Julius-Maximilians-Universität Würzburg
Institut für Geographie und Geologie
Lehrstuhl für Fernerkundung

Potential of multi-temporal remote sensing data for modeling tree species distributions and species richness in Mexico

Dissertation zur Erlangung
des naturwissenschaftlichen Doktorgrades
der Julius-Maximilians-Universität Würzburg

vorgelegt von Anna Cord

Würzburg, März 2012

Eingereicht am: 06.03.2012

1. Gutachter: Prof. Dr. Stefan Dech
2. Gutachter: Prof. Dr. Björn Reineking

1. Prüfer: Prof. Dr. Stefan Dech
2. Prüfer: Prof. Dr. Jürgen Rauh

Mentorin: Dr. Doris Klein

Tag der Disputation: 09.05.2012

Abstract

Current changes of biodiversity result almost exclusively from human activities. This anthropogenic conversion of natural ecosystems during the last decades has led to the so-called ‘biodiversity crisis’, which comprises the loss of species as well as changes in the global distribution patterns of organisms. These modifications in species functional traits available in a certain area are intimately connected with the loss of ecosystem services, from which we humans directly and indirectly benefit. To highlight this causal relationship between biodiversity and human well-being, the United Nations has declared the years 2011 to 2020 as the ‘Decade of Biodiversity’.

Species richness is unevenly distributed worldwide. Altogether, 17 so-called ‘megadiverse’ nations cover less than 10% of the earth’s land surface but support nearly 70% of global species richness. Mexico, the study area of this thesis, is one of those countries. Due to this special responsibility for global biodiversity, Mexico has put particular emphasis on assessing its biota. Already in 1992, the *National Commission for the Knowledge and Use of Biodiversity* (CONABIO) was established. The country today has a unique availability of species occurrence data from different sources including the *National Information System on Mexican Biodiversity* (SNIB) and the *National Forest Inventory* (INFyS). However, due to Mexico’s large extent and geographical complexity, it is impossible to conduct reliable and spatially explicit assessments of species distribution ranges based on these collection data and field work alone.

In the last two decades, *Species distribution models* (SDMs) have been established as important tools for extrapolating such *in situ* observations. SDMs analyze empirical correlations between geo-referenced species occurrence data and environmental variables to obtain spatially explicit surfaces indicating the probability of species occurrence. Remote sensing can provide such variables which describe biophysical land surface characteristics with high *effective* spatial resolutions. Especially during the last three to five years, the number of studies making use of remote sensing data for modeling species distributions has therefore multiplied. Due to the novelty of this field of research, the published literature consists mostly of selective case studies. A systematic framework for modeling species distributions by means of remote sensing is still missing.

This research gap was taken up by this thesis and specific studies were designed which addressed the combination of climate and remote sensing data in SDMs, the suitability of continuous remote sensing variables in comparison with categorical land cover classification data, the criteria for selecting appropriate remote sensing data depending on species characteristics, and the effects of inter-annual variability in remotely sensed time series on the performance of species distribution models. The corresponding novel analyses were conducted with the *Maximum Entropy* algorithm developed by Phillips et al. (2004).

In this thesis, a more comprehensive set of remote sensing predictors than in the existing literature was utilized for species distribution modeling. The products were selected based on their ecological relevance for characterizing species distributions. Two 1 km Terra-MODIS Land 16-day composite standard products including the *Enhanced*

Vegetation Index (EVI), *Reflectance Data*, and *Land Surface Temperature* (LST) were assembled into enhanced time series for the time period of 2001 to 2009. These high-dimensional time series data were then transformed into 18 phenological and 35 statistical metrics that were selected based on an extensive literature review.

Spatial distributions of twelve tree species were modeled in a hierarchical framework which integrated climate (WorldClim) and MODIS remote sensing data. The species are representative of the major Mexican forest types and cover a variety of ecological traits, such as range size and biotope specificity. Trees were selected because they have a high probability of detection in the field and since mapping vegetation has a long tradition in remote sensing.

The major findings of this thesis can be summarized as: Spatial autocorrelation (SAC, measured by Moran's I) was significantly lower for remote sensing than for interpolated climate data. Due to the inherent SAC of species occurrence data, a non-hierarchical model design would therefore not fully exploit the potential of remote sensing data. The hierarchical modeling approach implemented in this thesis outperformed purely climatic and remote sensing based models by improving the effective spatial resolution of the environmental predictors and integrating additional constraints on the species distributions. Due to the inclusion of remote sensing data, the *fractional predicted area* (FPA) was reduced by up to 80% per species. The comparison of continuous remote sensing data and an existing categorical land cover classification showed that the general overestimation of species distribution ranges based on categorical land cover information was influenced by how closely the spatial distribution patterns of a species could be linked to certain land cover types. A novel *land cover evenness index* was introduced which quantifies this species-land cover relationship and showed a linear correlation ($r^2 = 0.54$) with the model improvement measured by the increase in AUC (*area under curve*) scores.

Beyond this, the relevance of different remote sensing products to the model results was determined: Vegetation indices are the most widely used but not necessarily the best predictors of species distributions. Instead, the feature ranking indicated a high potential of LST – in particular for temperate species. Spectral bands in turn were only of minor importance, except for the mangrove species occurring in coastal wetlands. The use of longer remotely sensed time series generally improved the performance of species distribution models. However, the impact greatly differed from 0.6% up to 30.0% reduction in model deviance between species. Wide-ranging species (≥ 300 presence records) generally featured less variation in model deviance.

In addition, the so-called *stacked species distribution modeling* approach was applied to compile the first Mexican pine richness map based on 40 species of the tree genus *Pinus* (family Pinaceae). Mexico has the greatest number of pine species of any region of similar size in the world. The species distribution ranges modeled with the hierarchical combination of both bioclimatic and remote sensing data yielded the best accuracy with the lowest overall deviance (0.169) from independent presence-absence data across all species. A site-specific evaluation further showed that the combined model was significantly better in predicting species richness than the purely climatic model. The analysis thus confirmed the potential of remote sensing data not only for obtaining better estimates of individual species distributions but also for predicting species richness.

Zusammenfassung

Sämtliche aktuell zu beobachtenden Veränderungen in der Biodiversität lassen sich fast ausschließlich auf menschliche Aktivitäten zurückführen. In den letzten Jahrzehnten hat insbesondere die anthropogene Umwandlung bisher unberührter, natürlicher Ökosysteme zur sogenannten ‚Biodiversitätskrise' geführt. Diese umfasst nicht nur das Aussterben von Arten, sondern auch räumliche Verschiebungen in deren Verbreitungsgebieten. Daraus wiederum ergeben sich Veränderungen in den jeweiligen funktionalen Eigenschaften (*functional traits*) der Arten, welche eng mit der Verfügbarkeit von Ökosystemdienstleistungen in der jeweiligen Region verknüpft sind, von denen wir Menschen direkt und indirekt profitieren. Um diesen kausalen Zusammenhang zwischen Biodiversität und menschlichem Wohl zu unterstreichen, wurde der Zeitraum von 2011 bis einschließlich 2020 von den Vereinten Nationen zur ‚Dekade der Biodiversität' erklärt.

Global gesehen ist der Artenreichtum ungleich verteilt. Nur insgesamt 17 sogenannte ‚megadiverse' Länder, welche 10% der globalen Landoberfläche umfassen, beherbergen fast 70% der weltweiten Artenvielfalt. Mexiko, das Studiengebiet dieser Arbeit, ist eine dieser außerordentlich artenreichen Nationen. Aufgrund seiner besonderen Verantwortung für die globale Biodiversität wurde daher bereits im Jahr 1992 die *Nationale Kommission für die Erforschung und Nutzung von Biodiversität* (CONABIO) eingerichtet. Heute verfügt Mexiko über einzigartige Datenbanken über die Verteilung von Fundorten der einzelnen Arten im Land, wie beispielsweise das *Nationale Informationssystem zur mexikanischen Biodiversität* (SNIB) und das *Nationale Forstinventar* (INFyS). Aufgrund seiner großen Ausdehnung und geographischen Komplexität kann eine verlässliche und detaillierte räumliche Erfassung von Artverbreitungsgebieten in Mexiko jedoch nicht nur auf Basis dieser Datenbanken sowie von Feldarbeiten erfolgen.

In den letzten beiden Jahrzehnten haben sich Artverbreitungsmodelle (*Species distribution models*, SDMs) als wichtige Werkzeuge für die räumliche Interpolation solcher *in situ* Beobachtungen in der Ökologie etabliert. Artverbreitungsmodelle umfassen die Analyse empirischer Zusammenhänge zwischen georeferenzierten Fundpunkten einer Art und Umweltvariablen mit dem Ziel, räumlich kontinuierliche Vorhersagen zur Wahrscheinlichkeit des Vorkommens der jeweiligen Art zu treffen. Mittels Fernerkundung können solche Umweltvariablen mit Bezug zu den biophysikalischen Eigenschaften der Landoberfläche in hohen effektiven räumlichen Auflösungen bereitgestellt werden. Insbesondere in den letzten drei bis fünf Jahren ist daher die Verwendung von Fernerkundungsdaten in der Artverbreitungsmodellierung sprunghaft angestiegen. Da es sich hierbei jedoch immer noch um ein sehr neues Forschungsfeld handelt, stellen diese meist nur Einzelstudien mit Beispielcharakter dar. Eine systematische Untersuchung zur Modellierung von Artverbreitungsgebieten mit Hilfe von Fernerkundungsdaten fehlt bisher.

Diese Forschungslücke wurde in der vorliegenden Arbeit aufgegriffen. Hierzu wurden spezifische Untersuchungen durchgeführt, welche insbesondere folgende Aspekte betrachteten: die sinnvolle Verknüpfung von Klima- und Fernerkundungsdaten im Rahmen von Artverbreitungsmodellen, den quantitativen Vergleich von kontinuierlichen Fernerkundungsdaten und einer bestehenden kategorialen Landbedeckungsklassifikation, die Identifizierung von Kriterien zur Auswahl geeigneter Fernerkundungsprodukte, welche die Eigenschaften der Studienarten berücksichtigen, sowie der Einfluss inter-annueller Variabilität in fernerkundlichen Zeitreihen auf die Ergebnisse und Leistungsfähigkeit von Artverbreitungsmodellen. Die entsprechenden neuen Analysen wurden mit Hilfe des von Phillips et al. (2004) entwickelten *Maximum Entropy* Algorithmus zur Artverbreitungsmodellierung durchgeführt.

Im Rahmen dieser Arbeit wurde ein umfangreicherer Datensatz an Fernerkundungsvariablen als in der bisherigen Literatur verwendet. Die entsprechenden Fernerkundungsprodukte wurden spezifisch aufgrund ihrer Eignung für die Beschreibung ökologisch relevanter Parameter, die sich auf die Verbreitungsgebiete von Arten auswirken, ausgewählt. Für den Zeitraum von 2001 bis 2009 wurden zwei Terra-MODIS Standardprodukte mit 1 km räumlicher und 16-tägiger zeitlicher Auflösung zu geglätteten, kontinuierlichen Zeitreihen zusammengefügt. Diese Produkte beinhalten den verbesserten Vegetationsindex (*Enhanced Vegetation Index*, EVI), Reflexionsgrade und die Landoberflächentemperatur (*Land Surface Temperature*, LST). Diese hochdimensionalen Zeitreihendaten wurden in insgesamt 18 phänologische sowie 35 statistische Maßzahlen überführt, welche auf der Basis einer umfassenden Sichtung der vorhandenen Literatur zusammengestellt wurden.

Die Verbreitungsgebiete von zwölf Baumarten wurden mit Hilfe eines hierarchisch aufgebauten Ansatzes, welcher sowohl Klimadaten (WorldClim) als auch Fernerkundungsdaten des MODIS-Sensors berücksichtigt, modelliert. Die Studienarten sind repräsentativ für die in Mexiko vorkommenden Waldtypen und decken eine breite Spannweite ökologischer Eigenschaften wie Größe des Verbreitungsgebietes und Breite der ökologischen Nische ab. Als Studienobjekte wurden Bäume ausgewählt, weil sie im Feld mit hoher Wahrscheinlichkeit richtig erfasst werden und außerdem die fernerkundungsbasierte Kartierung von Vegetation bereits auf eine Vielzahl an Studien zurückgreifen kann.

Die wesentlichen Ergebnisse dieser Arbeit können wie folgt zusammengefasst werden: Räumliche Autokorrelation (*Spatial autocorrelation*, SAC) wurde anhand des Moran's I Index bestimmt und war für die Fernerkundungsdaten signifikant niedriger als für die interpolierten Klimawerte. Da auch die Daten zu den Fundorten einer Art typischerweise räumliche Autokorrelation aufweisen, würde ein nicht-hierarchischer Modellierungsansatz das Potential von Fernerkundungsdaten für die Artverbreitungsmodellierung nicht voll ausschöpfen. Der im Rahmen dieser Arbeit umgesetzte hierarchische Ansatz zeigte im Allgemeinen eine höhere Genauigkeit als ausschließlich auf Klimadaten beruhende oder rein fernerkundungsbasierte Modelle. Hierzu trugen insbesondere die Verbesserung der effektiven räumlichen Auflösung und die Einbindung zusätzlicher Einschränkungen bezüglich der Artverbreitung auf Basis von Fernerkundungsvariablen bei. Durch die Integration von Fernerkundungsdaten konnte die vorhergesagte Fläche (*fractional predicted area*, FPA) um bis zu 80% pro Art reduziert werden. Der Vergleich der Eignung von kontinuierlichen Fernerkundungsdaten und einer bestehenden kategorialen Landbedeckungsklassifikation zeigte, dass kategoriale Landbedeckungsdaten generell dazu neigen, Artverbreitungsgebiete zu überschätzen. Allerdings war der Grad dieser Überschätzung abhängig davon, wie genau die bekannten Fundorte der Studienart mit einem oder wenigen bestimmten Landbedeckungsklassen verknüpft werden konnten. Um diesen Zusammenhang zwischen Art und Landbedeckung zu quantifizieren, wurde ein neuer *land cover evenness index* eingeführt. Dieser wies einen linearen Zusammenhang ($r^2 = 0.54$) mit der am Anstieg der AUC (*area under curve*)-Werte gemessenen Verbesserung des entsprechenden Artverbreitungsmodells auf.

Darüber hinaus wurde die Bedeutung verschiedener fernerkundlicher Produkte für die Artverbreitungsmodellierung untersucht. Wie die Ergebnisse zeigten, sind Vegetationsindices zwar die am häufigsten verwendeten, aber nicht unbedingt die besten Variablen zur Vorhersage von Artverbreitungsgebieten. Stattdessen erwies sich auch LST als geeignete Umweltvariable, insbesondere für Arten, die in temperaten Bereichen vorkommen. Reflexionsgrade waren nur für die Vorhersage der Mangrovenart, welche in Feuchtgebieten entlang der Küste zu finden ist, von Bedeutung. Insgesamt wirkte sich die Verwendung von längeren fernerkundlichen Zeitreihen positiv auf die Modellergebnisse aus, der Einfluss variierte allerdings stark zwischen den einzelnen Arten und führte je nach Art zu einer Verringerung der Modellabweichung zwischen 0.6% und 30.0%. Weit verbreitete Arten, die durch eine Vielzahl an Fundorten (≥ 300 Datensätze) repräsentiert wurden, wiesen insgesamt geringere Schwankungen in der Abweichung der Modelle in Abhängigkeit von der Länge der Zeitreihe auf.

Mittels eines sogenannten *stacked species distribution modeling approach*, das heißt der Aufsummierung der Verbreitungsgebiete einzelner Arten, wurde ferner im Rahmen dieser Arbeit die erste fernerkundungsbasierte Diversitätskarte für 40 mexikanische Kiefernarten (Genus *Pinus*, Familie Pinaceae) erstellt. Das Land weist, gemessen an seiner Fläche, weltweit die höchste Zahl an Kiefernarten auf. Wiederum wurden die Verbreitungsgebiete der einzelnen Arten gemäß dem oben beschriebenen hierarchischen Ansatz modelliert. Diese hierarchische Kombination von Klima- und Fernerkundungsdaten führte bezogen auf unabhängige Präsenz-Absenz-Daten zu der insgesamt niedrigsten Abweichung von 0.169. Die Evaluierung zeigte weiterhin, dass das zusammengesetzte Modell den Artenreichtum signifikant besser vorhersagen konnte als das ausschließlich auf Klimadaten basierende Modell. Fernerkundungsdaten sind daher nicht nur zur Modellierung der Verbreitungsgebiete einzelner Arten, sondern auch zur räumlichen Vorhersage von Artenreichtum geeignet.

Abbreviations

ANN	Artificial Neural Network
AUC	Area Under Curve of the receiver operating characteristic
BRDF	Bidirectional Reflectance Distribution Function
BRT	Boosted Regression Tree
CONABIO	National Commission for the Knowledge and Use of Biodiversity (*Comisión Nacional para el Conocimiento y Uso de la Biodiversidad*) of Mexico
CONAFOR	National Forestry Commission (*Comisión Nacional Forestal*) of Mexico
CTA	Classification Tree Analysis
CVMVC	Constrained View Maximum Value Compositing
DEM	Digital Elevation Model
DOY	Day of Year
EE	Environmental Envelope
ES	Ecosystem Service
ENFA	Ecological Niche Factor Analysis
ENSO	El Niño/Southern Oscillation
EOS	End of Season
ETM+	Enhanced Thematic Mapper
EVI	Enhanced Vegetation Index
FAO	Food and Agriculture Organization of the United Nations
FPAR	Fraction of absorbed Photosynthetically Active Radiation
GAM	Generalized Additive Model
GARP	Genetic Algorithm for Rule-set Production
GBIF	Global Biodiversity Information Facility
GDP	Gross Domestic Product
GIS	Geographic Information System
GLM	Generalized Linear Model
GPS	Global Positioning System
HDF	Hierarchical Data Format
INFyS	National Forest Inventory (*Inventario Nacional Forestal y de Suelos*) of Mexico
ITCZ	Inner-Tropical Convergence Zone
LAI	Leaf Area Index
LC	Land Cover
LCCS	Land Cover Classification System
LiDAR	Light Detection and Ranging
LP DAAC	Land Processes Distributed Active Archive Center
LSP	Land Surface Phenology
LST	Land Surface Temperature
LUT	Look-Up Table
MARS	Multivariate Adaptive Regression Splines
MDA	Mixture Discriminant Analysis
ME	Maximum Entropy
MEI	Multivariate ENSO Index

MIR	Middle Infrared
MODAPS	MODIS Adaptive Processing System
MODIS	Moderate Resolution Imaging Spectroradiometer
MRT	MODIS Reprojection Tool
MVC	Maximum Value Compositing
NDVI	Normalized Difference Vegetation Index
NIR	Near Infrared
NPP	Net Primary Productivity
QA-SDS	Quality Assurance Science Data Set
RF	Random Forest
ROC	Receiver Operating Characteristic
RS	Remote Sensing
SAC	Spatial Autocorrelation
SAVI	Soil-Adjusted Vegetation Index
SD	Standard Deviation
SDM	Species Distribution Model
SNIB	National Information System on Mexican Biodiversity (*Sistema Nacional de Información sobre Biodiversidad de México*)
SOS	Start of Season
SRE	Surface Range Envelope
SVM	Support Vector Machine
SWIR	Short-Wave Infrared
TIR	Thermal Infrared
TiSeG	Time Series Generator
TM	Thematic Mapper
VI	Vegetation Index
VIS	Visible Light
WDRVI	Wide Dynamic Range Vegetation Index

Table of contents

1 Introduction

1.1 Ecological, societal, and economic relevance of species distributions and biodiversity

“The diversity of life has occupied the interest of human beings ever since they first appeared on Earth” (Ramamoorthy et al., 1993: p. 1). For several centuries, mankind has observed and studied spatial patterns of species distributions through their interactions with the physical environment (Elith and Leathwick, 2009). This early scientific attention to species distributions was evoked by the interest in understanding the ecological relationships between a species and its (a)biotic environment and in testing biogeographical or phylogenetic hypotheses (Franklin, 2009). Especially the importance of climate to explain plant species distributions and vegetation patterns was recognized early on (von Humboldt and Bonpland, 1807). Building on his extensive observations and documentations during his expedition through tropical America (from Peru to Mexico) between 1799 and 1803, Humboldt laid the foundation for the science of physical geography and in particular the field of **‘botanical geography’** or **‘plant-geography’** (von Humboldt and Bonpland, 1807: *Ideen zu einer Geographie der Pflanzen*).

Since then, Humboldt’s autecological and biogeographical interest in spatial patterns of species distributions has given way to a broader perspective. About 30 years ago, species were recognized as a key component of **‘biological diversity’** or **‘biodiversity’** (United Nations, 1992). Accordingly, biodiversity has been defined as “the variability among living organisms from all sources including, inter alia, terrestrial, marine and other aquatic ecosystems and the ecological complexes of which they are part: this includes diversity within species, between species and of ecosystems” (United Nations, 1992: Article 2, p. 3). As we know today, species diversity has direct functional implications since the number and kinds of species occurring within a certain area determine the organismal **functional traits** that in turn affect ecosystem processes (Chapin III et al., 2000; Mace et al., 2005). Besides *species richness* (number of species per unit area), the availability of these functional traits is controlled by several other components of biodiversity such as *species composition*, *species evenness* (relative abundances), *interactions* among species, and spatio-temporal variations in these properties (Chapin III et al., 2000).

In addition to supporting ecosystem functioning, biodiversity also directly facilitates human existence and well-being through a number of goods (e.g. food and fiber) and services. These services have been summarized as **‘ecosystem services’** (ES) from which we humans directly (provisioning, regulating, and cultural ecosystem services) and indirectly (supporting ecosystem services) benefit (Millennium Ecosystem Assessment, 2005, Figure 1.1). To highlight this causal relationship between biodiversity and human well-being, the United Nations has declared the year 2010 and later on the entire decade 2011-2020 as **‘The International Year of Biodiversity’** and the **‘Decade of Biodiversity’**, respectively. This international recognition of the societal and economic value of biodiversity is crucial given the multitude of threats it is subject to, as outlined in the following.

Current changes of biodiversity result almost exclusively from human activities (Mace et al., 2005). Over the past two centuries, many people have benefited from the exploitation of biodiversity and the conversion of natural ecosystems, e.g. for agriculture (Millennium Ecosystem Assessment, 2005). It is estimated that between 20% and 50% of nine out of 14 biomes have already been transformed to croplands (Mace et al., 2005) to support human lives. In addition to land use change, rapid climatic change (Araújo et al., 2005; Keith et al., 2008), exploitation (Dulvy et al., 2003), pollution (Wilting et al., 2010), invasive species (Ficetola et al., 2007), and pathogens (Puschendorf et al., 2009; Murray et al., 2011) impact species distributions and ecosystem functioning worldwide (Millennium Ecosystem Assessment, 2005; see Figure 1.1).

This human alteration of the global environment has led to the **‘biodiversity crisis’** (Wilson, 1988) recognized in the last two to three decades, which represents the sixth major extinction event and has caused significant changes in the global distribution patterns of organisms (Chapin III et al., 2000). In particular due to habitat loss and degradation, the rate of species extinctions at present is estimated to be 100 to 1000 times higher than the average or so-called ‘background’ extinction rate (Mace et al., 2005). According to the IUCN Red List, between 12% and 39% of species within well-studied higher taxa are threatened by extinction today (IUCN, 2011). As biodiversity also controls the resistance and resilience of ecosystems to environmental change based on the ‘diversity-stability hypothesis’ (McNaughton, 1977; Chapin III et al., 2000), anthropogenically transformed ecosystems are prone to further modifications, e.g. species invasions (Figure 1.1). The human impact on the environment does not only affect species richness but also functional shifts due to changes in species composition (Loreau et al., 2001).

Both the dramatic decline of biodiversity and modifications in species functional traits are intimately connected with a **loss of ecosystem services** (Loreau et al., 2001; Dobson et al., 2006; Hector and Bagchi, 2007). One of the key findings of the Millennium Ecosystem Assessment (2005) is that, worldwide, 15 out of the 24 ecosystem services considered are declining. Changes in the provision of ecosystem services in turn have major ecological, societal, and economic consequences (Chapin III et al., 2000; Balmford, 2002; Millennium Ecosystem Assessment, 2005). For example, decreased species diversity can directly reduce sources of food, fuel, medicinals, and genetic resources (Chapin III et al., 2000).

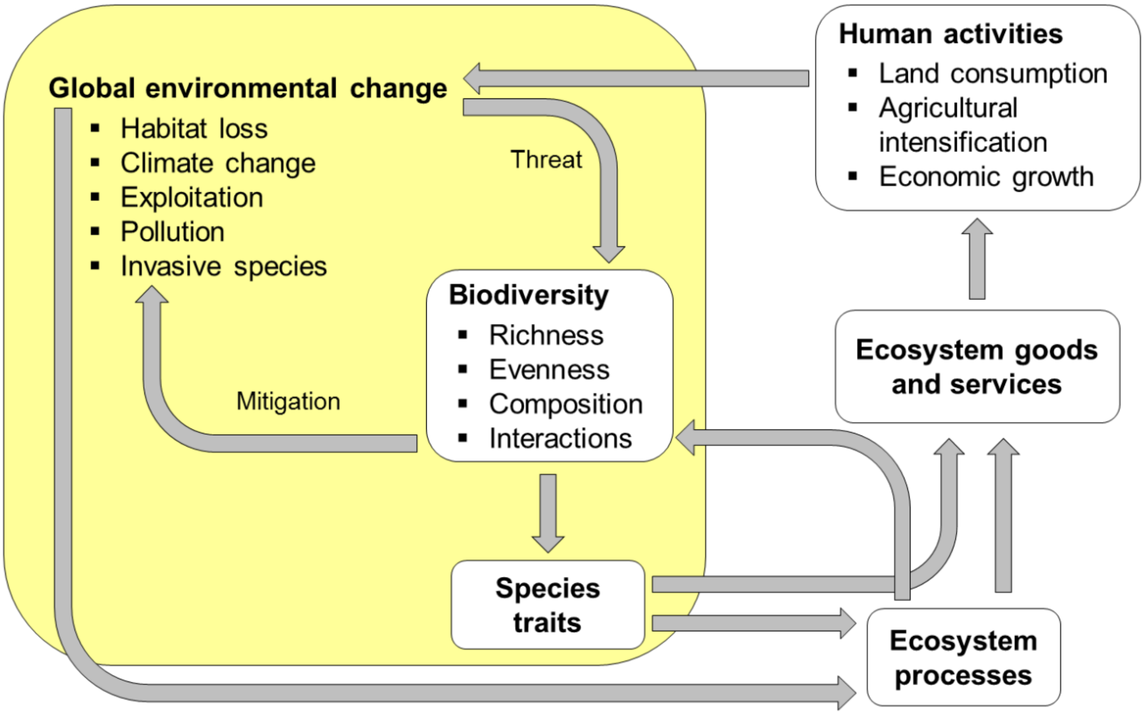


Figure 1.1: The role of biodiversity and species traits in the provision of ecosystem services and the mitigation of global environmental change (Chapin III et al., (2000); modified). Human activities are causing environmental change of global significance (here ordered according to their global importance; Millennium Ecosystem Assessment, 2005). Through a variety of mechanisms, global change leads to changes in biodiversity which in turn feed back on, for example, the susceptibility of ecosystems to species invasions. In addition, biodiversity changes have direct effects on ecosystem processes and services through altered species traits. Altered ecosystem processes again influence biodiversity (feedback mechanism) and ecosystem goods and services and hence human activities.

Accelerating loss of species and the novel view of ecosystem functioning – as based on the importance of species traits and their irreplaceability with regards to the provision of ecosystem services – has shifted the interest in and research on species distributions from the eco-physiological more to a conservation perspective (‘conservation biogeography’; Whittaker et al., 2005). It is mainly driven by the need to forecast the impact of environmental change and human management actions (e.g. species reintroductions, ecological restoration) on patterns of species distributions from local to global scales (Nagendra, 2001; Franklin, 2009).

As an attempt to globally prioritize conservation efforts, the **‘megadiversity’ concept** was developed (Mittermeier, 1988). **Mexico**, the study area of this thesis, is one of today’s 17 so-called ‘megadiverse’ (Mittermeier, 1988) nations that possess the greatest species richness worldwide and therefore carry special responsibility for global biodiversity. Already in 1992, right after the *Earth Summit* in Rio de Janeiro, Mexico established the *National Commission for the Knowledge and Use of Biodiversity* (*Comisión Nacional para el Conocimiento y Uso de la Biodiversidad*, CONABIO), which celebrates its 20th anniversary this year. The mission of CONABIO is to collect biodiversity data and to promote, coordinate, and carry out activities related to increasing awareness of biodiversity and its sustainable use. Just like worldwide, the greatest threat to biodiversity in Mexico is the loss of habitats due to increasing human land consumption especially in the last four to five decades (Toledo and Ordóñez, 1993; Sarukhán et al., 2010). Besides the immediate loss of habitats, land conversion also causes habitat fragmentation which reduces total genetic variation, hampers dispersal, and may impair key biotic interactions with pollinators and dispersal agents for plants (Skidmore et al., 2011).

Despite the multiplicity of efforts, our knowledge of the identity and the distribution of species and biodiversity is still incomplete and taxonomically biased (Ramamoorthy et al., 1993; Mace et al., 2005). While most estimates of the total number of species range from 5 million to 30 million (Mace et al., 2005), only 1.7 million species have been formally described so far (IUCN, 2011). Especially the knowledge of the spatial distribution patterns of species at global, regional, and even local scales is insufficient, a problem referred to as the *Wallacean shortfall* (Lomolino, 2004; Whittaker et al., 2005). In recent years, several novel data, tools, and methods have been initiated and developed for effectively predicting and monitoring species distributions as summarized in the following.

1.2 Novel data, tools, and methods: Species data portals, species distribution models, and remote sensing

Biogeographical research on species distributions has largely benefitted from the increasing concern about the impact of human disturbance on biodiversity as outlined in the previous section. In addition, recent advances in information, computer, and *Global Positioning System* (GPS) technology have opened up a new dimension for predicting spatial patterns of species distributions. The current ‘state of the art’ which forms the basis for this thesis is summarized in the following.

Species data portals. One of the key steps in the global-scale assessment of species distributions has certainly been the establishment of online databases of species occurrence records and biodiversity information such as the *Global Biodiversity Information Facility* (GBIF, http://www.gbif.org). GBIF was initiated in 2001 by several governments (1) to join existing activities and programs, (2) to facilitate the digitization of the enormous amount of biodiversity data already collected, and (3) to encourage the free and open access via the Internet (Edwards et al., 2000). By today, the GBIF network has become by far the largest portal of primary occurrence data (comprising information on taxon, locality, and date) with 319,699,379 indexed records (February 04, 2012).

Particularly for Mexico, CONABIO has established a national inventory of species occurrence records, the *National Information System on Mexican Biodiversity* (SNIB, *Sistema Nacional de Información sobre Biodiversidad de México*), with approximately six million indexed records. All geo-referenced species data collected in the context of projects that are carried out in cooperation with CONABIO as well as existing data from 190 Mexican and 240 international biological collections have been integrated into the

SNIB database under supervision of a designated database curator to prevent erroneous data. In addition, for Mexican plants a *National Forest Inventory* (*Inventario Nacional Forestal Inventario Nacional Forestal y de Suelos*, INFyS) was carried out by the *National Forestry Commission* (CONAFOR, *Comisión Nacional Forestal*) for a total of 24,659 sites between October 2004 and November 2007. Mexico therefore today belongs to those countries with a comparatively good availability of species collection data (i.e. occurrence points).

Species distribution models. As the availability of primary species observational data increased, the capacities of the computing and analytical tools to make use of this information amplified as well (Graham et al., 2004). While most of the early research that described biological patterns in terms of their relationships with environmental gradients (Grinnell, 1904; Murray, 1866) was largely qualitative, the first computer-based approaches within *Geographical Information Systems* (GIS) applied simple geographic or climatic envelopes to quantify species distributions (Nix, 1986; Box, 1981). Modern quantitative species distribution modeling began when the “two parallel streams of research activity converged” (Elith and Leathwick, 2009: p. 679): concepts from natural history, biogeography, and ecology were combined with the more recent developments in statistics, computing capacity, and information technology (Elith and Leathwick, 2009). So-called *Species distribution models* (SDMs) evolved which are numerical tools that describe empirical correlations between observations of species occurrence and environmental variables (Elith and Leathwick, 2009; Franklin, 2009). Based on statistically derived response surfaces (Guisan and Zimmermann, 2000), SDMs extrapolate *in situ* observations both in space and time to obtain spatially explicit and continuous surfaces indicating the probability of species occurrence (Franklin, 2009). This quantification of species-environment relationships via species distribution models is the most effective way to maximize the information content of the enormous information gathered in species point data collections.

In the last two decades, SDMs have been established as important tools in both basic and applied science (Franklin, 2009) and have been utilized to address various issues in ecology, biogeography, evolution, and – more recently – also in conservation biology and climate change research. Several different algorithms for predictive distribution modeling have been developed in the last years, with a trend towards complex machine-learning methods. In Chapter 3, the conceptual and methodological underpinnings of species distribution models as applied in this thesis are outlined in further detail.

Remote sensing. In addition to new algorithms for species distribution modeling and other methodological advances, more and more sources of environmental data with improved data quality have become freely available. Certainly, remote sensing variables today belong to the key data sources in environmental modeling. Remote sensing supports in particular the assessment of biophysical characteristics of landscapes, the detection of species habitats (and their loss), and the discovery of human-caused degradation up to the global scale (Strand et al., 2007; Mather, 2010). Biodiversity has been characterized based on remote sensing data at its multiple levels of ecological organization (individuals, species distributions, community compositions and species diversity) from local to global spatial scales (Franklin, 2009).

As remote sensors provide a systematic and consistent view of the earth at regular time intervals and comparatively low costs, they are useful for characterizing current species distributions as well as their temporal changes (Nagendra, 2001; Gillespie et al., 2008). Remote sensing “has heralded a new era in the ecological modelling of species habitats” (Buermann et al., 2008: p. 2) and is expected to receive even further attention from researchers in the field of species distribution modeling in the future (Gillespie et al., 2008). Given the dramatic environmental changes and the current biodiversity crisis, such reliable data on species distributions with high *effective* spatial resolutions (in comparison to interpolated climate data) are needed more than ever to make informed management decisions and to support sustainable conservation planning. Especially for Mexico and its unique biodiversity, it is impossible to conduct reliable and spatially explicit species assessments based on field work alone due to the enormous extent and geographical complexity of the country (Sarukhán et al., 2010).

The use of various remote sensing data for modeling species distributions “requires new rules and protocols” (Saatchi et al., 2008: p. 2015). The already recognized challenges concern (1) a more intensive utilization of continuous remote sensing data instead of satellite-derived land cover maps (Stickler and Southworth, 2008), (2) an assessment of the utility of spectral data or remote sensing products to quantify and map ecologically important features (Saatchi et al., 2008), and (3) detailed research on the effects of the spatial resolution of remote sensing data to delimit species requirements and model their distributions (Saatchi et al., 2008).

Especially during the last three to five years (that is, the period in which this thesis was accomplished), the number of studies making use of remote sensing data for modeling species distributions has multiplied. According to the published literature (see Chapter 4 for a review of the relevant studies), remote sensing predictors have been shown to **improve climatic species distribution models** at national to continental scales and to be sufficient in themselves to predict species occurrence from local to regional scales. However, the use of hierarchical frameworks to combine climatic and remote sensing based model predictions, which have been explicitly proposed for land cover data (Pearson et al., 2004), has as of yet not been implemented for remote sensing data. In addition, it was recognized that species distributions are determined not only by absolute levels of climate or vegetation variables, but also by subtle differences in their seasonality (Rogers et al., 2002). Most studies aimed at modeling species distributions thus have made use of **multi-temporal remote sensing data** (mostly time series of between two and four years). Especially multi-temporal vegetation index data have been utilized for modeling plant species distributions (Buermann et al., 2008; Prates-Clark et al., 2008; Saatchi et al., 2008; Evangelista et al., 2009; Tuanmu et al., 2010). Further, it has been shown that species distribution models developed from **phenological metrics** instead of the original vegetation index time series in general exhibit higher temporal transferability (Tuanmu et al., 2011). Finally, the existing literature has indicated that natural **inter-annual variability** in remotely sensed time series may impact temporal transferability and hence the validity of species distribution models (Tuanmu et al., 2011).

Due to the novelty of this field of research, the published literature consists mostly of selective case studies that examine certain aspects of an exemplary nature. What is missing are a sensitivity analysis and a systematic examination (e.g. regarding different remote sensing products or species traits) for modeling species distributions by means of remote sensing data. This research gap is addressed by this thesis.

1.3 Objectives and outline

The overall research objective of this dissertation is to develop a **systematic framework** of rules and standards for modeling species distributions which integrates **climate and multi-temporal remote sensing data** (Figure 1.2). In addition to the aspects related to modeling species distributions, the framework and findings are further transferred to predict spatial patterns of **species richness**, as "[t]he obvious extension to modeling individual species' distributions is to combine them to create maps of richness" (Stockwell and Peterson, 2003).

Beyond the selective case studies of an exemplary nature described in the previous section, there is a great potential in remote sensing data for modeling species distributions that has yet to be explored. This study takes up the approaches found in the recent literature (see Section 1.2). Methodological aspects are further developed with regards to several critical questions, some of which have already been raised by the latest studies (e.g. Tuanmu et al., 2011). The remaining gaps of knowledge (e.g. the selection of remote sensing predictors, the influence of inter-annual variability, and the comparison of continuous remote sensing and categorical land cover data) are explicitly investigated. Furthermore, the analysis of remote sensing variables in species distribution modeling still lacks a conceptual underpinning, a problem that is explicitly addressed in this thesis. In this sense, the interdisciplinary approach aims at helping to bridge the 'perception problem' between remote sensing researchers and biologists outlined by Turner et al. (2003) that has often hampered the use of remote sensing data in species distribution modeling studies.

The research conducted in this thesis addresses Mexican tree species distributions and tree species richness. **Mexico** has been selected as study area since (1) the country exhibits a great environmental and biological diversity that is reflected in an enormous variety of ecological processes and high levels of species richness and endemism (Sarukhán et al., 2010), (2) there is a unique availability in Mexico of species occurrence data from different sources, and (3) the work could build upon a long-standing cooperation with CONABIO, the *Mexican National Commission for the Knowledge and Use of Biodiversity*. This thesis focuses on modeling plant (in particular tree) species distributions due to several reasons. For Mexican plants, two independent sources of data, namely the *National Information System on Mexican Biodiversity* (SNIB) and the *National Forest Inventory* (INFyS) are available. Plants also belong to the groups with the highest proportion of species that have been formally described (Groombridge and Jenkins, 2002) and have a high probability of detection in the field. In addition, they are relatively slow dispersers compared with many other taxonomic groups (e.g. birds, mammals) and thus exhibit higher persistence in the landscape with less data errors

(Griffiths et al., 2000). Further, mapping vegetation as species assemblages or communities has a long tradition in remote sensing (Franklin, 1995; Xie et al., 2008). The twelve **tree species** studied in this thesis are representative of the major Mexican forest types and were chosen to capture a wide variety of ecological traits such as range size or biotope specificity. **Forests** belong to the vegetation types in Mexico with the highest species numbers and are particularly threatened by extensive transformation of natural ecosystems for agricultural, urban, and infrastructure activities (Burgos and Maass, 2004; Ricker et al., 2007; Sarukhán et al., 2010). In addition to this representative selection of tree species across several forest types, distribution ranges of all species of the tree genus ***Pinus*** (pines, family Pinaceae) are modeled. Mexico has the greatest number of pine species of any region of similar size in the world (Farjon, 2010) and Mexican pine-oak forests were recently added to the list of the 34 biologically richest and at the same time most endangered terrestrial ecoregions worldwide (Mittermeier et al., 2005). However, little is known of the exact distribution ranges of these pines, not to mention the spatial patterns of pine richness which are assessed in this thesis.

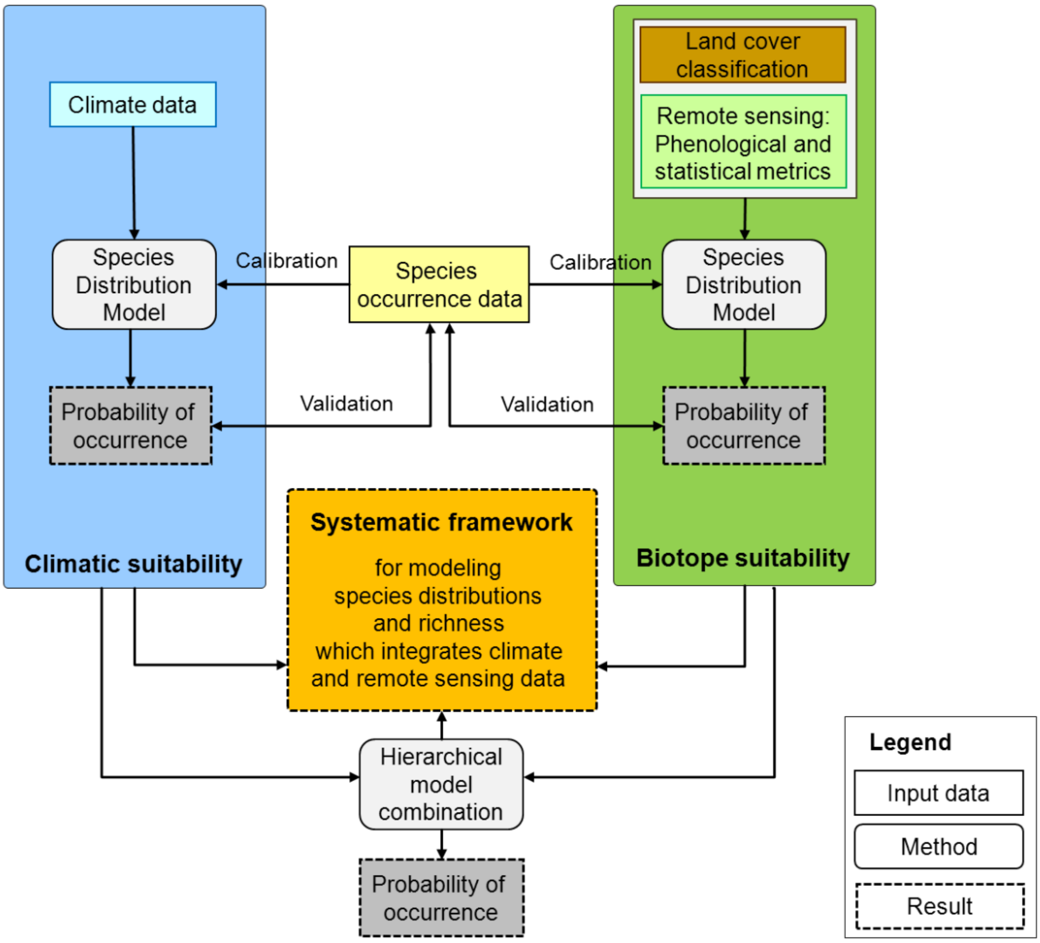


Figure 1.2: Overview of the conceptual framework and objectives of this thesis. The hierarchical modeling framework utilizes species occurrence data in combination with species distribution models to predict climatic suitability based on climate data and to predict biotope suitability based on remote sensing variables (or land cover data). At all stages of the hierarchical modeling framework, species traits (e.g. range size and biotope characteristics) are explicitly considered to develop a systematic framework for modeling species distributions in order to support reliable modeling of species distributions based on remote sensing data in the future.

In this thesis, a more comprehensive set of remote sensing predictors than in the existing literature is produced and utilized for species distribution modeling. This includes both the number of variables assessed and the length of the remotely sensed time series considered. For this purpose, continuous **remote sensing time series** are generated on the basis of freely available Terra-MODIS data for the time period of 2001 to 2009. In the last years, MODIS has emerged as one of the most important satellite time series sensors for monitoring vegetation phenology (Zhang et al., 2003; Xiao et al., 2006). As understanding and predicting species distributions across large spatial scales is highly important, the remote sensing data used cover the entire Mexican national territory. The **specific remote sensing products** utilized in the analyses have been selected based on the suitability of the products to characterize species distributions according to previous studies as well as on their relevance for ecosystem functioning in general. In particular, two different Terra-MODIS Land 16-day composite standard products (MOD13A2 and MOD11A2) including the *Enhanced Vegetation Index* (EVI), *Reflectance Data*, and *Land Surface Temperature* (LST) are analyzed. Low-quality remote sensing observations are temporally interpolated using the TiSeG software (Colditz et al., 2008) to generate enhanced and continuous time series. One of the main methodological areas is the computation of phenological and statistical metrics that characterize *Land Surface Phenology* (LSP) and are meaningful for species distributions. Moreover, the different sample densities as well as the spatial autocorrelation of interpolated climatic surfaces and remote sensing data are explicitly addressed.

Given the interdisciplinary character of this thesis, another core theme is the theoretical background and application of **species distribution models**. This implies a thoughtful examination of all aspects concerning the *ecological* model, *data* model, and *statistical* model (Austin, 2002). Important aspects are environmental variables representing those factors controlling species distributions, a statistical model linking species occurrences to the environmental predictors, methods for applying the model (e.g. coefficients, thresholds), and protocols for evaluating the error and uncertainty in the predictions. In addition, a detailed justification for selecting the *Maximum Entropy* algorithm (Phillips et al., 2006) used for species distribution modeling in this thesis is given.

In particular, this thesis addresses the following research questions and undertakes the corresponding **novel analyses**:

(1) **Hierarchical modeling framework** – How can the synergistic analysis of climate and remote sensing data in species distribution models be improved?

(2) **Categorical or continuous?** – Does the suitability of continuous remote sensing time series for modeling species distributions exceed an existing categorical land cover classification?

(3) **Which remote sensing data for which species?** – Which are the major criteria for selecting remote sensing data depending on species characteristics?

(4) **How many data?** – How does inter-annual variability in remotely sensed time series impact the predictive performance of species distribution models?

(5) **From species to species richness?** – Can the framework for modeling individual species distribution ranges also be transferred to predict species richness?

To guide the way for modeling species distributions regarding the aspects outlined above, specific individual studies are designed and accomplished. The methodological findings are intended to support reliable modeling of species distributions based on remote sensing data in Mexico and beyond. From the conservation perspective, the research and developments that are made can also be beneficent to biodiversity inventories and continuous monitoring as well as sustainable conservation planning. In this sense, remote sensing is not intended to substitute field data collections, but as a supporting method to investigate species distributions and biodiversity more accurately across large scales.

According to the research objectives outlined above, this thesis is divided into seven additional chapters. This introductory chapter is followed by the description of the study area Mexico with a special focus on climate as well as biodiversity and vegetation characteristics (Chapter 2, *Study area*). In Chapter 3 (*Species Distribution Models – Theoretical framework*), the theoretical background and conceptual underpinnings of modeling species distributions are presented. These include relevant terms and definitions, a detailed overview of concepts (e.g. ecological niche, geographical and environmental space), and methodological aspects (e.g. spatial autocorrelation, sampling bias, and threshold selection). All methodological issues are presented and discussed with respect to the objectives of this thesis. In addition, the *Maximum Entropy* algorithm applied in this thesis is compared with other methods available for species distribution modeling. The chapter concludes with a summary of typical applications of species distribution models in research and conservation illustrated by exemplary studies. The next chapter (Chapter 4, *Remote sensing for modeling plant species distributions and richness – State of the art and new challenges*) targets the different perceptions of environmental modelers as compared to remote sensing researchers and aims to provide a differentiated view regarding the potential of remote sensing data for modeling plant species distributions. This includes for example the transition from species mapping to distribution modeling and the ecological relevance of the various remotely sensed products analyzed in this thesis. In the following (Chapter 5, *Model input data and pre-processing*), the characteristics of the model input parameters used in this thesis, including species occurrence samples, climate variables, remote sensing data, and land cover parameters, are summarized. Chapter 6 (*Species distribution modeling with remote sensing data*) presents the findings of this thesis regarding (1) the synergistic analysis of climate and remote sensing data in species distribution models in a hierarchical framework and (2) the effects of different remotely sensed input variables on predictive model performance. Finally, based on the results described in Chapter 6, the framework developed to model individual species distribution ranges is transferred via a *stacked species distribution modeling* approach to predict species richness (Chapter 7, *From species to species richness: A case study for Mexican pines*). The results of the thesis are summarized in Chapter 8 (*Synthesis and outlook*) and discussed with regard to future research questions, remaining challenges, transferability, and their usefulness for conservation issues.

2 Study area

Mexico exhibits a great environmental diversity which is reflected in high levels of species richness and endemism, as mentioned in the previous section; it is therefore a unique region for studying species distributions. Barthlott et al. (2007) summarized this heterogeneity of geology, topography, soils, and climate as ‘geodiversity’ and identified it as one of the driving mechanism for habitat differentiation and hence species richness. This chapter aims to introduce the country’s key data and make familiar with both its physical-geographic and socio-economic conditions. With a size of 1,964,380 km² (FAO, 2011), Mexico is 5.5 times as large as Germany and borders the United States in the north and both Guatemala and Belize in the south. Politically, Mexico is divided into 31 states and a federal district (*Distrito Federal*, D.F.) which serves as the capital of the country. Mexico’s coastline has a total length of approximately 9,330 km (CIA, 2011) bordering the *North Pacific Ocean*, the *Caribbean Sea*, and the *Gulf of Mexico*. The country features a very complex geological setting (Section 2.1) and has a strong climatic north-south gradient (Section 2.2) as a result of its large latitudinal extension (from 32°43’ North to 14°32’ North). Due to its environmental heterogeneity and location on the boundary between two biogeographic regions, Mexico belongs to the global biodiversity hotspots (Section 2.3) and holds a unique flora (Section 2.4). However, this exceptional ‘natural capital’ (Sarukhán et al., 2010) is threatened by increasing human impact and land use/land cover change (Section 2.5).

2.1 Geology, topography, and soils

Mexico’s geological history has been one of the major evolutionary forces (Miranda and Hernández, 1963; Ramamoorthy et al., 1993) and is one of the main reasons for its remarkable phytodiversity (Cevallos-Ferriz and González-Torres, 2005). The country’s complex geological setting with its broad mountain ranges, large plateaus, and coastal lowlands can be explained by four major processes (Cevallos-Ferriz and González-Torres, 2005): (1) the formation of large magmatic provinces (*Sierra Madre Occidental* or *Trans-Mexican Volcanic Belt*), (2) the fragmentation and displacement of large continental segments which e.g. led to the opening of the *Gulf of California*, (3) marine regression that outlined the current shoreline, and (4) the folding of the eastern orogenic belt called *Sierra Madre Oriental*. Therefore, the whole spectrum of rock types can be found, namely sedimentary, volcanic, intrusive, and metamorphic rock (Ferrusquía-Villafranca, 1993).

The country is divided into eleven morpho-tectonic provinces (Figure 2.1) with distinct physical-geographic and geologic-tectonic features (Ferrusquía-Villafranca, 1993), namely (1) *Baja California Peninsula*, (2) *Northwestern Plains and Sierras*, (3) *Sierra Madre Occidental*, (4) *Chihuahua-Coahuila Plateaus and Ranges*, (5) *Sierra Madre Oriental*, (6) *Gulf Coast Plain*, (7) *Central Plateau*, (8) *Trans-Mexican Volcanic Belt*, (9) *Sierra Madre del Sur*, (10) *Sierra Madre de Chiapas*, and (11) *Yucatan Platform* (Figure 2.1). A detailed description of all geomorphological features is given in Sommerhoff and Weber (1999).

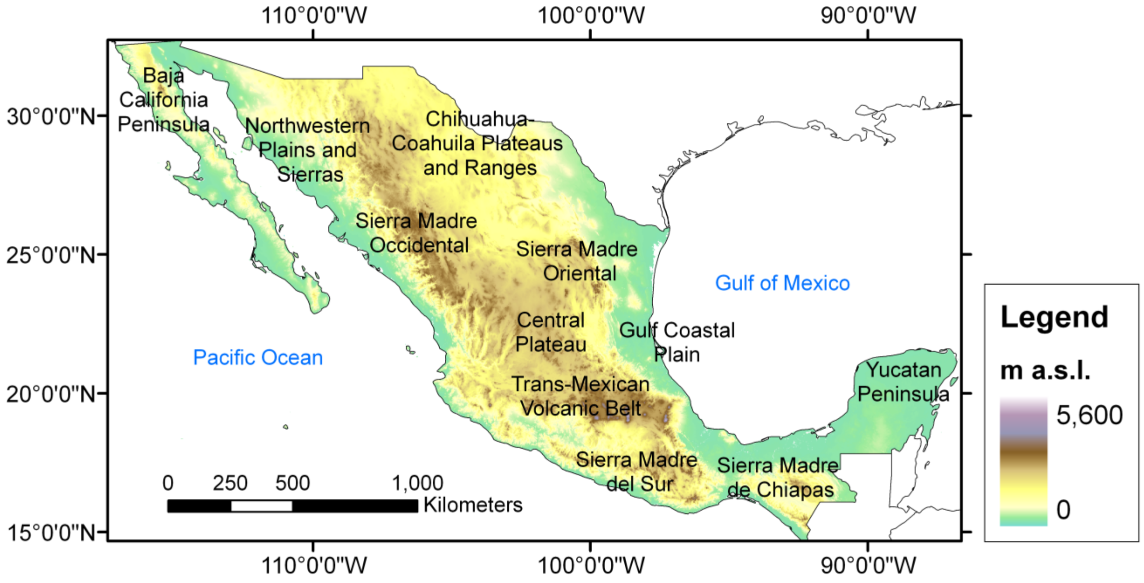


Figure 2.1: Major morpho-tectonic provinces of Mexico as described by Ferrusquía-Villafranca (1993). Topographic information was taken from the WorldClim data set (Hijmans et al., 2005).

Almost the whole Mexican territory is part of the *North American Plate*; only small portions of the *Baja California Peninsula* are assigned to the *Pacific* and *Cocos Plates* (Sommerhoff and Weber, 1999). Mexico’s west coast also belongs to the circum-Pacific ‘Ring of Fire’, an area of high tectonic, seismic, and volcanic activity (Sommerhoff and Weber, 1999) which has significantly contributed to the formation of the major mountain ridges in the country. Two-third of the country lies above 800 m a.s.l. (Ferrusquía-Villafranca, 1993) and the relief is in general very steep and dominated by several mountain ranges (*Sierras*) as shown in Figure 2.1. The highest peaks of the country are located in the *Trans-Mexican Volcanic Belt*, namely the *Pico de Orizaba* (5,650 m), the *Popocatepetl* (5,450 m), and the *Ixtaccihuatl* (5,280 m). Much of northern and central Mexico is covered by the *Central Plateau* and the *Chihuahua-Coahuila Plateaus and Ranges* with altitudes between 1,800 and 2,300 m a.s.l. (Sommerhoff and Weber, 1999). Large plains are located along the Pacific coast and the *Gulf of Mexico*.

According to SEMARNAP (1998) the most important soil types in the country are Leptopsols (24.0% of the area), Regosols (18.5%), and Calcisols (18.2%). Leptosols are azonal, very shallow soils over continuous rock and particularly common in mountainous areas (International Union of Soil Sciences, 2006). Regosols are weakly developed mineral soils and typically occur in eroding lands in mountainous terrain and under (semi-)arid conditions (International Union of Soil Sciences, 2006). Calcisols are especially dominant in the northeastern states of *Nuevo León*, *San Luis Potosí*, and *Zacatecas* (SEMARNAP, 1998). These soil types are common in calcareous parent material, represent soils with significant secondary accumulation of lime, and are typical of (semi-)arid environments (International Union of Soil Sciences, 2006).

2.2 Climate

In addition to geology and topography, also large-scale climatic gradients contribute to the great variety of environmental conditions in the country. Mexico is divided into a tropical (southern) and a sub-tropical (northern) part along the *Tropic of Cancer* (Sommerhoff and Weber, 1999). While the north experiences cooler air temperatures during winter, temperatures are fairly constant throughout the year in the southern parts and vary primarily as a function of altitude. The country's climate is mainly influenced by the following climatic circulation systems (Cavazos and Hastenrath, 1990; Sommerhoff and Weber, 1999): During the boreal winter (corresponding to the dry season), the *Inter-tropical Convergence Zone* (ITCZ) is far south and the upper tropospheric circulation is dominated by the sub-tropical westerly jet (Cavazos and Hastenrath, 1990). Due to the north-south orientation of the major mountain ranges, so-called *Nortes* (mid-latitude storms with cold, polar air) can penetrate far southward over central Mexico (Cavazos and Hastenrath, 1990). During the boreal summer, the sub-tropical westerly jet disappears and Mexico is affected by the *Passat circulation* which brings precipitation to most parts of the country. These dynamics lead to pronounced annual precipitation seasonality while the length of the rainy season decreases from south to north (Sommerhoff and Weber, 1999). Over 80% of the annual precipitation falls during the rainy season between May and October (Giddings et al., 2005). February and July are generally the driest and wettest months, respectively (Giddings et al., 2005). The sole exception is the *Baja California Peninsula* with Mediterranean climate conditions and precipitation mainly in winter (Sommerhoff and Weber, 1999).

Precipitation. The complex orographic structure of Mexico (Section 2.1) further contributes to the great diversity of rainfall regimes from local to regional scales (Cavazos and Hastenrath, 1990). The resulting humidity zones in the country are shown in Figure 2.2. As a result of orographic rainfall, annual precipitation amounts reach up to more than 5,500 mm along the mountain slopes in *Tabasco* and *Chiapas* (Velázquez et al., 2000). As opposed to this, precipitation rarely exceeds 400 mm/year in the northwestern and 600 mm/year in the northeastern areas near the US border (Magaña and Conde, 2000). Mainly related to El Niño/Southern Oscillation (ENSO) dynamics there is a large inter-annual variability in precipitation as well (Magaña and Conde, 2000). Large parts of the territory (35.4%) are exposed to extended droughts that have affected agricultural and livestock activities in the last two decades (Magaña and Conde, 2000), with the most heavily impacted areas being the regions of the *Baja California Peninsula*, *Chihuahuan Desert*, and the *Sierra Madre Occidental* (Sarukhán et al., 2010).

Temperature. Topography does not only have an impact on precipitation patterns, but also on temperature regimes. Based on annual mean temperatures, several temperature zones from *caliente* (hot) to *nevada* (nival) can be identified (see also Figure 2.8). Annual temperature ranges increase from south to north with less than 5 °C along the tropical Pacific and Carribean coast to more than 18-20°C in the northern interior (Sommerhoff and Weber, 1999). The highest average temperatures (28-30°C) are recorded in the *Balsas Depression* (located in the southwest, descending down to 200 m a.s.l.), while the nearby area around the top of *Pico de Orizaba* (in *Veracruz*) with an altitude of 5,650 m has the lowest mean temperature with -6°C (Velázquez et al., 2000). Mexico City located at 2,300 m a.s.l., has a semi-arid climate with an annual mean temperature of 15.6°C. For May, the warmest month, average daily highs and lows in the capital are 26°C and 12°C and for January, the coldest month, 19°C and 6°C.

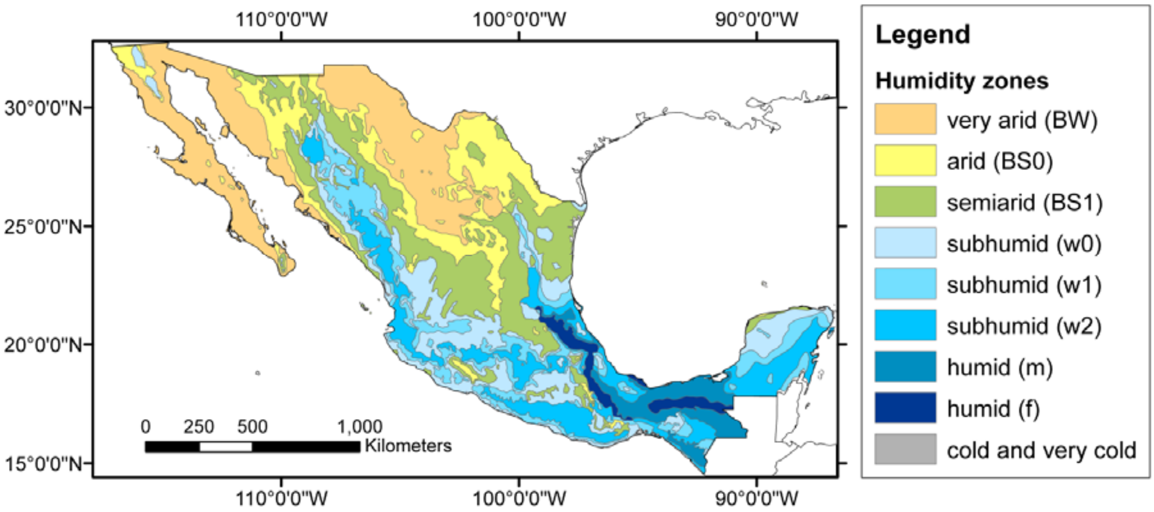


Figure 2.2: Humidity zones of Mexico (García, 1998), data available from CONABIO's online portal for geoinformation (CONABIO, 2011).

2.3 Mexico's biodiversity

In the following, a detailed description of Mexico's outstanding biodiversity is given to emphasize its importance as one of the global biodiversity hotspots (Section 2.3.1). Further, the major gradients and drivers of biodiversity and endemism in the country are highlighted and discussed (Section 2.3.2). Primarily due to human activities, we currently observe the phenomenon of an enormous biodiversity loss all over the planet (see Section 1.1). In this light, the major threats to biodiversity relating to Mexico are shortly described (Section 2.3.3). The existing network of protected areas in the country is characterized and discussed regarding its effectiveness for conserving the national biodiversity (Section 2.3.4).

2.3.1 Global context

Mexico is one of today's 17 so-called 'megadiverse' (Mittermeier, 1988) countries and hence belongs to a group of nations that possess the greatest species richness of animals and plants worldwide. The 'megadiversity' concept was first developed by Russell Mittermeier together with *Conservation International* (Mittermeier, 1988) as an attempt to prioritize global conservation efforts with a focus on prime habitats such as tropical rain forests and coral reefs. Altogether, the selected 'megadiverse' countries cover less than 10% of the global land surface but support nearly 70% of global species richness. Mexico occupies approximately 1.4% of the total land surface, but 10-12% of the recognized species (Sarukhán et al., 2010). Table 2.1 compares Mexico's position within the ten highest ranked nations of the 'megadiverse' countries. Mexico was further designated as a center of agricultural origin (Harlan, 1975; Sarukhán et al., 2010) meaning that many of the crop plants today used by mankind have their geographic origins here. Prominent examples are *Zea mays* (corn) which was domesticated approximately 5,000 years ago (Sluyter and Dominguez, 2006) or the legume genus *Phaseolus* that is used as a source of plant protein.

Table 2.1: Placing of Mexico's biodiversity (indicated by species richness) within the ten highest ranked nations of the 'megadiverse' countries (Llorente-Bousquets and Ocegueda, 2008). For each taxonomic group, countries with a higher rank than Mexico are highlighted grey; species numbers are absolute and not normalized by country area.

	Country	***Vascular plants***	***Mammals***	***Birds***	***Reptiles***	***Amphibians***
	Brazil	56,215	578	1,712	630	779
	Colombia	48,000	456	1,815	520	634
	China	32,200	502	1,221	387	334
	Indonesia	29,375	667	1,604	511	300
	Mexico	**23,424**	**535**	**1,096**	**804**	**361**
	Venezuela	21,073	353	1,392	293	315
	Ecuador	21,000	271	1,559	374	462
	Peru	17,144	441	1,781	298	420
	Australia	15,638	376	851	880	224
	Madagascar	9,505	165	262	300	234
Rank of Mexico		5	3	8	2	5

The global importance of Mexico regarding plant diversity was also highlighted by Barthlott et al. (2005) in their area-adjusted map of vascular plant richness (Figure 2.3). After Barthlott et al. (2005), Mexico features not only high absolute species numbers but also outstanding scores of plant species richness per unit area (4,000 to 5,000 species per 10,000 km²). Mexico is thus close to the top five regions with more than 5,000 species per 10,000 km² which are *Costa Rica-Chocó*, the tropical *Eastern Andes*, *Atlantic Brazil*, *Northern Borneo*, and *New Guinea* (Barthlott et al., 2005). However, species diversity is unevenly distributed within the country; the most relevant biodiversity gradients in Mexico are summarized in the following section.

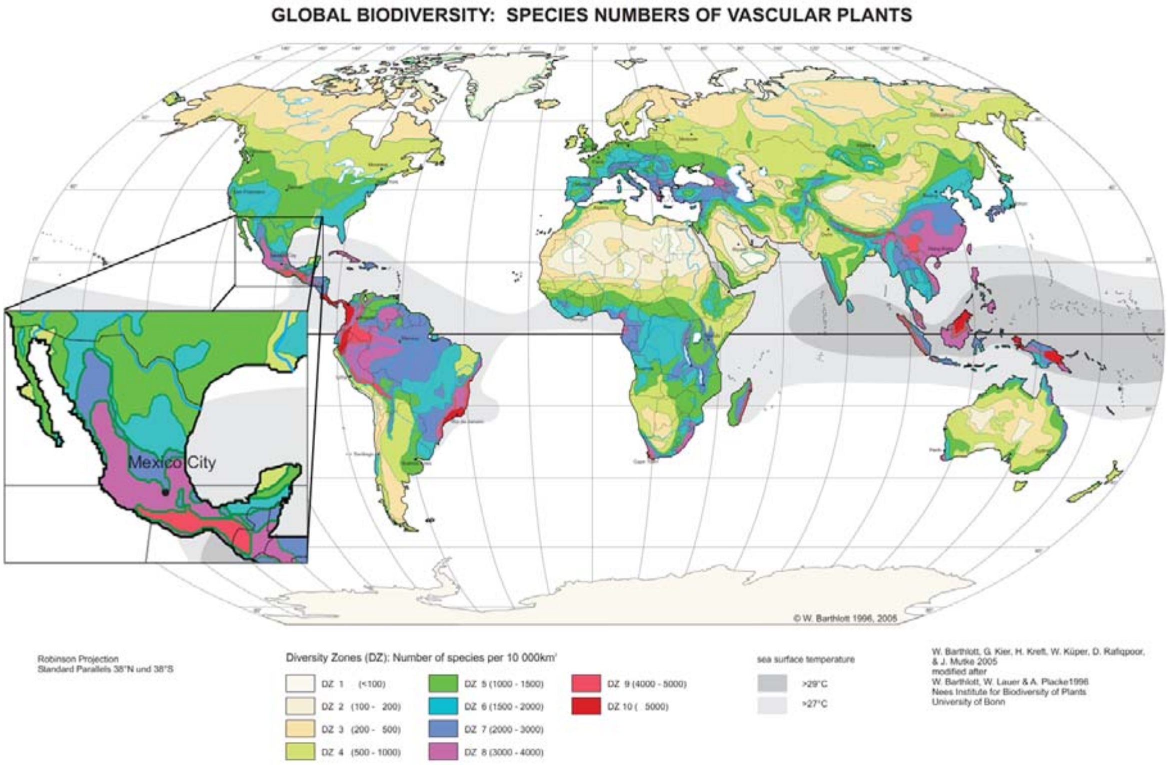


Figure 2.3: Global diversity zones measured by the species numbers of vascular plants per 10,000 km² (Bartlott et al., 2005; modified).

2.3.2 Spatial patterns of biodiversity and endemism

Not only total species numbers but especially their taxonomic composition, phylogenetic relations between species, and spatial patterns of distribution are of interest. Mexico is located on the boundary between two globally recognized biogeographic regions, the *Nearctic* and *Neotropical* (Barthlott et al., 2007) which means that both northern genera, which reach their southern limit (e.g. *Ulmus*), and southern genera, which reach their northern limit (e.g. *Pouteria*), overlap (Ricker et al., 2007). The southern part of the country is further located at the confluence of two migration routes through Central America and the Antilles which have been important through the Tertiary (Graham, 1993). In general, Mexico therefore has more species in its tropical than in its temperate parts (Sarukhán et al., 2010) – although there are some exceptions. For example, the Yucatan peninsula has a relatively poor flora (Figure 2.3, Rzedowski, 1993).

Biogeographical research further classifies into so-called 'ecoregions' which are areas that "contain a geographically distinct assembly of natural communities" (Saruhkan et al., 2010: p. 29), share the majority of species and taxa, and have similar environmental conditions and ecological dynamics. Several (principally consistent) classification systems of ecoregions or ecological zones have been proposed for Mexico based on the distribution of the two elements vegetation and climate (Toledo and Ordóñez, 1993; Morrone et al., 2002). The multi-level scheme adopted by the *Commission for Environmental Cooperation* (CEC, 1997; Wiken et al., 2011) identifies the following seven *Level I ecoregions* for Mexico: *Great Plains*, *Mediterranean California*, *North American Deserts*, *Southern Semiarid Highlands*, *Temperate Sierras*, *Tropical Dry Forests*, and *Tropical Wet Forests* (Figure 2.4). Ecoregions are important indicators of geographical limits of species mobility and thus affect species distribution ranges. For example, the mountainous areas of the *Temperate Sierras* ecoregion have acted as barriers to species dispersal, resulting in considerable differences in species composition between the Pacific and Carribean lowlands (Mittermeier et al., 2005).

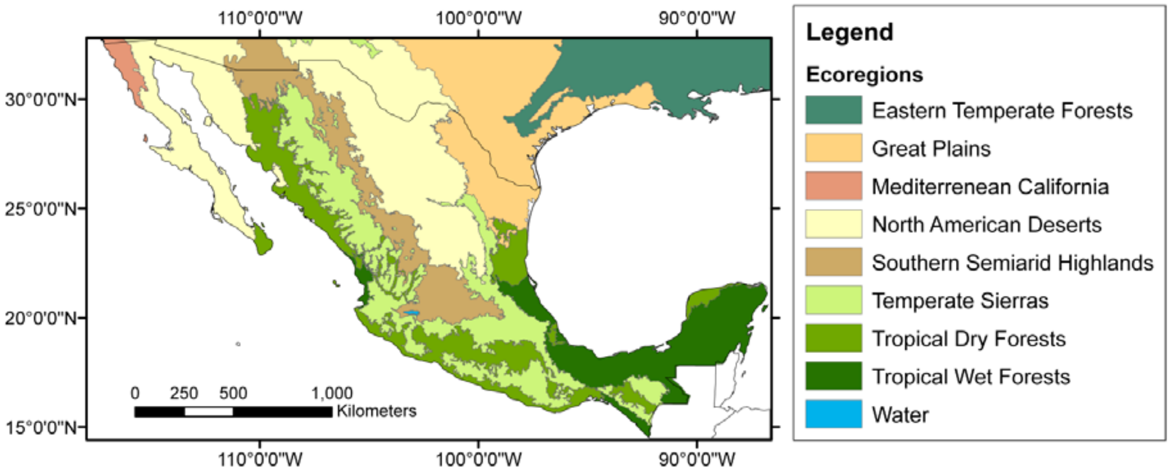


Figure 2.4: Level I ecoregions of Mexico and southern USA based on the multi-level scheme by the *Commission for Environmental Cooperation* (1997).

In addition, Mexico is not only known for its high overall number of species, but also for the wealth of endemic species (i.e. with distribution ranges confined to the country). Over 300 genera and about 50-60% of the species in Mexico are endemic (Ramamoorthy et al., 1993). The spatial distribution of endemism does not follow the same patterns as those of diversity (Sarukhán et al., 2010), for instance the tropical south is rich in species but comparatively poor in endemics (Rzedowski, 1993). In general, areas with specific environmental conditions that are distinct from their surroundings such as mountainous areas, e.g. the *Trans-Mexican Volcanic Belt* or the *Sierra Madre del Sur* (Ramamoorthy et al., 1993), or valleys, e.g. the Tehuacán-Cuicatlán region (Dávila et al., 2002), are considered as the major centers of endemism. Other important areas include the *Baja California Peninsula* as well as some of the distant islands, e.g. *Isla Guadalupe*, and some ‘ecological islands’ such as the *Balsas Depression*, highest mountain peaks, and regions that acted as refugia during the climatic changes of the Pleistocene (Rzedowski, 1993).

2.3.3 Threats to biodiversity

The global ‘biodiversity crisis’ discussed in the introductory chapter affects the Mexican species and taxa as well. According to the *NORMA Oficial Mexicana* NOM-059-

SEMARNAT-2010 (SEMARNAT, 2010), which is the recently updated official list of endangered species in Mexico, 2,606 species are classified as being threatened. These include 392 birds, 194 amphibians, 46 fungi, 49 invertebrates, 291 mammals, 204 fish, 443 reptiles, and 987 plants. In line with the global trend, the greatest threat to biodiversity in Mexico – across all ecosystems – is the loss of habitats (Toledo and Ordóñez, 1993), especially the deforestation of natural ecosystems for food production (Sarukhán et al., 2010). Current and future trends of land cover change and related habitat loss are discussed in more detail in Section 2.5.3 as these are the threats detectable by means of remote sensing and thus primarily concerned in this thesis. In addition to the loss of natural areas and as a response to indirect factors such as demographics, public policy, and technological developments (Sarukhán et al., 2010) other processes contribute to the changes in and the loss of biodiversity in Mexico (Figure 2.5). These include species overexploitation, the introduction of invasive species, environmental pollution, and climate change.

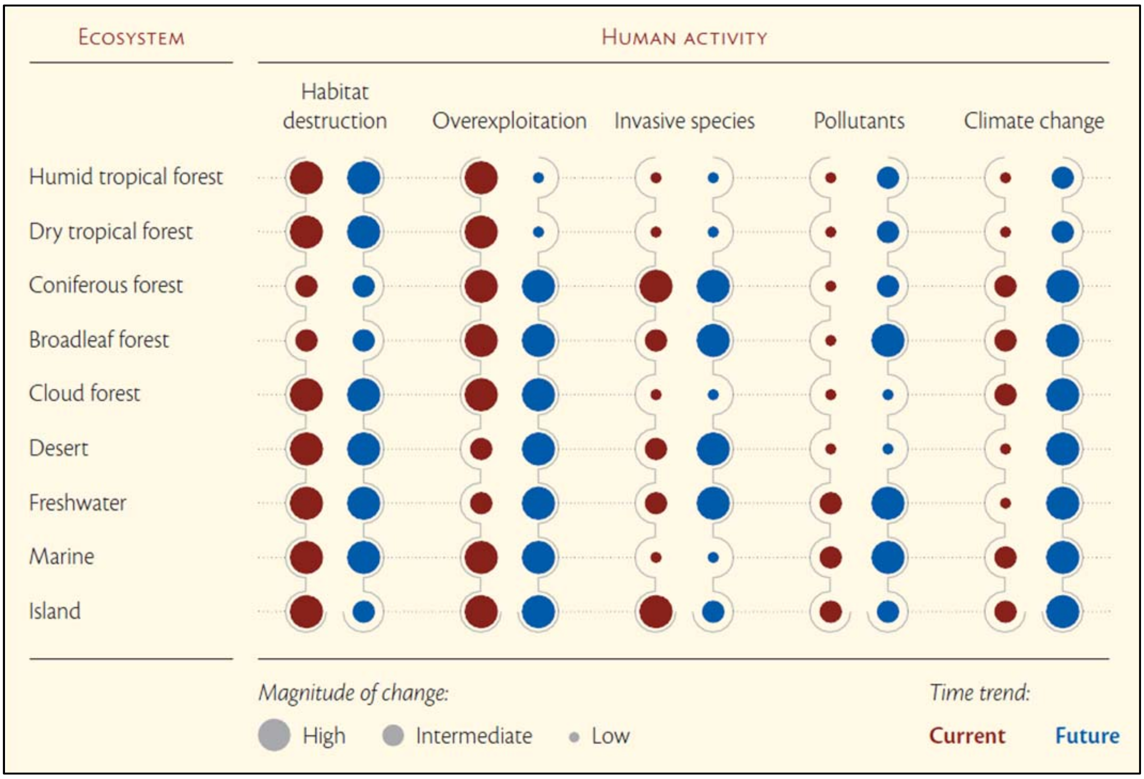


Figure 2.5: Major current threats to biodiversity in Mexico and their estimated future impact (Sarukhán et al., 2010). The magnitude of change for the different ecosystems is denoted by circles of different sizes.

Overexploitation of species is especially relevant in freshwater and marine environments (e.g. for fish or sea turtles) but has also been reported for plants (Olmsted and Alvarez-Buylla, 1995). Similarly, environmental pollution primarily affects aquatic habitats. Due to rising population numbers and related agricultural, infrastructural, and industrial activities, the impact of pollution is assumed to intensify in the future (Sarukhán et al., 2010). The introduction of invasive species is already, with an increasing trend, an important issue having detrimental effects on species richness. For Mexico, 721 invasive species – out of which 665 plants – have been reported (Sarukhán et al., 2010). Recently, the national strategy for the management of invasive species was published (National Advisory Committee on Invasive Species, 2010) and an information system on invasive species in Mexico (http://www.conabio.gob.mx/invasoras) was set up by CONABIO to facilitate exchange of information, discuss strategies for prevention and eradication of invaders, and define national priorities for counteractions.

The impact of climate change on overall biodiversity is very difficult to predict and different studies have aimed to obtain reliable estimates for certain taxa. Early on, it was concluded that out of 33 protected natural forest areas in Mexico only nine would remain in the same bioclimatic zone (Villers-Ruíz and Trejo-Vázquez, 1998). However, this assumed change in vegetation type does not necessarily imply large changes in species composition. For example, many of the same plant species found in the life form of trees in dry tropical forest, are also occurring as shrubs in thorny woodland (Rzedowski, 1978). Peterson et al. (2002) modeled distribution ranges of 1,870 Mexican species (birds, mammals, and butterflies) under changing climatic conditions. Their findings suggest that species extinctions and drastic range reductions are relatively rare, but species turnover in some areas very high (Peterson et al., 2002). Moreover, predictions depend very much on the climate change scenario used. This was confirmed by a study which modeled distribution ranges of cacti species in the *Tehuacán-Cuicatlán Biosphere Reserve* in central Mexico (Téllez-Valdés and Dávila-Aranda, 2003). This study found that under climate change scenarios distribution ranges of certain species were either drastically contracted, restricted only to the reserve, or remained similar in- and outside the reserve. Nevertheless, all threats mentioned have an impact on biodiversity and emphasize the need for reliable biodiversity monitoring.

2.3.4 Protected areas

Early on, the necessity to protect the country's unique biodiversity was recognized and led to the founding of CONABIO as long ago as 1992. The Mexican protected areas are today under the administration of the *National Commission of Natural Protected Areas* (*Comisión Nacional de Áreas Naturales Protegidas*, CONANP). They can be classified into different categories (Table 2.2) and various types of ownership: *public* (federal, state, and municipal), *social* (*ejido* - communal land shared by the people of the community), and *private*. Protected areas under social ownership comprise mainly regions that have long been kept with a low intensity of use. They often became protected to maintain water resources or for religious purposes, but also to conserve populations of particular species. Mexico's southern and coastal areas are also part of the trans-national *Mesoamerican Biological Corridor* (*Corredor Biológico Mesoamericano*) which was initiated in 1998 to protect critically endangered species by establishing a network between protected areas. The other participatory countries are Belize, Costa Rica, El Salvador, Guatemala, Honduras, Nicaragua, and Panama.

The proportion of protected areas compared to the national territory in Mexico is almost 13% (Table 2.2; compared to 3.6% in Germany for the year 2009 according to the *German Federal Agency for Nature Conservation*, BfN) whereas the effective protection status varies largely between the different types. About 65% of the area of federal protected areas have international recognition, e.g. as part of the *World Heritage Site* scheme or by inclusion as a biosphere reserve in the *UNESCO Man and the Biosphere Program* or the *Ramsar Convention on Wetlands* (Sarukhán et al., 2010). In 2005, protected areas were inhabited by almost 2.6 million people, equivalent to 2.5% of the total population, and 5.7% of the rural population (Sarukhán et al., 2010). The representation of indigenous groups in (often marginal) protected areas is comparatively high.

Table 2.2: Characteristics of protected areas in Mexico (CONANP; October, 2011).

Category	***Number***	***Area (ha)***	***Percentage of national territory***
Biosphere Reserves	41	12,652,787	6.44
National Parks	67	1,482,489	0.75
Natural Monuments	5	16,268	0.01
Protection areas of natural resources	8	4,440,078	2.26
Protection areas of flora and fauna	35	6,646,942	3.38
Sanctuaries	18	146,254	0.07
Total	174	25,384,818	12.91

A significant proportion of the protected areas in Mexico were established by circumstantial reasons, with little basis on scientific knowledge (Sarukhán et al., 2010) and many ecoregions are underrepresented. Further, a strong spatial bias exists since a greater proportion of highlands (higher than 2,800 m), which are less accessible to humans than lowlands, are protected (Figure 2.6). The current network of protected areas is hence not adequately designed to protect Mexico's biodiversity. Therefore, it has been claimed that in the future policy decisions should be based on reliable scientific data and meet the criteria relevant for the establishment of sustainable protected areas (Sarukhán et al., 2010). The findings of this thesis can effectively support such efforts.

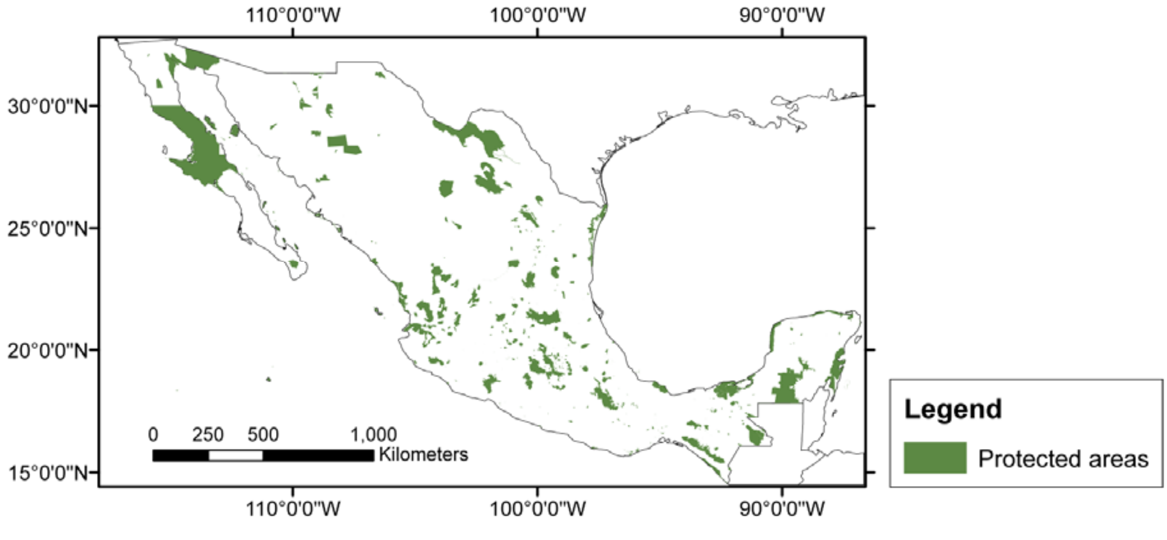


Figure 2.6: Distribution of protected areas in Mexico (IUCN and UNEP, 2010).

2.4 Vegetation characteristics

While so far biodiversity has been considered including all taxa, the following sections focus only on plants as this thesis addresses modeling distribution ranges of tree species. Mexico’s flora has approximately 30,000 plant species including 21,600 species and 2,500 genera of flowering plants (Rzedowski, 1993). In addition, Mexico’s vegetation includes a great variety of life forms. The main vegetation characteristics can be summarized based on the key botanical classification systems of Miranda and Hernández (1963) and Rzedowski (1978). In this section, first the potential vegetation (Section 2.4.1), describing pre-human conditions and thus indicative of potential plant species distributions, is characterized. These spatial distribution patterns of the potential vegetation can be resolved into hygrothermal altitudinal zones within the country which include all major vegetation communities but still ignore human modifications (see Figure 2.8). The second part (Section 2.4.2) focuses on current forest types, as reference for the biotopes that the target species of this thesis are associated with, and includes the effects of anthropogenic impact.

2.4.1 Potential vegetation and vegetation types

Based on his extensive studies on the Mexican vegetation, Rzedowski developed a map of ‘potential vegetation’ or ‘original cover’ (Ricker et al., 2007) in the country (Figure 2.7). The layer attempts to reconstruct pre-human vegetation type distributions based on the geological setting and the climatic regime. It “adequately reflects the vegetation types that would be available to species using primary habitats in the nineteenth and twentieth centuries” (Peterson et al., 2006: p. 230). The data set was already used as reference to quantify the loss of habitats (Peterson et al., 2006) and to estimate land cover change rates (Ricker et al., 2007). The percentage coverage of the potential distribution was estimated at 52.0% for forest, 38.5% for shrubs, 8.3% for natural grassland, and 1.2% for other coverage by Ricker et al. (2007).

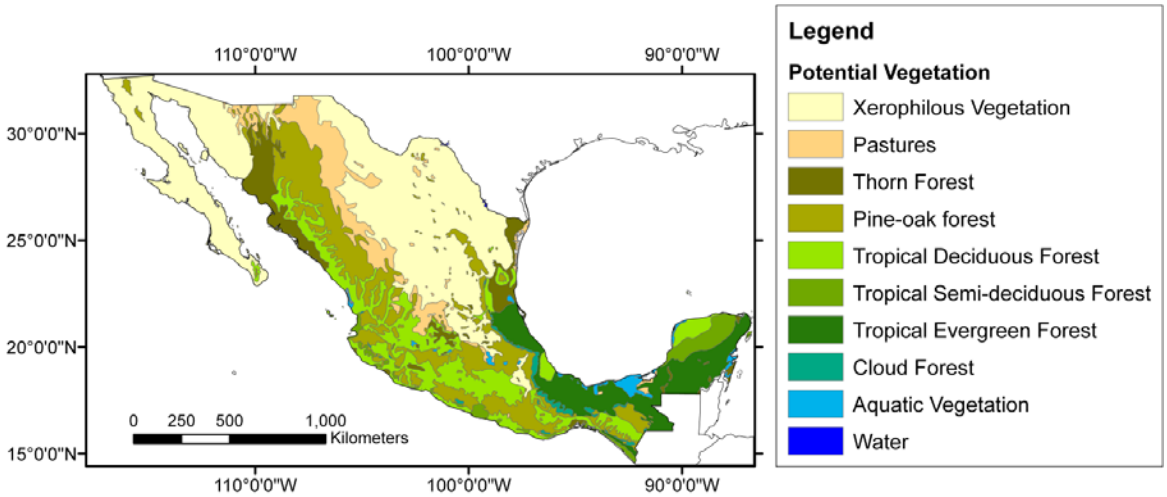


Figure 2.7: Potential vegetation in Mexico (Rzedowski, 1990), data available from CONABIO’s online portal for geoinformation (CONABIO, 2011).

The spatial distribution of the potential vegetation types (Figure 2.7) is mainly a result of precipitation and temperature gradients, where especially the length of the rainy season is important. The effect of these gradients and the resulting hygrothermal altitudinal zones are depicted with higher detail in Figure 2.8. The number of both arid and humid months per year determines a regular distribution of zonal vegetation types from the humid tropics in the southeast to the arid north (Sommerhoff and Weber, 1999). Exposition (Luv-Lee effects) also creates small-scale heterogeneity and complex vegetation patterns. Pictures of selected vegetation types are shown in Figure 2.9.

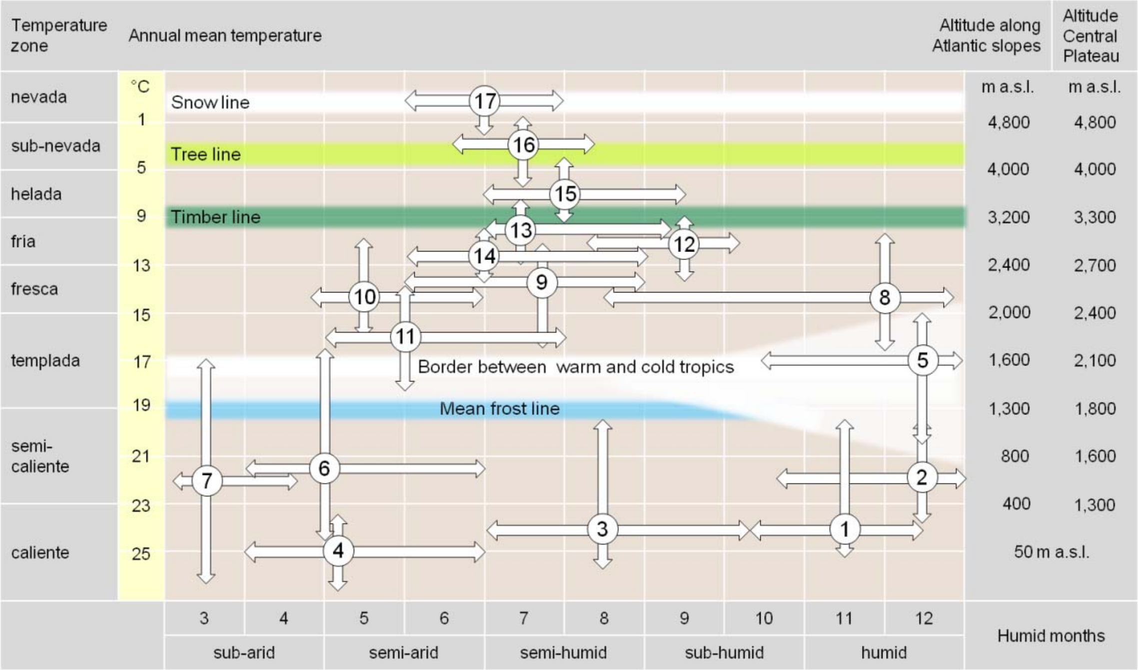


Figure 2.8: Hygrothermal altitudinal zones of Mexico and respective vegetation communities (Sommerhoff and Weber, 1999; modified). The superposition of climatic gradients and altitudinal zones creates a unique variety of vegetation types. 1: Rainforest (semi-evergreen), 2: Rainforest (evergreen), 3: Humid forest/savanna, 4: Dry forest/savanna, 5. Humid montane forest, 6: Montane thorny forest/savanna, 7: Thorny and succulent savanna, 8: Oak-pine cloud forest, 9: Pine-oak-forest (semihumid), 10: Pine-oak-forest (semiarid), 11: Deciduous oak forest, 12: Montane cloud forest, 13: Pine-fir forest, 14: Mixed pine forest, 15: Montane pine forest, 16: Montane grassland, 17: Periglacial vegetation.

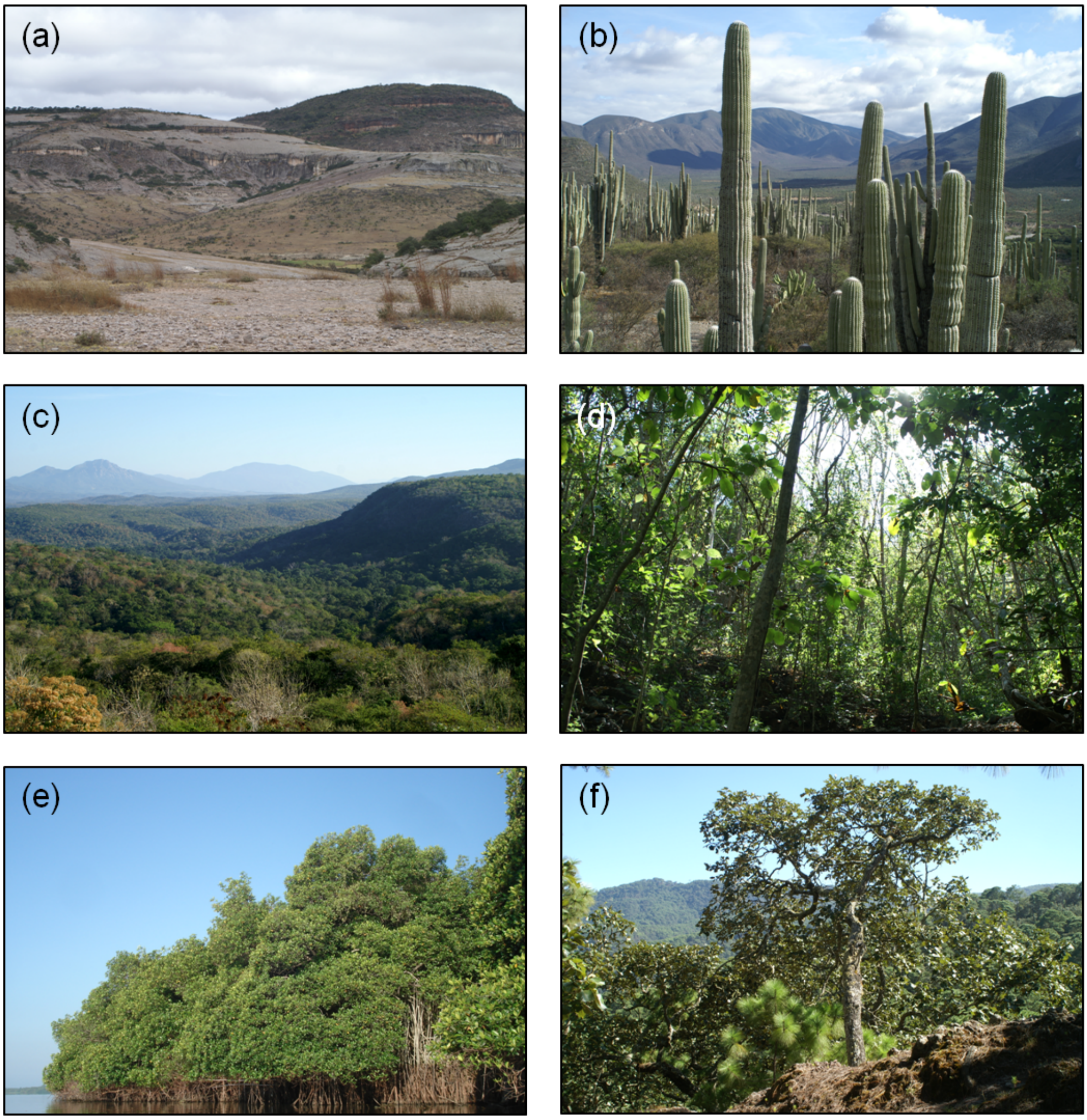


Figure 2.9: Example of typical vegetation types in Mexico (Pictures: A. Cord). (a) Sparse xerophytic vegetation, *Tehuacán-Cuicatlán Biosphere Reserve* (Puebla/Oaxaca), (b) Short xerophytic shrubland with columnar cacti, *Tehuacán-Cuicatlán Biosphere Reserve* (Puebla/Oaxaca), (c) and (d) Tropical deciduous forest, *Chamela-Cuixmala Biosphere Reserve* (Jalisco), (e) Mangroves, Pacific coast (Guerrero), and (f) Pine-oak forest, *Sierra de Manatlan Biosphere Reserve* (Jalisco).

2.4.2 Forest types

There is consensus that the loss of forests (both tropical and temperate) has major impacts on global biodiversity (Ramamoorthy et al., 1993). In the following, Mexico's key forest types including the study species of this thesis (see Sections 6.1 and 7.2) are characterized. Of Mexico's total area covered by vegetation, which is 141.8 million hectares, nearly 66 million hectares are woodlands and forests; of these, 21.6 million hectares have timber production potential (Sarukhán et al., 2010). In general terms, forests predominantly occur along the mountain ranges of the *Sierra Madre Occidental* and in the southern part of the country. Less dominant is the forest cover in northeastern Mexico along the mountain range of the *Sierra Madre Oriental* (Ricker et al., 2007). In Mexico, a distinction is made between *selva* and *bosque* (see Box 1). While *selva* refers to natural forests, *bosque* means forest in the sense of commercial forest including different intensities of human impact (Sommerhoff and Weber, 1999). Tree height is used for the distinction of different types such as *selva baja* (4 to 15 m), *selva mediana* (15 to 30 m), and *selva alta* with more than 30 m (INEGI, 2005).

Box 1. Forest types in Mexico based on Ricker et al. (2007)

TROPICAL FORESTS

1. Tropical or sub-tropical evergreen forests

- High and intermediate semi-evergreen lowland forest (*selva alta y mediana subperennifolia*): 20 m height or more, 25-50% of the trees shed their leaves during the dry season
- High and intermediate evergreen lowland forest (*selva alta y mediana perennifolia*): 20-35 m height
- Low semi-evergreen lowland forest (*selva baja subperennifolia*): 15-20 m; 25-50% of the trees shed their leaves during the dry season
- Low evergreen lowland forest (*selva baja perennifolia*): 4-15 m height, mostly in permanently flooded areas

2. Tropical or sub-tropical deciduous forests

- Low deciduous and subdeciduous lowland forest (*selva baja caducifolia y subcaducifolia*): 4-15 m height; over 50% of the trees shed their leaves during the dry season
- Intermediate deciduous and subdeciduous lowland forest (*selva mediana caducifolia y subcaducifolia*): 15-20 m height; over 50% of the trees shed their leaves
- Mesquite and wattle forest (*mezquital y huizachal*): tree vegetation of up to 12 m height dominated by *Prosopis* (‘‘mesquite’’) and/or *Acacia* (‘‘wattle’’)
- Low thorny lowland forest (*selva baja espinosa*): thorny trees of 4-15 m height

TEMPERATE FORESTS

3. Pine-oak forest

- Oak-pine or pine-oak forest (*bosque de encino-pino o pino-encino*): 5-20 m height, with both *Quercus* (‘‘oak’’) and *Pinus* (‘‘pine’’) being present, either genus being dominant
- Low open forest (*bosque bajo abierto*): 5-10 m height, transition zone between forest and natural grassland
- Oak forest (*bosque de encino*): 2-30 m height, dominated by *Quercus*

4. Coniferous forest

- Pine forest (*bosque de pino*): 8-25 m height, dominated by *Pinus*
- Juniper forest (*bosque de táscate*): 2-6 m height, open forest dominated by *Juniperus*
- Fir forest (*bosque de oyamel*): 20-40 m height, dominated by *Abies*, *Pseudotsuga*, *Picea*, and *Cupressus*

5. Moist montane/cloud forest

- *Bosque mesófilo de montaña*: 15-35 m height

OTHER FORESTED VEGETATION

6. Mangroves

- *Manglar*: forest flooded by (salty) sea water, generally of 3-15 m height, in Mexico dominated by the species *Rhizophora mangle*, *Avicennia germinans*, *Laguncularia racemosa*, and *Conocarpus erecta*

7. Palm forest

- *Palmar*: up to 40 m hight, mostly tropical vegetation dominated by palms (family Arecaceae)

8. Gallery forest

- *Vegetación de galería*: conspicuous tree vegetation along rivers and water bodies, with particularly high soil humidity

2.5 Socio-economic conditions

Mexico is considered a ‘newly industrialized country (NIC)’ meaning that it has not yet reached ‘First World’ status but that it is undergoing rapid (export-oriented) economic growth (Bozyk, 2006). Accompanied by growing population numbers, this gives rise to the increasing exploitation of natural resources. The concomitant land consumption for economical and infrastructural purposes leads to the fragmentation of the landscape and the loss of natural areas. These processes in turn have impact on species distribution ranges and biodiversity. In the following, Mexico’s population structure (Section 2.5.1), economy (Section 2.5.2), and human land use (Section 2.5.3) are briefly outlined.

2.5.1 Population density and structure

In July 2011, Mexico had a total estimated population of 113.7 million persons (CIA, 2011), with significant differences in population density throughout the country (Figure 2.10). As a result of favorable conditions for agriculture and commercial opportunities as well as settlement history, the major agglomerations are found in the central parts, in seaports, and along the US border (Sommerhoff and Weber, 1999). In 2009, Mexico-City had 19.3 million inhabitants (CIA, 2011) and is thus the second-largest urban agglomeration in the Western Hemisphere after Sao Paulo (Brazil). Other major cities in the country are (see Figure 2.10, assumed population for 2009 in parentheses): Guadalajara (4.3 million), Monterrey (3.8 million), Puebla (2.8 million), and Tijuana (1.6 million) (CIA, 2011). The national population growth rate is estimated to have been 1.3% between 2005 and 2010 (FAO, 2011) and 1.1% for the year 2011 (CIA, 2011). The major ethnic groups in Mexico are (CIA, 2011): 60% *mestizo* (Amerindian-Spanish), 30% Amerindian or predominantly Amerindian, 9% white, and 1% other.

Throughout the country, migration from rural to urban areas and urbanization processes are observed (see Figure 2.10). In 2010, 78% of the population was estimated to live in urban areas (CIA, 2011; FAO, 2011). This increasing agglomeration in urban centers creates several problems and challenges such as air pollution, waste disposal, insufficient sanitary conditions, and drinking water supply (Sommerhoff and Weber, 1999; Figure 2.11). Apart from that, although difficult to quantify, there are environmental recovery trends for areas that have been abandoned through depopulation as a result of migration from rural areas (Sarukhán et al., 2010).

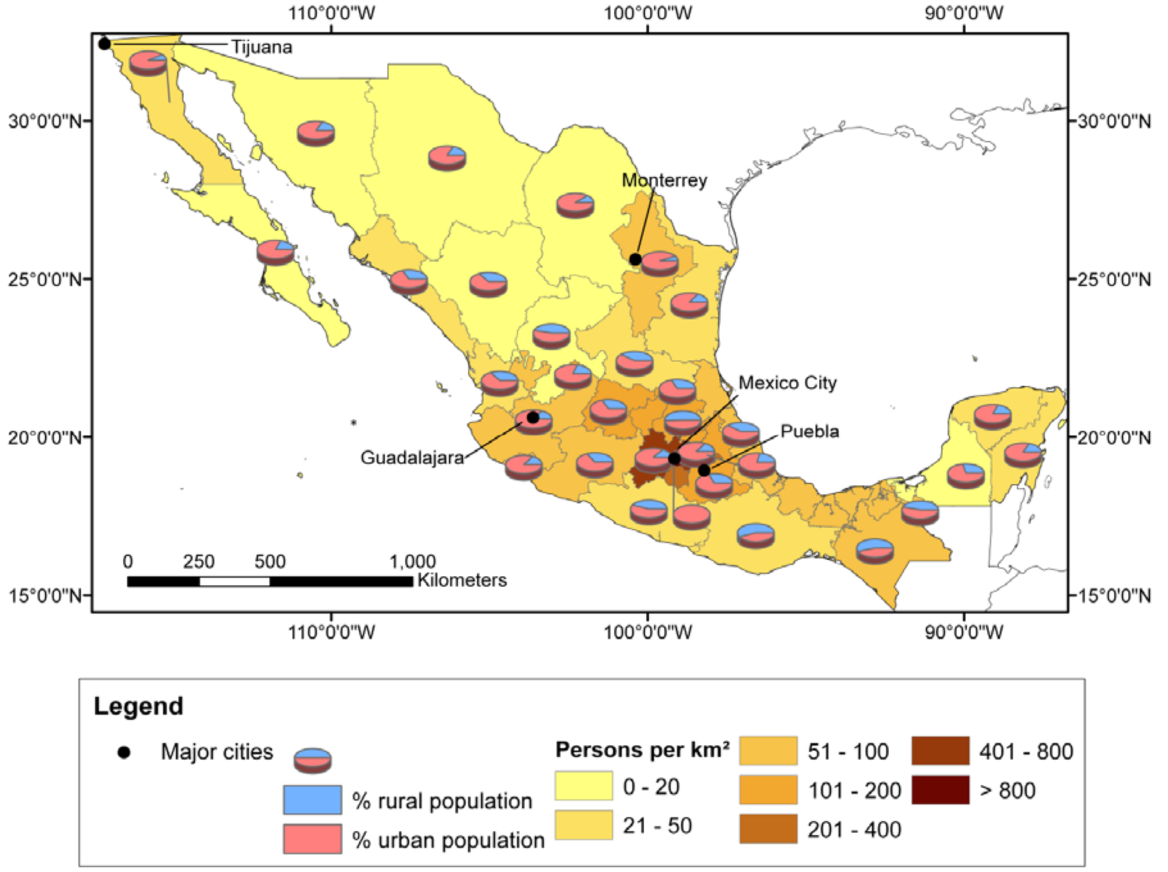


Figure 2.10: Population density and percentages of urban and rural population on the state level in Mexico for the year 2000 (CONABIO, 2006), data available from CONABIO's online portal for geoinformation (CONABIO, 2011).

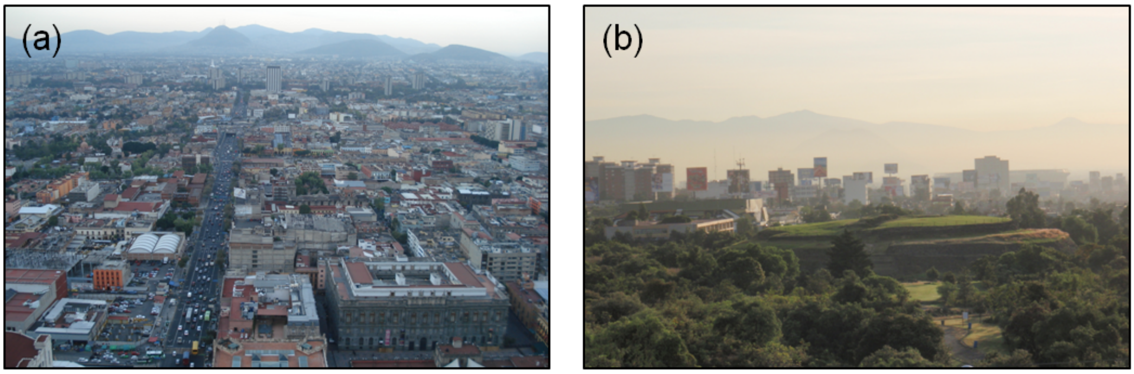


Figure 2.11: Mexico-City metropolitan area (Pictures: A. Cord). (a) Overview of the city center located in the enclosed *Valley of Mexico* and (b) Typical smog and air pollution conditions.

2.5.2 Economy

Land consumption is not only triggered by increasing population numbers as outlined above, but also by economic activities. Besides its biological capital, Mexico also has a variety of natural resources such as petroleum, silver, copper, gold, lead, zinc, natural gas, and timber (CIA, 2011) and belongs to the economically most important countries in Latin America (Sommerhoff and Weber, 1999). For example, Mexico had a daily oil production of 2.98 million barrel per day in 2010 and was thus the seventh largest global oil exporter (CIA, 2011). Mexico's *Gross Domestic Product* (GDP), which refers to the market value of all final goods and services produced within a country in a given period, is unevenly allocated to the different sectors (CIA, 2011) with 3.9% to the primary sector (agriculture, forestry and fishing, mining, and extraction of oil and gas), 32.6% to the secondary sector (industry), and 63.5% to the tertiary sector (services). The major agricultural products are corn, wheat, fruits, beans, soybeans, rice, cotton, and coffee (Sommerhoff and Weber, 1999; CIA, 2011). The main branches of industry are food and beverages, tobacco, chemicals, iron and steel, textiles, clothing, and motor vehicles; the two most important export partners are the US and Canada (FAO, 2011). Due to Mexico's rich cultural and natural resources the importance of tourism has been increasing since the 1980s (Sommerhoff and Weber, 1999) but was threatened along the Gulf coast due to the oil spill of the BP *Deepwater Horizon* drilling platform in 2010 (Blanke and Chiesa, 2011). International tourist arrivals in the country were estimated to be 21.4 million persons (for the year 2009); 4.3% of the total employment is in the tourism sector (Blanke and Chiesa, 2011).

2.5.3 Anthropogenic land use

Land transformation and the concomitant loss of habitats due to human activities is one of the major threats to biodiversity in Mexico (see Section 2.3.3) and can be quantified by means of remote sensing. Up to the 1980s, an extensive transformation of natural ecosystems for agricultural, urban, and infrastructure activities took place "while ignoring the consequences that this would have on the natural capital" (Sarukhán et al., 2010: p. 86). Especially Mexican forests, which are rich in species numbers and represent the focus vegetation community of this thesis, are threatened (Burgos and Maass, 2004; Sarukhán et al., 2010). Ricker et al. (2007) estimated these changes in cover percentages for the reference year 2000 as compared to the 'original cover' (Rzedowski, 1990) at -36% for forests, -29% for shrubland, and -47% for natural grasslands. According to the findings of Ricker et al. (2007) human-made grasslands increased by 12% and agricultural lands by 17%, respectively. Altogether, the extent of woody vegetation cover has therefore significantly been reduced or compromised by agriculture or the production of livestock (Sarukhán et al., 2010). Forest cover has decreased from 688,000 km² in 1994 to 649,600 km² in 2009 (FAO, 2011). Annual deforestation rates of 0.25% for temperate and 0.76% for tropical forests were reported (Mas et al., 2002) with only 27% of the original tropical dry forest remaining undisturbed (Burgos and Maass, 2004). Large portions of the remnant forests in the country are fragmented or disturbed and covered by secondary vegetation (Sarukhán et al., 2010).

Areas in use for agriculture or livestock production cover more than 50% of Mexico's territory with approximately one quarter used for arable farming and three quarters used for extensive ranching (Sommerhoff and Weber, 1999). In Mexico, agriculture is mainly limited by precipitation availability. Due to short rainy seasons and high inter-annual variability of precipitation in the northern parts of the country (see Section 2.2), agriculture is in these areas only possible with irrigation (Sommerhoff and Weber, 1999). For 2008, the area of irrigated agricultural land was estimated at 63,000 km² (CIA, 2011). Irrigation agriculture is mainly practiced along the northern Pacific coast in the states of *Sinaloa* and *Sonora* as well as close to the US border in northeastern Mexico (Sommerhoff and Weber, 1999). In addition to precipitation, steep relief with strong slopes limits the use for agriculture and leads to increased soil erosion (Sommerhoff and Weber, 1999). Peripheral mountain regions are used only for extensive agriculture, and farmers often practice the so-called 'Milpa'- or 'Ronza'-system (*milpa* = cornfield, *roza* = clearing) which is a form of forest-field-rotation (Sommerhoff and Weber, 1999). Burgos and Maass (2004) identified three different main pathways of vegetation change associated with agricultural land use in western Mexico: the replacement of forests by agriculture in flat areas, the establishment of pastures on slopes, and the extraction of wood on hill crests (see Figure 2.12). To summarize, human impact and land cover change in the country result in rapidly changing environmental conditions and the need to identify current biotope conditions as a baseline for any projection into the future.

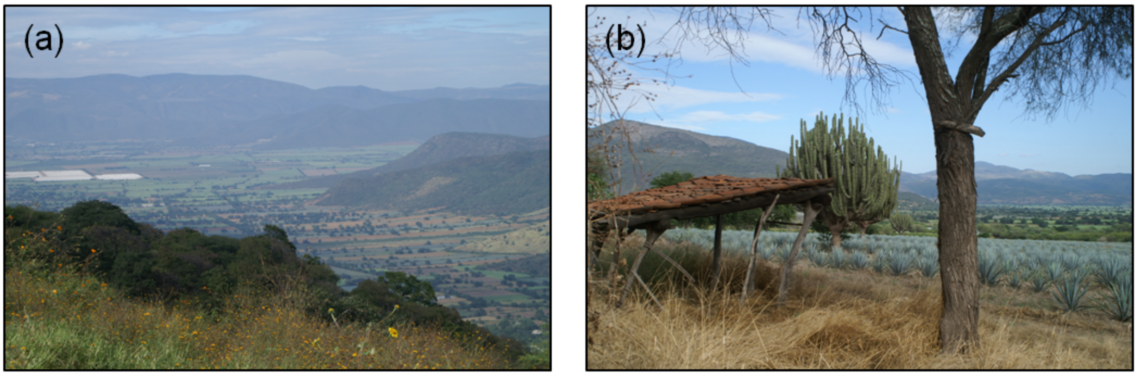


Figure 2.12: Agriculture in Mexico (Pictures: A. Cord). (a) Location of agricultural fields and greenhouses in flat areas and extensive pastures on the hill slopes (Jalisco), (b) Plantation of *Agave tequilana* for tequila production (Jalisco).

2.6 Summary

This overview of Mexico, the study area of this thesis, summarized the 'natural capital' of the country as well as recent trends in human-induced land use/land cover changes. Mexico's complex geological setting – with its broad mountain ranges, large plateaus, and coastal lowlands – as well as the different climatic regimes and soil types in the country have generated a large variety of ecosystems und a unique phytodiversity. The environmental heterogeneity has led to globally outstanding levels of species richness and endemism that made Mexico part of the 17 so-called 'megadiverse' countries and a unique opportunity for studying species distributions in this thesis. The chapter also included an overview of the distribution of potential vegetation, indicating pre-human conditions and hence potential species distribution ranges. In addition, an overview was given of the major forest types that the study species of this thesis are typical of.

Mexico is considered a ‘newly industrialized country (NIC)’ and is undergoing rapid (export-oriented) economic growth. This development leads to significant population growth and increasing agglomeration in urban centers. Further, an extensive transformation of natural ecosystems for agricultural and infrastructural activities takes place. According to this and consistent with the findings of the *Millennium Ecosystem Assessment*, habitat destruction is currently the major threat to biodiversity in the country. Especially the Mexican forests, which are rich in species numbers and the focus vegetation community of this thesis, are threatened. This loss of natural and pristine habitats and biotopes emphasizes the need for biodiversity assessments and environmental monitoring by means of remote sensing.

In addition, a significant proportion of the protected areas (with public, social, or private ownership) has been established by circumstantial reasons with little basis on scientific knowledge. The current network of protected areas is hence not adequately designed to protect Mexico’s biodiversity. Therefore, it has been claimed that policy decisions in Mexico should be based on reliable scientific data and meet the criteria relevant for the establishment of sustainable protected areas. The findings of this thesis can effectively support such efforts.

3 Species distribution models – Theoretical framework

In the last two decades, *Species distribution models* (SDMs) have established as important tools in both basic and applied science and now play a fundamental role in biogeography (Franklin, 2009). In general, there are two different approaches in which SDMs have been categorized, the so-called *mechanistic* or *correlative* models.

The *mechanistic* approach is based on the physiological relationship between the functional traits of the target species and its environment (Robertson et al., 2003; Kearney and Porter, 2009; Yates et al., 2010) and has considerably more power in testing ecophysiological hypotheses (Kearney and Porter, 2004). *Mechanistic* models aim to incorporate physiologically limiting mechanisms in a species' tolerance (Pearson, 2007) and evaluate species life-history together with behavioral, morphological and physiological traits (Kearney and Porter, 2009). They therefore require a detailed *a priori* understanding of the study species' ecophysiological constraints (e.g. measured in field or laboratory experiments) and are "difficult to develop for all but the most well understood species" (Pearson, 2007: p. 5). The physiological constraints identified within *mechanistic* approaches can then be used to display geographic regions of positive species fitness using *Geographic Information System* (GIS) technology (Soberón and Peterson, 2005).

This direct linking of physiological constraints and environmental variables (e.g. climate, terrain) poses significant challenges (Kearney and Porter, 2009). The majority of the species distribution modeling approaches is therefore *correlative* in the sense that the models statistically link geo-references species occurrence records to suites of environmental variables that can reasonably be expected to affect species distributions (Elith and Leathwick, 2009). As further outlined in the following sections, *correlative* approaches thus make use of empirical relationships and identify ecologically similar areas to those where the species is known to occur (Guisan and Zimmermann, 2000). These areas are in some sense identified by the model algorithm and displayed as areas of predicted presence on the map (Soberón and Peterson, 2005).

While the power of the *correlative* approaches lies in spatial *prediction,* the *mechanistic* approaches provides considerably more power in *explaining* patterns of distribution and abundance (Kearney and Porter, 2004). *Mechanistic* and *correlative* approaches thus answer different questions and, in this sense, are "complementary rather than alternative techniques" (Kearney and Porter, 2004: p. 3129). In this thesis, the *correlative* approach to modeling species distribution was applied as the focus of the work was to predict spatial patterns of species distributions (and richness) and to explore the potential of remote sensing variables. In the following, species distribution models are therefore considered only from the perspective of correlative approaches. A variety of textbooks and studies dealing with the concepts of species distribution modeling have been published (Guisan and Zimmermann, 2000; Guisan and Thuiller, 2005; Elith and Leathwick, 2009; Franklin, 2009). They make use of various terms and definitions to describe ecological theory, input parameters, data characteristics, and methodological issues. In order to provide a common basis for the concepts, methods and data used in this thesis, the most important terms and their usage in this work are defined in Appendix A.

The following chapter reviews relevant concepts and components of SDMs and provides a brief overview of the principle elements of a correlative species distribution model (Figure 3.1). It includes the **species occurrence data** in geographical space to calibrate the model and validate its performance, the **environmental predictors** representing those factors controlling species distributions, the **species distribution model** linking species occurrences to the environmental predictors, and methods and criteria for applying the model (e.g. coefficients, thresholds). The organization of the chapter follows the concept proposed by Austin (2002) who identified three important components in species distribution modeling: (1) the *ecological* model (Section 3.1) concerning the ecological theory to be used, (2) the *data* model (Section 3.2) in relation to candidate predictors and spatial scales, and (3) the *statistical* model including e.g. model algorithm and model thresholding (Sections 3.3 and 3.4). To point out the potential and limitations of species distribution models, typical applications in research and conservation are illustrated by exemplary studies (Section 3.5).

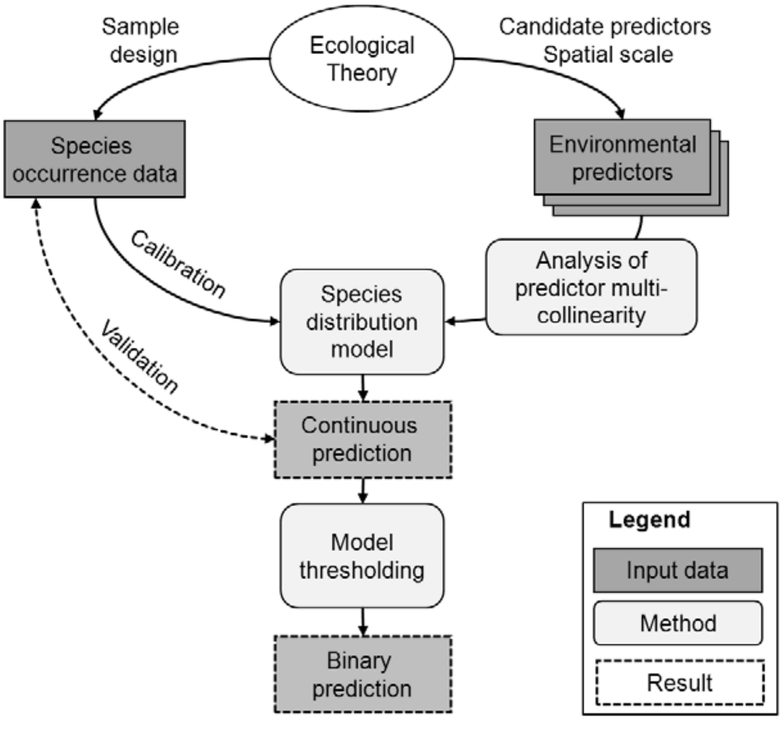


Figure 3.1: Theoretical framework and components of a correlative species distribution model (Franklin, 2009; modified).

3.1 Conceptual and ecological basis

The ecological roots of SDMs belong in those early studies (Murray, 1866; Grinnell, 1904) that described biological patterns in terms of their relationships with geographical and/or environmental gradients. The major aspects of ecological theory related to SDM, the current scientific consensus, and the assumptions made in this thesis are summarized in the following.

3.1.1 Niche concept

The species niche concept outlined in the following is central to ecology and in particular SDMs (Franklin, 2009). The concept has evolved over time and has different interpretations that have been classified (Leibold, 1995) into being driven by either the environmental **'requirements'** of the species as defined by Grinnell and Hutchinson (Grinnell, 1917; Hutchinson, 1957) or by the impact or **'functional role'** in the biotic community as described by Elton (Elton, 1927) and later on by MacArthur and Levins (MacArthur, 1967; Levins, 1968). As the latter, the 'functional role' concept, is more related to trophic levels and food web theory (Guisan and Thuiller, 2005), only the 'requirement' concept usually is considered in SDMs. Within this concept, Hutchinson (1957) distinguished two central viewpoints (Figure 3.2):

(1) The **fundamental** (potential) **niche**, defined as the physiological tolerances of the species in its environment in absence of biotic interactions – typically predicted from theoretically or experimentally derived physiological constraints.

(2) The **realized** (actual) **niche**, referring to the environmental dimensions in which species can survive and reproduce, including the effects of biotic interactions (e.g. competition, predation) – typically observed in field data on species occurrence.

The realized niche is thus often depicted as a subset of the fundamental niche (Figure 3.2), although this is not necessarily true, e.g. in the case of positive biotic interactions (facilitation, symbiosis) where the presence of another species is a requirement for the occurrence of the target species (Franklin, 2009). In species distribution modeling, the differentiation between fundamental and realized niche is important because it determines whether the distribution is predicted from theoretical physical constraints or field observations (Guisan and Zimmermann, 2000). The species observations used in this thesis, including field records and herbarium collection data, represent the realized niche.

A number of conceptual advancements of Hutchinson's niche concept have been made in the literature (Appendix B). According to these novel concepts, several other factors may result in a species being absent from suitable habitat and/or being found in unsuitable habitats. For example, the importance of measuring species fitness and population growth rates has been emphasized and the niche concept has been extended by meta-population theory (Hanski, 1999) and source-sink theory (Pulliam, 1988; Pulliam, 2000). However, as these concepts have mainly been applied by zoologists, they are not discussed in detail here. Plant ecologists have traditionally relied on the niche concepts of Grinnell and Hutchinson.

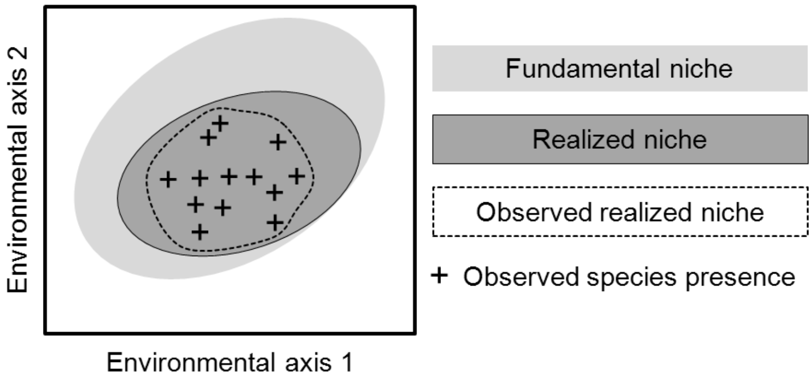


Figure 3.2: Hutchinson's (fundamental and realized) niche concept (Pearson, 2007; Franklin, 2009; modified). Both the observed realized niche and the observed species presence records often do not identify the full extent of the realized niche (e.g. due to incomplete sampling).

3.1.2 Environmental gradients and species response curves

Based on the niche concept and the results of his extensive field studies, Whittaker (1956) developed the key ideas for the changes of abundance of individual species along environmental gradients. Such species-specific correlations to environmental gradients have been termed 'species response curves' for plants (Austin, 2002) or 'resource selection functions' for animals (Boyce et al., 2002). They depict a function describing the relationship of species occurrence or abundance in relation to values of an environmental variable (Franklin, 2009).

The shape of these species response curves is dependent on ecological theory and data models applied (Oksanen and Minchin, 2002). Different hypothesized types of species response curves to an environmental gradient and possible explanatory processes are summarized in Figure 3.3. Niche theory assumes symmetric Gaussian-shaped unimodal species response curves for both plants and animals (Austin and Smith, 1989; Austin, 2007) that are usually not found when being estimated using field data (Franklin, 2009). Bi- or multimodal response curves typically result from competition as they assume that a superior competitor can displace a species from the optimum of its fundamental niche (Müller-Dombois and Ellenberg, 1974). Skewed responses can be expected for gradients with physiological stress such as temperature (Austin and Smith, 1989): At the "harsh end" (Franklin, 2009: p. 43) of the gradient, species fitness is often limited by physiological constraints and at the "benign end" (Franklin, 2009: p. 43) by inter-specific competition. In addition, species position along an environmental gradient has been shown to influence the shape of the response detected (Rydgren et al., 2003). However, reliable conclusions about the response curves of species can only be drawn if the sampled environmental gradient clearly exceeds the upper and lower limits of the species occurrence (Austin, 2007). Liebig's *Law of the minimum* (von Liebig, 1840) suggests that the true response of a species to a certain environmental gradient can only be detected when all other factors occur at non-limiting levels (Austin, 2007; Franklin, 2009).

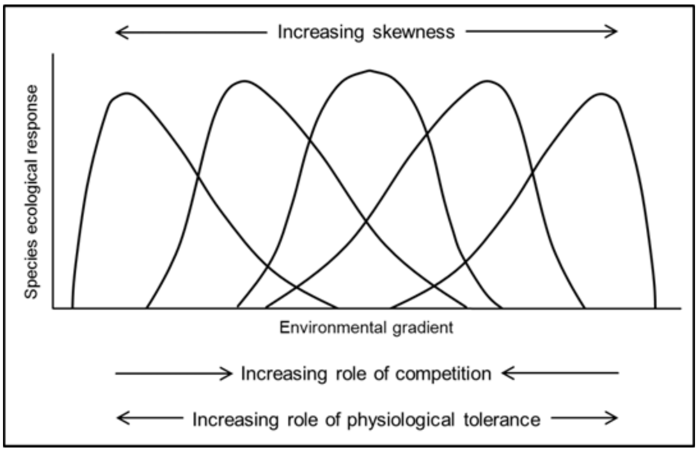


Figure 3.3: Hypothetical types of species response curves along an environmental gradient affected by inter-specific competition and physiological tolerance (Austin et al., 1990).

3.1.3 Environmental and geographical space

Both ecological niche theory and species response curves outlined in the previous sections, relate to a “conceptual space” (Pearson, 2007: p. 8), the so-called ‘environmental’ space. Based on Hutchinson’s characterization of the niche as “n-dimensional hypervolume” (Hutchinson, 1957: p. 416), environmental space in species distribution modeling is potentially multi-dimensional and defined by the respective environmental model predictors (Elith and Leathwick, 2009). This hypervolume in turn suggests multiple causal factors that influence species distributions (Franklin, 2009), discussed in more detail in Section 3.2. As an aside, this concept is analogous to the principle of the ‘spectral feature space’ which is applied in remote sensing to group observations according to their spectral properties.

Species distributions themselves are mostly represented in geographical space, i.e. by two-dimensional (x, y) map coordinates (Pearson, 2007). This distinction made between ‘environmental’ and ‘geographical’ space is one of the central concepts for the implementation of species distribution modeling algorithms (see Section 3.3). The corresponding dual representation of species distributions in environmental and geographical space is illustrated in Figure 3.4. Accordingly, species occurrence data are first mapped onto geo-referenced environmental data. The values of the environmental predictors at these localities are extracted and used to construct the ecological niche and species response curves. At this stage, the model is calibrated (i.e. built) in environmental space (Pearson, 2007) and thus effectively ignorant of geographic proximity (Elith and Leathwick, 2009). This non-spatial approximation of the climatic niche may then be projected back into geographical space based on species response curves (Section 3.1.2). The modeled prediction of the species distribution is again a spatially explicit representation in geographical space. Some (simple) approaches to predict species distributions are purely geographic assuming that rather geographic than environmental processes determine the distribution of the study species. Examples include convex hulls and kriging (Elith and Leathwick, 2009).

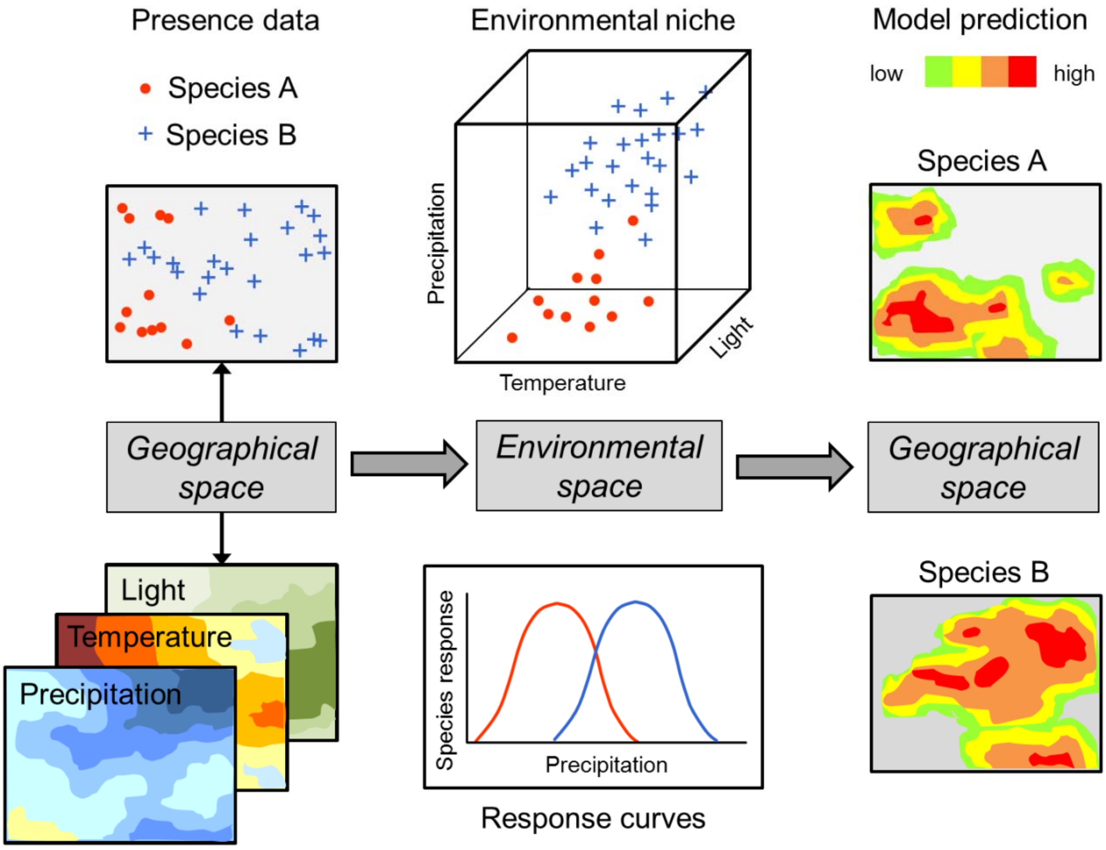


Figure 3.4: Interplay between environmental and geographical space (partly based on Elith and Leathwick, 2009). Species presence data of species A (occurring in colder, arid conditions) and B (occurring in warmer, humid conditions) are mapped onto geo-referenced environmental data (*Geographical space*, left). The values of the environmental predictors at the species localities are extracted and used to construct the ecological niche and species response curves (*Environmental space*, center). The approximation of the niche is then projected back into geographical space based on the previously analyzed species response curves (*Geographical space*, right). The output maps spatially depict the model prediction for both species separately.

3.1.4 What is predicted? – The informative value of SDMs

The previous section introduced the concept of model calibration in environmental space and model prediction in geographical space. But what is actually modeled during the projection back into geographical space – the fundamental or the realized niche, or the habitat suitability? The field of species distribution modeling is “plagued by loose and inconsistent application” (Kearney, 2006: p. 186) of the concepts of ‘niche’ and ‘habitat’ which has resulted in a substantial terminological confusion (Kearney, 2006; Soberón, 2007). Hence, different terminologies for predictive distribution models and their resulting maps were suggested (see Box 2, p. 38). Indeed, “modeling the ecological niche of a species is not necessarily the same as modeling its geographic distribution in a particular study region” (Phillips, 2008: p. 277) since the distribution may be influenced by factors such as dispersal ability or biotic interactions which are not represented in the definition of its ecological niche (Guisan et al., 2006; Zurell et al., 2009). Thus, “a good model of a species' niche may not result in a good model of its current distribution, and vice versa" (Phillips, 2008: p. 277).

Whether or not the niche or the habitat suitability is mapped in geographical space, one agrees that SDMs identify areas that have similar environmental conditions to where the species has been observed and currently maintains populations (Pearson et al., 2007). However, there is some dissent on the subset of the total niche or distribution that is captured by the models (Soberón, 2007): The most frequent simplification found in the literature is that SDMs quantify Hutchinson's realized (see Figure 3.2) niche since the observed species occurrences usually utilized for modeling are already constrained by biotic interactions and limited resources (Austin, 2002; Thuiller et al., 2004; Guisan and Thuiller, 2005) as well as stochastic effects (Guisan and Zimmermann, 2000) – an assumption also valid for this thesis. Therefore, the available species presence records do usually not reflect the whole range of habitats occupied by the species. The model is hence fit only to that portion of the niche that is represented by the observed records (Pearson, 2007; see Figure 3.2). The areas of predicted presence are thus typically larger than the realized distribution since only few species occupy all areas that satisfy their requirements (Phillips et al., 2006; Zimmermann et al., 2007).

Beyond this, both realized and potential distributions only refer to a certain point or particular period in time, usually the presence (Jiménez-Valverde et al., 2008). "This means that both potential and realized distribution vary with time but in different ways: The potential distribution varies geographically with the oscillations of climatic conditions but is environmentally stable. On the other hand, the "realized distribution of the same species will vary in both the geographical and environmental spaces" (Jiménez-Valverde et al., 2008: p. 886).

Given the variety of terms for predictive distribution models and their resulting maps (Box 2), in this thesis the term 'species distribution model' is used since the models are applied to predict the geographical distribution of the target species rather than to study the characteristics of their distributions in environmental (niche) space. Further, the model prediction is assumed to be the 'potential distribution of the realized niche' since the presence data available (see Section 5.1) are constrained by biotic interactions and limited resources. Depending on the environmental predictors used, two different information layers are generated within this work (Figure 6.1). While climate data determine the *climatic suitability*, the remote sensing predictors typically characterize the *biotope suitability*. Both model predictions are combined in a hierarchical modeling framework (see Section 6.2) to represent current potential species distributions.

Box 2. Different terminologies used for species distribution models

SDMs have been said to (1) describe the species niche and (2) the suitability of habitat to support a species (Franklin, 2009). Hence, the names used for the models vary widely between different authors and applications. What has been termed ‘species distribution model’ so far, has also been called ‘bioclimatic model’ (Gignac et al., 2000), ‘climate envelope model’ (Heikkinen et al., 2006), ‘climate response surface’ (Huntley, 1995), ‘ecological niche model’ (Peterson, 2006), ‘habitat suitability model’ (Hirzel et al., 2006), and ‘resource selection function’ (RSF) (Boyce et al., 2002). Each name has a different connotation and several authors have discussed the corresponding nomenclature (Graham et al., 2004; Peterson et al., 2008).

Terms such as ‘ecological/environmental niche model’, ‘bioclimatic model’, and ‘climate envelope model’ are often summarized as ‘ecological niche models’ (ENM) and usually refer to attempts to estimate the fundamental niche (Pearson, 2007). On the contrary, the use of the term ‘species distribution model’ implies that the aim of the study is to simulate the actual distribution of the species (Pearson, 2007). Peterson et al. (2006) also restricted ‘SDM’ to those models including biotic or accessibility predictors and/or being limited in spatial extent. They formalized the dichotomy between ENM and DM (‘Distribution Models’) and the ‘worlds’ the two approaches presume: While ENM would anticipate a full-range scale for the study, DM may aim to examine a smaller proportion of the species range (Peterson et al., 2006). In the lights of the controversy discussion on which components of the niche are represented by predictions of distribution models, the use of the more neutral terminology, SDM rather than ENM, seems preferable (Elith and Leathwick, 2009) and was used in the thesis (see also the remarks in the main text).

Similarly, the resulting maps have been called with different terms, e.g. ‘ecological response surfaces’ (Lenihan, 1993), ‘predictive maps’ (Franklin, 1995), ‘potentially occupied habitat’ (Guisan and Zimmermann, 2000), ‘habitat suitability’ (Hirzel and Guisan, 2002), ‘spatial predictions of species distribution’ (Austin, 2002), ‘predictions of occurrence’ (Rushton et al., 2004), or ‘biogeographical models of species distributions’ (Guisan et al., 2006).

3.1.5 Equilibrium postulate and environmental completeness

Since both species and environmental data are sampled within a limited period of time and/or space, models trained and fitted with these can only represent a snapshot of the underlying relationships (Guisan and Thuiller, 2005). For large-scale modeling – as in this thesis – and for projecting the model in space or time, two convenient and essential postulates are therefore presumed:

(1) **(Pseudo)equilibrium postulate.** The study species is assumed to be in (pseudo)equilibrium with its contemporary environment (Guisan and Theurillat, 2000). In this context, ‘equilibrium’ with the physical environment means that the observed distribution and abundance is indicative of resource requirements and environmental tolerances of the species (Franklin, 2009). Hence, the species occurs in all suitable areas, but is absent from all unsuitable areas (Pearson, 2007). ‘Pseudo-equilibrium’ can be interpreted as changes being slow – at least “relative to the life span of the studied biota” (Austin, 2002: p. 103). The degree of equilibrium in general depends on both biotic interactions (e.g. competitive exclusion from an area) and dispersal abilities – whereas organisms with higher dispersal ability are generally closer to equilibrium than organisms with lower dispersal ability (Araújo and Pearson, 2005).

(2) **Environmental completeness.** The relevant environmental gradients are assumed to have been adequately sampled (Elith and Leathwick, 2009). This presumption may be violated in cases where very few occurrence samples are available, e.g. due to limited survey effort (Anderson and Martínez-Meyer, 2004) or low probability of detection (Pearson et al., 2007). However, there is not necessarily a direct link between sampling in geographical space and in environmental space. Distances between sites in geographic space might be different from those in environmental space (Elith and Leathwick, 2009); poor sampling in geographical space may still result in good sampling in environmental space (Pearson, 2007). Environmental completeness has been shown to positively affect SDM performance (Kadmon et al., 2003).

The limitations of both assumptions are known in detail (Guisan and Zimmermann, 2000; Austin, 2002) and need to be considered in each SDM study. They appear to be less restrictive for species that are comparatively persistent and slowly react to variability of environmental conditions such as alpine or arctic species (Guisan and Zimmermann, 2000). In general, applications that fit models for species not in equilibrium with their environment (e.g. during species invasion or in climate change scenarios) are much more challenging, and results are more ambiguous (Elith and Leathwick, 2009).

3.2 Determinants of species distributions

According to the concept of the n-dimensional niche space (Hutchinson, 1957; see Section 3.1.3), species geographical distributions are determined by several factors. Candidate determinants of species distributions are not only abiotic (e.g. climate) factors as discussed so far, but also biotic interactions, dispersal constrains, anthropogenic effects, stochastic events, and other historical factors (Pulliam, 2000; Soberón and Peterson, 2005; Jiménez-Valverde et al., 2008). Soberón (2007) assembled these factors into three groups, namely (1) the dispersal capacities of the species, (2) the spatial distribution of environmental conditions favorable to the establishment, survivorship and reproduction by the individuals, and (3) the biotic environment, formed by competitors, predators, and pathogens. Since the focus of this thesis is modeling species distributions based on environmental predictors in the sense of Soberón's second group, in the following only concepts regarding their classification are considered.

3.2.1 Types of environmental predictors

Beyond the fundamental differentiation into *abiotic* and *biotic* factors, various conceptual frameworks for the classification of environmental variables used in SDMs have been proposed (Table 3.1). This classification supports the selection of candidate variables and a process-based understanding of predictor importance. However, the frameworks are all based on idealized and not necessarily mutually exclusive categories. Early on, Hutchinson (1978) distinguished between so-called *scenopoetic* (referring to environmental conditions) and *bionomic* (resource-related) predictors. Austin and

Franklin (Austin, 1980; Franklin, 1995) further differentiated *direct* (with physiological effect), *indirect* (without physiological relevance), and *resource* (consumed or used up) gradients. In addition, environmental variables may be classified as being either *proximal* or *distal* referring to their position in the chain of processes that link the predictor itself to its impact on the species (Austin, 2002). In general, SDMs based on *resource* and *proximal* or *direct* gradients will be the most robust and widely applicable (Austin, 2002). However, they will also be the least practical in terms of resources and time to measure the respective gradients.

Table 3.1: Conceptual frameworks proposed by various authors for the classification of environmental variables used in species distribution modeling.

Reference	***Type***	***Description***	***Example***
(Hutchinson, 1978; Soberón, 2007)	*Scenopoetic*	Environmental conditions, at broad scales	Solar radiation as a function of latitude
	Bionomic	Resource-related, at local scales	Radiation on a forest understory
(Austin, 1980; Franklin, 1995)	*Direct*	With physiological effect	Temperature, pH
	Resource	Consumed or used up, matter or energy used for growth	Water, light, nutrients, carbon dioxide, oxygen
	Indirect	Without physiological relevance	Latitude, longitude, altitude, slope, aspect
(Austin, 2002)	*Proximal* (causal)	Directly related to growth and survival	Nutrients
	Distal (proxy or surrogate)	Related to resources or regulators	Soil texture, pH

The climate data used in this thesis (see Section 5.2) mainly represent *scenopoetic* variables for the target tree species. However, the WorldClim data set also includes physiologically limiting (i.e. *direct* in the sense of Austin and Franklin) variables such as the *minimum temperature of the coldest month*. No terminology for the contribution of remote sensing data in characterizing species distributions exists so far. Certainly, remote sensing predictors are rather *indirect* or *resource* than *direct* variables. Their significance differs notably from the one of climate data and very generally indicates the suitability of the respective species community. Candidate predictors should generally be selected based on the ecological theory relevant for describing the geographical distribution of the target species (MacNally, 2000; Beaumont et al., 2005). Chapter 4 therefore particularly addresses the ecological relevance of the remote sensing variables utilized in this work for plants.

3.2.2 Spatial scales: Grain and extent

Strongly connected with different types of environmental variables as discussed above is the relevance of the spatial scale. For example, Hutchinson (1978) classified the predictors according to their relevant scales (see Table 3.1). Pearson and Dawson (2003) further introduced an explicitly hierarchical point of view suggesting that *abiotic* predictors may be more important at coarser spatial scales while *biotic* variables are relevant at finer scales to define species distributions.

One of the central and recurring problems in SDM is thus identifying the appropriate scale for modeling (Wiens, 2002). The adequate scale is dictated by the aims of the study, the system, and available input data (Elith and Leathwick, 2009). It is also closely related to the type of species considered, e.g. highly mobile or sessile organisms (Guisan and Thuiller, 2005). Scale has two important independent components, namely **resolution** (grain size) and **extent**. While resolution is the size of the sampling unit at which the data are recorded, extent refers to the area over which a study is carried out (Austin, 2007). In a review, Franklin (2009) summarized resolution and extent of 25 representative SDM studies and differentiated them into either 'ecological-scale' (high resolution, small extent) or 'biogeographical scale' (low resolution, large extent) studies. Given the resolution and extent of the environmental predictors used here (see Chapter 5), this thesis belongs to the 'biogeographical scale' studies. Grain size of the environmental predictors utilized should match the sampling resolution or spatial precision of the location data, e.g. plot size in field surveys or grid size in atlas surveys (Guisan and Thuiller, 2005; Baldwin, 2009). As studies covering large extents typically rely on herbarium or museum collection data with unknown positional accuracy, they necessarily employ environmental data with coarser resolutions. This guideline is applicable to some of the species occurrence data available for this thesis, too (see Section 5.1).

In addition, due to practical limitations such as the low effective resolution of many environmental data sets, smaller sampling units often need to be defined within larger modeling units (Guisan and Thuiller, 2005). The same environmental variable sampled at different resolutions can have varying meanings for a species and hence for its distribution (Guisan and Thuiller, 2005). For example, both Pearson et al. (2004) and Luoto et al. (2007) found that land cover data increased model performance at finer (1 km to 20 km) resolutions but did not improve or even decrease SDM performance at coarser (40 km to 80 km) resolutions.

Beyond this, it has been pointed out that the environmental factors controlling species distributions change with scale, making cross-scale predictions problematic (McPherson et al., 2006). A solution to the spatial scaling problem may therefore be a hierarchical multi-scale model approach: "[T]here is mounting evidence that scale-dependent hierarchies of environmental controls are apparent in a number of ecological systems" (Pearson et al., 2004: p. 285). A general hierarchical framework for the scales at which environmental factors affect species distributions was proposed by Pearson and Dawson (2003). They explicitly suggested appropriate spatial extents and environmental variables with reference to SDMs (Figure 3.5). Pearson and Dawson (2003) implicitly assume that large spatial extents are associated with coarse data resolutions, and small extents with fine grain sizes. According to this framework, climate can be considered the dominant factor from global to continental scales (Parmesan, 1996; Pearson and Dawson, 2003), whilst at local scales other factors such as topography and land cover type become increasingly important (Bailey et al., 2002; Pearson and Dawson, 2003; Pearson et al., 2004; Saatchi et al., 2008). Soil type and biotic interactions are only relevant at the smallest scales with correspondingly high resolutions (Pearson and Dawson, 2003). Compliant with this framework, a hierarchical approach combining climatic and remote sensing based models of species distributions was applied in this thesis. The different effective resolutions of both types of environmental variables were explicitly addressed (see Section 6.2).

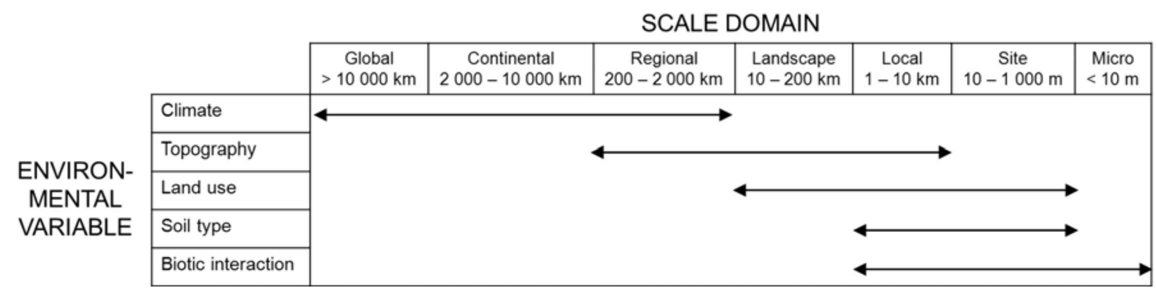


Figure 3.5: Schematic representation of how different environmental variables may affect the distribution of species across varying spatial scales (Pearson and Dawson, 2003).

3.3 Modeling algorithms

The methodological advancements of species distribution models made in the recent past have already been briefly touched in the introduction (Chapter 1) to this thesis. The chronological order in which certain algorithms arose in the species distribution modeling community was elaborated in detail by Elith and Leathwick (2009). In the following, a short overview of existing algorithms for modeling species distributions (Section 3.3.1) is given in order to understand why the *Maximum Entropy* (Maxent) approach (Section 3.3.2) was selected in this thesis.

3.3.1 Overview and classification of algorithms

A full suite of algorithms and techniques is now available to predict species distributions (Appendix C) – each with their certain advantages and drawbacks. The following classification scheme for the existing SDM algorithms was proposed by Pearson (2007) to facilitate algorithm selection based on data availability and data type:

(1) **Does the algorithm require data on observed species absence?** Some algorithms, so-called 'presence-absence' methods operate by contrasting presence sites with sites where the species has been recorded as being truly absent, e.g. *Generalized Linear Models* (GLM), *Generalized Additive Models* (GAM), and *Artificial Neural Networks* (ANN). Since reliable absence data are often not available, other so-called 'presence-only' methods have been developed. Here, three types are distinguished: (1) Methods that rely solely on presence records (e.g. *Environmental Envelopes*). (2) Methods that use 'background' environmental data for the entire study area (e.g. *Maximum Entropy*, *Environmental Niche Factor Analysis*). In this case, the occurrence

localities are also included as part of the background. (3) Methods that sample ‘pseudo-absence’ data from the study region that are used in place of real absence data. The important difference between ‘pseudo-absence’ and ‘background’ approaches is that ‘pseudo-absence’ algorithms do not include presence localities in the pseudo-absences (Pearson, 2007).

(2) **Is the model able to incorporate categorical environmental variables?** While some algorithms (e.g. GLM or GAM) require to treat each class of a categorical variable as a binary ‘dummy variable’ (Franklin, 2009), others such as *Classification Tree Analysis* (CTA) easily allow the association of categorical responses (Franklin, 2009). Maxent is able to directly integrate categorical predictors by the use of binary features (Phillips et al., 2006; Section 3.3.2).

(3) **How good is the algorithm to determine the relative influence of different input variables on the model’s fit or predictive capacity?** For example, ANNs have good predictive ability but identifying the relative contribution of each input variable to the prediction is difficult (Pearson, 2007); they are thus often termed ‘black box’. Maxent enables the user to run a jackknife procedure to determine which environmental variables contribute most to the model prediction (Phillips et al., 2006).

The selection of the appropriate algorithm therefore largely depends on the kind of species data available. Also the type of environmental predictors at hand needs to be considered (Guisan and Thuiller, 2005). Finally, the overall goal of the study is relevant (Guisan and Zimmermann, 2000): Whilst taking into account the type of presence and absence data used for model calibration, Jiménez-Valverde et al. (2008) arranged different modeling methods along the conceptual gradient between potential and realized species distributions (Figure 3.6, see Section 3.1.4 for the definitions of realized and potential distributions). By the integration of current remote sensing data, the aim of this thesis is to model realized species distributions.

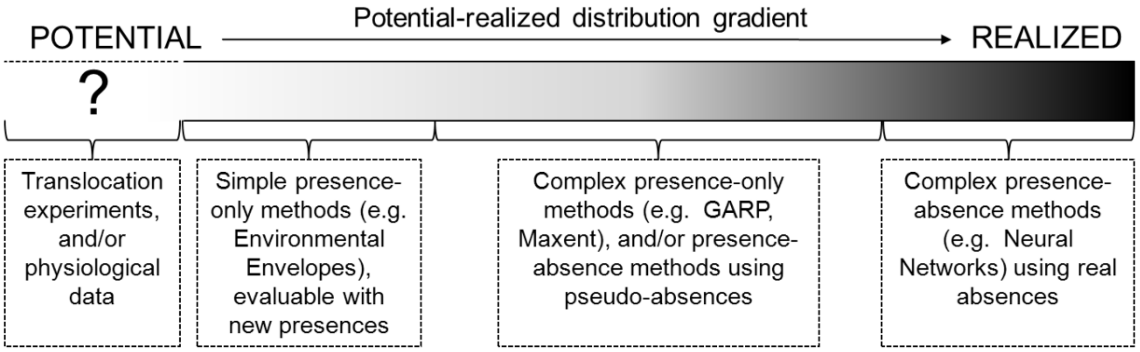


Figure 3.6: Characterization of different modeling techniques, according to the species data available, along a conceptual potential-realized distribution gradient (Jiménez-Valverde et al., 2008).

In practice, differences in predictive power and large variations in predicted range sizes of the above-mentioned modeling approaches have been demonstrated by a number of studies (Moisen and Frescino, 2002; Loiselle et al., 2003; Thuiller, 2003; Segurado and Araújo, 2004; Elith et al., 2006; Hernandez et al., 2006; Pearson et al., 2006; Tsoar et al., 2007; Elith and Graham, 2009). Slight differences between modeling techniques under current climatic conditions may be further accentuated in future projections (Thuiller, 2003) due to the assumptions made during model extrapolation (Pearson et al., 2006). Consistently, these comparative analyses found that novel, complex techniques being able (1) to establish more flexible relationships between the dependent and independent variables and (2) to identify complex relationships in the data, e.g. interactions among environmental variables, outperformed simple methods (Elith et al., 2006; Tsoar et al., 2007). Both Guisan et al. (2007) and Elith et al. (2006), which is the most comprehensive comparative study comparing 16 modeling methods over 226 species from six regions of the world, identified *Boosted Regression Trees* (BRT) and Maxent as the top two techniques, measured by predictive model performance. In the following, the *Maximum Entropy* algorithm used in this thesis is explained in detail.

3.3.2 *Maximum Entropy* algorithm

In this thesis, the *Maximum Entropy* (Maxent) algorithm as implemented in its software version 3.3.3e (Phillips et al., 2004; Phillips et al., 2006) was used. Maxent was chosen for the analyses in consequence of, among other things, the species occurrence data being available (see Section 3.3.1). The algorithm was designed especially for modeling species distributions based on presence-only herbarium and museum records, which are one of the two species data sources of this study (Section 5.1). These collections typically exhibit strong spatial sampling bias (Section 3.4.1) which may be compensated by targeting the ‘background’ points in Maxent (Phillips et al., 2009). In addition, categorical land cover data and continuous remote sensing variables were assessed regarding their usefulness for modeling species distributions in this thesis. Maxent can directly incorporate such categorical predictors and has a high computing efficiency enabling the use of large-scale data layers and the ability to model complex responses to environmental variables (Buermann et al., 2008; Saatchi et al., 2008) as required for this study. Thirdly, the aim of this thesis is to model realized species distributions based on current remote sensing data. Following the framework illustrated in Figure 3.6 (Jiménez-Valverde et al., 2008), Maxent was therefore an appropriate choice. Within several comparative studies as outlined above, Maxent outperformed many other algorithms including GARP, the *Genetic Algorithm for Rule-set Production* (which is suggested together with Maxent in the framework of Jiménez-Valverde et al., 2008). Several further advantages of Maxent in relation to its software implementation are summarized below.

General approach. All species distribution modeling algorithms face the problem of incomplete information given by the species samples that are available. The Maxent software application for species distribution modeling (Phillips et al., 2004; Phillips et al., 2006) solves this by using the *Maximum Entropy* principle as a general-purpose method for making predictions or inferences from incomplete information. A comprehensive overview of the principles and implementation of Maxent for modeling species distributions is given in Dudík (2007). The *Maximum Entropy* approach has already previously been incorporated in machine-learning techniques, e.g. in the field of natural language processing (Berger et al., 1996) or game theory (Grünwald, 2000).

Principle. The concept of entropy has its origins in thermodynamics describing the 'disorder' (Gibbs' entropy) in a system (Gibbs, 1902). The concept has been extended to its broad interpretation as entropy being "a measure of how much 'choice' is involved in the selection of an event or how uncertain we are of the outcome" (Shannon, 1948: Section 6). Shannon developed his theory in the field of communication theory – his term 'choice' is thus equivalent to 'information'. The *Maximum Entropy* principle was later introduced in the context of density estimation by E. T. Jaynes, one of Shannon's proponents, namely as a general answer to the question which is the best approximation for an unknown probability distribution (Jaynes, 1957). Accordingly, "the best approach is to ensure that the approximation satisfies any constraints on the unknown distribution, and that subject to those constraints, the distribution should have maximum entropy" (Phillips et al., 2006). This means that the distribution agrees with everything that is known, but carefully avoids assuming anything that is not known (Jaynes, 1990). Grendár and Grendár discussed the rationale of the *Maximum Entropy* theory in the realm of statistics in detail (Grendár and Grendár, 2001).

For species distribution modeling, the *Maximum Entropy* principle can be explained in two ways – either in geographical or environmental space (see also Section 3.1.3). While the first papers (Phillips et al., 2004; Phillips et al., 2006; Phillips and Dudík, 2008) described Maxent as estimating a distribution across geographical space, Elith et al. (2011) gave a different (but equivalent) characterization that focuses on comparing probability densities in covariate (environmental) space. In the following, the *Maximum Entropy* principle is elaborated from the geographical space perspective based on Phillips et al. (2006).

Implementation. When the *Maximum Entropy* principle is applied to the species distribution modeling problem, the 'unknown distribution' being estimated is the multivariate distribution of suitable conditions in environmental space (Dudík, 2007; Franklin, 2009). The *true* distribution of a species in a certain area can be represented by the probability distribution π over the sample space X, which represents the entire study area. The distribution π assigns a non-negative probability $\pi(x)$ to each point x (any pixel in the study area), and these probabilities sum to 1.0 (Phillips et al., 2006). The known information on π is characterized by the averages of the environmental data and functions thereof (called 'features', see below) under π. The *approximate* distribution of π is also a probability distribution, denoted by $\hat{\pi}$. The entropy H of $\hat{\pi}$ is defined as

$$H(\hat{\pi}) = -\sum_{x \in X} \hat{\pi}(x) \ln \hat{\pi}\,(x)$$

The entropy is therefore non-negative and is at most the natural logarithm of the number of elements in X (Phillips et al., 2006). The features represent a set of known functions f

on *X*, and each feature f_j assigns a real value $f_j(x)$ to each point *x* in *X*. The expectation of the feature f_j under π is denoted by $\pi[f_j]$ and defined as

$$\pi[f_j] = \sum_{x \in X} \pi(x)\, f_j(\mathrm{x})$$

The feature expectations $\pi[f_j]$ can be approximated using a set of sample points *x* drawn independently (with replacement) from *X* according to the probability distribution π. The empirical average of f_j is termed $\tilde{\pi}[f_j]$ and defined as

$$\frac{1}{m} = \sum_{i=1}^{m} f_j(\mathrm{x_i})$$

$\tilde{\pi}$ is the uniform distribution on the sample points and $\tilde{\pi}[f_j]$ is used as an estimate of $\pi[f_j]$. According to the *Maximum Entropy* principle, the aim is to find the *approximate* probability distribution $\hat{\pi}$ of maximum entropy subject to the constraint that each feature f_j has the same mean under $\hat{\pi}$ as observed empirically. The *approximate* probability distribution $\hat{\pi}$ can alternatively be described as a Gibbs distribution, which takes the general form

$$\hat{\pi} = q_\lambda(x) = \frac{e^{\lambda \cdot f(\mathrm{x})}}{Z_\lambda}$$

where λ is a vector of the feature weights, $f(x)$ is the vector of all *n* features, and Z_λ is a normalizing constant that ensures that q_λ sums to 1. In this form, the best $\hat{\pi}$ is found when λ maximizes the probability of occurrence of the sample points, which is equal to minimizing the negative log likelihood of samples (Phillips et al., 2006).

To summarize, the algorithm first characterizes the empirical distribution π by a set of features over the sample sites. *A priori*, a uniform distribution is approximated by setting λ to 0 for each feature. A number of iterations are performed in which λ, the vector of the feature weights associated with the environmental variables, is adjusted to maximize the average probability of the point localities and to decrease the negative log likelihood (Buermann et al., 2008). The algorithm continues to a threshold beyond which no substantial improvements can be made without violating the *Maximum Entropy* principle (Phillips et al., 2004; Phillips et al., 2006).

Regularization. As described so far, the *Maximum Entropy* algorithm can be prone to overfitting the model to the training data since the empirical features will not equal the true means, but only be close to their empirical values (Phillips et al., 2006). In the Maxent implementation, the constraint is thus relaxed by adding a constant factor β. This so-called *regularization* is equivalent to smoothing the model and making it more regular (Elith et al., 2011). The regularization parameter β affects how focused or closely fitted the output distribution is. It allows the estimated averages to deviate from the average of the empirical features within some bounds. Regularization forces Maxent to focus on the most important features; such models are less likely to be overfitted, because they have fewer parameters (Phillips et al., 2006). The default regularization parameters in Maxent are tuned to the sample size based on a comprehensive dataset of 226 species from 6 regions (Dudík et al., 2004; Phillips et al., 2004; Dudík, 2007). Smaller values than the default 1.0 will result in a closer fit to the presence records, but may lead to overfitting; larger regularization multipliers will give a more spread out distribution. Maxent models with inappropriate regularization parameters show not only reduced ability to infer habitat quality and variable importance but also decreased transferability to other time periods (Warren and Seifert, 2011).

Covariates and features. The unknown probability distribution is constrained by the values of the covariates (environmental data) at the presence localities, the so-called *features* in Maxent. Typically, the constraints on the target distribution require the expectation of every feature to match its empirical average (Dudík, 2007). Features are either the covariates themselves or functions thereof that account for nonlinear species responses to their environment (Austin, 1980) and can take different forms of functions called *feature types* (Phillips et al., 2004). Maxent version 3.3.3e used for this thesis has six feature types implemented: *linear*, *product*, *quadratic*, *hinge*, *threshold*, and *categorical* (Elith et al., 2011). *Linear* features are the covariates themselves. *Product* features are products of all possible pair-wise combinations of covariates, allowing interactions between the covariates to be fitted (Phillips et al., 2006). *Quadratic* features (squared values of the covariates) model the species' tolerance for variation from its optimal conditions (Phillips et al., 2006). *Threshold* features are equivalent to a piecewise constant spline and allow a 'step' in the fitted function; *hinge* features are similar except they allow a change in gradient of the response that is equivalent to a piecewise linear spline (Elith et al., 2011). Finally, *categorical* features are only calculated for categorical data, such as the land cover classification analyzed in this thesis, and are binary indicators showing the pixel-level memberships for each of the classes in a categorical predictor (Elith et al., 2011).

By default, the Maxent algorithm restricts the model to simple features if few presence data are available (*linear* is always used; *quadratic* ≥ 10 samples; *hinge* ≥ 15; *threshold* and *product* ≥ 80) since few samples provide limited information for determining the relationships between the species and its environment (Pearson et al., 2007; Elith et al., 2011). In Maxent, features "are formed 'behind the scenes'" (Elith et al., 2011: p. 46) and many features can be fitted for the same covariate giving a potentially complex function. The exponent in the Maxent model is a sum of the features used and usually defined over many features, meaning that in most models there will be more features than covariates (Elith et al., 2011).

Output formats. The Maxent software has three output formats implemented that are monotonically related, but are scaled differently and have different interpretations (Phillips, 2008; Phillips, 2009). The *raw* output is the Maxent exponential model itself. It indicates the probability (for the user) to be located within a certain pixel (geographical space) or observing a certain combination of environmental values (environmental space) when detecting the species. Hence, the probabilities modeled for each cell in the study area must sum up to one within the *raw* output and are thus dependent on the number of background and presence sites used in the model. In the *cumulative* format, values from 0 to 100 represent the range of probabilities predicted by the model across the entire study area. While being scale-independent, this format is not necessarily proportional to the probability of presence – for example in the case of a generalist species where probability values are similar across the entire area. The *logistic* format (which was used for the analyses conducted in this thesis) estimates the probability of presence for a certain pixel (Phillips and Dudík, 2008). Since prevalence cannot be determined from presence-only data, logistic probabilities are calibrated so that typical sampling sites have probabilities of presence of about 0.5 (Phillips, 2008; Elith et al., 2011).

Model gain. The model gain referring to the *raw* data model is closely related to deviance, a measure of fit used in GLM and GAM (Phillips, 2009). Maxent iteratively generates a probability distribution over pixels in the grid – starting from the uniform distribution and repeatedly improving the fit to the data (see the paragraph on Maxent implementation). The gain is defined as the average log probability of the presence samples, minus a constant that makes the uniform distribution have 0 again (Phillips et al., 2006). This means that it is an average measure of how much higher the assigned probability value of a true (test or training) presence pixel is compared to a random pixel. At the end of each model run, the gain indicates how closely the model is concentrated around the presence samples (Phillips, 2009). For example, if the gain is equal to 3, it means that the average likelihood of the presence samples is e^3 = 20.1 times higher than that of a random background pixel.

Variable contribution and response curves. While the Maxent model is being trained, it keeps track of which environmental variables are contributing to fitting the model (Phillips, 2009). To exactly quantify the relative contribution of all predictors, Maxent makes use of a jackknife (leave-one-out) procedure. Thus, in addition to the full model including all predictors, (1) models trained with all except one of the predictors as well as (2) models with only one predictor used in isolation are produced. For all cases, the model gain (see above) is recorded. In case the model without a certain predictor yields a higher gain or the use of that predictor in isolation leads to a very small or even negative model gain, presumably a better or at least more parsimonious model can be obtained without it (Phillips et al., 2006). In each of the iterations of the model training process, the observed increase or decrease in the model gain is assigned to the environmental variable(s) that the feature depends on. Model gain is added up for all model iterations and determines the importance of the respective variable.

In addition, species response curves are generated by either varying each environmental variable (while all other environmental variables are kept at their average sample values) or using each variable in isolation and disregarding all other variables (Phillips et al., 2006). Since the exponent in the Maxent model is a sum of the features used, also the shapes of the response curves depend on the type of features used. For example, if only threshold features are used, the sum of these features will be a threshold function, too, and the response curve will always be a step function (Phillips, 2009).

Background samples. To reduce computational time, Maxent uses only a representative sample of 10,000 random localities in the study area, the so-called *background* sample to estimate environmental conditions. After training the model based on these points, it is projected onto the entire extent of the study area using the established model coefficients. In this thesis, this default was specifically modified to use certain sites of the study region as background, the so-called *target-group background* approach (see Section 3.4.1).

Replication. Multiple model runs for the same species are implemented in the software using three different resampling techniques of the presence data (repeated subsampling, bootstrapping, and cross-validation) whereas only bootstrapping was utilized here.

Advantages. Besides its good performance compared to other algorithms, especially for small sample sizes (Hernandez et al., 2006; Pearson et al., 2007), Maxent has several other features that make it very useful for predicting species distributions: It can incorporate interactions between different variables via its built-in features (Phillips et al., 2006) and uses a deterministic framework that is guaranteed to converge to the optimal (*Maximum Entropy*) probability distribution (Phillips et al., 2006). The concise mathematical definition of Maxent means that the approach and underlying assumptions are transparent (Pearson et al., 2007). Maxent has been described as a *generative* modeling approach (as opposed to *discriminative* regression-based methods, Phillips and Dudík, 2008) which models the species distribution directly by estimating the density of environmental covariates conditional on species presence (Franklin, 2009). The algorithm is well-regularized and effectively avoids model overfitting (Phillips et al., 2006; Phillips and Dudík, 2008). Model outputs of Maxent are continuous allowing (1) fine distinctions to be made between the modeled suitabilities of different areas and (2) a great flexibility in the choice of threshold criteria if binary predictions are desired (Phillips et al., 2006). Maxent also has several aspects that support the interpretation of the model results (Buermann et al., 2008) such as automatically generated species response curves and a built-in jackknife option through which the importance of individual environmental data layers can be estimated.

Drawbacks. Because Maxent uses an exponential model for probabilities, it can give very large predicted values outside the range of conditions that were observed in the training data. However, this was not of relevance in this thesis since no spatial or temporal extrapolation was applied. Extrapolation should therefore be done carefully, though the problem may be alleviated by resetting these values to match the upper or lower values found in the study area (called *clamping*, Elith et al., 2011). Further, since Maxent is not a typical statistical method as GLM or GAM, there are fewer guidelines available for its use in general (Phillips et al., 2006).

Applications. Since becoming available in 2004, Maxent has been extensively used for modeling species distributions in a variety of studies. Elith et al. (2011) recently summarized published examples across many ecological, evolutionary, conservation, and biosecurity applications. Both governmental and non-governmental organizations have adopted Maxent for large-scale biodiversity mapping applications (Elith et al., 2011), including the *Point Reyes Bird Observatory* online application (http://www.prbo.org/) and the *Atlas of Living Australia* (http://www.ala.org.au/).

3.4 Selected methodological aspects

While the model algorithm is the 'core' component in SDM (Pearson, 2007), it is still only one aspect of the broader modeling process (see Figure 3.4). Several other key steps, components, and features important at various stages of the modeling procedure need to be considered (Guisan and Zimmermann, 2000; Guisan and Thuiller, 2005; Elith and Leathwick, 2009; Franklin, 2009). In the following, selected methodological aspects relevant for this thesis will be shortly outlined and discussed.

3.4.1 Spatial sampling bias and sample size

The richest – and in many cases the only – source of species occurrence localities are **natural history museum and herbarium collections** (Stockwell and Peterson, 2002; Graham et al., 2004). However, those collections typically include no information about the *failure* to observe the species at any given location (Phillips et al., 2004). To reliably confirm the absence of a species from a locality is very difficult (Gu and Swihart, 2004) and becomes almost unaffordable given the coarse-resolution grid cells that are used in most modeling studies (Jiménez-Valverde et al., 2008). Usually, the only reliable information on the distribution of organisms is hence their recorded presence (Jiménez-Valverde et al., 2008). Only in well-explored regions of the world, data sets exist in sufficient spatial resolution and sample density so that the absence of records of a certain species can be interpreted as *true* absence rather than being attributed to the lack of data or to insufficient sampling effort (Soberón and Peterson, 2005). Beyond this, species records often have been collected over several years, decades or even longer time periods, which may be problematic since species distributions may be dynamic over relatively short time scales (Franklin, 2009). In addition, the data provides no information about what proportion of the larger grid cell is occupied by the species (McPherson et al., 2006) and has inconsistent locational precision, especially of historical records collected before the age of GPS (Graham et al., 2004). Nevertheless, locality information in museum and herbarium collections is mostly supported by vouchers (specimen) so that taxonomic accuracy can be verified (Graham et al., 2004).

In addition, occurrence data are often recorded without or with different sampling strategies (Reddy and Davalos, 2003). Instead, they inevitably tend to be clustered around field stations, settlements, and already known species localities (Pearson et al., 2007) as well as areas close to roads or waterways which generally are easier to access (Reddy and Davalos, 2003; Kadmon et al., 2004) – termed *purposive* sampling (Edwards Jr. et al., 2006). This means that the presence data contain substantial **spatial bias** which may impact the assumption of data independence as roads and waterways are often correlated with certain topographical and vegetation features which in turn influence species distributions (Dudík et al., 2005). Especially in such cases where the species data exhibit a strong spatial sampling bias, the selection of background data can have a significant impact on the predicted species distributions (Costa et al., 2010). Various methods have been proposed for the selection of 'background' or 'pseudo-absence' sites (Engler et al., 2004; Chefaoui and Lobo, 2008; VanDerWal et al., 2009; Mateo et al., 2010). For Maxent, Phillips et al. (2009) suggested comparing the occurrences with background points reflecting the same spatial bias. The underlying idea is that a model based on presence and background data with the same bias will not focus on the sample selection bias but on any differentiation between the distribution of the occurrences and that of the background (Phillips et al., 2009). This *target-group background* approach (Ponder et al., 2001; Anderson et al., 2003) has successfully been applied to Maxent (Phillips et al., 2009; Mateo et al., 2010) and typically produces predictions which are spatially more differentiated within sampled areas and at the same time predictions more spread towards unsampled regions (Phillips et al., 2009). To account for the spatial sampling bias present in the species data available for this thesis (Section 5.1), the *target-group background* approach was applied here as well.

Further, small **sample size** (few presence localities) has been shown to be a significant source of model instability and errors in models (Guisan and Thuiller, 2005). In general, prediction accuracy increases with sample size (Carroll and Pearson, 1998; Stockwell and Peterson, 2002), but collecting new data is costly and thus needs to be optimized (Hirzel and Guisan, 2002). Some studies have aimed to identify the minimum sample size required for deriving robust predictions and have shown that different modeling methods and algorithms might require different minimum sampling sizes (Stockwell and Peterson, 2002; Kadmon et al., 2003). Maxent seems to perform better than GARP when sample size is small, i.e. less than ten presence points are available (Pearson et al., 2007; Costa et al., 2010).

3.4.2 Spatial autocorrelation

Many statistical models make the standard assumption that observations of the response and predictor variables are independent (Franklin, 2009; Skidmore et al., 2011). However, this violates the ‘first law of geography’ (Tobler, 1970) that “near things are more related than distant things” (Tobler, 1970: p. 236), i.e. have similar values. **Spatial autocorrelation** (SAC) is a typical phenomenon in ecology (Legendre and Fortin, 1989; Legendre, 1993) and defined as the covariation of variables with space so that their values are “more similar (positive autocorrelation) or less similar (negative autocorrelation) than expected for randomly associated pairs of observations” (Legendre, 1993: p. 1659). SAC occurs at all spatial scales (Dormann et al., 2007) and arises because variables are likely to either influence each other or to be influenced by the same pattern-generating processes (Miller et al., 2007; Franklin, 2009). These processes can be classified as (1) **endogenous** – generated by or associated directly with the response variable of interest, e.g. dispersal or competition, or (2) **exogenous** – independent of the variable of interest, e.g. disturbance, historical barriers or spatially structured environmental gradients.

Some potential consequences of SAC in the context of species distribution modeling are (based on Franklin, 2009): The precision of coefficients is decreased, resulting in higher likelihood of Type I (false positives) errors (Bahn et al., 2006). Further, variable selection may be predisposed towards more strongly auto-correlated predictors, the so-called ‘red-shift’ (Lennon, 2000). Similarly, broad-scale predictors are often selected over local or fine-scale predictors (Diniz-Filho et al., 2003). Several methods to test for SAC such as the *Moran’s I* and *Geary’s C* indices or semi-variograms (Dormann et al., 2007) have been proposed. Potential solutions to SAC include the (re)sampling of plots at sufficient spatial distance to avoid autocorrelation (Guisan and Zimmermann, 2000) or removing the spatial structure from the data, e.g. by culling observations that are within some pre-determined distance of other observations (Hawkins et al., 2007). However, less than 20% of species distribution modeling studies so far explicitly addressed SAC (Dormann, 2007). This thesis includes the analysis of SAC in climate and remote sensing predictors in the context of the hierarchical modeling framework (Section 6.2.2).

3.4.3 Predictor multi-collinearity and interactions

Different types of environmental variables and their significance for the characterization of species distributions have been discussed in Section 3.2.1. Further, the number of environmental predictors used has a large impact on the modeled sizes of species distribution ranges (Beaumont et al., 2005). A species-specific selection of appropriate predictors is therefore a crucial step in species distribution modeling and can significantly influence model results (Peterson et al., 2008; Phillips et al., 2009). Besides the selection of ecologically relevant predictors, two major methodological aspects need to be considered, namely (1) the multi-collinearity of environmental variables and (2) the interactions between different predictors. The inherent collinearity of explanatory variables hampers statistical analyses and the ecological interpretability of SDMs (Graham, 2003; Heikkinen et al., 2006; Dudík, 2007; Baldwin, 2009). In particular, ecologically more causal variables may be excluded from the models if other correlated predictors explain the variation of the response variable better in statistical terms (MacNally, 2000; Luoto et al., 2002). Nevertheless, predictor multi-collinearity has been neglected in nearly all previous studies which integrated remote sensing data into species distribution models (but see Buermann et al., 2008). Even though Maxent is largely robust to predictor collinearity regarding its effect on the spatial model predictions (maps) (Buermann et al., 2008), interpretations of species response curves should be made with caution when variables are correlated (Dudík, 2007).

Different solutions have been proposed and applied to this problem, such as the combination of environmental data into few orthogonal axes by *Principal Component Analysis* (Graham, 2003; Rushton et al., 2004; Dormann et al., 2008; Viña et al., 2010; Aranda and Lobo, 2011) or the selective removal of predictors that are too highly correlated with one or more others (Elith and Burgman, 2002; Riordan and Rundel, 2009; Giovanelli et al., 2010; Cord and Rödder, 2011). The combination of different predictors in principal components is data-dependent and thus changes through time (Schowengerdt, 2007).

Secondly, interactions between environmental predictors should be considered in the modeling process as the incorporation of predictor interactions has been shown to improve the model fit (Thuiller et al., 2003). Species are “likely to respond to multiple factors” (Pearson, 2007: p. 7); variable interactions therefore have too often been omitted from SDMs (Austin, 2002), since they exponentially increase the number of predictors (Guisan and Thuiller, 2005). In Maxent, interactions between environmental variables are explicitly addressed by the calculation of its internal *features* (see Section

3.3.2). In practice, thus the use of complex model algorithms capable of integrating several environmental variables such as Maxent is reasonable.

3.4.4 Model thresholding

To aid the interpretation and presentation of model results, it is often useful to identify a threshold value above which model outputs are classified as being ‘positive’, indicating species presence or suitable environmental conditions (Pearson et al., 2004). The choice of this (possibly subjective) threshold is of great significance since model predictions and accuracy as well as the predicted species prevalence (frequency of presences) may be very different depending on the value applied. Especially predictions of species with low prevalence and poor model quality are the most sensitive to the choice of this threshold (Freeman and Moisen, 2008). The threshold value should be selected based on the specific aim and application of the study considering the ‘costs’ of the different types of misclassification (Liu and Wu, 2005; Franklin, 2009). For example, if the aim is to identify potential (re)introduction sites for an endangered species, a relatively high threshold should be chosen in order to identify only the sites with the highest suitability.

By convention, threshold values were often set to the traditional default of 0.5 (Freeman and Moisen, 2008; Franklin, 2009). Because of the bias in probability outputs due to species prevalence, this fixed threshold value normally does not correspond to the actual threshold above which the species is more likely to be present than absent (Jiménez-Valverde and Lobo, 2007). This is especially true for Maxent as the logistic model output estimates the probability of presence by assigning a probability score of about 0.5 to typical presence localities (Phillips, 2008; Phillips, 2009; see Section 3.3.2). The use of the 0.5 threshold is therefore not reasonable for Maxent models. Alternatively, thresholds can be chosen to maximize the agreement between observed and modeled distributions judged by various criteria (Freeman and Moisen, 2008) or to minimize the false negative fraction (Pearson et al., 2004). Several recent studies (Liu and Wu, 2005; Jiménez-Valverde and Lobo, 2007; Freeman and Moisen, 2008) have evaluated the value of different threshold criteria for a wide range of sample sizes and species prevalence scores. Accordingly, they recommend either the use of combined model sensitivity and specificity approaches or the matching of predicted and observed prevalence to determine appropriate thresholds. Based on these findings, the *minimum training presence*, *10 percentile training presence* and *maximum sensitivity and specificity* thresholds were applied in this thesis depending on the usage of the respective binary map (see Section 6.2.4, Chapter 7).

3.4.5 Model evaluation

“Essentially, all models are wrong, but some are useful” (Box and Draper, 1987: p. 424). What Box and Draper (1987) intended to say is that all models are simplifications of the reality and thus have prediction errors. Accordingly, different sources of errors and uncertainty in species distribution modeling can be identified (Guisan and Thuiller, 2005). These include algorithmic errors, environmental and data errors (e.g. a failure to include relevant environmental predictors in the model), and biological errors (e.g. equilibrium theory, inefficient sampling scheme). In addition, “part of what is considered a mistake by standard evaluation procedures might actually be adequately explained by ecological theory” (Guisan and Thuiller, 2005: p. 1003), for example by source-sink-dynamics. Franklin (2009) summarizes the two concepts of model quality as *quantitative predictive performance* compared to *model credibility* (ecological realism). Most importantly, the kind of accuracy assessment used should thus have a close connection to the intended use of the model (Peterson, 2006; see Section 3.5 for exemplary applications) and the biology of the study species (Fielding, 2002).

Diverse opinions exist on what properties of a model are important (Anderson et al., 2003; Vaughan and Ormerod, 2005) and a wide range of metrics and techniques for statistically evaluating SDMs and their predictive performance have been developed (Fielding and Bell, 1997; Pearce and Ferrier, 2000; Guisan and Thuiller, 2005). In general, models are evaluated by their ability to predict values which were not used for calibrating the model. Ideally, the validation (test) samples are completely independent from the calibration (training) samples, e.g. by the use of herbarium collection and field inventory data as accomplished in this thesis (Chapter 7). Due to the common lack of such independent data, often some kind of data splitting provides for both validation and calibration samples (Fielding and Bell, 1997). Because species occurrence data are often sparse, data partitioning techniques such as bootstrapping (sampling with replacement) are used. A ‘rule of thumb’ for the proportion of testing data is given by

$$1/\,[1 + (p-1)^{\frac{1}{2}}]$$

where p is the number of predictors (Huberty, 1994).

As discussed in the previous section, either the continuous probabilistic predictions can be directly evaluated or a threshold value can be imposed to obtain binary data. Accordingly, different evaluation metrics (either threshold-dependent or threshold-independent) are available. The threshold-dependent measures largely overlap with the techniques for assessing the accuracy of thematic classifications of remotely sensed data (Congalton, 1991), e.g. the so-called ‘confusion matrix’ or the Kappa statistic. Besides these ‘global’ (Franklin, 2009) measures of model accuracy also spatial patterns of error distribution are important (Elith and Leathwick, 2009). In the following, the threshold-dependent and threshold-independent measures used in this thesis are described.

Threshold-dependent: *Fractional predicted area (FPA).* After applying a threshold value to the continuous prediction, the *fractional predicted area* (the fraction of the total study area predicted as species presence) can be estimated. It is a surrogate for the standard commission rate (fraction of false presences) in the case of presence-only data.

Threshold-independent: *Area under curve (AUC) of the receiver operating characteristic (ROC).* The “awkwardly-named” (Franklin, 2009: p. 222) *area under curve* (AUC) is a non-parametric accuracy measure developed in the fields of medical diagnosis and signal processing and is independent of prevalence (Zweig and Campbell, 1993). The AUC is calculated by summing the area under the *receiver operating characteristic* (ROC) plot. ROC plots are obtained by plotting all *sensitivity* values (true positive fraction) on the y-axis against their equivalent values (1-*specificity*, false positive fraction) on the x-axis to provide a quantitative representation of the trade-offs between omission and commission errors across all possible thresholds (Figure 3.7). AUC values can range from 0.5, indicating that the model performs not better than expected by chance, to 1.0 suggesting perfect discriminatory abilities. An approximate guide for classifying the model accuracy based on AUC scores was proposed: 0.90-1.00 = excellent, 0.80-0.90 = good, 0.70-0.80 = fair, 0.60-0.70 = poor and 0.50-0.60 = fail (Swets, 1988). If the species distribution covers the proportion *a* (e.g. 0.03) of all pixels in the study area, the maximum AUC is exactly 1-*a*/2 (Phillips et al., 2006), which is in this case 0.985. Therefore, more common and widespread species typically produce models with lower AUC scores. However, for presence-only data, the prevalence *a* is usually not known: Therefore, it cannot be judged how close to the optimal AUC score the result is (Phillips et al., 2006). In the case of presence-only data, training AUC describes the ability of the model to correctly score a randomly selected presence site higher than a randomly selected background site (Phillips et al., 2009). This is applicable also for Maxent which treats the random background pixels as negative instances and species presence pixels as positive instances. In Maxent, test AUC values describe the model's ability to predict a fraction of the species records omitted during model training (Phillips et al., 2006).

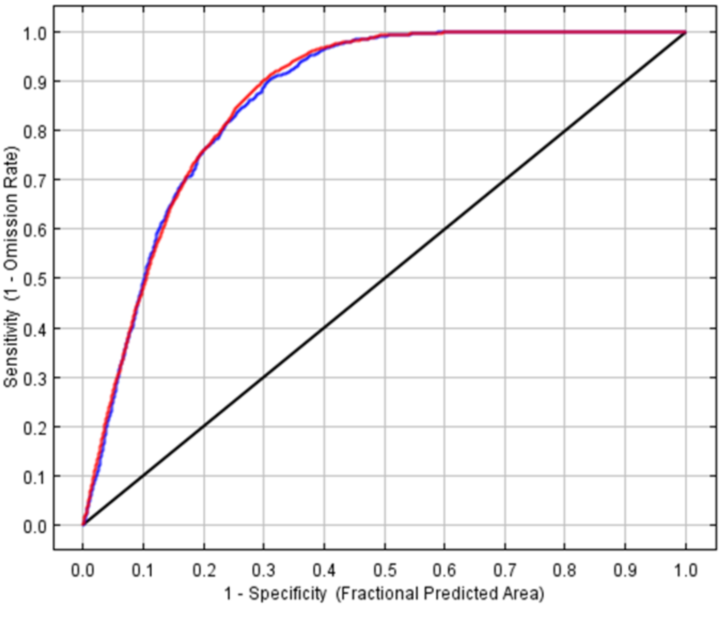


Figure 3.7: Typical receiver operating characteristic (ROC) plot produced by Maxent. The score of the *area under curve* (AUC) for training (red) and test (blue) data is AUC = 0.865 (training) and AUC = 0.862 (test), respectively. The random prediction or chance performance line (AUC = 0.5) is shown in black.

The AUC is not sensitive to transformations of the predicted probabilities that preserve the ranks (Reineking and Schröder, 2006) and to species prevalence (Manel et al., 2001) as the Kappa statistic is (McPherson et al., 2004; Vaughan and Ormerod, 2005). AUC therefore has become one of the standard methods for assessing performance of SDMs (Anderson et al., 2003). However, multiple assessments based on several measures should be preferred over reporting of a single measure (Fielding, 2002) – for example AUC does not cover all aspects of model errors (Austin, 2007; Lobo et al., 2008) and does not indicate the quality of model calibration (Reineking and Schröder, 2006).

Threshold-independent: *Statistical deviance*. The statistical deviance is a measure of goodness of fit between observed species occurrence data and fitted model values (Franklin, 2009). While AUC is only rank based, deviance takes into account the magnitude of the difference between the prediction (continuous) and the observation (binary presence-absence data). The larger the estimated deviance, the poorer is the fit of the model to the test data. Model deviance was in this thesis calculated as implemented in the R package ‘dismo’ (R Development Core Team, 2011).

3.5 Typical applications

Reliable maps of potential and actual species distributions or habitat suitability are required for many aspects of environmental research, resource management, and conservation planning (Franklin, 2009). Besides their importance as a research tool in autecology (for understanding the relationship between a species and its biotic and abiotic environment), SDMs have successfully been used to investigate a variety of scientific issues (Guisan and Thuiller, 2005). These include the selection of priority sites for conservation, habitat restoration, invasive species management, biodiversity assessments, population viability analysis, community and ecosystem modeling, and predictions of the effects of environmental change on species and ecosystems (Guisan and Thuiller, 2005; Peterson et al., 2006; Pearson et al., 2007; Franklin, 2009; Elith et al., 2011). Trends in SDM usage reveal subtle but important shifts in intention (Elith and Leathwick, 2009): While most early studies had a strong ecological focus on the drivers of species distributions and functional relationships between species and environment, many recent papers focus on spatial predictions of species distributions as useful products for conservation planning and land management (Guisan and Zimmermann, 2000; Elith and Leathwick, 2009). Another application that is currently much debated in the literature is the projection of species distributions based on climate change scenarios (Guisan and Thuiller, 2005). As a more detailed discussion is beyond the scope of this thesis, Table 3.2 summarizes characteristic types of SDM applications with their relevant references.

Table 3.2: List of potential uses of species distribution models in ecology and conservation biology (partly based on Guisan and Thuiller, 2005 as well as Peterson et al., 2006).

Type of use	***Sample references***
Understanding ecological requirements of species	(Luoto et al., 2006)
Evolutionary biology and phylogeography	(Hugall et al., 2002; Graham et al., 2004; Kozak and Wiens, 2006; Raxworthy et al., 2007; Schulte et al., 2012)
Identification of unknown populations or species	(Fleishman et al., 2002; Raxworthy et al., 2003; Engler et al., 2004; Bourg and McShea, 2005; Guisan et al., 2006)
Reserve and conservation planning, selection of priority sites	(Araújo and Williams, 2000; Ferrier, 2002; Araújo et al., 2004; Kremen et al., 2008; Thorn et al., 2009; Wilting et al., 2010)
Ecological restoration and species translocation or reintroduction	(Pearce and Lindenmayer, 1998; Hirzel et al., 2004; Martínez-Meyer et al., 2006)
Invasive species prediction and management	(Peterson, 2003; Andersen et al., 2004; Cord et al., 2010; Roura-Pascual et al., 2010; Rödder and Lötters, 2010)
Population viability analysis	(Beissinger and McCoullough, 2002)
Modeling biodiversity/species assemblages from individual species predictions	(Ferrier et al., 2002; Maes et al., 2003)
Climate change	(Araújo et al., 2004; Heikkinen et al., 2006; Araújo et al., 2006; Thuiller et al., 2006; Iverson et al., 2008)
Land cover change	(Pearson et al., 2004; Thuiller et al., 2004; Sánchez-Cordero et al., 2005; Peterson et al., 2006; Wisz et al., 2008)

3.6 Summary

This chapter summarized the conceptual and methodological underpinnings of correlative species distribution models as applied in this thesis. Starting from the ecological basis of species distribution modeling, first the ‘niche concept’ and its implications regarding the terminology used here were described. Since models in this thesis are applied to rather predict the *geographical* distribution of the target species (than to study the characteristics of their distributions in *environmental* niche space), the term ‘species distribution model’ is used. Model predictions developed from climate date or remote sensing data are termed ‘climatic suitability’ and ‘biotope suitability’, respectively. Further, the conceptual framework of model calibration in ‘environmental space’ and model projection into ‘geographical space’, which is one of the central principles for the implementation of species distribution modeling algorithms, was illustrated. The chapter included other relevant presumptions for this thesis such as the ‘(pseudo)equilibrium postulate’ and the assumption of environmental completeness of the species occurrence data.

Beyond the fundamental differentiation into *abiotic* and *biotic* factors, various conceptual frameworks for the classification of environmental variables used in species distribution models exist. This classification supports the selection of candidate variables and a process-based understanding of predictor importance. According to these frameworks, the climate data used in this thesis mainly represent so-called *scenopoetic* variables, but also include physiologically limiting *direct* variables. No terminology for the contribution of remote sensing data in characterizing species distributions exists so far, though remote sensing predictors are rather *indirect* or *resource* than *direct* variables. The above-mentioned environmental variables are relevant at different spatial scales (grain and extent). Hence, the use of hierarchical schemes to model species distributions has been proposed. Compliant with this framework, a hierarchical modeling approach combining purely climatic models with models developed from remote sensing data is applied in his thesis.

Special attention was also given to the criteria important for the selection of the appropriate modeling algorithm and in particular the *Maximum Entropy* (Maxent) algorithm utilized in this thesis. This algorithm was designed especially for modeling species distributions based on presence-only records and data with a strong spatial bias such as the species occurrence data available for this thesis. Further, Maxent can directly incorporate categorical predictors, such as the land cover classification analyzed in this thesis. The algorithm has a high computing efficiency enabling the use of large-scale data layers and the ability to model complex responses to environmental variables as required for this study. Within several comparative studies, Maxent outperformed many other algorithms.

Finally, selected methodological aspects of species distribution modeling (e.g. spatial bias, predictor multi-collinearity, and model thresholds) that were relevant for and considered in this thesis were briefly outlined. The chapter concluded with the manifold applications of species distribution models in ecological research and conservation management.

4 Remote sensing for modeling plant species distributions and richness – State of the art and new challenges

Remote sensing “has come of age” (Mather, 2010: p. 739) in environmental modeling and numerous studies have shown the suitability and unique information provided by remote sensing data and techniques for biodiversity-related issues. Remote sensing data have been used to describe biodiversity at multiple levels of ecological organization (individuals, species distributions, community compositions, and species diversity) from local to global scales (Franklin, 2009). Most remote sensing based studies of species richness address terrestrial habitats, although a few analyses on marine and freshwater ecosystems have been conducted (Nagendra, 2001). The *Sourcebook on Remote Sensing and Biodiversity Indicators* (Strand et al., 2007) provides a broad overview of the manifold applications of remote sensing for assessing the extent and status of selected ecosystems and protected areas or for quantifying habitat fragmentation. In addition, several review articles on the potential of remote sensing data to assess biodiversity and ecosystem functioning were published during the last decade (Nagendra, 2001; Kerr and Ostrovsky, 2003; Turner et al., 2003; Gillespie et al., 2008; Muchoney, 2008; Xie et al., 2008; Skidmore et al., 2011). The articles of Gillespie et al. (2008) and Xie et al. (2008) also include a detailed listing of the characteristics of passive and active sensors that can be reasonably used for mapping vegetation as well as measuring and modeling biodiversity.

Besides this general increase in the use of remote sensing for biodiversity research, also the number of studies particularly proposing or addressing the use of remote sensing data in species distribution modeling has lately multiplied (Figure 4.5). While this thesis was being composed, the quantity of such articles has considerably increased. However, the analysis of remote sensing variables in species distribution modeling is still lacking a conceptual underpinning as well as a systematic examination (e.g. regarding different remote sensing products or species traits). As one example, Bradley and Fleishman recently asked “Can remote sensing of land cover improve species distribution modelling?” (Bradley and Fleishman, 2008: p. 1158) and gave a collection of different examples where remote sensing or land cover data could (not) successfully been applied in species distribution models.

In the following, a more differentiated view regarding the potential of remote sensing data is given. In particular, important concepts and criteria for selecting remote sensing data in this thesis (and in general) are outlined. The aim of this chapter is to provide the theoretical background and conceptual framework for systematically modeling (plant) species distributions (Chapter 6) and species richness (Chapter 7) based on remote sensing data.

In this chapter, first the conceptual transition from species *mapping* to distribution *modeling* depending on the aim of the study and the spatial resolution and geographical extent of the remote sensing data used is outlined (Section 4.1). The underlying plant-biotope relationships for species distribution modeling as seen from a remote sensing perspective are addressed in Section 4.2. The following section (Section 4.3) gives an overview of previous studies that used the *Maximum Entropy* algorithm in combination with remote sensing data (mainly *Vegetation Indices*). In Section 4.4, the usefulness and ecological relevance of various remotely sensed products (*Enhanced Vegetation Index*, *Land Surface Temperature*, and *Reflectance*; see Chapter 5 for technical aspects and data specifications) that were utilized in this thesis are summarized. A specific view on the advantages and caveats of using multi-temporal vegetation index data is given in Section 4.4.3. Finally, the important background for using remote sensing data to predict species richness is summarized in Section 4.5.

4.1 Mapping or modeling species distributions? The trade-off between grain and extent

In general, three different approaches (Figure 4.1) to map or model species distributions have been distinguished (Nagendra, 2001; Kerr and Ostrovsky, 2003; Strand et al., 2007): (1) the **direct mapping or identification** of individuals, populations or mono-specific associations in large, contiguous units, (2) the **mapping of habitats or biotopes** to indirectly infer data on species distributions based on *a priori* knowledge from field surveys, and (3) the analysis of proxy variables for the use in predictive **species distribution models** (as done in this thesis). Remote sensing data can effectively be used for any of these approaches, but have been applied with different frequencies – most often for mapping individuals, habitats or biotopes. The three approaches further aim at predicting species distributions across different scales (grain and extent, see Section 3.2.2) and thus require remote sensing data acquired by differing sensors.

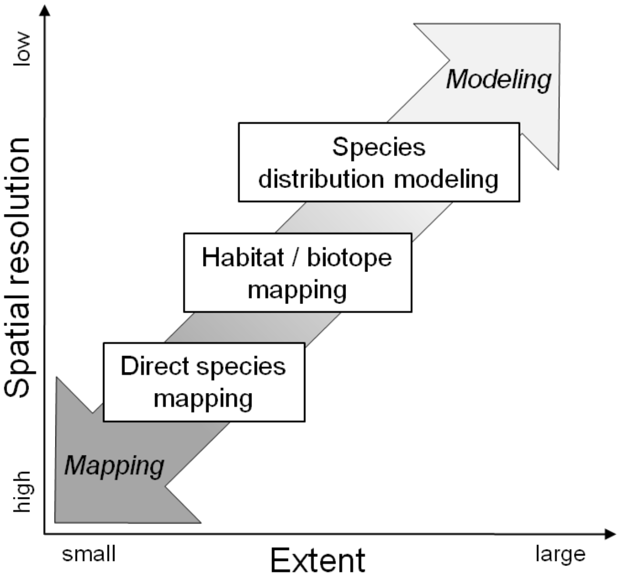


Figure 4.1: Conceptual framework of different approaches for predicting species distributions with remote sensing data. The transition from mapping to modeling is primarily a function of the spatial resolution and extent of the environmental data used but implies a change in underlying ecological theory.

Direct species mapping. Direct observations of individual plants are limited by the size of the study organism. Early approaches of direct species mapping therefore mainly used high-resolution aerial photography to identify individual plants, especially trees, to the species level (Gillespie et al., 2008). However, the ability to monitor plant species based on remotely sensed imagery is rare (Strand et al., 2007) as many species do not necessarily have unique spectral signatures (Franklin, 2009). Some invasive species show spectral profiles distinct from the native vegetation; they can directly be identified for example based on hyperspectral imagery (Hamada et al., 2007; Asner et al., 2008). Data typically analyzed for direct mapping plant species are high-resolution spaceborne IKONOS (Mills et al., 2006; Laba et al., 2010) or Quickbird (Wang et al., 2004; Everitt et al., 2006; Gómez-Garay et al., 2010) data as well as airborne hyperspectral (Wang et al., 2008) or LiDAR (Sellars and Jolls, 2007; Asner et al., 2008) data. The 'ideal' spatial resolution (pixel size) used for detecting species minimizes within-object (shade and sunlit leaves, bark, understory plants) variance and maximizes between-object (different individuals or species) variance (Meyer et al., 1996). Due to limitations in the resolution (spatial, spectral, and temporal) of the remote sensing data, the direct mapping approach is applicable only over smaller extents (typically several ha to few hundred km²). It will be generally supported by new sensor systems (Xie et al., 2008). Since species presence-absence data are typically point data and assigned to a certain grid cell of the remote sensing data, correct geo-location is especially important in these high-resolution mapping approaches.

Habitat/biotope mapping. The direct mapping of habitats (for animals) or biotopes (for plants) is one of the typical applications of remote sensing, mostly at landscape to regional scales. Hence, such mapping approaches typically represent only parts of entire species distribution ranges. Biotope or vegetation community distributions are often mapped based on medium-resolution Terra-ASTER (Nijland et al., 2009) or Landsat-TM/ETM+ (Thenkabail et al., 2003; Mehner et al., 2004; Yang, 2007) data. Mapping of vegetation types with coarser resolution Terra-MODIS data typically requires the use of higher resolution (e.g. Landsat-TM/ETM+) data to generate classification training data (Hüttich et al., 2009). In the biotope mapping approach, previously identified univariate relationships between species occurrences and vegetation communities (e.g. in field studies or from *a priori* knowledge) are transferred to indirectly deduce species presence or absence. However, this is only effective if there is a known one-to-one relationship between vegetation community type and species (Franklin, 2009). Additional errors may further be introduced through the biotope classification process (Strand et al., 2007). Moreover, biotopes are often not mapped for the purposes of the particular study. Instead, existing land cover or land use classifications derived from remote sensing data are interpreted with respect to the known ecological requirements of the target species. The drawbacks and caveats of using land cover data as a surrogate to draw conclusions on species distributions are further discussed in the context of predictive species distribution modeling in the following.

Species distribution modeling. To understand spatial patterns of species distributions, large proportions of or even entire geographical species distribution ranges are of interest. This means that large (continental to global) scales need to be examined. Together with the increasing extent of the study area, the grain of the remote sensing data recorded within a certain time period decreases and creates a trade-off between spatial resolution and repetition rate (Fisher and Mustard, 2007). As a side note, this relation can be overcome – though only at small scales – by satellite constellations (e.g. RapidEye). Predictive modeling of species distributions is therefore at the moment mainly applied based on coarse-resolution Terra-MODIS (Saatchi et al., 2008; Torres et al., 2010) or NOAA-AVHRR (Foody, 2005) data which cover large areas with high repetition rates. Combining species occurrence data collected in the field with such coarse-resolution remote sensing data implies a large difference in the spatial reference scales, just like for climatic species distribution models.

Modeling species distributions with remote sensing data is based on the assumption that “certain key environmental parameters, with remotely detectable biophysical properties, drive the distribution and abundance of species across landscapes” (Turner et al., 2003: p. 306). Remote sensing data may directly contribute to species distribution models by adding measured land surface characteristics beyond topographic and climatic conditions (Saatchi et al., 2008). In this sense, species distribution modeling with remote sensing data is different (1) from mapping species distributions since not the spectral signature of the species itself but rather its environment (i.e. vegetation community) are identified from the remotely sensed signal and (2) from mapping biotopes as certain algorithms that were initially developed for predicting climatic niches into geographical space (see Section 3.1.3) are applied. The aim of these algorithms is not to classify the landscape into certain categories as in the habitat/biotope mapping approach but to predict species occurrence as the single target variable. Remotely-sensed predictors also have their own data models, which need to be evaluated for compatibility with ecological models (Franklin, 1995; Austin, 2002).

Analyzing remote sensing data for large areas requires computational resources and specific skills and software. Therefore, many studies which aimed to integrate other factors than climate in species distribution models (Pearson and Dawson, 2003; Pearson et al., 2004) often made use of existing land cover or land use classifications that had previously been derived from remote sensing data. Many papers (Siriwardena et al., 2000; Maes et al., 2003; Eyre et al., 2004; Heikkinen et al., 2004; Luoto et al., 2006) have reported statistically significant relationships between land cover classification data and the regional or nationwide distribution and richness of species. However, the causal relationship between land cover and species distributions is often indirect (Thuiller et al., 2004). Especially for plants, land cover must be seen “more as a limiting factor (at least for human-oriented variables) than a factor having direct physiological impact“ (Thuiller et al., 2004: p. 359). However, land cover derived from remote sensing is still the most commonly used categorical predictor in species distribution modeling (Franklin, 2009). It has been proven to be valuable for predicting the distributions of both individual species and species assemblages (Kerr and Ostrovsky, 2003). For example, Pearson et al. (2004) integrated land cover data into species distribution models to uncouple effects of climate and habitat change in the interpretation and prediction of distribution changes of plants. Sánchez-Cordero et al. (2005) analyzed land cover data for limiting potential distributions based on ecological niche modeling to remnant ‘untransformed habitat’ for endemic mammals in Mexico.

The performance of land cover data for modeling species distributions depends on how closely species distributions can be linked to certain land cover types. The highest accuracies of land cover based models have generally been found for non-mobile species such as plants (Nagendra, 2001; Pearson et al., 2004). While ‘habitat’ is a species-specific concept (Hall et al., 1997), land cover – from the remote sensing perspective – is typically limited to the vegetation structure component (Gregorio and Jansen, 2005). Land cover classifications used for species distribution modeling must be of sufficient thematic detail (Kerr and Ostrovsky, 2003) and carefully designed to meet the objective of the study (Xie et al., 2008). The suitability of each land cover product to predict species distributions for a target species is hence based on the detail and the validity of its class definitions and legend. Recent comparative studies (Jung et al., 2006; Herold et al., 2008) highlighted the existence of major inconsistencies between different global land cover products, since they were embedded in different project structures, classification systems, and aims of research – especially in complex and heterogeneous landscapes and including large areas in Mexico. In addition, the spatial resolution of the land cover data is important (Thuiller et al., 2004; Luoto et al., 2007) as it affects classification accuracy (Foody, 2002). For example, land cover information did not improve the predictive power of species distribution models for several taxonomic groups (including plants) at a coarse 50 km resolution (Thuiller et al., 2004) but at finer 1 km and 10 km resolutions for plants (Pearson et al., 2004). In Section 6.3, the suitability of an existing Mexican land cover classification and continuous remote sensing time series for modeling biotope suitability are compared.

4.2 Plant-biotope relationships from a remote sensing perspective

The opportunity to capture biotope attributes relevant in species distribution mapping and modeling as described above is dependent on the characteristics of the remote sensing data (e.g. spatial and spectral resolution, acquisition date) but also of the target species itself and its relationship to the biotope as outlined in the following.

Vegetation strata. Optical remote sensing data is only capable of measuring the top canopy layer and thus provides little information on the lower strata (but see Tuanmu et al, 2010). It is therefore important that either the species itself (mapping approach) or the major factors which determine the distribution of the target species (modeling approach) can be identified from the remote sensing signal of the canopy. This may – for example – be true for canopy or epiphyte species but not necessarily for understory species. Dense or multi-layered vegetation is also susceptible to problems of canopy overlap, which may lead to confusion of different species and consequently over- or underestimation (Nagendra, 2001). The majority of studies on plant species diversity have therefore been carried out on tree species in the uppermost canopy or weeds in herbaceous vegetation (Nagendra, 2001).

Vegetation communities and floristic gradients. In remote sensing based studies, plants (as opposed to animals) are usually studied as assemblages or communities (Franklin, 1995). The theoretical foundations are the *continuum concept of vegetation science* (Gleason, 1926) and the results of *environmental gradient analysis* (Whittaker, 1956) as outlined in Section 3.1.2. These concepts assume that species exhibit individual responses to environmental gradients and that their probability of occurrence is dependent on how closely the environmental conditions at each site match the species' ecological optimum. From the remote sensing perspective, each pixel in medium to coarse resolution data (as the 1 km resolution MODIS data analyzed in this thesis) reflects the integrated response across the vegetation community including diverse species (Zhang et al., 2003). Remote sensing data typically characterize floristic gradients based on changing species composition and enclose – at least to a certain degree – biotic interactions. Other factors such as ground reflectance and understory vegetation components also contribute to the observed spectral signatures.

Floristic gradients may be either sharp (leading to discrete borderlines between assemblages, called *ecotone*) or smooth (leading to gradual transitions, called *ecocline*, Figure 4.2). While *ecotones* are typically (often unstable) zones of rapid change between two communities and most often created or maintained by human activities, *ecoclines* represent gradual changes corresponding to the spatial gradients of certain environmental factors (Kent et al., 1997). Remote sensing data can often serve as excellent indicators of *ecotones* related with human activities such as agricultural land use or deforestation (Figure 4.3). Particularly, remotely sensed variables allow for detecting biotope conversion or biotope loss which both are reliable and robust predictors of changes in species distributions.

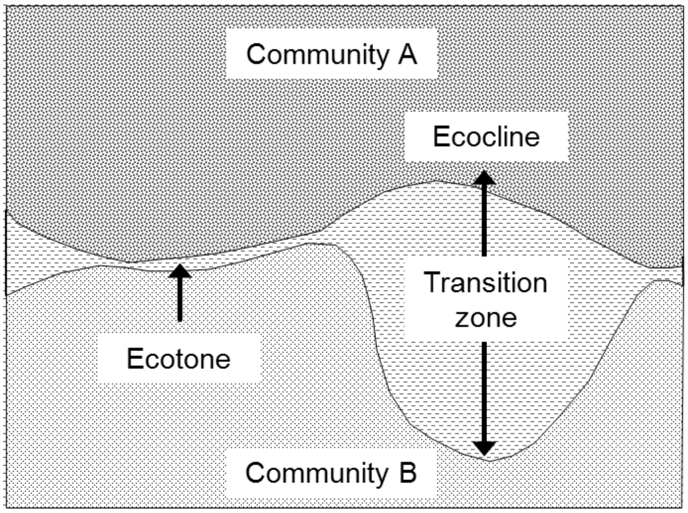


Figure 4.2: Linear representation of two species communities and their floristic gradients (Kent et al. (1997); modified). This gradient may be either sharp (*ecotone*) or gradual (*ecocline*).

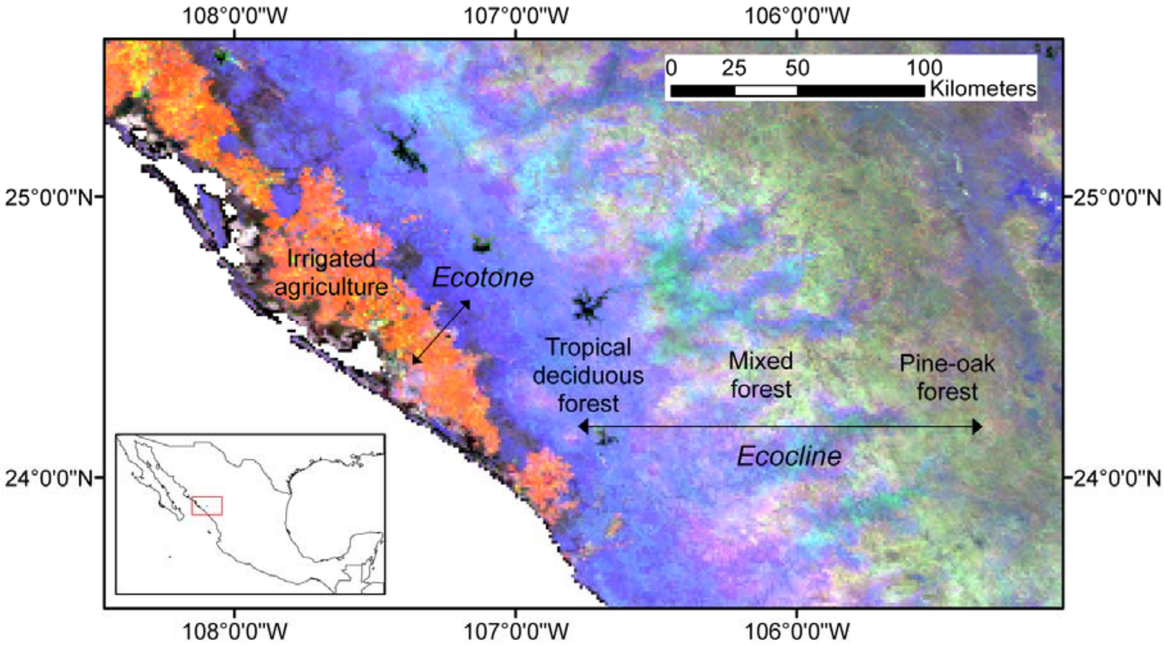


Figure 4.3: Example of a human-induced *ecotone* (agriculture – tropical deciduous forest) and an *ecocline* (tropical deciduous forest – mixed forest – pine-oak forest) as seen in remote sensing data. Underlying data is the MODIS-EVI time series of the year 2009 (R-G-B: DOY 65-145-273). Dark areas represent water bodies and the various colors indicate differences in vegetation conditions observed at the respective time periods. The red box in the inset map indicates the location of the main map.

Vegetation (mis)identification. Depending on the environmental completeness of the species records (Kadmon et al., 2003) – meaning that the relevant environmental gradients have adequately been sampled (Section 3.1.5) – and vegetation characteristics, modeling species distributions with remote sensing data can produce substantial over- or underestimation. The qualification of remote sensing data for modeling species distributions is therefore also subject to the environmental requirements of the species (examined in detail in Sections 6.3 and 6.4). Overestimation typically occurs either if the remotely sensed signal suggests the presence of the respective vegetation community but the target species is excluded due to competitive exclusion or limited dispersal capacities or if a different vegetation community features a similar remotely sensed signal. On the contrary, underestimation implicates that the remote sensing based model has misclassified the biotope or predicted its loss and therefore extirpation (local extinction) of the target species, even though the species is still present in reality (Strand et al., 2007). The likely reason of this error is that the definition of the biotope is too narrow (Strand et al., 2007). For plants, this may be the case for species locality data collected through *purposive* sampling (Edwards Jr. et al., 2006) in the core areas of distribution already known to botanical experts (so-called *sampling bias*, see Section 3.4.1).

Specialist or generalist? In addition to the above-mentioned potential misidentification of vegetation, also the environmental tolerance of the target species is of concern. For climatic species distribution models, it has been shown that specialist species with tightly defined ecological niches and a narrow range of tolerance can be predicted more precisely than generalist species (Brotons et al., 2004; Seoane et al., 2005; McPherson and Jetz, 2007; Evangelista et al., 2008; Syphard and Franklin, 2010). However, most remote sensing based studies so far ignored species' environmental tolerance – except Zimmermann et al. (2007) who found that remote sensing data especially improved distribution models for rare and early successional trees. In view of the findings of vegetation mapping studies (Thenkabail et al., 2003; Hüttich et al., 2009), biotope overestimation with remote sensing data is very likely for generalist species – which cannot unambiguously be assigned to a certain type of land surface cover with distinct spectral features. On the contrary, biotope requirements of specialist species, which have relatively tightly defined ecological niches and which have a narrow range of tolerance, can typically be better identified in the remotely sensed signal. However, if the species' requirements are too specific or only relevant at the smallest spatial scale, e.g. canopy gaps, these often cannot be recognized in coarse resolution remote sensing data. Environmental tolerance of the target species is hence especially important for remote sensing based models.

Species abundance. One of the major challenges in species distribution modeling from the remote sensing perspective is to deal with species occurrence data instead of parameters related to the predominant vegetation type or to vegetation structure. The training data may cover a wide range of species abundances from a single individual to dense mono-specific stands. The two exemplary plots (Figure 4.4) of the *National Forest Inventory* used as data source in this thesis (see Section 5.1) have completely differing plant cover percentages, vegetation height and age, vegetation structure, and understory composition. In a vegetation mapping campaign, the two sites would be classified into different categories in the field and – given that they exhibit spectrally different features – also in the remote sensing product. Here, as we are facing a 'single target class' (species presence) problem and thus a 'one-class classification' (Li and Guo, 2010) they are assigned to the same training data class, namely ‘presence’ for the target tree species *Liquidambar macrophylla*.

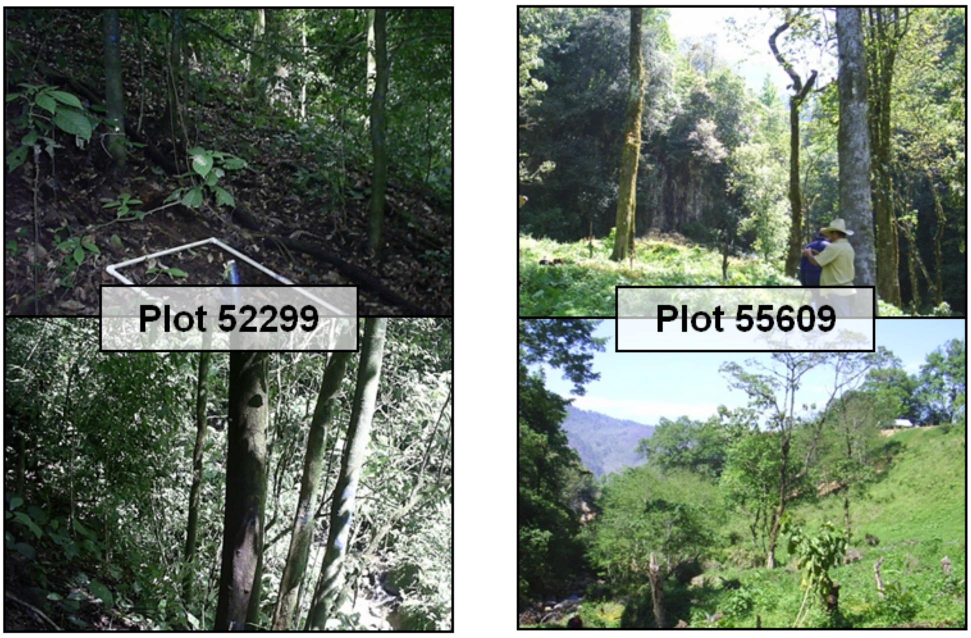


Figure 4.4: Reference pictures taken at two plots of the *National Forest Inventory* (INFyS) by CONAFOR. Both plots are presence sites for the tree *Liquidambar macrophylla*, one of the study species of this thesis, and located approximately 80 km distant from each other.

Stationarity. Last, the question remains whether species-biotope relationships are constant in space. This question again has been in detail discussed regarding climatic species distribution models (Pearman et al., 2007); both niche conservatism (Wiens and Graham, 2005) and niche shift have been observed (Broennimann et al., 2007; Rödder and Lötters, 2009); for remote sensing data stationarity has largely been ignored. While the opportunity to completely cover large areas with remote sensing data is often seen as a major advantage, the relationships between the variable of interest and the remotely sensed response are not necessarily stationary in space. The transferability between sites within the study region may thus be hampered (Gillespie et al., 2008).

4.3 Remaining gaps of knowledge

The use of various remote sensing data in relation to the assessment of species distributions has significantly increased in recent years – reflected in altogether 287 published articles (Figure 4.5). However, the majority of these studies mostly addressed biomass, yield or radiative transfer models and was hence not relevant for this thesis. Only 109 articles aimed at spatially predicting ecological patterns (species, communities, and biodiversity) by means of remote sensing data and are therefore considered hereafter. Still, these selected studies included many articles which proposed but not implemented the use of remote sensing in species distribution modeling.

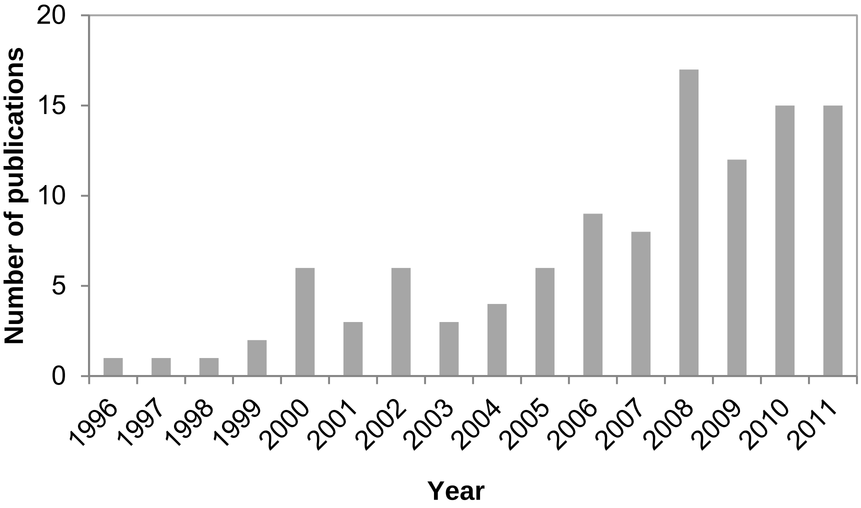


Figure 4.5: Results of a review based on an *ISI Web of knowledge* (http://www.webofknowledge.com) search of articles up to December, 2011 with the terms *remote sensing* and *species distribution/habitat/niche* in the title, abstract or keywords. Out of the altogether 287 studies returned by the search, only the 109 results relevant for the objectives of this thesis were considered.

In the following, every study that used – like this thesis – the *Maximum Entropy* (Maxent) algorithm for species distribution modeling (Section 3.3.2) in combination with remote sensing data is considered. As shown in Table 4.1, the number of studies applying the *Maximum Entropy* algorithm in combination with remote sensing data is sparse. Only five of these studies (Buermann et al., 2008; Prates-Clark et al., 2008; Saatchi et al., 2008; Evangelista et al., 2009; Tuanmu et al., 2010) addressed plant species distributions. However, the methodological findings of the remaining articles as well contribute to understanding the relevance of remote sensing data for modeling species distributions. The study of Li and Guo (2010) does not target species distributions, but was added here since it for the first time adopted the *Maximum Entropy* algorithm for classifying remote sensing data.

These 13 studies (Table 4.1) support the successful combination of remote sensing data with the Maxent algorithm from regional (Torres et al., 2010; Tuanmu et al., 2010) to continental scales (Buermann et al., 2008; Prates-Clark et al., 2008; Saatchi et al., 2008) comparable with the geographical extent of this thesis. Although the tendency towards overprediction of SDMs developed from only remote sensing data has been shown (Peterson et al., 2008), all studies carried out at local and regional scales (Evangelista et al., 2009; Li and Guo, 2010; Torres et al., 2010; Tuanmu et al., 2010; Pittman and Brown, 2011; Tuanmu et al., 2011) did not incorporate climatic information. In case both climate and remote sensing data were used as predictors (Buermann et al., 2008; Prates-Clark et al., 2008; Boubli and de Lima, 2009; Foley et al., 2009), the two different data sources were mostly combined into a single model ignorant of their different “sample density” (Franklin, 2009: p. 57). These studies consistently found that the use of remote sensing data improved the performance of climatic models.

Table 4.1: Overview of the use of the *Maximum Entropy* algorithm together with remote sensing data for modeling species distributions (until December, 2011; grouped according to study purpose and spatial scale). [a] WDRVI: *Wide Dynamic Range Vegetation Index*

Reference	***Journal***	***Study species***	***Study area***	***Remote sensing data***	***Combination with climate data***
Modeling tree species and richness (continental scale)					
(Buermann et al., 2008)	*Journal of Biogeography*	4 trees, 2 birds, and 2 mammals	South America	MODIS-LAI (2000 - 2004), MODIS-VCF (2001), QSCAT backscatter (2001), SRTM-DEM (2000)	Integration into same model
(Prates-Clark et al., 2008)	*Ecological Modelling*	3 trees	South America	MODIS-NDVI (2000 - 2001), MODIS-LAI (2000 - 2001), MODIS-VCF (2001), QSCAT backscatter (1999 - 2004), SRTM-DEM (2000)	Integration into same model
(Saatchi et al., 2008)	*Remote Sensing of Environment*	5 trees and tree richness categories	South America	MODIS-LAI (2000 - 2004), MODIS-VCF (2001), QSCAT backscatter (2000 - 2004), SRTM-DEM (2000), TRMM (1998 - 2006)	No combination
Predictive mapping of plant species (regional scale)					
(Evangelista et al., 2009)	*Remote Sensing*	1 invasive plant	Southwestern Colorado (USA)	Landsat ETM+: Reflectance, NDVI, SAVI (6 scenes between 1999 and 2003)	No climate data used
(Tuanmu et al., 2010)	*Remote Sensing of Environment*	2 bamboo species	Wolong Nature Reserve (China)	MODIS-WDRVI[a] (2001 - 2003)	No climate data used
Habitat modeling of animals (sub-continental scale)					
(Boubli and de Lima, 2009)	*International Journal of Primatology*	4 primate species	Amazonia	MODIS- EVI/NDVI (2007, single composite)	Integration into same model
(Cord and Rödder, 2011)	*Ecological Applications*	9 anurans	Southern USA, Mexico, Central America	MODIS-EVI (2001 - 2007) MODIS-LST (2001 - 2008)	Integration into same model and combination of model predictions

Habitat modeling of animals (regional scale)					
(Torres et al., 2010)	*Biological Conservation*	Chimpanzee	South-western Guinea-Bissau (West Africa)	Landsat TM: NDVI (1986,1994 and 2003; 4 scenes per year)	No climate data used
(Peck et al., 2011)	*International Journal of Primatology*	Spider monkey	Northern Ecuador	Landsat TM/ETM+: Reflectance to obtain forest classification (1988, 2001, 2006; 3 scenes)	Climate data only used for potential historical distribution
(Tuanmu et al., 2011)	*Journal of Biogeography*	Giant panda	Wolong Nature Reserve (China)	MODIS-WDRVI[a] (2001 - 2003, 2005 - 2007)	No climate data used
Modeling aquatic environments (regional scale)					
(Pittman and Brown, 2011)	*PloS One*	5 fish species	South-western Puerto Rico	LiDAR: Laser airborne depth sounder (LADS), 2006	No climate data used
Modeling the distribution of disease vectors (national scale)					
(Foley et al., 2009)	*Journal of Medical Entomology*	8 mosquitos	South Korea	AVHRR-NDVI (1992 - 1993)	Integration into same model
One-class classification of remote sensing imagery (local scale)					
(Li and Guo, 2010)	*International Journal of Remote Sensing*	Urban, tree, class, soil	City of El Cerrito, California (USA)	Aerial photography: Leica ADS40 (single data set)	No climate data used

The existing literature was analyzed regarding the specific remote sensing products used to identify the most useful set of variables for this thesis. Species distributions are determined not only by absolute levels of climate or vegetation variables, but also by subtle differences in their seasonality (Rogers et al., 2002) and may change with time (Torres et al., 2010; Peck et al., 2011). Except Boubli and de Lima (2009), Li and Guo (2010), and Pittman and Brown (2011) all studies summarized in Table 4.1 thus made use of **multi-temporal remote sensing data** (mostly time series of between two and four years). In this context, the studies of Li and Guo (2010) and Pittman and Brown (2011), that did not analyze multi-temporal data, are special cases with little direct methodological relevance for this thesis as they used not operationally available aerial photography and LiDAR data, respectively.

According to the literature review, large focus was given to **vegetation indices** (NDVI, EVI, SAVI, WDRVI, see Section 4.4.2), mostly obtained from the MODIS sensor (Prates-Clark et al., 2008; Boubli and de Lima, 2009; Tuanmu et al., 2010; Cord and Rödder, 2011; Tuanmu et al., 2011) or Landsat data (Evangelista et al., 2009; Torres et al., 2010). In addition, reflectance data (Evangelista et al., 2009; Torres et al., 2010; Peck et al., 2011) but also *Leaf Area Index* (LAI) together with radar backscatter information (Buermann et al., 2008; Prates-Clark et al., 2008; Saatchi et al., 2008) were widely utilized as environmental variables. However, the latter three studies largely overlap regarding their study region, the remote sensing data used and the target species modeled. Predictor selection in many of these studies often appears arbitrarily and – in case the distributions of several species were modeled – was never attuned to the ecological traits of each species separately.

Time series of the vegetation index were successfully analyzed for characterizing plant phenology (Hermance et al., 2007), photosynthetic vegetation activity (King et al., 2004), and primary productivity (Xiao et al., 2005). All five studies modeling plant species distributions (Buermann et al., 2008; Prates-Clark et al., 2008; Saatchi et al., 2008; Evangelista et al., 2009; Tuanmu et al., 2010) took advantage of the potential of **multi-temporal vegetation index data.** The studies hence successfully predicted plant species distributions based on vegetation characteristics from multiple observation periods. For example, Evangelista et al. (2009) found that model performance increased due to the inclusion of multi-temporal data as time series can better distinguish differences between invasive tamarisk (*Tamarix* spp.) and the native flora than a single-scene analysis. Previous studies further showed that **phenological metrics** derived from time series (instead of the ‘raw’ time series data) are useful for classifying land cover (DeFries et al., 1995; Hansen et al., 2000) and differentiating forest types (Townsend and Walsh, 2001). Compliant with this, most articles included in this review computed metrics such as annual or seasonal maximum or mean values (Prates-Clark et al., 2008) or annual standard deviation and range (Buermann et al., 2008; Saatchi et al., 2008) as input variables for their species distribution models. Tuanmu et al. (2010, 2011) used the so far most comprehensive set of altogether eleven phenological metrics. They were able to show that species distribution models developed with phenological metrics instead of the original vegetation index time series exhibited higher temporal transferability presumably – as Tuanmu et al. (2011) concluded – due to the reduction of model complexity and multi-collinearity among variables in phenological metrics. In this thesis, a data set consisting of 18 annual phenological metrics was compiled, based on a review of phenological metrics derived from time series of the vegetation index by previous studies (see Section 5.4.4, Appendix G).

To minimize the effects of any natural **inter-annual variability** known from vegetation mapping studies (Asner et al., 2000; Colditz et al., 2009; Brando et al., 2010), phenological metrics were typically averaged over several years of the study period in the existing studies (Buermann et al., 2008; Prates-Clark et al., 2008; Saatchi et al., 2008; Tuanmu et al., 2010; Cord and Rödder, 2011). Very recently, Tuanmu et al. (2011) published their results regarding the temporal transferability of species distribution models depending on the composition period. They compared and tested four species distribution models developed from remote sensing time series of either one or of three consecutive years and found that models based on variables derived from multi-year data had higher accuracies and higher transferability than those with the variables from single-year data.

To summarize, the literature review pointed out the following:

(1) Remote sensing predictors can **improve climatic species distribution models** at national to continental scales and are sufficient in themselves to predict species occurrence from local to regional scales.

(2) Especially **multi-temporal vegetation index data** (typically transformed into annual phenological metrics) are suitable for modeling plant species distributions.

(3) Natural **inter-annual variability** in remotely sensed time series may impact temporal model transferability and hence the validity of species distribution models.

What is obviously missing from the existing literature are a sensitivity analysis and a systematic framework of rules and standards for modeling species distributions by means of remote sensing data. The existing methodological recommendations for modeling species distributions found in the literature are taken up in this thesis and extended regarding some critical questions that had partly already been brought up by the most recent studies (e.g. Tuanmu et al., 2011). Particularly, the comprehensive analysis of several remote sensing products to cover a wider range of biotope characteristics, the computation of phenological and statistical metrics from remotely sensed time series data, and the examination of the effects of inter-annual variability are addressed in this thesis. The detailed methodology applied and the results obtained are described in Chapter 6. The remote sensing products analyzed in this thesis were selected based on the findings of the above-mentioned studies and based on their suitability to describe environmental gradients and ecosystem functioning as described in the following section.

4.4 Suitable remote sensing data and products

Some authors (Huete et al., 2002; Fensholt and Sandholt, 2003; John et al., 2008) have already questioned the use of vegetation index data alone for understanding ecosystem functioning. In this thesis, different remote sensing products were specifically selected (1) because they had successfully been utilized in previous species distribution modeling studies (Table 4.1) and (2) based on the suitability of the products to characterize species distributions. In the following, the information content of the remote sensing data particularly for modeling species distributions as well as their relevance for characterizing land surface characteristics are described. The different data sets are presented in the order of their processing levels and include *Reflectance* (Section 4.4.1), *Vegetation Indices* (Section 4.4.2), and *Land Surface Temperature* (Section 4.4.4). Section 4.4.3 concentrates on the new dimension for quantifying ecosystem dynamics opened up by the analysis of multi-temporal *Vegetation Index* data. The interplay between *Vegetation Indices* and *Land Surface Temperature* is outlined in Section 4.4.5.

4.4.1 Reflectance data

Plant species dominant in the vegetation canopy (Debinski et al., 1999) and their biochemical and structural properties (Kumar et al., 2001) play a major role in defining spectral reflectance patterns recorded by remote sensors. Reflectance and transmittance of leaves are determined by leaf pigments, e.g. through chlorophyll absorption (visible wavelengths, particularly blue and red), by internal scattering in the spongy mesophyll (near infrared, NIR), and by leaf water content (middle infrared, MIR) (Jensen, 2007) as illustrated in Figure 4.6. In general, canopies of broad leaved forests are spectrally brighter than those with needle leaves, especially at NIR wavelengths, because of their canopy geometry and their leaf angle distributions (Hall et al., 1992). NIR data have also been used to discriminate species which differ in their foliage content (Taylor, 1993). Since MIR bands typically respond to leaf water content, they may also be used for differentiating succulent and non-succulent plants (Everitt et al., 1987). The spectral profiles are further influenced by soil background reflectance in terrestrial ecosystems, and possibly water background reflectance in wetlands or temporarily inundated areas.

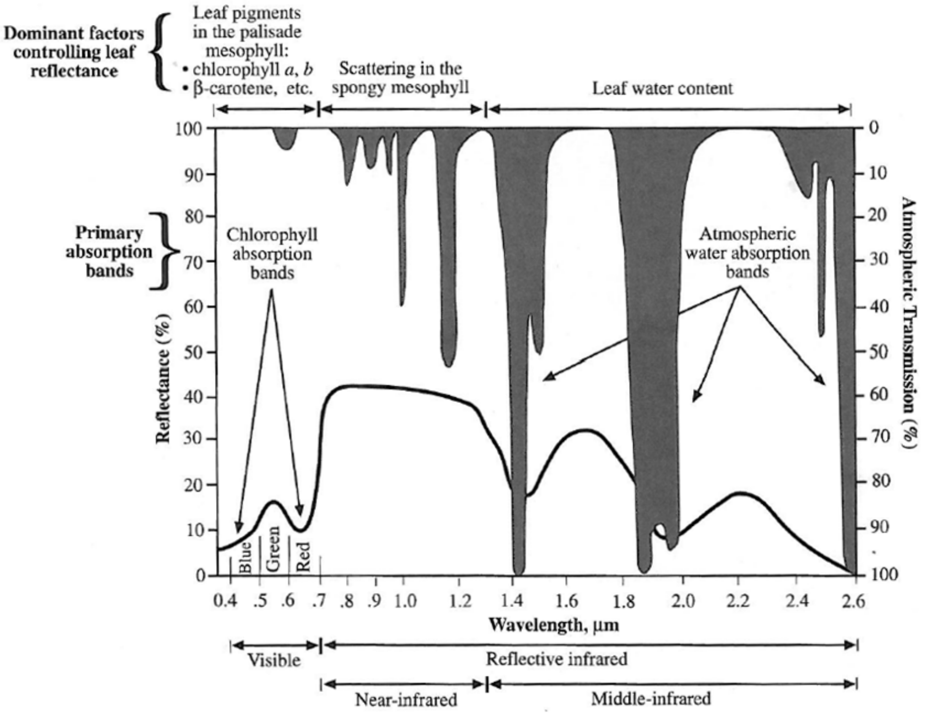


Figure 4.6: Spectral reflectance characteristics of green, healthy vegetation for the wavelength interval 0.4-2.6 μm (Jensen, 2007). The dominant factors controlling leaf reflectance are the leaf pigments in the palisade mesophyll (chlorophyll a and b; β-carotene), the scattering of the near infrared in the spongy mesophyll, and the water content of the plant affecting middle infrared reflectance.

4.4.2 Vegetation Indices

The above-mentioned relationship of red and NIR reflectance – based on chlorophyll absorbance and cellular reflectance (Tucker and Sellers, 1986; Jensen, 2007) – to the amount of vegetation present on the ground is well-known since the early beginnings of terrestrial remote sensing (Colwell, 1974). Accordingly, a variety of vegetation indices, for example the *Normalized Difference Vegetation Index* NDVI, the *Soil-Adjusted Vegetation Index* SAVI (Huete, 1988) or the *Enhanced Vegetation Index* EVI (Huete et al., 2002) have been developed using these wavelengths (for a full list see Jensen, 2007). Remotely sensed vegetation index products provide spatially and temporally consistent assessments of global vegetation conditions and ‘greenness’ that can be used to monitor terrestrial photosynthetic vegetation activity (King et al., 2004) and other land surface processes (Mora and Iverson, 1998). Standard vegetation index products are available from NOAA-AVHRR (since 1981), SPOT-VEGETATION (1998), and Terra-MODIS (1999) data.

Satellite vegetation index data have been shown to be correlated with *Leaf Area Index* (LAI) and the *Fraction of absorbed Photosynthetic Active Radiation* (FPAR) (Baret and Guyot, 1991). The NDVI also has a strong relationship with actual evapotranspiration (Box et al., 1989; Cihlar et al., 1991) and *Net Primary Productivity* (NPP) that is consistent across most ecosystems (Running and Nemani, 1988; Box et al., 1989; Running, 1990). This link between NDVI and NPP suggests the potential for characterizing water-energy dynamics related to plant species richness patterns (Fairbanks and McGwire, 2004). However, the distinction between forest types with different dominant species is hardly possible as multiple assemblages of plant species may produce the same vegetation index scores (Pettorelli et al., 2005). MODIS-EVI data has been shown to be sensitive to the difference between needle leaved and broad leaved canopy structures, with EVI scores over needle leaved forests being approximately one-half over those of broadleaf forests, though (Huete et al., 2002). In addition, there seems to be little correlation between vegetation indices and vegetation structure across different biomes (Box et al., 1989).

Besides species composition, several other factors affect vegetation index scores within a pixel such as plant structure and height, vegetation vigor, plant stress, and interactions within the canopy cover (Pettorelli et al., 2005). Different vegetation index values may thus be observed for the same vegetation communities under differing environmental conditions. To summarize, there is no direct relationship between vegetation indices and plant species presence or absence. However, as vegetation indices are typically related with certain biophysical variables such as LAI and NPP, they may be used to describe biotope characteristics, e.g. as predictors in species distribution modeling. The information content of multi-temporal vegetation index data for predicting species distributions is further discussed in the following section.

4.4.3 Multi-temporal Vegetation Index data

Based on the information content of mono-temporal vegetation index data outlined above, the analysis of multi-temporal (‘time series’) vegetation index data opens up a new dimension for quantifying ecosystem dynamics. At regional and larger scales, variations in community composition, micro- and regional climate regimes, and land management result in complex spatio-temporal variation in plant phenology (Zhang et al., 2003). Multi-temporal imagery thus allows for the extraction of these phenological and seasonal variations in vegetation cover and structure (Geerken et al., 2005) that affect and mirror geographical species distributions. However, a clear distinction needs to be made between *in situ* vegetation phenology and remotely-sensed *Land Surface Phenology*. Both parameters refer to different scales and imply a different view on and understanding of phenology.

***In situ* plant phenology.** Early on, a broad definition of phenology was provided (Lieth, 1974) which explained phenology as “the study of the timing of recurring biological events, the causes of their timing with regard to biotic and abiotic forces, and the interrelation among phases of the same or different species” (Lieth, 1974: p. 4). Despite this general definition (which is not restricted to plants), the topic has often been associated with agricultural events such as planting, emergence, fruiting, and harvest (Reed et al., 1994). Plant phenology is often studied in a hierarchical structure from plant organs (first leaf or flower), individual plants (% leaf expansion, % flowering), community (% of green vegetation) to landscape (Reed et al., 1994). In this sense, phenology is a key feature of plant the species’ niche because it determines the ability of the species to capture variable resources and it defines the seasonality and the duration of plant growth and reproduction (Schwartz, 2003; Chuine, 2010). In particular, it has been shown that (for the northern hemisphere) the northern limits of plant species ranges are often caused by the inability to undergo full fruit maturation, while the southern limits are determined by the failure of the plants to flower or to unfold leaves without the chilling temperatures required to break bud dormancy (Chuine and Beaubien, 2001; Chuine, 2010). The factors influencing phenology vary by species, but include photoperiod, soil moisture, soil temperature, air temperature, humidity, and solar illumination (Reed et al., 1994). Thus, phenology is very sensitive to changes in weather and climate. Even for a single plant species in a confined area, phenotypic plasticity and local adaptations thus create a wide range of phenological responses (Fisher and Mustard, 2007).

The satellite-based analysis of phenology is fundamentally different from these traditional ground-based observations (Reed et al., 1994). Due to the lack of comparable phenological metrics and the difficulties of ‘field-to-satellite scaling’ (e.g. positional errors, effects of other vegetation, and soil background characteristics), there has been little success in binding together field and satellite-based phenology research so far (Fisher et al., 2006). Especially in highly variable landscapes, comparing ground observations and satellite measurements may thus be difficult (Fisher and Mustard, 2007).

Satellite remote sensing of *Land Surface Phenology.* *Land Surface Phenology* (LSP) has been defined as “the study of vegetation phenology using remote sensing” (Reed et al., 2009: p. 232) or “the study of the spatio-temporal development of the vegetated land surface as revealed by synoptic spaceborn sensors“ (White et al., 2009: p. 2336). Dynamic phenological behavior has been observed based on remote sensing data also for ecosystems previously assumed ‘stable’ such as mature evergreen tropical forests (Xiao et al., 2005; Hermance et al., 2007). While *in situ* phenologies as outlined above are usually recorded from discrete events such as bud break, leaf emergence, or first flower (Reed and Xiao, 2009), current technology limits satellite observations, due to their coarse spatial resolutions, to the seasonality of the canopy cover (Fisher and Mustard, 2007) as indicator of ecosystem dynamics (Reed et al., 1994). The observed broad-scale changes may thus not directly be associated with phenological events of specific plants, but are descriptive of plant communities (Reed et al., 1994; Fisher and Mustard, 2007). Particularly, *Land Surface Phenology* data has been found to be sensitive to climate changes (Zhang et al., 2004; Morisette et al., 2009) and changes in vegetation due to natural (Eklundh et al., 2009) and human (Koltunov et al., 2009) disturbances.

Currently, no generally accepted definition of phenological metrics derived from satellite observations exists (White et al., 2009). However, different sets of phenological metrics (see Appendix G) – including for example *start of season* (SOS) or *end of season* (EOS) – have been developed and derived from multi-temporal vegetation index data (Reed et al., 1994; Zhang et al., 2003; Jönsson and Eklundh, 2004; White et al., 2009). Phenological metrics have been used to classify land cover (DeFries et al., 1995; Hansen et al., 2000), differentiate forest types (Townsend and Walsh, 2001), and to model species distributions (Tuanmu et al., 2010; Tuanmu et al., 2011). Tuanmu et al. (2011) showed that species distribution models developed from phenological metrics instead of vegetation index time series exhibited higher temporal transferability. The phenological metrics used in this thesis are described in Section 5.4.4 and Appendix H.

4.4.4 Land Surface Temperature

Temperature (along with water and available sunlight) belongs to the major influences on global vegetation distribution. On average, satellite-detected *Land Surface Temperature* (LST) follows patterns of insolation (Fisher et al., 2006), but yearly averages for air temperature are generally lower than LST (Julien et al., 2006). LST is defined as the “skin temperature of the ground” (Qin and Karnieli, 1999: p. 2368); it is hence a measurement of ‘how warm the land surface is to touch’ and different from the air temperature as land in general heats and cools more quickly than air (Qin and Karnieli, 1999).

Thermal radiation emitted from the earth has very low intensities (long wavelengths) and thus a low signal-to-noise ratio. Therefore, remotely sensed LST data typically has lower spatial resolution than reflectance data (e.g. in the visible or NIR bands). For bare soil, LST is equivalent to soil surface temperature while for dense vegetation it is determined by the temperature of the canopy. Over sparsely vegetated ground, LST is the composite average temperature of vegetation canopy, vegetation body, and the soil surface (Qin and Karnieli, 1999; Weng et al., 2004). Satellite-derived LST measures are controlled by the fractions of soil and vegetation in each pixel, surface wetness, and synoptic weather conditions (Nemani et al., 1993; Nemani and Running, 1997). They hence represent valuable information on biotope quality. Across different vegetation types, surface temperatures are – under dry surface conditions – linearly related to canopy densities (Nemani et al., 1993). Under wet conditions, the relationship between LST and vegetation cover is poor, but the maximum value compositing procedures applied to multi-temporal remote sensing data (Section 5.4.2) tend to favor dry surface conditions (Nemani and Running, 1997). The thermal responses of vegetation to air temperature changes depend on the biophysical properties of the vegetation itself (Quattrochi and Ridd, 1998). Similarly, the thermal properties of soils vary with soil type and its moisture content (Sandholt et al., 2002). LST is therefore very useful for modeling energy balance components and surface-atmosphere interactions (Parida et al., 2008).

Both human and natural disturbances often lead to changes in the ecosystem energy budget and reflected in modifications of LST characteristics (Quattrochi and Luvall, 1999; Kerr and Ostrovsky, 2003). Hence, LST data has been used for 'urban heat island' detection (Oke, 1982; Weng et al., 2004; Chen et al., 2006; Schwarz et al., 2011) or classification of agricultural land use (Parida et al., 2008). Despite this significant potential of LST data as a measure of ecosystem functioning (Quattrochi and Luvall, 1999) it is by far less frequently used than vegetation indices by ecologists (Kerr and Ostrovsky, 2003). For modeling plant species distributions, LST is a valuable parameter as it integrates vegetation presence and vegetation conditions with surface wetness. A recent study (Albright et al., 2011) also found that MODIS LST data were better predictors of changes in avian community structure due to heat waves than interpolated air temperature data. As with reflectance data and vegetation indices, LST is in vegetated areas always a measure of the upper canopy plants.

4.4.5 Relationship between Vegetation Index and Land Surface Temperature

Satellite-derived LST measures are, among other things, controlled by the fractions of vegetation coverage. Accordingly, a combined analysis of vegetation indices and LST facilitates the assessment of land surface processes and characteristics. The relation between vegetation indices and LST has been analyzed in two different ways in the literature – either in *space* related to spatial variation with the purpose to determine land surface parameters or in *time* with the aim to characterize vegetation changes based on temporal variations. Since the presence of vegetation moderates surface temperatures as compared to air temperatures, vegetation coverage determines the maximum measurable LST scores (Goward and Hope, 1989). This negative correlation between LST and vegetation indices has been observed for several land cover types (Goward et al., 2002; Kaufmann et al., 2003; Weng et al., 2004). Strength, sign, and the slope of their relation contain information of surface processes (Lambin and Ehrlich, 1996). However, the two variables are not redundant and respond to different biophysical processes (Lambin and Ehrlich, 1996). Changes in the seasonal trajectories of vegetation index and LST have therefore been used to map land cover (Weng et al., 2004; Yue and Tan, 2007) and detect land cover changes (Lambin and Ehrlich, 1997). Nemani and Running (1997) analyzed the link between LST and vegetation index data to classify land cover of the contiguous United States into four groups: *water limited* (shrubs, grass), *energy limited* (wetlands, boreal forests, snow, ice, water), *atmospherically decoupled* (aerodynamically smooth canopies, crops), and *atmospherically coupled* (aerodynamically rough canopies, forests). The combination of MODIS-LST and MODIS-NDVI has also been used to monitor drought in the southern Great Plains (USA) using the *Vegetation Temperature Condition Index* VTCI (Wan et al.,

2004) and in Senegal based on the *Temperature Vegetation Dryness Index* TVDI (Sandholt et al., 2002). Particularly the synergistic analysis of vegetation index and LST data hence improves vegetation monitoring and characterization (Nemani and Running, 1997).

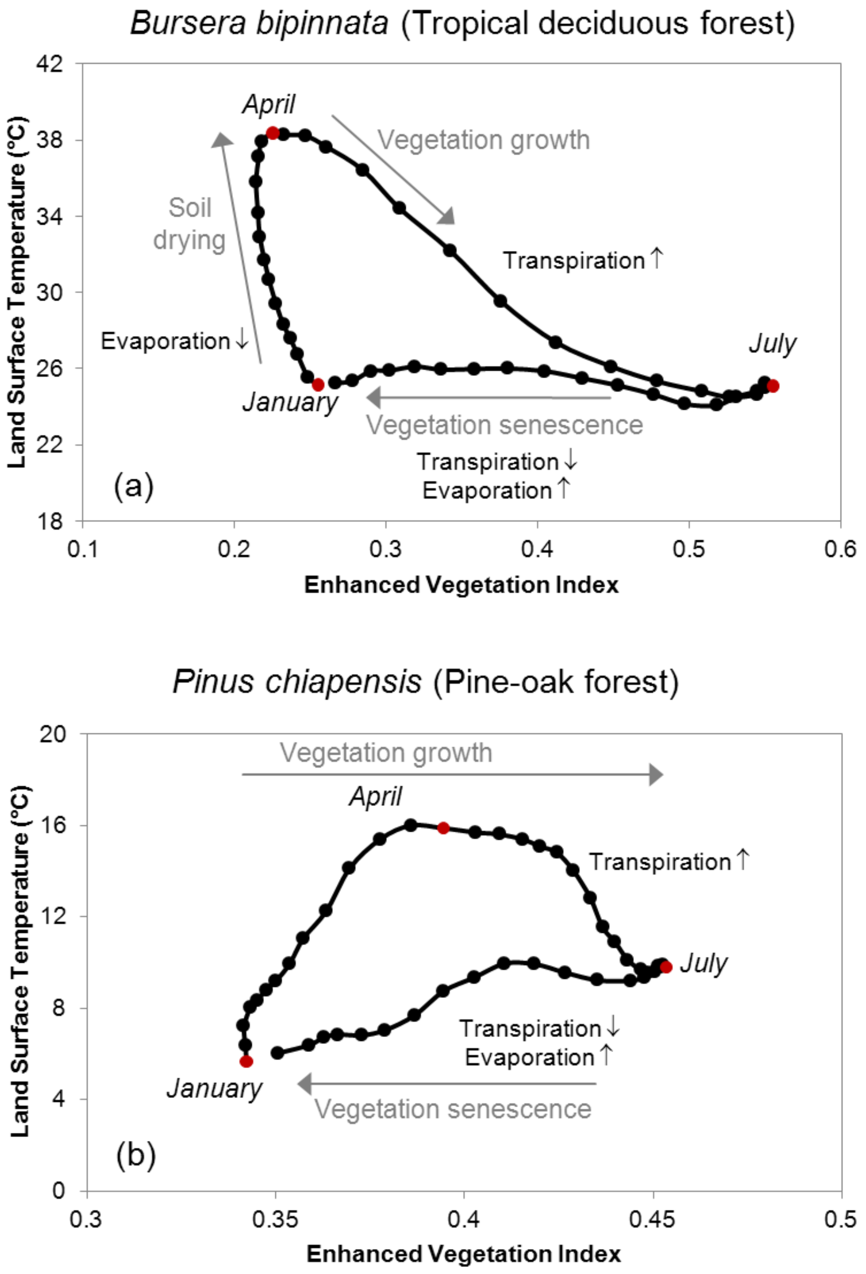


Figure 4.7: Seasonal trajectory of *Enhanced Vegetation Index* (EVI) and *Land Surface Temperature* (LST) averaged for presence sites of the tree species (a) *Bursera bipinnata* and (b) *Pinus chiapensis*. The vertices of the triangle represent (1) low air temperatures and low vegetation cover in January, (2) high air temperatures at the beginning of the rainy season (mid of April), and (3) peak of the rainy season in July/August. Note: The same MODIS composites are highlighted red in both figures. Both diagrams have different scales for EVI and LST.

The seasonal trajectories in the EVI/LST space for two study species are shown in Figure 4.7. Such trajectories typically show a triangular form (Sobrino and Raissouni, 2000). The vertices of this triangle coincide with the periods of maximum EVI (here July), minimum LST (January), and maximum LST (April). *Bursera bipinnata* (Figure 4.7 a) is a typical species of the tropical deciduous forest in the lowlands of western and southern Mexico with a pronounced phenological cycle including vegetation green-up and senescence. The seasonal trajectory shows the associated processes of transpiration and evaporation and their effects on the *Land Surface Temperature*. For *Pinus chiapensis* (Figure 4.7 b) occurring in tropical montane pine-oak forests in *Chiapas* (southeastern Mexico), the yearly trajectory amplitude of the EVI/LST space is by far lower and the vertices of the triangle less pronounced. In addition, the annual EVI range is significantly smaller due to the high proportion of evergreen vegetation in this forest type. As a result of the occurrence of *P. chiapensis* in montane areas (mainly between 1,200 m and 1,800 m) with lower air temperatures, LST scores are also considerably lower than *for B. bipinnata*, occurring in lowlands. Hence, the temporal profiles of LST and EVI do not only differ between different land cover classes (Weng et al., 2004) but also between different vegetation types. The *Maximum Entropy* algorithm (Section 3.3.2) applied for species distribution modeling in this thesis is able to account for variable interactions and is thus capable of integrating both vegetation index and LST scores.

4.5 Remote sensing data for predicting species richness

In the previous sections of this chapter, the theoretical background and conceptual framework for modeling spatial distributions of individual species were outlined. Particularly, the remote sensing products selected were characterized regarding their information content for modeling species distribution. The same remotely sensed variables are certainly also relevant for predicting species richness. The following section provides the theoretical background for predicting patterns of pine richness across Mexico by means of remote sensing data (Chapter 7).

Species richness – that is the number of taxa occurring in a defined geographic unit – is widely used to measure biodiversity (Whittaker, 1972). Assessing and mapping species richness has become technically feasible also at larger scales based on remotely sensed land surface properties (John et al., 2008). Two different approaches for modeling species richness exist (Ferrier and Guisan, 2006; Guisan and Rahbek, 2011): (1) **direct 'macroecological' modeling** (Guisan and Rahbek, 2011) of species numbers per geographic unit by statistically relating species richness to values of environmental variables and (2) the **stacking of individual species distribution models** (Guisan and Rahbek, 2011) to 'range-overlap maps' (Sorte and Hawkins, 2007) as illustrated in Figure 4.8. However, the latter – with the exception of this thesis – has so far not been implemented in combination with remote sensing data for large spatial extents.

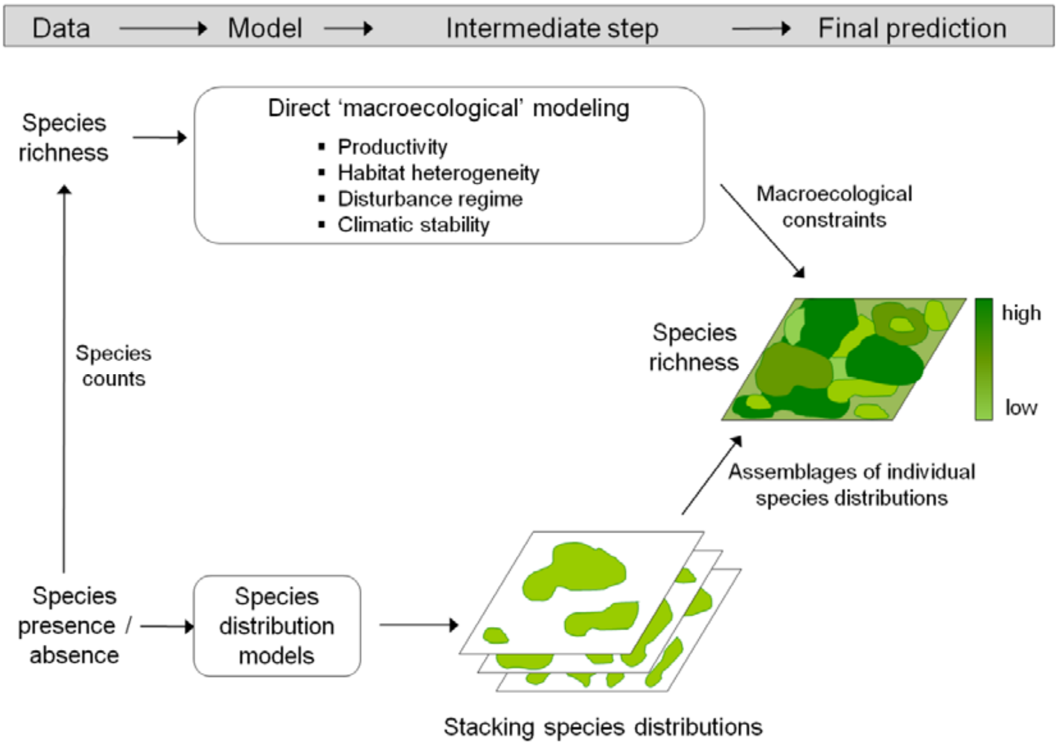


Figure 4.8: Illustration of the two theoretical approaches (direct macroecological modeling and stacking of individual species distribution models) for predicting species richness (Guisan and Rahbek, 2011; modified).

Direct 'macroecological' modeling of species richness. This approach directly predicts species richness based on either theoretical assumptions or hypothesized environmental factors that control the maximum number of species in a certain area (Guisan and Rahbek, 2011). Limiting factors of species richness which can be derived from remote sensing data are typically measures of ecosystem productivity and heterogeneity, but may also be related to the disturbance regime. Further, categorical land cover or vegetation type data derived from remote sensing have been used to indirectly extract information on plant species richness (Gould, 2000; Griffiths et al., 2000). Most direct modeling studies which involve remote sensing data conceptually refer to either (1) the *Spectral Variation Hypothesis* (Gould, 2000; Palmer et al., 2002) which predicates that within a sampling unit of given size, it is possible to find a relation between the variation of some spectral signal and the species richness or (2) the *Productivity Hypothesis* (Connell and Orias, 1964; Waide et al., 1999) assuming that species become more specialized when resources are abundant (which in turn leads to higher species numbers per unit area).

The *Spectral Variation Hypothesis* assumes that heterogeneity in spectral indices or land cover within a certain area is an indicator of habitat heterogeneity which allows more species to coexist (Gillespie et al., 2008). It has been applied to remote sensing data for predicting plant species richness (Gould, 2000; Rocchini, 2007; Feilhauer and Schmidtlein, 2009; Oldeland et al., 2009). Rocchini (2007) in detail investigated the effects of spatial scales concerning the relation between spectral heterogeneity and plant species richness by using satellite data with different spatial and spectral resolutions.

In the context of the *Productivity Hypothesis*, mainly data of the NDVI associated with NPP have been analyzed (Gillespie et al., 2008). Significant positive relationships between plant species richness and NDVI have been reported for boreal (Parviainen et al., 2009), temperate (Levin et al., 2007) and tropical (Gillespie, 2005; Cayuela et al., 2006) ecosystems. For the MODIS data used in this thesis, a polynomial relation between tree species richness and the EVI as a surrogate for ecosystem productivity was found at the ecoregion level in the contiguous United States (Waring et al., 2006). Similarly, John et al. (2008) predicted plant richness at the county level using linear regression with remotely sensed biophysical estimates of productivity and water content derived from MODIS data in semi-arid Mongolia. Empirical evidence shows that the direction and magnitude of the diversity-productivity relationship differ among ecosystems, taxonomic groups, and scales (Waide et al., 1999; Chalkraft et al., 2004).

Stacked species distribution models. This second approach was more recently introduced than the direct modeling approach. It has its origins in the attempts to forecast biodiversity changes due to climate and land cover changes (Guisan and Theurillat, 2000) as most species show individual responses to environmental change (Huntley, 1991). The approach focuses on aggregate properties of individual species (Guisan and Rahbek, 2011) and hence does not set a limit on the maximum number of species that can occupy a given habitat (Dubuis et al., 2011). In reality, constraints to this maximum number of species are for example the *carrying capacity* or the competitive exclusion of species. The stacked species distribution modeling approach has successfully been applied to model species richness of plants on the Tenerife island based on topo-climatic and soil data (Aranda and Lobo, 2011) as well as to model species richness of alpine plants in the Swiss Belalp area based on topo-climatic predictors (Dubuis et al., 2011) and topo-climatic variables in combination with aerial photography (Guisan and Theurillat, 2000). However, no study using a stacked species distribution modeling approach in combination with satellite remote sensing data over a large spatial extent has been published so far.

Advantages and drawbacks. The majority of published articles using remote sensing data have applied the direct modeling approach for predicting species richness. Models have been fitted using standard statistical methods including some form of averaging. They therefore predict species richness with symmetrically distributed errors and tend to underestimate high and to overestimate low species richness (Dubuis et al., 2011). The biggest drawback of this approach is, however, that it does not include information regarding the identity of the species and thus cannot characterize species compositions. In contrast, without rules to constrain maximum species richness per grid cell, stacked SDMs predict the entire range of potentially occurring species in a given area and thus typically tend to overpredict species richness (Diniz-Filho et al., 2009; Dubuis et al., 2011). Further, rare species which do not have a sufficient number of occurrences often need to be excluded from the analysis, whereas they can be included in the direct modeling approach. However, stacked SDMs allow deriving not only absolute species numbers, but also information of which species (and thus communities) occur in each grid cell. The performance of both approaches in comparison has been assessed by only very few studies (Algar et al., 2009; Newbold et al., 2009). A rigorous evaluation was recently done by Dubuis et al. (2011) who emphasized the complementary strength of both approaches and proposed frameworks combining both.

Applicability and input data requirements. Systematic field inventories of species diversity as required for the direct macroecological modeling of species richness are not available for many countries and ecosystems and for the majority of taxa. Richness data may be generated from available presence-absence observations (see Figure 4.8) but in this case represent coarser spatial resolutions which may not be sufficient for several applications including conservation management (see Section 3.5). In many cases, richness assessments therefore need to rely on the summation of individual species range maps. Exploring the utility of geo-referenced museum and herbarium collections and presence-only field observations is thus required. In Chapter 7, the stacked SDM approach is applied to herbarium collection and field data of Mexican pines and validated with independent inventory data.

4.6 Summary

The aim of this chapter was to provide the theoretical background and conceptual framework for systematically modeling (plant) species distributions (Chapter 6) and species richness (Chapter 7) based on remote sensing data. In particular, the chapter summarized the existing literature and the remaining challenges. In general, three different approaches to map or model species distributions can been distinguished, namely (1) the direct mapping or identification of individuals or populations, (2) the mapping of habitats or biotopes, and (3) the analysis of proxy variables for the use in species distribution models (as done in this thesis).

Modeling species distributions based on remotely sensed data is built on the assumption that certain key environmental parameters which have remotely detectable biophysical properties determine spatial distribution patterns of species. No terminology for the contribution of remote sensing data in characterizing species distributions exists so far though. In view of the findings of vegetation mapping studies, plant-biotope relationships from a remote sensing perspective need to incorporate for example floristic gradients (*ecotone, ecocline*), species abundance, and species environmental tolerance.

The ecological relevance of the remote sensing data (products) utilized in this thesis can be summarized as follows: *Reflectance* data effectively characterize land surface characteristics, e.g. through chlorophyll absorption, vegetation water content, and soil characteristics. *Vegetation Indices* are related with certain biophysical variables such as LAI and NPP and in this sense describe biotope characteristics. Further, multi-temporal vegetation index data can be used to extract information on phenological and seasonal variations in vegetation cover and structure (so-called *Land Surface Phenology*). Finally, *Land Surface Temperature* is a valuable parameter for modeling plant species distributions, as it integrates vegetation presence and vegetation conditions with surface wetness.

In addition, the specifications of the 13 existing articles that applied the *Maximum Entropy* algorithm utilized in this thesis in combination with remote sensing data were summarized. The findings pointed out (1) that remote sensing predictors can improve climatic species distribution models at national to continental scales, (2) that especially multi-temporal vegetation index data are suitable for modeling plant species distributions, and (3) that inter-annual variability in remotely sensed time series may impact the temporal transferability and the validity of species distribution models. What is missing from this existing literature are a sensitivity analysis and a systematic framework of rules and standards for modeling species distributions by means of remote sensing data. In this thesis, therefore the existing methodological recommendations learned from previous studies are taken up and extended regarding the critical questions that have partly already been brought up by the most recent studies. Particularly, the comprehensive analysis of several remote sensing products to cover a wider range of biotope characteristics, the computation of phenological and statistical metrics from remotely sensed time series data, and the examination of the effects of inter-annual variability are addressed (see Chapter 6).

The chapter concluded with a summary of the theoretical background for modeling species richness based on remote sensing data. Two conceptual approaches exist: the direct ‘macroecological’ modeling of species richness or the use of ‘stacked species distribution models’. In this thesis, the stacked species distribution modeling approach is applied to herbarium collection and field data of Mexican pines and extended by the utilization of remote sensing data as model predictors (see Chapter 7).

5 Model input data and pre-processing

After the theoretical framework for modeling species distributions was outlined in Chapter 3 and the ecological justification for the particular remote sensing data analyzed was given in Chapter 4, in the following, the characteristics of the specific model input data used as well as the pre-processing procedures applied in this thesis are described (Figure 5.1). The occurrence data of the study species (Section 5.1) stem from two existing species databases, namely the *National Forest Inventory* (INFyS) and the *National Information System on Mexican Biodiversity* (SNIB). In addition, three sources of environmental predictors (climate, land cover classification, and remote sensing data) relevant for characterizing the distributions of the target tree species are utilized. Climate data (Section 5.2) include eleven temperature and eight precipitation measures and are obtained from the WorldClim database (Hijmans et al., 2005). The land cover product (Section 5.3) has been produced within the *North American Land Change Monitoring System* project. Finally, the remotely sensed data (Section 5.4) include two Terra-MODIS standard products (MOD13A2 and MOD11A2).

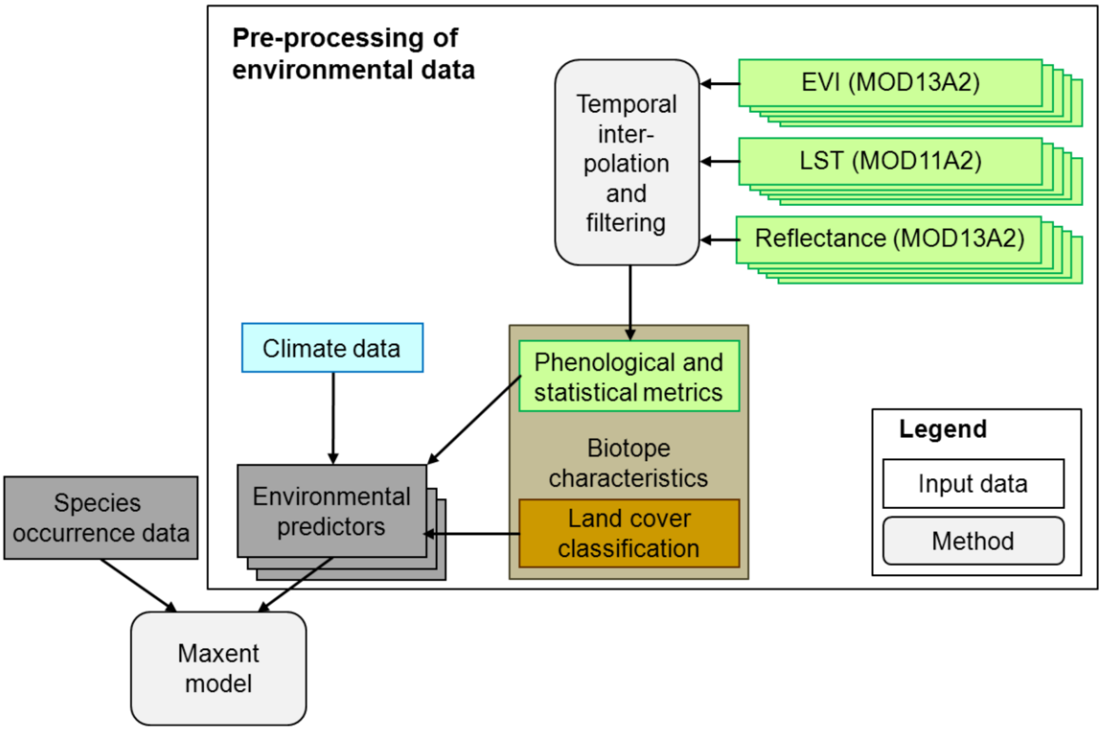


Figure 5.1: Overview of data sources used and pre-processing procedures applied in his thesis.

The multi-temporal remote sensing variables including the *Enhanced Vegetation Index* (EVI), *Land Surface Temperature* (LST), and *Reflectance* (blue, red, NIR, and MIR) are submitted to a quality assessment and assembled into enhanced continuous time series for Mexico (Section 5.4.3). These high-dimensional time series data are transformed into a set of 53 phenological and statistical metrics (Section 5.4.4) that are meaningful for describing species distributions. The application of the *Maximum Entropy* model and the settings chosen are further described in Chapter 6.

5.1 Species occurrence data

In the context of this thesis, the distributions of several tree species typical of certain forested vegetation communities (Section 2.4.1) were modeled. The ecological requirements of the study species are described in detail in Section 6.1 (for eleven tree species representative of different forest types and for one mangrove species) and in Section 7.2 (for 40 pine species typical of the *Madrean pine-oak forests*). The locality data of the target species stemmed from two existing databases, namely the *National Forest Inventory* (INFyS) and the *National Information System on Mexican Biodiversity* (SNIB).

The Mexican *National Forest Inventory* was carried out by the *National Forestry Commission* (CONAFOR, *Comisión Nacional Forestal*) between October 2004 and November 2007. Updating of the data base started in 2009 and is scheduled until 2014 with a regular annual revisit rate of 20% of the sites. Aim of the initiative is to monitor a total of 24,659 sites distributed across the whole country (Figure 5.2) based on a regular sampling scheme to obtain a detailed and reliable inventory of the national forest resources and conditions. The data set also includes soil samples which though have not been completely captured so far.

The reference area for each INFyS site (so-called *conglomerado*, Figure 5.3) is 1 ha (CONAFOR, 2009). Each *conglomerado* again includes four representative plots (*sitio*) of 400 m² for taxonomic identification and quantitative forestry measurements, e.g. diameter at breast height (DBH) and volume, of all tree individuals with a diameter of more than 7.5 cm. *Subsitio 1* is used to register saplings with a diameter of less than 7.5 cm and a height of minimum 25 cm. Taxonomic registration within each *subsitio 1* is only accomplished to the genus level though. A more detailed inventory is made within *subsitio 2* (nested within *subsitio 1*) where the principal occurrence of herbs, ferns, moss, and lichens is registered. However, species occurrence and abundance data are only collected for woody plants, primarily trees. More detailed information of the INFyS field sampling design can be found in CONAFOR (2009). To summarize, the information relevant for and utilized in this thesis is the registered species occurrence of woody plants with a diameter of more than 7.5 cm within the four *sitios* of 400 m² each.

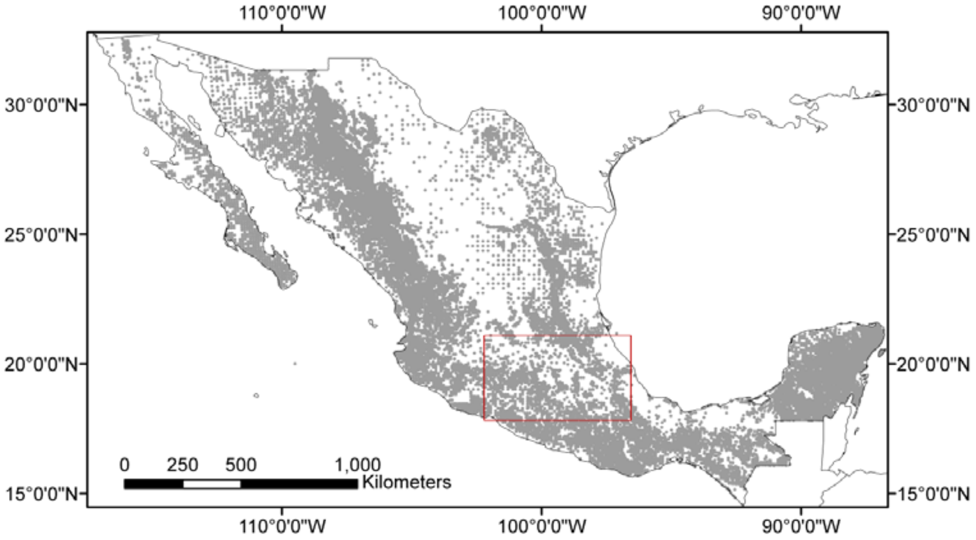


Figure 5.2: Spatial distribution of the Mexican *National Forest Inventory* (INFyS) sites. The focus area indicated with a red box is shown in Figure 5.4.

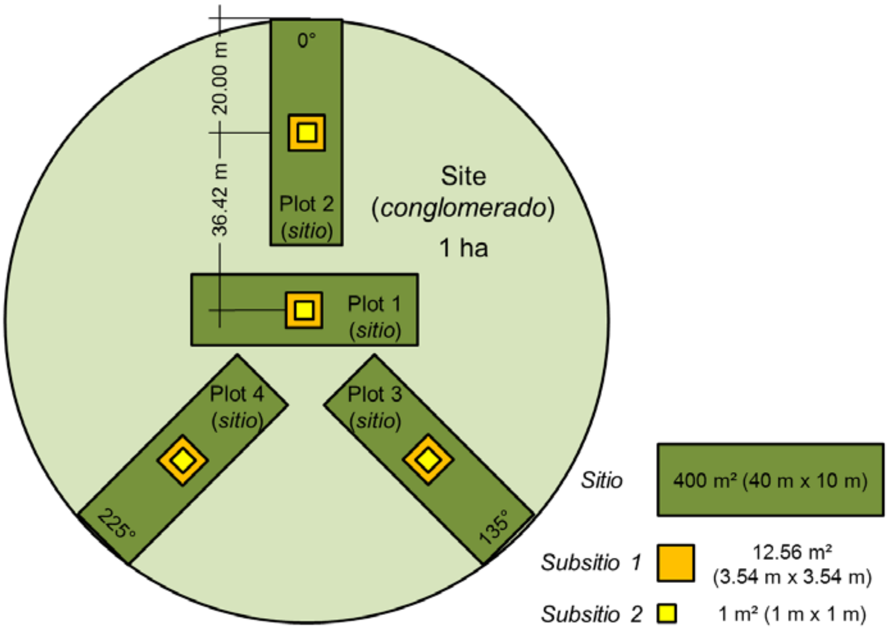


Figure 5.3: Multi-scale reference area for field data sampling at INFyS sites within the vegetation classes forest (*selva*), mangroves, and other aquatic vegetation (CONAFOR, 2009; modified). Note: For sites located within forest (*bosque*), (semi)arid vegetation, gallery forest, and palm trees, both *sitio* and *subsitio 1* have a circular shape but cover the same area.

The INFyS sampling scheme covers only natural vegetation complemented by forest plantations. The accuracy of the underlying land cover map used for stratification (Mas et al., 2002) was recently evaluated with global accuracy estimates between 64% and 78% among four eco-geographical areas defined for Mexico (Couturier et al., 2010). Distances between INFyS sites range from 5 km (for forests), through 10 km (dry forests, mangroves, wetlands) to 20 km (matorral). The data set is hence spatially biased towards forested sites (Figure 5.4). Given the spatial resolution of the environmental predictors used in this thesis, sites where a certain target species was not found, were not judged as *true* absences since non-detection of the species in the 1 ha reference area does not necessarily imply species absence within the corresponding 1 km² (see Elith et al., 2011). Instead, to account for the spatial sampling bias towards forested areas inherent in the INFyS data set, non-recorded presence was treated in the sense of the *target-group background* approach (Section see 3.4.1).

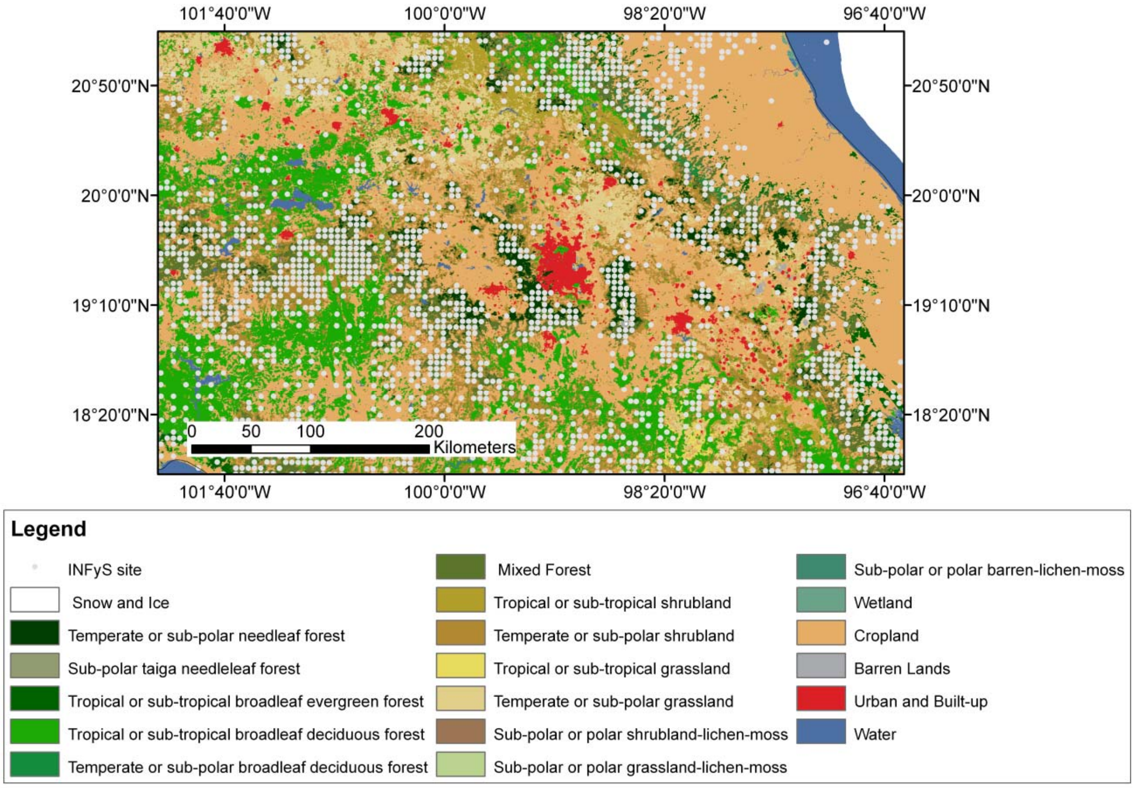


Figure 5.4: Stratification of sample sites of the Mexican *National Forest Inventory* (INFyS) depending on the respective land cover classes. The map illustrates the focus area highlighted in Figure 5.2. The land cover data shown in this figure originate from the NALCMS project (NALCMS, 2005). Note: For initially developing the INFyS sampling scheme, a different land cover data set (Mas et al., 2002) was used; the NALCMS classification is used here only for illustration purposes.

The second source of species data was the *National Information System on Mexican Biodiversity* (SNIB). This presence-only species database (containing approximately 6 million records) is maintained by CONABIO with the aim to establish a national information system of species occurrence records and provide assistance in regard to protecting biological diversity to all interested parties (government, research, social, and private sectors, see Section 1.2). All geo-referenced species data collected in the context of projects which have been funded by CONABIO (for all taxa, not restricted to plants) are integrated into SNIB under supervision of a designated database curator. In addition, existing data from 190 Mexican and 240 international biological collections have been integrated. All entries are submitted to a rigorous expert review for potential taxonomic and geolocation errors before being placed in the SNIB system. Herbarium and museum records included in SNIB date back to the late 18th century. Most plant data in SNIB have been collected between 1985 and 1990 when extensive field surveys had been carried out in Mexico; only a few records have been sampled after the year 2000. All SNIB data are distributed by CONABIO upon request at no cost. The majority of such natural collection data is based on "opportunistic" (Wisz et al., 2008: p. 764) or "purposive" (Edwards Jr. et al., 2008: p. 133) sampling which may introduce sampling biases (Figure 5.5). In this thesis, the SNIB data on Mexican pines were complemented by recent field surveys conducted by Dr. D. Gernandt (Institute of Biology, National Autonomous University of Mexico) and colleagues to enlarge sample size especially for some rare species.

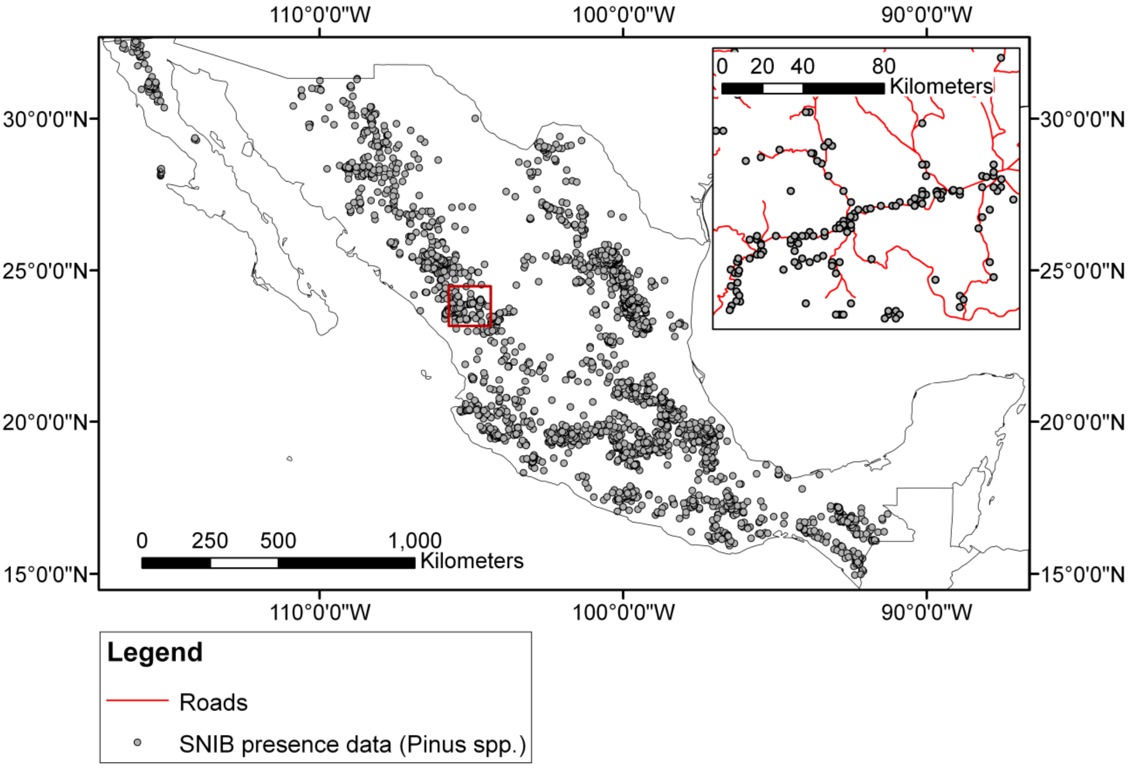


Figure 5.5: Spatial distribution of occurrence records of pine species included in CONABIO's *National Information System on Biodiversity* (SNIB). The red box indicates the area of detail shown in the inset map. The inset map illustrates the distribution of presence localities close to roads as a result of opportunistic sampling and is an indicator of their spatial bias.

Both data sources (INFyS and SNIB) were not combined in this thesis as they are based on different sampling designs and refer to other temporal sampling periods. Collection data of the INFyS are in general more reliable as they have excellent locational accuracy (based on GPS measurements) and are temporally consistent with the acquisition periods of both the land cover classification (Section 5.3) and the remote sensing variables (Section 5.4). Therefore, INFyS data were used to systematically assess the different aspects of the potential of remote sensing data for modeling species distributions (Chapter 6). On the contrary, the SNIB data have a substantial uncertainty in the localization of the records and exhibit the typical drawbacks of most biological databases that have been discussed in more detail in Section 3.4.1. As the SNIB data are representative of most natural history museum and herbarium collections, which are the richest – and in many cases the only – source of species occurrence localities (Stockwell and Peterson, 2002; Graham et al., 2004), they were used to explore the potential of such older presence-only data to obtain estimates of current species richness (according to the INFyS database) using a hierarchical framework of climate and recent remote sensing data (Chapter 7).

5.2 Climate data

Climatic variables including eleven air temperature and eight precipitation metrics (Table 5.1, Figure 5.6) that were derived from long-term measurements (1950-2000), were downloaded from the WorldClim data base (Version 1.4; http://www.worldclim.org/). The data rely on a global network of more than 46,000 climate stations from various sources such as the *Global Historical Climatology Network* (GHCN), the *World Meteorological Organization* (WMO), the *United Nations Food and Agricultural Organization* (FAO), the *International Center for Tropical Agriculture* (CIAT), and additional regional databases (Hijmans et al., 2005). These meteorological station data had been interpolated to monthly climatic surfaces at 30 arc sec (approximately 1 km) spatial resolution by using a thin-plate smoothing spline algorithm with latitude, longitude, and elevation data as independent variables by Hijmans et al. (2005).

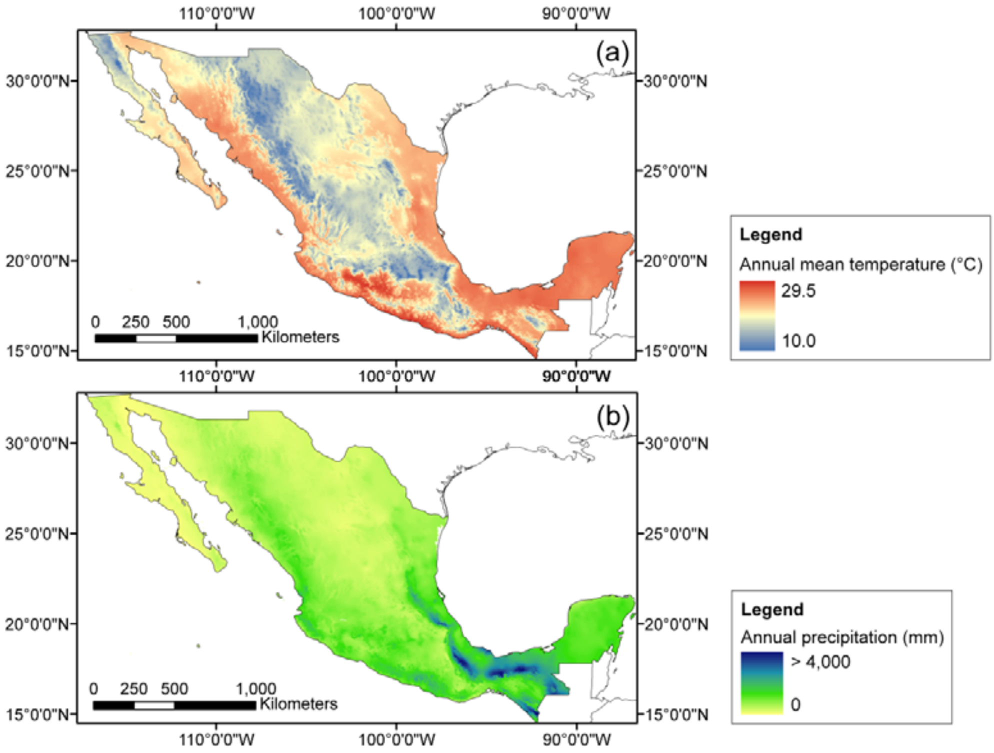


Figure 5.6: Exemplary climatic variables. (a) BIO1: *Annual Mean Temperature* and (b) BIO12: *Annual Precipitation*. Data source: WorldClim (Hijmans et al., 2005).

The monthly climatic surfaces further had been transformed into biologically meaningful variables for characterizing species distributions (Busby, 1991; Beaumont et al., 2005) which represent spatial variations in the seasonality and the extremes or certain interactions of climatic factors (Hijmans et al., 2005). The variables play a direct physiological role in limiting plant ranges and have been used in a variety of studies modeling plant distributions, also in combination with remote sensing data (Buermann et al., 2008; Saatchi et al., 2008). In this study, the readily available bioclimatic layers were resampled from their original resolution and gridded to pixel location, cell size (0.01 degree, approximately 1 km) and extent of the remote sensing data (see Section 5.4.3) to maintain spatial consistency using the ArcMap software (Version 9.3) and converted to ASCII files using the DIVA-GIS software (Version 7.3.0, http://www.diva-gis.org). The data set and the respective modeled species distribution ranges are termed 'BIOCLIM' hereafter.

Table 5.1: Full list of bioclimatic variables (referring to air temperature and precipitation) used for species distribution modeling. Data source: WorldClim (Hijmans et al., 2005).

Variable	***Description***
BIO1	Annual mean temperature
BIO2	Mean diurnal range (Mean of monthly (max temp - min temp))
BIO3	Isothermality (BIO2/BIO7) (* 100)
BIO4	Temperature seasonality (standard deviation *100)
BIO5	Maximum temperature of warmest month
BIO6	Minimum temperature of coldest month
BIO7	Temperature annual range (BIO5 - BIO6)
BIO8	Mean temperature of wettest quarter
BIO9	Mean temperature of driest quarter
BIO10	Mean temperature of warmest quarter
BIO11	Mean temperature of coldest quarter
BIO12	Annual precipitation
BIO13	Precipitation of wettest month
BIO14	Precipitation of driest month
BIO15	Precipitation seasonality (Coefficient of variation)
BIO16	Precipitation of wettest quarter
BIO17	Precipitation of driest quarter
BIO18	Precipitation of warmest quarter
BIO19	Precipitation of coldest quarter

5.3 Land cover classification

The land cover information for Mexico had been produced in the context of the *North American Land Change Monitoring System* project (NALCMS, 2005, Figure 5.7). The overall accuracy of the NALCMS land cover product was estimated at 82% for Mexico (Colditz et al., 2010). It is hence the most accurate data set currently available for this area; highest mapping accuracies were generally ascertained for forest classes, lowest accuracies for barren land and temperate shrubland (Colditz et al., 2010). In addition, the land cover data are temporally corresponding to species occurrence data from the INFyS database collected between 2004 and 2007, which is a required assumption for reliable species distribution modeling (Phillips et al., 2006).

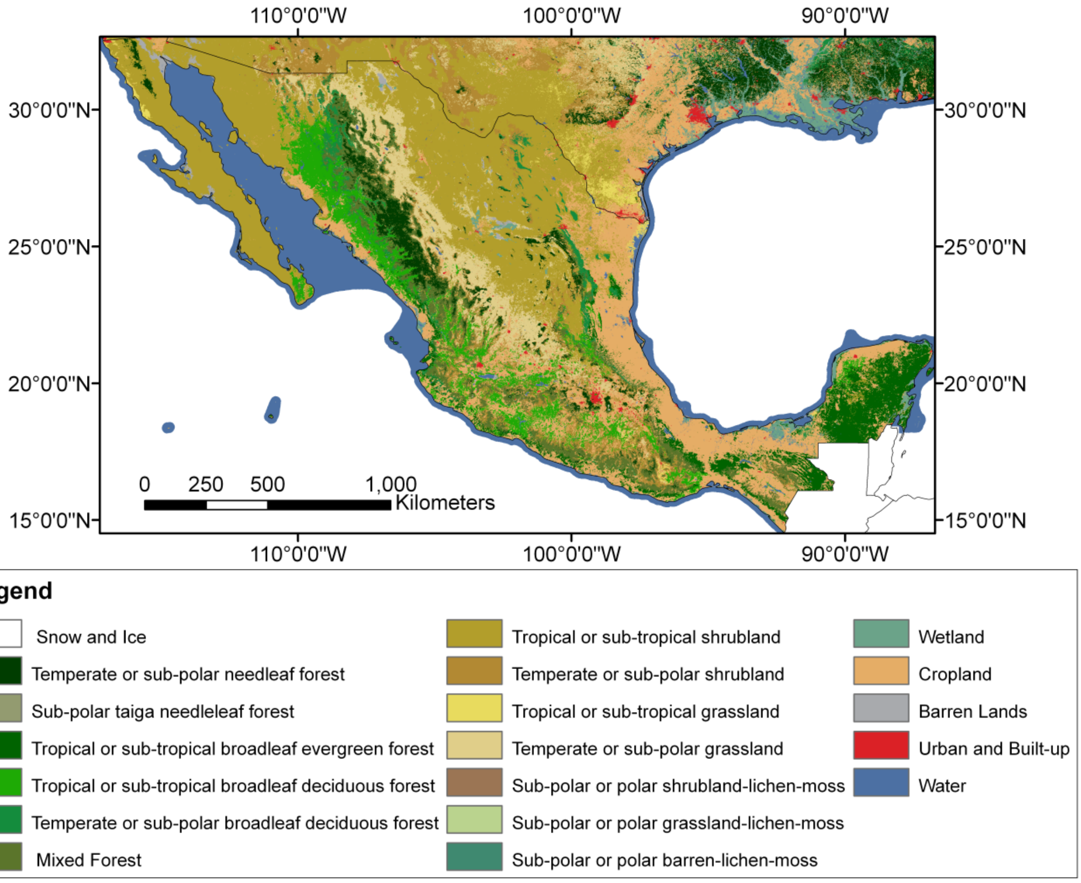


Figure 5.7: Land cover data used for modeling species distributions. Data source: *North American Land Change Monitoring System* project (NALCMS, 2005).

NALCMS is a collaborative initiative between governmental agencies of Canada, the United States, and Mexico coordinated by the *Commission for Environmental Cooperation* (CEC) to provide accurate, automatically generated annual land cover and land cover change products. The project is unique as it has been developed to meet the collective needs of the three North American countries and is carried out with the expertise from all three countries. For the NALCMS classification, monthly composites of the Terra-MODIS sensor for the year 2005 at 250 m spatial resolution together with ancillary data were used (Colditz et al., 2010). Specific procedures were developed in the project for pre-processing and spatial down-scaling of these remote sensing data (Luo et al., 2008). Using boosted decision trees, nineteen land cover classes (Level II, Figure 5.7) were mapped according to standards of the *Land Cover Classification System* (LCCS, DiGregorio and Jansen, 2005) of the *Food and Agriculture Organization* (FAO). Further information regarding the project and the land cover classification itself in digital format are distributed at the *Commission for Environmental Cooperation* website (http://www.cec.org/naatlas/nalcms).

5.4 Remote sensing variables

Remote sensing products were selected based on (1) their suitability to characterize species distributions according to the findings of previous studies (see Section 4.3) and (2) their relevance for ecosystem functioning. This ecological justification for the remote sensing data used here was given in Chapter 4. Particularly the analysis of multi-temporal ('time series') remote sensing data to quantify spatio-temporal ecosystem dynamics is reasonable for species distribution modeling (Section 4.4.3). The MODIS sensor (described below in Section 5.4.1) has emerged to one of the most important satellite time series sensors for monitoring vegetation phenology (Zhang et al., 2003; Xiao et al., 2006). In the following, the two MODIS Land standard products used (MOD13A2 and MOD11A2, Section 5.4.2) including the *Enhanced Vegetation Index*, *Reflectance*, and *Land Surface Temperature* are described. In addition, methods applied for time series interpolation and filtering (Section 5.4.3) and for the calculation of phenological and statistical metrics (Section 5.4.4) as predictors in species distribution models are outlined.

5.4.1 The MODIS instrument

MODIS represents one of the key sensors of NASA's *Earth Observing System* (EOS) aboard the *Terra* (formerly known as EOS-AM) and *Aqua* (EOS-PM) platforms (Barnes et al., 2003). The satellites were launched on December 18th, 1999 and May 4th, 2002, respectively, in a sun-synchronous, near-polar orbit (inclination of 98.2°) at an altitude of 705 km (Barnes et al., 2003). While *Terra* operates on a descending node with a local equator crossing time of 10:30 a.m., *Aqua* is on an ascending node with an equatorial overpass at 1:30 p.m. (Guenther et al., 2002). Only data acquired by *Terra* were used for this study, since (1) more data from this platform were available due to the earlier launch of *Terra* and (2) often cloud convection in the late morning and early afternoon (overpass time of *Aqua*) hinders ground observations in tropical regions. With a cross-track swath dimension of 2,340 km (Wolfe et al., 2002), global coverage can be achieved every two days and full coverage every day for latitudes higher than approximately 30° (Wolfe et al., 1998). The mean geolocation error of MODIS data was quantified to be 27 m along-track and 25 m along-scan (Wolfe et al., 2002).

The MODIS instrument provides high radiometric sensitivity (12 bit) in 36 spectral bands ranging in wavelength from 0.4 µm to 14.4 µm (Barnes et al., 1998; Guenther et al., 2002). The nadir resolutions of the bands are 250 m (Band 1: Red and Band 2: Near infrared), 500 m (Bands 3-7: Visible to short-wave infrared) and 1000 m (Bands 8-36: Visible to thermal infrared). Summary tables describing the MODIS spectral bands including spatial characteristics and signal-to-noise ratio can be found in Barnes et al. (1998) and Wolfe et al. (2002). Detailed auxiliary information on the MODIS instrument is summarized in two special issues published in *IEEE Transactions on Geoscience and Remote Sensing* (1998, Volume 36, Issue 4) and *Remote Sensing of Environment* (2002, Volume 83, Issues 1-2).

5.4.2 Characteristics of selected MODIS Land products

The product generation of MODIS imagery follows a well-defined procedure with several processing levels (Wolfe et al., 2002) that are generated by the *MODIS Adaptive Processing System* (MODAPS). The suite of available MODIS Land products was selected in a peer review process in 1992, based on scientific priorities established in the late 1980s “to meet the needs of global change research” (Justice et al., 1998: p. 1228). The Land products analyzed in this thesis are made available by the *Land Processes Distributed Active Archive Center* (LP DAAC) which distributes only data processed to level 2 or higher. These include Earth-gridded geophysical parameters (L2G and L3) and Earth-gridded model outputs (L4) (Justice et al., 1998). All high-level MODIS data are delivered in the integerized sinusoidal grid projection in non-overlapping ‘tiles’ each covering approximately 1,200 km x 1,200 km (Appendix D), which corresponds to an area of 10° x 10° at the equator (Huete et al., 2002). Nine MODIS tiles that cover the entire Mexican territory were processed and analyzed in the course of this thesis (Appendix D).

The MODIS standard products include not only the data products themselves but also the pixel-level *Quality Assessment Science Data Sets* (QA-SDS). This quality flags contain information on cloud coverage, atmospheric disturbances, viewing angle, and confidence levels of the product; they can be analyzed to produce high-quality time series (Colditz et al., 2008). Since the launch of *Terra* several MODIS data versions – the so-called ‘collections’ (Justice et al., 2002) – have been processed and distributed. In each of these irregularly updated data versions, the most recent scientific results regarding data processing utilities are included (Justice et al., 2002). Re-processing of collection 5 data (C5) used in this thesis started in 2006 and was completed in 2008; the reprocessing of C6 data is scheduled to start in summer 2012.

In the following, the characteristics of the particular MODIS Land products utilized here are outlined.

MOD13A2: *Enhanced Vegetation Index* (EVI) and *Reflectance* (blue, red, NIR, and MIR). MOD13A2 data are provided at 1 km nominal spatial resolution as a gridded Level 3 product in the HDF-EOS format. The product consists of twelve two-dimensional *Science Data Sets* (SDS) including vegetation indices, quality estimates, critical ancillary data such as view and sun zenith angles, and selected reflectance bands (see Appendix E). The vegetation indices are retrieved from daily, atmospherically-corrected, bidirectional surface reflectance data (Level 2).

The MOD13A2 product is made available at 16-day intervals using a per-pixel MODIS-specific compositing method that is dependent on the number of cloud-free observations available (Huete et al., 2002). Due to cloud contamination and extreme off-nadir viewing angles, the number of acceptable pixels over a 16-day period is typically lower than ten and often lower than five (Huete et al., 2002). For more than five cloud-free observations per pixel, *Bidirectional Reflectance Distribution Function* (BRDF) compositing (van Leeuwen et al., 1999) is used. For one up to five good pixels, the *constrained view maximum value compositing method* (CVMVC; Huete et al., 2002) is applied. For a single good observation, a simple vegetation index computation is done, and in case all observations have limitations the *maximum value compositing* (MVC) method (Holben, 1986) is used.

MOD13A2 includes the standard *Normalized Difference Vegetation Index* (NDVI) as the ‘continuity index’ (Huete et al., 2002; King et al., 2004) to the existing NOAA-AVHRR dataset. Further, it includes the *Enhanced Vegetation Index* (EVI) with improved sensitivity in high biomass regions and improved vegetation monitoring through a decoupling of the canopy background signal (Huete et al., 2002) and higher resistance to residual aerosol and atmospheric influences (Miura et al., 2001). For Mexico, it has also been shown that the EVI signal is more sensitive to temporal changes of greenness than the NDVI (Colditz, 2009). Hence, in this thesis the EVI rather than the NDVI was utilized. In addition, four reflectance data sets are included in the MOD13A2 product, namely Band 1 (red), Band 2 (near infrared), Band 3 (blue), and Band 7 (middle infrared) that were analyzed in this thesis as well.

MOD11A2: *Land Surface Temperature* (LST). Analogous to the MOD13A2 product, the MOD11A2 product is stored in the HDF-EOS format at nominal 1 km spatial resolution though as a Level 3 product. The MOD11A2 product is composed from the daily 1 km LST product (MOD11A1) and comprises the average values of clear-sky LSTs during an 8-day composite period. MOD11A2 comprises day-time and night-time LSTs retrieved under clear-sky conditions, quality assessment layers, observation times, viewing angles, bits of clear sky days and nights, and emissivities estimated in Bands 31 and 32 considering the land cover types of the IGBP classification of the MOD12 product (see Appendix F). The accuracy is better than 1 K for temperatures between 263 K and 323 K (Wan, 2003; Wan et al., 2004).

In this thesis, day-time land surface temperature was used as greater cloud contamination is statistically found in night-time LST maps (Neteler, 2010). Further, at night, only thermal infrared (TIR) band data can be used in the spectral tests needed for cloud mask production while other channels may also be used during day-time (Neteler, 2010). The LST product is generated based on a generalized split-window LST algorithm (Wan and Dozier, 1996) that computes the LST from the linear difference of the brightness temperature between Bands 31 and 32 depending on the respective land cover type derived from the MOD12 product. The algorithm uses coefficients derived by interpolation on a set of multi-dimensional look-up tables (LUT). The LUTs are obtained by linear regression of the MODIS simulation data from radiative transfer calculations over a wide range of surface and atmospheric conditions (Wan and Dozier, 1996; Wan et al., 2002). The respective coefficients account for viewing angle, water vapor, and atmospheric lower boundary temperature (Wan et al., 2002).

5.4.3 Compilation of time series data

To compile continuous time series data, the MODIS Land products MOD13A2 and MOD11A2 were downloaded via the NASA *Warehouse Inventory Search Tool* (WIST, https://wist.echo.nasa.gov/api/), which is going to be replaced by the *ECHO’s Reverb client* (http://reverb.echo.nasa.gov/reverb/) by the end of March, 2012. The nine tiles covering the Mexican territory (Appendix D) were obtained for the period from January, 2001 to December, 2009. These full volume of data comprised 23 (MOD13A2) and 46 (MOD11A2) individual data sets per year. Data of each composite period were mosaicked and re-projected from integerized sinusoidal grid projection to geographic coordinates (WGS 1984, pixel size: 0.01 degree equivalent to approximately 1 km) with the freely available *MODIS Reprojection Tool* software (MRT, Version 4.0).

Several methods have been proposed and have been implemented for the interpolation of remotely sensed time series (Pettorelli et al., 2005; Bradley et al., 2007; Colditz, 2007). In this thesis, the TiSeG (*Time Series Generator*) software package (Colditz et al., 2008) was applied which makes use of the MODIS product-specific and pixel-level *Quality Assurance Science Data Sets* (QA-SDS) to exclude low-quality data. Reduced data quality may be a result of e.g. cloud cover, atmospheric contamination or adverse viewing angles. TiSeG computes two critical quality indices per pixel over all time steps – the number of invalid observations and the maximum gap length between two valid observations (Colditz, 2007; Colditz et al., 2008). With a critical weighting between these data quality indicators and the necessary data quantity for meaningful interpolation (Colditz et al., 2008), low-quality observations can be excluded. In turn, TiSeG makes use of high-quality observations as vertices for pixel-level temporal or spatial interpolation. The approach has already successfully been applied for enhancing MODIS time series to classify agricultural land use in Central Asia (Conrad et al., 2007), to map land cover in the savanna ecosystems of southern Africa (Hüttich et al., 2009; Geßner, 2010; Hüttich et al., 2011), to estimate terrestrial carbon stocks in West Africa (Machwitz, 2010), and to analyze phenological characteristics in Mexico (Colditz, 2009; Colditz et al., 2009).

In this thesis, linear temporal interpolation of low-quality data was applied assuming that the remotely sensed variables changed steadily in time. To be able to interpolate possibly missing or low-quality data also at the beginning or end of each annual time series, three so-called ‘shoulder’ datasets (Colditz et al., 2008) were added from the previous and subsequent years, respectively. These additional time series layers were removed after interpolation. Besides the negative impact of low-quality data which was precluded by the application of TiSeG, also high-frequency fluctuations may easily be mistaken for the timing of real phenological events (Hird & McDermid, 2009). For studies aiming to derive metrics related to the timing of the growing season and vegetation development, additional noise reduction therefore has been recommended (Hird and McDermid, 2009). Hence, an adaptive local Savitzky-Golay filter (window size: 4) as implemented in the TIMESAT 3.0 software (Jönsson & Eklundh, 2004; Eklundh & Jönsson, 2009) was utilized. The Savitzky-Golay filter generally preserves higher values and is able to account for negatively-biased noise, a typical phenomenon for vegetation index data due to cloud coverage (Hird & McDermid, 2009). Savitzky-Golay filtering is also particularly suited for time series with minor noise level (Jönsson & Eklundh, 2002) – such as the enhanced time series computed with the TiSeG software implementation.

5.4.4 Phenological and statistical metrics

In high-dimensional time series data, each pixel value is necessarily highly correlated with its equivalent in the previous and subsequent time steps. Several studies have thus indicated the use of derived metrics – which are statistics (e.g. mean) or indices (e.g. ratios between bands or different periods of the time series) – for land cover classification (DeFries et al., 1995; Hansen et al., 2000). This transformation to metrics significantly reduces the feature space and allows deriving less correlated, more meaningful, and more robust input datasets (DeFries et al., 1995). A variety of methods for calculating phenological metrics from remotely sensed time series of the vegetation index has been proposed (Reed et al., 1994; Zhang et al., 2003; White et al., 2009).

One of the key methodological aspects of this thesis was the calculation of meaningful metrics that account for vegetation dynamics and seasonality as important dimensions for characterizing plant species composition and biotope characteristics (see Chapter 4). The use of such temporal metrics is reasonable as the presence or absence of a species in any area is often distinguished not only by absolute levels of vegetation or climate variables, but also by subtle differences in their seasonality (Rogers et al., 2002). A comprehensive data set consisting of 18 annual phenological metrics was compiled, based on a review of phenological metrics derived from time series of the vegetation index by previous studies (Appendix G), and implemented in the IDL programming language (ITT Visual Information Solutions, 2009).

These annual **phenological metrics** were derived from the EVI time series and its first derivative on the pixel level over the nine years of the study period (Figure 5.8, Figure 5.9, Appendix H). They included: **Temporal metrics**: Start of season (*date_SOS*), mid of season (*date_maximum*), end of season (*date_EOS*), dormancy (*date_dormancy*), length of season (*length_season*), (2) **Net primary productivity (NPP)-related metrics**: Vegetation index value at SOS (*value_SOS*), value at EOS (*value_EOS*), maximum value (*maximum*), minimum value (*minimum*), annual range (*range*), accumulated integral during vegetation period (*integral*), annual mean (*mean*), annual median (*median*) and (3) **Seasonality-related metrics**: rate of green-up (*rate_greenup*), rate of senescence (*rate_senescence*), shape of phenology curve (*skewness*), standard deviation (*standard_deviation*) and coefficient of variation (*CoV*). For temporal metrics referring to certain stages within the phenological cycle, the number of the corresponding MODIS composite (between 1 and 23 in accordance with the 16-day composite period of the MOD13A2 product, see Section 5.4.2) was assigned. The annual values were averaged over the nine years of the study period to mitigate the effects of inter-annual variability (Coops et al., 2009; Tuanmu et al., 2010) which do not necessarily reflect real changes in biotope conditions (Ichii et al., 2002). These effects of inter-annual phenological variability in remote sensing data on species distribution models were assessed in detail in Section 6.5.

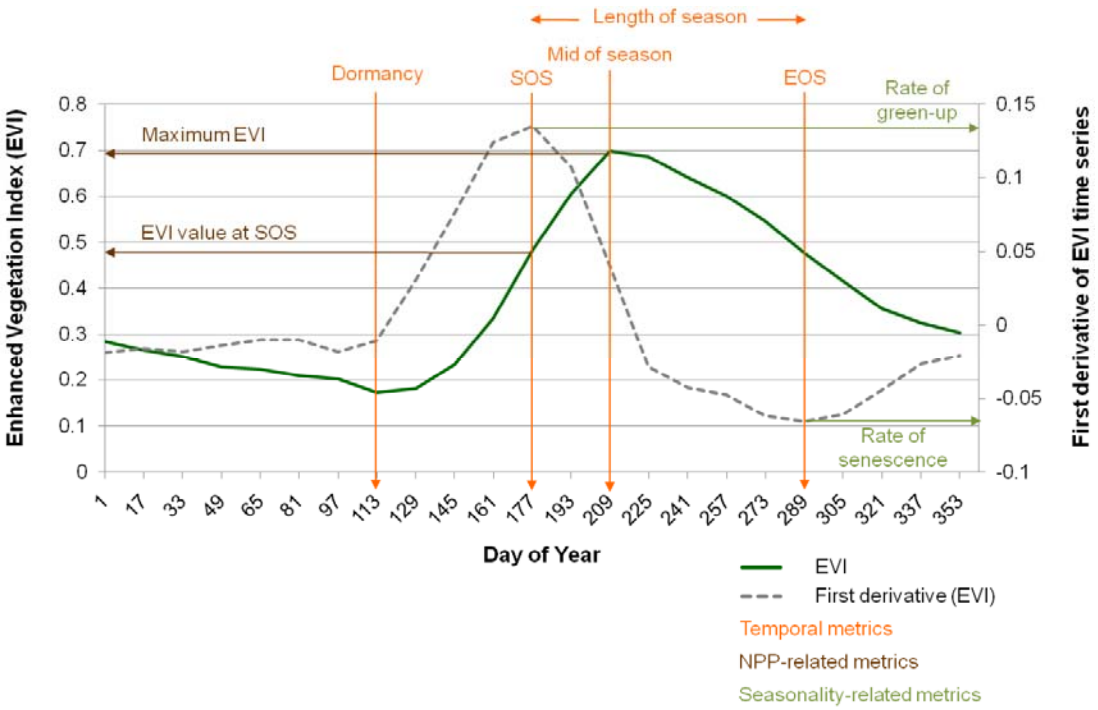


Figure 5.8: Typical interpolated and filtered MODIS-EVI time series profile (MOD13A2, year 2009) and its first derivative for tropical-deciduous forest areas in the state of *Sinaloa* (western Mexico). The figure gives a schematic representation of how the phenological metrics (temporal, NPP-related, and seasonality-related) were calculated (see also Appendix H).

The mapped spatial patterns of the phenological metrics (Figure 5.9) provided insights into the characteristics and the distribution of vegetation types within Mexico. For example, the analysis of the dates of mid of season (Figure 5.9a) illustrated that the Mediterranean climatic regime in *Baja California* with winter rain leads to a shifted vegetation phasing and a clear seasonal offset compared to the other parts of the country. Some regions along the eastern coast of the *Gulf of Baja California* and in the northeast close to the border of Texas, which are areas of contiguous irrigated agriculture (see also Section 2.5.3), exhibited conspicuously different vegetation seasonality (earlier onset and earlier maximum of the vegetation period). The data also showed a north-south gradient in the length of the vegetation period, which is primarily determined by an earlier onset and later end of the rainy season in the south (Figure 5.9b, Section 2.2). Both the EVI value at the start of the season (Figure 5.9c) and the integral under the EVI curve during the growing season (from SOS to EOS, Figure 5.9d) indicated the areas of high vegetation productivity in the country, i.e. mainly forested areas. The rate of vegetation green-up (Figure 5.9e) in turn highlighted those regions with rapid vegetation green-up rates such as the tropical deciduous forests and agricultural lands. The annual coefficient of variation (Figure 5.9f) finally allowed distinguishing between dynamic and stable areas. Stable conditions were observed (1) in the arid and desert areas on the *Baja California Peninsula* and in the *Central Plateau* which exhibit almost no seasonality and generally have very low levels of vegetation cover, (2) in the temperate forests along the mountain ranges, and (3) in the mainly evergreen tropical forests on the Yucatan peninsula.

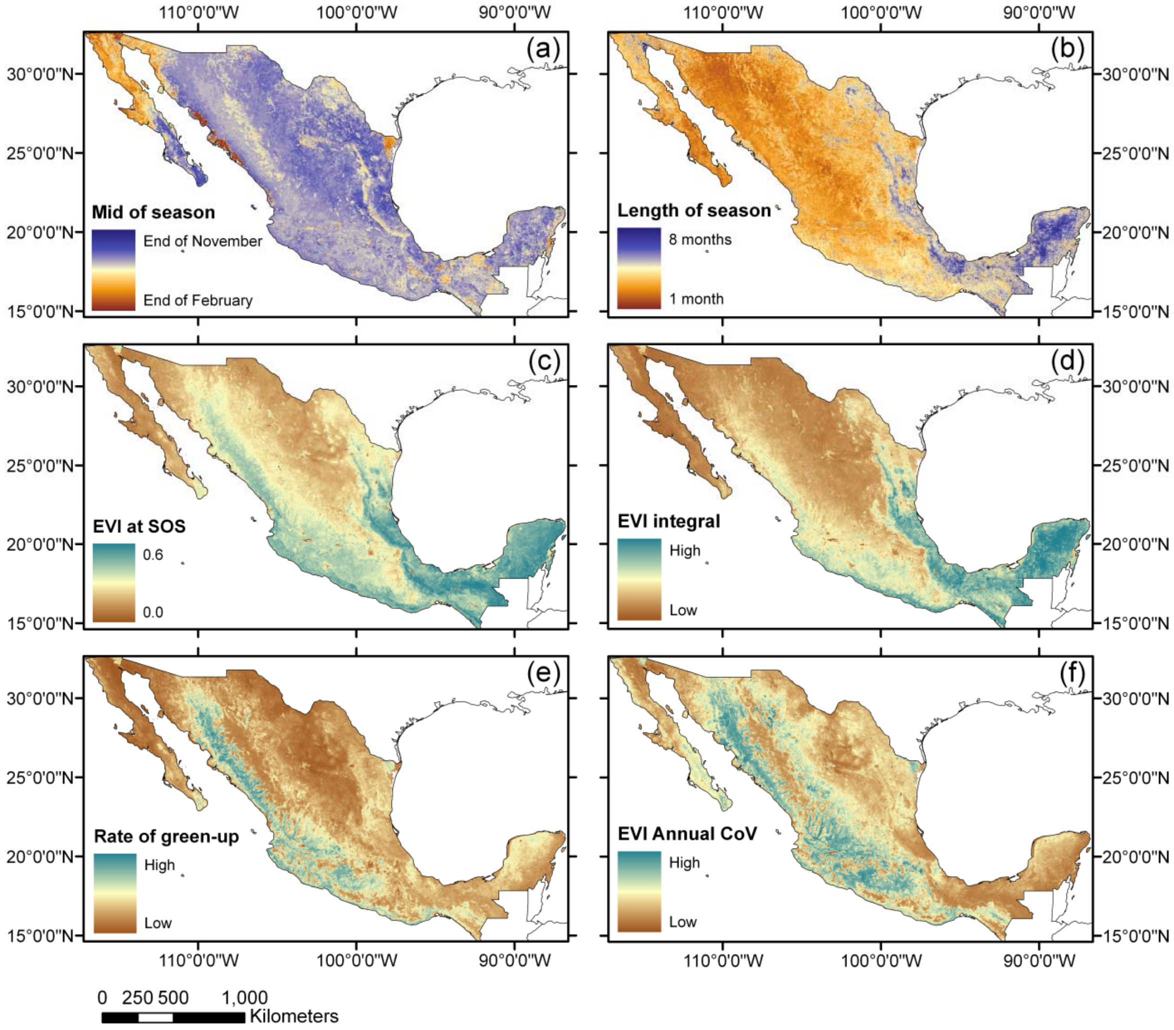


Figure 5.9: Selected phenological metrics derived from MODIS-EVI (MOD13A2) time series (see also Figure 5.8). The figures illustrate annual metrics that were averaged over the nine years of the study period including temporal metrics (a, b), net primary productivity (NPP)-related metrics (c, d), and seasonality-related metrics (e, f). The figures show (a) Day of year when the maximum annual EVI value is reached indicating the mid of season, (b) Length of the season or vegetation period, (c), EVI value at the start of the season, (d) Integral under the EVI curve during the growing season (from SOS to EOS), (e) Rate of vegetation green-up measured by the maximum of the first derivative of the EVI curve, and (f) Annual coefficient of variation (CoV) of the EVI time series as indicator of vegetation seasonality.

In addition to the phenological metrics, the characteristics of the LST time series and the four reflectance time series (blue, red, NIR, and MIR) were summarized each in a set of seven annual **statistical metrics** (*minimum*, *mean*, *median*, *maximum*, *range*, *standard deviation*, and *coefficient of variation*; Figure 5.10, Appendix H). Analogous to the phenological metrics, the annual values were averaged over the nine years of the study period. These 35 statistical metrics, in combination with the 18 phenological metrics, represent the set of remote sensing predictors used for species distribution modeling in this thesis. This data set and the respective modeled species distribution ranges based on it are termed ‘REMOTE’ hereafter.

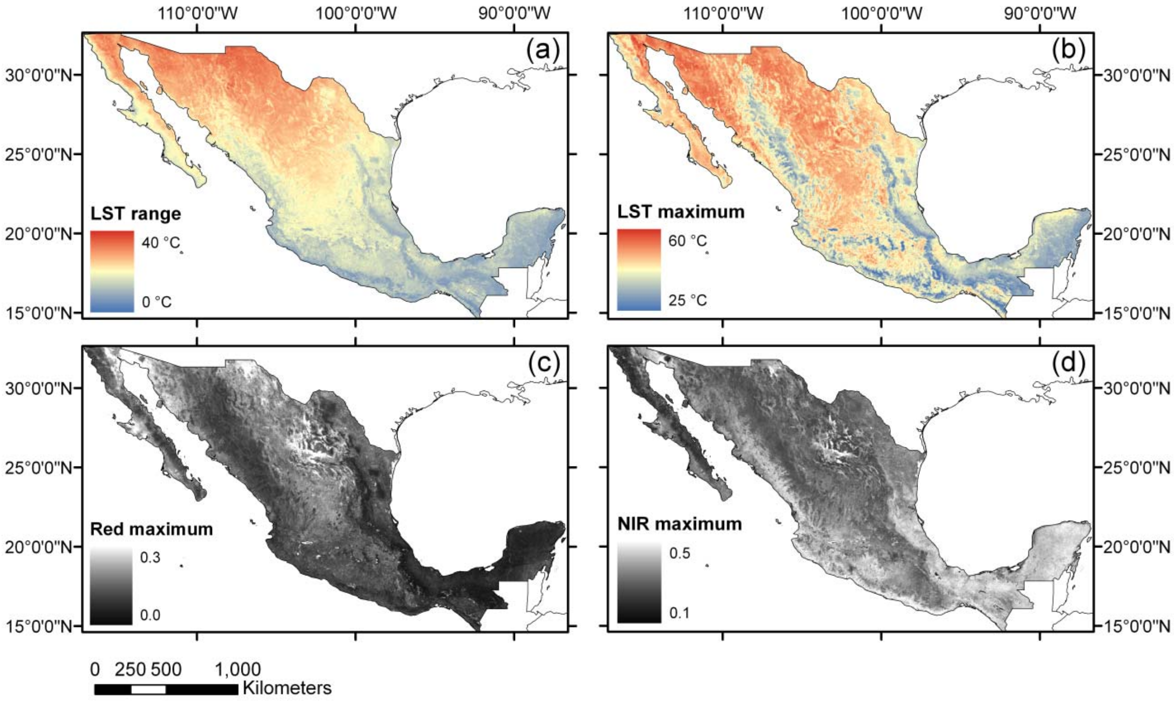


Figure 5.10: Selected statistical metrics derived from (a) and (b) MODIS-LST (MOD11A2) and (c) and (d) reflectance (MOD13A2) time series. (a) Annual range of measured LST values, (b) Annual mean maximum LST, (c) Annual mean maximum red reflectance, and (d) Annual mean maximum NIR reflectance.

The mapped spatial patterns of the statistical metrics derived from LST and reflectance time series (Figure 5.10) described the land surface characteristics in Mexico. The statistical LST metrics further illustrated the relationship between vegetation cover and measured LST scores as outlined in Section 4.4.5. For example, the annual LST range (Figure 5.10a) was closely related to the density of vegetation cover and featured very similar spatial patterns as the EVI integral during the growing season (Figure 5.9d). The general climatic north-south and altitudinal temperature gradient as seen in interpolated air temperature data (Figure 5.6a) was also apparent in measured maximum LST data (Figure 5.10b). In this case, the moderation effect of dense vegetation cover on LST was evident, e.g. along the forested mountain ranges and in the evergreen forests of the Yucatan peninsula. For areas with very low or missing vegetation cover, the annual maximum LST was typically higher than the air temperature (Figure 5.6 a; see Julien et al., 2006).

The blue, red, and MIR bands of MODIS time series obtained in this thesis were fairly stable most of the year, except for some variation in the red band over areas of irrigated agriculture and deciduous forests. The largest seasonal variation was observed in the NIR band as a result of vegetation seasonality (Huete et al., 2002). Figure 5.10c and Figure 5.10d show the maximum red and NIR reflectance values, respectively. Accordingly, red reflectance was highest in areas with a large fraction of bare soil, primarily in the north. In contrast, it was lowest in highly vegetated parts, especially in the south of Mexico, due to plant chlorophyll absorption in the red spectral range. On the contrary, NIR reflectance was highest for areas with highly dense vegetation (due to mesophyll reflection) but also in areas with large proportions of bare soil.

Given the manifold environmental conditions observed in Mexico, only the combination of several remotely sensed metrics together allowed characterizing the complex vegetation seasonality and dynamics as input for predictive species distribution models.

5.5 Summary

While the previous chapter (Chapter 4) had addressed the ecological relevance of the remote sensing data analyzed, this chapter introduced both input data characteristics and the methods applied for data pre-processing to model species distribution ranges and to predict species richness in this thesis. The species occurrence data stemmed from two existing databases, namely the systematic *National Forest Inventory* (INFyS) and the presence-only *National Information System on Mexican Biodiversity* (SNIB). Both data sources differed in their sampling designs and locational accuracy and referred to other temporal sampling periods; they were thus not combined to increase sample size per species.

Secondly, three sources of environmental predictors (climate, land cover classification, and remotely sensed phenological and statistical metrics) relevant for characterizing the distribution of the target tree species were described. The bioclimatic variables included eleven air temperature and eight precipitation metrics that had previously been interpolated based on long-term climatic measurements between 1950 and 2000 (Hijmans et al., 2005). The land cover information used in this thesis had been produced in the context of the *North American Land Change Monitoring System* project (NALCMS, 2005) and is the most accurate data set currently available for this area (Colditz et al., 2010). Further, the relevant steps and methods applied for pre-processing of remotely sensed time series and temporal interpolation were summarized. The respective remote sensing data included two Terra-MODIS standard Land cover products (MOD13A2 and MOD11A2). In particular, *Enhanced Vegetation Index* (EVI), *Land Surface Temperature* (LST), and *Reflectance* (blue, red, NIR, and MIR) were submitted to a rigorous quality assessment and assembled into enhanced continuous time series. These high-dimensional time series data were transformed into 18 phenological and 35 statistical metrics that had been selected based on a literature review conducted in this thesis. Given the manifold environmental conditions observed in Mexico (see Chapter 2), only the combination of several remotely sensed metrics together allowed characterizing the complex vegetation seasonality and dynamics as input for predictive species distribution models.

6 Species distribution modeling with remote sensing data

Remote sensing data have a significant potential for predicting species distributions (outlined in Chapter 4) based on the assumption that particular environmental variables with biophysical properties determine the spatial distribution of a species and can be detected by remote sensing (Turner et al., 2003). This chapter deals with (1) the synergistic analysis of climate and remote sensing data in species distribution models and (2) the effects of different remotely sensed input variables on model performance. Since the aim of this thesis is to develop a robust and transferable framework for modeling species distributions with multi-temporal remote sensing data, emphasis is given to methodological aspects and their implications regarding modeled species distributions. All analyses described in the following were implemented in the R programming language (R Development Core Team, 2011).

First, the characteristics (e.g. range size, vegetation type) of the twelve target tree species are described (Section 6.1). Hereafter, the **hierarchical modeling framework** applied in this thesis in order to integrate climate and remote sensing data is illustrated (Section 6.2). Both ecological theory and spatial autocorrelation are discussed in this context. The following sections focus on the selection of appropriate methods to model biotope suitability based on multi-temporal remote sensing data. Within all individual studies, the impact of the species' ecological traits (range size, biotope specificity) is explicitly considered as it may directly affect the usefulness of remote sensing data. Section 6.3 compares the suitability of **continuous remote sensing time series** and an existing **categorical land cover classification** for the prediction of biotope suitability. In Section 6.4, the **importance of different remote sensing products** to model tree species that are representative of different forest types (e.g. tropical compared to temperate climates; deciduous as opposed to evergreen vegetation communities) is assessed. Finally, the effect of **inter-annual variability** in time series of the vegetation index for modeling species distributional ranges is explored (Section 6.5). The chapter concludes with a summary of the major findings (Section 6.6).

6.1 Study species

Species' ecology has an impact on the accuracy and performance of climatic species distribution models (Hernandez et al., 2006; McPherson and Jetz, 2007; Pöyry et al., 2008; Syphard and Franklin, 2010) and may thus also influence the usefulness of remote sensing data for modeling species distributions (see Section 4.2). For example, remote sensing data are expected to be particularly valuable for widespread species which occur in comparatively homogenous climates (Buermann et al., 2008). The tree species studied in this thesis (Table 6.1) are representative of the major Mexican forest types (see Section 2.4.2) and were chosen to capture a wide variety of ecological traits such as range size or biotope specificity. An overview of biotope characteristics (distributional area, range size, climatic conditions, and vegetation type), sample size and conservation status of the target species is given in Table 6.1. Maps showing the spatial distribution of the presence sites across Mexico can be found in Appendix I. Species occurrence data used in all analyses described in this chapter stem from the recently accomplished *National Forest Inventory* (INFyS, Section 5.1) and are thus temporally consistent with the remote sensing and land cover data analyzed. The *National Forest Inventory* was carried out between 2004 and 2007, remote sensing data cover the period from 2001 to 2009 (Section 5.4), the land cover classification represent the year 2005 (Section 5.3).

Table 6.1: Overview of the target tree species. The corresponding occurrence records stem from the Mexican *National Forest Inventory*. ‘Phenology’ in this table refers to the study species and not to the corresponding vegetation type. ‘NOM’ is the *NORMA Oficial Mexicana* NOM-059-SEMARNAT-2010 (SEMARNAT, 2010), which is the recently updated official list of endangered species in Mexico.

Species	***Life form***	***Phenology***	***Distribution***	***Range size***	***Climate***	***Vegetation type***	***Records***	***Included in NOM***
Abies religiosa	conifer	evergreen	Trans-Mexican Volcanic Belt	restricted	temperate	Temperate needle leaved evergreen forest	132	no
Alnus acuminata	broadleaf	evergreen/ deciduous	Neotropical	wide	temperate	Temperate deciduous-evergreen forest	229	no
Arbutus xalapensis	broadleaf	evergreen	Neartic-Neotropical	wide	temperate	Temperate deciduous-evergreen forest	2,530	no
Astronium graveolens	broadleaf	deciduous	Neotropical	wide	tropical	Tropical or sub-tropical evergreen forest	438	yes
Avicennia germinans	broadleaf	evergreen	Neotropical	wide	tropical	Wetlands	54	yes
Bursera bipinnata	mimosoid	deciduous	Neotropical	intermediate	tropical	Tropical or sub-tropical deciduous forest	241	no
Bursera simaruba	broadleaf	deciduous	Neartic-Neotropical	wide	tropical	Tropical or sub-tropical evergreen forest	3,940	no
Cedrela odorata	broadleaf	deciduous	Neotropical	wide	tropical	Tropical or sub-tropical evergreen forest	381	yes
Guaiacum sanctum	broadleaf	deciduous	Southern Mexico	restricted	tropical	Tropical or sub-tropical deciduous-evergreen forest	118	no
Liquidambar macrophylla	broadleaf	deciduous	Southern Mexico	restricted	tropical montane	Moist montane/cloud forest	91	no
Liquidambar styraciflua	broadleaf	deciduous	Neartic-Neotropical	intermediate	tropical montane	Moist montane/cloud forest	127	no
Pinus chiapensis	conifer	evergreen	Sierra Madre del Sur	restricted	tropical montane	Temperate needle leaved evergreen forest	21	yes

6.2 Hierarchical combination of climate and remote sensing data

Various concepts for the classification of environmental variables used in species distribution modeling have been proposed (outlined in Section 3.2.1, summarized in Table 3.1). All these suggest that certain environmental variables determine species distributions at different spatial scales. Pearson and Dawson (2003) introduced an explicitly hierarchical point of view suggesting that climate may be more important at *global* to *regional* scales while land use is more relevant at *landscape* to *site* scales (Figure 3.5). This call for and the successful implementation of hierarchical frameworks (Pearson and Dawson, 2003; Pearson et al., 2004; Luoto et al., 2007) to integrate land cover information in climatic species distribution models has so far largely been ignored in studies making use of remote sensing data. Almost all remote sensing based modeling studies either did not incorporate climate data at all or combined both data sources in a single model (Table 4.1).

6.2.1 Objectives

The recommendation of using hierarchical modeling framework to integrate land cover characteristics in species distribution models (Pearson and Dawson, 2003; Pearson et al., 2004; Luoto et al., 2007) was taken up in this thesis and extended by using remote sensing variables instead of land cover classification data. The aim of this hierarchical modeling scheme is to refine areas “within the suitable climate space of the species” (Luoto et al., 2007: p. 39) using the results of the biotope suitability model to obtain spatially explicit maps of probability of species occurrence. The order of data integration into the hierarchical framework (first climate, second remote sensing) is based on the assumption that biotope availability is not sufficient for species survival if the climatic conditions with physiological relevance (e.g. temperature) are not suitable. Besides this ecological (scale-based) justification, also data characteristics suggest the combination of climate and remotely sensed information in a hierarchical manner. The relevant data characteristics (particularly *sample density* and *spatial autocorrelation*, Section 6.2.2) are specifically assessed in this analysis. The alternative non-hierarchical combination of climate and remote sensing data within the same species distribution model was tested, too, but yielded models with lower predictive power (measured by statistical model deviance) than the proposed hierarchical framework. In the following, only the implementation and results of the hierarchical approach are outlined.

The hierarchical scheme implemented in this thesis (Figure 6.1) first involves *Maximum Entropy* (Maxent) models for predicting continuous **climatic suitability** (BIOCLIM) based on WorldClim climate data (described in Section 5.2). Using a model thresholding approach, these continuous predictions are converted to binary climatic distribution ranges for each species (BIOCLIM_RANGE). Separately, **biotope suitability** (REMOTE) is modeled on the basis of either phenological and statistical metrics derived from multi-temporal MODIS data, namely *Enhanced Vegetation Index* (EVI), *Land Surface Temperature* (LST), and *Reflectance* data (see Section 4.4), or alternatively a categorical land cover classification (see Section 5.3). Both results are combined in a hierarchical manner by using the climatic distribution range for each species as a presence-absence mask to identify regions with both suitable climate and suitable land surface characteristics for the given species (BIOCLIM_REMOTE). This BIOCLIM_REMOTE model is the main outcome of the hierarchical framework based on a synergistic combination of climate and remote sensing data and used in this thesis to predict the spatial distributions of the target tree species in Mexico.

To assess predicted species range sizes for the different models, also the continuous predictions REMOTE and BIOCLIM_REMOTE are converted to binary presence-absence predictions using the same model thresholding approach as for the climatic model. The results are termed REMOTE_RANGE and BIOCLIM_REMOTE_RANGE, respectively. However, these binary predictions are only used to estimate variations of the different modeling approaches regarding the *fractional predicted area* (FPA, Section 3.4.5), i.e. the proportional area of the Mexican land surface that is predicted as species presence.

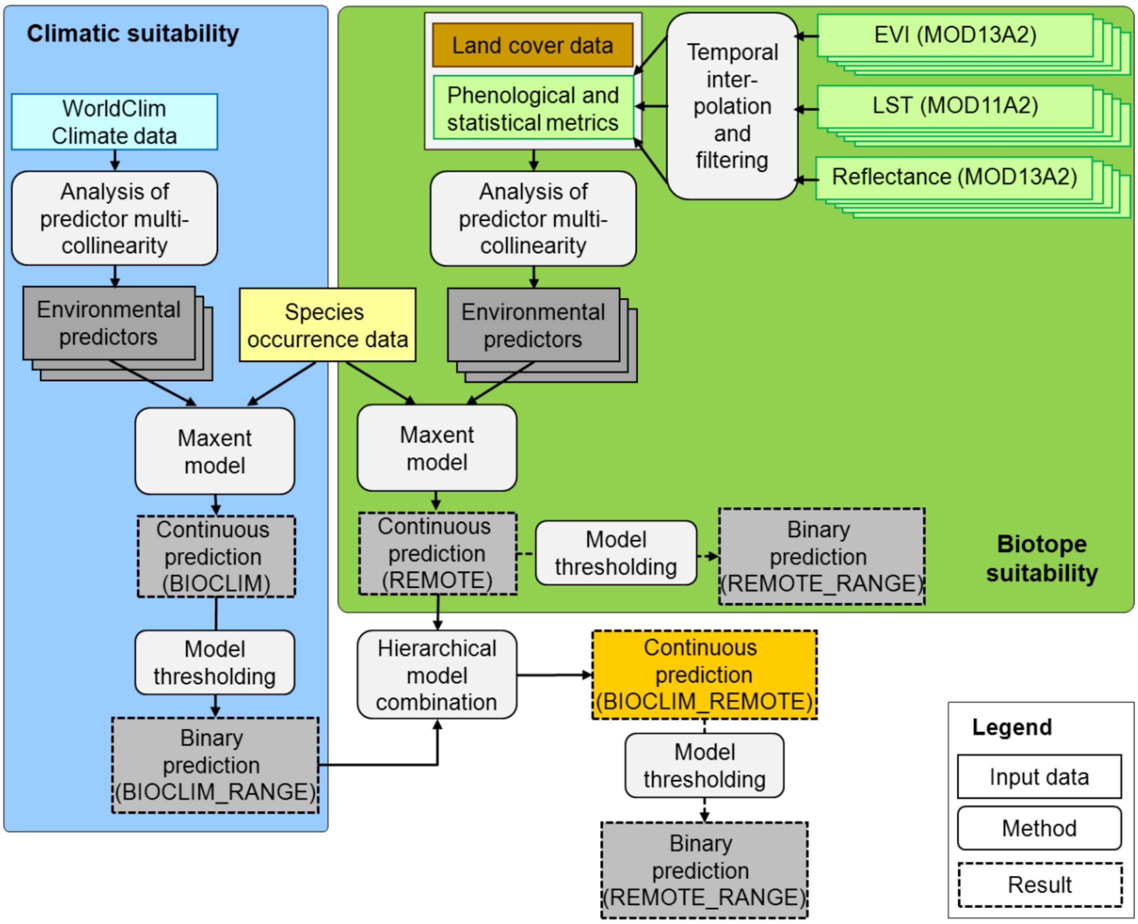


Figure 6.1: Hierarchical framework for modeling species distributions in this thesis. Climatic suitability (BIOCLIM) and biotope suitability based on remote sensing data (REMOTE) are modeled in separate *Maximum Entropy* (Maxent) models and combined to obtain continuous estimates of probability of species occurrence (BIOCLIM_REMOTE). The scheme further includes the analysis of multi-collinearity prior to the selection of environmental predictors.

The modeling framework implemented in this thesis does not only consider the hierarchical model structure, but also – based on the findings of previous studies accomplished without using remote sensing data – further relevant methodological aspects. These are in particular the analysis of multi-collinearity between environmental predictors and a species-specific selection of both climate and remote sensing model input parameters (Section 6.2.3). Both aspects have been poorly considered in remote sensing based species distribution modeling so far (but see Buermann et al., 2008).

The objective of the following sections is hence

(1) to illustrate the different **sample densities** and intensities of **spatial autocorrelation** in climate and remote sensing data (Section 6.2.2),

(2) to show that aspects such as **predictor multi-collinearity and variable selection**, so far only considered for climate data, are also relevant for remote sensing data (Section 6.2.3),

(3) and to assess the **model results** and compare the **performance** between the climatic (BIOCLIM), the remote sensing based (REMOTE), and the combined hierarchical (BIOCLIM_REMOTE) modeling approaches.

6.2.2 Sample density and spatial autocorrelation

Besides the ecological justification of hierarchical modeling frameworks outlined above, also data characteristics suggest the combination of climate and remotely sensed information in a hierarchical manner. Both data sources used in this thesis have the same *nominal* 1 km resolution; however, climatic data are limited in their *effective* spatial resolution. This effective resolution of climate data is typically coarser than their nominal resolution due to the limited number of climate station data available and the interpolation techniques applied to create continuous raster layers. Over large areas and in particular over regions with highly variable topography such as Mexico (see Section 2.1), the accuracy of interpolated climatic surfaces is therefore mainly sensitive to the station network density (Hijmans et al., 2005). The effective spatial resolution of environmental data sets can be summarized by the so-called *sample density* (Franklin, 2009) which is defined as the ratio of the total sampled area to the extent of the environmental data. Sample density is typically low for climatic surfaces based on data of widely dispersed climate stations (approximately 0.001 to 0.01), but can be very high for remotely sensed variables (up to 1.0) where each pixel is directly measured (Franklin, 2009). These instances lead to significant autocorrelation (defined as the covariation of variables with space, Section 3.4.2) in climate data.

Spatial autocorrelation (SAC, see Section 3.4.2) of both climate and remote sensing data was assessed based on the Moran's I index (Moran, 1950) as implemented in the R package 'raster' (R Development Core Team, 2011). Moran's I is a measure of global spatial autocorrelation and thus of the overall spatial clustering of the data. The index can take values from -1 (indicating the perfect dispersion or the clustering of dissimilar values) to +1 (indicating the perfect correlation or the clustering of similar values). Scores close to zero indicate a random spatial pattern (Anselin, 1995). The results (Figure 6.2) showed that spatial autocorrelation decreased for both climate and remote sensing data with larger window sizes – environmental heterogeneity thus increases with growing spatial distance. For all window sizes, Moran's I was significantly lower for remote sensing than for climate data ($p < 0.001$, Mann-Whitney U-test). Accordingly, climate variables showed a higher clustering of similar values due to the spatial interpolation techniques applied. On the contrary, remote sensing data as direct measurements reflected environmental variability detectable at the corresponding grain size. Remote sensing variables further featured a higher variability in Moran's I than climate data. Temporal phenological metrics (see Section 5.4.4) and statistical metrics derived from reflectance data generally showed the lowest SAC scores (data not shown here). The highest autocorrelation was observed for *Land Surface Temperature* metrics followed by *Net primary productivity* (NPP)-related metrics. No trend or pattern was found for climate variables.

Similar to environmental predictors, also species occurrence data featured spatial autocorrelation due to spatially biased sampling (Section 3.4.1). Reference species distribution models based on a non-hierarchical combination of climate and remote sensing data hence systematically overestimated the importance of climate data for all study species.

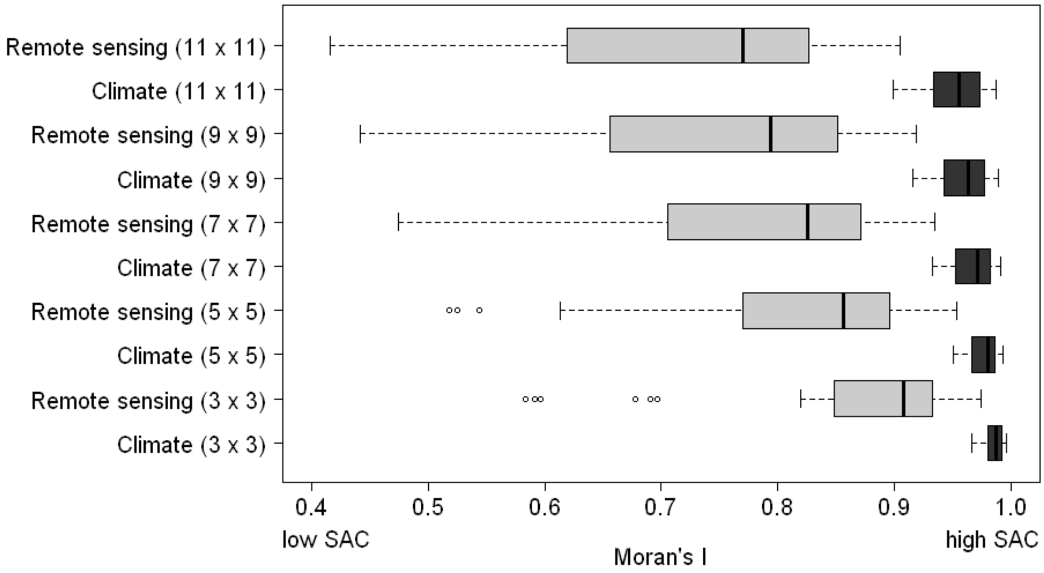


Figure 6.2: Spatial autocorrelation of the environmental raster data measured by the Moran's I index for different window sizes (e.g. 3 x 3). Moran's I values of 0.0 indicate random spatial patterns while values of 1.0 imply perfect autocorrelation.

6.2.3 Predictor multi-collinearity and variable selection

The inherent collinearity of explanatory variables hampers statistical analyses and the ecological interpretability of species distribution models (Graham, 2003; Heikkinen et al., 2006; Dudík, 2007; Baldwin, 2009). Different solutions have been proposed and applied to this problem (see Section 3.4.3). Nevertheless, predictor multi-collinearity was largely ignored in most previous studies that integrated remote sensing data into species distribution models (but see Buermann et al., 2008).

In the *Maximum Entropy* model used in this thesis, no straightforward way exists to identify and remove correlated predictors (Baldwin, 2009). Hence, prior to species distribution modeling, cross-correlation was computed within (1) climatic variables and (2) remote sensing parameters for each species to exclude redundant data. Pair-wise Spearman's rank correlation coefficients between environmental predictors were calculated based on species presence and pseudo-absence localities. For each species, all pairs of highly-correlated environmental predictors (Spearman's $r > 0.8$) were identified. Out of each pair of correlated predictors, the variable with the higher explanatory power according to a univariate *Generalized Linear Model* (GLM) was retained. The correlation analysis was implemented in the R programming language (R Development Core Team, 2011; code modified from C. Dormann). The flexible and species-specific procedure was applied for all target species (including the pine study species, see Chapter 7).

Redundancy was particularly high for closely related variables such as for example *Precipitation of the wettest month* (BIO13) and *Precipitation of the wettest quarter* (BIO16). For the remotely sensed phenological and statistical metrics, covariation was generally higher for metrics derived from reflectance than from LST or EVI data. In addition, it was particularly pronounced for statistically very similar predictors such as standard deviation and range or maximum and mean scores. Overall, predictor multi-collinearity was lowest for temporal metrics (e.g. *Start of season*). On average, 6.8 climatic predictors (35.5%) and 13.7 remote sensing variables (25.8%) were retained after the correlation analysis (Table 6.2). The non-correlated predictors selected for each species differed between species but no set of rules depending on species traits could be identified.

Table 6.2: Number of predictors retained from the 19 climatic and 53 remote sensing variables after the analysis of predictor multi-collinearity for each of the study species.

Species	***Climate***	***Remote sensing***
Abies religiosa	7	16
Alnus acuminata	7	13
Arbutus xalapensis	7	14
Astronium graveolens	5	13
Avicennia germinans	7	13
Bursera bipinnata	9	17
Bursera simaruba	5	11
Cedrela odorata	6	12
Guaiacum sanctum	7	13
Liquidambar macrophylla	7	13
Liquidambar styraciflua	7	14
Pinus chiapensis	7	15
Mean	**6.8**	**13.7**

A correlogram of the five most important (importance measured based on the increase in regularized model gain; see Section 3.3.2) climatic and remote sensing predictors for *Bursera simaruba* is exemplarily shown in Figure 6.3. Cross-correlation based on Spearman's r was determined based on species presence and pseudo-absence sites for both data sources (climate and remote sensing) independently; it may thus exceed the above-mentioned threshold of 0.8 between pairs of layers from the two different data sources. For example, the highest correlation of 0.81 in the correlogram was observed between annual minimum *Land surface Temperature* (*LST_min*) as measured by means of remote sensing and the *Minimum (air) temperature of the coldest month* (BIO6) included in the WorldClim climate data set.

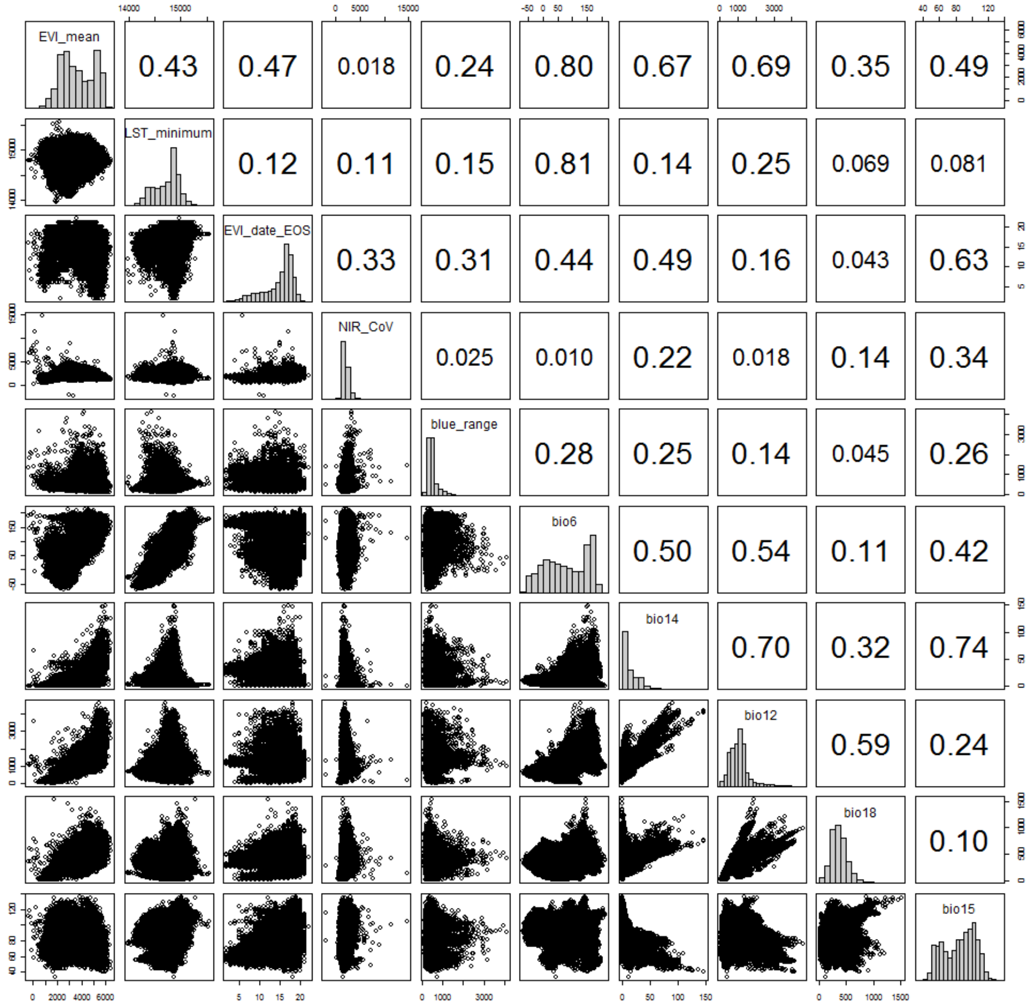


Figure 6.3: Correlogram of the five most important climate and remote sensing predictors for *Bursera simaruba*. The figure includes scatterplots (*lower triangle*) to visualize two-dimensional relationships between variables as well as histograms (*diagonal*) distributions for each variable. The *upper triangle* shows Spearman's r correlation coefficients of each pair of data sets to indicate the strength of variable correlation (the size of the numbers denotes the strength of correlation). Abbreviations: EVI_mean – Annual mean EVI, LST_minimum – Annual minimum LST, EVI_date_EOS – End of season, NIR_CoV – Annual coefficient of variation of NIR reflectance, blue_range – Annual range of blue reflectance, BIO6 – Minimum temperature of coldest month, BIO14 – Precipitation of driest month, BIO12 – Annual precipitation, BIO18 – Precipitation of warmest quarter, BIO15 – Precipitation seasonality.

6.2.4 Modeled species distribution ranges

Climatic suitability (BIOCLIM). As climatic model input variables the globally available WorldClim data base (Hijmans et al., 2005; Section 5.2) was used. To predict bioclimatic suitability maps for each species, *Maximum Entropy* models (Section 3.3.2) as implemented in its software version 3.3.3e (Phillips et al., 2004; Phillips et al., 2006) using only the previously identified non-correlated climatic predictors (Section 6.2.3) were run. The models included five replicates with replicate samples selected based on bootstrap resampling (Specific settings: replicates=5, replicatetype=bootstrap, auto features, randomtestpoints=25, applythresholdrule=minimum training presence, responsecurves, jackknife, addsamplestobackground, nodata=-3000, regularization multiplier=1, maximum iterations=500, convergence threshold=0.0001, writeplotdata, autorun, nodoclamp, nowritemess, nowriteclampgrid, nooutputgrids, writebackgroundpredictions, noprefixes, randomseed, redoifexists, notooltips). The proportion of test points was selected according to Huberty's 'rule of thumb' (Huberty, 1994; Section 3.4.5). Given the number of predictors retained after the correlation analysis (Table 6.2), test percentages of 29% for climate data and 22% for remote sensing were recommended. Herein, the mean proportion of 25% was used for both climatic and remote sensing based models. As an example, the modeled climatic suitability for *Bursera simaruba* across the Yucatan peninsula is shown in Figure 6.4 a.

Climatic distribution ranges (BIOCLIM_RANGE). At this stage of the hierarchical modeling framework (Figure 6.1), omission errors (false negatives) of species records were more serious than commission errors (false negatives). Therefore, the species-specific *minimum training presence threshold* (Section 3.4.4) was applied to convert the continuous probabilities into binary presence-absence masks. This threshold uses the lowest predicted value associated with an observed presence record (Aranda and Lobo, 2011) and guarantees that all presences are predicted as being suitable (Pearson et al., 2007). For each model replicate, all cells with modeled logistic probability of occurrence below this threshold value were classified as *absence* and vice versa as *presence*. Finally, all cells consistently predicted as *presence* by each of the five model replicates were classified as the composite climatic range for each species.

Biotope suitability (REMOTE). Analogous (i.e. using the same settings) to the climatic range models, *Maximum Entropy* models were run based on only the selected, non-correlated remote sensing predictors (REMOTE). The modeled probabilities of species occurrence based on either climate or remote sensing data are exemplarily shown for *Bursera simaruba* in Figure 6.4. According to this comparison, several landscape features which have an impact on the probability of presence for the target species (*Bursera simaruba* typically occurs in tropical evergreen forests) could be identified by means of remote sensing but not based on climate data. These landscape features with a reduced probability of species presence included, among other things, human-induced land surface elements such as agricultural areas or urban settlements but also the presence of other natural vegetation (e.g. mangroves or wetlands).

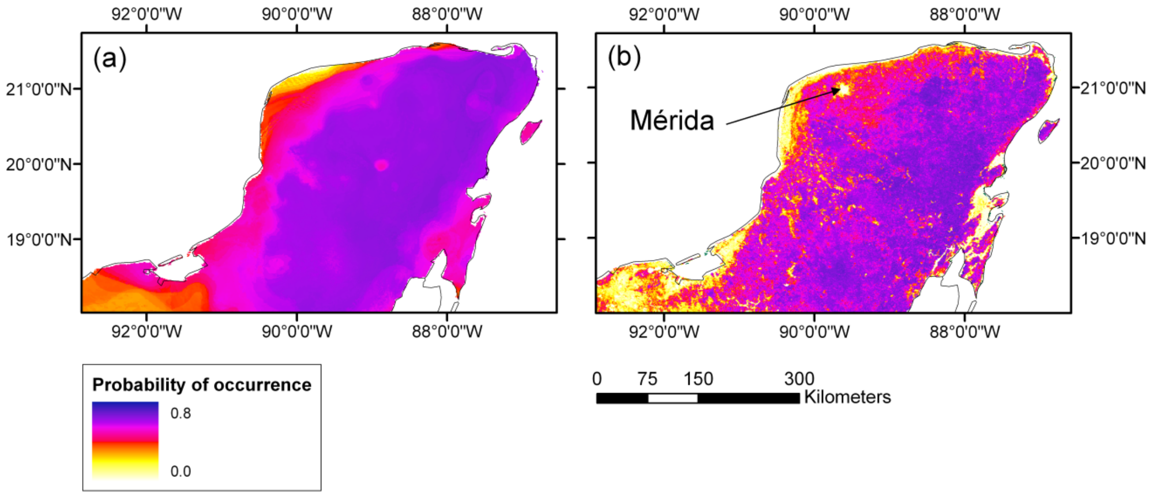


Figure 6.4: Species distribution models for *Busera simaruba* using (a) climate and (b) remote sensing data (Map section: *Yucatan* peninsula). Areas not predicted as being suitable by the remote sensing data based models include the city of Mérida and other towns, large agricultural areas in the southwest, the coastal strip and the *Sian Ka'an Biosphere Reserve* along the *Caribbean Coast* which is largely covered with mangroves and wetlands. In addition, the major roads typically surrounded by settlements and arable land can be identified.

For all species, both climatic and remote sensing based models predicted higher probabilities of species occurrence for presence than for pseudo-absence sites as exemplarily shown for four species covering a wide range of sample sizes (Figure 6.5). Despite the high correlation of both model predictions (Spearman's r > 0.75 for all target species, data not shown in Figure 6.5), for certain sites the climatic model predicted high values while the remote sensing based model predicted low probabilities (and vice versa) indicating the complementary information content of both datasets. This supports the usefulness of the hierarchical modeling framework, where the prediction of the models developed from remote sensing data was used to capture fine-scale biotope suitability for each tree species within its bioclimatic range.

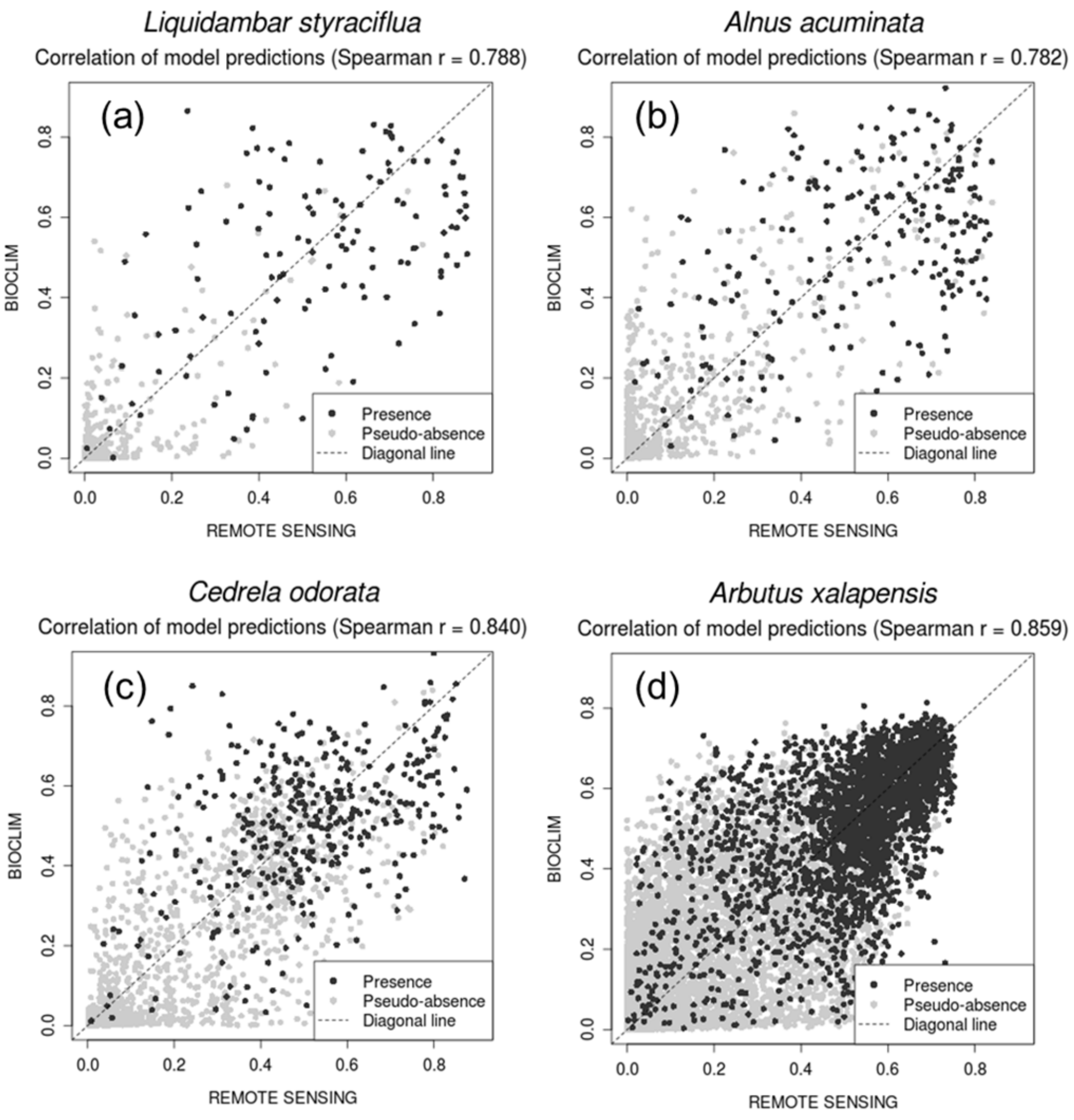


Figure 6.5: Correlation (Spearman's r) of climatic and remote sensing based model predictions for presence and pseudo-absence sites of (a) *Liquidambar styraciflua* (127 presence sites), (b) *Alnus acuminata* (229), (c) *Cedrela odorata* (381), and (d) *Arbutus xalapensis* (2,530). Note: For illustration purposes, for each species twice the number of presence records was randomly drawn from all pseudo-absence localities.

6.2.5 Hierarchical model combination

For each species, climatic and remote sensing based models were combined (BIOCLIM_REMOTE) by using the climatic distribution range (BIOCLIM_RANGE) as a presence-absence mask to identify regions with both suitable climate and suitable land surface characteristics (REMOTE) for the given species. Areas within the predicted climatic range were refined by assigning the values of the biotope suitability model to obtain spatially explicit maps of probability of occurrence (Figure 6.6). The hierarchical model combination allowed for the exclusion of areas with high modeled probabilities of occurrence based only on remote sensing data but with climatic conditions not suitable for species survival (Figure 6.6c and d).

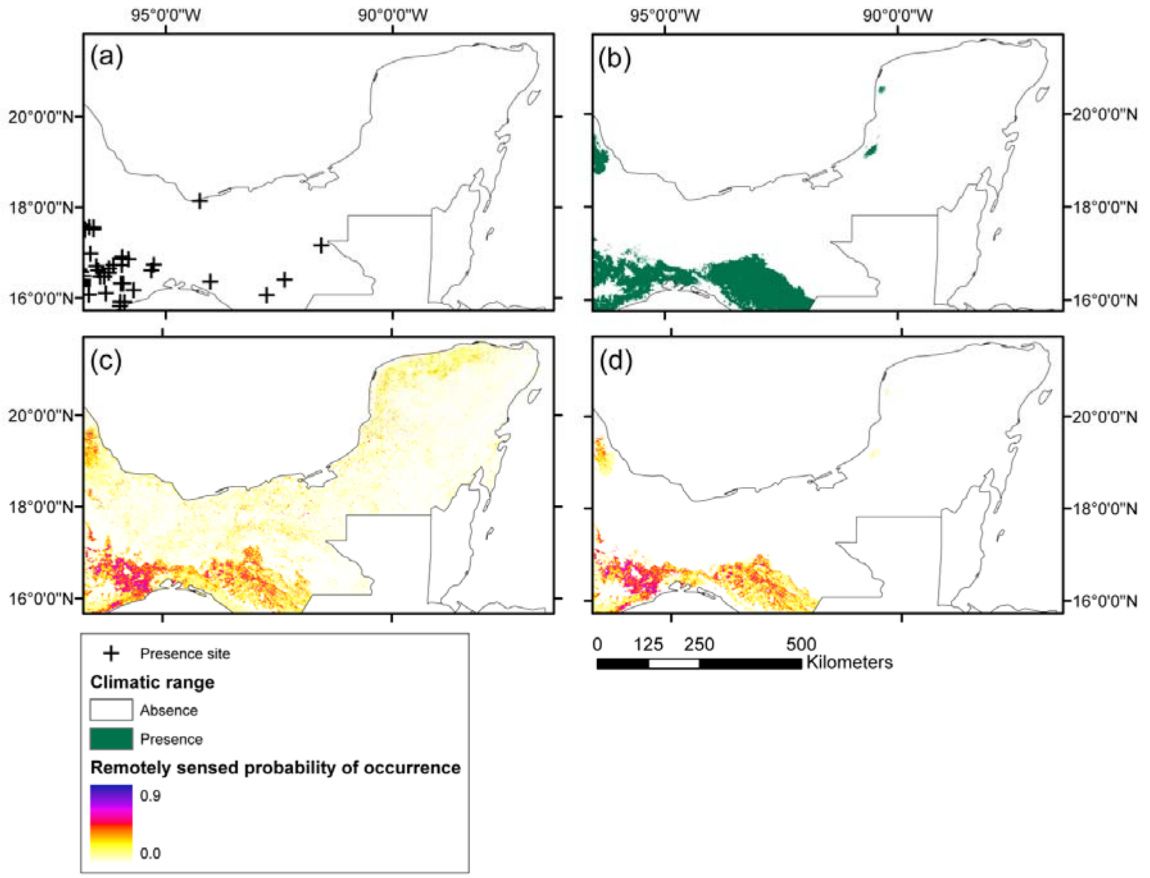


Figure 6.6: Intermediate steps and final result of the hierarchical modeling framework for *Bursera bipinnata*, here illustrated for the Yucatan peninsula. (a) Locations of presence sites, (b) Climatic distribution range (BIOCLIM_RANGE), (c) Biotope suitability based on remote sensing data (REMOTE), and (d) Result of the hierarchical combination of climatic and remote sensing based distribution models (BIOCLIM_REMOTE).

As illustrated in Figure 6.6, the hierarchical framework permitted the modeling of species distributions with higher spatial detail and the exclusion of areas which were not predicted suitable by either climate or remote sensing data. To assess predicted species range sizes of the different model approaches, also the biotope suitability layer (REMOTE) and the combined model prediction (BIOCLIM_REMOTE) were converted to binary presence-absence maps using the *minimum training presence threshold* as described above for the climatic suitability (BIOCLIM). The *fractional predicted area* (FPA, Section 3.4.5) and hence range size differed largely between both the different binary model predictions and between species (Table 6.3). For all species, FPA was lowest for the BIOCLIM_REMOTE model (mean 13.7%), medium for the REMOTE model (19.2%), and highest for the BIOCLIM model (19.7%). FPA showed a logarithmic correlation with sample size for all three model predictions (BIOCLIM_REMOTE: $r^2 = 0.87$; REMOTE: $r^2 = 0.86$; BIOCLIM: $r^2 = 0.82$). The percentage reduction in FPA due to the inclusion of remote sensing data (Table 6.3) was lowest for the most widespread species *Bursera simaruba* (13.5%) and highest for the mangrove species *Avicennia germinans* with very specific biotope requirements (53.4%).

Table 6.3: Species distributional range sizes measured by the *fractional predicted area* (FPA) of the BIOCLIM_RANGE, REMOTE_RANGE, and BIOCLIM_REMOTE_RANGE models (lowest values per species are printed in bold, highest in italics). The last column indicates the percentage decrease of FPA from climatically suitable (BIOCLIM_RANGE) to currently available biotopes after the inclusion of remote sensing data (BIOCLIM_REMOTE_RANGE).

Species	***BIOCLIM_ RANGE***	***REMOTE_ RANGE***	***BIOCLIM_REMOTE_ RANGE***	***% reduction***
Abies religiosa	0.018	*0.031*	**0.010**	44.4
Alnus acuminata	*0.141*	0.113	**0.074**	47.5
Arbutus xalapensis	0.449	*0.519*	**0.335**	25.4
Astronium graveolens	0.252	*0.264*	**0.213**	15.5
Avicennia germinans	*0.031*	0.022	**0.006**	80.7
Bursera bipinnata	0.177	*0.200*	**0.122**	31.1
Bursera simaruba	*0.517*	0.501	**0.447**	13.5
Cedrela odorata	*0.423*	0.255	**0.247**	41.6
Guaiacum sanctum	*0.123*	0.122	**0.067**	45.5
Liquidambar macrophylla	*0.092*	0.090	**0.047**	48.9
Liquidambar styraciflua	0.103	*0.107*	**0.048**	53.4
Pinus chiapensis	0.039	*0.077*	**0.025**	35.9
Mean	0.197	0.192	0.137	40.3

As threshold-independent measure (see Section 3.4.5), the statistical model deviance from observed presence-absence data (Table 6.6) was calculated as implemented in the R package ‘dismo’ (R Development Core Team, 2011). It was compared between the purely climatic (BIOCLIM), the remote sensing based (REMOTE), and the hierarchical (BIOCLIM_REMOTE) modeling approaches. Model deviance was lowest for the majority of species (except *Abies religiosa*, *Arbutus xalapensis*, and *Bursera simaruba*) for the BIOCLIM_REMOTE model (mean = 0.243), followed by BIOCLIM (mean = 0.277) and REMOTE (mean = 0.286). Besides these systematic differences dependent on the predictor set, model deviance generally varied more between species than between predictor sets for the same species. The observed variation in statistical model deviance was related to the number of presence records available per species. Model deviance showed a logarithmic increase upon higher sample sizes (Figure 6.7, $r^2 = 0.85$). Wide-ranging species with higher sample sizes could hence be modeled less accurately.

Table 6.4: Statistical model deviance for the climatic (BIOCLIM), remote sensing based (REMOTE), and hierarchical model (BIOCLIM_REMOTE) approaches. Highest scores for each species are printed in bold. The last column indicates the percentage decrease of model deviance from climatic suitable (BIOCLIM) to currently available biotopes after the inclusion of remote sensing data (BIOCLIM_REMOTE).

Species	***BIOCLIM***	***REMOTE***	***BIOCLIM_REMOTE***	***% reduction***
Abies religiosa	**0.042**	0.056	0.045	-7.1
Alnus acuminata	0.221	0.252	**0.201**	9.0
Arbutus xalapensis	**0.588**	0.615	0.605	-2.9
Astronium graveolens	0.369	0.314	**0.308**	16.5
Avicennia germinans	0.056	0.042	**0.027**	51.8
Bursera bipinnata	0.218	0.249	**0.191**	12.4
Bursera simaruba	**0.656**	0.661	0.659	-0.5
Cedrela odorata	0.491	0.429	**0.425**	13.4
Guaiacum sanctum	0.174	0.171	**0.137**	21.3
Liquidambar macrophylla	0.128	0.180	**0.119**	7.0
Liquidambar styraciflua	0.147	0.124	**0.100**	32.0
Pinus chiapensis	0.231	0.335	**0.102**	55.8
Mean	0.277	0.286	0.243	17.4

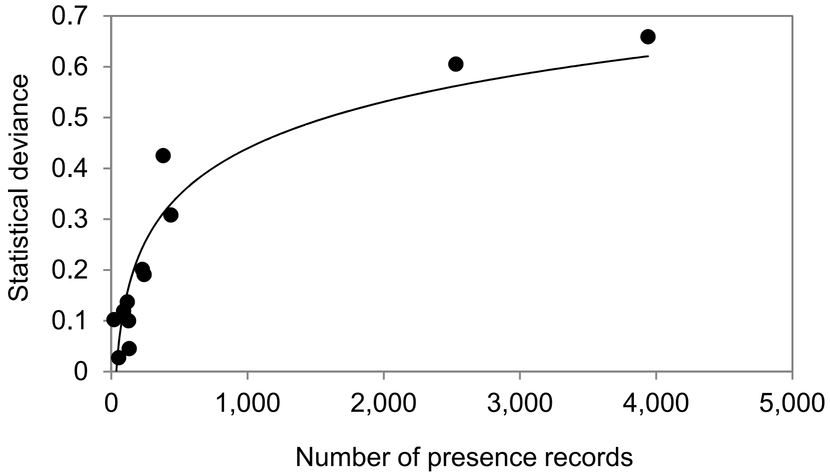


Figure 6.7: Logarithmic correlation between the statistical model deviance and the number of presence records per species ($r^2 = 0.85$).

6.2.6 Discussion and implications

Spatial autocorrelation. The analysis showed that climate variables (i.e. interpolated data) and remote sensing data (i.e. geo-referenced measurements) exhibited different strengths of spatial autocorrelation (Section 6.2.2) due to their distinct sample densities. Similar disparities are anticipated also for other geographical regions. The effect may be even stronger for areas with very few climate stations in relation to the comparatively high number of station data available for Mexico in the WorldClim data set (Hijmans et al., 2005). Given a constant number of climate stations, spatial autocorrelation is expected to further intensify in climate data with higher spatial resolutions due to increasing numbers of raster cells between the respective climate stations. In turn, spatial autocorrelation in remote sensing data is very much depending on the real landscape structure and the environmental heterogeneity. While spatial autocorrelation typically increases with higher resolution of the remote sensing data over homogeneous landscapes, it is expected to decrease in heterogeneous, small-scale structured landscapes. In an ideal situation (minimum spatial autocorrelation), each landscape element would be represented by an individual pixel of the remote sensing data. Similar to environmental predictors, also species occurrence data typically feature spatial autocorrelation, e.g. due to spatially biased sampling (Section 3.4.1). Test models developed in this thesis which integrated climate and remote sensing data in a non-hierarchical modeling framework systematically overestimated the importance of climate data for all species. This predisposed variable selection to predictors with higher SAC is known from other ecological models and has been termed the ‘red-shift’ (Lennon, 2000). A non-hierarchical model design would therefore not fully exploit the potential of the remote sensing variables.

Predictor multi-collinearity. Predictor covariation has as of yet largely been ignored for models including remote sensing data (but see Buermann et al., 2008). The results of the correlation analysis (Section 6.2.3) highlighted the significant amount of covariation within both climatic predictors and remote sensing variables and thus indicated data redundancy. High correlations between both climate data and remotely sensed phenological and statistical metrics could be explained by either the semantic or the statistical relationship between the respective variables. The predictor importance for each species (according to the predictor’s explanatory power assigned by a *Generalized Linear Model*) depending on species traits is addressed in detail in Section 6.4.

Impact of species range size. It has previously been shown that stenoecious species (with narrow or more specialized niches) and hence smaller range sizes can be better modeled based on topo-climatic data (Luoto et al., 2005; Guisan et al., 2007; Tsoar et al., 2007; Syphard and Franklin, 2010) or even based on a combination of climate and remote sensing data in a single, non-hierarchical model (McPherson and Jetz, 2007). As the Mexican *National Forest Inventory* is based on a systematic sampling design, the number of presence records per species is a direct indicator of species range size and linearly ranked between different species. The findings of this study regarding the correlation between statistical model deviance and number of presence records per species (Figure 6.7) support the conclusions of the above-mentioned studies: they are hence valid for models developed only from remote sensing data (REMOTE) and hierarchical frameworks integrating climate and remote sensing data (BIOCLIM_REMOTE) as well. Range size of the target species therefore has a strong impact on the performance of remote sensing based species distribution models and should be taken into account.

Hierarchical modeling framework. As the analysis of statistical model deviance showed (Table 6.4), the synergistic hierarchical modeling approach outperformed purely climatic and remote sensing based models in most cases. However, for three species

(25%), in particular species with very large sample sizes, climate based models showed lower deviance values than the combined model predictions. The integration of remote sensing predictors adds additional noise to the species-environment relationships since similar structural features may arise under different topo-climatic conditions and vice versa (Zimmermann et al., 2007). This might explain slightly higher values of statistical model deviance for the three species *Abies religiosa*, *Arbutus xalapensis*, and *Bursera simaruba* for the BIOCLIM_REMOTE than the BIOCLIM models as mentioned above.

The inclusion of remote sensing data generally improved the effective spatial resolution of the environmental predictors from interpolated surfaces (as for climate data) to direct geo-referenced measurements. Beyond this and analogous to the findings of Pearson et al. (2004) for land cover data, the additional use of remote sensing predictors reduced the *fractional predicted area* (Figure 6.6, Table 6.3). The inclusion of remote sensing data in the modeling framework produced particularly good results for *Avicennia germinans* (reduction of model deviance of 51.8%, Table 6.4) and *Pinus chiapensis* (55.8%) compared to climatic species distribution models. This can be explained as follows: The mangrove species *A. germinans* requires very distinctive biotope conditions in coastal wetlands which cannot be adequately characterized based on climate data. For *P. chiapensis*, only 21 presence records were collected in the *National Forest Inventory*. Five out of these samples are located in the *Sierra Madre Occidental* (Appendix I) even though the species is assumed to occur primarily in southern Mexico (Perry, 1991); they hence might represent introduced, planted or misidentified individuals. These records may lead to a too broadly defined climatic niche for *P. chiapensis* which can partly be compensated for by the integration of remote sensing data. The spatial overprediction of the REMOTE model for some species (Table 6.3) may be caused by the occurrence of similar vegetation characteristics and dynamics over different communities as well as by the non-detection of limiting factors (e.g. physiological constraints) in remote sensing data. The following sections (6.3, 6.4, and 6.5) deal with certain aspects of how to best predict biotope suitability within the proposed hierarchical modeling framework.

6.3 Categorical or continuous? – A comparison of land cover classification data and remote sensing variables to refine climatic species distribution models

The literature review (Chapter 4, in particular Sections 4.1 and 4.3) has shown the potential of land cover classification data as well as continuous remotely sensed land surface parameters for modeling species distributions. Both data sources have a different *measurement scale* (Franklin, 2009), namely *categorical* or *continuous*, as well as certain advantages and drawbacks. Their usefulness is therefore still subject to an active scientific discussion. Exemplarily, Bradley and Fleishman titled their commentary “Can remote sensing of land cover improve species distribution modelling?” (Bradley and Fleishman, 2008) and gave a collection of different examples where remote sensing or land cover data could (not) successfully been applied in species distribution modeling. This section carries out a differentiated comparison of the suitability of continuous remote sensing time series and an existing categorical land cover classification for modeling tree species distributions in Mexico.

6.3.1 Objectives

A number of different disciplines (geography, ecology, hydrology, land planning, etc.) use and refer to classified land cover (change) data as the most obvious indicator of land surface characteristics and human impact (Herold, 2009). There are currently a number of continental or global mapping activities ongoing and many land cover products are freely distributed, e.g. IGBP DISCover, GLC 2000 or GLOBCOVER (see Herold (2009) for a full description). The discrete representation of land surface characteristics in these products "has the advantages of concision and clarity" and "represents low data volume" (Lambin, 1999: p. 193). As land cover data are delivered in 'ready-to-use' raster formats including the required metadata information, ecologists increasingly integrate land cover classifications into their models. For example, Peterson et al. (2006) used historical and recent land cover data to track population extirpations of eleven endemic birds due to habitat loss in Mexico.

On the contrary, "[t]he real geographic variation is complex and continuous" (Goodchild et al., 1992: p. 90) and there is hence an increasing trend towards and interest in the direct integration of continuous spectral remote sensing data or derived vegetation indices (see Section 4.4) into species distribution models. Three main reasons suggest the use of continuous remote sensing data despite the higher pre-processing effort compared to readily available categorical land cover data, namely:

(1) Land cover mapping approaches are always designed for the **purpose of a specific study** and adapted to certain legend specifications. They may therefore not be representative and "detailed enough" (Bradley and Fleishman, 2008: p. 1158) for the target species.

(2) While conventional land cover classification data are based on discrete arbitrary classes which do not capture gradual changes in the landscape (Krishnaswamy et al., 2009), the use of remote sensing data preserves the **continuous characteristics of vegetation**.

(3) When directly using spectral remote sensing data or derived vegetation indices as model predictors, there is **no additional error** introduced as through the land cover classification process.

The objective of this study is therefore to directly compare the suitability of an existing categorical land cover classification and continuous remote sensing time series for the delineation of current biotope availability (Figure 6.8). Suitability of the two datasets to be examined is assessed based on model performance and related to the characteristics of the study species. The target tree species are selected based on their representativeness regarding different range sizes, biotope specificity, and associated vegetation types (Section 6.1). The NALCMS land cover information used (Section 5.3) is the most accurate data set currently available for Mexico and was produced based on MODIS data which were used as the continuous remote sensing variables in this analysis. The land cover data is temporally consistent with both the species occurrence data and the remote sensing data analyzed.

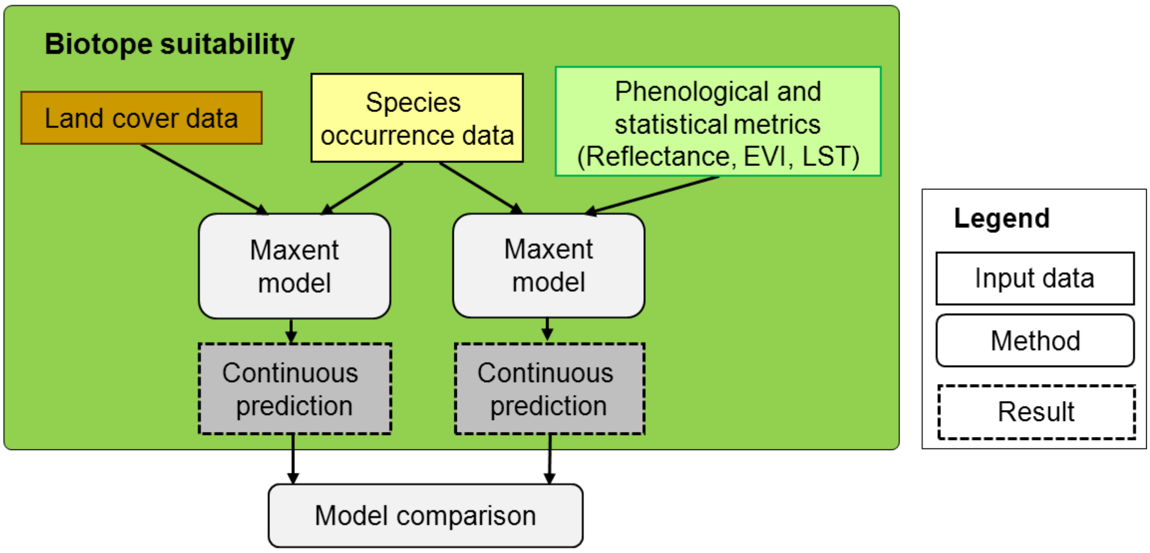


Figure 6.8: Modeling framework to compare the suitability of phenological and statistical metrics (derived from continuous remote sensing time series of *Reflectance Data*, *Enhanced Vegetation Index*, and *Land Surface Temperature*) and an existing categorical land cover classification for the delineation of biotope suitability.

6.3.2 Land cover evenness index

The suitability of land cover data for predicting species distributions is dependent on the applicability of the legend to characterize the biotope requirements of the target species. In case the locations of all occurrence records fall into the same or into very few land cover categories, this is a good indicator of the land cover data to be appropriate for the respective species. Hence, the distribution of land cover classes observed at the presence localities was assessed (results for three exemplary species are shown in Figure 6.9, plots for additional species discussed in further detail in the following can be found in Appendix J).

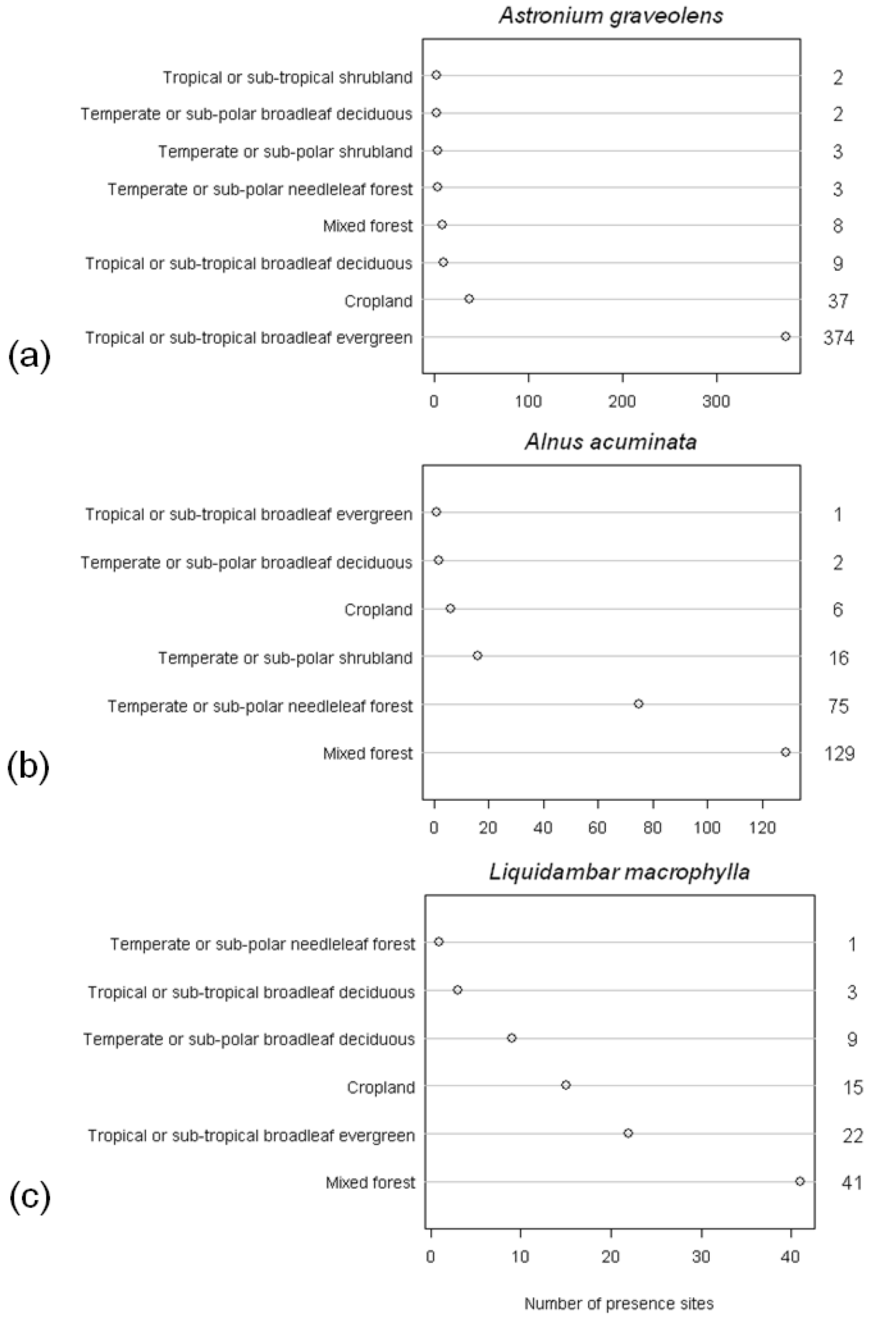


Figure 6.9: Frequency distribution of land cover classes observed at the presence localities of the study species. (a) *Astronium graveolens*, (b) *Alnus acuminata*, and (c) *Liquidambar macrophylla*. The numbers on the second y-axis on the right-hand side represent the exact number of species records per land cover class as shown on the x-axis.

To quantify the different distributions of land cover classes observed at the species presence localities (Figure 6.9), a *land cover evenness index*, analogous to a species evenness index used in biodiversity assessments, was computed for all study species introduced in Section 6.1. In this context, the *evenness* refers to how close in numbers the different land cover classes observed at the species presence locations are. Pielou's evenness index *J'* (Pielou, 1966) was used, which is the ratio of the observed Shannon diversity to the maximum diversity and given by

$$J' = \frac{H'}{H'_{max}}$$

where *H'* is the number derived from the Shannon diversity index (Shannon, 1948) and H'_{max} is the maximum value of *H'*.

H' is used to measure the diversity of categorical data and basically represents the information entropy of the distribution, treating *land cover classes* (equivalent to *species* in species diversity assessments) as symbols and their relative *frequencies* (*population sizes*) as the probability. *H'* is given by

$$\mathrm{H}' = -\sum_{i=1}^{S} (\mathrm{p_i} \ \ln \mathrm{p_i})$$

where *S* is the total number of *land cover classes* (*species*) and p_i the frequency of the *i*th *land cover class* (*species*).

$\mathrm{H'_{max}}$ is given by

$$\mathrm{H'}_{\max} = -\sum_{i=1}^{S} \frac{1}{\mathrm{S}} \ln \frac{1}{\mathrm{S}} = \ln \mathrm{S}$$

where S is again the total number of *land cover classes* (*species*).

The values of Pielou's evenness index range from 0 to 1 whereas higher values indicate that the *land cover classes* (*species*) are evenly distributed. This implies that the species records cannot be assigned to one or few land cover categories. Both Shannon diversity and Pielou evenness index were calculated as implemented in the R package 'vegan' (R Development Core Team, 2011), the results are shown in Table 6.5.

The lowest land cover evenness index scores were observed for *Astronium graveolens* (0.29; 438 presence records) and *Avicennia germinans* (0.33; 54 presence records). The geographical distributions of both species were characterized by only one dominant land cover class, namely *Tropical or sub-tropical broadleaf evergreen forest* (*A. graveolens*, Figure 6.9) and *Wetland* (*A. germinans*, Appendix J), respectively. On the contrary, the species with the highest land cover evenness index were *Liquidambar macrophylla* (91 presence records) and *Pinus chiapensis* (21 presence records). For both *L. macrophylla* and *P. chiapensis*, the land cover class *Mixed forest* was the most important category observed at their presence sites (Figure 6.9, Appendix J). Since *A. germinans* (with 54 presence records) was one of the species with the lowest land cover evenness index, the index was not simply a function of the number of presence records per species. The relationship between the land cover evenness index and the land cover class characteristics is further discussed in Section 6.3.4.

Table 6.5: Results for the *land cover evenness index*, computed analogous to Pielou's species evenness index (Pielou, 1966) based on the species occurrence localities.

Species	***Land cover evenness index***
Abies religiosa	0.53
Alnus acuminata	0.58
Arbutus xalapensis	0.52
Astronium graveolens	0.29
Avicennia germinans	0.33
Bursera bipinnata	0.69
Bursera simaruba	0.44
Cedrela odorata	0.57
Guaiacum sanctum	0.39
Liquidambar macrophylla	0.78
Liquidambar styraciflua	0.70
Pinus chiapensis	0.77

6.3.3 Model performance and spatial predictions

To assess and compare the suitability of categorical land cover and continuous remote sensing data, separate *Maximum Entropy* models based on either land cover information or remote sensing data were run (see Section 6.2.4 for the model settings). As summarized in Table 6.6, all models produced – according to the classification of Swets (1988) – 'fair' to 'excellent' model accuracies measured by AUC (*area under curve of the receiver operating characteristic*, Section 3.4.5). Both training and test AUC were higher for the remote sensing data based model except for *Pinus chiapensis* with a slightly higher test AUC score for land cover data (though with a very high standard deviation).

The direct comparison of (training) AUC scores per species for models developed from either remote sensing or land cover data is shown in Figure 6.10. For all study species, remote sensing based models revealed higher AUC scores than land cover classification based models. However, there was considerable variation in AUC scores between species. A very similar outcome was found for the statistical model deviance from reference presence-absence records (calculated as implemented in the R package 'dismo', R Development Core Team (2011); data not shown here) with higher deviance for all models developed from land cover data except for the species *Avicennia germinans*.

Table 6.6: Comparison auf AUC and standard deviation (SD) scores of the land cover and remote sensing based Maxent models. Training AUC describes the ability of the model to correctly score a randomly selected presence site higher than a randomly selected background site, test AUC describes the model's ability to predict a fraction of the species records omitted during model training (Phillips et al., 2006; Section 3.4.5). Highest scores for each species are printed in bold.

	Land Cover				***Remote Sensing***			
	Training		***Test***		***Training***		***Test***	
Species	***AUC***	***SD***	***AUC***	***SD***	***AUC***	***SD***	***AUC***	***SD***
Abies religiosa	0.888	0.011	0.886	0.013	**0.991**	0.002	**0.981**	0.003
Alnus acuminata	0.819	0.010	0.804	0.020	**0.955**	0.003	**0.928**	0.019
Arbutus xalapensis	0.832	0.005	0.834	0.004	**0.875**	0.002	**0.855**	0.003
Astronium graveolens	0.871	0.009	0.858	0.013	**0.938**	0.003	**0.919**	0.009
Avicennia germinans	0.979	0.005	0.963	0.037	**0.988**	0.005	**0.976**	0.015
Bursera bipinnata	0.852	0.014	0.839	0.024	**0.964**	0.002	**0.942**	0.009
Bursera simaruba	0.824	0.002	0.821	0.004	**0.866**	0.002	**0.857**	0.004
Cedrela odorata	0.821	0.007	0.809	0.014	**0.917**	0.006	**0.871**	0.002
Guaiacum sanctum	0.864	0.010	0.876	0.028	**0.971**	0.004	**0.938**	0.015
Liquidambar macrophylla	0.739	0.013	0.749	0.054	**0.957**	0.007	**0.935**	0.031
Liquidambar styraciflua	0.745	0.020	0.741	0.035	**0.977**	0.003	**0.949**	0.020
Pinus chiapensis	0.820	0.039	**0.802**	0.140	**0.932**	0.010	0.801	0.059
Mean	**0.838**	**0.012**	**0.832**	**0.032**	**0.944**	**0.004**	**0.912**	**0.016**

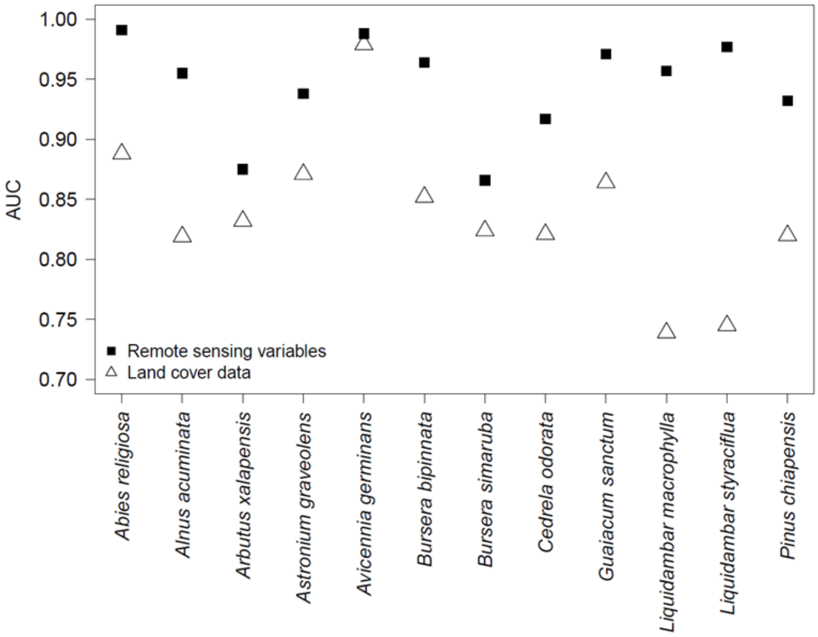


Figure 6.10: Comparison of (training) AUC scores for different species distribution models based on either continuous remote sensing variables or categorical land cover classification data. Training AUC describes the ability of the model to correctly score a randomly selected presence site higher than a randomly selected background site.

The difference between training AUC scores based on Maxent models developed from either remote sensing variables or land cover classification data could be related to the land cover evenness index score computed for each species. Figure 6.11 shows this linear relationship ($r^2 = 0.54$) between the improvements in AUC scores due to the use of remote sensing data and the corresponding land cover evenness index. A high land cover evenness index (e.g. for *Liquidambar macrophylla*, indicating an even proportional distribution of the species records on the different land cover classes) was hence related with a higher increase in model performance (AUC) due to the use of remote sensing instead of categorical land cover data (Figure 6.11). On the other hand, a low land cover evenness index (e.g. for *Astronium graveolens* and *Avicennia germinans*, see Table 6.5) was linked with small model improvements. The land cover evenness index was thus a direct indicator of the expected enhancement of species distribution models by replacing land cover classification data by continuous remote sensing variables.

In addition, a linear correlation ($r^2 = 0.52$ for training AUC, $r^2 = 0.56$ for test AUC) between AUC scores of the Maxent model developed from land cover data and the species-specific land cover evenness index was found. A low land cover evenness index therefore gives reason to expect high AUC scores for the respective land cover based model. On the contrary, no correlation was found between AUC scores of the Maxent models developed from remotely sensed variables and the land cover evenness index ($r^2 = 0.00$).

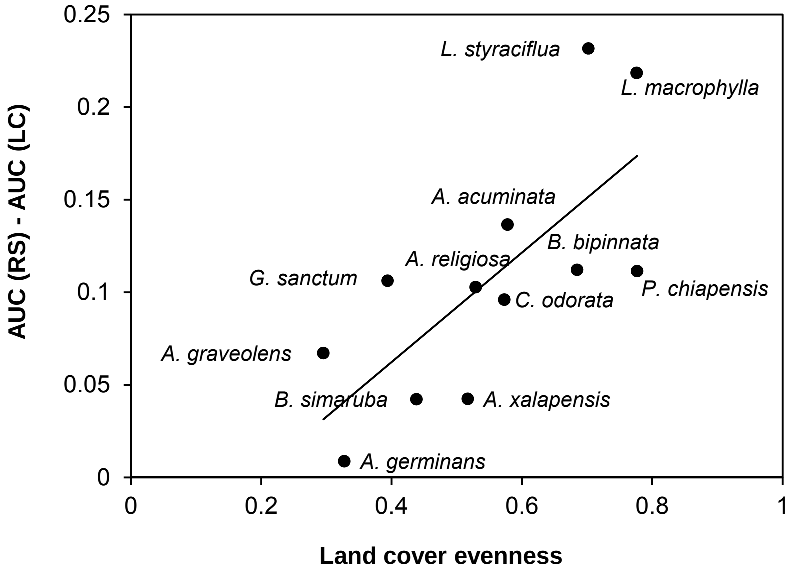


Figure 6.11: Relationship between model improvement measured by the increase in AUC scores (y-axis) and the land cover evenness index (x-axis) calculated for each species (Pearson's $r^2 = 0.54$, $y = 0.29x-0.06$).

Spatially explicit Maxent model predictions for *Astronium graveolens*, *Alnus acuminata*, and *Liquidambar macrophylla* as examples of species with low, medium, and high land cover evenness index are shown in Figure 6.12. Modeled distribution patterns were similar between models developed from land cover and remote sensing data for

Astronium graveolens, the species with the lowest land cover evenness index. This is consistent with the comparatively low difference in training AUC between land cover and remote sensing based models for this species (Figure 6.10). For *Liquidambar macrophylla*, considerable differences in the modeled probabilities of occurrence were observed, with a clear tendency of the land cover data based models to overestimate the species distribution range. Again, this overestimation was reflected in the distinct difference in training AUC scores for *L. macrophylla* (Figure 6.10). For *Alnus acuminata*, both the spatial agreement of remote sensing and land cover based models and the difference in training AUC values were intermediate, i.e. lower than for *Astronium graveolens* but higher than for *Liquidambar macrophylla*. Differences in spatial predictions between models developed from either remote sensing or land cover data could be related to improvements in AUC scores for the other species as well (data not shown).

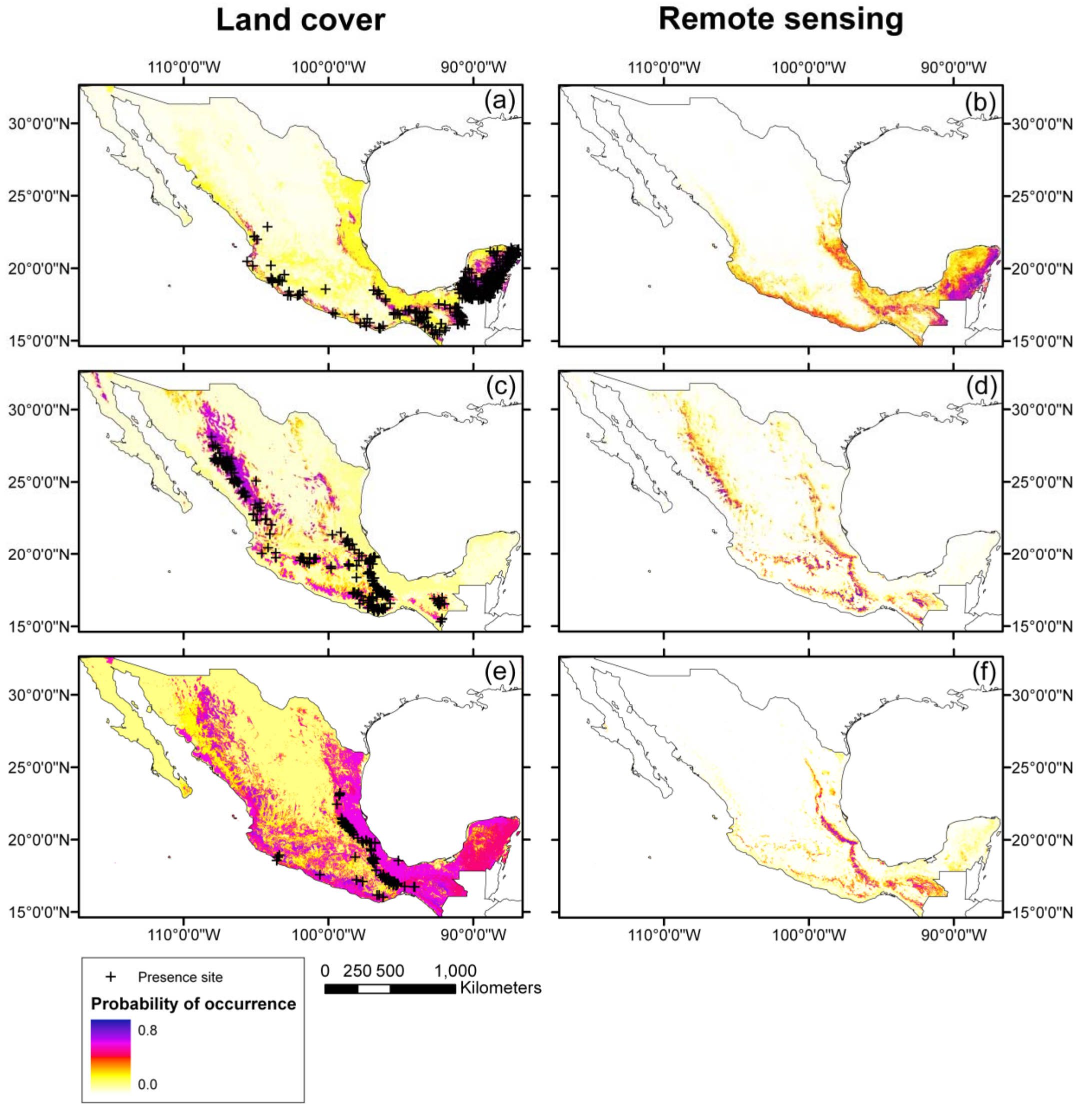


Figure 6.12: Modeled probability of species occurrence based on land cover data (left column) and remote sensing time series (right column) of (a) and (b) *Astronium graveolens* (land cover evenness index 0.29), (c) and (d) *Alnus acuminata* (evenness 0.58), and (e) and (f) *Liquidambar macrophylla* (evenness 0.78). Black crosses indicate species presence sites.

6.3.4 Discussion and implications

Trend towards overestimation of species distributions with categorical data. Models based on land cover data tended to overestimate species distribution ranges as no continuous geographic variation (Goodchild et al., 1992) or floristic gradients (Feilhauer et al., 2011) are represented or evident from these categorical data. In addition, the discrete representative of land surface characteristics in categorical land cover data is stable over a season (Lambin, 1999) whereas the analysis of phenological and statistical time series metrics allows for considering temporal dynamics. In addition, for species that do not occupy all suitable areas for any reason, land cover maps might predict only potential rather than actual species distributions (Kerr and Ostrovsky, 2003; Section 3.1.4). Land cover data further typically suffer from cartographic generalization and often lack sufficient spatial (Thuiller et al., 2004; Luoto et al., 2007) and thematic (Kerr and Ostrovsky, 2003) detail. For example, the difference in model training AUC between remote sensing and land cover based models was highest for the two species of the cloud forest (*Liquidambar* spp., Figure 6.10) which is not included as separate land cover class in the legend of the NALCMS data set (see Section 5.3). In this case, the failure of land cover data to model the species distributions is hence an indicator of insufficient thematic detail. Therefore, the suitability of each land cover product to predict species distributions is based on the detail (number of classes) and validity (significance of the class definitions) of its legend as well as its mapping accuracy. The NALCMS land cover information used here is the most accurate data set currently available for Mexico with an overall accuracy of 82% (Colditz et al., 2010).

Importance of species characteristics. The general trend towards overestimation of species distribution ranges based on categorical land cover information was found to be influenced as well by the specific characteristics of the target species. The performance of land cover data for modeling species distributions depends on how closely the spatial distribution patterns of a species can be linked to certain land cover types. Typically, species occur in more than one land cover class and may further be distributed patchily within each of the respective classes (Ferrier, 2002).

The species-land cover relationship can be characterized by the following crucial points which have implications regarding the usefulness of categorical land cover data for modeling species distributions. The suitability of a certain land cover product is dependent on (1) the land cover evenness index score based on the land cover categories observed at the species presence sites, (2) the proportions of the study area that are covered by the respective most important land cover class(es), and (3) the mapping accuracy of the dominant land cover class(es) observed at the majority of species presence sites.

In the following, these three points are exemplarily discussed in the light of the findings of this study. The geographical distributions of *Astronium graveolens* and *Avicennia germinans*, the two study **species with the lowest land cover evenness index**, were characterized by only one dominant land cover class, namely *Tropical or sub-tropical broadleaf evergreen forest* (*A. graveolens*, Figure 6.9) and *Wetland* (*A. germinans*, Appendix J), respectively. The proportions of both land cover classes in relation to the Mexican land surface were 7.2% (*Tropical or sub-tropical broadleaf evergreen forest*) and 1.0% (*Wetland*), respectively. Both species also considerably differed in model training AUC for the land cover based model (*A. graveolens*: 0.871; *A. germinans*: 0.979, Table 6.6). Consistent with the findings of Hernandez et al. (2006) and McPherson and Jetz (2007) for topo-climatic species distribution models, this difference in model performance was presumably a consequence of the more limited geographical range of *A. germinans* indicated by the lower number of presence records and the smaller proportional area of its associated land cover type *Wetlands* in the NALCMS dataset. Most importantly, the increase in model performance measured by AUC scores due to the use of remote sensing instead of land cover data was several times higher for *A. graveolens* than for *A. germinans* (Figure 6.10). This was mainly a consequence of the different mapping accuracies of the respective land cover classes which were 74.6% for *Tropical or sub-tropical broadleaf evergreen forest* and 96.4% for *Wetlands* (User's Accuracy, Colditz et al., 2010). Coastal wetlands had a spectral signature which was distinct from the other terrestrial land cover classes in Mexico and could thus reliably be mapped in the land cover classification procedure.

The species with the **highest land cover evenness index** were *Liquidambar macrophylla* (91 presence records, Table 6.1) and *Pinus chiapensis* (21 presence records). For both *L. macrophylla* and *P. chiapensis*, the land cover class *Mixed forest* was the most important category observed at their presence sites (Figure 6.9, Appendix J). Consistent with the results for *A. graveolens* and *A. germinans* discussed above as well as with previous studies (Brotons et al., 2004; Luoto et al., 2005; Tsoar et al., 2007), model performance in the land cover based model was higher for *P. chiapensis* (Table 6.6), the species with the lower prevalence . The obvious increase in model training AUC due to the use of remote sensing instead of land cover data for both species (Figure 6.10) was presumably the result of the comparatively low mapping accuracy of the class *Mixed Forest* (Producer's Accuracy: 80.1%, User's Accuracy: 62.9%; Colditz et al., 2010) and the insufficient significance of the class definitions for the target species. *L. macrophylla* occurs in tropical montane cloud forests which are not represented in the legend of the NALCMS land cover product but can be characterized based on continuous multi-temporal remote sensing data.

Implications for future studies. Certainly, since the pre-processing effort of remote sensing data is high compared to often readily available categorical land cover data, a trade-off situation is created between target model accuracy and processing effort. Based on the findings of this study, the following guidelines can be given: Models developed from remote sensing data in general outperformed land cover based models as measured by AUC scores and statistical deviance from presence-absence data. Typical indicators of a high suitability of land cover data to model the distribution of a target species are:

(1) a **low land cover evenness index** computed for the presence localities,

(2) a **low proportion of the study area** that is covered by the dominant land cover class(es) observed at the presence sites of each species,

(3) and a **high mapping accuracy** of the dominant land cover class(es) and the land cover product in general as indicated in the metadata.

To summarize, it is difficult to give a general answer whether to use land cover or remote sensing data for modeling biotope suitability, but the choice needs to be made case-by-case by the respective researcher, depending on the linkage between species records and land cover as well as the intended effort for data processing. The findings of this study, based on the application of the novel land cover evenness index, can provide the theoretical knowledge to decide which data set (land cover classification or remote sensing variables) is more suitable for modeling. In the following, the relevance of selected remote sensing products for modeling species distributions is assessed whilst taking into account species characteristics.

6.4 Which remote sensing data for which species? – The impact of species traits on selecting remote sensing predictors

Remotely sensed time series offer a variety of original or derived informative layers that can be used to describe vegetation characteristics; the selection of adequate variables is therefore crucial. The importance of a reasonable selection of predictors derived from MODIS time series has already been shown in vegetation mapping (Hüttich et al., 2009). In the context of species distribution modeling of plants, variable importance of remote sensing predictors has been considered only rarely. For example, Saatchi et al. (2008) identified the predictors (including MODIS data) with the highest importance for the distribution of five tree species in the Amazon basin. However, due to relatively similar biotope requirements of their target species, no clear trend in variable importance between species was recognized. In addition, Tuanmu et al. (2010) discussed differences in the values of phenological metrics (derived from MODIS data) between forests with and without the two target understory bamboo species. However, they did not relate these findings to the relevance of the predictors in the species distribution models. Therefore, despite the above-mentioned growing interest in understanding the importance of remotely sensed variables for modeling species distributions, no systematic assessment of predictors suitable for modeling plant species with certain traits exists so far.

6.4.1 Objectives

The aim of this study was to systematically examine the importance of different explanatory variables for the twelve target tree species in an ecological context (Figure 6.13). All parameters used here are suitable to characterize vegetation distribution as outlined in Section 4.4. This analysis thus focuses on the selection of the *most* suitable variable for a certain (group of) species. Especially considering the temporal and computational effort when dealing with remotely sensed time series, the study aims to highlight the major criteria for selecting remote sensing products as well as certain categories of phenological metrics depending on the biotope characteristics of the target species. The approach further aims to link recurring patterns of predictor importance to the characteristics of the respective vegetation community (e.g. tropical compared to temperate climates; deciduous as opposed to evergreen vegetation communities).

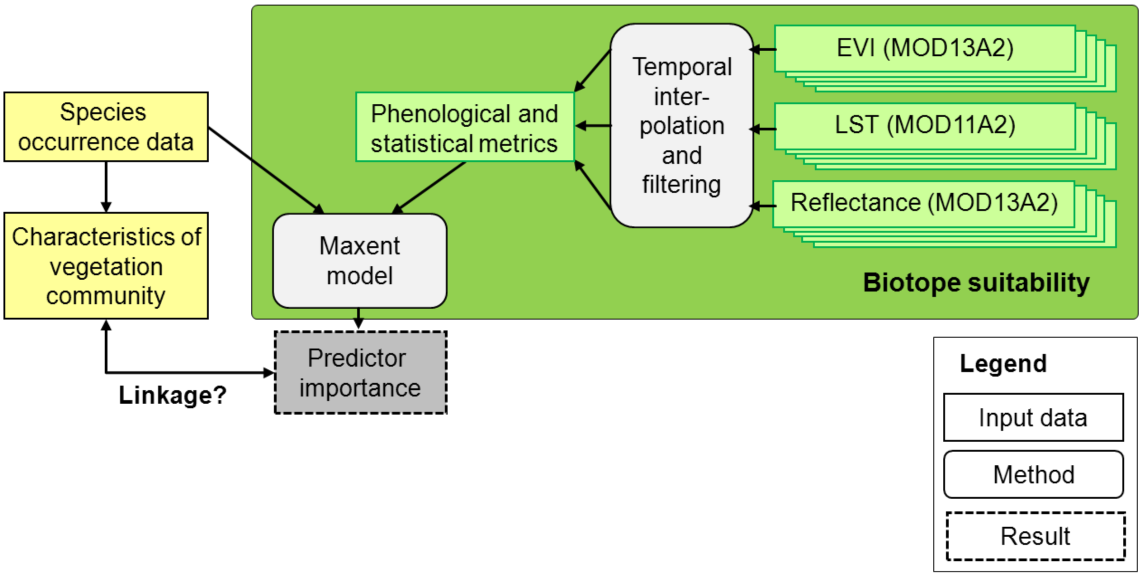


Figure 6.13: Modeling framework to examine the importance of different remotely sensed explanatory variables for modeling tree species in an ecological context. Recurring patterns of predictor importance are linked to the characteristics of the respective vegetation communities.

6.4.2 Species-specific variable importance

While the Maxent model is being trained, it keeps track of which internal features are contributing to fitting the model (Phillips, 2009). This increase in model gain is assigned to the environmental variable(s) that the feature depends on; it was used to assess predictor importance in this study (see Section 3.3.2 for a detailed description of how model gain and variable importance are computed in Maxent). As only non-correlated predictors based on their explanatory power according to a *Generalized Linear Model* were used here (see Section 6.2.3), variable importance scores could be directly compared among each other. Variable importance in this approach was determined by the contribution of each predictor to distinguish between species presence records and background sites. It is thus different from the techniques typically used in vegetation classification, where the multivariate distance between pairs of vegetation classes is measured by e.g. the Mahalanobis distance (Mahalanobis, 1936) or the Bhattacharyya distance (Bhattacharyya, 1943).

The results, namely the percentage contributions per species (mean scores across all model replicates) measured by the increase in regularized model gain, are shown in Table 6.7. Accordingly, species occurring in temperate or tropical montane climates (*Abies religiosa*, *Alnus acuminata*, *Arbutus xalapensis*, *Liquidambar macrophylla*, *Liquidambar styraciflua*, and *Pinus chiapensis*) were best predicted using *Land Surface Temperature* (LST) metrics, in particular mean, maximum, and minimum LST values. Reflectance data (*MIR_minimum*) had only a major explanatory power for one species, the mangrove *Avicennia germinans*, which typically occurs in coastal wetlands. The corresponding biotope features spectral characteristics distinct from the terrestrial background sites (see also Section 4.4.1). All species of the tropical evergreen forest (*Astronium graveolens*, *Bursera simaruba*, and *Cedrela odorata*) were best predicted by NPP-related phenological metrics, in this case consistently annual mean EVI across all species. Metrics related to vegetation seasonality were of major importance only for *Bursera bipinnata* occurring in tropical deciduous forests which exhibit a distinct phenological cycle between dry and rainy season. For this species, the annual standard deviation of EVI was the most important model predictor. Temporal EVI metrics comprised information relevant for modeling the distribution of all target species, albeit with comparatively low percentage contributions. Out of the variety of candidate remote sensing predictors, only few metrics appeared to be highly relevant for the species analyzed in this thesis, namely *LST_maximum*, *LST_mean*, *LST_minimum*, *LST_CoV*, *EVI_standard_deviation, EVI_mean*, and MIR_minimum. For the most important predictor variable per species, response curves were computed and analyzed as shown in the following section.

Table 6.7: Percentage contributions of remote sensing variables to the Maxent models measured by the increase in regularized model gain (mean scores across all model replicates). Notes: Explanations of the phenological and statistical metrics are given in Appendix H. Variable contributions higher than 40% are printed in bold. The number of predictors in each category of phenological and statistical metrics is indicated by N (see Section 5.4.4).

Species	***Climate***	***Vegetation type***	***EVI: Temporal (N=5)***	***EVI: NPP-related (N=8)***	***EVI: Seasonality-related (N=5)***	***LST (N=7)***	***Reflectance (N=28)***	***Most important metric***
Abies religiosa	temperate	Temperate needle leaved evergreen forest	3.1	0.0	9.2	**72.9**	14.8	LST_mean
Alnus acuminata	temperate	Temperate deciduous-evergreen forest	6.1	0.0	11.5	**71.3**	11.2	LST_mean
Arbutus xalapensis	temperate	Temperate deciduous-evergreen forest	18.6	0.0	11.1	**56.1**	14.2	LST_minimum
Astronium graveolens	tropical	Tropical or sub-tropical evergreen forest	3.9	**40.8**	6.7	17.0	31.6	EVI_mean
Avicennia germinans	tropical	Wetlands	17.5	0.0	4.7	12.1	**65.7**	MIR_minimum
Bursera bipinnata	tropical	Tropical or sub-tropical deciduous forest	18.1	2.2	**48.3**	15.3	16.2	EVI_standard_deviation
Bursera simaruba	tropical	Tropical or sub-tropical evergreen forest	4.4	**68.8**	0.7	23.0	3.1	EVI_mean
Cedrela odorata	tropical	Tropical or sub-tropical evergreen forest	3.7	**63.7**	3.8	6.9	22.0	EVI_mean
Guaiacum sanctum	tropical	Tropical or sub-tropical deciduous-evergreen forest	12.2	0.0	8.7	**56.1**	23.1	LST_CoV
Liquidambar macrophylla	tropical montane	Moist montane/cloud forest	23.2	0.0	4.4	**71.1**	1.2	LST_maximum
Liquidambar styraciflua	tropical montane	Moist montane/cloud forest	7.1	0.0	4.5	**70.4**	18.0	LST_maximum
Pinus chiapensis	tropical montane	Temperate needle leaved evergreen forest	10.0	0.0	5.0	**66.6**	18.4	LST_mean

6.4.3 Species response curves

While variable importance is a global measure of the explanatory contribution of each model predictor, species response curves allow for a more detailed insight how each environmental variable affects the model prediction (see Section 3.1.2 for the theoretical background of environmental gradient analysis and species response curves). Here, the species response curves (Figure 6.14) show how the logistic probability of occurrence predicted by the Maxent model changes as each environmental variable is varied, while all other environmental variables are kept at their average sample value (Phillips, 2009). The response curve shown here is thus slightly different from the species response computed by the Maxent model itself, as the model may take advantage of sets of variables changing together (Phillips et al., 2006). In other words, the curves show the effect of altering exactly the target variable while all other predictors remain at their mean value across all sample sites (Phillips, 2009). The prior removal of highly correlated predictors (see Section 6.2.3) supports the direct interpretation of the response curves; thereby less inherent correlations, which are not evident in the curves, affect the *Maximum Entropy* model itself.

The response curves of the target species (Figure 6.14) indicate the position of the ecological optimum along the respective environmental gradient. All target species – except *Pinus chiapensis* with an almost flat response – showed determinate (i.e. unimodal or monotone) response curves. *Pinus chiapensis* exhibited an almost flat response curve indicative of very small changes in probability of occurrence over a broad range of mean LST scores. The probability for fitting determinate response curves typically increases with species optima being closer to gradient extremes (Rydgren et al., 2003). Some of the monotone response curves (e.g. for *Astronium graveolens*, *Avicennia germinans* or *Guiacum sanctum*) featured very high slopes of decrease near the species optimum, so-called “tapering tails” (Oksanen and Minchin, 2002: p. 127). Accordingly, the probability of occurrence near the non-optimum end of the gradient was very low and it was therefore unlikely that a presence would be observed at the real limit of distribution (Oksanen and Minchin, 2002). Some of the response curves (e.g. for *Alnus acuminata*, *Cedrela odorata* and *Liquidambar styraciflua*) also appeared to be ‘spikey’. Huntley et al. (1995) explained these spikey shapes – typically observed for realized response curves built from observation data – by interactions with other species. They also argued that conventional smooth response curves are rather related to fundamental niche conditions hardly found in nature (Huntley et al., 1995). The shape of the response curves is further influenced by the feature types chosen (see Section 3.3.2). For example, the response curve of *Arbutus xalapensis* appears like a step function presumably due to the use of threshold features.

Many of the response curves computed for the same variable featured very similar shapes. For example, the response curves of *Bursera simaruba* and *Cedrela odorata* were both truncated as annual mean EVI scores in the predictor data set asymptotically converged (the highest *EVI_mean* value computed for Mexico was 0.68). *Liquidambar macrophylla* and *Liquidambar styraciflua*, the two species known to occur in cloud forests, exhibited almost identical optimum and tolerance ranges. Just like the importance of environmental variables could be related to the ecological requirements of the species, also the particular shapes of the response curves therefore matched what is known about the particular biotopes. Remote sensing data are hence useful for describing biotope characteristics.

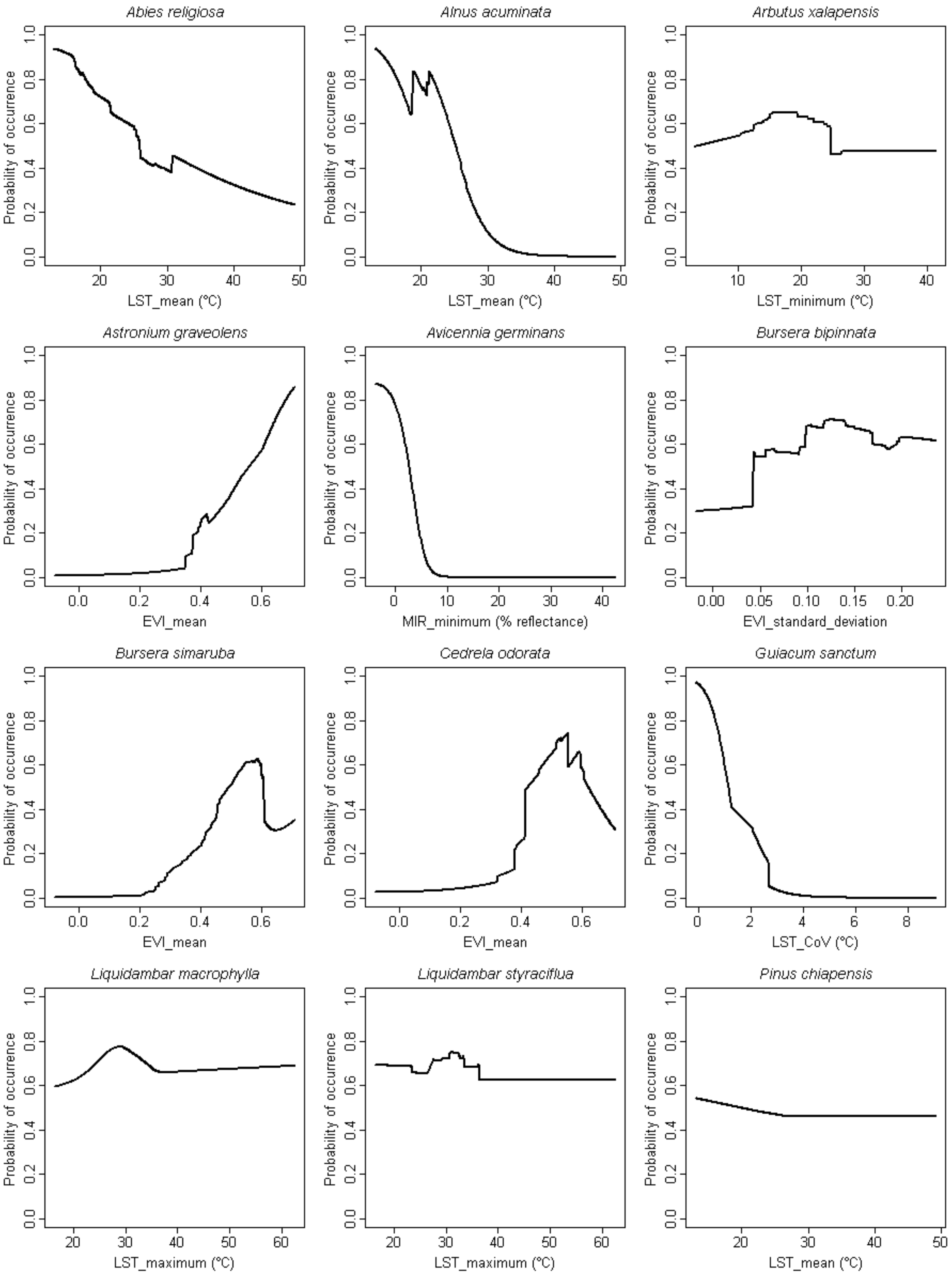


Figure 6.14: Species response curves for the most important predictor variable per species. The curves show the mean response of the five replicate model runs and illustrate how the logistic probability of occurrence predicted by the *Maximum Entropy* model (y-axis) changes as each environmental variable is varied (x-axis).

6.4.4 Discussion and implications

Variable importance. The results of the variable importance analysis showed that vegetation indices (in this case EVI) are the most widely used (see literature review in Section 4.2) but not necessarily the best predictors of species distributions. Variables derived from EVI time series were only of major importance for four out of the twelve study species, namely the species associated with tropical evergreen and tropical deciduous forests. Instead, the feature ranking also indicated a high potential of LST for modeling species distributions – in particular for temperate species. Spectral bands in turn were found to be of minor relevance except for the mangrove species *Avicennia germinans* occurring in coastal wetlands. Therefore, remote sensing products should be purposefully selected according to known biotope characteristics of the target species. On doing so, the outcomes and recommendations of previous vegetation mapping studies with a remote sensing background are very useful. For example, it was previously shown that forest exhibit lower LST scores than other vegetation in India (Parida et al., 2008). This is consistent with the recommendation given by this analysis, namely that LST is of major importance in explaining the distribution of species in temperate, evergreen forests. In addition, a high potential of spectral bands for the discrimination of vegetation classes was mainly found in landscapes with open vegetation cover, e.g. savannas (Stuart et al., 2006; Hüttich et al., 2009) or mountain rangelands (Cingolani et al., 2004). The target species of this analysis are primarily found in dense canopy forests (see Section 2.4.2) and therefore featured low importance of reflectance data for explaining their distributions. However, the *Maximum Entropy* implementation used in this thesis makes use of an internal calculation of features (i.e. interactions between environmental predictors). The high variable contributions assigned to certain predictors are hence not only a result of the variable itself but also of its interactions with other variables.

Species response curves. The analysis of species response curves for remotely sensed variables indicated mostly determinate (unimodal or monotone) response shapes. However, response curves based on remote sensing data make use of other assumptions than ecological niche theory which is relevant for species responses along climatic gradients (see Section 3.1.1). Instead, the remote sensing based response curves rather reflected the typical, remotely sensed properties at species occurrence localities. Further, it needs to be considered that the response curves (as calculated and shown in Figure 6.14) are not completely independent of the other predictors. This is consistent with ecological theory as species do not respond to single environmental factors but to complex gradients consisting of varying environmental factors (Whittaker, 1956). A general understanding of the relative importance of the factors governing the distribution of species response curves along environmental gradients (in particular species traits and gradient properties) has not been reached yet (Rydgren et al., 2003) and is subject to current scientific discussion.

Implications for future studies. The results gave valuable insights into predictor importance which are useful for a pre-selection of relevant input variables. The study could highlight the major criteria for selecting remote sensing products and certain categories of phenological metrics depending on biotope characteristics of the target species. This ecologically-based understanding of relevant remote sensing parameters supports future studies and calls for testing other available remote sensing products.

6.5 How many data? – The impact of inter-annual variability and length of the time series

Climate variables used in species distribution modeling are typically averaged scores over several years (e.g. the WorldClim data used in this thesis, Section 5.2). On the contrary, remote sensing predictors applied in species distribution models are often derived from single-year observations (Zimmermann et al., 2007) or sometimes even a single data set (Beck et al., 2005; Boubli and de Lima, 2009). However, inter-annual variability in vegetation characteristics as a result of varying climatic conditions is well-know from several remote sensing based vegetation mapping studies, e.g. in Brasil (Asner et al., 2000; Barbosa et al., 2006; Brando et al., 2010), southwestern USA (Huete and Didan, 2004), and Mexico (Salinas-Zavala et al., 2002; Colditz et al., 2009). Areas with high inter-annual variability might also not support stable (*source*) habitats or biotopes for species but rather represent *sink* conditions (Oindo et al., 2000). In the existing species distribution modeling literature (Buermann et al., 2008; Prates-Clark et al., 2008; Saatchi et al., 2008; Tuanmu et al., 2010; Cord and Rödder, 2011), phenological metrics were typically averaged over several years during the study period to minimize the effects of any natural inter-annual variability. Tuanmu et al. (2011) showed that the temporal transferability of species distribution models depends on the length of the time period considered. They compared and tested altogether four models developed from remote sensing time series of either one or three consecutive years and found that models based on variables derived from multi-year data had higher accuracies and higher temporal transferability than those with the variables from single-year data. Similar results were highlighted by Hüttich et al. (2009) who found that improved vegetation mapping accuracies could be achieved by the integration of inter-annual time series metrics in dry savanna ecosystems. These findings raise the question, which amount of data is necessary in species distribution modeling in order to capture the inter-annual variability of *Land Surface Phenology* (see Section 4.4.3) and at the same time reduce the number of input data as much as possible.

6.5.1 Objectives

This study aims to quantify the impact of inter-annual phenological variability of remotely sensed vegetation index data on the extent and accuracy of modeled species distribution ranges. Given the large amount of data available due to the increasing availability of continuous remotely sensed time series, this information is valuable for efficient data processing. The study design (Figure 6.15) includes the analysis of phenological metrics as single-year data sets and combined into multi-year average values (covering three, five, seven, and nine years; Table 6.8). In total, 25 different sets of remotely sensed phenological metrics covering different composite periods were tested for each of the target species. The study design, however, made use of continuous time series only meaning that consecutive years and not random combinations were analyzed. This approach is reasonable from both the ecological point of view (continuous environmental monitoring) and the perspective of data availability (analysis of consecutive, typically the most recent data). Beyond this, the analysis assumed stable land cover conditions throughout the study period from 2001 to 2009. Given the dramatic land cover changes in Mexico (see Section 2.5.3) this assumption may be violated and is not generally valid for the entire country, but may be anticipated for the field sites of the *National Forest Inventory* which are monitored on a regular basis.

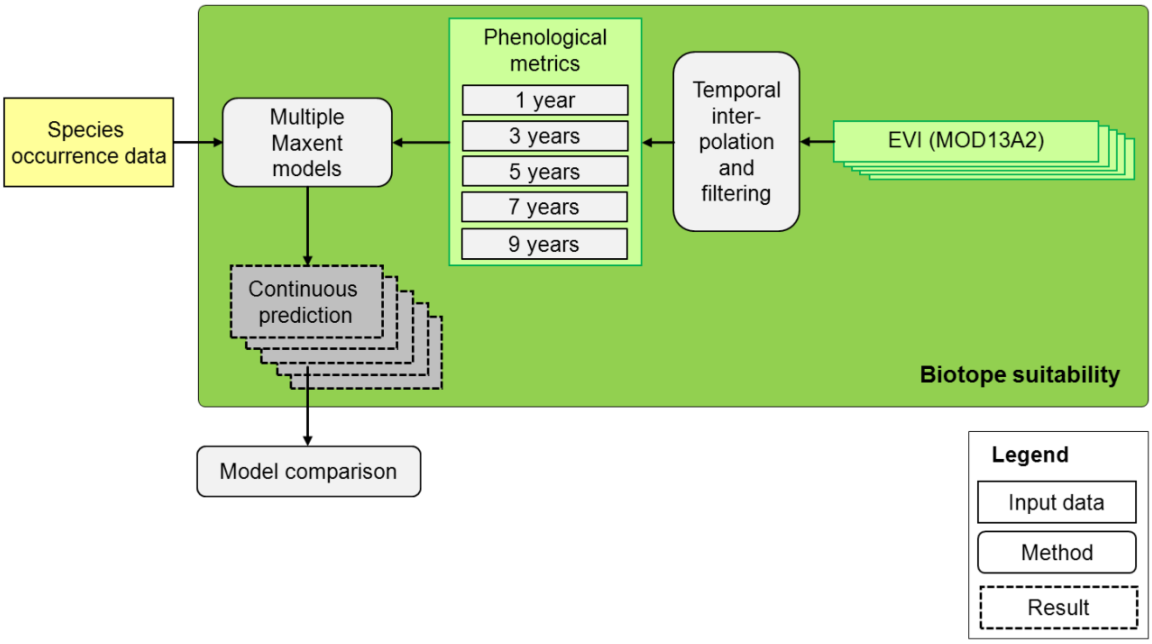


Figure 6.15: Modeling framework to assess the effect of inter-annual variability in time series of the vegetation index in species distribution modeling.

Table 6.8: Different composite periods of the phenological metrics analyzed to study the effect of inter-annual variability.

1 year	***3 years***	***5 years***	***7 years***	***9 years***
2001	2001-2003	2001-2005	2001-2007	2001-2009
2002	2002-2004	2002-2006	2002-2008	
2003	2003-2005	2003-2007	2003-2009	
2004	2004-2006	2004-2008		
2005	2005-2007	2005-2009		
2006	2006-2008			
2007	2007-2009			
2008				
2009				
			Total:	**25 data sets**

6.5.2 Inter-annual variability in remote sensing data

Inter-annual variability in the vegetation activity (measured by the *Enhanced Vegetation Index*, see Section 4.4.2) was quantified by the annual percentage variation normalized by the multi-annual mean between 2001 and 2009 (Figure 6.16). Spatial patterns of variability between different years were evident throughout the country, especially in the northern parts and along the *Baja California Peninsula* with annual variations of up to ±20%. These areas are generally drier and affected by higher inter-annual variability of precipitation (Magaña and Conde, 2000; see Section 2.2). Geographical patterns of vegetation activity further indicated considerable spatial heterogeneity of precipitation patterns within one year. On the contrary, the southern and mountainous regions (see Figure 2.1) showed lower inter-annual variation of vegetation activity (mostly between -5% and +5%). These areas are characterized by comparatively abundant vegetation cover (Figure 6.16, lower right panel) and include the different forest types that the study species are typical of (see Section 6.1). Consistent with the findings of Colditz et al. (2009), vegetation activity was conspicuously higher in a large proportion of Mexico in the year 2004 due to abundant precipitation. Spatial variability of vegetation activity in turn led to inter-annual differences in the scores of phenological metrics observed at the species presence sites (examples for *EVI_mean* are shown in Appendix K).

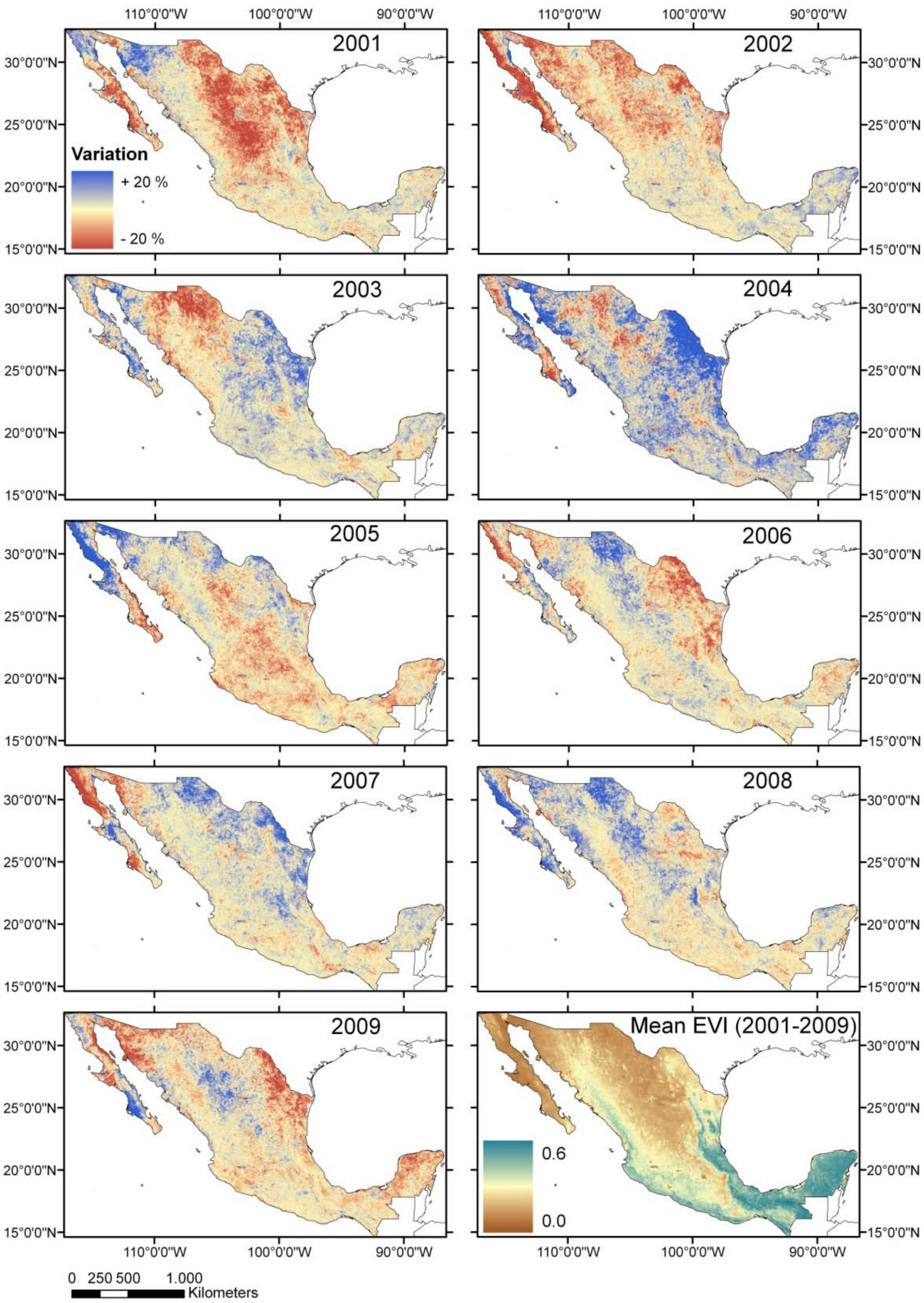


Figure 6.16: Annual percentage variation of vegetation activity (measured by *EVI_mean*) from the multi-annual mean over Mexico. The multi-annual (2001 to 2009) mean EVI used as reference data is shown in the lower right panel.

6.5.3 Model deviance from reference data

According to the study design with 25 different composite periods of the phenological metrics (nine annual and 16 multi-year average values, Table 6.8), species distributions were modeled independently for each composite period. The same settings for the *Maximum Entropy* algorithm were applied as described in Section 6.2.4. Statistical deviance from reference presence-absence data was calculated as implemented in the R package ‘dismo’ (R Development Core Team, 2011) and compared between the different composite periods for all study species.

The analysis showed that the use of longer time series in general improved the predictive performance of species distribution models (Figure 6.17) – although the impact greatly differed from 0.6% reduction in model deviance for *Pinus chiapensis* up to 30.0% for *Liquidambar macrophylla*. However, some models developed from single-year datasets could predict species occurrences with superior accuracies than the nine-year average (for example for *Alnus acuminata*, *Arbutus xalapensis* or *Pinus chiapensis*). In addition, the different input sets of phenological metrics resulted in substantial variability in model deviance depending on the time period under consideration. Especially short periods (one or three years) showed distinct variability (see for example *Abies religiosa* or *Avicennia germinans*). The effect was alleviated by the number of presence records. Wide-ranging species (e.g. *Bursera simaruba*) with large numbers of presence records generally featured less variation in model deviance (except for *Arbutus xalapensis*). No general pattern of certain years or composite periods (e.g. 2001, 2001-2003) that led to models with comparatively high or low deviance values across all species was found. However, consecutive year producing small deviance models (e.g. 2002, 2003, and 2004) also entailed distribution models with low deviance scores based on the respective composite period (e.g. 2002-2004). The importance of phenological metrics, however, as analyzed in Section 6.4 was not affected by the length of the composite period (Cord et al., 2011). Hence, highly important variables were relevant without exception independent of the length of the time series considered.

Because the analysis of model deviance pointed towards an impact of sample size, ‘virtual species data sets’ were created (analogous to the approach of Stockwell and Peterson, 2002) via re-sampling of occurrence records of *Bursera simaruba*, the study species with the most (3,940) samples. Sub-samples were created using random subsets of 20, 100, 500, 1,000, 1,500, and 2,000 records, respectively. The results (Figure 6.18) showed that model deviance was in general dependent on sample size and increased with the number of presence records used from less than 0.4 (20 records) to above 0.6 (≥ 1,500 records). In addition, variability of deviance varied more within the same composite length (e.g. 5 years) for models developed from fewer (20,100 or 500) records. By contrast, variability of deviance was lower for models based on higher number of records (1,000, 1,500 or 2,000). For all sample sizes except n = 20 with 17% percentage reduction, the use of nine years of remote sensing data led to comparatively low decrease in model deviance in comparison to models developed from single-year time series (between 0.1% for n = 100 and 4.1% for n = 1,000). In the following, the spatial model consensus across the different composite periods is assessed and compared.

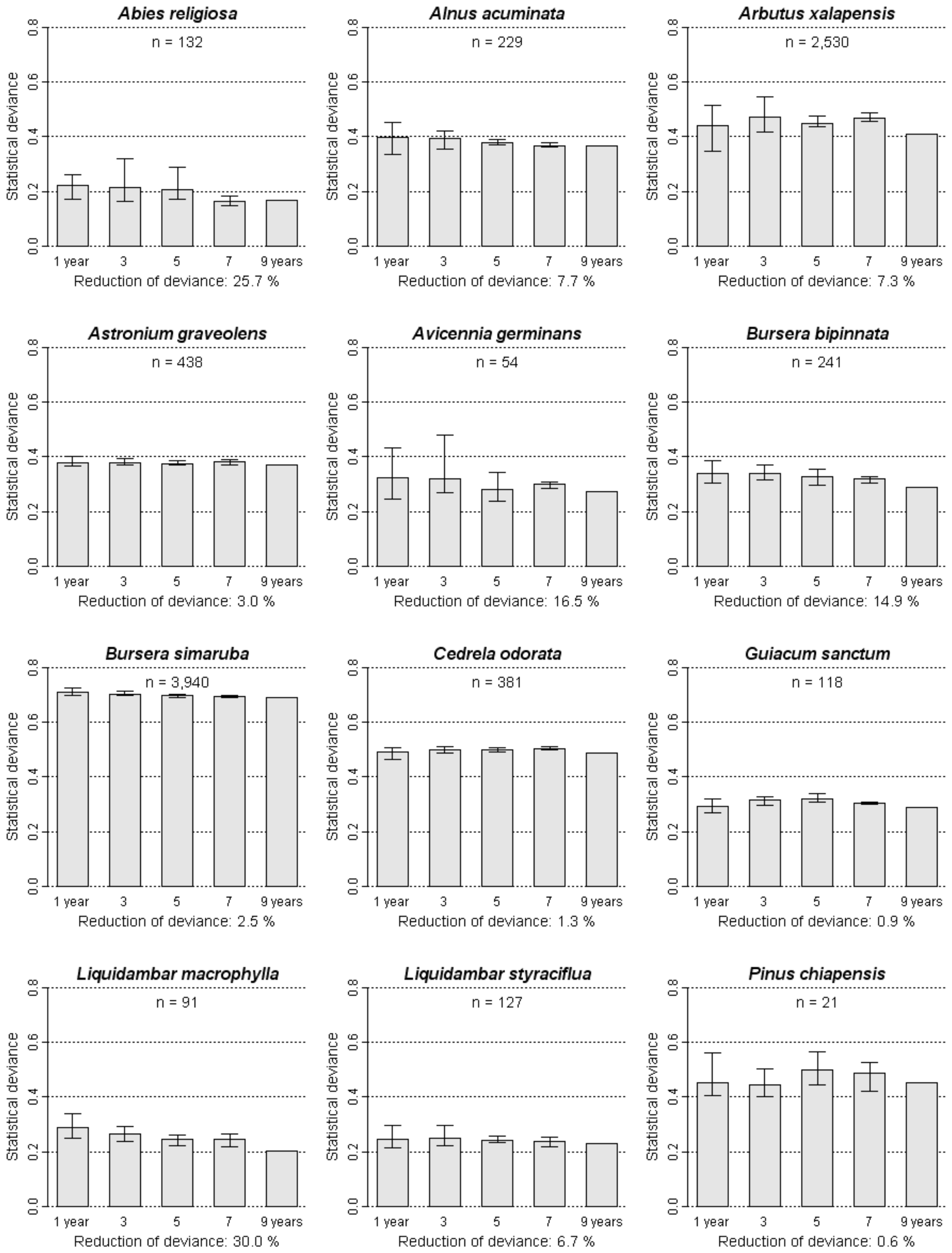


Figure 6.17: Model deviance from observed presence-absence data for the 12 study species. Error bars indicate the range (minimum and maximum) between different sets of composite periods of the respective length; *n* indicates the number of presence records per species; *reduction of deviance* refers to a comparison of model predictions based on single-year data or nine years of time series.

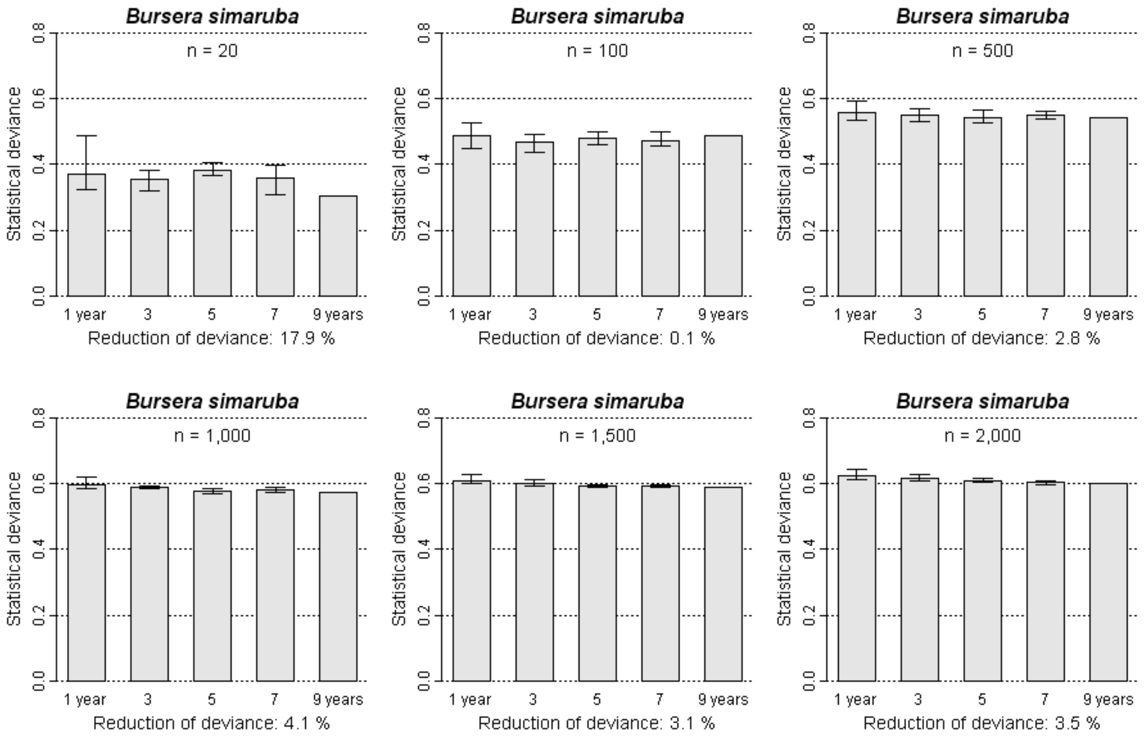


Figure 6.18: Model deviance for the ‘virtual species data set’ created via re-sampling of occurrence records of *Bursera simaruba*. For annotations see Figure 6.17.

6.5.4 Spatial model consensus

As an indicator of spatial agreement of the different distribution models, Figure 6.19 shows the accumulated probability of occurrence of the 25 species distribution models developed from remotely sensed time series of different composite periods. For illustration purposes, the two species with the highest reduction in model deviance based on single-year compared to nine consecutive years of time series, namely *Abies religiosa* (25.7%) and *Liquidambar macrophylla* (30.0%, see Figure 6.17) were selected. For *Abies religiosa*, a species typical of temperate evergreen forests, the highest probabilities of occurrence by all 25 models (*blue*) were predicted for the montane forests of the *Trans-Mexican Volcanic Belt* and the *Sierra Madre del Sur* (see Figure 2.1). Intermediate probabilities of occurrence (*green*) – due to either medium predicted probabilities across all models or partially high and partially low predicted probabilities – were modeled for some lower mountain regions. Finally, areas with low predicted probabilities for all models (*red*) were covered mainly by tropical deciduous forests and croplands. A different spatial pattern was observed for *Liquidambar macrophylla*, a tree typically occurring in moist montane cloud forests. Highest probabilities of occurrence (*blue*) were modeled along the eastern slopes of the *Sierra Madre Oriental* and the southern slopes of the *Sierra Madre del Sur*. The distinct boundary of the predicted distributional areas is obvious along the mountain ridge of the *Sierra Madre Oriental*, to the west of which forests are replaced by mainly sparse xerophilous vegetation and irrigated agriculture (shown in *red*).

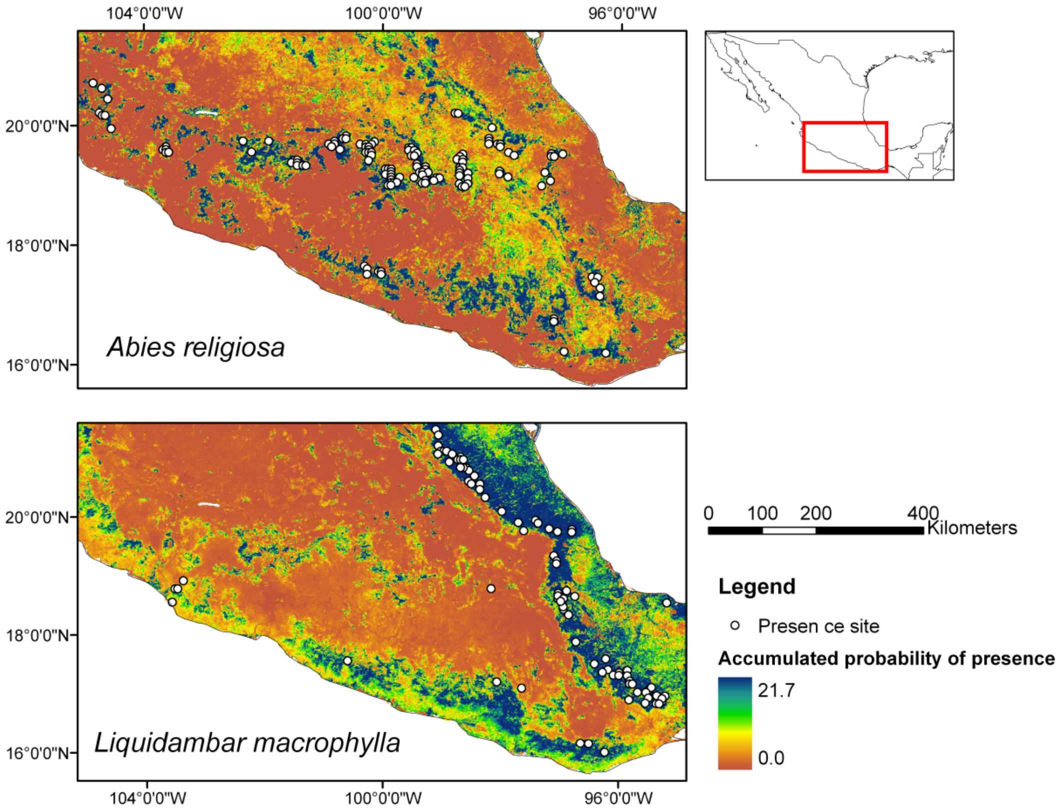


Figure 6.19: Accumulated probability of occurrence of 25 species distribution models developed from remotely sensed time series of different composite periods for *Abies religiosa* and *Liquidambar macrophylla*.

6.5.5 Discussion and implications

Inter-annual variability. The observed inter-annual variability in time series of the *Enhanced Vegetation Index* (Figure 6.16) is mainly a result of variation in precipitation which in large parts of Mexico is triggered by the El Niño/Southern Oscillation (ENSO) phenomenon. ENSO can be monitored based on the *Multivariate ENSO Index* (MEI, NOAA 2012). Positive variations from the MEI indicate El Niño, with generally warmer and wetter conditions in Mexico, negative variations point to La Niña, with cooler and drier conditions (Harris and Robinson, 2001; NOAA, 2012). ENSO greatly varies in intensity (Harris and Robinson, 2001) and no pronounced El Niño event occurred during the study period of this thesis from 2001 to 2009 (NOAA, 2012). However, the year 2004 featured very high vegetation activity throughout large portions of the country due to weak El Niño conditions, which led to higher accumulated precipitation than in all other years of the study period (see Colditz et al., 2009). Especially the arid and semi-arid regions in northern Mexico and areas with Mediterranean climate on the *Baja California Peninsula* (see Section 2.2) are governed by ENSO. For example, for the northeastern regions, it has been shown that the negative ENSO phase (La Niña) is associated with drier conditions with a delay of 4-6 months after the start of the event (Salinas-Zavala et al., 2002). In line with this, the years 2001, 2002, 2006, and 2009 showed comparatively low vegetation activity in the Northeast (Figure 6.16). Inter-annual variability was less pronounced in the southern parts of the country, which receive comparatively stable precipitation, even though also some of the tropical areas are affected by ENSO, for example through changes in hurricane activity (Harris and Robinson, 2001).

Importance of the length of the time series. The use of longer time series generally improved the predictive performance of species distribution models. Multi-year values reflect rather long-term conditions than annual fluctuations and are thus relevant for the establishment, survival, and reproduction of individual trees. Accordingly, especially short time periods (one or three years) showed the highest statistical model deviance. If little or even nothing is known about the inter-annual variability of remotely sensed variables within the distribution range of the target species, the use of longer time series is therefore recommended.

Impact of species characteristics. The reduction in model variance differed greatly between species. Wide-ranging species with a large sample size (e.g. *Bursera simaruba*, *Astronium geaveolens* or *Cedrela odorata*) and thus large numbers of presence records in general featured less variation in model deviance. This fact indicates that representative phenological metrics can be obtained by averaging in time (longer time series) or space (many presence sites) related to the concept of environmental completeness (Kadmon et al., 2003) discussed in Section 3.1.5. In this analysis, the wide-ranging species were at the same time typical of evergreen forest types – which certainly support the low levels of inter-annual variation in *Land Surface Phenology*. *Arbutus xalapensis*, with the second largest sample size, in this sense is an exceptional case. Even though the species has a high sample size (2,530 records), especially models developed from either one or three years of time series showed strong model deviance. The species exhibited this high variation presumably as its distribution range expands far north (see Appendix I); it is thus more affected by inter-variability than the other more southern species. Notably, the highest deviance value (0.52) for *A. xalapensis* was observed for the model developed from remote sensing data of the year 2004. As discussed before, this year showed significant positive variation in vegetation activity from the multi-year mean especially in the north (Figure 6.16). Beyond this, temperate species (*Abies religiosa*, *Alnus acuminata*, *Arbutus xalapensis*, and *Pinus chiapensis*) as well as the mangrove species *Avicennia germinans* generally featured higher variability in model deviance than tropical species. To summarize, the following criteria indicate that distribution models of a certain species are prone to be affected by inter-annual variability in remote sensing data: small sample size, species occurrence in temperate environments, and high environmental variability within the respective ecosystem or distributional range.

Implications for future studies. The observed improvement of predicted distribution ranges together with increasing length of the composite period of phenological metrics supports the need for continuous long-term time series of satellite remote sensing data. Longer (than nine years) composite periods may be analyzed in order to fully recognize the trade-off between data amount and model accuracy. As proposed by Tuanmu et al. (2011), the evaluation of temporal model transferability depending on the length of the remotely sensed time series is of high interest as well. Such analyses can be performed once species occurrence data from different sampling periods are available, e.g. in the form of updated data of the Mexican *National Forest Inventory*. In addition, the analysis of potential effects of El Niño/La Niña events and land cover change or degradation processes remains to be investigated. Even though the results showed consistent trends for several tree species, further analyses on other taxa and with different ecological requirements are needed to establish a common relationships between the length of the study period and model reliability. Especially studies with species that occur in the xerophilous vegetation communities in northern and central Mexico are of interest. Results may also vary depending on the size of the study region.

6.6 Summary

In this chapter, first of all a description of the twelve tree species (representative of the major Mexican forest types) studied here was given. Following this, the individual studies that were implemented and the respective methodological findings regarding the potential of multi-temporal remote sensing data for modeling species distributions were presented. In particular, the hierarchical modeling framework applied to integrate climate and remote sensing data based on ecological theory and observed spatial autocorrelation was illustrated. The studies focusing on the selection of appropriate methods to model biotope availability based on multi-temporal remote sensing data included (1) a comparison of the suitability of continuous remote sensing time series and an existing categorical land cover classification for the delineation of biotope availability, (2) an assessment of the importance of different remote sensing products for predicting species distribution ranges, and (3) a sensitivity analysis regarding the effects of inter-annual variability in time series of the vegetation index on modeling species distributional ranges. The major results of this chapter can be summarized as:

The **hierarchical modeling approach** implemented in this thesis outperformed purely climatic and remote sensing based models by the improvement of the effective spatial resolution of the environmental predictors and the integration of additional constraints on the species distributions. In addition, the analysis showed that climate variables (i.e. interpolated data) and remote sensing data (i.e. geo-referenced measurements) exhibited different strengths of spatial autocorrelation due to their distinct sample densities. As also species occurrence data typically feature spatial autocorrelation, test models which integrated climate and remote sensing data in a non-hierarchical model systematically overestimated the importance of spatially more autocorrelated climate data (termed the ‘red-shift’, Lennon, 2000). A non-hierarchical model design would therefore not fully exploit the potential of remote sensing data. Further, collinearity of explanatory variables hampers statistical analyses and the ecological interpretability of species distribution models (Graham, 2003; Heikkinen et al., 2006; Dudík, 2007; Baldwin, 2009). In this analysis, predictor covariation – which has as of yet largely been ignored for models including remote sensing data – was assessed. The results of the correlation analysis highlighted the significant amount of collinearity within both climatic predictors and remote sensing variables and thus indicated data redundancy. A framework for the species-specific selection of non-correlated predictors was implemented.

The comparison of **continuous remote sensing time series** and an existing **categorical land cover classification** for the delineation of biotope suitability showed that a species-specific view is required. The general trend towards overestimation of species distribution ranges based on categorical land cover information was found to be influenced as well by the specific characteristics of the target species. Typical indicators of a high suitability of land cover data to model the distribution of the target species could be identified, among other things, based on a novel *land cover evenness index*. This index characterizes the species-land cover relationship by quantifying different distributions of land cover classes observed at the species presence localities. In addition to the land cover evenness index, also the proportions of the study area that are covered by the respective most important land cover class(es), and the mapping accuracy of the dominant land cover class(es) are important. The findings of this study can therefore provide the theoretical knowledge to decide which data set (land cover classification or remote sensing variables) is more suitable given a certain target model performance.

The **importance of different remote sensing products** was measured by the increase in regularized model gain. The relevance of the predictors was largely determined by the different vegetation types that the target species were associated with. Variables derived from EVI time series were only of major importance for four out of the twelve study species, namely the species typical of tropical evergreen and tropical deciduous forests. Instead, the feature ranking also indicated a high potential of *Land Surface Temperature* for modeling species distributions – in particular for temperate species. Spectral bands in turn were found to be of minor relevance except for the mangrove species *Avicennia germinans* occurring in coastal *wetlands. The results therefore showed that veget*ation indices are the most widely used (see literature review in Section 4.3) but not necessarily the best predictors of species distributions. Further, the analysis of species response curves for remotely sensed variables revealed mostly determinate (unimodal or monotone) response shapes. The findings gave valuable insights into the variable importance that are useful for a pre-selection of relevant input variables and to develop an ecologically-based understanding of remote sensing parameters.

The observed **inter-annual variability** in time series of the *Enhanced Vegetation Index* was mainly a result of variation in precipitation which in large parts of Mexico is triggered by the El Niño/Southern Oscillation (ENSO) phenomenon. According to this analysis, the use of longer remotely sensed time series generally improved the predictive performance of species distribution models. However, the reduction in model variance varied greatly between species. Wide-ranging species with a large sample size (e.g. *Bursera simaruba*, *Astronium geaveolens* or *Cedrela odorata*) and thus large numbers of presence records in general featured less variation in model deviance. In addition to sample size, species occurrence in temperate environments indicated that distribution models of a certain species were prone to be affected by inter-annual variability in remote sensing data.

To summarize, in this chapter selected methodological aspects and their implications regarding the modeling of species distributions were assessed in order to develop a robust and transferable framework for modeling species distributions with multi-temporal remote sensing data. The respective findings were further applied to predict not only species distributions but also species richness of Mexican pines as outlined in the following chapter.

7 From species to species richness: A case study for Mexican pines

"The obvious extension to modeling individual species' distributions is to combine them to create maps of richness" (Stockwell and Peterson, 2003). This *stacked species distribution modeling* approach (Guisan and Rahbek, 2011) can make use of occurrence localities from presence-only data sources such as herbarium or field collections. It is thus an alternative to direct 'macroecological' modeling of species richness (see Section 4.5) and was introduced in consequence of the limited availability of systematic field inventories on species richness. However, the stacked species distribution model (SDM) approach focuses on aggregate distribution ranges of individual species (Guisan and Rahbek, 2011) and does not set a limit on the maximum number of species that can occupy a given habitat (Dubuis et al., 2011).

7.1 Objectives

This ignorance of limiting environmental conditions, for example the maximum *carrying capacity* or the competitive exclusion of species, makes the stacked SDM approach prone to overestimating species richness. In this study, remote sensing variables are incorporated as additional limiting environmental predictors into *Maximum Entropy* models, as described in the previous chapter, to obtain better estimates of current species distributions. The objective is to analyze whether the framework for modeling individual species distribution ranges (Chapter 6) can also be transferred to predict species richness.

The unique data setting available for this thesis is used to examine the potential of the older presence-only (primarily herbarium) collection data for predicting recently collected species richness information. Data of two completely independent data sets, namely the presence-only *National Information System on Mexican Biodiversity* (SNIB) and the presence-absence *National Forest Inventory* (INFyS, Section 5.1), are utilized. In this study, the first Mexican species richness map of the tree genus *Pinus* (family Pinaceae) based on 40 individual species range maps using climatic and remote sensing predictors was produced. Climatic suitability based on long-term climate data (BIOCLIM) and remotely sensed biotope availability (REMOTE) are modeled independently and then combined in a hierarchical framework (Section 6.2) to derive estimates of current species distributions (BIOCLIM_REMOTE). The evaluation is done with independently sampled species richness data (INFyS) that is temporally consistent with the acquisition period of the remote sensing data. Model performance of climatic and remote sensing based SDMs is compared and species-specific variable importance assessed. The effects of the inclusion of remote sensing predictors are analyzed by comparing estimates of individual species distribution ranges and species richness based on the BIOCLIM and BIOCLIM_REMOTE approaches with the respective reference data.

In particular, the methodological novel aspects of this analysis are:

(1) the modeling of species richness within a **less pronounced environmental gradient** than previous studies, namely for a single tree genus that is typical of a specific vegetation community,

(2) the comprehensive exploration of the value of **presence-only herbarium records** to predict **current species richness** based on remote sensing data,

(3) and the **application of a stacked species distribution modeling approach** to predict species richness based on **remote sensing data**.

7.2 Study species

Pines are ecologically relevant as a prominent, often dominant component of boreal, sub-alpine, temperate, and tropical forests as well as arid woodlands (Richardson and Rundel, 1998). Economically, pines are an important source of timber, paper, resins, charcoal, food (particularly seeds), and ornamentals (Le Maitre, 1998; Richardson and Rundel, 1998). The natural distribution of the genus is (except one population) confined to the Northern Hemisphere (Mirov, 1967) but several species are cultivated worldwide (Le Maitre, 1998).

Mexican pine-oak forests are among the richest ecoregions of the world for conifers (Dinerstein and Olson, 1999) and were recently added to the list of the 34 biologically richest and at the same time most endangered terrestrial ecoregions (Mittermeier et al., 2005). Approximately one quarter of all Mexican flowering plant species is found in the Madrean pine and pine-oak forests (Rzedowski, 1993). Today, the major threat to these forests is the conversion to agriculture and the exploitation for timber which both have significantly increased in recent years (Sarukhán et al., 2010). Pine-oak forests once covered about 21 percent of Mexico (Rzedowski, 1993), but the remnant forests are estimated to be no more than about eight percent (Mittermeier et al., 2005). The fragmented nature of the vegetation type makes it difficult to reliably assess the proportion of remaining pristine vegetation.

The genus *Pinus* is the major structural component of the Mexican pine-oak forests (Figure 7.1) and the largest existing genus of conifers with over 100 widely recognized species (Farjon, 2010). Mexico exhibits the greatest number of pine species of any region of similar size in the world with 40% of the known pine species (Farjon, 2010) and with 16 of these being endemic to the country. This high diversity is presumably the result of repeated southward migrations from North America during the Tertiary (Willyard et al., 2007; Gernandt et al., 2008) and subsequent local speciation in the climatically and geologically diverse mountainous areas (Farjon, 1996; see Sections 2.1 and 2.2).

Pines are found primarily at higher elevations in Mexico's major mountain ranges (*Sierra Madre Occidental*, *Sierra Madre Oriental*, *Trans-Mexican Volcanic Belt*, *Sierra Madre del Sur*, and *Sierra Madre de Oaxaca*) and isolated mountaintop islands in *Baja California*. Several types of communities can be identified which occur at different elevations, climates, and exposure conditions (Rzedowski, 1978): pine-oak and oak-pine forests, pure pine forests, and matorral interspersed with junipers. These pine-oak woodlands form an altitudinal transition zone between encinal woodland and pine forest (Marshall, 1957) and are composed of assemblages of several species of pine (*Pinus*) and oak (*Quercus*) or mixed stands with other conifers (Perry et al., 1998) such as Douglas-fir (*Pseudotsuga*) and fir (*Abies*). Between different Mexican pine species, strong differences regarding range size as well as preferred altitudinal range and climatic conditions required can be observed (Table 7.1). Rarity and conservation status vary from highly endangered species, e.g. *P. maximartinezii* or *P. jaliscana* which is currently known from only ten different localities of which half are on a few hectares (Pérez de la Rosa, 1983; Pérez de la Rosa, 2009), to widespread species not at risk such as *P. leiophylla*. In total, 20 out of the 40 species modeled in this thesis are included in the *NORMA Oficial Mexicana* NOM-059-SEMARNAT-2010 (SEMARNAT, 2010), which is the recently updated official list of endangered species in Mexico.

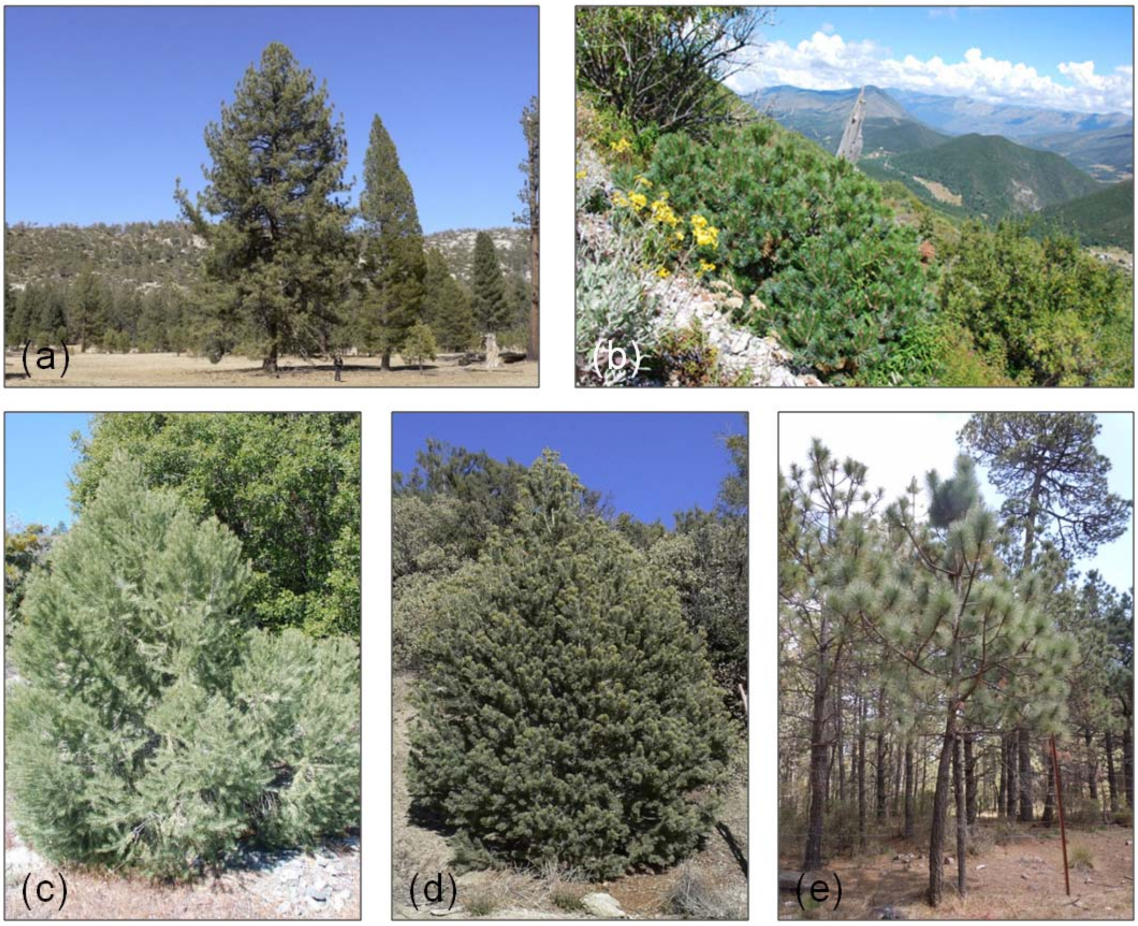


Figure 7.1: Examples of pine study species (Pictures: D. Gernandt). (a) *Pinus jeffreyi*, Baja California, (b) *Pinus culminicola*, Coahuila, (c) *Pinus californiarum*, California, (d) *Pinus quadrifolia*, Baja California, and (e) *Pinus hartwegii*, Veracruz.

7.3 Model training and reference data

Species occurrence records and pseudo-absence data (model training). For all 49 recognized Mexican pine species, presence-only locations of the *National Information System on Mexican Biodiversity* (SNIB, *Sistema Nacional de Información sobre Biodiversidad de México*, Section 5.1) based on geo-referenced locations of herbarium specimen and other field collections were assembled. The SNIB data were complemented by field surveys conducted by Dr. D. Gernandt (Institute of Biology, National Autonomous University of Mexico) and colleagues. All locations were extensively examined for errors and geographical outliers that may have resulted from incorrect geocoding. Species with less than ten presence records (namely *P. attenuata*, *P. caribaea*, *P. contorta*, *P. coulteri*, *P. flexilis*, *P. lagunae*, *P. luzmariae*, *P. maximartinezii* and *P. muricata*) were omitted due to their small sample sizes (Stockwell and Peterson, 2002). The distribution ranges of the remaining 40 species (Table 7.1, Figure 7.2a) were modeled.

Table 7.1: Description of pine species modeled. Range size categories were grouped based on the latitudinal and longitudinal range of the known presence localities (including records outside Mexico). Biogeographic regions and preferred climatic conditions were classified according to Richardson and Rundel (1998), which is based on several previous authors (Critchfield and Little, 1966; Mirov, 1967; Perry, 1991; Kindel, 1995).

Species	***Sample size***	***Range size***	***Biogeographic region***	***Climate***
Pinus arizonica	217	intermediate-wide	SW USA, N Mexico	temperate montane
Pinus ayacahuite	92	intermediate	Mexico, Central America	tropical montane
Pinus californiarum	25	wide	SW USA, Baja California	arid/temperate montane
Pinus cembroides	345	wide	Texas, NW Mexico	arid/temperate montane
Pinus chiapensis	46	wide	S Mexico, Guatemala	tropical montane
Pinus chihuahuana	111	intermediate	SW USA, NW Mexico	temperate montane
Pinus cooperi	58	intermediate	NW Mexico	temperate montane
Pinus culminicola	21	restricted	NE Mexico	temperate montane/ subalpine
Pinus devoniana	145	wide	Central and S Mexico, Guatemala	tropical montane
Pinus discolor	28	restricted-intermediate	SW USA, central and NW Mexico	arid/temperate montane
Pinus douglasiana	87	intermediate	W and central Mexico	temperate montane
Pinus durangensis	105	intermediate	NW Mexico	temperate montane
Pinus engelmannii	105	intermediate	W Mexico, SW USA	temperate montane
Pinus greggii	68	intermediate	NE Mexico	temperate montane
Pinus hartwegii	177	intermediate-wide	Mexico, Guatemala	tropical to subalpine

Pinus herrerae	84	intermediate	W Mexico	tropical montane/ temperate montane
Pinus jaliscana	19	restricted	W Mexico	tropical
Pinus jeffreyi	10	restricted	W USA, Baja California	temperate montane
Pinus johannis	19	restricted	NE Mexico	arid/temperate montane
Pinus lambertiana	11	restricted	Oregon, Nevada, California, Baja California	temperate montane
Pinus lawsonii	47	intermediate	S and central Mexico	tropical montane
Pinus leiophylla	292	wide	SW USA, Mexico	temperate montane
Pinus lumholtzii	86	intermediate	NW and central Mexico	temperate montane
Pinus maximinoi	94	wide	Mexico, Central America	tropical
Pinus montezumae	222	wide	Mexico, Guatemala	temperate montane
Pinus nelsonii	47	restricted	NE Mexico	arid/temperate montane
Pinus oocarpa	214	wide	Mexico, Central America	tropical
Pinus patula	116	intermediate	E Mexico	temperate montane/ tropical montane
Pinus pinceana	59	restricted	NE Mexico	arid/temperate montane
Pinus ponderosa	22	restricted-intermediate	W USA, N Mexico	temperate montane
Pinus praetermissa	10	restricted	W Mexico	tropical
Pinus pringlei	49	restricted-intermediate	S Mexico	tropical montane
Pinus pseudostrobus	366	wide	Mexico, Guatemala	temperate montane/ tropical montane
Pinus quadrifolia	42	restricted	California, Baja California	Mediterranean coastal
Pinus radiata	10	restricted	California, Baja California	Mediterranean coastal
Pinus remota	24	restricted	Texas, NE Mexico	Mediterranean coastal
Pinus rzedowskii	11	restricted	SW Mexico	tropical
Pinus strobiformis	133	wide	SW USA, N Mexico	temperate montane
Pinus tecunumanii	44	restricted	S Mexico, Central America	tropical montane
Pinus teocote	295	wide	Mexico	temperate montane
Total	**3,956**			

Due to the typically purposive sampling of herbarium presence-only collections and field data (Edwards Jr. et al., 2006), available species records were spatially (i.e. sampling bias towards roads and protected areas; see Figure 5.5) and thus presumably also environmentally biased. The influence of spatially biased samples in presence-only modeling can be reduced by comparing the occurrences with background points reflecting the same spatial bias rather than using a random background sample (see Section 3.4.1). According to this *target-group background* approach (Ponder et al., 2001; Anderson et al., 2003; Phillips et al., 2009) SNIB records of the tree genera *Acacia*, *Coccoloba*, *Bursera*, *Prosopis*, *Celtis*, *Caesalpinia*, *Salix*, *Ficus*, *Yucca*, and *Quercus* that had been collected using the same methods as the available pine presence data were used. After Phillips et al. (2009), 10,000 samples were randomly selected from these background data and used as pseudo-absence records for model training (Figure 7.2b).

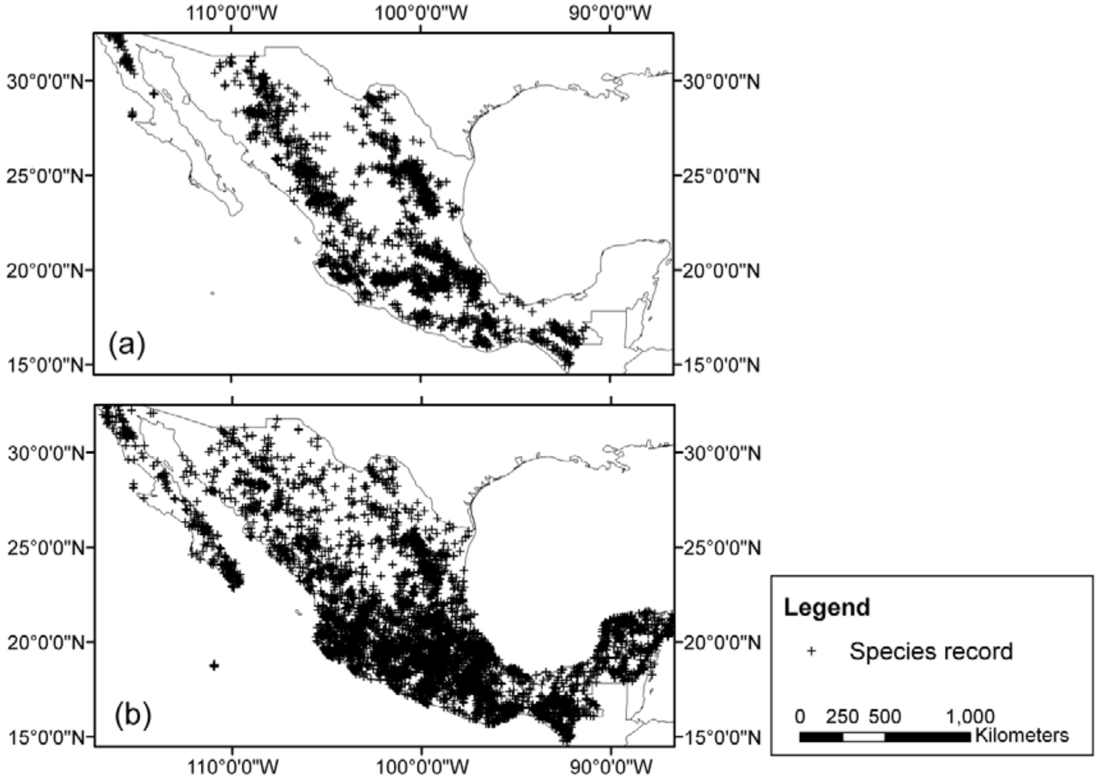


Figure 7.2: Spatial distribution of SNIB species records used for modeling pine distributions. (a) 3,956 presence sites (*Pinus* spp.), (b) 10,000 pseudo-absence sites (genera *Acacia*, *Coccoloba*, *Bursera*, *Prosopis*, *Celtis*, *Caesalpinia*, *Salix*, *Ficus*, *Yucca*, and *Quercus*).

Species richness data (reference). Independent pine species richness data from 19,319 available field sites (see Figure 5.2) of the Mexican National Forest Inventory (INFyS, *Inventario Nacional Forestal y de Suelos*, Section 5.1) were used for validation. Only the species included in Table 7.1 were considered in this richness assessment. The reference area for each INFyS site is 1 ha including four representative plots (*sitio*) for taxonomic identification (see Section 5.1). The inventory was carried out between October 2004 and November 2007 and is thus a recent and reliable estimate of tree species richness in Mexico.

7.4 Climatic distribution ranges

As climatic model predictors, the WorldClim data base (Version 1.4; Hijmans et al., 2005; Section 5.2) was used. For each species, all pairs of highly-correlated environmental predictors (Spearman's r > 0.8) were identified. Out of each pair of correlated variables, the predictor with the higher explanatory power according to a *Generalized Linear Model* (GLM) was retained. According to this analysis (see Section 6.2.3 for further aspects regarding the implementation), significant collinearity and redundancy was found between bioclimatic variables; only seven to eight bioclimatic (out of 19) variables per species were retained as input variables. To obtain bioclimatic suitability maps for each species, ten replicates of Maxent models using only those non-correlated climatic predictors with replicate sample sets selected based on bootstrap resampling were computed (Settings: replicates=10, replicatetype=bootstrap, randomtestpoints=25, applythresholdrule=maximum training sensitivity plus specificity, auto features, responsecurves, jackknife, addsamplestobackground, nodata=-3000, regularization multiplier=1, maximum iterations=500, convergence threshold=0.0001, writeplotdata, autorun, nodoclamp, nowritemess, noprefixes, nowriteclampgrid, nooutputgrids, writebackgroundpredictions, randomseed, redoifexists, notooltips). Various metrics (see Section 3.4.4) have been proposed to convert the resulting continuous probability surfaces of SDMs into binary presence-absence maps (Liu et al., 2005). Herein, the *maximum training sensitivity and specificity threshold*, which is specific for each species and each model replicate, was applied. This flexible, species-specific threshold has been shown to be superior to fixed thresholds for a wide range of sample sizes and prevalence rates (Manel et al., 2001; Jiménez-Valverde and Lobo, 2007). For each model replicate, all cells with modeled logistic probability of occurrence below the respective threshold value were classified as *absence* and vice versa for *presence*. Finally, all cells predicted as *presence* by all ten model replicates were classified as the composite bioclimatic range of each species (BIOCLIM, see also Section 6.2.4).

7.4.1 Predictor importance

While the Maxent model is being trained, it keeps track of which internal features are contributing to fitting the model (Phillips, 2009). This increase in model gain is assigned to the environmental variable(s) that the feature depends on (see Section 3.3.2). The results of this variable importance analysis (Figure 7.3) showed that BIO4 (*Temperature seasonality*) was the most important bioclimatic variable across all species with an average percentage contribution of 14.1%. BIO4 was followed by three precipitation variables (BIO15: *Precipitation seasonality*, BIO18: *Precipitation of warmest quarter*, and BIO16: *Precipitation of wettest quarter*) that contribute to explaining the distribution of pines. BIO1 (*Annual mean temperature*) was the least important bioclimatic predictor with a relative contribution of 0.5%.

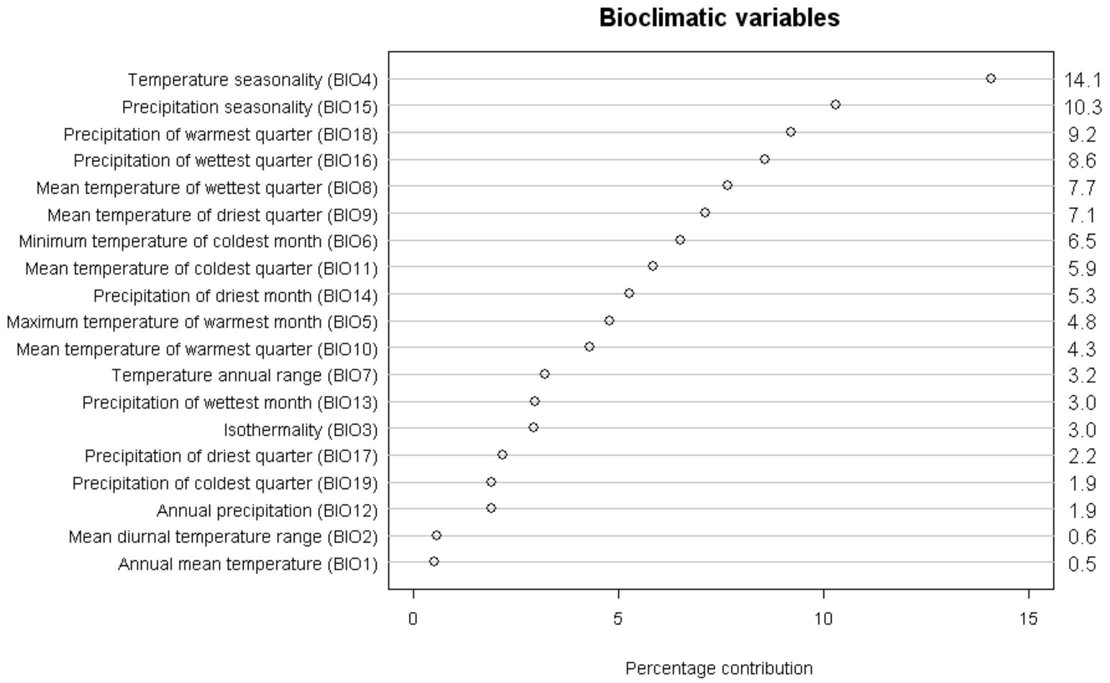


Figure 7.3: Percentage contributions (mean scores across all species and all model replicates) of WorldClim climate variables to the Maxent models measured by the increase in regularized model gain.

7.4.2 Model evaluation

The *area under curve* (AUC) is a non-parametric accuracy measure of model performance (see Section 3.4.5). According to the classification of Swets (1988), the BIOCLIM variable set produced ‘good’ AUC scores for 11 species and ‘excellent’ for the remaining 29 species (Table 7.2). On average, models developed with the BIOCLIM data exhibited a training AUC of 0.959 (SD ± 0.036) and a test AUC of 0.938 (SD ± 0.054). The observed similarities in both training and test data AUC underline the robust performance of the Maxent algorithm to predict a fraction of the species records omitted during model training and hence the algorithm’s ability to capture variations in environmental variables over point localities as already shown in previous studies (Saatchi et al., 2008). The modeled distribution ranges featured a good spatial concordance with known distributions of the target species according to literature (Farjon, 1996).

7.5 Remote sensing based models

Like the climate data, also the remote sensing variables showed significant collinearity and redundancy. Per species, only 13 to 18 (out of 53) variables were retained as predictors after the correlation analysis and subsequent exclusion of highly correlated data (see Section 6.2.3). Analogous to the climatic distribution models (i.e. using the same settings), Maxent models were run using only the selected, non-correlated remote sensing predictors (REMOTE) including phenological and statistical metrics derived form time series of the *Enhanced Vegetation Index* (EVI), *Land Surface Temperature* (LST), and *Reflectance* data. Again (see Section 7.4), the *maximum training sensitivity and specificity presence threshold* was applied for the conversion to presence-absence maps.

7.5.1 Predictor importance

Supporting the findings of the variable importance assessment of the climatic predictors (Figure 7.3), the annual coefficient of variation of the *Land Surface Temperature* was the most important parameter within the remote sensing predictors. LST metrics generally showed the highest contribution to the increase in regularized model gain and four out of the five variables with the highest relevance were derived from LST (Figure 7.4). In particular, the most important variables were the annual coefficient of variation (*LST_CoV*) and the annual minimum LST (*LST_minimum*) which together contributed more than 30% to the gain of the Maxent models across all species. Further, phenological metrics derived from EVI data were assigned with higher variable importance scores than statistical metrics based on spectral reflectance data. Only four reflectance predictors (*blue_CoV*, *blue_range*, *red_CoV*, and *NIR_maximum*) were among the 20 most important variables shown in Figure 7.4. The overall lowest contribution was assigned to the annual range of NIR reflectance (*NIR_range*) with a score of 0.2%. However, variable contributions varied strongly between species depending on the ‘contrast’ between values of the environmental predictors observed at presence and pseudo-absence sites. Even the most relevant variables were thus not important for all species. This emphasizes the importance of a species-specific variable selection, as applied in this thesis, based on the explanatory power of the predictors which has been neglected in remote sensing based species distribution modeling so far.

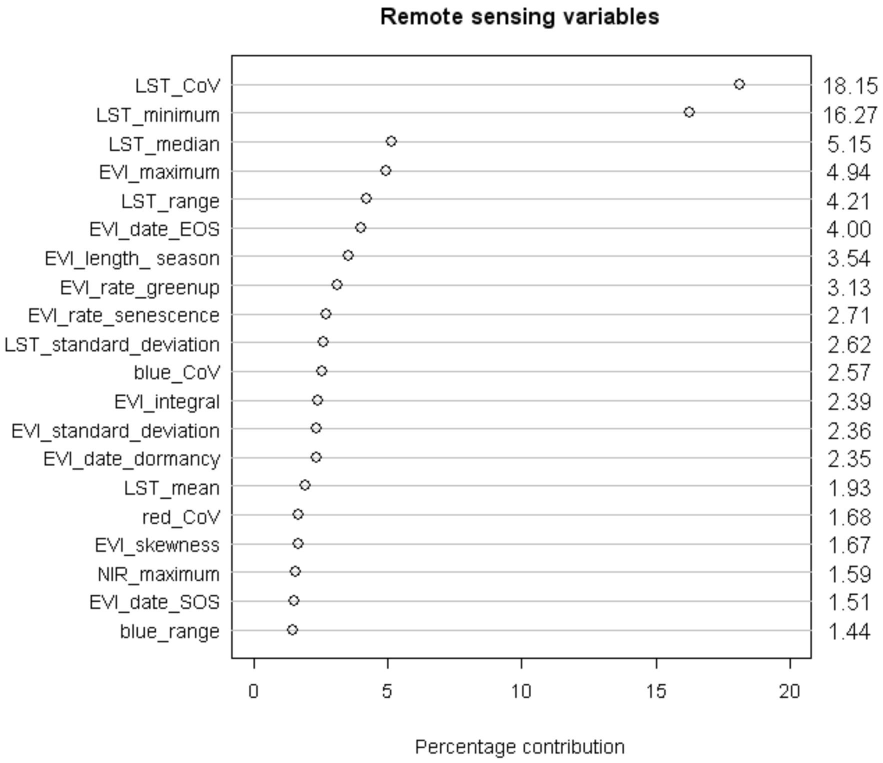


Figure 7.4: Percentage contributions of the 20 most important (according to mean scores across all species and all model replicates) remote sensing variables to the Maxent models measured by the increase in regularized model gain.

7.5.2 Model evaluation

Analogous to the findings of Prates-Clark et al. (2008), models developed with the REMOTE data set performed slightly worse (mean training AUC = 0.947; mean test AUC = 0.913) than those computed with the BIOCLIM predictor set (Table 7.2). According to the classification of Swets (1988), SDMs based on remote sensing predictors revealed ‘fair’ AUC values for 3 species (though only for test AUC), ‘good’ for 14 species, and ‘excellent’ for the other 23 species. In general, AUC scores differed more between species than between the BIOCLIM and REMOTE predictor sets for the same species and were dependent on the characteristics of the geographical distribution of the species. In line with Hernandez et al. (2006), accuracy was in general higher for species with restricted geographic ranges and small sample sizes (see Table 7.1), e.g. *P. jaliscana*, *P. jeffreyi*, *P. lambertiana*, and *P. radiata* as those species in general have narrow environmental tolerance. In addition, species which are typically difficult to identify in the herbarium (*P. maximinoi*, *P. lawsonii*, and *P. devoniana*) showed lower scores which could be an indicator of misidentification of some specimen (vouchers).

Table 7.2: AUC scores of the internal Maxent validation. ‘Excellent’ (0.9-1.0) values are printed in bold and ‘fair’ (0.7-0.8) values in italics. Training AUC describes the ability of the model to correctly score a randomly selected presence site higher than a randomly selected background site, test AUC describe the model’s ability to predict a fraction of the species records omitted during model training (Phillips et al., 2006; see Section 3.4.5).

	BIOCLIM		***REMOTE***	
Species	***Training***	***Test***	***Training***	***Test***
Pinus arizonica	**0.964**	**0.959**	**0.965**	**0.947**
Pinus ayacahuite	**0.924**	0.897	0.899	0.873
Pinus californiarum	**0.998**	**0.998**	**0.999**	**0.998**
Pinus cembroides	**0.946**	**0.933**	**0.942**	**0.913**
Pinus chiapensis	**0.945**	**0.910**	**0.938**	0.894
Pinus chihuahuana	**0.980**	**0.966**	**0.981**	**0.967**
Pinus cooperi	**0.983**	**0.976**	**0.975**	**0.960**
Pinus culminicola	**0.998**	**0.998**	**0.987**	**0.969**
Pinus devoniana	**0.905**	0.825	**0.912**	*0.783*
Pinus discolor	**0.990**	**0.987**	**0.987**	**0.980**
Pinus douglasiana	**0.914**	0.890	0.886	0.805
Pinus durangensis	**0.963**	**0.953**	**0.961**	**0.932**
Pinus engelmannii	**0.967**	**0.953**	**0.963**	**0.946**
Pinus greggii	**0.974**	**0.959**	**0.943**	**0.923**
Pinus hartwegii	**0.950**	**0.915**	**0.944**	0.892
Pinus herrerae	**0.931**	**0.909**	**0.912**	0.865
Pinus jaliscana	**0.997**	**0.995**	**0.992**	**0.983**
Pinus jeffreyi	**0.998**	**0.997**	**0.999**	**0.998**
Pinus johannis	**0.985**	**0.978**	**0.971**	**0.948**
Pinus lambertiana	**0.999**	**0.999**	**0.999**	**0.998**
Pinus lawsonii	**0.929**	0.891	0.868	*0.770*
Pinus leiophylla	**0.908**	0.874	**0.907**	0.837
Pinus lumholtzii	**0.950**	**0.940**	**0.945**	**0.922**
Pinus maximinoi	0.892	0.848	0.850	*0.785*
Pinus montezumae	0.894	0.852	**0.908**	0.825
Pinus nelsonii	**0.995**	**0.993**	**0.975**	**0.950**
Pinus oocarpa	0.896	0.841	**0.908**	0.801
Pinus patula	**0.964**	**0.941**	**0.946**	0.880
Pinus pinceana	**0.975**	**0.967**	**0.963**	0.952
Pinus ponderosa	**0.991**	**0.983**	**0.994**	**0.987**
Pinus praetermissa	**0.911**	0.877	**0.915**	0.883
Pinus pringlei	**0.969**	**0.964**	**0.921**	0.882
Pinus pseudostrobus	0.876	0.846	0.892	0.843
Pinus quadrifolia	**0.997**	**0.996**	**0.997**	**0.996**
Pinus radiata	**1.000**	**0.999**	**0.998**	**0.992**
Pinus remota	**0.996**	**0.995**	**0.991**	**0.985**
Pinus rzedowskii	**0.969**	**0.924**	**0.939**	**0.984**
Pinus strobiformis	**0.968**	**0.950**	**0.972**	**0.935**
Pinus tecunumanii	**0.982**	**0.978**	**0.955**	**0.903**
Pinus teocote	**0.985**	0.859	0.894	0.848
Mean	0.959	0.938	0.947	0.913
Standard deviation	0.036	0.054	0.041	0.069

7.6 Combination of climatic and remote sensing based models

The remotely sensed biotope suitability scores (REMOTE) were assigned to all cells within the previously determined climatic range (BIOCLIM, see Section 7.4). The usefulness of this hierarchical approach for modeling tree species distributions in Mexico was assessed in detail in Section 6.2. The resulting maps (both continuous and binary) indicated areas predicted as being suitable by both the bioclimatic and the remote sensing data sets (BIOCLIM_REMOTE). Statistical model deviance (see Section 3.4.5) from the observed presence-absence data was lowest for the majority of species for the BIOCLIM_REMOTE model (mean = 0.169), followed by BIOCLIM (mean = 0.210) and REMOTE (mean = 0.272). Besides these systematic differences depending on the predictor set, model deviance generally varied more between species than between predictor sets for the same species. Due to the comparatively better performance of the BIOCLIM model, results are only shown and compared between BIOCLIM and BIOCLIM_REMOTE models in the following.

To assess predicted range sizes, the number of predicted presence cells was compared between the different binary model predictions. Range size derived from the binary presence-absence maps according to the *maximum training sensitivity and specificity threshold* differed largely between species (Appendix L). The inclusion of remote sensing data in the hierarchical modeling framework led to a mean reduction in range size across all 40 study species of 65.1%, ranging from 18.8% (*Pinus engelmanii*) to 95.7% (*Pinus jeffreyi*). The mean *fractional predicted area* (FPA) of the BIOCLIM_REMOTE model was lower than the one of the BIOCLIM model for all species. The mean distribution range size was 3.7% (BIOCLIM) and 1.6% (BIOCLIM_REMOTE) of the Mexican land surface (see Appendix L).

7.7 Species richness assessments

By overlaying the spatial predictions of the presence-only models (*stacked species distribution model*; Graham and Hijmans, 2006; Guisan and Rahbeck, 2011; see Section 4.5) using (1) the BIOCLIM and (2) the BIOCLIM_REMOTE models, Mexican pine richness maps at 1 km² spatial resolution were created. For each validation site, the difference between modeled richness and independent species richness data was computed for the two different models. The maximum pine richness observed in the field according to the Mexican *National Forest Inventory* was eight species per reference plot. Both models over-estimated this maximum species richness with 14 (BIOCLIM) and 12 (BIOCLIM_REMOTE) species per cell. The richness maps (Figure 7.5) showed a significant spatial agreement with the major mountain ranges in Mexico (see Figure 2.1). Both models predicted the highest species richness in the *Sierra Madre Occidental* which is consistent with the data of the *National Forest Inventory*. Lower species numbers were observed and modeled for the *Sierra Madre Oriental*, *Trans-Mexican Volcanic Belt*, *Sierra Madre del Sur*, and *Sierra Madre de Oaxaca*. Since *P. caribaea*, the only pine species occurring on the Yucatan peninsula, was represented by too few samples and hence excluded from this analysis (see Section 7.3), no pines were modeled for the Yucatan peninsula. The representative results shown for the *Sierra Madre Occidental* (Figure 7.5c-e) demonstrate that the predicted patterns of species richness included more spatial detail and were closer to the independent validation data in case modeled species distribution ranges based on climatic and remote sensing data had been combined in the hierarchical framework.

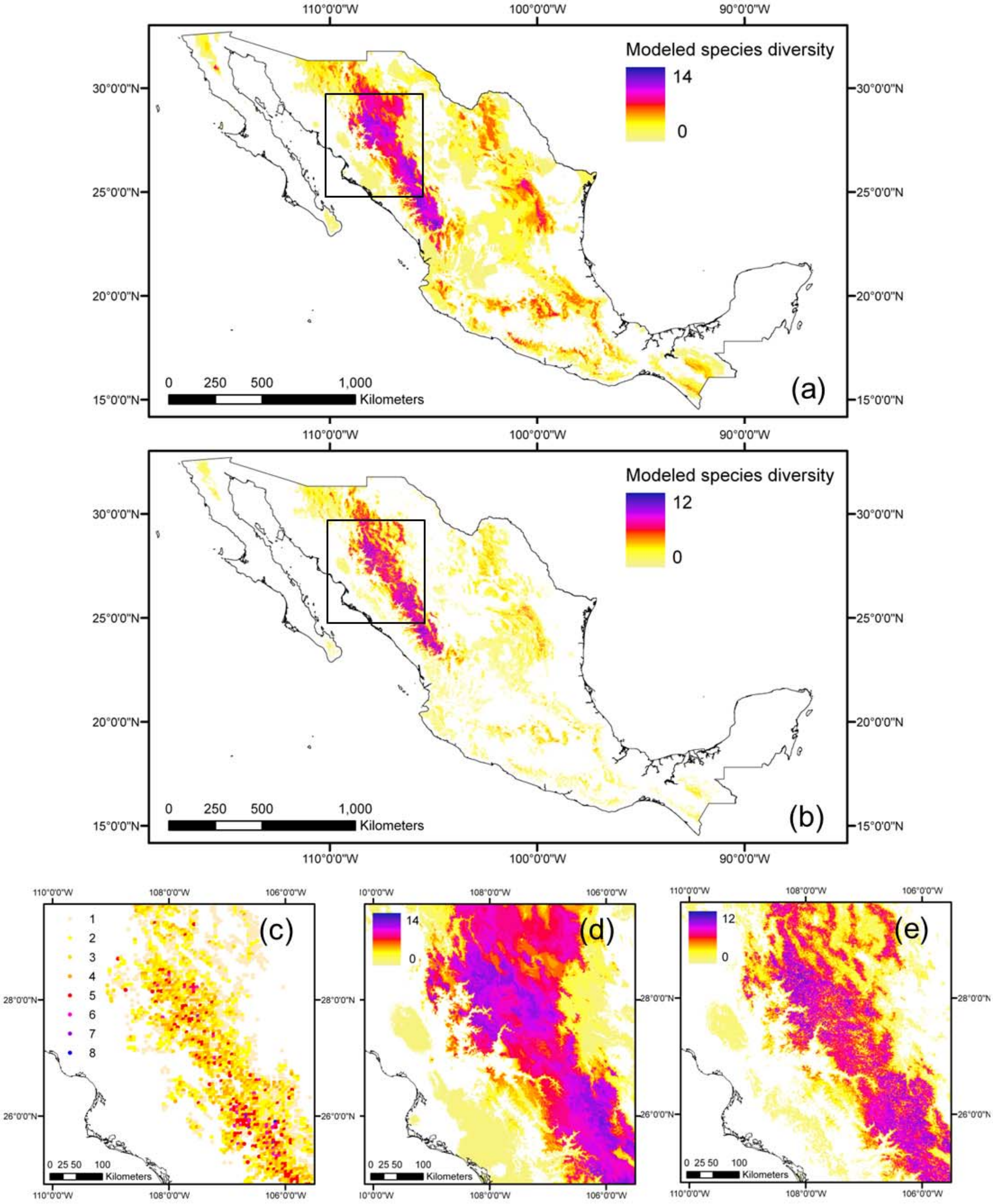


Figure 7.5: Modeled pine species richness in Mexico using (a) the BIOCLIM and (b) the BIOCLIM_REMOTE approaches. Figures (c), (d), and (e) show the area of detail with the highest pine species richness in the *Sierra Madre Occidental*: (c) Observed species richness based on field data, (d) Prediction of the BIOCLIM models, and (e) Prediction of the BIOCLIM_REMOTE models.

The systematic bias towards overestimation of species richness by the models was not only apparent from the respective richness maps (Figure 7.5), but also from the relative frequencies of species richness classes (Figure 7.6). Accordingly, the underestimation of cells without pine species presence (richness = 0) was higher in the BIOCLIM than in the BIOCLIM_REMOTE model. The inclusion of remote sensing data thus led to less underestimation of species-poor sites and less overestimation of species-rich sites. This greater overestimation of species richness based on stacked species distribution ranges modeled using only climate data (BIOCLIM) was consistent with the comparatively large modeled range sizes as shown in Appendix L.

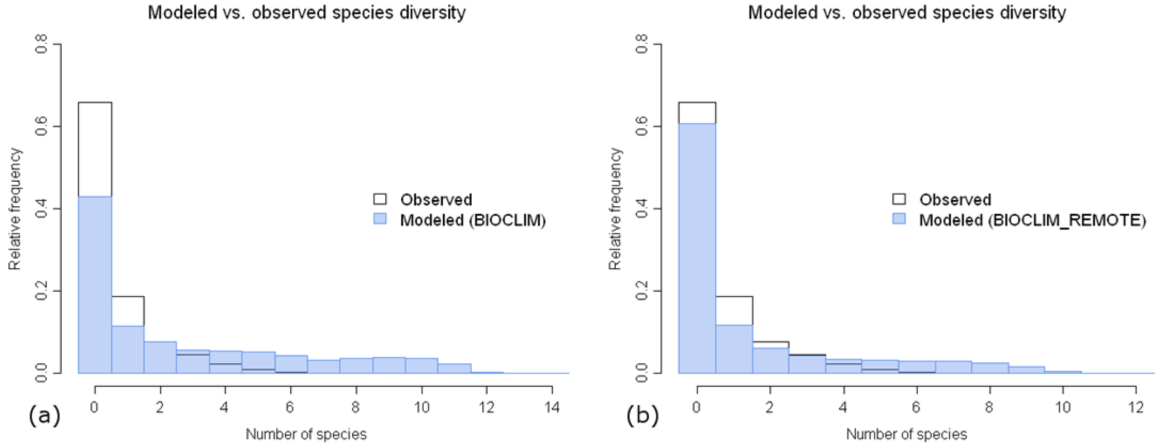


Figure 7.6: Relative frequencies of observed and modeled species richness for (a) the BIOCLIM and (b) the BIOCLIM_REMOTE model.

The modeled number of species was further compared with the independent species richness scores at the 19,319 reference sites (Figure 7.7). The results of this site-specific model evaluation support the superior performance of the BIOCLIM_REMOTE model which was significantly better in predicting species richness than the BIOCLIM model ($p < 0.001$, Mann-Whitney U-test). The mean difference to observed species richness data was 3.1 species per site (median = 2) for BIOCLIM and 1.5 species per site (median = 0) for BIOCLIM_REMOTE. The correlation between observed and modeled species richness scores was assessed by calculating the Spearman rank correlation coefficient (BIOCLIM: Spearman's $r^2 = 0.50$; BIOCLIM_REMOTE: Spearman's $r^2 = 0.51$). Correlation coefficients were comparatively low due to the small range of values encountered (from 0 to maximum 14 species).

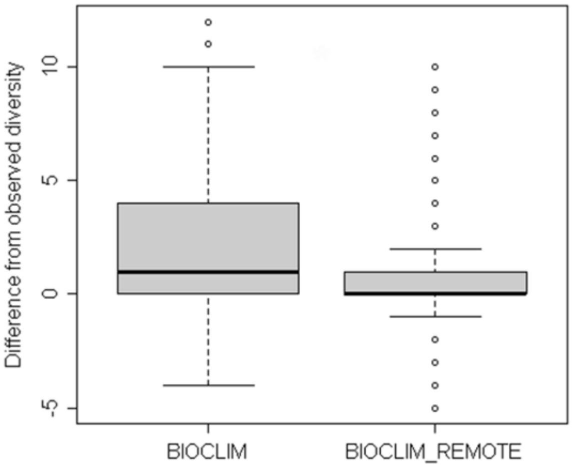


Figure 7.7: Difference between observed and modeled species richness for the BIOCLIM and BIOCLIM_REMOTE models.

7.8 Discussion and implications

Accuracy of species distribution models depending on sample size. Higher model accuracies for species with limited geographical ranges, as found in this analysis, were observed by previous studies for other taxonomic groups than plants (Hernandez et al., 2006; McPherson and Jetz, 2007) as well. Particularly for trees, higher model performance was found for species with narrow ecological niches and hence narrow geographical distribution in Switzerland (Guisan et al., 2007). This trend of increasing model accuracy with more specialized and restricted-range species is limited by the *minimum sample size* required for reliable species distribution modeling. Minimum sample size has been a subject of discussion in a variety of studies, e.g. Stockwell and Peterson (2002) who indicated a minimum level of ten sample points to achieve 90% of maximum model accuracy (see also Section 3.4.1). Even though Maxent is moderately sensitive to sample size compared to other algorithms (Pearson et al., 2007; Wisz et al., 2008; Costa et al., 2010), species with less than ten available records had therefore been excluded from this analysis. The availability of very few presence localities has been shown to be a significant source of model instability and errors in species distribution models (Guisan and Thuiller, 2005). With decreasing sample size, the levels of uncertainty associated with parameter estimates (e.g. means, medians, and predicted probabilities) increase and outliers carry more weight (Wisz et al., 2008). Especially for very rare species with few records it is thus in many cases not possible to reliably predict their ranges and include them in biodiversity assessments. A solution might be a multi-scale ensemble forecasting framework, which was particularly designed for modeling rare species and applied for an Iberian endemic plant (Lomba et al., 2010).

Predictor importance. Across all pine species studied, *Land Surface Temperature* metrics were the most important remote sensing predictors for explaining spatial patterns of species distributions, followed by *Enhanced Vegetation Index* and *Reflectance* metrics (Figure 7.4). Pines primarily occur along the mountain ranges in Mexico (see Section 7.2), mainly within temperate montane and tropical montane climates (see Table 7.1). At such higher elevations, LST showed an annual seasonal cycle which could be differentiated from the mean across the study area (see Figure 5.1) and was therefore suitable for characterizing pine species distributions. However, predictor importance varied between the different pine species depending on their preferred altitudinal ranges. For example, for tropical species (e.g. *P. oocarpa*, *P. jaliscana*, and *P. rzedowski*), occurring at lower altitudes than tropical montane or temperate montane species, metrics derived from time series of the EVI were of higher importance than the ones of LST. These findings were consistent with the results of the predictor importance analysis for the twelve tree species representative of the major Mexican forest types (see Section 6.4.2). For those species, the feature ranking indicated a high potential of LST for modeling the distributions of the temperate (as opposed to tropical) species as well. All species of the tropical forest types were best predicted by phenological metrics derived from EVI time series. This emphasizes the importance of a species-specific variable selection based on the explanatory power of the predictors which has been neglected in remote sensing based species distribution modeling so far.

Contribution of remote sensing data. The multi-temporal remote sensing imagery facilitated the identification of limiting environmental conditions beyond climatic suitability and allowed for the extraction of phenological and seasonal variations in vegetation cover that affect and mirror geographical species distributions. While bioclimatic models indicated suitable climatic conditions, the inclusion of remote sensing data allowed for identifying the remnant pine-oak forests; remote sensing variables may hence be interpreted as indicators of biotic conditions and/or human disturbance. Further, remotely sensed data improved the effective spatial resolution of the environmental predictors from interpolated surfaces (as for climate data) to direct geo-referenced measurements. However, the integration of remote sensing predictors also added additional noise or statistics to the species-environment relationships since similar vegetation features may arise under different topo-climatic conditions and vice versa (Zimmermann et al., 2007; Saatchi et al., 2008). The large variation in the reduction of the *fractional predicted area* (due to the inclusion of remote sensing data, Appendix L) could not be related to range size, sample size or preferred climatic conditions of the respective species.

Given the assumed distribution of remaining pine-oak forests in Mexico with about eight percent (Mittermeier et al., 2005), the prediction of 14.3% (*P. cembroides*) and 13.7% (*P. arizonica*) based on BIOCLIM predictors clearly over-estimated maximum potential species distribution. No overestimation was evident from the BIOCLIM_REMOTE models which predicted 7.9% (*P. cembroides*) and 6.6% (*P. arizonica*) of the land surface. Bioclimatic ranges thus indicated more accurate estimates of historical potential distribution ranges while the combined models including remote sensing data portrayed the current potential distribution. For example, modeled species richness according to the BIOCLIM model was similarly high in the core areas and margins of the mountain chains. In contrast, the BIOCLIM_REMOTE approach predicted lower scores of species richness along the margins. Along these foothills with moderate slopes and better accessibility, human impact and land transformation are significantly higher (Sommerhoff and Weber, 1999) leading to a reduction in pine species richness.

Modeling species richness. Previous studies have shown that overlaying species range maps based on climatic models generally leads to an overprediction of species richness (Graham and Hijmans, 2006; Algar et al., 2009; Newbold et al., 2009; Pineda and Lobo, 2009; Raes et al., 2009). Without rules (e.g. maximum *carrying capacity* or competitive exclusion) to constrain the maximum species richness per grid cell, stacked SDMs predict the entire range of potentially occurring species in a given area and thus typically tend to overpredict species richness (Diniz-Filho et al., 2009; Dubuis et al., 2011). This analysis found a similar systematic bias towards overestimation of species richness. Model overestimation may further be influenced by the threshold used for converting the continuous probability surfaces into binary presence-absence maps (Aranda and Lobo, 2011). In this analysis, two additional threshold criteria (*minimum training presence* and *10 percentile training presence*) had been applied, but both revealed higher disagreement of modeled species richness from the reference data than the selected *maximum training sensitivity and specificity threshold*. Thus, only the results according to the *maximum training sensitivity and specificity threshold* were shown here. However, in comparison to other studies, the overestimation especially of the BIOCLIM_REMOTE approach was rather small. For example, the best prediction of Aranda and Lobo (2011) who modeled diversity of 841 plants in Tenerife – based on stacked species distribution models but without remote sensing data – had a mean species richness error of 24%, while this study yielded an absolute mean error of 1.5 species per site for a total of 40 species.

The models predicted the greatest species richness of pines in the *Sierra Madre Occidental,* which has already previously been proposed as the major north-south migration route of pines (Perry et al., 1998) during the Tertiary (and presumably Cretaceous). During their migration the pines encountered warm temperate to subtropical climates – possibly one of the reasons why Mexico became a diversification center of pines (Perry et al., 1998). In addition to species richness, also species composition varies largely across the country. The complex topography (see Section 2.1) appears to be a natural dispersal barrier, since many pines have comparatively low dispersal abilities (Ledig, 1998). The functional traits (see Section 1.1) of the pine study species differ, for example regarding dispersal strategies (Anemochory, Zoochory) or specific adaptations to drought and fire. Even though pine species all share very similar structural characteristics within the respective plant communities, nature conservation is therefore interested in the identification and conservation of highly diverse spots. In addition to its effects on current functioning of ecosystems, species diversity further influences the resilience and resistance of ecosystems to environmental and climate change and to biological invasions (Chapin III et al., 2000; see Chapter 1). Unlike direct macroecological modeling of biodiversity (see Section 4.5), the stacked SDM approach not only allows producing spatial predictions of species richness but also mapping ecological traits associated with each species. For example in the context of climate change and its effects on species distributions, this knowledge of species characteristics, such as adaptation potential and dispersal strategies, is of high relevance.

Future research. While there has been significant progress concerning novel algorithms and environmental predictors in modeling the spatial distributions of single species, the prediction of species richness based on stacked SDMs remains challenging.

Especially for the numerous taxa and areas where systematic field inventories of species diversity are missing, the exploitation of geo-referenced museum and herbarium collections and presence-only field observations appears to be a very good alternative to expand the knowledge of species richness. In addition, further research should include assessments of compositional error, i.e. the percentage of species erroneously predicted present or absent from a cell. Guisan and Rahbek (2011) recently proposed a new theoretical framework called *SESAM* (*Spatially explicit species assemblage modeling*) that integrates stacked SDM with macroecological modeling, by integrating e.g. species dispersal and biotic filtering. Adding such macroecological constraints is expected to increase the prediction success of realized species richness and opens the way for new innovative research.

7.9 Summary

This chapter described the application of the so-called *stacked species distribution modeling* approach (Guisan and Rahbek, 2011; see Section 4.5) for modeling species richness based on presence-only data. Since two completely independent species occurrence data sets, namely the presence-only *National Information System on Mexican Biodiversity* and the presence-absence *National Forest Inventory* were available for this thesis, the unique data setting was used to examine the potential of older presence-only (primarily herbarium) collection data – that are the only available species data source for many countries and for the majority of taxa – to predict current plant species richness.

In this analysis, the first Mexican species richness map of the tree genus *Pinus* (family Pinaceae) was produced based on 40 individual range maps predicted in a series of presence-only species distribution models. Climatic suitability based on long-term climate data (BIOCLIM) as well as remotely sensed biotope availability (REMOTE) were modeled independently with the *Maximum Entropy* (Maxent) algorithm. Maxent models developed with the BIOCLIM data set performed slightly better (mean training AUC = 0.958; mean test AUC = 0.936) than those computed with the remote sensing predictor set (mean training AUC = 0.946; mean test AUC = 0.912). BIOCLIM and REMOTE predictions were then combined in a hierarchical framework to derive current pine species distributions (BIOCLIM_REMOTE). The species distribution ranges modeled with this hierarchical combination of both bioclimatic and remote sensing data yielded the best accuracy with the lowest overall deviance of 0.169 from presence-absence data across all species. Model evaluation of predicted species richness was accomplished with independently sampled species richness data that was temporally consistent with the acquisition period of the remote sensing data. A site-specific evaluation showed that the combined model was significantly better in predicting species richness than the purely bioclimatic model ($p < 0.001$, Mann-Whitney U-test). The combined species richness estimations showed a spatial agreement with the major mountain ranges in Mexico but over-estimated maximum species richness with 12 species per cell compared to eight species observed in a systematic field inventory of tree species diversity (the Mexican *National Forest Inventory*). The mean difference to observed species richness data was 3.1 species per site (median = 2) for BIOCLIM and 1.5 species per site (median = 0) for BIOCLIM_REMOTE. The analysis presented in this chapter hence supported the potential of remote sensing data not only for obtaining better estimates of individual species distribution ranges but also for predicting species richness.

8 Synthesis and outlook

As outlined in the beginning, the objective of this thesis was to develop a robust and transferable framework for modeling species distributions and species richness with multi-temporal remote sensing data. This chapter summarizes the major methodological findings and highlights the potential and limitations of remotely sensed data for the prediction of tree species distribution ranges and richness in Mexico (Section 8.1). Furthermore, the usefulness of the modeling framework developed in this thesis for biodiversity assessments, conservation planning, and land management are outlined (Section 8.2). Finally, the (taxonomic and geographical) transferability of the approach and new research perspectives building on the results of this thesis are opened up and discussed with respect to future sensor developments in remote sensing (Section 8.3).

8.1 Potential and limitations of remote sensing data for modeling tree species distributions and richness

Hierarchical combination of climate and remote sensing data. The analysis conducted in this thesis showed that climate data exhibited significantly higher spatial autocorrelation than remote sensing variables due to the lower *sample densities* (Franklin, 2009) of interpolated climatic surfaces. In accordance with theoretical frameworks concerning the importance of certain environmental determinants of species distributions (e.g. Pearson and Dawson, 2003), these findings made clear that climatic and remote sensing predictors should not be combined in the same species distribution model. As species occurrence data, too, typically feature spatial autocorrelation, test models which integrated climate and remote sensing data in a single, non-hierarchical approach systematically overestimated the importance of spatially autocorrelated climate data (a phenomenon termed the 'red-shift'; Lennon, 2000). A non-hierarchical model design would therefore not fully exploit the potential of remote sensing data.

Measured against the statistical model deviance from the observed presence and pseudo-absence data, the synergistic hierarchical modeling approach outperformed both purely climatic and remote sensing based models. Within the hierarchical framework, the multi-temporal remote sensing imagery facilitated the identification of limiting environmental conditions beyond climatic suitability and allowed for the extraction of phenological and seasonal variations in vegetation cover that affect and mirror geographical species distributions. In agreement with Saatchi et al. (2008), these contributions of remotely sensed imagery can be summarized as: (1) improving the effective spatial resolution of the environmental predictors from interpolated surfaces (as for climate data) to direct geo-referenced measurements and (2) amplifying the pool of environmental predictors by adding additional information layers in order to better characterize biotope characteristics. The integration of remote sensing predictors particularly allowed for the incorporation of disturbance factors, e.g. due to human activities.

Furthermore, analogous to the findings of Pearson et al. (2004) for land cover data, the inclusion of remote sensing predictors in the hierarchical modeling framework reduced the *fractional predicted area*. However, the reduction in predicted distribution range size varied largely between species, depending on their ecological requirements and the characteristics of their distributional areas. For example, the reduction was by far highest (80.7%) for the mangrove species *Avicennia germinans* which has very specific biotope requirements that cannot be characterized only by climate. For the 40 Mexican pines, the obvious (given the estimated size of remnant pine-oak forests in Mexico) overestimation of species ranges by climatic species distribution models could be reduced to reasonably sized predictions through the integration of remote sensing data. The omission of climate data in the hierarchical framework, though, led to a general overprediction of species distribution ranges. This overestimation was caused by the occurrence of similar vegetation characteristics and dynamics across different plant communities and by the non-detection of limiting factors (e.g. physiological constraints) in remote sensing data.

Within the hierarchical modeling scheme, a species-specific analysis of predictor correlation was carried out. Predictor covariation hampers the interpretability of species-environment relationships and significantly impacts model predictions (Heikkinen et al., 2006; Baldwin, 2009) but has been largely ignored for models including remote sensing data (but see Buermann et al., 2008). The results highlighted the significant amount of covariation within both climatic predictors and remote sensing variables and hence indicated data redundancy.

At all stages of the hierarchical modeling framework, species characteristics (particularly range size) had a large impact on model accuracy. The hierarchical modeling approach confirmed that range size affected the ability to accurately model the target species with climate data, remote sensing variables, as well as the hierarchical combination of both. This is consistent with previous studies which found that stenoecious species (those with narrow or more specialized niches) and species with restricted range sizes can be better modeled based on topo-climatic data (Luoto et al., 2005; Guisan et al., 2007; Tsoar et al., 2007; Syphard and Franklin, 2010) or even based on a combination of climate and remote sensing data in a non-hierarchical model (McPherson and Jetz, 2007).

Categorical or continuous? – A comparison of land cover classification data and continuous remote sensing variables. The analysis conducted in this thesis provided answers to the – in the existing literature only theoretically discussed – question under what circumstances it is better to use continuous remote sensing or categorical land cover data for modeling species distributions. This novel study was carried out for the twelve target tree species typical of the major Mexican forest types. The analysis compared the performance of species distribution models developed from either the

NALCMS land cover product (NALCMS, 2005), which is the most accurate data set currently available for Mexico with an overall accuracy of 82% (Colditz et al., 2010), or from phenological and statistical metrics derived from remote sensing time series. The results showed that models based on the categorical land cover data in general overestimated species distributions due to cartographic generalization and insufficient thematic detail of the land cover classification. The suitability of each respective land cover product to predict species distributions is therefore dependent on its mapping accuracy and on the detail (number of classes) and validity (significance of the class definitions) of its legend. In view of the generally higher mapping accuracies and greater thematic detail of regional land cover data, the use of regional or even continental rather than global land cover products is therefore recommended in species distribution modeling. Even though this analysis was conducted only for a specific land cover product, similar results can be expected for other land cover classifications.

The trend towards overestimation of species distribution ranges based on categorical land cover information was found to be influenced as well by the specific characteristics of the target species. In addition to the specifications of the land cover product, the performance of land cover data in species distribution modeling hence also depends on how closely observed species distribution can be linked to certain land cover types. As a measure to quantify this relationship between species occurrence and land cover, a novel *land cover evenness index* (following Pielou's species evenness index; Pielou, 1966) was introduced in this thesis. Species with a low land cover evenness index (implying the dominance of a single or of few land cover classes at the species localities), e.g. *Astronium graveolens* and *Avicennia germinans*, and with a high mapping accuracy of the respective dominant land cover type(s) exhibited the highest suitability of being modeled based on land cover data. A linear correlation between the land cover evenness index and model improvement measured by AUC (*area under curve*, see Section 3.4.5) was consistently observed for all study species.

Certainly, since the pre-processing effort of remote sensing data is high compared to often readily available categorical land cover data, a trade-off situation is created between target model accuracy and processing effort. Researchers should thus make their choice case-by-case depending on the linkage between species records and land cover, that is, on the ascertained score of the species-specific land cover evenness index.

Which data for which species? – The impact of species traits on selecting remote sensing predictors. Despite the growing interest in understanding the importance of remotely sensed variables for modeling species distributions, there exists as of yet no systematic assessment of predictors suitable for modeling plant species with certain traits.

This thesis could demonstrate that predictor importance varied significantly between species depending on their ecological traits. This underlines the importance of a species-specific variable selection based on the explanatory power of the predictors which has been neglected in remote sensing based species distribution modeling so far, but was implemented in this thesis. It was shown that vegetation indices are the most widely used but not necessarily the best predictors of plant species distributions. Instead, the feature ranking indicated a high potential of *Land Surface Temperature* (LST) data particularly for temperate as well as tropical montane tree species. Consistent with this, LST metrics were the most important remote sensing predictors for explaining the distributions across all 40 pine species modeled (see Chapter 7). Pines are found primarily in temperate and montane climates at higher elevations in Mexico's major mountain ranges, where LST showed an annual seasonal cycle which could be differentiated from the mean across the study area. *Net Primary Productivity* (NPP)-related phenological metrics derived from the *Enhanced Vegetation Index* (EVI) were the best predictors for species of the tropical evergreen forest (e.g. *Bursera simaruba* and *Cedrela odorata*). EVI metrics related to vegetation seasonality were of major importance only for *Bursera bipinnata* occurring in tropical deciduous forests with a distinct phenological cycle resulting from pronounced dry and rainy seasons. Except for the mangrove species *Avicennia germinans*, found in coastal wetlands, spectral bands were of minor significance for modeling tree species distributions.

On the basis of these results, remote sensing products should be purposefully selected according to the known characteristics of the respective vegetation community (e.g. tropical compared to temperate climates; deciduous as opposed to evergreen vegetation types). On doing so, the outcomes and recommendations of previous vegetation mapping studies with a remote sensing background (Cingolani et al., 2004; Yang, 2007; Hüttich et al., 2009) are very useful. The analysis of species response curves for remotely sensed variables revealed determinate (unimodal or monotone) response shapes and thus supports the usefulness of the remote sensing variables. The testing of further remote sensing products for developing an ecologically-based understanding of remote sensing parameters is hence explicitly suggested.

How many data? – Impact of inter-annual variability and length of remotely sensed vegetation index time series. The inter-annual variability in time series of the *Enhanced Vegetation Index* observed in Mexico was mainly a result of variation in precipitation, which in large parts of the country is triggered by the El Niño/Southern Oscillation (ENSO) phenomenon. The systematic study design, aimed at assessing the effects of inter-annual variability on species distribution models, included the analysis of phenological metrics as single-year data sets and combined into multi-year average values (covering three, five, seven, and nine consecutive years). In general, the use of longer time series improved the predictive performance of species distribution models. Such multi-year values reflect long-term conditions rather than annual fluctuations and are thus relevant for the establishment, survival, and reproduction of individual trees. In line with this, especially short time periods (one or three years) showed the highest statistical model deviance from validation data as well as considerable differences between composite periods of the same length. In addition, the reduction in model variance as a consequence of using longer remote sensing time series varied greatly between species. Wide-ranging species with high numbers of presence records (e.g. *Bursera simaruba*, *Astronium geaveolens* or *Cedrela odorata*) in general featured hardly improved model deviance scores. This fact indicates that representative phenological metrics can be obtained by averaging in time (longer time series) or space (many presence sites).

Furthermore, sample size had effects on the variation in model deviance within one composite period with higher numbers of species records indicating less variation. *Arbutus xalapensis* in this sense was certainly an exceptional case. Even though the species had the second largest sample size, especially models developed from either one or three years of time series showed strong differences in model deviance. The species showed this high variation presumably because its distribution range expands far to the north (see Appendix I), which is more affected by inter-annual variability of precipitation than Mexico's south.

Between remotely sensed time series of the same length, temperate species (*Abies religiosa*, *Alnus acuminata*, *Arbutus xalapensis*, and *Pinus chiapensis*) as well as the mangrove species *Avicennia germinans* featured higher variability in model deviance than tropical species. To summarize, the following criteria indicate that distribution models of a certain species are prone to be affected by inter-annual variability in remote sensing data: small sample size, occurrence in temperate environments, and generally high environmental variability within the respective ecosystem or distributional range.

From species to species richness? – A case study for Mexican pines. This thesis further investigated whether the hierarchical framework for modeling species distributions, which was implemented here, could be transferred to create maps of richness by overlaying individual species ranges as well. This analysis represents the first application of the so-called *stacked species distribution modeling* approach (Guisan and Rahbek, 2011) in combination with remote sensing data across a large spatial extent. Mexico features an enormous diversity of pine species; however, little is known about the exact distribution of pine diversity. The study generated the first Mexican species richness map of the tree genus *Pinus* (family Pinaceae) based on remote sensing data. Even though pine species all share very similar structural characteristics within the plant community, their functional traits (discussed in the introductory chapter) differ.

Without rules (e.g. in relation to the maximum *carrying capacity* or the competitive exclusion of species) to constrain the maximum species richness per grid cell, stacked species distribution models typically tend to overpredict species richness (Diniz-Filho et al., 2009; Dubuis et al., 2011). In accordance with previous climate based modeling studies (Graham and Hijmans, 2006; Algar et al., 2009; Newbold et al., 2009; Pineda and Lobo, 2009; Raes et al., 2009), the results of this thesis confirmed the systematic bias towards overestimation of species richness also for remote sensing based approaches. However, the species distribution ranges modeled with the hierarchical combination of both climatic and remote sensing data yielded the lowest overall model deviance from independent presence-absence data across all species. Beyond this, the site-specific evaluation of modeled species richness scores showed that the combined model was significantly better at predicting species richness than the purely climatic model. The inclusion of remote sensing data hence considerably improved the reliability of the species richness estimates. The applied methodology has two major advantages over 'traditional' approaches to predict species diversity based on spatial interpolation or macroecological modeling using remote sensing data (see Section 4.5): (1) the modeling framework can make use of presence-only collection data – which are the only available species data source for many countries and for the majority of taxa – and (2) the approach allows to derive information on actual species composition. Consistent with the known species records, the models predicted the greatest species richness of pines in the *Sierra Madre Occidental*, which has already previously been proposed as the major north-south migration route of pines (Perry et al., 1998).

Limitations. Modeling species distribution based on remote sensing data has limitations as well. First of all, the usefulness of remote sensing variables for modeling species distributions depends on their data quality; a rigorous quality assessment and meaningful interpolation of time series data as applied in this thesis is therefore mandatory. Over tropical areas, modeling based on multi-temporal data may be limited by persistent cloud coverage over longer timeframes – which, in this thesis, was of relevance only in the most southern areas. Further, there are limitations regarding spatial and temporal resolution of the remote sensing data depending on the given extent of a study area. Temporarily continuous remotely sensed time series as analyzed in this study are currently only operationally available at medium (250 m) to coarse (1 km) resolution, which may not be fine enough for certain conservation planning projects. This thesis showed that longer remotely sensed time series generally improved models, though this may only be assumed under comparatively stable environmental conditions at the species presence sites. Finally, this thesis confirmed that not only climatic species distribution models (Hernandez et al., 2006; Wisz et al., 2008) but also models developed from remote sensing data are sensitive to the number of species records available. The spatial distributions of common and widespread species are therefore more difficult to model – be it with climate or remote sensing data. On the contrary, modeling rare species is limited by the minimum sample size required by the modeling algorithm to produce reliable predictions. Given the findings of this thesis, one should also be skeptical about models based purely on remote sensing data over large areas due to the significant spatial overprediction of such models.

8.2 Usefulness for ecological research and nature conservation

It has been claimed that policy decisions in Mexico should be based on reliable scientific data and meet the criteria relevant for the establishment of sustainable protected areas (Sarukhán et al., 2010). The findings of this thesis can effectively support such nature conservation and land management efforts.

Biodiversity inventories and monitoring. The spatially more accurate prediction of species distribution ranges based on the synergistic combination of climatic and remote sensing data can significantly support species and biodiversity inventories in the future – in biodiversity hotspots such as Mexico and elsewhere. Given the progressive loss of natural areas due to the human alteration of the global environment (see Chapter 1), the establishment of baseline datasets regarding the current extent of suitable habitat for a given species using recent remote sensing data is required (Strand et al., 2007). On the basis of such reliable status-quo assessments, the increasing availability of temporally continuous remote sensing data may support the monitoring of biodiversity hotspots. The existence of remote sensing data with high temporal resolutions also allows for a detailed monitoring of the spatial extent and the conditions of habitats and biotopes, particularly in inaccessible areas that cannot be surveyed in the field (Cayuela et al., 2006; Saatchi et al., 2008). Especially the successful implementation of a *stacked species distribution modeling* approach (Guisan and Rahbek, 2011) for the prediction of species richness (here for Mexican pines) based on presence-only data opens up the opportunity for biodiversity inventories in countries and for taxa that have not been investigated yet.

Conservation planning and land management. "Vast gaps in available information on the spatial distribution of biodiversity pose a major challenge for regional conservation planning in many parts of the world" (Ferrier, 2002). The establishment of protected areas, however, is a 'cornerstone' (Margules and Pressey, 2000) or 'vital component' (Ferrier, 2002) of conservation strategies to mitigate the ongoing loss of biodiversity. The modeling framework developed in this thesis can facilitate conservation assessments as it provides information as to where anthropogenic threats and the distributional ranges of certain species of high biodiversity coincide. The respective methodological findings can effectively support conservation planning and land management in many respects – for example, a nation-wide optimization approach for the conservation of Mexican tree species as intended by Ricker et al. (2007). In their study, Ricker et al. (2007) used categorical vegetation type data in combination with assigned species presence or absence (in squares of 1° latitude by 1° longitude) to propose a reserve network. Applying the automated modeling framework of this thesis to the same species would yield modeled species distribution ranges (1) with a 10,000 times higher spatial resolution and (2) based on continuous remote sensing data instead of categorical vegetation type data. At the local scale, too, namely in Chiapas (Mexico), remote sensing data have already been used to define priority areas for conservation based on criteria related to maximized tree species richness (Cayuela et al., 2006). Such efforts could be brought up to a larger spatial extent using the modeling framework developed in this thesis. However, the approach is not limited to the delineation of the reserves but may also be valuable for monitoring their effectiveness and their surroundings, which may serve as buffer zones and thus also play a major supporting role in conserving biodiversity (Putz et al., 2001).

In addition, non-native, invasive species increasingly pose a challenge to land management as they threaten native biodiversity and impact the environment by affecting ecosystem processes (Ficetola et al., 2007; National Advisory Committee on Invasive Species, 2010). Given the ecological and economic consequences of invasive species, the prevention of establishment (Leung et al., 2002) and the prediction of the probability of successful establishment (Thuiller et al., 2005) of species with a high risk of becoming invasive are the most cost-effective ways. Remote sensing has a great potential for mapping and modeling invasive species distributions (Joshi et al., 2004; Morisette et al., 2005; Jones and Reichard, 2009; Kerns et al., 2009) and can thus support 'early warning systems' (Thuiller et al., 2005) concerning invasive species. The approach developed in this thesis was already successfully applied to predict areas prone to invasion by the plant species *Tamarix* spp. (tamarisk) in Mexico (Cord et al., 2010).

Guidance of new field surveys. Species distribution models hold a substantial potential for detecting unknown distributional areas, populations or species (Fleishman et al., 2002; Raxworthy et al., 2003; Engler et al., 2004; Bourg and McShea, 2005) and can thus be employed for the strategic identification of additional survey sites. Making use of the increased sampling efficiency, newly collected data can then be utilized to improve the initial model (Guisan et al., 2006). Remote sensing based approaches for species distribution modeling can contribute in particular as they are useful for identifying areas with suitable biotope conditions similar to those known to be occupied by a certain species (Strand et al., 2007). In this sense, remote sensing in combination with species distribution models offers “innovative potential for the discovery of unknown distributional areas and unknown species” (Raxworthy et al., 2003: p. 837). The hierarchical approach developed in this thesis can effectively facilitate such investigative efforts as it identifies areas that have both suitable climate conditions and similar biotope characteristics to those observed at the species presence sites. Certainly, remote sensing should aim at complementing rather than replacing ground surveys and field-based methodologies (Gillespie et al., 2008).

8.3 Future research needs

This thesis systematically assessed the potential of multi-temporal remote sensing data for modeling tree species distributions in Mexico and developed a framework for the optimized use of remote sensing data to predict species distributions and species richness. Future research perspectives building upon the major findings of this thesis include different stages, namely to examine the transferability of the framework as it is, to broaden the array of remote sensing data, to implement multi-scale models, and to incorporate species mobility.

Taxonomic, spatial, and temporal transferability. Since the target species of this thesis were selected to be representative of several forest types and to cover a wide range of ecological traits such as range size or biotope specificity, the findings of this study possess an inherent generalizability. However, the next step is surely the testing of its taxonomic transferability, i.e. the application of the modeling framework to a multiplicity of additional plant species (in particular other life forms such as shrubs and herbaceous vegetation) to confirm the preliminary findings. Especially studies on plant species that are more affected by inter-annual environmental variability than the tree species considered here, e.g. xerophilous vegetation, are required. Future research in this respect can make use of the unique availability of species occurrence data from different sources in Mexico (see Section 1.3). Regarding the prediction of species richness based on stacked species distribution models, it would furthermore be of high interest to utilize the framework for modeling species numbers across several plant families. The modeling framework has been specifically developed for plant species; the transferability to other taxa is assumed but remains to be verified. For animals, similar remote sensing products may be of relevance as animals also require the presence of certain vegetation communities, for example as a resource for foraging.

With regard to the spatial transferability, the use of multi-temporal remote sensing data certainly has several advantages. The data capture phenological variety throughout the year and therefore no prior knowledge or testing on the optimum dates to detect differences between species presence and background sites is required. In addition, Mexico exhibits a great environmental and biological diversity that is reflected in an enormous variety of ecological processes. It is precisely because of this topographical and climatological heterogeneity, that it would be interesting to transfer the framework and apply it in another – climatically very homogeneous – geographical region. As shown for the Yucatan peninsula with relatively uniform climatic conditions, remote sensing data were in this case particularly useful in providing additional constraints on the species ecological requirements. Similar findings – especially for widespread species – were obtained by the studies on tree species in the Amazon basin of Buermann et al. (2008) and Saatchi et al. (2008).

Inter-annual variability in remotely sensed time series data poses a great challenge in species distribution modeling. The improvement of model reliability together with the increasing length of the composite period of phenological metrics, as observed in this thesis, supports the need for continuous long-term time series of satellite remote sensing data. As proposed by Tuanmu et al. (2011), the evaluation of temporal model transferability depending on the length of the remotely sensed time series is therefore of high interest. Such analyses can be performed in the future once species occurrence data from different sampling periods become available, e.g. in the form of updated data of the Mexican *National Forest Inventory*.

Array of suitable remote sensing data. In view of the ‘perception problem’ between ecologists and remote sensing researchers mentioned in the introductory chapter, more studies on the ecological relevance of remote sensing variables regarding species distribution ranges and species diversity are urgently required. The array of available remote sensing products is by far not fully exploited in the existing literature on modeling species distributions. Besides the remote sensing products studied in this thesis, LAI (*Leaf Area Index;* Prates-Clark et al., 2008; Saatchi et al., 2008), FPAR (*Fraction of absorbed Photosynthetically Active Radiation*; Coops et al., 2009), and fractional vegetation cover (Buermann et al., 2008) have also been shown to be suitable for modeling species distributions. According to this thesis, there is a high potential especially in multi-annual remote sensing data for characterizing species distributions. Future research should thus focus on the potential of longer time series to predict species distribution ranges as well as certain seasonal dynamics or extreme events that can be detected by means of remote sensing and might be limiting species occurrences.

Finally, the synergistic combination of optical and radar data (see for example Buermann et al., 2008) could open up new possibilities for characterizing environmental gradients relevant for species distributions and diversity. Active radar measurements have already been used to estimate forest species composition and abundance (Wolter and Townsend, 2011) and provide complementary information for species distribution modeling as they explicitly quantify structural characteristics of vegetation (Bradley and Fleishman, 2008). LiDAR (*Light Detection and Ranging*) measurements are a very interesting alternative for characterizing canopy and understory vegetation structure

(Nilsson, 1996; Lefsky et al., 2002) and have been shown to be useful to predict coastal plant species (Sellars and Jolls, 2007) and forest habitat types (Bässler et al., 2011) at local scale. However, LiDAR data are only recorded from airborne sensors and hence not available across large extents as typically required for species distribution models. Despite their constraints in geographical coverage, upcoming multispectral (e.g. Sentinel-2) or hyperspectral (e.g. EnMAP) satellite missions will certainly enlarge the array of suitable remote sensing data. Novel remote sensors, indices, or data products should be utilized according to their ecological relevance for characterizing species distribution ranges and species richness.

Multi-scale models incorporating remote sensing data from different sensors. Besides the ecological relevance of remote sensing variables, effects of the spatial resolution of the remote sensing data on species distribution models are of high interest, too. The current state of the art (this thesis and the studies of Buermann et al., 2008; Prates-Clark et al., 2008; Foley et al., 2009; Cord and Rödder, 2011) is the use of climate data and remote sensing variables at equal or similar spatial resolutions (typically approximately 1 km). The proposed inclusion of higher resolution remote sensing data or even multi-scale modeling frameworks should allow improving species distribution models especially at landscape to regional scales. However, even though 250 m resolution MODIS products are readily available, it has to be considered that only Band 1 (red) and Band 2 (NIR) possess this original resolution of 250 m; all other bands are resampled. In addition, the following caveat need to be considered: environmental variables must be addressed at an appropriate scale (1) for characterizing species distributions (Pearson et al., 2004) and (2) in relation to the positional accuracy of the species records. The use of very high spatial resolution sensors (e.g. IKONOS, Quickbird, and GeoEye) thus seems pointless.

The use of such nested, multi-scale frameworks in species distribution modeling, in particular for plants, would cover novel technical as well as ecological aspects. From the technical point of view, insights into the sensitivity of species distribution modeling algorithms regarding the remotely sensed signatures could be gained. In other words, which proportion or possibly even spatial configuration of predicted presence in high spatial resolution data is required in order to predict the presence of the respective species also in the coarse resolution data? From the ecological perspective, such a hierarchical framework will lay the foundation of a new understanding of the determinants of species distributions – both across scales and within scales (e.g. regarding edge effects).

Incorporation of species mobility and modeling species migrations. Remote sensing data and derived information on land surface characteristics can hardly be predicted into the future (i.e. temporal extrapolation) in the way climate data can be. However, models developed with the methodological framework applied in this thesis can help identifying available species migration corridors to new habitats under climate change. One of the remaining challenges is hence to reliably incorporate species mobility or migrations into the prediction of geographical distribution ranges. Besides *biotic* and *abiotic* environmental conditions, species *mobility* is the third essential determinant of species distributions as it limits the area accessible to the target species

(*BAM* concept, Soberón and Peterson, 2005). In species distribution modeling studies incorporating remote sensing variables, dispersal capacities have been largely ignored even though the respective model predictions offer a unique opportunity to value current landscape connectivity and fragmentation (Strand et al., 2007). Predictions of species distributions developed from remote sensing data could thus effectively be used as a cost matrix for dispersal and migration scenarios, the definition of patches, and also for a more realistic estimation of ecological distances between patches in order to improve the estimation of dispersal success (Ferrier et al., 2002).

References

Albright, T. P., A. M. Pidgeon, C. D. Rittenhouse, M. K. Clayton, C. H. Flather, P. D. Culbert, and V. C. Radeloff (2011). Heat waves measured with MODIS land surface temperature data predict changes in avian community structure. *Remote Sensing of Environment 115*, 245 – 254.

Algar, A. C., H. M. Kharouba, E. R. Young, and J. T. Kerr (2009). Predicting the future of species diversity: Macroecological theory, climate change, and direct tests of alternative forecasting methods. *Ecography 32*, 22 – 33.

Andersen, M. C., H. Adams, B. Hope, and M. Powell (2004). Risk assessment for invasive species. *Risk Analysis 24*(4), 787 – 793.

Anderson, R. P., D. Lew, and A. T. Peterson (2003). Evaluating predictive models of species' distributions: Criteria for selecting optimal models. *Ecological Modelling 162*, 211 – 232.

Anderson, R. P. and E. Martínez-Meyer (2004). Modeling species' geographic distributions for conservation assessments: An implementation with the spiny pocket mice (*Heteromys*) of Ecuador. *Biological Conservation 116*, 167 – 179.

Anselin, L. (1995). Local Indicators of Spatial Association – LISA. *Geographical Analysis 27*, 93 – 115.

Araújo, M. B., M. Cabeza, W. Thuiller, L. Hannah, and P. Williams (2004). Would climate change drive species out of reserves? An assessment of existing reserve-selection methods. *Global Change Biology 10*, 1618 – 1626.

Araújo, M. B. and R. G. Pearson (2005). Equilibrium of species' distributions with climate. *Ecography 28*, 693 – 695.

Araújo, M. B., R. G. Pearson, W. Thuiller, and M. Erhard (2005). Validation of species-climate impact models under climate change. *Global Change Biology 11*, 1504 – 1513.

Araújo, M. B., W. Thuiller, and R. G. Pearson (2006). Climate warming and the decline of amphibians and reptiles in Europe. *Journal of Biogeography 33*, 1712 – 1728.

Araújo, M. B. and P. H. Williams (2000). Selecting areas for species persistence using occurrence data. *Biological Conservation 96*, 331 – 345.

Aranda, S. C. and J. M. Lobo (2011). How well does presence-only-based species distribution modelling predict assemblage diversity? A case study of the Tenerife flora. *Ecography 34*, 31 – 38.

Asner, G. P., M. O. Jones, R. E. Martin, D. E. Knapp, and R. F. Hughes (2008). Remote sensing of native and invasive species in Hawaiian forests. *Remote Sensing of Environment 112*, 1942 – 1955.

Asner, G. P., A. R. Townsend, and B. H. Braswell (2000). Satellite observation of El Niño effects on Amazon forest phenology and productivity. *Geophysical Research Letters 27*(7), 981 – 984.

Austin, M. (2007). Species distribution models and ecological theory: A critical assessment and some possible new approaches. *Ecological Modelling 200*, 1 – 19.

Austin, M. P. (1980). Searching for a model for use in vegetation analysis. *Vegetatio 42*, 11 – 21.

Austin, M. P. (2002). Spatial prediction of species distribution: An interface between ecological theory and statistical modelling. *Ecological Modelling 157*, 101 – 118.

Austin, M. P., A. O. Nicholls, and C. R. Margules (1990). Measurement of the realized qualitative niche: Environmental niches of five *Eucalyptus* species. *Ecological Monographs 60*(2), 161 – 177.

Austin, M. P. and T. Smith (1989). A new model for the continuum concept. *Vegetatio 83*, 35 – 47.

Bahn, V., R. J. O'Connor, and W. B. Krohn (2006). Importance of spatial autocorrelation in modeling bird distributions at a continental scale. *Ecography 29*, 835 – 844.

Bailey, S.-A., R. H. Haines-Young, and C. Watkins (2002). Species presence in fragmented landscapes: Modelling of species requirements at the national level. *Biological Conservation 108*, 307 – 316.

Baldwin, R. A. (2009). Use of maximum entropy modeling in wildlife research. *Entropy 11*, 854 – 866.

Balmford, A. (2002). Economic reasons for conserving wild nature. *Science 297*, 950 – 953.

Barbosa, H. A., A. R. Huete, and W. E. Baethgen (2006). A 20-year study of NDVI variability over the Northeast Region of Brazil. *Journal of Arid Environments 67*, 288 – 307.

Baret, F. and G. Guyot (1991). Potentials and limits of vegetation indices for LAI and APAR assessment. *Remote Sensing of Environment 35*, 161 – 173.

Barnes, W. L., T. S. Pagano, and V. V. Salomonson (1998). Prelaunch characteristics of the Moderate Resolution Imaging Spectroradiometer (MODIS) on EOS-AM1. *IEEE Transactions on Geoscience and Remote Sensing 36*(4), 1088 – 1100.

Barnes, W. L., X. Xiong, and V. V. Salomonson (2003). Status of Terra MODIS and Aqua MODIS. *Advances in Space Research 32*(11), 2099 – 2106.

Barthlott, W., A. Hostert, G. Kier, W. Küper, H. Kreft, J. Mutke, M. D. Rafiqpoor, and J. H. Sommer (2007). Geographic patterns of vascular plant diversity at continental to global scales. *Erdkunde 61*(4), 305 – 315.

Barthlott, W., J. Mutke, M. D. Rafiqpoor, G. Kier, and H. Kreft (2005). Global centres of vascular plant diversity. *Nova Acta Leopoldina 92*, 61 – 83.

Beaumont, L. J., L. Hughes, and M. Poulsen (2005). Predicting species distributions: Use of climatic parameters in BIOCLIM and its impact on predictions of species' current and future distributions. *Ecological Modelling 186*, 250 – 269.

Beck, P. S. A., E. Kalmbach, D. Joly, A. Stien, and L. Nilsen (2005). Modelling local distribution of an Arctic dwarf shrub indicates an important role for remote sensing of snow cover. *Remote Sensing of Environment 98*, 110 – 121.

Beissinger, S. R. and D. McCoullough (Eds.) (2002). *Population viability analysis*. University of Chicago Press, Chicago.

Berger, A. L., S. A. Della Pietra, and V. J. Della Pietra (1996). A maximum entropy approach to natural language processing. *Computational Linguistics 22*, 39 – 71.

Bhattacharyya, A. (1943). On a measure of divergence between two statistical populations defined by their probability distributions. *Bulletin of the Calcutta Mathematical Society 35*, 99 – 109.

Blanke, J. and T. Chiesa (2011). The travel & tourism competitiveness report 2011, Beyond the downturn. World Economic Forum, Geneva, Switzerland.

Boubli, J. P. and M. G. de Lima (2009). Modeling the geographical distribution and fundamental niches of *Cacajao* spp. and *Chiropotes israelita* in northwestern Amazonia via a maximum entropy algorithm. *International Journal of Primatology 30*, 217 – 228.

Bourg, N. A. and W. J. McShea (2005). Putting a CART before the search: Successful habitat prediction for a rare forest herb. *Ecology 86*(10), 2793 – 2804.

Box, E. O. (1981). Predicting physiognomic vegetation types with climate variables. *Vegetatio 45*, 127 – 139.

Box, E. O., B. N. Holben, and V. Kalb (1989). Accuracy of the AVHRR vegetation index as a predictor of biomass, primary productivity and net CO_2 flux. *Vegetatio 80*, 71 – 89.

Boyce, M. S., P. R. Vernier, S. E. Nielsen, and F. K. A. Schmiegelow (2002). Evaluating resource selection functions. *Ecological Modelling 157*, 281 – 300.

Bozyk, P. (2006). *Globalization and the transformation of foreign economic policy*. Ashgate Publishing Limited, Hampshire.

Bradley, B. A. and E. Fleishman (2008). Can remote sensing of land cover improve species distribution modelling? *Journal of Biogeography 35*, 1158 – 1159.

Bradley, B. A., R. W. Jacob, J. F. Hermance, and J. F. Mustard (2007). A curve fitting procedure to derive inter-annual phenologies from time series of noisy satellite NDVI data. *Remote Sensing of Environment 106*, 137 – 145.

Brando, P. M., S. J. Goetz, A. Baccini, D. C. Nepstad, P. S. A. Beck, and M. C. Christman (2010). Seasonal and interannual variability of climate and vegetation indices across the Amazon. *Proceedings of the National Academy of Sciences of the United States of America 107*(33), 14685 – 14690.

Broennimann, O., U. A. Treier, H. Müller-Scharer, W. Thuiller, A. T. Peterson, and A. Guisan (2007). Evidence of climatic niche shift during biological invasion. *Ecological Letters 10*(8), 701 – 709.

Brotons, L., W. Thuiller, M. B. Araújo, and A. H. Hirzel (2004). Presence-absence versus presence-only modelling methods for predicting bird habitat suitability. *Ecography 27*, 437 – 448.

Bässler, C., J. Stadler, J. Muller, B. Forster, A. Gottlein, and R. Brandl (2011). LiDAR as a rapid tool to predict forest habitat types in Natura 2000 networks. *Biodiversity and Conservation 20*, 465 – 481.

Buermann, W., S. Saatchi, T. B. Smith, B. R. Zutta, J. A. Chaves, B. Mila, and C. H. Graham (2008). Predicting species distributions across the Amazonian and Andean regions using remote sensing data. *Journal of Biogeography 300*, 1 – 17.

Burgos, A. and J. M. Maass (2004). Vegetation change associated with land-use in tropical dry forest areas of Western Mexico. *Agriculture, Ecosystems and Environment 104*, 475 – 481.

Busby, J. R. (1991). *Nature conservation: Cost effective biological surveys and data analysis*, Chapter BIOCLIM - A bioclimatic analysis and prediction system, pp. 64 – 68. CSIRO, East Melbourne.

Carpenter, G., A. N. Gillison, and J. Winter (1993). DOMAIN: a flexible modelling procedure for mapping potential distributions of plants and animals. *Biodiversity and Conservation 2*, 667 – 680.

Carroll, S. S. and D. L. Pearson (1998). The effects of scale and sample size on the accuracy of spatial predictions of tiger beetle (Cicindelidae) species richness. *Ecography 21*, 401 – 414.

Cavazos, T. and S. Hastenrath (1990). Convection and rainfall over Mexico and their modulation by the Southern Oscillation. *International Journal of Climatology 10*, 377 – 386.

Cayuela, L., J. M. R. Benayas, A. Justel, and J. Salas-Rey (2006). Modelling tree diversity in a highly fragmented tropical montane landscape. *Global Ecology and Biogeography 15*, 602 – 613.

CEC (1997). Ecological regions of North America: Toward a common perspective. Technical report, Commission for Environmental Cooperation, Montreal, Canada.

Cevallos-Ferriz, S. R. S. and E. A. González-Torres (2005). *Studies on Mexican paleontology*, Chapter Geological setting and phytodiversity in Mexico, pp. 1 – 18. Springer, Dortrecht.

Chalkraft, D. R., J. W. Williams, M. D. Smith, and M. R. Willig (2004). Scale-dependency in the species-richness-productivity relationship: The role of species turnover. *Ecology 85*(10), 2701 – 2708.

Chapin III, F. S., E. S. Zavaleta, V. T. Eviner, R. L. Naylor, P. M. Vitousek, H. L. Reynolds, D. U. Hooper, S. Lavorel, O. E. Sala, S. E. Hobbie, M. C. Mack, and S. Díaz (2000). Consequences of changing biodiversity. *Nature 405*, 234 – 242.

Chefaoui, R. M. and J. M. Lobo (2008). Assessing the effects of pseudo-absences on predictive distribution model performance. *Ecological Modelling 2*, 478 – 486.

Chen, X.-L., H.-M. Zhao, P.-X. Li, and Z.-Y. Yin (2006). Remote sensing image-based analysis of the relationship between urban heat island and land use/cover changes. *Remote Sensing of Environment 104*, 133 – 146.

Chuine, I. (2010). Why does phenology drive species distribution? *Philosophical Transactions of the Royal Society: Biological Sciences 365*, 3149 – 3160.

Chuine, I. and E. G. Beaubien (2001). Phenology is a major determinant of tree species range. *Ecology Letters 4*, 500 – 510.

CIA (2011). The world factbook 2011. Technical report, Central Intelligence Agency, Washington DC, USA.

Cihlar, J., L. St.-Laurent, and J. A. Dyer (1991). Relation between the Normalized Difference Vegetation Index and ecological variables. *Remote Sensing of Environment 35*, 279 – 298.

Cingolani, A. M., D. Renison, M. R. Zak, and M. R. Cabido (2004). Mapping vegetation in a heterogeneous mountain rangeland using Landsat data: An alternative method to define and classify land-cover units. *Remote Sensing of Environment 92*, 84 – 97.

Colditz, R. R. (2007). *Time series generation and classification of MODIS data for land cover mapping*. Ph. D. thesis, University of Würzburg, Germany.

Colditz, R. R. (2009). Phenology and phenological variability of Mexican ecosystems. In *Proceedings of the IEEE International Geoscience & Remote Sensing Symposium, July 12-17, 2009, Cape Town, South Africa*.

Colditz, R. R., C. Conrad, T. Wehrmann, M. Schmidt, and S. Dech (2008). TiSeG: A flexible software tool for time-series generation of MODIS data utilizing the Quality Assessment Science Data Set. *IEEE Transactions on Geoscience and Remote Sensing 46*(10), 20 – 30.

Colditz, R. R., A. Cord, C. Conrad, F. Mora, P. Maeda, and R. Ressl (2009). Analyzing phenological characteristics of Mexico with MODIS time series products. In *Proceedings of the MultiTemp2009, 5th International Workshop on the Analysis of Multitemporal Remote Sensing Images, July 28-30, 2009, Mystic, Connecticut, USA*.

Colditz, R. R., P. Maeda, G. López, I. Cruz, and R. Ressl (2010). Land cover classification of Mexico in the framework of the North American Land Change Monitoring System. In *Proceedings of 2010 ASPRS Annual Conference, April 26-30, 2010, San Diego, USA*.

Colwell, J. E. (1974). Vegetation canopy reflectance. *Remote Sensing of Environment 3*, 175 – 183.

CONABIO (2006). Densidad de Población por entidad federativa, 2000. Extraido del XII Censo de Población y Vivienda 2000 del Instituto Nacional de Estadística Geografía e Informática (INEGI). Technical report, Comisión Nacional para el Conocimiento y Uso de la Biodiversidad, Mexico-City, Mexico.

CONABIO (2011). Portal de Geoinformación, Sistema Nacional de Información sobre Biodiversidad, http://www.conabio.gob.mx/informacion/gis/.

CONAFOR (2009). Inventario Nacional Forestal y de Suelos, Manual y procedimientos para el muestreo de campo (RE-MUESTREO 2009). Technical report, Comision Nacional Forestal, Mexico-City, Mexico.

Congalton, R. G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment 37*, 35 – 46.

Connell, J. and E. Orias (1964). The ecological regulation of species diversity. *American Naturalist 98*, 399 – 414.

Conrad, C., S. W. Dech, M. Hafeez, J. Lamers, C. Martius, and G. Strunz (2007). Mapping and assessing water use in a Central Asian irrigation system by utilizing MODIS remote sensing products. *Irrigation and Drainage Systems 21*, 197 – 218.

Coops, N. C., M. A. Wulder, and D. Iwanicka (2009). Exploring the relative importance of satellite-derived descriptors of production, topography and land cover for predicting breeding bird species richness over Ontario, Canada. *Remote Sensing of Environment 113*, 668 – 679.

Cord, A., D. Klein, and S. Dech (2010). Remote sensing time series for modeling invasive species distribution: A case study of *Tamarix* spp. in the US and Mexico. In *Proceedings of the International Congress on Environmental Modelling and Software (iEMSs), July 5-8, 2010, Ottawa, Canada*.

Cord, A., D. Klein, and S. Dech (2011). The impact of inter-annual variability in remote sensing time series on modeling tree species distributions. In *Proceedings of the MultiTemp2011, 6th International Workshop on the Analysis of Multitemporal Remote Sensing Images, July 12-14, Trento, Italy*.

Cord, A. and D. Rödder (2011). Inclusion of habitat availability in species distribution models through multi-temporal remote sensing data? *Ecological Applications 21*(8), 3285 – 3298.

Costa, G. C., C. Nogueira, R. B. Machado, and G. R. Colli (2010). Sampling bias and the use of ecological niche modeling in conservation planning: A field evaluation in a biodiversity hotspot. *Biodiversity and Conservation 19*, 883 – 899.

Couturier, S., J.-F. Mas, E. López-Granados, J. Benítez, V. Coria-Tapia, and Á. Vega-Guzmán (2010). Accuracy assessment of the Mexican National Forest Inventory map: A study in four ecogeographical areas. *Singapore Journal of Tropical Geography 31*, 163 – 179.

Critchfield, W. and E. Little (1966). Geographic distribution of pines of the world. USDA Forest Service, Miscellaneous Publication, 991.

Dávila, P., M. del Coro Arizmendi, A. Valiente-Banuet, J. L. Villaseñor, A. Casas, and R. Lira (2002). Biological diversity in the Tehuacán-Cuicatlán Valley, Mexico. *Biodiversity and Conservation 11*, 421 – 442.

de Souza Muñoz, M. E., R. de Giovanni, M. F. de Siqueira, T. Sutton, P. Brewer, R. Scachetti-Pereira, D. A. Lange Canhos, and V. Perez Canhos (2011). openModeller: A generic approach to species' potential distribution modelling. *Geoinformatica 15*, 111 – 135.

Debinski, D. M., K. Kindscher, and M. E. Jakubauskas (1999). A remote sensing and GIS-based model of habitats and biodiversity in the Greater Yellowstone Ecosystem. *International Journal of Remote Sensing 20*, 3281 – 3291.

DeFries, R., M. Hansen, and J. Townshend (1995). Global discrimination of land cover types from metrics derived from AVHRR Pathfinder data. *Remote Sensing of Environment 54*, 209 – 222.

Dinerstein, E. and D. M. Olson (1999). *Terrestrial ecoregions of North America: A conservation assessment*, Chapter Globally outstanding biodiversity in our own backyard, pp. 93 – 99. Island Press, Washingon.

Diniz-Filho, J. A. F., L. M. Bini, and B. A. Hawkins (2003). Spatial autocorrelation and red herrings in geographical ecology. *Global Ecology and Biogeography 12*, 53 – 64.

Diniz-Filho, J. A. F., L. M. Bini, T. F. Rangel, R. D. Loyola, C. Hof, D. Nogués-Bravo, and M. B. Araújo (2009). Partitioning and mapping uncertainties in ensembles of forecasts of species turnover under climate change. *Ecography 32*, 897 – 906.

Dobson, A., D. Lodge, J. Alder, G. S. Cumming, J. Keymer, J. McGlade, H. Mooney, J. A. Rusak, O. Sala, V. Wolters, D. Wall, R. Winfree, and M. A. Xenopoulos (2006). Habitat loss, trophic collapse, and the decline of ecosystem services. *Ecology 87*(8), 1915 – 1924.

Dormann, C. F. (2007). Effects of incorporating spatial autocorrelation into the analysis of species distribution data. *Global Ecology and Biogeography 16*, 129 – 138.

Dormann, C. F., J. M. McPherson, M. B. Araújo, R. Bivand, J. Bolliger, G. Carl, R. G. Davies, A. Hirzel, W. Jetz, W. D. Kissling, I. Kühn, R. Ohlemuller, P. R. Peres-Neto, B. Reineking, B. Schröder, F. M. Schurr, and R. Wilson (2007). Methods to account for spatial autocorrelation in the analysis of species distributional data: A review. *Ecography 30*, 609 – 628.

Dormann, C. F., O. Purschke, J. R. G. Márquez, S. Lautenbach, and B. Schröder (2008). Components of uncertainty in species distribution analysis: A case study of the Great Grey Shrike. *Ecology 89*(12), 3371 – 3386.

Dubuis, A., J. Pottier, V. Rion, L. Pellissier, J.-P. Theurillat, and A. Guisan (2011). Predicting spatial patterns of plant species richness: A comparison of direct macroecological and species stacking modelling approaches. *Diversity and Distributions 17*, 1122 – 1131.

Dudík, M. (2007). *Maximum entropy density estimation and modeling geographic distributions of species*. Ph. D. thesis, Princeton University, USA.

Dudík, M., S. J. Phillips, and R. E. Schapire (2004). Performance guarantees for regularized maximum entropy density estimation. In *Proceedings of the 17th Annual Conference on Computational Learning Theory, July 1-4, 2004, Banff, Canada*.

Dudík, M., S. J. Phillips, and R. E. Schapire (2005). Correcting sample selection bias in maximum entropy density estimation. In *Proceedings of Neural Information Processing Systems, December 5-8, 2005, Vancouver, Canada*.

Dulvy, N., Y. Sadovy, and J. Reynolds (2003). Extinction vulnerability in marine populations. *Fish and Fisheries 4*, 25 – 64.

Edwards, J. L., M. A. Lane, and E. S. Nielsen (2000). Interoperability of biodiversity databases: Biodiversity information on every desktop. *Science 289*, 2312 – 2314.

Edwards Jr., T. C., D. R. Cutler, N. E. Zimmermann, L. Geiser, and G. G. Moisen (2006). Effects of sample survey design on the accuracy of classification tree models in species distribution models. *Ecological Modelling 199*, 132 – 141.

Eklundh, L., T. Johansson, and S. Solberg (2009). Mapping insect defoliation in Scots pine with MODIS time-series data. *Remote Sensing of Environment 113*, 1566 – 1573.

Elith, J. and M. Burgman (2002). *Predicting species occurrences: Issues of accuracy and scale*, Chapter Predictions and their validation: Rare plants in the Central Highlands, Victoria, Australia, pp. 303 – 313. Island Press, Washington.

Elith, J. and C. H. Graham (2009). Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models. *Ecography 32*, 66 – 77.

Elith, J., C. H. Graham, R. P. Anderson, M. Dudík, S. Ferrier, A. Guisan, R. J. Hijmans, F. Huettmann, J. R. Leathwick, A. Lehmann, J. Li, L. G. Lohmann, B. A. Loiselle, G. Manion, C. Moritz, M. Nakamura, Y. Nakazawa, J. M. Overton, A. T. Peterson, S. J. Phillips, K. Richardson, R. Scachetti-Pereira, R. E. Schapire, J. Soberón, S. Williams, M. S. Wisz, and N. E. Zimmermann (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography 29*, 129 – 151.

Elith, J. and J. R. Leathwick (2009). Species Distribution Models: Ecological explanation and prediction across space and time. *The Annual Review of Ecology, Evolution, and Systematics 40*, 677 – 697.

Elith, J., S. J. Phillips, T. Hastie, M. Dudík, Y. E. Chee, and C. J. Yates (2011). A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions 17*, 43 – 57.

Elton, C. (1927). *Animal ecology*. University of Chicago Press, Chicago.

Engler, R., A. Guisan, and L. Rechsteiner (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.

Evangelista, P. H., S. Kumar, T. J. Stohlgren, C. S. Jarnevich, A. W. Crall, J. B. Norman III, and D. T. Barnett (2008). Modelling invasion for a habitat generalist and a specialist plant species. *Diversity and Distributions 14*, 808 – 817.

Evangelista, P. H., T. J. Stohlgren, J. T. Morisette, and S. Kumar (2009). Mapping invasive tamarisk (*Tamarix*): A comparison of single-scene and time-series analyses of remotely sensed data. *Remote Sensing 1*, 519 – 533.

Everitt, J., D. Escobar, M. Alaniz, and M. Davis (1987). Using airborne middle-infrared (1.45 – 2.0 µm) video imagery for distinguishing plant species and soil conditions. *Remote Sensing of Environment 22*(3), 423 – 428.

Everitt, J. H., C. Yang, and C. J. Deloach (2006). Remote sensing of giant reed with QuickBird satellite imagery. *Journal of Aquatic Plant Management 43*, 81 – 85.

Eyre, M. D., S. P. Rushton, M. L. Luff, and M. G. Telfer (2004). Predicting the distribution of ground beetle species (Coleoptera, Carabidae) in Britain using land cover variables. *Journal of Environmental Management 72*, 163 – 174.

Fairbanks, D. H. K. and K. C. McGwire (2004). Patterns of floristic richness in vegetation communities of California: Regional scale analysis with multi-temporal NDVI. *Global Ecology and Biogeography 13*, 221 – 235.

FAO (2011). FAOSTAT. Technical report, Food and Agriculture Organization of the United Nations, http://faostat.fao.org, accessed on 10/18/2011.

Farjon, A. (1996). Biodiversity of *Pinus* (Pinaceae) in Mexico: Speciation and palaeo-endemism. *Botanical Journal of the Linnean Society 121*, 365 – 384.

Farjon, A. (2010). *A handbook of the world's conifers*. Brill Academic Publishers, Boston.

Feilhauer, H., U. Faude, and S. Schmidtlein (2011). Combining Isomap ordination and imaging spectroscopy to map continuous floristic gradients in a heterogeneous landscape. *Remote Sensing of Environment 115*, 2513 – 2524.

Feilhauer, H. and S. Schmidtlein (2009). Mapping continuous fields of forest alpha and beta diversity. *Applied Vegetation Science 12*(4), 429 – 439.

Fensholt, R. and I. Sandholt (2003). Derivation of a shortwave infrared water stress index from MODIS near- and shortwave infrared data in a semiarid environment. *Remote Sensing of Environment 87*, 111 – 121.

Ferrier, S. (2002). Mapping spatial pattern in biodiversity for regional conservation planning: Where to from here? *Systematic Biology 51*, 331 – 363.

Ferrier, S., M. Drielsma, G. Manion, and G. Watson (2002). Extended statistical approaches to modelling spatial pattern in biodiversity in northeast New South Wales. II. Community-level modelling. *Biodiversity and Conservation 11*, 2309 – 2338.

Ferrier, S. and A. Guisan (2006). Spatial modelling of biodiversity at the community level. *Journal of Applied Ecology 43*, 393 – 404.

Ferrusquía-Villafranca, I. (1993). *Biological diversity of Mexico: Origins and distributions*, Chapter Geology of Mexico: A synopsis, pp. 3 – 108. Oxford University Press, Oxford.

Ficetola, G. F., W. Thuiller, and C. Miaud (2007). Prediction and validation of the potential global distribution of a problematic alien invasive species – the American bullfrog. *Diversity and Distributions 13*, 476 – 485.

Fielding, A. (2002). *Predicting species occurences: Issues of accuracy and scale*, Chapter What are the appropriate characteristics of an accuracy measure?, pp. 271 – 280. Island Press, Washington.

Fielding, A. H. and J. F. Bell (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation 24*(1), 38 – 49.

Fisher, J. I. and J. F. Mustard (2007). Cross-scalar satellite phenology from ground, Landsat, and MODIS data. *Remote Sensing of Environment 109*, 261 – 273.

Fisher, J. I., J. F. Mustard, and M. A. Vadeboncoeur (2006). Green leaf phenology at Landsat resolution: Scaling from the field to the satellite. *Remote Sensing of Environment 100*, 265 – 279.

Fleishman, E., R. Mac Nally, and J. P. Fay (2002). Validation tests of predictive models of butterfly occurrence based on environmental variables. *Conservation Biology 17*(3), 806 – 817.

Foley, D. H., T. A. Klein, H. C. Kim, W. J. Sames, R. C. Wilkerson, and L. M. Rueda (2009). Geographic distribution and ecology of potential Malaria vectors in the Republic of Korea. *Journal of Medical Entomology 46*(3), 680 – 692.

Foody, G. M. (2002). Status of land cover classification accuracy assessment. *Remote Sensing of Environment 80*, 185 – 201.

Foody, G. M. (2005). Mapping the richness and composition of British breeding birds from coarse spatial resolution satellite sensor imagery. *International Journal of Remote Sensing 26*(18), 3943 – 3956.

Franklin, J. (1995). Predictive vegetation mapping: Geographic modelling of biospatial patterns in relation to environmental gradients. *Progress in Physical Geography 19*(4), 474 – 499.

Franklin, J. (2009). *Mapping species distributions: Spatial inference and prediction. Ecology, biodiversity and conservation*. Cambridge University Press, Cambridge.

Freeman, E. A. and G. G. Moisen (2008). A comparison of the performance of threshold criteria for binary classification in terms of predicted prevalence and kappa. *Ecological Modelling 217*, 48 – 58.

García, E. (1998). Climas. Classificación de Köppen (modificado), Escala 1:1 000 000. Comisión Nacional para el Conocimiento y Uso de la Biodiversidad, Mexico-City, Mexico.

Geerken, R., B. Zaitchik, and J. P. Evans (2005). Classifying rangeland vegetation type and coverage from NDVI time series using Fourier Filtered Cycle Similarity. *International Journal of Remote Sensing 26*(24), 5535 – 5554.

Geßner, U. (2010). *Räumliche und zeitliche Muster der Vegetationsstruktur in Savannen des südlichen Afrika*. Ph. D. thesis, University of Würzburg, Germany.

Gernandt, D. S., S. Magallon, G. G. Lopez, O. Z. Flores, A. Willyard, and A. Listonk (2008). Use of simultaneous analyses to guide fossil-based calibrations of Pinaceae phylogeny. *International Journal of Plant Science 169*(8), 1086 – 1099.

Gibbs, J. W. (1902). *Elementary principles in statistical mechanics, Developed with especial reference to the rational foundations of thermodynamics*. Yale University Press, New Haven.

Giddings, L., M. Soto, B. M. Rutherford, and A. Maarouf (2005). Standardized precipitation index zones for Mexico. *Atmosfera 18*, 33 – 56.

Gignac, L. D., L. A. Halsey, and D. H. Vitt (2000). A bioclimatic model for the distribution of *Sphagnum*-dominated peatlands in North America under present climatic conditions. *Journal of Biogeography 27*, 1139 – 1151.

Gillespie, T. W. (2005). Predicting woody-plant species richness in tropical dry forests: A case study from south Florida, USA. *Ecological Applications 15*, 27 – 37.

Gillespie, T. W., G. M. Foody, D. Rocchini, A. P. Giorgi, and S. Saatchi (2008). Measuring and modelling biodiversity from space. *Progress in Physical Geography 32*(2), 203 – 221.

Giovanelli, J. G. R., M. F. de Siqueira, C. F. B. Haddad, and J. Alexandrino (2010). Modeling a spatially restricted distribution in the Neotropics: How the size of calibration area affects the performance of five presence-only methods. *Ecological Modelling 221*, 215 – 224.

Gleason, H. A. (1926). The individualistic concept of the plant association. *Bulletin of the Torrey Club 53*, 7 – 26.

Gómez-Garay, A., S. Masselli, and A. Bueno (2010). Distribution of the genetic diversity of *Pinus ayacahuite* (Ehrenberg) in the communal forest of Totonicapán, Guatemala. *Bioremidiation, Biodiversity and Bioavailability 4*, 35 – 41.

Goodchild, M. F., S. Guoqing, and Y. Shiren (1992). Development and test of an error model for categorical data. *International Journal of Geographical Information Systems 6*(2), 87 – 104.

Gould, W. (2000). Remote sensing of vegetation, plant species richness, and regional biodiversity hotspots. *Ecological Applications 10*(6), 1861 – 1870.

Goward, S. and A. Hope (1989). Evapotranspiration from combined reflected solar and emitted terrestrial radiation: Preliminary FIFE results from AVHRR data. *Advances in Space Research 9*(7), 239 – 249.

Goward, S. N., Y. Xue, and K. P. Czajkowski (2002). Evaluating land surface moisture conditions from the remotely sensed temperature/vegetation index measurements, An exploration with the simplified simple biosphere model. *Remote Sensing of Environment 79*, 225 – 242.

Graham, A. (1993). *Biological diversity of Mexico: Origins and distributions*, Chapter Historical factors and biological diversity in Mexico, pp. 109 – 127. Oxford University Press, Oxford.

Graham, C. H., S. Ferrier, F. Huettman, C. Moritz, and A. T. Peterson (2004). New developments in museum-based informatics and applications in biodiversity analysis. *TRENDS in Ecology and Evolution 19*(9), 497 – 503.

Graham, C. H. and R. J. Hijmans (2006). A comparison of methods for mapping species ranges and species richness. *Global Ecology and Biogeography 15*, 578 – 587.

Graham, C. H., S. R. Ron, J. C. Santos, C. J. Schneider, and C. Moritz (2004). Integrating phylogenetics and environmental niche models to explore speciation mechanisms in dendrobatid frogs. *Evolution 58*(8), 1781 – 1793.

Graham, M. H. (2003). Confronting multicollinearity in ecological multiple regression. *Ecology 84*(11), 2809 – 2815.

Gregorio, A. D. and L. J. Jansen (2005). *Land Cover Classification System, Classification Concepts and User Manual, Software version 2*. Food and Agriculture Organization of the United Nations, Environmental and Natural Ressources Series, Rome.

Grendár, M. J. and M. Grendár (2001). Maximum Entropy: Clearing up mysteries. *Entropy 3*, 58 – 63.

Griffiths, G. H., J. Lee, and B. C. Eversham (2000). Landscape pattern and species richness; regional scale analysis from remote sensing. *International Journal of Remote Sensing 21*(13 - 14), 2685 – 2704.

Grinnell, J. (1904). The origin and distribution of the Chestnut-backed Chickadee. *Auk 21*, 364 – 365.

Grinnell, J. (1917). The niche-relationship of the California thrasher. *Auk 34*, 427 – 433.

Grünwald, P. (2000). Maximum entropy and the glasses you are looking through. In *Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence, June 30 - July 3, 2000, Stanford, USA*.

Groombridge, B. and M. Jenkins (2002). *World atlas of biodiversity*. University of California Press, Berkeley.

Gu, W. and R. K. Swihart (2004). Absent or undetected? Effects of non-detection of species occurrence on wildlife-habitat models. *Biological Conservation 116*, 195 – 203.

Guenther, B., X. Xiong, V. V. Salomonson, W. L. Barnes, and J. Young (2002). On-orbit performance of the Earth Observing System Moderate Resolution Imaging Spectroradiometer; first year of data. *Remote Sensing of Environment 83*, 16 – 30.

Guisan, A., O. Broennimann, R. Engler, M. Vust, N. G. Yoccoz, A. Lehman, and N. E. Zimmermann (2006). Using niche-based models to improve the sampling of rare species. *Conservation Biology 20*, 501 – 511.

Guisan, A., A. Lehmann, S. Ferrier, M. Austin, J. M. C. Overton, R. Aspinall, and T. Hastie (2006). Making better biogeographical predictions of species' distributions. *Journal of Applied Ecology 43*, 386 – 392.

Guisan, A. and C. Rahbek (2011). SESAM - a new framework integrating macroecological and species distribution models for predicting spatio-temporal patterns of species assemblages. *Journal of Biogeography 38*, 1433 – 1444.

Guisan, A. and J.-P. Theurillat (2000). Equilibrium modeling of alpine plant distribution: How far can we go? *Phytocoenologia 30*, 353 – 384.

Guisan, A. and W. Thuiller (2005). Predicting species distribution: Offering more than simple habitat models. *Ecology Letters 8*, 993 – 1009.

Guisan, A. and N. E. Zimmermann (2000). Predictive habitat distribution models in ecology. *Ecological Modelling 135*, 147 – 186.

Guisan, A., N. E. Zimmermann, J. Elith, C. H. Graham, S. Phillips, and A. T. Peterson (2007). What matters for predicting the occurrences of trees: Techniques, data, or species' characteristics? *Ecological Monographs 77*(4), 615 – 630.

Hall, F. G., K. F. Huemmrich, D. E. Strebel, D. E. Goetz, J. E. Nickeson, and K. D. Woods (1992). *Biophysical, morphological, canopy optical property, and product data from the superior National Forest*. NASA Technical Memorandum 104568, National Aeronautics and Space Administration, Goddard Space Flight Center, Greenbelt, Maryland, USA.

Hall, L. S., P. R. Krausman, and M. L. Morrison (1997). The habitat concept and a plea for standard terminology. *Wildlife Society Bulletin 25*(1), 173 – 182.

Hamada, Y., D. A. Stow, L. L. Coulter, J. C. Jafolla, and L. W. Hendricks (2007). Detecting tamarisk species (*Tamarix* spp.) in riparian habitats of Southern California using high spatial resolution hyperspectral imagery. *Remote Sensing of Environment 109*, 237 – 248.

Hansen, M., R. De Fries, J. Townshend, and R. Sohlberg (2000). Global land cover classification at 1 km spatial resolution using a classification tree approach. *International Journal of Remote Sensing 21*, 1331 – 1364.

Hanski, I. (1999). *Metapopulation Ecology*. Oxford University Press, Oxford.

Harlan, J. R. (1975). *Crops and Man*. Crop Science Society of America, Madison.

Harris, R. L. and S. Robinson (2001). Economy-wide effects of El Niño/Southern Oscillation ENSO in Mexico and the role of improved forecasting and technological change.

Hawkins, B. A., J. A. F. Diniz-Filho, L. M. Bini, P. D. Marco, and T. M. Blackburn (2007). Red herrings revisited: Spatial autocorrelation and parameter estimation in geographical ecology. *Ecography 30*, 375 – 384.

Hector, A. and R. Bagchi (2007). Biodiversity and ecosystem multifunctionality. *Nature 448*, 188 – 191.

Heikkinen, R. K., M. Luoto, M. B. Araújo, R. Virkkala, W. Thuiller, and M. T. Sykes (2006). Methods and uncertainties in bioclimatic envelope modelling under climate change. *Progress in Physical Geography 30*(6), 751 – 777.

Heikkinen, R. K., M. Luoto, R. Virkkala, and K. Rainio (2004). Effects of habitat cover, landscape structure and spatial variables on the abundance of birds in an agricultural-forest mosaic. *Journal of Applied Ecology 41*, 824 – 835.

Hermance, J. F., R. W. Jacob, B. A. Bradley, and J. F. Mustard (2007). Extracting phenological signals from multiyear AVHRR NDVI time series: Framework for applying high-order annual splines with roughness damping. *IEEE Transactions on Geoscience and Remote Sensing 45*(10), 2495 – 2513.

Hernandez, P. A., C. H. Graham, L. L. Master, and D. L. Albert (2006). The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography 29*, 773 – 785.

Herold, M. (2009). Assessment of the status of the development of the standards for the terrestrial climate variables, Land cover, T9 Report. GTOS Secretariat, Food and Agriculture Organization of the United Nations (FAO).

Herold, M., P. Mayaux, C. E. Woodcock, A. Baccini, and C. Schmullius (2008). Some challenges in global land cover mapping: An assessment of agreement and accuracy in existing 1 km datasets. *Remote Sensing of Environment 112*, 2538 – 2556.

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

Hijmans, R. J., L. Guarino, M. Cruz, and E. Rojas (2001). Computer tools for spatial analysis of plant genetic resources data: 1. DIVA-GIS. *Plant Genetic Resources Newsletter 127*, 15 – 19.

Hird, J. N. and G. J. McDermid (2009). Noise reduction of NDVI time series: An empirical comparison of selected techniques. *Remote Sensing of Environment 113*, 248 – 258.

Hirzel, A. and A. Guisan (2002). Which is the optimal sampling strategy for habitat suitability modelling. *Ecological Modelling 157*, 331 – 341.

Hirzel, A. H., J. Hausser, D. Chessel, and N. Perrin (2002). Ecological-niche factor analysis: How to compute habitat-suitability maps without absence data? *Ecology 83*, 2027 – 2036.

Hirzel, A. H., G. I. Le Lay, V. Helfer, C. Randin, and A. Guisan (2006). Evaluating the ability of habitat suitability models to predict species presences. *Ecological Modelling 199*, 142 – 152.

Hirzel, A. H., B. Posse, P.-A. Oggier, Y. Crettenand, C. Glenz, and R. Arlettaz (2004). Ecological requirements of reintroduced species and the implications for release policy: The case of the bearded vulture. *Journal of Applied Ecology 41*, 1103 – 1116.

Holben, B. N. (1986). Characterization of maximum value composites from temporal AVHRR data. *International Journal of Remote Sensing 7*(11), 1417 – 1434.

Huberty, C. J. (1994). *Applied discriminant analysis*. Wiley Interscience, New York.

Huete, A. (1988). A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment 25*, 295 – 309.

Huete, A. and K. Didan (2004). MODIS seasonal and inter-annual responses of semiarid ecosystems to drought in the southwest U.S.A. In *Proceedings of the Geoscience and Remote Sensing Symposium IGARSS, September 20-24, 2004, Anchorage, Alaska*, pp. 1538 – 1541.

Huete, A., K. Didan, T. Miura, E. P. Rodriguez, X. Gao, and L. G. Ferreira (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment 83*, 195 – 213.

Hugall, A., C. Moritz, A. Moussalli, and J. Stanisic (2002). Reconciling paleodistribution models and comparative phylogeography in the wet tropics rainforest land snail *Gnarosophia bellendenkerensis* (Brazier 1875). *Proceedings of the National Academy of Sciences of the United States of America 99*(9), 6112 – 6117.

Huntley, B. (1991). How plants respond to climate change: Migration rates, individualism and the consequences for plant communities. *Annals of Botany 67*, 15 – 22.

Huntley, B. (1995). Plant-species response to climate change - implications for the conservation of European birds. *Ibis 137*, 127 – 138.

Huntley, B., P. M. Berry, W. Cramer, and A. P. McDonald (1995). Modelling present and potential future ranges of some European higher plants using climate response surfaces. *Journal of Biogeography 22*(6), 967 – 1001.

Hutchinson, G. (1978). *An introduction to population ecology*. Ayle University Press, New Haven.

Hutchinson, G. E. (1957). Concluding remarks. *Cold Spring Harbor Symposia on Quantitative Biology 22*, 415 – 427. reprinted in 1991: Classics in Theoretical Biology. *Bulletin of Mathematical Biology 53*, 193 – 213.

Hüttich, C., U. Gessner, M. Herold, B. J. Strohbach, M. Schmidt, M. Keil, and S. Dech (2009). On the suitability of MODIS time series metrics to map vegetation types in dry savanna ecosystems: A case study in the Kalahari of NE Namibia. *Remote Sensing 1*, 620 – 643.

Hüttich, C., M. Herold, M. Wegmann, A. Cord, B. Strohbach, C. Schmullius, and S. Dech (2011). Assessing effects of temporal compositing and varying observation periods for large-area land-cover mapping in semi-arid ecosystems: Implications for global monitoring. *Remote Sensing of Environment 115*, 2445 – 2459.

Ichii, K., A. Kawabata, and Y. Yamaguchi (2002). Global correlation analysis for NDVI and climatic variables and NDVI trends: 1982–1990. *International Journal of Remote Sensing 23*, 3873 – 3878.

INEGI (2005). Guía para la interpretación de la cartografía uso del suelo y vegetación. Instituto Nacional de Estadística, Geografía e Informática.

International Union of Soil Sciences (2006). World reference base for soil resources 2006. A framework for international classification, correlation and communication, 2nd edition, World Soil Resources Reports, No. 103. Food and Agriculture Organization of the United Nations.

ITT Visual Information Solutions (2009). IDL Version 7.1.1.

IUCN (2011). IUCN Red list of threatened species. Version 2011.2. http://www.iucnredlist.org.

IUCN and UNEP (2010). The world database on protected areas (WDPA). UNEP-WCMC. Cambridge, UK. http://www.protectedplanet.net.

Iverson, L. R., A. M. Prasad, S. N. Matthews, and M. Peters (2008). Estimating potential habitat for 134 eastern US tree species under six climate scenarios. *Forest Ecology and Management 254*, 390 – 406.

Jaynes, E. (1957). Information theory and statistical mechanics. *Physical Reviews 106*, 620 – 630.

Jaynes, E. T. (1990). *Maximum Entropy and Bayesian methods*, Chapter Notes on present status and future prospects, pp. 1 – 13. Kluwer, Dordrecht.

Jensen, J. (2007). *Remote sensing of the environment: An earth resource perspective*. Pearson Education, Inc., New York.

Jiménez-Valverde, A. and J. M. Lobo (2007). Threshold criteria for conversion of probability of species presence to either-or presence-absence. *Acta Oecologica 31*, 361 – 369.

Jiménez-Valverde, A., J. M. Lobo, and J. Hortal (2008). Not as good as they seem: The importance of concepts in species distribution modelling. *Diversity and Distributions 14*, 885 – 890.

John, R., J. Chen, N. Lu, K. Guo, C. Liang, Y. Wei, A. Noormets, K. Ma, and X. Han (2008). Predicting plant diversity based on remote sensing products in the semi-arid region of Inner Mongolia. *Remote Sensing of Environment 112*, 2018 – 2032.

Jones, C. C. and S. Reichard (2009). Current and Potential Distributions of Three Non-Native Invasive Plants in the Contiguous USA. *Natural Areas Journal 29*(4), 332 – 343.

Jönsson, P. and L. Eklundh (2004). TIMESAT – A program for analyzing time-series of satellite sensor data. *Computers & Geosciences 30*, 833 – 845.

Joshi, C., J. de Leeuwa, and I. C. van Durena (2004). Remote sensing and GIS applications for mapping and spatial modelling of invasive species. In *Proceedings of the International Society for Photogrammetry and Remote Sensing*.

Julien, Y., J. A. Sobrino, and W. Verhoef (2006). Changes in land surface temperatures and NDVI values over Europe between 1982 and 1999. *Remote Sensing of Environment 103*, 43 – 55.

Jung, M., K. Henkel, M. Herold, and G. Churkina (2006). Exploiting synergies of global land cover products for carbon cycle modeling. *Remote Sensing of Environment 101*, 534 – 553.

Justice, C. O., J. R. G. Townshend, E. F. Vermote, E. Masuoka, R. E. Wolfe, N. Saleous, D. P. Roy, and J. T. Morisette (2002). An overview of MODIS Land data processing and product status. *Remote Sensing of Environment 83*, 3 – 15.

Justice, C. O., E. Vermote, J. R. G. Townshend, R. DeFries, D. P. Roy, D. K. Hall, V. V. Salomonson, J. L. Privette, G. Riggs, A. Strahler, W. Lucht, R. B. Myneni, Y. Knyazikhin, S. W. Running, R. R. Nemani, Z. Wan, A. R. Huete, W. van Leeuwen, R. E. Wolfe, L. Giglio, J.-P. Muller, P. Lewis, and M. J. Barnsley (1998). The Moderate Resolution Imaging Spectroradiometer (MODIS): Land remote sensing for global change research. *IEEE Transactions on Geoscience and Remote Sensing 36*(4), 1228 – 1249.

Kadmon, R., O. Farber, and A. Danin (2003). A systematic analysis of factors affecting the performance of climatic envelope models. *Ecological Applications 13*(3), 853 – 867.

Kadmon, R., O. Farber, and A. Danin (2004). Effect of roadside bias on the accuracy of predictive maps produced by bioclimatic models. *Ecological Applications 14*(2), 401 – 413.

Kaufmann, R. K., L. Zhou, R. B. Myneni, C. J. Tucker, D. Slayback, N. V. Shabanov, and J. Pinzon (2003). The effect of vegetation on surface temperature: A statistical analysis of NDVI and climate data. *Geophysical Research Letters 30*(22), 2147.

Kearney, M. (2006). Habitat, environment and niche: What are we modelling? *Oikos 115*(1), 186 – 191.

Kearney, M. and W. Porter (2004). Mapping the fundamental niche: Physiology, climate, and the distribution of a nocturnal lizard. *Ecology 85*, 3119 – 3131.

Kearney, M. and W. Porter (2009). Mechanistic niche modelling: Combining physiological and spatial data to predict species' ranges. *Ecology Letters 12*, 334 – 350.

Keith, D. A., H. R. Akçakaya, W. Thuiller, G. F. Midgley, R. G. Pearson, S. J. Phillips, H. M. Regan, M. B. Araújo, and T. G. Rebelo (2008). Predicting extinction risks under climate change: Coupling stochastic population models with dynamic bioclimatic habitat models. *Biology Letters 4*, 560 – 563.

Kent, M., W. J. Gill, R. E. Weaver, and R. P. Armitage (1997). Landscape and plant community boundaries in biogeography. *Progress in Physical Geography 21*(3), 315 – 353.

Kerns, B. K., B. J. Naylor, M. Buonopane, C. G. Parks, and B. Rogers (2009). Modeling Tamarisk (*Tamarix* spp.) Habitat and Climate Change Effects in the Northwestern United States. *Invasive Plant Science and Management 2*, 200 – 215.

Kerr, J. and M. Ostrovsky (2003). From space to species: Ecological applications for remote sensing, Review. *TRENDS in Ecology and Evolution 18*(6), 299 – 305.

Kindel, K.-H. (1995). *Kiefern in Europa*. Gustav Fischer Verlag, Stuttgart.

King, M., J. Closs, S. Wharton, M. Myers, and C. Parkinson (2004). *EOS Data Products Handbook, Volume 1*. Goddard Space Flight Center, Greenbelt, Maryland, USA.

Koltunov, A., S. L. Ustin, G. P. Asner, and I. Fung (2009). Selective logging changes forest phenology in the Brazilian Amazon: Evidence from MODIS image time series analysis. *Remote Sensing of Environment 113*, 2431 – 2440.

Kozak, K. H. and J. J. Wiens (2006). Does niche conservatism promote speciation? A case study in North American salamanders. *Evolution 60*(12), 2604 – 2621.

Kremen, C., A. Cameron, A. Moilanen, S. J. Phillips, C. D. Thomas, H. Beentje, J. Dransfield, B. L. Fisher, F. Glaw, T. C. Good, G. J. Harper, R. J. Hijmans, D. C. Lees, E. E. Louis Jr, R. A. Nussbaum, C. J. Raxworthy, A. Razafimpahanana, G. E. Schatz, M. Vences, D. R. Vieites, P. C. Wright, and M. L. Zjhra (2008). Aligning conservation priorities across taxa in Madagascar with high-resolution planning tools. *Nature 320*, 222 – 226.

Krishnaswamy, J., K. S. Bawa, K. N. Ganeshaiah, and M. C. Kiran (2009). Quantifying and mapping biodiversity and ecosystem services: Utility of a multi-season NDVI based Mahalanobis distance surrogate. *Remote Sensing of Environment 113*, 857 – 867.

Kumar, L., K. Schmidt, S. Dury, and A. Skidmore (2001). *Imaging spectroscopy. Basic principles and prospective applications*, Chapter Imaging spectrometry and vegetation science, pp. 111 – 155. Kluwer Academic Publishers, Dordrecht.

Laba, M., B. Blair, R. Downs, B. Monger, W. Philpot, S. Smith, P. Sullivan, and P. C. Baveye (2010). Use of textural measurements to map invasive wetland plants in the Hudson River National Estuarine Research Reserve with IKONOS satellite imagery. *Remote Sensing of Environment 114*, 876 – 886.

Lambin, E. F. (1999). Monitoring forest degradation in tropical regions by remote sensing: Some methodological issues. *Global Ecology and Biogeography 8*, 191 – 198.

Lambin, E. F. and D. Ehrlich (1996). The surface temperature-vegetation index space for land cover and land-cover change analysis. *International Journal of Remote Sensing 17*(3), 463 – 487.

Lambin, E. F. and D. Ehrlich (1997). Land-cover changes in Sub-Saharan Africa (1982-1991): Application of a change index based on remotely sensed surface temperature and vegetation indices at a continental scale. *Remote Sensing of Environment 81*, 181 – 200.

Le Maitre, D. C. (1998). *Ecology and biogeography of Pinus*, Chapter Pines in cultivation: A global view, pp. 407 – 431. Cambridge University Press, Cambridge.

Ledig, F. (1998). *Ecology and biogeography of Pinus*, Chapter Genetic variation in *Pinus*, pp. 251 – 280. Cambridge University Press, Cambridge.

Lefsky, M., W. Cohen, G. Parker, and D. Harding (2002). Lidar Remote Sensing for Ecosystem Studies. *Bioscience 52*(1), 19 – 30.

Legendre, P. (1993). Spatial autocorrelation: Trouble or new paradigm? *Ecology 74*(6), 1659 – 1673.

Legendre, P. and M. J. Fortin (1989). Spatial pattern and ecological analysis. *Vegetatio 80*, 107 – 138.

Leibold, M. A. (1995). The niche concept revisited: Mechanistic models and community context. *Ecology 76*, 1371–1382.

Lenihan, J. M. (1993). Ecological response surfaces for North American boreal tree species and their use in forest classification. *Journal of Vegetation Science 4*, 667 – 680.

Lennon, J. J. (2000). Red-shifts and red herrings in geographical ecology. *Ecography 23*, 101 – 113.

Leung, B., D. Lodge, D. Finnoff, J. Shogren, M. Lewis, and G. Lamberti (2002). An ounce of prevention or a pound of cure: Bioeconomic risk analysis of invasive species. *Proceedings of the Royal Society of London, Series B, Biological Sciences 269*, 2407 – 2413.

Levin, N., A. Shmida, O. Levanoni, H. Tamari, and S. Kark (2007). Predicting mountain plant richness and rarity from space using satellite-derived vegetation indices. *Diversity and Distributions 13*, 692 – 703.

Levins, R. (1968). *Evolution in changing environments*. Princeton University Press, Princeton.

Li, W. and Q. Guo (2010). A maximum entropy approach to one-class classification of remote sensing imagery. *International Journal of Remote Sensing 31*(8), 2227 – 2235.

Lieth, H. (1974). *Phenology and seasonality modeling*, Chapter Purposes of a phenology book, pp. 3 – 19. Springer-Verlag, New York.

Liu, C., P. M. Berry, T. P. Dawson, and R. G. Pearson (2005). Selecting thresholds of occurrence in the prediction of species distributions. *Ecography 28*(3), 385 – 393.

Liu, W. and E. Wu (2005). Comparison of non-linear mixture models: Sub-pixel classification. *Remote Sensing of Environment 94*, 145 – 154.

Llorente-Bousquets, J. and S. Ocegueda (2008). *Capital natural de México, Vol.I: Conocimiento actual de la biodiversidad*, Chapter Estado del conocimiento de la biota de México, pp. 283 – 322. Comisión Nacional para el Conocimiento y Uso de la Biodiversidad, Mexico-City, Mexico.

Lobo, J. M., A. Jiménez-Valverde, and R. Real (2008). AUC: A misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography 17*, 145 – 151.

Loiselle, B. A., C. A. Howell, C. H. Graham, J. M. Goerck, T. Brooks, K. G. Smith, and P. H. Williams (2003). Avoiding pitfalls of using species distribution models in conservation planning. *Conservation Biology 17*(6), 1591 – 1600.

Lomba, A., L. Pellissier, C. Randin, J. Vicente, F. Moreira, J. Honrado, and A. Guisan (2010). Overcoming the rare species modelling paradox: A novel hierarchical framework applied to an Iberian endemic plant. *Biological Conservation 29*, 2647 – 2657.

Lomolino, M. (2004). *Frontiers of biogeography: New directions in the geography of nature*, Chapter Conservation biogeography, pp. 293 – 296. Sinauer Associates, Sunderland.

Loreau, M., S. Naeem, P. Inchausti, J. Bengtsson, J. P. Grime, A. Hector, D. U. Hooper, M. A. Huston, D. Raffaelli, B. Schmid, D. Tilman, and D. A. Wardle (2001). Biodiversity and ecosystem functioning: Current knowledge and future challenges. *Science 294*, 804 – 808.

Luo, Y., A. P. Trishchenko, and K. V. Khlopenkov (2008). Developing clear-sky, cloud and cloud shadow mask for producing clear-sky composites at 250-meter spatial resolution for the seven MODIS land bands over Canada and North America. *Remote Sensing of Environment 112*, 4167 – 4185.

Luoto, M., R. K. Heikkinen, J. Pöyry, and K. Saarinen (2006). Determinants of the biogeographical distribution of butterflies in boreal regions. *Journal of Biogeography 33*, 1764 – 1778.

Luoto, M., J. Pöyry, R. K. Heikkinen, and K. Saarinen (2005). Uncertainty of bioclimate envelope models based on the geographical distribution of species. *Global Ecology and Biogeography 14*, 575 – 584.

Luoto, M., T. Toivonen, and R. K. Heikkinen (2002). Prediction of total and rare plant species richness in agricultural landscapes from satellite images and topographic data. *Landscape Ecology 17*, 195 – 217.

Luoto, M., R. Virkkala, and R. K. Heikkinen (2007). The role of land cover in bioclimatic models depends on spatial resolution. *Global Ecology and Biogeography 16*, 34 – 42.

MacArthur, R. H. (1967). *Population biology and evolution*, Chapter The theory of the niche, pp. 156 – 176. Syracuse University Press, Syracuse.

Mace, G., H. Masundire, J. Baillie, T. Ricketts, T. Brooks, M. Hoffmann, S. Stuart, A. Balmford, A. Purvis, B. Reyers, J. Wang, C. Revenga, E. Kennedy, S. Naeem, R. Alkemade, T. Allnutt, M. Bakarr, W. Bond, J. Chanson, N. Cox, G. Fonseca, C. Hilton-Taylor, C. Loucks, A. Rodrigues, W. Sechrest, A. Stattersfield, B. J. van Rensburg, C. Whiteman, R. Abell, Z. Cokeliss, J. Lamoreux, H. M. Pereira, J. Thonell, and P. Williams (2005). *Ecosystems and human well-being: Current states and trends, Volume 1*, Chapter Biodiversity, pp. 77 – 122. Island Press, Washingon.

Machwitz, M. (2010). *Eine raum-zeitliche Modellierung der Kohlenstoffbilanz mit Fernerkundungsdaten auf regionaler Ebene in Westafrika*. Ph. D. thesis, University of Würzburg, Germany.

MacNally, R. (2000). Regression and model-building in conservation biology, biogeography and ecology: The distinction between - and reconciliation of - ’predictive’ and ’explanatory’ models. *Biodiversity and Conservation 9*, 655 – 671.

Maes, D., M. Gilbert, N. Titeux, P. Goffart, and R. L. H. Dennis (2003). Prediction of butterfly diversity hotspots in Belgium: A comparison of statistically focused and land use-focused models. *Journal of Biogeography 30*, 1907 – 1920.

Magaña, V. I. and C. Conde (2000). Climate and freshwater resources in northern Mexico: Sonora, a case study. *Environmental Monitoring and Assessment 61*, 167 – 185.

Mahalanobis, P. (1936). On the generalized distance in statistics. *Proceedings of the National Institute of Science of India 2*(1), 49 – 55.

Maignan, F., F.-M. Breon, C. Bacour, J. Demarty, and A. Poirson (2008). Interannual vegetation phenology estimates from global AVHRR measurements: Comparison with in situ data and applications. *Remote Sensing of Environment 112*, 496 – 505.

Manel, S., H. C. Williams, and S. J. Ormerod (2001). Evaluating presence-absence models in ecology: The need to account for prevalence. *Journal of Applied Ecology 38*, 921 – 931.

Margules, C. R. and R. L. Pressey (2000). Systematic conservation planning. *Nature 405*, 13 – 17.

Marshall, J. T. (1957). *Birds of pine-oak woodland in southern Arizona and adjacent Mexico*, Volume 32 of *Pacific Coast Avifauna*. Cooper Ornithological Society, Berkeley.

Martínez-Meyer, E., A. Peterson, J. L. Servín, and L. F. Kiff (2006). Ecological niche modelling and prioritizing areas for species reintroductions. *Oryx 40*(4), 411 – 418.

Mas, J. F., A. Velazquez, J. R. Diaz, R. Mayorga, C. Alcantara, R. Castro, and T. Fernandez (2002). Assessing land use/cover changes in Mexico. Instituto de Geografia, UNAM.

Mas, J. F., A. Velázquez, J. L. Palacio-Prieto, G. Bocco, A. Peralta, and J. Prado (2002). Assessing forest resources in Mexico: Wall-to-wall land use/cover mapping. *Photogrammetric Engineering & Remote Sensing 68*(10), 966 – 969.

Mateo, R. G., T. B. Croat, A. M. Felicísimo, and J. Muñoz (2010). Profile or group discriminative techniques? Generating reliable species distribution models using pseudo-absences and target-group absences from natural history collections. *Diversity and Distributions 16*, 84 – 94.

Mather, P. M. (2010). Is there any sense in remote sensing? *Progress in Physical Geography 34*(6), 739 – 740.

McNaughton, S. J. (1977). Diversity and stability of ecological communities: A comment on the role of empiricism in ecology. *The American Naturalist 111*(979), 515 – 525.

McPherson, J. M. and W. Jetz (2007). Effects of species' ecology on the accuracy of distribution models. *Ecography 30*, 135 – 151.

McPherson, J. M., W. Jetz, and D. J. Rogers (2004). The effects of species' range sizes on the accuracy of distribution models: Ecological phenomenon or statistical artefact? *Journal of Applied Ecology 41*, 811 – 823.

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

Mehner, H., M. Cutler, D. Fairbarn, and G. Thompson (2004). Remote sensing of upland vegetation: The potential of high spatial resolution satellite sensors. *Global Ecology and Biogeography 13*, 359 – 369.

Meyer, R., K. Staenz, and K. I. Itten (1996). Semi-automated procedures for tree species identification in high spatial resolution data from digitized colour infrared-aerial photography. *ISPRS Journal of Photogrammetry & Remote Sensing 51*, 5 – 16.

Millennium Ecosystem Assessment (2005). *Ecosystems and human well-being: Biodiversity synthesis*. World Resources Institute, Washington.

Miller, J., J. Franklin, and R. Aspinall (2007). Incorporating spatial dependence in predictive vegetation models. *Ecological Modelling 202*, 225 – 242.

Mills, H., M. E. J. Cutler, and D. Fairbairn (2006). Artificial neural networks for mapping regional-scale upland vegetation from high spatial resolution imagery. *International Journal of Remote Sensing 27*(11), 2177 – 2195.

Miranda, F. and X. Hernández (1963). Los tipos de vegetación de México y su clasificación. *Boletín de la Sociedad Botánica de México 28*, 29 – 179.

Mirov, N. T. (1967). *The genus Pinus*. The Ronald Press Company, New York.

Mittermeier, R. A. (1988). *Biodiversity*, Chapter Primate diversity and the tropical forest: Case studies from Brazil and Madagascar and the importance of the megadiversity countries, pp. 145 – 154. National Academic Press, Washington.

Mittermeier, R. A., J. Schipper, G. Davidse, P. Koleff, J. Soberón, M. Ramírez, N. Goettsch, and C. G. Mittermeier (2005). *Hotspots revisited: Earth's biologically richest and most endangered terrestrial ecoregions*, Chapter Mesoamerica, pp. 103 – 112. CEMEX, Mexico-City.

Miura, T., A. R. Huete, H. Yoshioka, and B. N. Holben (2001). An error and sensitivity analysis of atmospheric resistant vegetation indices derived from dark target-based atmospheric correction. *Remote Sensing of Environment 78*, 284 – 298.

Müller-Dombois, D. and J. Ellenberg (1974). *Aims and methods of vegetation ecology*. John Wiley, New York.

Moisen, G. G. and T. S. Frescino (2002). Comparing five modelling techniques for predicting forest characteristics. *Ecological Modelling 157*, 209 – 225.

Mora, F. and L. R. Iverson (1998). On the sources of vegetation activity variation, and their relation with water balance in Mexico. *International Journal of Remote Sensing 19*(10), 1843 – 1871.

Moran, P. A. P. (1950). Notes on continuous stochastic phenomena. *Biometrika 37*, 17 – 23.

Morisette, J. T., C. S. Jarnevich, A. Ullah, W. Cai, J. A. Pedelty, J. E. Gentle, T. J. Stohlgren, and J. L. Schnase (2005). A tamarisk habitat suitability map for the continental United States. *Frontiers in Ecology and Environment 4*(1), 11 – 17.

Morisette, J. T., A. D. Richardson, A. K. Knapp, J. I. Fisher, E. A. Graham, J. Abatzoglou, B. E. Wilson, D. D. Breshears, G. M. Henebry, J. M. Hanes, and L. Liang (2009). Tracking the rhythm of the seasons in the face of global change: Phenological research in the 21st century. *Frontiers in Ecology and the Environment 7*(5), 253 – 260.

Morrone, J. J., D. Espinosa Organista, and J. Llorente-Bousquets (2002). Mexican biogeographic provinces: Preliminary scheme, general characterizations and synonyms. *Acta Zoológica Mexicana 85*, 83 – 108.

Muchoney, D. M. (2008). Earth observations for terrestrial biodiversity and ecosystems. *Remote Sensing of Environment 112*, 1909 – 1911.

Murray, A. (1866). *The geographical distribution of mammals*. Day & Son, London.

Murray, K. A., R. W. R. Retallick, R. Puschendorf, L. F. Skerratt, D. Rosauer, H. I. McCallum, L. Berger, R. Speare, and J. VanDerWal (2011). Assessing spatial patterns of disease risk to biodiversity: Implications for the management of the amphibian pathogen, *Batrachochytrium dendrobatidis*. *Journal of Applied Ecology 48*, 163 – 173.

Nagendra, H. (2001). Using remote sensing to assess biodiversity, Review article. *International Journal of Remote Sensing 22*(12), 2377–2400.

NALCMS (2005). North American Land Cover at 250 m spatial resolution. Produced by: Natural Resources Canada/Canadian Centre for Remote Sensing (NRCan/CCRS), United States Geological Survey (USGS), Insituto Nacional de Estadística y Geografía (INEGI), Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (CONABIO), and Comisión Nacional Forestal (CONAFOR), available from: http://www.cec.org/naatlas/nalcms/.

National Advisory Committee on Invasive Species (2010). *National strategy on invasive species in Mexico, prevention, control and eradication*. Comisión Nacional para el Conocimiento y Uso de la Biodiversidad, Comisión Nacional de Áreas Protegidas, Secretaría de Medio Ambiente y Recursos Naturales, Mexico-City, Mexico.

Nemani, R., L. Pierce, S. Running, and S. Goward (1993). Developing satellite-derived estimates of surface moisture status. *Journal of Applied Meteorology 32*, 548 – 557.

Nemani, R. and S. Running (1997). Land cover characterization using multitemporal red, near-IR and thermal-IR data from NOAA/AVHRR. *Ecological Applications 7*(1), 79 – 90.

Neteler, M. (2010). Estimating daily land surface temperatures in mountainous environments by reconstructed MODIS LST data. *Remote Sensing 2*, 333 – 351.

Newbold, T., F. Gilbert, S. Zalat, A. El-Gabbas, and T. Reader (2009). Climate-based models of spatial patterns of species richness in Egypt's butterfly and mammal fauna. *Journal of Biogeography 36*(11), 2085 – 2095.

Nijland, W., E. A. Addink, S. M. De Jong, and F. D. Van der Meer (2009). Optimizing spatial image support for quantitative mapping of natural vegetation. *Remote Sensing of Environment 113*, 771 – 780.

Nilsson, M. (1996). Estimation of tree heights and stand volumes using an airborne LIDAR system. *Remote Sensing of Environment 56*, 1 – 7.

Nix, H. A. (1986). *Atlas of elapid snakes of Australia. Australian flora and fauna*, Chapter A biogeographic analysis of Australian elapid snakes, pp. 4 – 15. Australian Government Publishing Service, Canberra.

NOAA (2012). Multivariate ENSO Index (MEI). National Oceanic and Atmospheric Administration, Earth System Research Laboratory, available from: http://www.esrl.noaa.gov/psd/enso/mei/.

Oindo, B. O., R. A. de By, and A. K. Skidmore (2000). Interannual variability of NDVI and bird species diversity in Kenya. *International Journal of Applied Earth Observation and Geoinformation 2*(3/4), 31 – 53.

Oke, T. R. (1982). The energetic basis of the urban heat-island. *Quarterly Journal of the Royal Meteorological Society 108*(455), 1 – 24.

Oksanen, J. and P. R. Minchin (2002). Continuum theory revisited: What shape are species responses along ecological gradients? *Ecological Modelling 157*, 119 – 129.

Oldeland, J., D. Wesuls, D. Rocchini, M. Schmidt, and N. Jürgens (2009). Does using species abundance data improve estimates of species diversity from remotely sensed spectral heterogeneity? *Ecological Indicators 10*(2), 390 – 396.

Olmsted, I. and E. R. Alvarez-Buylla (1995). Sustainable harvesting of tropical trees: Demography and matrix models of two palm species in Mexico. *Ecological Applications 5*(2), 484 – 500.

Palmer, M. W., P. G. Earls, B. W. Hoagland, P. S. White, and T. Wohlgemuth (2002). Quantitative tools for perfecting species lists. *Environmetrics 13*, 121 – 137.

Parida, B. R., B. Oinam, N. R. Patel, N. Sharma, R. Kandwal, and M. K. Hazarika (2008). Land surface temperature variation in relation to vegetation type using MODIS satellite data in Gujarat state of India. *International Journal of Remote Sensing 29*(14), 4219 – 4235.

Parmesan, C. (1996). Climate and species' range. *Nature 382*, 765 – 766.

Parviainen, M., M. Luoto, and R. K. Heikkinen (2009). The role of local and landscape level measures of greenness in modelling boreal plant species richness. *Ecological Modelling 220*, 2690 – 2701.

Pearce, J. and S. Ferrier (2000). Evaluating the predictive performance of habitat models developed using logistic regression. *Ecological Modelling 133*, 225 – 245.

Pearce, J. and D. Lindenmayer (1998). Bioclimatic analysis to enhance reintroduction biology of the endangered Helmeted Honeyeater (*Lichenostomus melanops cassidix*) in southeastern Australia. *Restoration Ecology 6*(3), 238 – 243.

Pearman, P. B., A. Guisan, O. Broennimann, and C. F. Randin (2007). Niche dynamics in space and time. *TRENDS in Ecology and Evolution 23*(3), 149 – 158.

Pearson, R. G. (2007). Species' distribution modeling for conservation educators and practitioners. Center for Biodiversity and Conservation & Department of Herpetology, American Museum of Natural History, New York. Available from: http://ncep.amnh.org.

Pearson, R. G. and T. P. Dawson (2003). Predicting the impacts of climate change on the distribution of species: Are bioclimate envelope models useful? *Global Ecology and Biogeography 12*, 361 – 371.

Pearson, R. G., T. P. Dawson, and C. Liu (2004). Modelling species distributions in Britain: A hierarchical integration of climate and land-cover data. *Ecography 27*, 285 – 298.

Pearson, R. G., C. J. Raxworthy, M. Nakamura, and A. T. Peterson (2007). Predicting species distributions from small numbers of occurrence records: A test case using cryptic geckos in Madagascar. *Journal of Biogeography 34*, 102 – 117.

Pearson, R. G., W. Thuiller, M. B. Araújo, E. Martínez-Meyer, L. Brotons, C. McClean, L. Miles, P. Segurado, T. P. Dawson, and D. C. Lees (2006). Model-based uncertainty in species range prediction. *Journal of Biogeography 33*, 1704 – 1711.

Peck, M., J. Thorn, A. Mariscal, A. Baird, D. Tirira, and D. Kniveton (2011). Focusing conservation efforts for the critically endangered brown-headed spider monkey (*Ateles fusciceps*) using remote sensing, modeling, and playback survey methods. *International Journal of Primatology 32*, 134 – 148.

Perry, J. (1991). *The pines of Mexico and Central America*. Timber Press, Portland.

Perry, J., A. Graham, and D. Richardson (1998). *Ecology and biogeography of Pinus*, Chapter The history of pines in Mexico and Central America, pp. 137 – 149. Cambridge University Press, Cambridge.

Peterson, A. T. (2003). Predicting the geography of species' invasions via ecological niche modeling. *The Quarterly Review of Biology 78*(4), 419 – 433.

Peterson, A. T. (2006). Uses and requirements of ecological niche models and related distributional models. *Biodiversity Informatics 3*, 59 – 72.

Peterson, A. T., M. A. Ortega-Huerta, J. Bartley, V. Sánchez-Cordero, J. Soberón, R. H. Buddemeier, and D. R. B. Stockwell (2002). Future projections for Mexican faunas under global climate change scenarios. *Letters to Nature 416*, 626 – 629.

Peterson, A. T., M. Papes, and J. Soberón (2008). Rethinking receiver operating characteristic analysis applications in ecological niche modeling. *Ecological Modelling 213*, 63 – 72.

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

Pettorelli, N., J. O. Vik, A. Mysterud, J.-M. Gaillard, C. J. Tucker, and N. C. Stenseth (2005). Using the satellite-derived NDVI to assess ecological responses to environmental change. *TRENDS in Ecology and Evolution 20*, 503 – 510.

Phillips, S. J. (2008). Transferability, sample selection bias and background data in presence-only modelling: A response to Peterson et al. (2007). *Ecography 31*, 272 – 278.

Phillips, S. J. (2009). A brief tutorial on Maxent. http://www.cs.princeton.edu/schapire/maxent/.

Phillips, S. J., R. P. Anderson, and R. E. Schapire (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling 190*, 231 – 259.

Phillips, S. J. and M. Dudík (2008). Modeling of species distributions with Maxent: New extensions and a comprehensive evaluation. *Ecography 31*, 161 – 175.

Phillips, S. J., M. Dudík, J. Elith, C. H. Graham, A. Lehmann, J. Leathwick, and S. Ferrier (2009). Sample selection bias and presence-only distribution models: Implications for background and pseudo-absence data. *Ecological Applications 19*(1), 181 – 197.

Phillips, S. J., M. D. Dudík, and R. E. Schapire (2004). A maximum entropy approach to species distribution modeling. In *Proceedings of the 21st International Conference on Machine Learning, Banff, Alberta, Canada, July 4-8, 2004*.

Pielou, E. C. (1966). The measurement of diversity in different types of biological collections. *Journal of Theoretical Biology 13*, 131 – 144.

Pineda, E. and J. M. Lobo (2009). Assessing the accuracy of species distribution models to predict amphibian species richness patterns. *Journal of Animal Ecology 78*, 182 – 190.

Pittman, S. J. and K. A. Brown (2011). Multi-scale approach for predicting fish species distributions across coral reef seascapes. *PLoS ONE 6*(5), e20583.

Ponder, W., G. Carter, P. Flemons, and R. Chapman (2001). Evaluation of museum collection data for use in biodiversity assesment. *Conservation Biology 15*(3), 648 – 657.

Pöyry, J., M. Luoto, R. K. Heikkinen, and K. Saarinen (2008). Species traits are associated with the quality of bioclimatic models. *Global Ecology and Biogeography 17*, 403 – 414.

Prates-Clark, C. D., S. S. Saatchi, and D. Agosti (2008). Predicting geographical distribution models of high-value timber trees in the Amazon Basin using remotely sensed data. *Ecological Modelling 211*, 309 – 323.

Pérez de la Rosa, J. A. (1983). Una nueva especie de pino de Jalisco, México. *Phytologia 54*(5), 289 – 298.

Pérez de la Rosa, J. A. (2009). *Pinus georginae* (Pinaceae), a new species from western Jalisco, Mexico. *Brittonia 61*(1), 56 – 61.

Pulliam, H. R. (1988). Sources, sinks, and population regulation. *The American Naturalist 132*(5), 652 – 661.

Pulliam, H. R. (2000). On the relationship between niche and distribution. *Ecology Letters 3*, 349 – 361.

Puschendorf, R., A. C. Carnaval, J. VanDerWal, H. Zumbado-Ulate, G. Chaves, F. Bolaños, and R. A. Alford (2009). Distribution models for the amphibian chytrid *Batrachochytrium dendrobatidis* in Costa Rica: Proposing climatic refuges as a conservation tool. *Diversity and Distributions 15*, 401 – 408.

Putz, F., G. Blate, K. Redford, R. Fimbel, and J. Robinson (2001). Tropical forest management and conservation of biodiversity: An overview. *Conservation Biology 15*, 7 – 20.

Qin, Z. and A. Karnieli (1999). Progress in the remote sensing of land surface temperature and ground emissivity using NOAA-AVHRR data. *International Journal of Remote Sensing 20*(12), 2367 – 2393.

Quattrochi, D. A. and J. C. Luvall (1999). Thermal infrared remote sensing for analysis of landscape ecological processes: Methods and applications. *Landscape Ecology 14*, 577 – 598.

Quattrochi, D. A. and M. K. Ridd (1998). Analysis of vegetation within a semi-arid urban environment using high spatial resolution airborne thermal infrared remote sensing data. *Atmospheric Environment 32*, 19 – 33.

R Development Core Team (2011). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna.

Raes, N., M. C. Roos, J. W. F. Slik, E. E. van Loon, and H. ter Steege (2009). Botanical richness and endemicity patterns of Borneo derived from species distribution models. *Ecography 32*, 180 – 192.

Ramamoorthy, J. P., R. Bye, A. Lot, and J. Fa (1993). *Biological diversity of Mexico: Origins and distributions*, Chapter Introduction, pp. 1 – 2. Oxford University Press, Oxford.

Raxworthy, C. J., C. M. Ingram, N. Rabibisoa, and R. G. Pearson (2007). Applications of ecological niche modeling for species delimitation: A review and empirical evaluation using day geckos (*Phelsuma*) from Madagascar. *Systematic Biology 56*(6), 907 – 923.

Raxworthy, C. J., E. Martínez-Meyer, N. Horning, R. A. Nussbaum, G. E. Schneider, M. A. Ortega-Huerta, and A. T. Peterson (2003). Predicting distributions of known and unknown reptile species in Madagascar. *Nature 426*, 837–841.

Reddy, S. and L. M. Davalos (2003). Geographical sampling bias and its implications for conservation priorities in Africa. *Journal of Biogeography 30*, 1719 – 1727.

Reed, B. C., J. F. Brown, D. VanderZee, T. R. Loveland, J. W. Merchant, and D. O. Ohlen (1994). Measuring phenological variability from satellite imagery. *Journal of Vegetation Science 5*, 703 – 714.

Reed, B.C., S.-M. and X. Xiao (2009). *Phenology of ecosystem processes*, Chapter Remote sensing phenology: Status and the way forward, pp. 232 – 247. Springer Verlag, Wien.

Reineking, B. and B. Schröder (2006). Constrain to perform: Regularization of habitat models. *Ecological Modelling 193*, 675 – 690.

Richardson, D. and P. Rundel (1998). *Ecology and biogeography of Pinus*, Chapter Ecology and biogeography of pines: An introduction, pp. 3 – 46. Cambridge University Press, Cambridge.

Ricker, M., I. Ramírez-Krauss, G. Ibarra-Manríquez, E. Martínez, C. H. Ramos, G. González-Medellín, G. Gómez-Rodríguez, J. L. Palacio-Prieto, and H. M. Hernández (2007). Optimizing conservation of forest diversity: A country-wide approach in Mexico. *Biodiversity and Conservation 16*, 1927 – 1957.

Riordan, E. C. and P. W. Rundel (2009). Modelling the distribution of a threatened habitat: The California sage scrub. *Journal of Biogeography 36*, 2176 – 2188.

Robertson, M. P., C. I. Peter, M. H. Villet, and B. S. Ripley (2003). Comparing models for predicting species' potential distributions: A case study using correlative and mechanistic predictive modelling techniques. *Ecological Modelling 164*, 153 – 167.

Rocchini, D. (2007). Effects of spatial and spectral resolution in estimating ecosystem alpha-diversity by satellite imagery. *Remote Sensing of Environment 111*, 423 – 434.

Rödder, D. and S. Lötters (2009). Niche shift versus niche conservatism? Climatic characteristics of the native and invasive ranges of the Mediterranean house gecko (Hemidactylus turcicus). *Global Ecology and Biogeography 18*, 674 – 687.

Rödder, D. and S. Lötters (2010). Potential distribution of the alien invasive brown tree snake *Boiga irregularis* (Reptilia: Colubridae). *Pacific Science 64*(1), 11 – 22.

Rogers, D. J., S. E. Randolph, R. W. Snow, and S. I. Hay (2002). Satellite imagery in the study and forecast of malaria. *Nature 415*, 710 – 715.

Roura-Pascual, N., J. M. Bas, and C. Hui (2010). The spread of the Argentine ant: Environmental determinants and impacts on native ant communities. *Biological Invasions 12*, 2399 – 2412.

Running, S. (1990). *Remote sensing of biosphere functioning*, Chapter Estimating primary productivity by combining remote sensing with ecosystem simulation, pp. 65 – 86. Springer Verlag, Wien.

Running, S. W. and R. R. Nemani (1988). Relating seasonal patterns of the AVHRR vegetation to simulated photosynthesis and transpiration of forests in different climates. *Remote Sensing of Environment 24*, 347 – 367.

Rushton, S. P., S. J. Ormerod, and G. Kerby (2004). New paradigms for modelling species distributions? *Journal of Applied Ecology 41*, 193 – 200.

Rydgren, K., R. H. Økland, and T. Økland (2003). Species response curves along environmental gradients. A case study from SE Norwegian swamp forests. *Journal of Vegetation Science 14*, 869 – 880.

Rzedowski, J. (1978). *Vegetación de México*. Limusa, Mexico-City.

Rzedowski, J. (1990). Vegetación potencial. Map IV.8.2 of the Atlas Nacional de México. Vol. II. Scale 1:4000000. Instituto de Geografía, Universidad Nacional Autónoma de México, Mexico-City, Mexico.

Rzedowski, J. (1993). *Biological diversity of Mexico: Origins and distributions*, Chapter Diversity and origins of the phanerogamic flora of Mexico, pp. 129 – 144. Oxford University Press, Oxford.

Saatchi, S., W. Buermann, H. ter Steege, S. Mori, and T. B. Smith (2008). Modeling distribution of Amazonian tree species and diversity using remote sensing measurements. *Remote Sensing of Environment 112*, 2000 – 2017.

Salinas-Zavala, C. A., A. V. Douglas, and H. F. Diaz (2002). Interannual variability of NDVI in northwest Mexico. Associated climatic mechanisms and ecological implications. *Remote Sensing of Environment 82*, 417 – 430.

Sandholt, I., K. Rasmussen, and J. Andersen (2002). A simple interpretation of the surface temperature/vegetation index space for assessment of surface moisture status. *Remote Sensing of Environment 79*, 213 – 224.

Sarukhán, J., P. Koleff, J. Carabias, J. Soberón, R. Dirzo, J. Llorente-Bousquets, G. Halffter, R. González, I. March, A. Mohar, S. Anta, and J. de la Maza (2010). *Natural capital of Mexico. Synopsis: Current knowledge, evaluation, and prospects for sustainability*. Comisión Nacional para el Conocimiento y Uso de la Biodiversidad, Mexico-City, Mexico.

Schowengerdt, R. A. (2007). *Remote sensing: Models and methods for image processing, Third edition*. Academic Press, Burlington.

Schulte, U., A. Hochkirch, S. Lötters, D. Rödder, S. Schweiger, T. Weimann, and M. Veith (2012). Cryptic niche conservatism among evolutionary lineages of an invasive lizard. *Global Ecology and Biogeography 21*(2), 198 – 211.

Schwartz, M. D. (2003). *Phenology: An integrative environmental science.* Kluwer Academic, Dordrecht.

Schwarz, N., S. Lautenbach, and R. Seppelt (2011). Exploring indicators for quantifying surface urban heat islands of European cities with MODIS land surface temperatures. *Remote Sensing of Environment 115*, 3175 – 3186.

Segurado, P. and M. B. Araújo (2004). An evaluation of methods for modelling species distributions. *Journal of Biogeography 31*, 1555 – 1568.

Sellars, J. D. and C. L. Jolls (2007). Habitat modeling for *Amaranthus pumilus*: An application of Light Detection and Ranging (LIDAR) data. *Journal of Coastal Research 235*, 1193 – 1202.

SEMARNAP (1998). Mapa de suelos dominantes de la República Mexicana (Primera aproximación 1996). Escala 1:4000000. Secretaría de Medio Ambiente, Recursos Naturales y Pesca, Subsecretaría de Recursos Naturales, Mexico-City, Mexico.

SEMARNAT (2010). Norma Oficial Mexicana NOM-059-SEMARNAT-2001, Protección ambiental – Especies nativas de México de flora y fauna silvestre – Categorías de riesgo y especificaciones para su inclusión, exclusión o cambio – Lista de especies en riesgo. Mexico-City.

Seoane, J., L. M. Carrascal, C. L. Alonso, and D. Palomino (2005). Species-specific traits associated to prediction errors in bird habitat suitability modelling. *Ecological Modelling 185*, 299 – 308.

Shannon, C. E. (1948). A mathematical theory of communication. *Reprinted with corrections from The Bell System Technical Journal 27*, 379 – 423, 623 – 656.

Siriwardena, G. M., H. Q. P. Crick, S. R. Baillie, and J. D. Wilson (2000). Agricultural land-use and the spatial distribution of granivorous lowland farmland birds. *Ecography 23*, 702 – 719.

Skidmore, A. K., J. Franklin, T. P. Dawson, and P. Pilesjö (2011). Geospatial tools address emerging issues in spatial ecology: A review and commentary on the Special Issue. *International Journal of Geographical Information Science 25*(3), 337 – 365.

Sluyter, A. and G. Dominguez (2006). Early maize (*Zea mays* L.) cultivation in Mexico: Dating sedimentary pollen records and its implications. *Proceedings of the National Academy of Sciences of the United States of America 103*(4), 1147 – 1151.

Sánchez-Cordero, V., P. Illoldi-Rangel, M. Linaje, S. Sarkar, and A. T. Peterson (2005). Deforestation and extant distributions of Mexican endemic mammals. *Biological Conservation 126*, 465 – 473.

Soberón, J. (2007). Grinnellian and Eltonian niches and geographic distributions of species. *Ecology Letters 10*, 1115 – 1123.

Soberón, J. and A. T. Peterson (2005). Interpretation of models of fundamental ecological niches and species' distributional areas. *Biodiversity Informatics 2*, 1 – 10.

Sobrino, J. A. and N. Raissouni (2000). Toward remote sensing methods for land cover dynamic monitoring: Application to Marocco. *International Journal of Remote Sensing 21*, 353 – 363.

Sommerhoff, G. and C. Weber (1999). *Mexiko*. Wissenschaftliche Buchgesellschaft, Darmstadt.

Sorte, F. A. L. and B. A. Hawkins (2007). Range maps and species richness patterns: Errors of commission and estimates of uncertainty. *Ecography 30*, 649 – 662.

Stickler, C. M. and J. Southworth (2008). Application of multi-scale spatial and spectral analysis for predicting primate occurrence and habitat associations in Kibale National Park, Uganda. *Remote Sensing of Environment 112*, 2170 – 2186.

Stockwell, D. and D. Peters (1999). The GARP modelling system: Problems and solutions to automated spatial prediction. *International Journal of Geographical Information Science 13*(2), 143 – 158.

Stockwell, D. and A. T. Peterson (2003). Comparison of resolution of methods used in mapping biodiversity patterns from point-occurrence data. *Ecological Indicators 3*, 213 – 221.

Stockwell, D. R. B. and A. T. Peterson (2002). Effects of sample size on accuracy of species distribution models. *Ecological Modelling 148*, 1 – 13.

Strand, H., R. Höft, J. Strittholt, L. Miles, N. Horning, E. Fosnight, and W. Turner (2007). Sourcebook on remote sensing and biodiversity indicators, CBD Technical Series No. 32. Secretariat of the Convention on Biological Diversity, Montreal, Canada.

Stuart, N., T. Barratt, and C. Place (2006). Classifying the Neotropical savannas of Belize using remote sensing and ground survey. *Journal of Biogeography 33*, 476 – 490.

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

Syphard, A. D. and J. Franklin (2010). Species traits affect the performance of species distribution models for plants in southern California. *Journal of Vegetation Science 21*, 177 – 189.

Taylor, J. E. (1993). Factors causing variation in reflectance measurements from bracken in eastern Australia. *Remote Sensing of Environment 43*, 217 – 229.

Téllez-Valdés, O. and P. Dávila-Aranda (2003). Protected areas and climate change: A case study of the cacti in the Tehuacán-Cuicatlán Biosphere Reserve, México. *Conservation Biology 17*(3), 846 – 853.

Thenkabail, P. S., J. Hall, T. Lin, M. S. Ashton, D. Harris, and E. A. Enclona (2003). Detecting floristic structure and pattern across topographic and moisture gradients in a mixed species Central African forest using IKONOS and Landsat-7 ETM+ images. *International Journal of Applied Earth Observation and Geoinformation 4*, 255 – 270.

Thorn, J. S., V. Nijman, D. Smith, and K. A. I. Nekaris (2009). Ecological niche modelling as a technique for assessing threats and setting conservation priorities for Asian slow lorises (Primates: *Nycticebus*). *Diversity and Distributions 15*, 289 – 298.

Thuiller, W. (2003). BIOMOD - optimizing predictions of species distributions and projecting potential future shifts under global change. *Global Change Biology 9*, 1353 – 1362.

Thuiller, W., M. B. Araújo, and S. Lavorel (2004). Do we need land-cover data to model species distributions in Europe? *Journal of Biogeography 31*, 353 – 361.

Thuiller, W., M. B. Araújo, and S. Lavorel (2003). Generalized models vs. classification tree analysis: Predicting spatial distributions of plant species at different scales. *Journal of Vegetation Science 14*, 669 – 680.

Thuiller, W., L. Brotons, M. B. Araújo, and S. Lavorel (2004). Effects of restricting environmental range of data to project current and future species distributions. *Ecography 27*, 165 – 172.

Thuiller, W., S. Lavorel, M. T. Sykes, and M. B. Araújo (2006). Using niche-based modelling to assess the impact of climate change on tree functional diversity in Europe. *Diversity and Distributions 12*, 49 – 60.

Thuiller, W., D. M. Richardson, P. Pysek, G. F. Midgley, G. O. Hughes, and M. Rouget (2005). Niche-based modelling as a tool for predicting the risk of alien plant invasions at a global scale. *Global Change Biology 11*, 2234 – 2250.

Tobler, W. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography 46*(2), 234 – 240.

Toledo, V. M. and M. J. Ordóñez (1993). *Biological diversity of Mexico: Origins and distributions*, Chapter The biodiversity scenario of Mexico: A review of terrestrial habitats, pp. 757 – 777. Oxford University Press, Oxford.

Torres, J., J. C. Brito, M. J. Vasconcelos, L. Catarino, J. Gonçalves, and J. Honrado (2010). Ensemble models of habitat suitability relate chimpanzee (*Pan troglodytes*) conservation to forest and landscape dynamics in western Africa. *Biological Conservation 9*, 416 – 425.

Townsend, P. A. and S. J. Walsh (2001). Remote sensing of forested wetlands: Application of multitemporal and multispectral satellite imagery to determine plant community composition and structure in southeastern USA. *Plant Ecology 157*, 129 – 149.

Tsoar, A., O. Allouche, O. Steinitz, D. Rotem, and R. Kadmon (2007). A comparative evaluation of presence-only methods for modelling species distribution. *Diversity and Distributions 13*, 397 – 405.

Tuanmu, M.-N., A. Viña, S. Bearer, W. Xu, Z. Ouyang, H. Zhang, and J. Liu (2010). Mapping understory vegetation using phenological characteristics derived from remotely sensed data. *Remote Sensing of Environment 114*, 1833 – 1844.

Tuanmu, M.-N., A. Viña, G. J. Roloff, W. Liu, Z. Ouyang, H. Zhang, and J. Liu (2011). Temporal transferability of wildlife habitat models: Implications for habitat monitoring. *Journal of Biogeography 38*, 1510 – 1523.

Tucker, C. J. and P. J. Sellers (1986). Satellite remote sensing of primary production. *International Journal of Remote Sensing 7*, 1395 – 1416.

Turner, W., S. Spector, N. Gardiner, M. Fladeland, E. Sterling, and M. Steininger (2003). Remote sensing for biodiversity science and conservation, Review article. *TRENDS in Ecology and Evolution 18*(6), 306–314.

United Nations (1992). Convention on Biological Diversity. Rio de Janeiro, Brazil.

van Leeuwen, W., A. R. Huete, and T. W. Laing (1999). MODIS vegetation index compositing approach: A prototype with AVHRR data. *Remote Sensing of Environment 69*, 264 – 280.

VanDerWal, J., L. P. Shoo, C. Graham, and S. E. Williams (2009). Selecting pseudo-absence data for presence-only distribution modeling: How far should you stray from what you know? *Ecological Modelling 220*, 589 – 594.

Vaughan, I. P. and S. J. Ormerod (2005). The continuing challenges of testing species distribution models. *Journal of Applied Ecology 42*, 720 – 730.

Velázquez, A., V. M. Toledo, and I. Luna (2000). *North American terrestrial vegetation*, Chapter Mexican temperate vegetation, pp. 573 – 592. Cambridge University Press, Cambridge.

Villers-Ruíz, L. and I. Trejo-Vázquez (1998). Climate change on Mexican forests and natural protected areas. *Global Environmental Change 8*(2), 141 – 157.

Viña, A., M.-N. Tuanmu, W. Xu, Y. Li, Z. Ouyang, R. DeFries, and J. Liu (2010). Range-wide analysis of wildlife habitat: Implications for conservation. *Biological Conservation 143*, 1960 – 1969.

von Humboldt, A. and A. Bonpland (1807). *Ideen zu einer Geographie der Pflanzen nebst einem Naturgemälde der Tropenländer*. F.G. Cotta, Tübingen.

von Liebig, J. (1840). *Die organische Chemie in ihrer Anwendung auf Agrikultur und Physiologie.* Vieweg, Braunschweig.

Waide, R. B., M. R. Willig, C. F. Steiner, G. Mittelbach, L. Gough, S. I. Dodson, G. P. Juday, and R. Parmenter (1999). The relationship between productivity and species richness. *Annual Review of Ecology and Systematics 30*, 257 – 300.

Wan, Z. (2003). Monitoring thermal status of ecosystems with MODIS land-surface temperature and vegetation index products. In *Proceedings of the SPIE*, pp. 280 – 288.

Wan, Z. and J. Dozier (1996). A generalized split-window algorithm for retrieving land-surface temperature from space. *IEEE Transactions on Geoscience and Remote Sensing 34*(4), 892 – 905.

Wan, Z., P. Wang, and X. Li (2004). Using MODIS Land Surface Temperature and Normalized Difference Vegetation Index products for monitoring drought in the southern Great Plains, USA. *International Journal of Remote Sensing 25*(1), 61 – 72.

Wan, Z., Y. Zhang, Q. Zhang, and Z. Li (2002). Validation of the land-surface temperature products retrieved from Terra Moderate Resolution Imaging Spectroradiometer data. *Remote Sensing of Environment 83*, 163 – 180.

Wang, C., B. Zhou, and H. L. Palm (2008). Detecting invasive sericea lespedeza (*Lespedeza cuneata*) in Mid-Missouri pastureland using hyperspectral imagery. *Environmental Management 41*, 853 – 862.

Wang, L., W. P. Sousa, P. Gong, and G. S. Biging (2004). Comparison of IKONOS and QuickBird images for mapping mangrove species on the Caribbean coast of Panama. *Remote Sensing of Environment 91*, 432 – 440.

Waring, R. H., N. C. Coops, W. Fan, and J. M. Nightingale (2006). MODIS enhanced vegetation index predicts tree species richness across forested ecoregions in the contiguous U.S.A. *Remote Sensing of Environment 103*, 218 – 226.

Warren, D. L. and S. N. Seifert (2011). Ecological niche modeling in Maxent: The importance of model complexity and the performance of model selection criteria. *Ecological Applications 21*(2), 335 – 342.

Weng, Q., D. Lu, and J. Schubring (2004). Estimation of land surface temperature-vegetation abundance relationship for urban heat island studies. *Remote Sensing of Environment 89*, 467 – 483.

White, M. A., K. de Beurs, K. Didan, D. Inouye, A. D. Richardson, O. Jensen, J. O'Keefe, G. Zhang, R. R. Nemani, W. van LeeuWen, J. Brown, A. de Wit, M. Schaepman, X. Lin, M. Dettinger, A. Bailey, J. S. Kimball, M. D. Schwartz, D. D. Baldocchi, J. Lee, and W. Lauenroth (2009). Intercomparison, interpretation, and assessment of spring phenology in North America estimated from remote sensing for 1982 - 2006. *Global Change Biology 15*, 2335 – 2359.

Whittaker, R. H. (1956). Vegetation of the Great Smoky Mountains. *Ecological Monographs 26*, 1 – 80.

Whittaker, R. H. (1972). Evolution and measurements of species diversity. *Taxon 21*, 231 – 251.

Whittaker, R. J., M. B. Araújo, P. Jepson, R. J. Ladle, J. E. M. Watson, and K. J. Willis (2005). Conservation biogeography: Assessment and prospect. *Diversity and Distributions 11*, 3 – 23.

Wiens, J. and C. Graham (2005). Niche conservatism: Integrating evolution, ecology, and conservation biology. *Annual Reviews in Ecology and Systematics 36*, 519 – 539.

Wiens, J. A. (2002). *Predicting species occurences: Issues of accuracy and scale*, Chapter Predicting species occurrences: Progress, problems, and prospects, pp. 739 – 749. Island Press, Washington.

Wiken, E., F. Jiménez Nava, and G. Griffith (2011). North American Terrestrial Ecoregions - Level III. Technical report, Commission for Environmental Cooperation, Montreal, Canada.

Willyard, A., J. Syring, D. S. Gernandt, A. Liston, and R. Cronn (2007). Fossil calibration of molecular divergence infers a moderate mutation rate and recent radiations for *Pinus*. *Molecular Biology and Evolution 24*(1), 90 – 101.

Wilson, E. (Ed.) (1988). *Biodiversity*. National Academy Press, Washington.

Wilting, A., A. Cord, A. J. Hearn, D. Hesse, A. Mohamed, C. Traeholdt, S. M. Cheyne, S. Sunarto, M.-A. Jayasilan, J. Ross, A. C. Shapiro, A. Sebastian, S. Dech, C. Breitenmoser, J. Sanderson, J. W. Duckworth, and H. Hofer (2010). Modelling the species distribution of flat-headed cats (*Prionailurus planiceps*), an endangered South-East Asian small felid. *PLoS ONE 5*(3), e9612.

Wisz, M., N. Dendoncker, J. Madsen, M. Rounsevell, M. Jespersen, E. Kuijken, W. Courtens, C. Verscheure, and F. Cottaar (2008). Modelling pink-footed goose (*Anser brachyrhynchus*) wintering distributions for the year 2050: Potential effects of land-use change in Europe. *Diversity and Distributions 14*, 763 – 773.

Wisz, M. S., R. J. Hijmans, J. Li, A. T. Peterson, C. H. Graham, and A. Guisan (2008). Effects of sample size on the performance of species distribution models. *Diversity and Distributions 14*, 763 – 773.

Wolfe, R. E., M. Nishihama, A. J. Fleig, J. A. Kuyper, D. P. Roy, J. C. Storey, and F. S. Patt (2002). Achieving sub-pixel geolocation accuracy in support of MODIS land science. *Remote Sensing of Environment 83*, 31 – 49.

Wolfe, R. E., D. P. Roy, and E. Vermote (1998). MODIS land data storage, gridding, and compositing methodology: Level 2 grid. *IEEE Transactions on Geoscience and Remote Sensing 36*(4), 1324 – 1338.

Wolter, P. T. and P. A. Townsend (2011). Multi-sensor data fusion for estimating forest species composition and abundance in northern Minnesota. *Remote Sensing of Environment 115*, 671 – 691.

Xiao, X., S. Hagen, Q. Zhang, M. Keller, and B. Moore (2006). Detecting leaf phenology of seasonally moist tropical forests in South America with multi-temporal MODIS images. *Remote Sensing of Environment 103*, 465 – 473.

Xiao, X., Q. Zhang, S. Saleska, L. Hutyra, P. D. Camargo, S. Wofsy, S. Frolking, S. Boles, M. Keller, and B. Moore (2005). Satellite-based modeling of gross primary production in a seasonally moist tropical evergreen forest. *Remote Sensing of Environment 94*, 105 – 122.

Xie, Y., Z. Sha, and M. Yu (2008). Remote sensing imagery in vegetation mapping: A review. *Journal of Plant Ecology 1*(1), 9 – 23.

Yang, X. (2007). Integrated use of remote sensing and geographic information systems in riparian vegetation delineation and mapping. *International Journal of Remote Sensing 28*(2), 353 – 370.

Yates, C. J., A. McNeill, J. Elith, and G. F. Midgley (2010). Assessing the impacts of climate change and land transformation on Banksia in the South West Australian Floristic Region. *Diversity and Distributions 16*, 187 – 201.

Yue, W. and W. Tan (2007). The relationship between land surface temperature and NDVI with remote sensing: Application to Shanghai Landsat 7 ETM+ data. *International Journal of Remote Sensing 28*(15), 3205 – 3226.

Zhang, X., M. A. Friedl, C. B. Schaaf, and A. H. Strahler (2004). Climate controls on vegetation phenological patterns in northern mid- and high latitudes inferred from MODIS data. *Global Change Biology 10*, 1133 – 1145.

Zhang, X., M. A. Friedl, C. B. Schaaf, A. H. Strahler, J. C. F. Hodges, F. Gao, B. C. Reed, and A. Huete (2003). Monitoring vegetation phenology using MODIS. *Remote Sensing of Environment 84*, 471 – 475.

Zimmermann, N. E., T. C. Edwards Jr., G. G. Moisen, T. S. Frescino, and J. A. Blackard (2007). Remote sensing-based predictors improve distribution models of rare, early successional and broadleaf tree species in Utah. *Journal of Applied Ecology 44*, 1057 – 1067.

Zurell, D., F. Jeltsch, C. F. Dormann, and B. Schroder (2009). Static species distribution models in dynamically changing systems: How good can predictions really be? *Ecography 32*, 733 – 744.

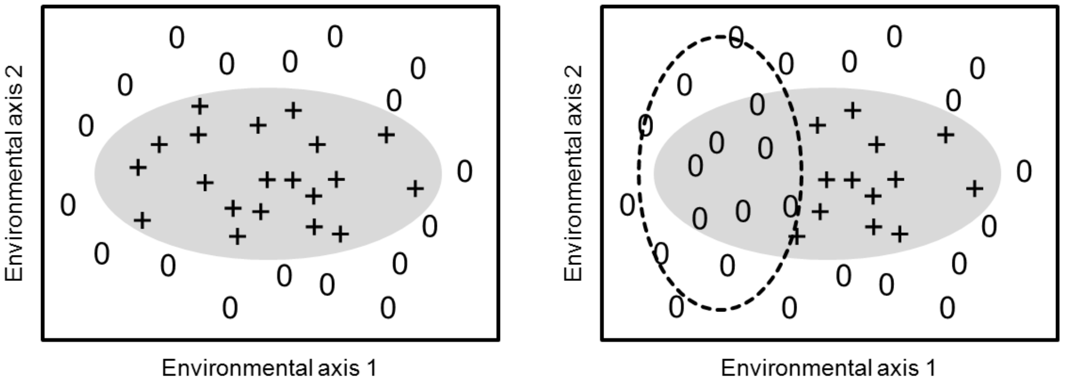
Zweig, M. H. and G. Campbell (1993). Receiver operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine. *Clinical Medicine 39*(4), 561 – 577.

Appendix

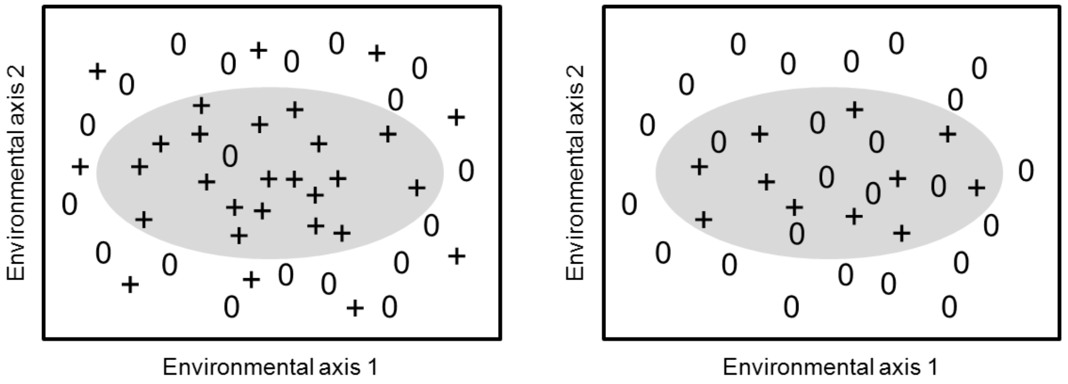
A. Glossary related to species distribution modeling.

Absence	Location where the study species was not observed or collected
Biotope	Area with uniform environmental conditions providing a living place for a specific species assemblage (typically plants)
Environmental predictor	Dependent variable used to predict species presence or absence
Environmental space	Multi-dimensional space defined by the set of environmental predictors used
Extent	Geographical size of the study area
Extrapolation	Prediction to new environments in space or time
Fundamental niche	Response of the species to its environment (resources) in absence of biotic interactions
Geographical space	Two-dimensional (x, y) map coordinates
Grain	For environmental data: grid cell or polygon size; for species records: spatial accuracy
Habitat	Description of a physical place at a particular scale in space and time, where an organism (typically an animal) actually or potentially lives
Niche	Requirements of a species to maintain positive population growth rates (see fundamental and realized niche)
Multi-collinearity	Common feature of environmental variables to be correlated with each other
Presence	Location where the study species was observed or collected, equivalent to species record or locality
Prevalence	Frequency of occurrence of the target species
Pseudo-absence	Location where the species is assumed not to occur and at which predictors are sampled, variably viewed as a sample of the 'background' or an implied absence
Range	Geographical area where a species occurs
Realized niche	Environmental dimensions in which the species can survive and reproduce, including the effects of biotic interactions
Sample size	Number or species records
Spatial auto-correlation	Common feature of environmental variables meaning that values sampled at nearby locations are not independent of each other
Spatial bias	Tendency of data not to be randomly sampled in geographical space and thus not being representative in environmental space (especially common for species presence data, e.g. due to sampling along roads or in protected areas)
Spatial scale	Geographical dimension, includes **grain** (resolution) and **extent**
Test data	(Independent or resampled) data to evaluate the model
Training data	Data (species records and predictors) used to fit the model

B. Four different views of the relationship between niche and species distribution (Pulliam, 2000).



(a) Grinnellian niche **(b) Hutchinsonian realized niche**



(c) Source-sink dynamics **(d) Dispersal limitation**

▨ Fundamental niche
⬚ Niche of competitor
\+ Observed species presence
0 Observed species absence

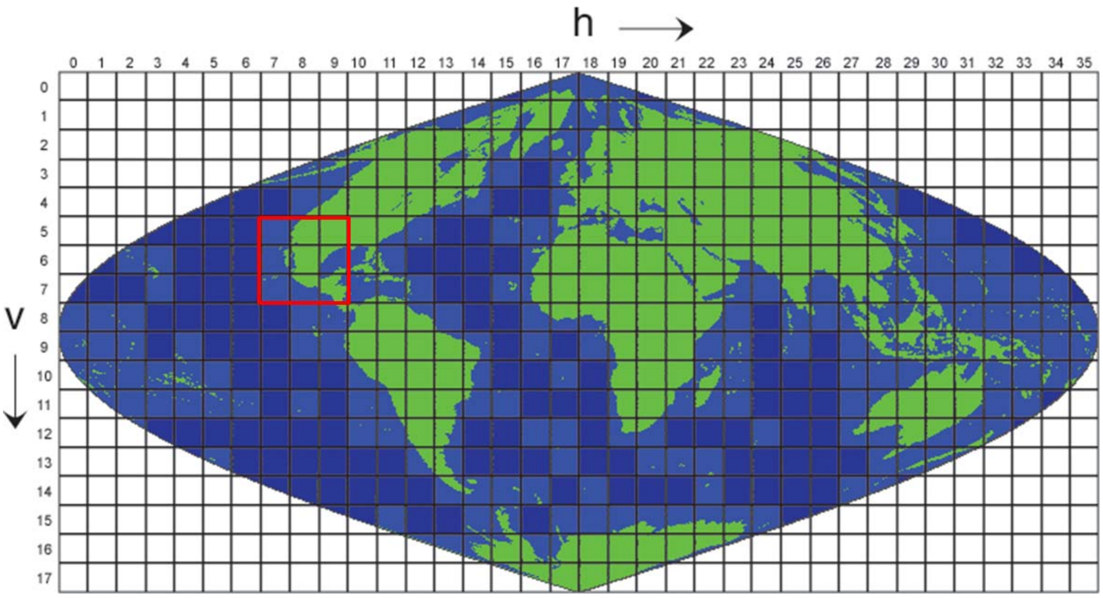
The solid oval refers to the fundamental niche or the combination of the two environmental axes displayed for which the species has a finite rate of increase greater than or equal to 1.0. The ‘pluses’ indicate the presence of the species in a patch of habitat and the ‘zeroes’ similarly indicate the absence of the species. Pulliam describes them as (a) the Grinnellian niche, where a species occurs everywhere that conditions are suitable and nowhere else, and (b) Hutchinson’s realized niche, where the species is absent from the portions of the niche that are used by a dominant competitor. According to (c) source-sink-theory, a species may also commonly occur in a sink habitat where its population growth rate is smaller than 1.0. Finally, (d) meta-population dynamics and dispersal limitations suggest that a species is frequently absent from suitable habitats because of recurring local extinction events and limited dispersal ability preventing full re-colonization.

C. Overview of published algorithms for species distribution modeling including software name, algorithms implemented, reference, and online link (partly based on Franklin, 2009; Guisan and Thuiller, 2005; Elith et al., 2006; Pearson, 2007).

Software	***Algorithms implemented***	***Key reference***	***URL***
BIOCLIM	EE	(Busby, 1991)	http://www.diva-gis.org
BIOMAPPER	ENFA	(Hirzel et al., 2002)	http://www2.unil.ch/biomapper/
BIOMOD	SRE, GLM, GAM, MARS, CTA, MDA, BRT, ANN, RF, GLM, GAM, SVM	(Thuiller, 2003)	http://r-forge.r-project.org/projects/biomod/
DIVA	EE	(Hijmans et al., 2001)	http://www.diva-gis.org
DOMAIN	EE	(Carpenter et al., 1993)	http://www.cifor.cgiar.org/docs/_ref/research_tools/domain/index.htm
DesktopGarp	Genetic algorithm	(Stockwell and Peters, 1999)	http://www.nhm.ku.edu/desktopgarp/
Maxent	ME	(Phillips et al., 2006)	http://www.cs.princeton.edu/~schapire/maxent/
OpenModeller	EE, ENFA, ANN, GARP, RF, SVM	(de Souza Muñoz et al., 2011)	http://openmodeller.sourceforge.net/

ANN: Artificial Neural Network, BRT: Boosted Regression Tree, CTA: Classification Tree Analysis, EE: Environmental (climatic) Envelope, ENFA: Ecological Niche Factor Analysis, GLM: Generalized Linear Model, GAM: Generalized Additive Model, MARS: Multivariate Adaptive Regression Splines, MDA: Mixture Discriminant Analysis, ME: Maximum Entropy, RF: Random Forest, SRE: Surface Range Envelope, SVM: Support Vector Machine.

D. MODIS tiles for processing levels 2G, 3 and 4. The nine tiles (h07v05, h08v05, h09v05, h07v06, h08v06, h09v06, h07v07, h08v07, and h09v07) that were processed are highlighted.



E. Data layers included in the MOD13A2 standard product.

Science Data Sets	***Units***	***Bit type***	***Fill***	***Valid range***	***Multiply by scale factor***
1 km 16 days NDVI	NDVI	16-bit signed integer	-3,000	-2,000-10,000	0.0001
1 km 16 days EVI	EVI	16-bit signed integer	-3,000	-2,000-10,000	0.0001
1 km 16 days VI Quality detailed QA	Bits	16-bit unsigned integer	65,535	0-65,534	NA
1 km 16 days red reflectance (Band 1)	Reflectance	16-bit signed integer	-1,000	0-10,000	0.0001
1 km 16 days NIR reflectance (Band 2)	Reflectance	16-bit signed integer	-1,000	0-10,000	0.0001
1 km 16 days blue reflectance (Band 3)	Reflectance	16-bit signed integer	-1,000	0-10,000	0.0001
1 km 16 days MIR reflectance (Band 7)	Reflectance	16-bit signed integer	-1,000	0-10,000	0.0001
1 km 16 days view zenith angle	Degree	16-bit signed integer	-10,000	-9,000-9,000	0.01
1 km 16 days sun zenith angle	Degree	16-bit signed integer	-10,000	-9,000-9,000	0.01
1km 16 days relative azimuth angle	Degree	16-bit signed integer	-4,000	-3,600-3,600	0.1
1 km 16 days composite day of the year	Julian day of year	16-bit signed integer	-1	1-366	NA
1 km 16 days pixel reliability summary QA	Rank	8-bit signed integer	-1	0-4	NA

F. Data layers included in the MOD11A2 standard product.

Science Data Sets	**Units**	**Bit type**	**Fill**	**Valid range**	**Multiply by scale factor**
LST_Day_1km: 8-Day day-time 1 km grid land surface temperature	Kelvin	16-bit unsigned integer	0	7,500-65,535	0.02
QC_Day: Quality control for day-time LST and emissivity	Bit Field	8-bit unsigned integer	see QA note	0-255	NA
Day_view_time: Average time of day-time land surface temperature observation	Hours	8-bit unsigned integer	255	0-240	0.1
Day_view_angle: Average view zenith angle of day-time land surface temperature	Degree	8-bit unsigned integer	255	0-130	1 (-65)
LST_Night_1km: 8-Day night-time 1 km grid land surface temperature	Kelvin	16-bit unsigned integer	0	7,500-65,535	0.02
QC_Night: Quality control for night-time LST and emissivity	Bit Field	8-bit unsigned integer	see QA note	0-255	NA
Night_view_time: Average time of night-time land surface temperature observation	Hours	8-bit unsigned integer	255	0-240	0.1
Night_view_angle: Average view zenith angle of night-time land surface temperature	Degree	8-bit unsigned integer	255	0-130	1 (-65)
Emis_31: Band 31 Emissivity	none	8-bit unsigned integer	0	1-255	0.0020 (+0.49)
Emis_32: Band 32 Emissivity	none	8-bit unsigned integer	0	1-255	0.0020 (+0.49)
Clear_sky_days: the days in clear sky conditions and with valid LSTs	None	8-bit unsigned integer	0	1-255	NA
Clear_sky_nights: the nights in clear sky conditions and with valid LSTs	None	8-bit unsigned integer	0	1-255	NA

G. Literature review of phenological metrics derived from remotely sensed time series of the vegetation index for various purposes. All metrics except the number of seasons per year (which is only one in Mexico except for agricultural areas) were implemented in this thesis. Blank cells indicate that the respective metric was not used in the study.

see next page

Authors	**(Reed et al., 1994)**	**(Zhang et al., 2003)**	**(Jönsson and Eklundh, 2004)**	**(Bradley et al., 2007)**	**(Fisher and Mustard, 2007)**	**(Maignan et al., 2008)**	**(Hird and McDermid, 2009)**	**(White et al., 2009)**	**This thesis**
Topic/Objectives	Utility of phenological metrics for vegetation mapping	New methodology to monitor global vegetation phenology	Three different filtering methods for processing time-series data	Curve fitting methodology for remotely sensed time series	Comparison of phenology derived from Landsat and MODIS data	Methodology to derive global phenological parameters	Comparison of six noise-reduction techniques	Comparison of 10 methods to derive start of season	
Study area	USA	New England	Africa	Great Basin (USA)	New England (USA) – local study	Europe/ global	Canada (local study)	North America	
Sensor and data	AVHRR - NDVI	MODIS - EVI	AVHRR - NDVI	AVHRR - NDVI	Landsat; MODIS – MOD09A	AVHRR – Red/NIR (DVI/NDVI)	MODIS - NDVI	AVHRR - NDVI	
Resolution	1 km	1 km	8 km	1 km	30 m; 500 m	8 km	250 m	8 km	
Time period	1989 - 1992	2000 - 2001	1998 - 2000	1990 - 2002	Landsat: 1984 – 2002; MODIS: 2000 - 2005	1982 - 1999	2003 - 2005	1982 – 2006	
Temporal metrics									
Start of season (SOS)	X	X	X	X	X	X	X	X	X
Mid of season (MOS)	X	X	X				X		X
End of season (EOS)	X	X	X			X	X		X
Dormancy		X							X
Length of season	X		X				X		X

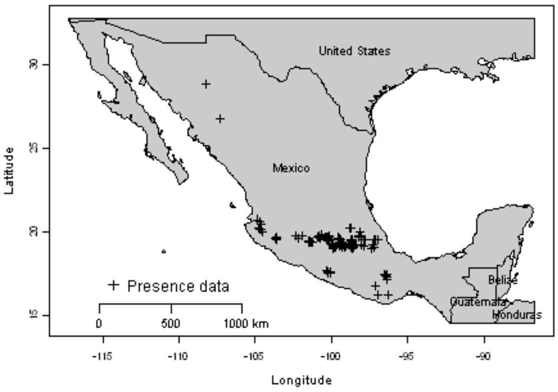
Authors	**(Reed et al., 1994)**	**(Zhang et al., 2003)**	**(Jönsson and Eklundh, 2004)**	**(Bradley et al., 2007)**	**(Fisher and Mustard, 2007)**	**(Maignan et al., 2008)**	**(Hird and McDermid, 2009)**	**(White et al., 2009)**	**This thesis**
NPP-related metrics									
Value at SOS	X								X
Value at MOS	X		X				X		X
Value at EOS	X								X
Amplitude	X		X				X		X
Integral	X		X				X		X
Minimum			X						X
Mean							X		X
Median									X
Seasonality-related metrics									
Rate of green-up	X						X		X
Rate of senescence	X								X
Shape of curve			X						X
Number of seasons	X		X						
Standard deviation									X
Coefficient of variation									X

H. Phenological and statistical metrics used for species distribution modeling. All values are annual scores and hence refer to one year. Abbreviations: EVI – *Enhanced Vegetation Index*, LST – *Land Surface Temperature*

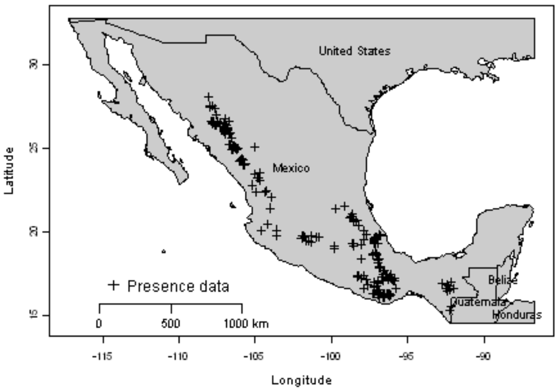
Short name	***Description***
EVI_date_SOS	Start of season
EVI_date_maximum	Mid of season
EVI_date_EOS	End of season
EVI_date_dormancy	Dormancy
EVI_length_season	Length of season
EVI_value_SOS	EVI score at start of season
EVI_maximum	Maximum EVI
EVI_value_EOS	EVI score at end of season
EVI_range	EVI range
EVI_integral	Accumulated sum of EVI values during growing season
EVI_minimum	Minimum EVI
EVI_mean	Mean EVI
EVI_median	Median EVI
EVI_rate_greenup	Rate of greenup
EVI_rate_senescence	Rate of senescence
EVI_skewness	Shape of annual phenology curve
EVI_CoV	Coefficient of variation of EVI
EVI_standard_deviation	Standard deviation of EVI
LST_CoV	Coefficient of variation of LST
LST_maximum	Maximum LST
LST_minimum	Minimum LST
LST_mean	Mean LST
LST_median	Median LST
LST_range	LST range
LST_standard_deviation	Standard deviation of LST
blue_CoV	Coefficient of variation of blue reflectance
blue_maximum	Maximum blue reflectance
blue_minimum	Minimum blue reflectance
blue_mean	Mean blue reflectance
blue_median	Median blue reflectance
blue_range	Blue reflectance range
blue_standard_deviation	Standard deviation of blue reflectance
red_CoV	Coefficient of variation of red reflectance
red_maximum	Maximum red reflectance
red_minimum	Minimum red reflectance

Short name	***Description***
red_mean	Mean red reflectance
red_median	Median red reflectance
red_range	Red reflectance range
red_standard_deviation	Standard deviation of red reflectance
NIR_CoV	Coefficient of variation of NIR reflectance
NIR_maximum	Maximum NIR reflectance
NIR_minimum	Minimum NIR reflectance
NIR_mean	Mean NIR reflectance
NIR_median	Median NIR reflectance
NIR_range	NIR reflectance range
NIR_standard_deviation	Standard deviation of NIR reflectance
MIR_CoV	Coefficient of variation of MIR reflectance
MIR_maximum	Maximum MIR reflectance
MIR_minimum	Minimum MIR reflectance
MIR_mean	Mean MIR reflectance
MIR_median	Median MIR reflectance
MIR_range	MIR reflectance range
MIR_standard_deviation	Standard deviation of MIR reflectance

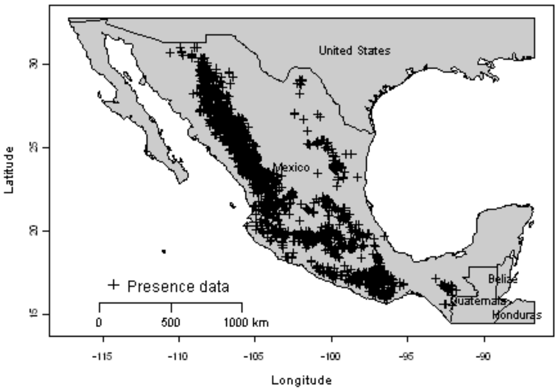
I. Spatial distribution of species presence sites (data of the *National Forest Inventory*). The twelve tree species are representative of the major Mexican forest types.



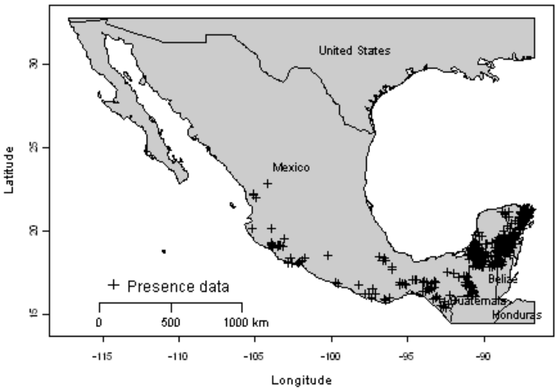
Abies religiosa



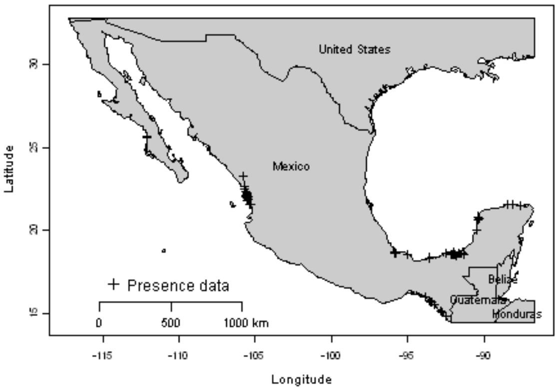
Alnus acuminata



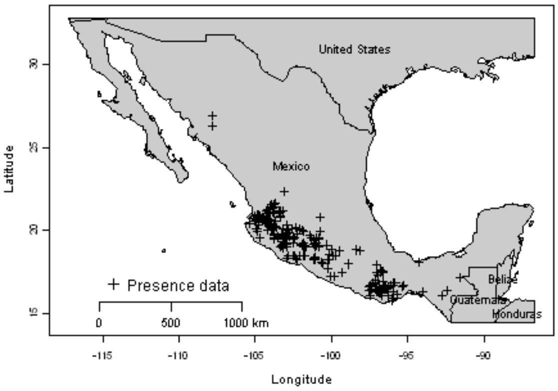
Arbutus xalapensis



Astronium graveolens

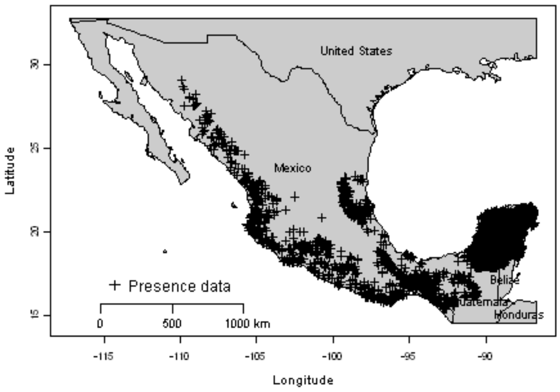


Avicennia germinans

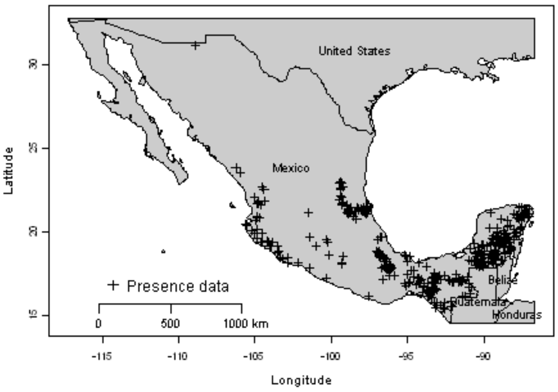


Bursera bipinnata

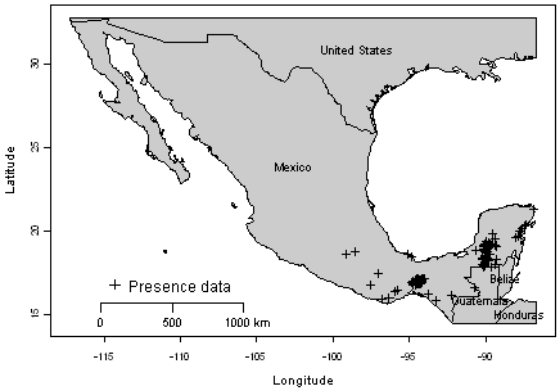
Bursera simaruba



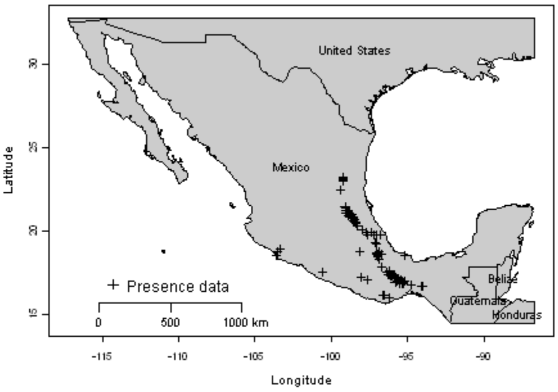
Cedrela odorata



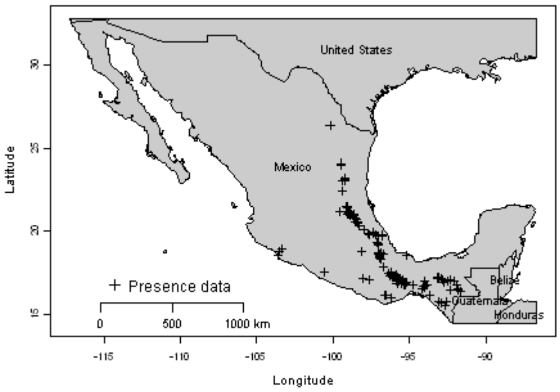
Guaiacum sanctum



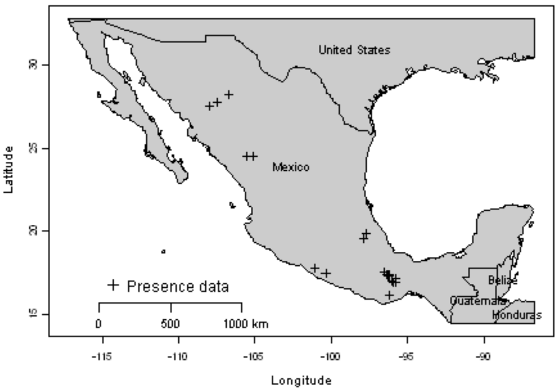
Liquidambar macrophylla



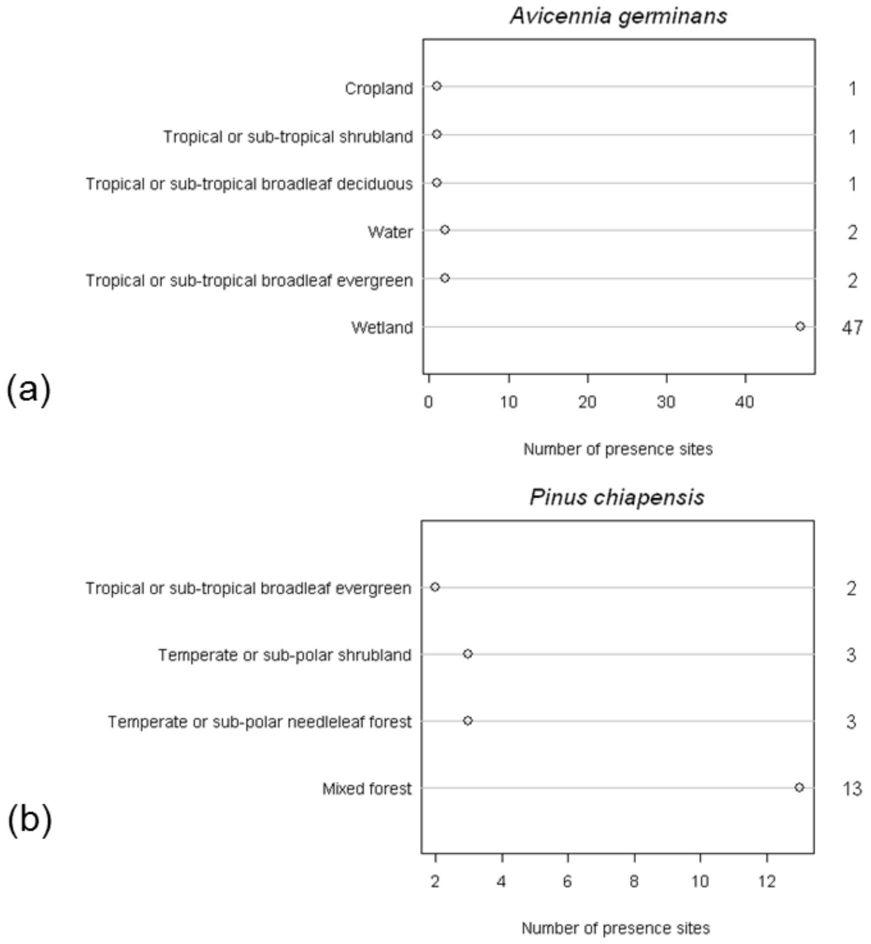
Liquidambar styraciflua



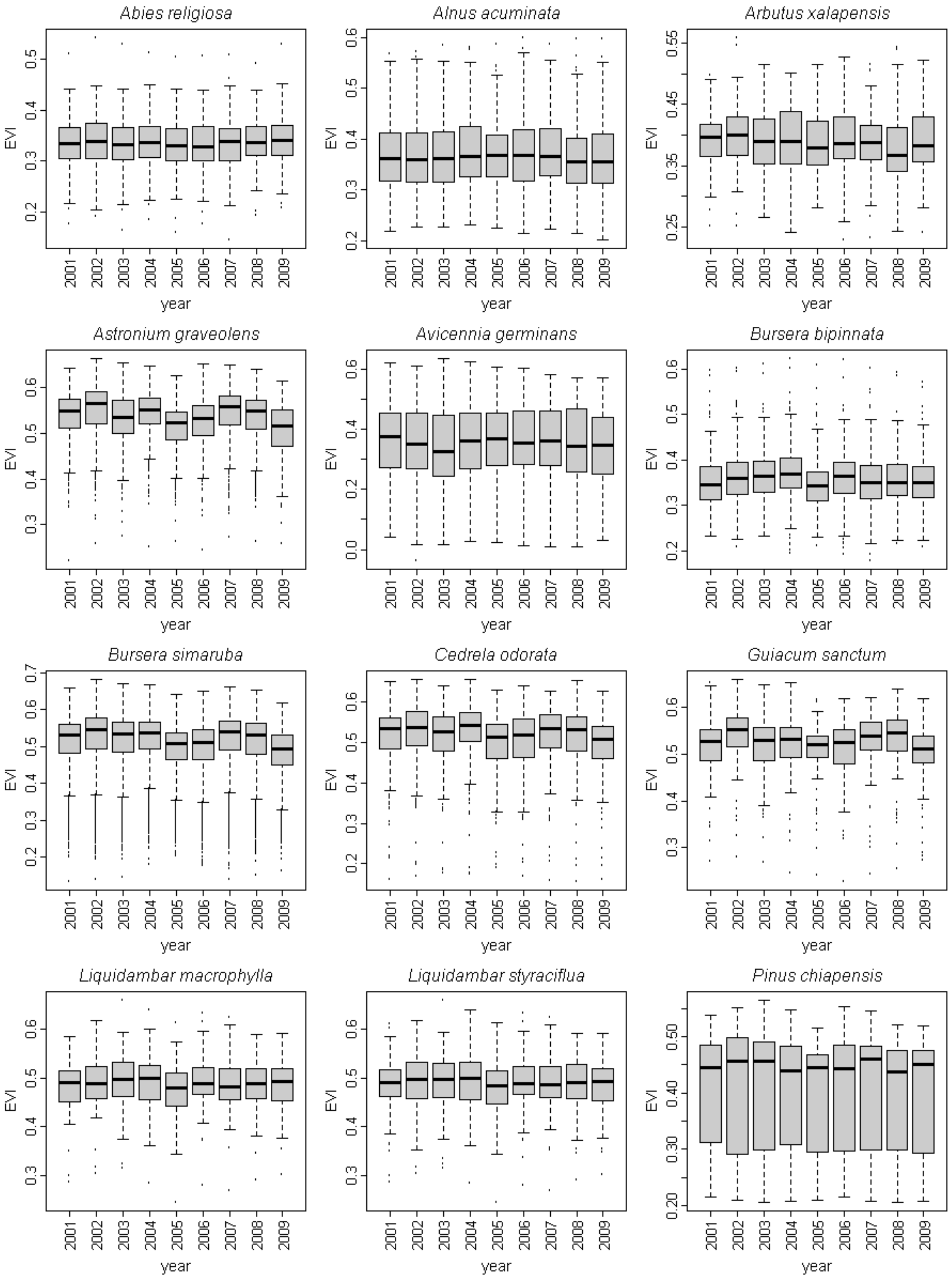
Pinus chiapensis



J. Distribution of land cover classes observed at presence localities of selected study species. (a) *Avicennia germinans*, (b) *Pinus chiapensis*.



K. Inter-annual differences in the scores of the mean *Enhanced Vegetation Index* (*EVI_mean*) observed at the species presence sites.



L. Range sizes of Mexican pine species measured by the *fractional predicted area* (FPA, Section 3.4.5) of the BIOCLIM and BIOCLIM_REMOTE models. The last column indicates the percentage decrease of *fractional predicted area* from climatic suitable to currently available biotopes after the inclusion of remote sensing data.

Species	***BIOCLIM***	***BIOCLIM_REMOTE***	***% decrease***
Pinus arizonica	0.1372	0.0663	51.7
Pinus ayacahuite	0.0248	0.0080	67.7
Pinus californiarum	0.0029	0.0007	75.9
Pinus cembroides	0.1428	0.0790	44.7
Pinus chiapensis	0.0193	0.0035	81.9
Pinus chihuahuana	0.0871	0.0448	48.6
Pinus cooperi	0.0345	0.0208	39.7
Pinus culminicola	0.0007	0.0001	85.7
Pinus devoniana	0.0180	0.0010	94.4
Pinus discolor	0.0657	0.0323	50.8
Pinus douglasiana	0.0281	0.0111	60.5
Pinus durangensis	0.0612	0.0437	28.6
Pinus engelmannii	0.0967	0.0785	18.8
Pinus greggii	0.0249	0.0071	71.5
Pinus hartwegii	0.0129	0.0033	74.4
Pinus herrerae	0.0191	0.0117	38.7
Pinus jaliscana	0.0024	0.0011	54.2
Pinus jeffreyi	0.0023	0.0001	95.7
Pinus johannis	0.0183	0.0030	83.6
Pinus lambertiana	0.0002	0.0001	50.0
Pinus lawsonii	0.0094	0.0023	75.5
Pinus leiophylla	0.0621	0.0204	67.1
Pinus lumholtzii	0.0867	0.0436	49.7
Pinus maximinoi	0.0460	0.0076	83.5
Pinus montezumae	0.0184	0.0053	71.2
Pinus nelsonii	0.0086	0.0029	66.3
Pinus oocarpa	0.0480	0.0032	93.3
Pinus patula	0.0077	0.0013	83.1
Pinus pinceana	0.0978	0.0258	73.6
Pinus ponderosa	0.0470	0.0192	59.1
Pinus praetermissa	0.0325	0.0111	65.8
Pinus pringlei	0.0083	0.0014	83.1
Pinus pseudostrobus	0.0484	0.0127	73.8
Pinus quadrifolia	0.0054	0.0031	42.6
Pinus radiata	0.0002	0.0001	50.0
Pinus remota	0.0199	0.0055	72.4
Pinus rzedowskii	0.0005	0.0001	80.0
Pinus strobiformis	0.0815	0.0350	57.1
Pinus tecunumanii	0.0054	0.0014	74.1
Pinus teocote	0.0420	0.0146	65.2
Mean	**0.0369**	**0.0158**	**65.1**

List of figures

List of tables

Curriculum Vitae

Anna Cord
Ulrichstr. 12
97074 Würzburg
born April 02, 1982 in Saarbrücken

Education

since 10/2007	**Ph.D. student**, University of Würzburg, in cooperation with the German Remote Sensing Data Center (DFD) and CONABIO (National Commission for the Knowledge and Use of Biodiversity), Mexico
08/2004 - 07/2007	**Diploma in Biology**, University of Würzburg (Grade: 1.0) Thesis: "Classifying land cover using FAO-LCCS standards in a tree savannah of Burkina Faso with ASTER and IKONOS Data" Major subject: Plant eco-physiology and vegetation ecology Minor subjects: Animal ecology and tropical biology, molecular plant physiology and biophysics
04/2005 - 07/2006	**Intermediate Diploma in Geography**, University of Würzburg (Grade: 1.5) Major subjects: Physical geography, human geography Minor subjects: Zoology, botany Specialization: Remote sensing, Geographic Information Systems (GIS)
01/2004 - 07/2004	**Semester abroad**, University of Umeå, Sweden
10/2001 - 12/2003	**Intermediate Diploma in Biology**, University of Würzburg (Grade: 1.3)
09/1992 - 06/2001	**School education**, Spessart-Gymnasium Alzenau (Grade: 1.0)

Professional Experience

since 10/2007	**Research assistant**, German Remote Sensing Data Center (DFD), German Aerospace Center (DLR) in cooperation with the Department of Remote Sensing, University of Würzburg
03/2008 - 09/2010	**Scientific assistant** to Prof. S. Dech, Department of Remote Sensing, University of Würzburg
04/2005 - 07/2005 04/2006 - 07/2006	**Undergraduate teaching assistant,** Faculty of Biology, University of Würzburg
06/2005 - 07/2005	**Undergraduate research assistant,** German Remote Sensing Data Center (DFD), German Aerospace Center (DLR)
10/2002 - 02/2003	**Undergraduate teaching assistant,** Faculty of Biology, University of Würzburg
2002 - 2005	**Student mentor**, Faculty of Biology, University of Würzburg

Awards and Grants

07/2010	**Best student paper award**, International Environmental Modelling and Software Society (iEMSs) conference in Ottawa, Canada
10/2006 - 11/2006	**Travel grant**, German National Academic Foundation ("Studienstiftung des Deutschen Volkes")
03/2004 - 07/2007	**Student scholarship**, German National Academic Foundation

Publications

Journal papers

Cord, A. & Rödder, D. (2011). Inclusion of habitat availability in species distribution models through multi-temporal remote sensing data?. *Ecological Applications*, 21(8), 3285-3298.

Hüttich, C., Herold, M., Wegmann, M., **Cord, A.**, Strohbach, B., Schmullius, C. & Dech, S. (2011). Assessing effects of temporal compositing and varying observation periods for large-area land-cover mapping in semi-arid ecosystems: Implications for global monitoring. *Remote Sensing of Environment*, 115(10), 2445-2459.

Cord, A., Conrad, C., Schmidt, M. & Dech, S. (2010). Standardized FAO-LCCS land cover mapping in heterogeneous tree savannas of West Africa. *Journal of Arid Environments*, 74(9), 1083-1091.

Wilting, A.*, **Cord, A.***, Hearn, A., Hesse, D., Mohamed, A., Traeholdt, C., Cheyne, S.M., Sunarto, S., Jayasilan, M.-A., Ross, J., Shapiro, A.C., Sebastian, A., Dech, S., Breitenmoser, C., Sanderson, J., Duckworth, J.W. & Hofer, H. (2010). Modelling the species distribution of flat-headed cats (*Prionailurus planiceps*), an endangered South-East Asian small felid. *PLoS ONE*, 5(3), e9612.

* equal contribution

Conference proceedings

Cord, A., Klein, D. & Dech, S. (2011). The impact of inter-annual variability in remote sensing time series on modeling tree species distributions. MultiTemp, *Proceedings of the 6th International Workshop on the Analysis of Multi-Temporal Remote Sensing Images*, July 12-14, 2011, Trento/Italy.

Cord, A., Klein, D. & Dech, S. (2010). Remote sensing time series for modeling invasive species distribution: A case study of *Tamarix* spp. in the US and Mexico. *Proceedings of the International Congress on Environmental Modelling and Software (iEMSs)*, July 5-8, 2010, Ottawa/Canada.

Colditz, R.R., **Cord, A.**, Conrad, C., Mora, F., Maeda, P. & Ressl, R. (2009). Analyzing phenological characteristics of Mexico with MODIS time series products. MultiTemp, *Proceedings of the 5th international Workshop on the Analysis of Multi-Temporal Remote Sensing Images*, July 28-30, 2009, Mystic, Connecticut/USA.

Cord, A., Colditz, R.R., Schmidt, M. & Dech, S. (2009). Species distribution and forest type mapping in Mexico. *Proceedings of the IEEE International Geoscience & Remote Sensing Symposium*, July 12-17, 2009, Cape Town/South Africa.

Cord, A., Schmidt, M. & Dech, S. (2009). Potential and limitations of multi-temporal earth observation aata to improve modeled results of tree species distribution in Mexico. *Proceedings of the 33rd International Symposium on Remote Sensing of the Environment*, May 4-8, 2009, Stresa/Italy.

Machwitz, M., Landmann, T., Conrad, C., **Cord, A.** & Dech, S. (2008). Land cover analysis on sub-continental scale: FAO LCCS standard with 250 meter MODIS satellite observations in West Africa. *Proceedings of the IEEE International Geoscience & Remote Sensing Symposium*, July 6-11, 2008, Boston, Massachusetts/USA.

Landmann, T., Vlek, P. J., Schmidt, M., Dech, S. & **Cord, A.** (2007). An object-conditional approach for satellite remote sensing land cover mapping in African Savannas. *Proceedings of the Third International Workshop on Image Fusion and Geographic Information Systems*, May 21-25, 2007, St. Petersburg/Russia.

Oral presentations

Cord, A., Conrad, C. & Dech, S. (2010). Ableitung von standardisierten Landbedeckungsdaten in Savannen nach dem LCCS-Klassifikationsschema. *Tagung des AK Fernerkundung innerhalb der Deutschen Gesellschaft für Geographie (DGfG)*, October 7-8, 2010, Heidelberg/Germany.

Cord, A., Klein, D. & Dech, S. (2010). From mapping to modeling: Integrating multi-temporal remote sensing data into species distributions models at different scales. *40th Symposium of the Ecological Society of Germany, Austria and Switzerland (GfÖ)*, Book of Abstracts, p. 324, August 30 - September 2, 2010, Gießen/Germany.

Landmann, T., Machwitz, M., **Cord, A.**, Hüttich, C. & Dech, S. (2009). Using the Land Cover Classification System (LCCS) for interoperable determination of land cover and land cover change in Africa. *3rd Workshop of the EARSeL Special Interest Group on Land Use and Land Cover*, November 25-27, 2009, Bonn/Germany.

Language Skills

German	Native
English	Fluent
Swedish	Basic
Latinum, Graecum	

Computing Skills

Programing	R, IDL
Modeling	Maxent, BIOMOD
Statistics	R, SPSS
GIS	ArcGIS, DIVA-GIS, QGIS
Remote Sensing	ENVI, ERDAS Imagine, TIMESAT

Acknowledgements

This thesis could not have been seen through to its conclusion without the help of many people:

First and foremost, I would like to thank my supervisor Prof. Dr. Stefan Dech, who provided the unique opportunity for my PhD research in collaboration between the *German Remote Sensing Data Center* (DFD) of the *German Aerospace Center* (DLR) and the *University of Würzburg*. I am grateful for his scientific advice and informative feedback during all stages of this thesis and above all for his continuous personal support, guidance, and encouragement. Secondly, I would like to thank Prof. Dr. Björn Reineking (*University of Bayreuth*), the second member of the scientific committee, for the valuable discussions, criticism, and his suggestions about how my work might be developed or improved.

Special acknowledgement is also due to my scientific mentor Dr. Doris Klein (DLR) for her instructive comments and constructive criticism at different stages of my thesis as well as for her personal encouragement and guidance. Our long discussions helped me sort out many of the technical details of my work.

Further, I would like to thank Prof. Dr. Christopher Conrad for his scientific support, for fruitful discussions, and for creating an open and inspiring working atmosphere at the *Department of Remote Sensing* at the *University of Würzburg*.

In addition, I want to extend my appreciation to the numerous past and present members of this department with whom I have had the privilege to work over the course of my postgraduate studies: Sarah Asam, Sebastian Fritsch, Dr. Ursula Geßner, Vitus Himmler, Patrick Knöfel, Carina Kübert, Dr. Tobias Landmann, Sylvia Lex, Fabian Löw, Dr. Miriam Machwitz, Dr. Noellie Rüth, Gunther Schorcht, Martin Schmidt, Dr. Matthias Schramm, Michael Thiel, Yvonne Walz, Dr. Martin Wegmann, and Julian Zeidler. Special thanks to Dr. Doris Klein, Sarah Asam, Dr. Miriam Machwitz, Yvonne Walz, Carina Kübert, and Florian Niedlich for proof-reading and valuable suggestions that greatly improved the earlier manuscript. Further, I would like to thank Sarah Asam, Carina Kübert, Vitus Himmler, and Julian Zeidler for allowing me to occupy their computers in order to process the time series data and to run the species distribution models. Many thanks also to Julian Zeidler, Gunther Schorcht, and Christian Bäuerlein for their technical support regarding any sort of minor and major hardware and software problems. Moreover, I am grateful to my office colleagues Dr. Christian Hüttich, Dr. Miriam Machwitz, Dr. Noellie Rüth, and later on Sarah Asam and Carina Kübert for their companionship and their valuable help with a number of technical, contextual or layout problems and questions.

Furthermore, this thesis would not have been possible without the continuous support of my former colleagues Dr. Michael Schmidt and Dr. René Colditz, now working at the *Mexican National Commission for the Knowledge and Use of Biodiversity* (CONABIO). Special thanks also to Dr. Rainer Ressl for his logistic help during my stays in Mexico.

The thesis has largely benefited from the discussions with Dr. Franz Mora (CONABIO) and his long-term experience in modeling species distributions in Mexico. Also, I am indebted to several other colleagues at CONABIO, especially Isabel Trejo Aguilar, Mariana Munguía Carrara, Dr. Pilar Rodríguez Moreno, and Marco Zea Salgado for the provision of species occurrence data, many inspiring and fruitful discussions, technical support, and their warm welcome. In particular, I would like to acknowledge Yvonne Simms Del Castillo for welcoming me to her home, for her help with arranging my research stays in Mexico, and for our friendship.

My thanks are also due to Dr. David Gernandt and his research group (*National Autonomous University of Mexico*, UNAM) as well as to Jorge Pérez de la Rosa (*University of Guadalajara*) for granting me access to their data base of pine occurrence records and for sharing their knowledge concerning the enormous diversity of Mexican pines and their ecology with me. Many thanks also to David’s family for letting me stay with them and showing me around Mexico-City.

I am grateful as well to Dr. Oswaldo Tellez Valdés and his students (UNAM) for taking me along to their field work in the Tehuacán-Cuicatlán valley and sharing all their knowledge concerning the Mexican flora. In addition, I would like to thank Dr. Jorge Vega Rivera, head of the field station in the *Chamela-Cuixmala Biosphere Reserve*, for the opportunity to stay at this unique place. Special thanks also to the PhD students Rodrigo Méndez Alonzo and César Manrique Asencio for sharing their field experience and taking me with them on their Sunday trips to Careyes.

In addition, I would like to thank all of my friends for their patience and encouragement during the last years, as well as my underwater rugby team and scout group for taking my mind off work.

Finally, many thanks are due to Florian Niedlich, who discussed my work with me at all possible and impossible times and provided me with help and encouragement when it was most needed. Most importantly, I would like to thank my parents Dr. Ute and Dr. Bernhard Cord, who have always encouraged and believed in me. Without their continuous support, I would not have been able to realize this thesis.

Anna Cord

Eidesstattliche Erklärung

Ich erkläre hiermit, dass die von mir eingereichte Dissertation zum Thema *Potential of multi-temporal remote sensing data for modeling tree species distributions and species richness in Mexico* selbstständig und ausschließlich unter Verwendung der angegebenen Literatur und Hilfsmittel verfasst wurde. Alle den angeführten Quellen wörtlich oder sinngemäß entnommenen Stellen habe ich als solche kenntlich gemacht.

Diese Arbeit ist keiner anderen Prüfungsbehörde vorgelegt worden.

Würzburg, den 06.03.2012

Anna Cord