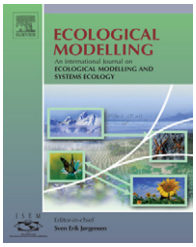
ECOLOGICAL MODELLING XXX (2007) XXX–XXX

available at www.sciencedirect.com

ScienceDirect

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



Predicting geographical distribution models of high-value timber trees in the Amazon Basin using remotely sensed data

Cássia Da Conceição Prates-Clark[a,b,*], *Sassan S. Saatchi*[b,c], *Donat Agosti*[d]

[a] *School of Geographical Sciences, University of Bristol, University Road, Bristol, BS8 1SS, UK*
[b] *Center for Tropical Research, University of California, Los Angeles, CA 90095, USA*
[c] *Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109, USA*
[d] *Naturmuseum der Burgergemeinde Bern, 3005 Bern, Switzerland*

ARTICLE INFO

Article history:
Received 13 May 2006
Received in revised form
17 September 2007
Accepted 21 September 2007

Keywords:
Amazon Basin
Climate variables
Ecological niche
Remote sensing data
Species distribution model
Timber trees

ABSTRACT

Species distribution models were developed for three high economic value timber trees (*Calophyllum brasiliensis*, *Carapa guianensis and Virola surinamensis*) that are heavily harvested in the Amazon Basin. A combination of habitat measurements extracted from remote sensing data (MODIS, QSCAT and SRTM) and bioclimatic surfaces was examined to ascertain the most influential factors determining the occurrence of these tree species. The prediction of species' occurrence rates was tested separately for each species distribution model and the results were examined for their ability to accurately map the spatial distribution of these tree species. By evaluating the omission and commission rates we concluded that species distribution models based on remote sensing data contributed significantly in quantifying environmental properties used to summarize the ecological niche of each tree species. Specific vegetation characteristics (such as percentage of tree cover, vegetation moisture and roughness, annual NDVI and mean LAI during the dry LAI) showed the dependence of these species' occurrence in more densely vegetated forests. Areas with high leaf area (even during the dry months) and areas with high vegetation moisture were predicted as potential species habitat for *C. brasiliensis*. The density vegetation during the dry season and vegetation phenology were strongly correlated with climate differences, such as variations in air temperature and precipitation seasonality for *V. surinamensis*. Lower elevation areas with more exuberant vegetation and a high greenness index were among the most important factors accounting for the geographical distribution of *C. guianensis*. Species distribution models are increasingly important in many fields of research and conservation. The potential of remotely sensed data to monitor environmental changes in tropical areas, along with the understanding of ecosystem function, are both critical for conservation of biodiversity and the long-term process of sustaining ecosystems.

---

[*] *Corresponding author at*: School of Geographical Sciences, University of Bristol, University Road, Bristol, BS8 1SS, UK.
E-mail address: Cassia.Prates-Clark@bristol.ac.uk (C.D.C. Prates-Clark).

0304-3800/$ – see front matter
doi:10.1016/j.ecolmodel.2007.09.024

1. Introduction

Selective logging is the process where a limited number of specifically targeted, marketable timber trees (rather than all trees within a specific logged area), are cut down (Veríssimo et al., 1992). However, little is known about the extent and the impact of selective logging in the world's tropical forests, or the impact of such logging on the environment. Selective logging causes widespread collateral damage to the trees that remain, to the sub-canopy vegetation and the surrounding soil—all of which, in turn, has an impact on hydrological processes, soil erosion, the heightened probability of forest fires, carbon storage and, of course, on forest-based animals and plants (Uhl and Vieira, 1989; Nepstad et al., 1999; Asner et al., 2005).

In tropical rainforests, economically valuable timber trees can be diverse and abundant (Connell, 1978; Nepstad et al., 2004). Tropical forests in the Amazon Basin contain a large number and a wide range of commercially valuable trees, such as those from the Clusiaceae, Meliacaeae, Myristicaceae and Bignoniaceae, families (Mori and Prance, 1990b). Because of their high economic value, such timber is often heavily harvested and, consequently, the more valuable tree species are becoming increasingly scarce in those areas which have been most heavily logged. For this reason, particular species of timber trees often have been earmarked for special attention in conservation and forest management practices (Carvalho et al., 2002; Nepstad et al., 2004).

To date, apart from maps of species' occurrences obtained from field guide information (de Oliveira and Mori, 1999; Fisch and dos Santos, 2001; Vinson et al., 2005), there has been no research conducted into the systematic distribution ranges of timber trees with high economically valuable. The development of predictive distribution models for these trees may help researchers answer questions such as how and why timber trees are spatially distributed, which factors limit their dynamic ranges, and how these species are likely to respond to both human disturbances and global environmental changes. Timber trees, due their high economic value, also serve as drivers for deforestation and logging. Predicting the geographical ranges of such tree species can therefore help researchers determine which areas are most likely to be exploited at some future date. Information about the geographical distribution and abundance of such species would also help prevent subsequent logging by locating areas to be protected or conserved—information that would be of obvious benefit to governmental and non-governmental institutions such as the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES), the Forest Stewartship Council (FSC), and the World Wildlife Fund's (WWF) timber certification, which promotes programs to conserve the world's forests and biodiversity.

As a consequence of the aforementioned points, the objective of this study was to answer the following questions: (1) how widely distributed are three high economic value timber trees (*Calophyllum brasiliensis*, *Carapa guianensis* and *Virola surinamensis*) across the Amazon Basin? (2) Which biophysical attributes (derived from satellite imagery) comprise the ecological niche of these tree species? (3) Which ecological (and other geographical) factors have an impact on the geographical distribution of these three tree species? In order to begin to answer these questions we first developed predictive distribution models for each of the three tree species we studied based on the estimated probability of species' occurrence and then we analysed the results of these models. We also evaluated the accuracy of these models using threshold-independent measures and re-sampling techniques, such as bootstrap. Subsequently, we investigated the historical processes that may explain evident differences between the 'potential' and 'true' geographical ranges (Soberón and Peterson, 2005) of these tree species. Finally, we discuss how environmental changes such as forest loss due to deforestation might affect their occurrence.

2. Methods

2.1. Study species

Three commercial timber tree species *C. brasiliensis* (Clusiaceae), *C. guianensis* (Meliaceae) and *V. surinamensis* (Myristicaceae) from the herbarium collection of the 2000–2002 New York Botanical Gardens were used to generate the species distribution models. Herbarium specimens represent primary occurrence data used to approximate the geographical distribution of a species. Total number of point localities used for each tree species was: 81 for *C. brasiliensis*, 79 for *C. guianensis* and 125 for *V. surinamensis*. (No duplicate point localities were included in the models.) Generally, more accurate predictions can be made with a larger number of point localities (Guisan et al., 2007). However, significant model degradation might occur due to error on georeferencing data points (Engler et al., 2004). Hence, it is essential to evaluate the trade-off between the number of occurrence samples available at herbarium or atlas museums for the modelling and the location accuracy of data.

Habitats where these trees were observed (Fig. 1) ranged from lowland to highland forests, a variety of riverine forests, and forests ranging from secondary succession to old growth. *C. brasiliensis* (jacareúba) is a canopy tree typical of the humid tropical forests that extend from southern Mexico to the northern parts of South America (Record and Hess, 1943). It can be found from sea level up to 1500 m altitude in areas with an annual rainfall above 3000 mm and a temperature of 24–28 °C (Marques and Joly, 2000; Fisch and dos Santos, 2001). This species frequently grows in evergreen lowlands, mountain forests, gallery forests and sloped areas with alluvial or clay soils, even when these areas are very humid such as in the Amazon 'varzeas' (i.e. seasonally flooded lowland forests) (King, 2003). Typically, *C. brasiliensis* is 40–50 m in height, 1.80 m diameter at breast height (dbh) and has a straight, cylindrical trunk with a dense crown and irregular branches. This species produces fruits annually, which mature between October and November. *C. brasiliensis* is one of the most important of the Latin America tree species and has been heavily harvested because of the quality of its wood which is often used as an alternative to mahogany (*Swietenia mahagoni*) (King, 2003). The wood is heavy, with a specific weight of 0.45–0.69 g/cm^3 (Carpenter et al., 2004).

C. guianensis (andiroba) is a climax tree that occurs in the West Indies, through Central America down to the Amazonian

Please cite this article in press as: Prates-Clark, C.C., et al., Predicting geographical distribution models of high-value timber trees in the Amazon Basin using remotely sensed data, Ecol. Model. (2007), doi:10.1016/j.ecolmodel.2007.09.024

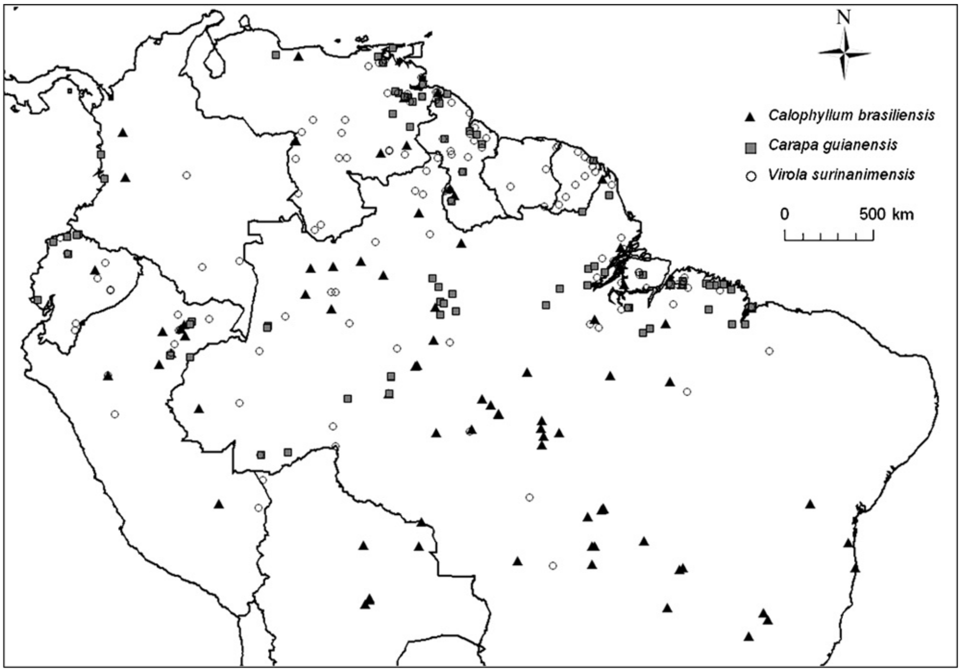


Fig. 1 – Data points of species' occurrence for *Calophyllum brasiliensis*, *Carapa guianensis* and *Virola surinamensis*.

rainforest including the overflow delta lands of the Orinoco in Venezuela (Hall et al., 1994; Forget, 1996). This tree species commonly establishes itself on rich soils, along streams and in the periodically inundated swamp forests and uplands forests along the rivers of the Amazon Basin (Guariguata and Dupuy, 1997; Guariguata et al., 2000). It is also cultivated in Brazil in the Islands region, Tocantins, Rio Solimões and near shorelines (Vianna, 1982). It has been observed in altitudes ranging from 0 to 1200 m, in areas with mean annual rainfall 1500–3200 mm and a mean annual temperature of 24–26 °C (Hall et al., 1994). *C. guianensis* is a tall rainforest tree that grows approximately 40 m in high and has a diameter of up to 1.20 m. It is one of the large-leafed trees of the rainforest and can be identified by its large and distinctly textured leaves (Guariguata et al., 2002). The main stems from *C. guianensis* are straight with low ramparts (Hall et al., 1994) and its timber is often exported for use in fine-wood products as it is in the same family as mahogany and has even been called the 'Brazilian mahogany' (Vinson et al., 2005).

The geographical distribution of *V. surinamensis* extends from the Lesser Antilles throughout northern South America down to Bolivia and around Northeast Brazil (Howe et al., 1985; Fisher et al., 1991). This species grows on humid habitats (average annual precipitation is 2500 mm), from sea level to 1100 m altitude, with a mean monthly temperature ranging from 26 to 27 °C. *V. surinamensis* (ucuúba) prefers swampy, fertile habitats, periodically flooded riverbanks that occur in varzea forests along the Amazon River and its muddy tributaries (Lopes et al., 2004). It generally occurs in open forest gaps (Howe, 1990). This tree species grows slowly until it attains 25–35 cm of dbh, after which it accelerates considerably as, at this dbh, the crown is probably finding its place in the sun (Rodrigues, 1980). *V. surinamensis* flowers twice a year with fruiting in May–June and November–December (Bena, 1960). *V. surinamensis* is also an economically high-value species and has been harvested because of its resistant timber and pulp paper purposes (Howe, 1990; Lopes et al., 2004).

The development of species distribution models requires species point localities and environmental data layers containing the possible range of environmental properties (e.g. surface elevation, percent tree cover, air temperature, rainfall) that provide the potential species' habit requirements. The environmental layers used to model the potential distribution of *C. brasiliensis*, *C. guianensis* and *V. surinamensis* are described below.

2.2. *Remote sensing environmental layers*

Remote sensing data represent a source of direct measurements of environmental variables that can be used in ecological studies. The environmental data layers extracted from remote sensed data used in this study cover a range of vegetation and landscape variables such as topography, vegetation greenness, leaf area, moisture, roughness, and seasonality. These variables are listed in Table 1.

2.2.1. *Vegetation type, seasonality and productivity*

These layers provide a measure of above-ground biomass, vegetation cover types, seasonal changes in the greenness of vegetation and vegetation canopy characteristics estimated from remote sensing data from Moderate Resolution Imaging Spectroradiometeter (MODIS) satellite on board the Terra National Aeronautics and Space Administration (NASA) spacecraft. The monthly (2000–2004) MODIS data were converted into annual normalized difference of vegetation index (NDVI) and leaf area index (LAI) metrics. In ecological studies, NDVI data have been used to estimate leaf biomass and the net primary productivity because NDVI values are associated

Please cite this article in press as: Prates-Clark, C.C., et al., Predicting geographical distribution models of high-value timber trees in the Amazon Basin using remotely sensed data, Ecol. Model. (2007), doi:10.1016/j.ecolmodel.2007.09.024

Table 1 – List of remote sensed data and products used in this study

Data record	Instrument	Vegetation/landscape parameter	RS metrics at 5 km resolution
Monthly NDVI (2000–2001)	MODIS	Vegetation type and seasonality	NDVI-1: maximum NDVI NDVI-2: annual mean NDVI NDVI-3: mean NDVI wet months NDVI-4: mean NDVI dry months
Monthly (2000–2001) leaf area index (LAI)	MODIS	Vegetation type, seasonality, productivity	LAI-1: maximum LAI LAI-2: annual mean LAI LAI-3: mean LAI wet months LAI-4: mean LAI dry months
Percent tree cover (2000–2001)	MODIS	Forest cover and heterogeneity	VCF: continuous field product
Scatterometer backscatter monthly composites at 1 km (1999–2004)	QuickSCAT	Vegetation moisture, leaf/wood density	QSCAT-H: mean backscatter HH QSCAT-V: mean backscatter VV QSCAT-SH: std. backscatter HH QSCAT-SV: std. backscatter VV
Digital elevation (100 m resolution) 2000	SRTM	Surface elevation	SRTM-HGT: mean elevation SRTM-STD: ruggedness factor

to the photosynthetically active radiation of plant canopies (Tucker, 1979; Tucker et al., 1983). NDVI metrics also have provided valuable information about vegetation phenology because NDVI values are influenced by the loss of vegetation due to relative seasonal changes in vegetation (Boone et al., 2000). LAI data layers provide a measure of foliage density of vegetation and structure of the vegetation canopy (Myneni et al., 1997, 2002). Each LAI layer was converted into a grid cell at the same spatial resolution, i.e. 5 km^2. The resulting NDVI layers show the vegetation greenness values within each 5 km pixel (range −1 to 1).

2.2.2. *Forest cover*

Forest cover and heterogeneity information was obtained from the MODIS satellite 500 m global vegetation continuous fields (VCF) data set (Hansen et al., 2000, 2002). This layer includes proportional estimates of forest canopy cover derived from a global data set collected between 2000 and 2001 (DeFries et al., 2002). The MODIS VCF Percent Tree Cover was re-sampled to a 5 km^2 spatial resolution to provide both an estimate of forest cover and a threshold to separate areas of dense from sparse deforested and fragmented vegetation.

2.2.3. *Vegetation moisture leaf/wood density*

Vegetation surface properties (such as moisture content, leaf size and branch orientation) were obtained from monthly composite scatterometer image data (i.e. QuickSCAT), The SeaWinds scatterometer aboard the QuickSCAT satellite provides near-daily global coverage of the Earth at intrinsic resolutions generally between 25 and 50 km^2, over incidence angles ranging from 20 to 55°, making both vertically and horizontally polarized measurements. A scatterometer transmits radar pulses and receives backscatter data that is less sensitive to atmospheric effects than the passive microwave sensors (Lin, 1998; Long and Drinkwater, 2000). As the intensity of the backscattering depends on the roughness and dielectric properties of a particular target (Long et al., 2001), if the surface being studied is the vegetation cover, preferential alignment of surface scatters, canopy surface roughness, moisture content, leaf size and density will influence the intensity of the backscattering. Although, the QuickSCAT layers used in this study are based on monthly composite data at a 1 km^2 spatial resolution obtained during 1999–2004 period at dual polarization (Long et al., 2001), these layers were also aggregated to a 5 km^2.

2.2.4. *Surface elevation*

This data layer provides topography and standard deviation surface elevation as an index of ruggedness. Elevation data on a near global scale were obtained from the Shuttle Radar Topography Mission (SRTM; NASA, 2006). Using the median value, the high-quality and high spatial resolution provided by SRTM (i.e. 90 m resolution digital elevation; Jarvis et al., 2004) was aggregated to a 5 km spatial resolution to maintain consistency with other environmental layers.

2.3. *Climate data*

The climate data set used for this study was obtained from the free access web site at http://www.wordclim.org. The set of global climate layers (i.e. WorldClim) includes annual time-series with mean monthly data for precipitation, and mean, minimum and maximum temperatures obtained from over 4000 weather stations between 1950 and 2000 (Hijmans et al., 2005). This data set, which comprises climate data from a number of spatial resolutions, was spatially interpolated with latitude, longitude and surface elevation to produce climate data on grids. A number of different statistical approaches have been used to generate interpolated climate surfaces (New et al., 1999, 2002; Daly et al., 2002). Compared to other global climate data sets, WordClim has the advantage of providing climate data from a larger number of stations. WordClim uses an algorithm that considers every station (instead a subset of stations) as a data point. Duplicate records were removed since only station records further than 5 km away from stations already in the data set was included. A second-order spline with latitude, longitude and elevation data at finer spatial resolution (i.e. SRTM) as independent variables was used to fit the model. As a result, the lowest overall cross-validation errors compared to other settings (i.e. third-order

Please cite this article in press as: Prates-Clark, C.C., et al., Predicting geographical distribution models of high-value timber trees in the Amazon Basin using remotely sensed data, Ecol. Model. (2007), doi:10.1016/j.ecolmodel.2007.09.024

spline or elevation as a covariate) were achieved (Hijmans et al., 2005).

To maintain consistency with the RS-data sets, each climate original layer obtained from WordClim was gridded to the 5 km spatial resolution data layers used in this study. The 19 bioclimatic variables used in this study are those one produced by Hijmans et al. (2005), i.e.: 'annual mean temperature', 'mean diurnal range', 'isothermality', 'temperature seasonality', 'maximum temperature of warmest month', 'minimum temperature of coldest month', 'temperature annual range', 'mean temperature of wettest quarter', 'mean temperature of driest quarter', 'mean temperature of warmest quarter', 'mean temperature of coldest quarter', 'annual precipitation', 'precipitation of wettest month', 'precipitation of driest month', 'precipitation seasonality', 'precipitation of wettest quarter', 'quarter precipitation of driest quarter', 'precipitation of warmest quarter', 'precipitation of coldest quarter'.

2.4. *Model generation*

For each species, the 'Potential Species Distribution Model' (PSDM) was developed from a set of environmental variables for a set of grid cells, together with a set of data point locations where the species had been observed. The 'Maximum Entropy Approach' within the Maxent computer program for modelling species geographic distributions (v.2.1; Phillips et al., 2006) was employed to generate the PSDM for the three tree species studied.

Maxent is a general-purpose program that generates predictions (or inferences) from incomplete information. Maxent enables researchers to estimate a target probability distribution by finding the probability distribution of maximum entropy, subject to a set of constraints that represent the incomplete information about the target distribution (Della Pietra et al., 1997). When the maximum entropy approach is applied to species distribution models, the pixels with known species occurrence records from the study area (training data points) produce the region on which the probability of maximum entropy is defined. The available information about the target distribution often presents itself as a set of environmental variables. The environmental layers are also used to produce assertions which constrain the probability distribution that can be computed (Phillips et al., 2006).

The variable constraints are the expected values of each feature that should match its empirical average (i.e. the average value for a set of training points taken from the target distribution (Phillips et al., 2004). The predictive performance of the PSDM is influenced by the choice of feature types and regularization constants, with quadratic versions of variables being used to capture possible non-linear responses (Phillips et al., 2006).

The probability of species' occurrence is displayed in terms of "gain"—the log of the number of grid cells minus the log loss (i.e. the average of the negative log probabilities of the sample locations). The gain increases, iteration by iteration, either until the change from one iteration to the next falls below the intrinsic accuracy measure (convergence threshold) or until the maximum number of iterations have been performed (Anderson et al., 2003; Dudik et al., 2004).

Initially, each environmental variable used as input data was potentially an important predictor variable to develop the PSDM. As part of Maxent internal procedures, the 'jackknife test' re-sampling method (Efron, 1979; Peterson and Cohoon, 1999) was applied, both to reduce the bias of correlated environmental variables and to diagnose which environmental variables were the most important predictor variables to create the PSDM. The environmental variables with the highest gain, when used in isolation to generate the model, are displayed as bar charts. The longer the length of the bar charts the higher the relative importance of variables that, potentially, contribute to generating the species distribution model (Phillips et al., 2004).

The influence of each environmental variable on Maxent's prediction is displayed in response curve diagrams. As the Maxent model is an exponential model (Della Pietra et al., 1997), the probability assigned to a pixel is proportional to the exponential contribution of each environmental variable. The response curves show the Maxent exponent prediction changes as each environmental variable is varied, when all other environmental variables are maintained at their average sample value (Phillips et al., 2006).

2.5. *Model validation*

The predictive species distribution models were tested by using two different sets of data: (1) a data set for model building (i.e. the training data set), and (2) a data set for model validation (i.e. the test data set). Following Maxent procedures, two types of evaluation tests were internally applied: the threshold-*dependent* and the threshold-*independent* evaluation.

A low omission rate of species' occurrence is necessary for developing a potential model for predicting species' distribution ranges (Anderson et al., 2003). After applying a threshold, model performance can be investigated using both: (i) the "extrinsic omission rate" (i.e. the fraction of the test data points that fall into pixels not predicted as suitable for the species); (ii) the "proportional predicted area" (i.e. the fraction of all the pixels predicted as being suitable for the species). A one-tailed test (i.e. a tool for assessing whether the omission rate is either lower or higher than random; Tukey, 1977) was used to determine whether the model could significantly predict the test localities.

A second way in which model performance was compared was via the use of 'Receiver Operating Characteristics' (ROC) curves. ROC analysis is the standard approach used to evaluate the sensitivity and the specificity of diagnostic procedures (Swets and Pickett, 1992). Sensitivity (also known as the 'true positive rate') represents the absence of omission error, and the quantity 1-specificity (also known as the 'false positive rate') represents the commission error (Cantor et al., 1999). By tradition, the ROC curve is a representation of the trade-offs between the omission and commission error. This curve is obtained by plotting sensitivity on the y axis and 1-specificity on the x axis for all possible thresholds (Swets, 1988). If the ROC curve rises rapidly towards the upper left hand corner of the resulting graph, the sensitivity rate is high and the 1-specificity rate is low. If the ROC curve follows a diagonal path from the lower left hand corner to the upper right hand cor-

Please cite this article in press as: Prates-Clark, C.C., et al., Predicting geographical distribution models of high-value timber trees in the Amazon Basin using remotely sensed data, Ecol. Model. (2007), doi:10.1016/j.ecolmodel.2007.09.024

ner, every improvement in the 1-specificity rate is matched by a corresponding decline in the sensitivity rate.

The area under the ROC curve (AUC) represents a particularly important metric for evaluating diagnostic procedures because it represents the average sensitivity over all possible specificities. The AUC provides a single measure of model performance, independent of any particular choice of threshold value (Swets and Pickett, 1992). It is possible to quantify how quickly the ROC curve rises to the upper left corner of the graph by measuring the area under the curve. The larger the AUC, the higher is the sensitivity rate and the lower is the 1-specificity rate. An AUC equal to 1.0 represents an ideal diagnostic test because it achieves both perfect 100% sensitivity and 100% specificity. If the AUC is 0.5, the test has 50% sensitivity and 50% specificity which indicates that sensitivity (omission error) and 1-specificity (commission error) rates are high (Cantor et al., 1999; Liu et al., 2005).

When only presence data are available, it would appear that ROC curves are inapplicable since, without absences, there seems to be no source of negative instances via which a researcher is able to measure specificity. However, this problem can be circumvented by distinguishing a sample of positive instances (the presence localities) from a sample of pseudo-absence instances or background points (i.e. pixels localities where the species has not been found and selected uniformly at random from the entire geographical space covering the study area) (Stockwell and Peters, 1999; Graham et al., 2004). The total numbers of background points used for each species were: 10,061 for *C. brasiliensis*, 10,060 for *C. guianensis* and 10,094 points for *V. surinamensis*.

Table 2 – The potential predictor variables used to generate the RS and Climate distribution models

RS	Climate
Calophyllum brasiliensis	
LAI of driest quarter and its seasonality (LAI-4 and LAI-2)	Annual precipitation
Vegetation moisture and roughness (QSCAT-H)	Precipitation of the wettest month
Percent of tree cover	Precipitation of the wettest quarter
	Precipitation of coldest quarter
	Temperature seasonality
Carapa guianensis	
Elevation (SRTM)	Temperature annual range
Vegetation moisture and roughness (QSCAT-H)	Mean diurnal range
Maximum NDVI	Minimum temperature of coldest month
LAI of driest quarter (LAI-4)	Temperature seasonality
Virola surinamensis	
Elevation (SRTM)	Mean diurnal range
LAI of driest quarter (LAI-4)	Temperature seasonality
Canopy moisture and roughness (QSCAT-H, -V and -SV)	Annual precipitation
	Precipitation of the wettest month

3. Data analysis

The species distribution models for each of the three tree species studied were based on point localities (83, 64 and 88, respectively, *C. brasiliensis*, *C. guianensis* and *V. surinamensis*) as well as 34 environmental variables obtained from both bioclimatic surfaces and satellite imagery (Table 1).

For each tree species, three different PSDM were generated using three different suites of input data: bioclimatic surfaces, RS-data sets, and a combination of RS and bioclimatic variables. There were two principal reasons for treating these input data sets separately. First, in order to ascertain which data sets (i.e. bioclimatic and RS-data) would produce the best predictive species distribution model (i.e. the highest rates of model predictions where the species were sampled) and also to identify the environmental variables that comprises its species ecological niche. Second, in order to determine if a model with a better potential to predict species' occurrence could be developed by using a combination of the potential predictor bioclimatic and remote sensed variables selected from the previous two species distribution models.

The first model (henceforth, the 'Climate-model') was generated solely from only bioclimatic variables (i.e. the 19 climate variables described in Hijmanns et al., 2005). Subsequently, the re-sampling jackknife test was applied to identify the important predictor environmental variables used to generate this model. A second model (the 'RS-model') was generated using the environmental variables extracted from the remote sensed data set listed in Table 1; and the jackknife test within the Maxent package was applied to determine the potential predictor variables for generating this model. Another approach to map the geographical range of the tree species was to use those predictor variables that potentially predict species' occurrence from the previous two models as input data to generate a third PSDM ('RS and Climate-model'). Again, the jackknife test was applied to this model.

To assess whether the models' performance obtained for the three initial models could possibly be improved, we also considered the success rate of species prediction via a reduced input data set. The criteria for selecting a reduced number of environmental variables was based on the environmental variables that the jackknife test indicated as being the most influential predictor variables to determine the probability of species' occurrence in the three initial models. Because models for different species included different predictor variables, the set of new input data for which predictions could be made was not identical for the three tree species.

Maxent was used to plot the omission against threshold, the predicted area against threshold, the ROC curve and the AUC. The threshold-dependent binomial test (based on omission/commission error and predicted area) was also used to verify whether the model generated from the use of training and test data performed significantly better than random.

To investigate the appropriate size of the test sample to be used to evaluate each model's performance, 10 levels of test sample sizes (i.e. 5, 10, 15, 20, 25, 30, 35, 40, 45 and 50% of the sample records) were used as thresholds. If more points are dedicated to testing models, the number of points available for building the models become reduced and may affect the sta-

Please cite this article in press as: Prates-Clark, C.C., et al., Predicting geographical distribution models of high-value timber trees in the Amazon Basin using remotely sensed data, Ecol. Model. (2007), doi:10.1016/j.ecolmodel.2007.09.024

tistical power of the tests (Parra et al., 2004; Phillips et al., 2006). Even so, for each of these test sample sizes, one data set for training was created by randomly sampling a specified number of presences (i.e. point localities) and background points without replacement from the original presences and background points, respectively. Ten species distribution models were generated using each created training data set, and each of the resulting models was applied to the test data sets for the three suites of input data. This procedure was repeated three times—i.e. once for each tree species. For each set of input data, the adequate test sampling size was determined by models' evaluation, including *p*-values, AUC (by a single number) and fraction of predicted area.

To assess the potential of bioclimatic variables and remote sensed data for modelling the geographical distribution of species for each tree species, we compared separately the species distribution models derived solely from climate variables ('Climate-model'), RS-data sets ('RS-model'), and the combination of data of the two data sets ('RS and Climate-model'). The threshold-independent and threshold-dependent analyses were then applied to evaluate the performance of each model. Subsequently, the performance of the best models for each tree species produced was evaluated by extracting the cumulative probabilities at each point depicted as being the most appropriate habitat (i.e. where species were observed). Thresholds of 30, 40, 50, and 75% probability of species' occurrence were established and the number of species above a given threshold was calculated. This provided information which threshold made the greatest contribution to the model, and also indicated which model should be retained as the best potential predicting species distribution model. Finally, for each tree species, the model based on (i) the lowest omission/commission rates, (ii) highest AUC, (iii) higher percentage of probability of species' occurrence at point localities, and (iv) a set of predictor variables biologically meaningful to summarize the ecological niche of its species, was selected as the model that would most accurately predict the tree species' potential geographical distribution.

4. Results

4.1. Contribution of input data layers

Three approaches using different suite of input data ('RS', 'Climate', and 'RS and Climate') were used to generate potential predicting species' occurrence models for *C. brasiliensis*, *C. guianensis* and *V. surinamensis*. The comparisons between the RS-, Climate-, and RS and Climate-models' performance were based on both the threshold-dependent and threshold-independent tests and the entire point localities data for each species. Three indicators were used to evaluate the accuracy of the species distribution models: (i) balance threshold for optimizing omission and commission rates; (ii) the fraction of predicted area; (iii) the AUC.

The success rates for model predictions evaluated by AUC were relatively similar for the different suites of input data and were significantly better than a random prediction ($p < 0.001$; one tailed) for each tree species (Table 3). The model performance using bioclimatic variables produced higher AUC values when compared with species distribution models generated using RS-data sets for all tree species. However, the reductions in the omission and commission rates for RS-models were at a level of only 0.02%. The increase in AUC achieved by combining RS- and Climate data sets (i.e. RS and Climate-models) for *C. brasiliensis* and *V. surinamensis* was, again, not significantly higher than the AUC obtained for Climate- and RS-models—i.e. a 0.002–0.15% improvement when environmental features were combined (Table 3). In addition, for *C. guianensis*, the combined RS and climate data layers did not perform as strongly as individual data layers and produced significantly lower AUC when RS and climate layers were used independently. We attributed this result to the fact that the combined data sets had a large number of correlated and redundant data layers that had an adverse impact on the performance of Maxent.

Table 3 – Comparisons of 'RS', 'Climate' and 'RS and Climate' models performances for *Calophyllum brasiliensis*, *Carapa guianensis* and *Virola surinamensis*

	Balance threshold[a]	Fraction of predicted area	AUC
Calophyllum brasiliensis			
RS	1.501	0.888	0.728
Climate	2.007	0.858	0.751
RS and Climate	1.537	0.575	0.877
Carapa brasiliensis			
RS	0.500	0.678	0.880
Climate	4.504	0.440	0.898
RS and Climate	1.502	0.625	0.825
Virola surinamensis			
RS	2.001	0.624	0.855
Climate	1.250	0.625	0.852
RS and Climate	1.250	0.575	0.877

[a] Balance threshold was based on training omission rate and fraction of predicted area.

The Maxent predictions by comparing the balance threshold and the fraction of predicted area provided addition information on the performance of the models. The balance threshold minimizes a combination of training omission rate, cumulative threshold and the fraction of predicted area. A lower balance threshold refers to an overall better performance. For all tree species, the models' performances were significantly higher ($p < 0.001$) for RS-data sets. In contrast to the similarity of AUC values between models, values of the fraction of the predicted area were notably higher (sometimes 34% and 24% higher for, respectively, *C. brasiliensis* and *C. guianensis*; see Table 3) for RS-data sets when compared with Climate and RS and Climate data sets. For *C. brasiliensis* and *C. guianensis*, the balance threshold showed higher performance for RS-models (1.5 and 0.5, respectively) than for the climate variables (2.0 and 4.5, respectively). However, the combined RS and Climate data layers showed an improved result for *V. surinamensis*.

Following the Maxent evaluation, an independent test on model performance was undertaken. This involved extracting the predictive probabilities of species' occurrence from the pixels of point localities and estimated the percentage of the area where probabilities were higher of certain thresholds

Please cite this article in press as: Prates-Clark, C.C., et al., Predicting geographical distribution models of high-value timber trees in the Amazon Basin using remotely sensed data, Ecol. Model. (2007), doi:10.1016/j.ecolmodel.2007.09.024

Table 4 – Species' occurrence probabilities (%) on areas where the species have been observed

	RS	Climate	RS and Climate
Calophyllum brasiliensis			
Average	56.05	55.78	55.40
>30	74.68	76.83	83.54
>40	69.62	67.07	68.35
>50	60.76	57.32	55.70
>75	35.44	26.83	32.91
Carapa guianensis			
Average	55.56	55.44	52.72
>30	75.68	82.43	77.03
>40	67.57	70.27	64.86
>50	60.81	55.41	56.76
>75	25.68	21.62	21.62
Virola surinamensis			
Average	54.03	54.78	53.06
>30	74.38	71.90	76.03
>40	68.60	66.12	66.94
>50	60.33	56.20	58.68
>75	27.27	25.62	24.80

(e.g. 30, 40, 50, and 75%). These results are summarized in Table 4. For all three species, the RS-models performed better than both climate- and combined RS and climate-models, with 60% of point localities being found on the 50% threshold. As expected, the predictive probabilities decreased for high probability thresholds, suggesting that a smaller number of point localities were predicted by the model in higher probability classes. It appears that 50% was a reasonable threshold for evaluating model predictions for each tree species because the values in Maxent output are representative of a cumulative probability rather than a pure probability of occurrence (Phillips et al., 2006). This threshold was used to represent the spatial distribution of Maxent predictions.

4.2. *Data layers reduction*

The jackknife test results from Maxent for the model runs associated with RS and climate data were compared to examine the importance of environmental variables to generate the predictive species distribution models. The jackknife test provides gain values for two scenarios: (1) identifying the environmental variable with the highest gain when used in isolation to predict the distribution; (2) the environmental variable that decreases the gain when it is omitted. These two gain values highlight the most important environmental variables necessary for model development and allow their identification with no independent information.

Four precipitation variables (i.e. 'annual precipitation', 'precipitation of the wettest month', 'precipitation of the wettest quarter', 'precipitation of coldest quarter') had the highest gain for *C. brasiliensis* geographical distribution. This suggests that the annual precipitation of wettest and coldest periods were the main climatic variables for *C. brasiliensis* geographical distribution. Temperature seasonality, to some extent, also contributed to the geographical range of this tree species. However, it was not one of the top contenders. Remote sensing data layers had a more uniform gain distribution among input variables. Leaf area index and its seasonality (LAI-4), canopy moisture and architecture (QSCAT), and percent of tree cover (VCF) data were among the most important variables (Table 2).

The variables contributing to the distribution of *C. guianensis* (Table 2) were very different. Among climate variables, jackknife tests showed temperature variables ('temperature annual range', 'mean diurnal range', 'minimum temperature of coldest month', 'temperature seasonality') having the highest gain. Both diurnal and seasonal magnitude and ranges of temperature were important in identifying the species location. Precipitation layers had little impact on the Maxent results. Among remote sensing data, elevation (from SRTM) was the most important layer, followed by QSCAT, maximum NDVI, and LAI of driest quarter.

Both climate ('temperature annual range', 'mean diurnal range', 'minimum temperature of the coldest month', 'temperature seasonality') and remote sensing data layers mean elevation (SRTM-HGT), canopy moisture and roughness (QSCAT), LAI during the dry months (LAI-4) had a more uniform gain distribution among environmental variables for *V. surinamensis*. Diurnal and seasonal range temperature and the annual and wettest-month's precipitation showed the highest gain.

Following these analyses, model performances were compared when the potential predictor variables to generate models using RS and climate variables were used as input data, i.e. using a reduced number of input variables. The results of model simulations with reduced input data layers (i.e. the potential predictor variables selected to generated models using RS, Climate and RS and Climate data sets) are presented in Supplementary data. For all three species, comparison of AUC values from Table 3 (results from the entire data sets) and Supplementary data showed models degradation (lower AUC values) as less important variables were removed.

4.3. *Validating spatial distributions*

Validating Maxent prediction of species' occurrence requires an independent data set (Phillips et al., 2006). This validation was performed by selecting a percentage of the species point localities for training data and the remaining point localities for testing model performances. Three sample sizes of 15, 25 and 45% of the original point localities (selected randomly by Maxent) were used, and the results are presented in Supplementary data. While the test sampling size changed, the ranking of different suites of input data remained stable according to the fractional of predicted area, but varied according to the sensitivity and specificity (AUC). Also, there was a trend for the three species with respect to the most adequate test sampling size. The one-tailed binomial probability test (based on the omission of test points and the fraction of predicted area) was significantly higher for the testing sample of 25% of data points. At the 25% level, the fraction of predicted area covered was much higher than other levels of test sampling (i.e. 5, 10, and higher than 35%), which, often, was not even statistically significant. In addition, it can be noticed that the test sample size of 25% of the point localities produced the best omission and commission rates (Supplementary data) which is close to values obtained when all training data sets were used to generate

Please cite this article in press as: Prates-Clark, C.C., et al., Predicting geographical distribution models of high-value timber trees in the Amazon Basin using remotely sensed data, Ecol. Model. (2007), doi:10.1016/j.ecolmodel.2007.09.024

the species distribution models (Table 3). This implies that for all tree species, overall model performance was significantly higher ($p < 0.001$; one tailed) with test points being classified as 'presence' more than expected by random, given the proportion of pixels predicted 'present' by the species distribution model.

Although the models' performance for testing size of 25% were similar for all tree species, the fraction of predicted area was much lower at that testing level for *C. guianensis* (Supplementary data). This probably was due to the small number of point localities (64 points) used to model this tree species' occurrence. According to Phillips et al. (2006), an optimal species distribution model would incorporate training data from all available records of the species (e.g. models performance presented in Table 2) and, when the number of point localities is insufficient, it may reduce the species distribution models' performance.

4.4. *Species distribution maps*

Model predictions for the three species were mapped in terms of the cumulative probability over the region of study at a 5 km spatial resolution. The estimated probabilities of occurrence were produced as continuous predictions with values ranging from 0 to 100. To generate the maps, six colours were used to illustrate the strength of each individual map pixel. One colour for probability greater than 50 was used to indicate the most reliable range of species' occurrence. However, the colours of lower probabilities assisted to compare maps with different input variables. For a biological interpretation of results, we focused primarily on areas that had a probability of greater than 50, and assumed that the lower probability areas were less significant in defining the ecological niche of these tree species.

Only the broad patterns of *C. brasiliensis*, *C. guianensis* and *V. surinamensis* could reliably be extracted from the Climate-driven models (i.e. the species distribution models generated using the 19 bioclimatic layers). Bioclimatic variables have a very coarse spatial resolution as they were produced from available distribution of precipitation and rainfall data over the region and interpolated using SRTM elevation data. Regardless of the high resolution of SRTM data, the interpolated climate variables are limited by their original spatial resolution. Using a combination of bioclimatic and RS-data sets to generate species distribution models (RS and Climate-models) produced species distribution maps that were even more difficult to interpret in biological terms. The spatial resolution of these distribution maps was degraded and landscape features were again neglected. The approach based on the usage of RS-data sets to generate the species distribution models produced species distribution maps containing more detailed information about the landscape and vegetation patterns in the Amazon Basin, generally without sharp differences between adjacent areas due to the smoothly varying fine-scale of environmental variables extracted from RS-data. The natural resolution of these data layers varied from 100 m to 1 km and, hence, they were more suitable for capturing the regional (and sometimes local) scale variability of the landscape, even at the 5 km spatial resolution maps generated by the models.

The three species distribution maps for *C. brasiliensis*, *C. guianensis* and *V. surinamensis* are presented in Figs. 2–4,

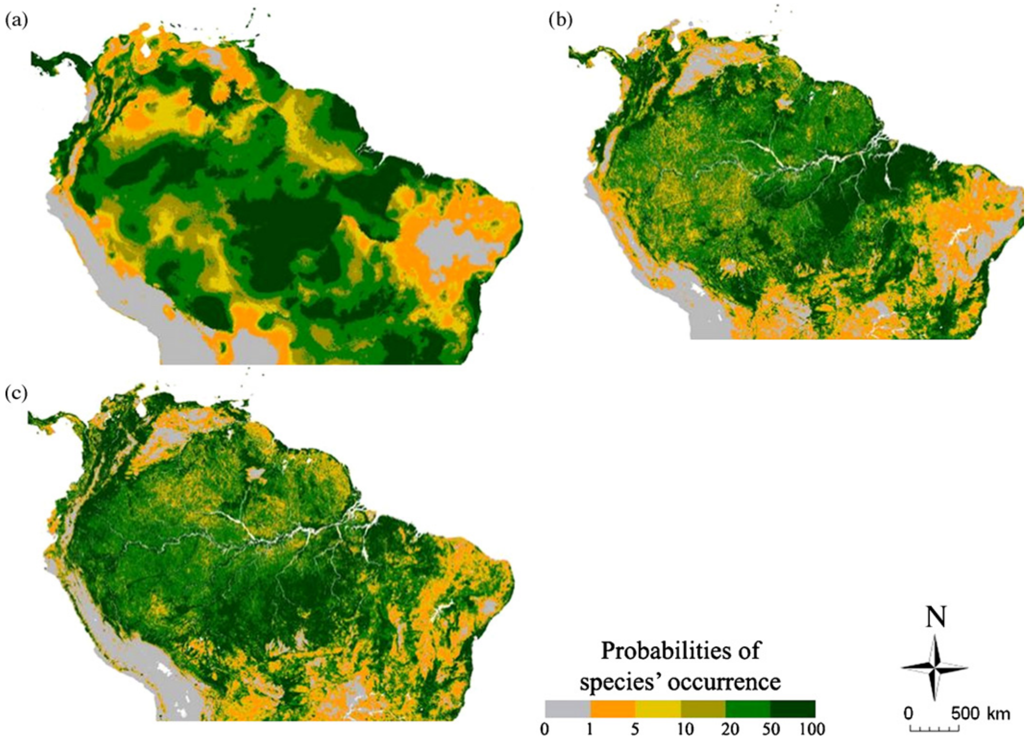


Fig. 2 – The predictive potential geographical maps for *Calophyllum brasiliensis* based on: (a) 'Climate-model', (b) 'RS-model' and (c) 'RS and Climate'.

Please cite this article in press as: Prates-Clark, C.C., et al., Predicting geographical distribution models of high-value timber trees in the Amazon Basin using remotely sensed data, Ecol. Model. (2007), doi:10.1016/j.ecolmodel.2007.09.024

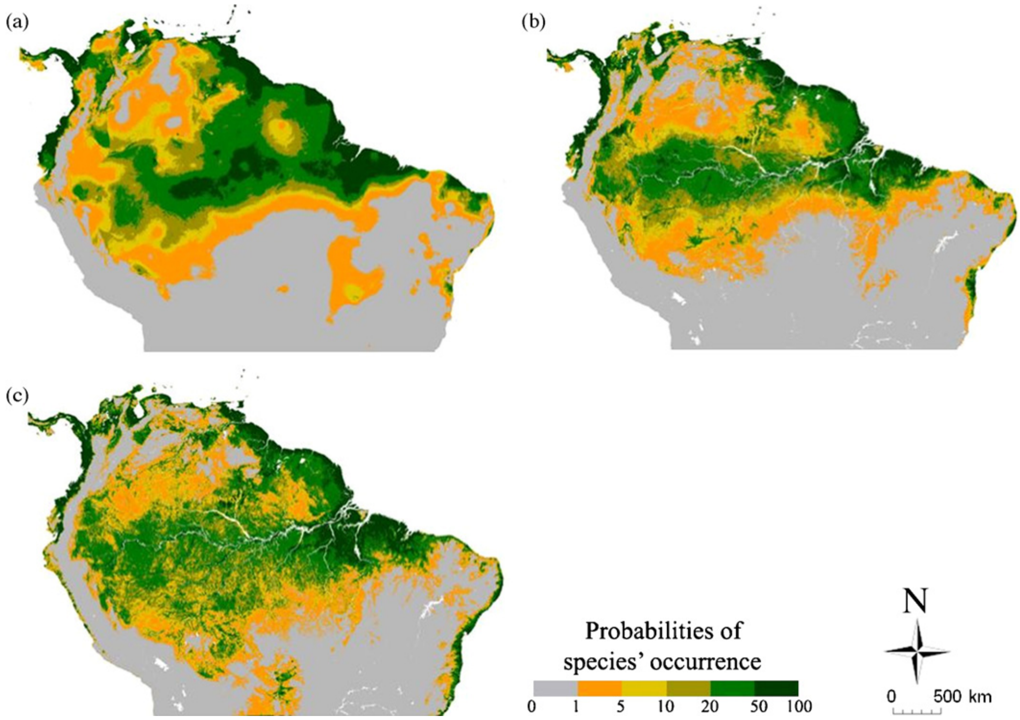


Fig. 3 – The predictive potential geographical maps for *Carapa guianensis* based on: (a) 'Climate-model', (b) 'RS-model' and (c) 'RS and Climate'.

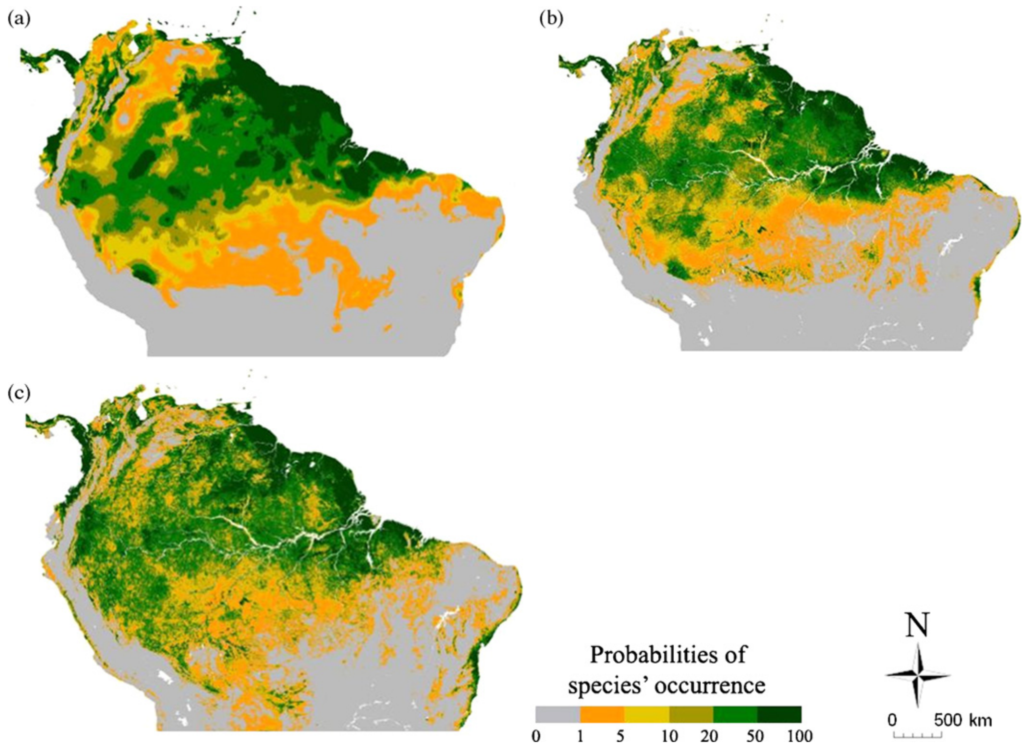


Fig. 4 – The predictive potential geographical maps for *Virola surinamensis* based on: (a) 'Climate-model', (b) 'RS-model' and (c) 'RS and Climate'.

Please cite this article in press as: Prates-Clark, C.C., et al., Predicting geographical distribution models of high-value timber trees in the Amazon Basin using remotely sensed data, Ecol. Model. (2007), doi:10.1016/j.ecolmodel.2007.09.024

respectively. Each figure included three panels for Climate, RS and combined RS and Climate species distribution maps.

5. Discussion

The species distribution models generated using 'Climate', 'RS' and, 'RS and Climate' variables produced different results, with the RS-models showing higher model performance to predict species' occurrence across the landscape as well as indicating the set of environmental variables that summarize each tree species ecological niche.

For each species (except *C. brasiliensis*), the results obtained for omission and commission rates were lower for RS-models. The potential distribution model for *C. brasiliensis* had a relatively lower statistical power (Table 3) compared to the other two tree species' models. Furthermore, the distribution map for *C. brasiliensis* failed to cover a possible true geographical distribution. This created some overprediction—e.g. in some deforested areas in the Brazilian Amazon both of which are areas where this tree species is highly unlikely to be present. The high rates of commission error (overprediction) for *C. brasiliensis* may be explained by inadequate sampling (Karl et al., 2000, 2002). Sample data points are clearly an integral part of species distribution models. However, inherent error margins of data might occur since the exact location of the collecting site is not very precise. In addition, widespread species such as *C. brasiliensis* are often associated with intrinsic and extrinsic commission error due to the relatively moderate number of point localities (for a widespread species) used to develop the species distribution model (Peterson and Viglais, 2001; Anderson and Martínez-Meyer, 2004). One possible way of avoiding some of the species distribution model's discrepancies would be to use a larger number of data points. Although the difference between models was relatively small when compared with the area under the Receiver Operating Characteristic curves (Table 3), RS-models showed a higher performance than Climate- and RS and Climate-models when evaluating values of probabilities of species' occurrence in point localities which the species was found (Table 4). The RS-models retained a high precision of over 60% in correctly predicting the species' occurrence for all three species. It was noticeable that that the addition of bioclimatic layers did not increase the models' accuracy. Indeed, the species distribution maps based on RS and Climate-models were even more difficult to interpret in biological terms. This implies that data sets at a very coarse spatial resolution (e.g. bioclimatic surfaces) should not be analysed (or even processed) together with RS-data sets (at finer resolution) in the same model. For *C. guianensis*, the model developed from the combined RS and climate layers removed some of the detailed landscape features obtained by the RS-model. The combined model suggested that for certain species such as *C. guianensis*, the combination of environmental variables at different spatial resolutions might cause the loss of important features that describe the ecological niche of this species. For this reason, bioclimatic variables are only important for general characteristics and should be treated with caution when employed in distribution models.

It is also important to point out that remote sensing data may be correlated with climate variables. For example, the RS-model is interrelated with variables used to develop the Climate-model. In the context of vegetation stress, Maximum Temperature or Potential Evapotranspiration (PET) is an indication of a favourable environment for photosynthesis since photosynthesis, of course, requires heat and light. Likewise, due to the minimum amount of precipitation, during dry years, the extended length of the dry season and amount of radiation, there will be less evapotranspiration in tropical rainforests (like those in the Amazon Basin; Shabanov et al., 2002). This, in turn, may generate a decrease in photosynthesis which can be captured in the thermal bands used to generate LAI and NDVI metrics. Also, since moisture condition may restrict plant growth and vegetation phenology, LAI and NDVI metrics may show temporal variations related to seasonal changes of moisture conditions (Myneni et al., 1995, 1997). Furthermore, a change in vegetation density affects the vegetation surface roughness, and vegetation surface roughness also depends on the LAI (Chase et al., 1996; Myneni et al., 1998).

Knowledge of meso-scale (climate) patterns is also important when attempting to model species' geographical ranges since climate parameters have an impact on plant and animal growth (Hubbell et al., 1999; O'Brien et al., 2000; Pausas et al., 2003). In our analysis, an interdependence link between environmental variables extracted from RS-data sets and climate variables was also evident. The potential geographical distribution map for *C. brasiliensis* based on the RS-model (Fig. 2b) shows some overlap with the distribution map generated using the Climate-model. However, the distribution map based on the RS-model incorporates entirely new areas (e.g. forest areas in Colombia and Ecuador) and excluded others areas such as deforested lands in North Para (Brazilian Amazon) and open savannas in Venezuela. For *V. surinamensis*, 'temperature seasonality' (Climate-model), LAI-5, LAI-7 and QSCAT-H (RS-model) were the key predictor variables for generating the models. Environmental variables (such as LAI and NDVI) have been recognized to be strongly conditioned by the behaviour of precipitation, air temperature and daily radiation the observed area (Shabanov et al., 2002). These metrics, which are sensitive to the length of dry season, assume that the potential of species' occurrence may decline with the increasing harshness of environmental conditions. This is deemed by researchers to be important, particularly for predicting tree species ranges. For example, fewer species should physiologically be equipped to tolerate annual (i.e. short-term) variations in climate conditions (Fraser and Currie, 1996). Although *V. surinamensis* is found in shade (Lopes et al., 2004) and forest gaps (Whitmore, 1988), Howe (1990) has shown that this tree species is vulnerable to the seasonal droughts occurring in forest gaps. Thus, deforested areas or seasonal vegetated regions were successfully captured by LAI metrics and consequently excluded from the distribution map presented in Fig. 4b for *V. surinamensis*. This finding is supported by other studies showing that tropical drought deciduous woodlands can be stratified from tropical humid forests, for example, by using LAI and NDVI brightness metrics (Hansen et al., 2000; DeFries et al., 2002).

Please cite this article in press as: Prates-Clark, C.C., et al., Predicting geographical distribution models of high-value timber trees in the Amazon Basin using remotely sensed data, Ecol. Model. (2007), doi:10.1016/j.ecolmodel.2007.09.024

Although, RS-data sets and climate variables are co-occurring and mutually dependent, they distinctly represent surface characteristic of habitats. The distribution models for *C. brasiliensis*, *C. guianensis* and *V. surinamensis* showed that climate variables are an inadequate or incomplete means of accurately describing the ecological niche of these tree species or portraying these species' geographical distribution. For example, the potential distribution map derived from bioclimatic layers for *C. brasiliensis* (Fig. 2a) predicted areas of high precipitation and shorter dry season in lowlands areas as potential habitat while several areas in Western and Northwestern regions were wrongly predicted as having a lower probability of species' occurrence. To a large extent, these patterns were due to artifacts in the bioclimatic data and, more specifically, as a result of the interpolation with the elevation data. Conversely, an interesting feature of the RS-model is the accurate prediction of Atlantic Coastal forest of Brazil as a potential habitat of *C. guianensis*. This species is known to be present in areas from the Atlantic and Pacific coasts of Central America, Guyanas, and the Atlantic coastal forest of Brazil (de Granville, 1988; Hall et al., 1994; Callado et al., 2001). Furthermore, the distribution map using RS data also reflected predictions with an equal balance between omission and commission errors avoiding areas unlikely for this species' occurrence, such as deforested areas in the Brazilian Amazon, drier areas in the middle-Northeast of Brazil, and open savannas in Venezuela. Therefore, when focusing on determinants or potential predictor variables to model species' occurrence by quantifying biophysical attributes associated with species ecological niche, data extracted from remote sensed data at a finer spatial resolution and in temporal basis represents the most convenient and reliable source of data.

The ecological niche for *C. brasiliensis*, *C. guianensis* and *V. surinamensis* varied between the three tree species in spite of broadly overlapping geographic distributions. However, densely vegetated forests as reported by high leaf area (even during the dry months), vegetation greenness, branch orientation and canopy structure and high moisture (possibly because they contain a larger number of emergent trees) were identified as being necessary for the occurrence of these three tree species across the Amazon Basin. Alternatively, deforested or seasonal vegetated forests identified by mean and seasonality LAI and NDVI metrics during the dry season were excluded from the distribution maps of these tree species. This has major implications for the occurrence of these species and other flora and fauna correlated, especially in those areas where deforestation mainly for selective logging continues at high rates. When these forests are removed, the complex forest system is destroyed and, in cleared areas, the physical environment is very different from that of the original forests. In particular, the ground is exposed to greater extremes of rainfall, temperature and solar radiation. Soil erosion and the humidity within these areas are also significantly different. The leaf litter dries out rapidly and the associated symbiotic mycorrizae are lost, which limits the ability of trees to extract nutrients from the soil. Conservation of these tree habitats should become a high priority in the Amazon Basin since tree species and the Amazon ecosystem as a whole are entirely depended on one another.

Although RS-data sets have been shown to be a useful operational tool for predicting habitat suitability, for *C. brasiliensis*, *C. guianensis* and *V. surinamensis*, an inherent limitation associated with predicting species distribution models based on species suitable habitat relates to how the difference between life zone (i.e. the potential predicted range) and the actual species' distribution can be explained (Soberón and Peterson, 2005). The selection of an appropriate method to model species' geographical ranges should be based on statistical considerations, techniques for testing accuracy of the model and plausible biological agreement of the models. Sometimes the species' geographical range is greatly influenced by historic processes such as dry periods during the Pleistocene or the uplift of the Andes (de Oliveira and Mori, 1999). Despite climate change, there is evidence to suggest that the Amazonian rainforest has retracted to cover areas smaller than today and has then expanded when conditions became more favourable (Hays et al., 1976; Haffer, 1969; Gentry, 1982a,b; Haffer and Prance, 2001). The repeated expansion and contraction of the rainforest, where large continuous sections of forests became fragmented and isolated for long periods and then reconnected, produced local differentiation in climate and biogeography (Prance, 1982; Haffer, 1993). These events could explain why a species does not grow in an area predicted by the predictive species distribution model. For example, species of the genus Cariniana (Brazilian nut family) are not found in the North Eastern Amazon where the conditions are favourable for their occurrence (personal communication with Dr. Scott Mori). Conditions for forest growth also have a strong local component. For example, despite the low quality of much Amazonian soil (Sanchez and Buol, 1975; Sanchez et al., 1982; Malhi et al., 2004), vigorous tree diversity grows in the Amazon Basin (de Oliveira and Mori, 1999) and is attributed to the rapid cycling of nutrients (Connell and Slatyer, 1977; Brown and Lugo, 1990), a diverse and abundant population of decomposers (Janos, 1980; Eggleton et al., 1998) and high nutrients stocks on its vegetation (Stark and Jordan, 1978).

Finally, the models developed to potentially predict tree species' occurrence on this paper are static, i.e. are based on biophysical properties extracted from RS data at the time that satellite imagery was obtained across the Earth surface of the study area. Nevertheless, these models carry important information: the geographical distributions of high-value timber trees were mapped and several environmental properties that define the ecological niche of these species were determined from data acquired from remotely sensed data. Furthermore, traditional methods such as data collection using ground surveys are impractical for large spatial analysis and require an inordinate amount of time. RS-data sets can be updated on a regular basis for environmental monitoring—a practice which is becoming increasingly important in tropical areas because of rapid, human-induced changes.

6. Conclusion

There are many possible approaches researchers can take when attempting to determine species' distribution ranges—e.g. habitat characteristics and estimates of species' physiological tolerances. However, there is, as yet, no comparative study of the use of remote sensed data and bioclimatic surfaces to modelling species' occurrence. Certainly, the interpretation of data extracted from remote sensing offers a tool for the temporal and spatial assessment of variables (such as percentage of tree cover, vegetation greenness, canopy moisture and roughness) to map, monitor and understand the dynamics of species' occurrence and factors that have an impact on their distribution.

In our study, environmental properties extracted from RS-data sets made a significant contribution in mapping habitat suitability for high-value timber tree species occurring in the Amazon Basin. The methodology we used may also be useful in studies of other plant or animal species as well as other ecosystems. We have shown that relationships exist between the ecological niche of the three tree species studied and environmental variables extracted from RS-data sets. We have also demonstrated that the usefulness of remote sensed data to predict tree species distributions is limited due to several other processes that may determine or explain the existence of a particular tree species. Nonetheless, our approach provides species distribution maps that can be used in *post hoc* analysis to maximize the agreement between observed and predicted species distributions. These analyses provide a measure of how environmental change such as land degradation, forest loss and forest fragmentation may compromise the occurrence of these species. In sum, our approach, we believe, provides a step forward in quantifying the biophysical properties underlying the interactions between species' occurrence and its habitat.

Acknowledgements

This research was funded by NASA grant number MNG04GM27G. We thank Dr. Chuck Peters for providing the field data. We are thankful to Dr. Raphael Sagarin, Brian Zutta, Levanto Schachter, Dr. Tomas Gillespie, Benjamin Wang, and Dr. Scott Mori for helpful comments on this manuscript. Special thanks to Natálio Prates for his continuous, everpresent support.

Appendix A. Supplementary datadata

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ecolmodel.2007.09.024.

REFERENCES

Anderson, R.P., Lew, D., Peterson, A.T., 2003. Evaluating predictive models of species' distributions: criteria for selecting optimal models. Ecol. Model. 162, 211–232.

Anderson, R.P., Martínez-Meyer, E., 2004. Modelling species' geographic distributions for conservation assessments: an implementation with the spiny pocket mice (Heteromys) of Ecuador. Biol. Conserv. 116, 167–179.

Asner, G.P., Knapp, D.E., Broadbent, E.N., Oliveira, P.J.C., Keller, M., Silva, J.N., 2005. Selective logging in the Brazilian Amazon. Science 310, 480–482.

Bena, P., 1960. Famille des Lécythidacées in essences forestières de Guyane. Bureau Agricole et Forestier Guyanais, Paris.

Boone, R.B., Galvin, K.A., Smith, N.M., Lynn, S.J., 2000. Generalizing El Nino effects upon Maasai lovestock using hierarchical clusters of vegetation patterns. PHOTOGRAMM ENG REM S 66 (6), 737–744.

Brown, S., Lugo, A.E., 1990. Tropical secondary forests. J. Trop. Ecol. 6, 1–32.

Callado, C.H., da Silva Neto, S.J., Scarano, F.R., Costa, C.G., 2001. Periodicity of growth rings in some flood-prone trees of the Atlantic rainforest in Rio de Janeiro, Brazil. Trees 15, 492–497.

Cantor, S.B., Sun, C.C., Tortolero-Luna, G., Richards-Kortum, R., Follen, M., 1999. A comparison of C/B ratios from studies using receiver operating characteristic curve analysis. J. Clin. Epidemiol. 52, 885–892.

Carpenter, L.F., Nicholos, J.D., Sandi, E., 2004. Early growth of native and exotic trees planted on degraded tropical pasture. For. Ecol. Manage. 196, 367–378.

Carvalho, G., Nepstad, D., McGrath, D., Diaz, M.del C.V., Santilli, M., Barros, A.C., 2002. Frontier expansion in the Amazon: balancing development and sustainability. Environment 44, 34–45.

Chase, T.N., Pielke, R.A., Kittel, T.G.F., Nemani, R., Running, S.W., 1996. The sensitivity of a general circulation model to large-scale vegetation changes. J. Geophys. Res. 101, 7393–7408.

Connell, J.H., Slatyer, R.O., 1977. Mechanisms of succession in natural communities and their role in community stability and organization. Am. Nat. 111, 1119–1144.

Connell, J.H., 1978. Diversity in tropical rain forests and coral reefs—high diversity of trees and corals is maintained only in a non-equilibrium state. Science 199, 1302–1310.

Daly, C., Gibson, W.P., Taylor, G.H., Johnson, G.L., Pasteris, P., 2002. A knowledge-based approach to the statistical mapping of climate. Climate Res. 22, 99–113.

de Granville, J.J., 1988. Phytogeographical characteristics of the Guianan forests. Taxon 37, 578–594.

de Oliveira, A.A., Mori, S.A., 1999. A central Amazonian terra firme forest. I. High tree species richness on poor soils. Biodiv. Conserv. 8, 1219–1244.

DeFries, R.S., Houghton, R.A., Hansen, M.C., Field, C.B., Skole, D., Townshend, J., 2002. Carbon emissions from tropical deforestation and regrowth based on satellite observations for the 1980s and 90s. Proc. Nat. Acad. Sci. 99, 14256–14261.

Della Pietra, S., Della Pietra, V., Lafferty, J., 1997. Inducing features of random fields. Trans. Pattern Anal. Mach. Intell. 19 (4).

Dudik, M., Phillips, S.J., Schapire, R.E., 2004. Performance guarantees for regularized maximum entropy density estimation. In: Proceedings of the Seventeenth Annual Conference on Computational Learning Theory, pp. 472–486.

Efron, B., 1979. Bootstrap methods: another look at the jackknife. Ann. Stat. 7, 1–26.

Eggleton, P., Homathevi, R., Jeeva, D., Jones, D.T., Davies, R.G., Maryati, M., 1998. The species richness and composition of termites (Isoptera) in primary and regenerating lowland dipterocarp forest in Sabah, east Malaysia. Ecotropica 3, 119–128.

Engler, R., Guisan, A., Rechsteiner, L., 2004. An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. Journal of Applied Ecology 41, 263–274.

Fisch, E., dos Santos, F.A.M., 2001. Demography, phenology and sex of *Calophyllum brasiliensis* (Clusiaceae) trees in the Atlantic forest. J. Trop. Ecol. 7 (6), 903–909.

Fisher, B.L., Howe, H.F., Wright, J.S., 1991. Survival and growth of *Virola surinamensis* yearlings: Water augmentation in gap and understory. Oecologia 86 (2), 292–297.

Please cite this article in press as: Prates-Clark, C.C., et al., Predicting geographical distribution models of high-value timber trees in the Amazon Basin using remotely sensed data, Ecol. Model. (2007), doi:10.1016/j.ecolmodel.2007.09.024

Forget, P.M., 1996. Removal of seeds of *Carapa procera* (Meliaceae) by rodents and their fate in rainforest in French Guiana. J. Trop. Ecol. 12, 751–761.
Fraser, R.H., Currie, D.J., 1996. The species richness-energy hypothesis in a system where historical factors are thought to prevail: coral reefs. Am. Nat. 148, 138–159.
Gentry, A.H., 1982a. Neotropical floristic diversity: phytogeographical connections between Central and South America, Pleistocene climatic fluctuations, or an ancient of the Andean orogeny? Ann. Mo. Bot. Garden 69, 557–593.
Gentry, A.H., 1982b. Patterns of neotropical plant diversity. Evol. Biol. 15, 1–84.
Graham, C.H., Ferrier, S., Huettman, F., Moritz, C., Peterson, A.T., 2004. New developments in museum-based informatics and application in biodiversity analysis. Trends Ecol. Evol. 19, 497–503.
Guariguata, M.R., Dupuy, J.M., 1997. Forest regeneration in abandoned logging roads in lowland Costa Rica. Biotropica 29, 15–28.
Guariguata, M.R., Rosales-Adame, J.J., Finegan, B., 2000. Seed removal and fate in two selectively logged lowland forests with contrasting protection levels. Conserv. Biol. 14, 1046–1054.
Guariguata, M.R., Claire, H.A., Jones, G., 2002. Tree seed fate in a logged and fragmented forest landscape, Northeastern Costa Rica. Biotropica 34, 405–415.
Guisan, A., Graham, C.H., Elith, J., Huettmann, F., and the N NCEAS Species Distribution Modelling Group, 2007. Sensitivity of predictive species distribution models to change in grain size. Diversity and Distributions 13, 332–340.
Haffer, J., 1969. Speciation in Amazonian forest birds. Science 165, 131–137.
Haffer, J., 1993. Time's cycle and time's arrow in the history of Amazonia. Biogeographica 69, 15–45.
Haffer, J., Prance, G.T., 2001. Climatic forcing of evolution in Amazonia during the Cenozoic: on the refuge theory of biotic differentiation. Amazoniana 16, 579–607.
Hall, P., Orrell, L.C., Bawa, K.S., 1994. Genetic diversity and mating system in a tropical tree, *Carapa guianensis* (Meliaceae). Am. J. Bot. 81 (9), 1104–1111.
Hansen, M.C., DeFries, R.S., Townshend, J.R.G., Sohlberg, R., 2000. Global land cover classification at 1 km spatial resolution using a classification tree approach. Int. J. Remote Sens. 21, 1331–1364.
Hansen, M., DeFries, R., Townshend, J.R.G., Sohlberg, R., Dimiceli, C., Carrol, M., 2002. Towards an operational MODIS continuous field of percent tree cover algorithm: examples using AVHRR and MODIS. Remote Sens. Environ. 3, 303–319.
Hays, J.D., Imbrie, J., Shackleton, N.J., 1976. Variations in the Earth's orbit: pacemaker of the ice ages. Science 194, 1121–2113.
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. Int. J. Climatol. 25, 1965–1978.
Howe, H.F., Schupp, E.W., Westley, L.C., 1985. Early consequences of seed dispersal for a Neotropical tree (*Virola surinamensis*). Ecology 66, 781–791.
Howe, H.F., 1990. Survival and growth of juvenile *Virola surinamensis* in Panama: effects of herbivory and canopy closure. J. Trop. Ecol. 6, 259–280.
Hubbell, S.P., Foster, R.B., O'Brien, S.T., Harms, K.E., Condit, R., Wechsler, B., Wright, S.J., De Lao, S.L., 1999. Light-gap disturbances, recruitment limitation and tree diversity in a neotropical forest. Science 283, 554–557.
Janos, D.P., 1980. Vesicular-arbuscular mycorrhizae affect lowland tropical rainforest plant growth. Ecology 61 (1), 151–162.
Jarvis, A., Rubiano, J., Nelson, A., Farrow, A., Mulligan, M., 2004. Practical use of SRTM data in the tropics: comparisons with digital elevation models generated from cartographic data. Working document 198. International Center for Tropical Agriculture, California, Columbia.
Karl, J., Heglund, P.J., Garton, E.O., Scott, J.M., Wright, N.M., Hutto, R.L., 2000. Sensitivity of species habitat-relationship model performance to factors of scale. Ecol. Appl. 10, 1690–1705.
Karl, J., Vancara, W.L.K.S., Eglung, P.J.H., Right, N.M.W., Coot, J.M.S., 2002. Species commonness and the accuracy of habitat-relationship models. In: Scott, J.M., Heglund, P.J., Morrison, M.I. (Eds.), Predicting Species Occurrences: Issues of Accuracy and Scale. Island Press, Washington, DC, pp. 573–580.
King, R.T., 2003. Succession and micro-elevation effects on seedling establishment of *Calophyllum brasiliensis* Camb. (Clusiaceae) in an Amazonian River meander forest. Biotropica 35, 462–471.
Lin, C.C., 1998. Emerging scatterometer applications. European Space Agency Spec. Publ. ESA SP-424, 25–31.
Liu, C., Berry, P.M., Dawson, T.P., Pearson, R.G., 2005. Selecting threshold of occurrence in the prediction of species distributions. Ecography 28, 385–393.
Long, D.G., Drinkwater, M.R., 2000. Azimuth variation in microwave scatterometer and radiometer data over Antarctica. IEEE Trans. Geosci. Remote Sens. 38 (4), 1857–1870.
Long, D.G., Drinkwater, M.R., Holt, B., Saatchi, S., Bertoia, C., 2001. Global ice and land climate studies using scatterometer image data. EOS, Trans. Am. Geophys. Union 82 (43), 503.
Lopes, N.P., Yoshida, M., Kato, M.J., 2004. Byosynthesis of Tetrahydrofuran lignans in *Virola surinamensis*. Brazil. J. Pharm. Sci. 40 (1), 54–59.
Malhi, Y., Baker, T., Phillips, O.L., Almeida, S., Alvarez, S., Arroyo, L., Chave, J., Czimczik, C.I., Di Fiore, A., Higushi, N., Killeen, T.J., Laurance, S.G., Laurance, W.F., Lewis, S.L., Montoya, L.M.M., Monteagudo, A., Neill, S.A., Vargas, P.N., Patino, S., Pitman, N.C.A., Quesada, C.A., Salomao, R., Silva, J.N.M., Lezama, A.T., Martinez, R.V., Terborgh, J., Vicenti, B., Lloyd, J., 2004. The above-ground coarse wood productivity of 104 Neotropical forest plots. Global Change Biol. 10, 1–29.
Marques, M.C.M., Joly, C.A., 2000. Seed germination and growth of *Calophyllum brasiliensis* (Clusiaceae), a typical species of flooded forests. Acta Botanica Brasilica 14 (1), 113–120.
Myneni, R.B., Hall, F.B., Sellers, P.J., Marshak, A.L., 1995. The interpretation of spectral vegetation indices. IEEE Trans. Geosci. Remote Sens. 33, 481–486.
Myneni, R.B., Keeling, C.D., Tucker, C.J., Asrarm, G., Nemani, R.R., 1997. Increased plant growth in the northern high latitudes from 1981 to 1991. Nature 386, 698–702.
Myneni, R.B., Tucker, C.J., Keeling, C.D., Asrar, G., 1998. Interannual variation in satellite-sensed vegetation index data from 1981 to 1991. J. Geophys. Res. 103 (D6), 6145–6160.
Myneni, R.B., Hoffman, S., Knyazikhin, Y., Privette, J.L., Glassy, J., Tian, Y., Wang, Y., Song, X., Zhang, Y., Smith, G.R., Lotsch, A., Friedl, M., Morisette, J.T., Votava, P., Nemani, R.R., Running, S.W., 2002. Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data. Remote Sens. Environ. 83, 214–231.
Mori, S.A., Prance, G.T., 1990b. Taxonomy, ecology and economic botany of the Brazil nut (Bertholletia excelsa Humb. & Bonpl: Lecythidaceae). Adv. Econ. Bot. 8, 130–150.
National Aeronautics and Space Administration (NASA), 2006. In NASA web site, www.jpl.nasa.gov/srtm Nepstad, D.C., Veríssimo Alencar, A., Nobre, C., Lima, E., Lefebvre, P., Schlesinger, P., Potter, C., Moutinho, P., Mendoza, E., Cochrane, M., Brooks, V., 1999. Large-scale impoverishment of Amazonian forests by logging and fire. Nature 398,505–508.
Nepstad, D.C., Veríssimo, Alencar, A., Nobre, C., Lima, E., Lefebvre, P., Schlesinger, P., Potter, C., Moutinho, P., Mendoza, E., Cochrane, M., Brooks, V., 1999. Large-scale impoverishment

Please cite this article in press as: Prates-Clark, C.C., et al., Predicting geographical distribution models of high-value timber trees in the Amazon Basin using remotely sensed data, Ecol. Model. (2007), doi:10.1016/j.ecolmodel.2007.09.024

of Amazonian forests by logging and fire. Nature 398, 505–508.

Nepstad, D., Ramos, C.A., Lima, E., McGrath, D., Pereira, C., Merry, F., 2004. Managing the Amazon timber industry. Conserv. Biol. 18 (2), 575–577.

New, M., Hulme, M., Jones, P., 1999. Representing twentieth-century space-time climate variability. Part I: Development of a 1961–90 mean monthly terrestrial climatology. J. Climate 12, 829–856.

New, M., Lister, D., Hulme, M., Makin, I., 2002. A high-resolution data set of surface climate over global land areas. Climate Res. 21, 1–25.

O'Brien, E.M., Field, R., Whittaker, R.J., 2000. Climatic gradients in woody plant (tree and shrub) diversity: water-energy dynamics, residual variation, and topography. Oikos 89, 588–600.

Parra, J.L., Graham, C.C., Freile, J.F., 2004. Evaluating alternative data sets for ecological niche models of birds in Andes. Ecography 27, 350–360.

Pausas, J.G., Carreras, J., Ferré, A., Font, X., 2003. Coarse-scale plant species richness in relation to environmental heterogeneity. J. Veg. Sci. 14, 661–668.

Peterson, A.T., Cohoon, K.P., 1999. Sensitivity of distributional prediction algorithms to geographic data completeness. Ecol. Model. 117, 159–164.

Peterson, A.T., Viglais, D.A., 2001. Predicting species invasions using ecological niche modelling. BioScience 51, 363–371.

Phillips, S.J., Dudik, M., Schapire, R.E., 2004. A maximum entropy approach to species distribution modelling. In: Proceedings of the Twenty-first Century International Conference on Machine Learning. ACM Press, New York, pp. 655–662.

Phillips, S.J., Anderson, R.P., Schapire, R.E., 2006. Maximum entropy modeling of species geographic distributions. Ecol. Model. 190, 231–259.

Prance, G.T., 1982. Forest refuges: evidence from woody angiosperms. In: Prance, G.T. (Ed.), Biological Diversification in the Tropics. Columbia University Press, New York, pp. 137–158.

Record, S.J., Hess, R.W., 1943. Timbers of the New World. Yale University Press, New Haven (reprinted by Arno Press, New York, 1972).

Rodrigues, W., 1980. Revisao taxonomica das especies de *Virola* Aublet. Acta Amazonica 1, 1–127.

Sanchez, P.A., Buol, S.W., 1975. Soils of the tropics and the world food crisis. Science 188, 598–603.

Sanchez, P.A., Bandy, D.E., Villachica, J.H., Nicholaides III, J.J., 1982. Amazon Basin soils: management for continuous crop production. Science 216, 821–827.

Shabanov, N.V., Zhou, L., Knyazikhin, Y., Myneni, R.B., 2002. Analysis of interannual changes in Northern vegetation activity observed in AVHRR data during 1981 to 1994. IEEE Trans. Geosci. Remote Sens. 40 (1), 115–130.

Stark, N., Jordan, C.F., 1978. Nutrient retention by the root mat of an Amazonian Rainforest. Ecology 59, 434–437.

Soberón, J., Peterson, A.T., 2005. Interpretation of models of fundamental ecological niches and species' distributional areas. Biodiv. Inf. 2, 1–10.

Stockwell, D.R.B., Peters, D., 1999. The GARP Modeling System: problems and solutions to automated spatial prediction. Int. J. Geogr. Inf. Sci. 13 (2), 143–158.

Swets, J.A., 1988. Measuring the accuracy of diagnostic systems. Science 240, 1285–1293.

Swets, J.A., Pickett, R.M., 1992. Evaluation of Diagnostic Systems: Methods from Signal Detection Theory. Academic Press, New York.

Tucker, C.J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. Remote Sens. Environ. 8, 127–150.

Tucker, C.J., Vanpraet, C., Boerwinkel, E., Gaston, A., 1983. Satellite remote sensing of total dry matter production in the Senegalese Sahel. Remote Sens. Environ. 13, 461–474.

Tukey, J., 1977. Exploratory Data Analysis. Addison-Wesley.

Uhl, C., Vieira, I., 1989. Ecological impacts of selective logging in the Brazilian Amazon: a case study of the Paragominas region of the State of Para. Biotropica 21, 98–106.

Veríssimo, A., Barreto, P., Mattos, M., Tarifa, R., Uhl, C., 1992. Logging impacts and prospects for sustainable forest management in an old Amazonian frontier: the case of Paragominas. For. Ecol. Manage. 55, 169–199.

Vianna, N.G., 1982. Conservacao de sementes de anriroba (*Carapa guianensis* Aubl.) Circular Tecnico 34, CPATU-EMBRAPA, Belem/PA, Brasil, p. 10.

Vinson, C.C., Azevedo, V.C.R., Sampaio, I., Ciampi, A.Y., 2005. Development of microsatellite markers for *Carapa guianensis* (Aublet), a tree species from the Amazon forest. Mol. Ecol. Notes 5 (3), 33–34.

Whitmore, T.C., 1988. The influence of tree population dynamics on forest species composition. In: Davy, A.J., Hutchings, M.J., Watkinson, A.R. (Eds.), Plant Population Ecology. Blackwell Scientific Publications, Oxford, pp. 271–292.

Please cite this article in press as: Prates-Clark, C.C., et al., Predicting geographical distribution models of high-value timber trees in the Amazon Basin using remotely sensed data, Ecol. Model. (2007), doi:10.1016/j.ecolmodel.2007.09.024