



**UNIVERSITY  
OF TURKU**

# **USING REMOTE SENSING TO MAP TREE SPECIES DISTRIBUTIONS AND FLORISTIC PATTERNS OF TREES IN PERUVIAN AMAZONIA AT DIFFERENT EXTENTS**

---

**Pablo Pérez Chaves**







UNIVERSITY  
OF TURKU

# **USING REMOTE SENSING TO MAP TREE SPECIES DISTRIBUTIONS AND FLORISTIC PATTERNS OF TREES IN PERUVIAN AMAZONIA AT DIFFERENT EXTENTS**

---

Pablo Pérez Chaves

## University of Turku

---

Faculty of Science and Engineering  
Department of Biology  
Biosciences / Biodiversity Research  
Doctoral programme in Biology, Geography and Geology

## Supervised by

---

Professor Hanna Tuomisto  
Department of Biology  
University of Turku, Finland

Dr. Risto Kalliola  
Department of Geography  
University of Turku, Finland

Dr. Kalle Ruokolainen  
Department of Biology  
University of Turku, Finland

## Reviewed by

---

Dr. Bruce Nelson  
National Institute of Amazonian  
Research (INPA), Brazil

Dr. Sassan Saatchi  
Jet Propulsion Laboratory, California  
Institute of Technology, USA

## Opponent

---

Professor Duccio Rocchini  
University of Bologna, Italy

The originality of this publication has been checked in accordance with the University of Turku quality assurance system using the Turnitin OriginalityCheck service.

Cover Image: Landsat TM/ETM+ composite (bands 4, 7, 3) covering a part of northern Peruvian Amazonia

ISBN 978-951-29-8289-9 (PRINT)  
ISBN 978-951-29-8290-5 (PDF)  
ISSN 0082-6979 (Print)  
ISSN 2343-3183 (Online)  
Painosalama Oy, Turku, Finland 2020

UNIVERSITY OF TURKU

Faculty of Science and Engineering

Department of Biology

Biosciences / Biodiversity Research

PABLO PEREZ CHAVES: Using remote sensing to map tree species distributions and floristic patterns of trees in Peruvian Amazonia at different extents

Doctoral Dissertation, 128 pp.

Doctoral Programme in Biology, Geography and Geology

November 2020

## ABSTRACT

Amazonia, the largest and most diverse tropical forest in the world, continues facing pressures and losing forest cover while vast areas remain biologically poorly known. Knowing where a species occurs, and the distribution of biologically unique areas is key in conservation and natural resource management. Collecting biological information is a basic requirement for determining such biologically important areas but it is time-consuming and expensive, especially in tropical and remote areas such as Peruvian Amazonia. Many forest institutions and forestry companies have already collected systematic tree data in Peruvian Amazonia, covering large and even remote areas. Also, remote sensing provides continuous information that is useful for biodiversity assessments over large extents. Therefore, combining existing field data on trees with available remote sensing and environmental layers would allow predicting spatial biodiversity patterns in large areas, even where field data is still missing. Here, I use forest inventory and census data from the Peruvian Amazon together with Landsat-derived predictors (such as optical bands from Landsat TM/ETM+ and vegetation indices) and other environmental layers (e.g., elevation) to model the distribution of trees at local extents (2,500–6,000 km<sup>2</sup>) and to predict floristic patterns of trees at broader extents (40,000–800,000 km<sup>2</sup>). At local extents, where relevant environmental layers are often not available, I found that the average reflectance values of Landsat imagery and elevation are important variables for predicting the distribution of tree species. At broader extents, floristic patterns of trees were mainly correlated with the reflectance values derived from Landsat imagery but also with climate layers. This enabled producing a predictive map of the main floristic gradients of trees throughout Peruvian Amazonia. Since taxonomic inaccuracies might exist in forestry data, it is important to assess data consistency when using it for ecological studies. Here, I found that using the entire tree inventory data set and analyzing it at the genus level yielded more congruent floristic patterns than analyses using only those tree stems that were identified to the species level. The combination of forest inventory and census data with available remote sensing and environmental layers offers an efficient way of mapping continuous biodiversity patterns, such as tree species distributions and tree community composition, in still poorly known regions such as Peruvian Amazonia. Mapping tree species distributions and floristic patterns of trees at different geographical extents is useful for (i) characterizing biodiversity patterns and forest resources in areas where field work is missing, (ii) identifying suitable areas for forest management and conservation purposes and hence for (iii) delivering practical information for decision-makers.

## TURUN YLIOPISTO

Luonnontieteiden ja tekniikan tiedekunta

Biologian laitos

Biotieteet / Biologisen monimuotoisuuden tutkimus

PABLO PÉREZ CHAVES: Using remote sensing to map tree species distributions and floristic patterns of trees in Peruvian Amazonia at different extents

Väitöskirja, 128 s.

Biologian, maantieteen ja geologian tohtorihjelma

Marraskuu 2020

### TIIVISTELMÄ

Amazonia on maailman suurin ja monimuotoisin trooppinen metsäalue. Metsiä halutaan hyväksikäyttää mutta samalla niitä myös tuhoetaan ilman että niitä olisi biologisesti juurikaan tutkittu. Tietämys lajien esiintymisalueista ja lajistollisesti erityisistä alueista on perusedellytys sekä suojelun että luonnonvarojen käytön suunnittelun onnistumiselle. Tämän tietämyksen kerääminen on kuitenkin kallista ja työlästä erityisesti Perun Amazonian kaltaisilla syrjäisillä alueilla. Tästä huolimatta useat tutkimuslaitokset ja metsäyritykset ovat toteuttaneet puulajistoinventointeja alueen eri osissa. Arvokasta informaatiota metsien vaihtelusta saadaan myös tasaisen kattavista kaukokartoitusaineistoista. Yhdistämällä pistemäistä maastoinventointitietoa kaukokartoitusaineistoihin ja muihin kattaviin ympäristötietoihin voidaan arvioida biodiversiteetin tilaa sellaisillakin alueilla, joilta ei ole saatavissa suoria maastohavaintoja. Hyödynnän työssäni metsäinventointeja ja Landsat-satelliittikuvia (esimerkiksi Landsat TM/ETM+ -kuvien optisia kanavia ja kasvillisuusindeksejä) puulajien levinneisyyksien mallintamiseen paikallisesti (2500–6000 km<sup>2</sup>) ja alueellisesti Perun Amazoniassa. Landsat-kuvien keskimääräiset heijastusarvot ja alueen korkeus merenpinnasta osoittautuivat paikallisesti hyvin toimiviksi ennustemuuttujiksi. Laajemmassa alueellisessa tarkasteluissa (40 000–800 000 km<sup>2</sup>) puulajiston vaihtelu korreloi myös muihin ympäristömuuttujiin, esimerkiksi ilmastoon ja pinnanmuotoihin. Esitän työssäni koko Perun Amazonian kattavia puiden levinneisyyksien ennustemalleja. Metsäinventointiaineistoihin voi kuitenkin sisältyä lajien tunnistamisen epätarkkuutta ja sen merkitystä tulee ekologisten tutkimusten yhteydessä kriittisesti arvioida. Havaitsin, että ennustemallien laatua voidaan parantaa muuntamalla lajitasoista inventointitietoa sukutasoiseksi informaatioksi. Metsäinventointitietojen, kaukokartoituksen ja ympäristötietojen yhteiskäyttö tarjoaa tutkimukseni valossa tehokkaan keinon luonnon monimuotoisuuden kartoittamiseen Perun Amazonian kaltaisilla heikosti tunnetuilla seuduilla. Kun puustokartoituksia tehdään eri mittakaavatasoilla, ne tukevat (i) heikosti tunnettujen metsien monimuotoisuuden ja resurssien arviointia, (ii) luonnonvarojen käyttöön tai suojeluun soveltuvien alueiden määrittämistä ja (iii) vahvistavat käytännön päätöksenteossa tarvittavaa tietoperustaa.

# Table of Contents

<b>Table of Contents</b> .....	<b>5</b>
<b>Abbreviations</b> .....	<b>7</b>
<b>List of Original Publications</b> .....	<b>8</b>
<b>1 Introduction</b> .....	<b>9</b>
1.1 Usefulness of remote sensing in biodiversity studies.....	10
1.2 Using forest inventories and forest census data for ecological studies .....	13
1.3 Species distribution models of trees at local extents.....	14
1.4 Floristic patterns of trees in Amazonia.....	16
1.5 Aims of the thesis .....	18
<b>2 Methods</b> .....	<b>19</b>
2.1 Study area.....	19
2.2 Forest inventory and census data.....	21
2.2.1 Forest census data.....	22
2.2.2 Forest inventory data.....	23
2.3 Predictor layers .....	23
2.3.1 Landsat satellite imagery.....	23
2.3.2 Environmental layers.....	24
2.4 Data analysis.....	25
2.4.1 Data preparation and extraction .....	25
2.4.2 Modelling species distributions.....	26
2.4.3 Predicting floristic patterns .....	27
<b>3 Results and Discussion</b> .....	<b>29</b>
3.1 Average reflectance values of Landsat imagery are good predictors to model the distribution of trees at local extents....	29
3.2 Elevation improves distribution models of trees at local extents .....	30
3.3 Training species distribution models from elongated areas derive more consistent models and spatial predictions.....	31
3.4 Floristic patterns of trees are consistent at different taxonomic resolutions.....	32
3.5 Floristic dissimilarities of tree communities are strongly correlated with spectral distances.....	33
3.6 Bamboo forests are floristically different from adjacent non- bamboo forests .....	34

3.7	Floristic patterns of trees in Peruvian Amazonia are correlated with reflectance values derived from Landsat imagery and with other environmental layers .....	34
3.8	Future directions .....	35
<b>4</b>	<b>Conclusions .....</b>	<b>38</b>
<b>5</b>	<b>Acknowledgements .....</b>	<b>40</b>
	<b>References .....</b>	<b>42</b>
	<b>Original Publications.....</b>	<b>53</b>



# Abbreviations

ANOSIM	Analysis of similarities
AUC	Area under the curve
BIOCLIM	Bioclimatic variable
CEC	Cation exchange capacity
CHELSEA	Climatologies at high resolution for the earth's land surface areas
DBH	Diameter at breast height
DEM	Digital elevation model
ENM	Ecological niche modelling
ETM	Enhanced thematic mapper
ESA	European Space Agency
GDEM	Global digital elevation model
GPS	Global positioning system
HAND	Height above nearest drainage
HSM	Habitat suitability modelling
IBPP	Inventario de bosques de producción permanente
INFFS	Inventario nacional forestal y de fauna silvestre
MINAM	Peruvian ministry of the environment
MOL	La Molina National University herbarium
NDVI	Normalized difference vegetation index
NMDS	Non-metric multidimensional scaling
PCA	Principal component analysis
RF	Random forest
RMSE	Root-mean-square error
ROC	Receiving operative characteristic
SDM	Species distribution modelling
SERFOR	Servicio nacional forestal y de fauna silvestre
SRTM	Shuttle radar topographic mission
SVH	Spectral variability hypothesis
TM	Thematic mapper
USGS	United States Geological Survey

# List of Original Publications

This dissertation is based on the following original publications, which are referred to in the text by their Roman numerals:

- I Chaves, P.P.; Ruokolainen, K.; Tuomisto, H. Using remote sensing to model tree species distribution in Peruvian lowland Amazonia. *Biotropica* 2018, 50, 758–767, <http://doi.org/10.1111/btp.12597>
- II Chaves, P. P., Ruokolainen, K.; Van doninck, J. & Tuomisto, H. Assessing spatial variation: Modelling tree species distributions at local scales in Amazonia with remote sensing (manuscript)
- III Chaves, P. P., Reategui, E. N., Ruokolainen, K., Kalliola, R., Van doninck, J., Gómez Rivero, E., Zuquim, G. & Tuomisto, H. Using forestry inventories and satellite imagery to assess floristic variation in bamboo-dominated forests in Peruvian Amazonia. *Journal of Vegetation Science*, 2020, <https://doi.org/10.1111/jvs.12938>
- IV Chaves, P. P., Zuquim, G., Ruokolainen, K., Van doninck, J., Kalliola, R., Gómez Rivero, E., & Tuomisto, H. Mapping Floristic Patterns of Trees in Peruvian Amazonia Using Remote Sensing and Machine Learning. *Remote Sensing* 2020, 12(9), 1523. <https://doi.org/10.3390/rs12091523>

The original publications have been reproduced with the permission of the copyright holders.

## Author contributions

	I	II	III	IV
Original idea	PPC, KR, HT	PPC, HT, KR	PPC, KR, HT, NR	PPC, HT, KR
Data compilation	PPC	PPC	PPC	PPC
Data preparation	PPC, KR, HT	PPC, HT, KR	PPC, NR, KR	PPC, GZ
Analyses	PPC, KR, HT	PPC, KR, HT	PPC, KR, HT, NR	PPC, HT, GZ, KR
Writing	PPC, HT, KR	PPC, HT, K R, JVd	PPC, HT, KR, GZ, NR, JVd, RK, EG	PPC, HT, GZ, K R, JVd, RK, EG

Authors are listed in decreasing order of contribution for each paper.

Author abbreviations: EG=Elvira Gómez, GZ=Gabriela Zuquim, HT=Hanna Tuomisto, JVd=Jasper Van doninck, KR=Kalle Ruokolainen, NR=Natalia Reategui, PPC=Pablo Pérez Chaves, RK=Risto Kalliola.

# 1 Introduction

Amazonia, the largest and most diverse tropical forest in the world, continues facing several threats such as deforestation, habitat loss and fragmentation, forest degradation, climate change and forest fires. Amazonia experienced unprecedented forest fires in 2019 which led to hundreds of thousands of hectares of forest lost, mainly related to the expansion of the agricultural frontier and livestock farming. Recent reports have shown that, despite the current COVID-19 pandemic, forest fires are still on the rise in the region, especially in Brazilian Amazonia. At least 17% of Amazonia has already been deforested, while vast areas still remain botanically poorly known (Hopkins 2007; Schulman et al. 2007; Sousa-Baena et al. 2014). In the Peruvian Amazon the situation is slightly different. Even though deforestation rates have been slowly decreasing in the last couple of years and forest fires are not increasing when compared to Brazil, the annual average deforested area is still around 150,000 ha. The main drivers of deforestation in Peruvian Amazonia continue being small-scale agriculture and illegal gold mining (Zambrano et al. 2010; “MAAP #122” 2020; “MAAP #124” 2020).

Knowing where a species occurs, and the distribution of biologically unique areas, is key in conservation, natural resource management and land use planning, especially in a context where forest cover is continuously lost, and species go extinct on a yearly basis. Collecting biological information and characterizing forest resources are necessary steps for determining such biologically important areas, but they are time-consuming and expensive, especially in tropical and remote areas such as Amazonia (Köhl et al. 2006).

Fortunately, in Peruvian Amazonia there have been recent efforts, both by national forest institutions and forestry companies, in collecting systematic forest inventory and census data mainly in order to characterize the economic potential of their forests. Such field data covers large areas and is available even from remote areas, which makes them a potential source of information for ecological studies (Steege 1998; ter Steege et al. 2006). Such data often get used only for assessing the timber potential and once this has been fulfilled, the data remains stored within the owner’s organization, limiting the possibilities of using it for further scientific

studies. Here, I accessed such forestry data and explore its potential for ecological studies in Peruvian Amazonia.

To determine important areas for forest management and conservation purposes, not only field data is required but also spatially continuous predictions (maps) of important biodiversity patterns such as species distributions and floristic patterns. A powerful tool to derive such predictive maps over extensive areas is to combine the existing tree field data with available satellite images and environmental layers. Satellite images provide spatially and temporally continuous spectral information that have proved to be useful for predicting different aspects of biological diversity over extensive areas (Rocchini 2007; Saatchi et al. 2008; Saatchi et al. 2009; Rocchini et al. 2010; Baldeck & Asner 2013; Baldeck et al. 2014; van Ewijk et al. 2014; Turner 2014; He et al. 2015; Asner, Martin, Knapp, et al. 2017; Asner, Martin, Tupayachi, et al. 2017; Rocchini et al. 2018; Van doninck & Tuomisto 2018; Tuomisto et al. 2019). Similarly, evidence suggests that both edaphic and climatic variables are good predictors of tree species distributions and community composition in Amazonia (ter Steege et al. 2006; Toledo et al. 2011; Baldeck et al. 2013; Baldeck et al. 2016). Large number of environmental data layers, including climate (Fick & Hijmans 2017; Karger et al. 2017) and soils (Hengl et al. 2014; Hengl et al. 2017), are currently freely available, even though they can still contain accuracy problems (Soria-Auza et al. 2010; Moulatlet et al. 2017).

The combination of available forestry field data with freely available satellite imagery and environmental layers offers a window of opportunities to assess whether it is possible to use such layers to model the distribution of tree species at local extents and also to predict floristic patterns of trees at broader extents such as across the Peruvian Amazon. Having continuous spatial predictions of important tree species distributions and floristic patterns of trees over large areas would not only enable reducing research gaps and having predictions over areas where field data is not yet available but also would contribute in improving natural resources management and prioritizing conservation schemes in Peruvian Amazonia.

## 1.1 Usefulness of remote sensing in biodiversity studies

It is often the case that in tropical areas, such as in Amazonia, environmental data layers from ground measurements are either not available at all or suffer from coarse resolution and accuracy problems (Soria-Auza et al. 2010; Moulatlet et al. 2017). Remote sensing observations, such as those from Landsat satellites, provide spectrally diverse, spatially explicit and temporally frequent information that can be used for assessing forest biological and structural state (Rocchini et al. 2010; Turner 2014; He et al. 2015; Rocchini et al. 2016; Rocchini et al. 2018). Furthermore,

remotely sensed data have the great advantage compared to other available environmental layers, often based on interpolation techniques of sparse data, that they provide direct measurements and consistent data over vast areas.

Different kinds of surfaces such as soil, vegetation or even different species of plants, reflect the electromagnetic radiation in different ways (Jones & Vaughan 2010; Cavender-Bares et al. 2020). For instance, vegetation reflects a major proportion of incident radiation between the near-infrared and red spectrum (Jones & Vaughan 2010; Cavender-Bares et al. 2020) whereas other surfaces such as water bodies absorb most of the radiation in the same spectra. Even further, different plant species and vegetation types also reflect radiation differently (Cavender-Bares et al. 2020). For instance, grasses, such as bamboo, reflect consistently more radiation in the near-infrared spectrum than trees (Carvalho et al. 2013). These reflectance values captured by satellite images have been useful for predicting different aspects of biological diversity (Rocchini 2007; Rocchini 2007; Saatchi et al. 2008; Saatchi et al. 2009; Rocchini et al. 2010; Baldeck & Asner 2013; Baldeck et al. 2014; van Ewijk et al. 2014; Turner 2014; He et al. 2015; Asner, Martin, Knapp, et al. 2017; Asner, Martin, Tupayachi, et al. 2017; Rocchini et al. 2018; Van doninck & Tuomisto 2018; Tuomisto et al. 2019). For instance, remote sensing has been used for predicting local species richness (Rocchini et al. 2005; Rocchini et al. 2007; Rocchini 2007; Rocchini et al. 2010; Schmidtlein & Fassnacht 2017), variation in species composition (Tuomisto, Poulsen, et al. 2003; Thessler et al. 2005; He et al. 2009; Rocchini et al. 2009; Baldeck & Asner 2013; Baldeck et al. 2014; Leitão et al. 2015; Rocchini et al. 2018; Tuomisto et al. 2019; Maximiano et al. 2020), functional diversity (Asner et al. 2009; Asner & Martin 2009; Féret & Asner 2014; Asner et al. 2014; Asner et al. 2015; Asner, Martin, Knapp, et al. 2017; Asner, Martin, Tupayachi, et al. 2017) and spectral diversity (Féret & Asner 2014; Laliberté et al. 2020) in different regions.

Many of the uses of remote sensing in biodiversity studies build upon the spectral variability hypothesis (SVH) and the spectral decay model. The SVH states that the variability in the spectral signal from remote sensing data is correlated with environmental heterogeneity and could be used as a proxy for species diversity (Palmer et al. 2002). Several studies have supported this hypothesis (Rocchini et al. 2007; Rocchini 2007; Nagendra et al. 2010) even though it is sensitive to the characteristics of the remote sensing product (Rocchini et al. 2007; Nagendra & Rocchini 2008; Rocchini et al. 2016) and varies across spatial scales (Schmidtlein & Fassnacht 2017). On the other hand, the spectral decay model states that a pair of locations that are spectrally dissimilar are expected to share less species (be floristically more dissimilar), than a pair of locations that are spectrally similar. If environmental properties affect both canopy reflectance and species composition, then differences in reflectance can be used as indicators of compositional dissimilarities. This idea has been also been supported by different studies



(Tuomisto, Poulsen, et al. 2003; Rocchini et al. 2005; Rocchini & Cade 2008; Rocchini et al. 2009).

There are several key considerations when using remote sensing in biodiversity studies. If the spatial resolution (the dimension of the minimum unit) is too high with respect of the target object (e.g: canopy trees), the spectral variability is expected to increase and lead to noise (Nagendra 2001; Nagendra & Rocchini 2008; Rocchini et al. 2010). Therefore, the spatial resolution of the satellite image should match the size of the target object. In addition to this, the higher the spectral resolution (the number of bands and the range of electromagnetic radiation at which they acquire information), the higher the power to discriminate among objects that reflect the radiation in a different manner. Therefore, more spectral bands would provide more accurate predictions (Asner & Martin 2009; Wang et al. 2018). Another important factor is the geographic extent, since a larger window or area will inevitably have higher spectral and environmental heterogeneity and, hence, is expected to have higher species richness (Rocchini et al. 2004). Some authors have found higher explanatory power of spectral variation when increasing the spatial extent (Rocchini et al. 2007), whereas others found higher correlations at smaller extents (Schmidtlein & Fassnacht 2017). Here, I explore the potential of Landsat imagery, which obtains reflectance values for 6 different sections within the optical portion of the electromagnetic spectrum at 30-m spatial resolution, across different geographic extents in Peruvian Amazonia.

In Amazonian forests, reflectance values derived from Landsat satellite images have been useful for identifying land cover classes and vegetation types (Tuomisto et al. 1995; MINAM 2016b; MINAM 2019) and for predicting soil patterns (Sirén et al. 2013; Van doninck & Tuomisto 2018). It has also been particularly observed that differences in canopy reflectance can predict spatial patterns of species composition of understory plants (Tuomisto, Poulsen, et al. 2003; Tuomisto, Ruokolainen, et al. 2003; Salovaara et al. 2005; Higgins et al. 2011; Higgins et al. 2012; Muro et al. 2016; Tuomisto et al. 2019) and are also useful complementary predictors for modeling the distribution of tree species and of understory plants at broad extents (Prates-Clark et al. 2008; Saatchi et al. 2008; Cord et al. 2013; Van doninck, Jones, et al. 2020). Elevation, another remote sensing variable, has also been related to geological substrates and soil nutrients in Amazonia (Vormisto et al. 2004; Costa et al. 2005; Higgins et al. 2011) and other topographic variables derived from elevation, such as the vertical height above local drainage (HAND), have also been found to structure tree species distributions (Fortunel et al. 2018; Zuleta et al. 2018).

It still remains to be assessed whether canopy reflectance, derived from Landsat imagery, can be used to model the distribution of tree species at smaller geographic extents (**I** and **II**) and whether the reflectance values are also related to the floristic

composition of trees (III and IV) at broader extents such as across Peruvian Amazonia. The bulk of existing knowledge suggests that both reflectance values derived from Landsat imagery and elevation are ecologically informative remote sensing layers (Bradley et al. 2012; He et al. 2015; Leitão & Santos 2019) which can be potentially used also for modelling the distribution of canopy trees and predicting floristic patterns of trees at different geographic extents in Peruvian Amazonia.

## 1.2 Using forest inventories and forest census data for ecological studies

In Peruvian Amazonia, forests are divided into permanent production forests, local forests, reserve forests, protection forests, forests on indigenous and native communities' lands, and private forests according to their land-use potential ("Ley Forestal y de Fauna Silvestre"). The Peruvian National Forest and Wildlife Service (SERFOR) promotes the sustainable management of the Peruvian forest resources and aims at characterizing such potential. Two main national forest inventory programs provide systematic and quantitative information on forest resources, such as timber. One of these programs takes place within permanent production forests, which can be granted as forest concessions for timber or non-timber production. The other inventory program aims at characterizing timber, non-timber, wildlife and carbon stocks in all forest types throughout the country. Both the permanent production forest inventory ("Inventario Forestal en Bosques de Producción Permanente", IBPP) and the national forest and wildlife inventory ("Inventario Nacional Forestal y de Fauna Silvestre", INFFS) offer a great potential not only for assessing forest resources but also for ecological studies.

Within permanent production forests, concessions can be granted for timber production. The forest concession is often divided into annual management areas. Forest concessionaries are required to make a forest census at each annual management area before initiating any logging or harvesting activities. In those forest censuses, all commercial trees above 30 cm of diameter are tallied within the management area, and their local name, height, diameter and coordinates are recorded. Such forestry census data can cover large areas with high-resolution occurrence data of trees, which offer a unique opportunity to train species distribution models (SDM) at small extents in combination with Landsat imagery and other environmental layers of higher spatial resolution.

Both forest inventory and census data are often used for assessing the forest timber potential and once such objectives are fulfilled, the data remain stored within the owner's organization. Unfortunately, this limits the possibilities of exploring the use of such valuable forest field data for other relevant ecological studies. Sharing those datasets and making them accessible is a key step in increasing research and

improving the forest sector. Here, I requested access to the forest inventory data from SERFOR and to forest census data from Consolidado Otorongo forest concession in Southern Peruvian Amazonia. Fortunately, both institutions agreed on sharing such valuable field data for scientific purposes.

A clear advantage of forest inventory and census data is that they can cover extensive areas and be available for remote areas. This indeed makes them a potential source of information for ecological studies (Steege 1998; ter Steege et al. 2006). Nevertheless, there are also associated disadvantages that are important to identify, acknowledge and if possible, address. One of these is taxonomical, since forest census data typically focus on large trees and do not document species identifications with herbarium vouchers. Even though some forest inventories might document species identifications with herbarium vouchers, a lot of taxa remain unidentified due to the large number of taxa, morphological similarities and practical difficulties in reaching tree crowns (Köhl et al. 2006) in tropical areas like Peruvian Amazonia. Another challenge of such forest inventory and census data is that they might rely on vernacular names, which are not necessarily consistent with scientific tree species (Guitet et al. 2014) particularly in species-rich tropical forests. Using this data requires hence identifying, acknowledging and addressing possible taxonomic issues when analyzing and drawing conclusions for ecological studies.

Even though there are intrinsic taxonomic issues associated with forest inventory and census data, they still offer a great research potential. In this dissertation I explore such potential for characterizing biological patterns in a still poorly known region as Peruvian Amazonia. I use the forest census data together with remote sensing data to explore whether it is possible to model the distribution of tree taxa at local extents (I and II). I further explore the forest inventory data to investigate whether it is possible to predict floristic patterns of trees using remote sensing and environmental layers (III and IV).

### 1.3 Species distribution models of trees at local extents

Species distribution models (SDM) characterize the environmental conditions that are suitable for a species and then identify where those suitable environments are distributed in the geographical space (Guisan & Zimmermann 2000; Guisan & Thuiller 2005; Franklin & Miller 2010; Peterson et al. 2011; Guisan et al. 2017). When describing the species niche, SDMs have also been referred to as ecological niche models (ENM) and when describing the habitat suitability, they have been referred to as habitat suitability models (HSM) (Soberón & Nakamura 2009; Peterson et al. 2011; Peterson & Soberón 2012; Guisan et al. 2017). SDMs are used

to predict the occurrence of species in areas where field data are unavailable and to deliver predictive maps of species' distribution (Franklin & Miller 2010). This has many practical applications (Franklin & Miller 2010) such as reserve design and conservation planning (Rodríguez-Soto et al. 2011), natural resources management (I), risks and impacts of invasive species (Barbet-Massin et al. 2018; Chapman et al. 2019) and the effects of climate change on the distribution of species (Raghavan et al. 2019).

Different environmental factors, both abiotic and biotic, shape the distribution of species (Franklin & Miller 2010; Peterson et al. 2011). In fact, the term niche has been defined as the multidimensional environmental space that contains suitable conditions for a species to persist (Hutchinson 1957), and such conditions include both abiotic and biotic factors or also referred as non-linked and linked, or indirect and direct factors (Hutchinson 1978; Austin & Smith 1989; Soberón & Nakamura 2009; Peterson et al. 2011). These factors determining the distribution of a species are scale-dependent (Willis & Whittaker 2002; Pearson & Dawson 2003). Biotic factors, such as species interactions, often act at finer extents whereas abiotic factors, such as climate, often act at broader extents as a first environmental filter (Willis & Whittaker 2002). Accessibility or barriers of movement are also another limiting factor of a species' distribution (Soberón & Nakamura 2009; Barve et al. 2011).

In the past decade there has been several advances in summarizing and clarifying the terminology of SDMs (Guisan & Zimmermann 2000; Soberón & Nakamura 2009; Franklin & Miller 2010; Peterson et al. 2011; Peterson & Soberón 2012; Guisan et al. 2017), increasing the availability of both open source species occurrences and environmental layers, the development of modelling techniques (Barbosa & Schneck 2015) as well as of standards and guidelines for delivering (Araújo et al. 2019), reproducing (Feng et al. 2019) and reporting SDMs (Zurell et al.). Typically, environmental layers are often freely available at coarse spatial resolution, such as climatic (Hijmans et al. 2005; Fick & Hijmans 2017; Karger et al. 2017) and soil variables (Hengl et al. 2014; Hengl et al. 2017), whereas finer-scale environmental layers are still lacking. Therefore, most of the modelling exercises have been done at global to regional extents (Guisan & Thuiller 2005; Franklin & Miller 2010). Even though there has been some research in modeling the distribution of species at smaller extents (Thuiller et al. 2003; Franklin & Miller 2010; Oke & Thompson 2015; Barbosa & Schneck 2015; He et al. 2015; Chaves et al. 2018), there are still advances to make in this research line. Remote sensing data are among the potentially most interesting sources of information that could be used for modelling the distribution of species at local extents with higher resolution layers (He et al. 2015).

In Amazonian forests, the distributions of trees have been mainly modelled at regional to continental extents using climate only (OSINFOR 2013; OSINFOR

2016) and both climate and remote sensing layers (Prates-Clark et al. 2008; Saatchi et al. 2008; Cord et al. 2013). Remote sensing variables have been suggested to be useful complements of other environmental variables to model the distribution of species at broad extents (Prates-Clark et al. 2008; Saatchi et al. 2008; Bradley & Fleishman 2008; Bradley et al. 2012; Cord et al. 2013; He et al. 2015; Van doninck, Jones, et al. 2020), but not using remote sensing variables only. Since in Amazonian forests, reflectance values derived from Landsat imagery have been related to floristic variation and soil patterns (Tuomisto, Poulsen, et al. 2003; Tuomisto, Ruokolainen, et al. 2003; Sirén et al. 2013; Van doninck & Tuomisto 2018), remote sensing could provide ecologically informative variables for modelling the distribution of tree species at smaller extents. As far as investigated, only one study has modelled the distribution of trees in Amazonia at more local extents using different topographic variables and one vegetation index derived from remote sensing variables (Figueiredo et al. 2015). Nevertheless, the use of remote sensing variables only, such as the reflectance values derived from Landsat imagery, has not been explored yet. Here, I explore this potential by combining tree occurrence data from forest census data with medium-resolution Landsat satellite imagery and elevation (**I** and **II**). Modelling the distribution of tree taxa at local extents, made possible thanks to remote sensing data, could facilitate forest management practices and operative decisions both in natural resources management and conservation purposes.

## 1.4 Floristic patterns of trees in Amazonia

Some studies have suggested that soil fertility is the major gradient in shaping tree community composition followed by the dry season length (ter Steege et al. 2006) in Amazonia, whereas other studies have found that tree floristic variation is mainly associated with climatic gradients in comparison with edaphic variation (Toledo et al. 2011; Toledo et al. 2012). Even though different studies might cover different geographic extents and hence, different ranges of environmental gradients, they agree that edaphic, topographic and climatic variables are good predictors of tree species distributions and community composition of trees at different extents in Amazonia (Phillips et al. 2003; ter Steege et al. 2006; Ruokolainen et al. 2007; Toledo et al. 2011; Toledo et al. 2012; Baldeck et al. 2013; Baldeck et al. 2016). Specifically, in Peruvian Amazonia, it has been found that soils, topography and climate all contributed to explaining variation in tree community composition (Baldeck et al. 2016) at regional extents, whereas at landscape level, soil differences were more important in explaining floristic patterns (Phillips et al. 2003; Ruokolainen et al. 2007; Baldeck et al. 2016).



Floristic patterns of other plant groups in Amazonia have also been related to edaphic and climatic variables, such as for ferns and Melatomataceae (Tuomisto, Poulsen, et al. 2003; Tuomisto, Ruokolainen, et al. 2003; Higgins et al. 2011; Zuquim et al. 2012; Zuquim et al. 2014; Tuomisto et al. 2019) , Zingiberales (Figueiredo et al. 2018) and palms (Vormisto et al. 2004; Kristiansen et al. 2012; Cámara-Leret et al. 2017). Additionally, the compositional turnover of trees in Amazonia has been found to be strongly correlated with understory plant groups (Tuomisto et al. 1995; Ruokolainen et al. 2007; Higgins et al. 2011; Tuomisto et al. 2016). This reaffirms that soils and climate variables are consistently related to floristic patterns in Amazonia. Accordingly, it is reasonable to expect that it is possible to predict the floristic variation of trees in Amazonia on the basis of environmental variables.

Field data on environmental gradients, such as soil properties, would allow a more accurate assessment of relationships between environmental patterns and floristic patterns of trees. Unfortunately, the forest inventory field data accessed for this dissertation only includes tree data but not direct measurements of environmental variables. Fortunately, many environmental data layers, including climate and soils (Fick & Hijmans 2017; Hengl et al. 2017; Karger et al. 2017), are already freely available and are a potential source of information for predicting floristic patterns of trees in Peruvian Amazonia even though some might still suffer from accuracy problems (Moulatlet et al. 2017). Furthermore, it is already known that average canopy reflectance, derived from Landsat imagery, is related to soil patterns in Amazonia (Higgins et al. 2011; Sirén et al. 2013; Van doninck & Tuomisto 2018).

Even though it is known that the floristic patterns of trees are related to soils and climate and that satellite imagery is related to soils in Amazonia, so far, no research has derived spatial predictions of floristic variation of trees in Amazonia, meaning predicted maps of floristic patterns. Knowing the main drivers of floristic variation of trees in Peruvian Amazonia and identifying the availability of potential environmental and remote sensing predictors are important steps when aiming at spatially predicting those floristic patterns over vast areas. Having spatial predictions of tree floristic patterns maps would be useful and required for assessing conservation priorities or natural resources management at national level. Here, I use available remote sensing and environmental layers to assess whether it is possible to predict floristic patterns of trees derived from national forest inventory data (**III** and **IV**).

## 1.5 Aims of the thesis

The present dissertation aims at utilizing available existing forest inventory and census data in Peruvian Amazonia together with freely available Landsat imagery and environmental layers to model the distribution of important tree species at local extent and to predict floristic patterns of trees in broader extents, ultimately throughout all Peruvian Amazonia.

In **I**, the main questions were whether it is possible to use the raw digital numbers of Landsat TM bands to derive SDMs, whether the use of averaging moving filters improve SDM performance and, finally, whether NDVI and elevation further improve the distribution models. For this, I modelled the distribution of tree taxa at a local extent (6,000 km<sup>2</sup>) in Peruvian Southern Amazonia combining tree occurrence data from forest census data and remote sensing data as predictor variables.

In **II**, the main aim was to investigate and quantify to what degree SDMs differ when using different spatial configurations of the tree occurrence data and how such spatial configurations can improve the applicability of SDMs for tropical forest management at local extent (6,000 km<sup>2</sup>). For this, I also modelled the distribution of tree taxa at local extent following the framework suggested in **I**, but with five times more tree occurrence data which enabled building different spatial configurations of the training data.

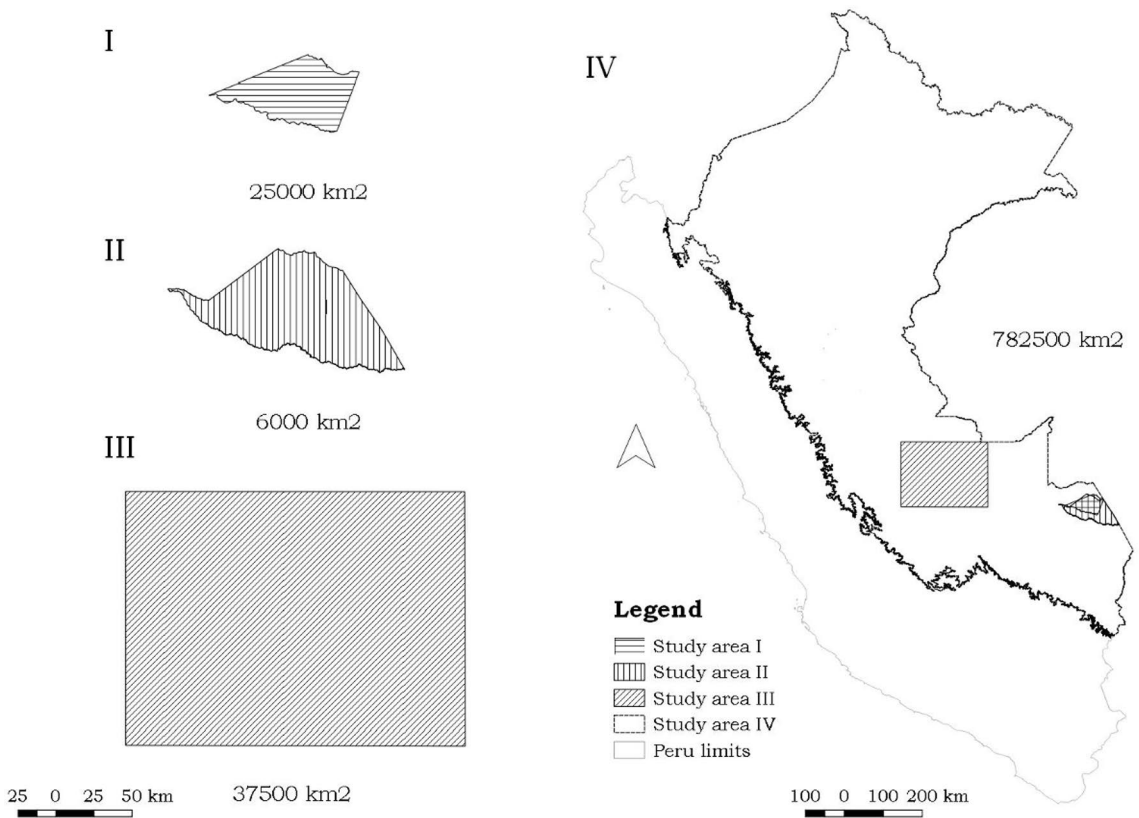
In **III**, the main aims were to assess whether the floristic composition of trees differ between bamboo-dominated forest and adjacent non-bamboo forests, whether the floristic patterns of trees can be predicted using Landsat data, and whether the results are sensitive to different taxonomic resolution and tree size classes. For this, I used regional forest inventory data (IBPP) and Landsat-derived remote sensing data and elevation.

In **IV**, the aims were to assess how floristic patterns of trees relate to canopy reflectance and environmental layers, to map floristic variation of trees across the Peruvian Amazon (~800,000 km<sup>2</sup>), and to derive a classification of floristic variation. For this, I used genus-level tree data from the Peruvian national forest inventory (INFSS), Landsat imagery and environmental layers (climate, soil and topography) to predict floristic patterns of trees throughout Peruvian Amazonia using random forest regression.

## 2 Methods

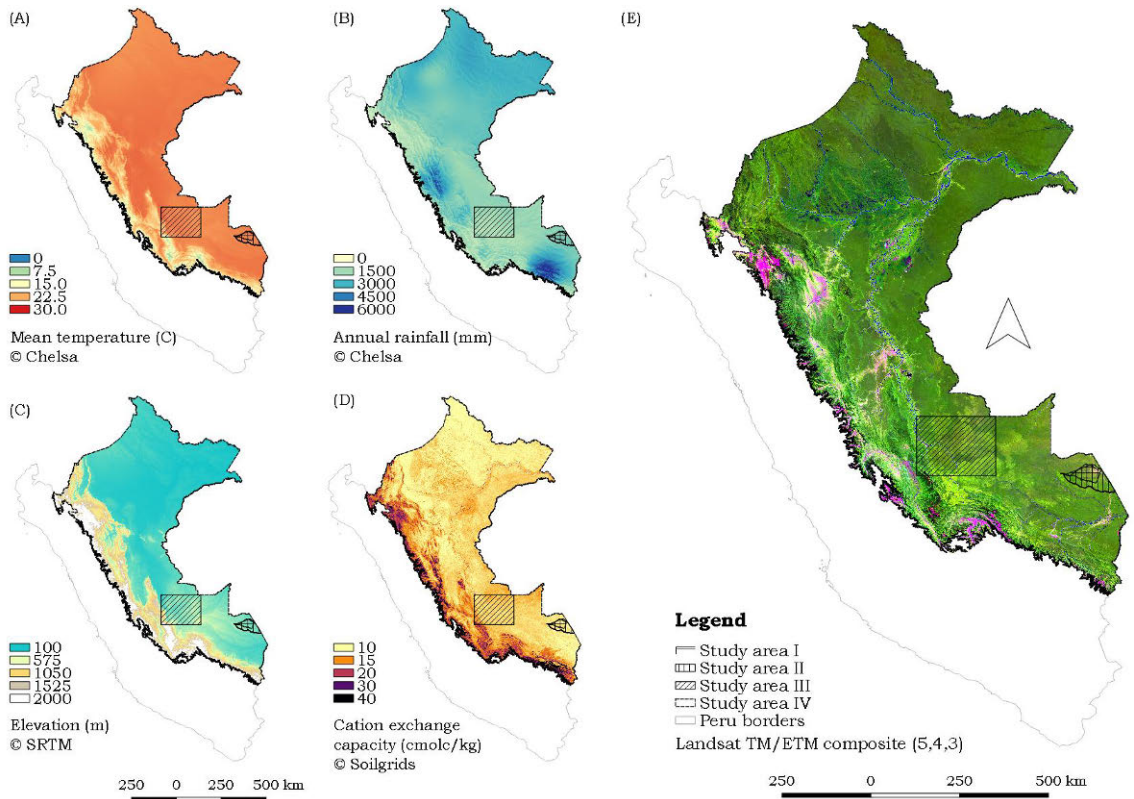
### 2.1 Study area

The study area in this dissertation comprehends different parts of Peruvian Amazonia (Figure 1) as defined by the Peruvian Ministry of Environment (MINAM 2016a). The study areas of the first and second chapters (**I** and **II**) are located in Madre de Dios region (Southern Peruvian Amazonia) covering approximately 2500 and 6000 km<sup>2</sup>, respectively, in mainly lowland forests. The third chapter (**III**) of this dissertation focuses on an area in Ucayali region (Central Peruvian Amazonia) covering almost 40,000 km<sup>2</sup>. Finally, the study area of the last chapter (**IV**) covers the whole Peruvian Amazonia, an area of more than 780,000 km<sup>2</sup>, including lowland forests, seasonally flooded areas and montane forests. The focus in **I** and **II** is to model the distribution of tree taxa at local extents whereas in **III** and **IV**, to predict floristic patterns at regional to national extent.



**Figure 1.** Scheme of the geographic extent of the study areas for each chapter throughout Peruvian Amazonia.

Throughout Peruvian Amazonia, the climate is mainly tropical and humid with a monthly average temperature of 24 °C (ranging between 15°C on the Andean slopes and 26 °C in the lowland forests) and an average annual rainfall of 2300 mm (ranging from 700 to more than 7000 mm) as extracted from Climatologies at High Resolution for the Earth’s Land Surface Areas CHELSA (Karger et al. 2017) (Figure 2A,B). The elevation ranges between 100 and 3000 m above sea level (Figure 2C) and the soils tend to richer closer to the Andes (Figure 2D) but in **IV** we limit the analyses up to 2000 meters of elevation



**Figure 2.** Geographical characteristics of the study area in Peruvian Amazonia (A) Mean annual temperature, (B) annual precipitation, (D) elevation, and (C) cation exchange capacity (CEC) in the study area. (D) Landsat TM/ETM+ composite (bands 5, 4, and 3 assigned to red, green, and blue, respectively) of Peruvian Amazonia.

## 2.2 Forest inventory and census data

The field data used in this dissertation came from a Peruvian national forest institution and a forest concession company. In chapters **I** and **II**, I used forest census data provided from Consolidado Otorongo timber forest concession in Peruvian Southern Amazonia. In chapter **III**, I used data from regional forest inventories (IBPP) whereas in chapter **IV**, I used data from the Peruvian national forest and wildlife inventory (INFFS). Both forest inventory datasets were provided by the Peruvian Forest and Wildlife Service (SERFOR). Below I describe the field data used on each of the chapters.



**Table 1.** Characteristics of the field datasets used in each study

CHAPTER	TYPE	NUMBER OF PLOTS / AREA (HA)	DATABASE	REGION	BOTANICAL GROUPS
I	Forest census	2500	Otorongo	Madre de Dios	Trees
II	Forest census	15,000	Otorongo	Madre de Dios	Trees
III	Forest inventory	25	IBPP	Ucayali	Trees and palms
IV	Forest inventory	157	INFFS	Peruvian Amazonia	Trees, tree ferns and palms

### 2.2.1 Forest census data

The forest census data used in **I** and **II** were produced by Consolidado Otorongo forest concession (Otorongo dataset) within their forest management areas. In **I**, I used forest census data from approximately 2500 ha whereas in **II**, more forest census data became available, including approximately 15,000 ha of the concession's management areas. Within those areas, all commercial trees (for timber) with diameter at breast height (dbh) at least 30 cm were registered along linear transects and their local name, height, dbh and geographical location were recorded. The beginning and end of each evaluation transect were georeferenced using GPS devices. Individual trees were georeferenced in the field using a relative system of X and Y coordinates. Geographical coordinates were later assigned to each tree by combining their field measured within-transect locations with the transect GPS coordinates.

The original forest census database consisted of approximately 30 tree taxa, of which only a few were selected based on two criteria: abundance and taxonomic consistency. The primary selection criterion was that the tree was abundant enough for robust species distribution modeling, and hence the number of presence records should be above the recommended minimum of 20–50 observations (Guisan et al. 2007; Merow et al. 2014; van Proosdij et al. 2016). The second selection criterion was that the taxonomic identification was considered reliable at the genus level. Identification of tropical trees to biological species is difficult (Köhl et al. 2006), especially without voucher specimens. Since the forest census was performed for commercial rather than scientific purposes and voucher specimens were not collected, the species identifications may not be entirely accurate. All the selected genera contain only one species within the forest census dataset, and external sources indicate that they have few species in the area in general. Nevertheless, I will be referring to them with their generic names and using the term 'taxon' instead of 'species'.

## 2.2.2 Forest inventory data

In **III**, I used regional forestry inventory data collected in February and March 2016 in 25 1-ha inventory plots by the Peruvian Forest and Wildlife Service (SERFOR). The main objective of the regional forestry inventory was to characterize the timber potential within permanent production forests (IBPP) which are delimited by the Peruvian government for timber production. The central point of each plot was georeferenced using a GPS device. Each cross-shaped plot (1 ha) consisted of four rectangular subunits of 0.25 ha. Each subunit was divided into 10 registration units (25 m x 10 m) for a total of 40 registration units per plot. Trees and palms with diameter at breast height (dbh) of at least 30 cm (“canopy trees”) were recorded in the entire plot (1 ha). In addition, trees with dbh between 10 cm and 30 cm (“pole trees”) were recorded in the two central registration units of each sub-unit, totaling 0.2 ha per plot. For each tree, its dbh (cm), total height (m), species and local name was recorded in the field. A complete set of vouchers is deposited in the herbarium of La Molina National Agrarian University (MOL) but their identifications have not been cross-checked.

In **IV**, I used part of the national forest and wildlife inventory (INFFS) collected between 2013 and 2018 in 157 sampling plots distributed across Peruvian Amazonia, which was stratified for sampling into lowland, hydromorphic (or also referred to as seasonally flooded) and montane forests (up to 2000 meters of elevation). The inventory plots were L-shaped and consisted of either 10 0.05-ha subunits totaling 0.5 ha (hydromorphic or mountain forests) or seven 0.1-ha subunits totaling 0.7 ha (lowland forests). The corner of each L-shaped plot was georeferenced using a GPS device. Trees (including palms and tree ferns) exceeding 30 cm in diameter at breast height (dbh) were recorded in all subunits of each plot. In addition, trees between 10 and 30 cm dbh were recorded in half of the subunits per plot. For each tree, its dbh (cm), height (m), scientific name, and local name were recorded in the field. The voucher specimens were deposited in La Molina National Agrarian University Herbarium (MOL), but voucher specimens were not cross-checked among botanists. Therefore, we used only genus-level identifications in the analyses even when species-level identifications were available in the inventory database.

## 2.3 Predictor layers

### 2.3.1 Landsat satellite imagery

Different Landsat satellite image products were used in all the chapters of this dissertation. The Landsat products consisted of six bands that correspond to different sections in the optical portion of the electromagnetic spectrum: 1 (blue), 2 (green), 3

(red), 4 (near infrared), 5, and 7 (the latter two shortwave infrared). In chapter **I**, a single Landsat 5 TM scene from 2011 was used since it covered the whole study area. The raw digital numbers (DN) of bands 3-7 were used as predictors in **I**. In chapters **II** and **IV**, the available Amazon Landsat TM/ETM+ composite (Van doninck & Tuomisto 2018) was used and cropped to each study area. The composite was based on all Landsat acquisitions from the dry season months of the 10-year period 2000–2009 (Van doninck & Tuomisto 2018). In chapter **III**, another Landsat ETM+ composite was produced over the study area by combining acquisitions of more than 1000 images of six Landsat scenes from a 3-year period (2014-2016). When producing both composites for **II**, **III** and **IV**, the Landsat images were transformed to surface reflectance and corrected for atmospheric and directional effects (Van doninck & Tuomisto 2015; Van doninck & Tuomisto 2017a). They were composited using the medoid method, which ensures that the reflectance values represent relatively stable ground cover characteristics that can be expected to be relevant for canopy trees (Van doninck & Tuomisto 2017b; Van doninck & Tuomisto 2018).

In chapters **I** and **II**, the values of the Landsat bands 3, 4, 5 and 7 were used as predictor variables. A normalized difference vegetation index (NDVI), calculated as  $NDVI = (band\ 4 + band\ 3) / (band\ 4 + band\ 3)$ , was used as an additional predictor. Reflectance values from each of the Landsat bands and NDVI were acquired at a 30-m spatial resolution. In chapters **III** and **IV** all six Landsat bands were used as predictors. In chapter **III**, a principal component analysis (PCA, based on a correlation matrix) of all the six Landsat bands (Landsat-PCA) was performed and the first two axes of the PCA were used as additional variables together with a NDVI layer.

In chapter **III**, an additional Landsat-based predictor was the percentage of pixels classified as bamboo within a 275-m buffer from the central coordinates of each inventory plot. To do so, bamboo and non-bamboo forests were mapped based on an time series analysis of yearly Landsat TM/ETM+ composite images from 1984 to 2018 over the study area (Van doninck, Westerholm, et al. 2020). Reflectance values of Landsat bands, NDVI, the Landsat-PCA layers and the bamboo percentage are collectively referred to as “Landsat layers” hereafter.

### 2.3.2 Environmental layers

In all the chapters, elevation was used as an environmental predictor. In **I**, **II** and **III**, a digital elevation model (DEM) was used and obtained from ASTER GDEM (Aster Global Digital Elevation Model), whereas in **IV**, it was obtained from the SRTM digital elevation model (Shuttle Radar Topography Mission). I acknowledge that ASTER GDEM has errors in tropical areas and therefore in the last chapter I decided

to use another digital elevation model source. All elevation layers were obtained at 30-m spatial resolution. Soil cation exchange capacity (CEC) at a 0.05-m depth was obtained from SoilGrids (Hengl et al. 2017) at a 250-m spatial resolution. Climatic data (19 bioclimatic variables) at approximately a 1-km spatial resolution were obtained from CHELSA (Karger et al. 2017). Elevation, soil, and climatic data are hereafter collectively referred to as “environmental layers”.

**Table 2.** Summary of the response and predictor variables and the methods used in each chapter.

CHAPTER	RESPONSE VARIABLES	PREDICTOR VARIABLES		METHOD
		LANDSAT	ENVIRONMENTAL	
I	Occurrence of tree taxa	Raw digital numbers of Landsat 5 TM image (bands 3, 4, 5 and 7) and NDVI	Elevation (DEM)	Maximum entropy (Maxent) algorithm
II	Occurrence of tree taxa	Reflectance values <sup>1</sup> of a 10-year Landsat composite TM/ETM+ (bands 3, 4, 5 and 7) and NDVI	Elevation (DEM)	Maximum entropy (Maxent) algorithm
III	Floristic dissimilarities	Reflectance values <sup>1</sup> of a 3-year Landsat composite ETM+ (bands 1, 2, 3, 4, 5 and 7), Landsat-PCA, NDVI and Bamboo (%)	Elevation (DEM)	Mantel and partial mantel test
IV	(a) Floristic dissimilarities (b) Floristic ordination axes (NMDS 1, 2 and 3)	Reflectance values <sup>1</sup> of a 10-year Landsat composite TM/ETM+ (bands 1, 2, 3, 4, 5 and 7)	Soil cation exchange capacity (CEC), 19 bioclimatic layers (BIOCLIM) and elevation (DEM)	(a) Mantel and partial mantel test (b) Linear regression analysis and random forest regression

<sup>1</sup> Surface reflectance and corrected for atmospheric and directional effects

## 2.4 Data analysis

### 2.4.1 Data preparation and extraction

In chapter I, averaging filters based on a 3x3 and 5x5 pixel window (90 x 90 and 150 x 150 m, respectively) were applied to the Landsat layers in order to reduce noise. Three sets of Landsat layers were used as predictors: with no filter, a 3x3 filter and

a 5x5 filter. In chapter **II**, only a 5x5 filter was applied to the Landsat layers. No moving averaging filters were applied to the Landsat layers in chapters **III** and **IV**.

In chapters **I** and **II**, the original 30-m spatial resolution Landsat and elevation layers were used for model prediction. In chapters **III** and **IV** the original 30-m resolution Landsat and elevation layers were used to extract the median values within a buffer around the central coordinates of each inventory plot, relative to each plot's dimensions. Buffers of 275 and 450 m were used for chapters **III** and **IV**, respectively in relation to the inventory plots dimensions. In chapter **IV**, climate and soil data were obtained for each inventory plot by extracting the required predictor value from the pixel corresponding to the inventory plot coordinates. For model prediction, in chapter **IV** all predictor layers (Landsat and environmental) were rescaled to 450-m spatial resolution since it corresponds to the inventory plot dimensions and previous studies found that spatial predictions at such spatial resolution in Amazonia are congruent with predictions at finer resolutions (Tuomisto et al. 2019) and computationally more efficient.

## 2.4.2 Modelling species distributions

In chapters **I** and **II**, the location of tree taxa from the forest census dataset was used together with Landsat and elevation layers to model their distribution within the study areas. In both chapters, the occurrence data was divided into training and test datasets that were spatially separated but covered similar ranges of the remote sensing layers. The census dataset had information about the presence of just a few tree taxa and only individuals of more than 30 cm of diameter, hence the use of presence-absence algorithms was not suitable as the modelling framework. Of the other possible available modelling algorithms, I chose the complementary log-log (cloglog) link function (Phillips et al. 2017) of MaxEnt algorithm, which considers presence-only data and background information to model species distributions (Phillips et al. 2017). A previous exercise that modelled the distribution of trees in Amazonia at local extents using remote sensing variables utilized MaxEnt (Figueiredo et al. 2015) and therefore, I opted the same framework for comparison purposes. Furthermore, MaxEnt has performed equally well or better than other modelling algorithms (Phillips et al. 2006; Elith et al. 2006; Hernandez et al. 2006; Wisz et al. 2008; Giovanelli et al. 2010; Merckx et al. 2011; Aguirre-Gutiérrez et al. 2013; Merow Cory et al. 2014; Villar Hernández & Pérez Elizalde 2015), deriving more consistent predictions across different calibration areas (Giovanelli et al. 2010) and being less sensitive to configuration settings (Hallgren et al. 2019). We used the same features for all the models to facilitate comparisons of model predictions among the different tree taxa and models sets.



In **II**, the training area used in the modelling procedure was divided in different spatial configurations of similar area but different shape. This was done in order to assess how much of the available environmental space from the study area they covered and how related it was with the model performance.

### 2.4.3 Predicting floristic patterns

Both in chapters **III** and **IV**, I investigated the relation between floristic patterns of trees at the community level with Landsat and environmental layers. In chapter **III**, floristic patterns were assessed at different tree size classes (pole trees - dbh 10-30 cm, canopy trees – dbh  $\geq$  30 cm, and all trees together – dbh  $\geq$  10 cm) and different taxonomic cleaning strategies (genus, identified species and consistent species). Lumping all trees to the genus-level allows including a higher proportion of the inventory database records at the expense of losing part of the ecological signal of species-environment relationships. Another common strategy is to only include those tree individuals which have been identified to species level (identified species). This enables maintaining the ecological signal but losing an important proportion of the database. A further step is to include only records that are consistent among brigades (consistent species). In chapter **IV**, floristic patterns were assessed using all trees (dbh  $\geq$  10 cm) at the genus level only.

An ordination (non-metric multidimensional scaling or NMDS) was used to summarize the main floristic variation among the forest inventory plots into two axes (NMDS 1 and 2) in **III** and three axes (NMDS 1-3) of variation in **IV**. In both chapters, the Sørensen dissimilarity index was used which considers only presence-absence data of the taxa. The extended (step-across) version of the index was used, because it provides ecologically realistic dissimilarity values between plots that share no genera (De'ath 1999; Tuomisto et al. 2012; Zuquim et al. 2012). In chapter **III**, the NMDS results were used to derive convex hulls and visualize the floristic dissimilarity patterns between bamboo and non-bamboo forests, and to determine differences in floristic composition between the two forest types using an analysis of similarity (ANOSIM). In chapter **IV**, the NMDS results were used to visualize floristic dissimilarities and to assess their relationship with Landsat and environmental variables through a linear regression analysis and a random forest regression.

In **III** and **IV**, Landsat and environmental distance matrices were obtained using pairwise Euclidean distances between the inventory plots, calculated for each variable separately. Geographical distances between inventory plots were calculated using their coordinates and the values were transformed to their natural logarithm before analysis. Mantel tests of the matrix correspondence (Legendre & Legendre 2012) were used to define whether floristic differences were correlated with Landsat

and environmental distances. The Pearson correlation method was used in the test with 999 permutations. Partial Mantel tests were used to assess residual correlations after controlling for the effect of geographic distances.

In **IV**, the floristic patterns of trees across Peruvian Amazonia were modelled using random forest (RF) regressions. Random forest is a machine learning algorithm that allows non-linear relationships in the data with high expected predicted power (Kuhn & Johnson 2013). Each of the NMDS axes 1–3 was used as the response variable whereas the Landsat and environmental variables were used as the predictor variables. Different cross-validation techniques and a feature forward selection (ffs) method was used to select the best RF models, which were then applied over the whole Peruvian Amazonia in order to produce predictive maps of each of the main floristic gradients (NMDS axes 1–3). Such predictive maps were classified in 10 classes using a k-means clustering method. An indicator analysis (De Cáceres et al. 2012) was performed based on indicator values (Dufrêne & Legendre 1997) to evaluate which taxa were associated to each of the 10 classes in **IV** and to bamboo and non-bamboo forests in **III**.

## 3 Results and Discussion

### 3.1 Average reflectance values of Landsat imagery are good predictors to model the distribution of trees at local extents

In **I** and **II**, I found that it is possible to use average values derived from Landsat imagery to model the distribution of tree taxa at local extents in Peruvian Amazonia. One of the main findings in **I** was that the use of averaging filters applied to the Landsat data consistently increased the SDMs performance in all tree taxa. It is important to note that species-specific spectral patterns are not being measured but rather the average spectral canopy signal as a measure of environmental patterns or habitat type. The size of a pixel in Landsat data is comparable to the crown of a single large canopy tree, but the spectral values obtained per pixel can be a mixed signal of different canopy crowns and canopy structure and shaded trees, since landscape heterogeneity does not follow pixel limits. By applying averaging filters, it is possible to suppress high-frequency variation, reduce the effects of local noise but retain landscape variation (Jones & Vaughan 2010). Furthermore, earlier studies in Amazonia have found that average filtering improves the correspondence between pixel data and floristic data (Rajaniemi et al. 2005; Salovaara et al. 2005; Thessler et al. 2005). Additionally, using averaging filters increase the likelihood that the field observation (coordinates of each tree) comes from the area covered in the spectral data. This is particularly relevant since the tree coordinates at the base of the trunk can be located several meters away from the actual position of the tree canopy, and the coordinates also have an associated spatial error.

A reason why averaging filters improved model performance is probably related to fact the average reflectance values of Landsat bands are strongly correlated both to soil properties (Higgins et al. 2012; Sirén et al. 2013; Van doninck & Tuomisto 2018) and also floristic patterns of understory plants (Tuomisto, Poulsen, et al. 2003; Tuomisto, Ruokolainen, et al. 2003; Salovaara et al. 2005; Higgins et al. 2012; Muro et al. 2016; Tuomisto et al. 2019) and canopy trees [**III**, **IV**]. Importantly, neither soils nor understory plants are directly visible to the satellite, so the predictive power comes from the fact that soils affect such properties of the forest canopy that in turn affect reflectance. These include species composition, tree architecture, leaf

mesophyll structure and leaf chemistry. As suggested by some studies, the inclusion of continuous remote sensing variables in SDM should be justified and they should characterize the potential habitat (Bradley et al. 2012; He et al. 2015; Leitão & Santos 2019). The spatially filtered average reflectance values do not reveal the occurrence of individual tree species but can serve as indicators of variation in ecologically relevant habitat conditions, which makes them ecologically informative layers for modelling the distribution of species.

### 3.2 Elevation improves distribution models of trees at local extents

Another important finding from **I** and **II** was that elevation is an important contributor to SDM performance of canopy trees in Southern Peruvian Amazonia. In **I**, the inclusion of elevation consistently increased the SDMs performance of all tree taxa and was considered one of the most contributing predictors. In **II**, elevation was the most contributing factor for all the taxa. It is important to point out that, since more field data became available in **II**, the elevational range covered by the training data naturally increased. Nevertheless, the elevational range both in **I** and **II** was too small (between 250 and 400 meters) to be related to other relevant environmental gradients, such as climate. At least one study modelled the distribution of trees at local extents using different topographic variables (Figueiredo et al. 2015) and found that elevation also contributed the most in the SDMs.

Elevation is an indirect environmental variable, meaning that it does not directly affect the performance of biological organisms. Nevertheless, elevation can be used as a proxy for other environmental variables that do affect species occurrences and are correlated with it. Elevation is generally related to soil drainage and water availability, since water tends to accumulate into depressions while soils at higher topographical positions remain drier. In addition, in Amazonia several studies have found that elevation is related to geological substrates and soil nutrient concentrations (Vormisto et al. 2004; Costa et al. 2005; Higgins et al. 2011). Soils are relevant factors for modelling the distribution of plant species in Amazonia (Figueiredo et al. 2018; Zuquim et al. 2020), but accurate soil data are often not available at local extents. Even though soil maps are available (Hengl et al. 2014; Hengl et al. 2017), they have been found to contain major inaccuracies in Amazonia (Moulatlet et al. 2017).

Topographic variables derived from elevation have been found to structure tree species distributions (Zuleta et al. 2018) and elevation has been found to be the most important predictor when modelling the distribution of canopy trees in Amazonia at local extents (Figueiredo et al. 2015). Other topographic variables, derived from elevation, are related to hydrology and sedimentation (Wittmann et al. 2004) and

more subtle changes in drainage can also be related to topography (Vormisto et al. 2004; Moulatlet et al. 2014) although slope is not necessarily a strong surrogate of soil moisture and drainage (Zuquim et al. 2009). In agreement with this, height above the nearest drainage (HAND) and slope contributed little in modelling the distribution of tree species in Amazonia at local extents (Figueiredo et al. 2015). Nevertheless, topographic variables such as HAND could indeed be used as surrogates of hydrological conditions in other study areas.

### 3.3 Training species distribution models from elongated areas derive more consistent models and spatial predictions

Since the field data used for the SDMs were highly concentrated in the study areas of **I** and **II**, the predictions made for areas close to the training data can be expected to be more accurate than those made for areas further away. Since more data had become available, an obvious question in **II** was to assess to what degree SDM performance and spatial predictions differed between the models that were trained using different configurations of the occurrence data. An important finding was that more elongated configurations of the training area often derived better and more robust SDMs, since they were more representative of the available environmental space and they often covered a larger proportion of it. For instance, the standard deviation of model performance (AUC) derived from training data occurring in more elongated configurations was half compared to models that use training data from less elongated or compact configurations. Another important result was that using a larger number of occurrences (4-fold) provided better distribution models for all tree taxa, which agrees with several previous studies (Stockwell & Peterson 2002; Hernandez et al. 2006; Wisz et al. 2008; Mateo et al. 2010; van Proosdij et al. 2016). For instance, model performance increased by 20% for some taxa when using four times more occurrence data for model training.

The predicted distributions using different spatial configurations of the training occurrences were rather stable for some taxa but in others, the spatial variation among models was higher. In **II**, a map of spatial variation is reported showing areas where there is more uncertainty associated with each taxon, which has been previously suggested in SDMs (Rocchini et al. 2011). Both the average suitability values and the spatial predictions differed when using different spatial arrangements of the training area. These findings address the importance of taking into consideration the spatial arrangement of the training area when using geographically highly concentrated occurrence data for modelling the distribution of trees at local extents.

Hopefully, forest management practitioners will use this framework in the future to estimate the distribution of important taxa within their management units even before forest census campaigns and for forest planning activities. If this practice is eventually adopted, having estimates of SDM performance variation, such as the standard deviation, would allow assessing uncertainty in the distribution potential of the tree taxa. Similarly, if SDMs are adopted for tropical forest management, using tree occurrence data from more elongated training areas will provide most robust results and species distribution estimates.

### 3.4 Floristic patterns of trees are consistent at different taxonomic resolutions

A key finding from **III** was the consistency of floristic patterns using different taxonomic cleaning strategies. By coarsening the taxonomic resolution to the genus level, it is possible to include a higher proportion of the available dataset but at the expense of blurring the relationship between species preferences and environmental gradients. Nevertheless, some studies have found a high degree of congruence between genus-level and species-level data (Higgins & Ruokolainen 2004; Cayuela et al. 2011; **III**). Additionally, excluding unidentified stems could include serious bias caused by separate identifiers if identifications are not cross-checked. Moreover, including only taxonomically consistent identifications could reduce such biases but at the expense of considerably reducing the usage of species and stems, and hence causing additional uncertainty in the analyses of species composition and its relationship with environmental gradients (Cayuela et al. 2011).

The floristic consistency found at different resolutions agrees with some studies that found a high degree of congruence between genus-level and species-level data (Higgins & Ruokolainen 2004; Honorio Coronado et al. 2009; Cayuela et al. 2011). This is particularly practical for ecological studies since it is possible to derive consistent floristic patterns of trees using legacy forestry inventory data without a detailed and time-consuming cross-checking of species-level identifications. In **III**, 97% of the original stems in the forest inventory database were used in the analyses since they were at least identified to the genus level, whereas only 56% were identified to the species level. By using genus-level data, it is possible to retain most of the database and at the same time derive consistent floristic results. This served as basis for analyzing genus-level patterns of trees in **IV**, in which 80% of the stems in the forest inventory database were identified to the genus level, whereas less than 45% were identified to the species level.

It is important to consider the proportion of monospecific genera in the forest inventory datasets, which might contribute to explaining the floristic similarities at different taxonomic resolutions. For instance, in **III** and **IV** forest inventory datasets,

44% and 38% of the genera, respectively, were monospecific. If a high proportion of taxa is monospecific in the inventory dataset, those taxa will behave in exactly the same manner as when they would be considered at the species level, and therefore the floristic patterns will remain similar when compared at species or genus-level resolution.

### 3.5 Floristic dissimilarities of tree communities are strongly correlated with spectral distances

In **III** and **IV**, floristic dissimilarities of tree communities were strongly correlated with differences in the spectral values derived from the Landsat data but also with other environmental layers. In particular, floristic patterns of trees were mainly correlated with the spectral values of the short-wave infrared spectrum of Landsat imagery (bands 5 and 7 from Landsat TM/ETM+ sensors). Previous studies have found that floristic similarity patterns of understory plants in Amazonia are also strongly correlated with the spectral values of red, infrared and short-wave infrared bands of the Landsat satellite (Tuomisto, Poulsen, et al. 2003; Tuomisto, Ruokolainen, et al. 2003; Higgins et al. 2012). It is also known that floristic patterns of understory plants and trees tend to be congruent in Amazonia (Tuomisto et al. 1995; Ruokolainen et al. 2007; Higgins et al. 2011; Tuomisto et al. 2016) so this reaffirms the potential of using Landsat bands to predict floristic dissimilarity patterns in Amazonia.

Interestingly, differences in the spectral values of Landsat bands had congruent roles in explaining the floristic dissimilarities at different geographic extents. For instance, in **IV** the geographic extent was more than 20 times bigger than in **III**, and still in both areas, Landsat bands explained over 40% of the floristic dissimilarities. In **III**, at least two forest types (bamboo and non-bamboo forests) were identified whereas in **IV**, at least “tierra firme” forests, bamboo forests, hydromorphic forests, montane forests and seasonally inundated forests. This suggests that we can consistently use differences in the reflectance values of the near-infrared spectrum to predict floristic dissimilarities at different extents and different forest types in Amazonia.

In **III**, floristic dissimilarity patterns were assessed at different taxonomic cleaning strategies, including genus-level and species-level resolution. An important finding was that the correlation between floristic dissimilarity and spectral distances (differences in reflectance values) was higher when using genus-level data than when using species-level data. This pinpoints another advantage of lumping the inventory data to the genus level (Section 3.4), since more ecologically meaningful results could be obtained when including more genus-level data than when excluding incompletely identified stems from the analyses.

### 3.6 Bamboo forests are floristically different from adjacent non-bamboo forests

In **III**, the focus was on assessing floristic differences between two adjacent forest types: bamboo and non-bamboo forests. Such floristic differences were assessed at different tree size classes and taxonomic cleaning strategies (genus, identified species and consistent species). One of the main findings was that bamboo-dominated forests were floristically different from adjacent non-bamboo forests, independent of the tree size class and the taxonomic resolution. These floristic differences between both forest types are congruent with previous findings in Peruvian Amazonia (Griscom et al. 2007) but here such differences were determined using forestry inventory data, which focuses mainly of the timber potential.

Since the forest inventory data used in **III** came from permanent production forests, selective logging for timber production occurred in the study area at least since 2002. It is known that logging activities may favor the establishment of non-commercial species and can affect forest structure and the tree community (Rockwell et al. 2014; Cazzolla Gatti et al. 2015). Nevertheless, since we found floristic differences also in a tree size class not directly affected by logging, the general floristic patterns documented are unlikely to be dominated by selective logging activities.

Another interesting floristic finding was the association of different palm taxa with bamboo and non-bamboo forests. In **III**, *Socratea exorrhiza* (Mart.) H.Wendl. was strongly associated with bamboo-dominated forests whereas *Iriartea deltoideia* Ruiz & Pav. was mainly associated with adjacent non-bamboo forests. The preferences of *Socratea* with poorly drained soils (Pacheco 2001) supports its association with bamboo-dominated forests (Griscom et al. 2007). Even though the causal effect of such associations (whether it is directly related to the presence-absence of bamboo or to soil characteristics) remains unclear, this finding confirms the association of certain palm taxa with each forest type.

### 3.7 Floristic patterns of trees in Peruvian Amazonia are correlated with reflectance values derived from Landsat imagery and with other environmental layers

Floristic patterns of trees at the genus level were strongly correlated with Landsat reflectance values and, to a lesser extent, with other environmental variables such as climate layers across Peruvian Amazonia. For instance, in a floristic ordination in **IV**, the first floristic axis was strongly correlated with the canopy reflectance values of all Landsat bands, especially within the short-wave infrared and near-infrared spectrum. The secondary axes were mainly correlated with climate layers. Since



canopy reflectance values have been found to predict soil properties rather well (Sirén et al. 2013; Van doninck & Tuomisto 2018), it is likely that the first floristic axis in **IV** is related to soil characteristics. Floristic patterns of trees were also correlated with different environmental variables, such as soil, topography, and climate, which agrees with many previous floristic studies of trees in Amazonia (Phillips et al. 2003; ter Steege et al. 2006; Ruokolainen et al. 2007; Toledo et al. 2011; Toledo et al. 2012; Baldeck et al. 2013; Baldeck et al. 2016).

Based on the correlates of floristic patterns of trees with Landsat and environmental layers, it was possible to build predictive models of the main floristic gradients of trees using random forest algorithm and apply them over the whole Peruvian Amazonia. In **IV**, the best models explained more than 60% of the variation along the first and second floristic gradients and 40% of the variation along the third floristic axis, which partially agrees with a similar approach used to map geocological patterns as indicated by fern species composition across all Amazonia (Tuomisto et al. 2019). Soil, climate and Landsat layers had congruent roles in predicting floristic spatial patterns even though the target groups (ferns at the species level and trees at the genus level) were phylogenetically distant, occupy different forest layers and were identified to different taxonomic resolutions. Furthermore, other studies have already found congruent floristic patterns between trees and understory plants in Amazonia and Central America (Tuomisto et al. 1995; Ruokolainen et al. 2007; Higgins et al. 2011; Jones et al. 2013; Tuomisto et al. 2016).

For the first time, in **IV** it was possible to visualize the predicted floristic variation of trees at the genus level in Peruvian Amazonia. The spatially predicted floristic patterns of trees are congruent with previous ecosystem and vegetation classifications (MINAM 2016b; MINAM 2019) and trait diversity in Peruvian Amazonia (Asner, Martin, Tupayachi, et al. 2017). Based on such spatial predictions, it was possible to discriminate between known ecosystems such as hydromorphic forest in Northern Peruvian Amazonia, bamboo forests in Southern Peruvian Amazonia and mountain tropical forests following the Andes. Having a map of predicted community composition is useful for reducing research gaps in a still unknown area like Peruvian Amazonia and it is useful also for identifying suitable areas for conservation purposes.

### 3.8 Future directions

Throughout this dissertation I explored the potential of Landsat imagery (**I**; **II**, **III** and **IV**), which has a global coverage at a medium 30-m resolution and a 16-day revisit time. Landsat has been also collecting information for several decades already, which enhances the possibilities of using it for forest monitoring. Different satellite sensors provide appropriate data for different purposes depending of the

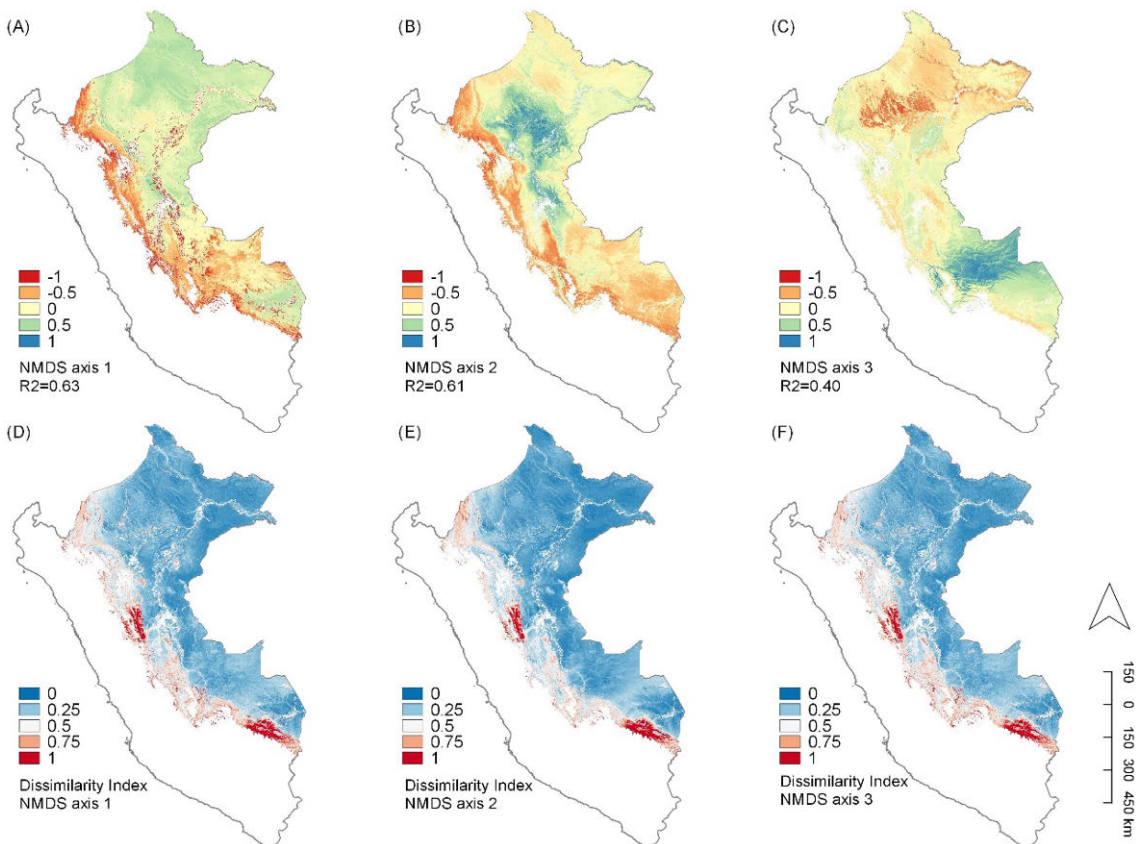
target object and the spatial, temporal and radiometric characteristics of the sensor (Wang et al. 2010). At broad extents, such as Peruvian Amazonia, analyzing 30-m resolution satellite imagery is already computationally challenging. Using higher resolution satellite products would allow identifying smaller objects but would be computationally even more complex. Obtaining reflectance values at different and narrower spectral range (higher spectral resolution) would allow a better spectral discrimination between different objects and surfaces. An obvious source of information worth further exploring is Landsat 8 satellite imagery as well as Sentinel-2A. Sentinel-2A, from the European Space Agency (ESA), is operative since 2015 and records information at 13 spectral bands (443–2190 nm) and at a spatial resolution of 10 m (for four visible and near-infrared bands), 20 m (six red edge and shortwave infrared bands) and 60 m (three atmospheric correction bands).

The spectral values used in the different analyses throughout the presented chapters were a spatial average (**I** and **II**) or median measure of canopy reflectance (**III** and **IV**). Average spectral values in **I** and **II** were proved to be good predictors for modelling the distribution of trees at local extents, since they measure habitat conditions and are good proxies of soil characteristics. In **III** and **IV**, the median reflectance values were strongly correlated with dissimilarity patterns of trees. Nevertheless, I would like to further explore other spectral measures, mainly related to spectral heterogeneity, such as the distance to the spectral centroid (Palmer et al. 2002; Rocchini et al. 2007; Rocchini et al. 2010), the variability in the reflectance values among pixels (Gillespie et al. 2008; Wang et al. 2018; Labiberté et al. 2020) and Rao's Q diversity index (Rocchini et al. 2017). It would also be interesting to explore the relationships between biodiversity patterns or species distributions and different measures of geodiversity (Record et al. 2020), including measures of topographic heterogeneity such as variance or fine scale texture of elevation. Similarly, I would further explore predicting not only variation in community composition (beta diversity) and floristic dissimilarities but also local species richness (alpha diversity) and spectral diversity.

The research conducted in **IV** was based on one of many campaigns of the national forest and wildlife inventory program in Peruvian Amazonia. The forest inventory data used in **IV** consisted of 157 inventory plots which were part of the first panel of a five-stage inventory program. The whole inventory program is planned to eventually include over 1800 inventory plots throughout Peru. Therefore, more inventory field data will become available in the near future, across different ecosystems in Peru, including Peruvian Amazonia. Additional information could be used as independent test data what would enable further validating and improving the current map of predicted floristic variation of trees in Peruvian Amazonia.

Even though the random forest regression in **IV** explained most of the variation of the floristic gradients, an estimate of spatial uncertainty would allow a better-

quality assessment of the predicted floristic maps. Recent existing frameworks allow quantifying spatial variation of uncertainty in random forest predictions (Meinshausen 2006; Wager et al. 2014; Coulston et al. 2016; Mentch & Hooker 2016; Meyer & Pebesma 2020) and hence, it would be recommended to address this important issue in future research. One of these frameworks allows estimating the area to which a prediction model can be reliably applied (Meyer & Pebesma 2020). This method, called “area of applicability” (AOA), derives a “dissimilarity index” (DI) that is based on the minimum distance to the training data in the predictor space. The following figure shows an example of the AOA method applied to the spatial random forest predictions on the floristic ordination axes (NMDS 1,2 & 3) in chapter IV.



**Figure 3.** Spatial patterns in the predicted community composition of tree genera across Peruvian Amazonia. (A, B and C) represent the predicted NMDS ordination axes 1, 2 and 3 respectively. (D, E, F) represent the dissimilarity index (Meyer & Pebesma 2020) for each of the NMDS predictions. Red colors in panels D, E and F show areas that are too different from the data used for model training to provide reliable results and, hence, model predictions should not be made for those areas.

## 4 Conclusions

Average reflectance values derived from Landsat imagery are good ecologically relevant predictors to model the distribution of tree taxa at local extents in Peruvian Amazonia. The average spectral values derived from Landsat imagery reflect habitat conditions and their underlying properties such as soil characteristics, and not species-specific spectral patterns. Including elevation as an additional variable consistently improves model performance when modelling the distribution of trees at local extents. Having maps of predicted distribution of tree taxa at local extents could help forest management practitioners to estimate the distribution of important tree taxa within their management areas even before collecting field data and even before requesting an area as a forest concession. This modelling framework could also be adopted by government forest institutions aiming at estimating timber potential as well as prioritizing or delimiting conservation or management areas for endangered tree species.

It is possible to use differences in the reflectance values of the near-infrared and short-wave infrared spectrum of Landsat imagery to consistently predict changes in the floristic composition of trees at different extents and different forest types in Amazonia. Floristic patterns of trees were strongly correlated with Landsat reflectance values and, to a lesser degree, to other environmental variables such as climate layers across Peruvian Amazonia. Based on those remote sensing and environmental correlates it was possible to derive the first map of predicted floristic variation of trees throughout all Peruvian Amazonia. This map offers a unique opportunity for government forest institutions to identify and prioritize suitable areas for conservation purposes.

Forest inventory and census data are sources of information of paramount importance for ecological studies since they cover vast areas from remote places. Combining legacy forestry data with available remote sensing and environmental layers offers an efficient way of mapping continuous biodiversity patterns, such as species distributions and floristic patterns, over large areas in Amazonia. Mapping tree species distributions and floristic patterns at different geographical extents are useful for practical applications such as characterizing biodiversity patterns and

forest resources in areas where field work is yet missing and identifying suitable areas for forest management and conservation purposes.

# 5 Acknowledgements

First of all, I would like to thank my main supervisor Hanna Tuomisto and my supervisors Kalle Ruokolainen and Risto Kalliola, as well as all my colleagues that are part of the Amazon Research Team of the University of Turku for the continuous technical support throughout my research. During the past 4 years I have learned a lot from the Amazon Research Team, which have not only enhanced my scientific and technical skills but have also made me feel like home in Turku.

I would also like to express my sincere gratitude to the doctoral programme in Biology, Geography and Geology (BGG), which granted me a 3-year funded position in the University of Turku. This funded position enabled me to focus continuously on my research. I feel really lucky to have received such funding.

Special thanks to the Peruvian National Forest and Wildlife Service (SERFOR) for sharing the forest inventory datasets from Peruvian Amazonia and to Consolidado Otorongo Forest Concession for sharing the forest census datasets used throughout this dissertation. This research would not have been possible if such forest institutions and companies would have not been interested in contributing to increasing research in the forest sector in Peruvian Amazonia. Thanks also to all the people involved in planning and collecting the forest inventory data from the National Forest and Wildlife Inventory (INFFS) and the regional forestry inventory in permanent production forests (IBPP) and the forestry census data from Consolidado Otorongo forest concession.

A big part of this research was computationally intensive, and many analyses could not have been possible without the use of super computers. Thanks to the IT Center for Science in Finland (CSC Finland) for giving such technological support and resources.

Thanks to all the co-authors that were involved in my dissertation: Elvira Gómez, Gabriela Zuquim, Hanna Tuomisto, Jasper Van doninck, Kalle Ruokolainen, Natalia Reategui and Risto Kalliola.

I was really lucky and honored to receive feedback from top quality researchers as pre-examiners of this dissertation. Thanks to Bruce Nelson and Sassan Saatchi for all the detailed comments and suggestions. I was able to improve many parts of this

dissertation and I also received insightful recommendations for future research in the use of remote sensing.

Thanks to Kalle Ruokolainen and Risto Kalliola for translating the abstract of this dissertation to Finnish.

Finally, I would like to thank my family and all my friends for the continuous emotional support and caring throughout all these years. To my Peruvian family, in particular to my parents Eduardo and Marifer. They have continuously fought for giving me and my siblings all the best within their reach. Due to different situations throughout my life I had the opportunity of having more than just two parents. I would like to thank Peteto and Patty for taking me as their own son for many years as well as my grandparents Fernando and Elizabeth. I would like to thank also my family in Finland, specially to Mirka for her continuous support during the last years.

# References

- Aguirre-Gutiérrez, J., Carvalheiro, L.G., Polce, C., Loon, E.E. van, Raes, N., Reemer, M., & Biesmeijer, J.C. 2013. Fit-for-Purpose: Species Distribution Model Performance Depends on Evaluation Criteria – Dutch Hoverflies as a Case Study. *PLOS ONE* 8: e63708.
- Araújo, M.B., Anderson, R.P., Barbosa, A.M., Beale, C.M., Dormann, C.F., Early, R., Garcia, R.A., Guisan, A., Maiorano, L., Naimi, B., O'Hara, R.B., Zimmermann, N.E., & Rahbek, C. 2019. Standards for distribution models in biodiversity assessments. *Science Advances* 5: eaat4858.
- Asner, G.P., Anderson, C.B., Martin, R.E., Knapp, D.E., Tupayachi, R., Sinca, F., & Malhi, Y. 2014. Landscape-scale changes in forest structure and functional traits along an Andes-to-Amazon elevation gradient. *Biogeosciences* 11: 843–856.
- Asner, G.P., & Martin, R.E. 2009. Airborne spectrometry: mapping canopy chemical and taxonomic diversity in tropical forests. *Frontiers in Ecology and the Environment* 7: 269–276.
- Asner, G.P., Martin, R.E., Anderson, C.B., & Knapp, D.E. 2015. Quantifying forest canopy traits: Imaging spectroscopy versus field survey. *Remote Sensing of Environment* 158: 15–27.
- Asner, G.P., Martin, R.E., Ford, A.J., Metcalfe, D.J., & Liddell, M.J. 2009. Leaf chemical and spectral diversity in Australian tropical forests. *Ecological Applications* 19: 236–253.
- Asner, G.P., Martin, R.E., Knapp, D.E., Tupayachi, R., Anderson, C.B., Sinca, F., Vaughn, N.R., & Llactayo, W. 2017. Airborne laser-guided imaging spectroscopy to map forest trait diversity and guide conservation. *Science* 355: 385–389.
- Asner, G.P., Martin, R.E., Tupayachi, R., & Llactayo, W. 2017. Conservation assessment of the Peruvian Andes and Amazon based on mapped forest functional diversity. *Biological Conservation* 210, Part A: 80–88.
- Austin, M.P., & Smith, T.M. 1989. A new model for the continuum concept.
- Baldeck, C.A., & Asner, G.P. 2013. Estimating Vegetation Beta Diversity from Airborne Imaging Spectroscopy and Unsupervised Clustering. *Remote Sensing* 5: 2057–2071.
- Baldeck, C.A., Colgan, M.S., Féret, J.-B., Levick, S.R., Martin, R.E., & Asner, G.P. 2014. Landscape-scale variation in plant community composition of an African savanna from airborne species mapping. *Ecological Applications* 24: 84–93.
- Baldeck, C.A., Harms, K.E., Yavitt, J.B., John, R., Turner, B.L., Valencia, R., Navarrete, H., Davies, S.J., Chuyong, G.B., Kenfack, D., Thomas, D.W., Madawala, S., Gunatilleke, N., Gunatilleke, S., Bunyavejchewin, S., Kiratiprayoon, S., Yaacob, A., Supardi, M.N.N., & Dalling, J.W. 2013. Soil resources and topography shape local tree community structure in tropical forests. *Proceedings of the Royal Society of London B: Biological Sciences* 280: 1–7.
- Baldeck, C.A., Tupayachi, R., Sinca, F., Jaramillo, N., & Asner, G.P. 2016. Environmental drivers of tree community turnover in western Amazonian forests. *Ecography* 39: 1089–1099.
- Barbet-Massin, M., Rome, Q., Villemant, C., & Courchamp, F. 2018. Can species distribution models really predict the expansion of invasive species? *PLOS ONE* 13: e0193085.
- Barbosa, F.G., & Schneck, F. 2015. Characteristics of the top-cited papers in species distribution predictive models. *Ecological Modelling* 313: 77–83.



- Barve, N., Barve, V., Jiménez-Valverde, A., Lira-Noriega, A., Maher, S.P., Peterson, A.T., Soberón, J., & Villalobos, F. 2011. The crucial role of the accessible area in ecological niche modeling and species distribution modeling. *Ecological Modelling* 222: 1810–1819.
- Bradley, B.A., & Fleishman, E. 2008. Can remote sensing of land cover improve species distribution modelling? *Journal of Biogeography* 35: 1158–1159.
- Bradley, B.A., Olsson, A.D., Wang, O., Dickson, B.G., Pelech, L., Sesnie, S.E., & Zachmann, L.J. 2012. Species detection vs. habitat suitability: Are we biasing habitat suitability models with remotely sensed data? *Ecological Modelling* 244: 57–64.
- Cámara-Leret, R., Tuomisto, H., Ruokolainen, K., Balslev, H., & Munch Kristiansen, S. 2017. Modelling responses of western Amazonian palms to soil nutrients. *Journal of Ecology* 105: 367–381.
- Carvalho, A.L. de, Nelson, B.W., Bianchini, M.C., Plagnol, D., Kuplich, T.M., & Daly, D.C. 2013. Bamboo-Dominated Forests of the Southwest Amazon: Detection, Spatial Extent, Life Cycle Length and Flowering Waves. *PLOS ONE* 8: e54852.
- Cavender-Bares, J., Gamon, J.A., & Townsend, P.A. (Eds.). 2020. *Remote Sensing of Plant Biodiversity*. Springer International Publishing, Cham.
- Cayuela, L., de la Cruz, M., & Ruokolainen, K. 2011. A method to incorporate the effect of taxonomic uncertainty on multivariate analyses of ecological data. *Ecography* 34: 94–102.
- Cazzolla Gatti, R., Castaldi, S., Lindsell, J.A., Coomes, D.A., Marchetti, M., Maesano, M., Di Paola, A., Paparella, F., & Valentini, R. 2015. The impact of selective logging and clearcutting on forest structure, tree diversity and above-ground biomass of African tropical forests. *Ecological Research* 30: 119–132.
- Chapman, D., Pescott, O.L., Roy, H.E., & Tanner, R. 2019. Improving species distribution models for invasive non-native species with biologically informed pseudo-absence selection. *Journal of Biogeography* 46: 1029–1040.
- Chaves, P.P., Echeverri, N.R., Ruokolainen, K., Kalliola, R., Doninck, J.V., Rivero, E.G., Zuquim, G., & Tuomisto, H. 2020. Using forestry inventories and satellite imagery to assess floristic variation in bamboo-dominated forests in Peruvian Amazonia. *Journal of Vegetation Science* n/a:
- Chaves, P.P., Ruokolainen, K., & Tuomisto, H. 2018. Using remote sensing to model tree species distribution in Peruvian lowland Amazonia. *Biotropica* 50: 758–767.
- Cord, A.F., Meentemeyer, R.K., Leitão, P.J., & Václavík, T. 2013. Modelling species distributions with remote sensing data: bridging disciplinary perspectives (R. Whittaker, Ed.). *Journal of Biogeography* 40: 2226–2227.
- Costa, F.R.C., Magnusson, W.E., & Luizao, R.C. 2005. Mesoscale distribution patterns of Amazonian understorey herbs in relation to topography, soil and watersheds. *Journal of Ecology* 93: 863–878.
- Coulston, J.W., Blinn, C.E., Thomas, V.A., & Wynne, R.H. 2016. Approximating prediction uncertainty for random forest regression models. *Photogrammetric Engineering & Remote Sensing* 82: 189–197.
- De Cáceres, M., Legendre, P., Valencia, R., Cao, M., Chang, L.-W., Chuyong, G., Condit, R., Hao, Z., Hsieh, C.-F., Hubbell, S., Kenfack, D., Ma, K., Mi, X., Supardi Noor, Md.N., Kassim, A.R., Ren, H., Su, S.-H., Sun, I.-F., Thomas, D., Ye, W., & He, F. 2012. The variation of tree beta diversity across a global network of forest plots: Beta diversity across a network of forest plots. *Global Ecology and Biogeography* 21: 1191–1202.
- De'ath, G. 1999. Extended dissimilarity: a method of robust estimation of ecological distances from high beta diversity data. *Plant Ecology* 144: 191–199.
- Dufrêne, M., & Legendre, P. 1997. Species assemblages and indicator species: The need for a flexible asymmetrical approach. *Ecological Monographs* 67: 345–366.
- Elith, J., H. Graham, C., P. Anderson, R., Dudík, M., Ferrier, S., Guisan, A., J. Hijmans, R., Huettmann, F., R. Leathwick, J., Lehmann, A., Li, J., G. Lohmann, L., A. Loiselle, B., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., McC. M. Overton, J., Townsend Peterson, A., J. Phillips, S., Richardson, K., Scachetti-Pereira, R., E. Schapire, R., Soberón, J., Williams, S., S. Wisz, M., & E.

- Zimmermann, N. 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29: 129–151.
- van Ewijk, K.Y., Randin, C.F., Treitz, P.M., & Scott, N.A. 2014. Predicting fine-scale tree species abundance patterns using biotic variables derived from LiDAR and high spatial resolution imagery. *Remote Sensing of Environment* 150: 120–131.
- Feng, X., Park, D.S., Walker, C., Peterson, A.T., Merow, C., & Papeş, M. 2019. A checklist for maximizing reproducibility of ecological niche models. *Nature Ecology & Evolution* 3: 1382–1395.
- Féret, J.-B., & Asner, G.P. 2014. Mapping tropical forest canopy diversity using high-fidelity imaging spectroscopy. *Ecological Applications* 24: 1289–1296.
- Fick, S.E., & Hijmans, R.J. 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology* 37: 4302–4315.
- Figueiredo, S.M. de M., Venticinque, E.M., Figueiredo, E.O., & Ferreira, E.J.L. 2015. Predicting the distribution of forest tree species using topographic variables and vegetation index in eastern Acre, Brazil. *Acta Amazonica* 45: 167–174.
- Figueiredo, F.O.G., Zuquim, G., Tuomisto, H., Moulatlet, G.M., Balslev, H., & Costa, F.R.C. 2018. Beyond climate control on species range: The importance of soil data to predict distribution of Amazonian plant species. *Journal of Biogeography* 45: 190–200.
- Fortunel, C., Lasky, J.R., Uriarte, M., Valencia, R., Wright, S.J., Garwood, N.C., & Kraft, N.J.B. 2018. Topography and neighborhood crowding can interact to shape species growth and distribution in a diverse Amazonian forest. *Ecology* 99: 2272–2283.
- Franklin, J., & Miller, J.A. 2010. *Mapping species distributions: Spatial inference and prediction*. Cambridge University Press.
- Gillespie, T.W., Foody, G.M., Rocchini, D., Giorgi, A.P., & Saatchi, S. 2008. Measuring and modelling biodiversity from space. *Progress in Physical Geography* 32: 203–221.
- Giovanelli, J.G.R., de Siqueira, M.F., Haddad, C.F.B., & Alexandrino, J. 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.
- Griscom, B.W., Daly, D.C., & Ashton, M.S. 2007. Floristics of bamboo-dominated stands in lowland terra-firma forests of southwestern Amazonia. *The Journal of the Torrey Botanical Society* 134: 108–125.
- Guisan, A., Graham, C.H., Elith, J., Huettmann, F., & the NCEAS Species Distribution Modelling Group. 2007. Sensitivity of predictive species distribution models to change in grain size. *Diversity and Distributions* 13: 332–340.
- Guisan, A., & Thuiller, W. 2005. Predicting species distribution: offering more than simple habitat models. *Ecology Letters* 8: 993–1009.
- Guisan, A., Thuiller, W., & Zimmermann, N.E. 2017. *Habitat Suitability and Distribution Models: With Applications in R*. Cambridge University Press, Cambridge.
- Guisan, A., & Zimmermann, N.E. 2000. Predictive habitat distribution models in ecology. *Ecological Modelling* 135: 147–186.
- Guitet, S., Sabatier, D., Brunaux, O., Hérault, B., Aubry-Kientz, M., Molino, J.-F., & Baraloto, C. 2014. Estimating tropical tree diversity indices from forestry surveys: A method to integrate taxonomic uncertainty. *Forest Ecology and Management* 328: 270–281.
- Hallgren, W., Santana, F., Low-Choy, S., Zhao, Y., & Mackey, B. 2019. Species distribution models can be highly sensitive to algorithm configuration. *Ecological Modelling* 408: 108719.
- He, K.S., Bradley, B.A., Cord, A.F., Rocchini, D., Tuanmu, M.-N., Schmidtlein, S., Turner, W., Wegmann, M., & Pettorelli, N. 2015. Will remote sensing shape the next generation of species distribution models? *Remote Sensing in Ecology and Conservation* 1: 4–18.
- He, K.S., Zhang, J., & Zhang, Q. 2009. Linking variability in species composition and MODIS NDVI based on beta diversity measurements. *Acta Oecologica* 35: 14–21.

- Hengl, T., de Jesus, J.M., MacMillan, R.A., Batjes, N.H., Heuvelink, G.B.M., Ribeiro, E., Samuel-Rosa, A., Kempen, B., Leenaars, J.G.B., Walsh, M.G., & Gonzalez, M.R. 2014. SoilGrids1km — Global Soil Information Based on Automated Mapping. *PLoS ONE* 9: e105992.
- Hengl, T., Jesus, J.M. de, Heuvelink, G.B.M., Gonzalez, M.R., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M.N., Geng, X., Bauer-Marschallinger, B., Guevara, M.A., Vargas, R., MacMillan, R.A., Batjes, N.H., Leenaars, J.G.B., Ribeiro, E., Wheeler, I., Mantel, S., & Kempen, B. 2017. SoilGrids250m: Global gridded soil information based on machine learning. *PLOS ONE* 12: e0169748.
- Hernandez, P.A., Graham, C.H., Master, L.L., & Albert, D.L. 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography* 29: 773–785.
- Higgins, M.A., Asner, G.P., Perez, E., Elespuru, N., Tuomisto, H., Ruokolainen, K., & Alonso, A. 2012. Use of Landsat and SRTM Data to Detect Broad-Scale Biodiversity Patterns in Northwestern Amazonia. *Remote Sensing* 4: 2401–2418.
- Higgins, M.A., & Ruokolainen, K. 2004. Rapid Tropical Forest Inventory: a Comparison of Techniques Based on Inventory Data from Western Amazonia. *Conservation Biology* 18: 799–811.
- Higgins, M.A., Ruokolainen, K., Tuomisto, H., Llerena, N., Cardenas, G., Phillips, O.L., Vásquez, R., & Räsänen, M. 2011. Geological control of floristic composition in Amazonian forests. *Journal of Biogeography* 38: 2136–2149.
- Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., & Jarvis, A. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* 25: 1965–1978.
- Honorio Coronado, E.N., Baker, T.R., Phillips, O.L., Pitman, N.C.A., Pennington, R.T., Vásquez Martínez, R., Monteagudo, A., Mogollón, H., Dávila Cardozo, N., Ríos, M., García-Villacorta, R., Valderrama, E., Ahuite, M., Huamantupa, I., Neill, D.A., Laurance, W.F., Nascimento, H.E.M., Soares de Almeida, S., Killeen, T.J., Arroyo, L., Núñez, P., & Freitas Alvarado, L. 2009. Integrating regional and continental scale comparisons of tree composition in Amazonian terra firme forests. *Biogeosciences Discussions* 6: 1421–1451.
- Hopkins, M.J.G. 2007. Modelling the known and unknown plant biodiversity of the Amazon Basin. *Journal of Biogeography* 34: 1400–1411.
- Hutchinson, G.E. 1978. *An Introduction to Population Ecology*. Yale University Press.
- Hutchinson, G.E. 1957. Concluding remarks. *Cold Spring Harbor Symposia on Quantitative Biology* 22: 415–427.
- Jones, M.M., Ferrier, S., Condit, R., Manion, G., Aguilar, S., & Pérez, R. 2013. Strong congruence in tree and fern community turnover in response to soils and climate in central Panama (G. Zotz, Ed.). *Journal of Ecology* 101: 506–516.
- Jones, H.G., & Vaughan, R.A. 2010. *Remote Sensing of Vegetation: Principles, Techniques, and Applications*. Oxford University Press, Oxford ; New York.
- Karger, D.N., Conrad, O., Böhrner, J., Kawohl, T., Kreft, H., Soria-Auza, R.W., Zimmermann, N.E., Linder, H.P., & Kessler, M. 2017. Climatologies at high resolution for the earth's land surface areas. *Scientific Data* 4: 170122.
- Köhl, M., Magnussen, S.S., & Marchetti, M. 2006. *Sampling Methods, Remote Sensing and GIS Multiresource Forest Inventory*. Springer-Verlag, Berlin Heidelberg.
- Kristiansen, T., Svenning, J.-C., Eiserhardt, W.L., Pedersen, D., Brix, H., Munch Kristiansen, S., Knadel, M., Grández, C., & Balslev, H. 2012. Environment versus dispersal in the assembly of western Amazonian palm communities. *Journal of Biogeography* 39: 1318–1332.
- Kuhn, M., & Johnson, K. 2013. *Applied Predictive Modeling*. Springer-Verlag, New York.
- Laliberté, E., Schweiger, A.K., & Legendre, P. 2020. Partitioning plant spectral diversity into alpha and beta components. *Ecology Letters* 23: 370–380.
- Legendre, P., & Legendre, L. 2012. *Numerical ecology*. Elsevier, Amsterdam, Oxford.

- Leitão, P.J., & Santos, M.J. 2019. Improving Models of Species Ecological Niches: A Remote Sensing Overview. *Frontiers in Ecology and Evolution* 7.
- Leitão, P.J., Schwieder, M., Suess, S., Catry, I., Milton, E.J., Moreira, F., Osborne, P.E., Pinto, M.J., van der Linden, S., & Hostert, P. 2015. Mapping beta diversity from space: Sparse Generalised Dissimilarity Modelling (SGDM) for analysing high-dimensional data. *Methods in Ecology and Evolution* 6: 764–771.
- Ley Forestal y de Fauna Silvestre. *SINIA | Sistema Nacional de Información Ambiental*. MAAP #122: Amazon Deforestation 2019. 2020. *MAAP*.
- MAAP #124: Deforestation Hotspots 2020 in the Peruvian Amazon. 2020. *MAAP*.
- Mateo, R.G., Felicísimo, Á.M., & Muñoz, J. 2010. Effects of the number of presences on reliability and stability of MARS species distribution models: the importance of regional niche variation and ecological heterogeneity. *Journal of Vegetation Science* 21: 908–922.
- Maximiano, M.F. de A., d’Horta, F.M., Tuomisto, H., Zuquim, G., Doninck, J.V., & Ribas, C.C. 2020. The relative role of rivers, environmental heterogeneity and species traits in driving compositional changes in southeastern Amazonian bird assemblages. *Biotropica* 52: 946–962.
- Meinshausen, N. 2006. Quantile Regression Forests. *Journal of Machine Learning Research* 7: 983–999.
- Mentch, L., & Hooker, G. 2016. Quantifying Uncertainty in Random Forests via Confidence Intervals and Hypothesis Tests. *Journal of Machine Learning Research* 17: 1–41.
- Merckx, B., Steyaert, M., Vanreusel, A., Vincx, M., & Vanaverbeke, J. 2011. Null models reveal preferential sampling, spatial autocorrelation and overfitting in habitat suitability modelling. *ECOLOGICAL MODELLING* 222: 588–597.
- Merow Cory, Silander John A., & Warton David. 2014. A comparison of Maxlike and Maxent for modelling species distributions. *Methods in Ecology and Evolution* 5: 215–225.
- Merow, C., Smith, M.J., Edwards, T.C., Guisan, A., McMahon, S.M., Normand, S., Thuiller, W., Wüest, R.O., Zimmermann, N.E., & Elith, J. 2014. What do we gain from simplicity versus complexity in species distribution models? *Ecography*. doi: 10.1111/ecog.00845
- Meyer, H., & Pebesma, E. 2020. Predicting into unknown space? Estimating the area of applicability of spatial prediction models. *arXiv:2005.07939 [cs, stat]*
- MINAM. 2016a. Mapa de propuesta técnica de límites Amazónicos.
- MINAM. 2016b. Mapa Nacional de Cobertura Vegetal (Memoria descriptiva). *SINIA | Sistema Nacional de Información Ambiental*.
- MINAM. 2019. Mapa Nacional de Ecosistemas del Perú. *SINIA | Sistema Nacional de Información Ambiental*.
- Moulatlet, G.M., Costa, F.R.C., Rennó, C.D., Emilio, T., & Schietti, J. 2014. Local Hydrological Conditions Explain Floristic Composition in Lowland Amazonian Forests. *Biotropica* 46: 395–403.
- Moulatlet, G.M., Zuquim, G., Figueiredo, F.O.G., Lehtonen, S., Emilio, T., Ruokolainen, K., & Tuomisto, H. 2017. Using digital soil maps to infer edaphic affinities of plant species in Amazonia: Problems and prospects. *Ecology and Evolution* 7: 8463–8477.
- Muro, J., Van doninck, J., Tuomisto, H., Higgins, M.A., Moulatlet, G.M., & Ruokolainen, K. 2016. Floristic composition and across-track reflectance gradient in Landsat images over Amazonian forests. *ISPRS Journal of Photogrammetry and Remote Sensing* 119: 361–372.
- Nagendra, H. 2001. Using remote sensing to assess biodiversity. *International Journal of Remote Sensing* 22: 2377–2400.
- Nagendra, H., & Rocchini, D. 2008. High resolution satellite imagery for tropical biodiversity studies: the devil is in the detail. *Biodiversity and Conservation* 17: 3431.
- Nagendra, H., Rocchini, D., Ghate, R., Sharma, B., & Pareeth, S. 2010. Assessing Plant Diversity in a Dry Tropical Forest: Comparing the Utility of Landsat and Ikonos Satellite Images. *Remote Sensing* 2: 478–496.

- Oke, O.A., & Thompson, K.A. 2015. Distribution models for mountain plant species: The value of elevation. *Ecological Modelling* 301: 72–77.
- OSINFOR. 2016. Modelamiento de la Distribución Potencial de 18 Especies Forestales en el Departamento de Loreto.
- OSINFOR. 2013. Modelamiento espacial de nichos ecológicos para la evaluación de presencia de especies forestales maderales en la Amazonia Peruana.
- Pacheco, M.A.W. 2001. Effects of Flooding and Herbivores on Variation in Recruitment of Palms between Habitats. *Journal of Ecology* 89: 358–366.
- Palmer, M.W., Earls, P.G., Hoagland, B.W., White, P.S., & Wohlgemuth, T. 2002. Quantitative tools for perfecting species lists. *Environmetrics* 13: 121–137.
- Pearson, R.G., & Dawson, T.P. 2003. Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? *Global Ecology and Biogeography* 12: 361–371.
- Peterson, A.T., & Soberón, J. 2012. Species Distribution Modeling and Ecological Niche Modeling: Getting the Concepts Right. *Natureza & Conservação* 10: 102–107.
- Peterson, A.T., Soberón, J., Pearson, R.G., Anderson, R.P., Martínez-Meyer, E., Nakamura, M., & Araújo, M.B. 2011. *Ecological Niches and Geographic Distributions (MPB-49)*. Princeton University Press.
- Phillips, S.J., Anderson, R.P., Dudík, M., Schapire, R.E., & Blair, M.E. 2017. Opening the black box: an open-source release of Maxent. *Ecography* 40: 887–893.
- Phillips, S.J., Anderson, R.P., & Schapire, R.E. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190: 231–259.
- Phillips, O.L., Vargas, P.N., Monteagudo, A.L., Cruz, A.P., Zans, M.-E.C., Sánchez, W.G., Yli-Halla, M., & Rose, S. 2003. Habitat association among Amazonian tree species: a landscape-scale approach. *Journal of Ecology* 91: 757–775.
- Prates-Clark, C.D.C., Saatchi, S.S., & Agosti, D. 2008. Predicting geographical distribution models of high-value timber trees in the Amazon Basin using remotely sensed data. *Ecological Modelling* 211: 309–323.
- van Proosdij, A.S.J., Sosef, M.S.M., Wieringa, J.J., & Raes, N. 2016. Minimum required number of specimen records to develop accurate species distribution models. *Ecography* 39: 542–552.
- Raghavan, R.K., Barker, S.C., Cobos, M.E., Barker, D., Teo, E.J.M., Foley, D.H., Nakao, R., Lawrence, K., Heath, A.C.G., & Peterson, A.T. 2019. Potential Spatial Distribution of the Newly Introduced Long-horned Tick, *Haemaphysalis longicornis* in North America. *Scientific Reports* 9: 498.
- Rajaniemi, S., Tomppo, E., Ruokolainen, K., & Tuomisto, H. 2005. Estimating and mapping pteridophyte and Melastomataceae species richness in western Amazonian rainforests. *International Journal of Remote Sensing* 26: 475–793.
- Record, S., Dahlin, K.M., Zarnetske, P.L., Read, Q.D., Malone, S.L., Gaddis, K.D., Grady, J.M., Costanza, J., Hobi, M.L., Latimer, A.M., Pau, S., Wilson, A.M., Ollinger, S.V., Finley, A.O., & Hestir, E. 2020. Remote Sensing of Geodiversity as a Link to Biodiversity. In Cavender-Bares, J., Gamon, J.A., & Townsend, P.A. (eds.), *Remote Sensing of Plant Biodiversity*, pp. 225–253. Springer International Publishing, Cham.
- Rocchini, D. 2007. Effects of spatial and spectral resolution in estimating ecosystem  $\alpha$ -diversity by satellite imagery. *Remote Sensing of Environment* 111: 423–434.
- Rocchini, D., Andreini Butini, S., & Chiarucci, A. 2005. Maximizing plant species inventory efficiency by means of remotely sensed spectral distances: Maximizing plant species inventorying by remote sensing. *Global Ecology and Biogeography* 14: 431–437.
- Rocchini, D., Balkenhol, N., Carter, G.A., Foody, G.M., Gillespie, T.W., He, K.S., Kark, S., Levin, N., Lucas, K., Luoto, M., Nagendra, H., Oldeland, J., Ricotta, C., Southworth, J., & Neteler, M. 2010. Remotely sensed spectral heterogeneity as a proxy of species diversity: Recent advances and open challenges. *Ecological Informatics* 5: 318–329.

- Rocchini, D., Boyd, D.S., Féret, J.-B., Foody, G.M., He, K.S., Lausch, A., Nagendra, H., Wegmann, M., & Pettorelli, N. 2016. Satellite remote sensing to monitor species diversity: potential and pitfalls. *Remote Sensing in Ecology and Conservation* 2: 25–36.
- Rocchini, D., & Cade, B.S. 2008. Quantile Regression Applied to Spectral Distance Decay. *IEEE Geoscience and Remote Sensing Letters* 5: 640–643.
- Rocchini, D., Chiarucci, A., & Loiselle, S.A. 2004. Testing the spectral variation hypothesis by using satellite multispectral images. *Acta Oecologica* 26: 117–120.
- Rocchini, D., He, K.S., & Zhang, J. 2009. Is spectral distance a proxy of beta diversity at different taxonomic ranks? A test using quantile regression. *Ecological Informatics* 4: 254–259.
- Rocchini, D., Hortal, J., Lengyel, S., Lobo, J.M., Jiménez-Valverde, A., Ricotta, C., Bacaro, G., & Chiarucci, A. 2011. Accounting for uncertainty when mapping species distributions: The need for maps of ignorance. *Progress in Physical Geography: Earth and Environment* 35: 211–226.
- Rocchini, D., Luque, S., Pettorelli, N., Bastin, L., Doktor, D., Faedi, N., Feilhauer, H., Féret, J.-B., Foody, G.M., Gavish, Y., Godinho, S., Kunin, W.E., Lausch, A., Leitão, P.J., Marcantonio, M., Neteler, M., Ricotta, C., Schmidtlein, S., Vihervaara, P., Wegmann, M., & Nagendra, H. 2018. Measuring  $\beta$ -diversity by remote sensing: A challenge for biodiversity monitoring. *Methods in Ecology and Evolution* 9: 1787–1798.
- Rocchini, D., Marcantonio, M., & Ricotta, C. 2017. Measuring Rao's Q diversity index from remote sensing: An open source solution. *Ecological Indicators* 72: 234–238.
- Rocchini, D., Ricotta, C., & Chiarucci, A. 2007. Using satellite imagery to assess plant species richness: The role of multispectral systems. *Applied Vegetation Science* 10: 325–331.
- Rockwell, C.A., Kainer, K.A., d'Oliveira, M.V.N., Staudhammer, C.L., & Baraloto, C. 2014. Logging in bamboo-dominated forests in southwestern Amazonia: Caveats and opportunities for smallholder forest management. *Forest Ecology and Management* 315: 202–210.
- Rodríguez-Soto, C., Monroy-Vilchis, O., Maiorano, L., Boitani, L., Faller, J.C., Briones, M.Á., Núñez, R., Rosas-Rosas, O., Ceballos, G., & Falcucci, A. 2011. Predicting potential distribution of the jaguar (*Panthera onca*) in Mexico: identification of priority areas for conservation. *Diversity and Distributions* 17: 350–361.
- Ruokolainen, K., Tuomisto, H., Macía, M.J., Higgins, M.A., & Yli-Halla, M. 2007. Are floristic and edaphic patterns in Amazonian rain forests congruent for trees, pteridophytes and Melastomataceae? *Journal of Tropical Ecology* 23: 13–25.
- Saatchi, S., Buermann, W., ter Steege, H., Mori, S., & Smith, T.B. 2008. Modeling distribution of Amazonian tree species and diversity using remote sensing measurements. *Remote Sensing of Environment* 112: 2000–2017.
- Saatchi, S., Malhi, Y., Zutta, B., Buermann, W., Anderson, L.O., Araujo, A.M., Phillips, O.L., Peacock, J., ter Steege, H., Lopez Gonzalez, G., Baker, T., Arroyo, L., Almeida, S., Higuchi, N., Killeen, T., Monteagudo, A., Neill, D., Pitman, N., Prieto, A., Salomão, R., Silva, N., Vásquez Martínez, R., Laurance, W., & Ramírez, H.A. 2009. Mapping landscape scale variations of forest structure, biomass, and productivity in Amazonia. *Biogeosciences Discuss.* 6: 5461–5505.
- Salovaara, K.J., Thessler, S., Malik, R.N., & Tuomisto, H. 2005. Classification of Amazonian primary rain forest vegetation using Landsat ETM+ satellite imagery. *Remote Sensing of Environment* 97: 39–51.
- Schmidtlein, S., & Fassnacht, F.E. 2017. The spectral variability hypothesis does not hold across landscapes. *Remote Sensing of Environment* 192: 114–125.
- Schulman, L., Ruokolainen, K., Junikka, L., Sääksjärvi, I.E., Salo, M., Juvonen, S.-K., Salo, J., & Higgins, M. 2007. Amazonian biodiversity and protected areas: do they meet? *Biodiversity and Conservation* 16: 3011–3051.
- Sirén, A., Tuomisto, H., & Navarrete, H. 2013. Mapping environmental variation in lowland Amazonian rainforests using remote sensing and floristic data. *International Journal of Remote Sensing* 34: 1561–1575.

- Soberón, J., & Nakamura, M. 2009. Niches and distributional areas: Concepts, methods, and assumptions. *Proceedings of the National Academy of Sciences* 106: 19644–19650.
- Soria-Auza, R.W., Kessler, M., Bach, K., Barajas-Barbosa, P.M., Lehnert, M., Herzog, S.K., & Böhner, J. 2010. Impact of the quality of climate models for modelling species occurrences in countries with poor climatic documentation: a case study from Bolivia. *Ecological Modelling* 221: 1221–1229.
- Sousa-Baena, M.S., Garcia, L.C., & Peterson, A.T. 2014. Completeness of digital accessible knowledge of the plants of Brazil and priorities for survey and inventory. *Diversity and Distributions* 20: 369–381.
- Steege, H.T. 1998. The use of forest inventory data for a National Protected Area Strategy in Guyana. *Biodiversity & Conservation* 7: 1457–1483.
- ter Steege, H., Pitman, N.C.A., Phillips, O.L., Chave, J., Sabatier, D., Duque, A., Molino, J.-F., Prévost, M.-F., Spichiger, R., Castellanos, H., Hildebrand, P. von, & Vásquez, R. 2006. Continental-scale patterns of canopy tree composition and function across Amazonia. *Nature* 443: 444–447.
- Stockwell, D.R.B., & Peterson, A.T. 2002. Effects of sample size on accuracy of species distribution models. *Ecological Modelling* 148: 1–13.
- Thessler, S., Ruokolainen, K., Tuomisto, H., & Tomppo, E. 2005. Mapping gradual landscape-scale floristic changes in Amazonian primary rain forests by combining ordination and remote sensing. *Global Ecology and Biogeography* 14: 315–325.
- Thuiller, W., Araújo, M.B., & Lavorel, S. 2003. Generalized models vs. classification tree analysis: Predicting spatial distributions of plant species at different scales. *Journal of Vegetation Science* 14: 669–680.
- Toledo, M., Peña-Claros, M., Bongers, F., Alarcón, A., Balcázar, J., Chuviña, J., Leño, C., Licona, J.C., & Poorter, L. 2012. Distribution patterns of tropical woody species in response to climatic and edaphic gradients. *Journal of Ecology* 100: 253–263.
- Toledo, M., Poorter, L., Peña-Claros, M., Alarcón, A., Balcázar, J., Chuviña, J., Leño, C., Licona, J.C., ter Steege, H., & Bongers, F. 2011. Patterns and Determinants of Floristic Variation across Lowland Forests of Bolivia. *Biotropica* 43: 405–413.
- Tuomisto, H., Moulatlet, G.M., Balslev, H., Emilio, T., Figueiredo, F.O.G., Pedersen, D., & Ruokolainen, K. 2016. A compositional turnover zone of biogeographical magnitude within lowland Amazonia. *Journal of Biogeography* 43: 2400–2411.
- Tuomisto, H., Poulsen, A.D., Ruokolainen, K., Moran, R.C., Quintana, C., Celi, J., & Cañas, G. 2003. Linking floristic patterns with soil heterogeneity and satellite imagery in Ecuadorian Amazonia. *Ecological Applications* 13: 352–371.
- Tuomisto, H., Ruokolainen, K., Aguilar, M., & Sarmiento, A. 2003. Floristic patterns along a 43-km long transect in an Amazonian rain forest. *Journal of Ecology* 91: 743–756.
- Tuomisto, H., Ruokolainen, K., Kalliola, R., Linna, A., Danjoy, W., & Rodriguez, Z. 1995. Dissecting Amazonian Biodiversity. *Science* 269: 63–66.
- Tuomisto, H., Ruokolainen, L., & Ruokolainen, K. 2012. Modelling niche and neutral dynamics: on the ecological interpretation of variation partitioning results. *Ecography* 35: 961–971.
- Tuomisto, H., Van doninck, J., Ruokolainen, K., Moulatlet, G.M., Figueiredo, F.O.G., Sirén, A., Cárdenas, G., Lehtonen, S., & Zuquim, G. 2019. Discovering floristic and geoeological gradients across Amazonia. *Journal of Biogeography* 46: 1734–1748.
- Turner, W. 2014. Sensing biodiversity. *Science* 346: 301–302.
- Van doninck, J., Jones, M.M., Zuquim, G., Ruokolainen, K., Moulatlet, G.M., Sirén, A., Cárdenas, G., Lehtonen, S., & Tuomisto, H. 2020. Multispectral canopy reflectance improves spatial distribution models of Amazonian understory species. *Ecography* 43: 128–137.
- Van doninck, J., & Tuomisto, H. 2018. A Landsat composite covering all Amazonia for applications in ecology and conservation. *Remote Sensing in Ecology and Conservation* 4: 197–210.

- Van doninck, J., & Tuomisto, H. 2015. BRDF correction for Landsat TM/ETM+ data over Amazonian forests. In *Proceedings of the 2nd EARSeL International Workshop on Temporal Analysis of Satellite Images, 17.–19.6.2015, Stockholm, Sweden*, pp. 349–354.
- Van doninck, J., & Tuomisto, H. 2017a. Evaluation of directional normalization methods for Landsat TM/ETM+ over primary Amazonian lowland forests. *International Journal of Applied Earth Observation and Geoinformation* 58: 249–263.
- Van doninck, J., & Tuomisto, H. 2017b. Influence of Compositing Criterion and Data Availability on Pixel-Based Landsat TM/ETM+ Image Compositing Over Amazonian Forests. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 10: 857–867.
- Van doninck, J., Westerholm, J., Ruokolainen, K., Tuomisto, H., & Kalliola, R. 2020. Dating flowering cycles of Amazonian bamboo-dominated forests by supervised Landsat time series segmentation. *International Journal of Applied Earth Observation and Geoinformation* 93: 102196.
- Villar Hernández, B. de J., & Pérez Elizalde, S. 2015. MODELOS BAYESIANOS PARA MODELAR DISTRIBUCIONES DE ESPECIES CON REGISTROS DE SOLO PRESENCIAS. *Tropical and Subtropical Agroecosystems* 18:.
- Vormisto, J., Tuomisto, H., & Oksanen, J. 2004. Palm distribution patterns in Amazonian rainforests: What is the role of topographic variation? *Journal of Vegetation Science* 15: 485–494.
- Wager, S., Hastie, T., & Efron, B. 2014. Confidence Intervals for Random Forests: The Jackknife and the Infinitesimal Jackknife. *Journal of Machine Learning Research* 15: 1625–1651.
- Wang, K., Franklin, S.E., Guo, X., & Cattet, M. 2010. Remote Sensing of Ecology, Biodiversity and Conservation: A Review from the Perspective of Remote Sensing Specialists. *Sensors* 10: 9647–9667.
- Wang, Z., Lu, C., & Yang, X. 2018. Exponentially sampling scale parameters for the efficient segmentation of remote-sensing images. *International Journal of Remote Sensing* 39: 1628–1654.
- Willis, K.J., & Whittaker, R.J. 2002. Species Diversity--Scale Matters. *Science* 295: 1245–1248.
- Wisz, M.S., Hijmans, R.J., Li, J., Peterson, A.T., Graham, C.H., Guisan, A., & NCEAS Predicting Species Distributions Working Group. 2008. Effects of sample size on the performance of species distribution models. *Diversity and Distributions* 14: 763–773.
- Wittmann, F., Junk, W.J., & Piedade, M.T.F. 2004. The várzea forests in Amazonia: flooding and the highly dynamic geomorphology interact with natural forest succession. *Forest Ecology and Management* 196: 199–212.
- Zambrano, A.M.A., Broadbent, E.N., Schmink, M., Perz, S.G., & Asner, G.P. 2010. Deforestation drivers in Southwest Amazonia: Comparing smallholder farmers in Iñapari, Peru, and Assis Brasil, Brazil. *Conservation and Society* 8: 157.
- Zuleta, D., Russo, S.E., Barona, A., Barreto-Silva, J.S., Cardenas, D., Castaño, N., Davies, S.J., Detto, M., Sua, S., Turner, B.L., & Duque, A. 2018. Importance of topography for tree species habitat distributions in a terra firme forest in the Colombian Amazon. *Plant and Soil*. doi: 10.1007/s11104-018-3878-0
- Zuquim, G., Costa, F.R.C., Prado, J., & Braga-Neto, R. 2009. Distribution of pteridophyte communities along environmental gradients in Central Amazonia, Brazil. *Biodiversity and Conservation* 18: 151–166.
- Zuquim, G., Costa, F.R.C., Tuomisto, H., Moulatlet, G.M., & Figueiredo, F.O.G. 2020. The importance of soils in predicting the future of plant habitat suitability in a tropical forest. *Plant and Soil* 450: 151–170.
- Zuquim, G., Tuomisto, H., Costa, F.R.C., Prado, J., Magnusson, W.E., Pimentel, T., Braga-Neto, R., & Figueiredo, F.O.G. 2012. Broad Scale Distribution of Ferns and Lycophytes along Environmental Gradients in Central and Northern Amazonia, Brazil. *Biotropica* 44: 752–762.
- Zuquim, G., Tuomisto, H., Jones, M.M., Prado, J., Figueiredo, F.O.G., Moulatlet, G.M., Costa, F.R.C., Quesada, C.A., & Emilio, T. 2014. Predicting environmental gradients with fern species composition in Brazilian Amazonia. *Journal of Vegetation Science* 25: 1195–1207.



- Zurell, D., Franklin, J., König, C., Bouchet, P.J., Dormann, C.F., Elith, J., Fandos, G., Feng, X., Guillera-Arroita, G., Guisan, A., Lahoz-Monfort, J.J., Leitão, P.J., Park, D.S., Peterson, A.T., Rapacciuolo, G., Schmatz, D.R., Schröder, B., Serra-Diaz, J.M., Thuiller, W., Yates, K.L., Zimmermann, N.E., & Merow, C. A standard protocol for reporting species distribution models. *Ecography* n/a:



**UNIVERSITY  
OF TURKU**

ISBN 978-951-29-8289-9 (PRINT)  
ISBN 978-951-29-8290-5 (PDF)  
ISSN 0082-6979 (Print)  
ISSN 2343-3183 (Online)