 165–172.

Zweig, M.H. & Campbell, G. 1993. Receiver-operating characteristics (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clinical Chemistry* 39: 561–577.